

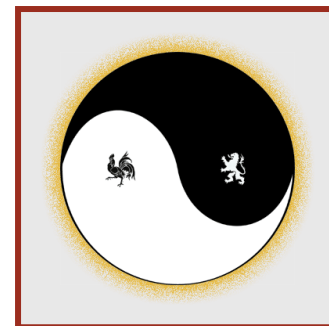
Korreláció, Lineáris regresszió

Dr. Sinkovits György és Dr. Prohászka Zoltán

Semmelweis Egyetem, Belgyógyászati és Hematológiai Klinika

https://semmelweis.hu/kutlab/oktatas/klinikai_biostatiztika_2023

2023.02.16.



I. rész
Elméleti alapok áttekintése

Két változó közötti kapcsolat

A változók típusa szerint

❖ Diszkrét változók:

- ❖ nem és SLE kialakulása
- **Khi-négyzet próba**

❖ Folytonos változók:

- **Korreláció és lineáris regresszió**



Két változó közötti kapcsolat

A kapcsolat jellege szerint

- ❖ **Függvényszerű (determinisztikus) kapcsolat**
 - ❖ az egyik változó és a kapcsolat egyértelműen meghatározza a másik változót
 - ❖ pl.: Celsius fok és Fahrenheit átváltás
 - ❖ **Stochasztikus kapcsolat**
 - ❖ a függvényszerű kapcsolaton kívül még egyéb tényezőktől, pl. a véletlentől is függ
 - ❖ pl.: koleszterin és triglicerid szintje
 - ❖ **Függetlenség**
 - ❖ nincs kapcsolat
 - ❖ pl.: koleszterinszint és az irányítószám
-



Két változó közötti kapcsolat

Megválaszolható kérdések

➤ példák

- ❖ **Összefügg-e** az életkor és az éhomi vércukorszint?
- ❖ **Milyen szoros az összefüggés** az éhomi vércukor és a HgbA1c szint között?
- ❖ **Hogyan függ** a betegek éhomi vércukorszintje a BMI-jüktől?



Két változó közötti kapcsolat

Korreláció és regresszió

- ❖ A két változó közötti kapcsolat léte és erőssége
 - **Korreláció**

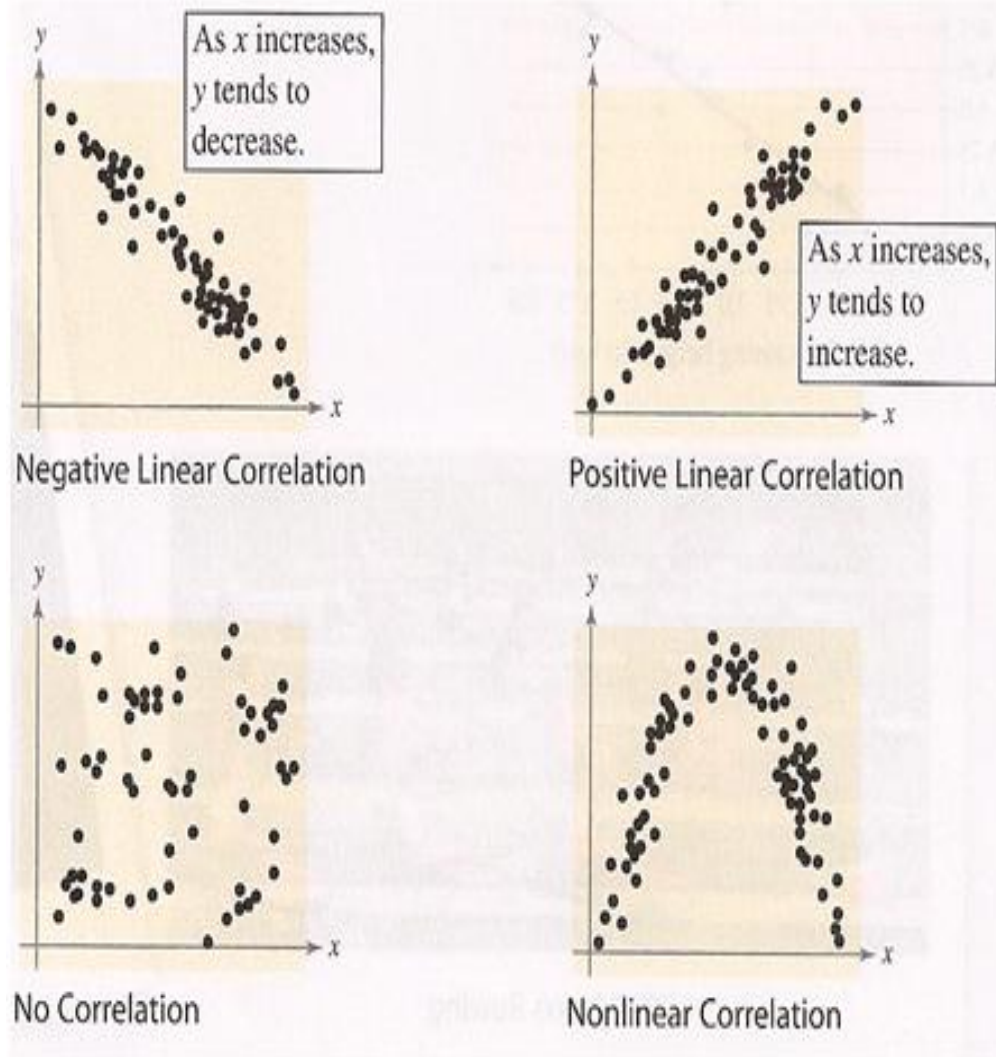
- ❖ A kapcsolat minőségének jellemzése, a kapcsolatban lévő tendenciák kifejezése függvények formájában
 - **Regresszióanalízis**



Korreláció

Ábrázolás

- ❖ Az összetartozó (x, y) értékpárok ábrázolása pontként
- ❖ Ha létezik egy **képzeltbeli egyenes**, amely mentén helyezkednek el a pontok
 - **lineáris korreláció**
- ❖ Az összefüggés **irányától** függően **pozitív vagy negatív**
- ❖ Ha a pontok **görbét** írnak le
 - **nemlineáris korreláció**
- ❖ Ha nincs ilyen vonal, a változók **korrelálatlanok** (de nem feltétlenül függetlenek!)



Korreláció

Jellemzői

❖ Skálafüggetlen

- azonos számmal szorozva vagy azonos számot hozzáadva nem változik, pl. független a mértékegységtől

❖ Szimmetrikus

- x korrelációja y -nal = y korrelációja x -szel

❖ Az összefüggés **nem** egyenlő az **okási** kapcsolattal

- ❖ pl. fagyifogyasztás és szívinfarktus

- az okási kapcsolatot logikai vagy kísérleti úton bizonyítani kell! (Hill kritériumok)

❖ A **lineáris** korreláció a lineáris összefüggést méri, nem az összefüggést általában



Lineáris korreláció

A korrelációs együttható

- A korreláció szorosságát a korrelációs együtthatóval (r) számszerűsíthetjük.
- ❖ Az r értéke -1 és 1 közötti:
 - ❖ minél szorosabb a kapcsolat, az $|r|$ annál közelebb van 1 -hez (azonos esetszám!)
 - ❖ ha előjele $+$: = pozitív korreláció, egyenes arányosság,
 - ❖ ha előjele $-$: negatív korreláció, fordított arányosság
 - ❖ ha -1 vagy 1 : függvényszerű, lineáris, determinisztikus a kapcsolat (minden pont azonos egyenesen)
 - ❖ ha 0 : korrelálatlanság (nem feltétlenül függetlenség!)



Lineáris korreláció

A korrelációs együttható szignifikanciája

- ❖ Kíváncsiak vagyunk két valószínűségi változó (pl. a testsúly és a magasság) korrelációjára a **populációban**
 - ezt az **elméleti korrelációs együttható** írja le: ρ – rho
 - Ennek becslésére:
 - ❖ véletlen mintavétellel mintát veszünk
 - ❖ meghatározzuk a mintában a **korrelációs együtthatót** (r),
 - ❖ ebből becsüljük a ρ -t
 - ❖ meghatározzuk a becslés hibáját, konfidencia-intervallumát
 - ❖ a becsült r és a hiba alapján számolom a szignifikanciát
-



Lineáris korreláció

A korrelációs együttható szignifikanciája

- ❖ H_0 : a korrelációs együttható a populációban 0
- ❖ H_1 : a $\rho \neq 0$

- ❖ N-2 szabadságfokú t-statisztika felhasználása
- ❖ Döntés a p-érték szerint
 - ha p kisebb, mint a szignifikancia-küszöb, elvetjük a H_0 -t, vagyis $\rho \neq 0$



Lineáris korreláció

A korrelációs együttható „erőssége”

Korrelációs koefficiens (r)	A kapcsolat erőssége
0,00-0,25	Nincs vagy igen gyenge
0,25-0,50	Gyenge
0,50-0,75	Mérsékelten erős
0,75-1,00	Igen erős

- ❖ Természettudományos kutatásokban gyakran ha az $r > 0,7$, jelentős összefüggésnek vélelményezik, "clinically meaningful"
- ❖ Biológiai kapcsolatoknál a $r > 0,95$ „gyanús”. Biztos nem áll fent determináltság=egyiket a másiktól számítom? Pl.: MCV, Hct



Lineáris korreláció

A determinációs együttható

- ❖ Azt fejezi ki, hogy az x változó az y változó varianciájának hány százalékát magyarázza
- ❖ Az r négyzetével (r^2) egyenlő
- pl. ha $r=0,6$ (mérsékelten erős korreláció), akkor $r^2=0,36$, vagyis egyik változó a másik varianciájának 36%-át magyarázza, a variancia további 64%-át más tényezők (további változók, mérési hiba) eredményezik

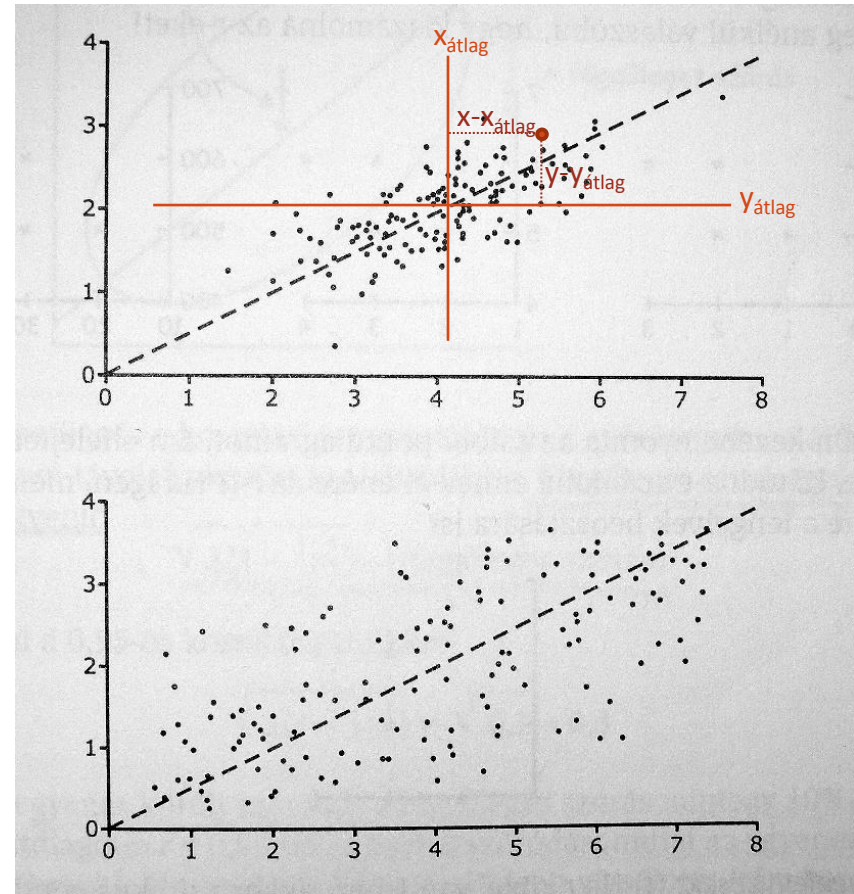


Lineáris korreláció

A számítás elve

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1) s_x s_y},$$

- ❖ Felsővonásos betűk – tapasztalati várható érték
- ❖ s_x, s_y – tapasztalati korrigált szórásnégyzet
- ❖ az $x - x_{\text{átlag}}$ értékeket az $x_{\text{átlag}}$ szórásával, az $y - y_{\text{átlag}}$ értékeket az $y_{\text{átlag}}$ szórásával osztjuk
- ❖ a standardizált értékeket pontonként összeszorozzuk
- ❖ a szorzatok összegét osztjuk a szabadsági fokkal



