PÉCSI TUDOMÁNYEGYETEM
EGÉSZSÉGTUDOMÁNYI KAR

ÁCS PONGRÁC

SPORTTUDOMÁNYI KUTATÁSOK MÓDSZERTANA

Pécsi Tudományegyetem EGÉSZSÉGTUDOMÁNYI Kar

Fizioterápiás- és Sporttudományi Intézet

SPORTTUDOMÁNYI KUTATÁSOK MÓDSZERTANA

ÁCS Pongrác

Pécs, 2015

PÉCSI TUDOMÁNYEGYETEM
UNIVERSITY OF PÉCS

SPORTTUDOMÁNYI KUTATÁSOK MÓDSZERTANA

Szerkesztette: Ács Pongrác

KIADJA A

PÉCSI TUDOMÁNYEGYETEM

EGÉSZSÉGTUDOMÁNYI KAR

Az első kiadást lektorálták:

Dr. habil. Rétsági Erzsébet

Dr. Herman Sándor

Dr. habil. Rappai Gábor

A bővített és átdolgozott kiadást lektorálta:

Dr. habil Ihász Ferenc

Címlapterv és szerkesztés:

Varga Gábor

Második, bővített kiadás

ISBN 978-963-642-881-5

A kézikönyv a TÁMOP-4.1.2. E-15/1/KONV-2015-0003. cím projekt keretében készült

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE

TARTALOMJEGYZÉK

ELŐSZÓ

A hazai sportszakember-képzésben az összes felsőoktatási intézmény régóta kiemelt helyen kezeli a sporttudományi kutatásokhoz kapcsolódó, Bevezetés a sporttudományos kutatásba című kurzust. Maga a tantárgy nagy hagyományokkal rendelkezik, hiszen az 1940-es évek elejétől a mostani Semmelweis Egyetem Testnevelési és Sporttudományi Karának (TF) jogelődjén már oktatták, Tudományos kutatás alapjai néven és egészen biztosan a mostani önálló Testnevelési Egyetemen is oktatásra kerül. Elmondható, hogy a tantárgy az utóbbi időkben még inkább fontos helyet foglal el a sportszakember képzésben. A „*Bologna-folyamat*” következtében átalakuló magyar felsőoktatási rendszer a sporttudományi képzési területet is érintette, s minden hazai sporttudományt oktató intézményben kötelező, alapozó lett ez a tantárgy. Megtalálható volt a Testnevelő-edző, valamint a Sportszervező (Sportmenedzser), illetve Rekreáció képzésben, a BSc és MSc szinten egyaránt és kikerülhetetlen lesz a képesítési jegyzékről szóló új kormányrendelet szerint kialakításra kerülő alap és mesterképzések vonatkozásában is. Vélelmezzük, hogy a jövőben leginkább az MSc képzésben fog még hangsúlyosabban megjelenik a kutatói és vezetői profil, mely nem nélkülözheti ezt a fajta elméleti és gyakorlati tudást.

A sportban megjelenő teljesítmény orientáció egyre jobban a sporttudományi kutatások fontosságára hívja fel a figyelmet, hiszen e kutatások nélkül a mai világban már „világraszóló” sporteredmények nem érhetők el. Emellett a növekvő tendenciát mutató szabadidő megjelenése, eltöltése, az egészséges életmóddal foglakozó kutatások elméleti és gyakorlati fontosságára hívják fel a figyelmet.

Kétségtelen tény, hogy az egyes területek, sportágak eredményei szinte „kimeríthetetlen kincsesbányái” a kutatóknak, valamint a sport aktív résztvevőinek, a versenyzőknek, az edzőknek és a menedzsereknek egyaránt. A rendelkezésre álló és elérhető gazdag adatbázis felveti annak szükségességét, hogy az adatokban rejlő információkat vizsgáljuk és elemezzük, és az eredményeket, következtetéseket kiadjuk, publikáljuk.

Egészen biztosan kijelenthetjük, hogy az elmúlt négy évben a sportot stratégiai ágazatként kezeli a kormányzat, így ennek tükrében számos a sport társadalmi, jogi és gazdasági környezetét befolyásoló változás történt. Gondoljunk csak a mindennapos testnevelés megjelenésére, vagy a sportfinanszírozást alapjaiban megváltoztató Tao megjelenésére. Mindezen intézkedések tovagyűrűző hatással, multiplikatív tényezőként fejtik ki hatásukat.

Mindezen fontos tények tükrében 2008-ban gondoltuk először, hogy fontos lenne egy olyan tankönyv megírása, mely az elméleti kutatás-módszertani alapokon túl a hallgatóknak gyakorlatban is használható segítséget tud nyújtani a tudományos igényű munkáik elkészítése során. Elmondható, hogy az azóta eltelt évek során megtapasztaltuk a tankönyv gyakorlati oktatásban betöltött szerepét és hasznát, valamint láthatóvá váltak mindazon részek is, melyek a hallgatók számára nehézséget okoznak. Azt gondoltuk, hogy ezeket a gyakorlati alkalmazás során felgyülemlett tapasztalatokat, illetve a statisztikai adatelemzéshez leggyakrabban alkalmazott informatikai szoftverek (Excel, SPSS) változásai is okot adnak arra, hogy a 2008-ban megírt tankönyvünket átírjuk, és bővítsük. A hallgatói kéréseket is integrálva egy teljesen új feladatgyűjteménnyel is bővítsük, hiszen ennek segítségével gyakorolhatnak és tudásuk szintjét is önállóan ellenőrizhetik. A második átdolgozott, bővített kiadás a SPSS program 22, illetve az Excel 2010-es verziójának felhasználásával készült.

A két nagyobb egységre tagolódó könyv a sporttudományi kutatáshoz kapcsolódó BSc és MSc képzésben megjelenő tantárgyak teljes tananyagát foglalja magába, illetve hasznára válhat a doktori képzést folyatató hallgatók számára is. Ugyanakkor haszonnal forgathatják azok is, akik más képzési területen tanulnak, ám tantervük előírja, hogy önálló kutatásokat végezzenek. A tankönyv felépítésében igyekeztünk az egyszerűbb módszertani ismeretektől a bonyolultabbak felé haladni, így biztosítva az érthetőség, gyakorolhatóság megtartását. Törekedtünk arra, hogy az eljárások egymásra épüljenek, így a kutatásokkal most ismerkedő olvasó is könnyen szerezhet érthető, „kézzel fogható” új ismereteket. Az életszerű, való életből hozott példák segítségével, a gyakorló szakemberek számára is hasznos lehet módszertani tankönyvünk.

A tankönyv szerkezeti felépítését átgondoltnak ítéljük, melyet a hallgatói visszajelzések is megerősítettek, így abban érdemi változtatásokat nem eszközöltünk. Mindenképpen fontosnak ítéltük azt, hogy az első könyvhöz képest az adatbázisok számát redukáljuk, és lehetőség szerint minimális számú a valós életből hozott adatbázissal szemléltessünk. Ezt szem előtt tartva a tankönyv első részében egy általunk az egyetemi populáción felmért adatbázist használtunk fel, melynek alapját képezték a Magyar Diáksport Szövetség által az általános iskolai mintára kiválóan meghatározott és jól alkalmazható fittségi tesztek, próbák. Fontosnak ítéljük azt, hogy egyfajta mindenki számára érthető, egységes mérőeszközökön definiált adatbázis segítségével mutassuk be a statisztika adatelemzéseket, hiszen így akár a gyakorló szakemberek a saját méréseik nyomán gyakorolhatják ezen statisztikai módszereket.

A tankönyv első fejezetében bemutatjuk a tudomány, illetve a sporttudomány meghatározásait, illetve a tudományok rendszerében való elhelyezését, ezzel keretet adunk a könyvnek. Kitérünk a sporttudomány tudománnyá válásának folyamatára, történetére, megismertetjük az olvasót a sporttudományos kutatás alapmodelljével is.

A második fejezet első részében bemutatjuk a kutatási terv készítésének menetét, foglalkozunk a témaválasztással, az irodalmazással, a kutatási hipotézisekkel, a vizsgálati minta meghatározásával, valamint a kutatás lefolytatásához szükséges kellékékkel is. Erősen támaszkodtunk oktatási tapasztalatainkra, melynek bizonyítékaként szólunk a hivatkozások, irodalomjegyzékek elkészítésének szakszerű módjáról, melyet a mai egyetemi és főiskolai hallgatók körében feltűnően problémás területnek ítélünk. Ebben a részben a kutatási eszközök megjelenítésénél alapos képet adunk a kérdőívek szerkesztéséről, napi aktuális sport példákon illusztrálva.

A fejezet második részében alaposan foglalkozunk az adatok feldolgozásával, a statisztikával, ahol igyekeztünk a leíró és következtetéses, valamint a sokváltozós statisztikai elemzések sokszínű világába bepillantást adni.

A számítástechnika fejlődése, a PC-k megjelenése a mindennapi életben azt eredményezte, hogy a hagyományos tollal papíron történő statisztikai számolások mellett a tudományos kutatómunka során szívesen alkalmazott számítógépes szoftverekkel (SPSS, Excel) is szemléltessük a tömegesen előforduló jelenségek mérését, leírását, modellezését. Itt felhívjuk a figyelmet arra, hogy a számítógépes példáknál alapfokú Excel ismeretekre támaszkodunk, ezt azért is tehetjük, mert ezzel a programmal már az általános iskolában megismerkednek a gyerekek, illetve a hallgatók gyakorlatilag 100%-ban hozzáférnek

A fejezet végén bemutatjuk azt is, hogy a legújabb kutatási eredmények közlése, illetve prezentációja milyen módon történjen.

A tankönyv megírásánál mindvégig törekedtünk az elmélet és a gyakorlat helyes arányának megválasztására. A módszertani fejtegetéseket próbáltuk közérthető sportpéldákkal, valós már publikált saját adatbázisokon „könnyebben emészthetővé” tenni. Az átdolgozott ás bővített tankönyvünkhöz szemben az első tankönyvvel nem tartozik DVD melléklet, hiszen a tankönyv, felhasznált adatbázisok, és az elektronikus munkafüzet is ingyenesen elérhető a Pécsi Tudományegyetem Egészségtudomány Kar honlapján (www. etk.pte.hu).

A szerző külön köszönettel tartozik kollégáinak a név szerint Harsányi László†, Prisztóka Gyöngyvér, Pótó Zsuzsanna és Kehl Dániel, akik tanácsaikkal, javaslataikkal, célravezető ötleteikkel segítették az első kiadás létrejöttét. Természetesen a második kiadás alkalmával több kollégától kapott a szerző támogatást melyért mérhetetlenül hálás. Feltétlenül szeretnénk kiemelni és megköszönni Betlehem Józsefnek, Oláh Andrásnak, Cselik Bencének és Varga Gábornak a támogatásukat és szakmai segítségüket.

Szeretnék köszönetet mondani kollégáimnak, munkatársaimnak, akik bátorítottak, és ötleteikkel inspirálták a könyv átdolgozását és bővítését és nem utolsó sorban hallgatóknak, akiktől számtalan javító szándékú megjegyzést kaptam. Köszönet illeti *Herman Sándor, Rappai Gábor, Rétsági Erzsébet, Ihász Ferenc* lektorok lelkiismeretes és alapos munkáját, akik olyan hiányosságokra hívták fel a figyelmemet, melyeknek kijavítása minden bizonnyal hasznára vált a tankönyvnek.

A szerző szeretettel ajánlja a tankönyvet, Dr. Farkas Ferenc TANÁR ÚRNAK, akinek a támogatását mind a sportolói, mind az egyetemi évei alatt megkapta.

Meg szeretném kérni Kollégáimat, Hallgatóimat, hogy észrevételeikkel, javaslataikkal támogassák a könyv folyamatos karbantartását, aktualizálást, tegyék lehetővé, hogy az esetleges bennmaradó (és csak a szerzőt terhelő) hibákat javíthassuk.

Végezetül remélem, hogy a tankönyv hozzájárulhat ahhoz, hogy nemzetünk ismét eredményesen sportoló nemzet lehessen, melynek elengedhetetlen és nélkülözhetetlen indikátorai a minőségi sporttudományi kutatások.

Kozármisleny, 2015. augusztus 7.

Ács Pongrác
szerző

1. TUDOMÁNY, A SPORTTUDOMÁNY HELYE A TUDOMÁNYOK RENDSZERÉBEN

„Csak az az igazi tudomány, amely világra szól; s ezért ha igazi tudósok és – amint kell – jó magyarok akarunk lenni, úgy a tudomány zászlóját olyan magasra kell emelnünk, hogy azt határainkon túl is meglássák és megadhassák neki az illő tiszteletet." (Eötvös Loránd)

A magyar nyelvhasználatban szinte mindennap találkozunk a tudomány kifejezésével, mára elengedhetetlen az élet területein a tudományos szemszögből való közelítés, a tudományos megalapozás. A tudomány egyidős az emberiséggel, hiszen a természet megismerését, és a törvényszerűségek összefüggéseit már az ősember is megfigyelte, gyűjtötte, tárolta (pl. barlangrajzok), és az így szerzett tudást felhasználta a közösség érdekében.

Ezt követően az új ismereteket homogén egységként a *filozófia* integrálta. Kezdetben a tudomány (filozófia) a megismerő tevékenység minden formáját „átkarolta", a vallásos tanokat, a mitológiát, a művészetet, a világnézeti gondolkodást, a tapasztalatokat, megfigyeléseket, elmélkedéseket. Gyakorlatilag az első filozófusok, ennek köszönhetően az első csillagászok, matematikusok, fizikusok voltak.

A tudomány, meghatározására álljon itt néhány példa:

„A természet, a társadalom és a gondolkodás objektív összefüggéseiről szerzett igazolható ismeretek rendszere."[1]

„A tudományon értjük a mai szinten: törvényszerűségek és összefüggések felderítésére, meghatározására igazolására irányuló tevékenységet, az igazolt ismeretek rendszerét, eme ismeretek tárolására, közzétételére, alkalmazására, valamint az egész tudomány irányítására szolgáló intézményeket és szerveket."[2]

„A tudomány valamely tárgyra vonatkozó tudásunknak rendszeres egésze; maguk az egyes igazságok, csak így rendszeresítve, egységbe kapcsolva adják a tudomány, összetartozó tárgyak körének megismerését. A rendszeres forma tehát lényeges követelmény minden

[1] Magyar Értelmező Kéziszótár (1973). 378.o.

[2] Hepp-Nádori (1971). 10. o.

tudományra nézve. Rendszeres forma alatt pedig értjük a tárgyra vonatkozó tételeknek objektív rendjét, melynél fogva a tételeket az őket megillető mellé- és alárendelésükben és teljességükben fölfogjuk és megértjük."[3]

„ A tudomány kísérletezik és felfedez, mér és megfigyel, elméleteket alkot, amelyek a dolgok hogyanját és miértjét magyarázzák. Technikai módszereket és szerszámokat gondol ki, javaslatokat vet fel és vet el. Hipotéziseket alkot és próbál ki, kérdésekkel fordul a természethez (tegyük hozzá- a társadalmi valósághoz), és kikényszeríti a választ, véleményt nyilvánít, cáfol, igazol és tagad, az igazságot megkülönbözteti a tévedéstől, az értelmeset az értelmetlentől, megmondja, hogyan jutunk el oda, ahová el akarunk jutni, hogyan kell csinálni azt, amit csinálni akarunk. A tudós olyan ember, mint bárki más, de különbözik is a többiektől, mivel tudja, hogy mindezt hogyan kell csinálni. Szigorúan tudományos kiképzést kapott, s makacs, magabiztos értelmes emberré vált… Mi több: a tudós azzal a ritka előjoggal rendelkezik, hogy a saját fejét használhatja, az önálló gondolkodás nagy és magányos művészetét gyakorolhatja. És mégis ahhoz az egyetemes közösséghez tartozik, amely egyetemes nyelven beszél."[4] (Wartofsky [1977] 13. old.)

A magyar nyelvben a *tudomány* szóra számos jól elkülöníthető jelentést, jelentéstartalmat találunk, melyeket a fenti tudományos meghatározásokban is láthattunk:

1. A világegyetem és saját magunk megismerésének egyik legfontosabb útját, mint cselekvési folyamatot, társadalmi tevékenységet, *a tudományos kutatást*, melyet jól meghatározható módszertan segítségével végeznek.
2. Tudományon érthetjük a tevékenységet végző emberek csoportját, a nemzetközi tudományos közösséget is. Ide sorolhatjuk mindazon elkötelezett személyeket, akik a saját szakterületükön tudományos kutatással foglalkoznak, illetve az eredményeiket hivatalos helyen közreadják, publikálják. Hazánkban ide tartoznak mindazon személyek, akik államilag meghatározott képzésen túl vannak, annak keretében felkészültségüket szakplénum előtt bizonyították, melynek pozitív hozadékaként az erre illetékes akkreditált testület számukra tudományos fokozatot ítélt. Magyarországon például ennek módja a felsőoktatási törvényben meghatározott doktori (Ph.D) eljárás, majd a fokozat egyetemi tanácsok általi odaítélése, valamint az államilag előírt doktori eskü és a hivatalos doktorrá avatási ceremónia.

[3] Pallas Nagy Lexikona. (http://www.mek.iif.hu/porta/szint/egyeb/lexikon/pallas)

[4] Szabó (2003)

3. Tudományon azonban leggyakrabban magát a produktumot értjük, melyet a már említett közösség hoz létre az emberiség számára. [5]

A tudományterületek és tudományágak meghatározása a (169/2000. [IX. 29.] és 154/2004. [X. 14.] Korm. Rendelet alapján):

1/1. táblázat

Tudományterületek		**Tudományágak (db)**
1.	Természettudományok	7
2.	Műszaki tudományok	11
3.	Orvostudományok	5
4.	Agrártudományok	6
5.	Társadalomtudományok	10 (Sporttudomány)
6.	Bölcsészettudományok	9
7.	Művészetek	7 (Művészeti ágak)
8.	Hittudomány	

Látható, hogy a kormányrendelet értelmében a **sporttudomány** a társadalomtudományok között szerepel másik kilenc tudományággal, melynek részletes felsorolását a következő táblázatban olvashatjuk:

1/2. táblázat

Társadalomtudományok	
1.	Gazdálkodás- és szervezéstudományok
2.	Közgazdaságtudományok
3.	Állam- és jogtudományok
4.	Szociológiai tudományok
5.	Pszichológiai tudományok
6.	Neveléstudományok
7.	***Sporttudományok***
8.	Politikatudományok
9.	Hadtudományok
10.	Multidiszciplináris társadalomtudományok

[5] http://hu.wikipedia.org

Amikor a sporttudományt kívánjuk meghatározni mindenképpen szem előtt kell tartani, hogy a sporttudomány tárgya nem más, mint *az ember egy speciális cselekvési körének vizsgálata, megismerése.*

„*A sporttudomány definíciója*: az emberi társadalom egyetemes kultúrájának részterületeként, a testkultúrának leképzésére szolgáló eszmerendszer – tudományosan igazolt, rendszerezett, általánosított elvek, tételek, törvények és törvényszerűségek, elméletek és módszerek együttese. Kutatási célja a társadalom testkulturális értékeinek (mint az egyetemes kultúra szubkultúrájának) gyarapítása, ezek segítségével az egyének, s ezen keresztül a társadalom totális fejlődésének elősegítése. A fizikai aktivitást tudatosan végző embernek, mint biológiai- pszichikai szociális egységnek a vizsgálata."(Bíróné N. E. 2004. 18. o.)[6].

Természetesen a sporttudományon belül a megismerésére való törekvés során is elkülönül a kétféle (Babbie, 2000. 33. o.) valóság. A *konszenzuális valóságról* beszélünk, amikor a dolgokat elfogadjuk, mivel az emberek többsége is elfogadja. Valami azért valóságos, mivel azt mondják. Csapatsportoknál elfogadjuk az edző szavait (pl.: a védők alacsonyak, fejjel gólveszélyesek leszünk), noha még soha nem játszottunk az ellenféllel. A *tapasztalati valóságnál* a megismerés a közvetlen tapasztalásból ered. A sport területén döntő fontosságú, hogy a váratlan, „megtapasztalt" helyzetekre, hogyan és milyen gyorsan reagálunk (pl.: a védők alacsonyak, de kétszer akkorát emelkednek, mint a csatáraink).

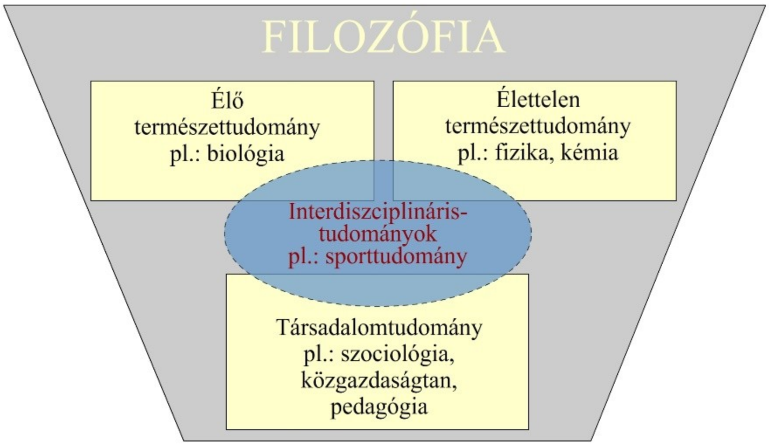
A ***testnevelés- és sporttudomány egy interdiszciplináris tudomány***, vagyis olyan határterületi tudományág, mely a két klasszikus tudományterület (*természet – és társadalomtudomány*) módszertani érintkezései mentén jött létre és egyszerre több tudományágat is közösen érint (1. ábra). A tudományterületeknél elkülönítjük az *élő- és élettelen* **természettudományokat**, illetve a **társadalomtudományokat**. A **természettudomány** az élő és élettelen természet objektumainak tanulmányozásával foglalkozó tudományágak gyűjtőneve. A **társadalomtudományok** az emberrel, mint társadalmi lénnyel foglalkoznak, valamint magát az ember által létrehozott társadalmat, és e kettő viszonyát tanulmányozzák.

„*A testnevelés egyike a legkomplexebb tudományoknak*. Alkotórészeként és horizontjában többé-kevésbé a természet és a társadalomtudományok legnagyobb része megtalálható. Véleményünk szerint, *a testnevelés-tudomány tartalmát és célkitűzéseit tekintve*

---

[6] Bíróné N. E. (2004). 18.o.

*társadalomtudomány, módszereiben pedig természettudomány*os jelleg az uralkodó. E kettőt szétválasztani persze igen nehéz és ez a szétválasztás inkább csak a fogalmak tisztázását szolgálja.” (Hepp F.- Nádori L. 1971. 20. o.)



1/1. ábra: A sporttudomány helye a tudományok rendszerében

Forrás: saját szerkesztés

1.1 A sporttudomány kialakulása

A sporttudomány intenzív fejlődésének kezdetét az 1950-es évek elejére tehetjük, a két szuperhatalom (Szovjetunió és az Egyesült Államok) élsportban való rivalizálásának köszönhetően. Ezt az időszakot megelőzően egészségügyi tudományos megfigyelésekkel találkozhatunk, melyekben a testnevelés és sportmozgások szervezetre kifejtett hatásait vizsgálták. A nagyhatalmak versenyében a kezdetektől megfigyelhető különbség abban volt, hogy a Szovjetunióban a sporttudományos kutatások szinte kivétel nélkül az élsport köré koncentrálódtak, addig az Egyesült Államokban az élsport mellett a rekráció, rehabiltáció és gyógytestneveléssel foglalkozó tudományos kutatások is kezdetüket vették. A nemzetközi elismertségét nagyban segítette, hogy az 1956 olimpiától kezdődően kísérőprogramként megjelennek a sporttudományos konferenciák. Az intézményrendszer kialakulásában elsőként 1928-ban megalakult a Nemzetközi Sportorvos Szövetség (FIMS), majd 1960-ban a Nemzetközi és Sporttudományi Tanács (ICSSPE).

A hazai sporttudomány fejlődését a Testnevelési Főiskola (TF) megalakulásához (1925) köthetjük, melyet követett az 1952-ben felállított Országos Testnevelési és Sporttudományi Intézet (OTSI). 1954-ben a TF-en megalakult a Testnevelési és Sporttudományi Tanács, majd 1959-ben létrehozták az önálló Testnevelési Tudományos Kutató Intézetet (TTKI), mely a 1969-től, mint TF Kutató Intézet működött tovább. Először 1975. szeptember 1-jétől a Magyar Népköztársaság Elnöki Tanácsa törvényerejű rendelete értelmében a Testnevelési

Főiskola, mint egyetemi jellegű főiskolaként működött, majd 1986-ban egyetemi státuszt kapott. A Magyar Népköztársaság Elnöki Tanácsa rendelete alapján 1989. szeptember 1-től a TF új elnevezése Magyar Testnevelési Egyetem. 1997-ben a Magyar Akkreditációs Bizottság elfogadta a Magyar Testnevelési Egyetem PhD. (Doctor of Philosophy) programját. Ezt követően a Magyar Tudományos Akadémia Orvostudományi Osztály Megelőző Orvostudományi Bizottságán belül megalakult a Sporttudományi Albizottság.[7] 2000-ben a fennállásának 75. évfordulóját ünneplő intézmény már a Semmelweis Egyetem Testnevelési és Sporttudományi Karaként (TF) szolgálja a magyar testnevelés és sport ügyét. 8

Újabb mérföldkő 2014 július 4-én, amikor az Országgyűlés döntött a nemzeti felsőoktatási törvény módosításáról és egy új intézmény létrejöttéről, vagyis önálló intézményként, Testnevelési Egyetem néven működik szeptember 1-től a Semmelweis Egyetemről leváló Testnevelési és Sporttudományi Kar.

A hazai sporttudomány meglétét számos érv támasztja alá:

- **Elfogadott felsőoktatási intézményrendszerek** (Budapest, Pécs, Szombathely, Eger, stb.)
- **Tudományos Társaság** (A hazai sporttudományt a Magyar Sporttudományi Társaság – *MSTT* – fogja össze, melyekben tudományos szakbizottságok működnek)
- **Tudományos folyóiratok** - Testnevelés (1928); Testneveléstudomány (1950); Sport és Tudomány (1956-1964); TF Tudományos Közlemények később Kalokagathia (1959); Testnevelés- és Sportegészségügyi Szemle, később Sportorvosi Szemle (1960); Sportélet (1965); A Sport és Testnevelés Időszerű Kérdései (1969-1982); Testnevelés Tanítása (1965); Mesteredző (1991); Sporttudomány (1998); Magyar Edző (1998) stb.-

A sporttudomány minden tekintetben megfelel a tudományok akadémiai rendszerébe való besorolhatóságának kritériumainak, hiszen:

- A sporttudománynak önálló tárgya van.
- A sporttudomány saját tudományos kutatási módszerekkel rendelkezik. A sporttudományok módszerei között jelentős átfedéseket, adaptációkat találunk más tudományok módszereiből.

[7] Harsányi L. (1998)

[8] Istvánfi Cs. (2000)

- A sporttudomány kialakította a saját terminológiáját, nyelvezetét, fogalmi rendszerét.
- A sporttudomány létrehozta és megfogalmazta sajátos elméleteit.
- A sporttudomány önálló intézményrendszerrel rendelkezik.

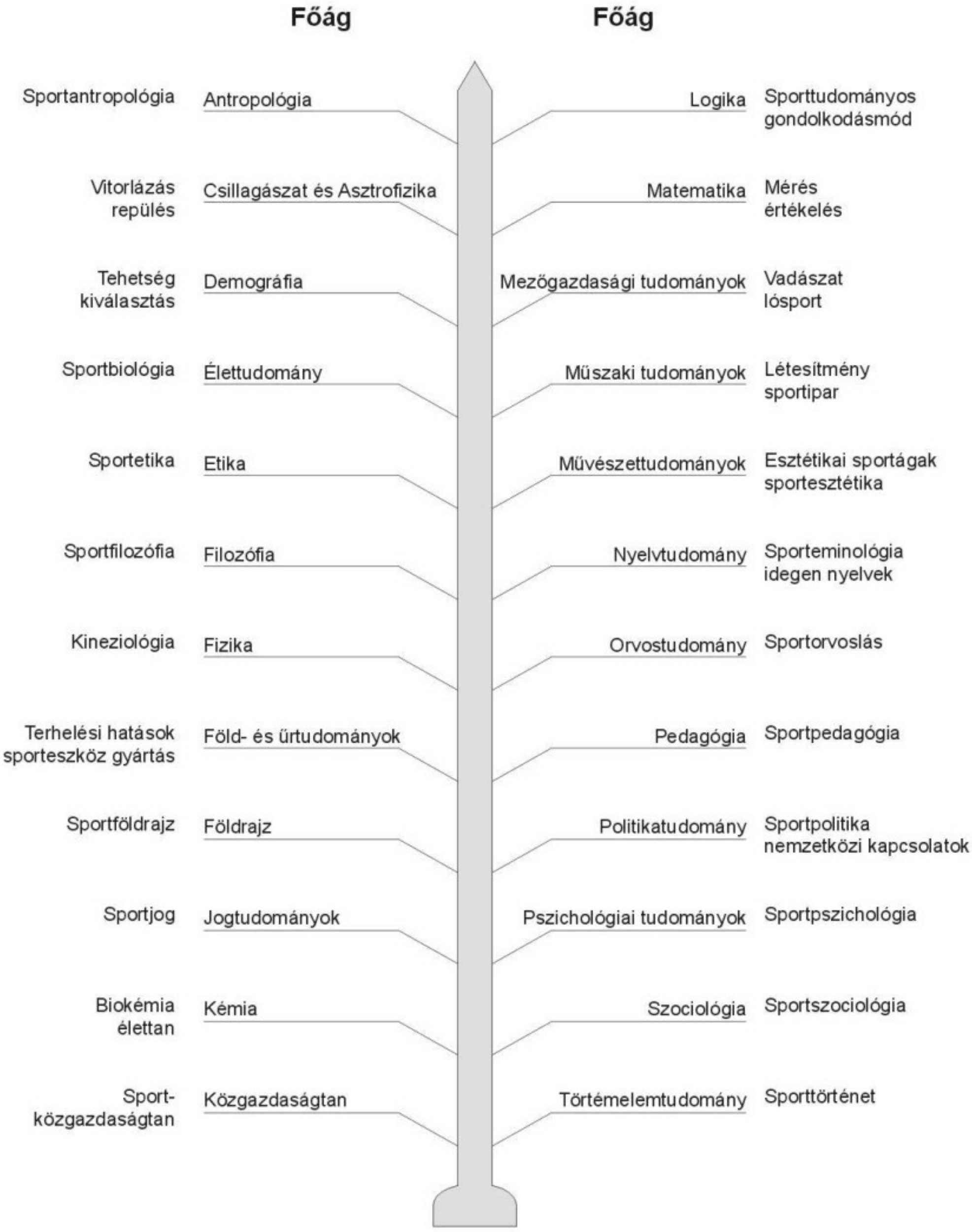
Kijelenthető, hogy a sporttudomány még a mai napig fejlődő önálló tudomány, mely kialakulásának fázisait Istvánfi (2000) a következőképpen írja le:

- empirikus fázis
- diszciplináris fázis
- interdiszciplináris fázis

Empirikus fázis (1930-)

A testkulturális tevékenységek, valamint az egyes sportágak ok- okozati összefüggéseinek felismerése, értelmezése, elemzése szakkönyvbe, tankönyvbe rendezése pl. Kerezsi Endre (1953): Torna; Csanádi Árpád (1955): Labdarúgás.

A sporttevékenység belső lényegi jegyeinek megismerése, specifikumok előtérbe kerülésére irányuló törekvések. Megkezdődik a szubdiszciplinák kialakulása (az anyatudományról való leválás) pl. Kereszty Alfonz (1954): Élettan, sportélettan; Nemessuri Mihály (1960): Sportanatómia .



1/2. ábra: A sporttudomány és a fő tudományterületek kapcsolata

Forrás: Zsolnai (1996), Vass, 2005; Horváth-Prisztóka, 2005

Diszciplináris fázis (1960-)

Ebben a fázisban a testnevelés és a sport gyakorlati problémáit próbálják a tudományos ismeretekre alapozottan vizsgálni. A vizsgálatokat többnyire önállóan végzik, de már kiscsoportos együttműködések is megjelentek, de a közös munkák szervezési feladattal nem jártak. Adaptációs fázisnak is nevezhetjük, hiszen ekkor adaptálódnak a más tudományokból a megfelelő módszerek, ekkor alakul ki a terminológia pl. Hepp Ferenc (1962): Sportwörterbuch in sieben Sprachen; Nádori László (1985): Sportlexikon I-II .

Differenciálódás szakasza: az empirikus szakaszban meginduló folyamatok erősödése eredményeként tankönyvekben is megjelennek a sporttudományok szubdiszciplinái:

- pedagógia	→	sportpedagógia
- pszichológia	→	sportpszichológia
- szociológia	→	sportszociológia
- rekreáció	→	sportrekreáció
- menedzsment	→	sportmenedzsment
- statisztika	→	sportstatisztika

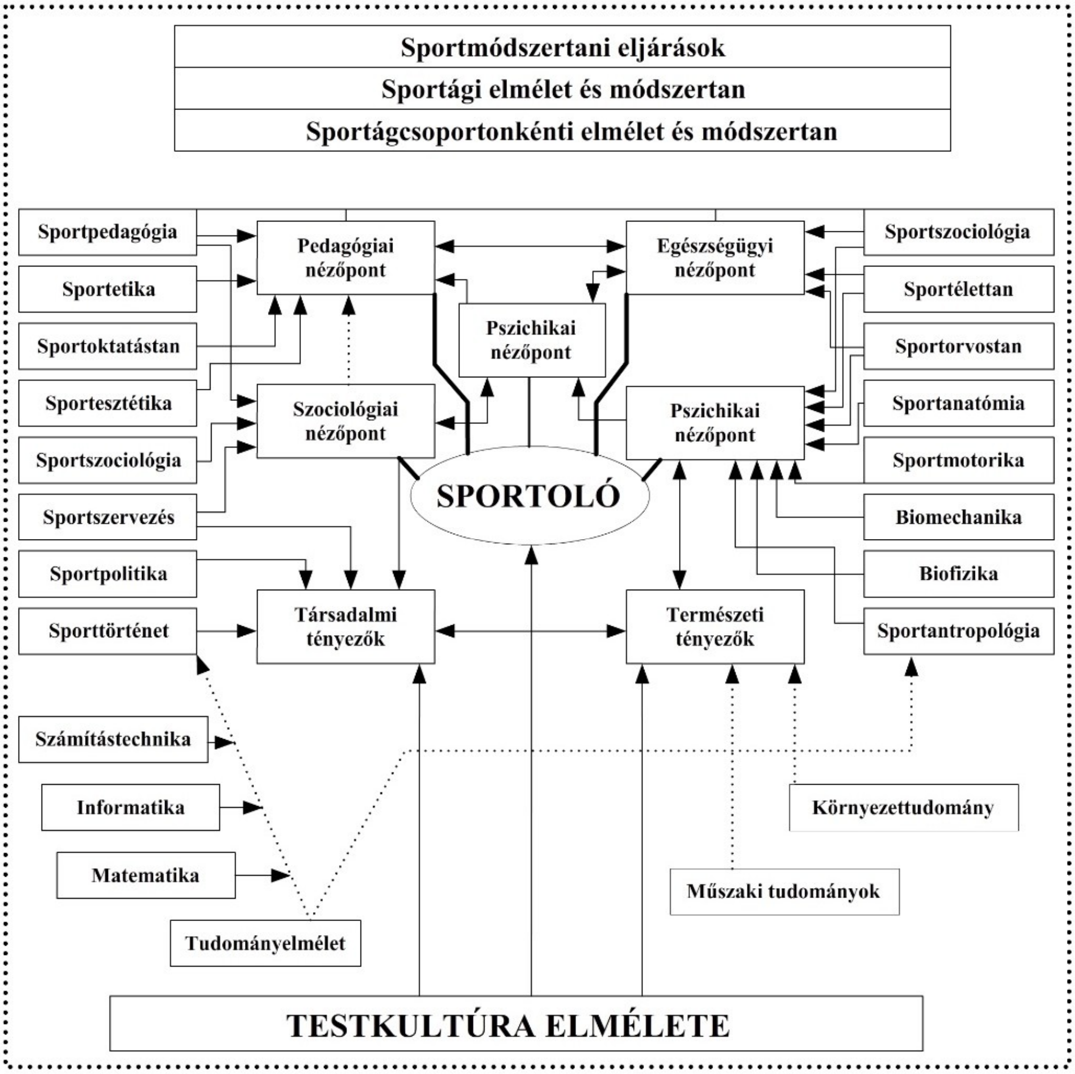
Integrálódás fázisa: a testkultúra egyes területein megtalálható módszerek összefüggő egészbe kapcsolása céljából próbálnak a gyakorlati problémákra megoldásokat találni pl. Koltai Jenő és Nádori László (1973): Sportképességek fejlesztése; Czirják József (1956): Testneveléselmélet.

Interdiszciplináris fázis (1990-

A tudományos problémák többoldalú megközelítése, melyet a több szempontúság jellemez. Ebben a fázisban jönnek létre a jól szervezett „team” munkák. Az interdiszciplináris fázis feltételei:

a) A kutatási stratégiák kijelölése.
b) Világos kutatási koncepciók és tisztázott feladatkörök meghatározása.
c) Együttműködő szubdiszciplinák harmóniája.
d) A résztvevők közötti harmónia.
e) Kutatási eredmények bemutatásának magas színvonala.
f) Felhasználók igényessége, kritikus magatartása.

Az interdiszciplináris „team-munkák” végeredménye az elméletképzés, amely a valóság általánosított rendszere. Összességében kijelenthető, hogy a sporttudomány középpontjában maga a sportoló ember áll, melynek vizsgálata természeti és társadalmi tényezők mentén, számos nézőpontból történik.



1/3. ábra: A sporttudomány komplex meghatározása

Forrás: Vass, 2005; Horváth-Prisztóka, 2005

1.2. A sporttudományos kutatómunka

A sporttudományi kutatások hazai megjelenése Hepp Ferenc nevéhez fűződik, hiszen 1941-ben a TF jogelődjének oktatott tárgyai között szerepel „A tudományos kutatás alapjai" elnevezésű tantárgy.

„Fontos és konkrét lépést jelentett, hogy a Magyar Testnevelési Főiskola tantervében az 1947-48. tanévben bevezették a testnevelési tudományos munkával való kötelező foglalkozást. Az I. és II. évben a „Testnevelés elmélete" c. tantárgy anyaga tartalmazta a testnevelési tudományos munka alapismereteit. Ennek szerves folytatását jelentette a III. és IV. évben a „Tudományos Kutatás" c. tantárgy. Ez utóbbiak keretében szaktanári irányítás alatt elkészítették a tudományos szakdolgozatukat, amely kötelező része volt a tanári szakvizsgának.” (Nádori, 1971. 37. o.). Ettől az időponttól datálható a sporttudományos munka alapelveinek és módszereinek széleskörű ismertetése, mely indukálta számos sporttudományos kutatás létrejöttét.

A sporttudományi kutatás igyekszik a törvényszerűségek feltárására, új ismeretek szerzésére, vizsgálatára, mely során számos módszert alkalmaz. A sporttudományi kutatásoknál a mai napig nem szabad élesen elhatárolni a tapasztalást és a jóval pontosabban mérhető célratörő tudományos kutatást. Nem szakadhat el a sporttudományi kutatás a gyakorlati tapasztalástól!

A tudományt és a sporttudományt is, a világról való ismeretszerzés egyéb módjaitól speciális tulajdonságai különböztetik meg, amelyek:

- általánosíthatóság,
- megismételhetőség,
- bizonyíthatóság,
- ellentmondás-mentesség (koherencia)
- analitikusság,
- egyszerűség (kompaktság, elegancia),
- fontosság (hasznosság),
- mélység (az új eredmény számos továbbihoz kapcsolható) (Csermely, Gergely, Koltay, Tóth, 1999.).

Véleményünk szerint, legtöbbször egy sporttudományos kutatásnak az alfája és omegája a tapasztalás, melyet igazolt ismeretek, módszerek segítségével próbálunk vizsgálni.

A sporttudományi kutatómunka jellegéből adódóan három fajta különböztethető meg, ami nem mindig különül el élesen egymástól, hiszen az alapkutatásoknál is egyre inkább előtérbe kerülnek az alkalmazási lehetőségek (projektek).

A tudományos kutatómunka fajtái:

1. alapkutatás (elméleti, ismeretszerzési),
2. alkalmazó (gyakorlati) kutatás,
3. fejlesztő kutatás.

Alapkutatás (elméleti, ismeretszerzési):

Az alapkutatások a világ jelenségeinek megismeréseivel, törvényszerűségeinek feltárásával foglalkozik. Általánosságban elmondható, hogy elméleti jellegűek, új ismeretek szerzésére irányulnak és nincsen gyakorlati felhasználási előfeltételeik, viszont a további tudományos munkák elméleti anyagát adják. Ez a fajta kutatási forma a sport területén viszonylag ritkább, hiszen elsősorban olyan kutatások folynak, ahol azonnali, gyakorlatban is hasznosítható eredményeket várnak.

Alkalmazó (gyakorlati) kutatás:

Az alkalmazó (gyakorlat) kutatás valamely alapkutatás elméleti eredményeinek, törvényszerűségeinek a gyakorlatra átültethető lehetőségeit vizsgálja. Jelen esetben a gyakorlati alkalmazást kell szem előtt tartani. Ha az ember keringési rendszerének változásait vizsgáltuk terhelés alatt (alapkutatás), akkor az alkalmazó kutatás során pl. a gyorsaságfejlesztés gyakorlati kérdéseit vizsgálhatjuk. Láthatóvá válik, hogy az elméleti tudomány eredményeinek felhasználásával próbáljuk a gyakorlati lehetőségeket keresni.

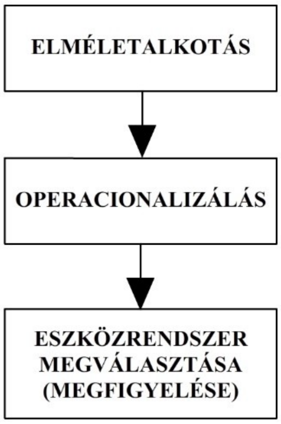
Fejlesztő kutatás:

A legtöbb kutatás valamilyen megbízásból történik, ahol a megbízó a napi szinten is alkalmazható (profitot hozó) megoldásokat (projekt terveket, eljárásokat, termékeket) vár, vagyis a közvetlen gyakorlati igény kielégítése a cél. Megállapíthatjuk, hogy a fejlesztő kutatás az alkalmazott kutatásból származó eredmények, törvényszerűségek továbbfejlesztése.

Általánosságban elmondhatjuk, hogy mindhárom fajtánál találkozunk kvantitatív és kvalitatív elemzésekkel is, a kettő élesen nem válik el egymástól. A *kvalitatív (minőségi) elemzés* a megfigyelések nem numerikus vizsgálata, amikor a cél az összefüggések mögöttes jelentéseinek megértése, feltárása, mely leginkább a sporttörténeti kutatásokra jellemző. A *kvantitatív (mennyiségi) elemzésnél* a vizsgálat numerikus alakban jelenik meg, és így is próbálja megmagyarázni a jelentéseket.

1.2.1. A tudományos kutatás alap modellje

A tudományosság működését három időben jól elkülöníthető részre (4. ábra) lehet bontani, ezek: *elmélet, operacionalizálás* és az *eszközrendszer megválasztása (megfigyelés)* (Babbie, 2000).



1/4. ábra: A tudomány alapmodellje

Forrás: saját szerkesztés, Babbie nyomán

Az ábrán jól látható, hogy az első lépésben az elméletalkotás található. Ez azt jelenti, hogy a kutatónak valami felkelti az érdeklődését a téma iránt. Kétfajta elméletalkotásról beszélhetünk, lehet *deduktív* vagy *induktív*.

A *deduktív elméletalkotás* során általános alapelvekből, törvényszerűségekből indulunk ki és alakítunk feltevéseket, hipotéziseket. Abból indulunk ki, hogy a labdarúgásban a büntetőt csak rosszul rúgni lehet, feltétezhetjük, ha a kapus megfogta, akkor azt rosszul rúgták.

Induktív elméletalkotás, amely során konkrét megfigyelések során általános elméleteket, alapelveket fogalmazunk meg. A kosárlabdázásban a centerek pontszerzéseit vizsgáltuk jegyzőkönyvek és vizuális eszközök segítségével, mely után kijelenthetjük, hogy a palánk alatt veszélyesek, de távolról csak ritkábban vállalnak dobásokat. Falus (2000) három egymásra épülő induktív kutatási stratégiát mutat be:

- *leíró*, amikor kizárólag kívülről figyeljük a folyamatot. Ezt többnyire akkor alkalmazzák, ha a valóságot kevésbé vagy egyáltalán nem ismerjük,
- *összefüggésfeltáró*, szintén beavatkozásmentes stratégia, amikor a folyamatok egymáshoz való viszonyait vizsgáljuk. Legtöbbször már ismerjük a leíró stratégia eredményeit,
- *kísérleti*, amikor a kutató beavatkozik a folyamatba, tudatosan változtatja a folyamat elemeit, hogy az összefüggésfeltáró stratégia eredményeit tovább magyarázza.

Időben az elméletalkotást az operacionalizálás követi. Ez azt jelenti, hogy a kívánt változó mérése, vizsgálata milyen lépesekből, eljárásokból és műveletekből fog állni. Ez egyfajta kutatási terv is lehet.

Legvégül a kutatási eszközrendszer (megfigyelés) következik, melyeket itt csak felsorolás szintjén ismertetünk, hiszen a későbbiekben, részleteiben találkozunk velük.

- megfigyelés,
- vizsgálat,
- kísérlet,
- interjú,
- kérdőíves módszer,
- szociometria, stb.

2. A SPORTTUDOMÁNYOS KUTATÓMUNKA FELÉPÍTÉSE, MENETE A KUTATÁSI TERV ALAPJÁN

„A sikerhez nemcsak energia és kitartás kell, hanem éppúgy hidegvér, gyors, biztos tájékozódás, megfigyelés, megfontoltság. Mindezekre szükség van az eredmények megszerzése és megtartása érdekében is, mert amit a sportban elértünk, gyorsan elvész, ha nem őrködünk felette. Ilyen formán a sport az emberi lélekbe az intellektus és erkölcs számos csíráját plántálja."(Pierre de Coubertin)

A kutatómunka felépítésének, menetének tervezése nagyon fontos feladat, hiszen ennek megfelelő színtű elkészítése nélkül, a kutatás kimenete kérdésessé válik. Az egyes tevékenységek tervezésénél az idődimenzió meghatározása kihagyhatatlan és elengedhetetlen. Jó, ha a kutatási terv készítésekor, már a részekre fordítható időt is megjelenítjük, mely sok változó függvénye (pl.: mintanagyság, adatfelvételi módszer, adatelemzési módszer, kérdezők személye, stb.). Természetesen egy kutatás lefolytatásának ideje, függ attól is, hogy *keresztmetszeti*, vagy *longitudinális* vizsgálatot készítünk.

A keresztmetszeti vizsgálatok, egyetlen időpontot reprezentáló (konkrét időpont) megfigyeléseken alapulnak. Az a kutató, aki országos felmérést készít, az igazolt férfi vízilabdázók agresszivitásának megítélése céljából, nagy valószínűséggel csak egy adott időbeli keresztmetszettel dolgozik.

Longitudinális vizsgálatoknak nevezzük, ahol a vizsgálat tárgyát képező megfigyelések hosszabb időn át folynak. Babbie (2000) három fontos típusát különbözteti meg a longitudinális vizsgálatoknak:

- *Trendvizsgálatok*, amely nagyobb populációkban vizsgálja az idővel bekövetkezett folyamatokat. Pl.: az igazolt sportolók számának változásai a népszámlálásokkor.
- *Kohorszvizsgálatoknál* kisebb populációkat vizsgál, azt tanulmányozza, hogy idővel, miként változnak ugyanazon kohorszok. Legtöbb esetben a kohorszok életkor szerint összetartozó csoportok. Pl. a 2000-ben serülő korú lány teniszezők.
- *Panelvizsgálatoknak* nevezzük azokat a longitudinális vizsgálatokat, amikor különböző időpontokban, mindig ugyanazon mintából (panelből), ugyanazon

emberektől vesznek adatokat, tehát a teljes népességből, vagy részpopulációból vett mintából mindig ugyanazon egyedeket vizsgálná időről-időre.

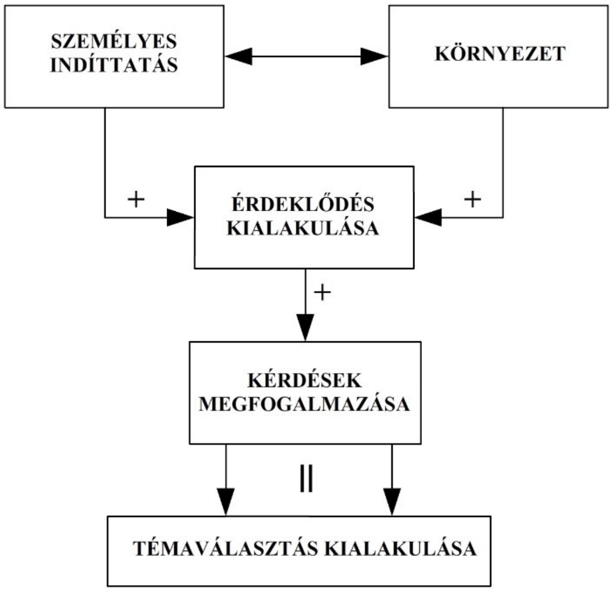
Leggyakrabban a keresztmetszeti vizsgálatokkal találkozhatunk, hiszen olcsóbb és gyorsabb, viszont a longitudinális vizsgálatokkal lehet pontosabb képet kapni az időbeli változásokról.

A kutatási terv ajánlott menete a következő (Falus, 2000):

1. A kutatási téma kiválasztása.
2. A témára vonatkozó szakirodalom elemzése.
3. A fő kutatási hipotézisek megfogalmazása.
4. A hipotézisek igazolását vagy elvetését biztosító kutatási módszerek, eszközök kiválasztása.
5. A vizsgálni kívánt minta kiválasztása.
6. A kutatás végrehajtása.
7. Az adatok elemzése, általánosítások megfogalmazása.
8. A kutatás eredményeinek közreadása, publikálása.

2.1. A kutatási téma kiválasztása

Minden tudományos kutatás kérdésekkel kezdődik, melynek eredménye az érdeklődés kialakulása. A kíváncsiság az érdeklődő embereknél alakul ki, ahogy a környezetükben lévő dolgokat különböző szempontokból vizsgálják. A sporttudomány területén az egyik leggyakoribb kérdésfeltevés „mivel tudnám a teljesítőképességemet, teljesítményemet tovább fokozni?”. Az ilyen jellegű kérdések, mindig az érdeklődésre vezethető vissza, melyet leginkább a kutató környezet determinál. Az egyetemi/ főiskolás hallgatók szakdolgozati kutatásának oka: a kényszer (szakdolgozat nélkül nem kap diplomát). A témája legtöbbször a dolgozatkészítő környezetétől is függ. A témaválasztás alkalmával meg kell határozni, hogy mit és miért vizsgálunk, és utalni kell arra is, hogy mi lehet a kutatás valószínűsíthető eredménye, valamint ha van, mi a gyakorlati hasznosíthatósága.
„A témaválasztásnál a legfontosabb az elméleti és gyakorlati felkészültség, a szakterület ismerete, ami lehetővé teszi a lényeges és lényegtelen jelenségek közötti eligazodást, a valódi problémák felismerését.”(Gyetvai-Kecskemétiné, 1997).



2/1. ábra: a témaválasztás kialakulásának folyamatábrája

Forrás: saját szerkesztés

2.2. A témára vonatkozó szakirodalom elemzése

A sporttudományi kutatásokban, mint az összes kutatásban nagy jelentősége van a már meglévő publikált irodalom összegyűjtésének, megismerésének. Sajnos sokszor találkozunk olyan kutatásokkal, cikkekkel melyeket már nemzetközileg és hazailag is sokszor vizsgáltak, mégsem említik az eredményeket. Számtalan alkalommal fordul elő, hogy olyan hipotézisekre keresik a választ a szakemberek, melyeket egy alapos szakirodalmi kutatás után, rögtön meg is tudnának válaszolni. Gyakorlatilag nagy munkával érnek el olyan eredményeket, melyet már *érvényes* [9]és *megbízható*[10] módszerekkel számos kutató *feltárt* és *leírt*. Éppen ezért, hogy az ilyen jellegű munkákat elkerülhessük, fel kell tárnunk a témára vonatkozó relevánsnak tekinthető szakirodalmat, meg kell ismernünk a hasonló kutatások megjelenített, elfogadott eredményeit.

Bármilyen szisztematikus információgyűjtés, kutatásnak minősülhet. Általában a kezdeti lépést akkor tesszük meg, amikor egy konkrét témát, vagy annak hátterét meg akarjuk ismerni. Az ember csak egy bizonyos mennyiségű információt tud megjegyezni az adott témában, így a kutatás első lépése az adott témakör irodalmazása. Annak ellenére, hogy úgy

[9] Érvényesség (validitás): jelenti, hogy a vizsgálati módszerünk vagy eszközünk mennyiben méri azt a jelenséget, illetve fogalmat, amit mérni kívánunk.

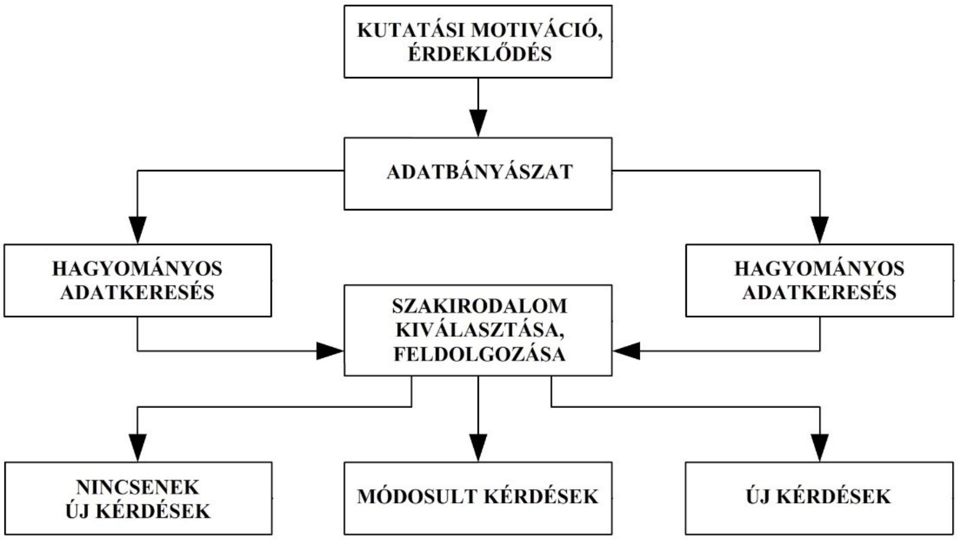
[10] Megbízhatóság (reliability): a vizsgálati eszközünknek vagy módszerünknek az a tulajdonsága, amely megmutatja, a vizsgálat ismétlése során megkapjuk-e ugyanazon eredményeket.

gondoljuk, hogy a mi ötletünk egyedi és újító, jobb „feltérképezni” a terepet, hiszen lehetséges, hogy a téma már mások érdeklődését is felkeltette. Legtöbb esetben az irodalmazás, a kutatási téma megismerése és a főbb cikkek átnézése akár évekig is tarthat, melyet befolyásol a kutatás célja. Más a kritériuma egy szakdolgozatnak vagy egy cikknek, nem is beszélve egy doktori disszertációról. Fontos, hogy kiválasszuk a lényeges információkat a rendelkezésre álló rengeteg anyagból. A számítógépek korában az irodalmazás mára megváltozott, hiszen az információhoz jutás felgyorsult.
„A szakirodalom áttekintésén általában azt értjük, hogy egy gondosan körülhatárolt témára vonatkozóan a meglévő ismeretek adott állapotát magas szinten kritikusan elemezzük, a témára vonatkozóan szintézist teremtünk. ” (Falus, 2004)

A kiválasztott témában lévő szakirodalom készítésekor figyelembe kell venni az alábbi kritériumokat:

1. **A forrás tudományos szintje megfeleljen az érdeklődési kör szintjének.** Minden esetben figyelemmel kell lenni a kutatás céljára, a várható eredmények megbízhatóságára, a közzététel helyére. Az irodalmazáskor a tudományos igényeket is kielégítő írásokat kell hangsúlyozni, de az egyéb információforrásokat sem szabad kifelejteni. Sokszor az edzők, a tanárok és a diákok, akiknek több gyakorlati problémájuk van, úgy gondolják, hogy a közkedveltebb magazinok ellátják őket megfelelő információkkal a kérdéseik megválaszolásához, de fontos felismerni, hogy habár az ilyen források nem mindig találkoznak a fontosabb tudományos eredményekkel, mégis néha fontos és hasznos információkat tudhatunk meg belőlük.
2. **El kell dönteni, hogy mennyire megbízható a forrás.** Törekedni kell mindig a legfrissebb kutatási eredmények interpretálására, ami nem jelentheti azt, hogy a legújabb kutatások, újabb források jobbak, mint a régiek. Amennyiben a témával kapcsolatban bőséges szakirodalom áll rendelkezésre, javasoljuk, hogy a cikkek absztraktjaival kezdjük a válogatást, hiszen rengeteg időt, fáradtságot takaríthatunk meg így.
3. **Meg kell keresni az alkalmas kifejezést, ami összekapcsolja a tárgykört az érdeklődési körrel.** Ez jelenti a kulcsszavak, kifejezések feltérképezését. Ezeknek a pontos meghatározása roppant fontos, hiszen segítségükkel időt, energiát spórolhatunk. A helytelen kiválasztás éppen ellenkezőleg hat, hiszen „információ robbanáshoz” vezethet.
4. **Figyelni kell a hasonló nézetekre, problémákra és fogalmakra is.** Fiatal kutatóknál, hallgatóknál előkerülő probléma, hogy a forrás használatának választásakor az új

információk végett, menet közben nézőpontot váltanak. A helyes egyensúly megtartása nagyfokú jártasságot, „rutint” követel.



2/2. ábra: Az irodalmazás folyamatábrája

Forrás: saját szerkesztés

A folyamatábra segítségével foglaljuk össze az irodalmazás menetét. Az első lépésben a kutatási motiváció (érdeklődés) tisztázása a feladat, melyet leggyakrabban a környezet indikál. Itt a két élesen elkülönülő csoportról beszélhetünk az egyik a **külső** (pl. kötelező, elvárják), a másik a **belső ösztönzésen** alapuló (pl. a saját magunk érdeklődése vezérel).

Az adatbányászat technikája mára két jól elkülöníthető részből áll: a *hagyományos (nyomtatott)* és *internetes* adatkeresésből. Mindkét technika a szekunder adatkeresés körébe tartozik, hiszen a célja a már meglévő, közölt ismereteket, adatok megtalálása, felhasználása. A *hagyományos adatkeresés* legfontosabb helyszínei a könyvtárak, levéltárak. A levéltárakat általában a történeti kutatások során veszik igénybe. A sporttudomány önálló intézményrendszerén érthetjük a saját könyvtárak meglétét is (pl. a Pécsi Tudományegyetem Testnevelési és Sporttudományi Intézetének saját könyvtára). Itt találhatók a sporttudomány területén íródott szakkönyvek, tankönyvek, jegyzetek, kutatási beszámolók, konferencia kiadványok, folyóiratok, szakcikkek, szakdolgozatok, napilapok és egyéb kiadványok (pl., Iskolai Testnevelés és Sport, Testnevelés, Testneveléstudomány, TF Tudományos Közlemények, Kalokagathia, Testnevelés- és Sportegészségügyi Szemle, Sportorvosi Szemle, A Sport és Testnevelés Időszerű Kérdései, Mesteredző, Sporttudomány, Magyar Edző, Asztalitenisz, stb.).

A számítógépek nagyban megkönnyítették az irodalmazást, mivel közvetlen hozzáférést biztosítanak az adatbázisokhoz. Ennél az adatbányászati technikánál konkrét helyszínt említeni nem tudunk, hiszen a világháló mára szinte mindenhonnan elérhető. A teljesség igénye nélkül néhány interneten található sportirodalom elérése: www. sporttudomany.hu; www.magyaredzo.hu; www.threef.hu; www.nupi.hu; www.hupe.hu (Kalokagáthia, régebben a TF Tudományos Közleményei); www.leistungssport.net. A számítógépes adatkeresés számos előnye (pl. gyors, olcsó), mellett fontos kiemelni, hogy az itt megjelenő ismeretek jó részét ellenőrzés nélkül (szaklektorálás elhagyásával) teszik elérhetővé, melynek következtében az ott közölt adatok hitelessége megkérdőjelezhető.

A szakirodalom kiválasztásánál, feldolgozásánál a hagyományos (nyomtatott) módszerrel történő adatbányászat után az adatok tömörítésére van szükség, melyet jegyzetkészítéssel tudunk elvégezni. Ennek a lényege, hogy az adatokat igyekezzünk a legminimálisabbra csökkenteni, de a visszakereshetőséget mindvégig megmaradjon, Attól függően, hogy hány irodalmat kívánunk feldolgozni, megkülönböztetjük a *kulcsszavas kiemelés (kivonatolás)*, illetve a *cédulázás* módszerét.

A *kulcsszavas kiemelés* során viszonylag kevés irodalmunk van, mellyel egyszeri alkalomra történő felkészülés a célunk (pl. dolgozat, előadás). Itt a legfontosabb gondolatokat, kulcsmondatokat emeljük ki, melyekből a feldolgozás követően sorrendet is készíthetünk.

A cédulázást (cetlizés) akkor alkalmazzuk, ha több irodalomból kívánunk információt megőrizni. Ilyenkor nem csak a kulcsszavakat, hanem magát a konkrét gondolatot vetjük papírra. Ha hosszabb gondolatokat, idézetek veszünk át, fontos a mű adatainak pontos megjelenítése is. Ennek okai: bármikor visszakereshető legyen, ne minősüljön részleges vagy teljes plágiumnak.

Tudomásul kell venni, hogy az irodalmi anyagok átnézése újabb kérdéseket eredményezhet, változtathat a régieken, vagy éppen válaszokat találhatunk, tehát megelégedhetünk, azaz nem teszünk fel újabb kérdéseket.

2.2.1. Irodalomjegyzék és hivatkozások készítése

Az irodalomjegyzék (bibliográfia) és a hivatkozások egyértelmű célja:

- az azonosíthatóság,
- és a visszakereshetőség.

A munka során felhasznált idézések, hivatkozások, irodalomjegyzék elkészítése kihagyhatatlan feladat, mely gyakran nagyon pontatlanul történik. Az irodalomjegyzék és hivatkozások készítését Harsányi László kéziratának a segítségével mutatjuk be.[11]

Idézésnek számít minden mástól átvett módszer, gondolat, ábra, táblázat. A tudományos életben kijelenthetjük, hogy mindenki szeretné a saját szellemi termékét védeni, éppen ezért az egyik legsúlyosabb etikai vétségnek számít a plagizálás.

Plágium: a szellemi alkotások részbeni vagy teljes eltulajdonítása, saját névvel való közzététele. Napjainkban számos eljárással próbálkoznak, hogy a plagizálásokat kiszűrjék. Óriási adatbázisokat hoznak létre (feltöltik szakcikkekkel, szakkönyvekkel, szakdolgozatokkal, disszertációkkal), melyet össze tudnak hasonlítani az újonnan megírt alkotással. Az összevetés esetén egyezőséget találnak és az eredeti szerzőkre nincs hivatkozás, akkor a plágium gyanúja állhat fenn. Éppen ezért a legtöbb felsőoktatási intézmény mára a szakdolgozatokat, diplomamunkákat a nyomtatott formán kívül már számítógépes adathordozón (CD lemez) is bekéri.

A szellemi alkotásról, a többi más termékhez hasonlóan törvény rendelkezik. Az 1999. évi LXXVI törvény rendelkezik az irodalmi, tudományos, művészeti alkotások szerzői jogáról. Mindenképpen szem előtt tartandó, a szabad felhasználásról és a szerzői jog más korlátairól alkotott törvénykezések, melyekről szintén a megnevezett törvény intézkedik.

AZ 1999. ÉVI LXXVI: TÖRVÉNY A SZERZŐI JOGRÓL

1. § (1) Ez a törvény védi az irodalmi, tudományos és művészeti alkotásokat.

A SZABAD FELHASZNÁLÁS ÉS SZERZŐI JOG MÁS KORLÁTAI

Általános szabályok

33. § (1) A szabad felhasználás körében a felhasználás díjtalan és ahhoz a szerző engedélye nem szükséges. Csak a nyilvánosságra hozott művek használhatók fel szabadon e törvény rendelkezésének megfelelően.

[11] Harsányi L. (2007)

33. § (4) E fejezet rendelkezéseinek alkalmazása szempontjából az iskolai oktatás célját szolgálja a felhasználás, ha az, az óvodai nevelésben, az általános iskolai, középiskolai, szakmunkásképző iskolai, szakiskolai oktatásban, az alapfokú művészetoktatásban vagy a felsőoktatásról szóló törvény hatálya alá tartozó felsőfokú oktatásban a tantervnek, illetve a képzési követelményeknek megfelelően valósul meg.

A szabad felhasználás esetei

34. § (1) A mű részletét –az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző megnevezésével bárki idézheti. Átvételnek minősül a mű olyan mértékű felhasználása más műben, amely az idézést meghaladja.

34. § (4) A célnak megfelelő módon és mértékig saját célra, valamint – vállalkozási tevékenységen kívüli – belső intézményi célra is készíthető másolat, ha az jövedelemszerzés vagy jövedelemfokozás célját közvetve sem szolgálja és

a) tudományos kutatáshoz szükséges,
b) saját példányról archiválásként tudományos célra vagy nyilvános könyvtári ellátás céljára készül, vagy
c) megjelent mű kisebb részéről, illetve újság- vagy folyóiratcikkről készül.

34. § (5) Könyvként kiadott mű egyes részei, valamint újság- és folyóiratcikkek az iskolai oktatás céljára egy-egy iskolai osztály létszámának megfelelően, illetve köz- és felsőoktatási vizsgákhoz szükséges példányszámban többszörözhetők.

34. § (6) Szabad felhasználás a mű ideiglenes többszörözése, ha kizárólag az a célja, hogy megvalósulhasson a műnek a szerző által engedélyezett, illetve e törvény rendelkezései alapján megengedett felhasználása, feltéve, hogy az ideiglenes többszörözés az ilyen felhasználásra irányuló műszaki folyamatnak elválaszthatatlan része, amelynek nincs önálló gazdasági jelentősége.[12]

[12] 1999. LXXVI. tv.

Minden dokumentumtípus leírásáról külön „szabvány” rendelkezik, melyek nagyon gyakran összekeverődnek. Az irodalomjegyzék, jegyzetek, hivatkozások alapegysége: a *bibliográfiai tétel*.

A bibliográfiai tétel könyvek esetében az alábbi kötelező adatokból tevődik össze: a szerző teljes neve (vezeték- és keresztnév), a mű címe (ha van alcíme), a kiadás száma, a megjelenés helye, a kiadó, a kiadás éve, oldalszám. Gyakran találkozunk olyan formulával is, melynél a szerző neve után következik a kiadás évszáma.

Rétsági Erzsébet: Kézikönyv a testnevelés tanításához (5-8. osztály). Budapest-Pécs, Dialóg Campus Kiadó, 2005. 352. o.

vagy

Rétsági E. (2005): Kézikönyv a testnevelés tanításához (5-8. osztály). Budapest- Pécs, Dialóg Campus Kiadó, 352. o.

Hepp Ferenc: A mozgásérzékelés kísérleti vizsgálata sportolókon. Pszichológia a gyakorlatban 22. kötet. Budapest, Akadémiai Kiadó, 1973.

Eco, U. (1991): Hogyan írjunk szakdolgozatot? Budapest, Gondolat, 256 o.

Hivatkozás: A hivatkozás formáit mindig az idézet végén kell feltüntetni. Két eljárás is elfogadott:

1. A leggyakoribb, amikor a szerző(k) nevét és a megjelenés évszámát tüntetjük fel. Pl.:(Nyerges, 1981). Nem kell ügyelni az irodalomjegyzékben található számozásra.
2. Csak egy szám pl.: (12), vagy a szerző(k) neve és sorszáma szerepel pl.: Dubecz (12). Itt figyelni kell, hogy a szám megegyezzen az szerző(k) irodalomjegyzékbeli sorszámával.

Példák:

- Berkes (2007)
- Ormai (1981, 126. o.)

Idézések során, csak a szó szerinti idézéseknél írjuk a szöveget idézőjelben!

Bibliográfiai tétel megadása folyóiratcikkeknél a következőképpen alakul: a szerző neve, a cikk címe, (elfogadott az egyenlőségjel, de ritka), folyóirat neve, évfolyam, év, hónap (vagy szám), oldalszám. Itt is gyakori, a kiadás évének szerepeltetése közvetlenül a szerző után zárójelben. Egy szerző több azonos évbeli írása esetén az évszámot követően az „abc” kis betűivel teszünk különbséget.

Katics L.-Petrekanits M.-Derzsy B.-Gedő D.-Römer I. (1992): Szabad-, versenyaerobic és akrobatikus gyakorlatok hatásai. Mester-edző, 4. 12-20. o.
Pintér J.–Rappai G. (2001): A mintavételi tervek készítésének néhány gyakorlati megfontolása. Marketing & Menedzsment 2001/4. 4-11. o.
Harsányi L. (1992 a): A blokkszerű terhelés az éves felkészülésben. Testnevelés- és Sporttudomány 3. 109-119. p.
Harsányi L. (1992 b): Die „Geheimnisse” der ungarischen Leistungen. Leistungsport, 6. 27-29. o.

Hivatkozás:

- Katics és mtsai. (1992) vagy (Katics és mtsai., 1992)
- Harsányi (1992 a)
- Harsányi (1992 b)

A magyar edzésmódszertan korai szerzői többen voltak (Vadas, 1927; Abád, 1962; Kutas, 1962; és Nádori, 1962, 1968, 1972).

Egy szerző több művére hivatkozás esetén:
A sportbeli felkészülés szakaszokra osztásával Nádori (1962, 1972, 1981), Harsányi (1992 ab) és Platonov (1987, 1999) foglalkozott.

Könyvfejezetek, megjelent konferencia előadások, gyűjteményes műveknél is készíthetünk bibliográfiát, illetve hivatkozhatunk is rá. A szerző neve, a cím, in, a kötet szerzői, a kötet címe, a kötet száma, hely kiadó, év, a feldolgozott rész oldalszámai.

Ács P. (2006): The analysis of the regional competitiveness of the Hungarian sport with multivariable statistical methods, In Hughes M. (szerk.): World Congress of Performance Analysis of Sports 7, Berzsenyi Dániel College, 299. – 309. o.

Hivatkozás:

- Ács (2006)

Heti és napilapoknál a következő a sorrendben írjuk. A szerző neve, cím, a folyóirat neve, megjelenés ideje, oldalszám.

Kulcsár Gy. (2008):A kapitány visszaszúr. Nemzeti Sport, szeptember 24. 20. o.

Hivatkozás

Kulcsár (2008)

Egyéb művek bibliográfiája és hivatkozásaikon értjük a kéziratokat, kutatási jelentéseket, szakdolgozatokat, diplomamunkákat, disszertációkat, értekezéseket, szabványokat, szabadalmakat. Itt a sorrend: szerző neve, cím, alcím, a mű jellege, esetlegesen az egyetem vagy főiskola megnevezése zárójelben, hely, évszám. A szabványoknál a szabványszám, hatályba lépés évszámát, és a szabvány címét kell megadni. A szabadalmaknál a szabadalom tulajdonosának a neve, a találmány címe, a feltalálók neve, az ország neve és a szabadalom száma, szabadalom megjelenésének kelte.

Rétsági E. (1996): Törekvések az iskolai testnevelés és a testnevelő tanárképzés megújítására. Kandidátusi értekezés. Magyar Tudományos Akadémia. Budapest. 186. p.

Mikroelektronikai Vállalat (Budapest). Kettős színhatású folyadékkristályos kijelző. Bence Gy-Seyfried É-Véghelyi T. HU 182-495. 1986. 06. 30

Ács P. (2007): A sportolói vándorlás, „Migráció – társadalmi összefüggések” című konferencia, Budapesti Corvinus Egyetem Szociológiai és Társadalompolitikai Intézete, valamint a Magyar Statisztikai Társaság Demográfiai szakosztálya. Budapest, 2007. október 19.

Hivatkozás:

- Rétsági (1996) azt írta, hogy…
- (SZAB)-HU 182-495 (1986)
- Ács (2007) szóban azt mondta, hogy…

A számítógépen (internetes) történt adatok, cikkek irodalomjegyzékbe, és hivatkozásként történő megjelenítése a szintén mára „kötelezővé” vált, még akkor is, ha közölt adatok hitelességét, visszakereshetőségét nem tudják garantálni. Éppen ezért a szerző neve, évszám, a mű címe, a megtalálás webcíme mögé a megtalálás időpontját is zárójelben fel kell tüntetni.

Soós I. (2002): A sportpedagógia, mint prevenciós eszköz a fiatalok egészségnevelésében. Kalokagathia. 2002. 1-2, 130-135. o. http:// http://www.hupe.hu/info/kg/cikkek/2002_1-2.pdf (2008. 07. 17)

Hivatkozás: megegyezik a többi információhordozóéval.

A bibliográfiában és a hivatkozás során előforduló formaságok közül kiemelendő, hogy a szerző(k) neve mellett soha nem tüntetjük fel a tudományos fokozatot, egyetemi rangot. Az irodalomjegyzékben jelenjen meg minden olyan mű, amelyre hivatkozás történt, de ne szerepeljen olyan mű, amire nem történt hivatkozás a szerző részéről. A több szerzős műveknél az írókat a művön közölt sorrendben kell feltüntetni, de a hivatkozásoknál elég az első szerző neve után azt, hogy mtsai (munkatársai) vagy et al. (többiek) feltüntetni. Külföldi szerzők műveinél is a vezetéknevet írjuk először, melyet a keresztnév követ legtöbbször csak kezdőbetűvel jelölve. Az irodalomjegyzékben fontos, hogy az első szerzők nevét szoros ABC sorrendbe rögzíteni. Ugyanazon szerző(k) műveit évszám szerint is rendezni kell, illetve az egy évben megjelent munkákat is meg kell különböztetni. Egyes műveknél hiányzik a kiadás helye (h.n. = hely nélkül), kiadás éve (é.n.= év nélkül) vagy a kiadó megjelölése (k.n. = kiadó nélkül). Az oldalszámok feltüntetésére is sokféle megoldással találkozhatunk, hiszen gyakran hazánkban is oldal latin megfelelőjét (pagina) alkalmazzák. Ha így látjuk, hogy p. 146 vagy 146 p., akkor azt jelenti, hogy a könyv 146 oldalból áll. Amennyiben ezt találjuk, hogy 45-49 pp. vagy pp. 45-49, akkor az idézés a 45. oldaltól a 49. oldalig tart. Javasolni tudjuk, hogy az oldalszámot magyarul, és a számadatok mögött tüntessük fel. A szellemi alkotás irodalomjegyzékének közlés során törekedjünk, mindig az azonosíthatóságra és a visszakereshetőségre. Az irodalomjegyzék elkészítését segíti, ha folyamatosan történik, vagyis a mű írása közben lépésről-lépésre bővül. Ha az irodalomjegyzékünk viszonylag hosszabb érdemes az irodalmat csoportosítani a következő sorrendben:

- *magyar nyelvű könyvek,*
- *magyar nyelvű cikkek,*
- *egyéb magyar nyelvű források,*
- *idegen nyelvű könyvek,*
- *idegen nyelvű cikkek*
- *internetes hivatkozások.*

Napjainkban a hagyományos irodalomkutatást egyre inkább felváltja a világhálón történő irodalmazás. Az *interneten való irodalomkutatás többféle forrásból valósulhat meg:*

- *Elektronikus könyvtári katalógusok*
- *Elektronikus úton elérhető bibliográfiák*
- *On-line adatbázisok*
- *Elektronikus könyvek, folyóiratok, cikkek*
- *Internetes keresőrendszerek*

A sporttudományban is egyre gyakrabban találkozhatunk további kiegészítő szolgáltatásokkal, melyek az irodalomkutatást segíthetik:

- *A Magyar Sporttudományi Társaság hírújságja (www.sporttudomány.hu)*
- *A Magyar Egyetemi és Főiskolai Sportszövetség honlapja (www.mefs.hu)*
- *A Magyar Asztalitenisz Szövetség hírleve (www.moatsz.hu)*

Az elektronikus keresőrendszerek és elektronikus könyvtárak használatát az olvasó részletesen megtalálja a Gyakorlati adatelemzés című könyvben (Ács, 2015). Az elektronikus könyvtárak közül ebben a tankönyven csak a Testnevelési Egyetem elektronikus könyvtárát kívánjuk bemutatni.

A Semmelweis Egyetem Testnevelési és Sporttudományi Kar Digitális Könyvtár gyűjteménye a testnevelési és sporttudományi területen megjelenő releváns szakirodalmak közzétételére fókuszál. A feldolgozott szakirodalmat a jobb áttekinthetőség céljából gyűjteményekbe rendezi, melyben folyóiratcikkek, könyvek, konferencia kiadványok, PhD-értekezések, régiségek (sportlapok), válogatott TF-szakdolgozatok is olvashatóak, emellett sporttudománnyal kapcsolatos videókat és képeket is tartalmaz. A Magyar Testnevelési Egyetem feldolgozott folyóiratai, mint a Kalokagathia közleményei és a Magyar Sporttudományi Szemle teljes szöveggel is elérhetőek digitális könyvtár jóvoltából.

Keresési lehetőségek

Alkalmazása során a keresés egyszerű és összetett módon történhet. Az egyszerű keresés a keresődobozba beírt kulcsszavak megadásával kezdődik. A keresés során alkalmazni lehet a csonkolást, a Boole-operátorokat, a zárójelet és proximity operátorokat. Összetett keresés történhet cím, szerző, tárgy, földrajzi fogalom, csak metaadat, és teljes szöveg szerint. Emellett a dátum intervallum, a gyűjtemények köre és a média típusa (kép, szöveg, hang, videó) is meghatározható.

Találatok megjelenítése

A találatok megjelenítése *Rövid formátum*, *Táblázatos megjelenítés* vagy *Teljes megjelenítés* formátumban tekinthető át. A rövid- és táblázatos formátumban jól látható a releváns találatok címe, a szerző(k) neve(i), és a PDF ikon. A címre kattintva a szerző mellett a tárgyszavakat, földrajzi fogalmat, megjelenés helyét és idejét, a szakirodalom terjedelmét és a kapcsolódó gyűjteményt tárja az érdeklődők elé (Forrás: Karamánné Pakai Annamária, Oláh András, 2015)

Hozzáférés

Semmelweis Egyetem Testnevelési és Sporttudományi Kar Digitális Könyvtár a http://tf.hu/oktatas/konyvtar/tf-digitalis-konyvtar/digitalis-dokumentumok/ oldalon keresztül érhető el.

2/1. képernyőnézet: Keresőfelület

2/2. képernyőnézet: Találati lista

2.3. A fő kutatási hipotézisek megfogalmazása.

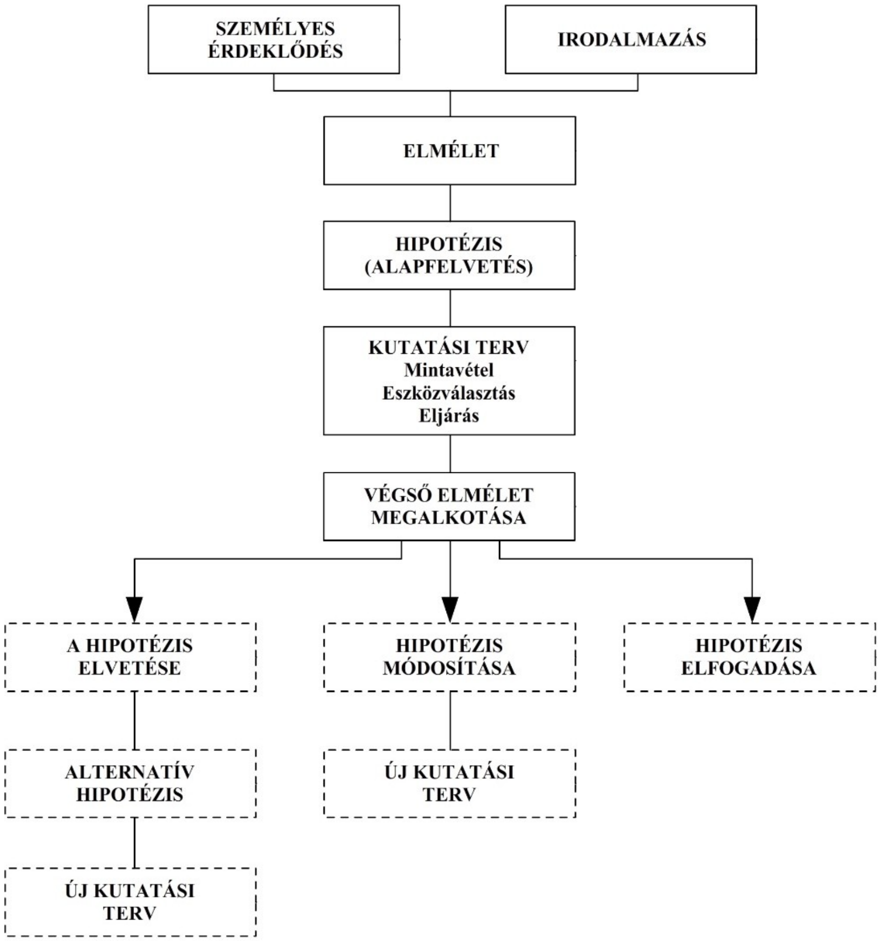
Általánosságban hipotézisnek *nevezzük, valamely jelenség természetére vonatkozó elméletből, gyakorlatból (tapasztalásból) levezetett feltevést, vagyis egy olyan kijelentés, mely a fő feltételezést fogalmazza meg a kutatásban szereplő változókra és kapcsolatikra.* <u>A fő kutatási hipotéziseinknek tartalmaznia kell a kutatásunk végső valószínűsíthető eredményét.</u>

Leggyakrabban már a témaválasztás alkalmával felmerülő kérdések vezetnek a tudományos feltevések kialakításához. Ezt segíti a szakirodalom áttekintése is, hiszen már az előkészítés (irodalmazás) során törekedni kell, hogy minél több információ birtokába jussunk.

A kutatók szerint a tudományos feltevések (hipotézisek) azok, melyek segítségével irányíthatjuk, értelmezhetjük, valamint megjósolhatjuk a világban zajló eseményeket, történéseket. Az alapvető feltevések és gondolatmenetek többnyire *megbízható teóriákra* támaszkodnak, és *vezérfonalként szolgálnak*. Ennek következtében egy fő kutatási feltevés(ek) (hipotézis) *feltételezésekből, meghatározásokból* és *érvelésekből* áll(nak), melyek logikus módon kölcsönösen egymáson alapulnak, megalkotva így egy fogalmi rendszer módszertani alapjait. Egy megbízható teória a következő kritériumoknak kell, hogy eleget tegyen:

- Tartalmazzon logikus kijelentéseket, melyek koherens egységet alkotnak és eseményei bizonyíthatóak.
- Egyesítse a kifejtést, és a fogalmat, melyekből a kutatási terv származtatható, megvizsgálható valamint megerősíthető illetve cáfolható.

Ennek tükrében célszerű, a kutatási tervet a fő kutatási hipotézisek megfogalmazása után elkészíteni (2/3. ábra.), hiszen nagymértékben befolyásolhatja azt.



