Matematika III. 6.

A statisztika alapfogalmai

Prof. Dr. Závoti , József

Matematika III. 6. : A statisztika alapfogalmai

Prof. Dr. Závoti , József
Lektor : Bischof , Annamária

Ez a modul a TÁMOP - 4.1.2-08/1/A-2009-0027 „Tananyagfejlesztéssel a GEO-ért” projekt keretében készült. A projektet az Európai Unió és a Magyar Állam 44 706 488 Ft összegben támogatta.

v 1.0

Publication date 2010

Kivonat

Ez a modul a statisztika alapfogalmaival ismerteti meg az olvasót.

Tartalom

6. fejezet - A statisztika alapfogalmai

1. 6.1 Bevezetés

Jelen modul a Matematika III. tárgy hatodik fejezete, modulja. Az itt következő ismeretek megértéséhez javasoljuk, hogy olvassa el a Tárgy korábbi moduljainál írottakat. Amennyiben ez még nem lenne elég a megértéshez, akkor forduljon a szerzőhöz segítségért.

Jelen modul célja, hogy az Olvasó megismerkedjen a statisztika alapfogalmaival, és képessé váljon azok összetettebb számítási feladatok megoldásában való felhasználására.

A valószínűségszámítás tárgyalása során megállapítottuk, hogy a véletlen jelenségek valószínűségi eloszlásokkal, illetve az eloszlást jellemző mennyiségekkel (várható érték, szórás stb.) írhatók le. A valóságban a véletlen jelenségek eloszlása és az eloszlást jellemző paraméterek nem ismertek, hanem azokat mérésekből, mintavételekből kell meghatározni.

2. 6.2 A statisztika alapfogalmai és főbb feladatai

2.1. 6.2.1 Leíró és következtetéses statisztika

A **matematikai statisztika** a véletlen tömegjelenségek statisztikai törvény-szerű-ségeit vizsgálja.

A statisztika egyszerűbb problémáit a **leíró statisztika** keretein belül lehet kezelni:

- Adatok ábrázolása
- Grafikonok szerkesztése
- Táblázatok készítése
- Egyszerű paraméterek számolása (átlagértékek, szóródás)
- Indexszámítás
- Koncentráció-számítás

A matematikai statisztika gyakorlati használhatósága a valószínűségszámítás elméleti alapjain nyugszik. **Minőségellenőrzés** során sokszor nincs mód a teljes sokaságot átvizsgálni, hanem csak egy n-elemű **minta** alapján kell következtetéseket levonnunk. A mintaelemekből célszerű olyan függvényeket konstruálni, amelyek jó információt nyújtanak az egész eloszlásra. Tapasztalati adatokból, u.n. mintából következtetünk események valószínűségeire, vagy valószínűségi változók ismeretlen eloszlás-és sűrűségfüggvényeire.

A statisztika másik nagy területe az **induktív statisztika** (következtetéses statisztika):

- Becslések
- Tesztek
- Döntéselmélet
- Többváltozós statisztikai módszerek

A statisztika felhasználási területe az **extrapoláció** (predikció): A jelenlegi adatok alapján a jövőre nézve statisztikai prognózisokat lehet készíteni, feltételezve, hogy a feltételek azonosak maradnak. Ilyen prognózisok készülnek a következő évi energia felhasználásra, az adóbevételre, a népességszám alakulására, a munkanélküliségre, stb.

2.2. 6.2.2 A statisztikai adatok forrásai és feldolgozása

Az **adatok forrása** szerint megkülönböztetünk:

- hivatalos statisztikai adatokat, amelyeket a Központi Statisztikai Hivatal (KSH) évkönyvben, folyóiratokban tesz közzé
- nem hivatalos statisztikai adatokat, ilyenek az ipari és kereskedelmi kamarák jelentései, a különböző közvélemény-kutató intézetek felmérései, a nagy vállalatok mérlegei.

