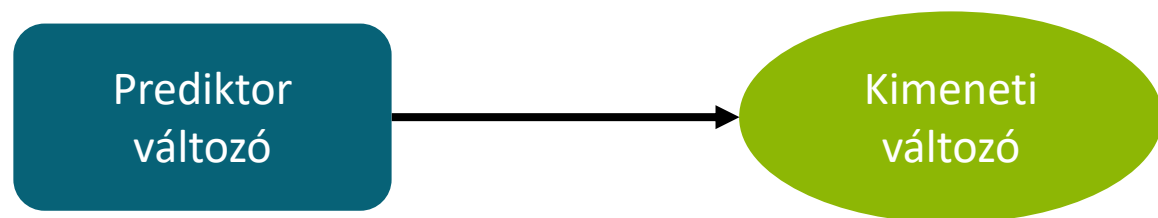


stathelp.hu

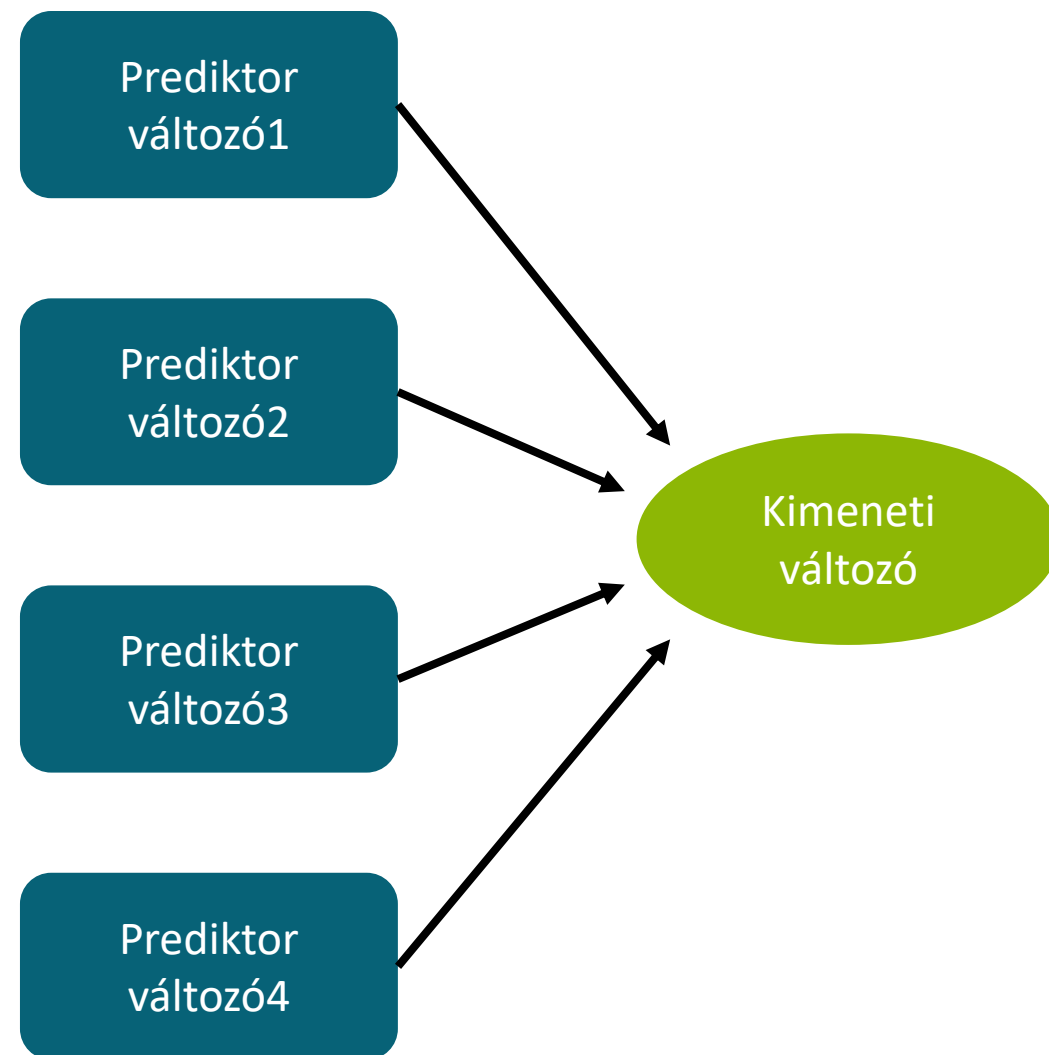
Készítette: Soltész-Várhelyi Klára

Többszörös lineáris regresszió

Egyszerű lineáris regresszió

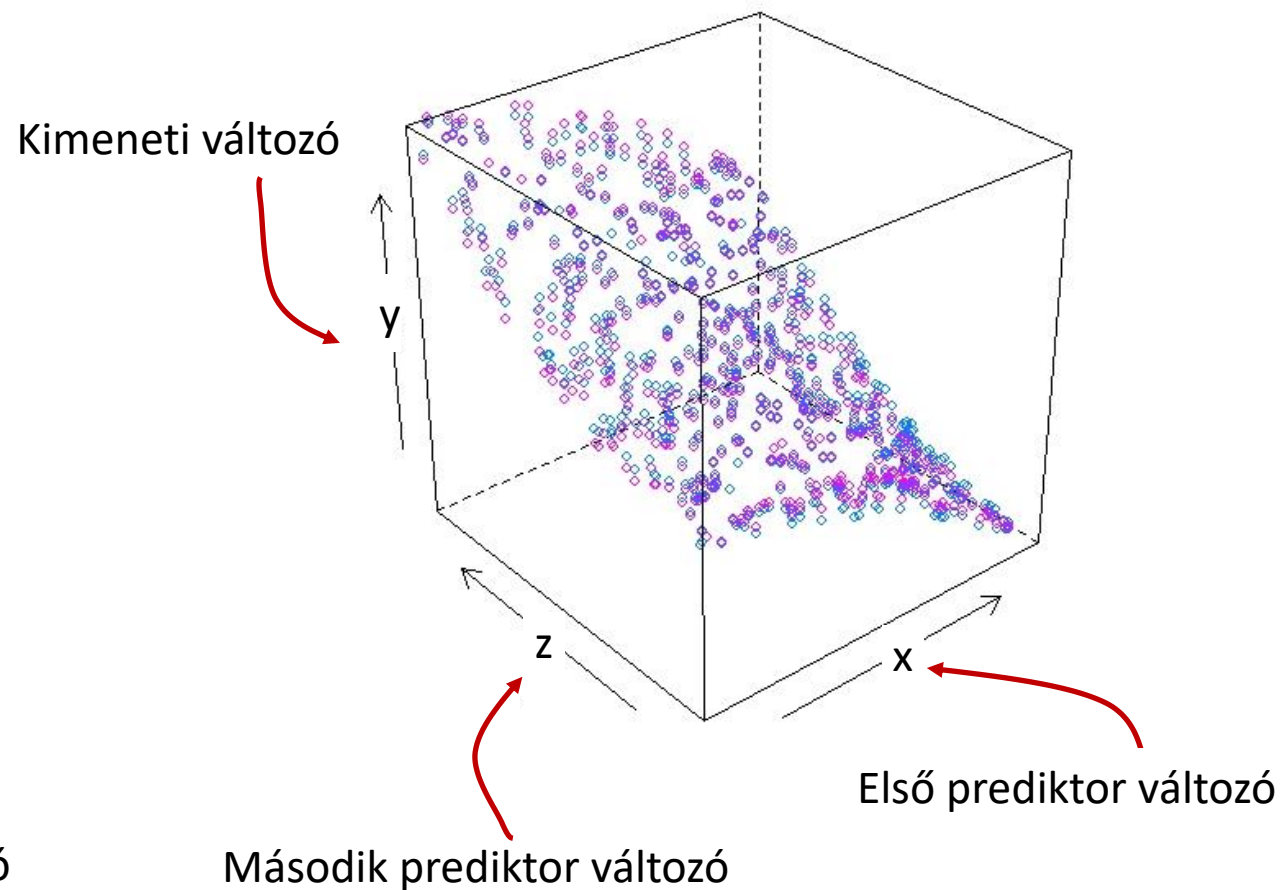
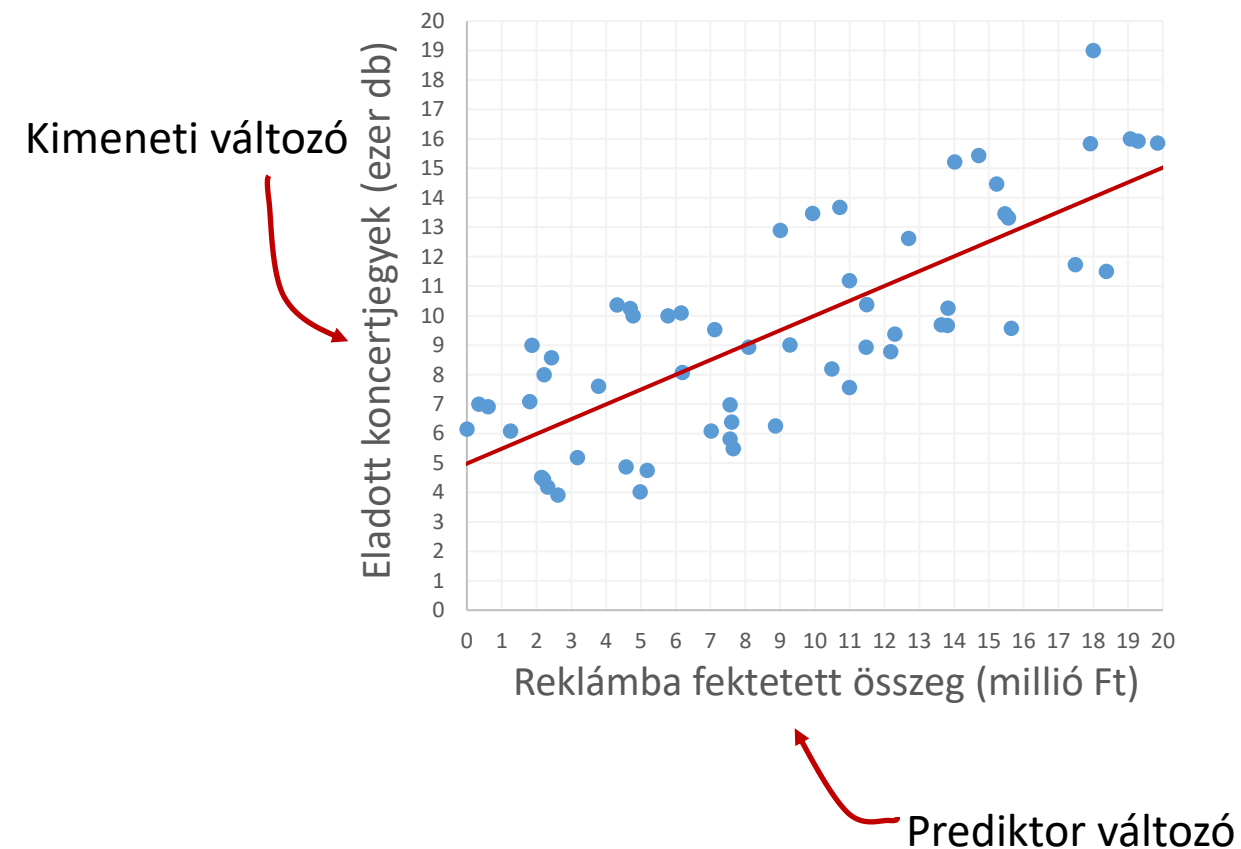


Többszörös lineáris regresszió

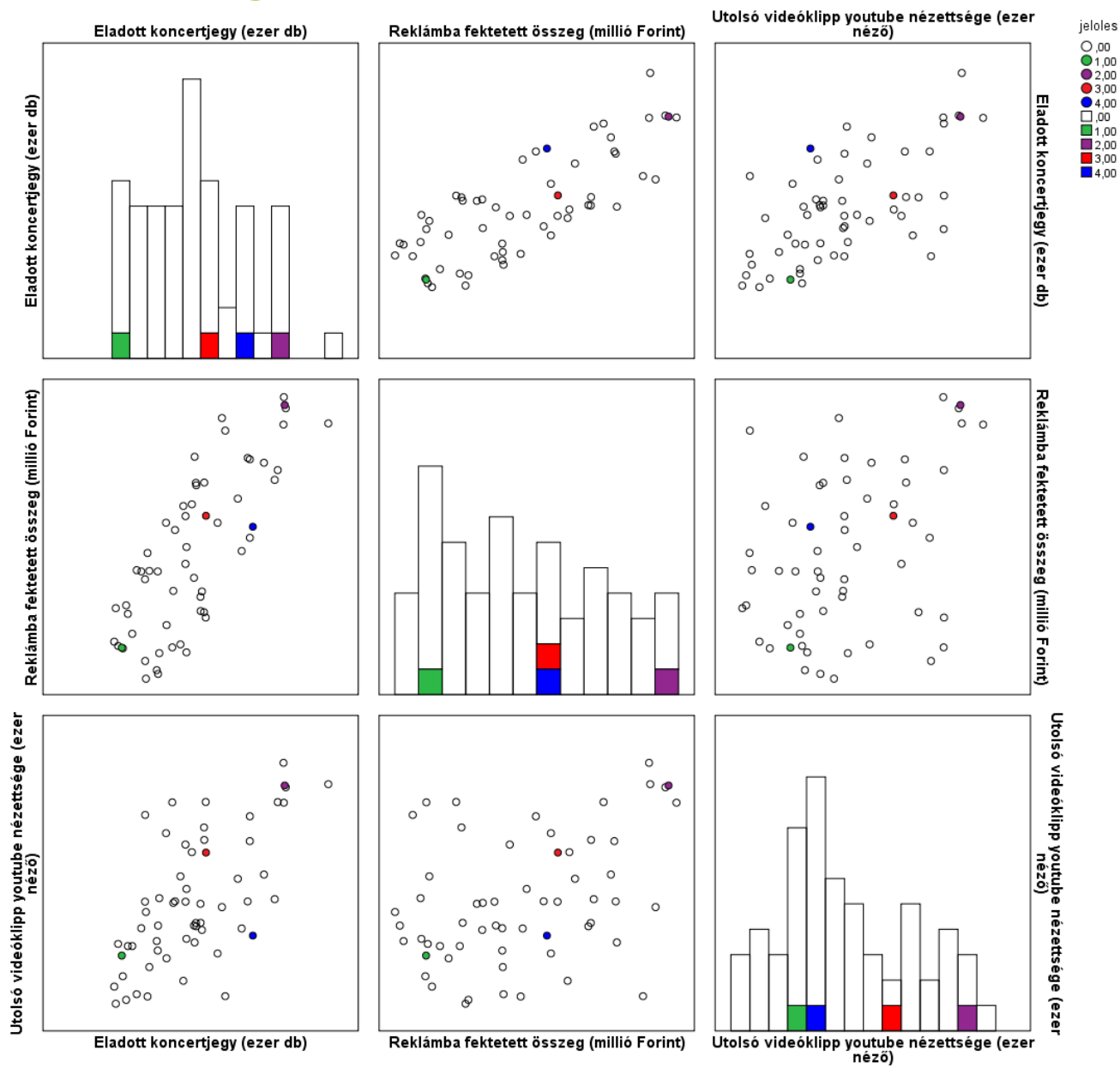
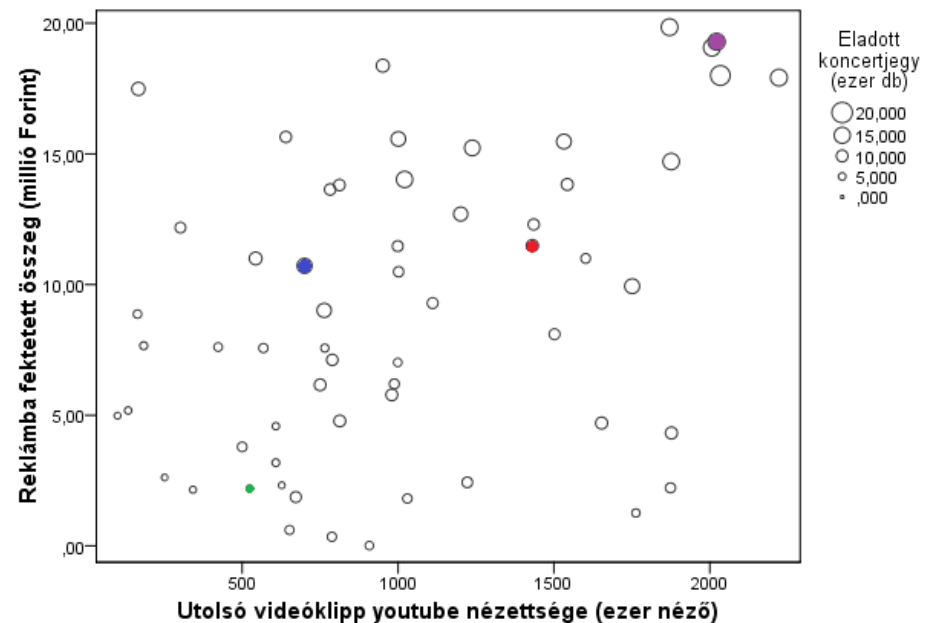
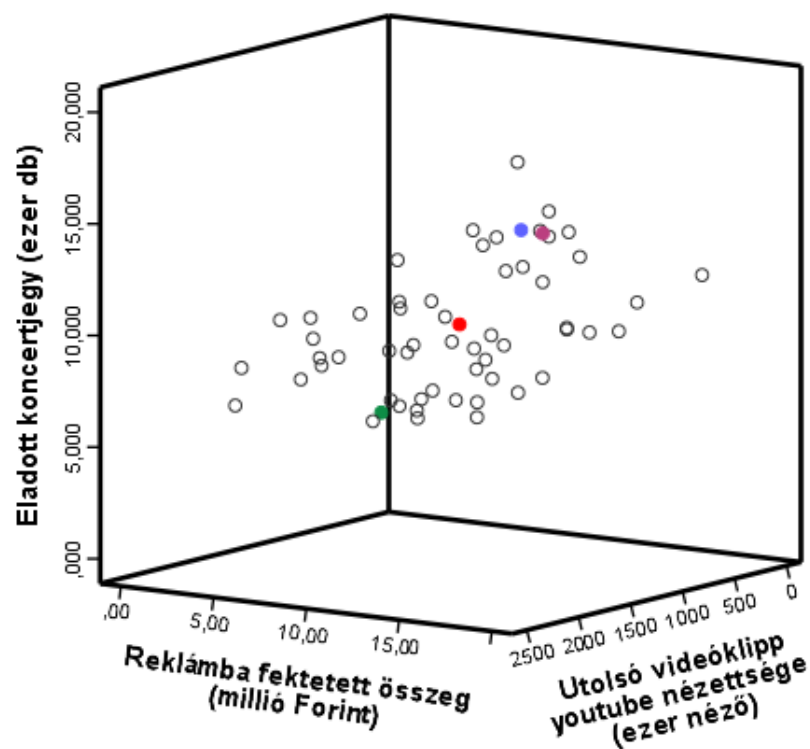


Egyszerű lineáris regresszió

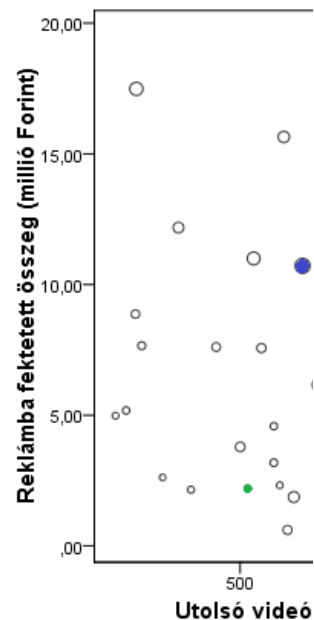
Többszörös lineáris regresszió



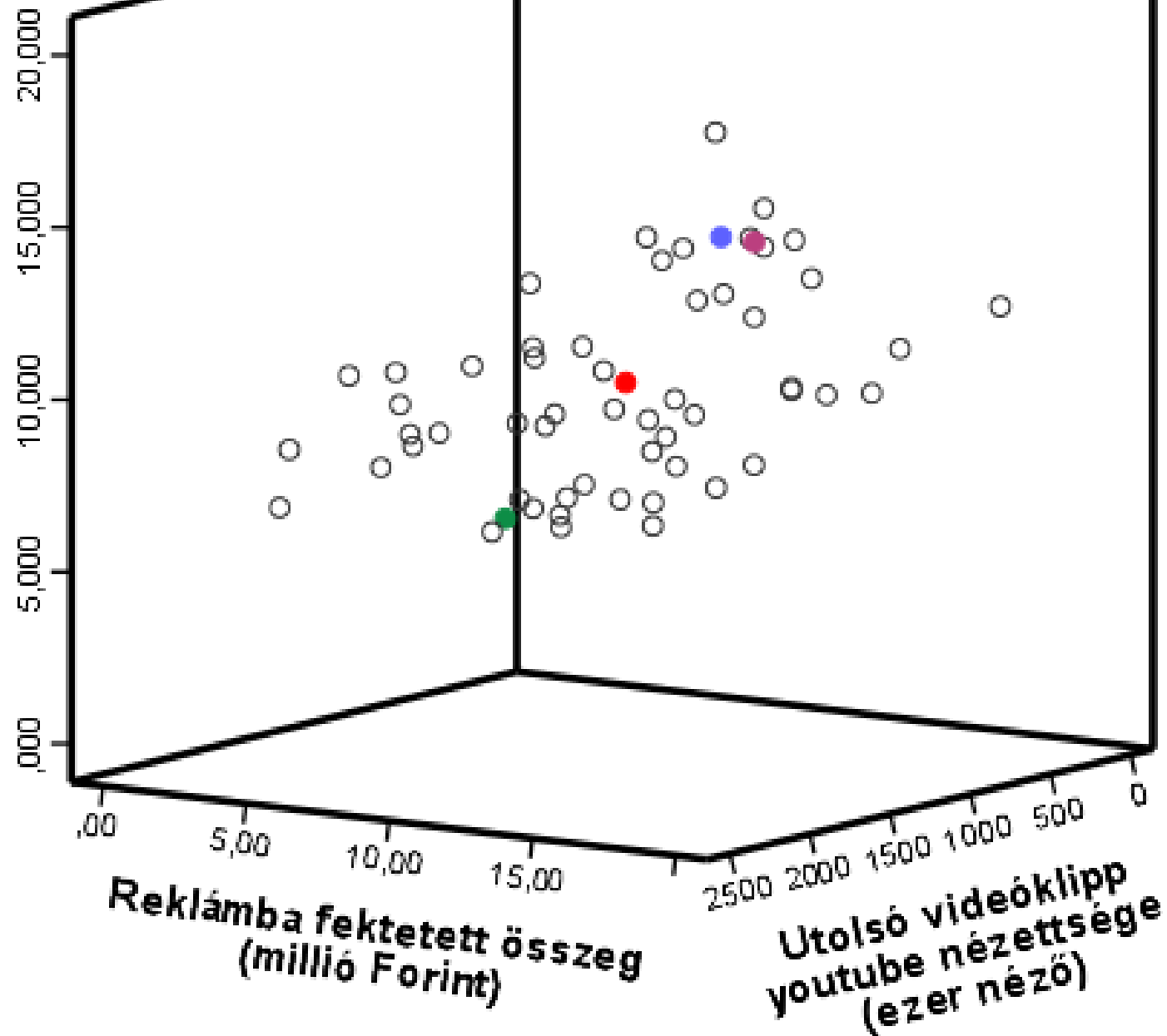
Azaz több prediktor változóval próbáljuk megbecsülni a kimeneti változó értékét



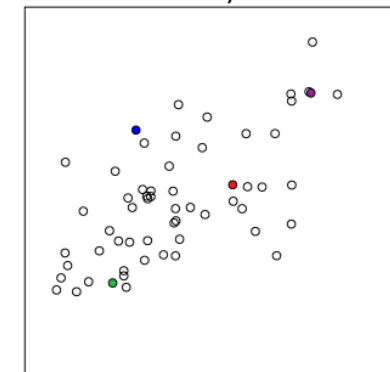
Itozóval próbáljuk
eti változó értékét



Eladott koncertjegy (ezer db)

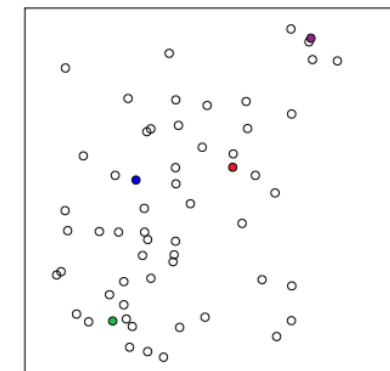


Utolsó videóklipp youtube nézettsége (ezer néző)

