BIOSTATISZTIKA

BIOSTATISZTIKA

SÁNDOR JÁNOS, ÁDÁNY RÓZA

PÉCSI TUDOMÁNYEGYETEM
UNIVERSITY OF PÉCS

Medicina Könyvkiadó Zrt • Budapest, 2011

A kiadvány a következő program keretében jelent meg:

TÁMOP-4.1.2-08/1/A-2009-0054

A kézirat lezárva: 2011. január 31.

Lektor:
Dr. Béres Judit, Dr. Bödecs Tamás

A kiadásért felel a Medicina Könyvkiadó Zrt. igazgatója

Felelős szerkesztő: Pobozsnyi Ágnes
Műszaki szerkesztő: Dóczi Imre
Ábrák száma: 22
Az ábrákat rajzolta: Olgyai Géza
Terjedelem: 11 (A/5) ív
Azonossági szám: 3598

Tartalomjegyzék

Előszó 7
Kísérletek és megfigyelések 9
A nem leíró jellegű vizsgálatok főbb lépései 11
A változók típusai 15
  Dichotóm, bináris adatok 15
  Nominális skálán mért adatok 16
  Ordinális skálán mért adatok 16
  Intervallumskálán mért adatok 17
  Arányskálán mért adatok 17
Leíró statisztika 19
  Hisztogram 21
  Centrális érték 22
  Szóródás 23
  Szabadsági fok 25
  Normális eloszlás 25
Megbízhatósági tartomány 29
Hipotézistesztelés 36
  Döntési küszöb 36
  Első- és másodfajú hiba 38
  Statisztikai tesztek 40
Csoportok közötti összehasonlítás 41
  Két csoport kvantitatív adatainak összehasonlítása 41
  Több csoport kvantitatív adatainak összehasonlítása 47
  Csoportok kvalitatív adatainak összehasonlítása 56
    Várható érték 56
    Csoportok közti különbség elemzése 62
    Párba rendezett kvalitatív adatok elemzése 67
Folytonos változók közötti kapcsolat elemzése 70
  Korreláció 70
  Lineáris regresszió 77

Determinációs koefficiens 86
Nem paraméteres próbák 88
Előjelteszt 88
Wilcoxon párosított teszt 90
Mann–Whitney U-teszt 92
Kruskal–Wallis H-teszt 94
Spearman-rangkorreláció 100
Többváltozós elemzések 103
Kétszempontos varianciaelemzés 104
Többváltozós lineáris regresszió 108
Oksági összefüggések 111
Az ok-okozati kapcsolat iránya 111
Koch-posztulátumok 112
Hill-szempontrendszer 113

Előszó

Azok a hallgatók, akik orvos- és egészségtudományi területen folytatják tanulmányaikat, meglehetős távolságtartással viszonyulnak a biostatisztikához. Ez a világ minden részén így van, és egyáltalán nem meglepő, ha a két tudományterület gondolkodásmódjának igen lényeges eltéréseibe belegondolunk. Nem is lehet reális cél, hogy megszerettesse valaki a biostatisztikát ebben a hallgatói körben. De…

Az orvos- és egészségtudományi területen hihetetlen mennyiségben keletkeznek kutatási eredmények, amelyek alapján a szakembereknek saját gyakorlatukat folyamatosan fejleszteni kell. Ehhez elengedhetetlen, hogy valamilyen szinten értsék a szakterületükön megjelenő új eredményeket előállító vizsgálatok belső logikáját, és pontosan tudják értelmezni azok végeredményeit. Ez elképzelhetetlen korrekt biostatisztikai ismeretek nélkül.

Az orvos- és egészségtudományi területen dolgozó szakembereknek meglehetősen nagy szabadsága van munkahelyükön a rájuk bízott területen a gyakorlat formálásában. Sok döntést kell önállóan meghozniuk. Előny, ha képesek a felmerülő problémákra tényeken alapuló megoldásokat javasolni. Ehhez sokszor a saját munkával kapcsolatban, saját munkahelyen keletkező adatok korrekt feldolgozása szükséges. Ez elképzelhetetlen korrekt biostatisztikai ismeretek nélkül.

Az orvos- és egészségtudományi területen dolgozó szakembereknek egyre gyakrabban kell bizonyítaniuk, hogy az a gyakorlat, amit követnek, hatékony. Ilyen feladatokat csak akkor tudnak színvonalasan ellátni, ha rendszeresen monitorozzák saját tevékenységüket, és a saját teljesítményt leíró adatokat korrekt módon prezentálják. Ez is elképzelhetetlen korrekt biostatisztikai ismeretek nélkül.

A tapasztalat az, hogy az orvos- és egészségtudományi területen dolgozó szakemberek gyakorlati problémák kezelésekor nem használják kellő gyakorisággal a statisztikai eszközöket, mert nem ismerik kellően őket. Ha pótolni akarják a hiányosságaikat, akkor pedig már a biostatisztikai könyvek bevezető fejezeténél feladják a küzdelmet a teljesen idegen terminológia miatt.

Ugyanakkor sok a statisztikai számítógépes program, amelyek használata bizonyos szinten egyre egyszerűbb, és egyre több az elektronikus adatbázisokba rendezett adat is az orvos- és egészségtudományi területen. Nagy a csábítás a két lehetőség összekapcsolására. Sok féligazság, félrevezető statisztika születhet, ha ezek a lehetőségek nem párosulnak a biostatisztikai alapelvek ismeretével.

A jegyzet arról szeretné meggyőzni a hallgatókat, hogy nem lehetetlen vállalkozás a biostatisztikai alapelvek és a legalapvetőbb eljárások menetének megértése, még akkor sem, ha valaki nem vonzódik a matematikához.

A jegyzet nem elriasztani akar, ezért nem elméleti igénnyel, nem a valószínűségszámítás elvont alapjairól indított magyarázatokkal, hanem ahol lehet, a biostatisztikai eljárások szemléletes értelmezésére törekedve próbál gyakorlati jelentőséggel bíró ismereteket átadni; és talán bátorítja az elmélyültebb, teoretikusan jobban megalapozott tanulmányokra a fogékonyabb hallgatókat.

A jegyzet a hallgatók előképzettségéhez igazodó, de a majdani szakemberek feladatait is szemmel tartó kompromisszum eredménye. De biztosan nem könnyű olvasmány! Sajnos igényli az elmélyült feldolgozást, egy-két gyors átolvasás helyett az alapos tanulmányozást.

Debrecen, 2011.

Kísérletek és megfigyelések

Ha olyan vizsgálatot végzünk, aminek része a statisztikai értékelés, általában egy specifikus kérdés (vagy másképpen hipotézis) megválaszolása a célunk. Ezek a kérdések (hipotézisek) felírhatók egy befolyásoló tényező és egy kiváltott hatás közti kapcsolatként: van-e hatása az adott befolyásoló tényezőnek az adott paraméterre. A befolyásoló tényező az egészségtudományok területén lényegében bármilyen, a szervezetre hatással levő faktor lehet. A kiváltott hatás pedig tulajdonképpen bármilyen biológiai, klinikai paraméter.

A kérdésekre adott válaszhoz szükséges adatokat alapvetően két módszerrel lehet összegyűjteni. Egyfelől lehetőség van arra, hogy a vizsgáló hozza létre az expozíciót (pl. szövettenyészetben dóziscsoportokat alakít ki, állatkísérletben kezelt és nem kezelt csoportokat hoz létre; klinikai vizsgálatban kezelt és placebocsoportba sorolja a betegeket), és a vizsgálati körülményeket is megpróbálja szabályozni. Ilyenkor kísérletről, experimentális vizsgálatról beszélünk. Másfelől vannak vizsgálatok, ahol nem befolyásoljuk a résztvevők viselkedését, nem mi hozzuk létre a vizsgálati körülményeket. Az adatgyűjtés ilyen esetben a tőlünk függetlenül zajló folyamatok megfigyelésén alapul.

Orvostudományi, egészségtudományi tudásunk egyik legfontosabb forrása a nem humán rendszerekben elvégzett kísérlet. In vitro kísérletekben, szövettenyészetekben vagy állatkísérletekben pontosan kezelhető és szabályozható a külső expozíció nagysága. Sőt, a célszervi dózis is mérhető, hiszen lényegében nincsenek korlátjai az invazivitásnak. A kialakuló biológiai elváltozás, betegség igen részletesen, invazív módszerek felhasználásával vizsgálható. A kísérlet során a folyamatban résztvevő biológiai rendszerekre ható egyéb tényezők hatása kontrollálható. A genetikai háttér zavaró hatásait megfelelően kiválasztott állattörzsekkel jelentősen csökkenteni lehet. A vizsgálatokban elvileg nagyszámú állatot is fel lehet használni, de erre gyakran nincs szükség, hiszen a kísérletekben a biológiai folyamatok természetes variabilitása szűk tartományon belül tartható. Ilyenkor pedig kis hatások is jól kimutathatók viszonylag kevés állat felhasználásával. De ha szükséges, az állatszám növelésével a precizitás javítható.

Az ilyen kísérletekben magas dózisokat használnak, hogy a biológiai válasz nagy százalékban fellépjen, illetve a válasz könnyen mérhető nagyságú legyen. Ezért, ha a kísérleti eredményeket humán viszonyok megértésére akarjuk felhasználni, számolnunk kell a következő problémákkal: (1) a humán expozíciók általában nagyságrendekkel

alacsonyabbak, mint a kísérletekben alkalmazott dózisok, (2) a kísérleti rendszer és az ember biológiai alapstruktúrái jelentősen különbözhetnek. Ezeknek megfelelően a kísérletes adatok humán felhasználásakor két extrapolációval élünk: (1) a magas kísérleti dózisok hatásai alapján becsüljük az alacsony dózisok hatásait, és (2) nem humán rendszerben mért adatokból következtetünk a humán hatásokra.

Amikor megfigyelésen alapuló humán vizsgálatokkal próbálunk a betegségek patológiai alapjainak természetéről adatokat szerezni, alapvetően más vizsgálati helyzetben találjuk magunkat. Itt a vizsgált személyek külső expozíciója csak ritkán adható meg pontosan, a célszervi dózisok meghatározásának pedig korlátot szab az invazív vizsgálatok korlátozott használata. Az expozíciók hatásainak detektálása az invazív vizsgálatok korlátozott használhatósága, a hosszú látenciaidő és a vizsgált személyekkel kapcsolatos nyilvántartás adminisztratív problémái miatt sokszor nem pontos. A zavaró tényezők kontrollálása sokszor csak részben, közelítő módszerekkel oldható meg. A genetikai háttér sokfélesége jelentős hibával terhelheti a vizsgálatot (bár vannak olyan

1. táblázat. Kísérletes és epidemiológiai vizsgálatok előnyös és hátrányos tulajdonságokból adódó kiegészítő természete

	Kísérlet		Megfigyelés	
Külső dózis	+	Beállítható, szabályozható	–	Nehéz pontosan becsülni
Célszervi dózis	+	Közvetlenül mérhető	–	Általában nem mérhető
Korai biológiai válasz	+	Invazívan vizsgálható	–	Indirekt markerekből becsülhető
Kifejlődött hatás	+	Részletesen vizsgálható	–	Diagnosztikus tévedések, nyilvántartási hibák miatt korlátozott pontosságú
Zavaró tényezők	+	Effektíven kontrollálható a vizsgálati elrendezésben, közvetlenül, invazívan mérhető	–	A vizsgálati elrendezésben, adatfeldolgozásban kezelhető, de gyakran csak indirekten mérhető
Érzékenységbeli variabilitás	+	Kezelhető (a kezelési csoportokban azonos érzékenység biztosítható)	–	Nehezen kezelhető
Vizsgált minta nagysága	+	Általában kis esetszámok	–	Általában nagy minták
Humán relevancia	–	Indirekt	+	Direkt

vizsgálati elrendezések, ahol ezzel gyakorlatilag nem kell számolnunk, például ikervizsgálatok esetén). A humán megfigyelésen alapuló vizsgálatokhoz gyakran nagy mintát kell összeállítani. Ez kényszer, hiszen csak így ellensúlyozhatók azok a közelítő jellegű becslésekből adódó hibák, amiket az előbb említettünk. Ha a vizsgálatot sikerül úgy kivitelezni, hogy a felsorolt hibaforrások által okozott torzítások bizonyos korlátok között maradjanak, akkor az eredmény extrapolációk nélkül alkalmazható, azaz (szemben a nem humán rendszerből származó adattal) közvetlenül humán releváns.

A fenti értelemben tehát a kísérletes és megfigyelésen alapuló vizsgálatok kiegészítik egymást *(1. táblázat)*. Ennek jó példája, hogy az egyes kémiai anyagokat csak akkor sorolják a bizonyítottan a humán karcinogének közé, ha ehhez mind a kísérletes, mind pedig az epidemiológiai bizonyítékok rendelkezésre állnak.

A nem leíró jellegű vizsgálatok főbb lépései

A vizsgálatok mindig gyakorlati jelentőséggel bíró probléma megoldására irányulnak. Annak a problémakörnek a jelentőségét, aminek az egyik részére fókuszál a vizsgálati kérdés, alapvetően a probléma előfordulásának gyakoriságával, az egészségkárosodás érintettek közti súlyosságával, illetve az intervenciós (preventív vagy gyógyító) potenciállal (van-e reális megelőzési lehetőség, milyenek a gyógyítás lehetőségei) tudjuk leírni.

A munkája során mindenki szembesül megoldatlan kérdésekkel. Ezek lehetnek alapvető biológiai folyamatokra vonatkozó, nagyon nehéz alapkutatási feladatok (Milyen módon lehet a daganatok kemoterápiáját célszövet-specifikusabbá tenni?), de lehetnek egészen egyszerű, a napi munkával kapcsolatos, gyakorlati jellegű kérdések. (A régebben vagy az újabban használt fertőtlenítő szer mellett alakul ki kevesebb sebfertőzés?) A vizsgálatok alapvetően mindegyik esetben ugyanazokat az elemi lépéseket tartalmazzák.

A nem leíró jellegű vizsgálatok alapkérdése mindig egy ok-okozati összefüggésre vonatkozik. A kutatási kérdés akkor kellően specifikus, ha abban mind a feltételezett ok (befolyásoló tényező, expozíció), mind pedig a vizsgálat középpontjában álló kiváltott hatás pontosan azonosítható. A kiváltott hatás széleskörűen értelmezendő: gyermekek elhízása, HiB okozta halálozás, dohányzás gyakorisága, influenzajárvány idején az időskorúak hospitalizációja, foglalkozási tüdőrák, hipertóniagondozás hatékonysága, emlőrák okozta halálozás, stroke-betegek halálozási esélye. A vizsgált befolyásoló tényezők is hasonlóan nagyon különböző természetű tényezők lehetnek: iskolai egészségfejlesztési program, tömeges védőoltás, alacsony társadalmi gazdasági státusz, in-

1. ábra. Vizsgálati modellek egy zavaró tényezővel

fluenzafertőzés, nikkel tartalmú füst belélegzése, orvosi team képzésének színvonala, szűrőprogram szervezési módja, intenzív osztály felszereltsége.

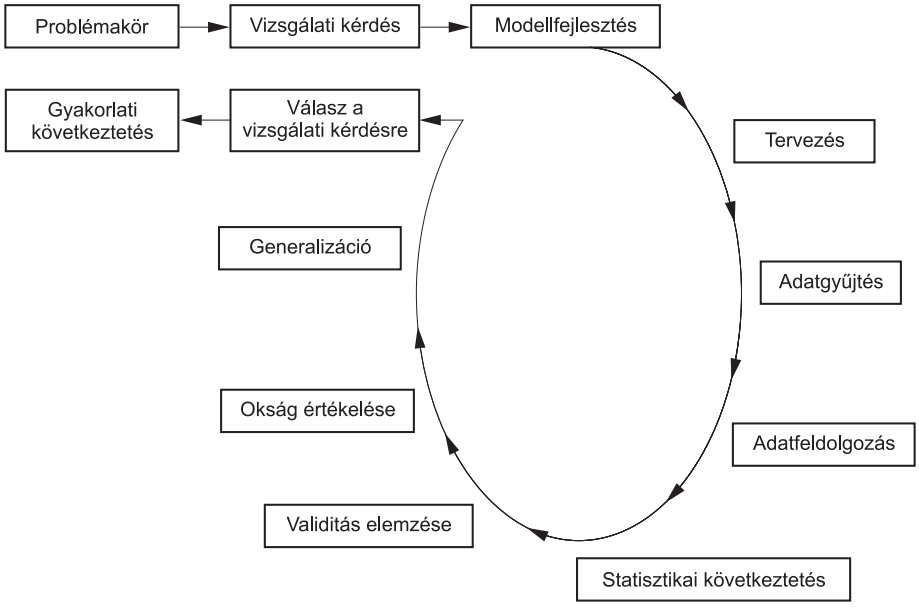
Ha csak a befolyásoló tényezőre és a kiváltott hatásra gyűjtünk adatot, akkor a biológiai folyamatok bonyolultsága miatt nem fogunk jól használható válaszokat kapni. Például ha az alapkérdésünk az, hogy a jelentős alkoholfogyasztás növeli-e a tüdőrák kialakulásának kockázatát, és csak az alkoholfogyasztásra és a tüdőrák jelenlétére gyűjtünk adatot, akkor az elemzésünk megmutatja, hogy az alkoholisták közt gyakrabban alakul ki a tüdőrák, mint a nem alkoholisták közt. Ezt viszont nem tudjuk az alkohol karcinogén hatásának bizonyítékaként elfogadni, hiszen nyilvánvaló, hogy az alkoholisták többet is dohányoznak, mint a nem alkoholisták, és köztük a tüdőráktöbbletet a több elszívott cigaretta is okozhatja. Ebben az esetben a dohányzás (ami a vizsgált befolyásoló tényezővel, az alkoholfogyasztással összekapcsolódik, és a vizsgált betegségnek önmagában is rizikófaktora) zavaró tényező. A zavaró tényező is sokféle faktor lehet az alapkérdéstől függően: a család étkezési szokásai, fertőzőbeteg-ellátás fejlettsége, dohánytermékek reklámozásának szabályozása, influenza elleni védőoltás elérhetősége, munkások dohányzási szokásai, hipertóniás betegek képzettsége, korábbi terhességek száma, lakóhely és intenzív ellátó székhelye közti távolság.

Lényegében az így kapott összefüggésrendszer tartalmazza az általunk feltett kérdés környezetéről eddig rendelkezésünkre álló tudást. Ennek a modellnek a felvázolásával összegezzük az eddigi ismereteket. A modell pontos felállítása azért meghatározó jelentőségű, mert a vizsgálat során a modell elemeire kell majd adatot gyűjteni.

A zavaró tényezők hatásától valamilyen módon meg kell tisztítani a vizsgálatunkat. (Kontrollálni kell a zavaró tényezőket.) Megfigyelésen alapuló vizsgálatok esetén lehetőségünk van arra, hogy a zavaró tényezőkre vonatkozóan adatot gyűjtsünk, és hatásukat a statisztikai elemzés során semlegesítsük. A kísérletek során pedig olyan vizsgálati körülményeket hozunk létre, hogy a kezelt és a kontrollcsoport ténylegesen csak a vizsgált befolyásoló tényező szempontjából különbözzön.

Az adatgyűjtés során arra kell törekedni, hogy a valóságot minél pontosabban tükröző adatbázist kapjunk, aminek az a feladata, hogy a vizsgálatunk számára számszerűen írja le a valós viszonyokat.

Ezt követően kerül sor a vizsgálati kérdésünknek megfelelő hipotézis tesztelésére. Ez tulajdonképpen abból áll, hogy kiszámítjuk annak a valószínűségét, hogy a valóság tényei a mi általunk feltételezett összefüggésnek megfelelően jönnek létre. Statisztikai eszközök segítségével leírjuk a hipotézis és a valóság közti összhang mértékét, és véleményt alkotunk a hipotézis helytállóságáról.



2. ábra. A vizsgálatok lépései

A vizsgálatok során gyűjtött adatok sohasem tükrözik pontosan a valóságot. Ennek megfelelően, a statisztikai értékelést követően meg kell vizsgálnunk, hogy az adatgyűjtés mennyire volt jó, az adatbázis és a belőle számított statisztikai eredmény mennyire megbízható, azaz valid. A validitást a minta megfelelősége, a zavaró tényezők megfelelő kontrollja és a mérések megbízhatósága mentén értékeljük.

Ha az elért eredmények a vizsgált kapcsolat meglétét vagy hiányát meggyőzően bizonyítják, akkor a kapcsolat ok-okozati jellegét külön kell értékelni. (Ha két jelenség kapcsoltan fordul elő, az még nem bizonyítja, hogy valamilyen mechanizmus révén az egyik befolyással van a másik alakulására. Lehet, hogy mindkét paraméterre hatással van egy harmadik tényező, ami látszólagos kapcsolatot eredményez köztük.)

Ha a statisztikai értékelés után a validitás értékelése kellően megbízhatónak minősíti a vizsgálatot, akkor kerül sor annak értékelésére, hogy a saját eredmény mennyire terjeszthető ki, milyen mértékben generalizálható.

Mindezek után tudunk válaszolni a vizsgálat alapkérdésére: van-e ok-okozati kapcsolat a befolyásoló tényező és a kiváltott hatás közt. Az ilyen módon megalapozott, kellően megbízható válasz birtokában tudunk konkrét következtetéseket levonni a vizsgálat alapját jelentő gyakorlati problémával kapcsolatban. Jó esetben gyakorlati beavatkozásokat tudunk megalapozni a körültekintően kivitelezett vizsgálat révén.

A változók típusai

Vizsgálatok során vagy kísérlet során generált, vagy megfigyelések során rögzített adatokat gyűjtünk össze. Az adatokat adatbázisba rendezzük, amelynek a feladata a valóság számszerű leképezése.

Az egyes jelenségeket különböző mérési technikákkal, különböző skálákkal lehet mérni. Ennek megfelelően különböző természetű adatokat fogunk kapni. Egyes mérések nagyon részletesen mutatják be a vizsgált jelenséget, míg mások durvább képet adnak csak. A mérési technika megválasztása a vizsgálatok tervezési fázisának feladata. Fontos döntés ez, hiszen az adatgyűjtés ez alapján zajlik majd, és ez fogja meghatározni a statisztikai feldolgozás során alkalmazható módszereket.

Az adattípusok közt hierarchia értelmezhető, amennyiben a legegyszerűbb típusok által hordozott információ mindig előállítható a hierarchiában felette állókból *(3. ábra)*. Fordított irányban ez lehetetlen. A tervezéskor ezért célszerű óvatosnak lenni, és kétség esetén inkább kicsit informatívabb adattípus mellett kell dönteni, ha erre lehetőségünk van. Redukálni lehet majd az adattípust, de ha az adatgyűjtés során egyszerűbb formát használunk, és a feldolgozás során derül ki, hogy részletesebb adatra lenne szükségünk ahhoz, hogy meg tudjuk válaszolni az alapkérdésünket, akkor már csak a vizsgálat újrakezdése áll nyitva előttünk. Ezt a nyilvánvaló hibát (pazarlást) el kell kerülni!

Dichotóm, bináris adatok

A legegyszerűbb adatgyűjtés az, amikor a vizsgálat résztvevője esetében egy tulajdonság meglétét vagy hiányát kell megállapítanunk. A kérdésünk mindösszesen annyi, hogy valaki lázas-e vagy nem, dohányzik vagy nem, hipertóniás-e vagy nem, elhízott-e vagy nem. Ugyanezek a kategóriák megfogalmazhatók egymást kölcsönösen kizáró pozitív megfogalmazások révén is. Azaz beszélhetünk lázas és normál hőmérsékletű résztvevőkről, dohányzókról és nem dohányzókról, hipertóniásokról és normotenzívekről. És vannak dichotóm kategóriák, amelyeknél eleve csak pozitív megfogalmazásokkal adjuk meg a kategóriák nevét: a nem lehet férfi vagy nő, a lakóhely jellege lehet vidéki vagy városi.

Ez az adattípus kvalitatív jellegű. Csak gyakorisági mutatók segítségével összegezhetőek. A résztvevők nemi aránya, a dohányzás prevalenciája, a hipertónia kialakulásának kumulatív incidenciája lehet a mintákat bemutató leíró statisztika.

Nominális skálán mért adatok

Ha a mérési skálánkon kettőnél több, egymást kizáró kategóriánk van, akkor az adatunk már több információt hordoz, a dichotóm változókhoz képest részletgazdagabban mutatja be a vizsgált jelleget. Ha ezek a kategóriák nem rendezhetők valamilyen elv alapján sorrendbe, akkor nominális adatról beszélünk. Például családi állapot (hajadon, házas, elvált, özvegy), foglalkozás, vallási irányultság mérésére alakíthatunk ki kategóriákat. A kategóriák egyikébe, és csak az egyikébe be kell tudnunk sorolni minden vizsgálati alanyt. Ez a kvalitatív jellegű adattípus is csak gyakorisági mutatók segítségével összegezhető. A résztvevők családi állapotának megoszlása, az egyes munkahelyeken dolgozók részaránya, az egyes vallási csoportokhoz tartozók aránya a mintában lehet a mintákat bemutató leíró statisztika.

Természetesen a kategóriákat valamilyen meggondolás alapján összevonhatjuk, az adatot dichotóm formába hozhatjuk, ezáltal redukálhatjuk az információtartalmat (pl. családi állapot esetében definiálhatunk egyedül élő és társas kapcsolatban élő csoportot).

Ordinális skálán mért adatok

Amennyiben kettőnél több, egymást kölcsönösen kizáró kategóriát definiálunk a skálánkon, olyan módon, hogy a kategóriák sorba rendezhetők, akkor tovább bővítettük az adataink információtartalmát. Ilyen skála mérheti a képzettségi kategóriákat, leírhatja az újszülött születés utáni állapotát (Apgar score) vagy a koponyasérülés utáni klinikai státuszt (Glasgow Coma Scale, GCS).

Ordinális skálán mért adatok esetben a sorrendiség ellenére is csak kvalitatív adatról beszélhetünk. A kategóriák egymáshoz viszonyított távolsága ugyanis nem egyforma. Ezért nem is szerencsés, ha az adatokat számként kódolt formában rögzítjük az adatbázisban. Például az a látszat keletkezhet, hogy az 1-gyel kódolt képzettségű vizsgálati résztvevő fele olyan képzett, mint a 2-vel kódolt, és harmadannyira, mint a 3-mal kódolt, ami természetesen nem igaz. (A tapasztalat szerint figyelmetlen feldolgozás során kvantitatív adatként kerülhet felhasználásra az adat!)

Ennél az adattípusnál definiálhatunk olyan dichotomizálási küszöböt, ami felett és alatt egyesítve a kategóriákat, dichotóm adattá redukálhatjuk az eredeti eredményeinket (pl. megkülönböztethetünk felsőfokú végzettséggel rendelkezőket és azzal nem rendelkezőket az adatbázisban lévő, a képzettségi szintet rendezett kategóriákkal mérő adatok alapján.)

A sorba rendezés lehetősége miatt ennél az adattípusnál, a gyakoriságok megoszlásán túlmenően, már összegezhetők a vizsgálati eredmények medián és kvantilisek segítségével is.

Intervallumskálán mért adatok

Ha olyan skálát alkalmazunk, amelyik sok sorba rendezhető kategóriát tartalmaz, és a sorba rendezett kategóriák közti különbség állandó, akkor már kvantitatív jellegű az adatunk. A kategóriák miatt diszkrét adatról beszélhetünk (pl. a testhőmérséklet értékét 1 Celsius fokos pontossággal mérjük, és az 1 Celsius fokos széles kategóriák egyikébe soroljuk be a vizsgálatban résztvevőket). Ha finomítjuk a mérésünket, javul a mérési pontosság, szűkülnek a kategóriák. Bizonyos mérési pontosság elérése után már nincs értelme kategóriákról beszélni, hiszen gyakorlatilag bármilyen érték lehet a mérés eredménye. Ez a változó már folytonos természetű. (A diszkrét, csak bizonyos értékeket felvevő és a folytonos változó közti megkülönböztetésnek nincs elméleti alapja. Gyakorlati szempontokat figyelembe véve, a konkrét vizsgálati helyzethez kell igazítani a döntésünket, hogy adott változót minek tekintünk. A besorolás természetesen nem öncélú, hanem azt határozza meg, hogy milyen statisztikai eszközzel lehet majd az adatot értékelni. Ha állást kell foglalni a két változótípus közti határról, akkor valószínűleg úgy nem tévedünk, ha a több mint 20 kategóriába sorolt változókat kezeljük folytonosként.) Az intervallumskála az egyes mérési eredmények közti távolságot értelmezhetővé teszi. Ezáltal lehetőségünk van a vizsgálati eredmények összegzésére átlagérték és szórás segítségével is. Az intervallumskálán azonban nincs kitüntetett kezdőpont, ezért nem értelmezhető az egyes eredmények hányadosa. A testhőmérséklet-csökkenés egy kezelés hatására két mérési eredmény közti különbségként értelmezhető. De nincs semmi értelme arról beszélni, hogy hány százalékkal csökkent a testhőmérséklet a kezelés során. Az intervallumskálán belül természetesen definiálhatunk ordinális kategóriákat, ezáltal redukálhatjuk kvalitatív jellegűvé az adatunkat. Sőt, ha egy dichotomizálási küszöböt határozunk meg, akkor a legegyszerűbb dichotóm adatot is előállíthatjuk.

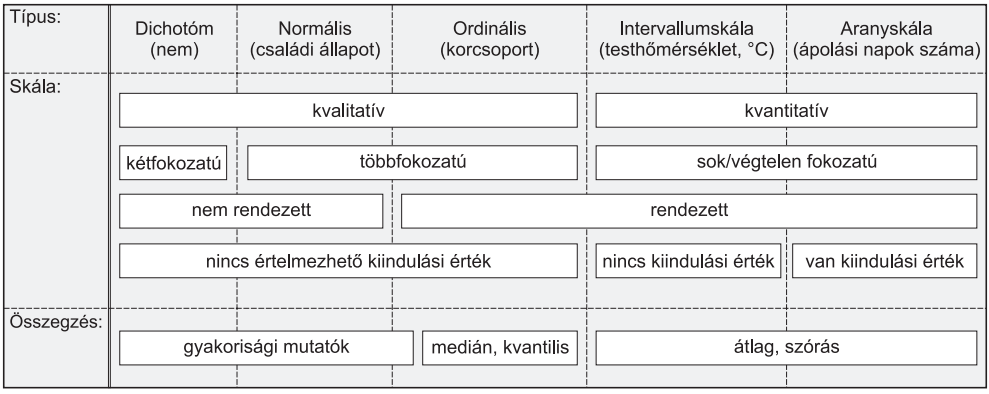
Arányskálán mért adatok

Ha olyan intervallumskálán mérünk, aminek van kezdő pontja, azaz, ahol a 0 érték értelmezhető, akkor arányskálánk van. Ez is lehet diszkrét vagy folytonos a mérési pontos-

ságtól függően. Az adatokat itt is átlag és szórás segítségével tudjuk összegezni. Ebben az esetben viszont már nem csak a mérési eredmények különbségeit tudjuk értelmezni, hanem azok arányát is. A túlélési idők elemzésekor értelmezhető az, hogy mennyi idővel lett hosszabb a túlélés egy új eljárás bevezetése után, de az is, hogy hányszorosára nőtt a túlélési idő.

Az adattípus meghatározza az alkalmazható statisztikai eljárást. Gyakran van arra szükség, hogy adatainkat az eljárás igényeihez igazítsuk. Az adattípusok hierarchiája határozza meg ilyenkor a lehetőségeinket. Csak az egyszerűsítés irányába tudunk átalakítani. Ez pedig kategóriákba sorolással, illetve kategóriahatárok definiálásával egyenértékű.

A kategóriahatárok, illetve a dichotomizálási küszöbök kijelölése történhet valamilyen biológiailag, klinikailag definiált normálérték segítségével (a ténylegesen mért vérnyomásértékek helyett használhatjuk a hipertóniás, normotenzív, vagy ennél részletesebb beosztásokat, amikhez a határértékeket a prognosztikai vizsgálatok eredményei alapján határozták meg). Ha erre nincs lehetőség, akkor statisztikai eszköz segítségével próbálkozhatunk, küszöbérték-definiálással. Például a kontrollcsoportban megfigyelt eloszlás alapján határozhatunk meg dichotomizálási küszöböt (egy újszerű laboratóriumi adat esetén, aminek még nem ismert pontosan a klinikai jelentősége és a normál tartománya, a kontrollcsoport 97,5 percentilis értékét használhatjuk küszöbként, ha a magas érték tűnik kórjelzőnek, és a 2,5 percentilist, ha az alacsony.



3. ábra. Statisztikai változók legalapvetőbb tulajdonságai

Leíró statisztika

A kísérleteken vagy megfigyelésen alapuló vizsgálatok eredményeinek értékelésekor statisztikai eszközök segítsége nélkül nem tudunk következtetéseket levonni. A vizsgálatok során keletkező adatokból levont következtetések nem csak a statisztikai elemzések eredményeire támaszkodnak, de a megalapozott statisztikai következtetések nélkül nem lehet a vizsgálatok végén érdemi konklúzióra jutni. Miért kötelező elemei a jó biostatisztikai elemzések minden kutatási, vizsgálati projektnek? A válasz tulajdonképpen roppant egyszerű, ha végiggondoljuk azt a két problémát, amivel akkor szembesülünk, ha egy kutatási kérdést kellő pontossággal meg szeretnénk válaszolni.

Az egyik az, hogy akármilyen körültekintően is járunk el a vizsgálati minta összeállításakor, és akármennyire is törekszünk az elemszámnövelésre, soha nem tudunk azon az elvi korláton átlépni, amit a minta és a populáció közti határvonal jelent. A populáció egészére jellemző, tehát a valóság tényleges paramétereinek a megismerésére törekszünk (azt szeretnénk tudni, hogy mennyi az egészséges emberek vörösvértestszáma, milyen szoros a kapcsolat az elhízás és a vastagbélrák kialakulásának kockázata között stb.), de csak a valóság egy kis szeletét jelentő mintát tudjuk ténylegesen megvizsgálni. A minta soha nem tudja pontosan megjeleníteni a valóságot, ezért a következtetéseink soha nem teljesen pontosak.

A másik probléma, hogy az élő szervezet rendkívül bonyolult. Nagyon sok tényező együttműködése révén állnak elő a legelemibb jelenségei is. A vérnyomás egyszerű élettani paraméter. Ha meg akarjuk érteni, hogy valakinek miért éppen annyi a vérnyomása, amennyit mértünk, akkor szembesülünk azzal, hogy ha csak a már megismert szabályozó rendszereket, és az ezekre ható külső tényezők egymásra hatását szeretnénk összefoglalni, akkor is szó szerint csak a könyvtárakban tárolt, egy ember által már valóban áttekinthetetlen szakirodalmat kellene ismernünk. Ehhez az is hozzájárul, hogy még a vérnyomással kapcsolatban sincs olyan érzése senkinek, hogy majdnem mindent tudunk róla (kell még hely a könyvtárban a jövőben feltárt ismereteknek).

A legegyszerűbb szervezetek elemi tulajdonságai is sok folyamat eredőjeként alakulnak ki. Ennek következtében a mérhető tulajdonságok változatos értékeket mutatva jelennek meg. Ha kérdésünk van ezekkel a tulajdonságokkal kapcsolatban, például azt szeretnénk tudni, hogy milyen szoros a kapcsolat az elhízás és a vastagbélrák kialakulásának kockázata között, akkor jelentős változékonyságot mutató testtömegindexű és

egymástól jelentős mértékben eltérő, egyéni megbetegedési kockázatot hordozó emberek adatai alapján kell választ adnunk. Úgy, hogy nem is tudjuk méréseinkkel lefedni az összes olyan tényezőt, ami alakítja a testtömeget és a megbetegedési kockázatot. Egyfelől a források elégtelenek arra, hogy minden ismert befolyásoló tényezőt ténylegesen mérjünk, másfelől nem is ismerünk minden tényezőt. Ha akarnánk, se tudnánk tehát mindenre kiterjedő vizsgálatot vezetni, amiben matematikai összefüggések révén és hiba nélkül tudnánk leírni a folyamatok dinamikáját és végeredményét.

Röviden összefoglalva, olyan a világ, hogy nem lehet benne a jelenségeket minden részlet vonatkozásában megismerni, és el kell fogadnunk, hogy ezek a jelenségek változatos formában vesznek minket körül.

Kezünket feltartva mégsem adhatjuk fel, hogy az emberi szervezet működésére vonatkozó kérdésekre válaszoljunk, mert azt azért valamilyen módon el kell érnünk, hogy ha nem is teljesen, de minél jobban értsük az egészséges és a beteg szervezet működési sajátosságait. Hiszen a betegeket szeretnénk gyógyultan, az egészségeseket pedig minél tovább egészségesnek látni. Ehhez pedig tudományosan megalapozott ismereteken nyugvó eljárások alkalmazására van szükség.