A statisztikai **adatfeldolgozás lépései**:

1. Tervezés
2. Mintavétel (elsődleges - másodlagos)
   - Kérdőív: olcsó – de általában kevés jön vissza
   - Interjú:drága – kvalifikált személyek szükségesek a felméré-sekhez
   - Megfigyelés: pl. forgalomszámlálás
   - Kísérlet: pl. a közgazdaságtanban az áruteszt
   - Automatikus rögzítés: vonalkódok a bevásárlóközpontokban vagy a telefonközpontok működése
3. Előkészítés: táblázat – grafikon szerkesztése
4. Analízis: matematikai statisztikai módszerek bevetése
5. Interpretáció: eredmények értékelése

Alapadatok: Sokaság (populáció)

Pl.: egy cég számlái 2009. szeptember 10-én, halálos balesetek száma 2008-ban

2.3. 6.2.3 A statisztikai sokaság

Sokaság: a vizsgálat tárgyát képező egységek összessége.

Csoportosítási lehetőségek:

1. Álló (időpont) és mozgó (időtartam) sokaság:

Pl.: Álló sokaság:

- Magyarország lakossága 2000 jan. 1.
- Raktár állománya 2009 szept. 10.
- Pénztári bevétel 2009 szept. 10.

Mozgó sokaság

- Születések száma 1 év alatt
- Egy bankba 1 nap alatt befizetett csekkek

1. Aggregált sokaság: Különböző fajta, minőségileg eltérő, de együtt vizsgált elemek

Aggregátum: értékben megadott mennyiség. Pl. húsfogyasztás 2009-ben.

1. Alapsokaságot és **részsokaságot** (például mintavétel).

Például: Magyarország összes háztartása mennyi mosóport használ?

2.4. 6.2.4 Statisztikai ismérvek

Definíció:

Ismérvek (karakterisztikus tulajdonságok): olyan vizsgálati szempontok, amelyek alapján egy sokaság át nem fedő részekre bontható.

Példa:

Ismérv:	Ismérvadat:
kor	év
nem	férfi
magasság	cím
kereset	Ft

Ismérvek fajtái:

- időbeli
- területi
- minőségi
- mennyiségi.

2.5. 6.2.5 Statisztikai skálák, információ szintek:

1. Nominális skála: nincs természetes sorrend, mellérendeltség

Pl.: vallások, nemek, színek

1. Ordinális (Rang, sorrendi) skála: van sorrend, létezik rendezés

Pl.: iskolai jegyek, futball bajnokság

1. Intervallum skála: nullpont választása önkényes

Pl.: hőmérséklet (20°C$\neq$2 $\cdot$ 10°C), időszámítás

1. Arány skála**:** létezik abszolút 0 pont

Pl.: magasság, kor, jövedelem

Az adatokat transzformációnak vethetjük alá úgy, hogy a meglévő viszonyok nem változnak.

2.6. 6.2.6 Diszkrét és folytonos ismérv adatok

Megkülönböztethetünk **diszkrét** és **folytonos ismérv** adatokat.

Diszkrét ismérv pl.: hallgatók száma, üzem dolgozói.

Folytonos ismérv: egy asztal hossza, súly stb.

2.7. 6.2.7 A matematikai statisztika fő területei:

1. **Becsléselmélet**: A valószínűségi eloszlások jellemzői mennyiségeinek meghatározását paraméterbecslésnek nevezzük.

Példa: a mintában található selejtarány alapján következtetünk az egész sokaságban valószínűsíthető selejtszámra.

1. **Hipotézisvizsgálat**: A valószínűségi változó eloszlására feltevéseket teszünk, azaz statisztikai hipotézist állítunk fel és matematikai statisztikai módszerekkel döntünk a hipotézis elfogadásáról vagy elvetéséről.

Példa: Egy adott gyártási technológia során meghatároztuk a gyártást jellemző paramétereket. Bizonyos idő elteltével azonban ellenőriznünk kell, hogy a gyártási feltételek megváltoztak-e vagy sem, azaz a paraméterek megegyeznek-e a korábbi értékekkel.

1. **Konfidencia intervallum becslés**: Mivel a becsléssel kapott érték általában nem azonos a keresett elméleti értékkel, ezért műszaki biztonsági okokból szükséges, hogy alsó és felső határt adjunk meg a becsült paraméterre.

Példa: a mintaátlag körül nagy valószínűséggel milyen intervallumban található az elméleti várható érték.

1. **Illeszkedésvizsgálat**: adott mintabeli eloszlásfüggvény milyen elméleti eloszlásfüggvényhez illeszkedik kielégítően.
2. **Homogenitásvizsgálat**: Állítható-e két valószínűségi változóról, hogy egyforma eloszlású?
3. **Korrelációanalízis**: mérési eredmények alapján próbáljuk eldönteni, hogy milyen összefüggés áll fenn két valószínűségi változó között.

