Gondolatok (!) a valószínűségelmélet és a statisztika oktatásáról

Kovácsné Székely Ilona
habilitált főiskolai tanár
BGE KVIK Üzleti Elemzés Módszertan Tanszék
E-mail: kovacsneszekely.ilona@uni-bge.hu
DOI: 10.29180/978-615-6342-61-4_20

Összefoglalás: A statisztika alapfogalmainak és módszereinek megértése és helyes alkalmazása elképzelhetetlen valószínűségszámítási alapok nélkül. Nagy jelentőségű a minta és a minta realizáció fogalma, a mintaközép eloszlásának ismerete, a függetlenség és a korrelálatlanság fogalma, továbbá az a „megállapítás", hogy a kétdimenziós normális eloszlás esetén a regressziós függvény lineáris. Igen hasznos a számítógéppel támogatott statisztika oktatása, de a kellő elméleti megalapozás hiányában, a hallgatóság egy jelentős része lemarad az oktatás folyamatában és önállóan képtelen egy ismeretlen adathalmazt elemezni, pedig a gazdasági szereplők ezt várják el a munkavállalóktól.

Kulcsszavak: valószínűségszámítás, függetlenség, korrelálatlanság, statisztika, oktatás.

Abstract: Understanding the basics of statistics and correctly applying statistical methods requires knowledge of fundamental probability theory. Such basics include the concept of random variables and their realizations leading to a sample of actual observations, the meaning of (in)dependence and (un)correlatedness, the central limit theorem for the distribution of the sample mean as well as the linearity of the regression function for two-dimensional normally distributed random variables. Teaching statistics in a software aided and application-oriented way is useful, but without necessary theoretical backgrounds, many students are facing severe difficulties over the learning process. In particular, students might not be able to analyse new real-life data, while this would be a primary expectation of future employers towards graduating students.

Keywords: probability theory, independence, uncorrelatedness, statistics, teaching

1. Gondolatok a valószínűségelmélet oktatásáról

Korpás Attiláné által szerkesztett ÁLTALÁNOS STATISZTIKA I; II könyvek szakmai lektora Csernyák László volt, aki nagyban hozzájárult a könyvek sikeréhez [1, 2]. Korrekt módon alapozza meg a szükséges valószínűségi hátteret. Érdemes ma is ezt a két könyvet lapozgatni és használni, mert jelentős segítséget nyújt a statisztika alapfogalmainak és módszereinek megértéséhez, melyek helyes alkalmazása elképzelhetetlen valószínűségszámítási alapok nélkül.

Van néhány fogalom, amelyet a hallgatók nehezen értenek meg és / vagy rosszul alkalmazzák azokat. Mivel külön tantárgyként nem jelenik meg a Valószínűségszámítás a tantervekben, a statisztika megalapozásához fontosnak tartom az alábbi elméleti kitekintés ismertetését előadáson, konkrét példák megmutatásával.
Az alábbiakban a matematikai kitekintést tartom fontosnak, hogy a hallgató lássa a Gazdasági matematikához a kapcsolódási pontot.

1.1 A statisztikai minta és a minta realizációja

X valószínűségi változóra vett minta legyen $(X_1,X_2, \ldots, X_n)$ ahol X_i-k független, azonos eloszlású valószínűségi változók
A minta realizációja az $(x_1,x_2, \ldots ,x_n)$ számértékek ahol $X_1=x_1$, $X_2=x_2$, … $X_n=x_n$ [3].

1.2 A centrális határeloszlás tétel alkalmazásai

Ha $X_1,X_2, \ldots , X_n$ azonos eloszlású független és véges szórású valószínűségi változók és
$E(X_i)=m \quad D(X_i)=\sigma \quad i=1,2,\ldots,n$

akkor $$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - nm}{\sigma\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-u^2}{2}}du \tag{1}$$

ahol $$E\left(\sum_{i=1}^{n} X_i\right) = nm \quad \text{és} \quad D^2\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2 \tag{2}$$