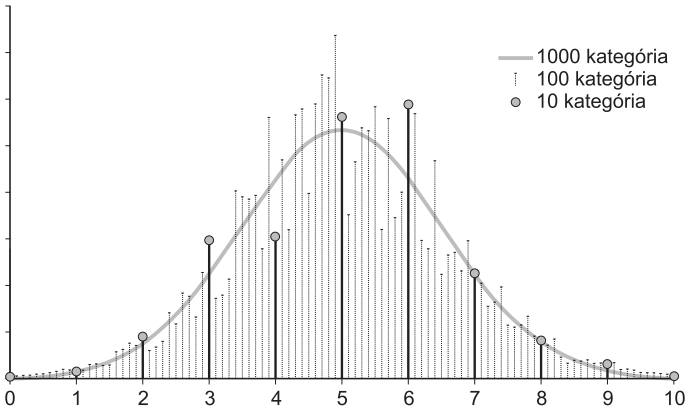
A megértés kényszere és a jelenségek változékonysága által berendezett terepen akkor tudunk előrehaladni, ha tömegjelenségként kezeljük a vérnyomást, a csontok kalciumtartalmát, a daganatos betegségek 5 éves túlélését, az inzulinszérum koncentrációját. Tömegjelenségként, amelyet nem mutat egy pontos érték (mint a fénysebesség esetében), hanem változékony. Emiatt nem egyszeri (de nagyon pontos) méréssel lehet őket meghatározni, hanem a szóródásuk meghatározásával. Utóbbi feladatra dolgozták ki a statisztikai módszereket. Ezek a vizsgálati eszközök teszik lehetővé, hogy leírjuk a biológiai tulajdonságokat azok szóródásának bemutatásán keresztül, és kapcsolatot keressünk a különböző tulajdonságok között, vagyis megértsük, hogy az egyes tulajdonságok változékonysága miként kapcsolódik a másik jelenség változékonyságához. A jelenségek szóródásának bemutatása a legelemibb statisztikai feladat, amit meg kell oldani az elemzéseink során. Aztán majd erre lehet felépíteni azokat a vizsgálatokat, amik összefüggést keresnek – szintén statisztikai eszközökkel – a szóródást mutató jelenségek közt.

A gondolatmenet zárásaként meg kell jegyezni, hogy a vizsgálataink tervezésekor igyekszünk olyan körülményeket teremteni, hogy a vizsgált jelenségek változékonysága minél kisebb mértékben korlátozzon minket, minél szűkebb legyen a variabilitás. A vizsgálati körülmények közt megmaradó teljes változékonyságot aztán alapvetően két részre bontjuk: az egyik rész, ami a nem ismert és az éppen kivitelezett vizsgálatban nem figyelt folyamatok révén áll elő; a másik, ami a már ismert, és a vizsgálatban

tekintetbe vett folyamatokkal kapcsolatos. Utóbbi változékonysági forrását próbáljuk a vizsgálatokban kiiktatni, hogy minél kisebb legyen a nem magyarázott része a variabilitásnak. Ezen a módon javul a képességünk arra, hogy megválaszoljuk vizsgálatunk alapkérdését. A kísérletek végzésekor arra törekszünk, hogy a bonyolult rendszer legmeghatározóbb elemeit standardizáljuk (azaz minden vizsgálati alany számára azonos körülményeket teremtsünk, a kezelési csoportok közti eltérések vizsgálatakor a variabilitás forrása csak a nem standardizált tényezők hatásaiból adódjon). Megfigyelésen alapuló vizsgálatok alkalmával pedig általában adatokat gyűjtünk a legfontosabb befolyásoló tényezőkre, és ezekre az adatokra támaszkodva korrigáljuk a megfigyelés eredményeit (eltávolítjuk valamilyen többváltozós statisztikai módszer segítségével a variabilitás ismert, befolyásoló faktorokkal kapcsolatos részét).

Hisztogram

A vizsgált jelenségek szóródását első lépésként érdemes ábra segítségével szemléltetni. A hisztogram az egyes megfigyelt értékek előfordulási gyakoriságát, valószínűségét ábrázolja *(4. ábra)*.



4. ábra. Egy biológiai paraméter (átlag: 5; szórás: 1,5) hisztogramja a mérés pontosságának függvényében az egész értékek (10 kategóriába osztott mérési eredmények), a tizedek (100 kategóriába osztott mérési eredmények) és a századok (1000 kategóriába osztott eredmények) mérésére képes mérőműszerek segítségével nyert adatok alapján (mindhárom sorozatban 1000 a ténylegesen kivitelezett mérések száma)

Diszkrét adatfajta esetében az eleve adott kategóriákat ábrázoljuk az x-tengelyen, és az egyes kategóriákhoz tartozó esetek számát mutató oszlopokból adódik a diagram. Folytonos változó esetén a kategóriahatárokat mesterségesen kell előállítani. Ezek számát az adott helyzethez igazodóan kell meghatározni, az adatok tartománya alapján egyenlő szélességű kategóriákat definiálva. Ha sok adat áll rendelkezésünkre, akkor a folytonos adatot egyre keskenyebb kategóriák definiálása után, egyre finomabb felbontásban ábrázolva, az oszlopdiagram átalakul folytonos vonallá. Ezek az ábrák önmagukban nem elégségesek a vizsgált jelenségek leírására, de sok információt hordoznak. Például megmutatják, hogy az élettani paraméterek döntő többsége olyan eloszlást mutat, aminek centruma van (ami közelében sűrűsödnek az egyes mérési eredmények), és aminek a centrumától távolodva fokozatosan egyre kevesebb lesz a mérési eredmények száma.

Centrális érték

A hisztogramok segítségével szemléltethető, hogy mi a tipikus vizsgálati alany jellemző értéke, mi az eloszlás centruma, amit x értékek N elemű mintája esetén legkézenfekvőbb $\overline{x}$ számtani átlagként megadni:

$$\overline{x} = \frac{\sum x}{N}.$$

Ez csak akkor szokott problémás lenni, ha az adataink között van néhány extrém érték, ami kilóg a hisztogramból is. Ezek az extrém értékek igen nagy hatással vannak az átlagra. Ha ilyen adatokkal együtt számolunk, akkor az lesz az érzésünk, hogy a számított átlag nem jól tükrözi a tipikus vizsgálati alanyt. Az extrém adattal kapcsolatban pedig, hogy valami oknál fogva speciális volt a szélsőséges értéket mutató vizsgálati alany, és jobb lenne megérteni a speciális érték mögött álló speciális körülményt, mint figyelembe venni a tipikus érték számításakor! Vagyis jobb lenne kizárni a szélsőséges adatot a feldolgozásból.

Extrém adatok jelenlétében a tipikus résztvevőt jobban bemutató mérőszámhoz jutunk, ha a vizsgálat eredményeit sorba rendezzük, és kiválasztjuk az éppen középen lévőt. Az így kapott érték a medián. Mivel a medián számításakor nincs jelentősége a szomszédos adatok közti különbség nagyságának, az extrém értékek hatása sem fog érvényesülni olyan nagymértékben, mint az átlag esetében. Ugyanezért viszont a medián nem tükrözi a tényleges adatok közti különbséget.

Az adatok eloszlásának jellegétől függően tehát használhatjuk az eloszlásról részletesebb információt adó átlagot, vagy az eloszlás részleteiről keveset mondó, de az extrém értékekre kevéssé érzékeny mediánt.

Inkább csak a teljesség kedvéért kell megemlíteni, hogy diszkrét adatoknál a legnagyobb számban előforduló érték, illetve folytonos adatnál a hisztogram csúcsához tartozó érték a módusz. A tipikus eset leírására ezt a mérőszámot egészségtudományi területen alig használjuk.

Szóródás

A centrális érték önmagában nem írja le azt a jelenséget, amivel foglalkozunk, hiszen a tipikus érték körül szóródnak a tényleges adataink. A szóródás pedig különböző adatok esetében jelentősen eltérő lehet. A szóródás számszerűsítése ugyanolyan alapfeladat, mint a centrális érték meghatározása.

A legegyszerűbben az adatok tartományát számíthatjuk (a minimum és maximum értékek alapján). Sajnos az egyszerű számíthatóság nem párosul komoly gyakorlati értékkel. Ez a mérőszám ugyanis mindösszesen két adatot, ráadásul két extrém adatot hasznosít. Egyáltalán nem szól az adatok túlnyomó többségéről.

Informatívabb szóródás-mérőszámokhoz jutunk, ha az adatok sorba rendezése után egyenlő darabszámú adatot tartalmazó tartományokat definiálunk, és ezek határait, azaz kvantiliseket adunk meg. Ha három, négy, öt vagy tíz egyenlő tartományt adunk meg akkor tercilisekről, kvartilisekről, kvintilisekről, illetve decilisekről beszélhetünk. A kvantilisek határai már többet mondanak a tartományon belüli eloszlásról. De ez a mutató sincs tekintettel arra, hogy milyen a kategóriákon belül az adatok eloszlása.

Érdemes olyan szóródás-mérőszámot kialakítani, ami az összes adat eloszlását tükrözi, úgy, hogy tekintetbe veszi az adatok közti tényleges különbségeket is. Ilyen mutató a centrális érték körüli szóródást kell, hogy szemléltesse, azaz az adatok alapján számított átlag körüli szóródást kell leírnia. Ehhez az első lépés az egyes adatok átlagtól való $(x-\bar{x})$ eltérésének számítása. Ha ezeket a távolságokat összegezzük, akkor összességében képet kapunk arról, hogy milyen mértékű a szóródás. Ha csak egyszerűen összeadjuk az egyes átlagtól való eltéréseket, akkor akármilyen is az adatok szóródása, az összeg éppen nulla lesz, mivel az átlag alatti és az átlag feletti adatokhoz tartozó eltérések éppen kioltják egymást: $\sum(x-\bar{x})=0$. Két mód is kínálkozik arra, hogy ezt az előjelproblémát megoldjuk. Az eltérések abszolút értékét vagy négyzetét használva csupa pozitív számot kapunk, amiket összegezve már olyan mutatókat kapunk (abszolút

eltérések összege: $\sum|(x-\bar{x})|$; négyzetes eltérések összege: $\sum(x-\bar{x})^2$), amelyek nagyobb értékei tükrözik a nagyobb szóródást. Ezek a mutatók félrevezetőek lehetnek, ha egy kicsi elemszámú és egy nagy elemszámú vizsgálat adatait hasonlítjuk össze, mert a nagyobb elemszám mellett több eltérésből adódó összegzett különbséget látunk. (Ha a nagy és a kicsi mintán ugyanolyan az adatok átlag körüli szóródása, akkor a nagy mintán nagyobbnak adódnak ezek a szóródást mérő kifejezések.) Valamilyen módon a vizsgálat méretét is figyelembe kellene tehát vennünk ahhoz, hogy ne csak az azonos nagyságú minták adatait tudjuk összehasonlítani. Kézenfekvő lenne egyszerűen az adatok számával osztani az abszolút vagy a négyzetes eltérések összegét. Ehelyett a vizsgálat méretét szabadsági fokkal írjuk le, ami az elemszám korrigált értéke. Így definiáljuk a V_x varianciát (az egyes adatok számtani átlagtól való négyzetes eltéréseinek átlagát) és a D_x átlagos eltérést (az egyes adatok számtani átlagtól való abszolút eltéréseinek átlagát):

$$V_x = \frac{\sum(x-\bar{x})^2}{N-1},$$

$$D_x = \frac{\sum|(x-\bar{x})|}{N-1}.$$

Az átlagos eltérés szemléletes értelmezése elég egyszerű. A varianciáé a négyzetre emelt érték (és a négyzetre emelt dimenzió) miatt nem ilyen egyszerű. A variancia gyöke viszont már szemléletes mérőszám (az egyes adatok számtani átlagtól való eltéréseinek átlaga, standard deviáció, SD_x):

$$SD_x = \sqrt{V_x} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

A szórás és az átlagos eltérés számszerűen közeli, de nem azonos érték. Önmagában mindegyik alkalmas a szóródás szemléletes leírására. A statisztikai feldolgozások további lépései viszont lényegében csak a szórást használják. Emiatt az átlagos eltérés nem tekinthető lényeges statisztikai mérőszámnak.

Szabadsági fok

A mérési eredmények feldolgozásakor statisztikai mutatókat, mérőszámokat határozunk meg. Vannak mérőszámok (pl. minta átlaga), amelyek kiszámítása közvetlenül a minta egyes elemeiből történik. Más esetekben nem csak magukból a vizsgálat során nyert elemi adatokból, hanem köztes mutatókból (pl. csoportátlagokból) számítjuk a mérőszámot (Jellemzően ilyen statisztikai tesztek mérőszámai a teszt-statisztikák). Ilyenkor a mérőszám meghatározásakor nem a minta nagyságával írjuk le a vizsgálat méretét, hanem a szabadsági fokkal (degree of freedom; df). Annyival lesz kevesebb a minta elemszámánál a mérőszám szabadsági foka, amennyi (mintából számított) köztes mérőszám értékét felhasználjuk a végső mutató értékének meghatározásához.

Minden statisztikai elemzés a teljes populációra vonatkozóan szeretne következtetést levonni az elemzett minta adatai alapján. A minta nagysága meghatározza, hogy mennyire megbízható a levont következtetés. Ezért minden statisztikai mutató értékelésekor figyelembe kell vennünk a vizsgálat méretét megadó szabadsági fokot.

Az egyes statisztikai mérőszámok szabadsági fokának számítási módját talán célszerűbb minden mérőszám esetében megjegyezni, mint a számítás menetének átgondolása révén kikövetkeztetni.

Ha egy kezelt és egy kontrollcsoport összehasonlítása révén szeretnénk értékelni egy beavatkozás hatékonyságát, akkor t-próba segítségével hasonlíthatjuk össze a beavatkozás eredményét jelző klinikai paraméter átlagát. A t-érték kiszámításához felhasználjuk a két csoport átlagait, ezért a statisztikai mutató értékelésekor nem azt vesszük figyelembe, hogy milyen nagy volt a vizsgált minta, hanem annak kettővel csökkentett értékét.

Normális eloszlás

Ha különböző mintákon megmérjük egy biológiai paraméter értékét, és elkészítjük a minták hisztogramját, akkor azt látjuk, hogy a hisztogramok alakja nagyon hasonló. Általános az a tapasztalat, hogy a biológiai paraméterek jellemző eloszlásmintázattal rendelkeznek.

Szintén az a tapasztalat, hogy a legtöbb paraméter eloszlása szimmetrikus, harang alakú görbét ír le. A harangalak abból adódik, hogy az eredmények az átlagérték körül sűrűsödnek, attól távolodva fokozatosan ritkulnak. Ezt az általánosan megfigyelhető haranggörbe-eloszlásmintát (Gauss-görbét) matematikai függvényként is le lehet írni.

A vizsgálati eredmény és annak előfordulási valószínűsége közti kapcsolatot leíró függvény a normális eloszlás sűrűségfüggvénye:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

ahol x a vizsgálati eredmény, $f(x)$ a vizsgálati eredmény előfordulási valószínűsége, μ a vizsgált paraméter átlaga, σ a vizsgált paraméter szórása (μ-vel és σ-val a vizsgálati paraméter populációs szinten jellemző, tehát valódi átlagát és standard deviációját jelöli, amiket meg kell különböztetnünk a mintán mért $\bar{x}$ átlagtól és SD_x szórástól).

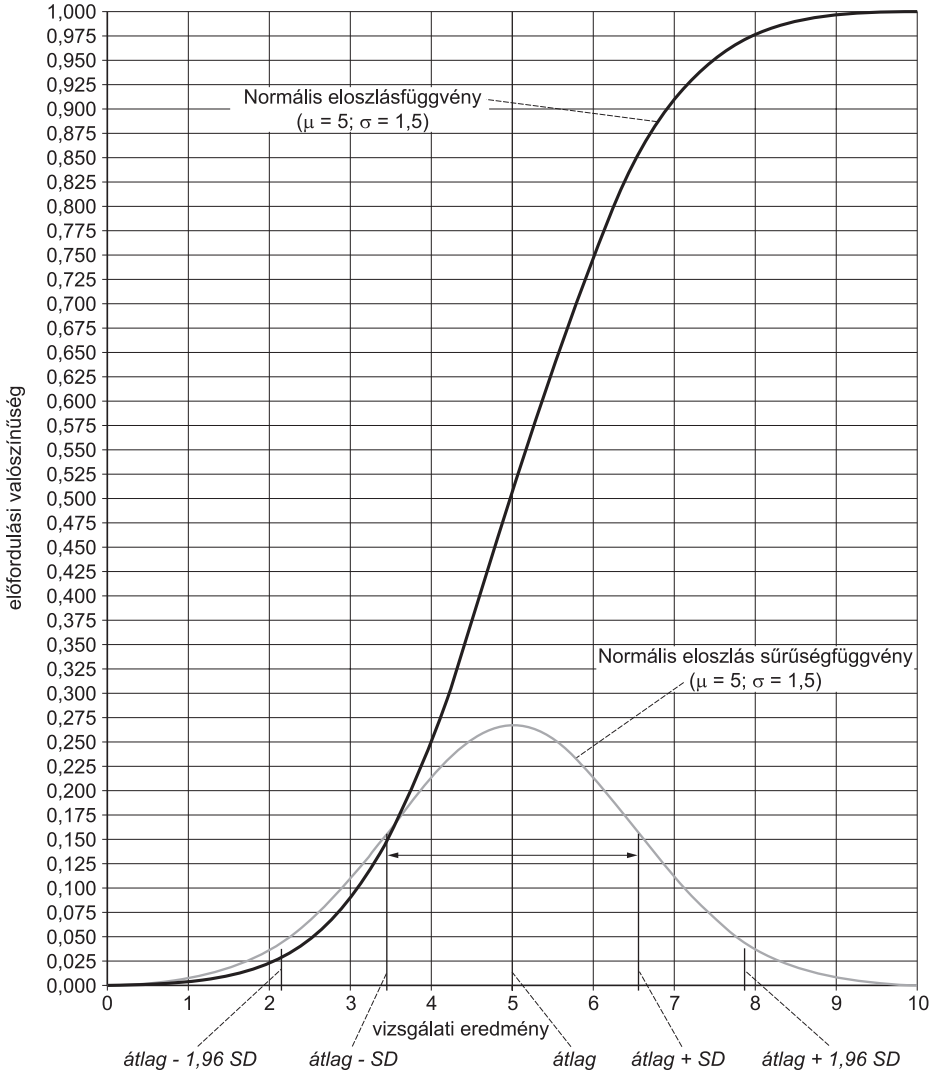
Ha azt ábrázoljuk, hogy egy adott mérési eredménynél kisebb mérési eredmény milyen valószínűséggel fordul elő, akkor a normális eloszlásfüggvényt kapjuk.

A normális eloszlás szimmetrikus, középpontja az átlag (ami itt egyenlő a mediánnal). A haranggörbe inflexiós pontjaihoz (ahol összekapcsolódnak a sűrűségfüggvény kifelé domború és homorú szakaszai) tartozó vizsgálati eredmény és az átlag közti távolság a szórás. Az átlaghoz 1 szórásnyinál közelebb van a vizsgálati eredmények 68,26%-a, 1,96 szórásnyira pedig a 95%-a. Azokat az eredményeket, amik az átlagnál legalább 1,96 szórásnyival kisebbek, extrémen alacsonynak tekintjük. Az összes adat 2,5%-a tartozik ebbe a csoportba. Az átlagnál legalább 1,96 szórásnyival nagyobb eredményeket tekintjük extrémen magasnak. Ezek is 2,5%-át adják az összes mérési eredménynek *(5. ábra)*.

Különböző vizsgált jelenségek eltérő átlaggal és szórással rendelkeznek, ezért maga a sűrűségfüggvény minden adatra más és más lesz. A függvények alakja azonban független az eloszlási paraméterektől (az átlagtól és a szórástól).

Ha az egyes mérési eredményeinkről szeretnénk véleményt mondani (mennyire tekinthető az szélsőségesnek vagy a centrumhoz tartozónak), akkor kényelmesebb, ha nem az eredeti adatok segítségével készített, hanem a standardizált vizsgálati eredmények segítségével felvett sűrűségfüggvényt használjuk. Azaz minden mérési eredményből kivonva az átlagot, eltoljuk a haranggörbét úgy, hogy alakját megtartva, annak középpontja pont 0 legyen. Majd ezeket az átlaggal csökkentett vizsgálati eredményeket a szórással osztva, a görbe alakját úgy alakítjuk, hogy annak szórása éppen 1 legyen. Ilyenkor a mérési mértékegységtől, annak átlagától és szórásától függetlenül, maga a standardizált érték (z érték) leírja a centrumhoz való viszonyt. A standardizált z érték átlaga mindig 0, a –1,96-nál alacsonyabb értékek extrémen alacsonynak, az 1,96-nál nagyobbak szélsőségesen magasnak tekinthetők.

$$z = \frac{x - \bar{x}}{SD_x}$$



5. ábra. Normális eloszlás és sűrűségfüggvény

Az ábrán szereplő adat normális eloszlású, átlaga 5, szórása 1,5. Ha *A* és *B* vizsgálati résztvevő esetében a mérés szerint x_A = 8x és x_B = 2,5, akkor mit tudunk mondani a két résztvevő mérési eredményéről? Az eredeti mérési eredményeknek megfelelően leolvashatjuk az eloszlásfüggvény értékét ennél a két pontnál. x_A magasabb annál a pontnál, ahol az eloszlásfüggvény értéke 0,975, azaz extrém magasnak tekinthető. x_B magasabb, mint az az érték, ahol az eloszlásfüggvény értéke 0,025, ezért ezt nem tekinthetjük extrémen alacsonynak. Kényelmesebb azonban a standardizált értékeket számítani:

$$z_A = \frac{x - \bar{x}}{SD_x} = \frac{8-5}{1{,}5} = 2 \qquad z_B = \frac{x - \bar{x}}{SD_x} = \frac{2{,}5-5}{1{,}5} = -1{,}66$$

Mivel z_A nagyobb, mint 1,96, az *A* résztvevő adata extrémen magasnak tekinthető. *B* résztvevő viszont a többséghez sorolható, mert a z_B = –1,66 nem kisebb, mint –1,96.

Megbízhatósági tartomány

Statisztikai elemzést azért végzünk, hogy a vizsgálat kérdését segítsük megválaszolni. Az alapkérdés lehet egyszerűen egy jelenség gyakoriságának a meghatározása (mennyi folsavat visznek be átlagosan táplálkozás során a felsőfokú végzettséggel rendelkező nők a terhesség korai szakaszában?), vagy lehet két jelenség közti kapcsolat igazolása (rontja-e a sejtmembrán áteresztőképességét, ha a szövettenyészethez nagy dózisú béta-karotint adunk?)

Ha a vizsgálati eredményünk ismeretében átnézzük a szakirodalmat, ha konferencián beszélgetünk a kollégáinkkal, akik hasonló kérdéseket tanulmányoznak, mint mi, akkor azt fogjuk látni és hallani, hogy mások hasonló, de a mienktől számszerűen különböző eredményeket kaptak. Mindaddig nyugodtak maradunk, amíg az eredmények alapján levont következtetések ugyanazok, mint amire mi is jutottunk. Miért is van ez?

Ha valóban vizsgálatot akarunk végezni a „mennyi folsavat visznek be átlagosan táplálkozás során a felsőfokú végzettséggel rendelkező nők a terhesség korai szakaszában?" kérdés megválaszolására, akkor gyorsan rájövünk, hogy csak a fizikai lehetőségeink által megszabott korlátok közt tudjuk azt a vizsgálati csoportot összeállítani, aminek a táplálkozási szokásait majd fel fogjuk mérni. Bármennyire is szeretnénk nagyon nagy elemszámú mintát vizsgálni, meg kell majd elégednünk valamilyen kompromisszumos megoldással. Természetesen azért szeretnénk minél nagyobb mintát vizsgálni, mert azt gondoljuk, ha nagy a minta, akkor megbízható a belőle származó eredmény, és tisztában vagyunk azzal is, hogy a mintánk nem képes az összes elvileg lehetséges vizsgálati alanyra jellemző viszonyokat tökéletesen reprezentálni. Vagyis tudjuk, hogy a mintán kapott eredményeink csak több-kevesebb pontossággal tükrözik a valós, megállapítani kívánt paramétert. Ezért nem lepődünk meg, ha egy másik mintán végrehajtott (egyébként a mienkhez teljesen hasonló és kifogástalan módszertani következetességgel kivitelezett) vizsgálat eredménye csak hasonlít a mi eredményeinkre. Ha több eredményt látunk, akkor azokat fogjuk megbízhatóbbnak érezni, melyek nagyobb mintát vizsgáltak.

Összefoglalva mindezt elmondható, hogy a valóságra jellemző paraméterek megállapítására törekszünk, de ehhez nem tudjuk az összes potenciálisan szóba jöhető vizsgálati alanyt, azaz a teljes populációt megvizsgálni. Ehelyett csak egy reprezentatív mintát tanulmányozunk. A mintán kapott eredmények szoros kapcsolatban lesznek ugyan a valós paraméterekkel, de azzal számszerűen nem egyeznek meg, csak több-kevesebb pontossággal tükrözik a valós viszonyokat.

Adódik a kérdés, hogy akkor érdemes-e bármit vizsgálni? A válasz igen, ha képesek vagyunk arra, hogy az eredmények bizonytalanságát számszerűen kifejezzük, ezáltal adva lehetőséget az eredmények gyakorlati hasznosítására. A kvantifikáláshoz végezzünk el gondolatban egy kísérletet a folsavbeviteli kérdés megválaszolására! A populáció, amit meg szeretnénk vizsgálni, az összes korai terhesség időszakában lévő nő. Ennek a populációnak természetesen van átlaggal és szórással jellemezhető folsavbevitele. Objektív, létező szám mindkettő. A gondolatkísérletben az átlagot próbáljuk majd meghatározni.

A teljes populációból véletlenszerűen válogassunk ki sok, azonos N elemszámú mintát, és ezeken mérjük fel a folsavbevitelt, majd számoljuk ki az átlagokat. (Nyilvánvalóan ez a valóságban kivitelezhetetlen!) Nem lepődünk meg azon, hogy a kapott átlageredmények szóródást mutatnak, és normális eloszlásúak. A szóródás centruma az a valós átlagos folsavbevitel, aminek a meghatározása egyébként a célunk. Az átlagértékek szóródásának a mértéke attól függ, hogy mekkora volt az egyes nők folsavbevitelének szórása. Ha széles tartományon belül variálódik az egyéni folsavbevitel, akkor a gondolatkísérletben kapott mintánkénti átlagok is szélesen szóródnak. (Bár nyilvánvalóan nem annyira szélesen, mint az eredeti adatok!) Azt is gondoljuk, hogy minél nagyobb a gondolatkísérletben használt mintanagyság, annál megbízhatóbbak lesznek az egyes minták átlagai, ez annyit jelent, hogy jobban tükrözik a valóságot, közelebb helyezkednek a valós átlaghoz, szűkebb tartományon belül szóródnak. Végső soron azt tudjuk megállapítani, hogy a szóródás mértéke egyenesen arányos az elemi adatok standard deviációjával, és fordítottan arányos az elemszámmal, pontosabban (nem érdemes itt a matematikai magyarázatra kitérni, de) annak gyökével. A nevezéktani káoszt elkerülendő a standard deviációt fenntartjuk az egyéni szinten meghatározott vizsgálati eredmények eloszlásának leírására (SD_x), a gondolatkísérletből származó mintaátlagok $SD_{\bar{x}}$ standard deviációját pedig átkereszteljük SE standard hibának. (De tudjuk, hogy egyébként ez a mintaátlagok Gauss-görbéjének a szórásával egyenlő!) A szemléletes jelentés matematikai megalapozásától eltekintve, meg tudjuk állapítani, hogy:

$$SE = \frac{SD_x}{\sqrt{N}}.$$

Tudjuk, hogy az átlagok eloszlása normális, aminek a középpontjában az általunk keresett valódi $\bar{X}$ átlagos folsavbevitel van (nagybetűvel a való populációra jellemző, kis betűvel a mintán megfigyelt adatokat szoktuk jelölni), és hogy ilyen eloszlás mellett az adatok 95%-a az átlag 1,96 szórásnyi közelében van. Azaz, meg tudjuk adni annak

a *T* tartománynak a szélességét, amin belül jellemzően (95%-os valószínűséggel) előfordulnak a gondolatkísérlet eredményei:

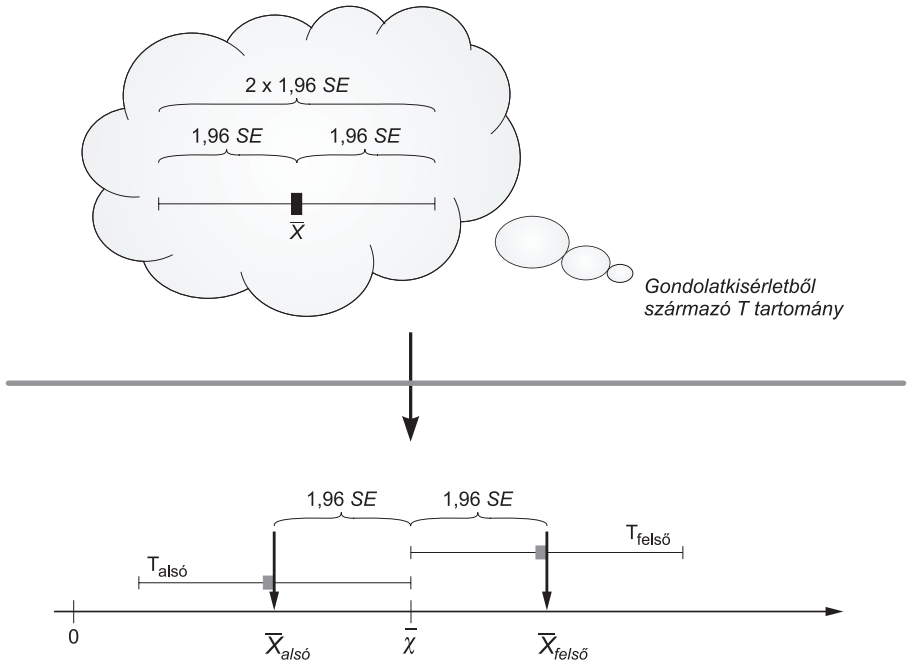
$$T = \overline{X} \pm 1{,}96SE = \overline{X} \pm 1{,}96\frac{SD_x}{\sqrt{N}}.$$

A gondolatkísérletnek az volt az értelme, hogy levezessük ezt az összefüggést, amely szerint *T* tartomány kiterjedése $2 \times 1{,}96\frac{SD_x}{\sqrt{N}}$. Ha ez a tartomány széles, akkor a minták átlagai széles tartományon belül variálódnak, ami alapján az egyes vizsgálati eredményeket kevéssé megbízhatónak érezzük. Ha viszont a *T* tartomány szűk, akkor az eredmények is szűk tartományon belül variálódnak, amiből azt következik, hogy a gondolatkísérletből származó mintaátlagok közel helyezkednek el a valós átlaghoz, és ezért megbízhatóbbnak érezzük őket. A tartománynak nem csak az a pozitívuma, hogy kifejezi, hogy mennyire megbízható egy vizsgálatnak az eredménye, hanem az is, hogy olyan adatoktól függ (elemszám és szórás) amelyeket egyetlen minta segítségével is meg tudunk állapítani. (Ezt a kijelentést majd a továbbiakban pontosítani kell!)

Ha tehát egyetlen vizsgálatot végzünk el, akkor annak segítségével ez a *T* tartomány meghatározható. Az egyetlen komoly gond a *T* tartománnyal, hogy nem tudjuk, hol helyezkedik el a számegyenesen, csak azt, hogy milyen széles. Ha rá tudnánk illeszteni a számegyenesre, akkor nem csak a határait, hanem a középpontját is pontosan (értsd: hiba nélkül!) meg tudnánk adni. Ez annyit jelentene, hogy a valóságnak ezt a paraméterét pontosan meg tudtuk határozni. A bevezetőben pont arról beszéltünk, hogy ez elvileg nem lehetséges!

Milyen a viszony a számegyenes és *T* tartomány közt? Ennek megválaszolásában segít minket az az adat, amit az egyetlen saját vizsgálatunkból kapunk, de ebben a gondolatmenetben még nem is vettünk figyelembe: a saját mintánk adataiból számított $\overline{x}$. Erről csak annyit tudunk, hogy egy a lehetséges nagyon sok, a gondolatkísérletben elvégzett vizsgálat eredményéből, azaz valahol benne van a *T* tartományban (pontosabban 95%-os valószínűséggel van a *T* tartományban). Segítségével hozzákapcsolhatjuk a *T* tartományt a számegyeneshez, mert azt úgy kell illeszteni, hogy akárhol is lehet, de legyen benne az $\overline{x}$, ami eleve a számegyenesen is van. Ez annyit jelent, hogy a *T* tartomány két szélsőértékét ($T_{alsó}$ és $T_{felső}$) tudjuk felvenni, azaz rögzíteni tudjuk a *T* tartományt a számegyenesre, de nem egy fix helyen, hanem két szélsőérték közt.

A 6. ábrából elég világosan kitűnik, hogy a *T* tartomány két szélső helyzete a középpontjukban levő valós átlag szélső pozícióit is ($\overline{x}_{alsó}$; $\overline{x}_{felső}$) meghatározza, mely pozíciók a saját minták átlagától 1,96 *SE* távolságra vannak.



6. ábra. Megbízhatósági tartomány határainak értelmezése

Ezzel a módszerrel meg tudjuk határozni, hogy milyen szélsőértékek közt helyezkedik el a folsavbeviteli átlag valós értéke. Meg tudunk adni egy tartományt, amin belül van valahol (nem tudjuk, hogy hol!) a valós átlag. Ezzel az intervallummal kvantifikáltuk a saját vizsgálati mintákon meghatározott átlag megbízhatóságának mértékét, megnyitva a lehetőséget az eredmény gyakorlati alkalmazására. (Nem mellesleg úgy tettük ezt, hogy harmóniában maradtunk azokkal a filozófiai gondolatmenetekkel is, amelyek szerint teljes részletességgel nem lehet megismerni a valóságot!) A származtatott intervallumot az átlag 95%-os megbízhatósági tartományának definiáljuk:

$$MT_{95\%,\bar{x}} = \bar{x} \pm 1{,}96 \frac{SD_x}{\sqrt{N}}.$$

A képletben szereplő standard deviációval kapcsolatban nyitva hagytunk azonban egy kérdést! A gondolatkísérletben csak annyit említettünk vele kapcsolatban, hogy a saját mintánkon számított szórást használjuk a számításainkhoz, azt feltételezve, hogy

ez a minta alapján meghatározott SD_x szórás a populációra jellemző σ valós szórással azonos. Pedig nyilvánvaló, hogy a szórások is variábilisak lennének a gondolatkísérletben. A mintánk adatai alapján számított szórás csak több-kevesebb pontossággal közelítené a valós, populációra jellemző σ szórást.

A 95%-os megbízhatósági tartomány képletének valamilyen korrekciójára van ezért szükség. A levezetett képlet alapján ugyanis szűkebb T tartományt kapunk, mint ami valóban tartalmazza a gondolatkísérletből származó mintaátlagok 95%-át. A korrekció az 1,96-os konstans módosításával végezhető el, aminek a matematikai részleteit nem tárgyaljuk, csak az eredményét állapítjuk meg. A képletben szereplő konstanst az $N - 1$ szabadsági fokú t-eloszlásból számított értékkel helyettesítjük. Így az átlag 95%-os megbízhatósági tartományát az alábbi módon tudjuk saját vizsgálatunk végén számítani:

$$MT_{95\%,\bar{x}} = \bar{x} \pm t_{[N-1;\alpha=0,05]} SE = \bar{x} \pm t_{[N-1;\alpha=0,05]} \frac{SD_x}{\sqrt{N}}.$$

Egyetlen vizsgálat végén tehát meg tudjuk mondani, hogy a valóság egy paramétere milyen határok közt helyezkedik el 95%-os megbízhatósággal. A vizsgálat eredménye tulajdonképpen az, hogy be tudjuk szorítani a korábban nem ismert értéket ebbe a tartományba. Nyilvánvaló, hogy minél nagyobb mintát vizsgálunk, annál szűkebb lesz ez a tartomány, ami teljesen összhangban van azzal a gondolatkísérlet előtt megfogalmazott érzésünkkel, hogy, minél nagyobb mintát vizsgáltunk, annál megbízhatóbb az eredményünk.

