KAPCSOLATVIZSGÁLATI MÓDSZEREK

Sokváltozós regressziós vizsgálatok (Multiple Regression Analysis)

Két ismérv (x és y) közötti kapcsolat

- A két ismérv **független egymástól**, ha x ismérv szerinti hovatartozás nem ad semmiféle többlet információt az y szerinti hovatartozásról.
- A két ismérv között **sztochasztikus összefüggés** van, ha az egyik ismérv változathoz való tartozásból tendenciaszerűen, *valószínűségi jelleggel* következtethetünk a másik ismérv szerinti hovatartozásra.
- A két ismérv **függvényszerű kapcsolatban** áll egymással, ha a vizsgált egységek x szerinti hovatartozásának ismeretében teljesen egyértelműen megmondható azok y szerinti hovatartozása is.

A függetlenség kölcsönös

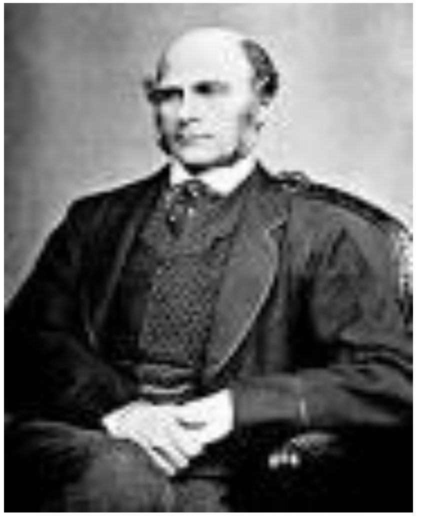
FONTOS:

Ha Y független X-től,
akkor X is független Y-tól

Regresszió (Regression)

- **Általános jelentése:** visszaesés, hanyatlás, visszafelé mozgás.
- **Orvosi területen (regrediál):** a betegség javulása.

Sir Francis Galton



Born: February 16, 1822, Birmingham,
United Kingdom
Died: January 17, 1911,

$$\hat{y} = a \pm b \cdot x + \varepsilon$$

Hiba (0 átlagú)

Eredményváltozó (prediktandusz)

Magyarázó változó (prediktor)

Egyváltozós lineáris modell

Regressziós felületek

- Egy változó esetén: regressziós egyenes.
- Két változó esetén: regressziós sík.
- Több változó esetén: regressziós felület.

Pearson féle lineáris korreláció

(Bivariate correlation)

Carl (Karl) Pearson

Born: March 27, 1857, Islington, London, United Kingdom
Died: April 27, 1936, Capel, United Kingdom

Parciális korreláció (partial correlation):

A parciális korreláció arra ad választ, hogy milyen mértékű lenne X és Y változók között a kapcsolat, ha kiszűrnénk Z változó hatását az X és Y változókra. Az X változót tekintjük itt a független változónak, az Y-t pedig a függő változónak.

Szemiparciális korreláció (semipartial correlation):

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 az éppen vizsgált változót.

R^2 (coefficient of determination) *meghatározottsági együttható*

$$SSR = Q_{reg} = \sum_{i=1}^{n}\left(\tilde{Y}_i - \overline{Y}\right)^2 \qquad SSTO = Q = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$

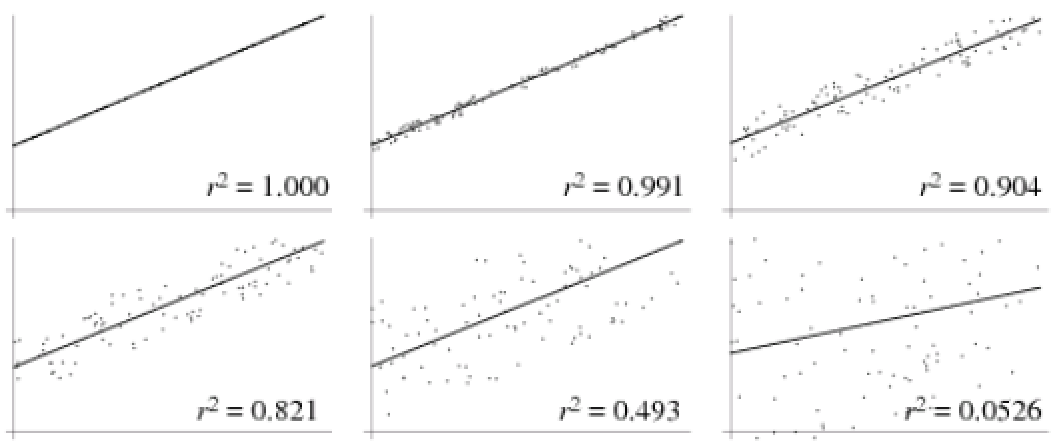
Ha csak egy magyarázó változó van, akkor R² éppen a korrelációs együttható négyzete!

$$\ldots\frac{SR}{TO} = 1 - \frac{SSE}{SSTO}, \quad 0 \le R^2 \ldots$$

Megmutatja, hogy a lineáris regresszióval a célváltozó varianciájának mekkora hányadát lehet magyarázni

$$R = \pm\sqrt{\frac{SSR}{SSTO}} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \ldots\right.}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}, \quad -1 \le R \le 1.$$

Az R^2 érték megmutatja a lineáris kapcsolat mértékét



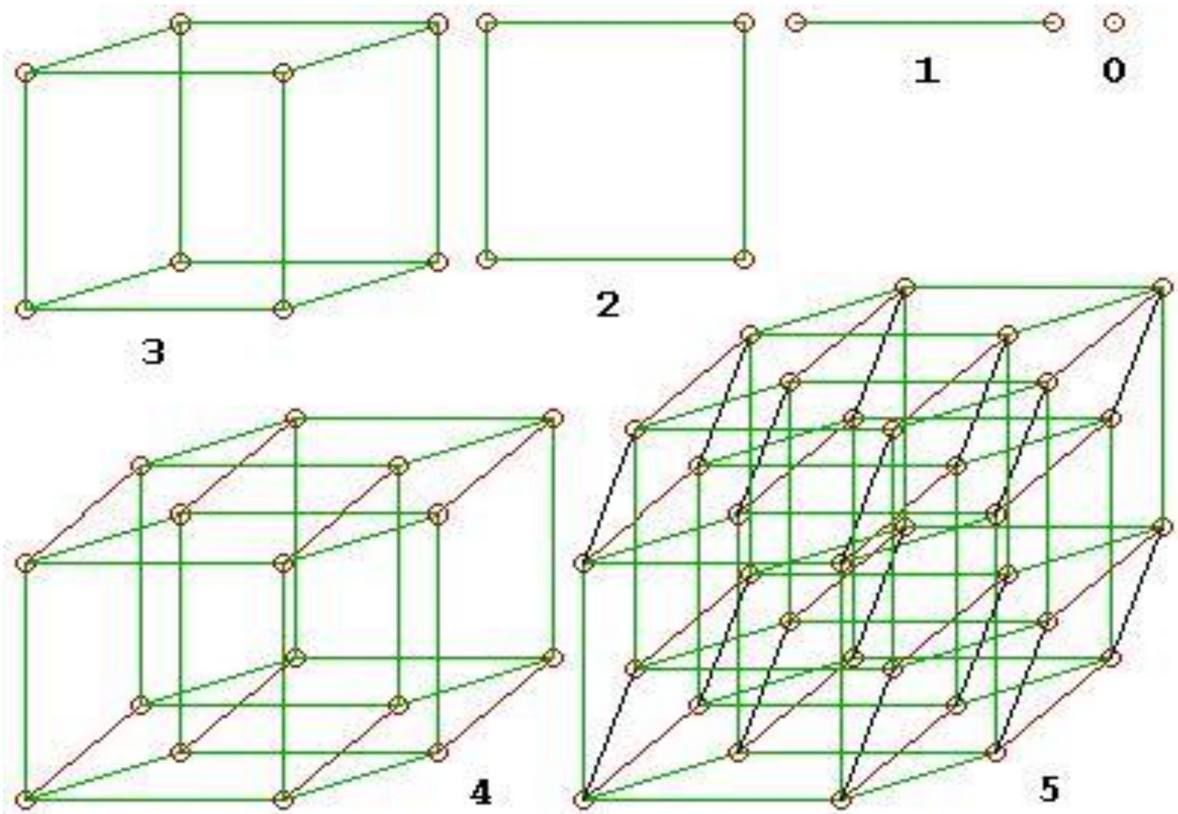
Korrigált (adjusztált) meghatározottsági mutató

$$R_{adj}^{2} = 1 - \left(1 - R^{2}\right)\frac{n-1}{n-p-1} = 1 - \frac{SSE/(n-p-1)}{SSTO/(n-1)}$$

p a független változók száma

A korrekció azért szükséges, mert újabb változók bevonásával R^2 automatikusan nő, és túl optimista képet mutat a modell illeszkedéséről. Az adjusztált változatban „büntetjük" a túl sok változó bevonását a modellbe. p=1 esetben nem korrigálunk.

Dimenzió: *méret*, *kiterjedés*. Hétköznapi használatban a dimenzió a fizikai tér, a testek különféle méreteinek, nagyságfajtáinak (szélesség, hosszúság, magasság) összefoglaló neve.



Az egyes térdimenziók bemutatása

Forrás: Wikipédia

Lineáris regresszió: olyan paraméteres regressziós modell, mely feltételezi a magyarázó- (X) és a a célváltozó (Y) közti (paramétereiben) lineáris kapcsolatot: az adatok pontfelhőjére egy egyenest illesztünk.

Komplex modell=Hierarchikus lineáris regresszió: csoportosított adatok hierarchikus szintekre osztása esetén használjuk. Pl.

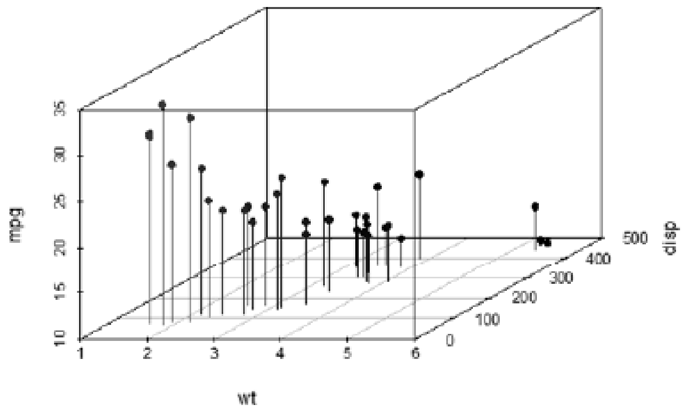
a) Oktatásban, ahol az adatok gyakran a tanuló, az osztály és az iskola szintjeire oszthatóak.
b) Ismételt méréses kutatásnál, a más időpontokban és más kondíciók között gyűjtött adatot minden résztvevő személynél önmagához mérjük.

A többváltozós lineáris regresszió

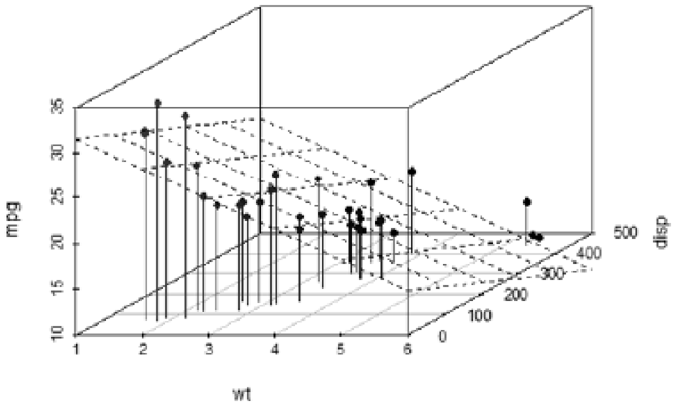
A regressziós egyenlet együtthatóinak a keresése a **legkisebb négyzetek elve** alapján történik. Geometriailag az (r+1)-dimenziós térben adott pontokhoz **legjobban illeszkedő (hiper) síkot** keressük.

A megoldás matematikai módszere: mátrix-egyenletek megoldása.

Háromdimenziós ábra



Regressziós sík



A lineáris regressziós modellek becslésénél feltétel, hogy a prediktor változók varianciája állandó legyen, és ne függjön más változótól.

