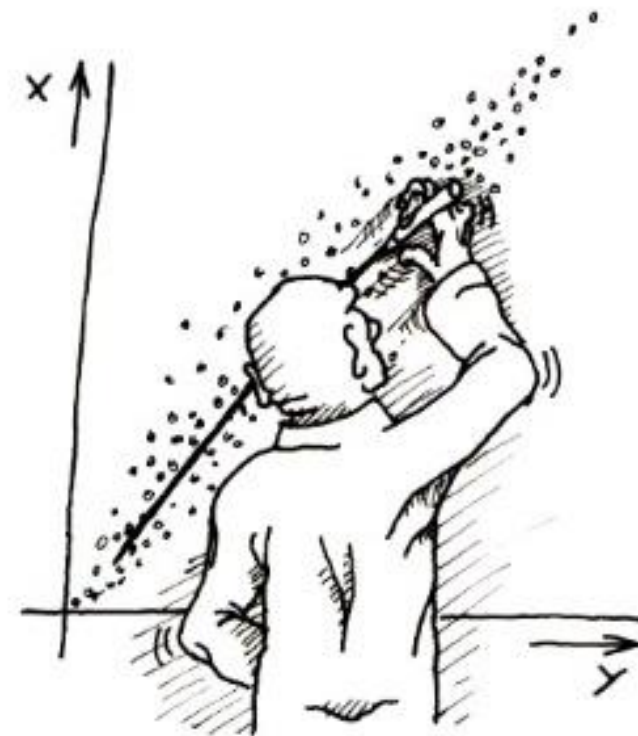


Többváltozós statisztika, regressziós eljárások (BMNPS07600M)

Készítette: Soltész-Várhelyi Klára

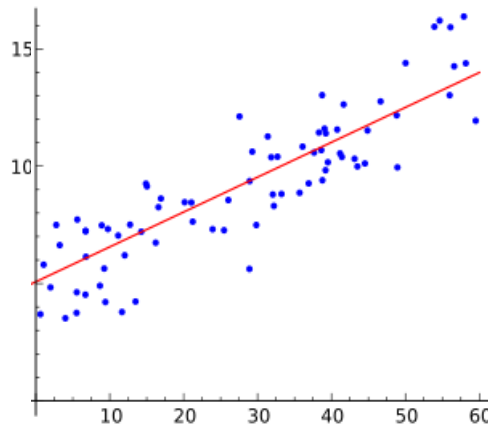
Lineáris regresszió

Elmélet



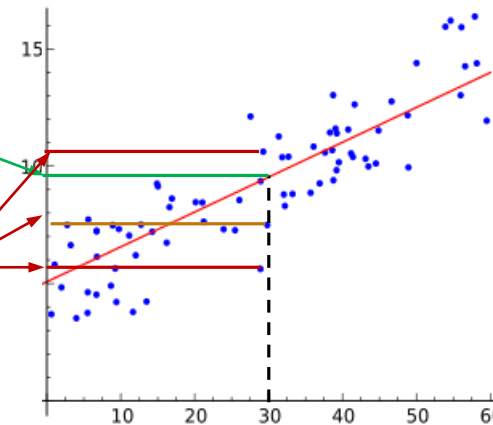
Lineáris Regresszió

- Modellt építünk az adatainkra, mely segítségével bejósoljuk a prediktor (X) változókból a kimenetet (Y)
 - A predikált érték számítása, a lineáris regresszió egyenlete
 - $Y_{\text{pred},i} = b_0 + b_1 * X_i$
 - **Kimeneti érték** = Model segítségével a **Prediktor érték**hez **Predikált érték** + **Hiba**
 - $Y_i = Y_{\text{pred},i} + \epsilon_i = b_0 + b_1 * X_i + \epsilon_i$



30-as prediktor értékhez
tartozó predikált érték

30-as prediktor értékhez
tartozó mért értékek

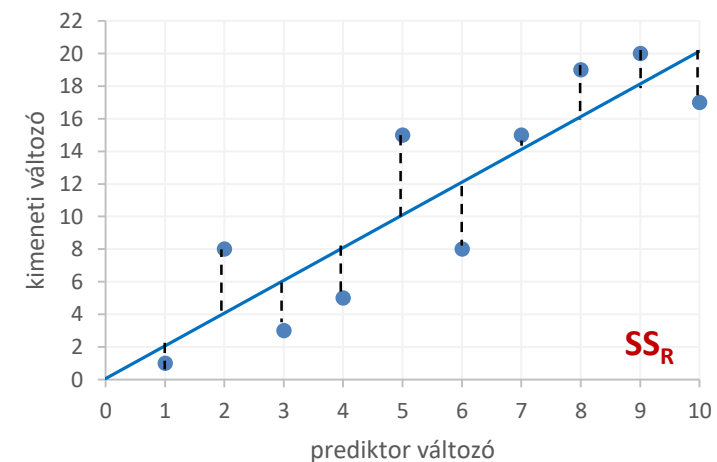
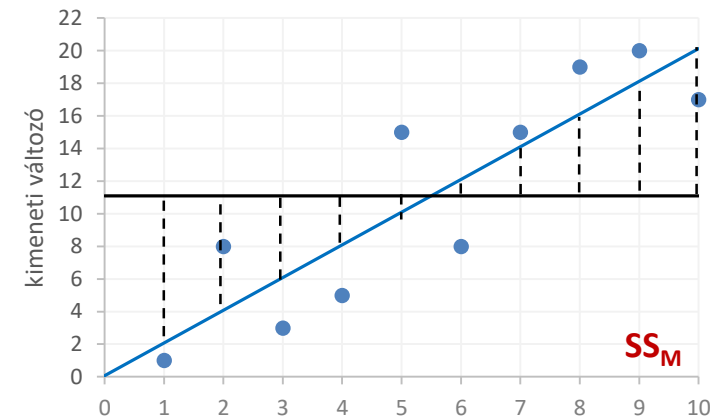
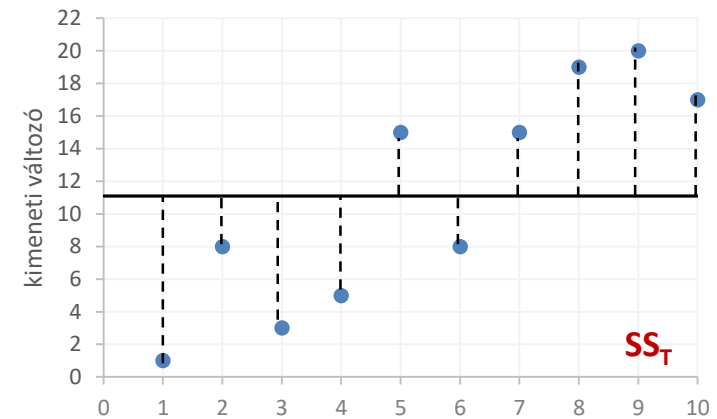


- **Típusok:**
 - Egyszerű regresszió: egyetlen független változóból próbáljuk megjósolni a függőt
 - Többszörös regresszió (multiple) – több független változóból jósoljuk meg a függőt.

R^2 és F érték

- **Mennyire jó a modellünk?**
- **Effect-size mutató (R^2)**– hatás nagysága, mennyit tudunk megmagyarázni a kimeneti változó variáciájából
 - SS_T = a mért pontok átlagtól való távolsága, a kimeneti változó teljes változatossága
 - SS_R = a pontok modellünktől való távolsága, a modell pontatlansága, a kimeneti változóból meg nem magyarázott változatosság
 - SS_M = a modellünk átlagtól való távolsága, a modell magyarázóereje, a kimeneti változóból megmagyarázott változatosság
 - Az effect-size a megmagyarázott és teljes variancia aránya $R^2 = \frac{SS_M}{SS_T}$
- **Statisztikai érték (F)** és a hozzá tartozó **szignifikancia** – hatás és hiba aránya, mennyiben következtethetünk a populációra
 - df_M = a modell szabadságfoka (a modell összetettségének mutatója)
 - df_R = a hiba szabadságfoka (N - predikált paraméterek száma)
 - $MS_M = SS_M / df_M$ = hatás
 - $MS_R = SS_R / df_R$ = hiba
 - A statisztikai erő a hatás és hiba aránya

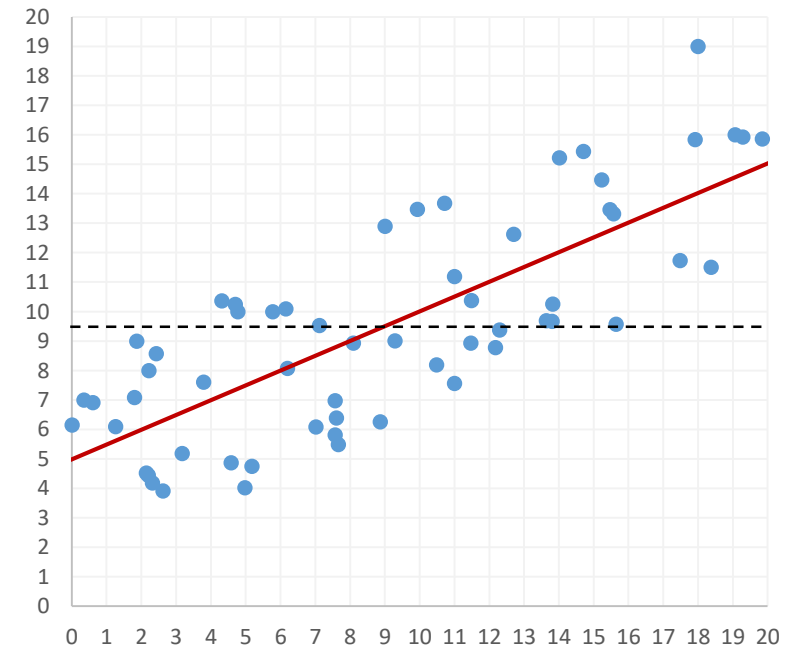
$$F = \frac{MS_M}{MS_R}$$



Regressziós együtthatók

- A **b0 regressziós együttható** megadja, hogy mennyi lenne a kimeneti változó értéke, ha (minden) prediktor értéke 0 lenne, ez az y tengely metszéspontja
- A **b1 regressziós együttható** megadja, hogy a prediktor változó egységnyi változása mekkora változást hoz létre a kimeneti változóban
 - A meredekségekhez tartozik egy SE érték is, mely megadja a paraméter bizonytalanságát
 - A regressziós egyenes vízszintes lesz (tehát $b_1 = 0$) akkor, ha a prediktor változó változásának nincs hatása a kimeneti változóra – ergo a legjobb predikció, amit tehetünk, a kimeneti változó átlaga
 - Feltételezhetjük, hogy egy prediktornak hatása van a kimeneti változóra, ha a meredekség szignifikánsan eltér a nullától

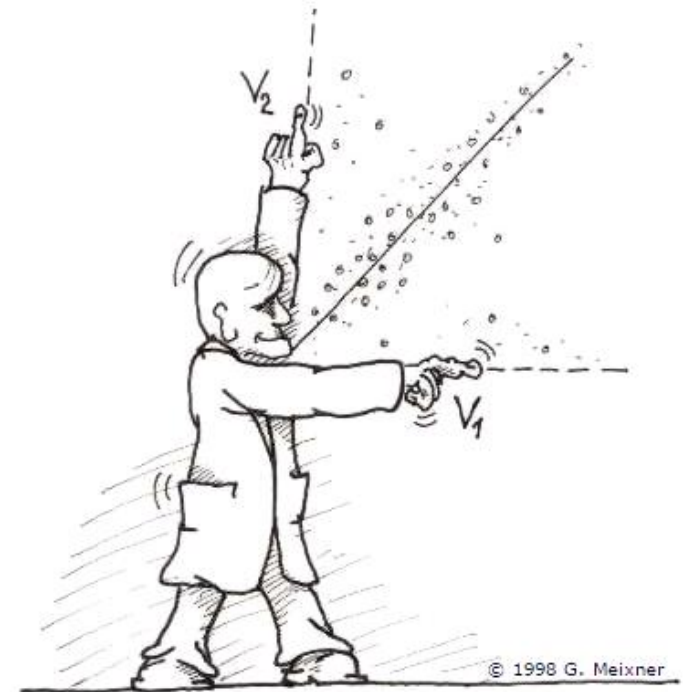
- Ezt t-teszttel ellenőrizzük:
$$t = \frac{b_{mért} - b_{nullhip}}{SE_b} = \frac{b_{mért}}{SE_b}$$



Regressziós mutatók összegzése

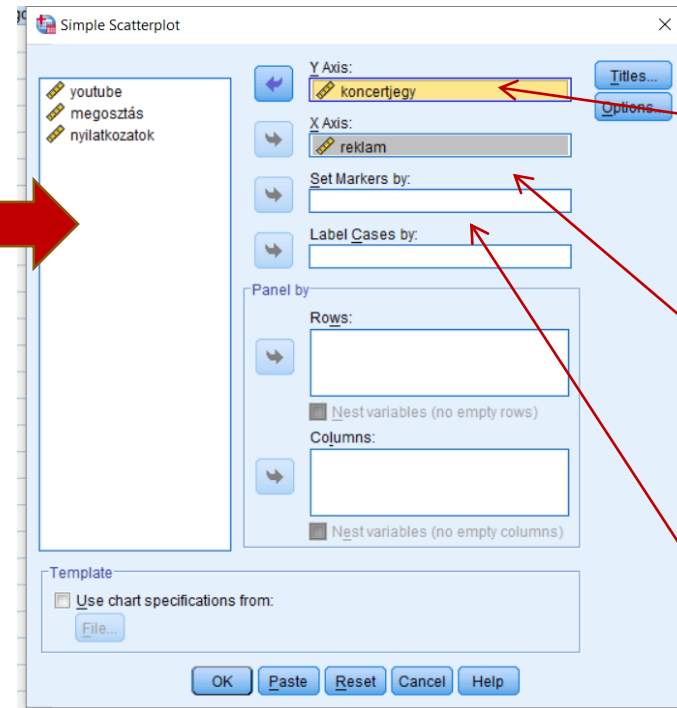
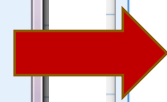
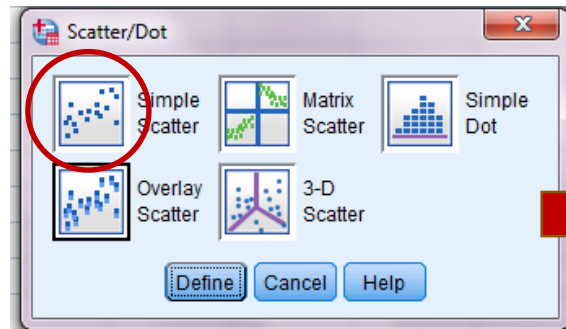
- **Regresszió elemzés** konyhanyelven
 - A prediktor változó(k)ból megpróbáljuk bejósolni a kimeneti változót
 - A prediktorok változó változása részben magyarázza a kimeneti változó változását
- **R^2** – effect size mutató
 - A kimeneti változóban változatosságának mekkora része magyarázható a prediktor változókkal
- **F** – statisztikai érték
 - milyen a hatás és hiba aránya
 - Az F-értékből számolható a szignifikancia értéke
- **b és hozzá tartozó t-próba** – a regressziós egyenes kezdőpontja és meredeksége
 - ha a független változó változása (növekedése/csökkenése) változást okoz a függő változóban (növekedést/csökkenést), akkor az egyenesnek van meredeksége.
 - T-próbával ellenőrizzük, hogy a meredekség szignifikánsan eltér-e a nullától

Egyszerű lineáris regresszió SPSS-ben



SPSS-ben Scatterplot

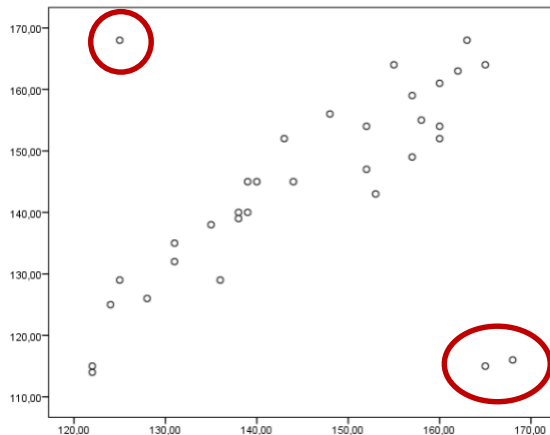
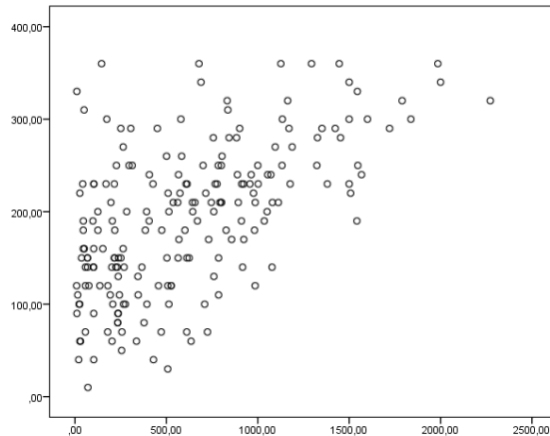
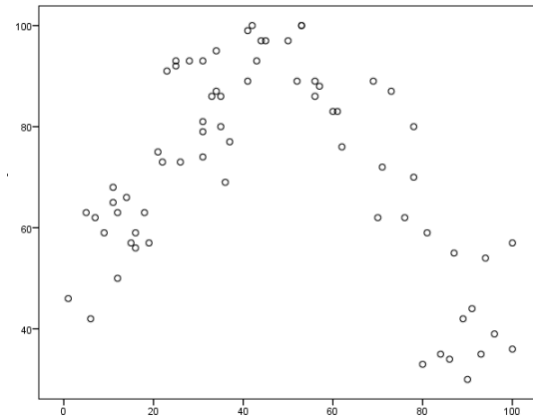
- Adatfájl: tobbval04_regresszio_koncertjegy.sav
- Először nézzünk rá az adatokra!
 - Elemezzük, mit látunk
 - Külön figyeljünk arra, teljesül-e a linearitás, a szóráshomogenitás, nincsenek-e (többdimenziós) outlierok
- Graph / Legacy Dialogs / Scatter\Dot / Simple Scatter