Meg kell jegyezni még, hogy a 95%-os megbízhatóság a precizitás és a megbízhatóság közti gyakorlati kompromisszum. A kutatási programok eredményeinek sikeres gyakorlati alkalmazása bizonyítja, hogy ez a megbízhatósági szint általában megfelelő. (Ha nem is véd meg minden tévedéstől! Részletesebben követjük majd ezt a gondolatmenetet a hipotézistesztelés értelmezésekor.) Esetenként azonban nem elégszünk meg a 95%-os megbízhatósággal, és szeretnénk megadni azt a tartományt, amin belül 99%-os valószínűséggel van a valós átlag. A nagyobb megbízhatóságnak az lesz az ára, hogy szélesíteni kell a megbízhatósági tartományt. A t-érték $\alpha = 0,01$ -re megadott értékeit használva már tudjuk a 99%-os megbízhatósági tartományt is számítani.

A mintánkon végzett vizsgálat során nem csak átlagértéket tudunk meghatározni, hanem bármilyen leíró statisztikai vagy összefüggés-elemzést szolgáló statisztikai mérőszámot. A fenti gondolatmenet végigvezethető bármelyik mérőszámra. A végeredmény mindig egy megbízhatósági tartomány lesz, aminek a képlete is teljesen azonos lesz a levezettel. Egyetlen különbséget fogunk találni, ami a meghízhatósági tartományokat mérőszám-specifikussá teszi. Ez a standard hiba számításának a módja. Az egyes mérő-

számokhoz tartozó standard hiba ismeretében már megadható bármelyik *M* mérőszám $(1 - \alpha)$%-os megbízhatósági tartománya *N* elemű minta vizsgálata után:

$$MT_{(1-\alpha)\%,M} = M \pm t_{[N-1;\alpha]} SE_M .$$

A 95%-os megbízhatósági tartomány gyakorlati alkalmazása egyszerű. Komoly előnye a rájuk alapozott statisztikai következtetéseknek, hogy a vizsgálati eredményeket a megbízhatóságukkal együtt kezelik és mutatják be.

Két vérnyomáscsökkentő gyógyszer (*A, B*) hatékonyságát szeretnék összehasonlítani. Az *A* gyógyszert 45 betegen próbálták ki, ahol a vérnyomáscsökkenés átlagosan 23 Hgmm volt (SD_A = 4,2 Hgmm). A 67 betegen vizsgált *B* gyógyszer esetében az átlagos terápiás hatás 27 Hgmm volt (SD_B = 8 Hgmm). Az átlagos terápiás hatás 95%-os megbízhatósági tartományait számították:

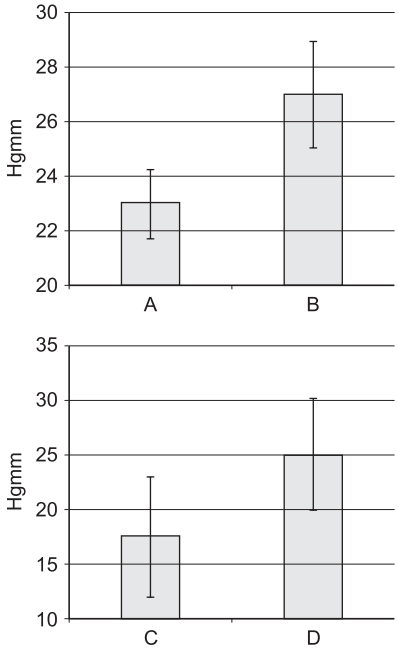
$$MT_{95\%,\bar{x},A} = \bar{x}_A \pm t_{[N_A-1;\alpha=0,05]} \frac{SD_A}{\sqrt{N_A}} = 23 \pm 2{,}015 \frac{4{,}2}{\sqrt{45}} = 21{,}74;24{,}26 ,$$

$$MT_{95\%,\bar{x},B} = \bar{x}_B \pm t_{[NB-1;\alpha=0,05]} \frac{SD_B}{\sqrt{NB}} = 27 \pm 1{,}997 \frac{8}{\sqrt{67}} = 25{,}05;28{,}95 .$$

Az *A* gyógyszerre valójában jellemző, átlagos terápiás hatás tehát nagyobb, mint 21,74 Hgmm, és kisebb, mint 24,26 Hgmm. *B* gyógyszer átlagos hatékonysága nagyobb, mint 25,05 Hgmm, és kisebb, mint 28,95 Hgmm. A két 95%-os megbízhatósági tartomány nem fed át egymással, azaz annak ellenére, hogy nem tudjuk pontosan egyik gyógyszer hatását sem meghatározni, azt a kérdést, hogy melyik jobb, mégis meg tudjuk válaszolni: *B* gyógyszer terápiás hatása jobb. Gyakorlati értékkel bíró megállapítást tudtunk tehát tenni, azaz kellően pontosak tudtunk lenni a vizsgálat során, Tulajdonképpen ez a jól kivitelezett vizsgálatok egyik legfontosabb ismérve: kellően pontosak, de nem fecsérlik az erőforrásokat olyan mértékű precizitás elérésére, amire már nincs is szükség az éppen vizsgált kérdés megválaszolásához *(7. ábra)*.

Egy másik helyzetben *C* és *D* gyógyszerek hatását hasonlítják össze. Itt az adatok az alábbiak: $N_C = 9$; x^-_C = 17,5 *Hgmm*; SD_C = 7,2 *Hgmm*; N_D = 16; x^-_D = 17,5 *Hgmm*; SD_D = 7,2 *Hgmm*. A megbízhatósági tartományok átfedték egymást:

$$MT_{95\%,\bar{x},C} = 11{,}97;23{,}03 ,$$



7. ábra. Négy vérnyomáscsökkentő gyógyszer (A, B, C, D) átlagos terápiás hatásának 95%-os megbízhatósági tartománya

$$MT_{95\%,\bar{x},D} = 19{,}88;30{,}12\ .$$

Az átfedés miatt lehet, hogy olyan valós értéke van a két gyógyszernek, ami egyforma hatáserősséget jelez; az is lehet, hogy a *D* gyógyszer a hatásosabb, hiszen sok olyan pontja van a *D* megbízhatósági tartománynak, amihez választható kisebb érték a *C* megbízhatósági tartományból; de az is lehet, hogy *C* a hatékonyabb, mert vannak olyan értéke is a *C* megbízhatósági tartománynak, amihez választható nála kisebb érték a *D* megbízhatósági tartományból. Összegezve vagy egyformák a gyógyszerek, vagy *C*, vagy *D* a hatékonyabb. (Ennyit árul el nekünk a vizsgálat! Ezt tudtuk korábban is, ezért nem volt érdemes belefogni a vizsgálatba!) A statisztikai következtetésünk az lesz, hogy nem találtunk különbséget a két gyógyszer hatáserősségében.

Hipotézistesztelés

Statisztikai elemzésre minden megfigyelésen és kísérleten alapuló vizsgálat esetében szükség van. Ennek oka, hogy vizsgálataink során a valóságra vonatkozó, általános érvénnyel bíró megállapítást szeretnénk tenni, ami megalapoz valamilyen gyakorlati beavatkozást. (Nem lehet eléggé hangsúlyozni a statisztikai eljárások részleteinek tanulmányozásakor sem, hogy a vizsgálatok nem öncélúak. Valós, gyakorlati jelentőséggel bíró kérdések megválaszolására szolgálnak. Az egész vizsgálati folyamat, benne a statisztikai értékelés is, végső soron csak az alapprobléma megoldásához való hozzájárulás szempontjából értékelhető.) A teljes valóság viszont egy-egy vizsgálat számára nem ragadható meg. Nem csak gyakorlati akadályok miatt (korlátozott a vizsgálatba vonható résztvevők, az elvégezhető kísérletek száma), hanem (meggyőző tudományfilozófiai tételek alapján) elvileg sem.

A statisztikai eljárások révén egy minta segítségével teszünk megállapítást a teljes populációra vonatkozóan. Mivel a minta csak közelítően reprezentálja a populációt, ez a megállapítás soha nem lehet 100%-ig biztos. A statisztikai eljárások menetét azért kell jól érteni, mert ismernünk kell az eljáráshoz kapcsolódó problémákat, azaz értenünk kell, hogy milyen szempontból informatívak, és milyen szempontból nem hasznosíthatók a statisztikai elemzések eredményei. A mindig meglevő korlátokkal együtt kell értelmezni az eredményeket.

Döntési küszöb

A teljesen biztos véleményalkotás hiánya miatt a statisztikai eljárások konzervatív szemléletűek. Egy megállapítást mindaddig nem tartunk megalapozottnak, amíg nagy valószínűséggel meg nem győznek minket az eredmények. Ugyanakkor, mindaddig azt gondoljuk, hogy a megállapításunk, elképzelésünk nem igaz, amíg erre akár csak kicsi esélyt is látunk. A helyzet analógiája az ártatlanság vélelme. Mindaddig ártatlannak minősítünk valakit, amíg erős bizonyíték nincs a bűnösségére. Ugyanakkor, ha gyenge kétség merül fel a bűnösséggel kapcsolatban, akkor már elmarad az ítélet. A bűnösséget kell meggyőzően bizonyítani, nem az ártatlanságot – azt eleve feltételezzük. Statisztikai elemzések esetén eleve feltételezzük, hogy az elképzelésünk nem igaz, és a vizsgálati

eredményektől várjuk, hogy meggyőző bizonyítékot adjanak arra, hogy mégis megalapozott a feltevésünk. Ha a meggyőző bizonyítékot nem tudjuk előállítani, akkor kénytelenek vagyunk annak megfelelően cselekedni, hogy nem volt igaz az elképzelésünk.

Vegyünk egy egészen konkrét példát erre! Azt gondoljuk, hogy az általunk kidolgozott eljárással javítani lehet a betegek gyógyulási esélyét. Vizsgálatot végzünk ennek igazolására, azaz adatokat gyűjtünk, amiket feldolgozunk. Ha a statisztikai eredmények szerint nagy biztonsággal állítható, hogy az alkalmazott eljárás javította a betegek állapotát, akkor kezdjük széleskörűen alkalmazni a módszer. Ha kis bizonytalanság marad a vizsgálat végén a gyógyítási hatékonyságot illetően, akkor viszont nem kezdeményezzük a korábbi gyakorlat módosítását, az új módszer alkalmazását – maradunk a régi gyakorlatnál, konzervatív módon.

A bírói gyakorlatban vannak nagyon könnyen megítélhető ügyek. Ha kétség merül fel a bűnösséget illetően, akkor az egyébként meglévő terhelő bizonyítékok ellenére sem ítélik el a vádlottat. Könnyű eset az is, ha az összes bizonyíték a bűnösség mellett szól. Az olyan átmeneti eset kezelhető nehezen, amikor vannak bőven bizonyítékok a bűnösség mellett, de az ártatlanság mellett is szólnak érvek. A statisztikai eredmények értékelésekor is hasonló a helyzet. A könnyen megítélhető extrém helyzetek közti zónában nehéz döntéseket hozni. Segítségünkre van az, hogy a statisztikai eljárás a valóságot reprezentáló adatok segítségével számszerűsíti annak a valószínűségét, hogy az elképzelésünkben szereplő kapcsolatok, összefüggések mentén jönnek létre a valóság tényei. Ezt a számszerűsített valószínűséget lehet egy kritikus valószínűségi küszöbértékhez hasonlítani: ha nagyobb az állítás igazságának valószínűsége, mint a küszöb, akkor igaznak tekintjük; ha kisebb, mint a küszöb, akkor nem tekintjük igaznak.

A szürke zónában a jól megválasztott döntési küszöb segít eligazodni. A küszöb is egy számérték, amit úgy kell megválasztani, hogy kellően nagy biztonságot jelentsen a nem megalapozott elképzelések elfogadása ellen, de azért ne állítson lehetetlenül szigorú feltételt a vizsgálatot végzők számára. Az előző azért fontos, mert a rossz elképzelés alapján bevezetett eljárásokat betegeken fogják alkalmazni, akiknek a gyógyulási esélyét ez jobb esetben nem javítja, rosszabb esetben kifejezetten rontja. Utóbbi pedig azért fontos, mert a gyógyító eljárásokat fejleszteni kell, akkor is, ha elvileg lehetetlen 100%-ig biztos tudományos bizonyítékok előállítása.

Hol található ennek a két feltételnek megfelelő döntési küszöb? Sajnos, nem lehet valamilyen elméletből levezetni ezt a küszöbvalószínűséget! De a gyakorlati tapasztalat az, hogy a biomedicinális vizsgálatok számára az 5%-os hibahatár megfelelő. Ez annyit jelent, hogy csak abban az esetben gondolunk egy elképzelést megalapozottnak, ha a statisztikai értékelés szerint az legalább 95%-os valószínűséggel igaz. Ennél kisebb va-

lószínűség esetén nem. Másképpen fogalmazva, amíg 5%-nál nagyobb a valószínűsége annak, hogy nem helyes a feltevésünk, akkor nem fogadjuk el azt. A döntési filozófia természetét jól szemlélteti, hogy ha a vizsgálatunk szerint 8% a valószínűsége annak, hogy nem megalapozott az elképzelésünk, és ennek megfelelően 92% annak a valószínűsége, hogy megalapozott, akkor ezt a vizsgálati eredményt nem tekintjük elég meggyőző adatnak az elképzelésünk mellett. Ezért nem is fogjunk a gyakorlatot az elképzeléseinknek megfelelően alakítani. Igen komoly meggyőző erőt kell tehát felmutatni ahhoz, hogy valaminek a megváltoztatásába belekezdjünk.

Elképzeléseinket hipotézisek formájában tudjuk megfogalmazni. Ha egy gerincsérült betegekkel foglalkozó szakember megoldatlan problémát lát, és van valamilyen elképzelése az ellátás hatékonyságának javításával kapcsolatban, akkor hipotézise (H_1) az lesz, hogy a hagyományos eljárásnál hatékonyabb az általa kidolgozott módszer. A statisztikai hipotézistesztelés általában nem közvetlenül ezt a H_1-et vizsgálja, hanem indirekt módon azt próbálja kizárni, hogy H_1 ellentettje az igaz. A H_1 ellentéte az a megállapítás, hogy a régi és az új módszer ugyanolyan hatékony. Ezt hívjuk nullhipotézisnek (H_0), és ennek valószínűségét adják meg a statisztikai tesztek *p*-érték formájában. (A *p*-érték tehát azt fejezi ki, hogy a vizsgált minta adatai alapján mekkora annak a valószínűsége, hogy a két módszer egyforma hatékonyságú.)

Ha a p-érték szerint 5%-nál kisebb a valószínűsége ($p < 0{,}05$) annak, hogy igaz H_0, akkor azt gondoljuk, hogy ez már elég kicsi valószínűséggel igaz ahhoz, hogy ne gondoljuk helyesnek. H_0 elutasítása viszont maga után vonja az ellentétes állítás elfogadását. Ha nem hisszük, hogy a két eljárás egyforma hatékonyságú, akkor azt gondoljuk, hogy eltérő a két hatékonyság: elfogadjuk, megalapozottnak gondoljuk H_1-et. Végső soron az lesz a véleményünk, hogy nem magyarázható pusztán a véletlennel, hogy a mintákban a két módszer hatékonysága eltérő volt. Röviden, szignifikáns teszteredményről és szignifikáns különbségről beszélünk.

Ha viszont 5%-nál nem kisebb H_0 valószínűsége ($p \geq 0{,}05$), akkor nem tartjuk elég meggyőzőnek a vizsgálatot ahhoz, hogy H_0-t elutasítsuk, és ennek következtében a H_1-et gondoljuk helyesnek. Ilyen esetben a rövid értékelés nem szignifikáns teszteredmény és nem szignifikáns különbség.

Első- és másodfajú hiba

A tapasztalat az, hogy összességében és általában a fenti döntési mechanizmussal később helyesnek bizonyuló gyakorlati lépéseket tudunk megalapozni. De sajnos nem

mindig! Az 5%-os hibahatár nem egy elméletileg megalapozott, a tévedések ellen biztos védelmet nyújtó eszköz.

Előfordulhat, hogy a vizsgálatunk $p = 0{,}08$ eredménnyel zárul, azaz 8%-nak találjuk a H_0 valószínűségét. Ezt a szigorú küszöb miatt még nem utasítjuk el – bár természetesen nem hisszük el azt sem, hogy igaz! Ebben a helyzetben előfordulhat, hogy valójában a H_1 volt a jó hipotézis (ami nem meglepő, hiszen mégiscsak 92%-nak találtuk annak a valószínűségét, hogy igaz!), és hibát követtünk el, amikor nem utasítottuk el H_0-t. A hiba azért jöhet létre, mert nem volt szerencsénk a minta összeállításakor (a sok lehetséges mintából pont egy szélsőséges összetételűt sikerült kiválasztanunk), vagy azért, mert kicsi volt a minta elemszáma (ennek részletes magyarázatát az egyes statisztikai teszteknél külön tárgyaljuk). Amikor nem találjuk megalapozottnak az egyébként igaz H_1-et, akkor másodfajú hibát követünk el: vizsgálatunk álnegatív eredményre vezet.

Olyan helyzet is adódhat, amikor a vizsgálatunk eredménye $p < 0{,}02$, tehát elég kicsinek tűnik a H_0 valószínűsége, és elvetjük (egyben elfogadjuk H_1-et), annak ellenére, hogy H_0 volt igaz. Ha az egyébként megalapozatlan H_1-et elfogadjuk, akkor elsőfajú hibát követünk el: vizsgálatunk álpozitív eredményre vezet. Ennek magyarázata szintén a szélsőséges mintaösszetétel.

Az első és másodfajú hibával kapcsolatban le kell szögezni, hogy ezek nem a vizsgálatot végző által elkövetett hibák, hanem a döntési folyamat természetéből következnek. Egy minden szakmai szabályt tiszteletben tartó, lelkiismeretesen elvégzett vizsgálat eredménye is lehet hibás statisztikai következtetés.

Teljes populáció vizsgálata helyett szelektált mintán végzett vizsgálat generálja a hibát! Ha nagyon sok azonos számú mintát választunk ki a teljes populációból, akkor ezek többsége hasonló (de nem azonos) összetételű lesz, mint maga a populáció, de pusztán a véletlennek köszönhetően, akad néhány szélsőséges összetételű minta is. A legjobban felkészült kutatócsoport is találkozhat extrém mintával, és hozhat rossz döntést a helyesen számított statisztikai eredményekre támaszkodva, az abszolút korrekt vizsgálata végén. Azt ugyanis egy vizsgálat elvégzése után nem tudjuk megmondani, hogy az éppen vizsgált minta tipikus volt vagy szélsőséges.

Ezért nem lehet önmagában egy vizsgálat eredménye alapján egy kutatási kérdést megnyugtatóan megválaszolni, még akkor sem, ha nagyon jól vezetett a vizsgálat. Ugyanazt a kérdést többször is meg kell vizsgálni (megerősítő jellegű vizsgálatokat kell végezni) ahhoz, hogy képet alkossunk az eredmények eloszlása alapján arról, hogy milyen is H_1 megalapozottsága.

Statisztikai tesztek

A p-érték generálására statisztikai tesztek szolgálnak. Ezek egy speciális vizsgálati helyzetre (pl. két átlagérték összehasonlítására, csoportokban megfigyelt gyakoriságok összehasonlítására, két folytonos változó közötti kapcsolat erősségének meghatározására stb.) kidolgozott statisztikai mérőszámot (teszt-statisztikát) állítanak elő. A mérőszám nagysága attól függ, hogy véletlenszerű vagy lényeges a vizsgált csoportok közti eltérés, vagy a vizsgált paraméterek közti kapcsolat. A mérőszámok egy ponton (kritikus érték) túl már a szignifikáns hatást jelzik. A kritikus értékeket táblázatokban találjuk meg (az internetről bármelyik könnyen elérhető), vagy számítógépes program segítségével számíthatjuk.

A mérőszámokat úgy dolgozták ki, hogy azok valamilyen ismert eloszlást (pl. t-, χ^2-, F-eloszlást) kövessenek. Az eloszlásfüggvények segítségével határozhatjuk meg azokat a pontokat, amin túl már véletlennel nem magyarázható, azaz 5%-nál kisebb valószínűséggel fordulnak elő a teszteredmények, ha a H_0 igaz. Ha számítógépes programot használunk (alapvetően ez a helyzet), akkor a számítások mennyisége nem korlátoz minket. Ezért nem csak a kritikus értékhez tudjuk hasonlítani a teszteredményt, hanem azt is számíthatjuk, hogyH_0 esetén a teszt milyen valószínűséggel vezet a számított értéknél extrémebb eredményre. Vagyis megadhatjuk, hogy mekkora a valószínűsége annak, hogy ha nincs szignifikáns különbség a vizsgált csoportok között, vagy nincs szignifikáns kapcsolat a vizsgált paraméterek között, akkor a teszt eredménye nagyobb lesz annál, mint amit a mintánk adatai alapján számítottunk. (Például, ha χ^2-próbát végzünk, akkor a minta jellemzői alapján egy χ^2-függvényt tudunk megadni. Ez a függvény azt mutatja meg, hogy ha nagyon sokszor elvégeznénk ugyanazt a vizsgálatot, mint ami a χ^2-értéket eredményezte, csak mindig más mintán, akkor a különböző teszteredmények milyen gyakorisággal fordulnának elő, ha H_0 igaz. Ennek a függvénynek az általunk számított χ^2-helyen vett értéke alapján kapjuk meg a keresett p-valószínűséget.) Ennek alapján természetesen pontosabb statisztikai következtetést tudunk levonni, mintha csak táblázatokat használnánk.

Csoportok közötti összehasonlítás

A kísérletek vagy megfigyelések során gyűjtött adatokat legalapvetőbben két csoportra lehet osztani a biostatisztikai feldolgozás menete szempontjából. Egyfelől kvantitatív (valamilyen folytonos skálán mért) adatokkal dolgozhatunk (vérnyomás, szérum HDL-szint, FEV1 stb.). Másfelől (kettő vagy több) kategóriába rendezett, kvalitatív adataink lehetnek (dohányzási szokások gyakorisága, a gyógyszerrel szemben allergiások részaránya, 5 éves túlélők százaléka stb.) Folytonos változók elemzésekor általában abból a körülményből indulunk ki, hogy ezek az adatok normális eloszlást mutatnak. Ez az eloszlás átlaggal és szórással leírható, mely paraméterek vizsgálatával oldjuk meg a csoportok közti különbségek (véletlennel magyarázható, vagy véletlennel nem magyarázható) természetének értékelését. Kvalitatív adatok esetében nem az eloszlás paramétereinek elemzése, hanem az egyes kategóriákba sorolt esetszámok, illetve az esetszámokhoz tartozó gyakorisági adat közvetlen értékelése a biostatisztikai eljárás alapja. Ilyen változótípusnál maguk a gyakorisági adatok összegzik a vizsgálati eredmények eloszlását – amint a folytonos változók esetében ezt az eloszlás-paraméterek teszik.

Két csoport kvantitatív adatainak összehasonlítása

Az egyik legáltalánosabb biostatisztikai feladat az, amikor két, valamilyen szempontból eltérő (különböző kezelésben részesülő, a betegség különböző stádiumában levő, különböző genetikai érzékenységű) csoport közti biológiai, klinikai különbséget szeretnénk értékelni. Ha a vizsgált biológiai/klinikai jelleg folytonos változóként mérhető (pl. vérnyomás, szérum-kreatininszint, carotis-falvastagság), akkor ezt legegyszerűbben a csoportok (*a, b*) átlagainak összehasonlítása révén tehetjük meg:

$$k = \bar{x}_a - \bar{x}_b \,.$$

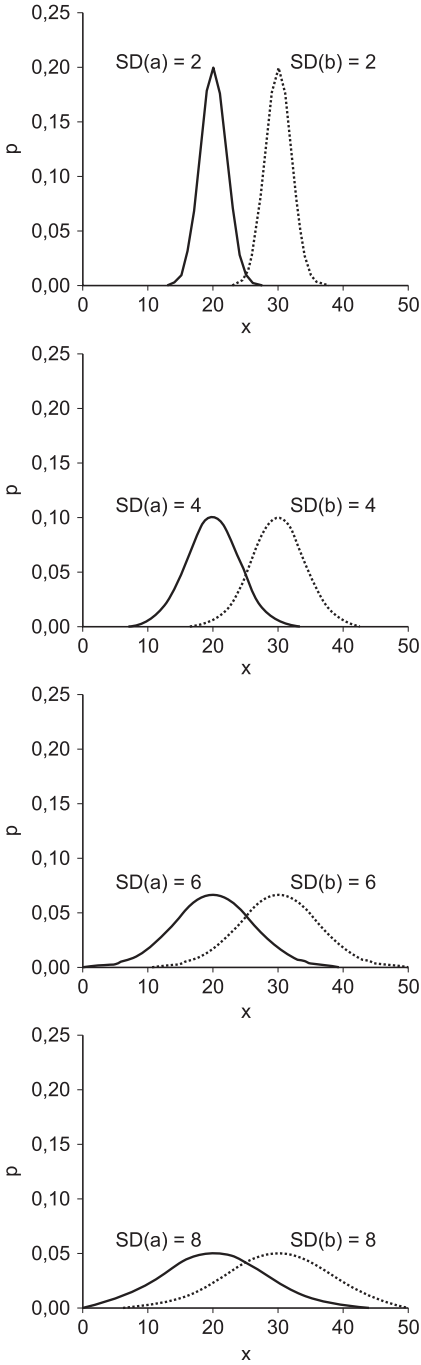
A vizsgálat természetesen két mintán zajlik. Emiatt az *a* és *b* kezelésben részesültekre, expozíciónak kitettekre stb. valóban jellemző átlagot csak több-kevesebb pontatlansággal tükrözik az átlagok. Emiatt az átlagok közti különbség is csak közelíti az *a*-ra és *b*-re jellemző valós átlagok közti különbséget. Márpedig természetesen mi mindig a

valós különbségre vagyunk kíváncsiak, hiszen a vizsgálat alapkérdésének megválaszolásával valamilyen gyakorlati beavatkozást szeretnénk megalapozni.

Két mintát vizsgálva szinte soha nem kapunk számszerűen egyenlő átlagot. Önmagában ez nem győz meg minket arról, hogy a különbség általában is megvan. (Valamilyen különbséget mindig látunk!) Szeretnénk tudni, hogy ha sokszor megismételnénk az adatgyűjtést, akkor is hasonló irányú és mértékű lenne-e az átlagok közti különbség.

A sokszor megismételt kísérlet során annál szélesebb tartományon belül variálódnak az átlagértékek, minél nagyobb szóródásúak az eredeti vizsgálati eredmények. Ha nagy a csoportokra jellemző variabilitás, akkor még viszonylag jelentős különbség esetén sem gondoljuk azt, hogy ismételt vizsgálat esetén is biztosan hasonló előjelű és mértékű eltérést látnánk. Ha kicsi a variabilitás, akkor már szerény eltérést is a csoportok közt meglévő valódi különbség jeleként tudunk értelmezni. Vagyis a két csoport átlaga közti különbség csak az átlagok különbségének variabilitása függvényében értékelhető *(8. ábra)*.

Ugyanakkora átlagkülönbség kis szórású adatoknál kevés kétséget hagy a csoportok közt meglévő lényeges eltérést illetően. A szóródás növekedésével egyre bizonytalanabbá válik a benyomásunk. Amikor a 8. ábrán szereplő hisz-



8. ábra. Az átlagok változatlan különbsége (k = 10) eltérő szórás mellett

togramok már jelentős mértékben átfednek, kevés kétségünk lesz afelől, hogy nincs lényeges eltérés a vizsgált csoportok közt.

Az átlagok *k* különbségét a szóródás mértékéhez viszonyító mérőszám származtatásához az *a* és a *b* csoport vizsgálati eredményeinek varianciáit (V_a, V_b) használjuk:

$$V_a = \frac{\Sigma(x_a - \bar{x}_a)^2}{(n_a - 1)} \qquad V_b = \frac{\Sigma(x_b - \bar{x}_b)^2}{(n_b - 1)} \quad .$$

Mivel a *k* két csoport adatainak együttes felhasználásával előállított statisztikai mérőszám, variabilitását a két csoport varianciáinak mintanagysággal (pontosabban a minták szabadsági fokával) súlyozott közös varianciájával kell számítanunk (V_k):

$$V_k = \frac{(n_a - 1)V_a + (n_b - 1)V_b}{(n_a - 1 + n_b - 1)},$$

$$V_k = \frac{(n_a - 1)\dfrac{\sum(x_a - \bar{x}_a)^2}{(n_a - 1)} + (n_b - 1)\dfrac{\sum(x_b - \bar{x}_b)^2}{(n_b - 1)}}{(n_a - 1 + n_b - 1)},$$

$$V_k = \frac{\sum(x_a - \bar{x}_a)^2 + \sum(x_b - \bar{x}_b)^2}{(n_a + n_b - 2)}.$$

A közös variancia mutatja meg, hogy sokszor ismételt vizsgálat *k* eredményei milyen mértékben szóródnának. (Az ismételt vizsgálatok alkalmával *k* szóródását nagyobb mértékben határozná meg az a csoport, aminek nagyobb a létszáma. Ez tükröződik a szabadsági fok szerinti súlyozásban.)

A varianciák és a minták elemszáma határozza meg a csoportátlagok standard hibáját:

$$SE_a = \sqrt{\frac{V_a}{n_a}} \quad \text{.....} \quad SE_b = \sqrt{\frac{V_b}{n_b}} \quad .$$

Ehhez hasonlóan a *k* különbség standard hibáját a közös variancia és a két csoport elemszáma együttesen határozza meg:

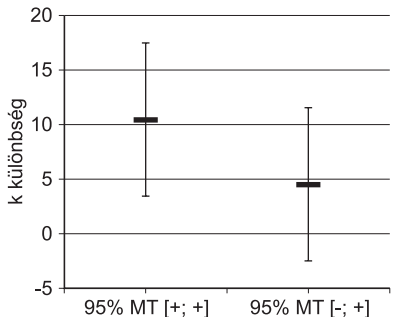
$$SE_k = \sqrt{\frac{V_k}{n_a} + \frac{V_k}{n_b}} = \sqrt{V_k \left(\frac{1}{n_a} + \frac{1}{n_b}\right)} = SD_k \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}.$$

A k különbség 95%-os megbízhatósági tartományát ezek alapján megadhatjuk (a t-érték szabadsági foka a számítások során a vizsgálati eredmények mellett felhasznált két csoportátlag miatt $n_a + n_b - 2$):

$$MT_{95\%,k} = k \pm t_{\alpha;df} SE_k .$$

Ha ez a tartomány tartalmazza a 0-át (alsó határa negatív szám, a felső határa pozitív), akkor a vizsgálatunk alapján nem tarthatjuk kizártnak, hogy: (1) a valós eltérés a két csoport átlaga közt éppen ellenkező előjelű, mint amit a saját vizsgálatunkban láttunk, (2) valójában nincs semmi eltérés a két csoport közt (erre utal, hogy a vizsgálat alapján a $k = 0$ a valószínű eredmények közt van). Ha a 95%-os megbízhatósági tartomány nem tartalmazza a 0-át, akkor a különbség nagyságáról ugyan nem lesz pontos képünk (arról csak annyit tudunk, hogy az nagy valószínűséggel a 95%-os megbízhatósági tartományon belül van valahol), de azt azért meg tudjuk állapítani, hogy valóban van eltérés a két csoport közt (hiszen minden valószínű vizsgálati eredmény szerint hasonló előjelű az eltérés a csoportok közt, mint amit a saját vizsgálatunkban is találtunk, *9. ábra)*.

A különbséget közvetlenül is viszonyíthatjuk saját varianciájához. Ha ez a statisztikai mérőszám kellően nagy, akkor az meggyőz minket arról, hogy az általunk talált k különbség nem a véletlennek tulajdonítható, hanem a csoportok közt meglevő lényeges



9. ábra. Az első vizsgálati eredmény szerint a 95% MT végig pozitív tartományban van. Minden valószínű ***k*** érték arra utal, hogy pozitív előjelű az eltérés a csoportok átlagai közt: biztosak lehetünk abban, hogy a saját vizsgálatunkban kapott pozitív különbség nem a véletlennek köszönhető. A második vizsgálatban a 95% MT a két csoport közti különbség hiányára utaló 0-t és a két csoport közti negatív előjelű különbséget is valószínűsítő negatív értékeket tartalmaz: ebben az esetben azt gondoljuk, hogy pusztán a véletlennek köszönhető, hogy saját vizsgálatunkban pozitív előjelű volt a különbség

különbségnek. Ha ez a hányados nem elég nagy, akkor azt gondoljuk, hogy ismételt vizsgálatokban éppen ellenkező előjelű eltérést is láthatnánk. A két csoportot nem tekinthetjük eltérőnek.

Az elég nagy különbség értelmezését a *k* standardizálásával tehetjük egyszerűvé, amikor a standard hiba révén a különbséget egységnyi variabilitás-mértékegységben fejezzük ki! A standardizált különbség, azaz a t-érték:

$$t = \frac{k}{SE_k}$$

$$t = \frac{\bar{x}_a - \bar{x}_b}{SD_k \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}}.$$

Ha a standardizált különbség nagyobb a kritikus $t_{\alpha;df}$ értéknél, akkor meggyőzően nagynak tekintjük az eltérést. Ha ennél kisebb a statisztikai mérőszám, akkor még nem tekintjük szignifikánsnak a különbséget. Az adott döntési küszöbnek és szabadsági foknak megfelelő t-értéket táblázatból kereshetjük ki, vagy a számítógép segítségével számíthatjuk ki. A t-érték kiszámításán alapuló hipotézistesztelés a t-próba (vagy Student-próba). A fentebb ismertetett formája a kétmintás t-próba, ami arra utal, hogy két külön minta vizsgálata során nyert átlagokat hasonlítottunk össze a segítségével.

Az asztmás gyerekek képzésének segítségével elérhető, hogy a gyerekek jobban értsék állapotukat, és saját maguk gondozásában részt vállaljanak. Az oktatási protokollok hatékonysága azonban különböző, ezért vállalkoztak egy új képzési protokoll hatékonyságának tesztelésére. 10 gyermeket vettek fel a képzési programba (a), 12, oktatásban nem részesülő gyermek képezte a kontrollcsoportot (b). A gyermekek súlyos rosszulléteinek számát a szülők naplói segítségével állapították meg. A *2. táblázat* mutatja a 12 hónap alatti rosszullétek számát.

2. táblázat

Oktatás nem volt (b)	Oktatásban részesültek (a)
3	2
4	2
6	4
8	8
5	6
1	7
8	5
7	1
7	6
8	3
6	
3	

Az oktatásban részesültek átlagos rosszullétszáma 4,4 volt (SD = 2,37; variancia = 5,60). A kontrollcsoportban pedig 5,36 (SD = 2,32; variancia = 5,36). A képzett csoportban

1,1-gyel kisebb volt az átlag. A közös variancia 5,47, a különbség standard hibája 1,001 volt.

$$V_k = \frac{(n_a - 1)V_a + (n_b - 1)V_b}{(n_a - 1 + n_b - 1)} = \frac{(10-1)\times 5{,}60 + (12-1)\times 5{,}36}{(10-1+12-1)} = 5{,}47$$

$$SE_k = \sqrt{\frac{V_k}{n_a} + \frac{V_k}{n_b}} = \sqrt{\frac{5{,}47}{10} + \frac{5{,}47}{12}} = 1{,}001$$

$$t = \frac{k}{SE_k} = \frac{1{,}1}{1{,}001} = 1{,}098$$

A t = 1,098 kisebb volt, mint a kritikus érték ($t_{[0,05;20]}$ = 1,098), azaz az eredményt nem tudjuk az oktatási protokoll hatékonyságának bizonyítékaként felhasználni, még akkor sem, ha látjuk, hogy kevesebb volt a képzésben részesülő gyerekek közt a komplikációk száma.

Gyakran előfordul, hogy egy kezelés hatékonyságát úgy tudjuk vizsgálni, hogy a betegek kezelés előtti és utáni állapotát mérjük fel. Elvileg lehetőség volna arra, hogy a kezelés előtti és utáni adatokat a kétmintás t-próba segítségével hasonlítsuk egymáshoz, és ezáltal értékeljük a kezelés hatékonyságát. Az önkontrollos vizsgálatban azonban ennél hatékonyabb értékelésre is van lehetőség. A hatékonyságnövekedés (kisebb hatások kimutatására való képesség) alapja, hogy párba rendezett adatok, adatpárok alapján számolt eltéréseket ($\bar{d}$) értékelünk:

$$\bar{d} = \frac{\sum d}{N}.$$

Egyszerűen, az általános képlet alapján adható meg a varianciája, standard hibája, illetve maga a t-érték, a standardizált különbségátlag:

$$V_{\bar{d}} = \frac{\sum (d - \bar{d})^2}{N-1},$$

$$SE_{\bar{d}} = \sqrt{\frac{V_{\bar{d}}}{N}},$$

$$t = \frac{\bar{d} - 0}{SE_{\bar{d}}} = \frac{\bar{d}}{SE_{\bar{d}}}.$$

A t teszt-statisztikai értékelése teljesen azonos módon zajlik, mint a két független csoport összehasonlításán alapuló elemzésnél. Az összes eltérés az, hogy itt a szabadsági fok N-1 (hiszen itt csak 1 átlagértéket használtunk a t-érték kiszámításához).