2/3. ábra: A kutatási hipotézisek megfogalmazásának időrendi ábrája

Forrás: saját szerkesztés G. Tenenbaum és M. Driscoll (2005) nyomán.

A jó kutatási hipotézisekkel szembeni elvárásokat Falus (2000) a következőképpen fogalmazza meg:

1. A hipotézisnek magyarázó erővel kell rendelkeznie, vagyis mindenki számára elképzelhető legyen a javasolt feltételezés, összefüggés.
2. A jó hipotézis jelzi a változók kapcsolatát. Pl.: a kosárradobások gyakorlási idejének növekedésével nő a ponterősség.

3. A hipotézis, egyértelműen igazolható vagy elvethető. Az előző példát folytatva az adott csapat mérkőzéseinek jegyzőkönyvei igazolják.
4. A hipotézisek igazolása vagy elvetése során felhasznált módszerek, technikák, eljárások, számítások, mérések kivitelezhetőek legyenek. A gyakorlási idő stopperórával (pl.: percben), a ponterősség a dobott pontok számával mérhető.
5. A hipotézis egyértelmű, világos, operatív terminusokban legyen megfogalmazva. Sokszor találkozunk olyan kutatással, melyben a hipotézisekben kizárólag az élsportolókra vonatkozó feltevés szerepelnek. Pl.: a profi sportolók kevesebbet dohányoznak. Ez itt nem helyes, hiszen konkretizálni kell a profi sportoló fogalmát.
6. A hipotézisnek támaszkodnia kell a meglévő ismereteinkre. Általában a feltevéseinket úgy fogalmazzuk meg, hogy igaz válaszokat kapjunk. Ez nem jelenti azt, hogy kellő megalapozottsággal ne vizsgálhatnánk más által igazoltnak hitt állításokat. Ezekkel főleg akkor találkozunk, ha valamely új sportági eszközt, vagy technika megjelenését vizsgáljuk.
7. A kiinduló feltevéseket a lehető legérthetőbben, legegyszerűbben és legtömörebben kell megfogalmazni. Az összetettebb problémáknál feltehetünk alhipotéziseket is.
8. A hipotézisek összességének választ kell adnia a témaválasztás során választott kutatási problémára.

Azt, hogy a kutatásunk során, hány fő feltevést (hipotézist) vizsgálunk, mindig függ a kutatás céljától és körülményeitől is. Az biztosan nem célravezető, ha ez a szám túl magas. Pl.: egy szakdolgozat írásakor nem előnyős, ha pl. ötnél több hipotézisre keressük a választ, hiszen ennél többet megfelelő minőségben vizsgálni, a megadott szempontrendszerek megtartásával nehézkes.

2.4. A hipotézisek igazolását vagy elvetését biztosító kutatási módszerek, eszközök kiválasztása.

A tudományos megfigyelések abban különböznek a hétköznapi megfigyelésektől, hogy céltudatosan, tervszerűen és rendszerességgel végezzük őket.

A tudományos kutatás módszereinek megválasztása mindig a kutató privilégiuma. A kutatási módszer jelenti azt az eljárást, ami azt mutatja, hogy hogyan végeztük az információk gyűjtését, rendszerezését, tárolását, valamint a feldolgozáshoz milyen eljárásokat alkalmaztunk.

A kutatási módszerek közül megkülönböztetjük:

I. Feltáró módszereket,
II. Feldolgozó módszereket

<u>A feltáró módszerek</u> többnyire kvalitatív technikákat használnak, melyek közül a leggyakrabban használt módszereket mutatjuk be.

A **megfigyelés** során közvetlenül nem avatkozunk bele az események menetébe, hanem természetes körülmények között tanulmányozzuk a jelenségeket, és összefüggéseiket. A megfigyelés sikerességét meghatározza a megfigyelő felkészültsége. Fontos a megfigyelés tárgyának, személyének helyes megválasztása. A megfelelő eljárás hatékonyságát a tudatosság és a tervszerűség határozza meg. A sporttudományi kutatások során nagyon gyakran használnak a megfigyelési módszerekhez segédeszközöket, melyek az észlelést és adatrögzítést hívatottak szolgálni. A legkedveltebb segédeszközök: diktafon, kamera, fénykép, videó, tükör, jegyzőkönyv, stb. Az egyéni sportágakat űzőknél sokszor találkozhatunk azzal, hogy a saját mérkőzéseiket, mozgásaikat, saját élményeiket kamerával rögzítik, ezt *introspekciónak* (önmegfigyelésnek) hívják. Ennek a célja legtöbbször, hogy a mozgást utólag megtekintve elemezzék, ezt az eljárást *retrospekciónak* nevezik. Mások megfigyelésére (objektív megfigyelés) legtöbbször mérkőzések előtt történik taktikai célzattal.

Előnyei: önmagában is teljes értékű lehet, a közvetlenség rugalmassá teszi, az oktatás eredménye mellett a folyamatát is megmutatja.

Hátrányai: időigényes, nagy szervezési nehézségek, szubjektív hibalehetőségek, a megfigyelő empátiája, megfigyelési észlelési hibák lehetségesek.

A **vizsgálat** lényege a megfigyelni, elemezni kívánt jelenség létrejöttének feltételeit a kísérletező szándékosan teremti meg (előnye, hogy a jelenség megismételhető, kontrollálható). Talán a legmegbízhatóbb adatokat a sporttudományi kutatások számára a kísérletek szolgáltatják. Megkülönböztetünk laboratóriumi (pl.: dopping labor), és természetes kísérleteket. Többféle módszert különböztetünk meg, melyek közül a legközvetlenebb a beszélgetés (exploráció), míg a legegységesebbek a tesztek (pl. IQ teszt), hiszen nagyrészük „standardizált".

A **kikérdezés** módszerének lényege, hogy valamely kutatás keretében kérdések segítségével információkat gyűjtsünk. A kikérdezés alkalmas egyének, esetleg csoportok együttes ismereteinek, véleményének, attitűdjeinek, élményeinek, motívumainak, életmódjának a felderítésére. A módszer alapváltozatai: a *szóbeli* és az *írásbeli* kikérdezés.

A szóbeli kikérdezés során személyes interakció jön létre a kérdező és a kérdezett között. A kikérdezettek száma szerint lehet *egyéni* vagy ***csoportos***.

Az *egyéni kikérdezések* alkalmával leggyakrabban a sporttudományi kutatások során a *mélyinterjút*, a ***narratív interjút***, valamint az *explorációt* alkalmazzák.

Mélyinterjú: az interjúnak ez a variációja a legbelsőségesebb dolgok feltárására is alkalmas. Általában neves sportolókkal, szakvezetőkkel készülő hosszabb beszélgetés, melyben a kérdezett olyan élményekről is beszél melyről másnak még nem tette. Ebben az interjúváltozatban mind a kérdezőnek, mind az interjúalanynak nagy a szabadsága, de a felelőssége is.

Narratív interjú: az interjúnak ennek a változatában a kérdezett kizárólag olyan eseményekről, élményekről beszél, amelyek vele történtek meg, ő élte át őket.

Az exploráció, a személyes, irányított kikérdezés lényege, hogy kutatás és gyakorlati tevékenység során korábban kialakult, irodalomból ismerhető és a gyakorlatban használt kész kérdéssorok együttesét alkalmazzuk, kiegészítve azt az adott kutatáshoz konkrétan kapcsolódó speciális kérdésekkel. A vizsgálatot vezető személy közvetlen kapcsolatba kerül a kísérleti alannyal. A kikérdezés segítségével megkísérelhetjük valamely összefüggés feltárását, szabályszerűség igazolását. Ebben az esetben követelmény a reprezentativitás (hiszen mintával dolgozunk), a reliabilitás (megbízhatóság) és a validitás (érvényesség) biztosítása.

A <u>reprezentativitás</u> a minta jellemzője, hogy a benne lévő tulajdonságok megoszlása, megfeleljen az alapsokaságban lévőével (cseppben a tenger), így a kikérdezésbe bevont személyek a kutatás szempontjából fontos jegyekben (amelyek feltárhatók, megragadhatók) tükrözik a vizsgálni kívánt népességcsoport egészét, és ez statisztikai eljárásokkal vizsgálható, bizonyítható. Így valószínűsíthető, hogy a mintából kapott eredmények az alapsokaságra vonatkoztatva hasonló tulajdonságokkal, eredményekkel bír. A mintavételi

eljárások közül a valószínűségi eljárások növelik a reprezentativitást.
A **csoportos interjú** alkalmas a „kollektív tudás” megismerése, a csoport véleményének feltárására, a csoporton (csapaton) belüli egyéni érvényesülések bemutatására.
A csoportos interjú a kérdezőt természetesen nagy feladat elé állítja, amely megoldásának egyik alapfeltétele, hogy bele tudjon érezni, bele tudjon helyezkedni a csoport kommunikációs stílusába. Legkedveltebb formája a fókuszcsoportos interjú.

A **fókuszcsoportos kutatás** a kvalitatív kutatási megoldások legnépszerűbb formája. Célja, hogy a meghívott kutatási alanyok véleménye mellett feltárja érzéseiket, attitűdjeiket. A kutatás során a megfelelően elkészített szűrőkérdőívvel gyűjtött és kiválasztott 8-12 személy beszélget korábbi tapasztalatairól, osztja meg egymással véleményét, érzéseit, meggyőződéseit egy adott témára fókuszálva. A beszélgetést moderátor vezeti, ő a beszélgetés koordinátora. Két feladata van: egy olyan beszélgetés létrejöttének segítése, amely során kiegyensúlyozott, szabad és nyitott véleménycsere alakul ki a csoport tagjai között, és az, hogy beszélgetés közben a megbízóval egyeztetett kutatási vezérfonal minden lényeges elemét érintsék. Mindezt olyan légkörben, amelyben a válaszadók gátlásai feloldódnak, megnyílhatnak egymás előtt. A kutatás rögzítésre kerül, így a kutatásban résztvevő szakemberek mellett, a megbízó is megkaphatja a beszélgetést, hogy így személyes élménye legyen saját célcsoportja viselkedéséről. A fókuszcsoportos kutatás célja: általános háttér-információk gyűjtése egy adott témában, új program, termék, szolgáltatás bevezetése során jelentkező potenciális problémák vizsgálata, új, kreatív ötletek nyerése, információt szerezhetünk az új ötletek főbb erősségeiről és gyenge pontjairól.
Ezt a technikát általában egy új termék vagy szolgáltatás bevezetése előtt alkalmazzák, pl.: egy sportcentrum, vagy wellness-szálloda nyitásakor.
A fókuszcsoportos kutatás menetét a következő ábrán láthatjuk:

2/4. ábra: a fókuszcsoportos kutatás folyamatábrája

Forrás: saját szerkesztés

A **felméréses vizsgálatok** mára az alapvizsgálatoknak számítanak a sporttudomány területén, hiszen leginkább az alkati (pl.: testmagasság, testsúly), és készségi (pl.: tenyeres pörgetés végrehajtásának értékelése 1-5-ig) és képességi (pl.: Cooper teszt) tulajdonságokra irányulnak. A legfontosabb a módszer alkalmazása során a megfelelő próbák, illetve a kvantitatív mérés és értékelés pontos kritériumrendszerének kiválasztása.

Az **írásbeli kikérdezésnél** a kérdező személyes jelenléte nem szükségszerű, hiszen az információszerzés írásban közvetített feladatokra (pl. postai kikérdezés, számítógépes kérdőív) történik. Az írásbeli megkérdezés lényeges eleme, hogy a kérdőívet egy udvarias magyarázó kísérőlevéllel együtt juttatjuk el a megkérdezetthez, az ugyanis a kérdezőt helyettesíti, annak kell ösztönöznie a személyt a kitöltésre. Alapvető adatokat és információkat kell tartalmaznia: a kutatást végző cég neve, bonyolult kérdőív mellé kitöltési útmutató. A megkérdezés sikere, a visszajuttatási arány nagyban függ a kísérőlevéltől. Gyakran nyereményszelvényekkel ösztönzik a megkérdezetteket a válaszadásra és visszajutatásra.

Napjainkban a *kérdőíves kutatások módszere* a legelterjedtebb, hiszen számos fajtája létezik. Népszerűségét annak is köszönheti, hogy gyors, gazdaságos, általa nagy adatmennyiség szerezhető és viszonylag objektív. A módszer előre meghatározott összetételű személyeket kérdezz meg, valamilyen „segédeszköz” vagy személy közbeiktatásával.

A kérdőív egy olyan dokumentum, amely információk megszerzésére álló kérdések halmazából tevődik össze. A kérdőíves módszernek vannak előfeltételei:

1. A vizsgált téma alkalmas legyen a megkérdezésre.
2. Pontosan, egyértelműen meg kell határozni a kérdezés célját. Tudatosan akkor lehet a megkérdezés módszereivel élni, ha előre egészen pontosan tudjuk, mit akarunk kérdezni és miért, hiszen ennek alapján állítjuk össze a kérdőívet.
3. Biztosított legyen a megkérdezettek kompetenciája. A kutatónak el kell döntenie, hogy azok a személyek vagy csoportok, akiket meg fogunk kérdezni, tudnak-e érdemben válaszolni a kérdésekre.
4. Biztosított legyen a reprezentatív minta tagjainak szabályszerű kiválasztása. A jó kiválasztás lényege, hogy a mintasokaság összetételében feleljen meg az alapsokaság összetételének, és így az általuk adott válaszok általánosíthatók az alapsokaság egészére.
5. Megfelelően elő kell készíteni a megkérdezést. A kérdőívvel történő megkérdezés munkaidő- és munkaerő-igényes. A megkérdezés zökkenőmentes lebonyolítása érdekében időtervet, programot kell készíteni. Fontos, hogy a feladatnak megfelelő kérdező és feldolgozó apparátus álljon rendelkezésre.
6. A megkérdezés eredményei alkalmasak legyenek a feldolgozásra. A vizsgálatnál általában nagyszámú mintával dolgoznak. Ezért fontos, hogy a válaszok számszerűsíthetők legyenek, még akkor is, ha nyitott kérdés is van a kérdőívben.
7. Legyen meg a megfelelő tudás, program az adatok feldolgozásához.

A megkérdezéssel lényegében a különböző módszerek alapján kiválasztott mintasokaságtól előre meghatározott kérdésekre számszerűsíthető, feldolgozásra alkalmas válaszokat nyerünk. A megkérdezés formái változatosak, s ezért az elérendő célnak megfelelően kell kiválasztani a legalkalmasabbat.

Kérdőívet lekérdezhetünk *írásban, szóban, telefonon* és a *számítógép* (internet) segítségével is. Minden kutatás témája más, mindig változik a célcsoport, a kutatás mélysége, ezért a kérdőívek is különbözőek. Vannak szabályok, amelyeknek betartása kötelező, másoké ajánlott, mindez természetesen sok év tapasztalatai alapján, minden kutatónál egyénileg alakul ki.

Írásbeli kérdőív egy tipikus kvantitatív technika. Leggyakoribb formája a postai úton történő megkérdezés, amelynek lényege, hogy a megkérdezett a kérdezőbiztos jelenléte, segítsége, befolyása nélkül tölti ki a postai úton vagy személyesen megkapott kérdőíveket, és postai úton vagy egy meghatározott helyre juttatja azt vissza.

Az írásbeli megkérdezésnél a posta bevonásával bérmentesített - a kutatást végző cég, személy, intézmény nevére megcímzett - válaszborítékot is mellékelni ajánlott. Napjainkban egyre elterjedtebb kérdezésfajta az újságban, folyóiratban elhelyezett kérdőív. Olcsó, gyors, de kevésbé hatékony, annak ellenére, hogy a kitöltést és visszaküldést ajándékozással bátorítják. Bevált módszer a boltokban, kiállításokon átadott kérdőív, amely viszonylag rövid, jól megválaszolható, a helyszínnel összefüggő témájú kérdéseket tartalmaz.

Előnyei: nagyszámú sokaság esetén jól alkalmazható. Gyorsan eljuttatható, van idő az átgondolásra, kitöltésre, a válaszoló őszintébb lehet, mert a kérdezőbiztos nem befolyásolja a megkérdezettet. Alacsony költségigényű.

Hátrányai: alacsony a visszaérkezési arány (8-25%), ebből adódóan a reprezentáció nem megfelelő. Emlékeztető akciók szükségesek. Viszonylag kevés kérdés tehető fel. Nagyszámú a hibásan kitöltött, félreértett kérdőív, (visszaérkezett kérdőívek 8-10%-a) amely abból adódik, hogy a kérdezőbiztos nincs jelen.

A „szóbeli kérdőív”, a legnépszerűbb, hiszen, alaposabb, pontosabb és az elképzelt reprezentációnak megfelelő válaszolást biztosít. A szóbeli megkérdezésnél fontos szerepe van a kérdőívnek és a szakszerű eligazításnak. A jó eredmények elérésében a legnagyobb szerepe a kérdezőbiztosnak van.

Fajtái:

1. hagyományos forma (PP „paper and pencil”- papírral és tollal), mely történhet otthon, vagy valamely direkt erre a célra elkülönített helységben (pl. egy áruházban külön bérelt teremben)
2. CAPI (Computer Assisted Personal Interviews) világszerte új technológia kérdezőbiztosok laptoppal dolgoznak.- Magyarországon most terjed el.

Előnyei: 100%-os kitöltés, amely a kérdezőbiztos jelenlétéből adódik. A válaszok kontrollja a kérdezés során megtörténik és az esetleges félreértések tisztázódnak.

Hátrányai: A személyes kontaktus veszélyei, mert a kérdezőbiztos személye pozitív vagy negatív irányba befolyásolja a megkérdezettet. A megkérdezettek esetleg nem merik bevallani tájékozatlanságukat a kérdezőnek, és úgy válaszolnak, ami nem fedi a valóságot.

A kérdezőbiztosok díjazása kérdőívenként történik, ezért az adatok megbízhatósága csak gyakori, szúrópróbaszerű ellenőrzésekkel biztosítható.

Telefonos kérdezés, azokban az országokban, ahol a telefonellátottság országosan megfelelő, igen kedvelt módszer. A válaszokat a kérdező a vonal mögött viszi fel a kérdőívre. Több esetben a válaszokat azonnal számítógépbe táplálják, így a feldolgozás gyors, az eredmények néhány percen belül rendelkezésre állnak.

Előnyei: Gyors, viszonylag olcsó, a szóbeli megkérdezés technikai eszközzel bővített mása. Fontos a kérdező hangja, stílusa.

Hátrányai: Szokatlan, családoknál kötött időben lehet alkalmazni. Így inkább intézményekben, boltok esetében népszerű. Gyakran tolakodó. Nem mindenki rendelkezik készülékkel.

*Számítógépes megkérdezés***,** összekapcsolt hálózat esetén, a képernyőn megjelenített kérdőív segítségével történik. A reprezentatívságot nehéz tartani, mivel még nem mindenki rendelkezik a világhálóval, ettől függetlenül a jövőben várhatóan óriási teret fog hódítani.

Előnye: gyors, olcsó, kényelmes, folyamatosan terjed

Hátránya: Nem biztos, hogy őszintén válaszolnak a megkérdezettek.

A kérdőívszerkesztés előkészítése során még egyszer pontosan meg kell határozni az adott megkérdezéssel elérendő célt, a kérdezéssel nyerhető válaszokat.

A **kérdőívszerkesztés**nél vannak íratlan **irányelvek**, melyeket nem árt betartani, ezek a következők:

1. Lehetőleg egyszerű, könnyen megválaszolható kérdéssel kezdjük. Jól alkalmazható ez esetben a zárt kérdés, melyre szóban, igen röviden; írásban, pedig többnyire aláhúzással lehet válaszolni.
2. Kevés kérdést tegyünk fel. Ha kérdőív hosszú, a kérdések száma növekszik és egyre bonyolultabb kérdések vannak, akkor az tapasztalható, hogy a megválaszolt kérdések pontossága is csökken.
3. Röviden, egyszerűen, egyértelműen kérdezzünk, és változatos kérdéstípusokat alkalmazzunk. A kérdőívben használt szavak, kifejezések mindenki számára érthetőek legyenek. Kerülni kell az idegen szavak, fogalmak használatát. Akkor tehetünk kivételt, ha egyedi, homogén csoportot kérdezünk meg. Ilyenkor használhatunk szakmai kifejezéseket.
4. A kérdéseknél minden lehetséges válaszlehetőséget tüntessünk fel. Ha kifejtős kérdést teszünk fel, hagyjunk elegendő szabad helyet a válaszadó véleményének kibontására.
5. A kérdések ne befolyásolják a válaszadót.

6. Fel kell hívni a figyelmet – a kérdőív felső szélén vagy a kísérőlevél szövegében – hogy a válaszadás önkéntes. Ugyanez elhangzik személyes megkérdezés esetén is. Tiszteletben kell tartani a válaszadó névtelenségét.
7. A válaszok alkalmasak legyenek a számszaki feldolgozásra.

A beérkezett válaszokat önmagukban vagy más kérdésekkel kombinálva összesítenünk kell, ezért célszerű, - pl. közvetlen számítógépre vitelnél - ha a kérdőíven kódkockákat alkalmazunk. Így egyszerűbb lesz a feldolgozás.

8. Könnyű legyen a kérdések megválaszolása. Egyszerű, ha aláhúzással, sorszámozással, egy-egy szó beírásával lehet kitöltetni.
9. A kérdőív felépítése logikus legyen. Nemcsak a kérdések megfogalmazása, de a sorrendje is fontos. Legyenek átvezető részek a kevésbé hangsúlyos és a fontosabb részek és témák között. A fő téma a kérdőív 2/3-nál legyen. Objektív és rutinszerű adatok, statisztikák a végére kerüljenek. Pl. a jövedelemmel kapcsolatos kérdések is a végén szerepeljenek.
10. A kérdőívnek fel kell keltenie az érdeklődést. Fontos szerepet játszik a jól szerkesztett kísérlővel is.

Ezután kerül sor a kérdőív kérdéseinek tartalmi és formai kimunkálására, a kérdések számának sorrendjének meghatározására, a kérdések cél szerinti csoportosítására, mely által a kérdőív logikusabb felépítésű lehet.

A *bevezető kérdések* a kérdőívek elején vannak, egyszerűek, hiszen céljuk, hogy a válaszadót felkészítsék a témára. Ezután az *átvezető kérdésekkel* tudunk az egyes logikai részek között váltani. A *szűrőkérdések* feladata, hogy logikailag összetartozó ismérveket összehozza (faktorba rendezze). Ezekkel a kérdésekkel a majdani feldolgozást és elemzést könnyíthetjük. A legfontosabb és leghangsúlyosabb a *tárgyköri kérdések* csoportja, melyeket az ellenőrző kérdésekkel is vizsgálhatunk. Az *ellenőrző kérdések* alkalmasak arra is, hogy megbizonyosodjunk, hogy a válaszadó valóban körültekintően dolgozott vagy igyekezett minél hamarabb letudni a feladatot.

A kérdőívekben szereplő kérdések típusainál két fő csoportot tudunk elkülöníteni: a *zárt* és *nyitott* kérdéseket.

Zárt kérdés típusúnak nevezzük azokat a kérdéseket, ahol a válaszlehetőségeket és válaszformákat előzetesen meghatározzák. Lehet: kétkimenetű, többkimenetű, fontossági skála, szemantikus differciál, Likert skála.

A *kétkimenetű zárt* kérdéstípusnál a megadott válaszlehetőségek száma csak és kizárólag kettő. Leggyakrabban a válaszadó nemére vonatkozó kérés és az igen-nem típusú válaszok tartoznak ide. Szeretett kérdéstípus, hiszen az adatrögzítése, kódolása (0, 1) egyszerű!

pl.: szereti Ön a FTC labdarúgó csapatát? A válaszát húzza alá!

Igen Nem

Többkimenetű zárt kérdések megegyeznek a kétkimenetű kérdéstípusoknál leírtakkal, csak itt több válaszlehetőséget adnak meg. Veszélye abban van, hogy jól átgondoltan kell a lehetőségeket megadni, figyelni kell arra, nehogy kimaradjon válaszlehetőség. Erre a kérdéstípusra nagyon kell figyelni, amikor a kérdőívet teszteljük. A kérdéstípusnak az előnye szintén az egyszerű adatrögzítésben rejlik.

pl.: Hol él? (tegyen „x”-et a megfelelő kockába!)

Megyeszékhely ☐
Város ☐
Falu ☐

Fontossági skála a kutató által kiválasztott tulajdonság fontosságának fokozatait tartalmazza, illetve ide tartoznak azok a kérdések is, melyek rangsorolást várnak. Hátránya, hogy a rangsor indokát nem adja meg, és talán a legnehezebb megválaszolni.

pl.: Mennyire fontos Önnek a mérkőzésen hordott sportfelszerelése?

Nagyon fontos ☐
Fontos ☐
Nem annyira fontos ☐
Egyáltalán nem fontos ☐

Rangsorolja a következő sportmárkákat 1-5-ig! (1-es a legrosszabb, 5-ös a legjobb)

Puma ☐
Reebok ☐
Adidas ☐
Nike ☐
Asics ☐

Minősítő skála valamely tulajdonság fokozatait minősíti a gyengétől a kiválóig. Előnye, hogy képet ad az érzelmek erősségéről. Hátránya, ha a skála rosszul kerül kialakításra, akkor a lehetőségek nem mindig értelmezhetőek.

pl.: Mennyire szeret úszni? (1-es utál; 7-nagyon szeret) ☐

A szemantikus differenciál a fogyasztói hozzáállás, attitűd mérésére szolgál oly módon, hogy *a* skála két végén ellentétes értelmű szavak vagy mondatok (állítások) szerepelnek. Az ellentétpárokra értelmezhető skála grafikusan is ábrázolható, a tesztszemély pedig kijelöli érzésének irányát és intenzitását. (A skálán szereplő értékek közül a megkérdezett teljesen szabadon választhat.) Az alapgondolat az volt, hogy az ellentétpárokban megtestesülő tulajdonságok, mint: szép–csúnya, jó–rossz, öreg–fiatal, hideg–meleg stb. rendelkeznek az általános érthetőség tulajdonságával. A szemantikus skálák kiválóan alkalmasak arra, hogy a kutatók képet kapjanak az egyes fogyasztók (csoportok) egy-egy termékről, reklámról, szolgáltatásról alkotott véleményéről. Mivel az attitűd, vélemény mérésére meghatározott ellentétpárok szolgálnak, rendkívül lényeges, hogy azok a termék, szolgáltatás, stb. szempontjából relevánsak legyenek. Leggyakrabban hétfokozatú skálával találkozunk.

pl.: Kérjük, helyezze el az alábbi skálán az Ön által kipróbált „X” nevű sportcipőt!

olcsó						drága
1	2	3	4	5	6	7

gyenge minőség					kiváló minőség	
1	2	3	4	5	6	7

A *Likert-skála* lényegében a szemantikus differenciál egyik speciális változata, „egyetértő skálának” is nevezik és a leggyakrabban alkalmazzák. E kérdéstípus esetén egy meghatározott témakörhöz kapcsolódó „állításlistát” kell a kérdezettnek értékelni, abból a szempontból, hogy az egyes kijelentésekkel milyen mértékben ért egyet vagy utasítja el őket. E skálatípust többnyire életmódkutatásoknál, piacszegmentálásnál, vállalati arculat kialakításánál stb. hasznosítják a kutatók. A skála lehet páros vagy páratlan kimenetszámú, a leggyakoribb és legtanácsosabb az 5 és a 7 fokozatú skála. A skáláról bővebben KEHL-RAPPAI [2006] tanulmányában olvashat az érdeklődő.

pl.: Kérjük, jelölje be a skálán a véleményével egyező pontokat! (1= egyáltalán nem ért egyet; 5= teljesen egyetért)

Egy sportcipőnél előnyös, hogy kényelmes legyen!

1	2	3	4	5

Egy sportcipőnél előnyös, hogy könnyű legyen!

1	2	3	4	5

vagy

Jelöld meg azt a választ, amelyik legközelebb áll az érzéseidhez!

	Határozottan egyetértek	*Egyetértek*	*Semleges*	*Nem értek egyet*	*Nagyon nem értek egyet*
Napjaim érdekesen telnek	☐	☐	☐	☐	☐
Magányosnak érzem magam	☐	☐	☐	☐	☐
Általában élvezem, amit csinálok	☐	☐	☐	☐	☐
Úgy érzem, szeretnek, és szükség van rám	☐	☐	☐	☐	☐
Képes vagyok lazítani	☐	☐	☐	☐	☐
Vannak terveim, céljaim a jövőmmel kapcsolatban	☐	☐	☐	☐	☐

A kérdőívben szereplő kérdések másik csoportját a *nyitott kérdések* alkotják. Meg lehet különböztetni: a *teljesen nyitott kérdéseket*, a *szótársításos (asszociációs) kérdéseket*, *mondat kiegészítéses kérdéseket*, *kép-kiegészítéses teszt jellegű kérdéseket*, *tematikus teszt jellegű kérdéseket.*

Teljesen nyitott kérdéseknél: a kérdés a válaszadónak teljes szabadságot ad a véleménye megfogalmazásában.

pl. : Mi a véleménye a doppingolásról?

Szótársítás (asszociáció) során a válaszadónak azt a szót/szavakat kell kimondania, amely az elhangzott szó/szavak hallatán először az eszébe jut.

pl.: Mi jut eszébe a dopping szó hallatán?

Mondat kiegészítéses kérdés: A megkezdett mondatot kell folytatnia a válaszadónak.

pl.: Azért nem doppingolnék soha, mert………

Történet kiegészítéses kérdés: A válaszadónak egy befejezetlen történetet kell folytatnia.

pl.: : Elmentem egy világversenyre és az öltözőben azt láttam, hogy az egyik sportoló doppingszereket fogyaszt, melyről egyből azok a gondolatok jutottak eszembe, hogy……

Képkiegészítéses teszt: A válaszadó előtt elhelyezett képen található két ember közül az egyik állít valamit, a válaszadónak az üres részbe kell írni a másik ember véleményét.

Tematikus észlelési teszt (TAT): Arra kérik a válaszadót, hogy mondja el, hogy mit lát az elé tett képen, mi történik.

A kérdések összeállításánál ügyeljünk tehát az alábbiakra:

1. átgondolt kérdéseket tegyünk fel
2. konkrét kérdéseket alkalmazzunk
3. használjunk egész mondatokat
4. kerüljük a rövidítéseket
5. kerüljük a szlenget
6. kerüljük a szakmai kifejezéseket
7. kerüljük az előítéletes kifejezéseket
8. kettős kérdéseket ne tegyünk fel
9. negatív kérdéseket ne tegyünk fel

A kérdőívet éles „használat” előtt feltétlenül *ki kell próbálni*, ha kell *módosítani*, majd *véglegesíteni*, hiszen lehetnek kérdések, melyek esetleg nem egyértelműek és nem világosak. A felülvizsgálat azt a célt szolgálja, hogy megállapítjuk, alkalmas-e a kérdőív a vélemények, szándékok, indítékok, előítéletek felszínre hozatalára. A feldolgozási terv alapján még egyszer ellenőrizni kell, hogy minden olyan kérdés, összefüggés szerepel-e a kérdőívben, amelyre feltétlenül szükség van a kutatás szempontjából. A többszöri ellenőrzés után lehet csak a kérdőívet véglegesíteni és sokszorosítani, és a megkérdezés tényleges műveletét megkezdeni.

Összefoglalva elmondható, hogy egy kérdőív készítése három lépésben történik :

1. Előkészítő részben: ismételten meg kell pontosan határozni a kutatás célját és a várhatóeredményt, valamint össze kell gyűjteni azokat az információkat, amelyek a kérdőív szakszerű összeállításához szükséges.

2. A fő rész a kérdőív szerkesztése, a kérdések sorrendiségének tükrében:.

- a legáltalánosabb résszel kezdünk, mert ezzel elkerülhető, hogy egy nehezen megválaszolandó kérdéssel már az elején összezavarjuk a megkérdezettet
- a bevezetést gondosan, udvariasan és a témához illeszkedően kell összeállítani, mert az meghatározza a további menetet
- a kérdőív felépítése olyan legyen, hogy az egyik részből (témakörből) legyen egy átvezetés a másikba
- az egyes kijelentések és állítások úgy következzenek egymás után, hogy az megfeleljen a válaszadó logikájának és könnyen értelmezhető legyen
- a legkényesebb részeket a kérdőív lezáró, összefoglaló részébe tegyük
- a válaszadóra vonatkozó személyes adatokat (szegmentációs ismérveket) a leggyakrabban az elején vagy a végén kérdezzük meg. Ha a végén kérdezzük, és már nem töltik ki, akkor esetlegesen a többi adat még használható.

3. **A záró rész, a kérdőív kipróbálása, véglegesítés, sokszorosítása.**

A másik nagy módszercsoport: a **feltáró, értékelő módszerek** csoportja, ahol a feltáró módszerekkel nyert információkat két fő szempont szerint értékelhetjük. A minőségi (kvalitatív) értékelésnél az eredmény tartalmát, és azok sajátosságait vizsgálva, a kapott eredmények pontos rögzítése és tartalmi kategorizálás révén. A mennyiségi (kvantitatív) értékelés során, az eredményekben rejlő mennyiségi mozzanatokat mutatjuk be, a leíró és következtetéses statisztika módszereivel.

A *kvalitatív értékelés* során eredmények ritkán számszerűsíthetők, nem mérhetők. A kvalitatív értékelést általában akkor alkalmazható sikeresen, amikor a különböző viselkedésformák, magatartásbeli sajátosságok mozgatórugóit igyekeznek feltárni.

A *kvantitatív értékelések* azon alapulnak, hogy akár az emberi hozzáállás, magatartás is mérhető, tehát számszerűsíthető, továbbá az így nyert adatok statisztikai módszerekkel elemezhetők, értékelhetők.

2.5. A vizsgálni kívánt minta meghatározása (Pintér-Rappai-Herman-Rédei nyomán)

A tudományos kutatás tárgyát képző egyedeket, a statisztika módszereivel számba vehetjük, megfigyelhetjük. Amennyiben a sokaság egészének megfigyelését valamilyen előre elhatározott szempont szerint végezzük, teljes körű megfigyelést hajtunk végre (klasszikusan teljes körű megfigyelés, a népszámlálás). Ha a megfigyelés a sokaságnak csak meghatározott egyedeire terjed ki, részleges megfigyelést végzünk. A részleges megfigyeléseken belül kiemelkedően fontos szerepe van a reprezentatív megfigyelésnek. A részleges, de különösen a reprezentatív megfigyelés során az alapsokaság valamennyi eleméről nincsen információnk, így az ismert részsokaságból (mintából) próbálunk közelítő megállapításokat tenni az alapsokaság egészére[13]. Könnyen belátható, hogy annál sikeresebb lesz ez a következtetés, minél inkább „emlékeztet” a kiválasztott minta az alapsokaságra. Mindezen fejtegetés két új fogalom bővebb kifejtését igényli: **sikeresség** és **emlékeztetőség**. A következtetés sikerességét a statisztika tudománya két – egymásnak, mint később látni fogjuk, ellentmondó – kategóriával méri:

- a siker egyik „látható” jele a **megbízhatóság**, azaz az esetek nagy százalékában „bejön”, vagyis a minta alapján tett állítás vagy meghatározott érték a valóságban sokszor bizonyul helytállónak;
- a másik siker-kritérium a **pontosság**, vagyis az a kívánalom, hogy a minta alapján nyert információ legyen valóban informatív, a becslés eredményét jelentő értékek egy szűk sávban mozogjanak.

Ebben a tankönyvben nem térünk ki arra, hogy mely tulajdonságok összessége végett mondható el egy mintáról, hogy reprezentatív, az érdeklődő ennek részletes leírását megtalálhatja PINTÉR-RAPPAI [2001] tanulmányában.

Jogosan kérdezhetjük, miért nem vizsgáljuk az alapsokaságot (pl. az igazolt sportolókat), hiszen a legpontosabb eredményt így kapnánk. Sokféle válasz közül a leggyakoribbak, hogy nagyon időigényes és költségigényes lenne, illetve jogilag is akadályokba ütközhetünk. Mindenképpen meg kell jegyezni, hogy a sport területén az ilyen jellegű alapstatisztikai adatok nagyon hiányosak (pl. nincs adat az igazolt sportolók számáról), ezért nagyon nehéz (szinte lehetetlen) a reprezentativitást biztosítani.

---

[13] Pintér- Ács (2006)

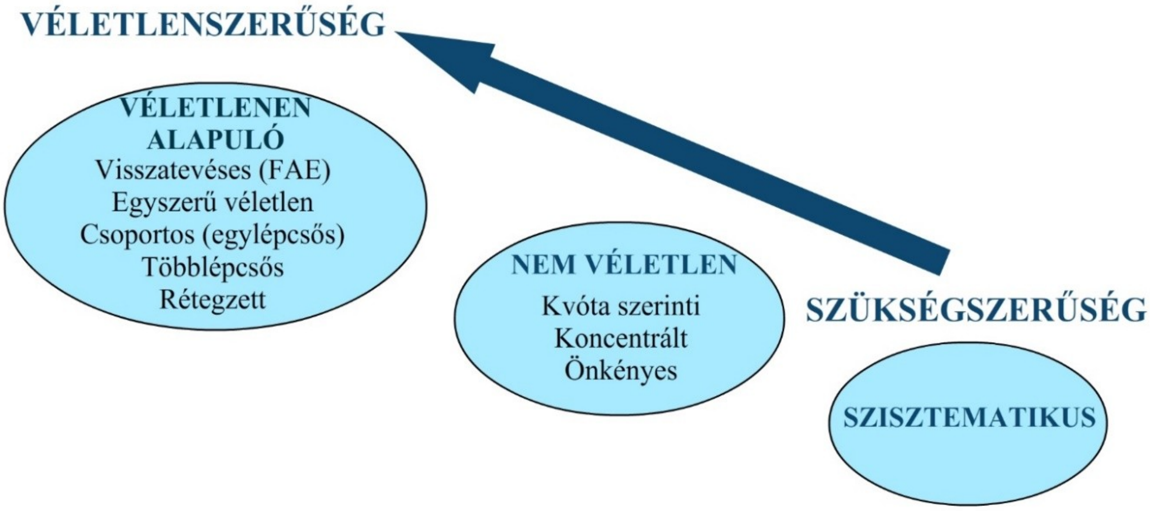
A mintavétel esetén mindig statisztikai hibákkal kell számolni, mely a mintavételi és nem mintavételi hibákból állnak. A mintavételi hiba abból adódik, hogy a sokaság egésze helyett, annak egy részét vizsgáljuk. A nem mintavételi hibák leggyakrabban az adatfelvételből keletkeznek (válaszadási hiba, végrehajtási hiba, feldolgozási hiba, stb.)

A mintavételi eljárásokat sokféleképpen csoportosítja a szakirodalom egy része (Babbie, 2000) beszél a valószínűségi és nem valószínűségi mintavételi módszerekről.

Valószínűségi minta, a valószínűség-számításnak megfelelően - leggyakrabban a véletlen kiválasztásos-módszerrel vett minták általános kifejezése. Pl.: egyszerű véletlen, rétegzett mintavétel.

Nem valószínűségi minta, mely során nem a valószínűségi mintavételi szabályokat követjük. Pl.: kvótás, hólabdás (görgetett) mintavétel.

A reprezentatív mintavételnek számos módja (módszere) ismert, melyek különböznek részben a lebonyolítás módja, részben a mintavételben érvényesülő véletlenszerűség alapján. Az alábbi ábra a mintavételi eljárások egyfajta csoportosítását mutatja.



2/5. ábra: A leggyakoribb reprezentatív mintavételi eljárások

Forrás: Pintér-Rappai: Statisztika (2007)

A mintavételi eljárások csoportosítása a fenti ábrán a véletlenszerűség csökkenése, mint rendező elv, alapján történt. Ugyanakkor az ábrán „jobbra-lefelé” haladva azt érzékeltetjük, hogy bizonyos egyedek mintába kerülése egyre inkább szükségszerű. Tekintsük át a leggyakrabban alkalmazott mintavételi módokat!

Egyszerű véletlen (EV) mintavétel, legtöbbször hivatkozott eljárás. Itt a minta, elemeit azonos valószínűséggel és egymástól független módon választottunk ki. A populáció minden tagjának egyforma esélye van a mintába való kerülésre. Először is a teljes populáció listájára van szükségünk, majd kalapba helyezve a neveket, vagy a véletlen számok táblázatát felhasználva erről a „listáról” kiválasztjuk a szükséges számú mintát. Ennek egy fajtája a *visszatevés nélküli* mintavétel, ahol egy elem egynél többször nem kerülhet a mintában. pl.: lottósorsolás .

A **visszatevéses mintavétel**, olyan minta, melynél minden elemet visszatettünk a következő elem mintába kerülése előtt. Az eljárás során szükségünk van egy az alapsokaság valamennyi egyedét felsoroló lajstromra, amelyből teljesen véletlenszerűen kiválasztjuk az első egyedet. Erre az eljárásra a gyakorlatból példát hozni nehéz, de ilyen mintavétel történik, amikor a szóbeli vizsgák alkalmával az oktató minden felelet után visszateszi a tételt. Előfordulhat, hogy egy tétel többször is elhangzik az adott vizsganap alkalmával.

Szisztematikus mintavételi technika, lényege, hogy a felsorolás pl. minden 5-ödik elemét veszik ki az alapsokaságból a mintába. (ez a mintavételi intervallum: az elemek közötti távolság) Azt, hogy ne lehessen bizonyos torzítás a mintában, az első számot véletlen módon választjuk ki jelen esetben az első 5 számból, utána pedig minden 5-ödik kerül a mintába. A kiválasztási arány a mintába kerülő elemek aránya az alapsokasághoz képest, itt: 1/5. A módszer veszélye: ha a mintavételi keret valamilyen sorrend szerint csoportokból áll, periodikus. Pl.: Kosárcsapatokat kérdeznek meg, és a csapatokon belül posztok szerint vannak felsorolva a játékosok, és mi minden ötödiket választjuk ki, akkor mind egy posztról kerül majd a mintába, pl. mind center lesz.

A rétegzett mintavétel az egyszerű véletlen és a szisztematikus mintavétel módosítása, mellyel tovább növelhető a minta reprezentativitása. A lényege az, hogy megfelelő számban kerülhessenek elemek az alapsokaság bizonyos vizsgálni kívánt egynemű részcsoportokból. A legegyszerűbb – bár nem az egyetlen – módja a rétegzett mintavételnek, ha a rétegképző ismérv alapján csoportosított sokaság minden csoportjából előre meghatározott arányban, egyszerű véletlen módszerrel kiválasztunk adott elemszámot. A rétegzés attól függ, milyen információink vannak a listáról. pl. ha kézilabdázókat nézünk, hogy meccsenként hány gólt dobnak, de előtte posztok szerint rétegzést készítünk, hiszen a gólok száma függ(het) a poszttól.

A **csoportos (egylépcsős) kiválasztás** az EV mintavétel azon kellemetlen tulajdonságát kívánja kiküszöbölni, miszerint az alapsokasági egyedek lajstromba foglalása nehézkes. Ez esetben a gyakorlathoz közelítünk annyiban, hogy nem a ténylegesen felmérendő egyedekről, hanem csak azok bizonyos lényeges tulajdonságáról (ismérvéről, hovatartozásáról) kell teljes körű információval rendelkeznünk. Az ezen, pontosan ismert tulajdonsággal rendelkező ún. **elsődleges mintavételi egység**ekből egyszerű véletlen mintát veszünk, majd az így – véletlenszerűen – kiválasztott csoportok valamennyi egyedét felmérjük (anélkül, hogy a mintavétel előtt tudtunk volna a létezésükről!). Pl.: Az Önkormányzati és Területfejlesztési Minisztérium sportért felelős szakállamtitkárának megbízásából feladatot kaptunk, hogy felmérjük a sportolók iskolai végzettségét. Ahhoz, hogy ezt el tudjuk végezni kérdőívet szerkesztünk, melyet a sportolók 5%-ával akarunk kitöltetni. Tudjuk, hogy 2007-ben Magyarországon 2 280 000 sportoló volt, ám őket név szerint nem ismerjük. Amit tudunk, hogy ebben az évben 2780 sportegyesületet tartottak számon. A mintavételt ezért csoportos módszerrel bonyolítjuk: első lépcsőben kiválasztunk egyszerű véletlen módszerrel 139 sportegyesületet ($139 \cong 0{,}05 \times 2780$), és ezekben az összes sportolóval megíratjuk a tesztet. A mintavétel során az elsődleges mintavételi egységekről, vagyis az egyesületekről kell felsorolással rendelkeznünk, a sportolókról nem!

Többlépcsős (kétlépcsős) mintavételi technika azt jelenti, hogy az elemek csoportjaiból veszünk mintát, majd második lépésben az így kapott csoportokon belül veszünk újabb mintát. A többlépcsős csoportos mintavétel jelentése, hogy egymás után többször készítünk listát és alkalmazzuk a kiválasztást.

A **kvóta szerinti kiválasztás** nagymértékben hasonlít a rétegzett mintavételre, hiszen ebben az esetben is használunk segédinformációkat. Lényege, hogy területileg bontva meghatározzák adott lakókörzetek lakók szerinti összetételét és a legfontosabb jellemzők szerint kvótát (listát) állítanak össze a felkeresendő alanyokról. A mintavétel végrehajtója a kvóta ismeretében valamennyi olyan egyedet felméri, amelyik megfelel a kijelölt tulajdonságoknak, mindaddig, amíg az előre adott kvótája nem teljesül.

A **koncentrált kiválasztás** még tovább csökkenti a véletlen szerepét, illetve növeli a kiválasztó személy „felelősségét” a mintavétel során. Az eljárás lényege, hogy súlyozzuk az egyedeinket, és azokat kérdezzük le, akik véleményvezetőnek (opinon leader) számítanak.

Az **önkényes (szakértői) mintavétel** esetében a kérdezőbiztos maga határozza meg, hogy mely egyedek kerülnek a mintába. Az egyetlen megkötés, amely befolyásolja a döntését, hogy mekkora elemszámot kell kiválasztania. Pl. a vezérszurkolókat akarom megkérdezni valamiről, és ezért mindig nagydarab, kigyúrt, kopasz személyeket választok a mintába, mert sejtéseim szerint ők azok.

Egyszerűen elérhető alanyokra vonatkozó mintavétel: lényege: a kutató csak olyan elemeket (embereket) vizsgál, akik könnyen elérhetőek. Reprezentatív adatokat ritkán érünk el vele, csak akkor jó, ha ténylegesen azok az emberek érdekelnek, akiket vizsgálok. Pl.: egy meccs előtt sorban állva megkérdezett 100 szurkoló véleménye, akik épp szembe jöttek a pénztárnál. (Nem reprezentatív az összes nézőre nézve)

„Hólabda” mintavétel a nehezen hozzáférhető populációknál a minta "ismeretségi lánc" nyomán gyarapodik.

2.6. A kutatás végrehajtása

A kutatás végrehajtásáról, tervezéséről és lefolytatásáról az előzőekben már bőven esett szó, ezért ebben a részben további gyakran használt gyakorlati módszer (*Swot-analízis*) részletes bemutatása következik

A Swot-analízis, az egyik legkedveltebb módszer, melyet a sporttudományos kutatások egyes területein is gyakran használnak. Leginkább valamely intézmény, szervezet helyzetelemzésére, problémamegoldásra hasznosítható, de számos egyéb területen, helyzetértékelés céljából (kiindulópont) alkalmaznak.

Erősségek	**S**trength
Gyengesége	**W**eaknesses
Lehetőségek	**O**pportunities
Veszélyek	**T**hreats

Ez a négy szó által négy szempont jön létre, mely segítségével csoportosítani lehet mindazon információkat, amelyek rendelkezésünkre állnak. A módszer segítségével a kutató megvizsgálhatja a működésben érintettek véleményét, értékelheti a jelenlegi helyzetét, és megállapításokat tehet a jövővel kapcsolatos lehetőségekről, veszélyekről is.

Az adatgyűjtés tipikusan „brainstorming” technikával történik, de a gyakorlatban kérdőíves vagy interjú-technikával begyűjtött adatok elrendezésére is alkalmazható ez az eljárás. A lényeges szempont, hogy itt is a megkérdezett kiválasztásánál csak azok kerüljenek a mintába, akik az adott kérdéskörbe kompetensek, tapasztalataik vannak, és a kutató számára megfelelő információkat, ismereteket tudnak adni. Pl. A magyar labdarúgásról helyzetéről szeretnénk egy kutatást végezni, akkor a Swot-analízis során ne a röplabda edzők információit használjuk.

Az erősségeket és gyengeségeket közösen jellemzi, hogy a kutatás tárgyának belső jellemzőiből fakadnak, míg a lehetőségek és fenyegetések a környezetéből erednek.

Erősségek: Mit tartanak a megkérdezettek a kutatás tárgyával kapcsolatosan jónak, mi lehet az erőssége! Leggyakrabban olyan jellemzők kerülnek ide, mint: tapasztalat, tudás, felszereltség, hagyomány, szervezettség, stb. Közös jellemzőjük, hogy a jövőben rájuk lehet építeni.

Gyengeségek: Az itt található adatok képezik a fejlesztendő területeket, ahol a megkérdezettek szükséges változtatásokat említenek. Pl. nehéz megközelítés, rossz felszereltség, sportszerek hiánya, stb.

Lehetőségek: Ennek a megállapításához elengedhetetlen a kompetens válaszadók megtalálása, hiszen ez a környezet kiváló ismeretét igényli. Pl. új edzőtermek megjelenése, új sportszolgáltatás igénye, új wellness központ nyitása, stb. A lehetőségek feltérképezése, jelentősen befolyásolhatja a a jövőbeli eredményeket.

Veszélyek (fenyegetések): Ez a lehetőségekhez hasonlóan egy független tényező, hiszen az előző példánál maradva pl.: az új sportszolgáltatás iránti érdeklődés hirtelen csökkenése.

Célszerű az információk rendszerezéséhez adatlapot használni.

Erősségek	**Gyengeségek**
Lehetőségek	**Veszélyek**

2/1. táblázat: a Swot-analízis táblázata

A SWOT-táblázat első két dobozának feltöltése általában nem okoz túl nagy nehézséget. A legtöbb buktató a táblázat *lehetőségek* és *veszélyek* dobozainak kitöltésében rejlik, hiszen gyakran lehetőségek közé olyan jellemzők is bekeverednek, amelyeknek valójában az erősségek között kellene, hogy szerepeljen. A különbség az, hogy belső (pl. szervezeten belüli), vagy a környezetből jövő külső, még kihasználatlan, de beépíthető forrásról van szó. Természetesen hasonló helyzet lehet a veszélyek és gyengeségek között is. Tudatosítani kell, hogy a gyengeségek a belső tulajdonságokból fakadnak.

2/6. ábra: A Swot-analízis folyamatábrája

Forrás: saját szerkesztés

I. *Az előkészítés fázisa*:

a) Ki kell választani a megismerésre kijelölt kutatási területet
b) El kell dönteni, hogy kiknek a véleményét akarjuk feltérképezni ezzel a módszerrel. pl. edzői vélemény, klubvezetés hivatalos álláspontja, stb.
c) Meg kell tervezni a swot-analízis adatfelvételének körülményeit: időpontját, helyszínét, a résztvevők számát, a munka egyéb feltételeit.
d) Fel kell kérni együttműködésre a résztvevőket: tájékoztatni kell őket az adatgyűjtés céljáról, a további felhasználásról.
e) Konkrét időpontot kell egyeztetni.
f) Be kell rendezni a helyiséget a szükséges módon, majd elő kell készíteni a megfelelő eszközöket.

II. Az információgyűjtés fázisa:

Első lépésben az előre kiválasztott - minden szempontnak megfelelő- jelenlévők leültetése történik egyénileg vagy csoportokban. Ezután mindenki elé helyezzünk megfelelő íróeszközt és papírt, esetlegesen magát az előre sokszorosított swot-táblázatot. Ezt követően történik az összejövetel céljainak és módszereinek előszóval történő bemutatása, illetve a táblázat kitöltésének ismertetése. Ezután a táblázatot részenként kezdjük feltöltetni, oly módon, hogy először az egyszerűbb szempontokra (erősségek- gyengeségek) legyünk kíváncsiak, melyet kövessen a környezetről (lehetőségek, veszélyek) alkotott vélemények lekérdezése. Ilyenkor a kutató személyes hangvétellel tegyen fel „gondolatokat-ébresztő" kérdéseket. Kérjük meg a válaszadókat az őszinte válaszokra, véleményekre, valamint javasoljuk, hogy a négyzetekbe hasonló számú információt tüntessenek fel.

III. A kiértékelés fázisa

A kiértékelés fázisa az adatok összesítésével veszi kezdetét, mely történhet azonnal, illetve egy későbbi időpontban is. A későbbi időpontban történő kiértékelés időigényesebb és az eredményekről a kutatásban résztvevő alanyok sem értesülhetnek.

A helyszínen történő összesítés oly módón történhet, hogy valamely segédeszközön (pl. mágnestábla) jól látható swot-táblázatba feltüntetjük a véleményeket. Általános szabályként leszögezhető: a kérdések megválaszolásakor, a dobozok kitöltése során törekednünk kell arra, hogy minél konkrétabb, megfoghatóbb legyen minden értékelő kijelentésünk.
Legegyszerűbb, ha többször elhangzó megállapításokat egyszerűen súlyozzuk a gyakorisággal, hiszen a majdani tényleges megoldási javaslatokat a súlyozásnak megfelelően priorizálhatjuk.

2.7. Az adatok elemzése, általánosítások megfogalmazása.

Ebben a részben a sporttudományi kutatások alkalmával gyűjtött információk feldolgozásának, elemzésének statisztikai alapmódszereit kívánjuk bemutatni. Napjainkban már a kutatások adatainak értékelése, elemzése elképzelhetetlen a számítógépek használata nélkül - és valljuk meg roppant mód, meg is könnyíti a munkát- ezért ebben a fejezetben az egyszerűbb kézi számításokon túl, a leggyakoribb Excel és SPSS adatelemzési technikákat, és outputértelmezéseket is tárgyalni fogjuk.
A **statisztika** a tömegesen előforduló jelenségekre, folyamatokra vonatkozó információk összegyűjtésének, leírásának, elemzésének, értékelésének és közlésének tudományos módszertana. A nemzetközi irodalomban elterjedt csoportosítás szerint megkülönböztethetjük *a leíró statisztikát, a következtetéses statisztikát* és a *statisztikai döntéselméletet*.
A **leíró statisztika** alapvetően a numerikus információk összegyűjtését, az információk összegezését, és tömör jellemzését szolgáló módszereket foglalja magában. A leíró statisztika legfontosabb területei: az adatgyűjtés, az adatok ábrázolása, az adatok csoportosítása, osztályozása, az adatokkal végzett egyszerűbb aritmetikai műveletek, az eredmények megjelenítése. E területen jellemzően egyszerűbb statisztikai módszereket használunk és az alapsokaságra vonatkozó, a vizsgálat szempontjából fontos adatokat maradéktalanul használjuk fel. A leíró statisztikai eljárásokat gyakran az Excel program segítségével számítják, hiszen kezelése könnyű és világos.
A **következtetéses statisztika** segítségével a jelenségekre, folyamatokra vonatkozóan olyan megállapításokat tehetünk, amelyek nem csak a közvetlen megfigyelésen alapulnak. Igen leegyszerűsítve, alkalmazásával közvetlenül nem mérhető, csak összetett statisztikai, matematikai eljárásokkal megszerezhető számszerű információkat nyerhetünk. A következtetéses statisztika szorosan épít a matematikai- statisztikára és a valószínűségelméletre, fontos, hogy a következtetések itt mindig valamely mintából

(részsokaságból) származnak. Két általunk tárgyalandó része: a becslés és a hipotézisellenőrzés, melyet az Excel és az SPSS programmal is szemléltetünk.

A **statisztikai döntéselmélet** a véletlen környezet által bekövetkező események figyelembevétele mellett, több cselekvési lehetőség közül az optimálisnak vélt kiválasztásához ad számszerű információkat. Az empirikus statisztikai megfigyeléseken, és a következtetéseken túl szubjektív, szakértői értékítéleteknek is teret enged. A valószínűségelmélet és a játékelmélet elemeit kombinálja a statisztikai megfigyelések eredményeivel.

2.7.1. Statisztikai alapfogalmak, skálák

Statisztikai sokaságnak nevezzük a statisztikai megfigyelés tárgyát képező egyedek összességét. A sokaság legkisebb részeit, **egyedeit** *megfigyelési egységeknek* nevezzük. A statisztikai sokasággal szervesen összefüggő fogalom az ismérv.

Statisztikai **ismérvnek** nevezzük a statisztikai sokaság egyedeire vonatkozó tulajdonságokat, jellemzőket. Az ismérv lehetséges kimenetelei az *ismérvváltozatok.* Általánosságban az ismérvek lehetnek: *mennyiségi, minőségi, időbeli, és területi* ismérvek. Amennyiben egy ismérv csupán két ismérvváltozattal rendelkezik, **alternatív** ismérvről beszélhetünk (pl.: a nem csak két ismérvváltozattal rendelkezik úgy, mint férfi vagy nő). Az ismérvek megjelölésére gyakran használják az változó kifejezést is. A **mennyiségi ismérv vagy mennyiségi változó** (méréses jellemző) az ismérvváltozatokat számokkal, míg a **minőségi ismérv vagy minőségi változó** (minősítéses jellemző) minőségi jegyekkel, fogalmakkal, szavakkal jellemzi. Tudva levő, hogy a kutatások során keletkező adatokat leggyakrabban mérés útján kapjuk. A gyakorlati adatelemzéshez, főleg a számítógéppel támogatott feldolgozások esetén a változók mérési szintjét, illetve a mérési skálákat pontosan meg kell határozni.

A mérési skálák típusai:

- nominális (névleges) skála,
- ordinális (sorrendi) skála,
- intervallum skála,
- arányskála.

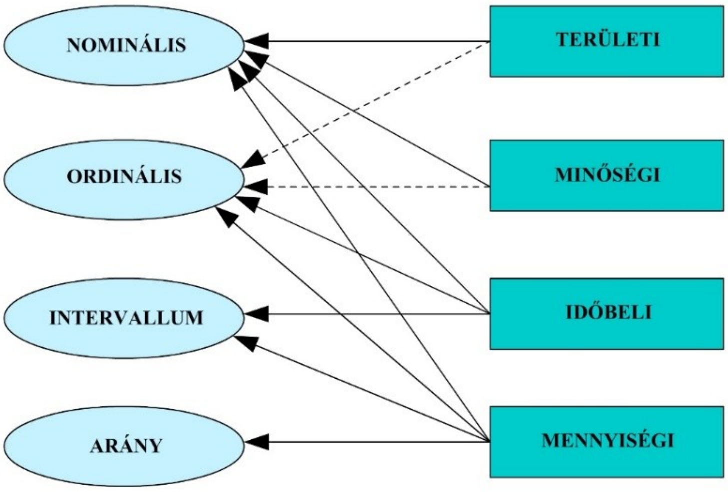
Nominális (névleges) skála a legegyszerűbb skála, mely kevés információt szolgáltat. Az egyedek osztályozása, csoportosítása csak megkülönböztetésre szolgál. A skálán az értelmezhető, hogy a megfigyelési egyedek egyenlők, vagy különbözők és a skálához tartozó értékekkel műveleteket elvégezni (pl. kivonás, osztás) nem lehet. A megfigyelési egységhez rendelt kódokat önkényesen választják.

Ordinális (sorrendi) skála megkülönböztet és sorrendet is mutat. A sorrendiségre vonatkozó relációk alapján rangsorba rendezzük a megfigyelt objektumokat, egyedeket. A sorrendi különbséget megállapítja, azonban az egymástól mért pozíciókat (pl. mennyivel jobb az egyik labdarúgó, mint a másik) nem mutatja.

Intervallum skálát metrikus skálának is nevezik. A skála alkalmas arra, hogy mérje, a különbséget a két érték között, így megmondhatja, hogy az egyik mennyivel nagyobb, jobb, vagy szebb a másiknál. Az intervallum-skálának nagyon jellegzetes tulajdonsága, hogy nem rendelkezik igazi zéró ponttal, vagyis a zéró érték nem jelenti a tulajdonság hiányát.

Arányskála, szolgáltatja a legtöbb információt, legmagasabb mérési szintet. Bármely két skálaérték arány értelmezhető. A skálának van igazi zéró pontja, ami azt jelenti, hogy a nulla érték a tulajdonság hiányát egyértelműen jelzi. A skálán bármely matematikai művelet elvégezhető.

A két metrikus (intervallum - és arányskála) skála között különbség a zérus meghatározásából áll. Hiszen az intervallum skálán a zérus meghatározás önkényes. Gondoljunk a gólkülönbségre, hiszen mindenki tudja, hogy mit jelent a +3-as gólkülönbség (a lőtt gólok száma hárommal több, mint a kapott gólok száma), és mit jelent a 0-s (pl. 20 gól lőttünk, és 20-at kaptunk), ez alapján létezhet negatív gólkülönbség is (kapott gólok száma meghaladja a lőtt gólokét).



2/7. ábra: Az ismérvek és mérési skálák összefüggése

Forrás: (Pintér-Rappai, 2007, 31. o.)

Fontos megjegyezni, hogy a legfejlettebb az arányskála, melyet az elvégezhető aritmetikai műveleteknek a száma is mutat. Belőlük bármely skálatípus transzformálható, előállítható. Minél fejlettebb egy skála, annál részletes elemzés, összehasonlítás elvégzésére alkalmas. „A skálatípusok azonosítása nagyon fontos, ugyanis egyértelműen meghatározza, hogy milyen elemzéstípusokat tudunk végezni, tehát az, hogy független vagy függő változók vannak metrikus vagy nem metrikus skálán mérve, jelentős eltéréseket okozhat."(Sajtos L.-Mitev A., 2007, 25. o.)

2.7.2. Leíró statisztikai elemzések

A leíró statisztika (descriptive statistics) legfontosabb területei: az adatgyűjtés, az adatok ábrázolása, az adatok csoportosítása, osztályozása, az adatokkal végzett egyszerűbb aritmetikai műveletek, az eredmények megjelenítése. Célja, hogy adott pillanatban leírjuk valaminek az állapotát.

Ebben a részben a következőket kívánjuk részletesebben áttekinteni (egyváltozós elemzések):

- viszonyszámok,
- információsűrítés középértékekkel (átlagszámítás, helyzeti középértékek),
- szóródás és szimmetria,
- adatprezentáció eszközei.

2.7.2.1. Viszonyszámok

Viszonyszámnak nevezzük **két,** egymással kapcsolatban álló **statisztikai adat hányadosát**.

A viszonyszám általános definíciója:

$$V = \frac{A}{B}$$

ahol: V – viszonyszám,

A – viszonyított adat,

B – viszonyítási alap.

A viszonyszámok következő fajtáit különböztetjük meg:

- *Megoszlási viszonyszámok* a rész és egész viszonyát fejezik ki, ahol a részsokaság elemszámát (gyakoriságát) viszonyítjuk a teljes sokaság elemszámához (pl.: a baranyai kézilabda csapatok száma az ország összes kézilabdacsapatához viszonyítva).

- *Koordinációs viszonyszám,* melyben a két rész-adat viszonyítjuk egymáshoz (pl. egy egyesületben 10 felnőtt korú sportolóra jutó ifjúsági korú sportoló száma).
- *Az intenzitási viszonyszám* két különböző, de egymással kapcsolatban álló statisztikai adat hányadosa. (pl. 100 km-re jutó üzemanyag fogyasztás)
- *Dinamikus viszonyszám*, az időbeli összehasonlításoknál kedvelt elemzési eszköz. Azt az időszakot, melyhez viszonyítunk *bázisidőszaknak*, a vizsgálat tárgyát képező időszakot *tárgyidőszaknak* nevezzük. Két különböző fajtája van: *bázisviszonyszám* és a *láncviszonyszám*.

A *bázisviszonyszámnál* a bázisidőszak mindig egy választott, állandó érték, legtöbbször egy kiemelkedően fontos, nevezetes dátum, vagy az idősor első értéke.
A bázisviszonyszám képlete, ha az idősor kezdő időpontját tekintjük bázisnak.

$$b_i = \frac{y_i}{y_0}$$

A *láncviszonyszám* bázisidőszaka állandóan változik, hiszen a tárgyidőszakot megelőző időszak értéke.

A láncviszonyszám képlete:

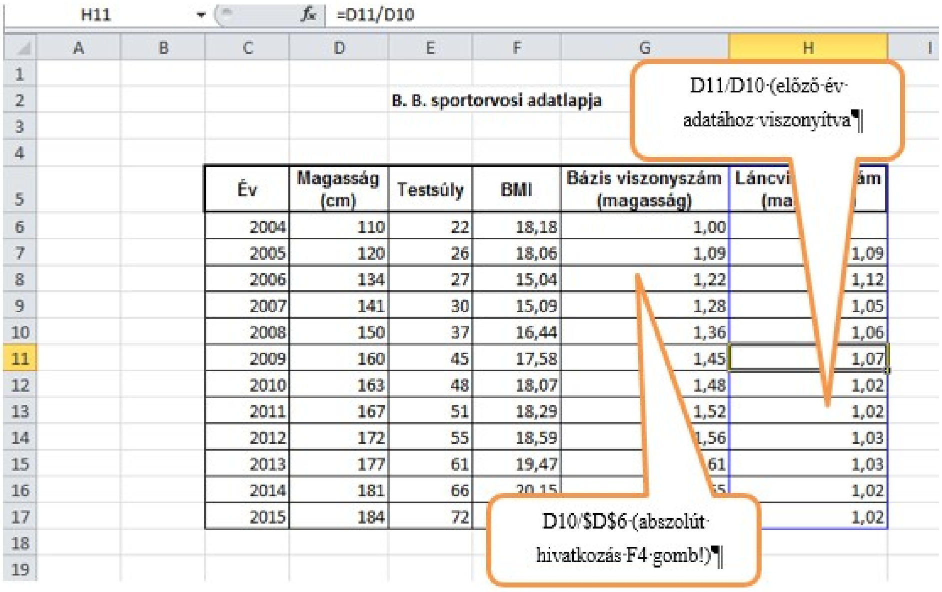
$$\mathrm{l_i} = \frac{\mathrm{y_i}}{\mathrm{y_{i-1}}}$$

Az alábbi képletek segítségével a bázisviszonyszámokból osztással láncviszonyszámot számíthatunk, míg a láncidexek bármely évvel bezárólag számított szorzata, a záróév bázisviszonyszámát adja.

$$\frac{\mathrm{b_m}}{\mathrm{b_{m-1}}} = \mathrm{l_m}$$

$$\mathrm{l_1 \times l_2 \times l_3 \times \ldots \times l_m = b_m}$$

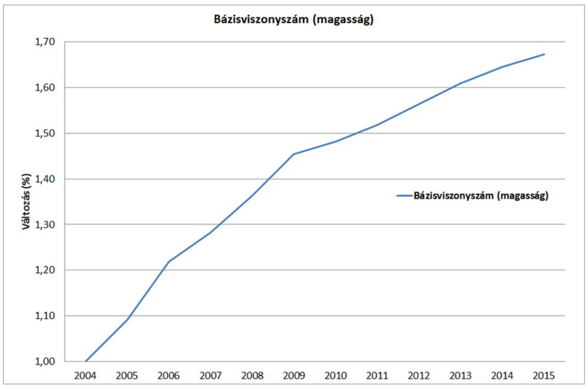
Az összefüggéseket felhasználva határozzuk meg a bázis és láncviszonyszámokat. A következőkben az excel program segítségével bemutatjuk a nevezetes (bázis- és láncviszonyszámokat)



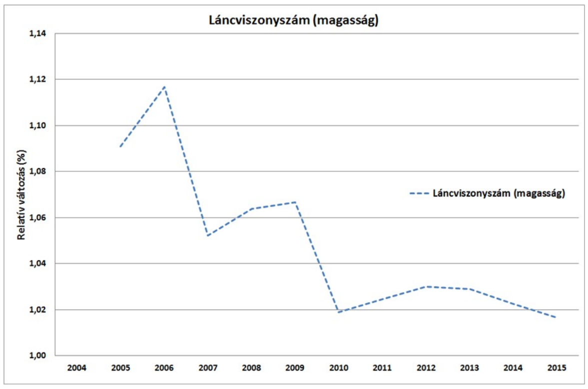
Év	Magasság (cm)	Testsúly	BMI	Bázis viszonyszám (magasság)	Láncvi[illegible]ám (ma[illegible])
2004	110	22	18,18	1,00	
2005	120	26	18,06	1,09	1,09
2006	134	27	15,04	1,22	1,12
2007	141	30	15,09	1,28	1,05
2008	150	37	16,44	1,36	1,06
2009	160	45	17,58	1,45	1,07
2010	163	48	18,07	1,48	1,02
2011	167	51	18,29	1,52	1,02
2012	172	55	18,59	1,56	1,03
2013	177	61	19,47	[illegible]	1,03
2014	181	66	[illegible]	[illegible]	1,02
2015	184	72	[illegible]	[illegible]	1,02

2/3. képernyőnézet: bázis és láncviszonyszámok

A bázisviszonyszámok (2/8 ábra) a fejlődés relatív mérését mutatják, míg a láncviszonyszámok (2/9. ábra) a változás ütemét illusztrálják (Forrás: viszonyszámok.xls).



2/8. ábra: bázisviszonyszám



2/9. ábra: láncviszonyszám

2.7.2.2. Információsűrítés középértékekkel (számtani átlag, módusz, medián)

A nagy mennyiségben keletkező kvantitatív adatok áttekintése, értékelése gyakran nehézkes. Kezdetleges elemzés lehet az adatok leszámlálása, rangsorba rendezése, összegzése, de felmerül az igény, hogy a tömeges előforduló jelenséget, kevés számú adattal jelenítsük meg. Szükséges egy olyan számadat, amely az egyedek közös jellemzőjeként elfogadható. Ennek az információsűrítésnek a legfontosabb eszköze a középérték számítás. Több olyan érték is van - nem mindig esnek egybe -, melyek törekszenek az egyedek közös lényegre mutató információi kifejezésére. A statisztika három követelményt támaszt az e fajta mutatószámokkal szemben: legyen *közepes*, *robusztus*, és *tipikus*.

A középértékek lehetnek számított (átlag), vagy helyzeti középértékek. *A számított középértékek (számtani-, mértani-, harmonikus-, négyzetes- átlag)* közül a számtani átlaggal foglalkozunk, a többi számított középértékről részletesebben DR JÁNOSA A. (2005) könyvében olvashat az érdeklődő. A *helyzeti középértékek közül a móduszt és a mediánt* mutatjuk be. Először egyszerű gyakorisági sorok alkalmazásával illusztráljuk ezeket a középértékeket, majd osztályközös gyakorisági sorokon is.

A statisztikai módszertanok bemutatását az általunk felmért primer adatbázisunk (forrás: fittségi 57 tisztított adat.xlsx) segítségével tesszük meg. A kutatásunk során egyetemi hallgatóságon (n=57 fő) mértük fel a Magyar Diáksport Szövetség által megalkotott NETFIT tesztrendszert (Kaj és társai: Kézikönyv a Nemzeti Egységes Tanulói Fittségi Teszt (NETFIT) alkalmazásához. 2014).

Az adatbázisban az alábbi felmérések adatait láthatjuk:

1. **Testösszetétel és tápláltsági állapot**

A testösszetétel jellemzése a testtömeg-index (Body Mass Index - BMI), és egy bioimpedancia-analizátor (OMRON BF 511) segítségével a testzsír-százalék, nyugalmi anyagcsere, viszcerális zsírszint mérését végeztük el.

2. **Ütemezett hasizom teszt**

Cél: A hasizmok erő-állóképességének mérése.

Eszköz: a teszt hanganyaga és a hanganyagot lejátszó eszköz, szőnyeg, mérőcsík.

Kiinduló helyzet: A hallgató hanyattfekszik egy szőnyegen, térdeit kb. 140º-os szögben behajlítja, talpait a talajon tartja kissé nyitott helyzetben, karjai a törzse mellett vannak, tenyere a talaj felé néz, az ujjai nyújtva, feje a talajon. A kiinduló helyzetet felvétele után a társa egy mérőcsíkot helyez el a tanuló behajlított lábai alatt úgy, hogy ujja éppen érintse a mérőcsík hozzá közelebbi szélét.

Feladat: A tanulónak a lehető legtöbb szabályos hasprést kell végeznie a hanganyag által diktált ütemben (1 db hasprés / 3 mp) hajlított térdekkel úgy, hogy lapockáit a talajról elemelve, ujjait a talajon előre csúsztatva érinti meg a mérőcsík távolabbi szélét, miközben sarkai folyamatosan érintik a talajt. A teszt addig tart, amíg a tanuló el nem éri a maximális ismétlésszámot, vagy már nem képes több hasprést teljesíteni a helyes technikával, illetve másodjára téveszti el az ütemet. A helyes végrehajtások számát a társa számolja, és figyelmezteti a hibákra.

Értékelés: A teszt eredményét a teljes ismétlések száma adja.

3. **Törzsemelés teszt**

Cél: A hátizmok erejének mérése.

Eszköz: centiméteres beosztású mérőrúd, szőnyeg, jelölőpont.

Kísérletek száma: kettő.

Kiinduló helyzet: A hallgató hason fekvésben helyezkedik el egy szőnyegen úgy, hogy a lábujjai a talajon vannak, kezeit pedig a combjai alá teszi.

Feladat: A tanuló lassú ütemben megemeli a törzsét, és miközben folyamatosan a szemével egy vonalban elhelyezett jelölőtárgyat nézi, tehát fejét egyenes vonalban tartja a törzs meghosszabbításaként. A törzs emelt helyzetét egészen addig megtartja, amíg társa le nem méri egy vonalzó vagy mérőrúd segítségével talaj és a tanuló álla közti távolságot. Ügyelni kell arra, hogy a hátrahajlítás ne legyen túlzott mértékű!

Értékelés: A tesztet kétszer kell végrehajtani és a jobbik eredményt kell feljegyezni centiméteres pontossággal. (A 30 cm-nél jobb eredményt is 30 cm-ként kell rögzíteni!)

4. **Ütemezett fekvőtámasz teszt**

Cél: A vállövi és a karizmok dinamikus erő-állóképességének mérése.

Eszköz: a teszt hanganyaga és a hanganyagot lejátszó eszköz.

Kiinduló helyzet: A hallgató vállszélességű fekvőtámaszban helyezkedik el, kezei előre néznek, ujjai nyújtottak, lábai egyenesek és kissé nyitottak, lábujjain támaszkodik. (A teszt hanganyagának indításáig térdelőtámaszba ereszkedhet!)

Feladat: A hallgató maximális számú karhajlítás-nyújtást igyekszik végrehajtani a hanganyag által diktált ütemben (1 db karhajlítás-nyújtás / 3 mp), törekedve arra, hogy teste folyamatosan egyenes vonalban maradjon, és a karhajlítás mélysége minden ismétlésnél kb. 90°-os legyen. A teszt addig tart, amíg a hallgató el nem éri a maximális ismétlésszámot (86 db), vagy már nem képes több karhajlítás-nyújtást helyes technikával végrehajtani, magáll, illetve másodjára téveszti el az ütemet. A helyes végrehajtások számát a társa számolja, és figyelmezteti a hibákra.

Értékelés: A teszt eredményét a teljes ismétlések száma adja, az eredmények értékelését egy összesítő táblázat segíti.

5. **Kézi szorítóerő mérése**

Cél: Az alkar izmainak maximális erejének mérése.

Eszköz: állítható markolatú kézi dinamométer.

Kísérletek száma: kettő + kettő.

Kiinduló helyzet: A hallgató az ügyesebbik kezébe veszi a kézi dinamométert, karját mélytartásba engedi úgy, hogy a kézfej és az alkar egyenes vonalban legyen.
Feladat: A hallgató az ügyesebbik kezével megpróbálja összeszorítani a kézi dinamóméter markolatát maximális erőkifejtéssel és megtartani két másodpercig.
A tesztet egyenes csuklóval és egyenletes, határozott mozdulattal kell végrehajtani, gyors rángató mozdulatok nélkül, illetve nem szabad a teszt végrehajtása közben a kart felemelni a és/vagy a mérőeszközt a testéhez szorítani. A tesztet kétszer kell végrehajtani az ügyesebbik kézzel, a két kísérlet között egy rövid szünettel, majd ugyanilyen módon az ügyetlenebbik kézzel is.
Értékelés: A két kísérlet közül a jobbik eredményt kell feljegyezni 1 kg-os pontossággal.

6. **Helyből távol ugrás teszt**

Cél: A láb dinamikus erejének mérése.
Eszköz: mérőszalag.
Kísérletek száma: kettő.
Kiinduló helyzet: A hallgató a kijelölt vonal mögött áll, térdei hajlítottak, a karjai a test előtt, párhuzamosan a talajjal, a lábujjak éppen érintik az elugró vonalat.
Feladat: A hallgató a kar hátra majd előrelendítésével vegyen lendületet és rugaszkodjon el arra törekedve, hogy a lehető legmesszebb érjen talajt.
Értékelés: A hallgató két kísérletet tehet és a jobbik eredményt kell rögzíteni az elért távolságot centiméterben megadva. Az ugrás távolságát az elugró helyhez közelebb eső sarok talajra érkezési pontjáig mért legrövidebb távolság adja.

7. **Hajlékonysági teszt**

Cél: Az ízületi mozgásterjedelem és a térdhajlító izmok nyújthatóságának vizsgálata.
Eszköz: mérőskálával ellátott stabil mérőeszköz.
Kísérletek száma: kettő + kettő.
Kiinduló helyzet: A hallgató nyújtott ülésben helyezkedik el a mérőskálával ellátott stabil mérőeszközzel szemben úgy, hogy egyik térdét behajlítja és a talpát a talajon tartja, másik lábának talpát pedig a mérőeszköz oldalához illeszti.
Feladat: A hallgató három előrehajlítást követően, kezét a mérőeszköz tetején lévő mérőskálán előrecsúsztatva maximális mértékű előrenyúlást végez az előírt testhelyzet megtartásával. A tesztet lábtartáscserével az ellenkező oldalra is meg kell ismételni.
Értékelés: Mindkét oldali végrehajtás eredményét 0,5 cm pontossággal kell rögzíteni.
A teszt abban különbözik a hagyományos „sit and reach" teszttől, hogy a hallgatót egy időben csak az egyik oldalt méri, így könnyebben kiszűrhető a két oldal közötti eltérés.

8. **Állóképességi ingafutás teszt (20méter)**

Cél: Az aerob kapacitás mérése.

Eszköz: a teszt hanganyaga és a hanganyagot lejátszó eszköz, jelzőbóják.

Kiinduló helyzet: a hallgatók felsorakoznak a számukra kijelölt pályák rajthelyeinél, párjaik pedig mögöttük helyezkednek el törökülésben, és figyelik társaik teljesítményét.

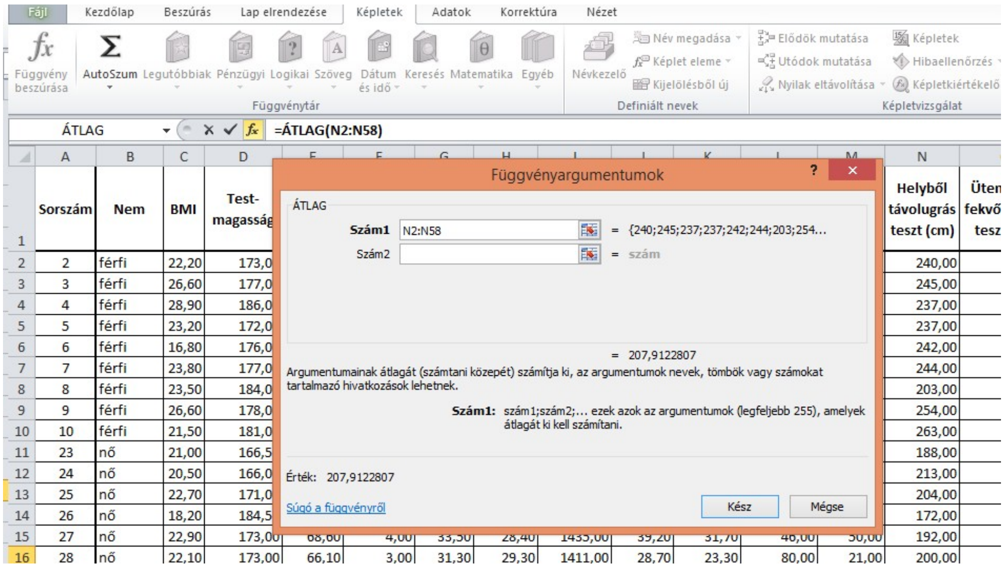
Feladat: A hallgatóknak maximális számú 20 méteres szakasz megtételére kell törekedniük a futás sebességét a hanganyag által diktált iramhoz igazítva. A tesztet a progresszív intenzitás jellemzi, azaz a teszt eleje könnyű és fokozatosan nehezedik. A teszthez kapcsolódó hanganyagban percenként emelkedően 21 szint különül el: az első szinten 9 másodperc áll rendelkezésre a 20 méteres táv teljesítésére, ami szintenként 1,5 másodperccel csökken. Az egyes szintek közötti váltásra háromszori hangjelzés figyelmezteti a hallgatókat, jelezve hogy gyorsabb tempóra kell váltaniuk.

Az adott szakasz akkor tekinthető teljesítettnek, ha a hallgató legkésőbb a hangjelzéssel egy időben legalább egy lábbal érint a 20 méteres szakaszt jelző vonalat vagy áthalad rajta. Ha a hallgató a hangjelzés előtt éri el a vonalat, meg kell várnia a jelzést, és csak a jelzés elhangzása után indulhat el visszafelé. A teszt akkor ér véget, ha a hallgató a második hibáját véti, azaz nem éri el a vonalat a hangjelzésre vagy nem tudja folytatni a futást.

Értékelés: A tesztet a teljesített szakaszok alapján értékeljük, egy táblázat segítségével megállapítva, hogy a hallgató hányadik szintet érte el, és méterben is megadhatjuk a hallgató által megtett távolságot.

A leggyakrabban alkalmazott középérték a **számtani átlag (mean)**, mely egy számított középérték. Gyakorlatilag az a szám, amelyet az átlagolandó értékek helyébe írva, azok összege változatlan marad. Általánosságban úgy számíthatjuk, hogy az adatok összegét osztjuk az adatok számával (forrás: fittségi 57 tisztított adat.xlsx).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{i=1}^{n} x_i}{n}$$



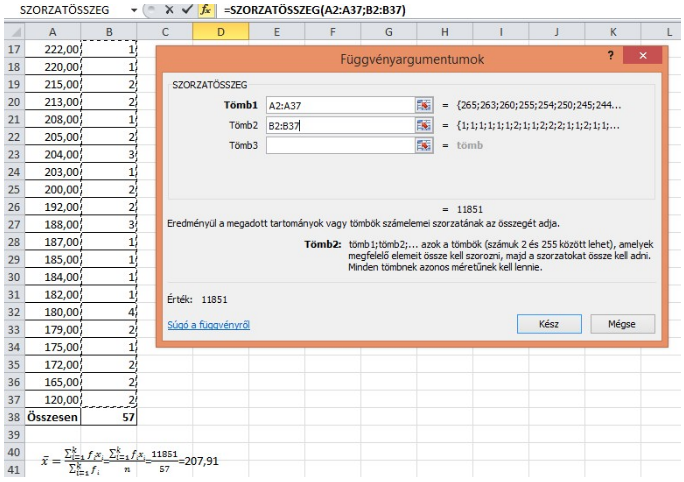
2/4. képernyőnézet: A számtani átlag számításának menete

Az Excel program számos lehetőségét kínál a számtani átlag meghatározására. Az egyik legkedveltebb módszer, hiszen más függvényeknél is gyorsan használható: a függvényvarázsló. A függvényvarázslót az Excel programban a képletek menü függvény beszúrása moduljából érhetjük el.

Számtalanszor találkozhatunk azzal, hogy gyakorisági sorból számtani átlagot számítani. Ilyenkor előfordulhat, hogy bizonyos ismérvértékek többször fordulnak elő, amit a gyakoriságok jeleznek számunkra. A számtani átlag számításához itt fel kell használni ezeket a gyakoriságokat, amit a **súlyozott számtani átlag** formulájával érünk el.