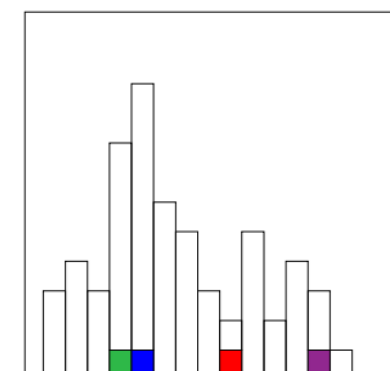


Eladott koncertjegy (ezer db)

jeloles
○ 0,00
● 1,00
● 2,00
● 3,00
● 4,00
□ 0,00
■ 1,00
■ 2,00
■ 3,00
■ 4,00



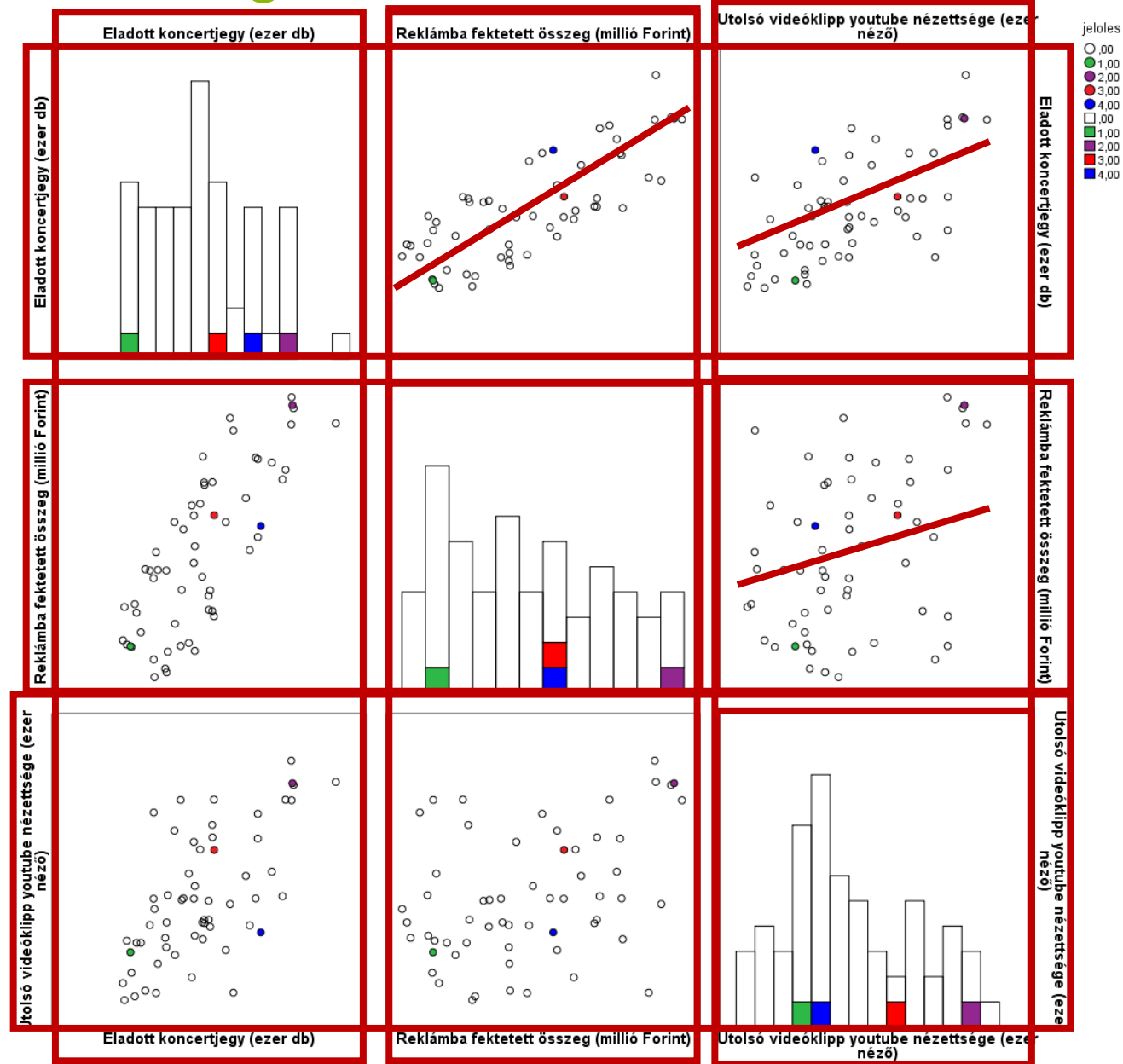
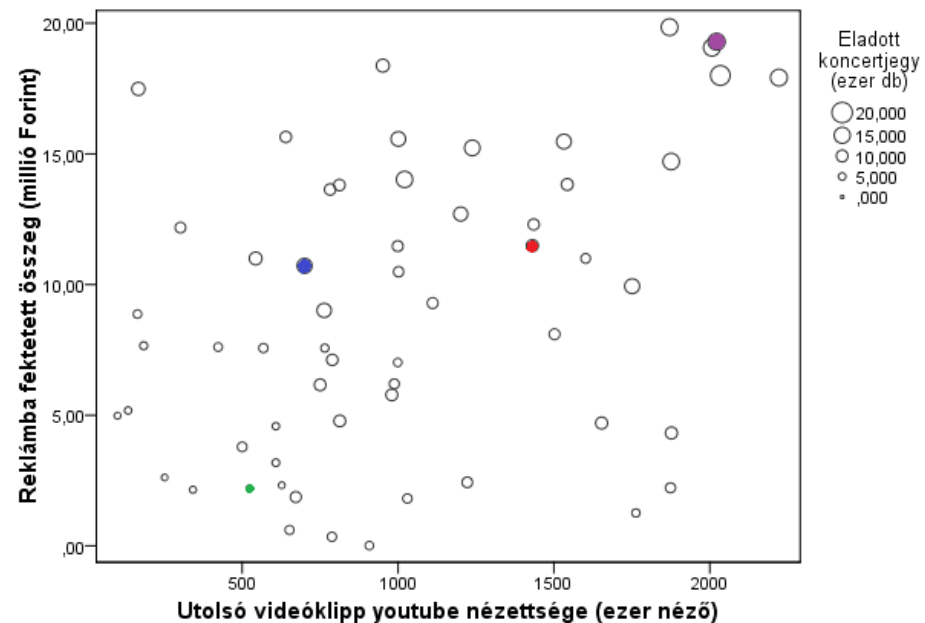
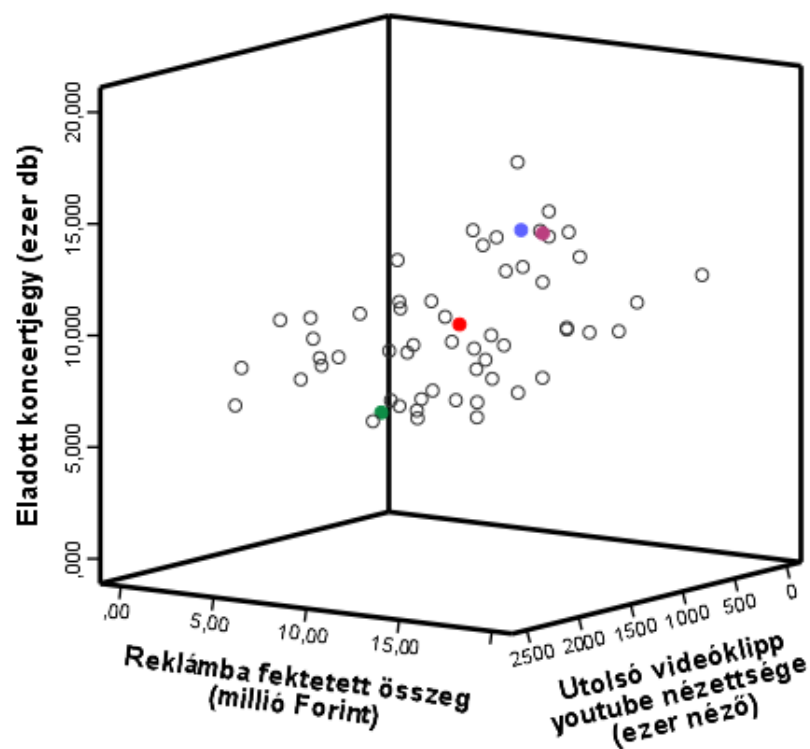
Reklámba fektetett összeg (millió Forint)



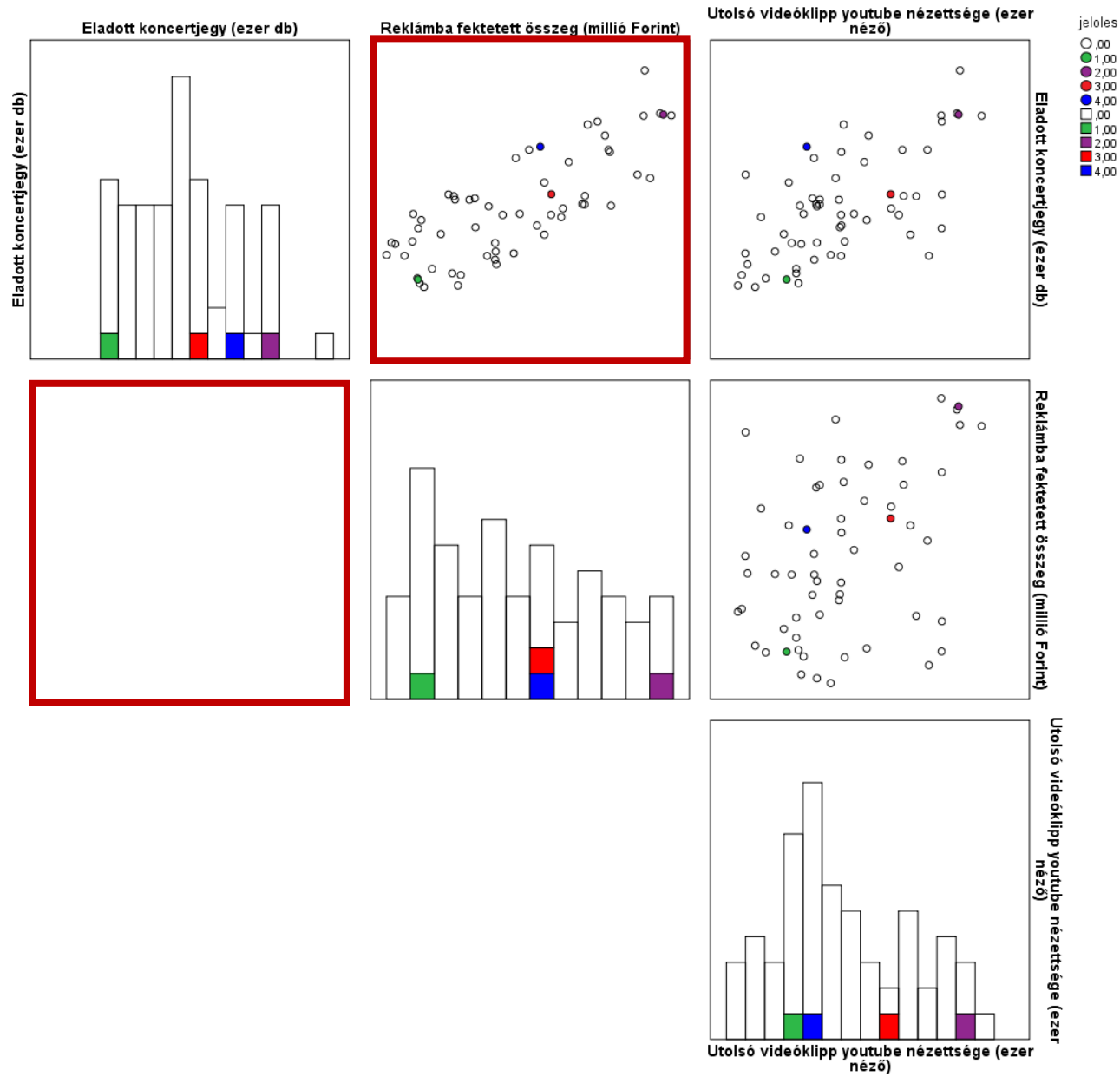
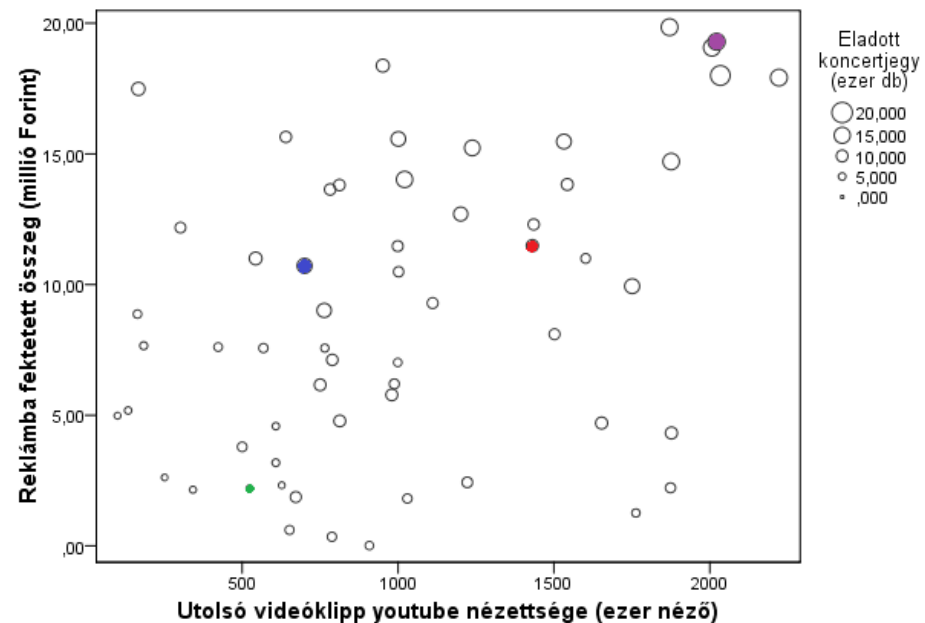
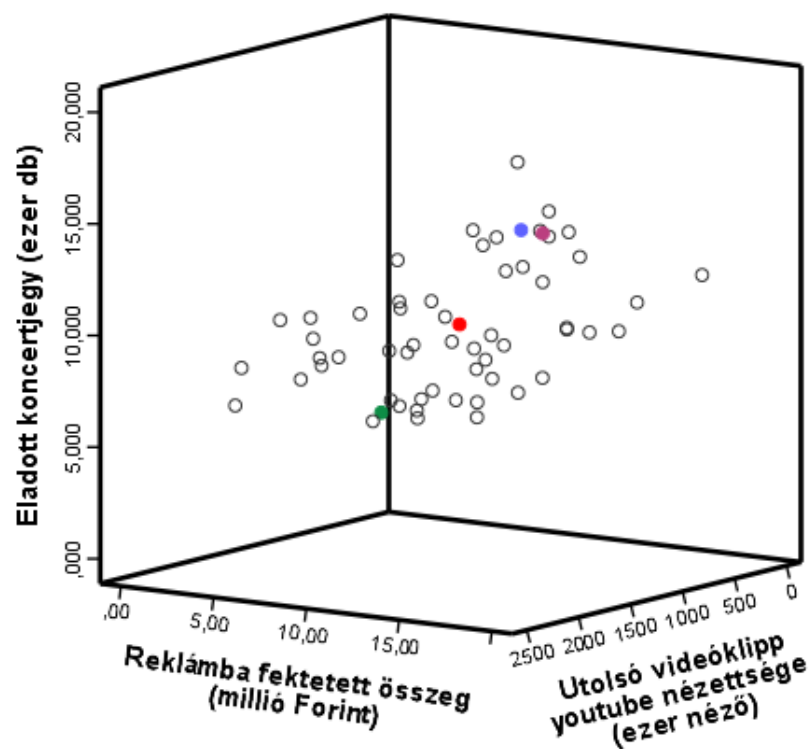
Utolsó videóklipp youtube nézettsége (ezer néző)

Utolsó videóklipp youtube nézettsége (ezer néző)

Azaz több prediktor változóval próbáljuk megbecsülni a kimeneti változó értékét



Azaz több prediktor változóval próbáljuk megbecsülni a kimeneti változó értékét



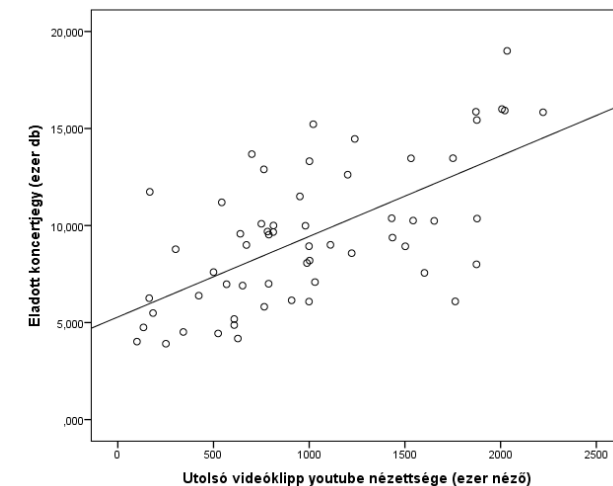
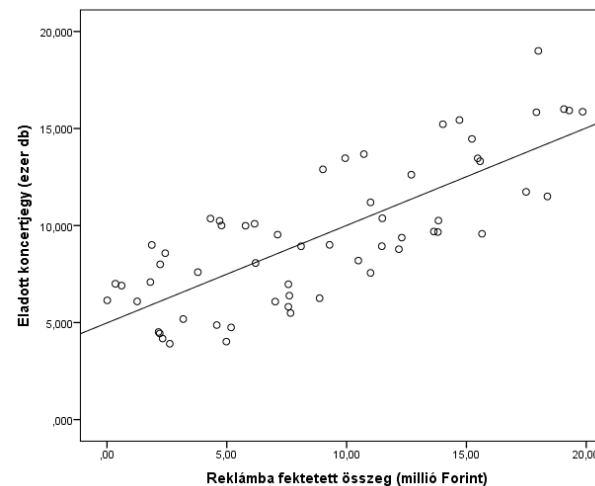
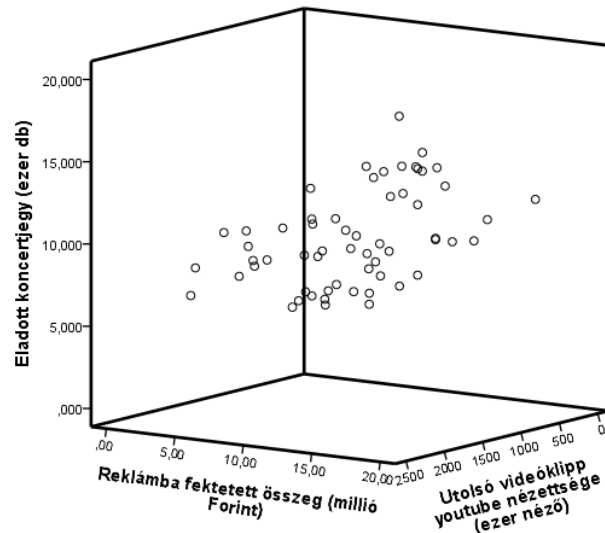
Többszörös regresszió

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.



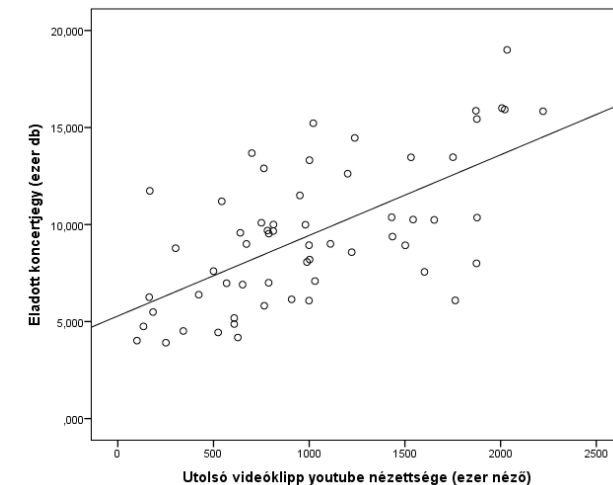
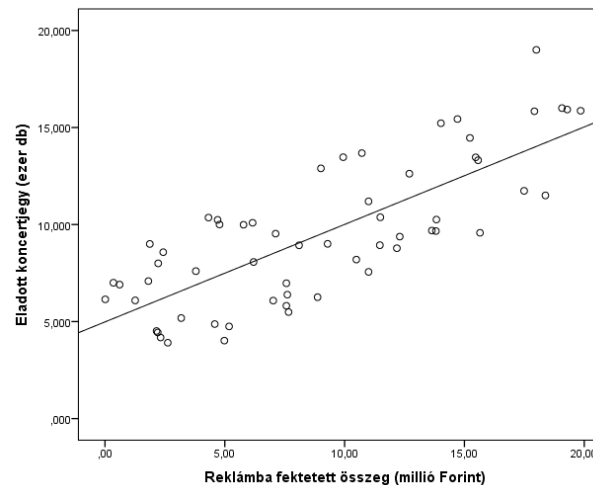
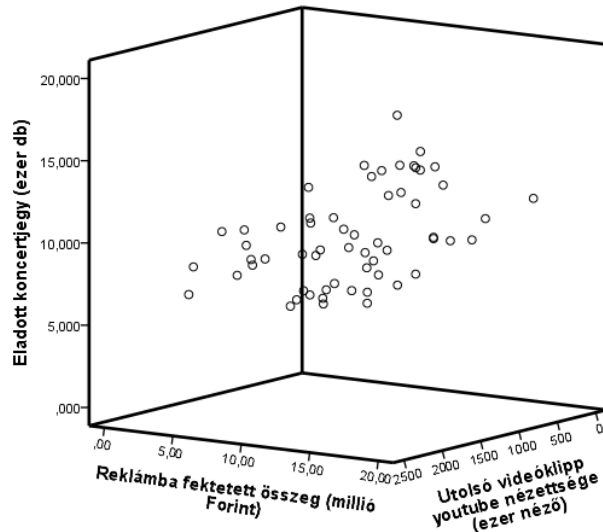
Többszörös regresszió

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

- **F** - A hatás és hiba aránya
 - Ehhez rendeljük a szignifikancia értéket
- **R²** - Effect-size mutató
 - a modellünk mennyit magyaráz a kimeneti változó varianciájából
- **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
- **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

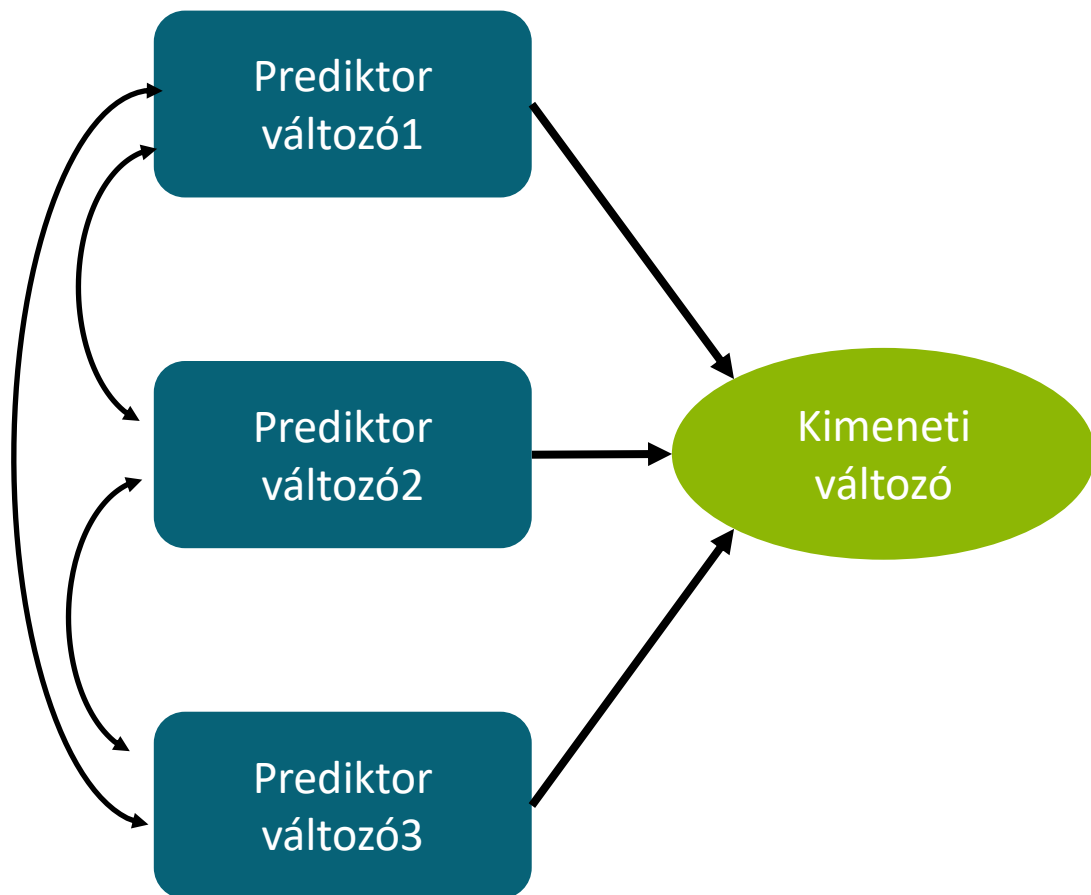
R = .662

Prediktor (X)	Kimeneti (Y)	Predikált (Y _{pred})	Hiba (ε)
4.38	10.22	9.78	0.44
6.27	10.09	9.59	0.50
2.65	9.23	9.95	-0.72
14.32	8.57	8.79	-0.22

R = 1

Prediktor (X)	Kimeneti (Y)	Predikált (Y _{pred})	Hiba (ε)
5.50	11.00	11.00	0
6.12	12.24	12.24	0
6.35	12.70	12.70	0
8.46	19.92	19.92	0

