Huzsvai László

BIOMETRIAI MÓDSZEREK AZ SPSS-BEN
SPSS ALKALMAZÁSOK

Debreceni Egyetem, Mezőgazdaságtudományi Kar
DEBRECEN
2004-2011

Ez a munka kéziratként kezelendő. A tartalma lefedi a mezőgazdasági és gazdálkodástudományi doktori iskolák Statisztika és számítógépes adatfeldolgozás, valamint Statisztika, kutatásmódszertan című tantárgyak anyagát. Több fejezete azonban még kidolgozás alatt áll.

TARTALOMJEGYZÉK

Bevezetés

Az SPSS program magába foglalja a legmodernebb statisztika eljárásokat az adatbázis kezeléstől, a leíró statisztikán keresztül a legbonyolultabb többváltozós matematikai statisztikai eljárásokig. A programcsomag nagy előnye, hogy nem csak előre elkészített statisztikai teszteket használhatunk, hanem saját magunk is készíthetünk egyszerű modelleket, és ezek statisztikai elemzését is el tudjuk végezni. Ennek előnye, hogy sok előre beépített statisztikai teszt igen szigorú alkalmazási feltételeket kíván meg, amik a talaj – növény – légkör rendszerben nem teljesülnek, de az általunk, a vizsgált jelenség sajátosságaihoz igazított eljárások már sokkal megbízhatóbb eredményeket szolgáltatnak.

Az adatok az alábbi típusba tartozhatnak:

- nominális
- ordinális
- intervallum
- skála

Az alacsony mérési szintű adatok:

a) nominális (középértéke a módusz, a leggyakrabban előforduló adat)

b) ordinális (középértéke a medián, a középen elhelyezkedő adat)

Magas mérési szintű adatok:

a) intervallum

b) skála típusú adatok (középértéke az átlag)

		Függő változó (y)	
		Alacsony	Magas
Független változó (x)	Alacsony	Kontingencia táblázat, Chi-négyzet	Két mintás t-próba, szórás-elemzés, GLM
	Magas	Diszkriminancia-analízis	Korreláció-, regresszió analízis (többváltozós esetekben: Faktoranalízis, stb

1. táblázat: *Két változó kapcsolatának elemzése*

Magas mérési szintű adatokból lehet előállítani alacsony mérési szintű adatot, fordítva nem.

FILE MENÜ

Read Text Data:

text típusú adatok beolvasása, pl. automata meteorológia állomás adatait. *.dat kiterjesztéssel. *Fixed widht*, a felső sor tartalmazza a változók neveit. A változók régi neveinek újakat adhatunk. Mentsük el a fájl formátumát későbbi munkák számára *.tpf kiterjesztéssel.

Apply Data Dictionary:

Az SPSS-be már beolvasott adatok oszlop, címke, stb. kiegészítő adatait már meglévő adatbázisból is beolvashatjuk a fenti paranccsal, *.sav kiterjesztésű fájlt választva.

Display Data Info:

Lemezen tárolt adatbázis tulajdonságait, változóit, címkéit listázza ki.

Érdemes néha *.por, portable formátumba menteni az adatokat, mert ezt még a DOS-os programok is el tudják olvasni, mivel majdnem szöveg fájlként menti. Excelből 4.0-s munkalapként kell menteni az adatokat.

SZERKESZTÉS (EDIT) MENÜ

Options…

Charts: A grafikonok formátumát, kinézetét lehet megadni. A mintát (template) előre szerkesztett formátumban, fájlban megőrizve is megadhatjuk. Figyeljünk arra, hogy a megadott könyvtárban ott legyen a *.sct kiterjesztésű fájl. Ha töröljük, a program indítása után hibajelzést kapunk. Betűtípusokat, színeket, vonalakat, mintázatot határozhatunk meg. A grafikon keretét, rácsozatát állíthatjuk be interaktív módon.

Alapbeállítások: Edit – Options – General, Output Labels, Data

Nézet (View) menü

Status Bar

A táblázat alján található információs sávot jeleníthetjük meg vagy rejthetjük el.

Toolbars…

A menüsor alá különböző ikonokat rakhatunk ki, amelyek így gyors billentyűként szolgálnak. A leggyakrabban használt eljárásokat érdemes itt megjeleníteni. (Show Toolbars). A beállítás paranccsal (Customize…) elvégezhetjük a szükséges beállításokat. Az Edit Tool… billentyűvel még az ikonokat is átrajzolhatjuk kívánság szerint. Bal egér gombbal fogjuk meg az ikonokat és vigyük a kívánt helyre. Az ikonok törlését is hasonló módon végezhetjük, egyszerűen vontassuk ki az ikon területről.

Fonts…

Meghatározhatjuk a betű típusát (Arial, Courier, stb., stílusát (normál, dőlt, félkövér, félkövér dőlt), méretét (8-72). Kiválaszthatjuk az alkalmazott írásrendszert (Közép-európai, Nyugati, Görög, stb.).

Grid Lines

Az adatbázis ablakban a rácsozatot tudjuk ki, illetve bekapcsolni.

Adatok (Data) menü

Define Variable…

Az aktív adat editor ablakban a kiválasztott változó leíró fejléc adatait lehet megváltoztatni, vagy új adatbázis változóit lehet definiálni.

Templates…

Ha több változónak egyszerre akarjuk beállítani a tulajdonságait, akkor ezt a parancsot kell használni. Előzetesen az aktív editor ablakban a módosítandó változókat lenyomott egérbillentyűvel ki kell jelölni

Insert Variable

Új változó (oszlop) beszúrását végzi az aktív változó után.

Insert Case

Egy új eset (sor) beszúrását végzi az aktív eset után.

Goto Case…

Megkeresi az adott esetet. Ha nem az adat ablak az aktív, akkor ennek a parancsnak hatására azzá válik. A kereső dobozt a kívánt eset megkeresése után a Close gomb megnyomásával lehet lezárni.

Sort Cases…

Az adatmátrix sorai csökkenő vagy növekvő sorrendbe rendezhetők. A parancsdobozban meghatározhatjuk, hogy melyik legyen az elsődleges, másodlagos, stb. kulcs.

Transpose…

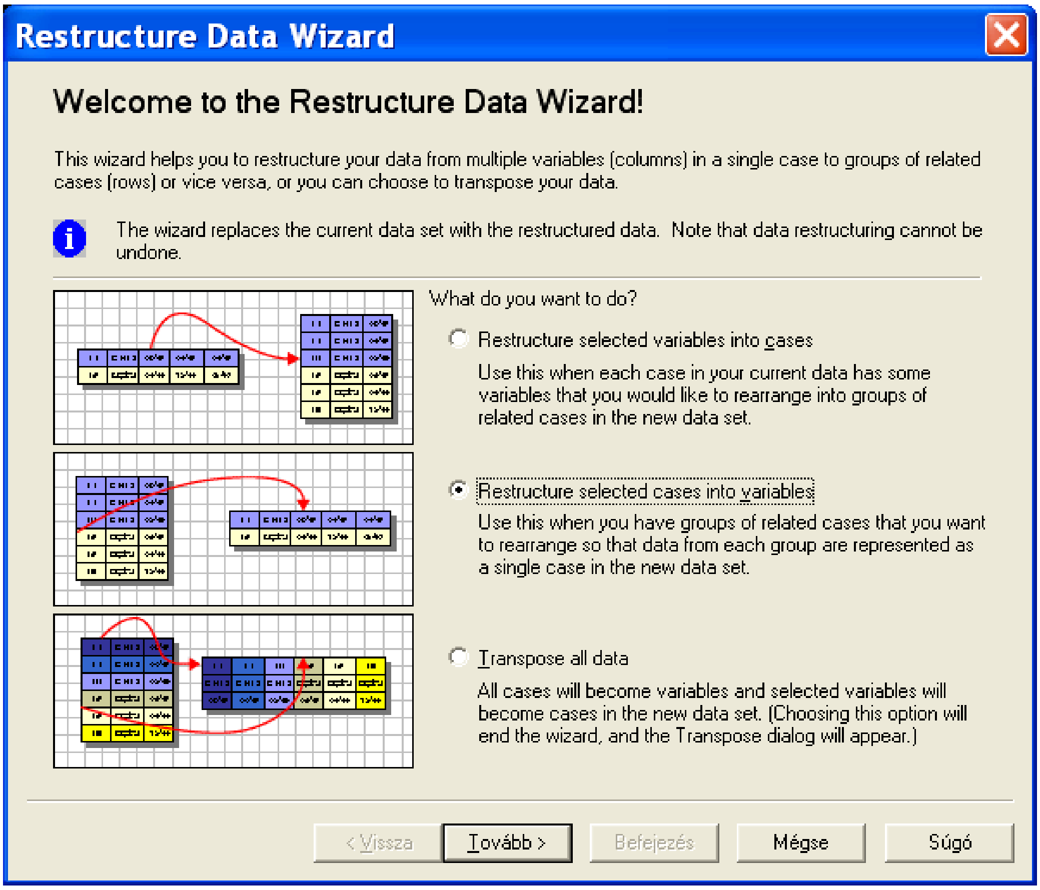
Az adatmátrix sorainak és oszlopainak felcserélése, ezzel az esetek és változók szerepei is felcserélődnek. A régi változók nevei a legelső új változó esetei lesznek, a többi új változó neve case_1, case_2, … stb. lesznek.

Restructure…

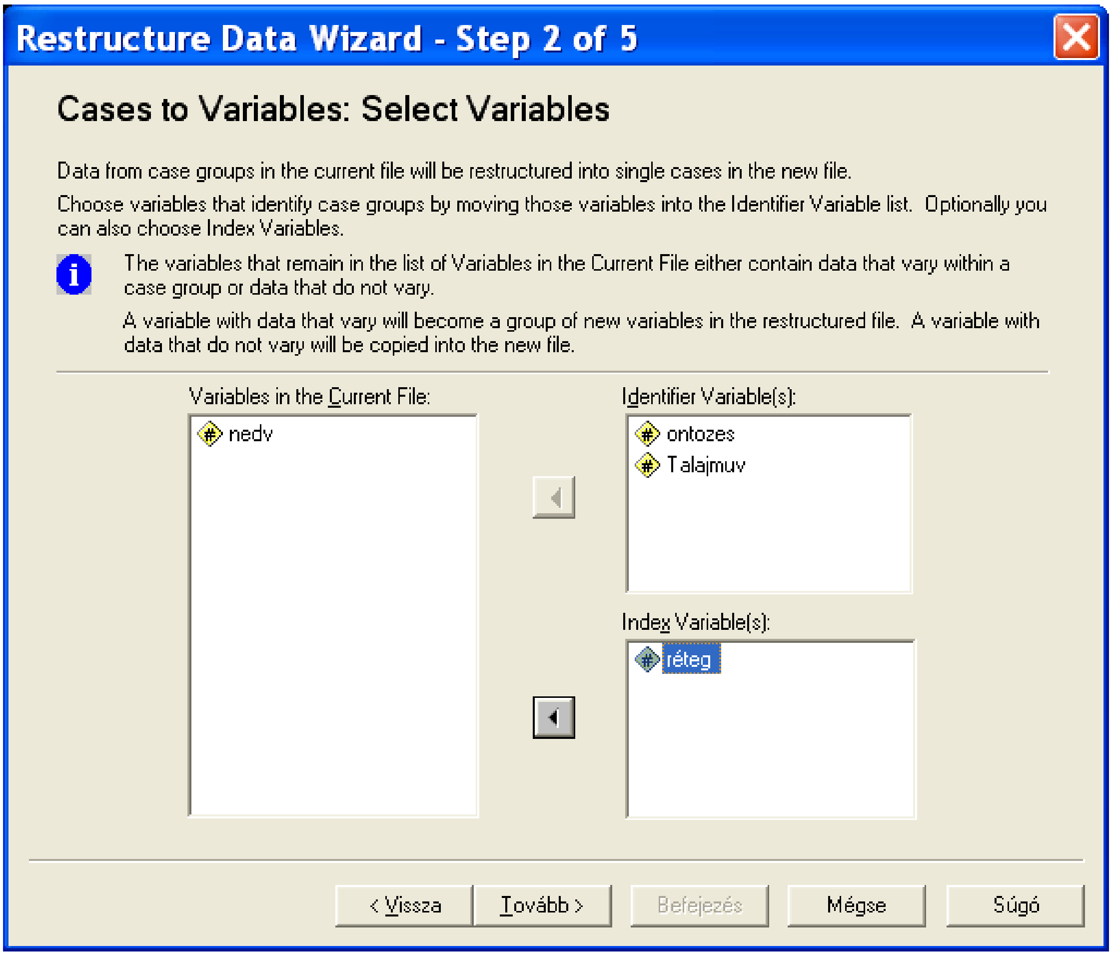
Itt az adatbázisok szerkezetét tudjuk megváltoztatni. Ezt átstrukturálásnak is nevezik. Vegyük az alábbi egyszerű adatbázist, és változtassuk meg a szerkezetét. A mért nedvességi értékek rétegenként kerüljenek új változókba.

Öntözés	Talajművelés	Réteg	Nedvesség
1,00	1,00	1,00	14,00
1,00	1,00	2,00	15,00
1,00	1,00	3,00	16,00
1,00	2,00	1,00	17,00
1,00	2,00	2,00	18,00
1,00	2,00	3,00	19,00
2,00	1,00	1,00	20,00
2,00	1,00	2,00	21,00
2,00	1,00	3,00	22,00
2,00	2,00	1,00	23,00
2,00	2,00	2,00	24,00
2,00	2,00	3,00	25,00

Data, Restructure… parancs után az alábbi párbeszédpanelt kapjuk. Itt kiválaszthatjuk, hogy a változókból csináljuk eseteket vagy fordítva, a kiválasztott esetekből legyenek új változók. A harmadik esetben az adatbázist transzponáljuk.



A Tovább billentyű után meg kell adni az új adatbázis szerkezetét. Szerintem, ez a párbeszédablakokban kissé nehézkes, sokkal egyszerűbb programból megadni. A baloldali ablakban láthatók a jelenlegi adatbázis változói (Variables in the Current File). Azonosító változóknak adjuk meg az öntözés és talajművelés változókat. Ezek külön sorokban fognak megjelenni az új adatbázisban. Index változónak jelöljük ki a réteg változót. Ez az adatbázis oszlopaiban fog megjelenni új változóként. Mivel három réteg van a nedvesség három új változóban fog megjelenni.



Öntözés	Talajművelés	Nedvesség_1	Nedvesség_2	Nedvesség_3
1,00	1,00	14,00	15,00	16,00
1,00	2,00	17,00	18,00	19,00
2,00	1,00	20,00	21,00	22,00
2,00	2,00	23,00	24,00	25,00

Merge Files

Fájlok bővítése, összekapcsolása. Új megfigyelésekkel (esetekkel) vagy új változókkal bővíthetjük az adatbázist. Az esetek bővítésével újabb megfigyeléseket csatolhatunk az adatainkhoz. Új változókkal történő bővítéskor több választási lehetőségünk is van, elő tudjuk állítani, pl. két fájl kombinációját egy kulcs változó felhasználásával. Legyen a *termés.sav* fájlnak három változója: *év, npk, termés*. Összesen 84 megfigyelt terméseredményünk van, öt-öt 1990-től 2003-ig. Legyen a *csapadék.sav* fájlnak két változója: *év és csapadék*. Összesen 14 megfigyelésünk (rekord) van, 1990-től 2003-ig. Ki szeretnénk bővíteni a *termés.sav* fájlunkat a csapadék értékekkel, hogy minden megfigyeléshez a megfelelő csapadékérték tartozzon. Nyissuk meg a *termés.sav* fájlt, és rendezzük növekvő sorrendbe az évek szerint. Válasszuk az Add Variables parancsot, a fájl megnyitás párbeszédpanelből válasszuk ki a csapadék.sav fájlt. Új párbeszédpanelt kapunk, amiben a két fájl információi láthatók. Válasszuk a Match cases on key variables in sorted files lehetőséget, és a rádiógombok közül External file is keyed table. A külső adatbázis lesz a kulcsmező tábla, ez tartalmazza a kulcsmezőt. A kulcsmező csak egyszer fordulhat elő a táblában. Az Excluded Variables: mezőben jelöljük ki az évváltozót, és húzzuk a Key Variables: mezőbe. Az OK gomb lenyomása után figyelmeztetést kapunk: ha nincsenek a fájlok a kulcsmező szerint sorba rendezve, rossz eredményt kapunk. Ez a lehetőség nagyon jól használható a logikailag összetartozó különböző táblák időszakos összekapcsolására, és elemzési feladatok elvégzésére. Ez nem más, mint az egy a többhöz kapcsolat megteremtése egy relációs adatbázisban. Ennek két feltétele van, hogy mindkét fájlban legyen azonos kulcsmező, ami alapján össze lehet kapcsolni a két adatbázist, és mindkét fájl a kulcsmező szerint sorba legyen rendezve.

Aggregate Data

Break Variables: az a változó, ami szerint az összegzés ill. statisztika készüljön. Aggregate Variables: változó, amit összegezni szeretnénk. Create new data file: ezt választva egy új aggr.-sav kiterjesztésű fájl készül az aggregált adatokkal.

Orthogonal Design

Generate…

Többtényezős kísérletek számára lineárisan független kezelés-kombináció tervet készíthetünk véletlen szám generátor segítségével. A tényező nevének (Factor Name) és címkéjének (Factor Label) megadása után az Add billentyűvel felvesszük a tényezők ablakba. Az egérrel kiválasztva a tényezőt definiálni kell a kezelésszintek számát (Define Values…), és el is lehet nevezni, pl. műtrágyából 1…3, nem trágyázott, 30 kg nitrogén, 60 kg nitrogén, stb.

Műtrágyázás	Öntözés	Status	Kártya
N 60	nem öntözött	Design	1
N 30	nem öntözött	Design	2
N 30	öntözött	Design	3
nem trágyázott	öntözött	Design	4
nem trágyázott	nem öntözött	Design	5
N 60	öntözött	Design	6

Split File…

Lehetőségünk van az adatbázist felosztani és az elvégzett analíziseket így elvégezni. Három lehetőség közül választhatunk:

1. Minden esetet megvizsgálunk, nem képezünk csoportokat.
2. A csoportokat hasonlítjuk össze.
3. Az analízisek eredményét csoportonként jelenítjük meg.

Select Cases…

eseteket választhatunk ki az adatbázisból. Négy lehetőség közül választhatunk:

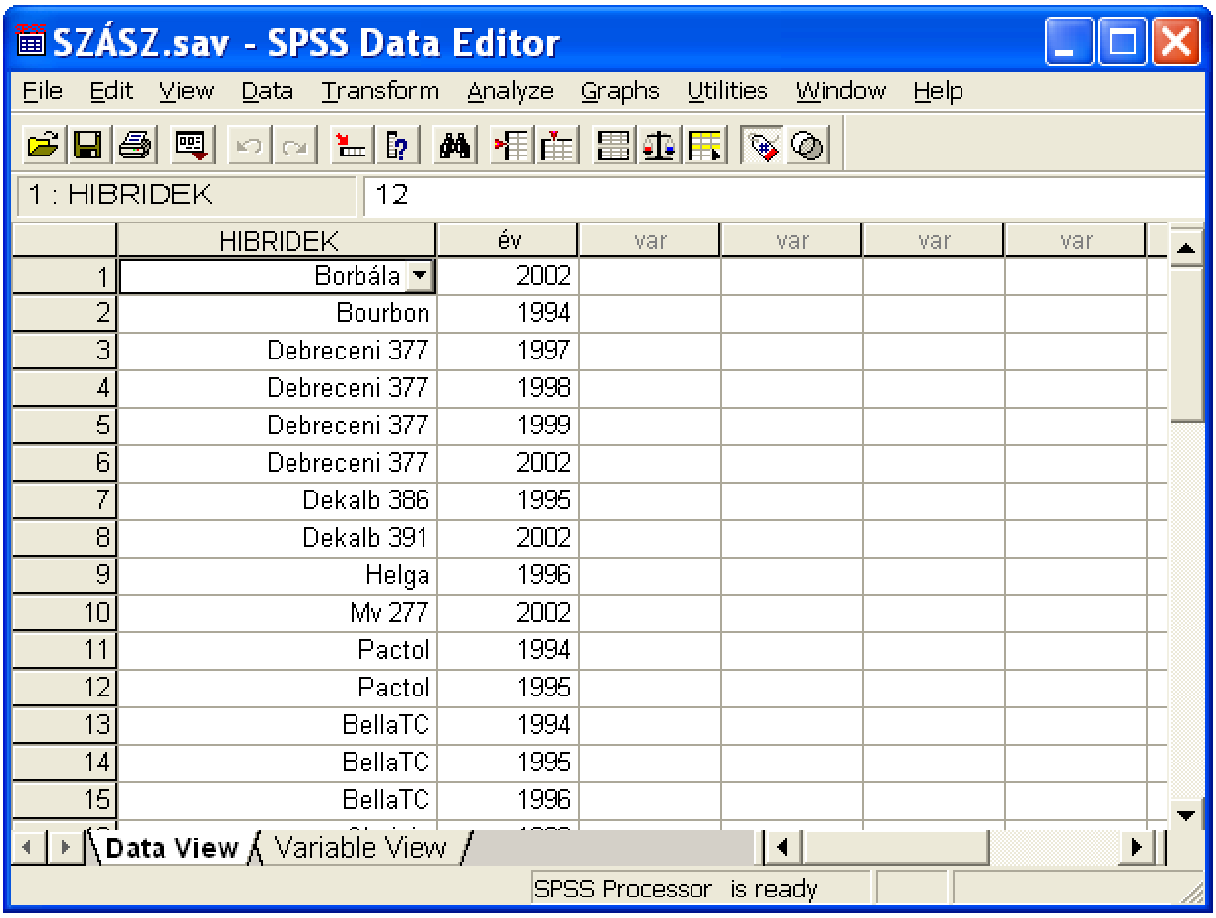
1. Minden eset részt vegyen az analízisben.
2. Ha valamilyen feltétel teljesül (if then)
3. Véletlen minta az esetekből
4. Kijelölhetjük az esetek bizonyos tartományát, az első és utolsó eset megjelölésével
5. Használhatunk szűrő változót

Mi legyen a ki nem választott esetek sorsa? Lehet szűrni és törölni őket az adatbázisból.

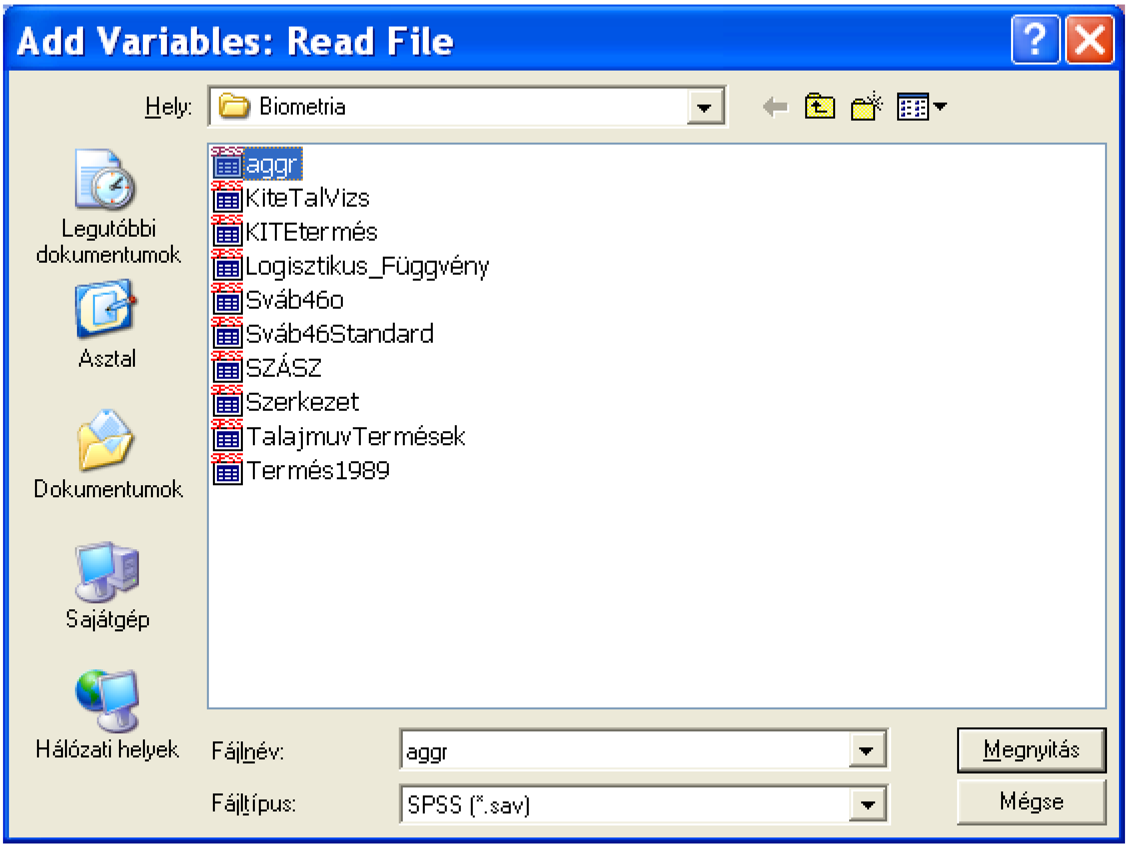
Nagymennyiségű adat lekérdezése

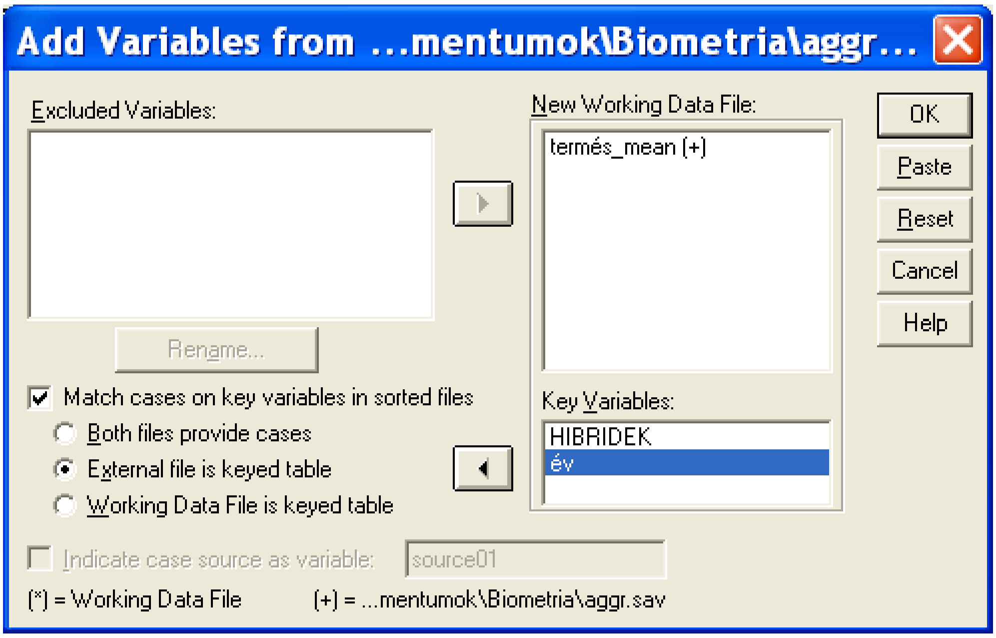
Egy viszonylag nagy adatbázisból nagy mennyiségű adatot különbözőképpen kérdezhetünk le. Az egyik legegyszerűbb megoldás az adatok szűrése (select cases) parancs használata, azonban nagy mennyiségű adat, illetve többszempontos lekérdezéskor nagyon sokat kell írni, és bonyolult logikai kifejezéseket kell megalkotni. Nagy a hibázási valószínűség. A másik nagyon hatékony megoldás, ha készítünk egy lekérdező adatbázist, és ehhez kapcsoljuk a nagy adatbázisból az adatokat az összekapcsol utasítással (merge files, add variables).

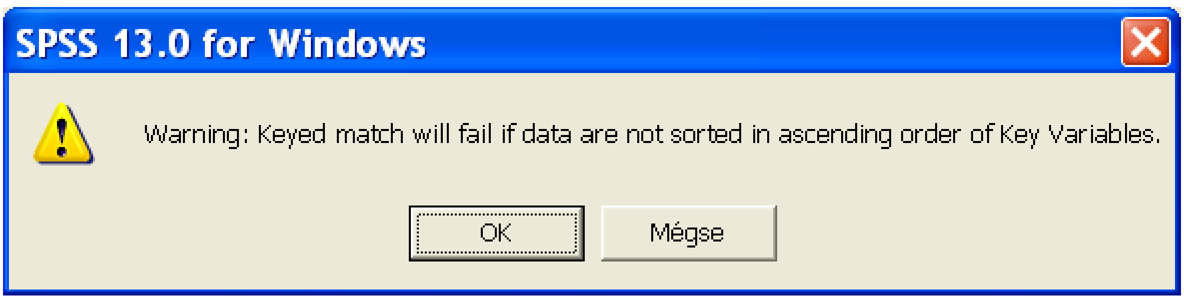
Pl.: a nagy adatbázis harminc év különböző kukorica hibridjeinek terméseit tartalmazza. Készítsük el az előre kiválasztott harminc hibrid egy-két vagy több éves terméseredményét. Az első lépés, alkossuk meg a lekérdező adatbázist. Rendezzük növekvő sorrendbe az adatokat a hibridek és év szempont alapján (Data, Sort Cases…).

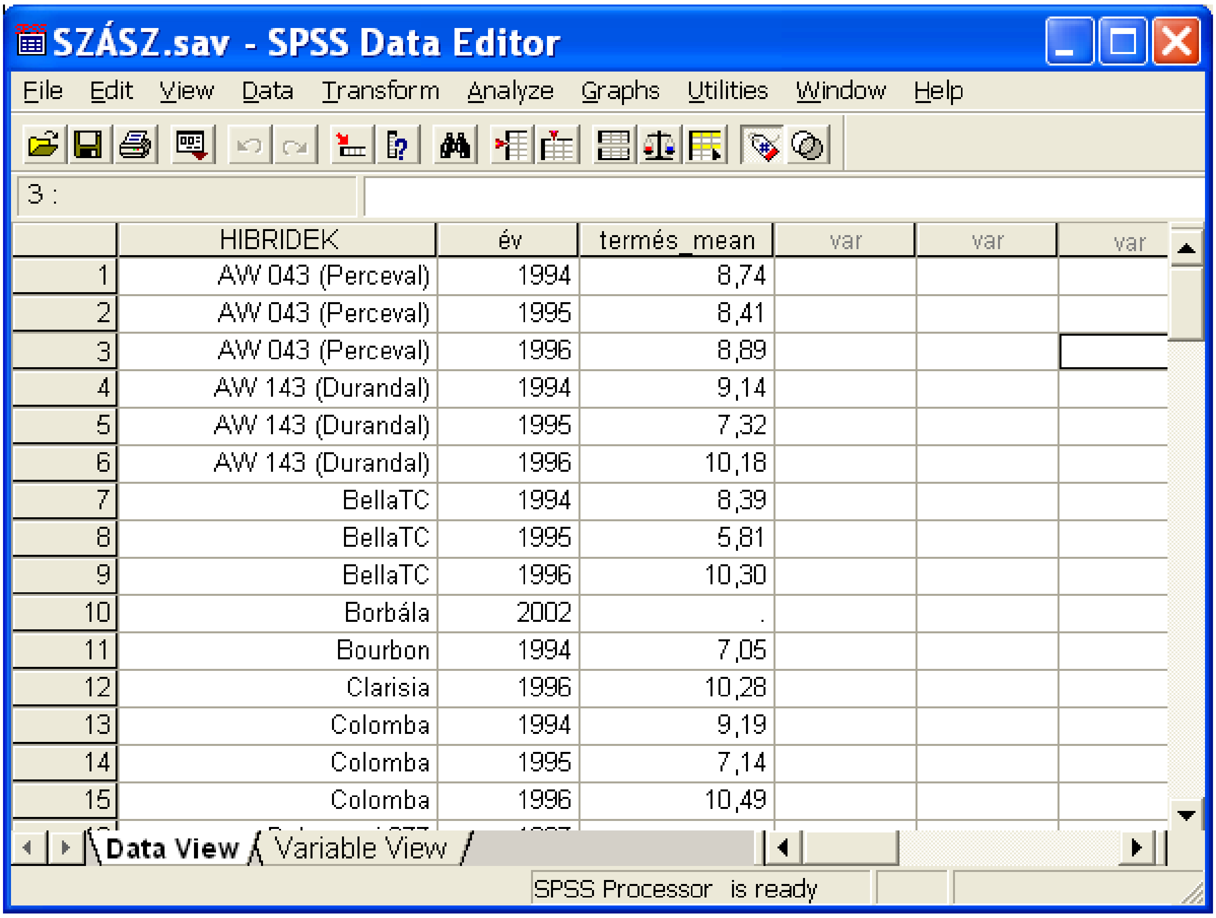


A második lépésben kapcsoljuk hozzá a terméseredményeket a nagy adatbázisból.


Add Variables: Read File
Hely:
Biometria
aggr
KiteTalVizs
KITEtermés
Logisztikus_Függvény
Sváb46o
Sváb46Standard
SZÁSZ
Szerkezet
TalajmuvTermések
Termés1989
Legutóbbi dokumentumok
Asztal
Dokumentumok
Sajátgép
Hálózati helyek
Fájlnév:
aggr
Fájltípus:
SPSS (*.sav)
Megnyitás
Mégse


Add Variables from ...mentumok\Biometria\aggr...
Excluded Variables:
New Working Data File:
termés_mean (+)
OK
Paste
Reset
Cancel
Help
Rename...
Match cases on key variables in sorted files
Both files provide cases
External file is keyed table
Working Data File is keyed table
Key Variables:
HIBRIDEK
év
Indicate case source as variable:
source01
(*) = Working Data File
(+) = ...mentumok\Biometria\aggr.sav


SPSS 13.0 for Windows
Warning: Keyed match will fail if data are not sorted in ascending order of Key Variables.
OK
Mégse



Weight Cases…,

Alul vagy túl reprezentált minták esetében lehet súlyzótényezőt alkalmazni. Ha több ismérv alapján is alul vagy túl reprezentált a minta, akkor egyenként kell a súlyzótényezőket kiszámítani, és az egyenkénti súlyzótényezőket össze kell szorozni. (Ez a szociológiai és társadalomkutatásban elfogadott eljárás.) Pl.: 60 megfigyelésből 50 férfi és 10 nő. A férfiak túl reprezentáltak ebben a mintában ezért a két súlyzó tényező férfiak esetében 10/60, nők esetében 50/60.

ÁTALAKÍTÁSOK (TRANSFORM) MENÜ

Az adatmátrix elemeit lehet megváltoztatni, illetve új változókat lehet előállítania régi változók segítségével. Átkódolhatjuk a régi esetek értékeit akár új, akár a régi változókba. Az esetek rangszámait is kiszámíthatjuk.

Compute Variable:

Számított változó létrehozása. Meg kell adni a célváltozó nevét és a numerikus kifejezést. Lehetőség van arra is, hogy valamilyen logikai kifejezést is beállítsunk, és ilyenkor csak azoknál az eseteknél képződik a számított érték, amelyeknél a logikai érték igaz. A többi helyre *system missing value* kerül.

Gyakran előforduló feladat, hogy idősort kell előállítani, vagy meglévő idősort kell különböző szempontok szerint átalakítani. A talaj-növény-atmoszféra modellekben az időt az aktuális év január elsejétől eltelt napok számával jelölik (julian dátum). Havonkénti, negyedévenkénti összesítést ill. kimutatást így elég nehéz elvégezni. A program a különböző dátum függvényekkel lehetőséget biztosít az átalakításokra. Pl. DATE.YRDAY(év, az év napja) segítségével rendes dátumot lehet előállítani. A számított új változónak természetesen dátum típust kell megadni. A DATE.* függvényekkel számokból lehet különféle dátumot előállítani, az XDATE.* függvények pedig dátumból számokat, pl. napok száma, hónap száma, negyedévek száma, stb. Az így elkészített attribútumokkal különféle szempontok szerint csoportosíthatjuk az adatokat, készíthetünk statisztikákat, elemzéseket. (ld. *esztendő2002.sav*).

Véletlen számokat is elő tudunk állítani a beépített eloszlásfüggvények segítségével. Pl. RV.NORMAL(mean, stddev) normáleloszlás ismert középérték és szórás esetén.

Random Number Seed:

A számítógéppel generált u.n. pszeudó-véletlen számok előállításakor a kiindulási szám megadása. Csak sokszámjegyű, páratlan szám adható meg. Amennyiben sokszor generálunk véletlen számokat, időnként célszerű átállítani, nehogy ismétlődés lépjen fel a véletlen számok között.

Count:

Egy olyan új változó hozható létre, amelyben a változólistára felvitt változók együttes előfordulásait lehet regisztrálni.

Recode:

Előfordulhat, hogy ugyanazt a hibridet szintaktikailag kétféle módon rögzítettük, pl. Pelican és Pelikán. Az automatikus újrakódolás során két különböző szám fog hozzárendelődni a két megnevezéshez. Hogyan lehet ezt kijavítani?

Az újrakódolás során választhatjuk, hogy ugyanabba a változóba (Into Same Variables) vagy új változóba (Into Different Variables) kerüljenek az új értékek. Válasszuk, hogy ugyanabba a változóba kerüljenek az értékek. Fel kell sorolnunk a régi és új értékeket, és fel kell venni őket a listába, majd OK. Az újrakódolás megtörténik. Meg kell jegyezni, hogy a régi értékek, amelyek most már nem szerepelnek az adatbázisban, címkéi továbbra is megőrződnek.

A régi felesleges címkéket az Automatikus Újrakódolással (Autamatic Recode) törölhetjük. Összefoglalásként: Automatic Recode → Recode Into Same Variable → Automatic Recode.

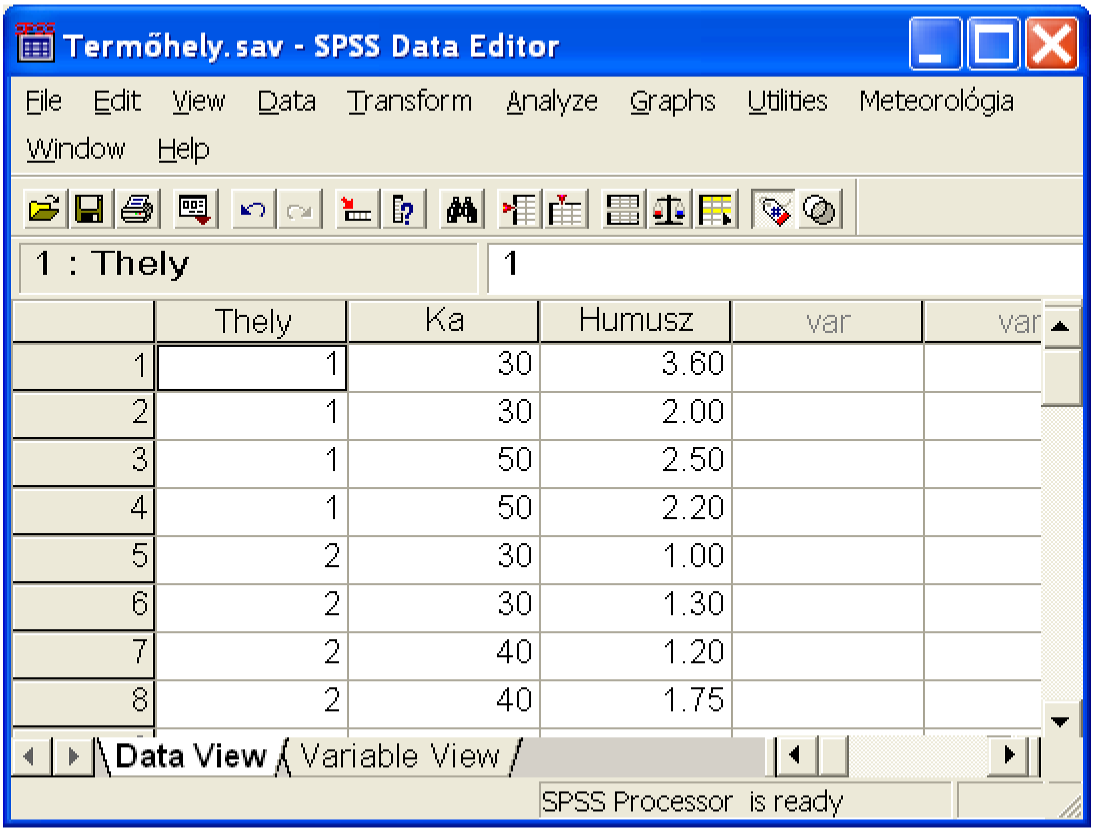
Ezzel a funkcióval kategorizálhatunk egy folytonos változót. A mezőgazdaságban pl. a talaj tápanyag-ellátottságának megítélését a talaj humusztartalmával szokták becsülni. A tápanyag-ellátottság azonban a humusz mellett függ a talaj típusától és egyéb, pl. a talaj kötöttségétől is. Ezeket korábban un. határérték táblázatokban foglalták össze. A táblázatok felhasználásával az SPSS-ben könnyen el tudjuk végezni a humusztartalom minősítését. Humusztartalom alapján: igen gyenge(1), gyenge(2), közepes(3), jó(4) és igen jó(5) ellátottságot lehet megkülönböztetni.

Egy ilyen táblázat részlete látható lent.

2. táblázat: A talajok humusztartalmának határértékei

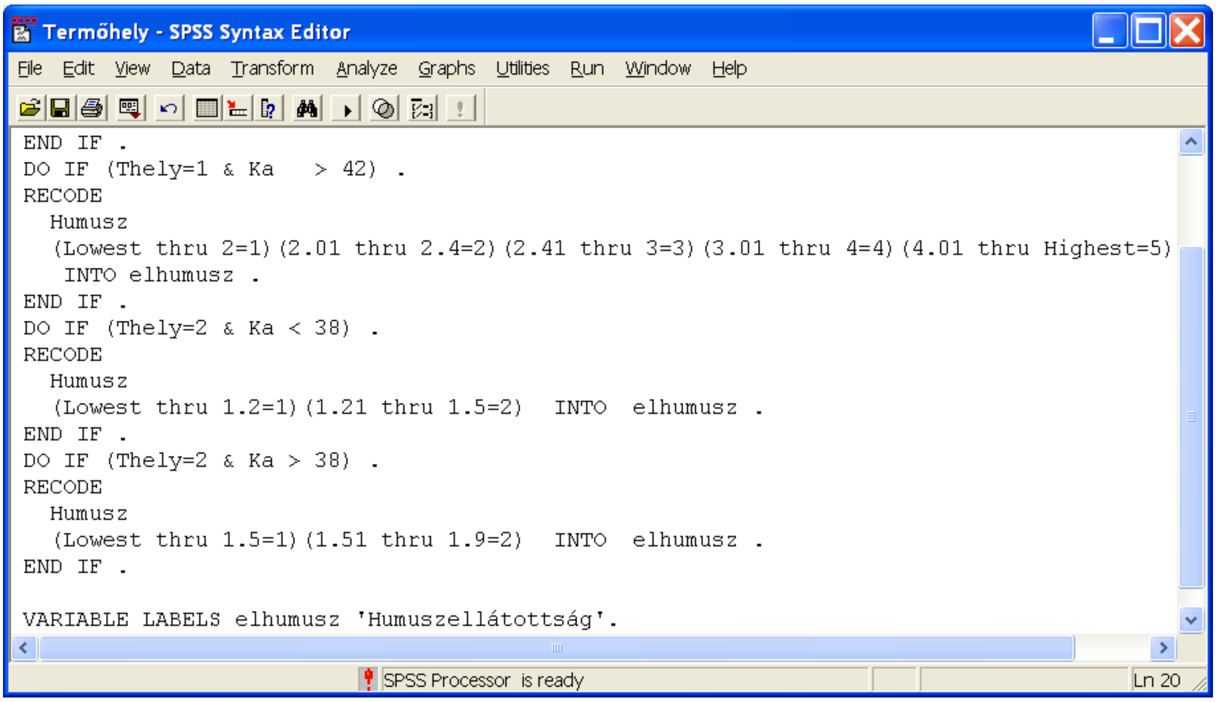
Szántóföldi termőhely	Ka	Humusz %				
		igen gyenge	gyenge	közepes	jó	igen jó
I.	>42	<2,00	2,01-2,40	2,41-3,00	3,01-4,00	>4,01
	<42	<1,50	1,51-1,90	1,91-2,50	2,51-3,50	>3,51
II.	>38	<1,50	1,51-1,90	1,91-2,50	2,51-3,50	>3,51
	<38	<1,20	1,21-1,50	1,51-2,00	2,01-3,00	>3,01

Az adatbázis ablakban látható a talajmintavétel után kapott eredményekből összeállított adatbázis.



1. ábra: A talajok humusztartama (%)

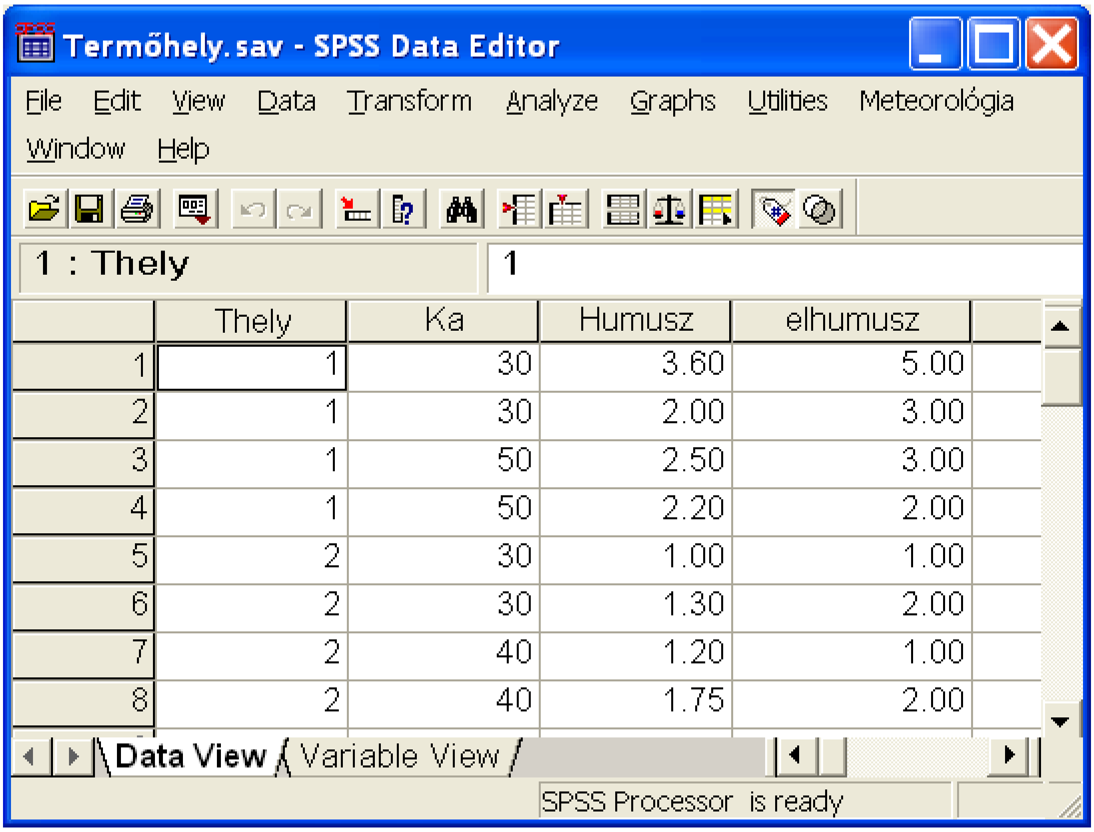
A 2. táblázat alapján bekódolhatjuk a talaj humusztartamát, és szövegesen jellemezhetjük a nitrogénellátottságot. Ehhez a Recode Into Same Variable parancsot kell használni. A párbeszédablakban a táblázat egyetlen sorát tudjuk csak megadni a feltételekkel együtt. Az első sor feltétele: Thely=1 és Ka>42. Ezután kell megadni a régi és új értékeket. A párbeszéd ablak tehát csak egyetlen sorra jó. Ilyenkor nagyon hasznos az SPSS syntax editora, mert segítségével automatizálhatjuk a kódolást. A Paste billentyű megnyomásával a párbeszédpanelen beállított utasítások a Syntax editor ablakban jelennek meg. Itt egyszerű másolással kibővíthetjük a feltételeket és megadhatjuk a kategória értékeket.



2. ábra: Program a humuszellátottság megítéléséhez

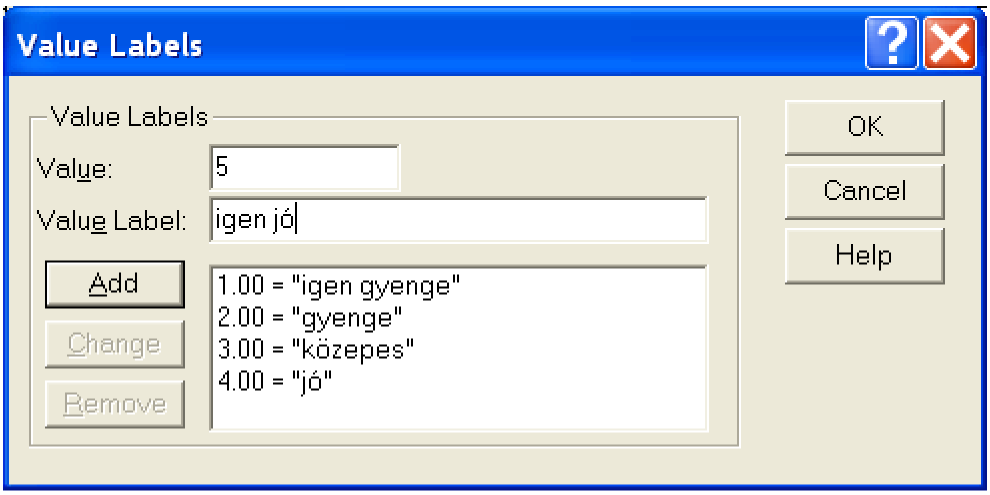
A programot a menüsor Run parancsával, Run All tudjuk futtatni. A futtatás után egy „elhumusz" változóban tárolódnak ez ellátottság kategóriái. (a második termőhely nem tartalmazza az összes kategóriát)

Az adatbázis a program futása után így alakul.



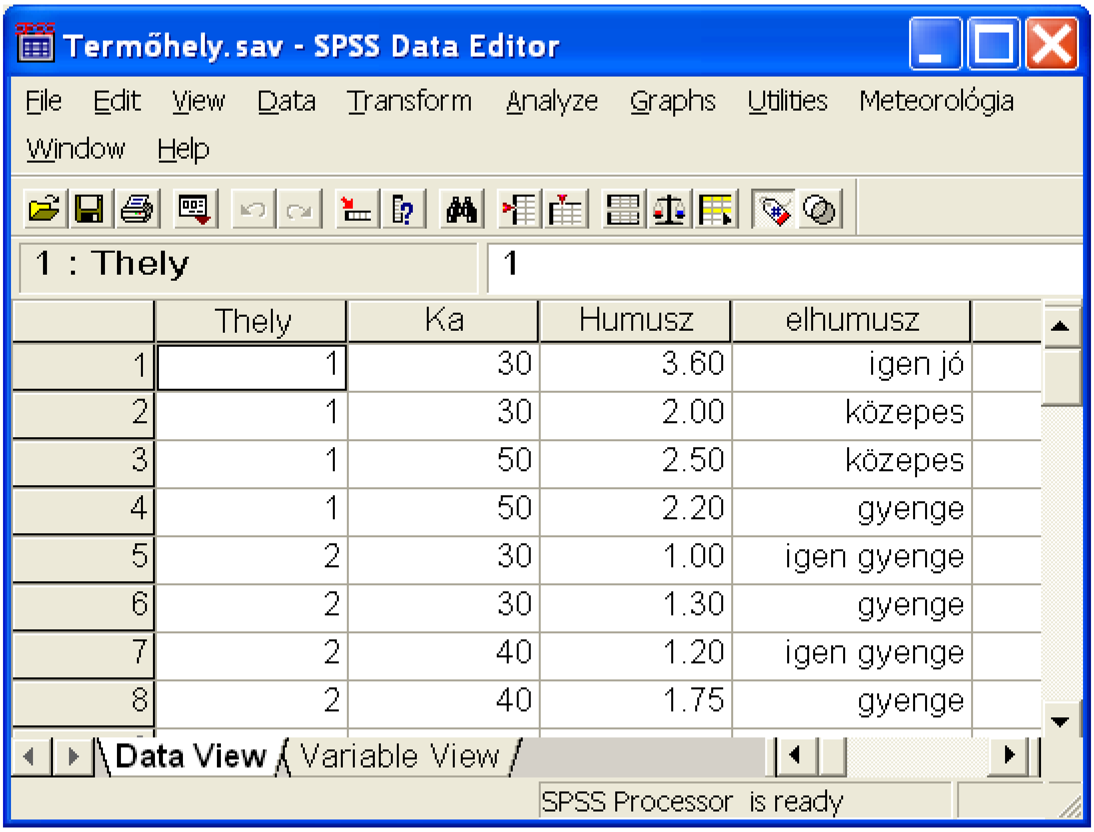
3. ábra: Humuszellátottság megítélése

Az „elhumusz” változóban az értékekhez címkéket kell rendelni.



4. ábra: Értékcímkék megadása

Adatbázis az értékcímkék beállítása után:



5. ábra: Humuszellátottság megítélése

Categorize Variables:

Egy változó tartományát lehet felosztani kategóriákra, alapállapotban négy kategóriát ajánl fel a program, de lehet változtatni.

Rank Cases:

Egy változó értékeinek a nagyság szerinti sorrendben elfoglalt helyzetének megfelelő rangszámát generálja egy új változóba. Ha két egyforma érték áll a változóban, megfelezi a sorszámot, pl. 1,5 és 1,5.

Automatic Recode:

Változókat lehet automatikusan újrakódolni. A változó listából válasszuk ki az újrakódolandó változót, a New Name ablakba írjuk be az új változó nevét és nyomjuk meg a New Name gombot. OK után automatikusan újrakódolja a változót. Text típusú változó esetében, ha a változó különböző csoportokat jelöl nem érdemes a szöveget mindenegyes rekordban tárolni, elég csak a kódokat. Ezzel az adatfájlt mérete jelentősen csökken. A kódok numerikus értékek lesznek. Az újrakódolt változóban a számokhoz címkék (labels) kapcsolódnak, melyek az eredeti text típusú változó tartalmát veszik fel.

Run Pending Transforms:

A felfüggesztett transzformációs parancsokat hajtja végre. Főként a syntax-ok futtatásakor használjuk. amelyeket a transzformációs opciókat használva a Preferences parancsdobozban felfüggesztettünk.

ELOSZLÁSOK

UNIFORM(max) = egyenletes eloszlású pszeudó véletlen számok előállítása a 0 és max tartományban.

RV.UNIFORM(min, max) = egyenletes eloszlású pszeudó véletlen számok előállítása min és max között.

RND(numexpr) = egész rész függvény

Kockadobások szimulálása:

RND(UNIFORM(6)+0.5), egyenletes eloszlás 1-től 6-ig. ábrázolni a gyakoriságot oszlopdiagramon.

Hat új egyenletes eloszlású változó létrehozása, összeg kiszámítása. Ábrázoljuk az összeget!

Az adatok standardizálása, Analyze, Descriptive Statistics, Descriptives…, Save standardized values as variables. Ábrázolás.

Csebisev tétel

Az adatoknak $1-\frac{1}{K^2}$ -ad része mindig közelebb van az átlaghoz, mint K standard eltérésen kívül. Ez független a változó eloszlásától. K egy egynél nagyobb pozitív szám.

Pl. K=2 esetén, ¾ (75%) az adatoknak 2 standard eltérésen belül helyezkedik el, K=3 esetén 8/9 (89%) 3 standard eltérésen belül található.

ANALÍZISEK

Riportok

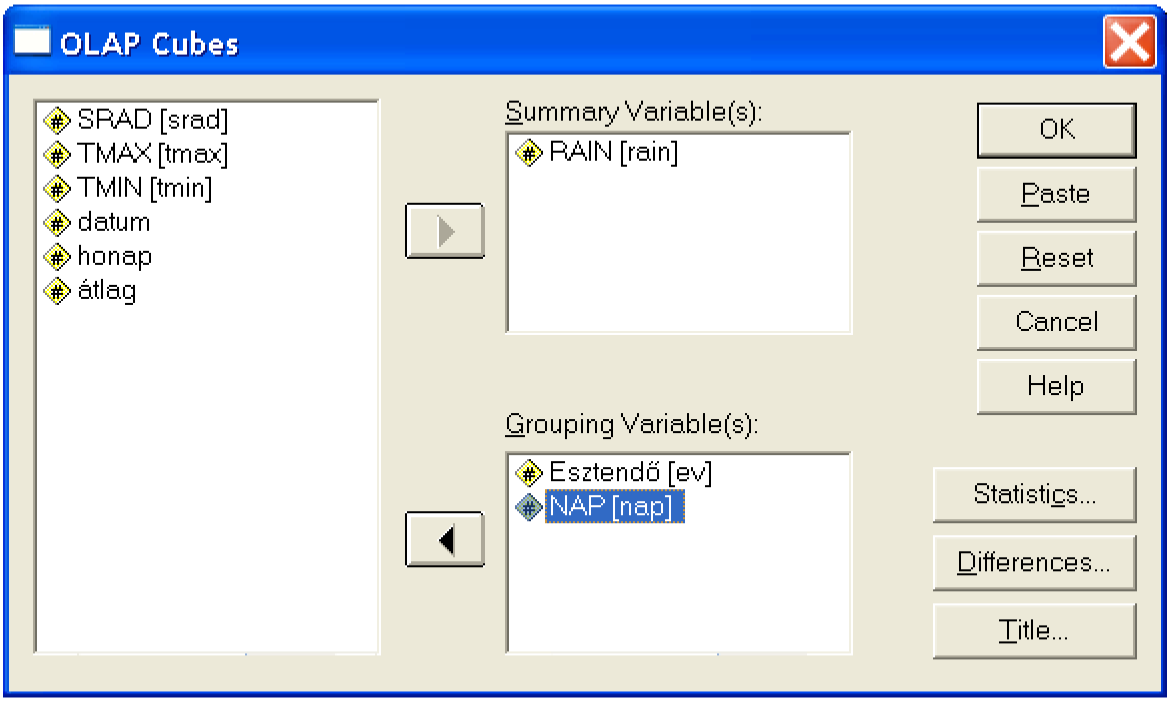
Adatbázisunkról különböző szempontok alapján készíthetünk kimutatásokat táblázatos formában.

OLAP Cubes…

Kimutatásokat, kimutatás táblázatokat készíthetünk skála típusú adtokkal (Olap Cubes), Pivot tábla formátumban. OLAP (Online Analytical Processing). Réteg (layer), sor (row) és oszlop (column) változók szerint csoportosíthatjuk az adatainkat. Különböző statisztikákat jeleníthetünk meg, centrális mutatókat, szóródási és terjedelmi jellemzőket.

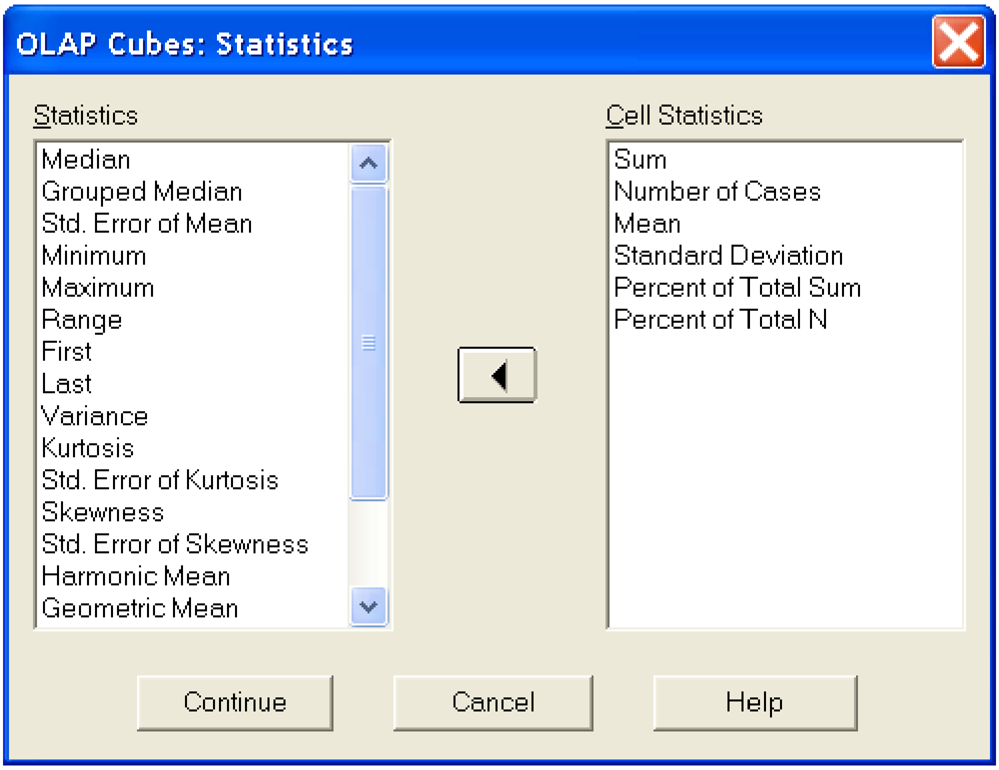
Analyze, Reports, OLAP Cubes…

Az elemezni kívánt skála típusú adatot vagy adatokat a



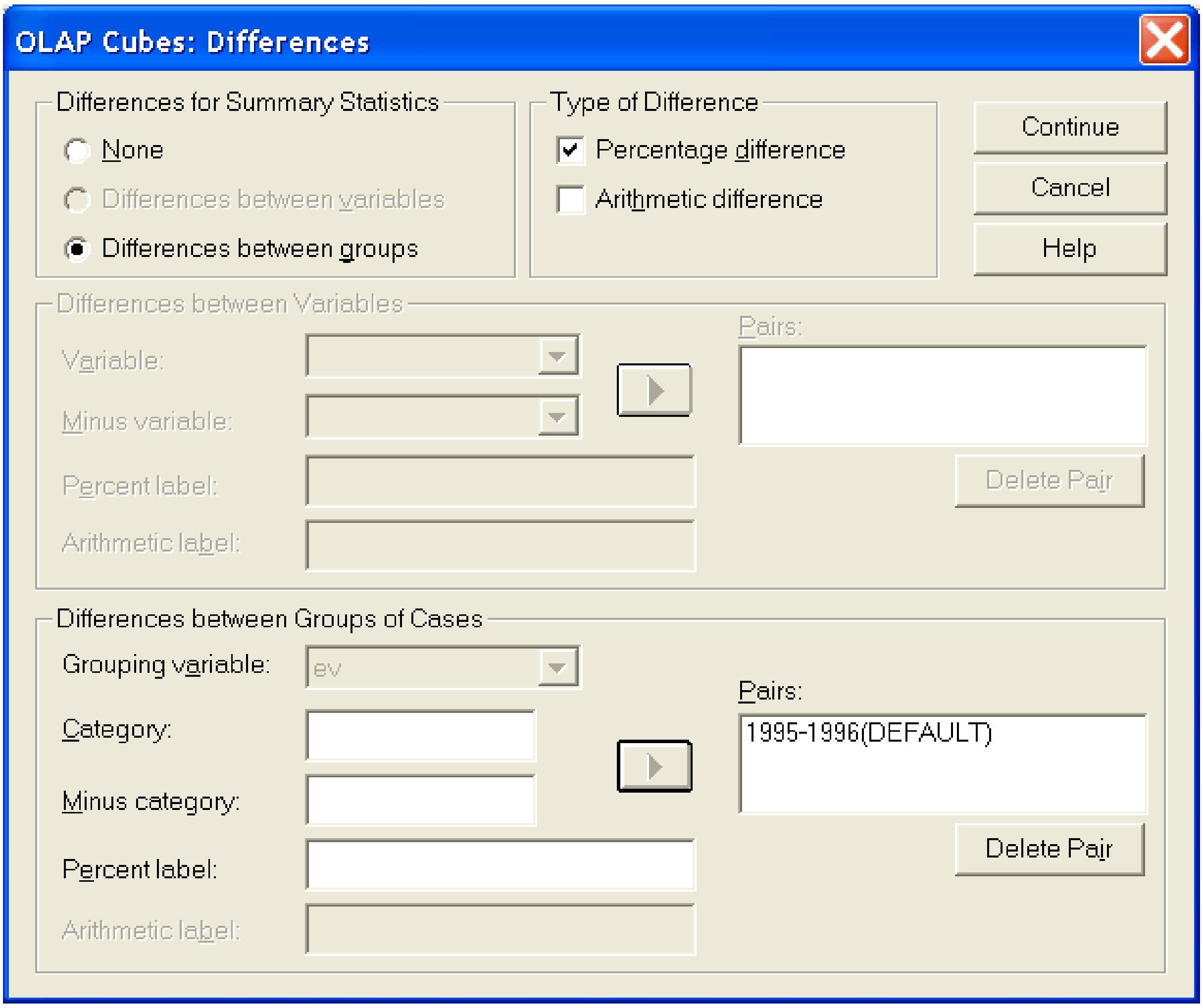
6. ábra: Kimutatás varázsló párbeszédablaka

Summary Variable(s): ablakba tegyük. A csoportképző változókat a Grouping Variable(s) ablakba. A Statistics… gombra kattintva különböző statisztikai jellemzőket választhatunk.



7. ábra: A kimutatásban megjeleníthető statisztikák

Differences… gomb a változók, ill. csoportok közötti különbségeket jeleníti meg.



8. ábra: A kimutatásban megjeleníthető különbségek

Az OK gomb lenyomása után az Output ablakban megjelenik az eredmény összezárt formában, azaz minden csoportképző változó a rétegekben (layer) kerül.

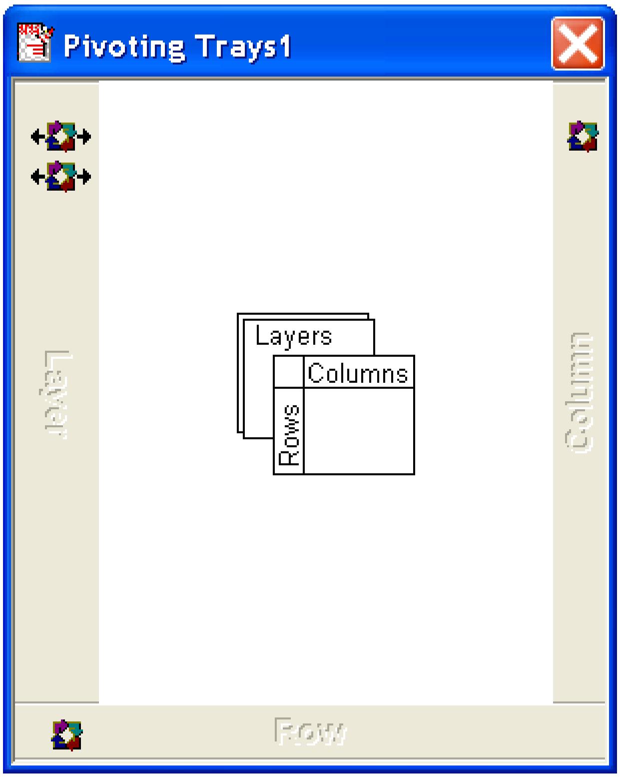
OLAP Cubes

Esztendõ: Total

NAP: Total

	Sum
RAIN	5101.1

A kimutatást tetszés szerinti formába önthetjük, a rétegeket sorokba illetve oszlopokba húzhatjuk. Ehhez kattintsunk kettőt a táblázatban az egér balgombjával. A felső menüsoron megjelenik a Pivot parancs, melyben a Pivoting Trays parancs megnyitja a szerkesztési lehetőséget.