Megegyezés szerint
a feltételezett függő
változó

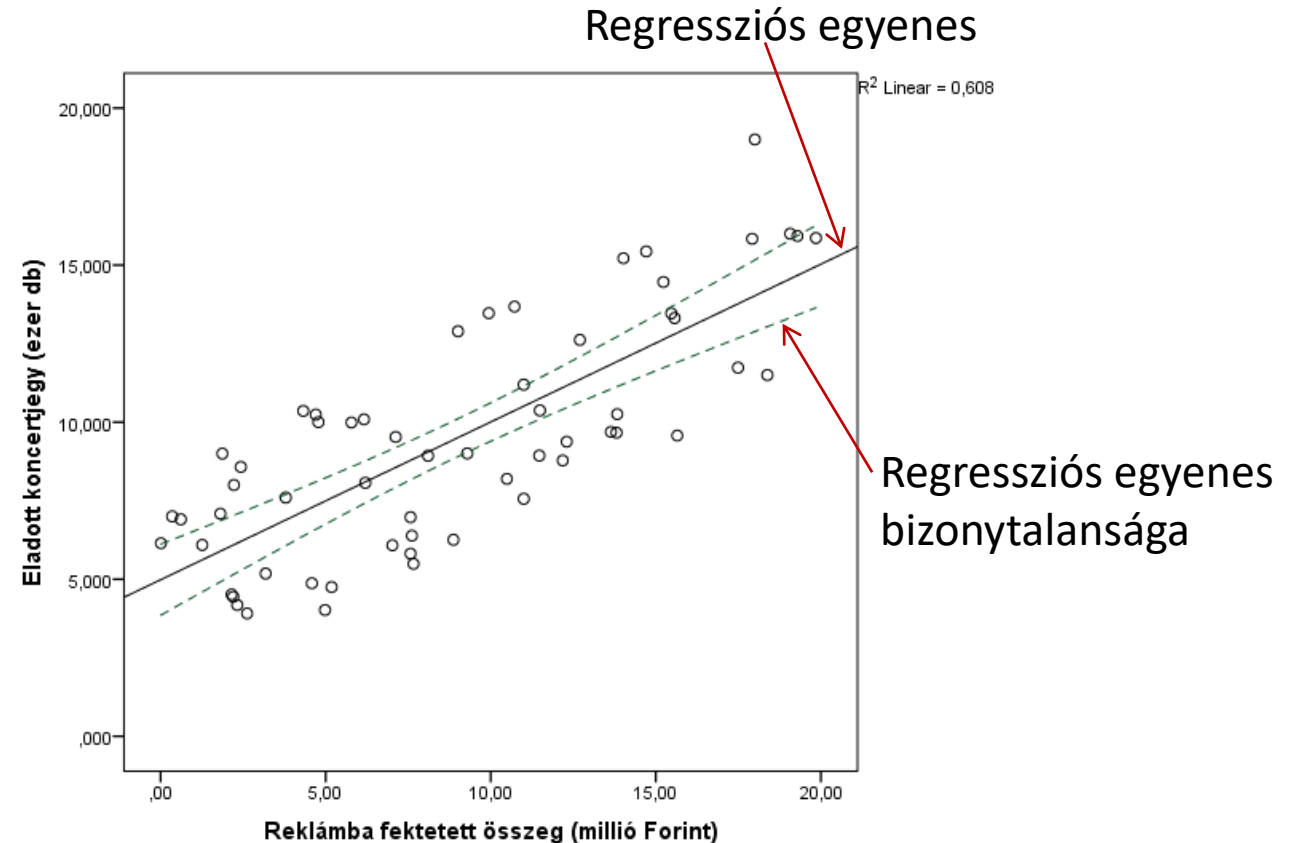
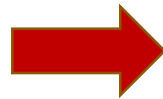
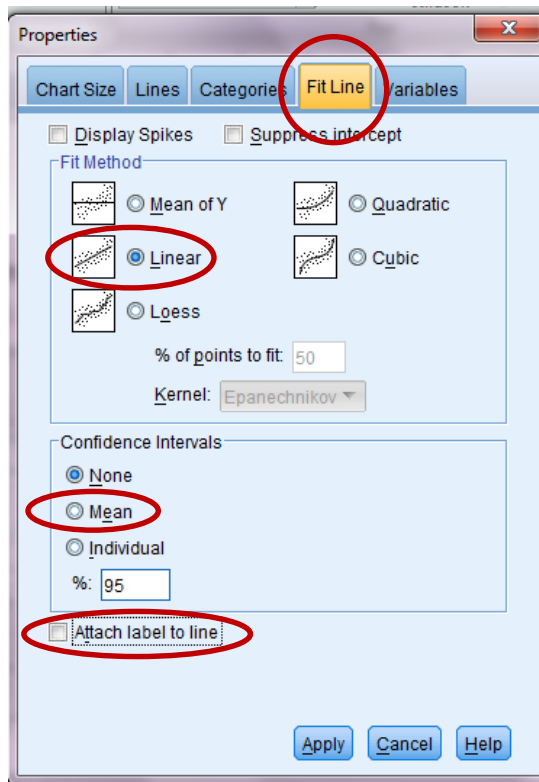
Megegyezés szerint
a feltételezett
független változó

Csoportosító tényező
(ha van) – a diagramon
külön színnel fog
megjelenni



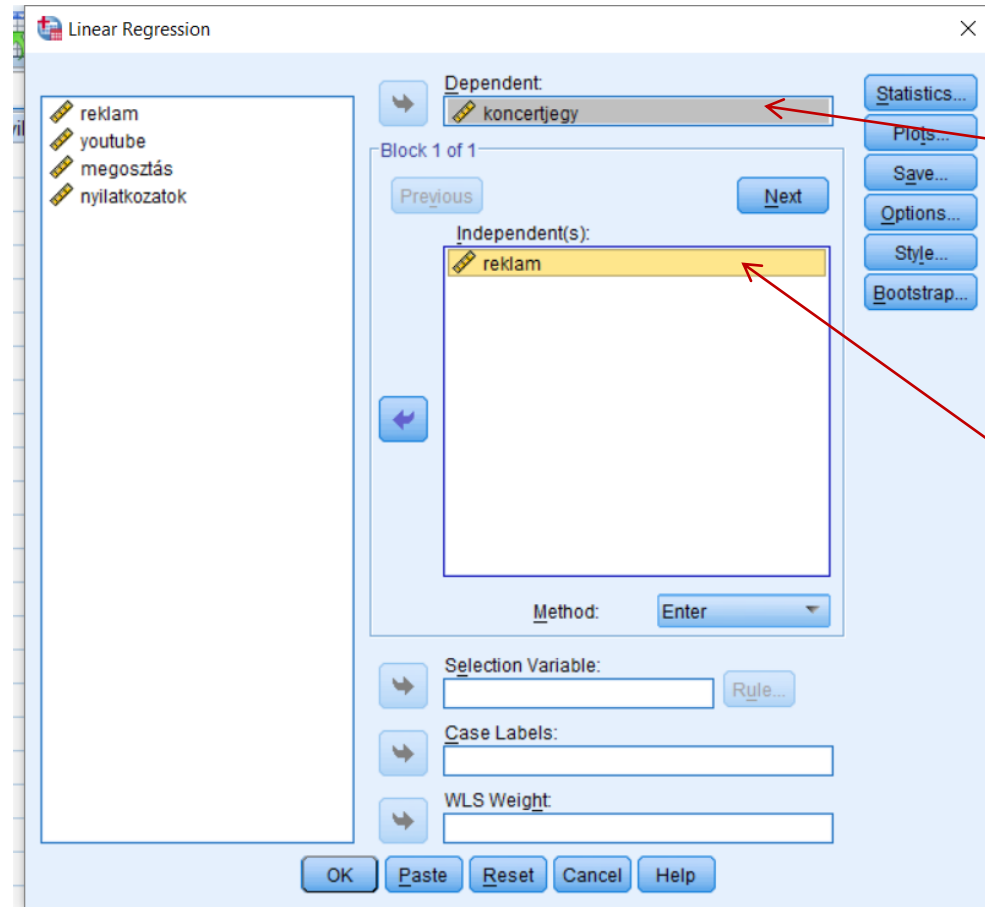
Scatterplot

- Felteheted a regressziós egyenest is: két klikk a diagrammra / Elements / Fit line at total (vagy Fit line at subgroups)
- Kipróbálhatsz nem lineáris modelleket is: lineáris = elsőfokú modell (egyenes), kvadratikuss = másodfokú modell (egy görbület), kubikus = harmadfokú modell (két görbület)
- Felteheted a grafikonra a modell bizonytalanságát jelző konfidencia intervallumot is.



Lineáris regresszió SPSS-ben

- Egyszerű regresszióban a feltételek megegyeznek a Pearson korreláció feltételeivel, gondold őket végig!
- Analyze / Regression / Linear
- A további beállítási lehetőségeket majd többszörös regresszióban átnézzük



Függő változó

Független változó
Egyelőre egyszerű lineáris
regressziót végzünk, ezért egy van

Lineáris regresszió SPSS-ben

- 1. Effect size – mennyit magyaráz meg a modell?

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,780 ^a	,608	,601	2,309072

a. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

R – predikált és kimeneti érték korrelációja. Egy változó esetén az megegyezik a prediktor és kimeneti változó korrelációs együtthatóval

R² – a modell a kimeneti változó varianciájának hány százalékát magyarázza. Itt 60,8%-ot tudunk magyarázni

Adjusted R² – mekkora magyarázóerőre számíthatunk a populációban

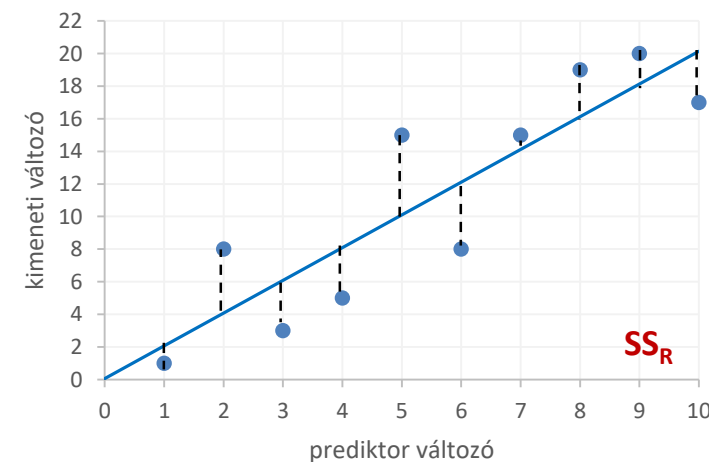
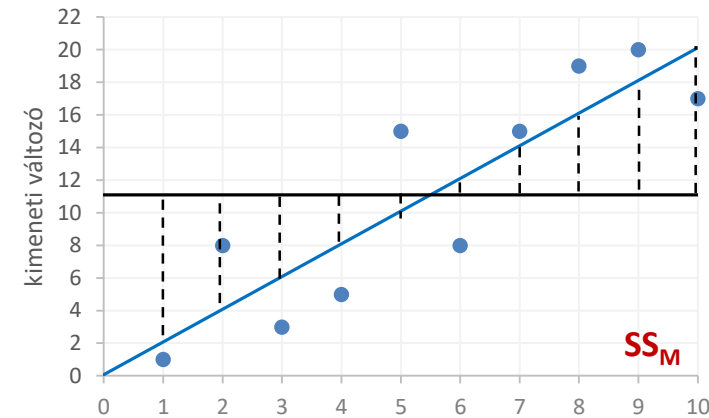
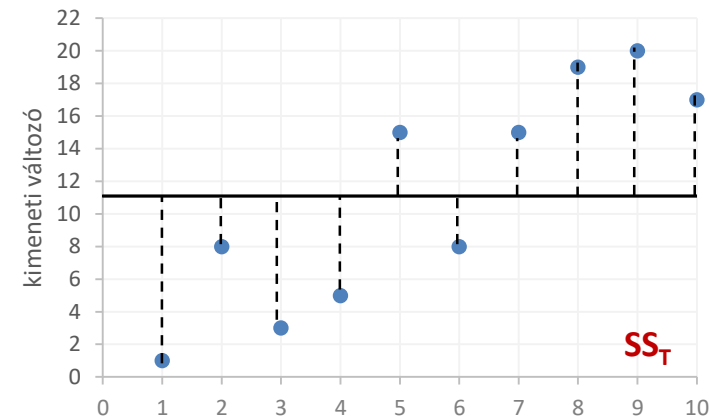
- 2. Statisztikai érték & szignifikancia érték – mekkora a hatás és hiba aránya és valószínűsége?

	SS _T	SS _R	SS _M	df _R	df _M	ANOVA ^a	MS _R	MS _M
Model	Sum of Squares	df	Mean Square	F	Sig.			
1	Regression	471,583	1	471,583	88,447	,000 ^b		
	Residual	303,913	57	5,332				
	Total	775,496	58					

a. Dependent Variable: Eladott koncertjegy (ezer db)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

F(1,57) = 88.447 p < .001**
A modellünk szignifikáns, a null hipotézist (nincs összefüggés a populációban) elvethetjük



Lineáris regresszió SPSS-ben

- Mit jósol a modellünk?

b0: konstans kezdőérték: ennyi lenne az Y a modell alapján, ha X nulla lenne

b1: ennyivel nő X egységeiként Y értéke

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,983	,565		8,817	,000
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000

a. Dependent Variable: Eladott koncertjegy (ezer db)

t-próba arra, hogy b_0 és b_1 szignifikánsan eltér-e a nullától

- Ez a modell alapján a következő jóslatot tehetjük:

- Ha nem fektetünk be a reklámokba, akkor:

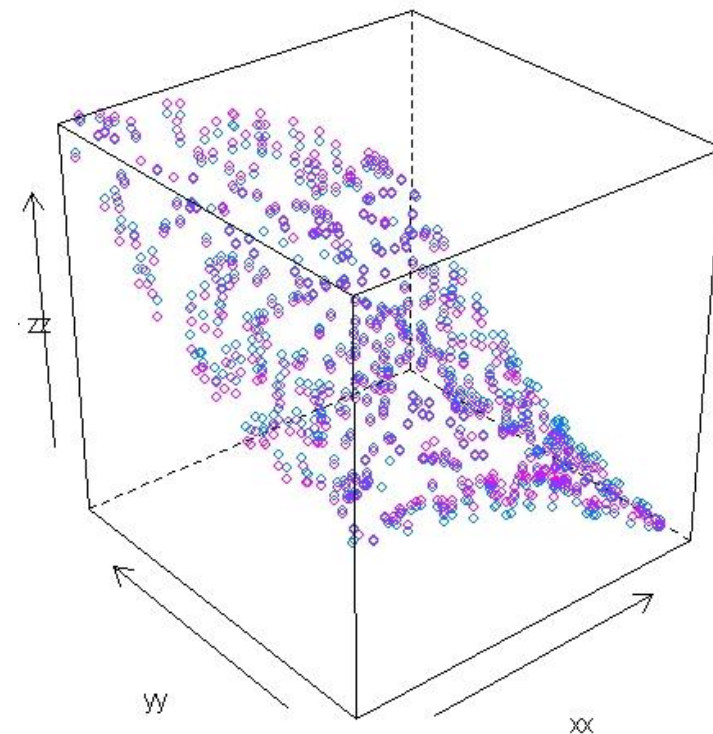
- $Y_{pred,i} = b_0 + b_1 * X_i$ -> Predikált jegyeladás = $4.983 + 0.502 * 0 = 4.983$ <ezer db>

- Ha befektetünk a reklámba 6millió Ft-ot, akkor:

- $Y_{pred,i} = b_0 + b_1 * X_i$ -> Predikált eladás = $4.983 + 0.502 * 6 = 7.995$ <ezer db>

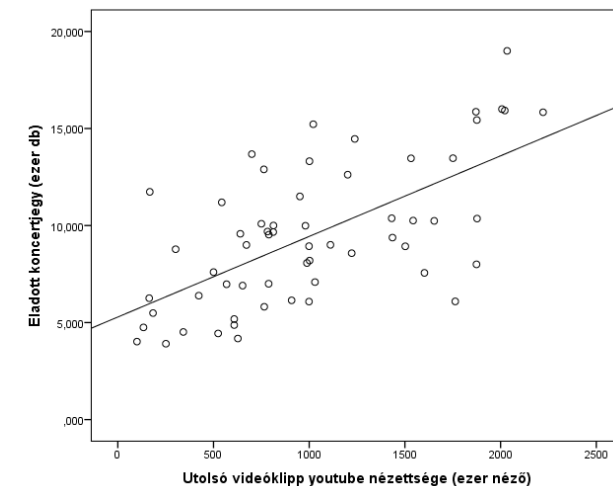
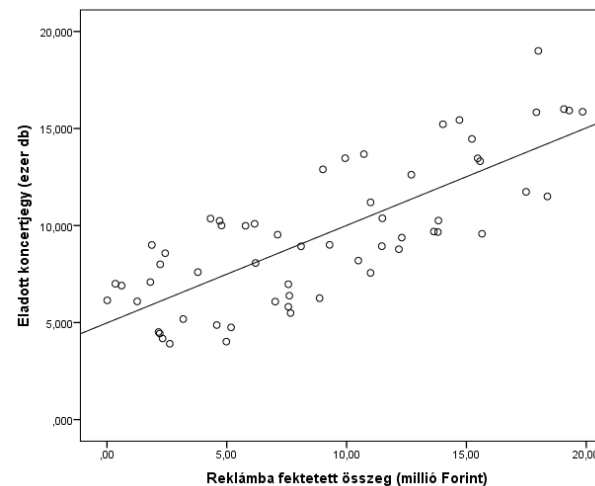
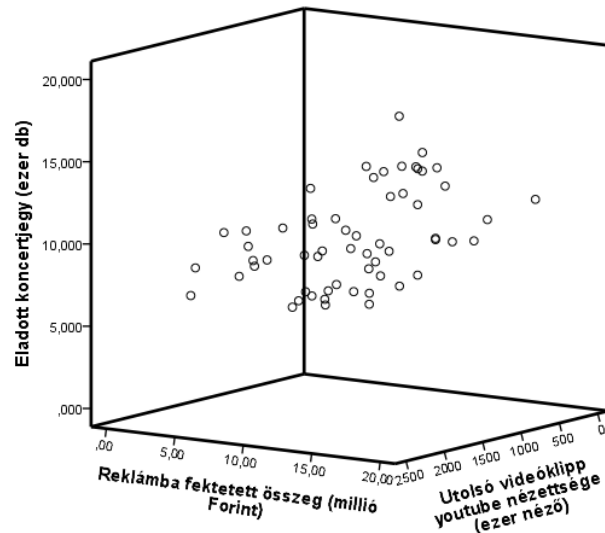
- DE: vigyázat: a modell csak 60,8%-ot magyaráz!

Többszörös lineáris regresszió elmélet



Többszörös regresszió

- K darab független prediktor változó esetén:
 - A pontfelhő **k+1 dimenziós térben** képzelhető el, ahol k a prediktorok száma, a plusz 1 pedig a kimeneti változó tengelye. A regressziós egyenes ebben a többdimenziós térben halad.
- Együtthatók:
 - A regressziós egyenes továbbra is egy pontban metszi az Y tengelyt (ott, ahol minden prediktor értéke 0), tehát **egyetlen b_0 konstans** tartalmaz az egyenlet.
 - A regressziós egyenes meredekségét viszont dimenzióként vizsgálhatjuk, azaz **minden prediktorhoz külön b_1 érték** fog tartozni.