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$

A gyakorlatban először az értékösszegeket (ismérvérték és a gyakoriságok szorzatának összege) kapjuk meg, mely önálló jelentéssel is bír (összesen ugrott távolság cm-ben). Miután ezt megkaptuk, osztjuk az elemszámmal. Ha az adatsorunk csoportosítja az egyedeket és az ismérvérték gyakorisága van megadva, akkor érdemes így számolni. A fenti számítást két lépésben végezzük. Először kiszámoltuk az értékösszeget, (szorzatösszeg 11851 cm), majd osztjuk az elemszámmal (57).



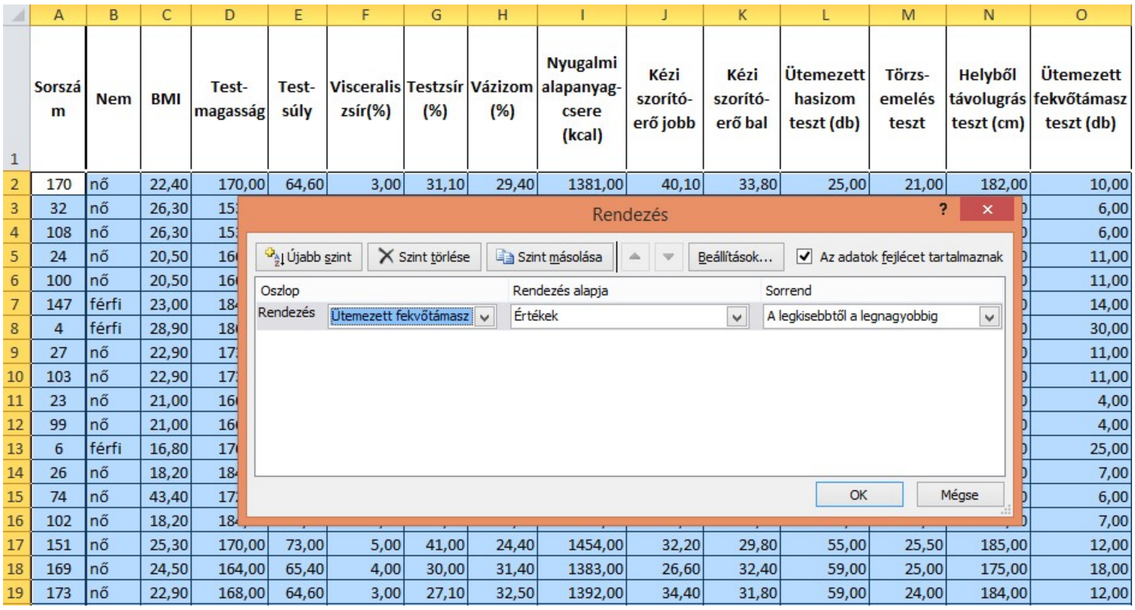
2/5. képernyőnézet: A szorzatösszeg számításának menete

Látható, hogy mindkét átlagszámítással egyforma eredményt kaptunk. A vizsgált 57 hallgató átlagosan 207,91 cm-re tudott elugrani helyből. Az átlagszámításhoz választott módszert a kiinduló adatok határozzák meg. Tudva lévő, hogy az eredmény értelmezése gyakran nehézkes (pl. 24, 11 db ütemezett fekvőtámasz), és sajnos a számtani átlagformula érzékeny a kiugró értékekre, pl. az „elgépelt” hibás adatokra, ezért valamely más középérték alkalmasságát is vizsgáljuk meg.

A **medián** a mennyiségi ismérvnek azon értéke, amelynél 50% kisebb, 50% nagyobb érték fordul elő (ugyanannyi kisebb, mint amennyi nagyobb). Meghatározása egyszerű gyakorisági sor esetén könnyű, hiszen a rangsorba rendezett középső tag értéke. A medián meghatározásának első lépéseként a számértékeket rangsorba rendezzük és

- ha n páratlan, akkor az (n+1)/2. sorszámú egyed ismérvváltozatának értéke lesz a medián ($Me = x_{\left(\frac{n+1}{2}\right)}$);
- ha n páros, akkor az n/2. és (n/2)+1. egyed ismérv-változatainak egyszerű számtani átlaga lesz a medián: $Me = \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$.

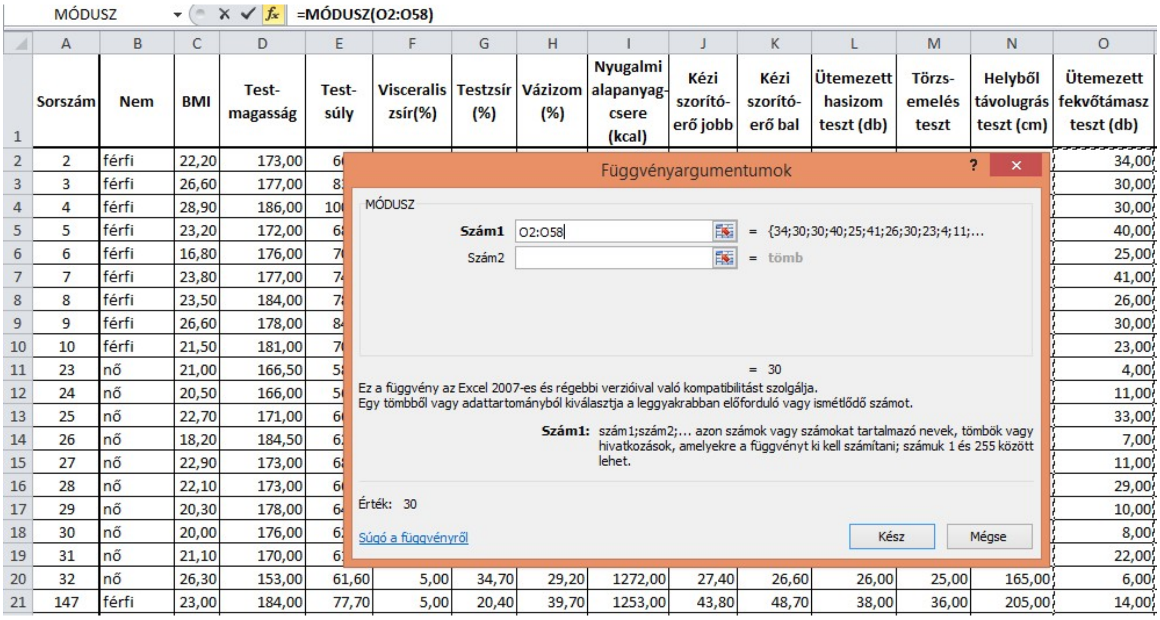
Az Excel programban természetesen megtalálható a medián függvény is, melyet előhívhatunk a függvényvarázslóból vagy egyszerűen szövegesen is megadhatunk (= medián). A két lépés szemléltetése végett vegyük a „bonyolultabb” esetet. Az ütemezett fekvőtámasz teszt mediánjának meghatározásához először rendezzük sorba az adatainkat, melyet kiválaszthatunk az adatok menüből a sorba rendezés parancs kiadásával. A megjelenő menü egyszerűsége végett (kezdőlap/rendezés és szűrés/egyéni sorrend) eltekintünk a részletes bemutatástól, csak a végeredményt közöljük. Javaslatként azonban el tudjuk mondani, hogy a sorba rendezések alkalmával az egyéni sorrend az egyszerű beállítási tényezői végett a legnépszerűbb modul.



2/6. képernyőnézet: A (sorba) rendezés menete

A keletkező rangsorból a fent ismertetett képlet segítségével történik a medián meghatározása. Tudjuk, hogy 57 megfigyelésünk van, tehát a rangsor (57+1)/2 elemének számtani átlaga lesz a medián, tehát a Me= 26. Ez azt jelenti, hogy a hallgatóknak a fele kevesebb, fele több volt, mint 26 db fekvőtámaszt tudott elvégezni. Itt most viszonylag könnyen tudtuk az eredményünket értelmezni, de gyakran jutunk nem egész számhoz, mely értelmezési problémát jelent.

A **módusz** az ismérvértékek tipikus, leginkább jellemző értékét jelöli, legtipikusabb érték. Diszkrét értékekkel rendelkező mennyiségi ismérv módusza a sokaságban leggyakrabban előforduló ismérvérték. Folytonos mennyiségi ismérv módusza ott található, ahol az előforduló értékek legjobban sűrűsödnek, ahol a gyakorisági görbe (lásd később) maximuma van.



2/7. képernyőnézet: A módusz számításának menete

A függvényvarázsló segítségével meghatározható a módusz is, mely jelen esetben 30. Tehát a felmért hallgatók között a leggyakoribb a 30 db ütemezett fekvőtámasz volt. Vigyázni kell, hiszen a módusz nem minden esetben lesz tipikusan középső érték. Látszik, hogy nem mindegyik középérték felel meg minden kritériumnak, tehát a követelményeknek maradéktalanul eleget tevő középérték nincs. A kiválasztását mindig a kutatási cél határozza meg, és az értelmezhetősége.

A **középértékek meghatározása az osztályközös gyakorisági sorokból** már kissé összetettebb feladat. Az osztályközös gyakorisági sorokat, akkor érdemes képezni, amikor az ismérvváltozatok számának csökkentése szükséges, hiszen nagy mennyiségű adattal kívánunk dolgozni. Ilyenkor az változóértékekből intervallumokat, vagyis osztályokat hozunk létre. Természetesen a legpontosabb eredményeket az egyedi adataink felhasználásával nyerhetjük és a számítógépes programok viszonylag könnyen kezelik a nagyszámú egyedi adatokat is. Az osztályközös gyakorisági sorok leggyakrabban akkor kerülnek elő, amikor az alapadatainkat már eleve így közlik, vagy amikor a szekunder adatbázis eleve osztályközökben rendezett. Felhívjuk a figyelmet arra, ha rendelkezésre állnak az egyedi adataink, akkor nem szükséges osztályközös gyakoriságokat képezni, hiszen a kapott eredmény pontatlanabbá válhat. A következő példa éppen ezért csak szemléltető jellegű, hiszen meglévő egyedi adatokból képezzünk osztályközös gyakorisági sort, mely a pontosság rovására megy, viszont a megjelenített példa alkalmas arra, hogy a módszert bemutassa.

A következőkben rendezzük az 57 egyetemista testmagasság (cm) adatát osztályközös gyakorisági sorba. Az alapadatokat a következő táblázat mutatja (Forrás: osztályközös gyakoriság. xlsx).

Első lépésben az osztályokat kell kialakítani. Számos lehetőségét ismer a statisztika módszertana, ezek közül most egy viszonylag gyorsan és egyszerűen alkalmazhatót kívánunk szemléltetni.

Az osztályok száma priori információ vagy képlet segítségével is meghatározható:

$$r = \sqrt{\sum f_j} + 1 \text{. vagy } r = 1 + (3{,}3 \times \lg n)$$

Az osztályközök hossza: $h = \frac{x_{max} - x_{min}}{r}$

Ezek szerint, ha a hallgatók átlagos testmagasságát kívánjuk meghatározni, először az osztályközös gyakorisági sorba rendezzük az adatainkat.

$$r = x_{max} - x_{min} = 188{,}5 - 153 = 35{,}5$$

Itt a használhatóak a min. és a max. függvények, melyek a gyakorisági sor legkisebb és legnagyobb értékét adják meg. A terjedelem (*range*) a maximális és minimális érték különbsége, mely egy szóródási mérőszám.

$r = \sqrt{57} + 1 = 8{,}55$, ilyenkor érdemesebb felfelé kerekíteni, tehát 9 osztályba soroljuk be az értékeket. Az intervallumok hossza: $h = \frac{188{,}5 - 153}{9} = 3{,}94$; kerekítetve 4

Ezután elkészíthető az osztályközös gyakorisági sor:

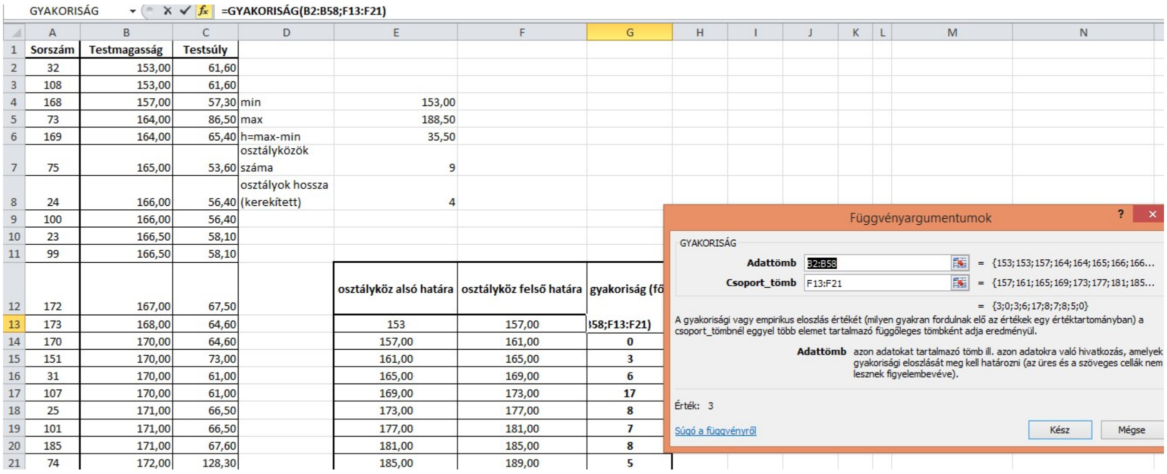
2/2. táblázat: A 9 „osztályba” rendezett osztályközös gyakorisági sor

osztályköz alsó határa	**osztályköz felső határa**	**gyakoriság (fő)**
153	157,00	3
157,00	161,00	0
161,00	165,00	3
165,00	169,00	6
169,00	173,00	17
173,00	177,00	8
177,00	181,00	7
181,00	185,00	8
185,00	189,00	5

Forrás: saját számítás

Megvizsgálva az elkészített osztályközös gyakorisági sort azt tapasztaljuk, hogy van négy osztályköz, ahol nincsen, vagy alig van érték, így azt feltételezzük, hogy jobb lenne a besorolást 5 osztályba rendezni. Az eredeti adataink is egy tizedes pontossággal vannak megadva, ezért az intervallum hosszát is egy tizedesig kell kerekíteni (függvényvarázsló: kerek fel függvény)

Az intervallumok hossza: $h = \dfrac{188{,}5 - 153}{5} = 7{,}1$; erősen felfelé kerekítetve **8** (törekedni kell az „osztályok” minimalizálására)



GYAKORISÁG =GYAKORISÁG(B2:B58;F13:F21)

	A	B	C	D	E	F	G
1	Sorszám	Testmagasság	Testsúly				
2	32	153,00	61,60				
3	108	153,00	61,60				
4	168	157,00	57,30	min	153,00		
5	73	164,00	86,50	max	188,50		
6	169	164,00	65,40	h=max-min	35,50		
7	75	165,00	53,60	osztályközök száma	9		
8	24	166,00	56,40	osztályok hossza (kerekített)	4		
9	100	166,00	56,40				
10	23	166,50	58,10				
11	99	166,50	58,10				
12	172	167,00	67,50		osztályköz alsó határa	osztályköz felső határa	gyakoriság (fő
13	173	168,00	64,60		153	157,00	;58;F13:F21)
14	170	170,00	64,60		157,00	161,00	0
15	151	170,00	73,00		161,00	165,00	3
16	31	170,00	61,00		165,00	169,00	6
17	107	170,00	61,00		169,00	173,00	17
18	25	171,00	66,50		173,00	177,00	8
19	101	171,00	66,50		177,00	181,00	7
20	185	171,00	67,60		181,00	185,00	8
21	74	172,00	128,30		185,00	189,00	5

2/8. képernyőnézet: A gyakoriság meghatározása

Az adattömb értékben az alapadatok, míg a csoporttömb értékben az osztályközök felső határai szerepelnek. Ezután kijelöljük (kiterjesztjük a többi osztályra is), hogy a többi intervallumra is számolja ki az eredményt, majd a megnyomjuk az F2 billentyűt, amit a shift-ctrl-enter gomb megnyomása követ.

Megvizsgálva az elkészített osztályközös gyakorisági sort azt tapasztaljuk, hogy van négy osztályköz, ahol nincsen, vagy alig van érték, így azt feltételezzük, hogy jobb lenne a besorolást 5 osztályba rendezni. Az eredeti adataink is egy tizedes pontossággal vannak megadva, ezért az intervallum hosszát is egy tizedesig kell kerekíteni (függvényvarázsló: kerek fel függvény)

Az intervallumok hossza: $h = \dfrac{188{,}5 - 153}{5} = 7{,}1$; erősen felfelé kerekítetve **8** (törekedni kell az „osztályok” minimalizálására)

Ennek segítségével megkapjuk az új osztályközös gyakorisági sorunkat, melynek helyes megítélésében segít a relatív és a kumulált relatív gyakorisági sor. Az osztályközös gyakorisági sorok készítésénél fokozott figyelmet igényel az osztályközhatárok megállapítása is. Általános alapelvként fogalmazhatjuk meg, hogy a határok mindenkor tegyék lehetővé az egyértelmű besorolást. Kifejezésre kell juttatni, hogy egy adott határérték mely osztályközbe tartozik[14]. Ez különösen a folytonos ismérvértékek használata esetén okozhat gondot, itt fokozott óvatossággal kell eljárni. Sokat segíthet a folytonos ismérvértékek kerekítése is, amely végül egyértelművé teheti a besorolást.

2/3. táblázat: Az 5 osztályba rendezett osztályközös gyakorisági sor munkatáblája

osztályköz alsó határa	osztályköz felső határa	gyakoriság (fő)	Relatív gyakoriság (gi)	Kumulált relatív gyakoriság (gi')
–	161	3	5,26%	5,26%
161	169	9	15,79%	21,05%
169	177	25	43,86%	64,91%
177	185	15	26,32%	91,23%
185	–	5	8,77%	100,00%
Összesen		**57**	**100,00%**	

Forrás: saját számítás

Az ismételten elkészített táblázatunkból a gyakoriságok mellett látható a *relatív gyakorisági sor* is, mely gyakorlatilag megoszlási viszonyszámként képezhető és értelmezhető. Azt jelenti, hogy a felmérésben résztvevők 5,26 %-a alacsonyabb 161 cm-nél. *A kumulált (felfelé) gyakorisági* sor azt mutatja meg, hogy a 64,91 %-a a hallgatóknak alacsonyabb, mint 177 cm.. Amennyiben a relatív gyakorisági sorban a két kezdő osztályon kívül nincsenek kiugróan különböző értékek, úgy az osztályok számát elfogadhatónak ítélhetjük meg.

A fenti táblában leírt gyakorisági sorban az alsó és a felső intervallum ún. **nyitott intervallum**. Ezt a megoldást általában akkor szokták használni, ha az adatok között extrém, kiugró szélsőértékek szerepelnek. Példánkban ez a tény nem áll fenn, azonban egy megismételt vizsgálat esetén –ha a többi ezer adatot is vizsgáljuk, nem kizárt szélsőségesebb értékek keletkezése – a fenti csoportosítás jó lehetőséget ad az összehasonlításra. A további számítások, elemzések során ezeket a nyitott intervallumokat úgy kezeljük, mintha zártak

[14] Ilyenkor a megfigyeléseknél is pontosabb osztályközhatárokat adunk meg, melyeket éppen ennek megfelelően „technikai számokként" kezelünk.

lennének, az első intervallumot ugyanolyan hosszúságúnak tételezzük fel, mint az azt követőt; az utolsót pedig olyan hosszúnak, mint az azt megelőzőt.

A következő lépésben az osztályközös gyakorisági sorral „megbecsüljük" a felmért gyerekek átlagos testmagasságát, amihez az osztályközepeket kell első lépésben meghatározni. Az **osztályközepek (x_i)**, - amelyek fontos szerepet töltenek be a további számításokban is - az intervallumok alsó és felső határainak átlagolásával képezhetők (kiszámításuk során már nem vesszük figyelembe a csak az egyértelmű besorolás érdekében megkülönböztetett, technikai alsó- és felső osztályköz-határokat. Az osztályközepek és a gyakorisági sorok szorzataként keletkező adatsort **értékösszeg-sornak** (s_i) nevezzük. Ezt követően az értékösszeg-sor összegét osztjuk az elemszámmal.

2/4. táblázat: a számtani átlag meghatározásának munkatáblája

osztályköz alsó határa	osztályköz felső határa	gyakoriság (fő)	osztályközép (x_i)	értékösszeg ($s_i=f_i*x_i$)
–	161	3	157	471
161	169	9	165	1485
169	177	25	173	4325
177	185	15	181	2715
185	–	5	189	945
Összesen		57		9941

Forrás: saját számítás

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{n} = \frac{9941}{57} = 174{,}4$$

Az eredmény azt jelenti, hogy a vizsgált 57 hallgató átlagosan 174,4 cm magas. Miután itt osztályközepekkel számoltunk így egy becsült adatot kaptunk, a tényleges adatokból végzett számítás szerint is az átlag 174, 3 cm, tehát a két számítás eredménye között nem volt jelentős az eltérés.

A **medián meghatározása osztályközös gyakorisági sorból szintén** közelítő eljárással történik - kumulált gyakorisági sor felhasználásával-, a következő képlet alkalmazásának segítségével:

$$Me = x_{me,a} + \frac{s - f'_{me-1}}{f_{me}} \times h$$

ahol: $x_{me,a}$ - mediánt magába foglaló osztályköz alsó (nem technikai) határa

s - n/2 - a medián sorszáma

f'_{me-1} - a mediánt megelőző osztályköz kumulált gyakorisága

f_{me} - a mediánt tartalmazó osztályköz gyakorisága

h - mediánt tartalmazó osztályköz hossza.

2/5. táblázat: a medián meghatározásának munkatáblája

osztályköz alsó határa	**osztályköz felső határa**	**gyakoriság (fő)**	**Kumulált gyakoriság (f_i')**
–	161	3	3
161	169	9	12
169	177	25	37
177	185	15	52
185	–	5	57
Összesen		**57**	

Forrás: saját számítás

$$Me = 169 + \frac{28{,}5 - 12}{25} \times 8 = 174{,}28\ cm.$$

Ez azt jelenti, hogy a hallgatók fele nagyobb, fele kisebb, mint 174,28 cm. Amennyiben nem osztályközös gyakorisági sorból számoljuk a mediánt, akkor az érték 173 cm lesz. A medián képlete alapján könnyen meghatározhatók a nevezetes kvantilisek is, hiszen egy ilyen nevezetes kvantilis maga a medián is. Ha a rangsorba rendezett sokaságot 2, 3, 4, ...,k **egyenlő részre osztjuk**, az osztópontoknak megfelelő ismérvértékeket **kvantiliseknek** hívjuk. Másképpen a kvantilis értékek azok az ismérvértékek, amelyeknél az összes előforduló érték

$$\frac{1}{k}, \frac{2}{k}, ..., \frac{k-1}{k},\ \text{röviden}\ \frac{j}{k}\ (j = 1,2,..., k-1)\text{-ad}$$

része kisebb, illetve 1-(j/k)-ad része nagyobb. Néhány fontosabb kvantilis értéknek sajátos elnevezése is van:

2/6. táblázat: Néhány nevezetes kvantilis

K	Elnevezés	Jele
2	Medián	M_e
3	Tercilis	T_j
4	Kvartilis	Q_j
5	Kvintilis	K_j
10	Decilis	D_j
100	Percentilis	P_j

A kvantilis értékek közül gyakran találkozunk a kvartilis értékekkel (Q_j). A kvartilisek használata során általában a ***felső kvartilis**re* (Q_3), illetve *alsó kvartilisre* (Q_1) gondolunk, mivel Q_2 = medián. Ezek szerint a kvartiliseket a mediánnál kisebb, illetve nagyobb értékek mediánjaiként is felfoghatjuk. Az alsó kvartilisnél az előforduló ismérvértékek egynegyede kisebb, háromnegyede nagyobb értéket; a felső kvartilisnél az előforduló ismérvértékek háromnegyede kisebb, egynegyede nagyobb értéket vesz fel. Tehát előző példánál maradva az n=57 tagszám megfelelő j/k- ad hányadait kell először kiszámítani. A tagszám fele 28,5, egynegyede 14,25 és háromnegyede 42,75. A kumulált gyakorisági sort nézve láthatjuk, hogy a 7,5 a 13,1 osztályközhatárt lépi át, tehát, - mint a medián- az alsó kvartilis 13,1-14,6 osztályközben helyezkedik el.

$$Q_1 = 169 + \frac{14{,}25 - 12}{25} \times 8 = 169{,}72.$$

$$Q_3 = 177 + \frac{42{,}5 - 37}{15} \times 8 = 179{,}93\ .$$

Tehát a hallgatók egynegyede alacsonyabb, mint 169,72 cm, viszont egynegyede magasabb, mint 179,93 cm, vagy úgy is fogalmazhatnánk, hogy háromnegyedük alacsonyabb, mint 179,93 cm.

Módusz meghatározás egyszerű gyakorisági sorból nem okoz különösebb nehézséget, hiszen a legnagyobb gyakorisággal rendelkező ismérvértéket kell választani. Osztályközös gyakorisági sor esetében a módusz közelítő meghatározása az ún. **modális osztályköz** meghatározásával kezdődik. A modális osztályköz, –amely a móduszt magában foglalja – a legnagyobb gyakorisággal rendelkező osztályköz. Természetesen csak egyenlő hosszúságú osztályközök esetén tudjuk a modális osztályközt rögtön meghatározni. Ha az osztályközök nem egyenlők, akkor korrekciót kell végezni, át kell számolni a gyakoriságokat azonos hosszúságú osztályközre.

A modális osztályköz kijelöléséhez, valamint a további számításokhoz ezeket a korrigált gyakoriságokat használjuk:

A módusz meghatározása az alábbi képlet segítségével végezhető el:

$$Mo = x_{mo,a} + \frac{k_1}{k_1 + k_2} \times h$$

Ahol: $x_{mo,a}$ – a modális osztályköz alsó határa,

k_1 – a modális osztályköz és a megelőző osztályköz gyakoriságának különbsége,

k_2 – a modális osztályköz és az azt követő osztályköz gyakoriságának különbsége,

h – a modális osztályköz hossza

A modális osztályköz meghatározásában segítenek a gyakoriságok, hiszen a modális osztályközben a legmagasabb a gyakoriságok száma. Előfordulhat, hogy két szomszédos osztályhoz, ugyanazon „legmagasabb" gyakoriság tartozik, ilyenkor a közös osztályhatárt kell választani. Ennek bizonyítását az érdeklődő megtalálja Pintér-Rappai (2007, 131. o.) könyvében.

$$Mo = 169 + \frac{25-9}{(25-9)+(25-15)} \times 8 = 181{,}8 cm.$$

Azt jelenti, hogy a leggyakrabban előforduló testmagasság a hallgatók körében 181,8 cm.

2.7.2.3. Szóródás és szimmetria

Láthatóvá vált, hogy a középértékek egyike sem felel meg tökéletesen minden elvárásnak ezért, kívánatos lehet a további, árnyaltabb statisztikai elemzés is. Előfordulhat, olyan eset is, hogy a középértékek megegyeznek, mégis az ismérvértékek (sokaság egyedei) jelentősen eltérnek. Szóródásnak nevezzük a sokaság egyedeinek különbözőségét, mely az alapsokaság heterogenitásából adódik. Belátható, hogy minél „egyneműbb" (homogén) egy sokaság annál kisebb az egyedek szóródása a vizsgált ismérv szempontjából.

A szóródás jelenségét számtalan mérőszám vizsgálja, ezek közül a legelterjedtebbek:

- szóródás terjedelme (R),
- interkvartilis terjedelem (TQ),
- szórás (σ), variancia (σ^2)
- relatív szórás (V)

A **szóródás terjedelme** az előforduló legnagyobb és legkisebb érték különbsége:

$$R = x_{\max} - x_{\min}$$

A sportolók felkészülése során nem elhanyagolható információ a legjobb egyéni eredménytől (világcsúcstól) való eltérés nagysága. Fontos mutatószám a csúcsforma meghatározásához, hiszen a mutatószám csökkenése jelzi a sportoló formájának javulását.

Az **interkvartilis terjedelem** azt az intervallumot jelöli, ahol az összes érték középső 50 %-a helyezkedik el. Az interkvartilis terjedelem képlete:

$$TQ = Q_3 - Q_1$$

Az előző példát folytatva TQ=179,93-169,72= 10,21 cm. azt jelzi, hogy hallgatók testmagasságának 50%-a 10,21 cm-en belül helyezkedik el.

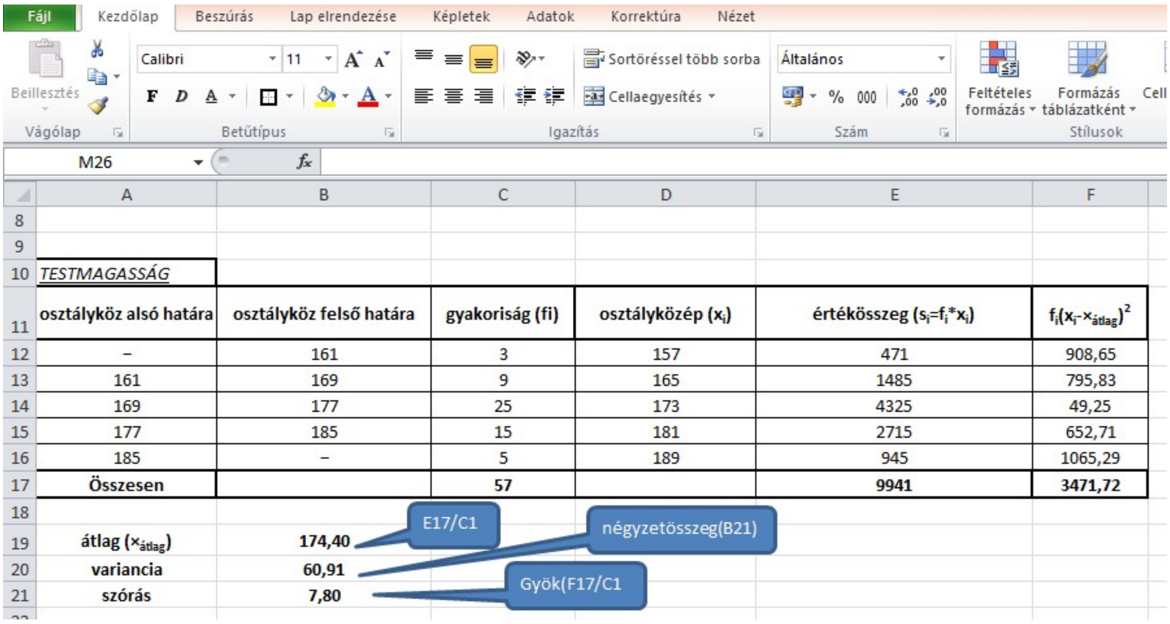
A leggyakrabban alkalmazott szóródási mérőszám a **szórás**, ami az egyes értékek számtani átlagtól való eltéréseinek négyzetes átlagát jelenti.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Ha gyakorisági sorból számítjuk a szórást, a mutatószám **súlyozott** formáját kell alkalmazni:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f}} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i\left(x_i - \bar{x}\right)^2}$$

A szórás négyzetét **varianciá**nak (σ^2) hívjuk. Önálló tartalommal nem bír, bizonyos statisztikai eljárásokban kulcsfontosságú mutatószám. Az előző példát folytatva szemléltetjük a szórás és variancia számítás menetét osztályközös gyakorisági sorból.



	A	B	C	D	E	F
10	TESTMAGASSÁG					
11	osztályköz alsó határa	osztályköz felső határa	gyakoriság (fi)	osztályközép (x_i)	értékösszeg (s_i=f_i*x_i)	$f_i(x_i-x_{átlag})^2$
12	–	161	3	157	471	908,65
13	161	169	9	165	1485	795,83
14	169	177	25	173	4325	49,25
15	177	185	15	181	2715	652,71
16	185	–	5	189	945	1065,29
17	Összesen		57		9941	3471,72
19	átlag ($x_{átlag}$)	174,40				
20	variancia	60,91				
21	szórás	7,80				

2/9. képernyőnézet: A szórás meghatározása

Az egyes testmagasságok átlaga átlagosan 7,8 cm-rel tér el az átlagtól (174,4 cm). A szóródás eddig megismert mérőszámai a mennyiségi ismérv mértékegységében fejezik ki a szóródás nagyságát. Sok esetben szükség lehet arra, hogy elvonatkoztassunk a mértékegységektől és ezáltal összehasonlíthatóvá tegyük a különböző jelenségek, különböző mértékegységben kifejezett szóródását, erre való a **relatív szórás**, ami kifejezi, hogy az egyes értékek átlagosan hány %-kal térnek el az átlagtól.

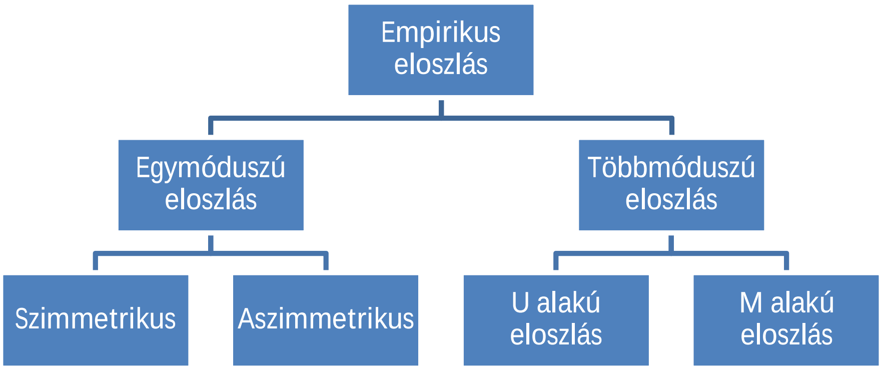
$$V = \frac{\sigma}{\bar{x}} = \frac{7,8}{174,4} = 4,47\%$$

Látható, hogy a relatív szórás viszonylag alacsony, hiszen az átlagos eredmények 4,47 %-a. (Az átlag és a szórás hányadosának értékét Müller a teljesítmény állandóság indexeként definiálta)[15].

A gyakorisági soraink vizsgálata során láthattuk, hogy a középérték magának az eloszlásnak a helyét, míg a szóródási mutatószámok a sokaság kiterjedését „sűrűségét" jellemzik. A sokaság látszólagos véletlenségéről, természetéről nyerhetünk többletinformációt, ha a gyakorisági eloszlásokat is vizsgáljuk. Ha azt tapasztaljuk, hogy az ismérvek a legnagyobb és legkisebb érték között nem egyformán „sűrűsödnek", akkor grafikus ábrákkal könnyen látható, hogy mely területen található viszonylag több érték. **Az eloszlás** arra keresi a választ,

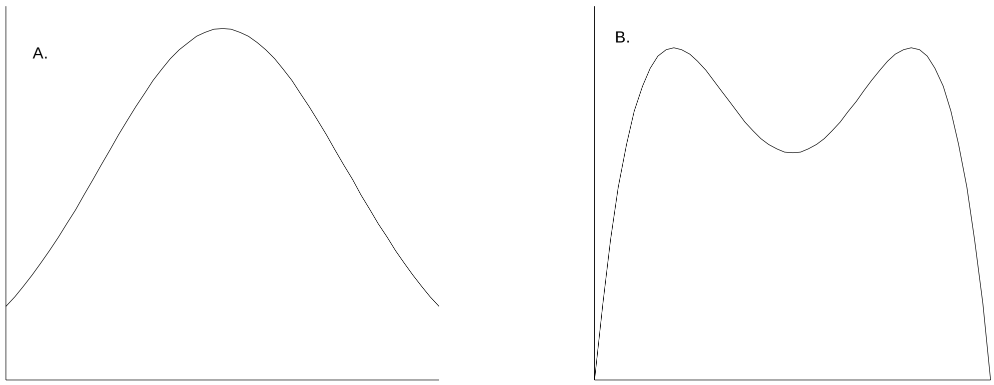
[15] Müller A.(2004), 103.o.

hogy a vizsgált ismérvértékek milyen gyakorisággal fordulnak elő a terjedelem adta nagyságrenden belül, azaz a kisebb vagy a nagyobb értéktartományba esik- e több érték. Ennek megfelelően az eloszlásokat osztályozhatjuk is (Pintér- Ács, 2006, 95.o).



2/10. ábra: Az empirikus eloszlások fajtái

Az egymóduszú-, illetve a többmóduszú eloszlások görbéjének egy-egy típusát a következő grafikus ábrán szemléltetjük.



2/11. ábra

A fenti ábra **A**. részében egy egymóduszú, **szimmetrikus** eloszlás görbéjét vázoltuk fel. A többmóduszú eloszlások (B rész) általában heterogén[16] sokaságot jellemeznek. A gyakorisági görbe helyi maximumai a homogénebb részsokaságok móduszainál jönnek létre. Jellemző példa lehet – az ábra B. részében szereplő – M alakú eloszlásra pl.: egy kézilabda csapat játékos-béreinek eloszlása, ami a képzettség, elismertség (élvonalbeli mérkőzések

[16]Heterogénnek tekintjük a sokaságot, ha valamely ismérv szerint homogénebb, egyneműbb részekre bontható.

száma), poszt stb. szerint különböző helyi (jelen esetben kettő) maximumokkal bírhat.[17] Viszonylag ritkán fordul elő a gyakorlatban egy olyan kétmóduszú, U alakú eloszlás, amelynek jellemzője, hogy a két módusz egyben a két szélsőérték. Gyakorló oktatóként gyakran hallható az egyetemeken található alsóbb éves évfolyamokon vizsgák után, hogy az oktató a szóbeli vizsgán csak elégtelent, vagy csak jelest adott és ez mennyire igazságtalan. Amennyiben a hallgatók észrevételei a valóságon alapulnának, akkor az érdemjegyek U alakú eloszlást vennének fel.

Az egymóduszú szimmetrikus eloszlásnál (A. ábra) a középértékek (számtani átlag, medián, módusz) ugyanazon helyre esnek (gyakorlatban csak közelítőleg): $Mo = Me = \bar{x}$. Ilyenkor a leggyakrabb előforduló ismérv a sokaság középső értéke.

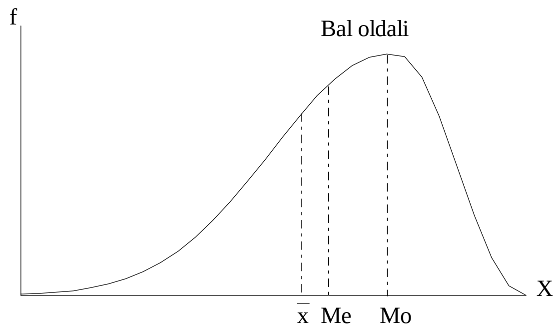
Az aszimmetrikus eloszlásoknál ez az egyezés nem áll fenn. Beszélünk jobb és baloldali aszimmetriáról, amit a statisztikai tankönyvek nem egységesen közölnek, jelen tankönyvben a „pécsi statisztikai alma mater”[18]-nek megfelelően tárgyaljuk.

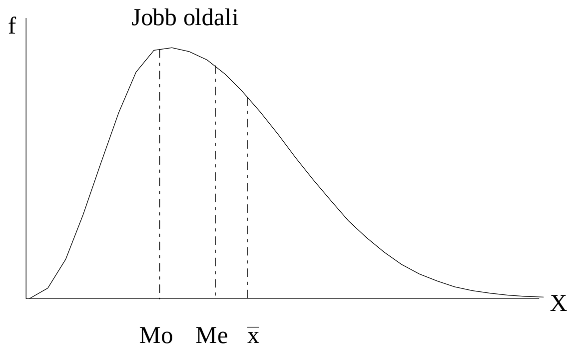
Baloldali aszimmetria esetén:

$$\mathrm{Mo} > \mathrm{Me} > \bar{\mathrm{x}}$$

Jobboldali aszimmetria esetén a nagyságrendi reláció:

$$\mathrm{Mo} < \mathrm{Me} < \bar{\mathrm{x}}$$





2/12. ábra

[17] M alakú eloszlásra leggyakrabban emlegetett példa egy évfolyamon (férfiak és nők) a testmagasság eloszlása.

[18] Pécsett az angolszász terminológiát veszik alapul, amelyet a statisztikai szoftverek többsége használ.

Aszimmetrikus gyakorisági görbék két típusa

A ferde, aszimmetrikus eloszlások közül a sportban is a jobboldali aszimetriát mutató eloszlás a gyakoribb, a sportolók teljesítményének eloszlása is jobboldali aszimmetriájú, ugyanis kevesen érnek el kimagasló eredményeket, a többségük „csak” szerényebb teljesítményre képes.

Baloldali aszimmetria esetén a viszonylag magasabb értékek nagyobb gyakorisággal fordulnak elő, pl.: a nem megfelelő edzésterheléseket - a kitűzött edzési szinteket-a sportolók többsége túlteljesíti. Ha egy gyakorlatot elemzünk, és az azt tapasztaljuk, hogy a grafikus ábrán a görbe bal irányba tolódik, akkor a edzésen végett feladat túl könnyű, tehát a terhelési összetevőkön (terjedelem, intenzitás) a fejlődés érdekében változtatni kell. Amennyiben helyesen határoztuk meg a terhelési összetevőket, ha a próbák elemzése során a normális eloszláshoz hasonló görbét kapjuk, hiszen mintából (véges számú esetből) az eloszlás csak közelíti a normálist[19] (Gauss-görbe).

Az aszimmetria számszerűsítése során gyakorlatban széles körűen alkalmazzák az ún.

Pearson-féle A mutatót, amelynek képlete: $A = \frac{\bar{x} - Mo}{\sigma}$

A szimmetrikus eloszlásnál az Pearson-féle A mutató értéke nulla, pozitív előjel esetén jobb oldali, míg negatív előjel esetén bal oldali aszimmetriát találunk. A mérőszámnak abszolút értékben nincs felső korlátja, azonban 1-nél nagyobb abszolút érték már erőteljes aszimmetriát jelez.

A példánkban *bal oldali* aszimmetriát találunk, hiszen: $A = \frac{174{,}4 - 181{,}8}{7{,}8} = -0{,}9$

A másik fontos mérőszám az F mutató: $F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$

Jobb oldali aszimmetriánál pozitív, míg bal oldali aszimmetriánál negatív lesz a mutató értéke, amely szimmetrikus eloszlású gyakorisági sor esetén zéró értékkel bír. A mutató határozott alsó és felső határokkal bír: $-1 \le \mathrm{F} \le 1$

$$F = \frac{(179{,}93 - 174{,}28) - (174{,}28 - 169{,}72)}{(179{,}93 - 174{,}28) + (174{,}28 - 169{,}72)} = 0{,}11$$

---

[19] Normáleloszlásnál az adatok úgy szóródnak, hogy az átlagtól egy szórásnyi távolságon belül lefelé és felfelé az adatok 68%-a ,helyezkedik el, egy hatoda a szóráson kívül alul és felül, és kb. az adatok 2,3 /-ának távolsága az átlagtól meghaladja két szórásnyi távolságot, és 0,1%-uk a három szórásnyi távolságot is meghaladja.

Az F mutatóval vizsgálva is *jobb oldali* aszimmetriát találunk. A különbség adódik a mutatószámok során használt különböző középérték fajtákból.

Az Excel program az egyedi adatokra a korrigált az S' (skewness) mutatót használja a „ferdeség" megállapítására, melynek értelmezése megegyezik az A mutatónál leírtakkal.

$$S' = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$

A szimmetria, ferdeség mellett másik használt jelenség a csúcsosság (kurtosis). A K-mutató értéke 3, normális eloszlás esetén, ha kisebb az érték az eloszlás lapult (egyenletes eloszlást közelíti), különben csúcsos (átlag körül tömörül).

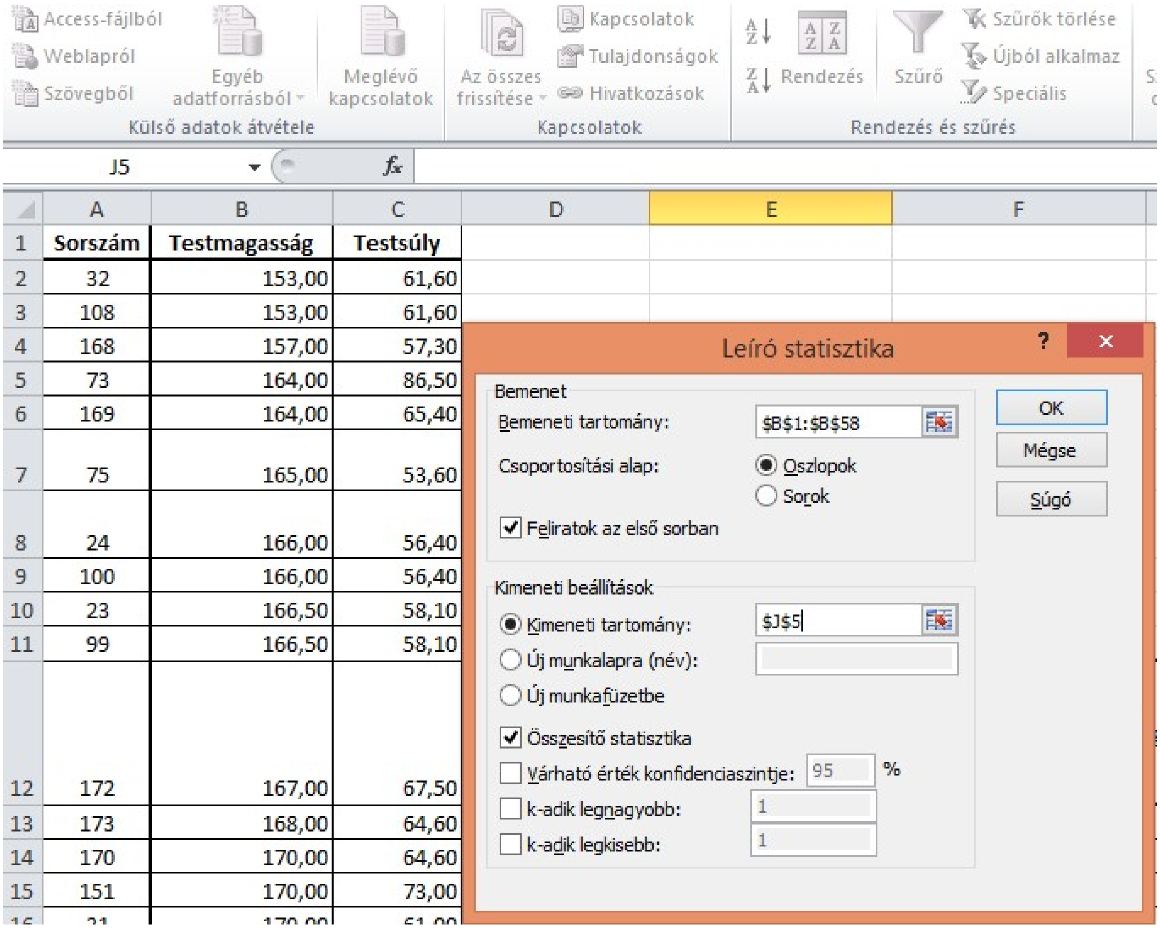
$$K = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$$

Az Excel programban beépített csúcsossági együttható K' számol a kismintás korrekcióval is, ezért értékelésénél ellentétesen a pozitív érték csúcsos, negatív érték lapult eloszlást jelent.

Láthattuk, hogy módunkban áll a számítógép segítségével lépésről-lépésre (pl.: függvényvarázsló) az egyes leíró statisztikai elemzéseket elvégezni, de ezt megtehetjük az adatok menüpontban található adatelemzés alpont, leíró statisztika módul alkalmazásával is.

Ez a módul alapesetben nem áll rendelkezésre, szükséges hozzá a bővítménykezelő (File menüpontban/ beállítások modul/bővítmények almodul) Analysis ToolPak moduljának bekapcsolása. (Forrás: osztályközös gyakoriság.xlsx)

Ezután kérhetjük az összesítő statisztikát!



2/10. képernyőnézet: Az összesítő statisztika beállításai

Érdemes a modulba a változók nevét szerepeltetni (feliratok az első sorban), hiszen így az elemzéseink során mindig tudni fogjuk, hogy miről kértük az összesítő statisztikát. Az eredmények egybevágnak az előbbiekben közöltekkel, a kisebb fajta eltéréseket annak köszönhetjük, hogy a mi adataink osztályközös gyakorisági sorból származtatott „becsült" adatok voltak, addig az Excel programban a tényleges gyakorisági sor adataival számolt. A számítógép alapértelmezésben mintának tekinti az adatokat, így korrekciós tényezőt használ a számításoknál. Pl. a szórásnál a korrigált szórást használja, ahol nevezőben n-1- található.

2/7. táblázat

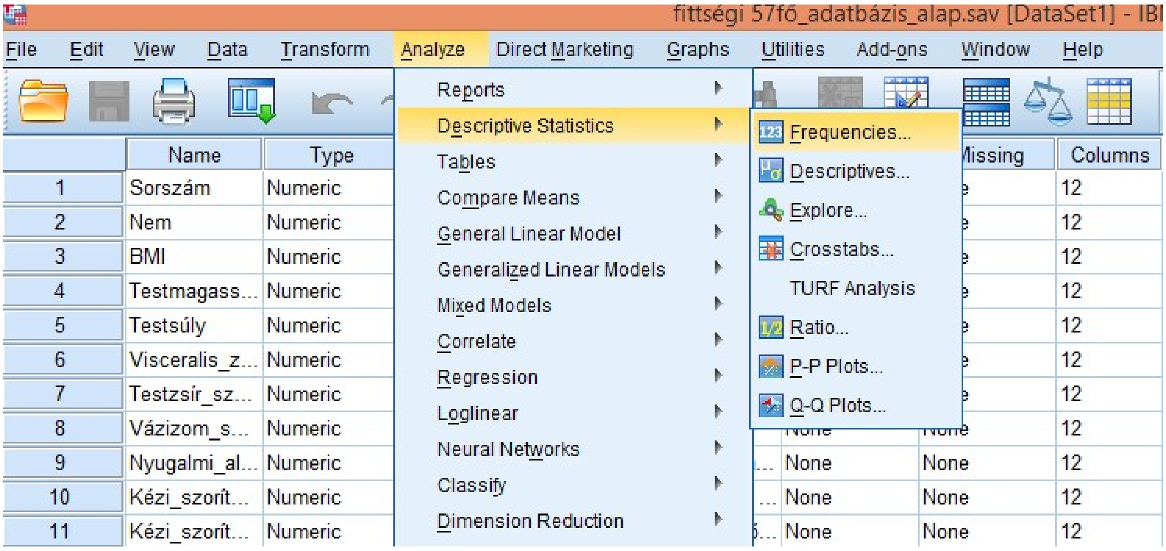
Testmagasság	
Várható érték	174,27
Standard hiba	1,07
Medián	173,00
Módusz	173,00
Szórás	8,05
Minta varianciája	64,80
Csúcsosság	0,36
Ferdeség	-0,47
Tartomány	35,50
Minimum	153,00
Maximum	188,50
Összeg	9933,50
Darabszám	57,00

Az elsőként a számtani átlagot látjuk, melyet várható értékként nevez a program, míg a tartomány címszó alatt a szórás terjedelme látható.

Az SPSS programmal három elérési útvonalon keresztül van lehetősünk a leíró statisztikai elemzés elkészítésére az ANALYZE/DESCRIPTIVE STATISTICS/DESCRIPTIVE vagy az ANALYZE/DESCRIPTIVE STATISTICS/FREQUENCIES vagy az ANALYZE/DESCRIPTIVE STATISTICS/EXPLORE modulok segítségével.

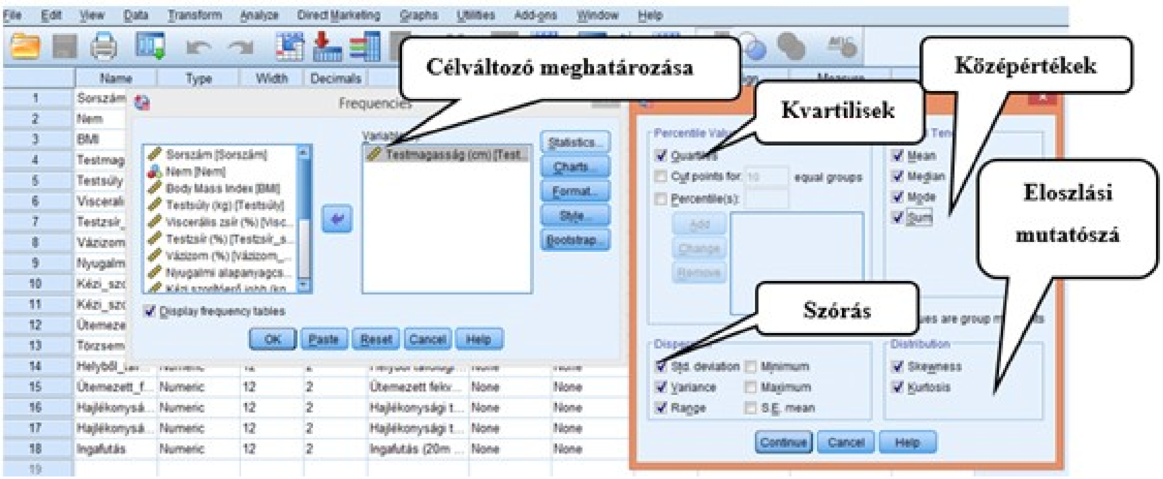
Annak eldöntésére, hogy melyik módszerrel számítjuk a leíró statisztikai módszereket kötelezően alkalmazandó szabályt mondani nem tudunk. A DESCRIPTIVE modult leggyakrabban az intervallum vagy arányskálán (SPSS-ben skála) mért változók esetében használjuk, abban az esetben, ha gyakorisági táblára nincsen szükségünk. A FREQUENCIES modul leginkább nominális vagy ordinális skálán mért változók esetében használandó, amikor gyakorisági táblára és grafikus ábrázolásra is szükségünk van. Természetesen ebben az opcióban is vizsgálhatunk arány vagy intervallumskálán mért változókat, azonban talán az eredményközlés során készített összefoglaló táblázat nem olyan látványos, mint a DESCRIPTIVE modulban. AZ EXPLORE modulban a már megismert módszertanokon túl, leíró statisztikai mutatókat is számoltathatunk, ahol a mintát részsokaságokra bonthatjuk. Az érdeklődő bővebben a Gyakorlati adatelemzés c. könyvben olvashat (Ács, 2015).

Vizsgáljuk meg a fenti példánkat a SPSS program segítségével (forrás: fittségi 57fő_adatbázis_alap.sav) Először a gyakorisági modullal készítsük el az elemzésünket (Analyze/Deccriptive Statistics/Frequencies)



2/11. képernyőnézet: Gyakorisági modul indítása.

Elsőként a „beválasztsztó panelben” válasszuk be a megfelelő változót (testmagasság), majd a Statistics gomb megnyomásával válasszuk ki a megfelelő mutatószámokat.



2/12. képernyőnézet: A gyakorisági modul beállításai az SPSS programban

A CONTINUE gomb megnyomását követően, további beállításokat nem eszközölve nyomjuk meg az OK-ot. Az eredmény (Output View) a következő lesz:

2/8. táblázat

Statistics

Testmagasság (cm)

N	Valid	57
	Missing	0
Mean		174,2719
Median		173,0000
Mode		173,00
Std. Deviation		8,04955
Variance		64,795
Skewness		-,469
Std. Error of Skewness		,316
Kurtosis		,364
Std. Error of Kurtosis		,623
Range		35,50
Sum		9933,50
Percentiles	25	170,0000
	50	173,0000
	75	180,0000

A számított eredmények megegyeznek az Excel program által leíró statisztikában közölt eredményekkel. Ezt követően a gyakorisági táblázat is alapbeállításként megjelenik.

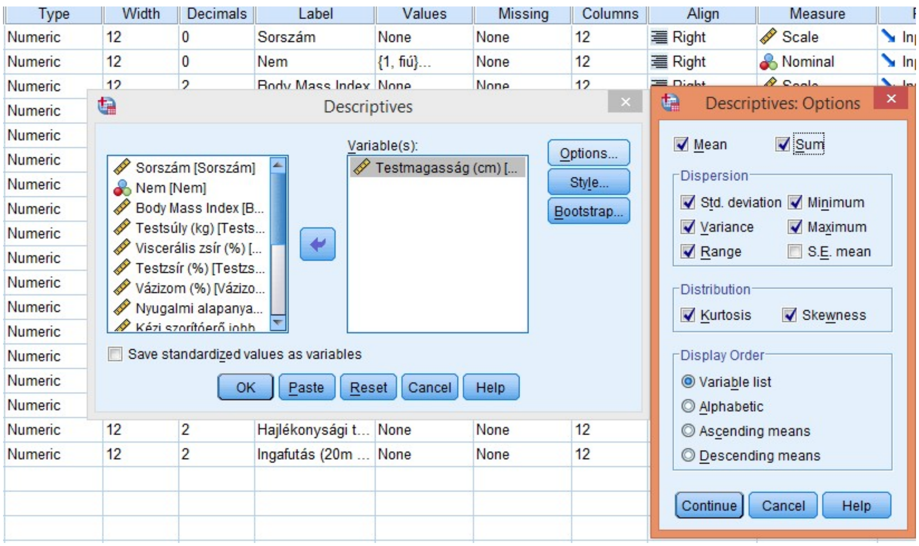
2/9. táblázat: a testmagasság gyakorisági táblázata (részlet)

Testmagasság (cm)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	153,00	2	3,5	3,5	3,5
	157,00	1	1,8	1,8	5,3
	164,00	2	3,5	3,5	8,8
	165,00	1	1,8	1,8	10,5
	166,00	2	3,5	3,5	14,0
	166,50	2	3,5	3,5	17,5
	167,00	1	1,8	1,8	19,3
	168,00	1	1,8	1,8	21,1
	170,00	4	7,0	7,0	28,1
	171,00	3	5,3	5,3	33,3
	172,00	3	5,3	5,3	38,6
	172,50	1	1,8	1,8	40,4
	173,00	6	10,5	10,5	50,9
	174,00	2	3,5	3,5	54,4
	176,00	4	7,0	7,0	61,4
	177,00	2	3,5	3,5	64,9
	178,00	5	8,8	8,8	73,7
	179,00	1	1,8	1,8	75,4
	181,00	1	1,8	1,8	77,2
	182,00	2	3,5	3,5	80,7
	183,00	1	1,8	1,8	82,5
	183,50	1	1,8	1,8	84,2
	184,00	2	3,5	3,5	87,7
	184,50	2	3,5	3,5	91,2

A modulnak a legnagyobb előnye, hogy itt lehetőség van egyből a grafikus ábrázolásra (charts) is (lásd később).

Ha a másik modult választjuk (Analyze/Descreptive Statisics/Descriptives) szintén hasonló végeredményhez juthatunk. Természetesen először itt is a változót kell beválasztani, majd az Options gomb lenyomását követően tudjuk a leíró statisztikai elemzések közül a kutatás szempontjából fontosnak vélt eljárásokat beválasztani (forrás: fittségi 57fő_adatbázis_alap.sav).



2/13. képernyőnézet: A leíró statisztika modul beállításai

A Continue, majd az Ok lenyomását követően megkapjuk a számított eredményeket, melyeket a következő táblában láthatunk:

2/10. táblázat

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Testmagasság (cm)	57	35,50	153,00	188,50	9933,50	174,2719	8,04955	64,795	-,469	,316	,364	,623
Valid N (listwise)	57											

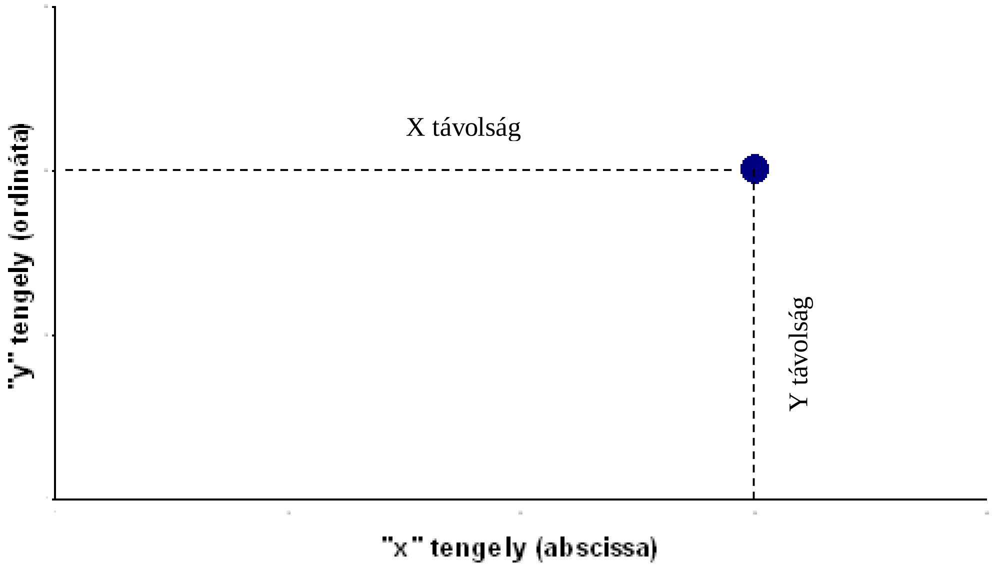
Természetesen az eredmények itt is megegyeznek az előzőekben közöltekkel.

2.7.2.4. Az adatprezentáció eszközei

Amikor az adatok száma meghaladja azt az értéket, mely egyszerűen és könnyen kezelhető, szokás az adatokat a szemléltetés és a gyors áttekinthetőség céljából tömöríteni. Ennek megfelelően hatásos és elterjedt adatprezentációs eszköz: *a grafikus ábrázolás* és az adatok *statisztikai táblázatba* rendezése.

Grafikus ábrák legfontosabb szerepe, hogy a vizsgált jelenségek fő vonásait, arányait, tendenciáit, és összefüggéseit igyekszik vizuálisan megjeleníteni. Célja az egyszerű adatközléstől a bonyolultabb kapcsolatok feltárásáig széles skálán mozoghat.

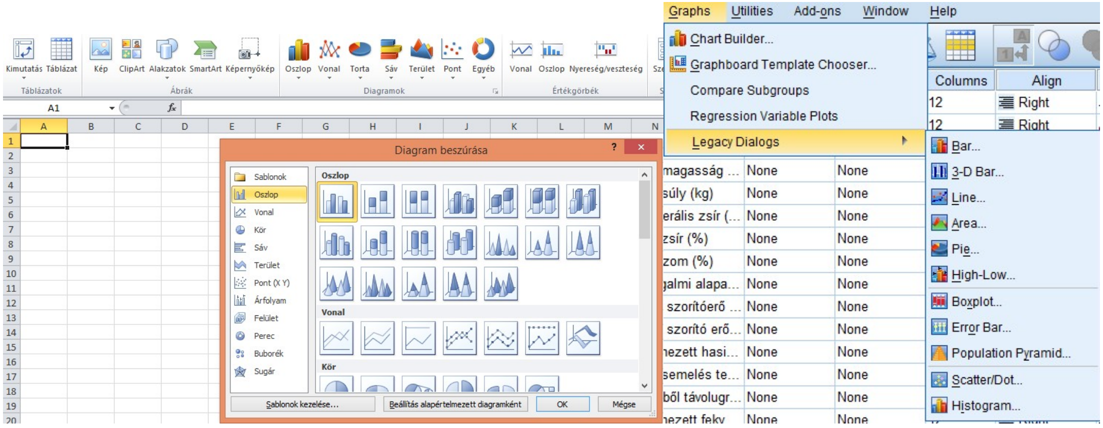
Megkülönböztetünk *egyszerű* és *összetett* statisztikai ábrákat. Az egyszerű ábrák lehetnek: pont (xy)-, oszlop-, kör-, és szalagdiagramok. Az összetett ábrák, - melyek mindig valamely statisztikai, illetve matematikai művelet eredményeként jönnek létre-, többnyire a gyakorisági sorok elemzésére szolgálnak pl.: poligon, hisztogram, ogiva, Box- plot, Lorenz-görbe, dendogram. A grafikus ábrázolás alapja a derékszögű koordináta rendszer, melynek lényegét már általános iskolában tanítják. A rendszer egy függőleges „y" tengely és egy vízszintes „x" tengelyből áll melyek a 0 pontban metszik egymást. Egy pont helyzete a két tengelytől való távolsággal meghatározható (ez a pontdiagram alapelve).



2/13. ábra: a grafikus ábrázolás alapja (derékszögű koordináta- rendszer)

A legfontosabb, hogy az „egyszerű ábra" típusok különböző adattípusok esetén alkalmazandók. Ez nem zárja ki, hogy bizonyos ábrákat más fajta adatok szemléltetésére ne lehessen használni, viszont vannak ábrák, melyek nem alkalmazhatók bizonyos esetekben. Ennek megfelelően a következő példák is csak ajánlás jelleggel mutatják be az grafikus ábrákat, legfontosabb vezérelv volt, hogy az elkészítés menetét illusztrájuk. Fontos, hogy az ábrák valóban a kutatásnak megfelelő adatokat jól illusztrálják, az egyes értékek azonosíthatók, értelmezhetők legyenek. A grafikus ábrázolásról HUNYADI 2001-es cikkében olvashat részletesebben az érdeklődő.

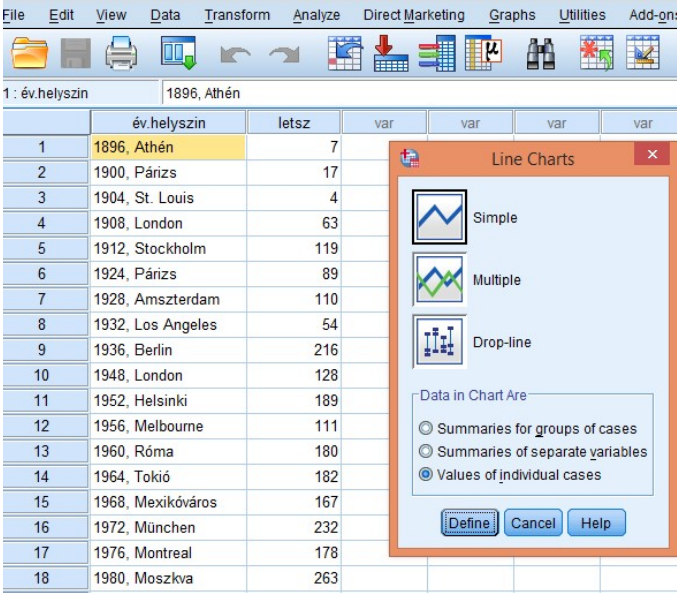
Természetesen az Excel (a. ábra) és az SPSS (b. ábra) programcsomag is rendelkezik ábrakészítő modullal.



2/14. képernyőnézet: A grafikus ábrázolás moduljai

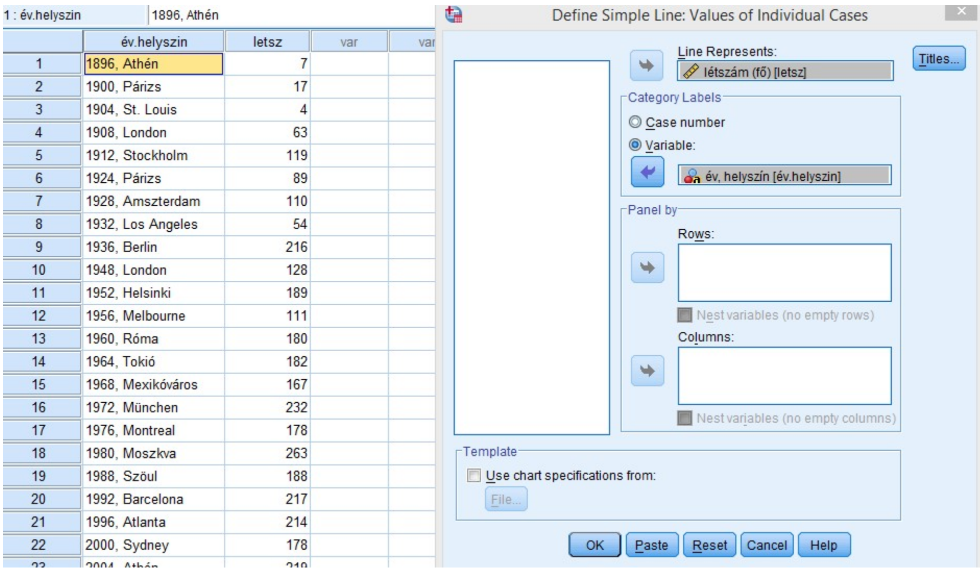
A bal oldali ábrán az Excel program diagram beszúrás modulja látható, mely a Beszúrás menüpont, Diagram almenüjéből érhető el. A jobb oldali ábra az SPSS program Graps menüpontjának lehetőségeit mutatja be. Minden grafikus ábrázolást elérhető ezekről a helyekről.

Elsőként a vonaldiagramot mutatjuk be, mely jól alkalmazható az idősorok és a dinamikus viszonyszámok elemzésénél. Az adatbázisunkat magyar sportolók olimpiákon való részvételének adatai szolgáltatják (Forrás: vonaldiagram.adat.sav). Az SPSS program Graph moduljából válasszuk a Line (vonal) opciót, mely után a következő panel látható:



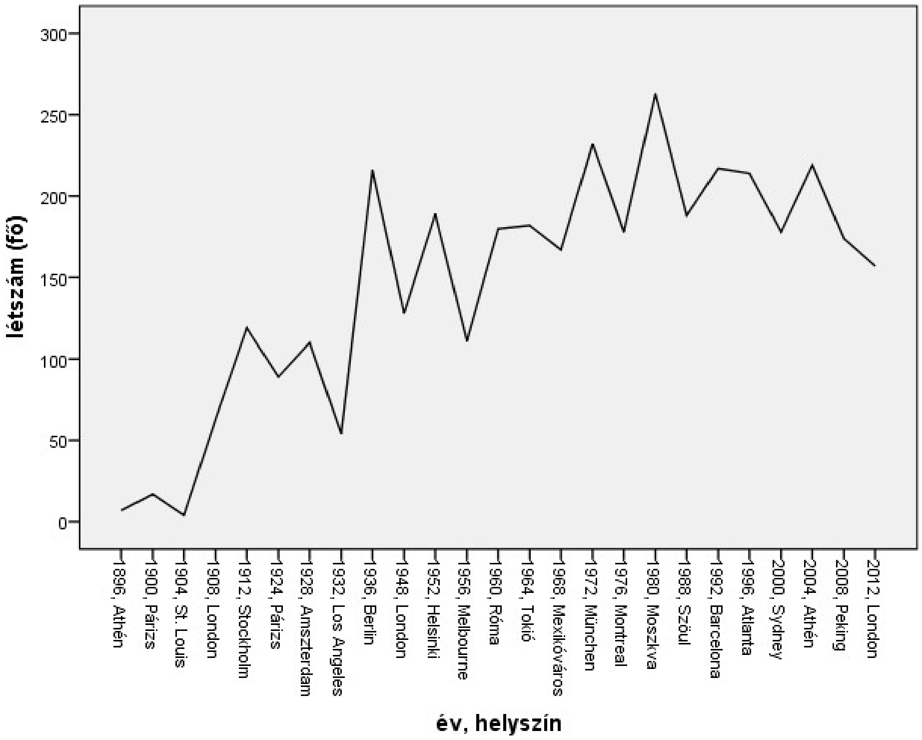
2/15. képernyőnézet: A vonaldiagram opció választása az SPSS programban

A lehetséges opciók közül most válasszuk az egyszerűt (Simple) és a egyedi adatokat tartalmazó értékeket (Values of induvidual cases), majd a Define gomb segítségével menjünk tovább.



2/16. képernyőnézet: A vonaldiagram beállításai az SPSS- ben

Ebben az ablakban a változókat kell elhelyezni, felülre kerül a létszámadat, hiszen ez fogja „diagramvonalat” adni, a kategóriaváltozó (itt az x tengely) értékeit, az évszámok és a helyszín adatok szolgáltatják.

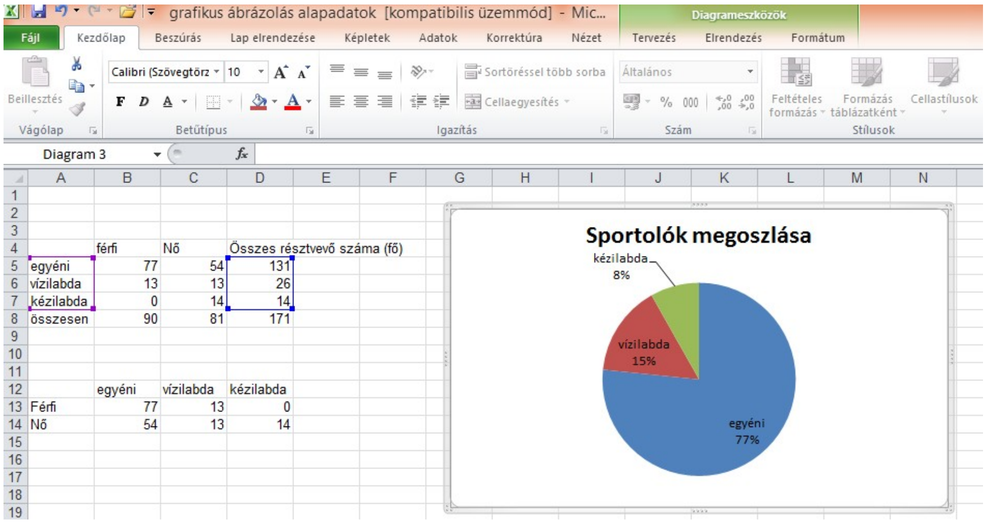


2/14. ábra

A vonaldiagram alkalmas arra, hogy megállapítsuk, hogy az olimpiákon résztvevő magyar sportolók száma tendenciájában növekvő képet mutat, bár az utolsó olimpiákon a létszámadat ismét csökkenő irányultságot vesz fel.

A kördiagram segítségével a viszonyításokat tehetünk, hiszen az egészet osztja részekre, „szeletekre”. Leggyakrabban a megoszlási viszonyszámoknál és a minőségi ismérv alapján rendezett adatok estén használható. A kördiagram az SPSS programban teljesen megegyezik az előbb leírtakkal, csak a Graph modulban most a Pie lehetőséget kell választani.

Most az Excel program segítségével vizsgáljuk azon 2008-as olimpián résztvevők sportágak (egyéni, vízilabda, kézilabda) szerinti megoszlását, mely alapadatok megtalálhatók: http://www.mob.hu (2008. július 26.) címen. Forrás: grafikus ábrázolás alapadatok.xls.



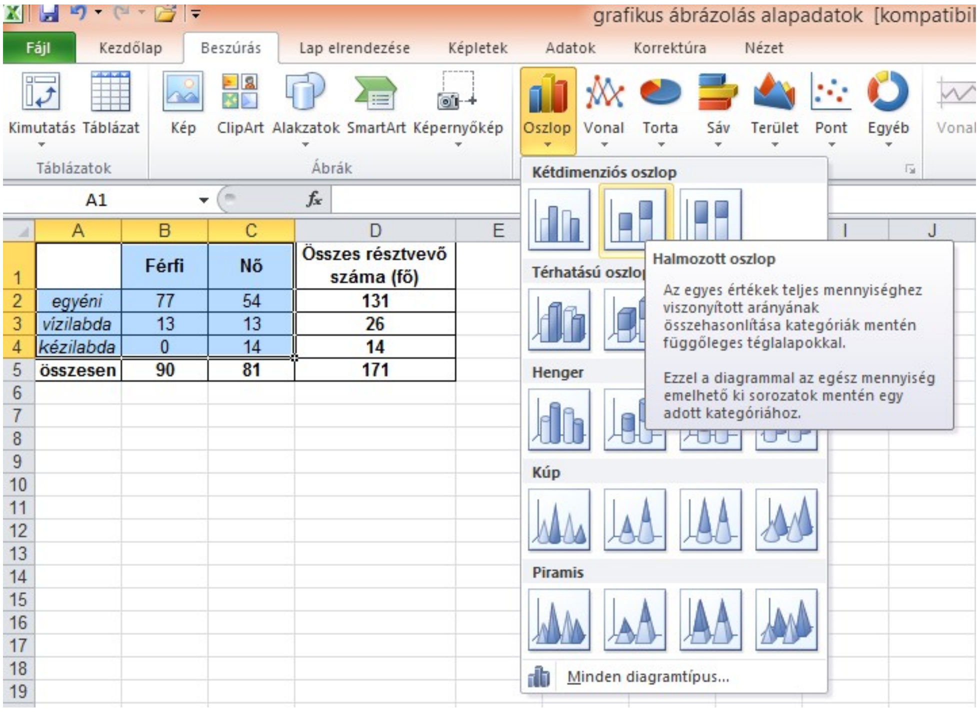
2/17. képernyőnézet: A kördiagram beállítási modulja az Excel programban

A végeredményből látszik, hogy a mintánkban az egyéni sportolók aránya a legmagasabb (77%), melyet a vízilabda (15 %) és a kézilabda (8%) követ. Ennek felhasználásával kijelenthető, hogy az egyéni sportolóink részvételi aránya (77%) magasabb, mint a csapatsportágakban (23%) érdekelteké.

Az oszlop diagram talán a legelterjedtebb grafikus ábrázolási mód, hiszen szintén alkalmas idősorok[20] megjelenítésre, illetőleg megoszlások, viszonyítások (pl. területek[21], összehasonlításokra is. prezentálására is. Vizsgáljuk meg a nemek közötti megoszlást a különböző sportágak tekintetében! Ezt az összehasonlítást oszlopdiagramon fogjuk bemutatni, melyet most a diagramvarázslóból választunk ki. Itt válasszuk a második altípust a halmozott oszlopdiagramot. A következő lépésben jelöljük, hogy az adatsoraink sorokban vannak rendezve. (Forrás: grafikus ábrázolás alapadatok.xls.)

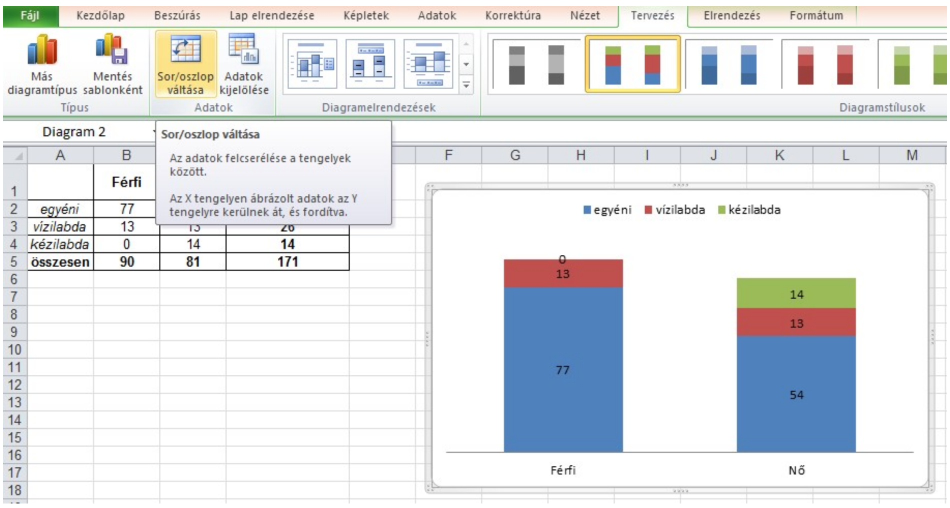
[20] Leginkább tartam idősorok esetén, amikor az adatok időtartamhoz köthetők.

[21] Területi összehasonlításoknál elterjedt a szalagdiagram alkalmazás.



2/18. képernyőnézet: Az oszlopdiagram beállítási modulja az Excel programban

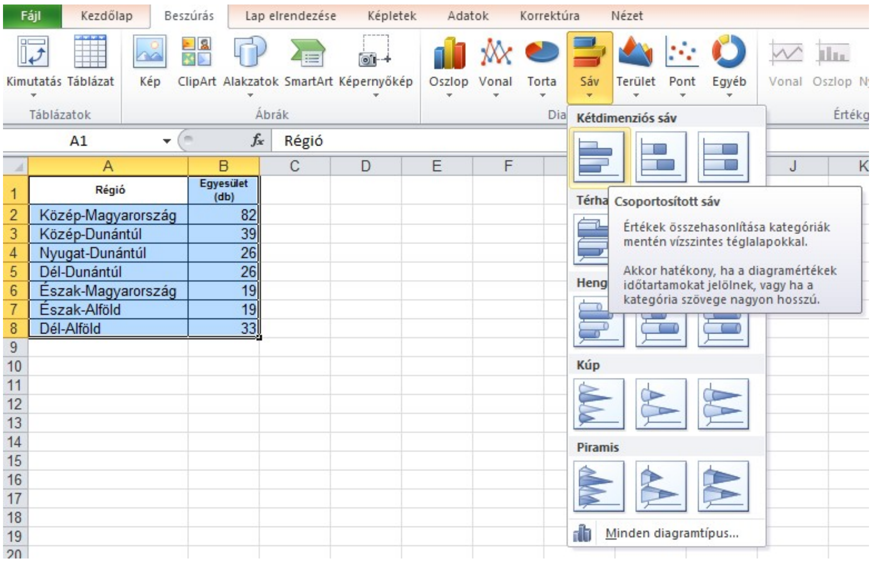
A következő panelen, kérjük, hogy az értékeket jelezze ki, majd eztán a befejezésre klikkeljünk.



2/19. képernyőnézet:

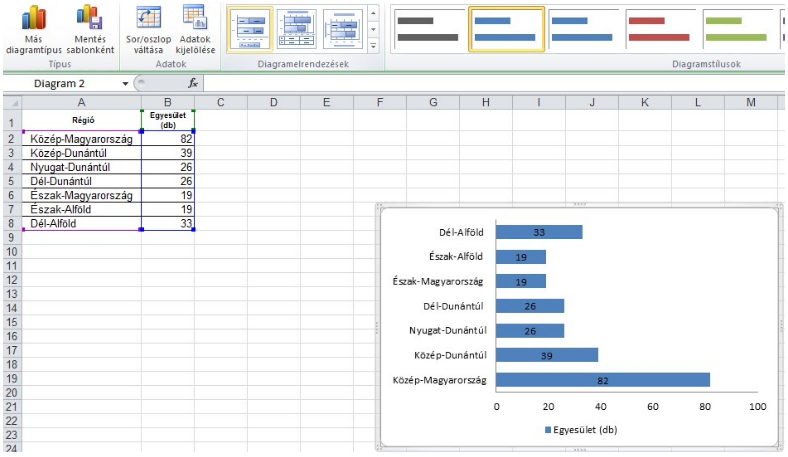
Megállapíthatjuk, hogy a végeredmény alkalmas a célként kitűzött összehasonlításra.

A következő szalagdiagramon az olimpai- keret sportolókat adó egyesületek régiók szerinti területi elhelyezkedését mutatjuk be (Forrás: grafikus ábrázolás alapadatok.xlsx.). A diagramvarázslóban a sáv diagramtípust válasszuk.