A keresett egyenlet általános alakja:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 x_2 + \dots + b_n x_n$$

Az alábbi hipotéziseket vizsgáljuk:

H₀: nincs kapcsolat az x_i és y változók között vagy H_0: $b_i = 0$.
H₁: van kapcsolat az x_i és y változók között vagy H_1: $b_i \neq 0$

Az eljárás arra is választ ad, hogy az x_i változók közül melyek az y szempontjából fontos változók, melyek azok, amelyek tényleges befolyásolják az értékét. Ki lehet szűrni a fontos x_i változókat.

Az együtthatók értelmezése azonos az egyszerű lineáris regressziónál tanultakkal!

Feltételezzük, hogy valamennyi változóra *n* számú megfigyelésünk van, amelyeket célszerűen vektorokba, illetve mátrixba rendezhetünk:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 x_{11} & x_{21} & \cdots & x_{k1} \\ 1 x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots \quad \vdots & \vdots & \ddots & \vdots \\ 1 x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}, \quad B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad és \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Az együtthatók meghatározása a legkisebb négyzetek módszerével:

$$\frac{\partial Q}{\partial \underline{b}} = -2\underline{\underline{X}}^T \underline{Y} + 2\underline{\underline{X}}^T \underline{\underline{X}}\,\underline{b} = \underline{0}$$

$$\underline{\underline{X}}^T \underline{\underline{X}}\,\underline{b} = \underline{\underline{X}}^T \underline{Y} \;\Rightarrow\; \underline{b} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{Y}$$

Szórásanalízis (ANOVA) a modell érvényességének eldöntésére

$$H_0: \quad \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$\frac{SSR/(k-1)}{SSE/(n-k)} \in F_{k-1,n-k}$$

A nullhipotézis az, hogy a független változók mindegyike 0, vagyis egyik prediktor változó sem magyarázza a célváltozót!

F-próbával dönthetünk a nullhipotézisről.

Béta-együtthatók

$$BETA_i = b_i \cdot \frac{S_i}{S_y} \quad (i = 1,2,\dots,k)$$

b_i az i-edik regr

S_i az i-edik válto

S_y a célváltozó standard szórása.

A béta-együtthatók egyfajta szempontból minősítik a változók fontosságát a lineáris összefüggésben. Ha egy változónak nagy az együtthatója abszolút értékben, akkor fontos, ha kicsi, kevésbé fontos .

A modell kimeneti értékét meghatározza a prediktor változók:	1. kiválasztása és száma 2. egymáshoz való viszonya 3. modellbe építés sorrendje
A prediktorok helyes mennyisége	Ha túl sok a prediktor, nő a modell bonyolultsága, nő a becsült paraméterek száma, nő a szükséges elemszám (és ez rossz). A modell túlspecifikált lesz, nem írja le a populációt.

- Ökölszabály (thumb of law):

  a szükséges esetek száma > 8*a bevont változók száma.

Többszörös regresszió használatához a feltételek

- **Linearitás a függő változóval:** ha ez nincs, akkor alábecsüljük az y-t, pontatlan a modell.
- **Mintaszám:** kis elemszám növeli a β-hibát. Ökölszabály betartása többváltozós vizsgálatoknál.
- **Nincsenek több dimenziós extrém (outlier) értékek:** az együtthatók torzítását eredményezik.
- **Nincs multikollinearitás:** az összefüggő változók nem értelmezhetők. Ki kell hagyni a kevésbé fontos változót.
- **Minden változónak van varianciája:** nincs konstans változónk.
- **Nincs kovariáns** (külső befolyásoló változó).
- **Független hibák:** a belső korrelációk a CI-t, a szignifikancia értékeket torzítják.
- **Hiba normális eloszlása:** a normalitás sérülése, más feltételek sérüléseként keletkezik.
- **Változók típusai:** dummy-változó is engedett (pl. beteg neme).
- **Homoszkedaszticitás** (varianciák homogenitása vagy szóráshomogenitás): a heteroszkedaszticitás rontja a konfidencia-intervallumokat, torzítja a szignifikancia értékeket.

Miért nem teljesülhetnek a feltételek?

- *Multikollinearitás:* a magyarázó változók nem lineárisan függetlenek
- *Autokorreláció:* a hibatagok lineárisan nem függetlenek
- *Heteroszkedaszticitás:* a hibák szórásnégyzete nem konstans

- A **multikollinearitás** úgy is megfogalmazható, hogy a magyarázó változók között korreláció van.
- Multikollineáris esetén mind a becslés, mind a paraméterek értelmezése megnehezedik, hiszen a magyarázó változók hatásait nem lehet egyértelműen szétválasztani.
- Minden változó hatása minden más változóban is megjelenik, a becslések bizonytalanná válnak.

Multikollinearitás ellenőrzése

- $|R|=0$ multikollinearitás,
  $|R|=1$ a vált. függetlenek.

- Többszörös determinációs együtthatóval:
  R^2

A multikollinearitás mérőszámai:

- **kondíciós index (CI)**
- **variancia hányad**
- **variancia infláló faktor (VIF)**
- **tolerancia**

A multikollinearitás mérőszámai 1.

A kondíciós index (CI) a magyarázó változók korrelációs mátrixának sajátértékeiből számolt statisztika. A legnagyobb és legkisebb sajátértékek hányadosának négyzetgyöke. A CI>15 esetében megállapítható az erős kollinearitás.

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

Variancia hányad is utalhat multikollinearitásra. Ha egy-egy nagy kondíciós index sorában több regressziós együtthatónak van magas variancia hányada. A regressziós együtthatók varianciáit a sajátértékek között szétosztjuk.

A multikollinearitás mérőszámai 2.

VIF (Variance Inflation Factors)

Varianciainflációs tényező azt mutatja, hogy a j-edik változó becsült együtthatójának varianciája hányszorosa annak, ami a multikollinearitás teljes hiányakor lenne. Ezért ezt a mutatószámot a j-edik változóhoz tartozó variancianövelő tényezőnek nevezzük.

$$VIF_j = \frac{1}{1 - R_j^2}$$

VIF

- Minimális értékét, az 1-et akkor veszi fel, amikor a j-edik magyarázó változó nem korrelál a többivel.
- Ahogy nő az R^2_j, úgy nő a VIF értéke is, jelezve, hogy a kollinearitás hányszorosára növeli a varianciával mért becslési hibát.
- Ha $R^2_j = 1$ a VIF mutató nem értelmezhető, ez a teljes vagy extrém multikollinearitás.
- A VIF reciprokát toleranciamutatónak nevezik.

VIF-mutató

$$VIF_j = \frac{1}{1 - R_j^2}$$

- $1 \leq VIF \leq \infty$
- VIF=1 ha $R_j^2=0$ (amikor a j. magyarázó változó nem korrelál a többi magyarázó változóval)
- $VIF \Rightarrow \infty$ $R_j^2=1$ (a j. magyarázó változó pontosan kifejezhető a többi lineáris kombinációjaként)

- $1 < VIF \leq 2$ - gyenge multikollinearitás

  $2 < VIF \leq 5$ - erős zavaró multikollinearitás

  $5 < VIF$ - nagyon erős, káros multikollinearitás

A multikollinearitás mérőszámai 3.

- A *VIF_j -mutató reciprokát toleranciamutatónak nevezzük.*

$$Tolerancia = \frac{1}{VIF_j}$$

- ***É*rtéke:** 0 ≤ Tolerancia ≤ 1.
- Minél nagyobb a multikollinearitás mértéke annál közelebb van a mutató értéke a nullához (<0.2).

Autokorreláció

- A hibatagok lineárisan nem függetlenek
- Az autokorreláció különböző rendű lehet, attól függően, hogy a hibatag *i*-edik értéke melyik értékkel van kapcsolatban. Ha a hibatag *i*-edik értéke közvetlenül az előtte lévő értékkel áll korrelációs kapcsolatban, akkor *elsőrendű autokorreláció*-ról beszélünk. Az elsőrendű autokorreláció modellje:

$$e_i = \rho \cdot e_{i-1} + \nu_i$$

Az elsőrendű autokorreláció mérése

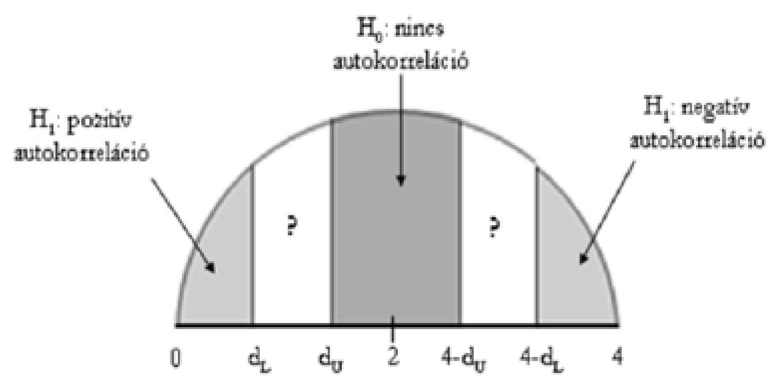
Durbin-Watson teszt:

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

Lineáris autokorrelációs együttható:

$$\hat{\rho} = \frac{\sum_{i=2}^{n} e_i \cdot e_{i-1}}{\sqrt{\sum_{i=2}^{n} e_i^2} \cdot \sqrt{\sum_{i=2}^{n} e_{i-1}^2}}$$

A Durbin-Watson teszt döntési szabálya

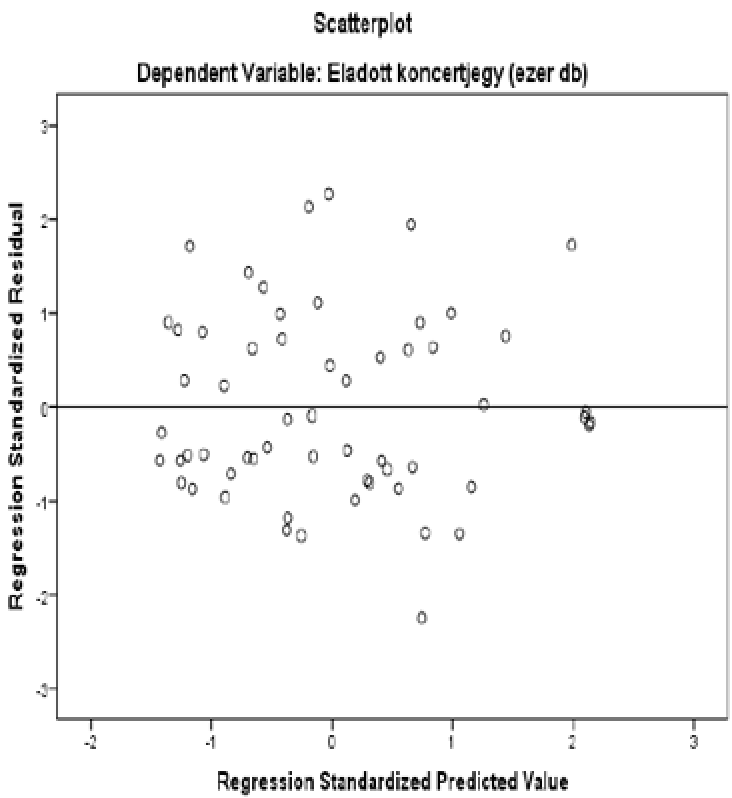


Homoszkedaszticitás és linearitás ellenőrzése

Leginkább azonos szórásúságnak lehetne fordítani.

A standardizált predikált érték és a standardizált reziduális hibák pontdiagramja

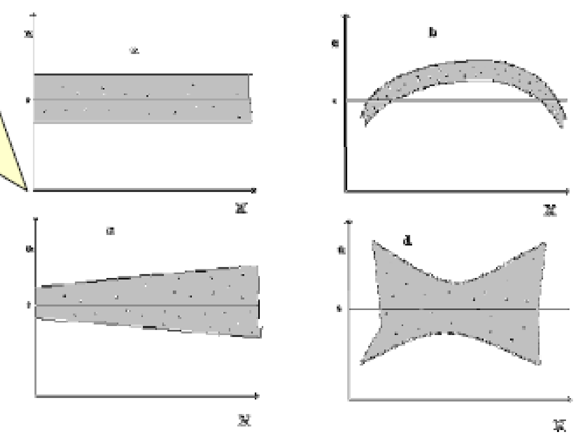
- Egyenletes pontfelhőt várunk – minden prediktor értéknél hasonló mértékű hiba kell hogy legyen
- Homoszkedaszticitás: nem változik a pontfelhő szélessége
- Linearitás: nincs a pontfelhőnek iránya, vagy görbülete



A maradéktagok (reziduálisok) elemzése

Heteroszkedaszticitás:

A maradéktagok nulla szint körüli szóródásának lehetséges típusai
a.) a szóródás megfelel a lineáris modellnek,
b.) nem a lineáris modellhez tartoznak a maradéktagok,
c.) a szóródások nem azonosak,
d.) a hibatagok nem függetlenek egymástól.