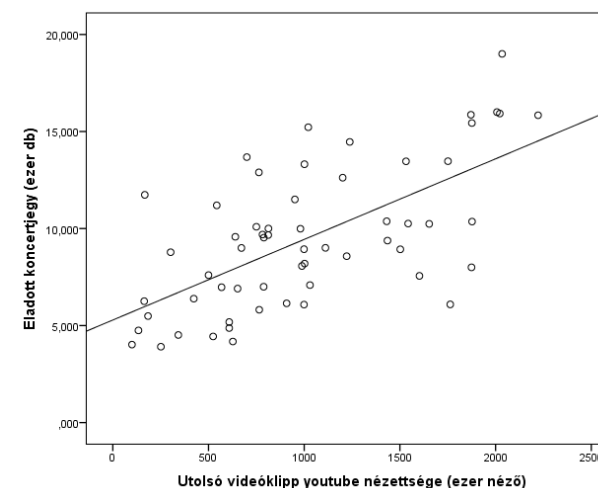
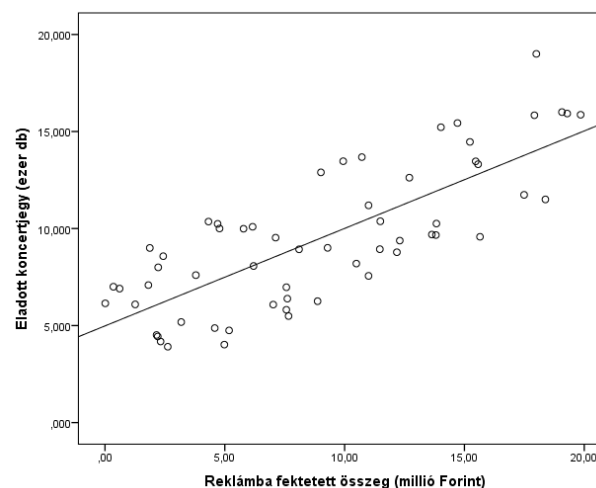
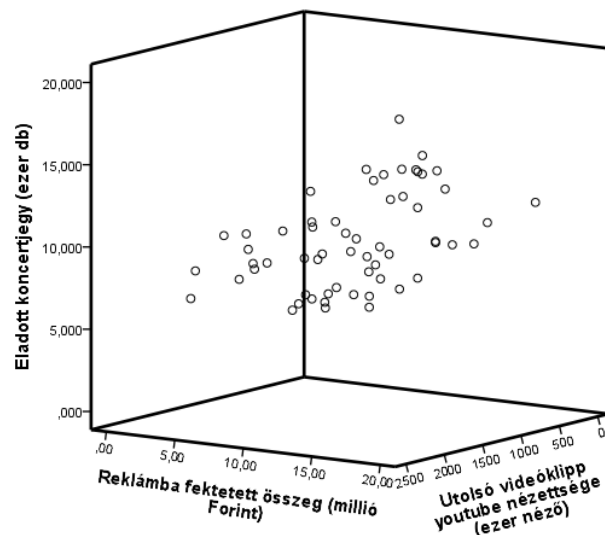
Többszörös regresszió

- A regressziós egyenes képlete:

- $Y_{\text{pred},i} = b_0 + b_1 * X_{1,i} + b_2 * X_{2,i} + \dots + b_k * X_{k,i}$, ahol i az egyedek, k pedig a prediktorok indexe
- Több prediktor változó lineáris kombinációja hozza létre a kimeneti változó predikált értékét
- Példa: Y_i zenekar várható jegyeladása = $b_0 + b_{\text{reklám}} * X_i$ zenekar reklámjába fektetett összeg + $b_{\text{youtube}} * X_i$ koncert nézettsége a youtube-on

- A kimeneti változó értéke:

- $Y_i = Y_{\text{pred},i} + \varepsilon_i = (b_0 + b_1 * X_{1,i} + b_2 * X_{2,i} + \dots + b_k * X_{k,i}) + \varepsilon_i$
- Y_i zenekar koncertjére eladott jegyek = Y_i zenekar várható jegyeladása + tévedés_i zenekar jegyeladásának becslésében



Többszörös regresszió

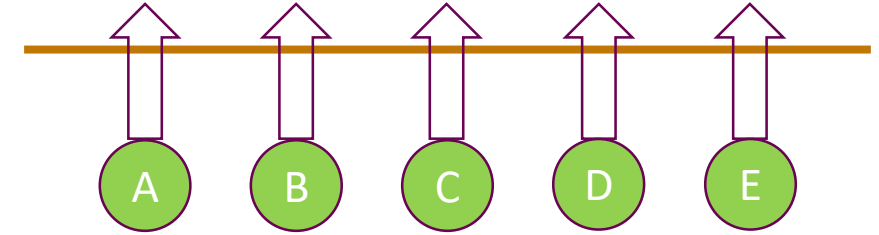
- A mutatók bonyolultabbak a többdimenziós tér miatt, de jelentésük ugyanaz
 - **R**
 - Korreláció a modell által predikált értékek és a valós értékek között
 - Ha tökéletesen be tudnánk jósolni a kimeneti változót, akkor nem lenne hiba a jóslatban, a predikált és mért értékek korrelációja 1 lenne
 - **R²**
 - Effect-size mutató, a modellünk mennyit magyaráz a függő változó varianciájából
 - **F és szignifikancia**
 - Statisztikai érték, a hatás és hiba aránya
 - A statisztikai értékhez rendeljük a szignifikancia értéket
 - **b₀**
 - Konstans – ha minden prediktor értéke nulla lenne, ennyi lenne a kimeneti változó értéke (itt metszi a regressziós egyenes az y tengelyt)
 - **b₁-ek és t-próbák**
 - Prediktor változónként meredekségek
 - Prediktor változók hatása a kimeneti változóra

Modell felépítésének módja(i)

- **Módok:**

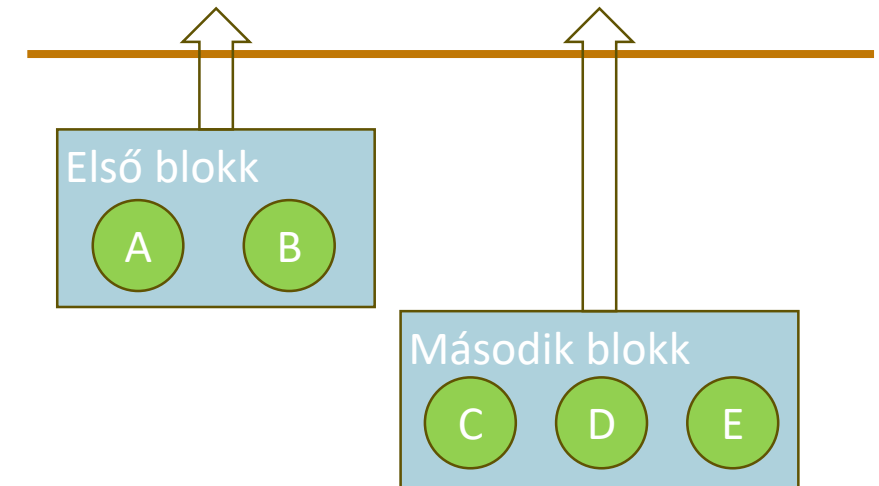
- **Enter (forced entry)**

- A modellbe egyszerre lépnek be a változók



- **Hierarchikus (blokkonként – blockwise model)**

- A változók sorrendjét a kutató adja meg (előzetes tudása, sejtése alapján)
 - Először azok a változók kerülnek a modellbe,
 - melyek a hipotézisek szempontjából fontosabban
 - melynek már ismerjük a hatását
 - melynek várhatóan legnagyobb a hatása
 - Egy blokkba
 - több változó is tartozhat,
 - és ezeknek külön megadhatjuk a belépési módját



Modell felépítésének módja(i)

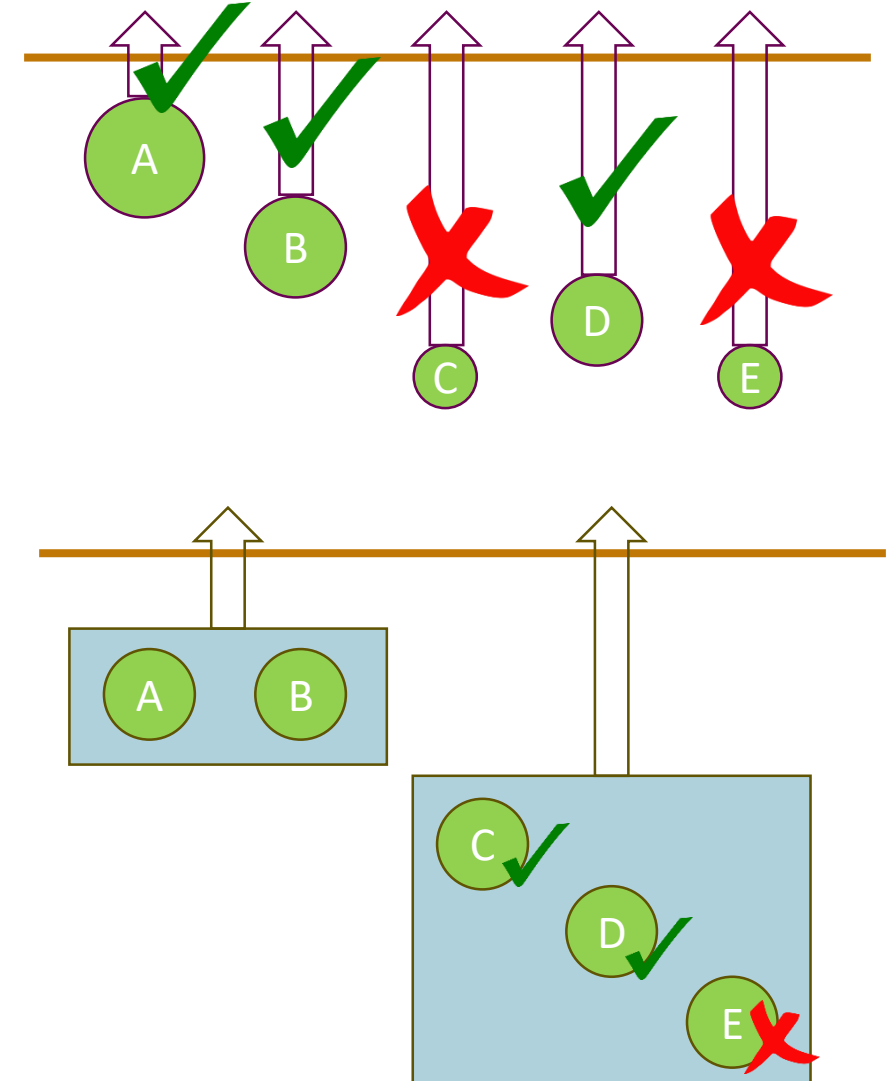
- **Módok:**

- **Stepwise módszerek**

- A modellbe való kerülést matematikai kritériumok szabják meg
- Ha egy prediktor nem tud szignifikáns mértékben hozzáadni a modellhez, nem kerül bele
- Több fajta is létezik:
 - Forward, Backward, Stepwise

- **Összetett modellek**

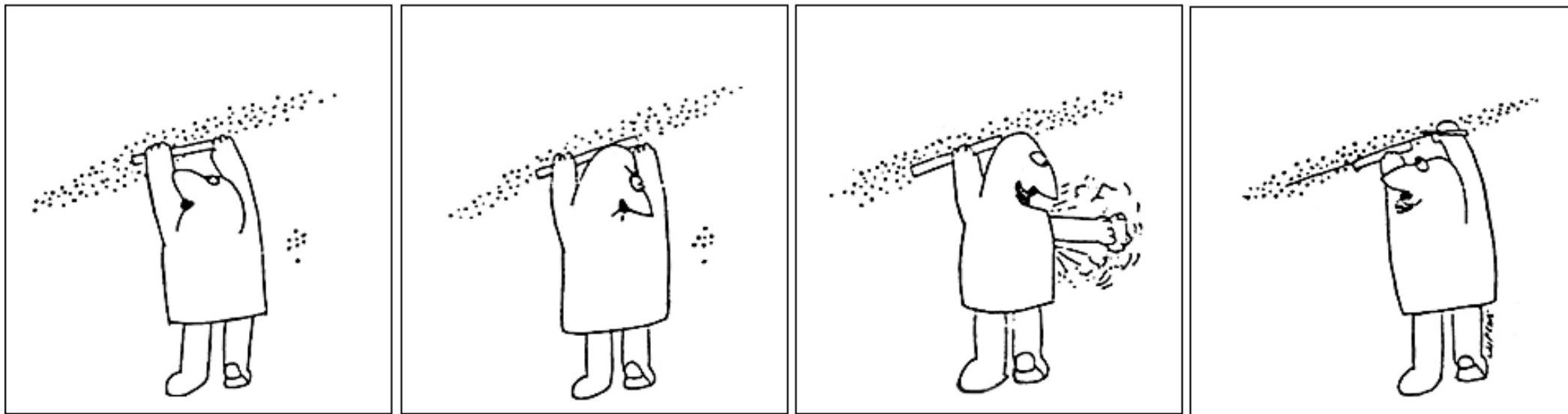
- A beléptetés módjait lehet kombinálni
- Pl. egyik blokkban Enter módszerrel lépnek be a változók, másik blokkban stepwise-zal



Blockwise, enter vagy stepwise?

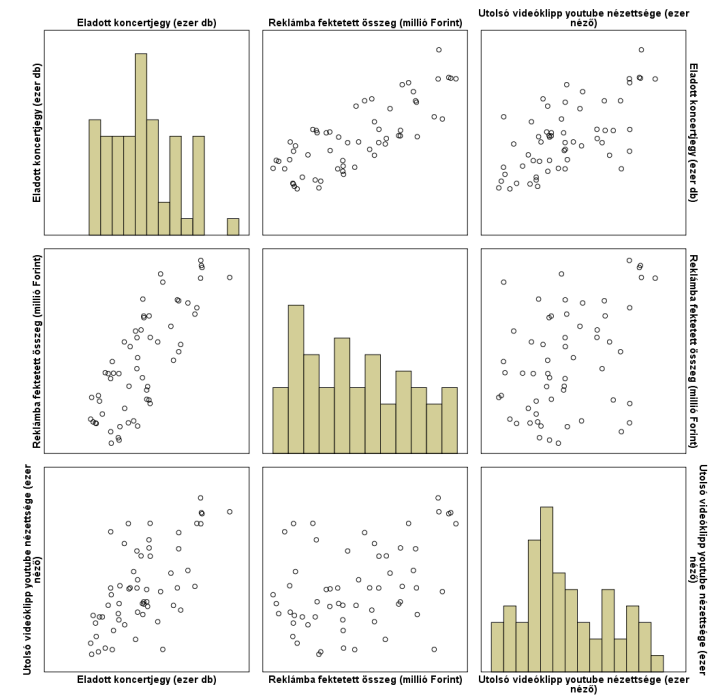
- **Ha van tudományos elképzelésed, használj blockwise (hierarchikus) módszert**
 - A blockwise módszer lehetővé teszi, hogy az előzetes tudásod is számításba vedd, a számodra fontosabb változókat kiemeld
- **Ha feltáró vizsgálatot végzel, használj enter módszert**
 - A változó modellbe való kerülését nem befolyásolják az előzetes elképzeléseid, tudások
 - Pontosabb képet kapsz a prediktor változók és a kimeneti változó kapcsolatáról, plusz a prediktor változók egymással való kapcsolatáról is
- **Ha „csak” a leggazdaságosabb modellt keresed, használj stepwise módszert**
 - Megkeresi azt a modellt, ahol a lehető legkevesebb változóval a lehető legpontosabb jóslatot lehet tenni
 - Például, ha célod, hogy lehetőleg kevés prediktor változóval kapj pontos becslést (mert a prediktor változók felvétele pénzbe kerül)
 - Hátránya, hogy a változókról önmagunkban kevés információt ad
 - Fennáll az overfitting veszélye
 - **Sokat támadják**, mert tisztán matematikai és véletlenül alapuló döntések határozzák meg a módszert – minden fontos döntést „kivesz” a kutató kezéből. Van folyóirat, mely nem fogad el cikket stepwise módszerrel :S

Feltételek



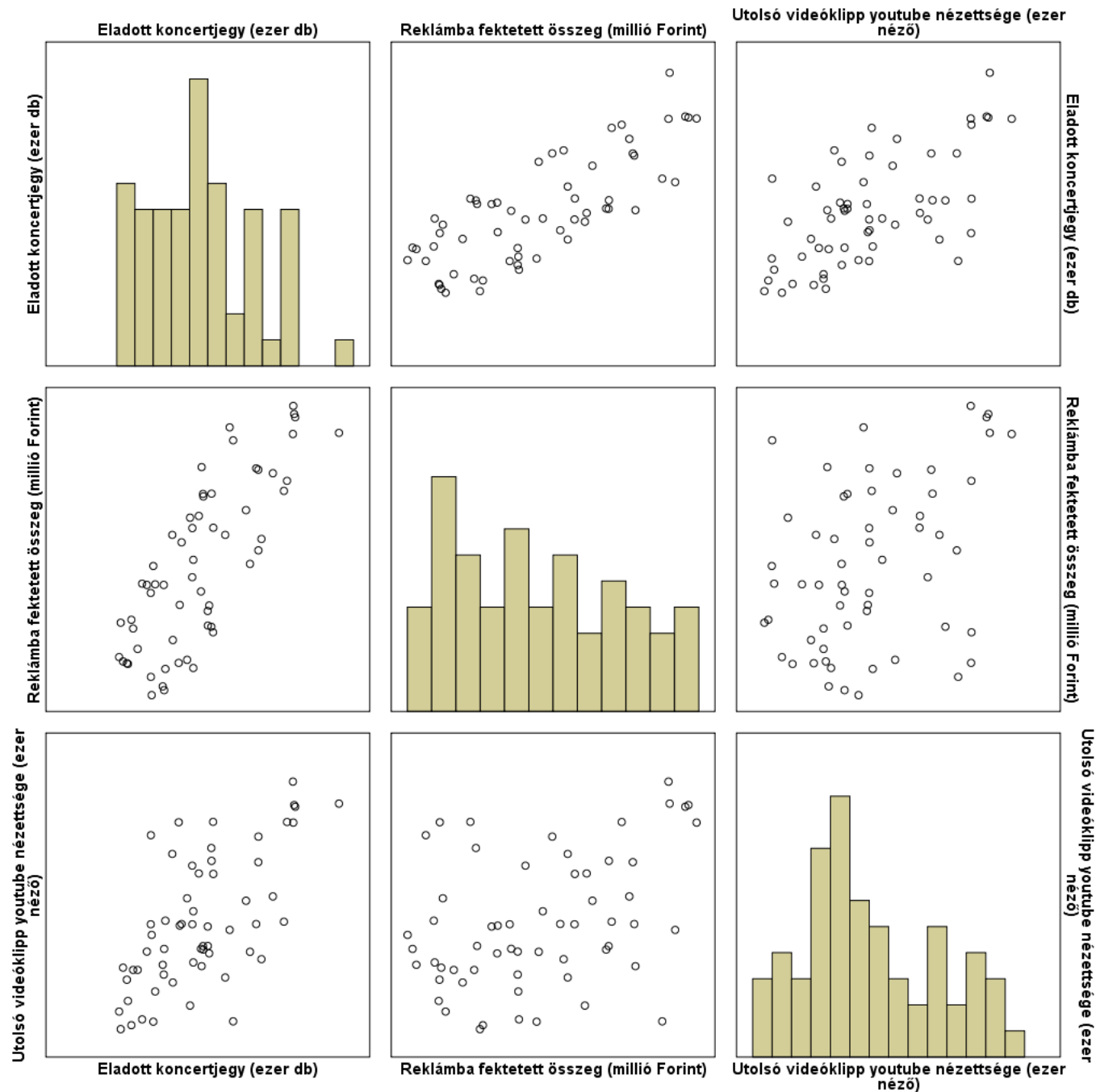
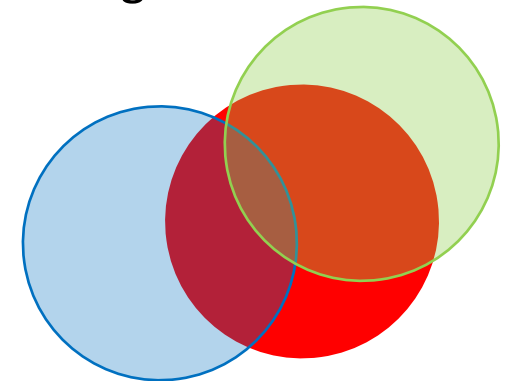
- A regresszió feltételeit nézd át az előadás anyaga alapján!
 - ① **Változók típusa**
 - ② **Nincs nulla varianciájú változó** – variancia nélküli változónak nincs információtartalma
 - ③ **Nincs (multi)kollinearitás** – (multi)kollinearitásban résztvevő változók hatása nem értelmezhető egyértelműen
 - ④ **Nincs külső változó** – egy külső változó illuzórikus összefüggéseket hozhat létre
 - ⑤ **Homoszkedaszticitás / Varianciák homogenitása / Szóráshomogenitás** – a heteroszkedaszticitás a konfidencia intervallumokat, így a szignifikancia értékeket torzítja
 - ⑥ **Linearitás** – ha nincs linearitás, pontatlan lesz a modell, és a magyarázóerőket alulbecsüljük
 - ⑦ **Független hibák** – a belső korrelációk a konfidencia intervallumokat, így a szignifikancia értékeket torzítják
 - ⑧ **Hiba normális eloszlása** – a normalitás sérülése általában más feltételek sérüléseként jön létre
 - ⑨ **Elemzés** – kis elemszám növeli a másodfajú hiba valószínűségét
 - ⑩ **Nincsenek (több)dimenziós outlierok** – a modell együttthatóit torzítják

Többváltozós regresszió SPSS-ben



Többváltozós regresszió SPSS-ben

- Kezdjük egy olyan elemzéssel, ahol először a reklámot, majd a youtube-ot léptetjük a modellbe!
- Grafikus megjelenítés:
- A kimeneti változóval a reklám és a youtube is pozitív kapcsolatban van
- A prediktorok között nem látható jelentős korreláció – úgy tűnik viszonylag függetlenek – így a kimeneti változó variáciájának is más részét magyarázzák majd, nem igazán „versengenek” egymással – ergo viszonylag tiszta képet látunk majd róluk



