

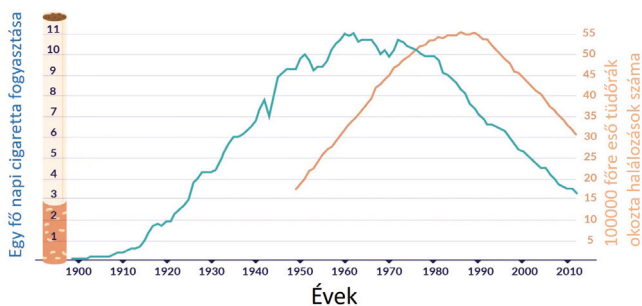
TANÍTSUNK STATISZTIKUS SZEMLÉLETET!

A szlovák középiskolai tantárgyi követelményeket nem ismerem. A magyar NAT szerencsére tartalmaz elemi statisztika ismereteket (minta, átlag, variancia, adatok ábrázolása). Ezek az ismeretek nagyon fontos részei az általános műveltségnek, de véleményem szerint ezen túl elengedhetetlenül szükséges lenne a statisztikus gondolkodásmód elemeinek az elsajátítására is. A modern tudomány és így azok az ismeretek, melyek birtokában kényelmes, kiszámítható életet élünk, számtalan ponton épít a természeti és társadalmi jelenségek, terápiás eljárások adatainak statisztikai elemzésére. Azonban erre az általam ismert középiskolai tankönyvek nem mutatnak rá. Valami nagyon fontos hiányzik az érettségivel rendelkező tanulók szemléletéből, és ez a tudáshiány megágyaz a kurzuslónak, oltáselleneseknek és általában a tudománytagadóknak (lásd korábbi írásomat a Katedrában).

A statisztika egy nemszeretem tantárgy az egyetemeken is, mert magasabb szinten annak elsajátítása komoly erőfeszítést, sok gyakorlást igényel. Ám az alapgondolatok a középiskolásoknak is jól átadhatók, egyúttal a diákok egy izgalmas szellemi kaland részesei is lehetnek. A következőkben a statisztikai gondolkodásmód fontosságát és hasznát fogom bemutatni néhány ismertebb vagy a cikk írása közben kitalált példán keresztül. Eközben alapvető statisztikai fogalmakat és kísérleti módszereket is megismerünk.

A DOHÁNYZÁS NÖVELI A TÜDŐRÁK KIALAKULÁSÁNAK AZ ESÉLYÉT

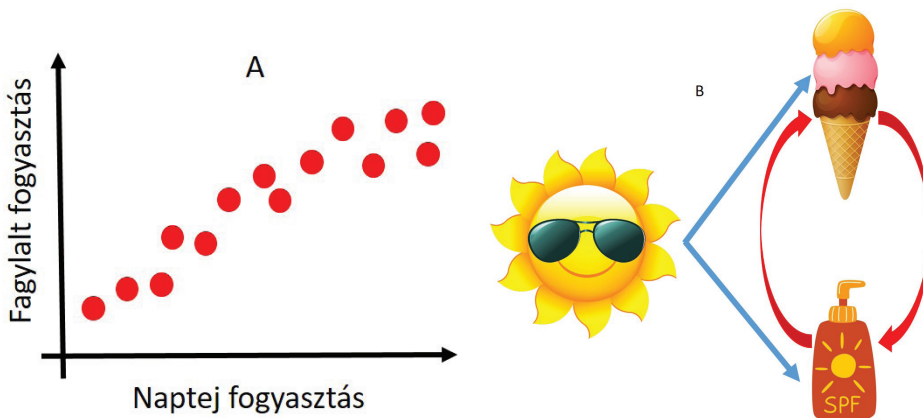
A címben szereplő állítást szinte mindenki alapvető tényként fogadja el. Talán meglepő, de ez a tény még a múlt évszázad hatvanas éveiben is heves vita tárgya volt. Ekkorra az adatok elemzése egyértelműen kimutatta, hogy a dohányzás és tüdőrák okozta halálozási esetek között a kapcsolat (szakszóval korreláció) nagyon erős (1. ábra), ám két mennyiség közötti kapcsolat még nem jelent szükségszerűen ok-okozati kapcsolatot is.