A „legjobb” modell kritériumai

- Nem létezik egyetlen „legjobb” modell mérőszám
- A végső modell kiválasztása a becslési pontosság és az egyszerűség kompromisszuma
- „Ockham borotvája” filozófiai elv

A többszörös lineáris regressziószámítás lépései

1. Modellalkotás, változók bevonása
2. Illeszkedés vizsgálata
3. Korrelációs index, determinációs együttható
4. Variancia-analízis, F-próba
5. Együtthatók t-próbája
6. Validálás

Filozófiai elv: két, az adott jelenséget egyformán jól leíró magyarázat közül azt érdemes választani, amelyik az egyszerűbb.

„Borotvával hasítsuk ketté a
szükségtelen hipotéziseket! ”
(William Ockham)

Modellek vizsgálata

- Beágyazott modellek vizsgálata ANOVA-val
- AIC (Akaike information criterion).
- A kisebb érték jelenti az adekvátabb modellt. Nem feltétel, hogy a modell beágyazott legyen.

Beágyazott modellek:

1. Modell: $Se\_hdl = b_0+b_1Nem+b_2Tsuly$

2. Modell: $Se\_hdl = b_0+b_1Nem+b_2Tsuly+b_3Se\_trig+b_4Se\_chol$

Mérések, megfigyelések korrigálása

1. Töröljük a befolyásos értékeket
2. Transzformáljuk a változókat
3. Töröljünk vagy adjunk hozzá változókat
4. Használjunk másik regressziós modellt

1. Megfigyelések törlése

- A kiugró értékek törlése sokszor javít a normális eloszlás feltételén.
- A befolyásos értékek törlése után a modellt újból becsülni kell.
- Az újabb diagnosztika lehet, hogy újabb befolyásos vagy kiugró értékeket jelez. Ekkor addig kell ismételni a modell becslést, amíg elfogadható eredményt nem kapunk.

2. Változók transzformálása

1. A modell nem teljesíti a normális eloszlást
2. A linearitás feltétele nem teljesül
3. Heteroszkedasztikus a modell

2.1. Nem normális eloszlás esetén

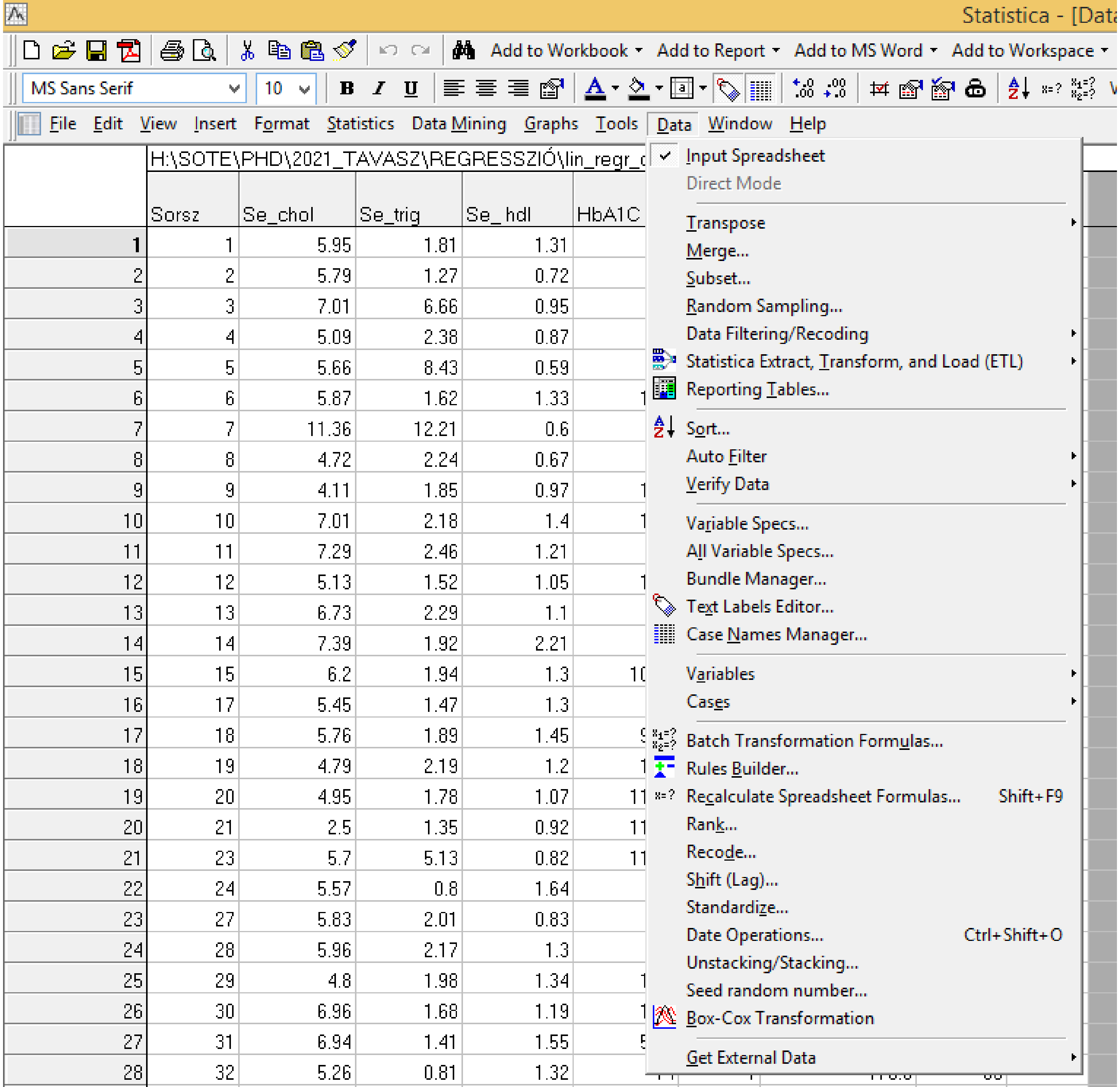
- A függő változó (y) transzformálása segíthet.
- y értékét y^{λ} értékére cserélhetjük.
- Arány, hányados esetén logit transzformáció
- λ gyakori értékei: -2; -1; -0,5; 0,5; 2 vagy log(y) balra ferde eloszlás esetén
- EZ NEVEZIK BOX-COX TRANSZORMÁCIÓNAK

2.2. Nemlinearitás esetén

- A magyarázó változók (x) transzformálása segíthet.
- A hatványkitevők becslése a
- BOX-TIDWELL TRANSZFORMÁCIÓ

2.3. Heteroszkedasztikus modell esetén

Függő változó (y) transzformációja, hatványkitevő keresése.


Statistica - [Data
Add to Workbook
Add to Report
Add to MS Word
Add to Workspace
MS Sans Serif
10
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help
H:\SOTE\PHD\2021_TAVASZ\REGRESSZIÓ\lin_regr_
Sorsz Se_chol Se_trig Se_hdl HbA1C
1 1 5.95 1.81 1.31
2 2 5.79 1.27 0.72
3 3 7.01 6.66 0.95
4 4 5.09 2.38 0.87
5 5 5.66 8.43 0.59
6 6 5.87 1.62 1.33
7 7 11.36 12.21 0.6
8 8 4.72 2.24 0.67
9 9 4.11 1.85 0.97
10 10 7.01 2.18 1.4
11 11 7.29 2.46 1.21
12 12 5.13 1.52 1.05
13 13 6.73 2.29 1.1
14 14 7.39 1.92 2.21
15 15 6.2 1.94 1.3
16 17 5.45 1.47 1.3
17 18 5.76 1.89 1.45
18 19 4.79 2.19 1.2
19 20 4.95 1.78 1.07
20 21 2.5 1.35 0.92
21 23 5.7 5.13 0.82
22 24 5.57 0.8 1.64
23 27 5.83 2.01 0.83
24 28 5.96 2.17 1.3
25 29 4.8 1.98 1.34
26 30 6.96 1.68 1.19
27 31 6.94 1.41 1.55
28 32 5.26 0.81 1.32
Input Spreadsheet
Direct Mode
Transpose
Merge...
Subset...
Random Sampling...
Data Filtering/Recoding
Statistica Extract, Transform, and Load (ETL)
Reporting Tables...
Sort...
Auto Filter
Verify Data
Variable Specs...
All Variable Specs...
Bundle Manager...
Text Labels Editor...
Case Names Manager...
Variables
Cases
Batch Transformation Formulas...
Rules Builder...
Recalculate Spreadsheet Formulas... Shift+F9
Rank...
Recode...
Shift (Lag)...
Standardize...
Date Operations... Ctrl+Shift+O
Unstacking/Stacking...
Seed random number...
Box-Cox Transformation
Get External Data

3. Változók törlése vagy hozzáadása

- Törlés – Multikollinearitás
- VIF (variance inflation factor) változása.

4. Egyéb modellek

- Multikollinearitás esetén – ridge regresszió
- Sok kiugró vagy befolyásos érték esetén – robusztus regresszió
- Nem normális eloszlás esetén – nemparaméteres regresszió
- Nemlinearitás esetén – nemlineáris regresszió
- Maradékok nem függetlenek – többlépcsős regresszió

A modell általánosíthatósága

- Hogyan működne a modellünk a valóságban, milyen pontos előrejelzést tehetünk vele?
- Szigorúbb validálási eljárások

Kereszt validálás (cross-validation)

- Különböző adatokon végezzük el a regressziós modell illesztését és a validációt.
- Paraméterek becslése („training sample”)
- Validáció („hold-out sample”)

- **Ridge regresszió:** torzított, de kisebb varianciájú becslőfüggvényt ad, mint a legkisebb négyzetek becslőfüggvénye.

Többdimenziós outlier kezelése	1. Mahalanobis-távolság, értéke mindig 16,27 2. Cook-távolság: 1 kritérium 3. Centered Leverage 4. DfBeta érték +/- 2 kritérium

- A négy módszer nem mindig hoz azonos eredményt, ekkor mérlegelni kell, kihagyjuk-e az adott elemet

Modellépítés

- Prediktorok kiválasztása
  - Hipotézis szerint (nem összevissza!)
  - Ha túl sok a prediktor, értelmezhetetlen lesz a modell.
  - Ha a prediktorok erősen korrelálnak egymással, használhatatlan lesz a modell.
- A modell felépítése
  - Jobb kisebb modellel kezdeni
  - Fokozatosan bővíteni

Modellépítés módszerei

- Enter
  - Minden prediktor egyszerre kerül a modellbe.
- Hierarchikus Enter
  - Több modell egymás után, egyre több prediktorral.
  - Először a hipotézis szerint a legerősebb hatású prediktor, azután a várhatóan kevésbé erős hatásúak

Forward módszer (lényegében a hierarchikus Entry módszer automatikus változata)

1. A szoftver kiválasztja azt a prediktort, ami a legerősebben korrelál a outcome-mal.
2. További prediktorokat aszerint választja, hogy milyen erős a szemiparciális korrálációjuk az outcome-mal. Erősebb előbb.
3. Addig folytatja, amíg nem marad olyan prediktor, ami szignifikánsan korrelál az outcome-mal.

Backward módszer

1. A szoftver az összes prediktort beteszi a modellbe, és kiszámolja a hatásukat.
2. Kiveszi azokat a prediktorokat, amiknek a hatása egy adott szint alatt van.
3. Addig folytatja, amíg a valamennyi modellben maradt prediktornak szignifikáns hatása van.

Stepwise módszer

A Forward és Backward kombinációja:

1. A szoftver egyenként adja a modellhez a prediktorokat.
2. Minden lépés után leteszteli, hogy a modell valamennyi prediktora szignifikáns-e.
3. Ha valamelyik prediktor már nem szignifikáns, kiveszi a modellből.

Milyen eljárást használjunk: Blockwise, Enter vagy Stepwise?