Ugyanezt a számítási módot követjük, ha a kezelés előtti és utáni állapotot leíró eredmények nem ugyanabból a személyből, kísérleti állatból, in vitro rendszerből származnak, de a párba rendezett vizsgálati alanyok kellően hasonlóak ahhoz, hogy a kezelt vs. nem kezelt különbséget a párok közt meglevő egyéb eltérések ne befolyásolják. Ilyen vizsgálati elrendezésben is a párok közti eltérés alapján vonunk le statisztikai következtetést.

A t-próbák kiindulási lépése a varianciák számítása, amit a normális eloszlást mutató változókra vonatkozó képlet segítségével tudunk számítani. A t-próba alkalmazásának ezért alapfeltétele, hogy az adatok normális eloszlásúak legyenek. (Ezt ellenőrizni is kell az alkalmazás előtt; pl. Kolmogorov–Szmirnov-próbával.) Ha két független minták van, abban az esetben további feltétel, hogy a két csoportban gyűjtött adatok varianciája egyenlő legyen, azaz homoszkedasztikus legyen a vizsgálati helyzet. (Ezt is ellenőrizni kell az alkalmazás előtt; pl. F-próbával.) Ha két variancia nem egyezik, akkor a közös variancia számolásakor nem használható a szabadsági fokokkal súlyozott varianciaátlag. Helyette a csoportonként külön-külön számított varianciák összegét kell számítanunk. Vagyis az alapstatisztika, a t-érték számítása némileg változik:

$$t = \frac{k}{SE_k} = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{V_a}{n_a} + \frac{V_b}{n_b}}}.$$

Az így kapott t-értékhez kapcsolódó szabadsági fok:

$$df = \frac{\left(\frac{V_a}{n_a} + \frac{V_b}{n_b}\right)^2}{\frac{\left(\frac{V_a}{n_a}\right)^2}{n_a - 1} + \frac{\left(\frac{V_b}{n_b}\right)^2}{n_b - 1}} - 2.$$

Több csoport kvantitatív adatainak összehasonlítása

Ha több csoport eredményeinek elemzése révén szeretnénk valamilyen következtetést levonni, akkor tapasztaljuk, hogy a két csoport összehasonlítására kidolgozott módszerek tulajdonképpen alkalmazhatók. Ha például 5 csoportot vizsgálunk, akkor páronként elemezhetjük a csoportok közti különbségeket: 10 t-próba árán tudunk következtetést levonni. [k-csoport esetén $k(k-1)/2$ párelemzést kell végrehajtani]. Azonban elvégezve a teszteket nehézségbe ütközünk a szignifikáns eredmények értelmezésekor. Nem lesz

egyértelmű, hogy a pozitív eredmény valóban lényeges csoportok közti különbségnek tulajdonítható, vagy csak a több statisztikai teszt összeadódó hibájaként létrejövő, álpozitív eredmény.

A statisztikai jellegű következtetéseink 5%-os elsőfajú hibával terheltek. Ahány statisztikai következtetést vonunk le egy adott vizsgálat során, annyiszor felmerül ennek az 5%-os hibának a lehetősége. 10 vizsgálat esetén már 10 × 5% = 50% az elsőfajú hiba valószínűsége (50% annak a valószínűsége, hogy akkor is szignifikáns különbséget látunk a tesztek végeredményeként, ha egyébként semmilyen hatást nem fejt ki a vizsgált befolyásoló tényező). Emiatt, lényeges eltérésre utaló teszteredmény esetén, két magyarázat is felmerül: (1) vagy tényleges eltérést mutatott ki a teszt, (2) vagy a sok tesztelés miatt megnövekedett hibalehetőség miatt látunk statisztikailag szignifikáns eredményt. Előző esetben a gyakorlati következtetésünk az, hogy a két csoport közti eltérésre való tekintettel beavatkozásokat kezdeményezhetünk. Második esetben az, hogy a különbség csak látszólagos és gyakorlati következtetést nem alapozhatunk a vizsgálatunk eredményére. A kérdést ebben a formában tehát nem tudtuk megoldani ezzel a módszerrel.

Megoldási lehetőséget jelent, ha a tesztelt párok számához igazodva szigorúbb döntési küszöböket alkalmazunk. Ha 5 csoportunk van, és 10 pár közti eltérést kell tesztelünk, akkor 0,05/10 = 0,005 döntési küszöböt alkalmazva minden egyes t-próbánál, összességében a vizsgálat elsőfajú hibája 5% lesz. Emiatt az álpozitív teszteredmények megnövekedett valószínűsége nem terheli a következtetéseinket. Ennek az ára, hogy jelentősen csökkent az egyes párok közti különbség kimutatására jutó statisztikai érzékenység: kisebb az esélyünk arra, hogy a valóban meglevő különbséget észlelni tudjuk. A többszörös hipotézistesztelés önálló fejezete a biosztatisztikának. Ennél az egyszerű döntési küszöb korrekciónál (Bonferoni-korrekció) hatékonyabb módszerek is rendelkezésre állnak, de ezek ismertetése meghaladja az alapozó tanfolyamok keretét.

A párok elemzése helyett érdemes visszanyúlnunk ahhoz a ponthoz, hogy valójában miért is kezdeményeztük a vizsgálatot? Voltaképpen nem az az alapkérdésünk, hogy egyes kezelési csoportok közt vannak-e eltérések, hanem az, hogy az alkalmazott eljárás hatékony-e. Ha a hatékonyságra már van bizonyítékunk, akkor lehet feltenni következő kérdésként, hogy melyik dózisban hatékony, és melyikben nem. Az első kérdés megválaszolásához az előző fejezetben tárgyalt módszerekhez képest új statisztikai eljárást kell kialakítanunk, melynek alapgondolata, hogy azt kell megértenünk és elemeznünk, hogy milyen tényezők hatására alakul ki az egyes vizsgálati résztvevők mérési eredménye.

Teljesen természetesnek vesszük, hogy még a teljesen azonos dózisokkal kezelt betegek esetében sem ugyanazt a terápiás hatást látjuk. Természetesnek vesszük, mert

tudjuk, hogy a biológiai rendszerek rendkívül összetettek, sok szabályozottan működő rendszer együttes hatásának az eredője minden élettani paraméter. Mivel a vizsgálati alanyok különbözőek, a válaszaik is különbözőek lesznek. Ez az egyéni variabilitás jellemző minden biológiai rendszerre.

Ha egy betegcsoportot kezelünk, akkor a hatékony kezelés hatására elmozdul a betegekben a terápiás hatást kifejező paraméter: a csoporton belül szóródó mérési eredmények átlaga kedvező irányba fog eltolódni. Ezért, ha a beavatkozás hatását szeretnénk igazolni, akkor az átlagok elmozdulására kell bizonyítékot találnunk.

Minél nagyobb az egyéni válaszok variabilitása, annál szélesebb a mérési eredmények szóródása, és annál nehezebb az átlag elmozdulását bizonyítani. Kicsi variabilitás esetén a viszonylag kicsi átlagérték-eltolódás is meggyőz minket arról, hogy a beavatkozás hatékony volt, megváltoztatta a mért paramétert; nagy variabilitás esetén már csak nagyobb átlagérték-változás esetén lesz az a benyomásunk, hogy volt hatása a kezelésnek. A változékonyság függvényében tudjuk tehát értékelni az átlagok eltérését.

Ha egy gyógyszer különböző dózisaival (D_a, D_b, D_c, D_d, D_a) kezelünk azonos klinikai állapotú betegeket, akkor minden csoporton belül bizonyos variabilitást mutatnak a kezelési eredmények. Ezt a kezelési csoporton belüli variabilitást a csoporton belüli átlagtól ($\overline{x}_a$, $\overline{x}_b$, $\overline{x}_c$, $\overline{x}_d$) való eltérés négyzetének összegével tudjuk legegyszerűbben leírni (négyzetes eltérések összege, sum of square; SS):

$$SS_a = \sum (x_a - \overline{x}_a)^2 \qquad SS_b = \sum (x_b - \overline{x}_b)^2 \qquad SS_c = \sum (x_c - \overline{x}_c)^2 \qquad SS_d = \sum (x_d - \overline{x}_d)^2 .$$

Ha a kezelés nem hatékony, akkor a csoportokon belüli átlagok nem térnek el egymástól. Ilyen esetben az összes kezelt betegre számított átlag ($\overline{x}_t$) fejezi ki a vizsgált klinikai paraméter átlagértékét:

$$\overline{x}_a = \overline{x}_b = \overline{x}_c = \overline{x}_d = \overline{x}_t.$$

Ebben az esetben a csoportokon belül számított négyzetes eltérések összege is számítható úgy, hogy a csoporton belüli átlagot a teljes betegcsoport átlagával helyettesítjük:

$$SS_a = \sum (x_a - \overline{x}_t)^2 \qquad SS_b = \sum (x_b - \overline{x}_t)^2 \qquad SS_c = \sum (x_c - \overline{x}_t)^2 \qquad SS_d = \sum (x_d - \overline{x}_t)^2 .$$

Mivel minden egyes beteg esetében meghatároztuk az átlagtól való eltérés négyzetét, és ugyanazt az átlagot használtuk mindig, a négy csoportban mért négyzetes eltéré-

sek összege ugyanazt a számot adja, mintha az egyes betegek eredményeit a teljes vizsgálati csoport átlagához viszonyítva számítottuk volna a négyzetes eltérések összegét:

$$SS_t = \sum (x_{a,b,c,d} - \bar{x}_t)^2 = SS_a + SS_b + SS_c + SS_d .$$

Ha viszont a kezelés hatékony, akkor az átlagértékek nem egyenlők (pontosabban vannak köztük különbözőek, ami miatt még lehetnek köztük egyenlők is). Ilyen esetben a csoportokon belül a négyzetes eltérésösszegek számításakor nem használhatjuk a teljes betegcsoportra jellemző átlagot, és a csoportokon belül számított négyzetes eltérések összege nem is lesz egyenlő az összes vizsgálatba vont beteg eredményének és a teljes csoport átlagának felhasználásával számított teljes variabilitással:

$$SS_t \neq SS_a + SS_b + SS_c + SS_d .$$

Az egyes betegek vizsgálati eredményei részben a kezeléstől függenek, részben mindazoktól az egyéni jellemzőktől, amelyek befolyásolják a vizsgált klinikai paramétert, de amelyet a vizsgálatunkban nem vettünk figyelembe. A vizsgálati eredmények teljes variabilitása ebből a két forrásból származik. A kezelési csoportokon belül megfigyelhető variabilitás független a kezeléstől, hiszen a csoporton belül minden beteg azonos ellátásban részesült. A csoportok közti átlagok különbségei fejezik ki a kezelés okozta variabilitást. A teljes variabilitás tehát felbontható a csoportokon belüli (S_i) és a csoportok közti (S_e) komponensre:

$$SS_t = SS_i + SS_e .$$

A csoportokon belüli és a teljes vizsgált betegcsoportra vonatkozó négyzetes eltérésösszeget számíthatjuk a fentebb már említett képletek segítségével. A csoportok közti variabilitást pedig egyszerűen ezek különbségeként kapjuk meg:

$$SS_e = SS_t - SS_i .$$

A variabilitás elemzését azért kezdtük, mert arra gondoltunk, hogy ha a csoportok közti átlagok közti különbség nagy a vizsgált csoportokon belüli adat variabilitáshoz képest, akkor az meggyőz minket arról, hogy a beavatkozás hatékony. Ha az átlagok közti eltérés kicsi az adat variabilitásához viszonyítva, akkor ez amellett szóló érv, hogy a kezelés nem volt hatékony, és a látott eltérés pusztán véletlennel magyarázható.

Felhasználva a variabilitással kapcsolatos gondolatmenet eredményeit, azt mondhatjuk, hogy akkor győz meg minket a vizsgálat a beavatkozás hatékonyságáról, ha a csoportok közti variabilitás nagyobb, mint a csoportokon belüli. Ha a csoportok közti variabilitás kicsi a csoporton belüliekhez viszonyítva, akkor ez amellett szól, hogy nem befolyásolja a kezelés a klinikai eredményt.

Valamilyen módon tehát a különböző variabilitási adatokat egymáshoz kell hasonlítanunk. Lehetne egyszerűen a négyzetes eltérések összegének hányadosait használni. A megoldás kézenfekvő, de nem jó! A négyzetes eltérések összege ugyanis nem csak attól függ, hogy milyen mértékű a vizsgálati eredmények szóródása, hanem attól is, hogy hány beteg volt egy-egy csoportban. (Nagy csoport kicsi variabilitású eredményei ugyanakkora négyzetes eltérés összeget adhatnak, mint a kis csoport nagy variabilitású adatai. Hasonlóan elmondható ez a csoportok közti variabilitásról is: sok csoport közti kis eltérés ugyanakkora négyzetes eltérés összeget adhat, mint a kevés csoport közti jelentős eltérés.) Valamilyen módon tehát figyelembe kell venni a csoportnagyságot, pontosabban azt a számot, ami megmutatja, hogy hány független adat felhasználásával kaptuk a variabilitási mérőszámokat, azaz a szabadsági fokokat.

A csoportokon belüli variabilitás számítása esetén egyszerű a helyzet. Itt az átlagértékhez viszonyítva számítottuk a négyzetes eltérések összegét. Emiatt az utolsó beteg esetén az előző betegek adataiból és az átlagból már pontosan tudjuk az eredményt. Ennek az egy betegnek az adata már nem ad új információt az elemzéshez (ha az ő adata elveszne, de az átlag megmaradna, akkor ennek az egy betegnek a mérési eredménye még pontosan kiszámítható lenne). A szabadsági fok tehát a példánkban szereplő négy csoportra:

$$df_a = n_a - 1 \qquad df_b = n_b - 1 \qquad df_c = n_c - 1 \qquad df_d = n_d - 1 .$$

A csoporton belüli teljes variabilitás számításakor (S_i) ezeket a csoportonkénti szabadsági fokokat kell összegezni. A csoporton belüli variabilitás szabadsági foka:

$$df_i = n_a - 1 + n_b - 1 + n_c - 1 + n_d - 1 .$$

Általában, ha a vizsgált csoportok száma k, akkor k alkalommal kell kivonni 1-et az előző egyenletben, és az egyes csoportelemszámok összességében a teljes minta nagyságát (n_t) adják. Ezért az általános helyzetre vonatkozó szabadsági fok:

$$df_i = n_t - k .$$

A csoporton belüli szabadsági fokhoz viszonyított négyzetes eltérések összege (ami egyszerűen variancia, de hívják még a négyzetes eltérések átlagának is, mean squares, MS):

$$MS_i = \frac{SS_a + SS_b + SS_c + SS_d}{n_t - k}.$$

A teljes variabilitásra a szabadsági fok ugyanazzal a gondolatmenettel vezethető le, mint a csoporton belüli variabilitás esetén. Itt sem ad már egy beteg adata információt a mutató kiszámításához. Ezért a szabadsági fok és a variancia ebben az esetben:

$$df_t = n_t - 1$$

$$MS_t = \frac{SS_t}{n_t - 1}.$$

A csoportok közti variabilitás számításakor k csoport átlagárértékeinek főátlagtól való eltérését értékeljük. A k átlagból is egyet veszíthetnénk el büntetlenül, ha a teljes vizsgálat átlagos eredménye megmaradna. Ezért a csoportok közti variabilitás szabadsági foka és a variancia:

$$df_e = k - 1$$

$$MS_e = \frac{SS_e}{k - 1}.$$

A varianciák értéke már a vizsgált elemek számától függetlenül kifejezi, hogy mekkora az elemi adatok szóródása. Ezek segítségével közvetlenül viszonyítható egymáshoz a csoportok közti és a csoportokon belüli variabilitás, egyszerűen a varianciák hányadosát kell számítanunk. Ha ez a hányados kellően nagy, akkor az a kezelés hatékonyságának a bizonyítéka. Ha ez a hányados kicsi, akkor ez az ellen szól, hogy a kezelés hatékony volt.

Már csak azt a kérdést kell megválaszolnunk, hogy mikor tekinthető egy varianciahányados kellően nagynak? Ennek a megválaszolása olyan matematikai ismereteket igényel, amelyeket eddig próbáltuk megkerülni, amikor a biostatisztikai problémák megoldási lehetőségeiről gondolkodtunk. Itt is elkerüljük a matematikai válasz ismertetését, és csak utalni tudunk a valószínűség-számítással foglalkozó tankönyvekre. Ugyanakkor felhasználjuk az ilyen könyvek végén található táblázatot, ami közli a kritikus határvonalnak számító értéket a szabadsági fokok és a döntési küszöb függvényében (http://www.socr.ucla.edu/Applets.dir/F_Table.html). A kritikus értéknél nagyobb varianciahányadost értelmezzük kellően nagynak.

A kritikus F-értéket természetesen számítógép segítségével is számíthatjuk. Igazából nem is nagyon érdemes a táblázatokat forgatni, amikor egy konkrét feladatot oldunk meg. A varianciaelemzés gondolatmenetéhez viszont a táblázatok hozzátartoznak, ezért ezek használatát érteni kell.

A fenti varianciaelemzés (analysis of variance, ANOVA) eredményeit szokás táblázatos formában összefoglalni *(3. táblázat)*. Ez a táblázat kiegészül még az adott helyzetre vonatkozó kritikus F-értékkel (ami azt mutatja meg, hogy a szóban forgó, n_t elemszámú vizsgálatnál k csoportba rendezett vizsgálati alany esetén legalább mekkorának kell lennie a csoportok közti varianciának a csoporton belülihez viszonyítva, ahhoz, hogy legalább 95%-os bizonyossággal állíthassuk, hogy hatásos a kezelés). Ha szigorúbb döntési küszöböt követel meg a vizsgálati helyzet, akkor magasabb lesz a kritikus F-érték (a csoportok közti különbségnek nagyobbnak kell lennie, mintha csak 95%-os bizonyossággal szerettünk volna valamit megállapítani.)

Számítógépek segítségével természetesen nem csak kritikus F-értéket tudunk számítani, hanem azt is meg tudjuk határozni, hogy az elemzés során számított varianciahányados (F-érték) milyen döntési küszöb esetén számít kritikus értéknek. (Milyen kritikus értékre számított F-táblázatból tudnánk kiolvasni a vizsgálatunkhoz tartozó F-értéket!) Ebben az esetben már egy p-érték számításáról, és F-értékek összehasonlítása helyett F-próbáról beszélhetünk. (Ahol két variancia hányadosát a nekik megfelelő szabadsági fokok függvényében összevetve megállapítjuk, hogy a számlálóban lévő variancia szignifikánsan nagyobb-e, mint a nevezőben lévő.)

Összességében tehát elmondhatjuk, hogy ha kettőnél több csoportot alakítottunk ki a vizsgálatunk során, és arra vagyunk kíváncsiak, hogy a kezelés (amit különböző dózisokban alkalmaztunk) hatékony-e, akkor a varianciaelemzéssel olyan módon tudjuk

3. táblázat. Az ANOVA-tábla általános szerkezete

Variabilitás forrása	Variabilitás (négyzetes eltérések összege, SS)	Szabadsági fokok	Variancia (átlagos négyzetes eltérés)	Varianciahányados (F)
Csoportok közti variabilitás (e)	$SS_e = SS_t - SS_i$	$df_e = k - 1$	$MS_e = \frac{SS_e}{k-1}$	$F = \frac{MS_e}{MS_i}$
Csoporton belüli variabilitás (i)	$SS_i = SS_a + SS_b + ... + SS_k$	$df_i = n_t - k$	$MS_i = \frac{SS_a + SS_b + ... + SS_k}{n_t - k}$	
Teljes variabilitás (t)	$SS_t = \sum (x_{a,b,...,n} - \bar{x}_t)^2$	$df_t = n_t - 1$		

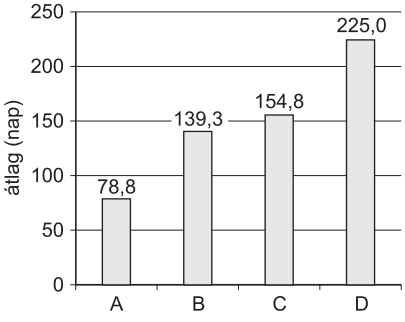
a kérdést megválaszolni, hogy fel sem merül a többszörösen alkalmazott t-próbával kapcsolatban említett döntési hiba kérdése. Ilyen helyzetekben tehát egyértelműen ez a preferálandó biostatisztikai módszer.

Ha a varianciaelemzés szerint nincs hatása a kezelésnek, akkor az adott kezelésről valóban azt gondolhatjuk, hogy nem képes befolyásolni a betegek állapotát. De ha a varianciaelemzés szerint van hatása a kezelésnek, akkor adódik a következő kérdés, hogy milyen csoportokban volt hatásos, és milyen csoportokban nem volt hatásos a kezelés. Ezt a szignifikáns eredményre vezető alapelemzést követően feltett kérdést célszerű már a Bonferoni-korrekcióval végzett t-próba segítségével tisztázni (post-hoc elemzés).

A varianciaelemzés alkalmazásának az alapadatok eloszlására vonatkozó előfeltételei vannak. Gondolatmenetének bemutatásakor abból indultunk ki, hogy a vizsgált adatok normális eloszlást mutatnak, és ennek megfelelően választottunk képleteket a variabilitás és a variancia számításához. Emiatt a normális eloszlás ellenőrzése a varianciaelemzés előfeltétele.

Amikor a csoportokon belüli variabilitást összegeztük, és egyszerűen összeadtuk a csoportonkénti négyzetes különbségösszegeket, akkor nem beszéltünk arról, hogy ha jelentős az eltérés az egyes csoportokon belüli varianciában, akkor ezek összevonása kerülendő. Ezért a varianciaelemzés másik feltétele, hogy legyenek egyformák a csoportokon belüli varianciák. (Ennek a feltételnek nem kell szigorúan megfelelni, de a nagy varianciaeltérések esetén tulajdonképpen az a legfontosabb kérdés, hogy mi lehet az oka, milyen mechanizmus révén, milyen folyamat miatt alakul ki jelentős csoportok közti varianciakülönbség. Az egyik dóziscsoportban miért szóródik jobban a betegek reakciója, mint a másikban? Nem kerültek-e a vizsgálatba speciális állapotú betegek? Nincs-e valamilyen interakció egyes dózistartományokban olyan gyógyszerekkel, genetikai fogékonysággal, környezeti expozícióval, aminek a vizsgálatára egyébként nem terjedt ki a kísérletvezetők figyelme?)

Új értágító eszközök kialakításához 4 új fémötvözetet fejlesztettek ki. A hagyományos ötvözetet kiegészítették egy adalékanyag különböző koncentrációját alkalmazva (A: 0,1%, B: 0,5%, C: 1%, D: 5%), és ugyanolyan körülmények közt építik be a belőlük készült stenteket. Ezt követően az erek megfelelő kapacitású



10. ábra

átjárhatóságának idejét (nap) regisztrálták. A 4. táblázat az egyes stentek esetében regisztrált napok számát összegzi. A csoportok közti átlagos időtartamok dózis-hatás kapcsolatot mutattak a grafikonon *(10. ábra)*.

A varianciaelemzés során nyert adatok szerint *(5. táblázat)* a relatív csoportok közti variancia (F = 3,116) nagyobb, mint a kritikus érték ($F_{(0,05;3;75]}$ = 2,727). A csoportok közt tehát nagyobb a különbség, mint amit pusztán a véletlennel meg lehetne magyarázni, azaz az adalékanyag koncentrációja hatással volt a kezelések eredményére. A magasabb koncentrációkhoz társuló jobb eredmények alapján érdemes használni az adalékanyagot.

4. táblázat

A	B	C	D
98	14	86	25
126	73	124	401
102	70	82	81
98	378	169	86
76	38	211	552
150	65	275	321
52	33	13	236
20	47	433	167
16	28	45	37
75	157	143	169
132	258	10	537
43	5	46	156
174	10	68	116
70	491	583	177
101	656	375	757
77	22	38	212
89	124	13	129
46	39	62	30
19		165	109
45			202
90			
34			

5. táblázat

	Négyzetes eltérések összege	Szabadsági fokok	Variancia	(F)
Csoportok közti variabilitás	226 240	3	75 413,2	3,116
Csoporton belüli variabilitás	1 815 229	75	24 203,1	
Teljes variabilitás	2 041 469	78		

Csoportok kvalitatív adatainak összehasonlítása

Gyakran kell olyan kérdésekkel foglalkoznunk, ahol nem folytonos változók, kvantitatív adatok segítségével írjuk le a kiváltott hatásokat, hanem két vagy több kategóriába sorolt, kvalitatív adatokkal. Vizsgálhatjuk például, hogy adott gyógyszer hatására emelkedik-e a gyógyult betegek aránya; passzív dohányzók gyerekei gyakrabban szenvednek-e középfülgyulladástól; a határérték feletti szérum-trigliceridszint esetén magasabb-e a szívinfarktus kialakulásának kockázata; egy ellátási protokoll következetes alkalmazása javítja-e a kedvező klinikai kimenetel valószínűségét.

A kvalitatív adatok esetében az egyes kategóriákban regisztrált esetek száma összegzi a vizsgálat eredményét. A vizsgálati beszámolók leíró-statisztikai táblázataiban is ezek az esetszámok, illetve a nekik megfelelő gyakorisági adatok szerepelnek.

Várható érték

Az RhD-antigén jelenléte alapján Rh-pozitív és Rh-negatív vércsoportokat különböztetünk meg. A tulajdonság autoszomális domináns öröklésmenetet mutat. Rh-negatív csak az lehet, aki mindkét szülőtől ezt a jelleget örökli. Ha heterozigóta szülőknek gyereke születik, akkor 25% valószínűséggel lesz a gyermek Rh-negatív (homozigóta). A gyerekek 75%-a Rh-pozitív lesz (50% heterozigóta, 25% homozigóta). Ha 60 heterozigóta szülőpártól származó gyereket vizsgálunk, akkor azt várjuk, hogy $60 \times 0{,}25 = 15$ gyerek lesz Rh-negatív és $60 \times 0{,}75 = 45$ Rh-pozitív.

Ha egy monogénes genetikai betegség esetén azt tételezzük fel, hogy az öröklésmenet autoszomális domináns, és megvizsgáljuk 60 heterozigóta szülőtől származó gyermek egészségi állapotát, akkor azt várjuk, hogy a gyerekek negyede lesz egészséges. Ha azt látjuk, hogy a vizsgálatunk nem pontosan a várt esetszámokat (15 egészséges és 45 beteg) találta, akkor kicsi eltérés esetén egyszerűen a véletlen hatásával magyarázzuk a megfigyelt és a várt érték közti eltérést. Az eredményt pedig az alapfeltevésünk mellett szóló érvként használjuk. Ha viszont az eltérés nagy, akkor az merül fel bennünk, hogy a betegség nem az általunk feltételezett autoszomális domináns módon öröklődik.

Az öröklésmenettel kapcsolatos példák azt szemléltetik, hogy vannak olyan gyakorisági adatok, amik levezethetőek valamilyen biológiai, orvostudományi elméletből, vagy egy saját magunk által kidolgozott elképzelésből. (Például, ha feltételezzük, hogy a nemek közt nincs különbség egy gyógyszerfajta hatékonyságában, akkor a férfiak és a nők közt ugyanolyan gyakorisággal várjuk a kezelésre adott kedvező reakciót. Emiatt

azt feltételezzük, hogy a kezelt betegek nemi arányát fogjuk látni a sikeresen kezelt betegek közt is.)

Ebben az értelemben beszélhetünk arról, hogy az egyes vizsgálati kategóriákban az elemszámok előre megjósolható arányban fordulnak elő. Adott vizsgálat esetén, a vizsgálatba vontak számát ezeknek az elméletileg megalapozott arányoknak megfelelően tudjuk felosztani. Az így kapott esetszámok előfordulását várjuk a vizsgálat végén majd. Ezért definiáljuk őket várható értéknek (*E*). A vizsgálat végén kapott tényleges esetszámokat pedig megfigyelt esetszámnak (*O*).

Abban az esetben, ha a megfigyelt és a várható esetszámok közti eltérés kicsi, akkor tulajdonképpen az elmélet és a tapasztalat közti összhangra kaptunk biostatisztikai bizonyítékot. Ilyen esetben a feltevésünk kiállta a valóság próbáját. Ha viszont a különbség kellően nagy, akkor nyilvánvaló, hogy nincs összhang a tapasztalat és a várakozásaink közt. Ha pedig a valóság próbáját nem állta ki az elképzelésünk, akkor be kell látnunk, hogy tévedtünk.

Az elképzelésünk és a valóság közti különbség számszerűsítésére első közelítésben a megfigyelt és várható értékek közti különbség megfelelő: (*O*–*E*). Az összes vizsgált kategória adatainak megfigyelt-várható értékpárjai adják majd együttesen a tapasztalat és az elképzeléseink közti kapcsolat mérőszámát: $\sum(O–E)$. Az egyszerű összegzés azonban az eltérő előjelek miatt nem használható. Helyette a négyzetes eltérések összegét számítjuk kategóriánként, és ezeket összegezzük:

$$\sum (O-E)^2 .$$

Adott vizsgálati helyzetben minél nagyobb a négyzetes különbségek összege, annál nagyobb az eltérés tapasztalat és valóság közt. Ez az összeg azonban nem csak akkor lehet nagy, ha jelentős a tapasztalat és a hipotézis közti távolság, hanem akkor is, ha nagy elemszámú a vizsgálat. (Ha az összegzett négyzetes eltérések ugyanakkora eltérést mutatnak egy 20 fős vizsgálatnál, mint egy 2000 fősnél, akkor a 20 fősnél ez már igen jelentős eltérés lehet, míg a 2000 fősnél elenyésző.) A vizsgálat nagyságához kell tehát viszonyítanunk a négyzetes különbségek összegét, amit a várható esetszámokkal adunk meg. A vizsgálat nagyságához viszonyított négyzetes eltérések összege minden kategóriában összegezve olyan statisztikai mérőszám, ami már megbízhatóan és a vizsgálat méretétől függetlenül írja le a tapasztalat és a várakozásaink (kutatási hipotézisünk) közti különbséget. Ez a mutató a χ^2:

$$\chi^2 = \sum \frac{(O-E)^2}{E} .$$

Ha a χ^2 kicsi, akkor lényegében ugyanazt láttuk a valóságban, amit a nullhipotézis alapján vártunk, ezért a nullhipotézist elfogadjuk. Ha a χ^2 kellően nagy, akkor a nullhipotézist elvetve saját kutatási hipotézisünket tartjuk megalapozottnak. A kicsi és kellően nagy χ^2 érték közti határt jelentő kritikus értéket a χ^2-függvény segítségével számíthatjuk, vagy táblázatból kereshetjük ki. Mivel a χ^2 értéke a megfigyelt-várható értékpárokból számított eredmények összege, ezért egy kategória adatai alapján számított relatív négyzetes különbség összeg már nem ad külön információt a vizsgálathoz. (Ha elvesztenénk egy ilyen részösszeget, akkor az a többi részösszeg és a χ^2 segítségével pontosan számítható lenne.) A szabadsági fok tehát a kategóriák számánál egyel kevesebb lesz.

Vegyük észre, hogy szemben a megfigyelt értékekkel, amik minden esetben egész pozitív számok, a várható értékek a csoportlétszám-arányok és a vizsgált minta elemszámának szorzataként adódnak, azaz általában tört számok. Ennek megfelelően a leggyakrabban előforduló helyzet az, hogy ha a valóság tényei és az elképzelésünk közt teljes az összhang, akkor is látunk némi különbséget a megfigyelt és a várható értékek közt. Kicsit nagyobbnak adódik a χ^2, mint amilyennek valójában lennie kellene. Emiatt a kritikus értéket egy kicsit könnyebben lépi át a χ^2, és egy kicsit könnyebben jutunk arra a következtetésre, hogy a vizsgálat szignifikáns eredményre vezetett. A felülbecsült χ^2 miatt növekszik az elsőfajú hiba valószínűsége.

A felülbecslés mértéke az egyes megfigyelt-várható értékpárok esetében maximum 0,5 lehet. (A és $A + 1$ = B egész számok közt levő számok közül $A + B / 2 = A + 0{,}5$ van legmesszebb A-tól és B-től) Ha minden egyes kategóriában számított (O–E) különbség abszolút értékét 0,5-del csökkentjük, akkor maximálisan kompenzáltuk a hibát. Az így számolt χ^2 biztosan nem becsli felül a pontos értéket. Ugyanakkor, a legnagyobb hiba lehetőségére korrigáltunk, emiatt viszont az ilyen módon számított χ^2–nél biztosan nagyobb a valós érték. Ezzel a megoldással alulbecsültük a χ^2–t, ami annyit jelent, hogy egy kicsit nehezebben lép át a statisztikai mutató a kritikus értéken, és egy kicsit nehezebben jutunk arra a következtetésre, hogy szignifikáns eltérés volt a megfigyelt és várható esetszámok közt. A számítás menetének ilyen módon elvégzett módosítása a Yates-korrekció:

$$\chi^2 = \sum \frac{(|O - E| - 0{,}5)^2}{E}.$$

Nyilvánvaló, hogy a korrigálatlan és a korrigált között van valahol a pontos χ^2, és akármelyik számítási módszert is használjuk, hibát követünk el. A két hiba közül az utóbbit kell elvállalnunk, mivel a biostatisztikai gondolkodás menetében egyértelműen annak adunk prioritást, hogy lehetőleg elkerüljük a nem szignifikáns eltérések szig-

nifikánsként való értékelését, még azon az áron is, hogy nem veszünk észre meglevő szignifikáns eltérést. (Inkább nem változtatjuk a gyakorlatot megalapozatlanul – azaz konzervatív szemlélettel közelítünk a problémákhoz.)

Egy vizsgálatban Magyarország 50–64 éves népességére reprezentatív lakossági mintájára van szükség. A minta reprezentativitását χ^2 érték számításával értékelték a két nemre külön-külön *(6. táblázat)*. A szabadsági fok mindkét nemnél 2 volt, mivel 3 korcsoportot értékeltek. A kritikus $\chi^2_{[0,05;2]} = 5{,}991$-nél mindegyik esetben kisebb volt a számított χ^2, ami alapján nem tudták igazolni, hogy a referencianépesség összetételétől eltért a saját minta összetétele. A mintát reprezentatívnak tekintették.

A megfigyelt és a várható érték közti eltérés arra hívta fel a figyelmünket, hogy valami nincsen rendben azzal a χ^2 képlettel, amit levezettünk az előzőekben. Láttuk, hogy értelmezési problémák is felmerülnek a végeredménnyel kapcsolatban. Ennek oka tulajdonképpen egyszerű, de részleteiben itt most nem tárgyalható. Elégedjünk meg annyival, hogy az általunk levezetett képletek közelítő jellegűek. Csak akkor adnak megbízható eredményt, ha az egyes kategóriákban a várható esetszám legalább 5. Ez alatt a közelítő és a pontos megfigyelt-várható különbségek közt már jelentős az eltérés. (Csak a teljesség kedvéért jegyezzük meg, hogy ilyenkor már csak a Fisher-exact eljárás alkalmazható.)

A fenti eljárás során egy elméletileg vagy valamilyen kutatási hipotézisnek megfelelően meghatározott esetszám-eloszláshoz viszonyítottuk a vizsgálatunk során megfigyelt esetszám-eloszlást. Az eljárást a valóság és a kutatási hipotézis illeszkedésének a vizsgálataként értelmezhetjük (goodness of fit test).