1.2.1 A mintaközép eloszlása
Ha az X valószínűségi változó várható értéke $E(X)=m$, szórása $D(X)=\sigma$ létezik és az n mintanagyság elég nagy akkor az X-re vonatkozó $(X_1, X_2, \ldots ,X_n)$ mintából számított

$$\overline{X}=\frac{\sum_{i=1}^{n} X_i}{n} \tag{3}$$

mintaközép *közel normális eloszlású* valószínűségi változó [3].
A mintaközép paraméterei

$$E(\overline{X}) = m \quad \text{és} \quad D(\overline{X}) = \frac{\sigma}{\sqrt{n}} \tag{4}$$

1.2.2 Standardizálás alkalmazása a mintaátlagra, Z-próbastatisztika
Legyen $(X_1,X_2,\ldots,X_n)$ minta, ahol X_i-k független, azonos eloszlású valószínűségi változók és

$$E(\overline{X}) = m \quad \text{és} \quad D(\overline{X}) = \frac{\sigma}{\sqrt{n}} \tag{5}$$

A mintaátlag standardizáltja a z próbastatisztikát adja meg várhatóértékre.

$$Z=\frac{\overline{X}-E(\overline{X})}{D(\overline{X})} \tag{6}$$

$$Z=\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \qquad (7)$$

ahol E(Z)=0 és D(Z)=1 (8)

1.2.3 Moivre-Laplace tétel

Végezzünk n független kísérletet, amelynek lehetséges kimenetelei az A és az $\bar{A}$ események

$$P(A)=p \quad \text{és} \quad P(\bar{A}) = 1 - p \qquad (9)$$

$Y_n=X_1+X_2+\ldots+X_n$ ahol X_k az A esemény indikátor változói. (10)

$E(X_k)=p$ és $D^2(X_k)=p(1-p)$ $k=1,2,\ldots,n$ (11)

$E(Y_n)=np$ és $D^2(Y_n)=np(1-p)$ (12)

Y_n binomiális eloszlású valószínűségi változó, ha n elég nagy akkor közelíthető normális eloszlással.

$$\lim_{n\to\infty} P\left(\frac{Y_n-np}{\sqrt{np(1-p)}} < x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-u^2}{2}}du \qquad (13)$$

1.2.4 Standardizálás alkalmazása relatív gyakoriságra

$\frac{Y_n}{n}$ változó standardizáltja a Z-próbastatisztikát eredményezi, arányra.

$$Z=\frac{\frac{Y_n}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \qquad E(Z)=0 \text{ és } \quad D(Z)=1 \qquad (14)$$

ahol

$$E\left(\frac{Y_n}{n}\right) = \frac{1}{n}\cdot np = p \quad \text{és} \quad D^2\left(\frac{Y_n}{n}\right) = \frac{1}{n^2}\cdot np(1-p) = \frac{p(1-p)}{n} \qquad (15)$$

1.3 A függetlenség és korrelálatlanság fogalma

1.3.1. Példa

Az (X,Y) vektorvalószínűségi változó egyenletesen veszi fel az (1,1);(1,0);(0,1);(1,2) értékeket.

a. Független-e X és Y valószínűségi változó?

b. Határozzuk meg ρ(X,Y) korrelációs együttható értékét!

X/Y	**0**	**1**	**2**	**pi**
0	0	0,25	0	0,25
1	0,25	0,25	0,25	0,75
qj	0,25	0,50	0,25	1

a. $P(X=0, Y=0)=0 \neq P(X=0)\cdot P(Y=0)=0{,}25\cdot 0{,}25$ (16)

X és Y valószínűségi változók nem függetlenek.

b. $E(X)=0{,}75 \quad E(X^2)=0{,}75 \quad D^2(X)=0{,}1875$ (17)

$E(Y)=1 \quad E(Y^2)=1{,}5 \quad D^2(Y)=0{,}5 \quad E(XY)=0{,}75$ (18)

$$\rho(X,Y)=\frac{E(XY)-E(X)E(Y)}{D(X)D(Y)}=\frac{0{,}75-0{,}75\cdot 1}{\sqrt{0{,}1875\cdot 0{,}5}}=0 \qquad (18)$$

Mivel $\rho=0$ az X és Y valószínűségi változók korrelálatlanok.