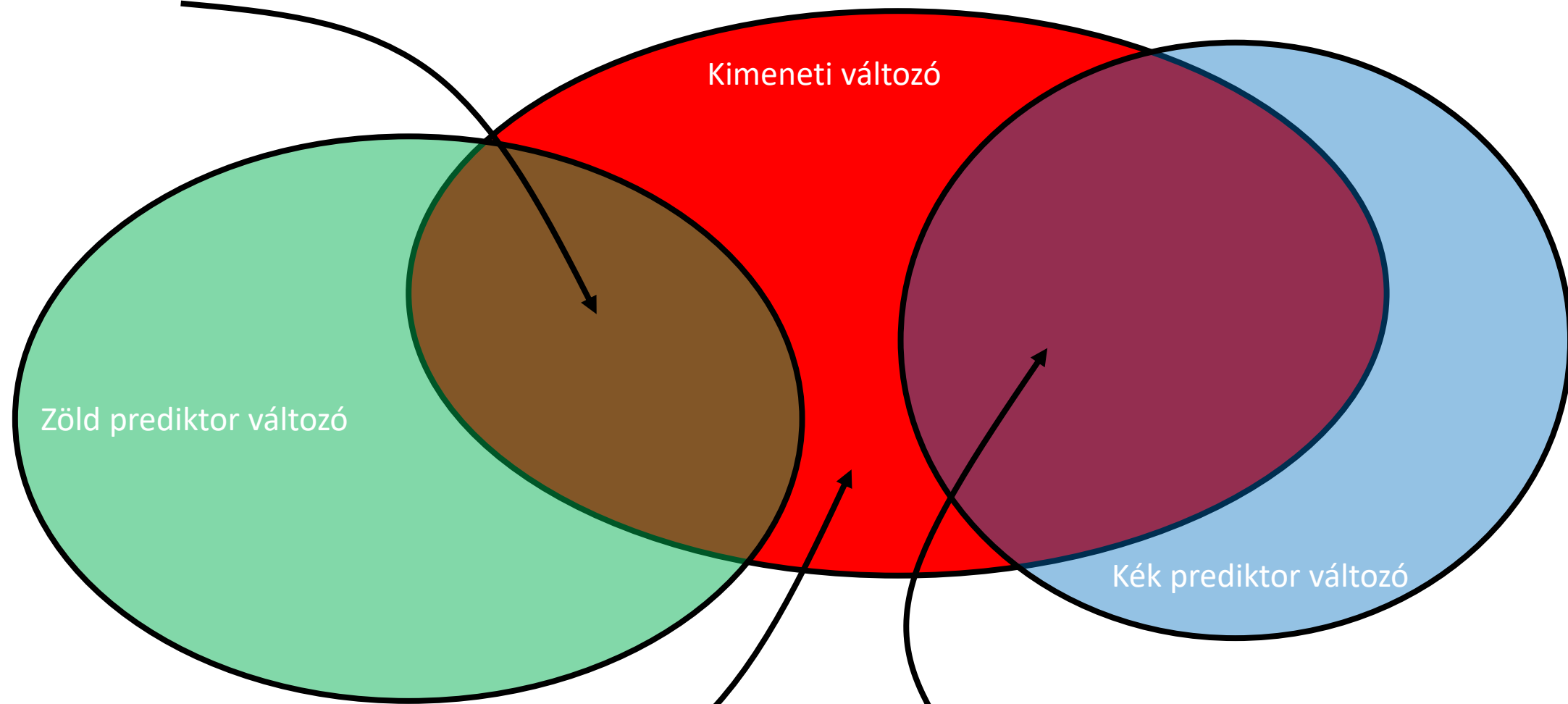
Modell felépítése & a modell szerkezete



A modell kimenetét meghatározza a **prediktor változók**

1. egymáshoz való **viszonya**
2. **kiválasztása és száma**
3. a **modellbe építésüknek sorrendje** (a modell felépítésének módja).

Zöld prediktor segítségével megmagyarázott rész



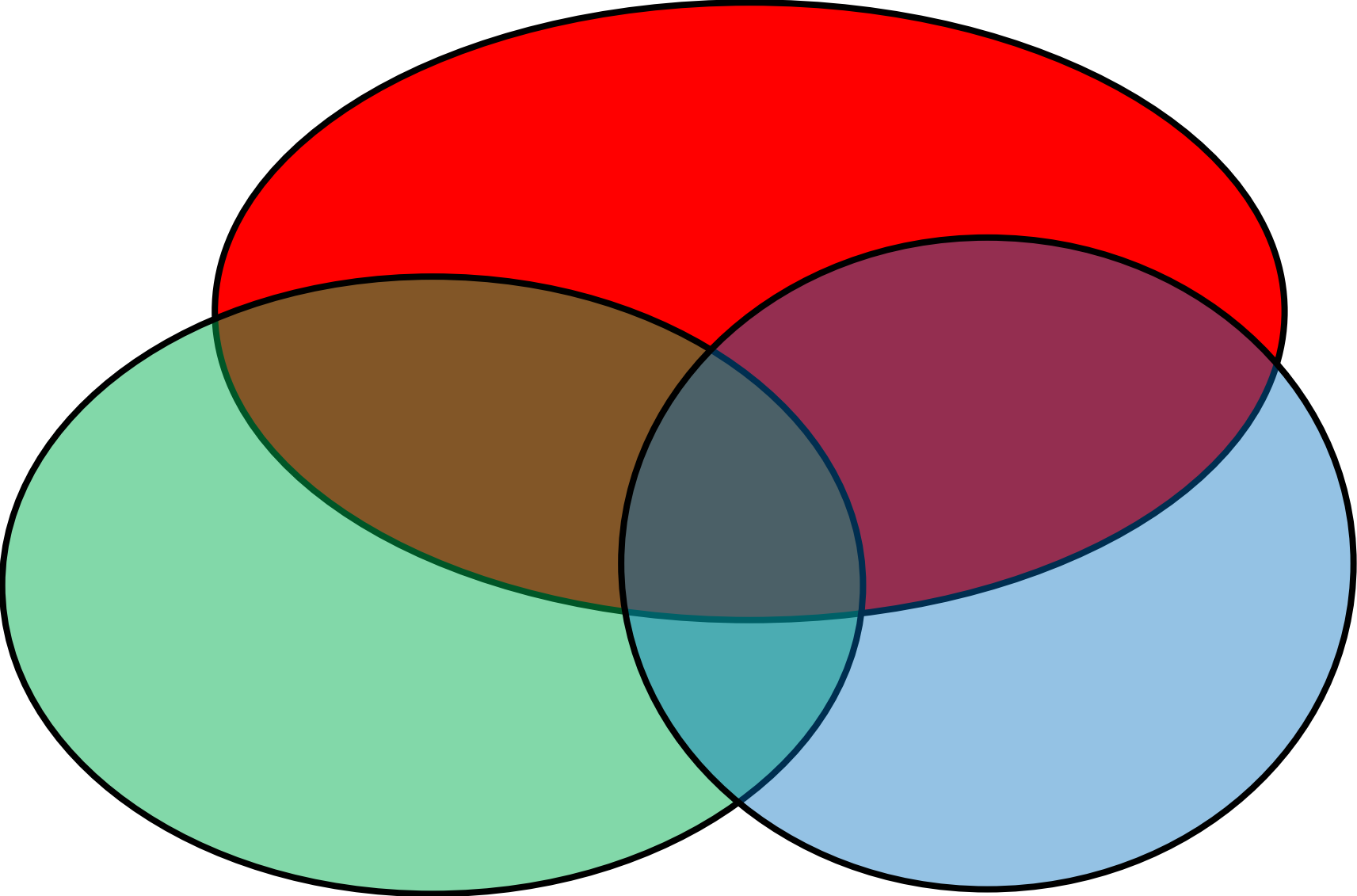
Kimeneti változó

Zöld prediktor változó

Kék prediktor változó

A kimeneti változó változatosságának
azon része, amit egyik prediktor sem magyaráz

Kék prediktor segítségével megmagyarázott rész



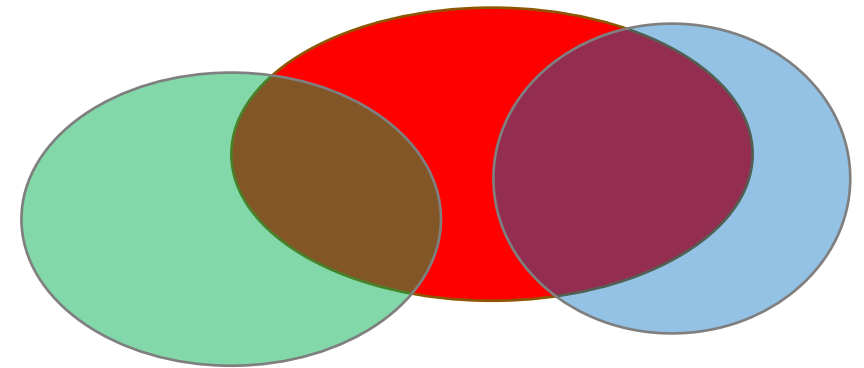
Prediktorok viszonya

A prediktorok egymáshoz való **viszonya**

$$Y_{\text{pred},i} = b_0 + b_1 * X_{1,i} + b_2 * X_{2,i} + \dots + b_k * X_{k,i}$$

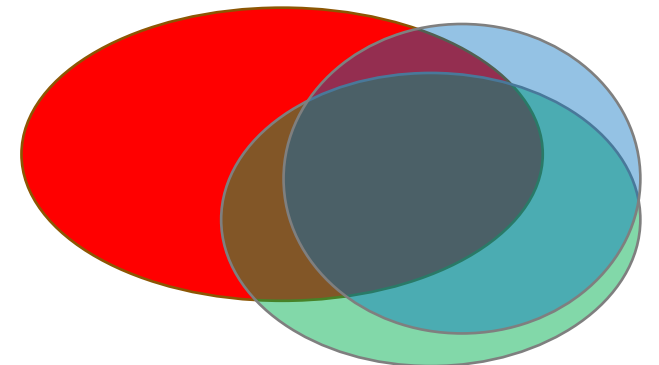
Ha a **változókim korrelálatlanok**,

- akkor minden prediktor más részt magyaráz a kimeneti változó variáciájából,
- tehát predikált érték a **prediktorok hatásának lineáris kombinációjaként** fog összeállni,
- azaz az egyéni hatások egyszerűen összeadódnak



Korreláló prediktor változóknál

- a prediktorok **részben ugyanazt a részt magyarázzák** a kimeneti változóból,
- ezért „**versengeni fognak**” a kimeneti változó variáciájáért



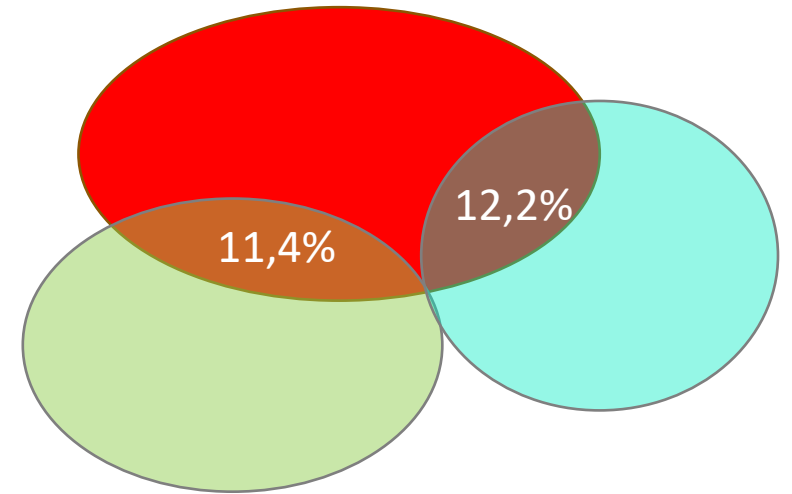
Prediktorok viszonya

Példa **korrelálatlan prediktor változók** viselkedésére

- **Prediktor1** és **Prediktor2** : $r = .000338$ $p = .998$
- **Kimeneti** és **Prediktor1**: $r = -.337$ $p = .017$
- **Kimeneti** és **Prediktor2**: $r = .350$ $p = .013$

Regressziós modellek:

- A **magyarázó erők összeadódnak** $R^2_{\text{pred1}} + R^2_{\text{pred2}} = R^2_{\text{közös}}$
- A **meredekségek ugyanazok maradnak**



Model Summary					
R Square					
,114					
ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4363,257	1	4363,257	6,152	,017 ^b
Residual	34044,123	48	709,253		
Total	38407,380	49			
Coefficients ^a					
Model	Unstandardized Coefficients	Standardized Coefficients			
	B	Beta	t	Sig.	
(Constant)	60,138		7,933	,000	
Prediktor 1	-,339	-,337	-2,480	,017	

Model Summary					
R Square					
,122					
ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4703,528	1	4703,528	6,699	,013 ^b
Residual	33703,85	48	702,164		
Total	38407,38	49			
Coefficients ^a					
Model	Unstandardized Coefficients	Standardized Coefficients			
	B	Beta	t	Sig.	
(Constant)	24,376		2,903	,006	
Prediktor 2	,349	,350	2,588	,013	

Model Summary					
R Square					
,236					
ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	9069,852	2	4534,926	7,265	,002 ^b
Residual	29337,528	47	624,203		
Total	38407,380	49			
Coefficients ^a					
Model	Unstandardized Coefficients	Standardized Coefficients			
	B	Beta	t	Sig.	
(Constant)	40,694		4,055	,000	
Prediktor 1	-,339	-,337	-2,645	,011	
Prediktor 2	,349	,350	2,746	,009	

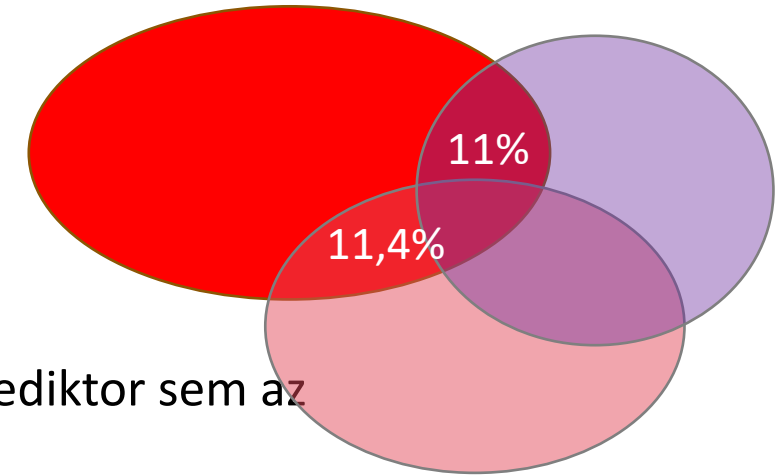
Prediktorok viszonya

Példa **korreláló prediktor változók** viselkedésére

- **Prediktor3** és **Prediktor4** : $r = .258$ $p = .070$
- **Kimeneti** és **Prediktor3**: $r = .331$ $p = .019$
- **Kimeneti** és **Prediktor4**: $r = .338$ $p = .016$

Regressziós modellek:

- A **magyarázó erők nem adódnak össze**: $R^2_{\text{pred1}} + R^2_{\text{pred2}} \neq R^2_{\text{közös}}$
- A **meredekségek mérséklődnek**, sőt bár a modell szignifikáns, egyik prediktor sem az



Model Summary					
R Square					
,110					
ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4208,463	1	4208,463	5,907	,019 ^b
Residual	34198,917	48	712,477		
Total	38407,380	49			
Coefficients ^a					
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	
	B	Beta			
(Constant)	24,546		2,795	,007	
Prediktor 3	,333	,331	2,430	,019	

Model Summary					
R Square					
,114					
ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4387,942	1	4387,942	6,191	,016 ^b
Residual	34019,438	48	708,738		
Total	38407,380	49			
Coefficients ^a					
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	
	B	Beta			
(Constant)	24,652		2,875	,006	
Prediktor 4	,386	,338	2,488	,016	

Model Summary					
R Square					
,178					
ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	6832,121	2	3416,06	5,085	,010 ^b
Residual	31575,259	47	671,814		
Total	38407,380	49			
Coefficients ^a					
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	
	B	Beta			
(Constant)	13,273		1,294	,202	
Prediktor 3	,263	,261	1,907	,063	
Prediktor 4	,309	,271	1,976	,054	

Prediktorok száma

Helytelen **rengetek véletlenszerűen kiválasztott változót** a modellbe dobni

Regressziós egyenes szempontjából

- Nő a becsült paraméterek (b_1 értékek) száma, és az esetleges átfedések mértéke a prediktorok között (nő „**verseny**”, amivel a prediktorok szembesülnek)

Statisztikai szempontból

- Nő a modell bonyolultsága.
- Nő a modell **specificitása**, és így csökken a **generalizálhatóság**.
- A túlspecifikált modell lehet nagyon pontos a mintán, de nem feltétlenül az a populációban is (nem generalizálható),
- mert a modell „rátanul” a mintában lévő zajra is.

Gyakorlati szempontból

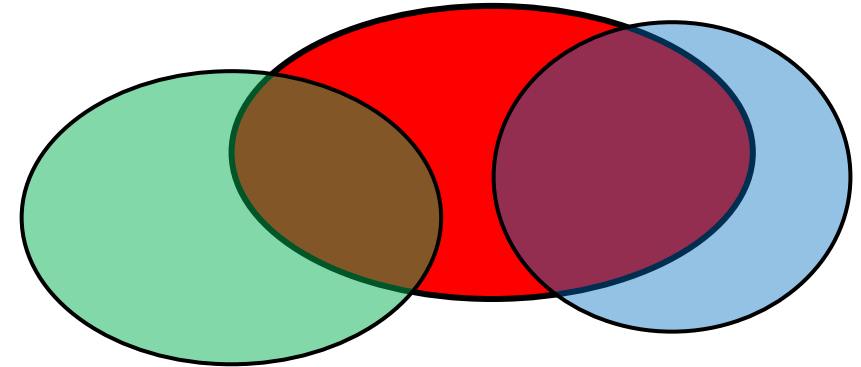
- A változók számának növelésével nő a szükséges elemszám.

Prediktorok sorrendje

A prediktorok modellbe való építésének **sorrendje**

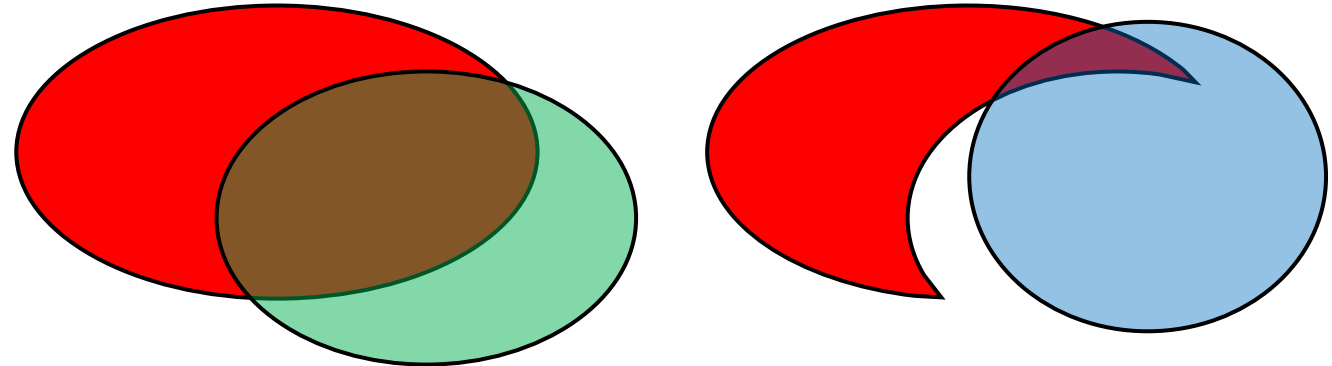
Ha a **változóim korrelálatlanok**,

- akkor **nem számít** a modellbe építés **sorrendje**,
- mert minden prediktor más részt magyaráz a kimeneti változó variáciájából,
- tehát a hatások összeadódnak,
- az összeadás elemei pedig felcserélhetőek



Korreláló prediktor változóknál

- **fontos** a modell kialakításának módja
- mert a prediktorok **részben ugyanazt a részt magyarázzák** a kimeneti változóból,
- ezért „**versengeni fognak**” a kimeneti változó variáciájáért

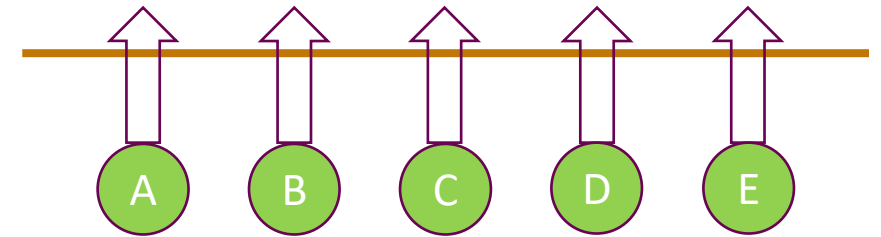


A modell felépítésének módja(i):

- Enter (forced entry)
- Hierarchikus (blockwise)
- Stepwise módszercsalád

Enter (forced entry)

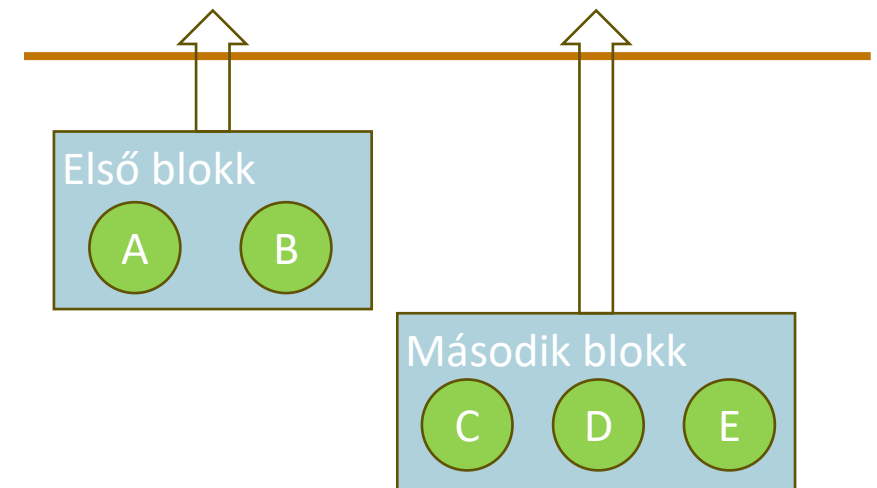
- A modellbe egyszerre lépnek be a változók
- Minden változó a modellbe kerül
- Egy prediktor hatását csak az összes többi prediktor kontextusában tudjuk értelmezni



Prediktorok sorrendje

Hierarchikus (blokkonként – blockwise model)

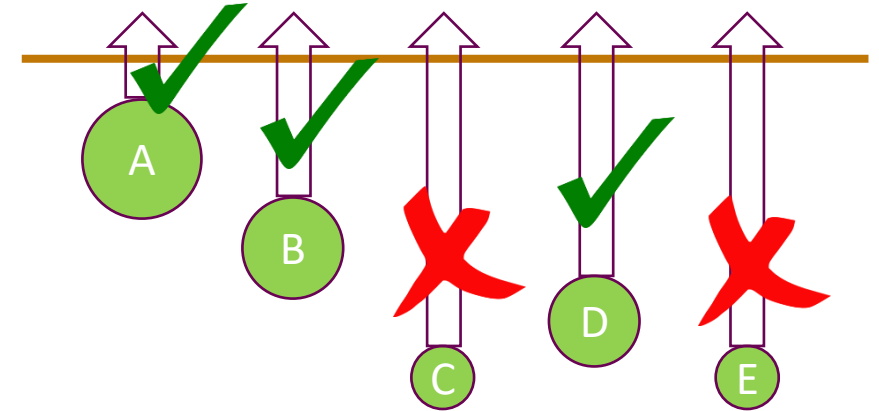
- A változók sorrendjét a kutató adja meg (előzetes tudása, sejtése alapján)
- Először azok a változók kerülnek a modellbe,
 - melyek a hipotézisek szempontjából fontosabban
 - melynek már ismerjük a hatását
 - melynek várhatóan legnagyobb a hatása
 - vagy amelyekre kontrollálni akarunk
- Egy blokkba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



Prediktorok sorrendje

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



Blockwise, enter vagy stepwise?

Ha van tudományos elképzelésed, használd blockwise (hierarchikus) módszert

- A blockwise módszer lehetővé teszi, hogy az előzetes tudásod is számításba vedd, a számodra fontosabb változókat kiemeld.