6. táblázat

Nem	Korcsoport	Referencia esetszámok	Referencia gyakoriságok	Saját minta esetszámai (O)	Várható esetszámok (E)	$(\lvert O-E\rvert-0{,}5)^2/E$
Férfiak	50–54	515406	0,414	8378	8427,4	0,284
	55–59	419839	0,337	6797	6864,8	0,660
	60–64	310858	0,249	5200	5082,8	2,679
	összesen	1246103	1	20375	20375,0	$\chi^2=3{,}623$
Nők	50–54	584629	0,410	9800	9943,7	2,062
	55–59	474511	0,333	8089	8070,8	0,039
	60–64	366198	0,257	6354	6228,5	2,509
	összesen	1425338	1	24243	24243,0	$\chi^2=4{,}610$

A várható esetszám-eloszlást eloszlásfüggvény alapján is generálhatjuk, ha ismerjük, hogy milyen a vizsgált változónk eloszlása, és annak melyek az eloszlás-paraméterei. Például, ha tudjuk, hogy az adat, amivel dolgozunk, Poisson-eloszlású, akkor a teoretikusan várható esetszám-eloszlást azelőtt kiszámíthatjuk, mielőtt még a saját adatainkat ténylegesen feldolgoznánk. A Poisson-eloszlás egyparaméteres: ha egy jelenség teljes vizsgálatunkban megfigyelhető gyakoriságát számítjuk, akkor megkaptuk az eloszlás paraméterét. Ennek segítségével számíthatjuk már, hogy egyes esetszámok előfordulásának mekkora az elméletileg várható valószínűsége.

Problémát fog jelenteni ennél a megközelítésnél, hogy a Poisson-eloszlás a 0 esetszám előfordulási valószínűségének számításától indul, de tulajdonképpen végtelen nagy esetszámig folytatódik. Korábban már láttuk, hogy az általunk levezett χ^2-érték számítása csak addig megbízható, amíg a kategóriákban legalább 5 a várható esetek száma. A várható értékek számítását ezért a végtelenségig semmiképpen nem folytatjuk. (Nem kell minden nagy esetszámra kiszámítani az előfordulás várható valószínűségét.) Ráadásul, eleve csak addig a határig kellene folytatni a valószínűségi értékek számítását, amennyi a vizsgálati minta elemszáma. Ha 100 az elemszámunk, akkor nem fordulhat elő, hogy 101 vizsgálati alany produkál egy bizonyos reakciót. Az utolsó kategória, amire még számítjuk a várható esetszámot, általában összevont (pl. annak a valószínűsége, hogy 15 vagy annál több esetben kapunk pozitív reakciót), ami még legalább 5 esetet tartalmaz. Az összevonás esetenként a kicsi esetszámok felé is indokolt lehet, ha ott is 5 alatti várható esetszámokat számítunk.

A várható-megfigyelt értékpárokra számított χ^2 a korábban megbeszélteknek megfelelően értelmezhető. Annyi specialitása van ennek az illeszkedésvizsgálatnak, hogy a Poisson-eloszlás paraméterét is a saját minta adatai alapján számítottuk. Emiatt a szabadsági fokot nem csak azért kell eggyel csökkentenünk, mert az utolsó relatív négyzetes eltérés összeg már nem ad új információt az elemzéshez, hanem azért is, mert a Poisson-eloszlás paramétere is számított érték. Összességében a szabadsági fok itt kettővel lesz kisebb a kategóriák számánál.

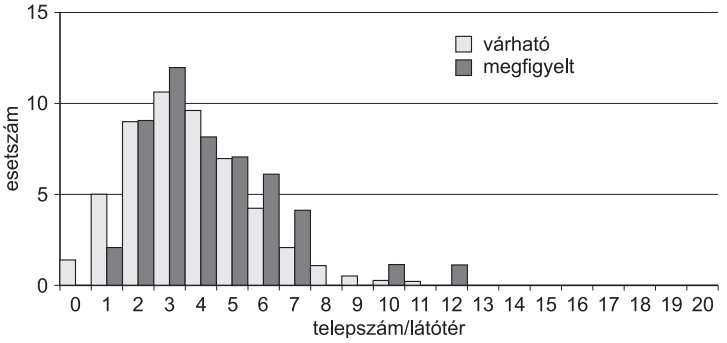
A látóterenkénti baktériumtelepek száma Poisson-eloszlást követ egy standardizált vizsgálati rendszerben, ahol a látóterenkénti átlagos telepszámnak 3,6-nak kell lennie ahhoz, hogy jól működjön a vizsgálat. Minden vizsgálat elején ellenőrizni kell, hogy megfelelő-e a tenyészet. 50 látóteret olvastak le. A *7. táblázat* tartalmazza a látóterekben megszámolt telepek számának eloszlását és a 3,6-es átlag alapján számított várható esetszámokat. A 0 és az 1 telepszámhoz már kevesebb, mint 5 eset tartozott, ezért a kezdő kategória a „0 vagy 1 telep/látótér" lett. A 6 az első magasabb esetszám, amihez már 5 alatti várható esetszám tartozott, ezért a „6 vagy annál több telep/látótér" lett a

7. táblázat

Telepek száma/látótér	Várható esetszámok (E)	Megfigyelt esetszámok (O)	$(\lvert O-E\rvert-0{,}5)^2/E$
0 vagy 1	6,3	2	2,279
2	8,9	9	0,014
3	10,6	12	0,072
4	9,6	8	0,118
5	6,9	7	0,021
6 vagy annál több	7,8	12	1,762

legfelső elemzett kategória. A számított $\chi^2 = 4{,}267$ kisebb volt, mint a kritikus $\chi^2_{[0,05;4]} = 9{,}488$-nél. (A szabadsági fok a 6 csoport miatt volt 4.) A telepszám-eloszlás tehát megfelelt a vártnak, nem tért el attól szignifikánsan. A vizsgálati rendszer alkalmas volt a feladat elvégzésére.

Gyakran kell azt ellenőrizni, hogy egy folytonos változó normális eloszlású-e. Ilyenkor is alkalmazhatjuk a fenti megoldásokat. Azaz a minták adatai alapján számítjuk adataink eloszlás-paramétereit (átlagot és szórást). Ezt követően kategóriákat határozunk meg (felosztjuk a folytonos skálát kategóriákra). A kategóriák számának, szélességének kialakításakor az 5-ös várható elemszámszabályra kell tekintettel lennünk. A szélső kategóriát pedig itt is végtelenig bővítve kell definiálni. A χ^2 számítása ettől kezdve követi a fenti mintát. A kritikus érték megállapításakor azt kell figyelembe venni, hogy itt kettő eloszlás-paramétert határozunk meg a minta adatai alapján, emiatt a szabadsági fok a kategóriák számánál hárommal lesz kisebb.



11. ábra. A látóterenkénti baktériumtelepek megfigyelt és várható száma

Csoportok közti különbség elemzése

Kvalitatív adatok elemzésekor azonban a leggyakrabban megválaszolandó kérdés az, hogy adott jelleg előfordulása eltérést mutat-e különböző csoportokban? (Különböző-e a betegség gyakorisága dohányosok közt és nem dohányzók közt, egy bőrelváltozás gyakoribb-e egy foglalkozási csoportban, mint más foglalkozásúak közt stb.) Ilyen vizsgálatok adatfeldolgozása során első lépésként az elemi adatokat (a vizsgálati csoportokon belül megfigyelt kategóriánkénti esetszámokat) táblázatba rendezzük. Az oszlopokba a befolyásoló tényezőt (a vizsgálat során összehasonlítandó csoportok megnevezését), a sorokba a kiváltott hatás kategóriáit szokás elhelyezni. Az így kapott kontingenciatáblázat minimum 2 × 2-es, de akárhány oszlopból, illetve sorból is állhat *(8. táblázat)*.

A kontingenciatáblázat adatainak felhasználásával kell válaszolni a vizsgálat alapkérdésére: van-e kapcsolat a befolyásoló tényező és a kiváltott hatás között? Nullhipotézisünk itt is az, hogy nincs ilyen kapcsolat, azaz a vizsgált befolyásoló tényező és a kiváltott hatás független egymástól.

A nullhipotézis alapján várható értékeket tudunk számítani a kontingenciatáblázat minden cellájára, azaz a „Hányan lennének adott csoporton belül az adott kategóriában?" kérdést minden cella esetében meg tudjuk válaszolni. Ha ugyanis a csoportok közt nincs különbség a kiváltott hatás szempontjából, akkor az egyes kiváltott hatás kategóriákban összesen megfigyelt esetszám ($\Sigma_{K1} = N_{11} + N_{21}$ és $\Sigma_{K2} = N_{12} + N_{22}$) a csoportok létszámának arányában oszlik meg (B_1-csoport részaránya Σ_{B1}/N; B_2-csoporté Σ_{B2}/N). Ha például két csoportot vizsgálunk, és a két csoportba összesen 100 beteg tar-

8. táblázat. 2 x 2-es kontingenciatáblázat szerkezete

Kategóriák (kiváltott hatás)	Csoportok (befolyásoló tényező) B1	B2	Mind
K1	N_{11}	N_{21}	$\Sigma_{K1} = N_{11} + N_{21}$
K2	N_{12}	N_{22}	$\Sigma_{K2} = N_{12} + N_{22}$
Mind	$\Sigma_{B1} = N_{11} + N_{12}$	$\Sigma_{B2} = N_{21} + N_{22}$	N

9. táblázat

	Megfigyelt esetszámok				Várható esetszámok	
	B1	B2	Mind		B1	B2
K1	N_{11}	N_{21}	$\Sigma_{K1} = N_{11}+N_{21} = 100$	K1	$= 100 \times 0,5 = 50$	$= 100 \times 0,5 = 50$
K2	N_{12}	N_{22}	$\Sigma_{K2} = N_{12}+N_{22} = 400$	K2	$= 400 \times 0,5 = 200$	$= 400 \times 0,5 = 200$
Mind	$\Sigma_{B1} = N_{11}+N_{12} = 250$	$\Sigma_{B2} = N_{21}+N_{22} = 250$	N = 500			
Csoportok részaránya	$\Sigma_{B1}/N = 250/500 = 0,5$	$\Sigma_{B2}/N = 250/500 = 0,5$				

tozik ($\Sigma_{K1} = N_{11} + N_{21} = 100$), a két csoport pedig 250-250 fős ($\Sigma_{B1} = 250$ és $\Sigma_{B2} = 250$), akkor, mivel a csoportok létszáma egyforma, a betegek is egyforma számban fordulnak elő a két csoportban a nullhipotézis alapján. Ebben az esetben a várható érték a két cellára (N_{11}-re és N_{12}-re) 50-50 lesz. Ebben az 500 fős vizsgálatban az egészségesek száma $\Sigma_{K2} = 500 - 100 = 400$, akiket a csoportlétszámok arányában szétosztva 200-es várható esetszámot kapunk N_{12}-re és N_{22}-re is *(9. táblázat)*.

Ha ugyanebben a helyzetben a csoportok létszáma $\Sigma_{B1} = 400$ és $\Sigma_{B2} = 100$ lenne, akkor a 100 beteg is ennek megfelelően oszlana meg: 80 beteg lenne a nagyobbik és 20 beteg a kisebbik csoportban. Az egészségesek száma is ebben az arányban fordulna elő *(10. táblázat)*.

10. táblázat

	Megfigyelt esetszámok				Várható esetszámok	
	B1	B2	Mind		B1	B2
K1	N_{11}	N_{21}	$\Sigma_{K1} = N_{11}+N_{21} = 100$	K1	$= 100 \times 0,8 = 80$	$= 100 \times 0,2 = 20$
K2	N_{12}	N_{22}	$\Sigma_{K2} = N_{12}+N_{22} = 400$	K2	$= 400 \times 0,8 = 320$	$= 400 \times 0,2 = 80$
Mind	$\Sigma_{B1} = N_{11}+N_{12} = 400$	$\Sigma_{B2} = N_{21}+N_{22} = 100$	N = 500			
Csoportok részaránya	$\Sigma_{B1}/N = 400/500 = 0,8$	$\Sigma_{B2}/N = 100/500 = 0,2$				

A várható esetszámok képzésénél az egyes kategóriákban összesen megfigyelt esetszámokat tehát szétosztjuk a csoportnagyságnak megfelelően. Az így kapott várható értékek azt fejezik ki, hogy milyen lenne az esetek eloszlása csoportokon belül kategóriánként, ha a csoportképző faktor nem befolyásolná a betegség kialakulását.

Ezek után már a χ^2 szokásos módon számítható, és segítségével a tapasztalat és az elméleti feltevés (nullhipotézis) közti kapcsolat természete már értékelhető. Az itt kapott χ^2 kritikus értékének meghatározásához szükségünk van még a szabadsági fokra. Ha egy n × m kontingenciatáblázat sor- és oszlopösszegeit tudjuk, akkor a várható értékek közül nagyon sokat nem kell már felhasználni a számításaink során. Minden sorban 1 adat van, ami ha elveszne, akkor a sor egyéb tagjaiból és a sorösszegből még számítható lenne. Ugyanez igaz az oszlopokra is. Ennek megfelelően a χ^2 számításakor a szabadsági fok:

$$df = (n-1)(m-1)$$

A cisztás fibrózis korszerű gondozása jelentősen javítja a betegek életminőségét. A diagnózis felállítása után ezért alapvetően fontos a gondozási központokba való bekerülés. Sokan azonban alapellátási körülmények között kerülnek ellátásra. A cisztás fibrózisos gyermekek szülei által létrehozott civil szervezet próbálja felkutatni és a szakellátó központok felé irányítani az érintetteket. A szervezet azonban nem épült ki az egész országban. Megvizsgálták, hogy a szervezet által lefedett és a szervezet által le nem fe-

11. táblázat

	Megfigyelt esetszámok			Várható esetszámok	
	Civil szervezet működési területe	Civil szervezet által le nem fedett terület	Mind	Civil szervezet működési területe	Civil szervezet által le nem fedett terület
Alapellátásban gondozott gyermekek száma	155	209	364	175,15	188,85
Szakellátó központban gondozott gyermekek száma	318	301	619	297,85	321,15
Mind	473	510	983		
Csoportok részaránya	0,48	0,52			

dett területek között milyen a gondozásba vétel hatékonysága *(11. táblázat)*. A számított $\chi^2 = 7{,}095$ nagyobb a kritikus $\chi^2_{[0{,}05;1]} = 3{,}841$ küszöbnél, ezért a megfigyelt és a várható értékek közti különbség szignifikáns. Mivel a szakellátásban gondozottak aránya 67% a civil szervezet által ellátott, és 59% az általuk nem ellátott területen, a szervezet működését kedvező hatásúnak értékeljük az eredmény alapján. (A $\chi^2 = 7{,}095$-hoz tartozó $p = 0{,}008$ értéket számítógép segítségével kapjuk.)

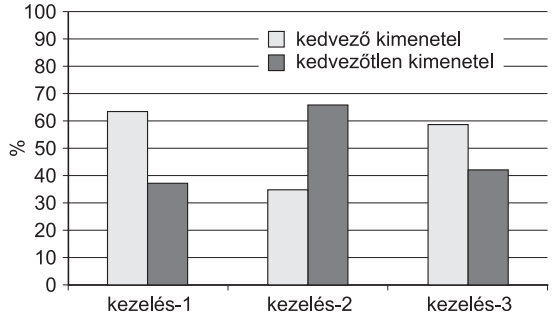
Ha 2 × 2-es a kontingenciatáblázat, akkor az alapkérdésünkre közvetlenül válaszolni tudunk a χ^2-próba segítségével. Ha azonban ennél nagyobb a táblázat (több kategóriába soroltuk a befolyásoló tényezőt és/vagy a kiváltott hatást), akkor a vizsgálati kérdés csak az lehet, hogy van-e egyáltalán hatása a csoportképző faktornak a vizsgált paraméterre? Erre a χ^2-próba választ is ad. Arra viszont nem kapunk így választ, hogy melyik csoportok közt és melyik kiváltott hatás szempontjából jelentkezik valójában az eltérés? A táblázat méretétől függően nagyon sok specifikus kérdést lehet az ilyen post hoc elemzések során feltenni. A specifikus kérdés megválaszolásakor a döntési küszöböt kell csökkentenünk annak érdekében, hogy összességében a vizsgálatban az első fajú hiba szintje ne emelkedjen 5% fölé, azaz, hogy ne értékeljünk a sokszoros hipotézistesztelés következtében látszólagos szignifikáns eredményt valóban szignifikánsnak.

Szívinfarktuson átesett betegek ellátására egy ellátási körzetben 3 lehetőség áll rendelkezésre. A háromféleképpen kezelt betegek 1 éven belüli állapotjavulását vizsgálták, annak érdekében, hogy megállapítsák, melyik ellátási forma szolgálja hatékonyabban a betegek érdekeit *(12. táblázat)*.

12. táblázat

	Megfigyelt esetszámok				Várható esetszámok		
	Kezelés-1	Kezelés-2	Kezelés-3	Mind	Kezelés-1	Kezelés-2	Kezelés-3
Kedvező kimenetel	41	21	104	166	35,38	33,20	97,42
Kedvezőtlen kimenetel	24	40	75	139	29,62	27,80	81,58
Mind	65	61	179	305	65	61	179
Csoportok részaránya	0,21	0,20	0,59				

A χ^2-próba szignifikáns eredményre vezetett (p = 0,002). Ennek alapján meg tudták állapítani, hogy egyáltalán nem mindegy, hogy melyik protokoll szerint kezelték a betegeket *(12. ábra)*. A preferálandó, illetve kerülendő ellátási forma azonosítása érdekében részletes elemzést végeztek. Páronként az alábbi teszteredményeket kapták: p_{1vs2} = 0,001 *(13. táblázat)*; p_{1vs3} = 0,484 *(14. táblázat)*; p_{2vs3} = 0,001 *(15. táblázat)*. Megállapították, hogy a 2-es protokoll helyett, az egymással egyenértékű 1-es és 3-as protokollt kell a jövőben használni, mert a p_{1vs2} = 0,001 és a p_{2vs3} = 0,001 még a 0,05/3 = 0,017 döntési küszöbnél is szignifikánsnak jelezte a hatékonyságbeli különbségeket.



12. ábra

13. táblázat

	Megfigyelt esetszámok			Várható esetszámok	
	Kezelés-1	Kezelés-2	Mind	Kezelés-1	Kezelés-2
Kedvező kimenetel	41	21	62	31,98	30,02
Kedvezőtlen kimenetel	24	40	64	33,02	30,98
Mind	65	61	126	65,00	61,00
Csoportok részaránya	0,52	0,48			

14. táblázat

	Megfigyelt esetszámok			Várható esetszámok	
	Kezelés-1	Kezelés-3	Mind	Kezelés-1	Kezelés-3
Kedvező kimenetel	41	104	145	38,63	106,37
Kedvezőtlen kimenetel	24	75	99	26,37	72,63
Mind	65	179	244		
Csoportok részaránya	0,27	0,73			

15. táblázat

	Megfigyelt esetszámok			Várható esetszámok	
	Kezelés-2	Kezelés-3	Mind	Kezelés-2	Kezelés-3
Kedvező kimenetel	41	104	125	31,77	93,23
Kedvezőtlen kimenetel	40	75	115	29,23	85,77
Mind	61	179	240		
Csoportok részaránya	0,25	0,75			

Párba rendezett kvalitatív adatok elemzése

Kvalitatív vizsgálatoknál is hatékonyságnövelő, ha nem két, egymástól független csoportot vizsgálunk, hanem párba rendezett adatokat állítunk elő. Vagy előtte-utána típusú vizsgálatot hajtunk végre, ahol az egy-egy vizsgálati alany kezelés/expozíció előtti és utáni adatai közötti eltérést vizsgáljuk, vagy néhány zavaró tényező szempontjából párba rendezett, de egy csoportok közötti különbséget meghatározó, csoportképző faktor szempontjából eltérő csoport tagjain hajtjuk végre a vizsgálatot. Mindkét esetben adatpárok adják az induló információt.

A legegyszerűbb helyzetet alapul véve (amikor két csoportot és kétféle kategóriába sorolt vizsgálati eredményt elemzünk), a párokat négy csoportba sorolhatjuk: a pár hasonló státuszú (mindkettőnél kialakult a vizsgált hatás: N_{++}, vagy mindkettőnél elmaradt a hatás kifejlődése: N_{--}), eltérő státuszú (az egyik csoporthoz tartozó tagban kifejlődött a hatás, a másik csoportba tartozónál nem: N_{+-}, és fordítva: N_{-+}) lehet.

Ha a csoportképző faktor befolyásoló szerepét akarjuk értékelni, akkor érdemi információval azok a párok szolgálnak, akiknél eltérő volt a kiváltott hatás (N_{+-}, N_{-+}). Ha két kategóriában hasonló a párok száma, akkor nincs különbség a két csoport közt a kiváltott hatás szempontjából (mindkét csoport ugyanolyan gyakran adja a pozitív választ produkáló tagot). Ha viszont jelentősen eltér egymástól ez a két elemszám, akkor az egyik csoportba tartozók gyakrabban alakítanak ki pozitív választ, jobban reagálnak a befolyásoló tényezőre, mint a másik csoport tagjai. Emiatt az alapkérdés megválaszolásához ennek a két esetszámnak a különbségét kell értékelni.

Az értékelés menete megfelel a fentebb már több alaphelyzetben követett mintának. A két esetszám négyzetes különbségét viszonyítjuk a vizsgálat méretéhez:

$$\chi^2 = \frac{(N_{+-} - N_{-+})^2}{(N_{+-} + N_{-+})} .$$

A statisztikai eljárás neve McNemar-teszt. Mivel összesen két kategóriát vizsgáltunk, a szabadsági fok 1 lesz. A vizsgálatot itt is érdemes Yates-korrekció segítségével jobban interpretálhatóvá tenni:

$$\chi^2 = \frac{(|N_{+-} - N_{-+}| - 1)^2}{(N_{+-} + N_{-+})} .$$

Egy üzemben 424 alkalmazott dolgozik. A gyártás két fázisban zajlik. Az elsőben alkalmazott vegyszerek allergiás reakciót okozhatnak. Megvizsgálták, hogy az első fázisban dolgozók ténylegesen veszélyeztetettebbek-e. Az alkalmazottakat 212 párba sorsolták véletlenszerűen, és rögzítették beosztásukat és allergiás anamnézisüket *(16. táblázat)*. Azt látták, hogy gyakrabban fordul elő, hogy az első fázisban dolgozó allergiás, de a második fázisban dolgozó párja nem, mint az, hogy a második fázisban dolgozó allergiás, de az első fázisban dolgozó párja nem. Az eltérés véletlenszerűségére McNemar-tesztet végeztek:

$$\chi^2 = \frac{(|N_{+-} - N_{-+}| - 1)^2}{(N_{+-} + N_{-+})} = \frac{(|52 - 32| - 1)^2}{(52 + 32)} = 4{,}298 \cdot$$

Az eredmény nagyobb, mint a $\chi^2_{[0,05;1]} = 3{,}841$ kritikus érték. Az elsődleges benyomásokat igazolta a vizsgálat. Valóban szignifikáns a kapcsolat az első munkafázis okozta expozíciók és az allergiás betegségek kifejlődése közt.

16. táblázat

Első fázisban dolgozó	Második fázisban dolgozó	N
Allergiás betegség +	Allergiás betegség +	48
Allergiás betegség +	Allergiás betegség –	52
Allergiás betegség –	Allergiás betegség +	32
Allergiás betegség –	Allergiás betegség –	80

Folytonos változók közötti kapcsolat elemzése

Ha két intervallum vagy arányskálán mért folytonos változó közti kapcsolat természetét szeretnénk leírni (arra vagyunk kíváncsiak, hogy a szérumban mérhető gyógyszerszint milyen kapcsolatban van a gyógyuláshoz szükséges idővel; az elfogyasztott alkohol mennyisége milyen mértékben rontja a reakcióidőt; az ivóvízben levő szennyező anyag milyen mértékben rontja a vese kiválasztó képességét; a felnőttek életkora milyen kapcsolatban van a csontok kalciumtartalmával stb.), akkor tulajdonképpen azt a függvényt keressük, amelyik a lehető legpontosabban megmutatja, hogy az egyik paraméter változása milyen mértékben módosítja a másik paraméter értékét.

A biológiai rendszerek tagjainak egymásra hatása rendkívül változatos, emiatt sokfajta függvénnyel lehet csak a folyamatokat leírni. Ezek közül legegyszerűbb a lineáris kapcsolat. A következőkben a folytonos változók közötti lineáris kapcsolat elemzésekor használható statisztikai alapismereteket tekintjük át. Nem foglalkozunk nem lineáris kapcsolatok elemzésével azon túl, hogy megemlítjük: sok nem lineáris viszony elemzésekor lehetőség van arra, hogy a változók megfelelő transzformálásával linearizáljuk a kapcsolatot, ami után semmi akadálya nincs az alábbiakban összefoglalt eljárások alkalmazásának.

Korreláció

A legegyszerűbb kérdés, amelyet a fenti kérdések vizsgálatakor meg kell válaszolnunk, hogy két jelenség között van-e egyáltalán kapcsolat, az egyik jelenség változása maga után vonja-e a másik változását? Ilyen kérdés megválaszolását célszerű a lehető legegyszerűbb módon kezdeni: ábrázoljuk a megfigyelt adatpárokat egy diagramon. A két tengelyen szerepelnek a vizsgált változók. Így minden vizsgálati alanynak egy pont felel meg a koordinátarendszerben. Az ilyen szórásdiagram sokszor könnyen értékelhető kapcsolatot tár fel, mert szemmel is jól értékelhető pozitív vagy inverz kapcsolat olvasható le a számpárok elhelyezkedéséből. Természetesen valós adatoknál csak trendet látunk, az adatok soha nem illeszkednek egy egyenesre. A trendnek megfelelő vonal körül kisebb-nagyobb mértékben szóródnak a ténylegesen mért értékek *(13. ábra)*.

Általában a vizuális benyomásunk csak nagyon erős kapcsolat esetén nyújt megbízható támpontot. Gyakoribb eset, hogy ilyen módon még a trend irányára vonatkozóan is

nehéz a véleményünket megfogalmazni. Természetes tehát, hogy a kapcsolat erősségének a leírására valamilyen kvantitatív paramétert kell használni.

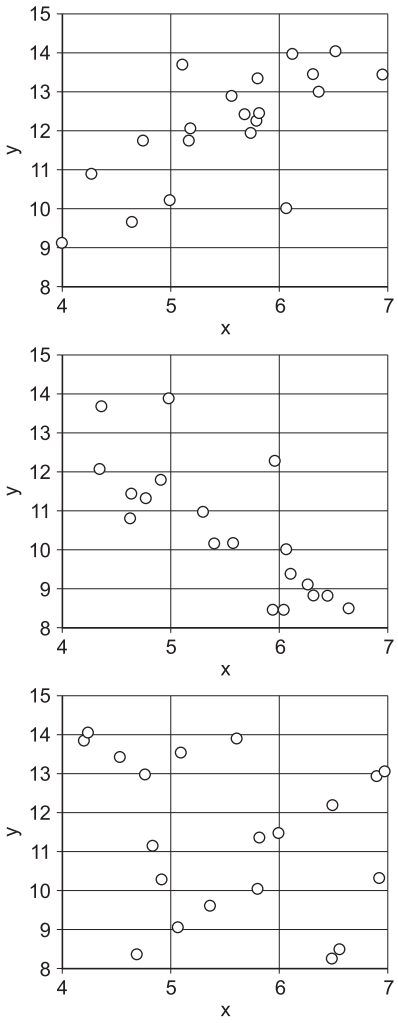
Két paraméter kapcsolt változékonyságát ugyanazzal a gondolatmenettel tudjuk számszerűsíteni, amit egyetlen változó esetében alkalmaztunk. Egy változó (x) esetén N_x mintanagyság mellett a variabilitás leírása az átlag ($\bar{x}$) körüli variancia (V_x), szórás (SD_x) meghatározását jelentette, ahol az átlagtól való négyzetes eltérések összege (S_x) volt a variabilitás elemi mérőszáma:

$$SS_x = \sum (x - \bar{x})^2$$

$$V_x = \frac{SS_x}{N_x - 1} = \frac{\sum (x - \bar{x})^2}{N_x - 1},$$

$$SD_x = \sqrt{\frac{SS_x}{N_x - 1}} = \sqrt{\frac{\sum (x - \bar{x})^2}{N_x - 1}}.$$

Két változó variabilitásának kapcsoltságát kovarianciával (C_{xy}) írhatjuk le. Ennek származtatása nem az egyik paraméter átlagától, hanem mindkét paraméter átlagától indul el: azaz nem $\bar{x}$ és nem $\bar{y}$ körüli elhelyezkedést, hanem a megfigyelt ($x;y$) értékpárok ($\bar{x}\,;\bar{y}$) értékpárhoz viszonyított elhelyezkedését vizsgáljuk. Ahhoz, hogy az éppen értékelt ($x;y$) pontba jussunk a ($\bar{x}\,;\bar{y}$) pontból, két lépést kell megtennünk. Először x-tengely mentén kell elmozdulnunk ($x–\bar{x}$) távolságra, utána y-tengely mentén ($y–\bar{y}$) távolságra. [Ha a vizsgált x nagyobb, mint az $\bar{x}$, akkor az ($x–\bar{x}$) pozitív előjelű lesz, és pozitív irányba kell mozdulni. Ha a vizsgált x kisebb, akkor az ($x–\bar{x}$) negatív előjelű lesz, és negatív irányba kell mozdulni. Hasonlóan értelmezzük az y-tengely menti mozgást is.] A két távolság növekedésével együtt jár a két vizsgált pont közti eltérés növekedése. A két távolság szorzata pedig olyan érték, ami összességében fejezi ki az adatpár távolságát a tipikus vizsgálati alanytól. Ezek a távolságok minden



13. ábra. Két folytonos változó (x, y) közötti szóródás kapcsolt szórásdiagramon

adatpárra meghatározhatók. Összegük annál nagyobb lesz, minél jelentősebb a tipikus pont körüli szóródás. Ennek a variabilitási mérőszámnak az értéke azonban nem csak attól függ, hogy milyen az adatok elhelyezkedése a szórásdiagramon, hanem attól is, hogy hány elemű a minta. Minél több az adat, annál nagyobb lesz a mutató. Ezt a problémát a mutató elemszámhoz (N) viszonyított értékének használatával lehet kezelni. Pontosabban a szabadsági fokhoz viszonyított értéket, ami a tipikus pontszámításokban való alkalmazása miatt 1-gyel kisebb lesz, mint az elemszám.

$$C_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{N - 1}$$

A kovariancia már sokat elárul a két paraméter közti kapcsolatról. Ha a tipikus pont központjával négy kvadránsra osztjuk a szórásdiagramot, akkor a négy mezőben elhelyezkedő pontokhoz tartozó kovarianciatag $(x–\bar{x})$ $(y–\bar{y})$ előjele a jobb felső és a bal alsó mezőben pozitív, a bal felső és a jobb alsó mezőben negatív lesz. Értelemszerűen a kovariancia negatív értékű is lehet (szemben a varianciával, ami soha nem negatív szám). Ha mindezek mellett figyelembe vesszük, hogy az adatainkra illesztett trendvonal mindig illeszkedik az $(\bar{x}\,;\bar{y})$ tipikus pontra, akkor megállapíthatjuk, hogy, ha teljesen illeszkednek az adataink egy emelkedő trendvonalra, akkor az összes pont a bal alsó és a jobb felső mezőbe esik, azaz az összes kovarianciatag pozitív szám lesz, ami miatt maga a kovariancia is pozitív értéket vesz fel. Azonban, ha csökkenő trendvonalra illeszkednek a vizsgálati eredmények, akkor a bal felső és a jobb alsó mezőben találjuk a negatív előjelű tagokat, amelyek negatív előjelű kovarianciát eredményeznek. Ha az egyik paraméter változása nem kapcsolódik a másik paraméter valamilyen módosulásához, akkor a szórásdiagramon egy vízszintes vonalra illeszkedő trendvonalat látunk. Ilyen esetben az adatok a négy kvadráns közt egyenletesen oszlanak meg, ami a negatív és pozitív tagok egymást kioltó hatása miatt összességében nulla lesz. A három szélsőséges helyzet természetesen soha nem fordul elő valós biológiai rendszerekben. Az eredmények nem illeszkednek tökéletesen a trendvonalakra. A szóródásuk miatt emelkedő trend esetén is látunk bal felső és jobb alsó mezőben adatpárokat *(14. ábra)*.

Minél jelentősebb a szóródás, annál több adat jelenik meg ebben a két, negatív előjelű kovarianciatagot eredményező mezőben, és a több negatív kovarianciatag miatt kisebb lesz a számított kovariancia. (A szóródásnövekedés végpontja az, amikor már kicsit sem több a pozitív kovarianciatagok súlya, és kiegyenlítődnek a negatív és pozitív részösszegek. Hasonló gondolatmenet vezethető le a csökkenő trendekre vonatkozóan is.) Ezek alapján megállapíthatjuk, hogy a kovariancia nem csak a trendvonal emel-

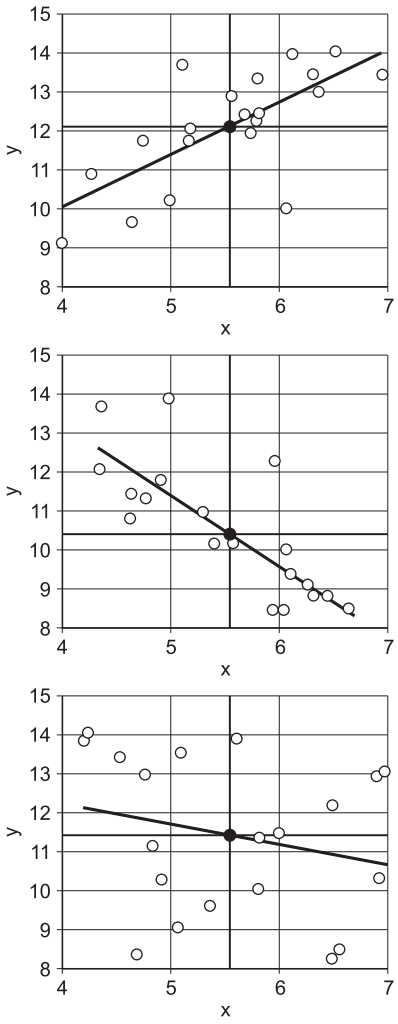
kedő, csökkenő vagy vízszintes jellegéről ad felvilágosítást, de arra is alkalmas, hogy értékelje a trendvonal körüli szóródás mértékét, azaz a két paraméter közti kapcsolat szorosságát. Minél nagyobb pozitív vagy negatív szám a kovariancia, annál szorosabb a kapcsolat.

A kovariancia mértékegységgel rendelkező mutató. Ha egy toxin koncentrációját és egy immunológiai marker koncentrációját vizsgáljuk, akkor a két koncentráció mértékegységének a szorzata lesz a kovariancia dimenziója. Ha a toxinkoncentráció nmol/ml, a markeré pedig IU/l, akkor a kovariancia mérőszáma nmol × IU/ml × l, amit nem egyszerű értelmezni. Ha a mértékegységeket változtatjuk, hogy könnyebb legyen a kovariancia értelmezése, akkor a kovariancia mérőszáma is módosul. Ez kényelmetlenné teszi a kovariancia alapú kapcsolatelemzést. Szerencsére a változók standardizálása révén dimenzió nélküli, a paraméter egységnyi szórásához viszonyított, ezért a különböző paraméterek esetében összehasonlítható (más vizsgálatok eredményeivel is egyszerűen összehasonlítható) mérőszámhoz juthatunk.

Minden x helyett x/SD_x-et, és minden y helyett y/SD_y-t (és a tipikus értékekre is $\bar{x}/SD_x$-et illetve $\bar{y}/SD_y$-t) használva, az eredeti kovarianciaképlet módosul:

$$r = \frac{1}{(N-1)} \frac{\sum (x-\bar{x})(y-\bar{y})}{SD_x SD_y} .$$

Így már nem csak a kapcsolat irányát, hanem a szorosságát is jól leíró mérőszámhoz jutunk, amit korrelációs koefficiensnek hívunk (r). Abban az esetben, ha emelkedő trendvonalat láttunk a szórásdiagramon, és az összes adat ehhez a trendvonalhoz illeszkedett, akkor a korrelációs koefficiens értéke 1 lesz. Ilyenkor teljes pozitív



14. ábra. Vizsgálati eredmények tipikus ($\bar{x};\bar{y}$) pont körüli elrendeződése a trendvonalakkal

korrelációról beszélünk. Ebben a helyzetben ugyanis az egyik paraméter pontosan meghatározza másik paraméter értékét (adott x-hez csak egy adott y tartózhat, és fordítva.) Ha csökkenő trendvonalra illeszkednek tökéletesen a vizsgálati eredmények, akkor tökéletes inverz korrelációról beszélünk, amit a korrelációs koefficiens –1-es értéke jelez. Ha a koefficiens értéke éppen nulla, akkor az illesztett trendvonal vízszintes, és ez jelzi a két vizsgált paraméter tökéletes függetlenségét (akármilyen értéket is vesz fel x, ahhoz mindig ugyanaz a jellemző y-érték fog hozzátartozni; x változása nem hozza magával y módosulását). Az interpretációs szabályok értelmezéséhez ezek a szélsőséges helyzetek hozzásegítenek, de valós mintákon soha nem találkozunk velük. Emelkedő vagy csökkenő trendvonal körül többé-kevésbé szóródó eredményeket látunk a szórásdiagramon. A korrelációs koefficiensek pedig valahol 0 és –1 között, illetve 0 és +1 között helyezkednek el. Minél szorosabb a korreláció, annál távolabb kerül nullától a korrelációs koefficiens.