2/20. képernyőnézet: A szalagdiagram beállítási modulja az Excel programban

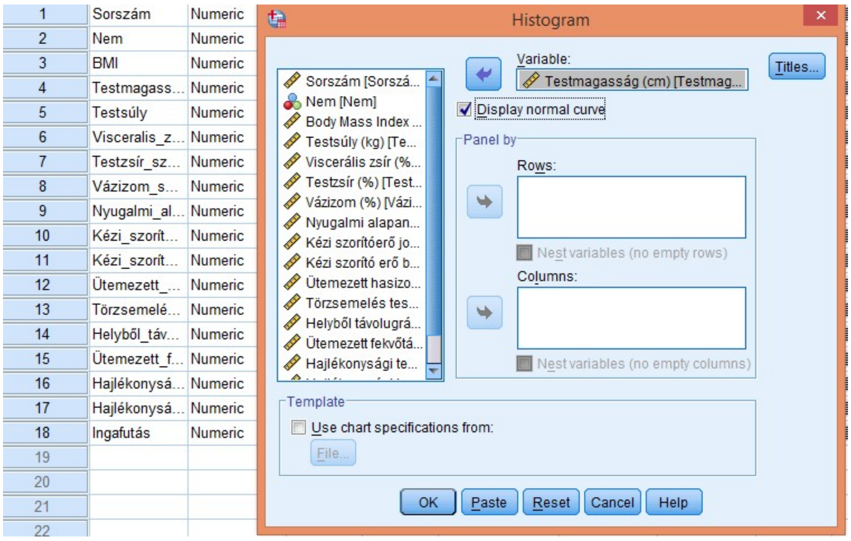
Ezt követően, ha a varázsló előírásait követjük, eljutunk a végeredményhez.



2/21. képernyőnézet

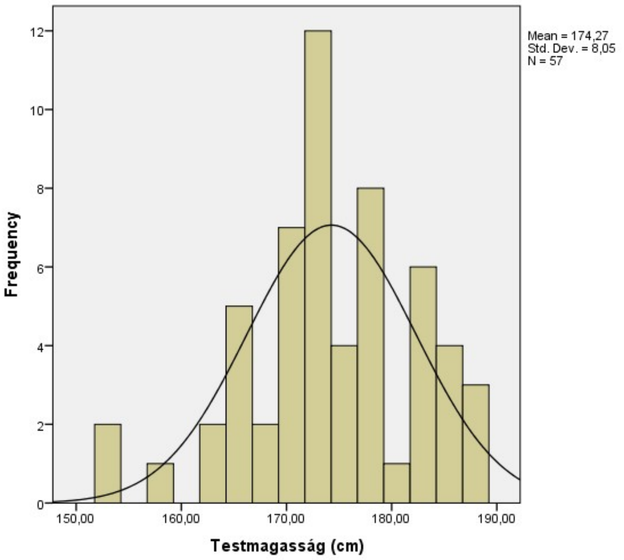
Megállapíthatjuk, hogy a közép- magyarországi régióban található a legtöbb olimpiai-keret sportolót adó egyesület, a legkevesebb az észak-alföldi, valamint az észak-magyarországi régióban van.

A hézag nélküli oszlopdiagramot hisztogramnak nevezik, és kiemelt helyen kezelik a grafikus ábrázolások között. A hisztogram a gyakorisági sorok alapvető ábrázolási módszere, oszlopainak területe arányos a gyakoriságokkal, egyenlő hosszúsági osztályközök esetén csak az oszlopok magasságára kell figyelni, amit a gyakoriságok mutatnak. Az SPSS programban a leíró statisztikai modulban (Decriptive Statiscics, Frequencies, Charts), illetve külön a grafikus ábráknál is megtalálható. Előnye, hogy a program felkínálja a normál eloszlás ábrájának az ábrázolását is. A következő példán a felmért magasugrások eredményeinek hisztogramját készítjük el. (forrás: fittségi 57fő_adatbázis_alap.sav)



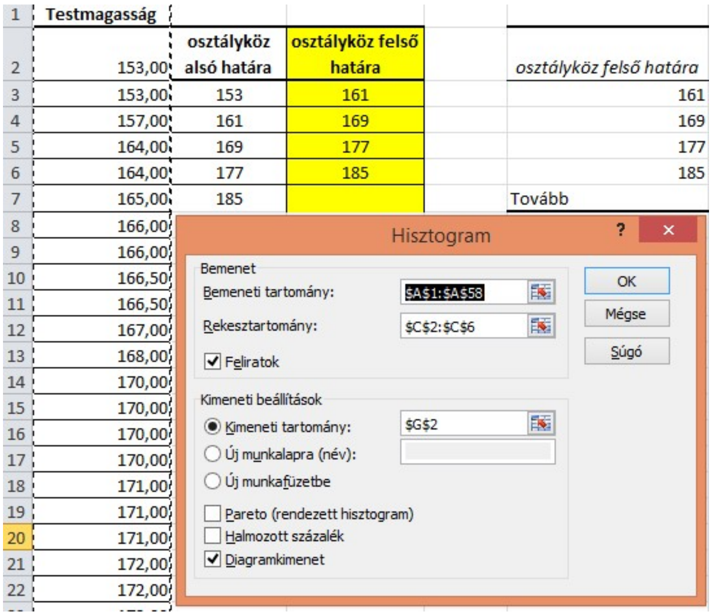
2/22. képernyőnézet: Hisztogram készítése az SPSS program segítségével

A végeredmény a következő ábrán látható.



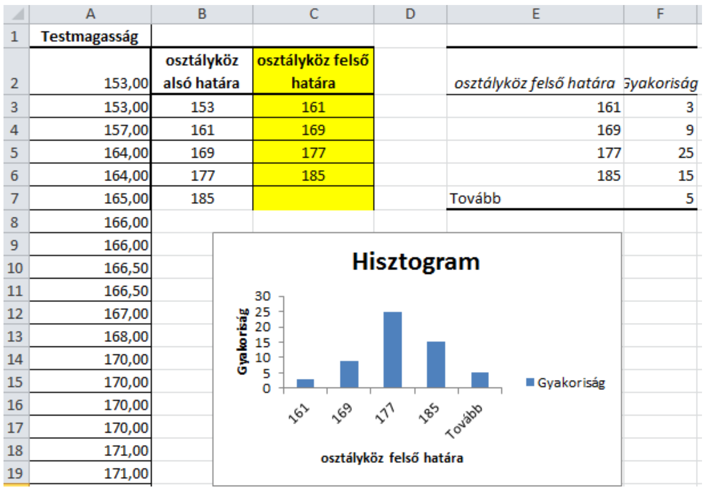
2/15. ábra

Hisztogram az Excel programban is könnyen készíthető az adatelemzés almenüből elérve, külön modul foglalkozik vele. Előnye, hogy a gyorsan osztályközös gyakorisági sor képezhető a segítségével, hiszen a bemeneti tartományon kívül a rekesztartomány is kijelölhető, amely az osztályközeink meghatározására szolgál (forrás: osztályközös gyakoriság. xlsx)



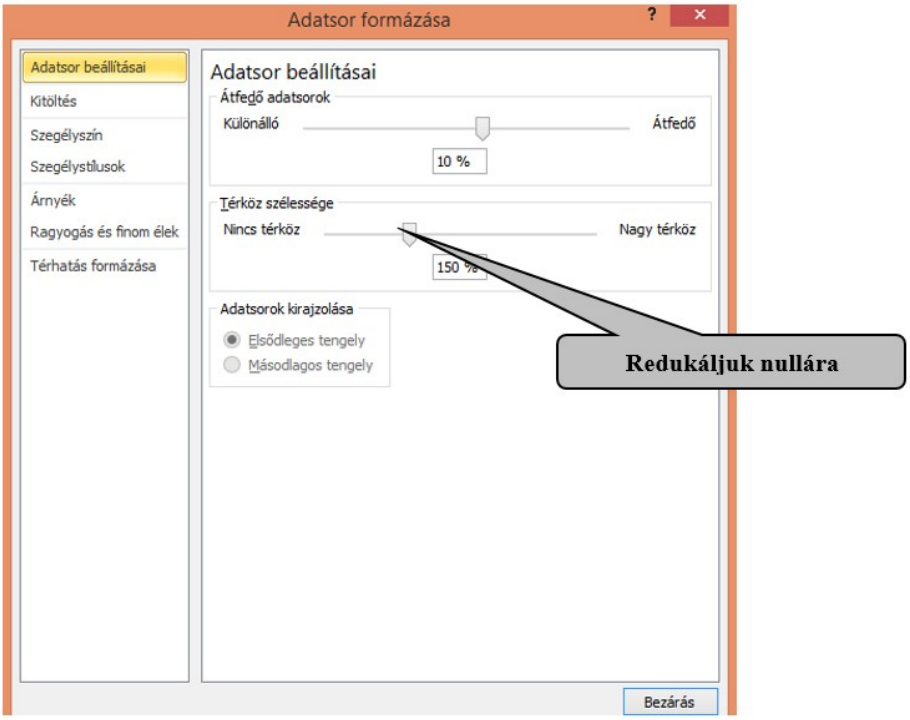
2/23. képernyőnézet: Hisztogram készítése az Excel program segítségével

Jelöljük ki a diagramkimenet és a rekesztartományba tüntessük fel a számított osztályközeink **felső határát**. Az így keletkező eredmény a következő ábrán látható.



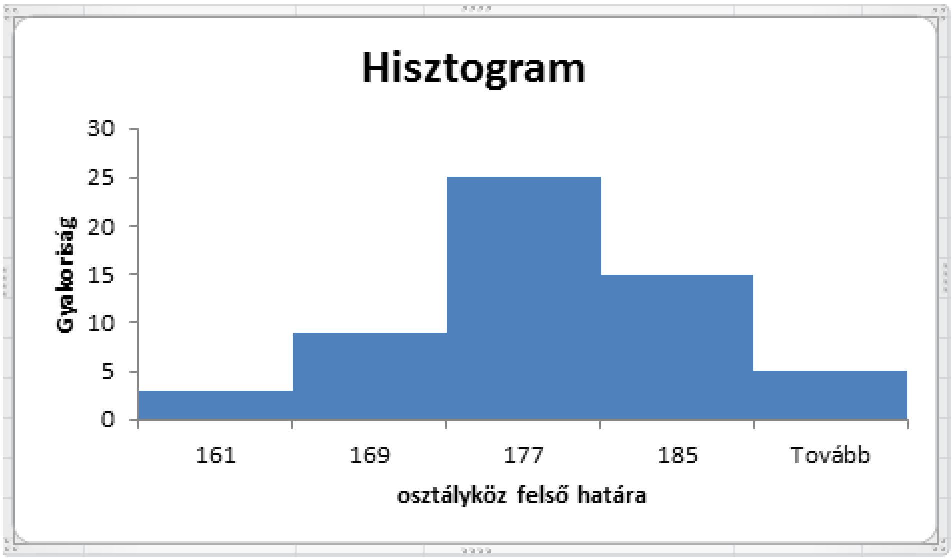
2/24. képernyőnézet

Láthatóvá válik, hogy az egyes osztályokhoz (rekeszekhez) tartozó gyakoriságok megegyeznek az előzőekben tárgyaltakkal, viszont a grafikus ábrázolás oszlopdiagramot ad, és nem hisztogramot, mivel vannak hézagok. Ez is könnyen megoldható, ha kétszer valamelyik oszlopra kattintunk. A felnyíló ablakban válasszuk a beállítások menüt, majd itt a köznél változtatható értéket vegyük vissza nullára.



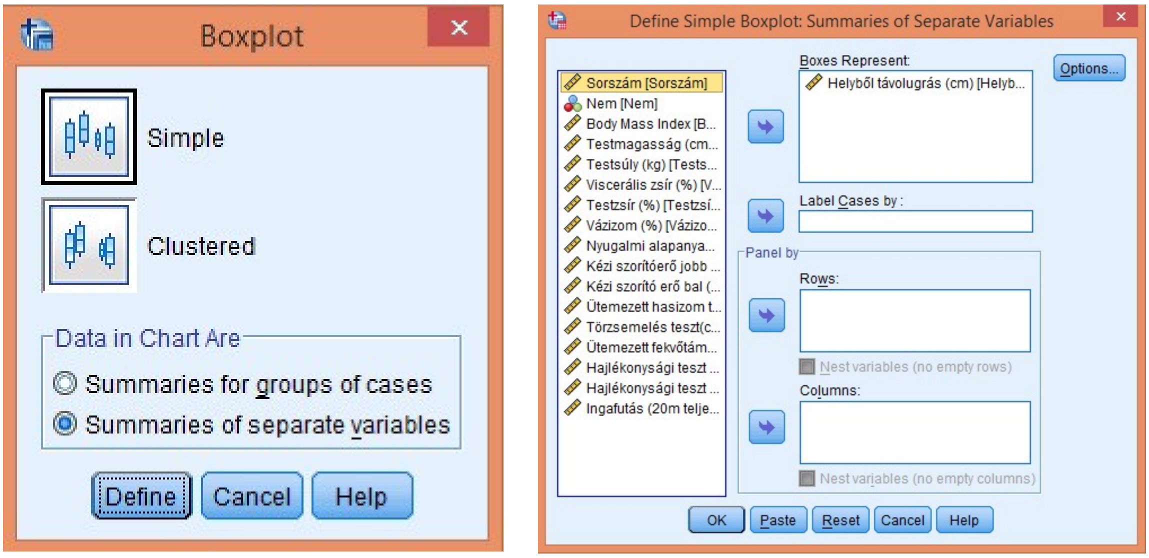
2/25. képernyőnézet: Az adatsorok formázása

Az így keletkező diagram lesz a hisztogram, melyet a következőkben szemléltetünk.



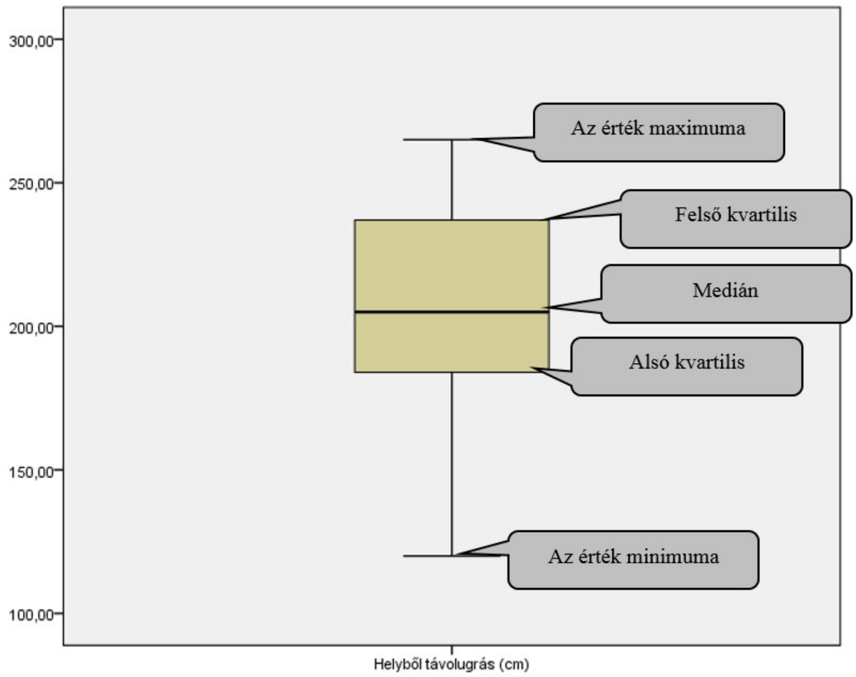
2/16. ábra

Az összetett ábrázolások közül népszerű a „Box-plot” ábra, hiszen a mennyiségi sor fontosabb jellemzőit (átlagot, kvartiliseket, terjedelmet, kiugró értékeket) ábrázolja. Ez az ábrázolási mód csak az SPSS programban (Graph/ Boxplot) érhető el. A helyből távolugrás ábráját láthatjuk. (forrás: fittségi 57fő_adatbázis_alap.sav)



2/26-27. képernyőnézet: A „Box-plot” beállítási moduljai

Az első beállításnál válasszuk az egyszerű típust és a Summaries of separate variables-t, hiszen az adataink nincsenek csoportosítva. A Define gomb lenyomást követően célváltozónak adjuk meg a helyből távolugrás változót. Az Ok gomb megnyomásával következő az ábrához jutunk.



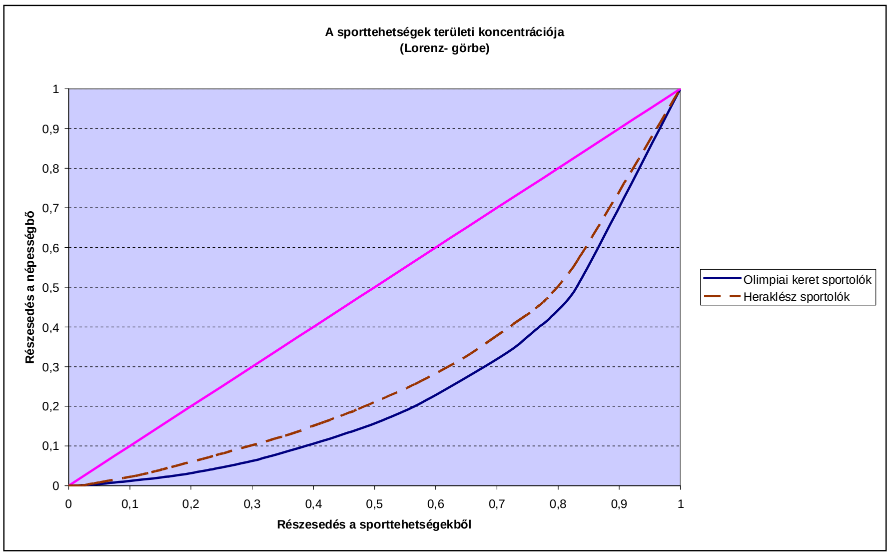
2/17. ábra: Box- plot

Egy másik összetett grafikus ábrázolási mód a Lorenz-görbe, melyet a koncentráció elemzésekor szoktak alkalmazni. A területi koncentráció elemzésekor az egyik leggyakrabban használt eljárás a Lorenz-görbe ábrázolása, ami tulajdonképpen a koncentrációs tábla grafikus megjelenítése.

Ez egy egységnyi oldalú négyzetben elhelyezett ábra, mely a kumulált relatív gyakoriságok (g_i') függvényében ábrázolja a kumulált relatív értékösszegeket (z_i'). Mindenképpen fel kell hívni a figyelmet, hogy a Lorenz- görbe grafikus megjelenítésre és összehasonlításra használt módszer során nem javasolt egyetlen görbét feltüntetni, mivel így teljes bizonyossággal nem lehet megállapítani, hogy a vizsgált jelenség területi egyenlőtlensége, milyen mértékű.

Az eljárást gyakran alkalmazzák, mivel ugyanazt a jelenséget, ha több időpontban ábrázoljuk, könnyen kapott információhoz juthatunk a területi egyenlőtlenség (koncentráció) változásáról.

A Lorenz-görbe elkészítése: a grafikus ábrázolás előtt, egy adott relatív mutató szerint, csökkenő vagy növekvő sorrendben kell állítani a vizsgált adatokat, jelen esetben a térségeinket. Ha az adatokat növekvő sorrendbe rendezzük, akkor a görbe az átlónk alá kerül, csökkenő sorrend esetén az átló fölé. Ács egy kutatása alkalmával (Ács, 2007) a sportolói tehetségek hazai területi koncentrációját vizsgálta, hipotézise szerint az idősebb sportolóknál (olimpiai-keret sportolók) nagyobb területi koncentráció tapasztalható, mint a fiatal tehetségeknél (Heraklész sportolók), mely leginkább a pénznek - sportolói fizetésnek- köszönhető.



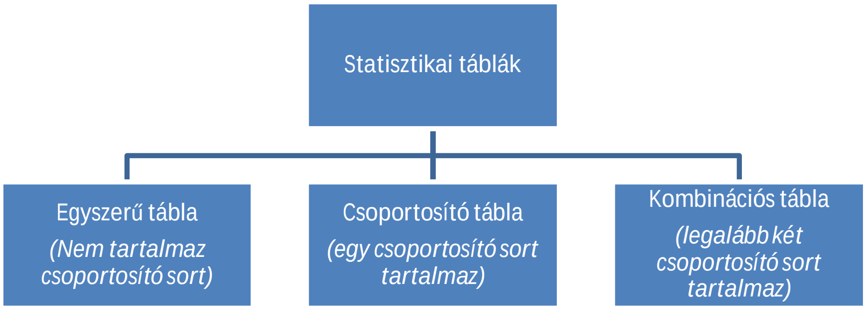
2/18. ábra: Lorenz-görbe

Forrás: Ács P. (2007)

A görbe értelmezése: ha létezne egy olyan területegység, amelyik a vizsgált ismérv értékösszegének nagy hányadát lekötné, vagyis a relatív gyakoriságok és a relatív értékösszegek igen nagymértékben eltérnének egymástól, akkor a görbe az átlótól távol esne, illetve a teljes koncentráció esetén a görbe az egységnyi oldalú négyzet oldalaival esne egybe.[22] Amennyiben az egységeknek az értékösszegből való részesedése azonos, a kumulált relatív gyakoriságok és a kumulált relatív értékösszegek megegyeznek ($g_i' = z_i'$), ilyenkor a görbe az átlóval egybeesik, ami jelzi nekünk a koncentráció hiányát, vagyis az abszolút egyenlőséget, koncentrálatlanságot.

Az ábra szemlélteti, hogy az olimpiai-keret sportolók területi elhelyezkedésében nagyobb koncentráció tapasztalható. A görbe jól szemlélteti ugyan a területi koncentrációt, de számadattal nem szolgál a koncentráció nagyságáról. A statisztikai tábla szintén gyakran előforduló adatprezentációs eszköz, hiszen a célja az adatok rendezése, tömörítése.

Statisztikai tábla a statisztikai sorok rendszere, melyben az adatok egy, illetve több ismérv szerint lehetnek felsorolva. A statisztikai táblák statisztikai sorokat (idő-, területi-, minőségi, mennyiségi sor) tartalmaznak. A táblákat általában két szempont szerint szokás tipizálni. A *dimenziószám szerint* leginkább két vagy háromdimenziós táblákkal találkozhatunk. Ennek eldöntése a táblában található ismérvek (változók) számától függ. A másik tipizálást az ismérvek felsorolásának a célja határozza meg, hiszen történhet összehasonlítás vagy csoportosítás kedvéért. Ennek megfelelően *típus szerint* a következő táblákat különböztetjük meg:



2/19. ábra: A statisztikai táblák csoportosítása

A statisztikai táblák többsége kombinációs tábla. Abban az esetben, ha a táblában gyakorisági sorok szerepelnek, vagyis a felsorolt adatok gyakoriságok, **kontingencia tábláról** beszélünk. A következő tábla, egy háromdimenziós kontingencia táblát mutat be.

---

[22] Hajdu (1997)

2/11. táblázat: Néhány csapatsportág régiók és nem szerint (2005/2006)

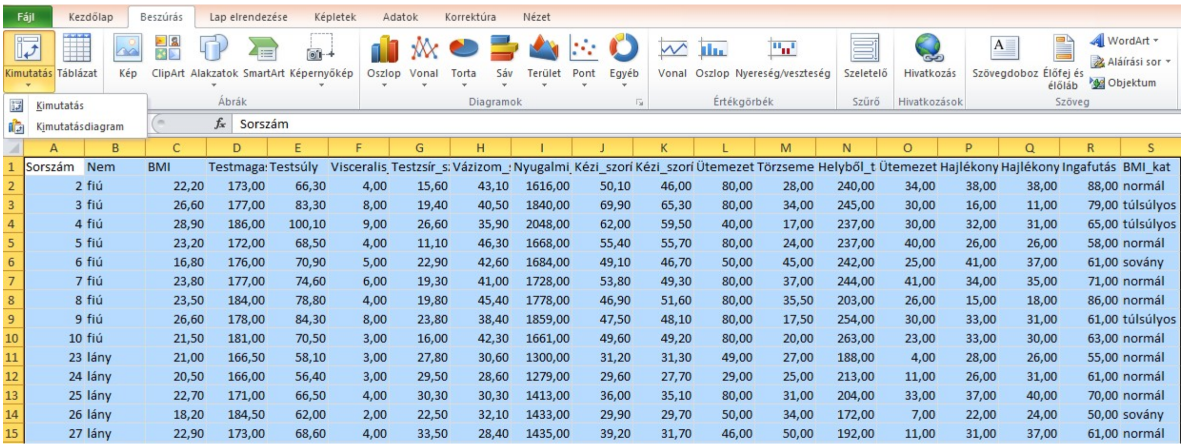
Régió	Kézilabda			Kosárlabda			Röplabda			Együtt		
	férfi	nő	összesen	férfi	nő	összesen	férfi	nő	összesen	férfi	nő	összesen
Közép- Magyarország	2	4	6	1	2	3	0	4	4	3	10	13
Közép- Dunántúl	3	2	5	1	0	1	2	0	2	6	2	8
Nyugat- Dunántúl	1	1	2	4	4	8	0	0	0	5	5	10
Dél-Dunántúl	1	0	1	4	2	6	2	0	2	7	2	9
Észak-Magyarország	1	0	1	0	1	1	1	1	2	2	2	4
Észak- Alföld	2	1	3	3	1	4	2	2	4	7	4	12
Dél- Alföld	2	3	5	1	1	2	1	1	2	4	5	9
Összesen	**12**	**11**	**23**	**14**	**11**	**25**	**8**	**8**	**16**	**34**	**30**	**64**

Forrás: Pintér-Ács (2006)

A statisztikai táblákkal szemben fontos formai meghatározások vannak, melynek hiánya csökkentheti a kutatások (diplomamunkák, szakdolgozatok) megítélését. Ezek a formai követelmények: a cím, a forrás és a magyarázó szövegek feltüntetése. Tartalmi követelmény (teljes körűség, besorolhatóság), hogy minden egyednek kell találni kizárólag egy helyet, ahová el tudjuk a rá vonatkozó adatok alapján helyezni.

A statisztikai táblákat mindkét program segítségével készíthetünk. Először nézzük, hogy az Excel program segítségével, miként készíthetünk gyorsan statisztikai táblát a fittségi adatbázisból. (Forrás: fittségi 57fő_adatbázis_alap_bmikat. xlsx).

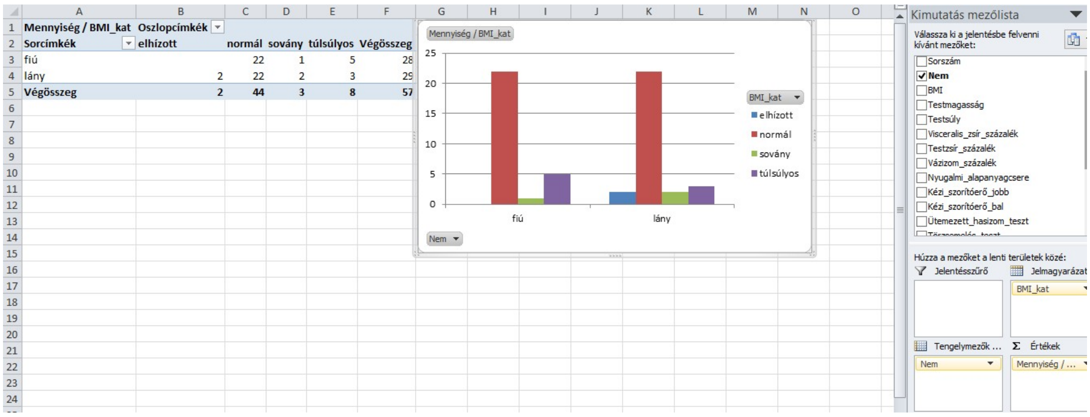
Készítsünk statisztikai táblázatot a hallgatók neme és BMI kategóriái segítségével (sovány, normál, elhízott, túlsúlyos). Két lehetőségünk van a szekunder adatokból a kimutatás, táblázat elkészítésére. A beszúrás menü kimutatás moduljában két lehetőséget is felkínál a program (kimutatás, kimutatásdiagram).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Sorszám	Nem	BMI	Testmaga:	Testsúly	Visceralis_	Testzsír_s:	Vázizom_:	Nyugalmi_	Kézi_szorí	Kézi_szorí	Ütemezet	Törzseme	Helyből_t	Ütemezet	Hajlékony	Hajlékony	Ingafutás	BMI_kat
2	2	fiú	22,20	173,00	66,30	4,00	15,60	43,10	1616,00	50,10	46,00	80,00	28,00	240,00	34,00	38,00	38,00	88,00	normál
3	3	fiú	26,60	177,00	83,30	8,00	19,40	40,50	1840,00	69,90	65,30	80,00	34,00	245,00	30,00	16,00	11,00	79,00	túlsúlyos
4	4	fiú	28,90	186,00	100,10	9,00	26,60	35,90	2048,00	62,00	59,50	40,00	17,00	237,00	30,00	32,00	31,00	65,00	túlsúlyos
5	5	fiú	23,20	172,00	68,50	4,00	11,10	46,30	1668,00	55,40	55,70	80,00	24,00	237,00	40,00	26,00	26,00	58,00	normál
6	6	fiú	16,80	176,00	70,90	5,00	22,90	42,60	1684,00	49,10	46,70	50,00	45,00	242,00	25,00	41,00	37,00	61,00	sovány
7	7	fiú	23,80	177,00	74,60	6,00	19,30	41,00	1728,00	53,80	49,30	80,00	37,00	244,00	41,00	34,00	35,00	71,00	normál
8	8	fiú	23,50	184,00	78,80	4,00	19,80	45,40	1778,00	46,90	51,60	80,00	35,50	203,00	26,00	15,00	18,00	86,00	normál
9	9	fiú	26,60	178,00	84,30	8,00	23,80	38,40	1859,00	47,50	48,10	80,00	17,50	254,00	30,00	33,00	31,00	61,00	túlsúlyos
10	10	fiú	21,50	181,00	70,50	3,00	16,00	42,30	1661,00	49,60	49,20	80,00	20,00	263,00	23,00	33,00	30,00	63,00	normál
11	23	lány	21,00	166,50	58,10	3,00	27,80	30,60	1300,00	31,20	31,30	49,00	27,00	188,00	4,00	28,00	26,00	55,00	normál
12	24	lány	20,50	166,00	56,40	3,00	29,50	28,60	1279,00	29,60	27,70	29,00	25,00	213,00	11,00	26,00	31,00	61,00	normál
13	25	lány	22,70	171,00	66,50	4,00	30,30	30,30	1413,00	36,00	35,10	80,00	31,00	204,00	33,00	37,00	40,00	70,00	normál
14	26	lány	18,20	184,50	62,00	2,00	22,50	32,10	1433,00	29,90	29,70	50,00	34,00	172,00	7,00	22,00	24,00	50,00	sovány
15	27	lány	22,90	173,00	68,60	4,00	33,50	28,40	1435,00	39,20	31,70	46,00	50,00	192,00	11,00	31,00	37,00	61,00	normál

2/28. képernyőnézet: A statisztikai táblázat (kimutatás) elkészítésének menüpontjai

Jelen esetben válasszuk a „kimutatásdiagramot”, amikor a táblázat elkészítésén túl egy diagram ábra is készül. Ezt az opcióra választva, rákérdez a program arra, hogy az adatokat Excel alkalmazásból vagy külső forrásból kívánjuk exportálni, illetve az adatok tartományának kiterjesztésére. Itt maradhat az alapbeállítás és nyomhatjuk az OK opciót.



2/29. képernyőnézet: A kimutatásdiagram beállításai

Első lépésben a sorcímkéket (tengelymező), illetve az oszlopcímkéket (jelmagyarázat) kell megadni. Ezt úgy tudjuk megtenni, hogy a képernyőn jobba alsó sarkában található négy négyzetben a megfelelő változókat jelenítjük meg. A jelmagyarázat négyzetbe kerül BMI_kategóriák, tengelymező négyzetbe a Nem változó, míg az érték mezőben a BMI mennyiségek kerülnek.

Eredményeknél jól látszik, hogy a táblázat mellett létrejött az oszlopdiagram is, mely a nemek vonatkozásában mutatja az egyes BMI kategóriákba tartozó egyedek értékét (számát).

Az SPSS programban a táblázatkészítést az Analyze menüből a Custom Tables modulon keresztül érhető el. Első lépésben a változókat kell a sor és oszlop helyekre mozgatni. Ezt úgy tehetjük meg, hogy a változót kijelölve a sor vagy oszlop mezőig húzzuk őket. Ezt kijelölve továbblépünk, akkor a következő képernyőhöz jutunk.

2/30. képernyőnézet: Táblázat készítése az SPSS programmal.

A sor és csoportösszegzés a Totals modullal érhető el, ahol jelöljük ki a Table-margin totals lehetőséget. A Format modulban kérjük, hogy a hiányzó értékek helyére nullát írjon, válasszuk a lehetőségek közül a zero opciót. A Continue és Ok gomb lenyomását követően megkapjuk a kért statisztikai táblázatot.

2/12. táblázat

		BMI kategóriák			
		sovány	normál	túlsúlyos	elhízott
		Count	Count	Count	Count
Nem	fiú	1	22	5	0
	lány	2	22	3	2

Természetesen a táblázatunkat tovább formázhatjuk, szépíthetjük, amennyiben a táblázatra állva először a baloldali egér billentyűt kétszer, illetve a jobb oldali billentyűt egyszer megnyomjuk. Ugyancsak van lehetőségünk a táblázat dimenziószámának növelésére, a kategóriák bővítésére. A táblázatok elkészítése mindenképpen gyakorlást igényel, hiszen itt most az alapbeállításokat mutattuk be. Azon érdeklődő, aki további információt szeretne kapni az SPSS-ben való táblázatok készítéséről erre vonatkozó fejezettel találkozhat Ács Gyakorlati adatelemzés c. könyvében.
Látható, hogy mindkét programmal viszonylag könnyen és gyorsan elkészíthetők az alaptáblázatok.

2.7.3. Kétváltozós kapcsolatok elemzése

A bennünket környező világ megismerése során a statisztikai változók kapcsolatainak megismerése igen fontos, és az így nyert információk fontos szerephez jutnak egy-egy döntés előkészítése, illetve a döntések hatásának mérése kapcsán. A sport területén vizsgálódva is jobban megérthetjük a vizsgált jelenségeket, ha nem kizárólag önmagukban nézzük őket, hanem egymáshoz való viszonyukat, más jelenségekkel való összefüggéseiket is vizsgáljuk. Vizsgálhatjuk egy csapat eredményességét - a bajnokságban szerzett pontjait- az eddig megismert módon (pl. átlag, szórás, stb.), de pontosabb képet kapunk, ha kiterjesztjük vizsgálódásunkat más egyéb befolyásoló tényezőkre (változókra) is. Ha bevonjuk a vizsgálatokba a játék helyszínét is (hazai pálya, idegen pálya), valószínűleg pontosabb képet kapunk, hiszen köztudott, hogy a hazai pályán a csapatok nagy része támadóbb taktikával játszik.

A jelenségeket, folyamatokat leginkább a következőképpen tipizálhatjuk:

- az ismérvek egymástól **függetlenek**, ha az egyik ismérvből semmilyen következtetés nem vonható le a másik ismérvre nézve,
- **sztochasztikus kapcsolatnál,** a statisztikai ismérvek között tendenciaszerűen, valószínűségi jelleggel érvényesülő kapcsolatot értünk. Számunkra ez lehet a legérdekesebb eset, hiszen az azt jelenti, hogy az ismérvek körében nagy eséllyel van valamiféle kapcsolat, „együtt alakulás”,
- **függvényszerű, determinisztikus** kapcsolatnál az egyik ismérv meghatározza, hogy a másik ismérv milyen értékeket vehet fel.

A kapcsolatokat megkülönböztetjük a bennük szereplő ismérvfajták alapján, ezek szerint három típust különbözetünk meg:

- Mindegyik (mindkét) ismérv minőségi (kategóriás ismérv, hiszen az ismérvértékek kategóriákat képeznek), azaz nominális típusú, akkor **asszociációs kapcsolatról** beszélünk.
- **Vegyes kapcsolatnál**, az ok szerepét minőségi, de az okozat szerepét mennyiségi ismérv tölti be.
- Ha mindegyik (mindkét) ismérv mennyiségi, akkor **korrelációs kapcsolatról** beszélünk.

Mindhárom típusú kapcsolatot valamilyen mutatószám segítségével célszerű számszerűsíteni. Fontos szerepet töltenek be az ún. kapcsolat intenzitását kifejező mérőszámok, amelynek általános sémáját az alábbiakkal írhatjuk le (az intenzitást általánosságban mérő mutatószám jele legyen T):

$$0 \leq T \leq 1$$

Általánosságban a T mutató abszolút értékére fogalmazzuk meg a fenti intervallumot, de bizonyos esetben – főleg a korrelációs kapcsolatokban - az előjel is fontos információ hordozója, ugyanis a kapcsolat pozitív vagy negatív irányát mutatja. Természetesen ilyen esetben a mutatószám intervalluma: $\left[-1;1\right]$.

A mutatószámok értelmezése mindig függ az adott problémától; ismerni kell a vizsgált összefüggés jellegét, természetét. Az általános elemzéshez ad segítséget az alábbi séma:

2/13. táblázat

$T = 0$	nincs kapcsolat
$0 < T < 0,3$	gyenge a kapcsolat
$0,3 \leq T \geq 0,7$	közepes szorosságú a kapcsolat
$0,7 < T < 1$	erős a kapcsolat
$T = 1$	függvényszerű vagy determinisztikus

2.7.3.1. Asszociációs kapcsolat vizsgálata

A kapcsolat alapfeltétele, hogy az ismérvek mindegyike minőségi legyen. Az asszociációs kapcsolat esetében az adatokat egy kombinációs táblába rendezzük, ha a kombinációs táblában gyakoriságok szerepelnek, akkor – ún. kontingencia tábláról beszélünk. Ha a kérdőívekben a zárt kérdések magválaszolásához csak kettő (alternatív ismérv), egymást kizáró ismérvváltozat van megadva pl. nő-férfi, igen-nem, van-nincs stb, akkor a kétdimenziós kontingencia tábla általános alakja a következő lesz:

2/14. táblázat: A kétdimenziós kontingencia tábla általános alakja

A ismérv Változatai	B ismérv változatai		**Összesen:**
	B_1	B_2	
A_1 A_2	f_{11} f_{21}	f_{12} f_{22}	S_1 S_2
Összesen:	O_1	O_2	n

n – az összes elemszám,

f_{11} – az A ismérv első és a B ismérv első változatához rendelt gyakoriság (hasonlóan értelmezhetők a többi cella gyakoriságai is!),

S_1 – az első sor (az A ismérv első változatához tartozó) gyakoriságok összege,

O_1 – az első oszlop (a B ismérv első változatához tartozó) gyakoriságok összege.

Belátható az alábbi összefüggés:

$$S_1+S_2 = O_1+O_2 = n$$

A sorok, illetve az oszlopok összegeit **peremgyakoriságoknak** nevezzük.

Alternatív ismérvek esetén a kapcsolat mérésére alkalmazhatjuk az ún. **Yule-féle mutatót**, ami a táblában szereplő gyakoriságok "kereszt-szorzataiból" állítható elő:

$$Y = \frac{f_{11} \times f_{22} - f_{12} \times f_{21}}{f_{11} \times f_{22} + f_{12} \times f_{21}}$$

A mutatószám – mivel két adat különbségének és ugyanazon adatok összegének hányadosa – minden esetben -1 és +1 közötti értéket vesz fel.

A következő példa egy valós kutatásból[23] származik, melyet a Pécsi Tudományegyetem Testnevelési- és Sporttudományi Intézetben folytattak az Önkormányzati és Területfejlesztési Minisztérium Sport Szakállamtitkárságának megbízásából 2008-ban.

A példa adatait a következő két kérdés válaszai szolgáltatták:

MILYEN NEMŰ VAGY? (tegyél „x”-et a megfelelő kockába!)

Fiú ☐ Lány ☐

FOGYÓKÚRÁZTÁL-E VALAHA? (tegyél „x”-et a megfelelő kockába!)

Igen ☐ Nem ☐

A válaszok eredményeit a kontingencia táblából olvashatjuk le:

2/15. táblázat

Count

		Fogyokuraztal valaha?		
		igen	nem	összesen
MILYEN NEMŰ VAGY?	fiú	53	261	314
	leány	117	165	282
összesen		170	426	596

A képletbe behelyettesítve a következő eredményt kapjuk:

$$Y = \frac{53 \times 165 - 261 \times 117}{53 \times 165 + 261 \times 117} = \text{-0,55}$$

Az eredmény abszolút értéke 0,3 és 0,7 közé esik, tehát erősségét tekintve közepes erősségű kapcsolatot találtunk a két ismérv között, vagyis a „fogyókúrázottak” számának

[23] A dél-dunántúli régióra nézve reprezentatív kutatás címe: **Fizikai aktivítás és életminőség a serdülő korúaknál.** A kutató team tagjai: Dr. Rétsági Erzsébet, Pótó Zsuzsanna, Dr. Ács Pongrác.

növekedésével csökken a fiú válaszadók száma. A mutatószám alkalmazása során azonban fokozottan figyelni kell arra, hogy valamennyi átlóban lévő elem különbözzön nullától. Ha csak egy esetben nulla a gyakoriság, a mutatószám akkor is determinisztikus kapcsolatot jelez, ha az egyébként nem is áll fenn.

Kettőnél több ismérvváltozat esetén más mérőszámot kell alkalmazni. A **Cramer-együttható** feloldja az alternatív ismérvek dilemmáját és ugyanakkor érzéketlen a kirívó (egyik cellában nulla értékkel bíró) esetekkel szemben. Az alapgondolat azt vizsgálja, miként alakulnának az egyes gyakoriságok, ha az ismérvek között nem lenne kapcsolat, tehát függetlenek lennének, vagyis valamely ismérvérték nem vonzaná a másik ismérv valamely adott értékét. A kiinduláshoz itt is a kontingencia táblát használjuk.

A számítás alapgondolata: amennyiben a független viszonyt feltételező gyakoriságok és a tényleges gyakoriságok között eltéréseket találunk, akkor a sztochasztikus kapcsolat meglétére gondolhatunk. A feltételezett gyakoriságok számítása tehát azt jelenti, hogy a sokaságot a peremeloszlások alapján osztjuk szét. Ha a táblát a feltételezett gyakoriságokkal töltjük ki, minden sor megoszlása ugyanolyan lesz, ami megfelel a két ismérv függetlenségének. A 2x2-es kontingencia táblát használva, a peremgyakoriságok segítségével meghatározhatjuk a függetlenség esetén feltételezett gyakoriságokat, amelyeket *-gal különböztetünk meg:

$$\frac{S_1 \times O_1}{n} = f^*_{11} \qquad \frac{S_1 \times O_2}{n} = f^*_{12}$$
$$\frac{S_2 \times O_1}{n} = f^*_{21} \qquad \frac{S_2 \times O_2}{n} = f^*_{22}$$

Elsőként minden cellában kiszámítjuk az alábbi relatív különbséget:

$$\frac{(f_{ij} - f^*_{ij})^2}{f^*_{ij}}$$

ahol: az f_{ij} az i-edik sorának és j-edik oszlopának gyakorisága.

A sztochasztikus kapcsolat létezését jelzi tehát az, ha a ténylegesen megfigyelt és a függetlenség esetére feltételezett gyakoriságok nem egyeznek meg. A kétféle gyakoriság eltérése közötti különbségeket egy mérőszámban kell kifejezni, amit a négyzetes kontingencia mutatójával az ún. χ^2-értékkel végezhetünk el.

$$\chi^2 = \sum\sum \frac{(f_{ij} - f^*_{ij})^2}{f^*_{ij}}$$

A χ^2 önmagában még nem felel meg a sztochasztikus kapcsolatok mérőszámaival szemben megfogalmazott feltételnek. Alsó határa ugyan 0, de felső határa jelentősen meghaladhatja az 1-et. Ezt a dilemmát oldja fel a Cramer-féle mutatószám, amelynek képlete:

$$C = \sqrt{\frac{\chi^2}{n \times (s-1)}}$$

Ahol a tört nevezőjében az **s** a változók ismérvváltozatainak minimumát (a kevesebb ismérvváltozat számát) jelöli.

A következőkben a korábban -a táblázatok készítésekor- bemutatott példán keresztül kívánjuk illusztrálni a Cramer- együttható számszerűsítését. (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)

2/16. táblázat: A kontigencia táblázat alapadatai:

Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0	22	1	5	28
lány	2	22	2	3	29
Összesen	2	44	3	8	57

Forrás: saját felmérés

A peremgyakoriságok segítségével a függetlenség esetén feltételezett gyakoriságok:

$$\frac{28 \times 2}{57} = 0{,}98 \qquad \frac{28 \times 44}{57} = 21{,}61 \qquad \text{stb.}$$

A függetlenség esetén feltételezett gyakoriságokat a következő táblázatban láthatjuk:

2/17. táblázat: A függetlenség esetén feltételezett gyakoriságok

Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0,98	21,61	1,47	3,93	28
lány	1,02	22,39	1,53	4,07	29
Összesen	2	44	3	8	57

Elsőként minden cellában kiszámítjuk a relatív különbségeket:

$$\frac{(0-0{,}98)^2}{0{,}98} = 0{,}98 \text{ stb.}$$

Az így létrejött újabb munkatáblát láthatjuk:

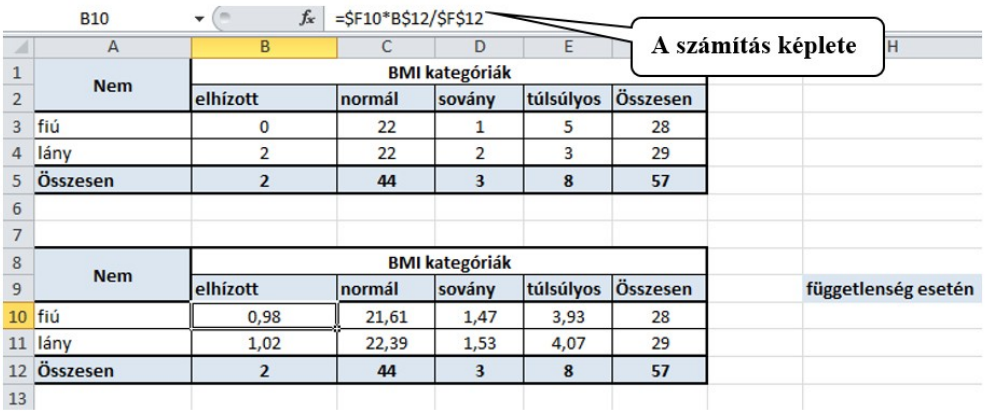
2/18. táblázat: A χ^2-érték számítása

Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0,98	0,01	0,15	0,29	28
lány	0,95	0,01	0,15	0,28	29
Összesen	2	44	3	8	57

A Cramer-féle mutatószám példánkban:

$$C = \sqrt{\frac{2{,}82}{57 \times (2-1)}} = 0{,}22$$

A mérőszám szerint a nem és a BMI kategóriák típusa közötti sztochasztikus kapcsolat gyengének mondható. A C^2 mérőszám is értelmezhető, amely azt mutatja meg, hogy – esetünkben – a nem típusa mintegy 4,94 %-ban determinálja a BMI kategóriák típusát.
Az Excel program segítségével - meglévő kontingencia táblából - a következő lépések során jutunk el a Cramer-féle mutatószámhoz. (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)



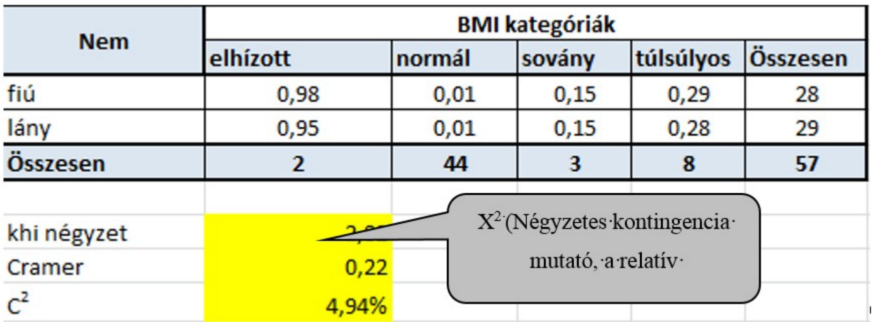
Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0	22	1	5	28
lány	2	22	2	3	29
Összesen	2	44	3	8	57

Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0,98	21,61	1,47	3,93	28
lány	1,02	22,39	1,53	4,07	29
Összesen	2	44	3	8	57

2/31. képernyőnézet: A Cramer mutató számításának menete az Excel programban

Az előző számítás képleteit felhasználva a megfigyelt gyakoriságok táblázatából készítjük el a fiktív (függetlenséget feltételező) gyakorisági táblánkat. Csak az első gyakoriságot határozzuk meg képlettel, a többinél másolni fogjuk a képletet. Ez a megoldás egyben az abszolút és relatív hivatkozások gyakorlására is kitűnő lehetőséget ad. A peremgyakoriságokat felhasználva kapjuk meg a fiktív gyakoriságainkat, mely segítségével láthatóvá válik a jelenségben szereplő ismérvek típusa . A peremgyakoriságokat az összesen sorokban találjuk ezért, az „oszlop- abszolút” és a „sor –abszolút” hivatkozásokat kell a képletben szerepeltetni. A képletben szereplő elemszám viszont „abszolút- abszolút” hivatkozással szerepel. Az így keletkező képlet a szerkesztőlécben látható.

Láthatjuk, hogy az így számított fiktív gyakoriságok nem egyeznek meg (bár nagyságrendileg hasonlóak) a megfigyelt (valós) gyakoriságokkal (nincs függetlenség és nincsen determinisztikus kapcsolat sem), tehát valamilyen mértékű sztochasztikus kapcsolattal van dolgunk. Végül a tényleges és fiktív gyakoriságokat vetjük össze, melynek végeredményeként a négyzetes kontingencia mutató kapjuk, melynek segítségével könnyen számítható a Cramer- együttható.

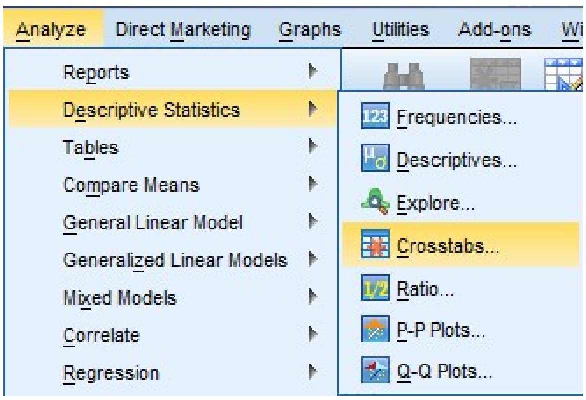


Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0,98	0,01	0,15	0,29	28
lány	0,95	0,01	0,15	0,28	29
Összesen	2	44	3	8	57
khi négyzet	[illegible]				
Cramer	0,22				
C^2	4,94%				

2/32. képernyőnézet: A négyzetes kontingencia mutató számításának menete az Excel programmal

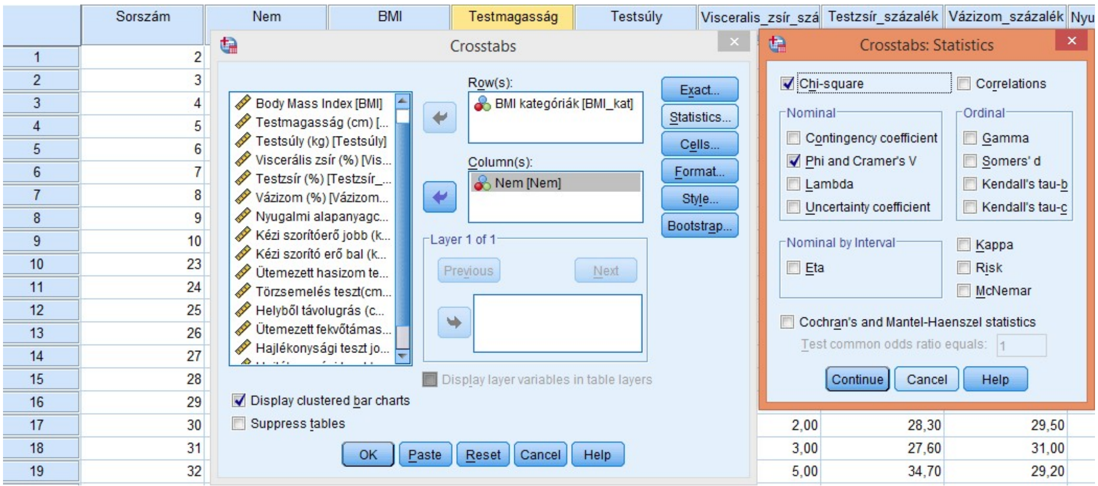
Az SPSS program segítségével az asszociációs kapcsolat elemzése a kereszttábla (Crosstabs) modul segítségével készíthető, mely az Analyze főmenü Decriptive Statistics almenü részéből érhető el. (forrás: fittségi 57fő_adatbázis_alap_bmikat.sav)

Az SPSS program segítségével az asszociációs kapcsolat elemzése a kereszttábla (CROSSTABS) modul segítségével készíthető, mely az ANALYZE főmenü DECRIPTIVE STATISTICS almenü részéből érhető el.



2/33. képernyőnézet: A kereszttábla készítésének elérési útvonala

Ezt követően a változók beállításainak ablaka következik. Első lépésben a vizsgált változók kijelölése következik, át kell mozgatni a sor (ROW(S)) és oszlop (COLUMN(S)) ablakba. Arra vonatkozóan kötelező szabályok nem léteznek, hogy melyik változónk legyen a sor és melyik az oszlop, ez a mindenkori kutató szakmai kompetenciájára van bízva. Segítségként és javaslatként megemlíthető, hogy a gyakorlati társadalomtudományokban az oszlopváltozó (COLUMN(S)) a függő (amelynek eloszlására kíváncsiak vagyunk), míg sorváltozó (ROW(S)) a független (úgy véljük, hogy meghatározó szerepet játszik a függő változóra változó.



2/34. képernyőnézet: Az asszociációs kapcsolat (kereszttábla) beállításai az SPSS programban

Az első lépésben a sor (row) és az oszlop (column) változókat kell megadni. Ezután a Statistics menüpont alatt számos lehetőséggel találkozunk, hiszen itt a különböző skálatípusoknál alkalmazható mérőszámok is szerepelnek. Itt jelen esetben elégséges lesz Chi- square-t (khi négyzetet) és a Phi and Cramer’s V mutatót kérni. A többi statisztikai mutatószámról egy rövid összegzést közlünk, mely segítheti a felhasználókat.

- <u>Kontingencia- együttható</u> (CONTINGENCY COEFFICIENT): a mutatószám alkalmazható bármekkora változószámú kereszttábla esetén, akár a speciális 2X2-es esetében is, azonban értelmezési nehézségek miatt ritkán használjuk, helyette a Cramer együttható (CRAMER V) ajánlott. A számításnál a mintanagyságot használja fel. $C = \sqrt{\dfrac{\chi^2}{\chi^2 + N}}$

- Phi (PHI) mutatószám (Φ): a kereszttáblák speciális esetében a 2X2-es tábláknál használjuk, mivel itt könnyen értelmezhető, hiszen ilyenkor a felső korlát maximális értéke 1. A számítás során a khi négyzet értékét a mintanagysággal korrigáljuk. Nem javasolt több ismérvváltozat során alkalmazni, mivel ilyenkor nincsen felső korlátja, így értelmezése nem egyszerű. $\phi = \sqrt{\frac{\chi^2}{N}}$. Amennyiben azt a mutatószámot használjuk, akkor a folytonossági Yates korrekciós együttható (CONTINUITY CORRECTION) nem alkalmazható, ami a khi- négyzet mutatószám 2X2-es táblánál alkalmazott korrekciója.
- Cramer V együttható (CRAMER's V): a leggyakrabban alkalmazott mérőszám, könnyen értelmezhető. Leggyakrabban a kettőnél több ismérvváltozattal rendelkező változók esetén alkalmazandó. $V = \sqrt{\frac{\chi^2}{n \times (s-1)}}$. Fontos megemlíteni, hogy vannak statisztikusok, akik azt tartják, hogy a khi- négyzet alapú mutatószámok – így a Cramer együttható is- nem alkalmazható abban az esetben, ha az ismérvváltozók közül (cellák), több, mint 20%-ban az érték 5 alatti.
- Lambda (LAMBDA): aszimmetrikus mutatószám. A mutató értelmezése során megkapjuk azt a százalékos értéket, amely megmutatja, hogy a független változó mekkora mértékben képes a függő változót előre jelezni. A számított érték megmutatja, hogy hány százalékkal csökken az előrejelzési hiba, amennyiben az előrejelzés során a feltételezhető okot képező független változót is bevonjuk. $\lambda = \frac{SUM(f_i - f_d)}{N - f_d}$.
- A lambdához hasonlóan értelmezzük a Goodman és Kruskal tau, valamint a bizonytalansági mutatókat. A maximális értéke 1 azt jelenti, független változó ismérvértékeinek ismeretében 100 %-ban (hibátlanul) becsülhető a függő változó értéke.

A modul jobb oldalában található négyzetben az **ordinális skálán** mért változók esetében használt leggyakoribb mutatószámok kerültek összegyűjtésre.

2/19. táblázat

Mutatószám (ordinális skálák esetén)	**Táblatípusok**
Gamma	Bármely táblatípus és méret esetén (könnyen értelmezhető)
Sommers'd	Bármely táblatípus és méret esetén (értelmezése nehézkes)
Kendall tau-b	Szimmetrikus táblák esetén
Kendall tau-c	Nem szimmetrikus táblák esetén

A változók sorrendjei között keressük az összefüggéseket, hiszen itt az egyes ismérv-kategóriák sorrendje jelentőséggel bír, így itt a kapcsolatok szorosságán kívül az irány is jelentőséggel bír. Pozitív az előjel abban az esetben, ha az egyik változó növekvő értéke növekvő változást idéz elő a másik változóban is. Amennyiben az egyik változó növekedése a másik változóban csökkenést vált ki, negatív kapcsolatról beszélünk. Összességében elmondható, hogy az ordinális skálán mért változók esetében az asszociációs mérőszámok esetpárok összehasonlítását célozza. Abban az esetben, amikor egy esetpárban az egyik eset minden változója magasabb, mint a másiké összhangban lévő, egyező (konkordáns) párról beszélünk. Ha azonos értékek szerepelnek az esetpárokban kötöt (tied) esetpárnak nevezzük. Eltérő, vagy széthúzó (diszkordáns) esetpárról abban az esetben beszélünk, ha az egyik változó értéke nagyobb, a másik változó értéke kisebb, mint a pár másik tagjáé. A számítás menete a konkordáns és diszkordáns párok különbségére épül, amelyek a különböző párok között számszerűsítünk. A pozitív asszociációs kapcsolat feltételezi, hogy párok többsége egyező (konkordáns), míg negatív kapcsolat esetében a párok többsége széthúzó (diszkordáns).

- <u>Gamma (GAMMA) együttható[24]</u>: bármely ordinális adatok és táblák összefüggéseinek vizsgálatánál alkalmazható és viszonylag könnyen értelmezhető. Az értéke -1 és 1 közé esik, ahol a 0 érték a változók függetlenségét mutatja. Ez jelenti, hogy mekkora valószínűséggel találjuk a konkordenciát, vagy a

[24] Gyakran szerepel a mutatószám Goodman és Kruskal féle Gamma néven is.

diszkordenciát meghatározónak a kutatott jelenségben. $\gamma = \frac{S - D}{S + D}$. Ha egy gyakorlati példánál a konkordencia az uralkodó jelleggel (γ >0), az azt jelenti, hogy az egyik változó magasabb kategóriája a másik változó magasabb kategóriájának együtt járását váltja ki (pl. a szülők iskolai végzettségének vizsgálatakor a pozitív értékét jelentheti, hogy az apa magasabb iskolai végzettsége az anya magasabb iskolai végzettségét eredményezi, tehát magasabb végzettségű párt választ).

- Somers féle d együttható: az ordinális változók közötti kapcsolatot -1 és +1 értéktartományok között méri, bármely tábla esetében, hasonlóan a gamma mutatóhoz. Az abszolút értékben 1-hez közeli érték szoros kapcsolatot jelent, azonban értelmezése a gamma mutatóhoz képest komplikáltabb.
- Kendall tau-b (KENDALL'S TAU-B): Szimmetrikus táblák, változók között használandó. A mutató -1 és + 1 között értékeket vesz fel, ahol a +1 jelenti, hogy a párok sorrendje hasonló, megegyezik (konkordáns), míg a -1-es értéknél a párok sorrendje ellentétes (diszkordáns).
- Kendall tau-c (KENDALL'S TAU-C): Aszimmetrikus tábláknál használandó, értelmezése megegyezik a Kendall tau-b mutatószáméhoz.
- Kappa (KAPPA)[25]egy egyetértési mutatószám, mely az értékek (értékelők) egyetértését hivatott mérni. A mérőszám értelmezését Landis- Koch (1977) nyomán a következő:

2/20. táblázat

K	Értelmezés
<0	Szegényes egyetértés
0-0,2	Enyhe egyetértés
0,21-0,4	Ígéretes egyetértés
0,41-0,6	Mérsékelt egyetértés
0,61-0,8	Megalapozott egyetértés
0,81-1	Csaknem tökéletes egyetértés

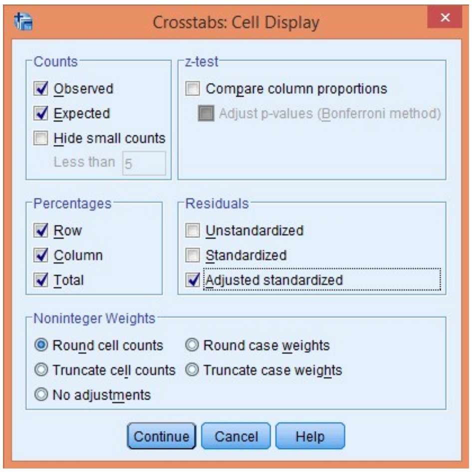
Forrás: Landis és Koch nyomán Jánosa

[25] Gyakran szerepel a mutatószám Cohan- féle kappa néven is.

Amennyiben a túlzó kategória számot összevonva kívánjuk értelmezni, akkor azt is mondhatjuk, hogy 0,4 alatt gyenge egyetértésről, 0,4-0,8 között elfogadható egyetértésről, míg 0,8 felett kiváló egyetértésről beszélhetünk. Szimmetrikus tábláknál és az értékelők ugyanazon skálán mért véleményei során használható.

- Kockázati hányados (RISK) 2X2-es tábláknál alkalmazott kapcsolat-szorossági mutatószám. A számított eredmény értelmezésekor a mutatószám mellett (0 alsó határral és felső határ nélkül) egy konfidencia intervallumot is ad eredményül. Ha a mutató eredménye 1 és ezt az értéket a konfidencia intervallum is tartalmazza, akkor nem beszélhetünk összefüggésről. Amennyiben 0 vagy 1-nél nagyobb akkor feltételezzük a változók közötti kapcsolatot. Összegezve a vizsgálat a relatív kockázatot és esélyhányadost számít 2X2-es tábláknál dichotóm változók esetében. A vizsgálat során az egyik ismérv okként, a másik egy eseményként értelmezhető.
- McNemar teszt (McNEMAR) egy dichotóm változók közti kapcsolatot vizsgáló mutatószám, amely ugyanazon vizsgálati csoporton végzett kutatási mérés közötti változást hívatott vizsgálni. Azt méri, hogy a válaszadók hány százaléka választotta ugyanazon opciót a két mérés alkalmával. Gyakorlatban vélemények összehasonlítását kívánja mérni két időszakban (pl: fogyasztói vélemények, választások stb).
- Cochran és Manthel- Haenzszel féle mutatószám (COCHRAN AND MANTEL-HANSSZEL STATISTICS) két dichotóm változó összefüggését vizsgálja kontrolváltozók együttes hatását feltételezve. A mutatószám előnye, hogy egyszerre veszi figyelembe az összes kontrollváltozó hatását.

A beállításokat elvégezve nyomjuk OK gombot és válasszuk a Cells almenüt.



2/35. képernyőnézet: Az asszociációs kapcsolat (kereszttábla) beállításai az SPSS programban (Cells)

A bal felső nézetben az adatokra vonatkozó beállításokat láthatjuk, ahol az OBSERVED az egyedi (valós) megfigyelt értékeket, míg az EXPECTED a fiktív, függetlenség esetén feltételezett gyakoriságokat jelenti.

Az alatta lévő dobozban a százalékos arányszámokat kérhetjük (sorszázalék= ROW; oszlopszázalék=COLUMN; teljes százalék=TOTAL).

A sorszázalék azt jelenti, hogy a cellában tartalmazott gyakoriság a sor gyakoriság hány %-a. Az oszlopszázalék jelenti, hogy a cellában tartalmazott megfigyelt gyakoriság az oszlop gyakoriság hány százaléka. Az összes gyakoriság a cellagyakoriság „sorösszesen”-„oszlopösszesen” és mintanagyság hányadosa.

A reziduális értékek (RESIDUALS) dobozában található mutatók a tényleges és a függetlenség esetén feltételezett gyakoriságok különbségeiből kerülnek meghatározásra. Amennyiben az érték negatív, akkor a tényleges gyakoriság kisebb, mint a függetlenség esetén indokolt lenne. A három mutató közül az ADJUSTED RESIDUAL érték talán a leghasznosabb, hiszen a kapcsolat meglétén túl, megjeleníti a kategóriák (ismérvváltozók) közül azokat, amelyek a okozzák a kapcsolatokat. Amennyiben a számított érték abszolút értéke 2-nél nagyobb, akkor a két ismérvváltozat között szignifikáns kapcsolat van. Ha a számított érték abszolút értéke kettőnél kisebb, akkor a két kategória között nincsen szignifikáns kapcsolat.

Az output nézetben az elemzések elején egy összesítést (case processing summary) kapunk arról, hogy mekkora volt a megvizsgált esetek száma (N), hány valós (valid) és hány hiányzó értékünk (missing) volt és mindezek relatív megoszlása (percent) miként alakult.

2/21. táblázat

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Nem * BMI kategóriák	57	100,0%	0	0,0%	57	100,0%

Látható, hogy az elemszám 57 és nincsen hiányzó adat, vagyis a 100%-ot tudjuk értékelni.

A következő táblázat, a tényleges kereszttáblát tartalmazza, amelynek segítségével könnyen összehasonlíthatjuk a tényleges és a függetlenség esetén feltételezett (fiktív) gyakoriságokat. Itt is látható, hogy az értékek ugyan nem egyeznek meg, de nagy eltéréseket nem tartalmaz, ami sztochasztikus kapcsolat hiányára utal. A számított táblázatban szereplő értékek megegyeznek a fent számított értékeinkkel. A táblázatban a következő adatok szerepelnek:

- A megfigyelések valós érték, gyakorisága (COUNT)
- A függetlenség esetén feltételezett (fiktív) gyakoriságok (EXPECTED COUNT)
- A sormegoszlás (% WITHIN BMI KATEGÓRIÁK)
- Az oszlopmegoszlás (% WITHIN Nem)
- A teljes mintanagyság szerinti megoszlás (% OF TOTAL)
- A standardizált, korrigált reziuduum (ADJUSTED RESIDUAL)

A következő táblában számunkra ismert érték tűnik fel a Pearson Chi-Square mutató mellett, hiszen ez megegyezik az általunk számolt négyzetes kontingencia mutató értékével.

2/22. táblázat: A számított khi-négyzet értéket tartalmazó táblázat

BMI kategóriák * Nem Crosstabulation

			Nem		Total
			fiú	lány	
BMI kategóriák	sovány	Count	1	2	3
		Expected Count	1,5	1,5	[illegible]
		% within BMI kategóriák	33,3%	66,7%	[illegible]
		% within Nem	3,6%	6,9%	[illegible]
		% of Total	1,8%	3,5%	[illegible]
		Adjusted Residual	-,6	,6	
	normál	Count	22	22	44
		Expected Count	21,6	22,4	44,0
		% within BMI kategóriák	50,0%	50,0%	100,0%
		% within Nem	78,6%	75,9%	77,2%
		% of Total	38,6%	38,6%	77,2%
		Adjusted Residual	,2	-,2	
	[illegible]	Count	5	3	8
		Expected Count	3,9	4,1	8,0
		% within BMI kategóriák	62,5%	37,5%	100,0%
		% within Nem	17,9%	10,3%	14,0%
		% of Total	8,8%	5,3%	14,0%
		Adjusted Residual	,8	-,8	
	elhízott	Count	0	2	2
		Expected Count	1,0	1,0	2,0
		[illegible]	0,0%	100,0%	100,0%
		[illegible]	0,0%	6,9%	3,5%
		[illegible]	0,0%	3,5%	3,5%
		[illegible]	-1,4	1,4	
Total		Count	28	29	57
		Expected Count	28,0	29,0	57,0
		% within BMI kategóriák	49,1%	50,9%	100,0%
		% within Nem	100,0%	100,0%	100,0%
		% of Total	49,1%	50,9%	100,0%

50% (22/44) a fiú aránya az összes normál BMI-vel

78,6 % (22/28) az összes fiú közül, aki normál BMI kategóriába tartozik

A mintába került összes egyed közül 38,6 % (22/57) aki fiúés normál BMI testalkat kategóriával

Ha mutató +2 feletti értéket vesz fel, akkor ott biztosan van szignifikáns összefüggés .ha -2 alattit

2/23. táblázat

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,817[a]	3	,421
Likelihood Ratio	3,600	3	,308
Linear-by-Linear Association	,040	1	,842
N of Valid Cases	57		

a. 6 cells (75,0%) have expected count less than 5. The minimum expected count is ,98.

A df a szabadságfokot jelöli, melynek számítása df=(sor-1)*(oszlop-1). A szabadságfoknak fontos szerepe van az elméleti érték megállapításakor. Hiszen a tapasztalati érték a négyzetes kontingencia mutató (χ^2) és az elméleti érték összevetése után tudjuk megállapítani, hogy a nullhipotézist elfogadjuk vagy elvetjük.[26] A χ^2 eloszlástáblázat a szabadságfok (3) és egy adott hibavalószínűség (0,05) mellett megadja viszonyítási értéket, mely esetünkben 7,81 (Excelben: =inverz.khi (0,05;3). Mivel a tapasztalati érték alacsonyabb, mint az elméleti, ezért a nullhipotézist elfogadjuk, vagyis nincsen kapcsolat a két ismérv között. Ez azt jelenti, hogy a hallgatók neme nincsen összefüggésben a BMI kategeriákkal, vagyis nem a nem határozza meg a BMI kategóriákat. Ez a szignifikanciára vonatkozó táblázatból (DIRECTIONAL MEASURES, SYMMETRCIC MEASURES) értékeiből is leolvasható, hiszen nagyobb, mint az általunk választott 5 százalék.

2/24. táblázat: A számított kapcsolat-szorossági mérőszámok

Symmetric Measures

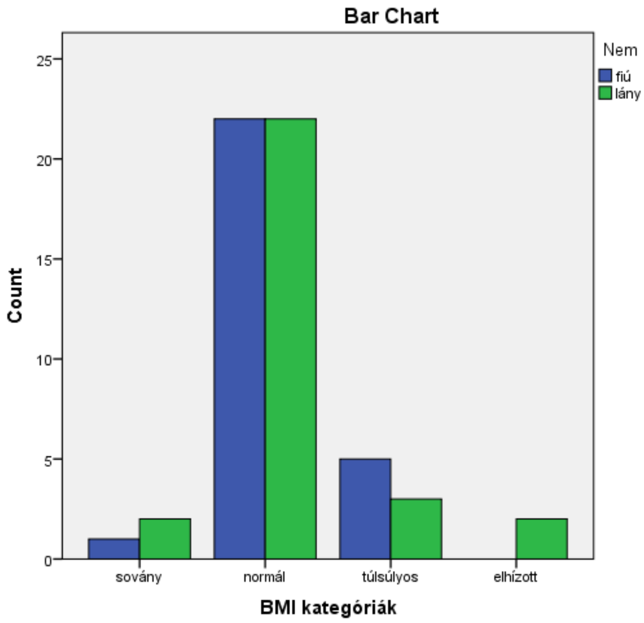
		Value	Approx. Sig.
Nominal by Nominal	Phi	,222	,421
	Cramer's V	,222	,421
N of Valid Cases		57	

A Cramer- együttható értékéből leolvasható, hogy gyenge kapcsolat van a két ismérv között. A Phi mutatót általában alternatív ismérvek esetén érdemes alkalmazni, mivel értelmezésnél csak ekkor lesz a felső korlát 1, különben nincsen korlát és ilyenkor

[26] Részletesen a későbbiekben.

problémás az értelmezése. A Phi együttható értéke a khi-négyzetnek a mintanagysággal (N) korrigált értéke: $\phi = \sqrt{\frac{\chi^2}{N}}$

A kereszttáblák készítésénél felkínálja a program a grafikus ábrázolást (Display clustered bar charts) lehetőségét is. Alaplehetőségként oszlopdiagramon keresztül szemlélteti az asszociációs kapcsolatot.



2/20. ábra

Fontos ismét nyomatékosítani, a fenti példa a szemléltetés kedvéért került bemutatásra, mivel tudva lévő, hogy nem szerencsés a khi- négyzet alapú mutatók alkalmazása abban az esetben, ha az ismérvváltozók közül (cellák), több mint 20%-ban az érték 5 alatti.

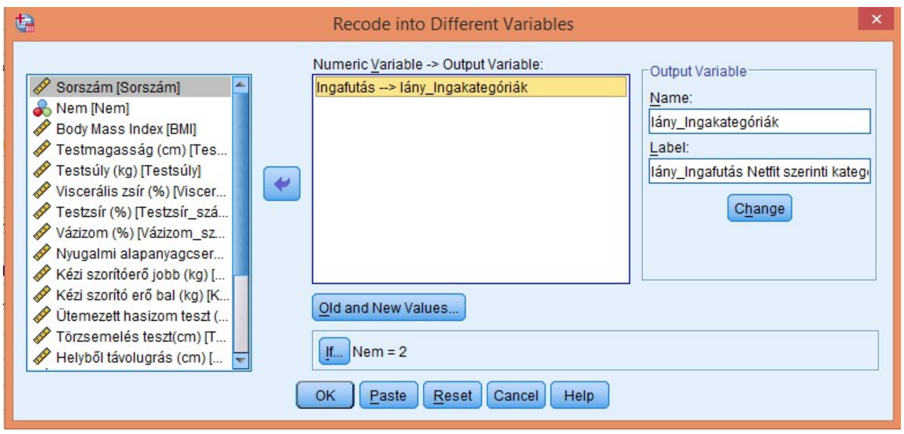
Az asszociációs kapcsolatot grafikus szemléltetésére alkalmazzuk jelen esetben a **korrespondencia-analízist**. „A korrespondencia-analízis lehetővé teszi, két nominális változó kapcsolatának grafikus megjelenítését egy többdimenziós, de a szemléletesség és a könnyű értelmezhetőség kedvéért kis dimenziószámú térben (általában síkban). Az egymáshoz hasonló kategóriák ezekben az ábrázolásokban is közel kerülnek egymáshoz. Az eredmények értelmezése az alkalmazott normalizáló eljárástól függ. A SPSS-ben az alapértelmezett normalizálás a sor- és az oszlopváltozók kapcsolatát elemzi.” (Ketskeméty-Izsó 2005, 417.o.)

Vizsgáljuk meg, hogy a nők esetében az állóképességi ingafutás kategóriái és a BMI kategóriák között létezik-e összefüggés, és az esetlegesen összetartozó kategóriákat grafikus módon is jelenítsük meg (forrás: fittségi 57fő_adatbázis_alap_bmikat.sav). A kategóriák kalibrálása a NEFTIT kézikönyv segítésével történt.

2/25. táblázat

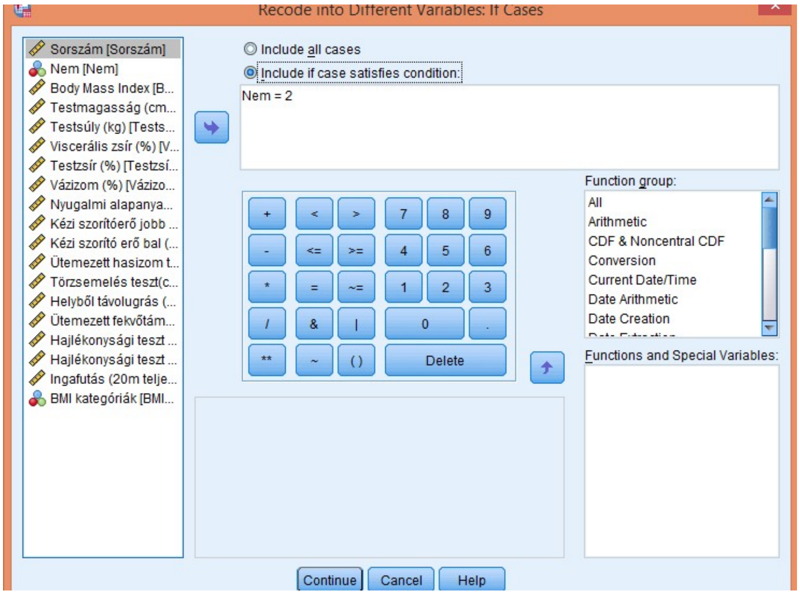
BMI					20 méteres állóképességi ingafutás			
	Lányok					Lányok		
Életkor	*sovány*	*egészség zóna*	*fejlesztés szükséges*	*fokozott fejlesztés szükséges*	*Életkor*	*Egészség zona*	*fejlesztés szükséges*	*fokozott fejlesztés szükséges*
18-	<18,5	18,6-24,9	25,0-29,9	30,0<	18-	38<	29-37	<28

Annak érdekében, hogy a korrespondencia-analízist el tudjuk végezni, fontos a meglévő folytonos változókból a kategóriás változókat elkészíteni. Miután az adatbázisunk tartalmazza a BMI kategóriákat, így elegendő meghatározni a 18 év feletti hallgató lányok 20 méteres állóképességi ingafutás értékeiből a kategóriákat. Ezt megtehetjük a TRANSFORM menü RECODE INTO DIFFERENT VARIABLES almenü segítségével. Ezt követően egy meglévő változó (Ingafutás) felhasználásával, hozzunk létre egy új változót lány_ingakategóriák néven. Ez azt jelenti, hogy az Ingafutás változót a bal változókat tartalmazó dobozból a középen található nyíl segítségével mozgassuk a középső (Numeric Variable) dobozba. Ezt követően nevezzük el az új változót (lány_ingakategóriák) és címkézzük is fel (lány_ingafutás Netfit kategóriák). Ha az új változót elneveztük minden esetben a CHANGE megnyomásával tudjuk a programmal elfogadtatni. Miután megnyomtuk a CHANGE opciót, a középső dobozban a régi változó neve mellett megjelenik az új változónk neve is.



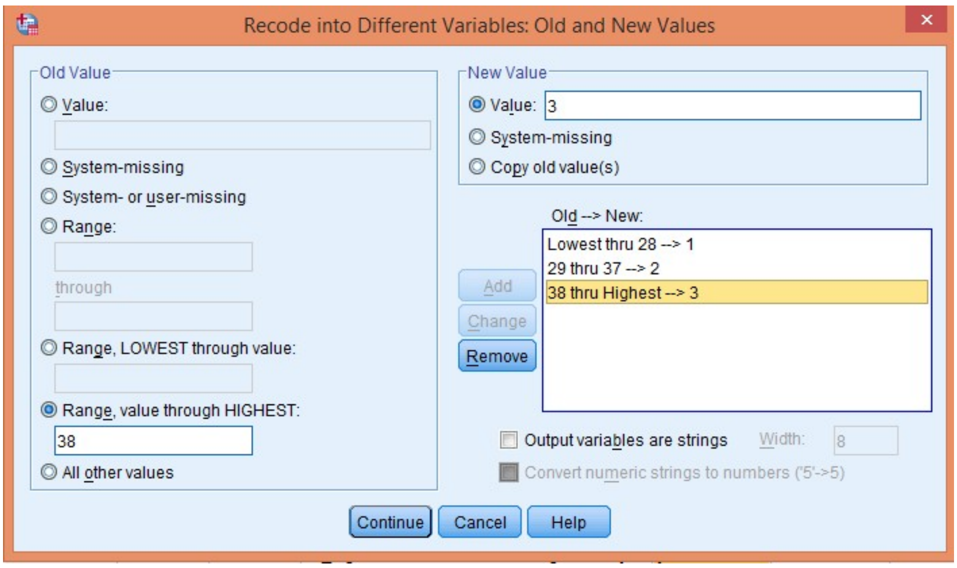
2/35. képernyőnézet: a női ingafutás változó átkódolása

Ezt követően állítsuk be a szűrőfeltételt is, hogy a kategóriák meghatározása csak a nőkre fog vonatkozni. A szűrőfeltétel beállítását az IF doboz megnyomásával tehetjük meg.



2/36. képernyőnézet: Szűrőfeltétel megadása

Jelöljük ki az Include if case satisfies condition és a Nem változót mozgassuk a mozépső dobozba. Itt jelöljük, hogy a nem változó ismérvváltozatai közül csak a nőket (2) vegye figyelembe. Ezt követően nyomjuk a Continue dobozt. Ezt követően tudjuk az értéktartományokat kijelölni, melyet az Old and New Values opcióval tehetünk meg.



2/37. képernyőnézet: Az értéktartományok jelölése

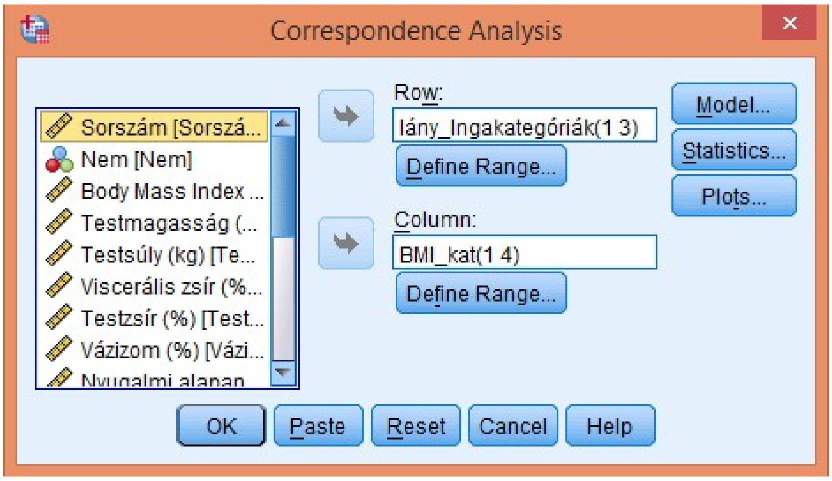
A felugró ablakot két részre lehet osztani, hiszen a baloldalon a kiinduló, a jobboldalon az új változó értékei láthatóak. Amennyiben a VALUE opciót jelöljük be, akkor egyesével megadhatjuk a régi értékeket. A SYSTEM-MISSING vagy a SYSTEM –OR USER MISSING-opciók segítségével azokat a tételeket tudjuk kizárni, melyek a feltételeinknek nem felelnek meg. A RANGE opció beállításainak segítségével tudjuk a különböző az osztályközök feltételeit megszabni. Az első alternatíva, amikor a RANGE ….THROUGH…….. választásával az osztályközök alsó és felső határát tudjuk az értéktartomány beírásával megadni (pl: 29 és 37 közöttiekből 2-es csoportot) . A RANGE LOWEST THROUGH segítségével az alulról nyitott osztályközt tudjuk megadni, míg a RANGE …..THROUGH HIGHEST opció segítségével a felülről nyitott osztályközt definiálhatjuk (pl: 38 értéktől 3-es kódót kapnak). A jobb oldalon található új értékek megadására, akkor van lehetőségünk, ha a régi értékeket az előzőekben ismertetett módón megadtuk. A régi értékhez tartozó új értéket a VALUE szövegdobozba rögzítjük, majd az ADD gombra nyomva véglegesítjük. Ezt követően a változó kalibrálása megjelenik az OLD→NEW ablakban. Ha a kategóriákat módosítani szeretnénk, akkor a CHANGE, ha törölni akkor a REMOVE gombot válasszuk. Amennyiben az új kategóriák szöveges formátumúak, akkor a OUTPUT VARIABLES ARE STRINGS jelölőnégyzetet kipipáljuk. A beállításokat elvégezve nyomjuk a CONTINUE gombot, majd az OK gomb lenyomását követően a három kategória létrejön.

Miután a három kategória, csak egy numerikus értéket tartalmaz, ezért a kategóriák elnevezését a változó nézetben a VALUE LABELS lehetőségnél megtehetjük.

