

BEVEZETÉS A SZÁMÍTÓGÉPES STATISZTIKÁBA

Csendes Tibor

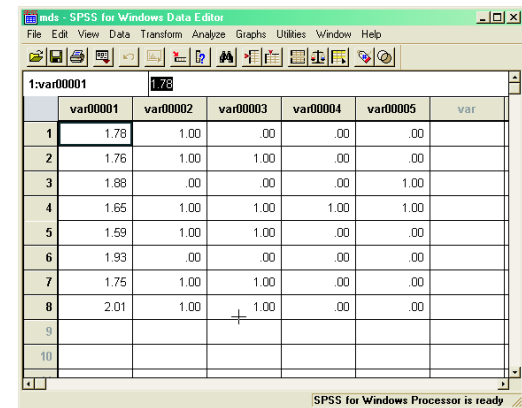
Szeged, 2001.

Lektorálta:

Dr. Boda Krisztina tudományos főmunkatárs, Szegedi Tudományegyetem,
Orvosi Informatikai Intézet,

Dr. Petres Tibor egyetemi docens, Szegedi Tudományegyetem, Gazdaság-
tudományi Kar és

Dr. Viharos László egyetemi docens, Szegedi Tudományegyetem, Bolyai
Intézet.



The screenshot shows the SPSS for Windows Data Editor window. The title bar reads "mds - SPSS for Windows Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data table with the following content:

	var00001	var00002	var00003	var00004	var00005	var
1	1.78	1.00	.00	.00	.00	
2	1.76	1.00	1.00	.00	.00	
3	1.88	.00	.00	.00	1.00	
4	1.65	1.00	1.00	1.00	1.00	
5	1.59	1.00	1.00	.00	.00	
6	1.93	.00	.00	.00	.00	
7	1.75	1.00	1.00	.00	.00	
8	2.01	1.00	1.00	.00	.00	
9						
10						

The status bar at the bottom right indicates "SPSS for Windows Processor is ready".

1. ábra: Az SPSS statisztikai program beolvasott adattal.

Előszó

A jelen jegyzet¹ a Szegedi Tudományegyetemen több tárgy: a TTK-s hallgatók Bevezetés az Informatikába II (Statisztikai programcsomagok) speciálkollégiuma, a közgazdász hallgatók Bevezetés az Informatikába II. féléves és a biológus hallgatók A Számítástechnika alkalmazásai című tárgyának közös anyagát tartalmazza. A tárgy ideális esetben heti egy óra előadást és egy óra gyakorlatot jelent, de mindenképpen legalább a fele időt számítógépes teremben, aktív gyakorlással kell tölteni.

A jegyzet nem támaszkodik más tárgyakra az érintett hallgatók különböző előképzettsége miatt, de nagyon hasznos a matematikai statisztika vagy a statisztika tárgyak anyagának előzetes ismerete. Aktuális változata egy része, a hozzá kapcsolódó feladatok, gyakorlatok és adataik elérhetők a

<http://www.inf.u-szeged.hu/~csendes/stat.ps.gz>

címen, illetve anonim ftp-vel az

[ftp.jate.u-szeged.hu](ftp:jate.u-szeged.hu)

címen a /pub/math/optimization könyvtárban.

A tárgy olyan tudást kíván adni, amely elegendő egyszerűbb statisztikai munkák elvégzéséhez, és amelyet önálló gyakorlással továbbfejlesztve egy-egy szakterület teljes statisztikai feldolgozását is végre lehet hajtani. Mivel a programcsomagok gyakran változnak, az anyag főleg az állandó vagy kevésbé változó ismereteket tartalmazza. Részben emiatt nem is csak egy statisztikai programot ismertet, hanem az SPSS (1. ábra) részletesebb tárgyalása után más könnyen elérhető statisztikai programokat is bemutat röviden. A kapcsolódó statisztikai ismeretek persze mindegyikre érvényesek.

¹Minden megjegyzést szívesen látok és előre is köszönök, különösen, ha hibákra hívják fel a figyelmem. Az email címem: csendes@inf.u-szeged.hu

A rendelkezésre álló rövid idő (kb. 14×1.5 óra) nem elég a valószínűségszámítás és a matematikai statisztika alapfogalmainak részletes tárgyalására sem, ezért a legfontosabb definíciókat, összefüggéseket az érintett statisztikai eljárások tárgyalása előtt csak a feltétlenül szükséges terjedelemben ismertetem. A teljesen önálló statisztikai munkához ez persze nem elegendő. Ennek ellenére bízom benne, hogy a tárgyalt anyag segít a leggyakoribb hibákat elkerülni, és a viszonylag könnyen kezelhető programok segítségével (támaszkodva a mind több esetben rendelkezésre álló kiterjedt súgó, tanácsadó varázslókra) önálló munkával is lehetséges a további szükséges eljárások megismerése. A teljes itt közreadott anyag több, mint amit egy féléves kurzusban át lehet adni, ez némi rugalmasságot követel az előadótól, illetve a gyakorlatvezetőtől.