Annak eldöntésére, hogy az adott korrelációs koefficiens eltérése a semleges nullától véletlennek tulajdonítható-e, vagy pedig annak, hogy a két változó valóban kapcsolatban van (szignifikáns-e a trendvonal által sugallt látszólagos kapcsolat), a korrelációs koefficiens konfidencia-intervallumának megadásával tudunk válaszolni. A korrelációs koefficiens nagysága ugyanis önmagában nem elég annak eldöntésére, hogy van-e kapcsolat egyáltalán a paraméterek közt. Egy biológiai rendszerben ugyanazt a hatást több befolyásoló tényező is képes előidézni. A különböző faktorok különböző hatékonysággal képesek változást generálni. Ebben az értelemben vannak gyenge és vannak erős determinánsok. Az erős determináns esetében nullától távoli korrelációs koefficienseket kapunk, a gyenge determinánsoknál pedig a nullához közelebb levőt. Ugyanakkor mindegyik ténylegesen faktor kapcsolatban van a kiváltott hatással.

A korrelációs koefficiens standard hibája (SE_r) és 95%-os megbízhatósági tartománya $MT_{95\%;r}$, ahol (N-2) a szabadsági fokot, t pedig a t-eloszlás megfelelő értékét jelöli:

$$SE_r = \sqrt{\frac{(1-r^2)}{(N-2)}}$$

$$MT_{95\%;r} = r \pm t_{[0,05;N-2]} SE_r .$$

Ez a megbízhatósági tartomány tartalmazza 95%-os valószínűséggel azt a korrelációs koefficienst, ami pontosan leírja a két paraméter közti kapcsolat erősségét. Ha ez az intervallum teljes egészében pozitív tartományban van, akkor minden valószínű korrelációs koefficiens pozitívnak mutatja a változók viszonyát, vagyis szignifikáns-

17. táblázat

Életkor, év (x)	Szisztolés vérnyomás, Hgmm (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
53	170	–12	31,6	–379,2
54	143	–11	4,6	–50,6
54	89	–11	–49,4	543,4
55	138	–10	–0,4	4
57	113	–8	–25,4	203,2
59	98	–6	–40,4	242,4
61	126	–4	–12,4	49,6
63	160	–2	21,6	–43,2
65	175	0	36,6	0
65	133	0	–5,4	0
68	95	3	–43,4	–130,2
68	208	3	69,6	208,8
68	114	3	–24,4	–73,2
69	112	4	–26,4	–105,6
70	170	5	31,6	158
70	171	5	32,6	163
72	131	7	–7,4	–51,8
72	116	7	–22,4	–156,8
77	189	12	50,6	607,2
80	117	15	–21,4	–321

nak interpretálhatjuk a kapcsolatot. (Negatív tartományban, inverz kapcsolatok esetén hasonló a helyzet.) Ha a megbízhatósági tartomány pozitív és negatív értékeket is tartalmaz, azaz valószínűnek látunk pozitív és negatív trendet leíró értékeket egyaránt, akkor a nyilvánvaló ellentmondás miatt nem állíthatjuk, hogy bizonyítékot találtunk a két változó közötti kapcsolatra.

A korrelációs koefficiensek is értékelhetők döntési küszöböt és mintanagyságot figyelembe vevő kritikus értékeket tartalmazó táblázatok segítségével. Ezek a táblázatok csak a pozitív kritikus értékeket tartalmazzák. Negatív korrelációs koefficiensek eseté-

ben az abszolút értéket kell a kritikus értékhez viszonyítani (http://www.gifted.uconn.edu/siegle/research/Correlation/corrchrt.htm).

Egy nefrológiai centrumban a betegek megvizsgálták, hogy milyen kapcsolat van a betegek életkora és szisztolés vérnyomása között. Az átlagéletkor 65 év (SD = 7,83), az átlagos vérnyomás 138,4 Hgmm (SD = 33,55) volt (17. táblázat). A szórásdiagram *(15. ábra),* a kovariancia és a korrelációs koefficiens is pozitív trend meglétét jelezte:

$$C_{xy} = \frac{\sum (x-\bar{x})(y-\bar{y})}{N-1} = \frac{868}{20-1} = 45{,}68 évHgmm,$$

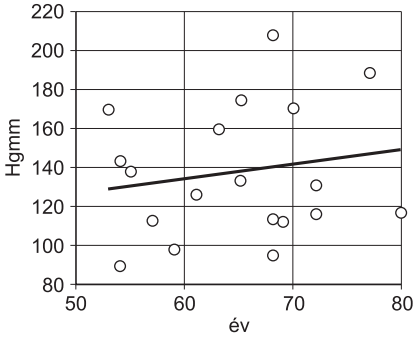
$$r = \frac{1}{(N-1)} \frac{\sum (x-\bar{x})(y-\bar{y})}{SD_x SD_y} = \frac{1}{(20-1)} \times \frac{868}{7{,}83 \times 33{,}55} = 0{,}17.$$

A 95%-os megbízhatósági tartomány azonban tartalmazta a nullát, ezért a trendet nem lehet az életkor és a szisztolés vérnyomás közti kapcsolat bizonyítékaként értelmezni.

Táblázatot használva, a 18-as szabadsági fokhoz és az 5%-os döntési küszöbhöz a kritikus érték 0,444. Ennél kisebb koefficienst számítottunk az eredményeinkre, ezért nem tekintjük elég meggyőzőnek a vizsgálatot annak bizonyítására, hogy az életkorral emelkedik a vesebetegek szisztolés vérnyomása, hiába volt ez a benyomásunk a szórásdiagram tanulmányozásakor.

$$SE_r = \sqrt{\frac{(1-r^2)}{(N-2)}} = \sqrt{\frac{(1-0{,}17^2)}{(20-2)}} = 0{,}23$$

$$MT_{95\%;r} = r \pm t_{[\alpha;N-2]} SE_r = 0{,}17 \pm 2{,}306 \times 0{,}23 = -0{,}36;0{,}71$$



15. ábra. Vesebetegek szisztolés vérnyomása az életkor függvényében

Összességében tehát a korrelációs koefficiens számítása révén meg tudjuk állapítani, hogy két változó között (1) van-e kapcsolat, (2) milyen a kapcsolat iránya, sőt meg tudjuk mondani, hogy (3) mennyire szoros a kapcsolat.

A kapcsolat erőssége a trend egyenese körüli szóródás mértékétől függ. Minél jobban szóródnak az adatok, annál kisebb a korreláció. Ez azt eredményezi többek között, hogy két, eltérő meredekségű trend esetén is hasonló lehet a korrelációs koefficiens. Vagyis, van egy olyan adat (a meredekség), ami fontos a változók kapcsolatának leírásakor, de nem szól róla a korrelációs koefficiens.

Lineáris regresszió

A korrelációs koefficiens származtatásához kiindulópontként használt szórásdiagram trendjének meredeksége számszerűsíti, hogy az *x* egységnyi változása mekkora *y* változást von maga után. Túl azon, hogy ennek a meredekségnek a vizsgálatával is válaszolni tudunk arra a kérdésre, hogy a két vizsgált változó kapcsolatban van-e, a regressziós egyenes arra is felhasználható, hogy egy tetszőleges x értékhez tartozó y értéket megadjunk anélkül, hogy azt megmérnénk. Az ilyen predikciónak nagy jelentősége van például olyan esetekben, amikor egy viszonylag könnyen mérhető paraméter szoros kapcsolatban van egy nehezen mérhetővel. Ilyenkor a vizsgálati szakaszban kell csak párhuzamosan mérni mindkét paramétert, aztán a vizsgálat végeredményeként kapott regressziós egyenletet használva már elég lesz csak az x-et mérni, és az y-t csak számolni kell.

Az x-tengelyen ábrázolt paramétert regressziós elemzés esetén magyarázó változónak, illetve független változónak nevezzük, megkülönböztetve az y-tengelyen ábrázolt függő változótól. Ennek oka, hogy a regressziós elemzés során a paraméterek közti kapcsolatnak iránya van. Nem azt vizsgáljuk, hogy két változó milyen kapcsolatban van egymással, hanem azt, hogy az egyik (magyarázó, független) változó miként befolyásolja a másik (függő) változó értékét. A kapcsolat irányát a vizsgált paraméterek közti kapcsolat feltételezett mechanizmusa alapján határozhatjuk meg. Ha az életkor és a csontok kalciumtartalma közti kapcsolatot korrelációs elemzéssel vizsgáljuk, akkor azokat a kérdéseket tudjuk megválaszolni, hogy van-e kapcsolat a két paraméter közt, pozitív vagy negatív a kapcsolat iránya, milyen szoros a paraméterek közti kapcsolat. Utóbbi alatt azt értjük, hogy milyen mértékben határozza meg az életkor a kalciumtartalmat, vagy milyen kapcsolatban van a csontok kalciumtartalma az életkorral. Számszerűen ugyanaz a korrelációs koefficiens adja meg a választ mindkét kérdésre. Ha ugyanezt a kapcsolatot lineáris regressziós elemzéssel próbáljuk értékelni, akkor nem

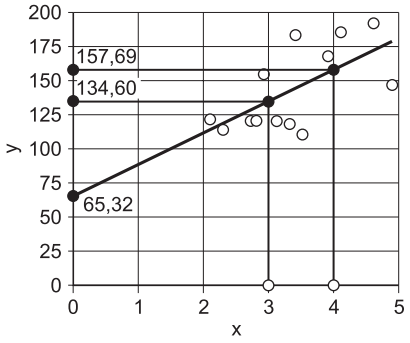
mindegy, hogy milyen irányú kapcsolatot tesztelünk. Mert az illesztett trendvonal meredeksége nem ugyanaz akkor, ha az életkor a magyarázó változó, és ezt ábrázoljuk a x-tengelyen, vagy ha a csontok kalciumtartalmát tekintjük magyarázó tényezőnek. (Természetesen az utóbbi kérdésfelvetés megalapozatlan, és csak az előbbinek van értelme.)

Az x független változó és y függő változó közti kapcsolatot szemléltető szórásdiagram pontjaira illesztett trendet leíró egyenes (regressziós egyenes) egyenlete:

$$y = bx + a,$$

ahol *b* a trendvonal meredeksége (regressziós koefficiens). Számszerűen megadja, hogy az *x* egységnyi változása esetén az *y* hogyan módosul. Mértékegységgel rendelkező mutató, aminek értéke a mértékegységek átváltásakor megváltozik. Szemben a korrelációs koefficienssel, ez nem standardizált mutató, ezért értéke nem (–1; +1) intervallumon belüli szám, hanem a mértékegységtől függően bármekkora pozitív szám lehet, emelkedő, és bármekkora negatív szám, csökkenő trend esetén. Nulla az értéke, ha nem befolyásolja a függő változó értékét a magyarázó változó.

Az *a* érték a regressziós egyenes y-tengellyel való metszéspontja, azaz az egyenlet megoldása $x = 0$ esetre. Szintén dimenzióval rendelkező paraméter, aminek konkrét értéke függ az éppen használt mértékegységtől. Nem mindig van biológiailag értelmezhető szemléletes jelentése. Ha például a naponta elszívott cigaretták számával írjuk le a dohányzás intenzitását egy dohányzás hatásait vizsgáló epidemiológiai tanulmányban (a dohányzás intenzitása az x változó), akkor az *a* metszéspont a nemdohányzókra jellemző y értéket jelenti. Ha azonban a BMI-vel mért elhízás és a vérnyomás közti

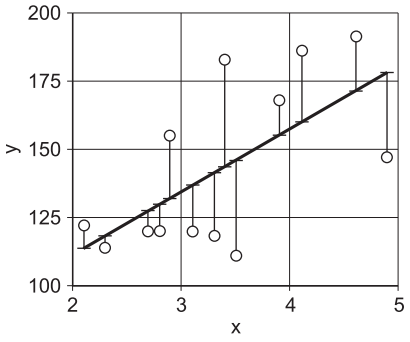


16. ábra. A regressziós egyenes egyenletének meredeksége (b = 157,69 – 134,60 = 23,09) és y tengelymetszete

kapcsolatot vizsgáljuk, akkor nyilván nincs szemléletesen értelmezési lehetősége az *a* metszéspontnak, mivel nincsenek olyan emberek, akiknek 0 a BMI-je. Utóbbi esetben csak a matematikai értelmezés lehetséges *(16. ábra)*.
A trendvonal, a regressziós egyenes, pontosabban a regressziós egyenes paramétereinek (*a; b*) meghatározásához a legkisebb négyzetek elvet használjuk *(17. ábra)*.

Azt az egyenest kell megtalálnunk, amelyik a vizsgálatban előforduló magyarázó változó értékeknél (*x*) olyan y_r-értékeket vesz fel, amelyek eltérése a tényleges függő változó értékektől (*y*) a lehető legkisebb. Amikor a devianciák ($y - y_r$) minimálisak, akkor kapjuk az adatainkra legjobban illeszkedő, tehát a kapcsolatot legpontosabban leíró egyenest, a regressziós egyenest. A devianciák egyaránt lehetnek pozitív és negatív számok, ezért ezek egyszerű összeadása során a pozitív értékek kioltanák a negatívakat, és nem kapnánk képet arról, hogy mekkora is a trendvonal és a vizsgálati eredményeket megjelenítő pontok közti távolság. Ennek elkerülésére az eltérések négyzetét összegezzük (ami mindig pozitív), vagyis az egyes megfigyelt értékek regressziós egyenestől mérhető távolságainak négyzeteit összegezzük, és keressük azt az egyenest, ami esetében ez az összeg minimális: $\sum (y - y_R)^2$. Szerencsére, az egyenes paramétereinek számításához nem kell különböző egyenesek illeszkedését elemeznünk. Egyszerűen számíthatók a négyzetes eltérések összegét minimalizáló (*a; b*) paraméterek.

A meredekség megadható a kovariancia és a magyarázó változó varianciájának hányadosával, aminek ismeretében az y-tengellyel való metszéspont azért számítható, mert kihasználjuk, hogy a tipikus vizsgálati alanyt reprezentáló ($\bar{x}; \bar{y}$) ponton mindig átmegy a regressziós egyenes:



17. ábra. Regressziós egyenes illesztése a legkisebb négyzetek elve alapján

$$b = \frac{C_{xy}}{V_x} = \frac{\frac{\sum (x-\bar{x})(y-\bar{y})}{N-1}}{\frac{\sum (x-\bar{x})^2}{N-1}},$$

$$\bar{y} = b\bar{x} + a,$$

$$a = \bar{y} - b\bar{x}.$$

Ezt követően már csak arra kell válaszolnunk, hogy a trend, a két változó közötti kapcsolat magyarázható-e a véletlennel vagy nem. (A semleges viszonyt leíró nullától szignifikánsan eltér-e a megfigyelt meredekség?)

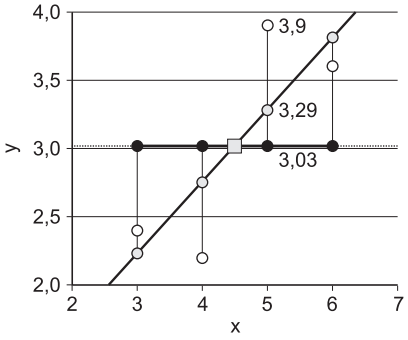
Varianciaelemzéssel tudunk válaszolni a kérdésre. Követnünk kell azt az utat, amit a tipikus ($\bar{x}; \bar{y}$) vizsgálati alanytól indulva a tényleges eredmény ($x; y$) értékpárjáig bejárunk *(18. ábra)*.

Az út első szakasza a regressziós egyenes mentén az ($x; y_R$) pontig tart.

$$y_R = a + bx$$

A második szakasz az x helyen, a trendvonal által jelzett tipikus y_R ponttól tart függőlegesen pozitív vagy negatív irányba a vizsgálat ($x; y$) eredményéig. Az y-tengely vetületében nézve a mozgás két szakasza is számszerűsíthető (minden vizsgálati alanyra, azaz minden x értékre):

$$|y - \bar{y}| = |y_R - \bar{y}| + |y - y_R|.$$



18. ábra. A lineáris regresszióval magyarázható (3,03 → 3,29) és nem magyarázható (3,29 → 3,9) variabilitás származtatása

Az útnak megfelelően, első lépésként a függő változó variabilitását két részre osztjuk. Meghatározzuk azt a részt, amit a regressziós összefüggéssel meg tudunk magyarázni, vagyis azt a változékonyságot, amit a magyarázó változótól való függése idéz elő. Ennek és a teljes varianciának a különbségeként pedig megadhatjuk a regressziós kapcsolattal nem magyarázható variabilitást. A variabilitás felosztását legegyszerűbben a négyzetes eltérések összegeinek (*SS*) segítségével tudjuk értelmezni. A függő változó teljes variabilitása (SS_y) és a regresszióval magyarázható variabilitás (SS_R) ezek alapján:

$$SS_y = \sum (y - \bar{y})^2 ,$$

$$SS_R = \sum (y_R - \bar{y})^2 .$$

A variabilitásnak az a része (SS_E), ami pedig nem magyarázható a regresszióval (a két változó közötti kapcsolattal):

$$SS_E = \sum (y - y_R)^2 .$$

Utóbbi egyszerűen különbségként is felírható:

$$SS_E = SS_y - SS_R .$$

A teljes variabilitás regresszióval magyarázható és azzal nem magyarázható részeinek aránya mutatja meg számunkra, hogy lényeges befolyással van-e a magyarázó változó a függő változó alakulására. (Ha van befolyása, akkor a trend nem látszólagos. Ebben az esetben az egyes vizsgálati eredmények értékét elég nagymértékben a magyarázó változó határozza meg, és a regressziós hatásra kialakuló érték csak tovább variálódik, a vizsgálatban nem értékelt egyéb tényezők hatására. A trend csak látszólagos, ha nincs lényeges hatása a magyarázó változónak. Ilyen esetben a függő változó teljes varianciájának túlnyomó része független lesz a regressziós kapcsolattól, a vizsgálati eredmények alapvetően olyan tényezők miatti szóródást mutatnak, amit nem a magyarázó változó, hanem más, a vizsgálatban nem értékelt faktorok hoznak létre.)

A négyzetes eltérések összegei alapján a varianciák számíthatók *(18. táblázat)*. A szabadsági fokok az egyes komponensekhez a következők: regressziós összefüggéssel magyarázott varianciára 1; teljes variancia szabadsági foka (N-1), a nem magyarázott rész szabadsági foka (N-2). Így a regresszióval magyarázható (V_R) és nem magyarázható variancia (V_E):

$$V_R = \frac{SS_R}{1}, \qquad V_E = \frac{SS_E}{N-2}.$$

Ezek segítségével a varianciák hányadosa már megadható:

$$F = \frac{SS_R}{SS_E}.$$

Ha a hányados kellően nagy szám, akkor a regresszióval magyarázható varianciarész nagy a regresszióval nem magyarázható részhez képest, vagyis (lineáris) kapcsolat van a két paraméter között. Ha ez az arány kicsi szám, akkor a regresszióval nem tudjuk lényeges részben magyarázni a teljes varianciát, azaz nincs kapcsolat a két változó között. Az F kritikus értékét, ami felett már szignifikáns a regressziós kapcsolat magyarázóereje, táblázatból lehet meghatározni, ahol a két szabadsági fok F számításának megfelelően 1 és (N-2).

A kapcsolat szignifikanciáját, a trend valódiságát vagy látszólagos jellegét a regressziós koefficiens megbízhatósági tartománya segítségével is értékelhetjük:

$$MT_{95\%;b} = b \pm t_{[\alpha;N-2]} SE_b,$$

ahol α az elsőfajú hibát, t pedig a t-eloszlás α és N-2 szabadsági foknak megfelelő értékét jelöli, és ahol a koefficiens standard hibája SE_b:

$$SE_b = \sqrt{\frac{V_E}{\sum (x-\bar{x})^2}}.$$

A megbízhatósági tartomány értékeléséhez ugyanazokat a szempontokat kell figyelembe venni, amiket a korrelációs koefficienseknél már említettünk. Ha az 5%-os megbízhatósági tartomány tartalmazza a nullát, azaz egyik része negatív, másik pozitív tartományban van, akkor a vizsgálati eredményeink alapján nem zárhatjuk ki, hogy a valódi regressziós koefficiens (ami ténylegesen leírja a magyarázó és a függő változó közti kapcsolat jellegét) negatív szám (ami szerint a magyarázó változó növekedése a függő változó csökkenését vonja maga után), vagy pozitív szám (ami szerint a magyarázó változó növekedése a függő változó növekedésével jár együtt), vagy nulla (ami szerint a magyarázó változó hatására nem változik a függő változó). Akármilyen is volt a szórásdiagramon a trend, ilyen eredmény birtokában nem állíthatjuk, hogy valóban kapcsolat van a két paraméter között. Ha viszont a megbízhatósági tartomány nem tar-

18. táblázat

Variabilitás forrása	Variabilitás (négyzetes eltérések összege, SS)	Szabadsági fokok	Variancia (átlagos négyzetes eltérés)	Variancia hányados (F)
Regressziós kapcsolattal magyarázható (R)	$SS_R = \sum (y_R - \bar{y})^2$	$df_R = 1$	$V_R = \frac{SS_R}{1}$	$F = \frac{SS_R}{SS_E}$
Regressziós kapcsolattal nem magyarázható (E)	$SS_E = SS_y - SS_R$	$df_E = N - 2$	$V_E = \frac{SS_E}{N-2}$	
Teljes (y)	$SS_y = \sum (y - \bar{y})^2$	$df_y = N - 1$		

talmazza a nullát, azaz teljes egészében vagy negatív, vagy pozitív tartományban helyezkedik el, akkor a diagram trendje valódi, irányának megfelelő módon valóban kapcsolat van a vizsgált paraméterek közt.

Miután elvégeztük a megfigyeléseinket, a regressziós egyenlet segítségével a magyarázó változó tetszőleges x_p értékéhez számíthatjuk a várható függő változó értéket, pontosabban egy legvalószínűbb y_p várható értéket:

$$y_p = a + b \times x_p .$$

Az így számított érték megbízhatóságának értékelésekor azonban figyelembe kell vennünk, hogy a regressziós egyenes meredekségét és y-tengelymetszetét csak több-kevesebb pontatlansággal tudtuk a minta vizsgálata után megállapítani. Ennek következtében minél távolabb kerülünk a vizsgálat során használt minta $\bar{x}$ átlagos magyarázó változó értékétől, annál pontatlanabbak lesznek a becsléseink. (A valódi és a minta alapján becsült regressziós egyenes a vizsgált intervallum átlagánál lesz egymáshoz a legközelebb, attól távolodva egyre távolabb futnak egymástól.)

Ráadásul nem csak a becsült meredekséggel van a probléma, hanem maga a vizsgálat során meghatározott egyenes sincs jó helyen. Ennek oka, hogy a vizsgálati eredmények feldolgozásakor a minta alapján képzett tipikus ($\bar{x}\,;\bar{y}$) vizsgálati alanyhoz kötöttük a regressziós egyenest. Mindezekhez a bizonytalanságokhoz adódik hozzá, hogy eleve csak egy befolyásoló tényező hatásait vizsgáltuk. A nem vizsgált többi faktor hatása véletlenszerű variabilitás forrásaként okozta a vizsgálati eredmények regressziós egyenes körüli szóródását, és ugyanez a hatás fokozza a jósolt érték bizonytalanságát.

Ezeket a hibákat figyelembe véve adható meg a jósolt érték, amihez a paraméterek becslését biztosító minta átlagos magyarázó értékétől távolodva, fokozatosan romlik a megbízhatósága, ami a jósolt értékhez kapcsolódó, fokozatosan bővülő, 95%-os megbízhatósági tartományokban tükröződik:

$$MT_{95\%;y_p} = y_p \pm t_{[\alpha;N-2]} \times \sqrt{\frac{SS_y - SS_R}{(N-2)}} \times \sqrt{1 + \frac{1}{N} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}}\ .$$

Jósolt értéket csak a paraméterek megállapítására használt minta magyarázó értékeinek tartományában lehet számítani. Az eredeti tartományon kívülre extrapolálni nagyon kockázatos. Egyáltalán nem biztos, hogy a vizsgált tartományon kívül ugyanolyan a paraméterek közti kapcsolat, mint a vizsgált tartományon belül. (Számos példa van arra, hogy különböző dózistartományokban ugyanaz a hatás más intenzitással idéz elő biológiai válaszokat.) Sőt, az is előfordulhat, hogy a vizsgált tartományon kívül értelmetlen már maga a magyarázó változó is. (Ha a testmagasság és a testsúly közti kapcsolatot vizsgáljuk, akkor szignifikáns trendet látunk. De a matematikai lehetőség ellenére nyilvánvalóan értelmetlen a 10 m magas emberek várható testtömegét számítani.)

19. táblázat

Edzések száma (x)	Fittségi teszt (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
27	122	–5,69	–20,77	118,22
30	155	–2,69	12,23	–32,93
39	192	6,31	49,23	310,53
38	186	5,31	43,23	229,46
27	114	–5,69	–28,77	163,76
28	120	–4,69	–22,77	106,84
39	147	6,31	4,23	26,69
27	120	–5,69	–22,77	129,61
33	118	0,31	–24,77	–7,62
31	120	–1,69	–22,77	38,53
36	168	3,31	25,23	83,46
35	183	2,31	40,23	92,84
35	111	2,31	–31,77	–73,31

Egy 13 fős labdarúgó csapatnál értékelik a tavaszi szezonban végzett munkát. Többek között elemzik, hogy a teljesített edzések száma milyen kapcsolatban volt egy fittségi teszten elért pontszámmal. Az átlagos edzésszám 32,69 (SD = 4,64), az átlagos fittségi teljesítmény 142,78 (SD = 30,49) pont volt *(19. táblázat)*.

A lineáris trendvonal paramétereit számították *(19. ábra)*.

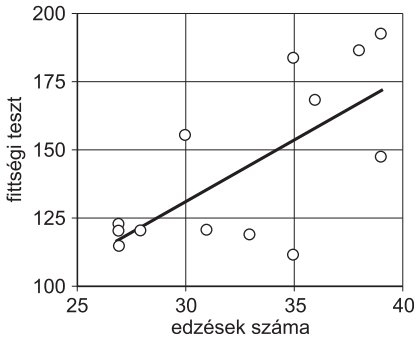
A kovariancia 98,84 volt. Meghatározták a trend szignifikanciáját.

$$b = \frac{C_{xy}}{V_x} = \frac{98{,}84}{4{,}64^2} = 4{,}58$$

$$a = \bar{y} - b\bar{x} = 142{,}78 - 4{,}58 \times 32{,}69 = -7{,}07$$

$$SE_b = \sqrt{\frac{V_E}{\sum (x - \bar{x})^2}} = \sqrt{\frac{519{,}62}{258{,}77}} = 1{,}42$$

$$MT_{95\%;b} = b \pm t_{[\alpha;N-2]} SE_b = 4{,}58 \pm 2{,}20 \times 1{,}42 = 1{,}46;7{,}70$$



19. ábra. Lineáris regressziós kapcsolat az edzések száma és a fittségi teszten elért teljesítmény között

20. táblázat

Variabilitás forrása	Variabilitás (négyzetes eltérések összege)	Szabadsági fokok	Variancia (átlagos négyzetes eltérés)	Variancia hányados
Regressziós kapcsolattal magyarázható	5436,42	1	5436,42	10,46
Regressziós kapcsolattal nem magyarázható	5715,89	11	519,62	
Teljes	11 152,3	12		

21. táblázat

Sportoló neve	Edzések száma	Jósolt fittségi teszt (95%-os megbízhatósági tartomány)	
TH	29	126	(73; 179)
ZV	34	149	(97; 201)
ESz	37	163	(109; 216)

A regressziós koefficiens standard hibája 1,42 (95%-os megbízhatósági tartománya 1,46–7,70) volt *(20. táblázat)*.

Ezek alapján megállapították, hogy: (1) a diagram alapján úgy tűnik, a több edzéssel magasabb teljesítmény érhető el; (2) a megbízhatósági tartomány teljesen a pozitív tartományban volt, emiatt az emelkedő trend véletlennel nem magyarázható, tényleges kapcsolat van az edzések száma és a teljesítmény között; (3) minden edzés 4,58 pontos teljesítményjavulást eredményezett. Annak a 3 csapattagnak, akik nem tudtak részt venni a fittségi teszten, de akik év közben rendszeresen látogatták az edzést, megjósolták a fittségi pontszámukat *(21. táblázat)*.

Determinációs koefficiens

A trend valódiságának vagy látszólagos természetének értékelésekor használt varianciaelemzés alkalmával definiáltuk a teljes variabilitás leírásához SS_y-t, melynek regresszióval magyarázható része SS_R volt. Az SS_R/SS_y hányadossal tudjuk leírni, hogy az y függő változó teljes varianciájának hányadrészét képes megmagyarázni az x magyarázó változó a köztük lévő kapcsolat következtében, ezért ezt definiáljuk determinációs koefficiensként. (Az SS_E/SS_y hányados pedig a magyarázó változótól független hányadát adja a teljes függő változó variabilitásnak.) Az SS_R/SS_y hányados a korrelációs koefficiens négyzetével egyenlő:

$$r^2 = \frac{SS_R}{SS_y}.$$

A sportolók adatainak elemzésekor azt találták, hogy a fittségi index 49%-ban az edzéseken való részvétel gyakoriságával magyarázható:

$$r^2 = \frac{SS_R}{SS_y} = \frac{5436{,}42}{11152{,}3} = 0{,}49.$$

Standardizált regressziós koefficiensek

A regressziós egyenes meghatározásával jól jellemezhetünk egy egyszerű kapcsolatot. A valóságban azonban nagyon ritkán fordul elő, hogy egy függő változó értékét alapvetően egyetlen független változó határozza meg. A biológiai jelenségek sokkal bonyolultabb kölcsönhatások eredményeként alakulnak ki. Ezért, ha valós problémákat akarunk eredményesen vizsgálni, akkor az egyváltozós elemzések helyett olyan módszerre van szükségünk, amellyel egyszerre több determináns hatását is vizsgálhatjuk.

Ezt elvégezhetnénk úgy, hogy az összes determinánsra külön-külön kiszámítjuk a regressziós koefficienseket. Ezek azonban dimenzióval rendelkező számok, amik közvetlenül nem hasonlíthatók össze. (Az egyéves életkor-növekedésre eső vérnyomás-emelkedés nem hasonlítható össze az egységnyi testtömegindex-növekedésre eső vérnyomás-emelkedés mértékével. A regressziós koefficiensek értéke nem segíti annak megítélését, hogy melyik faktor bír nagyobb befolyással a vérnyomás meghatározásában.) Emiatt az egyes független változók relatív súlyát ilyen módon nem tudjuk meghatározni.

Egymással közvetlenül összehasonlítható, dimenzió nélküli, regressziós koefficienseket úgy kaphatunk, ha standardizáljuk a vizsgálati eredményeinket. Ilyenkor az éppen használatos mértékegységek elvesztik jelentőségüket, és a standardizált regressziós koefficiensek (β) már közvetlenül összehasonlíthatók: a nagyobb standardizált regressziós koefficiens erősebb hatást jelez. A standardizálást utólag is elvégezhetjük:

$$\beta = b \times \frac{SD_x}{SD_y}.$$

Ezzel a megközelítéssel mindaddig semmi gondunk nem is lesz, amíg az egyes magyarázó változók egymástól függetlenek. Amikor viszont ezek egymással is kapcsolatban vannak (nagyon gyakran ez a helyzet), akkor az adott független változókra számított regressziós koefficiens nem csak az adott magyarázó változó saját hatását fogja tartalmazni, hanem az egyéb determinánsok rajta keresztül kifejtett hatásait is. Ilyen esetben a standardizálás helyett olyan eljárásra van szükségünk, amely a determinánsok önálló hatását számszerűsíti. (Ez az eljárás a többváltozós lineáris regressziós elemzés.)

Nem paraméteres próbák

A kvantitatív adatok elemzésekor korábban olyan elemzési technikákat ismertünk meg, amelyek az adatok normális eloszlását feltételezve a normális eloszlás átlagát, szórását (eloszlás-paramétereit) elemezték valamilyen gondolatmenet mentén. A normalitást azonban kifejezetten azért kell ellenőrizni, mert nem minden esetben érvényesül. Ha nem normális eloszlásúak az adataink (például kiugró értékek vagy aszimmetrikus eloszlás miatt), akkor nem lehet alkalmazni t-próbát, varianciaelemzést, korrelációs koefficiens és regressziós együttható számítást. Ezekben az esetekben olyan statisztikai eljárásokat alkalmazunk, melyek nem feltételezik a normális eloszlást, és amelyek menete nem igényli eloszlás-paraméterek számítását. Ezek a nem paraméteres eljárások. Alkalmazásuk egyszerűségének az ára a kisebb hatékonyság. Az alapvető paraméteres próbák nem paraméteres alternatívái általában könnyen alkalmazhatók, de a regressziós elemzések nem helyettesíthetők nem paraméteres módszerekkel (hiszen ennél a statisztikai eszköznél éppen az egységnyi független változó növekedéshez kapcsolódó függő változó változásának megállapítása van a fókuszban.)

Alapszabályként elmondható, hogy ha lehetőség van rá, akkor törekedni kell olyan vizsgálattervezésre, ami a normalitás kritériumának megfelelő adatokat eredményez. Ha az adataink nem normális eloszlásúak mégsem, akkor meg kell próbálni valamilyen transzformáció segítségével normalizálni őket.

Előjelteszt

Ha párba rendezett kvantitatív adataink vannak, melyek kezelés előtti és utáni állapotot, vagy zavaró tényezők szempontjából hasonló két személy reakcióit vizsgálják, melyekre nem teljesül a normalitás feltétele, akkor legegyszerűbb, ha a párokon belül az adatok különbségének irányát elemezzük.

Ha például a kezelés előtt (H) és a kezelés után (U) megfigyelt adatpárokat nézzük, akkor megállapíthatjuk a kezelés hatására javuló (N_U), illetve romló (N_E) állapotú betegek számát. Természetesen lehetnek betegek, akik nem számolnak be változásról – őket a mostani elemzésben figyelmen kívül is hagyjuk!

Amennyiben a kezelés valójában nem befolyásolta a beteg állapotát, akkor a javuló és romló állapotú betegek száma hasonló lesz. Ha a kezelés javítja a kimenetelt, akkor

az N_U meggyőző mértékben meghaladja az N_E-t. A meggyőző mérték megadásához, azaz a kritikus statisztikai mérőszám kiszámításához ebben az esetben is a χ^2 számítását hívjuk segítségül, ahol a várható érték természetesen $(N_U + N_E)/2$ (tükrözve azt az állapotot, amikor nincs semmilyen hatása a kezelésnek, és ezért a javuló és romló betegek száma egyenlő), a szabadsági fok pedig 1:

$$\chi^2 = \frac{N_U - \frac{N_U + N_E}{2}^2}{\frac{N_U + N_E}{2}} + \frac{N_E - \frac{N_U + N_E}{2}^2}{\frac{N_U + N_E}{2}}.$$

Yates-korrekcióval érdemes ezt a számítást is kiegészíteni a konzervatív szemlélet érvényesítése érdekében:

$$\chi^2 = \frac{\left|N_U - \frac{N_U + N_E}{2}\right| - 0{,}5^2}{\frac{N_U + N_E}{2}} + \frac{\left|N_E - \frac{N_U + N_E}{2}\right| - 0{,}5^2}{\frac{N_U + N_E}{2}}.$$

Kritikus értéknél nagyobb teszteredmény arra utal, hogy jelentős az eltérés a javuló és romló állapotú betegek száma között. Számítógép segítségével pedig meg tudjuk határozni, hogy 1-es szabadsági fok esetén a számított χ^2 milyen döntési küszöb esetén felel meg a kritikus értéknek (azaz kiszámíthatjuk a p-értéket).

12 beteget kezelnek, és egy életminőség-skálán mérik a kezelés hatását *(22. táblázat)*. A kezelés előtti és utáni értékek nem normális eloszlásúak. 8 beteg javulásról, 4 pedig az életminősége romlásáról számolt be. A számított χ^2 kisebb, mint a kritikus $\chi^2_{df:1} = 3{,}841$.