Lineáris korreláció

A számítás elvégzésének feltételei

➤ Általános

- ❖ véletlenszerű mintavétel
- ❖ változópárok (minden esethez 1-1 érték)
- ❖ független megfigyelések
- ❖ két mintából származó méréseket nem szabad kombinálni

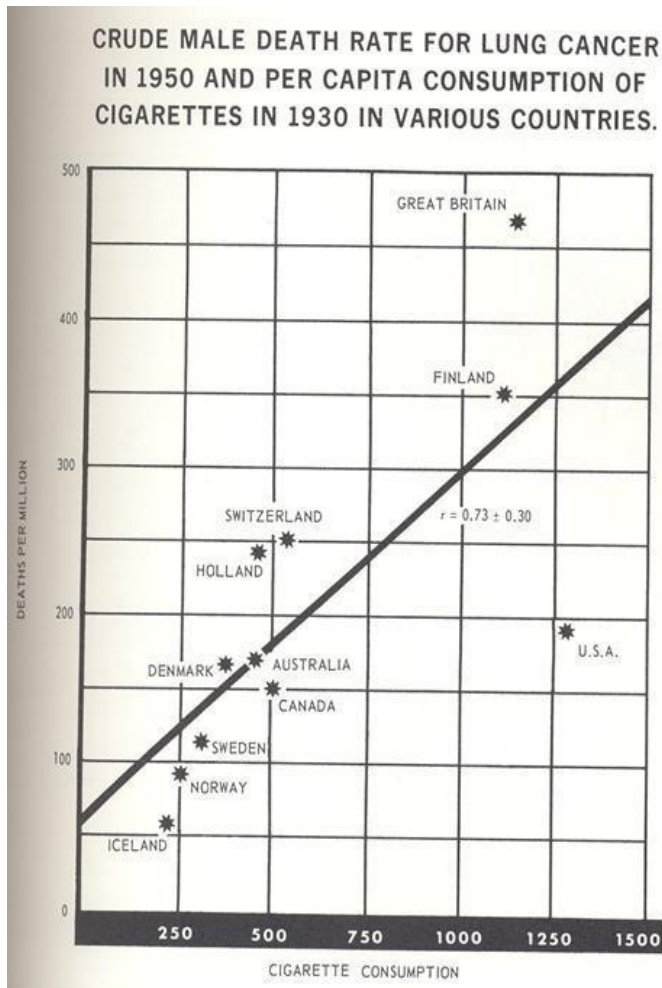
➤ Lineáris korrelációra specifikus

- ❖ változók folytonosak
 - ❖ változók normál eloszlásúak
 - ❖ nincsenek kiugró értékek
 - ❖ változók közötti kapcsolat lineáris
-



Lineáris korreláció

A számítás elvégzésének feltételei – példa

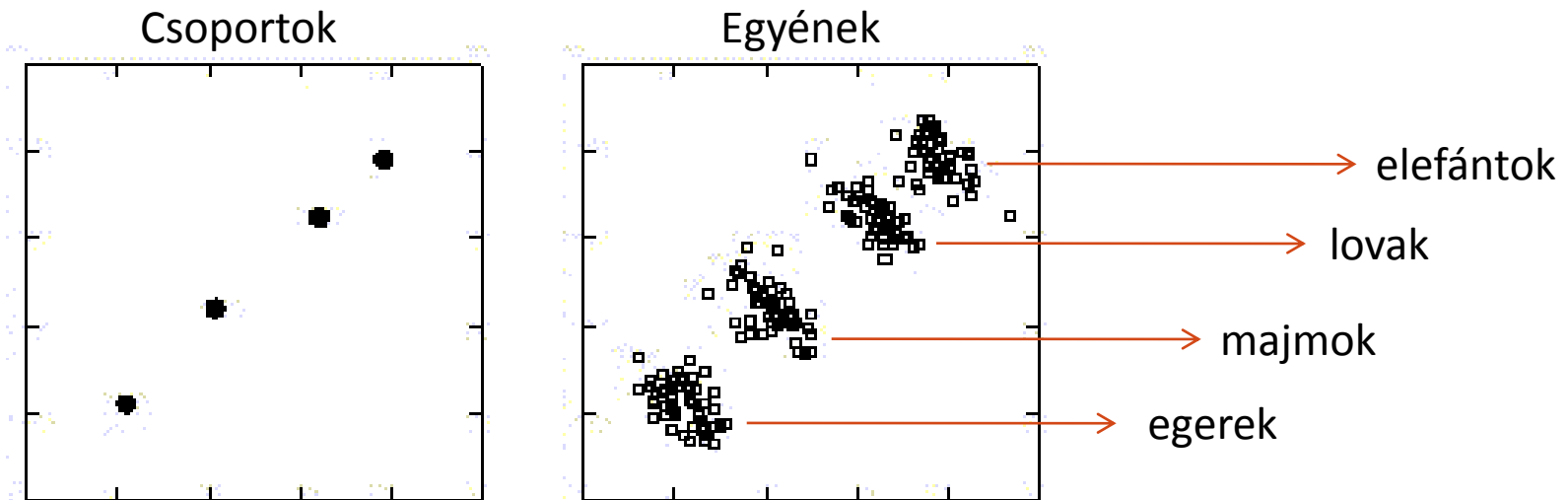


- R. Doll, „Etiology of lung cancer”, *Advances in Cancer research vol. 3 (1955) 1-50*
- Vizsgálat a dohányzás és a tüdőrák kapcsolatára
- Pontdiagram – egy főre jutó cigarettafogyasztás és a tüdőrák miatti elhalálozások arányszáma 11 országban
- A korreláció 0,73 volt a 11 pontra
- Országok betegszenek meg, vagy emberek?

Lineáris korreláció

A számítás elvégzésének feltételei – példa kicsit másképp

➤ Ökológiai korreláció: fül átmérője vs. élettartam



- ❖ Kerülendő! Sohasem szabad két populációból származó mintát keverni egy analízisben!
- ❖ Nem az egyének adatait használja, hanem csoportok átlagain vagy arányszámokon alapul. Jellemzően eltúlozza az összefüggés erősségét.



Outlierek

Az outlierek észlelése/kimutatása

- ❖ Nincs egyértelmű definíció → Szubjektív
 - általában 2-3 SD-n (vagy ferdült eloszlás esetén a medián-kvartilis távolság 2-3-szorosán) kívül eső értékek
- ❖ Számításos módszer:
 - ❖ értékek standardizálása után a 2-3-nál nagyobb abszolút értékkel rendelkező értékek (normál eloszlás esetén)
- ❖ Grafikus módszer:
 - ❖ box-and-whiskers plot alapján (részletek később)



Outlierek

Lehetséges okok tisztázása

❖ Téves adatrögzítés

- helyes adat kikeresése, vagy korrigálás, ha értelemszerű (pl. 167 m-es testmagasság esetén)

❖ Mérési hiba

- mérés megisméltése, ha lehetséges (pl. 0-s vércukorszint)

❖ Valós kiugró érték

- egyedi mérlegelés
-



Outlierek

Kiugró értékek kezelése

- ❖ Jelentőségük lehet, ezt **egyedileg** ellenőrizni/mérlegelni kell
 - automatikus kizárásuk nem javasolható
- Lehetséges megoldások:
 - ❖ kizárás
 - ❖ helyettesítés a legközelebbi, nem kiugró értékkel
 - ❖ helyettesítés átlaggal
 - ❖ nemparáméteres teszt végzése (kevésbé érzékeny)
 - ❖ paraméteres teszt elvégzése a kiugró érték ellenére
 - ❖ paraméteres teszt végzése a kiugró értékkel és anélkül is



Normalitás

Az eloszlás normalitásának vizsgálata

❖ Statisztikai tesztek:

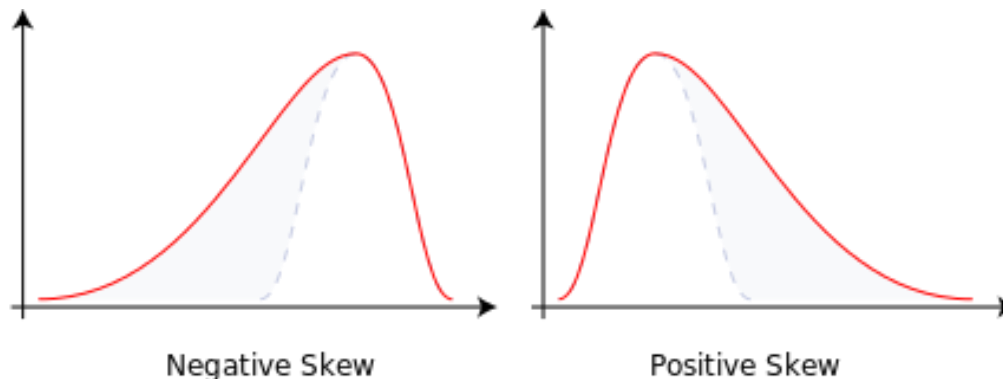
❖ Shapiro-Wilk teszt és Kolmogorov-Smirnov teszt

➤ ha szignifikáns a teszt eredménye, az eloszlás **nem** normáeloszlás

❖ Eloszlásfüggvény megtekintése:

❖ Milyen irányban ferdült?

❖ Mennyire ferdült?



Linearitás, normalitás

Teendők nem lineáris összefüggés esetén

- ❖ Ha nem monoton (nem azonos irányú) az összefüggés, nincs értelme a korrelációnak. Ha monoton:
 - Lehetséges megoldások:
 - ❖ Transzformációval normál eloszlásúvá tenni
 - ❖ Nemparaméteres teszt végzése
 - ❖ Paraméteres teszt elvégzése ennek ellenére
 - ❖ Paraméteres és nemparaméteres teszt elvégzése, eredményeinek összevetése



Transzformáció

Módszerek az eloszlás normalizálására

- a **ferdültség iránya és mértéke** alapján javasolt képletek

	Pozitív	Negatív
Gyenge	$y=\sqrt{x}$	$y=\sqrt{x_{\max}+1-x}$
Közepesen erős / Erős	$y=\log(x)$	$y=\log(x_{\max}+1-x)$
Közepesen erős / Erős	$y=1/x$	$y=1/(x_{\max}+1-x)$

- a logaritmizálás a legtöbb esetben segít, viszonylag könnyű értelmezni



Transzformáció

Hátrányok

- ❖ A transzformáció sem garantálja, hogy az új változó eloszlása normáleloszlás lesz
- ❖ Az eredmények nehezen értelmezhetőek:
 - Ha a fehérvérsejtszám reciproka korrelál a CRP logaritmusával, akkor mi mondható el a fehérvérsejtszám és a CRP kapcsolatáról?



