

R Commander kézikönyv a “Biostatisztika” tankönyv példáival

© Harnos Andrea
harnos.andrea@aotk.szie.hu

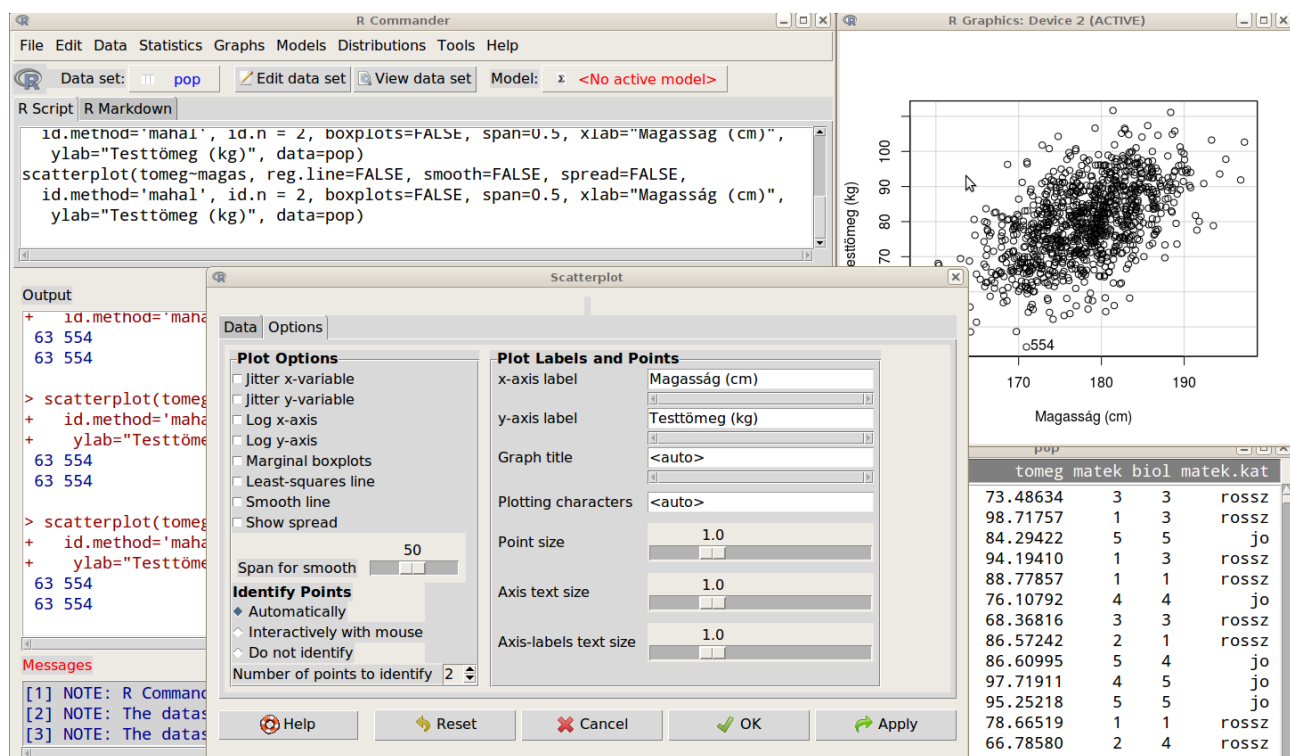
2014. március 17.

Tartalomjegyzék

1. Bevezetés	4
2. Az R Commmander installálása és futtatása	5
2.1. Linux	5
2.2. MS Windows	5
3. Az R Commander struktúrája	6
3.1. Menüsor	6
3.2. Eszköztár	7
3.3. Ablakok	7
4. A példákban használt adattáblázatok	8
4.1. pop: elseoves.txt	8
4.2. regr.kurz: regr.kurz.csv	8
4.3. ozmeret: ozmeret.csv	8
4.4. oz: oz.csv	8
4.5. lepke: lepke.txt	9
5. Adatbeolvasás	10
5.1. Adatbeolvasás szöveges állományból	10
5.2. Adatbeolvasás Excel állományból	12
5.3. Az adatok megtekintése, szerkesztése	12
6. Adattábla műveletek	12
6.1. Rész adattábla leválogatása	13
6.2. Aggregált táblázat készítése	13
6.3. Sorok eltávolítása	15
6.4. Több változó összefűzése egy változóba (Stack variables...)	15
6.5. Hiányzó értékeket tartalmazó esetek eltávolítása	15
6.6. Az aktív adattábla mentése	16
6.7. Az aktív adattábla exportálása szöveges táblázat formátumba	16

7. Műveletek változókkal	17
7.1. Változók átkódolása	17
7.2. Származtatott változók számítása	19
7.3. Numerikus változók faktorra alakítása	19
7.4. További műveletek változókkal	20
8. Leíró statisztikák	21
8.1. Aktív adattábla összegzése	21
8.2. Leíró statisztikák numerikus változókra	21
8.3. Gyakorisági eloszlások	22
8.4. Hiányzó adatok száma	23
8.5. Leíró statisztikák két faktor szerinti bontásban	24
8.6. Korrelációs mátrix	24
8.7. Korrelációs együttható tesztelése	25
8.8. Normalitás tesztelése	26
9. Kontingencia-táblák elemzése	26
9.1. Kétdimenziós kontingencia-táblák elemzése	26
9.2. Többdimenziós táblázatok	28
9.3. Kézzel begépelte kontingencia-tábla elemzése	29
10. Hipotézisvizsgálatok: átlagok elemzése	30
10.1. Egymintás t -próba	30
10.2. Két, független mintás t -próba	31
10.3. Két, párosított mintás t -próba	32
10.4. Egytényezős ANOVA	33
10.5. Többtényezős ANOVA	35
11. Arányokra vonatkozó próbák	36
11.1. Egy arány (valószínűség) vizsgálata	37
11.2. Két arány hasonlítása független minták esetén	38
12. Varianciák elemzése	40
12.1. Két, független mintás F -próba	40
12.2. Levene-próba	41
13. Nemparaméteres próbák	42
13.1. Két, független mintás Wilcoxon–Mann–Whitney próba	42
13.2. Két, párosított mintás Wilcoxon-próba	43
13.3. Több, független mintás Kruskal–Wallis-féle H -próba	45
14. Modellek illesztése	45
14.1. Regressziószámítás	45
14.2. A lineáris modell	47
15. Grafikus lehetőségek	48
15.1. Színpaletta beállítása	49
15.2. Index ábra	49
15.3. Hisztogram	50
15.4. Simított hisztogram	51
15.5. Stem and leaf ábra	52

15.6. Boxplot	53
15.7. QQ-ábra	53
15.8. Szórásdiagram	55
15.9. Szórásdiagram mátrix	56
15.10. Vonaldiagram	57
15.11. Feltételes szórásdiagram	57
15.12. Átlagok ábrája	59
15.13. Pontábra	60
15.14. Oszlop- és kördiagram	61
15.15. 3D szórásdiagram	62
15.16. Ábrák mentése	63
16. Modellekkel kapcsolatos műveletek	64
16.1. Az aktív modell kiválasztása	64
16.2. A modell összegzése	64
16.3. Modellből számított statisztikák	64
16.4. Információs kritériumok	65
16.5. Lépésenkénti modellszelkeció	65
16.6. Részmodell kiválasztása	66
16.7. Konfidencia-intervallumok	66
16.8. Hipotézis vizsgálatok	66
16.8.1. ANOVA-tábla	66
16.8.2. Modellek összehasonlítása	67
16.8.3. Lineáris hipotézisek tesztelése	67
16.9. Diagnosztikai eszközök	68
16.9.1. Numerikus diagnosztikák	68
16.9.2. Grafikus diagnosztikák	68
17. Eloszlások	69
17.1. Folytonos eloszlás: normális	69
17.1.1. Adott valószínűséghez tartozó kvantilisek meghatározása	69
17.1.2. Adott kvantilisekhez tartozó valószínűségek meghatározása	70
17.1.3. Sűrűség-, illetve eloszlásfüggvény ábrázolása	70
17.1.4. Mintavétel	71
17.2. Diszkrét eloszlás: binomiális	72
17.2.1. Adott valószínűségekhez tartozó kvantilisek meghatározása	72
17.2.2. Adott kvantilisekhez tartozó széli valószínűségek meghatározása	73
17.2.3. Adott kvantilisekhez tartozó valószínűségek meghatározása	73
17.2.4. Eloszlás ábrázolása	74
17.2.5. Mintavétel	75
18. Eszközök	76
19. A Biostatisztika tankönyv R Commander-rel megoldható példái	77
Irodalomjegyzék	79



1. ábra. Az R Commander

1. Bevezetés

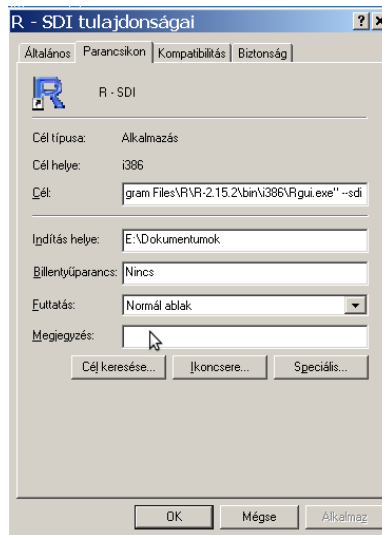
Ez az útmutató kiegészítésként készült a „Biostatistika nem statisztikusoknak”¹ [1] című tankönyvhöz, azzal a céllal, hogy a tankönyv példáit azok is meg tudják oldani, akik nem szeretnék R kódokat gépelni. Éppen ezért, a statisztikai példák magyarázatát, és az outputok értelmezését itt nem közöljük, csak azt, hogy a tankönyvben hol található meg a kedves olvasó. Az R Commander² egy GUI az R-hez, amellyel menük és párbeszédablakok segítségével generálhatunk R kódokat, amelyek újra futtathatók. Kezdő felhasználóknak, rövidebb statisztikai kurzusokon kiváló eszköz a statisztikai elemzésekhez. Az elkészült kódok tanulmányozása jó alap lehet az R tanulásához. Alapvetően egy R csomag (Rcmdr), amelyet az R indítása után kell betölteni. A csomag komolyabb statisztikai elemzésekre nem alkalmas, bár egyre több plug-in jelenik meg hozzá, szintén R csomaggént, különböző témákban (túléléselemzés, valószínűségszámítás, idősor analízis stb.). Az R Commandert John Fox (McMaster University)[2] fejlesztette bevezető statisztikai kurzusok hallgatóinak. Mi magunk is bevezető kurzusokon oktatjuk. A segédlet elkészítéséhez az Ubuntu 12.04-es operációs rendszerre installált R 3.02-es és az R Commander 2.0-0-ás verzióját használtuk. A windows-os verzió tartalmilag megegyezik a linuxossal, kinézetre a linux alapú kontrasztosabb és a így a képernyőképek jobb minőségűek, ezért ezek szerepelnek az útmutatóban.

Az R Commandernek – kevés kivételtől eltekintve – azokat a lehetőségeit ismertetjük, amelyek a tankönyvben is szerepelnek. Így kihagytunk néhány statisztikai tesztet (pl. Friedman-teszt), a modellillesztések közül az általánosított, ordinális és multinomiális modelleket, valamint a többváltozós elemzéseket (Dimensional analysis). Beletettük azonban néhány olyan adatmanipulációs lehetőség ismertetését, amelyek a tankönyvben nem szerepelnek, de nagyon hasznosak lehetnek.

Az egyes módszerek ismertetésénél megadjuk a tankönyv megfelelő fejezetének, példájának számát, illetve a kézikönyv utolsó fejezetében (19. fejezet) megadjuk, hogy a tankönyv pél-

¹<http://biostatkonyv.hu/>

²<http://rcommander.com/>



2. ábra. Az R parancsikon tulajdonságainak beállítása

dáihoz mely kézikönyvbeli fejezet tartozik. Egyes példák egy az egyben ugyanazok, mint a tankönyvben, de a legtöbb esetben csak a megfelelő módszert adjuk meg. Minden módszer ismertetése után megadjuk, hogy a tankönyv (*Tk.*) mely példáját lehet megoldani, illetve melyik ábráját lehet elkészíteni az adott módszerrel. Az oldalszámok a 3. kiadásnak felelnek meg, de ezek általában ugyanazok, mint a 2. kiadás oldalszámai. A fejezetek, példák, ábrák számozása ugyanaz a különböző kiadásokban.

2. Az R Commander installálása és futtatása

Az R Commander installálásához szükség van egy telepített R környezetre³ (Tk. 12.1. fejezet).

2.1. Linux

Az R-t elindítva, ki kell adni az `install.packages("Rcmdr")` parancsot élő internet kapcsolat esetén. Az R indítása után (például terminál ablakban) a `library(Rcmdr)` parancsot kell futtatni.

2.2. MS Windows

Windowsos környezetben, ahhoz, hogy jól működjön az R Commander, az R-t SDI (Single-Document Interface) módban kell futtatni. Ezt például a következőképpen lehet elérni:

- Készítsen egy parancsikont az R-hez az Asztalon!
- Az ikonra kattintva a jobb egérgombbal, válassza ki a Tulajdonságok menüpontot!
- Kattintson a Parancsikon fülre!
- A Cél mezőt írja át úgy, hogy a cím végéhez hozzáfűzi a `--sdi` szöveget! (Vigyázzon arra, hogy csak egy szóközt helyezzen el az első kötőjel előtt! (A Cél mező tartalma esetenként más és más lehet.)

³<http://cran.r-project.org/>

A parancsikont célszerű értelemszerűen átnevezni. Például: R-SDI-re. Ezt az Általános menüpontban teheti meg. Ezután kell installálni és betölteni az R Commander-t.

- Indítsa el az R-t!
- Installálja az Rcmdr csomagot az `install.packages("Rcmdr")` futtatásával, vagy az RGui menüjéből!
- Töltse be az R Commandert a `library(Rcmdr)` parancs futtatásával!

Az R Commander első indításakor felajánlja, hogy letölti és installálja az összes hiányzó, működéshez szükséges csomagot. Windows Vista vagy Windows 7-es operációs rendszer esetén, ha a Program Files könyvtárba szeretnénk telepíteni az R csomagokat, akkor az R-t rendszergazda (administrator) minőségben kell futtatni. (Jobb egér kattintás az R ikonján, majd a Futtatás rendszergazdaként kiválasztása). Ha az R-t nem a Program Files könyvtárba installáljuk, vagy rendszergazdai jogosultságokkal használjuk a Windows-t, akkor ezzel nincs probléma, egyébként pedig érdemes az R-t egy másik könyvtárba installálni, és akkor megint csak nem okoz problémát a telepítés.

Ha az R Commander-t automatikusan szeretnénk indítani az R futtatásakor, akkor az R etc könyvtárában lévő `Rprofile.site` fájlhoz írjuk hozzá a következőket:

```
local({
old <- getOption("defaultPackages")
options(defaultPackages = c(old, "Rcmdr"))
})
```

3. Az R Commander struktúrája

A betöltés után megjelenik az R Commander GUI (3. ábra). A felső sorban egy menüsor, alatta egy eszköztár sor különböző gombokkal, majd három ablak található.

3.1. Menüsor

File Munkakönyvtár (*Working directory*) váltás, szkriptek betöltése, mentése, R Markdown fájlok megnyitása, mentése, output fájlok megnyitása, mentése, R munkaterület betöltése, mentése, kilépés az R Commander-ből, illetve az R-ből.

Edit Szokásos szerkesztő funkciók (kivágás, másolás, beillesztés), kereső és kiválasztó funkciók. Visszavonás, ismétlés és ablak tartalmak törlése.

Data Adattáblák létrehozására, beolvasására és módosítására szolgáló funkciók.

Statistics Statisztikai elemzések.

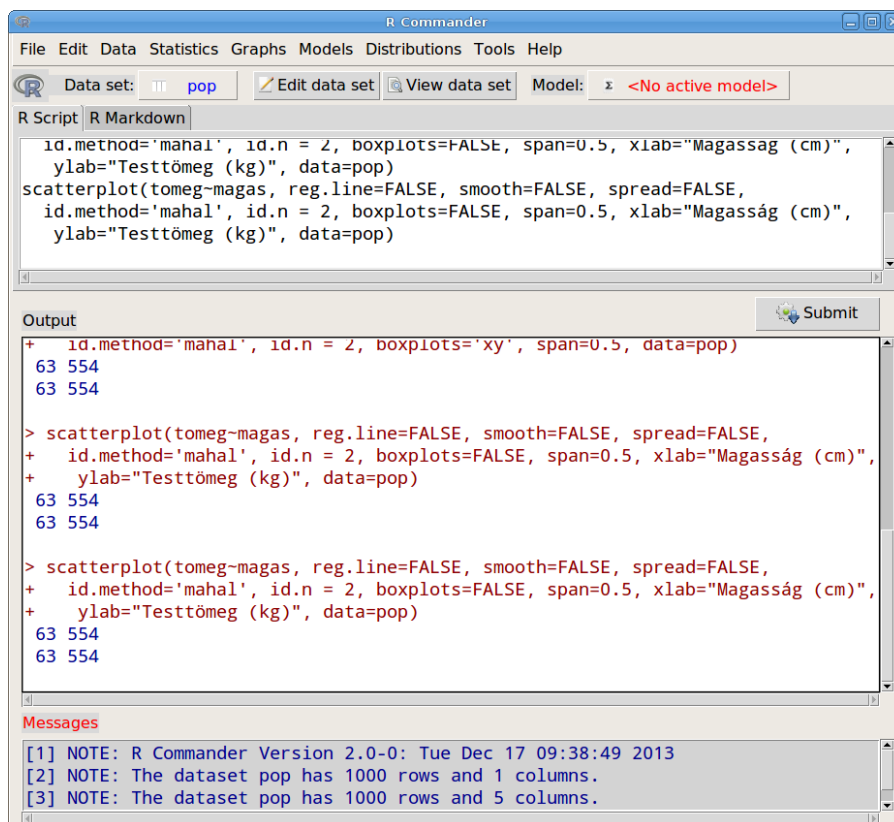
Graphs Ábrák készítése és mentése.

Models Statisztikai modellekkel kapcsolatos összesítések, tesztek, konfidencia intervallumok, modellszelekciós eszközök, numerikus és grafikus diagnosztikák.

Distributions Gyakran használt folytonos és diszkrét eloszlásokból valószínűségek és kvantilis számolása, mintavételezése, és ábrázolásuk.

Tools R csomagok és R Commander kiegészítések (plug-in) betöltése. R Commander beállítási lehetőségek.

Help R Commander-rel és R-rel kapcsolatos súgó.



3. ábra. Az R Commander ablakai

3.2. Eszköztár

Az eszköztár gombjai:

Data set A legbaloldalibb gomb mutatja az aktív adattábla nevét. Kezdetben nincs ilyen, később, ha már több adattáblánk is van betöltve, akkor erre a gombra kattintva választhatunk belőlük.

Edit data set Adatszerkesztő, melynek segítségével módosíthatjuk az aktív adattáblát.

View data set A gomb megnyomásával megnézhetjük, de nem szerkeszthetjük az aktív adattáblát.

Model A statisztikai modellezés során illesztett modellek közül választhatunk a gomb megnyomásával, illetve az utoljára illesztett modellt mutatja (aktív modell).

3.3. Ablakok

R Script Az R Commander által generált kódok jelennek meg benne. Magunk is gépellhetünk be ide kódokat, illetve módosíthatjuk a már elkészült kódokat. Innen futtathatjuk is a kódokat ha kijelöljük és megnyomjuk a **Submit** gombot.

Output A futtatások eredményeinek megjelenítése.

Messages Hibüzenetek (piros), figyelmeztetések (zöld), illetve egyéb információk (kék).

Az R Markdown fülre kattintva az R Markdown nyelvre fordítja a kódokat és az outputot, amiből egy kattintással létrehozható egy html dokumentum.

4. A példákban használt adattáblázatok

Ebben a fejezetben a kézikönyvben többször előforduló adatállományok leírását adjuk meg. Ezeket az adattáblázatokot használtuk a tankönyv példáiban is. Az adatállományok letölthetők a tankönyv⁴ honlapjáról. A statisztikai teszteknel egy-egy esetben használt adattáblázatok leírását az adott tesztnél ismertetjük.

4.1. pop: `elsoeves.txt`

Elsőéves férfi biológus hallgatók egy (elképzelt) 1000 fős populációja:

magas magasság cm-ben
tomeg testtömeg kg-ban
matek biomatematika vizsgajegy
biol biológia vizsgajegy
matek.kat jo: 4, 5, rossz: 1, 2, 3

4.2. regr.kurz: `regr.kurz.csv`

Egy regressziós kurzuson megkértük a hallgatóinkat – összesen 21-et –, hogy írják fel egy papírra a következő adataikat:

NEM a hallgató neme: `no`, `ferfi`
MAGASSAG magasság cm-ben
TOMEG testtömeg kg-ban
SZULHOSSZ születés kori hossz
SZULTOMEG születés kori tömeg
SZEMSZIN szemszín: `barna`, `kek`
CIPOMERET cipóméret

4.3. ozmeret: `ozmeret.csv`

Vadászok által kilőtt őzek testméret adatai. Az őzeket vadbiológusok mérték le (Kabai és mtsai, 2007). A felnőtt bakok adatait használjuk.

TOMEG teljes tömeg (kg)
ZSIGTOMEG zsigerelt tömeg (kg)
TESTH testhossz (cm)
MARMAG marmagasság (cm)
OVMERET övméret (cm)

4.4. oz: `oz.csv`

Vadászok által kilőtt őzek adatai.

TERULET kilövés helye: `Babat`, `Pitvaros`, `Arpadhalom`
SEX ivar: `suta`, `bak`
TOMEG teljes tömeg (kg)
TESTH testhossz (cm)

⁴<http://biostatkonyv.hu/>

4.5. lepke: lepke.txt

Egy kísérletben farkasalmalepkék (*Zerynthia polyxena*) hernyóinak fejlődését vizsgálták. (Kis J., Kassai F., Peregovits L., nem közölt adatok. A kísérlet leírását Kis János készítette el.) A hernyókat Csévharaszt közelében gyűjtötték 2004 júniusának első hetében. Egyetlen farkasalma növényről csak egyetlen 3. lárvastádiumban lévő hernyót gyűjtöttek, azért hogy azonos tojáscsomóból származó testvérek a kísérletben ne szerepeljenek. A hernyókat háromféle hőmérsékleti kezelésnek és kétféle táplálékellátottság-kezelésnek tették ki laboratóriumban az 1. táblázatban bemutatott elrendezés szerint.

1. táblázat. Kísérleti elrendezés

		Hőmérséklet (HOM)		
		Alacsony (hutott)	Közepes (szobahom)	Magas (melegített)
Táplálék (TAP)	Nem limitált (adlibitum)	AH	AS	AM
	Limitált (limitalt)	LH	LS	LM

A közepes hőmérséklet a szobahőmérséklet volt, ehhez képest hűtötték vagy melegítették időszakosan a többi csoport egyedeit. A hernyók fele mindig bőségesen kapott farkasalmalevelet, míg a másik csoporttól ezt időszakosan megvonták. A hernyókat véletlenszerűen választották az egyes kezelési csoportokba. A kezelés megkezdése előtt lemérték a hernyók tömegét.

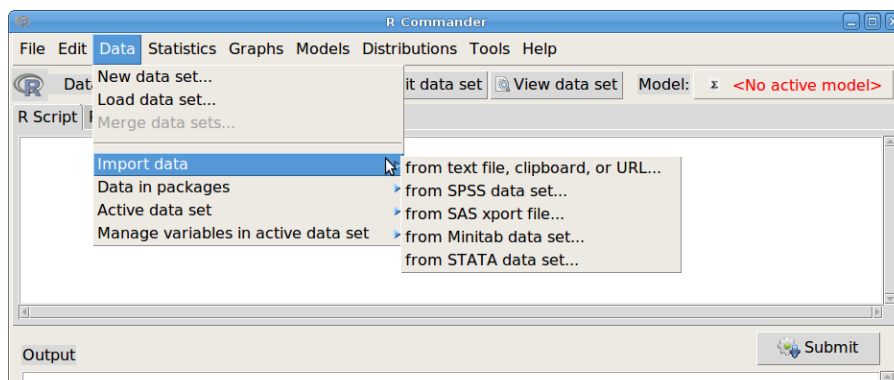
TAP táplálékellátottság kísérleti tényező

HOM hőmérséklet kísérleti tényező

BABTOMEG a bábok testtömege egy héttel a bebábozódás után (g)

TOMEGO a hernyók kezelés előtti tömege (g)

KEZELES az 1. táblázatnak megfelelő kezelés kombinációk



4. ábra. Adat importálás: Data → Import Data

5. Adatbeolvasás

Adatokat többféleképpen bevihetünk az R Commander-be.

- Begépelhetünk kis adattáblázatokat: Data → New data set...
- Beolvashatunk RData formátumú adattáblákat: Data → Load data set...
- Importálhatunk adatokat egyszerű szöveges állományokból, a vágólapról, vagy webcímről, illetve különböző statisztikai csomagokból (Minitab, SPSS, SAS vagy Stata), illetve Windows esetén Excel-ből, Access-ből, vagy dBase-ből: Data → Import data (4. ábra)
- Lehetőség van arra is, hogy R csomagokból olvassunk be adatokat: Data → Data in packages

A fenti lehetőségek közül a szöveges és Excel állományok beolvasását mutatjuk be.

5.1. Adatbeolvasás szöveges állományból

Példaként a tankönyv `elsoeves.txt` állományát töltjük be. Szöveges állományok esetén tudni kell a következőket az állományról (5. ábra).

Variable names in file Vannak-e változónevek az első sorban

Missing data indicator Hiányzó adatok jelölése

Location of Data file Az adattáblázat helye

Local file system Helyi fájl rendszer

Clipboard Vágólap

Internet URL Internet cím

Field Separator Mezőhatároló karakter

Spaces Egy vagy több szóköz, tabulátor vagy sorelemelés

Tabs Tabulátorok

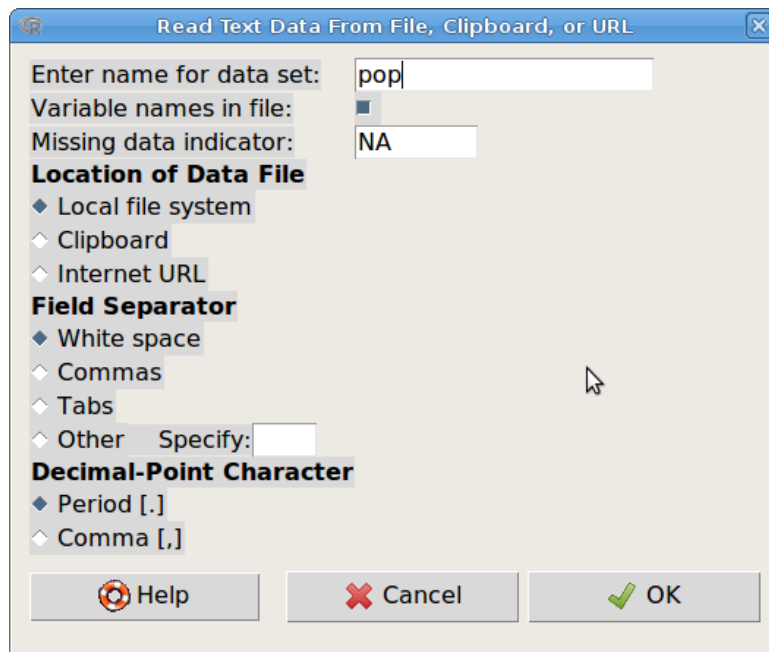
Commas Vesszők

Other Specify Egyéb Specifikálás

Decimal-Point Character Tizedes jel

Period [.] Tizedes pont

Comma [,] Tizedes vessző



5. ábra. Adatbeolvasás szöveges állományból: Data → Import Data → from text file, clipboard, or URL

Az adattáblát beolvasás előtt mindenképpen meg kell nyitni valamilyen egyszerű szöveg-szerkesztővel (pl. Kate, Jegyzettömb stb.), hogy megtudjuk a fent felsoroltakat. Esetünkben az adattábla első 5 sora:

```
magas tomeg matek biol matek.kat
183 73 3 3 rossz
176 99 1 3 rossz
179 84 5 5 jo
180 94 1 3 rossz
```

Látható, hogy vannak változónevek, nincsenek hiányzó adatok, tehát ezek beállításával nem kell foglalkozni. A mezőhatároló a szóköz, így az alapbeállításokon nem kell változtatni (5. ábra).