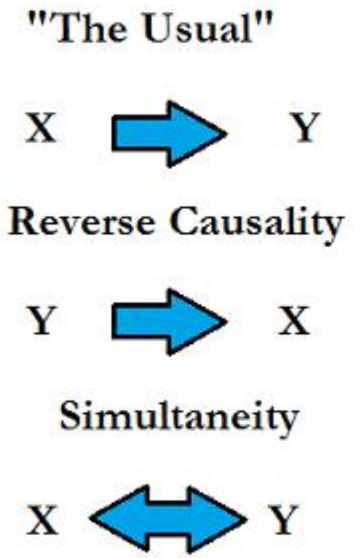
- **Ha előzetesen van elképzelésünk a megoldásról, akkor Blockwise (hierarchikus) módszer ajánlott:**
  - a számunkra fontos változókat hangsúlyozottan használhatjuk
- **Exploratív vizsgálatnál Enter módszer ajánlott:**
  - a változó modellbe kerülését nem befolyásolja az előzetes tudás vagy elképzelés,
  - pontos képet kapunk a függő és független változók kapcsolatáról, illetve a független változók kapcsolatrendszeréről.
- **Takarékos vagy gazdaságos (parsimonious) modell keresésénél Stepwise:**
  - a lehető legkevesebb prediktorral a lehető legjobb becslést tenni,
  - hátránya: a változókról önmagukban kevés információt kapunk,
  - overfitting veszély!
  - támadják: véletlen és matematikai döntéseken múlik, a kutatónak „nincs beleszólása” a modellbe.

Shrinkage módszerek

- **Ridge regresszió**
- **Lasso**
- **Legkisebb szög regresszió (2004)**
- **Bilineáris regresszió**

Megjegyzés: multikollinearitás esetén érdemes kipróbálni a Stepwise regresszió-analízist standardizált változókkal.

Az epidemiológiában az ún. ***instrumentális változók (IV, instrumental variables)*** módszerét használják az ok-okozati összefüggések becslésére, ha az ellenőrzött kísérletek nem megvalósíthatók, vagy ha egy kezelést nem végeznek el sikeresen minden egység számára egy randomizált kísérlet során.
Az IV-eket akkor alkalmazzuk, amikor egy „érdekes" magyarázó változó korrelál a hibataggal, amely esetben az ismert OLS (legkisebb négyzetek) és az ANOVA torzított eredményt adnak. Egy valós IV indukálja a magyarázó változó változásait, de nincs független hatása a függő változóra. A módszer lehetővé teszi a magyarázó változó tényleges oksági hatását a függő változóra.



Feladat: 2-es típusú diabateses betegeket vizsgáltak. A vizsgálat célja, hogy megállapítsák, a HDL (védő) koleszterin értékre a felsorolt változók milyen hatást fejtenek ki: testsúly, testmagasság, SE_chol, SE_trig, HbA1c, betegek neme, CRP értékek.
**Forrás: Prof. Dr. Füst György engedélyével.*

Megjegyzés: megemlítendő, hogy a független változók *(x_i)* bizonyos esetekben ordináris vagy nominális változók is lehetnek (*dummy* változó, például a „*nem*” mint a jelen példában).

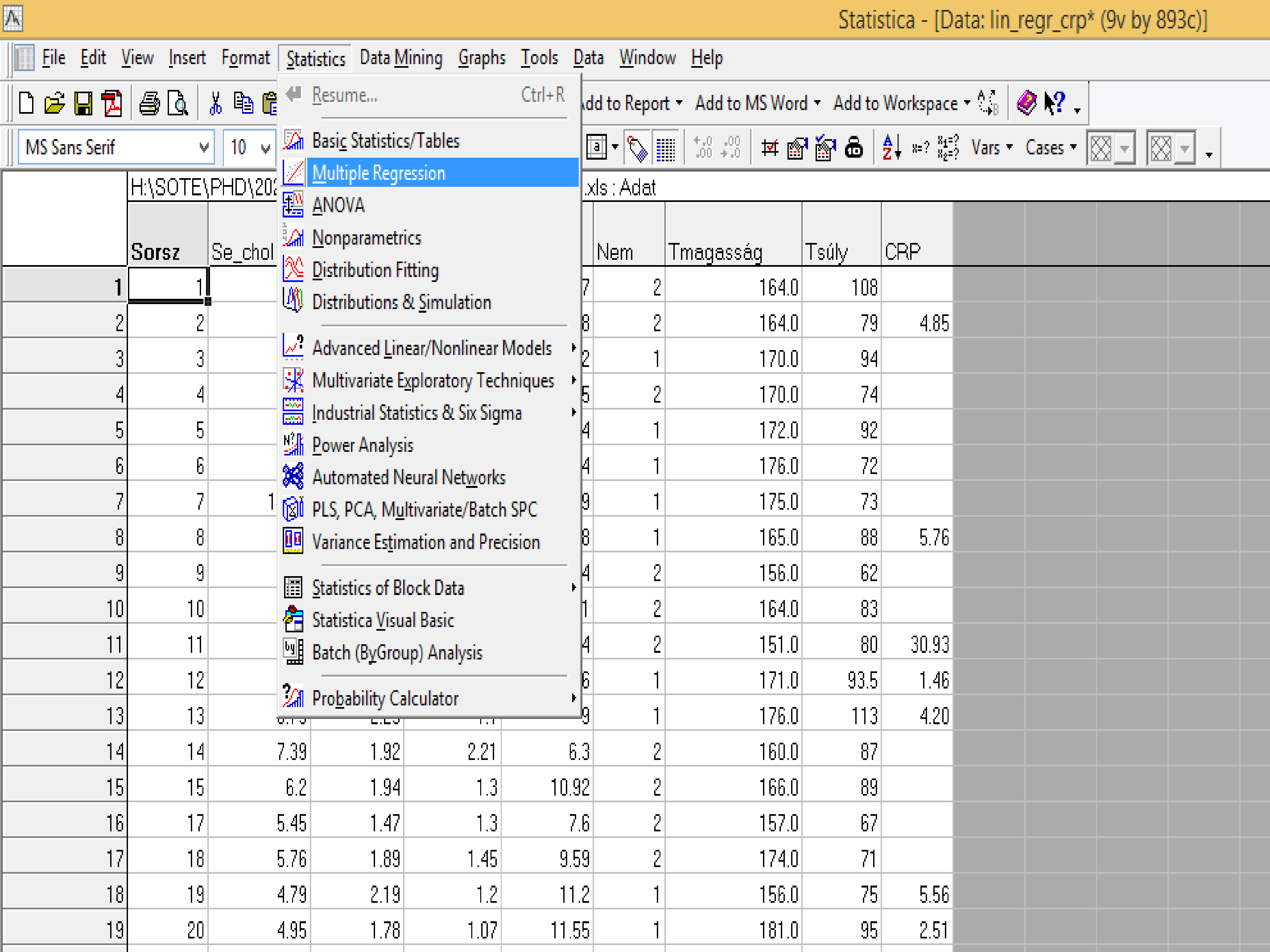
**Magyar Kardiovaszkuláris Konszenzus ajánlása, 2009*

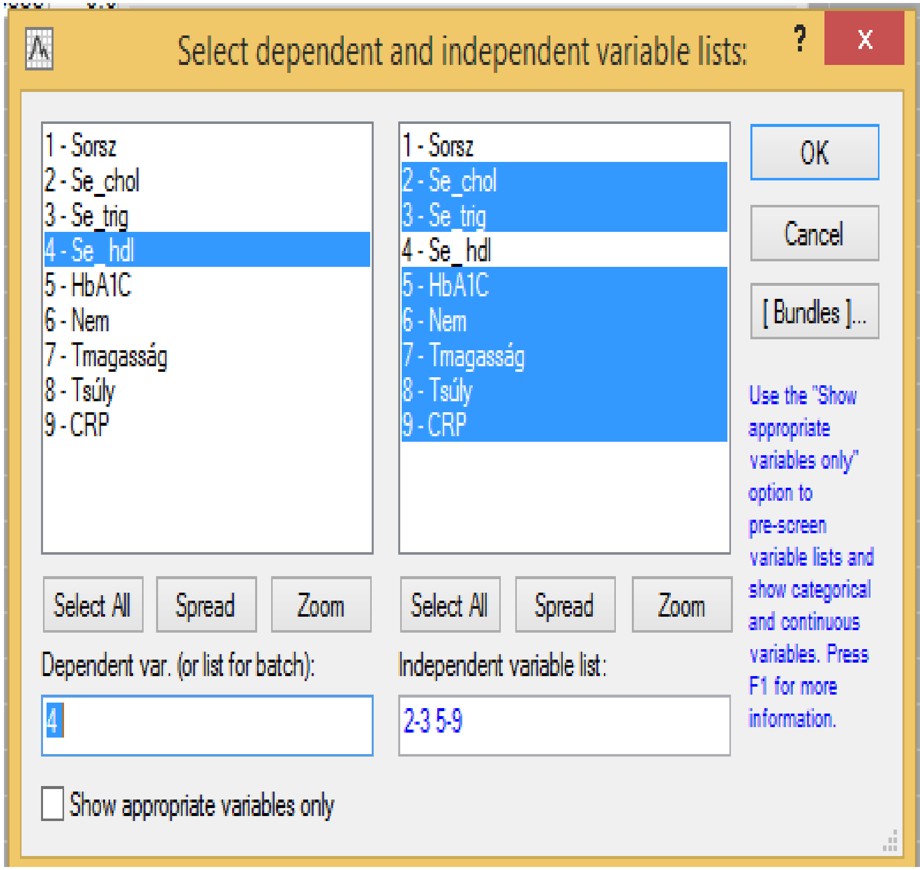
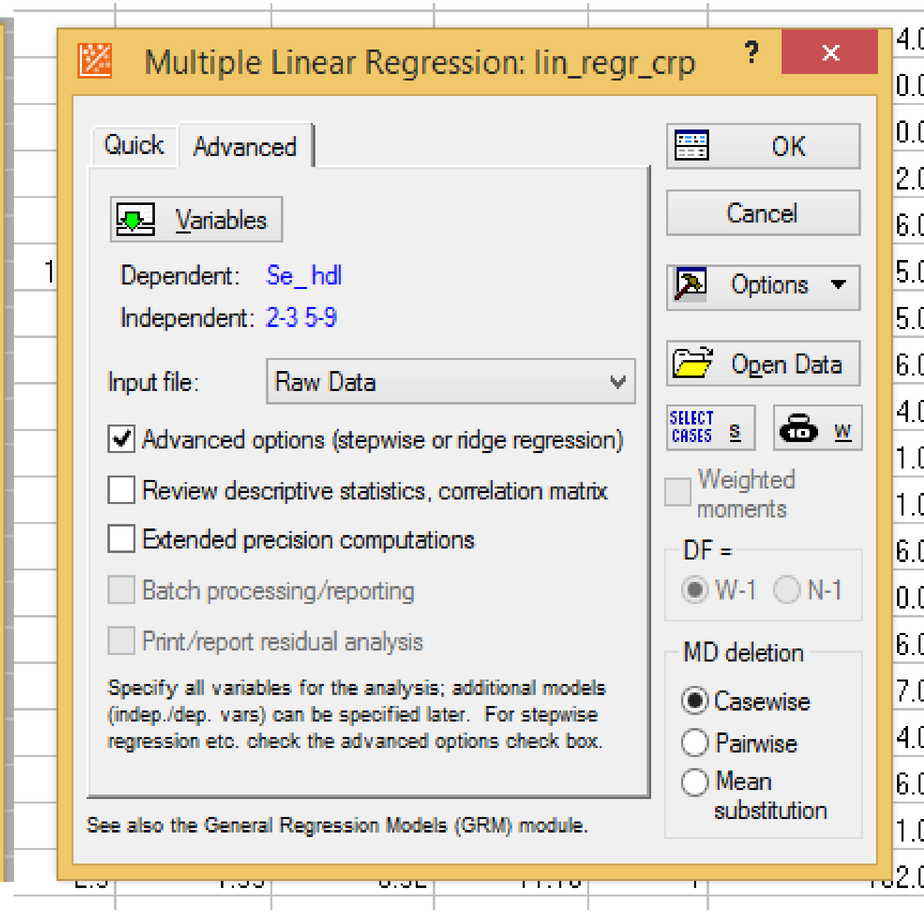
Optimális koleszterinértékek*