Nem értékelhetjük a vizsgálati eredményt úgy, hogy a beavatkozás javítja a betegek életminőségét, és a vizsgálati eredmény nem ad alapot arra, hogy a vizsgált eljárást alkalmazzák.

$$\chi^2 = \frac{\left|8 - \frac{8+4}{2}\right| - 0{,}5^2}{\frac{8+4}{2}} + \frac{\left|4 - \frac{8+4}{2}\right| - 0{,}5^2}{\frac{8+4}{2}} = 0{,}750.$$

Az előjelteszt inkább didaktikai szempontból érdekes. Egyszerűsége, ami a számítási igény és az eredmény (χ^2) értelmezése szempontjából is egyértelmű, barátságossá teszi a módszert. Gyakorlati alkalmazására ma már csak ritkán kerül sor. Ennek oka,

22. táblázat

Beavatkozás előtti indikátor (E)	Beavatkozás utáni indikátor (U)	Változás mértéke (U–E)	Vátozás előjele
2,11	3,68	1,57	+
2,12	3,63	1,51	+
2,14	4,05	1,91	+
2,83	3,56	0,73	+
2,85	2,38	–0,47	–
3,40	3,21	–0,19	–
3,88	3,35	–0,53	–
6,66	5,95	–0,71	–
6,87	9,08	2,21	+
7,96	10,17	2,21	+
8,75	20,63	11,88	+
13,95	16,45	2,50	+

hogy a teszt egyáltalán nem veszi figyelembe, hogy milyen nagyok voltak azok a különbségek, amelyek alapján a párokon belüli különbségek irányát meghatároztuk. Nyilvánvalóan komoly információ vész el ezzel a megoldással. Ráadásul olyan információ, amelyet az adatgyűjtés során egyszer már megszereztünk. A vizsgálatok statisztikai támogatása pedig nem szólhat az általános vizsgálati hatékonyság csökkentéséről. A mai számítástechnikai feltételek mellett pedig a számítás egyszerűsége és a végeredmény értelmezésének egyszerűsége nem lehet szempont a módszerek megválasztásakor.

Wilcoxon párosított teszt

A párba rendezett, nem normális eloszlású kvantitatív adatok értékelésekor alkalmazható eljárás a Wilcoxon párosított teszt, ami a párok közti eltérésnek nem csak az irányát veszi figyelembe, hanem a párok közt meglevő különbség nagyságát is. Mivel a vizsgálati eredmények nem normális eloszlásúak, ezért a megfigyelt adatok közti különbség nagysága helyett (amely t-próba esetén volna a statisztikai számítás kiindulási pontja) azok egymáshoz viszonyított sorrendjét (rangját) használjuk. Ennél a pontnál nyilván-

valóan információt vesztünk a tényleges különbségek nagyságához viszonyítva, de a veszteség nem olyan jelentős, mint az előjelteszt esetén.

A vizsgálat során az adatpárok, az E- és U-csoporthoz tartozó vizsgálati alanyok adatai közti különbséget számítjuk, a különbség irányát figyelmen kívül hagyva, azaz az eltérés abszolút értékét határozzuk meg. Az abszolút értékek alapján növekvő sorrendbe, rangsorba állítjuk a párokat. A legkisebb különbséget mutató pár lesz az első a sorban, ez a pár kapja az 1-es rangot. Előfordul, hogy több abszolút különbség is számszerűen azonos. Ilyenkor a rangok átlagát kapja mindegyik érték. Pl. ha az 5. és 6. adatpár azonos abszolút különbséget mutatott, akkor mindkét párt 5,5-es ranggal [(5 + 6)/2 = 5,5] vesszük majd figyelembe a további számítások során. Ha a 8–10. helyeken kaptunk azonos értékeket, akkor mindhárom pár a 9-es rangot [(8 + 9 + 10)/3 = 9] fogja kapni. Ha a vizsgálat során olyan adatpárt is találunk, ahol a két érték egyforma, akkor ezeket a párokat a feldolgozásból ki kell zárni.

Az így számított rangok előjelet kapnak: az eredeti különbségek előjelét. Ha egy adatpárnál az E-adat nagyobb volt, mint az U, akkor a pozitív előjelű eltérés miatt a rang előjele is pozitív lesz. Ha az E-érték volt a kisebb, akkor az eltérés negatív előjelét kapja az adatpár rangszáma is. Ha a két csoport közt nem volt lényeges eltérés (például nem volt hatékony az alkalmazott kezelés), akkor a különbségek véletlenszerűen oszlanak meg, emiatt a különbségek negatív és a pozitív előjelű rangszámainak összege közel azonos lesz. Ha az előjeles rangokat összeadjuk, akkor 0 körüli értéket kapunk. Ha a két csoport közt lényeges különbség van (például a kezelés hatásos volt), akkor az adatpárok különbségei eltolódnak pozitív irányba, több lesz a pozitív előjelű rangszám, illetve nagyobb lesz a pozitív rangszámok összege. Ha kellően nagy az eltérés a pozitív és negatív előjelű rangszámok összege közt, akkor tekintjük a véletlennel már nem magyarázhatónak a két csoport közti különbséget. Vagyis eljutottunk ismét ahhoz a ponthoz, ahol valamilyen statisztikai mérőszám segítségével számszerűsítenünk kell egy eltérést, és ennek az eltérésnek a kritikus értékét is meg kell tudnunk határozni.

A különböző előjelű rangszámok összege közti különbség természetesen nem csak az adatpárok közti különbségektől függ, hanem attól is, hogy hány adatpárt elemeztünk. Minél nagyobb elemszámú a vizsgálat, annál nagyobb rangszámösszeget kapunk. A vizsgálat méretét a minta elemszámával adhatjuk meg. Ennek (és természetesen a döntési küszöbnek) a figyelembe vételével értékeljük majd a teszt eredményét.

A teszt eredménye Wilcoxon-tesztnél a kisebbik rangszámösszeg. (A pozitív és negatív előjelű rangszámösszegek közül a kisebbik abszolút értékűt választjuk ki.) Ha a két csoport közt lényeges a különbség, akkor a különbségek eltolódása miatt az egyik oldalon a sok rangszám nagy rangszámösszeget eredményez, a másik oldalon viszont

23. táblázat

Beavatkozás előtti érték (E)	Beavatkozás utáni érték (U)	Változás mértéke (U–E)	Változás előjele	Változás abszolút értéke |U–E|	Rang-szám	Előjeles rangszám	Pozitív eltérések rangszámai	Negatív eltérések rangszámai
2,11	3,68	1,57	+1	1,57	7	7	7	
2,12	3,63	1,51	+1	1,51	6	6	6	
2,14	4,05	1,91	+1	1,91	8	8	8	
2,83	3,56	0,73	+1	0,73	5	5	5	
2,85	2,38	–0,47	–1	0,47	2	–2		2
3,4	3,21	–0,19	–1	0,19	1	–1		1
3,88	3,35	–0,53	–1	0,53	3	–3		3
6,66	5,95	–0,71	–1	0,71	4	–4		4
6,87	9,08	2,21	+1	2,21	9,5	9,5	9,5	
7,96	10,17	2,21	+1	2,21	9,5	9,5	9,5	
8,75	20,63	11,88	+1	11,88	12	12	12	
13,95	16,45	2,5	+1	2,5	11	11	11	

kevés különbség lesz (ritkán fordul elő, hogy az általában hatásos kezelés alkalmazása után egyes betegek állapota romlik), abszolút értékét tekintve kicsi rangszámösszeget eredményez. Minél kisebb ez a kisebbik rangszámösszeg, annál biztosabbak lehetünk a csoportok közti eltérés szignifikanciájában. A kritikus értékeket táblázatokból tudjuk megkeresni (http://www.sussex.ac.uk/Users/grahamh/RM1web/WilcoxonTable2005.pdf), vagy számítógép segítségével tudjuk számítani.

Az előjelteszt példájának adatait felhasználva a pozitív rangszámok összege 68, a negatív rangszámoké 10. Összesen 12 adatpárt vizsgáltunk. A kritikus érték 12 adatpárnál és 5%-os döntési küszöbnél 14. A különbség tehát elég nagy a két csoport közt ahhoz, hogy meggyőzzön minket arról, hogy a beavatkozás hatásos volt *(23. táblázat)*.

Mann–Whitney U-teszt

Abban az esetben, ha két független csoport folytonos változóinak különbségét szeretnénk értékelni, de az adatok nem normális eloszlásúak, vagy a csoportokon belüli variabilitás jelentősen eltér, akkor az adataink eloszlására az átlagérték és a szórás nem használható. Helyette a medián segítségével tudjuk leírni az adatok centrális értékét, és

rangszámok segítségével az egyes adatok elhelyezkedését. A különbség értékelésére a Mann–Whitney U-tesztet alkalmazhatjuk.

A teszt a két vizsgálati csoport (*A, B*) adatainak összevonásával és az egyesített adatsoron belüli rangsorolásával kezdődik. (A legkisebb érték kapja az 1-es sorszámot.) Az egyesített mintában kapott sorszámokat a két csoportban külön-külön összegezve láthatjuk, hogy melyik csoportban fordulnak elő jellemzőbben a magas, és melyikben inkább az alacsony rangszámok. Ha nincs lényeges különbség a két csoport közt, akkor a rangsorban is véletlenszerűen szóródnak a csoportok adatai, és a csoportonkénti rangszámok összege (T_A, T_B) is hasonló lesz. Ha jelentős a két csoport közti különbség, akkor az egyik csoport adatai eltolódnak a kicsi sorszámok felé, a másik csoport adatai pedig inkább a nagyobb sorszámok irányába. Emiatt lesz nagy a két rangszámösszeg közti eltérés. Ha ez a különbség kellően nagy, akkor értékeljük szignifikánsnak, véletlennel nem magyarázhatónak a csoportok közti különbséget.

A statisztikai mérőszám önmagában nem lehet a rangszámösszegek különbsége, mert ez az érték nem független a vizsgálat elemszámától. Minél nagyobb az egyes csoportok létszáma, annál nagyobbak a rangszámok összegei is, emiatt tendenciaszerűen nagyobb lesz a két rangszámösszeg különbsége.

Az elvileg lehetséges minimális rangszámösszeggel korrigált rangszámösszeget kell helyette alkalmaznunk. (Ha az egyik csoportban kapott adatok mind kisebbek, mint a másik csoportban, akkor az első csoport tagjai kapják 1-től kezdve folyamatosan a legkisebb sorszámokat. Ebben az extrém helyzetben 1, 2, 3, 4, …, n lesz a csoport tagjainak sorszáma. A rangszámok összege az n elemű számtani sorok összege, azaz n(n-1)/2 lesz.) Az ezzel a korrekcióval kapott U-értékek minimuma 0, azaz a számított U-értékek mindegyikét ugyanahhoz a kiindulási ponthoz tudjuk viszonyítani.

$$U_A = T_A - \frac{n_A(n_A+1)}{2} \qquad U_B = T_B - \frac{n_B(n_B+1)}{2}.$$

U-értéket kapunk mindkét csoport esetében. A statisztikai mérőszám végül csak a kisebbik U-t hasznosítja, a nagyobbikat figyelmen kívül hagyjuk. Ennek az elemszámokra és a döntési küszöbre vonatkozó kritikus értékét táblázatokból tudjuk megkeresni. (http://www.lesn.appstate.edu/olson/stat_directory/Statistical%20procedures/Mann_Whitney%20U%20Test/Mann-Whitney%20Table.pdf)

Ha az U kisebb, mint a táblázatban szereplő határérték, akkor tekinthetjük a két csoport közti különbséget szignifikánsnak. Természetesen egy sor statisztikai számítógépes program segít a számítások kivitelezésében és a teszt eredményének értékelésében.

Cukorbetegek HgbA1c-szintjeit vizsgálják két terápiás protokoll (A, B) alkalmazása mellett *(24. táblázat)*. A 18 A és 17 B protokollal kezelt beteg mediánértéke 7,87%, illetve 7,72% volt. A rangszámok összege a B-csoportban volt magasabb: TA = 261, TB = 369. Az A-csoport U-értéke alapján lehetett a csoportok közti különbség szignifikanciáját értékelni:

$$U_A = 261 - \frac{18(18+1)}{2} = 90, \qquad U_B = 369 - \frac{17(17+1)}{2} = 216.$$

Mivel 5%-os döntési küszöb mellett és 18-as, illetve 17-es csoportlétszámoknál a kritikus érték 93, aminél kisebb az U_A, a két csoport közt véletlennel nem magyarázható a különbség. Lényegesen alacsonyabbak a HgbA1c-eredmények az A-protokoll szerint gondozott betegek közt.

Érdemes megemlíteni, hogy a Mann–Whitney U-teszt nem a mediánok különbségeit vizsgálja, amint azt eseteként mondani szokták. A fenti példa is jól szemlélteti, hogy lényegében azonos mediánok mellett is lehet lényegesen eltérő az adatok eloszlása. (Akkor szokták a mediánok közti különbség teszteléseként ismertetni a Mann–Whitney U-tesztet, amikor a kétmintás t-próbával való hasonlóságot magyarázzák, ami az átlagok közti különbséget teszteli. Mint minden hasonlat, ez is csak egy bizonyos pontig segíti a megértést a tanulás idején. A tanulás célja viszont végső soron a statisztikai eljárás gondolatmenetének tisztánlátása.)

Kruskal–Wallis H-teszt

Amennyiben kettőnél több csoportba rendezett kvantitatív adatainak elemzésével szeretnénk egy befolyásoló tényező hatására vonatkozóan megállapítást tenni, de az adatok nem normális eloszlásúak, és emiatt az egyszempontos varianciaelemzés nem használható, akkor a nem paraméteres alternatívát a Kruskal–Wallis H-teszt jelenti. Az eloszlás módjára tekintet nélkül az egyes mérési eredmények sorba rendezése révén nyert rangsor segítségével értékelhetjük a csoportok közti eltérést. Az eredmények közti tényleges különbség nagyságára vonatkozó információt emiatt elveszítjük, de ilyen módon elkerüljük az a csapdát, amit az alkalmazási feltételek teljesülése nélkül alkalmazott paraméteres eljárás alkalmazása során nyert eredmények állítanak! Sokszor ez a nem paraméteres teszt jelenti a leghatékonyabb megoldást.

A vizsgálat nullhipotézise, hogy a vizsgálati csoportok ugyanabból a populációból származnak, nincs köztük a vizsgált jelleg szempontjából lényeges különbség. (Nem a

24. táblázat

Csoport	Eredmény	Rang	Rangszámösszeg-A	Rangszámösszeg-B
A	6,09	1	1	
A	6,41	2	2	
A	6,43	3	3	
A	6,60	4	4	
A	6,63	5	5	
A	6,77	6	6	
A	6,79	7	7	
A	6,91	8	8	
B	7,08	9		9
B	7,10	10		10
B	7,12	11		11
B	7,33	12		12
B	7,42	13		13
B	7,44	14		14
B	7,70	15		15
B	7,71	16		16
B	7,72	17		17
A	7,80	18	18	
A	7,94	19	19	
A	8,19	20	20	
A	8,37	21	21	
A	8,43	22	22	
A	8,67	23	23	
A	9,07	24	24	
A	9,08	25	25	
A	9,16	26	26	
A	9,17	27	27	
B	9,20	28		28
B	9,46	29		29
B	9,53	30		30
B	9,62	31		31
B	9,62	32		32
B	9,73	33		33
B	9,74	34		34
B	9,91	35		35

csoportok átlagai közti különbség természetéről szól a vizsgálat, de még csak nem is a mediánok közti eltérésekről!)

A nullhipotézist követve az *N* elemű csoportokat egyesítjük, és az így nyert mintán belül sorba rendezzük az eredményeket. A legkisebb eredmény kapja az 1. sorszámot. A rangszámokat aztán csoportonként összegezzük. A csoportonkénti rangszámösszegek (*T*) felhasználásával pedig számíthatjuk *H*-t:

$$H = \frac{12}{N(N+1)} \left(\frac{T_A^2}{N_A} + \frac{T_B^2}{N_B} + \ldots + \frac{T_k^2}{N_k} \right) - 3(N+1).$$

A képlet szemléletes értelmezése talán nem a legegyszerűbb feladat, de azért arra fel kell hívni a figyelmet, hogy az egyes csoportok létszámához viszonyított négyzetes rangszámösszegek határozzák meg adott vizsgálatnagyság mellett a statisztikai mérőszám értékét. Ha feltételezzük, hogy minden részcsoport egyforma, akkor könnyű belátni, hogy a

$$\frac{T_A^2}{N_A} + \frac{T_B^2}{N_B} + \ldots + \frac{T_k^2}{N_k}$$

kifejezés akkor minimális, ha a rangszámösszegek egyformák, azaz, ha teljesen véletlenszerűen oszlanak meg az egyes mérési eredmények az egyesített adathalmaz rangsorában. Minél egyenetlenebb a csoportonkénti adatok eloszlása az egyesített rangsorban, annál nagyobb az eltérés a csoportonkénti rangszámösszegekben, és ez a négyzetre emelés miatt összességében a kifejezés értékének az emelkedését vonja maga után. Ha kellően nagy a számított H, akkor a különbséget már nem magyarázhatja véletlen.

Ha 3 csoportba soroltuk az adatainkat (*A, B, C*), és mindegyik csoport ugyanakkora ($N_A = N_B = N_C = n$), akkor az említett kifejezés egyszerűsödik:

$$\frac{T_A^2}{N_A} + \frac{T_B^2}{N_B} + \ldots + \frac{T_k^2}{N_k} = \frac{T_A^2 + T_B^2 + T_C^2}{n}.$$

Ha a csoportok közt nincs különbség, és az eredmények véletlenszerűen oszlanak el az egyesített rangsorban, akkor az azonos elemszámú csoportok mindegyikében ugyanannyi lesz a rangszámok összege ($T_A = T_B = T_C = T$). A kifejezés tovább egyszerűsíthető:

$$\frac{T_A^2}{N_A} + \frac{T_B^2}{N_B} + \ldots + \frac{T_k^2}{N_k} = \frac{T_A^2 + T_B^2 + T_C^2}{n} = \frac{3T^2}{n}.$$

Ha a csoportok között van szignifikáns eltérés, akkor a rangszámok összege nem változik, de egyes csoportok rangszámösszege csökken, a másik csoporté pedig növekszik.

Tehát a rangszámösszegek (feltételezve, hogy az egyik csoport rangszámösszege nem változik) felírhatók az alábbi formában:

$$T_A = (T - x); T_B = T; T_C = (T + x),$$

és emiatt a vizsgált kifejezés is egyszerűsödik:

$$\frac{T_A^2 + T_B^2 + T_C^2}{n} = \frac{(T-x)^2 + T^2 + (T+x)^2}{n} = \frac{T^2 - 2Tx + x^2 + T^2 + T^2 + 2Tx + x^2}{n} = \frac{3T^2 + 2x^2}{n} = \frac{3T^2}{n} + \frac{2x^2}{n}.$$

Jól látható, hogy minél egyenetlenebb a csoportok egyesített rangsoron belüli helyzete, és ennek következtében minél jobban eltér a csoportok rangszámösszege egymástól (minél nagyobb *x* értéke), annál nagyobb lesz az egyes csoportok rangszámösszeg négyzeteinek összege.

Összességében a nagyobb H-érték szól a csoportok közti különbségek mellett. Az értékeléshez használható küszöbérték a csoportok számánál eggyel kisebb szabadsági fokú χ^2-tel kellő pontossággal megadható (ha minden csoportban legalább 5 volt a vizsgált elemek száma). Kisebb elemszámoknál részben táblázatok (http://www.watpon.com/table/kruskalwallis.pdf), részben a statisztikai szoftverek segítségével tudunk statisztikai következtetést levonni.

Ezt a képletet addig lehet használni, amíg a rangszámok kiosztásakor nem találunk egyenlő eredményeket. Ilyenkor a rangszámok átlagát kell az egyes adatok rangszámaként képezni (ami ezért lehet törtérték is!) Az ilyen módon kapcsolt rangszámok miatt azonban a H-értéket is korrigálni kell. Ha egy érték *t* esetben fordul elő, akkor a *K* korrekciós tag

$$K = t^3 - t$$

módon számítható. Minden kapcsolt rangszámra külön ki kell számítani ezeket a korrekciós tagokat. A korrigált H_K ezek után már számítható:

$$H_K = \frac{H}{1 - \frac{\sum K}{N^3 - N}}.$$

Ha a vizsgálat végeredménye szerint szignifikáns különbség van a csoportok közt, akkor ennek pontosabb értelmezéséhez a két-két csoportra kiterjedő Mann–Whitney U-teszt ad segítséget. A részletes elemzés során a többszörös hipotézistesztelés miatt

csökkenteni kell az egyes részelemzések döntési küszöbét. Ha *k* vizsgálati csoportunk volt, akkor a lehetséges párok száma $\frac{k(k-1)}{2}$, az egyedi részelemzések döntési küszöbe pedig $\frac{0{,}05}{\frac{k(k-1)}{2}} = \frac{0{,}1}{k(k-1)}$ lesz.

Zaj hatására romló reakcióidők tanulmányozásakor 30 önkéntessel oldatnak meg ügyességi feladatot különböző zajszinten (A-csoport: nincs zavaró hang; B-csoport: kisforgalmú közúti zajszint; C-csoport: nagy forgalmú közút zajszintje). A reakcióidőket a *25. táblázat* tartalmazza (ms).

Az egyesített mintában kiosztott rangszámok csoportonkénti összege 100,5 az A-csoportban, 150,5 a B-csoportban és 214 a C-csoportban. A 71,1 ms-os reakcióidő két alkalommal fordult elő, a 93,4 ms-os pedig 3 alkalommal. *A 26. táblázat* már a kapcsolt rangszámokat tartalmazza.

25. táblázat

A	B	C
60,3	70,7	93,4
71,1	95,3	75,8
52	93,4	109
42,4	58,5	90,7
49,6	77,2	99,5
63,5	76,3	84
53,5	71,1	97,4
93,4	89,9	68
117,2	86,8	92,8
75,4	66,4	105,4

26. táblázat

Csoport	Eredmény (ms)	Rang	Rang–A	Rang–B	Rang–C
A	42,4	1	1	0	0
A	49,6	2	2	0	0
A	52	3	3	0	0
A	53,5	4	4	0	0
B	58,5	5	0	5	0
A	60,3	6	6	0	0
A	63,5	7	7	0	0
B	66,4	8	0	8	0
C	68	9	0	0	9
B	70,7	10	0	10	0
A	*71,1*	*11,5*	11,5	0	0
B	*71,1*	*11,5*	0	11,5	0
A	75,4	13	13	0	0
C	75,8	14	0	0	14
B	76,3	15	0	15	0
B	77,2	16	0	16	0
C	84	17	0	0	17
B	86,8	18	0	18	0
B	89,9	19	0	19	0
C	90,7	20	0	0	20
C	92,8	21	0	0	21
A	*93,4*	*23*	23	0	0
B	*93,4*	*23*	0	23	0
C	*93,4*	*23*	0	0	23
B	95,3	25	0	25	0
C	97,4	26	0	0	26
C	99,5	27	0	0	27
C	105,4	28	0	0	28
C	109	29	0	0	29
A	117,2	30	30	0	0

A korrekció nélkül számított H érték:

$$H = \frac{12}{30(30-1)}\left(\frac{100{,}5^2}{10}+\frac{150{,}5^2}{10}+\frac{214^2}{10}\right) - 3(30+1) = 8{,}350.$$

A korrekciós tag a két 11,5 ms-os ($t_1 = 2$) és a három 93,4 ms-os ($t_2 = 3$) kapcsolt rang miatt:

$$K = (2^3 - 2) + (3^3 - 3) = 30.$$

A korrigált H_K pedig:

$$H_K = \frac{8{,}350}{1-\frac{30}{30^3-30}} = 8{,}359.$$

Mivel a csoportok száma 3 volt, a 2-es szabadsági fokú $\chi^2 = 5{,}991$ a kritikus értéke a teszt eredményének. A teszt-statisztika értéke nagyobb, mint a kritikus érték, azaz a csoportok közti rangeloszlás véletlennel nem magyarázható, a zajnak van hatása a reakcióidőre.

Spearman-rangkorreláció

Folytonos változók közti kapcsolat erősségének vizsgálatára alapvetően a korrelációs együtthatók számítását használjuk. Ha a paraméteres próbák alkalmazására az adatok normalitásának hiánya miatt nincs lehetőség (a Pearson korrelációs koefficiens számítása nem végezhető el), akkor az eredeti függő és független változók helyett azok rangsorát értékelő rangkorrelációs módszereket kell használnunk. Legelterjedtebb a Spearman rangkorrelációs együttható (ρ) meghatározása.

A rangkorreláció értékelésekor egyszerűen a legkisebbtől a legnagyobb felé haladva külön-külön rangszámot adunk mind a független, mind a függő változó értékeinek. Az ilyen módon kapott rangszámok közti korrelációt már ugyanúgy végezzük el, mint a Pearson korrelációs koefficiens számításakor. Legalábbis elvileg.

A tényleges adatok helyett használt rangokkal sok információt vesztünk, leegyszerűsítjük a vizsgálati helyzetet. Ez baj abból a szempontból, hogy romlik a vizsgálatunk hatékonysága, de jó abból a szempontból, hogy egyszerűsíthető a számítás menete. (A számítógépek segítségével végzett elemzéseknél ez persze nem valódi előny!) Ha az

egyes adatpárok rangszámai közti különbséget (d) számítjuk, akkor a koefficiens az alábbi egyszerű módon is számítható:

$$\rho = 1 - \frac{6\sum d^2}{N(N^2-1)}.$$

Ez a koefficiens is –1 és +1 közti értéket vehet fel. Teljesen független paraméterek esetén 0 az értéke. Interpretációs szabályai megegyeznek a Pearson korrelációs koefficiens értékelésénél ismertetettel. Ugyanazt a táblázatot is lehet használni a kritikus érték megállapítására. Ha ρ abszolút értéke nagyobb, mint a kritikus érték, akkor értékeljük a két változó közti kapcsolatot szignifikánsnak, és minél nagyobb ez az abszolút érték, annál erősebb a kapcsolat. Pozitív koefficiens esetén a két rangsor közt közvetlen, negatív előjel esetén fordított kapcsolat van.

Abban az esetben, ha a változók sorba rendezésekor azonos értékeket találunk, akkor kapcsolt rangszámok kiosztására kerül sor, és a számítás menete is összetettebbé válik. Ha egy érték t esetben fordul elő, akkor a K korrekciós tag

$$K = \frac{t^3 - t}{12}$$

módon számítható. Minden kapcsolt rangszámra külön ki kell számítani ezeket a korrekciós tagokat, és külön a független (x) és függő változóra (y) is összegezni kell (K_x, K_y). A korrigált koefficiens (aminek értelmezési szabályai a korrigálatlan mutatóéval megegyeznek) ezek után számítható:

$$\rho_k = \frac{\frac{1}{16}N(N^2-1) - (K_x + K_y) - \sum d^2}{\sqrt{\frac{1}{6}N(N^2-1) - 2K_x \quad \frac{1}{6}N(N^2-1) - 2K_y}}.$$

Egy háziorvos felmérte saját idős betegei közt a diabetesgondozás hatékonyságát *(27. táblázat)*. Többek között a betegség fennállásának időtartama (T) függvényében vizsgálta a diasztolés vérnyomásértékeket (V). A rangszámok kiosztásakor figyelembe vette, hogy 3 betegét gondozta 20 éve, két-két betegének volt a vérnyomása 104, illetve 111 Hgmm.

A rangkülönbségek négyzeteinek összege 235,5 volt, és pozitív előjelű korrekció nélküli koefficienst kapott eredményül:

$$\rho = 1 - \frac{6 \times 235{,}5}{15(15^2-1)} = 0{,}579.$$

27. táblázat

Gondozás tartama (T)	Diasztolés vérnyomás (V)	Rang–T	Rang–V	Rangkülönbség (d)	Rangkülönbség négyzet (d^2)
17	88	3	1	2	4
18	90	4	2	2	4
14	92	1	3	–2	4
19	98	5,5	4	1,5	2,25
16	104	2	5,5	–3,5	12,25
21	104	10	5,5	4,5	20,25
25	106	13	7	6	36
26	107	14	8	6	36
20	111	8	9,5	–1,5	2,25
23	111	12	9,5	2,5	6,25
19	114	5,5	11	–5,5	30,25
20	119	8	12	–4	16
28	120	15	13	2	4
22	123	11	14	–3	9
20	130	8	15	–7	49

A kapcsolt rangszámok miatt a korrekciós tagokat meg kellett határozni:

$$K_x = \frac{3^3 - 3}{12} = 2\,, \qquad K_y = \frac{2^3 - 2}{12} + \frac{2^3 - 2}{12} = 1\,.$$

A korrigált koefficiens pontos értékét ezek után lehetett számítani:

$$\rho_k = \frac{\frac{1}{16}N(N^2-1) - (K_x + K_y) - \sum d^2}{\sqrt{\frac{1}{6}N(N^2-1) - 2K_x \quad \frac{1}{6}N(N^2-1) - 2K_y}} = \frac{\frac{1}{16}15(15^2-1) - (2+1) - 235{,}5}{\sqrt{\frac{1}{6}15(15^2-1) - 4 \quad \frac{1}{6}15(15^2-1) - 2}} = 0{,}577\,.$$

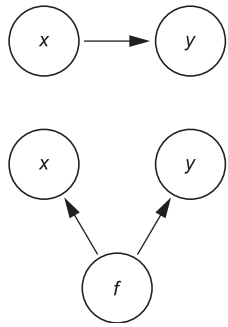
A táblázat alapján 0,482 az ehhez az elemszámhoz tartozó kritikus érték. Mivel ennél nagyobb eredményre vezetett az elemzés, a véletlennel nem magyarázható kapcsolatra talált bizonyítékot a háziorvos. Megállapította, hogy minél hosszabb a gondozásban eltöltött idő, annál magasabb a betegei diasztolés vérnyomása.

Többváltozós elemzések

A korábbi fejezetekben bemutatott statisztikai elemzések kérdése mindig az volt, hogy két paraméter kapcsolatban van-e egymással (ha t-próba segítségével vizsgáltuk, hogy a férfiak közt magasabb-e a szérum-trigliceridszint vagy a nők közt, akkor a nem és a szérum-trigliceridszint közti kapcsolatot értékeltük; ha egy prognosztikai faktor jelenlétében alacsonyabbnak találtuk az 5 éven túl még élő betegek részarányát χ^2-tesztben, akkor a prognosztikai faktor és a betegség kimenetele közti kapcsolatot elemeztük; ha azt analizáltuk, hogy az életkor és a nyaki verőér falvastagsága hogyan korrelál egymással, akkor az életkor és az ér falvastagsága közti kapcsolatot vizsgáltuk). A kapcsolat igazolása kikerülhetetlen lépés a vizsgált probléma megértése szempontjából. Ezen a módon el tudjuk különíteni az egymással valamilyen kapcsolatban levő (tehát valamilyen mechanizmus révén egymásra ható) és az egymástól független (egymással semmilyen mechanizmus révén nem kapcsolódó) faktorokat. Meg tudjuk határozni azokat a tényezőket, amiknek valamilyen szerepe van a vizsgált jelenség alakításában.

A szereplők ismerete után a tényleges funkciók megértése a feladat. Ehhez érdemes végiggondolnunk, hogy a kapcsolt előfordulás milyen módokon jöhet létre! *(20. ábra.)*

Ha igazolni tudtuk, hogy *x* paraméter kapcsolatban van *y*-nal, akkor erre alapvetően két magyarázatot adhatunk. Vagy ténylegesen hat *x* az *y*-ra, vagy van egy faktor (*f*), ami egyaránt befolyással van *x*-re és *y*-ra. Ha az *f* hatására *x* és *y* egyforma irányba változik (mindkettő értékét növeli, vagy mindkettőét csökkenti), akkor *x* és *y* közt úgy jön létre kapcsolat (*x* és *y* pozitív korrelációt fog mutatni), hogy közöttük semmilyen



20. ábra. Közvetlen és közvetett kapcsolat a magyarázó változó és a függő változó között

fizikai kapcsolat nincs, és x változása nem befolyásolja y értékét. Ha a két faktorra ellentétesen hat f, akkor azok inverz korrelációt fognak mutatni. Gyógyszerek támadáspontjának meghatározásakor nyilvánvalóan nem mindegy, melyik kapcsolatrendszer áll az $x \rightarrow y$ kapcsolat hátterében. Az első modellnél x-en ható gyógyszer hatásos lehet, a második modellben biztosan hatástalan lenne. A két helyzet megkülönböztetése ezért elengedhetetlen.

Megoldható a feladat olyan statisztikai eszközökkel, amelyek nem csak x-re és y-ra, hanem f-re gyűjtött adatokat is egy elemzésen belül értékelik. Ha több befolyásoló tényező hatását értékeljük egyszerre, akkor többváltozós statisztikai eljárásról beszélünk. A korábbi fejezetekben bemutatott tesztek mind egyváltozós eljárások voltak; egy befolyásoló tényező kapcsolatát vizsgálták a függő változóval. A többváltozós elemzéseknek az a célja, hogy x-nek y-ra kifejtett hatását f hatásától függetlenül számszerűsítse.

Elvileg a legegyszerűbb módja annak, hogy f zavaró hatásától védjük a vizsgálatunkat, ha olyan módon válogatjuk megfigyelésen alapuló vizsgálatainkhoz a mintát, hogy az f faktorral rendelkezők ne kerüljenek a vizsgálatba. Vagy ennek analógiájára, olyan kísérleti körülményeket teremtünk, amikor az f nincs jelen. Ezeknél a módszereknél nem is gyűjtünk adatot f-re.

Az egyváltozós statisztikai tesztek közt több is van, amit párosított adatok elemzésére fejlesztettek ki. Önkontrollos elemzésekben adott, hogy aki a beavatkozás előtt rendelkezett f faktorral, az utána is. A $x \rightarrow y$ kapcsolatot értékelő teszt eredményét nem befolyásolhatja f. Ha független mintákat illesztünk egymáshoz, akkor mindazoknak a faktoroknak a hatásától megtisztítjuk a vizsgálatot, amiket a párok összekapcsolásakor figyelembe vettünk. (A pontosság kedvéért megjegyzendő, hogy ezeknek feltétele, hogy x és f közt ne legyen interakció, x hatása f-től független legyen.)

Kétszempontos varianciaelemzés

Az egyszempontos varianciaelemzés is kiterjeszthető több befolyásoló tényező együttes vizsgálatára. Legegyszerűbb eljárás a kétszempontú varianciaelemzés ismétlés nélkül. Ezt olyan esetben használhatjuk, ha a kísérletet két (kategorizált változóként értékelhető) befolyásoló tényező hatásának együttes elemzésére szeretnénk felhasználni. Az ismétlésnélküliség arra utal, hogy a két kategorizált befolyásoló tényező lehetséges kombinációinak egy-egy vizsgálati alanyt teszünk ki. A kísérlet végén a *28. táblázat* mintájára tudjuk a mérési eredményeinket összefoglalni.

28. táblázat

		x magyarázó változó				
		x_1	x_2	…	x_n	f átlagok
f faktor	f_1	y_{11}				$\overline{f}_1$
	f_2		y_{22}			$\overline{f}_2$
	⋮			y_{xf}		
	f_m				y_{nm}	$\overline{f}_m$
	x átlagok	$\overline{x}_1$	$\overline{x}_2$		$\overline{x}_n$	$\overline{Y}$

Az x magyarázó tényezőt n-féle, az f-et m-féle dózisban alkalmaztuk. Minden lehetséges befolyásoló tényező kombinációhoz egy y_{xf} vizsgálati eredményt kaptunk a vizsgálat során. Az egyes x_n dózisokhoz kapcsolódó átlagos eredmények ($\overline{x}_n$) közti különbség azt fogja megmutatni, hogy milyen mértékben befolyásolja x a függő változó értékét. Hasonlóan, f hatását az egyes f_m dózisokhoz tartozó átlagos $\overline{f}_m$ eredmények közti eltérés fogja leírni.

Ha x tényleges hatással bír, akkor az egyes dózisaihoz kapcsolódó átlagos eredmények közti különbség kellően nagy lesz. Ha x nem hatásos, akkor ezek az átlagok csak kis mértékben különböznek egymástól. Hasonlóan értelmezhető az egyes f dózisokhoz tartozó átlagos vizsgálati eredmények közti variabilitás is. Természetesen az is előfordulhat, hogy mindkét magyarázó változó befolyásolja a függő változót, vagy egyik sem.