Ha feltáró vizsgálatot végzel, használd enter módszert

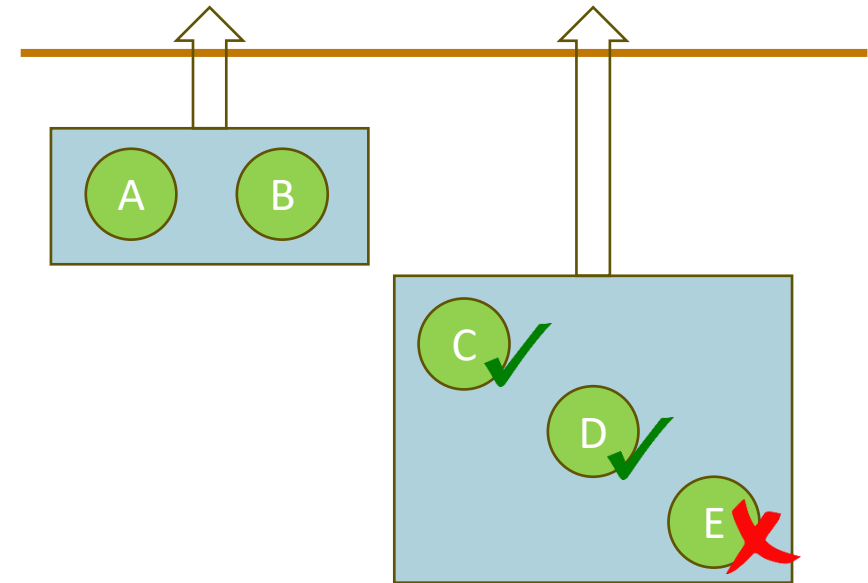
- A változó modellbe való kerülését nem befolyásolják az előzetes elképzeléseid, tudásod.
- Pontosabb képet kapsz a prediktor változók és a kimeneti változó kapcsolatáról, plusz a prediktor változók egymással való kapcsolatáról is.

Ha „csak” a leggazdaságosabb modellt keresed, használd stepwise módszert

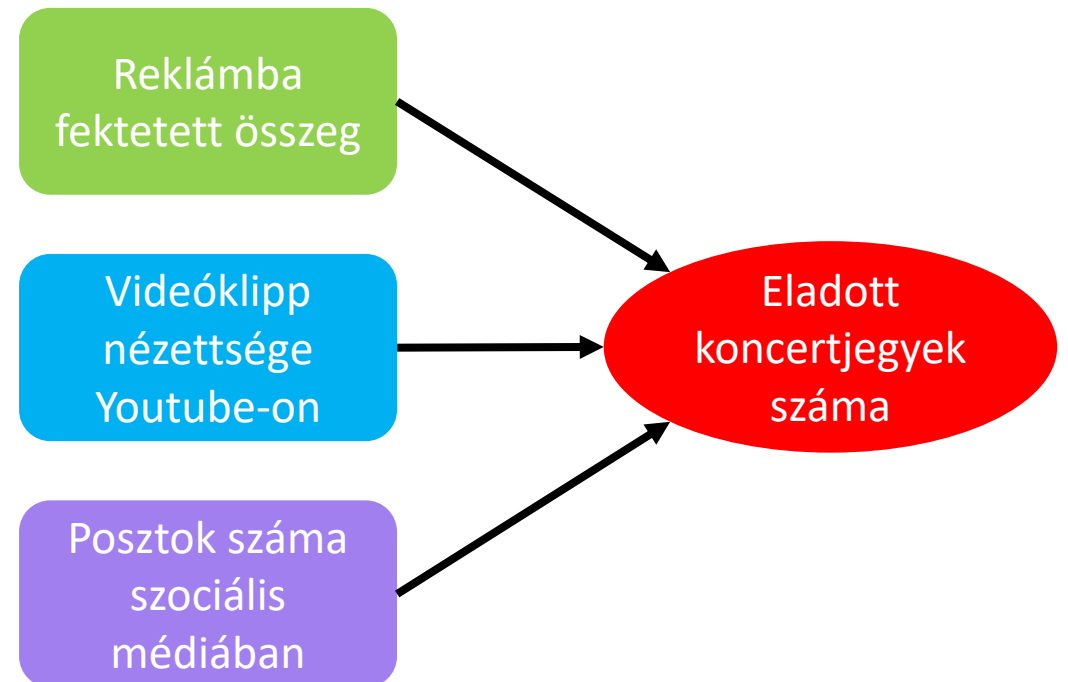
- Megkeresi azt a modellt, ahol a lehető legkevesebb változóval a lehető legpontosabb jóslatot lehet tenni. Például, ha célod, hogy lehetőleg kevés prediktor változóval kapj pontos becslést (mert a prediktor változók felvétele pénzbe kerül).
- Hátránya, hogy a változókról önmagunkban kevés információt 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.

Összetett modellek

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



Többszörös lineáris regresszió a gyakorlatban



Példa modell felépítésére

Jósoljuk be az eladott **koncertjegyek** számát (kimeneti változó) abból, hogy

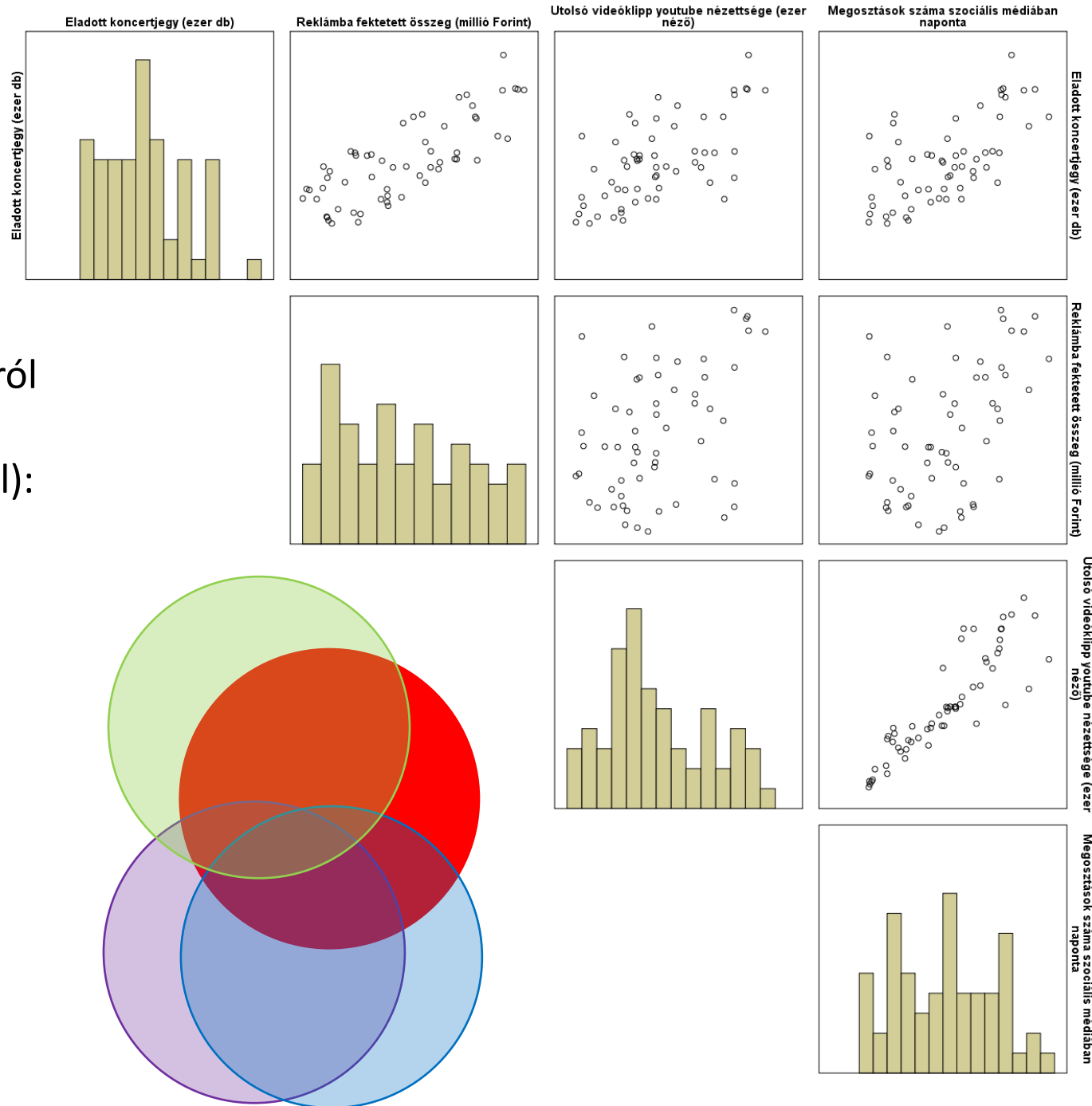
- mennyit fektetnek a koncert **reklámozás**ába
- mekkora a zenekar nézettsége **Youtube**-on
- naponta átlagosan hány **poszt** születik a zenekarról

Korrelációk a **koncertjegyek**kel (kimeneti változóval):

- **Reklám**: $r = .780$ $p < .001$ (~60%)
- **Youtube**: $r = .643$ $p < .001$ (~41%)
- **Poszt**: $r = .639$ $p < .001$ (~41%)

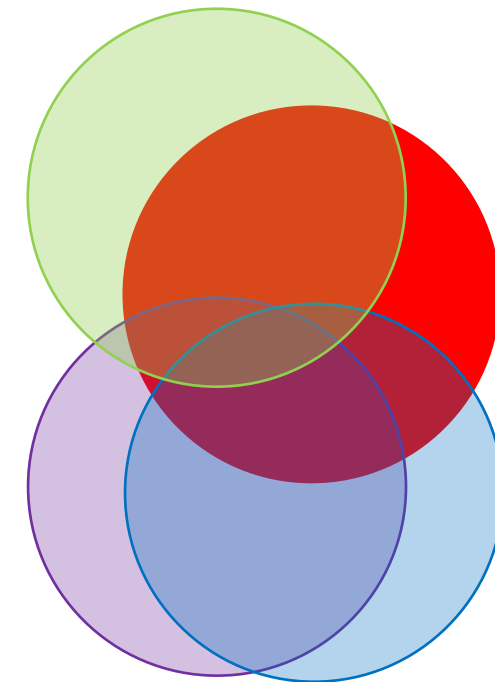
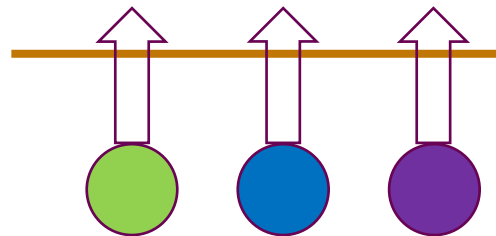
Korreláció a prediktorok között:

- **Reklám** x **Youtube**: $r = .393$ $p = .002$ (~15%)
- **Reklám** x **Poszt**: $r = .435$ $p = .001$ (~18%)
- **Poszt** x **Youtube**: $r = .884$ $p < .001$ (~78%)



Példa modell felépítésére

- Mi történik ha **ENTER** módszert választunk?
- Mindhárom változó egyszerre kerül a modellbe.
 - A **reklám** hatását viszonylag pontosan látni fogjuk, hiszen viszonylag független a többi prediktortól
 - A **youtube** és **poszt** hatását nem tudjuk igazán felmérni, mert túl nagy az átfedés közöttük, az általuk megmagyarázott varianciarész túl nagy részén kell egymással osztozniuk
 - A **poszt**nak, ami 78%-ban osztozik a **youtube**-bal és 18%-ban a **reklámmal**, szinte semmilyen önálló variancia rész nem jut, így nem is lesz szignifikáns – bár önállóan 41%-ot magyarázott volna



Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,091	,671		4,607	,000
	Reklámba fektetett összeg (millió Forint)	,396	,049	,615	8,102	,000
	Utolsó videóklipp youtube nézettsége (ezer néző)	,002	,001	,329	2,248	,029
	Megosztások száma szociális médiában naponta	,003	,006	,082	,546	,587

a. Dependent Variable: Eladott koncertjegy (ezer db)

Pearson's Correlations

Variable		Koncertjegyek	Reklam	Youtube	Megosztas
1. Koncertjegyek	Pearson's r	—			
	p-value	—			
2. Reklam	Pearson's r	0.780	—		
	p-value	< .001	—		
3. Youtube	Pearson's r	0.643	0.393	—	
	p-value	< .001	0.002	—	
4. Megosztas	Pearson's r	0.639	0.435	0.884	—
	p-value	< .001	< .001	< .001	—

Majd megtanuljuk, hogy itt a kollinearitás hiányának feltétele sérül egyébként

Linear Regression

Model Summary - Koncertjegyek

Model	R	R ²	Adjusted R ²	RMSE
H ₀	0.000	0.000	0.000	3.657
H ₁	0.862	0.743	0.729	1.904

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₁	Regression	576.185	3	192.062	53.000	< .001
	Residual	199.311	55	3.624		
	Total	775.496	58			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	9.483	0.476		19.921	< .001
H ₁	(Intercept)	3.091	0.671		4.607	< .001
	Reklam	0.396	0.049	0.615	8.102	< .001
	Youtube	0.002	9.447e-4	0.329	2.248	0.029
	Megosztas	0.003	0.006	0.082	0.546	0.587

Példa modell felépítésére

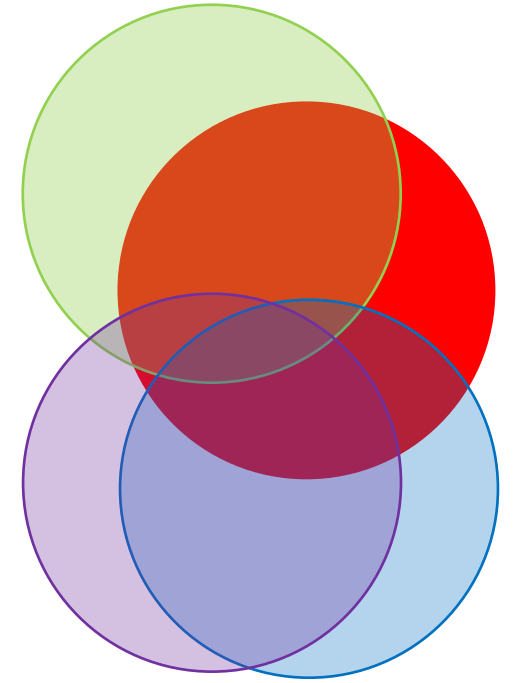
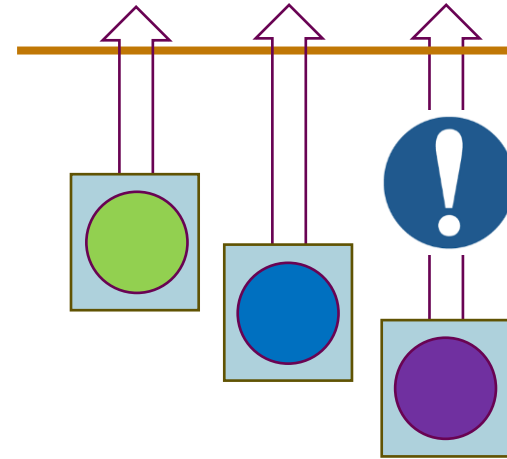
Mi történik ha **BLOCKWISE** módszert választunk?

- Itt te döntheted el, melyik változók fontosak számodra.
- Az, ahogy a modellt felépíted, egyfajta történetet mesél el. A hierarchikus elrendezésben te határozod meg, hogy melyik modell mondja el a leginformatívabb történetet.

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)	3,091	,671		4,607	,000
	Reklámba fektetett összeg (millió Forint)	,396	,049	,615	8,102	,000
	Utolsó videóklipp youtube nézettsége (ezer néző)	,002	,001	,329	2,248	,029
	Megosztások száma szociális médiában naponta	,003	,006	,082	,546	,587

a. Dependent Variable: Eladott koncertjegy (ezer db)



- Meg fogod kapni mind a három modellt, és az R^2 változásából látni fogod, hogy egy modellhez képest a következő jelent-e szignifikáns növekedést a magyarázóerőben.
- Eldöntheted, hogy helyik modellt fogod értelmezni.
- Így kaptad a legteljesebb képet, és kontrolláltad a változók sorrendjét is

Linear Regression ▾

Model Summary - Koncertjegyek ▾

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
H ₀	0.780	0.608	0.601	2.309	0.608	88.447	1	57	< .001
H ₁	0.861	0.742	0.732	1.892	0.133	28.929	1	56	< .001

Note. Null model includes Reklam

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₀	Regression	471.583	1	471.583	88.447	< .001
	Residual	303.913	57	5.332		
	Total	775.496	58			
H ₁	Regression	575.104	2	287.552	80.357	< .001
	Residual	200.392	56	3.578		
	Total	775.496	58			

Note. Null model includes Reklam

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4.983	0.565		8.817	< .001
	Reklam	0.502	0.053	0.780	9.405	< .001
H ₁	(Intercept)	3.290	0.560		5.876	< .001
	Reklam	0.402	0.048	0.623	8.439	< .001
	Youtube	0.003	4.775e −4	0.397	5.379	< .001

Linear Regression

Model Summary - Koncertjegyek

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
H ₀	0.780	0.608	0.601	2.309	0.608	88.447	1	57	< .001
H ₁	0.848	0.719	0.709	1.971	0.111	22.202	1	56	< .001

Note. Null model includes Reklam

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₀	Regression	471.583	1	471.583	88.447	< .001
	Residual	303.913	57	5.332		
	Total	775.496	58			
H ₁	Regression	557.865	2	278.932	71.774	< .001
	Residual	217.631	56	3.886		
	Total	775.496	58			

Note. Null model includes Reklam

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4.983	0.565		8.817	< .001
	Reklam	0.502	0.053	0.780	9.405	< .001
H ₁	(Intercept)	2.748	0.677		4.062	< .001
	Reklam	0.399	0.051	0.619	7.872	< .001
	Megosztas	0.016	0.003	0.370	4.712	< .001

Linear Regression

Model Summary - Koncertjegyek

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
H ₀	0.780	0.608	0.601	2.309	0.608	88.447	1	57	< .001
H ₁	0.848	0.719	0.709	1.971	0.111	22.202	1	56	< .001

Note. Null model includes Reklam

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₀	Regression	471.583	1	471.583	88.447	< .001
	Residual	303.913	57	5.332		
	Total	775.496	58			
H ₁	Regression	557.865	2	278.932	71.774	< .001
	Residual	217.631	56	3.886		
	Total	775.496	58			

Note. Null model includes Reklam

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4.983	0.565		8.817	< .001
	Reklam	0.502	0.053	0.780	9.405	< .001
H ₁	(Intercept)	2.748	0.677		4.062	< .001
	Reklam	0.399	0.051	0.619	7.872	< .001
	Megosztas	0.016	0.003	0.370	4.712	< .001

Linear Regression ▾

Model Summary - Koncertjegyek ▾

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
H ₀	0.848	0.719	0.709	1.971	0.719	71.774	2	56	< .001
H ₁	0.862	0.743	0.729	1.904	0.024	5.055	1	55	0.029

Note. Null model includes Reklam, Megosztas

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₀	Regression	557.865	2	278.932	71.774	< .001
	Residual	217.631	56	3.886		
	Total	775.496	58			
H ₁	Regression	576.185	3	192.062	53.000	< .001
	Residual	199.311	55	3.624		
	Total	775.496	58			

Note. Null model includes Reklam, Megosztas

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	2.748	0.677		4.062	< .001
	Reklam	0.399	0.051	0.619	7.872	< .001
	Megosztas	0.016	0.003	0.370	4.712	< .001
H ₁	(Intercept)	3.091	0.671		4.607	< .001
	Reklam	0.396	0.049	0.615	8.102	< .001
	Megosztas	0.003	0.006	0.082	0.546	0.587
	Youtube	0.002	9.447e-4	0.329	2.248	0.029

Linear Regression ▾

Model Summary - Koncertjegyek ▾

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
H ₀	0.780	0.608	0.601	2.309	0.608	88.447	1	57	< .001
H ₁	0.861	0.742	0.732	1.892	0.133	28.929	1	56	< .001

Note. Null model includes Reklam

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₀	Regression	471.583	1	471.583	88.447	< .001
	Residual	303.913	57	5.332		
	Total	775.496	58			
H ₁	Regression	575.104	2	287.552	80.357	< .001
	Residual	200.392	56	3.578		
	Total	775.496	58			

Note. Null model includes Reklam

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4.983	0.565		8.817	< .001
	Reklam	0.502	0.053	0.780	9.405	< .001
H ₁	(Intercept)	3.290	0.560		5.876	< .001
	Reklam	0.402	0.048	0.623	8.439	< .001
	Youtube	0.003	4.775e-4	0.397	5.379	< .001

Linear Regression

Model Summary - Koncertjegyek

Model	R	R ²	Adjusted R ²	RMSE	R ² Change	F Change	df1	df2	p
H ₀	0.861	0.742	0.732	1.892	0.742	80.357	2	56	< .001
H ₁	0.862	0.743	0.729	1.904	0.001	0.298	1	55	0.587

Note. Null model includes Reklam, Youtube

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₀	Regression	575.104	2	287.552	80.357	< .001
	Residual	200.392	56	3.578		
	Total	775.496	58			
H ₁	Regression	576.185	3	192.062	53.000	< .001
	Residual	199.311	55	3.624		
	Total	775.496	58			

Note. Null model includes Reklam, Youtube

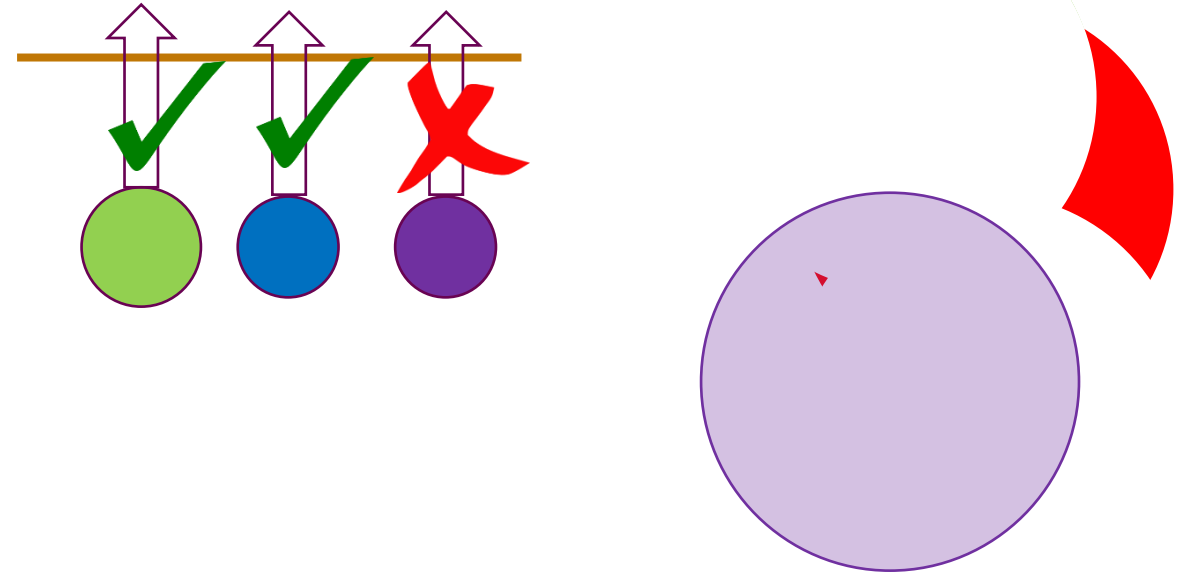
Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	3.290	0.560		5.876	< .001
	Reklam	0.402	0.048	0.623	8.439	< .001
	Youtube	0.003	4.775e-4	0.397	5.379	< .001
H ₁	(Intercept)	3.091	0.671		4.607	< .001
	Reklam	0.396	0.049	0.615	8.102	< .001
	Youtube	0.002	9.447e-4	0.329	2.248	0.029
	Megosztas	0.003	0.006	0.082	0.546	0.587

Példa modell felépítésére

Mi történik ha **FORWARD** módszert választunk?