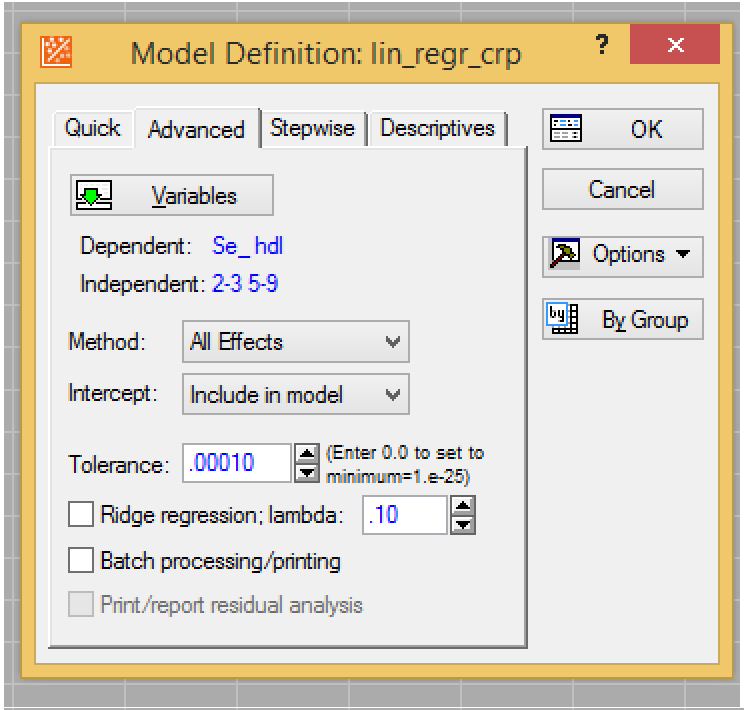
- összkoleszterin mennyiség: 5 mmol/l alatt
- LDL-koleszterin: 3 mmol/l alatt
- trigliceridek: 1,7 mmol/l alatt
- HDL-koleszterin:
  férfiak esetében 1 mmol/l felett
  nők esetében 1,3 mmol/l felett

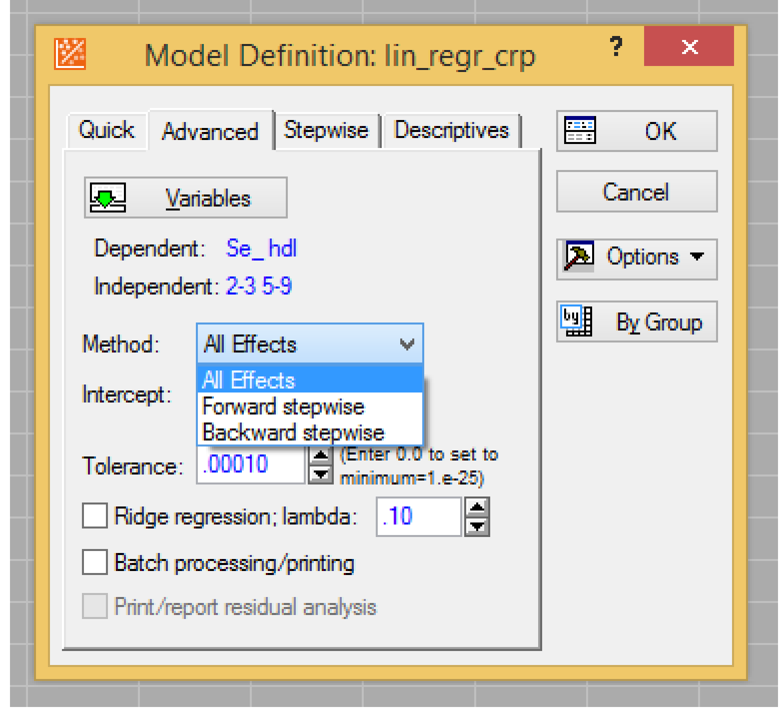
Intelmek (Remonstrances)

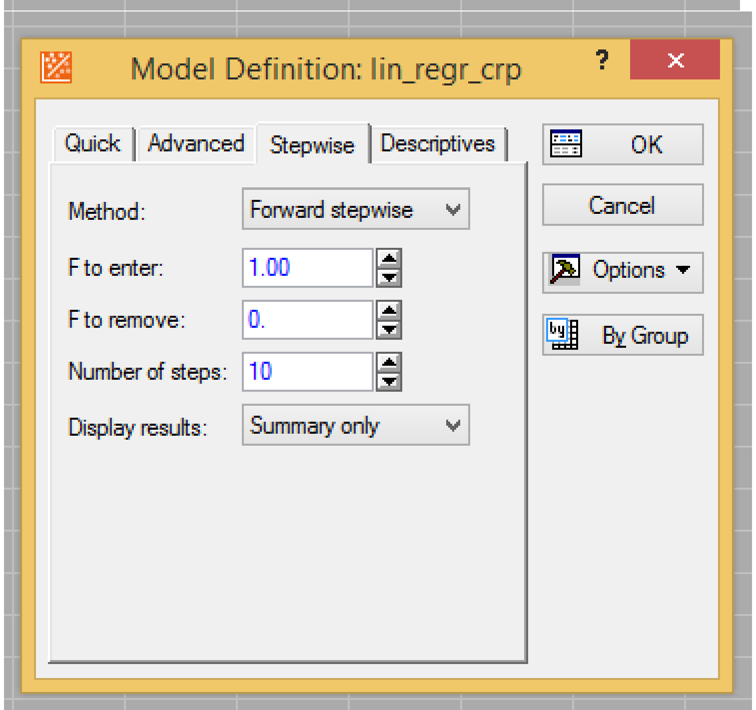
- Nincs recept a sokváltozós analízisre.
- Tapasztalatnak nagy a szerepe.
- A szoftverek összetettek. Excel is használható (korlátozottan).
- Nagy a felhasználói szabadság.
- Tengernyi adatot adnak a szoftverek.
- Szakmai <u>kontroll</u> a végeredményre!!!

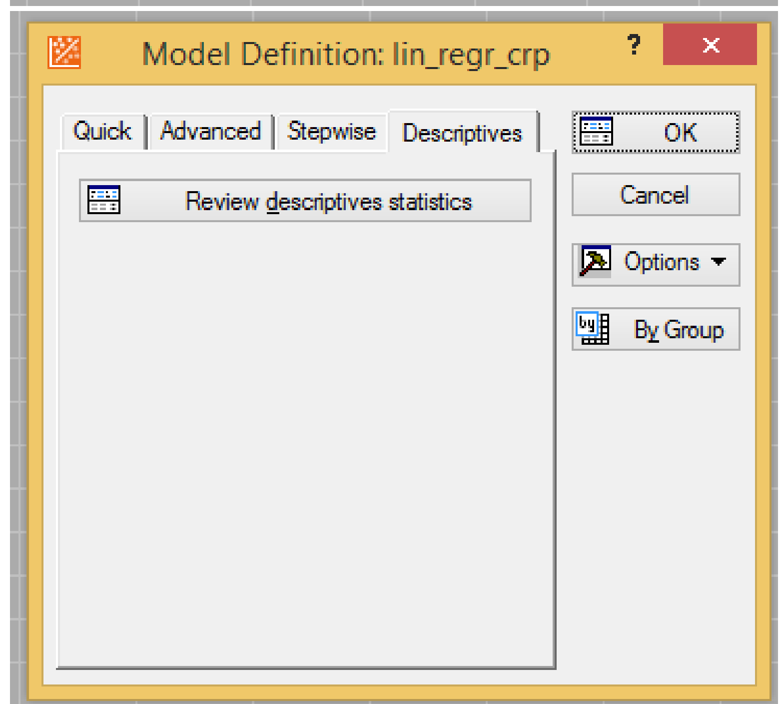

Statistica - [Data: lin_regr_crp* (9v by 893c)]
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help
Resume... Ctrl+R
Basic Statistics/Tables
Multiple Regression
ANOVA
Nonparametrics
Distribution Fitting
Distributions & Simulation
Advanced Linear/Nonlinear Models
Multivariate Exploratory Techniques
Industrial Statistics & Six Sigma
Power Analysis
Automated Neural Networks
PLS, PCA, Multivariate/Batch SPC
Variance Estimation and Precision
Statistics of Block Data
Statistica Visual Basic
Batch (ByGroup) Analysis
Probability Calculator
Add to Report
Add to MS Word
Add to Workspace
MS Sans Serif
10
Vars
Cases
H:\SOTE\PHD\202
.xls : Adat
Sorsz
Se_chol
Nem
Tmagasság
Tsúly
CRP


Select dependent and independent variable lists:
1 - Sorsz
2 - Se_chol
3 - Se_trig
4 - Se_hdl
5 - HbA1C
6 - Nem
7 - Tmagasság
8 - Tsúly
9 - CRP
1 - Sorsz
2 - Se_chol
3 - Se_trig
4 - Se_hdl
5 - HbA1C
6 - Nem
7 - Tmagasság
8 - Tsúly
9 - CRP
OK
Cancel
[Bundles]...
Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous variables. Press F1 for more information.
Select All
Spread
Zoom
Select All
Spread
Zoom
Dependent var. (or list for batch):
4
Independent variable list:
2-3 5-9
Show appropriate variables only

Multiple Linear Regression: lin_regr_crp
Quick
Advanced
Variables
Dependent: Se_hdl
Independent: 2-3 5-9
Input file: Raw Data
Advanced options (stepwise or ridge regression)
Review descriptive statistics, correlation matrix
Extended precision computations
Batch processing/reporting
Print/report residual analysis
Specify all variables for the analysis; additional models (indep./dep. vars) can be specified later. For stepwise regression etc. check the advanced options check box.
See also the General Regression Models (GRM) module.
OK
Cancel
Options
Open Data
SELECT CASES
Weighted moments
DF =
W-1
N-1
MD deletion
Casewise
Pairwise
Mean substitution

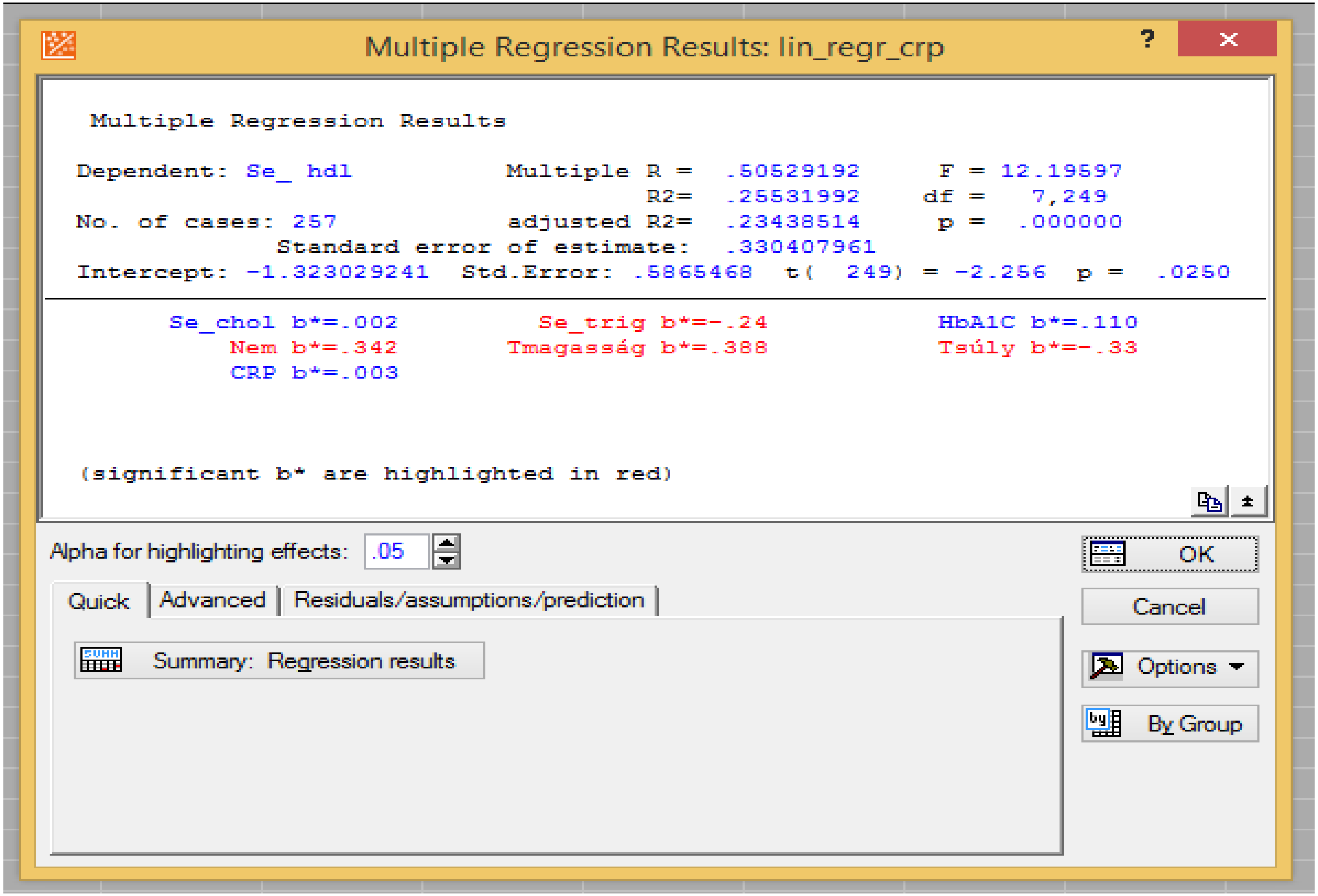

Model Definition: lin_regr_crp
Quick
Advanced
Stepwise
Descriptives
OK
Cancel
Options
By Group
Variables
Dependent: Se_hdl
Independent: 2-3 5-9
Method: All Effects
Intercept: Include in model
Tolerance: .00010
(Enter 0.0 to set to minimum=1.e-25)
Ridge regression; lambda: .10
Batch processing/printing
Print/report residual analysis