Varianciák elemzésével tudjuk meghatározni, hogy a sor- és oszlopátlagok közti variabilitás elég nagy-e ahhoz, hogy azt már ne értelmezhessük pusztán véletlen hatásként, vagy még nem érte el a variabilitás a kritikus értéket, és a vizsgált tényező nem fejt ki szignifikáns hatást. Az összes vizsgálati eredmény alapján főátlagot ($\overline{Y}$) és ennek segítségével a teljes variabilitást leíró négyzetes eltérések összegét (SS_y) számítjuk ki első lépésben:

$$SS_y = \sum (y_{xf} - \overline{Y})^2.$$

Az x által előidézett variabilitást az egyes dózisokhoz kapcsolódó átlagok és a fő átlag közti négyzetes eltérések összegével tudjuk leírni. Önmagában ez az összeg még nincs tekintettel arra, hogy mennyi adat felhasználásával számoltuk az átlagot. Az x-hez kap-

csolódó variabilitás ezért az elemszámot (amit *f* faktor dózisainak száma határoz meg) is tekintetbe vevő négyzetes eltérésösszeg az alábbi lesz:

$$SS_x = m\sum(\bar{x}_n - \bar{Y})^2 .$$

Hasonlóan írjuk le az *f* dózisai által előidézett variabilitást:

$$SS_f = n\sum(\bar{f}_m - \bar{Y})^2 .$$

A teljes variancia *x* és *f* dóziskülönbségeivel magyarázható variabilitás egészül ki a két tényezőtől függetlenül, tehát a más befolyásoló tényezők hatására létrejövő variabilitással (SS_E), melyet a teljes variabilitás segítségével tudunk számítani:

$$SS_E = SS_y - SS_x - SS_f .$$

A szabadsági fokok figyelembe vételével az *x*-nek és *f*-nek tulajdonítható variancia, illetve a két tényező által nem magyarázott variancia számítható ki. A nem magyarázott varianciához viszonyítva kapjuk meg azokat az F-értékeket *(29. táblázat),* amelyek kifejezik az egyes faktorok variabilitást magyarázó képességét, és amelyekhez kritikus

29. táblázat

Variabilitás forrása	Variabilitás (Négyzetes eltérések összege, SS)	Szabadsági fokok	Variancia (átlagos négyzetes eltérés)	Variancia hányados (F)
Vizsgált magyarázó változó (*x*)	$SS_x = m\sum(\bar{x}_n - \bar{Y})^2$	$df_x = n-1$	$MS_x = \frac{SS_x}{n-1}$	$F_x = \frac{MS_x}{MS_E}$
Vizsgált magyarázó változó (*f*)	$SS_f = n\sum(\bar{f}_m - \bar{Y})^2$	$df_f = m-1$	$MS_f = \frac{SS_f}{m-1}$	$F_f = \frac{MS_f}{MS_E}$
Nem magyarázott variabilitás (*E*)	$SS_E = SS_y - SS_x - SS_f$	$df_E = (n-1)(m-1)$	$MS_E = \frac{SS_E}{(n-1)(m-1)}$	
Teljes variabilitás (*y*)	$SS_y = \sum(y_{xf} - \bar{Y})^2$	$df_y = (n\times m)-1$		

értéket is meg tudunk határozni. (A lépésekhez tartozó részletesebb magyarázat a korábban ismertetett egyszempontos varianciaelemzés leírásánál található.)

Ha F_x nagyobb, mint amit még véletlen hatással magyarázni tudunk, akkor az *x* magyarázó változó és az *y* függő változó közt kapcsolatot tudtunk kimutatni. Ugyanígy értékeljük az *f* faktort is. (A kétszempontos, ismétlés nélküli varianciaelemzésnek alkalmazási feltétele az adatok normális eloszlásán túlmenően, hogy ne legyen interakció a két befolyásoló tényező között.)

Két gyógyszer kombinációjának hatékonyságát szeretnék meghatározni. Ennek érdekében egy in vitro kísérleti modellben 4-4 dózist alkalmaznak mindegyikből. A kísérletek során egy reaktív metabolit mennyiségét mérik. Az eredményt a *30. táblázat* foglalja össze.

F kritikus értéke mindkét vizsgált gyógyszer esetén $F_{[0,05;3;9]} = 3{,}863$, aminél *a* esetében kisebb, *b* esetében nagyobb a számított teszt-statisztika *(31. táblázat)*.

30. táblázat

		a gyógyszer különböző dózisai				
		a_1	a_2	a_3	a_4	*b* átlagok
b gyógyszer különböző dózisai	b_1	39,7	34,8	22,5	26,6	30,9
	b_2	28,5	39,9	28,5	29,6	31,7
	b_3	41,8	35,7	33,6	36,7	37,0
	b_4	41,0	41,4	35,2	39,3	39,2
	a átlagok	37,7	38,0	30,0	33,1	34,7

31. táblázat

Variabilitás forrása	Variabilitás (négyzetes eltérések összege, SS)	Szabadsági fokok	Variancia (átlagos négyzetes eltérés)	Variancia hányados (F)
Vizsgált magyarázó változó (*a*)	180,88	3	60,29	3,544715
Vizsgált magyarázó változó (*b*)	198,68	3	66,23	3,893627
Nem magyarázott variabilitás	153,08	9	17,01	
Teljes variabilitás	532,65	15		

Ezért a vizsgálat elegendő bizonyítékot szolgáltatott arra, hogy *b* gyógyszer hatással van a reaktív metabolit termelődésére, míg *a* gyógyszer esetén a vizsgálat nem volt ebből a szempontból teljesen meggyőző. Ugyanakkor ezen a példán jól demonstrálható, hogy milyen problémát okoz, ha mereven alkalmazzuk az 5%-os döntési küszöböt a statisztikai következtetések levonásakor. A két számított F-érték és a hozzájuk tartozó p-értékek ugyanis alig különböznek egymástól: $p_a = 0{,}061$; $p_b = 0{,}049$. Részben azt látjuk, hogy a két gyógyszer lényegében nem különbözik a metabolit termelődését befolyásoló képesség szempontjából. Az *a* gyógyszer esetében annak a valószínűsége, hogy a különböző dózisban alkalmazott hatóanyag nem eredményezett változást a metabolit termelődésében, 6,1%. Ugyanez a valószínűség *b* gyógyszerre 4,9%. Nyilván nem lehet a véleményünk ezek után az, hogy a két gyógyszer közt lényeges eltérés van!

Többváltozós lineáris regresszió

A lineáris regressziós elemzés is alkalmas arra, hogy egyszerre több magyarázó változó hatását értékelje. Egyváltozós elemzéssel külön-külön elemezve egy-egy magyarázó változó és a vizsgált függő változó közti kapcsolatot, majd standardizálva a regressziós koefficienst, egymáshoz képest értékelhetjük az egyes befolyásoló tényezők hatáserősségét. Ilyen módon egészen addig helyes statisztikai következtetéseket tudunk levonni, amíg a magyarázó változók egymástól függetlenek. Ha ezek közt valamilyen kapcsolat van (egymással is korrelálnak a magyarázó változók, azaz multikollinearitás áll fenn), akkor az egyváltozós elemzésben számított regressziós koefficiens értékét nem csak az éppen tesztelt magyarázó változó határozza meg, hanem a vizsgált magyarázó változóval együtt alakuló más (az adott tesztben éppen nem értékelt) magyarázó változó is. Többváltozós lineáris regressziós elemzés képes arra, hogy egy elemzésbe egyszerre több befolyásoló tényező hatását értékelje olyan módon, hogy minden egyes magyarázó változó önálló (más változók hatásától független) hatását számszerűsítse. A módszer megértéséhez legegyszerűbb a két magyarázó változót egyszerre elemző lineáris regressziós elemzést végigkövetni.

Ha *x* és *z* magyarázó változók hatását vizsgáljuk *y*-ra, akkor hasonlóan járunk el, mint az egyváltozós regressziós elemzéseknél tettük. Először ábrázoljuk a számhármasokat egy háromdimenziós koordinátarendszerben, és keressük azt a regressziós egyenletet, ami a legjobban leírja az ábrázolt pontok trendjét. Az illesztésmódszer itt is a legkisebb négyzetek elve.

$$y = b_x x + b_z z + a \,.$$

Az eredeti egyenlethez képest annyi csak a különbség, hogy mindkét független változóra külön regressziós koefficienst (b_x, b_z) kapunk. Ezek a koefficiensek csak azt a hatást fejezik ki, amit az adott független változó önmagában kifejt. Vagyis már nem érvényesül ezekben a számokban a független változók egymáshoz fűződő kapcsolata. Emiatt az egyes független változókra kapott regressziós koefficienseket külön-külön interpretálhatjuk az egyváltozós elemzésnél elmondottak alapján.

A varianciák elemzésekor is hasonlóan járunk el, mint egyváltozós elemzéseknél tettük. Felbontjuk az *y* teljes (a négyzetes eltérések összegével számszerűsített) variabilitását az *x* és *z* paraméterekkel magyarázható, és ezekkel nem magyarázható részre (SS_E):

$$SS_y = SS_x + SS_z + SS_E \,.$$

Hasonlóan az egyváltozós elemzésekhez, itt is számíthatjuk a korrelációs koefficiens négyzetét minden egyes független változóra. Az így kapott parciális korrelációs koefficiens azt fejezi ki, hogy önmagában az egyes magyarázó változók milyen mértékben (hány százalékban) határozzák meg a függő változó értékét.

$$r_x^2 = \frac{SS_x}{SS_y}$$

$$r_z^2 = \frac{SS_z}{SS_y}$$

A két független változóval magyarázott négyzetes eltérések összege és a mintára jellemző teljes négyzetes eltérés hányadosa a teljes modellre vonatkozó determinációs koefficienst adja meg. Ez fejezi ki, hogy a két változó együttesen milyen mértékben határozza meg a vizsgálati modellben a függő változó értékét, azaz a két változó révén mennyire vagyunk képesek megérteni a vizsgált paraméter értékének alakulását.

$$r^2 = \frac{SS_x + SS_z}{SS_y}$$

Ugyanezeket az elveket követve jutunk el az igazi, többváltozós, lineáris regressziós elemzéshez. A konkrét számítások elvégzése ilyen esetekben már biztosan valamilyen software feladata, ezért ennél részletesebben nem foglalkozunk a képletek megadásával.

A feladat kettőnél több független változó esetén már elveszti azt a szemléletes fogódzkodót is, amit az eredeti adatok koordinátarendszerben történő ábrázolása révén nyertünk. És bár a többdimenziós regressziós egyenletek megadása is az említett alapelveket követi, a feladat matematikailag alapvetően másképpen alakul, mint egy- vagy kétváltozós elemzéseknél.

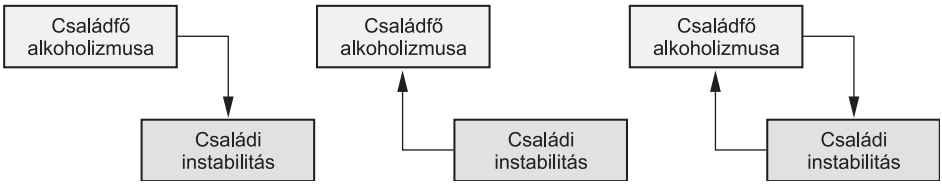
Oksági összefüggések

A jól kivitelezett vizsgálatokban valódi problémára fókuszáló kérdésre kapunk kellően megbízható, a magyarázó változó és a kiváltott hatás közti kapcsolat véletlenszerűségéről, statisztikai eszközök segítségével levont, következtetésen alapuló választ. A kutatási kérdésre adott válasznak azonban csak egyik alapja a statisztikai elemzés eredménye. Ugyanilyen fontos, hogy a vizsgálat körülményeit kellően mérlegeljük: nem lehet-e az eredményt a mintaválasztás hibájával, a zavaró tényezők nem megfelelő kontrolljával, vagy a mérések, az adatgyűjtés pontatlanságával magyarázni. Ezen túlmenően azt is értékelni kell, hogy az eredményeink összhangban vannak-e a vizsgálatunkkal kapcsolatos elméletekkel.

A statisztikai eljárások eredményeként alapvetően arról adnak információt, hogy a különböző paraméterek között van-e kapcsolat. A kapcsolat természetére, a kapcsolatot létrehozó mechanizmusra vonatkozóan nem. Emiatt a véletlennel (és a vizsgálati hiányossággal) nem magyarázható kapcsolatok feltárása esetén sem állíthatjuk, hogy ok-okozati összefüggést találtunk. Sőt, esetenként még azt sem tudjuk értelmezni pusztán a statisztikai eredmények segítségével, hogy adott vizsgálatban melyik volt a magyarázó és melyik a függő változó.

Az ok-okozati kapcsolat iránya

Ha egy vizsgálatban értékeljük a családfő alkoholizmusa és a család instabilitása közti kapcsolatot, akkor minden bizonnyal könnyen tudjuk bizonyítani a két jelenség véletlennel nem magyarázható kapcsolatát, de távolról sem ilyen egyértelmű annak eldöntése, hogy a kapcsolat alapját a családfő alkoholizmusa miatt kialakuló családi diszfunkció, vagy a család működési zavarának következtében felépülő családfői alkoholista karrier, esetleg mindkét mechanizmus egyszerre jelenti *(21. ábra)*. A valódi kiváltó ok pontos meghatározása azért fontos, mert csak az oksági láncolat megszakításával lehet eredményt elérni. A három modellhez három beavatkozási stratégia tartozik. A hatékony segítségnyújtás megszervezéséhez alapvetően fontos a modellek közti helyes választás, a folyamatot valóban elindító faktor kezelésére koncentrálva érhetünk el csak sikert. Rosszul értékelt helyzet rossz beavatkozási pontokat határoz meg, ami csak az erőforrások elherdálásához és a beavatkozástól várt nyereség elmaradásához vezet.

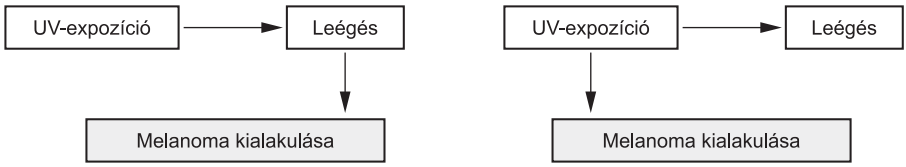


21. ábra. Lehetséges kapcsolódási irányok a családfő alkoholizmusa és a család instabilitása között

Ha konzekvens kapcsolatot látunk a napozás utáni leégés és a melanoma kialakulásának kockázata közt, akkor elvileg önmagában a bőrt érő UV-expozíció vagy a leégés patológiai folyamatai is magyarázatul szolgálhatnak az összefüggésre *(22. ábra)*.

A kétféle modellhez kétféle beavatkozási stratégia tartozik. Az első esetben a leégés megelőzése vagy kezelése egyaránt hatékony preventív beavatkozás. A második modellben a leégés kezelését nem, csak az UV-expozíció kerülését tekintjük preventív eszköznek.

A hatékony beavatkozási pont azonosítása a valódi oksági láncolat meghatározásán alapul. Minél összetettebb problémával állunk szemben, annál nehezebb az oksági láncolatok feltárása, de soha nem tekinthetünk el ettől, vizsgálatunk kötelező része az oksági értelmezés.



22. ábra. Az UV-expozíció, a leégés és a melanoma kialakulása közötti kapcsolat hátterében álló hatásmechanizmusok

Koch-posztulátumok

Számos olyan fertőző betegségeket ismerünk, amik jellemző klinikai képpel járnak, és amiket egy kórokozó képes csak kiváltani. Az ilyen monokauzális betegségek vizsgálatára dolgozta ki *Koch* azokat a kritériumokat, amiknek meg kell felelnie a kórokozónak ahhoz, hogy azt a betegség okozójának tekintsük:

1. A kórokozó a beteg szervezetéből kimutatható, de nem mutatható ki az egészségesekből.
2. A betegből izolált kórokozó tenyészthető.
3. A kitenyésztett kórokozóval megbetegíthető az egészséges szervezet.
4. Az inokulált kórokozó kitenyészthető a tenyésztett kórokozóval megbetegített szervezetből.

Számos járványügyi problémát sikeresen oldottak meg az ezekre az elvekre épített kutatásokra alapozva. Ugyanakkor már a tuberkulózis vizsgálata során világossá vált, hogy a posztulátumai nem tekinthetők általában érvényesnek. (Hiszen tünetmentes szervezetből is kitenyészthető a Mycobacterium, az egészséges szervezetbe juttatott Mycobacterium pedig nem feltétlenül vált ki betegséget.) A Koch-posztulátumok akkor működnek jól, ha a kórokozó szükséges (kórokozó nélkül nem alakul ki adott betegség) és önmagában elégséges (a fertőzöttekben mindig kifejlődik az adott kórkép) oka is a betegségnek. A Mycobacterium-fertőzés szükséges a tuberkulózis kialakulásához (nélküle nincs tuberkulózis), de önmagában nem elégséges, hiszen nem minden fertőzött lesz beteg (egyéb tényezőknek is jelen kell lenni a betegség kialakulásához.)

Hill-szempontrendszer

A betegségek többsége multikauzális, hátterükben számos kiváltó hatás áll. A legtöbb daganatra például igaz, hogy nincs olyan ágens, ami nélkül biztosan nem fejlődik ki, azaz nincs szükséges oka. Ezzel szemben több olyan külső hatás is létezik, ami képes elindítani a malignus klón kialakulását, amelyek viszont önmagukban nem elegendőek a betegség kifejlődéséhez. Az összetett biológiai folyamatok bonyolult oksági hálózatba rendeződnek, melyeken belül az egyes hálózati csomópontok (faktorok) közti kapcsolat azonosítása (egyik faktor változása kapcsolatban van a másik változásával) még távolról sem jelenti a köztük lévő ok-okozati kötelék bizonyítását és a beavatkozási pont azonosítását.

Sajnos nem rendelkezünk olyan kritériumrendszerrel, aminek a segítségével a kapcsolatok ok-okozati vagy nem ok-okozati természetét hiba nélkül meg tudnánk ítélni. Ugyanakkor szükség van olyan szempontrendszerre, amely gyakorlati kérdések megoldásakor általában helyes döntést eredményez. Ennek az igénynek megfelelő rendszert dolgozott ki *Austin Bradford Hill.* Az általa javasolt szempontok alapján értelmezni kell minden vizsgálati eredményt, de hangsúlyozni kell, hogy (az időbeliségtől eltekintve)

egyik szempont sem perdöntő önmagában az okságot illetően. Maga Hill is a kutatási eredmények értelmezését segítő szempontrendszernek és nem az okság megítélését szolgáló kritériumrendszernek tartotta az ajánlását.

1. Az összefüggés erőssége

Általában igaz, hogy szoros összefüggést mutató mérőszámokat akkor eredményez egy vizsgálat, ha ok-okozati kapcsolat van a vizsgált paraméterek közt.

A HPV méhnyakrákot okozó hatása melletti alapos érv, hogy daganatepidemiológiában szokatlanul magas kapcsolati mutatókat (100 feletti esélyhányadosokat, incidenciahányadosokat) találnak az epidemiológiai vizsgálatokban.

De számos példa van arra, hogy egy magyarázó változó ténylegesen befolyással van egy függő változóra, de annak értékét csak kis mértékben változtatja meg. Az ilyen gyenge befolyásoló tényezők vizsgálata komoly nehézséget jelent, mert a nem megfelelően kontrollált zavaró tényezők okozta torzító hatás is előidézheti a gyenge, de szignifikáns kapcsolatot. A passzív dohányzásról ma már jól tudjuk, hogy megnöveli a tüdőrák kialakulásának kockázatát. Ez a hatás azonban gyenge, ezért nagyon hosszú időt vett igénybe a megfelelő bizonyítékok előállítása, olyan vizsgálatok végrehajtása, amelyekben a gyenge hatás észleléséhez szükséges nagy mintaelemszám biztosítható volt, és ahol a tüdőrák kialakulását segítő, egyéb kockázatnövelő tényezők kontrollja is hatékonyan megvalósítható volt.

Ugyanakkor arra is számos példa akad, hogy szoros kapcsolatban levő tényezők közt nincs ok-okozati összefüggés. Ha azt vizsgáljuk, hogy a családon belül hányadik gyermeknek a legnagyobb az esélye arra, hogy Down-szindrómás legyen, akkor azt látjuk, hogy a születési sorrenddel fokozatosan emelkedik a betegség kockázata. Igen erős a statisztikai kapcsolat. A kapcsolat hátterében azonban nem a születési sorrend, hanem a születési sorrenddel szintén szoros kapcsolatot mutató anyai életkor áll. Valójában az anyai életkorral emelkedik a Down-szindróma kockázata, és ettől teljesen függetlenül, a gyermekek születési sorrendje és az anya életkora szintén szoros kapcsolatot mutat a betegség kialakulásával.

Az erős kapcsolat inkább csak annak kizárására ad alapot, hogy valamilyen ismeretlen, vagy a vizsgálatban nem figyelt zavaró tényező torzító hatásának tulajdonítható a paraméterek közt leírt összefüggés.

2. Konzisztencia

Saját vizsgálatunk eredményét össze kell hasonlítani a mások által kivitelezett vizsgálatok eredményeivel. Ha a hasonló kérdésre fókuszáló, de más körülmények

közt elvégzett vizsgálatok és a saját eredményeink közt összhang van, akkor beszélünk konzisztenciáról. Minél több vizsgálat által rajzolódik ki egy konzisztens kép, annál inkább meggyőződünk az oksági kapcsolatról.

Teljesen mindegy, hogy milyen módszerrel vizsgálják a HPV jelenlétét a méhnyakrákos szövetekben, mindig arra az eredményre jutnak, hogy a vírus kockázatnövelő jellegű. Mivel nagyon sok vizsgálatot hajtottak már végre nagyon sok országban, ez a konzisztencia az okság melletti erős érvet jelent. Nincs is irodalmi adat, ami ezzel ellentétes megállapításra jutott volna.

A konzisztenciahiány azonban egyáltalán nem zárja ki az oksági kapcsolatot. Előfordulhat, hogy egy faktor önmagában nem képes kiváltani a vizsgált hatást, csak ha együttesen hat más tényezőkkel. Azokban a vizsgálatokban, ahol a kiegészítő hatás is jelen van, látni fogják a kapcsolatot; ahol a kiegészítő tényező nincs jelen, ott nem látják. Az eredmények tehát nem lesznek konzisztensek.

Ha azt vizsgálták, hogy a transzfúzió növeli-e a HIV-fertőzés kockázatát, akkor azokon a helyeken, ahol a vérkészítmények biztonságosak voltak, nem látták a kapcsolatot; ahol nem vizsgálták a vérkészítményeket, és a HIV-hordozó tünetmentesség prevalenciája nagy volt, ott észlelték a véletlennek nem tulajdonítható kapcsolatot.

3. Specificitás

Ha egy befolyásoló tényező csak egy elváltozást, betegséget képes kiváltani, akkor beszélünk a kapcsolat specifikus jellegéről.

A több mint százféle HPV nem csak méhnyakrákos szövetekben mutatható ki, hanem egy sor más daganatnál is kockázatnövelőnek tűnik. Legalább 15 (de elsősorban 2) HPV-szerotípus kapcsolható a méhnyakrákhoz. Az egyes szövettani típusokhoz viszont sajátos HPV-spektrum tartozik, ami a kapcsolat specificitását mutatja.

A fertőző betegségeknél tipikus erős specificitás nagyon ritkán fordul elő nem fertőző betegségek esetében, ezért a vizsgálati eredmények értékelésekor ez a szempont általában nem segíti az értelmezést. Inkább kivételesnek mondható az a helyzet, amikor ez a szempont segíti a helyes következtetés levonását.

A dohányzás talán a legjobb példa arra, hogy egy tényező képes nagyon sok ponton károsítani az élettani funkciókat. Az egyes funkcionális veszteségek és a dohányzás közti kapcsolatot értékelő vizsgálatok eredményei akkor is ok-okozati természetűek, ha ez a specificitás egyáltalán nem érvényesül.

4. Időbeliség

Az oknak meg kell előzni az okozatot. Ez az egyetlen szempont, aminek mindig teljesülnie kell. Különösebben ezt nem is kell értelmezni.

A szűrési programok során nyert minták vizsgálata alapján jól ismert, hogy évekkel a méhnyakrák kialakulásának megkezdődése előtt már HPV-fertőzött a méhnyak hámja.

Talán azt érdemes csak megjegyezni, hogy ha egy feltételezett ok később jelenik csak meg, mint az okozat, akkor ez nem jelenti azt, hogy az oki tényező nem képes kiváltani az elváltozást. Azaz, ha fordított időbeliséget látunk a vizsgálatunkban, az még nem zárja ki az ok-okozati kapcsolat lehetőségét. (Vannak vesebetegek, akiknek a hipertóniája a vesebetegség szövődményeként alakul ki. Ez még nem jelenti annak a bizonyítékát, hogy a hipertónia nem tud vesebetegségeket előidézni.)

5. Dózis-hatás kapcsolat

Ha azt tapasztaljuk, hogy a magasabb dózisokhoz intenzívebb kiváltott hatások tartoznak, akkor ez az ok-okozati jelleg melletti erős érv, az egyébként kimutatottan létező kapcsolat esetében. Elsősorban a dózissal monoton módon erősödő válasz a meggyőző.

Bár egy fertőző ágens esetében nem biztos, hogy a magasabb vírusszámtól magasabb kockázatot várunk, a manapság elindított és kvantitatív vírus DNS-mérésen alapuló vizsgálatok azt mutatják, hogy a HPV magasabb dózisaihoz fokozottabb méhnyakrák-kialakulási kockázat társul.

A monotonitás azonban nem mindig teljesül. Nagyon sok toxikus anyag csak akkor fejti ki a hatását, ha egy bizonyos mennyiségnél több kerül a szervezetbe. A hatás kifejlődése valamilyen küszöbdózishoz kötődik. Az ilyen küszöbökhöz igazítják például a környezet-egészségügyi határértékeket.

Az alkohol és halálozás közti kapcsolat sem monoton jellegű. A moderált alkoholfogyasztók közt alacsonyabb a halálozás, mint az absztinenseknél. A legmagasabb halálozási adatokat a sok alkoholt fogyasztók között látjuk (J-alakú dózis-hatás görbe). Mivel a görbe egyes részeihez más-más értelmezés tartozik, valójában ez a vizsgálati eredmény is az alkohol egészségkárosító hatását bizonyítja. (Az absztinencia oka meglehetősen gyakran egy életet veszélyeztető alapbetegség. Az absztinensek közti emelkedett halálozás nem az alkohol protektív hatásának elmaradását tükrözi, hanem az absztinenciát is megalapozó alapbetegség következménye.)

Másfelől monoton dózis-hatás kapcsolat esetén sem kell feltétlenül ok-okozati kapcsolat meglétére gondolnunk. Magyarországon a 90-es években folyamatosan emelkedett az antidepresszánsfogyasztás, és folyamatosan csökkent az öngyilkosság miatti halálozás. A két trend közt nagyon szoros a kapcsolat. Magasabb gyógyszerfo-

gyasztáshoz konzekvensen alacsonyabb halálozás párosul. A kapcsolat önmagában mégsem bizonyíték az oksági viszonyra, mert nagyon sok társadalmi-gazdasági tényező is monoton módon változott ugyanebben az időszakban, amik elvben szintén magyarázhatják az öngyilkosság csökkenő trendjét.

6. Biológiai plauzibilitás

Ha kapcsolatot állapítunk meg két tényező közt, és egyéb vizsgálatok alapján ismerjük meg azt a biológiai mechanizmust, ami mentén a kapcsolat ténylegesen működik, akkor ez erős érv a feltárt kapcsolat ok-okozati jellege mellett.

Nagyon részletesen ismertek azok a molekuláris mechanizmusok, amik a HPV-infekció következtében indulnak meg, és a méhnyakrák kialakulásához vezetnek.

Az ismert pathomechanizmus hiánya viszont önmagában nem vonja maga után az oksági kapcsolat elvetését, hiszen lehet, hogy egyszerűen eddig fel nem tárt biológiai folyamat adja az alapját a statisztikai eszközökkel leírt kapcsolatnak. *Semmelweis Ignác* sem tudta biológiai alapfolyamatok mentén értelmezni, hogy miért hatékony a gyermekágyi láz megelőzésére a klórmeszes kézmosás, de látta az összefüggést. Az ok-okozati jelleg mellett pedig kiváló érvet szolgáltatott a kézfertőtlenítés hatékonysága. *John Snow* úgy tudott hatékonyan beavatkozni a kolera megelőzésének érdekében, hogy azt sem tudta, hogy léteznek baktériumok. Csak annyit tudott, hogy kapcsolat van bizonyos helyekről származó ivóvíz fogyasztása és a betegség terjedése közt. *James Lind* táplálkozási emberkísérletei után (citromlével kiegészített táplálék fogyasztása esetén meggyógyultak a skorbutos tengerészek) a matrózok citromlevet kaptak a hosszú hajóutakon, hogy megelőzzék a betegség kialakulását, pedig senki nem tudott még akkor a C-vitaminról.

7. Koherencia

Saját eredményeinket össze kell vetnünk a vizsgálat-elváltozásról szóló egyéb ismeretekkel. Ha összhang van a kiváltott hatással kapcsolatos régebbi kutatási eredmények, illetve az elméleti megfontolások és a saját vizsgálatban leírt összefüggés közt, akkor beszélünk koherenciáról.

A méhnyakrák a szexuális úton terjedő betegségekre jellemző epidemiológiai jellegzetességgel bír. Ennek teljesen megfelel a HPV, mint etiológiai ágens, ami szexuális úton terjed.

(A koherenciát önálló szempontként nevezte meg annak idején *Hill*. De azért a koherencia és a plauzibilitás közti jelentős átfedés egészen nyilvánvaló.)

8. Kísérletes bizonyítékok

Ha olyan vizsgálat alapján vonjuk le két tényező kapcsolatára vonatkozó következtetésünket, ahol a magyarázó változó eltávolítása után helyreálló élettani funkciót mérjük, akkor ez – *Hill* értelmezésében – kísérletes bizonyíték arra, hogy a magyarázó változó oki ágens.

A HPV elleni vakcinálás után csökkenő méhnyakrák-incidencia adja a kísérleti jellegű bizonyítékot a HPV etiológiai szerepére.

Nem vitatva, hogy sok esetben ez az érvelés helytálló, nem tekinthető ez a szempont sem általános kritériumnak; és nem csak azért, mert a befolyásoló faktor kísérletes körülmények közti eltávolítása nem mindig oldható meg.

A malária elnevezése is arra utal, hogy sokáig azt gondolták, a mocsarak rossz levegőjének belélegzése okozza a betegséget. Ha erre egy mocsár-lecsapolási kísérletet hajtottak volna végre, akkor látszólag igazolni lehetett volna az oksági kapcsolatot, hiszen a mocsarak lecsapolás után a betegség visszaszorul. Ennek ellenére nincs ok-okozati összefüggés a mocsár levegője és a malária közt. A mocsár lecsapolásával a kórokozót terjesztő szúnyog veszti el élőhelyét, és emiatt csökkent a betegség gyakorisága.

9. Analógia

Az analógiák (saját vizsgálati kérdésünkre hasonlító helyzetekkel kapcsolatos ismeretek) keresése nagyon fontos a hipotézisek felállításakor, de erősen kétséges, hogy saját eredményeink ok-okozati jellegének elbírálásakor ezt a szempontot komolyan lehet-e egyáltalán venni. (Mindenesetre *Hill* ezt a szempontot is javasolta. Arra hívta fel a figyelmet, hogy ha egy gyógyszerrel kapcsolatban igazoltuk, hogy az fejlődési rendellenességet képes okozni, akkor ezt más gyógyszerről is el tudjuk képzelni.)

A HPV-cervixrák etiológiai viszonnyal kapcsolatban említhető analógiák az alábbiak: (1) vannak más DNS-vírusok is (pl. HBV), melyek képesek humán daganatot indukálni; (2) vannak olyan állati papillómavírusok, melyek képesek daganatot előidézni; (3) vannak olyan karcinogén vírusok, amelyek hasonló elemi hatásokat váltanak ki, mint a HPV (interakció p53- és RB-fehérjékkel).

Tárgymutató

A, Á

abszolút érték 74, 89
adat 13
~, dichotóm 13
~, nominális skálán mért 14
~, ordinális skálán mért 14
~, diszkrét 15
~, kvalitatív 54, 60
~, kvantitatív 39
adatbázis 11, 13
adatgyűjtés 11
adatpár 65
analógia 116
ANOVA-tábla 51
Apgar score 14
arányskála 15
átlag 16, 39, 40
átlagkülönbség 40
átlagos eltérés 22

B

befolyásoló tényező 7, 10
biológiai plauzibilitás 115
Bonferoni-korrekció 46, 52

C

centrális érték 20

Cs

csoportvariabilitás 49, 50

D

determinációs koefficiens 84
deviancia 77
dichotomizálási küszöb 14–16
dózis-hatás kapcsolat 114
döntési küszöb 34–36

E, É

egyéni variabilitás 47
eloszlás 39
eloszlásfüggvény 38, 58
eloszlás-paraméter
előjel 89
előjelteszt 86–88
elsőfajú hiba 36, 37, 46
értékpár 58

esetszám-eloszlás 57
extrém érték 20

F

F-érték 51
Fisher-exact eljárás 57
folytonos változó 20, 68, 69
F-próba 45, 51
független változó 75, 85, 107
függő változó 75, 79, 80, 85, 101, 106

G

Glasgow Coma Scale 14
goodness of fit test 57

H

haranggörbe-eloszlásminta (Gauss-görbe) 23
hatékonyságnövekedés 44
Hill-szempontrendszer 111
hipotézis 7
hipotézistesztelés 11, 34, 36
hisztogram 19

I

időbeliség 113
intervallumskála 15

J

jósolt érték 82

K

kapcsolat 101
kategória 59
kétmintás t-próba 43, 92
kétszempontos varianciaelemzés 102
kísérlet 7, 8
kísérletes bizonyíték 116
kiváltott hatás 7, 9
Koch-posztulátumok 110
koherencia 115
Kolmogorov–Szmirnov-próba 45
konfidencia-intervallum 72
kontingenciatáblázat 60
konzisztencia 112, 113
konzisztenciahiány 113
korreláció 68
~ teljes pozitív 72
~ inverz 72
korrelációs koefficiens 71–75, 84
kovariancia 69–71, 83
kritikus érték 38, 56, 74, 87
~ statisztikai mérőszám 87
Kruskal–Wallis H-teszt 92
kutatási kérdés 9
különbség 39, 40
küszöbérték-definiálás 16
küszöbdózis 114

L

legkisebb négyzetek elve 77, 106
leíró statisztika 17
lineáris kapcsolat 68
~ regresszió 75, 78, 83

M

magyarázó változó 79, 80, 101, 16
Mann–Whitney U-teszt 90–92
másodfajú hiba 36, 37
McNemar-teszt 66
medián 20, 90, 92
megbízhatósági tartomány 27–35, 72, 73, 80
megfigyelés 7, 8
megfigyelt esetszám 55
mérési technika 13
módusz 21
monotonitás 114

N

négyzetes eltérés átlaga 50
~ ~ összege 47, 48, 79, 103
~ különbségek összege 55
nem paraméteres eljárás 86
normális eloszlás 23, 25, 86
nullhipotézis 56, 60

O

ok 113, 114
ok-okozati kapcsolat 12, 109, 112, 114
okozat 113, 114
oksági összefüggés 109
~ láncolat 109, 110
~ kapcsolat 113

Ö

összegzés 55
összefüggés erőssége 112

P

párelemzés 45
Pearson korrelációs koefficiens 98, 99
p-érték 36, 38, 87
Poisson-eloszlás 58

R

rangszám 89
rangszámösszeg 89, 90
regresszió 79
regressziós egyenes 76, 77, 85
~ elemzés 75
~ koefficiens 80, 84
reprezentatív minta 27

S

Spearman rangkorrelációs együttható 98
Spearman-rangkorreláció 98
specifitás 113
standard deviáció 28, 30
~ hibaszámítás 31
standardizált érték 24
~ különbség (t-érték) 43, 45
~ regressziós koefficiens 85
statisztikai teszt 38
~ változó 16
sűrűségfüggvény 24, 25

Sz

szabadsági fok 22, 23, 45, 49, 87
szórás 16, 39
szórásdiagram 68, 69
szóródás 18, 70
teljes variabilitás 50

T

teszt-statisztika 98, 105
t-érték 23
többszörös hipotézistesztelés 46
többváltozós elemzés 101
~ lineáris regresszió 106
~ ~ regressziós elemzés 85
t-próba, Student-próba 23, 43, 45

V

várható érték 54
~ esetszám 55
variabilitás 18, 40, 49, 52, 78, 79
variancia 22, 41, 52
varianciaelemzés 51–53, 78
varianciahányados 50, 80
vizsgálat lépései 11

W

Wilcoxon párosított teszt 88–90

Y

Yates-korrekció 56, 66, 87