Nemparaméteres (rang-) korreláció

A számítás elve

- ❖ értékeket sorbarendezzük és ún. rangszámot rendelünk hozzá
 - azonos értékek esetén a rangszámok átlagát adjuk rangszámként

Adatpárok sorszama	Változó 1 értéke	Változó 1 rangsáma	Változó 2 értéke	Változó 2 rangsáma	Rangszámok különbsége
1	1800	7	19	7	0
2	550	2,5	11	3	0,5
3	780	4	16	5,5	1,5
4	550	2,5	9	2	0,5
5	1120	6	16	5,5	0,5
6	960	5	14	4	1
7	200	1	6	1	0



Nemparaméteres (rang-) korreláció

A számítás elvégzésének feltételei

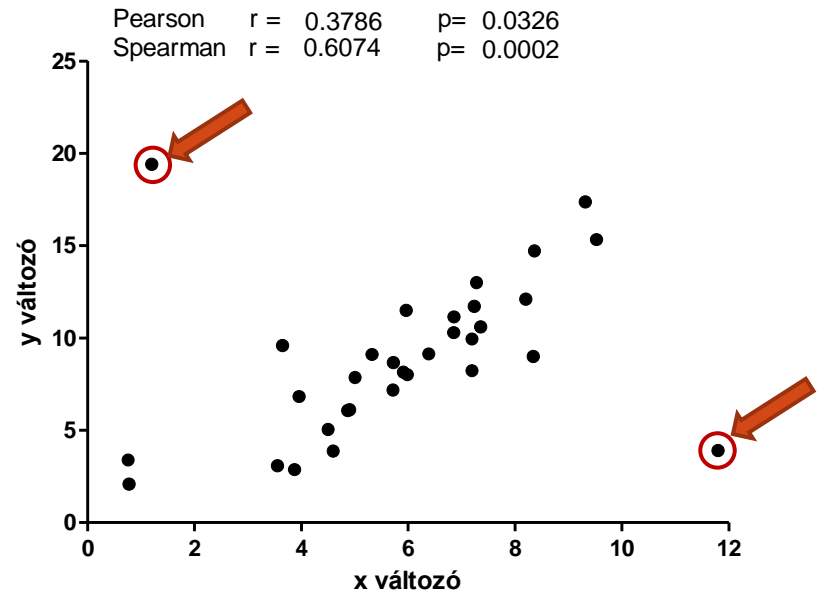
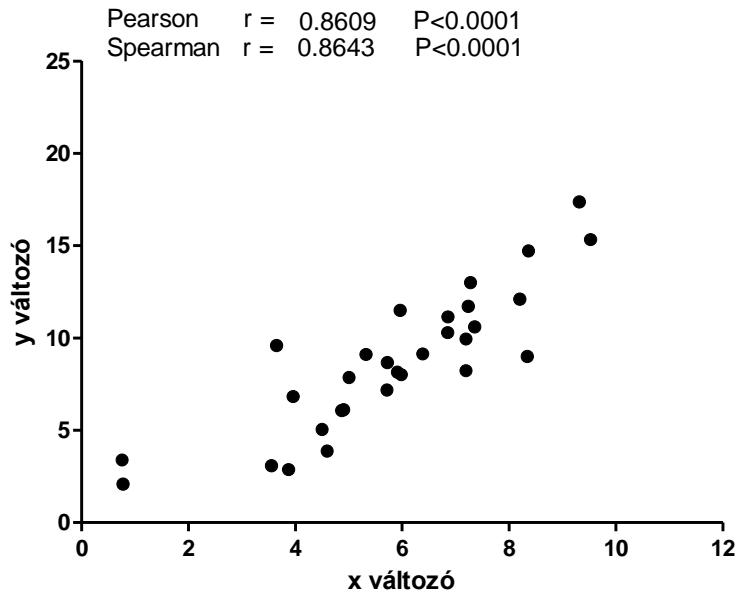
- Általános (azonos)
 - ❖ véletlenszerű mintavétel
 - ❖ változópárok (minden esethez 1-1 érték)
 - ❖ független megfigyelések
 - ❖ két mintából származó méréseket nem szabad kombinálni
 - Nemparaméteres korrelációra specifikus
 - ❖ változók folytonosak vagy ordinálisak (nem csak folytonos)
 - ❖ változók közötti kapcsolat **monoton** (nem kell lineáris)
 - kiugró értékekre és az eloszlás normalitására kevésbé érzékeny
 - (cserébe kevésbé szenzitív)
-



Nemparaméteres (rang-) korreláció

Érzékenység

- outlierok hatása a paraméteres és nemparaméteres tesztek eredményére



Lineáris regresszió

Jelentése, jellemzői

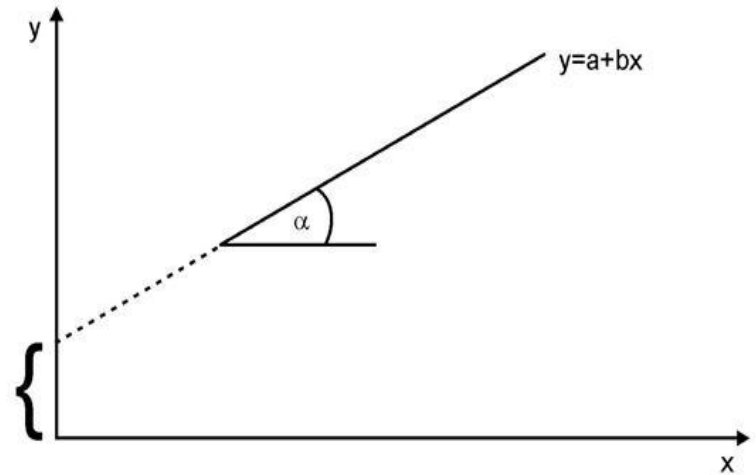
- A változók közötti kapcsolatot egy **egyenest leíró** függvénnyel fejezzük ki
- ❖ Egy kitüntetett változót (függő változó) a független változó(k) függvényében fejezzük ki
 - Nem szimmetrikus! Nem felcserélhető!
- ❖ Független változóból több is lehet
 - **többváltozós lineáris regresszió** (következő előadás)
- ❖ Az egyik változó értékéből megjósolhatjuk a másik értékét
 - a megfigyelési tartományban
 - ez természetesen nem lesz pontos (véletlen ingadozás)



Lineáris regresszió

Regressziós egyenes

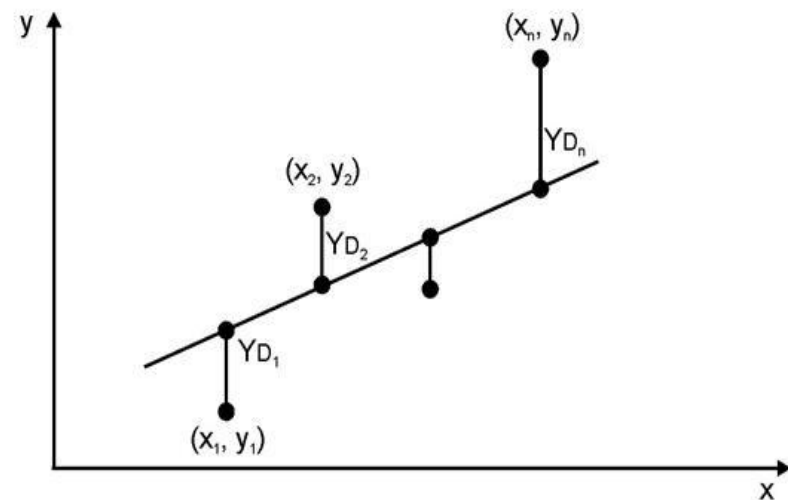
- A változók közötti kapcsolatot egy **egyenest leíró** függvénnyel fejezzük ki
- A pontokhoz legjobban illeszkedő egyenes az ún. **regressziós egyenes**
- A függő változót leíró képlet:
$$Y_i = \beta_0 + \beta_1 x_1 (+ \dots + \beta_n x_n) + \varepsilon$$
 - **y** - függő változó
(amit magyarázok, amit „feltárok”)
 - **x** - független változó(k)
(amivel magyarázok)
 - **b₀** - y tengelymetszete
 - **b_n** - meredekség (tangens α)
 - ε - reziduális, vagy illeszkedési hiba



Lineáris regresszió

Legkisebb négyzetek módszere

- Az a legjobban illeszkedő egyenes, amelyiknél a megfigyelések egyenestől mért négyzetes távolságösszege a lehető legkisebb
- Ezek a távolságok a reziduálisok (ε). Minél szorosabb az összefüggés, annál kisebbek
- Az egyenes együtthatóiból standard hiba határozható meg, ebből számítható a konfidencia intervallum



Lineáris regresszió

A számítás elvégzésének feltételei

➤ Általános

- ❖ véletlenszerű mintavétel
- ❖ változópárok (minden esethez 1-1 érték)
- ❖ független megfigyelések
- ❖ két mintából származó méréseket nem szabad kombinálni

➤ Lineáris regresszióra specifikus

- ❖ a függő és független változók folytonosak, normál eloszlásúak
 - ❖ nincsenek kiugró értékek
 - ❖ a függő és független változók közötti kapcsolat (közel) lineáris
 - ❖ a reziduálisok normál eloszlásúak
 - ❖ a reziduálisok mértéke ne függjön a független változók mértékétől (homoszkedaszticitás)
-



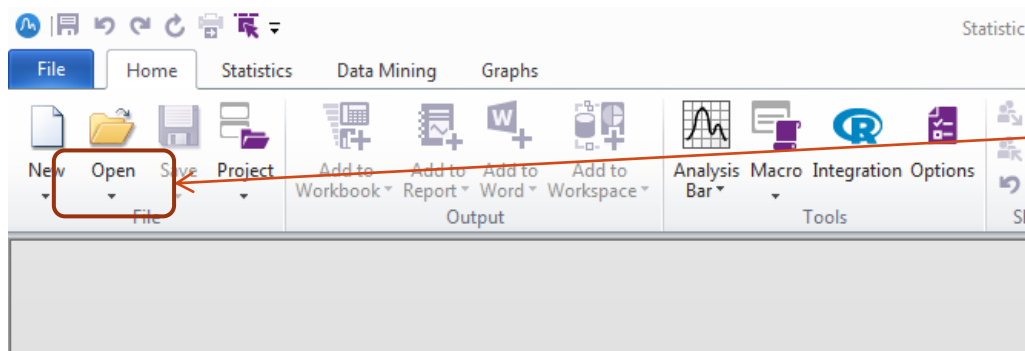
II. rész

Problémák megoldása a STATISTICA szoftver segítségével

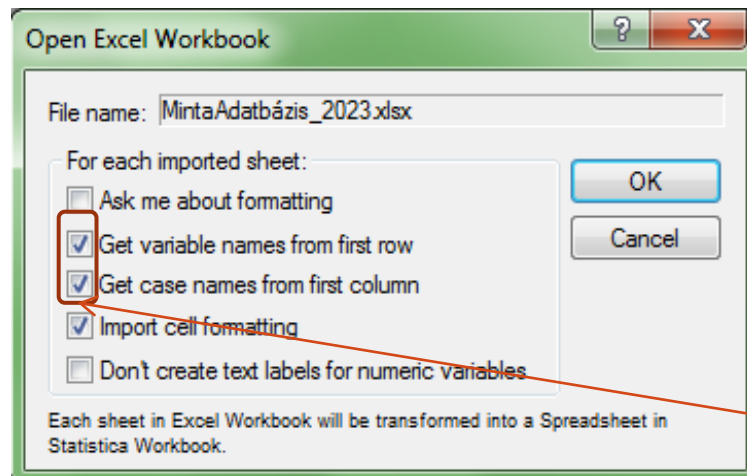


Adatbázis importálása

➤ Open/Open document (Ctrl + O)



legördülő menüből
„Open document”
majd keressük meg
az adatbázist (excel)



ha csak egy munkalapon van adat, válasszuk
az „Import all sheets to a Workbook” opciót



ha az első sorban van a változók,
illetve az első oszlopban az esetek neve,
ezek legyenek bepipálva, aztán OK

Ábrák, változók választása

➤ Graphs/Box/Quick/Variables

The screenshot displays the Statistica software interface. The 'Graphs' menu is open, and the 'Box' option is selected. The '2D Box Plots' dialog box is open, showing the 'Variables' tab. The 'Variables' button is highlighted with a red box. The background shows a data table with columns for patient codes and various medical parameters.