Többváltozós regresszió SPSS-ben

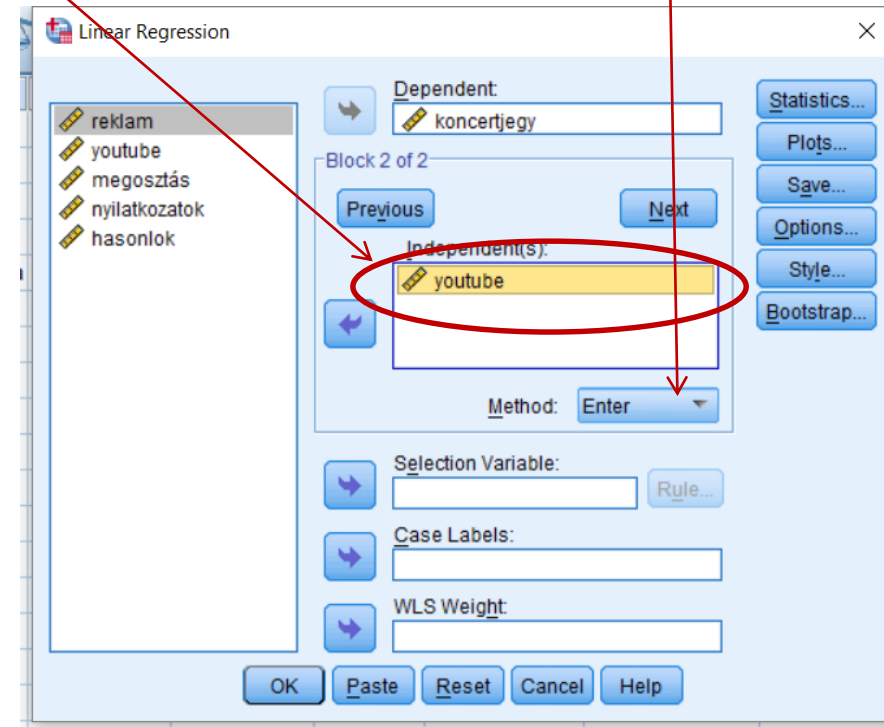
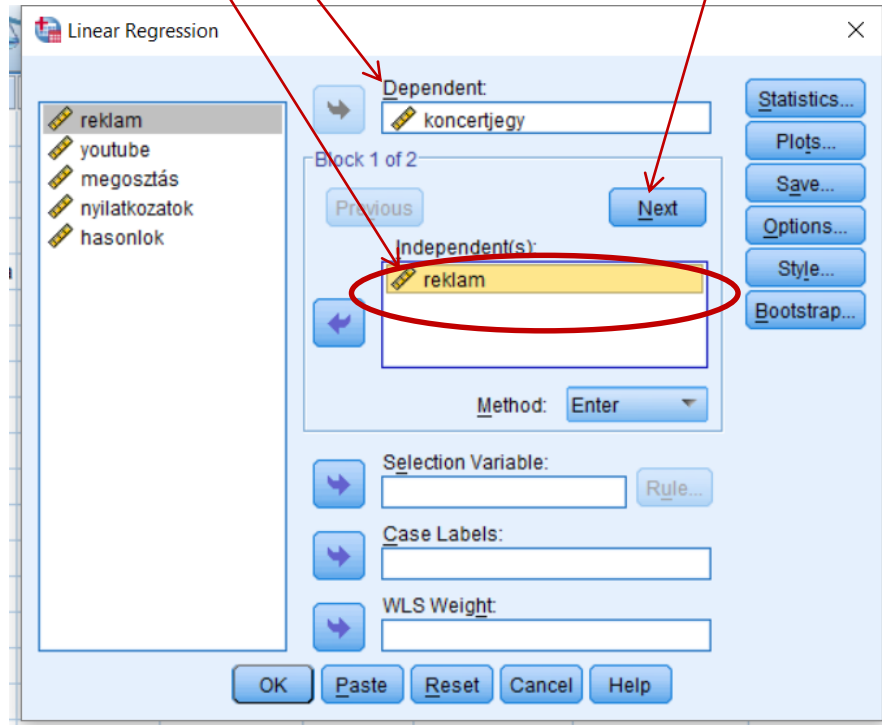
- Modell felépítése: első blokkba kerüljön a reklám, majd második blokkba a youtube!

Függő változó

Nexttel lépek át a következő blokkba

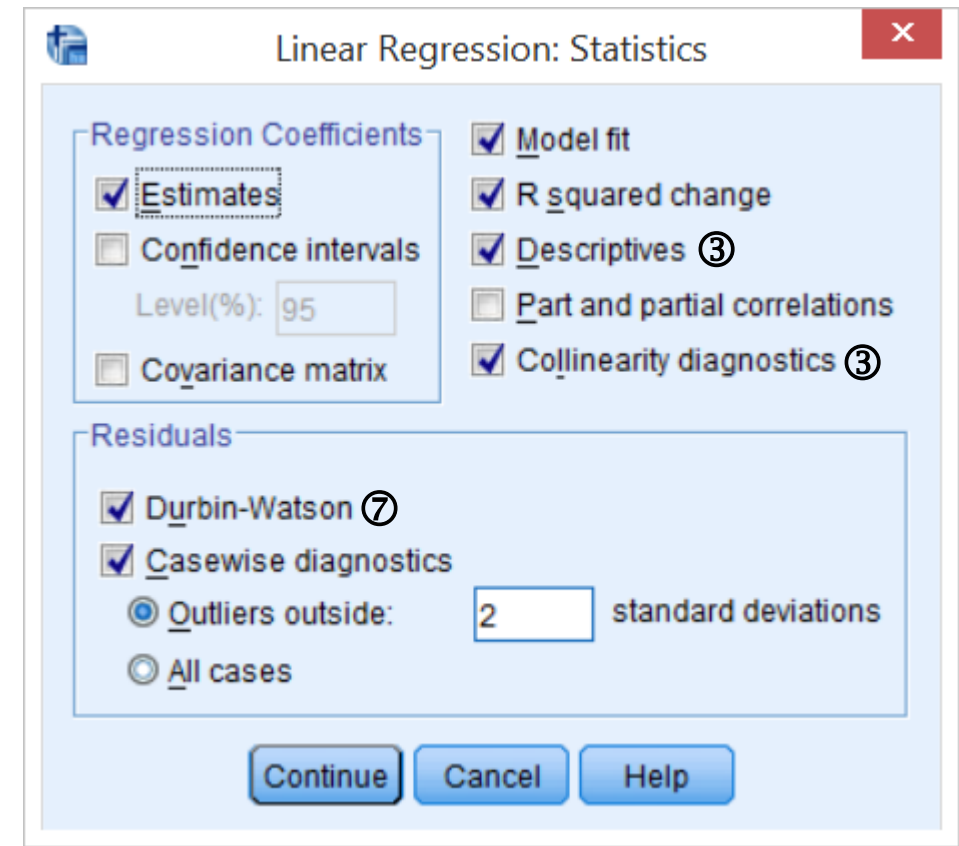
Első blokkomba kerüljön a reklám, mert az a hipotézisek szempontjából fontosabb

A youtube kevésbé fontos, ezért a második blokkba teszem be.
Bár most nem számít, blokkon belül Enter módszert választok



Többváltozós regresszió SPSS-ben

- **Statistics**-en belül:
 - **Estimate**: ez fogja megadni a regressziós együtthatókat (b_0 és b_1 értékeket) + a hozzá tartozó t-próbákat.
 - **Model fit**: a modell statisztikai értékét (F-érték) és magyarázóerejét (R^2) adja meg
 - **R^2 change**: megadja, hogy az újabb változó(k) modellbe építésekor mennyivel nő az R^2
 - **③ Descriptive**: deskriptív statisztikák + korrelációs mátrix (prediktor változók korrelációjának erőssége, ezáltal a kollinearitás ellenőrizhető vele)
 - **③ Collinearity diagnostics**: multikollinearitást ellenőrzi, többek között a tolerancia értékeket adja ki
 - **⑦ Durbin-Watson teszt**: a hibák függetlenségét ellenőrzi
 - **Casewise diagnostics**: gyanús elemeket sorolja fel, azokat, akiket nem tudtunk jól bejósolni (ahol a reziduális hiba túl nagy). Állítsuk 2 SD-re, így a minta kb. 5%-át sorolja fel!



Többváltozós regresszió SPSS-ben

- **Plots-on belül:**
 - **Plot:** homoszkedaszticitást és linearitást ellenőrizhetem
 - két plotot kérünk ki, ⑤, ⑥ a predikált érték és reziduálisok kapcsolatát és ⑥ a predikált érték és a függő változó kapcsolatát (ez utóbbit ritkábban szoktuk használni)
 - **⑧ Hisztogram:** a reziduális hiba eloszlását ellenőrizhetem

Reziduális hiba standardizálva

⑤

⑥

Linear Regression: Plots

Scatter 1 of 1

DEPENDNT

- *ZPRED
- *ZRESID
- *DRESID
- *ADJPRED
- *SRESID
- *SDRESID

Y:

X:

Standardized Residual Plots

☒ Histogram ⑧

☐ Normal probability plot

Produce all partial plots

Continue Cancel Help

Next

⑥

Linear Regression: Plots

Scatter 2 of 2

DEPENDNT

- *ZPRED
- *ZRESID
- *DRESID
- *ADJPRED
- *SRESID
- *SDRESID

Y:

X:

Standardized Residual Plots

☒ Histogram

☐ Normal probability plot

Produce all partial plots

Continue Cancel Help

Függő változó

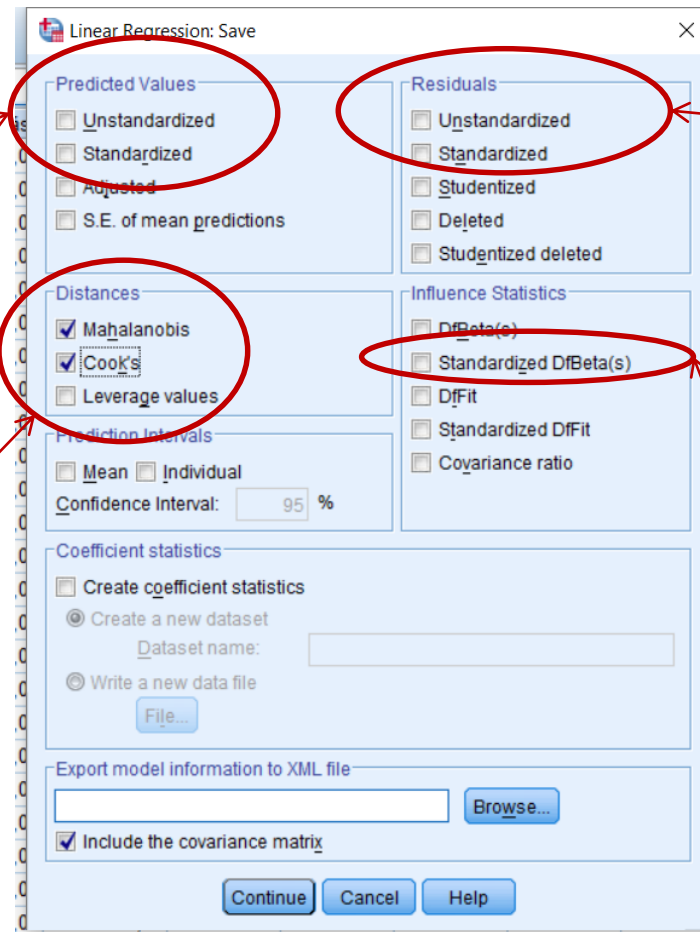
Függő változó predikált értéke standardizálva

Többváltozós regresszió SPSS-ben

- **Save**-en belül:
 - A többdimenziós outlierok kereséséhez tartozó új változókat itt hozhatom létre
 - Ha szükség van rá, akkor a predikált változót is menthetem standardizálatlan (skálafüggő) és standardizált formában is (jól jön majd a riportolás során)

Predikált érték
Lehet standardizált és
standardizálatlan is

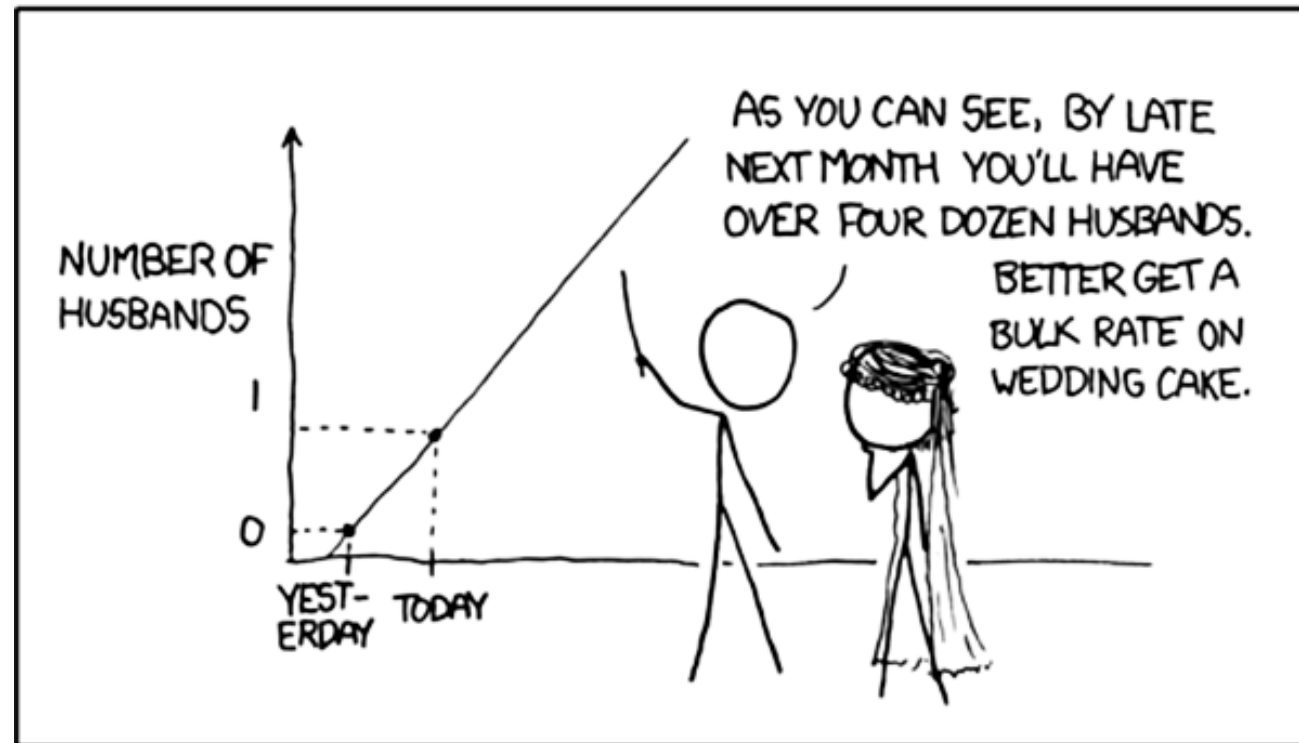
⑩ Többdimenziós outlierok keresésére:
Mahalanobis távolság
Cook távolság
Leverage/befolyás érték



⑧ Ha a hibák normál eloszlását
tesztesen szeretném ellenőrizni, itt
menthetem a hibákat standardizált
vagy standardizálatlan formában

⑩ Standardizált DfBeta

Többszörös lineáris regresszió értelmezés



Többszörös lineáris regresszió értelmezés

- **Descriptive**

- Deskriptív statisztikák a változókról

Descriptive Statistics

	Mean	Std. Deviation	N
Eladott koncertjegy (ezer db)	9,48331	3,656585	59
Reklámba fektetett összeg (millió Forint)	8,9583	5,67590	59
Utolsó videóklipp youtube nézettsége (ezer néző)	1010,46	565,769	59

- **③Correlations**

- Kollinearitás hiányát ellenőrizzük: ha a prediktor változók között van olyan, mely .8-nál erősebben korrelál, akkor nem teljesül a kritérium.
- Ekkor vagy ki kell venni a két változó közül az egyiket, vagy főkomponens elemzéssel össze kell őket egyetlen változóba gyúrni

Correlations

		Eladott koncertjegy (ezer db)	Reklámba fektetett összeg (millió Forint)	Utolsó videóklipp youtube nézettsége (ezer néző)
Pearson Correlation	Eladott koncertjegy (ezer db)	1,000	,780	,643
	Reklámba fektetett összeg (millió Forint)	,780	1,000	,393
	Utolsó videóklipp youtube nézettsége (ezer néző)	,643	,393	1,000
Sig. (1-tailed)	Eladott koncertjegy (ezer db)	.	,000	,000
	Reklámba fektetett összeg (millió Forint)	,000	.	,001
	Utolsó videóklipp youtube nézettsége (ezer néző)	,000	,001	.
N	Eladott koncertjegy (ezer db)	59	59	59
	Reklámba fektetett összeg (millió Forint)	59	59	59
	Utolsó videóklipp youtube nézettsége (ezer néző)	59	59	59

Többszörös lineáris regresszió értelmezés

- **Variables entered/removed**

- Modell felépülésének útja
 - Először a reklám került be Enterrel (ez volt az első blokk), majd a youtube szintén Enterrel (ez volt a második blokk)
- Információ tartalma stepwise módszereknél van, ahol a változók modellbe való be és esetleges kilépésének folyamatát foglalja össze

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Reklámba fektetett összeg (millió Forint) ^b	.	Enter
2	Utolsó videóklipp youtube nézettsége (ezer néző) ^b	.	Enter

a. Dependent Variable: Eladott koncertjegy (ezer db)

b. All requested variables entered.

Többszörös lineáris regresszió értelmezés

- **Model Summary**

- **R Square:** mennyit tudtunk a modellel megmagyarázni az adatokból. Az első modell csak 60,8%-ot, a második 74,2%-ot
- **Adjusted R Square:** Mennyire számíthatnánk az egész populációt mérve. Ha közel van a sima R^2 -hez, akkor jó
- **Change Statistics:** Az új változók beemelése mennyiben tette jobbá a modellt (R Square Change), és egy F-próba arra, hogy a javulás szignifikáns-e
- **⑦Durbin-Watson:** Hibák korrelálatlanságát ellenőrzi, ha 2 körül van jó, ha 1 alatt vagy 3 felett, akkor baj van
- A táblázat alatti megjegyzések segítenek az értelmezésben

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin- Watson ⑦
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,780 ^a	,608	,601	2,309072	,608	88,447	1	57	,000	1,891
2	,861 ^b	,742	,732	1,891674	,133	28,929	1	56	,000	

a. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Utolsó videóklipp youtube nézettsége (ezer néző)

c. Dependent Variable: Eladott koncertjegy (ezer db)

Többszörös lineáris regresszió értelmezés

- **ANOVA** tábla
 - **F-érték és Sig.:** hatás és hiba aránya – következtethetőség a populációra
 - $F([df_{reg}], [df_{res}]) = [F\text{-érték}] p = [p\text{-érték}]$
 - Az első modell: $F(1, 57) = 88.447 p < .001$ és a második modell: $F(2, 56) = 80.357 p < .001$
 - **Sum of Squares és Mean Square** oszlopok tartalmazzák az SS_M , SS_R , SS_T és MS_M , MS_R értékeket – vedd össze az elmélettel

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	471,583	1	471,583	88,447	,000 ^b
	Residual	303,913	57	5,332		
	Total	775,496	58			
2	Regression	575,104	2	287,552	80,357	,000 ^c
	Residual	200,392	56	3,578		
	Total	775,496	58			

a. Dependent Variable: Eladott koncertjegy (ezer db)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

c. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Utolsó videóklipp youtube nézettsége (ezer néző)

Gondold végig, vajon az első modellben
miért 1 a modellhez tartozó szabadságfok?
miért 58 a teljes mintához tartozó szabadságfok?
miért 57 a hibához tartozó szabadságfok?
Segítség: $N=59$

Többszörös lineáris regresszió értelmezés

- **Coefficients (Regressziós együtthatók)**

- **B:** b_0 -át (konstans kezdőérték) és b_1 -et (meredekség) adja meg minden változóra úgy, hogy mi lenne a hatásuk, ha a többi kontrollálnánk
 - b_0 - ha a független változók értéke 0, mennyi lenne a függő változó értéke
 - b_1 – a független változó egységnyi változása mekkora változást idéz elő a függőben
- **Beta:** a b_1 standardizált értékei – így már összehasonlíthatóak egymással
- **t és sig:** b_0 és b_1 értéke szignifikánsan eltér-e a 0-tól (a 0 azt jelentené, hogy az adott változónak nincs hatása a függő változóra)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4,983	,565		8,817	,000		
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000	1,000	1,000
2	(Constant)	3,290	,560		5,876	,000		
	Reklámba fektetett összeg (millió Forint)	,402	,048	,623	8,439	,000	,845	1,183
	Utolsó videóklipp youtube nézettsége (ezer néző)	,003	,000	,397	5,379	,000	,845	1,183

a. Dependent Variable: Eladott koncertjegy (ezer db)