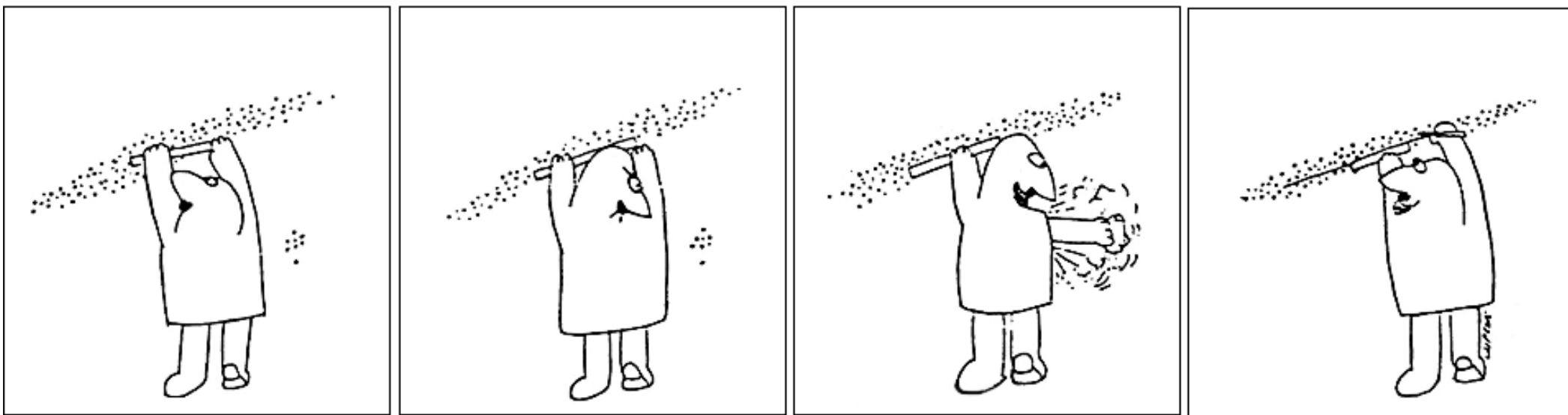
- Először bekerül a **reklám**, mert ő magyaráz meg legtöbbet a kimeneti változóból
- Ezt követően megnézzük, a maradék megmagyarázatlan variáciát, melyik változó magyarázza jobban, és ez szignifikánsan hozzá tesz-e a modellhez. Így kerül be a **youtube**
- Végül megvizsgáljuk, hogy a maradék megmagyarázatlan variációhoz a **poszt** hozzá tud-e tenni. Mivel nem, ezért ő nem kerül be a modellbe
- Eredményként a **reklám**ot és **youtube**-ot viszonylag pontosan látjuk, a **posztról** viszont semmit nem fogunk tudni.
- Azt, hogy a második lépésben a **youtube** vagy a **poszt** kerül a modellbe, nem mi döntöttük el, hanem a Forward algoritmus



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

a. Dependent Variable: Eladott koncertjegy (ezer db)

Többszörös lineáris regresszió feltételei



① Változók típusa

② Nincs nulla variáciájú változó

③ Nincs kollinearitás és nincs multikollinearitás

④ Nincs külső változó

⑤ Homoszkedaszticitás

⑥ Linearitás

⑦ Független hibák

⑧ Hiba normális eloszlása

⑨ Elemszám

⑩ Nincsenek (több)dimenziós outlierok

① Változók típusa

- Kimeneti változó skála típusú lehet csak.
- Prediktor változó skála típusú vagy dichotóm lehet.
- Ha több kategóriával rendelkező nominális változót szeretnél a modellbe helyezni, azt dummy változókká alakítással érheted el.
- Bizonyos ordinális változókkal egyes kutatók megengedőbbek, ha elég sok szintje van a változónak, pl. iskolai végzettség.
- Nem tesztel kell ellenőrizni, hanem a mérőeszköz/kérdőív kialakításánál kell gondoskodni arról, hogy a változód megfelelő mérési szintűek legyenek.
- Ha nem teljesül, legtöbbször váltani kell logisztikus vagy ordinális regresszióra, Spearman korrelációk vagy Khi-négyzet próbák sorozatára, esetleg diszkriminancia elemzésre.

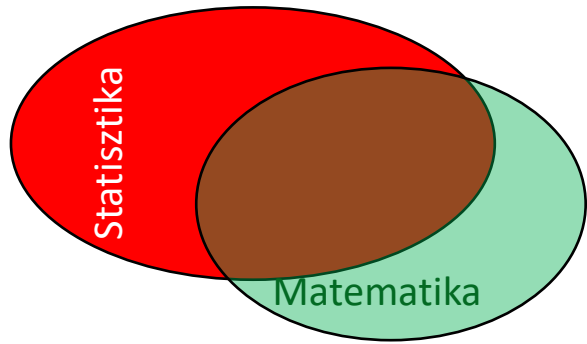
②Nincs nulla varianciájú változó

- Csak olyan változó kerüljön a modellbe, melynek van varianciája.
- Evidensnek tűnhet, mégis kezdő kutatóknál gyakran látok olyan modelleket, ahol például nemi hatást vizsgálnak egy csupa-női mintában.
- A deskriptív statisztikák alapján ellenőrizendő.
- Ha egy változóban nincs változatosság, az semmilyen elemzésbe nem helyezhető be...

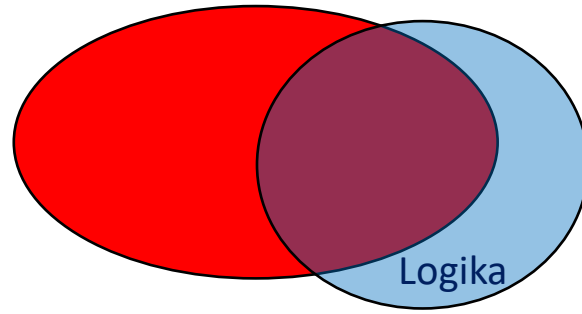
③Nincs kollinearitás és nincs multikollinearitás

- Mindkettő a prediktorok közötti együttjárásra vonatkozik.
- Kollinearitás: a prediktor változók közötti tökéletes /nagyon erős együttjárás.
- Multikollinearitás: több prediktor változó közötti közepes/erős korreláció.

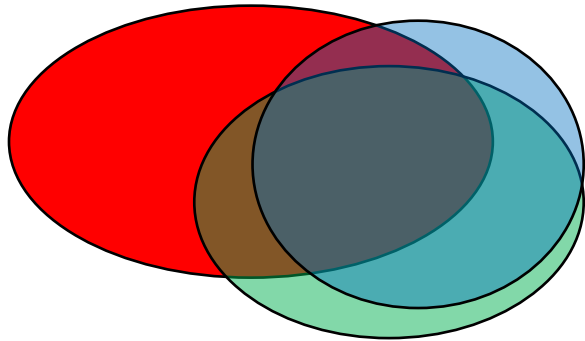
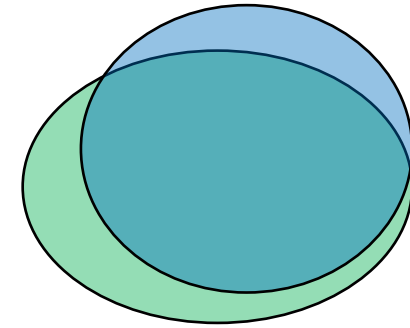
Kollinearitás



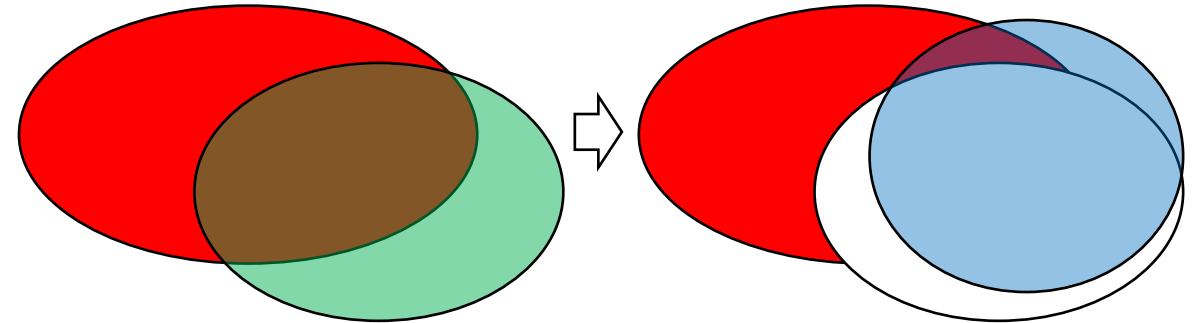
A statisztikai tudást mind a matematikai tudás, mind a logikai készségek viszonylag jól magyarázzák.



A baj az, hogy matematika tudás és a logikai készségek között erős a korreláció.



Ezért a statisztika tudás variációjából majdnem teljesen ugyanazt a rész magyarázzák. Ezért a regresszióban tévesen alulbecsülnénk a magyarázóerejüket.



Sőt Stepwise modellnél az is előfordulhat, hogy ha a matematika tudás lép először a modellbe, a logikai készségeknek már „nem marad további variancia”, amit magyarázhatna, és be sem kerül a modellbe.

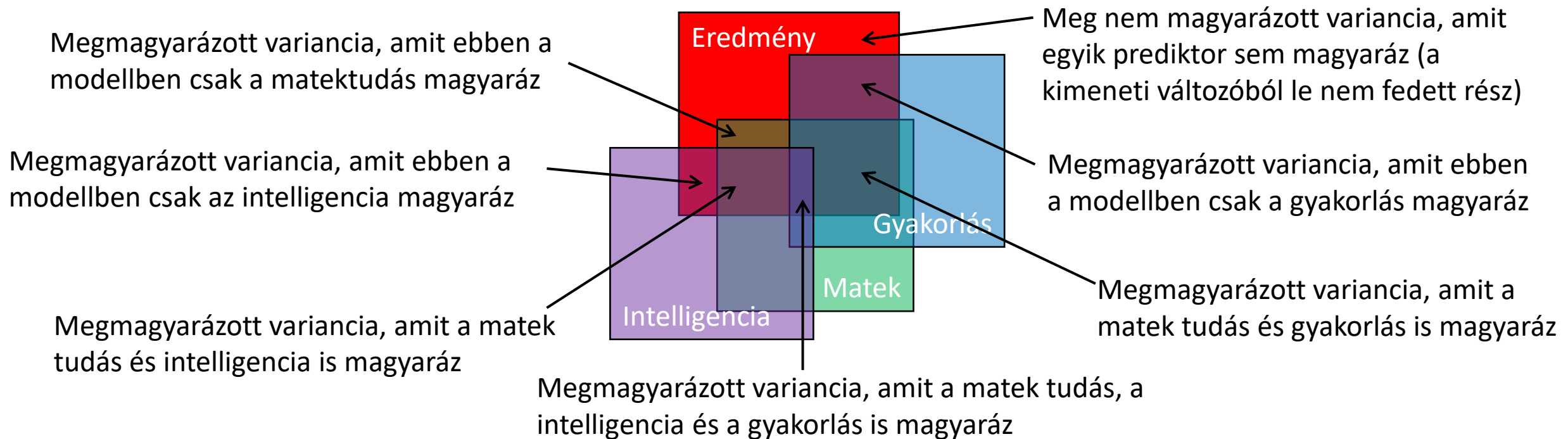
③Nincs kollinearitás és nincs multikollinearitás

- Mindkettő a prediktorok közötti együttjárásra vonatkozik.
 - Kollinearitás: a prediktor változók közötti tökéletes /nagyon erős együttjárás.
 - Multikollinearitás: több prediktor változó közötti közepes/erős korreláció.
-
- A modell (F és R^2) értelmezhető magad, de csökken egy-egy változó értelmezésének lehetősége.
 - Legnagyobb nehézség lineáris regresszió értelmezésénél.
-
- A kollinearitás ellenőrizhető korrelációs táblákkal.

Multikollinearitás

Több prediktor változó közötti közepes/erős korreláció.

A probléma ugyanaz, mint a kollinearitásnál, csak nehezebb észrevenni, mert nem két változó korrelál egymással nagyon erősen, hanem sok közepesen erősen.

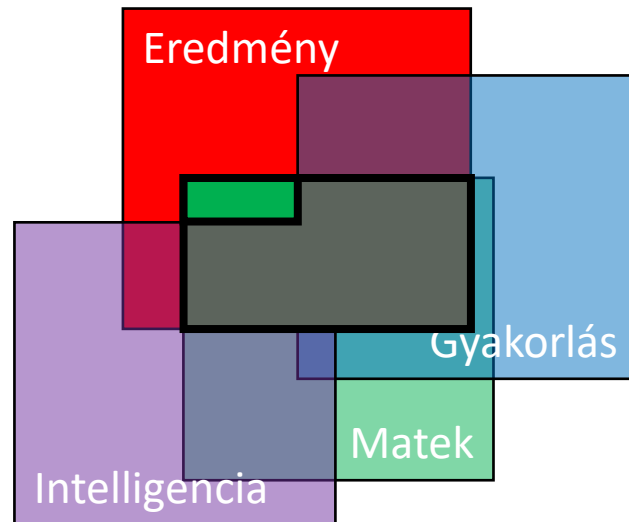


Multikollinearitás

Több prediktor változó közötti közepes/erős korreláció.

A probléma ugyanaz, mint a kollinearitásnál, csak nehezebb észrevenni, mert nem két változó korrelál egymással nagyon erősen, hanem sok közepesen erősen.

Akár egy erős prediktor hatása is teljesen kimutathatatlaná válik több, vele közepesen korreláló prediktor jelenlétében. Gyenge prediktorok hatása gyakran válik kimutathatatlaná akár enyhe multikollinearitás esetén is.



③Nincs kollinearitás és nincs multikollinearitás

- Mindkettő a prediktorok közötti együttjárásra vonatkozik.
- Kollinearitás: a prediktor változók közötti tökéletes /nagyon erős együttjárás.
- Multikollinearitás: több prediktor változó közötti közepes/erős korreláció.
- A probléma a kollinearitás és multikollinearitás, a feltétel pedig az, hogy ezek ne legyenek jelen a regressziós modellünkben.
- A modell (F és R^2) értelmezhető magad, de csökken egy-egy változó értelmezésének lehetősége.
- Legnagyobb nehézség lineáris regresszió értelmezésénél.
- A kollinearitás ellenőrizhető korrelációs táblákkal.
- A multikollinearitás ellenőrizhető a tolerancia értékei segítségével.

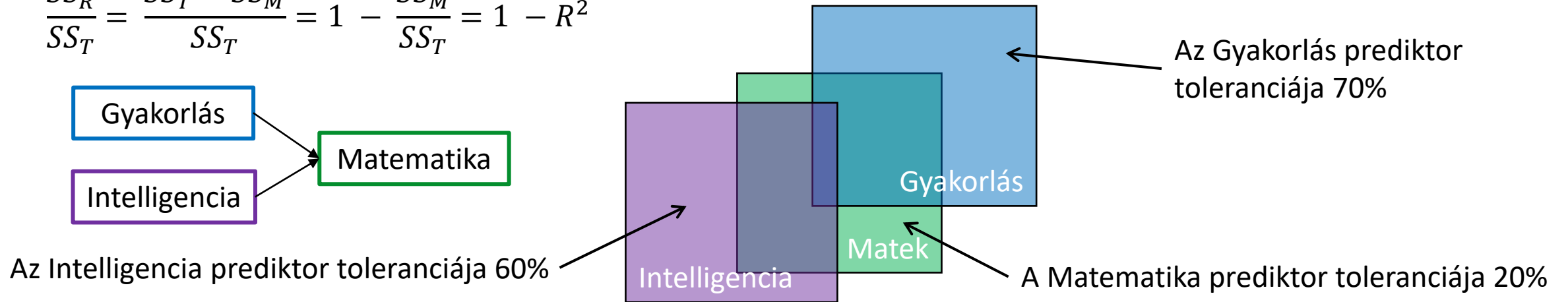
Tekintsünk el egy pillanatra a kimeneti változótól, és nézzük csak a prediktor változókat.

Nézzük meg regressziók sorozatával minden egyes prediktor változóra, hogy a többi prediktor változó milyen mértékben magyarázza a változatosságukat, és milyen mértékben függetlenek a többi prediktortól.

- Teljes változatosság (SS_T): prediktor változó teljes változatossága
- Közös variancia (SS_M): a prediktor változó változatosságának az a része, amit a többi prediktor magyarázni tud.
- Egyéni variancia (SS_R): a prediktor változó változatosságának az a része, amit a többi prediktor nem tud megmagyarázni.

Tolerancia: a prediktor változó változatosságának hány százaléka egyéni variancia.

$$\frac{SS_R}{SS_T} = \frac{SS_T - SS_M}{SS_T} = 1 - \frac{SS_M}{SS_T} = 1 - R^2$$

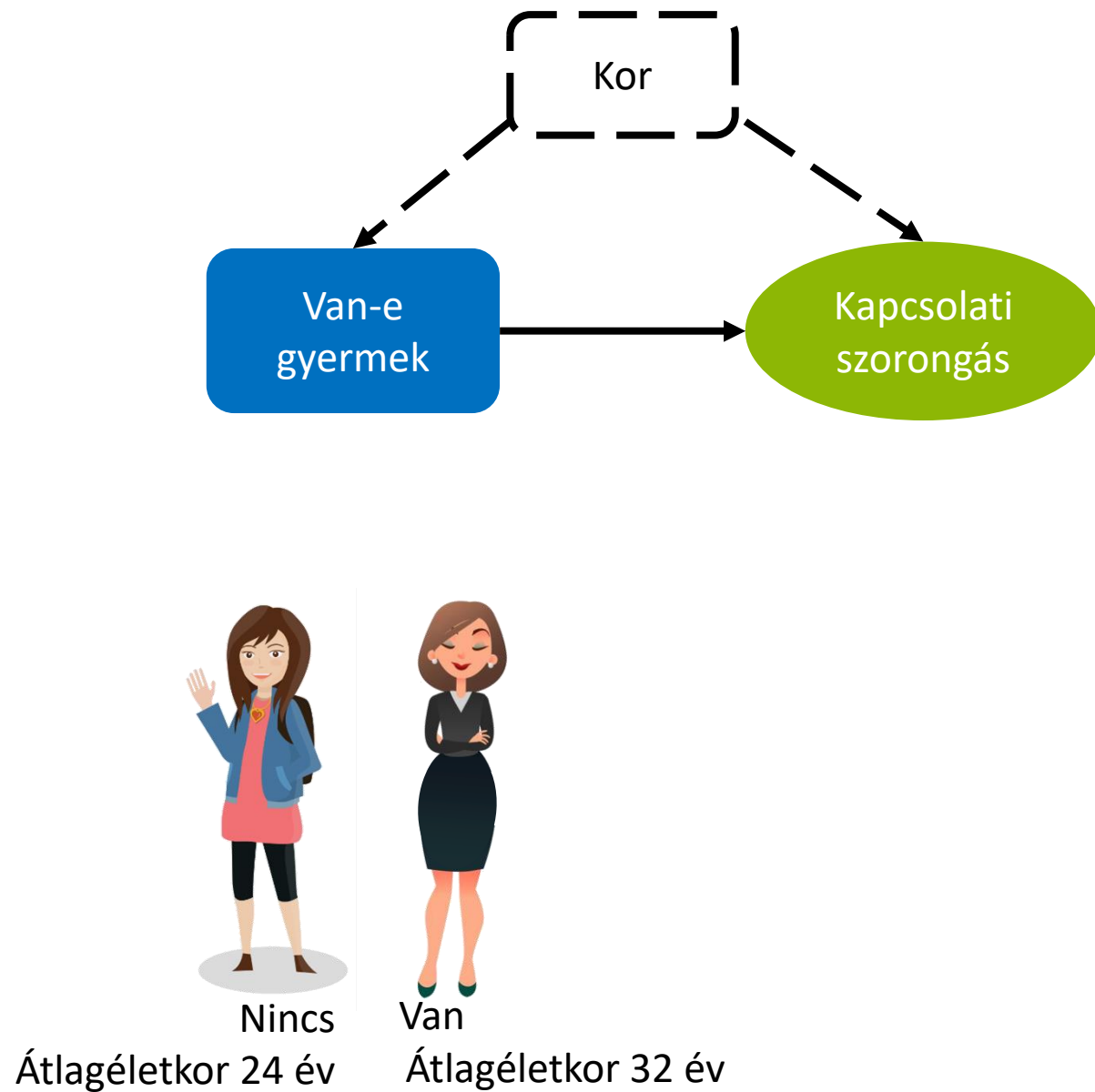
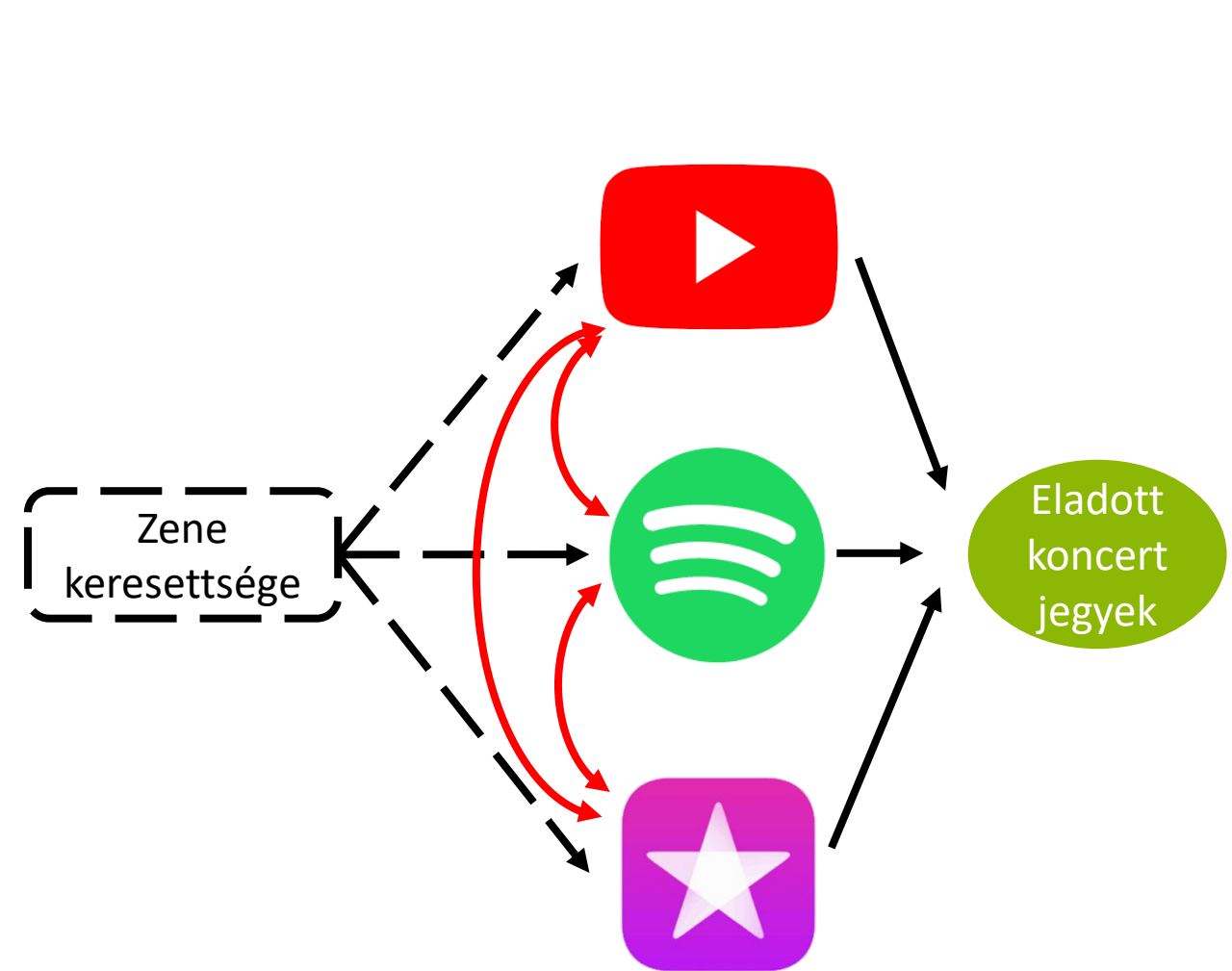


③Nincs kollinearitás és nincs multikollinearitás

- Mindkettő a prediktorok közötti együttjárásra vonatkozik.
- Kollinearitás: a prediktor változók közötti tökéletes /nagyon erős együttjárás.
- Multikollinearitás: több prediktor változó közötti közepes/erős korreláció.
- A probléma a kollinearitás és multikollinearitás, a feltétel pedig az, hogy ezek ne legyenek jelen a regressziós modellünkben.
- A modell (F és R^2) értelmezhető magad, de csökken egy-egy változó értelmezésének lehetősége.
- Legnagyobb nehézség lineáris regresszió értelmezésénél.
- A kollinearitás ellenőrizhető korrelációs táblákkal.
- A multikollinearitás ellenőrizhető a tolerancia értékei segítségével.
- A megoldás a helyes modellépítésben rejlik, illetve az összefüggésrendszer több modell segítségével való körüljárásában.
- Legtöbbször ki kell venni a kollinearitásban résztvevő változók közül egyet vagy többet vagy főkomponens elemzéssel összevonni őket.

④Nincs külső, modellbe be nem emelt fontos tényező

- Nincs olyan, a modellbe be nem emelt külső tényező,
 - mely a prediktorok között erős korrelációt hoz létre, mert az kollinearitást vagy multikollinearitás okozhat,
 - magyarázza a kimeneti változó és prediktor(ok) korrelációját, mert az illuzórikus összefüggést hozhat létre a kimeneti változó és a prediktorok között.



④Nincs külső változó

- Nincs olyan, a modellbe be nem emelt külső tényező,
 - mely a prediktorok között erős korrelációt hoz létre, mert az kollinearitást vagy multikollinearitás okozhat,
 - magyarázza a kimeneti változó és prediktor(ok) korrelációját, mert az illuzórikus összefüggést hozhat létre a kimeneti változó és a prediktorok között.
- A kutatás tervezésekor kell átgondolni milyen prediktor változók fontosak a modell szempontjából.
- A modell alapos átgondolával kerülhető el.
- Szerencsés esetben az esetleges külső hatásokat mediáló és moderáló elemzéssel lehet átlátni.
- Szerencsétlen esetben az egész elemzés értelmezhetetlenné válhat.