9. ábra: A kimutatás szerkezetének megváltoztatása

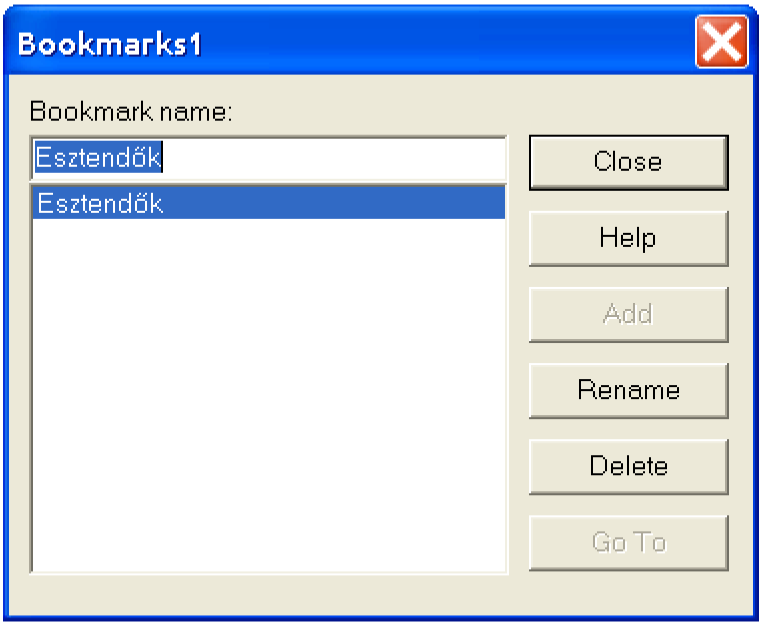
A baloldalon a réteg (layer), alul a sor (row) és jobboldalon az oszlop (column) található. A változókat az egérrel húzzuk a kívánt helyre, pl.

OLAP Cubes

NAP: Total

Sum		
RAIN	1995	414.0
	1996	573.0
	1997	397.0
	1998	635.0
	1999	637.0
	2000	359.0
	2001	585.0
	2002	411.5
	2003	520.9
	2004	568.7
	Total	5101.1

A táblázat mindenegyes elemét formázhatjuk, és elmenthetjük a kimutatás egyes változatait. Ehhez nyissuk meg a Bookmarks (könyvjelzők) parancsot.



10. ábra: A könyvjelzők megadása

Adjunk nevet az aktuális kimutatás változatnak, és az Add gombbal adjuk hozzá a könyvjelzőt.

A View menüparancsban válasszuk a Toolbars… lehetőséget, ekkor megjelennek a segédeszközök (tolltartó), melyek segítségével hasznos eszközök állnak rendelkezésünkre a kimutatások további elemzéséhez, formázásához.

11. ábra: Segédeszközök a kimutatások formázásához

Itt megtalálhatók a könyvjelzők is, amivel a kimutatások különböző változatai könnyen áttekinthetők.

Case summaries…

Nagyon hasonlít a pivot táblához, csak sokkal egyszerűbb formátumban jeleníti meg az adatokat. Jól használható a bevitt adatok ellenőrzésére, különböző csoportosítások szerinti adatmegjelenítéshez.

Case Summaries

Mean

HONAP	TMAX	TMIN
1,000	-1,268	-7,496
2,000	-1,629	-12,093
3,000	7,677	-2,019
4,000	15,650	3,213
5,000	26,100	12,616
6,000	28,057	14,823
7,000	27,352	15,674
8,000	29,671	13,008
9,000	22,063	9,453
10,000	12,165	3,248
11,000	10,490	2,657
12,000	2,652	-3,039
Total	15,005	4,431

3. táblázat

Report summaries in Rows…

Report summaries in Columns…

A meteorológia adatbázisból minden kimutatás elvégezhető ezzel az eljárással. A **Data Columns** párbeszéd ablakban kell megadni az elemzendő változókat. Minden változóhoz különböző statisztikát rendelhetünk, sőt ugyanazt a változót többször is felvehetjük különböző számítási eljárásokkal. Pl. a hőmérséklet-változóból az átlagot, minimumot, maximumot így egy táblázaton (kimutatáson) belül egyszerűen ki tudjuk számítani. A csoportképző változót a **Break Columns** ablakban kell megadni. Választhatunk növekvő, ill. csökkenő kiíratás között. A kimutatás rtf formátumban készül. Nagyon jól használható az aggregált adatok megjelenítéséhez.

Leíró statisztikák (Descriptive Statistics)

<u>Centrális mutatók:</u> Átlag (várható érték), Medián (középső adat, gyakran helyettesíti a számtani közepet), Módusz (leggyakrabban előforduló elem)

<u>Szóródási mutatók:</u> Helyzeti és számított, Maximum (standardizált értéke), Minimum (standardizált értéke), Terjedelem (max.-min., range), Kiugró értékek, Kvartilisek (negyedelők), Interkvartilis (Q3-Q1)/2, Szórás (standard eltérés), Variancia (szórásnégyzet), Standard hibája az átlagnak,Standard hibája a mediánnak

<u>Az eloszlás alakjának jellemzése:</u> Ferdeség (skewness, jobbra-balra ferde eloszlások), Csúcsosság (kurtosis, 0 normális még –2, +2 között), Boxplot ábrázolás

Trimelt, csonkított, robosztus leíró statisztika, a kiugró értékek elhagyása.

Gyakoriságok (Frequencies…)

A megfigyelt változók relatív és kumulatív eloszlását tudjuk elemezni, ill. ábrázolni. Megjeleníthetjük a gyakorisági táblázatot (Display frequency tables). A százalékos értékeken belül (Percentile Values): a kvartiliseket, ahol az adatok 25, 50 és 75%-a található. Feloszthatjuk az adatokat egyenlő csoportokra (2-től 100-ig) (Cut points for x equal groups) valamint tetszőlegesen megadott százalékok alapján is megjeleníthetjük az adatok eloszlását. A centrális mutatók közül az átlagot (mean), mediánt, móduszt valamint a megfigyelések összegét (sum), az eloszlási mutatók közül a szórást (std. Deviation), a varianciát, a terjedelmet (range), a minumum és maximum értékeket valamint az átlag hibáját (S.E. mean) tudjuk kiszámítani.

Statistics

t/ha

N	Valid	18208
	Missing	0
Mean		9.86786
Std. Deviation		3.05116
Skewness		-.474
Std. Error of Skewness		.018
Kurtosis		-.207
Std. Error of Kurtosis		.036
Percentiles	5	4.59100
	25	7.75800
	50	10.22800
	75	12.22500
	95	14.20255

4. táblázat

Meghatározhatjuk az eloszlás jellemző paramétereit is. Az eloszlás szimmetriáját a ferdeségi mutatóval (skewness) jellemezhetjük. A normáleloszlás szimmetrikus és a ferdesége nulla. Pozitív ferdeségi érték mellett az eloszlásnak hosszú jobboldali része, farka van (right tail), ekkor balra ferdül, negatív érték esetében jobbra ferdül az eloszlás. Amennyiben a ferdeség értéke nagyobb, mint egy, az eloszlás nem normál. Az adatok középpont körüli csoportosulását a csúcsossági mutatóval (kurtosis) mérhetjük. Normál eloszlás esetén az értéke ennek is nulla. A csúcsosság pozitív értéke azt mutatja, hogy az adatok szélesebb csoportban helyezkednek el, az eloszlás két széle hosszú. Negatív érték esetében kisebb csoportban helyezkednek el az adatok, az eloszlás két széle rövidebb. A példa a kukorica termésének (t/ha) eloszlását mutatja be.

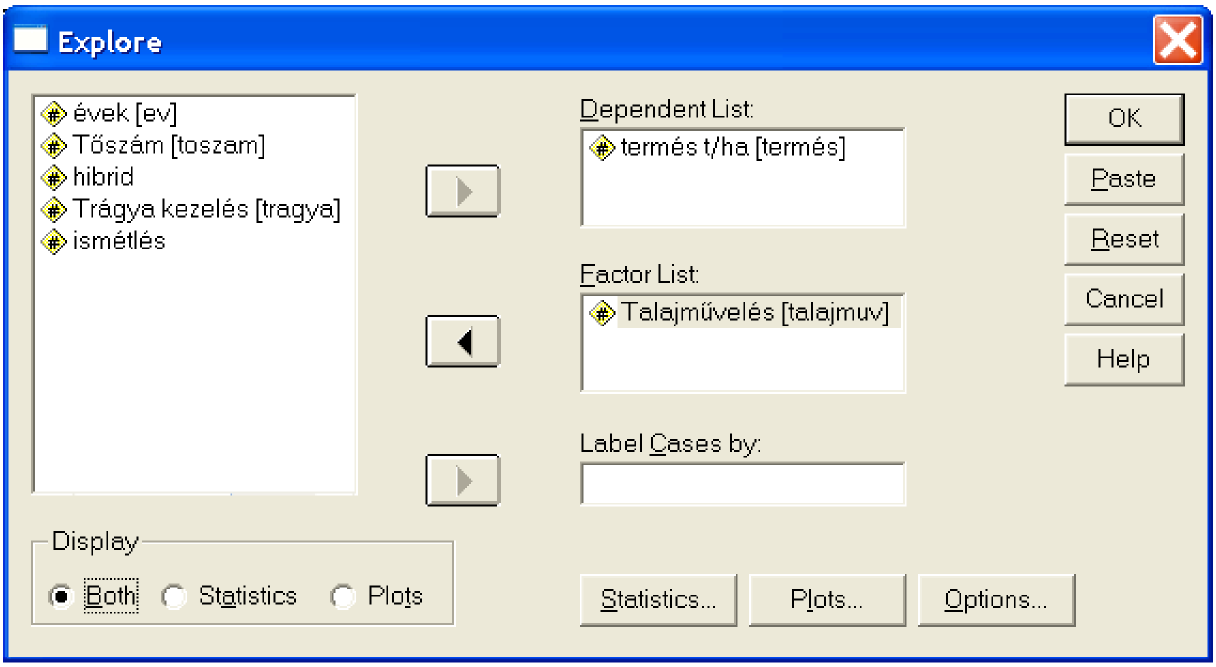
Ábrázolhatjuk az adatokat oszlop és kör diagramon, valamint hisztogram formájában is. A diagramokon ábrázolhatjuk a gyakoriságokat vagy a megfigyelések százalékos értékeit.

12. ábra

Descriptives…

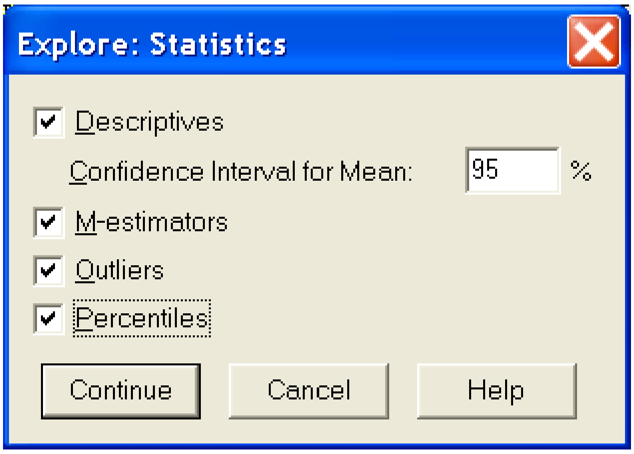
Explore…

Itt exploratív adatanalízist végezhetünk. Ez különösen fontos nagy adatbázisok esetében az adatok alapos megismerésére, felderítésére.



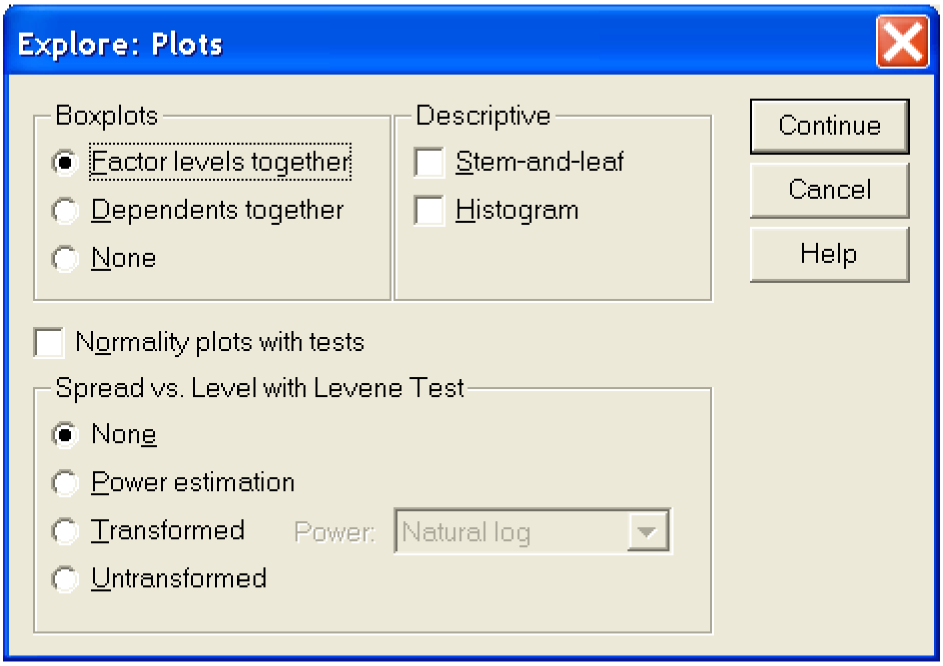
13. ábra

A Statistics… gombra kattintva különböző statisztikákat számíthatunk ki. Leíró statisztikák (Descriptives): átlag, medián, módusz, 5%-os csonkolt átlag, az átlag hibája, variancia, szórás, minimum, maximum, terjedelem, interkvartilisek, ferdeség, csúcsosság.



14. ábra

Robosztus centrális mutató meghatározása maximum-likelihood módszerrel (M-estimators). Négyféle módszerrel lehet meghatározni a centrális mutatót, mely torz eloszlás vagy extrém, kiugró értékek esetén jobb becslést ad, mint az átlag.



15. ábra

Az öt legnagyobb és legkisebb érték kijelzése (Outliers), ezeket az eredménylistában extrém értékként láthatjuk.

16. ábra

A megfigyelések százalékos eloszlását határozhatjuk meg, 5, 10, 25, 50, 75, 90, 95% (Percentiles).

Ábrák készítése, eloszlások tesztelése. Box-plot grafikonok (kvartilis vagy doboz ábra): a független változók függvényében készíthetünk kvartilis ábrát. A kiugróértékeket külön jelzi a program. A kvartilis ábrán a nagyság szerint rendezett adatok négy egyenlő részre vannak osztva. A minimális értéktől haladva a maximálisig Q_1, Q_2, Q_3 jelöli az adatok 25, 50, 75%-t. A Q_3 egyben a medián. Interkvartilis IQR= Q_3-Q_1, ezt doboznak is nevezik. A program kiugró értéknek (kör vagy csillag) jelöli a doboztól 1,5IQR-nél nagyobb távolságra elhelyezkedő adatokat (Q_3+1,5IQR illetve Q_1-1,5IQR).

Az adatok eloszlásának leírása (Descriptive):

Stem-and-leaf grafikon: stem=szár, leaf=levél skála típusú adatok felbontása, hogy a fő értéket a szár, az utolsó jegyeket a leaf adja. Pl. 7.18 t/ha stem=7, leaf=1.

```
termés t/ha Stem-and-Leaf Plot for
TALAJMUV= őszi szántás

 Frequency    Stem &  Leaf

     2.00        7 .  99
     6.00        8 .  002458
     6.00        9 .  013699
     3.00       10 .  009
     5.00       11 .  02278
     8.00       12 .  00035679
    13.00       13 .  1223346666668
     3.00       14 .  233
     2.00 Extremes    (>=113.5)

 Stem width:   1.000
 Each leaf:       1 case(s)
```

Hisztogram készítése (Histogram):

17. ábra

Normáleloszlás tesztelése Kolmogorov-Smirnov és Shapiro-Wilk próbával.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Talajmûvelés	Statistic	df	Sig.	Statistic	df	Sig.
termés t/ha	õszi szántás	.127	48	.050	.916	48	.002
	tavaszi szántás	.227	48	.000	.845	48	.000
	tárcsás	.263	48	.000	.817	48	.000

a. Lilliefors Significance Correction

Shapiro és Wilk's W-próba

Normális eloszlás tesztelésére szolgáló módszer, értéke maximum 1 lehet. Ennél jóval kisebb érték esetén nem normális az eloszlás. Szignifikancia vizsgálata megoldott, α = 0,05. Akkor érdemes kiszámolni, ha a minta elemszáma nem haladja meg az 50-et.

Kereszttáblák (Crosstabs…)

A meteorológiai alapadatok ellenőrzését is el lehet végezni vele. Minden nap 24 darab nulla, negyed, fél és háromnegyed órás mérésnek kell lennie. Adjuk meg a napokat sorként, a negyedórákat oszlopként.

A hónap napja * Perc Crosstabulation

Count

		Perc				Total
		0	15	30	45	
A hónap napja	1	23	24	24	24	95
	2	24	24	24	24	96
	3	24	24	24	24	96
	4	24	24	24	24	96
	5	24	24	24	24	96
	6	24	24	24	24	96
	7	24	24	24	24	96
	8	24	24	24	24	96
	9	24	24	24	24	96
	10	24	24	24	24	96
	11	24	24	24	24	96
	12	24	24	24	24	96
	13	24	24	24	24	96
	14	24	24	24	24	96
	15	24	24	24	24	96
	16	24	24	24	24	96
	17	24	24	24	24	96
	18	24	24	24	24	96
	19	24	24	24	24	96
	20	24	24	24	24	96
	21	24	24	24	24	96
	22	24	24	24	24	96
	23	24	24	24	24	96
	24	24	24	24	24	96
	25	24	24	24	24	96
	26	24	24	24	24	96
	27	24	24	24	24	96
	28	24	24	24	24	96
	29	24	24	24	24	96
	30	24	24	24	24	96
Total		719	720	720	720	2879

5. táblázat

Négymezős Chi^2-próba függetlenség és homogenitás vizsgálatra

Osszunk fel egy véletlen minta alapján kiválasztott 100 személyt két alternatív ismérv szerint: nemek szerint és dohányzási szokás szerint.

	Nem dohányzó	Dohányzó	∑
Nők	33	20	53
Férfiak	9	38	47
∑	42	58	100

	–	+	∑
–	a	b	$a+b = n_1$
+	c	d	$c+d = n_2$
∑	a+c	b+d	a+b+c+d = n

Függetlenség esetén:

a/ $n_1 = c/n_2 = (a+c)/n$ vagy

b/ $n_1 = d/n_2 = (b+d)/n$ stb

$$Chi^2 = \frac{(n-1)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$Chi^2 = \frac{99(33*38-20*9)^2}{(33+20)(9+38)(33+9)(20+38)} = 18{,}819$$

DF = 1

Kritikus Chi^2-értékek 5%-on: 3,841

Példa:

Kukorica fajták csövesedése:

FAJTA * CSÖVESD Crosstabulation

Count

		CSÖVESD		
		Egy csõ	Legalább két csõ	Total
FAJTA	A fajta	73	23	96
	B fajta	48	8	56
Total		121	31	152

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.038[b]	1	.153		
Continuity Correction[a]	1.486	1	.223		
Likelihood Ratio	2.123	1	.145		
Fisher's Exact Test				.210	.110
Linear-by-Linear Association	2.025	1	.155		
N of Valid Cases	152				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.42.

A Yates korrekcióval korrigált különbség négyzetéből számított Chi-négyzetet a Continuity Correction mutatja. A két kukoricafajta a vizsgált tulajdonság szempontjából egyforma.

Custom Tables

Középérték összehasonlítás (Compare Means)

A kezelésátlagok közötti különbségek megbízhatóságának igazolására többféle teszt ismeretes. Az összehasonlítás során, vagy két átlag különbségére vagyunk kíváncsiak, vagy a kezelésszintjeinket akarjuk összehasonlítani egymással, sorban tesztelve, hogy melyik kettő vagy több kezelés átlag tér el a többitől (szimultán vagy többszörös összehasonlítás). A kétféle eljárás kétféle összehasonlítási módszer csoportot takar. Az első módszer a páronkénti-tesztek csoportja a második a többszörös összehasonlító tesztek csoportja.

Középértékek (Means…)

A függő változók (Dependent List) különböző statisztikai mutatóit lehet kiszámítani a független változók (Independent List) függvényében. Elkészíthetjük a variancia-táblázatot, tesztelhetjük az összefüggés linearitását és az összefüggés szorosságára az R és *eta* paraméter nagyságából következtethetünk. Az R-érték, ill. R^2 a függő változó megfigyelt és becsült értékei közötti lineáris kapcsolat erőségét méri. Értéke 0,0 – 1,0 terjedhet. Kis érték esetében a függő és független változó között gyenge a kapcsolat vagy nem lineáris. Az *eta* paraméter a korrelációs koefficienshez hasonlít, de itt a független változó nem folytonos, hanem kategóriaváltozó.

Report

termés t/ha

Talajmûvelés	Mean	N	Std. Deviation
õszi szántás	11.50673	48	2.06058
tavaszi szántás	10.30987	48	2.06889
tárcsás	9.56033	48	2.28744
Total	10.45898	144	2.27357

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
termés t/ha * Talajmûvelés	Between Groups	(Combined)	92.524	2	46.262	10.087	.000
		Linearity	90.923	1	90.923	19.825	.000
		Deviation from Linearity	1.601	1	1.601	.349	.556
	Within Groups		646.657	141	4.586		
	Total		739.181	143			

Measures of Association

	R	R Squared	Eta	Eta Squared
termés t/ha * Talajmûvelés	-.351	.123	.354	.125

6. táblázat

A fenti táblázat első része a kezelések hatására kialakuló termésátlagokat és szórásokat tartalmazza. A legnagyobb termést őszi szántás után takaríthatjuk be. Az ANOVA tábla a variancia-analízis eredményét mutatja. A Between a csoportok közötti hatást. Ez fel van bontva lineáris és nem lineáris hatásra. A Sig. alapján a lineáris hatás szignifikáns, a nem lineáris hatás nem szignifikáns. Az utolsó táblázat a talajművelés és termés közötti összefüggés szorosságát mutatja. Az R és R^2-nek ebben az esetben nincs értelme, mivel a független változó (talajművelés) nem skála típusú változó, hanem nominális. Ez vegyes kapcsolat, ami az Eta ill. a négyzete mutat. Ennek természetesen nincs iránya csak nagysága. Minél közelebb van egyhez, annál szorosabb a két tényező közötti kapcsolat. Az Eta-négyzet a csoportok közötti eltérés-négyzetösszeg és az Összes eltérés-négyzetösszeg hányadosa. Ebben az esetben 92,5/739,2=0,125. Ennek a gyöke az Eta, ami ebben a példában közepes erősségű összefüggést mutat (0,354).

Egymintás t-teszt (One Sample T Test…)

Egymintás t-próba. Tesztelhetjük, hogy a valószínűségi változónk értéke megegyezik-e egy konkrét értékkel. Megválaszthatjuk a konfidencia intervallum nagyságát is.

Feltétel:

Normális eloszlású populáció, szigma ismeretlen és n>30.

$$z=\frac{\bar{X}-\mu_0}{s/\sqrt{n}}$$

DF = n-1

A minta elemszámának növekedésével a t – eloszlás egyre jobban közelíti a standard normális eloszlást.

Az X középértékű minta abban az esetben származhat a mű középértékű populációból ha t próbastatisztika abszolút értéke kisebb, mint az adott valószínűséghez tartozó kritikus t – érték.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
termés t/ha	144	10.45898	2.27357	.18946

One-Sample Test

	Test Value = 10					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
termés t/ha	2.423	143	.017	.45898	8.45E-02	.83349

7. táblázat

Egymintás z-próba

A minta középértékének összehasonlítása egy feltételezett középértékkel. Származhat-e az X középértékű minta egy μ_0 középértékű populációból? H_0 hipotézis:

H0: $\mu = \mu_0$

Feltétel:

Normális eloszlású populáció, és ismert szórás,

Vagy tetszőleges eloszlású populáció, és n>30.

A minta alapján számított X középérték standardizált érték felírható az alábbi formában:

$$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Ahol:

z a próbastatisztika minta alapján meghatározott értéke

X a minta középértéke,

μ a populáció feltételezett középértéke (adott középérték),

σ a populáció (ismert) szórása,

n a minta elemszáma.

A minta abban az esetben származhat az mű középértékű populációból, ha minta alapján meghatározott z próbastatisztika értéke kisebb az adott valószínűségi szinthez tartozó kritikus z -értéknél. Egyoldalú hipotézis esetén alfánál, kétoldalú hipotézis esetén alfa/2-nél kell kikeresni.

z < kritikus z

Két független minta középértékének összehasonlítása (Independent-Samples T Test…)

Származhat-e a két független megfigyelés, minta azonos középértékű populációból?

Azonosnak tekinthető-e a két populáció középértéke, amelyekből a minták származnak? A két populáció, amelyekből a minták származnak, $\mu 1$, ill. $\mu 2$ várható értékének becslésére a minták középértékei szolgálnak, $E(\overline{X}_1) = \mu_1$, ill. $E(\overline{X}_2) = \mu_2$.

H_0: $\mu_1 = \mu_2$

A középértékek összehasonlítására szolgáló statisztikai próbák – az egymintás próbákhoz hasonlóan – némileg eltérőek attól függően, hogy mekkora az egyes minták elemszáma, ill. hogy ismert-e az alappopulációk szórása.

Két független minta középértékének összehasonlítása. Feltétel:

Két független minta,

Normális eloszlású sokaságok,

A varianciák ismeretlenek, de azonosak

És n<30 (n nem elég nagy a kétmintás z – próba alkalmazásához)

Ha a varianciák ismeretlenek, akkor azokat a mintákból számított szórásnégyzetekből becsülhetjük. A próbastatisztika értéke:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_p \sqrt{(1/n_1) + (1/n_2)}}$$

DF = n1 + n2 – 2

A nevezőben az s_p a két minta összevont varianciájának (pooled variancie) négyzetgyökét jelenti, melyet a két minta összevont szórásának nevezzük és az alábbi képlettel számítjuk ki:

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

A két populáció középértéke, amelyekből a minták származnak, abban az esetben tekinthetők azonosnak, ha:

$$|t| \leq t^*$$

A próba statisztika kritikus t – értékét kétoldali alternatív hipotézis esetén $\alpha/2$-nél, egyoldali alternatív hipotézis esetén, α-nál kell a táblázatból meghatározni. Ha a két populáció ismeretlen szórásnégyzete korábbi ismeretek, ill. a mintákból számított szórásnégyzetek alapján nem tekinthető azonosnak, akkor a t – próba helyett a Welch-próbát kell alkalmazni, mely igen hasonló a t-próbához, a különbség a szabadságfokok meghatározásában van.

A t-teszt alkalmazásakor előre tudni kell, hogy a két csoport szórása megegyezik-e, tehát tesztelni kell a csoportok szórását (Levene-póba). Amennyiben a szórások egyenlők, akkor a vizsgálatba vont összes csoportból kell a varianciát becsülni (pooled variancia). A próba valószínűségi változója t-eloszlású, így a középértékek különbségének szignifikanciája a t-érték táblázatból megállapítható.

Ha a két csoport szórása szignifikánsan különbözik, ilyenkor a két összehasonlítandó csoport varianciáját súlyozni kell a variancia becsléséhez (separate variancia). A módosított variancia becslés az alábbi:

$$S_d = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

A próba valószínűségi változója ebben az esetben nem t-eloszlású, ezért nem a t-táblázatot, hanem a **Bonferroni-módosított** szignifikancia értékeket kell használni a középértékek különbözőségének elbírálásakor.

8. táblázat

Group Statistics

	Trágya kezelés	N	Mean	Std. Deviation	Std. Error Mean
termés t/ha	nem trágyázott	48	7.66106	1.23444	.17818
	nitrogén 120	48	11.77213	1.08695	.15689

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
termés t/ha	Equal variances assumed	.472	.494	-17.317	94	.000	-4.11106	.23740	-4.58243	-3.63969
	Equal variances not assumed			-17.317	92.518	.000	-4.11106	.23740	-4.58253	-3.63959

Kétmintás z-próba

Feltétel:

Normális eloszlású független sokaságok, a variancia ismert,

Vagy tetszőleges eloszlású, mindkét mintában n>30.

Az X_1 és X_2 középértékek különbsége akkor normális, ill. közelítőleg normális eloszlású, ha a sokaságok – amelyekből a minták származnak – normális eloszlásúak, illetve tetszőleges eloszlásúak, de a mintaelemek száma mindkét populációban nagyobb, mint 30.

A próbastatisztika:

$$z=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{(\sigma_1^2/n_1)+(\sigma_2^2/n_2)}}$$

A két populáció középértéke, amelyekből a minták származnak, abban az esetben tekinthetők azonosnak, ha:

$$|z|\leq z^*$$

A próba statisztika kritikus z-értékét kétoldali alternatív hipotézis esetén $\alpha/2$-nél, egyoldali alternatív hipotézis esetén, α-nál kell a táblázatból meghatározni.

Párosított t-próba (Paired-Samples T Test…)

Párosított t-próba, két összefüggő minta középértékének összehasonlítására szolgál.

Ugyanazon egyeden két különböző időpontban mérünk egy tulajdonságot, vagy valamilyen csoportképző tulajdonság alapján párokat tudok képezni.

A két minta középértékének azonossága helyett a párosított minták d (előjeles) különbségének középértékére is megfogalmazhatjuk a H_0 hipotézist:

H_0: $d_{átlag} = 0$

Az előző eljárásokhoz hasonlóan itt is z-. ill. t-próbát alkalmazhatunk attól függően, hogy ismert-e a d különbségek eloszlása és szórása, illetve mekkora a minta elemszáma?

Feltétel:

- a *d* különbségek eloszlása normális, és σ_d ismeretlen (a mintából számított), és n<30.

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \qquad \text{DF} = n-1$$

A képletben s_d a párosított minták különbségének szórása, amelyet a minta alapján becsüljük.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Maximális hõmérséklet (C)	15.133	365	10.034	.525
	Minimális hõmérséklet (C)	5.710	365	7.868	.412

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Maximális hõmérséklet (C) & Minimális hõmérséklet (C)	365	.900	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	Maximális hõmérséklet (C) - Minimális hõmérséklet (C)	9.424	4.521	.237	8.958	9.889	39.824	364	.000

9. táblázat

Átlag, az esetek száma, szórás, az átlag hibája. A két csoport közötti lineáris korrelációs együttható. Párosított t-próba eredmény táblázata: a két csoport különbségének átlaga, szórása, az átlag hibája, az átlag 95%-os konfidencia intervalluma, t-érték, szabadságfok, kétoldalú szignifikancia szint.

Egytényezős variancia-analízis (One-Way ANOVA…)

Egytényezős variancia-analízis. Segítségével egy tényező hatását lehet vizsgálni a függő változó mennyiségi alakulására. A tényező, faktor valamilyen csoportképző ismérvvel rendelkezik, a függő változó pedig legtöbbször skála típusú adat. Egyszerre több függő változót is kijelölhetünk az analízis számára. A teszt során a nullhipotézis, hogy az átlagok egyenlők, nincs közöttük különbség. Ez a technika a kétmintás t-teszt általánosítása, kiterjesztése több mintára.

Közös szórásnégyzet (variancia) =Vizsgált tényezők + Hiba

A számítás során az összes eltérés-négyzetösszeget bontjuk fel csoportok közötti (kezelés hatás) és csoporton belüli (hiba) hatásokra.

$$x_{ij} = \mu + A_i + \varepsilon_{ij}$$

μ = a kísérlet főátlaga
A_i = (Kezelés$_{\text{átlagok}}$-μ)
ε_{ij} = hiba

A hiba normális eloszlású, független a blokk és kezeléshatástól. Mi van, ha nem teljesül? Lehet transzformálni az alapadatokat, logaritmikus vagy egyéb transzformációval. A blokk (sokszor ez az ismétlés), kezelés és hiba hatások összege nulla.

Alkalmazási feltételek:

- Független megfigyelések
- Normális eloszlású sokaságok
- Azonos szórások

Amennyiben az analízis az átlagok közötti egyenlőséget nem igazolja, szükséges az átlagok közötti különbségek kimutatása. A variancia-analízist kiegészítő középérték összehasonlító teszteknek kétféle típusa létezik:

1. előzetes, un. a priori kontrasztok és
2. az analízis után elvégezhető, un. post hoc analízisek

A kontrasztokat tehát a kísérleti adatok elemzése előtt kell előállítani, és így elvégezni az elemzést.

Az alábbi statisztikák készülnek: minden csoportról az esetek száma, átlag, szórás, az átlag hibája, minimum, maximum és az átlag 95%-os konfidencia intervalluma.

A csoportok varianciájának egyezőségét Levene's teszttel vizsgáljuk. Minden függő változóra elkészülnek a variancia-táblázatok. A post hoc range és többszörös középérték összehasonlító tesztek: Bonferroni, Sidak, Tukey's honestly szignifikáns differencia, Hochberg's GT2, Gabriel, Dunnett, Ryan-Einot-Gabriel-Welsch F test (R-E-G-W F), Ryan-Einot-Gabriel-Welsch range teszt (R-E-G-W Q), Tamhane's T2, Dunnett's T3, Games-Howell, Dunnett's C, Duncan's multiple range test, Student-Newman-Keuls (S-N-K), Tukey's b, Waller-Duncan, Scheffé, and least-significant difference.

Szimultán vagy többszörös összehasonlítás (multiple comparison) a köztudatban a szórásanalízis kiegészítője, fejlődését főleg felhasználói igények indították útjára. Jelentősége azonban jóval nagyobb, különösen a nem paraméteres esetben, ahol szórásanalízisre, e normalitást feltételező eljárásra, nem kerülhet sor. Ha az egyszempontos szórásanalízis F-próbája szignifikáns, kíváncsiak vagyunk, mely populációk miatt nem homogén a minta. Eleinte csak páronként az összes lehetséges csoport párra kétmintás t-próbát hajtottak végre. Előfordulhat azonban, hogy adott α-szinten szignifikáns F-próba esetén egyik csoport pár sem mutat szignifikáns t-értéket az adott α-szint mellett. A szimultán hipotézis vizsgálatok nemcsak az egyszempontos szórásanalízisben hódítottak teret, hanem mindenütt, ahol egyidejű döntésre van szükség, pl. regresszió, kovariancia, többszempontos szórásanalízis, stb.

Szimultán döntés, ha kettőnél több összehasonlítandó mintám van. Olyan állításokat fogalmaznak meg, amelyek egyidejűleg érvényesek. Ezek lehetnek:

- Egyidejűleg érvényes konfidencia intervallumok vagy
- Szimultán végzett statisztikai próbák.

A többszörös statisztikai próbák zöme paraméteres, a normális eloszlásra épülő eljárás. Sorozatos statisztikai összehasonlítások végzésekor halmozódik a próbaként vállalt elsőfajú hiba (kockázat). A szimultán összehasonlítási módszerek fő célkitűzése ennek a halmozódásnak a csökkentése, illetve megszüntetése. Ennek eredményeként az egyes összehasonlítások konzervatív irányba tolódnak el: az egyenkénti próbákra érvényes elsőfajú hiba kisebb a vállalt (névleges) kockázatnál. Ez azonnal szembeötlik a többszörös összehasonlítások azon csoportjánál, amelyek az ún. Bonferroni-egyenlőtlenség alapján dolgoznak. Az első ilyen javaslat Fisher könyvében (1935) található. A lényege, hogy *m* összehasonlítás estén, az egyes összehasonlításokat a névleges α szint helyett α/m valószínűségi szinten hajtják végre. A valószínűség szubadditív tulajdonsága miatt, ha az összehasonlításonként vállalt α_i kockázatok összege olyan nagy, mint a teljes sorozatra vállalt α valószínűségi szint, akkor annak valószínűsége, hogy *m* elvégzett összehasonlítás után valahol elkövetjük az elsőfajú hibát, legfeljebb α:

$$P(H) \le \alpha = \sum_{i=1}^{m} \alpha_i$$

ahol: H esemény azt jelenti, hogy az állítások közt legalább egy hibás. Ha az egyes állítások (valószínűség-számítási értelemben) függetlenek lennének, akkor a fenti becslés helyett az

$$1 - P(H) = \prod_{i=1}^{m} (1 - \alpha_i)$$

egyenlőséget alkalmazhatnánk, ami azt mutatja, hogy az állítások között nincs hibás. Miller (1966) megmutatta, hogy a szimultán konfidencia-intervalumokra a fenti egyenlőség helyett mindig a $\geq$ érvényes. A szimultán vizsgált minták között végezhető összehasonlítások nem függetlenek. Legyen valamennyi α_i valószínűsége egyforma: $\alpha_i = \alpha_m = \alpha/m$, akkor az összehasonlítások nem független természetét figyelembe véve, a szimultán próbák együttes kockázata:

$$P(H) <= 1-(1-\alpha_m)^m$$

A levezetésből látszik, hogy az egyes szintek egyformaságának semmiféle szerepe nincs. Megtehetjük tehát, hogy a fontosabb összehasonlítások számára magasabb szintet jelölünk ki, ezzel biztosítva számukra a nagyobb erőt.

Kontrasztok: a csoportok közötti eltérés-négyzetösszeget (sums of squares) fel lehet bontani trend komponensekre, vagy előzetesen megadhatunk általunk definiált kontrasztokat is. A trendek között különböző hatvány függvényekkel leírható trendösszetevőket tesztelhetünk.

A kontrasztok az egyes csoportok várható értékeinek lineáris kombinációi. A súlyok segítségével meg lehet adni a csoportviszonyokat, akár több kontrasztot is egyidejűleg. Ilyen csoportviszonyok a mezőgazdaságban, pl. műtrágyadózis kísérletekben nagyon könnyen értelmezhetőek. A lineáris összehasonlító függvények elméletével több szerző is foglalkozott, magyar nyelven ÉLTETŐ Ö.-ZIERMANN M. 1964 megjelent művében található meg. A módszer lényege, hogy egy olyan lineáris függvényt kell alkotni, mint pl.:

$$\lambda_g = c_{g1}x_{1.} + c_{g2}x_{2.} + ... + c_{gp}x_{p.}$$

és ha teljesül a

$$c_{g1} + c_{g2} + ... c_{gp} = 0$$

feltétel, akkor ez egy lineáris összehasonlító függvény. A fenti definícióból következően végtelen számú λ_g létezik.

A kontrasztokra vonatkozó nullhipotézis: H_g: $\lambda_g = 0$,

Az ellenhipotézis: A_g: $\lambda_g \neq 0$.

Ha pl. egy tényező hatását T1, T2, T3, T4 szinten vizsgálunk, akkor a (T1, T2) csoport egybevetését a (T3, T4) csoporttal a λ_g

$= x_{1.} + x_{2.} - x_{3.} - x_{4.}$ függvény segítségével végezhetjük el (itt 1+1-1-1=0).

A fenti összehasonlítás a variancia-analízis által szolgáltatott *pooled* variancia felhasználásával történik, ezért követelmény, hogy a csoportok szórásai megegyezzenek, így gyakran a variancia-analízis kiegészítő részét képezi. A kontraszt fejezetben a hatótényezők sokféle csoportosítása útján kapott átlagok különbözőségét lehet vizsgálni, pl. műtrágyázás esetén, a feltételezésem az, hogy az őszi búza a legnagyobb termést a 120 kg nitrogén adag mellet éri el. Vizsgálhatom az ez alatti adagokat, mintát véve, vagy az e feletti adagokat, szintén mintát véve, véletlenszerűen, ha nem 120 kg-t alkalmazok, vajon milyen eredmény születne.

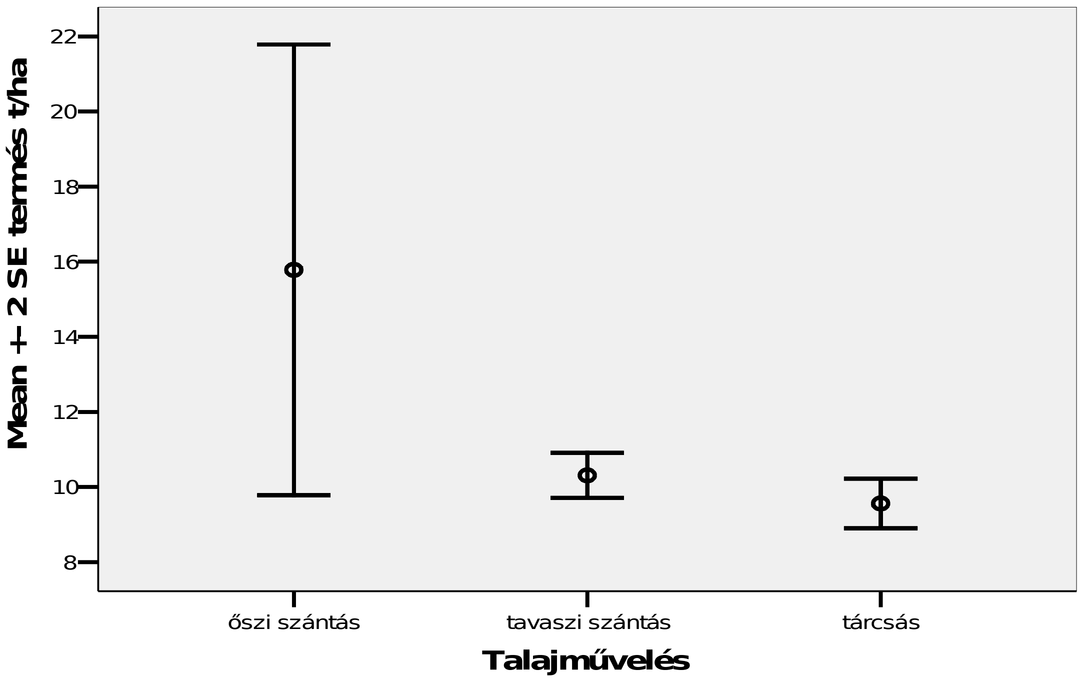
Az egyszempontos szórásanalízis F-próbája akkor ad α-szinten szignifikáns eredményt, ha ezen a szinten létezik szignifikáns kontraszt a csoportok között.

<u>Feladat:</u>

Nyissuk meg a *Termés1989.sav* fájlt, és vizsgáljuk meg, hogy a talajművelésnek milyen hatása volt ebben az évben a kukorica termésére.

A legelső lépésben ábrázoljuk talajművelési változatonként a termések átlagának hibáját (standard error), pontosabban az átlagot ± az átlag hibájának kétszeresét. Ebbe a tartományba fog legalább 95%-os valószínűséggel a valódi átlag beleesni.

Graph, Error Bar…, Simple, Summaries for groups of cases, Define, Variable: termés t/ha, Category Axis: Talajművelés, Bars Represent: Standard error of mean, Multiplier: 2.



18. ábra: Az átlag és az átlag hibájának kétszerese

Az őszi szántásos kezelés adataival valami probléma lehet, mert túlságosan nagy az átlag hibája, magába öleli a másik két kezelést is. Egyelőre hagyjuk így, és végezzük el a variancia-analízist.

Analyze, Compare Means, One-Way ANOVA, Dependent List: termés t/ha, Factor: Talajművelés, Options…: Homogeneity of variance test.

Test of Homogeneity of Variances

termés t/ha

Levene Statistic	df1	df2	Sig.
5.107	2	141	.007

10. táblázat

A Levene-teszt azt mutatja, hogy a csoportokon belül a varianciák nem egyenlők. Ezek szerint valószínűleg az őszi szántásos parcellák terméseinek szórása szignifikánsan nagyobb, mint a másik kettőé. A variancia-analízis alkalmazásának egyik feltétele nem teljesül, ezért a lenti táblázat eredményét fenntartásokkal kell kezelni. Az elvégzett analízis a talajművelés szignifikáns hatását igazolja látszólag, sig. < 0,05. Ezt csak akkor fogadhatnánk el, ha a varianciák megegyeznének.

ANOVA

termés t/ha

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1106.779	2	553.390	3.758	.026
Within Groups	20763.765	141	147.261		
Total	21870.544	143			

11. táblázat

Végezzük el a középértékek többszörös összehasonlítását (Post Hoc analízis). A többszörös összehasonlító teszteknek két nagy csoportja van: 1. a varianciáknak egyenlőknek kell lenni, 2. nem feltétel a varianciák egyenlősége. Válasszuk ki mindkét csoportból az elsőt!

One-Way ANOVA, Post Hoc…, Equal Variances Assumed: LSD, Equal Variances Not Assumed: Tamhane's T2.

Multiple Comparisons

Dependent Variable: termés t/ha

	(I) Talajmûvelés	(J) Talajmûvelés	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
LSD	õszi szántás	tavaszi szántás	5.470354*	2.477068	.029	.57336	10.36735
		tárcsás	6.219896*	2.477068	.013	1.32290	11.11689
	tavaszi szántás	õszi szántás	-5.470354*	2.477068	.029	-10.36735	-.57336
		tárcsás	.749542	2.477068	.763	-4.14745	5.64653
	tárcsás	õszi szántás	-6.219896*	2.477068	.013	-11.11689	-1.32290
		tavaszi szántás	-.749542	2.477068	.763	-5.64653	4.14745
Tamhane	õszi szántás	tavaszi szántás	5.470354	3.015757	.211	-1.99077	12.93148
		tárcsás	6.219896	3.019043	.128	-1.24821	13.68800
	tavaszi szántás	õszi szántás	-5.470354	3.015757	.211	-12.93148	1.99077
		tárcsás	.749542	.445175	.260	-.33289	1.83197
	tárcsás	õszi szántás	-6.219896	3.019043	.128	-13.68800	1.24821
		tavaszi szántás	-.749542	.445175	.260	-1.83197	.33289

*. The mean difference is significant at the .05 level.

12. táblázat

Az LSD-teszt az őszi szántás és tavaszi szántás, valamint az őszi szántás és tárcsás talajművelés között 5%-os szignifikáns különbséget mutat. A Tamhane teszt egyik kezelés pár között sem mutat szignifikáns különbséget. Mivel a varianciák különbözősége miatt LSD tesztet nem csinálhatunk, a Tamhane-teszt eredményét kell elfogadni, és kideríteni, hogy miért nem tudjuk kimutatni a talajművelés okozta hatást.

Vizsgáljuk meg az őszi szántásos kezelések adatait!

Analyze, Descriptive Statistics, Explore…, Dependent List: termés t/ha, Factor List: Talajművelés, Statistics…, Outliers.

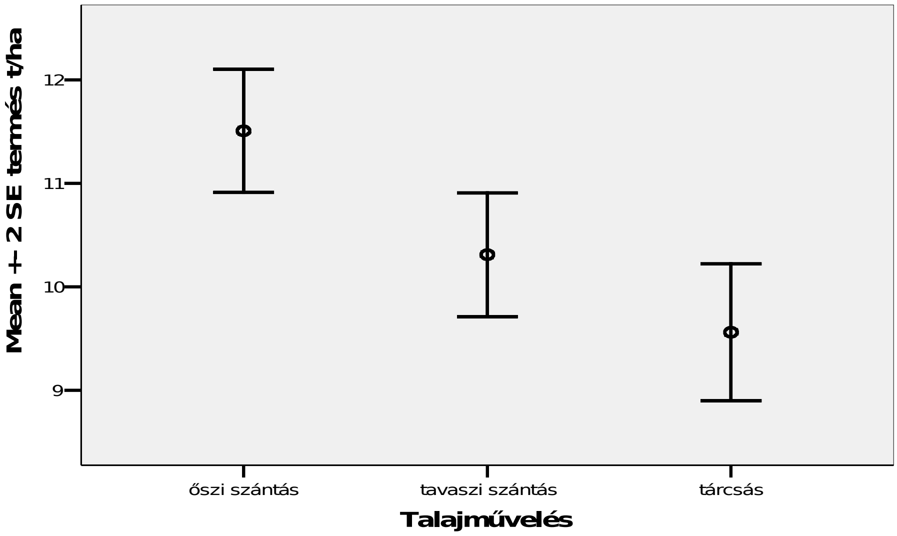
Az eredménylistából csak a kiugró értékek táblázatát mutatom be.

Extreme Values

			Case Number	Talajmûvelés	Value
termés t/ha	Highest	1	9	õszi szántás	114.41
		2	10	õszi szántás	113.51
		3	32	õszi szántás	14.395
		4	35	õszi szántás	14.392
		5	36	õszi szántás	14.286
	Lowest	1	135	tárcsás	5.355
		2	134	tárcsás	5.421
		3	136	tárcsás	5.652
		4	122	tárcsás	5.697
		5	124	tárcsás	6.059

13. táblázat

Jól látható, hogy a 9. és 10. megfigyelés adatrögzítési hiba miatt egy nagyságrenddel nagyobb, mint a többi. Javítsuk ki a hibás adatokat és ismételjük meg az analízist a legelső lépéstől kezdődően!



19. ábra: Az átlag és az átlag hibájának kétszerese

Az átlagok ebben az esetben már jól elkülönülnek egymástól. Az átlagok hibáiból képzett intervallumok már kevésbé érnek egymásba.

Test of Homogeneity of Variances

termés t/ha

Levene Statistic	df1	df2	Sig.
1.096	2	141	.337

14. táblázat

A variancia-analízis alkalmazási feltétele, a csoporton belüli varianciák egyezősége teljesül, tehát lehet variancia-analízist csinálni.

ANOVA

termés t/ha

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	92.524	2	46.262	10.087	.000
Within Groups	646.657	141	4.586		
Total	739.181	143			

15. táblázat

A variancia-analízis a talajművelés szignifikáns hatását mutatja. Az elvégzett többszörös középérték összehasonlító tesztek most már hasonló eredményt adnak. Mivel a varianciák megegyeznek, az LSD-teszt eredményét érdemes figyelembe venni, mert ennek a tesztnek ebben az esetben nagyobb a próba ereje. Ez azt jelenti, hogy a meglévő valódi különbséget nagyobb biztonsággal tudja kimutatni, mint a Tamhane teszt.

Multiple Comparisons

Dependent Variable: termés t/ha

			Mean Difference			95% Confidence Interval	
	(I) Talajmûvelés	(J) Talajmûvelés	(I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
LSD	õszi szántás	tavaszi szántás	1.196854*	.437141	.007	.33266	2.06105
		tárcsás	1.946396*	.437141	.000	1.08220	2.81059
	tavaszi szántás	õszi szántás	-1.196854*	.437141	.007	-2.06105	-.33266
		tárcsás	.749542	.437141	.089	-.11466	1.61374
	tárcsás	õszi szántás	-1.946396*	.437141	.000	-2.81059	-1.08220
		tavaszi szántás	-.749542	.437141	.089	-1.61374	.11466
Tamhane	õszi szántás	tavaszi szántás	1.196854*	.421463	.017	.17227	2.22144
		tárcsás	1.946396*	.444371	.000	.86591	3.02689
	tavaszi szántás	õszi szántás	-1.196854*	.421463	.017	-2.22144	-.17227
		tárcsás	.749542	.445175	.260	-.33289	1.83197
	tárcsás	õszi szántás	-1.946396*	.444371	.000	-3.02689	-.86591
		tavaszi szántás	-.749542	.445175	.260	-1.83197	.33289

*. The mean difference is significant at the .05 level.

16. táblázat

Az SPSS post hoc analízisei

Legkisebb szignifikáns differencia (LSD)

R.A.Fisher 1935-ben úgy módosította az egyszerű t-próbát, amennyiben a szórásanalízis F-próbája szignifikáns, akkor alkalmazhatjuk a legkisebb szignifikáns különbség (LSD) próbát, amelyben a közös hiba négyzetösszeg osztva a szabadságfokával (error mean square) becsli a varianciát. A mezőgazdasági kutatásban, a kísérletek kiértékelésben, a legrégebben használt módszer a kezelésszintek különbségének vizsgálatára. A variancia-analízis szolgáltatta Hiba MQ-ból kiszámolt SZD$_{p\%}$ -ból (

$SzD_{p\%} = t_{p\%} s_d$) levont következtetések azonban csak akkor érvényesek, ha az analízis előtt véletlenül választunk ki két kezelésátlagot, és ennek a különbségét teszteljük. Általában a legnagyobb és legkisebb értéket adó kezelések közötti különbségek akkor is nagyobbak, mint az $SZD_{p\%}$, ha a kezelések véletlen minták ugyanabból a sokaságból, tehát nincs közöttük valódi különbség. Erre a következtetésre jutott Sváb, 1981 is és a fenti hátrányok kiküszöbölésére a Duncan-tesztet említi, de az értékelés körülményes voltára hivatkozva nem foglalkozik vele. Sajnos a mezőgazdasági kutatásban is sokszor tévesen alkalmazzák az $SZD_{p\%}$-t és gyakorlatilag sorba tesztelik a kezelésszinteket, és azt nézik, hogy melyik két kezelés közötti különbség nagyobb, mint az $SZD_{p\%}$. Az így kimutatott szignifikáns különbségek igen kétes értékűek, mivel az α-hiba valószínűsége (az elsőfajú hiba elkövetésének kockázata) az összehasonlítások során halmozódik. A legkisebb szignifikáns differencia tehát a t-eloszlást alkalmazza.

Newman-teszt

D. Newman (1939) dolgozta ki az első, studentizált terjedelmeken alapuló többszörös összehasonlító tesztet. Erre az eloszlásra először ő állított fel táblázatokat, később Pearson és Hartley (1943) részletesebb táblázatot készített. Ha a próba érték szignifikáns, akkor elhagyják valamelyik szélső értéket, és a következő terjedelmet vizsgálják tovább. Newman a próbát Student (alias W.S. Gosset) (1927) cikke alapján dolgozta ki. Statisztikája:

$$q = \frac{\bar{x}_k - \bar{x}_1}{s}$$

k, ν paraméterekkel, ahol *k* a normál eloszlású populációk száma és

$$s^2 = \sum_{i,j} (x_{ij} - \bar{x}_{i.})^2$$

négyzetösszeg, amelynek szabadságfoka *ν*=*k*(*m-1*), ahol *m* a minta elemszáma.

Bonferroni-teszt

Páronkénti átlagok különbségének vizsgálatára használható, a két csoport elemszáma lehet különböző is. Lényege, hogy az α-hibához tartozó t-értéket korrigálja a független összehasonlítások számának megfelelően.

$$L = t(táblázatbeli)\sqrt{S_p^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Az α-hiba a kísérlet egészére igaz, az egyenkénti összehasonlításkor α/m az elsőfajú hiba valószínűsége. Ez a teszt is tehát a t-eloszlást alkalmazza.

Tukey-teszt, J.W. Tukey (1953)

Studentizált terjedelem tesztjében a *p*-elemű részcsoportokat ugyanazzal a kritikus értékkel hasonlítja össze. Itt a teljes vizsgálat elsőfajú hibája rögzített, és az egyes összehasonlítások elsőfajú hibája *n* növekedésével csökken, és így a másodfajú hiba nő. A Tukey teszt (1953) alapesetben egyforma minta nagyságú csoportok átlagainak különbségét tudja tesztelni, és a következő null-hipotézist vizsgálja:

$$H_0: \mu_1 = \ldots = \mu_{k,}.$$

Ezt felbontja a következő hipotézisek metszetére:

$$t_{ij} = (x_{i.} - x_{j.})\frac{\sqrt{m\nu}}{\sqrt{2s^2}}$$

$$H_{ij} = \mu_i - \mu_j = 0,$$

Ellenhipotézis: A_{ij}: $\mu_i - \mu_j \neq 0$. Mivel a minták azonos elemszámúak: $n_i = m$, ezért $\nu = k(m-1)$. Tehát a páronkénti egyenlőségeket szimultán teszteli. Statisztikája:

A H_{ij} hipotézist elfogadja α-szinten, ha $t_{ij} < t_{\alpha 1}$, ahol

$$P(t_{ij} < t_{\alpha 1} \mid H) = 1 - \alpha$$

Annak a valószínűsége, hogy a számított érték kisebb a táblázati értéknél, ha a nullhipotézis igaz, tehát a teljes elsőfajú hiba α. A $\sqrt{2}t_{\alpha 1}$ értékre α, k, ν függvényében studentized range néven táblázatokat készítettek.

A fenti összefüggésből H.Scheffé (1959) készített λ_g-re konfidencia-intervallumot. A H_g hipotézist akkor fogadjuk el, ha a konfidencia-intervallum tartalmazza a 0-t. Ezt kiterjesztett Tukey-próbának vagy T-módszernek nevezik.

Tukey és tőle függetlenül C.Y.Kramer (1956) javaslata alapján kiterjesztették nem egyenlő elemszámra. Ez a módszer a **Tukey-Kramer** módosított teszt. A teszt a Newman-Keuls-teszttel kiszámított legnagyobb különbséggel egyenlő, ezért is hívják Tukey's Honestly (őszinte, becsület) Significant Difference-nek. Később általánosították tetszőleges kontrasztokra is.

$$L = q(táblázatbeli)\frac{s_p}{\sqrt{n}}$$

Dunett (1980a) cikkében számítógépes szimulációval több szerző hasonló eljárását hasonlította össze és ezek közül a Tukey-Kramer próbát találta a legjobbnak, azaz a különböző elsőfajú hibák mellett a konfidencia-intervallum hosszát a legrövidebbnek.

H. Scheffé (1953) Scheffe-teszt

A hagyományos tesztek közé tartozik. Ő már valóban a H_g hipotéziseket vizsgálta. Az egyszerű F-próba akkor utasítja el a H_0-hipotézist, ha létezik egy a<>0 vektor, amelynél a konfidencia-intervallum nem tartalmazza a 0-t. Ha k darab összehasonlítandó csoportom van akkor $k(k$-1)/2 összehasonlítást kell végeznem. A statisztikája:

$$L = \sqrt{s_p^2(k-1)F_{(táblázatbeli)}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Ez a teszt tehát az F-eloszlást alkalmazza. A teszt tetszőleges elemszámok esetén érvényes, és a paraméterek valamennyi kontrasztjának (lineáris kombinációjának) egyidejű vizsgálatára alkalmas. A kontrasztok szimultán vizsgálata legtöbbször a szimultán konfidencia intervallumok felírásával történik, és vizsgáljuk, hogy azok tartalmazzák-e a nullát vagy nem. Mivel a kontrasztok száma végtelen, a Scheffé által kezdeményezett kiterjesztés igen lényeges általánosítást jelent. Ez a módszer a legáltalánosabb, egyedül ennek van meg az a tulajdonsága, hogy ekvivalens a szórásanalízissel. Az olyan vizsgálatokat azonban, amelyek megfelelnek a Tukey vagy Dunett-módszer eredeti kérdésfelvetésnek (egyenlő elemszámú csoportok közötti különbségek vizsgálata ill. ezek egy kontrollal való összehasonlítása a cél) érdemes ezekkel a módszerekkel elvégezni, mert erejük ilyenkor nagyobb a Scheffé-módszer erejénél. A Scheffé-módszer ereje a Bonferroni-egyenlőtlenség alapján kiterjesztett t-próbákénál is kisebb mindaddig, míg az elvégzett összehasonlítások *m* száma lényegesen meg nem haladja az elvégezhető összehasonlítások dimenzióját (Miller, 1966) *k* független csoport egyszempontos összehasonlításakor ez a dimenzió (*k*-1).

A Scheffe-teszt gyakorlati alkalmazásához nyújt nagy segítséget O BRIEN R. R. 1983 megjelent műve és LOTHAR SACHS 1985.

Dunnett-teszt

A Dunnett-teszt (1955) egy kijelölt csoportot (kontroll) hasonlít össze a többivel. Eredetileg egyenlő elemszámokra volt érvényes, de később elkészült az általánosítása egyenlőtlen elemszámokra is. Lényegét tekintve páronkénti t-tesztet végez többször, de meg kell adni egy kezdő, kontroll csoportot, és ehhez hasonlítja a többi csoport átlagát. Statisztikája:

$$\bar{x}_i - \bar{x}_o \pm |d| s_p \sqrt{\frac{2}{n}}$$

$\bar{x}_o$= kontrol csoport átlaga

Statisztikája megegyezik Tukey statisztikájával, elfogadási tartománya viszont nem.

$$P(t_{ik} < t_{\alpha 2} \mid H) = 1 - \alpha$$

Ehhez a statisztikához J. P. Shaffer készített konfidencia intervallumot, λ_g-re. Itt is a H_g:λ_g=0 hipotézist elfogadják, ha az intervallum tartalmazza a 0-t. Ezt nevezik kiterjesztett Dunett-próbának.

Student-Newman-Keuls próba

M.Keuls (1952) Módosította a Newman próbát. A próba a studentizált range eloszlást használja. A minták elemszámai egyenlők. A statisztikája megegyezik Newmanéval, az elsőfajú hiba összehasonlításonként rögzített, ezért a teljes vizsgálat elsőfajú hibája *n*-nel együtt nő.

$$w_r = q_\alpha, r, \nu \sqrt{\frac{s_p}{n}}$$

A próba teszteli, hogy mely kezelés kombinációk tartoznak egy homogén csoportba. Kiszámítása bonyolultabb, ezért célszerű számítógéppel elvégezni. Az eredmény grafikusan ábrázolható és könnyen értelmezhető. Legtöbb számítógépes program először az átlagokat sorba rendezi, kicsitől a nagy felé és vízszintes vagy függőleges vonallal jelzi a homogén csoportokat, ahol nincs szignifikáns különbség a kezelés kombinációk között. Véleményem szerint a kezelés kombinációk sorba tesztelésére a mezőgazdaságban is az egyik legjobban használható próba.

Duncan többszörös rang teszt (1955, 1965)

Itt is a homogén csoportok képzése a cél. Studentizált terjedelem eloszlást alkalmazza. Napjainkban az egyik legjobbnak tartott többszörös összehasonlító teszt. Itt is a grafikus megjelenítés nagyban segíti a kapott eredmények interpretációját. A mezőgazdasági kutatásban is potenciálisan nagy jelentőséggel bíró teszt.

Általános lineáris modell (General Linear Model)

Az általános lineáris modell a hagyományos variancia-analízis és a lineáris regresszió-analízis ötvözete. Egyetlen táblázatban jelenik meg a szórás elemzés és regresszió-analízis eredménye (17. táblázat). Napjainkban a variancia-analízisnek nagyon sokféle technikája létezik, amik lehetővé teszik a feladat sajátosságainak figyelembevételével a legalkalmasabb értékelési módszer kiválasztását. Az elemzés megbízhatósága a hiba (error) meghatározásának módjától függ, ami tulajdonképpen az eltérés négyzetösszeg (SQ) számítási technikájának függvénye. Az SPSS lehetővé teszi a kísérleti elrendezéshez hű, a felhasználó által megalkotott lineáris modell megbízható értékelését.

Tests of Between-Subjects Effects

Dependent Variable: X

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	119.248[a]	3	39.749	4.706	.006
Intercept	20563.279	1	20563.279	2434.723	.000
FAJTA	119.248	3	39.749	4.706	.006
Error	439.184	52	8.446		
Total	21121.710	56			
Corrected Total	558.431	55			

a. R Squared = .214 (Adjusted R Squared = .168)

17. táblázat

A 17. táblázat 1-4 oszlopának értelmezése logikai sorrendben:

- Total: az alapadatok négyzet összege (21 121), $\sum x^2$, szabadságfok (56). A szabadságfok itt megegyezik az adatok számával.
- Intercept: az alapadatok összegének négyzete osztva az adatok számával (20 563) $\frac{\left(\sum x\right)^2}{n}$, szabadságfok (1) valamint átlaga (20 563). Amennyiben az adatok egyáltalán nem szórnak (minden adat megegyezik), akkor a fenti két kifejezés értéke megegyezik. Az Intercept SS értékét Sváb könyveiben korrekciós tényezőként („C") említi, mely nem más, mint a kísérlet főátlagának négyzetösszege, $\sum x^2$
- Corrected Total: egyenlő Total – Intercept (558), vagyis $\sum x^2$ - $\frac{\left(\sum x\right)^2}{n}$, ez tulajdonképpen az alapadatok eltérésnégyzet-ös-

szege. Sváb könyveiben ez jelentette az „Összesen” sort. Szabadságfok (55).

- Error: ebben a példában a négy FAJTA csoporton belüli eltérés négyzetösszege (439) $\sum_i \sum_j (x_{ij} - \bar{x}_i)^2$, szabadságfok (52), valamint ennek átlaga (8,446), ami gyakorlatilag a csoporton belüli varianciák átlaga. Sváb könyveiben a Hiba, a véletlen hatása, a meg nem magyarázott hatások. Minden FAJTA csoportban 14-14 megfigyelés van. Ebből az értékből gyököt vonva megkapjuk a csoporton belüli átlagos szórás nagyságát, a kísérlet hibáját.
- FAJTA: a kezelés okozta hatás, a négy fajta átlagának eltérése a főátlagtól. $r\sigma^2_{fajta}$
- Corrected Model: a lineáris modellel becsült és a megfigyelt értékekre illesztett lineáris függvény jóságát mutatja. Eldönthető, hogy az alkalmazott modell megfelelő-e. $SS_R = \sum \left(\hat{Y}_i - \bar{Y}\right)^2 = \frac{SP^2_{xy}}{SS_x}$
- r-négyzet értéke Corrected Model SS/Corrected Total SS, (119/558).

Ha az általános lineáris modell alkalmazása során a becsült (predicted values) értékeket is elmentjük, elvégezhetjük a lineáris regresszió-analízist (18. táblázat). A regresszió eredménye megkönnyíti a GLM táblázatának újbóli értelmezését.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.462[a]	.214	.199	2.8518

a. Predictors: (Constant), Predicted Value for X

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	119.248	1	119.248	14.662	.000[a]
	Residual	439.184	54	8.133		
	Total	558.431	55			

a. Predictors: (Constant), Predicted Value for X

b. Dependent Variable: X

18. táblázat: A lineáris regresszió-analízis eredménye

A lineáris függvény illesztése során kapott eltérés négyzetösszegek teljesen megegyeznek a GLM-vel kapott értékekkel. A lineáris regresszió-analízis táblázatának (ANOVA) értelmezése:

– Total: az alapadatok eltérés négyzetösszege, szabadságfoka. Ez megegyezik a GLM Corrected Total értékével. $SS_y = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$

– Regression: a lineáris modellel becsült és a megfigyelt értékekre illesztett lineáris függvény jóságát mutatja. Eldönthető, hogy az alkalmazott modell megfelelő-e. $SS_R = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = \frac{SP_{xy}^2}{SS_x}$

– Residual: maradékok négyzetösszege, szabadságfok, négyzetösszeg átlagok. A lineáris egyenessel meg nem magyarázott hatás.

Az r-négyzet értéke 0,214. Ez a Regression SS/Total SS hányadosa (119/558).

A GLM alkalmazhatósági feltételi megegyeznek a variancia-analízis alkalmazhatósági feltételeivel:

- Független megfigyelések (a maradéknak kell függetlennek lennie a függő változó megfigyelt és becsült értékétől. Ezt ábrázolással és regresszió-analízissel tudjuk leellenőrizni.)

- Normális eloszlású sokaságok (a maradéknak kell nulla várható értékű normáleloszlást mutatni. Ezt hisztogrammal és Kolmogorov-Smirnov teszttel vizsgálhatjuk.)
- Azonos szórások (a maradékok varianciáinak meg kell egyeznie minden kezelés kombinációban, azaz cellában. Ezt Levene-teszttel ellenőrizhetjük.)

Egyváltozós variancia-analízis (Univariate…)

Segítségével egy tényező hatását lehet vizsgálni a függő változó mennyiségi alakulására. A tényező, faktor valamilyen csoportképző ismérvvel rendelkezik, a függő változó pedig legtöbbször skála típusú adat. Egyszerre több függő változót is kijelölhetünk az analízis számára.

A variancia-analízis során négyféleképpen tudjuk kiszámítani az eltérés négyzetösszegeket (SS). Római számokkal jelölöm a négy típust (I-IV.). A programban kezdőértékként a III. jelenik meg, ezt használhatjuk az egy vagy többtényezős, kiegyensúlyozott (balanced) vagy kiegyensúlyozatlan (unbalanced), teljes, azaz nincs hiányzó parcella adatú kísérletek kiértékelésekor (ez a leggyakoribb). Ez a módszer megegyezik a széles körben ismert Yates-féle módszerrel. A Yates módszer lényegében az átlagok súlyozott eltérésnégyzet technikáját használja a négyzetösszegek számításakor. Ez a módszer jól ismert a mezőgazdasági kutatásban, mivel Sváb könyveiben a variancia-analízis ismertetésekor ezt a technikát mutatja be.

Type I: ezt kell használni, ha a kezelésekben nem egyezik meg a megfigyelések száma, hiányzó parcellaadat van.

Type II: kiegyensúlyozott (balanced) egymásba ágyazott (nested) kísérleteknél kell alkalmazni.

Type IV: kiegyensúlyozott vagy kiegyensúlyozatlan kísérleteknél, ha hiányzó adat van.

Amennyiben az analízis az átlagok közötti egyenlőséget nem igazolja, szükséges az átlagok közötti különbségek kimutatása.