2/38. képernyőnézet: A létrejött új csoportok elnevezése, címkézése

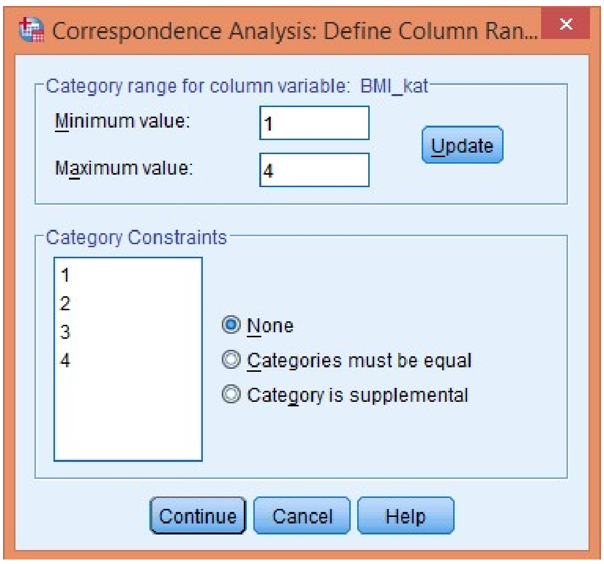
Az adatok korrekciójánál gyakran fordul elő, hogy nem kell egy teljesen új változót létrehozni, elégséges a meglévő változót javítani, változtatni (pl. abban az esetben is ezt az opciót javasoljuk használni, amikor a skála irányultságát kívánjuk megfordítani). Ebben az esetben a TRANSFORM főmenü RECODE INTO SAME VARIABLES-re kattintva tudjuk a szükséges beállításokat elvégezni. Mindezen adat-transzformációkat követően tudjuk a korrespondencia-analízist elvégezni.

Ez az eljárás a DATA REDUCTION főmenü a CORRESPONDENCE ANALYSIS almenüjében érhető el. (forrás: fittségi 57fő_adatbázis_alap_bmikat.sav).



2/39. képernyőnézet: A korrespondenica-analízis beállításai az SPSS programban

Először jelöljük ki a sor- (row) és oszlopváltozókat (column). Ezután minden egyes ismérvet definiálni kell, a benne szereplő ismérvváltozatok számának segítségével.



2/40. képernyőnézet: Az ismérvváltozatok számának kijelölése

Jelen esetben a BMI kategóriákat definiáltuk 1-től 4-ig, az ismérvváltozatok számának függvényében. Miután mindkét ismérvet meghatároztuk, a többi beállításon ne változtassunk és nyomjuk meg az Ok gombot. Az eredményeket a következő táblázatokban láthatjuk.

Az első táblázat (Correspondence Table) a tényleges gyakoriságokat tartalmazza, a másodikban az összesítő eredmények olvashatók. Itt látható, hogy a kapcsolat szignifikáns (p=0,00), illetve a létrejövő két dimenzió alkalmas a megjelenítésre, hiszen az értékek szóródásának 100%-át magyarázza. A következő két táblázat az egyes ismérvváltozatok koordinátáit tartalmazzák az alapbeállításként szereplő két dimenzió mentén. Talán a legszemléletesebb lehet számunkra a grafikus megjelenítés (Biplot), amely segítségével az összetartozó értékek két dimenzió mentén láthatóvá válnak.

2/25. táblázat

Correspondence Table

lányok_Ingafutás Netfit szerinti kategóriái lányokná	BMI kategóriák				
	sovány	normál	túlsúlyos	elhízott	Active Margin
fokozott fejlsztés szükséges	0	1	1	2	4
fejlesztés szükséges	0	2	2	0	4
egészségzóna	2	19	0	0	21
Active Margin	2	22	3	2	29

2/26. táblázat

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	,778	,605			,712	,712	,124	,090
2	,495	,245			,288	1,000	,231	
Total		,850	24,653	,000[a]	1,000	1,000		

a. 6 degrees of freedom

2/27. táblázat

Overview Row Points[a]

lányok_Ingafutás Netfit szerinti kategóriái lányokná	Mass	Score in Dimension		Inertia	Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
		1	2		1	2	1	2	Total
fokozott fejlsztés szükséges	,138	1,961	-,804	,457	,682	,180	,903	,097	1,000
fejlesztés szükséges	,138	,681	1,673	,241	,082	,780	,207	,793	1,000
egészségzóna	,724	-,503	-,166	,152	,236	,040	,936	,064	1,000
Active Total	1,000			,850	1,000	1,000			

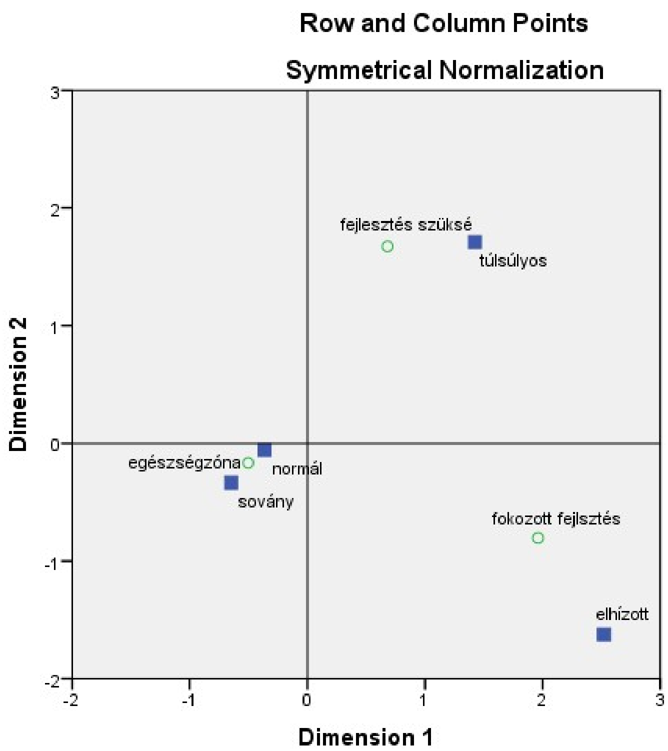
a. Symmetrical normalization

2/28. táblázat

Overview Column Points[a]

BMI kategóriák	Mass	Score in Dimension 1	Score in Dimension 2	Inertia	Contribution: Of Point to Inertia of Dimension 1	Of Point to Inertia of Dimension 2	Of Dimension to Inertia of Point 1	Of Dimension to Inertia of Point 2	Total
sovány	,069	-,647	-,334	,026	,037	,016	,855	,145	1,000
normál	,759	-,365	-,055	,080	,130	,005	,986	,014	1,000
túlsúlyos	,103	1,424	1,712	,313	,270	,612	,521	,479	1,000
elhízott	,069	2,521	-1,624	,431	,564	,367	,791	,209	1,000
Active Total	1,000			,850	1,000	1,000			

a. Symmetrical normalization



2/21. ábra

Az ábra jól szemlélteti az összetartozó ismérvértékeket. Látszik, hogy leginkább az egészségzónába a normál és a sovány BMI kategóriával tartozó hallgatók kerültek, míg a állóképesség fokozott fejlesztése az elhízott BMI kategóriába tartozó egyedeknek javasolt.

A grafikus megjelenítésnél a dimenziók elnevezése kulcsfontosságú, hiszen a megértés segítése a célja, mely a kutató önálló, saját feladata. Az itt látható dimenziókat elnevezését is az olvasóra bízzuk.

2.7.3.2. Vegyes kapcsolat vizsgálata

Korábban már említésre került, hogy a vegyes kapcsolatokban az ok szerepét mindig a minőségi ismérv tölti be, míg az okozat(ok)ét a mennyiségi ismérv(ek). A vegyes kapcsolat vizsgálata azt célozza, hogy megállapítsuk a kvantitatív változóban rejlő információ mekkora része határozható meg a minőségi ismérv szerinti csoportosítással. A mennyiségi ismérv lehetőséget teremt arra, hogy a számítási eljárásokat kibővítsük. A vegyes kapcsolat szorosságának mérése a szórás felbontásának összefüggésén alapul. A szórásnégyzetek közötti összefüggést a szórásnégyzet összetevőkre bontásának nevezünk. E szerint a teljes szórásnégyzet a belső szórásnégyzet és a külső szórásnégyzet összegeként írható fel.

$$\sigma^2 = \sigma_B^2 + \sigma_K^2$$

A belső és külső szórásnégyzet kiszámításának módja:

$$\sigma_B^2 = \frac{\sum_{j=1}^{m} n_j \sigma_j^2}{n}$$

$$\sigma_K^2 = \frac{\sum_{j=1}^{m} n_j (\bar{x}_j - \bar{x})^2}{n}$$

A szórásnégyzetek összefüggését kifejező képletet átrendezhetjük, ha elosztjuk az egyenlőség mindkét oldalát a teljes szórásnégyzettel:

$$1 = \frac{\sigma_B^2}{\sigma^2} + \frac{\sigma_K^2}{\sigma^2}$$

A csoportosító (minőségi) ismérv – ami egyben a sztochasztikus kapcsolat ok szerepét is betölti – hatását a külső szórás közvetíti. Könnyen belátható, hogy amennyiben a külső szórás nulla, a minőségi ismérvnek semmilyen mérhető hatása nincs; a két ismérv (a minőségi és mennyiségi) független. Az ellenkező szélsőséges esetben – amennyiben a belső szórás nulla – a külső szórás megegyezik a teljes szórással, tehát a kapcsolat determinisztikus. Ezek alapján a külső és a teljes szórás segítségével számszerűsíthető a vegyes kapcsolatot mérő **szóráshányados** mutatója:

$$H = \frac{\sigma_K}{\sigma} = \sqrt{\frac{\sigma_K{}^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma^2{}_B}{\sigma^2}}$$

Az eltérés négyzetösszegekre is igaz, hogy a teljes eltérésnégyzet- összeg a külső és a belső eltérésnégyzet-összeg összegeként adódik:

$$SS = SS_B + SS_K = \sum_{j=1}^{m}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2 = \sum_{j=1}^{m} f_j(\bar{x}_j - \bar{x})^2 + \sum_{j=1}^{m}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2$$

Így az is igaz:

$$H^2 = \frac{SS_K}{SS}$$

Nézzünk most egy konkrét példát: A www.nemzetsport.hu (2008-08-02) oldalról szerzett adatok alapján. Arra vagyunk kíváncsiak, hogy a sportágak típusa, mennyire befolyásolja a csapatban nevezett játékosok létszámadatát (mennyi játékost kell alkalmazni), vagyis van- e kapcsolat a sportágak fajtái és a csapatban szereplő játékosok száma között. (Forrás: vegyes kapcsolat.xls)

2/29. táblázat: A vegyes kapcsolat alapadatai

Csapat neve	Sportág neve	Létszám (fő)
Siófok	Labdarúgás	26
Budapest Honvéd	Labdarúgás	22
DVSC	Labdarúgás	35
Paks Fc	Labdarúgás	24
MTK	Labdarúgás	22
Kaposvár	Labdarúgás	21
Pick- Szeged	Kézilabda	18
MKB Veszprém Kc	Kézilabda	16
Komlói KBSK	Kézilabda	17
Dunaferr SE	Kézilabda	18
Albacomp	Kosárlabda	12
Atomerőmű SE	Kosárlabda	12
Falco KC	Kosárlabda	13
PVSK Expo Center	Kosárlabda	14
BVSC	Vízilabda	18
Domino BHSE	Vízilabda	15
PVSK- Füszért	Vízilabda	16

A csoportosított az adatokat a következő munkatáblában láthatjuk:

2/30. táblázat: Csoportosított adatok

Sportág neve	Vizsgált csapatok száma (f_i)	A létszámadatok átlaga	A létszámadatok szórása
Labdarúgás	6	25,00	4,76
Kézilabda	4	17,25	0,83
Kosárlabda	4	12,75	0,83
Vízilabda	3	16,33	1,25
összesen	17		

A csoportonkénti átlagok segítségével (is) kiszámíthatjuk a 17 megfigyelés főátlagát (súlyozott átlagszámítás):

$$\bar{x} = \frac{6*25 + 4*17{,}25 + 4*12{,}75 + 3*16{,}33}{17} = 18{,}76 \text{ fő}$$

A belső szórásnégyzet és a belső szórás, a csoportszórásokat felhasználva az alábbi módon számítandó:

$$\sigma_B^2 = \frac{6\times 4{,}76^2 + 4\times 0{,}83^2 + 4\times 0{,}82^2 + 3\times 1{,}25^2}{17} = 8{,}60$$

$$\sigma_B = \sqrt{8{,}60} = 2{,}93$$

A külső szórásnégyzet és szórás:

$$\sigma_K^2 = \frac{6\times(25-18{,}76)^2 + 4\times(17{,}25-18{,}76)^2 + 4\times(12{,}75-18{,}76)^2 + 3\times(16{,}33-18{,}76)^2}{17} = 23{,}82$$

$$\sigma_K = \sqrt{23{,}82} = 4{,}88$$

A teljes sokaságra vonatkozó szórásnégyzet és szórás (az additív összefüggést felhasználva):

$$\sigma^2 = 8{,}6 + 23{,}82 = 32{,}42$$

$$\sigma = \sqrt{32{,}42} = 5{,}69$$

A létszámok együttes (teljes) szórása: 5,69 fő. A belső szórás megmutatja, a létszámadatok szórása a saját csoportjának (sportágnak) átlagától, átlagosan 2,93 fővel tér el. A külső szórás azt szemlélteti, hogy a sportáganként kialakult átlagos létszámadatok a vizsgálatba bevont összes csapat átlagától (főátlagtól), átlagosan 4, 88 fővel térnek el.

A fenti adatok lehetővé teszik a kapcsolat szorosságának mérését a szóráshányados segítségével.

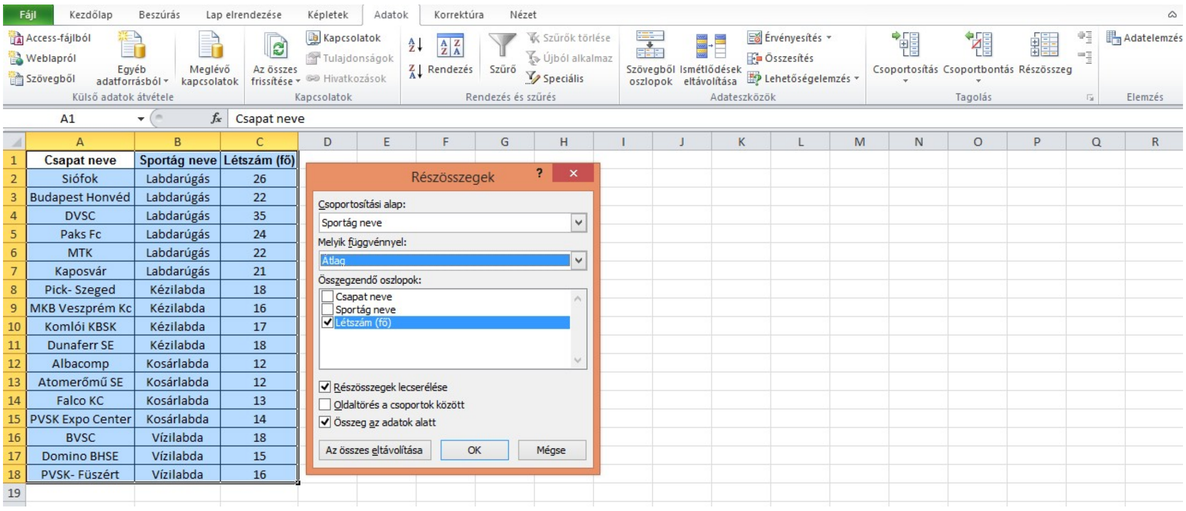
$$H^2 = \frac{23{,}82}{32{,}42} = 0{,}73$$

$$H = \frac{4{,}88}{5{,}69} = \sqrt{\frac{23{,}82}{32{,}42}} = \sqrt{1 - \frac{8{,}6}{32{,}42}} = 0{,}86$$

A szóráshányados mutató segítségével megállapíthatjuk, hogy szoros kapcsolat van a sportágak típusa és a csapatban nevezett (létszám között) játékosok létszámadata között. A szórásnégyzet- hányados megmutatja, hogy a sportágak típusa a létszámadat (játékosok számának) 73 %-át magyarázza.

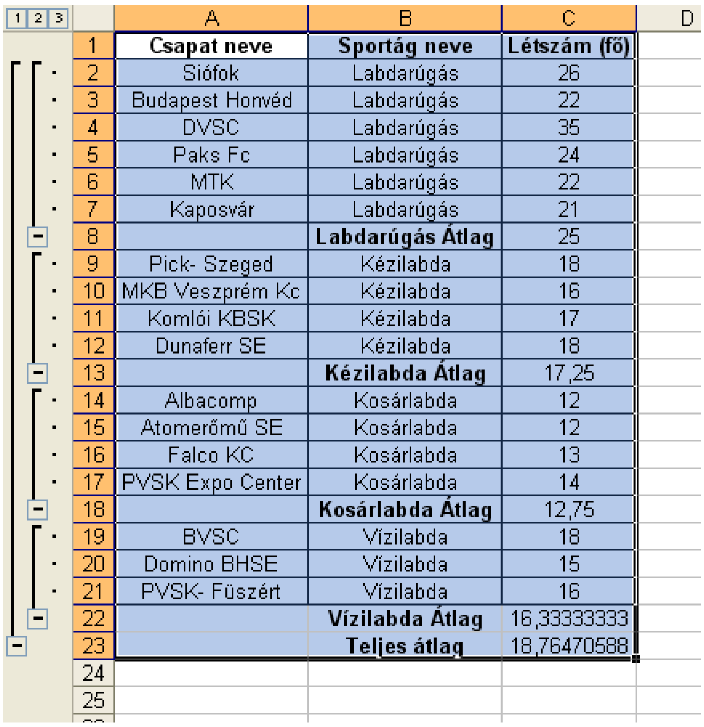
Most nézzük, miként lehet az Excel program segítségével számszerűsíteni a vegyes kapcsolat szorosságát.

Mivel a részátlagokra szükségünk lesz, először azokat kell meghatározni. Ezt a legegyszerűbben az Adatok menüpont, Részösszeg moduljával tehetjük, ahol jelöljük ki a következőket:



2/41. képernyőnézet: A vegyes kapcsolat számítása az Excel program segítségével

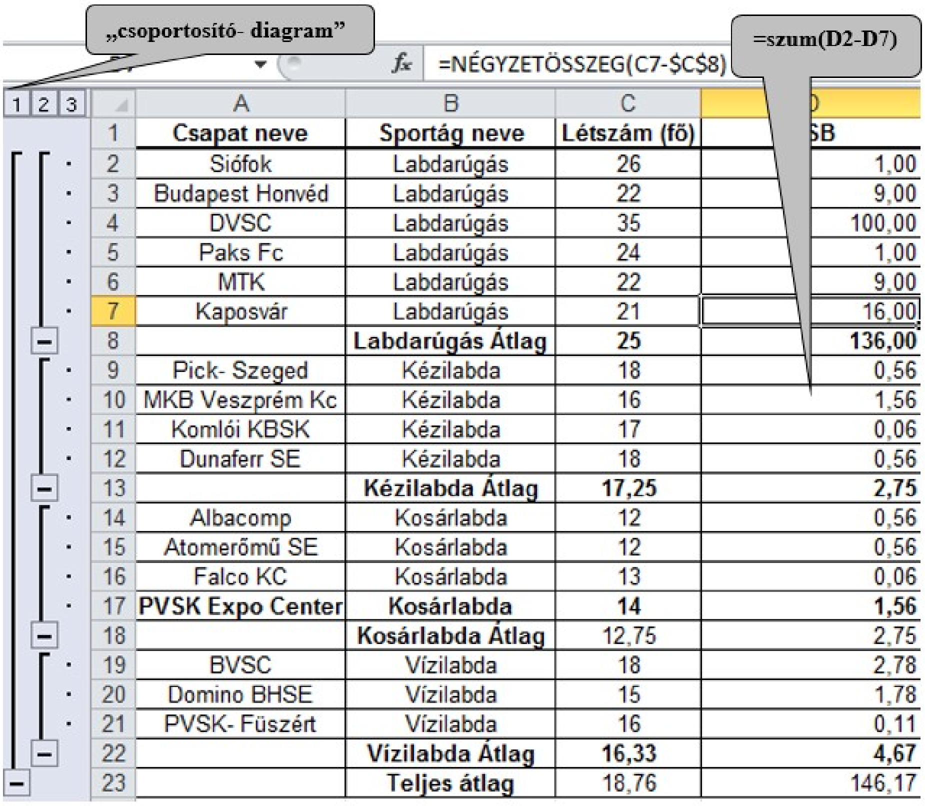
A csoportosítási alap jelen esetben a sportág neve, a jellemző érték az átlag, melyet a létszámadatokra számítunk. Ennek végeredményeként a program minden csoport átlagát kiszámítja, melyet a csoportok után beszúrva fel is tüntet.



	A	B	C
1	Csapat neve	Sportág neve	Létszám (fő)
2	Siófok	Labdarúgás	26
3	Budapest Honvéd	Labdarúgás	22
4	DVSC	Labdarúgás	35
5	Paks Fc	Labdarúgás	24
6	MTK	Labdarúgás	22
7	Kaposvár	Labdarúgás	21
8		**Labdarúgás Átlag**	25
9	Pick- Szeged	Kézilabda	18
10	MKB Veszprém Kc	Kézilabda	16
11	Komlói KBSK	Kézilabda	17
12	Dunaferr SE	Kézilabda	18
13		**Kézilabda Átlag**	17,25
14	Albacomp	Kosárlabda	12
15	Atomerőmű SE	Kosárlabda	12
16	Falco KC	Kosárlabda	13
17	PVSK Expo Center	Kosárlabda	14
18		**Kosárlabda Átlag**	12,75
19	BVSC	Vízilabda	18
20	Domino BHSE	Vízilabda	15
21	PVSK- Füszért	Vízilabda	16
22		**Vízilabda Átlag**	16,33333333
23		**Teljes átlag**	18,76470588
24			
25			

2/42. képernyőnézet: A vegyes kapcsolat résszámításai az Excelben

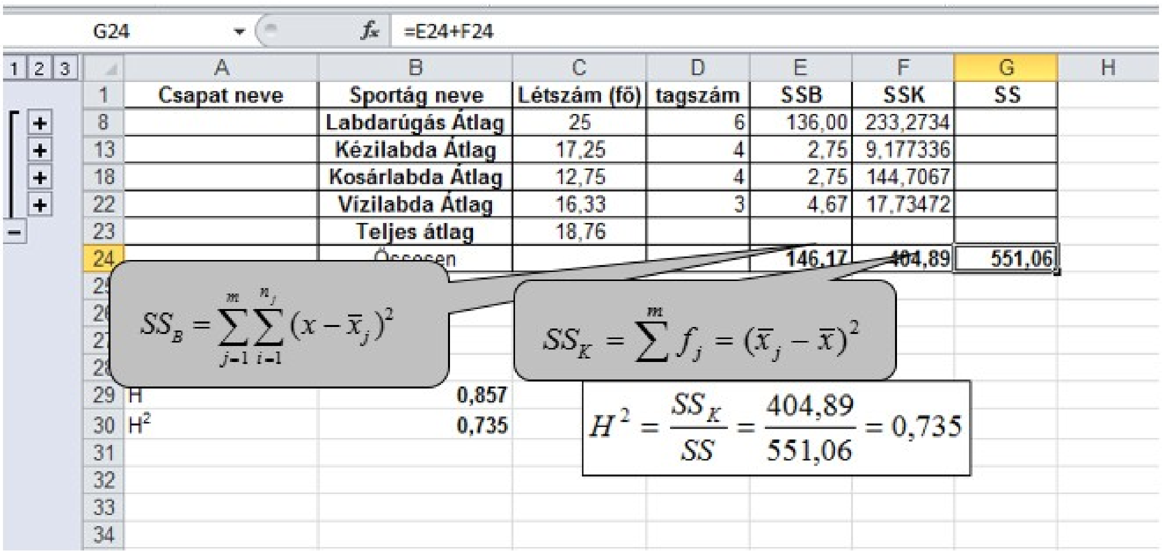
Ezután az eltérésnégyzet-összeget felhasználva számolunk tovább. Meghatározzuk a belső eltérésnégyzet-összeget és a külső eltérésnégyzet-összeget, melyek összegeként a teljes eltérésnégyzet-összeg számítható. Először készítsünk egy új oszlopot, ahol a csoportok tagszámát tüntetjük fel, ezt a darab függvény segítségével is megtehetjük. Az eredményeket a sportágak átlagainak sorában tüntessük fel. Ezután a következő oszlopban már a belső eltérésnégyzet-összegeket számítjuk ki, a négyzetösszeg függvényt felhasználva.



	A	B	C	D
1	Csapat neve	Sportág neve	Létszám (fő)	SB
2	Siófok	Labdarúgás	26	1,00
3	Budapest Honvéd	Labdarúgás	22	9,00
4	DVSC	Labdarúgás	35	100,00
5	Paks Fc	Labdarúgás	24	1,00
6	MTK	Labdarúgás	22	9,00
7	Kaposvár	Labdarúgás	21	16,00
8		**Labdarúgás Átlag**	**25**	**136,00**
9	Pick- Szeged	Kézilabda	18	0,56
10	MKB Veszprém Kc	Kézilabda	16	1,56
11	Komlói KBSK	Kézilabda	17	0,06
12	Dunaferr SE	Kézilabda	18	0,56
13		**Kézilabda Átlag**	**17,25**	**2,75**
14	Albacomp	Kosárlabda	12	0,56
15	Atomerőmű SE	Kosárlabda	12	0,56
16	Falco KC	Kosárlabda	13	0,06
17	**PVSK Expo Center**	**Kosárlabda**	**14**	**1,56**
18		**Kosárlabda Átlag**	12,75	2,75
19	BVSC	Vízilabda	18	2,78
20	Domino BHSE	Vízilabda	15	1,78
21	PVSK- Füszért	Vízilabda	16	0,11
22		**Vízilabda Átlag**	**16,33**	**4,67**
23		**Teljes átlag**	18,76	146,17

2/43. képernyőnézet: A belső eltérésnégyzet- összeg résszámításai

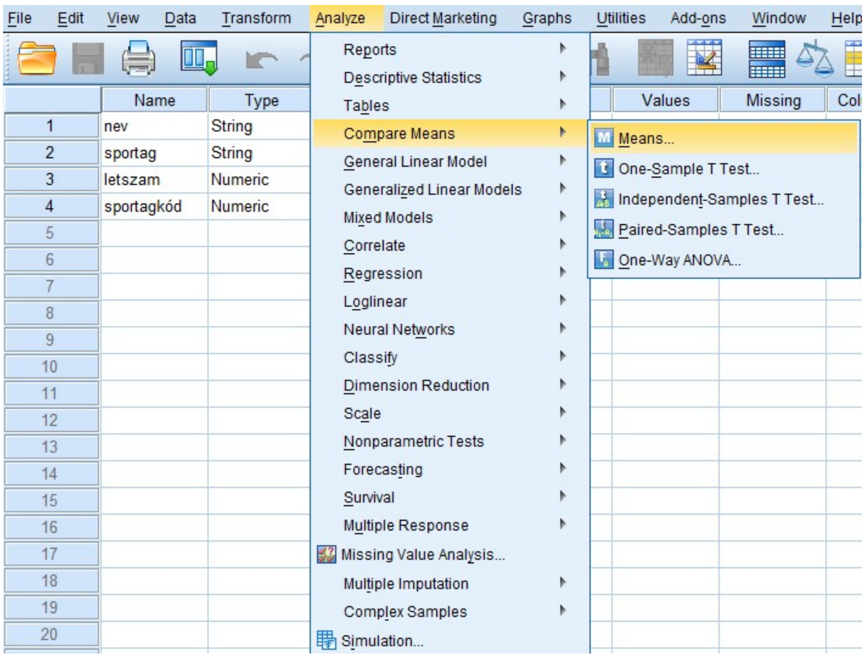
Ügyeljünk arra, ha az adatokat másolni kívánjuk, használjunk abszolút- abszolút hivatkozást a négyzetösszeg függvények során. A következő lépésben az egyes sportágak belső eltérésnégyzet-összegeit összegezzük egyesével. Miután végeztünk, már csak a részadatokra lesz szükségünk, tehát nyomjuk meg a „csoportosító- diagram” 2 gombját.



2/44. képernyőnézet: A vegyes kapcsolat végeredménye az Excel programban

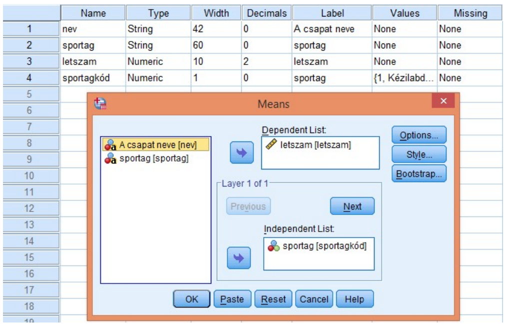
Látható, hogy a program segítségével is megkaptuk a szórásnégyzet-hányadost (0,73), mely a kerekítés miatt módosult egy kissé.

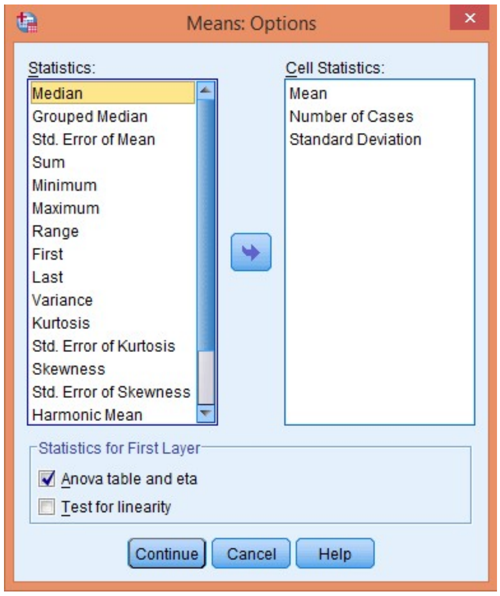
Az SPSS program segítségével még egyszerűbb dolga van a kutatónak, hiszen viszonylag gyorsan elvégezhetők a beállítások, melynek segítségével az eredményekhez juthatunk. (Forrás: vegyes kapcsolat.sav)



2/45. képernyőnézet: A vegyes kapcsolat számszerűsítése

Az érdemi beállításokat az Analyze menü, Compare means almenüjének Means moduljában végezhetjük el. Elő lépésben a függő és a független változókat kell megadni, majd az Options gomb használatával a kért műveleteket tudjuk beállítani.





2/46. és 47. ábra: A vegyes kapcsolat beállításai az SPSS programban

A legfontosabb beállítás az Anova table and eta doboz kijelölése. Az eta (η) a tulajdonképpen a már tanult módon írható fel:

$$\eta = \sqrt{\frac{\sigma^2_K}{\sigma^2}}$$

A Continue majd az Ok gombok lenyomását követően kapjuk meg a számított eredményeket, melyek most is egy összesítő táblázattal kezdődnek. A következő táblázatban (Report) a részátlagokat és a tagszámot láthatjuk.

2/31. táblázat

Report

letszam

sportag	Mean	N	Std. Deviation
Kézilabda	17,2500	4	,95743
Kosárlabda	12,7500	4	,95743
Labdarúgás	25,0000	6	5,21536
Vízilabda	16,3333	3	1,52753
Total	18,7647	17	5,86866

Ezt követi az Anova tábla, mely a belső és külső eltérésnégyzet-összegeket és a szignifikanciára vonatkozó eredményeket közli.

2/32. táblázat

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
letszam * sportag	Between Groups	(Combined)	404,892	3	134,964	12,004	,000
	Within Groups		146,167	13	11,244		
	Total		551,059	[illegible]			

Belső eltérésnégyzet- összeg

Külső eltérésnégyzet- összeg

Ezt követően megkapjuk a kapcsolat-szorossági mérőszámokat is.

2/33. táblázat

Measures of Association

	Eta	Eta Squared
letszam * sportag	,857	,735

H

H^2

2.7.3.3. Korrelációs kapcsolat vizsgálata

Amennyiben mind az ok(ok) mind az okozat szerepét is mennyiségi ismérvek közvetítik, **korrelációs kapcsolat**ról beszélünk. A továbbiakban elsősorban egy *tényező*, vagy *magyarázó változó* (X)- és egy *eredményváltozó* (Y) közötti kapcsolat mérését mutatjuk be, viszont hangsúlyozni kívánjuk, hogy a valóságban általában nem egy, hanem több tényező együttes, igen összetett hatására alakul ki egy-egy jelenség, folyamat. A korrelációs kapcsolat mérése során azonban több ok együttes hatásának vizsgálatát is viszonylag könnyen meg lehet oldani.

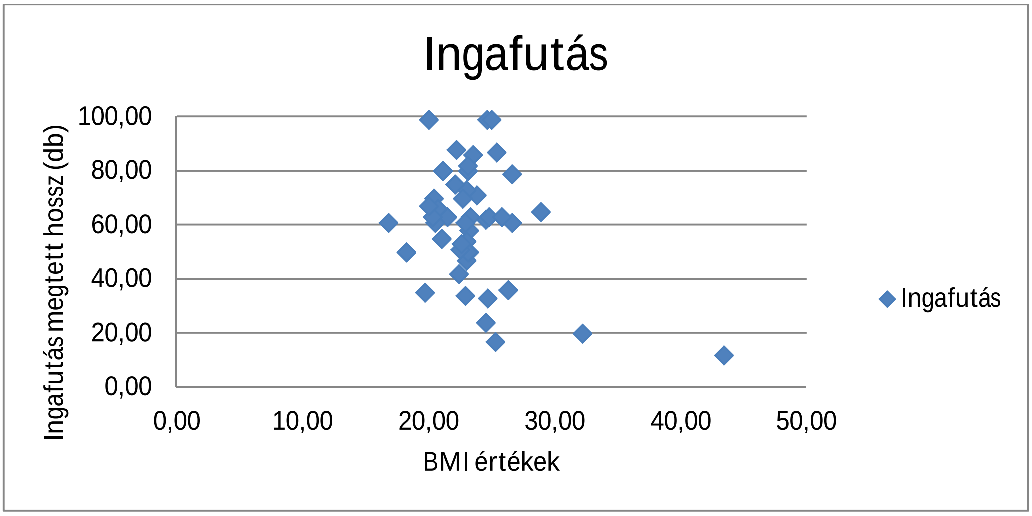
A korreláció természete szerint a változók között az alábbi kapcsolatok értelmezhetők: monoton kapcsolat, illetve ezen belül lineáris kapcsolat. Mindkét kapcsolat lehet pozitív, vagy negatív irányú, melynek megítélést segíti a grafikus megjelenítés. Két mennyiségi ismérv között meglévő kapcsolatot jól ábrázolhatjuk a derékszögű koordináta rendszerben, ún. pontdiagram segítségével. Bővebben az érdeklődő PINTÉR-RAPPAI (2007): Statisztika című könyvében olvashat.

A korrelációs kapcsolat mérésének legelterjedtebb mutatószáma a **lineáris korrelációs együttható (**jele**: r),** amelynek alkalmazása során feltételezzük a változók közötti lineáris kapcsolatot, illetve ha a linearitás feltevése nem áll távol a vizsgált problémától. A korrelációs együttható kiszámítása a változók együttmozgását jellemző kovariancia mérőszáma és a változók szórása segítségével történik, az alábbi algoritmus segítségével:

$$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}, ahol\ a\ kovariancia: C_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum d_x d_y}{n} = \frac{\sum xy}{n} - \overline{xy}$$

ahol: σ_X és σ_Y a változók szórásai.

Most a korrelációs kapcsolat szorosságát vizsgáljuk meg a BMI értékek és a 20 méteres ingafutás példáján. Első lépésben pontdiagram segítségével ábrázoljuk a mért eredményeket. (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)



2/22. ábra: Pontdiagram ábrája

A grafikus ábra alapján feltételezhető a vizsgált változók közötti negatív irányú lineáris kapcsolat. Növekszik a BMI értéke, csökken az ingafutás alatt megtett hosszak száma. A kapcsolat szorosságának meghatározásához szükséges adatokat és részszámításokat a következő munkatábla tartalmazza:

	A	B	C	D	E	F
1	sorszám	BMI	Ingafutás	dx_i	dy_i	dx_idy_i
2		22,20	88,00	-0,97	27,02	-26,16
3		26,60	79,00	3,43	18,02	61,83
4		28,90	65,00	5,73	4,02	23,03
5		23,20	58,00	0,03	-2,98	-0,09
6		16,80	61,00	-6,37	0,02	-0,11
7		23,80	71,00	0,63	10,02	6,33
8		23,50	86,00	0,33	25,02	8,30
9		26,60	61,00	3,43	0,02	0,06
10		21,50	63,00	-1,67	2,02	-3,37
11		21,00	55,00	-2,17	-5,98	12,97
12		20,50	61,00	-2,67	0,02	-0,05
13		22,70	70,00	-0,47	9,02	-4,22
14		18,20	50,00	-4,97	-10,98	54,57
15		22,90	61,00	-0,27	0,02	0,00
16		22,10	75,00	-1,07	14,02	-14,98
17		20,30	63,00	-2,87	2,02	-5,79
18		20,00	67,00	-3,17	6,02	-19,07
19		21,10	80,00	-2,07	19,02	-39,34
20		26,30	36,00	3,13	-24,98	-78,23
21		23,00	47,00	-0,17	-13,98	2,35
22		23,00	73,00	-0,17	12,02	-2,02
23		23,00	54,00	-0,17	-6,98	1,18
24		24,70	33,00	1,53	-27,98	-42,86

H	I	J	K	L	M
sorszám	BMI	Ingafutás	dx_i	dy_i	dx_idy_i
Összeg	1320,6	3476			-1547,93
Átlag	23,16842	60,98246			-27,1567
Szórás	3,757386	19,53359			0

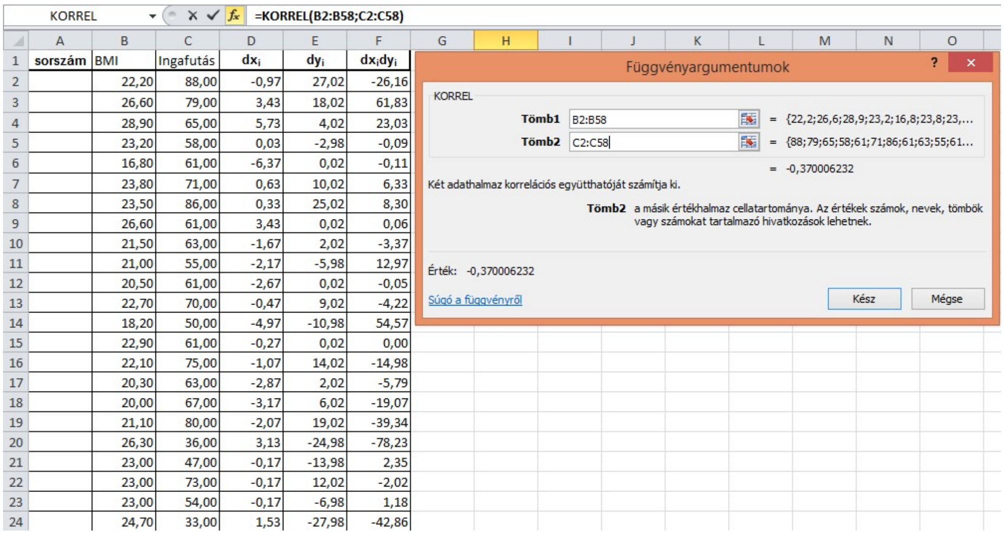
2/48. képernyőnézet: A korrelációs kapcsolat munkatáblája

A munkatábla adatainak felhasználásával a lineáris korrelációs együttható meghatározása:

$$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} = \frac{-27{,}16}{3{,}75 \times 19{,}53} = \text{-}0{,}37, \text{ vagyis a determinációs együttható: } r^2{}_{xy} = 0{,}1369 \text{ azaz } 13{,}69\%.$$

Az eredmény elárulja, hogy a síkfutás és a magasugrás eredményei között közepes szorosságú negatív kapcsolatot áll fent. A determinációs együttható alapján kijelenthető, hogy a BMI értékek eredményei 13,69%-ban határozzák meg az ingafutás eredményeit, vélhetően a fennmaradó többi részt más tényezők befolyásolják (pl: edzettség, stb) A negatív kapcsolat kissé zavaró lehet, de ez úgy értelmezendő: ahogyan növekednek a BMI érték, (kedvezőtlen), úgy csökken a magasugrás az ingafutás eredményessége (lefutott hosszak száma)

A korrelációs kapcsolat szorossága az Excel program segítségével könnyen számítható, hiszen egy előre programozott függvény előhívása kell csak hozzá, melyet a függvényvarázslóból is megtehetünk. Az Excel programban a Korrel függvény és a Pearson függvénnyel is ugyanarra az eredményre juthatunk.



	A	B	C	D	E	F
1	sorszám	BMI	Ingafutás	dx_i	dy_i	$dx_i dy_i$
2		22,20	88,00	-0,97	27,02	-26,16
3		26,60	79,00	3,43	18,02	61,83
4		28,90	65,00	5,73	4,02	23,03
5		23,20	58,00	0,03	-2,98	-0,09
6		16,80	61,00	-6,37	0,02	-0,11
7		23,80	71,00	0,63	10,02	6,33
8		23,50	86,00	0,33	25,02	8,30
9		26,60	61,00	3,43	0,02	0,06
10		21,50	63,00	-1,67	2,02	-3,37
11		21,00	55,00	-2,17	-5,98	12,97
12		20,50	61,00	-2,67	0,02	-0,05
13		22,70	70,00	-0,47	9,02	-4,22
14		18,20	50,00	-4,97	-10,98	54,57
15		22,90	61,00	-0,27	0,02	0,00
16		22,10	75,00	-1,07	14,02	-14,98
17		20,30	63,00	-2,87	2,02	-5,79
18		20,00	67,00	-3,17	6,02	-19,07
19		21,10	80,00	-2,07	19,02	-39,34
20		26,30	36,00	3,13	-24,98	-78,23
21		23,00	47,00	-0,17	-13,98	2,35
22		23,00	73,00	-0,17	12,02	-2,02
23		23,00	54,00	-0,17	-6,98	1,18
24		24,70	33,00	1,53	-27,98	-42,86

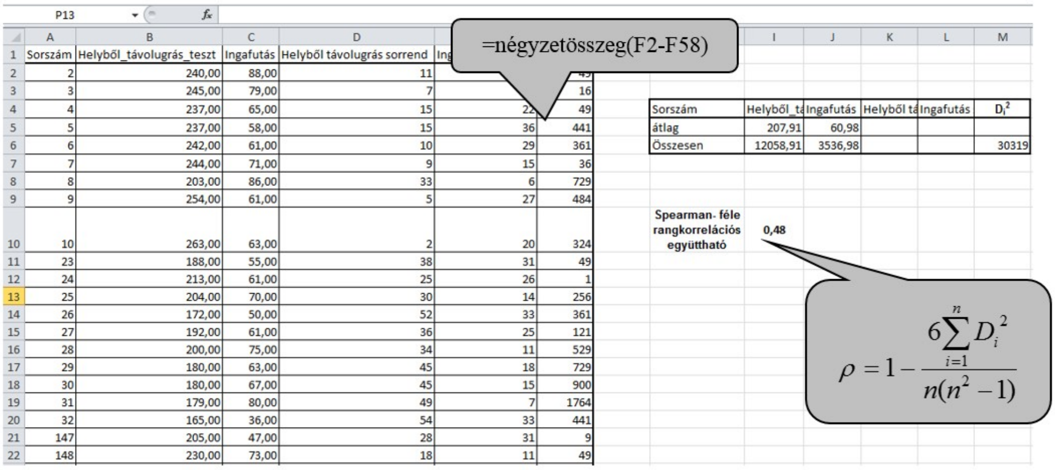
2/49. képernyőnézet: A korrelációs együttható számítása az Excel program segítségével

Eredményül megkapjuk az előbb már kiszámított értéket. Egy kicsit másképpen járunk el, ha a vizsgálatba bevont eredmények rangsorba vannak rendezve. Monoton kapcsolat esetén[27] a kapcsolat szorosságát a Spearman- féle rangkorrelációs együtthatóval mérjük, ami egy robusztusabb mérőszám, azaz érzéketlenebb a kiugró értékekre, hiszen az intervallum- vagy arányskála helyett ordinális skálát használ. Ez azt jelenti, hogy egy magasabb rendű skáláról alacsonyabb rendű skálára alakítjuk át az adatokat. A rangkorrelációs együttható képlete:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left[R(y_i) - R(x_i)\right]^2}{n(n^2-1)} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$$

A sorrendiség felhasználásával vizsgáljuk, hogy milyen szoros kapcsolat van az ingafutások hossza és a helyből távolugrás között. (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)

[27] Monoton kapcsolatnál Y változásának mértéke X egységnyi elmozdulása esetén nem konstans.



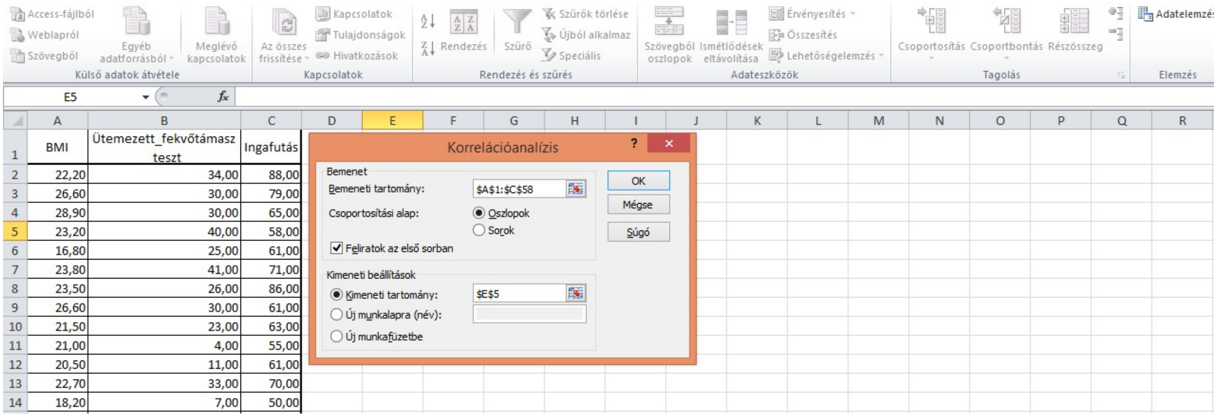
	Sorszám	Helyből_távolugrás_teszt	Ingafutás	Helyből távolugrás sorrend	Ing	
2	2	240,00	88,00	11		[illegible]
3	3	245,00	79,00	7		16
4	4	237,00	65,00	15	22	49
5	5	237,00	58,00	15	36	441
6	6	242,00	61,00	10	29	361
7	7	244,00	71,00	9	15	36
8	8	203,00	86,00	33	6	729
9	9	254,00	61,00	5	27	484
10	10	263,00	63,00	2	20	324
11	23	188,00	55,00	38	31	49
12	24	213,00	61,00	25	26	1
13	25	204,00	70,00	30	14	256
14	26	172,00	50,00	52	33	361
15	27	192,00	61,00	36	25	121
16	28	200,00	75,00	34	11	529
17	29	180,00	63,00	45	18	729
18	30	180,00	67,00	45	15	900
19	31	179,00	80,00	49	7	1764
20	32	165,00	36,00	54	33	441
21	147	205,00	47,00	28	31	9
22	148	230,00	73,00	18	11	49

Sorszám	Helyből_tá	Ingafutás	Helyből tá	Ingafutás	D_i^2
átlag	207,91	60,98			
Összesen	12058,91	3536,98			30319

2/50. képernyőnézet: A Spearman-féle rangkorreláció munkatáblája

Megállapítható, hogy az ingafutás sorrendisége és a helyből távolugrás sorrendisége között közepes szorosságú pozitív kapcsolat van, illetve az ingafutás eredménye 22,69%-ban (determinációs együttható) meghatározza távolugrás eredményét.

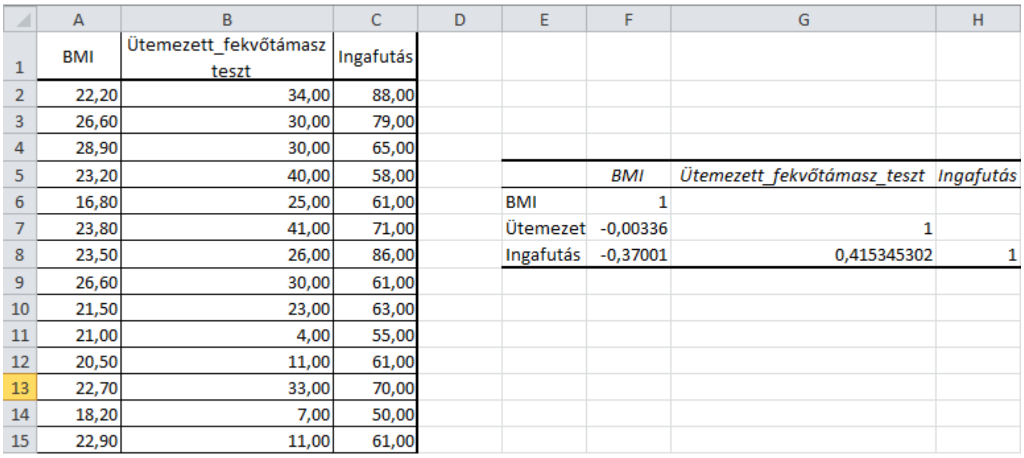
Amennyiben több változó kapcsolatot szeretnénk megvizsgálni, használjuk az Eszközök menü, Adatelemzés almenüjének, Korrelációanalízis beépített modulját. Jelen esetben kibővítettük a hallgatók ütemezett fekvőtámasz eredményeivel (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)



	BMI	Ütemezett_fekvőtámasz teszt	Ingafutás
2	22,20	34,00	88,00
3	26,60	30,00	79,00
4	28,90	30,00	65,00
5	23,20	40,00	58,00
6	16,80	25,00	61,00
7	23,80	41,00	71,00
8	23,50	26,00	86,00
9	26,60	30,00	61,00
10	21,50	23,00	63,00
11	21,00	4,00	55,00
12	20,50	11,00	61,00
13	22,70	33,00	70,00
14	18,20	7,00	50,00

2/51. képernyőnézet: Korrelációanalízis az Excel program segítségével

A bemeneti tartománynak adjuk meg a vizsgálatba bevonni kívánt adatokat. Miután a változók nevét is bevontuk, így jelöljük ki a feliratok az első sorban található lehetőséget. Az ok opció lenyomását követően jutunk a végeredményhez, ami tulajdonképpen egy korrelációs-mátrix. A mátrix a totális és a kétváltozós korrelációs együtthatókat tartalmazza.



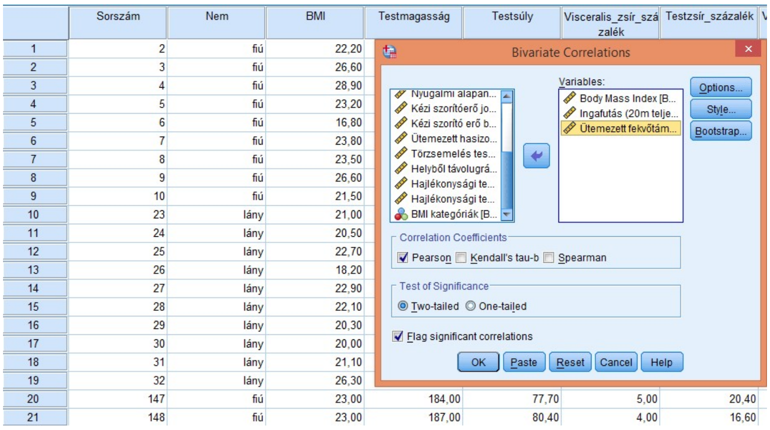
	BMI	Ütemezett_fekvőtámasz teszt	Ingafutás
2	22,20	34,00	88,00
3	26,60	30,00	79,00
4	28,90	30,00	65,00
5	23,20	40,00	58,00
6	16,80	25,00	61,00
7	23,80	41,00	71,00
8	23,50	26,00	86,00
9	26,60	30,00	61,00
10	21,50	23,00	63,00
11	21,00	4,00	55,00
12	20,50	11,00	61,00
13	22,70	33,00	70,00
14	18,20	7,00	50,00
15	22,90	11,00	61,00

	BMI	*Ütemezett_fekvőtámasz_teszt*	*Ingafutás*
BMI	1		
Ütemezet	-0,00336	1	
Ingafutás	-0,37001	0,415345302	1

2/52. képernyőnézet: A korrelációs mátrix eredménye

Az átlóban szereplő értékek értelemszerűen az egyes számértéket veszik fel, hiszen a változók önmagukkal determinisztikus kapcsolatban vannak. Új eredményként látható, hogy a BMI értékek és az ütemezett fekvőtámasz értékek eredményei is gyenge negatív kapcsolatban vannak (-0,01), míg a fekvőtámasz és az ingafutás között közepes szorosságú kapcsolatot (0,42) tapasztalunk.

Az SPSS programban az Analyze menüben a Correlate menüpont Bivariate moduljával érhetjük el a korrelációszámítást. A változók dobozba, helyezzük el a három változónkat (BMI, ingafutás, ütemezett fekvőtámasz). (forrás: fittségi 57fő_adatbázis_alap_bmikat.sav)



2/53. képernyőnézet: A korreláció számítás az SPSS program segítségével

Látható, hogy a Pearson- féle együttható alapbeállításként szerepel, de örömmel tapasztaljuk, hogy lehetőség van a Spearman- féle rangkorellációs értékek számítására is. Végeredményként egy korrelációs- mátrixot kapunk, mely teljes mértékben megegyezik az Excel által kiszámított értékekkel.

2/34. táblázat

Correlations

		Body Mass Index	Ingafutás (20m teljes távok száma)	Ütemezett fekvőtámasz (ism.szám.)
Body Mass Index	Pearson Correlation	1	-,370**	-,003
	Sig. (2-tailed)		,005	,980
	N	57	57	57
Ingafutás (20m teljes távok száma)	Pearson Correlation	-,370**	1	,415**
	Sig. (2-tailed)	,005		,001
	N	57	57	57
Ütemezett fekvőtámasz (ism.szám.)	Pearson Correlation	-,003	,415**	1
	Sig. (2-tailed)	,980	,001	
	N	57	57	57

**. Correlation is significant at the 0.01 level (2-tailed).

Az értékek mögött található csillagok a szignifikáns kapcsolatokat jelentik, melynek értelmezését a későbbiekben alaposan tárgyaljuk. Az azonban leolvasható, hogy egyedül az ütemezett fekvőtámasz teszt és a BMI index értékei között nem találtunk szignifiánsnak a kapcsolatot.

2.7.3.4. Kétváltozós lineáris regresszió

A mennyiségi ismérvek közötti összefüggések vizsgálatára a korrelációszámítás mellett a leggyakrabban alkalmazott statisztikai módszer a regressziószámítás. A regressziószámítás a jelenségek tendenciáit vizsgálja, megpróbálja a kapcsolat természetét valamilyen jól megfogható és értelmezhető függvény formájában megragadni. Ezeket a függvényeket regresszió-függvényeknek nevezzük. A regressziószámítás bemutatását az alapesettel, a kétváltozós lineáris regresszióval tesszük meg. Jelen esetben a vizsgálni kívánt változó, az eredményváltozó alakulása (növekedése vagy csökkenése) sztochasztikusan függ az egyetlen magyarázó változótól. Fontos megjegyezni, hogy a változók közötti kapcsolat a gyakorlatban sokszor nem lineáris. Ilyen esetben mind a szorosság mérésének, mind a kapcsolat törvényszerűségét felíró matematikai modellnek a felépítése viszonylag bonyolult matematikai-statisztikai eljárásokat igényel, melyekkel most a jelen tankönyvben nem foglalkozunk.

Amennyiben a változók között lineáris sztochasztikus kapcsolatot tételezünk fel, egy viszonylag egyszerű matematikai modellel, egy lineáris függvény (egyenes) paramétereinek meghatározásával jól felhasználható regresszió-függvényt számszerűsíthetünk:

$$\hat{Y} = b_0 + b_1 X$$

Az egyenes konstans paramétereinek becslését az ún. legkisebb négyzetek módszerével (LNM)[28] végezhetjük el.

A két paraméter az alábbi képletekkel határozható meg:

$$b_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2}$$

$$b_0 = \overline{Y} - b_1 \overline{X} = \frac{\sum Y}{n} - \frac{b_1 \sum X}{n}$$

A gyakorlatban a tényezőváltozó paraméterének (b_1) különösen kiemelt a szerepe, regressziós együtthatónak nevezi a statisztika, míg a b_0 paramétert tengelymetszetnek vagy konstansnak hívják. A regressziós együttható a tényezőváltozó egységnyi növekedését kísérő várható eredményváltozó- változást számszerűsíti az eredetileg megadott mértékegységben, vagyis a magyarázó változó értékének egy egységnyi változása az eredményváltozót b_1 egységnyivel változtatja.

A következőkben egy gyakorlati példán keresztül kívánjuk bemutatni a kétváltozós lineáris regressziót.

Vizsgáljuk meg, a helyből távolugrás és a BMI kapcsolatát. számszerűsítsük, hogy a BMI egységnyi változása mekkora változást idéz elő a helyből távolugrás eredményében.

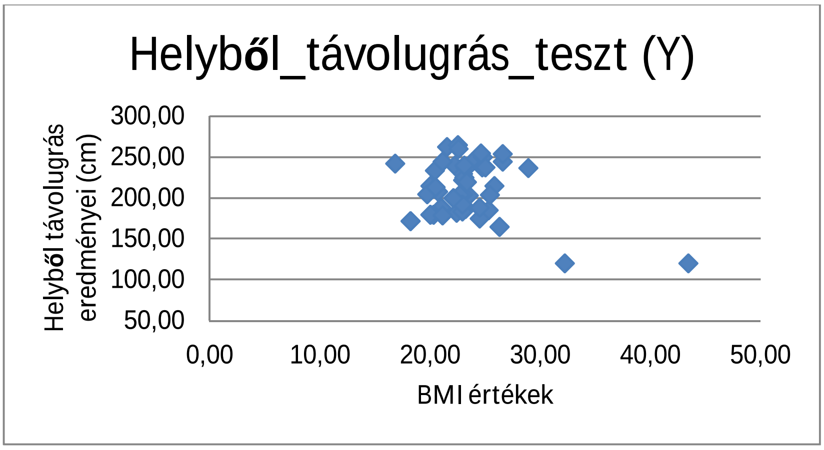
A következő táblázat a helyből távolugrás értékeit, valamint a BMI index értékeit tartalmazza. (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)

[28] A becslési módszer elvi leírásától tananyagunkban eltekintünk, csupán a módszer alkalmazásával nyert megoldóképleteket használjuk fel.

2/35. táblázat: Alapadatokat tartalmazó munkatábla (részlet)

	A	B	C	D	E	F
1	Sorszám	BMI (X)	Helyből_távolugrás _teszt (Y)	XY	x^2	y^2
2	2	22,20	240,00	5328	492,84	57600
3	3	26,60	245,00	6517	707,56	60025
4	4	28,90	237,00	6849,3	835,21	56169
5	5	23,20	237,00	5498,4	538,24	56169
6	6	16,80	242,00	4065,6	282,24	58564
7	7	23,80	244,00	5807,2	566,44	59536
8	8	23,50	203,00	4770,5	552,25	41209
9	9	26,60	254,00	6756,4	707,56	64516
10	10	21,50	263,00	5654,5	462,25	69169
11	23	21,00	188,00	3948	441	35344
12	24	20,50	213,00	4366,5	420,25	45369
13	25	22,70	204,00	4630,8	515,29	41616
14	26	18,20	172,00	3130,4	331,24	29584
15	27	22,90	192,00	4396,8	524,41	36864
16	28	22,10	200,00	4420	488,41	40000
17	29	20,30	180,00	3654	412,09	32400
18	30	20,00	180,00	3600	400	32400
19	31	21,10	179,00	3776,9	445,21	32041

Az alapadatokat célszerű először pontdiagram segítségével megjeleníteni, melyből láthatóvá válik a kapcsolat típusa.



2/23. ábra: A vizsgált változók (távolugrás, BMI) kapcsolatának összefüggése

A becsült paraméterek számítását a következő munkatábla tartalmazza.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sorszám	BMI (X)	Helyből_távolugrás _teszt (Y)	**XY**	**x^2**	**y^2**	Sorszám	BMI	Helyből_távol ugrás_teszt	**XY**	**x^2**	**y^2**
2	2	22,20	240,00	5328	492,84	57600	**Összesen**	1320,60	11851,00	272679,40	31400,94	2523819,00
3	3	26,60	245,00	6517	707,56	60025	**átlag**	23,17	207,91	4783,85	550,89	44277,53
4	4	28,90	237,00	6849,3	835,21	56169	**szórás**	3,76	32,40			
5	5	23,20	237,00	5498,4	538,24	56169						
6	6	16,80	242,00	4065,6	282,24	58564						
7	7	23,80	244,00	5807,2	566,44	59536						
8	8	23,50	203,00	4770,5	552,25	41209						
9	9	26,60	254,00	6756,4	707,56	64516						

2/54. képernyőnézet: A regresszió munkatáblája (részlet)

A fent közölt képleteket alkalmazva, a paraméterek a következők lesznek:

$$b_1 = \frac{57 \times 272679{,}4 - 1320{,}6 \times 11851}{57 \times 31400{,}94 - 1320{,}6^2} = -2{,}35$$

$$\mathrm{b}_0 = 207{,}91 - (-2{,}35) \times 23{,}17 = 262{,}31$$

A regresszió-függvény: $\hat{Y} = 262{,}31 - 2{,}35X$

A regressziós együttható ismeretében elmondható, hogy a BMI egységnyi növekedése a helyből távolugrás hosszát átlagosan várhatóan 2,35 (cm)-rel csökkenti.

A modell segítségével megbecsülhetjük egy 21 BMI értékkel rendelkező hallgató várható távolugrás eredményét: $\hat{Y} = 262{,}31 - 2{,}35 \times 21 = 212{,}96$

A regressziós együttható ismerete lehetővé teszi, hogy lineáris összefüggés esetén is kvantifikáljuk az elaszticitást (a rugalmasságot), amely a változás relatív (százalékban kifejezett) mértékét fejezi ki. A mutató választ ad, arra, hogy a X magyarázó változó 1%-os változásához az Y eredményváltozó hány %-os változása járul várhatóan. Az átlagos elaszticitás a változók átlagai segítségével az alábbi módon határozható meg:

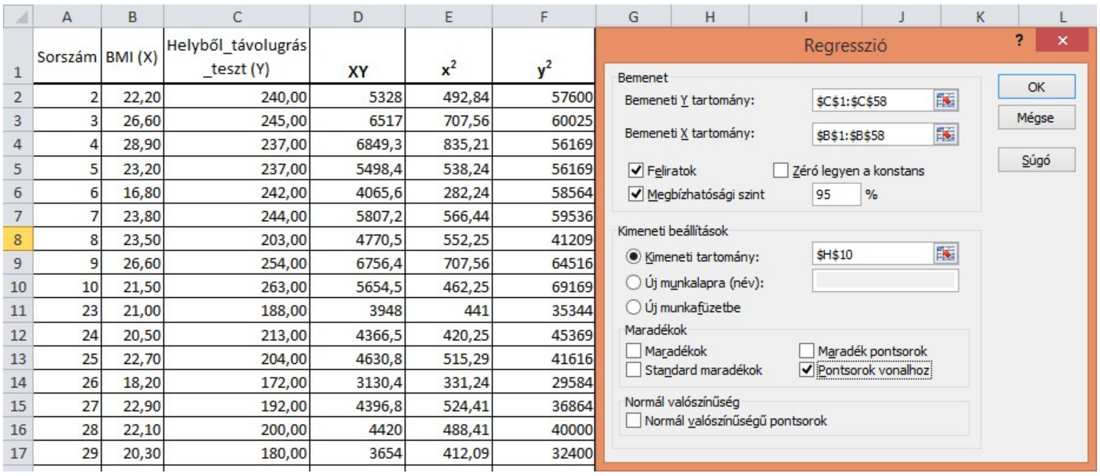
$$\mathrm{El} = \mathrm{b}_1 \frac{\overline{\mathrm{x}}}{\overline{\mathrm{y}}}$$

Az átlagos elaszticitási mutató példánkban: $El = -2{,}35 \dfrac{23{,}17}{207{,}91} = -0{,}26\ \%$

Ez azt jelenti, hogy az távolugrás értéke a BMI változásokra kevésbé erősen reagál.

A regressziószámítás bőséges irodalmát és módszereit az olvasó megtalálhatja PINTÉR-RAPPAI (2007), MUNDRUCZÓ (1981), RAMANATHAN (2003) könyveiben.

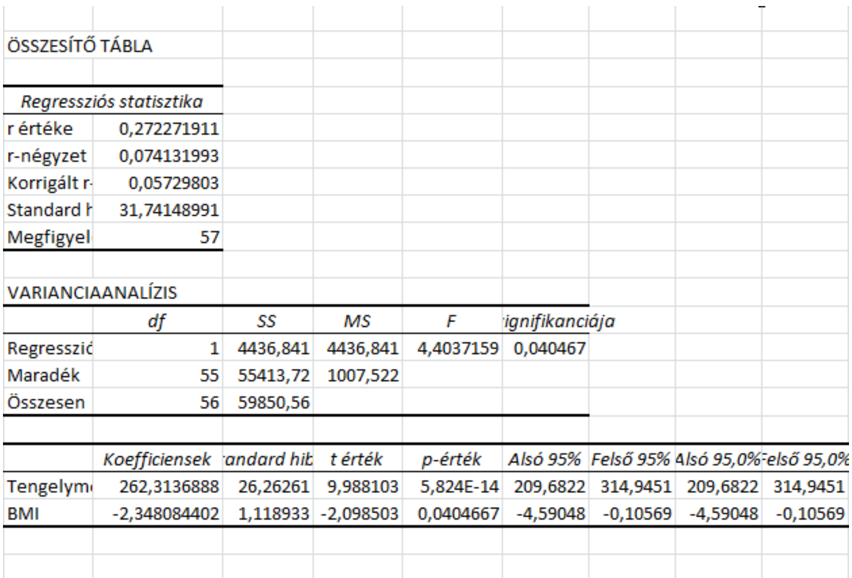
Az Excel programban az Adatok menü Adatelemzés almenüjének Regresszió moduljában könnyen elkészíthetjük az elemzéseket.



	A	B	C	D	E	F
1	Sorszám	BMI (X)	Helyből_távolugrás_teszt (Y)	XY	x^2	y^2
2	2	22,20	240,00	5328	492,84	57600
3	3	26,60	245,00	6517	707,56	60025
4	4	28,90	237,00	6849,3	835,21	56169
5	5	23,20	237,00	5498,4	538,24	56169
6	6	16,80	242,00	4065,6	282,24	58564
7	7	23,80	244,00	5807,2	566,44	59536
8	8	23,50	203,00	4770,5	552,25	41209
9	9	26,60	254,00	6756,4	707,56	64516
10	10	21,50	263,00	5654,5	462,25	69169
11	23	21,00	188,00	3948	441	35344
12	24	20,50	213,00	4366,5	420,25	45369
13	25	22,70	204,00	4630,8	515,29	41616
14	26	18,20	172,00	3130,4	331,24	29584
15	27	22,90	192,00	4396,8	524,41	36864
16	28	22,10	200,00	4420	488,41	40000
17	29	20,30	180,00	3654	412,09	32400

2/55. képernyőnézet: A regresszió számítás az Excelben

A bemeneti Y tartományba a helyből távolugrás (cm) értékei, míg az X tartományba a BMI adatokat jelöljük ki. Miután a feliratokat is kiválasztottuk, így a Feliratok dobozt is válasszuk, illetve a megbízhatósági szintet, valamint kérhetünk egy grafikus ábrát a pontsorok a vonalhoz négyzet kijelölésével. A beállításokat követően a következő eredményhez jutunk:



ÖSSZESÍTŐ TÁBLA

Regressziós statisztika	
r értéke	0,272271911
r-négyzet	0,074131993
Korrigált r-	0,05729803
Standard h	31,74148991
Megfigyel	57

VARIANCIAANALÍZIS

	df	*SS*	*MS*	*F*	*ignifikanciája*
Regresszió	1	4436,841	4436,841	4,4037159	0,040467
Maradék	55	55413,72	1007,522		
Összesen	56	59850,56			

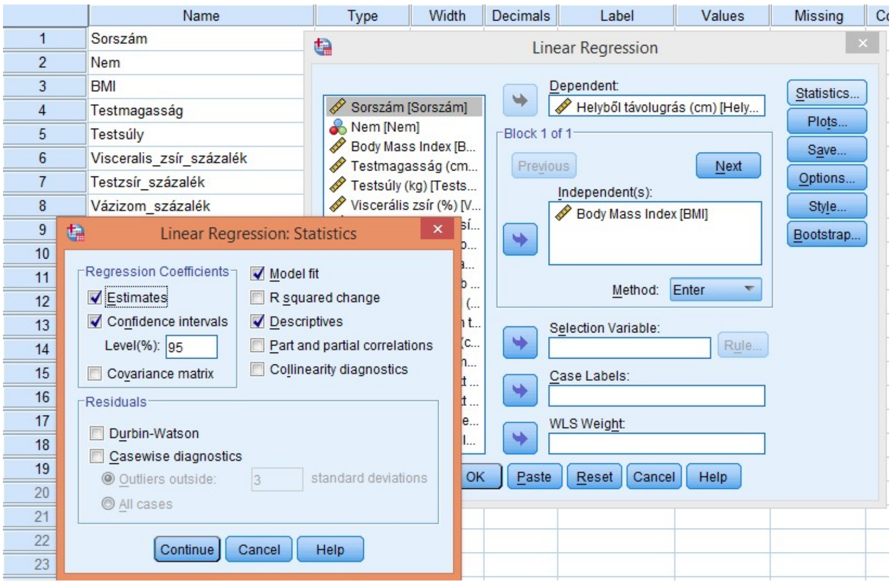
	Koefficiensek	*andard hib*	*t érték*	*p-érték*	*Alsó 95%*	*Felső 95%*	*Alsó 95,0%*	*Felső 95,0%*
Tengelym	262,3136888	26,26261	9,988103	5,824E-14	209,6822	314,9451	209,6822	314,9451
BMI	-2,348084402	1,118933	-2,098503	0,0404667	-4,59048	-0,10569	-4,59048	-0,10569

2/56. képernyőnézet: A regresszió számítás összesítő eredményei

A korábban részletesen számított eredményeket találjuk a táblában. Az „r-értéke” azt mutatja, hogy gyenge kapcsolat van az a két vizsgált változó között (BMI, helyből távolugrás értéke). Az BMI 7,41%-ban meghatározza a helyből távolugrás eredményeit (cm). Fontos megemlíteni, hogy az r^2 magas értéke jelezheti, hogy a egyenesünk jól illeszkedik a ponthalmazra.

Az általunk becsült paraméterek (b_0; b_1) a Koefficiens oszlopban olvashatók. Itt látható, hogy a b_0 paraméter a Tengelymetszet nevet kapja.[29] A paraméterbecslés rész végén lehetőség nyílik az általunk beállított megbízhatósági szinten a paraméterekre való intervallumbecslés értelmezésére is (lásd később). A többváltozós lineáris regresszió számítása is hasonló módón történik.

Az SPSS program segítségével is gyorsan juthatunk ugyanezen eredményhez, ha az Analyze menüpont, Regression almenüjének, Linear modulját választjuk. (forrás: fittségi 57fő_adatbázis_alap_bmikat.sav)



2/57. képernyőnézet: Regresszió számítás az SPSS program segítségével

A függő változónak a helyből távolugrás eredményeit, míg független változónak BMI értékeket jelöljük, majd a STATISTICS gomb lenyomását követően az alapbeállításokhoz válasszuk még a CONFIDENCE INTERVALS, MODEL FIT és a DESCRIPTIVES lehetőségeket. A CONTINUE és OK gombok választását követően az számítási eredményeket az Output View nézetben olvashatjuk.

[29] A t-próbák a paraméterek 0-val való egyezőségét teszteli.

2/36. táblázat

Descriptive Statistics

	Mean	Std. Deviation	N
Helyből távolugrás (cm)	207,9123	32,69190	57
Body Mass Index	23,1684	3,79079	57

2/37. táblázat

Correlations

		Helyből távolugrás (cm)	Body Mass Index
Pearson Correlation	Helyből távolugrás (cm)	1,000	-,272
	Body Mass Index	-,272	1,000
Sig. (1-tailed)	Helyből távolugrás (cm)	.	,020
	Body Mass Index	,020	.
N	Helyből távolugrás (cm)	57	57
	Body Mass Index	57	57

Az első táblázatban a leíró statisztikai adatok láthatóak, melyet a korrelciós mátrix követ. Ezt követően láthatóak a regressziós modellre vonatkozó összegző táblázatok:

2/38. táblázat

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,272[a]	,074	,057	31,74149

a. Predictors: (Constant), Body Mass Index

2/39. táblázat

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4436,841	1	4436,841	4,404	,040[b]
	Residual	55413,720	55	1007,522		
	Total	59850,561	56			

a. Dependent Variable: Helyből távolugrás (cm)

b. Predictors: (Constant), Body Mass Index

A modellre vonatkozó összefoglaló első táblázatban PEARSON féle korrelációs együtthatót láthatjuk, mely alapján gyenge kapcsolatról beszélhetünk (r=0,272). Ezt követően látható a kapcsolat erejét számszerűsítő determinációs együttható (r^2 =0,074), mely azt jelenti teljes szórás közel 7,4%-át tudja a független változó magyarázni, vagyis a BMI értékek a helyből távolugrás eljesítményének változásában a tömeg csupán 7%-ban játszik szerepet. Fontos megemlíteni, hogy minél magasabb az r^2, annál jobban illeszkedik az egyenesünk a ponthalmazra. Ezt követi a becslés standard hibája, mely az előrejelzés pontosságának proxy mutatója (minél magasabb, annál kevésbé képes a modell jól becsülni). Ezt követően az ANOVA táblázat következik, mely az F- próba értékét és a kapcsolat meglétét igazoló ($p<0,05$) szignifikancia értéket tartalmazza. A táblázat alapján megállapítható, hogy a két változó közötti kapcsolat létezik és nem a véletlen műve. Ezt követően a regressziós egyenes paramétereit is tartalmazó koefficiens táblázat látható.

2/40. táblázat

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	262,314	26,263		9,988	,000	209,682	314,945
	Body Mass Index	-2,348	1,119	-,272	-2,099	,040	-4,590	-,106

a. Dependent Variable: Helyből távolugrás (cm)

A modell értelmezése előtt látható, hogy mindkét változóhoz tartozó t- érték, illetve az ezekhez tartozó szignifikancia értékek ($p<0,05$) is a modell meglétét igazolják. A paraméterbecslés rész végén lehetőség nyílik az általunk beállított megbízhatósági szinten a paraméterekre való intervallumbecslés értelmezésére is.

A nem standardizált koefficiensek segítségével leolvasható a regressziós egyenes egyenlete:

$$b_1 = -2,35$$
$$b_0 = 262,31$$

A regresszió-függvény:

$$\hat{Y} = 262,31 - 2,35X$$

2.7.4. Következtetéses statisztikai elemzések

A sportéletben is, mint az élet egyéb más területén nagyon sokszor fordul elő, hogy egy eseményről, jelenségről nem rendelkezünk teljes körű információval

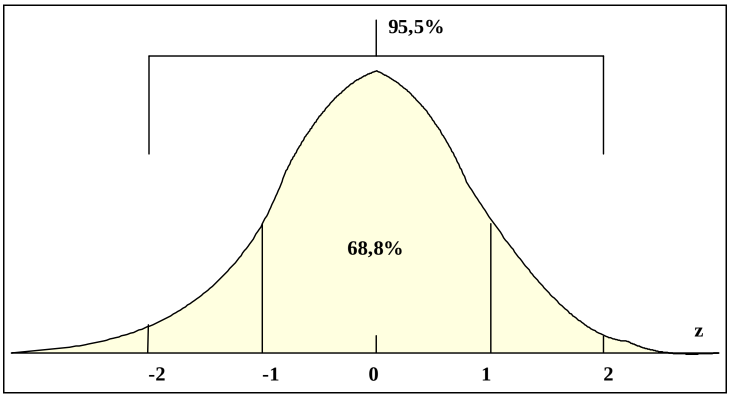
A következtetéses statisztikai módszerek mindazon eljárások, melyek segítségével a megfigyelhető egyedek egy részének felmérése után, a sokaság valamennyi egyedére vonatkozó megállapításokat teszünk. Elmondható, hogy az így létrejövő adatbázis nem teljes körű, hanem csak egy bizonyos technikával (mintavételi technika) kiválasztott részsokaságra, mintára vonatkozik. Ennek köszönhetően a következtetéses statisztikai eljárások során a "bizonytalanság" mindig jelen van. A következtetéses statisztikai eljárások két nagy csoportjával a becslési és a hipotézis ellenőrzési módszerekkel kívánunk foglalkozni. A következtetéses statisztika matematikai alapjait a valószínűségelmélet szolgáltatja, melyet az érdeklődő Hunyadi (2001) művében részletesen megismerhet.

A társadalmi és gazdasági jelenségek, valamint a sportteljesítmények jelentős köréről feltesszük, hogy folytonos, normális eloszlású valószínűségi változóként viselkednek. A folytonos valószínűségi változók tulajdonsága, hogy egy adott intervallumban végtelen számú értéket vehetnek fel, és annak valószínűsége, hogy egy X változó pontosan x értékét veszi fel, zérus. A valószínűségi eloszlások fontos „azonosítója" a várható érték (μ) és a variancia, szórásnégyzet (σ^2). A normális eloszlás könnyen azonosítható a várható érték és a szórás segítségével, jele: $N(\mu,\sigma)$. A normalitás feltételezésével élünk pl. a súly, a térfogat, magasság, hosszúság, és a teljesítmények esetében.

A várható értékek és a szórások, az elemzés tárgyától függően, igen sokféle értéket vehetnek fel, ami a munkát sokszor megnehezíti, hiszen nagyságuk a változók dimenziójától függ. A **standardizálás** segítségével azonban ez a probléma megoldható, ami azt jelenti, hogy a várható értéket kivonjuk a valószínűségi változó értékéből, és a különbséget elosztjuk a szórással, így egy **standard normális eloszlású valószínűségi változó**t (jele: z) kapunk eredményül. Képletben:

$$z = \frac{x-\mu}{\sigma}$$

A standardizálás eredményeként kapott standard normális eloszlású valószínűségi változó várható értéke zérus, szórása egységnyi, azaz N (0,1). Mindkettő - a normális és a standard normális eloszlású- valószínűségi változó sűrűségfüggvénye ún. harang-görbével, **Gauss-görbével (2/76. ábra)** jellemezhető. Standard normális eloszlás esetén mind a valószínűségi változók, mind a hozzájuk rendelhető valószínűségek táblázatba foglalhatók, melyek segítségével a kapott értékek könnyen és viszonylag gyorsan felhasználhatók gyakorlati problémák megoldására, bár a számítógépek korában az ehhez hasonló táblázatok szép lassan kikopnak a gyakorlati alkalmazások repertoárjából.



2/24. ábra: Fontosabb valószínűségek a z függvényében

A várható értéktől egységnyi szórással eltérő intervallum – és ez nemcsak a standard, hanem az általános normális eloszlás esetére érvényes – és a valószínűségi görbe által bezárt terület 68,8 %-os valószínűséget reprezentál. A kétszeres szórás által meghatározható intervallumhoz tartozó valószínűség 95,5 %; míg a háromszoros szórással lefedhetjük a vízszintes tengely és a görbe által meghatározható teljes területet, szinte a teljes valószínűséget (99,9 %). Mivel a standard normális valószínűségi változó sűrűségfüggvénye szimmetrikus, így elegendő a 0 és a pozitív végtelen közé eső számokhoz tartozó valószínűségi értéket meghatározni.

A statisztikai középértékek – különösen a számtani átlag – kiemelt fontossággal bírnak a következtetéses statisztikában is. Közvetlenül adódik annak igénye, hogy a reprezentatív módon kiválasztott minták átlagai és szórásai, valamint az alapsokaság átlaga és szórása között valamilyen összefüggést keressünk. Hangsúlyozni kell, hogy a **centrális határeloszlás** tétele értelmében bármilyen eloszlással rendelkező alapsokaságból egyszerű véletlen mintavétel segítségével nyert minta átlaga valószínűségi változó, mivel értéke mintáról mintára ingadozik, ugyanakkor az **átlagok normális eloszlású valószínűségi változók**. Mindez természetesen fokozottan aláhúzza a normális eloszlás gyakorlati hasznosíthatóságát, elterjedtségét.

A következtetéses statisztika igényli különböző összefüggések felismerését a mintaátlagok, azok szórása és az alapsokasági átlag és szórás között. Könnyen belátható, hogy amennyiben ismerjük valamennyi minta átlagát, a minták átlagából képzett átlag megegyezik az alapsokasági átlaggal. A mintaátlagok szórása azonban eltér az alapsokaság szórásától:

Létezik azonban – bizonyítás nélkül közöljük – egy olyan összefüggés, amelynek segítségével közvetlen kapcsolat írható fel az alapsokasági szórás (szórásnégyzet) és a mintaátlagok szórása (szórásnégyzete) között:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$$

ahol: *n* a mintaelemek száma és *N* az alapsokaság elemeinek száma.

Itt jegyezzük meg, hogy a kifejezés második tagját, az $\left[\frac{N-n}{N-1}\right]$ tényezőt korrekciós tényezőnek vagy véges szorzónak hívja az irodalom. A visszatevés nélküli kiválasztás[30] esetén játszik fontos szerepet, visszatevéses mintavétel alkalmazása során nem szerepel a képletben. Itt kell szólni arról, hogy a korrekciós tényezőt elhagyhatjuk visszatevés nélküli kiválasztás, azaz egyszerű véletlen mintavétel esetén is, amennyiben az alapsokaság (N) nagysága jelentősen eltér a minta (n) nagyságától, mivel ilyen esetekben a tényező 1-hez közeli értékkel bír.

A mintaátlag szórásnégyzete $\sigma_{\bar{x}}^2$, egy olyan átlagos négyzetes hiba, amelyet akkor követünk el, amikor következtetéseink során a sokasági várható értéket mindig a mintaátlaggal helyettesítjük. A statisztikai módszerek között kiemelkedő fontossággal bír a mintaátlag szórása ($\sigma_{\bar{x}}$), amit a mintaátlag **standard hibájának** neveznek. Az alapsokaság szórásának ismeretében tehát könnyen kiszámítható a mintaátlagok szórása.

A véletlen minta elemei véletlen változók, ezért bármely transzformációjuk, így a belőlük számított számtani átlag is, véletlen változó lesz. Ha a sokasági eloszlás normális, akkor a mintaátlag is normális eloszlású, függetlenül a minta elemszámától. Ezt figyelembe véve, a mintaátlagok is standardizálhatók a $z = \frac{x-\mu}{\sigma}$ képlet alapján[31].

2.7.4.1. A statisztikai becslések

A statisztikai becslés az ismeretlen alapsokaság valamely konstans paraméterének közelítő jellegű meghatározása. Ilyen paraméterek: várható érték (véges alapsokaságnál, átlag), szórás és az arány.

---

[30] A visszatevés nélküli mintavétel (pl. egyszerű véletlen mintavétel) a gyakorlatban igen népszerű, mivel alkalmazása nem jár információveszteséggel.

[31] Pintár- Ács (2006)

Láttuk azonban, hogy az alapsokaság átlaga, valamint a mintaátlagok között közvetlen, a szórás és a mintaátlagok szórása között is jól kifejezhető összefüggés írható fel. Különösen fontos szerepet tölt be a standard hiba, a mintaátlagok szórása. Ez a szóródási mérőszám lehetőséget ad arra, hogy a becslésünket egy olyan intervallummal adjuk meg, aminek a bekövetkezése, adott valószínűségi szinten, garantálható.