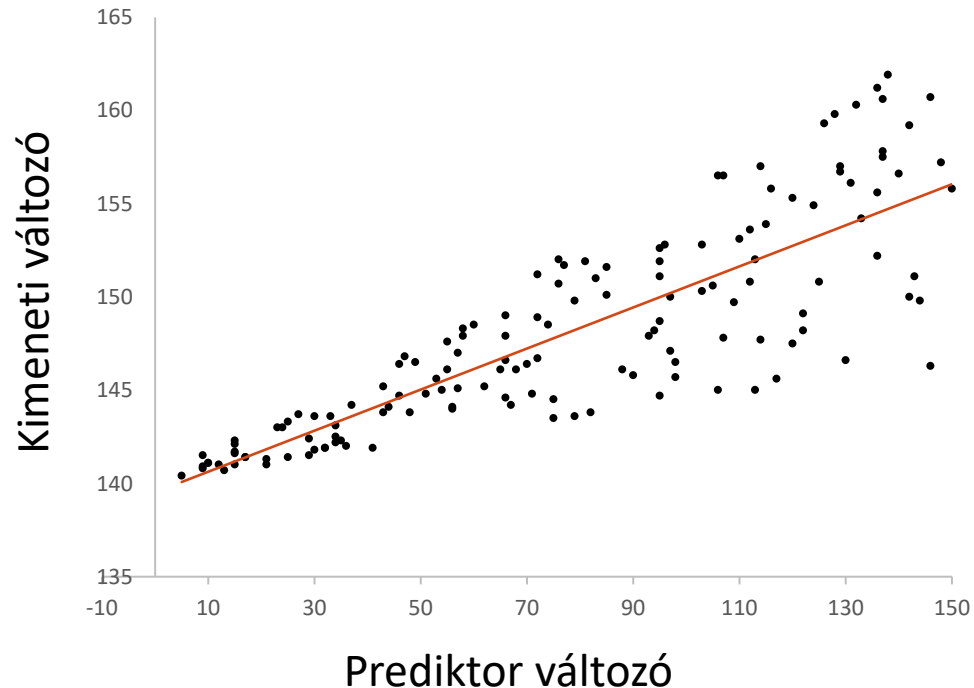
⑤ Homoszkedaszticitás / Varianciák homogenitása / Szóráshomogenitás

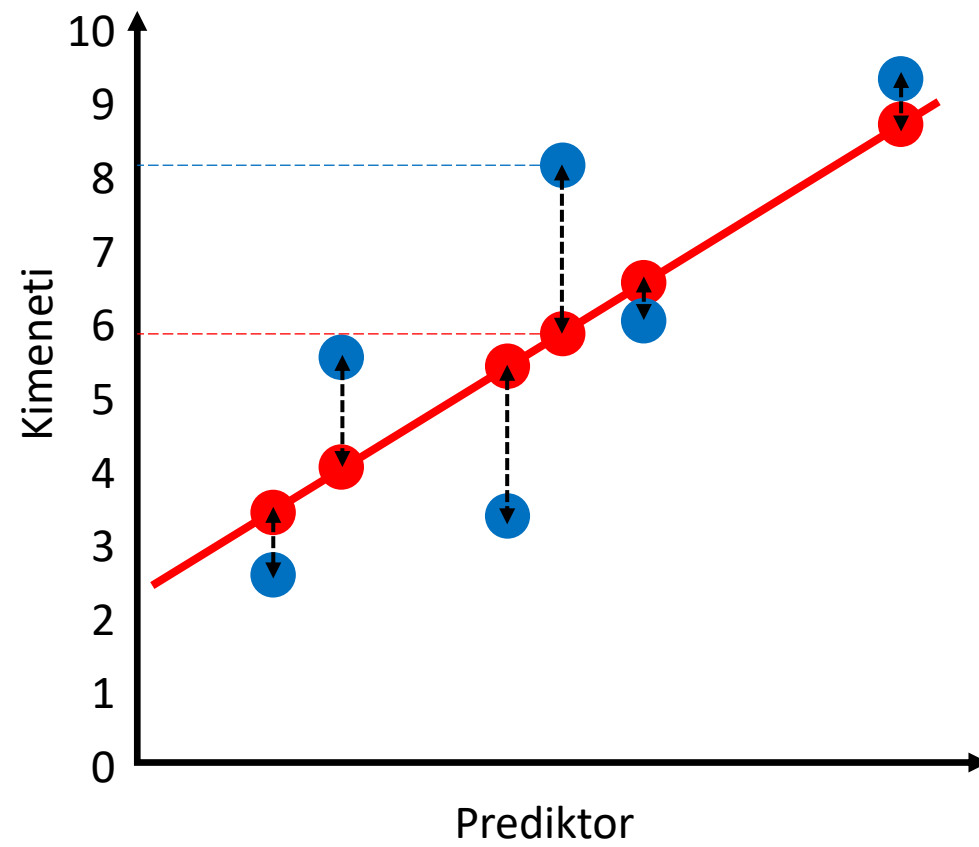
Egyetlen prediktor (2D) esetén: a prediktor minden szintjén hasonló nagyságú a kimeneti változó szórása.

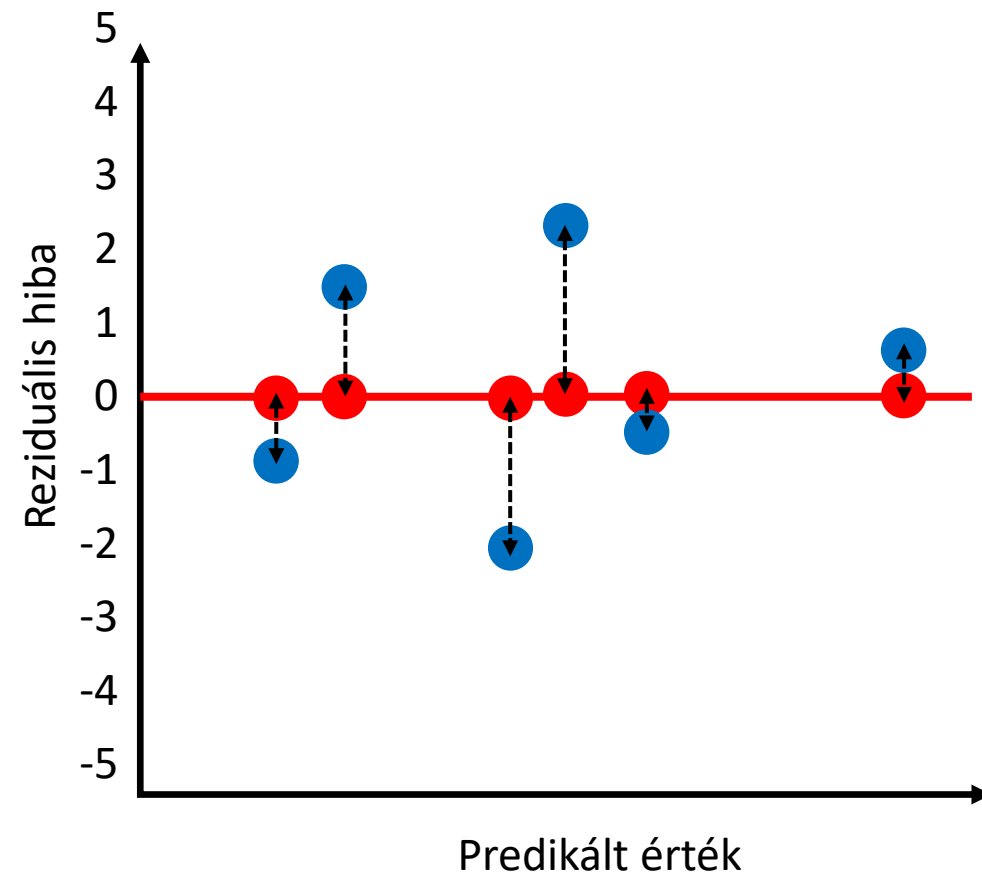
Fogalmazzuk át úgy, hogy több prediktorra (3D és felette) is szemléletes maradjon!

A prediktor változó helyett (amiből most több van) a predikált érték mentén haladunk, és azt figyeljük meg, hogy a mért értékek milyen távol vannak a predikáltaktól, azaz mekkora a reziduális hiba mértéke

A predikált értékek minden szintjén hasonló a reziduális hibák nagysága, azaz a hibák nagysága homogén.







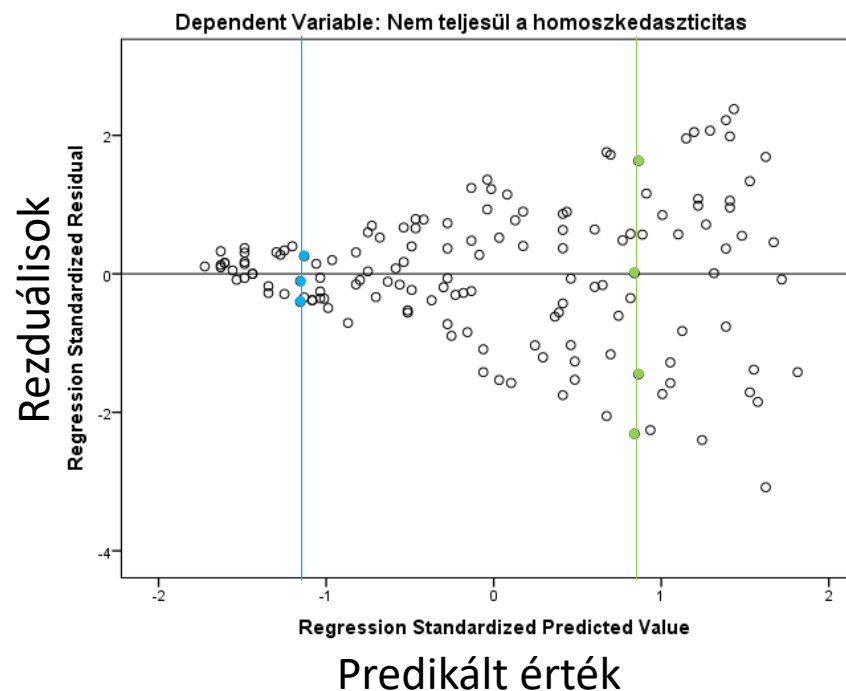
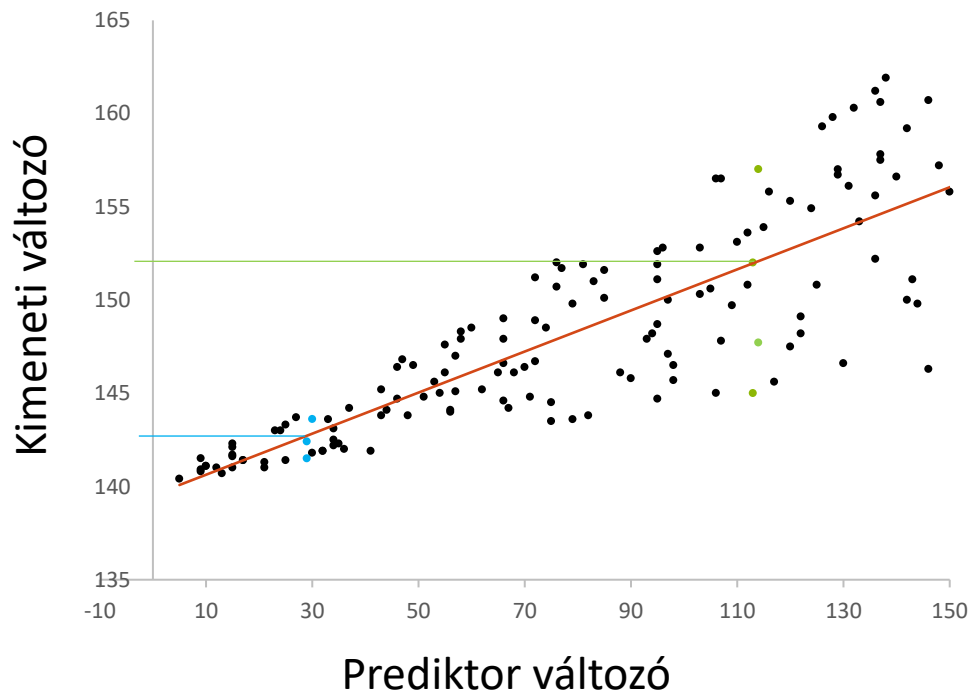
⑤ Homoszkedaszticitás / Varianciák homogenitása / Szóráshomogenitás

Egyetlen prediktor (2D) esetén: a prediktor minden szintjén hasonló nagyságú a kimeneti változó szórása.

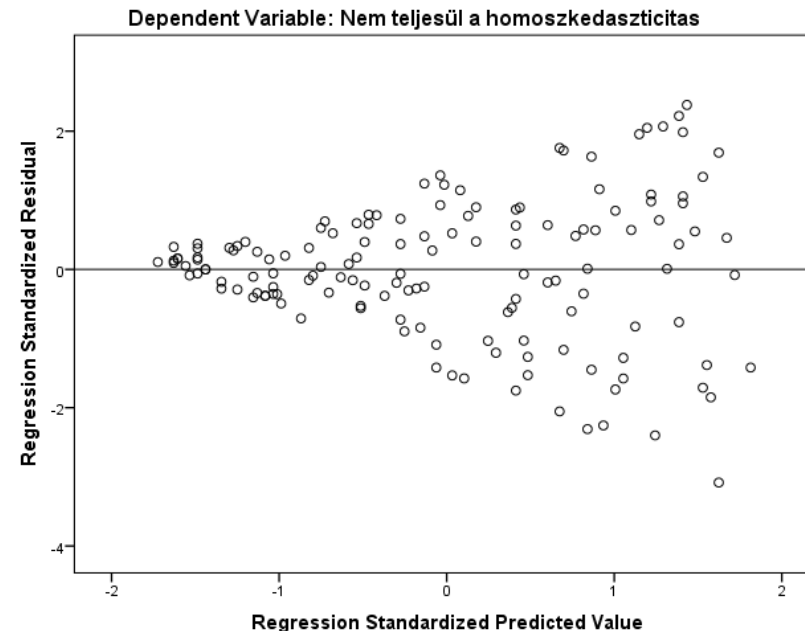
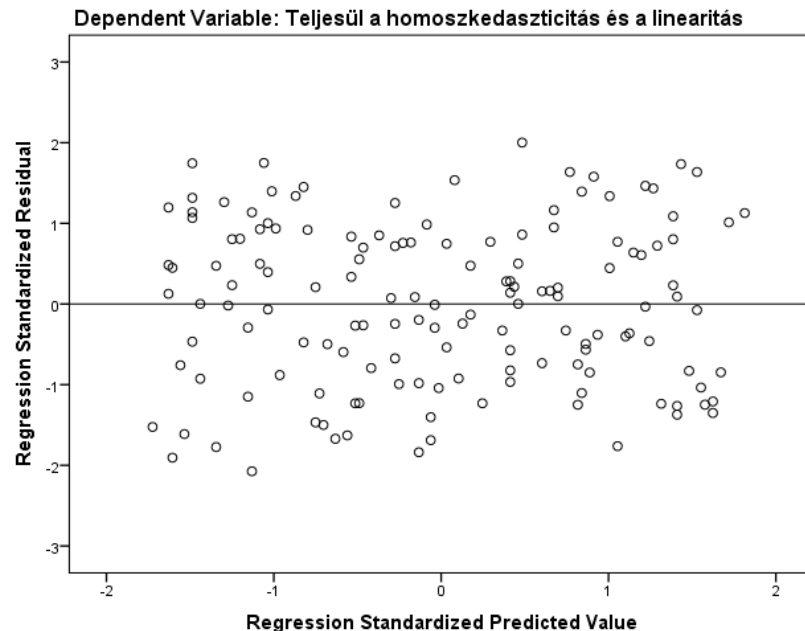
Fogalmazzuk át úgy, hogy több prediktorra (3D és felette) is szemléletes maradjon!

A prediktor változó helyett (amiből most több van) a predikált érték mentén haladunk, és azt figyeljük meg, hogy a mért értékek milyen távol vannak a predikáltaktól, azaz mekkora a reziduális hiba mértéke

A predikált értékek minden szintjén hasonló a reziduális hibák nagysága, azaz a hibák nagysága homogén.

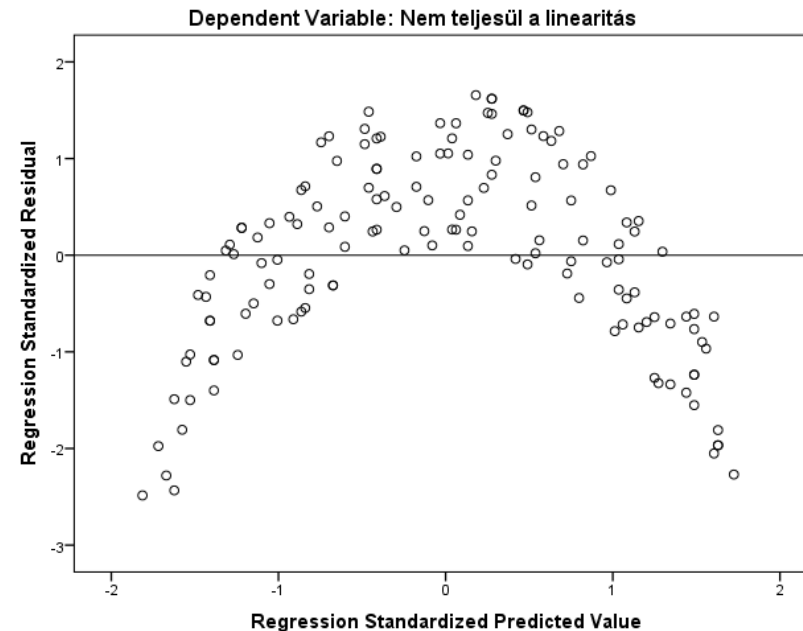
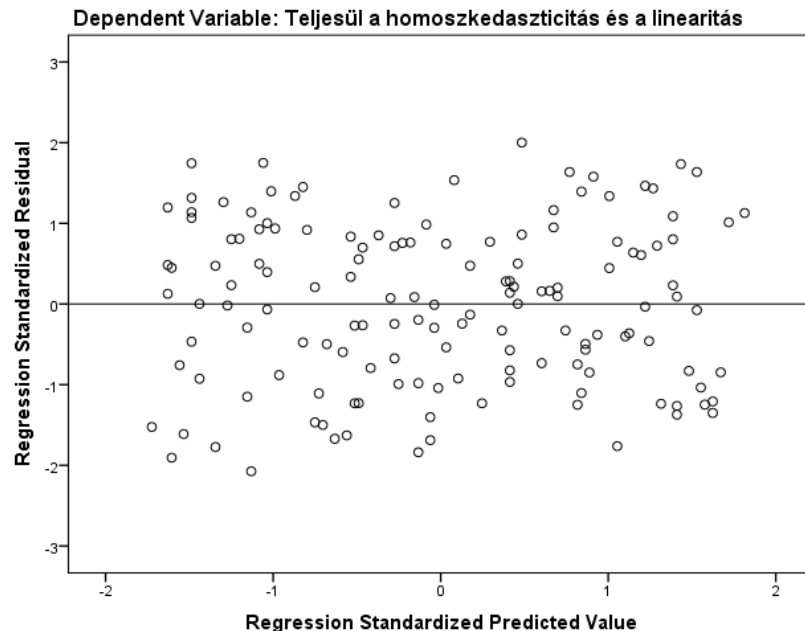


- ⑤ Homoszkedaszticitás / Varianciák homogenitása / Szóráshomogenitás
 - A predikált érték minden szintjén hasonló a reziduális hibák nagysága / a hibák nagysága homogén.
 - Ha nem teljesül a homoszkedaszticitás (ergo heteroszkedaszticitás van), a regresszióval tett predikció nem mindenhol ugyanolyan mértékben bízható meg (ahol nagyon a hibák, ott pontatlanabb)
 - Ellenőrizhető az elemzés során a **predikált** és **reziduális** értékek plottolásával
 - Ha az előző hibát mutat, az megkereshető az elemzés során a **prediktor változók** és a **reziduális értékek** plottolásával prediktoronként egyenként

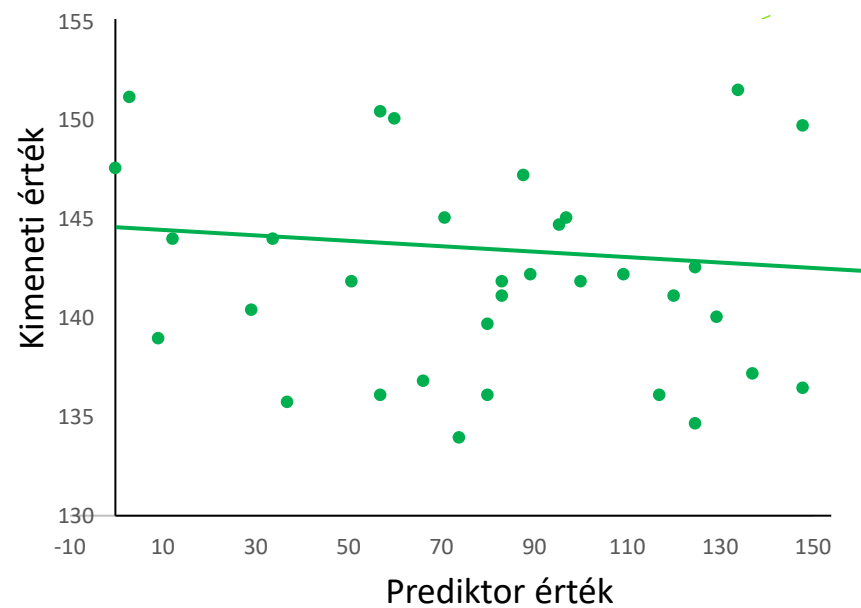


- ⑥Linearitás

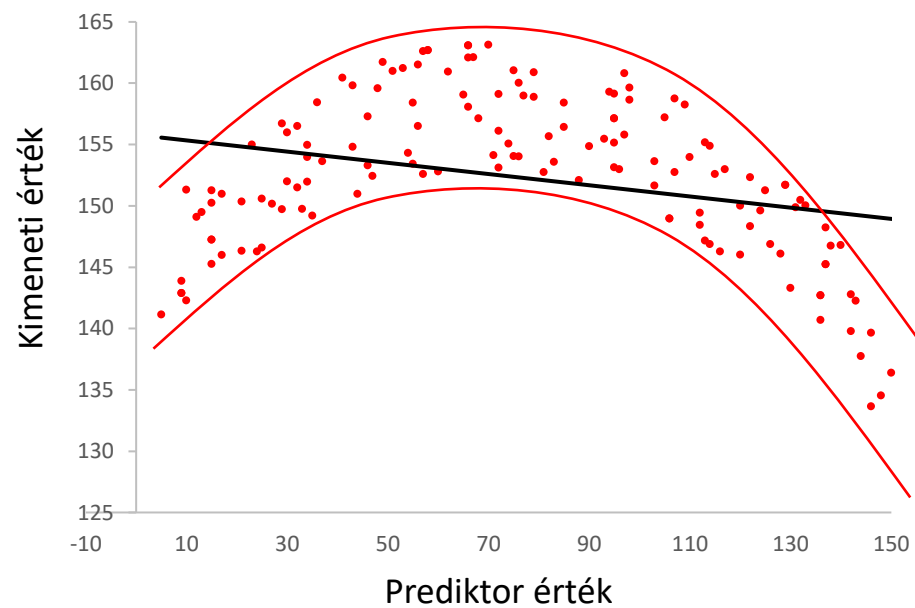
- Lineáris kapcsolat van minden prediktor változó és a kimeneti változó között (a többi prediktor változóra kontrollálva)
- Ellenőrizhető az elemzés során a **predikált értékek és reziduális értékek**, vagy a **predikált értékek és kimeneti változó** közötti kapcsolat plottolásával (elég az egyiket kikérni)
- Ha a linearitás sérülésére gyanakszunk, ellenőrizhető előzetesen a **prediktor változó és kimeneti változó** közötti kapcsolat plottolásával prediktoronként egyenként



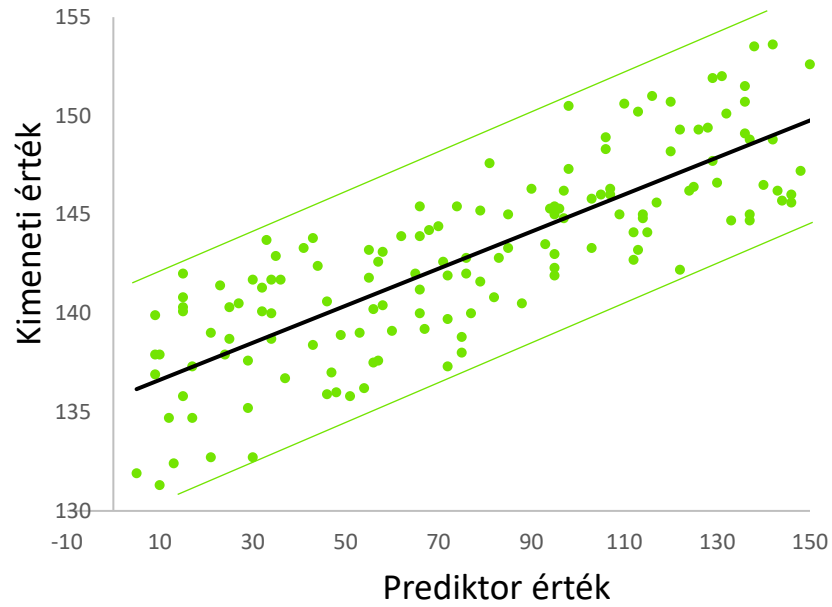
Teljesül a homoszkedaszticitás és linearitás