Kód	Munka1	Kor	Neu	HT	DM	Krónikus szivbete	Krónikus tüdőbete	Malignit	CRP (mg/L)	WBC (G/L)	Neu (G/L)	Ly (G/L)	Súlyosság mintavételkor: 1 - ambuláns, 2 - hospitalizált, O2	Elhunyt	Túlélés (nap)
(71	1	1	0	1	1	0	0	171,1	11,58	10,3	0,77	4	1	
(68	1	1	0	1	1	0	0	250,4	6,01	5	0,91	4	0	
(59												4	0	
(79												4	1	
(85												3	0	
(67												3	0	
(67												2	0	
(53												2	0	
(32												2	0	
(65												3	0	
(33												4	0	
(78												3	0	
(72												4	1	
(63												3	0	
(69												2	0	
(53												2	0	
(83												3	0	
(86												3	0	
(51												2	0	
(18												2	0	
(78												4	1	
(57												3	0	
(54												2	0	
(70												3	1	
(76												4	1	
(67												1	0	
(34	0	0	0	0	0	0	0	0	7,47	4,0	2,21	1	0	
(80	1	1	0	0	1	0	1	24,7	6,69	5,1	0,99	3	0	
(44	0	0	0	0	0	0	0	0,5	9,13	6,23	2,35	1	0	
(34	1	0	0	0	0	0	0	0	9,16	7,28	1,53	1	0	
(31	0	0	0	0	0	0	0	0,1	6,28	3,97	1,66	1	0	
(75	1	0	0	0	0	0	1	107,1	6,94	5,16	0,85	4	0	
(78	1	1	0	0	0	0	0	149,1	6,8	5,95	0,33	4	1	
(64	1	0	0	0	0	1	0	0,1	5,22	2,66	2,02	1	0	

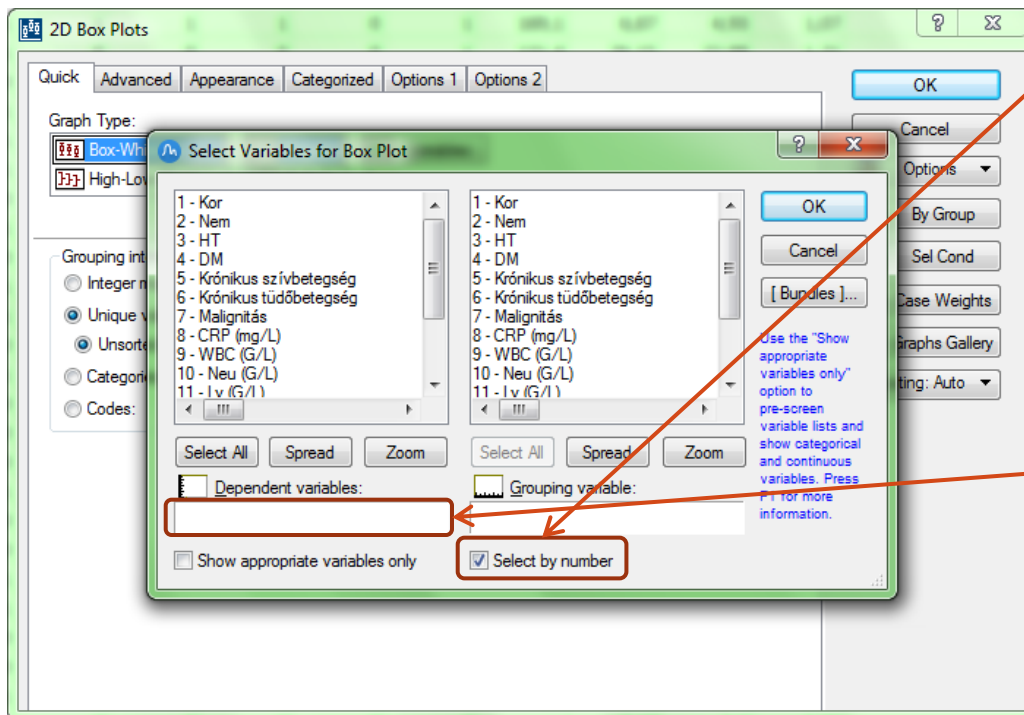
Változók kiválasztása

➤ Graphs/Box/Quick/Variables

a legújabb verzióban
ezt be kell pipálni!

ezt követően a Ctrl gomb
nyomvatartásával tudunk
több változót kijelölni, illetve
ezek kijelölését megszüntetni

alternatív megoldásként
a változók sorszámait
(szóközökkel elválasztva)
manuálisan be is írhatjuk
(vagy gyakran használt
kombinációk esetén
bemásolhatjuk)



Esetek alcsoportjának kiválasztása

➤ Graphs/Box/Quick/Sel Cond (Select cases)

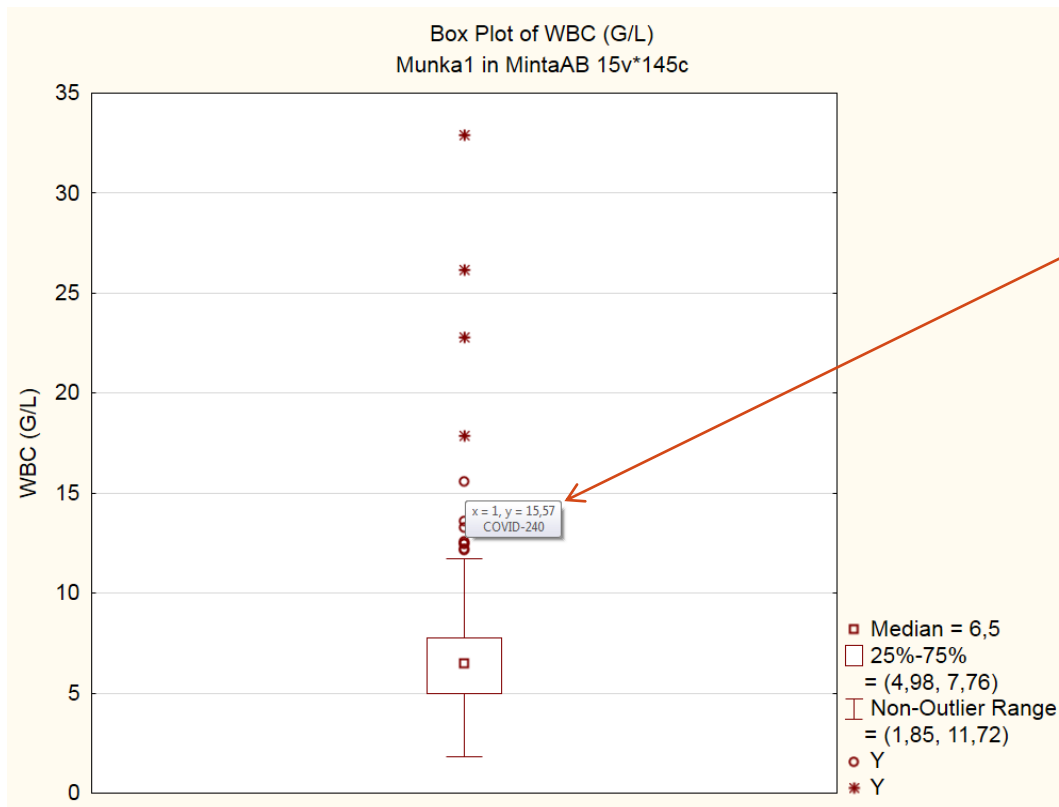
The image shows two overlapping dialog boxes from SPSS. The top dialog is '2D Box Plots' with tabs for Quick, Appearance, Categorized, Options 1, and Options 2. Under 'Graph Type', 'Box-Whiskers' is selected. Under 'Grouping intervals', 'Integer mode' is selected with 'Auto' checked. The 'Sel Cond' button is highlighted with a red box. The bottom dialog is 'Analysis/Graph Case Selection Conditions'. The 'Enable Selection Conditions' checkbox is checked. The 'Review Variables...' button is highlighted with a red box. Under 'Include cases', 'Specific, selected by:' is selected, and the 'By expression' field contains 'v12>1', which is also highlighted with a red box. A red arrow points from the 'Sel Cond' button in the top dialog to the 'Review Variables...' button in the bottom dialog. Another red arrow points from the 'v12>1' expression to the text on the right.