Model Definition: lin_regr_crp
Quick
Advanced
Stepwise
Descriptives
OK
Cancel
Options
By Group
Variables
Dependent: Se_hdl
Independent: 2-3 5-9
Method: All Effects
All Effects
Forward stepwise
Backward stepwise
Intercept:
Tolerance: .00010
(Enter 0.0 to set to minimum=1.e-25)
Ridge regression; lambda: .10
Batch processing/printing
Print/report residual analysis


Model Definition: lin_regr_crp
Quick
Advanced
Stepwise
Descriptives
OK
Cancel
Options
By Group
Method: Forward stepwise
F to enter: 1.00
F to remove: 0.
Number of steps: 10
Display results: Summary only


Model Definition: lin_regr_crp
Quick
Advanced
Stepwise
Descriptives
OK
Cancel
Options
By Group
Review descriptives statistics

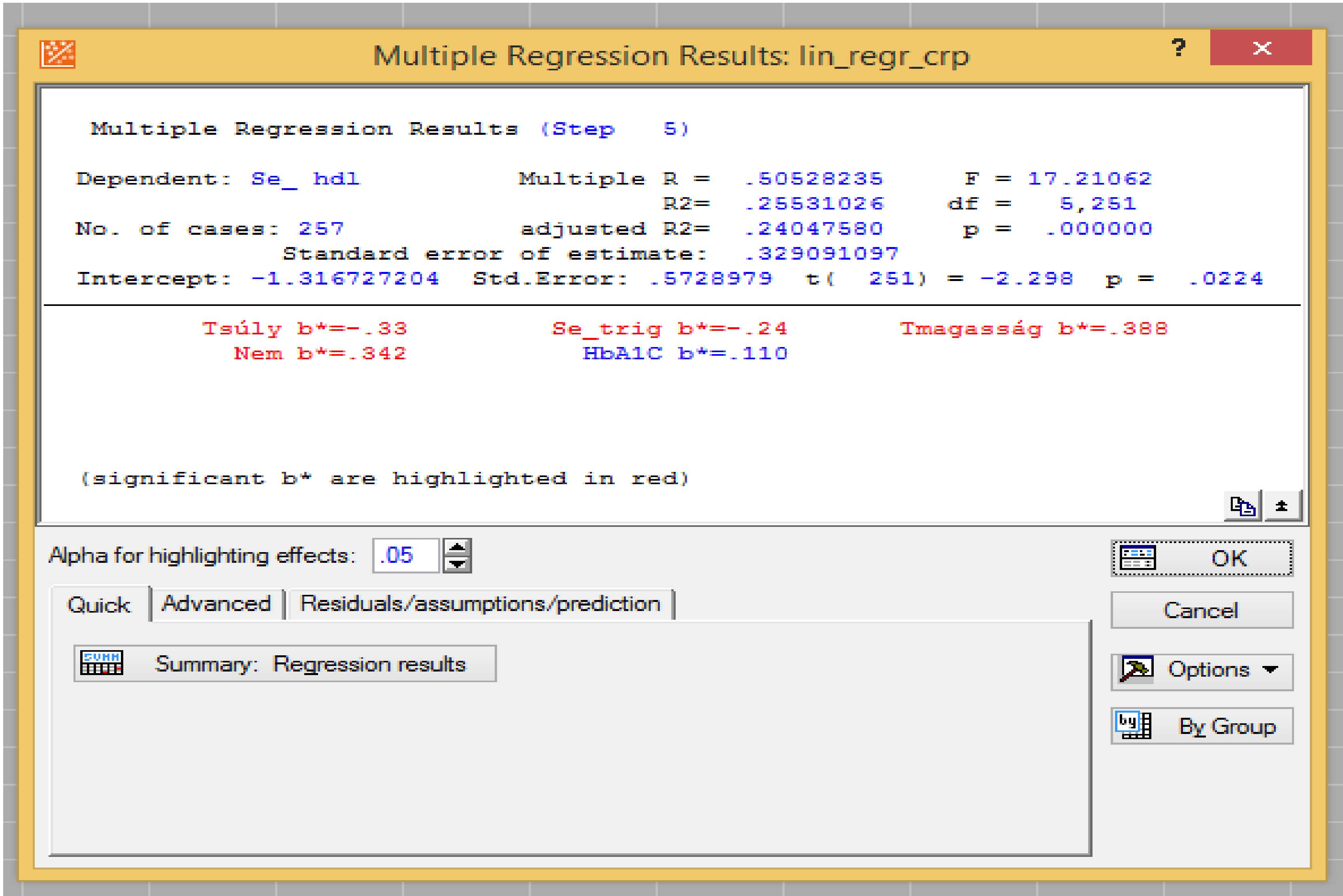
All effects



Analysis of Variance; DV: Se_ hdl (lin_regr_crp)

Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	9.31999	7	1.331427	12.19597	0.000000
Residual	27.18319	249	0.109169		
Total	36.50317				

Stepwise regression



N=257	Regression Summary for Dependent Variable: Se_ hdl (lin_regr_crp) R= .50529192 R2= .25531992 Adjusted R2= .23438514 F(7,249)=12.196 p<.00000 Std.Error of estimate: .33041					
	b*	Std.Err. of b*	b	Std.Err. of b	t(249)	p-value
Intercept			-1.32303	0.586547	-2.25562	0.024963
Se_chol	0.001990	0.067716	0.00057	0.019549	0.02939	0.976576
Se_trig	-0.241563	0.067505	-0.03141	0.008777	-3.57843	0.000415
HbA1C	0.110348	0.056421	0.01500	0.007670	1.95580	0.051607
Nem	0.341754	0.075754	0.25793	0.057173	4.51139	0.000010
Tmagasság	0.388073	0.077893	0.01592	0.003196	4.98216	0.000001
Tsúly	-0.329105	0.062992	-0.00701	0.001342	-5.22456	0.000000
CRP	0.002585	0.055289	0.00010	0.002033	0.04675	0.962747

Variable	Variables currently in the Equation; DV: Se_ hdl (lin_regr_crp)						
	b* in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(249)	p-value
Se_chol	0.001990	0.001863	0.001607	0.652213	0.347787	0.02939	0.976576
Se_trig	-0.241563	-0.221158	-0.195694	0.656288	0.343712	-3.57843	0.000415
HbA1C	0.110348	0.123003	0.106957	0.939479	0.060521	1.95580	0.051607
Nem	0.341754	0.274884	0.246715	0.521150	0.478850	4.51139	0.000010
Tmagasság	0.388073	0.301081	0.272460	0.492922	0.507078	4.98216	0.000001
Tsúly	-0.329105	-0.314313	-0.285716	0.753705	0.246295	-5.22456	0.000000
CRP	0.002585	0.002963	0.002557	0.978355	0.021645	0.04675	0.962747

Variable	Redundancy of Independent Variables; DV: Se_ hdl (lin_regr_crp) R-square column contains R-square of respective variable with all other independent variables			
	Toleran.	R-square	Partial Cor.	Semipart Cor.
Se_chol	0.652213	0.347787	0.001863	0.001607
Se_trig	0.656288	0.343712	-0.221158	-0.195694
HbA1C	0.939479	0.060521	0.123003	0.106957
Nem	0.521150	0.478850	0.274884	0.246715
Tmagasság	0.492922	0.507078	0.301081	0.272460
Tsúly	0.753705	0.246295	-0.314313	-0.285716
CRP	0.978355	0.021645	0.002963	0.002557

Standardizált regressziós koefficiens b* (általában Beta, path coefficient) lehetővé teszi a regressziós koefficiensek korrekt összehasoníthatóságát. Az a független változó gyakorolja a legnagyobb hatást a függő változóra, melyre ez az érték a legmagasabb. A „szabványosított” együtthatók a mértékegység nélküliek.

A mértékegység megváltoztatásával ezek is változhatnak.
Ettől a jelenségtől szeretnénk megszabadulni. Azaz **standardizálunk**, ami azt jelenti, hogy 0 várható értékűvé, és 1 szórásúvá transzformáljuk az együtthatókat.

Első lehetőség: az adatbázis standardizálása, majd a becslések lefuttatása, melynek eredményeképp megkapjuk a standardizált regressziós együtthatókat.

Második lehetőség: ehelyett érvényes a

$$b_i^* = b_i \cdot \sqrt{\frac{Var(X_i)}{Var(Y)}}$$

összefüggés, ahol $Var(.)$ a változó szórásnégyzetét jelöli. Azaz nem kell standardizálni a teljes adatbázist a standardizált együtthatók előállításához.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	9.31999	1.33143	12.20	<.0001
Error	249	27.18319	0.10917		
Corrected Total	256	36.50317			

Root MSE	0.33041	R-Square	0.2553
Dependent Mean	1.27786	Adj R-Sq	0.2344
Coeff Var	25.85635		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	-1.32303	0.58655	-2.26	0.0250	.	0
Se_chol	1	0.00057457	0.01955	0.03	0.9766	0.65221	1.53324
Se_trig	1	-0.03141	0.00878	-3.58	0.0004	0.65629	1.52372
HbA1C	1	0.01500	0.00767	1.96	0.0516	0.93948	1.06442
Nem	1	0.25793	0.05717	4.51	<.0001	0.52115	1.91883
Tmagasság	1	0.01592	0.00320	4.98	<.0001	0.49292	2.02872
Tsúly	1	-0.00701	0.00134	-5.22	<.0001	0.75371	1.32678
CRP	1	0.00009504	0.00203	0.05	0.9627	0.97835	1.02212

SAS output ellenőrzésre

A *Variance Inflation* oszlop alapján a *Nem*, a *Tmagasság* változók nem szerepelhetnek egy modellben, mert nem függetlenek egymástól.

Variable	Variables currently in the Equation; DV: Se_ hdl (lin_regr_crp)						
	b* in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(249)	p-value
Se_chol	0.001990	0.001863	0.001607	0.652213	0.347787	0.02939	0.976576
Se_trig	-0.241563	-0.221158	-0.195694	0.656288	0.343712	-3.57843	0.000415
HbA1C	0.110348	0.123003	0.106957	0.939479	0.060521	1.95580	0.051607
Nem	0.341754	0.274884	0.246715	0.521150	0.478850	4.51139	0.000010
Tmagasság	0.388073	0.301081	0.272460	0.492922	0.507078	4.98216	0.000001
Tsúly	-0.329105	-0.314313	-0.285716	0.753705	0.246295	-5.22456	0.000000
CRP	0.002585	0.002963	0.002557	0.978355	0.021645	0.04675	0.962747

Sweep mátrix használata a VIF érték megállapítására

	Current Status of Sweep Matrix; DV: Se_ hdl (lin_regr_crp)							
	Se_chol	Se_trig	HbA1C	Nem	Tmagasság	Tsúly	CRP	Se_ hdl
Se_chol	-1.53324	0.86208	0.05356	0.12405	-0.02398	0.15402	0.07723	0.001990
Se_trig	0.86208	-1.52372	0.03827	-0.14154	-0.06625	0.09355	-0.05232	-0.241563
HbA1C	0.05356	0.03827	-1.06442	0.21183	0.02617	0.20606	-0.00722	0.110348
Nem	0.12405	-0.14154	0.21183	-1.91883	-1.22421	-0.16409	-0.12234	0.341754
Tmagasság	-0.02398	-0.06625	0.02617	-1.22421	-2.02872	0.45128	-0.14329	0.388073
Tsúly	0.15402	0.09355	0.20606	-0.16409	0.45128	-1.32678	-0.06858	-0.329105
CRP	0.07723	-0.05232	-0.00722	-0.12234	-0.14329	-0.06858	-1.02212	0.002585
Se_ hdl	0.00199	-0.24156	0.11035	0.34175	0.38807	-0.32911	0.00258	0.744680

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	9.31999	1.33143	12.20	<.0001
Error	249	27.18319	0.10917		
Corrected Total	256	36.50317			

Root MSE	0.33041	R-Square	0.2553
Dependent Mean	1.27786	Adj R-Sq	0.2344
Coeff Var	25.85635		

SAS output

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	-1.32303	0.58655	-2.26	0.0250	.	0
Se_chol	1	0.00057457	0.01955	0.03	0.9766	0.65221	1.53324
Se_trig	1	-0.03141	0.00878	-3.58	0.0004	0.65629	1.52372
HbA1C	1	0.01500	0.00767	1.96	0.0516	0.93948	1.06442
Nem	1	0.25793	0.05717	4.51	<.0001	0.52115	1.91883
Tmagasság	1	0.01592	0.00320	4.98	<.0001	0.49292	2.02872
Tsuly	1	-0.00701	0.00134	-5.22	<.0001	0.75371	1.32678
CRP	1	0.00009504	0.00203	0.05	0.9627	0.97835	1.02212

VIF meghatározása

A magyarázó változók korrelációs mátrixának inverzéből. A főátló elemei.