1. ábra

Az Egyesült Államok egy főre jutó cigarettafogyasztása és a százezer főre jutó tüdőrák okozta halálesetek alakulása 1900-tól 2010-ig. Jól látszik, hogy a cigaretta fogyasztásának felívelését majd a 80-es évektől a csökkenését durván 30 év késéssel követi a tüdőrákos esetek növekedése majd csökkenése is. Forrás: <https://healthmatch.io/blog/smoking-and-lung-cancer-how-big-is-the-problem-today>

Egy pillanatra félretéve a dohányzás meg a tüdőrák témát, világítsuk meg ezt a kérdést egy hétköznapi példa segítségével. A kiskereskedelmi adatok alapján kimutatható, hogy az egy havi naptejfogyasztás és ugyanazon hónapban a fagyaltfogyasztás között nagyon szoros kapcsolatot találunk, valami olyasmit, amit a 2.A. ábra mutat. Bár a kapcsolat nagyon szoros, biztosan érezzük, hogy ebből sem az nem következik, hogy a fagyalt fogyasztása miatt használunk több naptejet, sem az, hogy a fokozott naptejhasználat váltja ki a fagyalt iránti vágyódásunkat. Valójában van egy közös ok, az adott hónapban az átlaghőmérséklet és a napsütéses órák száma (vagy a legmagasabb napi hőmérséklet), az ami meghatározza mind a naptej, mind a fagyalt fogyasztását (2.B. ábra).



2. ábra

A naptej és a fagyalt fogyasztása közötti kapcsolat. A: E két mennyiség szoros kapcsolatban (korrelációban van), ami mögött a közös ok, hogy mennyire van meleg az adott hónapban. B: A piros nyilak mutatják a lehetséges közvetlen ok-okozati kapcsolatot a két mennyiség között, de a kapcsolatot valójában a kék nyilakkal jelölt közös ok okozza.

Visszatérve a dohányzás és a tüdőrák kialakulása közötti kapcsolatra, gondolhatunk arra itt is (és persze gondoltak is a kutatók annak idején), hogy valami közös ok van a háttérben: bizonyos emberek olyan géneket hordoznak, melyek egyszerre a nikotinfüggőségre és a tüdőrákra is hajlamosítanak. De azt sem lehetett kizárni, hogy a rákos sejtek korai megjelenése okozza a nikotin iránti vágyat. Azóta egyértelműen kiderült, hogy e két feltételezés nem állja meg a helyét. Kiderült viszont, hogy a cigaretta füstjében számos olyan vegyület van, melyek a tüdőszövetek sejtjeinek genetikai állományát támadják, így a sejtek könnyen rákossá válhatnak. Ezt az összefüggést azóta rengeteg közvetlen és közvetett kísérlet is igazolta. Bizonyított, hogy a tüdőrákos megbetegedések 90%-át a dohányzás okozza. Ne dohányozzunk!

Azt, hogy egy vegyület hatására rákos burjánzásnak indul egy szövettenyészet vagy sem, azt megbízhatóan ki lehet mutatni. Az állatok és emberek viselkedése viszont rengeteg ismeretlen, így a megfigyelő számára véletlenszerű elemet tartalmaz. Például kíváncsiak vagyunk arra, hogy meg tudnak-e a patkányok különböztetni szögletes formákat kerekded formáktól, vagy az emberek képesek-e arra, hogy megállapítsák, hogy a tea vagy a tej került-e előbb a csészébe. Mivel ilyenkor a patkányok vagy az emberek akár véletlenül is adhatnak helyes választ, a kérdés eldöntéséhez kicsit el kell mélyedjünk a valószínűségszámításba, és meg kell ismerkedjünk a nullhipotézis és a szignifikanciaszint fogalmával.

A HÖLGY ÉS A TEA, AVAGY VAK TYÚK IS TALÁL SZEMET, AZONBAN NEM MINDEGY, HOGY MENNYIT

A fejezet címéhez tartozó történetnek mint a tudományos anekdotáknak általában számos változata van, de mindegyik változat közös vonása, hogy valamikor az 1920-as évek Angliájában játszódik. A történetben Muriel Bristol alkagutató kollégáival társalognak kifejt, hogy ő úgy szereti a teát, ha előbb a tejet, majd a teát öntik a csészébe. Mert az nem mindegy, érvel Muriel, a teának más az íze, ha más sorrendben kerülnek a csészébe ezek az összetevők. A társaság tagjai kételkednek ebben, hiszen semmilyen kémiai vagy fizikai okot nem tudnak elképzelni, ami alapján más lenne a tea és a tej keveréke a csészébe töltés sorrendjének a hatására. A társaságban jelen van Ronald Fisher, a XX. század egyik legkiemelkedőbb statisztikusa és elméleti biológusa (későbbi Sir Ronald Fisher), aki némi gondolkodás után egy kísérletre tesz javaslatot.