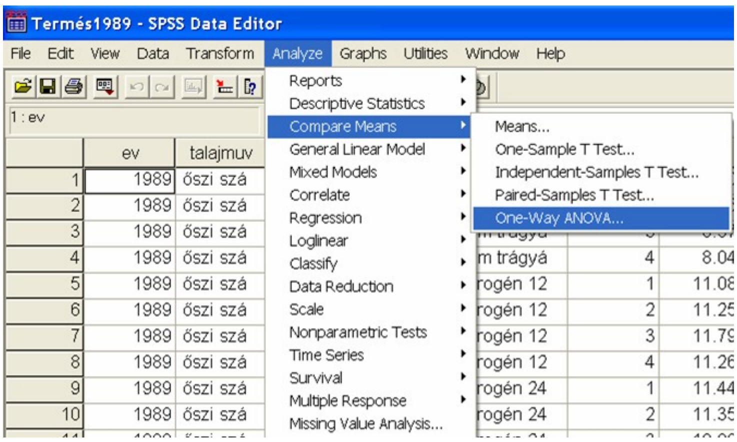
A variancia-analízist kiegészítő középérték összehasonlító teszteknek kétféle típusa létezik:

1. előzetes, ún. a priori kontrasztok és
2. az analízis után elvégezhető, ún. post hoc analízisek

A kontrasztokat tehát a kísérleti adatok elemzése előtt kell előállítani, és így elvégezni az elemzést.

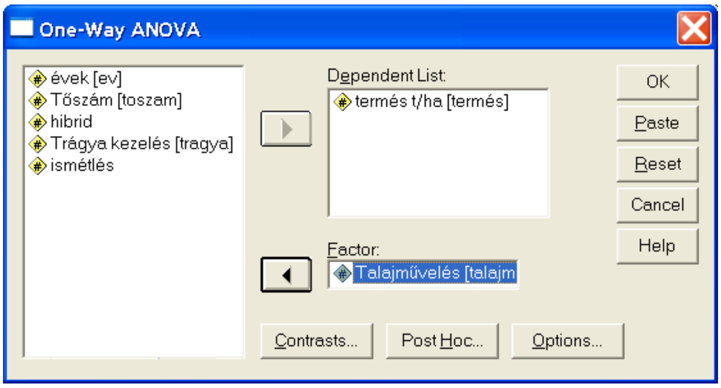
Az egyszempontos szórásanalízis F-próbája akkor ad α-szinten szignifikáns eredményt, ha ezen a szinten létezik szignifikáns kontraszt a csoportok között.

Példa:
vizsgáljuk meg, hogy három talajművelési változatban hogyan alakul a kukorica termése. Az egytényezős variancia-analízis alkalmazásához kattintsunk az ANALYZE menüpont COMPARE MEANS almenüjében az ONE-WAY ANOVA parancsra (20. ábra).



20. ábra: Egytényezős variancia-analízis

A statisztikai számítás elvégzéséhez a vizsgált függő változót helyezzük a DEPENDENT LIST ablakba, míg FACTOR-ént definiáljuk a talajművelést, hiszen a termésnek a talajművelési változatok közötti különbségét próbáljuk igazolni. Amennyiben a variancia-analízis a talajművelés szignifikáns hatását igazolja, kíváncsiak leszünk, hogy a három talajművelési változat közül vajon melyik között van lényeges (szignifikáns) különbség. A variancia-analízis után elvégzendő középérték összehasonlító tesztek helyes alkalmazásához azonban tudni kell, hogy a csoporton belüli variancia vajon megegyezik-e. Ezeket a teszteket a későbbi fejezetekben mutatjuk be.



21. ábra: A változók és a tényező megadása

19. táblázat: A variancia-analízis eredménye

ANOVA

termés t/ha

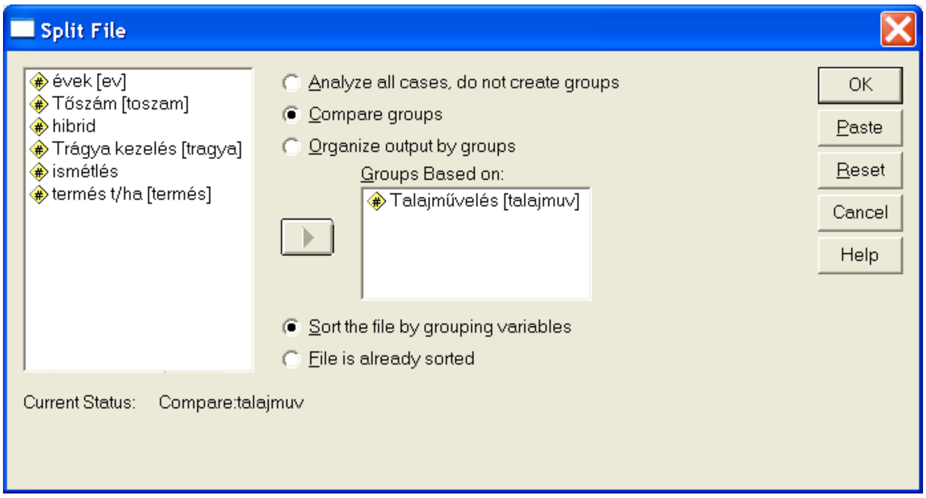
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	92.524	2	46.262	10.087	.000
Within Groups	646.657	141	4.586		
Total	739.181	143			

A fenti táblázat a szórás-elemzés eredményét mutatja. 5%-os elsőfajú hibát választva, megállapítható, hogy a talajművelési változatokban a kukoricatermése szignifikánsan különbözik. Hangsúlyozzuk, hogy az F-próba eredménye csak akkor fogadható el, ha a hiba normál eloszlású és a csoportokon belüli varianciák megegyeznek.

A modell érvényességének vizsgálata

Normalitás vizsgálat

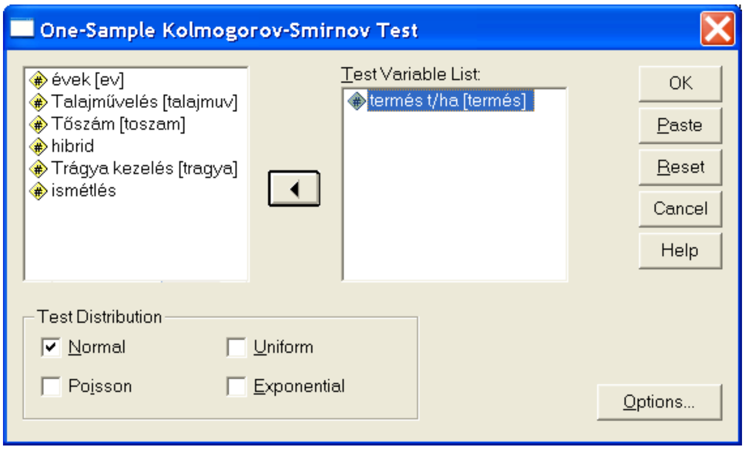
A variancia-analízis alkalmazhatóságának feltétele, hogy a függő változó normális eloszlású legyen, pontosabban a különböző kezelések mintáinak



22. ábra. Az adatbázis megosztása

(lényegében azonban a hibának, vagy eltérésnek) kell normáleloszlásúnak lenni. A kezeléscsoportok elkülönített elemzéséhez meg kell osztani az adatbázist a Data, majd a Split File… kiválasztása után a megjelenő ablakban válasszuk a Compare groups rádiógombot, és a Groups Based on: ablakba helyezzük a „talajművelés” változót. Az Ok gomb megnyomása után térjünk vissza az adatbázis ablakhoz.

Normalitás vizsgálatot az SPSS-ben többféleképpen is végezhetünk, pl. Analyze/Nonparametric test/1-Sample K-S… a megjelenő párbeszédablakban (23. ábra) adjuk meg a vizsgálandó változót, és jelöljük be a normáleloszlást (alapesetben ez van megjelölve). A nullhipotézisünk ennek megfelelően az lesz, hogy a vizsgált változó eloszlása nem különbözik a normális eloszlástól. Válasszuk a szignifikancia szintet 5%-osra, és végezzük el az analízist az OK gomb megnyomásával. Az eredmény a táblázatban látható.



23. ábra. Az egymintás Kolmogorov-Smirnov teszt

20. táblázat: A változó eloszlásának vizsgálata Kolmogorov-Smirnov próbával

One-Sample Kolmogorov-Smirnov Test

Talajművelés			termés t/ha
őszi szántás	N		48
	Normal Parameters[a,b]	Mean	11.50673
		Std. Deviation	2.060577
	Most Extreme Differences	Absolute	.127
		Positive	.095
		Negative	-.127
	Kolmogorov-Smirnov Z		.882
	Asymp. Sig. (2-tailed)		.418
tavaszi szántás	N		48
	Normal Parameters[a,b]	Mean	10.30988
		Std. Deviation	2.068890
	Most Extreme Differences	Absolute	.227
		Positive	.148
		Negative	-.227
	Kolmogorov-Smirnov Z		1.574
	Asymp. Sig. (2-tailed)		.014
tárcsás	N		48
	Normal Parameters[a,b]	Mean	9.56033
		Std. Deviation	2.287441
	Most Extreme Differences	Absolute	.263
		Positive	.136
		Negative	-.263
	Kolmogorov-Smirnov Z		1.821
	Asymp. Sig. (2-tailed)		.003

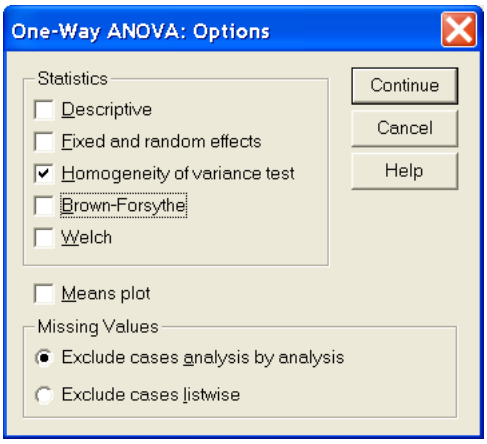
a. Test distribution is Normal.

b. Calculated from data.

Az ASYMP. SIG. (2-TAILED) sort tanulmányozva elmondható, hogy az őszi szántásos parcellák kukoricatermése normáleloszlású (*p>0,05*), azonban a másik két talajművelési változat (tavaszi szántás, tárcsás) nem normáleloszlású, mert *p*<*0,05*, vagyis elvetjük a nullhipotézist. A kapott eredmény alapján ebben az esetben nem szabadna variancia-analízissel értékelni a kísérletet. Vajon mi lehet ennek az oka? Sokszor a kiugró értékek, vagy adatrögzítési hiba okozza.

Homogenitás vizsgálat

A varianciák homogenitásának ellenőrzésére az OPTIONS parancsgomb megnyomása után, a HOMOGENITY OF VARIANCE megjelölésével történik (24. ábra). A homogenitást Levene-teszttel állapíthatjuk meg. Visszatérve a variancia-analízis párbeszédablakhoz, és az OK gomb megnyomása után megkapjuk az eredményeket.



24. ábra. Homogenitás vizsgálat

Amennyiben a szignifikancia szintet előzetesen 5%-on rögzítettük, a talajművelés esetén megtartjuk a nullhipotézisünket (*p*>*0,05*) (21. táblázat). Ez azt jelenti, hogy majd a kezelésátlagok összehasonlítása során nyugodtan alkalmazhatjuk az egyenlő varianciákat feltételező teszteket. Abban az esetben, ha a Levene-teszt a varianciák különbözőségét igazolja (22. táblázat), nem használhatjuk a Fischer-féle tesztet. Ilyenkor robosztusabb próbát kell választani, pl. Brown-Forsythe vagy Welch próbát (WELCH, 1938).

21. táblázat: A talajművelési változatokon belüli varianciák egyenlőségének ellenőrzése

Test of Homogeneity of Variances

termés t/ha

Levene Statistic	df1	df2	Sig.
1.096	2	141	.337

22. táblázat: A talajművelési változatokon belüli varianciák egyenlőségének ellenőrzése

Test of Homogeneity of Variances

termés t/ha

Levene Statistic	df1	df2	Sig.
5,144	2	141	,007

A statisztika panelen különböző kiegészítő számításokat kérhetünk. Leíró statisztika (Descriptive): esetek száma, átlag, szórás, az átlag hibája, minimum, maximum, 95%-os konfidencia intervallum mindenegyes csoportra. Fix és véletlen hatások (Fixed and random effects):

Brown-Forsythe próba

Ezt a próbát BROWN-FORSYTHE 1974-ben közölte először. A szórások különbözősége esetén meg kell vizsgálni, miért különbözik a szórás, milyen szakmai magyarázatot lehet rá adni. Ha a szórások különbözőségének semmilyen logikai vagy szakmai okát nem tudjuk megadni, nagy valószínűséggel a szórások véletlenül vagy valamilyen kísérleti hiba miatt különböznek.

A Welch és Brown-Forsythe-próba mezőgazdasági alkalmazásával még nem találkoztunk, ezért a több éves kutatómunka tapasztalatai alapján itt ragadjuk meg az alkalmat, hogy a használatukhoz néhány tanácsot adjunk. Ha a csoporton belüli szórás négyzetek (varianciák) nem egyformák nyugodtan használhatjuk a kezelésátlagok egyenlőségének tesztelésére bármelyiket a kettő közül. A legjobb, ha mindkettőt kipróbáljuk és összehasonlítjuk az eredményeket.

Válasszuk ki az OPTIONS párbeszédablakban (24. ábra) a Brown-Forsythe és Welch próbákat és futtassuk le a programot újból. A kapott eredményeket lentebb láthatjuk.

23. táblázat: A kezelés középértékek összehasonlítása robosztus tesztekkel

Robust Tests of Equality of Means

termés t/ha

	Statistic[a]	df1	df2	Sig.
Welch	3,238	2	83,571	,044
Brown-Forsythe	3,725	2	49,027	,031

a. Asymptotically F distributed.

Ebben az esetben a két teszt ugyanazt az eredményt adta, ha különbség lett volna a két eredmény között, tovább kell folytatni az értékelést. Ilyenkor szélsőséges esetben a Welch-próba szignifikáns különbséget mutathat a kezelés átlagok között, míg a Brown-Forsythe-próba nem. Mi lehet ennek az oka? Ez akkor következik be, ha a csoportok varianciája nagyon nagymértékben különbözik egymástól. Ilyenkor az elkülönített (separate) variancia tesztek a szabadságfok csökkentésével válaszolnak, és ezzel rontják a teszt eredményét. A varianciák nagyon nagy mértékű különbözőségét legtöbbször a csoportokon belüli kiugró értékek okozzák. A kiugró értékek zavaró hatását többféleképpen szűrhetjük ki. Az egyik hatásos eszköz a csonkított (trimmed) teszt, amikor mindenegyes csoportból elhagyjuk a legnagyobb és legkisebb érték 15%-át. A csonkolás mértékét szakmai megfontolások miatt tetszőlegesen megváltoztathatjuk. A csonkolás után megismételt Brown-Forsythe próbában a szabadságfokok száma nőni fog és a teszt eredménye javul (24. táblázat). A fenti feltételek esetén a szórás hagyományos meghatározása helyett a *ROBUST SD* és *WINSORIZED SD* kiszámítása jobb becslést ad a csoporton belüli szórás nagyságára. Ezek a próbák kevésbé érzékenyek a kiugró értékekre. A különböző módon kiszámított szórások összehasonlítása közvetett módon, a csoporton belüli varianciák egyenlőségére vagy egyenlőtlenségére is rámutat. Szántóföldön tőszám kísérleteknél, ahol a varianciák egyezősége nem várható, a Welch vagy

Brown-Forsythe- által kidolgozott variancia-analízist kell alkalmazni.

24. táblázat: A kezelés középértékek összehasonlítása robosztus tesztekkel a csonkolás után

Robust Tests of Equality of Means

termés t/ha

	Statistic[a]	df1	df2	Sig.
Welch	9,905	2	93,797	,000
Brown-Forsythe	10,087	2	139,613	,000

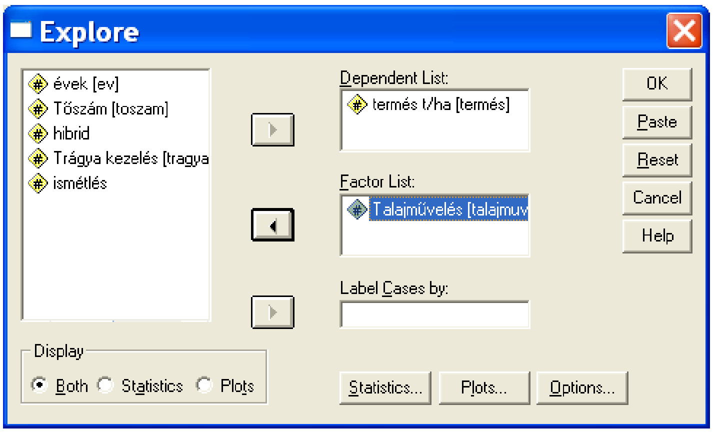
a. Asymptotically F distributed.

Kiugró értékek vizsgálata

Az előző fejezetben láttuk, hogy a kiugró értékek milyen nagymértékben tudják megzavarni a variancia-analízis eredményét. Ezért a statisztikai elemzések első és egyik legfontosabb lépése a kiugró értékek ellenőrzése. Az SPSS ennek ellenőrzésére is kínál lehetőséget.

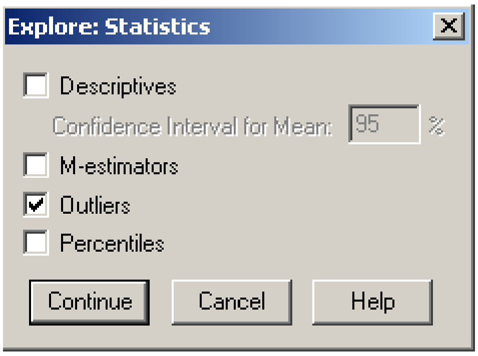
Az elemzés első lépéseként nézzük meg, hogy a kukoricatermés adataiban találunk-e kiugró értéket. Van-e vajon adatrögzítési, gépelési hiba?

A kiugró értékek ellenőrzése az ANALYZE / DESCIPTRIVE STATISTICS blokkjában az EXPLORE parancs szolgál.



25. ábra. Az kiugró értékek vizsgálata

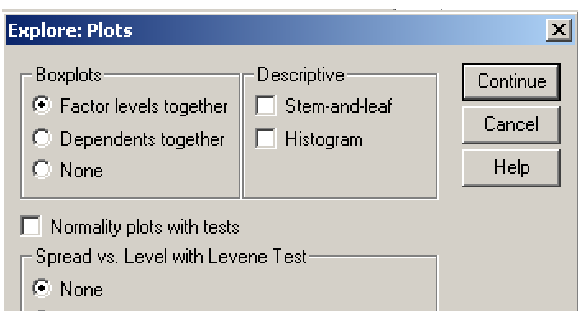
A DEPENDENT LIST mezőbe helyezzük a vizsgálni kívánt változót (változókat). Mivel mind a három talajművelésre ellenőrizni kívánjuk, hogy a kukoricatermés adatok tartalmaznak-e kiugró értéket, a FACTOR LIST mezőbe helyezzük a „*termés*” változót. Ezzel érjük el, hogy a program a kiugró értékeket talajművelési változatonként külön-külön és nem összevont állományon ellenőrizze. A STATISTICS nyomógombra kattintva a megjelenő beszédpanelban válasszuk ki az OUTLIERS lehetőséget (26. ábra).



26. ábra. Az kiugró értékek vizsgálata

A beállítások elvégzése után futtassuk le a programot. A 25. táblázat talajművelési változatonként az öt legnagyobb és legkisebb értéket tartalmazza. A kiugró értékek ugyanis biztos, hogy itt keresendők, hiszen azok vagy sokkal nagyobbak, vagy sokkal kisebbek, mint a többi érték a mintában. A táblázatból jól látszik, hogy az őszi szántásos adatokban a 9. és 10. adat kiugró érték, adatrögzítési hiba miatt a tizedesvessző eggyel jobbra csúszott. A másik két talajművelésnél nem találunk kiugró eseteket. Hasonlóan végezhetjük el más változó esetében is a vizsgálatot.

A kiugró értékek ellenőrzésének egy másik lehetséges módja az adatok grafikus megjelenítése. Az EXPLORE párbeszédpanel ablakából válasszuk a PLOTS lehetőséget (25. ábra). A BOXPLOTS panelrészben jelöljük



27. ábra. A kiugró esetek grafikus ábrázolása meg a FACTOR LEVELS TOGETHER lehetőséget, majd a CONTINUES gombra kattintva menjünk vissza a főablakba, és ott az OK gomb megnyomásával hagyjuk jóvá a statisztika számítását. Ezt a vizsgálatot mindhárom talajművelésre futtatva a 28. ábrat kapjuk.

A megjelenő ábra legalsó és legfelső vonala a különböző talajművelésekben – a kiugró értékeket nem számítva – a mért legnagyobb és legkisebb értékeket jelölik. Az ábrán e két vonal alatt és fölött jelennek meg a kiugró értékek (9. 10. adat őszi szántás) pontok formájában ábrázolva[1].

[1] A kiugró értékek ellenőrzésére az SPSS még számos lehetőséget kínál, ezek további bemutatására most nem kerül sor.

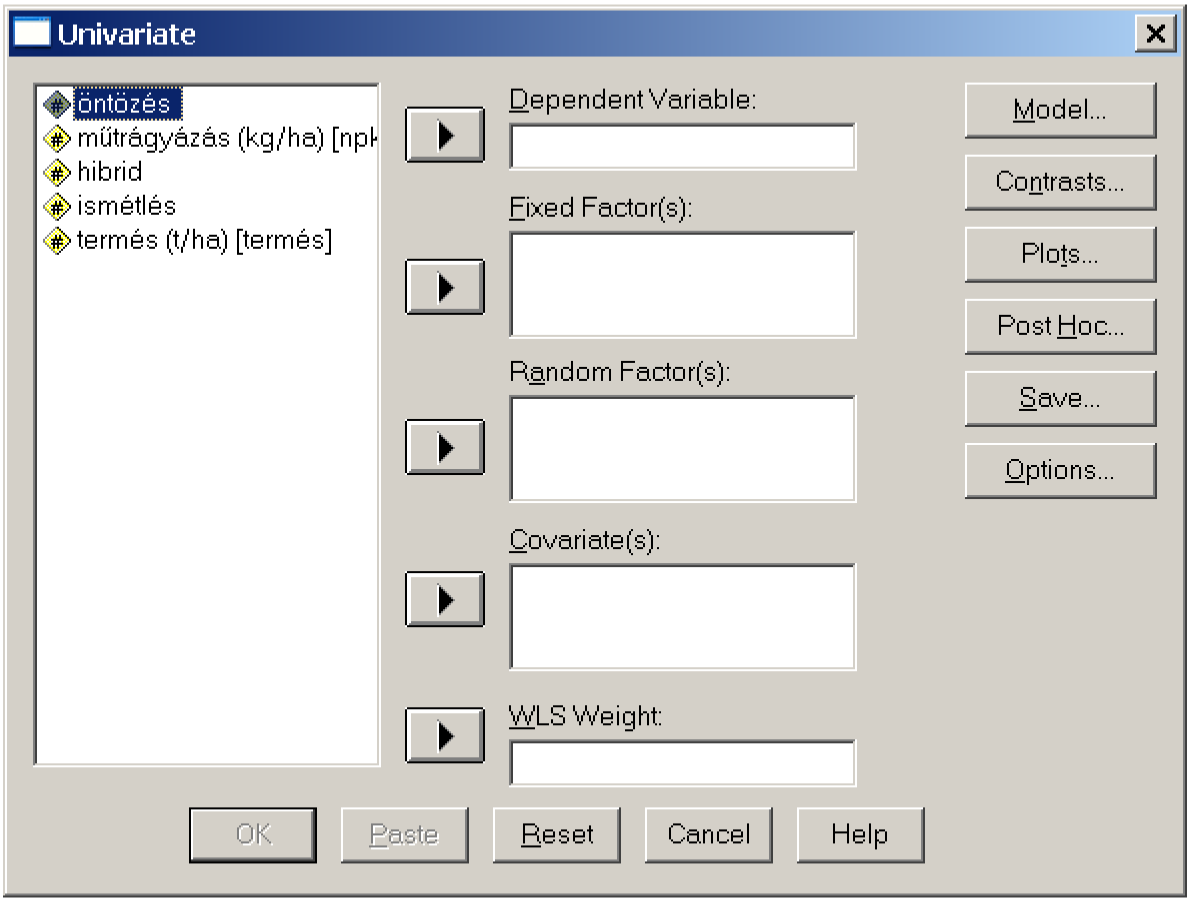
25. táblázat: A kiugró értéket ellenőrző táblázat

Extreme Values

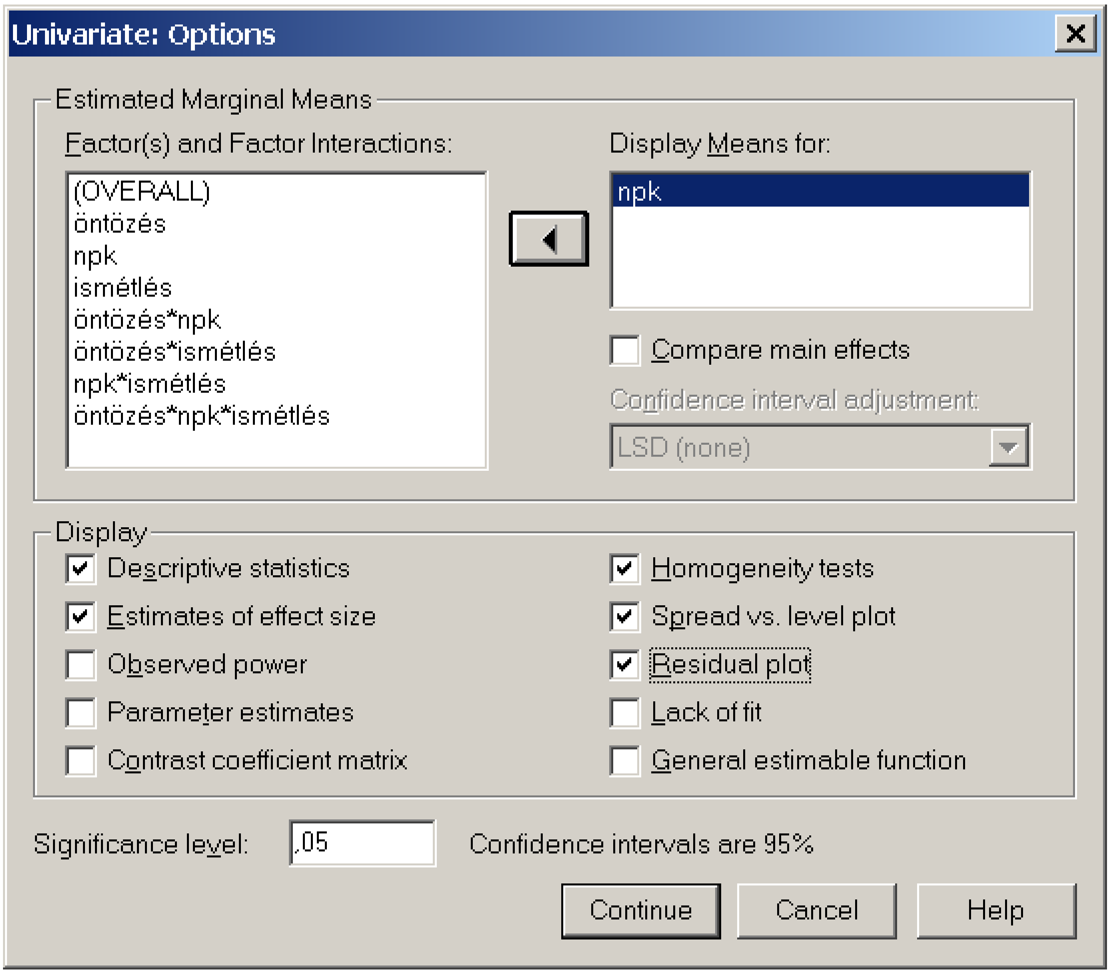
Talajművelés				Case Number	Value
termés t/ha	őszi szántás	Highest	1	9	114,41
			2	10	113,51
			3	32	14,395
			4	35	14,392
			5	36	14,286
		Lowest	1	1	7,906
			2	2	7,957
			3	13	8,010
			4	4	8,043
			5	38	8,248
	tavaszi szántás	Highest	1	84	13,118
			2	83	12,859
			3	78	12,746
			4	82	12,672
			5	80	12,569
		Lowest	1	88	6,715
			2	87	6,865
			3	85	6,929
			4	49	7,078
			5	62	7,099
	tárcsás	Highest	1	118	12,070
			2	115	11,966
			3	117	11,936
			4	129	11,859
			5	119	11,801
		Lowest	1	135	5,355
			2	134	5,421
			3	136	5,652
			4	122	5,697
			5	124	6,059

28. ábra. A kiugró értékek grafikus ellenőrzése (box-plot ábra)

Egyváltozós lineáris modell (GLM Univariate…



Lehetőségek (Options)



A lineáris modell validálásához kapunk segítséget a lehetőségek panelen. A legfontosabb lehetőségek:

- Leíró statisztika (descriptive statistics)
- A kezelés nagyságának becslése (estimates of effect size): ez a parciális eta meghatározását jelenti. Megmutatja, hogy a független változó (ami gyakran nominális típusú változó) hány százalékban befolyásolja a függő változó varianciáját. A számítás menete: a kezelés SS (eltérés négyzetösszeg) osztva az összes SS-vel. A mutató értéke 0-1.00 között lehet. A nulla függetlenséget az 1.00 determinisztikus kapcsolatot jelent. A köztes értékek a sztochasztikus kapcsolat erősségét becsülik. Minél közelebb van egyhez, annál erősebb a két változó közötti összefüggés. Ennek a mutatónak természetesen nincs előjele, így a kapcsolat irányát nem lehet vele jellemezni.

- Homogenitás teszt: a kezeléskombinációk csoporton belüli varianciáját teszteli. A variancia-analízis egyik alkalmazhatósági feltételét tesztelhetjük vele. A H_0: a varianciák egyformák.
- Maradékok ábrázolása (residual plot): a standardizált maradékokat ábrázolja a megfigyelt, ill. becsült értékek függvényében.

Többváltozós variancia-analízis, (Multivariate…)

Több kvantitatív tulajdonság együttes figyelembe vétele alapján kívánjuk kimutatni a kezelések hatása közötti különbségeket. Két kezelés közötti különbség szignifikanciájának vizsgálata, D^2 általánosított távolság tesztelése F-próbával. A DA a MANOVA határesete. Hotelling T^2.

Ismételt mérési modellek (Repeated Measures)

Variancia komponensek (Variance Components)

Kísérletek tervezése és értékelése általános lineáris modellel

Az alábbi fejezetekben a mezőgazdasági, földművelési, növénytermesztési, nemesítési, fajta összehasonlító, stb. kísérletek laboratóriumi és különböző szántóföldi kisparcellás elrendezéseinek értékelését mutatom be a teljesség igénye nélkül. Az ismertetésre kerülő klasszikus elrendezések tanulmányozása és megértése segítséget nyújt a jövőbeli kísérletek megtervezéséhez és kiértékeléséhez. A fejezetekben az elrendezés rövid ismertetése után megadom a kísérlet vázrajzát, a matematikai modell leírását és a GLM-táblázat szerkezetét valamint a kiértékeléshez szükséges parancsokat, amit a parancsszerkesztő (syntax editor) ablakban lehet futtatni. Az elrendezéshez hű kiértékelés legfontosabb parancsa a DESIGN, ezért ezt a GLM-táblázat szerkezetében is megadom. Ezt követi a mintapélda GLM-táblázata, melyben a tényezők, négyzetösszegek, szabadságfokok, átlagos négyzetösszegek, F-próbák eredményei, valamint a szignifikancia szintek láthatók.

A variancia-analízis modellben a függő változókat magyarázzuk független változó(k) segítségével. A magyarázat a függő változó teljes heterogenitásának[2] két részre bontását jelenti. A teljes heterogenitás egyik része az, amelynek „okai” a független változók, a másik heterogenitás-rész pedig az, amelynek „okait” az egyéb, általunk nem vizsgált tényezők tartalmazzák. Ez utóbbit sokszor a véletlen hatásaként is emlegetik. A heterogenitás mérésére többféle mérőszám szolgál:

(1) *range* (terjedelem); a legnagyobb és legkisebb érték közötti távolság

(2) *átlagos eltérés;* ;

(3) *szórás;* ;

(4) *variancia- vagy szórásnégyzet;* .

[2] A változó heterogenitása azt jelenti, hogy az adott változó nem konstans.

Ebből látszik, hogy a függő változónak magas (intervallum- vagy arányskála) mérési szintűnek kell lenni. Attól függően, hogy a független változók alacsony vagy magas mérési szintű-ek, eltérő magyarázó modelleket kell felépíteni. Ha ugyanis a független változóink nominális vagy ordinális mérési szintűek, akkor ***variancia-analízissel***l kereshetjük a magyarázatot a függő változó „viselkedésére”. Ha a független változók is magas mérési szintűek, akkor ***regresszió-analízist*** alkalmazhatunk. (Ha a függő változó alacsony mérési szintű, a magyarázatra szolgáló változók pedig magas mérési szintűek, akkor ***diszkriminancia-analízist*** használhatunk.)

A variancia-analízis során kettőnél több sokaság középértékeinek minta alapján történő összehasonlítása történik. Ezért nevezik a kétmintás t-próba általánosításának.

A variancia-analízis modellek olyan rugalmas statisztikai eszközök, amelyek alkalmasak valamely kvantitatív (numerikus vagy intervallum skálájú) változónak (függő változónak) egy vagy több nem feltétlenül kvantitatív változóval (független változók) való kapcsolata elemzésére. Arra vagyunk kíváncsiak, hogy van-e hatása a független változóknak a függő változóra, és a hatás különbözik-e vagy egyforma? A hatás, kapcsolat függvényszerű leírása azonban nem célunk, még akkor sem, ha a független változók kvantitatívak. A regresszió-analízistől két szempont különbözteti meg a variancia-analízist:

- A vizsgált független változók kvalitatívak is lehetnek (pl. a vizsgált személy neme, lakhelye stb.). Ebben az esetben ugyanis regresszió-analízis nem alkalmazhatunk.
- Még ha a függő változók kvantitatívak is, nem cél a független változóval való kapcsolat természetének feltárása. A szórásanalízist tekinthetjük a regresszió-analízis vizsgálat megelőző vizsgálatának, ha ugyanis pozitív összefüggést kapunk a függő és független változó kapcsolatára, akkor van értelme vizsgálni az összefüggés jellegét.
-

Alapfogalmak

Nézzük át azokat az alapfogalmakat, amelyeket a variancia-analízis során használunk.

a) *Faktor:* Faktornak nevezzük a vizsgálatba bevont független változókat, pl. különböző kezeléseket, tényezőket.

b) *Faktor szint:* A faktor értékkészletének az eleme, mely beállítása mellett vizsgálhatjuk meg a függő változónkat. A kezelések szintjei, pl. műtrágyaadagok.

c) *Kvalitatív és kvantitatív faktorok:* Ha a faktorszintek nem numerikusak vagy intervallum skálájúak, akkor kvalitatív, ellenkező esetben kvantitatív faktorokról beszélünk.

d) *Kezelések (cellák)*: Egyfaktoros esetekben a kezelések megfelelnek a faktorok szintjeinek, többfaktoros esetben a figyelembe vett faktorok szintjeiből előálló kombinációk a kezelések. Pl. amikor a 2 faktor műtrágyaadagok és öntözési módok, akkor a kezelések a (műtrágyaadagok, öntözési módok) összes lehetséges kombinációjából áll.

e) *Interakció*: Két változó kapcsolatában akkor áll fenn interakció (kölcsönhatás), ha x_1 változó hatása függ az x_2 változó szintjétől és fordítva.

f) *Egy szempontos variancia-analízis*: Variancia-analízis, ahol csak egy faktor van.

g) *Több szempontos variancia-analízis*: Variancia-analízis, ahol kettő vagy több faktor van.

h) *Egyváltozós variancia-analízis*: ANOVA technika, amely egy függő változót használ.

i) *Többváltozós variancia-analízis*: ANOVA technika, amely kettő vagy több függő változót használ.

A variancia-analízis alkalmazásának feltételei

A variancia-analízis adott n számú populáció középértékeinek minták alapján történő összehasonlítására szolgál (a kétmintás t-próba általánosításának tekinthető). A középértékre vonatkozó hipotézisek a következők:

H_0: azoknak a populációknak a középértékei, amelyekből a minták származnak azonosak: ; H_A: legalább egy olyan középérték pár van, ahol a középértékek nem tekinthetők azonosnak: legalább egyszer .

A variancia-analízis adatait a szokásos jelölésekkel 26. táblázat tartalmazza.

A statisztikai mintára alapozott variancia-analízis a következő lépésekben végezhető el:

1. A variancia-analízis modell felállítása.
2. A variancia-analízis kiszámítása, az F-próba.
3. A modell érvényességének ellenőrzése.
4. Amennyiben az F-próba szignifikáns, középértékek többszörös összehasonlítása.

26. táblázat. A variancia-analízis adatai.

Sorszám	Populáció		Minta			
	várhatóérték	variancia	elemszám	mintaelemek	középértékek	variancia
1	μ_1		r_1	$X_{11}\ X_{12} \ldots X_{1r_1}$	$\overline{X}_1$	s_1^2
2	μ_2		r_2	$X_{21}\ X_{22} \ldots X_{2r_2}$	$\overline{X}_2$	s_2^2
.	.	.	.	.	.	.
n	μ_n		r_n	$X_{n1}\ X_{n2} \ldots X_{nr_n}$	$\overline{X}_n$	s_n^2

1. A variancia-analízis modell felállítása

A módszer alapgondolata szerint a modellben a mérési, megfigyelési értékeket összegként tekintjük. A k megfigyelés mindegyikére egy-egy modellegyenlet írható fel, amelynek alapján a mintaelemeken mért, ill. megfigyelt X_{ij} értékek felbonthatók a modell által meghatározott részekre és a hibára. A modell által meghatározott rész a szisztematikus hatásokat tartalmazza, a hibakomponens pedig a véletlen hatást jelenti.

A variancia-analízis legegyszerűbb modelljében a vizsgálatban szereplő n számú populációból egyszerűen véletlen mintát veszünk, majd a mintánkénti középértékeket hasonlítjuk össze, ezt nevezzük *egyszempontos variancia-analízisnek* (kísérlet esetén teljesen véletlen elrendezésnek). Az elrendezés modell-egyenlete:

$$X_{ij} = \mu + A_i + e_{ij}$$

ahol X_{ij} az i -edik minta j -edik eleme $(i = 1,...,n \quad j = 1,...,r_i)$; $\bar{\mu}$ a kísérlet vagy minta főátlaga; A_i az i-edik mintához tartozó populáció hatása (növelheti vagy csökkentheti a főátlagot); e_{ij} véletlen hatás. Ebben a modellben a modell által meghatározott rész, csak az i -edik mintához tartozó populáció várható értékét tartalmazza, tehát szisztematikus különbséget csak a populációk várható értékei között feltételezhetünk. A véletlen okozta hatásokat a hibakomponens tartalmazza. Amennyiben teljesülnek a variancia-analízis alkalmazásának feltételei, akkor A_i összege nulla, és e_{ij} normális eloszlású nulla várhatóértékű sokaság, és független a blokk és kezeléshatástól.

A variancia-analízis alkalmazásának feltételei:

(a) **Az egyes kezelésekhez tartozó mintáknak függetleneknek kell lenniük.** Ezt leginkább a kísérleti elrendezéssel, randomizálással biztosíthatjuk A kísérleti elrendezésekről a vonatkozó fejezetben szólunk.

(b) **Az e_{ij} maradék normális eloszlású, nulla várható értékű, független a blokk, kezeléshatástól és a függő változótól.** Attól, hogy egy normáleloszlású mintához egy konstans értéket hozzáadunk, vagy abból levonunk, az eloszlás és a minta szórása nem változik. A normalitás vizsgálatát korábban ismertetett módszerek valamelyikével ellenőrizhetjük. (Megjegyezzük, hogy a matematikai-statisztikai kézikönyvek az ANOVA-t robosztus eljárásnak tekintik, s azt állítják, hogy a függő változónak nem kell normális eloszlásúnak lennie). Ha matematikailag korrekt módon akarjuk az ANOVA-t használni, akkor a függő változót normális eloszlásúvá transzformálhatjuk.

(c) **A minták szórásnégyzetei egyezzenek meg .** (Az SPSS programnál ezt a homogenitást a Levene teszt alapján tesztelhetjük: ANALYZE/COMPARE MEANS/ONE-WAY ANOVA menüben az OPTIONS alatt jelölhetjük ki.)

Egytényezős kísérletek

Teljesen véletlen elrendezés

Az ismétlésekből nem lehet vagy nem célszerű blokkot képezni, annak ellenére, hogy minden kezelésből azonos számú ismétléssel rendelkezünk. A kísérleti vázrajzot felesleges megadni. A példa húsz tyúk tojástermelését mutatja be 4 kezeléses (v) 5 ismétléses (r) kísérletben.

Az elrendezés matematikai modellje:

$Y_{ij} = m + K_j + e_{ij}$

ahol:

Y_{ij} = egy tyúk tojástermelése (db/tyúk)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

K_j = a kezelés hatása a tyúkok tojástermelésére

e_{ij} = a kísérlet hibája, a csoporton belüli szórás

Tényező	SS	df	MS	F	Sig.	DESIGN
Korrigált modell		?				
Eltérés		1				
Kezelés (csop. között)		v-1				kezelés
Hiba (csoporton belül)		v(r-1)				
Összesen		rv				
Korrigált összesen		rv-1				

27. táblázat: *A GLM-táblázat szerkezete teljes véletlen elrendezésben*

```
UNIANOVA
  tojás  BY kezelés
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = kezelés .
```

28. táblázat: A teljesen véletlen elrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: Tojástermelés

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3784.000[a]	3	1261.333	3.859	.030
Intercept	133824.8	1	133824.8	409.469	.000
KEZELÉS	3784.000	3	1261.333	3.859	.030
Error	5229.200	16	326.825		
Total	142838.0	20			
Corrected Total	9013.200	19			

a. R Squared = .420 (Adjusted R Squared = .311)

29. táblázat: A teljesen véletlen elrendezés eredménytáblázata, v=4, r=5

Véletlen blokkelrendezés

Egy csoportba minden kezelésből egy megfigyelési egység szerepel. Ez a csoport a blokk, amely egyben az ismétlést is jelenti. A blokkok száma így egyenlő az ismétlések számával. A blokkokon belül a kezeléseket randomizálni kell.

Műtrágyázás

1	2	3	4	5	6	7	(5)	ismétlés
2	7	5	6	7	3	1	(4)	
5	6	2	3	1	7	4	(3)	
3	1	4	5	7	6	2	(2)	

4	3	1	7	2	5	6	(1)

29. ábra: Véletlen blokk elrendezés terve 7 kezelés (v) 4 ismétlésben (r)

Az elrendezés matematikai modellje:

$Y_{ij} = m + R_i + M_j + e_{ij}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitása, hogyan változik a talaj termékenysége fentről lefelé haladva

M_j = a műtrágyázás hatása a cirok termésére

e_{ij} = a kísérlet hibája

Tényező	SS	df	MS	F	Sig.	DESIGN
Korrigált modell		?				
Eltérés		1				
Ismétlés		r-1				ismétlés
Kezelés (csop. között)		v-1				kezelés
Hiba		(r-1)(v-1)				
Összesen		rv				
Korrigált összesen		rv-1				

30. táblázat: *A GLM-táblázat szerkezete véletlen blokk elrendezésben*

```
UNIANOVA
  termés  BY ismétlés kezelés
  /METHOD = SSTYPE(3)
```

```
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/DESIGN = ismétlés kezelés.
```

31. táblázat: A véletlen blokkelrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: cirok kg/parcella

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	124.943[a]	8	15.618	7.416	.000
Intercept	5424.027	1	5424.027	2575.511	.000
ISMÉTLÉS	17.993	5	3.599	1.709	.193
KEZELÉS	106.950	3	35.650	16.928	.000
Error	31.590	15	2.106		
Total	5580.560	24			
Corrected Total	156.533	23			

a. R Squared = .798 (Adjusted R Squared = .691)

32. táblázat: A véletlen blokkelrendezés eredménytáblázata, v=4, r=6

Latin négyzet elrendezés

4, 5, 6, 7 és 8 kezelés összehasonlítására alkalmas kísérleti elrendezés, ha az ismétlések száma azonos a kezelések számával. Alkalmazható háromtényezős kísérletek elrendezésére is, ha a kezelések között nincs kölcsönhatás.

1	2	3	4	5	6
6	1	2	3	4	5
5	6	1	2	3	4
4	5	6	1	2	3
3	4	5	6	1	2
2	3	4	5	6	1

30. ábra: 6x6 latinnégyzet vázrajza

Az elrendezés matematikai modellje:

$Y_{ijk} = m + S_i + O_j + K_k + e_{ijk}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

S_i = blokk ill. ismétlés hatás a talaj heterogenitása, hogyan változik a talaj termékenysége fentről lefelé haladva

O_j = blokk ill. ismétlés hatás a talaj heterogenitása, hogyan változik a talaj termékenysége jobbról balra haladva

K_k = kezeléshatás

e_{ijk} = a kísérlet hibája

Tényező	SS	df	MS	F	Sig.	DESIGN
Korrigált modell		?				
Eltérés		1				
Sor		r-1				sor
Oszlop		r-1				oszlop
Kezelés (csop. között)		v-1				kezelés
Hiba		(r-1)(v-2)				
Összesen		rv				
Korrigált összesen		rv-1				

33. táblázat: *A GLM-táblázat szerkezete latin négyzet elrendezésben*

```
UNIANOVA
  termés  BY sor oszlop kezelés
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = sor oszlop kezelés .
```

34. táblázat: A latinnégyzet elrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: TERMÉS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4512.271[a]	12	376.023	22.332	.000
Intercept	108860.4	1	108860.4	6465.188	.000
SOR	2326.386	4	581.597	34.541	.000
OSZLOP	901.374	4	225.344	13.383	.000
KEZELÉS	1284.510	4	321.128	19.072	.000
Error	202.055	12	16.838		
Total	113574.7	25			
Corrected Total	4714.326	24			

a. R Squared = .957 (Adjusted R Squared = .914)

*35. táblázat: 5*x5 latinnégyzet eredménytáblázata

Latin tégla elrendezés

A latin négyzet elrendezés kiterjesztése 8, 9, 10, 12, 14, 15, 16, 18 kezelés összehasonlítására, ha a kezelések száma kétszer vagy háromszor annyi, mint az ismétlések száma.

Csoportosított elrendezés

Egytényezős kísérletekben, ha sok kezelést hasonlítunk össze, gyakran olyan kezeléscsoportokat képezünk, melyeken belül a kezelések összehasonlítását pontosabban szeretnénk elvégezni. Burgonya fajtakísérletben 11 burgonyafajta három érési csoportba sorolható 4, 4 illetve 3. Az öt ismétléses kísérlet kiértékelését az alábbiak szerint végezhetjük.

ismétlés			
(1)	III 10, 9, 11	I 1, 3, 4, 2	II 8, 6, 7, 5

(2)	II 5, 6, 8, 7	III 11, 9, 10	I 2, 1, 4, 3
(3)	I 4, 1, 3, 2	II 6, 7, 5, 8	III 10, 11, 9
(4)	III 9, 11, 10	I 1, 2, 4, 3	II 7, 5, 8, 6
(5)	II 6, 8, 7, 5	III 10, 9, 11	I 2, 3, 4, 1

31. ábra: Csoportosított elrendezés terve, 11 kezelés (v), 3 csoport (cs), 5 ismétlés (r)

Az elrendezés matematikai modellje:

$Y_{ijk} = m + R_i + C_j + e_{ij} + K_k + e_{ijk}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitása

C_j = az éréscsoportok termésre gyakorolt hatása

e_{ij} = az éréscsoportok közötti hiba

K_k = a fajták hatása a burgonya termésére

e_{ijk} = a kísérlet hibája

Tényező	SS	df	MS	F	Sig.	DESIGN
Eltérés		1				
Ismétlés		r-1				ismétlés
Csoportok között		cs-1				csoport
Hiba (cs)		(r-1)(cs-1)				ismétlés*csoport

Kezelés csp. belül	v-cs	kezelés
Hiba (v)	(r-1)(v-cs)	

36. táblázat: *A GLM-táblázat szerkezete csoportosított elrendezésben*

```
UNIANOVA
  termés  BY ismétlés csoport kezelés
  /METHOD = SSTYPE(1)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /RANDOM = ismétlés
  /DESIGN = ismétlés csoport csoport*ismétlés kezelés .
```

37. táblázat: : A csoportosított elrendezés SPSS parancsai

Az eltérés négyzetösszeget az első típus szerint kell számítani.

Tests of Between-Subjects Effects

Dependent Variable: Termés kg/parcella

Source		Type I Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	214531,364	1	214531,364	1096,328	,000
	Error	782,727	4	195,682(a)		
ismétlés	Hypothesis	782,727	4	195,682	3,708	,056
	Error	409,209	7,754	52,772(b)		
csoport	Hypothesis	4158,403	2	2079,202	39,761	,000
	Error	418,339	8	52,292(c)		
Hiba (cs)	Hypothesis	418,339	8	52,292	1,579	,170
	Error	1059,733	32	33,117(d)		
kezelés	Hypothesis	2520,433	8	315,054	9,513	,000
	Hiba (v)	1059,733	32	33,117(d)		

a MS(ismétlés)
b 1,025 MS(ismétlés * csoport) - ,025 MS(Error)
c MS(ismétlés * csoport)
d MS(Error)

38. táblázat: A csoportosított elrendezés eredménytáblázata, cs=3, v=11, r=5

A csoportok közötti szignifikancia vizsgálatkor, ha a ’Hiba (cs) MS’ kisebb, mint a ’Hiba (v) MS’, akkor a csoportok közötti tényezőt az F-próbában a Hiba (v)-hez viszonyítjuk.

Kéttényezős kísérletek

Véletlen blokkelrendezés

Egyik legegyszerűbb kéttényezős kísérleti elrendezés. Egyforma pontossággal tudjuk megítélni a kezelés kombinációk közötti különbségeket.

ismétlés	kezelések					
(1)	a_1b_1	a_1b_2	a_2b_1	a_2b_2	a_3b_1	a_3b_2
(2)	a_2b_1	a_1b_1	a_1b_2	a_3b_1	a_3b_2	a_2b_2
(3)	a_3b_1	a_2b_1	a_1b_1	a_3b_2	a_2b_2	a_1b_2
(4)	a_2b_2	a_3b_2	a_3b_1	a_1b_2	a_1b_1	a_2b_1
(5)	a_3b_2	a_3b_1	a_2b_2	a_2b_1	a_1b_2	a_1b_1

32. ábra: 5 ismétléses véletlen blokkelrendezésű 2x3-as faktoriális kísérlet elrendezése

Az elrendezés matematikai modellje:

$Y_{ijk} = m + R_i + A_j + B_k + AB_{jk} + e_{ijk}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitását mutatja

A_j = az „A” tényező termésre gyakorolt hatása

B_k = a „B” tényező termésre gyakorolt hatása

AB_{jk}= a két tényező kölcsönhatása

e_{ijk} = a kísérlet hibája

Tényező	SS	df	MS	F	Sig.	DESIGN
Korrigált modell		?				
Eltérés		1				
Ismétlés		r-1				ismétlés
A tényező		a-1				atényező
B tényező		b-1				btényező
AxB kölcsönhatás		(a-1)(b-1)				atényező*btényező
Hiba		(r-1)(ab-1)				
Összesen		rab				
Korrigált összesen		rab-1				

39. táblázat: *A GLM-táblázat szerkezete kéttényezős véletlen blokkelrendezésben*

```
UNIANOVA
  termés  BY ismétlés atényező btényező
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = ismétlés atényező btényező atényező*btényező
```

40. táblázat: A kéttényezős véletlen blokkelrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: Termés kg/parcella

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	62695[a]	9	6966.130	18.867	.000
Intercept	1393207	1	1393207	3773.414	.000
ISMÉTLÉS	2720	4	679.917	1.842	.160
A TÉNYEZÕ	57965	2	28982.500	78.497	.000
B TÉNYEZÕ	3	1	2.700	.007	.933
A x B	2008	2	1003.900	2.719	.090
Error (Hiba)	7384	20	369.217		
Total	1463287	30			
Corrected Total	70080	29			

a. R Squared = .895 (Adjusted R Squared = .847)

41. táblázat: A kéttényezős véletlen blokkelrendezés eredménytáblázata, a=3, b=2, r=5

Osztott parcellás (split-plot) elrendezés

Akkor alkalmazzuk, amikor az egyik vizsgált tényező (A) változatai közötti különbségek elbírálása nem elsődleges cél. Vagy kisparcellán lehetetlen beállítani a kezelést. A kísérlet főként a második tényező (B) változatainak értékelésére és az A x B kölcsönhatásra irányul.

(1) ismétlés

Fő parcella	A1	A2	A3
Osztó terület	B1	B2	B3
	B2	B1	B1
	B3	B3	B2

(2) ismétlés

Fő parcella	A2	A1	A3
Osztó terület	B3	B2	B1
	B1	B3	B2
	B2	B1	B3

33. ábra: Kéttényezős osztott parcellás kísérlet terve

Az elrendezés matematikai modellje:

$Y_{ijk} = m + R_i + A_j + e_{ij} + B_k + AB_{jk} + e_{ijk}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitását mutatja

A_j = az „A” tényező termésre gyakorolt hatása

e_{ij} = az „A” tényező hibája

B_k = a „B” tényező termésre gyakorolt hatása

AB_{jk}= a két tényező kölcsönhatása

e_{ijk} = a „B” tényező hibája

Tényező	SS	df	MS	F	Sig.	DESIGN
Eltérés		1				
Ismétlés		r-1				ismétlés
A tényező		a-1				atényező
Hiba (a)		(r-1)(a-1)				atényező*ismétlés
B tényező		b-1				btényező
AxB kölcsönhatás		(a-1)(b-1)				atényező*btényező
Hiba (b)		a(r-1)(b-1)				

42. táblázat: *A GLM-táblázat szerkezete kéttényezős osztott parcellás elrendezésben*

```
UNIANOVA
  napok  BY ismétlés atényező btényező
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /RANDOM = ismétlés
  /DESIGN = ismétlés atényező atényező*ismétlés btényező
    atényező*btényező .
```

43. táblázat: A kéttényezős osztott parcellás elrendezés SSPS parancsai

Tests of Between-Subjects Effects

Dependent Variable: Napok száma vetéstől hímvirágzásig

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept5	Hypothesis	675220,417	1	675220,417	58292,410	,000
	Error	46,333	4	11,583(a)		
ismétlés	Hypothesis	46,333	4	11,583	1,590	,240
	Error	87,400	12	7,283(b)		
A tényező	Hypothesis	1704,183	3	568,061	77,995	,000
	Error	87,400	12	7,283(b)		
Hiba (a)	Hypothesis	87,400	12	7,283	,495	,902
	Error	470,667	32	14,708(c)		
B tényező	Hypothesis	9168,233	2	4584,117	311,668	,000
	Error	470,667	32	14,708(c)		
A*B kölcsön-hatás	Hypothesis	19,767	6	3,294	,224	,966
	Hiba (b)	470,667	32	14,708(c)		

a MS(ismétlés)
b MS(ismétlés * atényező)
c MS(Error)

44. táblázat: A kéttényezős osztott parcellás elrendezés eredménytáblázata, a=4, b=3, r=5

Az „A” tényező közötti szignifikanciavizsgálatkor az „A” tényező MS-t akkor kell osztani a Hiba (a) MS-vel, ha ez az érték nagyobb, mint a Hiba (b) MS. Egyéb esetben a Hiba (b)-hez kell viszonyítani az „A” tényező hatását.

Sávos elrendezés

Sváb (1973) szerint a kéttényezős kísérletek kevésbé javasolható módszere. Mégis gyakran az egyetlen megoldás. Főként szántóföldi kísérletekben, ha a parcellaméret olyan kicsi, hogy azon technikai nehézségek miatt a tényezők egyike sem vizsgálható. Ilyenkor a pontosság rovására a belső ismétlések feláldozásával mindkét tényezőt főparcellákon helyezzük el. Előnye viszont, hogy a kölcsönhatást pontosabban lehet becsülni.

(1) ismétlés (2) ismétlés

	A1	A2	A3
B1			
B2			
B3			

	A2	A1	A3
B3			
B1			
B2			

34. ábra: Kéttényezős sávos elrendezés terve

Az elrendezés matematikai modellje:

$Y_{ijk} = m + R_i + A_j + e_{ij} + B_k + e_{ik} + AB_{jk} + e_{ijk}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitását mutatja

A_j = az „A” tényező termésre gyakorolt hatása

e_{ij} = az „A” tényező hibája

B_k = a „B” tényező termésre gyakorolt hatása

e_{ik} = a „B” tényező hibája

AB_{jk} = a két tényező kölcsönhatása

e_{ijk} = a „B” tényező hibája

Tényező	SS	df	MS	F	Sig.	DESIGN
Eltérés		1				
Ismétlés		r-1				ismétlés
A tényező		a-1				atényező
Hiba (a)		(r-1)(a-1)				atényező*ismétlés
B tényező		b-1				btényező
Hiba (b)		(r-1)(b-1)				btényező*ismétlés

AxB kölcsönhatás	(a-1)(b-1)	atényező*btényező
Hiba (a x b)	(r-1)(a-1)(b-1)	

45. táblázat: *A GLM-táblázat szerkezete kéttényezős sávos elrendezésben*

```
UNIANOVA
  termés  BY ismétlés atényező btényező
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /RANDOM = ismétlés
  /DESIGN = ismétlés atényező atényező*ismétlés btényező bté-
nyező*ismétlés atényező*btényező
```

46. táblázat: A kéttényezős sávos elrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: c.répa termés kg/10 m2

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	34133,333	1	34133,333	967,559	,000
	Error	105,833	3	35,278(a)		
Ismétlés	Hypothesis	105,833	3	35,278	2,988	,134
	Error	59,652	5,053	11,806(b)		
A tényező	Hypothesis	133,667	3	44,556	4,367	,037
	Error	91,833	9	10,204(c)		
Hiba (a)	Hypothesis	91,833	9	10,204	1,406	,257
	Error	130,667	18	7,259(d)		
B tényező	Hypothesis	1113,167	2	556,583	62,812	,000
	Error	53,167	6	8,861(e)		
Hiba (b)	Hypothesis	53,167	6	8,861	1,221	,341
	Error	130,667	18	7,259(d)		
AxB kölcsön-hatás	Hypothesis	86,333	6	14,389	1,982	,122
	Hiba (axb)	130,667	18	7,259(d)		

a MS(ismétlés)
b MS(ismétlés * atényező) + MS(ismétlés * btényező) - MS(Error)
c MS(ismétlés * atényező)
d MS(Error)
e MS(ismétlés * btényező)

47. táblázat: A kéttényezős sávos elrendezés eredménytáblázata, a=4, b=3, r=4

Három- és többtényezős kísérletek

Véletlen blokkelrendezés

Három tényező vizsgálatakor ez az elrendezés főként laboratóriumi vagy tenyészedény kísérletekben előnyös, mivel minden kombináció azonos pontossággal hasonlítható össze. Szántóföldi kísérletekben ritkán alkalmazzák, mivel a sok kombinációhoz nagy blokkokat kell képezni, és egyes kezelések beállítása technikailag szinte lehetetlen pl. talajművelés, öntözés, stb.

Az alábbi példa egy tőszám (2), hibrid (3) és műtrágyázási (3) kísérlet kiértékelését mutatja be.

Az elrendezés matematikai modellje:

$Y_{ijkl} = m + R_i + A_j + B_k + C_l + AB_{jk} + AC_{ji} + BC_{ki} + ABC_{jkl} + e_{ijkl}$

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitását mutatja

A_j = az „A” tényező termésre gyakorolt hatása

B_k = a „B” tényező termésre gyakorolt hatása

C_l = a „C” tényező termésre gyakorolt hatása

AB_{jk} = a két tényező kölcsönhatása

AC_{ji} = a két tényező kölcsönhatása

BC_{ki} = a két tényező kölcsönhatása

ABC_{jkl} = a három tényező kölcsönhatása

e_{ijkl} = hiba

Tényező	SS	df	MS	F	Sig.	DESIGN
Korrigált modell		?				
Eltérés		1				
Ismétlés		r-1				ismétlés
A tényező		a-1				toszam
B tényező		b-1				hibrid
C tényező		c-1				tragya
AxB kölcsönhatás		(a-1)(b-1)				hibrid*toszam
AxC kölcsönhatás		(a-1)(c-1)				toszam*tragya
BxC kölcsönhatás		(b-1)(c-1)				hibrid*tragya
AxBxC		(a-1)(b-1)(c-1)				hibrid*toszam*tragya
Hiba		(r-1)(abc-1)				
Összesen		rabc				
Korrigált összesen		rabc-1				

48. táblázat: *A GLM-táblázat szerkezete háromtényezős véletlen blokkelrendezésben*

```
UNIANOVA
  termes  BY ismetles toszam hibrid tragya
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = ismetles toszam hibrid tragya hibrid*toszam
toszam*tragya hibrid
 *tragya hibrid*toszam*tragya .
```

49. táblázat: A háromtényezős véletlen blokkelrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: termés t/ha

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	204.634[a]	20	10.232	14.208	.000
Intercept	6769.019	1	6769.019	9399.807	.000
ISMETLES	1.098	3	.366	.508	.678
TOSZAM	16.872	1	16.872	23.430	.000
HIBRID	23.914	2	11.957	16.604	.000
TRAGYA	147.237	2	73.618	102.230	.000
TOSZAM * HIBRID	3.873	2	1.936	2.689	.078
TOSZAM * TRAGYA	8.104	2	4.052	5.627	.006
HIBRID * TRAGYA	1.438	4	.360	.499	.736
TOSZAM * HIBRID * TRAGYA	2.098	4	.525	.728	.577
Error	36.726	51	.720		
Total	7010.379	72			
Corrected Total	241.360	71			

a. R Squared = .848 (Adjusted R Squared = .788)

50. táblázat: A háromtényezős véletlen blokkelrendezés eredménytáblázata, a=2, b=3, c=3, r=4

Kétszeresen osztott parcellás (split-split-plot) elrendezés

Ez az elrendezés technikailag igen előnyös háromtényezős elrendezés, főként szántóföldi kísérletekben, mert a főparcellák, az elsőrendű és a másodrendű alparcellák eltérő méretei különböző tulajdonságú kezelések kombinációját teszik lehetővé. A kevésbé fontos tényezőt (kezelést) a főparcellán (A) helyezzük el, a legfontosabbat a másodrendű alparcellákon (C). Az „A” tényező változatainak ismétlése megegyezik a valódi ismétlés számával (r). A „B” tényező ismétlés száma *ra*, amiből *a* a belső ismétlés. A „C” tényező ismétlése *rab*, amiből *ab* belső ismétlés. Amennyiben a kölcsönhatások nem szignifikánsan, a belső ismétlések is valódi ismétlést jelentenek.

Egy debreceni kísérletben a főparcellán a tőszámot (A), az elsőrendű alparcellán a hibridet (B) és a másodrendű alparcellán a műtrágyakezeléseket (C) helyezték el négy ismétlésben (r).

(1) ismétlés (2) ismétlés

Fő parcella	A1		A2		A2		A1	
Al parcella	B1	B2	B2	B1	B2	B1	B1	B2
	c1	c4	c3	c2	c2	c3	c4	c1
	c2	c2	c4	c1	c1	c4	c2	c2
	c3	c3	c1	c4	c4	c1	c3	c3
	c4	c1	c2	c3	c3	c2	c1	c4
	Osztó területek				Osztó területek			

35. ábra: Háromtényezős kétszeresen osztott parcellás elrendezés terve

Az elrendezés matematikai modellje:

Y_{ijkl} = m + R_i + A_j + e_{ij} + B_k + AB_{jk} + e_{ijk} + C_l + AC_{ji} + BC_{ki} + ABC_{jkl} + e_{ijkl}

ahol:

Y_{ij} = egy parcella termése (kg/parcella)

m = a kísérlet becsült, számított átlaga, a kísérlet legjellemzőbb értéke

R_i = blokk ill. ismétlés hatás a talaj heterogenitását mutatja

A_j = az „A” tényező termésre gyakorolt hatása

e_{ij} = az „A” tényező hibája

B_k = a „B” tényező termésre gyakorolt hatása

AB_{jk}= a két tényező kölcsönhatása

e_{ijk} = a „B” tényező hibája

C_l = a „C” tényező termésre gyakorolt hatása

AC_{ji} = a két tényező kölcsönhatása

BC_{ki} = a két tényező kölcsönhatása

ABC_{jkl} = a három tényező kölcsönhatása

e_{ijkl} = hiba

Tényező	SS	df	MS	F	Sig.	DESIGN
Eltérés		1				
Ismétlés		r-1				ismétlés
A tényező		a-1				toszam
Hiba (a)		(r-1)(a-1)				ismetlés*toszam
B tényező		b-1				hibrid
AxB kölcsönhatás		(a-1)(b-1)				hibrid*toszam
Hiba (b)		a(r-1)(b-1)				toszam(hibrid*ismetles)
C tényező		c-1				tragya
AxC kölcsönhatás		(a-1)(c-1)				toszam*tragya
BxC kölcsönhatás		(b-1)(c-1)				hibrid*tragya
AxBxC		(a-1)(b-1)(c-1)				hibrid*toszam*tragya
Hiba (c)		ab(r-1)(c-1)				

51. táblázat: *A GLM-táblázat szerkezete háromtényezős kétszeresen osztott parcellás elrendezésben*

```
UNIANOVA
  termes  BY ismetles toszam hibrid tragya
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /RANDOM = ismetles
  /DESIGN = ismetles toszam ismetles*toszam hibrid hibrid*toszam
toszam(hibrid*ismetles) tragya toszam*tragya hibrid*tragya
hibrid*toszam*tragya .
```

52. táblázat: Háromtényezős kétszeresen osztott parcellás elrendezés SPSS parancsai

Tests of Between-Subjects Effects

Dependent Variable: termes

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	6769,019	1	6769,019	18498,464	,000
	Error	1,098	3	,366(a)		
Ismetles	Hypothesis	1,098	3	,366	,336	,803
	Error	3,266	3	1,089(b)		
Toszam	Hypothesis	16,872	1	16,872	15,498	,029
	Error	3,266	3	1,089(b)		
Hiba (a)	Hypothesis	3,266	3	1,089	1,865	,189
	Error	7,005	12	,584(c)		
Hibrid	Hypothesis	23,914	2	11,957	20,482	,000
	Error	7,005	12	,584(c)		
Toszam * Hibrid	Hypothesis	3,873	2	1,936	3,317	,071
	Error	7,005	12	,584(c)		
Hiba (b)	Hypothesis	7,005	12	,584	,794	,653
	Error	26,455	36	,735(d)		
Tragya	Hypothesis	147,237	2	73,618	100,181	,000
	Error	26,455	36	,735(d)		
Toszam * Tragya	Hypothesis	8,104	2	4,052	5,514	,008
	Error	26,455	36	,735(d)		
Hibrid * Tragya	Hypothesis	1,438	4	,360	,489	,743
	Error	26,455	36	,735(d)		
Toszam * Hibrid * Tragya	Hypothesis	2,098	4	,525	,714	,588
	Hiba (c)	26,455	36	,735(d)		

a MS(ismetles)
b MS(ismetles * toszam)
c MS(toszam(ismetles * hibrid))
d MS(Error)

53. táblázat: Háromtényezős kétszeresen osztott parcellás elrendezés eredménytáblázata, a=2, b=3, c=3, r=4

ÖSSZEFÜGGÉS-VIZSGÁLATOK

Két változó között háromféle kapcsolat jöhet létre:

A *két változó független egymástól*, ha az egyik változó semmilyen információt nem szolgáltat a másik változóról.

Ha az egyik változó hat a másik változó alakulására, de a hatást „véletlenszerű" események zavarják (következtetés szintű és csak közelítőleg becsülhető), akkor *sztochasztikus* a kapcsolat a két változó között.

Függvényszerű kapcsolatról akkor beszélünk, ha az egyik változó egyértelműen befolyásolja a másik változót.