Új modell

N=892	Regression Summary for Dependent Variable: Se_ hdl (lin_regr_crp) R= .44393381 R2= .19707723 Adjusted R2= .19436465 F(3,888)=72.653 p<0.0000 Std.Error of estimate: .37154					
	b*	Std.Err. of b*	b	Std.Err. of b	t(888)	p-value
Intercept			1.625968	0.080170	20.28151	0.000000
Nem	0.156605	0.032157	0.129734	0.026639	4.86999	0.000001
Se_trig	-0.251311	0.031006	-0.043717	0.005394	-8.10526	0.000000
Tsúly	-0.220937	0.032896	-0.004945	0.000736	-6.71614	0.000000

Variable	Redundancy of Independent Variables; DV: Se_ hdl (lin_regr_crp) R-square column contains R-square of respective variable with all other independent variables			
	Toleran.	R-square	Partial Cor.	Semipart Cor.
Nem	0.874393	0.125607	0.161287	0.146440
Se_trig	0.940527	0.059473	-0.262459	-0.243723
Tsúly	0.835533	0.164467	-0.219864	-0.201953

Keresett modell:

$$\hat{y} = 1.626 + (0.13 \cdot Nem) - (0.044 \cdot \mathrm{Se\_trig}) - (0.005 \cdot \mathrm{Tsuly})$$

Következtetések:

1. A triglicerid és a HDL között fordított kapcsolat van: növekedett trigliceridszint mellett csökken a HDL szintje vagy fordítva.
2. Testsúlyra vonatkozóan: hasonló a fentihez a megállapítás.
3. Nemekre vonatkozóan: nők esetében magasabb a HDL érték.

R-Square = 0.1377

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	21.01847	10.50924	70.97	<.0001
Error	889	131.64760	0.14809		
Corrected Total	891	152.66608			

	Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
	Intercept	1.62010	0.08303	56.37631	380.70	<.0001
*	Nem	0.13796	0.02757	3.70735	25.04	<.0001
*	Tsuly	-0.00622	0.00074498	10.33287	69.78	<.0001

* Forced into the model by the INCLUDE= option

Error tagban a változás: 0.01!
Megfontolandó a Se_trig modellbeli szerepe!

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	30.08701	10.02900	72.65	<.0001
Error	888	122.57907	0.13804		
Corrected Total	891	152.66608			

Root MSE	0.37154	R-Square	0.1971
Dependent Mean	1.33650	Adj R-Sq	0.1944
Coeff Var	27.79930		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	1.62597	0.08017	20.28	<.0001	0
Nem	1	0.12973	0.02664	4.87	<.0001	1.14365
Tsúly	1	-0.00495	0.00073634	-6.72	<.0001	1.19684
Se_trig	1	-0.04372	0.00539	-8.11	<.0001	1.06323

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Proportion of Variation			
			Intercept	Nem	Tsúly	Se_trig
1	3.41360	1.00000	0.00199	0.00686	0.00357	0.02761
2	0.47184	2.68972	0.00287	0.02278	0.00212	0.88103
3	0.09912	5.86854	0.00688	0.48690	0.18605	0.08790
4	0.01544	14.86703	0.98826	0.48346	0.80826	0.00346

Se-Trig szerepe a modellban

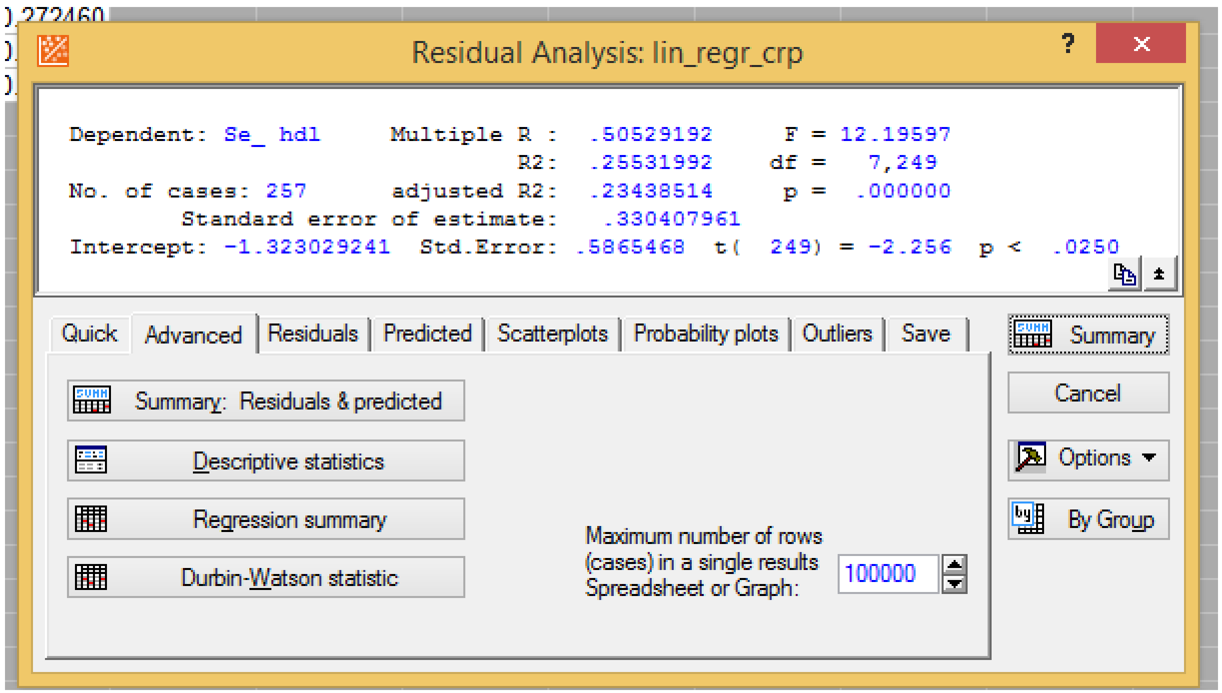
Variable	Descriptive Statistics (lin_regr_crp) Valid N	Mean	Confidence -95.000%	Confidence 95.000%	Median	Minimum	Maximum	Std.Dev.
Modell1	892	1.332039	1.319875	1.344202	1.351830	0.04786	1.661500	0.185096
Modell2	893	1.338274	1.328217	1.348332	1.336400	0.76600	1.672180	0.153133
Differencia	892	-0.006229	-0.012902	0.000444	0.016340	-1.11240	0.140240	0.101551

Modell1:

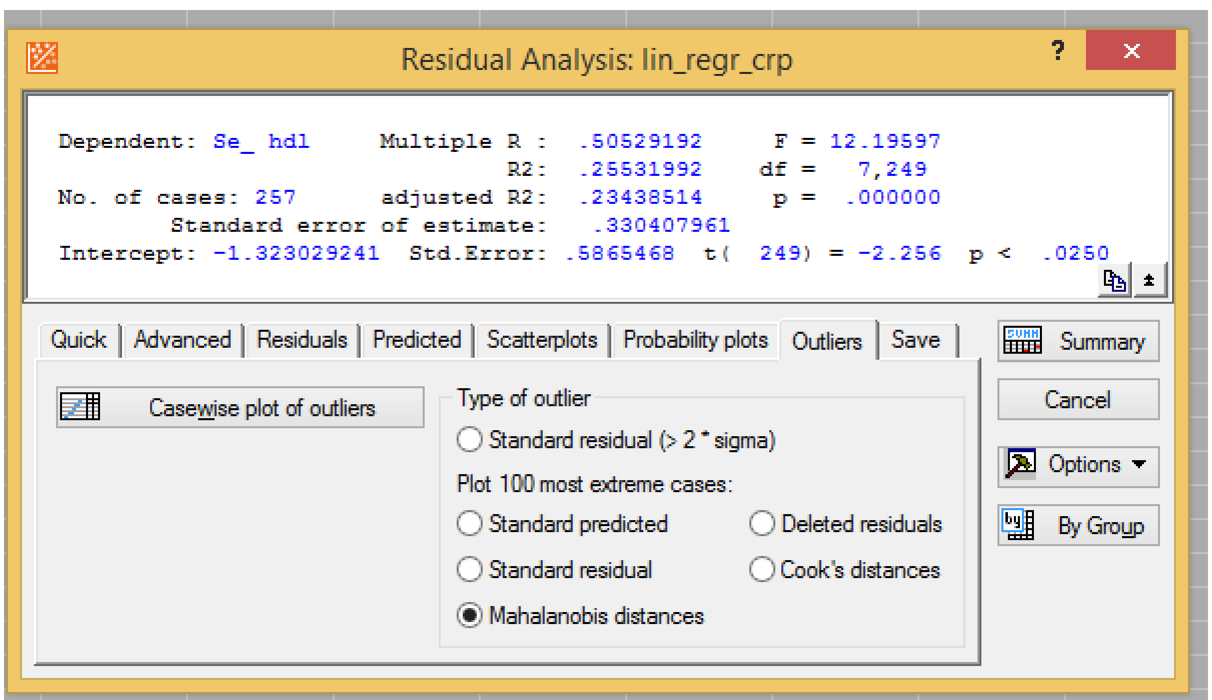
$$\hat{y} = 1.626 + (0.13 \cdot Nem) - (0.044 \cdot \mathrm{Se\_trig}) - (0.005 \cdot \mathrm{Tsuly})$$

Modell2:

$$\hat{y} = 1.62 + (0.138 \cdot \mathrm{Nem}) - (0.0062 \cdot \mathrm{Tsuly})$$



	Durbin-Watson d (lin_regr_crp) and serial correlation of residuals	
	Durbin-Watson d	Serial Corr.
Estimate	1.656376	0.164376