További cél segítséget nyújtani a statisztikai feldolgozáshoz olyanok számára is, akik ezt a hagyományos képzés keretében nem tanulták. Így a jegyzet alapján az egyszerűbb feladatok esetén az olvasó elegendő útmutatást kap ahhoz, hogy a feladatát úgy fogalmazza meg, illetve írja át, hogy az a rendelkezésre álló szoftverrel hatékonyan megoldható legyen.

A jelen jegyzet a korábbi speciálkollégiumok és gyakorlatok során csiszolódott anyagot is tartalmazza. Itt mondok köszönetet korábbi hallgatóimnak és munkatársaimnak a jegyzet létrejöttéhez, illetve a javításához nyújtott segítségükért. Külön köszönet illeti a lektorokat, akik alapos és gyors munkát végeztek, és számos hasznos tanáccsal segítettek munkámat. Várom a további véleményeket és javaslatokat is.

Szeged, 2001. október

a szerző

Jelölések

Itt a legfontosabb, szinte mindig a megadott formában használatos jelöléseket adjuk meg, de ezektől helyenként — ahol a tárgyalás ezt megköveteli — eltérhetünk.

A, B, C	események
α	szignifikancia-szint
D_1, \dots, D_9	decilisek
$E(x)$	az x valószínűségi változó várható értéke
$F(x) = P\{\xi < x\}$	eloszlásfüggvény
$f(x)$	sűrűségfüggvény
H_0	a nullhipotézis
H_1	az alternatív hipotézis
Me	medián
μ	(elméleti) átlag
$N(0, 1)$	standard normális eloszlás
$N(\mu, \sigma)$	normális eloszlás
p, q	valószínűségek
$P(A)$	egy A esemény valószínűsége
P_1, \dots, P_{99}	percentilisek

Q_1, Q_2, Q_3	a kvartilisek
σ	(elméleti) szórás
s	korrigált szórás
SD	szórás
SE	standard hiba
x, y, z	valószínűségi változók. A matematikai statisztikában inkább görög betűk használatosak (mint pl. ξ, ζ).
\bar{x}	az x_1, x_2, \dots, x_n mintaelemek számtani átlaga

1. Fejezet

Bevezetés

A statisztikával kapcsolatban két gyakran idézett mondás a: “*van hazugság, nagy hazugság és statisztika*”, illetve az “*elegendő számú adatból statisztikával bármit ki lehet mutatni*”. Ezek mögött az az igazság rejlik, hogy a statisztikai eljárások nem elég gondos, nem elegendően körültekintő használata esetén megkérdőjelezhetetlennek tűnő hibás eredményeket kaphatunk. A statisztikai programcsomagok ismertetése során a leggyakoribb hibalehetőségeket is megtárgyaljuk az elkerülésükhöz szükséges lépésekkel.

A jegyzet címében a számítógépes jelző arra utal, hogy közvetlenül nem a statisztika fogalmaival, összefüggéseivel foglalkozunk, hanem statisztikai eljárások, próbák, mutatók konkrét adatokra való meghatározásával. Márpedig ezeket ma a legegyszerűbbek kivételével csak számítógépen hajtják végre. Néhány statisztikai eljárás más jellegű programban is elérhető, így például gyakran táblázatkezelő programban, vagy általános numerikus programcsomagokban is találunk ilyeneket. Jegyzetünkben ilyen statisztikai eljárásokat a StarOffice és az Excel programokkal kapcsolatban ismertetünk röviden (lásd a 4. Fejezetet). Bár ezek a programok kevés statisztikai tesztet, algoritmust bocsátanak rendelkezésre, mégis elterjedtségük miatt és viszonylag könnyű kezelhetőségük révén fontos eszközök. Ide sorolható még a hálózatról letölthető shareware As-Easy-As táblázatkezelő program is (bármely keresőprogrammal megtalálható, csak egy hónapig használható).

Az egyszerűbb statisztikai programok, mint a SigmaStat is, csak egyváltozós statisztikákat képesek kiszámolni, cserében viszont könnyen kezelhetők és kisebb kapacitású gépen is futtathatók, olcsóbbak.

A statisztikai eljárások közel teljes körét rendelkezésre bocsátó professzionális programokból sok van, ezeket főleg PC-n vagy munkaállomásokon használhatjuk. Ide tartozik a részletesen tárgyalt SPSS mellett például a

StatGraphics, a Statistica, a BMPD és az SAS. Ezen osztály által kínált algoritmusok köre nem nagyon tér el, és bár a használatuk nagyon különböző lehet, a céljainkra elegendő ezek közül egyet ismertetni.

Linux operációs rendszerhez is számos programot lehet találni. Egy bő lista van ezekről a <http://chps06.ch.unito.it/linux/A/3> internetes címen (további linkekkel és rövid ismertetéssel minden programról).

Az általános célú numerikus programok közül a NAG programcsomagját, a Maple és a Mathematica szimbolikus algebrarendszereket kell megemlíteni, de ide tartozik a Matlab is. Az ilyen programokat nem tárgyaljuk, mert még rövid ismertetésük is aránytalanul sok időbe kerülne.

1.1 Statisztikai alapfogalmak

A *statisztika* olyan eljárásokkal foglalkozik, amelyek mérési adatok, felmérésekre kapott válaszok vagy más véletlen eseményektől függő adatok jellemzőit vagy összefüggésük mértékét és jellegét határozzák meg. Ide tartozik a kapott eredmények olyan megjelenítése is, amely az adatok értelmezését megkönnyíti. Ezt a diszciplínát szokás *általános statisztikának* is hívni, szemben a matematikai alapjait tisztázó *matematikai statisztikával*. De röviden statisztikának szokás nevezni a *statisztikai függvényeket*, a mintaelemekből számított értékeket is, mint amilyen például az átlag.

A *valószínűség*: egy 0 és 1 közötti szám ($0 \leq p \leq 1$), amely azt jellemzi, hogy egy esemény bekövetkezése milyen eséllyel, gyakorisággal várható. Az 1 valószínűség csaknem biztos bekövetkezést, a nulla valószínűség csaknem lehetetlen előfordulást jelent². A kísérletezés során tapasztalt relatív gyakoriságok megközelítik az elméleti valószínűséget.

Az adatokat általában egy táblázatban célszerű elrendezni. Az *eset* az összetartozó statisztikai adatok olyan egysége, amelyek amiatt képeznek egységet, mert egy egyedre, vagy mérési kísérletre vonatkoznak (pl. a kísérletben résztvevő személy, állat, vegyület stb.). Az eseteket általában egy-egy számítógépes rekordban, rendszerint a táblázat soraiban adjuk meg.

A tulajdonságokat, jellemzőket az egyes egyedekre vonatkozóan a *valószínűségi változók* (röviden változók) tartalmazzák. Az esetekre vonatkozó változóértékek alkotják a *statisztikai mintát*, vagy röviden *mintát*. Sok esetben jellemző az, hogy a teljes sokaságból csak kevés egyedre vonatkozó adat áll rendelkezésre.

²A köznyelvben itt használhatunk biztos, illetve lehetetlen előfordulást is, a “csaknem” a matematikai pontosság kedvéért áll itt.

magyar, ékezetes betűket is használhatunk — amennyiben ezt az operációs rendszer, illetve a rendelkezésre álló betűtípusok megengedik. Az ábra kimentésénél figyeljünk arra, hogy egyes formátumok csak az ábrákra érhetőek el (Chart only), az egész output ablakra (Output document) nem. Szintén említésre méltó, hogy bár az így létrejött encapsulated postscript file (.eps) használható, például a ghostview program meg tudja mutatni, és ki tudja nyomtatni, de a létrejött eps állomány illegális karakterekkel kezdődik és végződik, amire egyes programok (pl. a \LaTeX) érzékenyek. Ebben az esetben ezeket töröljük ki a file elejéről és végéről. A létrehozott grafikont a 4. ábra tartalmazza.

2.20 Feladatok

2.20.1 Egy dolgozat feladatai

1. Olvassa be a zh.txt file-ből az Elso, Masodik, Harmadik, Negyedik és a Szazalek nevű változókat (az elválasztó jel az &). Adja meg a 9, illetve a 99 értékeket mint hiányzóadat kódokat. Írja le a beolvasott utolsó sor tartalmát! 1 pont
2. Határozza meg a min, max, átlag és szórás értékeket a Szazalek változóra azon esetekre vonatkozóan, amelyeket az Elso, Masodik, ..., Negyedik változók 1 értékei jelölnek ki (4-4 érték). Írja ide a kapott értékeket! 1 pont
3. Vizsgálja meg klaszter analízis segítségével, az Elso — Negyedik változók összetartozásának szorosságát, kérjen dendrogrammot is! Foglalja össze itt az eredményt szavakban! 1 pont
4. Vizsgálja meg, hogy a Szazalek változó értékét az Elso — Negyedik változók milyen lineáris függvénye írja le (lineáris regresszió)! Értelmezze az eredményt! 1 pont
5. Generáljon egy új változót Ujszazalek néven, amelynek tartalmát úgy kapjuk, hogy a Szazalek változóhoz a standard normális eloszlással generált véletlen számok ötszörösét adjuk, majd ezt egészsre kerekítjük. Adja meg a kapott új értékeket! 1 pont
6. Ismétlje meg a 4. feladat vizsgálatát az 5. feladatban kapott változóra, és értelmezze az eredményt! 1 pont

Az első feladatban említett zh.txt nevű adatállomány tartalma:

```
0 & 1 & 0 & 1 & 20
1 & 0 & 0 & 0 & 63
0 & 1 & 1 & 0 & 37
1 & 1 & 0 & 0 & 33
0 & 0 & 0 & 0 & 67
1 & 0 & 0 & 0 & 66
0 & 1 & 1 & 0 & 33
0 & 0 & 0 & 1 & 50
1 & 1 & 0 & 1 & 45
0 & 0 & 1 & 0 & 55
```

2.20.2 Egy másik dolgozat feladatai

1. Olvassa be a <http://www.inf.u-szeged.hu/~csendes/statzhfile-t>. Töltse be belőle a BUX, MOL, MATAV, OTP és BP-Növekedési nevű változókat (az elválasztó jel a szóköz) az SPSS programba. Adja meg a megfelelő beállításokat (címké, szélesség, tizedes jegyek száma). Adja meg a 6789 értéket mint hiányzó adatkódot minden változóra. Írja ide a beolvasott utolsó sor tartalmát! 1 pont
2. Határozza meg a min, max, átlag és szórás értékeket a MOL, MATAV és OTP változókra azon esetekre vonatkozóan, amelyekre a BUX változó 5300-nál nagyobb értéket vesz fel (4-4 érték). Írja ide a kapott értékeket! 1 pont
3. Vizsgálja meg klaszter analízis segítségével, a BUX, MOL, MATAV, OTP és BP-Növekedési változók összetartozásának szorosságát, kérjen dendrogrammot is! Foglalja össze az eredményt szavakban! 1 pont
4. Vizsgálja meg, hogy a BP-Növekedési változó értékét a BUX, MOL, MATAV és OTP változók milyen lineáris függvénye írja le (többváltozós lineáris regresszió)! Értelmezze az eredményt! 1 pont
5. Generáljon 5 új változót, amelyek tartalmát úgy kapjuk, hogy a BUX, MOL, MATAV, OTP és BP-Növekedési változók értékét elosztjuk az első napi értékkel és beszorozzuk 100-al (százalékos relatív érték). Ábrázolja az új változókat egy megfelelő grafikonon! Ismétlje meg

a 3. – 4. feladat vizsgálatát a kapott új változókra, és értelmezze az eredményt!
3 pont

Az első feladatban említett statzh nevű adatállomány tartalma:

Nap	BUX	MOL	MATAV	OTP	BP-Novekedesi
XI.1.	5489	4890	1279	10955	1.8789
XI.2.	6789	5120	1343	11034	1.8894
XI.3.	5834	5235	1355	11870	1.9292
XI.4.	5589	6789	1312	11345	1.9111
XI.5.	5309	5102	1298	6789	1.9034
XI.6.	5103	5023	1176	10766	1.8943
XI.7.	5302	5301	1212	11034	1.9102
XI.9.	5332	5395	1235	11203	1.9302
XI.10.	5401	5378	1320	11301	1.9422
XI.11.	5578	5567	1349	11650	1.9388
XI.12.	5632	6789	1377	11703	1.9401
XI.13.	5276	5423	1336	11329	1.9213
XI.14.	5303	5546	6789	11543	1.9324

tanácsadó olyan statisztikai eljárásokból álló sort javasol, amely szakmailag korrekt kiértékelést tesz lehetővé. Ennek követése különösen a kezdő alkalmazó dolgát könnyíti meg. Ha már tudjuk, hogy mely tesztekre van szükség, akkor természetesen ezek közvetlenül is végrehajthatók. Érdemes a jelentések angol nyelvű szövegét is gondosan tanulmányozni, kiértékelni, ezen az úton indokolatlan következtetések levonását lehet elkerülni.

A tanácsadó közvetlenül, a menü nélkül is elérhető az eszközsorban lévő izzólámpát ábrázoló ikonnal. A sűgő ikonja e mellett található, egy kérdőjelet tartalmaz.

3.7 Feladatok SigmaStat-tal való megoldásra

Az alábbi feladatokhoz készítsünk olyan file-t, amely a tesztekhez tartozó jelentések lényegét tartalmazza!

1. Olvassunk be egy adatfile-t az Import utasítással úgy, hogy a benne lévő adatok a táblázat megfelelő oszlopába kerüljenek! Javítsuk ki az adódó eltéréseket!
2. Generáljunk egy táblázatot a Transforms eljárások segítségével úgy, hogy két oszlop normális eloszlású, a köztük lévő különbség pedig szignifikáns legyen!
3. Kérjünk leíró statisztikákat egy adathalmazról, majd az elkészült jelentést szerkesszük át magyar nyelvre!
4. Generáljunk olyan adathalmazt, amelyben a 2. oszlop az elsőnek regresszióval jól leírható transzformáltja — de az összefüggés nem lineáris! Értelmezzük a jelentésben szereplő korrelációs együttható és hibanégyzet értékeit!
5. Generáljunk két normális eloszlást követő oszlopot úgy, hogy a köztük levő korreláció (kb.) -0,5 legyen!
6. Olvassuk be a MINTA.TXT állományt, hozzuk létre belőle a megfelelő táblázatot, és végezzük el rajta a Describe Data, Before and After nevű eljárásokat! (Mind a négy, az állományban lévő táblázat 3., 4., 6. és 8. oszlopoi azonos jellegű mennyiségeket tartalmaznak kezelés nélkül, illetve különböző kezeléseket után.)

7. Generáljunk egy lineáris, polinomiális, illetve nemlineáris összefüggést jelentő adathalmazt kis véletlen, normális eloszlású zajjal terhelve, majd keressük meg az összefüggést regressziós módszerrel!
8. Adjunk meg egy olyan nemlineáris regressziós feladatot, amelynek több lényegesen különböző helyi optimuma van, majd illusztráljuk a SigmaStat lehetőségeit a feladat teljes megoldására!

A 6. feladatbeli MINTA.TXT file tartalma:

```
S5  4  0.17  0.16  94  0.17  100  0.17  100  \\  
S7  4  0.25  0.24  96  0.24  96  0.25  100  \\  
S10 4  0.35  0.35  100 0.35  100  0.36  103  \\  
H3  3  0.48  0.22  46  0.21  44  1.07  223  \\  
H6  6  5.84  3.66  63  3.09  53  16.30  279  \\  
GP  2  558.57 449.94 81 474.15 85 81 089.76 14 517  \\  
SHCB 2  0.07  0.07  100 0.06  86  0.09  129  \\  
RCOS 2  0.04  0.04  100 0.03  75  0.04  100  \\  
RB  2  0.01  0.01  100 0.01  100  0.01  100  \\  

```

3.7.1 Egy dolgozat feladatai

1. Olvassa be (az import utasítással) a statzh file 5. számokat tartalmazó oszlopát (cím: <http://www.inf.u-szeged.hu/~csendes/statzh>)! Ügyeljen arra, hogy a számokon belül ne maradjon szóköz! Törölje az üres sorokat! 1 pont
2. Az adatokban talált 6789-as számokat nyilvánítsa hiányzóadat kódnak! 1 pont
3. Generáljon egy új (a 6.) oszlopba annyi standard normális eloszlású véletlen számot, ahány sor van a táblázatban! 1 pont
4. Képezze az 5. oszlop és a 6. oszlop 0,1-szeresének összegét, és tegye az eredményt a 7. oszlopba! Mentse el az eredmény táblázatot ZHA néven! 1 pont
5. Vizsgálja meg a 4. és a 7. oszlop lineáris függését az 1. oszlop adatától lineáris regresszióval! Az eredményekhez fűzzön magyar nyelvű rövid értelmezést, és mentse el őket ZHA1.RPT és ZHA2.RPT néven! 2 pont

Az első feladatban említett statzh nevű adatállomány tartalma:

Nap	BUX	MOL	MATAV	OTP	BP-Novekedesi
XI.1.	5489	4890	1279	10955	1.8789
XI.2.	6789	5120	1343	11034	1.8894
XI.3.	5834	5235	1355	11870	1.9292
XI.4.	5589	6789	1312	11345	1.9111
XI.5.	5309	5102	1298	6789	1.9034
XI.6.	5103	5023	1176	10766	1.8943
XI.7.	5302	5301	1212	11034	1.9102
XI.9.	5332	5395	1235	11203	1.9302
XI.10.	5401	5378	1320	11301	1.9422
XI.11.	5578	5567	1349	11650	1.9388
XI.12.	5632	6789	1377	11703	1.9401
XI.13.	5276	5423	1336	11329	1.9213
XI.14.	5303	5546	6789	11543	1.9324

Irodalomjegyzék

- [1] A Klinikai Biostatistikai Társaság Jubileumi Évkönyve, 2001.
<http://kbt.zona.hu>
- [2] Boda Krisztina, Eller József, Nyári Tibor: Biostatistika oktatási segédlet. Szeged, 1998-99.
<http://www.szote.u-szeged.hu/dmi/educat/stat/index.htm>
- [3] Hunyadi László és Vita László: Statisztika I., Aula Kiadó, Budapest, 1991.
- [4] Hunyadi László, Mundruczó György és Vita László: Statisztika, Aula Kiadó, Budapest, 2000.
- [5] Katona Tamás és Lengyel Imre (szerk.): Statisztikai Ismerettár, szerzők: Csendes Tibor, Katona Tamás, Kovacsicsné Nagy Katalin, Lengyel Imre, Petres Tibor, Pukli Péter és Vavró István, JATEPress, Szeged, 1999.
- [6] Ketskeméty László és Izsó Lajos: Az SPSS for Windows programrendszer alapjai, SPSS Partner Bt, Budapest, 1996.
- [7] Obádovics J. Gyula: Valószínűségszámítás és Matematikai Statisztika, SCOLAR Kiadó, Budapest, 1998.
- [8] Prékopa András: Valószínűségelmélet műszaki alkalmazásokkal, Műszaki Könyvkiadó, Budapest, 1962.
- [9] SPSS Base 7.5 User Guide. SPSS Inc. 1997.
- [10] Tandori Károly: Matematikai statisztika, JATE Szeged, 1974.
- [11] Tandori Károly: Valószínűségszámítás, JATE Szeged, 1974.

Az irodalomjegyzék persze nem teljes, csak néhány fontosabb, illetve a tananyaghoz közvetlenül kapcsolódó könyvet, jegyzetet adunk meg. Magyar nyelven is elég kiterjedt szakirodalom érhető el, a legtöbb könyvtárban kulcsszó alapján ezek megtalálása minden nehézség nélkül megoldható.

Az itt megadott a könyvek, jegyzetek kiegészítésül ajánlhatók jegyzetünkhöz, de a félév anyagához nem feltétlenül szükségesek. A matematikai alapok alapos megismeréséhez Tandori Károly két jegyzetét [10, 11] javaslom. Az SPSS program korábbi változatait kicsit más megközelítésben tárgyalja Ketskeméty László és Izsó Lajos oktatási segédkönyve [6]. Tudomásom szerint jelenleg ez az egyetlen magyar nyelvű ismertetés az SPSS-ről. Ennek a programnak a teljes dokumentációja számos kötetből áll, az átlagos felhasználó legtöbb problémájára megoldást tud találni a felhasználói leírásban [9]. Ez a jelenlegi szerződések alapján minden hazai egyetemen legalább egy példányban elérhető, a Szegedi Tudományegyetemen például a Számítóközpontban. Az általános statisztikát és alkalmazásait tárgyalják például az [3, 4, 7, 8] könyvek, illetve jegyzetek. Az irodalomjegyzékben speciális szerepet tölt be a [5] ismerettár: ez szótárszerűen összefoglalja a legfontosabb statisztikai fogalmakat képletekkel, rövid magyarázattal és az Excel valamint SPSS használatát segítő útmutatóval.

Tárgymutató

adatbevitel SPSS-be 23
 szöveges állományból 29
 adatbázis állományból 32
adatok mentése SPSS-ben 35
adattípusok 10, 25
alternatív hipotézis 18
arányskála 11
As-Easy-As 137

általános statisztika 8
átlag 13
 számítási átlag 13

bináris változó 10
binomiális eloszlás 15

cellaformátumok 27
centrum 92
cenzorált minta 12
célértékkeresés 132
célfüggvény 87
címké 24
comma adattípus 25
custom currency adattípus 26

csúcosság 14

date adattípus 25
dichotom változó 10
diszkrét valószínűségi változó 9

egyenletes eloszlás 15
egyvoldali próba 18

elemszám 12
eloszlásfüggvény 14
elsőfajú hiba 18
eltérésfüggvény 91
eset 8
érvényes esetek száma 12
Excel 135

faktoranalízis 99
ferdeségi együttható 14
folytonos valószínűségi változó 9
függetlenség 12

globális optimalizálás 87
görbeillesztés 85

hasonlósági mérték 80
hiányzóadat kód 12
 SPSS-ben 26
hierarchikus klaszterezés 95
hisztogram 12

intervallum skála 11

kapcsolt rang 53
kétoldali próba 18
khi-négyzet eloszlás 16
klaszter 91
k-means klaszterezés 92
konfidencia intervallum 17
korreláció 19
középtérték 13
különbségi skála 11

kvalitatív adattípus 10
kvantilis 14
kvantitatív adattípus 10

lapultság 14

matematikai statisztika 8
másodfajú hiba 18
medián 13
megbízhatósági intervallum 17
megbízhatósági szint 17
megbízhatósági tartomány 17
mérési skálák 10
 SPSS-ben 27
minta 8
a minta eloszlása 12
módusz 13

nem szignifikáns 17
nemlineáris regresszió 87
nemparaméteres eljárás 16
névleges skála 10
nominális skála 10
normális eloszlás 12, 16
nullhipotézis 17
numeric adattípus 25

ordinális skála 11

paraméter 16
 paraméteres módszer 16
 paraméteres próba 16
paraméterbecslés 84
percentilis 14
Poisson-eloszlás 15
populáció 12

rangskála 11
regresszió 20
relatív szórás 14
reprezentativitás 12

scientific notation adattípus 25
SigmaStat 113
sokaság 12
Solver 136
sorrendi skála 11
SPSS 21
 File menüsor 34
 Window menüsor 58
standard hiba 14
standard normális eloszlás 16
StarOffice 127
 célértékkeresés 132
 grafikonok 133
 statisztikai függvények 128
statisztika 8
 matematikai statisztika 8
 statisztikai becslés 17
 statisztikai függvény 8
 statisztikai minta 8, 12
 statisztikai próba 17
string adattípus 26
sűrűségfüggvény 14

szabadságfok 17
szignifikancia 17
 szignifikancia-szint 17
 szignifikáns eltérés 17
szórás 13
szórásnégyzet 14
szöveg beolvasása az SPSS-be 29

távolságfüggvény 91
terjedelem 14
többdimenzionális skálázás 103

valószínűség 8
valószínűségi változó 8
változók típusa 25
várható érték 14

Magyar-angol szószeret

Itt a leggyakoribb szakkifejezéseket gyűjtöttem össze azok angol nyelvű változatával. Ez remélhetőleg segít majd az inkább az angol kifejezéseket ismerőknek, és fordítva, megkönnyíti majd az angol szakszöveg, illetve programutasítások olvasását, felismerését azoknak, akik nem ismerik az angol szakirodalmat.

Az nehezíti a dolgot, hogy valószínűleg marketing okokból a statisztikai programcsomagok írói mintha tudatosan igyekeznének egyéni szóhasználatot kialakítani, néhány alapvető kifejezés kivételével a programok gyakran más-más szót használnak ugyanarra a fogalomra (talán úgy gondolják, hogy így az egyszer megszerzett felhasználójuk nehezebben tér át más programra).

adattáza	database
adathiányos eset	missing case
arány skála	continuous scale, ratio scale
aszimmetria	skewness
ábra	graph
átlag	mean, average
centrum	center
célérték keresés	GoalSeek
célfüggvény	loss function, objective function
csoport	group, class
csökkenő	descending
csúcsosság	kurtosis
diagram	chart, plot, graph

egyenesillesztés	linear regression
együttható	coefficient
eloszlás	distribution
eltérés	dissimilarity, deviation
eltérésfüggvény	dissimilarity measure
eset	case
exponenciális	exponential
érték	value
értékcimke	value label
érvényes esetek	valid cases
fadiagram	dendrogram
faktoranalízis	factor analysis
ferdeség	skewness
független változó	independent variable
függő változó	dependent or response variable
görbeillesztés	curve fitting
gyakoriság	frequency
hasonlóság	similarity
hiba	error
hisztogram	histogram
illeszkedés	fit
illesztési utasítás	fit statement
intervallum	interval
jelentés	report
jósolt értékek	predicted values
khi-négyzet próba	chi-square test
kiugró értékek	extremes, outliers
klaszter	cluster
konfidencia intervallum	confidence interval
korlátozás nélküli feladat	unconstrained problem
korlátozó feltétel	constraint

korrelációs együttható	correlation coefficient
köbös	cubic
kördiagram	pie chart
lapultság	kurtosis
legjobb részalmaz regresszió	best subset regression
leíró statisztikák	description statistics
lépés	step
lépésenkénti módszer előre	stepwise forward method
lépésenkénti módszer visszafelé	stepwise backward method
lineáris regresszió	linear regression
magyarázó változó	predictor, explanatory variable
maximum	maximum (plural: maxima)
medián	median
mérési skála	measurement scale
minimum	minimum (plural: minima)
minta	sample
mintaelemszám	sample size
modell függvény	model function, equations
módusz	mode
nemlineáris	nonlinear
négyzetes	quadratic
négyzetösszeg	sum of squares (SS)
névleges vagy nominális skála	nominal scale or observations in categories
növekvő	ascending
optimalizálási feltételek	constraints
ordinális skála	ordinal scale, rank scale
oszlopdiagram	bar chart
osztályközép	midpoint
osztott kördiagram	pie chart
összeg	sum, total
paraméter	parameter
percentilis	percentile

p-érték	p-value
polinomiális regresszió	polynomial regression
pontdiagram	scatter plot
predikció	prediction
próba	test
rangskála	ordinal scale, rank scale
regresszió	regression
regressziós egyenlet	regression equation
reziduum	residual
sorrendi skála	ordinal scale, rank scale
standard hiba	standard error or standard error of the mean
statisztika	statistics
Student reziduum	studentized residual
Student törölt reziduum	studentized deleted residuals
súly	weight
szabadságfok	degree of freedom
szórás	standard deviation
szöveges adat	text data, string
távolság	distance
távolságfüggvény	distance function
terjedelem	range
többváltozós lineáris	multiple linear
t-statisztika	t-ratio
valószínűségi változó	variable
variancia-analízis	analysis of variance (ANOVA)
variancia növelési faktor	variance inflation factor
variációs koefficiens	coefficient of variation
válasz változó	dependent or response variable
változó	variable
változókra vonatkozó korlát	bound constraint
vonaldiagram	line chart

Tartalomjegyzék

Előszó	3		
Jelölések	5		
1. Bevezetés	7		
1.1 Statisztikai alapfogalmak	8		
1.1.1 Változótipusok	9		
1.1.2 Adattípusok	10		
1.1.3 Mérési skálák	10		
1.1.4 A minta jellemzői	12		
1.1.5 Eloszlások	14		
1.1.6 Az eloszlásokkal kapcsolatos alapfogalmak	16		
1.2 Statisztikai próbák	17		
1.2.1 A statisztikai próbákkal kapcsolatos további alapfogalmak	18		
1.2.2 Statisztikai próba végrehajtása	18		
1.2.3 Változók összefüggése	19		
1.2.4 Egy dolgozat feladatai	20		
2. Az SPSS programcsomag	21		
2.1 Az adatok bevitele	23		
2.1.1 A változók beállításai	24		
2.1.2 Gyakorlat	27		
2.2 Az adatok kimentése	28		
2.3 Adat beolvasása szöveges állományból	29		
2.3.1 Gyakorlat	31		
2.4 Adat beolvasása adatbázis állományból	32		
2.4.1 Gyakorlat	33		
2.5 A File menü további utasításai	34		
2.5.1 Új file	34		
		2.5.2 Korábban létrehozott állomány megnyitása	35
		2.5.3 Adataink mentése	35
		2.6 A szerkesztési és a nézet menü	36
		2.6.1 A szerkesztési menü	36
		2.6.2 Beállítások	36
		2.6.3 A nézet menü	37
		2.6.4 Gyakorlat	38
		2.7 Az adatok beállításai, rendezése és mozgatása	38
		2.7.1 Gyakorlat	46
		2.8 Az adatok módosítása	49
		2.8.1 Új változó kiszámítása	50
		2.8.2 Előfordulások megszámlálása	50
		2.8.3 Újrakódolás	51
		2.8.4 Változók kategorizálása	53
		2.8.5 Az esetek rangsorolása	53
		2.8.6 Automatikus újrakódolás	54
		2.8.7 Idősorok létrehozása	55
		2.8.8 A hiányzóadat kódok pótlása	55
		2.8.9 Gyakorlat	56
		2.9 A statisztikai eljárások	58
		2.10 Jelentések és Leíró statisztikák	59
		2.10.1 Eset összefoglalás	60
		2.10.2 A Gyakoriságok (Frequencies) parancs	61
		2.10.3 Példa	63
		2.10.4 Gyakorlat	65
		2.11 Az output ablakok tartalmának szerkesztése	66
		2.11.1 Az output ablak tartalmának nyomtatása	68
		2.12 Táblázatok	71
		2.12.1 A Statistics... nyomógomb párbeszédéses ablaka	72
		2.12.2 További beállítások	73
		2.12.3 Példa	74
		2.13 Átlagok összehasonlítása	75
		2.13.1 Párosított t -próba	75
		2.13.2 Példa	78
		2.13.3 Gyakorlat	79
		2.14 Korreláció	80
		2.14.1 Páronkénti korreláció	80
		2.14.2 Távolságok	82
		2.15 Regresszió	84

Tartalomjegyzék	149
2.15.1 Paraméterbecslés	84
2.15.2 Nemlineáris regresszió	87
2.16 Osztályozás	91
2.16.1 A K-közép klaszterezés	92
2.16.2 Hierarchikus klaszterezés	95
2.17 Az adat tömörített jellemzése	99
2.17.1 Faktoranalízis	99
2.18 Skálázás	103
2.18.1 Többdimenzionális skálázás	103
2.19 Grafikonok	106
2.19.1 Példa	107
2.20 Feladatok	109
2.20.1 Egy dolgozat feladatai	109
2.20.2 Egy másik dolgozat feladatai	110
3. SigmaStat eljárások	113
3.1 Az adatállományok kezelése	114
3.2 Szerkesztési utasítások	116
3.3 Az adatok átalakítása	116
3.4 Statisztikai eljárások	118
3.4.1 Leíró statisztikák	119
3.4.2 Korreláció	120
3.4.3 Regresszió	120
3.4.4 Nemlineáris regresszió	121
3.5 Grafikonok	123
3.6 Sűgő	123
3.7 Feladatok SigmaStat-tal való megoldásra	124
3.7.1 Egy dolgozat feladatai	125
4. StarOffice, Excel és As-Easy-As	127
4.1 StarOffice	127
4.1.1 Statisztikai függvények	128
4.1.2 Célértékkeresés	131
4.1.3 Grafikonok	133
4.2 Excel	135
4.3 As-Easy-As	137
Irodalomjegyzék	139
Tárgymutató	140

150	Tartalomjegyzék
Magyar-angol szószedet	143
Tartalomjegyzék	147