1.3.2. A függetlenség és korrelálatlanság fogalma kétdimenziós normális eloszlás esetén

Az (X,Y) vektorvalószínűségi változó normális eloszlású, ha sűrűségfüggvénye

$$h(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2}-2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right]\right\} \qquad (19)$$

Ha az X és Y változók korrelálatlanok $\rho=0$

$$h(x,y)=\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left\{-\frac{(x-m_1)^2}{2\sigma_1^2}\right\}\cdot\frac{1}{\sigma_2\sqrt{2\pi}}\exp\left\{-\frac{(y-m_2)^2}{2\sigma_2^2}\right\} \qquad (20)$$

$h(x,y)=f(x)\cdot g(y)$ X és Y valószínűségi változók függetlenek

$X \sim N(m_1,\sigma_1)$ és $Y \sim N(m_2,\sigma_2)$ (21)

A kétdimenziós normális eloszlás esetén, ha a változók korrelálatlanok akkor függetlenek is. Ha a peremeloszlások normálisak nem feltétlenül következik, hogy az együttes eloszlásuk kétdimenziós normális [4].

1.4 Regressziós függvény kétdimenziós normális eloszlás esetén

A kétdimenziós normális eloszlás esetén a regressziós függvény lineáris. Ismert, hogy az alábbi minimum feladat megoldása a feltételes várhatóérték.

$E(Y-g(x))^2=\min$ (22)

$g(x)=E(Y|X=x)$ (23)

A feltételes várható érték meghatározható a feltételes sűrűségfüggvényből.

$$g(y|x) = \frac{h(x,y)}{f(x)} \quad (24)$$

Az Y változó X-re vonatkozó feltételes várható értéke

$$E(Y|X=x)=m_2 + \rho\frac{\sigma_2}{\sigma_1}(X - m_1) \quad (25)$$

$$g(y|x) = \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[[Y - (m_2 + \rho\frac{\sigma_2}{\sigma_1}(X - m_1)\right]^2\right. \quad (26)$$

A regressziós függvény lineáris $E(Y|X=x)=m_2 + \rho\frac{\sigma_2}{\sigma_1}(X - m_1)$ (27)