a „v12” a 12. változót jelöli

a változók sorszámát a „Review Variables” gombra kattintva ellenőrizhetjük

Outlierek kimutatása

➤ Graphs/Box/Quick/OK

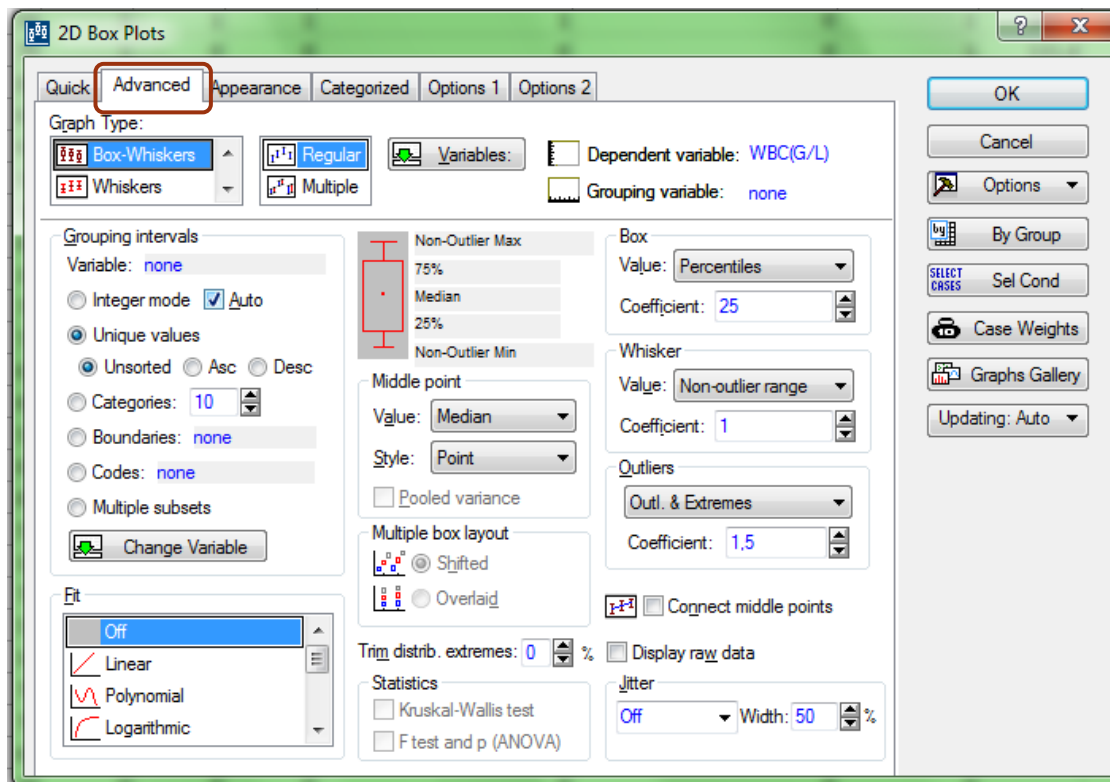


outlier leolvasható



Outlierek kimutatása

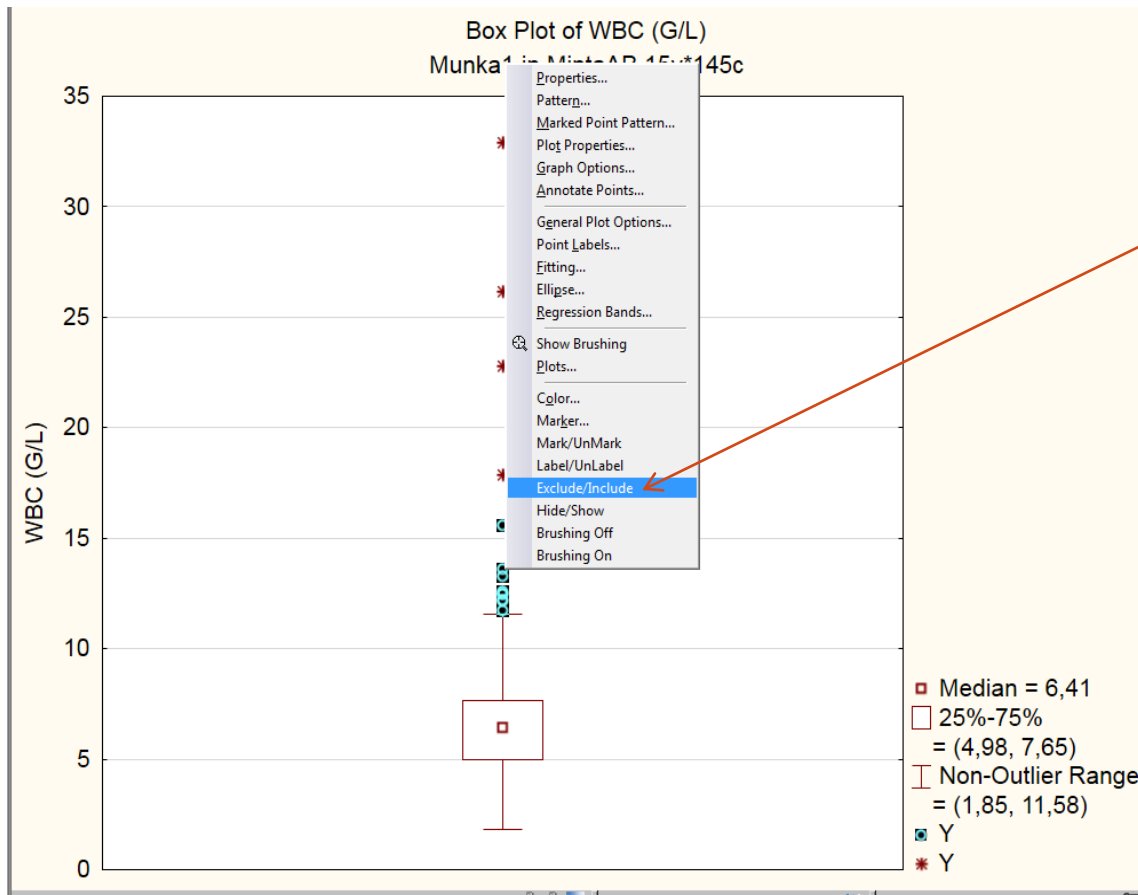
➤ Graphs/Box/Advanced



itt lehet módosítani
a box-whiskers plot beállításait és
az outlierek feltételeit

Outlierek kizárása

➤ Graphs/Box/Quick/OK/jobb-klick



outlierek kizárhatók
vagy visszavehetők

Outlierek kizárása

➤ Adatbázisban: esetre (sorra) jobb-klick

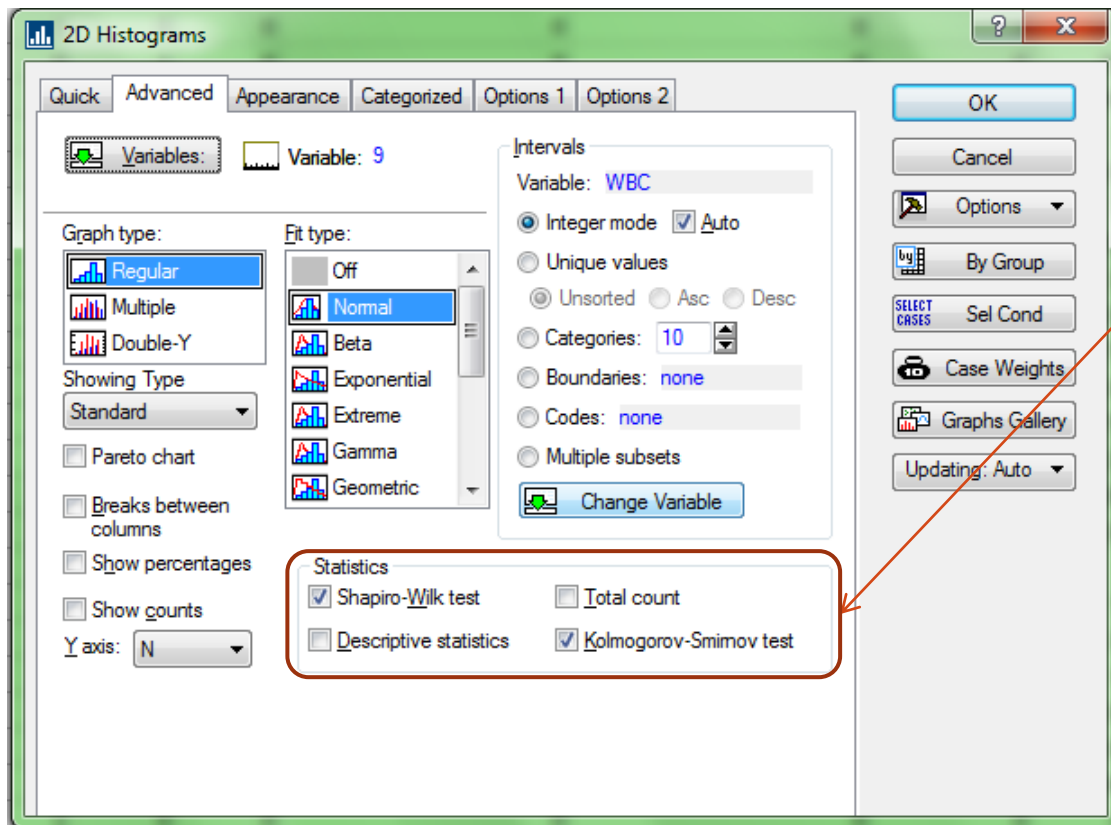
Kód	Munka1				
	Kor	Nem	HT	DM	Krónikus szí
COVID-2	71	1	1	0	
COVID-3	68	1	1	0	
COVID-4	59	1	1	1	
COVID-5	79	1	0	0	
COVID-6	85	0	1	0	
CC			1	0	
CC			0	1	
CC			0	0	
CC			0	0	
CC			0	1	
CC			0	0	
CC			1	0	
CC			0	0	
CC			0	0	
CC			0	0	
CC			1	1	
CC			1	0	
CC			1	1	
CC			0	0	
CC			0	0	
CC			1	0	
CC			0	0	
CC			1	0	
CC			1	1	
CC			0	0	
CC			1	0	
CC			0	0	
CC			0	0	
CC			0	0	
CC			0	0	
COVID-55	75	1	0	0	
COVID-56	78	1	1	0	
COVID-63	64	1	0	0	
COVID-66	60	1	1	0	

a kizáráshoz az off-ra kell kattintani

a kizárt értékeket piros jel jelzi

Normalitás vizsgálata

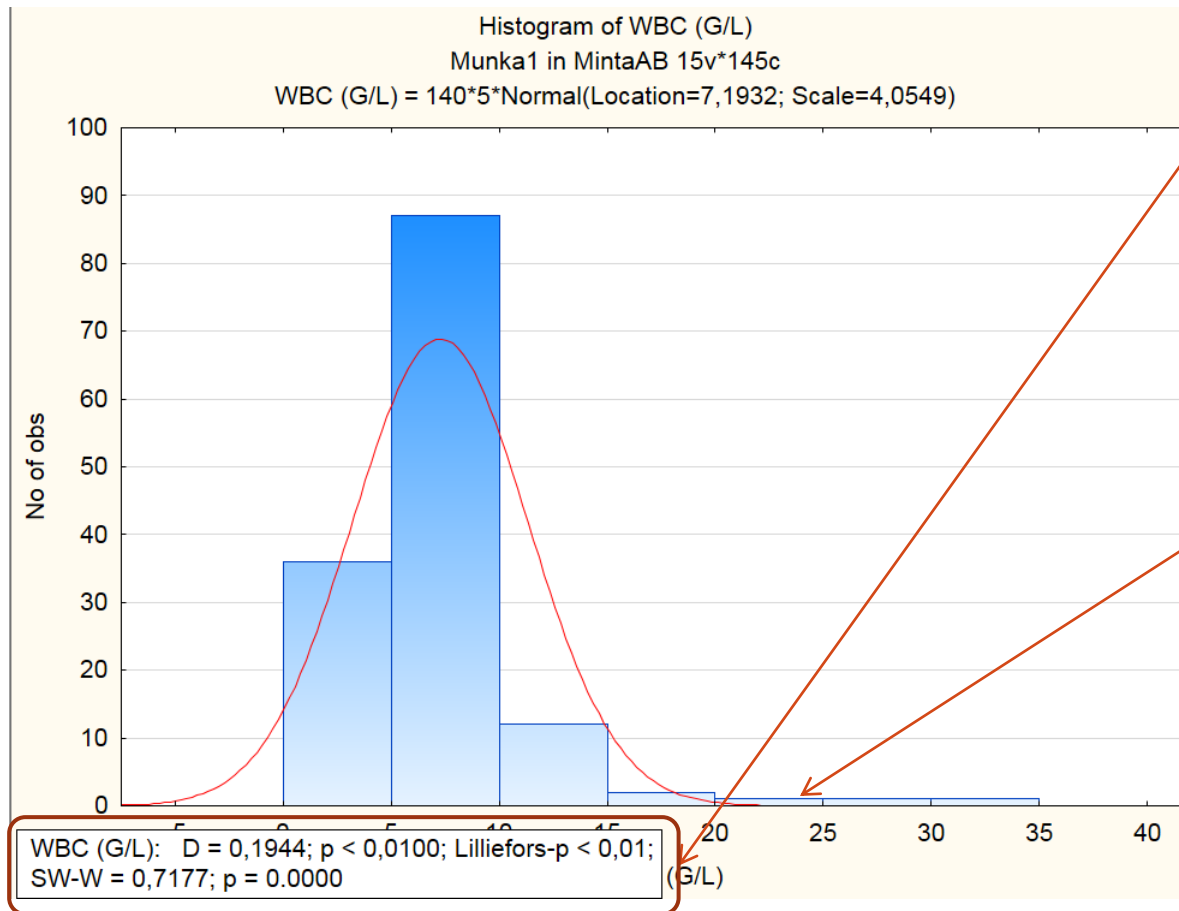
➤ Graphs/Histogram/Advanced



Kolmogorov-Smirnov és
Shapiro-Wilk
teszteket itt kell kiválasztani

Normalitás vizsgálata

➤ Graphs/Histogram/OK



Kolmogorov-Smirnov és Shapiro-Wilk teszt eredménye szignifikáns: nem normál eloszlás

az eloszlás jobbra (pozitívan) ferdült

Transzformálás (normalizálás?)