A Script ablakban a következő kód jelenik meg:

```
>pop <-
read.table("/home/andrea/work/rcmdr/elsoeves.txt",
header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
```

A `read.table()` függvény olvassa be az R-ben a szöveges állományokat. A különböző paraméter beállításokat a párbeszéd ablak beállításainak megfelelően generálta le a program.

Ha Excelből kimentett CSV (*Comma Separated Value*) állományt olvasunk be (pl. `oz.csv`), akkor annak formátuma a Windows magyar területi beállításai esetén:

```
TERULET;SEX;TOMEG;TESTH
Babat;suta;1,9;58
Pitvaros;suta;2,85;59
Pitvaros;suta;2,78;60
Pitvaros;suta;2,59;60
Pitvaros;suta;4,32;64,5
```

Látható, hogy a mezőhatároló karakter a pontosvessző, és a tizedesjel a tizedesvessző (5. ábra). Angol területi beállítások esetén a mezőhatároló a vessző, a tizedesjel pedig a pont.

5.2. Adatbeolvasás Excel állományból

Data → Import data → from Excel, Access or dBase data set...

Adja meg az adattábla nevét, majd keresse meg fájlkezelőben a betöltendő Excel állományt! Több munkalap esetén, válassza ki azt, amelyikre éppen szüksége van!

Javasolt inkább az Excel tábla CSV formátumban történő mentése, majd beolvasása szöveges állományként!

Excel állományok beolvasására menüből MS Windows esetén van lehetőség.

5.3. Az adatok megtekintése, szerkesztése

Az adatokat megnézhetjük a View data set gomb megnyomásával. Szerkeszteni az Edit data set gomb megnyomásával lehet, de nem javasolt.

6. Adattábla műveletek

Az adattábla egészén a Data → Active data set menüvel (6. ábra) különböző műveleteket végezhetünk:

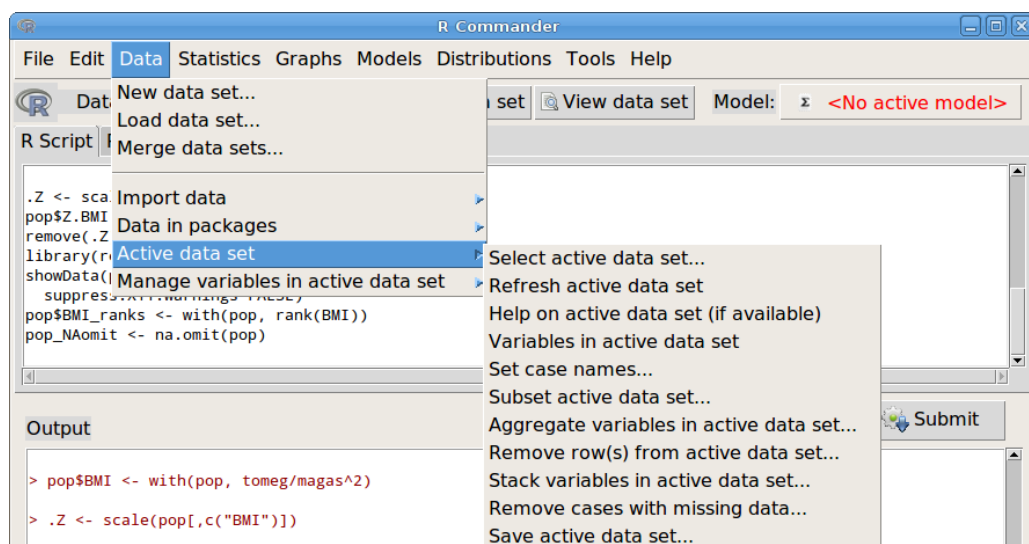
Select active data set Aktív adattábla kiválasztása

Refresh active data set Aktív adattábla frissítése, ha valamilyen változtatást hajtottunk végre rajta

Help on active data set (if available) Az aktív adattáblához tartozó súgó betöltése, ha van (általában R csomagokból betöltött adatállományok esetén van erre lehetőség)

Variables in active data set Az aktív adattábla változói

Set case names... Beállíthatjuk, hogy az aktív adattábla mely oszlopa tartalmazza a megfigyelések (esetek) megnevezését.



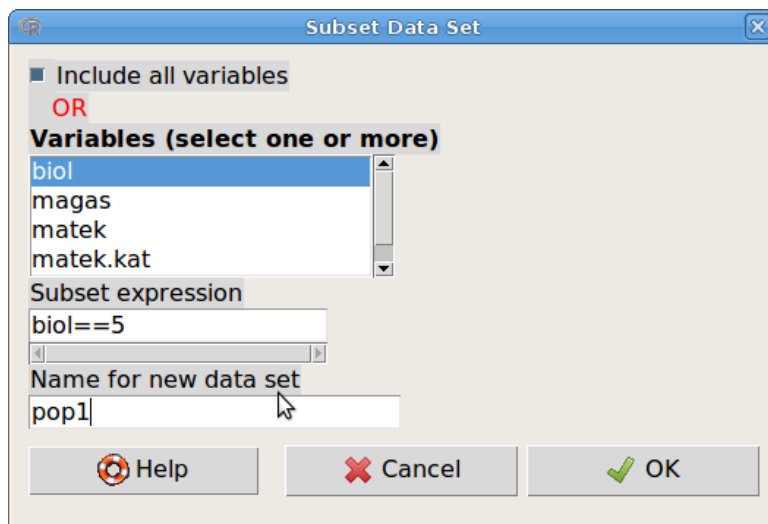
6. ábra. Adattábla műveletek: Data → Active data set

6.1. Rész adattábla leválogatása

Az aktív adattáblából leválogathatunk eseteket, illetve változókat a Data → Active data set → Subset active data set... menüpont segítségével (7. ábra). Alapértelmezésben az összes változót leválogatjuk (Include all variables), vagy kiválaszthatjuk közülük a szükségeseket (Variables (select one or more)).

Az esetek leválogatásához egy logikai kifejezést kell megadnunk. Ez általában úgy néz ki, hogy megadjuk, hogy egy adott változó milyen értékeket vehet fel, illetve több ilyen összekapcsolhatunk 'és'-sel illetve 'vagy'-gyal. A felhasználható operátorokat a 2. táblázat tartalmazza. Faktor illetve szöveges változó esetén az értékeket idézőjelbe, vagy aposztrófok közé kell tenni. Lássunk néhány példát a pop adattáblázzal (ld. 8. oldal)!

<code>biol==5</code>	biológiából 5-öst kapottak
<code>magas>=170</code>	legalább 170 cm magasak
<code>matek.kat=='rossz'</code>	rossz matekosok
<code>biol==5 & matek==5</code>	matekból is és biológiából is 5-öst kapottak
<code>biol==1 matek==1</code>	matekból és/vagy biológiából bukottak



7. ábra. Rész tábla leválogatása: Data → Active data set → Subset active data set...

Végül meg kell adni a rész adattábla nevét (Name for new data set).

A Script ablakban például a következő kód jelenik meg:

```
pop1 <- subset(pop, subset=biol==5)
```

6.2. Aggregált táblázat készítése

Aggregáláskor valamely kategóriás változó vagy változók kategóriái szerint csoportosított adatok összesítő statisztikáit számoljuk ki. Példaként a lepke táblázat BABTOMEG illetve TOMEGO változóinak számítsuk ki az átlagait a HOM és TAP csoportokban (8. ábra).

Name of aggregated data set Az aggregált táblázat neve

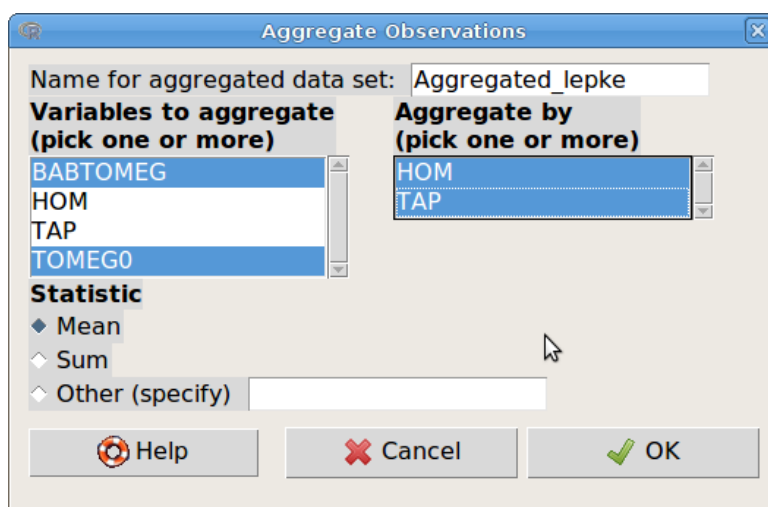
Variables to aggregate Aggregálandó változók

Aggregate by Csoportosító változók

Statistic Összesítő statisztika

2. táblázat. Résztabla leválogatáshoz használható operátorok

Operátor	Leírás
<	kisebb
<=	kisebb vagy egyenlő
>	nagyobb
>=	nagyobb vagy egyenlő
==	egzaktul egyenlő
!=	nem egyenlő
!x	Nem x
x y	x vagy y
x & y	x és y
isTRUE(x)	X igaz-e



8. ábra. Aggregált táblázat készítése: Data → Active data set → Aggregate variables in active data set ...

	HOM	TAP	BABTOMEG	TOMEGO
1	hutott	adlibitum	0.3038000	0.00780000
2	melegített	adlibitum	0.3104167	0.01441667
3	szobahom	adlibitum	0.3008750	0.01191250
4	hutott	limitalt	0.1996667	0.02322222
5	melegített	limitalt	0.1906000	0.02200000
6	szobahom	limitalt	0.2080000	0.02928571

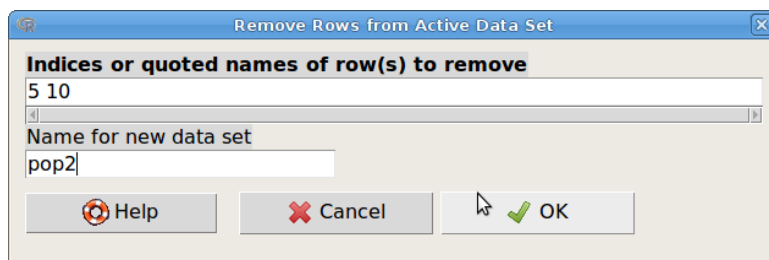
9. ábra. Az aggregált lepke táblázat

Az aggregálás eredménye az `Aggregated_lepke` táblázat lett, amelyet a View data set gomb megnyomásával megnézhetünk (9. ábra).

A *Script* ablakban a következő kód jelenik meg:

```
>Aggregated_lepke <- aggregate(AggregatedData[,c("BABTOMEG", "TOMEGO"),
  drop=FALSE], by=list(HOM=AggregatedData$HOM,
  TAP=AggregatedData$TAP), FUN=mean)
```

6.3. Sorok eltávolítása



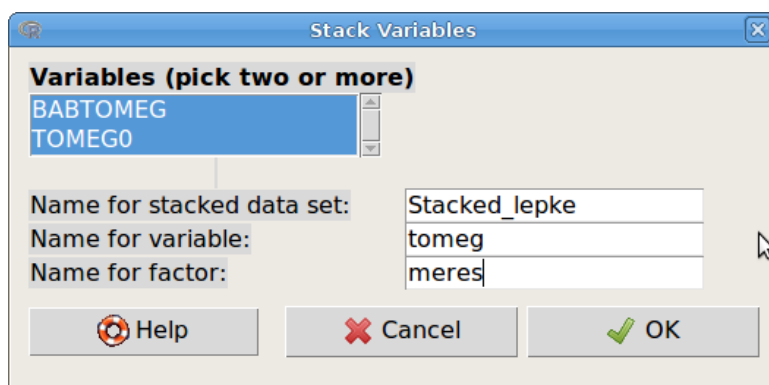
10. ábra. Sorok eltávolítása: Data → Active data set → Remove row(s) form active data set...

Adott sorszámú, vagy megnevezésű sorok eltávolítása az adattáblázatból (10. ábra).

```
>pop2 <- pop[-c(5,10),]
```

6.4. Több változó összefűzése egy változóba (Stack variables...)

Ezzel a lehetőséggel fűzhetünk össze több oszlopban elrendezett adatokat egy oszlopba. Például, ha a lepke táblázat TOMEGO és BABTOMEG változóiba elrendezett tömegértékeket egy tomeg változóba akarjuk rendezni úgy, hogy egy meres nevű új változóban tüntetjük fel, hogy melyik mérésről van szó, akkor a 11. ábrán látható módon kell kitölteni a párbeszéd ablakot. Eredményül a 12. ábrán látható táblázatot kapjuk.



11. ábra. Több változó összefűzése: Data → Active data set → Statck variables in active data set...

6.5. Hiányzó értékeket tartalmazó esetek eltávolítása

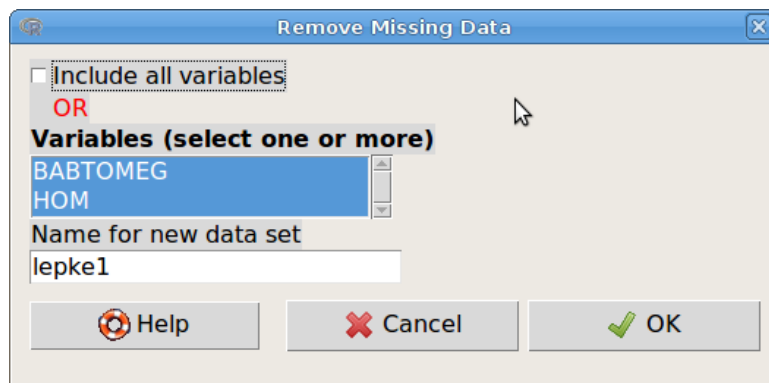
Eltávolíthatjuk a hiányzó értékeket tartalmazó sorokat a teljes táblázatból, vagy kiválasztott változókból (az új táblázatban csak a kiválasztott változók fognak szerepelni). A 13. ábra szerinti beállítások esetén a lepke táblázat BABTOMEG és HOM változóiból eltávolítjuk a hiányzó értékeket tartalmazó sorokat, és ezekből a változókból elkészítjük a lepke1 táblázatot.

```
>lepke1 <- na.omit(lepke1[,c("BABTOMEG", "HOM")])
```

(Tk. 45. oldal, 2.4.4. fejezet)

	tomeg	meres
50	0.3110	BABTOMEG
51	0.2910	BABTOMEG
52	0.3290	BABTOMEG
53	0.2300	BABTOMEG
54	0.3140	BABTOMEG
55	0.3750	BABTOMEG
56	0.3160	BABTOMEG
57	0.0070	TOMEGO
58	0.0130	TOMEGO
59	0.0150	TOMEGO
60	0.0150	TOMEGO
61	0.0170	TOMEGO
62	0.0250	TOMEGO

12. ábra. Összefűzött BABTOMEG és TOMEGO változók a lepke táblázatból



13. ábra. Hiányzó adatokat tartalmazó esetek eltávolítása: Data → Active data set → Remove cases with missing data...

6.6. Az aktív adattábla mentése

A Data → Active data set → Save active data set... menüponttal az aktív adattáblát menthetjük el az R saját adatformátumában (.RData).

(Tk. 404. oldal, 12.5. fejezet)

6.7. Az aktív adattábla exportálása szöveges táblázat formátumba

Az aktív adattáblát szöveges állományként menthetjük (14. ábra). Az exportáláshoz meg kell adnunk a következőket:

Write variable names Változónevek kiírása

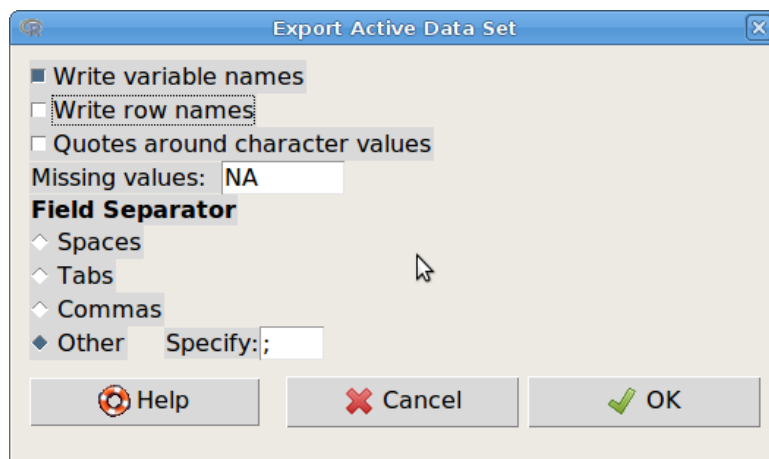
Write row names Sornevek kiírása

Quotes around character values Szöveges mezők idézőjelben

Missing values Hiányzó adat jelölése

Field separator Mezőhatároló karakter:

Spaces Szóközök



14. ábra. Adattábla exportálása: Data → Active data set → Export active data set...

Tabs Tabulátorok

Commas Vesszők

Other Specify Egyéb, megadandó

Magyar beállítású Excel táblázatkezelőbe importáláshoz célszerű a 14. ábrán bemutatott beállításokat alkalmazni, és .csv kiterjesztést adni az exportált táblázatnak.

```
>write.table(lepke1, "/home/andrea/munkakonyvtar/lepke1.csv",
  sep=";", col.names=TRUE, row.names=FALSE, quote=FALSE, na="NA")
```

(Tk. 404. oldal, 12.5. fejezet)

7. Műveletek változókkal

A Data → Manage variables in active data set menü (15. ábra) segítségével készíthetünk származtatott változókat, átkódolhatjuk, törölhetjük stb. változóinkat.

7.1. Változók átkódolása

Általában numerikus illetve faktor változók értékeinek kombinálásával új faktor vagy numerikus változó létrehozásához használható. Az alábbi példában az oz táblázat TESTH (testhossz) változójából hozunk létre egy olyan új változót, amely a 80 cm-nél rövidebbek esetén 'kicsi', a 81–110 cm-es tartományba esőknél 'kozepes', a 110 cm-nél hosszabbak esetén pedig 'nagy' értéket vesz fel (16. ábra). Az átkódolt változó az adattáblázat utolsó oszlopban látható (17. ábra).

Az átkódolás párbeszédablak kitöltendő mezői:

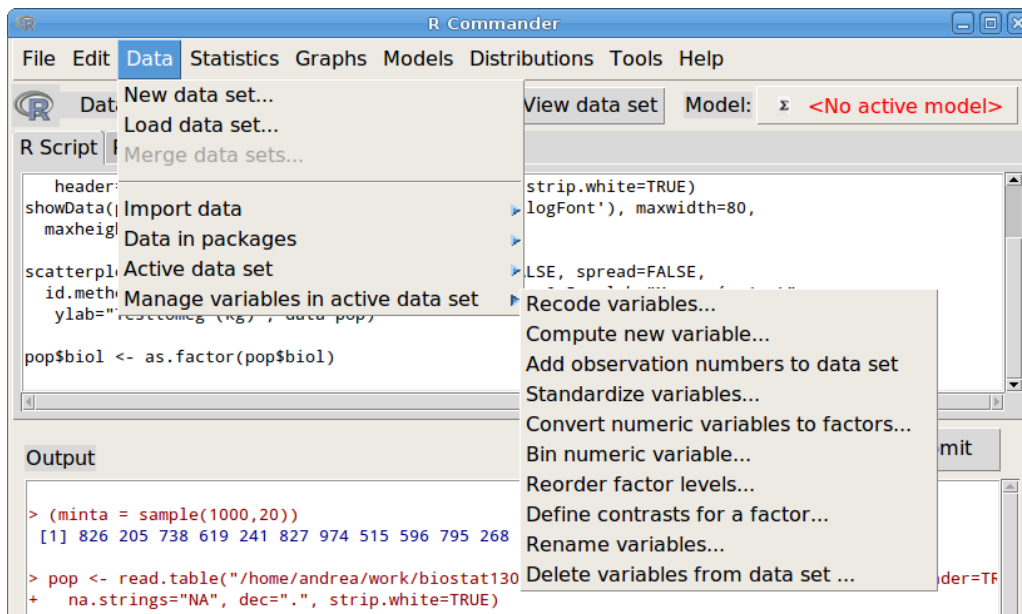
Variables to recode Az átkódolandó változó(k) kiválasztása

New variable name or prefix for multiple recodes Az új, átkódolt változó neve, vagy egy előtag több változó egyszerre történő átkódolása esetén

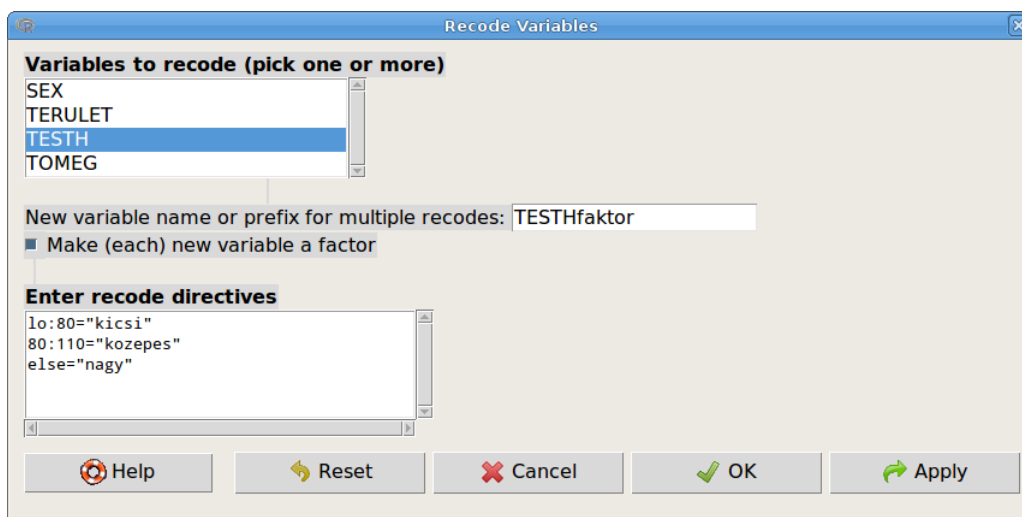
Make (each) new variable a factor Az új változók faktorrá alakítása

Enter recode directives Átkódolási szabályok

- Az egyes szabályokat külön sorba kell írni, vagy pontosvesszővel elválasztani.
- Mindegyik 'input=output' alakú (az aposztrófot nem kell beleírni!).



15. ábra. Változó manipuláció: Data → Manage variables in active data set



16. ábra. Változók átkódolása: Data → Manage variables in active data set → Recode variables

- Az 'NA' megengedett érték.
- A faktorszinteket idézőjelek közé kell tenni.
- Felsorolás esetén az értékeket vessző választja el: '7,8,9,10 = "keves"'
- Tartományokat kettősponttal adunk meg: '7:10="keves"'
- Ha nem tudjuk az alsó, vagy felső határt, használhatjuk a 'lo' illetve 'hi'-t helyettük: 'lo:10="keves"'
- Az 'else' minden egyébre vonatkozik, amit addig nem adtunk meg: 'else="sok"'

(Tk. 370. o. 11.12. példa)

	TERULET	SEX	TOMEG	TESTH	TESTHfaktor
1	Babát	suta	1.90	58.0	kicsi
2	Pitvaros	suta	2.85	59.0	kicsi
3	Pitvaros	suta	2.78	60.0	kicsi
4	Pitvaros	suta	2.59	60.0	kicsi
5	Pitvaros	suta	4.32	64.5	kicsi
6	Babát	bak	3.05	65.0	kicsi
7	Pitvaros	bak	4.55	67.5	kicsi
8	Arpadhalom	suta	5.40	72.0	kicsi
9	Pitvaros	suta	6.41	74.0	kicsi
10	Pitvaros	bak	6.00	74.5	kicsi
11	Babát	suta	9.50	77.5	kicsi
12	Arpadhalom	suta	7.50	77.5	kicsi
13	Pitvaros	suta	7.22	78.0	kicsi
14	Arpadhalom	suta	7.50	79.0	kicsi
15	Arpadhalom	suta	11.00	79.0	kicsi
16	Babát	bak	10.50	79.0	kicsi
17	Pitvaros	suta	7.70	80.5	kozepes
18	Arpadhalom	bak	9.00	81.5	kozepes
19	Arpadhalom	bak	8.00	82.0	kozepes
20	Pitvaros	suta	8.40	82.0	kozepes
21	Pitvaros	suta	8.12	83.0	kozepes
22	Arpadhalom	bak	10.00	84.5	kozepes
23	Arpadhalom	suta	11.50	84.5	kozepes

17. ábra. Az őzes adattábla az átkódolt TESTHfaktor változóval

7.2. Származtatott változók számítása

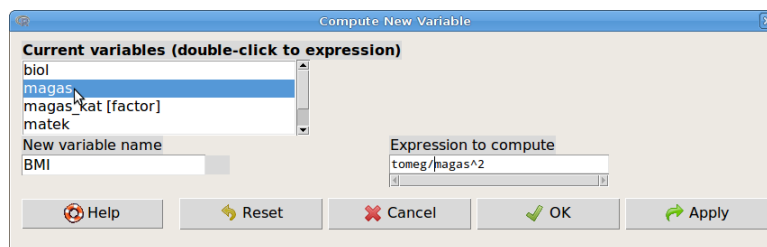
A meglévő változókból készíthetünk új változókat (18. ábra). A párbeszédablak kitöltése:

Current variables (double click to expression) A meglévő változók, kétszer rájuk kattintva átkerülnek az Expression to compute mezőbe

New variable name Új változó neve

Expression to compute A kiszámítandó kifejezés

A példában a BMI-t számítjuk ki ($BMI = \frac{tomeg}{magas^2}$) a testmagasság és tömeg változókból. (TK. 40. o. 2.3. példa)

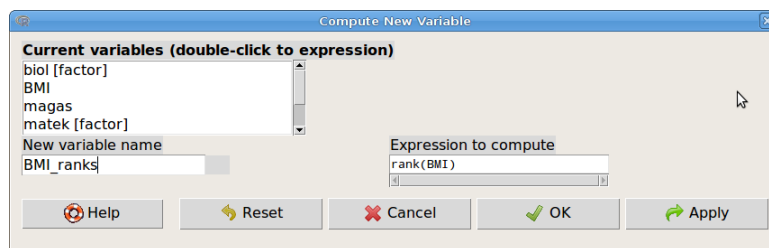


18. ábra. Származtatott változó készítése: Data → Manage variables in active dataset → Compute new variable...

At Expression to compute mezőbe függvényeket is írhatunk. Második példánkban egy numerikus változó értékeit rangokká transzformáljuk a rank() függvény felhasználásával (19. ábra). (Tk. 39. o. 2.4.3. fejezet)

7.3. Numerikus változók faktorrá alakítása

Sokszor előfordul, hogy kategóriás változókat numerikusan kódolnak. Ekkor, az elemzések korrekt elvégzése érdekében, faktorrá kell alakítanunk az adott változókat (20. ábra). A párbeszédablak kitöltése:



19. ábra. Rangok kiszámítása: Data → Manage variables in active data set → Compute new variable

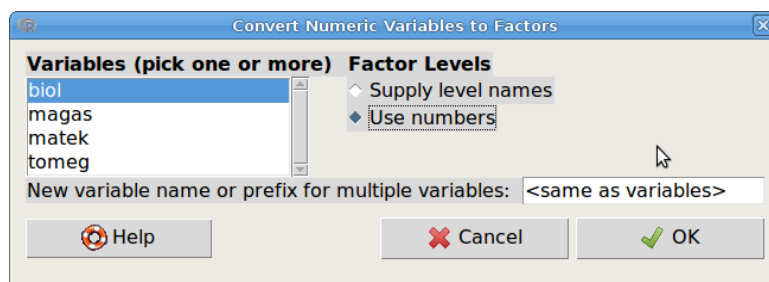
Variables Faktorrá alakítandó változó(k) kiválasztása

Factor levels Faktor szintek megadása

Supply level names Faktorszintek megnevezése

Use numbers Az eredeti számokat tartjuk meg faktorszintekként (szöveges típusú lesz)

New variable name or prefix for multiple variables Új változó név, vagy előtag több változó esetén



20. ábra. Faktorrá alakítás: Data → Manage Variables in Active Dataset → Convert Numeric Variables to Factors

7.4. További műveletek változókkal

Add observation numbers to data set Létrehoz egy sorszám változót az aktív adattáblázatban

Standardize variables... Változó standardizálása

Bin numeric variable... Faktor változót készít egy numerikus változó értéktartományának intervallumokra (bin) osztásával

Reorder factor levels Egy faktor szintjeinek átrendezése

Define contrasts for a factor... Kontraszt megadása a kiválasztott faktorhoz

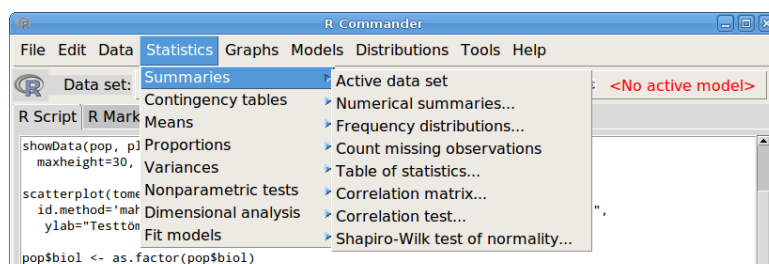
Rename variables... Változók átnevezése

Delete variables from data set... Változók törlése

(Tk. 43. o. 2.4. fejezet)

8. Leíró statisztikák

A leíró statisztikák kiszámítására szolgáló eljárásokat a **Statistics** → **Summaries** menüben találjuk (21. ábra).



21. ábra. Leíró statisztikák: **Statistics** → **Summaries**

8.1. Aktív adattábla összegzése

Az aktív adattábla változóinak alapvető leíró statisztikáit írathatjuk ki (**Statistics** → **Summaries** → **Active data set**). Numerikus változók esetén a minimum (**Min.**), alsó kvartilis (**1st Qu.**), a medián (**Median**), az átlag (**Mean**), a felső kvartilis (**3rd Qu.**) és a maximum (**Max.**) értékeket. Faktor változók esetén pedig a kategóriák gyakoriságait.

```
> summary(pop)
```

magas	tomeg	matek	biol	matek.kat
Min. :158	Min. : 44.00	1:232	1:243	jo :382
1st Qu.:174	1st Qu.: 72.00	2:198	2:165	rossz:618
Median :178	Median : 80.00	3:188	3:218	
Mean :178	Mean : 79.81	4:191	4:144	
3rd Qu.:182	3rd Qu.: 88.00	5:191	5:230	
Max. :197	Max. :112.00			

8.2. Leíró statisztikák numerikus változókra

Az alapvető leíró statisztikákat a **Statistics** → **Summaries** → **Numerical summaries...** segítségével egy kategóriás változó kategóriáira bontva is kiszámíthatjuk. A **lepke** táblázat TAP kategóriái esetén mutatjuk be a funkció használatát (22.–24. ábrák). A 22. ábrán látható ablakban kell kiválasztani a változó(ka)t, illetve a **Summarize by groups** gomb megnyomása után a kategóriás változót (23. ábra). Ezután, rákattintva a **Statistics** gombra, kiválaszthatjuk a kiszámítandó statisztikákat (24. ábra):

Mean Átlag

Standard Deviation Szórás

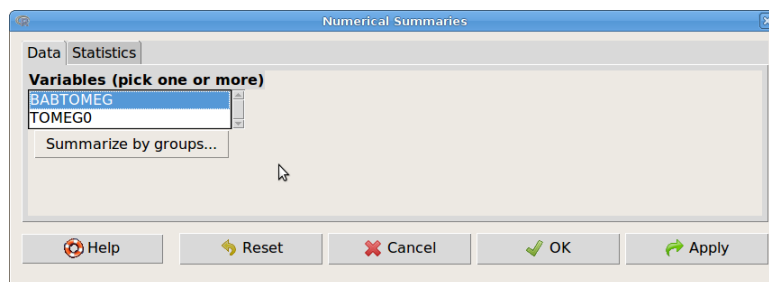
Coefficient of Variation Relatív szórás v. variációs együttható

Skewness Ferdeség

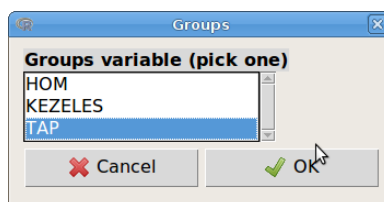
Kurtosis Csúcsosság

(E két utóbbi nem szoktuk használni.)

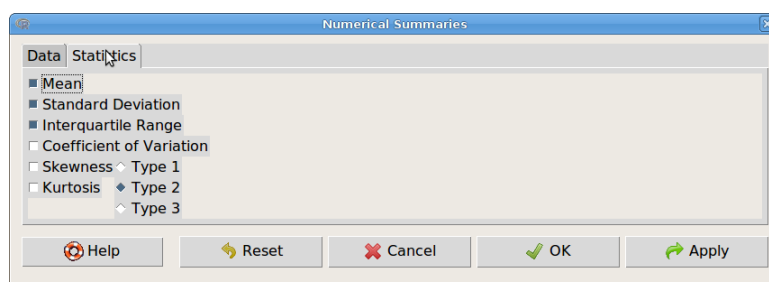
Az alábbi outputban az átlag (**Mean**), a szórás (**sd**), az interkvartilis terjedelem (**IQR**), a variációs együttható (**cv**), a kvartilisek (0%: minimum, 25%: alsó kvartilis, 50%: medián, 75%: felső kvartilis, 100%: maximum) és a kategóriák gyakoriságai (**data:n**) szerepelnek.



22. ábra. Leíró statisztikák numerikus változókra: Statistics → Summaries → Numerical summaries...



23. ábra. Csoportok beállítása: Statistics → Summaries → Numerical summaries... → Summarize by groups



24. ábra. Leíró statisztikák kiválasztása: Statistics → Summaries → Numerical summaries... → Statistics

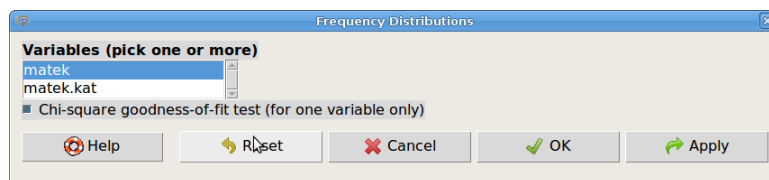
```
> numSummary(lepke[, "BABTOMEG"], groups=lepke$TAP,
  statistics=c("mean", "sd", "IQR", "quantiles", "cv"),
  quantiles=c(0, .25, .5, .75, 1))
```

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	data:n
adlibitum	0.306	0.0295	0.0260	0.0964	0.230	0.291	0.309	0.318	0.375	30
limitalt	0.198	0.0269	0.0323	0.1354	0.143	0.183	0.195	0.215	0.252	26

(Tk. 104. o. 4.2. fejezet, 341. o. 11.1.2. fejezet)

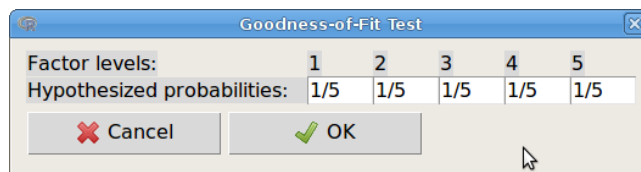
8.3. Gyakorisági eloszlások

Kategóriás (faktor, ha numerikusan kódolt, faktorrá kell először alakítani. ld. 7.3. fejezet) változók esetén gyakoriság táblázatot, illetve az eloszlás vizsgálatára szolgáló Khi-négyzet-próbát a Statistics → Summaries → Frequency distributions... párbeszédablak előhívásával íratathatunk ki, illetve végezhetünk (25. ábra). Ki kell választanunk a vizsgálandó változót (variables (pick one or more)). Ha Khi-négyzet próbát is szeretnénk végezni, akkor be kell jelölnünk a Chi-square goodness-of fit test (for one variable only) opciót.



25. ábra. Gyakorisági táblázat: Statistics → Summaries → Frequency distributions

Az OK gomb megnyomása után felugró (26. ábra) meg kell adnunk az egyes kategóriákba tartozás hipotetikus valószínűségét (alapbeállítás: egyenletes eloszlás).



26. ábra. Hipotetikus valószínűségek beállítása Khi-négyzet próbához: Statistics → Summaries → Frequency distributions

Az output első részébe a gyakoriságok, majd a százalékos gyakoriságok, végül a Khi-négyzet próba eredménye kerül kiíratásra.

```
> .Table <- table(pop$matek)
> .Table # counts for matek

1  2  3  4  5
232 198 188 191 191

> round(100*.Table/sum(.Table), 2) # percentages for matek

1  2  3  4  5
23.2 19.8 18.8 19.1 19.1

> .Probs <- c(0.2,0.2,0.2,0.2,0.2)
> chisq.test(.Table, p=.Probs)
```

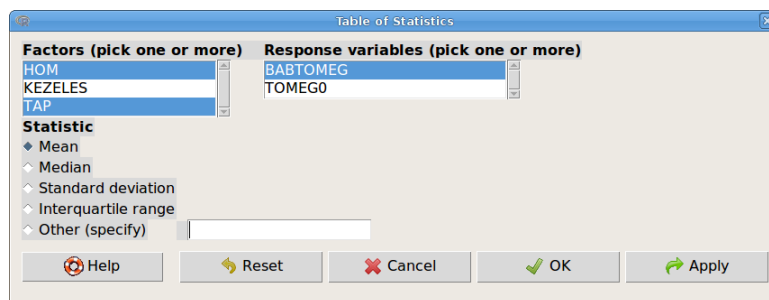
Chi-squared test for given probabilities

```
data: .Table
X-squared = 6.67, df = 4, p-value = 0.1544
```

(Tk. 36. o. 2.2. példa; 88. o. 4.1.1. fejezet; 208. o. 7.3.1. fejezet)

8.4. Hiányzó adatok száma

A Statistics → Summaries → Count missing observations opcióval az aktív táblázat hiányzó adatait számoltathatjuk meg változónként.



27. ábra. Leíró statisztikák táblázata: Statistics → Summaries → Table of statistics...

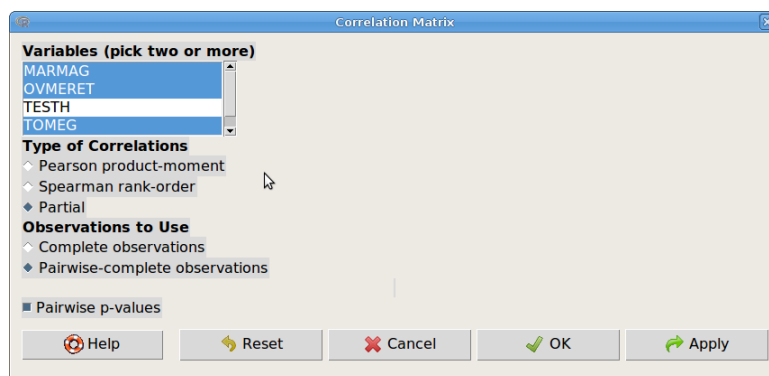
8.5. Leíró statisztikák két faktor szerinti bontásban

A leíró statisztikákat egyesével, két kategóriás változó szerinti bontásban is kiszámíthatjuk (27. ábra). Be kell állítanunk a faktorokat (Factors (pick one or more)) illetve a numerikus változókat (Response variables (pick one or more)), valamint ki kell választanunk a statisztika típusát (Statistic).

```
> tapply(lepke$BABTOMEG, list(HOM=lepke$HOM, TAP=lepke$TAP), mean,
        na.rm=TRUE)
```

HOM	TAP	
	adlibitum	limitalt
hutott	0.3038000	0.1996667
melegített	0.3104167	0.1906000
szobahom	0.3008750	0.2080000

(Tk. 339. o. 11.1.2. fejezet)



28. ábra. Korrelációs mátrix: Statistics → Summaries → Correlation matrix...

8.6. Korrelációs mátrix

Több numerikus változó páronkénti Pearson, Spearman, illetve parciális korrelációját számíthatjuk ki, illetve tesztelhetjük (kétoldali tesztek) a Statistics → Summaries → Correlation matrix funkcióval (28. ábra). Meg kell adnunk két, vagy több változót (Variables (pick two or more)), a korrelációs együtttható típusát, valamint azt, hogy a hiányzó adatokat hogy kezelje a program (Observations to Use). A Complete observations lehetőség választása esetén, a program kihagyja az összes olyan esetet, amelyben bármelyik kiválasztott változó esetén hiányzik az adat. Ha a Pairwise-complete observations lehetőséget választjuk, akkor minden változó pár esetén a lehető legtöbb adatot felhasználja. A Pairwise p-values bejelölése esetén a korrelációkat teszteli is.

Az output első részében a korrelációs együtthatókat, majd a mintaelemszámokat, utána a korrelációs együttható tesztelésére kapott p-értékeket, majd a többszörös tesztelés miatt a Holm módszerrel korrigált p-értékeket Adjusted p-values (Holm's method) láthatjuk.

```
> partial.cor(oz1[,c("MARMAG", "OVMERET", "TOMEG")], tests=TRUE,
  use="pairwise.complete")
```

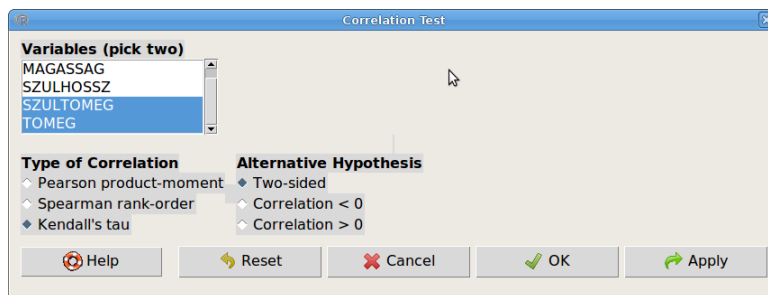
```
Partial correlations:
      MARMAG OVMERET  TOMEG
MARMAG 0.00000 0.19510 0.52434
OVMERET 0.19510 0.00000 0.50112
TOMEG   0.52434 0.50112 0.00000
```

```
Number of observations:
      MARMAG OVMERET  TOMEG
MARMAG   109     107   109
OVMERET   107     107   107
TOMEG     109     107   109
```

```
Pairwise two-sided p-values:
      MARMAG OVMERET  TOMEG
MARMAG      0.045  0.000
OVMERET 0.045      0.000
TOMEG    0.000  0.000
```

```
Adjusted p-values (Holm's method)
      MARMAG OVMERET  TOMEG
MARMAG      0.045  0.000
OVMERET 0.045      0.000
TOMEG    0.000  0.000
```

(Tk. 275. o. 9.10. példa, 245. o. 8. fejezet)



29. ábra. Korreláció tesztelése: Statistics → Summaries → Correlation test...

8.7. Korrelációs együttható tesztelése

Két változó Pearson-, Spearman- illetve Kendall-féle korrelációját tesztelhetjük. Egyoldai tesztek is végezhetőek (29. ábra). Ehhez a következőket kell beállítani:

Variables (pick two) Változók kiválasztása

Type of Correlation A korrelációs együttható típusa

Alternative Hypothesis Az ellenhipotézis típusa

Two-sided $H_1 : \text{Correlation} \neq 0$

Correlation<0 $H_1 : \text{Correlation} < 0$

Correlation>0 $H_1 : \text{Correlation} > 0$

```
> cor.test(regr.kurz$SZULTOMEG, regr.kurz$TOMEG,
           alternative="two.sided", method="kendall")
```

Kendall's rank correlation tau

```
data: regr.kurz$SZULTOMEG and regr.kurz$TOMEG
z = 2.3089, p-value = 0.02095
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.3725535
```

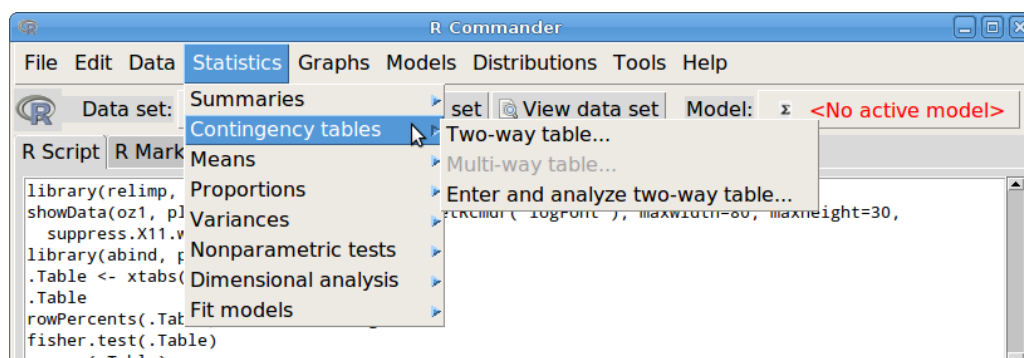
(Tk. 251. o. 8.4. példa)

8.8. Normalitás tesztelése

Egy változó esetén végezhető el a Shapiro-Wilk-féle normalitás vizsgálat (Statistics → Summaries → Shapiro-Wilk test of normality...).

9. Kontingencia-táblák elemzése

Keresztábra, vagy más néven kontingencia-táblázatok elemzésére szolgáló eljárásokat a Statistics → Contingency tables menüben találunk (30. ábra).



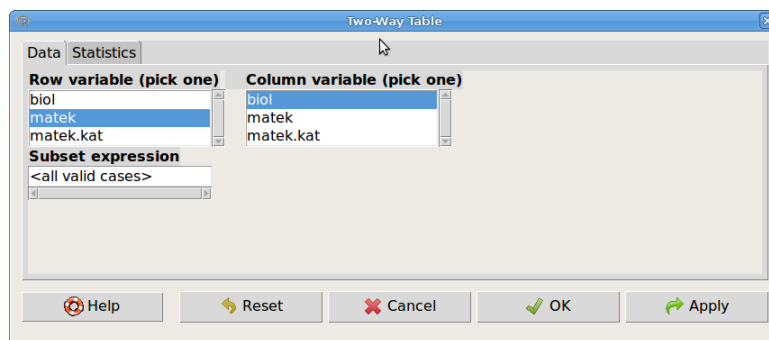
30. ábra. Kontingencia-táblázatok: Statistics → Contingency tables

Készíthetünk és elemezhetünk adattáblázatból kétdimenziós (Two-way table), többdimenziós táblázatot (Multi-way table), illetve elemezhetünk kézzel begépett táblázatot (Enter and analyze two-way table...).

9.1. Kétdimenziós kontingencia-táblák elemzése

Kétdimenziós táblázat elemzéséhez meg kell adnunk a sorokba (Row variable (pick one)), illetve az oszlopokba kerülő **faktor** változót (Column variable (pick one)). Lehetőség van arra is, hogy a táblázatnak csak egy részére végezzük el az elemzéseket, ehhez a 6.1. fejezetben leírt módon meg kell adni a résztáblázat eseteit meghatározó logikai kifejezést (31.ábra).

A változók megadása után a Statistics fülre kattintva beállíthatjuk, hogy milyen formában kérjük a kontingencia-táblát, illetve hogy milyen tesztekkel végezzon el a program (32. ábra).



31. ábra. Kétdimenziós kontingencia-táblák: Statistics → Contingency tables → Two-way table...



32. ábra. Kontingencia-táblák készítése és elemzése: Statistics → Contingency tables → Two-way table... → Statistics

Compute Percentages Százalékos arányok

Row percentages Soronkénti százalékos arányok

Column percentages Oszloponkénti százalékos arányok

Percentages of total Százalékos arányok az összgyakorisághoz képest

No percentages Nem számol százalékos arányokat

Hypothesis Tests Függelenségvizsgálatok

Chi-square test of independence Khi-négyzet teszt

Components of chi-square statistic A khi-négyzet statisztika komponensei

Print expected frequencies Várt gyakoriságok kiírása

Fisher's exact test Fisher-féle egzakt próba

A pop adattábla esetén, az output első részében a kontingencia-táblát láthatjuk, majd a megfelelő százalékos arányokat, végül a függetlenségvizsgálatok (esetünkben Khi-négyzet próba) eredményét. Az elemzés előtt a matek és biol változók faktorrá lettek alakítva.

```
> .Table <- xtabs(~matek+biol, data=pop)
> .Table
```

```
biol
matek  1  2  3  4  5
  1 121 55 56  0  0
  2  80 38 43 37  0
  3  42 29 40 32 45
  4  0 43 39 40 69
  5  0  0 40 35 116
```

```
> totPercents(.Table) # Percentage of Total
```

	1	2	3	4	5	Total
1	12.1	5.5	5.6	0.0	0.0	23.2
2	8.0	3.8	4.3	3.7	0.0	19.8
3	4.2	2.9	4.0	3.2	4.5	18.8
4	0.0	4.3	3.9	4.0	6.9	19.1
5	0.0	0.0	4.0	3.5	11.6	19.1
Total	24.3	16.5	21.8	14.4	23.0	100.0

```
> .Test <- chisq.test(.Table, correct=FALSE)
> .Test
```

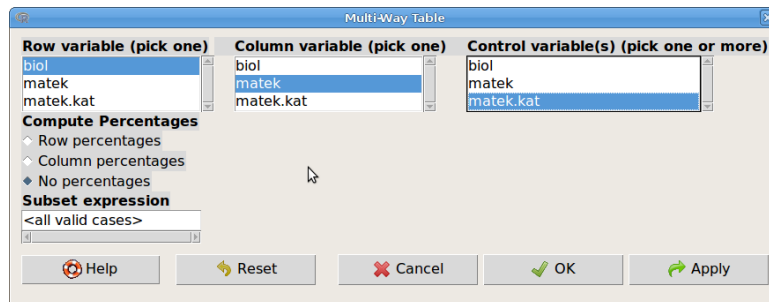
Pearson's Chi-squared test

```
data: .Table
X-squared = 509.2416, df = 16, p-value < 2.2e-16
```

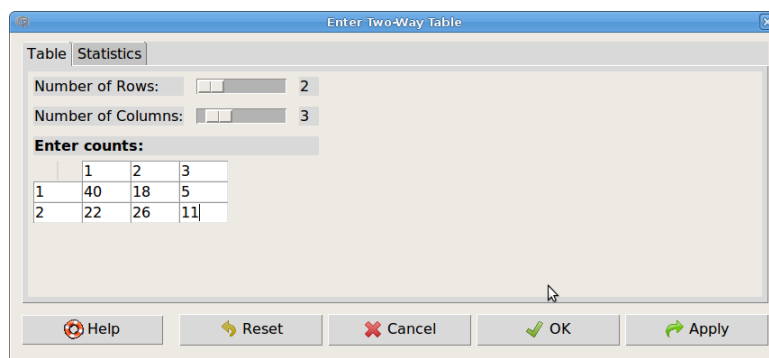
(Tk. 98. o. 4.1.2. fejezet, 4.3. példa)

9.2. Többdimenziós táblázatok

Kettőnél több faktor változó esetén, a sor (Row variable), illetve oszlop változóként (Column variable) megjelölt változók kontingencia-tábláit a többi faktor változó (Control variables) kategóriái szerinti bontásban írhatjuk ki százalékos formában is (33. ábra).



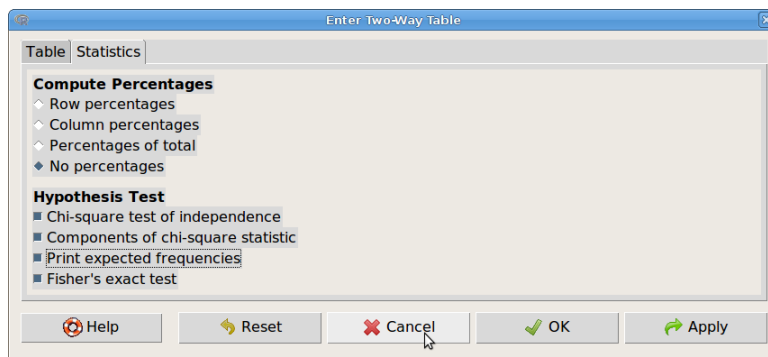
33. ábra. Többdimenziós kontingencia-tábla: Statistics → Contingency tables → Multi-way table...



34. ábra. Kétdimenziós kontingencia-tábla kézi bevitele: Statistics → Contingency tables → Enter and analyze two-way table...

9.3. Kézzel begépelte kontingencia-tábla elemzése

Lehetőségünk van kézzel is begépelni kontingencia-táblát (34. ábra). Ehhez először be kell állítani a táblázat méretét, azaz a sorok számát (Number of Rows) és az oszlopok számát (Number of Columns), majd be kell gépelni a gyakoriságokat. Ezután a Statistics fülre kattintva be lehet állítani az előzőekhez hasonló százalékos arányokat, illetve tesztek (35. ábra).



35. ábra. Kézzel felvitt kétdimenziós tábla elemzési beállításai: Statistics → Contingency tables → Enter and analyze two-way table... → Statistics

A 35. ábrán látható beállításoknak megfelelő output első részében a kontingencia-táblát kapjuk meg, majd a Khi-négyszet teszt eredményét, utána a várt gyakoriságokat, majd a Khi-négyszet komponenseit, végül pedig a Fisher-próba eredményét.

```
> .Table <- matrix(c(40,18,5,22,26,11), 2, 3, byrow=TRUE)
> rownames(.Table) <- c('1', '2')
> colnames(.Table) <- c('1', '2', '3')
> .Table # Counts
  1  2  3
1 40 18  5
2 22 26 11

> .Test <- chisq.test(.Table, correct=FALSE)
> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 8.8087, df = 2, p-value = 0.01222

> .Test$expected # Expected Counts
  1      2      3
1 32.01639 22.72131 8.262295
2 29.98361 21.27869 7.737705

> round(.Test$residuals^2, 2) # Chi-square Components
  1      2      3
1 1.99 0.98 1.29
2 2.13 1.05 1.38

> fisher.test(.Table)

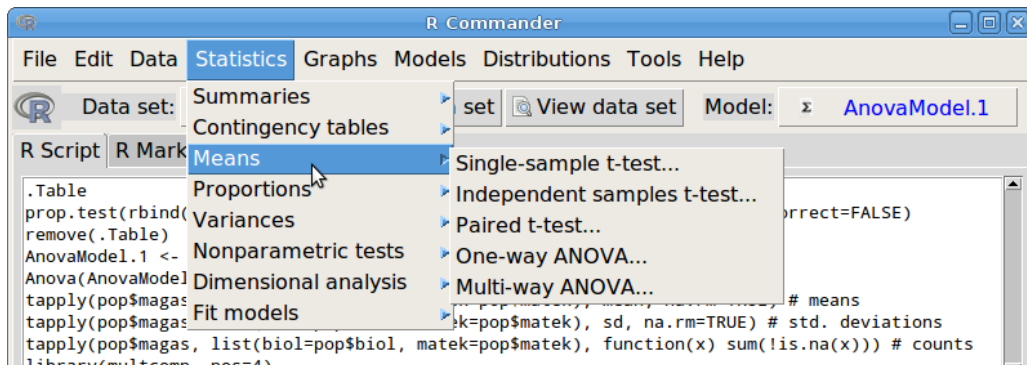
      Fisher's Exact Test for Count Data

data:  .Table
p-value = 0.01073
alternative hypothesis: two.sided
```

(Tk. 214-218. o., 7.3.2. fejezet, 7.6.- 7.10. példák)

10. Hipotézisvizsgálatok: átlagok elemzése

Átlagok elemzésére szolgáló eljárásokat a Statistics → Means menüben találunk (36. ábra).



36. ábra. Átlagok elemzése: Statistics → Means

10.1. Egymintás t -próba

Példánkban az vizsgáljuk egymintás t -próbával (Statistics → Means → Single sample t-test...), hogy az elsőéves hallgatók átlagos tömege szignifikánsan több-e, mint 78 kg (37. ábra). Ehhez meg kell adnunk a következőket:

Variable (pick one) A vizsgálandó változó

Alternative Hypothesis) Az ellenhipotézis típusa

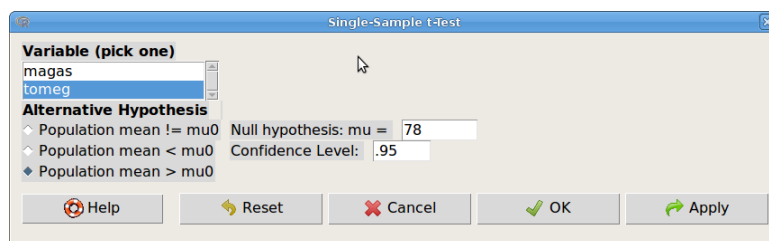
Population mean \neq μ_0 $H_1 : \mu \neq \mu_0$

Population mean $<$ μ_0 $H_1 : \mu < \mu_0$

Population mean $>$ μ_0 $H_1 : \mu > \mu_0$

Null hypothesis: $\mu =$ A tesztelendő hipotetikus érték (μ_0)

Confidence level A mintából becsült populáció átlagra vonatkozó konfidencia-intervallum megbízhatósági szintje



37. ábra. Egymintás t -próba: Statistics → Means → Single sample t-test...

A teszt outputjában megkapjuk a t -statisztika értékét, a szabadsági fokot (df) és a p -értékek (p -value). Ezenkívül, kapunk egy – az alternatív hipotézis típusának megfelelő – konfidencia intervallumot, valamint a mintaátlagot.

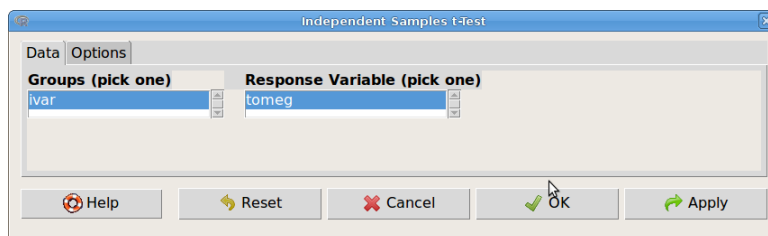
```
> t.test(pop$tomeg, alternative='greater', mu=78, conf.level=.95)
```

One Sample t-test

```
data: pop$tomeg
t = 5.238, df = 999, p-value = 9.895e-08
alternative hypothesis: true mean is greater than 78
95 percent confidence interval:
 79.24247      Inf
sample estimates:
mean of x
 79.812
```

10.2. Két, független mintás *t*-próba

Példánkban az vizsgáljuk kétmintás *t*-próbával (Statistics → Means → Independent samples t-test...), hogy bizonyítják-e az alábbi minták, hogy a bikaborjak (b: bika) átlagos születéskori testtömege nagyobb, mint az üszőké (u: üsző)?(38. ábra). Ehhez meg kell adnunk a következőket (borjak.csv).



38. ábra. Kétmintás *t*-próba: Statistics → Means → Independent samples t-test...

Groups (pick one) Csoportosító változó (2 szintű faktor lehet)

Response variable (pick one) A vizsgálandó változó

Az **Options** fülre kattintva a megjelenő párbeszéd ablakban (39. ábra) pedig a következőket:

Difference A különbség

Two-sided $H_1 : \mu_1 - \mu_2 \neq 0$

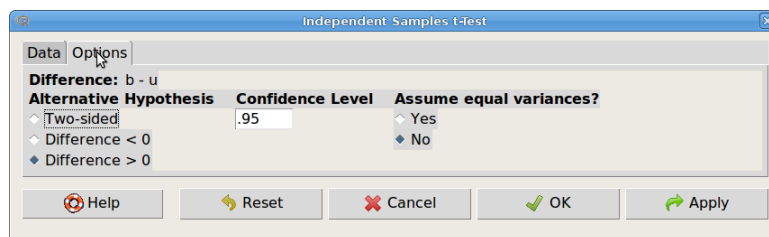
Difference < 0 $H_1 : \mu_1 - \mu_2 < 0$

Difference > 0 $H_1 : \mu_1 - \mu_2 > 0$

Confidence level A mintákból becsült, populációs átlagok különbségére vonatkozó konfidencia-intervallum megbízhatósági szintje.

Assume equal variances? Feltételezzük-e a populációs varianciák egyezőségét? Ha nem, No (alapbeállítás, hagyjuk így!), akkor a Welch-próbát végzi el a program.

A teszt outputjában megkapjuk a *t* statisztika értékét, a szabadsági fokot (df) és a *p*-értékek (p-value). Ezenkívül kapunk, egy – az alternatív hipotézis típusának megfelelő – konfidencia intervallumot a populációs átlagok különbségére, valamint a mintaátlagokat.



39. ábra. Kétmintás t -próba: Statistics → Means → Independent samples t-test... → Options

```
> t.test(tomeg~ivar, alternative='greater', conf.level=.95,
+ var.equal=FALSE, data=borjak)
```

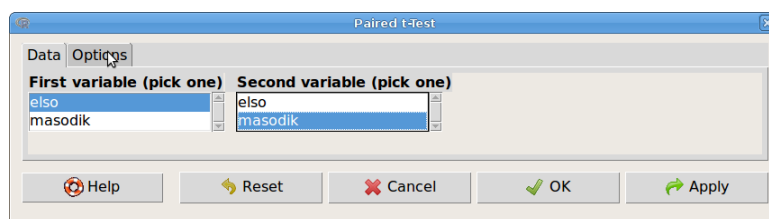
Welch Two Sample t-test

```
data: tomeg by ivar
t = 0.9912, df = 11.736, p-value = 0.1708
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -2.099368      Inf
sample estimates:
mean in group b mean in group u
   39.28571      36.66667
```

(Tk. 200. o. 7.1.2. fejezet, 7.2. példa)

10.3. Két, párosított mintás t -próba

Példánkban az vizsgáljuk páros t -próbával (Statistics → Means → Paired t-test...), hogy bizonyítják-e az adatok, hogy a második gyermek születéskori testtömege meghaladja az elsőét? (40. ábra, gyermek.csv). Ehhez meg kell adnunk a következőket:



40. ábra. Páros t -próba: Statistics → Means → Paired t-test...

First variable (pick one) Az egyik adatsort tartalmazó változó

Second variable (pick one) A másik adatsort tartalmazó változó

Az **Options** fülre kattintva a megjelenő párbeszéd ablakban pedig a következőket (41. ábra).

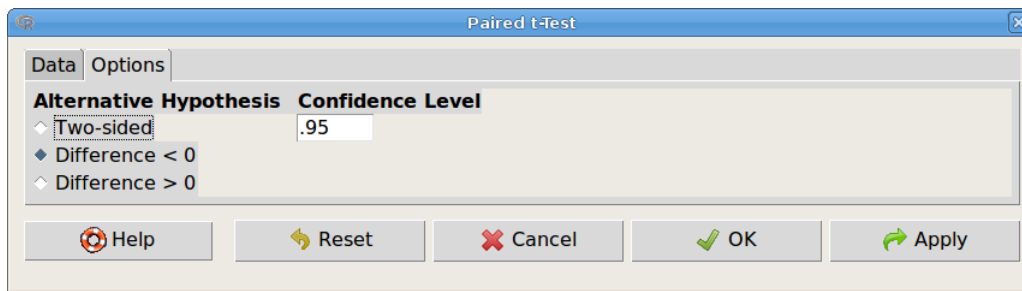
Alternative Hypothesis Alternatív hipotézis típusa

Two-sided $H_1 : \mu_1 - \mu_2 \neq 0$

Difference < 0 $H_1 : \mu_1 - \mu_2 < 0$

Difference > 0 $H_1 : \mu_1 - \mu_2 > 0$

Confidence level A mintákból becsült populációs átlagok különbségére vonatkozó konfidencia-intervallum megbízhatósági szintje.



41. ábra. Páros t -próba: Statistics → Means → Paired t-test... → Options

A teszt outputjában megkapjuk a t -statisztika értékét, a szabadsági fokot (df) és a p -értékek (p -value). Ezenkívül kapunk, egy – az alternatív hipotézis típusának megfelelő – konfidencia intervallumot a populációs átlagok különbségére, valamint a különbségek átlagát.

```
t.test(gyermek$elso, gyermek$masodik, alternative='less',
+ conf.level=.95, paired=TRUE)
```

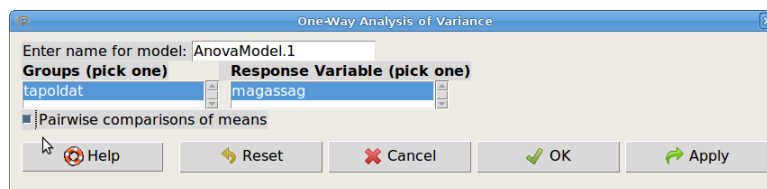
Paired t-test

```
data: gyermek$elso and gyermek$masodik
t = -1.6692, df = 9, p-value = 0.06471
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 12.47327
sample estimates:
mean of the differences
 -127
```

(Tk. 201. o., 7.1.3. fejezet, 7.3. példa)

10.4. Egytényezős ANOVA

Több átlag összehasonlítását varianciaelemzéssel végezzük el (Statistics → Means → One-way ANOVA...),. Példánkban egy kísérletben egy tápoldatot tesztlünk! A kísérletet 12 növényel végezzük, amelyek közül sorsolással eldöntjük, hogy melyik kapjon tiszta vizet, és melyiket öntözzük tömény, illetve híg oldattal. A növények magasságát vizsgáljuk. (tápoldat.csv). Az elemzéshez meg kell adnunk a következőket (42. ábra).



42. ábra. Egytényezős ANOVA: Statistics → Means → One-way ANOVA...

Enter name of model: A modell elnevezése

Groups (pick one) Csoportosító változó

Response variable (pick one) A vizsgálandó célváltozó

Pairwise comparisons of means Páronkénti összehasonlítások elvégzése

A teszt outputjában megkapjuk az ANOVA-táblázatot a p -értékkel ($Pr(>F)$). Ezenkívül kapunk egy táblázatot a mintaátlagokkal, szórásokkal és mintaelemszámokkal.

```

> AnovaModel.1 <- aov(magassag ~ tapoldat, data=adat)
> summary(AnovaModel.1)

Df Sum Sq Mean Sq F value    Pr(>F)
tapoldat      2  303.5   151.75    18.84 0.000607 ***
Residuals     9    72.5     8.06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> numSummary(adat$magassag , groups=adat$tapoldat, statistics=c("mean", "sd"))

      mean      sd data:n
hig    56.75 1.258306      4
tomeny 61.75 3.304038      4
viz    49.50 3.415650      4

```

A páronkénti összehasonlítások eredményeként tesztek és konfidencia intervallumokat kapunk a páronkénti különbségekre, a homogén csoportokat (ahol azonos betű van, azok a csoport-átlagok nem különböznek szignifikánsan), valamint egy ábrát a különbségekkel és konfidencia-intervallumaikkal (43. ábra).

```

> .Pairs <- glht(AnovaModel.1, linfct = mcp(tapoldat = "Tukey"))
> summary(.Pairs) # pairwise tests

```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = magassag ~ tapoldat, data = adat)

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
tomeny - hig == 0	5.000	2.007	2.491	0.0800 .
viz - hig == 0	-7.250	2.007	-3.612	0.0139 *
viz - tomeny == 0	-12.250	2.007	-6.104	<0.001 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```

```

> confint(.Pairs) # confidence intervals

```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = magassag ~ tapoldat, data = adat)

Quantile = 2.7923

95% family-wise confidence level

Linear Hypotheses:

	Estimate	lwr	upr
tomeny - hig == 0	5.0000	-0.6040	10.6040
viz - hig == 0	-7.2500	-12.8540	-1.6460
viz - tomeny == 0	-12.2500	-17.8540	-6.6460

```

> cld(.Pairs) # compact letter display

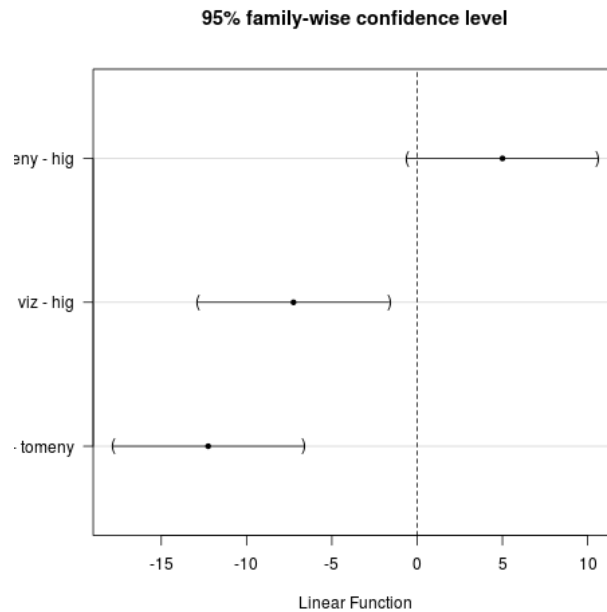
```

```

hig tomeny viz
  "b"   "b"   "a"

```

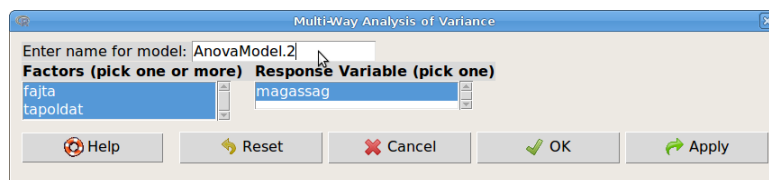
(Tk. 322. o. 10.1. fejezet, 10.1.-2. példa)



43. ábra. Páronkénti különbségek konfidencia-intervallumokkal

10.5. Többtenyezős ANOVA

Az előző fejezet tápoldatos kísérletet megismételték úgy is, hogy a szóban forgó növény két fajtáját kezelték az oldatokkal (`tapoldat2.csv`). A kiértékelést a többtenyezős ANOVA elemzéssel végezzük el (**Statistics** → **Means** → **Multi-way ANOVA...**). (A `fajta` változót faktorrá kell alakítani!) Az elemzéshez meg kell adnunk a következőket (44. ábra).



44. ábra. Többtenyezős ANOVA: **Statistics** → **Means** → **Multi-way ANOVA...**

Enter name of model: A modell elnevezése

Factors (pick one or more) Tényezők (faktorok)

Response variable (pick one) A vizsgálandó célváltozó

A teszt outputjában megkapjuk az ANOVA-táblázatot a p -értékekkel ($\text{Pr}(>F)$). Ezenkívül kapunk egy-egy táblázatot a kezelés kombinációkénti mintaátlagokkal, szórással és minta elemszámokkal.

```
> AnovaModel.2 <- (lm(magassag ~ fajta*tapoldat, data=adat))
> Anova(AnovaModel.2)
```

Anova Table (Type II tests)

Response: magassag

	Sum Sq	Df	F value	Pr(>F)	
fajta	42.67	1	5.4857	0.03087	*
tapoldat	777.58	2	49.9875	4.481e-08	***
fajta:tapoldat	13.08	2	0.8411	0.44751	
Residuals	140.00	18			

```

> tapply(adat$magassag, list(fajta=adat$fajta, tapoldat=adat$tapoldat),
  mean, na.rm=TRUE)
+ # means

tapoldat
fajta  hig tomeny  viz
  1  56.75  61.75 49.50
  2  55.25  60.00 44.75

> tapply(adat$magassag, list(fajta=adat$fajta, tapoldat=adat$tapoldat),
  sd, na.rm=TRUE)
+ # std. deviations

tapoldat
fajta  hig  tomeny  viz
  1  1.258306  3.304038  3.41565
  2  3.403430  2.160247  2.50000

> tapply(adat$magassag, list(fajta=adat$fajta, tapoldat=adat$tapoldat),
  function(x) sum(!is.na(x))) # counts

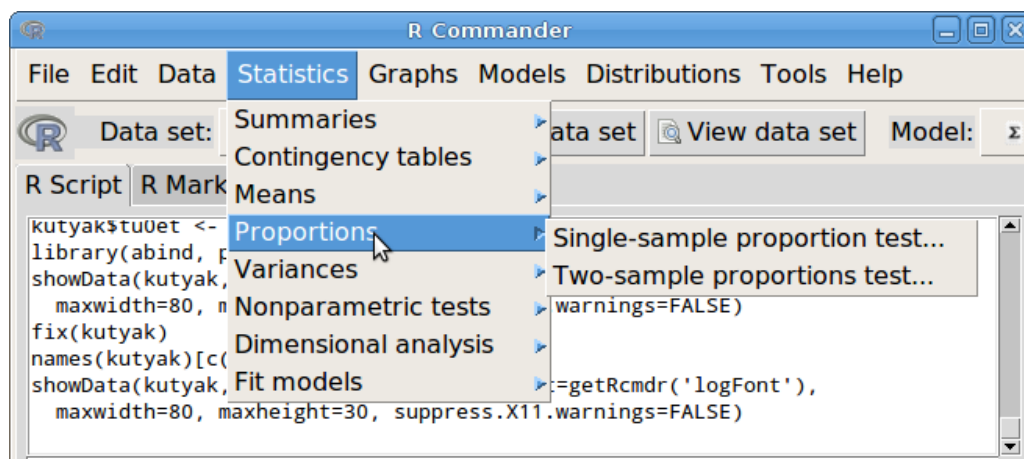
tapoldat
fajta hig tomeny viz
  1  4  4  4
  2  4  4  4

```

(Tk. 328. o. 10.3. fejezet, 10.3. példa)

11. Arányokra vonatkozó próbák

Arányokra (valószínűségekre) vonatkozó próbákat a Statistics → Proportions menüben találunk (45. ábra).

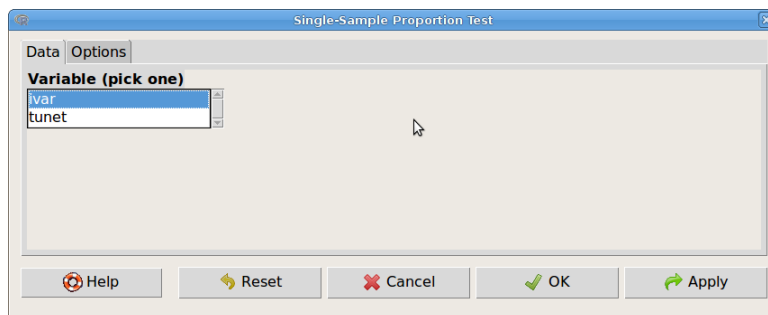


45. ábra. Arányokra vonatkozó próbák: Statistics → Proportions

A próbák használatát egy olyan táblázat adatain keresztül mutatjuk be, amelyben kutyák adatai (kutyak.csv) vannak egy bizonyos betegség tüneteinek meglétéről (tunet, 0: nincs, 1: van), illetve a kutyák ivaráról (ivar, 0: hím, 1: szuka). A 0,1-es kódolás szükséges a kétmintás próbához, és mindkét változót faktorra kell alakítani az elemzés előtt. Mindkét faktor két szintű kell, hogy legyen.

11.1. Egy arány (valószínűség) vizsgálata

Vizsgáljuk meg először azt, hogy az ivararány megfelel-e az 1:1-es aránynak? Egzakt binomiális próbával végezzük el az elemzést. Ehhez meg kell adni a következőket (46. ábra):



46. ábra. Egy arány (valószínűség) vizsgálata: Statistics → Proportions → Single-sample proportion test...

Variable (pick one) A vizsgálandó változó

Az **Options** fülre kattintva a megjelenő párbeszéd ablakban (47. ábra) pedig a következőket:

Alternative Hypothesis Az alternatív hipotézis típusa

Population proportion $\neq p_0$ $H_1 : p \neq p_0$

Population proportion $< p_0$ $H_1 : p < p_0$

Population proportion $> p_0$ $H_1 : p > p_0$

Type of test A teszt típusa

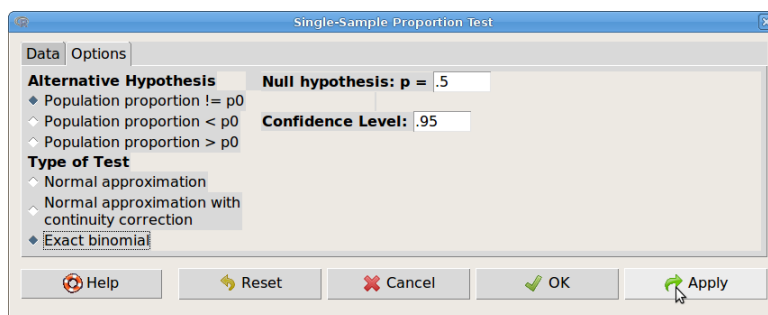
Normal approximation Normális közelítés korrekció nélkül

Normal approximation with continuity correction Normális közelítés folytonossági korrekcióval

Exact binomial Egzakt binomiális próba

Null hypothesis: p = .5 A tesztelendő hipotetikus valószínűség (p_0)

Confidence Level A mintából becsült populációs arányra vonatkozó konfidencia-intervallum megbízhatósági szintje



47. ábra. Egy arány vizsgálatának beállításai: Statistics → Proportions → Single-sample proportion test... → Options

A teszt outputjában megkapjuk az ivarok gyakoriságait, a p -értéket (p -value). Ezenkívül kapunk egy – az alternatív hipotézis típusának megfelelő – konfidencia intervallumot a populációs arányra, valamint az arány becslését.

```

> .Table <- xtabs(~ ivar , data= kutyak )
> .Table

ivar
 0  1
22 25

> binom.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.95)

```

Exact binomial test

```

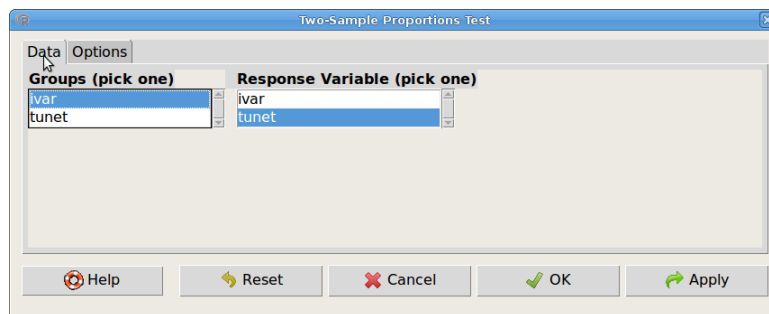
data: rbind(.Table)
number of successes = 22, number of trials = 47, p-value = 0.7709
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3211153 0.6192221
sample estimates:
probability of success
 0.4680851

```

(Tk. 221. o., 7.4.1. fejezet)

11.2. Két arány hasonlítása független minták esetén

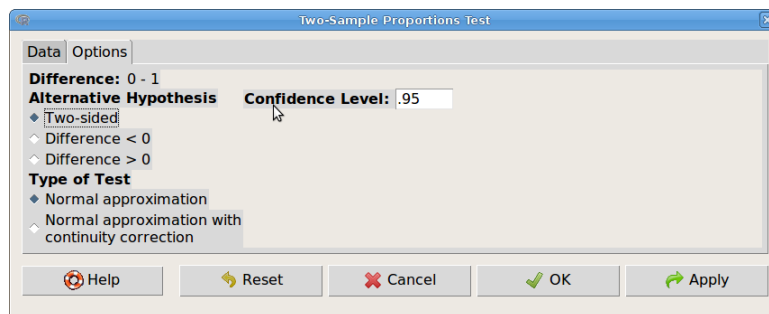
Példánkban azt vizsgáljuk, hogy a tünet előfordulási arány azonosnak tekinthető-e a hímek és szukák esetén. Korrekció nélküli normális közelítéssel végezzük el az elemzést. Ehhez meg kell adni a következőket (48. ábra):



48. ábra. Két arány vizsgálata: Statistics → Proportions → Two-sample proportion test...

Groups (pick one) Csoportosító változó

Response variable (pick one) A vizsgálandó változó



49. ábra. Két arány összehasonlítása: Statistics → Proportions → Two-sample proportion test... → Options

Az **Options** fülre kattintva a megjelenő párbeszéd ablakban (49. ábra) pedig a következőket:

Difference Különbség

Alternative Hypothesis Az alternatív hipotézis típusa

Two-sided $H_1 : p_1 \neq p_2$

Difference < 0 $H_1 : p_1 < p_2$

Difference > 0 $H_1 : p_1 > p_2$

Type of test A teszt típusa

Normal approximation Normális közelítés korrekció nélkül

Normal approximation with continuity correction Normális közelítés folytonossági korrekcióval

Confidence Level A mintából becsült populációs arányok különbségére vonatkozó konfidencia-intervallum megbízhatósági szintje

A teszt outputjában megkapjuk a tünet százalékos előfordulási arányát a két ivar esetén, valamint az ivarok gyakoriságait. A teszt outputjában megkapjuk a Khi-négyzet statisztika (X-squared) értékét, a szabadsági fokot (df), valamint a p -értéket (p-value). Ezenkívül kapunk egy – az alternatív hipotézis típusának megfelelő – konfidencia-intervallumot a populációs arányok különbségére, valamint az arányok becslését.

```
> .Table <- xtabs(~ivar+tunet, data=kutyak)
> rowPercents(.Table)
```

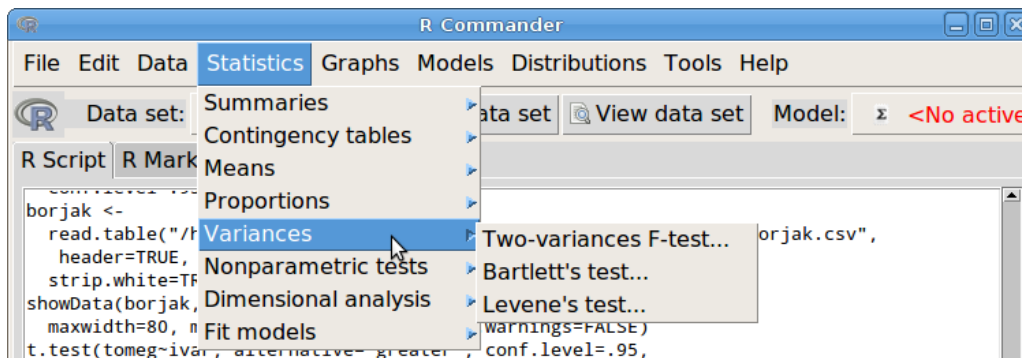
```
      tunet
ivar   0    1 Total Count
  0 22.7 77.3    100     22
  1 80.0 20.0    100     25
```

```
> prop.test(.Table, alternative='two.sided', conf.level=.95,
+ correct=FALSE)
```

```
      2-sample test for equality of proportions without continuity
      correction
```

```
data: .Table
X-squared = 15.4168, df = 1, p-value = 8.622e-05
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.8077818 -0.3376728
sample estimates:
 prop 1    prop 2
0.2272727 0.8000000
```

(Tk. 224. o., 7.4.2. fejezet)



50. ábra. Varianciák elemzése: Statistics → Variances

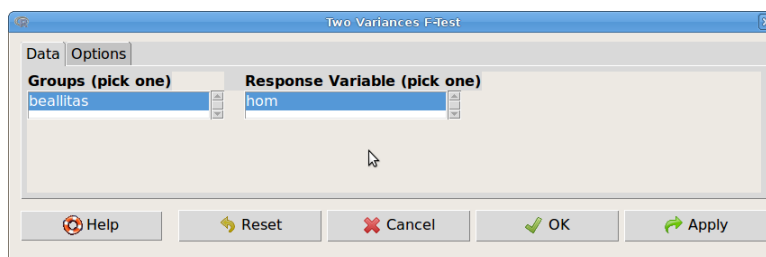
12. Varianciák elemzése

Varianciák elemzésére szolgáló eljárásokat a Statistics → Variances menüben találunk (50. ábra).

A Bartlett-próba (**Bartlett's test**) az F -próba többmintás változata, a Levene-próbához hasonlóan kell paraméterezni (a középérték (Center) beállításától eltekintve).

12.1. Két, független mintás F -próba

Példánkban az vizsgáljuk kétmintás F -próbával (Statistics → Variances → Two variances F-test...), hogy egy keltetőgép felülvizsgálatakor ugyanolyan jó-e, ha alacsonyabb (36°C), illetve ha magasabb (44°C) hőfokot állítanak be rajta? (51. ábra, kelteto.csv). Ehhez meg kell adnunk a következőket:



51. ábra. F -próba: Statistics → Variances → Two variances F-test...

Groups (pick one) Csoportosító változó (2 szintű faktor lehet)

Response variable (pick one) A vizsgálandó változó

Az **Options** Az megjelenő párbeszéd ablakban pedig a következőket:

Difference A különbség

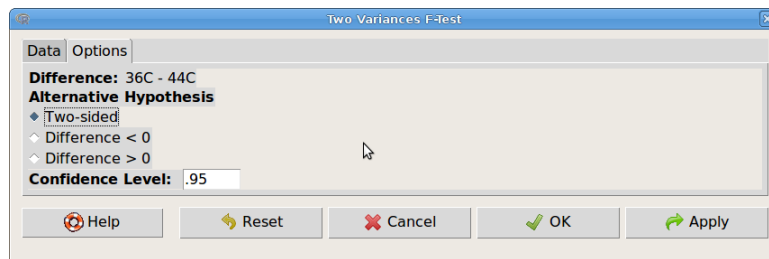
Two-sided $H_1 : \sigma_1^2 \neq \sigma_2^2$

Difference < 0 $H_1 : \sigma_1^2 < \sigma_2^2$

Difference > 0 $H_1 : \sigma_1^2 > \sigma_2^2$

Confidence Level A mintákból becsült populációs varianciák hányadosára vonatkozó konfidencia-intervallum megbízhatósági szintje

A teszt outputjában megkapjuk a F -statisztika értékét, a számláló (num df) és a nevező szabadsági fokát (denom df) és a p -értéket (p-value). Ezen kívül, kapunk egy – az alternatív hipotézis típusának megfelelő – konfidencia intervallumot a populációs varianciák hányadosára, valamint a mintából számolt varianciák hányadosát.



52. ábra. F -próba beállítások: Statistics → Means → Independent samples t-test... → Options

```
> var.test(hom ~ beallitas, alternative='two.sided',
           conf.level=.95, data=kelteto)
```

F test to compare two variances

```
data: hom by beallitas
F = 0.2412, num df = 9, denom df = 9, p-value = 0.04565
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.05990248 0.97093703
sample estimates:
ratio of variances
 0.241167
```

(Tk. 204. o. 7.2.2. fejezet, 7.4. példa)

12.2. Levene-próba

Az előző (12.1). fejezetben szereplő példánkat elemezzük Levene-próbával (Statistics → Variances → Levene's test...)(53. ábra, *kelteto.csv*). Ehhez meg kell adnunk a következőket:

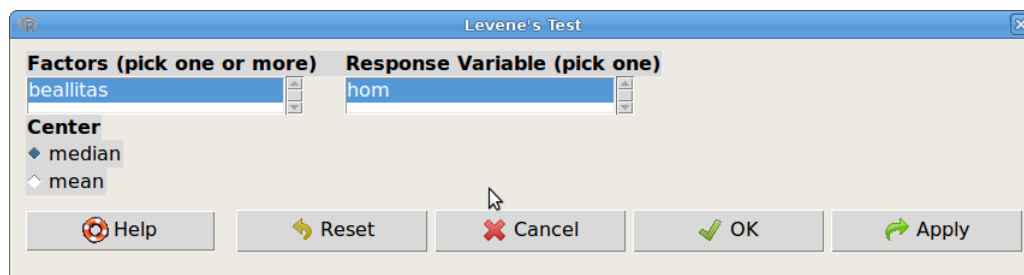
Factors (pick one) Csoportosító változó

Response variable (pick one) A vizsgálandó változó

Center Középvérték típusa

Median Medián, robusztusabb teszt, inkább ezt használjuk!

Mean Átlag, eredeti Levene-próba



53. ábra. Levene-próba: Statistics → Variances → Levene's test...

A teszt outputjában megkapjuk a szabadsági fokokat (df), az F statisztika értékét, és a p -értéket ($\Pr(>F)$).

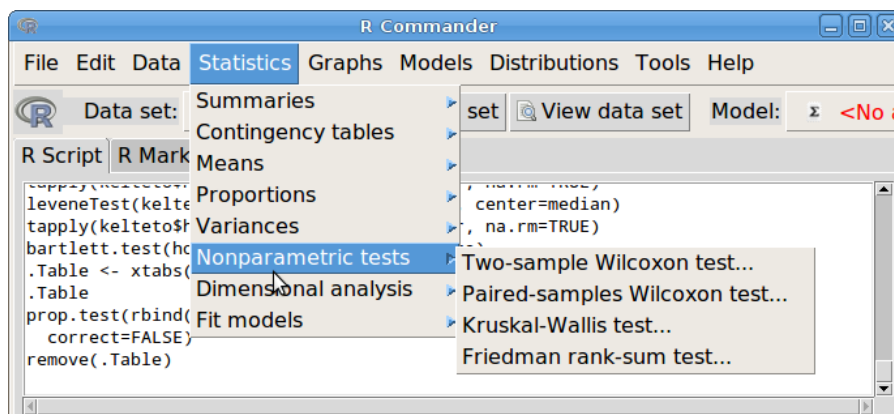
```
> leveneTest(kelteto$hom, kelteto$beallitas, center=median)
```

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1 8.0907 0.01076 *
      18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Tk. 207. oldal, 7.2.2. fejezet)

13. Nemparaméteres próbák

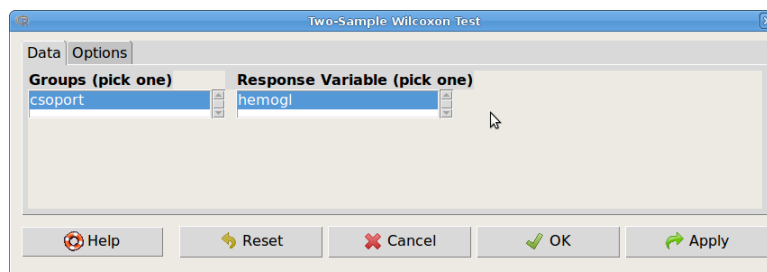
Nemparaméteres próbákat a Statistics → Nonparametric tests menüben találunk (36. ábra).



54. ábra. Nemparaméteres próbák: Statistics → Nonparametric tests

13.1. Két, független mintás Wilcoxon–Mann–Whitney próba

Példánkban azt vizsgáljuk egy kétmintás próbával (Statistics → Nonparametric tests → Two-samples Wilcoxon test...), hogy egy kísérletben, melyben enyhe vérszegénység vaskészítménnyel való kezelését tesztelték 10 kezelttel és 10 placebo-kontrollal, a kísérleti egyedeket a két csoportba véletlenszerűen besorolva, hogy a kezelt csoport hemoglobinszintje (g/dl) magasabb lett-e?. A kontrollcsoportban az egyik mérés nem sikerült, ezért ott csak 9 érték van.?(55. ábra, hemoglobin.csv). Ehhez meg kell adnunk a következőket:



55. ábra. Kétmintás Wilcoxon–Mann–Whitney próba: Statistics → Nonparametric tests → Two-samples Wilcoxon test...

Groups (pick one) Csoportosító változó (2 szintű faktor lehet)

Response variable (pick one) A vizsgálandó változó

Az **Options** fülre kattintva megjelenő párbeszéd ablakban (56. ábra) pedig a következőket:

Difference Eltolás

Alternative Hypothesis Az alternatív hipotézis típusa

Two-sided $H_1 : \text{eltolás} \neq 0$

Difference < 0 $H_1 : \text{eltolás} < 0$

Difference > 0 $H_1 : \text{eltolás} > 0$

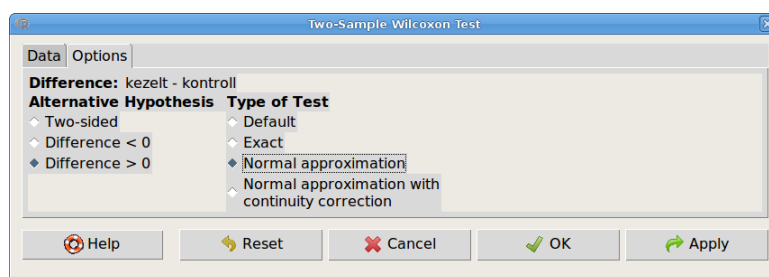
Type of test A teszt típusa

Default Alapbeállítás

Exact Egzakt módszer

Normal approximation Normális közelítés korrekció nélkül

Normal approximation with continuity correction Normális közelítés folytonossági korrekcióval



56. ábra. Kétmintás Wilcoxon–Mann–Whitney próba: Statistics → Nonparametric tests → Two-samples Wilcoxon test... → Options

A teszt outputjában megkapjuk a minták mediánját, normális közelítést használva a W statisztika értékét és a p -értéket (p -value).

```
> tapply(hemogl$hemogl, hemogl$csoport, median, na.rm=TRUE)
```

```
kezelt kontroll  
10.45    9.20
```

```
> wilcox.test(hemogl ~ csoport, alternative='greater', exact=FALSE,  
+ correct=FALSE, data=hemogl)
```

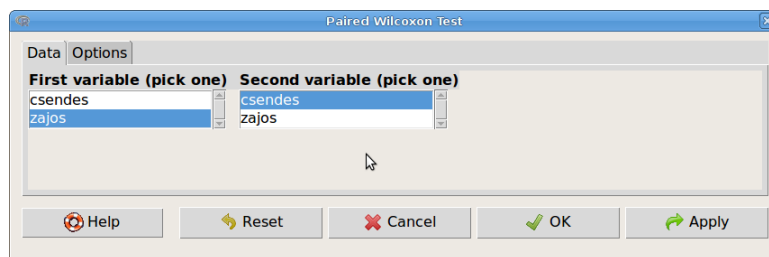
```
Wilcoxon rank sum test
```

```
data: hemogl by csoport  
W = 76.5, p-value = 0.00499  
alternative hypothesis: true location shift is greater than 0
```

(Tk. 237. o. 7.6.2. fejezet, 7.18. példa)

13.2. Két, párosított mintás Wilcoxon–próba

Példánkban az vizsgáljuk egy páros próbával (Statistics → Nonparametric tests → Paired-samples Wilcoxon test...), hogy tíz kísérleti személynek ugyanazzal a módszerrel mérve a reakcióidejét csendes és zajos környezetben, bizonyíthatóan nagyobb-e a reakcióidő zajos környezetben? (57. ábra, `reakcio_ido.csv`). Ehhez meg kell adnunk a következőket:



57. ábra. Páros Wilcoxon-próba: Statistics → Nonparametric tests → Paired-samples Wilcoxon test...

First variable (pick one) Egyik adatsort tartalmazó változó

Second variable (pick one) Másik adatsort tartalmazó változó

Az **Options** fülre kattintva megjelenő párbeszéd ablakban (58. ábra) pedig a következőket:

Alternative Hypothesis Az alternatív hipotézis típusa

Two-sided H_1 : a különbségek mediánja $\neq 0$

Difference < 0 H_1 : a különbségek mediánja < 0

Difference > 0 H_1 : a különbségek mediánja > 0

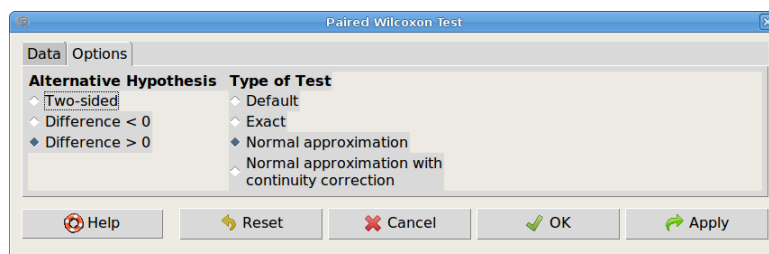
Type of test A teszt típusa

Default Alapbeállítás

Exact Egzakt módszer

Normal approximation Normális közelítés korrekció nélkül

Normal approximation with continuity correction Normális közelítés folytonossági korrekcióval



58. ábra. Páros Wilcoxon-próba beállításai: Statistics → Nonparametric tests → Paired-samples Wilcoxon test... → Options

A teszt outputjában (normális közelítést használva) a W statisztika értékét és a p -értéket (p-value) kapjuk meg.

```
> wilcox.test(reakcio$zajos, reakcio$csendes, alternative='greater',
  correct=FALSE, exact=FALSE, paired=TRUE)
```

Wilcoxon signed rank test

data: reakcio\$zajos and reakcio\$csendes

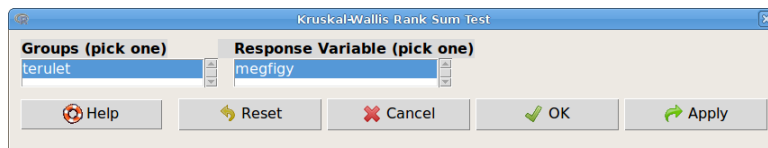
$V = 38.5$, p-value = 0.0289

alternative hypothesis: true location shift is greater than 0

(Tk. 234. o. 7.6.1. fejezet, 7.17 példa)

13.3. Több, független mintás Kruskal–Wallis-féle H-próba

Példánkban azt vizsgáljuk (Statistics → Nonparametric tests → Kruskal-Wallis test...), hogy négy terület mindegyikén 5-5 véletlenszerűen kiválasztott azonos méretű kvadrátban megszámlolt pipacsok alapján, van-e különbség a négy terület között a pipacsok gyakoriságát tekintve. (59. ábra, pipacs.csv). Ehhez meg kell adnunk a következőket:



59. ábra. Kruskal–Wallis-féle H-próba: Statistics → Nonparametric tests → Kruskal-Wallis test...

Groups (pick one) Csoportosító változó (faktor!)

Response variable (pick one) A vizsgálandó változó

A teszt outputjában megkapjuk a minta mediánokat, a Khi-négyzet statisztika (chi-squared) értékét a hozzá tartozó szabadsági fokkal (df) és a p -értéket (p-value).

```
> tapply(pipacs$megfigy, pipacs$terulet, median, na.rm=TRUE)
```

```
 1  2  3  4
14 28  8 48
```

```
kruskal.test(megfigy ~ terulet, data=pipacs)
```

```
      Kruskal-Wallis rank sum test
```

```
data:  megfigy by terulet
```

```
Kruskal-Wallis chi-squared = 11.483, df = 3, p-value = 0.009381
```

(Tk. 240. o. 7.6.3. fejezet, 7.19. példa)

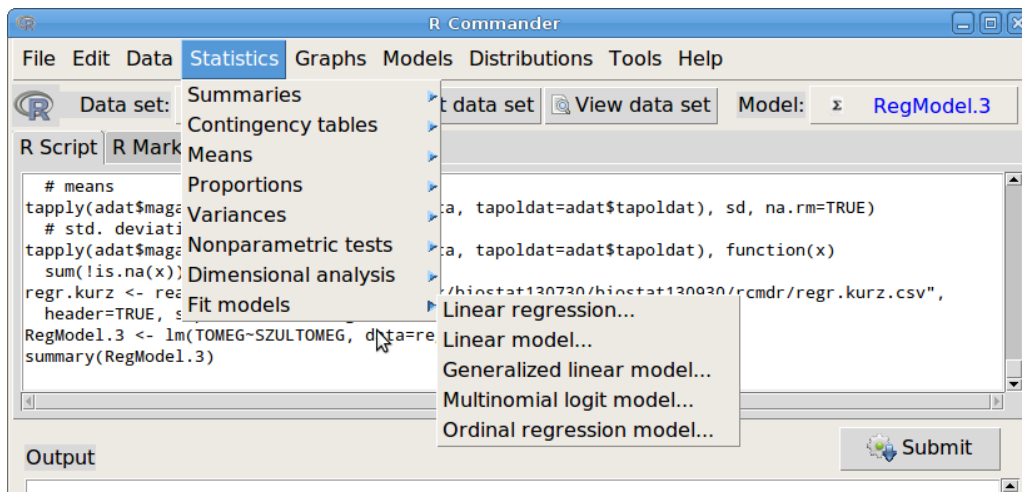
14. Modellek illesztése

Különböző statisztikai modelleket a Statistics → Fit models menüben találunk (60. ábra). Lehetőség van lineáris regresszió (Linear regression), általános lineáris modell (Linear model), általánosított lineáris modell (Generalized linear model), polinomiális logit modell (Multinomial logit model) és ordinális regressziós modell (Ordinal regression model) illesztésére. Itt most a lineáris regressziós és az általános lineáris modell illesztését mutatjuk be.

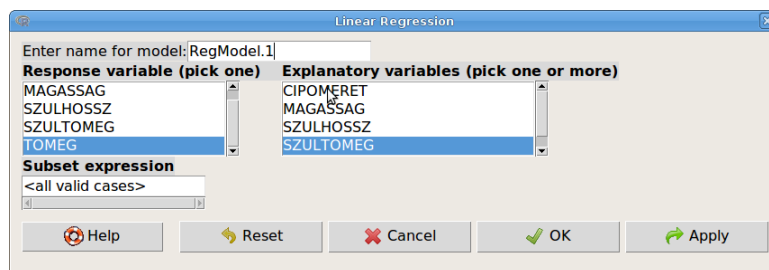
14.1. Regressziószámítás

Regressziós modelleket vagy a Statistics → Fit models → Linear regression... vagy a Statistics → Fit models → Linear models... párbeszédablakban állíthatunk be (60. ábra). Mindkettő ugyanazt az `lm()` függvényt hívja meg, de míg az előbbiben csak numerikus magyarázó változókat adhatunk meg, addig ez utóbbiban faktorok is lehetnek.

A regressziós kurzus példáján mutatjuk be a beállításokat (`regr.kurz.csv`). A testtömegszületéskori tömeg adatokra végezzük el a regressziós elemzést. Az elemzéshez meg kell adnunk a következőket (61. ábra).



60. ábra. Modellek illesztése: Statistics → Fit models



61. ábra. Lineáris regresszió: Statistics → Fit models → Linear regression...

Enter name of model: A modell elnevezése

Response variable (pick one) Függő változó

Explanatory variables (pick one or more) Magyarázó változó(k)

Subset expression Rész adattáblázatot definiáló logikai kifejezés (ld. 6.1. fejezet)

A teszt outputjában megkapjuk a modell összegzését (**summary**): a reziduumokra vonatkozó statisztikákat (**residuals**), valamint a paraméterbecsléseket (**Estimate**), a becslések standard hibáit (**Std. Error**) és a *t*-próbák eredményeit (*t*- és *p*-értékek). Az output utolsó részének első sorából kiderül, hogy mekkora a reziduumok szórása (**Residual standard error**), és mekkora a szabadsági foka (**degrees of freedom**). A második sorból olvashatjuk ki a determinációs együttható értékét (**Multiple R-squared**) és a korrigált determinációs együttható értékét (**Adjusted R-squared**) Az utolsó sorban a modell egészére vonatkozó *F*-próba eredményét láthatjuk.

```
> RegModel.1 <- lm(TOMEG~SZULTOMEG, data=regr.kurz)
> summary(RegModel.1)
```

Call:

```
lm(formula = TOMEG ~ SZULTOMEG, data = regr.kurz)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-14.0247  -7.1101  -0.7684   6.9753  18.4024
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
```

```

(Intercept) 33.099147 14.380794 2.302 0.0328 *
SZULTOMEG 0.009146 0.004029 2.270 0.0350 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.807 on 19 degrees of freedom
Multiple R-squared: 0.2134, Adjusted R-squared: 0.172
F-statistic: 5.153 on 1 and 19 DF, p-value: 0.03503

```

(Tk. 259. o. 9.4. fejezet, 9.2.-3. példa)

14.2. A lineáris modell

Lineáris modelleket a **Statistics** → **Fit models** → **Linear models** párbeszédablakban állíthatunk fel (62. ábra). A példánkban lepkék bábtömegét (**BABTOMEG**) modellezzük a hőmérsékleti (**HOM**), a táplálékellátottság (**TAP**) kezelések, valamint a kezdeti hernyótömeg (**TOMEG0**) függvényében. A párbeszédablakban a következőket kell beállítani:

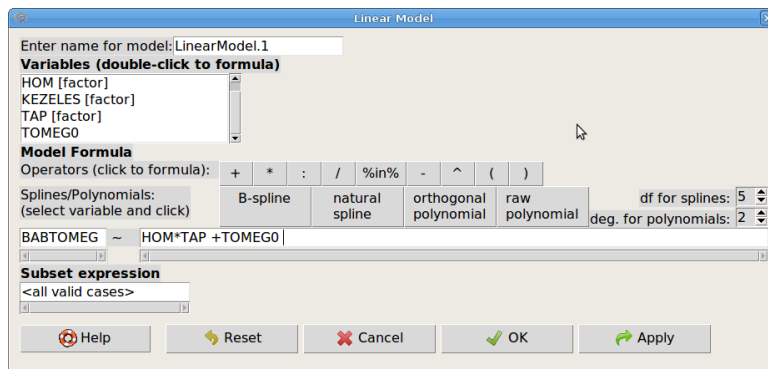
Enter name of model: A modell elnevezése

Variables (double-click to formula) Az adattábla változói. A faktorok mögé szögletes zárójelben bekerül a [factor] szó. Dupla egér kattintással a modell formulába helyezhetők

Model Formula A modell formula megadása (ld. alább)

Subset expression Rész adattáblázatot definiáló logikai kifejezés (ld. 6.1. fejezet)

A modell formulába a változókat és operátorokat be lehet kézzel is írni, de a változó lista adott elemére dupla kattintással, illetve a megfelelő operátor gombokra kattintással is beletelhetők. A bal oldalra kell a célváltozót írni. A jobb oldalra a magyarázó változókat a megfelelő operátorokkal összekapcsolva. Lehetőség van regressziós spline-ok és polinomiális tagok beillesztésére is (nem részletezzük).



62. ábra. Lineáris modell: Statistics → Fit models → Linear model...

Outputként a lineáris modell összegzését kapjuk meg. További modellekhez kapcsolódó műveletek (ANOVA-tábla, konfidencia intervallumok, diagnosztika stb.) a **Models** menüben található (16. fejezet).

```
> LinearModel.1 <- lm(BABTOMEG ~ HOM*TAP +TOMEGO, data=lepke)
> summary(LinearModel.1)
```

Call:

```
lm(formula = BABTOMEG ~ HOM * TAP + TOMEGO, data = lepke)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-0.081285 -0.013193  0.002068  0.013066  0.057270
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.302322	0.009497	31.834	< 2e-16 ***
HOM[T.melegített]	0.005363	0.012586	0.426	0.672
HOM[T.szobahom]	-0.003704	0.013808	-0.268	0.790
TAP[T.limitált]	-0.107056	0.014198	-7.540	9.7e-10 ***
TOMEGO	0.189542	0.319718	0.593	0.556
HOM[T.melegített]:TAP[T.limitált]	-0.014198	0.018371	-0.773	0.443
HOM[T.szobahom]:TAP[T.limitált]	0.010889	0.020064	0.543	0.590

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02898 on 49 degrees of freedom

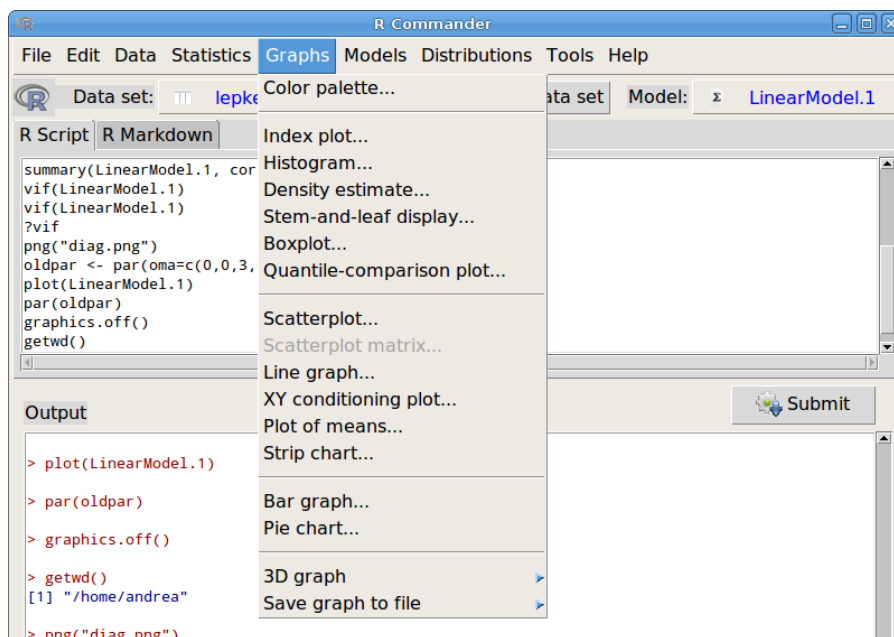
Multiple R-squared: 0.7977, Adjusted R-squared: 0.7729

F-statistic: 32.2 on 6 and 49 DF, p-value: 2.176e-15

(Tk. 352.-355. o. 11.1.-5. fejezet)

15. Grafikus lehetőségek

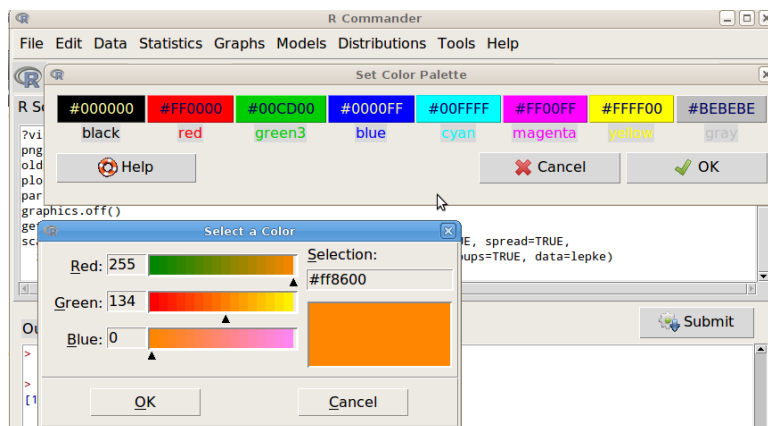
Néhány kivételtől eltekintve (modellekhez kapcsolódó ábrák: ld. 16. fejezet), az R Commander összes ábrakészítési lehetősége ebben a menüben található (63. ábra). Az ábrákat a Graphs → Save graph to file menüpontban található lehetőségekkel menthetjük.



63. ábra. Graphs menü: Graphs

15.1. Színpaletta beállítása

Lehetőség van az a grafikonoknál használt, előre definiált színek megváltoztatására, amelyek az adott színre kattintva, csúszkával állíthatók (64. ábra).



64. ábra. Színpaletta beállítása: Graphs → Color palette

15.2. Index ábra

Egy változó egyedi értékeinek grafikusán áttekintésére szolgál az index ábra (Graphs → Index plot). Először ki kell választanunk az ábrázolandó változót (Variable (pick one)), majd be kell állítani az ábra megjelenítési lehetőségeit az Options gombra kattintva (65. ábra). A értékek ábrázolási stílusán kívül beállíthatjuk, hogy automatikus azonosítson-e be néhány kiugró értéket, vagy az egér segítségével kézi úton történjen a beazonosítás. A beazonosításkor az értékek sorszáma rákerül az ábrára (egér bal gombja), illetve az azonosítás kikapcsolásakor (jobb egér gomb) az outputba. Ezenkívül beállíthatjuk az ábra feliratait. A 65. ábra szerinti beállításokkal készült a 66. ábra.

Plot options Ábra beállítások

Style of plot Az értékek ábrázolásának stílusa

Spikes Tüskék

Points Pontok

Identify points Pontok beazonosítása

Automatically Automatikus

Interactively with mouse Interaktívan egérrel

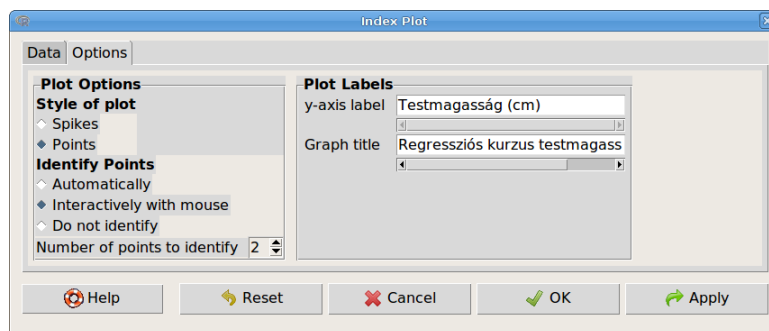
Do not identify Nincs pont beazonosítás

Number of points to identify Automatikusan beazonosítandó pontok száma

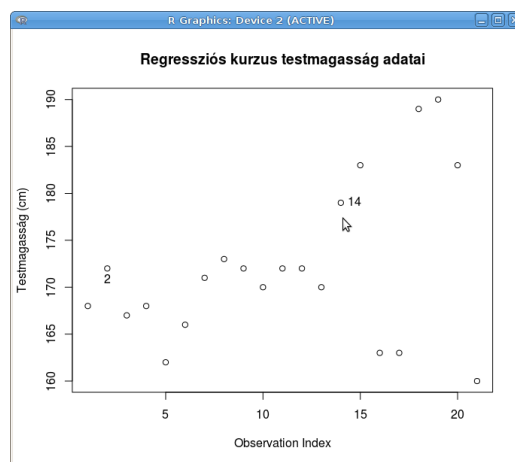
Plot labels Ábra feliratok

y-axis label y-tengely címe

Graph title Az ábra címe



65. ábra. Index ábra beállítása: Graphs → Index plot → Options



66. ábra. Testmagasság adatok index ábrája

15.3. Hisztogram

Egy változó hisztogramját készíthetjük el a **Graphs → Histogram** menüponttal. Először ki kell választanunk az ábrázolandó változót (**Variable (pick one)**), majd be kell állítani az ábra megjelenítési lehetőségeit az **Options** fülre kattintva (67. ábra). Beállíthatjuk az osztályintervallumok (**bins**) számát és azt, hogy mit ábrázoljon a program az y-tengelyen. Ezenkívül beállíthatjuk az ábra feliratait. A 67. ábra szerinti beállításokkal készült a 68. ábra.

Plot options Ábra beállítások (**Number of bins:**) Osztályintervallumok száma

Axis Scaling y-tengely skálája

Frequency counts Gyakoriságok

Percentages Százalékos arányok

Densities Sűrűségek

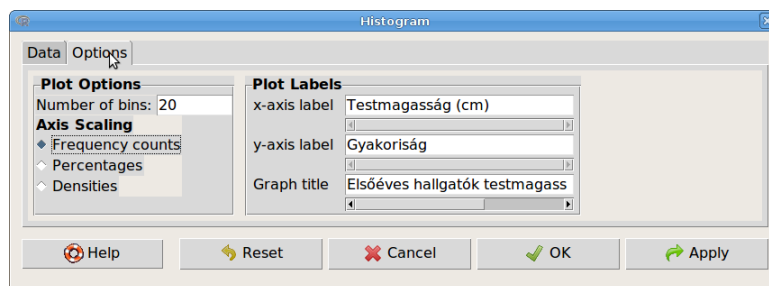
Plot labels Ábra feliratok

x-axis label x-tengely címe

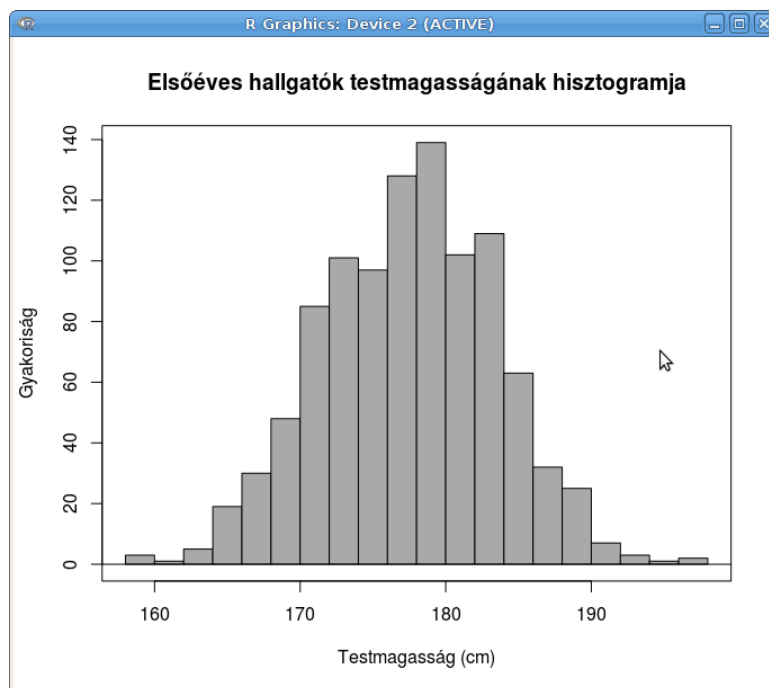
y-axis label y-tengely címe

Graph title Az ábra címe

(Tk. 93. o. 4.1.1. fejezet, 4.5-8. ábra)



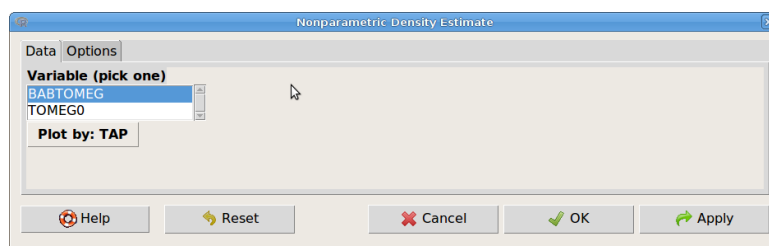
67. ábra. Hisztogram beállítása: Graphs → Histogram → Options



68. ábra. Testmagasság adatok hisztogramja (elsoeves.csv)

15.4. Simított hisztogram

Egy változó simított hisztogramját készíthetjük el a Graphs → Density estimate menüponttal több csoportra. Először ki kell választanunk az ábrázolandó változót (Variable (pick one)), illetve azt, hogy mely faktor szerinti csoportonként (Plot by groups) kérjük a sűrűségfüggvény becslését (69. ábra). Az Options fülre kattintva (70. ábra) be lehet állítani azt, hogy a sűrűség becslést milyen módszerrel végezze a program, valamint beállíthatjuk az ábra feliratait. A 70. ábra szerinti beállításokkal készült a 71. ábra.



69. ábra. Simított hisztogram változóinak beállítása: Graphs → Density estimate...

Kernel Functions Kernel függvények

Gaussian Gauss-féle

Epanechnikov Epanechnikov

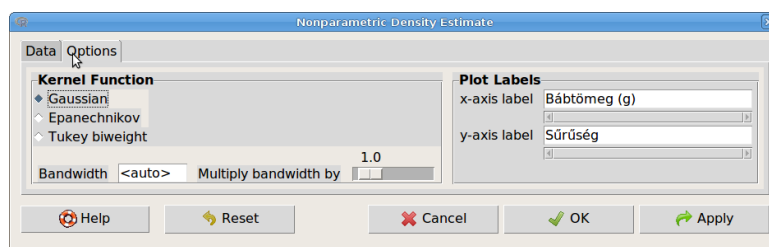
Tukey biweight Tukey biweight

Bandwidth A kernel szélessége

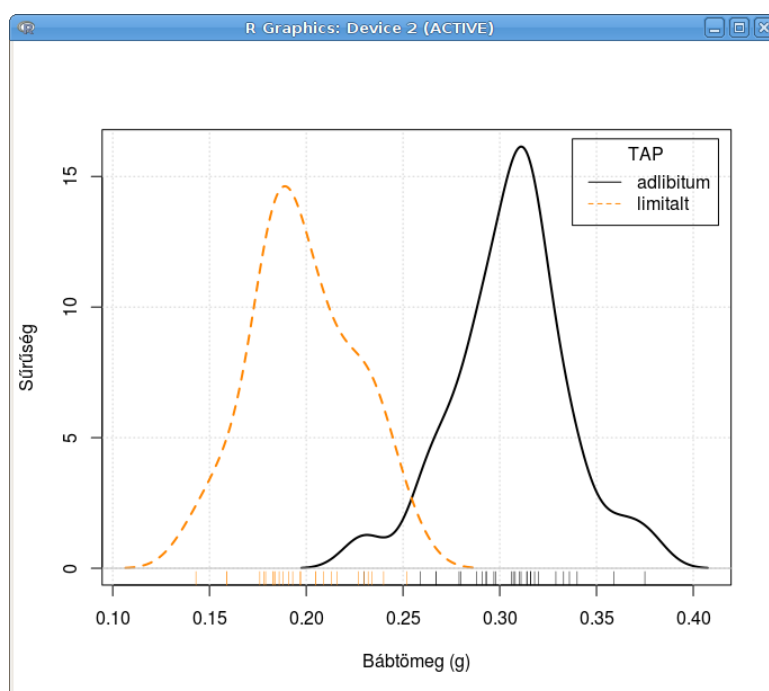
Plot labels Ábra feliratok

x-axis label x-tengely címe

y-axis label y-tengely címe



70. ábra. Simított hisztogram beállítása: Graphs → Density estimate... → Options



71. ábra. Bábtömeg adatok simított hisztogramja a két tápanyagellátottsági csoportban (lepke.txt)

(Tk. 95. o. 4.1.1. fejezet, 4.7.b. ábra)

15.5. Stem and leaf ábra

Korszerűbb ábrázolási módszerek miatt már nem nagyon használatos, nem tárgyaljuk.

15.6. Boxplot

Egy változó boxplotját készíthetjük el a **Graphs** → **Boxplot...** menüponttal több csoportra. Először ki kell választanunk az ábrázolandó változót (**Variable (pick one)**), illetve azt hogy mely faktor szerinti csoportonként (**Plot by groups**) kérjük a az ábrát. Az **Options** fülre kattintva (72. ábra) be lehet állítani azt, hogy a kiugró értékeket hogy azonosítsa a program, valamint beállíthatjuk az ábra feliratait. A 72. ábra szerinti beállításokkal készült a 73. ábra.

Identify Outliers Kiugró pontok beazonosítása

Automatically Automatikus

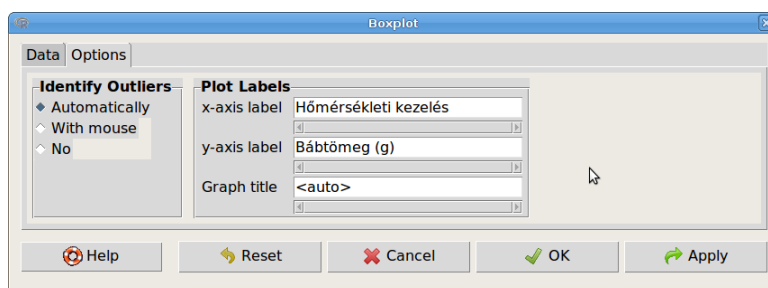
With mouse Interaktívan egérrel

No Nincs pont beazonosítás

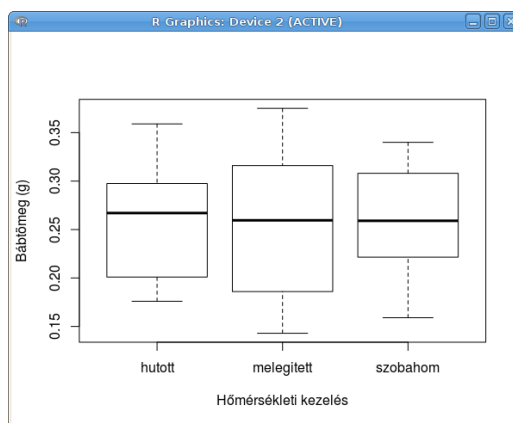
Plot labels Ábra feliratok

x-axis label x-tengely címe

y-axis label y-tengely címe



72. ábra. Boxplot beállítása: **Graphs** → **Boxplot...** → **Options**



73. ábra. Bábtömeg adatok boxplotjai a három hőmérséklet kezelési csoportban (`lepke.txt`)

(Tk. 97. o. 4.1.1. fejezet, 4.9.a. ábra, 342. o. 11.2. fejezet, 11.1. és 11.4. ábra)

15.7. QQ-ábra

Egy változó QQ-ábráját készíthetjük el a **Graphs** → **Quantile comparison plot...** menüponttal. Először ki kell választanunk az ábrázolandó változót (**Variable (pick one)**). Az **Options** fülre kattintva (74. ábra) be lehet állítani, hogy mely eloszláshoz hasonlítsa a beállított változó eloszlását (az eloszlások paramétereit meg kell adni), illetve azt, hogy pontokat hogyan azonosítsa a program. Ezenkívül beállíthatjuk az ábra feliratait. A 74. ábra szerinti beállításokkal

készült a 75. ábra. A program automatikusan rárajzolja az ábrára a pontok konfidencia sávokat (*confidence envelope*).

Plot options Ábra beállítások

Distribution Eloszlás

Normal Normális

t $df =$ szabadsági fokú *t*-eloszlás

Chi-square $df =$ szabadsági fokú Khi-négyzet-eloszlás

F Numerator $df =$ **Denominator** $df =$ A számláló (Numerator) és a nevező (Denominator) szabadsági fokának megfelelő *F*- eloszlás

Other Egyéb eloszlás

Identify Points Pontok beazonosítása

Automatically Automatikus

Interactively with mouse Interaktívan egérrel

Do not identify Nincs pont beazonosítás

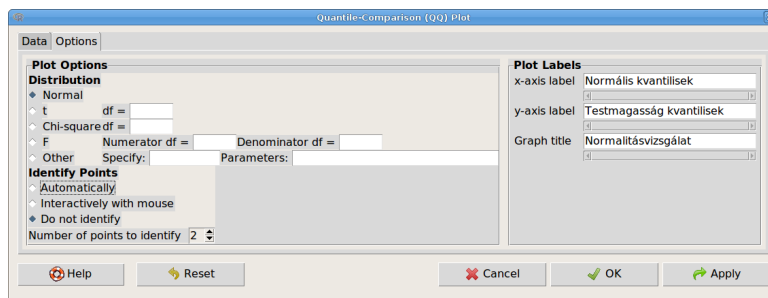
Number of points to identify Automatikusan beazonosítandó pontok száma

Plot labels Ábra feliratok

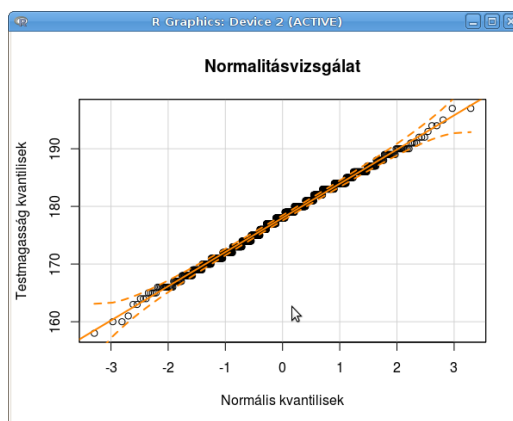
x-axis label x-tengely címe

y-axis label y-tengely címe

Graph title Az ábra címe

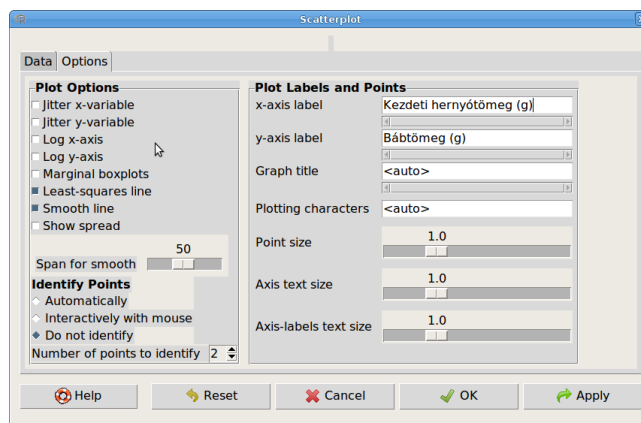


74. ábra. QQ-ábra beállítása: Graphs → Quantile comparison plot... → Options



75. ábra. Testmagasság változó normalitásvizsgálata QQ-ábrával (elsoeves.txt)

(Tk. 208. o. 7.3.1. fejezet)



76. ábra. A szórásdiagram beállítása: Graphs → Scatterplot... → Options

15.8. Szórásdiagram

Két numerikus változó szórásdiagramját készíthetjük el a Graphs → Scatterplot... menüponttal. Először ki kell választanunk az x-tengelyen (x-variable (pick one)) illetve az y-tengelyen ábrázolandó változókat, illetve azt hogy mely faktor szerinti csoportonként (Plot by groups) ábrázolja a pontokat, és megadható egy rész adattáblát leválogató logikai kifejezés is (ld. 6.1. fejezet). Az Options fülre kattintva (76. ábra) be lehet állítani, hogy a pontok egyszerű ábrázolásán túl milyen illesztéseket, simításokat tegyen rá a program az ábrára, illetve azt, hogy pontokat hogy azonosítsa a program. Ezenkívül beállíthatjuk az ábra feliratait, valamint a pontok és feliratok méretét. A 76. ábra szerinti beállításokkal készült a 77. ábra.

Plot options Ábra beállítások

Jitter x-variable x változó értékeinek szétszórása

Jitter y-variable y változó értékeinek szétszórása

Log x-axis x-tengely logaritmus skálázása

Log y-axis y-tengely logaritmus skálázása

Marginal boxplot A változók boxplot-jai a margókra

Least squares line Egyenes illesztés a legkisebb négyzetek módszerével

Smooth line Simítás

Show spread Az adatok szóródásának bemutatása

Span for smooth Simítási paraméter beállítása

Identify Points Pontok beazonosítása

Automatically Automatikus

Interactively with mouse Interaktívan egérrel

Do not identify Nincs pont beazonosítás

Number of points to identify Automatikusan beazonosítandó pontok száma

Plot labels and point Ábra feliratok és pontok beállítása

x-axis label x-tengely címe

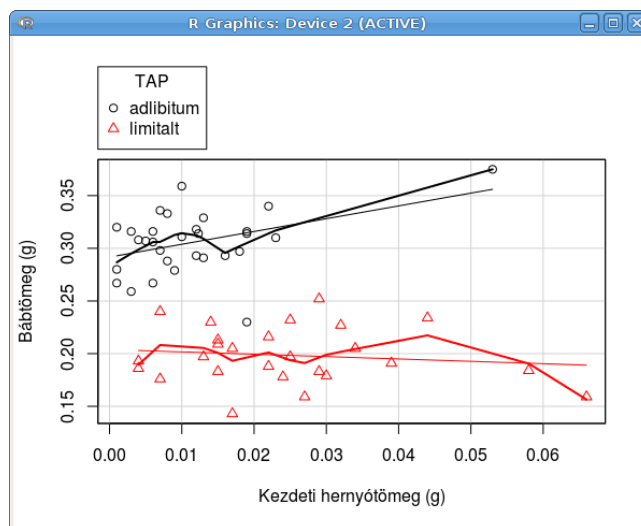
y-axis label y-tengely címe

Graph title Az ábra címe

Plotting characters Pontokat jelölő karakterek

Axis text size Tengely felirat mérete

Axis-labels text size Tengely címek mérete



77. ábra. Szórásdiagram csoportonkénti egyenes illesztéssel és simítással (lepke.txt)

(Tk. 248. o. 8. fejezet, 8.2. ábra, 345. o. 11.1.2. fejezet 11.6. ábra)

15.9. Szórásdiagram mátrix

Három vagy több numerikus változó szórásdiagram mátrixát készíthetjük el a **Graphs** → **Scatterplot matrix...** menüponttal. Először ki kell választanunk a változókat (**Select variables (three or more)**), illetve azt hogy mely faktor szerinti csoportonként (**Plot by groups**) ábrázolja a pontokat. Az **Options** fülre kattintva (78. ábra) be lehet állítani, hogy a fődiagonális elemek helyén mit mutasson, valamint hogy a pontok egyszerű ábrázolásán túl milyen illesztéseket, simításokat tegyen rá a program az ábrára, illetve azt, hogy hány pontot, és hogy azonosítsa a program. Ezenkívül beállíthatjuk az ábra címét. A 78. ábra szerinti beállításokkal készült a 79. ábra.

On Diagonal A diagonális tartalma

Density plots Simított hisztogramok

Histograms Hisztogramok

Boxplots Boxplotok

One-dimensional scatterplots Egydimenziós szórásdiagramok

Normal QQ plots Normális QQ-ábrák

Nothing (empty) Üres

Other Options Egyéb beállítások

Least squares line Egyenes illesztések a legkisebb négyzetek módszerével

Smooth line Simítások

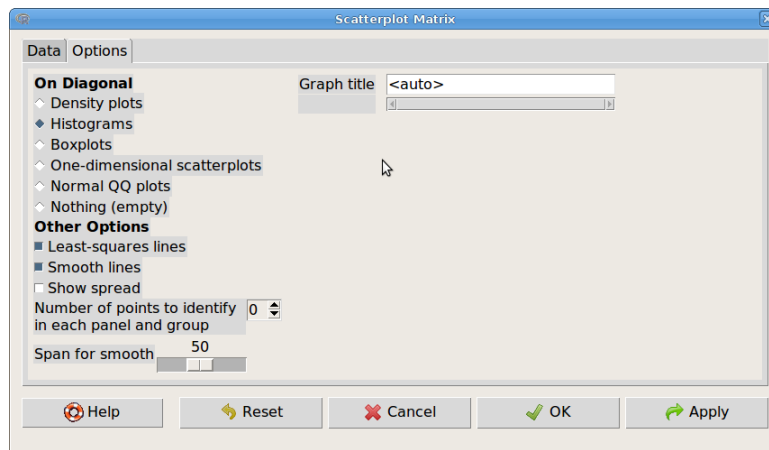
Show spread Az adatok szóródásának bemutatása

Number of points to identify Beazonosítandó pontok száma csoportonként és panelenként

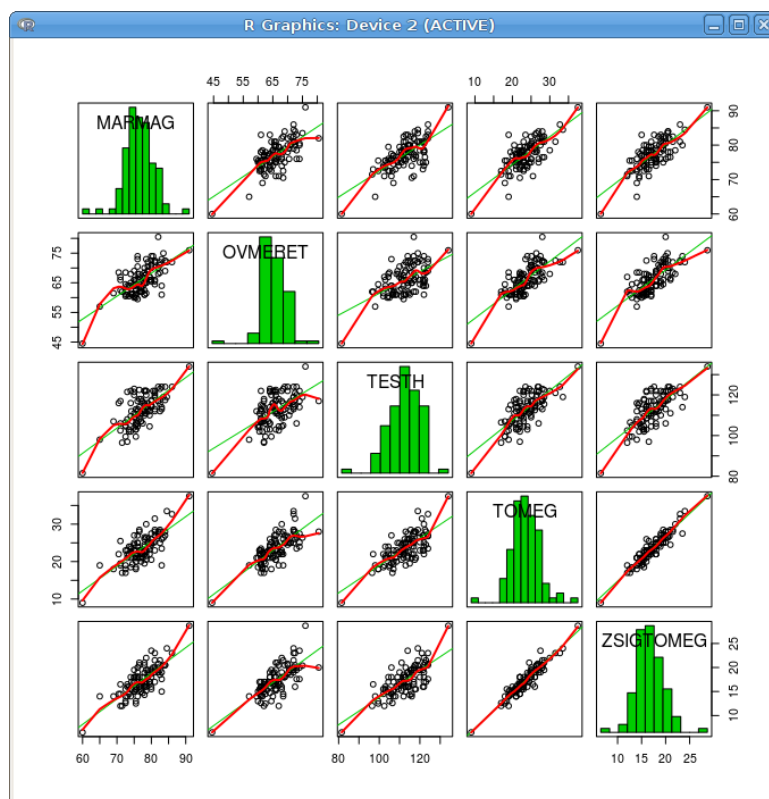
Span for smooth Simítási paraméter beállítása

Graph title Az ábra címe

(Tk. 277. o. 9.7. fejezet 9.10. ábra)



78. ábra. A szórásdiagram mátrix beállítása: Graphs → Scatterplot matrix... → Options



79. ábra. Őz testméretek szórásdiagram mátrixa egyenes illesztéssel és simítással (ozmeret.csv)

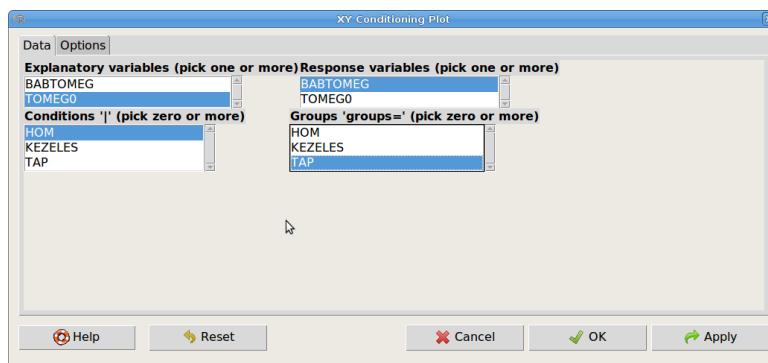
15.10. Vonaldiagram

Ha a táblázatban egymást követő értékek vannak, például egy idősor adatai, akkor célszerű lehet az egymást követő értékeket vonallal ábrázolni. Ezt lehet megtenni a vonaldiagrammal (Graphs → Line graph...). Használatához egyszerűen meg kell adni az X és Y változókat.

15.11. Feltételes szórásdiagram

Ha az adattáblázat eseteit egy vagy több faktor szerint csoportokba lehet sorolni (pl. kezelési csoportok), akkor célszerű lehet az egyes csoportokban a szórásdiagramokat egyedileg elkészíteni. A feltételes szórásdiagrammal több változó összefüggését páronként és csoportonként

külön-külön panelen lehet ábrázolni egy grafikus ablakban (Graphs → XY conditioning plot...). Először ki kell választanunk az x-tengelyen ábrázolandó magyarázó (Explanatory variables (pick one or more)), illetve az y-tengelyen ábrázolandó célváltozókat (Response variables (pick one or more)), illetve azt hogy mely faktor szerinti csoportonként (Conditions '|' (pick zero or more)) ábrázolja a pontokat külön-külön panelen, illetve, hogy egy panelen belül mely faktorok szerint ábrázolja más-más színnel a pontokat (Groups 'groups=' (pick zero or more)) (80. ábra). Az Options fülre kattintva (81. ábra) be lehet állítani a panelek elrendezését, a pontok típusát, valamint az ábra feliratait. A 81. ábra szerinti beállításokkal készült a 82. ábra.



80. ábra. A feltételes szórásdiagram változóinak beállítása: Graphs → XY conditioning plot...

Plot options Ábra beállítások

Automatically draw key Automatikus címkék

Different panels for different $x \sim y$ combinations külön panelek a különböző X és Y változó pároknak

Plot Type (one or both) A pontok típusa (egyik vagy mindkettő lehet)

Points Pontok

Lines Vonalak

X-axis Scales in Different Panels x-tengely skálája a különböző paneleken

Identical Azonos

Free Az adott panelen ábrázolt értékeknek megfelelő

Same range Ugyanaz a tartomány

Y-axis Scales in Different Panels y-tengely skálája a különböző paneleken

Identical Azonos

Free Az adott panelen ábrázolt értékeknek megfelelő

Same range Ugyanaz a tartomány

Layout Panelek elrendezése

number of columns oszlopok száma

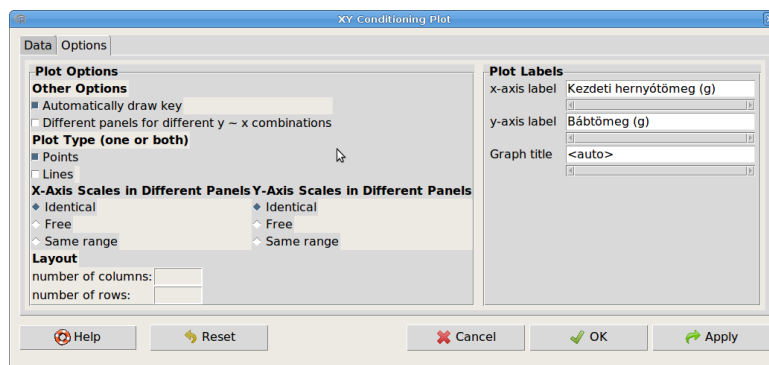
number of rows sorok száma

Plot Labels Ábra feliratok

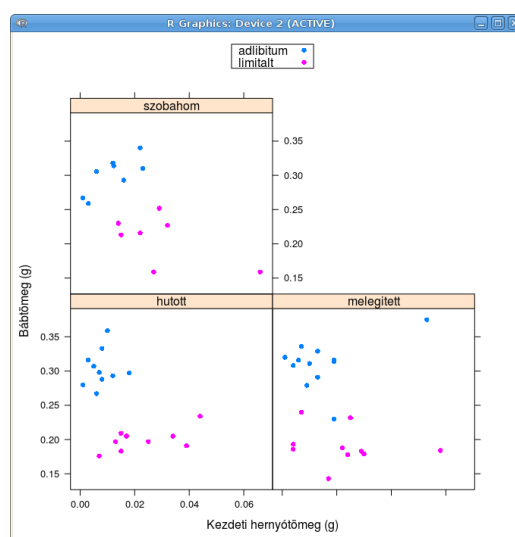
x-axis label x-tengely címe

y-axis label y-tengely címe

Graph title Az ábra címe



81. ábra. A feltételes szórásdiagram beállítása: Graphs → XY conditioning plot... → Options



82. ábra. Lepke bábtömeg-kezdeti tömeg kapcsolatok feltételes szórásdiagramon (lepke.txt)

15.12. Átlagok ábrája

Egy vagy két faktor szerint csoportosított adatokra ábrázolhatjuk az átlagokat kiegészítve különböző szóródást mutató hiba oszlopokkal (*error bars*). Először ki kell választanunk faktorokat (Factors (pick one or two)) illetve a célváltozót (Response Variable (pick one)). Az Options fülre kattintva (83. ábra) be lehet állítani a hibaoszlop típusát, valamint az ábra feliratait. A 83. ábra szerinti beállításokkal készült a 84. ábra.

Error bars Hiba oszlopok

Standard errors Standard hibák

Standard deviations Szórások

Confidence intervals **Level of confidence:** Konfidencia-intervallum, konfidencia-szint:

Plot Labels Ábra feliratok

x-axis label x-tengely címe

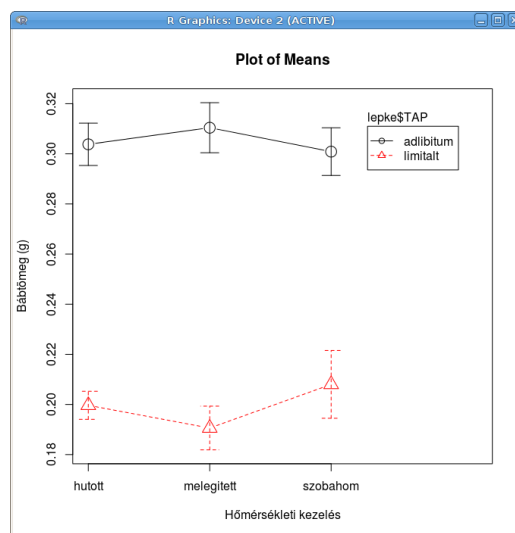
y-axis label y-tengely címe

Graph title Az ábra címe

(Tk. 104. o. 4.1. fejezet, 4.15. ábra; 343. o. 11.1.2. fejezet 11.2. ábra)



83. ábra. Átlagok ábrája beállítása: Graphs → Plot of means... → Options



84. ábra. Lepke bábtömeg átlagok standard hibákkal a kezelési csoportokban (lepke.txt)

15.13. Pontábra

Egy vagy két faktor szerint csoportosított adatokra ábrázolhatjuk az egyedi értékeket. Először ki kell választanunk faktorokat (Factors (pick one or two)) illetve a célváltozót (Response Variable (pick one)). Az Options fülre kattintva (85. ábra) be lehet állítani, hogy az azonos értékeket hogy ábrázolja, valamint az ábra feliratait. A 85. ábra szerinti beállításokkal készült a 86. ábra.

Duplicate values Azonos értékek

Stack Egymásra rajzolva

Jitter Megszórva

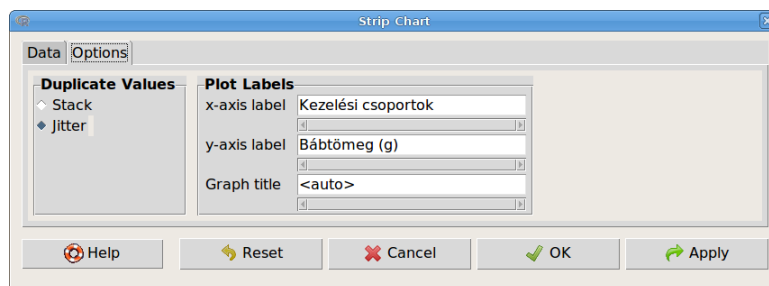
Plot Labels Ábra feliratok

x-axis label x-tengely címe

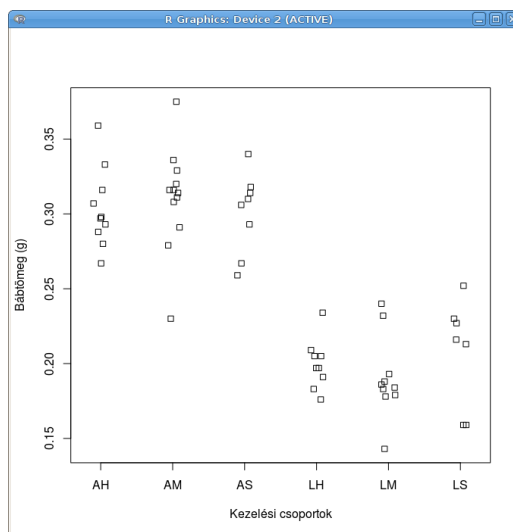
y-axis label y-tengely címe

Graph title Az ábra címe

(Tk. 92. o. 4.1. fejezet 4.4. ábra, 4.5.a. ábra)



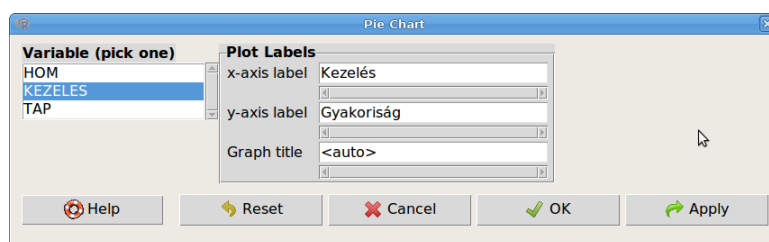
85. ábra. Pontábra beállítása: Graphs → Strip chart... → Options



86. ábra. Lepke bábtömegek a kezelési csoportokban (lepke.txt)

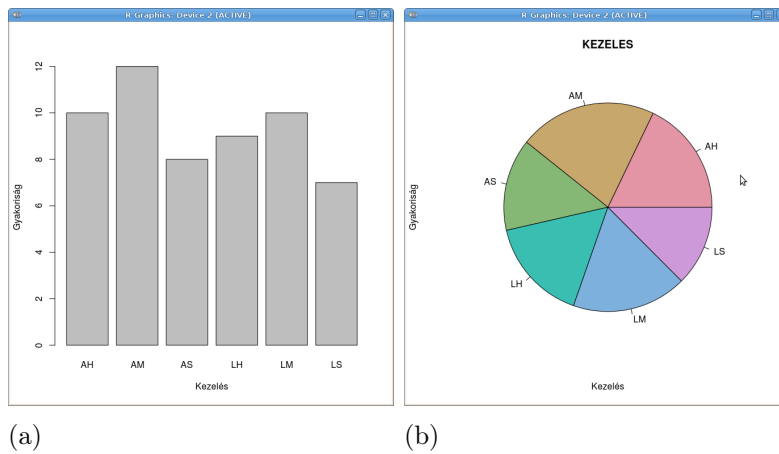
15.14. Oszlop- és kördiagram

Kategóriás változó gyakoriságainak ábrázolására szolgálnak (Bar graph..., Pie chart...). Először ki kell választani a faktort (Variables (pick one)), majd beállítani az ábra feliratait. A 87. ábra szerinti beállításokkal készült a 88. ábra.



87. ábra. Kördiagram beállítása: Graphs → Pie chart... → Options

(Tk. 91-92. o. 4.1. fejezet 4.1.-3. ábra)



88. ábra. Lepke gyakoriságok a kezelési csoportokban oszlop- és kördiagramon

15.15. 3D szórásdiagram

Három numerikus változó kapcsolatát vizsgálhatjuk ezzel az ábrával (Graphs → 3D scatterplot...). Meg kell adni két magyarázó változót (Explanatory variables (pick two)) és egy függőváltozót (Response variable (pick one)), valamint az esetleges csoportosító változót (Plot by groups). Ez utóbbi beállítása esetén meg kell adni, hogy az illesztendő regressziós síkok párhuzamosak legyenek-e (Parallel regression surfaces). Az Options fülre kattintva (89. ábra) be lehet állítani a következőket:

Show axis scales Mutassa-e a tengelyek skáláit

Show surface gridline Mutassa-e az illesztett felület rácsvonalait

Show squared residuals Mutassa-e négyzet alakban a reziduum négyzetösszegeket

Surface to Fit Az illesztendő felület

Linear least-squares Sík

Quadratic least-squares Kvadratikus legkisebb négyzetes

Smooth regression Simított felület

Additive regression Additív modellel illesztett

Plot 50% concentration ellipsoid Az adatok 50%-át tartalmazó ellipszoid kirajzolása

Background colors Háttérszínek

Black Fekete

White Fehér

Identify Points Pontok beazonosítása

Automatically Automatikus

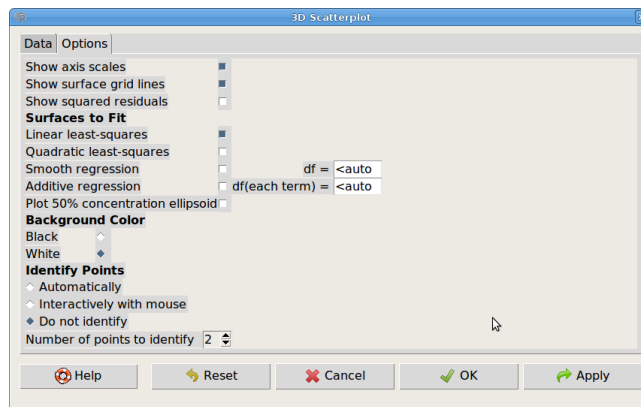
Interactively with mouse Interaktívan egérrel

Do not identify Nincs pont beazonosítás

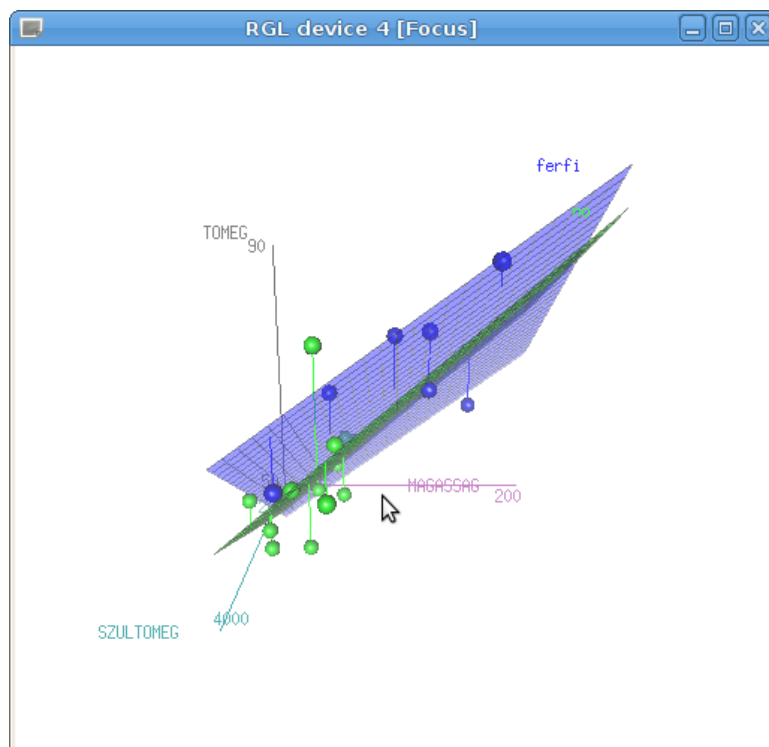
Number of points to identify Automatikusan beazonosítandó pontok száma

A 89. ábra szerinti beállításokkal készült a 90. ábra. Az elkészült ábra forgatható. A Graphs → 3D scatterplot... menü Save graph to file menüpontjával file-ba menthető.

(Tk. 273. o. 9.7. fejezet 9.8. ábra)



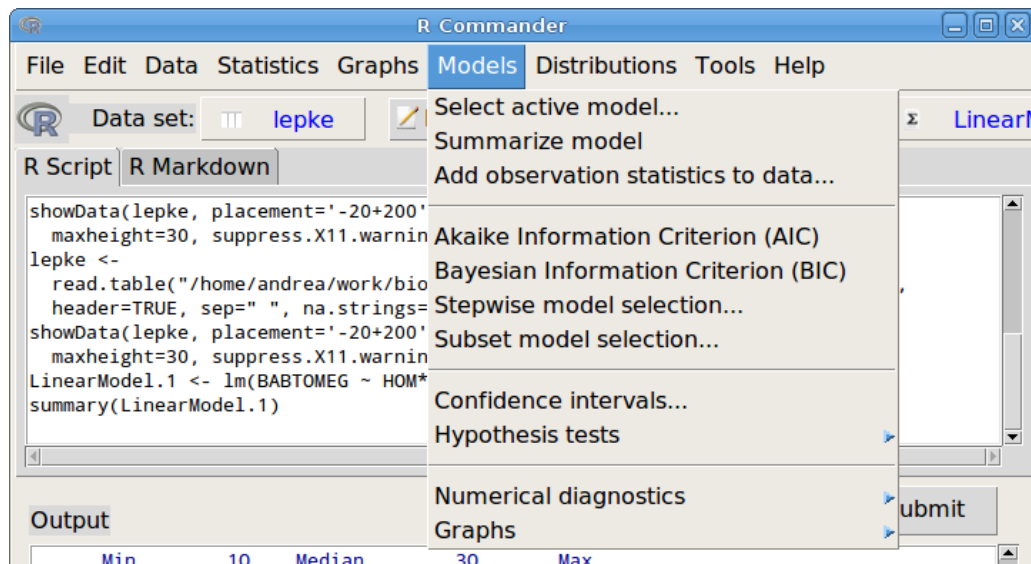
89. ábra. 3D szórásdiagram beállítása: Graphs → 3D scatterplot... → Options



90. ábra. Regressziós kurzus testtömeg adatai a születéskori tömeg és a magasság függvényében (regr.kurz.csv)

15.16. Ábrák mentése

A Graphs → Save graph to file menüponttal lehet a már elkészült, grafikus ablakban megnyitott ábrákat különböző formátumokba menteni. Windows operációs rendszer esetén, az elkészült ábrára kattintva a jobb egérgombbal, szintén menthetjük azt különböző formátumokban.



91. ábra. Models menü: Models

16. Modellekkel kapcsolatos műveletek

Az éppen aktív modellel kapcsolatos műveleteket a **Models** menüben találunk (91. ábra). Mindig csak az aktuális modell típusnak megfelelő menüpontok elérhetők. Ebben a fejezetben a lineáris modellekhez kapcsolódó fontosabb lehetőségeket részletesen bemutatjuk, a többit csak felsorolásszerűen.

16.1. Az aktív modell kiválasztása

Select active model... Az aktív modell általában az utoljára készített modell, de átállíthatjuk ezzel lehetőséggel, vagy az eszköztár jobb szélső – az éppen aktív modell nevét mutató – gombjának megnyomásával.

16.2. A modell összegzése

Summarize model Az aktív modell összegzése: `summary()` (ld. 14.1. fejezet output része).

(Tk. 262. o. 9.4.2. fejezet 9.3. példa)

16.3. Modellből számított statisztikák

Add observation statistics to data... Az adattáblához, új változóként hozzáfűzhetjük a következő, az aktív modellel számított értékeket (92. ábra):

Fitted values Becsült értékek

Residuals Reziduumok

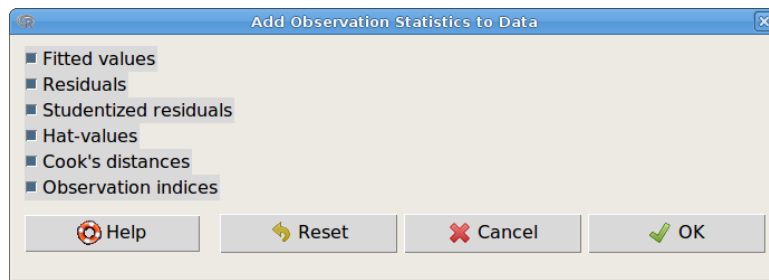
Studentized residuals Studentizált (jackknife) reziduumok

Hat-values Hatóerő értékek

Cook's distances Cook-féle távolságok

Observation indices A megfigyelés sorszám

(Tk. 287. o. 9.10.3. fejezet)



92. ábra. Modellből számított értékek hozzáfűzése az adattáblázathoz: Models → Add observation statistics to data...

16.4. Információs kritériumok

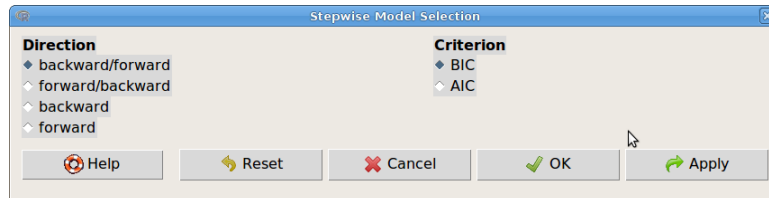
Akaike Information Criterion (AIC) A modell AIC értéke

Bayesian Information Criterion (AIC) A modell BIC értéke

(Tk. 373. o. 11.7.3. fejezet 11.13. példa)

16.5. Lépésenkénti modellszelkeció

Lépésenkénti modellszelkeció (**Stepwise model selection**). Megadandó a szelekció iránya (**Direction**) és az információs kritérium, ami alapján történik a szelekció (**Criterion**) (93. ábra). Példánkban először létrehozzuk a következő modellt: $BABTOMEG \sim TAP * HOM * TOMEG0$. Majd elvégezzük a modellszelekciót.



93. ábra. Modellből számított értékek hozzáadása az adattáblázathoz: Models → Stepwise model selection...

Célszerű a futtatás után a szkript ablakban egy új modellt létrehozni (`redmod`) szelekció eredményeként, amellyel a továbbiakban dolgozhatunk:

```
> LinearModel1.2 <- lm(BABTOMEG ~ TAP*HOM*TOMEG0, data=lepke)
> stepwise(LinearModel1.2, direction='backward/forward',
  criterion='AIC')
> redmod=stepwise(LinearModel1.2, direction='backward/forward',
  criterion='AIC')
```

A redukált modell összegzése (**Summarize model**):

```
> summary(redmod, cor=FALSE)
```

Call:

```
lm(formula = BABTOMEG ~ TAP + TOMEG0 + TAP:TOMEG0, data = lepke)
```

Residuals:

```
      Min          1Q      Median          3Q          Max
```

-0.084718 -0.016629 -0.001001 0.017550 0.055207

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.291655	0.007637	38.190	< 2e-16	***
TAP[T.limitalt]	-0.087752	0.012789	-6.862	8.23e-09	***
TOMEGO	1.213812	0.502142	2.417	0.0192	*
TAP[T.limitalt]:TOMEGO	-1.438542	0.617372	-2.330	0.0237	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02724 on 52 degrees of freedom
Multiple R-squared: 0.8104, Adjusted R-squared: 0.7994
F-statistic: 74.07 on 3 and 52 DF, p-value: < 2.2e-16

(Tk. 375. o. 11.8. fejezet 11.14. példa)

16.6. Részmodell kiválasztása

A Subset model selection Egy olyan ábrát produkál, amelyen különböző paraméterszámú részmodelleket lehet összehasonlítani.

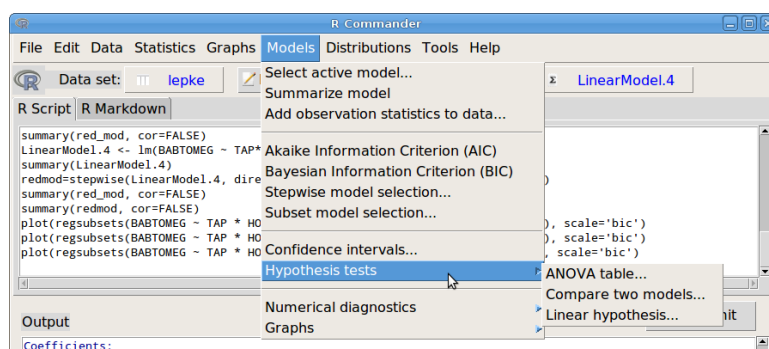
16.7. Konfidencia-intervallumok

Az adott modell együtthatóira számol konfidencia intervallumokat (Models → Confidence intervals...). A konfidencia szint állítható (**Confidence level**).

(Tk. 263. o. 9.4.3. fejezet, 367. o. 11.5.5. fejezet 11.10. példa)

16.8. Hipotézis vizsgálatok

A modellel kapcsolatos hipotézisvizsgálatok (94. ábra).

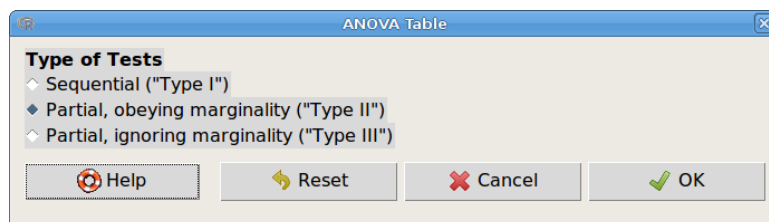


94. ábra. Hipotézisvizsgálatok: Models → Hypothesis tests

16.8.1. ANOVA-tábla

A modell ANOVA-táblázatát írja ki. Meg lehet adni, hogy melyik négyzetösszeggel számoljon (95. ábra). (Megjegyzés: Nem az R `anova()` függvényét hívja meg, hanem a `car` csomag `Anova()` függvényét!).

A modellszelekcióval redukált modell ANOVA-táblája II-es típusú négyzetösszeggel számolva:



95. ábra. Négyzetösszeg típusának kiválasztása: Models → Hypothesis tests → ANOVA table

```
> Anova(redmod, type="II")
```

Anova Table (Type II tests)

Response: BABTOMEG

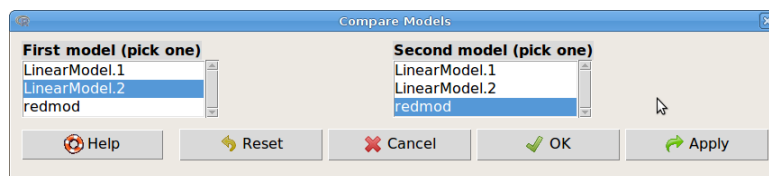
	Sum Sq	Df	F value	Pr(>F)
TAP	0.134792	1	181.7155	< 2e-16 ***
TOMEGO	0.000597	1	0.8053	0.37366
TAP:TOMEGO	0.004027	1	5.4294	0.02372 *
Residuals	0.038572	52		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Tk. 364. o. 11.5.3. fejezet 11.8. példa)

16.8.2. Modellek összehasonlítása

Két beágyazott modell összehasonlítása. Ki kell választani a megjelenő párbeszédablakban az összehasonlítandó modelleket (96. ábra).



96. ábra. Modellek összehasonlítása: Models → Hypothesis tests → Compare two models...