Tegyünk a hölgy elé nyolc csésze teát, mondja Fisher. Négy csészébe a tejet, a másik négyben a teát töltjük először, de persze minden csészében gondosan megkeverjük az összetevőket. A csészéket teljesen véletlenszerűen sorba rakjuk az asztalon. Ezután a (különleges ízleléssel kérkedő) hölgynek mondjuk is el az imént ismertetett kísérleti beállítás részleteit. Majd kérjük meg, hogy ízlelje meg a nyolc csésze tartalmát, és mondja meg, hogy mely csészékbe került a tej és melyekbe a tea először. Fisher alap- vagy nullhipotézise az, hogy a hölgy nem tudja megmondani, hogy a tea vagy a tej került-e előbb a csészébe. Természetesen véletlen szerencse folytán lehetnek helyes tippjei, ám a kérdés az, hogy a 2, 3, esetleg 4 helyes tippnek mi a valószínűsége az adott kísérleti beállításban. Ezen a ponton némileg önkényes, ám azóta is szokássá vált feltevessel él Fisher: ha a hölgy tippelése olyan eredményre vezet, ami pusztán véletlen szerencse folytán 5%-nál kisebb eséllyel ($p < 0,05$) valósul meg, akkor a hölgy valóban rendelkezik a megkülönböztetés képességével, azaz a nullhipotézis hamis. Ellenkező esetben a nullhipotézis igaz. Persze ezt a korlátot megszabhatnánk 1% vagy akár 0,1% százalékban is, és sok esetben ezeket a sokkal szigorúbb, úgynevezett szignifikanciaszinteket állítják be a kísérletekben.

Kérdés, ezek után mondhatjuk-e, hogy Muriel Bristol valóban érzékeli az összetevők betöltésének a sorrendjét, ha mind a nyolc csésze esetén eltalálja a helyes sorrendet. Mondhatjuk-e ugyanezt, ha hármát talál el a négyből, és ha kettőt?

A válaszhoz ki kell számoljunk, hogy ezek az események milyen valószínűséggel következnek be, ha a hölgy (vagy bárki más) csak vakon találgat. Ehhez némi elemi matematikát be kell vessünk. A kísérleti alanyunk két négyes csoportra kell osztani a nyolc csészét. Ha négyet az egyik csoportba sorolunk (pl. tea először csoport), akkor ezzel a másik négyet meg besoroltuk a másik csoportba. Számoljunk ki először, hogy hányféleképpen lehet 4 különböző csészét kiválasztani a nyolcból (pl. a tea először csoportba). Az elsőt 8 közül, majd a másodikat 7, a harmadikat 6 és a negyediket 5 közül választhatjuk, ez $8 \cdot 7 \cdot 6 \cdot 5 = 1680$ lehetőség. Ám így bármely kiválasztott négy csészét az összes lehetséges választási sorrendben figyelembe vettünk, az meg ugye nem számít (mindegy, hogy a sorban harmadik csészét elsőre vagy negyedikre választom ki). Mivel a 4 csészét (vagy bármit) $4 \cdot 3 \cdot 2 \cdot 1 = 24$ -féleképpen lehet sorba rakni, így $1680/24 = 70$ -féleképpen lehet nyolc csészéből négyet kiválasztani. Ha Muriel mind a nyolc csészéről meg tudja mondani, hogy mit töltöttek bele előbb, azt csak egyféleképpen teheti meg, tehát annak az esélye, hogy ez csupán a véletlen műve, $1/70 \approx 0,014$. Azaz, ha mind a négyet eltalálja, akkor a $p = 0,05$ -ös szignifikanciaszint alapján a nullhipotézis hamis, a hölgy rendelkezik a mágikus képességgel. Na de mi van, ha hármát eltalál, de egyet téved. Ekkor mind a két csoportban egy-egy hibás találat lesz, hiszen két csésze besorolása fel van cserélve. Nézzük most a „tea először” csoportot, ahova három „tea először” csészét sorolt be a négyből. Ezt négyféleképpen teheti meg ($(4 \cdot 3 \cdot 2)/(3 \cdot 2)$), ettől függetlenül ugyanez igaz a „tej először” csoportra. Azaz összesen $4 \cdot 4 = 16$ összeállításban történhet meg a három találat. Ennek az esélye $16/70 \approx 0,23$, azaz pusztán véletlen találgatással durván 23% az esélye, hogy a négyből három csészét a jó csoportba sorol az ízlelő kísérleti alany. Pedig úgy érzésre meggyőző eredmény lenne három csészét jól besorolni, ám e kis számolás után nyugodt szívvel mondhatjuk, hogy aki erre képes, az valójában nem tudja megkülönböztetni, hogy mit töltöttek először a csészébe, pusztán mérsékelt szintű szerencséje volt.