Mahalanobis distances: Se_ hdl (lin_regr_crp)
Sorted

Case	Mahalanobis distances 5.55 . . . . . 123.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
317	. . . | . . *	1.210000	1.269818	-0.059818	-0.04215	-0.18104	0.230407	123.4927	-0.116442	0.007550
234	. . . | . *. .	0.980000	0.801504	0.178496	-2.49657	0.54023	0.207097	99.5784	0.294000	0.038882
30	. . . * | . . .	0.800000	0.775970	0.024030	-2.63039	0.07273	0.149741	51.5839	0.030241	0.000215
228	. . *. | . . .	1.030000	0.508131	0.521869	-4.03413	1.57947	0.131862	39.7776	0.620735	0.070269
258	. . *. | . . .	1.600000	0.823338	0.776662	-2.38214	2.35062	0.131469	39.5348	0.922756	0.154358
58	. .* . | . . .	0.850000	1.239896	-0.389896	-0.19897	-1.18004	0.113306	29.1091	-0.441858	0.026289
77	. * . | . . .	1.210000	1.165127	0.044873	-0.59083	0.13581	0.106195	25.4489	0.050043	0.000296
252	. *. . | . . .	1.060000	0.749832	0.310168	-2.76738	0.93874	0.095448	20.3673	0.338408	0.010943
329	. *. . | . . .	2.000000	1.588533	0.411467	1.62823	1.24533	0.089897	17.9548	0.444362	0.016737
134	. * . . | . . .	1.090000	1.268152	-0.178152	-0.05088	-0.53919	0.086812	16.6763	-0.191362	0.002895
109	. * . . | . . .	0.640000	0.952421	-0.312421	-1.70562	-0.94556	0.083183	15.2298	-0.333563	0.008075
267	. * . . | . . .	1.050000	1.017131	0.032869	-1.36647	0.09948	0.083061	15.1822	0.035086	0.000089
296	. * . . | . . .	1.720000	1.216158	0.503842	-0.32338	1.52491	0.081339	14.5185	0.536347	0.019962
261	. * . . | . . .	0.560000	1.414955	-0.854955	0.71851	-2.58757	0.080449	14.1806	-0.908834	0.056068
123	. * . . | . . .	1.220000	0.851142	0.368858	-2.23642	1.11637	0.080071	14.0385	0.391872	0.010326
129	. * . . | . . .	1.240000	1.028765	0.211235	-1.30550	0.63932	0.079804	13.9384	0.224321	0.003361
131	.* . . | . . .	0.800000	1.301298	-0.501298	0.12284	-1.51721	0.078668	13.5160	-0.531423	0.018331
262	.* . . | . . .	1.380000	1.440587	-0.060587	0.85285	-0.18337	0.077493	13.0860	-0.064113	0.000259
177	.* . . | . . .	1.900000	1.781567	0.118433	2.63992	0.35845	0.077394	13.0500	0.125309	0.000986
70	.* . . | . . .	0.910000	0.963598	-0.053598	-1.64704	-0.16222	0.076518	12.7337	-0.056636	0.000197
278	.* . . | . . .	1.090000	0.950036	0.139964	-1.71812	0.42361	0.076197	12.6188	0.147826	0.001331
11	.* . . | . . .	1.210000	1.092186	0.117814	-0.97312	0.35657	0.076009	12.5516	0.124398	0.000938
190	.* . . | . . .	1.260000	1.174352	0.085648	-0.54248	0.25922	0.075851	12.4954	0.090412	0.000493
90	.* . . | . . .	1.760000	1.616961	0.143039	1.77722	0.43292	0.074898	12.1587	0.150787	0.001338
288	.* . . | . . .	1.610000	1.412363	0.197637	0.70493	0.59816	0.074397	11.9833	0.208193	0.002516
310	.* . . | . . .	1.400000	1.668786	-0.268786	2.04884	-0.81350	0.073650	11.7239	-0.282840	0.004551
106	.* . . | . . .	1.000000	0.926397	0.073603	-1.84201	0.22276	0.073142	11.5490	0.077396	0.000336

Multiple Regression Results: lin_regr_crp

```
Ridge Regression Results

Dependent: Se_ hdl          Multiple R =  .46846790      F = 10.00154
                                   R2=  .21946217     df =   7,249
No. of cases: 257           adjusted R2=  .19751934      p =  .000000
            Standard error of estimate:  .338269322
Intercept:  -.574667697  Std.Error: .5303409  t(  249) = -1.084  p =  .2796

      Se_chol b*=-.02        Se_trig b*=-.22          HbA1C b*=.099
          Nem b*=.265      Tmagasság b*=.286          Tsúly b*=-.28
          CRP b*=-.00

(significant b* are highlighted in red)
```

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

Summary: Regression results

OK

Cancel

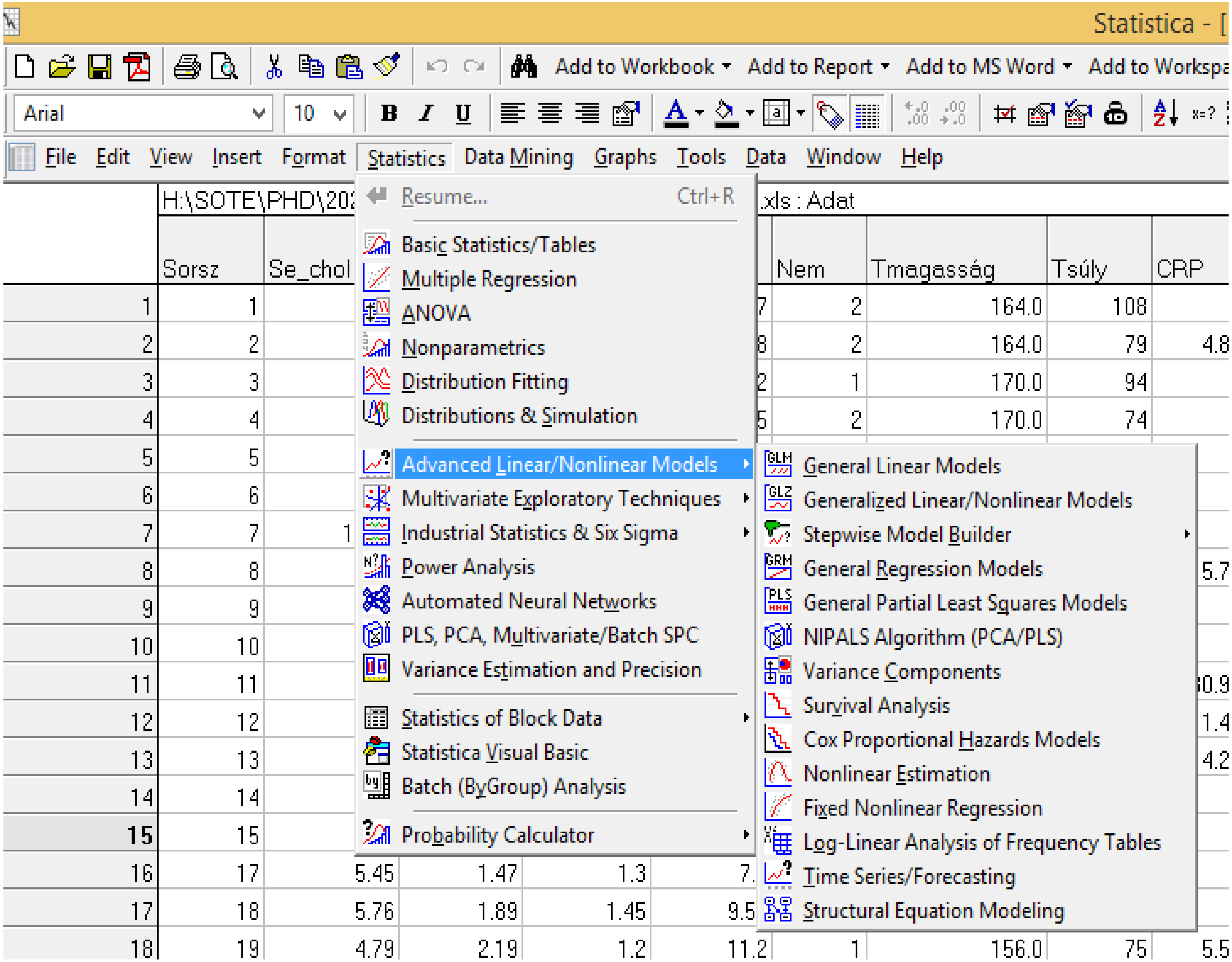
Options

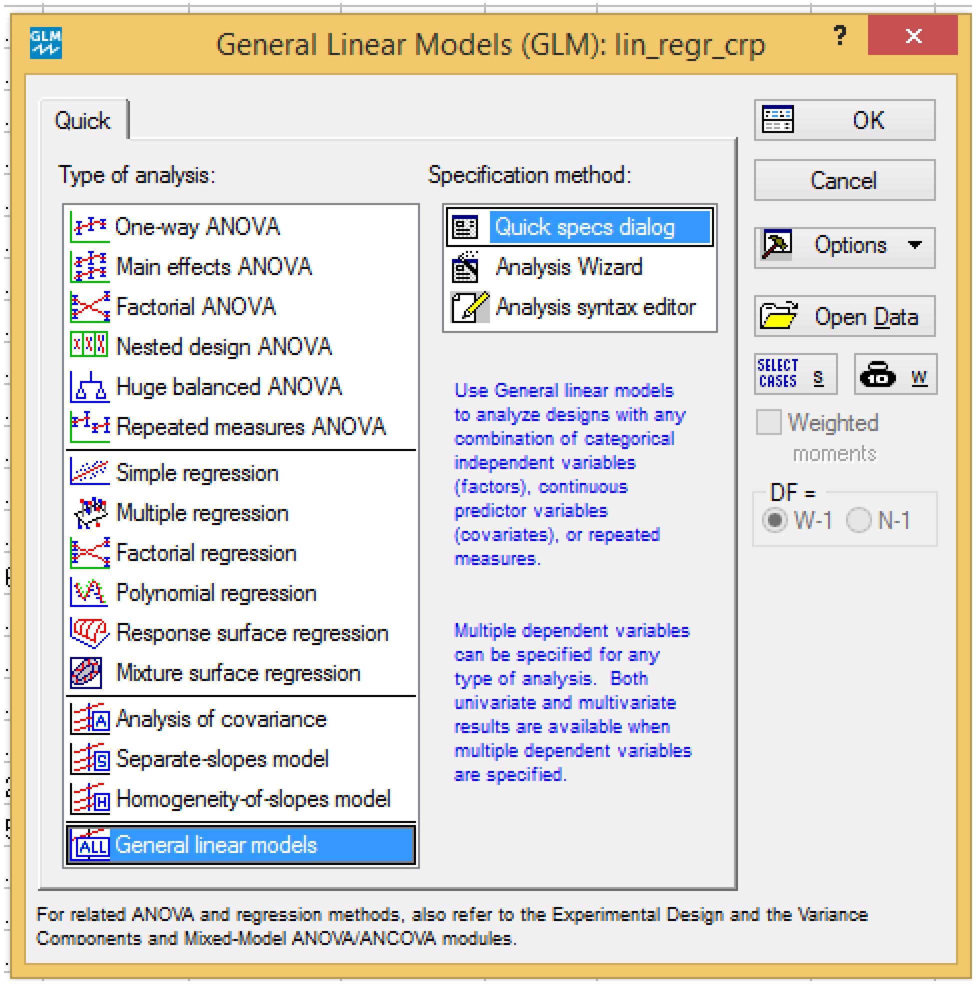
By Group

Ridge Regression Summary for Dependent Variable: Se_ hdl (lin_regr_crp)
l=.10000 R= .46846790 R2= .21946217 Adjusted R2= .19751934
F(7,249)=10.002 p<.00000 Std.Error of estimate: .33827

N=257	b*	Std.Err. of b*	b	Std.Err. of b	t(249)	p-value
Intercept			-0.574668	0.530341	-1.08358	0.279598
Se_chol	-0.015616	0.063280	-0.004508	0.018268	-0.24678	0.805283
Se_trig	-0.217658	0.063124	-0.028301	0.008208	-3.44811	0.000663
HbA1C	0.099391	0.054731	0.013512	0.007440	1.81600	0.070574
Nem	0.265417	0.068869	0.200315	0.051977	3.85395	0.000148
Tmagasság	0.286452	0.070458	0.011754	0.002891	4.06555	0.000064
Tsúly	-0.283179	0.060031	-0.006033	0.001279	-4.71724	0.000004
CRP	-0.001704	0.053824	-0.000063	0.001979	-0.03167	0.974764

Bonyolultabb eljárások




General Linear Models (GLM): lin_regr_crp
Quick
Type of analysis:
One-way ANOVA
Main effects ANOVA
Factorial ANOVA
Nested design ANOVA
Huge balanced ANOVA
Repeated measures ANOVA
Simple regression
Multiple regression
Factorial regression
Polynomial regression
Response surface regression
Mixture surface regression
Analysis of covariance
Separate-slopes model
Homogeneity-of-slopes model
General linear models
Specification method:
Quick specs dialog
Analysis Wizard
Analysis syntax editor
Use General linear models to analyze designs with any combination of categorical independent variables (factors), continuous predictor variables (covariates), or repeated measures.
Multiple dependent variables can be specified for any type of analysis. Both univariate and multivariate results are available when multiple dependent variables are specified.
OK
Cancel
Options
Open Data
SELECT CASES S
W
Weighted moments
DF =
W-1
N-1
For related ANOVA and regression methods, also refer to the Experimental Design and the Variance Components and Mixed-Model ANOVA/ANCOVA modules.

Köszönöm figyelmüket!