- Változóra (oszlop) jobb-klick/Add variables
- vagy: Edit/Variables+/Add

How many: 1

After: v9

Name: logWBC

Type: Double

MD code: -99999998

Length: 8

Display format

General

Number

Date

Time

Scientific

Currency

Percentage

Fraction

Custom

Decimal places: 4

Long name (label or formula with Functions):

=log(v9)

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

v9 az eredeti változó (WBC)

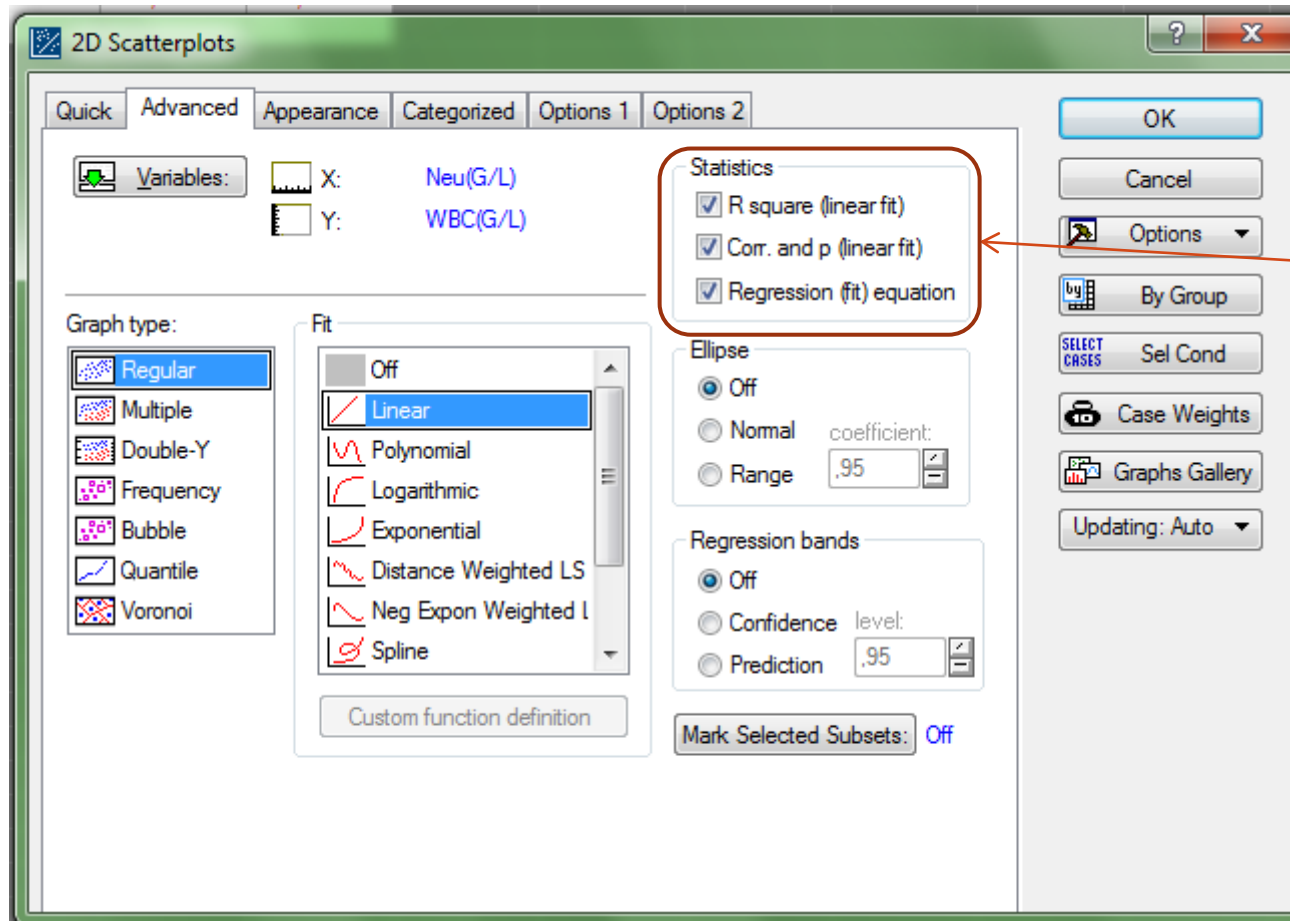
értelmesen nevezzük el
az új változót

logaritmizálásnál
a tizedesjegyek fontosak

itt lehet beírni a képletet
=log₁₀(v9) lenne a tízes alapú logaritmus
képlete helyesen

Linearitás vizsgálata

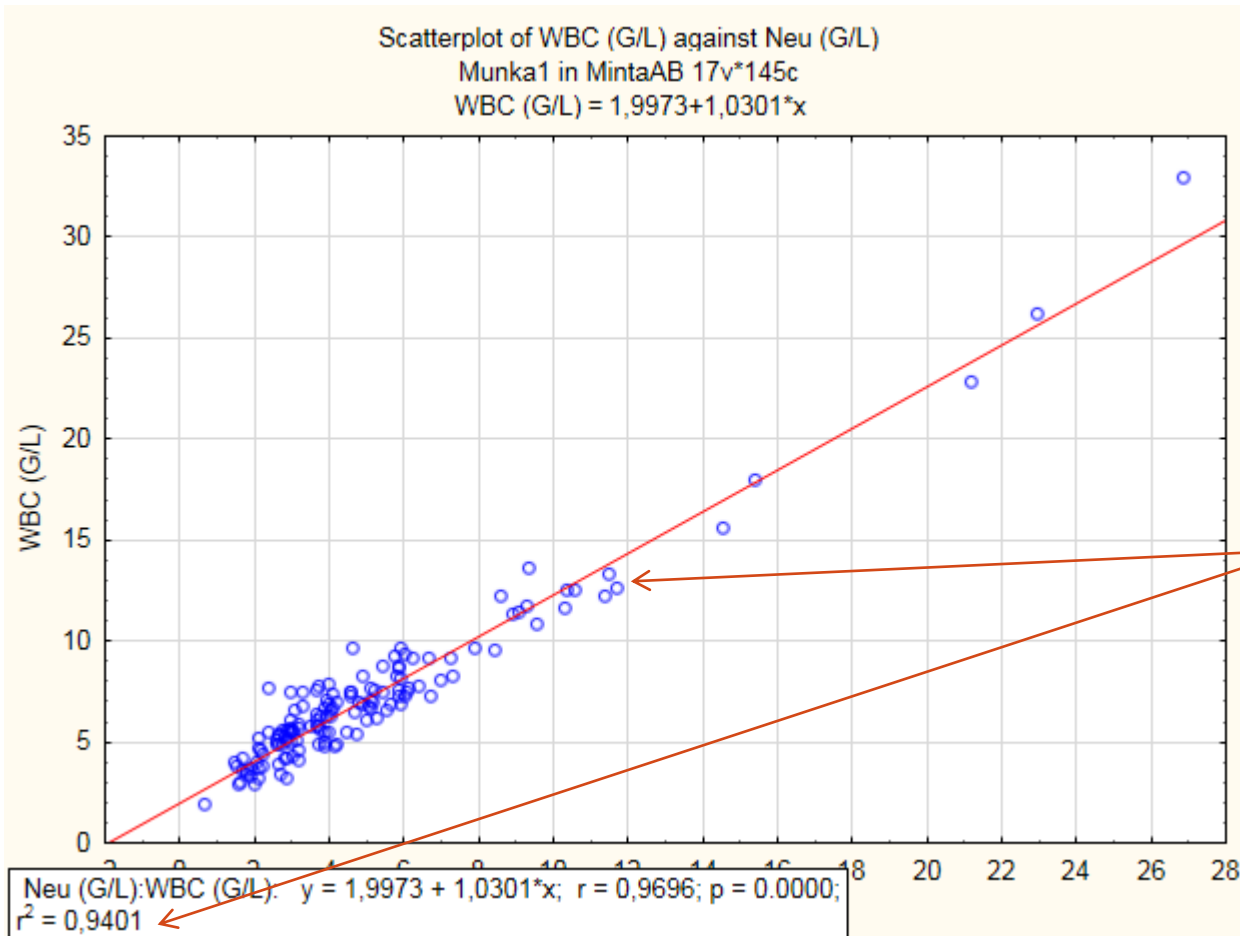
➤ Graphs/Scatterplot/Advanced



itt is kérhető
korreláció és
regresszió
vizsgálata is

Linearitás vizsgálata

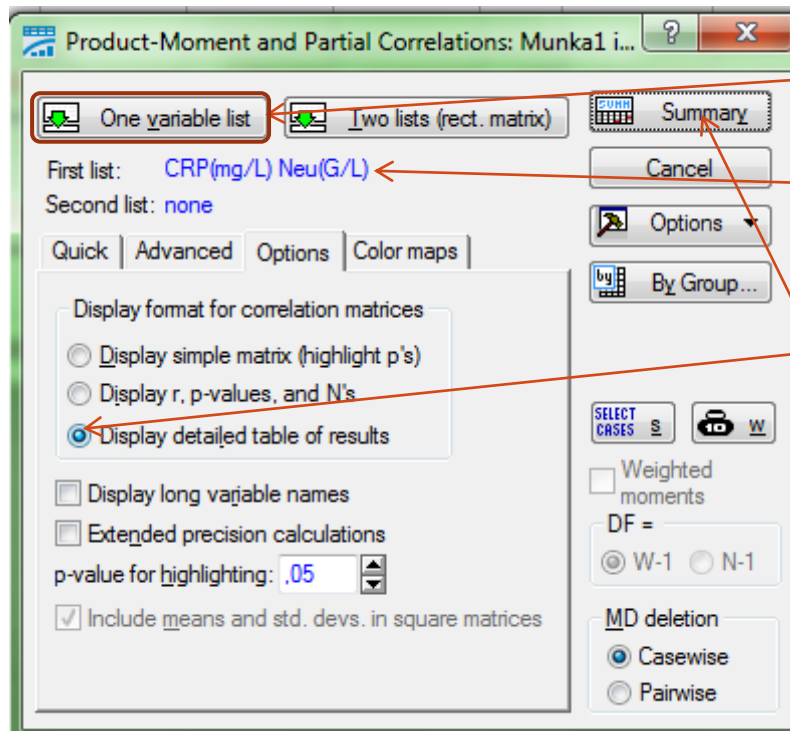
➤ Graphs/Scatterplot



gyanúsán szép
lineáris kapcsolat
(ugyanis a változók
nem függetlenek!)