Bár már eddig is elég hosszan írtam erről a kísérletről, még nem zárom le az elemzését, mert vannak még itt érdekességek. Tegyük fel, hogy nem nyolc, hanem csak hat csészével hajtjuk végre a kísérletet (háromban a tea, másik háromban a tej volt először beöntve). A korábbi számolást 6 csészére elvégezve kiderül, hogy $(6 \cdot 5 \cdot 4)/(3 \cdot 2) = 20$ -féleképpen lehet két csoportra osztani a hat csészét. Tegyük fel, hogy kísérleti alanyunk tökéletesen tippeli meg, hogy melyik csésze melyik csoportba tartozik, amit most is csupán egyféleképpen lehet megtenni. Ám ennek a valószínűsége $1/20 = 0,05$, ami alapján még a hibamentes osztályozás is 5% eséllyel történik meg, pusztán a véletlen folytán. Azaz a kísérlet nincs jól megtervezve, mert így nem lehet eldönteni a kísérleti személy képességeit. És mi van, ha nem nyolc, hanem 12 csészénk van 6-6 mind a

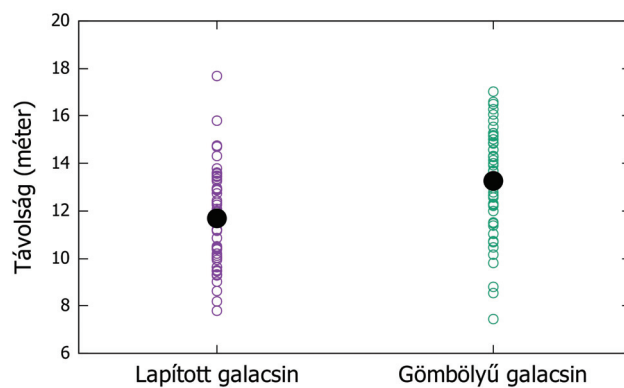
két beállításból. Ilyenkor összesen 924-féleképpen lehet két hatos csoportot csinálni (a számolás hasonló, mint a 8 csészénél). Tegyük fel, hogy a kísérleti alany öt csészét jól csoportosít, egyet-egyét meg rosszul. Ez 36-féleképpen történhet meg, azaz, annak az esélye, hogy ötöt jól, egyet rosszul csoportosít valaki pusztán véletlenszerű választással, $36/924 \approx 0,039$, ami kisebb, mint 0,05. Tehát, ebben az esetben a nullhipotézisünk hamis lenne.

Nézzünk ezután egy másik, első látásra a valós problémáktól elég távolinak tűnő kérdést, amiről aztán kiderül, hogy nagyon általános problémát mutat be, és csak statisztikus szemlélettel lehet a kérdésre választ adni.

A LAPÍTOTT VAGY A GÖMBÖLYŰ PAPÍRGALAC SIN REPÜL MESSZEBBRE?

Tegyük fel, hogy egy nagy terem egyik sarkában 200, körülbelül hógolyó méretű papírgalacs sin van. Színben, a papír vastagságában, méretben is különbözöek. Azt a feladatot kapjuk, hogy kísérletes úton mutassuk meg, hogy a lapos és a gömbölyű papírgalacs inok nem repülnek ugyanolyan messzire. A feladat megoldásához kapunk egy nagyobb csúzlit, amit egy stabil állványra szereltek fel, meg egy 2 méteres mérőszalagot, 1 mm-es beosztással.