Többszörös lineáris regresszió értelmezés

- **Coefficients (Regressziós együtthatók)**

- **③ VIF és Tolerancia**

- multikollinearitást ellenőrizzük vele
- $VIF = 1/\text{tolerancia}$, ezért csak az egyiket kell értelmezni
- A tolerancia megadja, hogy a prediktor változó magyarázó erejének hány százaléka független, tehát nem fedhető le a többi prediktor változó által megmagyarázott résszel.
- Ha a tolerancia 10%-alatt van (vagy $VIF > 10$) a változó redundánsnak tekinthető, és érdemes kihagyni az elemzésből

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics ^③	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4,983	,565		8,817	,000		
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000	1,000	1,000
2	(Constant)	3,290	,560		5,876	,000		
	Reklámba fektetett összeg (millió Forint)	,402	,048	,623	8,439	,000	,845	1,183
	Utolsó videóklipp youtube nézettsége (ezer néző)	,003	,000	,397	5,379	,000	,845	1,183

a. Dependent Variable: Eladott koncertjegy (ezer db)

Többszörös lineáris regresszió értelmezés

- **Excluded Variables**

- A modellbe még be nem tett, vagy onnan már kiejtett változók statisztikái

Excluded Variables^a

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Utolsó videóklipp youtube nézettsége (ezer néző)	,397 ^b	5,379	,000	,584	,845	1,183	,845

a. Dependent Variable: Eladott koncertjegy (ezer db)

b. Predictors in the Model: (Constant), Reklámba fektetett összeg (millió Forint)

- **Collinearity Diagnostics**

- További tábla a kollinearitásról,de amit akartunk, már megtudtuk a tolerancia értékekből

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Reklámba fektetett összeg (millió Forint)	Utolsó videóklipp youtube nézettsége (ezer néző)
1	1	1,847	1,000	,08	,08	
	2	,153	3,472	,92	,92	
2	1	2,709	1,000	,02	,03	,02
	2	,166	4,042	,15	,96	,25
	3	,125	4,646	,83	,01	,73

a. Dependent Variable: Eladott koncertjegy (ezer db)

Többszörös lineáris regresszió értelmezés

- **Casewise Diagnostics**

- Gyanús esetek az adataink között:
 - 2-nél nagyobb standard reziduális hibával rendelkeznek
 - őket tudjuk legkevésbé bejósolni, ezért érdemes rájuk nézni
- Természetes, hogy van néhány ilyen érték. Mivel a hiba normál eloszlást követ,
 - a minta 5%-ánál várható, hogy a hibatag kívül esik a ± 1.96 -os Z-értéken (szóráson), és a minta 1% esik kívül a ± 2.5 -ös határon
 - (59 fős a minta, tehát kb. 3 illetve fél esetet várunk, ez szerencsére teljesül)

Casewise Diagnostics^a

Case Number	Std. Residual	Eladott koncertjegy (ezer db)	Predicted Value	Residual
36	-2,253	7,560	11,82272	-4,262718
53	2,129	12,895	8,86850	4,026504
55	2,266	13,680	9,39354	4,286459

a. Dependent Variable: Eladott koncertjegy (ezer db)

Többszörös lineáris regresszió értelmezés

⑩ Residuals Statistics

- Többdimenziós outlierok ellenőrzése
- Mahalanobis:** a táblázatból kikeresve 16.27 a kritérium
- Cook's distance:** 1 a kritérium
- Centered Leverage:**
 - itt $2 \cdot (2+1)/59 = 0.10$ vagy $3 \cdot (2+1)/59 = 0.15$ a kritérium
- Dfbeta értékek:** ± 2 a kritérium
- Új változók jöttek létre, ha van, ott kell kikeresni a problémás értékeket (pl. az Extreme values táblával)
- A négy módszer nem mindig hoz azonos eredményt, ekkor mérlegelni kell, kihagyjuk-e az adott elemet

Residuals Statistics^a

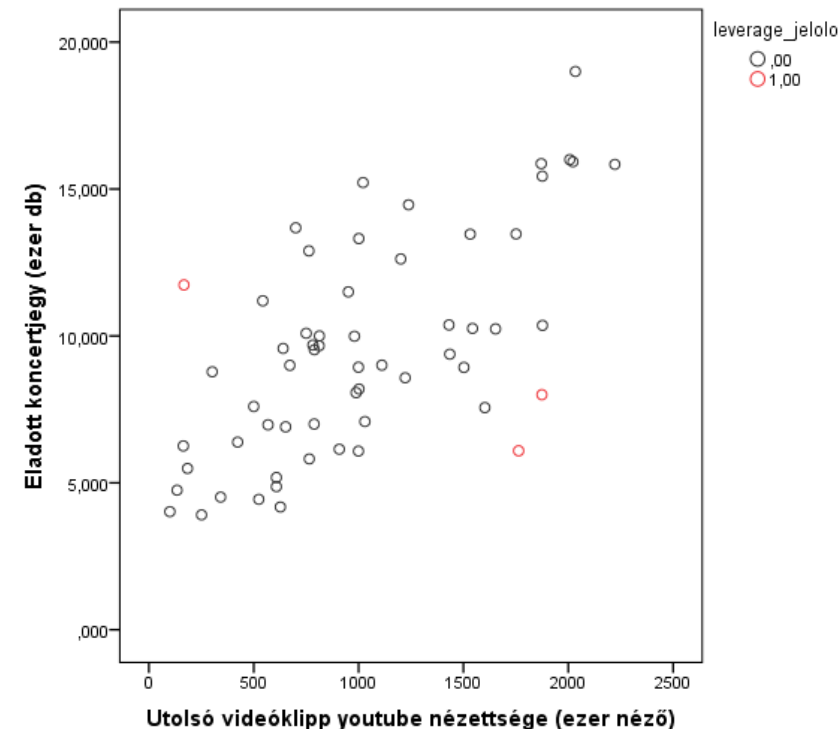
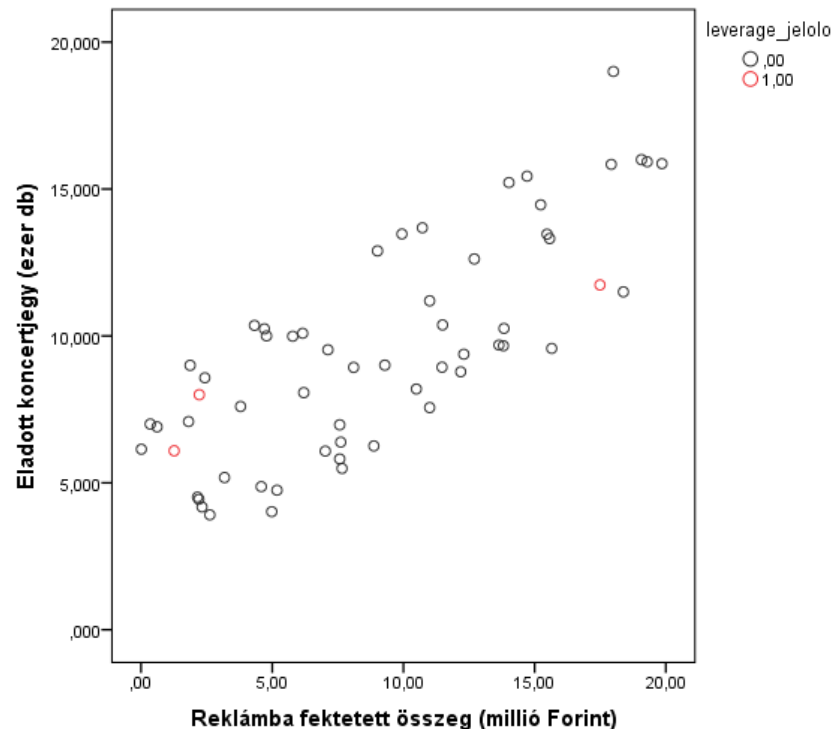
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4,98680	16,23129	9,48331	3,148902	59
Std. Predicted Value	-1,428	2,143	,000	1,000	59
Standard Error of Predicted Value	,250	,719	,411	,114	59
Adjusted Predicted Value	5,04957	16,26439	9,48521	3,142641	59
Residual	-4,262718	4,286459	,000000	1,858773	59
Std. Residual	-2,253	2,266	,000	,983	59
Stud. Residual	-2,295	2,298	,000	1,003	59
Deleted Residual	-4,421225	4,408982	-,001910	1,938978	59
Stud. Deleted Residual	-2,389	2,393	,003	1,019	59
Mahal. Distance	,032	7,388	1,966	1,685	59
Cook's Distance	,000	,106	,014	,019	59
Centered Leverage Value	,001	,127	,034	,029	59

a. Dependent Variable: Eladott koncertjegy (ezer db)

	koncertjegy	reklam	youtube	megosztás	nyilatkozatok	hasonlok	MAH_2	COO_2
1	10,000	4,78	813	180,00	43,00	6	,54609	,01892
2	10,090	6,16	750	172,00	28,00	5	,32704	,01268
3	9,000	1,87	672	102,00	35,00	6	1,57315	,04696
4	9,575	15,65	640	100,00	33,00	6	2,87015	,01921
5	9,665	13,81	812	260,00	44,00	5	1,28906	,00625
6	15,220	14,02	1021	312,00	19,00	3	,92582	,04431
7	13,465	15,47	1532	390,00	20,00	4	1,57813	,00000
8	12,620	12,70	1201	354,00	22,00	3	,44168	,00315

Többszörös lineáris regresszió értelmezés

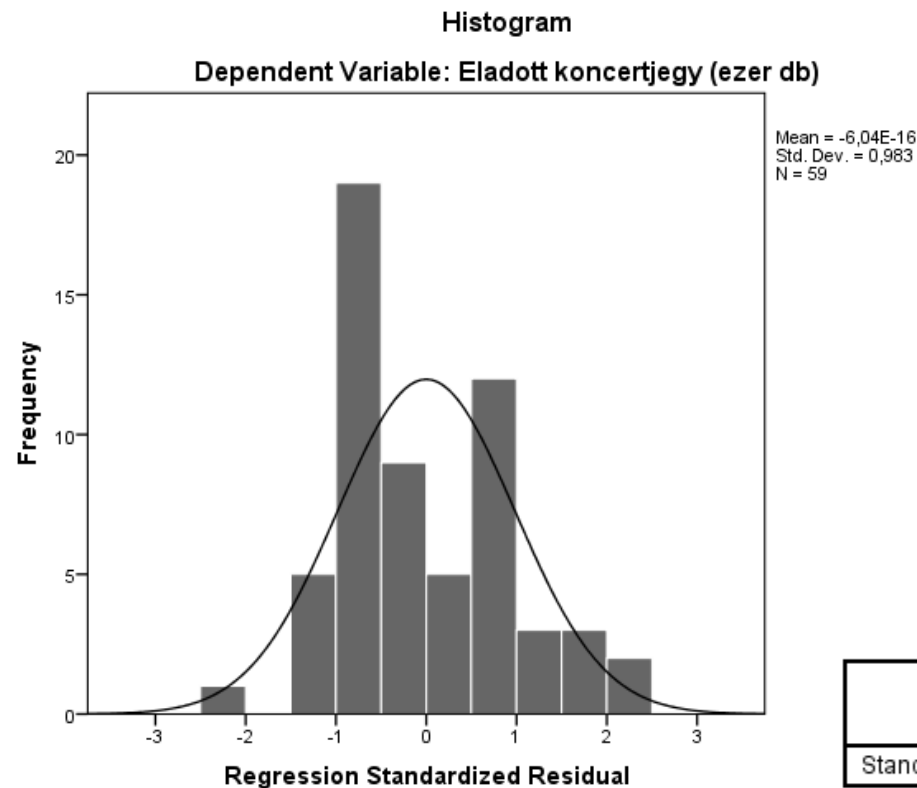
- **Grafikus segítség az outlierok kidobásának mérlegeléséhez**
 - A mintában csak a Centered leverage alapján találtunk outliereket. Ezek könnyen megtalálhatóak, az Explore menü Extreme values táblájával.
 - Tovább segíthet a grafikus megjelenítés. Ehhez Transform / Recode into different variable-ben hozzunk létre egy új változót, mely 0, ha leverage ≤ 0.10 és 1, ha leverage ≥ 0.10
 - Scatter plot-on ábrázoljuk a prediktor változók és a kimeneti változó kapcsolatát, ezt az új változót Marker-ként beállítva.
- **Ha kiveszünk outliert, az elemzést újra kell futtatni!!!**



Többszörös lineáris regresszió értelmezés

• ⑧ Hisztogram

- a standardizált reziduálisok eloszlása normál eloszlást kell kövessen
- Ha a hisztogram alapján bizonytalan vagy, a regresszió kikérésekor a Save opción belül elmentettük a standardizált reziduális hibák értékét, és az így létrejövő változón elvégezheted a tanult normalitás teszteket



ZRE_1
1,42844
1,26853
1,70915
-,86942
-,66468
1,94351
,01407



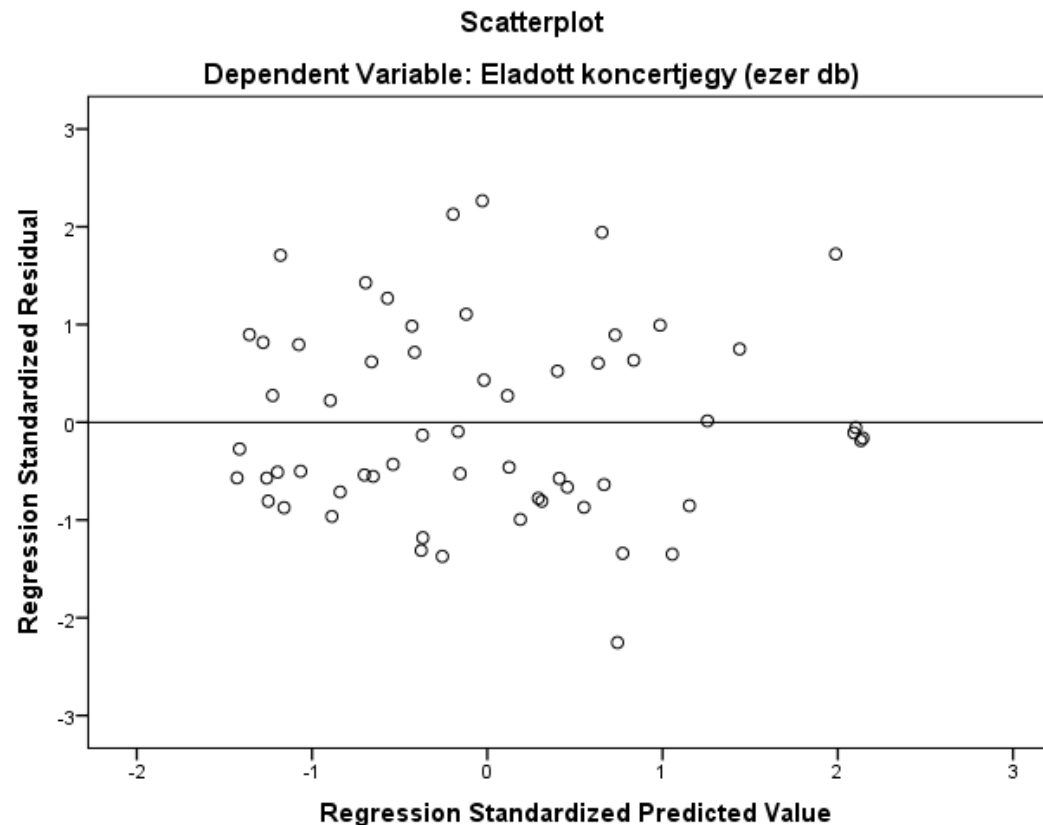
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	,127	59	,019	,968	59	,128

a. Lilliefors Significance Correction

Többszörös lineáris regresszió értelmezés

- ⑤,⑥ **Homoszkedaszticitás és linearitás ellenőrzése**
- 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



Többszörös lineáris regresszió értelmezés

- ⑤ **Homoszkedaszticitás**
- Ha a pontdiagram alapján nem vagy biztos a döntésedben, elvégezheted a Breusch-Pagan vagy Koenker tesztet, melyek makrója letölthető a weblapról (a makrórt Marta Garcia-Granero kódolta és David Marso módosította)
- A makró három részből áll:
 - 1. az elemzés definiálása – ezt csak EGYSZER a legelején kell lefuttatni
 - 2. próbafájl – ellenőrizhető, hogy sikerült-e az előző lépés, kihagyható
 - 3. a BPK teszt - A következő kódot kell szerkeszteni:
BPKTEST x1 19 x2 TO x20.
 - **x1** helyére írd be a KIMENETI változót (pl.: **koncertjegy**)
 - **19** helyére írd be, **hány PREDIKTOR** változó van
 - **x2 TO x20** helyére sorold fel vesszővel elválasztva a PREDIKTOR változókat (pl.: **reklam youtube nyilatkozatok**)
 - Ennek megfelelően a ki kódunk így fog kinézni:
 - **BPKTEST koncertjegy 3 reklam youtube nyilatkozatok.**

```
*****
*ezt a részt futtasd EGYSZER a legelején!
*****

DEFINE bpktest(
  !POSITIONAL !TOKENS(1)
  /!POSITIONAL !TOKENS(1)
  /!POSITIONAL !CMDEND).
* Regression to GET the residuals and residual plots.
REGRESSION
  /STATISTICS R ANOVA
  /DEPENDENT !1
  /METHOD=ENTER !3
  /SCATTERPLOT=(*ZRESID,*ZPRED)
  /RESIDUALS HIST(ZRESID) NORM(ZRESID)
  /SAVE RESID(residual) .
DO IF $casenum=1.
PRINT /"Examine the scatter plot of the residuals to detect"
      /"model misspecification and/or heteroscedasticity"
      /""
*****

*ezzel a résszel futtathatod a saját adataidon.
*x1 helyére írd be a saját KIMENETI változód nevét!
*ezt követően add meg, hány PREDIKTOR változód van.
*végül sorold fel a prediktorváltozóidat.
*ha kevés van, megadhatod mindegyiknek külön a nevét.
*ha sok van, akkor add meg az elsőt, majd TO, majd az utolsót.

* x1 is the dependent and x2 TO x20 the predictors.

* (3) MACRO CALL (select and run).

BPKTEST x1 19 x2 TO x20.
```

Többszörös lineáris regresszió értelmezés

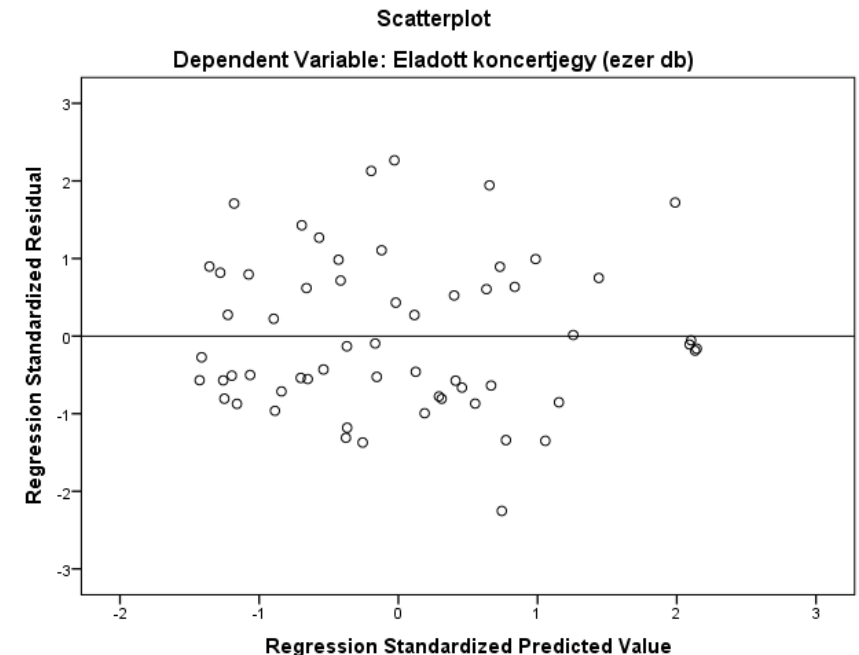
- A kapott outputnak csak az utolsó sorai kellenek:
 - Mind a Breusch-Pagan, mind a Koenker azt teszteli, hogy a reziduális hiba varianciája függ-e a prediktor változók értékétől (ekkor nem teljesül a szóráshomogenitás). A Koenker teszt általánosabb, a Breusch-Pagan tesztnél pontosabban működik kisebb mintákon illetve a normalitás sérülése esetén.
 - Ha a teszt NEM szignifikáns, akkor teljesül a szóráshomogenitás feltétele.

```
Breusch-Pagan test for Heteroscedasticity (CHI-SQUARE df=P)  
      ,478
```

```
Significance level of Chi-square df=P (H0:homoscedasticity)  
      ,9237
```