A $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$ képlet alapján szükségünk van az alapsokasági szórás ismeretére, ha mintánk van, akkor a korrigált mintabeli szórást használjuk, melyet előre programozva az Excelben a szórás függvénnyel hívhatunk elő, melynek képlete:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

A korrigált mintabeli szórás segítségével felírható a gyakorlatban jól használható standard hiba képlete is, melynél a véges szorzót $\left(1 - \frac{n}{N}\right)$, akkor használjuk, ha a mintánk nagysága meghaladja az alapsokaság nagyságának 5%-át.:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Hangsúlyoznunk kell, hogy a fenti standard hiba képlete csupán az átlagok szóródását jellemzik. Más paraméterekre pl. értékösszeg, arány is felírhatók a megfelelő szórások, más néven standard hibák.

Azokat a mintából származó statisztikákat, melyeket az alapsokasági paraméterek közelítő meghatározására használnak, becslőfüggvénynek nevezik. A becslőfüggvény egy adott mintára vonatkozó konkrét értékét, pontbecslésnek hívják. A becslés során elkövethető véletlen hiba átlagos nagyságát a standard hiba (becslőfüggvény szórása) szolgáltatja. A következő táblázat a leggyakrabban használt alapsokasági paraméterbecslések fő jellemzőit tartalmazza.

2/41. táblázat: Legfontosabb sokasági paraméterek becslőfüggvényi és azok jellemzői

Alapsokasági paraméter	Torzítatlan becslőfüggvény	Standard hiba	Becslőfüggvény eloszlása
várható érték	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$	kis minta (n<50) t- eloszlás nagy minta (n≥50) normális
arány	$p = \frac{k}{n}$	$S_p = \sqrt{\frac{p(1-p)}{n}}$	kis minta (n<50) binomiális nagy minta (n≥50) normális

A gyakorlatban jól használható információt nyerünk azonban akkor, ha intervallumbecslést végzünk. Az intervallumbecslés során felhasználjuk azt, hogy a minta-paraméterek valamilyen ismert eloszlású valószínűségi változók, és így az adott eloszlás értékének felhasználásával egy **adott megbízhatósági szinten** állapíthatunk meg egy intervallumot. Ezt az intervallumot **konfidencia intervallum**nak hívjuk. Az intervallumok meghatározásához szükséges kritikus érték – a normális eloszlás szimmetrikus voltából adódóan- a 0-ra szimmetrikusan helyezkedik el. A pontbecslés, a standardhiba és az eloszlás típusának ismeretében a konfidencia intervallumot (ez egy pontbecslés, amely köré mindkét irányba felvesszük a hibahatárt) már felírhatjuk. A hibahatár tartalmazza az általunk pozitív és negatív irányba tolerált maximális „pontatlanságot”. Az átlagbecslés esetén a konfidencia intervallum:

$$\bar{x} \pm z \times \sigma_{\bar{x}}$$

ahol: *z* a standard normális eloszlás adott értéke, melyek közül a fontosabbakat az alábbiak:

2/42. táblázat: Gyakran használt kritikus értékek

α	1-α	$Z_{(\alpha/2)}$	$Z_{(1-\alpha/2)}$
0,01	0,99	-2,576	2,576
0,05	0,95	-1,96	1,96
0,1	0,9	-1,645	1,645

Ennek tükrében a becslések gyakorlata 6 lépésből áll. Az első három elméleti feladat (mintavétel, becslőfüggvény konstruálása, becslőfüggvény megítélése), míg a második három a gyakorlati munka (pontbecslés, standard hiba meghatározás, intervallum-becslés)[32].

[32] Rappai G. (2001)

Ennek segítségével nézzünk egy konkrét példát:

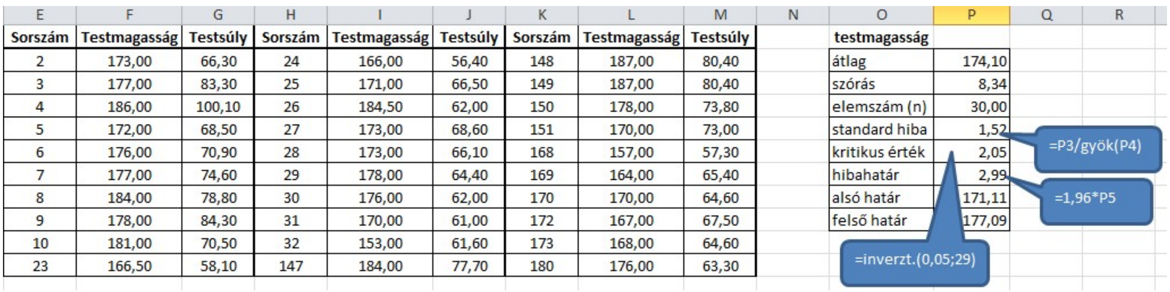
Ismerjük a felmérésünkből a 57 egyetemi hallgatónk testmagasság eredményeit, , melyből egyszerű véletlen mintavétellel kiválasztottunk 30 főt. Becsüljük meg 95 %-os megbízhatóság mellett a hallgatók testmagasság értékét (cm) (forrás: fittségi 57fő_adatbázis_alap_bmikat.xlsx)!

2/43. táblázat: Véletlen mintavétellel kiválasztott hallgató testmagasság és testsúly paraméterei

Sorszám	Testmagasság	Testsúly	Sorszám	Testmagasság	Testsúly	Sorszám	Testmagasság	Testsúly
2	173,00	66,30	24	166,00	56,40	148	187,00	80,40
3	177,00	83,30	25	171,00	66,50	149	187,00	80,40
4	186,00	100,10	26	184,50	62,00	150	178,00	73,80
5	172,00	68,50	27	173,00	68,60	151	170,00	73,00
6	176,00	70,90	28	173,00	66,10	168	157,00	57,30
7	177,00	74,60	29	178,00	64,40	169	164,00	65,40
8	184,00	78,80	30	176,00	62,00	170	170,00	64,60
9	178,00	84,30	31	170,00	61,00	172	167,00	67,50
10	181,00	70,50	32	153,00	61,60	173	168,00	64,60
23	166,50	58,10	147	184,00	77,70	180	176,00	63,30

Az átlag (mintaátlag) számítása: $\overline{\mu} = \dfrac{\sum x}{n} = 174{,}1$ és a szórás (korrigált szórás, hiszen mintáról van szó): $s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} = 8{,}34$ eredményeit felhasználva kiszámíthatjuk a standard hibát is (Véges szorzót nem használunk, mivel a minta nagysága az alapsokaság nagyságának 5%-át nem haladja meg): $s_{\overline{x}} = \dfrac{s}{\sqrt{n}} = 1{,}52$. Mivel kis mintánk van (n<50) ezért a t- eloszlást táblaértékeiből keressük ki a kritikus értéket: z=2,05. ($t^{szf,\ megbízhatóság}$)

A hibahatár értéke($z \times \sigma_{\overline{x}}$): 2,99, mely után a végeredmény a következő: 174,1±2,99



E	F	G	H	I	J	K	L	M	N	O	P
Sorszám	Testmagasság	Testsúly	Sorszám	Testmagasság	Testsúly	Sorszám	Testmagasság	Testsúly		testmagasság	
2	173,00	66,30	24	166,00	56,40	148	187,00	80,40		átlag	174,10
3	177,00	83,30	25	171,00	66,50	149	187,00	80,40		szórás	8,34
4	186,00	100,10	26	184,50	62,00	150	178,00	73,80		elemszám (n)	30,00
5	172,00	68,50	27	173,00	68,60	151	170,00	73,00		standard hiba	1,52
6	176,00	70,90	28	173,00	66,10	168	157,00	57,30		kritikus érték	2,05
7	177,00	74,60	29	178,00	64,40	169	164,00	65,40		hibahatár	2,99
8	184,00	78,80	30	176,00	62,00	170	170,00	64,60		alsó határ	171,11
9	178,00	84,30	31	170,00	61,00	172	167,00	67,50		felső határ	177,09
10	181,00	70,50	32	153,00	61,60	173	168,00	64,60			
23	166,50	58,10	147	184,00	77,70	180	176,00	63,30			

2/58. képernyőnézet: A becslés munkatáblája

Tehát 95%-os megbízhatóság mellett megállapíthatjuk, hogy a hallgatók testmagasságának átlagos értéke minimum 171,11 és maximum 177,09 cm.

A kritikus értékek meghatározása az Excelben előre programozott inverz.stnorm(valószínűség)[33], illetve az inverz.t[34](valószínűség, szabadságfok) függvények segítségével történhet.

Amennyiben értékösszegbecslést kell végeznünk, akkor a használhatók a következő összefüggések: $x' = N \times \bar{x}$ *, illetve* $\sigma_{x'} = N \times \sigma_{\bar{x}}$. Ebben a példában ez nem értelmezhető.

Az átlagbecsléséhez hasonlóan becsülhető az alapsokaság valamilyen ismérv szerinti aránya (megoszlási viszonyszáma) is. Valamely tulajdonsággal bíró egyed arányát jelöljük az alapsokaságban P-vel. A P arány pontbecslése:

$$p = \frac{k}{n}$$

ahol: *k a mintában* az adott tulajdonsággal bíró egyedek száma, n a minta elemszáma.

A mintabeli aránynak a mintából számítható standard hibája:

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)}$$

Nagy minta esetén joggal feltételezzük, hogy p eloszlása közelíthető a normális eloszlással, ezért a konfidencia intervallum szerkesztéséhez felhasználhatjuk a standard normális eloszlás értékeit.

A konfidencia intervallum:

$$p \pm z \times \sigma_p$$

Az előző felmérés adatait ismét használva a 30 fős véletlen mintánk segítségével 95,5 %-os megbízhatóság mellett határozzuk meg a hallgatók hány százaléka magasabb, mint 180 cm

[33] Inverz.stnorm: a standard normális eloszlásból származó kritikus értéket ad eredményül. Inverz.stnorm(α/2) az 1-α megbízhatósághoz tartozó értéket adja.

[34] Inverz.t(valószínűség, szabadságfok):a t-eloszlásból az általunk megadott valószínűség értéket egyből felezi és így adja a kritikus értéket.

$$p = \frac{k}{n} = \frac{7}{30} = 0{,}233$$

$$0{,}233 \pm 2 \times \sqrt{\frac{0{,}267 \times 0{,}767}{30}}$$

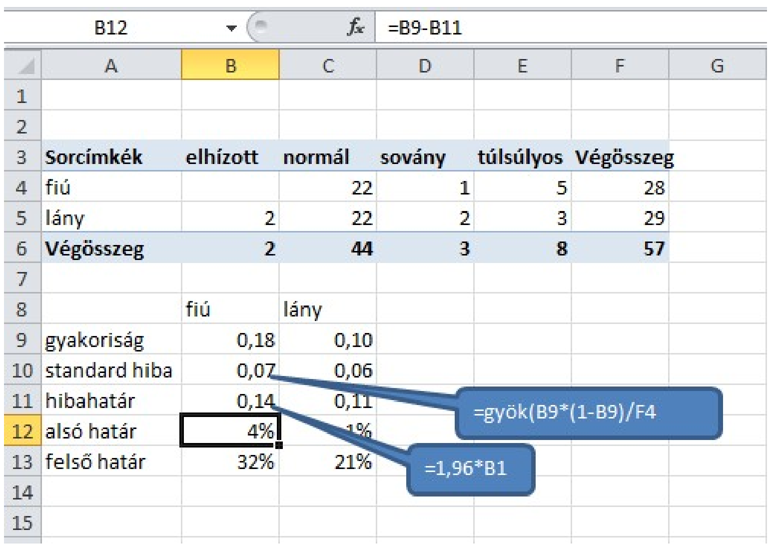
$$0{,}233 \pm 0{,}165$$

6,8% 39,8,%

Tehát 95,5%-os megbízhatósági szint mellett megállapíthatjuk, hogy a hallgatók minimum 6,8%-a és legfeljebb 39,8 %-a magasabb, mint 180 cm.

Egy kicsit más feladat az aránybecslés, hiszen itt a becslőfüggvény a mintabeli relatív gyakoriság - nem előre programozott-, de ha egy új változó képezzünk, akkor visszavezethető az átlagbecslés esetére. Az új változónk értéke legyen 1, ha a testmagasságok magasabbak, mint 180 cm, különben 0. Ezt a „ha” függvénnyel gyorsan elvégezhetjük (ha B2>180;1;0). Ezután a számítás menete megegyezik az előzőekben bemutatottal.

Gyakorló feladatként határozzuk meg 95 %-os megbízhatóság szinten a túlsúlyos fiúk és lányok arányát.


B12 =B9-B11

	A	B	C	D	E	F	G
1							
2							
3	**Sorcímkék**	**elhízott**	**normál**	**sovány**	**túlsúlyos**	**Végösszeg**	
4	fiú		22	1	5	28	
5	lány	2	22	2	3	29	
6	**Végösszeg**	**2**	**44**	**3**	**8**	**57**	
7							
8		fiú	lány				
9	gyakoriság	0,18	0,10				
10	standard hiba	0,07	0,06				
11	hibahatár	0,14	0,11				
12	alsó határ	4%	1%				
13	felső határ	32%	21%				
14							
15							

2/59. képernyőnézet: Az aránybecslés munkatáblája

Ezen kívül az Excel program Adatok menü, Adatelemzés almenüjének, Leíró statisztikai moduljában lehetőség van a hasonló egyszerű becslések gyors elkészítésére. Az ismert modulban egyetlen új beállítást kell alkalmaznunk, csak a várható érték konfidenciaszintjét kell beállítanunk. A beállítások után a következő eredményeket adja a számítógép:

2/44. táblázat: Statisztikai becslés eredménye az Adatelemzés almenü segítségével

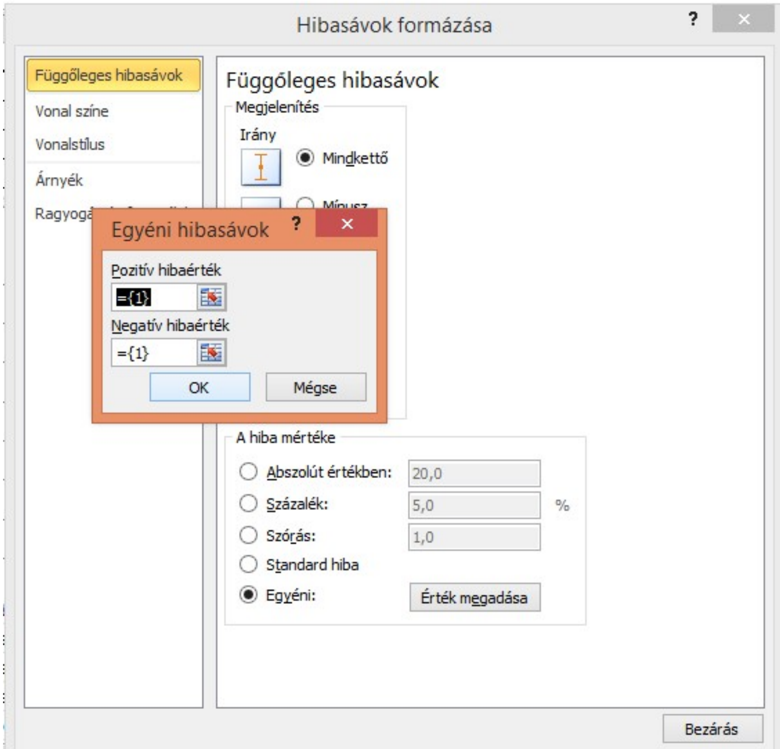
Testmagasság	
Várható érték	174,10
Standard hiba	1,52
Medián	174,50
Módusz	173,00
Szórás	8,34
Minta varianciája	69,59
Csúcsosság	0,31
Ferdeség	-0,48
Tartomány	34,00
Minimum	153,00
Maximum	187,00
Összeg	5223,00
Darabszám	30,00
Konfidenciaszint(95,0%)	3,12

A kis eltérések abból adódnak, hogy a számítógép mindig a megfelelő szabadságfokú t-eloszlásból számítja a kritikus értéket[35], ami nem hiba, hiszen a mintanagyság növelésével a t-eloszlás belesimul a standard normális eloszlásba.

Gyakorló feladatként grafikusan is ábrázoljuk a várható testmagasság értékeket és a megbízhatósági tartományt.

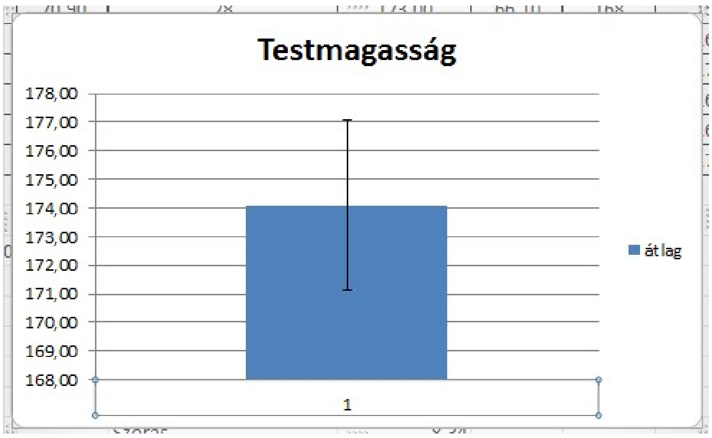
- Jelöljük ki az oszlopok neveit és a hozzájuk tartozó várható értékeket (Ctrl billentyű felhasználásával)
- Ezt követően a Beszúrás menü oszlopdiagram moduljával hozzuk létre a diagramot.
- Ezt követően a Diagrameszközök menü, Elrendezés almenü, Hibasávok moduljának, további hibasávok beállításainál tudjuk az egyéni hibasávok értékeinél (pozitív és negatív hibaértéknél) a hibahatár értékét megadni.

[35] Inverz.t(0,05;29)



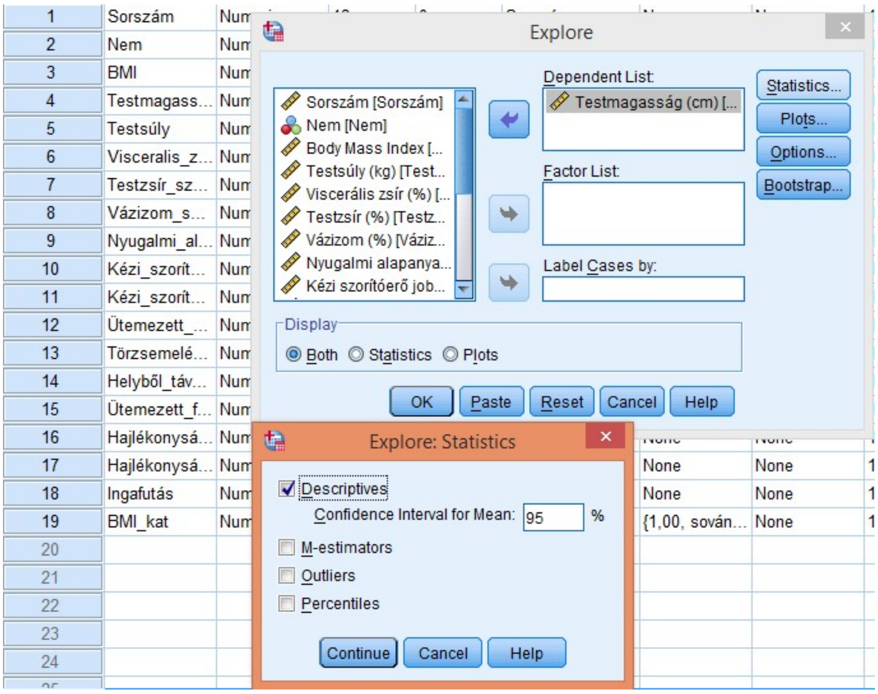
2/60. ábra: Hibahatár beállításai

Miután megadtuk a hibahatárokat (pozitív és negatív értéknél is 2,99), a program elhelyezi az oszlopdiagramon a konfidencia intervallumot.



2/25. ábra: A testmagasság becslésének (95%-os megbízhatósági szinten) grafikus ábrázolása

Az SPSS program is viszonylag gyorsan elvégzi az egyszerű becsléseket. az Analyze menü, Descriptive Statistics, Explore modulban. (forrás: fittségi 57fő_adatbázis_alap_bmikat.sav). Gyakorló feladatként becsüljük meg 95 %-os megbízhatóság mellett, mind az 57 hallgató testmagasság értékét felhasználva a testmagasság várható értékét.



2/61. képernyőnézet: Statisztikai becslés az SPSS program segítségével

A függő változó mezőbe a testmagasság kerüljön, majd a Statistics modulba a leíró statisztikán belül a megbízhatóságot írjuk. Ezt elvégezve kapjuk a következő output táblát:

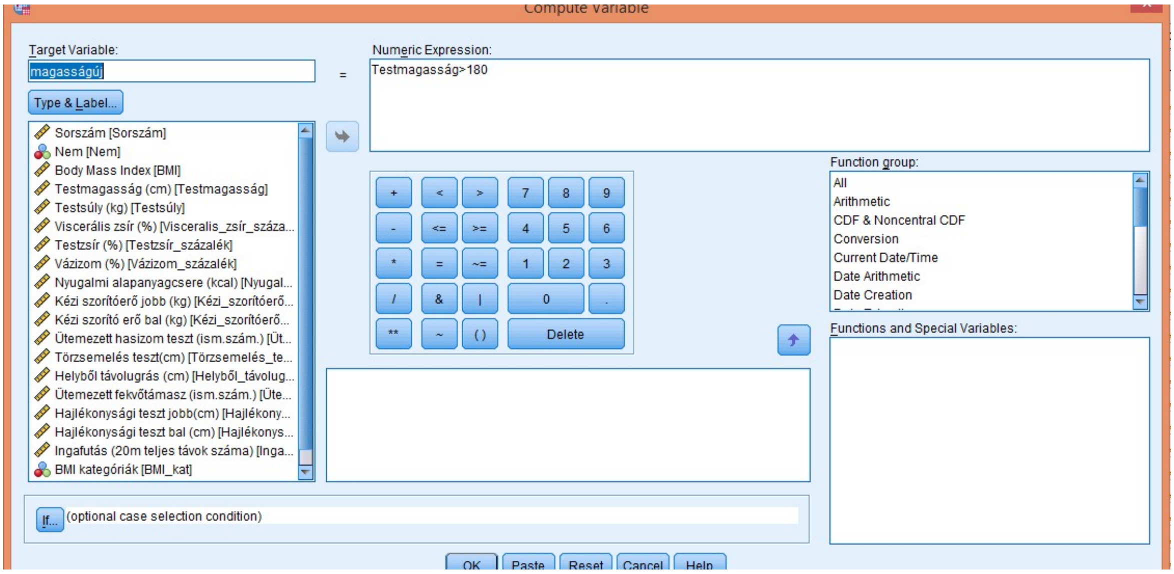
2/45. táblázat

Descriptives

			Statistic	Std. Error
Testmagasság (cm)	Mean		174,2719	1,06619
	95% Confidence Interval for Mean	Lower Bound	172,1361	
		Upper Bound	176,4078	
	5% Trimmed Mean		174,6511	
	Median		173,0000	
	Variance		64,795	
	Std. Deviation		8,04955	
	Minimum		153,00	
	Maximum		188,50	
	Range		35,50	
	Interquartile Range		10,00	
	Skewness		-,469	,316
	Kurtosis		,364	,623

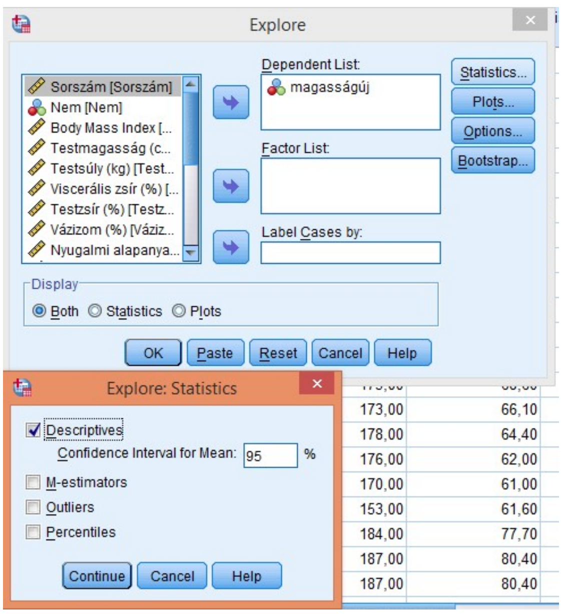
Jól látszik, hogy a program eredményei között minden korábban általunk számolt mutató megtalálható (felső és alsó hibahatár) is.

Az aránybecslésnél egy új változót (magasságúj) kell létrehozni a Transform menü Compute (új változó létrehozása) modullal.



2/62. képernyőnézet: Egy új változó létrehozása az SPSS program segítségével

Itt új változót hozunk létre egy régi, már meglévő változó felhasználásával. Nevezzük el az új változót (magasságúj) a Target Variable dobozban, majd a Numeric Expression dobozba adjuk meg a feltételünket (magasságúj>180). Az így létrejövő új változónál látható, hogy, akik a megadott értéknél (180 cm) magasabbak kaptak 1-est, akik kisebbek 0-át kaptak. Ezt követően a becslés hasonlóan folytatódik, mint az előző feladatnál.



2/63. képernyőnézet: Az aránybecslés beállításai

A beállításoknál már ezt az új változót jelöljük függő változóként, és a megbízhatóságnál is 95%ot válasszunk. Ezt követően az eredmény a következő:

2/46. táblázat

Descriptives

			Statistic	Std. Error
magasságúj	Mean		,2456	,05752
	95% Confidence Interval for Mean	Lower Bound	,1304	
		Upper Bound	,3608	
	5% Trimmed Mean		,2173	
	Median		,0000	
	Variance		,189	
	Std. Deviation		,43428	
	Minimum		,00	
	Maximum		1,00	
	Range		1,00	
	Interquartile Range		,50	
	Skewness		1,214	,316
	Kurtosis		-,546	,623

2.7.4.2. Hipotézisellenőrzés

A következtetéses statisztika egyik leggyakrabban alkalmazott módszereinek összefoglaló neve. A hipotézisellenőrzés (feltevés-vizsgálat) olyan statisztikai módszer, mely alkalmas egy választott statisztikai próba (teszt) segítségével egy-egy feltevés elfogadásáról vagy elvetéséről való döntés meghozatalában. Tehát a feltevések (hipotézisek), egy-egy sokaság jellemzőjét (átlagát, arányát stb.), eloszlási paraméterét (pl. várható érték), az alapsokaság eloszlását (pl. normális eloszlás) tartalmazzák többnyire egzakt matematikai-statisztikai formában. Így lehetővé válik az, hogy a hipotéziseket a matematikai-statisztika eszközeivel, meghatározott valószínűség figyelembevétele mellett ellenőrizzük és végezetül a feltevést elfogadjuk, vagy elvessük.

A hipotézisellenőrzés mindig egy adott hipotézisrendszerre vonatkozik, amely mindig egy nullhipotézisből (H_0) - kiinduló feltevés- és egy vele szemben álló alternatív hipotézisből (H_1) áll. A hipotézisellenőrzés végeredménye mindig egy igen-nem (elfogadom- elvetem) típusú döntés, mely csak egy adott hibavalószínűség mellett érvényes. Egy dolgot feltétlenül tartsunk szem előtt, hogy a döntés kizárólag a nullhipotézisre vonatkozik. A későbbiekben jelöljük Θ-val (theta) az ismeretlen alapsokasági értéket és Θ_0-val a feltételezett értéket!

A sporttudományok területén is gyakran a kísérleti módszer alkalmazásánál valósul meg, ami a gyakorlatban sokszor azt jelenti, hogy valamely új, vagy más módszerről kívánjuk eldönteni, hogy van- e valamilyen hatása a vizsgálat tárgyát képező egyedekre. Két vagy több csoportot képezünk és azt vizsgáljuk, hogy a csoportoknál az eltérő foglalkoztatás (edzésmódszer) eredményeként milyen változás tapasztalható. Például az úszóknál két csoportot képezünk és két különböző edzésmódszerrel foglalkoztatjuk őket (hagyományos, és pl. „Széchy- módszer"). A nullhhipotézis azt mondja, hogy a két csoport eredményei között nincs lényeges eltérés, vagyis a két módszer fejlesztő hatásában nincsen különbség. Kiinduló *nullhipotézisünket* az alábbi módon írhatjuk fel:

$$H_0 \colon \Theta = \Theta_0$$

Természetesen ez a kifejezés önmagában még nem értelmezhető, meg kell fogalmazni ellentét-párját, azaz az *alternatív hipotézis*, így lesz hipotézisrendszer, amely az egész „eseményteret" lefedi. Az előző példánál maradva: a két csoport eredményei között van különbség (nem a véletlen hatása), viszont azt nem tudjuk melyik a jobb, ilyenkor az alternatív hipotézis *kétoldalú*:

$$H_1 \colon \Theta \neq \Theta_0$$

illetve *egyoldalú***,** ha állást foglalunk, hogy az eredmények alapján melyik csoport eredményei jobbak, melyik módszer a jobb :

$$H_1 \colon \Theta \langle \Theta_0$$

vagy

$$H_1 \colon \Theta \rangle \Theta_0$$

A fent megfogalmazott hipotézisek ellenőrzését matematikai függvények ún. *próbafüggvények* segítségével végezhetjük el. A függvény lehetővé teszi az ismert statisztikai eloszlástípusoknak megfelelő elméleti értékkel való összevetést. Egy adott valószínűségi szint ún. **szignifikancia szint** mellett a számított értéket az elméleti értékkel összehasonlítva, a hipotézist vagy elvetjük, vagy elfogadjuk; ezáltal teszteljük az adott alapsokaságra megfogalmazott állításunkat.

A vizsgálat menete így négy lépésben folyik:

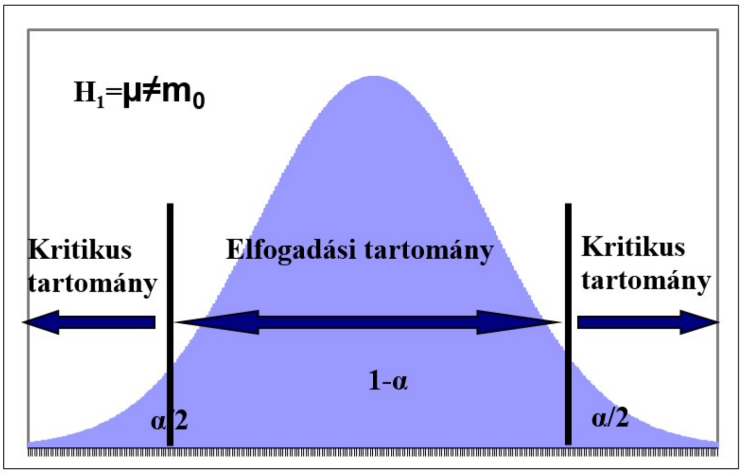
1. Az első lépésben fel kell állítani a hipotézisrendszert (H_0 és a H_1 meghatározása).
2. A megfelelő próbafüggvény kiválasztása.
3. A mintaelemek alapján számított (empirikus) próbafüggvény-érték meghatározása.
4. Döntés.

A próbafüggvény kiválasztása szempontjából fontos az alapsokaság eloszlása, a mintavétel módja és a minta nagysága. Leggyakrabban független azonos eloszlású (FAE) mintát feltételezünk, viszont az alapsokaság eloszlásáról gyakran csak elképzelésünk van (ennek vizsgálata: illeszkedésvizsgálat).

A hipotézisellenőrzés során a próbafüggvény érték-tartományát logikailag két egymást kizáró tartományra osztjuk: az elfogadási tartományra és a kritikus (elutasítási) tartományra. Meg tudjuk mondani, hogy a próbafüggvényünk értéke milyen valószínűséggel esik az elfogadási tartomány határai közé. Döntést úgy hozunk, ha a próbafüggvény a kritikus tartományba esik, akkor a nullhipotézist elvetjük, különben elfogadjuk. A próbafüggvény kritikus (elutasítási) tartományba esésének valószínűségét nevezzük **szignifikancia-szintnek**.

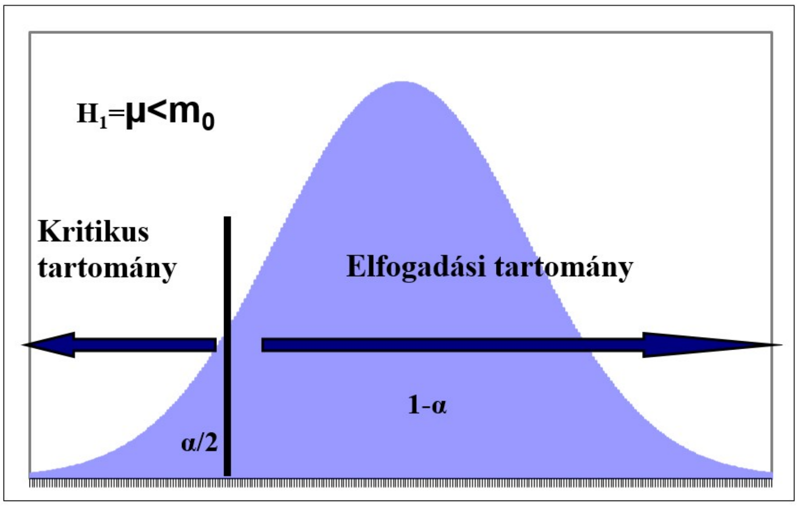
A döntéshozás másik módszere a **szignifikancia- érték (p- érték)** alapján történik, ami azt mutatja meg, hogy az nullhipotézis elvetése milyen valószínűséggel okoz hibát. A döntés, ha a p- érték alacsony szám, akkor kicsi az elsőfajú hiba elkövetésének valószínűsége, ezért célszerű a elutasítani a nullhipotézist. Ez szemben, ha a p- érték nagy elfogadjuk a nullhipotézist. A tartományok elhelyezkedése többféle lehet, mely az alternatív hipotézistől függ. Ha a kritikus tartományba esés valószínűsége α, akkor az elfogadási tartományba esésé: $1-\alpha$.

Tételezzük fel, hogy a hipotézisünk a várható érték (μ) és egy feltételezett érték (m_0) egyenlőségére vonatkozik. Az egyik leggyakrabban alkalmazott hipotézisvizsgálati probléma annak vizsgálata, hogy a sokasági várható érték egy előre adott kontanssal egyezik-e, az ilyen próbát **egymintás várható érték** próbának nevezzük. Ilyenkor egy sokaság várható értékének egy konkrét számmal történő egyezőségét teszteljük, különböző alternatív hipotézisekkel szemben.

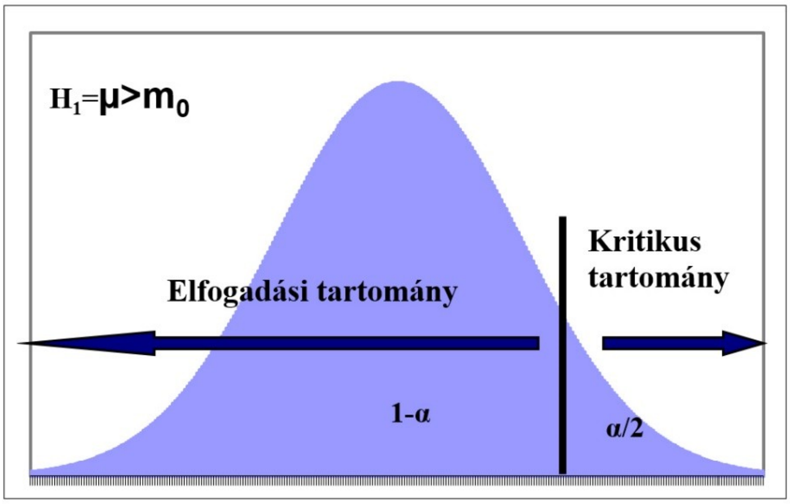


2/26. ábra: Elfogadási és kritikus tartomány kétoldali (two tailed) alternatív hipotézis

A kritikus tartományba esés valószínűsége α, mivel két egyenlő nagyságú részből áll a kritikus tartomány ezért, egyes részekbe α/2 valószínűséggel esik a függvény. Ha a nullpihotézissel szemben azt állítjuk, hogy a várható érték nemcsak, hogy nem egyenlő, hanem nagyobb vagy kisebb, akkor egyoldalas jobb széli (right tailed), vagy bal széli (left tailed) kritikus tartományt kapunk.



2/27. ábra: Elfogadási és kritikus tartomány bal oldali alternatív hipotézis esetén



2/28. ábra: Elfogadási és kritikus tartomány jobb oldali alternatív hipotézis esetén

A próbák leggyakrabban egy- vagy kétmintásnak nevezzük és vonatkozhatnak a sokasági várható értékekre, szórásra, illetve arányra is, ennek megfelelve a leggyakoribb egymintás tesztek próbafüggvényei:

2/47. táblázat

Nullhipotézis	Nagyminta (100≤n)	Kisminta (n<100)
$H_0 : \mu = \mu_0$	$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \mid H_0 \sim N(0;1)$	$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \mid H_0 \sim {}_{n-1}t$
$H_0 : P = P_0$	$z = \frac{P - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \mid H_0 \sim N(0;1)$	
$H_0 : \sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \mid H_0 \sim {}_{n-1}\chi^2$	

Nézzünk most a próbát a gyakorlati alkalmazás során!

Egy szakdolgozat írása során egy hallgató egy külföldi tudományos cikkben[36] (Bai és munkatársai, 2013) azt olvasta, hogy a Tibetből származó kínai egyetemisták BMI indexének átlaga 21,35. A saját adatbázisunk segítségével ismerjük, a mintába került magyar egyetemisták (n=57) átlagos BMI értékét, 23,17 és a korrigált szórását: 3,79 Kíváncsiak vagyunk, hogy van- e a Tibeti és magyar hallgatók BMI értékében különbség!

---

[36] Bai Jingya és társai: Quantitative Analysis and Comparison of BMI among Han, Tibetan, and Uygur University Students in Northwest China. The Scientific World Journal Volume 2013 (2013), Article ID 180863, 6 pages. http://www.hindawi.com/journals/tswj/2013/180863/

Elfogadhatjuk- e, hogy a magyar egyetemisták BMI értéke a kínai egyetemisták mért testtömeg- index eredményét nem haladhatja meg?

$$H_0: \mu=21{,}35$$

$$H_1: \mu>21{,}35$$

A nullhipotézisben tehát azt feltételezzük, hogy a magyar hallgatók BMI értékének várható átlagos értéke megegyezik a kínai egyetemisták várható testtömeg index értékével (a kínai átlagtól való eltérés csak véletlen tényezőknek tekinthető). Az alternatív hipotézisben pedig azt fogalmazzuk meg, hogy ez az érték nagyobb lehet 21,35 -nál, amiből arra következtethetünk, hogy az eltérés valamilyen szisztematikus dologgal magyarázható, ami a testsúlyra fejt ki kedvezőtlen hatást[37] (pl: egészségtelen táplálkozás túl nagy volumene, vagy a mozgásszegény életmód).

A gyakorlati esetek során legtöbbször nem áll módunkban nagy elemű minta segítségével a hipotéziseinket ellenőrizni, hanem kis mintával kell dolgoznunk. Kis minta esetén a standard normális eloszlás nem alkalmazható, ilyenkor a **Student-féle t-eloszlás**t és ennek az eloszlásnak a táblázatát kell alkalmaznunk. A t-eloszlás alkalmazása során figyelembe kell venni az ún. **szabadságfok**ot, amely a minta elemszámának 1-gyel csökkentett értéke. Egy adott rendszer szabadságfokán azt a számot értjük, ahány érték ebben a rendszerben szabadon megválasztható (t- és χ^2 – eloszlás esetén egy, F- eloszlásnál két szabadságfokot határozunk meg).

$$t = \frac{23{,}17 - 21{,}35}{\frac{3{,}79}{\sqrt{57}}} = 3{,}62$$

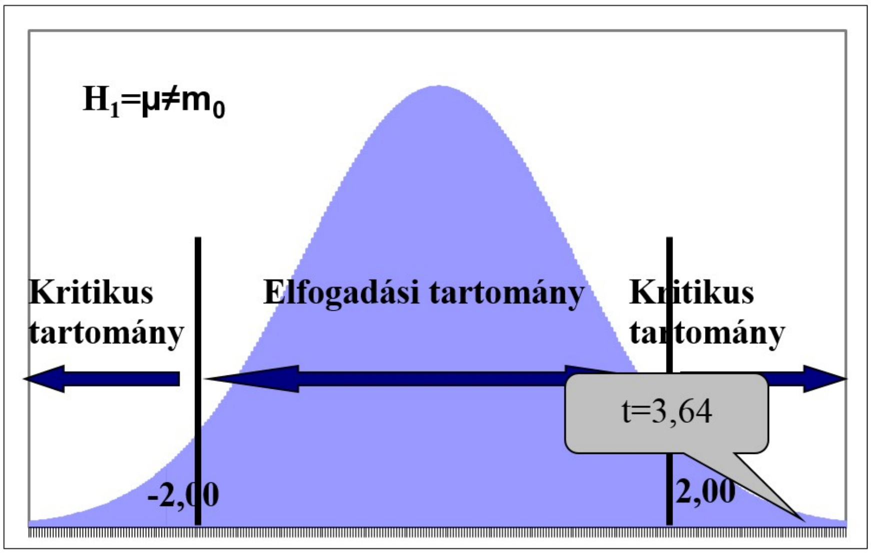
A t-eloszlás kritikus értéke 5 %-os szignifikancia szinten 56 szabadságfok (n-1) mellett a t-eloszlást tartalmazó táblázatból olvasható:1,67

Mivel a számított értékünk nagyobb, mint a táblabeli, ezért elutasítjuk a nullhipotézist (nincs okunk rá, hogy elfogadjuk), tehát elfogadjuk az alternatív hipotézist (a grafikus ábrán az elutasítási tartományban jelenik meg). A mintabeli és az elvárt érték közötti eltérést – 5 %-os szignifikancia szinten – valamely szisztematikus tényező okozhatta.

[37] Ezt a feltételezést szakértői véleményünkre alapozzuk, melynek irodalmi kutatásától jelen esetben eltekintünk.

Ha csak arra lennénk kíváncsiak, hogy az általunk mért érték megegyezik-e vagy sem a tibeti hallgatók átlagértékével, akkor kétoldalú hipotézis-ellenőrzést hajtanánk végre. Ilyenkor az alternatív hipotézis a következő lesz:

$$H_1: \mu \neq 23{,}17$$



2/29. ábra: A döntést segítő grafikus ábra kétoldalú alternatív hipotézis esetén

A fenti alternatív hipotézis esetén a sűrűségfüggvény mindkét oldalát figyelembe kell venni, így a kritikus érték (5 %-os szignifikancia szinten): ±2,00. Ehhez viszonyítva is el kell vetni a nullhipotézist.

Hasonlóan kell eljárnunk, ha nem az alapsokasági átlagra, hanem az alapsokasági arányra vonatkozóan fogalmazunk meg feltevést. Itt is meg kell jegyeznünk, hogy csak nagy minta esetén használható a tesztelésre a standard normális eloszlás.

$$z = \frac{p - P_0}{\sqrt{\dfrac{P_0(1 - P_0)}{n}}}$$

Az asszociációs kapcsolatok során kontingencia táblázatba rendeztük a hallgatókat a BMI kategóriákba tartozás vonatkozásában.

2/48. táblázat: BMI kategóriák a nemek vonatkozásában

Nem	BMI kategóriák				
	elhízott	normál	sovány	túlsúlyos	Összesen
fiú	0	22	1	5	28
lány	2	22	2	3	29
Összesen	2	44	3	8	57

A hallgatók közül 10 fő tartozik a túlsúlyos és elhízott kategóriába.

Vizsgáljuk meg, hogy elvárható-e az egyetemista populáció vonatkozásában a 15%-os túlsúlyos és elhízott arány.

$$H_0: P=0{,}15$$

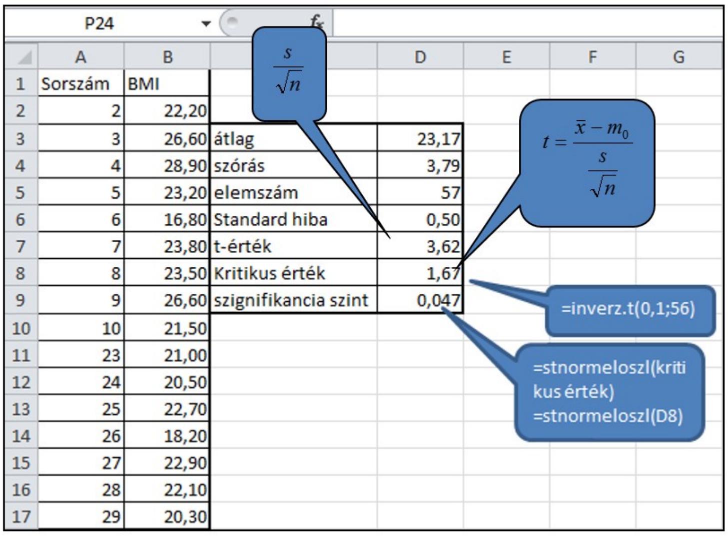
$$H_1: P<0{,}15$$

Az alternatív hipotézisben azt a feltevést fogalmaztuk meg, amely szerint 15%-nál kisebb lesz a túlsúlyos és elhízott arány.

$$p = \frac{10}{57} = 0{,}176$$

$$z = \frac{0{,}176 - 0{,}15}{\sqrt{\frac{0{,}15 \times (1 - 0{,}15)}{57}}} = 0{,}55$$

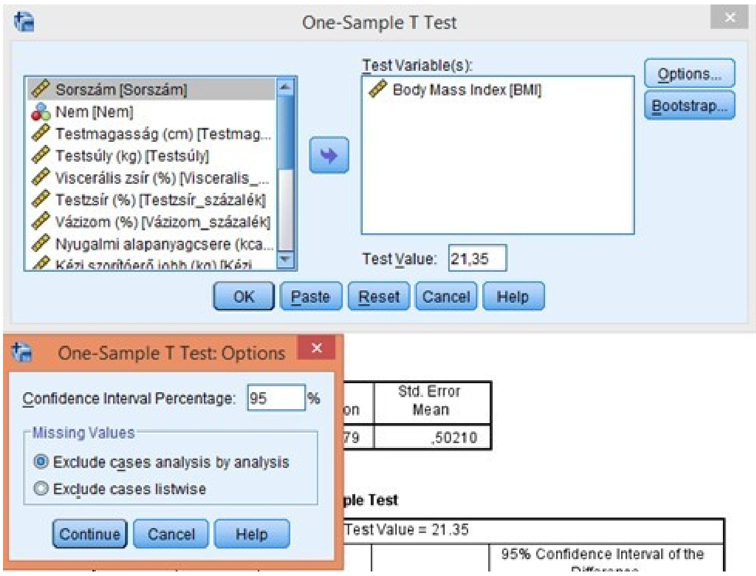
A táblabeli érték (a negatív oldalt figyelembe véve) -1,645. Az empirikus érték az elfogadási tartományba esik, tehát elfogadjuk a nullhipotézist, vagyis az egyetemisták körében a túlsúlyos és elhízott arány 15%-nál magasabb lesz.



2/64. képernyőnézet: A számítás menete az Excel programban

Mivel kis mintánk van, ezért t-eloszlás kritikus értékét kell meghatározni, azonban az Excel mindig kétoldali kritikus tartománnyal dolgozik, vagyis felezi a tartományt. A kritikus tartományt az inverz.t függvénnyel számítjuk, ha 5%-os szignifikancia szintet szeretnénk, akkor 10%-os értéket adunk meg.

Az SPSS programmal viszonylag egyszerűbben jutunk eredményhez. Az Analyze menü, Compare Means almenü, One- Sample T Test modulja segítségével. (Forrás: Egymintás t-próba alapadatai.sav)



2/65. képernyőnézet: Az egymintás t- próba az SPSS programmal (BMI)

Először válasszuk ki a változónkat, majd a Test Value dobozba BMI írjuk a nullhipotézisben megadott, az összehasonlítás alapjául szolgáló konstans értéket (21,35). Ezután az Options dobozba megadhatjuk a szignifikancia szintet. Ezt követően az Ok gomb megnyomásával jutunk el a végeredményhez az Output View-ban:

A végeredmény két táblából áll. Az elsőben a leíró statisztikai adatokat láthatjuk, melyeket az Excelben is megkaptunk.

2/49. táblázat

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Body Mass Index	57	23,1684	3,79079	,50210

A második tábla a hipotézis eldöntését segíti.

2/50. táblázat

One-Sample Test

	Test Value = 21.35					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Body Mass Index	3,622	56	,001	1,81842	,8126	2,8243

Látható, a számított t-érték, a szabadságfok, viszont nem látjuk a kritikus értéket. A döntést azonban a következő adat szolgáltatja, amely azt mutatja, hogy a t számított értéke milyen szinten szignifikáns. Általában a Sig=5 % (0,05) alatt a nullhipotézist elutasítjuk, így most a nullhipotézist elvetjük. A konfidencia- intervallum azokat a határértékeket mutatja, melyek közé a követelménytől vett eltérések értékei 95%-os valószínűséggel esnek.

Gyakorlatban nagyon sokszor fordul elő, hogy két különböző sokaságból veszünk véletlen és független mintát. Ilyen esetekben a két sokaság ugyanazon paramétereit hasonlítjuk össze, teszteljük különbségeiket, azonosságukat. A gyakorlati alkalmazások során számtalanszor találkozunk a két alapsokasági várható érték egyezőségének, minta alapján történő tesztelésével. Ahogyan már eddig is megszokhattuk, az állítást általánosságban nullhipotézisben, konkrét formában az alternatív hipotézisben található.

$$H_0: \mu_1 - \mu_2 = \delta$$

Az alternatív hipotézis különböző módon történő felírása lehetővé teszi a várható értékek nagyságrendi relációiról történő döntést is, vagyis:

$H_1: \mu_1-\mu_2<\delta$; $H_1: \mu_1<\mu_2$ (bal oldali)

$H_1: \mu_1-\mu_2>\delta$; $H_1: \mu_1>\mu_2$ (jobb oldali)

$H_1: \mu_1-\mu_2\neq\delta$; $H_1: \mu_1\neq\mu_2$ (két oldali)

Most nézzük a legelterjedtebb **kétmintás t- próbát**, melynek két előfeltétele van: mindkét sokaság eloszlása legyen normális (külső, egyéb információ szükséges), illetve az alapsokasági szórásnégyzetek legyenek egyenlők.

Ha normális eloszlású alapsokaságból vett kis mintánk van, és a sokasági szórás ismeretlen, de feltételezzük az azonosságot, akkor egyszerűen t- próbát használunk (Nem követünk el nagy hibát azonban, ha nagyobb minta esetén is ezt a próbát[38] használjuk):
Az alkalmazható próbafüggvény:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

A szabadságfok: $n_1 + n_2 - 2$

Ebben az ún. közös szórás (s_p) négyzetének képlete:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$

Amennyiben az alapsokasági szórásnégyzeteket nem ismerjük, akkor teszteljük. Ismeretes, hogy a mintabeli korrigált varianciák hányadosa –valószínűségi változó-, ha az alapsokasági varianciák egyenlő F-eloszlást követ, n_1-1;n_2-1 szabadságfok- párral.

$$F = \frac{s_1^{\,2}}{s_2^{\,2}} \Big| H_0 \approx F_{n_1 - 1; n_2 - 1}$$

A két várható érték különbözőségét a következő példán keresztül értelmezzük. Vizsgáljuk meg, hogy a lányok és fiúk testmagasságában 5%-os szignifikancia szinten van-e meghatározó különbség. A lányok (29 fő) testmagasságának átlaga 169,62 cm szórása 7,49, míg a fiúk (28 fő) testmagasságának átlaga 179,04 cm. szórása 5,37.

Mivel kétmintás próbáról van szó, először megvizsgáljuk, hogy a szórások egyenlőnek tekinthetők-e:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$F = \frac{7{,}49^2}{5{,}37^2} = 1{,}94$$

Korábban alkalmazott logikánk értelmében, kétoldalú próba és 5%-os szignifikancia-szint esetén a felső kritikus értéket a 2,5%-os szignifikancia értéknél olvastuk le, majd ebből határoztuk meg az „alsó” kritikus értéket. F-eloszlás esetén kicsit nehezebb a dolgunk,

[38]Amennyiben a t-eloszlás nagyobb szabadságfokú értékeit a standard normális eloszlás hasonló adataival összevetjük, szembetűnő a hasonlóság.

ugyanis – abból adódóan, hogy az F-eloszlás nem szimmetrikus és csak a pozitív tartományon értelmezett – az alsó kritikus érték meghatározása a következő képlet alapján történik:

$$_{n;m}F_{1-\alpha} = \frac{1}{_{m;n}F_{\alpha}}$$

Esetünkben a két kritikus érték: (az első a táblából, a második számítva):

$$_{28;27}F_{0,025} = 2{,}15$$

$$_{27;28}F_{0,975} = \frac{1}{_{11;12}F_{0,025}} = \frac{1}{2{,}18} = 0{,}46$$

Tehát elfogadjuk (5%-os szignifikancia szint mellett) a varianciák egyezőségét, hiszen az általunk kapott érték az elfogadási tartományba (a két kritikus érték közé) esik. Ezután következik a kétmintás t-próba:

$$H_0{:}\mu^1 = \mu^2$$

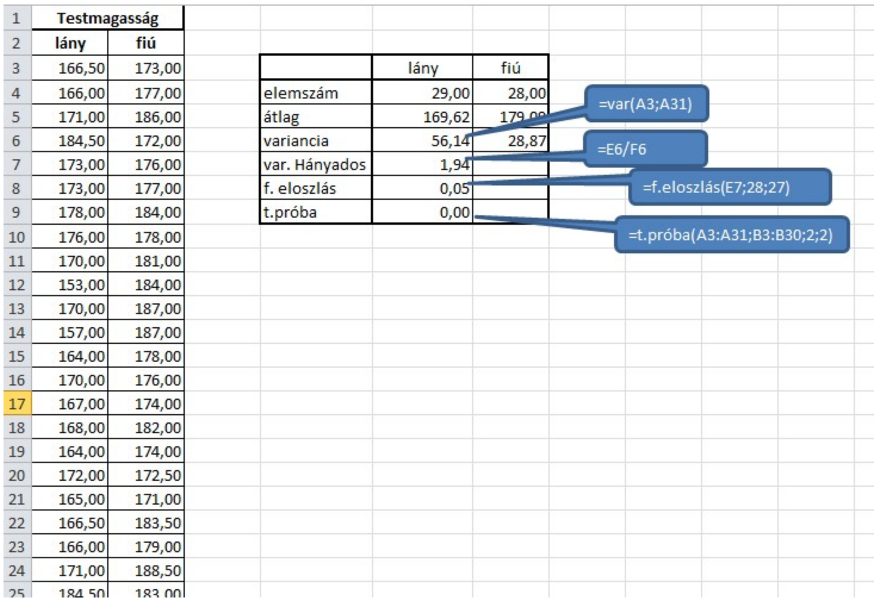
$$H_1{:}\ \mu^1 \neq \mu^2$$

$$s_p^2 = \frac{(29-1)\times 7{,}49^2 + (28-1)\times 5{,}37^2}{29+28-2} = 42{,}75$$

$$s_p = \sqrt{42{,}75} = 6{,}54$$

$$t = \frac{169{,}62 - 174{,}04}{6{,}54\times\sqrt{\frac{1}{29}+\frac{1}{28}}} = -5{,}5$$

A t-eloszlás táblabeli értéke 55-mas szabadságfok esetén, mivel kétoldali a hipotézisünk, 2,00. A számított érték az elutasítási tartományba esik, tehát 5-os szignifikancia szint mellett a nemek testmagassága szignifikánsan különbözik.

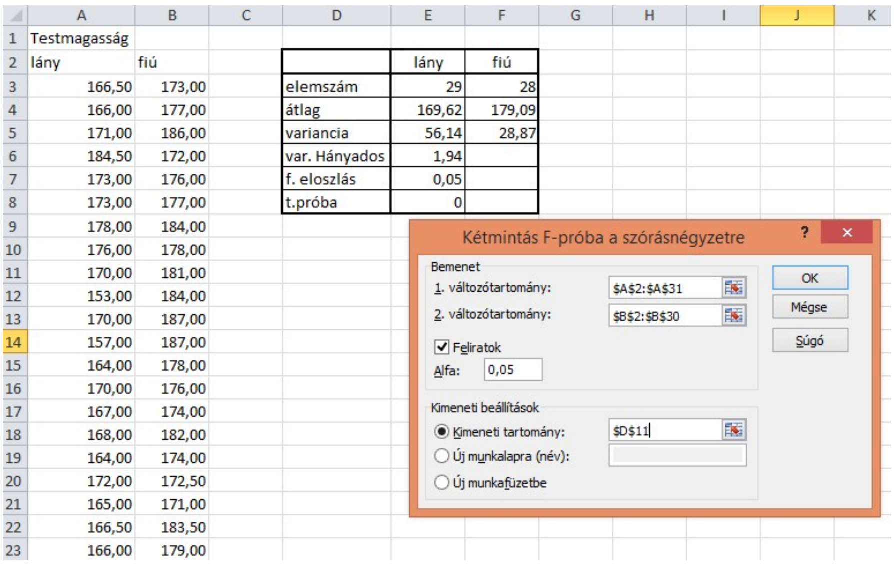


	Testmagasság	
2	lány	fiú
3	166,50	173,00
4	166,00	177,00
5	171,00	186,00
6	184,50	172,00
7	173,00	176,00
8	173,00	177,00
9	178,00	184,00
10	176,00	178,00
11	170,00	181,00
12	153,00	184,00
13	170,00	187,00
14	157,00	187,00
15	164,00	178,00
16	170,00	176,00
17	167,00	174,00
18	168,00	182,00
19	164,00	174,00
20	172,00	172,50
21	165,00	171,00
22	166,50	183,50
23	166,00	179,00
24	171,00	188,50
25	184,50	183,00

	lány	fiú
elemszám	29,00	28,00
átlag	169,62	179,09
variancia	56,14	28,87
var. Hányados	1,94	
f. eloszlás	0,05	
t.próba	0,00	

2/66. képernyőnézet: két mintás t-próba számítása

Az Excel programban az Eszközök menü, Adatelemzés almenüjének segítségével is két lépésben hajtható végre a kétmintás t-próba, hiszen először az előfeltételt kell tesztelnünk (Kétmintás F-próba a szórásnégyzetekre).



	A	B
1	Testmagasság	
2	lány	fiú
3	166,50	173,00
4	166,00	177,00
5	171,00	186,00
6	184,50	172,00
7	173,00	176,00
8	173,00	177,00
9	178,00	184,00
10	176,00	178,00
11	170,00	181,00
12	153,00	184,00
13	170,00	187,00
14	157,00	187,00
15	164,00	178,00
16	170,00	176,00
17	167,00	174,00
18	168,00	182,00
19	164,00	174,00
20	172,00	172,50
21	165,00	171,00
22	166,50	183,50
23	166,00	179,00

	lány	fiú
elemszám	29	28
átlag	169,62	179,09
variancia	56,14	28,87
var. Hányados	1,94	
f. eloszlás	0,05	
t.próba	0	

2/67. képernyőnézet: Az előfeltétel (kétmintás F-próba) tesztelése

A változótartományokba a vizsgálni kívánt csoportok adatait választottuk (felirattal), ennek megfelelően a feliratok dobozt is jelöltük, majd a kimeneti tartomány helyét határoztuk meg. Ennek eredményeként a következő számított adatokhoz jutottunk:

$$F = \frac{s_1^{\,2}}{s_2^{\,2}}$$

Kétmintás F-próba a szórásnégyzetr		
	lány	*fiú*
Várható érték	169,62	179,09
Variancia	56,14	28,87
Megfigyelések	29,00	28,00
df	28,00	27,00
F	1,94	
P(F<=f) egyszél	0,05	
F kritikus egysz	1,94	

2/68. képernyőnézet: Az előfeltétel tesztelésének eredményei

Látszólag meglepő, hogy az Excel egyoldalú próbát hajt végre, hiszen tankönyvünkben mi is kétoldalú próbát (nem egyenlő ellenhipotézist) sugalltunk, ám a program „takarékosan” jár el, kihasználja, hogy az F-eloszlású próbafüggvény esetén az egyik kritikus érték 1-nél nagyobb, a másik kisebb, így az empirikus próbafüggvény-érték nagyságrendje alapján kiválasztja, hogy az alsó, vagy a felső kritikus érték a releváns.

Döntésünk tehát úgy történik, hogy amennyiben a számított F-értékünk az Excel által megadott kritikus érték és 1 közé esik, akkor a nullhipotézist elfogadjuk, ellenkező esetben (túl kicsi, vagy túl nagy F-érték esetén) elvetjük[39]. Jelen esetben határértéken van a számítás, de most a szórásnégyzeteket egyezőnek tekintjük, így elvégezhetjük a kétmintás t-próbát egyenlő szórásnégyzeteknél (Ha nem lennének egyenlőek akkor is itt, az adatelemzés menüből kellene kiválasztani a kétmintás-t próba nem- egyenlő szórásnégyzeteknél nevű modult). A beállítások megegyeznek az F-próbánál bemutatottal.

Kétmintás t-próba egyenlő szórásnégyzeteknél		
	lány	*fiú*
Várható érték	169,62	179,09
Variancia	56,14	28,87
Megfigyelések	29,00	28,00
Súlyozott variancia	42,75	
Feltételezett átlagos eltérés	0,00	
df	55,00	
t érték	-5,47	
P(T<=t) egyszélű	0,00	
t kritikus egyszélű	1,67	
P(T<=t) kétszélű	0,00	
t kritikus kétszélű	2,00	

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}$$

t

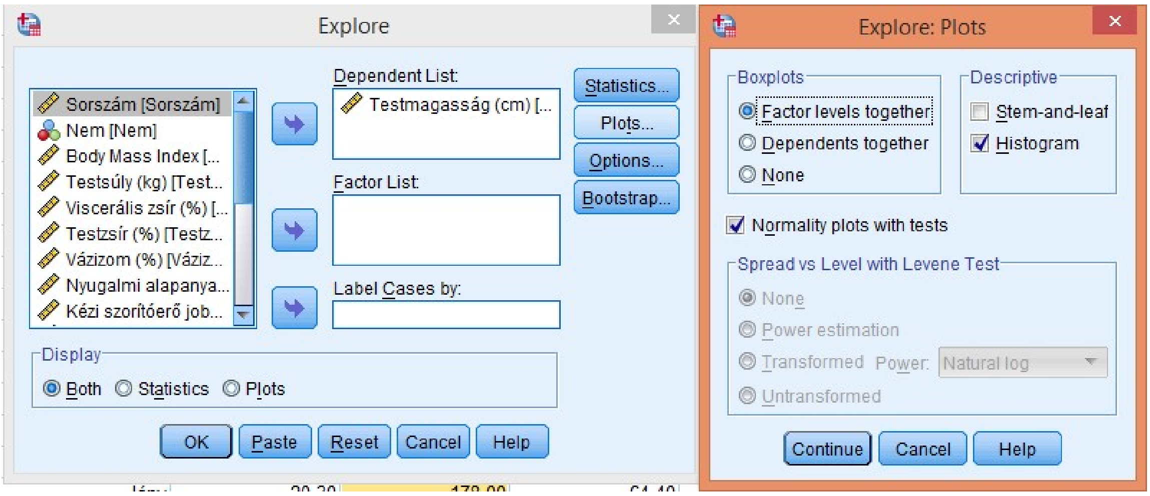
=inverz.t(05;55)

2/69. képernyőnézet: A kétmintás t-próba eredményei az Excel programban

[39] Pintér- Rappai 2007, 385.o.

Egyező eredményeket kaptunk, vagyis el kell utasítani a nullhipotézist. Ha a p-érték (szignifikancia- érték) nyomán döntünk, úgy a kapott maximális szignifikancia- szint, ameddig a nullhipotézist el kell fogadni (0,00), vagyis minden bizonnyal az elutasítás mellett döntünk.

Az SPSS program segítségével a kutatónak kényelmes dolga van, hiszen könnyen számítható az eljárás. Az előfeltételek közül első lépésben vizsgáljuk meg, hogy a BMI index normális eloszlást követ-e. A normalitásvizsgálat alkalmas arra, hogy két valószínűségi változó eloszlását összehasonlítsuk, vagy ellenőrizzük, hogy egy valószínűségi változónak az eloszlása, az általunk feltételezett eloszlásból (normális eloszlásból) származik-e.
A normalitás vizsgálatának léteznek kvantitatív (numerikus) és grafikus módszerei. A grafikus módszerek közül a leggyakrabban a normál eloszlás görbéjét tartalmazó hisztogrammal, valamint a kvantilis- kvantilis (Q-Q Plot) ábrázolással ismerkedünk meg.
A grafikus módszerek ábrája alapján akkor tekinthetünk egy változó normális eloszlásúnak, ha a változó eloszlásának alakja jól lefedi a normál eloszlás haranggörbéjét (Histogramm), valamint a Q-Q ábrán a hipotetikus normál eloszlás egyenesére is jól illeszkednek az adatok.
A numerikus módszerek közül a Kolmogorov- Szmirnov, valamint Shapiro-Wilk teszteket mutatjuk be. Ez utóbbit akkor érdemes alkalmazni, amikor viszonylag kis minta áll rendelkezésünkre, vagyis a minta elemszáma kevesebb, mint 50. Amennyiben a számított szignifikancia érték magasabb, mint 5 %, akkor elmondható, hogy a változó normál eloszlást követ. Adattranszformáció segítségével gyakran sikerül a normális eloszlástól eltérő változót normális eloszlásúvá alakítani.
A grafikus és numerikus vizsgálatokat megtaláljuk az ANALYZE menü DESCRIPTIVE STATISTICS menüpontjának EXPOLRE opciójában, illetve a Kolmogorov- Szmirnov tesztet az ANALYZE menü LEGACY DIALOGS elmenüjének 1-SAMPLE K-S opciójából indíthatjuk. Ez utóbbi esetben a TEST VARIABLES dobozba a testmagasság érték változót mozgassuk, majd az OPTIONS gombra kattintva kérjük a leíró statisztikát (DESCRIPTIVE). Ezt követően a CONTINUE és OK gombok lenyomása következik. Most részletesen nézzük meg a normalitásvizsgálat kvantitatív és grafikus beállításait (ANALYZE/DESCRIPTIVE STATISTICS/EXPLORE), eredményeit.



2/70. ábra: Normalitásvizsgálat beállításai

A függő változó dobozban a felső nyíl segítségével a BMI változót mozgassuk, majd a Plots gombra klikkelve a jobb oldali dobozban (DESCRIPTIVE) válasszuk a HISTOGRAM ábrát, valamint pipáljuk ki a NORMALITY PLOTS WITH TESTS négyzetet. Ezt követően nyomjuk meg a CONTINUE gombot. Miután további beállításokra most az alapbeállításokon túl nincsen szükségünk, nyomjuk meg az OK gombot.

Ezt követően az OUTPUT nézetben több számítási eredményeket tartalmazó táblázat és ábra is megjelenik, melyek közül, most a számunkra relevánsakat ismertetjük.

2/51. táblázat: A normalitás vizsgálat numerikus eredményei

Tests of Normality

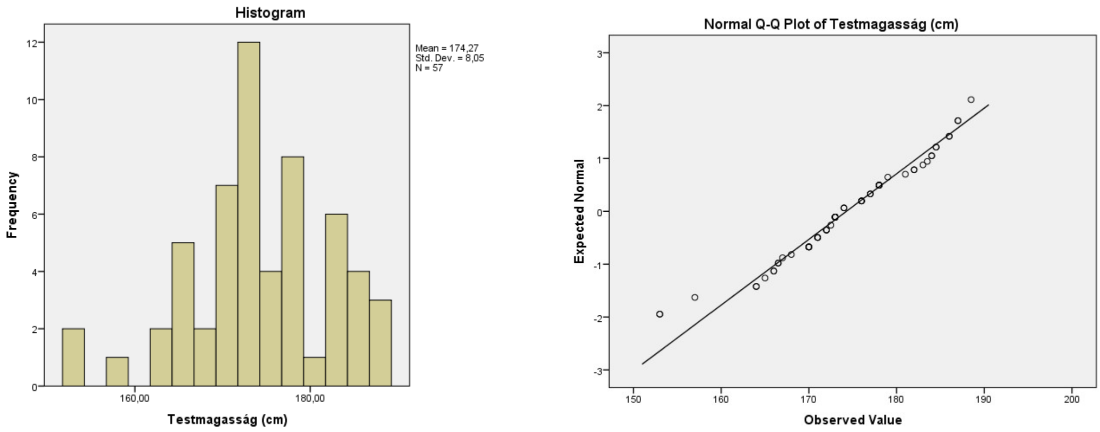
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Testmagasság (cm)	,087	57	,200*	,964	57	,087

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

a. Lilliefors Significance Correction

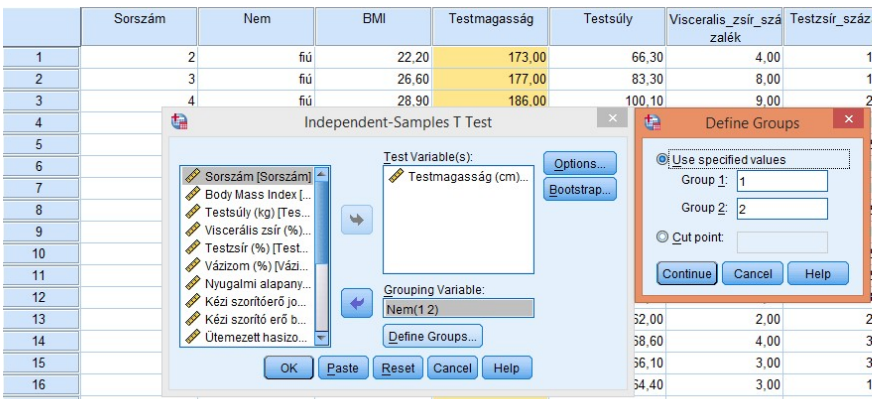
Látható, hogy a mindkét tesztnél (Kolmogorov- Szmirov és a Shapiro- Wilk) a szignifikancia értékek magasabbak, mint 0,05, így a normalitás teljesül.



2/30. ábra: A testmagasság normalitás vizsgálatának grafikus ábrái (hisztogramm, Q-Q ábra)

A grafikus ábrák megerősítik, hogy a változó eloszlása normálisnak nevezhető, hiszen a hisztogram ábrája lefedi a haranggörbét, valamint a Q-Q ábra elemei illeszkednek az egyenesre. Összegezve megállapítható, hogy a testmagasság változó a normális eloszlást követi (p<0,05), vagyis a paraméteres eljárásokat lehet alkalmazni, így a kétmintás- t próba alkalmazható a hipotézis eldöntésére.

Az ANALYZE főmenü, COMPARE MEANS almenü, INDEPENDENT- SAMPLES T TEST modulban hajtható végre a művelet. Az első lépésben a tesztváltozó dobozba a testmagasság értékek kerüljenek, majd a csoportosító változónak megadjuk nem változónkat. Itt jelölni kell az ismérvváltozatokat is, vagyis hogy milyen kódokkal (1 és 2) jelöltük a két nemet.



2/71. képernyőnézet: A kétmintás- t próba beállításai

A beállításokat követően a következő végeredményhez jutunk, melynek első táblázata ismét a leíró statisztikai adatokat tartalmazza:

2/52. táblázat: A leíró statisztikai adatok

Group Statistics

	Nem	N	Mean	Std. Deviation	Std. Error Mean
Testmagasság (cm)	fiú	28	179,0893	5,37321	1,01544
	lány	29	169,6207	7,49244	1,39131

A kétmintás t- próba előfeltétele a szórások azonossága, melyet az SPSS a Levene (Levin) teszttel ellenőriz. A sajátos F-próbaként felfogható teszt értékelése megegyezik az eddig tárgyaltakkal, hiszen itt is a nullhipotézis arra vonatkozik, hogy a szórások azonosak. Mivel a megfigyelt szignifikanciaszint értéke nagyobb, mint 0,05, ezért a szórások azonosságára vonatkozó nullhipotézist elfogadjuk, ilyenkor a táblázat felső sorát kell nézni. Ha elvetnénk a nullhipotézist (a szórások nem egyenlők), akkor a második sort lenne szükséges elemezni. A program kétszélű próbát végez, alternatív hipotézist ellenőriz.

2/53. táblázat

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Testmagasság (cm)	Equal variances assumed	,631	,430	5,466	55	,000	9,46860	1,73236	5,99688	12,94031
	Equal variances not assumed			5,497	50,821	,000	9,46860	1,72246	6,01032	12,92688

A további eredmények megegyeznek a már tárgyaltakkal, tehát a nullhipotézist elvethetjük. A t értéknél szereplő pozitív előjel arra enged következtetni, hogy a vizsgálatba szereplő első ismérvváltozat átlaga a magasabb, vagyis a fiúk szignifikánsan magasabb testmagasággal rendelkeznek, mint a lányok és ez nem a véletlennek köszönhető.

Amennyiben az előfeltételek során a normalitás sérülne, abban az esetben nem tudjuk a kétmintás t- próbát alkalmazni, ilyenkor nem- paraméteres vizsgálatok közül a Mann-Whitney próbát szükséges alkalmazni.

A következőkben betekintést nyerünk a **variancia-analízis** módszerébe, mely kettőnél több sokaság esetén is alkalmazható. A módszer segítségével megkíséreljük egy vagy több minőségi ismérv alapján képzett részmintákban a kiválasztott mennyiségi ismérv szerinti különbözőségét számszerűsíteni. A variancia- analízis (**An**alysis **O**f **Va**riance=Anova) célja az átlagok összehasonlítása, viszont eszköze a varianciák vizsgálata. A varianciaanalízis feltételezi, az alapsokaságon és valamennyi csoporton (részsokaságon) belül a mennyiségi ismérv normális eloszlását. A módszer másik előfeltétele: a varianciahomogenitás, vagyis a csoportok szórásai azonosak (egyenlők) legyenek.

A módszer alkalmazásának három legtipikusabb területe:

1. kettőnél több (rész) sokaság várható értékének egyezőségére vonatkozó hipotézis ellenőrzése;
2. homogenitás-vizsgálat;
3. vegyes kapcsolat (kvalitatív és kvantitatív változó közötti kapcsolat) szignifikáns voltának tesztelése.

A variancia- analízis modellje: $x_{ji} = \mu + \tau_j + \varepsilon_{ji}$

ahol a j-edik csoport i-edik eleme (x_{ji}), a teljes sokaságra vonatkozó várható érték (μ), a j-edik osztály csoporthatása (τ_j) és az ε_{ji} véletlen hatás összegeként adódik. A vizsgálat során a következő hipotézisrendszert teszteljük:

$$H_0 : \mu_1 = \mu_2 = \ldots\mu_m = \mu$$
$$H_1 : \mu_j \neq \mu$$

A nullhipotézis elfogadása a várható értékek egyezőségének, a részekre bontott sokaság homogenitásának, valamint a vegyes kapcsolat hiányának (függetlenség) kimondását jelenti. A csoportosított sokaságra vonatkoztatva, egy adott mintáról elmondható, hogy háromféle átlagtól vett eltérés számítható, mely az alábbi összefüggésből keletkezik:

$$\sum\sum(x_{ij} - \mu)^2 = \sum n_j(\mu_j - \mu)^2 + \sum\sum(x_{ij} - \mu_j)^2$$

,ahol a képlet a teljes eltérés- négyzetösszeget felbontja külső (csoportok közötti), illetve belső (csoportokon belüli) eltérés- négyzetösszegekre.

2/54. táblázat

Eltérés- négyzetösszeg	**Eltérés- négyzetösszeg típusa**
$SS = \sum_{j=1}^{m} \sum_{i=1}^{nj} (x_{ji} - \overline{x})^2$	Teljes eltérés- négyzetösszeg
$SS_K = \sum_{j=1}^{m} (\overline{x}_j - \overline{x})^2$	Külső eltérés- négyzetösszeg
$SS_B = \sum_{j=1}^{m} \sum_{i=1}^{nj} (x_{ji} - \overline{x}_j)^2$	Belső eltérés- négyzetösszeg

Az eltérésnégyzet- összegekből képezhető próbafüggvény F eloszlást követ, ahol a számláló szabadságfoka m-1 (m a csoportok száma), a nevező szabadságfoka n-m (n a sokaság tagszáma). A próbafüggvény, egyoldalú nagyobb alternatív hipotézist feltételezve alkalmas a variancia- analízis végrehajtására, vagyis ha F számított értéke nagyobb, mint a kritikus érték, akkor a nullhipotézist elvetjük.

$$F = \frac{\frac{SS_K}{m-1}}{\frac{SS_B}{n-m}}$$

A döntést segítendő összefüggéseket a következő táblában olvashatjuk:

2/55. táblázat

Tényezők	Eltérésnégyzet-összeg (SS)	Szabad-ságfok (df)	Átlagos eltérés-négyzet-összeg(MS)	F
Csoportok között	$SS_K = \sum_{j=1}^{m} n_j(\bar{x}_j -$	m-1	$\frac{SS_K}{m-1}$	$\frac{MS_K}{MS_B}$
Csoporton belül	$SS_B = \sum_{j=1}^{m}\sum_{i=1}^{n_j}(x_{ji} -$	n-m	$\frac{SS_B}{n-m}$	
Összesen	$SS = \sum_{j=1}^{m}\sum_{i=1}^{n_j}(x_{ji} -$	n-1	$\frac{SS}{n-1}$	

A következőkben arra vagyunk kíváncsiak, hogy van-e különbség a különböző BMI kategóriákban lévő távolugrás értékek között! A fenti kérdés megválaszolásához szolgáló véletlen mintavétel eredményeit az alábbiakban közöljük:

D	E	F	G
sovány	normál	túlsúlyos	Elhízott
242,00	240,00	245,00	120,00
172,00	237,00	237,00	120,00
172,00	244,00	254,00	
	203,00	165,00	
	263,00	185,00	
	188,00	215,00	
	213,00	204,00	
	204,00	165,00	
	192,00		
	200,00		
	180,00		
	180,00		
	179,00		
	205,00		
	230,00		
	222,00		
	238,00		
	187,00		
	175,00		
	182,00		
	188,00		
	184,00		

2/72. képernyőnézet: A távolugrás adatok részlet

Vizsgáljuk meg, hogy azonosnak tekinthető-e a különböző BMI kategóriákban szereplő hallgatók átlagos távolugrás értékei, vagyis függetlennek tekinthető-e a BMI kategória a helyből távolugrás értéktől, illetve homogénnek tekinthető-e a hallgatók helyből távolugrás értéke?

Amennyiben feltételezzük, hogy a hallgatók helyből távolugrás értékei normális eloszlást követnek, valamint hogy valamennyi BMI kategóriában egyenlő a távolugrások szórása, akkor alkalmazható a varianciaanalízis módszere (Forrás: fittségi 57fő_adatbázis_alap_bmikat. xlsx). A következő táblázat a részeredményeket közli, amelyeket a továbbiakban felhasználunk:

2/56. táblázat: A számítás részeredményei

		Helyből távolugrás (cm)		
		Átlag	szórás	darab
BMI kategóriák	sovány	195,33	40,41	3
	normál	212,61	26,88	44
	túlsúlyos	208,75	35,08	8
	elhízott	120,00	0,00	2
	Összesen	**207,91**	**32,69**	**57**

Először határozzuk meg a főátlagot:

$$\overline{x} = \frac{3\times195{,}33+44\times212{,}61+8\times208{,}75+2\times120}{57} = 207{,}91$$

Írjuk fel az eltérésnégyzet- összegeket:

$$SS_K = 3(195{,}33-207{,}91)^2 + 44(212{,}61-207{,}91)^2 + 8(208{,}75-207{,}91)^2 + 2(120-207{,}91)^2 = 16909{,}96$$

$$SS_B = 3\times(40{,}41)^2 + 44\times(26{,}88)^2 + 8\times(35{,}08)^2 + 2\times(0{,}00)^2 = 46526{,}77$$

$$F = \frac{\frac{SS_K}{m-1}}{\frac{SS_B}{n-m}} = \frac{\frac{16909{,}96}{4-1}}{\frac{46526{,}77}{57-4}} = 6{,}42$$

Az F kritikus érték (3;57) szabadságfok-párnál: ${}_{3;53}F_{0{,}05}=2{,}78$

Ezek szerint a nullhipotézist el kell vetni, az sportolók súlyai különböznek sportágak szerint. A korábbiakban tanultak szerint, a szóráshányados mutató is meghatározható:

$$H = \sqrt{\frac{SS_K}{SS}} = \sqrt{\frac{SS_K}{SS_K + SS_B}} = \sqrt{1-\frac{SS_B}{SS}} = \sqrt{\frac{16909{,}96}{63436{,}73}} = 0{,}51$$

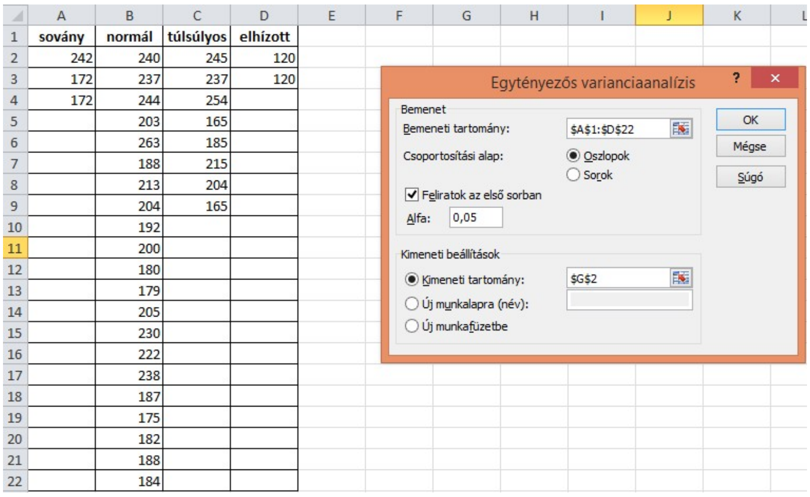
Vagyis a helyből távolugrás eredmények és a BMI kategóriák között az egyetemista populációban a kapcsolat közepes szorosságú, a távolugrás eredmények 26,66%-át, vagyis a szóródásának 26,66 százalékát (H^2) magyarázza meg a BMI kategóriája.

Az Excell programban az egytényezős varianciaanalízis gyorsan számítható, hiszen az adatelemzésbe menüpontban beépített modul áll rendelkezésre. A számításhoz feltétlenül szükséges, hogy az adatok összefüggő tartományt alkossanak, illetve a különböző részsokaságok sor vagy oszlop szerint is rendezve legyenek. Itt is abból a feltételezésből indulunk ki, hogy a testsúlyok normális eloszlást követnek.

A hipotézisrendszerünk változatlanul:

$$H_0 : \mu_1 = \mu_2 = \ldots\mu_m = \mu$$
$$H_1 : \mu_j \neq \mu$$

A beállításoknál a bemeneti tartományba kerül az oszloponként rendezett adathalmaz. Mivel a BMI kategóriák megnevezései is szerepelnek, ezért a feliratok az első sorban lehetőséget is ki kell jelölni. Az alfa paraméterben (szignifikancia- szint) az alapbeállítás maradhat (0,05), majd a kimeneti tartományként megadhatjuk annak a területnek a kezdő celláját, ahová az eredménytáblát helyezni szeretnénk.