A megoldáshoz a következő kísérleti eljárást javaslom: első lépésben bontsuk véletlenszerűen két csoportba a galacs inokat. Például úgy, hogy csukott szemmel megfogunk egy galacs inot, majd feldobunk egy pénzérmét. Ha fej, akkor a jobb oldali, ha írás, akkor a bal oldali kupacba tesszük. Ha ezzel megvagyunk, akkor kiválasztjuk ugyancsak pénzfeldobással, hogy melyik csoportban lévő galacs inokat fogjuk „kezelés alá” vonni. A kezelés az lesz, hogy ezeket a galacs inokat a tenyerünkbe fogva összelapítjuk. Igyekszünk minden galacs inot azonos vastagságú koronggá lapítani, mondjuk legyen ez a vastagság 10 mm, amit a mérőszalaggal gondosan meg is mérünk. Ezután a csúzliba behelyezett gömbölyded és lapított galacs inokat elkezdjük kilödni, ügyelve arra, hogy a csúzlit minden esetben azonos mértékig húzzuk ki. (Ebben szintén segít a mérőszalag.) Mivel a csúzli gumija elfárad a kísérlet során, és a terem hőmérséklete is változhat, ami hat a gumiszalag keménységére, ezért ebben a lépésben is ajánlatos pénzfeldobással véletlenszerűen választani a kezeletlen és a kezelt galacs inok közül. Minden lövés után lemérjük, hogy a galacs in milyen távolságra ért földet a csúzlitól. Ilyen körülmények között ezt a mérést a legnehezebb pontosan végrehajtani. Javítja a mérés pontosságát, ha van egy társunk, aki közelről figyeli, hogy hova esik a galacs in, és gyorsan meg is jelöli annak a helyét a padlón. Tegyük fel, hogy gondosan végrehajtottuk a kísérletet, aminek eredményét a 3. ábra szemlélteti. Nagyon nem lehetünk meglepödvé a sokféle mért távolságon, hiszen rengeteg nem kiküszöbölhető bizonytalanság volt a kísérletben. De mit kezdünk ezzel? Ha kiszámoljuk az átlagos repülési távolságokat, akkor az 11,68 méter lesz a lapított és 13,26 méter a gömbölyű galacs inok esetében.



genetikai háttér, a fejfájás oka, stb. Legalább annyira sokfélék a kísérlet alanyai, mint a gömbölyded galacsinok, sőt! Aztán kapnak egy kezelést (lapított galacsin az előző esetben, gyógyszer most), majd meg szeretnénk nézni a hatását a kezelésnek. Míg a galacsinok repülési távolságát, ha nem is nagyon pontosan, de egyértelműen tudjuk mérni, addig a fájdalomról a betegek számolnak be valamilyen szubjektív skálát használva. De még ha mérni is tudnánk, ahogy pl. a vércukorszintet tudjuk, akkor is jelentősen torzítja az eredményt a placebo hatás, azaz annak a tudata, hogy a beteg gyógyszert kap. Ez mérhetően növeli mind az érzést, mind a valódi hatását a gyógyszernek. Ez utóbbi azért következik be, mert a gyógyszer hatékonyságába jobb lelkiállapotba hozza a beteget, akinek ettől jobban működik az immunrendszere és az egész anyagcseréje, ami minden betegség esetén segíti a gyógyulást. Szóval a placebo hatást ki kell zárni valahogy. Idővel azt is észrevették a kutatók, hogy a kísérlet irányítói akaratlanul is gondosabbak azokkal a betegekkel, akik kapják az új szert, mint azokkal, akik nem kapják. Emberi dolog, ők is szeretnék, hogy valami igazán hatékony szer legyen a kezükben. És ez is befolyásolja a kísérlet eredményét. Ki kéne tehát zárni ezt a két jelentős torzító hatást.

Ehhez az úgynevezett *kettős vak próba* módszert kell alkalmazni. Természetesen most is véletlenszerűen két csoportra bontjuk a vizsgálandó betegeket, és mind a két csoport ugyanolyan módon kapja meg a kezelést, azonban az egyik csoport gyógyszerében nem lesz hatóanyag. Hogy ez melyik csoport (vagy aktuálisan melyik páciens), *azt sem a páciens, sem a kísérletben a páciensekkel kapcsolatban lévő szakember nem tudja*. Ezzel biztosítjuk, hogy mind a két irányból jövő torzítás megszűnjön. Ahhoz, hogy a szer hatékonyságát a lehető legnagyobb biztonsággal tudjuk vizsgálni, a kettős vak próba módszert kell alkalmazni, és a gyógyszerkutatásban minden esetben ezt is alkalmazzák. Fontos megjegyezni, hogy a részrehajlást (esetleges csalást) elkerülendő a gyógyszerek klinikai fázisban való tesztelését minden esetben kötelezően külső cégek végzik, akik nagyon szigorúan követik a protokollt. Őket aztán nem befolyásolja, hogy a gyógyszergyár mennyi pénzt fektetett egy fejlesztésbe, őket azért fizetik, hogy megbízhatóak és pontosak legyenek.