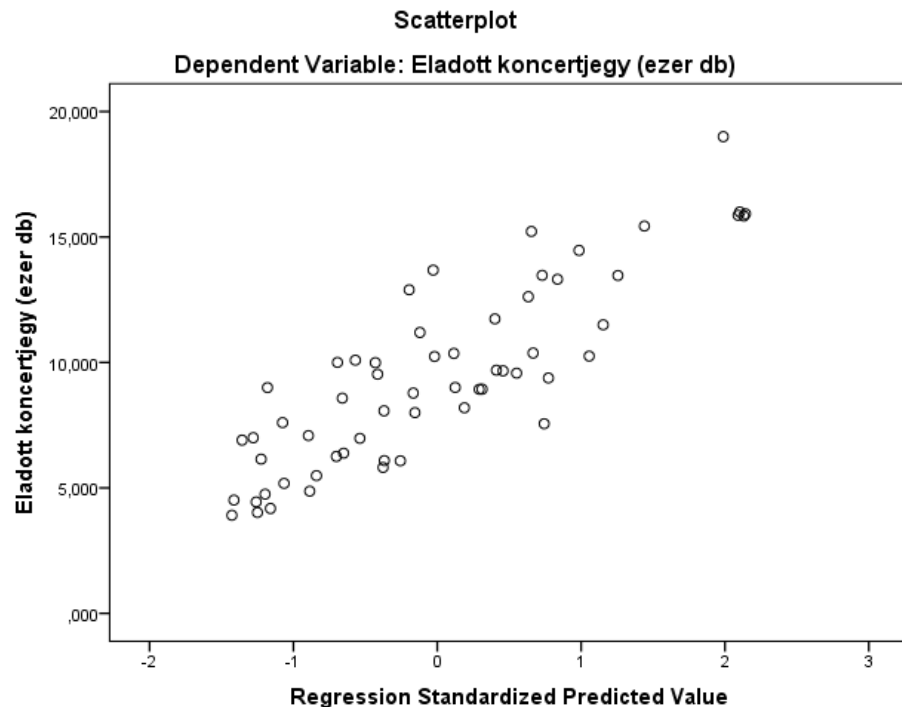
```
Koenker test for Heteroscedasticity (CHI-SQUARE df=P)  
      ,712
```

```
Significance level of Chi-square df=P (H0:homoscedasticity)  
      ,8703
```



Többszörös lineáris regresszió értelmezés

- ⑥ **A standardizált predikált és a mért kimeneti értékek pontdiagramja** (mivel nagyon hasonló információt tartalmaz, mint az előző, nem szokás mindkettőt kikérni)
 - Azt várjuk, hogy az általunk predikált érték kövessék a ténylegesen mért értékeket – pozitív, lineáris kapcsolatot várunk a kettő között. A pontfelhő vastagsága a becslés pontosságáról ad információt.
 - Könnyebben értelmezhető a plot, ha nem standardizálva jelenítjük meg a predikált értékeket. A Save opcióban menthetjük el a standardizálatlan predikált értékeket, majd Scatter plottal ábrázolhatjuk a mért értékekkel való kapcsolatát.



Linear Regression: Save

Predicted Values

- ☒ Unstandardized
- ☐ Standardized
- ☐ Adjusted
- ☐ S.E. of mean predictions

Residuals

- ☐ Unstandardized
- ☐ Standardized
- ☐ Studentized
- ☐ Deleted
- ☐ Studentized deleted

Distances

- ☒ Mahalanobis
- ☒ Cook's
- ☒ Leverage values

Prediction Intervals

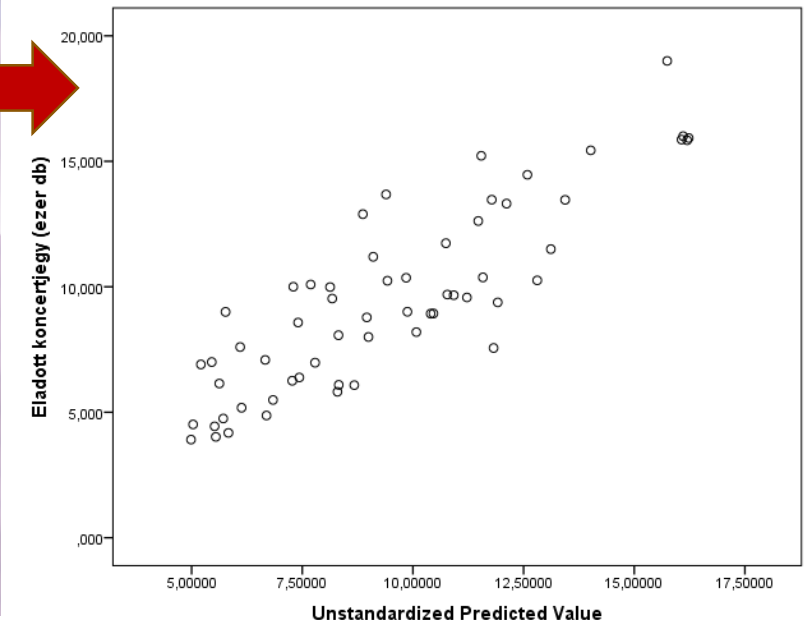
- ☐ Mean ☐ Individual
- Confidence Interval: 95 %

Influence Statistics

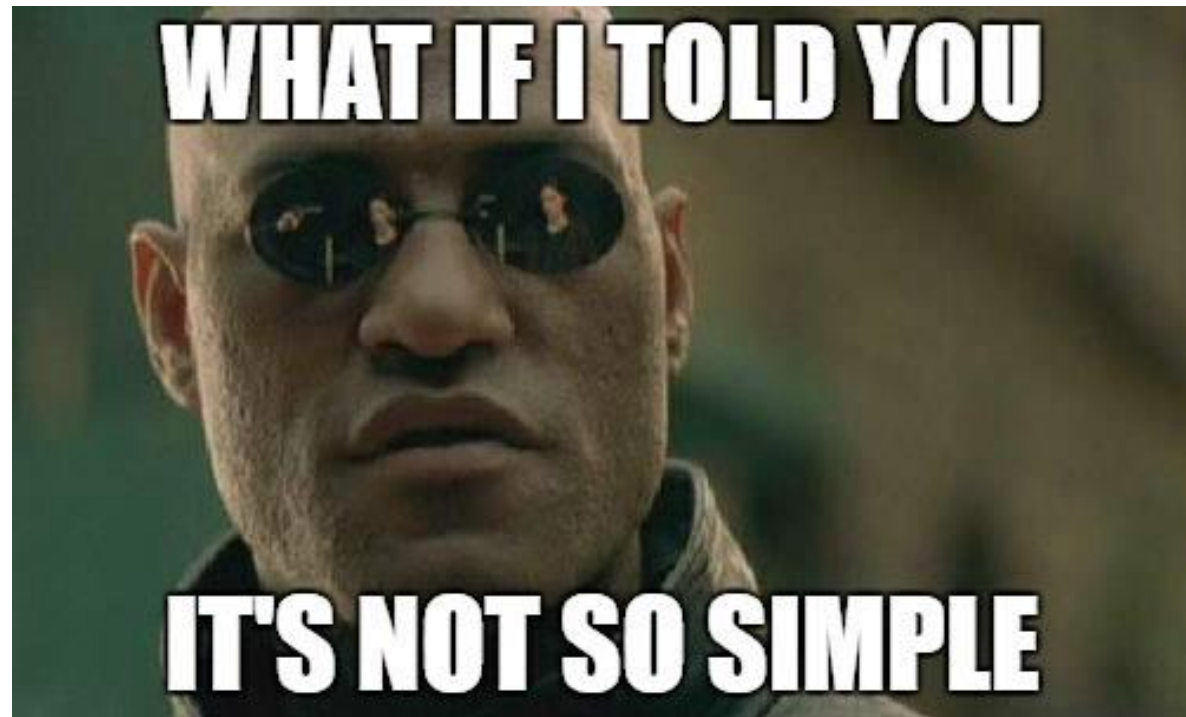
- ☐ DfBeta(s)
- ☐ Standardized DfBeta(s)
- ☐ DfFit
- ☐ Standardized DfFit
- ☐ Covariance ratio

Coefficient statistics

- ☐ Create coefficient statistics
- ☒ Create a new dataset
- Dataset name:

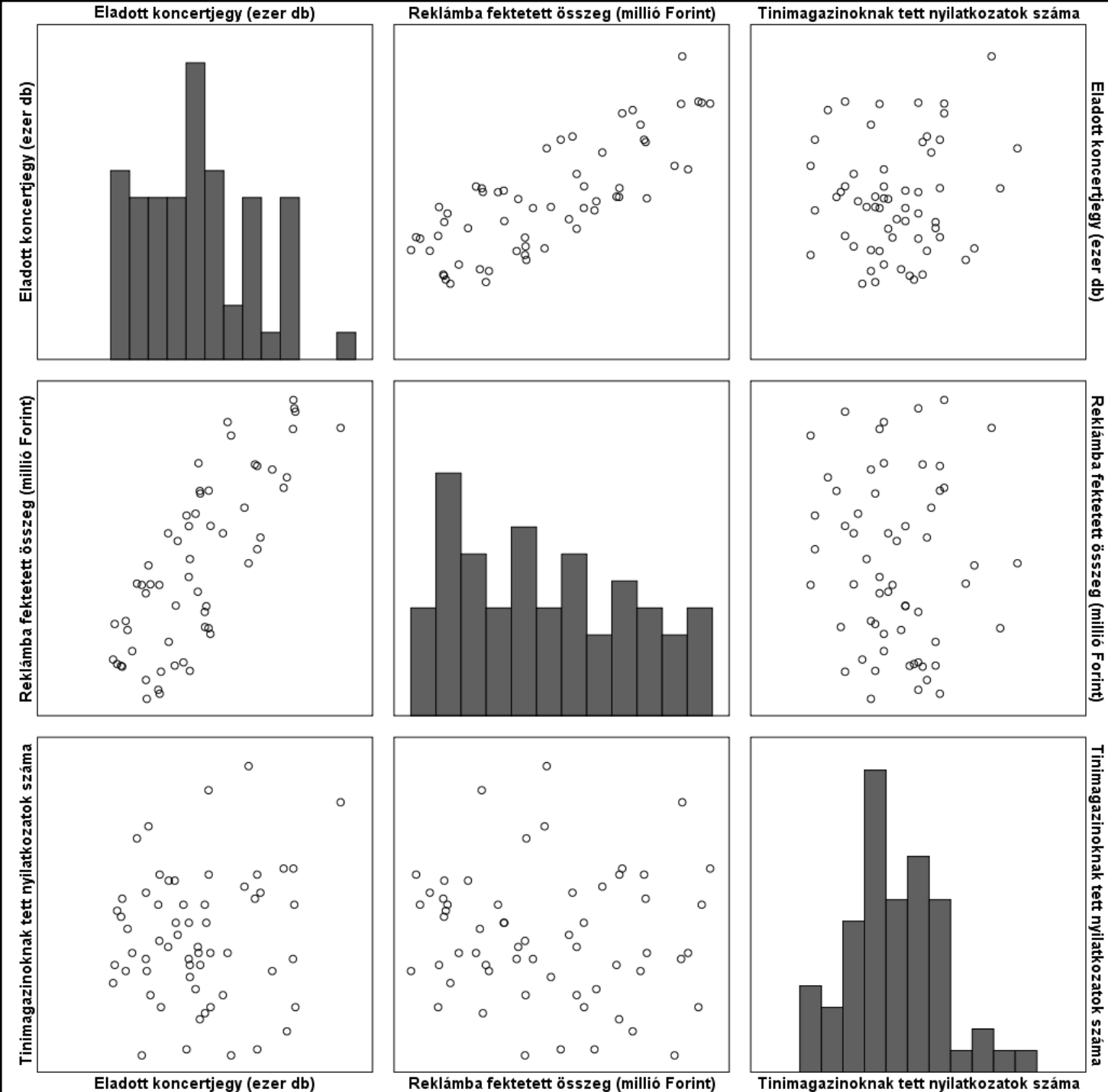
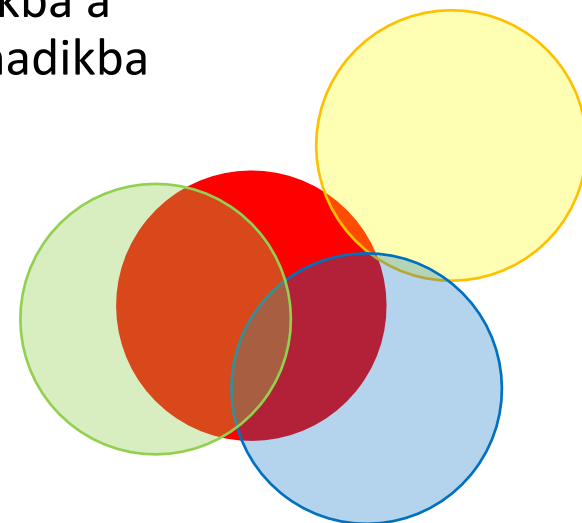


Amikor nem ilyen egyszerű



Változó, melynek nincs szignifikáns hatása

- Nézzünk példát arra, hogyan néz ki, ha egy változónak nincs a kimeneti változóra szignifikáns hatása.
- Ezt kétféleképp is kikérjük.
- Az első elemzésben az első modellbe kerüljön be a reklám, és a másodikba a youtube és a nyilatkozatok
- A második elemzésben, az első modellbe a reklám, a másodikba a youtube, és harmadikba a nyilatkozatok.



Variables Entered/Removed^a

Model	Variables Entered	Method
1	Reklámba fektetett összeg (millió Forint) ^b	Enter
2	Tinimagazinoknak tett nyilatkozatok száma, Utolsó videóklipp youtube nézettsége (ezer néző) ^b	Enter

a. Dependent Variable: Eladott koncertjegy (ezer db)

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,780 ^a	,608	,601	2,309072	,608	88,447	1	57	,000	
2	,869 ^b	,755	,741	1,859518	,147	16,446	2	55	,000	1,891

a. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Tinimagazinoknak tett nyilatkozatok száma, Utolsó videóklipp youtube nézettsége (ezer néző)

c. Dependent Variable: Eladott koncertjegy (ezer db)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	471,583	1	471,583	88,447	,000 ^b
	Residual	303,913	57	5,332		
	Total	775,496	58			
2	Regression	585,316	3	195,105	56,425	,000 ^c
	Residual	190,179	55	3,458		
	Total	775,496	58			

a. Dependent Variable: Eladott koncertjegy (ezer db)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

c. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Tinimagazinoknak tett nyilatkozatok száma, Utolsó videóklipp youtube nézettsége (ezer néző)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,983	,565		8,817	,000
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000
2	(Constant)	2,536	,704		3,604	,001
	Reklámba fektetett összeg (millió Forint)	,413	,047	,640	8,737	,000
	Utolsó videóklipp youtube nézettsége (ezer néző)	,002	,000	,379	5,157	,000
	Tinimagazinoknak tett nyilatkozatok száma	,039	,023	,116	1,719	,091

a. Dependent Variable: Eladott koncertjegy (ezer db)

- A tinimagazinoknak tett nyilatkozatok számának hatása nem szignifikáns.
- A modell ettől még szignifikáns és magas magyarázó erejű, mert a reklámba fektetett összeg és a youtube nézettség jól magyaráz.

Variables Entered/Removed^a

Model	Variables Entered	Method
1	Reklámba fektetett összeg (millió Forint) ^b	Enter
2	Utolsó videóklipp youtube nézettsége (ezer néző) ^b	Enter
3	Tinimagazinoknak tett nyilatkozatok száma ^b	Enter

a. Dependent Variable: Eladott koncertjegy (ezer db)

b. All requested variables entered.

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,780 ^a	,608	,601	2,309072	,608	88,447	1	57	,000	1,891
2	,861 ^b	,742	,732	1,891674	,133	28,929	1	56	,000	
3	,869 ^c	,755	,741	1,859518	,013	2,953	1	55	,091	

a. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Utolsó videóklipp youtube nézettsége (ezer néző)

c. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Utolsó videóklipp youtube nézettsége (ezer néző) száma

d. Dependent Variable: Eladott koncertjegy (ezer db)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	471,583	1	471,583	88,447	,000 ^b
	Residual	303,913	57	5,332		
	Total	775,496	58			
2	Regression	575,104	2	287,552	80,357	,000 ^c
	Residual	200,392	56	3,578		
	Total	775,496	58			
3	Regression	585,316	3	195,105	56,425	,000 ^d
	Residual	190,179	55	3,458		
	Total	775,496	58			

- Ha a modellt három lépésben építjük fel, és a tinimagazinokat a harmadik modellben léptetjük be, akkor láthatjuk, hogy az előző modellhez képest nem magyaráz szignifikánsan többet (74,2%-ról 75,6%-ra növekedett a magyarázóerő).
- Maga a harmadik modell attól még szignifikáns a reklám és youtube hatása miatt, de az F érték alacsonyabb, hiszen a modell bonyolultabbá vált (mivel kettő helyett három prediktor változót tartalmaz).

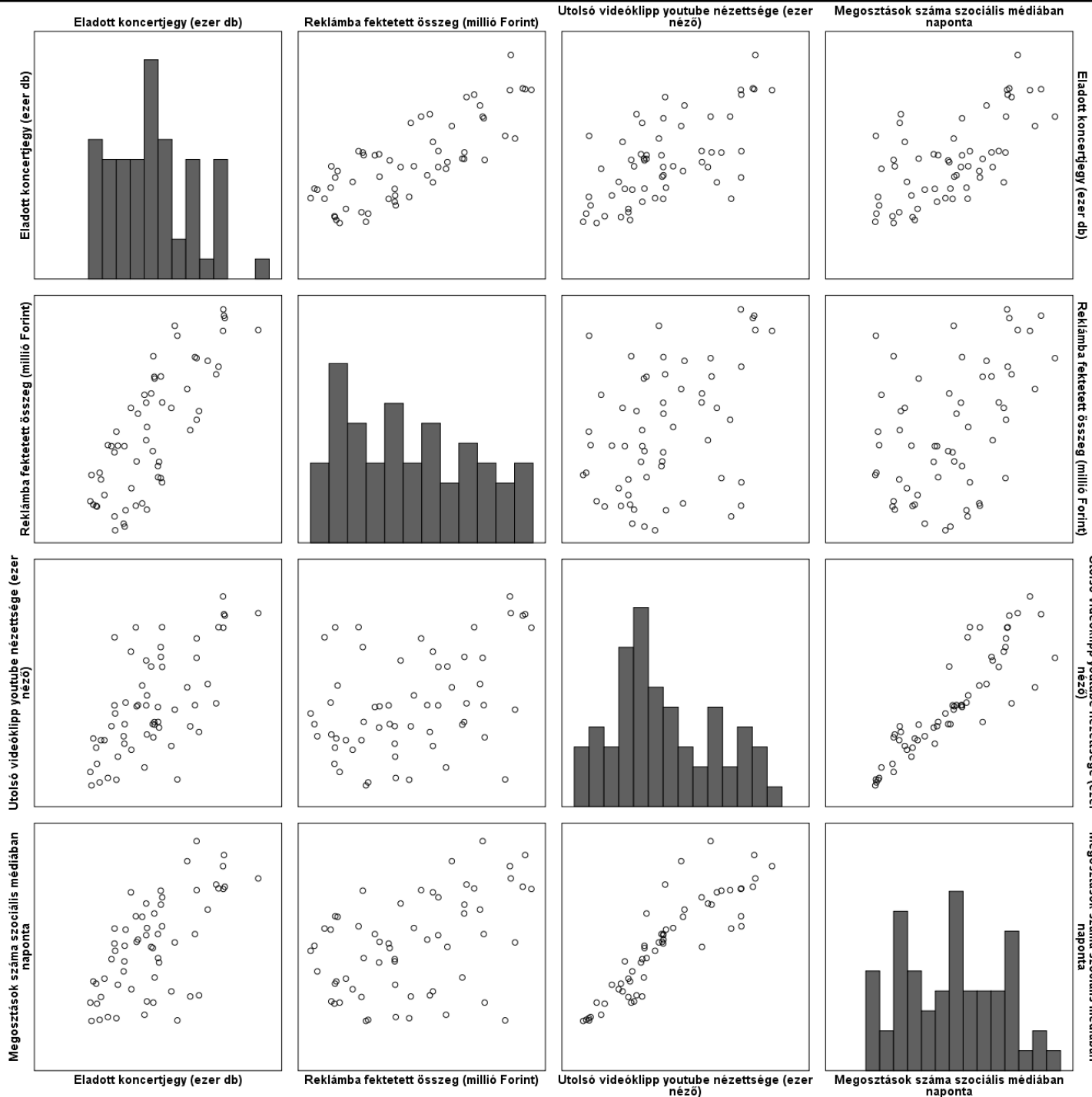
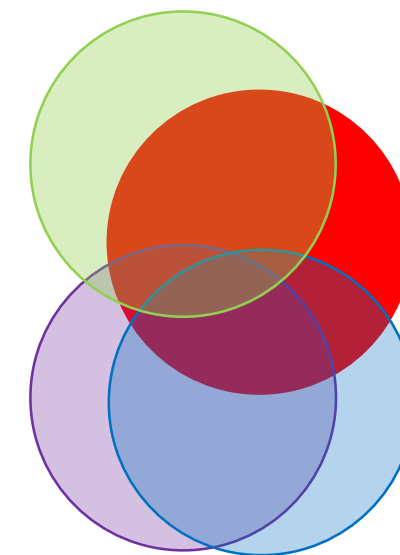
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,983	,565		8,817	,000
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000
2	(Constant)	3,290	,560		5,876	,000
	Reklámba fektetett összeg (millió Forint)	,402	,048	,623	8,439	,000
	Utolsó videóklipp youtube nézettsége (ezer néző)	,003	,000	,397	5,379	,000
3	(Constant)	2,536	,704		3,604	,001
	Reklámba fektetett összeg (millió Forint)	,413	,047	,640	8,737	,000
	Utolsó videóklipp youtube nézettsége (ezer néző)	,002	,000	,379	5,157	,000
	Tinimagazinoknak tett nyilatkozatok száma	,039	,023	,116	1,719	,091

Kollinearitás

Nézzünk példát arra, hogyan néz ki, ha két prediktor között kollinearitás van!