	A	B	C	D
1	sovány	normál	túlsúlyos	elhízott
2	242	240	245	120
3	172	237	237	120
4	172	244	254	
5		203	165	
6		263	185	
7		188	215	
8		213	204	
9		204	165	
10		192		
11		200		
12		180		
13		179		
14		205		
15		230		
16		222		
17		238		
18		187		
19		175		
20		182		
21		188		
22		184		

2/73. képernyőnézet: Egytényezős varianciaanalízis (részlet) az Excel programban

A helyesen elvégzett beállításokat követően az OK gomb megnyomásával az alábbi megoldáshoz jutunk:

Egytényezős varianciaanalízis						
ÖSSZESÍTÉS						
Csoportok	*Darabszám*	*Összeg*	*Átlag*	*Variancia*		
sovány	3	586	195,3333	1633,333		
normál	44	9355	212,6136	722,3356		
túlsúlyos	8	1670	208,75	1230,5		
Elhízott	2	240	120	0		
VARIANCIAANALÍZIS						
Tényezők	*SS*	*df*	*MS*	*F*	*p-érték*	*F krit.*
Csoportok között	16909,96292	3	5636,654	6,957115	0,000493	2,779114
Csoporton belül	42940,59848	53	810,2			
Összesen	59850,5614	56				

2/74. képernyőnézet: a varianciaanalízis végeredményei

Az eredmény első részében a BMI kategóriákra vonatkozóan egy alapstatisztikát láthatunk, melyben látható, hogy a három fő sovány BMI kategóriába tartozó átlagos távolugrás értéke 195,33 cm 1633,33 varianciával. Itt ismét meg kell jegyezni, hogy a program alapértelmezése -, mint a legtöbb statisztikai programcsomagnál- a korrigált szórás.[40] Könnyen belátható, hogy a korrigált szórás kissé meghaladja a szórást (felülről közelíti), ám

[40] $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})}{n-1}}$

különbségük főleg nagy elemszám esetén elenyésző. Ebből adódik, hogy a belső szórás értéke nem egyezik az általunk számolt értékkel, azonban ez az eltérés nem jelentős.
A számítások apróbb eltéréseit a szórás adatok okozzák, vagyis a próbafüggvény értéke 6,96, ami több mint a kritikus érték 2,78, tehát a nullhipotézist el kell vetni, vagyis a sportolók testsúlya heterogén a sportágak szerint. Hasonló eredményre jutunk a szignifikancia- érték alapján is, hiszen ha a nullhipotézist elvetjük, akkor nagyon kis valószínűséggel (0,1%) követünk el hibát.
Most nézzük meg az egytényezős varianciaanalízist az SPSS program segítségével is.

Első lépésben a folytonos változó eloszlását teszteljük, hiszen a normalitásvizsgálat a varianciaanalízis egyik előfeltétele. A normailtásvizsgálat alkalmas arra, hogy két valószínűségi változó eloszlását összehasonlítsuk, vagy ellenőrizzük, hogy egy valószínűségi változónak csakugyan az az eloszlása, amit feltételeztünk. Jelen esetben arra vagyunk kíváncsiak, hogy a helyből távolugrás változó normális eloszlásból származik-e. A változó kiugró/outlier értékeinek ellenőrzését követően vizsgáljuk a normalitást. Amennyiben a számított szignifikancia érték magasabb, mint 5 %, akkor elmondható, hogy a változó normál eloszlást követ. A felhasznált adatbázis elemszáma 57 fő, így a nem paraméteres eljárások közül ismét Kolmogorov- Szmirnov próbát alkalmazzuk a normalitás tesztelésére.

A megfelelő beállítások elvégzése után a következő eredményekhez jutunk:

2/57. táblázat

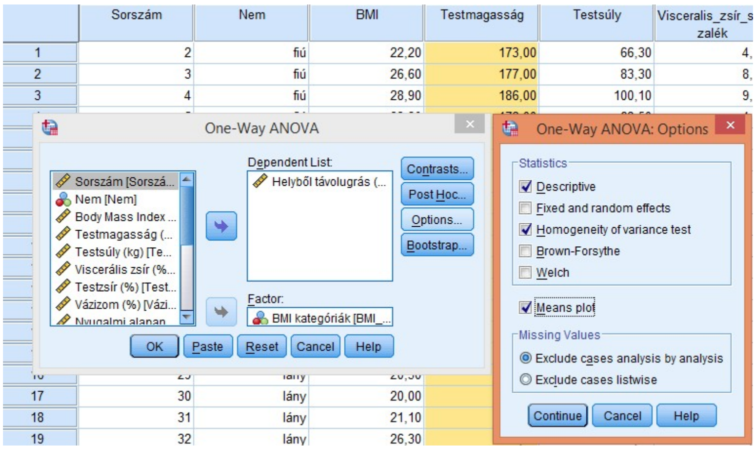
Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Helyből távolugrás (cm)	57	207,9123	32,69190	120,00	265,00

One-Sample Kolmogorov-Smirnov Test

		Helyből távolugrás (cm)
N		57
Normal Parameters[a,b]	Mean	207,9123
	Std. Deviation	32,69190
Most Extreme Differences	Absolute	,094
	Positive	,080
	Negative	-,094
Test Statistic		,094
Asymp. Sig. (2-tailed)		,200[c,d]

A Kolmogorov- Szmirnov Z értéke, valamint a hozzá tartozó szignifikancia szint alapján (p=0,20) megállapítható, hogy a változó normál eloszlást követ, így a varainciaanalízis, mint paraméteres próba elvégezhető. Abban az esetben, ha a Z értékhez tartozó szignifikancia kisebb, mint 5%, akkor nem paraméteres Kruskal- Wallis próbát kell végrehajtani (lásd Ács Pongrác: Gyakorlati adatelemzés c. könyve). Amennyiben problémát jelent a normalitásvizsgálat, akkor tekintsük úgy, mintha nem normál eloszlású lenne a mintánk, mert ha normál eloszlású adatokon nem parametrikus tesztet végzünk, gyakorlatilag a parametrikus teszttel azonos eredményt kapunk, míg fordított esetben ez nem áll fent!
Az elemzést az ANALYZE menü, COMPARE MEANS almenüjének, ONE- WAY- ANOVA moduljával indíthatjuk. Először az elemzésben megjelenő változókat jelöljük. A függő változó (DEPENDENT LIST) a Helyből távolugrás eredménye lesz, míg a független változónak (FACTOR), vagyis csoportosító változónak a BMI kategóriát jelöljük.



2/75. képernyőnézet: A varianciaanalízis beállításai az SPSS programban

Ezt követően a Options panelen jelöljük be a Descriptive, a Homogeneity of Variance, illetve a garfikus megjelenítést biztosító Means plot lehetőségeket. A HOMOGENEITY OF VARIANCE (varianciahomogenitás) opciót mindig használjuk, hiszen így tudjuk tesztelni a második előfeltételt, a varianciák egyezőségét. A kellő beállításokat követve (CONTINUE majd OK) eljutunk a következő eredményekhez.

A beállítások nyomán először a leíró statisztikát tartalmazó (elemszám, átlag, szórás, korrigált szórás, konfidencia intervallum alsó és felső értékei, minimum és maximum érték) táblázatot kapjuk meg:

2/58. táblázat

Descriptives

Helyből távolugrás (cm)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
sovány	3	195,3333	40,41452	23,33333	94,9381	295,7286	172,00	242,00
normál	44	212,6136	26,87630	4,05176	204,4425	220,7848	175,00	265,00
túlsúlyos	8	208,7500	35,07848	12,40212	179,4237	238,0763	165,00	254,00
elhízott	2	120,0000	,00000	,00000	120,0000	120,0000	120,00	120,00
Total	57	207,9123	32,69190	4,33015	199,2380	216,5866	120,00	265,00

A táblázatból látható, hogy a mintában 3 fő volt, aki a sovány kategóriába tartozik. Az átlagos távolugrás értékek a túlsúlyos BMI kategóriában lévő hallgatók esetében 208,75 cm, melyhez 35,08 cm szórás érték tartozott. Az is látható, hogy a távolugrás érték 95%-os konfidencia intervalluma 165 és 254 cm közé esik.

A következő táblázat a szóráshomogenitást vizsgálja az úgynevezett Levene- teszt segítségével.

2/59. táblázat

Test of Homogeneity of Variances

Helyből távolugrás (cm)

Levene Statistic	df1	df2	Sig.
2,720	3	53	,054

Amennyiben a teszt szignifikancia- szintje alacsony, mint 0,05, akkor el kell utasítani a szórások azonosságára vonatkozó hipotézist, ha azonban magasabb, mint 0,05, akkor fennáll a szórások azonossága. Amennyiben a szórások egyezősége nem állna fenn, úgy az Options panelből a Brown- Forsythe, valamint a Welch próbákat kellene alkalmazni, mivel az F-próba nem szolgáltatna releváns eredményt. Jelen esetben (p=0,054) a szórásokat azonosnak tekinthetjük. Ezt követően a varianciaanalízis táblázata következik.

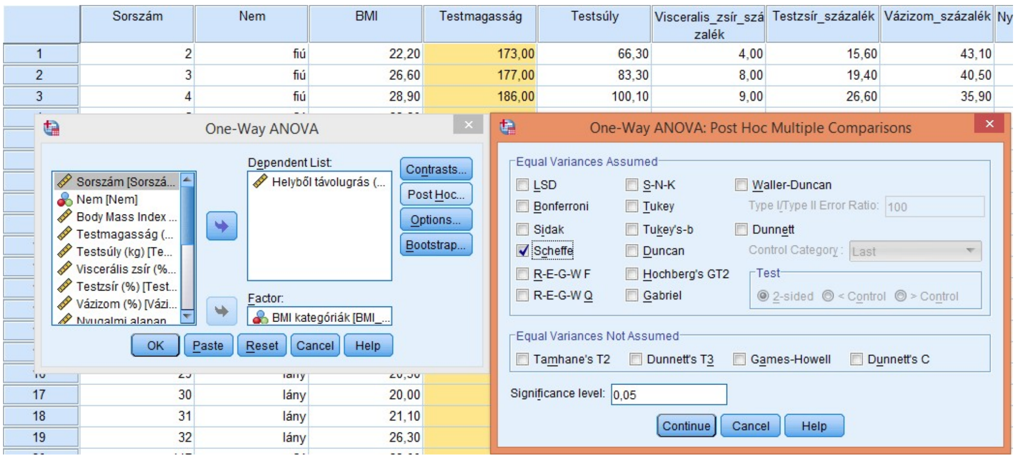
2/60. táblázat

ANOVA

Helyből távolugrás (cm)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16909,963	3	5636,654	6,957	,000
Within Groups	42940,598	53	810,200		
Total	59850,561	56			

A táblázat első oszlopában láthatóak a csoportok közötti, csoportokon belüli és a teljes eltérés négyzetösszegek. A második oszlop a szabadságfokokat tartalmazza, amelyekkel az eltérésnégyzet-összegeket elosztva a csoportok közötti és csoportokon belüli eltérésnégyzet-összegeket kapjuk. A F- értékét megkapjuk (F=6,96), ha a csoportok közötti és csoportokon belüli eltérésnégyzeteket egymáshoz hasonlítjuk. Itt is látható, hogy a szignifikanciaszint kisebb, mint 0,05, ezért a nullhipotézist elutasítjuk, vagyis a távolugrás értékek a BMI kategóriák tekintetében szignifikánsan eltérnek egymástól. A különbözőség tényét konstatálva, lehetőségünk van az egyes kategóriák átlagai közötti különbségek vizsgálatára, vagyis arra, hogy mely kategóriák különböznek egymástól. Ennek vizsgálatát segíti a Post Hoc (utólagos) modul, melynek feltétele, hogy legalább három kategóriánk legyen.



2/76. képernyőnézet: A Post Hoc tesztek beállítási modulja

A modult a ONE- WAY ANOVA panelből érhetjük el, ha ott a POST HOC lehetőségre klikkelünk. A csoportok közötti különbségek tesztelésére több utóteszt is létezik. Nincs egyetlen általánosan elfogadott eljárás, amit mindenki használ. A post-hoc tesztek elsősorban aszerint vannak csoportosítva, hogy a szórásegyezés feltétele teljesül-e. A próba kiválasztásánál két fontos szempontot kell figyelembe vennünk: mennyire könnyen lehet vele különbséget kimutatni (mennyire engedékeny), illetve mennyire megbízható. A Post Hoc tesztek első csoportja az egyenlő szórásnégyzeteknél alkalmazható teszteket tartalmazza.

A Post Hoc tesztek közül bemutatunk pár gyakran alkalmazott analízist. A szórások egyezőségekor gyakran használják a Bonferroni és Scheffe teszteket. A Bonferroni teszt páronkénti átlagok különbségének vizsgálatára használható, amikor a két csoport elemszáma különböző is lehet. Lényege, hogy az α-hibához tartozó t-értéket korrigálja a független összehasonlítások számának megfelelően. A Bonferroni teszt statisztikája:

$$L = t(táblázatbeli)\sqrt{S_p^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

A **Scheffe-teszt** a hagyományos tesztek közé tartozik. Ez már valóban a H_0 hipotéziseket vizsgálja. Az egyszerű F-próba akkor utasítja el a H_0-hipotézist, ha létezik egy a<>0 vektor, amelynél a konfidencia-intervallum nem tartalmazza a 0-t. Ha *k* darab összehasonlítandó csoport van, akkor *k(k-1)/2* összehasonlítást kell végezni. A statisztikája:

$$L = \sqrt{s_p^2(k-1)F_{(táblázatbeli)}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

A **DUNNETT-TESZT** egy kijelölt csoportot (kontroll) hasonlít össze a többivel. Eredetileg egyenlő elemszámokra volt érvényes, de később elkészült az általánosítása egyenlőtlen elemszámokra is. Lényegét tekintve páronkénti összehasonlítást végez szimultán, de meg kell adni egy kezdő, kontroll csoportot, és ehhez hasonlítja a többi csoport átlagát. Dunett-tesztet az ANALYZE, COMPARE MEANS, ONE-WAY ANOVA, POST HOC parancsok után érhetjük el. A teszt alkalmazása előtt ki kell választani a kontroll csoportot (CONTROL CATEGORY). A párbeszéd ablakból csak az első vagy utolsó csoportot tudjuk kiválasztani a legördülő listából. Továbbá meg kell adni, hogy az összehasonlítás egyoldalú vagy kétoldalú legyen. Alapbeállításként kétoldalú szimmetrikus összehasonlítás történik. Ebben az esetben nincs semmiféle előzetes információnk az összehasonlítandó párokról, bármelyik csoport lehet nagyobb, vagy kisebb, mint a kontroll. Egyoldalú próba esetében előzetesen már van információnk arról, hogy az összehasonlítandó csoport vagy csak nagyobb, vagy csak kisebb lehet, mint a kontroll csoport. Amennyiben nincs információnk a csoportok közötti relációról, mindig a kétoldalú próbát használjuk.

Statisztikája:

$$\bar{x}_i - \bar{x}_o \pm |d| s_p \sqrt{\frac{2}{n}}$$

$$\bar{x}_o = kontroll\ csoport$$

Amennyiben a szórásnégyzetek különböznek Tamhane-tesztet, illetve Dunnett’s T3 próbákat alkalmazhatunk.

Az adatbázisunk segítségével a Scheffe post- hoc próba eredményeinek értelmezése következik. A post- hoc beállítások közül jelen esetben a Scheffe- próbát választottuk, melynek legfontosabb adatait a következő output tábla mutatja.

2/61. táblázat: A variancia- analízis post hoc tesztjeinek táblázata

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Helyből távolugrás (cm)

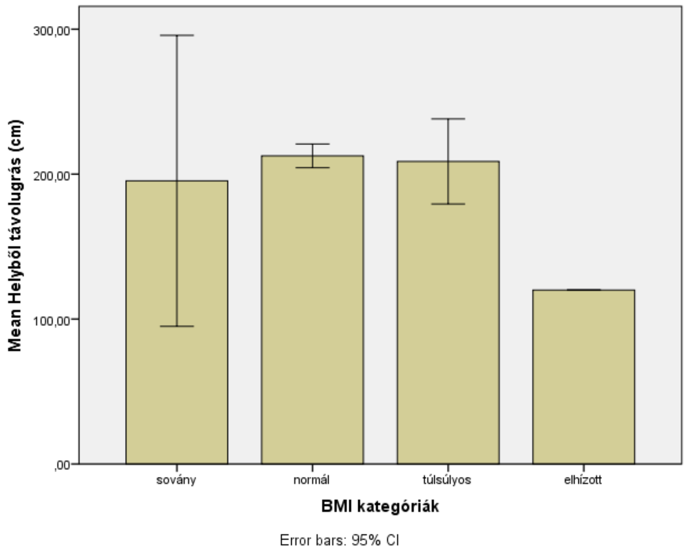
Scheffe

(I) BMI kategóriák	(J) BMI kategóriák	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
sovány	normál	-17,28030	16,98471	,793	-66,3227	31,7621
	túlsúlyos	-13,41667	19,27023	,922	-69,0584	42,2251
	elhízott	75,33333*	25,98397	,049	,3060	150,3606
normál	sovány	17,28030	16,98471	,793	-31,7621	66,3227
	túlsúlyos	3,86364	10,94023	,989	-27,7257	35,4530
	elhízott	92,61364*	20,57945	,001	33,1916	152,0357
túlsúlyos	sovány	13,41667	19,27023	,922	-42,2251	69,0584
	normál	-3,86364	10,94023	,989	-35,4530	27,7257
	elhízott	88,75000*	22,50278	,003	23,7745	153,7255
elhízott	sovány	-75,33333*	25,98397	,049	-150,3606	-,3060
	normál	-92,61364*	20,57945	,001	-152,0357	-33,1916
	túlsúlyos	-88,75000*	22,50278	,003	-153,7255	-23,7745

*. The mean difference is significant at the 0.05 level.

Az első oszlopban láthatók a viszonyítás alapját (BMI kategóriák I), a második oszlopban a viszonyítás tárgyát képező ismérvváltozók (BMI kategóriák J). A Scheffe post hoc analízis különbséget jelez a sovány és elhízott kategória között (p<0,049). A felhasználó segítségére van és a szignifikáns eltéréseket jelzi a harmadik oszlopban található (Mean Difference I-J) számított értékek mögött található apró csillag.

Az eredmények prezentációjánál jelen esetben az oszlopdiagram a javasolt megjelenítési forma, melyre érdemes egy konfidencia intervallumot illeszteni. A beállításokat a korábbiakban tárgyaltuk, így most a végeredményként megjelenő grafikus ábrát mutatjuk be.



2/31. ábra: A helyből távolugrás eredményei a BMI kategóriáinak vonatkozásában

Abban az esetben, ha az előfeltételek során a normalitás vizsgálat nem tejesül, vagyis a folytonos változó nem normál eloszlásból származik, a fent bemutatott paraméteres próbát (variancia analízist) nem alkalmazhatjuk. Amennyiben a mintánk nem az ismert normális eloszlásból származik és háromnál több csoportot tartalmaz akkor a *Kruskal-Wallis* nem paraméteres próbát alkalmazzuk. A nem paraméteres eljárásoknak a közös tulajdonsága, hogy nem tételezik fel, hogy az adatok egy adott populáció egy specifikus eloszlásához illeszkednek, szemben a paraméteres módszerekkel, melyek esetében fontos előfeltétel, hogy eloszlásuk a módszerben feltételezett tulajdonságokkal rendelkezzen. Ezért szokták ezeket a módszereket összevontan eloszlás-mentes módszereknek is nevezni.

Az egyszempontos varianciaanalízis nem paraméteres megfelelő módszereként általában a Kruskal-Wallis próbát szokták emlegetni. A Kruskal-Wallis próba végrehajtása hasonló a Mann-Whitney U próbához, sőt, két független csoport esetén mindkét módszer azonos eredményt adhat. Gyakorlatban mondható, hogy a Kruskal-Wallis próba, a Mann- Whitney teszt általánosítása három vagy több független minta esetén. A Kruskal-Wallis próba a mintákat egyesíti, kiszámítja a rangokat, majd csoportonként átlagolja. Ha a mediánok egyenlők, akkor a rangok átlagai nem térnek el lényegesen egymástól. Hátránya, hogy a post hoc teszteket a nem paraméteres próbáknál nem tudjuk elvégezni (Ács, 2015).

A gyakorlatban leginkább a többtényezős varianciaelemzéssel találkozhatunk, amely az egytényezőstől annyiban különbözik, hogy egyszerre több független változó hatását méri. A könyvünkben ennek tárgyalásától eltekintünk, az érdeklődő részletesen Sajtos László- Mitev Ariel (2007), valamint Székelyi Mária- Barna Ildikó (2005) könyvében olvashat.

3. SOKVÁLTOZÓS STATISZTIKAI ELEMZÉSEK

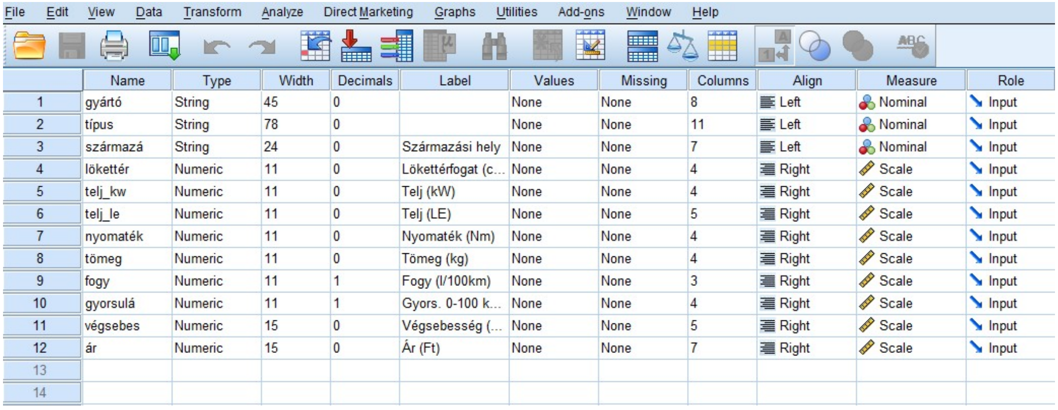
A sokváltozós statisztikai módszereket is gyakran alkalmazhatunk az önálló kutatásaink során. Az így keletkező adatbázisokat leggyakrabban az SPSS program segítségével dolgozzák fel, hiszen mára több mint 250 000 üzleti, akadémiai és kormányzati felhasználó támaszkodik az SPSS technológiára. Az SPSS nagy segítséget nyújt, amikor nagy mennyiségű adatokkal dolgozunk, hipotéziseket tesztelünk, összefüggéseket keresünk, hiszen rendkívül gyorsan juthatunk használható eredményekhez.

A leggyakrabban használt módszereket kívánjuk bemutatni egy újabb adatbázis (Forrás: motor.sav) felhasználásával. Az adatbázist Kehl Dániel a Pécsi Tudományegyetem Közgazdaságtudományi Karának adjunktusa állította össze abból célból, hogy az alapvető sokváltozós statisztikai módszerek viszonylag könnyen elsajátíthatók és érthetők legyenek. Az adatbázist úgy állította össze, hogy abban a világ legjelentősebb motor márkái, illetve a márkák legnépszerűbb típusai szerepeljenek. A végleges adatbázis ötvenhárom elemet tartalmaz, ezek jelentik a Magyarországon leginkább keresett típusokat. Az adatok szekunder módon a „Motor katalógus” 2003-as kiadványából lettek kigyűjtve.

Az egyes motorokról a következő adatok állnak rendelkezésre:

1. gyártó: általában egy gyártónak több terméke is bekerült az adatbázisba,
2. típus: az adott gyártók által használt típusjelzés,
3. származás: a gyártó cég nemzetisége,
4. lökettérfogat (cm^3)
5. teljesítmény (KW)
6. teljesítmény (lőerő, LE)
7. tömeg (kg)
8. fogyasztás (l/100km)
9. gyorsulás (0-100 km/h (s))
10. végsebesség (km/h)
11. ár (Ft).

A kiinduló adatbázis a következő képernyőn látható:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	gyártó	String	45	0		None	None	8	Left	Nominal	Input
2	típus	String	78	0		None	None	11	Left	Nominal	Input
3	származá	String	24	0	Származási hely	None	None	7	Left	Nominal	Input
4	lökettér	Numeric	11	0	Lökettérfogat (c...	None	None	4	Right	Scale	Input
5	telj_kw	Numeric	11	0	Telj (kW)	None	None	4	Right	Scale	Input
6	telj_le	Numeric	11	0	Telj (LE)	None	None	5	Right	Scale	Input
7	nyomaték	Numeric	11	0	Nyomaték (Nm)	None	None	4	Right	Scale	Input
8	tömeg	Numeric	11	0	Tömeg (kg)	None	None	4	Right	Scale	Input
9	fogy	Numeric	11	1	Fogy (l/100km)	None	None	3	Right	Scale	Input
10	gyorsulá	Numeric	11	1	Gyors. 0-100 k...	None	None	4	Right	Scale	Input
11	végsebes	Numeric	15	0	Végsebesség (...	None	None	5	Right	Scale	Input
12	ár	Numeric	15	0	Ár (Ft)	None	None	7	Right	Scale	Input
13											
14											

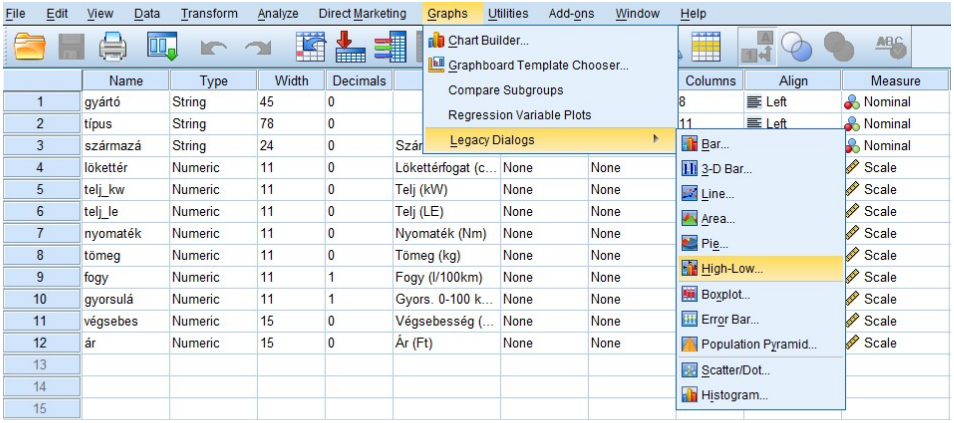
3/1. képernyőnézet: A kiinduló adatbázis

A folytonos változók vizsgálata előtt nem árt, ha a változók esetleges „kiugró" értékeit is vizsgáljuk. A következőkben röviden bemutatásra kerül ez a módszer.

Nagyon gyakran előkerülő probléma, hogy az adatok értékelése során vizsgálnunk kell azok értékeit és az esetlegesen megjelenő kiugró értékről el kell döntenünk, hogy létezhet-e valóságtartalma, vagy adatbeviteli hiba eredményeként keletkeztek. Ezt az adattisztítási módszert összefoglaló néven **kiugró (outlier) értékek vizsgálatának** nevezzük. A módszer eredménye nyomán a kutatónak a detektált kiugró értékekről kell eldöntenie, hogy kizárja, vagy az elemzés számára megtartja. Leginkább a folytonos változók esetében használható és érdemes a statisztikai analízisek előtt az extrém értékeket vizsgálni. A döntés meghozatalában a kutató szakértelme és a témában való jártassága, olvasottsága játszik priori szerepet. A gyakorlatban a kutatók azokat az értékeket szokták kizárni, melyek adatbeviteli, kódolási hibák végett jöttek létre vagy olyan indok nélküli esemény következménye, melyre a kutatónak konkrét, objektív magyarázata nincsen. Fontos, hogy meg kell tudni határozni azt a konkrét értéket, mely felett kiugró, extrém értéknek minősül az adat. Hibát követünk el, abban az esetben, ha a kiugró érték megtartása mentén az torzító hatással bír (pl. normalitásvizsgálat során), illetve akkor is, ha kizárás mellett döntünk, pedig az nem lenne indokolt, hiszen az érték egy valós megfigyelést tartalmaz és az általánosíthatóságot is segíti. Az egyik leggyakrabban alkalmazott módszertan az outlier (kiugró) értékek vizsgálatára az ún. boxplot diagram. Az ábra segítségével egyértelműen kiderülhet, hogy van-e kiugró értékünk és az mely egyednél fordul elő.

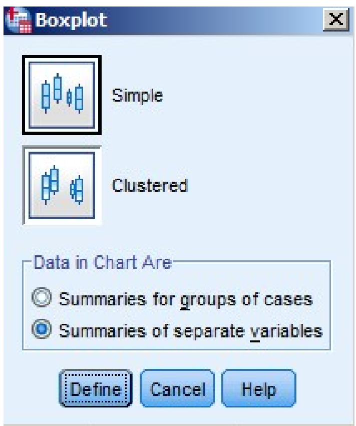
Vizsgáljuk meg, Boxplot ábra segítségével az adatbázisunkban található a motorok fogyasztását (motor. sav)

A grafikus ábra készítését megtaláljuk a GRAPHS menüpont LEGACY DIALOGS/BOXPLOT opciójában.



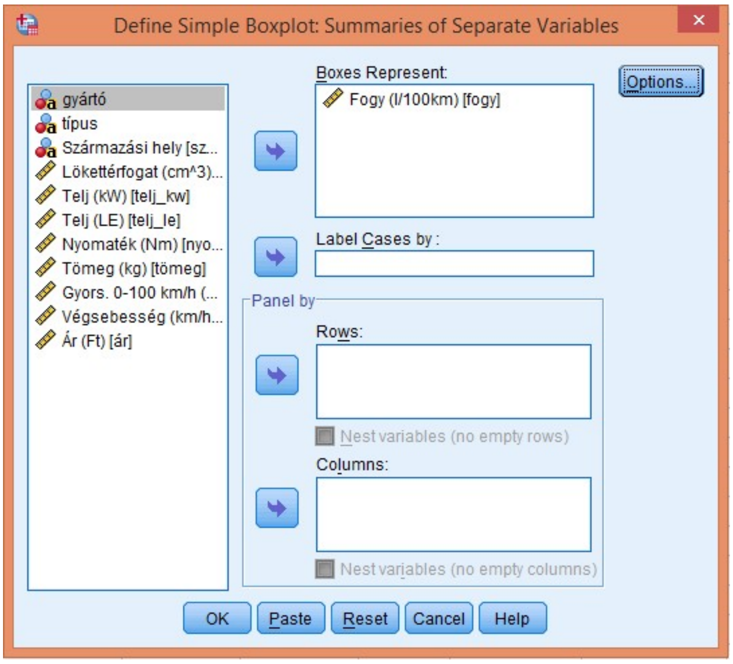
3/2. képernyőnézet: A Boxplot ábra elérési útvonala

A BOXPLOT menüpont alatt két alternatív megjelenítés áll rendelkezésre: egyszerű (SIMPLE) vagy a csoportosított (CLUSTERED) lehetőségek. Itt jelöljük az egyszerű (SIMPLE) megjelenési formát. Ezt követően az adatok típusánál (DATA IN CHART ARE) választható az eset (SUMMARIES FOR GROUPS OF CASES) vagy a változó szerinti (SUMMARIES OF SEPARATE VARIABLES) típus. Itt válasszuk a második lehetőséget. A kalibrálásunk nyomán az egyszerű típusban a változók típusa szerinti megjelenítés lesz a jó megoldás, hiszen így lehetőség van egy vagy több változó eloszlásának bemutatására. Amennyiben az esetek (CASES) szerinti ábrázolást választjuk, abban az esetben egy adott változót (pl: végsebesség) egy másik változó kategóriái (pl: származás hely) szerint ábrázolhatunk. A CLUSTERED opció segítségével, amennyiben a változók típusa szerint ábrázolunk, egyszerre minimum két folytonos változót jelenítünk meg (pl: teljesítmény, végsebesség) egy másik változó (pl. származás) kategóriák szerint.



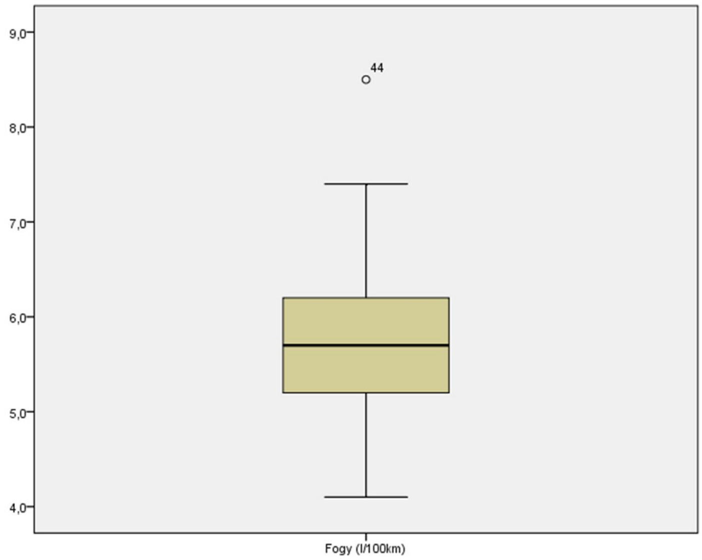
3/3. képernyőnézet: Boxplot típusának kalibrálása

Ezt követően a vizsgálatba bevont változóink kiválasztása következhet. A BOXES REPRESENT ablakban kell a fogyasztás változót a középen található nyíl segítségével mozgatni. Ezt követően további beállításokra nincsen szükség, így az OK gombra kell kattintani.



3/4. képernyőnézet. A vizsgálatba bevont változó(k) kijelölése

Az output ábrán megjelenő téglalapok (dobozok) szélei mutatják az alsó (25) és felső (75) kvartilis közötti távolságot, míg a középen megjelenő vonal a medián (50) értékét. Az interkvartilis (felső és alsó kvartilis különbsége) másfélszerese a dobozból felfelé és lefelé irányuló vonalak hossza. Ideális esetben az értékek ebben a terjedelemben helyezkednek el (normál eloszlás), melyet nyomatékosít a vonalak végén lévő vízszintes jelzés. Amennyiben az érték a dobot szélétől 1,5-3- interkvartilis terjedelemre van kiugró értékként jelöli a program (jele:O). Az ezt a három interkvartilis terjedelmet is meghaladó értéket extrém értékként kezeli és * jelöli.



3/1. ábra. A fogyasztás változó Box-plot ábrája

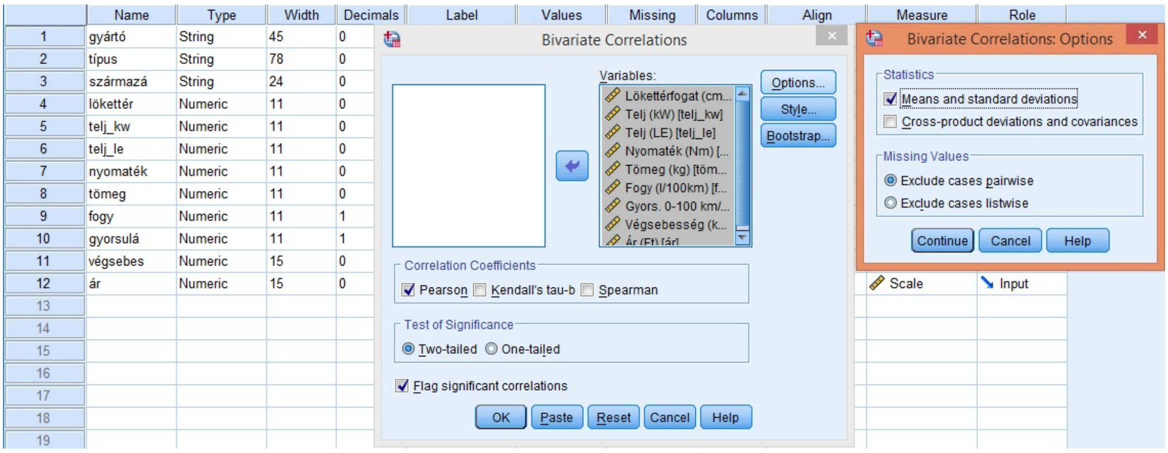
Az ábrán jól látható, hogy a fogyasztás változónál egy kiugró érték szerepel (44 sorszámú motor), viszont extrém kiugró érték nincsen. Az adatnézetben van lehetőségünk, hogy a 44-es motorhoz tartozó adatot megnézzük.

	gyártó	típus	származá	löket...	telj_kw	telj_le	nyomaték	tömeg	fogy	gyorsulá	végse...	ár
26	Honda	Fireblade	japán	954	110	150	104	199	5,3	3,0	277	3149000
27	Honda	VTR 1000 SP-2	japán	999	99	135	102	218	5,7	3,1	278	3348000
28	Honda	X-11 CBS	japán	1137	100	136	113	254	6,4	2,9	251	2590000
29	Honda	CBR 1100 XX	japán	1137	112	152	119	254	6,5	3,0	290	2890000
30	Honda	GL 1800 Gold ...	japán	1832	87	118	167	399	6,8	4,1	200	6490000
31	Kawasaki	ZZR 600	japán	599	72	98	66	221	6,1	3,8	236	2190000
32	Kawasaki	Ninja ZX-6R	japán	636	87	118	67	188	6,1	3,1	263	2550000
33	Kawasaki	Ninja ZX-9R	japán	899	105	143	100	211	5,3	2,9	276	3250000
34	Kawasaki	ZZR 1200	japán	1164	112	152	124	276	6,7	2,7	275	3350000
35	Kawasaki	Ninja ZX-12R	japán	1199	131	178	134	249	6,8	2,7	298	3630000
36	Kawasaki	VN 1500 Mean ...	japán	1471	53	72	114	314	6,2	4,8	182	3400000
37	Moto Guzzi	V11	olasz	1064	67	91	94	246	6,2	3,9	214	2699000
38	MV Agusta	F4 S EVO3	olasz	749	101	137	80	205	6,1	3,1	274	.
39	Suzuki	GSX-R 600	japán	600	85	116	68	188	5,5	3,1	254	2398000
40	Suzuki	GSX 750 F	japán	750	68	92	67	235	4,8	3,6	223	1798000
41	Suzuki	DL 1000 V-Storm	japán	996	72	98	100	235	5,1	3,4	200	2498000
42	Suzuki	GSF 1200 SsB...	japán	1157	72	98	91	244	4,7	3,2	235	1998000
43	Suzuki	GSX 1300 R Ha...	japán	1299	129	175	138	251	6,6	2,7	295	3198000
44	Suzuki	VL 1500 Intrude...	japán	1462	49	67	112	314	8,5	5,5	160	2748000
45	Triumph	Speed Triple	anglia	955	88	120	100	210	5,0	3,0	240	2998000
46	Triumph	Daytona 955i	anglia	955	108	147	100	212	5,3	3,1	260	3499000

3/5. képernyőnézet: Az adatnézetben a 44. motor fogyasztás értéke

Látható, hogy a 44 motorhoz tartozó fogyasztás 8,5 l/100 km, ami a lökettérfogat tükrében (1462) egy hihető eredmény, vagyis nem adatrögzítési hiba végett került az adatbázisba, így nem szükséges kizárni. Amennyiben adatrögzítési hiba lépne fel, akkor az adatot kik kell zárni. Az ilyen esetekben vagy a missing értéktartományt tudjuk definiálni, vagy adatszűrőt tudjuk alkalmazni (DATA menüpont SELECT CASES).

Az első lépésben azt vizsgáljuk meg, hogy az adatbázisban található mennyiségi ismérvek milyen kapcsolatban vannak egymással. A mennyiségi ismérvek (lökettérfogat, teljesítmény (kW), teljesítmény (LE), nyomaték, tömeg, fogyasztás, gyorsulás, végsebesség, ár) kapcsolatának vizsgálata a korrelációszámítással történik, melyet az ANALYZE menü, CORRELATE almenüjének, BIVARITE moduljával végzünk.



3/6. képernyőnézet: Korrelációszámítás beállításai

Első lépésben a változókat jelöljük ki, majd az OPTIONS gomb megnyomását követően válasszuk az átlag és a szórás megjelenítést. Ezt követően a CONTINUE és OK lenyomásával jutunk el a következő eredményekhez:

3/1. táblázat

Descriptive Statistics

	Mean	Std. Deviation	N
Lökettérfogat (cm^3)	1044,57	277,637	53
Telj (kW)	78,45	24,664	53
Telj (LE)	106,72	33,458	53
Nyomaték (Nm)	98,13	22,349	53
Tömeg (kg)	248,36	51,095	53
Fogy (l/100km)	5,779	,8092	53
Gyors. 0-100 km/h (s)	3,825	1,1437	53
Végsebesség (km/h)	225,94	42,198	53
Ár (Ft)	3440618,00	1343598,867	50

Az első táblázat leíró statisztikai adatokat (átlag, szórás, vizsgálatba bevont elemek száma) tartalmaz. Látható, hogy az ár adatainak a száma (N=50) eltér a többi változó elemszámától (N=53), ami azért lehetséges, mivel a forgalmazó három típusról csak külön érdeklődésre ad információt az árral kapcsolatosan.

A következő tábla a változók közötti korrelációs kapcsolat vizsgálatát mutatja, ahol a kapcsolat létezésén kívül (Sig. 2-tailed), a kapcsolat szorossága is könnyen leolvasható. A program egy, illetve két csillaggal jelzi a szignifikáns kapcsolatokat.

3/2. táblázat

Correlations

		Lökettérfogat (cm^3)	Telj (kW)	Telj (LE)	Nyomaték (Nm)	Tömeg (kg)	Fogy (l/100km)	Gyors. 0-100 km/h (s)	Végsebesség (km/h)	Ár (Ft)
Lökettérfogat (cm^3)	Pearson Correlation	1	-,087	-,086	,850**	,820**	,408**	,404**	-,335*	,607**
	Sig. (2-tailed)		,537	,539	,000	,000	,002	,003	,014	,000
	N	53	53	53	53	53	53	53	53	50
Telj (kW)	Pearson Correlation	-,087	1	1,000**	,401**	-,327*	,118	-,823**	,937**	-,004
	Sig. (2-tailed)	,537		,000	,003	,017	,400	,000	,000	,977
	N	53	53	53	53	53	53	53	53	50
Telj (LE)	Pearson Correlation	-,086	1,000**	1	,401**	-,327*	,119	-,823**	,937**	-,004
	Sig. (2-tailed)	,539	,000		,003	,017	,395	,000	,000	,979
	N	53	53	53	53	53	53	53	53	50
Nyomaték (Nm)	Pearson Correlation	,850**	,401**	,401**	1	,589**	,415**	-,044	,129	,537**
	Sig. (2-tailed)	,000	,003	,003		,000	,002	,757	,357	,000
	N	53	53	53	53	53	53	53	53	50
Tömeg (kg)	Pearson Correlation	,820**	-,327*	-,327*	,589**	1	,351**	,616**	-,542**	,658**
	Sig. (2-tailed)	,000	,017	,017	,000		,010	,000	,000	,000
	N	53	53	53	53	53	53	53	53	50
Fogy (l/100km)	Pearson Correlation	,408**	,118	,119	,415**	,351**	1	,104	,008	,221
	Sig. (2-tailed)	,002	,400	,395	,002	,010		,458	,954	,124
	N	53	53	53	53	53	53	53	53	50
Gyors. 0-100 km/h (s)	Pearson Correlation	,404**	-,823**	-,823**	-,044	,616**	,104	1	-,883**	,305*
	Sig. (2-tailed)	,003	,000	,000	,757	,000	,458		,000	,031
	N	53	53	53	53	53	53	53	53	50
Végsebesség (km/h)	Pearson Correlation	-,335*	,937**	,937**	,129	-,542**	,008	-,883**	1	-,191
	Sig. (2-tailed)	,014	,000	,000	,357	,000	,954	,000		,184
	N	53	53	53	53	53	53	53	53	50
Ár (Ft)	Pearson Correlation	,607**	-,004	-,004	,537**	,658**	,221	,305*	-,191	1
	Sig. (2-tailed)	,000	,977	,979	,000	,000	,124	,031	,184	
	N	50	50	50	50	50	50	50	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

A kapott eredmények alapján néhány jellegzetes dologra felfigyelhetünk:

- a lökettérfogat szorosan korrelál a nyomatékkal, hisz ez a robbanómotor elvéből következik: nagyobb hengerűrtartalom nagyobb robbanást jelent,
- a lökettérfogat gyakorlatilag nem korrelál a teljesítménnyel, ami a valóságban nem egészen így van: ha kizárólag sportmotorokat vizsgáltunk volna, ez a kapcsolat erősebb lenne,
- a tömeg és a végsebesség közötti közepes erősségű negatív irányú korrelációt figyelhetünk meg, mely nem szorul különösebb magyarázatra,
- a gyorsulás és a végsebesség közötti erős negatív kapcsolat azt jelenti, hogy a magasabb végsebességre képes motorok gyorsabban érik el a 100 km/h-s sebességet (ez annak tudható be, hogy mindkét értéket a motor teljesítménye határozza meg),
- az ár egyik tényezővel sem mutat közepesnél erősebb kapcsolatot, érdekesség, hogy a tömeggel mutatja a legerősebb kapcsolatot,

- a fogyasztás egyik tényezővel sincs közepesnél erősebb kapcsolatban, sőt a végsebességtől teljes mértékben független az adatbázis adatai szerint,
- a kilowattban és lóerőben mért teljesítmény természetesen függvényszerű kapcsolatban állnak egymással, hiszen 1 LE ≈ 0,735 kW.

A következőkben a leggyakrabban alkalmazott sokváltozós statisztikai módszerek (faktor- analízis, klaszter- analízis, diszkriminancia- analízis) rövid ismertetése történik.

3.1. Faktor- analízis

Az elmúlt időszakban a faktor- analízis módszere a sokváltozós elemzések gyakorlati alkalmazásai során megnőtt, a módszer adattömörítő és összefüggés-feltáró voltának köszönhetően. A módszer segítségével a nagyszámú változók, olyan faktorváltozókba vonhatók össze, amelyek közvetlenül nem megfigyelhetők. A nagyszámú sztochasztikusan összefüggő változók helyett, kisszámú faktorváltozókat keresünk, mely segítségével az adatok értelmezése és további elemzése egyszerűbb lesz, hiszen csökken a kiinduló változók száma.

Az így újonnan létrejövő faktorok egyáltalán nem korrelálnak egymással. A gyakorlati alkalmazása a kérdőíves kutatások előtérbe kerülésének köszönhető, hiszen a kérdőívek hajlamosak egy-egy kérdéskört (szokások, jellemzők, életstílusok, stb.) túlzóan is körüljárni, mely által az adatfeldolgozás nehézkes lehet. Ilyen esetekben előszeretettel alkalmazzák a kutatók ezt a módszert, hiszen a változók számának csökkentésével próbálja feltárni az egyes jellemzők kapcsolatrendszerét. A faktor-analízis egy struktúra- feltáró módszer, ami azt jelenti, hogy a függő és független változók nem előre meghatározottak, tehát a változók összefüggéseinek feltárására törekszik. (Sajtos L.- Mitev A. ,2007)

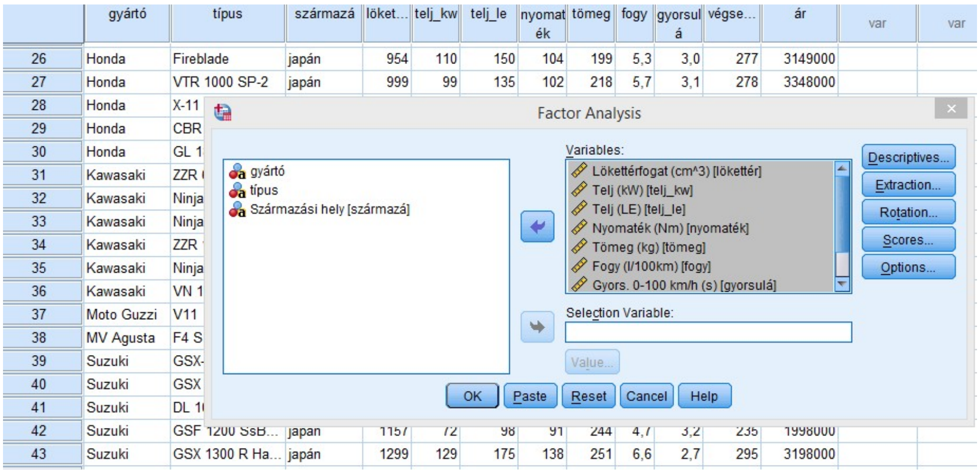
A faktor-analízis másik előnye, hogy a létrejövő új faktorok további sokváltozós elemzések során is felhasználhatók.

A faktor-analízis során előforduló leggyakoribb kérdések:

- Hogyan lehet a változók által közösen magyarázott információt kis számú, lehetőleg korrelálatlan faktorokkal kifejezni?
- A létrejövő új faktorok milyen mértékben magyarázzák az eredeti változókat?
- Mely változók vannak ugyanazon faktorokban?
- Mi lehet az egyes faktorok jelentése, illetve elnevezése?

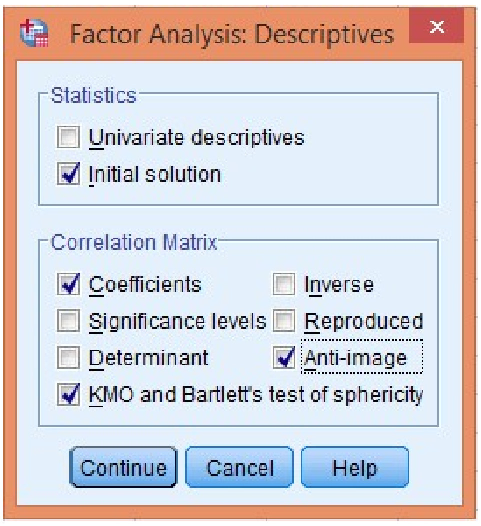
(Forrás: Ketskeméty- Izsó, 2005)

A faktor-analízist az ANALYZE menü, DATA REDUCTION almenüjének, FACTOR moduljával készíthetünk, melynek kijelölését követően a vizsgálatba bevont változókat kell kiválasztani és a VARIABLES ablakba mozgatni a nyíl segítségével. (Forrás: motor.sav)



3/7. képernyőnézet: A faktor- analízis beállításai

Ezt követően DESCREPTIVES doboz segítségével tudjuk tesztelni, hogy a fent bevont változók alkalmasak-e a faktor-analízisre. A STATISTICS menü alapbeállítása mellet kérhetünk egyváltozós leíró statisztikát is (UNIVARIATE DECREPTIVES), mely a fent már bemutatott táblát (átlag, szórás, elemszám) adja eredményül.



3/8. képernyőnézet: Az előfeltételek beállításai

A korrelációs mátrix itt is előállítható, mely fontos feltétele az elemzésnek, hiszen az egyes változók korrelációja alapfeltétele a faktor- analízisnek. A változók közti szoros korreláció, arra utal, hogy a bevont változók alkalmasak a faktorelemzésre. A COEFFICIENT doboz jelelősével a korrelációs mátrix korrelációs értékeit (koefficienseit) kapjuk.

3/3. táblázat

Correlation Matrix

		Lökettérfogat (cm^3)	Telj (kW)	Telj (LE)	Nyomaték (Nm)	Tömeg (kg)	Fogy (l/100km)	Gyors. 0-100 km/h (s)	Végsebesség (km/h)	Ár (Ft)
Correlation	Lökettérfogat (cm^3)	1,000	-,069	-,069	,850	,821	,429	,396	-,321	,607
	Telj (kW)	-,069	1,000	1,000	,421	-,319	,111	-,826	,937	-,004
	Telj (LE)	-,069	1,000	1,000	,421	-,319	,112	-,825	,937	-,004
	Nyomaték (Nm)	,850	,421	,421	1,000	,593	,424	-,052	,149	,537
	Tömeg (kg)	,821	-,319	-,319	,593	1,000	,385	,608	-,542	,658
	Fogy (l/100km)	,429	,111	,112	,424	,385	1,000	,122	,000	,221
	Gyors. 0-100 km/h (s)	,396	-,826	-,825	-,052	,608	,122	1,000	-,890	,305
	Végsebesség (km/h)	-,321	,937	,937	,149	-,542	,000	-,890	1,000	-,191
	Ár (Ft)	,607	-,004	-,004	,537	,658	,221	,305	-,191	1,000

Ez a táblázat megegyezik a fent tárgyalt táblázattal. A DESCREPTIVE dobozban a másik fontos előfeltétel tesztelélésre az ANTI- IMAGE doboz jelöltük meg. Ez abból indul ki, hogy a változók szórásnégyzete felbontható megmagyarázott és meg nem magyarázott szórásnégyzetre, melyet az anti- image kovariancia és variancia mátrixok mutatnak. A két mátrix közül az anti- image korrelációs mátrix átlóban lévő értékei az MSA értékek. Ezen értékek 0 és 1 között lehetnek és leginkább az átlóban található értékek fontosak számunkra, hiszen megmutatja, hogy az adott változó mennyire áll szoros kapcsolatba a többi bevont változóval. Az MSA értéke magas, akkor a változó jól illeszkedik a faktorszerkezetbe, ha alacsony (0,5 alatti), akkor nagy a valószínűsége, hogy ki kell majd a változót zárni az elemzésből.

3/4. táblázat

Anti-image Matrices

		Lökettérfogat (cm^3)	Telj (kW)	Telj (LE)	Nyomaték (Nm)	Tömeg (kg)	Fogy (l/100km)	Gyors. 0-100 km/h (s)	Végsebesség (km/h)	Ár (Ft)
Anti-image Covariance	Lökettérfogat (cm^3)	,050	7,482E-5	-3,695E-5	-,039	-,008	-,019	-,014	-,008	-,015
	Telj (kW)	7,482E-5	6,422E-5	-6,448E-5	-9,820E-5	,000	,001	,001	,000	-2,067E-5
	Telj (LE)	-3,695E-5	-6,448E-5	6,487E-5	5,116E-5	,000	-,001	-,001	4,589E-5	-1,933E-5
	Nyomaték (Nm)	-,039	-9,820E-5	5,116E-5	,038	-,013	,006	,012	,014	,010
	Tömeg (kg)	-,008	,000	,000	-,013	,177	-,070	-,030	,015	-,112
	Fogy (l/100km)	-,019	,001	-,001	,006	-,070	,723	-,045	-,018	,099
	Gyors. 0-100 km/h (s)	-,014	,001	-,001	,012	-,030	-,045	,156	,020	-,042
	Végsebesség (km/h)	-,008	,000	4,589E-5	,014	,015	-,018	,020	,043	,002
	Ár (Ft)	-,015	-2,067E-5	-1,933E-5	,010	-,112	,099	-,042	,002	,490
Anti-image Correlation	Lökettérfogat (cm^3)	,714[a]	,042	-,021	-,894	-,085	-,100	-,162	-,182	-,094
	Telj (kW)	,042	,729[a]	-,999	-,063	-,085	,112	,185	-,065	-,004
	Telj (LE)	-,021	-,999	,730[a]	,032	,085	-,115	-,182	,028	-,003
	Nyomaték (Nm)	-,894	-,063	,032	,662[a]	-,161	,034	,160	,346	,075
	Tömeg (kg)	-,085	-,085	,085	-,161	,894[a]	-,197	-,183	,172	-,380
	Fogy (l/100km)	-,100	,112	-,115	,034	-,197	,820[a]	-,134	-,103	,167
	Gyors. 0-100 km/h (s)	-,162	,185	-,182	,160	-,183	-,134	,917[a]	,239	-,153
	Végsebesség (km/h)	-,182	-,065	,028	,346	,172	-,103	,239	,922[a]	,016
	Ár (Ft)	-,094	-,004	-,003	,075	-,380	,167	-,153	,016	,858[a]

a. Measures of Sampling Adequacy(MSA)

Az MSA értékei jelen esetben 0,66 és 0,92 között vannak. A következő előfeltétel, amit, szinte minden faktor-analízis során tesztelünk: a KMO (Kaiser- Meyer- Olkin) kritérium és a Bartlett-teszt. A KMO kritérium segítségével tudjuk leginkább és legkönnyebben megállapítani, hogy a változók mennyire alkalmasak a faktor- analízisre. A KMO értékét az MSA értékek átlaga adja, amely az összes változót egyidejűleg teszteli. A KMO érték a faktor- analízis szempontjából a következőképpen írható le:

0,9 ≤KMO≤1 tökéletes

0,8 ≤KMO≤0,9 nagyon megfelelő

0,7 ≤KMO≤0,8 megfelelő

0,6 ≤KMO≤0,7 közepes

0,5 ≤KMO≤0,6 gyenge

KMO≤0,5 elfogadhatatlan, alkalmatlan

A Bartlett- próba nullhipotézise azt mondja ki, hogy a kiinduló változók között nincs korreláció, vagyis korrelálatlanok. Számunkra az lenne a jó, ha a nullhipotézist el tudnánk vetni, vagyis a változók korreláljanak egymással.

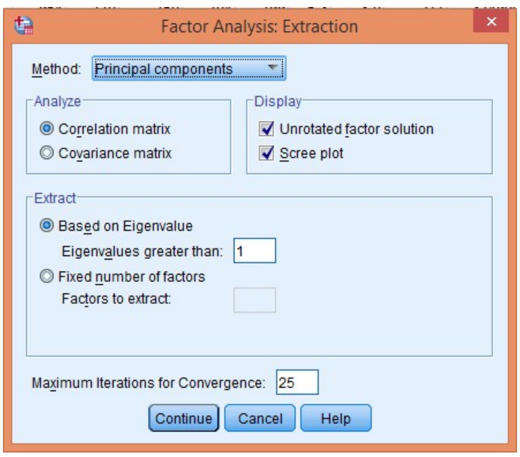
3/5. táblázat

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,796
Bartlett's Test of Sphericity	Approx. Chi-Square	901,966
	df	36
	Sig.	,000

Az eredmény alapján látszik, hogy a Bartlett-teszt szignifikancia értéke kisebb 0,05-nél, tehát a változók korrelálnak egymással, vagyis elvégezhető a faktor- analízis. Hasonló eredményt mutat a KMO értéke is (0,796), tehát a bevont változók megfelelőek a faktorelemzéshez.

A faktor- analízis párbeszédpanelében a következő ablak (EXTRACTION) segítségével választhatunk a módszerek közül, hiszen a faktorelemzés egy gyűjtőfogalom, amely több módszert tömörít.

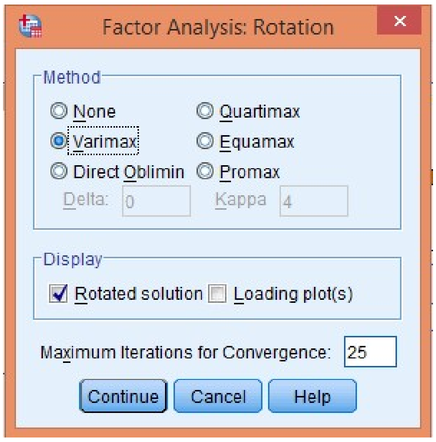


3/9. képernyőnézet: A módszer kiválasztása

A módszerek közül válasszuk a PRINCIPAL COMPONENTS (főkomponens- elemzés), hiszen ez a módszer a változók számát úgy csökkenti, hogy közben a legkevesebb információt veszíthetjük a sokaságról. Az EXTRACT dobozban beállíthatjuk a faktoraink számát. Ha a kutatónak létezik elképzelése a faktorok számának tekintetében, akkor a NUMBER OF FACTORS kijelölését követően ezt megteheti (a maximális faktorszám nem lehet több, mint a változóink száma). Alapbeállításként a KAISER- KRITÉRIUMOT (sajátérték) használja, mely szerint csak azokat a faktorokat veszi figyelembe, melynek sajátértéke minimum 1, hiszen ez alatt már az adott faktor kevesebb információt hordoz, mint egy változó.

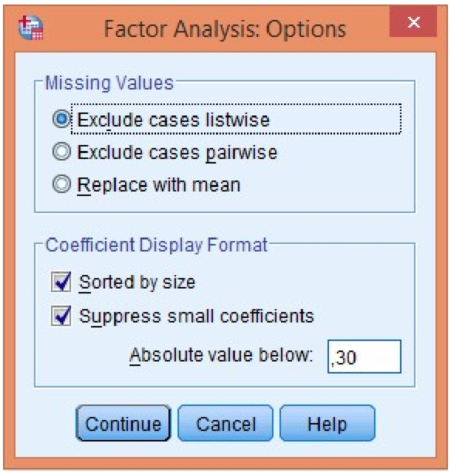
A SCREE PLOT (scree-teszt) grafikus ábra segítségével is képesek lehetünk a faktorok számát meghatározni. Ez az úgynevezett könyökszabály, mely azt mondja ki, hogy a faktorok számát ott kell meghatározni, ahol a meredekség csökken és egyenesbe kezd fordulni. Ennek értelmében lehetnek olyan faktorok is, melyek fontosak, bár sajátértéke 1 alatt van. Általában ez a szabály a Kaiser- kritériumhoz képest nem mér olyan szigorúan, és 1-3 faktorral többet engedélyez. Ennek eldöntése a mindenkori kutató feladata.

A CONTINUE GOMB lenyomását követően a ROTATION almenüben kell a faktorrotációt beállítani. Ez azt jelenti, hogy az egyszerűbb és könnyebb értelmezhetőség kedvéért a faktorok tengelyeit elforgatjuk. A faktorok forgatásának segítségével a faktorok által megmagyarázott variancia arányosabbá válik. A faktorelemzés több módszerei közül válasszuk a VARIMAX módszert, mely a leggyakrabban alkalmazott eljárás. A módszer előnye a többihez képest, hogy jobban szétválasztja a faktorokat, így az értelmezhetőség még könnyebbé válik.



3/10. képernyőnézet. A rotáció beállításai

A módszer kijelölését követően a DISPLAY keretben csak a ROTATED SOLUTIONS válasszuk, így most a komponenseket grafikus megjelenítése (LOADING PLOT) az elforgatott térben nem történik. Ezt követően az OPTIONS ALMENÜ beállításai következnek, ahol lehetőségünk van, a majdani faktorok értelmezését könnyíteni. Ha a SORTED BY SIZE lehetőséget kijelöljük, akkor a rotált faktorsúly-mátrixban a súlyok csökkenő sorrendben lesznek feltüntetve, így könnyebbé válik az értelmezés.



3/11. képernyőnézet: A rotált faktorsúly-mátrix beállításai

Szintén itt tudjuk kérni (SUPPRESS ABSOLUTE VALUES LESS THAN), hogy az általunk megadott faktorsúlyokat meghaladó értékeket írja ki. Jelöljük, hogy csak a 0,3-nál magasabb értékek szerepeljenek, ami által szintén gyorsabbá válik a faktorok értelmezése és elnevezése. Ezt követően, ha megfelelő faktorokat kaptunk, akkor elmenthetjük őket a SCORES menü SAVE AS VARIABLES opciója segítségével, így a további sokváltozós elemzések során (pl. klaszter- analízis) felhasználható.

Mindezen beállításokat elvégezve futassuk le az elemzést. Az output ablakban a következő eredményeket láthatjuk, melyek közül az első három táblázatról már esett szó.

A negyedik táblázat a változók kommunalitásának vizsgálatát mutatja. Itt el kell fogadni azt a „hüvelykujjszabályt”, hogy a végső kommunalitás értékének a 0,3-at meg kell haladnia, különben a változóknak nincsen elegendő magyarázó erejük.

3/6. táblázat

Communalities

	Initial	Extraction
Lökettérfogat (cm^3)	1,000	,894
Telj (kW)	1,000	,983
Telj (LE)	1,000	,982
Nyomaték (Nm)	1,000	,908
Tömeg (kg)	1,000	,890
Fogy (l/100km)	1,000	,331
Gyors. 0-100 km/h (s)	1,000	,894
Végsebesség (km/h)	1,000	,963
Ár (Ft)	1,000	,574

Extraction Method: Principal Component Analysis.

A táblázatban az Initial érték mindig a kezdeti 1-es érték, míg az Extraction oszlopban a faktor-analízist követő kommunalítások láthatók. Ennek értelmében nem kell változót kihagyni, hiszen mindegyik érték meghaladja a 0,3-at.

A következő táblázatban láthatjuk a faktorok által magyarázott varianciát. A táblázat három része a kezdeti (Initial), a faktor-analízist követő (Exraction Sums of Squared Loadings), illetve a forgatást követő (Rotation Sums of Squared Loadings) értékeket mutatja.

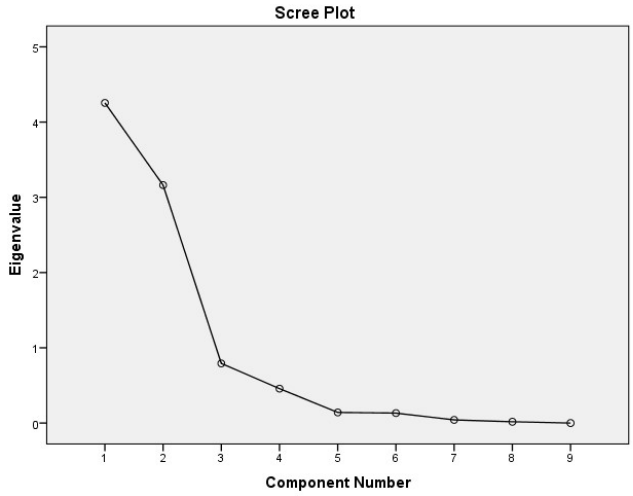
3/7. táblázat

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,255	47,281	47,281	4,255	47,281	47,281	4,035	44,834	44,834
2	3,162	35,136	82,417	3,162	35,136	82,417	3,382	37,583	82,417
3	,793	8,808	91,225						
4	,457	5,081	96,305						
5	,141	1,571	97,877						
6	,132	1,466	99,343						
7	,042	,464	99,807						
8	,017	,192	100,000						
9	3,229E-5	,000	100,000						

Extraction Method: Principal Component Analysis.

Számunkra a faktorelemzés utáni, illetve a forgatás utáni értékek fontosak, hiszen itt már csak az általunk beállított 1-nél nagyobb sajátértékű faktorokat tartalmazza. Elsőként a legnagyobb sajátértékű faktor jelenik meg (4,255/47,281). A legfontosabb számunkra, hogy a két létrejövő faktor összesített varianciája (Comulative %) magasabb, mint a kritériumnak tartott 60 százalék, hiszen 82,417 százalék, ami azt mutatja, hogy az információ csupán 17,583 %-át veszítettük el. Látható a forgatás utáni értékekből, hogy az összvariancia megmaradt csak ez eloszlása lett egyenletesebb. A következő ábra a Scree Plot, mely alapján az látszik, hogy a meredekség a harmadik faktor után csökken, és ettől kezdve kezd laposodni.



3/2. ábra: A faktor- analízis faktorszámának eldöntését segítő grafikus ábra

A könyökszabály értelmében a faktorok számát a laposodás kezdetén maximalizáljuk, tehát jelen esetben három faktort kellene létrehozni, vagyis a harmadik faktor is fontos lehet, bár sajátértéke egy alatt van. Ezt követően a forgatás nélküli faktorsúlyokat tartalmazó (Component Matrix), majd a forgatást követő faktorsúlyokat tartalmazó mátrixot kapunk. Nekünk a forgatási utáni mátrix lesz a jelentősebb.

3/8. táblázat

Rotated Component Matrix[a]

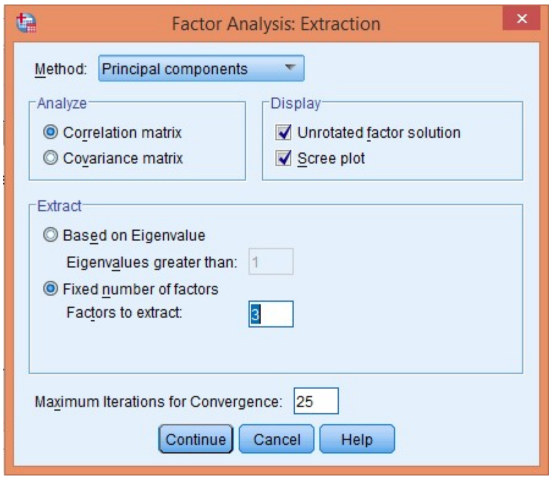
	Component	
	1	2
Telj (kW)	,984	
Telj (LE)	,984	
Végsebesség (km/h)	,970	
Gyors. 0-100 km/h (s)	-,903	
Lökettérfogat (cm^3)		,930
Nyomaték (Nm)	,323	,896
Tömeg (kg)	-,438	,835
Ár (Ft)		,749
Fogy (l/100km)		,571

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

A rotált mátrixban csak az általunk beállított (0,3) faktorsúlyoknál magasabb értékek szerepelnek. Minél magasabb az abszolút értéke egy faktorsúlynak annál fontosabb a szerepe az adott faktorban. Ez alapján az első faktorba tartozó változók: teljesítmény (kW), teljesítmény (LE), végsebesség, gyorsulás. Az összes többi változó a második faktorba került.

Most nézzük meg, miként alakulna ez az elemzés, három faktor esetén. A beállításoknál csak egy dolgot változtassunk meg, mely szerint kijelöljük, hogy három faktorba való rendezést kérünk.



3/12. képernyőnézet: A módszer és a faktorszám meghatározása

Ezt követően futassuk le az analízist, mely során látható, hogy a három faktor az összvariancia 91,225 százalékát magyarázza, tehát a három faktor alkalmazása során nagyon minimális információt fogunk veszíteni.

3/9. táblázat

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,255	47,281	47,281	4,255	47,281	47,281	4,000	44,448	44,448
2	3,162	35,136	82,417	3,162	35,136	82,417	3,069	34,098	78,546
3	,793	8,808	91,225	,793	8,808	91,225	1,141	12,679	91,225
4	,457	5,081	96,305						
5	,141	1,571	97,877						
6	,132	1,466	99,343						
7	,042	,464	99,807						
8	,017	,192	100,000						
9	3,229E-5	,000	100,000						

Extraction Method: Principal Component Analysis.

Végül a forgatás utáni faktorsúlyokat tartalmazó mátrix felhasználásával nevezzük el a keletkező három faktort.

3/10. táblázat

Rotated Component Matrix[a]

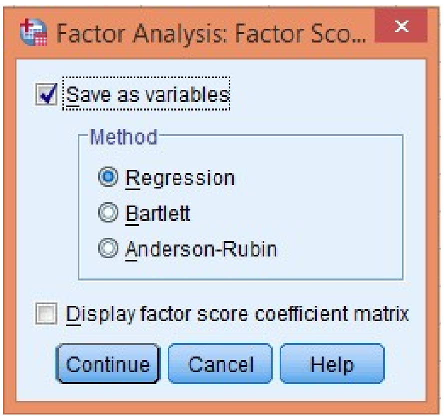
	Component		
	1	2	3
Telj (kW)	,987		
Telj (LE)	,987		
Végsebesség (km/h)	,964		
Gyors. 0-100 km/h (s)	-,897		
Lökettérfogat (cm^3)		,887	
Ár (Ft)		,846	
Nyomaték (Nm)	,346	,845	
Tömeg (kg)	-,415	,817	
Fogy (l/100km)			,947

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

- az első főkomponens a teljesítményekkel, a végsebességgel, és a gyorsulással áll szoros kapcsolatban. A leíró elemzésnél láthattuk már, hogy ezen változók között erős korrelációs kapcsolat van, ezért is kerülhettek a faktor-analízis során egy csoportba. Ha nevet szeretnénk adni ennek a főcsoportnak, talán a motor teljesítőképessége lenne a legmegfelelőbb. Ebben a komponensben a gyorsulás negatív értékkel ál, vagyis az ellentettje az igaz, tehát nem a minél magasabb másodperc szám a kedvező, hanem a minél alacsonyabb. Vagyis az a megfelelő, ha minél kevesebb időre (sec.) van szükség a 100 km/h sebesség eléréséhez.
- a második főkomponens a lökettérfogattal, az árral, a nyomatékkal, és a tömeggel és van összefüggésben. Ezt a komponenst nevezhetnénk motorikus jellemzőnek.
- a harmadik főkomponens a fogyasztással van szoros kapcsolatban. Ez az ismérv egyedül maradt a csoportban, ami a korrelációs elemzés tükrében nem meglepő, hisz a fogyasztás egyik jellemzővel sincs szoros kapcsolatban.

Miután ez a megoldás elfogadhatónak találjuk, elmenthetjük a keletkezett értékeket.



3/12. képernyőnézet: A faktorok elmentése

A mentést követően a VARIABLE VIEW ablakban jól járunk, ha rögtön a LABEL (címke) alatt elnevezzük a keletkező új faktorokat.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	gyártó	String	45	0		None	None	8	Left	Nominal	Input
2	típus	String	78	0		None	None	11	Left	Nominal	Input
3	származá	String	24	0	Származási hely	None	None	7	Left	Nominal	Input
4	lökettér	Numeric	11	0	Lökettérfogat (cm^3)	None	None	4	Right	Scale	Input
5	telj_kw	Numeric	11	0	Telj (kW)	None	None	4	Right	Scale	Input
6	telj_le	Numeric	11	0	Telj (LE)	None	None	5	Right	Scale	Input
7	nyomaték	Numeric	11	0	Nyomaték (Nm)	None	None	4	Right	Scale	Input
8	tömeg	Numeric	11	0	Tömeg (kg)	None	None	4	Right	Scale	Input
9	fogy	Numeric	11	1	Fogy (l/100km)	None	None	3	Right	Scale	Input
10	gyorsulá	Numeric	11	1	Gyors. 0-100 km/h (s)	None	None	4	Right	Scale	Input
11	végsebes	Numeric	15	0	Végsebesség (km/h)	None	None	5	Right	Scale	Input
12	ár	Numeric	15	0	Ár (Ft)	None	None	7	Right	Scale	Input
13	FAC1_1	Numeric	11	5	A motor teljesítőképessége	None	None	13	Right	Scale	Input
14	FAC2_1	Numeric	11	5	Motorikus jellemző	None	None	13	Right	Scale	Input
15	FAC3_1	Numeric	11	5	Fogyasztás	None	None	13	Right	Scale	Input
16											
17											

3/13. képernyőnézet: A faktorok elnevezése

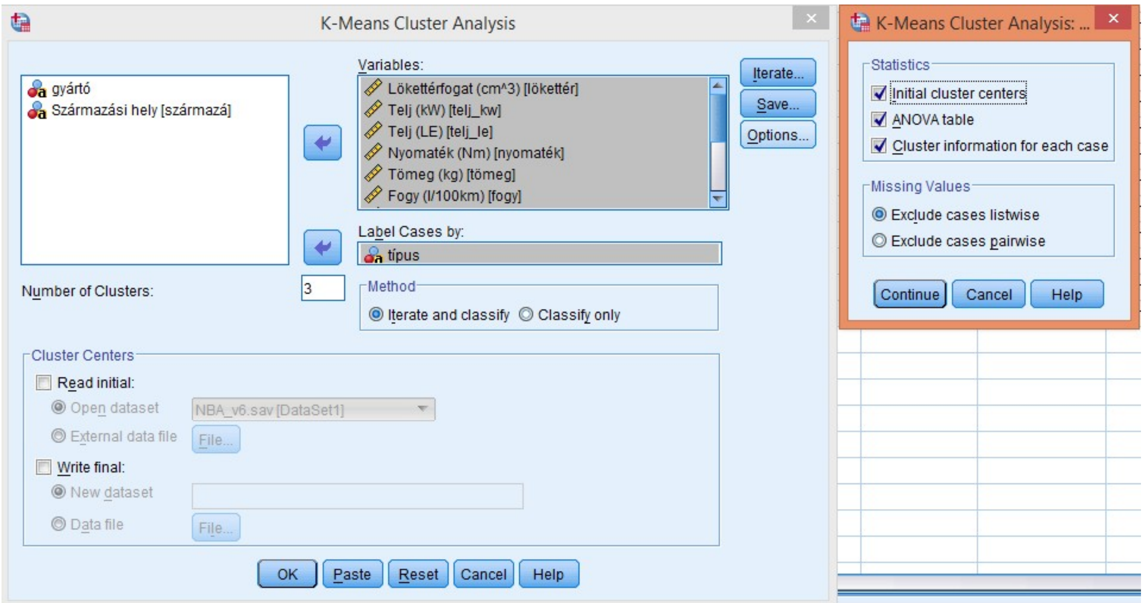
3.2. Klaszter- analízis

A klaszter- analízis a változók csoportosításával foglalkozó, dimenziócsökkentő módszer. Az analízis lényege, hogy a megfigyelési egységeket csökkentse (a faktor- analízis a változók számát csökkenti), összetartozó csoportokba rendezze, az elemzésbe bevont változó alapján. Az elemzés akkor sikeres, ha az egy csoportba, klaszterbe tartozók mindegyik vizsgált változó mentén közel vannak egymáshoz, viszont a többi csoporttól, klasztertől távol kerülnek.

A klaszter-analízisnek két nagy módszertani csoport mentén kategorizálják. Léteznek a hierarchikus (faszerű felépítés) és a nem hierarchikus (K-közép) eljárások. A hierarchikus módszereknél az úgynevezett összevonó klaszterelemzést (egyszerű-, teljes-, átlagos láncmódszer, ward módszer, centroid módszer) alkalmazzák leggyakrabban, ahol a folyamat megkezdésekor külön lévő elemeket (klasztereket) egyre nagyobb, majd legvégül egyetlen klaszterbe vonjuk össze. A módszert akkor alkalmazzák a kutatók, amikor előre nem tudják a klaszterszámot meghatározni. A nem hierarchikus K-közép eljárást nagyobb minták esetén érdemes alkalmazni, hiszen ilyen esetekben egyszerűbben értelmezhető, mint a hierarchikus eljárások. Az eljárás során a létrehozandó klaszterek számát előre rögzíteni kell!

Annak eldöntése, hogy melyik módszert válasszuk nehéz feladat, mely függ a kutató témában folytatott eddigi felméréseitől és hozzáértésétől. Éppen ezért leggyakrabban a két módszert egyszerre alkalmazzák. Első lépésben a hierarchikus módszerrel meghatározzák a klaszterek számát, majd a nem hierarchikus módszerrel elvégzik az elemzést, illetve a változók csoportosítását. Jelen esetben a nem hierarchikus módszert alkalmazzuk, mivel előzetes információval rendelkezünk a klaszterek számának tekintetében. Ennek megfelelően három klaszterbe fogjuk rendezni a típusokat. Megjegyezendő, ha a vizsgálatban bevont változóink különböző mérési skálán lennének, akkor először standardizálni[41] kellene az értékeket, majd ezt követően már elvégezhető a különböző skálákon mért adatok összehasonlítása. A legegyszerűbben a standardizálást a ANALYZE/DESCRIPTIVE STATISTICS/DESCRIPTIVES almenüjéből tudjuk megtenni, hiszen a változók standardizált értékeit tudjuk elmenteni SAVE STANDARDIZED VALUES AS VARIABLES dobozának jelölését követően.

A klaszter-analízist az ANALYZE menü, CLASSIFY almenü, K-MEANS CLUSTER moduljának segítségével készíthetjük el.



3/14. ábra: A klaszter-analízis beállításai

Az első lépésben a vizsgálatba bevonni kívánt változókat (lökettérfogat, teljesítmények, nyomaték, tömeg, fogyasztás, gyorsulás, végsebesség, ár) a nyíl segítségével mozgassuk be a VARIABLES dobozba. A LABEL CASES by dobozba kerüljön a típus, hiszen ez alapján szeretnénk címkézni. Ezt követően az OPTIONS modulban kérjük az ANOVA táblát és minden esetre vonatkozó klaszterinformációt is (CLUSTER INFORRMATION FOR EACH

[41] Az átlagot kivonjuk az egyes értékekből és elosztjuk a szórással, melynek eredményként a standardizált skála átlaga 0, szórása 1 lesz. Az SPSS-ben az Analyze/Classify/Hierarchial Cluster/Method/Transform Values/Standardize: Z Scores/ By Variable menüpont alatt tehetjük ezt meg.

CASE). Az ITERATE[42] dobozzal most nem foglakozzunk, hagyjuk meg az alapbeállításokat. Ezt követően a CONTINUE, majd az OK lenyomásával a következő eredményekhez jutunk:

3/11. táblázat

Initial Cluster Centers

	Cluster		
	1	2	3
Lökettérfogat (cm^3)	750	1298	1449
Telj (kW)	68	106	50
Telj (LE)	92	144	68
Nyomaték (Nm)	67	134	110
Tömeg (kg)	235	263	385
Fogy (l/100km)	4,8	4,9	5,4
Gyors. 0-100 km/h (s)	3,6	2,9	6,5
Végsebesség (km/h)	223	245	158
Ár (Ft)	1798000	3750000	7309000

A fenti első táblázat azt mutatja, hogy milyen középpontokból indult ki a program. Miután három klasztert kértünk, így természetesen ennyi középpontot hozott létre program, annyi változó mentén, amennyit bevontunk az elemzésbe.

A következő táblázat adatai alapján négy iterációra került sor.

3/12. táblázat

Iteration History

	Change in Cluster Centers		
Iteration	1	2	3
1	521368,4	86888,712	764600,0
2	78631,594	51211,558	340828,6
3	50000,000	64621,056	,000
4	,000	,000	,000

A Cluster Membership táblázatának segítségével láthatóvá válik, hogy az egyes típusokat mely klaszterben helyezte el a program. Itt a táblázat részletéből látszik a klaszter száma és a középpontjától vett távolság is. Ez alapján pl. az Aprilia RST 1000 Futura típusú motor az egyes klaszterben lesz.

---

[42] Az iteráció azt jelenti, hogy a program mindig újraszámolja a klaszterközéppontokat mindaddig, míg új elem kerül a klaszterhez. Ez egészen eltart addig, míg a középpontok nem változnak, vagyis stabil szerkezetet kapunk.

3/13. táblázat

Cluster Membership

Case Number	típus	Cluster	Distance
1	RST 1000 Futura	1	401000,0
2	RSV mille R	2	176621,1
3	Tornado 900 i.e.	2	722479,0
4	F 650 GS	1	266000,2
5	R 1100 S	2	378521,1
6	R 1150 RT	2	412479,0

Az ezt követő végleges klaszterközpontok táblázata nagyon fontos információkkal szolgál, hiszen segítségükkel jellemezhetjük és nevezhetjük el a keletkező klasztereket.

3/14. táblázat

Final Cluster Centers

	Cluster		
	1	2	3
Lökettérfogat (cm^3)	931	1071	1418
Telj (kW)	70	94	62
Telj (LE)	95	128	85
Nyomaték (Nm)	86	107	117
Tömeg (kg)	236	234	345
Fogy (l/100km)	5,7	5,7	6,1
Gyors. 0-100 km/h (s)	3,9	3,3	5,3
Végsebesség (km/h)	217	252	181
Ár (Ft)	2448000	3676521	6203571

Ennek alapján jól megkülönböztethető csoportokat lehet elkülöníteni:

1. klaszter (**„utcai motorok”**): ebbe a csoportba tartoznak a viszonylag olcsó, alacsony, illetve közepes teljesítményű motorok. Főleg az alacsonyabb lökettérfogatú (600-1000 cm^3) gépek alkotják ezt a csoportot. Közepes gyorsulással és végsebességgel rendelkeznek.
2. klaszter („**sport - túra motorok**”): ebbe a csoportba a nagy lökettérfogatú, nagy teljesítményű járművek tartoznak magas végsebességgel és nyomatékkal. Ezeket a járműveket általában a sportos beállítottságú, ám túrázni is kedvelő vásárlók választják.
3. klaszter („**országúti nehéz cirkálók**”): ebbe a csoportba tartoznak a nehéz, lassú, de nagy nyomatékkal, és rosszabb gyorsulással bíró motorok, óriási lökettérfogattal

és magas árral. Ők a tipikus nehéz cirkálók tulajdonosaik, akik egy külön „életérzéssel, életstílussal” is bírnak.

3/15. táblázat

Distances between Final Cluster Centers

Cluster	1	2	3
1		1228521	3755571
2	1228521		2527050
3	3755571	2527050	

A Distances between Final Cluster Centers táblázata azt bizonyítja, hogy a keletkezett klaszterek távol kerületek egymástól. A klaszterek közti távolságot mutatja ez a táblázat.

A következő táblázat hasonlít a már megismert Anova táblázatra, azonban hiányzik a már megszokott Sum of Squres és a Total oszlop. A tábla alatti magyarázó szöveg is felhívja a figyelmet arra, hogy nem egy hagyományos szignifikancia- tesztről van szó.