3. 6.3 Gyakorisági eloszlások

Legyen adott N elemű sokaság metrikus skálán.

Legyenek adottak a következő ismérvek: $x_1, x_2, \dots, x_k$

Jelölje $f_1, f_2, \dots, f_k,$ ahányszor x_i előfordul

Abszolút gyakoriság: $f_i \; i = 1, 2, \dots, k \; 0 \le f_i \le N \,|\, f_1 + f_2 + \dots + f_k = N$

Relatív gyakoriság: $g_i = \frac{f_i}{N} \; i = 1, 2, \dots, k \; 0 \le g_i \le 1$

Ekkor fennáll: $\sum_{i=1}^{k} g_i = 1$

Abszolút gyakorisági összeg (kumulált gyakorisági sor):

$$F_i = f_1 + \dots + f_i = \sum_{j=1}^{i} f_j \quad i = 1, 2, \dots, k$$

F_i azon elemek száma, amelyek legfeljebb x_i ismérvvel rendelkeznek.

Relatív gyakorisági összeg:

$$G_i = g_1 + \dots + g_i = \sum_{j=1}^{i} g_j \quad i = 1, 2, \dots, k$$

$$G_i = \frac{F_i}{N}$$

Kapcsolat a gyakorisági és a kumulált gyakorisági sorok között:

$$f_i = F_i - F_{i-1}$$

$$g_i = G_i - G_{i-1}$$

$$F_0 = G_0 = 0$$

4. 6.4 Gyakorisági eloszlás osztályokba sorolt adatokra

Ha nagyon sok ismérv van, vagy folytonos ismérvek vannak, ilyen esetben **osztályokat** célszerű **képezni** az ismérvekre.

Egy adatrendszer feldolgozásánál alapvető probléma, hogy hány osztály képezzünk? Általában célszerű 20 osztálynál kevesebbet választani. Ha sok osztályt választunk, töredezett a hisztogram, ha kevés osztállyal dolgozunk, akkor pedig durva lesz a felbontás.

Az osztályhatárokat úgy kell megválasztani, hogy minden elem belekerüljön valamelyik osztályba (teljes), minden elem csak egy osztályba kerüljön (diszjunkt), és lehetőleg homogén osztályok legyenek.

Az **osztályok számának (k) meghatározására** a szakirodalomban általában kétféle módszert javasolnak. Megállapodás kérdése, hogy melyiket választjuk. Kevés adatszámra mindkét módszer közel azonos osztályszámot szolgáltat.

1. legkisebb *k*, amelyre $2^k > N$
2. Sturges-képlet: $k = [1 + 3.3\lg N]$,

ahol N az osztályozni kívánt adatok száma (minta, vagy sokaság elemszáma)

Egyenközű osztályszélesség esetén minden **osztály hossza**: $\Delta x = \frac{x_{max} - x_{min}}{k}$.

De választhatunk különböző osztályszélességeket is.

Jelölések:

x_i^a : az i-edik osztály alsó határa

x_i^f : az i-edik osztály felső határa

Az **osztályhatárok meghatározása** történhet az alábbi szabályok szerint:

$$x_1^a = x_{min} - \frac{\text{adatpontosság}}{2} \quad x_1^f = x_1^a + \Delta x = x_2^a$$

$$x_k^a = x_1^a + (i-1) \cdot \Delta x \quad i = 2, 3, \ldots, k$$

$$x_k^f = x_k^a + \Delta x = x_1^a + k \cdot \Delta x$$

Az **osztályközép** képzési szabálya: $x_i' = \frac{x_i^f + x_i^a}{2} \quad i = 1, 2, \ldots, k$

5. 6.5 A statisztikai minta fogalma

Definíció: Valamely valószínűségi változóra vonatkozó véges számú független kísérlet vagy megfigyelés eredménye: véges sok azonos eloszlású valószínűségi változó.

Jelölés: Tekintsük a valószínűségi változót, ekkor a -re vonatkozó n elemű minta

$_1$, $_2$,...., $_n$

Az n számú kísérlet elvégzése során a i mintaelem egy-egy konkrét számértéket vesz fel:

$_1 = x_1$, $_2 = x_2$,..., $_n = x_n$

A statisztikai minta reprezentatív: a mintaelemek eloszlása megegyezik a vizsgált valószínűségi változó eloszlásával, hiszen mindegyik kísérletnél magát a valószínűségi változót figyeljük meg.