Az elemzést úgy építjük fel, hogy az első modellbe bekerül a reklám, a másodikba pedig a kollinearitásban résztvevő két prediktor, a youtube és a megosztások száma.



Variables Entered/Removed^a

Model	Variables Entered	Method
1	Reklámba fektetett összeg (millió Forint) ^b	Enter
2	Utolsó videóklipp youtube nézettsége (ezer néző), Megosztások száma szociális médiában naponta ^b	Enter

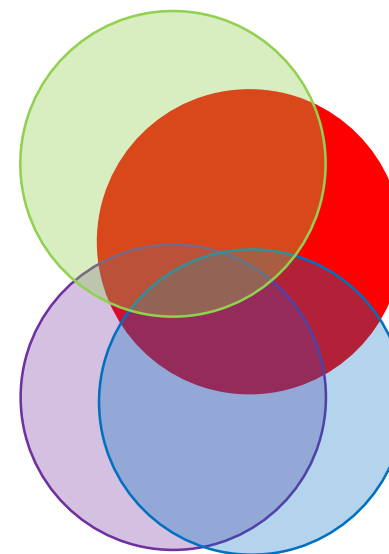
Correlations

		Eladott koncertjegy (ezer db)	Reklámba fektetett összeg (millió Forint)	Utolsó videóklipp youtube nézettsége (ezer néző)	Megosztások száma szociális médiában naponta
Pearson Correlation	Eladott koncertjegy (ezer db)	1,000	,780	,643	,639
	Reklámba fektetett összeg (millió Forint)	,780	1,000	,393	,435
	Utolsó videóklipp youtube nézettsége (ezer néző)	,643	,393	1,000	,884
	Megosztások száma szociális médiában naponta	,639	,435	,884	1,000

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4,983	,565		8,817	,000		
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000	1,000	1,000
2	(Constant)	3,091	,671		4,607	,000		
	Reklámba fektetett összeg (millió Forint)	,396	,049	,615	8,102	,000	,811	1,234
	Utolsó videóklipp youtube nézettsége (ezer néző)	,002	,001	,329	2,248	,029	,219	4,572
	Megosztások száma szociális médiában naponta	,003	,006	,082	,546	,587	,210	4,765

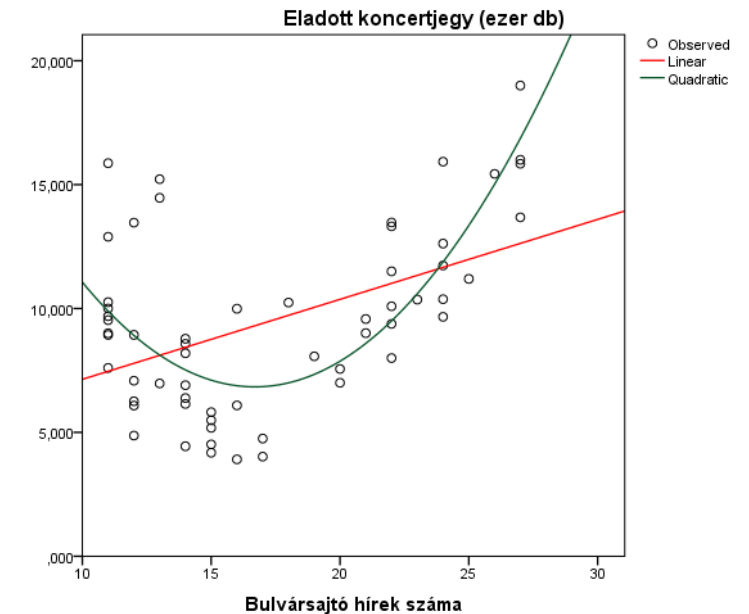
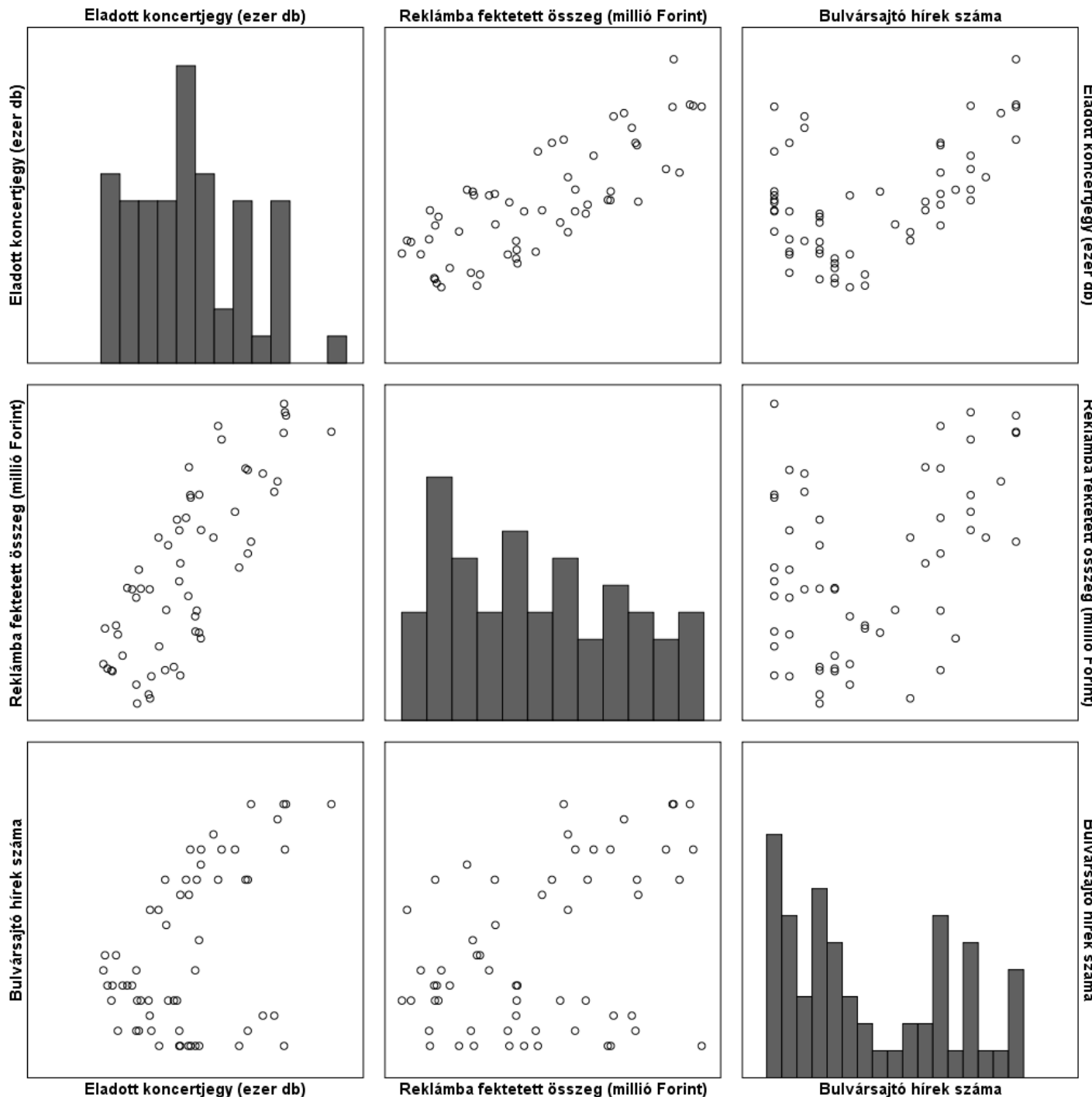
- A kollinearitás a grafikon mellett a korrelációs táblában is jól látszik.
- A modell egyébként szignifikáns, $F(3, 55) = 53.00$ $p < .001$, $R^2 = .743$
- További gyanús jel, hogy a kollinearitásban résztvevő két változó toleranciája nagyon alacsony.
- A két prediktor közül a youtube hatása erősen meggyengült, a megosztások pedig egyáltalán nem szignifikáns.



Linearitás hiánya

Nézzünk példát arra, hogyan néz ki, ha olyan prediktort helyezünk a modellbe, mely nincs lineáris kapcsolatban a kimeneti változóval.

Görbeillesztéssel látható (később tanuljuk), hogy a koncertjegy eladás varianciáját a bulvársajtó a lineáris modellel 22%-ban, a kvadratikus modellel 52.8%-ban magyarázza.



Variables Entered/Removed^a

Model	Variables Entered	Method
1	Reklámba fektetett összeg (millió Forint) ^b	Enter
2	Bulvársajtó hírek száma ^b	Enter

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change
1	,780 ^a	,608	,601	2,309072	,608
2	,797 ^b	,634	,621	2,250028	,026

a. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

b. Predictors: (Constant), Reklámba fektetett összeg (millió Forint), Bulvársajtó hírek száma

c. Dependent Variable: Eladott koncertjegy (ezer db)

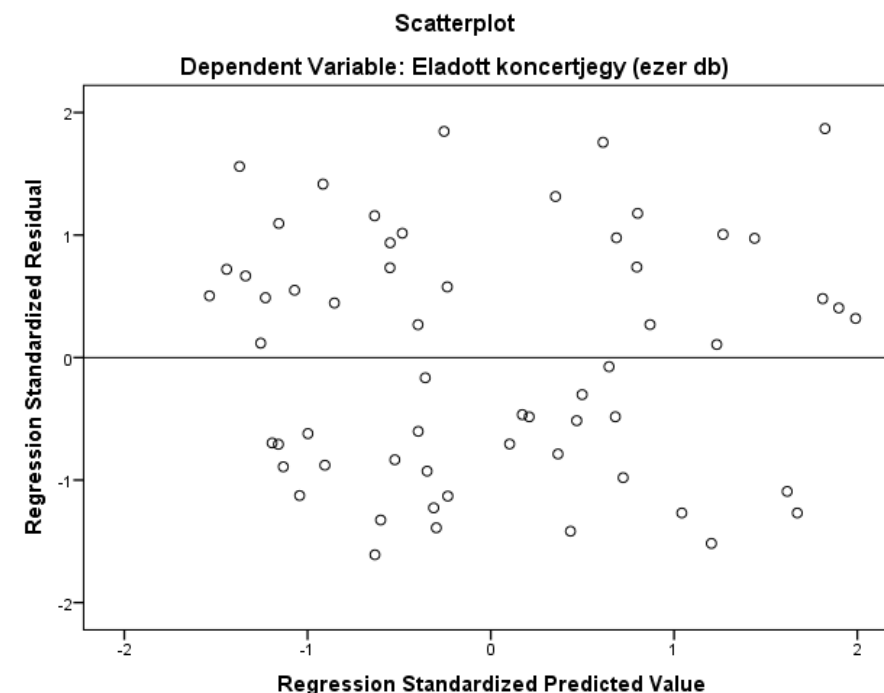
ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	471,583	1	471,583	88,447	,000 ^b
	Residual	303,913	57	5,332		
	Total	775,496	58			
2	Regression	491,989	2	245,994	48,590	,000 ^c
	Residual	283,507	56	5,063		
	Total	775,496	58			

Coefficients^a

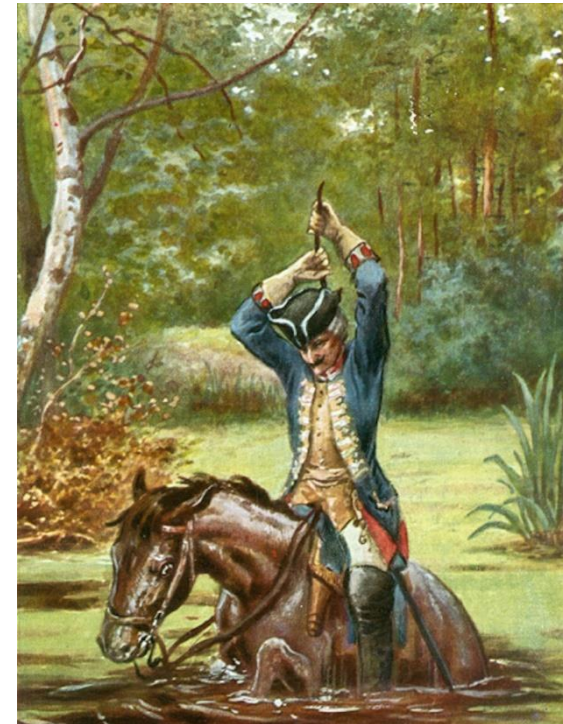
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4,983	,565		8,817	,000		
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000	1,000	1,000
2	(Constant)	3,308	1,001		3,306	,002		
	Reklámba fektetett összeg (millió Forint)	,456	,057	,708	8,013	,000	,836	1,196
	Bulvársajtó hírek száma	,121	,061	,177	2,004	,050	,836	1,196

- A modell ugyan szignifikáns (a reklám hatása miatt), de úgy tűnik, hogy a bulvárszerepléseknek csak nagyon kicsi, 2,6%-nyi plusz információtartalma van. E változó hatása nem is szignifikáns. Ez nem (multi)kollinearitás miatt van, hiszen a tolerancia értékek magasak.
- Ha egyenes helyett négyzetes görbét illesztünk a bulvársajtó hatására, 52%-os hatást kapnánk.
- Vigyázat! A linearitás sérülése nem mindig látszik az összes prediktor közös hatását ábrázoló pontdiagramon (ez történik most is), a regresszió előtt mindig érdemes megnézni az egyedi pontdiagramokat is!**



Béták összehasonlíthatósága

Meredekséghez tartozó konfidencia intervallumok számítása Bootstrappinggel



Különbség béták közt

- A **Bootstrap** egy a mintából való többszöri, ismételt mintavételezésen alapuló módszer
 - Becsülhető vele a végzett statisztika megbízhatósága
 - Csökkenthető az outlierok torzító hatása
 - Csökkenthető a normalitás sérüléséből adódó torzítás
 - Bootstrap csak az újabb SPSS-ekben van

Különbség béta értékek között

- Nézzük egy egymodelles regressziót reklámmal, youtube-bal, tinimagazinoknak tett nyilatkozatokkal!
- A Beta értékekből látjuk, hogy a reklámnak van a legnagyobb hatása, ezt követi a youtube, majd a nyilatkozatok

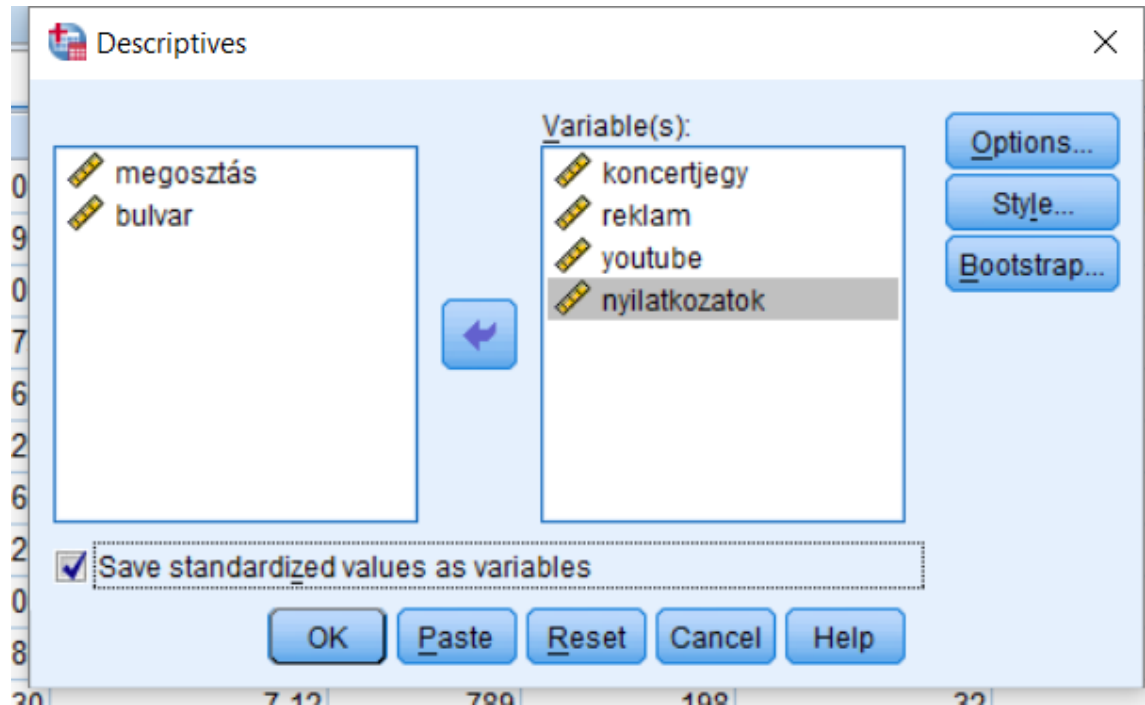
Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,536	,704		3,604	,001
	Reklámba fektetett összeg (millió Forint)	,413	,047	,640	8,737	,000
	Utolsó videóklipp youtube nézettsége (ezer néző)	,002	,000	,379	5,157	,000
	Tinimagazinoknak tett nyilatkozatok száma	,039	,023	,116	1,719	,091

a. Dependent Variable: Eladott koncertjegy (ezer db)

- Kérdés, hogy a hatások nagysága közötti különbség szignifikáns-e?
- Ennek megállapításához ki kell számolni a Béta értékekhez tartozó konfidencia intervallumokat. Két egymással maximum 50%-ban átfedő CI intervallum esetében a különbséghez $p < .05$ valószínűség tartozik, tehát ha két Béta értékhez tartozó konfidencia intervallum 50-nél kisebb százalékban fed át, szignifikáns különbséget feltételezhetünk.