A természettudományi kutatásban egy eseményt nagyon sok tényező befolyásol, sokszor egyes tényezők nagyon kis mértékben hatnak a vizsgált eseményre, ezért ezeket figyelmen kívül szoktuk hagyni. Így ezek a „kis" tényezők szinte véletlenszerű ingadozást okoznak a függőváltozó alakulásában. Szerencsés esetben ezek az ingadozások egy normáleloszlású, nulla várhatóértékű sokasággal reprezentálhatók, így sztochasztikus törvényszerűséget mutathatnak. E sajátosság miatt elsősorban tehát sztochasztikus összefüggés-vizsgálatokat szoktak végezni.

Mind az alacsony, mind a magas mérési szintű változók között lehet kapcsolat, amit különböző mérőszámokkal jellemezhetünk.

Két *nominális* változó közötti kapcsolatot az *asszociációs* mérőszámokkal jellemezzük.

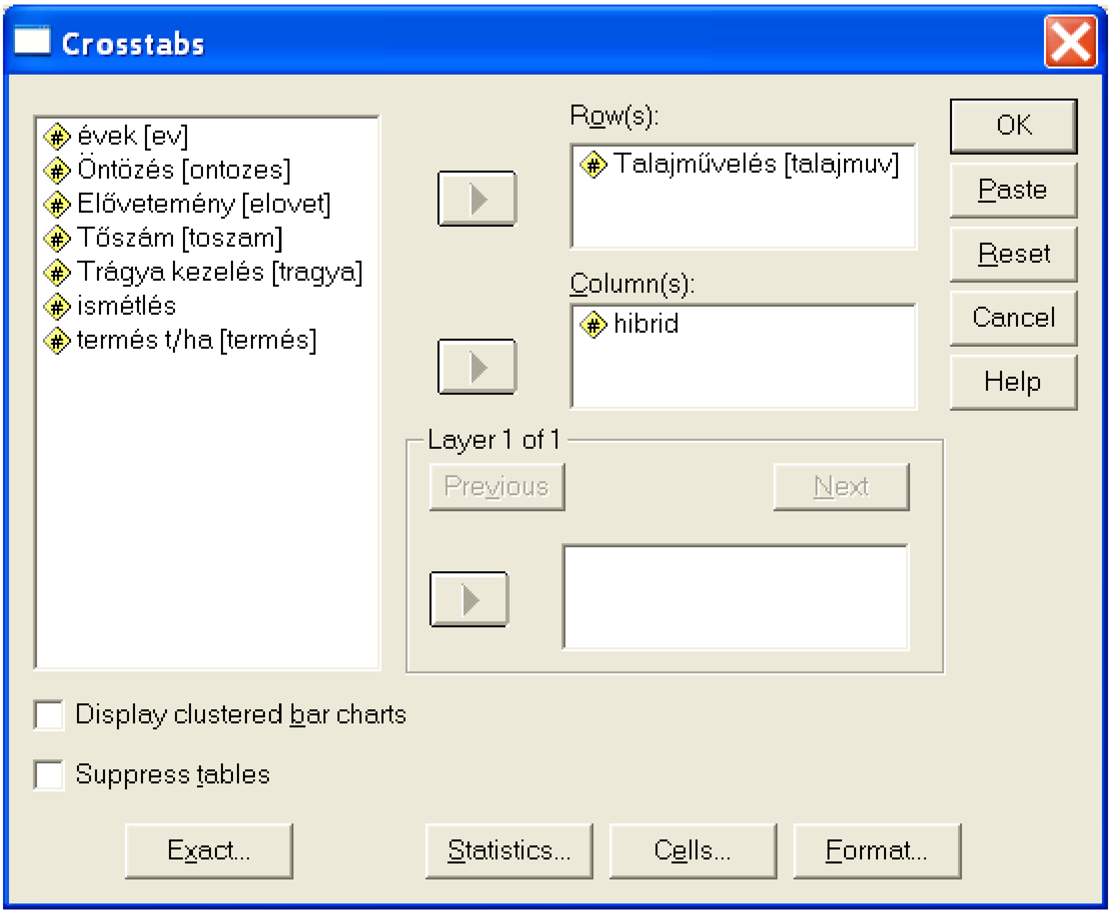
Ordinális típusú változók összefüggését a különböző *rangkorrelációs* mutatók mérik.

Skála típusú változók összefüggését *korreláció- és regresszióanalízissel* mutathatjuk ki.

Asszociáció

Amikor megkezdjük két nominális változó kapcsolatának vizsgálatát, és a kapcsolat meghatározását, elsőként az adatokat kereszttáblába (kombinációs táblába) rendezzük. A kereszttáblában az adatokat két (vagy több) szempont / két (vagy több) változó szerint rendezve látjuk. Ha a kontingencia táblázatban a gyakoriságok elhelyezkedése valamilyen szabályosságot mutat, akkor érdemes konkrét mutatószámmal kimutatni a kapcsolat szorosságát.

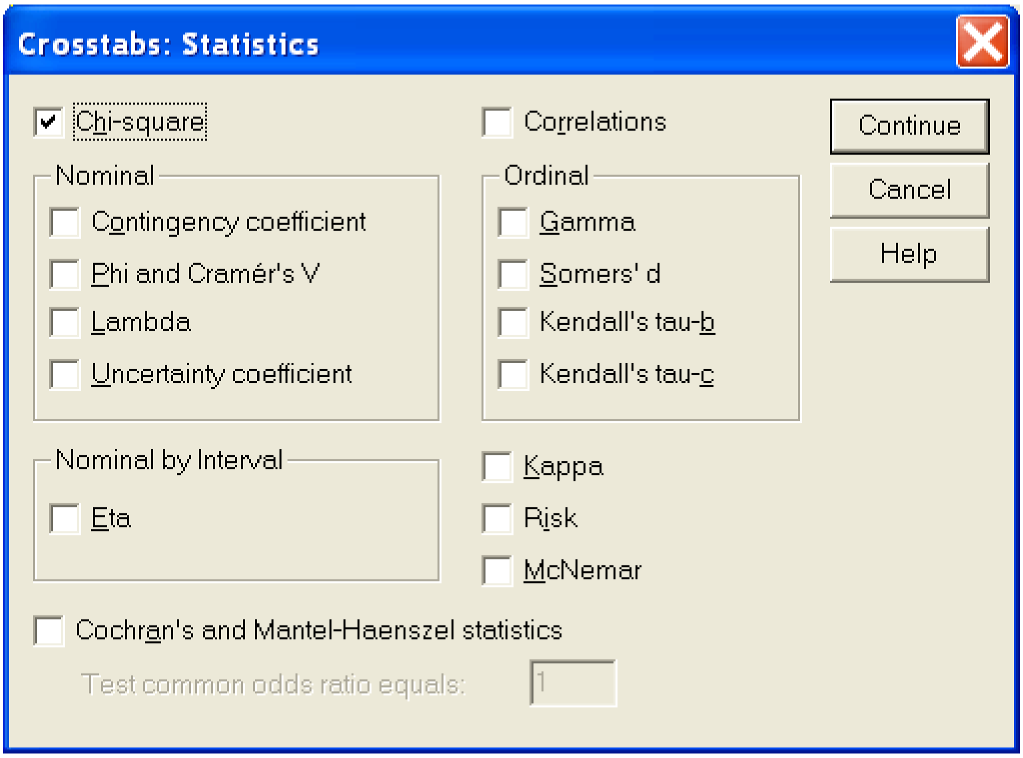
Az SPSS-ben a kereszttáblákat a Analyze, Descriptive Statistics, Crosstabs… parancsokon keresztül érhetjük el.



A χ^2 - próba

A próba két nominális változó közötti kapcsolat „valódiságának” az eldöntésére szolgál. A nullhipotézis itt is a változók függetlenségét jelenti. Ez a módszer önmagában nem mutatja meg a kapcsolat erősségét és irányát, csak arra ad választ, hogy a változók között van-e ténylegesen kapcsolat egy bizonyos valószínűség mellett.

Analyze, Descriptive Statistics, Crosstabs…, Statistics



A **Chi-square** bejelölésével, ha a táblában csak két sor és két oszlop szerepel (2*2 tábla), akkor kiszámításra kerül: Pearson-féle chí^2, likelihood-ratio chí^2, Fisher-teszt és Yates-féle korrigált chí^2. Tetszőleges sor és oszlop esetén kiszámításra kerül: Pearson-féle chí^2, likelihood-ratio chí^2. Amennyiben a tábla mindkét változója mennyiségi változó, kiszámításra kerül a linear-by-linear association test, aminek az eredményét csak kvantitatív változók esetében szabad figyelembe venni. A Fisher-féle teszt eredményét csak akkor kapjuk meg, ha a 2*2 táblában nincs hiányzó sor vagy oszlop, ill. nagyobb táblázatok esetén a várható cellagyakoriság kisebb, mint 5.

A **Correlations** négyzet bejelölésekor számítódik: ha a sor és oszlop változók ordinális típusúak, Spearman's correlation coefficient, rho, tehát mindkét változónak numerikus adatot kell tartalmaznia. Ez a mutató a rendezett mintában elfoglalt rangsor alapján becsüli a kapcsolat szorosságát és irányát. A kapcsolat milyenségének, pl. lineáris, másodfokú, stb. vizsgálata nem értelmezhető ordinális típusú adatoknál. Amennyiben mindkét változó skála típusú, akkor a Pearson-féle korrelációs koefficiens méri a lineáris kapcsolat szorosságát és irányát.

Összefüggés nominális változók között

A **Nominal** csoportba tartozó tesztek: akkor használjuk, ha mindkét változó nominális nem tudunk semmiféle valódi rangsort képezni, pl. vallási irányzatok és a nem kapcsolata. Ezek a tesztek mind a Khí^2 alapuló eljárások.

> **Kontingencia koefficiens:** 0-1 közötti értékkel méri a kapcsolat szorosságát. 0 függetlenséget, 1 szoros kapcsolatot jelent. A maximális értéke függ a sorok és oszlopok számától.
>
> **Phi and Cramer's V** (Cramer-féle asszociációs együttható): a Khí2 statisztika eredményét osztják a minta elemszámával és gyököt vonva kapják meg az eredményt. 0-1 közötti értékkel méri a kapcsolat szorosságát. 0 függetlenséget, 1 szoros kapcsolatot jelent.
>
> **Lambda**: kiszámításra kerül a szimmetrikus és nem szimmetrikus lambda értéke és a Goodman and Kruskal's tau statisztika. Ezek azt mérik, hogy milyen arányban tudjuk előre jelezni a független változóval a függő változót. Az 1 érték azt mutatja, hogy a független változóval tökéletes előre tudjuk jelezni a függő változót. 0 érték esetében a függő változó ismerete semmilyen segítséget nem ad a függő változó becslésére.
>
> **Uncertainty coefficient**: a bizonytalansági együttható megmutatja, hogy a változó előrejelzésének hibája mennyivel csökken, ha a másik változó értékeit használjuk becslésre. Pl. 0.83 azt jelenti, hogy az egyik változó ismerete 83%-kal csökkenti a másik változó előrejelzésének hibáját. A program kiszámítja a szimmetrikus és aszimmetrikus bizonytalansági együttható értékét.

Összefüggés ordinális változók között

Az **Ordinal** csoportba tartozó tesztek: akkor használjuk, ha mindkét változó ordinális típusú, tehát rangsorolhatók a vizsgált szempont alapján. Ilyen esetben első lépésként a változók megfigyelt értékeit rangsoroljuk és az egyes megfigyeléseknek a rangsoruknak megfelelő rangszámot adunk 1-től *n*-ig, ahol *n* a megfigyelési egységek száma. Azt vizsgáljuk, hogy a változók rangszámai az azonos megfigyelési egységeken mennyire egyeznek meg.

Gamma: két ordinális típusú változó közötti szimmetrikus kapcsolatot méri. Érétke -1 és 1 között van. Abszolút értékben egyhez közelítve erős kapcsolatot jelez a két változó között. Nullához közelítve gyenge kapcsolat vagy függetlenség áll fenn a két változó között. 2-utas táblák esetén zéró-rendű gamma, 3-utastól n-utas tábláig conditional gamma kerül kiszámításra.

Somers' d: két ordinális típusú változó közötti szimmetrikus kapcsolatot méri. Érétke -1 és 1 között van. Abszolút értékben egyhez közelítve erős kapcsolatot jelez a két változó között. Nullához közelítve gyenge kapcsolat vagy függetlenség áll fenn. Ez a teszt a gamma teszt aszimmetrikus kiterjesztése. Somer javasolta, hogy a vizsgálat során vegyék figyelembe azt is, hogy *B* függ *A*-tól, vagy fordítva. Kiszámításra kerül a két aszimmetrikus teszt és a szimmetrikus is.

Kendall's tau-b: Ordinális vagy rangsorolható változók közötti nem paraméteres mérce a változók korrelációjának mérésére. A koefficiens előjele a kapcsolat irányát, abszolút értéke a kapcsolat szorosságát mutatja. Értéke -1 és 1 között lehet. -1 vagy +1 értéket csak négyzetes táblázat esetén kaphatunk, ahol a sor és oszlop száma megegyezik. Ez a statisztika a számítások során figyelembe veszi a kötéseket. A kötés jelentése: előfordulhat amikor két vagy több megfigyelt eset között nem tudunk különbséget tenni, vagyis rangsorolásuk nem egyértelmű. Az ilyen esetek azonos rangszámot kapnak, így ezek a megfigyelések ún. kötésben állnak egymással. Egy rangsorban több kötés is előfordulhat.

Kendall's tau-c: Ordinális vagy rangsorolható változók közötti nem paraméteres mérce a változók korrelációjának mérésére. A koefficiens előjele a kapcsolat irányát, abszolút értéke a kapcsolat szorosságát mutatja. Értéke -1 és 1 között lehet. -1 vagy +1 értéket csak négyzetes táblázat esetén kaphatunk, ahol a sor és oszlop száma megegyezik. A mutató számításakor figyelmen kívül hagyják a kötéseket.

Összefüggés nominális és intervallum típusú változók között

Amennyiben az egyik változó kategóriákat tartalmaz és a másik változó mennyiségi változó, akkor az Eta mutatja a kapcsolat szorosságát. A kategorikus változót numerikusan kell kódolni az analízis előtt.

Eta: A sor és oszlop változók közötti kapcsolat szorosságát méri. Értéke 0 és 1 között lehet. 0 esetén nincs kapcsolat a két változó között, egyhez közeli értéke erős kapcsolatot jelez. A statisztika számításakor feltétel, hogy a függő változó skála típusú legyen, pl. fizetés, a független változó pedig korlátozott számú kategóriákat tartalmazzon, pl. nem. Két Eta értéket kapunk a teszt elvégzése után. Az egyik akkor érvényes, ha a sor tartalmazza az intervallum típusú változót, a másik akkor, ha az oszlop.

Egyéb összefüggéseket mérő tesztek

Kappa: Cohen-féle kappa

Kappa. Cohen's kappa measures the agreement between the evaluations of two raters when both are rating the same object. A value of 1 indicates perfect agreement. A value of 0 indicates that agreement is no better than chance. Kappa is only available for tables in which both variables use the same category values and both variables have the same number of categories.

Risk: 2 x 2 táblában méri az összefüggés szorosságát

Risk. For 2 x 2 tables, a measure of the strength of the association between the presence of a factor and the occurrence of an event. If the confidence interval for the statistic includes a value of 1, you cannot assume that the factor is associated with the event. The odds ratio can be used as an estimate or relative risk when the occurrence of the factor is rare.

McNemar: nem paraméteres teszt két kapcsolatban álló dichotóm változó esetén. Tests for changes in responses using the chi-square distribution. Useful for detecting changes in responses due to experimental intervention in "before-and-after" designs. Nagyméretű négyzetes táblák esetén a McNemar-Bowker test szimmetrikus változata kerül kiszámításra.

Cochran's and Mantel-Haenszel statisztika: Cochran's and Mantel-Haenszel statistics can be used to test for independence between a dichotomous factor variable and a dichotomous response variable, conditional upon covariate patterns defined by one or more layer (control) variables. Note that while other statistics are computed layer by layer, the Cochran's and Mantel-Haenszel statistics are computed once for all layers.

KORRELÁCIÓ-ANALÍZIS

A korreláció két (vagy több) véletlen változó közötti kapcsolat jellemzésére szolgál. Feltételezzük, hogy mindkét valószínűségi változó (*x* és *y*) normális eloszlású, és a közöttük lévő lineáris összefüggés mértékét a korrelációs együttható mutatja, melyet r-rel jelölünk. Értéke -1 és +1 közé eshet, a határokat is beleértve. Ha *r* pozitív, akkor *y* együtt növekszik, vagy csökken *x*-szel. Negatív *r* esetében ellentétes irányú a változás. Amennyiben az *r* értéke $|1|$, *x* és *y* között függvényszerű kapcsolat van, amelynél minden pont egy egyenesen helyezkedik el. A két változót, ill. ismérvet korrelálatlannak nevezzük, ha *r=0*.

Kétváltozós korreláció-analízis

Pearson-féle korrelációs együttható

Szorzatmomentum korreláció, a korrelációs együttható egyenlete:

$$r_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

ahol:

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Kendall-féle rangkorreláció

τ (tau) < r_s

τ értéke 1 ha $\mu_{ij} > \mu_{kj}$ és -1 ha $\mu_{ij} < \mu_{kj}$

Spearman-féle rangkorreláció

A XX. század eleje óta ismert, ezt alkalmazzák leggyakrabban. A szorzatmomentum korrelációs együtthatóból közvetlenül kiszámítható. Értéke {-1, +1}. Próba statisztikája: Student-eloszlású, *n-2* szabadságfokkal, t-próbát végezhetünk H_0 elfogadására vagy elvetésére. Jele: r_s.

Először növekvő sorrendbe rendezzük mind az x_i mind az y_i értékeket, majd mindegyik helyébe egy *1* és *n* közé eső rangszámot írunk. Azonos értékek esetén az átlagos rangszámot írjuk mindegyikhez. Mindkét számsorban legfeljebb a megfigyelések egyötöde lehet azonos rangszámú. Képezzük az x_i, y_i értékpárok rangjainak különbségét, amit jelöljünk D_i-vel.

$$r_s = 1 - \left\{ 6 \sum \frac{D_i^2}{n\left(n^2 - 1\right)} \right\}$$

Abban az esetben használhatjuk, ha:

- egyik vagy mindkét változó ordinális változó (pl. az alma íze és színezettsége közötti összefüggés)
- a két változó közötti összefüggés nem lineáris, de az összefüggést ábrázoló görbe nem hajlik vissza

Kanonikus korreláció (CANOCOR)

A változókat két csoportra osztjuk, és azt vizsgáljuk, hogy:

1. A változók egyik csoportja milyen szorosan függ össze a változók másik csoportjával?
2. Az összefüggésrendszerben az egyes változóknak milyen nagy a jelentőségük, súlyuk?

A kanonikus korreláció bővített többszörös regresszió-analízis. *P* számú *x* változót *q* számú *y* változóval hozzuk összefüggésbe. Az *x* és *y* változócsoportok közötti összefüggést több egyenlettel fejezzük ki, amelyek függetlenek egymástól. A számítás a két változócsoport közös sajátérték számításán keresztül történik. Annyi kanonikus korrelációs koefficiens van, amennyi a λ_j sajátértékek, azaz az egyenletek száma. Minden egyenletre szignifikancia számítható Chí^2-próbával.

REGRESSZIÓ-ANALÍZIS

Lineáris regresszió (Linear…)

Lineáris regresszió-analízis akkor használható, ha a függő változó skálatípusú, és lineáris kapcsolatban van egy vagy több magyarázó változóval.

Alkalmazhatósági feltételek

- a hiba normáleloszlású, nulla várható értékű
- a hibatagok varianciája állandó és független a modell változóitól. Amennyiben a hibatagok varianciája nem állandó, akkor a modell heteroszkedasztikus, és nem alkalmas az összefüggés leírására
- a hibatagok nagysága független a változóktól, nincs autokorreláció közöttük

A lineáris regressziós egyenes meghatározás a legkisebb négyzetek elve alapján:

$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \min.$ a legkisebb négyzetek elve, Gauss német matematikus, 1777-1855.

$$\hat{y} = a + bx$$

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \to \min.$$

$$\sum_{i=1}^{n}\left(y_i - (a + bx)\right)^2 \to \min.$$

A legjobban illeszkedő függvény, a közelítés hibája:

$$\delta_i = y_i - y_i'$$

A hibák négyzetének átlagos értéke:

$$\bar{S}_y^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

az „*a*” és „*b*” paraméter meghatározása:

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum\left[y-(a+bx)\right]^2=\sum\left[y-a-bx\right]^2$$

A zárójel felbontása és négyzetre emelés után:

$$f=\sum_{i=1}^{n}(y^2-ay-bxy-ay+a^2+abx-bxy+abx+b^2x^2)$$
$$f=\sum(y^2-2ay-2bxy+a^2+2abx+b^2x^2)$$

az eltérések négyzetösszege az „*a*” és „*b*” függvénye. A függvény szélsőértékének létezéséhez szükséges feltétel, hogy

$$\frac{\partial f}{\partial a}=0\text{, és }\frac{\partial f}{\partial b}=0$$

$$\frac{\partial f}{\partial a}=\sum\left(-2y+2bx+2a\right)=0$$
$$f'=2\sum\left(-y+bx+a\right)=0 \qquad /2$$
$$-\sum y+b\sum x+\sum a=0$$
$$1. \qquad \sum y=na+b\sum x$$

$$\frac{\partial f}{\partial b}=\sum\left(-2xy+2ax+2bx^2\right)=0$$
$$f'=2\sum\left(-xy+ax+bx^2\right)=0 \qquad /2$$
$$-\sum xy+a\sum x+b\sum x^2=0$$
$$2. \qquad \sum xy=a\sum x+b\sum x^2$$

a deriválás és átalakítás után kaptuk a Gauss-féle normál egyenletrendszert.

1. $\sum y=an+b\sum x$
2. $\sum xy=a\sum x+b\sum x^2$

Az 1. egyenletet n-nel elosztva és „*a*”-t kifejezve kapjuk az alábbi egyenletet:

$$3. \qquad a=\frac{\sum y}{n}-b\frac{\sum x}{n}$$

„*a*”-t behelyettesítve a 2. egyenletbe:

$$\sum xy = \left(\frac{\sum y}{n} - b\frac{\sum x}{n} \right) \sum x + b\sum x^2$$

$$\sum xy = \frac{\sum y \sum x}{n} - b\frac{\left(\sum x\right)^2}{n} + b\sum x^2$$

$$\sum xy = \frac{\sum y \sum x}{n} + b\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n} \right)$$

$$\boxed{b = \frac{\sum xy - \frac{\sum y \sum x}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{SP_{xy}}{SS_x}}$$

A 3. egyenletet ha jól megfigyeljük, nem más, mint:

$$\boxed{a = \bar{Y} - b\bar{X}}$$

Munkaképletek:

$$\boxed{SS_x = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}} \quad \boxed{SS_y = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}} \quad \boxed{SP_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}$$

Regresszió-analízis variancia táblázata

Tényező	DF	SS	MS	F
Összes (Total)	n-1	$SS_T = SS_Y$	$\frac{SS_T}{n-1}$	
Regresszió (Regression)	1	$SS_R = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = \frac{SP_{xy}^2}{SS_x}$	SS_R	$\frac{MS_R}{MS_E}$
Hiba (Error)	n-2	$SS_E = SS_T - SS_R$	$\frac{SS_E}{n-2}$	

Az F-próba megmutatja, hogy becslésre a modell jobb-e, mintha csak az átlagot használnánk.

Statisztikák (Statistics):

Regressziós koefficiens: a regressziós koefficiens becslése, konfidencia intervallumának meghatározása, kovariancia mátrix számítása.

Modell illesztése (Model fit): többszörös R, R-négyzet, korrigált R-négyzet meghatározása, illesztés hibája (standard error), ANOVA táblázat.

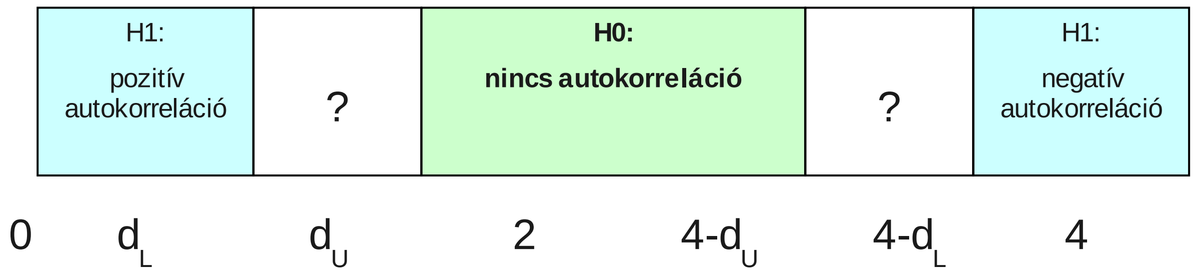
R-négyzet változása: többváltozós lineáris regresszió esetén adja meg a helyes r-négyzet értékét.

Leíró statisztika (Descriptives): átlag, szórás, korrelációs-mátrix és ennek egyoldalú próbája.

Parciális korrelációk meghatározása:

Kollinearitás meghatározása (Collinearity diagnostics): leteszteli, hogy egy vagy több független változó előállítható-e más független változók lineáris kombinációjaként. Többszörös lineáris regresszió-analízis esetén használható.

Maradékok (Residulals): a lineáris reláció feltételeit elemezhetjük vele. A maradékoknak (ε hibatagok) függetleneknek kell lenniük az x (független változó) értékeitől. Abban az esetben, ha az összefüggésben az x változó az idő, jól használható a Durbin-Watson féle d statisztika. A d statisztika értékei 0 és 4 közé esnek. Ha d megközelítően 2, akkor az eltérések nem függnek az időtől. Ha d megközelítően 0, az eltérések elsőrendű pozitív összefüggésben vannak az idővel, azaz az idő múlásával fokozatosan nőnek az eltérések. Ha d értéke megközelítően 4, negatív összefüggés áll fenn, azaz az idő múlásával csökkennek az eltérések.



36. ábra: A Durbin-Watson teszt döntési szabálya

Alternatív hipotézis	$H_0 : \rho = 0$		
	Elfogadjuk	Elvetjük	Nincs döntés
	$d > d_U$	$d < d_L$	$d_L \le d \le d_U$
	$d < 4 - d_U$	$d > 4 - d_L$	$4 - d_U \le d \le 4 - d_L$

54. táblázat: *A Durbin-Watson próba döntési táblája*

Esetek diagnosztikája (Casewise diagnostics): a kiugró eseteket tudjuk meghatározni. Azokat az eseteket, amelyeknél a standard maradék abszolút értékben 3-nál nagyobb, az SPSS kiugró értéknek (outlier) nyilvánítja. Ezt az értéket szabadon meg tudjuk változtatni.

Ábrák (Plots):

A lineáris összefüggés feltételeinek elemzését grafikus úton is elvégezhetjük. Ábrázoljuk a standardizált maradékokat a standardizált becsült értékek függvényében. Ha a lineáris feltételezés jó volt, az ábrán az y=0 egyenes körül szabálytalanul szóródó pontokat kell látni. Hibás feltevés esetén, az ábrán valamilyen szabályosság fedezhető fel. Ha a maradék x függvénye, az ábrán aszimmetrikus sűrűsödési területek láthatók.

Standardizált maradékok ábrája (Standardized Residual Plots): ábrázolhatjuk a maradékokat hisztogram formájában, valamint megadhatjuk az ábrán a normál-valószínűségű pontsor mintázatát (Normal probabilty plot).

Mentések (Save):

Becsült értékek (Predicted Values): becsült értékek, standardizált becsült értékek, korrigált becsült értékek, a becsült értékek standard hibája.

Maradékok (Residulas): maradékok, standardizált maradékok, studentizált maradékok, törölt maradékok (deleted residulas), studentizált törölt maradékok.

Távolságok (Distances): Mahalanobis távolság az x változó standardizált értékének négyzete. Ez a távolságérték nagyon jól használható az átlagostól jelentősen eltérő szélsőséges értékek kimutatására. Cook-távolság megmutatja, hogy egy pont elhagyása milyen mértékben változtatná meg a regressziós koefficienst. Ahol ez a távolság nagy, nagymértékben változtatja meg a regressziós koefficienst.

Megbízhatósági szint

Akkor jelöljük be a négyzetet, ha az összesítő eredménytáblában egy további szintet szeretnénk megjeleníteni. A mezőbe írjuk be azt a konfidenciaszintet, amelyet az alapértelmezésű 95%-os szint felett alkalmazni szeretnénk.

Zéró legyen a konstans?

Jelöljük be a négyzetet, ha azt akarjuk, hogy a regressziós egyenes átmenjen a koordinátatengelyek metszéspontján (az origón).

Maradékok

Ha bejelöljük, akkor a maradékok egy külön eredménytáblában jelennek meg.

Standard maradékok

Ha bejelöljük, akkor a normalizált maradékok egy külön eredménytáblában jelennek meg.

Alkalmazhatósági feltételek megvizsgálása:

- a hiba normáleloszlású, nulla várható értékű: ábrázoljuk a standardizált maradékokat (ZRESID) hisztogram formájában, valamint adjuk meg az ábrán a normál-valószínűségű pontsor mintázatát (Normal probabilty plot). Másik lehetőség a P-P ábra készítése.

- a hibatagok varianciája állandó és független a modell változóitól, nem heteroszkedasztikus a modell: ábrázoljuk a studentizált törölt maradékokat (SDRESID) a standard becsült értékek (ZPRED) függvényében!
- a hibatagok nagysága független a változóktól, nincs autokorreláció közöttük: Ábrázoljuk a standardizált maradékokat az x változó függvényében is! Végezzük el a Durbin-Watson tesztet!

Többszörös lineáris regresszió

Alkalmazhatósági feltételek

- ugyanaz, mint a kétváltozós lineáris regresszió-analízisnél, és még
- a független változók (x_i) függetlenek egymástól, nincs közöttük multikollinearitás

A kétváltozós lineáris regresszió-analízist ki kell egészíteni a független változók további vizsgálataival.

Az SPSS beállítása: Statistics: Estimates, Model fit, Part and Partial correlation, Collinearity diagnostics

A beállítás eredményeként a regressziós együtthatók eredménytáblázata az alábbi statisztikai jellemzőkkel egészül ki.

Coefficients[a]

t	Sig.	Correlations: Zero-order	Partial	Part	Collinearity Statistics: Tolerance	VIF
-,004	,996					
3,104	,002	,094	,216	,094	,990	1,010
3,520	,001	,177	,243	,106	,511	1,956
-,956	,340	,097	-,068	-,029	,516	1,938
29,275	,000	,891	,902	,884	,997	1,003

55. táblázat: *A kollinearitás vizsgálata*

Háromféle korrelációs koefficienst (Correlations) kapunk. A zero-order az egyszerű kétváltozós korrelációs együttható a függő (y) és független (x) változó között. A *Partial* és *Part* a parciális korrelációs együtthatók, amikor az előbbi korrelációs koefficiens számításakor a független változó hatásából eltüntetjük a többi független változó lineáris hatását (kontroláló változók figyelembevétele). A *Partial* számításakor a függő és független változónál egyaránt figyelembe vesszük a többi független változó lineáris hatását, azaz ezeket a „zavaró” hatásokat kiszűrjük a modellből. A *Part* számításakor csak a vizsgált független változóból szűrjük ki a többi független változó hatását. Ez megmutatja, hogy mennyivel változna az R^2 értéke, ha a lineáris egyenletbe bevonnánk a változót. Néha ezt szemiparciális korrelációnak is nevezik.

A parciális korrelációs koefficiensek számítását három változó esetében az alábbi módon kell elvégezni. (Több változó esetén a számítás hosszadalmas, ezért érdemes számítógéppel végezni.) A számításhoz ismerni kell a páronkénti korrelációs koefficiensek értékét. A parciális korrelációs koefficiensek képlete:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

$$r_{xz.y} = \frac{r_{xz} - r_{xy} r_{yz}}{\sqrt{(1-r_{xy}^2)(1-r_{yz}^2)}}$$

$$r_{yz.x} = \frac{r_{yz} - r_{xy} r_{xz}}{\sqrt{(1-r_{xy}^2)(1-r_{xz}^2)}}$$

A parciális koefficiensek indexében az első két betű mutatja azt a két változót, aminek az összefüggését az együttható kifejezi, a pont mögött a kikapcsolt változó jele látható.

Az utolsó két oszlop a független változók kollinearitásának mérőszámait tartalmazza.

A Tolerancia jelentése: a vizsgált független változó varianciájának azon százaléka, amit nem lehet a többi független változóval megmagyarázni. Számítása: $1\text{-}R^2$. R^2 a vizsgált független változó és a többi független változó közötti többszörös lineáris regresszió-analízis determinációs együtthatója. Alacsony tolerancia érték azt jelenti, hogy a vizsgált független változó varianciájának zömét a többi független változóval lehet megmagyarázni. Kritikus érték a 0,3-nál kisebb. Amennyiben a Tolerancia értéke közelíti a nullát, magas multikollinearitással kell számolni. Ebben az esetben a regressziós koefficiens hibája megnő.

VIF (variance inflation factor) a Tolerance reciproka. Amennyiben értéke meghaladja a kettőt probléma léphet fel a változók függetlenségével. A tolerancia érték ilyenkor kisebb, mint 0,5.

A kollinearitás létének kimutatását további analízisek segítik. Az alábbi táblázat faktoranalízis segítségével analizálja a regresszió-analízisben résztvevő független változók kapcsolatrendszerét.

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	nitráttartalom (mg/kg)	globálsugárzás (MJ/m2/év)	éves átlaghőmérséklet (C)	csapadék (mm/év)
1	1	4,868	1,000	,00	,00	,00	,00	,00
	2	,106	6,785	,00	,96	,00	,00	,02
	3	,023	14,397	,01	,03	,00	,01	,94
	4	,002	45,098	,47	,00	,00	,52	,03
	5	,001	77,738	,52	,00	,99	,46	,00

a. Dependent Variable: tarhonyatermés (t/ha)

56. táblázat: *A kollinearitás további vizsgálata faktoranalízissel*

Sajátérték (Eigenvalues) a független változók belső összefüggésrendszeréről ad információt. Megmutatja, hány független dimenzió létezik a változók között. Amennyiben számos sajátérték közelíti a nullát, akkor az azt jelent, hogy a magyarázó változók erősen összefüggnek, a belső korrelációjuk nagy (56. táblázat). Ebben az esetben az alapadatok kis változása, ingadozása nagy hatást gyakorol a regressziós együtthatók becslésére. Pontatlan lesz a becslés.

Condition Index: $\sqrt{\frac{\lambda_{\max}}{\lambda_i}}$, 15-nél nagyobb értéke esetén már probléma lehet, 30-nál nagyobb értéke esetén súlyos probléma lép fel, a magyarázó változók nem függetlenek.

Multikollinearitás esetében érdemes kipróbálni a Stepwise regresszió-analízist standardizált változókkal.

Probit-analízis, probit regresszió

A kifejezés az angol „probability unit” szavak összevonásából keletkezett. A dózis-válasz összefüggés tisztázására szolgáló statisztikai eljárás. Ez az eljárás azt vizsgálja, hogy a dózis milyen mértékben, százalékban vált ki hatást. A független változó (x) tehát valamilyen dózis. A dózist azonban tágan kell értelmezni. Ez egy olyan hatás, ami a megfigyelt esemény bekövetkezésére hat. Sokszor a dózis természetes, vagy tízes alapú logaritmusát veszik. Ekkor feltételezik, hogy az egyedek érzékenysége lognormál eloszlást követ. A függő változó (y) a dózisra reagáló (választ adó) gyakoriság. A függő változót probit transzformációval átalakítják, ami nem más, mint a kumulált standard normáleloszlás inverze. A z-érték a -4-től +4 értéket vehet fel leggyakrabban. Azért, hogy elkerüljék a negatív számokat a z-értékekhez hozzáadnak ötöt. Így a probit 5 érték az 50%-os reakciót jelenti, amikor az egyedek felénél reakciót vált ki a dózis. Növényvédelmi kísérletekben, gyomirtó, rovarirtó szereknél az LD50 (letális dózis 50%) vizsgálata.

Feltételezzük, hogy a transzformált függő változó és a logaritmizált független változó lineáris kapcsolatban van. Ezekre a változókra lineáris regresszió-analízissel illesztünk egyenest, illetve keressük meg az egyenes paramétereit. A lineáris egyenlet:

Probit(p) = a + b*log(x)

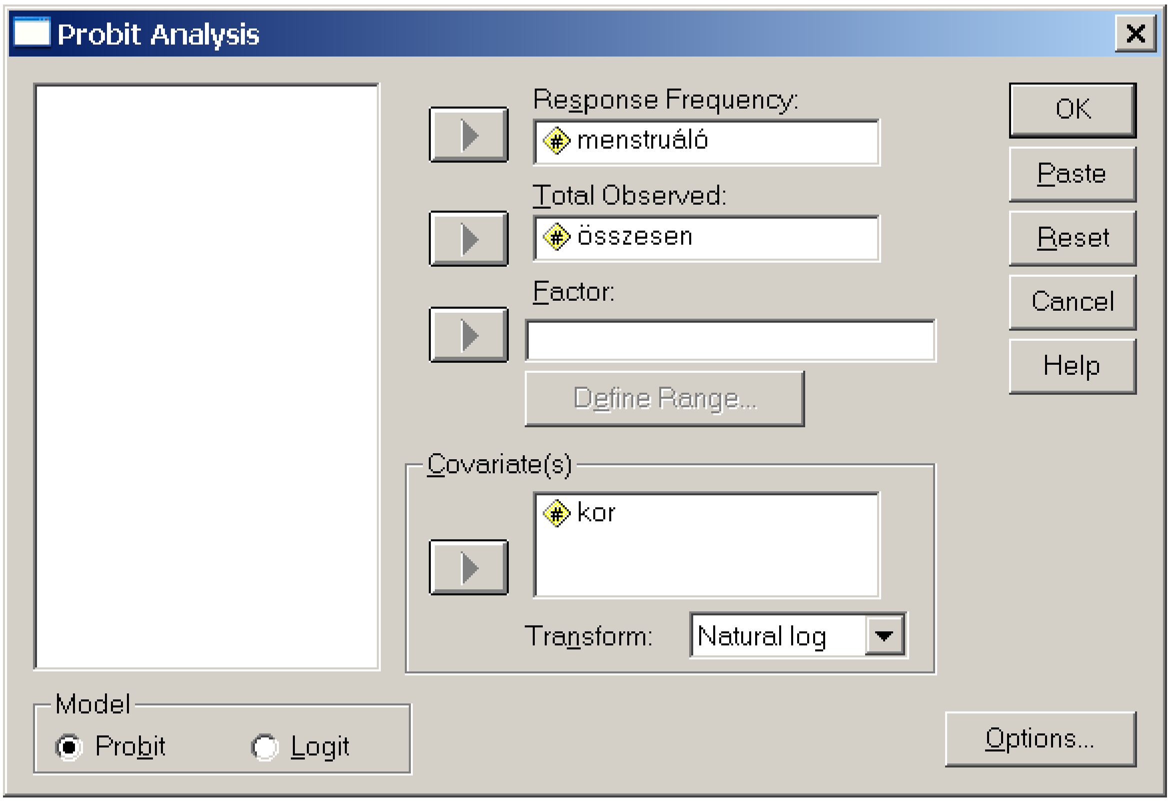
A regresszió-analízis során a maximum likelihood módszert használják, ami iterációkon keresztül közelíti meg a függvény paramétereit.

Az egyenes paramétereinek ismeretében meghatározhatjuk a becsült reakciók nagyságát, és az ehhez tartozó 95%-os konfidencia-intervallumot. Ebbe a tartományba fog esni a dózis okozta hatás 95%-os valószínűséggel.

A számítógépes megoldás menete függ a programtól. Az SPSS-ben az adatbázist úgy kell felépíteni, hogy az egyik változó tartalmazza a független változót (x, kovariáns, dózis), a másik a dózisra reagáló gyakoriságot, és a harmadik az összes megfigyelések számát. A kovariánst úgy is előállíthatjuk, hogy egy folytonos változót kategóriákba, osztályközökbe soroljuk.

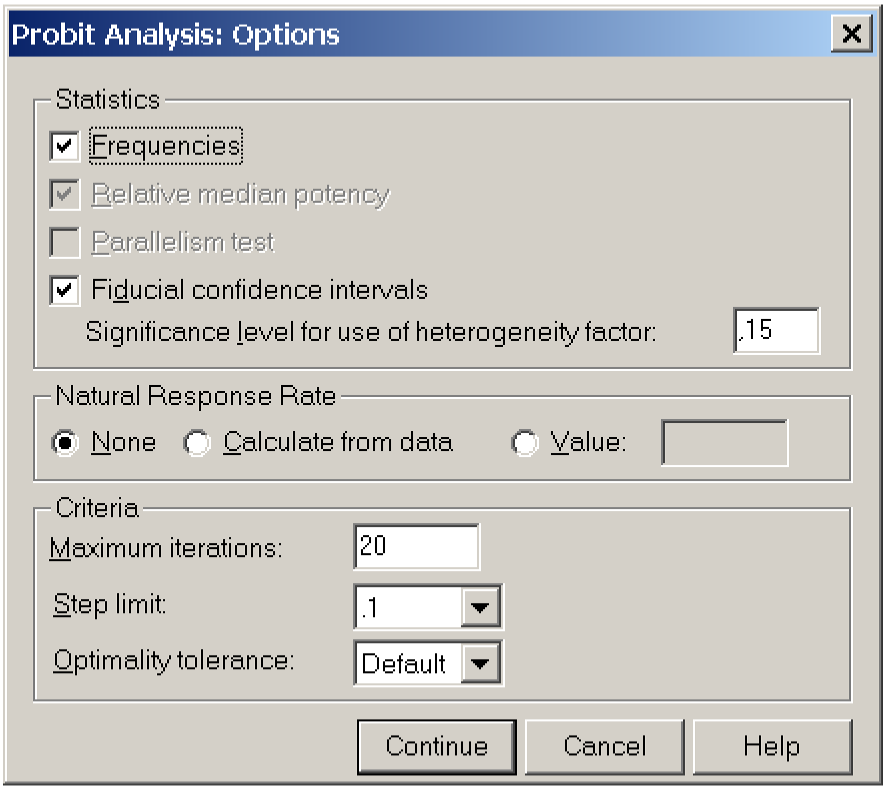
Az alábbi példa a magyarországi lányok nemi érését vizsgálta. A változók az alábbiak voltak: kovariáns az életkorcsoport, féléves bontásban. A második változó az erre adott reakció gyakorisága, a menstruáció száma az adott életkorcsoportban. A harmadik pedig a vizsgálatba vont leányzók számát tartalmazta.

A párbeszédpanel beállítása lent látható.



Az SPSS-ben is ki lehet választani, hogy a független változót transzformáljuk-e. Az analízist le lehet futtatni: nem transzformált, tízes alapú és természetes alapú logaritmus transzformációval is.

Az egyéb opciók az alábbi párbeszéd ablakból érhetők el.



A statisztikák részben kiválasztva a frekvenciák lehetőséget, megkapjuk minden egyes dózis kategóriában a tapasztalati és elméleti gyakoriságot, valamint a kettő különbségét. A konfidencia intervallum a dózisokhoz tartozó várható értéknek a 95%-os megbízhatósági tartományát adja meg. A heterogenitás vizsgálatakor beállíthatjuk az első fajú hiba nagyságát, ami ebben a vizsgálatban 10-15% szokott lenni.

A dózisokra adott válasz mértékét is meghatározhatjuk, ha rendelkezünk kontrollal, nulla dózissal. Ezt vagy a tapasztalati adatokból vagy korábbi szakirodalmi adatok alapján becsülhetjük.

A fenti példa megoldása:

```
* * * * * * * * *  P R O B I T   A N A L Y S I S  * * * * * * * * * * * *

 Parameter estimates converged after 7 iterations.
 Optimal solution found.

 Parameter Estimates (PROBIT model:  (PROBIT(p)) = Intercept + BX):
```

```
          Regression Coeff.  Standard Error     Coeff./S.E.

  kor               ,94813          ,05404        17,54368


                  Intercept  Standard Error  Intercept/S.E.

                  -11,99553          ,69002       -17,38431
```

A regressziós egyenes képlete: Y=0,94813 – 11,995.

```
 Pearson  Goodness-of-Fit  Chi Square =     12,634    DF = 7   P =  ,082

 Since Goodness-of-Fit Chi square is significant, a heterogeneity
 factor is used in the calculation of confidence limits.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- -
```

A fenti szöveg alapján a minta heterogén, ezért a konfidencia intervallum meghatározásakor ezt figyelembe kell venni.

```
* * * * * * * * * *  P R O B I T    A N A L Y S I S  * * * * * * * * * * *

 Observed and Expected Frequencies
```

kor	Number of Subjects	Observed Responses	Expected Responses	Residual	Prob
11,00	55,0	3,0	3,226	-,226	,05866
11,50	99,0	7,0	13,603	-6,603	,13741
12,00	192,0	67,0	51,512	15,488	,26829
12,50	146,0	57,0	64,646	-7,646	,44278
13,00	201,0	124,0	126,501	-2,501	,62936
13,50	134,0	108,0	105,775	2,225	,78936

14,00	143,0	129,0	128,618	,382	,89942
14,50	76,0	72,0	72,971	-,971	,96014
15,00	17,0	17,0	16,779	,221	,98701

A korcsoport kategóriák függvényében látjuk a megfigyelések számát, a tapasztalati gyakoriságot, a becsült gyakoriságot, a kettő közötti különbséget és a kumulált valószínűséget.

A függvény által becsült konfidencia intervallumok láthatók lent. A sárgával jelzett mutatja az 50%-os reakcióhoz tartozó “dózis” nagyságát.

```
* * * * * * * * ** * *  P R O B I T   A N A L Y S I S  * * * * * * * * * *
* *
```

Confidence Limits for Effective kor

Prob	kor	95% Confidence Limits Lower	Upper
,01	10,19817	9,61241	10,61150
,02	10,48568	9,96068	10,85775
,03	10,66810	10,18128	11,01435
,04	10,80533	10,34698	11,13240
,05	10,91695	10,48159	11,22860
,06	11,01196	10,59602	11,31064
,07	11,09526	10,69621	11,38270
,08	11,16985	10,78581	11,44734
,09	11,23768	10,86719	11,50623
,10	11,30013	10,94200	11,56053
,15	11,55866	11,25047	11,78664
,20	11,76413	11,49369	11,96830
,25	11,94040	11,70035	12,12615
,30	12,09870	11,88374	12,27009
,35	12,24539	12,05120	12,40595
,40	12,38458	12,20726	12,53772
,45	12,51925	12,35498	12,66847
,50	12,65179	12,49670	12,80081
,55	12,78433	12,63444	12,93712

,60	12,91900	12,77028	13,07976
,65	13,05819	12,90658	13,23128
,70	13,20488	13,04632	13,39486
,75	13,36318	13,19349	13,57503
,80	13,53945	13,35399	13,77903
,85	13,74492	13,53782	14,02007
,90	14,00345	13,76576	14,32672
,91	14,06590	13,82040	14,40120
,92	14,13373	13,87962	14,48225
,93	14,20832	13,94457	14,57153
,94	14,29162	14,01695	14,67141
,95	14,38663	14,09931	14,78551
,96	14,49825	14,19583	14,91980
,97	14,63548	14,31422	15,08517
,98	14,81790	14,47119	15,30540
,99	15,10541	14,71787	15,65324

Meg kell jegyezni, hogy napjainkban inkább a logit-analízist, logisztikus regressziót használják a probit helyett, mert az sokoldalúbb.A probit és a logit-függvény között formailag alig van eltérés, ezért nehéz eldönteni, hogy az adott probléma megoldására melyik módszer az alkalmasabb.

Nemlineáris regresszió

Analyze, Regression, Nonlinear…menü

Dependent: a függő változó megadása, válasszunk a változók listájáról

Model expression: itt kell megadni a függő változó becslésére alkalmas függvényt, aminek legalább egy független változót kell tartalmaznia. Help, See Also, Nonlinear Regression Common Models választva segítséget kaphatunk a leggyakrabban használt nem lineáris függvények szintaktikájáról:

Név	**Az egyenlet szintaktikája**
Asymptotic Regression v. Baule-Mitscherlich telítődési	b1 + b2 *exp(b3 * x)

függvény	
Asymptotic Regression	b1 -(b2 *(b3 ** x))
Density	(b1 + b2 * x)**(-1/ b3)
Gauss	b1 *(1- b3 *exp(-b2 * x **2))
Gompertz	b1 *exp(-b2 * exp(-b3 * x))
Johnson-Schumacher	b1 *exp(-b2 / (x + b3))
Log-Modified	(b1 + b3 * x) ** b2
Log-Logistic	b1 –ln(1+ b2 *exp(-b3 * x))
Metcherlich Law of Diminishing Returns	b1 + b2 *exp(-b3 * x)
Michaelis Menten	b1* x /(x + b2)
Morgan-Mercer-Florin	(b1 * b2 + b3 * x ** b4)/(b2 + x ** b4)
Peal-Reed	b1 /(1+ b2 *exp(-(b3 * x + b4 * x **2+ b5 * x **3)))
Ratio of Cubics	(b1 + b2 * x + b3 * x **2+ b4 * x **3)/(b5 * x **3)
Ratio of Quadratics	(b1 + b2 * x + b3 * x **2)/(b4 * x **2)
Richards	b1 /((1+ b3 *exp(- b2 * x))**(1/ b4))
Verhulst	b1 /(1 + b3 * exp(- b2 * x))
Von Bertalanffy	(b1 ** (1 - b4) - b2 * exp(-b3 * x)) ** (1/(1 - b4))
Weibull	b1 – b2 *exp(- b3 * x ** b4)
Yield Density	(b1 + b2 * x + b3 * x **2)**(-1)

exp = természetes logaritmus alapja (közelítőleg 2,71828182845904).

57. táblázat

Parameters: a modellben szereplő minden paramétereknek kezdeti értéket kell adni. Lehetőség van arra is, hogy a korábban elvégzett analízis során kapott paraméterek értékeit használjuk kezdeti, kiinduló értékeknek.

Constraints… Parameter Constraints: lineáris kifejezéssekkel egy vagy több paraméter értékének korlátozó feltételeket adhatunk meg.

Loss…Loss Function: a regressziós egyenlet meghatározásának módját határozhatjuk meg. Alapesetben a maradékok eltérés négyzetösszegének minimalizálásával folyik a regressziós egyenlet meghatározása. Lehetőség van általunk készített és definiált módszerrel is meghatározni a regressziós függvényt (User-defined loss function). Ha pl. a keresett függvény képe f = $p_1 + p_2$*x, akkor a legkisebb négyzetes eltérést így adhatjuk meg xloss = (y – ($p_1 + p_2$*x))**2, az illesztés során ezen értékek összegének a minimalizálása folyik. Ha valamelyik paraméterrel osztani szeretnénk, nyugodtan megtehetjük, pl. tőszám kísérletben Nedler 1963 és Mead szerint az illesztés pontosságára vonatkozó kritérium: RSS/Θ^2, vagyis a maradék (rezidium) négyzetösszegét osztani kell a Θ^2-tel. Az xloss-ba csak paraméterek is tartalmazó függvényt érdemes megadni, hisz a számítások során a p értékek változnak, és csak ezek tudnak konvergálni a megadott feltételek szerint. Az SPSS-ben a változók között megtalálhatók a becsült és maradék értékek, PRED_ és RESID_ jelöléssel.

Paste paranccsal az Spsss for Windows Syntax Editor-ba léphetünk be, ahol a párbeszédpaneleken megadott beállítások az Spss programnyelvén láthatók. Később a sokszor ismétlődő futtatások itt egyszerűbben és hatékonyabban állíthatók be, ill. futtathatók.

Save… paranccsal elmenthetjük a becsült, maradék, derivált és ha volt az xloss függvény értékeit. Ezek az értékek új változóként az adatbázisban megjelennek, és további elemzést végezhetünk rajtuk.

Lehetőségünk van logikai feltételeket is beépíteni a regressziós függvényünkbe. Például a modellünkben a vizsgált független változó egyenlő 0 ha X<=0, X ha 0<X<1, és 1 ha X>=1. A logikai kifejezés az alábbi lehet:

(X<=0)*0 + (X>0 & X < 1)*X + (X>=1)*1.

Az összetett logikai kifejezés mindenegyes tagja 1 (igaz) vagy 0 (hamis) értéket vehet fel. Megoldások:

Ha X<=0, a fenti kifejezés értéke 1*0 + 0*X + 0*1=0.

ha 0<X<1, a kifejezés értéke 0*0 + 1*X +0*1 = X.

ha X>=1, a kifejezés értéke 0*0 + 0*X + 1*1 = 1.

Ennél sokkal bonyolultabb logikai feltételeket is megadhatunk, sőt szöveges változókkal is dolgozhatunk, ilyenkor a szöveges változó nevét aposztróf közé kell tenni. pl. (város='Debrecen') *rezsi + (város='Derecske')*0.59*rezsi. Ennek a mezőnek az értéke egy kifejezés (a rezsi változó értéke) ha Debrecen és a rezsi változó értékének 59%-a, ha Derecske a város változó tartalma.

Kovariánsok alkalmazása a lineáris modellben

A kovariánsok alkalmazásakor még három további feltétellel kell kiegészíteni a varincia-analízis alkalmazhatóságának feltételrendszerét.

1. Minden csoportban a függő változó (amit vizsgálok) és a kovariáns (ami a függő változó varianciáját alapvetően meghatározza) között lineáris kapcsolat van.
2. A regressziós vonalak meredeksége minden csoportban egyforma, nem változik.
3. Az alkalmazott kezelések a kovariánst nem befolyásolják. Ez utóbbi feltétel nagyon fontos, vizsgálata sokszor nagy szakmai tudást és mérlegelést igényel.

OSZTÁLYOZÁSOK

Diszkriminancia-analízis két csoporttal DA

Két csoport A és B szétválasztására alkalmas módszer több változó együttes figyelembe vétele alapján. A többszörös regresszió-analízishez hasonló, y azonban kvalitatív változó, ezt tekintjük két csoportnak.

1. Több kvalitatív tényező együttes figyelembe vételével kimutatható-e szignifikáns különbség a két csoport között?
2. Az egyedek eredeti besorolása helyes-e?
3. Keresünk egy függvényt, amellyel eldönthető, hogy egy új megfigyelt egyed az A vagy B csoportba sorolható?
4. Több tulajdonság együttes figyelembe vételével egyetlen számszerű értékkel jellemezhetjük az egyedet „z”-érték. Az egyedeket egyetlen dimenzióban, egy közös grafikonon ábrázolhatjuk.
5. Z_a és Z_b középértéket számíthatunk, számszerűsíteni tudjuk a Z_a-Z_b különbséget.
6. Mennyire függ a két csoport Z_a-Z_b különbsége az egyes tulajdonságoktól? D^2, általánosított távolság felbontása.

A fenti eljárást több csoport egyidejű vizsgálatára is ki lehet terjeszteni, MDA.

ADATREDUKCIÓK

Főkomponens-analízis

Sajátérték számításon alapuló valódi többváltozós eljárás. Az x változó $s_i^2 = 1$ varianciáját bonjuk fel. Az eredetileg megfigyelt változókat korrelációjuk alapján kevesebb számú főkomponens változóvá vonjuk össze. Gyakran már 2-3 főkomponens változóval kielégítő pontossággal helyettesíthetjük a „p" számú megfigyelt változót. Minden megfigyelési egység annyi főkomponens értéket kap, ahány főkomponens-változót kiszámítunk.

A főkomponens-analízis (principal component analysis) a többváltozós módszerek közül a legfontosabb. Gyakorlati alkalmazásuk a bonyolult és számításigényes sajátérték számítás miatt csak számítógépen valósítható meg. A módszer előnyei:

1. A változók számának csökkentése, a jelentéktelen változók kiszűrése.
2. A vizsgált változók csoportosítása az egymás közötti korrelációjuk alapján. Megállapíthatjuk, hogy hány ilyen csoport van, és csoporton belül a változók kapcsolata milyen, pozitív vagy negatív.
3. Közös háttérváltozó ill. faktor felismerése, mely valamely változócsoporttal szoros összefüggésben van. (pl. levegő, talajhőmérséklet közötti kapcsolat, melynek közös háttérváltozója a napenergia)
4. A változók térbeli elhelyezkedését, csoportosulását lehet ábrázolni. A főkomponensek lesznek a koordinátarendszer tengelyei.
5. A főkomponensváltozók kiszámításával osztályozni tudjuk a megfigyelési egységeket több tulajdonság, ill. változó együttes figyelembevételével. Minden megfigyelés annyi főkomponensértéket kap, ahány főkomponensváltozót kiszámítunk. A főkomponens változók fogják képezni a két-, esetleg háromdimenziós ábrák tengelyeit.

6. A főkomponensváltozók és egy adott függőváltozó között kétváltozós vagy többszörös regresszióanalízist végezhetünk, ezt nevezik főkomponensregressziónak.

58. táblázat. Alapadatok

Fajta	farinográf érték	sikér terülés	sikér mennyiség	fehérje %
Mironovszkaja 808.	81.8	3.0	34.3	14.8
Fertődi 293.	75.9	6.4	39.3	16.1
Bezosztája	79.9	2.6	32.6	14.2
Martonvásári 1.	68.6	3.7	31.7	14.5
Martonvásári 2.	77.4	3.2	33.0	14.5
Martonvásári 16.	68.7	6.0	37.1	14.8
Martonvásári 24.	73.6	3.2	31.7	13.4
Jubilejnaja	73.3	2.1	31.4	14.5
Avróra	66.8	5.1	34.1	14.5
GK-Fertődi 2.	58.3	6.5	33.4	15.0
Kavkáz	61.2	5.1	33.3	14.5
Rannaja	59.6	2.9	30.4	15.1
Kiszombori	52.6	7.9	35.8	14.6
Burgas	44.2	10.8	36.1	14.0
Összesen:	941.9	68.5	474.2	204.5

SPSS Analyze, Descriptive Statistics, Descriptives…

Options… Mean, Std. Deviation

Save standardized values as variables

59. táblázat. Átlagok és szórások

Descriptive Statistics

	N	Mean	Std. Deviation
Farinograf érték	14	67.279	10.9255
Sikér terülés	14	4.893	2.4474
Sikér mennyisége	14	33.871	2.4703
Fehérje %	14	14.607	.6044
Valid N (listwise)	14		

Standardizálás után az alábbi értékeket kapjuk:

*60. táblázat. Standardizált adatok, **Z** mátrix*

Fajta	farinográf érték	sikér terülés	sikér mennyiség	fehérje %
Mironovszkaja 808.	1.33	-.77	.17	.32
Fertődi 293.	.79	.62	2.20	2.47
Bezosztája	1.16	-.94	-.51	-.67
Martonvásári 1.	.12	-.49	-.88	-.18
Martonvásári 2.	.93	-.69	-.35	-.18
Martonvásári 16.	.13	.45	1.31	.32
Martonvásári 24.	.58	-.69	-.88	-2.00
Jubilejnaja	.55	-1.14	-1.00	-.18
Avróra	-.04	.08	.09	-.18
GK-Fertődi 2.	-.82	.66	-.19	.65
Kavkáz	-.56	.08	-.23	-.18
Rannaja	-.70	-.81	-1.41	.82
Kiszombori	-1.34	1.23	.78	-.01
Burgas	-2.11	2.41	.90	-1.00
Összesen:	0	0	0	0

A standardizált értékek tulajdonságai: összegük, ill. az átlaguk egyenlő nullával, a szórásuk egy. A standardizálással egy nulla várhatóértékű, egy szórású sokaságot állítottunk elő.

SPSS Analyze, Data Reduction, Factor…
Descriptives, Correlation Matrix

Korrelációs mátrix meghatározása

*61. táblázat. Korrelációs mátrix, **R** mátrix*

Correlation Matrix

		Farinograf érték	Sikér terülés	Sikér mennyisége	Fehérje %
Correlation	Farinograf érték	1.000	-.774	-.126	.103
	Sikér terülés	-.774	1.000	.681	.087
	Sikér mennyisége	-.126	.681	1.000	.480
	Fehérje %	.103	.087	.480	1.000

Az U sajátvektor mátrix és a sajátértékek (λ_j) meghatározása

*62. táblázat. Sajátvektor mátrix és sajátértékek, **U** mátrix és **λ***

Változó	u_1	u_2	u_3	u_4
Farinograf érték	-.4787	.5312	.5045	.4838
Sikérterülés	.6560	-.2008	.1514	.7116
Sikér mennyiség	.5361	.4144	.5454	-.4933
Fehérje %	.2303	.7111	-.6520	.1270
sajátértékek (λj)	2.1524	1.3316	0.4989	0.0170

A sajátvektorok sor és oszlop irányban normáltak, azaz a négyzetösszegük egy sor-, ill. oszlopvektoron belül 1.

A sajátvektorok további tulajdonsága, hogy sorpáronkénti és oszloppáronkénti szorzatösszegük nulla, azaz a sorok és oszlopok páronként ortogonálisak (függetlenek egymástól). Az **U** mátrix ortonormált.

Ha a sajátértékeket összeadjuk, megkapjuk a változók számát, a mátrix rangját.

Főkomponens koefficiensek

A főkomponens koefficienseket (Component Score Coefficient) a sajátvektor mátrixból állítjuk elő súlyozással, tehát a sajátvektorokat osztjuk a hozzátartozó sajátértékek gyökével.

$$wu_{ij} = \frac{u_{ij}}{\sqrt{\lambda_j}}$$

*63. táblázat. Súlyozott főkomponens-koefficiensek **WU***

Component Score Coefficient Matrix

	Component			
	1	2	3	4
Farinograf érték	-,326	,460	,714	3,706
Sikér terülés	,447	-,174	,214	5,451
Sikér mennyisége	,365	,359	,772	-3,778
Fehérje %	,157	,616	-,923	,973

Extraction Method: Principal Component Analysis.
Component Scores.

Főkomponens változók

Főkomponens-változók kiszámítása: **Z** mátrix * Súlyozott főkomponens-koefficiensek.

*64. táblázat. Főkomponens-változók **C** mátrix*

Fajta	C_1	C_2	C_3	C_4
Mironovszkaja 808.	-,66600	1,00537	,62309	,36481
Fertődi 293.	1,20873	2,56748	,11257	,38011

Bezosztája	-1,08968	,09483	,84878	,46418
Martonvásári 1.	-,60646	-,28447	-,53313	,94043
Martonvásári 2.	-,76830	,31087	,40475	,82342
Martonvásári 16.	,68755	,64722	,90447	-1,67999
Martonvásári 24.	-1,13286	-1,15980	1,42983	-,24724
Jubilejnaja	-1,08353	-,01630	-,45970	-,56999
Avróra	,05812	-,11090	,22191	-,22301
GK-Fertődi 2.	,59406	-,16055	-1,19359	1,88714
Kavkáz	,10701	-,46317	-,39425	-,89894
Rannaja	-,52031	-,18406	-2,51418	-,94032
Kiszombori	1,27122	-,55915	-,08266	-1,24292
Burgas	1,94044	-1,68739	,63211	,94232
Összesen:	0	0	0	0

A főkomponens-változók középértéke nulla, szórásnégyzetük egyenlő eggyel. Tehát tulajdonságban hasonlítanak a Z-mátrixhoz, azonban van egy nagyon jelentős eltérés. A főkomponens-változók egymástól függetlenek, azaz az egymás közötti korrelációjuk nulla. (A kovariancia-mátrixa is ugyanígy néz ki.) A standardizált változók és a főkomponens-változók szórásnégyzeteinek összege, valamint a sajátértékek összege azonos.

65. táblázat. Főkomponens-változók korrelációs mátrixa

	C_1	C_2	C_3	C_4
C_1	1	0	0	0
C_2	0	1	0	0
C_3	0	0	1	0
C_4	0	0	0	1

A főkomponens változók ábrázolása

A vízszintes tengely a C_1, a függőleges a C_2 változó. A különböző őszi búzafajták főkomponens-változó értékeit az alábbi ábra mutatja.

37. ábra. A főkomponens-változók ábrázolása

Három főkomponens-változó kétdimenziós ábrázolásához válasszuk a Scatterplot Matrix menüpontot, és adjuk meg az első három főkomponens-változót.

38. ábra. Három főkomponens-változó ábrázolása

Az átló elemei a főkomponens-változók. Az első oszlopban az első változó az x-tengely, a másodikban a második, és így tovább. Az y-tengelyt a sorok mutatják.

A főkomponens súlyok meghatározása

A sajátvektorok elemeit megszorozzuk a hozzátartozó sajátérték négyzetgyökével, vagyis a szórással.

$$a_{ij} = u_{ij}\sqrt{\lambda_j}$$

66. táblázat. Főkomponenssúly mátrix, ***A****-mátrix*

Component Matrix[a]

	Component			
	1	2	3	4
Farinograf érték	-,702	,613	,356	6,317E-02
Sikér terülés	,962	-,232	,107	9,291E-02
Sikér mennyisége	,787	,478	,385	-6,44E-02
Fehérje %	,338	,821	-,461	1,658E-02

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

A főkomponens-súly mátrix tulajdonságai:

Számszerű értéke csak -1 és +1 között lehet. Az oszloponkénti négyzetösszeg egyenlő a hozzátartozó sajátértékkel.

A soronkénti négyzetösszeg egyenlő eggyel.

Tehát oszlop irányban a főkomponensek, sor irányban a megfigyelt változók varianciáját bontottuk fel.

A súlyok négyzeteinek főösszege egyenlő a mátrix rangjával, az egész rendszer összvarianciájával.

Kommunalitás, h^2. Ha sor irányban balról jobbra haladva összegezzük a főkomponens-súly négyzeteit, megkapjuk a kumulált értéküket, és ezeket nevezzük kommunalitásnak.

67. táblázat. Kommunalitások

Communalities

	Initial	Extraction
Farinograf érték	1,000	1,000
Sikér terülés	1,000	1,000
Sikér mennyisége	1,000	1,000
Fehérje %	1,000	1,000

Extraction Method: Principal Component Analysis.

Bármely két oszlop szorzata nulla. A főkomponenssúly vektorok ortogonálisak (függetlenek).

Bármely két sor szorzata a két változó kétváltozós korrelációs koefficiensét adja. Ha megszorozzuk az **A**-mátrixot a transzponáltjával, visszakapjuk az **R**-mátrixot, azaz az eredeti változók korrelációs koefficienseit.

Factor Analysis, Descriptives…, Correlation Matrix, Reproduced

68. táblázat. Korrelációs mátrix reprodukálása a főkomponessúlyokból, maradékok

Reproduced Correlations

		Farinograf érték	Sikér terülés	Sikér mennyisége	Fehérje %
Reproduced Correlation	Farinograf érték	1,000[b]	-,774	-,126	,103
	Sikér terülés	-,774	1,000[b]	,681	8,740E-02
	Sikér mennyisége	-,126	,681	1,000[b]	,480
	Fehérje %	,103	8,740E-02	,480	1,000[b]
Residual[a]	Farinograf érték		,000	1,665E-16	,000
	Sikér terülés	,000		-4,441E-16	-2,78E-17
	Sikér mennyisége	1,665E-16	-4,441E-16		1,110E-16
	Fehérje %	,000	-2,776E-17	1,110E-16	

Extraction Method: Principal Component Analysis.

a. Residuals are computed between observed and reproduced correlations. There are 0 (,0%) nonredundant residuals with absolute values greater than 0.05.

b. Reproduced communalities

A főkomponens-analízissel a varianciákat átrendeztük. A standardizált változóknál minden változó azonos jelentőséggel szerepel a variancia szempontjából. A főkomponens-analízisben az eredeti változók összefüggése miatt az első főkomponens varianciája magába foglalja az összes változó varianciájának legnagyobb közös részét, második főkomponens a maradék varianciák legnagyobb közös részét és így tovább, míg az utolsó főkomponensekre alig marad varianciarész. Ezért ezeket jelentéktelennek tekinthetjük, és elhanyagolhatjuk. Az átrendezett varianciákban figyelembe vettük az X változó összes varianciáját és egymás közötti korrelációját. A főkomponensek egymással már nem korrelálnak.

A λ sajátértékeket főkomponensenként kumulálva mutatja a 69. táblázat. Leolvasható, hogy a különböző főkomponensek hány százalékát értelmezik az összes varianciának.

69. táblázat. Az összes variancia felbontása

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,152	53,810	53,810	2,152	53,810	53,810
2	1,332	33,290	87,100	1,332	33,290	87,100
3	,499	12,473	99,574	,499	12,473	99,574
4	1,704E-02	,426	100,000	1,704E-02	,426	100,000

Extraction Method: Principal Component Analysis.