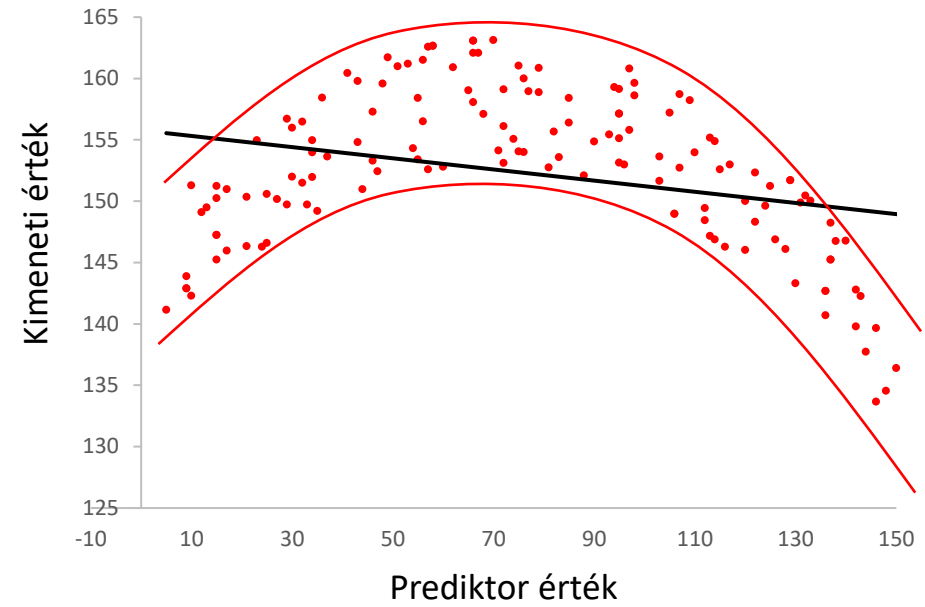
Nem teljesül a linearitás



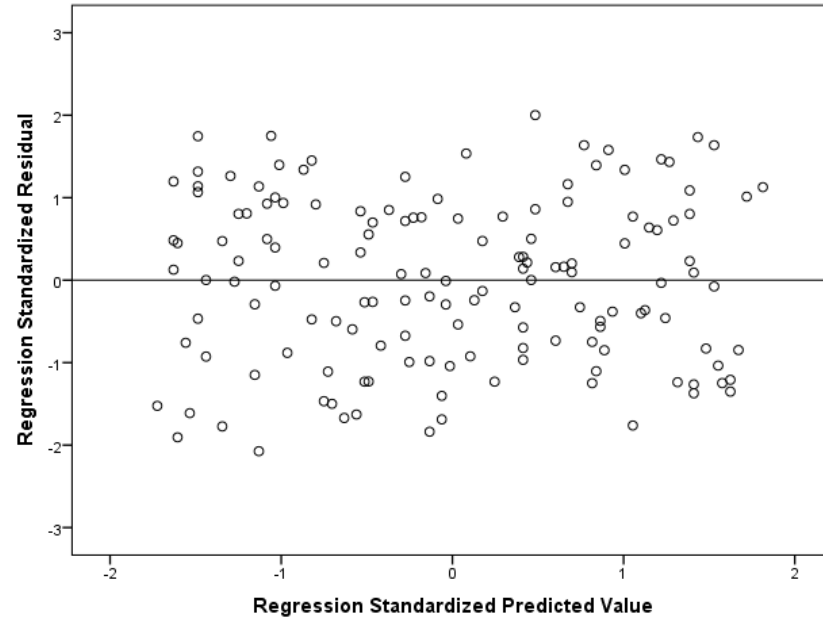
Teljesül a homoszkedaszticitás és linearitás



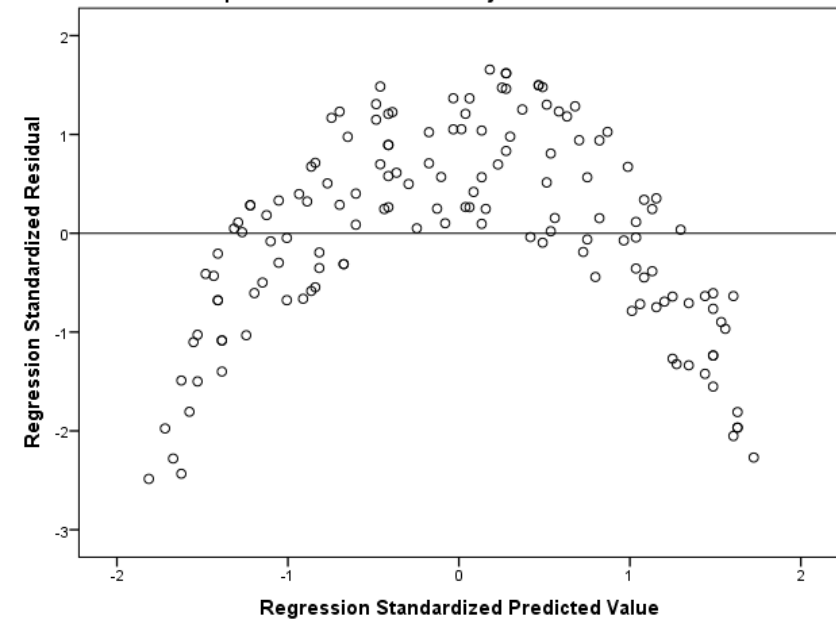
Nem teljesül a linearitás



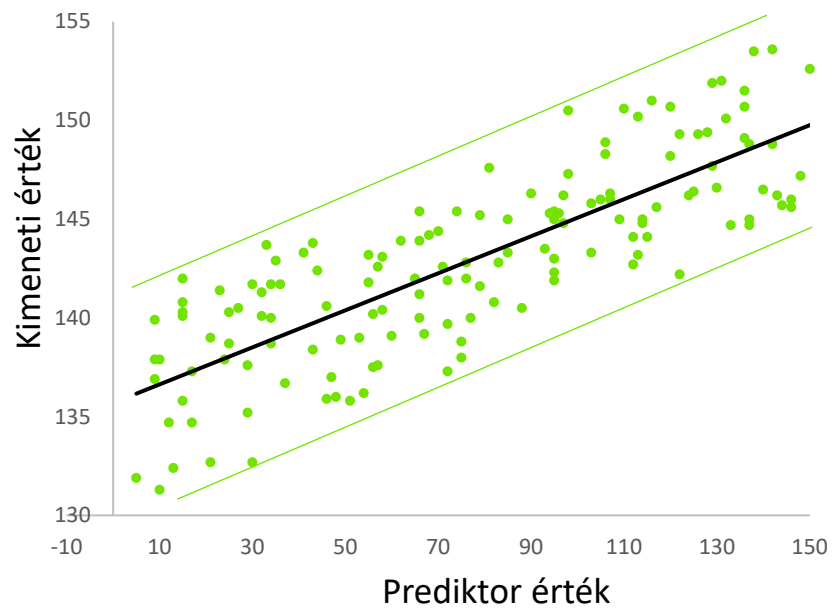
Dependent Variable: Teljesül a homoszkedaszticitás és a linearitás



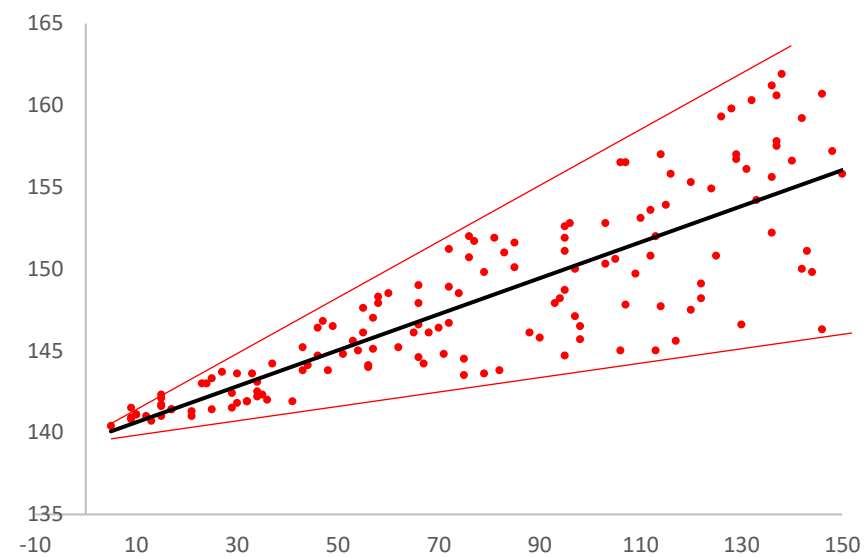
Dependent Variable: Nem teljesül a linearitás



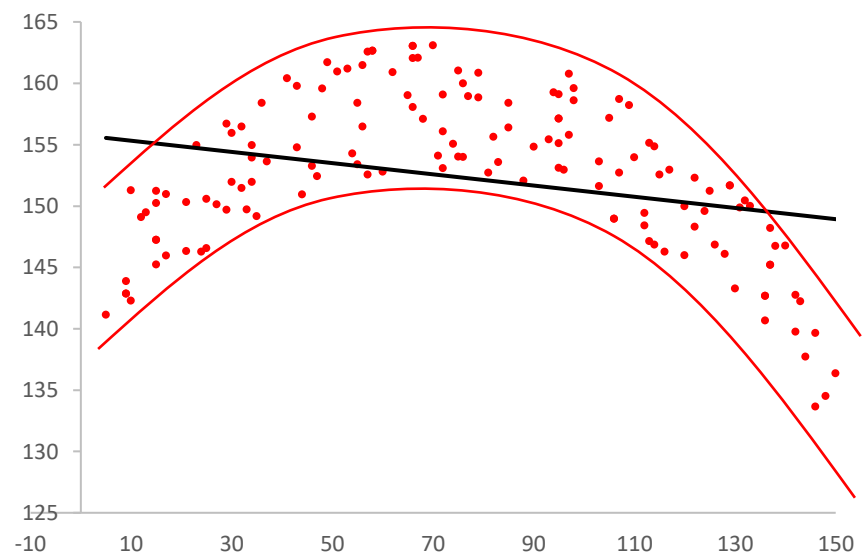
Teljesül a homoszkedaszticitás és linearitás



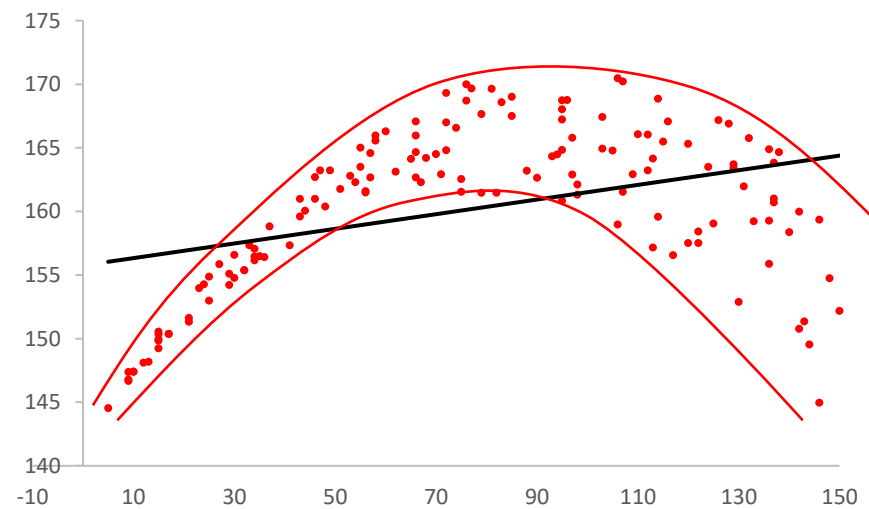
Nem teljesül a homoszkedaszticitás

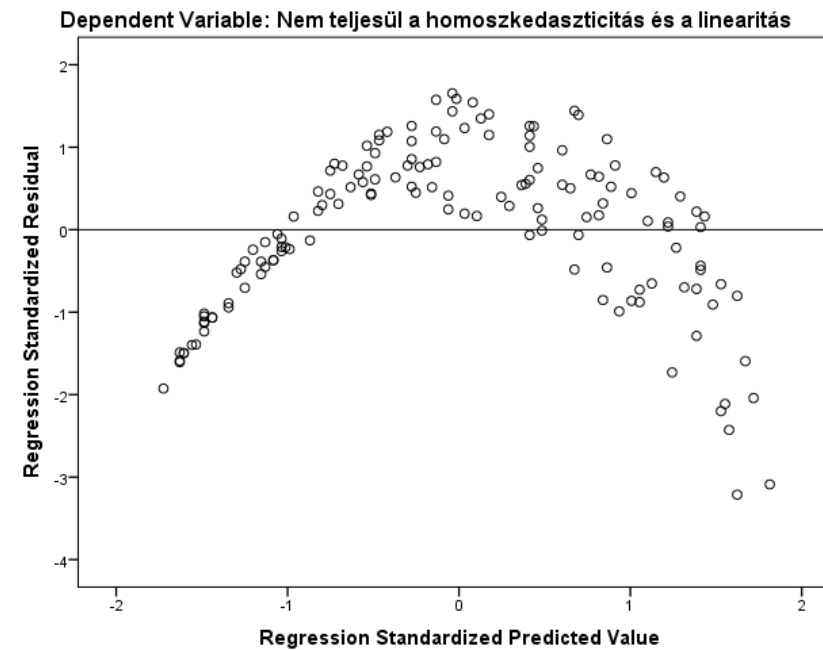
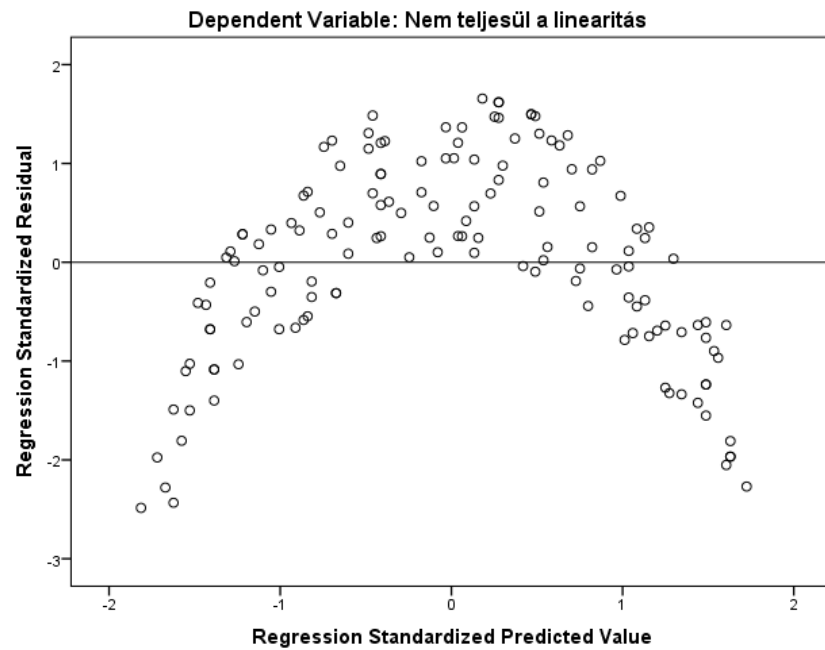
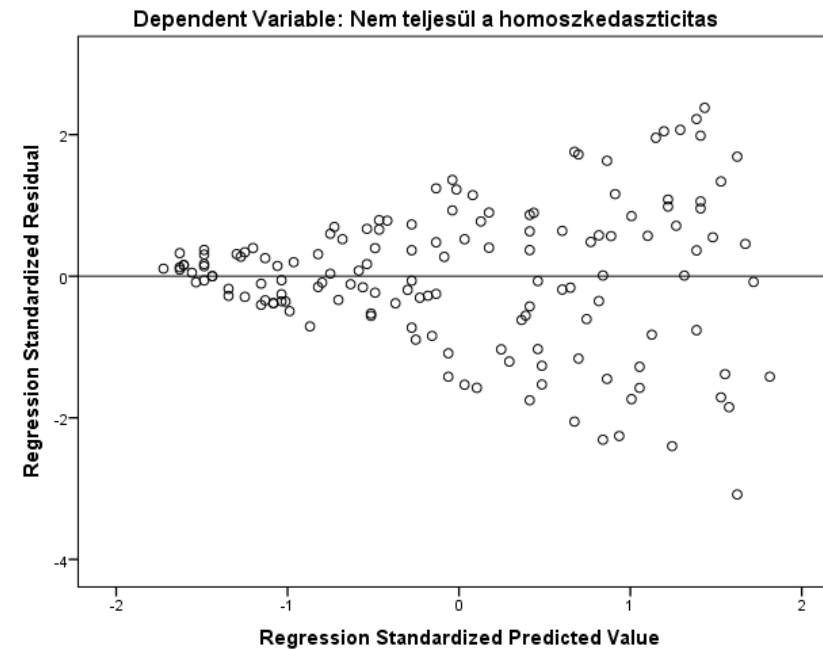
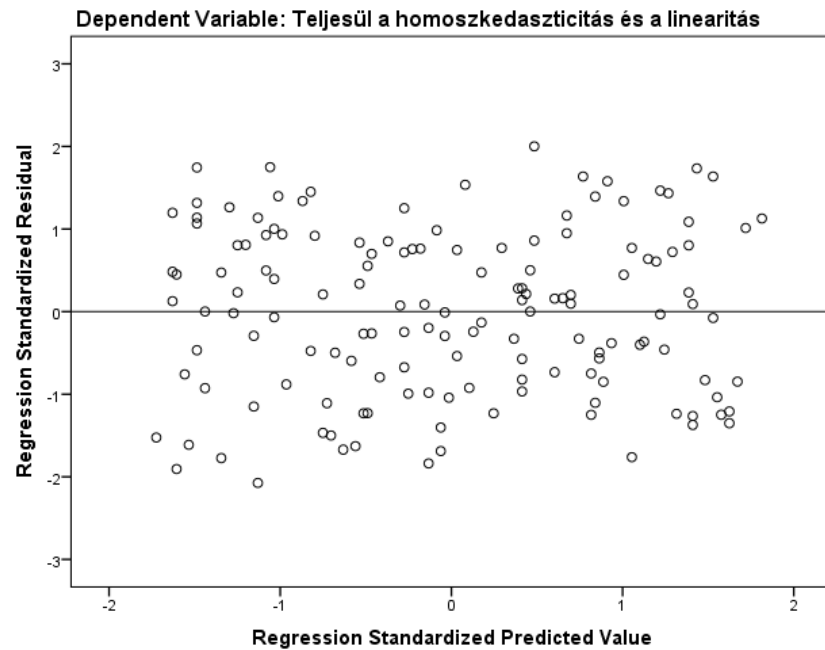


Nem teljesül a linearitás



Nem teljesül a homoszkedaszticitás és linearitás sem





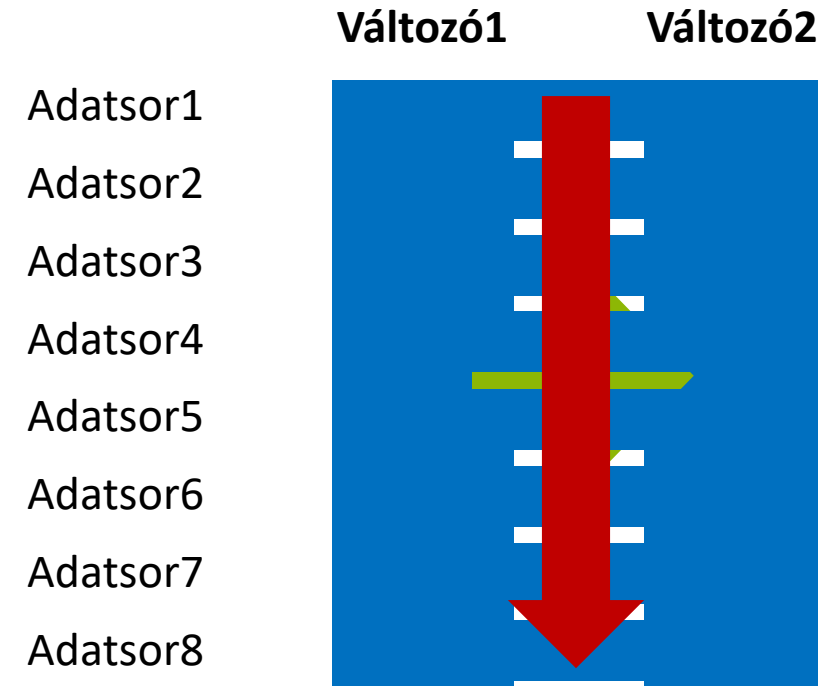
⑦ Független hibák

Feltétel, hogy a reziduális hibák ne korreláljanak

A korreláló hibatagok

- Idősorok elemzése során gyakran előforduló hiba (pszichológiában ritka)
- Rosszul specifikált modellt jelezhet (rosszul kiválasztott prediktor változók)
- Linearitás sérülésének jele is lehet

Ellenőrizhető az elemzés során a Durbin-Watson teszttel, melynek értéke 2 körül tökéletes, 1 alatt vagy 3 felett jelez bajt



⑧ Hiba normális eloszlása

A predikciónk hibájának függetlennek kell lennie a predikciónktól.
A hiba a véletlen műve, azaz zaj. Mivel soktényezős, ezért a CLT-nek megfelelően normáloszlást kell kövessen.

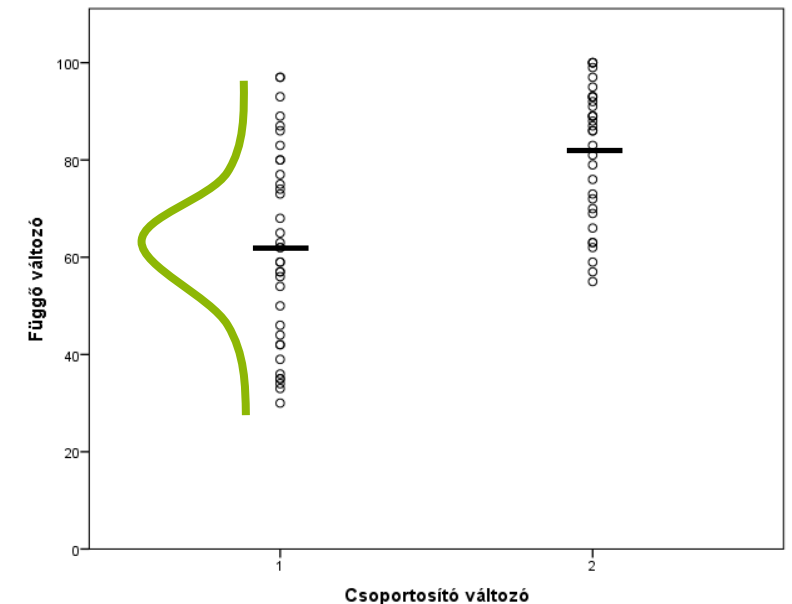
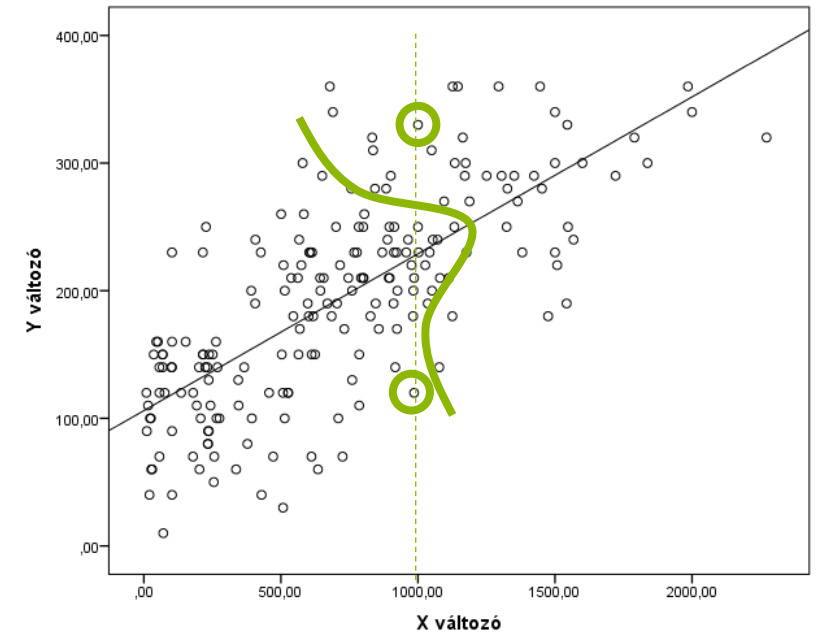
A konfidencia intervallumok és a regressziós együtthatókhoz tartozó tesztek kiszámításához szükséges.