2. Gondolatok a statisztika oktatásártól

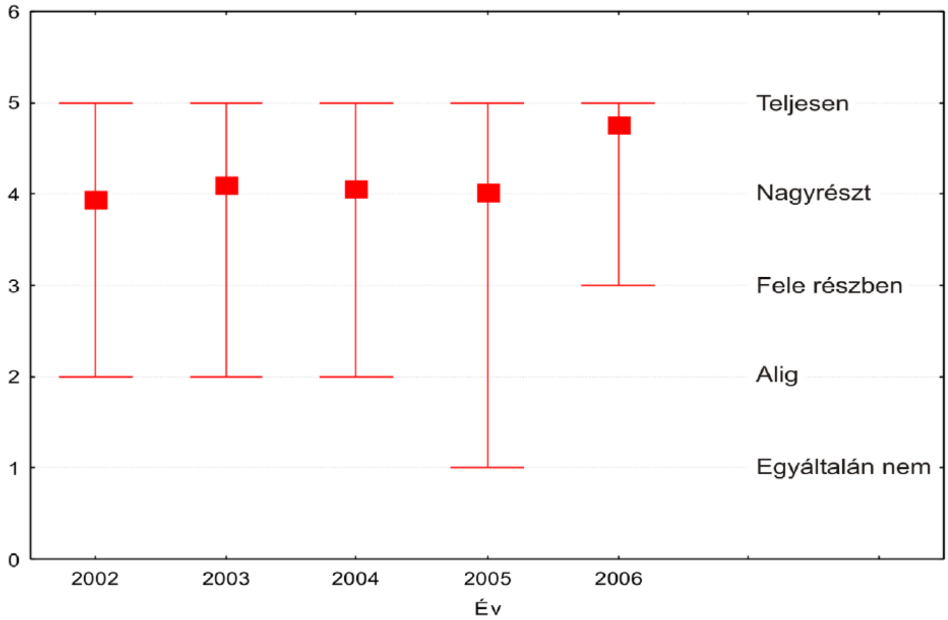
2.1 A statisztika oktatás célja

A statisztika oktatásának célja olyan szakember képzése, aki a gazdasági döntési folyamatban képes alkalmazni a statisztikai módszereket és a vizsgálat eredményeit alkotó módon felhasználni. További lényeges szempont, hogy a tárgy adatfeldolgozó képességeket alakítson ki a hallgatókban, nem pedig egyszerű mechanikus számolási és műveleti képességet.
Az előadásokon és a gyakorlatokon a hallgató egy mintát dolgoz fel, melynek alapján nem érzékeli a változékonyság problémáját és fontosságát. Nem tudatosul benne az a tény, hogy a statisztikák valószínűségi változók. A statisztikai módszerek alkalmazási feltételeinek ellenőrzése nem történik meg, ezeket legtöbbször megadják a feladatban. Ennek súlyos következménye van, ugyanis nem alakul ki az a képesség, hogy a sokaságra vonatkozóan, a hallgatók helyes következtetéseket tudjanak levonni. Igen hasznos a számítógéppel támogatott statisztika, de a kellő elméleti megalapozás hiányában, a hallgatóság egy jelentős része lemarad az oktatás folyamatában és önállóan képtelen egy ismeretlen adathalmazt elemezni.

2.2 A statisztika oktatásának módszertanáról

Bármelyik oktató, aki saját maga által kitalált és nem életszerű adatokkal szeretne megértetni, elmagyarázni valamit diákjai számára, a saját diákjait hozza hátrányos helyzetbe [5]. Ahhoz, hogy megtudjuk, mennyire járunk jóúton az oktatás metodikájában, szükség van olyan módszerek kidolgozására [6], amelyek segítségével mérni lehet az adott oktatási folyamat hatását a tanulók megértési és tanulási folyamatair
Az ELTE TTK-n 2003-ban bevezettem a statisztika programcsomaggal támogatott oktatását valós adatokon [7]. Öt éven át mértem az eredményeket. 2002-ben csak hagyományos oktatás volt majd 2003-2005 között kisebb, 2006-ban nagy metodikai módosításokat hajtottam végre.

Tapasztalataim szerint „a hatékony gyakorlatok alapjait körültekintő folyamat során kiválasztott feladatok jelentik. Amennyiben sokféle feladatot szerepeltetünk az órákon, a hallgatónak mindegyik esetében meg kell értenie a szakmai problémát, és rendelkeznie kell a szükséges szakmai ismeretanyaggal. Felméréseim azt bizonyítják, célszerű kevés számú – amennyiben lehet egy – feladatot választani, amely az adott szakterületről való, tudományos igényű, érdekes és rajta szinte minden, vagy lehetőség szerint minél több tanult statisztikai módszer alkalmazható. Ekkor csak egyszer kell a hallgatónak egy ilyen szakmai példának az ismeretanyagát, feltételrendszerét magáévá tenni, így kellőképpen a statisztikai fogalmakra, eszközökre és azok alkalmazásának feltételeire tud koncentrálni” [8]. A 2006-ban bevezetett nagy és összetett adathalmazon történő oktatás sikerességét az 1. ábra szemlélteti a hallgatóknak feltett kérdésre - „Mennyivel tartaná hatékonyabbnak tanulását, megalapozottabbnak tudását, ha a tárgyat szakmai példák alkalmazásán keresztül tanulná?” - adott válasszal.



1.ábra

Mennyivel tartaná hatékonyabbnak tanulását, megalapozottabbnak tudását, ha a tárgyat szakmai példák alkalmazásán keresztül tanulná? *forrás: [7]*

Tanulságos volt megismerni a hallgatók véleményét erről az oktatási formáról. Tekintsünk meg néhányat ezek közül [7]:

„A gyakorlat alapvető hasznának a szakmai problémák statisztikai kiértékelését látom. Ezáltal látásmódunkat, gondolkodásmódunkat a tárggyal és a szakmával kapcsolatban átértékeli, segít mélyebben megismerni. Külön értékének tekintem, hogy a statisztika elmélet fogalmait is a hallgatóság számára szemléletessé, kézzel foghatóvá teszi.”

„A példák nagyon jók, így az igazi egy tantárgyat tanulni, hogy nemcsak a tankönyv fiktív, steril példáit boncolgathatja az ember, hanem hogy végre szembe kerül valami „életszagúval”.
„Kifejezetten jó, hogy az órán szereplő feladatok és adatsorok igaziak, az életből szedett példákon alapszanak. A tárgy sokat segít a szakmai fejlődésben” [7].
Ez a 2003-ban lefolytatott kísérlet is rámutat arra, hogy mennyire fontos az adatok értelmezése, és a kiértékelés módja. Manapság gyakorlatokon Excelben oldunk meg kész problémát, de az eszközhasználat mögött „elveszik” a statisztikai látásmód.
A pandémia idején módszertanunk bővült kis videókkal, melyekben Excellel végezzük a kiértékelést. Ugyanakkor a feljövő generáció Excel készségeit is fejleszteni kell, így a statisztikai gondolkodásmód megintcsak sérül.

2.3 A megvalósítás eszközei és javaslatok

2.3.1. Oktatás
“*KOMBINÁLT” oktatási forma*: ami online előadást és hagyományos gyakorlatot jelent. Az online előadások nem sikeresek, a hallgatók figyelme elkalandozik, az oktató számára nincs visszajelzés. A gyakorlatokra teljesen felkészülés nélkül érkeznek a hallgatók.
Hallgatói feltétel: A hallgató rendelkezzen megfelelő számítástechnikai, valószínűségszámítási ismerettel. Ez a feltétel csak részben teljesül.
Személyi feltétel: Ez feltételezi a jól felkészült oktatót, aki vonzóvá tudja tenni a hallgatók által absztraktnak ítélt fogalmakat és módszereket valós feladatok megoldásával, elemzésével képes közelebb hozni a hallgatót választott szakmájához. Az oktatónak szükséges közvetítenie tanítványai felé, azt, hogy ne az legyen a hallgató egyetlen célja, valahogy teljesítse a tárgyat, hanem a tudását gyarapítsa, ismereteit sikeresen tudja alkalmazni és szakmai tovább fejlődéséhez ez adjon egy jó alapot.
Technikai feltétel: jól felszerelt, megbízhatóan működő számítógépes labor.

2.3.2. Tematika

Figyelembe véve a tárgytól elvárt követelményeket és a hallgatóink felkészültségét, hozott alapokat, a tanultak alkalmazásának esélyét, véleményem szerint a BA szakon kevesebbet kellene tanítani, de azt sokkal részletesebben. Az oktatás folyamatában fontos lenne visszatérni a megelőző gyakorlaton megismert módszerekre, visszajelzést kapni a hallgatóktól a tanulási és megértési problémáikra. Lényeges lenne bevezetni minden nagyobb fejezet után egy komplex feladat megoldását és elemzését gazdasági környezetben, azaz egy kisebb kutatást elvégezni. Nagy hangsúlyt kell / kellene helyezni a kapott eredmények szakmai környezetben történő kiértékelésére. Ma az alkalmazott online vizsgáztatásnál ez háttérbe szorul és ezt nagy hibának tartom.

Több olyan paraméteres próbát tanítunk, amelyek nem alkalmazhatók a normális eloszlás hiánya miatt. Ezért olyan nemparaméteres próbák bevezetését javaslom, melyek alkalmazásának nem feltétele a normális eloszlás.
Hibának tartom, hogy idő hiányában nem ellenőrizzük a különböző próbák alkalmazhatóságának feltételrendszerét és a regressziószámításnál a reziduumok vizsgálatát.
Ma már akár Excellel is meghatározható a próbák szignifikanciája azaz a p-érték. Javaslom minden alkalmazott próba esetén a döntést ennek alapján történjen.
A lelkesebb hallgatók a TDK dolgozat megírásánál és/vagy a szakdolgozat elkészítésénél szeretnék alkalmazni a tanult statisztikai módszereket. Ilyenkor kérnek segítséget és szembesülnek azzal, hogy a tanult statisztikai módszereket alig vagy egyáltalán nem tudják alkalmazni. Felmerül a kérdés, ennek mi lehet az oka? Véleményem szerint ilyen esetben nem elegendő csak a hallgatók gyenge tudás szintjére hivatkozni, mert a TDK-ra a jobb képességűek, elhivatottabbak és szorgalmasabbak jelentkeznek.
Hol van a hiba? Melyek azok a tények, feltételek, melyeken változtatni kellene? Ehhez egy közös gondolkodásra lenne szükség. A tematika és tantárgyleírás összeállításánál ezeket is figyelembe kellene venni. Nem kell mindent megtanítani a BA képzésen. A kevesebb lenne a több, de ezzel adhatnánk alkalmazható tudást a hallgatónak. Oktatásunkban azt is lényegesnek tartom, ne csak a lemorzsolódás megakadályozása legyen a cél.
Remélem a statisztika módszertanának folyamatos fejlesztésével, átgondolásával és igényes oktatásával sikerül közelítenünk a 2.1 pontban megfogalmazottakat, a statisztika oktatás céljait.

2.3.3.Idézet a „Könyvszemlé”-ből

Végül felhívnám a figyelmet egy igen értékes könyvszemlére, melynek minden mondatával egyetértek, nem kívánok hozzátenni semmit, csak idézni [9].
„Ez a tankönyv azzal a céllal íródott… Statisztika könnyen, gyorsan, nehézségek nélkül, matematika nélkül, formulák nélkül… A szlogen ismerős, ám a magyar nyelvű szakirodalomban tudomásom szerint ilyen kiadvány még nem jelent meg, ezért talán érdemes egy kicsit részletesebben megismerni a tartalmát.”
„Az oktatónak úgy hiszem nem az lenne a feladata, hogy idomuljon a környezet igénytelenségéhez, hanem oktassa, nevelje, tanítsa meg a hallgatókat tanulni, dolgozni. Ebből az e-book koncepcióból annyi racionális magot természetesen elfogadok, miszerint egy oktatási anyag lehetőleg kerülje az öncélú tudálékoskodást, próbáljon meg világosan, célratörően, tömören, a rendelkezésre álló technika adta lehetőségek maximális kihasználásával magyarázni, de ne sugallja azt, hogy kevés

munkával is el lehet jutni komoly tudásig. Tudomásul kell venni, hogy a tanulás (a hatékony tanulás) igen nehéz (egyesek szerint a legnehezebb) szellemi munka. Az egyetemi hallgatóknak a társadalomban elfoglalt kiváltságos helyzete mögött az van, hogy a társadalom tudomásul veszi ezt a nehéz munkát. Félő, hogy az ilyen és hasonló – a tudás látszatát biztosító, valójában áltudást adó – könyvek visszaélnek ezzel, és meggyőződésem szerint hosszú távon jelentős károkat okoznak a fiatal generáció fejlődésében és általában az oktatás társadalmi megítélésében." [9]

Irodalomjegyzék

[1] Havasy Gy., Molnár M.né., Tóth M.né., Szunyogh Zs.: *Általános statisztika I.*, Szerkesztette: Korpás A.né, Nemzeti Tankönyvkiadó, 1996;

[2] Sándorné Kriszt É., Varga E., Vitézné Kenyeres E.: *Általános statisztika II*., Szerkesztette: Korpás A.né, Nemzeti Tankönyvkiadó, 1997;

[3] Kovácsné Székely I.: *Understanding the spatiotemporal sample: a practical view for teaching geologist students*, Teaching Mathematics and Computer Science, .4/1, 89–99,2006; https://doi.org/10.5485/TMCS.2006.0104

[4] G.S. Maddala: *Bevezetés az ökonometriába*, Nemzeti Tankönyvkiadó, 2001;

[5] Bailar, J. C.,: *A Larger Perspective*. The American Statistician, 49, pp. 10–11.,1995; https://doi.org/10.1080/00031305.1995.10476103

[6] Mills J. D.,: *Using Simulation Methods to Teach Statistics*: Review of the Literature. Journal of Statistics Education [Online], 10(1) 2002; https://doi.org/10.1080/10691898.2002.11910548

[7] Kovácsné Székely I: *Programcsomaggal támogatott statisztika oktatása geológus hallgatók példáján*, Doktori (Ph.D.) disszertáció, pp.: 117, 2007;

[8] Kovácsné Székely I: *Programcsomaggal támogatott statisztika oktatása geológus hallgatók példáján*, Doktori (Ph.D.) disszertáció tézisei, pp.:18 2007;

[9] Hunyadi L. Könyvszemle Salkind, N. J.: *Statisztika olyanoknak, akik (azt hiszik) gyülölik a statisztikát Statistics for People Who (Think They) Hate Statistics*. (2nd Edition) Sage Publications, Inc. Thousand Oaks, London, New Delhi. 2004. Statisztikai Szemle, 89. évfolyam 7—8. szám.