```
> anova(LinearModel.2, redmod)
```

Analysis of Variance Table

Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Model 1: BABTOMEG ~ HOM * TAP * TOMEGO	44	0.032874				
Model 2: BABTOMEG ~ TAP + TOMEGO + TAP:TOMEGO	52	0.038572	-8	-0.0056984	0.9534	0.4839

(Tk. 364. o. 11.5.3. fejezet 11.7. példa, 376. o. 11.14. példa)

16.8.3. Lineáris hipotézisek tesztelése

Ebben a pontban van lehetőség arra, hogy a modellben kontrasztokat (a modellparaméterekből képzett lineáris kombináció) teszteljünk.

(Tk. 382. o. 11.11. fejezet)

16.9. Diagnosztikai eszközök

Lehetőség van numerikus és grafikus modelldiagnosztikák készítésére. Ezek közül csak a legáltalánosabban használtakat részletezzük.

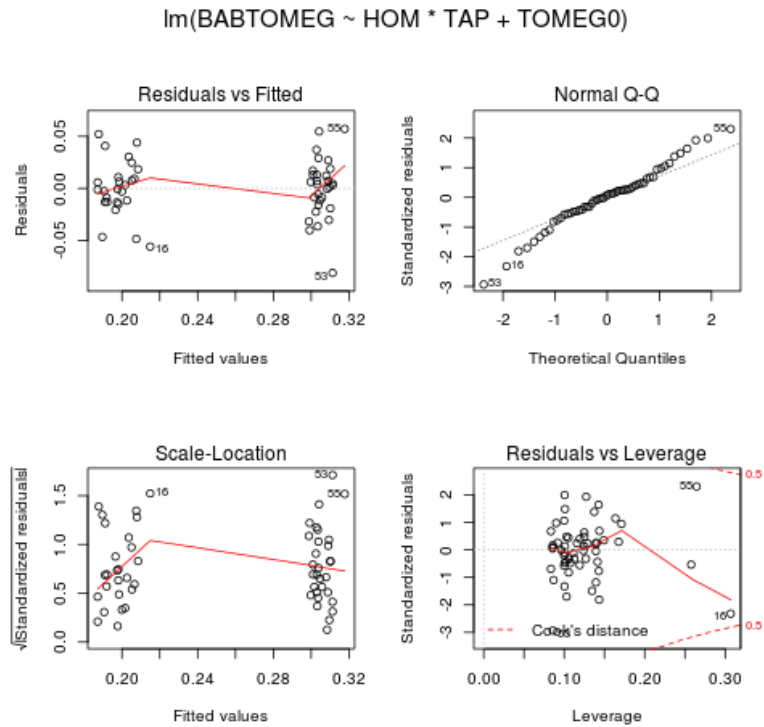
16.9.1. Numerikus diagnosztikák

A modell változóinak általánosított variancia infláció faktorát a Models → Numerical diagnostics → Variance-inflation factors választásával számíttathatjuk ki. A többi lehetőséggel tesztelhetjük a hibátág szórásának állandóságát, az autokorreláltságot, linearitást, illetve kiugró értékeket.

(Tk. 278. o. 9.9. fejezet 9.11. példa)

16.9.2. Grafikus diagnosztikák

A szokásos diagnosztikus ábrákat a Models → Graphs → Basic diagnostic plots pontban találjuk (97. ábra).

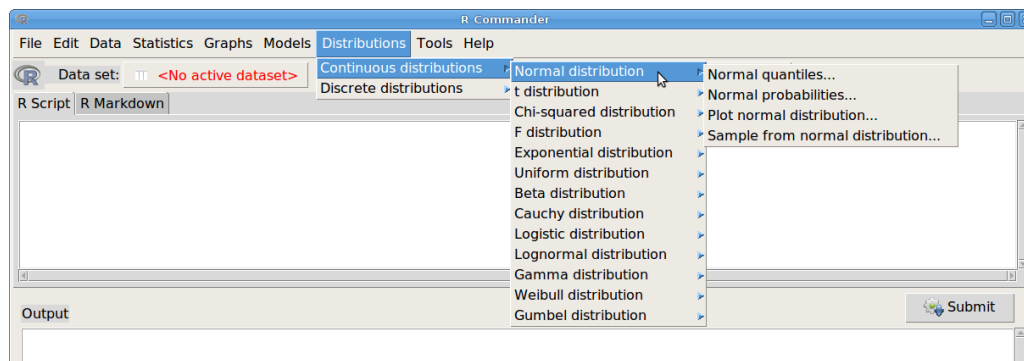


97. ábra. Diagnosztikus ábrák: Models → Graphs → Basic diagnostic plots

(Tk. 295. o. 9.10.4. fejezet 9.20. ábra, 368. o. 11.6. fejezet 11.11. példa)

17. Eloszlások

Ezzel a menüvel (Distributions) lehet folytonos (Continuous distributions) illetve diszkrét eloszlásokból (Discret distributions) valószínűségeket és kvantiliseket (p -kvantilis: a valószínűségi változó azon értéke, amelynél kisebb értékek hányada p) számoltatni, illetve kirajzoltatni a megfelelő eloszlást, sűrűségfüggvényt, eloszlásfüggvényt, illetve adott eloszlású véletlen számokat generálni. Az egyes eloszlások esetén nagyon hasonló a működése, ezért csak egy-egy példát mutatunk folytonos (normális), illetve diszkrét (binomiális) eloszlásra.



98. ábra. Normális eloszlás menü: Distributions → Continuous distributions → Normal distribution

17.1. Folytonos eloszlás: normális

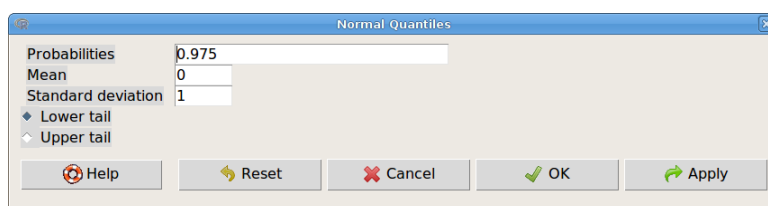
Normal quantiles... Normális eloszlás kvantilisei

Normal probabilities... Valószínűségek normális eloszlásból

Plot normal distribution... Normális eloszlás ábrázolása

Sample from normal distribution... Mintavétel normális eloszlásból

17.1.1. Adott valószínűséghez tartozó kvantilisek meghatározása



99. ábra. Normális eloszlás kvantilisének meghatározása: Distributions → Continuous distributions → Normal distribution → Normal quantiles

A kvantilisek meghatározásához a következőket kell megadni:

Probabilities Valószínűségek (vesszővel elválasztva)

Mean A normális eloszlás átlaga

Standard deviation A normális eloszlás szórása

Lower tail Az eloszlás alsó széle

Upper tail Az eloszlás felső széle

```
>qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
```

```
[1] 1.959964
```

17.1.2. Adott kvantilisekhez tartozó valószínűségek meghatározása

A valószínűségek meghatározásához a következőket kell megadni:

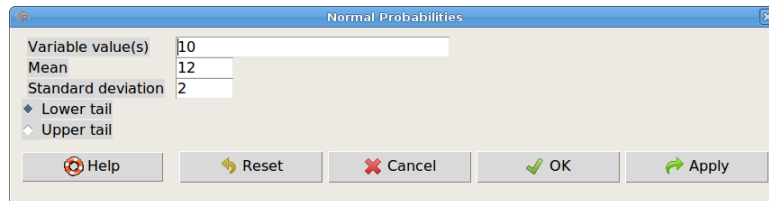
Variable value(s) A változó értékei (kvantilisek), vesszővel elválasztva

Mean A normális eloszlás átlaga

Standard deviation A normális eloszlás szórása

Lower tail Az eloszlás alsó széle

Upper tail Az eloszlás felső széle

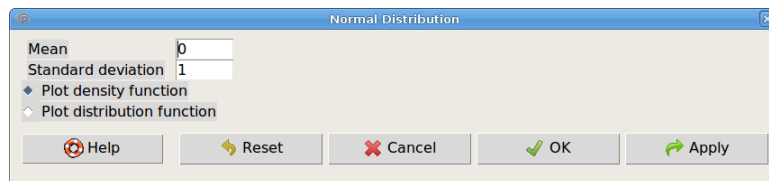


100. ábra. Valószínűség meghatározása adott kvantilishez: Distributions → Continuous distributions → Normal distribution → Normal probabilities

```
> pnorm(c(10), mean=12, sd=2, lower.tail=TRUE)
```

```
[1] 0.1586553
```

17.1.3. Sűrűség-, illetve eloszlásfüggvény ábrázolása



101. ábra. Sűrűség-, illetve eloszlásfüggvény ábrázolása: Distributions → Continuous distributions → Normal distribution → Plot normal distribution

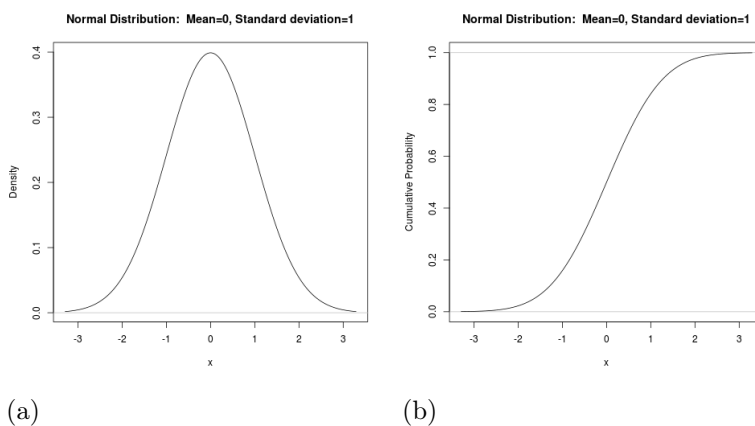
A függvények ábrázolásához a következőket kell megadni:

Mean A normális eloszlás átlaga

Standard deviation A normális eloszlás szórása

Plot density function Sűrűségfüggvény

Plot distribution function Eloszlásfüggvény



102. ábra. Normális eloszlás sűrűség- (a) és eloszlásfüggvénye (b)

17.1.4. Mintavétel

Ezzel a lehetőséggel véletlen normális eloszlású mintákat készíthetünk (Distributions → Continuous distributions → Normal distribution → Sample from normal distribution). A mintaelemeket a mintaátlagokkal, szórásokkal, és a minták összegével együtt (ha bejelöljük) egy adattáblázatba írja bele a program, melyet menthetünk 6.6. fejezet).

Megadandó:

Enter name of data set: Adattáblázat neve

Mean A normális eloszlás átlaga

Standard deviation A normális eloszlás szórása

Number of samples (rows) Minták (sorok) száma

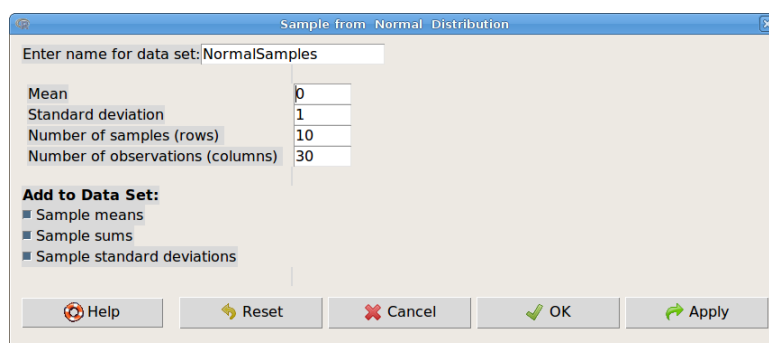
Number of observations (columns) Mintaelemek (oszlopok) száma mintánként

Add to Data Set Adattáblázatba kiírandó

Sample means Mintaátlagok

Sample sums Minta összegek

Sample standard deviations Minta szórások



103. ábra. Mintavétel normális eloszlásból: Distributions → Continuous distributions → Normal distribution → Sample from normal distribution

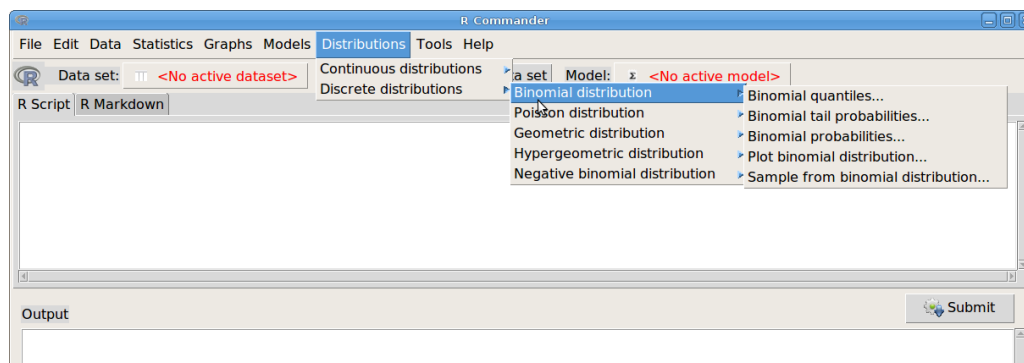
(Tk. 81.o. 3.5.4. fejezet 3.10. példa)

	bs28	obs29	obs30	mean	sum	sd
sample1	5974	-0.10025217	-0.86441835	-0.16091997	-4.8275990	0.8391831
sample2	5591	-1.27627444	-1.57122041	-0.42404978	-12.7214934	1.1822619
sample3	4907	1.03519542	-1.06980753	0.14623485	4.3870455	0.9903460
sample4	8763	0.48422733	-1.69341245	-0.26622665	-7.9867995	0.8240819
sample5	7229	0.96515576	-0.01753702	0.18637800	5.5913399	1.2626151
sample6	9572	-0.38972844	0.10177908	-0.15057803	-4.5171109	0.7415329
sample7	3669	-1.64407956	-0.81688385	0.11069913	3.3209738	1.0305830
sample8	8728	-0.63430423	-0.37422431	-0.01291352	-0.3874056	1.0129759
sample9	9345	-0.08308977	0.03849314	-0.32603863	-9.7811589	1.0014557

104. ábra. Minták normális eloszlásból

17.2. Diszkrét eloszlás: binomiális

A diszkrét eloszlások közül a – talán leggyakrabban használt – binomiális eloszlással kapcsolatos műveleteket mutatjuk be (105. ábra).



105. ábra. Binomiális eloszlás menü: Distributions → Discrete distributions → Binomial distribution

Binomial quantiles... Binomiális eloszlás kvantilisei

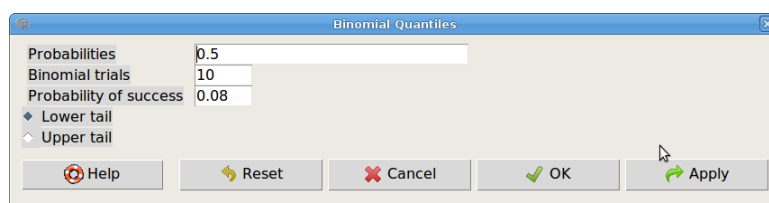
Binomial tail probabilities... Széli valószínűségek binomiális eloszlásból

Binomial probabilities... Valószínűségek binomiális eloszlásból

Plot binomial distribution... Binomiális eloszlás ábrázolása

Sample from binomial distribution... Mintavétel binomiális eloszlásból

17.2.1. Adott valószínűségekhez tartozó kvantilisek meghatározása



106. ábra. Binomiális eloszlás kvantilise: Distributions → Discrete distributions → Binomial distribution → Binomial quantiles

A kvantilis meghatározásához a következőket kell megadni:

Probabilities Valószínűségek (vesszővel elválasztva)

Binomial trials Kísérletek száma

Probability of success A bekövetkezés valószínűsége

Lower tail Az eloszlás alsó széle

Upper tail Az eloszlás felső széle

```
>qbinom(c(0.5), size=10, prob=0.08, lower.tail=TRUE)
```

```
[1] 1
```

17.2.2. Adott kvantilisekhez tartozó széli valószínűségek meghatározása

A valószínűségek meghatározásához a következőket kell megadni:

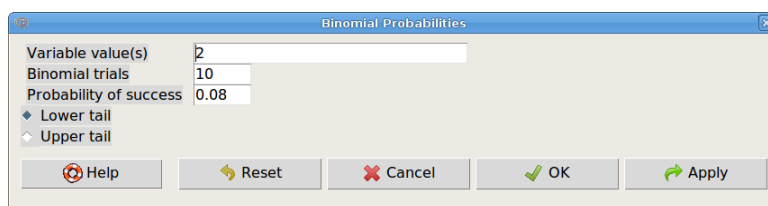
Variable value(s) A változó értékei (kvantilisek), vesszővel elválasztva

Binomial trials Kísérletek száma

Probability of success A bekövetkezés valószínűsége

Lower tail Az eloszlás alsó széle

Upper tail Az eloszlás felső széle



107. ábra. Széli valószínűség meghatározása adott kvantilishez: Distributions → Discrete distributions → Binomial distribution → Binomial tail probabilities

```
> pbinom(c(2), size=10, prob=0.08, lower.tail=TRUE)
```

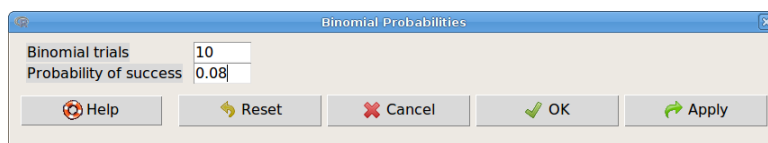
```
[1] 0.9599246
```

17.2.3. Adott kvantilisekhez tartozó valószínűségek meghatározása

A valószínűség meghatározásához a következőket kell megadni:

Binomial trials Kísérletek száma

Probability of success A bekövetkezés valószínűsége



108. ábra. Valószínűségek meghatározása adott kvantilisekhez: Distributions → Discrete distributions → Binomial distribution → Binomial probabilities

```

> .Table <- data.frame(Pr=dbinom(0:10, size=10, prob=0.08))
> rownames(.Table) <- 0:10
> .Table

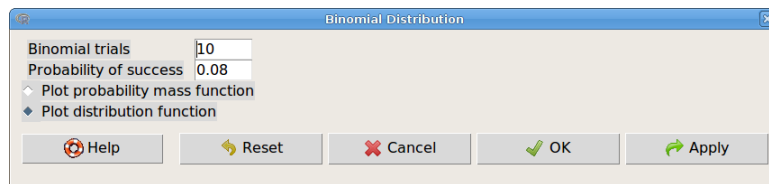
```

```

Pr
0 4.343885e-01
1 3.777291e-01
2 1.478070e-01
3 3.427410e-02
4 5.215623e-03
5 5.442389e-04
6 3.943760e-05
7 1.959633e-06
8 6.390106e-08
9 1.234803e-09
10 1.073742e-11

```

17.2.4. Eloszlás ábrázolása



109. ábra. Binomiális eloszlás ábrázolása: Distributions → Discrete distributions → Binomial distribution → Plot binomial distribution

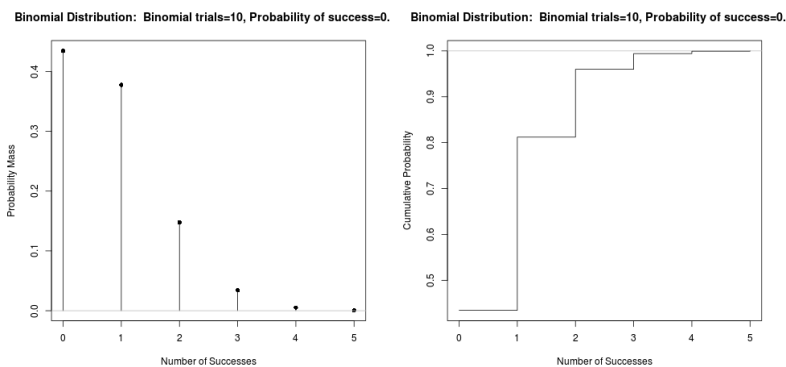
A függvények ábrázolásához a következőket kell megadni:

Binomial trials Kísérletek száma

Probability of success A bekövetkezés valószínűsége

Plot probability mass function Eloszlás

Plot distribution function Eloszlásfüggvény



(a)

(b)

110. ábra. Binomiális eloszlás (a) és eloszlásfüggvény (b)

17.2.5. Mintavétel

Ezzel a lehetőséggel véletlen binomiális eloszlású mintákat készíthetünk. A mintaelemeket a mintaátlagokkal, szórásokkal, és a minta összegekkel együtt (ha bejelöljük) egy adattáblázatba írja bele a program, melyet menthetünk 6.6. fejezet).

Megadandó:

Enter name of data set: Adattáblázat neve

Binomial trials Kísérletek száma

Probability of success A bekövetkezés valószínűsége

Number of samples (rows) Minták (sorok) száma

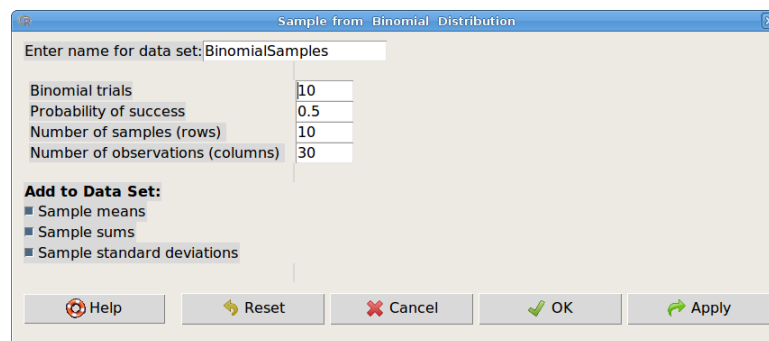
Number of observations (columns) Mintaelemek (oszlopok) száma mintánként

Add to Data Set Adattáblázatba kiírandó

Sample means Mintaátlagok

Sample sums Minta összegek

Sample standard deviations Minta szórások



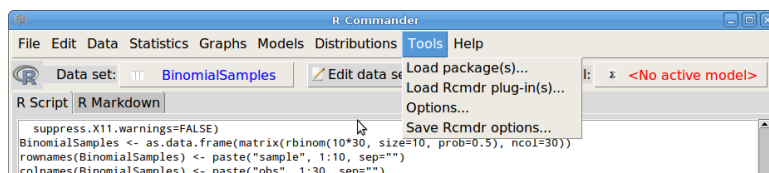
111. ábra. Mintavétel binomiális eloszlásból: Distributions → Discrete distributions → Binomial distribution → Sample from binomial distribution

	obs21	obs22	obs23	obs24	obs25	obs26	obs27	obs28	obs29	obs30	mean	sum	sd
sample1	5	7	4	5	2	3	6	6	5	4	4.433333	133	1.568732
sample2	7	4	4	5	4	5	4	5	5	6	5.433333	163	1.222866
sample3	3	5	6	7	5	7	5	6	3	5	5.200000	156	1.845778
sample4	4	5	5	4	2	4	8	5	5	4	4.633333	139	1.449931
sample5	4	3	2	4	5	5	3	5	3	7	4.733333	142	1.720732
sample6	4	6	3	5	5	6	6	2	6	4	5.066667	152	1.436791
sample7	2	6	4	6	6	7	5	7	6	3	5.200000	156	1.471570
sample8	8	4	6	6	4	5	7	5	4	4	5.133333	154	1.502488
sample9	3	6	3	6	6	8	5	5	4	5	5.366667	161	1.496740
sample10	4	4	4	4	8	3	4	6	6	5	4.900000	147	1.748891

112. ábra. Minták binomiális eloszlásból

(Tk. 74. o. 3.5.1. fejezet 3.7. példa)

18. Eszközök



113. ábra. Eszközök menü: Tools

Ebben a menüben vannak lehetőségek már installált R csomagok és R Commander kiegészítések (plug-in) – amik tulajdonképpen R csomagok – betöltésére, és az R Commander kinézetének beállítására. Csomagok és R Commander „plug-in”-ek installálása az R Commander installálásához hasonlóan történik (2. fejezet).

Load package(s)... R csomag betöltése

Load Rcmdr plug-in(s)... R Commander „plug-in”-ek betöltése

Options... Beállítások

Save Rcmdr options Beállítások mentése

Egyre több „plug-in” jelenik meg az R Commanderhez⁵. Jelenleg 29 van belőlük. Néhányat ezek közül érdemes megemlíteni:

RcmdrPlugin.coin A coin csomag R Commander-es implementációja (ld. tankönyv 7.6. fejezet)

RcmdrPlugin.plotByGroup Grafikus kiegészítés, lehetővé teszi, hogy például a QQ-ábrát csoportonként készítsük el

RcmdrPlugin.survival Túlélés elemzés

⁵<http://rcommander.com/>

19. A Biostatisztika tankönyv R Commander-rel megoldható példái

Példa száma	Címe	Oldal	Útmutató fejezet
2.1.	Véletlen számok mintavételhez	30	17.
2.2.	Matek osztályzatok – gyakorisági táblázat	36	8.3.
2.3.	Testtömeg-index kiszámítása R-rel	40	7.2.
2.4.	Studentizált értékek kiszámítása R-rel	43	7.2.
2.5.	Rangok kiszámítása R-rel	44	7.2.
3.5.	Balkezesek száma – hipergeometrikus eloszlás, dhyper() függvény	71	17.2.
3.7.	Balkezesek száma – binomiális eloszlás, dbinom() függvény	74	17.2.
3.9.	Percenkénti hívásszámok a mentőknél – Poisson-eloszlás	76	17.2.
3.10.	Számolás eloszlásfüggvényekkel	81	17.1.
4.1.	Kocák alomszáma – gyakoriságtáblázat és oszlopdia-gram	88	8.3
4.2.	Inszeminálások száma – gyakoriságtáblázatok és dia-gramok	89	8.3.
4.3.	Kontingencia tábla	98	16.8.3.
4.4.	Regressziós kurzus hallgatóinak adatai – szórásdia-gramok	100	15.8
4.5.	Regressziós kurzus hallgatóinak adatai – boxplot és átlag \pm szórás	102	15.6., 15.12.
4.6.	Átlagszámítás	105	8.2.
4.9.	Variancia kiszámítása	110	8.2.
4.11.	Szarvasmarhák testtömege – szórás és relatív szórás	113	8.2.
4.12.	Elsőéves hallgatók testmagassága – kvartilisek, percen-tilisek	114	8.2.
5.6.	Konfidencia-intervallum a populációátlagra normális eloszlás esetén	142	10.1.
7.1.	Vizsgapontszámok – egymintás z - és t -próbák	195	10.1.
7.2.	Bikák és üszők születéskori testtömege – kétmintás t -próba	198	10.2.
7.3.	Első és második gyermek születéskori testtömege – páros t -próba	201	10.3.
7.4.	Keltetőgép hőmérséklet-tartása – F -próba	204	12.1.
7.5.	Kockadobás – illeszkedésvizsgálat	208	8.3
7.6.	Kutyák daganatai – megfigyelt gyakoriságok	214	9.3.
7.7.	Kutyák daganatai – függetlenség esetén várt gyakorisá- gok	215	9.3.
7.8.	Kutyák daganatai – khi-négyzet-próba R-rel	216	9.3.
7.10.	Kutyák daganatai – Fisher-féle egzakt próba R-rel	218	9.3.
7.11.	Szabályos-e az érme – próbák R-rel	221	11.1.

Példa száma	Címe	Oldal	Útmutató fejezet
7.12.	Valószínűségek összehasonlítása	224	11.2.
7.17.	Reakcióidő megnövekedése – Wilcoxon-próba	234	13.2.
7.18.	Vérszegénység kezelése – Mann–Whitney-próba	237	13.1.
7.19.	Pipacsok számlálása – Kruskal–Wallis-próba	240	13.3.
8.1.	Regressziós kurzus – szórásdiagram és korrelációs együttható	245	8.6., 8.7.
8.2.	Regressziós kurzus – a korrelációs együtthatóra vonatkozó próba	247	8.7.
9.2.	Regressziós kurzus – egyszerű lineáris regresszió	257	14.1.
9.3.	Regressziós kurzus – summary()	260	14.1.
9.5.	Regressziós kurzus – origón átmenő regresszió	264	14.1.
9.7.	Regressziós kurzus – szórásdiagram-mátrix	269	15.9.
9.8.	Regressziós kurzus – többszörös lineáris regresszió 1	270	14.1.
9.9.	Regressziós kurzus – többszörös lineáris regresszió 2	274	14.1.
9.10.	őzek – parciális korreláció	275	8.6.
9.11.	Őzek – multikollinearitás	277	16.9.1.
9.13.	Standardizált jackknife-reziduumok	289	16.3.
9.14.	Torzító pontok detektálása	291	16.3.
9.15.	Logaritmikus összefüggés	297	14.1.
9.16.	Csibék – exponenciális összefüggés	297	14.1.
9.18.	Őzek – hatványfüggvény-összefüggés	301	14.1.
10.1.	Növények magassága: első tápoldat – egytényezős ANOVA	320	10.4.
10.2.	Növények magassága: új tápoldat – Post hoc-tesztek	323	10.4.
10.3.	Növények magassága – kéttényezős ANOVA	326	10.5.
11.1.	Lepke – hatás- és átlagmodell	351	14.2.
11.2.	Lepke – két faktoros modell	352	14.2.
11.3.	Lepke – interakció	352	14.2.
11.4.	Lepke – ANCOVA, párhuzamos egyenesek	354	14.2.
11.5.	Lepke – ANCOVA, különböző meredekségű egyenesek	355	14.2.
11.7.	Lepke – Az X -ek együttes tesztelése	362	16.2.
11.8.	Lepke – egyes magyarázó változók tesztelése	363	16.8.1
11.9.	Lepke – szimultán tesztelés	364	16.8.2.
11.10.	Lepke – konfidencia-intervallumok	365	16.7
11.11.	Lepke – diagnosztika	366	16.9.2.
11.12.	Őzek – linearitás vizsgálata	368	7.1.
11.13.	Lepke – információs kritériumok	372	??.
11.14.	Lepke – lépésenkénti (stepwise) szelekció	374	16.5.
11.16.	Lile – leíró statisztikák	381	8.2.
11.17.	Lile – modell	383	14.2.
A.1.	Csibék testtömege – konfidencia-intervallum az átlagra	420	10.1.
A.2.	Párosított és független mintás vizsgálatok	421	10.3.

Irodalomjegyzék

- [1] Reiczigel Jenő – Harnos Andrea – Solymosi Norbert (2014): Biostatisztika nem statisztikusoknak, Pars Kft., Nagykovácsi, 3. kiadás
- [2] Fox, J. (2005). The R Commander: A Basic Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14(9): 1–42.