Lineáris (Pearson) korreláció

➤ Statistics/Basic statistics/Correlation matrices/Options



kérdéssé változók kiválasztása

kiválasztott változók

részletes eredménytáblát kérjünk

ha minden kész, kattintsunk a Summary-re

Lineáris (Pearson) korreláció

➤ Statistics/Basic statistics/Correlation matrices/Summary

Var. X & Var. Y	Mean	Std.Dv.	r(X,Y)	r ²	t	p	N	Constant dep: Y	Slope dep: Y	Constant dep: X	Slope dep: X
CRP (mg/L)	61,34599	79,57665									
CRP (mg/L)	61,34599	79,57665	1,000000	1,000000			137	0,000000	1,000000	0,000000	1,000000
CRP (mg/L)	61,34599	79,57665									
Neu (G/L)	5,03923	3,85229	0,500959	0,250960	6,725372	0,000000	137	3,551506	0,02425	9,198570	10,34830
Neu (G/L)	5,03923	3,85229									
CRP (mg/L)	61,34599	79,57665	0,500959	0,250960	6,725372	0,000000	137	9,198570	10,34830	3,551506	0,02425
Neu (G/L)	5,03923	3,85229									
Neu (G/L)	5,03923	3,85229	1,000000	1,000000			137	0,000000	1,000000	0,000000	1,000000

r
korrelációs
koefficiens

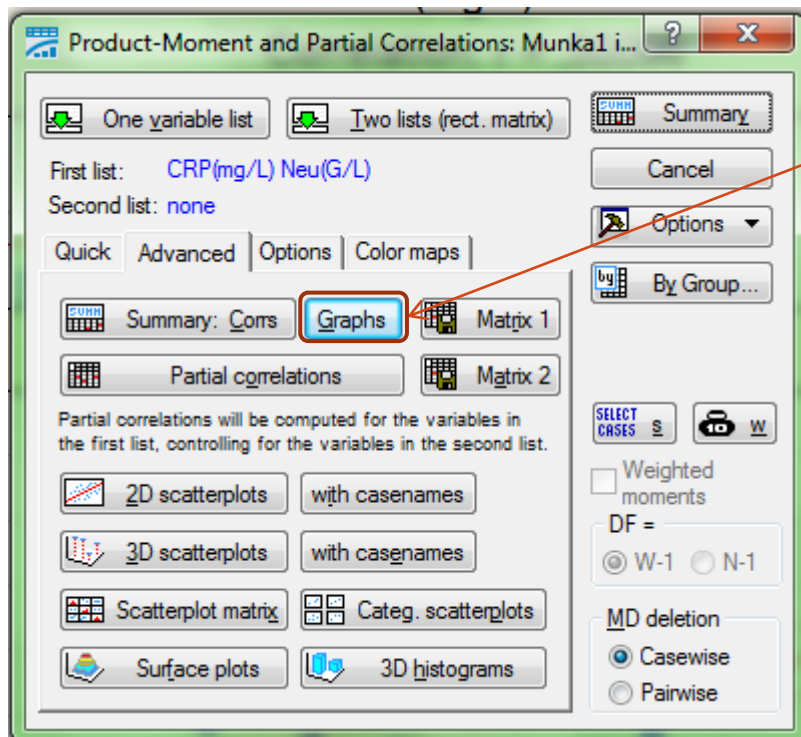
r²
determinációs
koefficiens

p-érték
szignifikancia



Normalitás vizsgálata menet közben

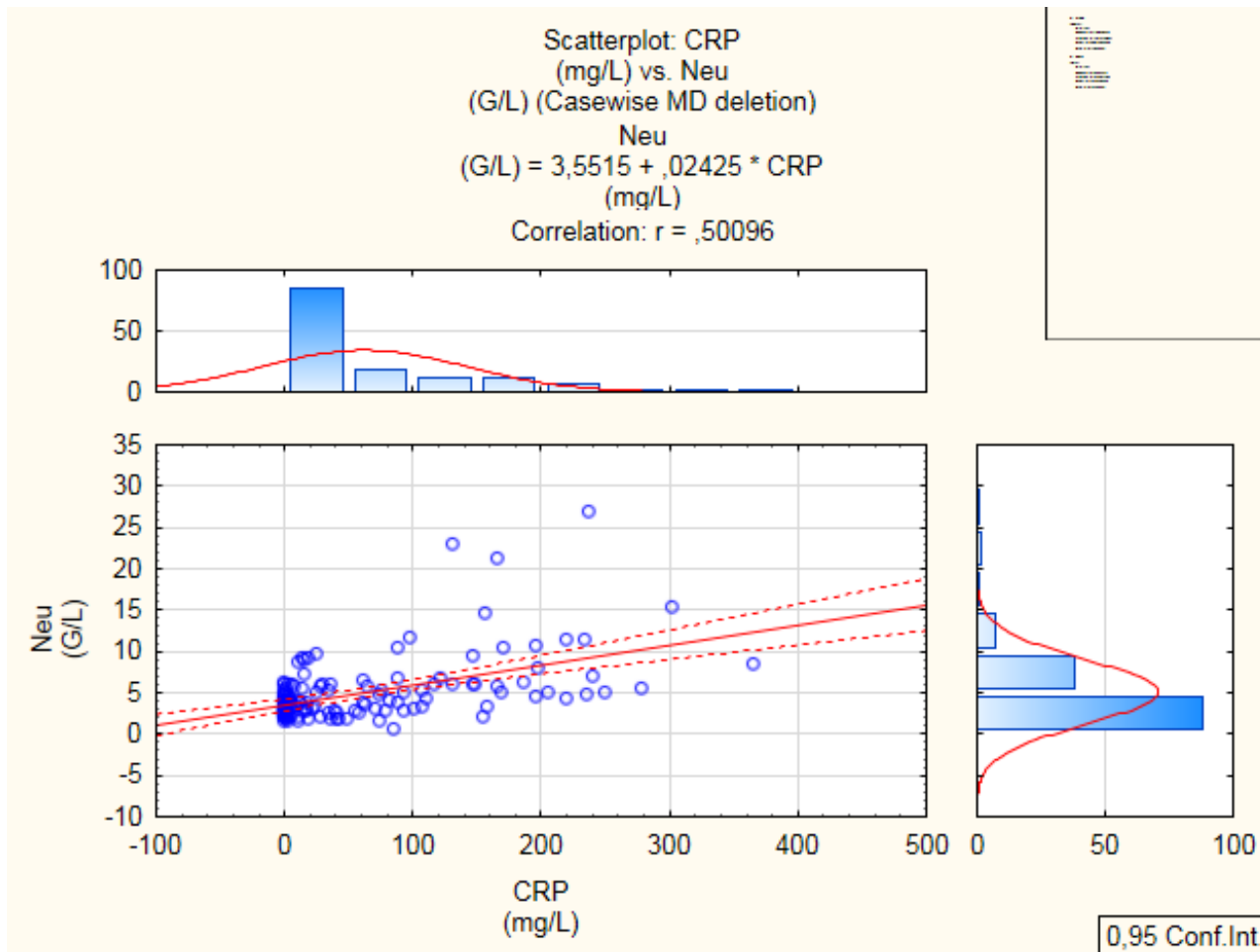
- Statistics/Basic statistics/Correlation matrices/Advanced



kérjünk grafikont

Normalitás vizsgálata menet közben

- Statistics/Basic statistics/Correlation matrices/Advanced/Graphs

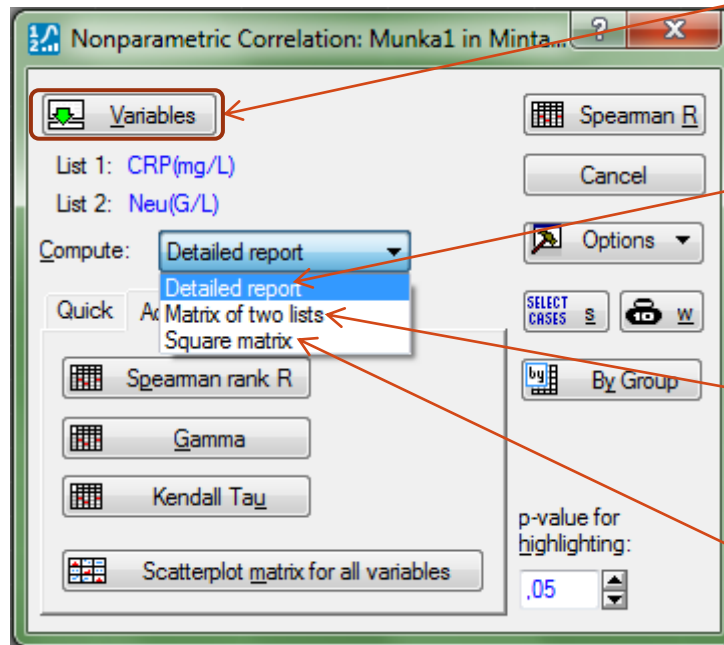


egyik változó sem normál eloszlású

végezzünk nemparaméteres tesztet!

Rang- (Spearman) korreláció

➤ Statistics/Nonparametrics/Correlations/Advanced



kérdéses változók kiválasztása

részletes elemzés,
egy változó összefüggései
egy listányi változóval
r, t és p-érték is

áttekintés,
csak az r, szignifikáns kiemelve
két oszlop tagjai egymással

áttekintés,
minden mindennel



Rang- (Spearman) korreláció

- Statistics/Nonparametrics/Correlations /Detailed report

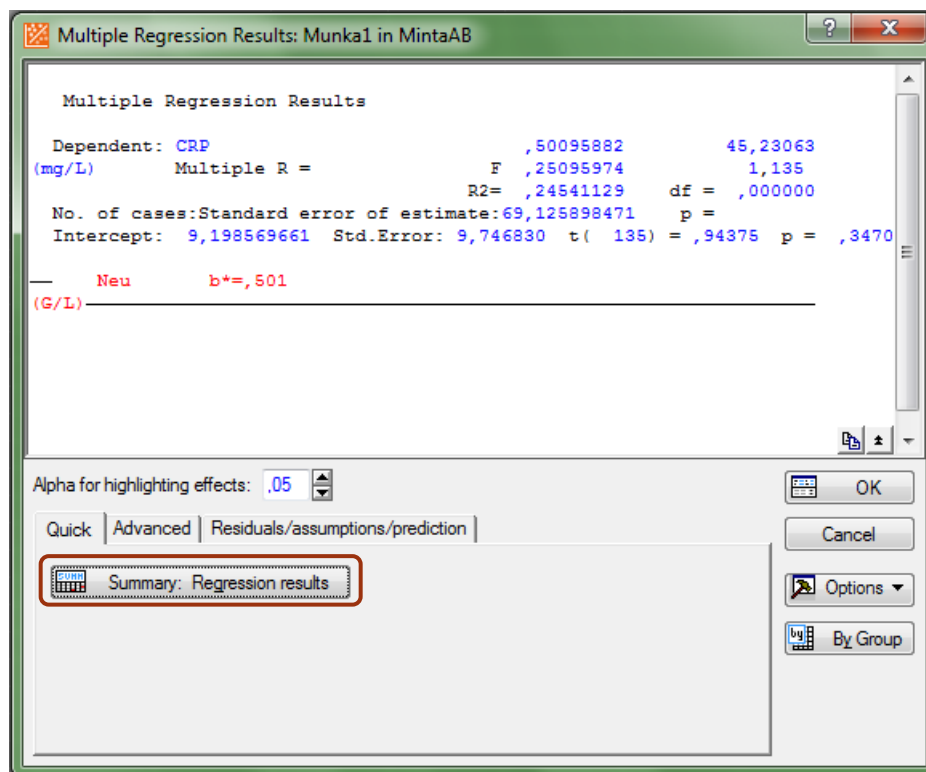
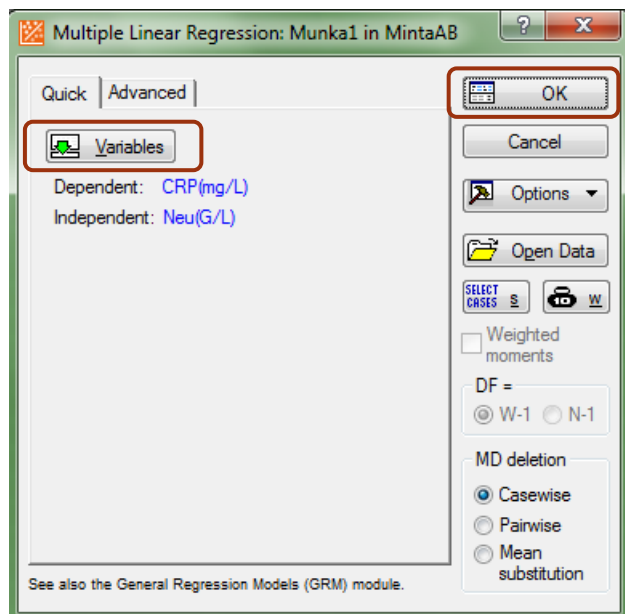
Pair of Variables	Spearman Rank Order Correlations (Munka1 in MintaAB) MD pairwise deleted Marked correlations are significant at $p < ,05000$			
	Valid N	Spearman R	t(N-2)	p-value
CRP (mg/L) & Neu (G/L)	137	0,441303	5,713963	0,000000

r
korrelációs
koefficiens

p-érték
szignifikancia

Lineáris regresszió

- Statistics/Multiple regression/(változók)/OK/Summary



Lineáris regresszió

➤ Statistics/Multiple regression/(változók)/OK/Summary

R négyzet – a modellre vonatkozik a függő változó variáciájának hány százalékát magyarázzák a független változó(k) – itt kb. 25%