Főkomponensek ábrázolása

A főkomponensek ábrázolása a főkomponenssúlyok alapján történik, ezért az **A** mátrixot főkomponensmintázatnak (pattern) is nevezik. Két legfeljebb háromdimenziós ábrát készíthetünk.

Factor Analysis, Rotation, Display, Loading plot(s)

39. ábra. A változók háromdimenziós konfigurációja

A főkomponenssúlyok gyakorlati értelmezése

A főkomponenssúlyok a megfigyelt változók és a főkomponens-változók közötti korrelációs koefficiensek, melyet a 70. táblázat mutat.

70. táblázat. A korrelációs koefficiensek, ill. főkomponenssúlyok

Correlations

		Farinograf érték	REGR factor score 1 for analysis 1	REGR factor score 2 for analysis 1	REGR factor score 3 for analysis 1	REGR factor score 4 for analysis 1
Farinograf érték	Pearson Correlation	1	-,702**	,613*	,356	,063
	Sig. (2-tailed)	,	,005	,020	,211	,830
	N	14	14	14	14	14
REGR factor score 1 for analysis 1	Pearson Correlation	-,702**	1	,000	,000	,000
	Sig. (2-tailed)	,005	,	1,000	1,000	1,000
	N	14	14	14	14	14
REGR factor score 2 for analysis 1	Pearson Correlation	,613*	,000	1	,000	,000
	Sig. (2-tailed)	,020	1,000	,	1,000	1,000
	N	14	14	14	14	14
REGR factor score 3 for analysis 1	Pearson Correlation	,356	,000	,000	1	,000
	Sig. (2-tailed)	,211	1,000	1,000	,	1,000
	N	14	14	14	14	14
REGR factor score 4 for analysis 1	Pearson Correlation	,063	,000	,000	,000	1
	Sig. (2-tailed)	,830	1,000	1,000	1,000	,
	N	14	14	14	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

A főkomponensekkel háttérváltozókat (okváltozókat) akarunk felderíteni. A főkomponenssúlyok azt fejezik ki, hogy milyen jelentősége és súlya van valamely főkomponensnek (háttérváltozónak) a megfigyelt változók varianciájában.

A megfigyelt változók közötti korrelációs koefficiensek felbontása. Az **A**-mátrix bármely két sorának skaláris szorzata megadja a két változó közötti korrelációs koefficienst. Két változó skaláris szorzata akkor lehet nagy, ha a két változó nagy főkomponenssúlyai ugyanazokban a főkomponensekben vannak és a különböző főkomponensekben a szorzatuk azonos előjelű. A korrelációs koefficienst így egymástól független tényezőkre bontottuk fel.

A főkomponenssúlyok csoportosulása. Ha kettőnél több nagy főkomponens ugyanabba a főkomponensben van, akkor a változók egymással páronként, ezáltal közösen, csoportosan korrelálnak. Közös háttérváltozót kereshetünk.

A változók ábrázolásakor legjobban a kör, ill. gömb kerületén elhelyezkedő változók korrelálnak a legszorosabban. Az egymással negatívan korreláló változókat az origóra középpontosan tükrözni lehet, hogy könnyebben felismerjük az összefüggést.

Mit jelent a nagy vagy kis főkomponenssúly? Sváb szerint, ha a változók között nincs korreláció, úgy p=2 esetén $\frac{1}{\sqrt{2}}$, kb. 0,7 körüliek az a_{ij} értékek, p=10 változó esetén $\frac{1}{\sqrt{10}}$, kb. 0,3 körüli értékeket kapunk véletlenszerű elosztásban, minthogy a négyzetek összege mindenképpen csak 1 lehet.

Hány főkomponens jelentős? A λ legalább egy, vagyis eléri az átlagot. Ezt alkalmazzák a statisztikai programcsomagokban is. Más ajánlás szerint az összes variancia legalább 80%-át magyarázzák a főkomponensek, azaz a kumulált λ százalék legalább 80% legyen. Sváb szerint ez, ha túl sok változó van, magas követelmény. Egyesek a faktoranalízisben képletet is megadnak, hogy legfeljebb hány faktort érdemes meghatározni.

$$q \approx \frac{(2p+1)-\sqrt{8p+1}}{2}$$

Főkomponens-analízis forgatással

A faktoranalízist jóval korábban fejlesztették ki, mint a főkomponens-analízist, így a forgatást is eredetileg a faktoranalízisre dolgozták ki. Mind a faktoranalízisben, mind a főkomponens-analízisben ugyanazokat a forgatási módszereket használjuk. A faktoranalízis kidolgozásakor az volt az elképzelés, hogy p számú X változó kevesebb q<p számú háttérváltozóval, faktorral értelmezhető, mert ugyanaz a faktor több X változót értelmez.

Az X változó faktorsúlyai azonban többnyire megoszlanak kettő vagy annál is több faktorra, annak ellenére, hogy az ábrázolás szerint a változók csoportokba tömörülnek. A csoportok ugyanis nem egyetlen tengely mentén, hanem a hipergömb több tengelye által közrefogott valamely szektorában fekszenek. Ilyenkor a tengelyek elforgatásával meg tudjuk tenni, hogy a tengelyek áthaladjanak a csoportokon. Az azonos csoportokba tartozó X változók faktorsúlyai a közös faktorban -1-hez és +1-hez lesznek közel, a többi faktorban viszont nullához közelítenek. Így az eredeti p számú változót q<p számú faktorral tudjuk leírni, amelynek nagy része közös faktor, és sokszor szakmailag értelmezhető háttérváltozót ismerhetünk fel benne. Ezt az eljárást nevezik *faktorextrahálásnak*.

Amennyiben a forgatással nem sikerül az X változókat kevesebb számú faktorral előállítani, ez azt jelenti, hogy egyáltalán nincs vagy csak kevés a változócsoport, és az egyes csoportok is csak kevés változóból tevődnek össze.

A forgatás lehet derékszögű és ferdeszögű. A forgatás szöge mindig tengelypáronként értendő. A derékszögű forgatáskor az új koordináta rendszer is derékszögű marad, ezért a faktorok függetlensége továbbra is megmarad, a forgatás tehát ortogonális.

Ferdeszögű forgatás esetén a faktorok nem lesznek függetlenek, a közöttük fennálló korreláció mértéke:

$$r_{I,\,II} = \cos(90^{o} + \alpha_{I} - \alpha_{II})$$

ahol: α_I : az I., vízszintes tengely elfordításának szöge

α_{II} : a II., függőleges tengely elfordításának szöge

Ez a megoldás az „elsődleges faktor” (primary factor) szerinti ferdeszögű forgatás. Ennek továbbfejlesztése a ferdeszögű „vetületi vektorok” (reference vector) szerinti forgatás.

Derékszögű forgatás Varimax módszerrel

Többféle derékszögű forgatás létezik. A legelterjedtebb eljárás H.F. Kaiser módszere a Variamax rotáció. Ez elégíti ki legjobban a Thurstone-féle egyszerű struktúra követelményeit. A Varimax kritérium: a főkomponenssúly négyzetek oszloponkénti varianciáinak összege (*V*) maximum legyen.

$$V = \sum_{j=1}^{q} s^2_{a_j^2} \quad \text{maximum}$$

Ez akkor a legnagyobb, ha a főkomponens súlyok |1| és 0-hoz közeli értékek. A forgatás során megváltoznak az a_{ij} értékek. Az új $\mathbf{A_0}$ mátrix is ortogonális, soronként a a_{ij}^2 összege továbbra is 1, azonban az oszloponkénti összegük módosul, többé már nem azonosak $\boldsymbol{\lambda_j}$-vel. Az ortogonális forgatás a főkomponensekben azonban csak átrendezi a varianciák változónkénti megoszlását, de az összes varianciát nem módosítja.

A fenti Varimax kritériumban minden változó azonos súllyal, jelentőséggel vesz részt. Ezért a számítások során a tényleges kritérium: az a_{ij}^2 értékeket az x_i változók h_i^2 kommunalitásával súlyozzák.

$$V = p\sum_{j=1}^{q}\sum_{i=1}^{p}\left(\frac{a_{ij}^2}{h_i^2}\right)^2 - \sum_{j=1}^{q}\left(\sum_{i=1}^{p}\frac{a_{ij}^2}{h_i^2}\right)^2$$

*71. táblázat. Főkomponenssúly mátrix Varimax rotáció után, **Ao** mátrix*

Rotated Component Matrix[a]

	Component			
	1	2	3	4
Farinograf érték	,997	-4,30E-02	5,821E-02	3,083E-02
Sikér terülés	-,753	,639	-6,40E-03	,155
Sikér mennyisége	-,102	,951	,291	-1,02E-02
Fehérje %	5,514E-02	,212	,976	-5,15E-04

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

*72. táblázat. Transzformáló mátrix, **T** mátrix*

Component Transformation Matrix

Component	1	2	3	4
1	-,691	,681	,238	,055
2	,587	,341	,734	-,017
3	,421	,645	-,636	,048
4	,028	-,063	,030	,997

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

73. táblázat. Fökomponens-koefficiensek forgatás után

Component Score Coefficient Matrix

	Component			
	1	2	3	4
Farinograf érték	,900	,162	-,084	3,704
Sikér terülés	-,168	,039	,004	5,474
Sikér mennyisége	,177	1,108	-,252	-3,717
Fehérje %	-,108	-,340	1,105	,924

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

*74. táblázat. Főkomponens-változók forgatás után, **C** mátrix*

Fajta	C_1	C_2	C_3	C_4
Mironovszkaja 808.	1,32	,27	,19	,34
Fertődi 293.	,73	1,75	2,11	,41
Bezosztája	1,18	-,19	-,72	,44
Martonvásári 1.	,05	-,91	,01	,88
Martonvásári 2.	,91	-,21	-,19	,79
Martonvásári 16.	,24	1,38	,01	-1,60
Martonvásári 24.	,70	-,23	-2,04	-,22
Jubilejnaja	,53	-1,00	,01	-,65
Avróra	-,02	,16	-,22	-,21
GK-Fertődi 2.	-,95	-,54	,84	1,86
Kavkáz	-,54	-,28	-,09	-,90
Rannaja	-,83	-1,98	1,31	-1,08
Kiszombori	-1,28	,70	-,09	-1,16
Burgas	-2,04	1,09	-1,15	1,11
Összesen:	0	0	0	0

75. táblázat. Az összes variancia felbontása

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,152	53,810	53,810	2,152	53,810	53,810	1,575	39,364	39,364
2	1,332	33,290	87,100	1,332	33,290	87,100	1,360	33,998	73,362
3	,499	12,473	99,574	,499	12,473	99,574	1,040	26,011	99,373
4	1,704E-02	,426	100,000	1,704E-02	,426	100,000	2,509E-02	,627	100,000

Extraction Method: Principal Component Analysis.

40. ábra. A főkomponensváltozók ábrázolása

41. ábra. A változók háromdimenziós konfigurációja

Faktor-analízis

X változó h^2 részének (közös faktorok) felbontása történik. A faktorok lehetnek:

1. közös faktor

1.1 általános faktor

1.2 csoport faktor

2. egyedi faktor

3. hiba faktor

a hiba faktor származhat a lineáris korrelációs közelítésből is, ill. egyéb zavaró hatásokból. Ebben az eljárásban alapesetben a faktorok nem korrelálnak egymással. Csak a közös faktorokat számítjuk ki. A korrelációs mátrix főátlójába a kommunalitásokat helyettesítjük be.

Kategorikus főkomponens-analízis

A kategorikus főkomponens-analízis (CATPCA) az egyszerű PCA általánosítása kevert mérési szintű változók összefüggés-rendszerének elemzésére. Ez nagyban hasonlít a többszörös korrespondencia-analízishez. CATPCA segítségével például meghatározhatjuk az autómárkák és az ár, tömeg, üzemanyag-fogyasztás, egyéb közötti kapcsolatokat. Vagy osztályokba sorolhatjuk a különböző típusú autókat több jellemvonás egyidejű figyelembevételével.

A főkomponens-analízis célja a változók eredeti számának csökkentése kisebb egymással nem korreláló komponensekre, amik hordozzák az eredeti adatok információinak jelentős részét. A standard főkomponens-analízisben feltételezzük, hogy a változók között lineáris kapcsolat van. A kategorikus főkomponens-analízisben a változók közötti nem lineáris kapcsolatot modellezzük.

Hogyan lehet használni? Analyze, Data Reduction, Optimal Scaling…, Selected Analysis. Itt választhatjuk ki, hogy milyen kategorikus analízist szeretnénk végezni. Három közül választhatunk:

1. Többszörös korrespondencia-analízis
2. Kategorikus főkomponens-analízis
3. Nemlineáris kanonikus korreláció

A megfelelő analízis kiválasztása az Optimal Scaling Level és a Number of Sets of Variables rádiógombok kombinációjával történik.

Optimal Scaling Level:

1. Minden változó többszörös nominális változó
2. Néhány változó nem többszörös nominális változó (egy vagy több változó skála típusú a többi többszörös nominális. Vagy lehetnek még egyszerű nominális ordinális és diszkrét értékek is.

Number of Sets of Variables: meg kell adni hogy a változók csoportjából hányat akarunk összehasonlítani más változó csoportokkal.

1. Egy csoport
2. Több csoport

	Minden változó többszörös nominális	**Néhány nem**
Egy csoport	Többszörös korrespondencia analízis	Kategorikus főkomponens-analízis
Több csoport	Nemlineáris kanonikus korreláció	Nemlineáris kanonikus korreláció

NEM PARAMÉTERES PRÓBÁK

Chi-négyzet teszt

A Chi-négyzet teszt a változókat kategóriákba rendezi, és utána számítja ki a statisztikát. A teszt során a megfigyelt és feltételezett relatív gyakoriságokat hasonlítja össze. Lehetőségünk van több csoport eloszlásának homogenitását tesztelni vagy egy megadott relatív gyakorisággal való egyezés tesztelésére.

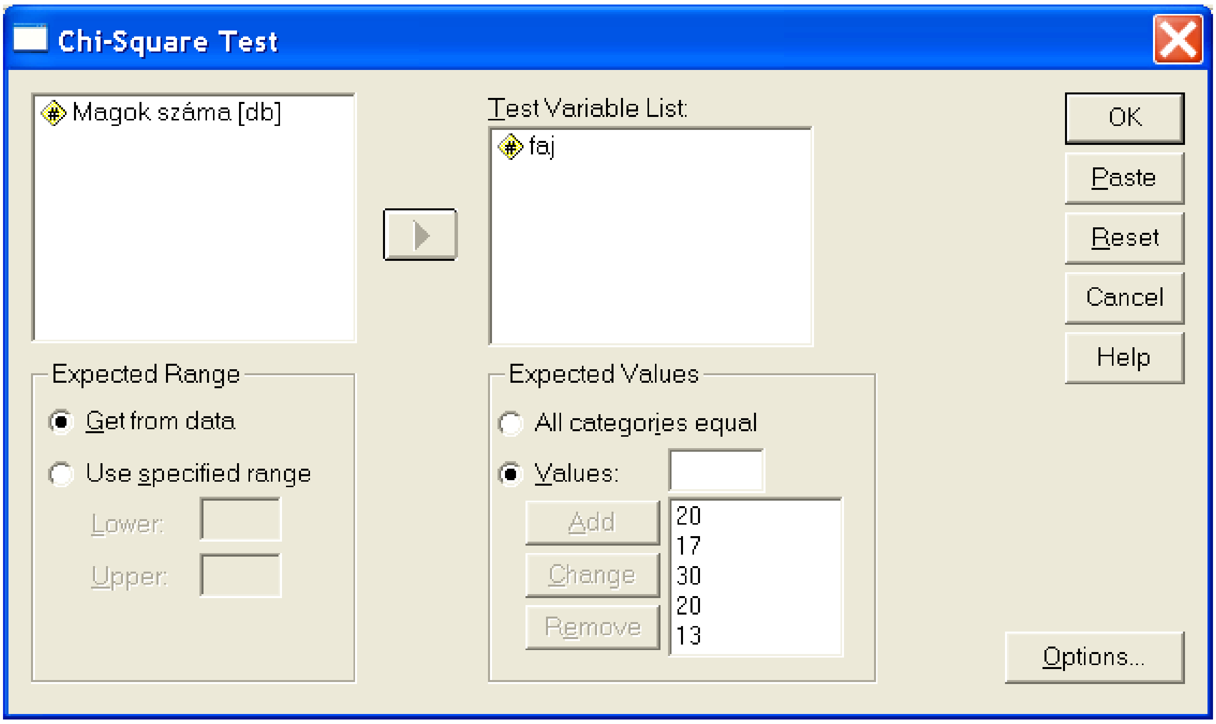
Feladat:

Származhat-e egy gyepmag keverék egy 20, 17, 30, 20, 13%-os összetételű keverékből? A mintavételezés során az alábbi eredményt kaptuk:

Faj	magok száma (db)
Réti perje	236
Angolperje	241
Réti komócsin	443
Réti csenkesz	252
Fehérhere	155
Összesen:	1 327

Az SPSS-ben a fenti adatbázissal csak akkor lehet gyakoriságokat számítani, ha a **Faj** változót súlyozzuk a **magok száma** változóval. Date, Weight Cases…

Analyze, Nonparametric Tests, Chi-Square.



FAJ

	Observed N	Expected N	Residual
1.00	236	265.4	-29.4
2.00	241	225.6	15.4
3.00	443	398.1	44.9
4.00	252	265.4	-13.4
5.00	155	172.5	-17.5
Total	1327		

Megfigyelt gyakoriság, várható gyakoriság és a kettő különbsége.

Test Statistics

	FAJ
Chi-Square[a]	11.827
df	4
Asymp. Sig.	.019

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 172.5.

A vetőmagkeverék aránya nem felel meg az előírásnak. Mi lehet ennek az oka?

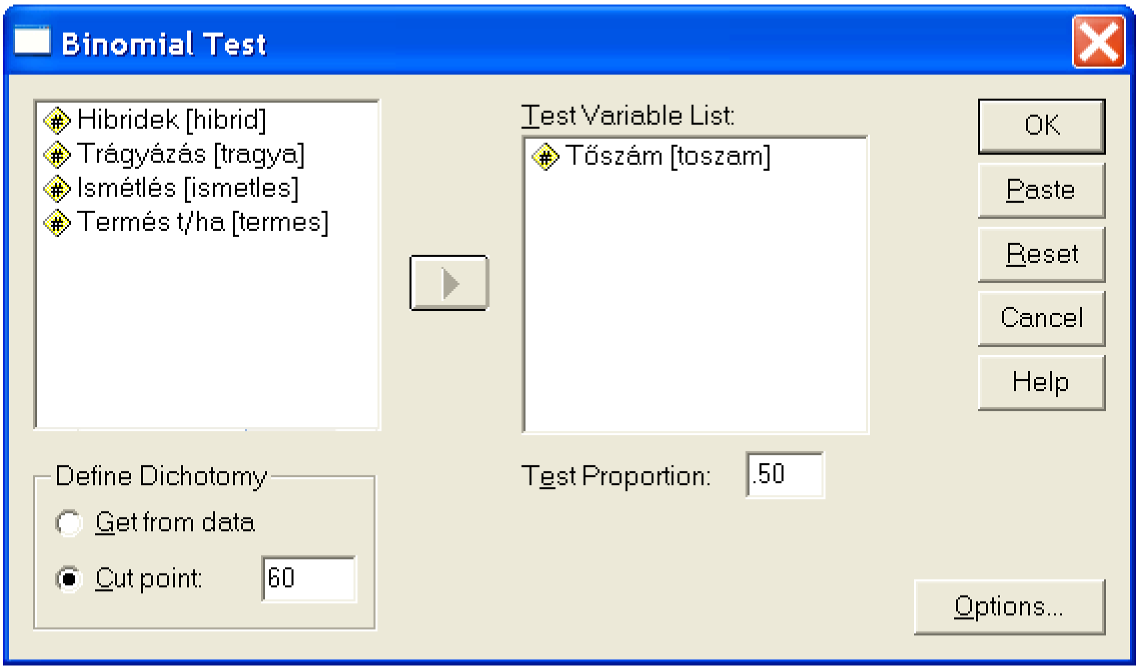
Binomiális teszt

Ezzel a teszttel két csoport (kategória) megfigyelt relatív gyakoriságát lehet összehasonlítani a binomiális eloszlás alapján. A valószínűségi változó kezdeti értéke mindkét csoportban 0,5. A valószínűség megváltoztatásakor az első csoport előfordulását tesztelhetjük. A második csoport előfordulásának valószínűsége 1 mínusz az első csoportra megadott valószínűség.

Feladat:

Megegyezik a 60 ezer alatti tőszám parcelláinak száma az e feletti parcellák számával?

Tőszám, Cut point 60. Test Proportion 0.5, OK.



Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
Tõszám	Group 1	<= 60	36	.50	.50	1.000[a]
	Group 2	> 60	36	.50		
	Total		72	1.00		

a. Based on Z Approximation.

A két csoport relatív gyakorisága tökéletesen megegyezik.

Runs Test

Sok statisztikai teszt alkalmazásának feltétele, hogy a mintában a megfigyelések függetlenek legyenek. Ezt csak akkor tudjuk leellenőrizni, ha ismert a mintaelemek kihúzásának időpontja vagy sorrendje, főként idősoros elemzésnél hasznos. Ezzel a teszttel leellenőrizhetjük, hogy a mintánk véletlen mintának tekinthető-e. Legelőször választani kell egy jellemző értéket – ami legtöbbször valamilyen centrális mutató – és ehhez hasonlítjuk a megfigyelt értékeket. Az eljárás során a változó mindenegyes értékét osztályozzuk, hogy a töréspont alatt vagy felett helyezkedik el. Ez után megállapítjuk, hogy van-e valamilyen szabályosság a sorozatban, hányszor ismétlődik egymásután ugyanabba az osztályba tartozó elem, azaz egy sorozat. Egy sorozatnak (run) legalább egy tagja van.

A teszt eredménye attól is függ, hogy mit választunk ki töréspontnak (medián, módusz, átlag, stb.).

Az eloszlás egy bimodális sokaságot mutat melynek két módusza van. Az SPSS a módusz meghatározásakor a nagyobbikat adja meg.

Descriptive Statistics

		Web Site Rating
N		32
Mean		9.94
Std. Deviation		2.368
Minimum		6
Maximum		14
Percentiles	25th	8.00
	50th (Median)	10.00
	75th	12.00

Runs Test

	Web Site Rating
Test Value[a]	10.00
Cases < Test Value	14
Cases >= Test Value	18
Total Cases	32
Number of Runs	10
Z	-2.283
Asymp. Sig. (2-tailed)	.022

a. Median

Ha a minta teljesen véletlen lenne, akkor a sorozatok száma 17 körüli lenne. Mivel a megfigyelt sorozatok száma csak 10, ezért a Z-statisztika értéke negatív. Túl alacsony a szignifikancia értéke, ezért nem tekinthető véletlennek a minta.

Rating	cut point 10
8	1
7	1
8	1
6	1
10	2
8	1
6	1
7	1
8	1

9	1
7	1
10	2
7	1
8	1
12	2
10	2
12	2
9	1
11	2
12	2
10	2
13	2
13	2
12	2
11	2
14	2
9	1
14	2
11	2
12	2
11	2
13	2

Runs Test 2

	Web Site Rating
Test Value[a]	12[b]
Cases < Test Value	22
Cases >= Test Value	10
Total Cases	32
Number of Runs	16
Z	.315
Asymp. Sig. (2-tailed)	.752

a. Mode

b. There are multiple modes. The mode with the largest data value is used.

A módusz alatt több, mint kétszer annyi elem fordul elő, mint felette. Ennek az az oka, hogy 12 felett az adatoknak csak a 25%-a helyezkedik el. Mivel a teszt a töréspont alatti ill. feletti elemeket különíti el, a várható sorozatok száma a törésponttól függ. Ebben az esetben a sorozatok várható száma 15 körüli. A számított érték nagyon közel van hozzá, ezért véletlennek tekinthető a minta, amit a szignifikancia értéke is megerősít.

Runs Test 3

	Web Site Rating
Test Value[a]	8
Total Cases	32
Number of Runs	11
Z	.000
Asymp. Sig. (2-tailed)	1.000

a. User-specified.

A bimodális eloszlás első móduszát választottuk töréspontnak. A sorozatok várható száma ekkor 11. A számított érték pontosan megegyezik a várható értékkel, ezért a minta véletlennek tekinthető.

Egy-mintás Kolmogorov-Smirnov teszt

Egy-mintás Kolmogorov-Smirnov teszt angolul One-Sample Kolmogorov-Smirnov Test. Milyen eloszlásba tartozik a minta? Normál, Poisson, egyenletes (Uniform) és exponenciális (Exponential) eloszlás tesztelése. A megfigyelt adatok kumulatív eloszlás függvényét (empirical cumulative distribution function, ECDF) hasonlítja össze a teoretikus eloszlás kumulatív függvényével. Adott *N* számú nagyság szerint sorba rendezett $Y_1, Y_2, \ldots, Y_n$ változó, akkor az ECDF az alábbi:

$$E_N = n(i)/N$$

ahol

n(i) : adatok száma, amelyek kisebbek, mint Y_i.

A lépésköz *1/N*, azaz minden egyes adat figyelembe kerül a meghatározás során.

A Kolmogorov-Smirnov Z-érték a megfigyelt és teoretikus kumulált eloszlás függvények közötti legnagyobb abszolút különbségből számítják, ezt az értéket szorozzák a megfigyelések négyzetgyökével. Statisztikája:

$$D = \max_{1 \leqslant i \leqslant 1} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right)$$

ahol:

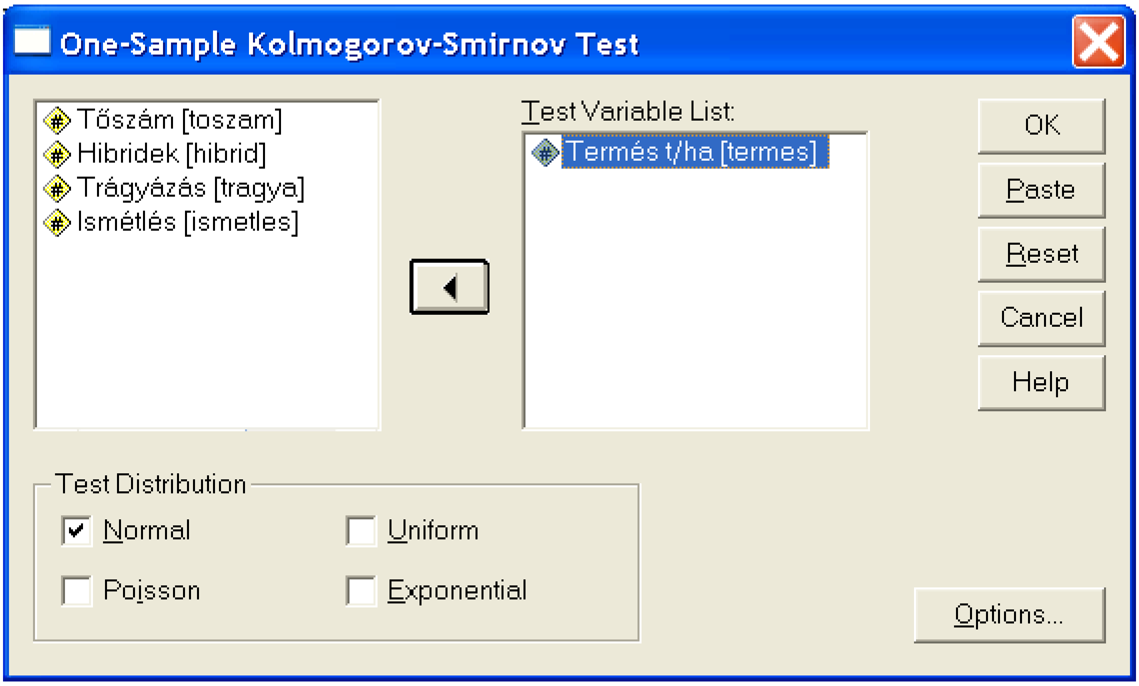
F : elméleti eloszlásfüggvény

A kritikus értékekre Birnbaum 1952-ben készített táblázatott.

- Birnbaum, Z. W. (1952). "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size", Journal of the American Statistical Association, 47, page 425.

$$Z = D\sqrt{N}$$

Sok paraméteres teszt megköveteli, hogy a változó normális eloszlású legyen.



Az eloszlás paramétereit a mintából becsüli, pl. normális eloszlás esetén ez a számtani átlagot és a szórást jelenti.

One-Sample Kolmogorov-Smirnov Test

		Termés t/ha
N		72
Normal Parameters[a,b]	Mean	9.69609
	Std. Deviation	1.843756
Most Extreme Differences	Absolute	.075
	Positive	.047
	Negative	-.075
Kolmogorov-Smirnov Z		.635
Asymp. Sig. (2-tailed)		.814

a. Test distribution is Normal.

b. Calculated from data.

A nullhipotézis: a mért változó normál eloszlású. A hipotetikus és mért eloszlás nem különbözik egymástól. A nullhipotézist megtartjuk, mivel nagyon kicsi az eltérés a kettő között, és a szignifikancia szint is magas.

A teszt alkalmazásának korlátai:

- csak folytonos eloszlású változóra alkalmazható
- a teszt sokkal érzékenyebb a centrum körüli adatokra, mint az eloszlás két szélén lévőkre
- az eloszlást teljes mértékben meg kell határozni. Az eloszlás paramétereit a mintából kell becsülni. A kritikus tartomány nem érvényes

A két utolsó hiányosság miatt a Anderson-Darling tesztet ajánlják helyette.

Független két-mintás tesztek

Angolul Two Independent Sampes Tests.

Wilcoxon W statisztika

Mann-Whitney U-próba

Két független minta medián egyezésének igazolására való eljárás (két-mintás t-teszt). A nullhipotézis, hogy a két sokaság ugyanabba az eloszlásba tartozik. Ordinális típusú adatoknál használható, vagy skála típusú adatoknál, ahol nem feltétel a normál-eloszlás. Csak az egyezésre ad elfogadható, megbízható eredményt. Ha ettől eltérő eredményt kapunk, nem tudhatjuk biztosan, hogy mi a valóság.

Alkalmazási feltétel:

- Hasonló alakú eloszlások (tesztelhető a két-mintás Kolmogorov-Smirnov próbával)
- Független minták

Null hipotézis: M(x) = M(y). A hipotézisvizsgálat céljára konstruált valószínűségi változó: n_1+n_2 elemű mintából egyetlen rangsor felállítása, „x” mintára vonatkozó rangszámok összege: R1 vagy W-érték.

$$m = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

A próba változójának eloszlása, ha n_1 és n_2 elég nagy, megközelítően N(m, σ).

Kolmogorov-Smirnov Z-próba

Két eloszlás összehasonlítására szolgáló eljárás. A nullhipotézis, hogy a két sokaság ugyanabba az eloszlásba tartozik. A Kolmogorov-Smirnov Z-értéket a két csoport kumulált eloszlás függvényei közötti legnagyobb abszolút különbségből számítják. A változóknak ezért illik folyamatos eloszlásúnak lenni. A két csoportban a megfigyelések számának nem kell megegyeznie. Nagyon rugalmas a teszt, nem kell az eloszlásoknak hasonló alakúnak lennie, hisz az eljárás ezt is teszteli.

Alkalmazási feltétel:

- Csak folytonos eloszlások hasonlíthatók össze.
- Független minták

A próba érzékeny a helyzeti különbségekre és az eloszlások alakjára. A helyzeti különbség azt jelenti, hogy a két eloszlás hol helyezkedik el a skálán. A Kolmogorov-Smirnov teszt akkor is különbözőnek mutatja a két eloszlást, ha az alakjuk (shape) megegyezik, de egymástól távol helyezkednek el. Ezek szerint két eloszlás akkor különbözik, ha vagy az alakjuk, vagy az elhelyezkedésük különbözik, vagy mindkettő. Amennyiben a két eloszlás helyzeti különbsége nem érdekel bennünket, toljuk el a skálát az origóra, aminek a legegyszerűbb módja az adatok standardizálása (ettől az eloszlások alakja semmit sem változik). A standardizálással skálaeltolást és skála transzformációt is végrehajtunk egyszerre.

Alternatívaként használhatjuk a Crosstabs eljárásokat is kettő vagy több ordinális vagy nominális változó közötti különbség kimutatására.

Amennyiben a t-teszt alkalmazásának feltételei teljesülnek, akkor azt kell használni.

Moses extreme reactions

Wald-Wolfowitz runs

Több független mintás teszt (K Independet Samples...)

Kruskal-Wallis H próba

Rendezett mintán alapuló, több mintás hipotézis vizsgálat, amelynek null hipotézise: minden minta azonos eloszlású sokaságból származik. A próba segítségével *„h”* darab *„nh”* elemszámú mintát vizsgálhatunk. Ezt ismételt Wilcoxon-póbákkal is elvégezhetnénk, de ebben az esetben az ismétlések megnövelik az elsőfajú hibát (analóg a középértékek többszörös összehasonlításának, szimultán próbák problematikájával).

Medián teszt

Két páronként összetartozó minták tesztjei (2 Related Samples…)

Wilcoxon teszt (Wilcoxon signed-rank test)

Két eloszlás egyezésének vizsgálatára alkalmas. Sokszor használják két várható érték egyezésének vizsgálatára is. A két minta elemei páronként összefüggnek. n_1+n_2 elemű mintából egyetlen rangsort képeznek. Konstruált valószínűségi változó *„u”*. A nullhipotézis: a páronkénti különbségek a nulla körül szimmetrikusan helyezkednek el.

Előjel próba (Sign)

Összetartozó elem párok vizsgálata. Hipotézis, hogy x_1 … x_n minta elemei nagyobb (vagy kisebb) értéket vesznek fel, mint y_1

… y_n elemei, ahol az azonos indexű minta elemek között valamilyen logikai kapcsolat van (pl. ugyanazon jelenség két különböző időpontban vagy helyen mért értékei).

Első lépésben meghatározzuk az x_i-y_i különbségek előjelét, utána megszámoljuk, hogy hány darab „-„ és „+” előjelű különbség adódott. Az előjel próba, ellentétben a rendezett mintás próbákkal szemben, kisebb elemszámokra erősebb. Így kétszeresen nem indokolt nagy elemszámok esetén az előjel próba használata: Nagyobb minták esetén relatíve gyengébb a próba ereje. Elvész az előjel próba jelentős előnye, a gyors alkalmazhatóság.

McNemar teszt

Kétértékű, bináris vagy dichotóm változók összehasonlítására szolgáló módszer. Tipikusan megismételt mérések esetében használható, amikor ugyanazon egyedeket figyeljük meg: bizonyos esemény bekövetkezése (pl. kezelés) megváltoztatja-e az egyedek állapotát (az esemény előtti és utáni állapot összevetése). Nullhipotézis: a kezelés utáni állapot egyenlő a kezdeti állapottal.

Ez a teszt főként nominális vagy ordinális változók tesztelésére alkalmas.

K számú összetartozó minta tesztjei (k Related Samples…)

Friedman teszt

Több eloszlás homogenitás vizsgálatára alkalmas, összetartozó több változó esetén. Paraméteres megfelelője a kéttényezős variancia-analízis. Feltételezzük, ha az eloszlás megegyezik a várható érték is megegyezik nagy valószínűséggel. Fordítva ez nem igaz. Null hipotézis: a *k* darab összetartozó változó ugyanabba a sokaságba tartozik.

F(x) = G(x) = … =K(x)

Alkalmazási feltétel: több rendezett minta azonos elemszámokkal, *g* és *h* elég nagy, ahol g a minta elemszáma a szempont egy szintjére (blokk), 'h' a szempontonkénti vagy szintenkénti minták száma (kezelés).

A próba változójának eloszlása *Chi-négyzet*, szabadságfoka *k-1*.

	1	r_{11}	…	r_{1k}	A minta elemeinek sorrendje az első szempont szerint
Blokk	⋮			⋮	
	g	r_{g1}	…	r_{gk}	A minta elemeinek sorrendje az utolsó szempont szerint
		$\mathbf{R_1}$		$\mathbf{R_k}$	A változók átlagos rangszámai.

Megjegyzés: a Friedman-teszt és a Kendall-féle konkordancia együttható ugyanannak a problémának a tesztelésére használható. Szignifikancia-szintjeik megegyeznek, mindkettő kéttényezős problémát tárgyal.

Kendall konkordancia együtthatója W

Kettőnél több „bíráló" rangsora áll rendelkezésre. Van-e különbség a bírák együttesét tekintve a közöttük lévő egyetértésnek, vagy van-e szignifikáns mértéke? Milyen az egyetértés (konkordancia) a rangsorok együttesében.

(egyáltalán nem egyezik a bírálók véleménye) $0 <= W <= 1$ (tökéletesen egyezik a bírálók véleménye). A próba változójának eloszlása Chi-négyzet, szabadságfoka *m-1*.

Pl.: több oktató a hallgatókat rangsorolja tudás szerint. Minden oktató sorba rendezi a hallgatót *1*-től *m*-ig, *m* a hallgatók száma. Az oktatók száma legyen *n*. Vajon megegyeznek az oktatók véleményei, van közöttük egyetértés?

Hallgatók Oktatók	Hallga- tó1	Hallga- tó2	Hallga- tó3	Hallga- tó4	Hallga- tó5	Hallga- tó6	Hallga- tó7	Hallga- tó8
Kovács	**6**	**2**	**3**	**5**	**1**	**4**	**8**	7
Kiss	**6**	**3**	**1**	**7**	**2**	**4**	**8**	5
Szabó	7	3	2	5	1	4	8	6

Test Statistics

N	3
Kendall's W[a]	.931
Chi-Square	19.556
df	7
Asymp. Sig.	.007

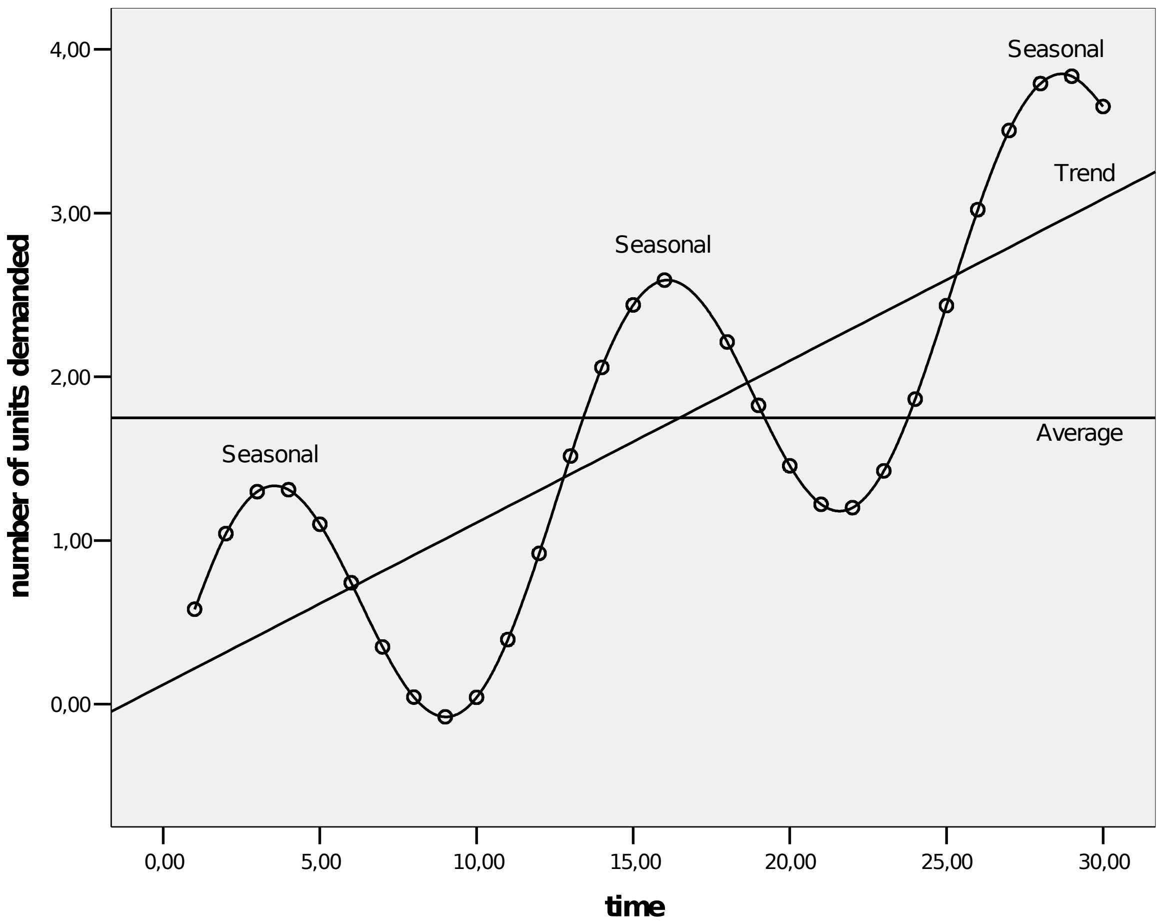
a. Kendall's Coefficient of Concordance

Cochran Q próba

A McNemar teszt kiterjesztése k számú mintára.

IDŐSOROK ANALÍZISE

Az idősornak a különböző időpontokban végzett megfigyeléseket nevezzük. Az adatok sorrendje nagyon fontos, a különböző rendezéseknek itt nincs értelme. A megfigyeléseket egyenlő időközökben végezzük. *t=1 ,2 , 3, … n*.



A mért értékeket u_1, u_2, u_3 … u_t-vel jelöljük.

t=0 időpontból kiindulva nemcsak előre, hanem hátra is haladhatunk, ekkor az indexeket -1, -2 stb. jelöljük.

Az idősorok elmélete más típusú adatokra is alkalmazhatók, pl. földsáv mentén különböző pontokban mért nitrogéntartalom, amelyben az időbeli változás helyébe térbeli változás lép fel. Műtrágyadózisok is felfoghatók idősornak. A módszer felhasználható olyan esetben, ahol egy valószínűségi változó egy „t” változótól függ, ahol „t” egyaránt vonatkozhat időre vagy lineáris térre.

Az „*u*” változó lehet diszkrét, pl. emberek száma, és lehet folytonos változó, pl. hőmérséklet, légnyomás, stb.

Az idősornak négy összetevője lehet:

1. Trend, hosszú időszakon keresztül érvényesülő változás
2. Szezonális ingadozás, rövid ideig tartó szisztematikus ingadozás
3. Periodikus ingadozás, mely hosszabb időtávon jelentkezik
4. Véletlen ingadozás

Az idősorok analízise során ezt a négy összetevőt kell elkülöníteni, ami sokszor elég nehéz feladat.

Trend

Regressziós technikával valamilyen alkalmas függvény illesztése az adatokra. Lehet lineáris ill. nem lineáris, pl. polinomok illesztése. A magasabb fokú polinomok illesztése azonban sokszor hátrányos, mivel sok számítást igényel, és újabb tagok csatolása esetében az illesztést elölről kell kezdeni.

Mozgóátlagolás. A leggyakoribb a 3, 5, 7, 9, 15 és 21 pontos mozgóátlagolás. Ezzel a módszerrel a szezonális hatások kiküszöbölhetők.

Rövid lejáratú szezonális és véletlen összetevők

Feltételezzük, hogy az idősor trendmentes, vagy a trendet már korábban kiszűrtük (detrendelés). Sorozatunk ekkor többé-kevésbé szabálytalanul ingadozik valamilyen középponti érték körül.

A sorozat véletlenszerűségének vizsgálata

Vizsgáljuk meg, hogy milyen sorozatot várhatunk abban az esetben, ha az ingadozás teljesen véletlenszerű, azaz ha az egymást követő tagok függetlenek, és a sorozat egy ismeretlen sokaságból származó minta véletlen elrendezéseként fogható fel. Az ettől az állapottól való eltérést különböző mérőszámokkal mérhetjük, pl.:

- Csúcspontok és mélypontok előfordulása a sorozatban
- A szomszédos tagok közötti korreláció

Csúcspontok és mélypontok előfordulása a sorozatban

$u_{t-1}<u_t>u_{t+1}$ csúcspont vagy $u_{t-1}>u_t<u_{t+1}$ mélypont. Mindkét esetben u_t fordulópont. Két fordulópont közötti intervallumot fázisnak nevezünk. Egy oszcilláló idősor véletlenszerűsége a fordulópontok számának meghatározásával jól vizsgálható, ez *n* tagú véletlen sorozatban $\frac{2}{3}(n-2)$ várható értékű és *(16n-29)/90* szórású.

A szomszédos tagok közötti korreláció, sorozatkorreláció

Egy sorozat szomszédos tagjai korrelációs együtthatóját elsőrendű autókorrelációs együtthatónak nevezzük. A *k* távolságra lévő tagok korrelációs együtthatójának elnevezése *k*-ad rendű autókorrelációs együttható.

$$\rho_k = \frac{\operatorname{cov}(u_t, u_{t+k})}{D(u_t)D(u_{t+k})}$$

Hosszú sorozatban $D^2(u_t)$ és $D^2(u_{t+1})$ gyakorlatilag azonosak, és emiatt a fenti képlet a következő módon adható meg:

$$\rho_k = \frac{\operatorname{cov}(u_t, u_{t+k})}{D^2(u_t)}$$

Rövid megfigyelési sorozatok esetében $D^2(u_t)$ becslésének jobb az egész sorozat (*n* tagból számított) szórásnégyzetét tekinteni, bár a kovarianciát csak *(n-k)* tagból határozzuk meg. Hasonlóképpen jobb u_t és u_{t+k} szorzatösszegének meghatározásánál az u eltéréseket a teljes sorozat számtani közepétől mérni.

Amennyiben a sorozat tagjait az összes tag számtani közepétől mérjük, akkor:

$$r_k = \frac{n}{n-k} \frac{\sum_{t=1}^{n-k} u_t u_{t+k}}{\sum_{t=1}^{n} u_t^2}$$

Amennyiben a sorozat véletlen jellegű, akkor ρ_k elméleti értéke minden k-ra nulla. Ennélfogva a sorozatkorrelációs együtthatók nullától való eltérését felhasználhatjuk a sorozat véletlenszerűségének vizsgálatára. Véletlen sorozatban nagy n-re ρ_k szórásnégyzete közelítően:

$$D^2(r_k) \approx \frac{1}{n-k}$$

A ρ_k autókorrelációs együtthatót k függvényében ábrázoló görbét korrelogrammnak nevezzük. Ennek segítségével megkülönböztethetők a harmonikus sorozatok és az autoregresszív sorozatok.

Periodogram-elemzés

Számos oszcilláló fizikai jelenség bizonyos számú „tiszta” harmonikus hullámra bontható fel, amely mindegyike egy-egy szinusz vagy koszinusz függvénnyel írható le. Egy tiszta oszcillátor időbeli mozgása az $A \sin\left(\alpha + \frac{2\pi}{\lambda} t\right)$ függvénnyel fejezhető ki, ahol λ a hullámhossz és A az amplitúdó. Az oszcillációs jelenség pedig gyakran állítható elő ilyen tagok összegeként:

$$u_t = A_1 \sin\left(\alpha_1 + \frac{2\pi}{\lambda_1} t\right) + A_2 \sin\left(\alpha_2 + \frac{2\pi}{\lambda_2} t\right) + \ldots +$$

Idősor periodicitásának keresése harmonikus analízis segítségével

Nincs zavar. Egy rádiókészülék behangolásával hasonlítható össze. Ismert hullámhosszú sorozatokat korrelálunk az adott sorozatokkal, ha összhangba jutnak, akkor intenzív korrelációt kapunk. Hibákkal terhelt, régi módszernek tartják napjainkban.

Autoregresszív sorozatok

Autoregresszív sorozatnak olyan sorozatot nevezünk, amely minden pontban az előző pontban felvett értékek, plusz egy zavar függvénye. Amennyiben a függvény lineáris, akkor lineáris autoregresszív függvényről beszélünk. A módszer figyelembe veszi, hogy zavar előfordulása esetén ez a rendszer változóiba beleolvad. Nem szabályos ingadozáshoz hasonlít, melyet néha meglöknek. A kilengések közötti idő nem állandó, valamint a kilengés sem mindig azonos mozgású. Nagyon hasonló ahhoz, ahogyan sok oszcilláló idősor viselkedik. Ebből kifolyólag az autoregresszív sorozatnak nincs szigorú értelemben vett periódusa. A csúcspontok közötti átlagos távolság teljesen különbözhet a korrelogram periódusától. Tegyük fel, hogy egy rendszer mozgását két tényező határoz meg:

A. Belső tulajdonságainak összessége, pl. rugalmasság, kényszer, ezek a külső hatás nélküli mozgást határozzák meg

B. Külső lökések sorozata

Az autoregresszív sorozatokban a két legfontosabb eset:

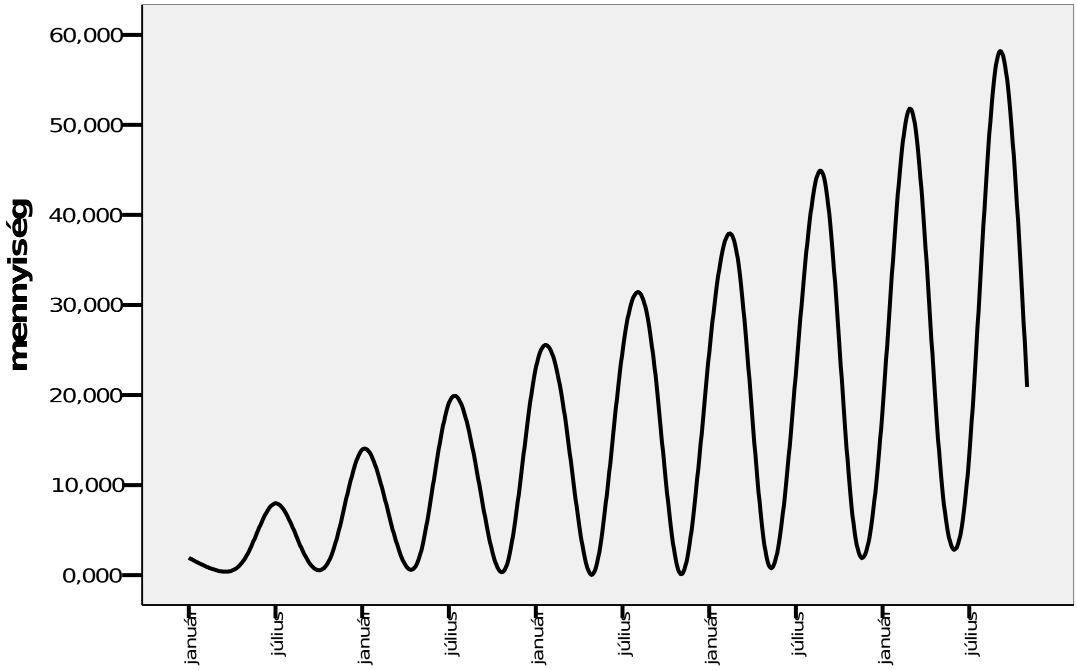
$$u_{t+1} = \mu u_t + \varepsilon_{t+1} \qquad (1.)$$

$$u_{t+2} + \alpha u_{t+1} + \beta u_t = \varepsilon_{t+2} \qquad (2.)$$

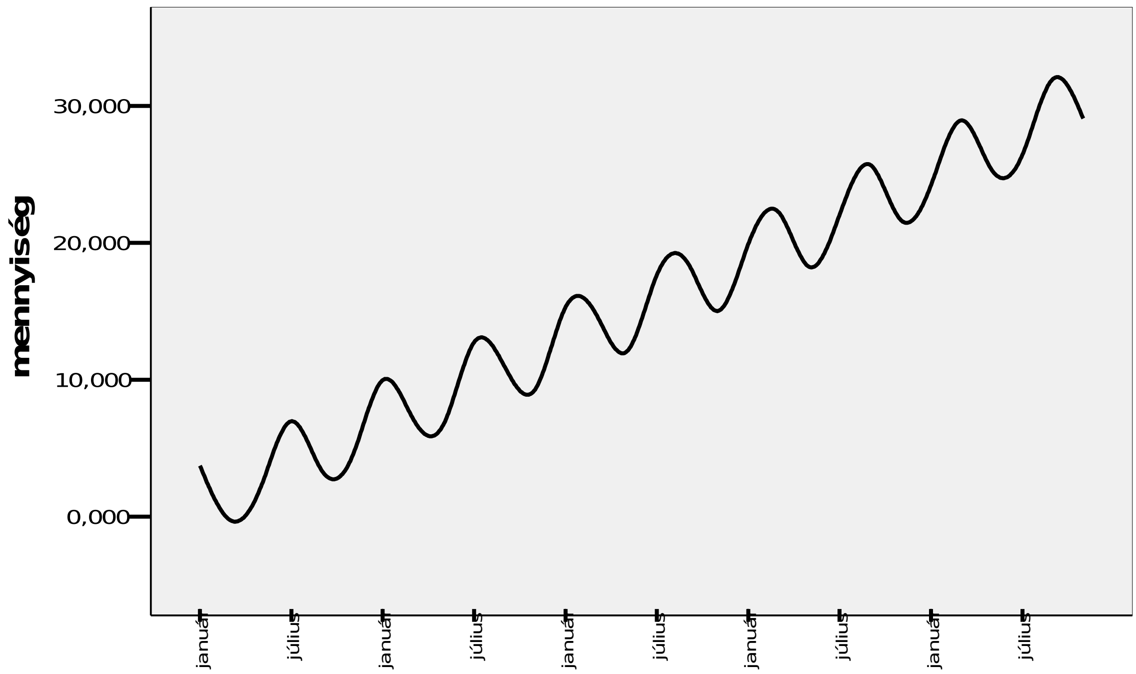
Exponenciális simítás

Négyféle modellt állíthatunk fel a trend és szezonalítás figyelembe vételének kombinációjával.

– Egyszerű (Simple) modell: nincs trend és nincs szezonális hatás, ill. változás.
– Holt modell: lineáris trend szezonális hatás nélkül.
– Winters modell: lineáris trend és multiplikatív szezonális hatás. Az ingadozás nagysága nő vagy csökken a sorozat értékétől függően
– Felhasználó által definiált modellek: a felhasználó állíthatja be a trend és szezonális hatásokat.



Multiplikatív szezonális dinamika, növekvő ingadozás



Additív szezonális dinamika

A fenti modellek négy paramétert használnak

Alfa (általános paraméter), minden modell használja, értéke 0,00-1,00. Ha az alfa 1, kizárólag a legfrissebb megfigyeléseket használjuk, ha alfa 0, akkor a régebbi megfigyelések is befolyásolják az aktuális érték alakulását.

Gamma Akkor használjuk, ha feltételezzük, hogy az idősornak van trendje. Értéke 0,00-1,00. A gammát csak lineáris vagy exponenciális trendnél, vagy aszimptotikus (damped, csillapított, amelynél a változás sebessége az idő folyamán egyre lassul) trendnél, ahol nincs szezonális komponens, használjuk. Egyszerű modell esetében nincs értelme.

Delta. Ez a paraméter a szezonalítást írja le. Értéke 0,00-1,00, egyhez közeli értéke magasabb súlyt jelent. Csak szezonális hatást tartalmazó modellben kerül meghatározásra, nem használjuk egyszerű, ill. Holt modell esetében.

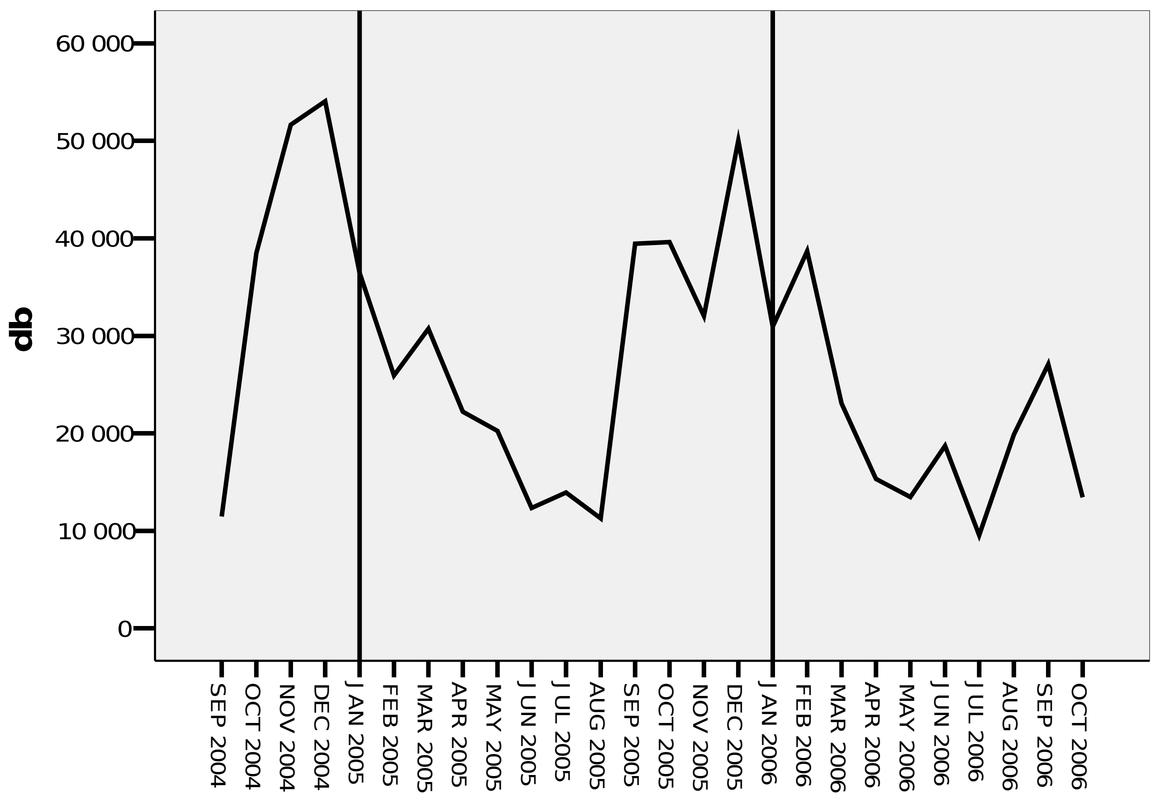
Phi. Az exponenciális simítás ezen paramétere kontrollálja, hogy a trend „damped”, csillapított vagy milyen gyorsan csökken a nagysága az idő függvényében. Értéke 0,00-1,00 (de soha sem éri el az egyet), egyhez közeli értéke nagyobb fokú csillapítást jelez. A phi csak csillapított trendet tartalmazó modellben használható, nincs értelme a szimpla, a Holt ill. a Winters modellben.

Az exponenciális simítás legelső lépése az ábrázolás, mivel a adatok időbeli alakulása segít a megfelelő modell kiválasztásában.

Van-e a sorozatnak egyáltalán trendje? Milyen a trend: változatlan vagy változik az idő függvényében?

Látható-e az adatokon szezonalítás? A szezonális ingadozások idővel nőnek, vagy változatlanok.

Válasszuk a grafikonok közül a Szekvenciális grafikonokat.

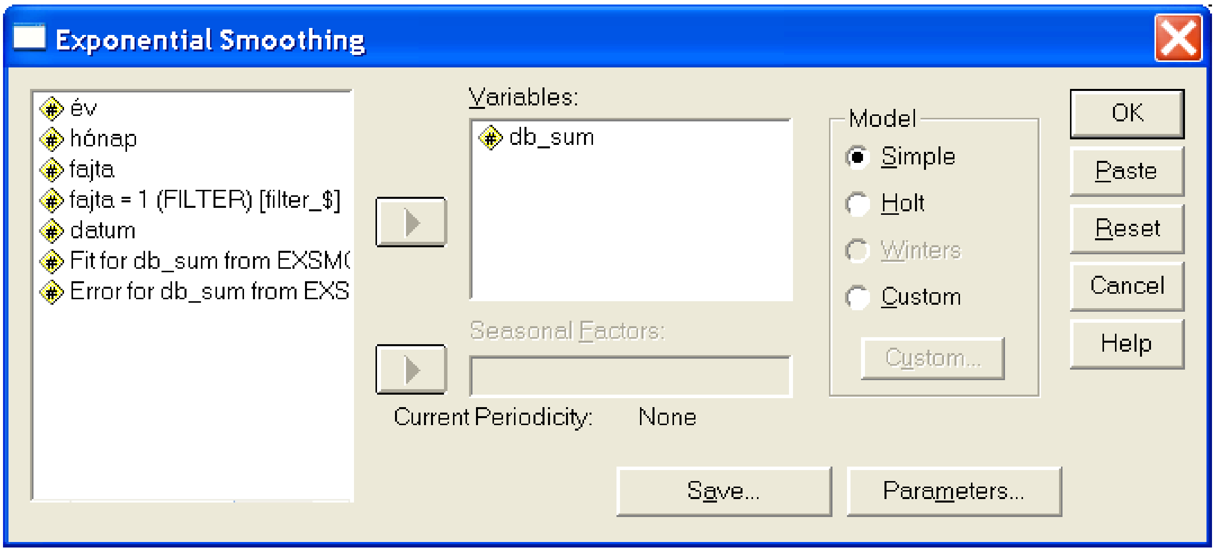


Filteres teafogyasztás

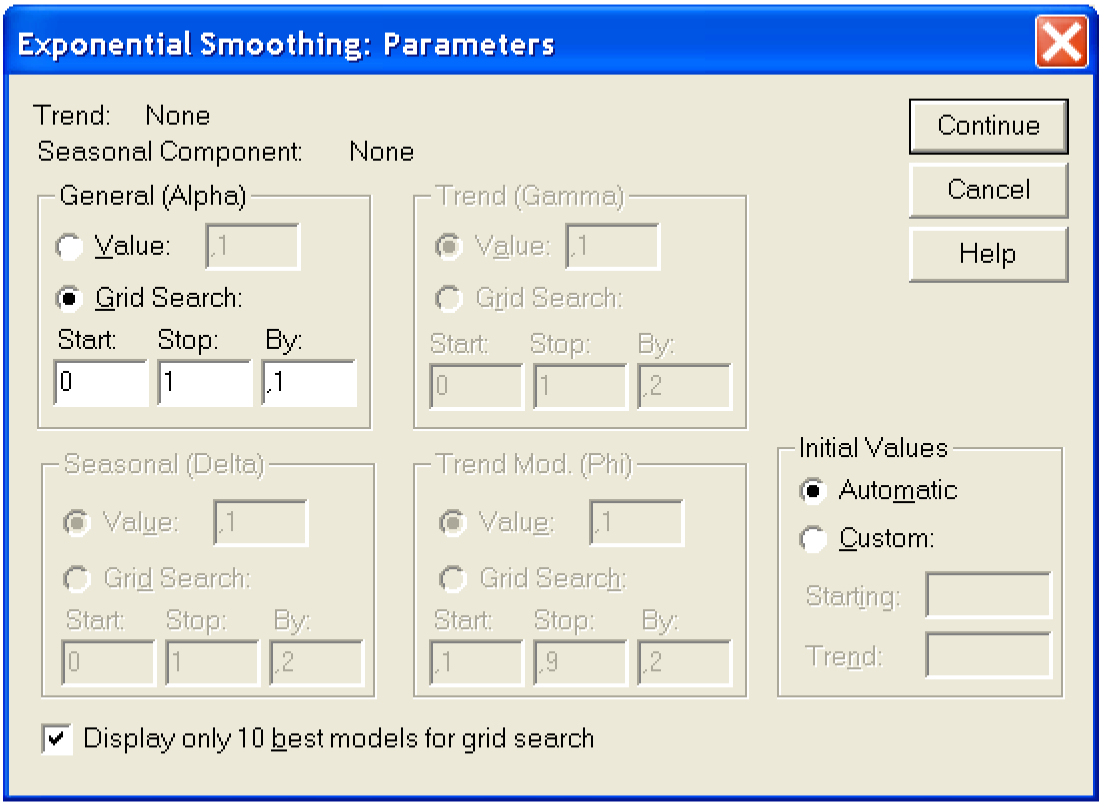
Trend nehezen ismerhető fel, vagy nincs, vagy enyhén csökkenő lineáris trendet feltételezhetünk. (nincs elég adat, hogy biztonságosan megítéljük konstans-e a trend).

Szezonális dinamika figyelhető meg: a hidegebb hónapokban több, a nyári időszakban kevesebb teát fogyasztanak az emberek.

A fentiek ismeretének ellenére legelőször a legegyszerűbb modellt állítsuk fel, ahol nincs trend és nincs szezonális hatás.



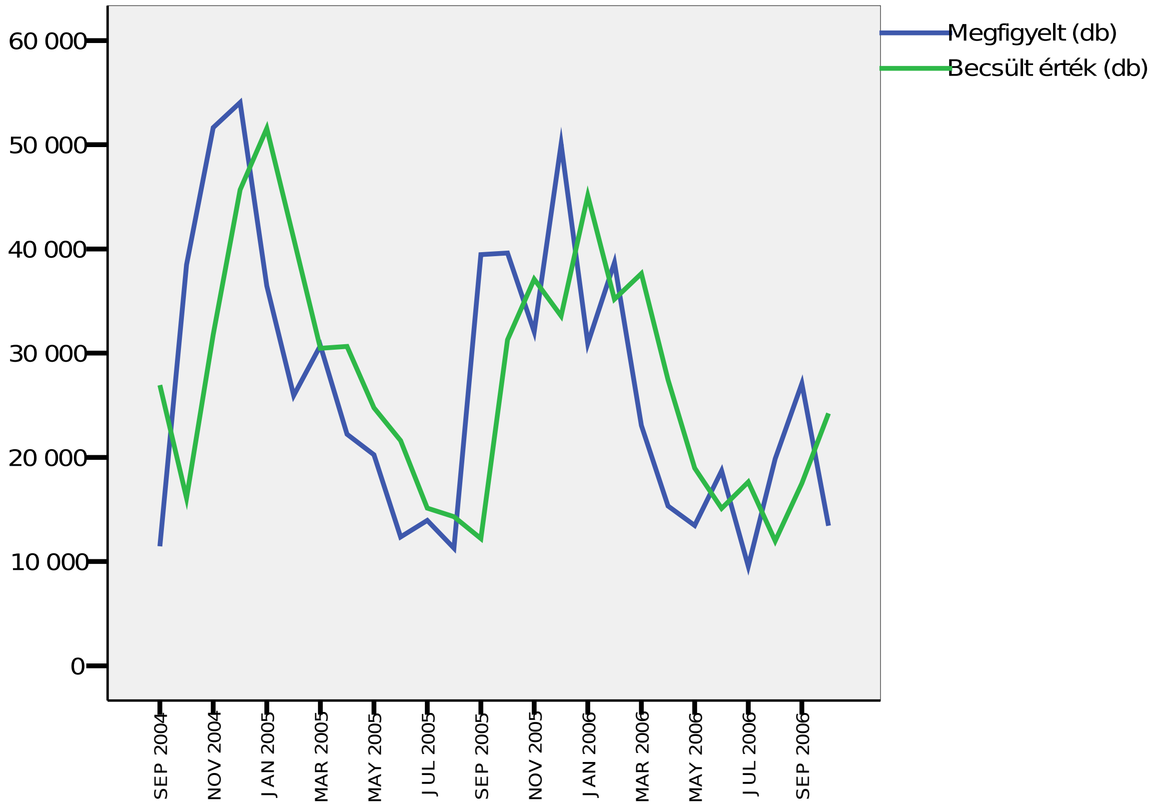
Az alfa paraméter meghatározását bízzuk a programra (válasszuk a Grid Search lehetőséget). Írassuk ki a legjobb 10 modell paramétereit.



A legjobb 10 modell alfa értéke az alábbi volt. A legpontosabb értéket alfa=0,7 értéknél kaptuk, ekkor volt az eltérés négyzet-összeg a legkisebb.