3/16. táblázat

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Lökettérfogat (cm	49108,695	2	55645,613	47	11,665	,000
Telj (kW)	4284,065	2	478,847	47	8,947	,001
Telj (LE)	7858,366	2	882,425	47	8,905	,001
Nyomaték (Nm)	3888,055	2	378,853	47	10,263	,000
Tömeg (kg)	36552,543	2	1238,199	47	29,521	,000
Fogy (l/100km)	,642	2	,683	47	,940	,398
Gyors. 0-100 km/	10,136	2	,991	47	10,226	,000
Végsebesség (km	14510,333	2	1299,967	47	11,162	,000
Ár (Ft)	,907E+013	2	,195E+011	47	178,009	,000

The F tests should be used only for descriptive purposes because the clusters ha maximize the differences among cases in different clusters. The observed signific corrected for this and thus cannot be interpreted as tests of the hypothesis that th

A Sig. alacsony értéke mutatja, hogy a klaszterközéppontok mindhárom klaszterképző mentén szignifikánsan különböznek. A táblázat adatai alapján azt tapasztaljuk, hogy a fogyasztás változótól eltekintve a többi változóban szignifikáns különbséget találunk. Ez alapján újra fogjuk futtatni az analízist a fent említett változó (fogyasztás) mellőzésével. A táblabeli F-értékek még jelezhetik számunkra, hogy mely változó mentén sikerült a legjobban elkülöníteni a klasztereket. Minél magasabb F-értéke, annál tökéletesebb az adott változó mentén a klaszter kialakítása, vagyis annál fontosabb a változó a klaszterezési eljárásban. Ez alapján az ár a legerősebb klaszterképző változó.

Ennek tudatában futassuk le ismét az analízist, immáron a fogyasztás változó nélkül. Az eddig magyarázott táblázatok értelmezése egyező. A létrejött új táblázatok közül az utolsóról még nem esett szó, amely a klaszterekben található egyedeknek a számát mutatja.

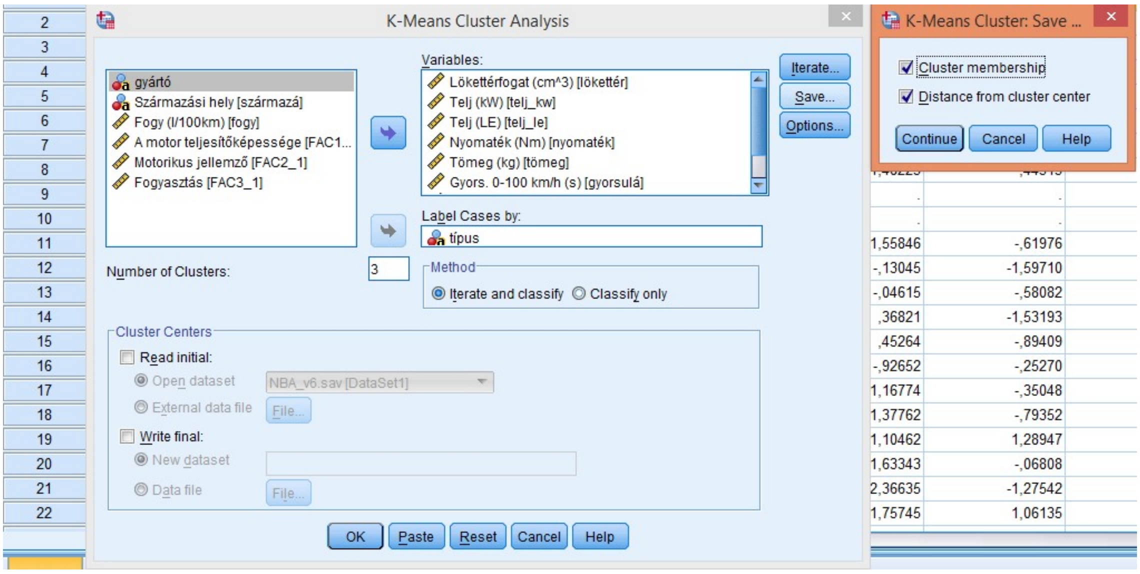
3/17. táblázat

Number of Cases in each Cluster

Cluster	1	24,000
	2	19,000
	3	7,000
Valid		50,000
Missing		3,000

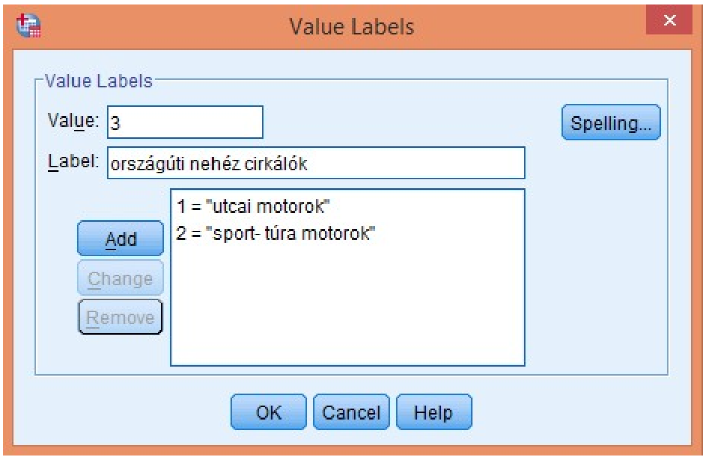
A program az ötven motort helyezett el három klaszter mentén. Három egyedet nem tudott a módszer besorolni, mert az áradatok nem ismertek. Az első klaszterbe (utcai motorok) 24 motor található, a másodikban (sport - túra motorok) 19, míg a harmadikban (országúti nehéz cirkálók) 7 darab.

A nagyobb gyártók sratégiájára is rávilágít ez az elemzés: a BMW öt terméke került be az adatbázisba, ebből egy „utcai motor”, egy „országúti nehéz cirkáló”, a többi pedig „sport - túra motor”, mint ahogyan azt vártuk. Az olasz Ducati csak az egyes klaszterbe tartozó motorokkal szerepel a vizsgálatban, míg az amerikai Harley-Davidson hat szereplő motorjából öt a hármas csoport tagja! Ne felejtsük el, hogy a hármas csoportnak mindössze hét eleme van. A Honda kilenc modellje között van egy „Harley-imitátor” (legalábbis a paramétereket tekintve), az összes többi azonban a másik két csoportba tartozik, ahogy a Kawasaki összes típusa is. A Suzuki szinte kivétel nélkül az egyes csoportba tartozó motorokat árusít, ahogy a Yamaha is (Mindez természetesen csak az adatbázisunk adataira vonatkozik.). Fontos, hogy a kialakult három klasztert elmentsük és a megfelelő névvel lássuk el őket. Jelöljük a SAVE dobozban a CLUSTER MEMBERSHIP és DISTANCE FROM CLUSTER CENTER, aminek következtében a klaszterek klaszterközepektől mért távolsága és a klaszterek száma mentésre kerül.

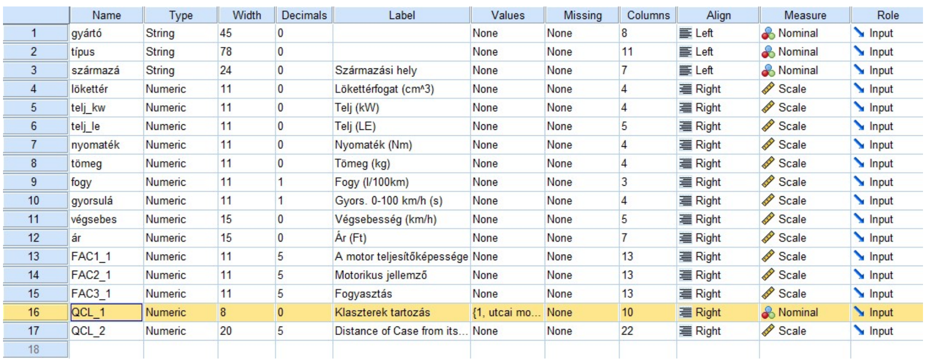


3/15. képernyőnézet: A létrejövő klaszterek mentése

A mentést követően a VARIABLE VIEW ablakban mentük el, a LABEL (címke) alatt a keletkező új faktorokat.



3/16. képernyőnézet: A klaszterek elnevezésének folyamata



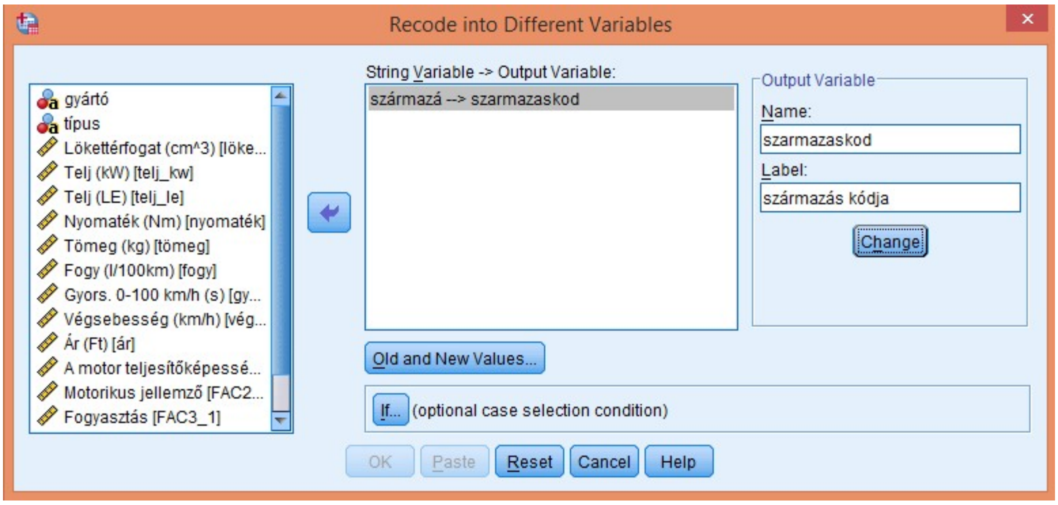
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	gyártó	String	45	0		None	None	8	Left	Nominal	Input
2	típus	String	78	0		None	None	11	Left	Nominal	Input
3	származá	String	24	0	Származási hely	None	None	7	Left	Nominal	Input
4	lökettér	Numeric	11	0	Lökettérfogat (cm^3)	None	None	4	Right	Scale	Input
5	telj_kw	Numeric	11	0	Telj (kW)	None	None	4	Right	Scale	Input
6	telj_le	Numeric	11	0	Telj (LE)	None	None	5	Right	Scale	Input
7	nyomaték	Numeric	11	0	Nyomaték (Nm)	None	None	4	Right	Scale	Input
8	tömeg	Numeric	11	0	Tömeg (kg)	None	None	4	Right	Scale	Input
9	fogy	Numeric	11	1	Fogy (l/100km)	None	None	3	Right	Scale	Input
10	gyorsulá	Numeric	11	1	Gyors. 0-100 km/h (s)	None	None	4	Right	Scale	Input
11	végsebes	Numeric	15	0	Végsebesség (km/h)	None	None	5	Right	Scale	Input
12	ár	Numeric	15	0	Ár (Ft)	None	None	7	Right	Scale	Input
13	FAC1_1	Numeric	11	5	A motor teljesítőképessége	None	None	13	Right	Scale	Input
14	FAC2_1	Numeric	11	5	Motorikus jellemző	None	None	13	Right	Scale	Input
15	FAC3_1	Numeric	11	5	Fogyasztás	None	None	13	Right	Scale	Input
16	QCL_1	Numeric	8	0	Klaszterek tartozás	{1, utcai mo...	None	10	Right	Nominal	Input
17	QCL_2	Numeric	20	5	Distance of Case from its...	None	None	22	Right	Scale	Input
18											

3/17. képernyőnézet: A klaszterek címkézése

3.3. Korrespondencia- analízis

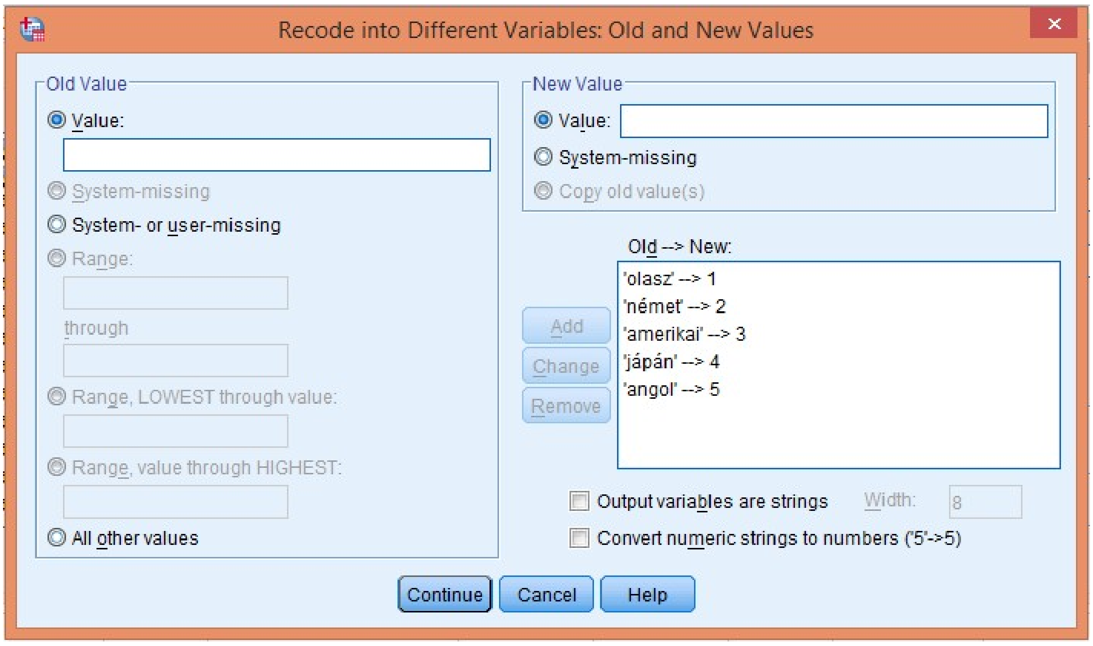
A módszer rövid leírása már az asszociációs kapcsolatoknál megtörtént, most ez tekinthetjük ezt újabb gyakorló példának is. Itt arra leszünk kíváncsiak, hogy a motort gyártó országok és a keletkező klaszterek között van- e összefüggés, vagyis hogy a gyártó nemzetisége mennyiben határozza meg a korábbi klaszter-analízis által nyújtott eredményeket. Kicsit másképpen fogalmazva, a motorokat gyártó cégek között megtörtént-e a piac szegmentálása?

Ehhez fel kell használni a klaszter- analízis során létrejövő klasztereinket, melyeknek már neveket is adtunk. Első lépésben át kell kódolni egy új változóba (származáskód) az adatbázisban szereplő országokat. A származási helyek kódolása a következőképpen történt: Az ANALYZE menüpont INTO DIFFERENT VARIABLE almenüjében tudjuk egy új változóba kódolni a származási helyeket.



3/18. képernyőnézet: A kódolás kiválasztása

A nyíl segítségével jelöljük ki azt a változót, melynek alapján az újat elkészítjük. Az OUTPUT VARIABLE dobozba nevezzük el az új változót (szarmazaskod), ügyeljünk arra, hogy az ékezet nem megengedett. A LABEL dobozban megadhatjuk a címke nevét is (itt használhatunk ékezetes karaktereket). Ezt követően a tényleges kódolást az Old and New Values opcióval tehetjük meg.



3/19. képernyőnézet: A származási helyek kódolása

Az OLD VALUE dobozba írjuk az eredeti változó nevét (ügyelve a helyesírásra), majd a NEW VALUE dobozba írjuk az új kódszámot. Ha ez megtörtént nyomjuk meg az ADD opciót, majd ismételjük meg mindaddig, amíg az összes változót át nem kódoltuk. Ha ezzel megvagyunk a CONTINUE és az OK opciók után létrejön az új változónk. Ezután nézzük meg létezik- e a kapcsolat a származási hely és a létrejött klaszterek között. Ennek megválaszolása a kereszttábla elemzés szimmetrikus mérőszámok táblázatából történik.

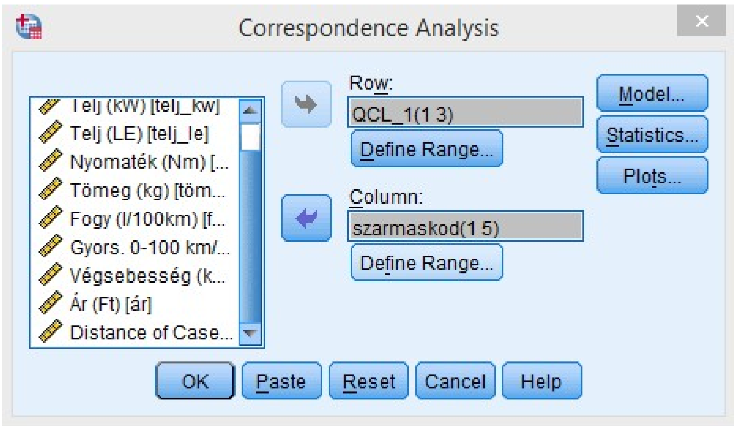
3/18. táblázat

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,803	,012
	Cramer's V	,568	,012
N of Valid Cases		20	

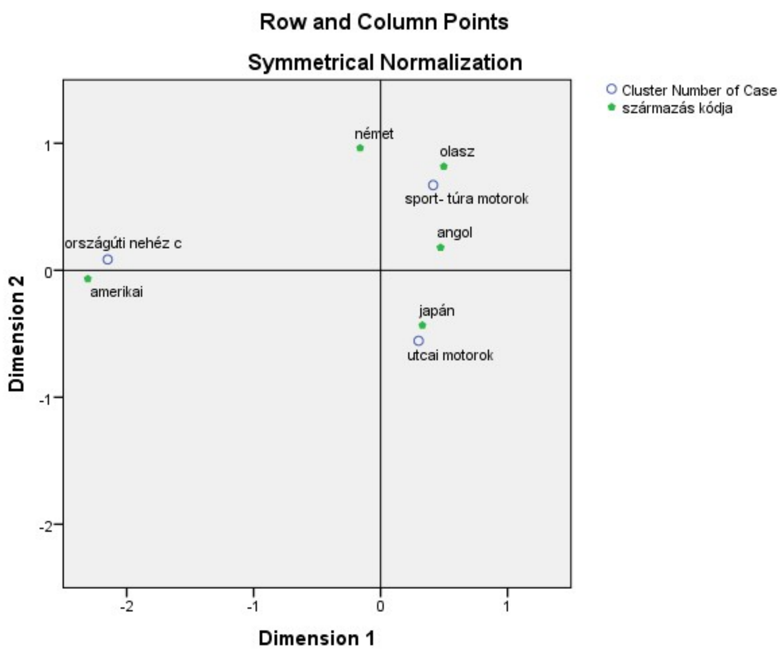
Ez alapján látható, hogy a kapcsolat létezik és közepes (0,57) erősségű. Ezt követően elkészítjük a korrespondencia- analízist.

Az eljárás a DIMENSION REDUCTION főmenü a CORRESPONDENCE ANALYSIS almenüjéből végezhető el, ahol először a sorváltozóba a keletkezett klaszterek kódját, míg az oszlopváltozónak az új származáskód változót jelöljük ki



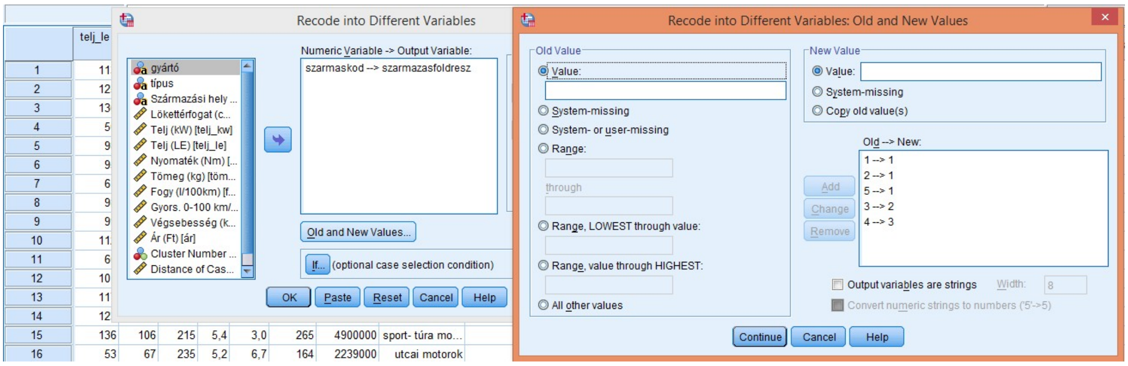
3/20. képernyőnézet: A korrespondencia- analízis beállításai

Ezután mindkét ismérvet definiálni kell, a benne szereplő ismérvváltozatok számának segítségével. A sor változót (klaszterek kódja) 1-től 3-ig, míg az oszlopváltozót (származáskód) 1-től 5-ig definiáljuk a már korábban bemutatott módón. A többi beállítást változatlanul hagyva futassuk le az elemzést. A keletkező eredmények közül a grafikus ábrázolást vizsgálva, láthatóvá válnak az összetartozó értékek.



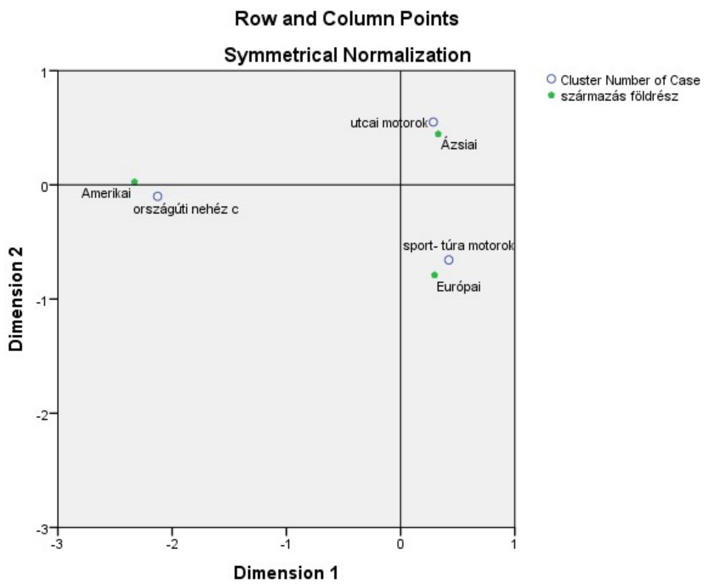
3/3. ábra: A korrespondencia- analízis grafikus megjelenítése

A dimenziók elnevezésétől függetlenül látszik, hogy a „sport- túra” motorok származási helye leginkább Európa, míg a „utcai” motorokat Japánban és az „országúti nehézcirkálókat” az Amerikában gyártják. Ennek tudatában felmerül az igény, hogy ábrázoljuk ezt a jelenséget a földrészek viszonyában is. Első lépésben ismét hozzunk létre egy új változó földrészkód névvel.



3/21. képernyőnézet: A földrészek szerinti kódolás

A kódoláskor az eddigi 1-es, 2-es és 5-ös kódszámú kapja az 1-es kódot (Európai), az eddigi 3-ad lesz a 2-es (Amerikai), míg a négyeseknek az új kódja legyen a kettes (Ázsiai). Az így létrejövő új változót a zárójelnek megfelelően címkézzük fel. Az így lefutatott korrespondencia- analízis grafikus ábrája a következő lesz:



3/4. ábra: A földrész és a klaszterek grafikus megjelenítése

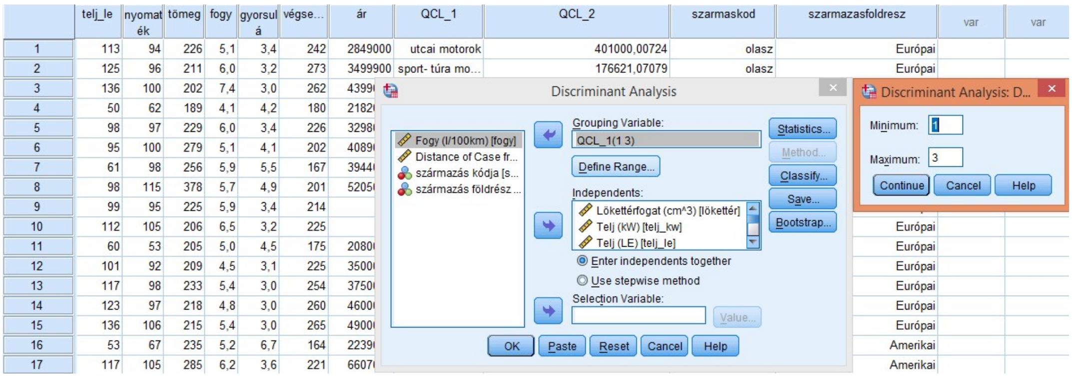
Jól látható, hogy az adatbázis adatai alapján a földrészek és a klaszterek jól megfeleltethetőek egymásnak. Az ábra jól mutatja, hogy a különböző életstílusoknak megfelelően a motorok gyártása, földrészek szerint szegmentálódott, gondoljunk csak az amerikai szokásokra. Talán jól ismerjük az amerikai autókat, és most láthatóvá vált, hogy a motorok terén sincs más ízlésük, illetve választási szokásuk az ott élő embereknek.

3.4. Diszkriminancia- analízis

A diszkriminancia-analízis olyan sokváltozós adatelemzési módszer, melyet leginkább a csoportok szétválasztására és a kategóriába tartozás előrejelzésére alkalmaznak. Megpróbálja a függő változók értékeit, a független változók értékeivel magyarázni, vagyis arra keresi a választ, hogy a csoporthoz tartozás előre becsülhető-e, és ha igen, hány százalékban az adott független változókkal. Ebben nem csak az a cél, hogy a változók közötti összefüggést felfedezzük, hanem az is, hogy a függő változók ismeretlen értékeit a független változók értékei alapján előre megmondjuk. A módszer hasonlít varianciaelemzéshez, illetve a sokváltozós regresszióhoz, az utóbbihoz főleg az egyenes illesztés problematikája miatt.

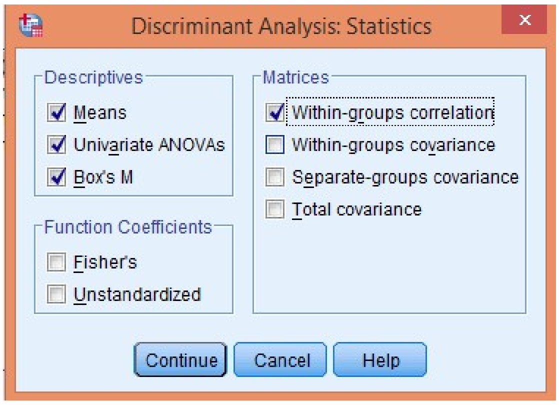
A diszkriminancia- analízis jóságáról nyerhetünk képet akkor, ha az analízis által feltételezett csoport hovatartozást összehasonlítjuk a valóságos hovatartozással. A diszkriminancia- analízishez hasonló a logisztikus regresszió is, melynek alkalmazásának nincsenek olyan szigorú előfeltételei. Míg a diszkriminancia- analízisnél a függő változót nominális, a független változót intervallum- vagy arányskálán mérjük, addig a logisztikus regressziónál a független változó között lehet nominális és ordinális skálán mért változó is.

Példánkat folytatva azt vizsgáljuk, hogy a motorok paramétereinek ismeretében (lökettérfogat, teljesítmény (kW), teljesítmény (LE), nyomaték, tömeg, gyorsulás, végsebesség, ár), megbecsülhető- e, hogy melyik klaszterhez (utcai motorok, sport- túra motorok, országúti nehézcirkálók) tartozik. A vizsgálatot az ANALYZE menü, CLASSIFY almenüjának, DISCRIMINANT moduljából érhetjük el. Először a csoportosító (függő változó) változóként adjuk meg a létrejött klasztereket, melyeket definiálnuk is kell (Define Range), annak megfelelően, hogy mennyi klaszterünk keletkezett. Itt adjuk meg minimum értékként az egyet, maximumként a hármat. A független változóinkat az Independents mezőbe mozgatjuk a nyíl segítségével. (Forrás: motor.sav)



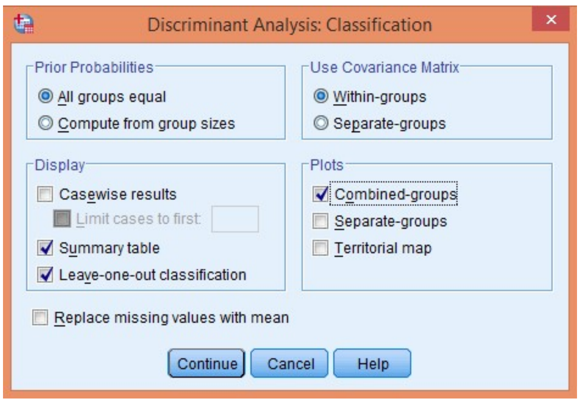
3/22. képernyőnézet: A diszkriminancia- analízis beállítása

Ezek után a STATISTICS menüpontban a DECREPTIVES lehetőségek közül jelöljük ki mindet, hiszen így az elemzés előfeltételeit tesztelhetjük.



3/23. képernyőnézet: Az előfeltételek beállításai

A MATRICES opciók közül a csoporton belüli korrelációt (Within- groups correlation) jelöljük. Legvégül a CLASSIFY menüben a következő lehetőségeket kell kijelölni:



3/24. képernyőnézet: Az analízis csoportosításainak beállításai

Az alapbeállításokat meghagyva a Display opciók közül kérjük az összesítő táblát (Summary table), mely a megfelelően elhelyezett esetekről közöl információt, illetve a Leave-one-out classification, amely szintén erről szolgáltat információkat. A grafikus megjelenítéshez a Combined- groups kérhetjük, amely a csoportok elhelyezkedését ábrázolja a keletkező diszkriminancia- függvények tükrében. Ezt követően lefuttatva az elemzést számtalan táblázatot kapunk, melyek közül a leglényegesebbeket tárgyaljuk részletesen.

Az első táblázat (Analysis Case Processing Summary) az egyszerű, alapstatisztikákat mutatja, mint az érvényes (50), és hiányzó (3) esetszámot. A következő táblázat (Group Statistics) az elemzésbe bevont összes változó csoportok szerinti és összesített átlagát, szórását, súlyát mutatja.

3/19. táblázat

Group Statistics

Cluster Number of Case		Mean	Std. Deviation	Valid N (listwise) Unweighted	Weighted
utcai motorok	Lökettérfogat (cm^3)	930,958	282,2981	24	24,000
	Telj (kW)	69,625	21,5786	24	24,000
	Telj (LE)	94,750	29,3054	24	24,000
	Nyomaték (Nm)	85,833	21,9934	24	24,000
	Tömeg (kg)	235,958	35,1388	24	24,000
	Gyors. 0-100 km/h (s)	3,917	1,1309	24	24,000
	Végsebesség (km/h)	216,750	38,1567	24	24,000
	Ár (Ft)	2448000,000	351868,6761	24	24,000
sport- túra motorok	Lökettérfogat (cm^3)	1070,789	160,4215	19	19,000
	Telj (kW)	94,474	23,1789	19	19,000
	Telj (LE)	128,421	31,4984	19	19,000
	Nyomaték (Nm)	107,474	14,3231	19	19,000
	Tömeg (kg)	234,053	31,9435	19	19,000
	Gyors. 0-100 km/h (s)	3,274	,7377	19	19,000
	Végsebesség (km/h)	252,158	36,2495	19	19,000
	Ár (Ft)	3676521,053	499190,3473	19	19,000
országúti nehéz cirkálók	Lökettérfogat (cm^3)	1418,429	230,6483	7	7,000
	Telj (kW)	62,286	18,8212	7	7,000
	Telj (LE)	84,857	25,3734	7	7,000
	Nyomaték (Nm)	117,000	22,3159	7	7,000
	Tömeg (kg)	345,286	43,6452	7	7,000
	Gyors. 0-100 km/h (s)	5,257	1,1088	7	7,000
	Végsebesség (km/h)	181,286	25,6886	7	7,000
	Ár (Ft)	6203571,429	705079,8671	7	7,000
Total	Lökettérfogat (cm^3)	1052,340	282,6103	50	50,000
	Telj (kW)	78,040	25,1826	50	50,000
	Telj (LE)	106,160	34,1637	50	50,000
	Nyomaték (Nm)	98,420	22,8492	50	50,000
	Tömeg (kg)	250,540	51,7649	50	50,000
	Gyors. 0-100 km/h (s)	3,860	1,1681	50	50,000
	Végsebesség (km/h)	225,240	42,8855	50	50,000
	Ár (Ft)	3440618,000	1343598,867	50	50,000

Az ezt követő táblázatban azt vizsgálhatjuk, hogy a független változók milyen mértékben járulnak hozzá a létrejövő függvényhez. A változók szignifikáns voltának tesztelésére az F-érték mellett, a Wilks'- Lambda statisztika is szerepel.

3/20. táblázat

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Lökettérfogat (cm^3)	,668	11,665	2	47	,000
Telj (kW)	,724	8,947	2	47	,001
Telj (LE)	,725	8,905	2	47	,001
Nyomaték (Nm)	,696	10,263	2	47	,000
Tömeg (kg)	,443	29,521	2	47	,000
Gyors. 0-100 km/h (s)	,697	10,226	2	47	,000
Végsebesség (km/h)	,678	11,162	2	47	,000
Ár (Ft)	,117	178,009	2	47	,000

Látható, hogy minden változónak szignifikáns hatása van. A Wilks'sLambda értéke 0 és 1 közé eső értékek, melyek közül a mindig a nullához közeli értékekhez tartozó változóknak van a legjelentősebb hatása diszkriminancia- függvényre.

3/21. táblázat

Pooled Within-Groups Matrices

		Lökettérfogat (cm^3)	Telj (kW)	Telj (LE)	Nyomaték (Nm)	Tömeg (kg)	Gyors. 0-100 km/h (s)	Végsebesség (km/h)	Ár (Ft)
Correlation	Lökettérfogat (cm^3	1,000	-,058	-,058	,841	,792	,289	-,280	,239
	Telj (kW)	-,058	1,000	1,000	,426	-,213	-,822	,933	,049
	Telj (LE)	-,058	1,000	1,000	,426	-,214	-,821	,933	,048
	Nyomaték (Nm)	,841	,426	,426	1,000	,637	-,145	,173	,252
	Tömeg (kg)	,792	-,213	-,214	,637	1,000	,432	-,408	,175
	Gyors. 0-100 km/h (	,289	-,822	-,821	-,145	,432	1,000	-,856	,018
	Végsebesség (km/h	-,280	,933	,933	,173	-,408	-,856	1,000	-,015
	Ár (Ft)	,239	,049	,048	,252	,175	,018	-,015	1,000

A következő két táblázatban két alapfeltevés tesztelése történik. A Pooled Within- Groups Matrices táblázat a multikollinearitást teszteli. A következő táblázat a variancia- kovariancia mátrixok homogenitását (homoszkedaszcticitás) teszteli a Box'M mutató segítségével.

A következő fontos táblázat (Eigenvalues), mely során először kapunk információt a keletkező függvényről.

3/22. táblázat

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	8,603[a]	89,5	89,5	,946
2	1,005[a]	10,5	100,0	,708

a. First 2 canonical discriminant functions were used in the analysis.

A táblázatból látszik, hogy két függvény keletkezett. A függvények számát megállapíthatjuk, ha a csoportok száma, illetve a független változók száma közül a kevesebbikből egyet kivonunk. A két függvény fontosságának megállapításában, a sajátérték segíti a kutatót. A táblázat sajátértékei és magyarázott variancia értékei alapján az első függvény lesz fontosabb számunkra. A kanonikus korreláció (0,946) azt jelenti, hogy az adott függvény igen számottevő részt magyaráz a teljes varianciából. A kapott érték négyzete megmutatja, hogy a függő változó varianciájának, hány százalékát magyarázzák a független változók csoportja (89,49%).

3/23. táblázat

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,052	130,133	14	,000
2	,499	30,604	6	,000

A megjelenő Wilks' Lambda táblázat a függvények szignifikanciájának tesztelését végzik. Láthatóan mindkét függvény szignifikáns, de az első hatása jelentősebb.

A következő táblázatban (Standardized Canonical Discriminant Function Coefficients), a standardizált együtthatók segítségével megállapíthatjuk, hogy melyik változók különböztetik meg leginkább a csoportokat.

A korrelációs együttható mátrixa (Structure Matrix) hasonlóan értelmezendő, mint a faktor-analízisnél a Component Matrix, hiszen a független változók és a diszkriminancia-függvények közti, csoportonként átlagolt (Pooled within- groups) Pearson- féle lineáris korrelációk.

3/24. táblázat

Structure Matrix

	Function	
	1	2
Ár (Ft)	,932*	,307
Lökettérfogat (cm^3)	,240*	,038
Végsebesség (km/h)	-,106	,613*
Telj (LE)[a]	-,032	,610*
Telj (kW)	-,031	,609*
Gyors. 0-100 km/h (s)	,150	-,491*
Tömeg (kg)	,355	-,415*
Nyomaték (Nm)	,190	,355*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

a. This variable not used in the analysis.

Ez alapján az első függvény az árat és a lökettérfogatot, míg a második az összes többit - kivétel a teljesítményt lóerőben- foglalja magában, mely alapján a kutató a dimenziókat elnevezheti (hasonlóan a faktor- analízishez).
A következő táblázat (Functions at Group Cetroids) a csoportok középpontértékeit tartalmazza.

3/25. táblázat

Functions at Group Centroids

Cluster Number of Case	Function 1	2
utcai motorok	-2,030	-,736
sport- túra motorok	,132	1,241
országúti nehézcirkálók	6,602	-,843

Unstandardized canonical discriminant functions evaluated at group means

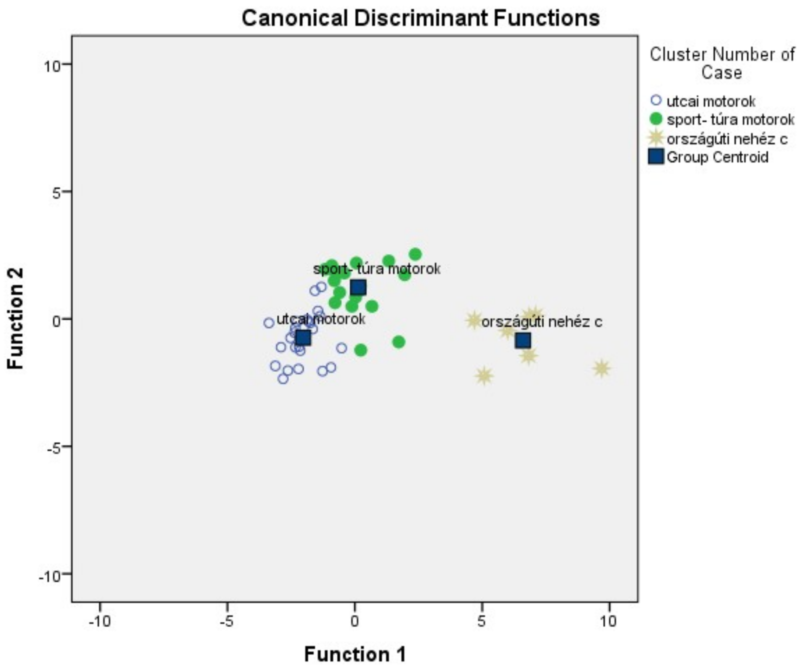
Megállapíthatjuk, hogy az első és harmadik csoport magas értékkel rendelkezik az első dimenzióban, míg a sport- túra motorok magas értékei a második dimenzió mentén jelentkeznek. A későbbi grafikus megjelenéshez ezeket a koordinátákat használja fel a program.
A következő részben a klasztifikációs statisztika következik, amely az analízisünk legfontosabb része. Az első táblázat (Prior Probabilities for Groups) a kiinduló értékeket tartalmazza.

3/26. táblázat

Prior Probabilities for Groups

Cluster Number of Case	Prior	Cases Used in Analysis Unweighted	Weighted
utcai motorok	,333	24	24,000
sport- túra motorok	,333	19	19,000
országúti nehézcirkálók	,333	7	7,000
Total	1,000	50	50,000

Látszik, hogy a csoportokba kerülés esélye 33,3 százalék volt. A következőben a grafikus ábrázolás történik, ahol a tengelyek maguk a függvények (dimenziók).



3/5. ábra: A diszkriminancia- analízis grafikus megjelenítése

Az ábra az analízisbe bevont egyedek értékeit és a centrumközéppontokat ábrázolja. A helyesen kategorizált csoporttagságok arányát a klasszifikációs eredmények elnevezésű táblázatban (Classification Results) láthatjuk.

3/27. táblázat

Classification Results[b,c]

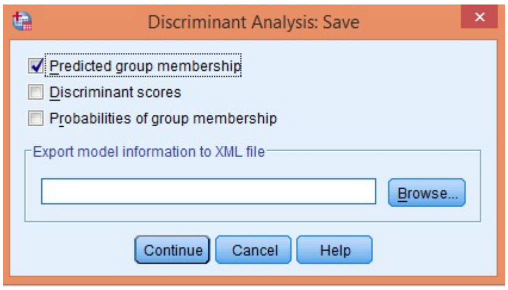
			Predicted Group Membership			
		Cluster Number of Ca	utcai motorok	sport- túra motorok	országúti nehézcirkálók	Total
Original	Count	utcai motorok	22	2	0	24
		sport- túra motorok	1	18	0	19
		országúti nehézcirkáló	0	0	7	7
	%	utcai motorok	91,7	8,3	,0	100,0
		sport- túra motorok	5,3	94,7	,0	100,0
		országúti nehézcirkáló	,0	,0	100,0	100,0
Cross-validated[a]	Count	utcai motorok	21	3	0	24
		sport- túra motorok	1	18	0	19
		országúti nehézcirkáló	0	0	7	7
	%	utcai motorok	87,5	12,5	,0	100,0
		sport- túra motorok	5,3	94,7	,0	100,0
		országúti nehézcirkáló	,0	,0	100,0	100,0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is the functions derived from all cases other than that case.

b. 94,0% of original grouped cases correctly classified.

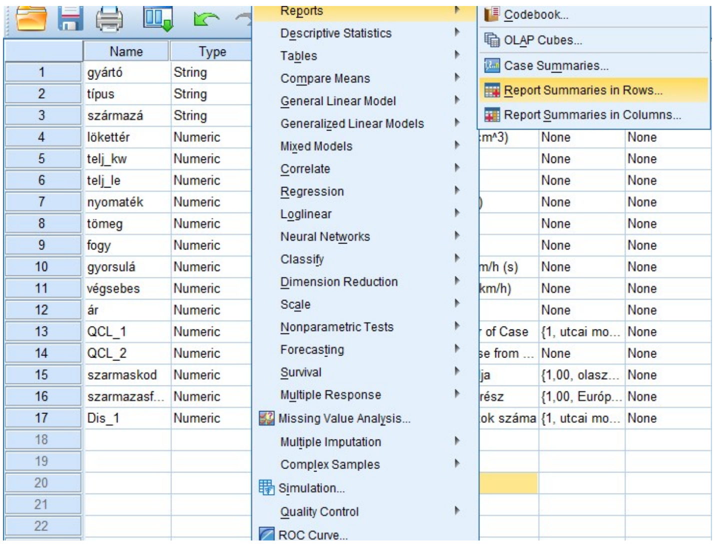
c. 92,0% of cross-validated grouped cases correctly classified.

A táblázat alján láthatjuk, hogy a modell 94%-ban tudta helyesen kategorizálni a megadott független változó mentén. Ezt az összevetést úgy végzi, hogy a kiinduló (original) csoportba tartozást hasonlítja a diszkrimináló függvény segítségével történő (Cross- validared) besorolással. Azt jelenti (átlókon elhelyezkedő értékeket nézve), hogy az utcai motorok (24 db) közül 21 került jó csoportba 3 nem, ami 87,5 %. A sport-túra motorok (19 db) közül 18 jó csoportba 1 nem megfelelőbe került (94,7%), míg az országúti nehézcirkálók közül az összes jó csoportba lett sorolva (100%). A három csoport helyes találati aránya 94%. A táblázat alatti harmadik állítás 92%-a, jelzi azt, hogy a CLASSIFY menüben bejelöltük a Leave-One-Out opciót, amely szintén az előző keresztérvényességet teszteli. Ez a százalék általában kisebb szokott lenni, mint a felette lévő, mivel szigorúbban mér. Ennek menete, egy- egy megfigyelési egység kihagyásával ismételten elvégzi az elemzést. Ezek után mentsük el (SAVE) a függvénnyel becsült csoportok számát.



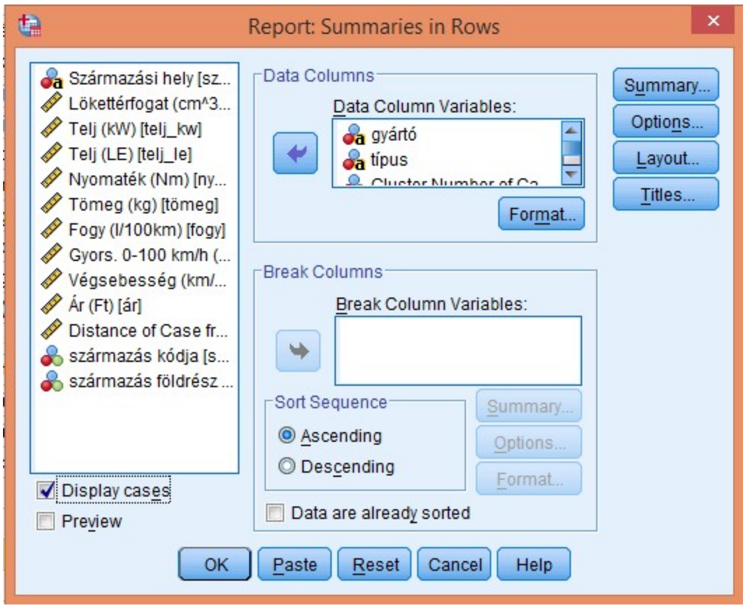
3/25. képernyőnézet: A becsült csoportok számának mentése

Ennek eredményeként a Data Editor ablakban létrejön egy új változó (Dis_1), melyet „címkézzünk” fel (Label), a „becsült csoportok száma” névvel.
Most listáztassuk ki az eredeti és becsült csoportba tartozásokat. Ezt többféleképpen is megtehetjük az ANALYZE menü REPORTS almenüjének segítségével. Először kérjünk egy leíró statisztikát sorba rendezve (Report Summaries in Rows).



3/26. képernyőnézet: Az eredeti és becsült csoportba tartozás megjelenítése

Az ezt követő beállításoknál a nyíl segítségével adjuk meg, hogy mely változók szerepeljenek az oszlopokban, vagyis kérjünk listát a keletkezett a gyártóról, a típusról, a klaszterek száma, illetve becsült csoportok száma változókról.



3/27. képernyőnézet: A listán szereplő változók beállításai

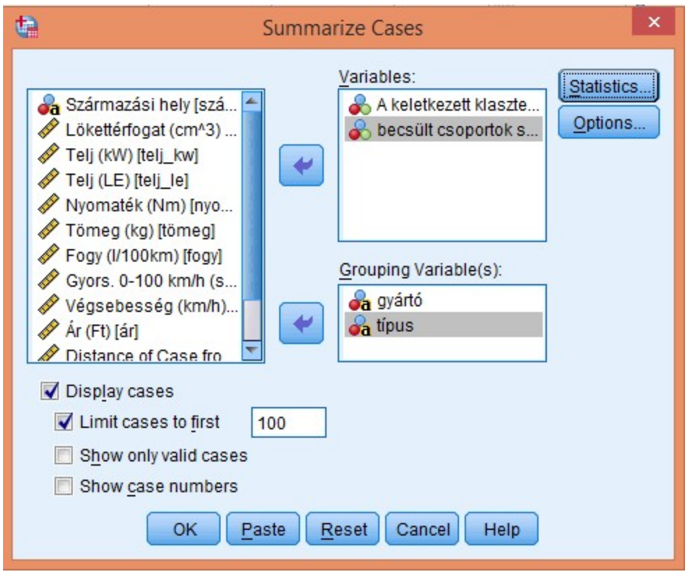
A többi lehetőséget most nem változtatva az OK gomb lenyomása után a következő eredményt kapjuk az Output ablakban:

3/28. táblázat

Page 1

GYÁRTÓ	TÍPUS	A keletkezett klaszterek száma	Becsült csoportok száma
Aprilia	RST 1000 Futura	1	1
Aprilia	RSV mille R	2	2
Benelli	Tornado 900 i.e.	2	2
BMW	F 650 GS	1	1
BMW	R 1100 S	2	2
BMW	R 1150 RT	2	2
BMW	R 1200 C Independent	2	2
BMW	K 1200 LT	3	3
Cagiva	Navigator 1000	.	.
Cagiva	Raptor 1000	.	.
Ducati	Monster 620 Dark	1	1
Ducati	Monster S4	2	2
Ducati	ST 4 S	2	2
Ducati	998	2	2
Ducati	999	2	2
Harley-Davidson	XL 883R Sportster	1	1
Harley-Davidson	VRSCA V-Rod	3	3
Harley-Davidson	Night Train	3	3
Harley-Davidson	Fat Boy	3	3
Harley-Davidson	Road King FLHR	3	3
Harley-Davidson	Electra-Gide Ultra Classic	3	3
Honda	Hornet 600	1	1
Honda	CBR 600 F	1	1
Honda	XRV 750 Africa Twin	1	1
Honda	VFR/CBS-ABS	1	1
Honda	Fireblade	2	2
Honda	VTR 1000 SP-2	2	2
Honda	X-11 CBS	1	1
Honda	CBR 1100 XX	1	2
Honda	GL 1800 Gold Wing	3	3

Az eredmény részletén is jól látható, hogy fent feltüntet ismérvek szerint egy egyszerű felsorolást végzett a program. Lényegesen szebb listázást is elvégezhetünk a REPORTS almenü, CASE SUMMERIES moduljával, hiszen itt egy vagy több csoportképző által megjelölt kategóriákon belüli statisztikákat kérhetünk táblázatos formában.



3/28. ábra: Összesítő táblázat beállításai

A változók dobozba a keletkezett klaszterek száma, illetve a becsült csoportok száma, míg a csoportosító változó dobozba a gyártó és a típus ismérvek kerüljenek. A következőben az így keletkező táblázatnak a részlete látható.

3/29. táblázat

Case Summaries[a]

					A keletkezett klaszterek száma	Becsült csoportok száma
gyártó	Aprilia	típus	RST 1000 Futura	1	utcai motorok	utcai motorok
			RSV mille R	1	sport- túra motorok	sport- túra motorok
	Benelli	típus	Tornado 900 i.e.	1	sport- túra motorok	sport- túra motorok
	BMW	típus	F 650 GS	1	utcai motorok	utcai motorok
			R 1100 S	1	sport- túra motorok	sport- túra motorok
			R 1150 RT	1	sport- túra motorok	sport- túra motorok
			R 1200 C Independent	1	sport- túra motorok	sport- túra motorok
			K 1200 LT	1	országúti nehézcirkálók	országúti nehézcirkálók
	Cagiva	típus	Navigator 1000	(missing) 9	.	.
			Raptor 1000	(missing) 10	.	.
	Ducati	típus	Monster 620 Dark	1	utcai motorok	utcai motorok
			Monster S4	1	sport- túra motorok	sport- túra motorok
			ST 4 S	1	sport- túra motorok	sport- túra motorok
			998	1	sport- túra motorok	sport- túra motorok
			999	1	sport- túra motorok	sport- túra motorok

Az így keletkező táblázatból könnyen leolvasható, hogy a diszkriminancia- analízis mely típusú motorokat sorolta az eredetivel nem egyező csoportba.

3.5. A kutatási eredmények, jelentések közzététele és prezentációja

A kutatási eredményeket általában a kutatási jelentésekben adják közzé. A kutatási beszámoló célja a kutatási eredmények, következtetések, javaslatok interpretálása a megbízó, illetve a meghatározott célközönség felé. Mivel az olvasó összetétele változó, így többféle kutatási jelentés is elképzelhető, attól függően, hogy kik, hol és mikor olvassák. Ennek megfelelően a beszámoló célja többféle lehet:

- elszámolás a kutatás támogatóival,
- a sportszakma tájékoztatása (pl. szakmai folyóiratban, szakcikk formájában),
- szakdolgozat, diploma, záródolgozat, tudományos fokozat szerzése.

A kutatási jelentésnek mindenképpen ki kell térnie:

1. a kutatási téma meghatározására, illetve az ezzel kapcsolatos hipotézisek megfogalmazására,
2. a kutatási módszerek megfogalmazására (minta, eszközök, kutatás lebonyolítása),
3. az adatok elemzésére,
4. a kutatás jelentőségére, újdonságára (várható eredmények, lehetséges alkalmazások).

A leggyakrabban előforduló hiba a kutatási jelentések készítésekor a terjedelem. A terjedelem helyes megítélése gyakran nehéz feladat, hiszen a túl rövid jelentés készítésekor az a benyomás alakulhat ki, hogy a készítő nem végezte el a feladatot tisztességesen. A másik probléma éppen az ellenkezője lehet, ha túlságosan is hosszúra sikerül a kutatási beszámoló, akkor előfordulhat, hogy végigolvasás időtartama miatt nem kerül rá sor. „Ez a jelentés már a terjedelménél fogva sincs kitéve a veszélynek, hogy bárki elolvassa."[43]

A készítőnek törekednie a hitelesség miatt is a korrekten és nyelvi hibáktól mentesen megfogalmazott közérthető mondatokra. Feltétlenül fejben kell tartani azt a lehetőséget, hogy nem biztos, hogy létezik egyetlen elmélet, ami magyaráz minden lehetséges kimeneti eredményt, viselkedést. Sajnos sok kutató egyetlen elmélet helyességében bízik, amikor megírja a kutatásának az összegzését.

[43] Winston Churchill

A kutatási beszámolót általában annak a bizottságnak a tagjainak szokták elküldeni, amely engedélyezte a kutatást, illetve azoknak a személyeknek, amelyek támogatták a kutatást, vagy érdeklődést mutattak a megvalósítás és az eredmények iránt, vagy akiknek kötelezően le kell adni (egyetemi, - főiskolai diploma vagy szakdolgozat formában).

Az első oldal tartalmazza a tanulmány, dolgozat címét, a kutatók nevét, címét és elérhetőségét (a sorrendben még a kutatás lefolytatása előtt meg kell állapodni).

Néha a harmadik és ötödik oldal közötti részt az előszónak szentelik a szerzők. Az előszóban kapnak helyet azok a személyek is, akik bármilyen formában hozzájárultak a kutatás véghezviteléhez és nem lettek megemlítve a szerzők között. Általában a kutatók hálájukat fejezik ki a résztvevőknek, a technikai stábnak és a programok, tanszékek, karok és intézmények igazgatóinak. Az előszóban a szerző(k) ismertetik az általuk vizsgált kérdésekre vonatkozó ötleteiket és gondolataikat, hogy miért olyan fontosak ezek a kérdések. Ez a rész egy személyesebb módon mutatja be az olvasónak a motivációkat és okokat, amelyek a tanulmány témaválasztásához, megírásához vezettek.

A következő oldalon következik a tényleges kutatás leírása, mely a klasszikus módón tagolódik: bevezetés, tárgyalás, befejezés (konklúzió, javaslatok). A kutatás ténymegállapításainak (felfedezéseinek) interpretálása mindig a kutatás összes hipotézisei magyarázatának, bizonyításának vagy megcáfolásának „integrálásával", egy egységbe rendezésével kell hogy végződjön. Tehát ezt a szakaszt egy integrált holisztikus (teljeskörű, mindenre kiterjedő) szemlélet zárja le, amely magában foglalja a ténymegállapítások összegzését, és kijelöli a jövőbeni irányokat egyaránt. Közvetlenül a konklúzió és a javaslatok után következik a felhasznált irodalmak listája, amely tartalmazza a teljes tanulmány során felhasznált összes irodalmat. Az irodalomjegyzék után a függelék következik, ahol megtalálható az eredeti kérdőív, dokumentumokról, felszerelésről készült fényképek, és minden olyan egyéb anyag, amely fontos lehet a részletek iránt érdeklődő olvasók és kutatók számára.

A kutatási jelentések általában lerövidítésre és átdolgozásra kerülnek a folyóiratokban és könyvben történő publikálás céljára. A legfontosabb elv, amelyet be kell tartanunk, hogy kevesebb szót használva, de a lényegi mondanivalót meghagyva kell átalakítanunk a kutatási jelentést. Ha az eredeti jelentés sok analízist, táblázatot, ábrát, interjút, idézetet vagy észrevételt tartalmaz, javasolt két-három publikáció készítése, melyek címe hasonló, de I., II. vagy III-as számmal vannak megjelölve, jelezve, hogy a két vagy három tanulmány egy tanulmánysorozat része. A rövidebb változat ugyanazt a felosztást követi, mint a hosszabb kutatási jelentés, azonban csak a főbb elméleti okfejtést, a hipotéziseket, módszereket, eredményeket, következtetéseket, és az irodalomjegyzéket kell tartalmaznia.

A kutatásról készülő **rövid összegzést, absztraktnak** nevezzük. Ezt az összegzést, amely 250-500 szót tartalmaz, a tanulmány befejezése után kell megírni. A következő logikai sorrendet követi: a tanulmány célja, röviden a módszertanról (kutatási terv), főbb eredmények, konklúzió, majd következtetés. Általában hasznos, ha megjelölünk egy vagy két megbízható tanulmányt, amelyet alapként használtunk. Nagyon fontos lefektetni, hogy az eredmények alátámasztották-e az alapfeltevéseket és amennyiben nem, egy-két mondatban tegyünk javaslatot a jövőben követendő teendőkre.

A kutatási módszereket leggyakrabban **szóbeli prezentáció** kíséri. Az orális prezentáció a következő lépésekből áll:

1. Célmeghatározás.
2. A hallgatóság feltérképezése.
3. A struktúra meghatározása.
4. Segédanyagok megírása.
5. Vizuális eszközök előkészítése.
6. Prezentáció elpróbálása.
7. Prezentáció helyének előkészítése.
8. Hang bemelegítése.
9. Prezentálás.
10. Kérdések megválaszolása.
11. Utógondozás.

Forrás: Sajtos L.- Mitev A (2007), 381. o.

Régebben egy befejezett kutatást egy egyszerű előadás követett, mely először a jegyzetekből való felolvasást jelentette, majd a technika fejlődésével megjelentek a táblák, írásvetítők, diavetítők, majd a videók, mint segédeszközök. Manapság a számítógépek korában a leggyakoribb szemléltetőeszköz a laptop és a projektor. A számítógépes prezentációk olyan felhasználóbarát szoftverekkel készülnek, mint például a PowerPoint. A PowerPointban történő bemutató- készítés néhány előnye a hagyományos szóbeli prezentációval szemben:

- a gyermekeket már korán (általános, illetve középiskolában) megismertetik az alkalmazásával,
- az elkészített prezentációk könnyen megőrizhető, módosítható,
- a szövegek megjelenítése időzíthető,

- a minősége sokkal jobb, mint az írásvetítőé,
- egyszerűen tudunk megjeleníteni diagramokat, ábrákat, képeket, videókat, animációkat,
- stb.

Általában a szóbeli prezentációk időbeli korlátokhoz kötöttek. A prezentáló számára rendelkezésre álló idő átlagosan 10 és 45 perc között mozoghat, függően az előadás jellegétől (konferencia, diplomavédés, projektzáró előadás, stb.), a prezentáló státuszától és egyéb tényezőktől, melyeket minden alaklommal vegyünk figyelembe.

A prezentációk általános formai szempontjai a következőek:

- lehetőség szerint azonos hátteret, betűtípust, formátumot, animációt alkalmazzunk.
- a betűk méreteit a terem adottságaihoz kell kialakítani, de általában a 36 betűméret feletti méreteket érdemes alkalmazni, így a terem végében ülő hallgatóság is látni fogja.
- érdemes multimédiás elemeket is alkalmazni (animáció, grafikon, diagram, táblázat, mozgókép, hang),
- felsorolások alkalmazása során törekedjünk az egyetlen szavas felsorolásokra,
- átlagosan 15-25 szó egy dián
- 6X6 szabály (maximálisan 6 szó 6 sorban)

Az általánosan követendő irányvonalak a számítógépes szemléltetés elkészítése során a következők:

- A nyitó dia tartalmazza a tanulmány címét, a szerzőket és elérhetőségeiket
- 1-3 dián keresztül összegezzük az alapvető elméleti pontokat és a hipotéziseket. Nagyon fontos, hogy minden dia csak minimális mennyiségű szót tartalmazzon, a túl sok információ rontja a prezentációt.
- 4-5 dián vázoljuk a kutatási tervet, ha lehet grafikus formában, pl. résztvevők mintavétele, eszközhasználat, kutatás kivitelezése, stb.
- 6-8 diát szenteljünk az eredmények (ténymegállapítások) ismertetésének. Minden dia csak egy ábrát vagy táblázatot tartalmazzon (preferáljuk az ábrát a táblázattal szemben). Az ábrák érthetőbbek és sokkal többet mondanak.
- 9-10 dia legyen a konklúzió, eredmények és azok újszerűségre, javaslatok.

- 11. dia hivatkozott irodalom, elérhetőség, egyéb (opcionális)
- 12. figyelem megköszönése

Lehetőségekhez mérten a papírról való felolvasást kerüljük, hiszen az érdeklődés rovására megy, és az előadó kompetenciáját csorbítja. A szóbeli prezentáció során a grafikus megjelenítéseket kell előnyben részesíteni, de kerüljük az apró betűs táblázatokat és ábrákat. A táblázatok és grafikonoknál a formai követelmények megtartása mellett a vizsgálatba bevont egyedek elemszámát (N) is ajánlott tüntetni. Kiemelten figyeljünk a betűméret beállítására, hiszen a nem megfelelő betűnagyság alkalmazásával az egész prezentáció „súlytalanná" válik.

Válasszuk meg azokat a kifejezéseket és mondatokat, amelyek meghatározzák a prezentációnkat. Változtassuk a hangszínünket, hogy elkerüljük a monotonitást. Legyünk meggyőződve arról, hogy a tanulmányunk és eredményeink helytállóak, mutassunk lelkesedést. ***Tartsunk előtte egy próbát!***

A szóbeli prezentáció a kutatási folyamat egyik, ha nem a legfontosabb része, éppen ezért még alaposabb tervezést és előkészítést igényel. A jó prezentáció során az előadó felkészültségét sugallja, ha meggyőzően és könnyen fejezze ki magát. A megfelelő felkészültség abban is rejlik, hogy a prezentáló tekintettel van a hallgatóságban található különböző személyiségekre, érdeklődésre. Fontos, hogy a kezdés hangsúlyos, érthető legyen, melyet a tények bemutatása követ. Fontos, hogy a figyelmet az előadás végéig megragadja a prezentáló, hiszen az összefoglalás és az újdonságok bemutatása meg kell hogy ragadjon a hallgatóban.

Összegezve néhány általános ajánlást kívánunk megfogalmazni, melyek segítségre lehetnek a prezentációk elkészítésében és bemutatásában.

<u>Diák elkészítése:</u>

1. Javasolt oldalszámokat feltüntetni a diákon (legelső kivételével), ráadásul úgy, hogy a jelenlegi diaszám mellett az összes diaszám is feltüntetésre kerül. Ez segítséget nyújt a hallgatóságnak is abban, hogy mennyi van még vissza az előadásból, illetve a felmerülő kérdések során a kérdező, könnyebben tud hivatkozni a diára.
2. Lehetőség szerint kerülni szükséges a diákon a rövidítéseket, mozaikszavakat.
3. A diák az előadás vázlatát tartalmazzák, ezért kerülendő a körmondatok által történő fogalmazás. A diák tartalmának felolvasása nagyon nem javasolt.

4. A túlzásba vitt sok animáció kerülendő, mivel zavaró lehet a közönség számára, illetve az időkorlát rovására is mehet.
5. Egyértelműen derüljön ki az előadásból, hogy mi volt a megoldandó feladat. Célszerű „Feladatkitűzés”, vagy „Problémafelvetés” címmel legalább egy diát szentelni ennek a résznek.
6. Egyértelműen derüljön ki az előadásból, hogy mik a saját eredmények, és ajánlott azok összevetése más kutatók eredményeivel is.
7. Fontos legalább egy diát a konklúzió, összegzés résznek szentelni.
8. Mivel az utolsó kivetített dia marad fenn legtovább a kivetítőn, a hallgatóság ezt nézi legtovább, így ezt kulcsdiának nevezhetjük. Javasolt egy informatívabb összefoglalást és a további munkákra vonatkozó elképzeléseket feltüntetni, semmiképpen nem a „Köszönöm a megtisztelő figyelmet” szerepeltetni..
9. A hallgatóság a célja, hogy a problémafelvetést, a megoldást és az eredményeket megismerje, ezért a legapróbb részletek bemutatásától eltekinthetünk, mivel fontos, hogy az idővel megfelelően tudjunk gazdálkodni.
10. Az elkészített diákat javasolt megmutatni a témában jártas szakembernek és kikérni a véleményét bemutatás előtt. A szakdolgozat vagy diplomamunka védés előtt feltétlenül a témavezetővel, mentorral közösen érdemes átnézni a prezentációt.
11. Előfordulhat egy-két olyan kérdés, amikre előre számítani lehet, az ezekre adandó válaszokat, illetve az ezzel kapcsolatos diákat érdemes beletenni az előadásba az utolsó előadás dia után, így ha felmerül az adott kérdés (pl.: szakdolgozatvédés alkalmával), akkor rögtön fel lehet vetíteni az előre elkészített diát.

Előadás, prezentáció:

1. Az előadás elején köszöntsük a közönséget, az előadás végén - lehetőleg szóban- célszerű a "Köszönöm a megtisztelő figyelmet" mondattal jelezni, hogy az előadásunk véget ért.
2. Az előadást egyszer elgyakorolva lemérhető, hogy belefér-e az időkeretbe.. További gyakorlással 1-3 perc faragható az időből. Az idő tartása azért fontos, mivel sokszor „büntetik” az idő túllépést, vagy egyszerűen nem hagyják tovább mondani az előadást.
3. Az előadás során ne fordítsunk lehetőleg senkinek sem hátat, célszerű a közönség, bizottság, zsűri felé fordulni. Kerülendő a zsebre tett kézzel tartott előadás.
4. Az öltözék legyen az alkalomhoz illő.

5. Az előadásra jóval előbb illik megérkezni, hogy legyen idő az előadási anyag feltöltésére.
6. Hatástalanok maradnak a látványos és jól megírt diák, ha az előadó túlságosan fáradt, ezért célszerű kipihenten menni az előadásra.

4. FÜGGELÉK (TÁBLÁZATOK)

4.1. Standard normális eloszlás

Sűrűségfüggvény értékei

z	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,496	0,492	0,488	0,484	0,480	0,476	0,472	0,468	0,464
0,1	0,460	0,456	0,452	0,448	0,444	0,440	0,436	0,433	0,429	0,425
0,2	0,421	0,417	0,413	0,409	0,405	0,401	0,397	0,394	0,390	0,386
0,3	0,382	0,378	0,374	0,371	0,367	0,363	0,359	0,356	0,352	0,348
0,4	0,345	0,341	0,337	0,334	0,330	0,326	0,323	0,319	0,316	0,312
0,5	0,309	0,305	0,302	0,298	0,295	0,291	0,288	0,284	0,281	0,278
0,6	0,274	0,271	0,268	0,264	0,261	0,258	0,255	0,251	0,248	0,245
0,7	0,242	0,239	0,236	0,233	0,230	0,227	0,224	0,221	0,218	0,215
0,8	0,212	0,209	0,206	0,203	0,200	0,198	0,195	0,192	0,189	0,187
0,9	0,184	0,181	0,179	0,176	0,174	0,171	0,169	0,166	0,164	0,161
1,0	0,159	0,156	0,154	0,152	0,149	0,147	0,145	0,142	0,140	0,138
1,1	0,136	0,133	0,131	0,129	0,127	0,125	0,123	0,121	0,119	0,117
1,2	0,115	0,113	0,111	0,109	0,107	0,106	0,104	0,102	0,100	0,099
1,3	0,097	0,095	0,093	0,092	0,090	0,089	0,087	0,085	0,084	0,082
1,4	0,081	0,079	0,078	0,076	0,075	0,074	0,072	0,071	0,069	0,068
1,5	0,067	0,066	0,064	0,063	0,062	0,061	0,059	0,058	0,057	0,056
1,6	0,055	0,054	0,053	0,052	0,051	0,049	0,048	0,047	0,046	0,046
1,7	0,045	0,044	0,043	0,042	0,041	0,040	0,039	0,038	0,038	0,037
1,8	0,036	0,035	0,034	0,034	0,033	0,032	0,031	0,031	0,030	0,029
1,9	0,029	0,028	0,027	0,027	0,026	0,026	0,025	0,024	0,024	0,023
2,0	0,023	0,022	0,022	0,021	0,021	0,020	0,020	0,019	0,019	0,018
2,1	0,018	0,017	0,017	0,017	0,016	0,016	0,015	0,015	0,015	0,014
2,2	0,014	0,014	0,013	0,013	0,013	0,012	0,012	0,012	0,011	0,011
2,3	0,011	0,010	0,010	0,010	0,010	0,009	0,009	0,009	0,009	0,008
2,4	0,008	0,008	0,008	0,008	0,007	0,007	0,007	0,007	0,007	0,006
2,5	0,006	0,006	0,006	0,006	0,006	0,005	0,005	0,005	0,005	0,005
2,6	0,005	0,005	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
2,7	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003
2,8	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
2,9	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,001	0,001	0,001
3,0	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001

Kritikus érékek különböző szignifikancia- szintek esetén

	Szignifikancia-szint (α)					
Egyoldalú	0,1000	0,0500	0,0250	0,0225	0,0100	0,0050
Kétoldalú	0,2000	0,1000	0,0500	0,0450	0,0200	0,0100
z	1,280	1,645	1,960	2,000	2,330	2,587

4.2. Student féle t-eloszlás

Student féle t-eloszlás kritikus értékei különféle szignifikancia-szint mellett

Szabadság-fok	**Szignifikancia-szint**				
	0,1	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,656
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756

30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
70	1,294	1,667	1,994	2,381	2,648
80	1,292	1,664	1,990	2,374	2,639
90	1,291	1,662	1,987	2,368	2,632
100	1,290	1,660	1,984	2,364	2,626
150	1,287	1,655	1,976	2,351	2,609
200	1,286	1,653	1,972	2,345	2,601

4.3. $\varkappa^2$-eloszlás

${}_{n}\chi^2$-eloszlás kritikus értékei különféle szignifikancia-szintek mellett

Szabadságfok	Szinifikancia-szint					
	0,9900	0,9500	0,9000	0,1000	0,0500	0,0100
1	0,000	0,004	0,016	2,706	3,841	6,635
2	0,020	0,103	0,211	4,605	5,991	9,210
3	0,115	0,352	0,584	6,251	7,815	11,345
4	0,297	0,711	1,064	7,779	9,488	13,277
5	0,554	1,145	1,610	9,236	11,070	15,086
6	0,872	1,635	2,204	10,645	12,592	16,812
7	1,239	2,167	2,833	12,017	14,067	18,475
8	1,647	2,733	3,490	13,362	15,507	20,090
9	2,088	3,325	4,168	14,684	16,919	21,666
10	2,558	3,940	4,865	15,987	18,307	23,209
11	3,053	4,575	5,578	17,275	19,675	24,725
12	3,571	5,226	6,304	18,549	21,026	26,217
13	4,107	5,892	7,041	19,812	22,362	27,688
14	4,660	6,571	7,790	21,064	23,685	29,141
15	5,229	7,261	8,547	22,307	24,996	30,578
16	5,812	7,962	9,312	23,542	26,296	32,000
17	6,408	8,672	10,085	24,769	27,587	33,409
18	7,015	9,390	10,865	25,989	28,869	34,805
19	7,633	10,117	11,651	27,204	30,144	36,191
20	8,260	10,851	12,443	28,412	31,410	37,566
21	8,897	11,591	13,240	29,615	32,671	38,932
22	9,542	12,338	14,041	30,813	33,924	40,289
23	10,196	13,091	14,848	32,007	35,172	41,638
24	10,856	13,848	15,659	33,196	36,415	42,980
25	11,524	14,611	16,473	34,382	37,652	44,314
26	12,198	15,379	17,292	35,563	38,885	45,642
27	12,878	16,151	18,114	36,741	40,113	46,963
28	13,565	16,928	18,939	37,916	41,337	48,278
29	14,256	17,708	19,768	39,087	42,557	49,588
30	14,953	18,493	20,599	40,256	43,773	50,892
31	15,655	19,281	21,434	41,422	44,985	52,191
32	16,362	20,072	22,271	42,585	46,194	53,486
33	17,073	20,867	23,110	43,745	47,400	54,775
34	17,789	21,664	23,952	44,903	48,602	56,061
35	18,509	22,465	24,797	46,059	49,802	57,342

36	19,233	23,269	25,643	47,212	50,998	58,619
37	19,960	24,075	26,492	48,363	52,192	59,893
38	20,691	24,884	27,343	49,513	53,384	61,162
39	21,426	25,695	28,196	50,660	54,572	62,428
40	22,164	26,509	29,051	51,805	55,758	63,691
50	29,707	34,764	37,689	63,167	67,505	76,154
60	37,485	43,188	46,459	74,397	79,082	88,379
70	45,442	51,739	55,329	85,527	90,531	100,425
80	53,540	60,391	64,278	96,578	101,879	112,329
90	61,754	69,126	73,291	107,565	113,145	124,116
100	70,065	77,929	82,358	118,498	124,342	135,807
150	112,668	122,692	128,275	172,581	179,581	193,207
200	156,432	168,279	174,835	226,021	233,994	249,445
250	200,939	214,392	221,806	279,050	287,882	304,939

4.4. F-eloszlás

F-eloszlás kritikus értékei 5%-os egyoldalú (10%-os kétoldalú) szignifikancia-szint mellett

Nevező szf.	Számláló szabadságfoka															
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	50	100
2	18,5	19,0	19,1	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,63	8,62	8,58	8,55
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,77	5,75	5,70	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,52	4,50	4,44	4,41
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,83	3,81	3,75	3,71
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,40	3,38	3,32	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,11	3,08	3,02	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,89	2,86	2,80	2,76
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,73	2,70	2,64	2,59
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,60	2,57	2,51	2,46
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,50	2,47	2,40	2,35
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,41	2,38	2,31	2,26
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,34	2,31	2,24	2,19

15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,28	2,25	2,18	2,12
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,23	2,19	2,12	2,07
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,18	2,15	2,08	2,02
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,14	2,11	2,04	1,98
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,11	2,07	2,00	1,94
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,07	2,04	1,97	1,91
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,18	2,10	2,05	2,01	1,94	1,88
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07	2,02	1,98	1,91	1,85
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,13	2,05	2,00	1,96	1,88	1,82
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03	1,97	1,94	1,86	1,80
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,09	2,01	1,96	1,92	1,84	1,78
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99	1,94	1,90	1,82	1,76
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,06	1,97	1,92	1,88	1,81	1,74
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96	1,91	1,87	1,79	1,73
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,03	1,94	1,89	1,85	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,88	1,84	1,76	1,70

35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	1,96	1,88	1,82	1,79	1,70	1,63
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,78	1,74	1,66	1,59
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,73	1,69	1,60	1,52
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,69	1,65	1,56	1,48
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,80	1,71	1,65	1,61	1,52	1,44
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,68	1,62	1,57	1,48	1,39
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,72	1,62	1,56	1,52	1,41	1,32

F-eloszlás kritikus értékei 2,5%-os egyoldalú (5%-os kétoldalú) szignifikancia-szint mellett

Nevező	Számláló szabadságfoka															
szf.	1	2	3	4	5	6	7	8	9	10	15	20	25	30	50	100
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5	14,4	14,3	14,2	14,1	14,1	14,0	14,0
4	12,2	10,7	10,0	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,66	8,56	8,50	8,46	8,38	8,32
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,43	6,33	6,27	6,23	6,14	6,08
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,27	5,17	5,11	5,07	4,98	4,92
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,57	4,47	4,40	4,36	4,28	4,21
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,10	4,00	3,94	3,89	3,81	3,74
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,77	3,67	3,60	3,56	3,47	3,40
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,52	3,42	3,35	3,31	3,22	3,15
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,33	3,23	3,16	3,12	3,03	2,96
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,18	3,07	3,01	2,96	2,87	2,80
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,05	2,95	2,88	2,84	2,74	2,67
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	2,95	2,84	2,78	2,73	2,64	2,56
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,86	2,76	2,69	2,64	2,55	2,47
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,79	2,68	2,61	2,57	2,47	2,40
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,72	2,62	2,55	2,50	2,41	2,33
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,67	2,56	2,49	2,44	2,35	2,27
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,62	2,51	2,44	2,39	2,30	2,22
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,57	2,46	2,40	2,35	2,25	2,17
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,53	2,42	2,36	2,31	2,21	2,13
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,50	2,39	2,32	2,27	2,17	2,09
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,47	2,36	2,29	2,24	2,14	2,06
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,44	2,33	2,26	2,21	2,11	2,02
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,41	2,30	2,23	2,18	2,08	2,00
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,39	2,28	2,21	2,16	2,05	1,97
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,36	2,25	2,18	2,13	2,03	1,94
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,34	2,23	2,16	2,11	2,01	1,92
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,32	2,21	2,14	2,09	1,99	1,90
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,31	2,20	2,12	2,07	1,97	1,88
35	5,48	4,11	3,52	3,18	2,96	2,80	2,68	2,58	2,50	2,44	2,23	2,12	2,05	2,00	1,89	1,80
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,18	2,07	1,99	1,94	1,83	1,74
50	5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46	2,38	2,32	2,11	1,99	1,92	1,87	1,75	1,66
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,06	1,94	1,87	1,82	1,70	1,60
75	5,23	3,88	3,30	2,96	2,74	2,58	2,46	2,37	2,29	2,22	2,01	1,90	1,82	1,76	1,65	1,54
100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,24	2,18	1,97	1,85	1,77	1,71	1,59	1,48
200	5,10	3,76	3,18	2,85	2,63	2,47	2,35	2,26	2,18	2,11	1,90	1,78	1,70	1,64	1,51	1,39

5. IRODALOMJEGYZÉK

1. 169/2000. [IX. 29.] és 154/2004. [X. 14.] Kormány Rendelet
2. Ács P. (2007): A területi egyenlőtlenségek feltérképezése során leggyakrabban alkalmazott mérőszámok bemutatása, a sporttehetségek területi elhelyezkedésének példáján, Egy életpálya három dimenziója- Tanulmánykötet Pintér József emlékére, Pécsi Tudományegyetem Közgazdaságtudományi Kar, Pécs, 10- 22. o.
3. Ács P. (2015): Gyakorlati adatelemzés. Pécsi Tudományegyetem Egészségtudományi Kar. Pécs
4. Bíróné. N. E. (2004): Sportpedagógia. Dialóg Campus Kiadó. Budapest- Pécs
5. Dr. Hepp F.- Dr. Nádori L. (1971): Bevezetés a tudományos kutatásba. Kézirat. Tankönyvkiadó. Budapest.
6. Dr. Jánosa A. (2005): Adatelemzés számítógéppel, Perfekt Kiadó. Budapest, 271. o.
7. Falus I. (szerk.) (2000): Bevezetés a Pedagógiai kutatás módszereibe. Pedagógus Könyvek. Budapest. Műszaki Könyvkiadó. 540. o.
8. Gyetvai Gy.- Kecskemétiné Petri A. (1997): Testkultúra elméleti- és kutatásmódszertani alapismeretek. Főiskolai jegyzet. Juhász Gyula Tanárképző Főiskola. Szombathely. 208. o.
9. Hajdu O. (1997): A szegénység mérőszámai. KSH. Könyvtár és Dokumentációs Szolgálat. Budapest
10. Hajdu O. (1987): Sokváltozós statisztikai módszerek gyakorlati alkalmazása. Prodinform Műszaki Tanácsadó Vállalat. Budapest
11. Harsányi L (1998): Jó úton a sporttudomány akadémiai elismerése. Sporttudomány. 1998.2. sz.
12. Harsányi L. (2007): Az irodalomjegyzék készítés, idézés, hivatkozás további szabályai. Kézirat. Pécs. 2007. január 25.
13. Horváth L.- Prisztóka Gy. (2005): A sportpedagógia és sportpszichológia alapkérdései (főiskolai tankönyv) Bessenyei György Könyvkiadó, Nyíregyháza 2005.
14. Hunyadi L. (2001): Statisztikai következtetéselmélet közgazdászoknak. KSH, Budapest
15. Hunyadi L. (2002): Grafikus ábrázolás a statisztikában, Statisztikai Szemle 2002/1. 22-53. old.

16. Istvánfi Cs. (2000): Gondolatok a sporttudományokról. Kalokagathia. 2000. 1-2 sz. 7-18. o.
17. Kaj M.- Csányi T.- Karsai I.- Marton O. (2014): Kézikönyv a Nemzeti Egységes Tanulói Fittségi Teszt (NETFIT) alkalmazásához. Testnevelés Módszertani Könyvek (Csányi T. főszerk.). Magyar Diáksport Szövetség. Budapest
18. Kecskeméty L- Izsó L. (2005): Bevezetés az SPSS programrendszerbe, ELTE-Eötvös Kiadó, Budapest, 460.o
19. Kehl D.- Rappai G. (2006): Mintaelem-szám tervezése Likert-skálát alkalmazó lekérdezésekben. Statisztikai Szemle 84. évfolyam 9. szám. 848- 876. o.
20. Kerlinger F. (1980): Analysis Of Covariance Structure Tests Of A Criterial Referents Theory Of Attitudes, Multivariate Behavioral Research, Volume 15, Issue 4 January 1980 , 403 – 422. o.
21. Mundruczkó Gy. (1981): Alkalmazott regressziószámítás, Akadémiai Kiadó, Budapest
22. Müller A. (2004): Mozgásvizsgálatok a mozgásegyenletesség és a teljesítménykonstancia példáján. Disszertáció Semmelweis Egyetem Doktori Iskola Nevelés- és Sporttudományok Doktori Iskolája (Sport és Társadalomtudomány).
23. Pintér J. – Rappai G. (2001): A mintavételi tervek készítésének néhány gyakorlati megfontolása. Marketing & Menedzsment 2001/4. 4-11. o.
24. Ramanathan R. (2003): Bevezetés az Ökonometriába alkalmazásokkal, Panem Kft. Budapest
25. Rappai G. (2001): Üzleti statisztika Excellel, Központi Statisztikai Hivatal, Budapest
26. Sajtos L.- Mitev A. (2007): SPSS kutatási és adatelemzési kézikönyv, Alinea Kiadó. Budapest, 402. o.
27. Szabó K. (2002): Kommunikáció felsőfokon. Kossuth Kiadó. Budapest. 2.Kiadás. 404 o.
28. Székelyi M.- Barna I. (2005): Túlélőkészlet az SPSS-hez, Typotex Kiadó, Budapest, 455.o.
29. Vass M. (2005): Nevelés a sportban: kompetenciák c. habilitációs nyilvános előadás, Veszprémi Egyetem Interdiszciplináris Bölcsészet- és társadalomtudományok (nyelvtudomány; neveléstudomány) Doktor Iskola, Veszprém, 2005. október 18.
30. www. tanulokozosseg.mindentudo.hu/s_doc_server.php?id=1271

31. Zsolnai J. (1996): A pedagógia új rendszere címszavakban Nemzeti Tankönyvkiadó. Budapest, 227. p.

TÁMOP-4.1.2. E-15/1/KONV-2015-0003
„SPORTTUDOMÁNYI KÉPZÉS FEJLESZTÉSE A DUNÁNTÚLON 2015”

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE