

**Ozsváth Károly, Ács Pongrác**

# **Bevezetés a sporttudományos kutatásba**

# Tartalomjegyzék

<b>1. Bevezetés</b> .....	<b>7</b>
<b>2. A tudományos kutatás alapfogalmai</b> .....	<b>8</b>
2.1. <i>Tudomány és kutatás</i> .....	8
2.1.1. A kutatás szintjei .....	10
<b>3. A tudományos munka menete</b> .....	<b>11</b>
3.1. <i>A munka előkészítése</i> .....	12
3.2. <i>Adatgyűjtés</i> .....	14
3.3. <i>Adatfeldolgozás</i> .....	15
3.4. <i>Publikálás</i> .....	16
3.4.1. Részletes tartalmi követelmények.....	17
3.4.2. Formai követelmények.....	18
3.4.3. Eredményközlés szinterei .....	19
3.4.3.1. Szemlézés, citációs index .....	20
3.4.3.1.1. EISZ (Elektronikus Információszolgáltatás) .....	21
3.4.3.2. Impakt faktor.....	23
3.4.3.3. Kongresszusok, konferenciák .....	24
3.4.3.4. További prezentációs formák.....	25

<b>4. Irodalomkezelés .....</b>	<b>26</b>
4.1. <i>Publikációs stílusok</i> .....	27
4.1.1. Hivatkozás könyveknél és folyóiratoknál .....	28
4.1.2. Hivatkozási alapszabályok .....	29
4.1.3. További jelölések .....	30
4.2. <i>Az irodalomjegyzék és szövegbeli hivatkozások összekapcsolása</i> .....	31
4.3. <i>Ábrák és táblázatok kezelése</i> .....	32
4.4. <i>ISBN, ISSN, DOI jelentése</i> .....	33
4.5. <i>Online anyagok kezelése</i> .....	34
<b>5. Elméleti alapok .....</b>	<b>35</b>
5.1. <i>A tesztekkel szemben támasztott alapkövetelmények, kritériumok</i> .....	35
5.1.1. Standardizálás .....	36
5.1.1.1. Érvényesség (validitás) .....	37
5.1.1.1.1. Kritériumvaliditás .....	38
5.1.1.1.2. A kritériumvaliditás formái .....	39
5.1.1.1.3. Az érvényesség jellemzése .....	40
5.1.1.2. Megbízhatóság (reliabilitás) .....	42
5.1.1.2.1. A megbízhatóság jellemzése .....	43
5.1.1.2.2. A megbízhatóság ellenőrzésének alapvető módszerei .....	44
5.1.1.3. Tárgyilagosság (objektivitás) .....	45
5.1.1.4. Gazdaságosság (ökonómikusság) és normatíválhatóság .....	46

<b>6.</b>	<b>Az adatfeldolgozás módszerei .....</b>	<b>47</b>
6.1.1.	Az adatrögzítés javasolt formája.....	48
6.2.	<i>Statisztikai programcsomagok</i> .....	49
6.2.1.	SPSS.....	50
6.2.2.	SAS.....	53
6.2.3.	StatSoft STATISTICA .....	54
6.2.4.	BMDP .....	56
6.3.	<i>Statisztikai alapfogalmak</i> .....	57
6.3.1.	Populáció és minta .....	58
6.3.2.	Adatok, skálák .....	59
6.3.2.1.	Az adatok jellege .....	59
6.3.2.2.	Az adatok értékkészlete.....	60
6.3.2.3.	Az adatok skálája .....	61
6.3.2.3.1.	Névleges skála.....	62
6.3.2.3.2.	Rendező skála.....	63
6.3.2.3.2.1.	Az iskolai osztályzatok.....	64
6.3.2.3.2.2.	Likert skála .....	65
6.3.2.3.3.	Intervallumskála .....	66
6.3.2.3.4.	Arányskála .....	67
6.3.2.3.5.	„Mérési szint” .....	68
6.3.2.4.	Az adatok skálatípus szerinti feldolgozhatósága.....	69
6.3.2.4.1.	„Dummyzás, vakváltozózás” .....	70
6.3.3.	Változó, paraméter .....	71
6.3.4.	Függő és független változók.....	72

6.3.5.	Hipotézisek, szignifikancia .....	73
6.3.5.1.	A hipotézisek formái .....	74
6.3.5.2.	A kiemelt forma: nullhipotézis .....	75
6.3.5.2.1.	Szignifikancia szintek .....	76
6.3.5.2.2.	Statisztikai hibák .....	77
6.4.	<i>Leíró statisztikák</i> .....	78
6.4.1.	Középértékek.....	79
6.4.2.	Az adatok változékonyságának mutatói.....	80
6.4.2.1.	Szóródás, terjedelem, szélsőértékek.....	80
6.4.2.2.	Kvantilisek .....	81
6.4.2.3.	Átlagos eltérés, variancia, szórás.....	82
6.4.2.3.1.	Szabadságfok .....	83
6.4.2.4.	Átlag hibája és variációs együttható.....	84
6.4.3.	Ábrázolás.....	85
6.4.4.	További leíró statisztikai elemek.....	86
6.4.5.	Összefüggések „leírása” .....	87
6.4.6.	Leíró statisztikai képletek áttekintése.....	88
6.4.6.1.	Kiindulási képletek és jelölések.....	89
6.4.6.2.	Két változóra kiterjesztett képletek.....	90
6.4.7.	Gyakorisági eloszlás, percentilisek.....	91
6.4.7.1.	Hisztogram .....	92
6.4.7.2.	A gyakorisági eloszlás alapvető jellemzői.....	93
6.4.7.3.	A normális eloszlás.....	94
6.4.7.3.1.	Transzformálás: standardizálás, standard normális eloszlás.....	95
1.1.1.1.1.1.	A standard normális eloszlás jellemzői.....	96

6.4.7.3.2.	További transzformációk.....	97
6.5.	<i>Gyakorló és demonstrációs példák StatSoft és SPSS programcsomagokkal.....</i>	<i>100</i>
6.5.1.	A Statistica és az SPSS számítási indító ablakai.....	102
6.5.2.	Adatellenőrzés: frekvencia táblázatok lehívása.....	105
6.5.3.	Leíró statisztikák számítása a statisztikai programokkal.....	108
6.5.3.1.	Általános indító műveleti ablakok (StatSoft).....	108
6.5.3.1.1.	A Select Cases opció használata.....	109
6.5.3.1.2.	A „Statistics by Groups” funkció használata leíró statisztikáknál.....	114
6.5.3.1.3.	A „By Group...” funkció közvetlen használata a műveleti ablakból.....	119
6.5.3.1.4.	Az eredménytáblázatok szerkesztése.....	122
6.5.3.2.	Indító menü és leíró statisztikák az SPSS-ben.....	124
6.5.3.2.1.	Az SPSS „Descriptive Statistics” és esetválasztó menüje.....	128
6.5.3.2.2.	Leíró statisztikák és hisztogram a „Frequencies” menüből (SPSS).....	134

## 1. Bevezetés

A tudományos kutatás és eszköztára az elmúlt fél évszázadban szerves részét képezte a felsőoktatás tananyagának. A kutatás-módszertani tárgyak a számítógépek elterjedésével egyre hangsúlyosabbá váltak a képzésben. A tudományos kutatással kapcsolatos alapismeretekre a hallgatóknak a szakirodalom tanulmányozásához, a különböző beadandó dolgozataik és prezentációik, valamint a szak- illetve diploma dolgozatuk elkészítéséhez feltétlen szükségük van. Sajnálatosan a kutatás-módszertani tantárgyakat a hallgatóság sokszor nem ebből a szempontból kezeli.

A tankönyv alapvetően a sporttudományi BSc képzéshez készült, azonban célunk, hogy az oktatás minden szintjén és színterén – így a TDK munkában is – használható legyen. A tárgyalta anyag ennek megfelelően meghaladja az alapképzés szintjét, és magába foglalja a legfontosabb többváltozós módszereket is. Hangsúlyozzuk azonban a tárgyalta módszerek eszköz jellegét, és kiemelten kezeljük a felsőoktatásban talán világszerte leggyakrabban használt két statisztikai program használatát. Kitekintést adunk ugyanakkor a legelterjedtebb táblázatkezelő program, az MS Excel statisztikai lehetőségeire is. A könnyebb érthetőség miatt a legtöbb esetben egy konkrét sporttudományi vizsgálat anyagát használjuk példánkainál. Reméljük, hogy hallgatónk felkészülését hatékonyan segíthetjük a kiadvánnyal.

Érd – Pécs, 2011.

## 2. A tudományos kutatás alapfogalmai

### 2.1. *Tudomány és kutatás*

Az alapfogalmakat a különböző kézikönyvek és lexikonok részletekbe menően tárgyalják. Jelen fejezetben a lehető legegyszerűbben, a lényegre fókuszálva kerül bemutatásra a kutatás-módszertani terminológia.

A magyar nyelv „**tudomány**” szava három jelentéstartalmat hordoz:

jelenti egyrészt a világ megismerésének egyik legfontosabb útját, aminek alapvető eszköze a kutatás folyamata és az ezzel kapcsolatos tevékenység;

jelenti másrészt a fenti tevékenységet végző embereket, a nemzetközi tudományos közösséget;

jelenti harmadrészt (és dominánsan) a tudományos közösség tevékenységének produktumát, a tudományos ismeretek szigorú elvek szerint ellenőrzött, megvitatott, meghatározott szabályok szerint közzétett (publikált), és a tudományos közösség által rendszerezett együttesét.



A különböző gondolkodók és tudományos iskolák azonban e három jelentéstartalmat is eltérő módon értelmezik. A tudomány fogalmának legegyszerűbb meghatározása: **az igazolt ismeretek rendszere**. Specifikum az „igazolás” módja (amelyben napjainkban kiemelkedő a statisztika szerepe). **A tudomány magába foglalja törvényszerűségeket, összefüggések meghatározását, közzétételét, tárolását és hozzáférhetőségének biztosítását (dokumentáció-információs rendszer), alkalmazását, valamint koordinációs szervezeteit.** A tudomány egyúttal módszeres megismerési tevékenység, valamint e tevékenység során szerzett tudás összessége.

A jelenségek felderítése, leírása, magyarázata empirikus és teoretikus szinten alapvetően jellemző a tudományra. Fő eszköze a **kutatás**, amely új ismeretek szerzésére és igazolására szolgál. A kutatás célirányos felderítés, probléma megoldás, a tudásbázis szisztematikus bővítése szigorúan ellenőrzött és reprodukálható körülmények között. Jellemzői a statisztikailag kiértékelt és megfelelően interpretált eredmények. Módszertana és eszköztára (pl. a statisztika) a logikailag elvárható és a ténylegesen megfigyelt vagy megmért események és adatok összehasonlításán alapulnak. **A kutatáshoz tehát mindenekelőtt adatokra van szükség!**

### 2.1.1. A kutatás szintjei

A kutatásnak 3 **szintjét** különböztetjük meg: alap-, alkalmazott, fejlesztő kutatás.

Az **alapkutatások** olyan új ismeretek feltárására irányulnak, amelyek közvetlen gyakorlati hasznosíthatósággal nem járnak, de bázisát képezik vagy képezhetik további kutatásoknak. Rendkívül eszközigényesek és drágák, ugyanakkor a tudományos, technikai-technológiai és társadalmi fejlődés, a világ jobb megismerésének alapját és lehetőségét hordozzák magukban. Fő céljuk az elméleti ismeretek bővítése.

Az **alkalmazott kutatások** az alapkutatások eredményeit felhasználva a gyakorlati hasznosítást és felhasználást célozzák. A kutatások többsége, sőt egyes tudományterületek is e kategóriába tartoznak. Fő céljuk az elméleti alapok gyakorlati alkalmazásának támogatása.

A **fejlesztő kutatások** már ismert tudományos eredmények felhasználásával a gyakorlati alkalmazás hatékonyságának, eredményességének növelését célozzák, és sok esetben új módszerek kidolgozásával járnak együtt. A gyakorlati bevezetés, illetve a fejlesztés megfelelő innovációt feltételez. Létezik azonban olyan nézet is, amely vitatja a fejlesztő kutatások céljaként az új ismeretek feltárását (a megismerést), és ezért a fejlesztést nem is tekinti „igazán” tudományos tevékenységnek.

### 3. A tudományos munka menete

Tanulmányaik során tudományos jellegű munkával a hallgatók többsége a szakdolgozat készítése vagy TDK munka kapcsán kerül közvetlen kapcsolatba. Kezdetnek témát (címet) és témavezetőt/konzulenszt keres, áttekinti a vonatkozó irodalmat, kialakítja az irodalomjegyzékét. Mindezek azonban csak az indulást, a tényleges tartalmi rész megalapozását jelentik. A folytatás intézménytől, témától és témavezetőtől függően eltérő lehet.

A tudományos igényű tevékenység a gyakorlatban 4 fő, egymásra épülő részre bontható: **előkészítés, adatgyűjtés, adatfeldolgozás, közzététel** (publikálás). A fő részek további elemekre bonthatók, időigényük sokszor közel azonos. Szerencsés esetben a „gyakorlati hasznosítás” nem merül ki a publikációban, hanem az eredmények további kutatásokban felhasználásra kerülnek, vagy akár konkrét gyakorlati alkalmazások részévé válnak.

### **3.1. A munka előkészítése**

Az előkészítés 3 nagyobb részre bontható: **problémafelvetés, irodalmi áttekintés, adatgyűjtés előkészítése.** Utóbbi lényegében a későbbiekben „anyag és módszer” elnevezéssel szereplő metodikai részt takarja.

Az előkészítés ténylegesen többnyire a probléma felvetéssel és a hozzá kapcsolódó célkitűzéssel, valamint a kérdésfeltevéssel kezdődik. A kérdésekből elvileg már következnek a rájuk adott feltételezett válaszok, a hipotézisek. A hipotézis (feltételezés) ennek megfelelően formájában mindig állítás. Kiindulásként „munkahipotézist” szokás megfogalmazni, ami későbbiekben finomításra és pontosításra kerülhet. A kiindulási munkahipotézisek sok esetben további részelemekre bonthatók. A statisztikai analízisek sajátossága a „nullhipotézis” – amit a későbbiekben tárgyalunk –, ennek alternatíváját célszerű még az adatgyűjtés előtt megfogalmazni.

Az előkészítés másik központi eleme a szakirodalom áttekintése és feldolgozása. Ennek során el kell készíteni az irodalomjegyzéket, aminek a téma alapvető irodalmát magába kell foglalnia, és a munka befejezéséig az időközben fellelt vagy újonnan megjelent anyagokkal folyamatosan bővíthet. Az irodalom kezelését fontossága miatt külön fejezetben tárgyaljuk.

Az előkészítés harmadik fő eleme az adatgyűjtés megtervezése, előkészítése és leszervezése. Meg kell határozni (identifikálni és definiálni) a rendelkezésünkre álló lehetőségek függvényében az adatgyűjtési/mérési eljárásokat, a vizsgálandó

tulajdonságokat, változókat. Ezt követi a mintaválasztás, a vizsgálati személyek/esetek behatárolása. Eldöntendő, hogy keresztmetszeti („cross-sectional study”) vagy hosszmetzeti (longitudinális) vizsgálatot hajtunk végre. Végül – fentiekből elvileg következik – már ekkor át kell gondolni az adatfeldolgozás módszereit. Ezt követően kerülhet sor az érdemi szervezésre, az adatgyűjtési/mérési eszközök beszerzésére/előkészítésére, az esetleges mérőszemélyzet kiválasztására és felkészítésére, az adatfelvétel helyének és időpontjának kitűzésére, egyeztetésére.

### **3.2. Adatgyűjtés**

Az adatgyűjtés és mérés, a vizsgálatok konkrét lebonyolítása többnyire időigényes és pontos végrehajtást feltételező, központi és meghatározó részét képezi a tudományos munkának. Legfontosabb eleme, hogy adataink pontossága és megbízhatósága egyforma legyen, az adatok keletkezési körülményei azonosak legyenek. Mérések esetén az eljárások forgatókönyvét, a mérési protokollt minden részletében szükséges betartani. Az adatlapokon szereplő értékeket, eredményeket célszerű mielőbb rögzíteni a későbbi adatfeldolgozáshoz.

### **3.3. Adatfeldolgozás**

A tudományos tevékenység harmadik nagy része az adatok feldolgozása **adatrögzítés, adatellenőrzés, és a tényleges számítások** (leíró statisztikák, valamint a célkitűzésnek megfelelő adatelemzési eljárások) végrehajtására tagolható. Fentieket a továbbiakban részletesen tárgyaljuk. Most csak annyit jelzünk előzetesen, hogy az **adatokat Excel táblázatban javasoljuk rögzíteni: az oszlopokban szerepeljenek a változók, a sorokban az esetek/személyek.**

Az adatrögzítést sokan „rabszolgamunkának” tekintik, és a monoton adatbevitel valóban tárháza a potenciális hibáknak. A mérési és adatrögzítési hibák kizárása, lehetséges korrekciója érdekében a tényleges számítások elvégzése előtt feltétlenül szükséges részletes adatellenőrzést végrehajtani.

A tényleges adatfeldolgozás, a számítások eredményei azonnal adják az értelmezés elvi lehetőségeit is. Ezek bővebb kifejtésére a publikációkban külön fejezetekben (diskusszió és következtetések) kerül sor.

### **3.4. Publikálás**

A publikációk szerkezete lényegében követi a tudományos tevékenység menetét. A tudományos igényű eredményközlés tartalmi és formai követelményeit részletesen előírják legtöbb esetben. A minimális tartalmi követelmények magyarul és angolul:

Cím/Title

Szerző/Author

Bevezetés/ Introduction

Cél /Purpose

Metodika/ Methods

Eredmények/Results

Megbeszélés/ Discussion

Következtetések/Conclusion

Összefoglalás/ Abstract

Irodalomjegyzék/References

Nagyobb terjedelmű anyagoknál a bevezetés előtt tartalomjegyzék feltüntetése elvárás, a legvégén pedig melléklet, függelék, ábra és táblázatjegyzék, esetleg tárgymutató szerepelhet. Tanulmányoknál és konferencia előadásoknál/poszttereknél az is előírás lehet, hogy a tartalmi összefoglaló (abstract, resume) az anyag elején, a bevezetés előtt szerepeljen.



### **3.4.1. Részletes tartalmi követelmények**

A tartalmi követelményeket tovább lehet részletezni, és az egyes fejezetek elnevezésében számos szinonima használatos. Szak- és diplomadolgozat, tudományos értekezés esetében szokásos részletesebb tartalmi követelmények:

- Cím/Szerző(k)/Témavezető

- Bevezetés

- Problémafelvetés

- A vizsgálat tárgya és célja

- Irodalmi áttekintés

- Kérdésfeltevés, hipotézis(ek)

- Anyag és módszer (metodika)

  - Vizsgálati anyag/személyek (férfi és női elemszámok, vizsgálat időpontja, helye, körülményei)

  - Vizsgálati módszerek (a változók részletesen, mérési dimenzióra és pontosságra, az eljárás technikai körülményeire kitérve)

  - Az adatfeldolgozás módszerei (az alkalmazott statisztikai eljárások felsorolása, szoftver megnevezése)

- Eredmények

- Diskusszió (megbeszélés, tárgyalás, megvitatás)

- Következtetések

- Összefoglalás

- Bibliográfia (irodalomjegyzék)

- Függelék/Melléletek/Jegyzetek

### **3.4.2. Formai követelmények**

Formai követelmények: intézménytől, kiadótól, szerkesztőségtől, konferencia szervezőitől függő, de általában részletesen szabályozott terjedelem és tipográfia (betű típusa, mérete, sorköz, ábrák-táblázatok, stb. vonatkozásában). A szakdolgozatokhoz, diplomamunkákhoz az egyetemek többnyire részletesen szabályozzák a tartalmi és formai követelményeiket, amit fentiekől és a későbbiektől függetlenül ellenőrizni szükséges!

### 3.4.3. Eredményközlés szinterei

Eredményközlés szinterei: könyv/monográfia/értekezés (lektorálás, opponálás), folyóiratban tanulmány (lektorálás, szemlézés, citációs index, impact factor) konferenciák/kongresszusok: előadás (nyitó, plenáris, szekció), poszter.

Legértékesebbnek a szakkönyveket és egyetemi tankönyveket tartják. Napjainkban ezeket sok esetben szerzői munkaközösségek írják. A könyveket külön bírálják, lektorálják, az észrevételek alapján a szöveget általában korrigálják. A lektor szerepe egyértelműen segítő, támogató szándékú. Monográfiának hívják egy tudományos témakör kimerítő tárgyalását tartalmazó könyvet.

Az értekezések (doktori értekezés) jellemzője a témavezető, és az opponenseknek nevezett bírálók (általában 2 személy). Az opponens szerepe „szembe helyezkedő”, elvileg kifogásokat kell keresnie a munkában. Az opponensi bírálatra a szerzőnek (jelöltnek, aspiránsnak, doktorandusznak) reagálnia kell, „meg kell védenie” értekezését. Amennyiben az opponensek elfogadják a választ, érdemben értékelhető az értekezés. Szak- és diplomadolgozatok esetében is előfordul hasonló eljárás és elnevezés.

Napjaink tudományos eredményeinek döntő többsége hagyományos és online szakmai-tudományos folyóiratokban, tudományos konferenciákon kerül közzétételre. A tanulmányok, szakcikk az „értékesebbek”, de a „jobb” konferenciák is megjelentetnek tanulmányköteteket. A tanulmányokat szintén lektorálni szokás, a szerkesztő bizottságok kizárólag a lektor által támogatott, a szükséges mértékben javított, megfelelő szintű szakcikkkel foglalkoznak érdemben.

#### **3.4.3.1. Szemlézés, citációs index**

A nívós folyóiratokat és a bennük szereplő tanulmányokat több szinten szemlézik, a bennük szereplő és a rájuk történő hivatkozásokat adatbázisokban is nyilvántartják. (Többnyire USA-beli tudományos központok speciális számítógépein.) A szerző(k) idézettségét (hivatkozások száma) külön jellemzik. A Science Citation Index 1964 óta használatos, a természet- és műszaki tudományok területére terjed ki. Napjainkra a társadalomtudományi (Social Sciences Citation Index), valamint a bölcsészettudományi és művészeti területre (Arts & Humanities Citation Index) is kiterjesztették. Sőt ma már szakterületekre kialakított indexek is léteznek (pl. BioSciences Citation Index, Chem Sciences Citation Index és a Clinical Medicine Citation Index). A legnagyobb bibliográfiai adatbázist az amerikai (USA) Thomson Reuters cég kezeli, formális elnevezése ISI (Institute for Scientific Information). A „Web of Knowledge” és „Web of Science” (WoS) néven is futó szolgáltatásokért elvileg fizetni kell, azonban a magyar egyetemi hálózaton belül minden oktató és hallgató részére ingyenesen hozzáférhető.

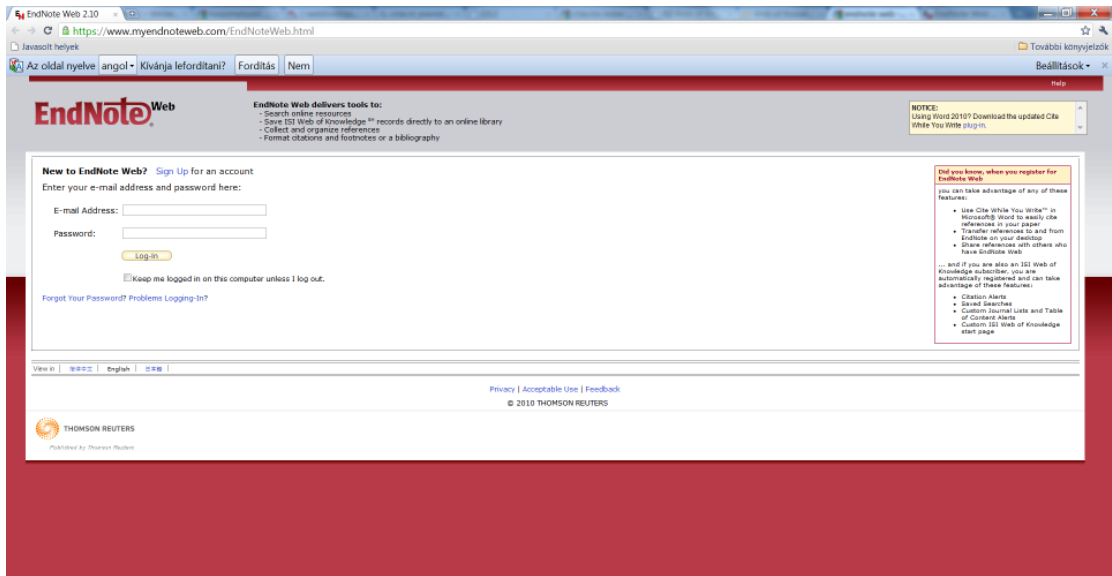
### 3.4.3.1.1. **EISZ (Elektronikus Információsztájtatás)**

Pusztán az EISZ (Elektronikus Információsztájtatás, [www.eisz.hu](http://www.eisz.hu), 1. ábra) szolgáltatásra kell regisztrálni, amihez hallgatóknál a diákigazolvány száma szükséges.



**1. ábra: Az EISZ nyitó ablaka**

A szolgáltatás otthonról nem (illetve nagyon körülményesen) használható, csak az egyetemi számítógépekről, illetve az egyetemi hálózatra csatlakoztatott laptopokról. További információ: <http://www.eisz.hu/main.php?folderID=848> oldalon található. A szolgáltatás bibliográfiakezelő alkalmazások használatát is ingyenesen engedi. Ezek közül az EndNote (<http://www.endnote.com/>) webes felületen már otthonról is elérhető, amennyiben az EISZ-en belül regisztráltunk rá (2. ábra).



2. ábra: Az EndNote bibliográfiakezelő webes felülete

### **3.4.3.2.      Impakt faktor**

Az „impact factor” (IF) a tudományos folyóiratok jellemzője. Pályázatokhoz és tudományos minősítésekhez szükséges szakirodalmi tevékenység irodalomjegyzékében célszerű szerepeltetni a folyóiratok IF értékét is. *Schubert A.* ([http://www.kfki.hu/library/imp/impakt\\_faktor.htm](http://www.kfki.hu/library/imp/impakt_faktor.htm)) alapján: „Az impakt faktor (leggyakoribb magyar fordításban hatástényező) a tudományos folyóiratok átlagos idézettsége alapján létrehozott mutatószám. Megalkotója Eugene Garfield, a philadelphiai (PA, USA) Institute for Scientific Information (ISI) alapító elnöke. A Science Citation Index (SCI) kiegészítő köteteként megjelenő Journal Citation Reports (JCR) kiadványban - a folyóiratokra jellemző más idézettségi adatokkal együtt - 1976-ban jelentek meg az impakt faktorok 1974. évi idézetek alapján kiszámított értékei. Azóta évenként jelennek meg a JCR kötetei a tárgyévi impakt faktorokkal - kezdetben nyomtatott kötetekben, majd mikrofilmen, CD-ROM-on és legújabban Interneten hozzáférhető adatbázis formájában (kizárólag előfizetők számára).”

### **3.4.3.3. Kongresszusok, konferenciák**

A konferenciák és a nagyobb kongresszusok alapvetően a szóbeli prezentáció színterei. Előfordul, hogy egy kiemelt szaktekintély nyitó előadásával kezdenek, ennek időtartama 30-60 perc között szokott lenni. A nagy konferenciákon plenáris és szekcióüléseket szerveznek. A plenáris ülés mindenkihez szól, ideje alatt más programot párhuzamosan nem szerveznek. A plenáris üléseken többnyire meghívott előadók szerepelnek, jellemzően 20-40 perces előadásokkal. A különböző témaköröket felölelő szekcióüléseket részben párhuzamosan szervezik, a legtöbb előadó itt szerepel. Az előadások szokásos időkerete 10 perc, amit 5 perc kérdések és vita követ. Az előadói időkeretet többnyire rendkívül szigorúan betartatják. Egyes esetekben szimpóziumokat is szerveznek, ami szűkebb körű tanácskozás egy meghatározott témáról meghatározott résztvevőkkel. Az előadások a szimpóziumokon is jellemzően 10-15 percesre tervezettek, de a vita és tanácskozás időkerete általában kevésbé kötött. Végül meg kell említeni a poszter („plakát”) szekciókat, ami sajátos átmenet az írásbeliség és a szóbeliség között. A poszter felépítése elvileg azonos az előadásokéval. A posztereket előre meghatározott helyre és időre kell kifüggeszteni. Fél napig vagy 1 napig lehet a helyén hagyni. A szerzőnek jellemzően 3-5 perce van poszterének bemutatására a hivatalos poszter szekció időtartama alatt. A szekcióülés (pontosabban „szekcióállás”) előtt és után azonban az érdeklődők bőségesen konzultálhatnak egymással.



#### **3.4.3.4. További prezentációs formák**

Számos esetben – akár konferenciákhoz kapcsolódóan – alkalmaznak további szóbeli prezentációs formákat. Ilyen pl. a „workshop” (műhely, munkaértekezlet, tanácskozás) vagy a „round table” (kerekasztal).

A „workshop” konkrét, többnyire szűk témára koncentráló, sok esetben egy behatárolt munkacsoport, team tevékenységét bemutató, gyakorlatorientált prezentációs forma, ahol az adott kérdés elemzésére és a megoldási javaslatok vitájára koncentrálnak. Rövidebb (60-90 perc) és hosszabb (3-6, 2x6 óra, 1-4 nap stb.) formái egyaránt előfordulnak, utóbbiak jellemzően tréninggel összekötve. A workshopok jellemzője a rugalmasság és a jelenlévők aktív részvétele az adott tevékenységben, elemzésben.

A „round table”, kerekasztal prezentációk 8-10 szakértő eszmecserejét jelentik egy adott témáról, egy moderátor vezetésével. Jellemzően 15 perces bemutatóval, és ezt követő 30 perces irányított beszélgetéssel és vitával összekötöttek – de ettől eltérő időkeretek is előfordulnak (pl. 90 perces időkeret). A hallgatóság célzott kérdéseket tehet fel, így egy adott témáról mélyreható vita alakulhat ki a hasonló érdeklődésű emberek között.

#### 4. Irodalomkezelés

Az irodalomjegyzék vagy bibliográfia fő elemei: **szerző** – **cím** – **pontos forrás** (azonosíthatósági/visszakereshetőségi/hozzáférhetőségi adatok: kiadó/folyóirat, év, szám, oldal). Alapszabály, hogy ami szerepel a szövegben (hivatkozás/„citáció”, idézet), annak a bibliográfiában is szerepelnie kell! Fordítva is igaz: az irodalomjegyzék nem tartalmazhat olyan tételt, amire a szövegben nincs utalás. Sorrend: szerzők vezetékneve szerinti abc sorrend, azon belül a megjelenés éve szerinti sorrend. Titulusok (dr, PhD, Prof. stb.) nem kellenek! A keresztnévet (neveket) csak első betűjükkal jelzik.

Itt utalnánk a szokásos hivatkozások és a pontos idézetek megkülönböztetésére. Utóbbiak mindig idézőjelben szerepelnek. Az egyszerű hivatkozások viszont csak tartalmi utalások (üggyelve a plágium elkerülésére). Mindkét esetben megadandó(k) a bibliográfiai azonosításhoz szükséges adatok: szerző(k) és évszám, vagy a bibliográfiai sorszám – ha utóbbi eljárást alkalmazzák.

#### **4.1. Publikációs stílusok**

Az irodalomkezelés tartalmi és formai előírásainak feladata a hivatkozások rendjének szabályozása. A cél egy könnyen kezelhető hivatkozási rendszer behatárolása, ami ugyanakkor egyértelműen azonosíthatóvá teszi a hivatkozott irodalmat. A kiadók, egyetemek, folyóiratok fentiek érdekében rengeteg „publikációs stílust”, bibliográfiai formátumot dolgoztak ki az elmúlt évtizedekben, amit esetenként újabb verziókkal, „kiadásokkal” frissítettek. Fő vonásaikat tekintve ezek hasonlóak egymáshoz, részleteikben és néhány formai előírásukban azonban különbözőek. A sporttudomány és a kapcsolódó tudományterületek esetében – különösen a társadalomtudományban – két fontosabb stílus terjedt el, az MLA (Modern Language Association, *Gibaldi és Aichert, 1984*) és az APA (*American Psychological Association, 1984*). A társadalomtudományi folyóiratok többsége, így a pedagógiai és pszichológiai folyóiratok is alapjaiban az APA stílust használják. A két vezető stílus tartalmi jegyeiben azonos, formai jegyeikben viszont eltérőek. Azonnal szembeötlő különbség a megjelenés évének jelzése. Az APA jellegű stílusoknál a megjelenés éve a szerző(k) neve után szerepel zárójelben, míg a másikonál az évszám a végén, a hozzáférhetőségi adatoknál szerepel. Fenti, szabványnak tekinthető formátumok részletes kézikönyvének, előírásának tárgyalása kereteinket meghaladja. Az egyetemek egyébként is szabályozzák a szakdolgozatok tartalmi és formai követelményeit, ezeket mindenhol be kell tartani. Az említett publikációs stílusok ráadásul angol nyelvterületen

kerültek kidolgozásra, ezért a magyar nyelvi sajátosságok indokolják a saját előírásokat. Fentieknek analógiájára tehát **két fő formát** jelzünk.

#### **4.1.1. Hivatkozás könyveknél és folyóiratoknál**

Könyvek esetében:

- a./ Ács P. (2009): *Sporttudományi kutatások módszertana*. PTE, Pécs. 291 p.<sup>1)</sup>
- b./ 1. Ács P.: *Sporttudományi kutatások módszertana*. PTE, Pécs, 2009.

Folyóiratnál:

- a./ Ozsváth K., Oláh Zs. (2009): TF hallgatók Eurofit értékelési normarendszere / Standardized system for the Eurofit evaluation of P.E. students. *Magyar Sporttudományi Szemle/Hungarian Review of Sport Science* 10. évf. 38. sz. 2009/2 43.p.
- b./ 1. Ozsváth K., Oláh Zs.: TF hallgatók Eurofit értékelési normarendszere / Standardized system for the Eurofit evaluation of P.E. students. *Magyar Sporttudományi Szemle/Hungarian Review of Sport Science* 10. évf. 38. sz. 2009/2 43.p.

---

<sup>1</sup> „page”, vagy „o.”, oldal

### **4.1.2. Hivatkozási alapszabályok**

Minden esetben az eredeti, teljes címnek kell szerepelni az irodalomjegyzékben. A bemutatott példánál szerkesztőségi előírás volt a kétnyelvű cím. Egyéb esetekben nem szabad több nyelven megadni a címet. A cím és a forrás betűtípusát pedig célszerű eltérően megadni – amelynek tekintetében a bemutatott példától eltérő előírások lehetségesek. Formai szempontból az APA jellegű előírásoknál a könyv címek, a folyóirat megnevezések, a szövegbeli hivatkozásoknál pedig a nevek dőltbetűsek.

Második példánknál sorszámot is feltüntettünk, ami szintén használatos sok helyen, és elvileg a szövegen belüli hivatkozásokat hivatott egyszerűsíteni (és nehezen áttekinthetővé tenni).

### 4.1.3. További jelölések

Az irodalomjegyzékkel kapcsolatban még két szokásos jelzésre és rövidítésre térünk ki. Példáink:

Ozsváth K. (1999): The types of motoric and Eurofit tests. *Actes du congres international de l'ASEP Neuchatel 1998.* (Ed.: J.C. Bussard / F. Roth ) 305-309.p.

Ozsváth K. (2009): Adalékok a motoros tehetség problematikájához. In.: *Tanulmányok a kiválasztás és a tehetséggondozás köréből* (Szerk.: Bognár J.). MSTT Budapest, 42-75.p.

A „Szerk.” vagy „Ed.” (Editor) a szerkesztő(k) megjelölése. Akkor használjuk, ha valamilyen nagyobb terjedelmű, mások által szerkesztett kiadványban szerepel az adott tanulmány. Amennyiben a kiadványnak vagy könyvnek csak egy önálló, nevesített szerző által írt fejezetére vagy önálló cikkére hivatkozunk, akkor használatos a cím után az „In.” rövidítés alkalmazása.

## **4.2. Az irodalomjegyzék és szövegbeli hivatkozások összekapcsolása**

Az irodalomjegyzék szerkezetileg általában utolsó része a publikációnak, nagy terjedelmű anyagoknál még követhetik mellékletek. A tényleges szövegben a hivatkozás a szerző(k) nevével és a megjelenés évszámával adható meg. A szerző neve része lehet a mondatnak: „...Ozsváth tanulmányában (1999) közölt eredmények ...”. Gyakoribb azonban a zárójelben feltüntetett azonosítás: „... a pécsi tankönyvben (Ács, 2009) kifejtésre került...”

Több szerző esetén a neveket elvileg vessző választja el. Célszerű azonban, ha két szerző nevét az "és" választja el egymástól: (Ozsváth és Oláh, 2009). Több név esetén pedig az "és" az utolsó két név között szerepel. Sok szerző esetén szokásos még az első szerző nevének kiírása, a többiekét „és mtsai” vagy „et al.” (et alii and others) jelezhetjük.

Az irodalomjegyzéket csak egyes szakterületeken szokás sorszámozni (orvostudományi területen gyakori). Ha előfordul, akkor a szövegbeli hivatkozásoknál esetleg csak egy szám szerepel zárójelben. A megoldás elvileg korrekt, azonban az azonosítást nehezítheti. A név és évszám jelzése hosszabb, de könnyebb az azonosítás, jobb a szöveg áttekinthetősége.

### **4.3. Ábrák és táblázatok kezelése**

Az ábrák és táblázatok kezelésénél is célszerű az APA stílus ajánlásait követni. Mindegyiket arab számokkal sorszámozzuk és megcímezzük, nevesítjük. A szövegben a sorszámok alapján hivatkozunk rájuk<sup>2</sup>. A sorszámozott címek helye az ábrák alatt, illetve a táblázatok felett legyen. Azaz ábrafelirat, kép aláírás alul, táblázat felirat felül! Ábráknál és táblázatoknál is ügyelni kell az áttekinthetőségre, érthetőségre, jelmagyarázatra, rövidítések megadására. Táblázatoknál az oszlopoknak és soroknak is legyen neve.

---

<sup>2</sup> Célszerű automatikus sorszámozást és kereszthivatkozásokat alkalmazni, ha ezt a használt szövegszerkesztő (pl. Word) lehetővé teszi.



#### **4.4. ISBN, ISSN, DOI jelentése**

A irodalom kezelésével foglalkozó fejezet végén meg kell említenünk néhány jelzést. Az ISBN (International Standard Book Number) könyvek és monográfiák nyilvántartására használt nemzetközi azonosító kódszám. Az ISSN (International Standard Serial Number) az időszaki kiadványok (folyóiratok, periodikák) nemzetközi azonosítója. A DOI (Digital Object Identifier) pedig digitális objektumazonosító kódszám, ami az utóbbi években került bevezetésre és digitális szövegek, képek, hanganyagok és audiovizuális művek azonosítására és kezelésére alakították ki.

#### **4.5. Online anyagok kezelése**

Végezetül említést kell tennünk a webes, online anyagok kezeléséről. Egyrészt e téren fellelhetők teljesen megbízható, lektorált, stabil források, pl. DOI azonosítóval rendelkező anyagok, online hozzáférésű könyvtárak anyagai stb. Más források kevésbé megbízhatók, fellelhetőségük is nagyobb idő távlatában nem feltétlenül biztosított. Az online anyagok hivatkozásának sajátossága, hogy a forrás webcímét (URL, Uniform Resource Locator) mindenképpen fel kell tüntetni. Célszerű a lehívás dátumát is jelezni zárójelben. Egyébként törekedni kell a szokásos szerző – cím – forrás megjelölés alkalmazására. A weben található anyagok egy részénél nem deríthető ki a szerző, és sok esetben címe sincs az anyagnak, ekkor csak az URL cím adható meg. A leírtakra két példát hozunk, a konkrét idézet akár mottója lehetne fejezetünknek: „Csak azért, mert valami egyszer nyomtatásra került, még nem biztos, hogy hiteles, míg az online anyagokra sem húzható rá a megbízhatatlanság.” (Miller-Cochran, S., 2008. In: [http://eduline.hu/hirek/20081207\\_kutatasi\\_anyagok\\_hitelessege.aspx](http://eduline.hu/hirek/20081207_kutatasi_anyagok_hitelessege.aspx))

Institute for Scientific Information: Web of Science.  
[http://thomsonreuters.com/products\\_services/science/science\\_products/a-z/web\\_of\\_science](http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science) (2010.08.26.)

## 5. Elméleti alapok

### 5.1. A tesztekkel szemben támasztott alapkövetelmények, kritériumok

A tesztek lényegüket tekintve mérőeszközök. A velük szemben támasztott követelményeket elsőként *Guilford* (1936) határozta be. A későbbiekben sokan foglalkoztak e mérőeszközökkel kapcsolatos elméleti alapkérdésekkel, közülük talán *Lienert* (1961) és *Magnusson* (1975) munkássága a legismertebb. A kérdéskör lényegileg a „mérce”, az „etalon” problematikáját fedi le, és a tesztek standardizálásának<sup>3</sup> tartalmi vonatkozásait foglalja magába. Az alapkérdés, hogy **„mit – hogyan – milyen pontosan mérünk”**? A standardizálás egy vizsgálati mód mérőeszközzé, tesztté válásának útja, és a teszttel szembeni követelmények rendszerének ellenőrzését és teljesítését jelenti. A magyar nyelvben a vizsgálati eljárásokat teszteknek vagy próbáknak nevezzük. Elvileg a nem standardizált eljárások a „próbák”, míg a standardizált eljárások a „tesztek”. A gyakorlatban ezt a finom megkülönböztetést ritkán használják. Az angol „test” szó eredeti magyar jelentése „próba”, de szótárak ma már a „teszt” fordítást is megjelenítik. A megalapozott szakmai következtetések levonásának előfeltétele a vizsgálati eljárások standardizálása.

---

<sup>3</sup> Nem keverendő össze a statisztikai standard értékekkel (Z vagy u), adataink statisztikai standardizálásával! (Lásd későbbiekben.)

## Standardizálás

A tesztek standardizálása lényegében a tesztkritériumok vizsgálatának és meghatározásának folyamata. Szakterületünkön Bős (1988, 2001) sportmotoros tesztekéről szóló kézikönyvei tekinthetők a legteljesebb és legkritikusabb anyagoknak, amelyek a tesztkritériumokat kiemelten kezelik. Itthon Nádori és mtsai (1984, 1989, 1998, 2006) kézikönyvében ugyancsak következetesen fellelhetők az egyes tesztek értékelési kritériumai.

### **A teszteknek az alábbi követelményeknek kell megfelelniük:**

- ***a teszt végrehajtásának állandósága tartalmi és formai szempontból (vizsgálati protokoll megléte és betartása);***
- ***a teszt eredményének összehasonlíthatósága, értékelésének azonossága;***
- ***a tesztekkel szemben támasztott feltételeknek, a tesztkritériumoknak való megfelelés.***

### **Tesztkritériumok:**

- ***fő kritériumok: érvényesség, megbízhatóság, tárgyilagosság;***
- ***mellék kritériumok: gazdaságosság és normatíválhatóság.***

A tesztekkel szemben támasztott legfőbb feltételek, **alapvető tesztkritériumok** tehát az **érvényesség** (validitás), a **megbízhatóság** (reliabilitás), és a **tárgyilagosság** (objektivitás). A mérések lebonyolításának és az eredmények gyakorlati feldolgozásának feltételeként, másodlagos kritériumként jelentkezik a normatíválhatóság (értékelhetőség) és gazdaságosság (ökonómikusság). (Lienert 1961, Magnusson 1975, Nádori és mtsai 1989)

### 5.1.1.1. Érvényesség (validitás)

A **validitás** vagy érvényesség a tesztek legfontosabb alapkritériuma, a mérési eljárással vizsgált jelenség – esetünkben tulajdonság, képesség, készség, kompetencia – meghatározását szolgálja. A validitás vizsgálatával a „**mit mérek**” kérdésre kaphatunk választ. Amikor tesztet dolgozunk ki pl. vívók specifikus (vívásra jellemző, azt meghatározó) mozgásos jellemzőinek mérésére, akkor annak érvényessége megközelítően abban jelentkezik, hogy más sportolók teszteredményei, teljesítményei elmaradnak a vívókétól. Jelentése tehát: a teszttel valóban azt a tulajdonságot, képességet mérjük-e, ami szándékunkban áll, és amelyre kidolgoztuk az eljárást.

Az elsődleges tesztkritériumokat számszerűen általában egy korrelációs együttható szorosságával jellemezzük. Kivétel a **logikai** vagy **tartalmi validitás**, ami egy teszt logikai úton belátható érvényességét jelöli és számszerűen nem fejezhető ki.

#### **5.1.1.1.1. Kritériumvaliditás**

Az érvényesség klasszikus meghatározási és ellenőrzési formája a kritériumvaliditás, amely viszont már számszerűen is vizsgálható. A **kritériumvaliditás** lényege annak meghatározása, hogy a teszt milyen információt tartalmaz a kritériumról. A mért teszteredményeket ez esetben egy kritérium (feltétel) értékeihez kell viszonyítani. A legegyszerűbb esetben ez a kritérium lehet egy másik teszt, amelyről már biztosan tudjuk, hogy mit és hogyan mér. Ilyenkor a két teszt eredményei közötti korreláció jellemzi a validitást. Általánosságban azt mondhatjuk, hogy statisztikai szemszögből a teszt eredménye és a kritérium közötti összefüggés szorosságát vizsgáljuk. Ezt általában a teszteredmények és a kritérium értékei között korrelációs együtthatóval fejezhetjük ki ( $r_{kx}$ ). *Letzelter* (1983) a sport területére kiterjesztett kritériumvaliditás lényegét úgy fogalmazza meg, hogy ***a magasabban kvalifikáltak a gyengébbektől egyértelműen különböznek***. A kritérium ez esetben tehát a sportolók minősítése, amely háttérében a motoros teljesítmény húzódik meg. A sporttudomány területén általánosságban is magát a mozgásos teljesítményt kell alapvető feltételként, kritériumként tekinteni saját mérőeszközeink, tesztheink kialakításához.

#### **5.1.1.1.2. A kritériumvaliditás formái**

A validitást a kritérium jellegzetességei alapján, több formában is kifejezhetjük:

**Gyakorlati validitás:** a kritérium értékei a jelenben ismertek, vagy a jelenben közvetlenül mérhető, illetve megállapíthatók. Tipikus esete egy olyan teszt kritériumként tekintése, amelyet korábban már igazoltak. Így a két teszteljárás eredménye közötti összefüggést vizsgálják. A leggyakrabban alkalmazott validitási forma.

**Predikciós validitás:** a kritérium csak a jövőben lesz ismert, és akkor közvetlenül mérhető vagy megállapítható lesz. Az eljárás hasonlít a gyakorlati validitás vizsgálatához, azonban ellenőrzése hosszú időt igényel. A sporttudományban a kiválasztásnál és a teljesítményprognózisnál alkalmazott mérési eljárásoknál jöhet számításba validitási formaként.

**Faktorális validitás:** a kritérium ez esetben a teszteredmények mögött meghúzódó összetett háttérváltozó, egy közvetlenül nem mérhető latens mennyiség, **hipotetikus faktor, ill. komponens**. Kizárólag faktoranalízis és komponensanalízis útján állapítható meg. Mértékét a teszt faktorsúlya – a faktorsúlyok tulajdonképpen korrelációs együtthatók – adja a kritériumfaktorban vagy -komponensben. Túl gyakran nem találkozunk vele.

### **5.1.1.1.3. Az érvényesség jellemzése**

A standardizálás folyamán bármelyik kritériumvaliditási forma alkalmazható. A kritérium és a teszteredmények közötti **korrelációval** jellemezhető egy teszt érvényessége.<sup>4</sup> Az érvényességet elsősorban az **összefüggés szorossága** jellemzi, amelynek értelemszerűen szignifikánsnak is kell lennie. (A „fordított” összefüggésre utaló negatív előjelű együtthatókat az előjel szempontjából is kell értelmezni.) A kielégítő szorosság tekintetében a szakirodalom nem teljesen egységes, de támpontként szolgálhatnak az 1. táblázat értékei.

Komplex **teszt együtteseknél az eredményt összességében kell értékelni**, ilyenkor az egyes tesztek elvileg elveszítik önállóságukat. A teszt battéria összesített eredménye – például pontszám – és a kritérium közötti összefüggés ez esetben „egyszerű” korrelációval jellemezhető. A teszt battéria elemeinek összefüggése a validitás kritériummal azonban a többszörös korrelációs koefficiens (R) alapján is megállapítható, és a validitás így is jellemezhető. (Meggjegyzem, hogy a többszörös korrelációnál gyakoriak a magas értékű, szoros összefüggésre utaló együtthatók.)

---

<sup>4</sup> Szóráselmzéssel is igazolható bizonyos körülmények között a validitás – bár a varianciaanalízis nem összefüggések vizsgálatára irányul.



### 1. táblázat: A validitási együttható értékelése

#### *A validitás minősítése*

<b><math>r_{kx}</math> értéke</b>	
<b>0,85 - 1,00</b>	<i>kiváló</i>
<b>0,80 - 0,84</b>	<i>jó</i>
<b>0,70 - 0,79</b>	<i>megfelelő</i>
<b>0,60 - 0,69</b>	<i>egyres tesztnél nem, tesztbatteria összetevőként elfogadható</i>
<b>0,00 - 0,59</b>	<i>nem megfelelő</i>

### 5.1.1.2. **Megbízhatóság (reliabilitás)**

A következő mérési alapfogalom a **megbízhatóság** (reliabilitás), ami lényegében a teszt mérési pontosságának alapvető jellemzője. A megbízhatóság megállapítására általában a teszt megismétlését alkalmazzák. Az ismételt teszteredmények az eredetivel azonosnak kell lennie. A megbízhatóság jelentése tehát: a megismételt tesztenél az eredmények nem változnak. A két mérésnél az eljárás, a mérés és értékelés módja, a vizsgálati személyek, a mérő személyzet, és a vizsgálati feltételek nem változhatnak. Azaz azonos vizsgálati személyeknél ugyanazon felmérő személyzet végzi az ismételt vizsgálatot.

A megbízhatóság függ:

- a mérési eljárás pontosságától,
- a vizsgálati személyek teljesítőképességének változásától, amelynek okai nem ismertek.

#### **5.1.1.2.1. A megbízhatóság jellemzése**

A megbízhatóságot is korrelációs koefficienssel ( $r_{xx}$ ) szokták kifejezni, amit a két mérés eredménye között mutatkozik. A mérések közötti különbséget egymintás t-próbával is ellenőrizni kell, az átlagok között nem lehet lényeges eltérés. Ha az összefüggés szoros és a t-próba nem szignifikáns, a teszt megbízhatónak minősíthető. Ha a két mérés közötti korreláció szoros, de a t-próba szignifikáns különbséget jelez, akkor az ismételt tesztvételt befolyásolta az első tesztelés közben szerzett jártasság, begyakorlás, vagy éppen elfáradás. (Azaz a megbízhatóság nem kielégítő.) A megbízhatóság ellenőrzésének alapvető módszerei a „teszt – retest” és a „felezés” módszer.

Motoros teszteknel a sportversenyek analógiájára gyakran előfordul, hogy több kísérlet közül a legjobb eredményt kell rögzíteni a mérési protokoll értelmében. (Tipikus példa erre a helyből távolugrás.) Ez elvileg kiváló lehetőség a megbízhatóság vizsgálatához, de ügyelni kell az egyes kísérletek eredményei közötti különbségre (t-próba).

A megbízhatóság minősítése a validitás tárgyalásánál bemutatott táblázat szerint történhet, de a 0,7-nél kisebb korrelációs együtthatók nem fogadhatók el. A megbízhatóság is növelhető a „teszthossz” változtatásával (Magnusson 1975). (Motoros tesztek esetében pl. a megengedett végrehajtások/kísérletek számának növelésével.)

### **5.1.1.2.2. A megbízhatóság ellenőrzésének alapvető módszerei**

#### **A „teszt – releszt” módszer:**

Az alkalmazott mérési eljárást viszonylag rövid időtartamon belül kétszer alkalmazzuk. Az eredeti és az ismételt tesztvétel eredményei közötti korreláció a stabilitás, időbeli állandóság mutatójaként is értelmezhető. A két mérés közötti teljes kipihenést kell biztosítani a vizsgálati személyeknek. Motoros próbáknál az is fontos lehet, hogy a két tesztvétel között a vizsgálati személyek ne kapjanak más jellegű fizikai terhelést.<sup>5</sup>

#### **A „felezéses” módszer:**

Az eljárás alapesetében a tesztvétel két részeredményre bontható. A teszt részeredményei közötti korreláció az alaki-tartalmi állandóság, más néven a konzisztencia mutatójaként is értelmezhető. A módszer akkor is alkalmazható, ha a teszt végeredménye több részeredmény összesítéséből áll. Így pl. a páros és páratlan sorszámú összetevők részeredményét viszonyítjuk egymáshoz.

---

<sup>5</sup> Egyes motoros tesztrendszereknél az egyedi tesztek, tesztitemek végrehajtási sorrendje többek között ezért is meghatározott. Pl. gyorsasági tesztelés előtt nem szabad állóképességi tesztet elvégezteni, mert a két terhelés „üti” egymást.

### 5.1.1.3. Tárgyilagosság (objektivitás)

A **tárgyilagosság** (objektivitás) azt jelenti, hogy a teszteredmények függetlenek a mérő-értékelő személyétől. Az objektivitás a megbízhatósághoz hasonlóan a mérési eljárás pontosságának egyik jellemzője, csak ezúttal ugyanazon mintán két mérőszemélyzetnek kell azonos eredményt produkálnia egymástól függetlenül. Az ismételt vagy egyidejű tesztvételnél a mérési mód, a vizsgálati személyek és a külső feltételek nem változhatnak. A két mérés során az előírt feltételeket (instrukciók, a végrehajtás módja stb.), azaz a mérési protokollt szigorúan be kell tartani. Az objektivitást a fentiek szerint keletkezett két adatsor közötti korrelációs együtthatóval jellemezzük /  $r_o$  /. Az objektivitás minősítésénél a megbízhatóságnál leírtakkal megegyezően kell eljárni. Itt is igaz, hogy a tárgyilagosság ellenőrzésénél sem elég pusztán a korrelációra hagyatkozni! A mérések eredményének azonosnak kell lennie, tehát az átlagok között sem lehet különbség. Ezt célszerű egymintás t-próbával ellenőrizni.

#### 5.1.1.4. **Gazdaságosság (ökonomikusság) és normatíválhatóság**

A **gazdaságosság** és **normatíválhatóság** a tesztek mellékkritériumai, és a tesztelés gyakorlati lebonyolíthatóságának és értékelésének általános feltételeként jelentkeznek. Ezek a másodlagos kritériumok nem jellemezhetők számszerűen úgy, mint a fő kritériumok.

A **teszt gazdaságossága, ökonomikussága** a mérés idő és energia ráfordításával áll kapcsolatban. Magába foglalja a végrehajtás és értékelés idő- és költségigényét, az eszköz- és műszerigényt, a mérőszemélyzet létszámát, a helyigényt, a tömeges, „forgószínpados” lebonyolítás lehetőségét, a teszteléssel nyert információk gyakorlati felhasználhatóságát. Áttételesen kapcsolódik a gazdaságossághoz a **normatíválhatóság**. A norma viszonyítási alap, **etalon** az értékeléshez. Viszonyítási alap nélkül nem lenne mihez hasonlítanunk a kapott eredményeket. A normák kialakítása reprezentatív mintát feltételez, és igen nagyszámú mérési adat eloszlása alapján valószínűsíthető meg. Az összehasonlításokhoz a szakirodalomban gyakran „csak” úgynevezett **„referencia értékeket”** adnak meg, amivel elkerülhetők a reprezentatív mintával és a normák kialakításával kapcsolatos esetleges szakmai-tudományos viták. A normákat és a referencia értékeket leggyakrabban táblázatokban és/vagy grafikonokon foglalják össze. A táblázatoknak minimálisan magukba kell foglalniuk a különböző szempontok szerint elkülönített csoportok (nem, életkor, esetleg sportág, minősítés stb.) középértékeit és szórásait. A „komolyabb” normarendszerek ennél jóval részletesebbek, jelzik a szélsőértékeket és a különböző percentiliseket, így megadják a „proporciókat” (magyarul arányokat, százalékos értékeket). A normák sok esetben minőségi kategóriákat is magukba foglalnak, illetve meghatároznak. Ilyenkor a kategória határok kialakításának szempontjait egyértelműen jelezni kell. Egyes esetekben a normák statisztikai modellek segítségével is megadhatók, e téren elsősorban a regressziós és a diszkriminatív modellek jöhetnek számításba.

## **6. Az adatfeldolgozás módszerei**

A számítógépek térhódításával egy időben jelentek meg a különböző statisztikai programok. Napjainkban már egyes irodai alkalmazásokat tartalmazó programok is tartalmaznak statisztikai függvényeket. Így például a Microsoft Office táblázatkezelője, az Excel is. Lehetőségei azonban nyilvánvalóan messze elmaradnak a célzott statisztikai programokétól.

### **6.1.1. Az adatrögzítés javasolt formája**

Részemről azt szoktam ajánlani, hogy a vizsgálati adatokat Excelben rögzítsék, de a tényleges adatfeldolgozáshoz valamilyen statisztikai programcsomagot használjanak. Az Excel ugyanis gyakorlatilag mindenki számára hozzáférhető, az adattáblázata nagyon egyszerűen kezelhető, és a grafikai lehetőségei is jók. A „komolyabb” statisztikai programcsomagok pedig kivétel nélkül kezelni, illetve konvertálni tudják az Excelben rögzített adatokat. Az is az Excel mellett szól, hogy a statisztikai programcsomagok gyakran időköddal védettek, és ennek lejárta után a speciális formátumban mentett adatbázisok nem lesznek hozzáférhetőek a továbbiakban. Az Excel esetén ez a veszély nem áll fenn.

A változók (variables) az oszlopokban, az esetek/vizsgálati személyek (cases) a sorokban szerepeljenek! (Tehát praktikusán az Excel első sorának cellái tartalmazzák a változók megnevezését. A következő sorokban a változókhoz tartozó egyéb, részletező információk is szerepelhetnek – pl. a kérdőív teljes konkrét kérdései és a lehetséges válaszok – a tényleges esetek előtt. Ezek a kiegészítő információkat hordozó „esetek” kizárhatók a későbbi feldolgozásból.)

Általánosságban javasolható, hogy mind az adatainkat, mind az adatfeldolgozás eredményeit több formátumban is mentjük el. Így a későbbiekben is biztosan hozzáférhetünk minden adatunkhoz és eredményünkhöz az időkorlátos licenszek lejárta után.



## 6.2. Statisztikai programcsomagok

Számos statisztikai programcsomagot fejlesztettek ki az utóbbi évtizedekben. A fejlesztők jellemzően amerikai egyetemek és tudományos kutatóintézetek közreműködésével a tudományos, mérnöki/ipari és üzleti statisztikai eljárások szoftvereit készítették el. Kezdetekben alapstatisztikák és grafikonok készítésére, és a „saját” tudományterületük jellemző statisztikai eljárásainak elvégzésére és adatelemzésére szolgáló programok készültek el. A statisztika azonban nem tudományág specifikus, így a programok egyre komplexebbek lettek, napjainkra jellemzően részben önálló modulokból épülnek fel. A kezdetekben néhány fős kis fejlesztő csoportok közül a legéletképesebbek nagy, tőkeerős, profitorientált cégekké nővekedtek.

A piacvezető szoftvereket folyamatosan fejlesztik, és egyre újabb verzióik kerülnek a piacra. Ezek napjainkra már annyira fejlettek, olyan sokat tudnak, hogy mellettük „újak” egyre kisebb valószínűséggel tudnak piacra kerülni. Bár a szoftvereknek az ára eléggé borsos, adataink feldolgozásához mégis célszerű lehetleg a piacvezető szoftverek valamelyikét választani. Nagyobb cégek, egyetemek egészen biztosan rendelkeznek legális statisztikai szoftverrel. A legnagyobb statisztikai szoftvercégek egyébként nonprofit oktatási-kutatási célokra általában kedvezményesen adják, esetenként reklámcélokból ingyenesen is hozzáférhetővé teszik programcsomagjaikat. Az egyetemi szférában mindezeket központi kormányzati projektek is támogatják. Meg kell jegyezni, hogy napjainkban e szoftverek fő piaci vadászterülete nem is a „hagyományos” statisztika, hanem az „adatbányászás” és újabban a „szövegbányászás”<sup>6</sup> – amivel elsősorban a nagy ipari, kereskedelmi és szolgáltató cégeket, bankokat célozzák meg.

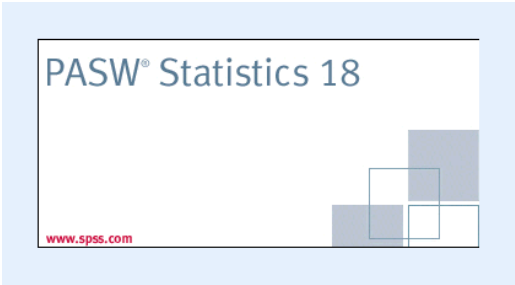
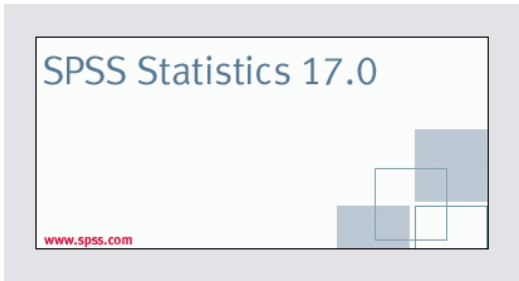
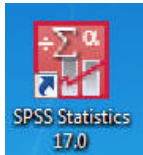
A továbbiakban a jelenleg Magyarországon legismertebb programcsomagokra térünk ki röviden.

---

<sup>6</sup> Nagy és összetett adatbázisok különféle elemzésére és folyamatellenőrzésére szolgáló statisztikai alapú, speciális eljárások. Elsősorban az üzleti életben használatosak. Az utóbbi években megjelentek és rohamosan fejlődnek a szövegelemzéssel foglalkozó statisztikai szoftverek is, amelyek az adatbányászati technikák analógiájára kerültek kifejlesztésre. (Data Mining, Text Mining)

### 6.2.1. SPSS

Az SPSS (Statistical Package for the Social Sciences) a világ piacvezető statisztikai szoftvereként hirdeti önmagát – nem teljesen alaptalanul. Nevének megfelelően eredetileg a társadalomtudományok területét célozta meg, és használata az egyetemi-akadémiai szférában világszerte elterjedt. Kezdetei 1968-ig nyúlnak vissza, amerikai-kanadai gyökerekkel. Statisztikai programjai teljes körűek, rendkívül jól kezelhető, a világon valóban mindenhol megtalálható, népszerű programcsomag. A statisztikai eljárások csoportosítása jól áttekinthető, egyértelmű és tiszta logikát követ. Az összetettebb, „haladó” eljárások megtalálása viszont a súgó használata nélkül nem mindig egyszerű. Beállítási lehetőségei rendkívül változatosak, ezért néha kissé nehézkesnek tűnhet, használatát mégis hamar meg lehet szokni. Súgó, oktató és „edző-tanácsadó” programrészei is igen jól használhatók és nagyon részletesek. Grafikája nem túl látványos, nem tartozik az erősségei közé, de tág határok között állítható. 2005-ig a magyar felsőoktatás oktatási-kutatási célokra – dátumkódos korláttal – ingyenesen használhatta a programcsomagot. 2005 őszén ezt a rendszert megszüntették, illetve teljesen átszervezték, de az egyetemeken továbbra is kedvezményesen juthatnak hozzá a programcsomaghoz. 2006 elején a 14.0 verziónál tartottak, azóta évente jön az újabb verzió, már piacra került a 19.0 változat is (2011). Az egyes verzióknál kisebb-nagyobb módosítások fellelhetők, a jól bevált alapokon azonban szerencsére nem változtattak. Az adat fájlok \*.sav vagy \*.por kiterjesztésűek, ugyanakkor számos más formátumot – közte a fő vetélytárs SAS formátumokat – támogat a programcsomag. Excel, dBase, Lotus formátumban is értelemszerűen minden adat megnyitható és menthető. A programcsomagról és a cégről aktuális információk a [www.spss.com](http://www.spss.com) és a [www.spss.hu](http://www.spss.hu) weboldalakon találhatóak.





3. ábra: Az SPSS ikonja és indító ablaka

### 6.2.2. SAS

Az SAS (Statistical Analysis System) talán a legnagyobb vetélytársa az SPSS-nek, a világ vezető üzleti analitikai szoftvereként hirdeti önmagát. Szintén amerikai eredetű, 1976-ban alapították a fejlesztő céget. Rendkívül komplex szolgáltatásokat tartalmazó, eredendően számítógépes hálózatokra, üzleti és banki alkalmazásokra fejlesztett programcsomag. Statisztikai programcsomag moduljai teljes körűek, de az átlagos felhasználók számára valószínűleg kissé nehezebben kezelhető vetélytársainál. Az egyetemi-akadémiai szférában nem igazán tud gyökeret verni, bár a SAS Egyetemi Programja hazánkban is megindult. A program magyarországi elindításával a hazai egyetemek számára is könnyen elérhetővé váltak a SAS szoftverei, amennyiben az egyetem azokat oktatási és tudományos munkájában kívánja használni. Az adatfájlok \*.sd2 vagy \*.xpt kiterjesztésűek. 2006 elején a 9.1.3 verziónál tartottak, 2010-ben pedig a 9.2 változatot használják. A programcsomagról aktuális információk a [www.sas.com](http://www.sas.com) és a [www.sas.com/offices/europe/hungary/](http://www.sas.com/offices/europe/hungary/) weboldalakon találhatóak.

### 6.2.3. StatSoft STATISTICA

Az 1984-ben alapított StatSoft Inc. társaság nemes egyszerűséggel STATISTICA néven forgalmazza programcsomagját. Az előzőekhez képest talán ez a leginkább felhasználóbarát programcsomag, nagyon szép és sokoldalú grafikai háttérrel és lehetőségekkel. Az összes programcsomag közül jelenleg ez idomul legjobban a Windows környezethez. Külön támogatja a kezdő felhasználókat a minden alkalmazásnál megtalálható, leegyszerűsített „Quick” kezelőfelületekkel. Ugyanakkor a statisztikai eljárások, analízisek csoportosítása kissé sajátos logikát követ. Az egyszerűbb eljárások könnyen megtalálhatók és könnyen kezelhetők (Basic Statistics), bár csoportosításuk már itt sajátos. Ezen túlmenően – és pl. az SPSS-hez viszonyítva – viszont igencsak „keresgélgni” kell az egyes eljárásokat, és nagyon könnyen valamilyen „komplikált” változatot sikerül elindítani. A programcsomag nagy előnye, hogy a futó analízisek bármelyik lépéséhez könnyedén vissza lehet térni, és az esetleges szükséges pontosításokat könnyű elvégezni. A programcsomag megengedi párhuzamosan több analízis megnyitását, de ettől függetlenül is „hajlamos” kissé túl sok ablak megnyitására. Az adat fájlok \*.sta kiterjesztésűek, és ez a programcsomag is számos más formátumot támogat. Excel, dBase, Lotus/Quattro formátumban értelemszerűen minden adat megnyitható és menthető. 2006-ban a 7.1 verziónál tartottak, 2009-ben kijött a 8.0 verzió, 2010-ben pedig a STATISTICA 9. Ezek a változatok – a korábbiakkal szemben – már teljes körűen képesek az SPSS és a SAS formátumok kezelésére is. A program egyetemi keretek között itt is időköddal védett. Az SPSS-hez hasonlóan az egyes verzióknál kisebb-nagyobb módosítások fellelhetők. A bevált alapokon azonban szerencsére eddig a StatSoft sem változtatott – így a korábbi tan- és szakkönyvek itt is jól használhatók. A programcsomagról aktuális információk a [www.statsoft.com](http://www.statsoft.com) és a [www.statsoft.hu](http://www.statsoft.hu) weboldalakon található.



**4. ábra: A StatSoft STATISTICA ikonja és indító ablaka**

#### **6.2.4. BMDP**

A BMDP (Biomedical Data Processing) is több mint 30 éve a statisztikai szoftverpiac szereplője, a PC-k előtti „nagygépes” korszak igen sikeres képviselője. A korrallal haladva Windows alapú változatai is kifejlesztésre kerültek. Nevének megfelelően az élettudományok képviselői alkalmazzák előszeretettel. Ugyanazokat a statisztikákat „tudja”, mint a többi programcsomag. Az SPSS és az SAS, valamint a leggyakoribb adatbázis-kezelőkben létrehozott adatfájlokat is tudja kezelni. A programcsomagról aktuális információk a [www.statsol.ie/bmdp/bmdp.htm](http://www.statsol.ie/bmdp/bmdp.htm) weboldalon található.



### **6.3. Statisztikai alapfogalmak**

A statisztika a tömegjelenségek leírásával és jellemzésével foglalkozó tudományág. Megállapításai és eredményei egyedi esetekre, egyénekre csak rendkívül korlátozottan és nagy hibahatárokkal vonatkoztathatók. A jelenségek leírásához többnyire elégséges a számtani alapműveletek használata. A jelenségek sokoldalú jellemzéséhez, összefüggéseik és sajátosságaik, törvényszerűségeik feltárásához azonban összetettebb matematikai módszerek alkalmazása is szükséges. Ez már a matematikai-statisztika területe. Központi eleme a becslés, és a becsléshez kapcsolódó valószínűségek, hibahatárok elemzése. Mindezeket különböző célú hipotézisvizsgálati eljárásokkal oldja meg. A biológiai jelenségek vizsgálatára szolgáló matematikai-statisztikai módszereket – pontosabban e módszerek alkalmazását biológiai jelenségekre – biometriának is szokás nevezni. (A kifejezés nem keverendő össze az egyedi azonosításra alkalmas biometrikus jellemzőkkel, mint pl. az ujjlenyomat.)

### 6.3.1. Populáció és minta

Az alapfogalmak közül elsőként a populáció és a minta meghatározásával kell kezdenünk. **A populáció – magyarul „alapsokaság” – valamilyen ismertető jegyek, tulajdonságok alapján összetartozó egyedek összességét jelenti.** A populáció általában nagy, igen sok egyedből áll. Legszélesebb értelmezésében minden korábban létezett, és a jövőben létező egyed és eset beletartozik. Teljes körű felmérésére tehát nem nyílik lehetőség, ezért csak mintát veszünk az alapsokaságból. **A minta a populáció vizsgált része.** Az azonban nem teljesen mindegy, hogy honnan, a populáció melyik részéből veszünk mintát. A mintának ugyanis jól kell képviselnie, jellemeznie a populációt. Bár a populációt valamilyen közös jellemzők, ismertető jegyek alapján minősítjük összetartozó egyedek összességének, azonban más – „alacsonyabb rendű” – szempontok szerint egymástól részben elkülöníthető részei is lehetnek. A populációból vett mintának e részeket is képviselnie, reprezentálnia kell. A reprezentatív minta előre meghatározott szempontok szerinti, meghatározott mintavételi eljárással kiválasztott, általában nagy elemszámú mintát jelent. Újabban a szükséges minta megválasztásának elősegítéséhez speciális modulokat is fejlesztenek a legnagyobb statisztikai szoftvercégek.

### 6.3.2. Adatok, skálák

A vizsgálatok során kapott/keletkezett adatok jellegzetességei behatárolják az adatfeldolgozás lehetőségeit és kereteit. Az adataink jellemzőivel ezért tisztában kell lennünk. Az adatok jellemzői több szempont szerint csoportosíthatók.

#### 6.3.2.1. Az adatok jellege

Az adatok **jellegük szerint** lehetnek:

- minőségi / megállapítható / kvalitatív / nem metrikus skálán elhelyezkedő, vagy
- mennyiségi / mérhető / kvantitatív / metrikus skálán elhelyezkedő adatok.

A két típust alapján az adatok mérhetősége különbözteti meg. Általánosságban elmondható, hogy a mennyiségi adatokból mindig képezhetők minőségi jellegű adatok (gyakoriságok, kategóriákba/osztályokba sorolások, rangsorok), míg fordítva ez alapesetben nem lehetséges. A mennyiségi adatok sokszor összevonhatók, átlagolhatók, a minőségi adatoknál ez többnyire értelmetlen rangsorok és kategóriák esetében egyaránt. (Pl. ha a férfiak=1, nők=2 kódolással megállapítható adatokat képezünk, ezek összege=3 értelmezhetetlen, átlagolásuk pedig pusztán azt mutatja, hogy milyen arányú a két nem aránya az adott mintában.)

### 6.3.2.2. Az adatok értékkészlete

Az adatok *értékük / értékkészletük* szerint lehetnek:

- bináris,
- diszkrét,
- folytonos adatok.

A bináris adatoknál mindig csak két (tetszőleges) érték fordulhat elő. Igen gyakori a „0-1” („nem - igen”) érték – bár ez a későbbi számításoknál gondot is okozhat, mivel nullával nem lehet osztani. Általános szabályt nehéz kimondani, de ha lehet, kerüljük el a nulla érték kategória változóként kezelését, esetleg csak az „igen-nem” típusú változóknál használjuk<sup>7</sup>.

A diszkrét adatok jellemzője, hogy „pontoszerűek”, az értékek között nincs folytonosság, a legtöbb esetben az értékek közötti tartomány nem is értelmezhető. Tipikus példái a rangsorok, darabszámok, évszámok, kategóriába sorolások stb. (Bár pl. a „kapcsolt rangok”, vagy az évszámok esetében részben értelmezhető több érték közötti tartomány is – lásd a későbbiekben.)

A folytonos adatok tetszőleges pontossággal megadhatók és bármely két érték közötti tartomány is értelmezhető. A „folytonosság” mérhető adatok értékkészletének jellemzője.

---

<sup>7</sup> Előfordulhatnak persze olyan esetek is, ahol éppen „0-1” értékek használata a kívánatos. Bővebben lásd a „dummy” változóknál a könyv 4.9 fejezetében.

## Az adatok skálája

Az adatok a **skála típusa szerint** is csoportosíthatók. Az alábbi skálák különböztethetők meg:

- nominális / névleges skála („igen-nem”, „egyezik - nem egyezik”, sorba nem rendezhető kategóriák) >>> kvalitatív
- ordinális / sorrendi / rendező skála („nagyság szerinti sorrend”) >>> kvalitatív jellegű
- intervallumskála („értékei között azonos intervallumok helyezkednek el”, a különbségek számszerűek, de diszkrét jellegűek, és nem ismert vagy értelmezhetetlen a zéró pontja; pl. pontrendszerek, IQ) >>> kvantitatív, de az arányoknak nincs feltétlenül érdemi információértéke, ugyanakkor hasonlít az arányskálára
- arányskála (ismert a „kezdőpont”, a „nulla” pont, és valamilyen „etalonhoz” viszonyít. Hasonlít az intervallumskálához, de annál általánosabb, és egyértelmű arányokat jelent. Tipikus példái bármelyik általánosan elfogadott mértékrendszerben mért adatok.) >>> kvantitatív, elvileg folytonos

Az első kettő nem metrikus skála, az utolsó kettő metrikus skála.

#### **6.3.2.2.1. Névleges skála**

A nominális (nominal, categorical) skála kategóriákat, csoportokat foglal magába. Eredendően nem számszerű, számszerűsítése csak kódolásnak tekinthető. A kódszámokkal értelmezhető számtani művelet nem végezhető. Gyakran csak bináris adatokat tartalmaz („két kategória”). Több kategória esetén az értékkészlete is nagyobb, de ez esetben is mindig diszkrét értékeket képez. Fontos, hogy a nominális skála értékei egymáshoz nem viszonyíthatók, nem adhatók össze, nem rendezhetők, nem átlagolhatók, nincs „kisebb-nagyobb”, „jobb-gyengébb” stb. A nominális skála mindig kvalitatív adatokat jelent, és ennek megfelelően soha nem tartalmaz folytonos eloszlású adatokat.

#### **6.3.2.2.2. Rendező skála**

Az ordinális (ordinal, ordered categorical) skála valamilyen szempont(ok) szerint rendezett több kategóriát, csoportot vagy eseteket foglal magába. Eredendően nem szám, de számszerűsítése egymáshoz való viszonyításnak tekinthető. Gyakran más számszerűsíthető vagy mért eredmény relatív általánosításának tekinthető, amely már független az „eredeti” abszolút különbségektől (pl. versenyek esetén). Tipikus esetei a rangsorok és a „növekvő” vagy „csökkenő” kategóriánkénti előfordulási gyakoriságok. Az ordinális skála is mindig diszkrét adatokat tartalmaz – a kapcsolt rang kivételével – és alapjaiban kvalitatív jellegű adatokat jelent. (Azért csak „alapjaiban kvalitatív”, mert bizonyos körülmények között az egyes kategóriák statisztikai osztályoknak, illetve adott esetben intervallumskálán elhelyezkedő változóknak is tekinthetők. Folytonos adatok is besorolhatók rendezetten „növekvő kategóriákba”, osztályokba. Mindezeknek azért van elvi jelentősége, hogy az eredetileg ordinális skálán elhelyezkedő adatok feldolgozásánál alkalmazhatunk-e paraméteres eljárásokat – amelyek eredendően kvantitatív adatokat feltételeznek.)

A rendező skála speciális esete a rangskála, ami csökkenő vagy növekvő „eredmények” szerinti rangsort (rank) jelent. A „holtversenyek” kezelésére használatos a „kapcsolt rang” használata, ami nem csak egész szám lehet, tizedes értékeket is felvehet. A kapcsolt rang (tie) vagy rangszám a sorba rendezett „azonos értékű” adatok rangszámainak „átlagolását” jelenti. (A 3. helyen kettős holtverseny esetén 3,5 mindkét esetről, ugyanez hármas holtverseny esetén mindegyiknél 4.) A rangskála azért fontos, mert – szemben a gyakorisági adatokkal – itt összefüggések vizsgálata (rangkorreláció) is jól megoldható.

#### **6.3.2.2.1. Az iskolai osztályzatok**

Tulajdonképpen az iskolai osztályzatok is ordinális skálán helyezkednek el, az „eredeti” és egymáshoz képest rendezett kategóriák: elégtelen, elégséges, közepes, jó, jeles. Ez a minősítés jól számszerűsíthető, de az így kapott adatok egyértelműen diszkrét és kvalitatív jellegűek. Más kérdés, hogy a pedagógiai kutatásokban az osztályzatokat – az „erősebb” paraméteres statisztikai eljárások alkalmazhatósága miatt – többnyire intervallumskálán elhelyezkedő értékeknek tekintik, lásd Falus I. (1993, 2000, 2004) munkáit. Nem is alaptalanul, mert elvileg lehetne pontosabban „mérni” a teljesítményeket, és egységes intervallumok vannak az egyes értékek között. Arról nem szólva, hogy a záró osztályzatok többnyire kerekítések eredményei. A pedagógiai értékelés pedig számos más, a nálunk általában használnál sokkal részletesebb, nagyobb terjedelmű skálát is használ. Más oldalról viszont pl. a 4-es osztályzat nem egyértelműen „kétszer jobb” a kettesnél, és pláne nem „négyeszer jobb” az elégtelennél, a skála zéró pontja pedig értelmezhetetlen – azaz tipikus intervallumskálaként is felfogható.



#### **6.3.2.2.2. Likert skála**

Alapjaiban tipikus ordinális skála a kérdőíves módszereknél gyakran alkalmazott, attitűd vizsgálatokhoz kidolgozott, eredetileg ötfokozatú Likert-skála. Néhány fokozatú terjedelme és diszkrét értékei miatt „alapjaiban” kvalitatív jellegű. De ez a skála lehetne százas, ezres vagy még nagyobb terjedelmű – amitől persze a „pontossága” nem feltétlenül javulna. Az adatfeldolgozás során hasonló esetekben általában már a paraméteres eljárásokat alkalmazzák, lényegében kvantitatívnak – és egyszerűen „csak” osztályba soroltnak, így intervallumskálán elhelyezkedőnek – tekintve a felmérési eredményeket.

### **6.3.2.2.3. Intervallumskála**

Az intervallumskála (interval) olyan pontosan behatárolható adatokat foglal magába, amelyek meghatározott feltételek között, egységes intervallumokon belül értelmezhetők. Eredendően számszerű, mennyiségi jellegű, de az esetek többségében diszkrét értékeket követ az egységes intervallumok miatt. (Elvileg nincs törtrésű intervallum, nincs „fél” alma, nincs „fél” hiba). Legfontosabb jellemzője, hogy az egyes intervallumok közötti számszerű különbségek nem jelentenek egyértelmű arányviszonyokat.<sup>8</sup> Másik fontos jellemzője, hogy nincs egyértelmű zéró pontja. Mindezek következtében a skála számszerű értékeinek összege vagy egymáshoz viszonyított aránya nem értelmezhető. (Szemben az arányskálával.)

Az intervallumskála ezzel együtt mindig kvantitatív adatokat tartalmaz. A kutatói gyakorlatban az intervallumskálán elhelyezkedő adatok feldolgozásánál bevett gyakorlat a paraméteres eljárások alkalmazása – jöllehet ez elvileg feltételezi a kvázi folytonos eloszlást. Ugyanakkor pl. pontszámok esetén nincs elvi akadálya a tizedes értékek használatának. Hibaszámoknál viszont értelmetlen törtrésű hibákról szólni, de azt sem lehet mondani, hogy kétszeres hibaszám egyértelműen kétszeres teljesítményromlást, negatív hatást okoz. Az intervallumskála lényegének megértése alapvető jelentőségű, mert a társadalomtudományok számos területén mindent elkövetnek, hogy a kutatások során kapott adatok ezen a skálán elhelyezkedőnek tekinthetők legyenek (lásd 4.9 fejezetet).

---

<sup>8</sup> A nem egyértelmű arányviszonyokra nagyon szemléletes *Szokolozsky* (2004) „a zseni és az idióta” példája: az IQ skálán 80 és 160 pontot elérő két személy esetében nem mondható az, hogy az egyik kétszer okosabb a másiknál.

#### **6.3.2.2.4. Arányskála**

Az arány- vagy arányos (proportional) skála tulajdonképpen hasonló az intervallumskálához, csak annál általánosabb és teljesen egyértelmű arányokat jelez. Valamely „etalonhoz” viszonyít, és a skálának egyértelmű a nulla pontja. Annyiban hasonló az intervallumskálához, hogy az „etalon” jelenti a skála alapintervallumát, ami a nagyságrendektől függően akár különböző dimenziókban is megragadható, tetszőleges helyi értékű pontossággal. A skála számszerű értékeinek összege, egymáshoz viszonyított aránya egyértelműen értelmezhető. Eredendően számszerű, és folytonos eloszlású, kvantitatív adatokat foglal magába. Az összes használatos mértékegységünk arányskálát képez. A mérési pontosság kizárólag technikai kérdés. Használatánál mindössze arra kell ügyelni, hogy az azonos jellegű, de különböző dimenziójú mértékegységek egymástól eltérő számrendszerűek lehetnek. A mértékegységek átváltásánál főleg az időadatoknál kell figyelni.

#### **6.3.2.2.5. „Mérési szint”**

A társadalomtudományok területén – elsősorban szociológusok – előszeretettel használják a „mérési szint” kifejezést, ami lényegében a skálatípusokat fedi le. A megközelítés kiindulás alapelve, hogy méréskor számokat rendelnek hozzá jelenségekhez (változókhöz) és tulajdonságaikhoz valamilyen szabály(rendszer) szerint. Az így kapott adatokkal a mérési szinttől függően különböző számtani műveletek hajthatók végre. Az alacsony vagy alsó mérési szintek a névleges és rendező skála, a magas vagy felső mérési szintek az intervallum és az arányskála. A legalsó mérési szint a névleges skála, a legmagasabb mérési szint az arányskála. Tehát a mérési szintek „sorrendje”: névleges, rendező, intervallum, arányskála. (Az alacsony mérési szinteken lényegében „kódszámokat” használunk, kiterjesztett értelemben még rangszámok esetében is.)

Az alacsony mérési szintek kezelésére a nemparaméteres (nonparametrics), a magas mérési szintek kezelésére a paraméteres eljárások szolgálnak. Utóbbiak az „erősebb” módszerek „számszerűségük” (kvantitatív) miatt.

### **6.3.2.3. Az adatok skálatípus szerinti feldolgozhatósága**

Az adatok csoportosítása szempontjából a skála típusa szerinti besorolás a legfontosabb. Ez ugyanis egyértelműen behatárolja az adatok feldolgozhatóságának kérdését. A névleges és rendező skálák esetében csak gyakoriságokat, százalékos arányokat vizsgálhatunk, rangskálák esetében még rangkorrelációt lehet számítani. A sporttudomány területén (itthon) az utóbbi időkben elfogadott nézet szerint szentségtörés számba menne például kérdőíves adatokra többváltozós paraméteres eljárásokat „ráereszteni”. Részemről, személy szerint ezt a hozzáállást szélsőségesnek tartom. E nézet képviselőinek ajánlom, hogy kissé nézzenek utána a szociológiai szakirodalomnak. Külön ajánlom figyelmükbe Székelyi M.- Barna I. (2005) SPSS-el kapcsolatos módszertani kézikönyvét – amely kizárólag többváltozós technikákat tárgyal. A kulcskérdés az, hogy milyen technikákkal lehet alapjaikban megállapítható adatokat intervallumskálán elhelyezkedőnek tekinteni, illetve intervallumskálára „forgatni”, transzformálni. A kérdés nem új keletű, hiszen a már említett Likert-skálát pont emiatt találták ki. Nyilván az sem véletlen, hogy az elvi vitákat megkerülő szejt tudományterületükre jellemző statisztikai terminológiát használnak a társadalomkutatók. Példaként az előző pontban említett „mérési szint” és a „dummy változó”, a „dummyzás” esetét hoznám fel.

(Dummy: ál-, formális, látszólagos. Az autók ütközési tesztjeinél alkalmazott tesztbábut is szokás „Dummy”-nak nevezni. A statisztikai zsargonban a „dummy variable” vakváltozóként ismeretes.)

### **6.3.2.3.1. „Dummyzás, vakváltozás”**

Társadalomkutatók számára e fogalmak nem ismeretlenek, míg a sporttudományt űzők jelentős részének újszerűek lehetnek. A mérési szinthez jelzőket is szokás kapcsolni, így pl. alacsony meg magas mérési szintet gyakran említene. A fogalom kapcsán lényegileg az alkalmazott skála típusáról van szó, és maga a besorolás is intervallum jellegű a jelzős nyelvtani szerkezet révén. A névleges skála alacsony mérési szintet, az intervallum és különösen az arányskála magas mérési szintet jelent. A társadalomkutatók tehát ebben az értelemben minden keletkezett adatukat valamilyen mérés eredményének tekintik. Függetlenül attól, hogy ténylegesen mért vagy megállapított adatokról van-e szó, hiszen arra a „mérési szint” eleve utal. A gyakorlatban még kérdőíves módszerekkel kapott gyakorisági értékeket is felhasználnak, „befogatnak” pl. egy 100 fokozatú Likert-skálába. Ebbe a gondolatmenetbe illeszkedik a „dummy” bináris változó, amely „lefordítva” olyan intervallumskálát jelent, amely egyetlen intervallumból áll – és kivételesen esetleg a nulla pontja is értelmezhető. (Vagy ha így valakinek jobban tetszik: az intervallumskála két pontszerű intervallumot foglal magába, a pontokon kívüli területek értelmezhetetlenek. Az esetlegesen számított, a két pont számszerű értéke közé eső „statisztikák” – pl. átlag – pusztán a két végpont előfordulási arányára utalnak.) A dummyzás a kérdőívek adatainak „igen-nem” szintű kezelését jelenti. Ha az intervallum kezdetét „0=nem”, az intervallum végét pedig „1=igen” képezi, a számszerűség miatt még regressziós modellben is értelmezhető eredményeket kaphatunk. Vegyük észre az analógiát a számítógépek működési alapelveivel! A dummyzás esetében természetesen tetszőleges két számmal kódolható az „igen-nem/van-nincs” esete, de a további számításokhoz a „0-1” a leghasználhatóbb, minden más megoldás csak értelmezési problémákhoz vezet. Pl. a hazánkban használatos nembeli kódnál „1=férfi”, „2=nő” használata az elmúlt közel 30 évben megszokottá vált. Ez csoportosítási változóként kitűnően használható, jelentését is gyakorlatilag mindenki tudja. Dummy változóként bevonva valamilyen többváltozós analízisbe viszont már értelmezési gondokat eredményezhet, ez esetekben célszerű legegyszerűbb lineáris transzformációként eggyel csökkenteni számszerű értékét (0=férfi, 1=nő). Ezzel együtt most is hangsúlyoznám, hogy a hazai sporttudományi gyakorlatban hasonló esetben nagy valószínűséggel azonnal elkezdenék vitatni az esetleg alkalmazott statisztika adekvát, megengedhető voltát.

### 6.3.3. Változó, paraméter

Adataink változókhoz, paraméterekhez tartoznak. A két fogalom jelentése hasonló, az általánosabb jelentésű a változó, pontosabban **valószínűségi változó**. A fogalom alatt az adott populációban vizsgált jelenség/objektum nem állandó értékű, hanem a valószínűségi törvények szerint változó, a véletlentől is függő, de azonos módon rögzített jellemzőjét értjük. **Paraméternek** a vizsgált objektum/jelenség mért, számszerű jellemzőjét, tulajdonságát nevezzük, amelynek az alábbiak a sajátosságai (*Fábián-Zsidegh* 1998):

- számszerű, mennyiségi jellegű,
- egyetlen számmal jellemezhető,
- egyértelmű,
- pontos, értelmezhető.

A változóval szemben nincsenek ilyen megkötések, általánosabban használható a fogalom, vagy ha fentieknek nem teljesen felel meg a vizsgált jelenség/objektum valamely jellemzője. A két fogalom közti különbségekre utal a statisztikában a paraméteres és nemparaméteres eljárások megkülönböztetése is. (Előbbi a mennyiségi, utóbbi a minőségi adatok feldolgozására szolgál.) Ugyanakkor figyeljünk fel arra, hogy a paraméter jelen meghatározásánál nem kikötés az arányskála használata, csak a mennyiségi jelleg. A fogalom meghatározása és megkülönböztetése szorosan kapcsolódik az előzőekben az intervallumskálával kapcsolatba említett problémakörhöz.

#### **6.3.4. Függő és független változók**

A statisztikában gyakran előfordul még a ***függő és független változók*** megkülönböztetése. A gyakorlatban ez azt jelenti, hogy egyik tulajdonság függvényében miként változik egy másik tulajdonság, ami értelemszerűen többváltozós esetekre is értelmezhető. Szűk értelemben csak a regresszió számításoknál használjuk. Tágabb értelmezésben az analízis tárgya a függő változó, amelyet az adott vizsgálati beállítás függvényében elemezünk. Különbségek elemzésénél például a kategóriák, csoportok tekinthetők független változónak, melyek „függvényében” vizsgáljuk a különbségek alakulását és jelentőségét, azaz szignifikanciáját. (Az adatbázisban a legtöbb esetben létre is kell hozni egy vagy több „csoportosítási”, besorolási változót – ami jellegében kvalitatív és diszkrét értékeket vesz fel.)



### 6.3.5. Hipotézisek, szignifikancia

A **hipotézis** feltételezést, feltevést jelent, formájában **igazolásra szoruló állítás, kijelentés**. A feltételezés a vizsgálat várható eredményére irányul. Az **igazolt hipotézis** pedig a **tézis**. (A vizsgálatok jelentős részénél több hipotézis is megfogalmazható.)

Fentiek – és negatív tapasztalataim – következtében fontosnak tartom hangsúlyozni, hogy a hipotézisek tulajdonképpen sohasem lehetnek kérdő mondatok! Mi van ugyanis előbb, a kérdés vagy felelet, azaz az állítás? Ha szembekerülünk egy problémával, az kérdéseket vet(het) fel. A megoldás várható irányát a feltételezés(ek) körvonalazzák, a kérdésre adott (igazolt) válasz(ok) pedig a tézis(ek). Függetlenül attól, hogy az „eredmény” negatív vagy pozitív, esetleg „semleges”, eldöntetlen. Függetlenül attól, hogy a várható eredmény „rutin” jellegű megoldás, vagy váratlan, szokatlan, esetleg ismeretlen eredetű események befolyásolják. Ez nemcsak a tudományban, hanem napi életünk során is így van. A különbség csak annyi, hogy a napi életben a problémafelvetést, kérdésfeltevést, hipotéziseket, és a megoldás eredményeképpen kapott téziseket általában nem szokás tételesen megfogalmazni. De lehetne, még egy egyszerű postai csekk esetében is.

#### **6.3.5.1. A hipotézisek formái**

A hipotézisek több formája megkülönböztethető a vonatkoztatási rendszer alapján, amelyeket különböző jelzős szerkezetekkel fejezünk ki. Így gyakran találkozhatunk a „munkahipotézis” kifejezéssel, ami tulajdonképpen előzetes feltételezést jelent, amely szerint a vizsgálatainkat elkezdtük. Ezek pontosítása a későbbiekben többnyire feltétlenül szükséges. Néha találkozhatunk az „alternatív hipotézis” kifejezéssel is, ami inkább elméleti, logikai jelentőséggel bír. Hipotézis és alternatív hipotézis ugyanis egymás ellentettje, egymás kiegészítője és egyúttal egymás kizárója. Az alternatív hipotézis az „eredeti” hipotézissel szemben támasztható állítások összességét magába foglalja (elvileg).

### 6.3.5.2. A kiemelt forma: nullhipotézis

A vizsgálataink során felállított hipotézisek többnyire alkalmatlanok konkrét statisztikai vizsgálatokhoz, ezeket „le kell fordítani” a statisztika nyelvezetére. Egy olyan formulát kell találni, amely általános, minden esetben alkalmazható és értelmezhető, számszerű, és mindig ugyanazt az értéket feltételezi. A feltételeknek egyetlen megoldás felel meg, ha a várható eredményt nullának feltételezzük. A statisztikában ezért kitüntetett szerepe van a **nullhipotézisnek**, azaz a várható változás, különbség, összefüggés egyenlő nullával. Ezt egy lehetséges változatként nem kell indokolni, szemben az alternatív hipotézissel, ahol a mérték és a nagyságrend számtalan, részben bizonytalan tényező függvénye lehet. A nullhipotézis tehát azt feltételezi, hogy nincs különbség, nincs változás, nincs összefüggés a vizsgált változóknál. A statisztikai analíziseknél mindig közvetve, a nullhipotézis elvetésével vagy megtartásával valószínűsíthetők az eredmények.

Itt egy igen lényeges ponthoz érkeztünk. A statisztika ugyanis soha semmit nem bizonyít, vagy nem vet el. A statisztika csak valószínűsít, valószínűségi alapon becslő értékeket. A statisztikai analízisek mindig becslések, amelyek hibahatárokkal rendelkeznek (a hibahatárokat bizonyos esetekben konfidencia – megbízhatósági – intervallumoknak nevezzük.) A nullhipotézis elvetése vagy megtartása is valószínűségi alapokon álló becslésnek tekinthető. Azt kell eldönteni, hogy valószínűségi alapon különböznek-e **jelentősen (szignifikánsan)** az eredmények nullától?

#### **6.3.5.2.1. Szignifikancia szintek**

Az angol „significant” kifejezésből eredően használjuk a tudományos életben a „szignifikáns” jelzőt. Tulajdonképpen egy konvenció és némi számmisztika eredménye. E konvenció szerint 95%-os, 99%-os és 99,9%-os valószínűségi szinten tekinthetők az eredmények jelentősnek. E szintek meglétét ellenőrizni, „próbázni” kell. A statisztikában a „probabilitás” jele a „p”, értéke 0 és 1, illetve 0% és 100% között változhat. A „p” ténylegesen valószínűségi szintet jelent a statisztikában. Az előzőekben jeleztem, hogy az analíziseknél a nullhipotézist vizsgáljuk, és ennek bekövetkezési valószínűségének megállapítása az analízis végső célja. Igazából azonban közvetve, a nullhipotézis elvetésével állapítható meg az eredmények szignifikanciája. A jelentőség kimondásához tehát a nullhipotézis fennállásának (fenntartásának) valószínűsége 5% alatt kell legyen, amit „maradék valószínűségnek” is szokás nevezni. A szokásos jelölések:

- $p > 0,05$  nem szignifikáns (n.sz.)
- $p < 0,05$  szignifikáns (sz.)
- $p < 0,01$  erősen szignifikáns (e.sz.)
- $p < 0,001$  igen erősen szignifikáns (i.e.sz.)

Gyakorlatilag ez azt jelenti, hogy ha a mintavételünk a populációból megfelelő volt, akkor  $p < 0,05$  szinten 100 esetből 95 esetben hasonló, 5 esetben eltérő eredményt várhatunk.

### **6.3.5.2.2. Statisztikai hibák**

A statisztikai analízisek mindig a nullhipotézis fennállására vonatkozó becslések, ezért magukban hordozzák a tévedés, a hibázás lehetőségét. Két fajta hiba különböztethető meg. Az első fajta hibánál elvetjük a nullhipotézist, pedig az igaz. Ez a téves elutasítás hibája. (Azaz szignifikánsnak tekintjük az eredményt, pedig nem az. Kifogtunk egyet az „ellentétes” eredmények közül.) Az első fajtu hiba ellen a szignifikancia szint emelésével lehet védekezni, és a szignifikancia szint egyúttal jelzi az első fajtu hiba bekövetkezésének valószínűségét. A második fajtu hiba az elsőnek az ellentéte: megtartjuk a nullhipotézist, pedig az hamis, téves. Ez a téves elfogadás hibája. (Azaz tévesen nem szignifikánsnak minősítjük az eredményt.) A második fajtu hiba ellen az elemszám növelése nyújthat védelmet, bekövetkezésének valószínűsége azonban nem határozható meg.<sup>9</sup> (Bár a szükséges minta megválasztásához kifejlesztett legújabb speciális statisztikai modulok, amelyeket elsősorban szociológiai, közvélemény és piackutatásokat céloznak, már tartalmaznak becslést a második fajtu hibára is.)

---

<sup>9</sup> Az elemszám korlátlan növelése ezzel egyúttal nem indokolt. Az analíziseknél kapott statisztikai eredményekre vonatkozó szignifikancia határok elemszám – pontosabban szabadságfok – függőek, magasabb elemszámoknál alacsonyabbak az ugyanazon szignifikanciához tartozó határértékek. A többváltozós analízisek pedig tényleges különbségek/összefüggések esetén különösen hatékonyan jelzik a szignifikáns eredményeket. Az célszerű, hogy az elemszám lehetőleg haladjon meg a 30-at vizsgálati csoportonként. Többváltozós esetekben pedig az elemszám jóval haladjon meg a változók számát. Az azonban a lényegét tekintve többnyire mindegy, hogy néhány száz vagy többtízezres elemszámú a mintánk. (Arra persze ügyelni kell, hogy a teljes vizsgálati mintát esetleg „almintákra” bontva maradjon elégséges elemszám a kialakított csoportokban.)

## 6.4. *Leíró statisztikák*

A leíró statisztikák (decriptives, basic statistics) a minta egyik változójának alapvető jellemzőit adják meg. Ahogy a nevében is benne van, leírják a mintát, a minta jellemzőit foglalják magukba. Szokás **alapstatisztikának** is nevezni. A kapott értékek a további elemzések, statisztikai próbák során felhasználásra kerülnek, kiindulási pontot jelentenek. Bár az alapstatisztikák nem tartalmaznak hipotézisvizsgálati elemeket, az eredmények mégis valamilyen hibával rendelkező becslésnek tekinthetők. A minta jellemzésével ugyanis a populáció egészére kívánunk következtetéseket levonni.

**A mintát alapvetően elemszáma, középértékei, és adatainak változékonysága jellemzi.** A vizsgált esetek/egyedek (cases) számát **elemszámnak** nevezzük, jelölése: N, n (number).

Egy konkrét vagy csoportba/osztályba sorolt adat előfordulási számát **gyakoriságnak** hívjuk. A gyakoriságok összessége azonos az elemszámmal.

### 6.4.1. Középértékek

A változékonny adatok egy számmal jellemzését a **középértékek** adják meg. Középértékek: medián, módusz, átlag (median, modus, mean). Közülük legfontosabb az átlag, de a másik két középérték is lényeges információkat hordoz. A különféle középértékek az egyes adatok elhelyezkedése, az adatok eloszlása alapján egymástól kissé eltérhetnek. Egyetlen esetben azonosak számszerűleg, ha az adatok a későbbiekben tárgyalandó ún. normális eloszlást pontosan követik.

A **medián** a nagyság szerint rendezett adatok közül a középső, „50%-os” érték, amelynél az ennél kisebb és nagyobb adatok száma azonos.

A **módusz** a leggyakrabban előforduló érték. A másik két középértékkel szemben a móduszból több is lehet, mert több érték is előfordulhat azonos gyakorisággal.

Az **átlag** vagy számtani közép az adatok összegének és elemszámának hányadosa.

Jelölése:  $\bar{X}$ ,  $\bar{x}$ , vagy  $M$ . Tetszőleges pontossággal megadható, de maximum a mérési pontosságot 1 helyi értékkel (1 tizedessel) meghaladó adattal szokás megadni.

Névleges és rendező skálán elhelyezkedő adatokra átlagot (és szórást) nem szabad számolni.

## 6.4.2. Az adatok változékonyságának mutatói

### 6.4.2.1. Szóródás, terjedelem, szélsőértékek

A középértékek önmagukban nem jellemzik kielégítően a mintát, ehhez ismerni kell az adatok tömörülését, az adatok változékonyságát mutató mérőszámokat is. Az adatok átlag körüli elhelyezkedése és tömörülése, szétszórtsága, azaz **szóródása** több értékkel is jellemezhető. Ezek közül legfontosabb és a további analízisek során is felhasználható mérőszám a szórás.

Az adatok változékonyságának „legdurvább” jellemzője a **terjedelem**, ami a **szélsőértékek** (minimum-maximum) közötti különbséget jelenti. A szélsőértékek között az egyes adatok előfordulási gyakorisága adja az **eloszlást**, ami tovább részletezhető.



#### 6.4.2.2. Kvantilisek

A nagyság szerint sorba rendezett adatok egyenlő darabszámú részekre bontását a **kvantilisek** jelentik. Az adatok tetszőleges számú egyenlő részre oszthatók, a gyakorlatban azonban főleg két kvantilissel találkozhatunk. A **kvartilisek** négy azonos előfordulási számú részre bontják az adatokat. Az alsó és felső kvartilisek a nagyság szerint sorba rendezett adatok 25 és 75 százalékos határát jelentik (a „harmadik” – pontosabban második – kvartilis a medián, az 50 %-os érték). A további tetszőleges pontosságú részletezést a „százalékos” értékek, a **percentilisek** nyújtják. Jelölésük „P” mellett egy szám (azaz a fentiekben tárgyalt értékek percentilis megfelelői: P0, P25, P50, P75, P100).

(Lehetne említeni fentiek analógiájára a deciliseket is, de ezek és további „-ilisek” a statisztikában nem használatosak.)

### 6.4.2.3. **Átlagos eltérés, variancia, szórás**

Az adatok változékonyságának, átlag körüli elhelyezkedésének egy számmal való jellemzése az előzőek ellenére szükséges. Erre szolgálhatna az átlagos eltérés, az adatok középértéktől számított abszolút értékű eltéréseinek átlagolása (szumma abszolút differencia / N). Ez a mérőszám azonban a további statisztikai elemzésekhez nem használható. Az átlagtól való eltéréseket azonban valahogyan nyilvánvalóan figyelembe kell venni az adatok szétszórtságának jellemzésénél. Az eltérések különböző előjelűek lehetnek, ennek kiküszöbölése is szükséges. A legegyszerűbb megoldást a **négyzetes eltérések** figyelembe vétele nyújtja, ami számításba veszi az eltéréseket, és egyúttal kiküszöböli a negatív előjeleket. A négyzetes eltérések kvázi átlagolása adja a varianciát vagy szórásnégyzetet. A **variancia** az átlagtól való eltérések négyzeteinek összege osztva (n-1)-el. Jelölése:  $s^2$ , V.

A variancia négyzetgyöke a **szórás**. (A statisztikában négyzetgyökvonásnál mindig csak a pozitív előjelű értéket vesszük figyelembe.) A szórást másképpen standard eltérésnek is nevezzük (standard deviation), jelölése: s, SD.

**A szórás az adatok változékonyságának általánosan használt mérőszáma a statisztikában.** A szórás négyzete a variancia, ami az adatok „variálódását” jelzi, és a legtöbb statisztikai módszer alkalmazásánál szerephez jut. Hangsúlyozni kívánom azonban, hogy a szórás nem egészen „kvázi átlagos eltérés”, mert alapját a négyzetes eltérések képezik – és ezek összegét nem az elemszámmal, hanem az úgynevezett szabadságfokkal osztjuk, ami a szórás esetében (n-1).<sup>10</sup> A szórás további alapstatisztikai mérőszámok kiinduló pontját is jelenti. Ezek az átlag hibája és a variációs együttható.

---

<sup>10</sup> Ha viszont a négyzetes eltérést az elemszámmal osztjuk és a kapott értékből négyzetgyököt vonunk, akkor az átlagos eltérést kapjuk meg. Ezt azért kell hangsúlyoznom, mert a négyzetre emelés önmagában keményen súlyozza, felnagyítja, „bünteti”, kiemeli a nagyobb eltéréseket. A gyökvonás ezt a hatást kiküszöböli. Téves tehát az a néha hallható vagy olvasható vélekedés, hogy a szórás a négyzetre emelés miatt az átlagos eltéréshez képest felnagyítja az adatok eltérését az átlagtól. A szórás valóban nagyobb számszerűen, de ezt kizárólag az osztóban szereplő szabadságfok és elemszám különbsége okozza.

#### **6.4.2.3.1. Szabadságfok**

A szórás kapcsán említett szabadságfok jelölése:  $df$  (degrees of freedom). A **szabadságfok** az egymástól függetlenül választható elemek számát jelenti. Ha a statisztika számítása során (a képletben) az elemek között érvényesül egy, vagy több összefüggés, akkor az összefüggés(ek) számát levonva az elemszámból kapjuk a szabadságfokot. Másképpen mindezek azt jelentik, hogy az elemszámból levonjuk az adott statisztika kiszámításhoz szükséges, az adatokból már meghatározott paraméterek számát.

Az átlag esetében a szabadságfok az elemszámmal azonos, mivel az átlag kiszámításakor csak a minta adatait használjuk, a képletben nincs az adatokból előzetesen már kiszámított érték, statisztikai paraméter. A mintának csak az adatai szerepelnek a számlálóban, a nevezőben pedig az elemszám.

A varianciánál, illetve a szórásnál viszont a már kiszámított átlaghoz viszonyítunk, az átlagtól való (négyzetes) eltéréseket összegezzük a számlálóban. Az átlag miatt egy összefüggés, meghatározottság érvényesül, ami a képletben is szerepel, így ez esetben a nevezőben szereplő szabadságfok  $(n-1)$ . (Ténylegesen egyébként az összegről – szumma  $x$  – van szó. Egy meghatározott  $n$  elemű összegből  $(n-1)$ -et választhatunk szabadon, amelyekből az „utolsó” kiszámítható. A szabadságfokot ennek megfelelően „az egymástól függetlenül összeadandó elemek számának” is szokás nevezni.)

A később tárgyalandó korrelációs számításnál pedig két átlag is „adott” ( $x$  és  $y$  változóké) amihez viszonyítjuk az eredeti adatokat, ennek megfelelően a statisztika szabadságfoka  $(n-2)$ .

#### 6.4.2.4. **Átlag hibája és variációs együttható**

Az **átlag hibáját** (standard error) más néven standard hibának, vagy az átlag szórásának is nevezik, esetenként középhibaként említik. Miután számításainkból végső soron az egész populációra kívánunk következtetni, ezért az elemszámoktól függően jelentkezik egy állandó hiba. Elvi jelentése az, hogy a populáció tényleges átlaga körül hogyan szóródnak a populációból vett különböző minták átlagai, illetve mennyire „pontos” az eredményünk.

Értékét a szórás és az elemszám négyzetgyökének hányadosa adja ( $s/\sqrt{n}$ ). Jelölése:  $s_{\bar{x}}$ , SE.

A **variációs együttható** (coefficient of variation) vagy más néven **relatív szórás** az átlaghoz viszonyított százalékos formában mutatja az adatok változékonyságát. Segítségével különböző dimenziójú és nagyságrendű változók szórása összevethető egymással. Motoros próbák, próbarendszerek esetében különösen célszerű a kiszámítása.

Értékét a szórás és az átlag hányadosa adja ( $s/\bar{x}$ ). Jelölése:  $v$ ,  $s\%$ , CV.

### **6.4.3. Ábrázolás**

Az eredmények ábrázolásakor a diagramokon az átlagot és az átlag hibáját, vagy a szórást szokták feltüntetni. A statisztikai programokban ezt általában külön be lehet állítani, egyes grafikon típusok pedig mindkét értéket képesek megjeleníteni. Az értékeket  $\pm$  értelemben értelmezzük és általában így is ábrázoljuk „bajuszként” az átlag körül.

#### **6.4.4. További leíró statisztikai elemek**

A leíró statisztikákhoz az előzőekben leírtakon túlmenően még a gyakorisági eloszlások és a standardizált értékek tartoznak, amit a következő önálló fejezetben tárgyalunk.

#### 6.4.5. Összefüggések „leírása”

A leíró statisztikák szinonimájaként használt „alapstatisztika” fogalmába beleérthető még az összefüggések kimutatására szolgáló **korrelációszámítás** is. (A StatSoft Statistica programja is a „Basic Statistics” menü második sorában, a „Descriptive statistics” után tartalmazza.) A korreláció tulajdonképpen két változó közötti összefüggést „írja le” egy mérőszám formájában, azonban itt már hipotézisvizsgálati eljárásról van szó. A korreláció szignifikancia vizsgálatától ugyanis nem lehet eltekinteni, így az eljárás lényegét tekintve a statisztikai analízisek, a statisztikai próbák közé tartozik. Az is kétségtelen viszont, hogy a korreláció számítása mutat némi analógiát a szórás kiszámításával. Az eljárás részleteit a későbbiekben tárgyaljuk.

#### 6.4.6. Leíró statisztikai képletek áttekintése

A könnyen hozzáférhető statisztikai programcsomagok korában alkalmazói szempontból nincs szükség a statisztikai képletek ismeretére, senki nem fog nekiállni „kézzel” számolni. Ráadásul a többváltozós eljárásoknál általában a mátrixműveletek sem mellőzhetők, amelyek pontos matematikai leírása az „egyszerű” alkalmazók többségét minden bizonnyal elriasztaná. Az alapstatisztikák esetében **a könnyebb megértést** mégis elősegítheti a vonatkozó képletek áttekintése (v.ö. a szabadságfoknál leírtakkal). A fejezet befejezéseként ezért röviden összefoglalom az eddig leírtakra vonatkozó képleteket. A számítógépeknél szokásos formulákat alkalmazom, a képletek egy sorba kerülnek, a szumma jelnél nem használom az indexeket ( $i=1, \dots, n$ ) stb.



### 6.4.6.1. Kiindulási képletek és jelölések

Elemzés:	$n$
Összeg:	$\sum x$
Négyzetösszeg	$\sum x^2$

Négyzetes eltérés („Summa Quadrata“):  $SQ = \sum (x_i - \bar{x})^2 = \sum x^2 - \bar{x} * \sum x = \sum x^2 - (\sum x)^2/n$

Átlag:  $\bar{x} = (\sum x)/n$

Szórás:  $s = \sqrt{SQ/(n-1)}$

Átlag hibája (standard hiba):  $s_{\bar{x}} = s/\sqrt{n}$

Varíációs együttható:  $v = s/\bar{x}$ ,  $v\% = s/\bar{x} * 100$

## 6.4.6.2. Két változóra kiterjesztett képletek

### Két változó esetén az előzőek analógiájára:

Összeg:  $\Sigma x$ ,  $\Sigma y$

Szorzatösszeg:  $\Sigma xy$

Négyzetösszeg  $\Sigma x^2$ ,  $\Sigma y^2$

Négyzetes eltérés:  $SQ_x$ ,  $SQ_y$

Keresztszorzat („Summa Productum”, SP,  $SQ_{xy}$ ):  $SQ_{xy} = \Sigma(x_i - \bar{x})(y_i - \bar{y})$

$$SQ_{xy} = \Sigma xy - \bar{x} * \Sigma y$$

$$SQ_{xy} = \Sigma xy - \bar{y} * \Sigma x$$

$$SQ_{xy} = \Sigma xy - (\Sigma x * \Sigma y) / n$$

Korreláció:

$$r = SQ_{xy} / \sqrt{SQ_x * SQ_y}$$

### 6.4.7. Gyakorisági eloszlás, percentilisek

**Gyakoriságon** azt értjük, hogy az egyes adataink hányszor fordulnak elő a mintában. Nagy adatterjedelem esetén az adatokat egyenlő intervallumokba, **osztályokba** sorolhatjuk<sup>11</sup>. Osztályba sorolt adatoknál az egy osztályban előforduló adatok száma jelenti a gyakoriságot, amit osztálygyakoriságnak is hívhatunk. A statisztikai programok a gyakoriságokat („frequencies”) minden egyes előforduló adatra, vagy tetszőlegesen beállított számú osztályra egyaránt megadják<sup>12</sup>. A frekvencia táblázatokat minden adatfeldolgozás első lépéseként le kell hívni<sup>13</sup>, hogy adataink ellenőrzéseként a legdurvább adatrögzítési és elírási hibákat korrigálni tudjuk. (Ezek ugyanis a szélsőértékeknel, nagyságrendi tévedésként szoktak a leggyakrabban előfordulni.)

---

<sup>11</sup> Ha nem programmal készítettjük az osztályokba sorolást, akkor ügyelni kell az osztályhatárok megállapítására. A határokat úgy kell meghúzni, hogy egy adat ne tartozhasson két osztályba, azaz a szomszédos osztályok felső és alsó határa ne legyen azonos. Praktikusán adatainknál egy helyiértékkel nagyobb pontosságú határok eleve kiküszöbölik ezt a hibázási lehetőséget.

<sup>12</sup> Alapbeállításként az SPSS minden előforduló értékre, a StatSoft Statistica pedig 10 intervallumra adja meg a frekvencia táblázatokat. Az adatellenőrzéshez általában szükséges a minden egyes értékre vonatkozó gyakorisági táblázat – és kifejezetten zavaró lehet az osztályba sorolt adatok kezelése.

<sup>13</sup> Igazából „vizuális” áttekintésként elsőként a grafikonos megjelenítést, a hisztogramokat érdemes lehívni.

#### 6.4.7.1. Hisztogram

A korábbiakban már jeleztük, hogy az egyes adatok előfordulási **gyakorisága** (frequency) valamilyen **eloszlást** követ. A gyakorisági eloszlás grafikusán is ábrázolható, ennek oszlopdiagramját **hisztogramnak** nevezzük. A hisztogram vízszintes (x) tengelyén a mért értékek helyezkednek el, míg a függőleges (y) tengelyen az előfordulási gyakoriságok. A gyakoriságok összessége értelemszerűen azonos a minta elemszámával (N). Megadható a **relatív gyakoriság** is, ha a minta elemszámához viszonyított százalékos értékeket adjuk meg az y tengelyen. Ha adataink nem csak néhány, hanem sokféle számszerű értéket vesznek fel, akkor célszerű osztályba sorolt adatokat feltüntetni a hisztogramon. A statisztikai programoknál ilyen esetekben az osztályok kívánt száma tág határok között beállítható.

#### **6.4.7.2. A gyakorisági eloszlás alapvető jellemzői**

Az eloszlások lehetnek folytonosak és diszkréték. A folytonos eloszlású adatoknál elvileg tetszőleges pontossággal, csak a mérés technikai korlátoktól függően fordulhatnak elő az egyes adatok. Ilyenek például az időeredmények. A diszkrét eloszlású adatoknál viszont csak egész számok fordulnak elő, és a két szám közötti tartomány nem mérhető. Ilyenek például a hibaszámok, darabszámok, ahol tizedes értékek nem fordulnak elő.

Sokféle eloszlás létezik. Amennyiben minden adat egyforma gyakorisággal fordulna elő, akkor az adatok **egyenletes eloszlást** követnének. Ez az eset azonban nem igen szokott előfordulni. A sokféle eloszlás közül a statisztikában, illetve a biológiai és társadalomtudományokban kiemelt jelentősége van a **normális eloszlásnak**. A normális eloszlás a legtöbb statisztikai számításnak elvi előfeltétele. A normális eloszlás a folytonos eloszlások közé tartozik, grafikonját **Gauss-görbének** is szokás nevezni.

### 6.4.7.3. A normális eloszlás

A természeti jelenségek jelentős része gyakorisági megnyilvánulásaiban a Gauss-görbét, vagy más néven a normális eloszlást követi. A normális eloszlás jellemzője, hogy szimmetrikus, alakja harang alakú, csúcsa kerekített, és gyorsan lelapuló ágai elvileg a végtelenbe tartanak. A görbe szélessége és magassága sokféle lehet, elvileg végtelen sok normális eloszlású görbe létezhet. A görbe szélességének és magasságának jellemzője a **lapultság** (kurtosis), míg a görbe szimmetriájának jellemzője a **ferdeség** (skewness). Az adatok mindig jelentős mértékben tömörülnek a középvértékek körül, míg a szélső értékek felé egyre kisebb gyakoriságok fordulnak elő. A görbe negatív és pozitív irányban is a végtelen felé tart – a matematikai absztrakt populációra vonatkoztatva. A gyakorlatban a populációból vett minta természetesen „véges”, az adatok a szélső értékek között helyezkednek el. Ugyancsak a gyakorlatban az eloszlás kisebb-nagyobb mértékben eltérhet a „tökéletes” normális eloszlástól, ami a ferdeség és lapultság mutatóival jellemezhető. (A ferdeségnek és lapultságnak a standard hibája is kiszámítható, illetve vizsgálható, hogy a kapott gyakorisági görbe eltér-e szignifikánsan a normális görbétől.)

A normális görbének legfontosabb jellemzője, hogy adatok **68,26 %-a a középvértéktől  $\pm 1$  szórásnyi** távolságra helyezkedik el. Középtől  **$\pm 2$  szórásnyi** távolságra az adatok **95,44%-a**, míg  **$\pm 3$  szórásnyi** távolságra az adatok **99,74%-a** helyezkedik el. A 3 szórásnyi távolságokon túlmenő, „végtelenbe nyúló” széleken már csak az adatok 0,26%-a található, amelyek akár „extrém” értékeknek is tekinthetők. A „tökéletes” normális görbénél a középvértékek (átlag, módusz, medián) teljesen egybeesnek, számszerűen azonosak. Mindezek következtében a normális görbénél pontosan meghatározhatók az egyes százalékos értékek, percentilisek is, illetve megadhatók az egész szórásnyi értékek százalékos megfelelői.

#### **6.4.7.3.1. Transzformálás: standardizálás, standard normális eloszlás**

A korábbiakban jeleztem, hogy számtalan formájú normális eloszlás fordulhat elő. A különböző változók pedig mind számszerű nagyságukban, mind dimenziójukban rendkívül eltérőek lehetnek. Az összehasonlításokhoz tehát ezeket valamilyen formában egységesíteni, **standardizálni** kell. Ehhez az eredeti mért értékeket kell valamilyen egységes matematikai szisztéma szerint megváltoztatni, transzformálni, mégpedig az eloszlás megváltoztatása nélkül. A változtatás természetesen az eredetileg mért dimenziót is megváltoztatja. Mindezeket hogy lehet minden esetre általános érvényűen megoldani? Nagyon egyszerűen: úgy kell a változót transzformálni, hogy várható középpértéke nulla, szórása = 1 legyen, és mindezek mellé vesszítse el dimenzióját, változzon dimenzió nélkülivé. **Az átlag=0, szórás=1, dimenzió nélküli eloszlást standard normális eloszlásnak nevezzük.** Bármilyen minta, bármilyen változó egyszerűen standardizálható, és ennek a standard értéknek a jelölése „z” vagy esetleg „u”. Az angol nyelvterületen – lásd a statisztikai programcsomagokat – többnyire „Zscore” jelöléssel látják el, és a programok fel is kínálják a standard értékek rögzítését<sup>14</sup>, mentését. Kiszámítása nagyon egyszerű:

**Z = (x<sub>i</sub> - átlag) / szórás**, másképpen:

$$Z = (x_i - \bar{x}) / s$$

Azaz minden egyes mért értékből kivonjuk az átlagot, és ezt a különbséget osztjuk a szórással.

---

<sup>14</sup> A standard értékek használata annyira jelentős, hogy a mai statisztikai programcsomagok például a regresszió számításoknál az eredménytáblázatban elsőként a standard értékekre vonatkozó „béta” regressziós koefficienseket tüntetik fel. A standardizált regressziós egyenletek „konstansa” nulla, ezért nem is tüntetik fel. Az eredménytáblázatokban csak ezt követően szerepelnek az eredeti dimenziókra vonatkozó regressziós együtthatók és a konstans. Bővebben lásd a regresszió számítások tárgyalásánál.

#### **1.1.1.1.1.A standard normális eloszlás jellemzői**

A képzett standard értékek átlaga nulla, szórása pedig 1 lesz. Normális eloszlás esetén ezen értékek fele negatív előjelet vesz fel. A dimenzió pedig azért esik ki, mert a fizikában használt képletek analógiájára a z érték kiszámítására szolgáló képlet számlálójában és a nevezőjében is ugyanaz a dimenzió szerepel, ami az egyszerűsítés során kiesik, a „z érték” már dimenzió nélküli lesz. Az 5. ábra mutatja, hogy a korábbiakban leírtak szerint hogyan helyezkednek el az eddig tárgyalt, a normális eloszláshoz is kapcsolható értékek. Az ábrán feltüntettem továbbá két „nevezetes” értéket: 1,645 (95%) és 1,96 (2,5% és 97,5%). Ezek egyrészt a hipotézisvizsgálatoknál jutnak szerephez, és az egy- illetve kétoldali próbák szignifikancia vizsgálatánál jelentik a konvencionális 5%-os határt. (Azaz a nullhipotézis „maradék valószínűségét”.) Nem véletlen, hogy például a Student-féle t-eloszlás 95%-os kritikus értéke „végtelen” elemszámnál 1,96. Tehát az ez alatti t-értékek elemszámtól függetlenül biztosan nem szignifikánsak. Másrészt az előzőeken túlmenően a megbízhatósági, vagy más néven konfidencia intervallumok meghatározásánál is szerephez jutnak a jelzett „nevezetes” standard értékek. Ha például az átlag hibájának (SE) 1,96 szorosát  $\pm$  hozzáadjuk az átlaghoz, akkor a kapott intervallumba 95 %-os biztonsággal beleesik a populáció tényleges átlaga. (Ez az opció a programok több grafikonjánál beállítható.)



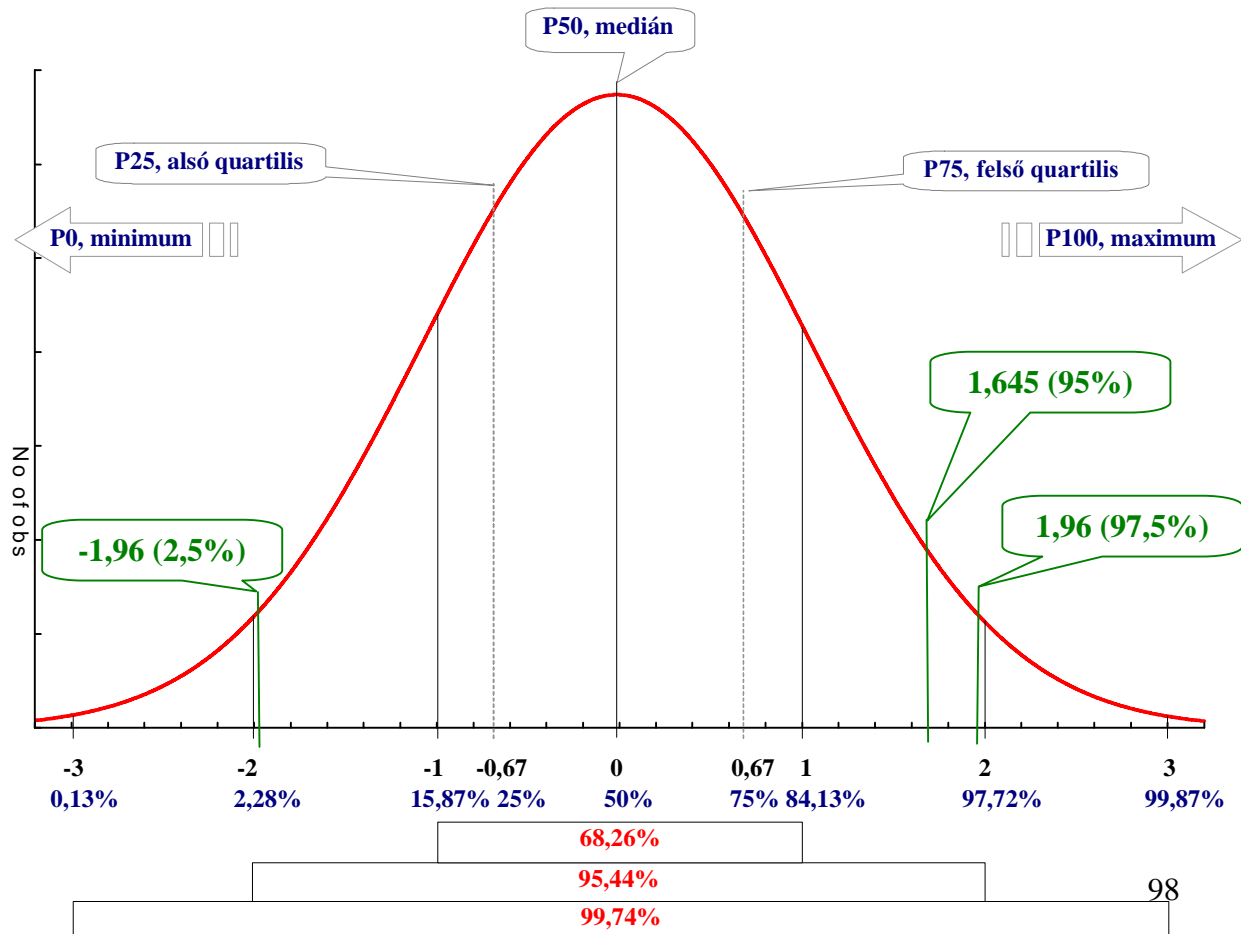
#### **6.4.7.3.2. További transzformációk**

A normális eloszlás a legtöbb statisztikai számításnak elvi előfeltétele. Elvileg a számítások előtt ellenőrizni kellene az adatok eloszlásának **normalitását** („normality”). A statisztikai programok erre természetesen lehetőséget nyújtanak, bár a különböző programok egymástól eltérő hangsúlyt fektetnek rá<sup>15</sup>. Szerencsére a statisztikai eljárások többsége túlzottan nem érzékeny a normalitási feltétel kisebb megsértésére. Szélsőségesen ferde eloszlások esetén azonban megfontolandó valamilyen **transzformáció** alkalmazása, amely megváltoztatja az eloszlást.

A transzformációk az eredeti adatok eloszlásának megváltoztatását jelentik valamilyen függvény, egyenlet szerint. Az előzőekben említett standardizálás is transzformáció, amely azonban az adatok eloszlását nem változtatja meg. A normalitási feltétel sérülése esetén ennek éppen az ellenkezőjére van szükség. Szakterületünkön különösen időeredményeknél, így a gyorsaságot mérő motoros teszteknel (pl. 60 m síkfutás) előfordulhat szélsőségesen „balra ferde” eloszlás. Ilyenkor a „reciprok transzformáció” ( $1/x$ ) segíthet az eloszlás normalizálásában. A további adatfeldolgozásnál ilyenkor a transzformált adatokat kell figyelembe venni. Ez azonban az értelmezést megnehezítheti, és különösen ügyelni kell az esetleges „visszatranszformálásra”.

---

<sup>15</sup> A StatSoft Statistica nagy hangsúlyt helyez a normalitás vizsgálatára, szinte „kikerülni” sem lehet a frekvencia táblázatok lekérésénél. Az SPSS esetében viszont kissé „eldugták” ezt a lehetőséget a leíró statisztikákon belül az „Explore: Plots” menübe.



**-3 SD      -2SD      -1 SD      átlag      +1 SD      +2 SD      +3 SD**

### **5. ábra: Standard normális eloszlás**

Kvantilisek: kvartilisek és percentilisek (kék), a hozzájuk tartozó standard z (vagy u) - értékekkel („Zscores”)

Az y tengelyen a gyakoriságok, az x tengelyen standardizált (fekete, átlag=0, szórás=1, dimenzió nélküli értékek), „százalékos” (kék, piros), illetve az „eredeti” (fekete, átlag, SD) adatok szerepelnek. Az ábrán feltüntetésre került a „nevezetes” 1,96 és 1,645 standard érték. Az ábra alapja (piros harang görbe) StatSoft Statistica-val készült.

## **6.5. Gyakorló és demonstrációs példák StatSoft és SPSS programcsomagokkal**

A következőkben bemutatott példák korábbi és folyamatban lévő vizsgálatok anyagából kerültek kiválogatásra. Arra törekedtem, hogy egy adatbázison minél több eljárást tudjak bemutatni. Ez a „központi” adatsor a Semmelweis Egyetem Testnevelési és Sporttudományi Kar (TF) tanári szak, nappali tagozat III. éves hallgatóinak Eurofit felmérése 2006 őszén (TFunisex2006\_gyak.sta ; \*.sav ; \*.xls). A felmérés eredményeinek publikálása a könyv kéziratának befejezése idején még csak éppen elkezdődött. Külön köszönöm kollégáimnak, Makszin Imrénnek, Oláh Zsoltnak és Woth Klárának, hogy hozzájárultak az adatok jelen prezentációs felhasználásához.

A gyakorló adatbázisok a Kiadó és a NYME ACSK honlapjairól szabadon letölthetők, amit hosszú időn át szeretnék elérhetővé tenni (<http://www.ak.nyme.hu/index.php?id=11067> ). A gyakorló adatbázisok személyi azonosításra alkalmas adatokat nem tartalmaznak. A gyakorlási és demonstrációs célból meghagyott ilyen jellegű részadatok véletlenszerűen össze lettek keverve, egymással nincsenek kapcsolatban.

A példánál alapvetően a StatSoft Statistica 9.0 verziójára támaszkodtam. Bemutatom azonban az SPSS megoldásait is (SPSS 17.0). Megjegyzem, hogy a programok előző verziói is lényegében azonos vagy nagyon hasonló műveleti ablakokat és eredménytáblázatokat produkálnak. Külön jelzem, ha valamelyik szoftver véleményem szerint jobban kezelhető, vagy egymástól eltérő megoldást nyújt. Megítélésem szerint a Statistica általában barátságosabb, jobban szerkeszthető, könnyebben kezelhető, különösen kezdők részére. De az SPSS is nagyon profi, és egyes megoldásaiban jobbnak tartom a Statisticánál. Meggyőződésem, hogy – különösen a doktori képzésben – rendkívül hasznos mindkét szoftver megoldásainak és lehetőségeinek ismerete.

A példák eredménytáblázatait több esetben, kisebb-nagyobb mértékben szerkesztenem kellett, hogy elférjenek a tankönyv oldalain. Ez főleg a tizedes értékek csökkentésében nyilvánul meg. Az is előfordul azonban néha, hogy a megértést segítő okokból töröltem bizonyos „lényegtelen” adatokat az eredménytáblázatokból, másokat meg kiemeltem. Ha tehát a gyakorló adatbázisokon elvégzik a számításokat, a fentiek következtében kissé részletesebb és kevésbé „színes” eredményeket kaphatnak.

### **6.5.1. A Statistica és az SPSS számítási indító ablakai**

A Statistica számításai a „Statistics” menüből, az SPSS számításai az „Analyse” menüből indíthatók. Az indítás után további ablakok nyílnak meg, ahol beállíthatók illetve kiválaszthatók a további műveleti paraméterek, a lekérendő statisztikák és egyéb opciók. Minden eljárásnál kezdetként a számításba bevonandó változókat kell kijelölni. Ugyancsak a kezdeti lépésekhez tartozik a számításokba bevonandó esetek, személyek kijelölése (Select cases), amit azonban később is bármikor megtehetünk, módosíthatunk. A szelekciós funkció használatára alapesetben nincs szükség, mert minden eset bevonásra kerül. Ha azonban valamilyen szelekciót egyszer már végeztünk és így mentettük el az adatbázist, az adatok következő megnyitásakor ez lesz az alaphelyzet. Tehát a szelekcióval „normál” esetben nem kell foglalkozni, de ha egyszer elkezdünk „babrálni” a szelekcióval, utána kifejezetten ügyelni kell rá.

A következő ábrákból látható, hogy a statisztikai programcsomagok milyen széles repertoárt kínálnak fel. Az is látható, hogy a két program egymástól eltérő logika alapján csoportosítja a számításokat, eljárásokat és analíziseket. Ettől a sokrétűségtől nem szabad megijedni, a gyakorlatban mindenkinél kialakul, hogy mely eljárásokat használ elsősorban. A továbbiakban csak a leginkább használatos eljárások kerülnek bemutatásra. A sok elvi lehetőségből a tényleges gyakorlatban többnyire csak néhányat használnak a legtöbben. Jelen keretek között nem cél a programok minden lehetőségének bemutatása, már csak azért sem, mert kifejezetten jó sűgővel rendelkeznek (igaz, csak angolul).

STATISTICA - Tfunisex2006\_gyak

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Resume... Ctrl+R Add to Report Add to MS Word

Arial 10 Basic Statistics/Tables

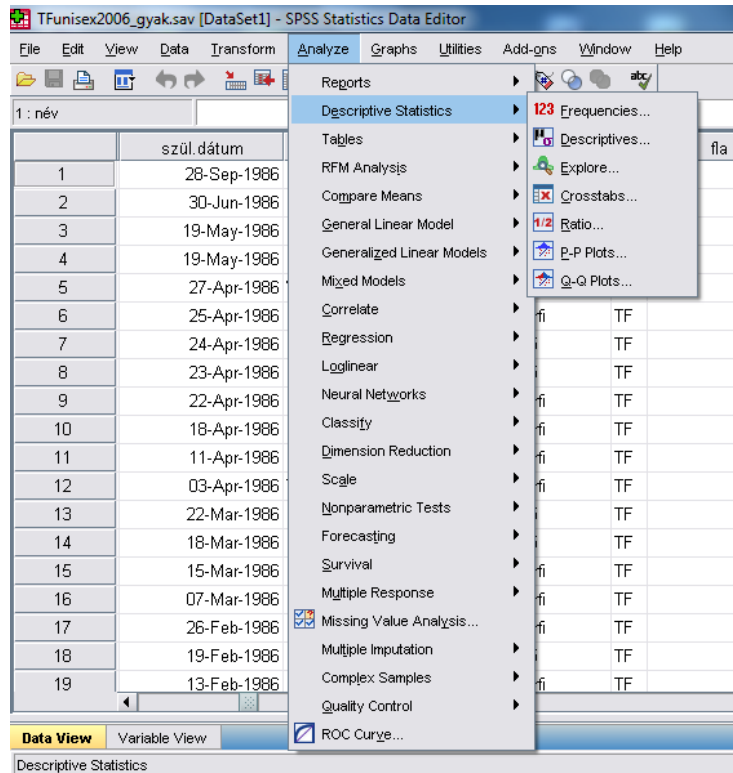
Data: Tfunisex2006\_gyak (45v by

- Multiple Regression
- ANOVA
- Nonparametrics
- Distribution Fitting
- Advanced Linear/Nonlinear Models
- Multivariate Exploratory Techniques
- Industrial Statistics & Six Sigma
- Power Analysis
- Automated Neural Networks
- PLS, PCA, Multivariate/Batch SPC
- Variance Estimation and Precision
- Statistics of Block Data
- STATISTICA Visual Basic
- Batch (ByGroup) Analysis
- Probability Calculator

10	11	12	13	14	15	16	17	18	19
AJL	FELÜL	HTU	FÜGG	SZORÉ	10x5m	20mINGA	TT	TM	BM
29	30	190	300	22	197	45	56	171	19
29	32	265	471	54	195	86			
16	29	232	450		168	108	97	190	26
33	29	242	450	50	193	102	72	183	21
32	26	285	462	43	184	109	76	185	22
25	44	225	700	57	187	9	70	173	23
35	26	180	80	20	189	60	56	162	21
38	26	210	170	24	188	53	62	170	21
25	38	270	480	58	196	53	83	184	24
22	30	252	500	55	190	101	77	185	22
35	35	260	780	56	190	122	71	184	20
32	38	230	450	48	190	53	85	186	24
32	27	213	330	28	202	56	62	167	22

Starts up Basic Statistics/Tables Tfunisex2006 C1.V23 39,7152388787924 Sel:OFF Weight:OFF CAP NUM REC

6. ábra: A StatSoft Statistica számítási műveleteinek indító ablaka



**7. ábra: Az SPSS számítási műveleteinek indító ablaka (17.0)**



### **6.5.2. Adatellenőrzés: frekvencia táblázatok lehívása**

Az adatellenőrzés a feldolgozás első lépése, gyakorlatilag az alapstatisztikákon, illetve a leíró statisztikákon belüli művelet. Hisztogramok és gyakorisági táblázatok segítségével ellenőrizendők az adatbázisban szereplő adatok. A lehívás módját a leíró statisztikákon belül mutatom be, most csak egy példát hozok fel. A korábbiakban említett TF-es Eurofit felmérésnél a lányok testmagassága a

2. táblázat szerint alakult. Az adatok 155-180 cm között szóródnak, reálisnak tűnnek, 1 fő adata hiányzik. A hasonló ellenőrzést az összes többi változóra elvégeztük, a lehetséges adatpótlásokat megejtettük. A további és lényegi adatfeldolgozásnak nincs akadálya.

**2. táblázat: Nők testmagasságának gyakorisági táblázata a mintapéldában**

Frequency table: <b>TM</b> (TFunisex2006_gyak)				
Include condition: nem=2				
Category	Count	Cumulative Count	Percent	Cumulative Percent
155	2	2	3,57	3,57
158	1	3	1,79	5,36
159	2	5	3,57	8,93
162	4	9	7,14	16,07
164	1	10	1,79	17,86
165	7	17	12,50	30,36
166	2	19	3,57	33,93
167	3	22	5,36	39,29
168	7	29	12,50	51,79
169	3	32	5,36	57,14
170	7	39	12,50	69,64
171	4	43	7,14	76,79
172	4	47	7,14	83,93
173	2	49	3,57	87,50
174	1	50	1,79	89,29
175	1	51	1,79	91,07
177	1	52	1,79	92,86
178	1	53	1,79	94,64
179	1	54	1,79	96,43
180	1	55	1,79	98,21
Missing	1	56	1,79	100,00

### **6.5.3. Leíró statisztikák számítása a statisztikai programokkal**

#### **6.5.3.1. Általános indító műveleti ablakok (StatSoft)**

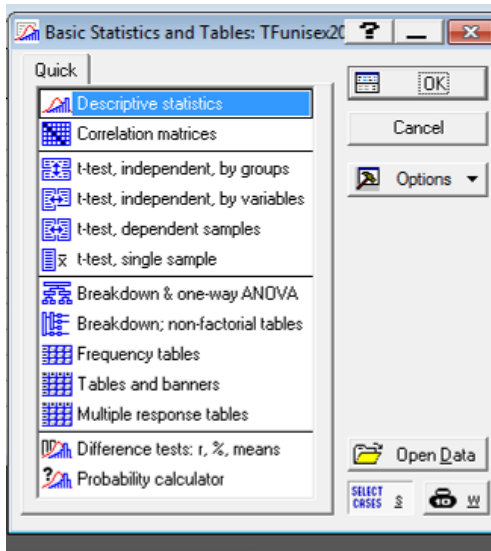
A leíró statisztikai mutatókra az eredmények közlésekor mindig szükség van valamilyen formában. Bár az értékek többnyire megjelennek vagy megjeleníthetők a statisztikai próbáknál is, számításainkat mindig célszerű az alapstatisztikák lekérdezésével indítani. A megoldási lehetőségek mindkét programcsomagnál szerteágazók – az eredmények természetesen azonosak. A számításokat mindkét programcsomagnál a változók kijelölésével kell kezdeni!

A 8. ábra és a 9. ábra a leíró statisztikák legáltalánosabb indító ablakait mutatja a beállításai lehetőségekkel a Statistica programcsomagban. A beállítástól függően a TF-es Eurofit felmérésnél a fiúk testtömeg, testmagasság és BMI alapstatisztikáinál a

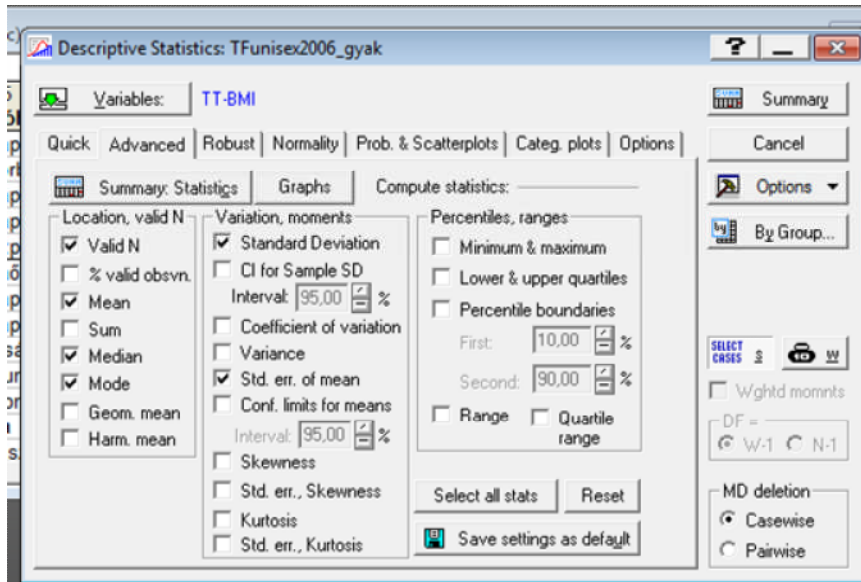
3. táblázat és a 4. táblázat szerinti értékeket kaptuk. Fontos beállítani az esetleg hiányzó adatok kezelésével kapcsolatos opciókat („Missing Data”, „MD deletion”) a 9. ábra jobb alsó sarka szerint. A „Pairwise” páronkénti elhagyást jelent, jelen esetben változóként különböző elemszámokat jelenthet. (Páronként összetartozó adatoknál, pl. korrelációnál, ha egy adatnak hiányzik a „párja”, akkor mindkettőt figyelmen kívül hagyja. Ez esetben is változó elemszámokat, illetve szabadságfokokat eredményezhet.) Jobb megoldás a „Casewise” opció, amely egyetlen hiányzó adat esetében is kizárja a további feldolgozásból az adott esetet (személyt). Ez a beállítás minden változónál azonos elemszámot eredményez. Természetesen, ha nincs hiányzó adatunk, akkor nincs jelentősége ennek a beállítási lehetőségnek.

#### **6.5.3.1.1. A Select Cases opció használata**

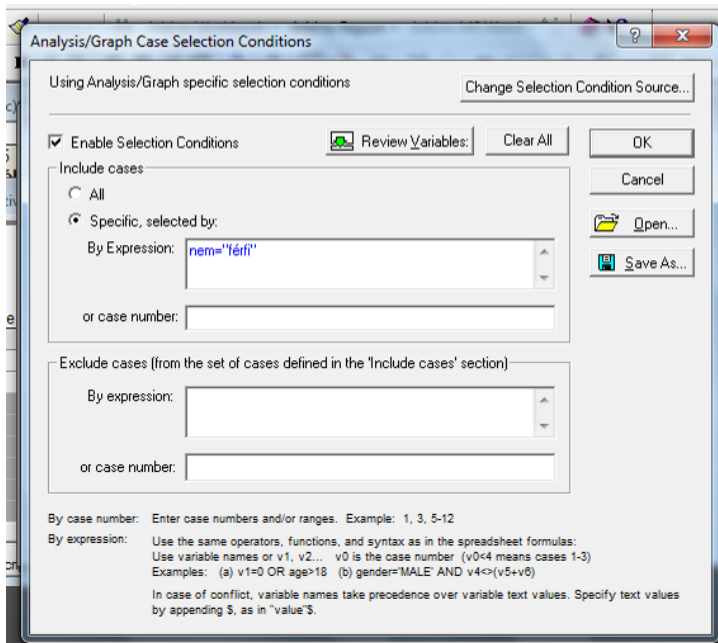
A „Select cases” opció (9. ábra, 10. ábra) használata szintén csak akkor szükséges, ha valamilyen ok miatt szűrni, szelektálni kell eseteinket. A „TFunisex2006” elnevezésű fájlok férfiak és nők adatait együtt tartalmazzák, ezért jelen esetben valamelyik nem adatait ki kell zárni a feldolgozásból. Ezt megtehetjük akár a „bevonás/Include”, akár a „kizárás/Exclude” ablakokban. A szelekciós változónak is megadhatjuk a számát vagy nevét (itt „v6” vagy „nem”), értékének pedig a vonatkozó kódszámot vagy idézőjelben a kódhoz kapcsolódó címke (Label) szövegét (itt „1” vagy „férfi”). Tehát jelen esetben a v6=1 vagy nem=„férfi”, illetve az ábra szerinti nem=1 azonos eredményre vezet.



**8. ábra: Az alapstatisztikák műveleti ablaka a StatSoft Statisticánál**



9. ábra: A leíró statisztikák kijelölési lehetőségei az „Advanced” ablakban (StatSoft)



10. ábra: A szelekciós feltételek beállíthatóságának ablaka (StatSoft)



**3. táblázat: Férfiak alapvető leíró statisztikai mutatói 3 paraméternél**

Descriptive Statistics (TFunisex2006_gyak) (Casewise Deletion of Missing Data) Include condition: nem="férfi"							
Variable	Valid N	Mean	Median	Mode	Frequency of Mode	Std.Dev.	Standard Error
<b>TT</b>	63	78,06	77	Multiple	5	7,94	1,00
<b>TM</b>	63	181,24	182	Multiple	6	6,28	0,79
<b>BMI</b>	63	23,76	23,46	Multiple	2	1,99	0,25

**4. táblázat: Férfiak további leíró statisztikai mutatói 3 paraméternél**

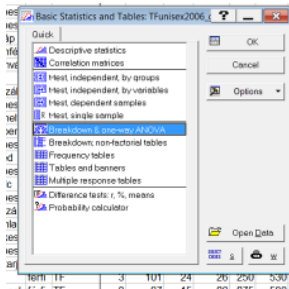
Descriptive Statistics (TFunisex2006_gyak) (Casewise Deletion of Missing Data) Include condition: nem="férfi"							
Variable	Valid N	Minimum	Maximum	Lower Quartile	Upper Quartile	Percentile 10,00000	Percentile 90,00000
<b>TT</b>	63	66,00	98,00	72,00	81,00	68,00	91,00
<b>TM</b>	63	165,00	197,00	176,00	185,00	173,00	190,00
<b>BMI</b>	63	19,93	29,34	22,28	24,90	21,50	26,73

#### **6.5.3.1.2. A „Statistics by Groups” funkció használata leíró statisztikáknál**

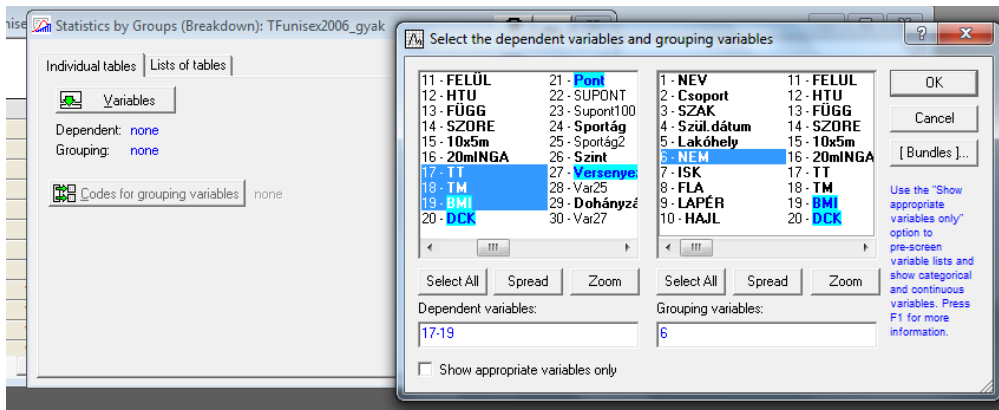
A leíró statisztikákat a „Basic statistics” menüpont „Breakdown/Statistics by Groups” ablakából is elérhetjük esetünkben (11. ábra- 14. ábra). Ebben az esetben nem szabad használni a „Select Cases” funkciót, mert a csoportosítási változónk („Grouping Var.”) a nem lesz! A 12. ábra szerint történhet a változók kijelölése, és a 13. ábra szerint kell megadni a csoportosítási változó értékeit a kódszám vagy a kód szerinti elnevezésekkel. Végül a 14. ábra szerint lehet kijelölni a lekérni kívánt leíró statisztikai értékeket. Eredményként az

5. táblázat: értékeit kapjuk, amelyben a korábbiakkal (

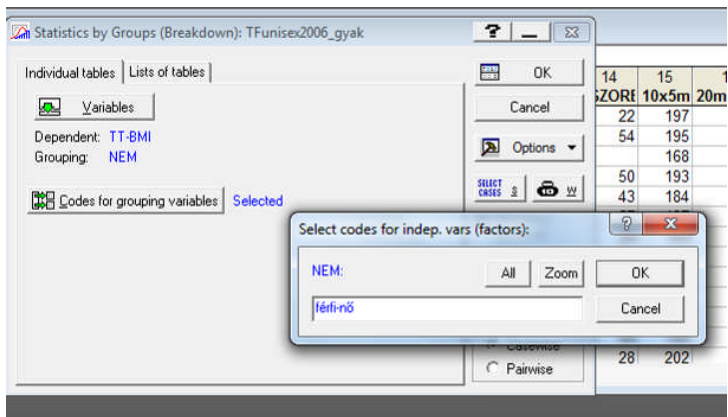
3. táblázat) azonos részeredmények szerepelnek, csak más az elrendezésük a nők adatainak szerepeltetése miatt.



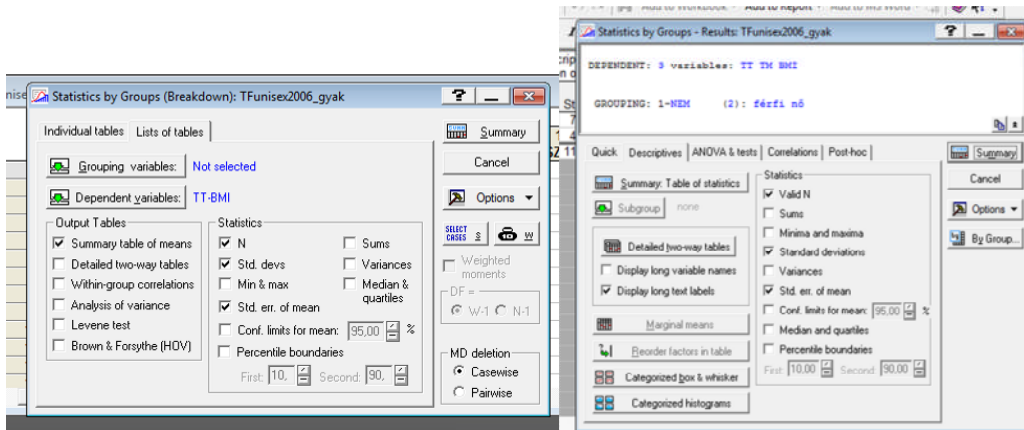
11. ábra: „Breakdown/Statistics by Groups” ablakból is lekérhetőek az alapstatisztikák



12. ábra: Változók kijelölése a „Breakdown/Statistics by Groups” ablakaiban



13. ábra: A csoportosítási változó értékeinek megadása



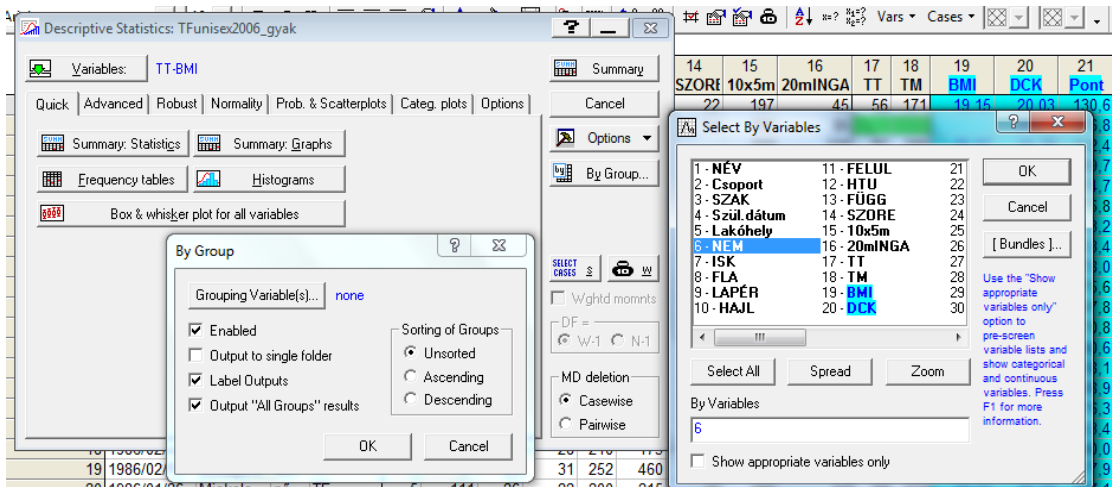
**14. ábra: A választható leíró statisztikák a csoportonkénti statisztikáknál (két ablakban is beállítható)**

**5. táblázat: Férfiak és nők átlagai és szórásai a választott 3 paraméternél**

Breakdown Table of Descriptive Statistics (TFunisex2006_gyak)									
N=118 (Casewise deletion of missing data)									
NEM	TT	TT	TT	TM	TM	TM	BMI	BMI	BMI
	Means	N	Std.Dev.	Means	N	Std.Dev.	Means	N	Std.Dev.
férfi	78,06	63	7,94	181,24	63	6,28	23,76	63	1,99
nő	59,87	55	4,71	168,00	55	5,35	21,21	55	1,31
All Grps	69,58	118	11,26	175,07	118	8,84	22,57	118	2,12

#### **6.5.3.1.3. A „By Group...” funkció közvetlen használata a műveleti ablakból**

A StatSoft Statistica a 7. verziótól bevezette a „By Group Analysis” menüpontot, ami a 8. verziónál már közvetlenül az egyes számítások műveleti ablakában található. Természetesen lekérhető az alapstatisztikák így is. Ez esetben a számítási feltételeket a 15. ábra: szerint adjuk meg és a 16. ábra: szerint kapjuk a csoportonkénti eredményeket. Az ábrán az „utolsó” csoport – beállítási lehetőség növekvő (Ascending), csökkenő (Descending) és semmi (Unsorted) – eredményei láthatók, a többi csoport, illetve „All Groups” eredményeit az eredményeket tartalmazó fájl (\*.stw) bal oldali sávjának megfelelő elemére kattintva lehet megjeleníteni.



15. ábra: A csoportokra vonatkozó statisztikák (By Group...) indító ablaka a Statisticában



STATISTICA - [Workbook1\* - Descriptive Statistics (TFunisex2006\_gyak)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Workbook Window Help

Arial 10 B I U

Workbook1\*

- Basic Statistics/Tables (TFunisex2006\_gyak)
  - Frequency tables dialog
    - Frequency table: TM (TFunisex2006\_gyak)
  - Descriptive statistics dialog
    - Descriptive Statistics (TFunisex2006\_gyak)
    - Descriptive Statistics (TFunisex2006\_gyak)
    - Descriptive Statistics (TFunisex2006\_gyak)
    - Descriptive Statistics (TFunisex2006\_gyak)
- Basic Statistics/Tables (TFunisex2006\_gyak)
  - Breakdown results dialog
    - Breakdown Table of Descriptive Statistics (TFunisex2006\_gyak)
- Basic Statistics/Tables (TFunisex2006\_gyak)
  - Descriptive statistics dialog
    - All Groups
      - Descriptive Statistics (TFunisex2006\_gyak)
    - NEM=nő
      - Descriptive Statistics (TFunisex2006\_gyak)
    - NEM=férfi
      - Descriptive Statistics (TFunisex2006\_gyak)

NEM=férfi  
Descriptive Statistics (TFunisex2006\_gyak)  
(Casewise Deletion of Missing Data)

Variable	Valid N	Mean	Std.Dev.
TT	63	78,0635	7,943090
TM	63	181,2381	6,275180
BMI	63	23,7566	1,985926

Descriptive Statistics (TFunisex2006\_gyak)

Ready TFunisex2006 C1.V1

16. ábra: Leíró statisztikák eredményei a csoportanalíziseknél

**6. táblázat: Leíró statisztikák eredmény táblázata a csoportanalíziseknél (nők)**

Variable	NEM=nő Descriptive Statistics (TFunisex2006_gyak) (Casewise Deletion of Missing Data)					
	Valid N	Mean	Minimum	Maximum	Std.Dev.	Standard Error
<b>TT</b>	55	59,87	47,00	72,00	4,71	0,64
<b>TM</b>	55	168,00	155,00	180,00	5,35	0,72
<b>BMI</b>	55	21,21	17,91	24,02	1,31	0,18

#### **6.5.3.1.4. Az eredménytáblázatok szerkesztése**

A leíró statisztikák befejezéseként felhívom a figyelmet egy lehetőségre, amit szintén a StatSoft egyik előnyének tartok. A Statisticában ugyanis szerkeszthetők, bővíthetők az eredménytáblázatok, amelyeket a programcsomag teljes értékű adattáblázatként kezel. Így például lekérhetők a relatív szórás eredményei. Igaz, ehhez ismerni kell a képletet ( $v=SD/\text{átlag}$ ), miután ezzel a statisztikával nem foglalkozik a program. A 17. ábra szerint bővíthető a megfelelő eredménytáblázat (7. táblázat).

STATISTICA - [Workbook1\* - Descriptive Statistics (TFUnisex2006\_gyak)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Workbook Window Help

Arial 10 B I U

Workbook1

- Basic Statistics/Tables (
  - Frequency tables di
  - Frequency table
  - Descriptive statistic:
    - Descriptive Stati
    - Descriptive Stati
    - Descriptive Stati
    - Descriptive Stati
- Basic Statistics/Tables (
  - Breakdown results c
  - Breakdown Tabl
- Basic Statistics/Tables (
  - Descriptive statistic:
    - All Groups
    - Descriptive S
    - NEM=nó
    - Descriptive S
    - NEM=férfi
    - Descriptive S

Variable	Valid N	Mean	Std.Dev.	Standard Error
TT	55	59,87	4,71	0,64
TM	55	168,00	5,35	0,72
BMI	55	21,21	1,31	0,18

Descriptive Statistics (TFUnisex2006\_gyak)  
(Casewise Deletion of Missing Data)

**Add Variables**

How many: 1

After: Standard

Name: Variációs együttható Type: Double

MD code: -99999998 Length: 8

Display format

General Number Date Time Scientific Currency Fraction Custom

Decimal places: 2

12.00%

Long name (label or formula with Functions):

=v3/v2

Formulas: use variable names or v1, v2, ..., v0 is case #.  
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; comment (after:)

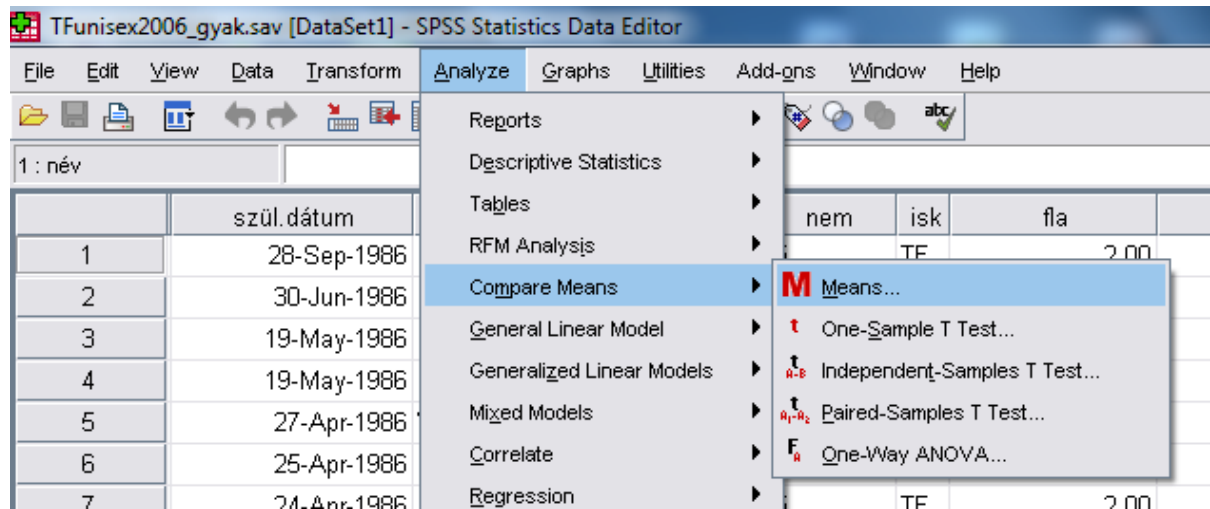
17. ábra: Eredménytáblázat bővítésének lehetősége a StatSoft Statisticában (variációs együttható)

**7. táblázat: Kibővített leíró statisztikai eredménytáblázat (Statistica, nők)**

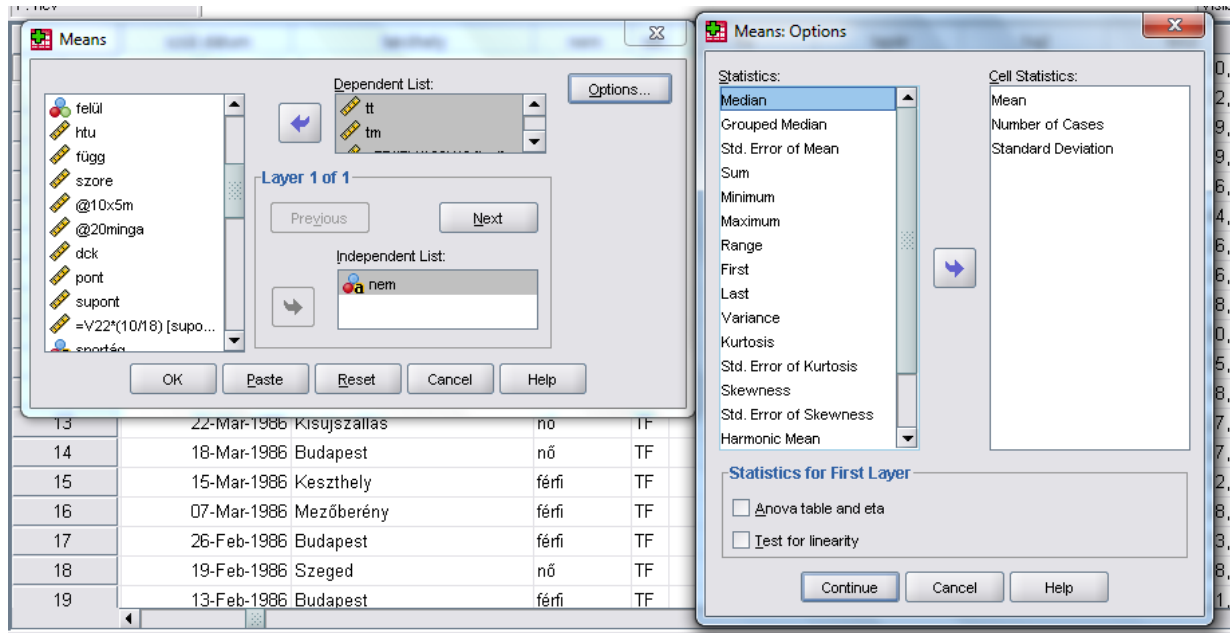
NEM=nő Descriptive Statistics (TFunisex2006_gyak) (Casewise Deletion of Missing Data)					
Variable	Valid N	Mean	Std.Dev.	Standard Error	Variációs együttható = $\sqrt{3}/\sqrt{2}$
<b>TT</b>	55	59,87	4,71	0,64	7,87%
<b>TM</b>	55	168,00	5,35	0,72	3,18%
<b>BMI</b>	55	21,21	1,31	0,18	6,19%

### **6.5.3.2. Indító menü és leíró statisztikák az SPSS-ben**

A következőkben nézzük meg az alapstatisztikák néhány lehívási lehetőségét az SPSS-ben. Itt is többféle úton kaphatjuk meg az eredményeket. Ha több csoportunk van érdemes a 18. ábra és 19. ábra szerinti megoldást választani a „Compare Means/Means” menüből. A felnyíló ablakok az SPSS jellegzetes logikáját követik, és természetesen itt is a változók kijelölésével kell kezdeni. Az opcióknál lehet kiválasztani a kért statisztikákat, és a 8. táblázat szerint kapjuk meg az eredményeket. Az SPSS eredménytáblázatai egyébként – szemben a StatSofttal – közvetlenül nem szerkeszthetők.



18. ábra: A minta jellemzőinek egyik legegyszerűbb lehívása az SPSS-ben



19. ábra: A változók és a kért leíró statisztikák kijelölése (SPSS)

**8. táblázat: Eredménytáblázat (SPSS)**

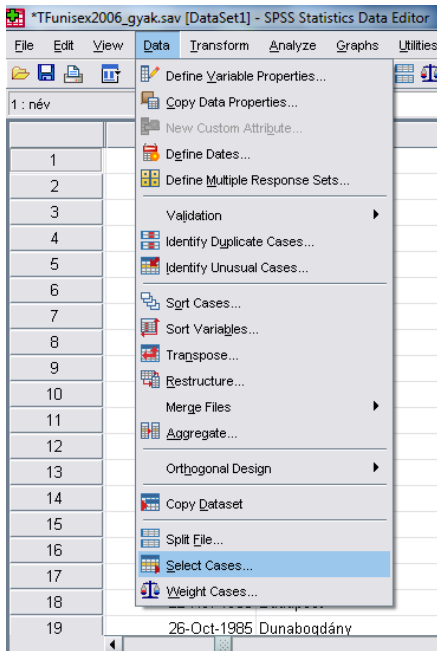
**Report**

nem		tt	tm	bmi
férfi	Mean	78,0635	181,2381	23,7566
	N	63	63	63
	Std. Deviation	7,94309	6,27518	1,98593
nő	Mean	59,9107	168,0000	21,2106
	N	56	55	55
	Std. Deviation	4,67596	5,35067	1,31336
Total	Mean	69,5210	175,0678	22,5699
	N	119	118	118
	Std. Deviation	11,23090	8,83537	2,12436

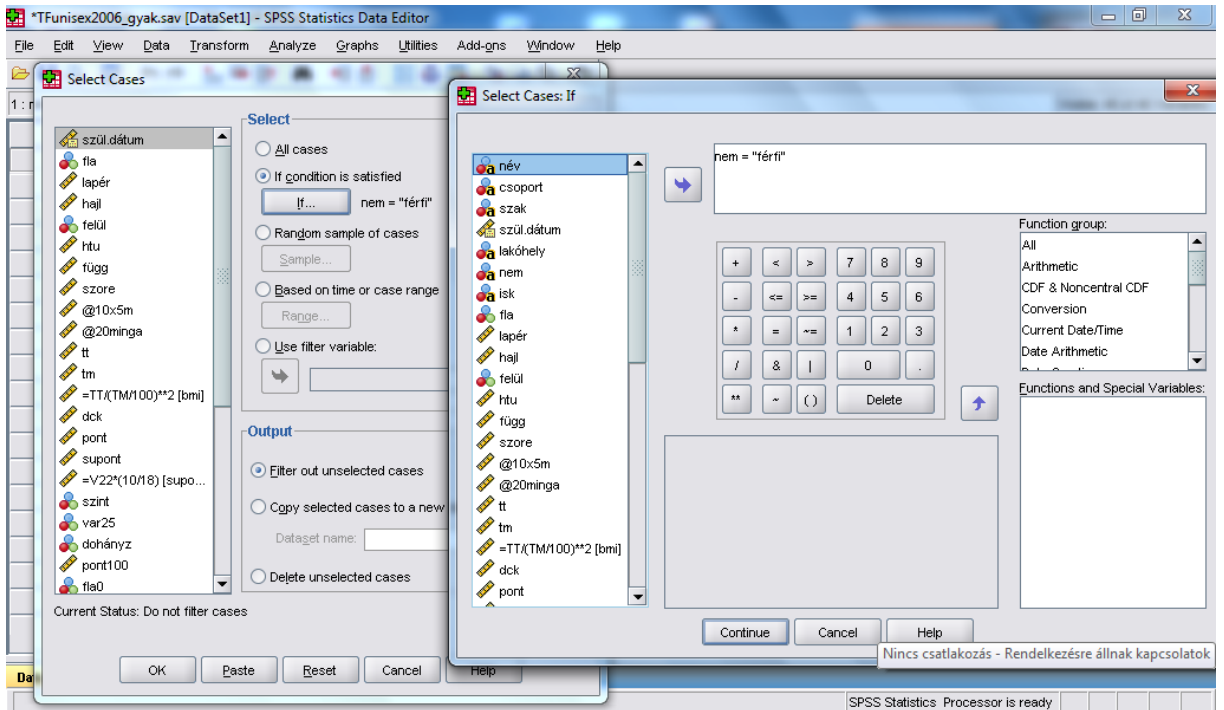
#### **6.5.3.2.1. Az SPSS „Descriptive Statistics” és esetválasztó menüje**

Az SPSS „Descriptive Statistics” menüje az előzőekhez hasonló megoldásokat kínál, csak leegyszerűsített formában és kevés opcióval. Tulajdonképpen ez képezi az alapfunkciót, és egyszerű adattáblázatoknál jól használható. Több csoportnál azonban használata előtt ki kell választani a számításba vonandó eseteket valamilyen kategóriaváltozó szerint. Esetünkben a két nem adatainak alapstatisztikáit külön kérjük le, és a nemenkénti kizárás/bevonás a „Data/Select Cases/If...” pontjai szerint kényelmesen beállítható (20. ábra és 21. ábra). Ezután indítható a leíró statisztikák ablaka (22. ábra és 23. ábra), és az opciók választása után megkapjuk az eredményt (9. táblázat). Sajnos a szelekciós feltételt nem tartalmazza az eredménytáblázat, ez csak a táblázat előtti „szintaxisból” derül ki – ha ennek kiírását előre beállítottuk a programcsomag egészének alapbeállításai között.





**20. ábra: A SPSS esetvlasztó funkciójának indítása**



21. ábra: Beállítási lehetőségek az SPSS Select Cases ablakaiban

\*TFunisex2006\_gyak.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : név

	szül. dátum	fla	lapér	hajl	felül	htu	
1	30-Jun-1986	1,00	116,00	29,00	32,00		
2	19-May-1986	12,00	105,00	16,00	29,00		
3	19-May-1986	7,00	103,00	33,00	29,00		
4	27-Apr-1986	6,00	100,00	32,00	26,00		
5	25-Apr-1986	5,00	91,00	25,00	44,00		
6	22-Apr-1986	fi	TF	2,00	101,00	25,00	38,00
7	18-Apr-1986	fi	TF	2,00	94,00	22,00	30,00
8	11-Apr-1986	fi	TF	6,00	110,00	35,00	35,00
9	03-Apr-1986	fi	TF	3,00	89,00	32,00	38,00
10	15-Mar-1986	fi	TF	0,00	88,00	40,00	32,00
11	07-Mar-1986	fi	TF	.	105,00	33,00	28,00
12	26-Feb-1986	fi	TF	6,00	101,00	14,00	33,00
13	13-Feb-1986	fi	TF	3,00	97,00	29,00	31,00
14	18-Jan-1986	fi	TF	2,00	92,00	43,00	34,00
15	31-Dec-1985	fi	TF	1,00	86,00	16,00	25,00
16	28-Nov-1985	fi	TF	3,00	101,00	24,00	26,00
17	22-Nov-1985	fi	TF	3,00	97,00	15,00	26,00
18	22-Nov-1985	fi	TF	0,00	99,00	33,00	36,00
19	26-Oct-1985	fi	TF	3,00	96,00	16,00	32,00

Visible: 46 of 46 Variables

Reports

- Descriptive Statistics
  - 129 Frequencies...
  - 1 Descriptives...
  - Explore...
  - Crosstabs...
  - Ratio...
  - P-P Plots...
  - Q-Q Plots...
- Tables
- RFM Analysis
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...

Data View Variable View

SPSS Statistics Processor is ready Filter On

Nincs csatlakozás - Rendelkezésre állnak kapcsolatok

22. ábra: Az SPSS „eredeti” leíró statisztikáinak indító menüje

The screenshot shows the SPSS Statistics Data Editor interface. The main window displays a list of variables and a data table. The Descriptives dialog box is open, showing the list of variables to be analyzed: tt, tm, and =TT(TM/100)\*\*2 [bmi]. The Descriptives: Options dialog box is also open, showing the settings for the descriptive statistics to be calculated.

**Descriptives Dialog Box:**

- Variable(s): tt, tm, =TT(TM/100)\*\*2 [bmi]
- Save standardized values as variables:

**Descriptives: Options Dialog Box:**

- Mean:  Mean,  Sum
- Dispersion:  Std. deviation,  Minimum,  Variance,  Maximum,  Range,  S.E. mean
- Distribution:  Kurtosis,  Skewness
- Display Order:  Variable list,  Alphabetic,  Ascending means,  Descending means

**Data Table:**

1: név	szül.dátum	fla	hajl	felül	htu	függ	szore	@10x5m
1								
2								
3								
4								
5								
6								
7								
8								
9								
10	15-Mar-1986	Keszthely	terfi	TF				
11	07-Mar-1986	Mezőberény	férfi	TF				
12	26-Feb-1986	Budapest	férfi	TF				
13	13-Feb-1986	Budapest	férfi	TF				
14	18-Jan-1986	Budapest	férfi	TF				
15	31-Dec-1985	Szekszárd	férfi	TF				
16	28-Nov-1985	Paks	férfi	TF				
17	22-Nov-1985	Budapest	férfi	TF				
18	22-Nov-1985	Budapest	férfi	TF				
19	26-Oct-1985	Dunabogdány	férfi	TF				

23. ábra: Beállítási lehetőségek az „eredeti” leíró statisztikáknál (SPSS)

**9. táblázat: Eredménytáblázat (SPSS)**

**Descriptive Statistics**

	N	Mean		Std. Deviation
	Statistic	Statistic	Std. Error	Statistic
tt	63	78,0635	1,00074	7,94309
tm	63	181,2381	,79060	6,27518
=TT/(TM/100)**2	63	23,7566	,25020	1,98593
Valid N (listwise)	63			

#### **6.5.3.2.2. Leíró statisztikák és hisztogram a „Frequencies” menüből (SPSS)**

A leíró statisztikai mutatók az SPSS-ben a „Frequencies” menüből is lekérhetőek. Könnyen kezelhető és jól áttekinthető ablakokban állíthatók be a lekérdezés feltételei. Amennyiben szükségünk lenne a percentilis értékekre, itt tetszés szerint beállíthatók – ebben a témában az SPSS jobban kezelhető, mint a StatSoft (24. ábra). Hasonló a helyzet a gyakorisági adatok diagramjaival, egyszerűen lekérhetőek a számunkra szükséges formában (25. ábra). A két ábra szerint beállított lekérdezések eredményeit a 10. és 11. táblázat, valamint a 26. ábra hisztogramja tartalmazza. Az összehasonlíthatóság kedvéért ugyanezt a hisztogramot a StatSoft Statisticával is elkészítettem (27. ábra). Ízlés kérdése, hogy kinek melyik tetszik jobban. Mindenesetre a StatSoft ábrája szerkeszthető bemásolás után még Wordben is, az SPSS esetén erre nincs lehetőség. Grafikában a StatSoft a jobb.

\*TFunisex2006\_gyak.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

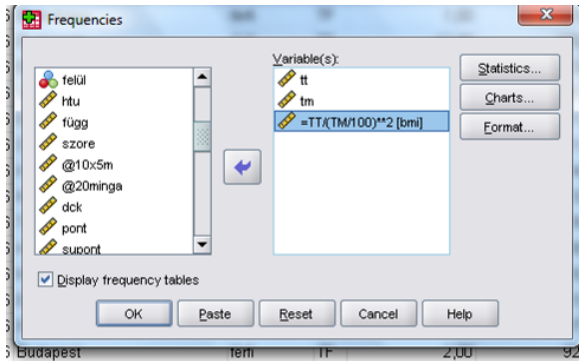
1 : név

	szül.dátum	fla
1	30-Jun-1986	
2	19-May-1986	
3	19-May-1986	
4	27-Apr-1986	
5	25-Apr-1986	
6	22-Apr-1986	
7	18-Apr-1986	
8	11-Apr-1986	
9	03-Apr-1986	
10	15-Mar-1986	
11	07-Mar-1986	
12	26-Feb-1986	
13	13-Feb-1986	
14	18-Jan-1986	
15	31-Dec-1985	
16	28-Nov-1985	
17	22-Nov-1985	
18	22-Nov-1985	
19	26-Oct-1985	

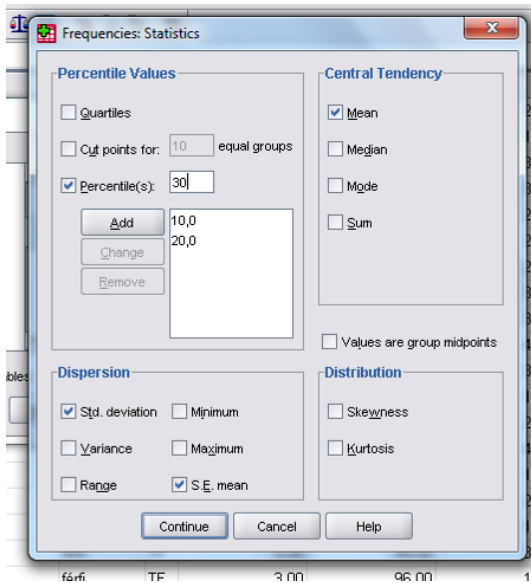
Reports  
 Descriptive Statistics **123** Frequencies...  
 Tables Descriptives...  
 RFM Analysis Explore... 1,0  
 Compare Means Crosstabs... 12,0  
 General Linear Model Ratio... 7,0  
 Generalized Linear Models P-P Plots... 6,0  
 Mixed Models Q-Q Plots... 5,0  
 Correlate fi TF 2,0  
 Regression fi TF 2,0  
 Loglinear fi TF 6,0  
 Neural Networks fi TF 3,0  
 Classify fi TF 0,0  
 Dimension Reduction fi TF  
 Scale fi TF 6,0  
 Nonparametric Tests fi TF 3,0  
 Forecasting fi TF 2,0  
 Survival fi TF 1,0  
 Multiple Response fi TF 3,0  
 Missing Value Analysis... fi TF 3,0  
 Multiple Imputation fi TF 0,0  
 Complex Samples fi TF 3,0  
 Quality Control  
 ROC Curve...

Data View Variable View

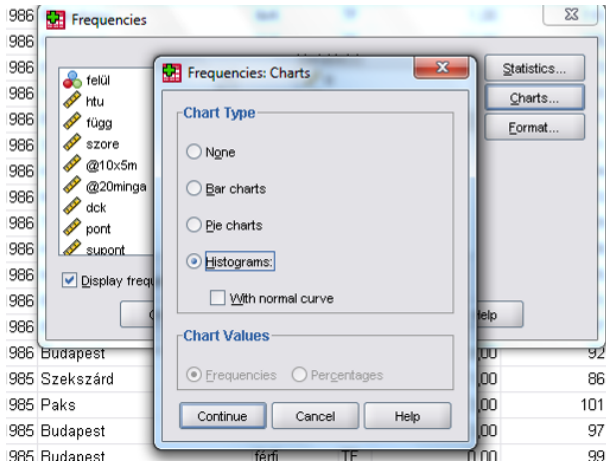
Frequencies...







**24. ábra: Percentilis értékek tetszőleges lekérési lehetősége a Frequencies menüben (SPSS)**



**25. ábra: Diagram lekérthetőség a Frequencies menüben (SPSS)**

**10. táblázat: SPSS eredménytáblázat a kiválasztott percentilisekkel (férfiak, testtömeg, testmagasság, BMI)**

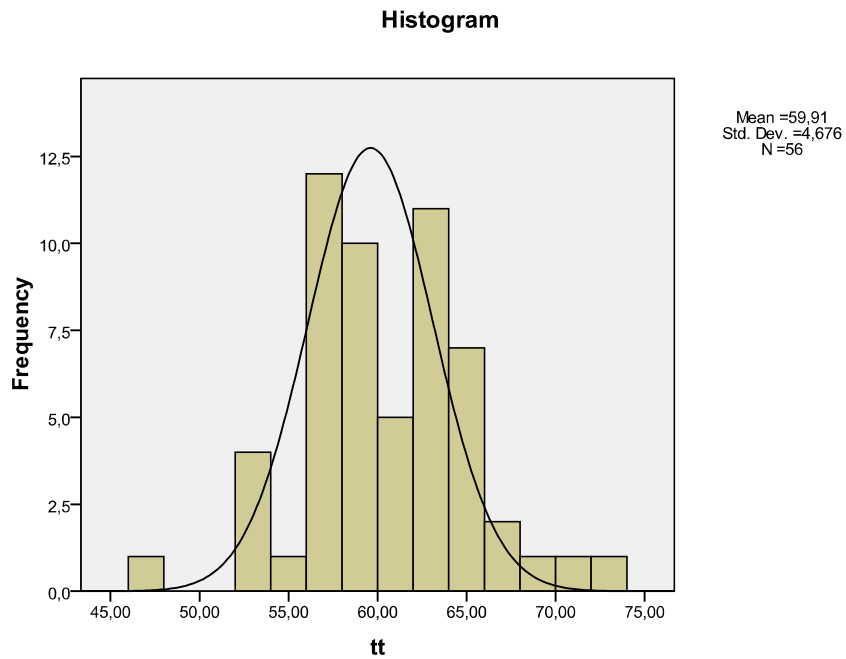
		<b>Statistics</b>		
		tt	tm	=TT/(TM/100)**2
N	Valid	63	63	63
	Missing	3	3	3
Mean		78,0635	181,2381	23,7566
Std. Error of Mean		1,00074	,79060	,25020
Std. Deviation		7,94309	6,27518	1,98593
Percentiles	10	67,4000	173,0000	21,4606
	20	71,8000	175,0000	21,9148

**11. táblázat: Férfiak testtömegének gyakorisági táblázata (SPSS)**

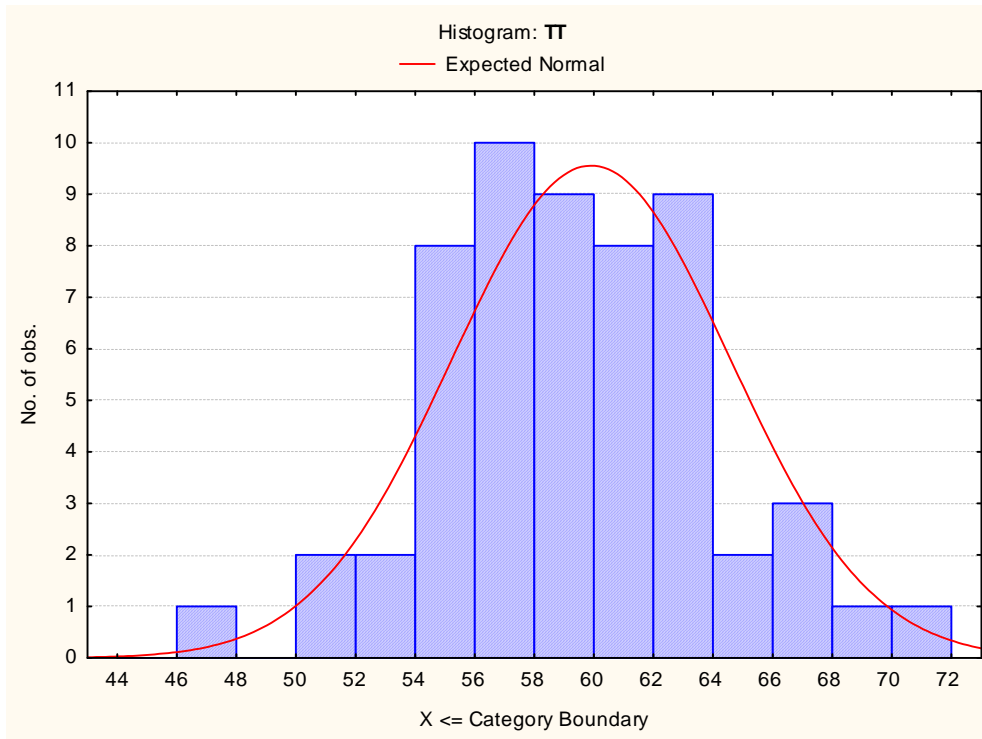
**tt**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	66,00	2	3,0	3,2	3,2
	67,00	4	6,1	6,3	9,5
	68,00	1	1,5	1,6	11,1
	70,00	2	3,0	3,2	14,3
	71,00	3	4,5	4,8	19,0
	72,00	4	6,1	6,3	25,4
	73,00	4	6,1	6,3	31,7
	74,00	4	6,1	6,3	38,1
	75,00	3	4,5	4,8	42,9
	76,00	3	4,5	4,8	47,6
	77,00	4	6,1	6,3	54,0
	78,00	2	3,0	3,2	57,1
	79,00	2	3,0	3,2	60,3

	80,00	5	7,6	7,9	68,3
	81,00	5	7,6	7,9	76,2
	82,00	1	1,5	1,6	77,8
	83,00	2	3,0	3,2	81,0
	85,00	2	3,0	3,2	84,1
	86,00	1	1,5	1,6	85,7
	88,00	1	1,5	1,6	87,3
	90,00	1	1,5	1,6	88,9
	91,00	2	3,0	3,2	92,1
	92,00	1	1,5	1,6	93,7
	94,00	1	1,5	1,6	95,2
	96,00	1	1,5	1,6	96,8
	97,00	1	1,5	1,6	98,4
	98,00	1	1,5	1,6	100,0
	Total	63	95,5	100,0	
Missing	System	3	4,5		
Total		66	100,0		



**26. ábra: A kiválasztott diagram, nők testtömegének histogramja (SPSS)**



**27. ábra: Nők testtömegének hisztogramja a Statisticaban**