A statisztikai minta elemei független valószínűségi változók, mivel a kísérleteket egymástól függetlenül végezzük.

A mintaelemekből tapasztalati jellemzőket, u.n. **statisztikát** konstruálunk.

A statisztika a mintaelemek valamely függvénye. A statisztika tehát maga is valószínűségi változó és eloszlásának meghatározása fontos feladat. A valószínűségszámítás tárgyalása során láttuk, hogy a valószínűségi változók eloszlása néhány számadattal (várható érték, szórás) kielégítően jellemezhető. A várható érték az eloszlás súlypontjáról, a szórás a változó értékeinek szétszórtságáról ad felvilágosítást. Ezekre az elméleti jellemzőkre a mintaelemekből igyekszünk következtetni úgy, hogy a $_1$, $_2$, ...,$_n$ mintából különböző függvényeket képezünk. Valamely $_n$ = $_n(_1,_2,....,_n)$ függvény minden konkrét minta esetén egyetlen számadatba tömöríti a mintaelemekben rejlő információt. Milyen függvényt konstruáljunk a mintaelemekből, hogy minél jobb közelítést kapjunk az elméleti várható értéknek, az elméleti szórásnak és egyéb paramétereknek?

6. 6.6 Statisztikák

Az előző fejezet jelöléseit alkalmazva:

Mintaközép

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$$

Tétel:

Ha a valószínűségi változó várható értéke μ, szórása σ, akkor a mintaközépre

$$E(\bar{\xi}) = \mu \qquad D(\bar{\xi}) = \frac{\sigma}{\sqrt{n}}$$

Rendezett minta:

A véletlen, az észlelés sorrendjében kapott mintaelemeket rendezzük nagyság szerint. Jelölje a nagyság szerint a legkisebbet ξ_1^*, a megmaradók közül a legkisebbet ξ_2^*, stb.

Ekkor

$$\xi_1^* \le \xi_2^* \le \ldots \le \xi_n^*$$

A rendezett mintaelemek már nem függetlenek és nem is azonos eloszlásúak.

Mintaterjedelem:

$$R = \xi_n^* - \xi_1^*$$

Medián:

Ha a mintanagyság páratlan, akkor a középső mintaelem a medián - páros mintanagyság esetén a két középső átlaga.

Tapasztalati (empirikus) szórásnégyzet:

A mintaközéptől vett eltérések négyzetének átlaga:

$$s_n^2 = \frac{\left(\xi_1 - \overline{\xi}\right)^2 + \left(\xi_2 - \overline{\xi}\right)^2 + \ldots + \left(\xi_n - \overline{\xi}\right)^2}{n}$$

Korrigált tapasztalati szórásnégyzet:

$$s_n^{*2} = \frac{\sum_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^2}{n-1} = \frac{n}{n-1} s_n^2$$

Variációs tényező (relatív szórás):

$$V = \frac{s_n}{\overline{\xi}}$$

Gyakorisági és sűrűséghisztogram

Gyakorisági hisztogramm szerkesztése

Tegyük fel, hogy az a,b intervallum lefedi a mintaterjedelmet.

Osszuk fel az a,b intervallumot n részre:

$$a = d_0 < d_1 < \ldots < d_n = b$$

A részintervallumok n számára nincs általános szabály, általában 6-12 részintervallumot képezzünk.

Adjuk meg az egyes d_{i-1},d_i részintervallumba eső mintaelemek k_i számát (i=1,2,...,n) és mindegyik részintervallumra rajzoljunk az oda eső mintaelemek gyakoriságával arányos magasságú téglalapot: az i-edik részintervallumra rajzolt téglalap magassága legyen

$$\frac{k}{d_i - d_{i-1}}$$

Ekkor a téglalapok területeinek összege n.

Sűrűséghisztogram szerkesztése

Az egyes intervallumokra rajzolt téglalapok magasságát az oda eső mintaelemek relatív gyakoriságával adjuk meg, azaz az i-edik részintervallum magassága legyen

$$\frac{k}{n \cdot (d_i - d_{i-1})}$$

Az így kapott lépcsős függvény a tapasztalati sűrűségfüggvény, amely közelíti az ismeretlen elméleti sűrűségfüggvényt. Ha ez a hisztogram egy haranggörbét közelít, akkor az eloszlást jó közelítésben normálisnak tekinthetjük.

Tapasztalati eloszlásfüggvény:

Az $x_1,x_2,...,x_n$ mintához tartozó tapasztalati eloszlásfüggvény az x-tengellyel párhuzamos szakaszokból álló lépcsős függvény, amelynek minden egyes felvett x_i értékénél 1/n ugrás van, ha x_i-t egyszer kaptuk a mintában; k/n ugrás van, ha k-szor fordul elő x_i a mintában. A minta eloszlásfüggvénye a minta elemszámának növelésével minden x-re egyenletesen konvergál az elméleti eloszlásfüggvényhez.

7. 6.7 A statisztikai becslések jellemzői

Torzítatlan becslés:

Ha a valószínűségi változó elméleti jellemzője az a paraméter, és az $\alpha_n(\xi_1, \xi_2, \dots \xi_n)$ statisztikai mintából kívánjuk becsülni, akkor elvárjuk, hogy az α_n statisztika értékei az 'a' szám körül ingadozzanak.

$$E(\alpha_n(\xi_1, \xi_2, \dots \xi_n)) = a$$

Konzisztens becslés:

A minta elemszámának növelésével az α_n statisztika egyre jobban közelítse meg az 'a' paramétert.

Elégséges becslés:

Az α_n statisztika tartalmazza az 'a' paraméterre vonatkozó összes információt.

Efficiens becslés:

A legkisebb szórású torzítatlan becslés.

8. 6.8 Becslési módszerek

Alapvető probléma, hogy egy adott valószínűségi eloszláshoz, hogyan található jó becslés. Létezik-e olyan általános matematikai elv, amely megadja, hogy milyen statisztikát számítsunk ahhoz, hogy a keresett paraméterek jó becslését kapjuk?

8.1. 6.8.1 Maximum likelihood módszer, azaz a legnagyobb valószínűség elve

Legyen $\xi^T = (\xi_1, \xi_2, \dots, \xi_n)$ valószínűségi vektor változó $y^T = (y_1, y_2, \dots, y_n)$ mintaelemeinek együttes sűrűségfüggvénye:

$$f(\boldsymbol{y}; \boldsymbol{x}) = \prod_{i=1}^{n} f(y_i; x_1, x_2, \dots, x_m)$$

- ahol $\boldsymbol{x}^T = (x_1, x_2, \dots, x_m) \in R^m$, és R^m a becslési tartomány.

A fenti összefüggés alapján a likelihood függvényre az alábbi kifejezést kapjuk:

$$L(y;\boldsymbol{x}) = \ln f(y;x) = \sum_{i=1}^{n} \ln f(y_i;x)$$

A likelihood becslés a parciális deriváltak nullával való egyenlőségének szükségességéből adódik:

$$\frac{\partial L(y;x)}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial \ln f(y_i;x_1,x_2,\dots,x_m)}{\partial x_j} = 0, \quad j = 1,2,\dots,m$$

Elméleti és gyakorlati szempontból két fontos esetet tárgyalunk:

8.2. 6.8.2 Normális eloszlás esetén

Az együttes sűrűségfüggvény:

$$f(y;\vartheta,\sigma) = \prod_{i=1}^{n} \varphi(y_i;\vartheta,\sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\vartheta)^2},$$

- ahol σ a szórás, ϑ a várható érték.

A likelihood függvény:

$$L(y;\vartheta,\sigma) = \ln c - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\vartheta)^2.$$

A szélsőérték létezésének szükséges feltétele alapján:

$$\frac{\partial L}{\partial \vartheta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\vartheta) = 0,$$

$$\frac{\partial L}{\partial \sigma} = -n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\vartheta)^2 = 0, \quad \sigma \neq 0$$

A fentiekből az ismeretlen paraméterek becslésére az alábbi összefüggések adódnak:

$$\hat{\vartheta} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)^2}{n}$$

Tehát a hagyományos becslési eljárás normális eloszlás esetén a várható értéket a számtani középpel, a szórásnégyzetet a tapasztalati (empirikus) szórásnégyzettel becsüli.

8.3. 6.8.3 Laplace eloszlás esetén

A kétoldalú exponenciális eloszlás, azaz a Laplace eloszlás sűrűségfüggvénye:

$$f(y;\vartheta,\sigma) = \frac{1}{2\sigma} e^{-\frac{|y-\vartheta|}{\sigma}}$$

Az együttes sűrűségfüggvény:

$$f(y;\vartheta,\sigma)=\frac{1}{(2\sigma)^n}\prod_{i=1}^{n}e^{-\frac{1}{\sigma}\sum_{i=1}^{n}|y_i-\vartheta|}$$

A likelihood függvény a következő:

$$L(y;\vartheta,\sigma)=\ln c-n\ln\sigma-\frac{1}{\sigma}\sum_{i=1}^{n}|y_i-\vartheta|$$

A szélsőérték létezésének szükséges feltétele alapján:

$$\frac{\partial L}{\partial\vartheta}=n^{+}-n^{-}=0$$

$$\frac{\partial L}{\partial\sigma}=-n+\frac{1}{\sigma}\sum_{i=1}^{n}|y_i-\vartheta|=0$$, $\sigma\neq 0$,

- ahol n^{+} és n^{-} a pozitív és negatív deriváltak száma.

A fenti egyenletekből az ismeretlen paraméterek becslésére az alábbi összefüggéseket kapjuk:

$$\hat{\vartheta}=\underset{i\in n}{\operatorname{med}}\,y_i$$, $$\hat{\sigma}=\frac{\sum_{i=1}^{n}|y_i-\vartheta|}{n}$$

Tehát Laplace eloszlás esetén a várható érték becslésére a medián, a szórás becslésére a legkisebb abszolút eltérés (LAD) adódik.

9. 6.9 Összefoglalás

1. Az Express újságban 1995. 10. 04.-én eladásra kínált 70 m2 körüli lakások ára (mFt):

2.0, 4.0, 3.1, 3.4, 4.2, 6.0, 3.6, 3.1, 2.6, 3.3, 3.4, 3.5, 2.4, 3.2, 3.8, 3.1, 5.3, 2.5, 3.6, 3.0, 3.5, 3.5, 4.1.

Végezze el az osztályba sorolást!

1. 48 db eladásra kínált lakás megoszlása a kínálati ár szerint

Ár (mFt)	Lakások száma (db)
2.0-2.9	6
3.0-3.9	19
4.0-4.9	11
5.0-5.9	6
6.0-6.9	3
7.0-7.9	3
Összesen	48

Számítsa ki és értelmezze a helyzetmutatókat (átlag, módusz, medián)!

1. Egy vállalkozásnál az azonos termékeket előállító dolgozók napi teljesítménye db-ban:

90, 98, 92, 94, 101, 103, 99, 96, 94, 100, 98, 92, 96, 91, 104, 100, 99, 97, 93, 102, 101, 96, 93, 88, 97

Végezze el az osztályba sorolást!

1. Egy közkedvelt gyorsétterem-hálózat egyik egységében megfigyelték a kiszolgálási időt (mp):

45	48	49	56	61	66	66	66	70	72	72	75	78	79	81	81	83	95	102	135

Határozza meg ugyanezen értékeket osztályozással!

Irodalomjegyzék

Hunyadi-Vita: *Statisztika közgazdászoknak,* KSH, Budapest, 2002

Keresztély, Sugár, Szarvas: *Statisztika példatár közgazdászoknak, BKE,* Nemzeti Tankönyvkiadó, 2005

Korpás A.: *Általános statisztika I-II.,* Nemzeti Tankönyvkiadó, Budapest, 1996

Csanády V., Horváth R., Szalay L.: *Matematikai statisztika,* EFE Matematikai Intézet, Sopron, 1995

Závoti, Polgárné, Bischof : *Statisztikai képletgyűjtemény és táblázatok,* NYME Kiadó, Sopron, 2009

Csernyák L.: *Valószínűségszámítás,* Nemzeti Tankönyvkiadó, Budapest, 1990

Obádovics J. Gy. : *Valószínűségszámítás és matematikai statisztika,* Scolar Kiadó, Budapest, 2003

Reimann J. - Tóth J. : *Valószínűségszámítás és matematikai statisztika,* Tankönyvkiadó, Budapest, 1991

Solt Gy.: *Valószínűségszámítás,* Műszaki Könyvkiadó, Budapest, 1971

Denkinger G.: *Valószínűségszámítás,* Nemzeti Tankönyvkiadó, Budapest, 1978