F-próba és a hozzá tartozó p-érték annak meghatározására, hogy ez a megmagyarázott hányad szignifikáns-e

Regression Summary for Dependent Variable: CRP (mg/L) (Munka1 in MintaAB)						
R= ,50095882 R2= ,25095974 Adjusted R2= ,24541129						
F(1,135)=45,231 p<,00000 Std. Error of estimate: 69,126						
	b*	Std. Err. of b*	b	Std. Err. of b	t(135)	p-value
N=137						
Intercept			9,19857	9,746830	0,943750	0,346983
Neu (G/L)	0,500959	0,074488	10,34830	1,538695	6,725372	0,000000

standardizált B – több változó esetén lehetővé teszi az összehasonlítást

meredekség

tengelymetszet

$$\text{CRP} = 10,3 * \text{Neu} + 9,2 + \epsilon$$

az adott modellen belül egy-egy független változó hatása (a regressziós együtthatók parciálisak – a többi független változó hatása kontroll alatt van)

Lineáris regresszió - feltételek

➤ Statistics/Multiple regression/Residuals.../

Multiple Regression Results

Dependent: CRP
(mg/L) Multiple R = ,50095882 45,23063
F ,25095974 1,135
R2= ,24541129 df = ,000000
No. of cases: Standard error of estimate: 69,125898471 p =
Intercept: 9,198569661 Std. Error: 9,746830 t(135) = ,94375 p = ,3470

Neu b* = ,501
(G/L)

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

Perform residual analysis

Descriptive statistics

Code generator

Predict values

Predict dependent variable

Compute confidence limits Alpha: .05

Compute prediction limits

Residual Analysis: Munka1 in MintaAB

Dependent: CRP
(mg/L) Multiple R : ,50095882 45,23063
F ,25095974 1,135
R2: ,24541129 df = ,000000
No. of cStandard error of estimate: 69,125898471 p =
Intercept: 9,198569661 Std. Error: 9,746830 t(135) = ,94375 p < ,3470

Quick | Advanced | Residuals | Predicted | Scatterplots | Probability plots | Outliers | Save

Summary: Residuals & predicted

Normal plot of residuals

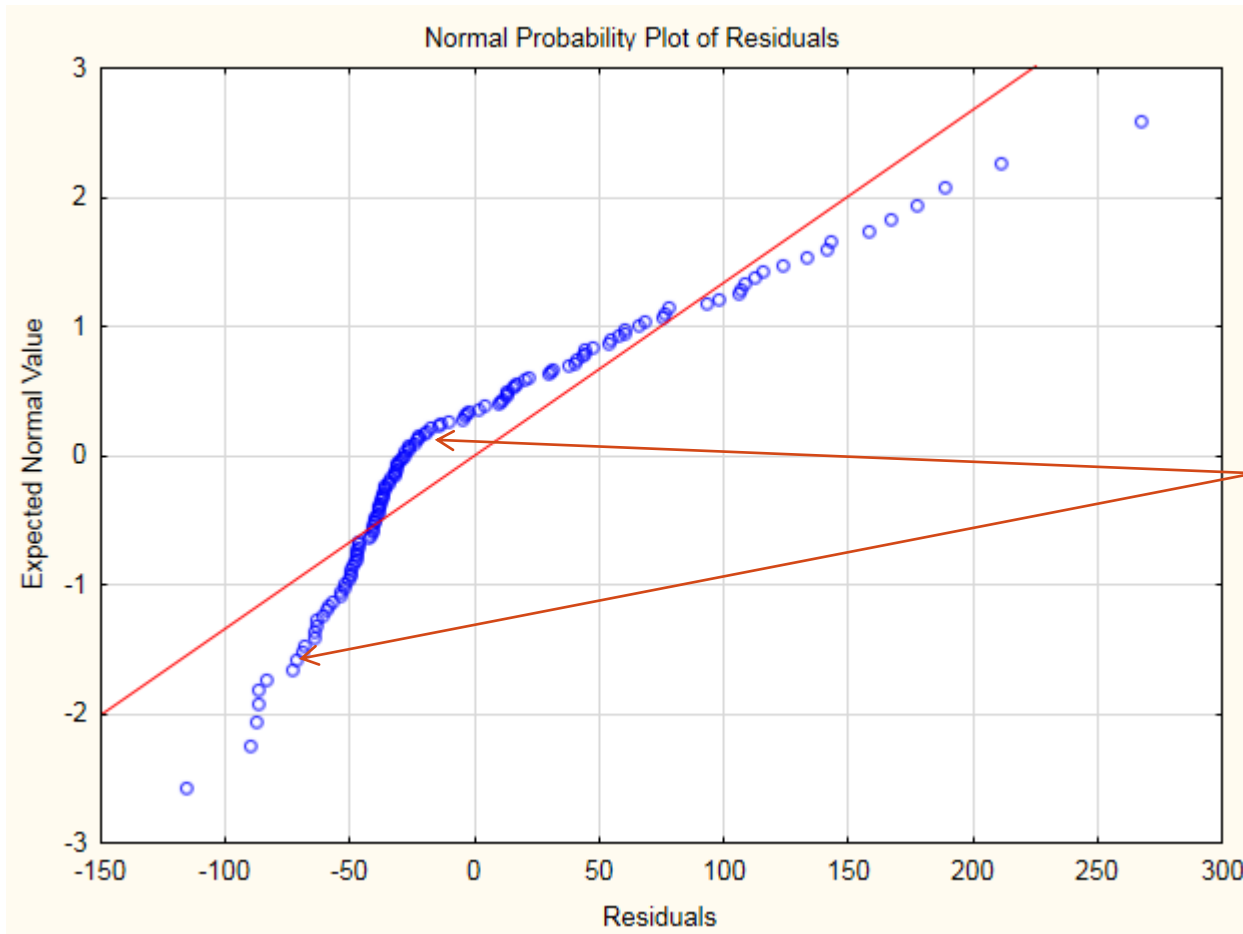
Cancel

Options

By Group

Lineáris regresszió - feltételek

➤ Statistics/Multiple regression/Residuals.../



a reziduálisok eloszlása
eltér a
normál eloszlástól

Lineáris regresszió - feltételek

➤ Statistics/Multiple regression/Residuals.../

Multiple Regression Results

Dependent: CRP (mg/L)

Multiple R = .50095882

F = 45.23063

R2 = .24541129

No. of cases: Standard error of estimate: 69,125898471

Intercept: 9,198569661 Std. Error: 9,746830 t(135) = ,94375 p = ,3470

Neu b* = ,501 (G/L)

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

Perform residual analysis

Descriptive statistics

Code generator

Predict values

Compute confidence limits

Compute prediction limits

Residual Analysis: Munka1 in MintaAB

Dependent: CRP (mg/L)

Multiple R : .50095882

F = 45.23063

R2: .24541129

No. of cStandard error of estimate: 69,125898471

Intercept: 9,198569661 Std. Error: 9,746830 t(135) = ,94375 p < ,3470

Quick | Advanced | Residuals | Predicted | Scatterplots | Probability plots | Outliers | Save | Summary

Predicted vs. residuals

Observed vs. squared residuals

Predicted vs. squared residuals

Residuals vs. deleted residuals

Predicted vs. observed

Bivariate correlation

Observed vs. residuals

Partial residual plot

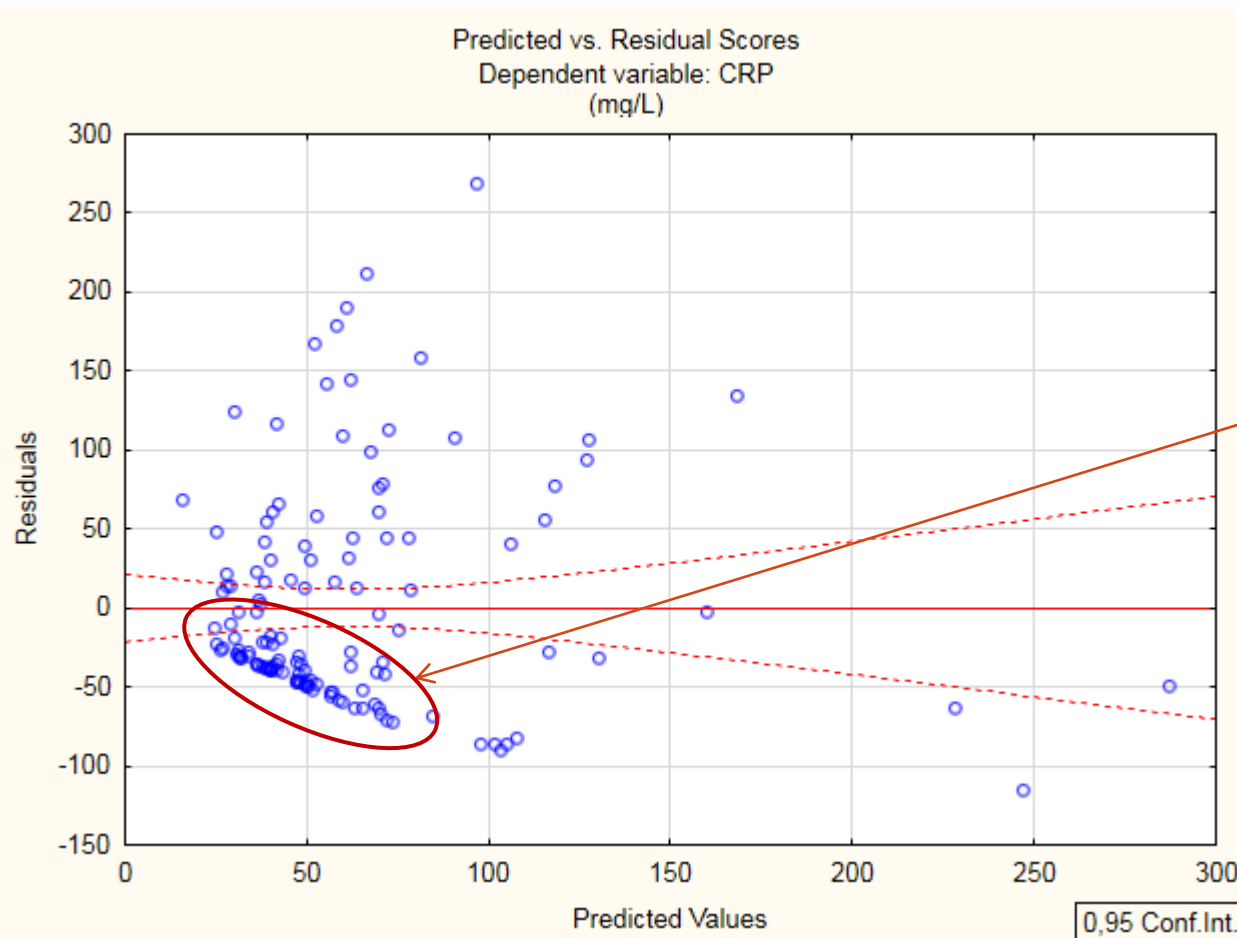
Cancel

Options

By Group

Lineáris regresszió - feltételek

➤ Statistics/Multiple regression/Residuals.../



a reziduálisok eloszlása
nem homogén
(heteroszkedaszticitás)