Különbség béta értékek között

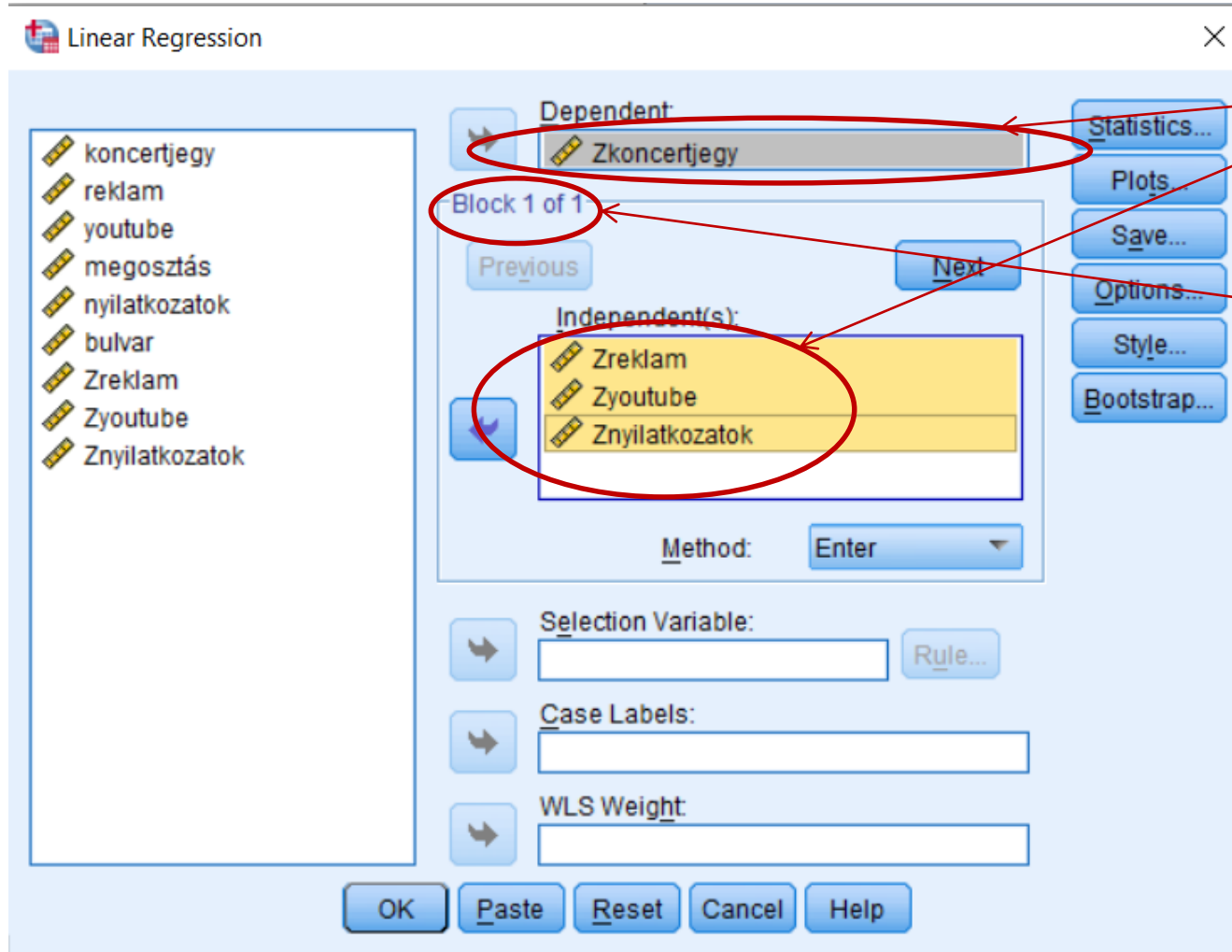
- A CI-k kiszámolásához először létre kell hozni a modellben szereplő változók standardizált értékét. Ezt az Analyze / Descriptive / Descriptive-ben tehetjük meg



Zkoncertjegy	Zreklam	Zyoutube	Znyilatkozatok
,14131	-,73615	-,34901	-1,17921
,16592	-,49302	-,46036	,21226
-,13217	-1,24884	-,59823	-,43709
,02508	1,17897	-,65479	-,25156
,04969	,85479	-,35078	-1,27197
1,56887	,89179	,01863	1,04714
1,08891	1,14725	,92183	,95437
,85782	,65923	,33678	,76884
-,51504	-,91057	-,90224	,86161
-,84732	-,23755	-1,03833	-,90091
,01277	-,32388	-,39143	-,15880
-,67913	-1,51664	-,39320	,95437
1,04789	1,16487	-,01672	,58331
-1,52418	-1,11671	-1,34235	-,71539
-1,35873	-1,19951	-1,18150	,58331
-,70511	-1,47083	-,63358	,49055
,05653	,82484	-,40380	-,43709
-,91296	-1,57654	-,18109	-,52986

Különbség béta értékek között

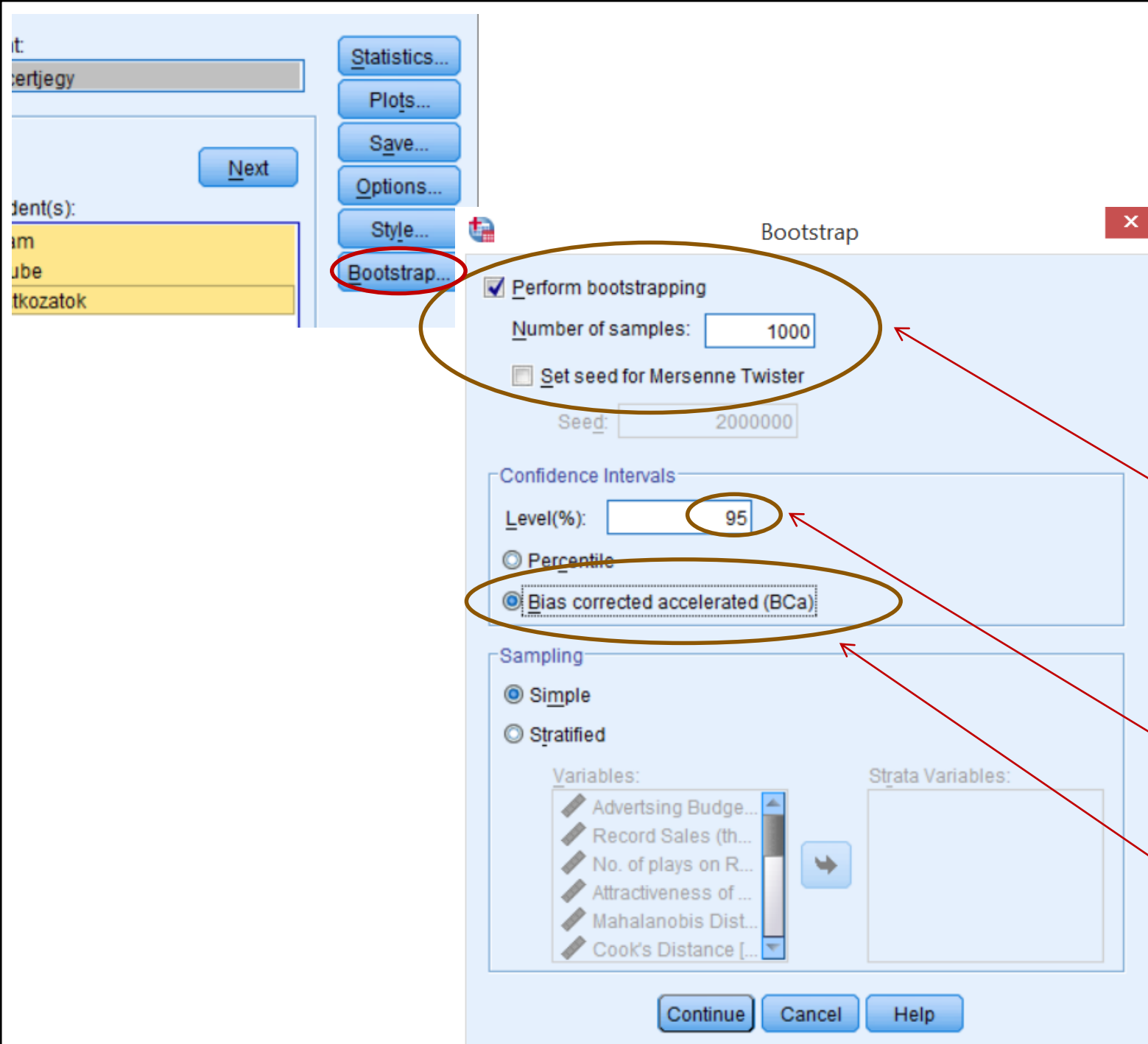
- Majd ugyanazzal a modellfelépítéssel, de az eredeti változók helyett a most létrehozott standardizált értékekkel újra futtatjuk a hierarchikus regressziót.



A modellben mindenhol kicseréltem az eredeti változókat a z-értékekre.

Ne feledkezz meg arról, hogy a hierarchikus regressziónak több blokkja is lehet, mindenhol cseréld ki az értékeket

Különbség béták közt



- Az elemzést Bootstrapping-et beállítanva kell lefuttatni.
(érdeemes a Bootstrapet kikéve minden opciót kikapcsolni)
- Ha erős a géped, 1000 helyett használhatsz 10 000 mintát is, de sokáig fog tartani. 500 mintavétel alá nem szabad, 10 000 fölé nem érdemes menni.
- Ha 95CI helyett mást szeretnél, itt állíthatod be
- Mindig Bias korrigált számítási módot érdemes használni.

Különbség béták között

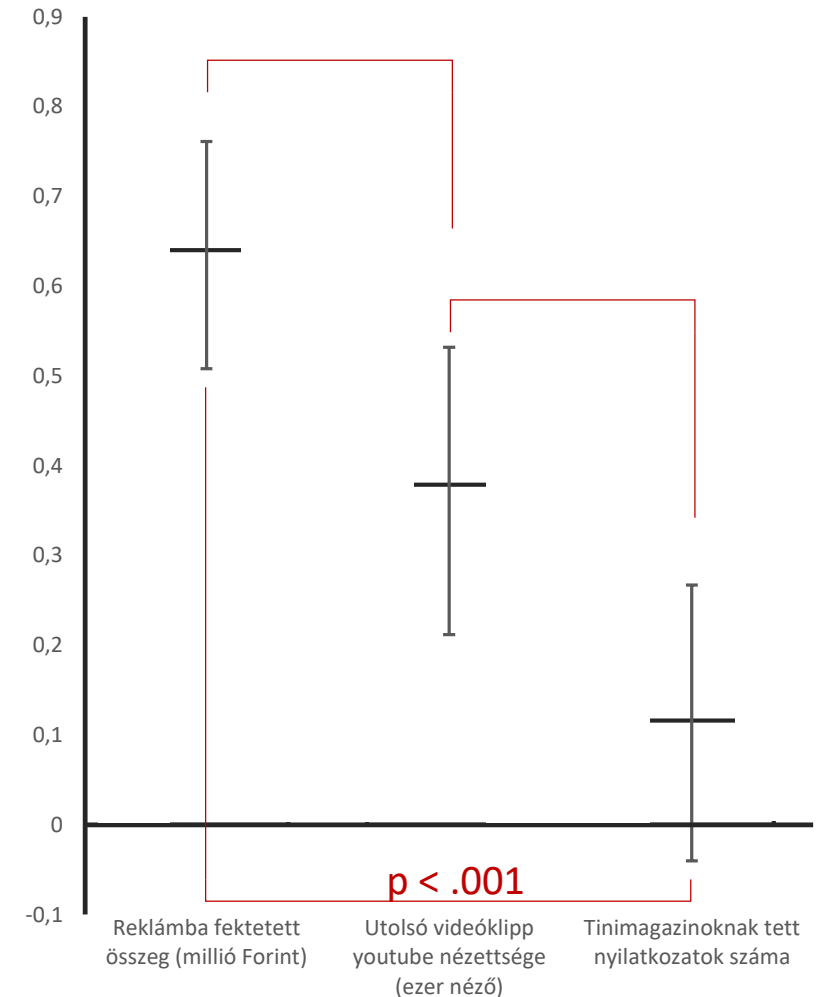
- Csak a Bootstrap táblát vesszük figyelembe, ahonnan a változókhoz tartozó B értékek 95%-os konfidencia intervallumának alsó és felső határát találjuk.
- Vedd észre, hogy a standardizált változók B értéke (standardizálatlan regressziós együttható) megegyezik a standardizálatlan változók β értékével (standardizált regressziós együttható).

Bootstrap for Coefficients						
Model	B	Bootstrap ^a				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
					Lower	Upper
1 (Constant)	1,034E+15	-,002	,068	1,000	-,122	,132
Zscore: Reklámba fektetett összeg (millió Forint)	,640	-,005	,070	,001	,508	,761
Zscore: Utolsó videóklipp youtube nézettsége (ezer néző)	,379	,000	,072	,001	,214	,532
Zscore: Tinimagazinoknak tett nyilatkozatok száma	,116	,000	,077	,138	-,040	,267

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Különbség béták között

- Ha két változó regressziós együtthatójához tartozó 95CI maximum 50%-ban fed át egymással, akkor a két regressziós együttható szignifikánsan különbözik egymástól $p < .05$ szinten.
- Ha nem fednek át, akkor szignifikánsan különböznek egymástól $p < .001$ szinten (Cumming és Finch, 2005)
- A három regressziós együtthatót és a konfidencia intervallumokat ábrázolva szemre is látszik, hogy reklám és nyilatkozatok CI intervalluma egyáltalán nem fed át egymással, tehát $p < .001$ szinten szignifikáns a különbség a meredekségek között.
- Ha az eredmény szemre is egyértelmű, nem kell a pontos értékeket kiszámolni.



Különbség béták között

- Ha szemre nem egyértelmű, ki kell számolni a két CI átfedését.
- A számítás szemléltetéséhez használjuk a következő értékeket:

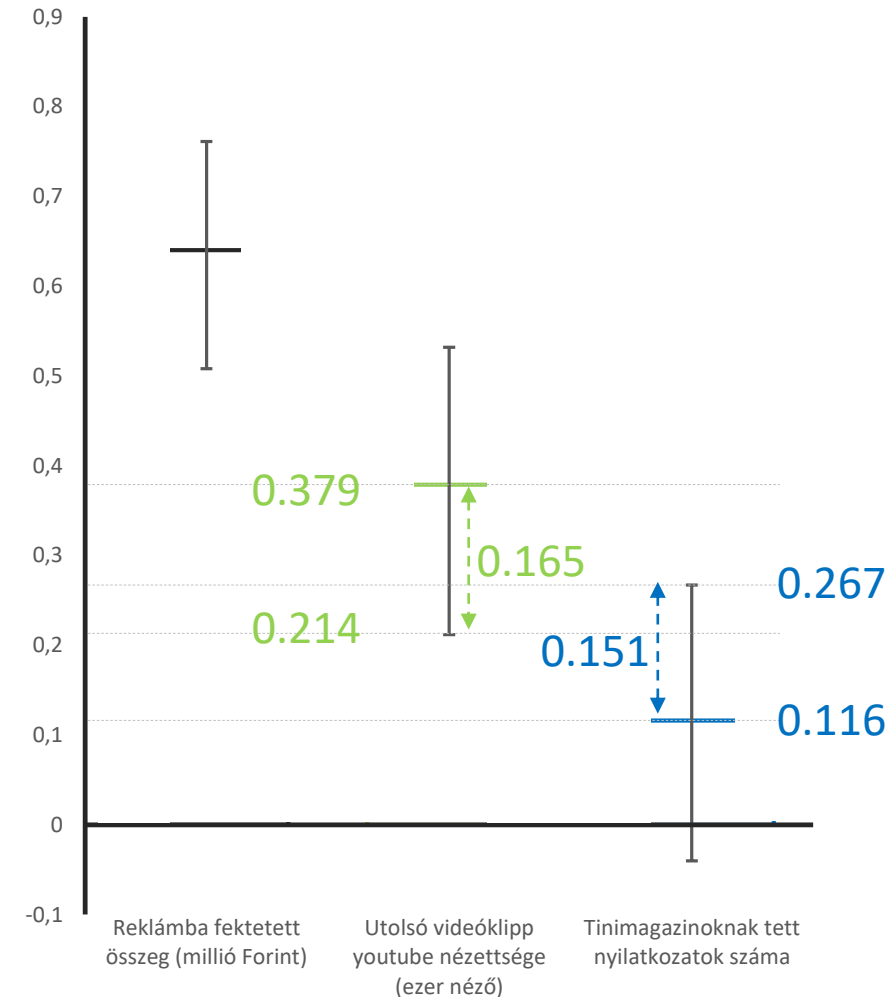
	95CI alsó határa	β értéke	95CI felső határa
X	0.214	0.379	0.532
Y	-0.040	0.116	0.267

- 1. lépés:
számítsuk ki a magasabb bétához tartozó konfidencia intervallum alsó felének szélességét, és az alacsonyabb béta értékhez tartozó konfidencia intervallum felső felét szélességét.

(a bootstrappinggel kapott CI-k nem feltétlenül szimmetrikusak)

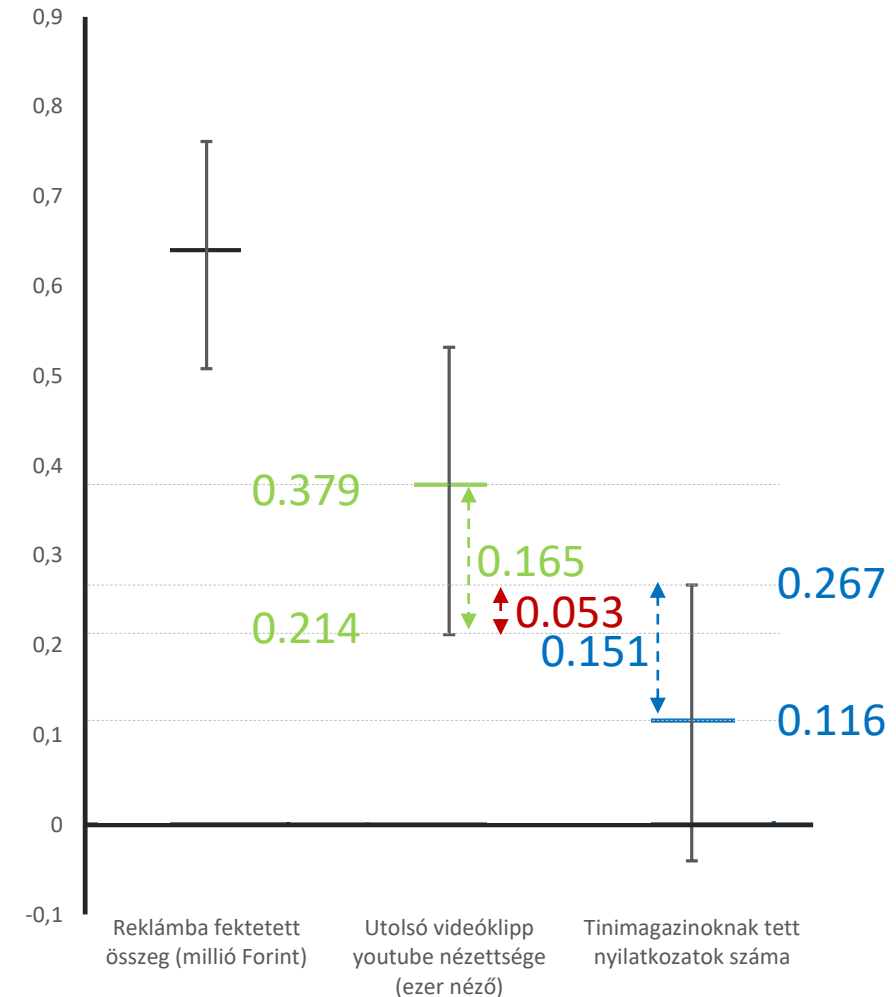
$$\text{CIszélesség}_X = 0.379 - 0.214 = 0.165$$

$$\text{CIszélesség}_Y = 0.267 - 0.116 = 0.151$$



Különbség béták között

- 2. lépés
Kiszámoljuk a két szélesség átlagát:
 $M_{CI} = (0.165 + 0.151) / 2 = 0.158$
- 3. lépés
Kiszámoljuk a két intervallum átfedő részét
Átfedés = $0.267 - 0.214 = 0.053$
- 4. lépés
Kiszámoljuk, hogy az átfedés hány százaléka a két CI átlagos szélességének.
 $\% = \text{Átfedés} / M_{CI} = 0.053 / 0.158 = 0.3354$
- Az átfedés a két konfidencia intervallum átlagos szélességének 33.54%-át teszi ki, tehát a két regressziós együttható $p < .05$ szinten szignifikánsan különbözik egymástól.



Különbség béták között

- Gyakorold a számításokat a reklám és youtube hatásának összehasonlításával.
- Aminek ki kell jönni: 16.84%, tehát $p < .05$ szinten különbözik egymástól a reklám és youtube hatása.