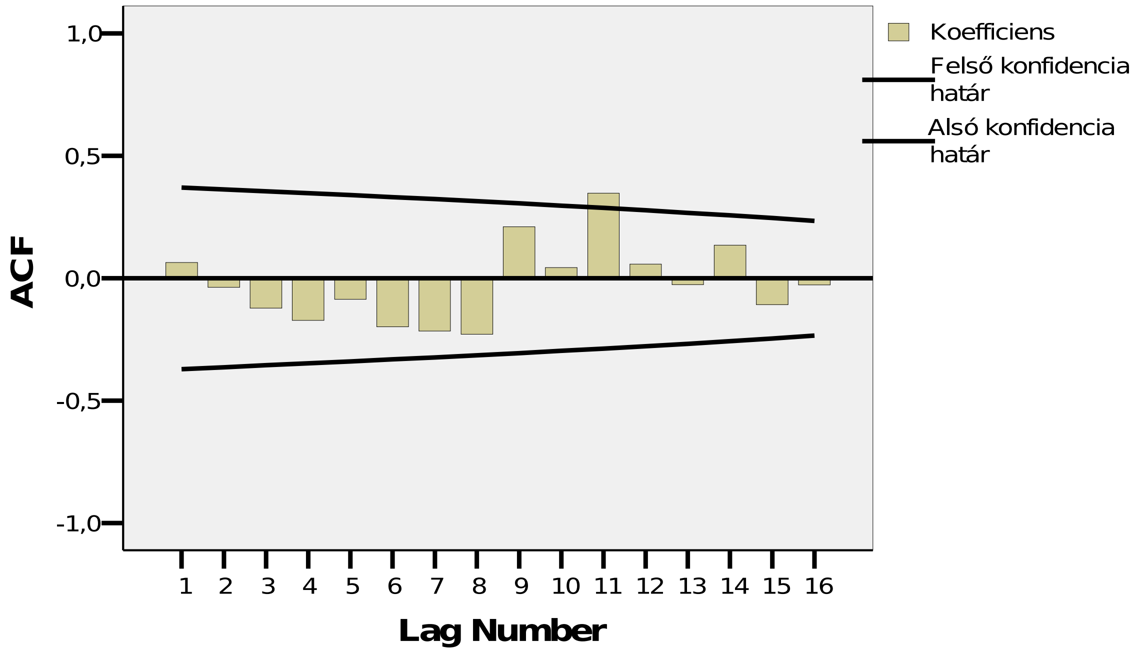
Smallest Sums of Squared Errors

Series	Model rank	Alpha (Level)	Sums of Squared Errors
db_sum	1	,70000	3907879011,54776
	2	,80000	3910177609,03472
	3	,60000	3965966440,05588
	4	,90000	3974185801,05280
	5	,50000	4081471836,26282
	6	1,00000	4103872124,85208
	7	,40000	4242310694,02872
	8	,30000	4418217451,93448
	9	,00000	4485032083,84616
	10	,20000	4560252585,14987



Az egyszerű modellel kapott becsült és tényleges forgalmi adatok

Az exponenciális simítás szépen leírja az adatokat, azonban van egy időbeli elcsúszás, ami több ezer darabos alá és felül becslést jelent. Ezért érdemes a maradékok autókorrelációs grafikonját elkészíteni, és megvizsgálni, hogy van-e valamilyen jól felismerhető szezonális hatás.



Az ábrán jól látható, hogy van egy 11 hónapos szignifikáns hatás, ami erős szezonális hatást jelent.

A becsült adatok rossz illeszkedése, és a 11 hónapos autókorrelációs együttható miatt az egyszerű exponenciális modell nem alkalmas a teakereslet pontos előrejelzésére.

A szezonális hatás felbontása

A szezonális analízis négy új változót hoz létre, amelyek az adatbázisban megjelennek, és az alábbi kezdőbetűkkel azonosíthatók:

SAF. Szezonális faktorok, a szezonális változásokat mutatják. A multiplikatív modellben az 1 érték a szezonális ingadozás hiányát mutatja. Az additív modellben ugyanezt a 0 érték jelenti. A szezonális faktorokat használhatjuk inputként az exponenciális simítás modelljeiben.

SAS. A szezonális hatástól megtisztított eredeti idősor. Ezzel a sorozattal trend-analízist, vagy más független szezonális összetevő kimutatását végezhetjük el. Trend meghatározása regresszió-analízis segítségével függő változóként lehet megadni. Autoregresszió számítása.

STC. Rövidebb trend-ciklus összetevők. Trend meghatározása regresszió-analízis segítségével függő változóként lehet megadni. Autoregresszió számítása.

ERR. Maradék tagok.

A szezonális hatások leválasztásával az egyszerű szezonális hatásokat távolíthatjuk el a ciklikus idősorokból.

KOVARIANCIA-ANALÍZIS

A kovariancia-analízis menete:

1. A vizsgált változó egytényezős variancia-analízise
2. t-tesztek és ezek próbája
3. A kovariáns és a vizsgált változó becsült átlagának kiszámítása
4. Lineáris egyenes illesztése regresszió-analízissel. A közelítés jóságának megítélése. (igen/nem).
5. A regresszió alapján a korrigált átlagok kiszámítása.
6. A korrigált átlagok egyenlőségének tesztje. (igen/nem).
7. A regressziós koefficiens egyenlő nullával? (igen/nem).
8. A csoportok regressziós koefficiensei egyenlők? (igen/nem).
9. Egy csoport és az összes többi csoport regressziós koefficiensének statisztikai próbája.
10. Ha a csoportok regressziós koefficiensei szignifikánsan különböznek egymástól, akkor ki kell számítani mindenegyes csoport regressziós koefficiensét, és ezt kell használni a továbbiakban.
11. A korrigált csoportátlagok t-tesztje, szignifikancia vizsgálata.
12. A megfigyelt és a korrekció után becsült értékek ábrázolása.

Grafikonok

Grafikon készítésekor nemcsak az alapadatokat ábrázolhatjuk, hanem a csoportképző változó szerint összesített, számított értékeket is. Pl. napi hőmérsékleti átlagok, minimumok, maximumok ábrázolása.

Oszlop diagramok (Bar Charts)

Egyszerű (Simple)

Csoportosított megfigyelések ábrázolása (Summaries for groups of cases):

A kategória tengelyen (Category Axis) a csoportképző változó szerepel, pl. a kezelés (öntözés, hibrid, trágyázás, stb.). Az oszlopok mutathatják a kezelésszintek megfigyeléseinek, eseteinek számát, kumulált értékeit és ezek százalékos részesedéseit. Egy függő változót kijelölve különböző statisztikai mutatókat ábrázolhatunk a csoportképző változó függvényében.

42. ábra: A kukorica termése (t/ha) különböző talajművelésekben

Különböző változók ábrázolása egy diagramon (Summaries of separate variables):

Csoportképző változó nélkül több változót, vagy ugyanannak a változónak a különböző statisztikai mutatóit ábrázolhatjuk a grafikonon.

43. ábra: A termés (t/ha) különböző statisztikai mutatói

A megfigyelt értékek ábrázolása (Values of individual cases):

A változó mindenegyes értékét ábrázolhatjuk. A megfigyelések száma nem lehet több, mint 3000. A megfigyeléseknek, eseteknek magyarázatokat, címkéket is adhatunk.

44. ábra: 2002. év napi csapadékadatai (mm)

Csoportosított (Clustered)

Csoportosított megfigyelések ábrázolása (Summaries for groups of cases):

Egy változó különböző statisztikai jellemzőinek ábrázolása két ismérv alapján. A kategória tengelyen a trágyázás, klaszterként az öntözés szerepel.

45. ábra: A trágyázás hatása a kukorica termésére nem öntözött és öntözött kezelésekben

Különböző változók ábrázolása egy diagramon (Summaries of separate variables):

Több változót, vagy ugyanannak a változónak a különböző statisztikai mutatóit ábrázolhatjuk a grafikonon. A csoportképző változó a kategória tengelyen jelenik meg.

46. ábra: Az elővetemény hatása a kukorica termésére és varianciájára

A megfigyelt értékek ábrázolása (Values of individual cases):

Több változó mindenegyes értékét ábrázolhatjuk. A megfigyelések száma nem lehet több, mint 3000. A megfigyeléseknek, eseteknek magyarázatokat, címkéket is adhatunk.

47. ábra: 2002. év január havi napi hőmérséklet és csapadékadatai

Halmozott (Stacked)

Egyetlen változó számított értékeinek ábrázolása vonaldiagram segítségével: (Graphs, Line Charts Simple)

… Summaries for groups of cases, Define. A változó kiválasztása után megadható az összesítés módja (statisztikája). A kategória tengelyen a csoportképző változót kell megadni.

… Summaries of separate variables

… Values of individual cases: a változó mindenegyes előfordulását ábrázolja, nem számít statisztikát.

Többszörös ábrázolás: (Graphs, Line Charts, Multiple)

… Summaries for groups of cases. Két csoportképző ismérv szerint ábrázolhatjuk a kiválasztott változókat, pl. kukorica terméseket a termőhely és idő függvényében. Az x-tengely (Category Axis) lehet az idő, pl. év, a vonalak (Define Lines by) pedig a termőhelyenkénti termések.

… Summaries of separate variables, Define. A Lines Represent ablakban az ábrázolandó változókat lehet megadni. A változókhoz különböző összesítési eljárások választhatók. Jelöljük ki egérrel a változót (kék szín) és Change Summary gombbal adjuk meg a számítási eljárást (átlag, medián, módusz, esetek száma, összeg, szórás, variancia, minimum, maximum, kumulatív összeg). Lehetőség van különböző százalékokban is megjeleníteni a változó értékeit.

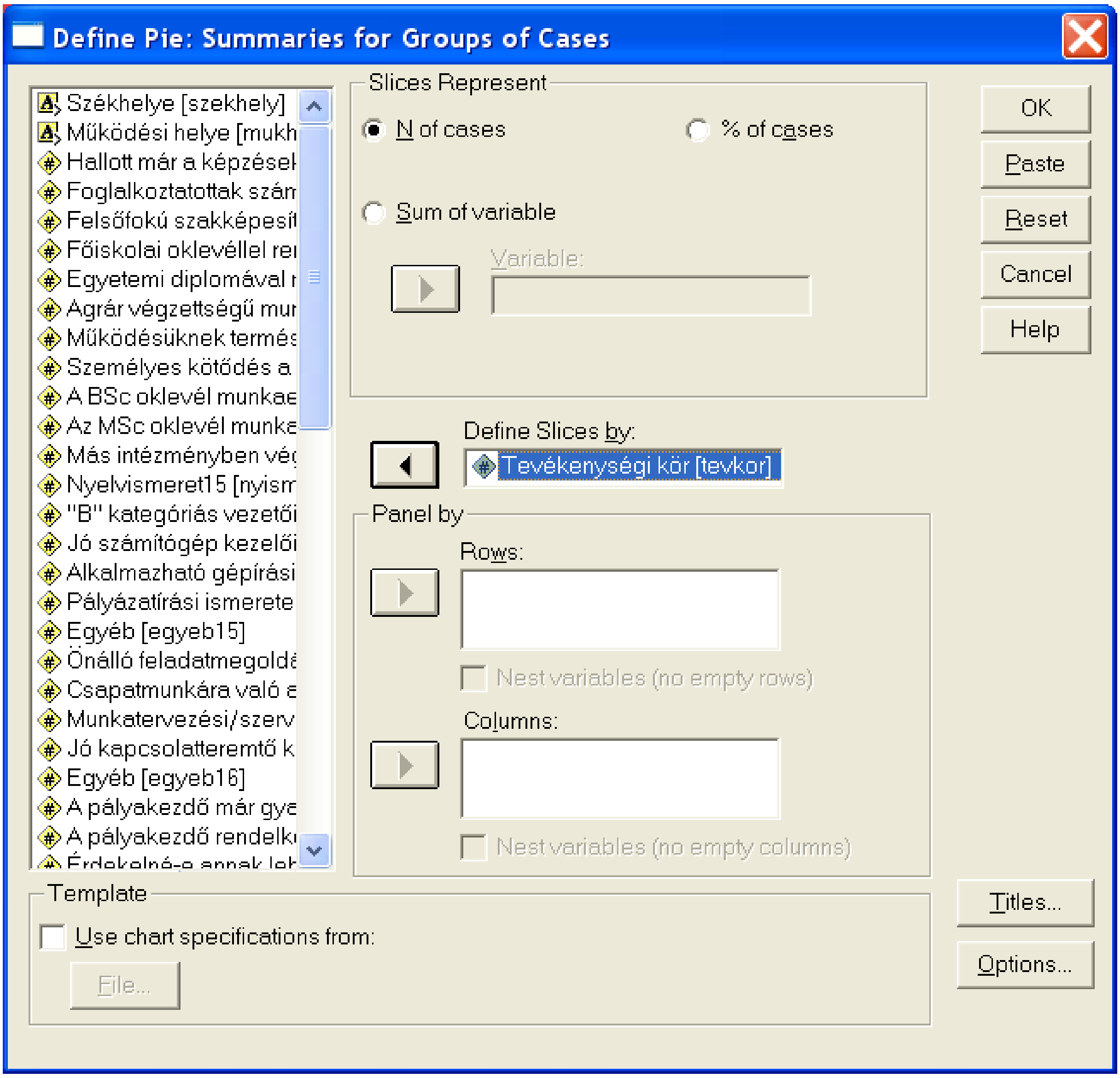
… Values of individual cases:

Minden grafikonnak azonos formátumot biztosíthatunk, ha a mintát (template) alkalmazunk. A minta *.sct fájlban található. Figyelem: bonyolult elérési útvonal esetén nem mindig találja meg a fájlt. Érdemes az SPSS alkönyvtárban tárolni ezeket a fájlokat.

Kördiagramok (Pie Charts)

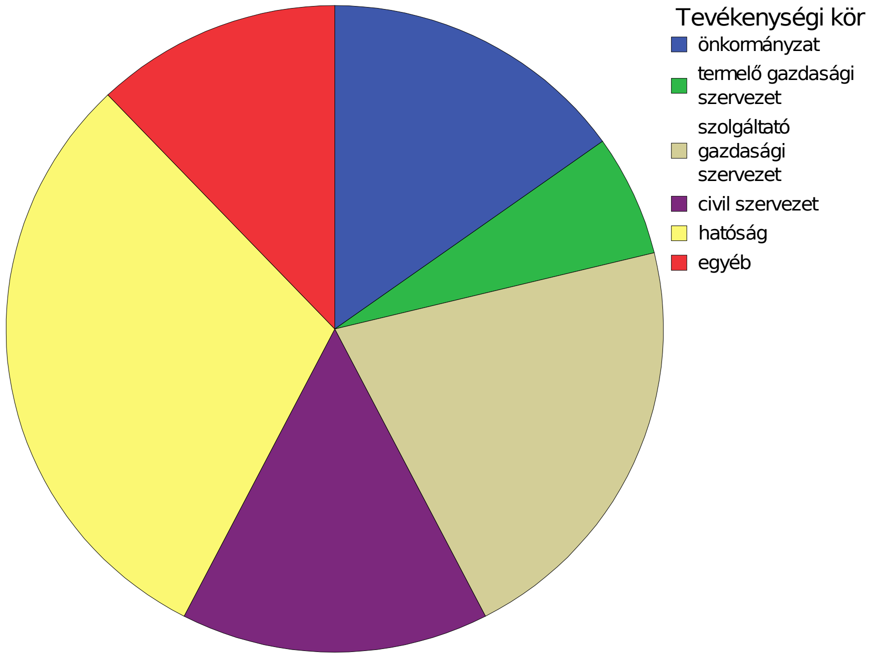
A megfigyelések csoportjainak ábrázolása (summaries for groups of cases)

Kördiagramot főként nominális változók gyakoriságának, vagy egy változó összetételének bemutatására használunk.

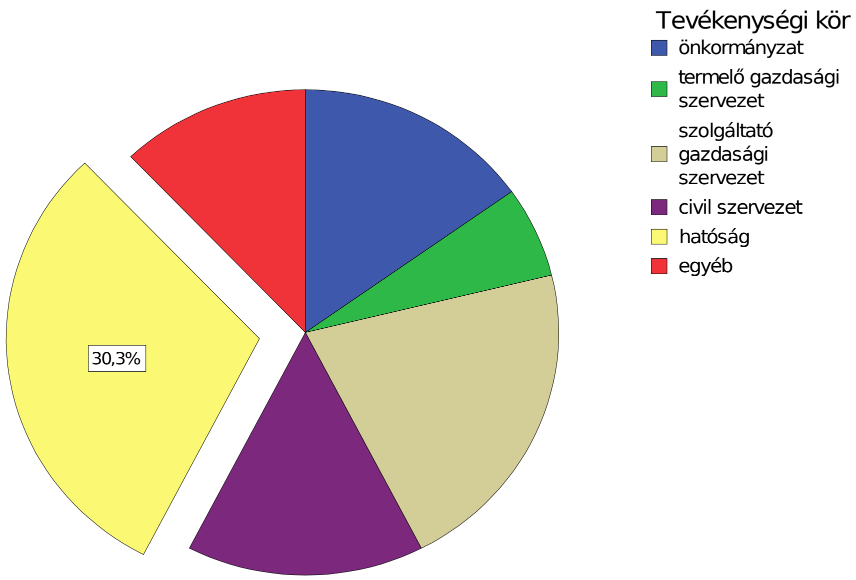


48. ábra: A kördiagram beállítása

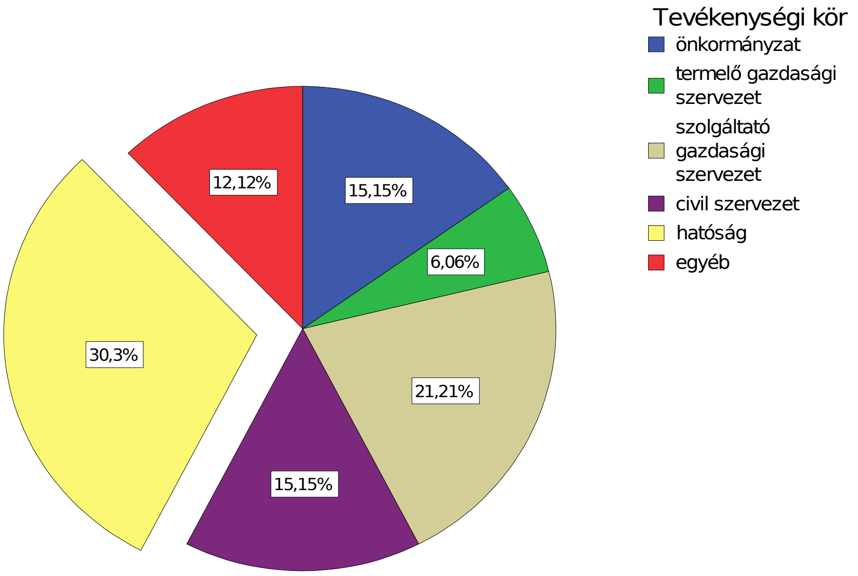
Készíthetünk egyszerű, úgynevezett robbantott ill. különböző információkkal kiegészített kördiagramokat.



49. ábra: Egyszerű kördiagram



50. ábra: Robbantott kördiagram, a leggyakoribb érték jelölésével



51. ábra: Robbantott kördiagram, a százalékok feltüntetésével

Kérdőívek tervezése

A tesztkészítéssel, tervezéssel és kiértékeléssel foglalkozó tudomány a tesztelmélet. Ebben a fejezetben nem akarok a teljességre törekedni, mert akkor több száz oldalt kellene igénybe venni. Csak az SPSS programhoz szükséges alapvető ismereteket tárgyalom, és megpróbálok némi gyakorlati útmutatót adni a helyes kérdőív kiértékeléshez.

A szakemberek nagyon sokféle kérdést különböztetnek meg, pl.:

- Igen/Nem kérdés
- Nyitott kérdés (1 soros válasz)
- Nyitott kérdés (több soros válasz)
- Nyitott kérdés (számjegyes válasz)
- Többszörös nyitott kérdés (számjegyes válasz).
- Egyszerű választás (egy válaszlehetőség)
- Többszörös választás (több válaszlehetőség)
- Mátrix-kérdés (soronként egy válaszlehetőség)
- Mátrix-kérdés (soronként több válaszlehetőség)
- Értékelő kérdés
- Többszörös értékelő kérdés
- Osztályozó kérdés 1-től 10-ig
- Többszörös osztályozó kérdés 1-től 10-ig A kérdőív létrehozása
- Időpontra vonatkozó kérdés
- Érték-relevanciakérdés

És még biztosan lehetne kitalálni még egy párat. A könnyű eligazodás érdekében le fogom egyszerűsíteni a kérdések csoportosítását, főként az adatbázis tulajdonságaik alapján, mivel a különböző típusú kérdéseket különbözőképpen kell beépíteni az adatbázisba. Vannak olyan kérdések, amikre csak egyetlen választ lehet adni, pl. a lenti kérdés (melyik korosztályba tartozik), és vannak többszörös válaszadásúak is. A válaszadó csak egyetlen korcsoportba tartozhat. Az ilyen típusú válaszokat rádiógombokkal szokták jelezni, ezzel is sugalmazva, hogy csak egyetlen választ vár a kérdést feltevő személy. Az adatbázisban ezt egyetlen nominális vagy ordinális típusú változóban tárolhatjuk. Érdemes számokkal kódolni az egyes korosztályokat, és címkéket használni a megnevezésükhöz.

Melyik korosztályhoz tartozik?

1. 0-20 év
2. 21-40 év
3. 41-60 év
4. 61-80 év
5. 80 felett

Az alábbi kérdésre (hogy van megelégedve a munkahelyével) is csak egyetlen válasz adható. Ez egy minősítő, eldöntendő kérdés. Az adatbázisban ezt is egyetlen változóban tároljuk, ordinális adatként. Értéke lehet szöveg vagy szám. Célszerű számokat megadni, és címkéket használni, mivel így sokkal kisebb méretű adatbázist kapunk.

„Hogy van megelégedve a munkahelyével?”

1. Nagyon
2. Közepesen
3. Kevésbé

Az előző két kérdéstípusra adott válaszokat tehát egyetlen változóban kell tárolni. Az olyan kérdéseket, ahol több válasz is lehetséges, kicsit bonyolultabb az adatbázisba elhelyezni. Ilyen típusú kérdés az alábbi:

„Van-e a lakásban?”

1. Vezetékes víz
2. Központi fűtés
3. Telefon
4. Színes televízió
5. Számítógép

A válaszoló akár mindet megjelölheti vagy egyiket sem. Ilyenkor a lehetséges válaszok mindegyikére egy-egy dichotóm (kétértékű, 0=nincs, 1=van) változót képezünk. Ezek a változók egy csoportot alkotnak, érdemes a változók nevével is jelezni a csoportba tartozást. Pl. ha a fenti kérdés a nyolcadik, akkor a válaszokat K8_1, K8_2, K8_3 stb. jelölhetjük. A többszörös válaszadások elemzése bonyolultabb, mint az egyszerű választó kérdések kiértékelése.

Milyen sorrendben tartalmazzák a kérdéseket a kérdőívek?

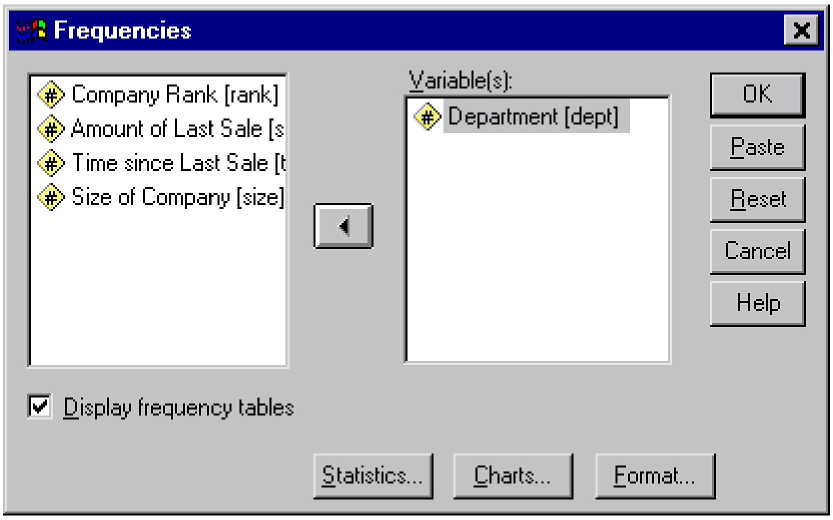
Ez elég szubjektív, függ a kérdőív típusától. Általában a közvélemény kutatásban használt kérdőívekben az elején egyszerű, figyelemfelkeltő kérdések vannak. A bizalmas jellegű, személyes kérdések a kérdőív végére kerülnek. Bizalmatlanságot és ellenkezést válthat ki az olyan kérdőív, aminek az első kérdése azt firtatja, hogy hány évesek vagyunk és mennyit keresünk, kövérek vagyunk-e vagy soványak. Egy jó kérdőív betartja az íratlan udvarlási szabályokat. Előbb az érdeklődést, szimpátiát, bizalmat kell megszerezni, és csak utána jöhetnek az esetleges bizalmas kérdések.

KÉRDŐÍVEK KIÉRTÉKELÉSE

Először az egyetlen választ adó kérdések értékelését mutatom be. Ezek a válaszok lehetnek nominális, ordinális és skála típusú adatok.

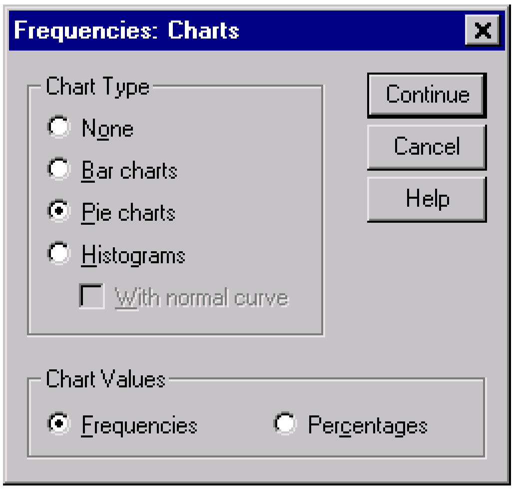
Nominális típusú adatok kiértékelése

Az elemzést az SPSS-hez mellékelt *contact.sav* adatbázison mutatom be. Nyissuk meg az adatbázist és válasszuk Analyze, Descriptive Statistics, Frequencies menü pontot. Ezután a Department változót tegyük a változók ablakba.



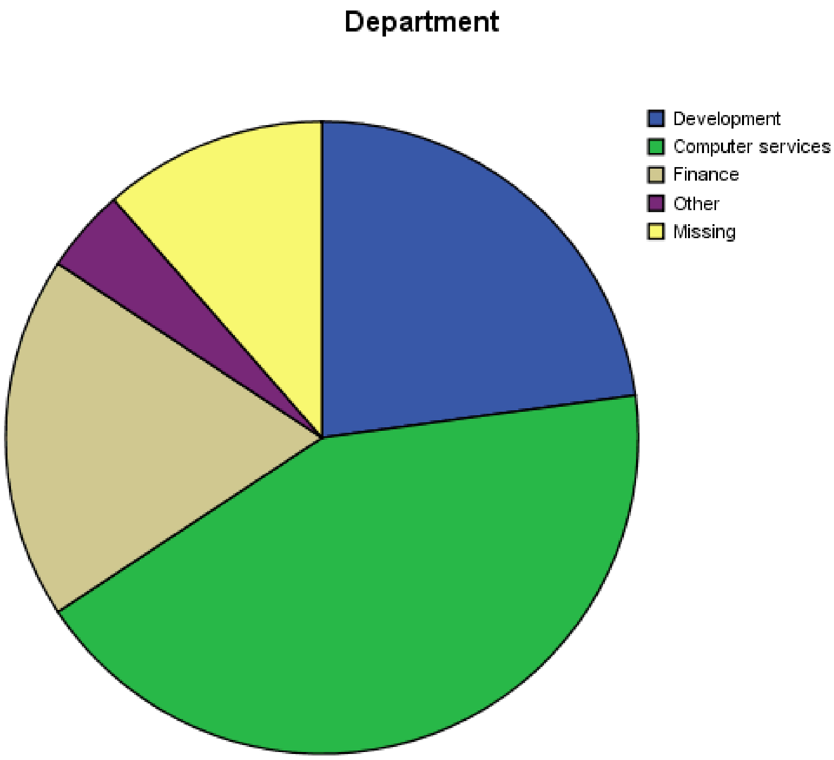
52. ábra: Gyakoriságok párbeszédpanel

Kattintsunk a Charts… gombra (grafikonok), és a grafikon típusának (Chart Type) adjuk meg a kördiagramot (Pie Charts). A diagram ábrázolhatja az adatok gyakoriságát (az előfordulás számát) vagy százalékát. Ezt a Chart Values területen tudjuk beállítani.



53. ábra: Grafikonok kiválasztása és beállítása

A folytatáshoz kattintsunk a Continue gombra.



54. ábra: Kördiagram, a kategóriák feltüntetésével

A kördiagram szemléletesen ábrázolja a különböző kategóriák relatív gyakoriságát a megfigyelések egészéhez viszonyítva. A gyakorisági táblázat pontosan megmutatja az egyes kategóriák gyakoriságát (Frequency), százalékban kifejezve (Percent), az érvényes megfigyelések százalékában kifejezve (Valid Percent) és a kumulatív eloszlást százalékban (Cumulative Percent). Az első oszlopban a Valid az érvényes megfigyeléseket, a Missing a hiányzó értékeket jelöli. Az adatbázisban összesen 70 megfigyelés szerepel. Ebben 62 érvényes és 8 hiányzó adat van. Általános érvényű következtetés levonásához mindig az érvényes adatokat kell figyelembe venni. Nominális típusú adatoknál a kumulatív eloszlás nem ad többlet információt, gyakorlatilag nem is lehet használni semmire. Ez inkább az ordinális típusú adatok esetén hasznos.

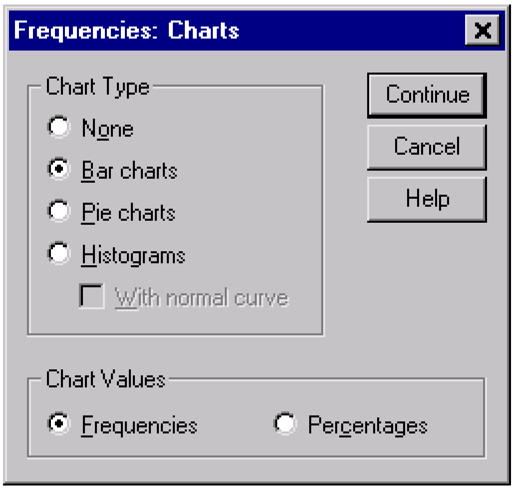
Department

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Development	16	22.9	25.8	25.8
	Computer services	30	42.9	48.4	74.2
	Finance	13	18.6	21.0	95.2
	Other	3	4.3	4.8	100.0
	Total	62	88.6	100.0	
Missing	Don't know	8	11.4		
Total		70	100.0		

A leggyakoribb adat a Computer Services, az érvényes adatok százalékában 48,4%.

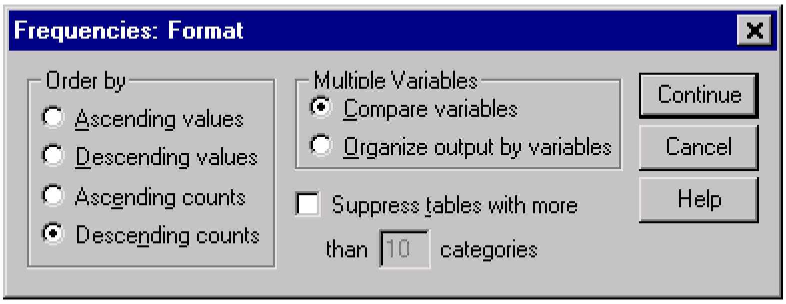
A kördiagram helyett oszlopdiagramot választva, amiben a kategóriákat csökkenő gyakorisággal ábrázoljuk, gyorsan megállapíthatjuk a sokaság móduszát (leggyakrabban előforduló kategória), illetve a relatív gyakoriságot szemléletesen ábrázolhatjuk.

Válasszuk újból a grafikonok menü pontot, és most az oszlop diagramot aktivizáljuk.



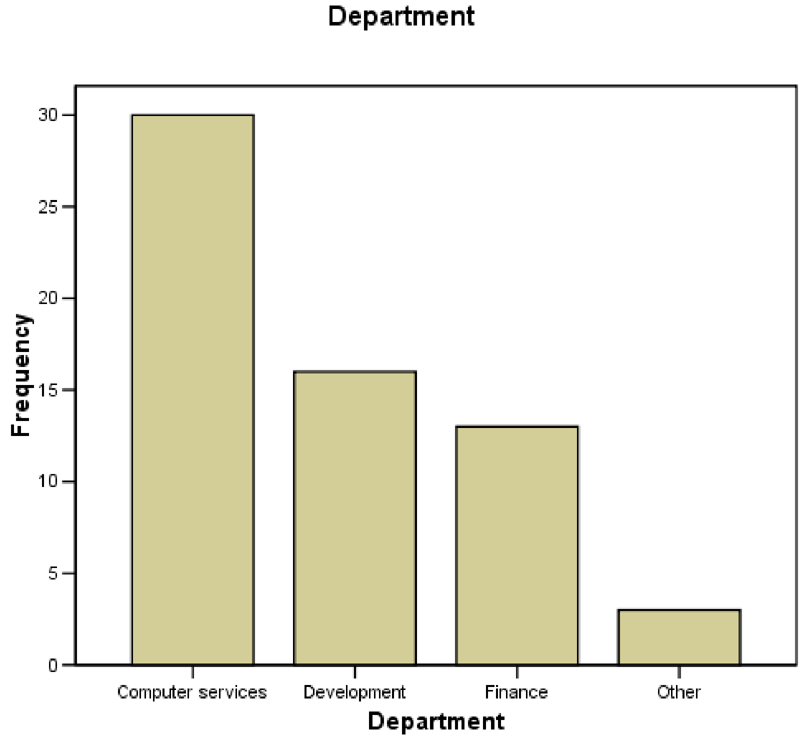
55. ábra: Oszlopdiagram kiválasztása

A folytatás után kattintsunk a Format gombra, ahol a frekvenciák táblázatos megjelenítésének és ábrázolásának módját lehet beállítani. A gyakoriságok megjelenítését végezzük a kategória-gyakoriságok csökkenő sorrendjében (Descending Counts). A Continue után megkapjuk az oszlopdiagramot.



56. ábra: A gyakoriságok megjelenítésének beállítása

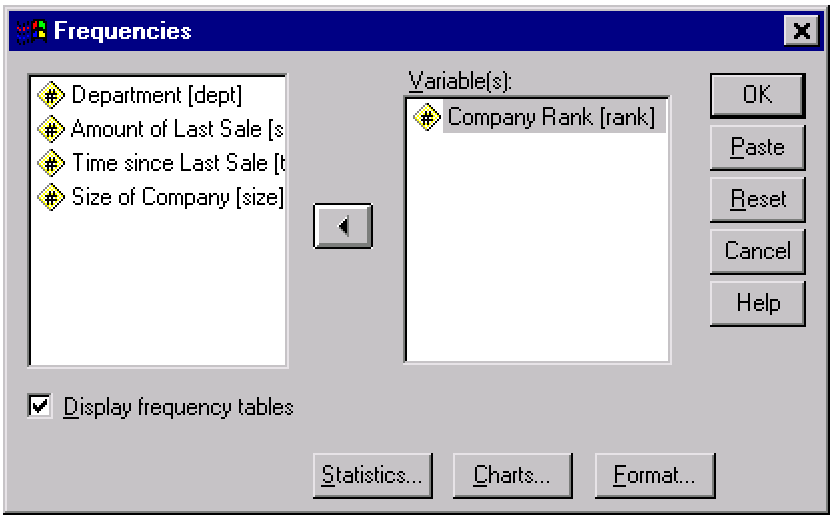
A gyakoriságok az előfordulások nagyságának csökkenő sorrendjében jelennek meg.



57. ábra: Oszlopdiagram, csökkenő gyakorisági sorrend

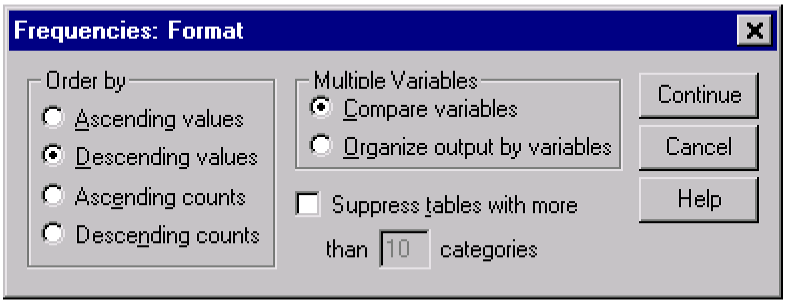
Ordinális típusú adatok kiértékelése

Olyan adatok kiértékelése, amik valamilyen szempont alapján sorba rendezhetők, valamilyen rangsor felállítható közöttük. Ugyanazt az eljárás fogjuk használni, mint az előbb, csak a beállítások lesznek mások. Az adatbázis alapján elemezzük a cég rangsor változót. Tegyük be a Company Rank változót a vizsgálati ablakba. Ez egy ordinális típusú változó, amit a változó definiálásakor nekünk kellett beállítani az SPSS adatbázis ablakában.



58. ábra: Gyakoriságok elemzése

Készítsünk oszlopdiagramot.. A Format… beállításai az alábbiak lesznek: csökkenő rendezés a változó értékei szerint. Tehát nem az előfordulás gyakorisága szerint, hanem a rangsorban elfoglalt értéke alapján fognak megjelenni a frekvenciák. A rangsorban legértékesebb kategória gyakorisága fog az első helyen (balra) megjelenni, és utána a többi.



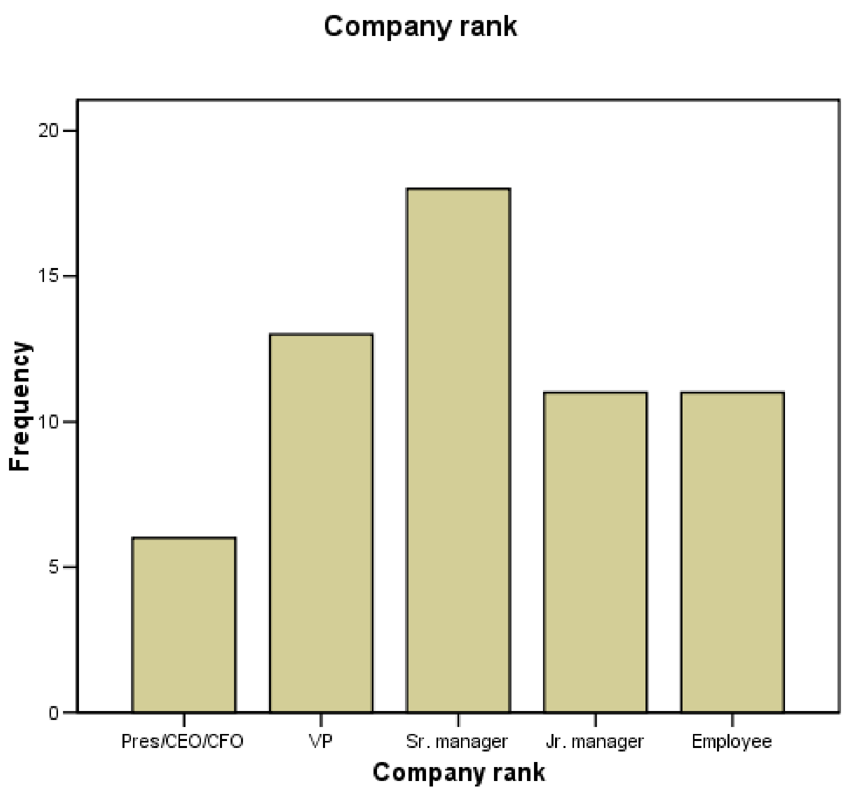
59. ábra: Gyakoriságok megjelenítése csökkenő értékek szerint

Ez alapján fog elkészülni a frekvenciatáblázat és az oszlopdiagram. A gyakorisági táblázat fentről lefelé csökkenő rangszámok alapján mutatja a frekvenciákat. Képzeljük el, hogy iskolai osztályzatokat értékeltünk. Ekkor a jeles áll az első helyen és az elégtelen az utolsón. Jeles – csak az érvényes megfigyeléseket figyelembe véve – 10,2%, jó 22, közepes 30,5%, elégséges 18,6% és végül elégtelen szintén 18,6%. A kumulatív eloszlásból (Cumulative Percent) egyéb értékes megállapítások is tehetők. Pl. a legalább közepes eredményt elért hallgatók aránya 62,7%. Sikeres vizsgák aránya 81,4% és így tovább.

Company rank

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Pres/CEO/CFO	6	8.6	10.2	10.2
	VP	13	18.6	22.0	32.2
	Sr. manager	18	25.7	30.5	62.7
	Jr. manager	11	15.7	18.6	81.4
	Employee	11	15.7	18.6	100.0
	Total	59	84.3	100.0	
Missing	Don't know	11	15.7		
Total		70	100.0		

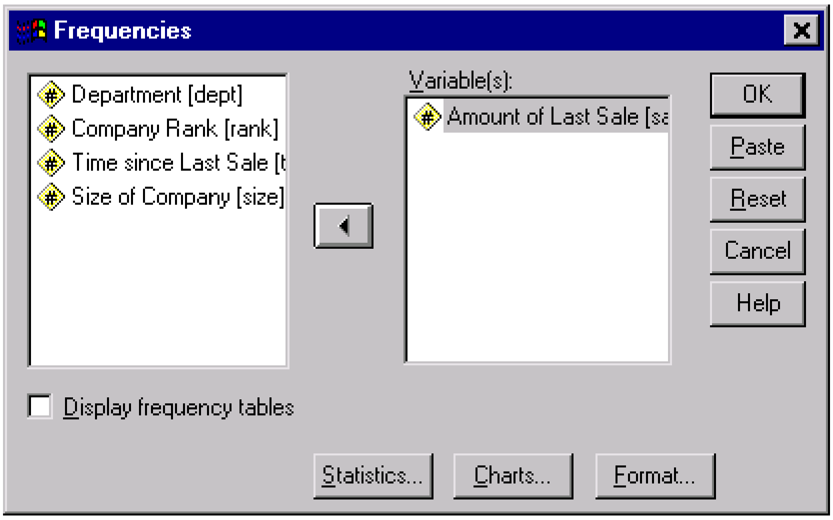
Az oszlop diagram a gyakoriságokat mutatja az „osztályzatok" csökkenő sorrendjében. Leggyakoribb osztályzat a közepes, jelesből van a legkevesebb, stb.



60. ábra: Oszlopdiagram, gyakoriságok csökkenő érték kategória szerint rendezve

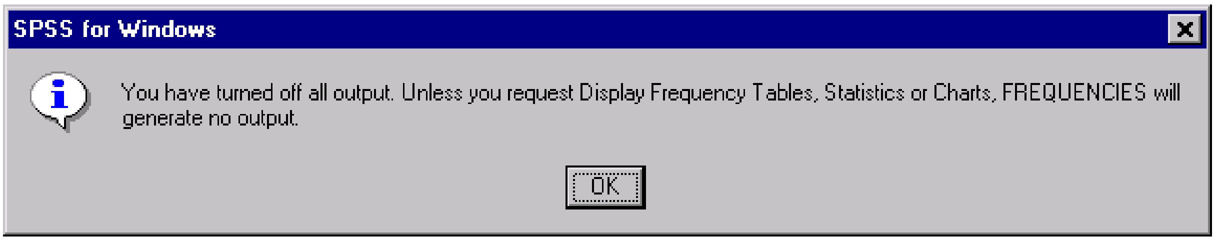
Skála típusú adatok kiértékelése

Itt is ugyanazt a programot fogjuk használni, mint az előbb. A skála típusú adat valamilyen fizikai mennyiséget jelöl, legtöbbször mértékegységgel is rendelkezik. Az előbbi adatbázis segítségével a legutóbbi értékesítési eredményeket fogjuk elemezni. Kattintsunk a Reset gombra, hogy alaphelyzetbe hozzuk a párbeszédablakot. Ezután válasszuk ki a Amount of Last Sale változót és tegyük be a változók ablakba. Ne készítsünk frekvencia táblázatot, mivel skála típusú adatnál nagyon sok „kategória" van, ezért a Display Frequency Tables jelölőnégyzetet töröljük.



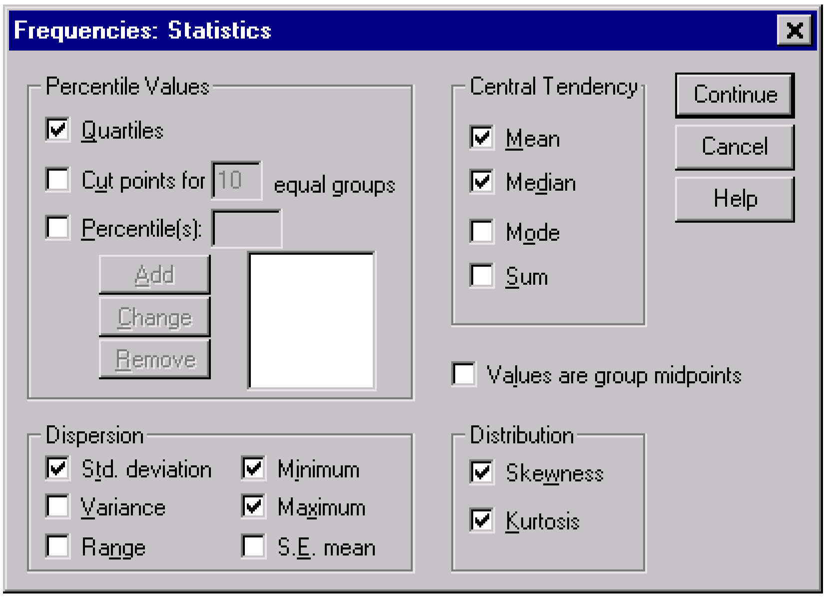
61. ábra: Skála típusú adat gyakoriságának elemzése

Valószínűleg kapunk egy hibaüzenetet, ami arra figyelmeztet, hogy minden outputot letiltottunk. Ez nem baj, majd a későbbiekben beállítjuk amire szükségünk van. Klikkeljünk az OK-ra.



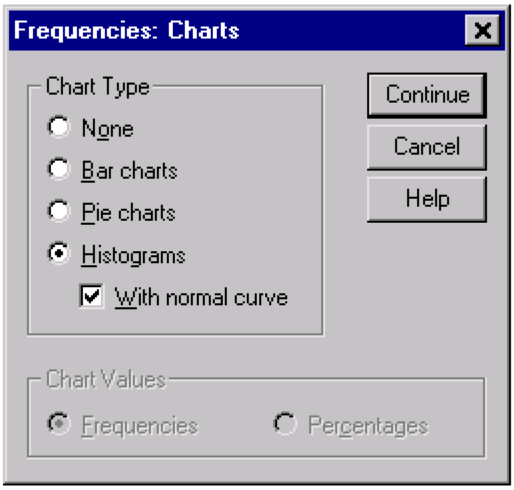
62. ábra: Az SPSS figyelmeztető üzenete

Kattintsunk a Statistics gombra a Frequencies párbeszédablakban. Válasszuk ki a kvartiliseket, a szórást, a minimumot és maximumot, az átlagot, a mediánt, a ferdeséget és végül az eloszlás csúcsosságát.



63. ábra: A különböző statisztikák beállítása

Klikk Continue. Válasszuk ki a grafikonok menüből a hisztogramot és jelöljük be a normáleloszlás jelölőnégyzetét. Continue.



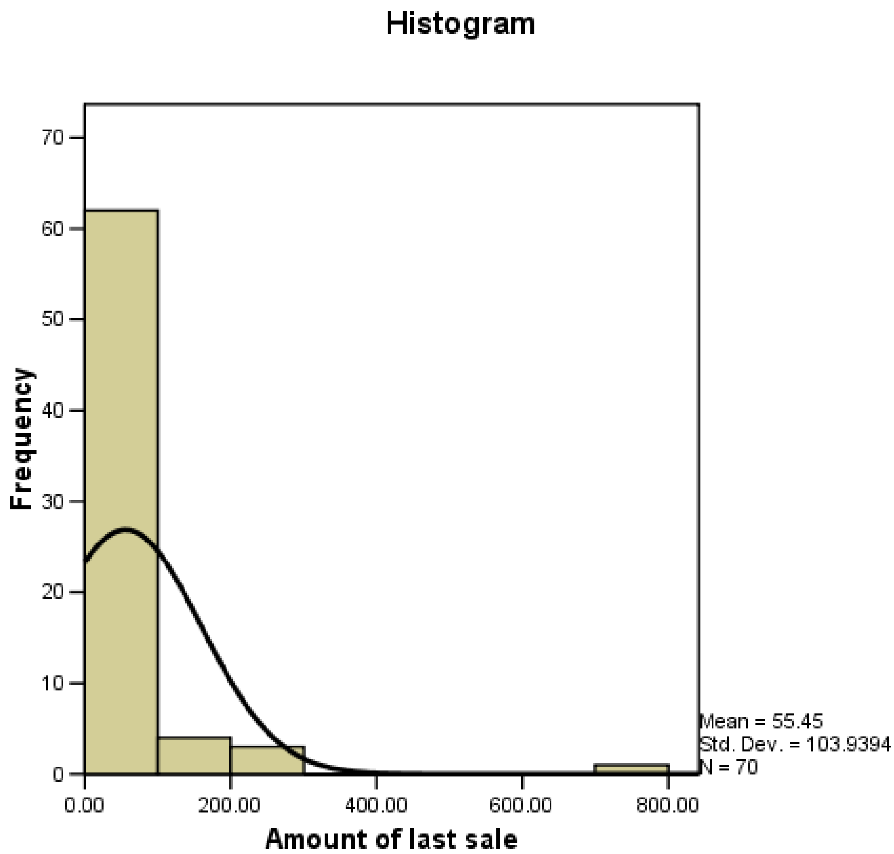
64. ábra: Hisztogram beállítása, a normáleloszlás görbéjének kiválasztása

Az elészült statisztikák az eladások jellemző értékeit mutatják. A megfigyelések száma 70, nincs hiányzó adat. Az átlag 55,45, a medián 24, a szórás 103,94. Az eloszlás ferdesége 5,33, csúcsossága 34,29. A legkisebb értéke 6, a legnagyobb 776,5. Az első kvartilis éréke 12, a másodiké 24, a harmadiké 52,88. Az eladások eloszlásának közepét a medián, illetve a második kvartilis mutatja, értéke 24. Az eloszlás középpontja körül az adatok fele 12 és 52,88 közé esik. Ezt az első és harmadik kvartilis mutatja, aminek a különbségét interkvartilisnek nevezik.

Amount of last sale

N	Valid	70
	Missing	0
Mean		55.4500
Median		24.0000
Std. Deviation		103.93940
Skewness		5.325
Kurtosis		34.292
Minimum		6.00
Maximum		776.50
Percentiles	25	12.0000
	50	24.0000
	75	52.8750

Az eladások két extrém értéke a minimum és a maximum. Az átlag és a medián nagyon különbözik, ami azt sugalmazza, hogy az eloszlás erősen aszimmetrikus. Ezt erősíti meg a ferdeségi mutató nagy pozitív értéke (5,33), ami azt mutatja, hogy az eladások eloszlásának hosszú jobboldali farka van, azaz balra ferde az eloszlás. Kis gyakorisággal nagyon nagy eladások is előfordulnak, szinte nincs felső határa az eladás nagyságának. Az alsó határa azonban csak nulla lehet, negatív eladás nem létezik. Az eladások ezen tulajdonsága okozza a balra ferde eloszlást, vagyis a pozitív ferdeséget. A nagy pozitív ferdeség jól látható a medián és átlag elhelyezkedésén is, az átlag jobbra található a mediántól. A szórás nagyon nagy 103,94. A nagy pozitív csúcsosság (34,29), a normáleloszlásnál csúcsosabb eloszlást jelez.



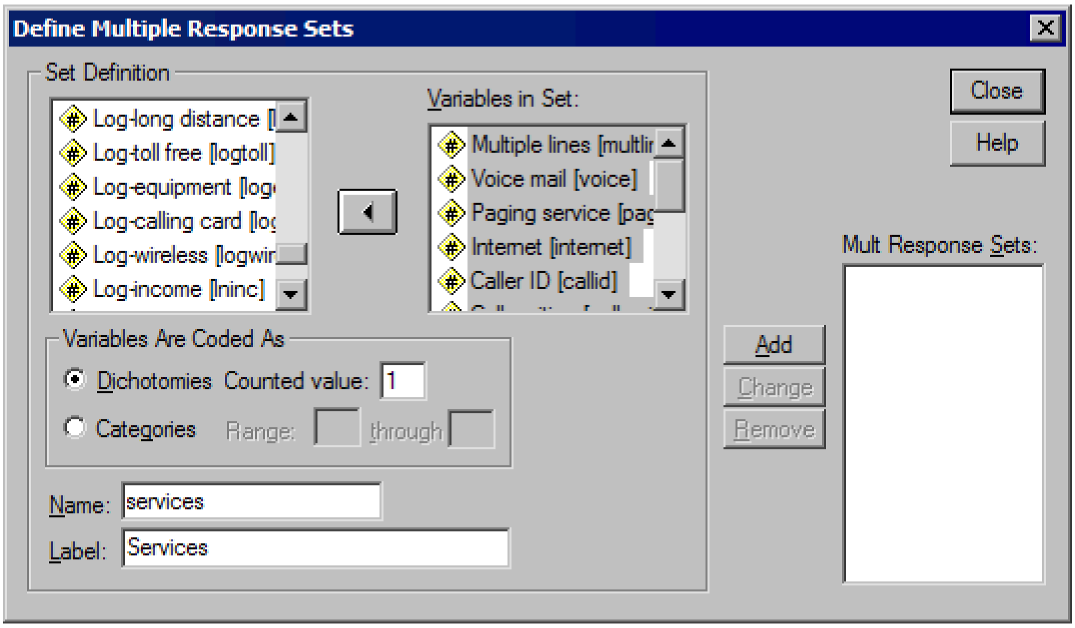
65. ábra: Az eladások hisztogramja

A hisztogram vizuálisan mutatja az eladások eloszlását. A normáleloszlás vonala segít a tényleges eloszlás tulajdonságainak megítéléséhez.

Többszörös válaszadások elemzése 1.

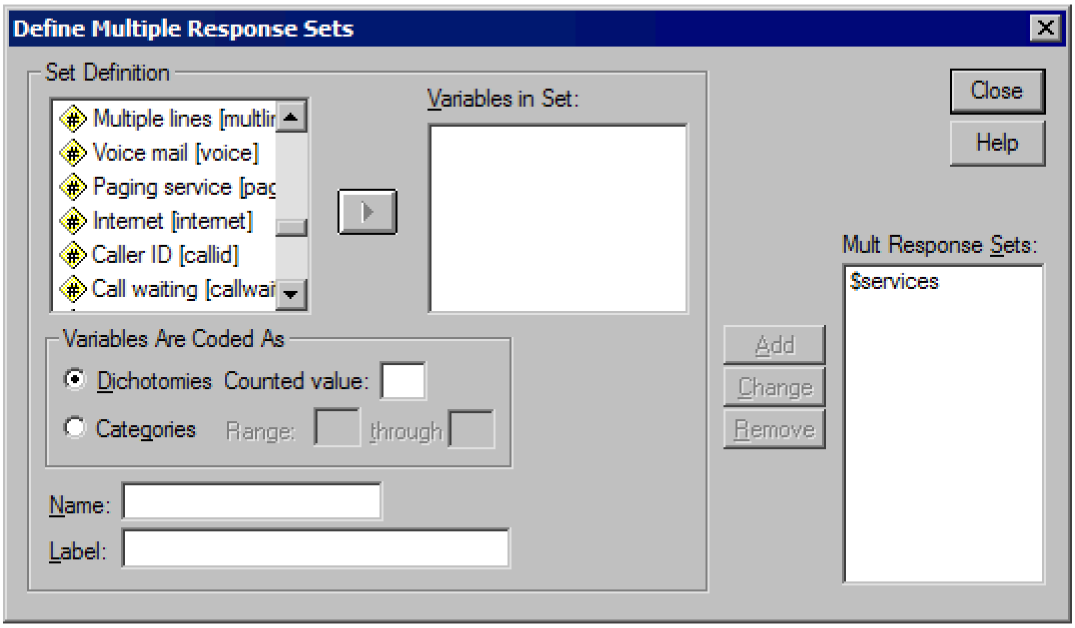
A többszörös válaszadások elemzése gyakorisági-, és kereszttáblázatok létrehozását jelenti az általunk előre definiált többszörös válaszadások csoportjai, szettjei alapján. A többszörös válaszadás szettje logikailag összetartozó változók együtteséből áll. Ezek a változók legtöbbször dichotóm, kétértékű vagy kategória változók. A többszörös dichotóm szett a gyakorlatban sokszor igen/nem (1/0 vagy true/false) típusú válaszok csoportját jelenti. Pl. milyen eszközökkel rendelkezik a válaszadó az alább felsoroltak közül? Többszörös kategória szettet akkor készítünk, amikor maximalizáljuk a válaszok számát. Ebben az esetben a megkérdezettek maximális válaszainak száma jelentősen kevesebb, mint a lehetséges válaszok száma.

A többszörös válaszadás szettjének létrehozásához válasszuk az Analyze, Multiple Response, Define Sets... parancsot. A példaadatbázist az SPSS-hez kapjuk. Ebben egy telekommunikációs kérdőív adatai szerepelnek, hogy milyen szolgáltatásokat vesznek igénybe a megkérdezettek. Az adatbázis változói közül válasszuk ki a logikailag összetartozókat, és tegyük bele Variables in Set ablakba. Összesen maximum 20 ilyen szettet tudunk megadni. Mindegyiknek egyedi névvel kell rendelkeznie.



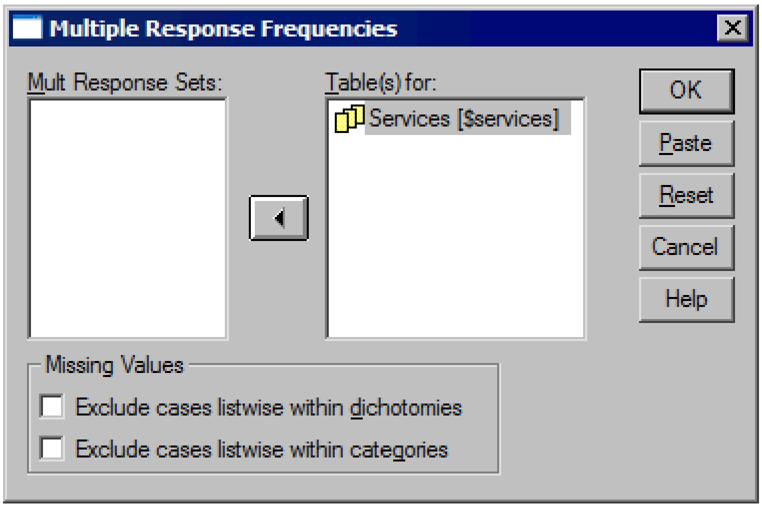
66. ábra: Többszörös válaszadások szettjének megadása

A változók most kétértékűek és az összeszámlálandó értéket az 1 jelenti. A szett neve ’servises’ és a címkéje ’Services’. Klikk Add gomb és Close.



67. ábra: Többszörös válaszadások neve és címkéje

A gyakorisági táblázat elkészítéséhez válasszuk az Analyze, Multiple Response, Frequencies... parancsot. A Services változót tegyük a Table(s) for: ablakba.



68. ábra: Gyakoriság elemzése többszörös választású változóval

Klikk OK. Az adatbázisban összesen 1 000 megfigyelés szerepel 111 hiányzó adattal. A hiányzó adat ebben a példában azokat a személyeket jelenti, akik egyetlen szolgáltatásra sem fizetnek elő. Érvényes válasznak tekintjük azt, ha legalább egy szolgáltatást igényben vesz az illető.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$services[a]	889	88.9%	111	11.1%	1000	100.0%

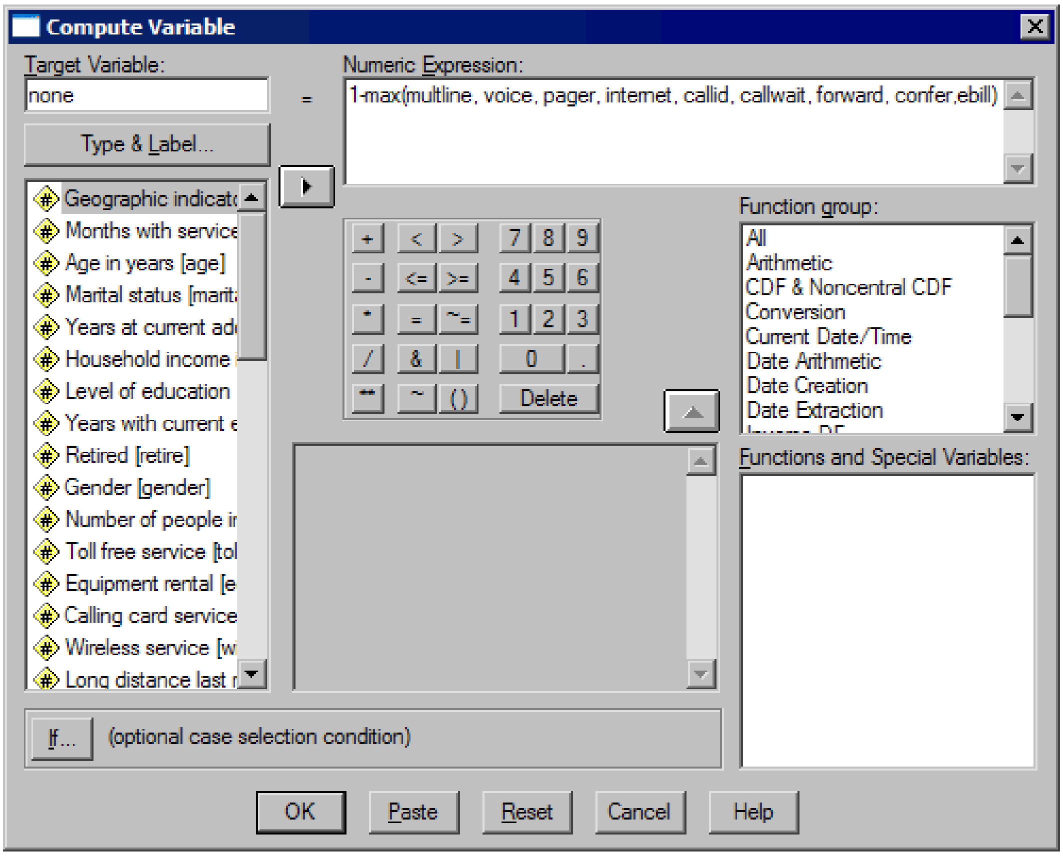
a. Dichotomy group tabulated at value 1.

A gyakorisági táblázat a megkérdezettek válaszait tartalmazza. Jól látható, hogy milyen szolgáltatást hányan vesznek igénybe. Pl. az 1 000 megkérdezett közül 368 használja az Internetet. Természetesen az összes válaszok száma meghaladhatja az 1 000, mivel egy válaszadó több szolgáltatást is használhat egyszerre. Ez a többszörös válaszadás lényege. Az egyszerű gyakorisági értékek mellett egyéb fontos információ is kiolvasható a táblázatból. Az N jelenti, hogy hányan használják vagy fizetnek elő a szolgáltatásra, a Total N az összes előfizetői szerződés számát mutatja. A százalék (Percent) oszlop az összes válaszok százalékában adja meg az igénybe vett szolgáltatások nagyságát. Ezt többféleképpen is lehet értelmezni. Pl. a válaszadók mindennapi tevékenységük során ilyen arányban használják a különböző szolgáltatásokat vagy a különböző szolgáltatások piaci részesedése. Ilyen kimutatást egy egyszerű gyakorisági táblázattal nem lehet készíteni csak többszörös gyakorisági táblázattal.

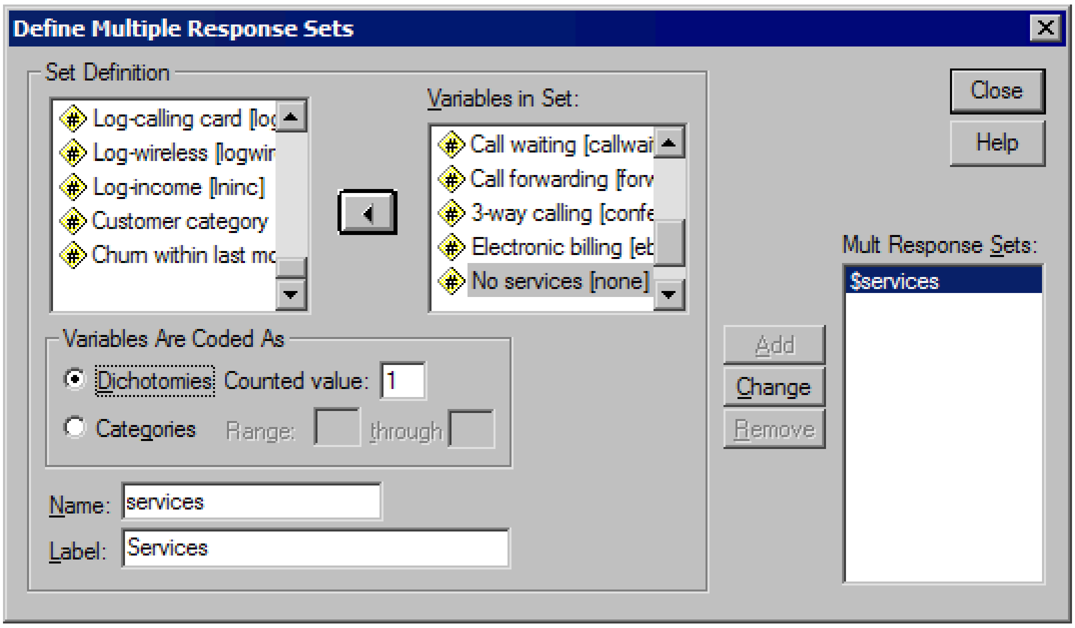
		Responses		Percent of Cases
		N	Percent	
Services[a]	Multiple lines	475	12.7%	53.4%
	Voice mail	304	8.1%	34.2%
	Paging service	261	7.0%	29.4%
	Internet	368	9.8%	41.4%
	Caller ID	481	12.9%	54.1%
	Call waiting	485	13.0%	54.6%
	Call forwarding	493	13.2%	55.5%
	3-way calling	502	13.4%	56.5%
	Electronic billing	371	9.9%	41.7%
Total		3740	100.0%	420.7%

a. Dichotomy group tabulated at value 1.

A Percen of Cases oszlopban az érvényes válaszok százalékában látjuk az adott szolgáltatás igénybevételi arányát. A 889 válaszoló 41,4%-a használja az Internetet. Ezek a számok azonban nem mutatják azokat a felhasználókat, akik egyetlen szolgáltatást sem választanak (ezek a hiányzó adatok). Amennyiben ezekre is kíváncsiak vagyunk elő kell állítani egy új változót, amit a Transform, Compute paranccsal végezhetünk el (69. ábra). Legyen a változó neve „none”. A függvény: 1-max(a szolgáltatások listája vesszővel elválasztva). Ennek az értéke 1, ha egyelten szolgáltatást sem vesz igénybe az illető. A változó címkéje legyen „No services”. Klikkelj a Continue gombra. Ezek után adjuk hozzá az új változót a már meglévő többszörös dichotóm szetthez (70. ábra).

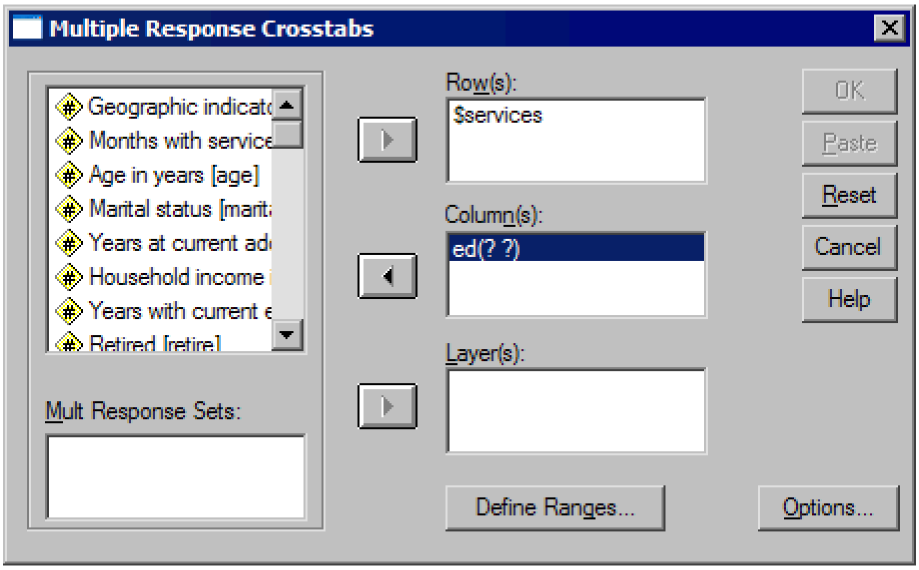


69. ábra: A szolgáltatást igénybe nem vevők számítása



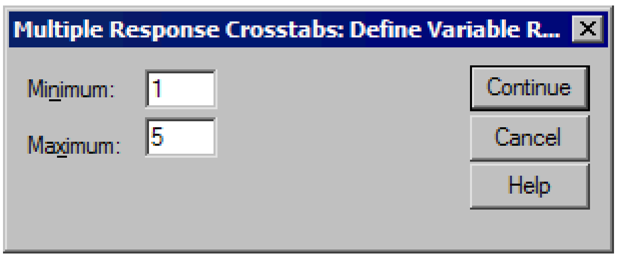
70. ábra: A többszörös választás bővítése

Van-e összefüggés az iskolai végzettség és a szolgáltatások kedveltsége között? Ehhez készítsünk kereszttáblázatot az igénybe vett telekommunikációs szolgáltatások gyakorisága és az azt igénybevevő ügyfelek iskolai végzettség változók felhasználásával. Analyze, Multiple Response, Crosstabs…



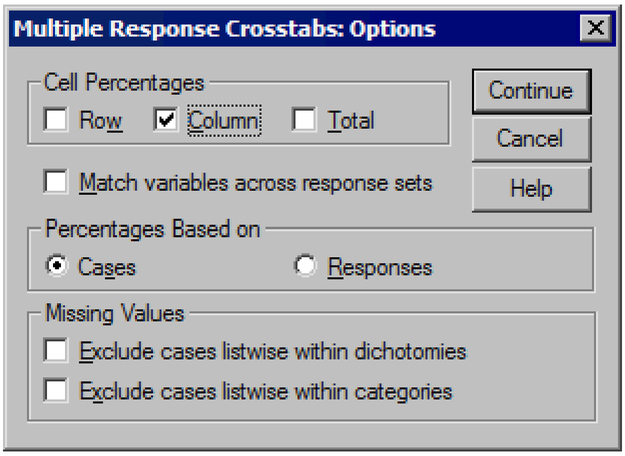
71. ábra: Kereszttáblázat készítése többszörös változóval

A sorba (Row) tegyük bele a 'services' változót, az oszlopba (Column) az 'ed' (level of education) változót. Az iskolai végzettség változónak meg kell adni az elemzésbe vont tartományát (Define Ranges). Jelen esetben 1-5 kategóriák vesznek részt az elemzésben.



72. ábra: Iskolai végzettség kategóriájának megadása

Ezek után állítsuk be, hogy a cellákban a százalékok oszloponként számítódjanak, vagyis iskolai végzettség szerint (73. ábra).



73. ábra: Cellák százalékszámítása oszlop szerint

A számítások elvégzése után az alábbi eredménytáblázatot kapjuk.

			Level of education					
			Did not complete hig	High school degree	Some college	College degree	Post-under graduate d	Total
Services	Multiple lines	Count	57	121	107	139	51	475
		% within ed	27.9%	42.2%	51.2%	59.4%	77.3%	
	Voice mail	Count	21	70	68	104	41	304
		% within ed	10.3%	24.4%	32.5%	44.4%	62.1%	
	Paging service	Count	14	61	53	101	32	261
		% within ed	6.9%	21.3%	25.4%	43.2%	48.5%	
	Internet	Count	14	71	86	145	52	368
		% within ed	6.9%	24.7%	41.1%	62.0%	78.8%	
	Caller ID	Count	87	142	103	121	28	481
		% within ed	42.6%	49.5%	49.3%	51.7%	42.4%	
	Call waiting	Count	92	145	101	116	31	485
		% within ed	45.1%	50.5%	48.3%	49.6%	47.0%	
	Call forwarding	Count	94	141	106	119	33	493
		% within ed	46.1%	49.1%	50.7%	50.9%	50.0%	
	3-way calling	Count	98	145	106	120	33	502
		% within ed	48.0%	50.5%	50.7%	51.3%	50.0%	
	Electronic billing	Count	14	84	88	141	44	371
		% within ed	6.9%	29.3%	42.1%	60.3%	66.7%	
	No services	Count	48	36	17	7	3	111
		% within ed	23.5%	12.5%	8.1%	3.0%	4.5%	
Total		Count	204	287	209	234	66	1000

Percentages and totals are based on respondents.

[a] Dichotomy group tabulated at value 1.

A sorokban a szolgáltatások, az oszlopokban az iskolai végzettségek láthatók. A legutolsó sor (Total Count) mutatja a megkérdezettek iskolai végzettségének megoszlását (204, 287…66). A táblázat azt sugalmazza, hogy összefüggés van az iskolai végzettség és a használt szolgáltatások között. Pl. a ’Caller ID’-től ’3-way calling’ szolgáltatások igénybevétele közel azonos a különböző iskolai végzettségű kategóriákban. A többi szolgáltatások használatának gyakorisága azonban nő az iskolai végzettséggel. Minél magasabb végzettséggel rendelkezik a válaszadó, annál több szolgáltatást használ. Az összefüggés pozitív. A szolgáltatást nem használó kategóriában (’No services’) az iskolai végzettség növekedésével csökkennek a százalékos értékek. A két változó közötti összefüggés negatív.

Többszörös válaszadások elemzése 2.

Az SPSS programban egyéb lehetőség is rendelkezésre áll a többszörös választásos kérdések kiértékelésére. Nagyon hasonló az előbb leírtakhoz az Analyze, Tables, Multiple Response Sets… menüpont. Ebben is dichotóm és kategória változókból készíthetünk összeállításokat, szetteket. A kategória változók kódolásának tökéletesen meg kell egyezniük. A kategória változókból készített szett előállítását nem írtam le részletesen korábban, most pótolom. Az alábbi kérdésre várjuk a válaszokat:

Jelölje be a legszimpatikusabb nemzeteket! Maximum hármat választhat.

1. magyar
2. olasz
3. angol
4. román

.

.

.

41. egyéb

A lehetséges válaszok száma 41, de ebből csak maximum hármat választhat a megkérdezett. Az adatbázisban ilyenkor három kategória változót kell definiálni. Az első, második és harmadik választásra. Mindhárom változóban 41 kategória van (magyartól-egyéb nációig).

CONJOINT-ANALÍZIS

A conjoint analízis a marketingkutatás hatékony eszköze, és igen intenzíven alkalmazzák a ’70-es évek óta. Feladata a termékekkel, szolgáltatással kapcsolatos fogyasztási szokások, elvárások egységes skálán való megjelenítése. A szolgáltatás, a terméktervezés szempontjából olyan alapvető kérdésekre adja meg a választ, mint pl.:

– Egy adott terméknek vagy szolgáltatásnak milyen tulajdonságai (attribútumai) fontosak (illetve nem fontosak) a fogyasztók számára?
– Az egyes szintjeit (értékeit) hogyan preferálja a fogyasztó?
– A tulajdonságok kombinációit (a termékváltozatokat) hogyan értékelik a fogyasztók?

A conjoint analízis lépései:

1. tulajdonságok meghatározása,
2. a tulajdonságok szintjeinek meghatározása,
3. adatgyűjtési módszer megválasztása,
4. az adatgyűjtés (megkérdezés) végrehajtása,
5. az eredmények értelmezése.

Az analízis során a tulajdonságokat faktoroknak (factor), ezek konkrét értékeit szinteknek (level) és a megkérdezettek válaszait preferenciának (preference) nevezzük. A conjoint eljárás során a faktorszintekhez regressziós eljárással rendelünk értékeket úgy, hogy a faktorkombinációkhoz tartozó eredmények vagy hasznosságok a kialakult eredeti preferenciákhoz a legjobban illeszkedjenek.

A faktorok értékeik jellegének megfelelően az alábbiak lehetnek:

- **diszkrét:** az értékek kategorizált vagy ordinális skáláról származnak, és nincs előzetes információnk, a faktorszintek és a preferenciák kapcsolatáról.
- **lineáris**: a faktorszintek és a preferenciák közötti kapcsolat lineáris és az adatok legalább intervallum szintűek.
- **négyzetes**: másodfokú kapcsolatot feltételezünk a faktorszintek és preferenciák között.

A preferenciákat függő a faktorszinteket független változónak tekintjük, és a közöttük lévő kapcsolatot a legkisebb négyzetes eltérés módszerével becsült regresszióval adjuk meg. A felállított modell jóságát a becsült és a mért preferenciák közötti korrelációval ellenőrizzük. A regresszió számításában részt nem vevő kártyák (holdout) segítségével számított korreláció gyakorlatiasabb, hihetőbb eredményt ad. Ez a modell jóságának megítélésére egy sokkal szigorúbb mérőszám.

Az SPSS CONJOINT analízisben használható parancsok:

```
CONJOINT [PLAN={* }]
            {file}
    [/DATA={* }]
          {file}
    /{SEQUENCE}=varlist
     {RANK }
     {SCORE }
```

```
[/SUBJECT=variable]

[/FACTORS=varlist['labels'] ([{DISCRETE[{MORE}
]}]
                           {    { {LESS} }
                         {LINEAR[{MORE}] }
                              { {LESS} }
                               {IDEAL }
                            {ANTIIDEAL }
            [values['labels']])]
            varlist...

[/PRINT={ALL** } [SUMMARYONLY]]
        {ANALYSIS }
        {SIMULATION }
        {NONE }

[/UTILITY=file]

[/PLOT={[SUMMARY] [SUBJECT] [ALL]}]
        {[NONE**] }
```

**Kezdeti érték, ha az parancs vagy kulcsszó hiányzik.

A conjoint analízis futtatásához két fájlra van szükség:

1. Tervek (PLAN) fájl
2. Adatok (DATA) fájl

A tervek fájlt általában az SPSS programcsomag kísérlettervező moduljával (ORTHOPLAN) állítjuk elő, vagy a felhasználó is megadhatja direkt módon. A fájlnak az alábbi kötelező változókat kell tartalmaznia:

- CARD_ (a kísérlet azonosító száma)
- STATUS_ (0, 1, 2 értékkel)
- FAKTOR1
- .
- .   faktorkombinációk
- .
- FAKTORN

Az adatfájl a megkérdezettek, megfigyelők faktorkombinációkról kialakított preferenciáit vagy rangsorait tartalmazzák. A fájl kötelező változói:

- Subject (a megfigyelő azonosítója)
- A többi változó tartalmazza a személyek válaszait. Ezek száma megegyezik a tervfájl sorainak ill. kártyáinak számával. Alapesetben az azonos rangszám megengedett.

Példa:

```
DATA LIST FREE /SUBJ RANK1 TO RANK15.
BEGIN DATA
01 3 7 6 1 2 4 9 12 15 13 14 5 8 10 11
02 7 3 4 9 6 15 10 13 5 11 1 8 4 2 12
03 12 13 5 1 14 8 11 2 7 6 3 4 15 9 10
04 3 6 7 4 2 1 9 12 15 11 14 5 8 10 13
05 9 3 4 7 6 10 15 13 5 12 1 8 4 2 11
50 12 13 8 1 14 5 11 6 7 2 3 4 15 10 9
END DATA.
```

```
SAVE OUTFILE=RANKINGS.SAV.

DATA LIST FREE /CARD_ WARRANTY SEATS PRICE SPEED.
BEGIN DATA
1 1 4 14000 130
2 1 4 14000 100
3 3 4 14000 130
4 3 4 14000 100
5 5 2 10000 130
6 1 4 10000 070
7 3 4 10000 070
8 5 2 10000 100
9 1 4 07000 130
10 1 4 07000 100
11 5 2 07000 070
12 5 4 07000 070
13 1 4 07000 070
14 5 2 10000 070
15 5 2 14000 130
END DATA.

CONJOINT PLAN=* /DATA=RANKINGS.SAV
/FACTORS=PRICE (ANTIIDEAL) SPEED (LINEAR)
WARRANTY (DISCRETE MORE)
/SUBJECT=SUBJ /RANK=RANK1 TO RANK15.
```

A példa első harmada az adatfájl. Az első oszlop (SUBJ) tartalmazza a megfigyelők azonosítóját (01-50). A második oszloptól a 16-ig oszlopig (RANK1-RANK15) a megfigyelők értékítélete, kialakított rangsorai találhatók. Minden megfigyelő 15 faktorkombinációt rangsorolt. Az adatfájlt a RANKINGS.SAV fájlba mentjük el, tehát az adatok egy külső fájlba kerülnek.

A példa középső felében látjuk a faktorkombinációkat, tervfájlt, ez különböző autók tulajdonságait tartalmazza. Az első változó CARD_ a faktorkombináció, ebben az esetben az autó azonosítója, a második a garancia ideje, harmadik az ülések száma, negyedik az ár, és végül az ötödik az autó végsebességét mutatja mérföld/óra mértékegységben.

A példa utolsó harmada conjoint-analízis parancsait tartalmazza. A tervfájl az éppen aktuális munkafájl, ami az SPSS DataEditor ablakában található. Az adatfájl a külső RANKING.SAV fájl. A /FACTORS paranccsal kell megadni a faktorok, tényezők tulajdonságait. Az ár pl. ANTIIDEAL, hátrányos tulajdonságú tényező, ami azt jelenti, hogy van egy minimális szint: ahogyan nő az ár, annál hátrább kerülhet a rangsorba. A /SUBJECT parancs a megfigyelő azonosítóját tartalmazó változót adja meg. A /RANK parancs jelzi, hogy az adatfájlban rangsorok találhatók. Minden adat egy rangszámot jelöl, ami egytől indul n-ig. A kisebb rangszám nagyobb előnyt, kedveltséget jelent.

A CONJOINT-analízis adatgyűjtési módjai:

SEQUENCE

A megkérdezett személyek sorba rendezést végeznek a tervkártyákkal. Ez a leggyakrabban alkalmazott módszer. A gyakorlatban a megkérdezett kiválasztja a kártyák közül a legkedveltebbet, a maradékból megint a legkedveltebbet és így tovább. Az adatfájlban a kártyák azonosítói szerepelnek. Az adatfájl változói: legkedveltebb kártya (first) … legkevésbé kedvelt kártya (last).

RANK

Az adatfájlban minden adat egy rangszám. A változók: első kártya (RANK1)…utolsó kártya (RANKn).

SCORE parancs

Az adatfájl ebben az esetben pontszámokat tartalmaz. Például a megkérdezetteknek egy 1-től 100-ig terjedő skálán kell értékelni az adott termékeket (pl. Likert skála használata). A magasabb pontszám nagyobb kedveltséget jelent.

Az adatfájl változói: első kártya (SCORE1) … utolsó kártya (SCOREn).

A fenti három lehetőség közül csak egy adható meg.

Likert-skála

A Likert-skála lényegében a szemantikus differenciál egyik speciális változata, „egyetértő skálának” is nevezik és a marketingkutatási felméréseknél nagyon gyakran alkalmazzák. E kérdéstípus esetén egy meghatározott témakörhöz kapcsolódó „állításlistát” kell a fogyasztónak értékelni, abból a szempontból, hogy az egyes „állításokkal” milyen mértékben ért egyet vagy utasítja el őket. E skálatípust többnyire életmódkutatásoknál, piacszegmentálásnál, vállalati arculat kialakításánál stb. hasznosítják a kutatók.

SUBJECT parancs

Ezzel adhatjuk meg azt a változót, ami a megkérdezettek azonosítja.

FACTORS parancs

Ezzel adjuk meg, hogy melyik változók tartalmazzák a rangszámokat vagy pontszámokat.

- Ha nem adjuk meg a FACTORS-t, akkor az összes tényezőre DISCRETE modellt feltételez a program.
- A FACTORS parancsot a változók listája követi és zárójelben a modell megadása. A modell leírja, hogy milyen kapcsolat van a pontértékek vagy rangsorok és a tényezők szintjei között.

- A modell megadása a modell nevével, DISCRETE vagy LINEAR és opcionálisan kiegészítve a MORE vagy LESS szavakkal, valamint az IDEAL és ANTIIDEAL kulcs szavakkal történik, ami a feltételezett összefüggés irányát mutatja. Érték és értékcímkéket is megadhatunk.
- A MORE és LESS kulcs szavak nem lesznek hatással a hasznosság becslésekor. Ezeket csak a feltételezett irány azonosításra használjuk.

A következő négy modell használható:

DISCRETE

Nincs semmilyen feltételezés. A tényezők szintjei kategóriák (pl. hajszín, nem, stb.), nincs semmilyen feltételezés a tényező és a pontszámok vagy rangsor közötti kapcsolatra. Ordinális típusú faktorok esetén a MORE kulcsszó a DISCRETE után azt mutatja, hogy a tényező magasabb szintjei várhatóan jobban preferáltak. A LESS kulcsszó a fordítottját feltételezi.

LINEAR

Lineáris kapcsolat. A rangszámok vagy pontszámok várhatóan lineáris kapcsolatban vannak a vizsgált tényezővel. A MORE kulcsszó a LINEAR után azt mutatja, hogy a tényező magasabb szintjei várhatóan jobban preferáltak. A LESS kulcsszó a fordítottját feltételezi.

IDEAL

Négyzetes összefüggés, csökkenő preferenciák. Másodfokú összefüggést feltételezünk a pontszámok vagy rangszámok és a tényező között. Itt azt feltételezzük, hogy van egy ideális szintje a tényezőnek, és ettől az ideális ponttól bármelyik irányba távolodva a preferencia csökken. Ebben a modellben a tényezőnek legalább három szinttel kell rendelkeznie. Konkáv parabola.

ANTIIDEAL

Másodfokú összefüggés, növekvő preferenciák. Másodfokú összefüggést feltételezünk a pontszámok vagy rangszámok és a tényező között. Itt azt feltételezzük, hogy van egy legrosszabb szintje a tényezőnek, és ettől a legrosszabb ponttól bármelyik irányba távolodva a preferencia nő. Ebben a modellben a tényezőnek legalább három szinttel kell rendelkeznie. Konvex parabola.

Azokra a változókra, amiket nem sorolunk fel a FACTORS parancsban, a DISCRETE modell az alapértelmezett.

Eredménylista

PRINT

/ANALYSIS

csak a kísérleti adatok eredményeit írja ki.

/SIMULATIONS

csak a szimulált adatok eredményét írja ki. A három szimulációs modell eredménye – maximum utility, Bradley-Terry-Luce (BTL), és a logit – kerül kijelzésre.

/SUMMARYONLY

csak a csoport egészének jellemzőit írja ki, az egyedi megfigyelések eredményét nem. Amennyiben nagyon sok a megfigyelések száma, nem érdemes az egyedi eredményeket kiíratni.

/ALL

a kísérleti és szimulált adatok eredményét írja ki. Ez az alapbeállítás.

/NONE

Az eredmény területre nem ír ki semmit se. Ez a parancs akkor hasznos, ha az eredményeket egy fájlba szeretnénk kiíratni (UTILITY parancs).

UTILITY

Egy SPSS fájlba írja ki az eredményeket. Ha nem adjuk meg a fájlt, nem lesz létrehozva automatikusan. A parancsot a fájl nevének kell követnie. A utility fájl minden megfigyelésre egy esetet tárol.

A fájlba a változók az alábbi sorrendben íródnak ki:

- SPLIT FILE változók
- SUBJECT változó
- A megfigyelés regressziós egyenesének konstansa, neve CONSTANT.
- A DISCRETE változók estében a becsült hasznossági értékek. A változók nevei az eredeti faktor név plusz a faktorszint sorszáma.
- LINEAR változók esetén a koefficiens. A változó neve az eredeti faktor plusz _L hozzákapcsolása. (A becsült értéket a faktorérték és a koefficiens szorzatával kapjuk meg.)
- IDEAL vagy ANTIIDEAL változóknál két koefficiens. Ebben az esetben az eredeti változó nevekhez a program _L és _Q karaktert kapcsol. (A becsült értékeket úgy kapjuk meg, hogy

megszorozzuk a lineáris koefficienst a faktor értékével és hozzáadjuk a második koefficienssel szorzott faktor négyzetet.)

A conjoint-analízishez szükséges tervkártyákat, faktorkombinációkat sokszor az SPSS ORTHOPLAN moduljával állítjuk elő. A program segítségével a teljes kísérleti tervek (full factorial design) helyett redukált faktoriális terveket készíthetünk. A tervezés során a faktorkombinációk számát és az óhatatlanul fellépő információveszteséget minimalizálják. E két kívánalom ellentétesen alakul, tehát a kombinációk számának csökkentésével az információveszteség nő.

A program által megadott tervkártyák jelölései és jellemzői:

0 az ezzel jelzett tervkártya közönséges kártya, részt vesz a becslésben és a modell jóságának megítélésében. A felmérésben szerepel. (status=design)

1 visszatartatott kártya „holdout”, nem vesz részt a becsült értékek meghatározásában csak a modell jóságának megítélésében. A felmérésben szerepel, rendelkezik a válaszadók minősítésével. (status=holdout)

2 szimulált kártya. A megkérdezés során nem szereplő kártya, a modell becsült értékei alapján értékelt kártya. (status=simulation)

Korlátozások

– A faktorok csak numerikus változók lehetnek.

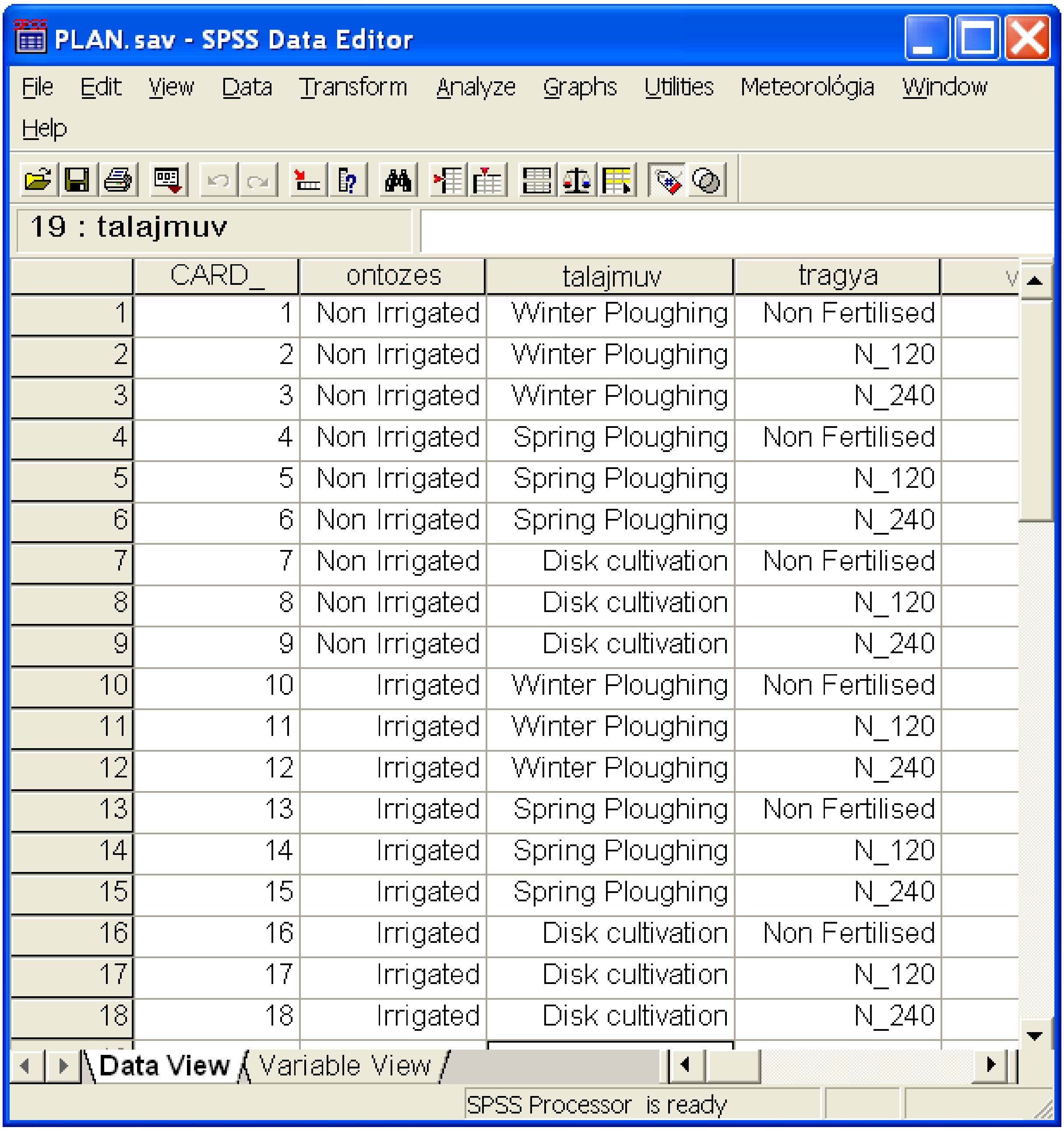
- A tervfájl nem tartalmazhat üres adatokat vagy súlyokat. A munka adatfájlban, ha bármelyik preferencia érték hiányzik (rangszám, súly, vagy kártyaszám), a megfigyelés ki lesz zárva az analízisből.
- A faktoroknak minimum két szinttel kell rendelkezniük. Minden egyes faktor maximum 99 szinttel rendelkezhet.

Termesztési tényezők értékelése conjoint-analízissel

Agronómiai kutatásokban – tudomásom szerint – eddig még nem alkalmazták ezt a módszert, pedig egy termesztési eljárás felfogható termék vagy szolgáltatásként is, melyet a termesztési tényezők kombinációjával lehet jellemezni. A módszer „full concept” megközelítést alkalmaz, ami azt jelenti, hogy a növényi növekedés, a termés nagysága a termesztési eljárás valamennyi jellemzőitől együttesen függ.

Bemutatom, hogyan lehet kísérleti adatok felhasználásával a legkedvezőbb termesztési eljárást, illetve az ehhez tartozó faktorkombinációt meghatározni.

Az adatok 1990-2009 között monokultúrában termesztett kukorica kísérletből származnak 18 különböző termesztési eljárásból (2 öntözés x 3 talajművelés x 3 trágyázás). Ez képezte a Conjoint-analízis tervfájlját (PLAN). A termesztési eljárások megítélésekor a vizsgált időszakot együttesen értékelem, de az eljárás lehetőséget ad jellemző évjáratok (aszályos, átlagos, csapadékos) elkülönítésére is.



	CARD_	ontozes	talajmuv	tragya
1	1	Non Irrigated	Winter Ploughing	Non Fertilised
2	2	Non Irrigated	Winter Ploughing	N_120
3	3	Non Irrigated	Winter Ploughing	N_240
4	4	Non Irrigated	Spring Ploughing	Non Fertilised
5	5	Non Irrigated	Spring Ploughing	N_120
6	6	Non Irrigated	Spring Ploughing	N_240
7	7	Non Irrigated	Disk cultivation	Non Fertilised
8	8	Non Irrigated	Disk cultivation	N_120
9	9	Non Irrigated	Disk cultivation	N_240
10	10	Irrigated	Winter Ploughing	Non Fertilised
11	11	Irrigated	Winter Ploughing	N_120
12	12	Irrigated	Winter Ploughing	N_240
13	13	Irrigated	Spring Ploughing	Non Fertilised
14	14	Irrigated	Spring Ploughing	N_120
15	15	Irrigated	Spring Ploughing	N_240
16	16	Irrigated	Disk cultivation	Non Fertilised
17	17	Irrigated	Disk cultivation	N_120
18	18	Irrigated	Disk cultivation	N_240

74. ábra: Tervfájl a conjoint-analízishez

A számításokat az SPSS 13.0 statisztikai programcsomaggal készítettem (SPSS, 2004). Mivel a Conjoint-analízis a párbeszédpanelen keresztül nem érhető el, ezért először meg kellett írni egy rövid programot, ami a syntax editor ablakban futtatható. A program összekapcsolja a terv és adatfájlt, valamint definiálja a tényezők és preferenciák tulajdonságait.

```
CONJOINT PLAN=*
/DATA='C:\Documents and Settings\Felhasználó\Dokumentumok\Pub-
likációim\Pub_konferencia\AlpsAdria2010\Data.sav'
     /SCORE=termes.1 TO termes.18
/SUBJECT=ev
/FACTORS=ontozes (DISCRETE MORE) talajmuv (DISCRETE) tragya
(IDEAL)
/PLOT ALL.
```

A conjoint-modell megalkotásakor az öntözést kategória tényezőnek tekintettem, és feltételeztem, hogy az öntözés növeli a termést (discrete more). A talajművelést szintén kategória változóként állítottam be a modellbe, azonban itt nem tettem semmilyen feltételezést a talajművelés és a termés kapcsolat között (discrete). A műtrágyázást a modellbe ideális tényezőnek (ideal) tekintettem, ami azt jelenti, hogy van egy ideális szintje, ahol a termés a legnagyobb, és ettől a szinttől lefelé ill. felfelé a termés csökken. A különböző termesztési eljárások eredményességét, preferenciáit a kukorica évenkénti termésével minősítettem, pontoztam (score), tehát subject változónak a termést adtam meg, ez volt a Conjoint-analízis adatfájlja (DATA).

	ev	termes.1	termes.2	termes.3	termes.4	termes.5	termes.6
1	1990	5.53	7.23	6.97	5.55	7.24	7.24
2	1991	9.77	11.54	11.21	10.50	11.53	11.02
3	1992	3.69	4.55	3.73	2.48	3.29	3.06
4	1993	5.86	8.86	8.55	4.88	7.28	7.47
5	1994	3.71	5.61	4.76	3.49	5.30	4.73
6	1995	2.91	1.55	.97	2.24	1.09	.61
7	1996	6.31	9.22	7.98	6.30	8.98	7.98
8	1997	6.41	11.24	10.95	6.06	10.42	10.37
9	1998	6.39	11.17	10.25	6.21	10.49	9.53
10	1999	7.64	11.12	11.48	7.59	9.94	10.25
11	2000	5.86	7.86	7.50	3.62	5.41	5.32
12	2001	7.86	10.92	11.22	4.89	5.48	6.79
13	2002	5.79	6.80	6.23	4.34	5.57	5.31
14	2003	5.35	8.16	7.70	5.05	5.68	4.90
15	2004	8.77	11.32	11.26	7.68	10.82	10.59
16	2005	8.74	10.49	10.80	7.22	10.92	10.81
17	2006	7.13	8.84	8.71	6.28	8.50	8.49
18	2007	1.51	.90	.39	1.58	1.13	.48
19	2008	11.97	13.05	12.57	12.87	12.58	12.19
20	2009	6.47	8.34	7.72	6.57	8.48	7.85

75. ábra: Adatfájl a conjoint-analízishez

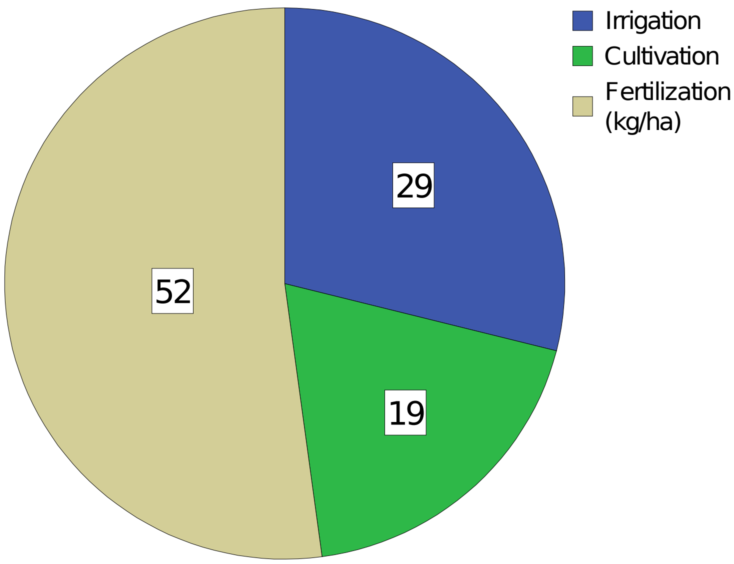
A syntax editor ablakban található program lefuttatása után az output ablakban jelenik meg az analízis eredménye.

A modell jóságát a becsült és mért preferenciák közötti korrelációs koefficiensekkel minősíti. A Pearson-féle a lineáris kapcsolat irányát és erősségét méri. Az egyhez közeli érték nagyon szoros összefüggést mutat, és a modell jóságát igazolja. A Kendall-féle koefficiens az előbbi nem paraméteres változata, ami az ordinális típusú adatok közötti kapcsolat szorosságát és irányát méri. A számított szignifikancia értékek kisebbek, mint 0,05, ezért a kapcsolat valósnak tekinthető.

A modell jósága:

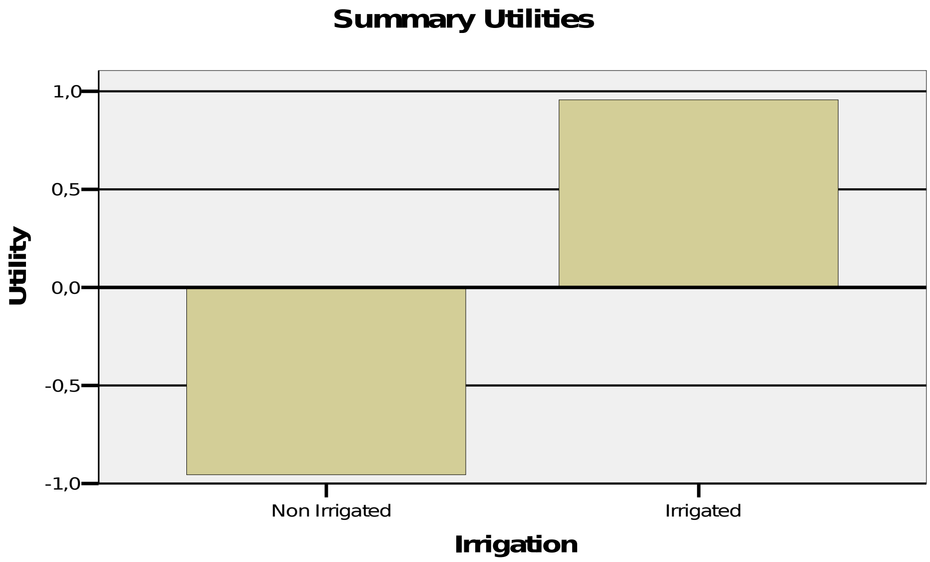
```
Pearson's R   =  0.961                 Significance =  .0000

Kendall's tau =  0.895                 Significance =  .0000
```

A termesztési tényezők átlagos fontosságát a vizsgált időszak alatt (1990-2009) a 76. ábra mutatja. A legfontosabb termesztési tényező a tápanyag-utánpótlás. 52%-ban befolyásolja az eljárások sikerességét. Második az öntözés, 29%. Ezek szerint a kísérleti hely klimatikus- és talajviszonyai között a műtrágyázás fontosabb, mint az öntözés, ezért csak a szakszerű tápanyag-visszapótlás megvalósítása után lehet az öntözésfejlesztésen gondolkodni. A legkevésbé fontos tényező a talajművelés (19%). Ez a jó szerkezetű, könnyen művelhető csernozjom talajnak köszönhető, amelyen a különböző talajművelési eljárások csak csekély mértékben módosítják a növények termését.



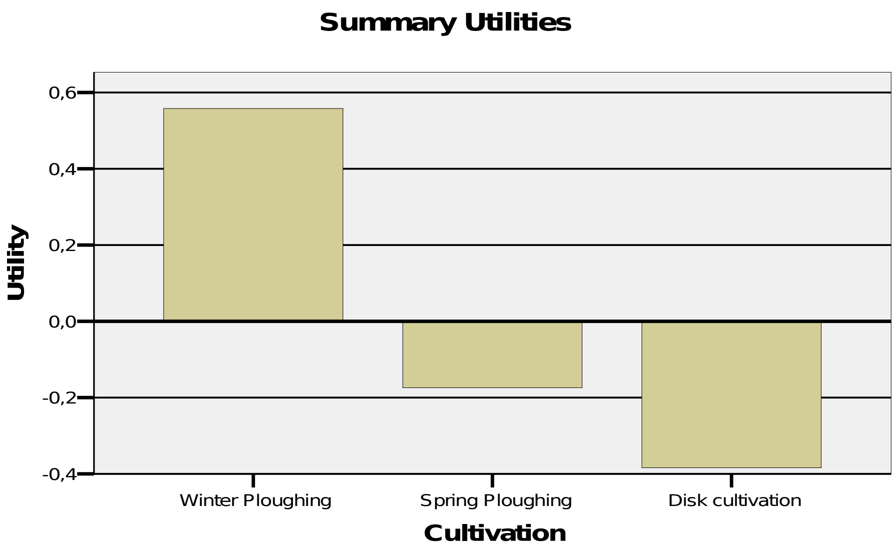
76. ábra: Átlagos fontosság, 1990-2009

A 77. ábra az öntözés két változatának hasznosságát mutatja. Mivel csak két szint van, és a conjoint-analízis szimmetrikusan adja meg a hatásokat, ezért az öntözés hasznossága a nem öntözött ellentettje. A hasznosság mérőszáma a kukoricatermés nagysága volt t/ha mértékegységben, így a függőleges tengelyen a t/ha-ban mért hatások láthatók. A kísérlet átlagtermését az öntözés 956 kg/ha-val növeli, ha elhagyjuk ugyanennyivel csökkenti. A nem öntözött és öntözött kezelések között tehát átlagban 1 912 kg/ha különbség van.



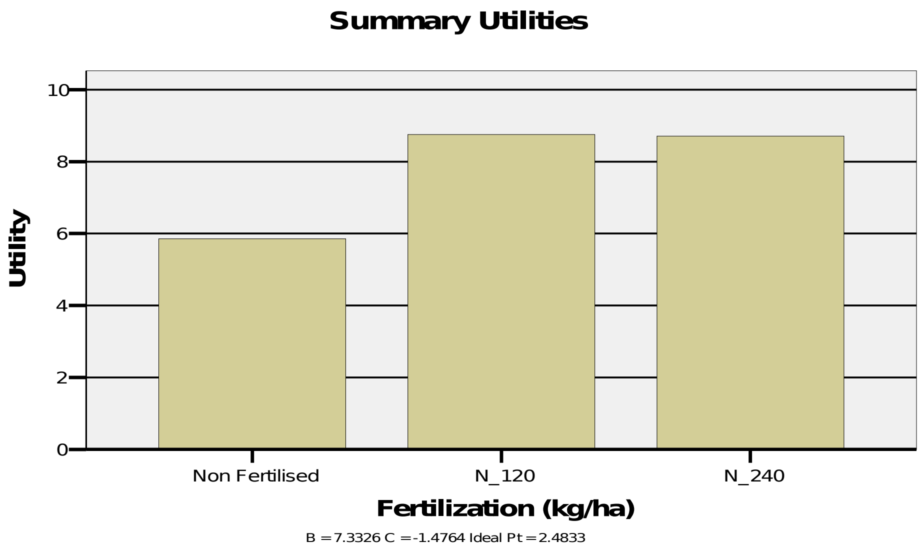
77. ábra: Az öntözés hasznossága, 1990-2009

A talajművelési változatok hasznosságát mutatja a 78. ábra. A leghasznosabb eljárás az őszi szántás, ez egyértelműen termésnövekedést okozott a vizsgált időszakban, az átlagos termést 558 kg/ha-val növeli. A másik két talajművelési eljárásnak negatív hatása van az átlagterméshez viszonyítva. A termés-csökkenés a tárcsás művelésnél a legnagyobb, -384 kg/ha.



78. ábra: A talajművelés hasznossága, 1990-2009

A 79. ábra a műtrágyázás hasznosságát mutatja. A modellben ideális tényezőként vettem figyelembe, ezért megkapjuk a másodfokú függvény paramétereit és a műtrágyázás optimális szintjét is. Az optimális szint itt a maximális kukoricatermésekhez tartozó műtrágyaadagot jelenti.

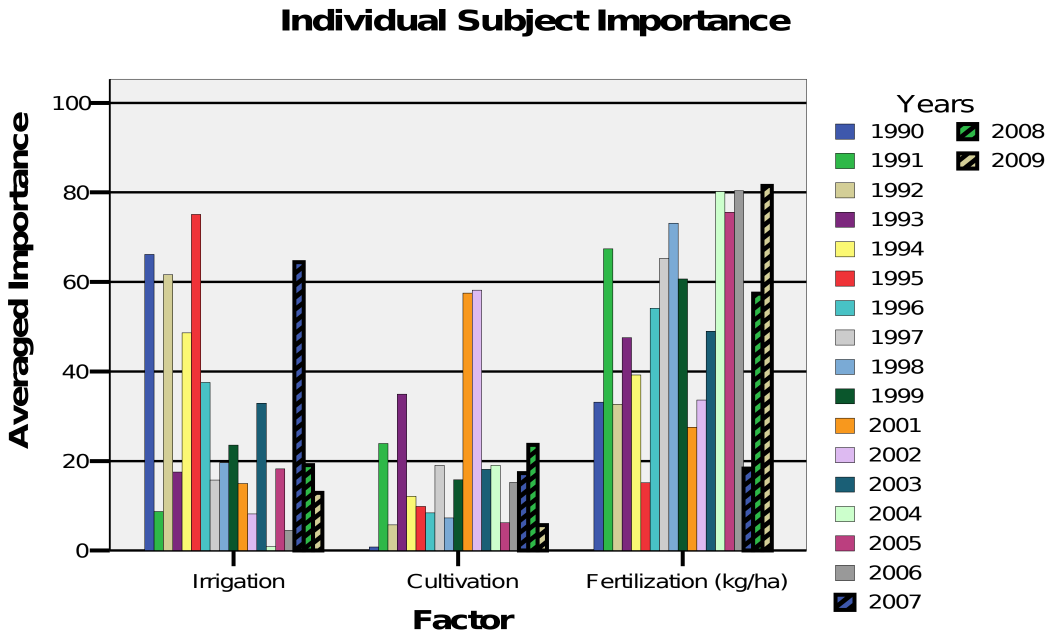


79. ábra: A műtrágyázás hasznossága, 1990-2009

Az optimális szint a műtrágyázás 2,4833 szintje, amely 178 kg/ha nitrogénnek felel meg a kísérlet összes kezelésének átlagában.

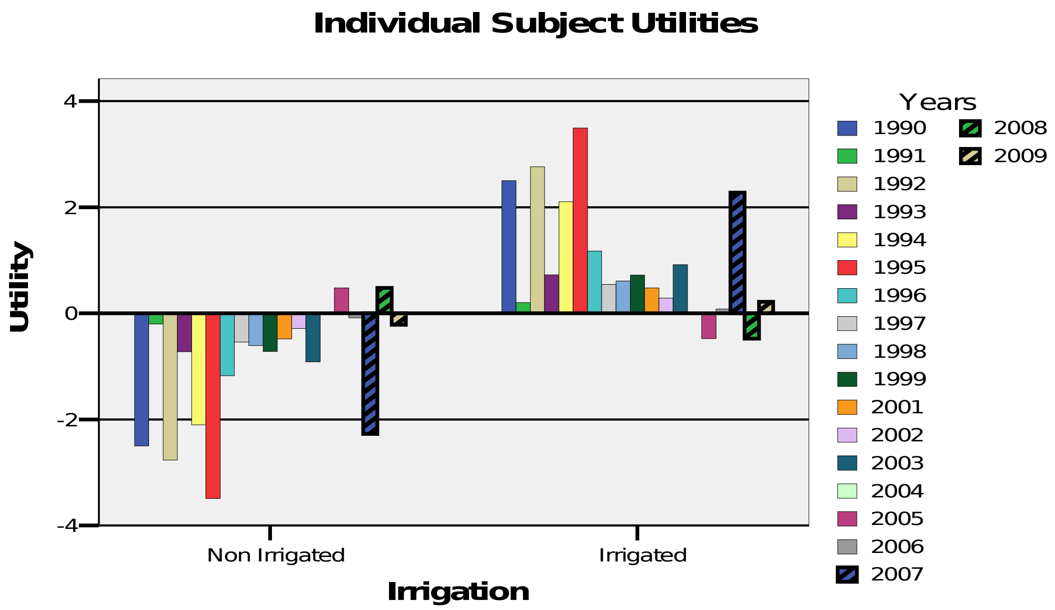
A fenti eredmények alapján a legeredményesebb termesztési eljárás: 178 kg/ha nitrogén, öntözött, őszi szántás. Ettől a kombinációtól való eltérés bármely irányban csökkenti a kukorica 1990-2009 közötti termését.

Megkapjuk az évenkénti eredményeket is, ami árnyaltabb elemzést tesz lehetővé.



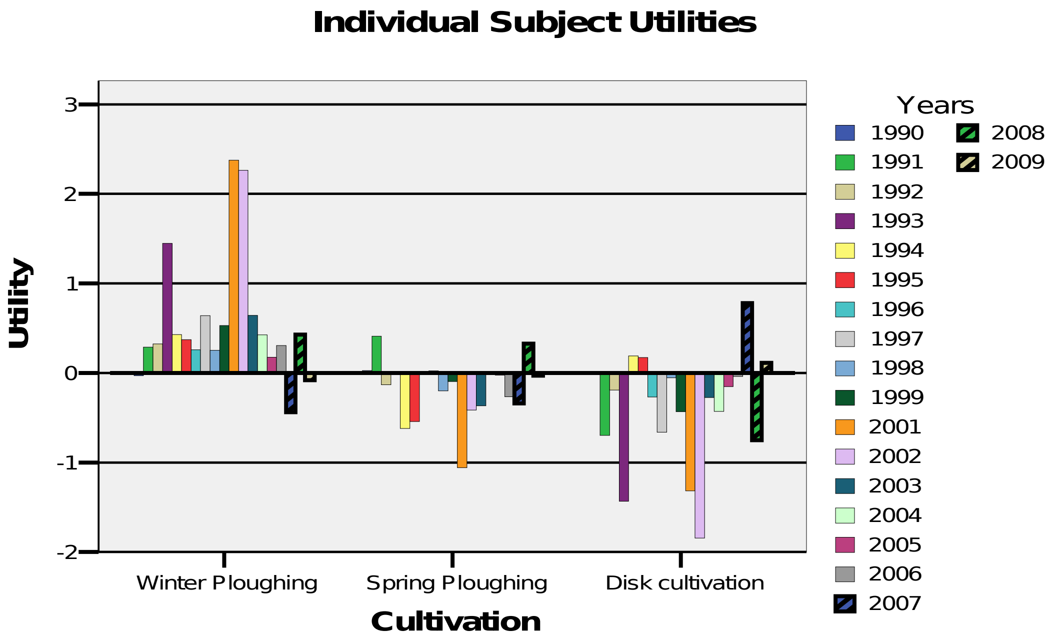
80. ábra: A termesztési tényezők évenkénti fontossága

Évenként a termesztési tényezők fontosságának összege 100%. Jól látható, hogy a három tényező fontossága évről-évre változik. Száraz évben, pl. 1995, az öntözés szerepe megnő.

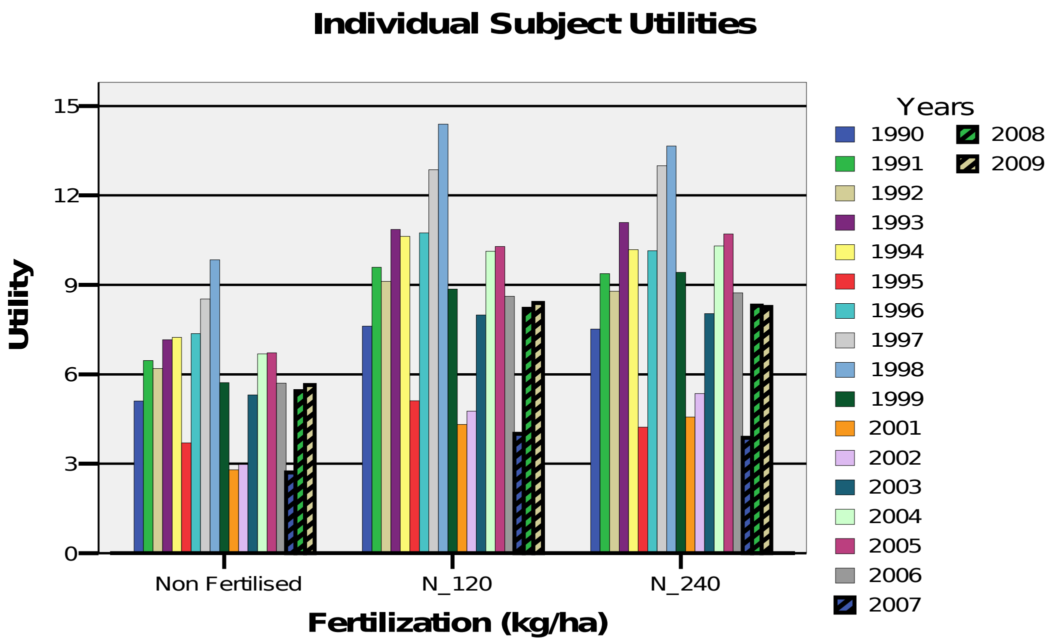


81. ábra: Az öntözés évenkénti hasznossága

Az öntözésről azt feltételeztem, hogy mindig növeli a kukorica termését. Azonban két évben ettől eltérő eredményt kaptam (2005, 2008). Három évben az öntözésnek gyakorlatilag nem volt hatása (1991, 2006, 2009).



82. ábra: A talajművelés évenkénti hasznossága



83. ábra: A műtrágyázás évenként hasznossága

A műtrágyázás termésnövelő hatása is évről-évre változik (83. ábra). Száraz években alig növeli a termést, csapadékos években kiemelkedő a hatása.

Egyetlen év eredménye látható lent.

```
SUBJECT NAME:  1990

Importance   Utility(s.e.)  Factor

                                ontozes     Irrigation
66.12      -2.5021( .1232)    -         Non Irrigated
             2.5021( .1232)     -        Irrigated


                               talajmuv    Cultivation
  .78        -.0320( .1742)             Winter Ploughing
              .0267( .1742)             Spring Ploughing
              .0053( .1742)             Disk cultivation

                              tragya      Fertilization (kg/ha)
33.11        5.1030( .7985)     ---       Non Fertilised
             7.6086(1.0881)     ----      N_120
             7.5169( .9054)     ----     N_240
           B =  6.4017(1.0563)  C = -1.2987( .2614)  Ideal Pt =  2.4647

             2.4976( .9302) CONSTANT

Pearson's R   =  .988                   Significance =  .0000

Kendall's tau =  .830                   Significance =  .0000
```

Az összesítő jelentésben tájékozódhatunk, hogy a modell mennyire volt megfelelő. Látható, hogy két esetben a feltételezés megfordult (reversal). Mindkettő az öntözésnél volt, az 2005 és 2008. évben.

```
SUBFILE SUMMARY

Reversal Summary:

     2 subjects had  1 reversals

Reversals by factor:

   ontozes   2
   tragya   0
   talajmuv   0
```

Reversal index:

Page	Reversals	Subject
1	0	1990
2	0	1991
3	0	1992
4	0	1993
5	0	1994
6	0	1995
7	0	1996
8	0	1997
9	0	1998
10	0	1999
12	0	2001
13	0	2002
14	0	2003
15	0	2004
16	1	2005
17	0	2006
18	0	2007
19	1	2008
20	0	2009

GYAKORLÓ FELADATOK

1. Meglévő adatbázis tulajdonságainak megtekintése, (Display Data Info)
2. Olvassa be a *meteorológia.txt* fájlt. Mentse el a fájl formátumát *.tpf kiterjesztéssel! Text típusú adatok beolvasása.
3. Egészítse ki a változókat további információkkal meglévő adatbázisból! (Apply Data Dictionary)
4. Rendezze növekvő, majd csökkenő sorrendbe a *termés1989.sav* fájlt!
5. Bővítse ki a *termés1989.sav* fájlt a *termés1995.sav* fájl eseteivel! (Merge Files, Add Cases…)
6. Kapcsolja össze a *termés.sav* fájlt a *csapadék.sav* fájllal! (Merge Files, Add Variables, Match cases on key variables in sorted files, External file is keyed table, Excluded Variables: év változó → Key Variables. OK) Őrizze meg az utasításokat!
7. Alakítsa át az *esztendő2002.sav* fájl „jdate” változóját hónapokra! (Transform, Compute, DATE.YRDAY(2002,jdate) dátum előállítása, XDATE.MONTH(datum) a hónap számainak képzése)
8. Agregálja az *esztendő2002.sav* fájlt változóit értelemszerűen, a hónapok alapján! (Data, Aggregate…, grad.sum; tmax.mean; tmin.mean; csapad.sum)
9. A *termés.sav* fájlban kódolja át az aktuális műtrágyaadagokat egy másik változóba (1→0, 60-180→90, 240-300→150). 1990 és 1991 esztendőben! (Transform, Recode, Into Different Variables)
10. Ossza fel a *termés.sav* fájl „termés” változóját négy egyforma számú kategóriára! (Transform, Categorize Variables)
11. Számítsa ki a *termés.sav* fájl „termés” változójának rangszámait és ábrázolja vonal diagrammal! (Transform,

Rank Cases), (Graphs, Line Charts, Category Axis: rank of termés, Variable: mean of termés)

12. Próbálja ki az automatikus újrakódolást saját készítésű text típusú adatbázison!

13. Számítsa ki *termés.sav* fájl „termés” változójának legfontosabb statisztikai mutatóit a különböző műtrágya kezelésekben! (Analyze, Reports, Case summaries, Variables: termés, Grouping Variables: műtrágyázás)

14. Számítsa ki az *esztendő2002.sav* fájl minden változójának jellemző éves értékét! (Analyze, Reports, Reports summaries in Columns…)

15. Állapítsa meg, hogy a *Termés1989.sav* fájlban hány parcellát figyeltünk meg a talajművelés x tőszám kombinációban! (Analyze, Descriptive Statistics, Crosstabs…)

16. Ábrázolja oszlopdiagrammal a *Termés1989.sav* fájlból a termés nagyságát a műtrágyázás függvényében! Graphs, Bar…, Simple, Summaries for groups of cases)

17. Ábrázolja oszlopdiagrammal az *esztendő2002.sav* fájlból a globálsugárzás, minimális, maximális és átlagos értékét! Graphs, Bar…, Simple, Summaries of separate variables)

18. Ábrázolja az *esztendő2002.sav* fájlból a globálsugárzás éves menetét! (Graphs, Line…, Simple, Values of individual cases, Line Represents: globálsugárzás, Category Labels Variable: az év napja)

19. A *termés1989.sav* fájlból számítsa ki a „termés” változó különböző statisztikai mutatóit a talajművelés függvényében! Állapítsa meg az összefüggés szorosságát és lineáris jellegét! Mennyiben határozza meg a talajművelés a kukorica termését ebben az esztendőben?

20. Állapítsa meg, hogy a kukorica termése 10t/ha volt-e az 1989-as esztendőben!

21. Van-e különbség a nem trágyázott és a 120 kg/ha nitrogénnel műtrágyázott kukorica termése között 1995-ben? (kétmintás t-próba)

Függelék

Az *esztendő2002.sav* fájl szerkezete:

```
Name
Position

JDATE     Az év napja
1
          Measurement Level: Scale
          Column Width: Unknown  Alignment: Right
          Print Format: F3
          Write Format: F3

GRAD      Globálsugárzás (MJ/m2 nap)
2
          Measurement Level: Scale
          Column Width: Unknown  Alignment: Right
          Print Format: F4.1
          Write Format: F4.1

TMAX      Maximális hőmérséklet (C)
3
          Measurement Level: Scale
          Column Width: Unknown  Alignment: Right
          Print Format: F4.1
          Write Format: F4.1

TMIN      Minimális hőmérséklet (C)
4
          Measurement Level: Scale
          Column Width: Unknown  Alignment: Right
          Print Format: F4.1
          Write Format: F4.1

CSAPADÉK  Napi csapadék (mm)
5
          Measurement Level: Scale
          Column Width: Unknown  Alignment: Right
          Print Format: F4.1
          Write Format: F4.1
```

A *csapadék.sav* fájl szerkezete:

```
Name
Position

ÉV          Esztendő
1
            Measurement Level: Scale
            Column Width: Unknown  Alignment: Right
            Print Format: F4
            Write Format: F4

CSAPADÉK    Éves csapadék (mm)
2
            Measurement Level: Scale
            Column Width: Unknown  Alignment: Right
            Print Format: F8.2
            Write Format: F8.2
```

A termés.sav fájl szerkezete:

```
Name
Position

ÉV
1
            Measurement Level: Scale
            Column Width: 4  Alignment: Right
            Print Format: F4
            Write Format: F4

NPK         műtrágyázás
2
            Measurement Level: Nominal
            Column Width: 7  Alignment: Right
            Print Format: F3
            Write Format: F3

            Value    Label

                1    nem trágyázott
               30    N30, P23, K27
```

```
          60    N60, P45, K53
          90    N90, P68, K80
         120    N120, P90, K106
         150    N150, P113, K133
         180    N180, P135, K159
         240    N240, P180, K212
         300    N300, P225, K265

TERMÉS    Termés (t/ha)
3
          Measurement Level: Scale
          Column Width: Unknown  Alignment: Right
          Print Format: F8.2
          Write Format: F8.2
```

A *termés1989.sav* fájl szerkezete:

```
Name
Position

EV        évek
1
          Measurement Level: Scale
          Column Width: 5  Alignment: Right
          Print Format: F4
          Write Format: F4

TALAJMUV  Talajművelés
2
          Measurement Level: Nominal
          Column Width: 5  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    őszi szántás
              2    tavaszi szántás
              3    tárcsás

TOSZAM    Tőszám
3
          Measurement Level: Scale
          Column Width: 4  Alignment: Right
```

```
          Print Format: F2
          Write Format: F2

HIBRID
4
          Measurement Level: Nominal
          Column Width: 6  Alignment: Right
          Print Format: F2
          Write Format: F2

          Value    Label

              1    De 351
              2    De 377
              3    De 382
              4    Dk 366
              5    Dk 373
              6    Dk 391
              7    Dk 471
              8    Dk 477
              9    DK 493
             10    Dk 524
             11    Dk 527
             12    Kanada
             13    Katinka
             14    LG 2298
             15    Lipesa
             16    Maya
             17    Mv 444
             18    Mv 484
             19    MV 487
             20    Occitán
             21    Pannónia
             22    Pelikán
             23    Sprinter
             24    Stira
             25    Szegedi 348
             26    Veronika (Sze 427)
             27    Volga SC
             28    Szegedi 463
             29    Dk 440
             30    Ella
             31    Hunor
```

```
TRAGYA      Trágya kezelés
5
            Measurement Level: Ordinal
            Column Width: 9  Alignment: Right
            Print Format: F1
            Write Format: F1

            Value    Label

                1    nem trágyázott
                2    nitrogén 120
                3    nitrogén 240

ISMÉTLÉS
6
            Measurement Level: Nominal
            Column Width: 5  Alignment: Right
            Print Format: F1
            Write Format: F1

TERMÉS      termés t/ha
7
            Measurement Level: Scale
            Column Width: 6  Alignment: Right
            Print Format: F6.3
            Write Format: F6.3
            Missing Values: -999.0
```

A *termés1995.sav* fájl szerkezete:

```
Name
Position

EV          évek
1
            Measurement Level: Scale
            Column Width: 5  Alignment: Right
            Print Format: F4
            Write Format: F4

ONTOZES     Öntözés
2
            Measurement Level: Ordinal
```

```
          Column Width: 9  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    Nem öntözött
              2    Öntözött

ELOVET    Elővetemény
3
          Measurement Level: Nominal
          Column Width: 7  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    Kukorica
              2    Búza

TALAJMUV  Talajművelés
4
          Measurement Level: Nominal
          Column Width: 5  Alignment: Right
          Print Format: F1
          Write Format: F1

          Value    Label

              1    őszi szántás
              2    tavaszi szántás
              3    tárcsás

TOSZAM    Tőszám
5
          Measurement Level: Scale
          Column Width: 4  Alignment: Right
          Print Format: F2
          Write Format: F2

HIBRID
6
          Measurement Level: Nominal
```

```
            Column Width: 6  Alignment: Right
            Print Format: F2
            Write Format: F2

            Value    Label

                1    De 351
                2    De 377
                3    De 382
                4    Dk 366
                5    Dk 373
                6    Dk 391
                7    Dk 471
                8    Dk 477
                9    DK 493
               10    Dk 524
               11    Dk 527
               12    Kanada
               13    Katinka
               14    LG 2298
               15    Lipesa
               16    Maya
               17    Mv 444
               18    Mv 484
               19    MV 487
               20    Occitán
               21    Pannónia
               22    Pelikán
               23    Sprinter
               24    Stira
               25    Szegedi 348
               26    Veronika (Sze 427)
               27    Volga SC
               28    Szegedi 463
               29    Dk 440
               30    Ella
               31    Hunor

TRAGYA      Trágya kezelés
7
            Measurement Level: Ordinal
            Column Width: 9  Alignment: Right
            Print Format: F1
            Write Format: F1
```

```
           Value    Label

               1    nem trágyázott
               2    nitrogén 120
               3    nitrogén 240

ISMÉTLÉS
8
           Measurement Level: Nominal
           Column Width: 5  Alignment: Right
           Print Format: F1
           Write Format: F1

TERMÉS     termés t/ha
9
           Measurement Level: Scale
           Column Width: 6  Alignment: Right
           Print Format: F6.3
           Write Format: F6.3
           Missing Values: -999.0
```

Paraméteres eljárások. Várható érték vagy szórás?

Várható érték?	
Várható érték egyezése adott értékkel?	
Szórás ismert?	
Igen	*Nem*
EGYMINTÁS U-PRÓBA	**EGYMINTÁS T-PRÓBA**
Két várható érték egyezése? Az elméleti szórások ismertek?	
Igen	*Nem*
KÉTMINTÁS U-PRÓBA	**KÉTMINTÁS T-PRÓBA**
Összetartozó adatpárok különbségének tesztelése? Az elméleti szórások ismertek?	
Igen	*Nem*
PÁRONKÉNTI T-TESZT	
Több várható érték egyezése? A mintavétel egy szempont szerint történik? Szórások egyenlők?	
EGYTÉNYEZŐS VARIANCIA-ANALÍZIS, WELCH,	

BROWN-FORSYTHE-PRÓBA	

Több várható érték egyezése? A mintavétel két szempont szerint történik? Szórások egyenlők?	
KÉTTÉNYEZŐS VARIANCIA-ANALÍZIS **BROWN-FORSYTHE-PRÓBA**	
Több várható érték egyezése? A mintavétel több szempont szerint történik? Szórások egyenlők?	
TÖBBTÉNYEZŐS VARIANCIA-ANALÍZIS	

SZÓRÁS?	
Két szórás egyezése?	*Több szórás egyezése? Minták elemszáma egyenlő?*
F-PRÓBA **LEVENE-TESZT**	**LEVENE-TESZT, MAX. F-PRÓBA** **COCHRAN-PRÓBA** **BARTLETT-PRÓBA, LEVENE-TESZT**

Nem paraméteres eljárások az eloszlások, vagy a mediánok tesztelésére alkalmasak.

Eloszlás egyezése egy adott eloszlással (egymin-tás próba)?	*Medián egyezése adott értékkel?*
CHI-NÉGYZET PRÓBA (RELATÍV GYAKORISÁ-GOK ÖSSZEHASONLÍTÁSA)	ELŐJEL-PRÓBA
Két eloszlás egyezése, homogenitás vizsgálat?	*Két várható érték egyezése?*
CHI-NÉGYZET PRÓBA	**ELŐJEL-PRÓBA, MANN-WHITNEY, WILCOXON-PRÓBA**
Két esemény függetlenségének tesztje?	*Két összetartozó minta egyezése?*
FÜGGETLENSÉG VIZSGÁLAT, CHI-NÉGYZET PRÓBÁVAL	**WILCOXON-TESZT, ELŐJEL-PRÓBA**
	Több várható érték egyezése? A mintavétel egy szempont alapján történik?
	KRUSKAL-WALLIS-PRÓBA (paraméteres: egytényezős variancia-analízis)
	Több várható érték egyezése? A mintavétel egy szempont alapján történik? Minta elemszámok azonosak?
	FRIEDMAN-TESZT (paraméteres: kéttényezős variancia-a-

	nalízis)

Ajánlott irodalom

SPSS:

Falus István – Ollé János: Statisztikai módszerek pedagógusok számára, OKKER Kiadó, 2000. (Excel és SPSS alkalmazásokkal)

Huzsvai L. (2004): Biometriai módszerek az SPSS-ben. http://www.agr.unideb.hu/~huzsvai

Katona Tamás - Lengyel Imre (szerk.): Statisztikai ismerettár - fogalmak, képletek, módszerek Excel és SPSS alkalmazásokkal. JATEPress, Szeged, 1999. 121 oldal, (közgazdász, jogász, kísérletes és társadalomtudomány)

Ketskeméty L. – Izsó L.: Bevezetés az SPSS programrendszerbe. Módszertani útmutató és feladatgyűjtemény statisztikai elemzésekhez. ELTE Eötvös Kiadó, Budapest, 2005.

Ketskeméty L. – Izsó L.: Az SPSS for Windows programrendszer alapjai, Felhasználói útmutató és oktatási segédlet. Budapest, 1996.

Moksony Ferenc: Gondolatok és adatok: Társadalomtudományi elméletek empirikus ellenőrzése. Budapest, Osiris Kiadó, 1999.

SPSS 13.0 Base User’s Guide (2004). ISBN 0-13-185723-1, Printed int he USA.

Székelyi Mária - Barna Ildikó: Túlélőkészlet az SPSS-hez. TYPOTEX, 2002, ISBN 963 9326 429

Statisztika:

Anscombe, F.J. (1973). Graphs in statistical analysis. American Statistician, 27, 17-21.

Baráth Cs.-né. - Ittzés A. - Ugrósdy Gy.: 1996. Biometria: módszertan és a MINITAB programcsomag alkalmazása. Mezőgazda Kiadó, Budapest

G.U. Yule – M.G. Kendall: Bevezetés a statisztika elméletébe. Közgazdasági és Jogi könyvkiadó, Budapest. 1964.

Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting,* 4, 1-28.

Harnos ZS. szerk.: 1993. Biometriai módszerek és alkalmazásaik MINITAB programcsomaggal. AKAPRINT, Budapest

Lehota József (2001): Marketingkutatás az agrárgazdaságban. http://www.tankonyvtar.hu/marketing/marketingkutatas-080905-134

Lothar Sachs.: Statisztikai módszerek. Mezőgazdasági Kiadó, Budapest, 1985.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman. 1997. *Forecasting: Methods and Applications.* New York: John Wiley & Sons.

Mérő, L.: 1986. A többdimenziós skálázás alapelvei. Pszichológia, (6), 3, 399-433.

Móri F.T. – Székely J.G.: Többváltozós statisztikai analízis. Műszaki Könyvkiadó, Budapest, 1986. (ISBN 963 10 6684 3)

Sváb J.: Biometriai módszerek a kutatásban. Mezőgazdasági Kiadó, Budapest, 1973. (második, átdolgozott, bővített kiadás)

Sváb J.: Többváltozós módszerek a biometriában. Mezőgazdasági Kiadó, Budapest, 1979. (ISBN 963 230 011 4)

Szűcs I. (szerk.)(2002): Alkalmazott statisztika. Tankönyv, Agroinform K. Budapest

GAUSS, CARL FRIEDRICH
(1777. 04. 30. - 1855. 02. 23.)

Német matematikus, csillagász és fizikus. Őt tartják minden idők egyik legnagyobb matematikusának. Így is nevezik: "A matematikusok fejedelme." Euler mellett ő a matematika legsokoldalúbb tudósa.

Braunschweigben született, édesapja nyergesmester volt. Már 6 éves korában kitűnt matematikai tehetségével. Tanítója egyszer azt a feladatot adta a kis tanulóknak, hogy adják össze a számokat 1-től 40-ig, mivel a tanító úr addig egy másik évfolyammal akart foglalkozni, és így akarta addig a kicsiket lefoglalni. De a kis Gauss hamarosan jelentkezett a jó eredménnyel. Csodálkozó tanítójának el is magyarázta, hogyan csinálta.

Párba állította a számokat 40 + 1 = 39 + 2 = 38 + 3 stb. Ezek a párok mindig 41-t adnak összegül, és mivel 20 ilyen pár van, az eredmény 820. Ez a gondolkozás megegyezik a számtani sorozat összegének meghatározásánál alkalmazottal. Tanítója felismerve a kisfiú rendkívüli képességeit, jelentette az esetet elöljáróinak. Így jutott el a híre braunschweig-i herceghez, aki felkarolta a kis Gauss-t. Gimnáziumba került, majd a göttingeni egyetemre.

Pályája töretlenül ívelt felfelé. Ismerte és barátjának nevezte Bolyai Farkast, ennek ellenére fiát Bolyai Jánost nem támogatta, és ezzel igen nagy csalódást okozott mindkettőjüknek. Sajnos Gauss mások elismerésével is fukarkodott. Így például Abel tehetséges norvég matematikussal kapcsolatban is. Lobacsevszkij orosz matematikust ugyan beajánlotta a Göttingeni Tudományos Társaságba, de a nem euklideszi geometria megalkotásának területén végzett munkásságának közvetlen elismerésétől tartózkodott, akárcsak Bolyai János esetében.

Gauss csillagászként is számottevőt alkotott. 1801-ben egy új és egyszerűbb módszert dolgozott ki a bolygó pályájának kiszámítására. 1820 körül geodéziával (földmérés) kezdett foglalkozni. Fizikai munkássága is említésre méltó. Göttingenben egy szobor ábrázolja őt és Wilhelm Webert a távíró 1833-ban történő feltalálása közben. Ő alkotja meg az első abszolút fizikai mértékegységrendszert. Még számológép fejlesztéssel is foglalkozott, Leibniz gépét tökéletesítette. Ez a gép az ő idejében népszerű volt egész Németországban.

Gauss békés, hosszú és elismert életet élt. Igazi zsenialitást még így is nehéz teljes egészében felmérni, mert nagyon sok felfedezését, elgondolását, így a nem euklideszi geometria felfedezése irányába tett gondolatait sem publikálta. Utolsó kívánsága az volt, hogy egyik korai és számára legkedvesebb felfedezésének, a 17 oldalú szabályos sokszög szerkesztésének emlékére sírkövére egy szabályos 17-szöget véssenek. Ezt ugyan nem teljesítették, de szülővárosában a tiszteletére emelt szobor talapzatán látható a szabályos 17 oldalú sokszög.

Matematikai munkásságáról:

Egyik legkedvesebb matematikai szakterülete a számelmélet volt. Tőle származik az a mondás, hogy: "A matematika a tudományok királynője, és a matematika királynője a számelmélet." 1791-ben, 14 éves korában becslést adott a prímszámok eloszlására, miszerint ezres számkörben a prímszámok száma fordítottan arányos a számok természetes alapú logaritmusával. Ezt ugyan később többen is pontosították, de ez semmit nem von le a fiatal Gauss érdemeiből.

Ő volt az, aki felfedezte, hogy kapcsolat van a prímszámok és a szabályos sokszögek szerkeszthetősége között. Egy "n" oldalszámú szabályos sokszög csak akkor szerkeszthető euklideszi szerkesztéssel ha "n" prímtényezős felbontásában csak a 2 szerepel tetszőleges nem negatív egész kitevőjű hatványon és az ún. Fermat-féle prímek (3,5,17,65537) első kitevőjű hatványon.

Azaz n = 2k * p1 * p2 *... * pk, ahol p1, p2, pk különböző Fermat-féle prímek. Tehát szerkeszthető a 3, 4, 5, 6, 8, 10, 12, 14, 15, 17, ..., 257 és 65537 oldalú szabályos sokszög, de nem szerkeszthető például a 7, 9 , ill. 11 oldalú. A 17 oldalú szabályos sokszög szerkesztésének a módját ő meg is oldotta.

Gauss foglalkozott a szakaszos tizedes törtekkel, és tisztázta, mikor kapunk tiszta vagy vegyes szakaszos tizedes törtet, és mekkora lehet a szakasz hosszúsága. 1799-ben a doktori értekezésében az "algebra alaptételét" igazolta, amely szerint minden algebrai egyenletnek van gyöke. Ezek a gyökök nem okvetlenül valósak, hanem lehetnek komplex számok is, és nem biztos, hogy ezek a gyökök mind különböznek egymástól. A gyökök száma (beleértve az azonosakat is) az egyenlet fokszámával egyenlő. 1827-ben jelent meg „A görbe felületekre vonatkozó általános vizsgálatok” című műve, amelynek eredményei geodéziai munkásságára vezethetők vissza. Gauss ötlete, hogy a komplex számokat a sík pontjaiként ábrázolhatjuk. 1837-ben megjelent értekezése a komplex számok algebráját és aritmetikáját tartalmazza. A nem euklideszi geometria megalkotásának területén végzett kutatásairól csak levelezéseiből tudunk, és feltételezhető, hogy ezen a területen is messzire jutott.

Melléklet

Models with an intercept (from Savin and White)

Durbin-Watson Statistic: 1 Per Cent Significance Points of dL and dU																				
	k'=1		k'=2		k'=3		k'=4		k'=5		k'=6		k'=7		k'=8		k'=9		k'=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	0.390	1.142	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.435	1.036	0.294	1.676	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.497	1.003	0.345	1.489	0.229	2.102	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.554	0.998	0.408	1.389	0.279	1.875	0.183	2.433	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.604	1.001	0.466	1.333	0.340	1.733	0.230	2.193	0.150	2.690	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.653	1.010	0.519	1.297	0.396	1.640	0.286	2.030	0.193	2.453	0.124	2.892	-----	-----	-----	-----	-----	-----	-----	-----
12	0.697	1.023	0.569	1.274	0.449	1.575	0.339	1.913	0.244	2.280	0.164	2.665	0.105	3.053	-----	-----	-----	-----	-----	-----
13	0.738	1.038	0.616	1.261	0.499	1.526	0.391	1.826	0.294	2.150	0.211	2.490	0.140	2.838	0.090	3.182	-----	-----	-----	-----
14	0.776	1.054	0.660	1.254	0.547	1.490	0.441	1.757	0.343	2.049	0.257	2.354	0.183	2.667	0.122	2.981	0.078	3.287	-----	-----
15	0.811	1.070	0.700	1.252	0.591	1.465	0.487	1.705	0.390	1.967	0.303	2.244	0.226	2.530	0.161	2.817	0.107	3.101	0.068	3.374
16	0.844	1.086	0.738	1.253	0.633	1.447	0.532	1.664	0.437	1.901	0.349	2.153	0.269	2.416	0.200	2.681	0.142	2.944	0.094	3.201
17	0.873	1.102	0.773	1.255	0.672	1.432	0.574	1.631	0.481	1.847	0.393	2.078	0.313	2.319	0.241	2.566	0.179	2.811	0.127	3.053
18	0.902	1.118	0.805	1.259	0.708	1.422	0.614	1.604	0.522	1.803	0.435	2.015	0.355	2.238	0.282	2.467	0.216	2.697	0.160	2.925
19	0.928	1.133	0.835	1.264	0.742	1.416	0.650	1.583	0.561	1.767	0.476	1.963	0.396	2.169	0.322	2.381	0.255	2.597	0.196	2.813
20	0.952	1.147	0.862	1.270	0.774	1.410	0.684	1.567	0.598	1.736	0.515	1.918	0.436	2.110	0.362	2.308	0.294	2.510	0.232	2.174
21	0.975	1.161	0.889	1.276	0.803	1.408	0.718	1.554	0.634	1.712	0.552	1.881	0.474	2.059	0.400	2.244	0.331	2.434	0.268	2.625
22	0.997	1.174	0.915	1.284	0.832	1.407	0.748	1.543	0.666	1.691	0.587	1.849	0.510	2.015	0.437	2.188	0.368	2.367	0.304	2.548
23	1.017	1.186	0.938	1.290	0.858	1.407	0.777	1.535	0.699	1.674	0.620	1.821	0.545	1.977	0.473	2.140	0.404	2.308	0.340	2.479
24	1.037	1.199	0.959	1.298	0.881	1.407	0.805	1.527	0.728	1.659	0.652	1.797	0.578	1.944	0.507	2.097	0.439	2.255	0.375	2.417
25	1.055	1.210	0.981	1.305	0.906	1.408	0.832	1.521	0.756	1.645	0.682	1.776	0.610	1.915	0.540	2.059	0.473	2.209	0.409	2.362
26	1.072	1.222	1.000	1.311	0.928	1.410	0.855	1.517	0.782	1.635	0.711	1.759	0.640	1.889	0.572	2.026	0.505	2.168	0.441	2.313
27	1.088	1.232	1.019	1.318	0.948	1.413	0.878	1.514	0.808	1.625	0.738	1.743	0.669	1.867	0.602	1.997	0.536	2.131	0.473	2.269
28	1.104	1.244	1.036	1.325	0.969	1.414	0.901	1.512	0.832	1.618	0.764	1.729	0.696	1.847	0.630	1.970	0.566	2.098	0.504	2.229
29	1.119	1.254	1.053	1.332	0.988	1.418	0.921	1.511	0.855	1.611	0.788	1.718	0.723	1.830	0.658	1.947	0.595	2.068	0.533	2.193
30	1.134	1.264	1.070	1.339	1.006	1.421	0.941	1.510	0.877	1.606	0.812	1.707	0.748	1.814	0.684	1.925	0.622	2.041	0.562	2.160
31	1.147	1.274	1.085	1.345	1.022	1.425	0.960	1.509	0.897	1.601	0.834	1.698	0.772	1.800	0.710	1.906	0.649	2.017	0.589	2.131
32	1.160	1.283	1.100	1.351	1.039	1.428	0.978	1.509	0.917	1.597	0.856	1.690	0.794	1.788	0.734	1.889	0.674	1.995	0.615	2.104
33	1.171	1.291	1.114	1.358	1.055	1.432	0.995	1.510	0.935	1.594	0.876	1.683	0.816	1.776	0.757	1.874	0.698	1.975	0.641	2.080
34	1.184	1.298	1.128	1.364	1.070	1.436	1.012	1.511	0.954	1.591	0.896	1.677	0.837	1.766	0.779	1.860	0.722	1.957	0.665	2.057
35	1.195	1.307	1.141	1.370	1.085	1.439	1.028	1.512	0.971	1.589	0.914	1.671	0.857	1.757	0.800	1.847	0.744	1.940	0.689	2.037

Durbin-Watson Statistic: 1 Per Cent Significance Points of dL and dU

	k'=1		k'=2		k'=3		k'=4		k'=5		k'=6		k'=7		k'=8		k'=9		k'=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
36	1.205	1.315	1.153	1.376	1.098	1.442	1.043	1.513	0.987	1.587	0.932	1.666	0.877	1.749	0.821	1.836	0.766	1.925	0.711	2.018
37	1.217	1.322	1.164	1.383	1.112	1.446	1.058	1.514	1.004	1.585	0.950	1.662	0.895	1.742	0.841	1.825	0.787	1.911	0.733	2.001
38	1.227	1.330	1.176	1.388	1.124	1.449	1.072	1.515	1.019	1.584	0.966	1.658	0.913	1.735	0.860	1.816	0.807	1.899	0.754	1.985
39	1.237	1.337	1.187	1.392	1.137	1.452	1.085	1.517	1.033	1.583	0.982	1.655	0.930	1.729	0.878	1.807	0.826	1.887	0.774	1.970
40	1.246	1.344	1.197	1.398	1.149	1.456	1.098	1.518	1.047	1.583	0.997	1.652	0.946	1.724	0.895	1.799	0.844	1.876	0.749	1.956
45	1.288	1.376	1.245	1.424	1.201	1.474	1.156	1.528	1.111	1.583	1.065	1.643	1.019	1.704	0.974	1.768	0.927	1.834	0.881	1.902
50	1.324	1.403	1.285	1.445	1.245	1.491	1.206	1.537	1.164	1.587	1.123	1.639	1.081	1.692	1.039	1.748	0.997	1.805	0.955	1.864
55	1.356	1.428	1.320	1.466	1.284	1.505	1.246	1.548	1.209	1.592	1.172	1.638	1.134	1.685	1.095	1.734	1.057	1.785	1.018	1.837
60	1.382	1.449	1.351	1.484	1.317	1.520	1.283	1.559	1.248	1.598	1.214	1.639	1.179	1.682	1.144	1.726	1.108	1.771	1.072	1.817
65	1.407	1.467	1.377	1.500	1.346	1.534	1.314	1.568	1.283	1.604	1.251	1.642	1.218	1.680	1.186	1.720	1.153	1.761	1.120	1.802
70	1.429	1.485	1.400	1.514	1.372	1.546	1.343	1.577	1.313	1.611	1.283	1.645	1.253	1.680	1.223	1.716	1.192	1.754	1.162	1.792
75	1.448	1.501	1.422	1.529	1.395	1.557	1.368	1.586	1.340	1.617	1.313	1.649	1.284	1.682	1.256	1.714	1.227	1.748	1.199	1.783
80	1.465	1.514	1.440	1.541	1.416	1.568	1.390	1.595	1.364	1.624	1.338	1.653	1.312	1.683	1.285	1.714	1.259	1.745	1.232	1.777
85	1.481	1.529	1.458	1.553	1.434	1.577	1.411	1.603	1.386	1.630	1.362	1.657	1.337	1.685	1.312	1.714	1.287	1.743	1.262	1.773
90	1.496	1.541	1.474	1.563	1.452	1.587	1.429	1.611	1.406	1.636	1.383	1.661	1.360	1.687	1.336	1.714	1.312	1.741	1.288	1.769
95	1.510	1.552	1.489	1.573	1.468	1.596	1.446	1.618	1.425	1.641	1.403	1.666	1.381	1.690	1.358	1.715	1.336	1.741	1.313	1.767
100	1.522	1.562	1.502	1.582	1.482	1.604	1.461	1.625	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717	1.357	1.741	1.335	1.765
150	1.611	1.637	1.598	1.651	1.584	1.665	1.571	1.679	1.557	1.693	1.543	1.708	1.530	1.722	1.515	1.737	1.501	1.752	1.486	1.767
200	1.664	1.684	1.653	1.693	1.643	1.704	1.633	1.715	1.623	1.725	1.613	1.735	1.603	1.746	1.592	1.757	1.582	1.768	1.571	1.779

	k'=11		k'=12		k'=13		k'=14		k'=15		k'=16		k'=17		k'=18		k'=19		k'=20	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
16	0.060	3.446	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
17	0.084	3.286	0.053	3.506	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
18	0.113	3.146	0.075	3.358	0.047	3.557	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
19	0.145	3.023	0.102	3.227	0.067	3.420	0.043	3.601	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
20	0.178	2.914	0.131	3.109	0.092	3.297	0.061	3.474	0.038	3.639	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
21	0.212	2.817	0.162	3.004	0.119	3.185	0.084	3.358	0.055	3.521	0.035	3.671	-----	-----	-----	-----	-----	-----	-----	-----
22	0.246	2.729	0.194	2.909	0.148	3.084	0.109	3.252	0.077	3.412	0.050	3.562	0.032	3.700	-----	-----	-----	-----	-----	-----
23	0.281	2.651	0.227	2.822	0.178	2.991	0.136	3.155	0.100	3.311	0.070	3.459	0.046	3.597	0.029	3.725	-----	-----	-----	-----
24	0.315	2.580	0.260	2.744	0.209	2.906	0.165	3.065	0.125	3.218	0.092	3.363	0.065	3.501	0.043	3.629	0.027	3.747	-----	-----
25	0.348	2.517	0.292	2.674	0.240	2.829	0.194	2.982	0.152	3.131	0.116	3.274	0.085	3.410	0.060	3.538	0.039	3.657	0.025	3.766
26	0.381	2.460	0.324	2.610	0.272	2.758	0.224	2.906	0.180	3.050	0.141	3.191	0.107	3.325	0.079	3.452	0.055	3.572	0.036	3.682

	k'=11		k'=12		k'=13		k'=14		k'=15		k'=16		k'=17		k'=18		k'=19		k'=20	
n	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**
27	0.413	2.409	0.356	2.552	0.303	2.694	0.253	2.836	0.208	2.976	0.167	3.113	0.131	3.245	0.100	3.371	0.073	3.490	0.051	3.602
28	0.444	2.363	0.387	2.499	0.333	2.635	0.283	2.772	0.237	2.907	0.194	3.040	0.156	3.169	0.122	3.294	0.093	3.412	0.068	3.524
29	0.474	2.321	0.417	2.451	0.363	2.582	0.313	2.713	0.266	2.843	0.222	2.972	0.182	3.098	0.146	3.220	0.114	3.338	0.087	3.450
30	0.503	2.283	0.447	2.407	0.393	2.533	0.342	2.659	0.294	2.785	0.249	2.909	0.208	3.032	0.171	3.152	0.137	3.267	0.107	3.379
31	0.531	2.248	0.475	2.367	0.422	2.487	0.371	2.609	0.322	2.730	0.277	2.851	0.234	2.970	0.193	3.087	0.160	3.201	0.128	3.311
32	0.558	2.216	0.503	2.330	0.450	2.446	0.399	2.563	0.350	2.680	0.304	2.797	0.261	2.912	0.221	3.026	0.184	3.137	0.151	3.246
33	0.585	2.187	0.530	2.296	0.477	2.408	0.426	2.520	0.377	2.633	0.331	2.746	0.287	2.858	0.246	2.969	0.209	3.078	0.174	3.184
34	0.610	2.160	0.556	2.266	0.503	2.373	0.452	2.481	0.404	2.590	0.357	2.699	0.313	2.808	0.272	2.915	0.233	3.022	0.197	3.126
35	0.634	2.136	0.581	2.237	0.529	2.340	0.478	2.444	0.430	2.550	0.383	2.655	0.339	2.761	0.297	2.865	0.257	2.969	0.221	3.071
36	0.658	2.113	0.605	2.210	0.554	2.310	0.504	2.410	0.455	2.512	0.409	2.614	0.364	2.717	0.322	2.818	0.282	2.919	0.244	3.019
37	0.680	2.092	0.628	2.186	0.578	2.282	0.528	2.379	0.480	2.477	0.434	2.576	0.389	2.675	0.347	2.774	0.306	2.872	0.268	2.969
38	0.702	2.073	0.651	2.164	0.601	2.256	0.552	2.350	0.504	2.445	0.458	2.540	0.414	2.637	0.371	2.733	0.330	2.828	0.291	2.923
39	0.723	2.055	0.673	2.143	0.623	2.232	0.575	2.323	0.528	2.414	0.482	2.507	0.438	2.600	0.395	2.694	0.354	2.787	0.315	2.879
40	0.744	2.039	0.694	2.123	0.645	2.210	0.597	2.297	0.551	2.386	0.505	2.476	0.461	2.566	0.418	2.657	0.377	2.748	0.338	2.838
45	0.835	1.972	0.790	2.044	0.744	2.118	0.700	2.193	0.655	2.269	0.612	2.346	0.570	2.424	0.528	2.503	0.488	2.582	0.448	2.661
50	0.913	1.925	0.871	1.987	0.829	2.051	0.787	2.116	0.746	2.182	0.705	2.250	0.665	2.318	0.625	2.387	0.586	2.456	0.548	2.526
55	0.979	1.891	0.940	1.945	0.902	2.002	0.863	2.059	0.825	2.117	0.786	2.176	0.748	2.237	0.711	2.298	0.674	2.359	0.637	2.421
60	1.037	1.865	1.001	1.914	0.965	1.964	0.929	2.015	0.893	2.067	0.857	2.120	0.822	2.173	0.786	2.227	0.751	2.283	0.716	2.338
65	1.087	1.845	1.053	1.889	1.020	1.934	0.986	1.980	0.953	2.027	0.919	2.075	0.886	2.123	0.852	2.172	0.819	2.221	0.789	2.272
70	1.131	1.831	1.099	1.870	1.068	1.911	1.037	1.953	1.005	1.995	0.974	2.038	0.943	2.082	0.911	2.127	0.880	2.172	0.849	2.217
75	1.170	1.819	1.141	1.856	1.111	1.893	1.082	1.931	1.052	1.970	1.023	2.009	0.993	2.049	0.964	2.090	0.934	2.131	0.905	2.172
80	1.205	1.810	1.177	1.844	1.150	1.878	1.122	1.913	1.094	1.949	1.066	1.984	1.039	2.022	1.011	2.059	0.983	2.097	0.955	2.135
85	1.236	1.803	1.210	1.834	1.184	1.866	1.158	1.898	1.132	1.931	1.106	1.965	1.080	1.999	1.053	2.033	1.027	2.068	1.000	2.104
90	1.264	1.798	1.240	1.827	1.215	1.856	1.191	1.886	1.166	1.917	1.141	1.948	1.116	1.979	1.091	2.012	1.066	2.044	1.041	2.077
95	1.290	1.793	1.267	1.821	1.244	1.848	1.221	1.876	1.197	1.905	1.174	1.943	1.150	1.963	1.126	1.993	1.102	2.023	1.079	2.054
100	1.314	1.790	1.292	1.816	1.270	1.841	1.248	1.868	1.225	1.895	1.203	1.922	1.181	1.949	1.158	1.977	1.136	2.006	1.113	2.034
150	1.473	1.783	1.458	1.799	1.444	1.814	1.429	1.830	1.414	1.847	1.400	1.863	1.385	1.880	1.370	1.897	1.355	1.913	1.340	1.931
200	1.561	1.791	1.550	1.801	1.539	1.813	1.528	1.824	1.518	1.836	1.507	1.847	1.495	1.860	1.484	1.871	1.474	1.883	1.462	1.896

Models with an intercept (from Savin and White)

Durbin-Watson Statistic: 5 Per Cent Significance Points of dL and dU																				
	k'=1		**k'=2**		**k'=3**		**k'=4**		**k'=5**		**k'=6**		**k'=7**		**k'=8**		**k'=9**		**k'=10**	
n	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**	**dL**	**dU**
6	0.610	1.400	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.700	1.356	0.467	1.896	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.763	1.332	0.559	1.777	0.367	2.287	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	-----	-----	-----	-----	-----	-----	-----	-----
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	-----	-----	-----	-----	-----	-----
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	-----	-----	-----	-----
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	-----	-----
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.258	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.681	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.796	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.079	1.891	1.015	1.978	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.876	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.197
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180

Durbin-Watson Statistic: 5 Per Cent Significance Points of dL and dU

	k'=1		k'=2		k'=3		k'=4		k'=5		k'=6		k'=7		k'=8		k'=9		k'=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.838	1.369	1.874	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.747	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.846	1.608	1.862	1.593	1.877
200	1.758	1.779	1.748	1.789	1.738	1.799	1.728	1.809	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

	k'=11		k'=12		k'=13		k'=14		k'=15		k'=16		k'=17		k'=18		k'=19		k'=20	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
16	0.098	3.503	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
17	0.138	3.378	0.087	3.557	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
18	0.177	3.265	0.123	3.441	0.078	3.603	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
19	0.220	3.159	0.160	3.335	0.111	3.496	0.070	3.642	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
20	0.263	3.063	0.200	3.234	0.145	3.395	0.100	3.542	0.063	3.676	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
21	0.307	2.976	0.240	3.141	0.182	3.300	0.132	3.448	0.091	3.583	0.058	3.705	-----	-----	-----	-----	-----	-----	-----	-----
22	0.349	2.897	0.281	3.057	0.220	3.211	0.166	3.358	0.120	3.495	0.083	3.619	0.052	3.731	-----	-----	-----	-----	-----	-----
23	0.391	2.826	0.322	2.979	0.259	3.128	0.202	3.272	0.153	3.409	0.110	3.535	0.076	3.650	0.048	3.753	-----	-----	-----	-----
24	0.431	2.761	0.362	2.908	0.297	3.053	0.239	3.193	0.186	3.327	0.141	3.454	0.101	3.572	0.070	3.678	0.044	3.773	-----	-----
25	0.470	2.702	0.400	2.844	0.335	2.983	0.275	3.119	0.221	3.251	0.172	3.376	0.130	3.494	0.094	3.604	0.065	3.702	0.041	3.790
26	0.508	2.649	0.438	2.784	0.373	2.919	0.312	3.051	0.256	3.179	0.205	3.303	0.160	3.420	0.120	3.531	0.087	3.632	0.060	3.724
27	0.544	2.600	0.475	2.730	0.409	2.859	0.348	2.987	0.291	3.112	0.238	3.233	0.191	3.349	0.149	3.460	0.112	3.563	0.081	3.658
28	0.578	2.555	0.510	2.680	0.445	2.805	0.383	2.928	0.325	3.050	0.271	3.168	0.222	3.283	0.178	3.392	0.138	3.495	0.104	3.592
29	0.612	2.515	0.544	2.634	0.479	2.755	0.418	2.874	0.359	2.992	0.305	3.107	0.254	3.219	0.208	3.327	0.166	3.431	0.129	3.528

	k'=11		k'=12		k'=13		k'=14		k'=15		k'=16		k'=17		k'=18		k'=19		k'=20	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
30	0.643	2.477	0.577	2.592	0.512	2.708	0.451	2.823	0.392	2.937	0.337	3.050	0.286	3.160	0.238	3.266	0.195	3.368	0.156	3.465
31	0.674	2.443	0.608	2.553	0.545	2.665	0.484	2.776	0.425	2.887	0.370	2.996	0.317	3.103	0.269	3.208	0.224	3.309	0.183	3.406
32	0.703	2.411	0.638	2.517	0.576	2.625	0.515	2.733	0.457	2.840	0.401	2.946	0.349	3.050	0.299	3.153	0.253	3.252	0.211	3.348
33	0.731	2.382	0.668	2.484	0.606	2.588	0.546	2.692	0.488	2.796	0.432	2.899	0.379	3.000	0.329	3.100	0.283	3.198	0.239	3.293
34	0.758	2.355	0.695	2.454	0.634	2.554	0.575	2.654	0.518	2.754	0.462	2.854	0.409	2.954	0.359	3.051	0.312	3.147	0.267	3.240
35	0.783	2.330	0.722	2.425	0.662	2.521	0.604	2.619	0.547	2.716	0.492	2.813	0.439	2.910	0.388	3.005	0.340	3.099	0.295	3.190
36	0.808	2.306	0.748	2.398	0.689	2.492	0.631	2.586	0.575	2.680	0.520	2.774	0.467	2.868	0.417	2.961	0.369	3.053	0.323	3.142
37	0.831	2.285	0.772	2.374	0.714	2.464	0.657	2.555	0.602	2.646	0.548	2.738	0.495	2.829	0.445	2.920	0.397	3.009	0.351	3.097
38	0.854	2.265	0.796	2.351	0.739	2.438	0.683	2.526	0.628	2.614	0.575	2.703	0.522	2.792	0.472	2.880	0.424	2.968	0.378	3.054
39	0.875	2.246	0.819	2.329	0.763	2.413	0.707	2.499	0.653	2.585	0.600	2.671	0.549	2.757	0.499	2.843	0.451	2.929	0.404	3.013
40	0.896	2.228	0.840	2.309	0.785	2.391	0.731	2.473	0.678	2.557	0.626	2.641	0.575	2.724	0.525	2.808	0.477	2.829	0.430	2.974
45	0.988	2.156	0.938	2.225	0.887	2.296	0.838	2.367	0.788	2.439	0.740	2.512	0.692	2.586	0.644	2.659	0.598	2.733	0.553	2.807
50	1.064	2.103	1.019	2.163	0.973	2.225	0.927	2.287	0.882	2.350	0.836	2.414	0.792	2.479	0.747	2.544	0.703	2.610	0.660	2.675
55	1.129	2.062	1.087	2.116	1.045	2.170	1.003	2.225	0.961	2.281	0.919	2.338	0.877	2.396	0.836	2.454	0.795	2.512	0.754	2.571
60	1.184	2.031	1.145	2.079	1.106	2.127	1.068	2.177	1.029	2.227	0.990	2.278	0.951	2.330	0.913	2.382	0.874	2.434	0.836	2.487
65	1.231	2.006	1.195	2.049	1.160	2.093	1.124	2.138	1.088	2.183	1.052	2.229	1.016	2.276	0.980	2.323	0.944	2.371	0.908	2.419
70	1.272	1.987	1.239	2.026	1.206	2.066	1.172	2.106	1.139	2.148	1.105	2.189	1.072	2.232	1.038	2.275	1.005	2.318	0.971	2.362
75	1.308	1.970	1.277	2.006	1.247	2.043	1.215	2.080	1.184	2.118	1.153	2.156	1.121	2.195	1.090	2.235	1.058	2.275	1.027	2.315
80	1.340	1.957	1.311	1.991	1.283	2.024	1.253	2.059	1.224	2.093	1.195	2.129	1.165	2.165	1.136	2.201	1.106	2.238	1.076	2.275
85	1.369	1.946	1.342	1.977	1.315	2.009	1.287	2.040	1.260	2.073	1.232	2.105	1.205	2.139	1.177	2.172	1.149	2.206	1.121	2.241
90	1.395	1.937	1.369	1.966	1.344	1.995	1.318	2.025	1.292	2.055	1.266	2.085	1.240	2.116	1.213	2.148	1.187	2.179	1.160	2.211
95	1.418	1.930	1.394	1.956	1.370	1.984	1.345	2.012	1.321	2.040	1.296	2.068	1.271	2.097	1.247	2.126	1.222	2.156	1.197	2.186
100	1.439	1.923	1.416	1.948	1.393	1.974	1.371	2.000	1.347	2.026	1.324	2.053	1.301	2.080	1.277	2.108	1.253	2.135	1.229	2.164
150	1.579	1.892	1.564	1.908	1.550	1.924	1.535	1.940	1.519	1.956	1.504	1.972	1.489	1.989	1.474	2.006	1.458	2.023	1.443	2.040
200	1.654	1.885	1.643	1.896	1.632	1.908	1.621	1.919	1.610	1.931	1.599	1.943	1.588	1.955	1.576	1.967	1.565	1.979	1.554	1.991

Models with no intercept (from Farebrother): Positive serial correlation

Durbin-Watson One Per Cent Minimal Bound																					
N	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14	K=15	K=16	K=17	K=18	K=19	K=20
2	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3	0.034	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4	0.127	0.022	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5	0.233	0.089	0.014	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Durbin-Watson One Per Cent Minimal Bound																					
N	**K=0**	**K=1**	**K=2**	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**	**K=10**	**K=11**	**K=12**	**K=13**	**K=14**	**K=15**	**K=16**	**K=17**	**K=18**	**K=19**	**K=20**
6	0.322	0.175	0.065	0.010	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.398	0.253	0.135	0.049	0.008	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.469	0.324	0.202	0.106	0.038	0.006	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.534	0.394	0.268	0.164	0.086	0.031	0.005	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.591	0.457	0.333	0.223	0.136	0.070	0.025	0.004	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.643	0.515	0.394	0.284	0.189	0.114	0.059	0.021	0.003	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
12	0.691	0.568	0.451	0.341	0.244	0.161	0.097	0.050	0.018	0.003	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
13	0.733	0.617	0.503	0.396	0.298	0.212	0.139	0.083	0.043	0.015	0.002	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----
14	0.773	0.662	0.552	0.448	0.350	0.262	0.185	0.121	0.072	0.037	0.013	0.002	0.000	-----	-----	-----	-----	-----	-----	-----	-----
15	0.809	0.703	0.598	0.496	0.400	0.311	0.232	0.163	0.107	0.063	0.032	0.011	0.002	0.000	-----	-----	-----	-----	-----	-----	-----
16	0.842	0.741	0.640	0.541	0.447	0.358	0.278	0.206	0.145	0.094	0.056	0.028	0.010	0.002	0.000	-----	-----	-----	-----	-----	-----
17	0.873	0.776	0.679	0.583	0.491	0.404	0.323	0.249	0.184	0.129	0.084	0.050	0.025	0.009	0.001	0.000	-----	-----	-----	-----	-----
18	0.901	0.808	0.715	0.623	0.533	0.447	0.366	0.292	0.225	0.166	0.116	0.075	0.044	0.023	0.008	0.001	0.000	-----	-----	-----	-----
19	0.928	0.839	0.749	0.660	0.572	0.488	0.408	0.333	0.265	0.204	0.150	0.105	0.068	0.040	0.020	0.007	0.001	0.000	-----	-----	-----
20	0.952	0.867	0.780	0.694	0.609	0.527	0.448	0.374	0.304	0.241	0.185	0.136	0.095	0.062	0.036	0.018	0.006	0.001	0.000	-----	-----
21	0.976	0.893	0.810	0.727	0.644	0.564	0.486	0.413	0.343	0.279	0.221	0.169	0.124	0.087	0.056	0.033	0.017	0.006	0.001	0.000	-----
22	0.997	0.918	0.838	0.757	0.677	0.599	0.523	0.450	0.381	0.316	0.257	0.203	0.155	0.114	0.079	0.051	0.030	0.015	0.005	0.001	0.000
23	1.018	0.942	0.864	0.786	0.709	0.632	0.558	0.486	0.417	0.352	0.292	0.237	0.187	0.143	0.104	0.073	0.047	0.027	0.014	0.005	0.001
24	1.037	0.964	0.889	0.813	0.738	0.664	0.591	0.520	0.452	0.387	0.327	0.270	0.219	0.172	0.131	0.096	0.067	0.043	0.025	0.013	0.004
25	1.056	0.984	0.912	0.839	0.766	0.693	0.622	0.553	0.486	0.421	0.361	0.304	0.251	0.203	0.160	0.122	0.089	0.062	0.040	0.023	0.012
26	1.073	1.004	0.934	0.863	0.792	0.722	0.652	0.584	0.518	0.454	0.394	0.336	0.283	0.233	0.189	0.148	0.113	0.083	0.057	0.037	0.022
27	1.089	1.023	0.955	0.886	0.817	0.749	0.681	0.614	0.549	0.486	0.426	0.368	0.314	0.264	0.218	0.176	0.138	0.105	0.077	0.053	0.034
28	1.105	1.040	0.974	0.908	0.841	0.774	0.708	0.643	0.579	0.517	0.457	0.400	0.345	0.294	0.247	0.204	0.164	0.129	0.098	0.071	0.050
29	1.120	1.057	0.993	0.929	0.864	0.798	0.734	0.670	0.607	0.546	0.487	0.430	0.376	0.324	0.276	0.232	0.191	0.154	0.120	0.091	0.067
30	1.134	1.073	1.011	0.948	0.885	0.822	0.759	0.696	0.635	0.574	0.516	0.460	0.405	0.354	0.305	0.260	0.217	0.179	0.144	0.113	0.086
31	1.147	1.088	1.028	0.967	0.905	0.844	0.782	0.721	0.661	0.602	0.544	0.488	0.434	0.383	0.334	0.288	0.244	0.205	0.168	0.135	0.106
32	1.160	1.103	1.044	0.985	0.925	0.865	0.805	0.745	0.686	0.628	0.571	0.516	0.462	0.411	0.362	0.315	0.271	0.230	0.193	0.158	0.127
33	1.173	1.117	1.060	1.002	0.944	0.885	0.826	0.768	0.710	0.653	0.597	0.542	0.489	0.438	0.389	0.342	0.298	0.256	0.218	0.182	0.149
34	1.185	1.130	1.075	1.018	0.961	0.904	0.847	0.790	0.733	0.677	0.622	0.568	0.516	0.465	0.416	0.369	0.324	0.282	0.243	0.206	0.172
35	1.196	1.143	1.089	1.034	0.978	0.923	0.867	0.811	0.755	0.700	0.646	0.593	0.541	0.491	0.442	0.395	0.350	0.308	0.268	0.230	0.195
36	1.207	1.155	1.102	1.049	0.995	0.940	0.886	0.831	0.777	0.723	0.669	0.617	0.566	0.516	0.467	0.421	0.376	0.333	0.292	0.254	0.218
37	1.217	1.167	1.116	1.063	1.010	0.957	0.904	0.850	0.797	0.744	0.692	0.640	0.590	0.540	0.492	0.446	0.401	0.358	0.317	0.278	0.241
38	1.228	1.178	1.128	1.077	1.026	0.974	0.921	0.869	0.817	0.765	0.713	0.663	0.613	0.564	0.516	0.470	0.425	0.382	0.341	0.302	0.265
39	1.237	1.189	1.140	1.090	1.040	0.989	0.938	0.887	0.836	0.785	0.734	0.684	0.635	0.587	0.540	0.494	0.449	0.406	0.365	0.325	0.288

Durbin-Watson One Per Cent Minimal Bound																					
N	**K=0**	**K=1**	**K=2**	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**	**K=10**	**K=11**	**K=12**	**K=13**	**K=14**	**K=15**	**K=16**	**K=17**	**K=18**	**K=19**	**K=20**
40	1.247	1.200	1.152	1.103	1.054	1.004	0.954	0.904	0.854	0.804	0.754	0.705	0.657	0.609	0.562	0.517	0.473	0.430	0.388	0.349	0.311
45	1.289	1.247	1.204	1.160	1.116	1.071	1.026	0.981	0.936	0.890	0.845	0.800	0.755	0.710	0.666	0.623	0.581	0.539	0.499	0.459	0.421
50	1.325	1.287	1.248	1.208	1.168	1.128	1.087	1.046	1.004	0.963	0.921	0.880	0.838	0.797	0.756	0.715	0.675	0.636	0.597	0.559	0.521
55	1.356	1.321	1.286	1.250	1.213	1.176	1.139	1.101	1.063	1.025	0.987	0.948	0.910	0.872	0.833	0.796	0.758	0.721	0.684	0.647	0.611
60	1.383	1.351	1.319	1.285	1.252	1.218	1.183	1.149	1.114	1.078	1.043	1.008	0.972	0.936	0.901	0.865	0.830	0.795	0.760	0.725	0.691
65	1.408	1.378	1.348	1.317	1.286	1.254	1.222	1.190	1.158	1.125	1.092	1.059	1.026	0.993	0.960	0.927	0.894	0.861	0.828	0.795	0.762
70	1.429	1.401	1.373	1.345	1.316	1.286	1.257	1.227	1.197	1.166	1.136	1.105	1.074	1.043	1.012	0.981	0.950	0.919	0.888	0.857	0.826
75	1.448	1.423	1.396	1.369	1.342	1.315	1.287	1.260	1.231	1.203	1.174	1.146	1.117	1.088	1.058	1.029	1.000	0.971	0.941	0.912	0.883
80	1.466	1.442	1.417	1.392	1.367	1.341	1.315	1.289	1.262	1.236	1.209	1.182	1.155	1.127	1.100	1.072	1.045	1.017	0.989	0.962	0.934
85	1.482	1.459	1.436	1.412	1.388	1.364	1.340	1.315	1.290	1.265	1.240	1.214	1.189	1.163	1.137	1.111	1.085	1.059	1.033	1.006	0.980
90	1.497	1.475	1.453	1.431	1.408	1.385	1.362	1.339	1.315	1.292	1.268	1.244	1.220	1.195	1.171	1.146	1.121	1.097	1.072	1.047	1.022
95	1.510	1.490	1.469	1.448	1.426	1.405	1.383	1.361	1.338	1.316	1.293	1.271	1.248	1.225	1.201	1.178	1.155	1.131	1.108	1.084	1.060
100	1.523	1.503	1.483	1.463	1.443	1.422	1.402	1.381	1.359	1.338	1.317	1.295	1.273	1.251	1.229	1.207	1.185	1.162	1.140	1.118	1.095
150	1.611	1.598	1.585	1.571	1.558	1.544	1.530	1.516	1.502	1.488	1.474	1.460	1.445	1.431	1.416	1.402	1.387	1.372	1.357	1.342	1.327
200	1.664	1.654	1.644	1.634	1.624	1.613	1.603	1.593	1.582	1.572	1.561	1.551	1.540	1.529	1.519	1.508	1.497	1.486	1.475	1.434	1.453

Models with no intercept (from Farebrother): Positive serial correlation

Durbin-Watson Five Per Cent Minimal Bound																					
N	**K=0**	**K=1**	**K=2**	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**	**K=10**	**K=11**	**K=12**	**K=13**	**K=14**	**K=15**	**K=16**	**K=17**	**K=18**	**K=19**	**K=20**
2	0.012	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3	0.168	0.006	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4	0.355	0.105	0.004	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5	0.478	0.248	0.070	0.002	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
6	0.584	0.358	0.180	0.050	0.002	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.677	0.462	0.275	0.136	0.037	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.754	0.556	0.371	0.217	0.106	0.029	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.820	0.635	0.460	0.303	0.175	0.085	0.023	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.877	0.706	0.539	0.385	0.251	0.143	0.069	0.019	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.927	0.768	0.610	0.460	0.326	0.211	0.120	0.058	0.016	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
12	0.972	0.823	0.674	0.530	0.397	0.279	0.180	0.101	0.049	0.013	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
13	1.012	0.872	0.731	0.593	0.464	0.345	0.241	0.154	0.087	0.042	0.011	0.001	-----	-----	-----	-----	-----	-----	-----	-----	-----
14	1.047	0.916	0.783	0.651	0.525	0.408	0.302	0.210	0.134	0.075	0.036	0.010	0.001	-----	-----	-----	-----	-----	-----	-----	-----
15	1.079	0.955	0.829	0.704	0.583	0.467	0.361	0.266	0.185	0.118	0.066	0.031	0.008	0.001	-----	-----	-----	-----	-----	-----	-----
16	1.109	0.992	0.872	0.752	0.635	0.523	0.418	0.322	0.237	0.164	0.104	0.058	0.028	0.007	0.000	-----	-----	-----	-----	-----	-----

Durbin-Watson Five Per Cent Minimal Bound																					
N	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14	K=15	K=16	K=17	K=18	K=19	K=20
17	1.136	1.024	0.911	0.797	0.684	0.575	0.472	0.376	0.288	0.211	0.146	0.093	0.052	0.025	0.007	0.000	-----	-----	-----	-----	-----
18	1.160	1.055	0.946	0.837	0.729	0.624	0.523	0.427	0.339	0.260	0.190	0.131	0.083	0.046	0.022	0.006	0.000	-----	-----	-----	-----
19	1.183	1.082	0.979	0.875	0.771	0.669	0.570	0.476	0.388	0.307	0.235	0.171	0.118	0.075	0.041	0.020	0.005	0.000	-----	-----	-----
20	1.204	1.108	1.010	0.910	0.810	0.711	0.615	0.523	0.436	0.354	0.280	0.213	0.156	0.107	0.067	0.037	0.018	0.005	0.000	-----	-----
21	1.224	1.132	1.038	0.942	0.846	0.751	0.657	0.567	0.481	0.400	0.324	0.256	0.195	0.142	0.097	0.061	0.034	0.016	0.004	0.000	-----
22	1.242	1.154	1.064	0.972	0.879	0.787	0.697	0.609	0.524	0.443	0.368	0.298	0.235	0.178	0.130	0.089	0.056	0.031	0.015	0.004	0.000
23	1.259	1.175	1.088	1.000	0.911	0.822	0.734	0.648	0.565	0.485	0.410	0.339	0.274	0.216	0.164	0.119	0.081	0.051	0.028	0.014	0.004
24	1.275	1.194	1.111	1.026	0.940	0.854	0.769	0.685	0.604	0.525	0.450	0.380	0.314	0.254	0.199	0.151	0.110	0.075	0.047	0.026	0.012
25	1.290	1.212	1.132	1.050	0.967	0.884	0.802	0.720	0.641	0.563	0.489	0.419	0.353	0.291	0.235	0.184	0.140	0.101	0.069	0.044	0.024
26	1.304	1.229	1.152	1.073	0.993	0.913	0.833	0.753	0.676	0.600	0.527	0.457	0.390	0.328	0.271	0.218	0.171	0.130	0.094	0.064	0.040
27	1.318	1.245	1.171	1.094	1.017	0.940	0.862	0.785	0.709	0.635	0.563	0.493	0.427	0.365	0.306	0.252	0.203	0.159	0.120	0.087	0.060
28	1.330	1.260	1.188	1.115	1.040	0.965	0.889	0.815	0.741	0.668	0.597	0.529	0.463	0.400	0.341	0.286	0.236	0.190	0.148	0.112	0.081
29	1.342	1.275	1.205	1.134	1.062	0.989	0.916	0.843	0.770	0.699	0.630	0.562	0.497	0.435	0.376	0.320	0.268	0.221	0.177	0.139	0.105
30	1.354	1.288	1.221	1.152	1.082	1.011	0.940	0.869	0.799	0.729	0.661	0.595	0.530	0.468	0.409	0.353	0.301	0.252	0.207	0.166	0.130
31	1.365	1.301	1.236	1.169	1.101	1.033	0.964	0.895	0.826	0.758	0.691	0.626	0.562	0.501	0.442	0.386	0.333	0.283	0.237	0.195	0.156
32	1.375	1.313	1.250	1.185	1.120	1.053	0.986	0.919	0.852	0.785	0.720	0.653	0.593	0.532	0.474	0.418	0.364	0.314	0.267	0.223	0.183
33	1.385	1.325	1.264	1.201	1.137	1.072	1.007	0.942	0.876	0.811	0.747	0.684	0.623	0.563	0.504	0.449	0.395	0.344	0.297	0.252	0.211
34	1.394	1.336	1.277	1.216	1.153	1.091	1.027	0.963	0.900	0.836	0.774	0.712	0.651	0.592	0.534	0.479	0.425	0.374	0.326	0.280	0.238
35	1.403	1.347	1.289	1.230	1.169	1.108	1.046	0.984	0.922	0.860	0.799	0.738	0.678	0.620	0.563	0.508	0.455	0.404	0.355	0.309	0.266
36	1.412	1.357	1.301	1.243	1.184	1.125	1.064	1.004	0.943	0.883	0.823	0.763	0.705	0.647	0.591	0.536	0.483	0.432	0.384	0.337	0.293
37	1.420	1.367	1.312	1.256	1.199	1.141	1.082	1.023	0.964	0.905	0.846	0.787	0.730	0.673	0.618	0.564	0.511	0.460	0.412	0.365	0.321
38	1.428	1.376	1.323	1.268	1.212	1.156	1.099	1.041	0.983	0.925	0.868	0.811	0.754	0.698	0.644	0.590	0.538	0.488	0.439	0.392	0.347
39	1.436	1.385	1.333	1.280	1.225	1.170	1.114	1.058	1.002	0.945	0.889	0.833	0.778	0.723	0.669	0.616	0.564	0.514	0.466	0.419	0.374
40	1.443	1.394	1.343	1.291	1.238	1.184	1.130	1.075	1.020	0.965	0.909	0.854	0.800	0.746	0.693	0.641	0.590	0.540	0.492	0.445	0.400
45	1.476	1.432	1.387	1.341	1.294	1.246	1.197	1.148	1.099	1.049	1.000	0.950	0.900	0.851	0.802	0.753	0.706	0.658	0.612	0.567	0.523
50	1.504	1.464	1.424	1.382	1.340	1.297	1.253	1.209	1.164	1.120	1.075	1.029	0.984	0.939	0.894	0.849	0.804	0.760	0.717	0.674	0.631
55	1.528	1.492	1.455	1.417	1.379	1.340	1.300	1.260	1.219	1.179	1.138	1.096	1.055	1.013	0.972	0.930	0.889	0.848	0.807	0.766	0.726
60	1.549	1.516	1.482	1.447	1.412	1.376	1.340	1.303	1.266	1.229	1.191	1.153	1.115	1.077	1.038	1.000	0.962	0.923	0.885	0.847	0.810
65	1.568	1.537	1.505	1.474	1.441	1.408	1.375	1.341	1.307	1.272	1.238	1.202	1.167	1.132	1.096	1.061	1.025	0.989	0.953	0.918	0.882
70	1.584	1.555	1.526	1.497	1.467	1.436	1.405	1.374	1.342	1.310	1.278	1.245	1.213	1.180	1.147	1.113	1.080	1.047	1.013	0.980	0.947
75	1.599	1.572	1.545	1.517	1.489	1.461	1.432	1.403	1.373	1.344	1.313	1.283	1.253	1.222	1.191	1.160	1.129	1.098	1.066	1.035	1.004
80	1.612	1.587	1.561	1.536	1.509	1.483	1.456	1.429	1.401	1.373	1.345	1.317	1.288	1.259	1.230	1.201	1.172	1.143	1.113	1.084	1.054
85	1.624	1.600	1.576	1.552	1.527	1.502	1.477	1.452	1.426	1.400	1.373	1.347	1.320	1.293	1.266	1.238	1.211	1.183	1.155	1.128	1.100
90	1.635	1.613	1.590	1.567	1.544	1.520	1.497	1.472	1.448	1.423	1.399	1.373	1.348	1.323	1.297	1.271	1.245	1.219	1.193	1.167	1.141

Durbin-Watson Five Per Cent Minimal Bound																					
N	**K=0**	**K=1**	**K=2**	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**	**K=10**	**K=11**	**K=12**	**K=13**	**K=14**	**K=15**	**K=16**	**K=17**	**K=18**	**K=19**	**K=20**
95	1.645	1.624	1.603	1.581	1.559	1.537	1.514	1.491	1.468	1.445	1.422	1.398	1.374	1.350	1.326	1.301	1.277	1.252	1.227	1.202	1.177
100	1.654	1.634	1.614	1.593	1.573	1.551	1.530	1.508	1.487	1.465	1.442	1.420	1.397	1.374	1.352	1.328	1.305	1.282	1.258	1.235	1.211
150	1.720	1.706	1.693	1.679	1.666	1.652	1.638	1.624	1.609	1.595	1.580	1.566	1.551	1.536	1.521	1.506	1.491	1.476	1.461	1.445	1.430
200	1.759	1.748	1.738	1.728	1.718	1.708	1.697	1.687	1.676	1.666	1.655	1.644	1.633	1.622	1.611	1.600	1.589	1.578	1.567	1.556	1.544

Models with no intercept (from Farebrother): Negative serial correlation

Durbin-Watson Ninety Five Per Cent Minimal Bound																					
N	**K=0**	**K=1**	**K=2**	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**	**K=10**	**K=11**	**K=12**	**K=13**	**K=14**	**K=15**	**K=16**	**K=17**	**K=18**	**K=19**	**K=20**
2	1.988	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3	2.761	0.994	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4	2.871	1.836	0.582	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5	2.857	2.178	1.267	0.380	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
6	2.844	2.320	1.655	0.917	0.266	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	2.828	2.398	1.871	1.283	0.690	0.197	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	2.805	2.453	2.008	1.521	1.017	0.537	0.151	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	2.783	2.483	2.110	1.687	1.251	0.823	0.429	0.120	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	2.762	2.501	2.181	1.816	1.427	1.044	0.678	0.350	0.097	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	2.742	2.511	2.231	1.913	1.569	1.218	0.881	0.567	0.291	0.080	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
12	2.723	2.516	2.268	1.987	1.682	1.364	1.049	0.752	0.481	0.245	0.068	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
13	2.705	2.518	2.296	2.044	1.771	1.484	1.193	0.911	0.649	0.413	0.210	0.058	-----	-----	-----	-----	-----	-----	-----	-----	-----
14	2.688	2.517	2.316	2.090	1.843	1.582	1.316	1.051	0.797	0.565	0.358	0.181	0.050	-----	-----	-----	-----	-----	-----	-----	-----
15	2.672	2.515	2.332	2.126	1.902	1.664	1.419	1.172	0.931	0.703	0.497	0.314	0.158	0.043	-----	-----	-----	-----	-----	-----	-----
16	2.657	2.512	2.344	2.155	1.950	1.732	1.506	1.276	1.049	0.829	0.624	0.439	0.277	0.139	0.038	-----	-----	-----	-----	-----	-----
17	2.644	2.508	2.353	2.179	1.990	1.789	1.580	1.367	1.153	0.944	0.743	0.557	0.391	0.246	0.124	0.034	-----	-----	-----	-----	-----
18	2.631	2.504	2.359	2.199	2.024	1.838	1.644	1.445	1.244	1.045	0.852	0.669	0.501	0.351	0.220	0.110	0.030	-----	-----	-----	-----
19	2.618	2.499	2.364	2.215	2.053	1.880	1.699	1.513	1.324	1.136	0.951	0.773	0.605	0.452	0.316	0.198	0.099	0.027	-----	-----	-----
20	2.607	2.494	2.368	2.228	2.077	1.916	1.747	1.573	1.395	1.216	1.040	0.868	0.704	0.550	0.410	0.286	0.179	0.090	0.025	-----	-----
21	2.596	2.489	2.370	2.239	2.098	1.947	1.789	1.625	1.457	1.289	1.120	0.955	0.796	0.644	0.502	0.373	0.260	0.162	0.081	0.022	-----
22	2.585	2.484	2.372	2.249	2.116	1.974	1.825	1.671	1.513	1.353	1.193	1.034	0.880	0.731	0.591	0.460	0.341	0.238	0.148	0.074	0.020
23	2.575	2.479	2.373	2.257	2.131	1.998	1.858	1.712	1.563	1.411	1.258	1.107	0.957	0.813	0.674	0.544	0.422	0.313	0.218	0.136	0.068
24	2.566	2.474	2.373	2.263	2.145	2.019	1.886	1.749	1.607	1.463	1.318	1.172	1.029	0.888	0.753	0.623	0.502	0.389	0.289	0.201	0.125
25	2.557	2.470	2.373	2.269	2.156	2.037	1.912	1.782	1.647	1.510	1.371	1.232	1.094	0.958	0.826	0.699	0.578	0.465	0.360	0.267	0.185
26	1.073	1.004	0.934	0.863	0.792	0.722	0.652	0.584	0.518	0.454	0.394	0.336	0.283	0.233	0.189	0.148	0.113	0.083	0.057	0.037	0.022
27	1.089	1.023	0.955	0.886	0.817	0.749	0.681	0.614	0.549	0.486	0.426	0.368	0.314	0.264	0.218	0.176	0.138	0.105	0.077	0.053	0.034

Durbin-Watson Ninety Five Per Cent Minimal Bound																					
N	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14	K=15	K=16	K=17	K=18	K=19	K=20
28	1.105	1.040	0.974	0.908	0.841	0.774	0.708	0.643	0.579	0.517	0.457	0.400	0.345	0.294	0.247	0.204	0.164	0.129	0.098	0.071	0.050
29	1.120	1.057	0.993	0.929	0.864	0.798	0.734	0.670	0.607	0.546	0.487	0.430	0.376	0.324	0.276	0.232	0.191	0.154	0.120	0.091	0.067
30	1.134	1.073	1.011	0.948	0.885	0.822	0.759	0.696	0.635	0.574	0.516	0.460	0.405	0.354	0.305	0.260	0.217	0.179	0.144	0.113	0.086
31	1.147	1.088	1.028	0.967	0.905	0.844	0.782	0.721	0.661	0.602	0.544	0.488	0.434	0.383	0.334	0.288	0.244	0.205	0.168	0.135	0.106
32	1.160	1.103	1.044	0.985	0.925	0.865	0.805	0.745	0.686	0.628	0.571	0.516	0.462	0.411	0.362	0.315	0.271	0.230	0.193	0.158	0.127
33	1.173	1.117	1.060	1.002	0.944	0.885	0.826	0.768	0.710	0.653	0.597	0.542	0.489	0.438	0.389	0.342	0.298	0.256	0.218	0.182	0.149
34	1.185	1.130	1.075	1.018	0.961	0.904	0.847	0.790	0.733	0.677	0.622	0.568	0.516	0.465	0.416	0.369	0.324	0.282	0.243	0.206	0.172
35	1.196	1.143	1.089	1.034	0.978	0.923	0.867	0.811	0.755	0.700	0.646	0.593	0.541	0.491	0.442	0.395	0.350	0.308	0.268	0.230	0.195
36	1.207	1.155	1.102	1.049	0.995	0.940	0.886	0.831	0.777	0.723	0.669	0.617	0.566	0.516	0.467	0.421	0.376	0.333	0.292	0.254	0.218
37	1.217	1.167	1.116	1.063	1.010	0.957	0.904	0.850	0.797	0.744	0.692	0.640	0.590	0.540	0.492	0.446	0.401	0.358	0.317	0.278	0.241
38	1.228	1.178	1.128	1.077	1.026	0.974	0.921	0.869	0.817	0.765	0.713	0.663	0.613	0.564	0.516	0.470	0.425	0.382	0.341	0.302	0.265
39	1.237	1.189	1.140	1.090	1.040	0.989	0.938	0.887	0.836	0.785	0.734	0.684	0.635	0.587	0.540	0.494	0.449	0.406	0.365	0.325	0.288
40	1.247	1.200	1.152	1.103	1.054	1.004	0.954	0.904	0.854	0.804	0.754	0.705	0.657	0.609	0.562	0.517	0.473	0.430	0.388	0.349	0.311
45	1.289	1.247	1.204	1.160	1.116	1.071	1.026	0.981	0.936	0.890	0.845	0.800	0.755	0.710	0.666	0.623	0.581	0.539	0.499	0.459	0.421
50	1.325	1.287	1.248	1.208	1.168	1.128	1.087	1.046	1.004	0.963	0.921	0.880	0.838	0.797	0.756	0.715	0.675	0.636	0.597	0.559	0.521
55	1.356	1.321	1.286	1.250	1.213	1.176	1.139	1.101	1.063	1.025	0.987	0.948	0.910	0.872	0.833	0.796	0.758	0.721	0.684	0.647	0.611
60	1.383	1.351	1.319	1.285	1.252	1.218	1.183	1.149	1.114	1.078	1.043	1.008	0.972	0.936	0.901	0.865	0.830	0.795	0.760	0.725	0.691
65	1.408	1.378	1.348	1.317	1.286	1.254	1.222	1.190	1.158	1.125	1.092	1.059	1.026	0.993	0.960	0.927	0.894	0.861	0.828	0.795	0.762
70	1.429	1.401	1.373	1.345	1.316	1.286	1.257	1.227	1.197	1.166	1.136	1.105	1.074	1.043	1.012	0.981	0.950	0.919	0.888	0.857	0.826
75	1.448	1.423	1.396	1.369	1.342	1.315	1.287	1.260	1.231	1.203	1.174	1.146	1.117	1.088	1.058	1.029	1.000	0.971	0.941	0.912	0.883
80	1.466	1.442	1.417	1.392	1.367	1.341	1.315	1.289	1.262	1.236	1.209	1.182	1.155	1.127	1.100	1.072	1.045	1.017	0.989	0.962	0.934
85	1.482	1.459	1.436	1.412	1.388	1.364	1.340	1.315	1.290	1.265	1.240	1.214	1.189	1.163	1.137	1.111	1.085	1.059	1.033	1.006	0.980
90	1.497	1.475	1.453	1.431	1.408	1.385	1.362	1.339	1.315	1.292	1.268	1.244	1.220	1.195	1.171	1.146	1.121	1.097	1.072	1.047	1.022
95	1.510	1.490	1.469	1.448	1.426	1.405	1.383	1.361	1.338	1.316	1.293	1.271	1.248	1.225	1.201	1.178	1.155	1.131	1.108	1.084	1.060
100	1.523	1.503	1.483	1.463	1.443	1.422	1.402	1.381	1.359	1.338	1.317	1.295	1.273	1.251	1.229	1.207	1.185	1.162	1.140	1.118	1.095
150	1.611	1.598	1.585	1.571	1.558	1.544	1.530	1.516	1.502	1.488	1.474	1.460	1.445	1.431	1.416	1.402	1.387	1.372	1.357	1.342	1.327
200	1.664	1.654	1.644	1.634	1.624	1.613	1.603	1.593	1.582	1.572	1.561	1.551	1.540	1.529	1.519	1.508	1.497	1.486	1.475	1.464	1.453

Models with no intercept (from Farebrother): Negative serial correlation

Durbin-Watson Ninety Nine Per Cent Minimal Bound																					
N	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14	K=15	K=16	K=17	K=18	K=19	K=20
2	1.999	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3	2.951	0.999	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4	3.221	1.967	0.586	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Durbin-Watson Ninety Nine Per Cent Minimal Bound																						
N	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14	K=15	K=16	K=17	K=18	K=19	K=20	K
5	3.261	2.462	1.359	0.382	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
6	3.235	2.682	1.878	0.983	0.268	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
7	3.198	2.776	2.177	1.459	0.740	0.198	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
8	3.166	2.817	2.347	1.776	1.158	0.576	0.153	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
9	3.133	2.837	2.448	1.983	1.465	0.937	0.460	0.121	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
10	3.101	2.847	2.514	2.121	1.684	1.224	0.773	0.375	0.098	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
11	3.071	2.847	2.560	2.220	1.842	1.441	1.035	0.647	0.312	0.081	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
12	3.043	2.843	2.592	2.294	1.961	1.607	1.244	0.885	0.549	0.263	0.069	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
13	3.017	2.836	2.612	2.349	2.054	1.737	1.410	1.082	0.764	0.471	0.225	0.059	-----	-----	-----	-----	-----	-----	-----	-----	-----	
14	2.992	2.828	2.626	2.391	2.127	1.842	1.544	1.244	0.948	0.666	0.409	0.195	0.051	-----	-----	-----	-----	-----	-----	-----	-----	
15	2.969	2.818	2.635	2.423	2.185	1.928	1.656	1.379	1.104	0.837	0.585	0.358	0.170	0.044	-----	-----	-----	-----	-----	-----	-----	
16	2.948	2.808	2.640	2.447	2.231	1.997	1.749	1.494	1.237	0.985	0.743	0.517	0.316	0.150	0.039	-----	-----	-----	-----	-----	-----	
17	2.927	2.797	2.643	2.466	2.269	2.055	1.827	1.591	1.351	1.114	0.883	0.664	0.461	0.281	0.133	0.035	-----	-----	-----	-----	-----	
18	2.908	2.787	2.644	2.480	2.299	2.102	1.893	1.675	1.451	1.227	1.007	0.796	0.597	0.413	0.251	0.119	0.031	-----	-----	-----	-----	
19	2.890	2.776	2.643	2.492	2.324	2.142	1.948	1.746	1.538	1.327	1.118	0.915	0.721	0.539	0.372	0.226	0.107	0.028	-----	-----	-----	
20	2.874	2.766	2.641	2.500	2.344	2.176	1.996	1.807	1.613	1.415	1.217	1.022	0.834	0.656	0.489	0.337	0.204	0.096	0.025	-----	-----	
21	2.858	2.756	2.638	2.506	2.361	2.204	2.036	1.861	1.678	1.492	1.305	1.119	0.937	0.763	0.598	0.446	0.307	0.185	0.087	0.023	-----	
22	2.842	2.746	2.635	2.511	2.375	2.228	2.071	1.907	1.736	1.561	1.384	1.207	1.032	0.862	0.700	0.548	0.408	0.280	0.169	0.080	0.021	
23	2.828	2.736	2.631	2.515	2.387	2.249	2.102	1.947	1.786	1.621	1.454	1.285	1.118	0.954	0.796	0.645	0.504	0.374	0.257	0.155	0.073	0
24	2.814	2.727	2.627	2.517	2.396	2.267	2.128	1.983	1.831	1.675	1.516	1.356	1.196	1.038	0.884	0.736	0.596	0.465	0.345	0.237	0.143	0
25	2.801	2.717	2.623	2.518	2.404	2.282	2.151	2.014	1.871	1.723	1.572	1.420	1.267	1.115	0.966	0.821	0.683	0.552	0.430	0.319	0.218	0
26	2.789	2.709	2.618	2.519	2.411	2.295	2.171	2.042	1.906	1.766	1.623	1.478	1.331	1.186	1.042	0.901	0.765	0.635	0.512	0.399	0.295	0
27	2.777	2.700	2.614	2.519	2.416	2.306	2.189	2.066	1.938	1.805	1.669	1.530	1.390	1.250	1.111	0.975	0.842	0.714	0.592	0.477	0.371	0
28	2.766	2.692	2.609	2.519	2.421	2.316	2.205	2.088	1.966	1.839	1.710	1.577	1.444	1.309	1.176	1.043	0.914	0.788	0.667	0.553	0.445	0
29	2.755	2.684	2.604	2.518	2.425	2.325	2.219	2.107	1.991	1.871	1.747	1.621	1.493	1.364	1.235	1.107	0.981	0.858	0.739	0.625	0.517	0
30	2.745	2.676	2.600	2.517	2.428	2.332	2.231	2.125	2.014	1.899	1.781	1.660	1.537	1.414	1.290	1.166	1.044	0.924	0.807	0.695	0.587	0
31	2.735	2.668	2.595	2.515	2.430	2.339	2.242	2.140	2.035	1.925	1.812	1.696	1.579	1.460	1.340	1.221	1.102	0.986	0.872	0.761	0.654	0
32	2.725	2.661	2.590	2.514	2.432	2.344	2.252	2.155	2.053	1.948	1.840	1.729	1.616	1.502	1.387	1.272	1.157	1.043	0.932	0.823	0.718	0
33	2.716	2.654	2.586	2.512	2.433	2.349	2.260	2.167	2.070	1.970	1.866	1.759	1.651	1.541	1.430	1.319	1.208	1.097	0.989	0.882	0.779	0
34	2.707	2.647	2.581	2.510	2.434	2.353	2.268	2.179	2.086	1.989	1.889	1.787	1.683	1.577	1.470	1.363	1.255	1.148	1.042	0.938	0.836	0
35	2.699	2.640	2.576	2.508	2.435	2.357	2.275	2.189	2.100	2.007	1.911	1.813	1.713	1.611	1.507	1.404	1.299	1.196	1.093	0.991	0.891	0
36	2.690	2.634	2.572	2.506	2.435	2.360	2.281	2.199	2.113	2.023	1.931	1.837	1.740	1.642	1.542	1.442	1.341	1.240	1.140	1.041	0.943	0
37	2.683	2.627	2.567	2.503	2.435	2.363	2.287	2.207	2.124	2.038	1.950	1.859	1.765	1.670	1.574	1.477	1.379	1.282	1.184	1.088	0.992	0
38	2.675	2.621	2.563	2.501	2.435	2.365	2.292	2.215	2.135	2.052	1.967	1.879	1.789	1.697	1.604	1.510	1.416	1.321	1.226	1.132	1.039	0

Durbin-Watson Ninety Nine Per Cent Minimal Bound

N	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14	K=15	K=16	K=17	K=18	K=19	K=20
39	2.667	2.615	2.559	2.499	2.435	2.367	2.296	2.222	2.145	2.065	1.982	1.898	1.811	1.722	1.632	1.541	1.450	1.358	1.266	1.174	1.083
40	2.660	2.609	2.555	2.496	2.434	2.369	2.300	2.229	2.154	2.077	1.997	1.915	1.831	1.746	1.659	1.570	1.482	1.392	1.303	1.213	1.124
45	2.628	2.583	2.535	2.484	2.430	2.374	2.315	2.253	2.190	2.124	2.056	1.986	1.914	1.841	1.767	1.691	1.614	1.537	1.459	1.381	1.302
50	2.600	2.559	2.516	2.471	2.424	2.374	2.323	2.269	2.214	2.157	2.098	2.037	1.975	1.911	1.847	1.781	1.714	1.646	1.578	1.509	1.439
55	2.575	2.538	2.500	2.459	2.417	2.373	2.327	2.280	2.231	2.180	2.128	2.075	2.020	1.964	1.907	1.849	1.790	1.730	1.669	1.608	1.546
60	2.553	2.519	2.484	2.448	2.409	2.370	2.329	2.286	2.242	2.197	2.151	2.103	2.054	2.004	1.954	1.902	1.849	1.796	1.742	1.687	1.631
65	2.534	2.503	2.470	2.437	2.402	2.366	2.329	2.290	2.250	2.210	2.168	2.125	2.081	2.036	1.990	1.944	1.896	1.848	1.799	1.750	1.700
70	2.516	2.487	2.458	2.427	2.395	2.361	2.327	2.292	2.256	2.219	2.181	2.142	2.102	2.061	2.020	1.977	1.934	1.891	1.846	1.802	1.756
75	2.500	2.473	2.446	2.417	2.387	2.357	2.325	2.293	2.260	2.226	2.191	2.155	2.118	2.081	2.043	2.005	1.965	1.926	1.885	1.844	1.802
80	2.486	2.461	2.436	2.408	2.380	2.352	2.323	2.293	2.262	2.231	2.198	2.165	2.132	2.098	2.063	2.027	1.991	1.954	1.917	1.879	1.841
85	2.473	2.449	2.425	2.399	2.374	2.347	2.320	2.292	2.263	2.234	2.204	2.174	2.143	2.111	2.079	2.046	2.012	1.979	1.944	1.909	1.874
90	2.460	2.438	2.415	2.391	2.367	2.342	2.317	2.291	2.264	2.237	2.209	2.181	2.152	2.122	2.092	2.061	2.030	1.999	1.967	1.935	1.902
95	2.449	2.428	2.406	2.384	2.361	2.338	2.314	2.289	2.264	2.239	2.212	2.186	2.159	2.131	2.103	2.075	2.046	2.016	1.986	1.956	1.926
100	2.438	2.418	2.398	2.377	2.355	2.333	2.310	2.287	2.264	2.240	2.215	2.190	2.165	2.139	2.113	2.086	2.059	2.031	2.003	1.975	1.946
150	2.363	2.349	2.336	2.322	2.308	2.294	2.279	2.265	2.250	2.235	2.220	2.204	2.188	2.173	2.156	2.140	2.124	2.107	2.090	2.073	2.056
200	2.317	2.307	2.296	2.286	2.276	2.265	2.255	2.244	2.233	2.222	2.211	2.200	2.189	2.177	2.166	2.154	2.142	2.131	2.119	2.106	2.094