Még ennél is átgondoltabb kísérlettervezést és az adatok még részletesebb statisztikai elemzését kívánja meg, ha arra is választ keresünk, hogy mennyire hatékony egy gyógyszer vagy oltás, és milyen mellékhatások, milyen gyakran figyelhetők meg a használata során.

A csodaszerekből, titkos terápiákból és „természetes” gyógymódokon meggazdagodó mágusok eljárásai soha nincsenek szigorúan tesztelve, sőt általában egyáltalán nincsenek tesztelve. Így aztán bízni bennük nem tanácsos. Az igazi természettudomány alapkövetelménye, hogy a kísérlet megismételhető s így ellenőrizhető legyen, és az alkalmazott statisztikai eljárás ne legyen hibás. Ezért aztán nem ritka, hogy sok olyan tudományos eredményt közölnek, melyről később kiderül, hogy vagy az adatgyűjtés vagy a statisztikai eljárás hibás volt, esetleg a cikk szerzői – eléggé el nem ítéhető módon – manipulálták az adatokat. Ezekon a buktatókon keresztül, de előre halad a tudomány.

HAMIS POZITÍV ÉS HAMIS NEGATÍV TESZTEK, AVAGY MIKOR AGGÓDJUNK

Amikor szűrővizsgálatra megyünk, jó, ha tudjuk, hogy a vizsgálat nem fogja teljes bizonyossággal megállapítani, hogy az adott betegségben szenvedünk vagy sem. Minden teszt kétféle hibát véthet: egyrészt megesik, hogy olyan pacienseket is betegnek ítélt, akik nem azok (hamis pozitív), másrészt bizonyos eséllyel nem tekinti betegnek azokat, aki valójában betegek (hamis negatív). Ennek a ténynek az a meglepő következménye, hogy még nagyon pontos tesztek esetén is sok egészséges páciens lesz a betegek közé sorolva az első teszt alapján, főleg, ha maga a betegség ritka.

Nézzünk egy példát. Tegyük fel, hogy a mammográfiai eljárás 90% eséllyel mutatja ki a tumort, amikor valóban ott van a betegben és 8% eséllyel jelez tumort akkor is, amikor nincsenek ott a tumorsejtek (hamis pozitív). Tudjuk, hogy a szűrésen résztvevők kb. 1%-a beteg, és tegyük fel, hogy 1000 nőt szűrnek ezzel a módszerrel, azaz átlagosan $10 (1000 \cdot 0,1)$ lesz beteg közülük. De a teszt csak 9-et $(10 \cdot 0,9)$ fog betegnek jelezni. Van továbbá 990 páciens, akik nem betegek, ám $990 \cdot 0,08 \approx 79$ esetben a teszt a betegek közé fogja őket sorolni. Tehát, ha valaki kap egy papírt, hogy az imént ismertett mammográfiai teszt alapján a lelete pozitív, akkor annak az esélye, hogy valóban van rákos elváltozás a mellében, csupán $10\% (9 / (79 + 9) \cdot 100)$! Tehát ilyen körülmények között egy teszt után még ne aggódjunk nagyon. Ám ha a teszt továbbra is 90% pontossággal mutatja ki a tumort, de csak 1% eséllyel mutat hamis pozitívat, akkor a betegnek jelzett páciensek már közel $48\%-a (9 / (9,9 + 9) \cdot 100)$ valóban beteg lesz. Minél ritkább egy betegség, annál fontosabb a pontos diagnózishoz, hogy a hamis pozitív eredmények esélye nagyon kicsi legyen!

ÖSSZEGÉS, KITEKINTÉS

Természetesen ez csupán egy rövidke ízelítő volt abból, hogy hogyan sajátíthatják el a középiskolás diákok a statisztikus szemléletmód alapjait. Anélkül, hogy az egyes statisztikai próbákkal fásasztanánk őket, megismerkedhetnek a korreláció és az ok-okozat különbségével, a nullhipotézissel és a szignifikanciaszinttel, a kettős vak próbával és a hamis pozitív és hamis negatív teszteredmények következményével. Mindehhez csupán elemi matematikai tudásra van szükség, és persze a lelkes tanár mellett némi tanszabadságra is.

A tanulmány bírálati folyamaton ment keresztül.

AJÁNLOTT IRODALOM

- Hans Peter Beck-Bernholdt, Hans-Hermann Dubben (2014): *A tojást rakó kutya*. Budapest, Magyar Könyvklub.
- Jordan Ellenberg (2014): *Hogy (ne) tévedjünk*. Budapest, Park Kiadó.