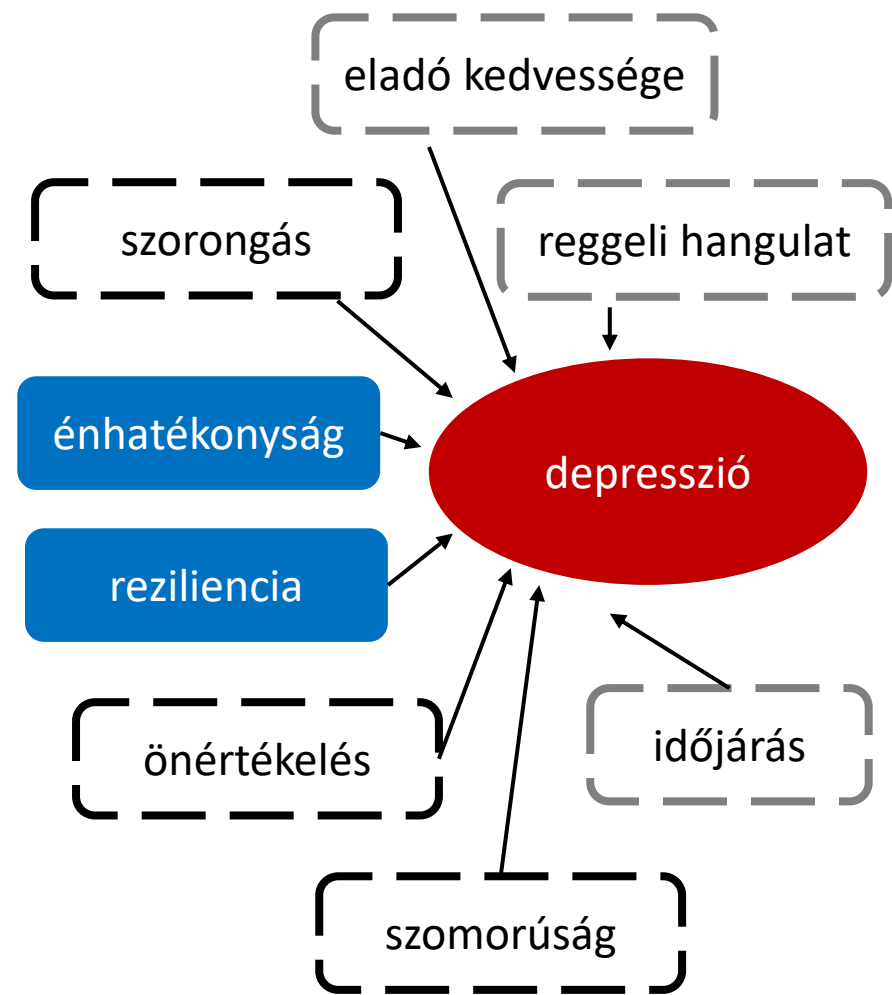
A prediktorok vagy kimeneti változó normál eloszlása NEM feltétele a lineáris regressziónak, de ha ezek nem követnek normál eloszlást, gyakran a hibák sem követnek.

A linearitás sérülése is okozhatja a hibák nem normális eloszlását

Ellenőrizhető az elemzés során a hiba hisztogramjának plottolásával vagy a hibaértékek mentésével, és az így létrejövő új változó normalitásának tesztelésével.

Ha nem teljesül, a linearitást, homoszkedaszticitást és a változók (főleg a kimeneti változó) normalitását érdemes ellenőrizni.





⑨Elemszám

Függ:

- (1) prediktorok számától,
- (2) prediktorok közötti összefüggések erősségétől,
- (3) a hatások nagyságától,
- (4) és attól, hogy milyen következtetést akarunk levonni.

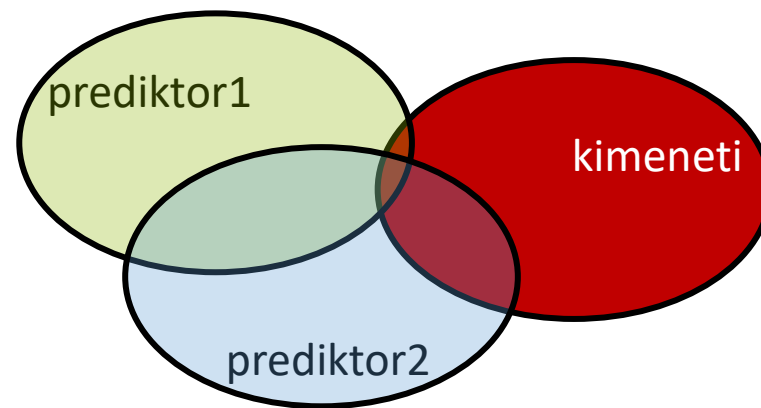
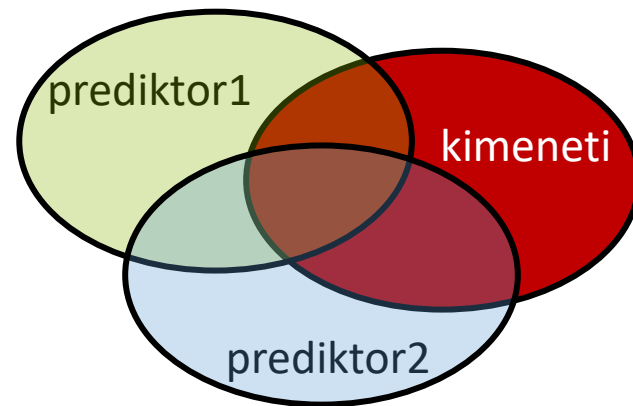
Ökölszabály: változók száma szorozva 10 vagy 15

Ha nagy a hatás várunk: 80 mindig elég (20 változóig)

Közepes hatást várunk: 200 mindig elég (20 változóig),

de kevés változóval is minimum 60

Kicsi hatást várunk: 600 kell (6 változóig)



$$\text{statisztikai érték} = \frac{\text{hatás}}{\text{hiba}} = \frac{\frac{SS_M}{df_M}}{\frac{SS_R}{df_R}} = \frac{MS_M}{MS_R} = F \rightarrow p$$

⑩Nincsenek (több)dimenziós outlierok

Önmagában nem feltétel, de a többi feltétel teljesülését sérti

Többdimenziós outlier külön-külön egyik dimenzió belül sem szélsőséges érték, csak az adott kombináció szélsőséges

Például 50kg-os és 190cm magas ember

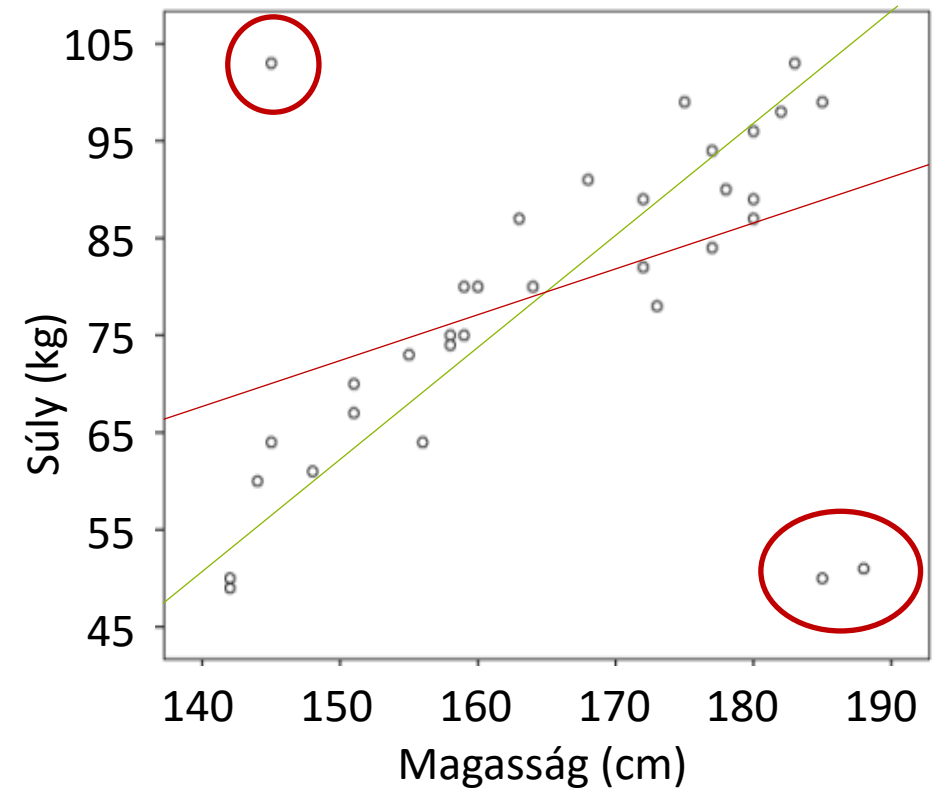
Ellenőrizhető az elemzés során

Cook távolság

Mahalanobis távolság

Leverage values

Standardizált reziduálisok



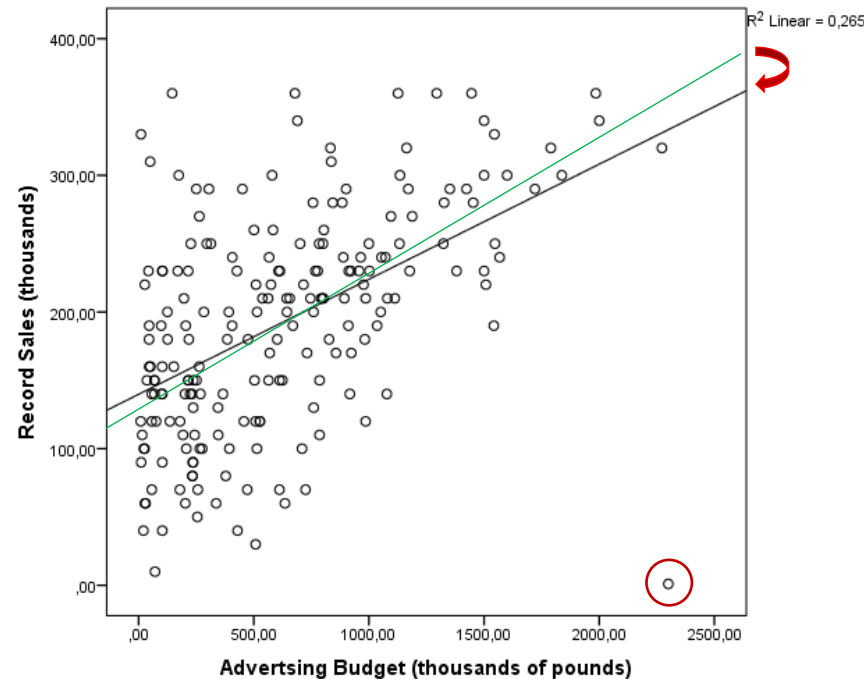
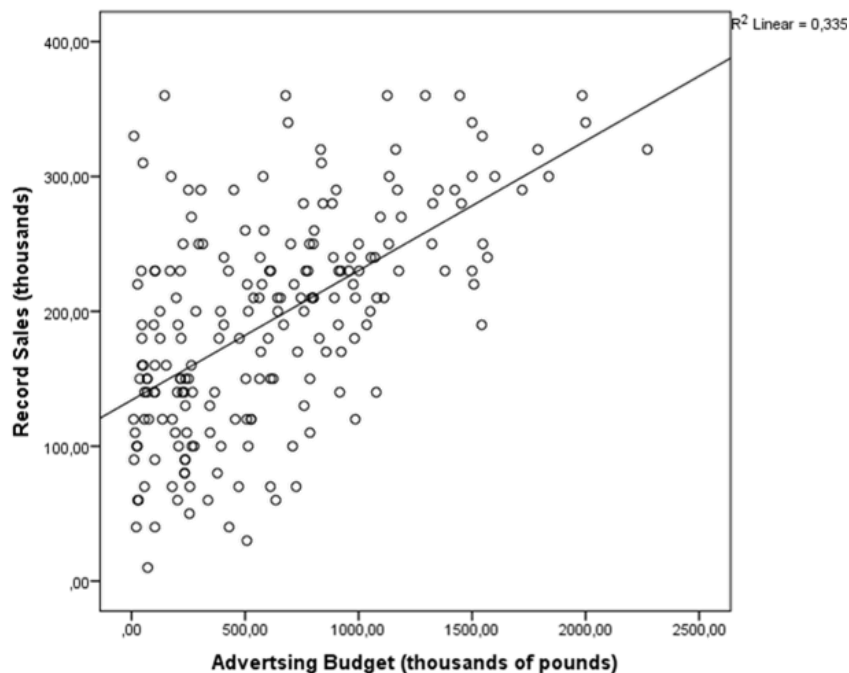
Cook távolság

Minden egyes érték regressziós modellre gyakorolt hatását becsli meg.

- Mennyire lenne más a regressziós modelled, ha az az érték nem lenne benne.
- A szélsőségesen értékek jobban el tudják téríteni az egyenest a maguk irányába

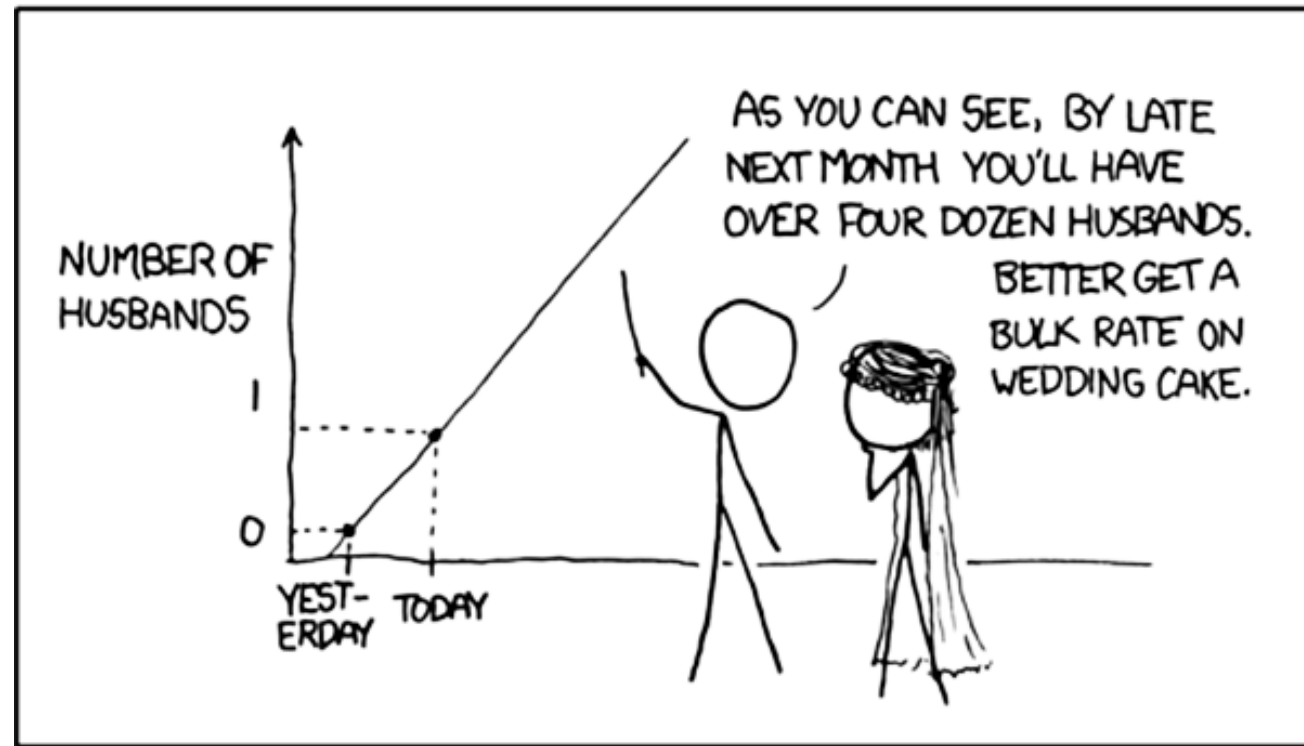
Mindenkihez külön Cook-érték tartozik, mely megadja, az adott személynek mekkora hatása van

Többdimenziós outlier az, akinek a Cook-értéke 1 felett van



Ralph Dennis Cook

Többszörös lineáris regresszió értelmezés



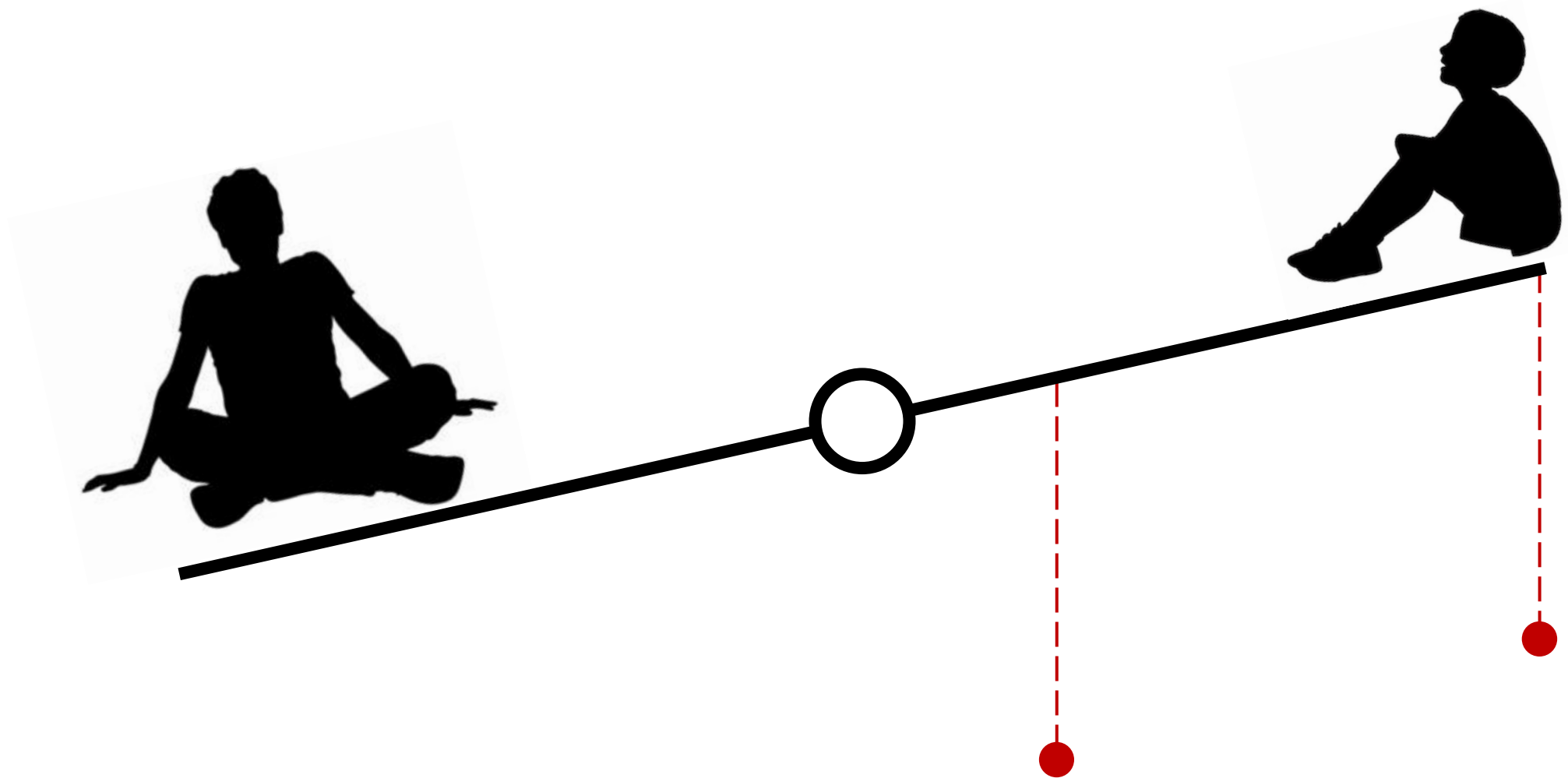
Többszörös lineáris regresszió értelmezés

A regressziós táblák értelmezése előtt a feltételek teljesülését ellenőrizzük.

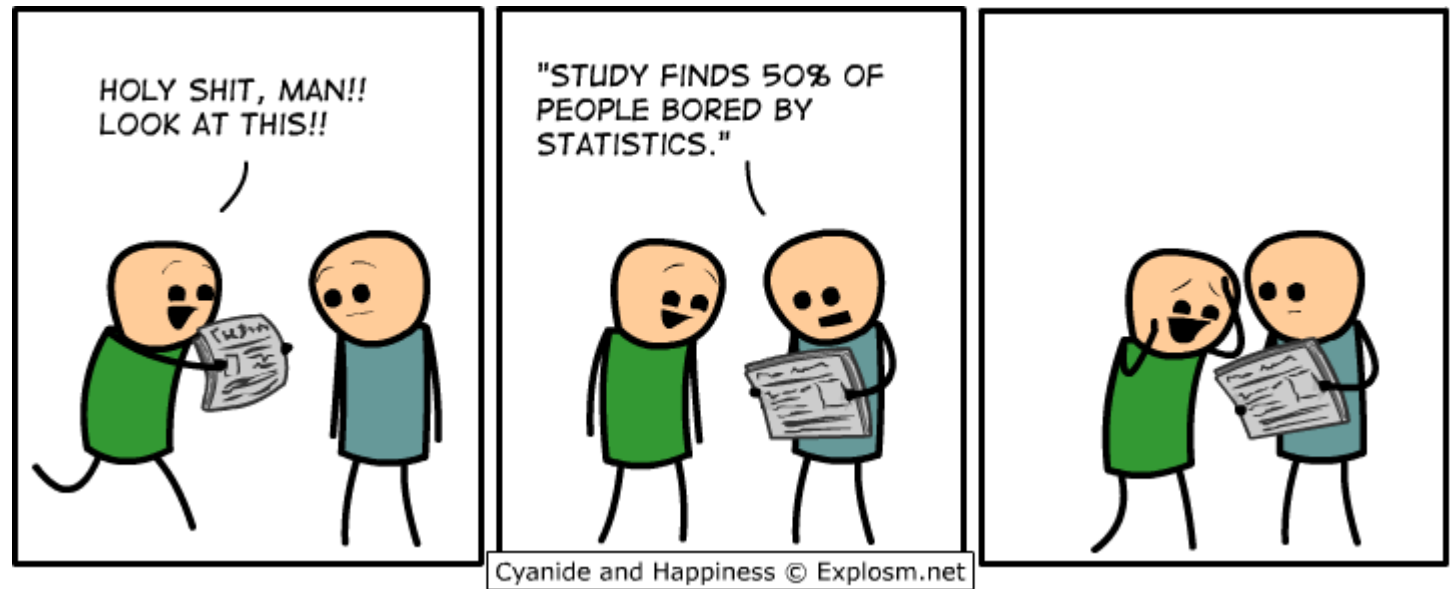
A feltételek egymásra is hatnak, ezért érdemes azokkal a feltételekkel kezdeni, melyek viszonylag függetlenek a többi teljesülésétől, vagy amelyek sérülése nagyon sok mindent „elronthat”.

Ajánlott sorrend:

- Nem ellenőrizendő, hanem a vizsgálat tervezése során kialakítandó feltételek:
 - ① Változók típusa
 - ④ Nincs külső változó
 - ⑨ Elemszám
- Elemzés során ellenőrizhető feltételek (és ajánlott sorrendjük):
 - ② Nincs nulla varianciájú változó
 - ⑩ Nincsenek (több)dimenziós outlierok
 - ③ Nincs (multi)kollelinearitás
 - ⑥ Linearitás
 - ⑤ Homoszkedaszticitás
 - ⑦ Független hibák
 - ⑧ Hiba normális eloszlása



Többszörös lineáris regresszió publikálása



Többváltozós regresszió publikálása

Kezdd a választott statisztika, a kimeneti változó és a prediktor változók megnevezésével!

A publikálás példájában a gyakorlaton elsőként elhangzott, és a gyakorlati pdf-ben részletesen értelmezett modell egy lehetséges publikálását adom meg, tehát a kimeneti változó a koncertjegy, a prediktorok a reklámba fektetett összes és a zenekar utolsó videoklipjének nézettsége Youtube-on.

Hierarchikus regresszió segítségével elemeztük, hogy egy koncertre eladott jegyek számát (továbbiakban KONCERTJEGY) milyen mértékben jósolja be a koncert reklámozásába fektetett pénz (továbbiakban REKLÁM) és a zenekar utolsó videoklipjének nézettsége a Youtube-on (továbbiakban YOUTUBE).

Többváltozós regresszió publikálása

Ha az adattisztítás során vettél ki értékeket az elemzésből, az itt leírhatod. Az egydimenziós adattisztítást, lehet, hogy leírtad már korábban, ebben az esetben nem kell itt is megismételni.

Az adatokban az outlier labelling rule (Tukey, 1977) alapján nem található egydimenziós outlier.

A többdimenziós outlier szűrésre azonban csak most került sor, tehát azt mindenképp itt kell közölni. Csak azokban a mutatóknak a kritériumszintjét közöld, amit használtál is (pl. ha csak Mahalanobis és Cook távolságot használtál, akkor csak azokat. Ha nincsen outlier, vagy úgy döntesz, nem veszed ki őket, akkor nem kell erre túl sok sort áldozni, ha viszont kivettél értéket, akkor érdemes azt a grafikonon is megmutatni.

A többdimenziós outlierok szűrésére a Cook távolságot számoltunk ki, és az adatokat az 1-es Cook-érték kritériumszintje mellett ellenőriztük (Cook és Weisberg, 1982).

Az eredmények azt mutatják, hogy az adatokban nincs többdimenziós outlier.

Többváltozós regresszió publikálása

Érdemes a változóidat bemutatni, vagy ha már korábban a módszertani résznél megtetted ezt, akkor legalább visszautalni arra a táblára.

Az adatok megtisztítását követően a regresszióba kerülő változók jellemzői a következők:

	Minimum	Maximum	Átlag	Szórás	Elemszám
Eladott koncertjegy	3.910	19.000	9.483	3.657	59
Reklámba fektetett összeg	0.010	19.85	8.958	5.676	59
Videoklip nézettsége	100	2222	1010.458	565.769	59

Eladott koncertjegy ezer darabban, a befektetett összeg millió Forintban, a videoklipp nézettsége ezer nézőben értendő

Többváltozós regresszió publikálása

Kollinearitás vizsgálatakor csak a prediktorok közötti korrelációkat figyeljük, és fölösleges betenni a teljes korrelációs táblát, elegendő csak a legkisebb és legnagyobb korrelációs együtthatót megnevezni, vagy azt írni például, hogy a prediktorok közötti korrelációs együtthatók gyengék / közepesek.

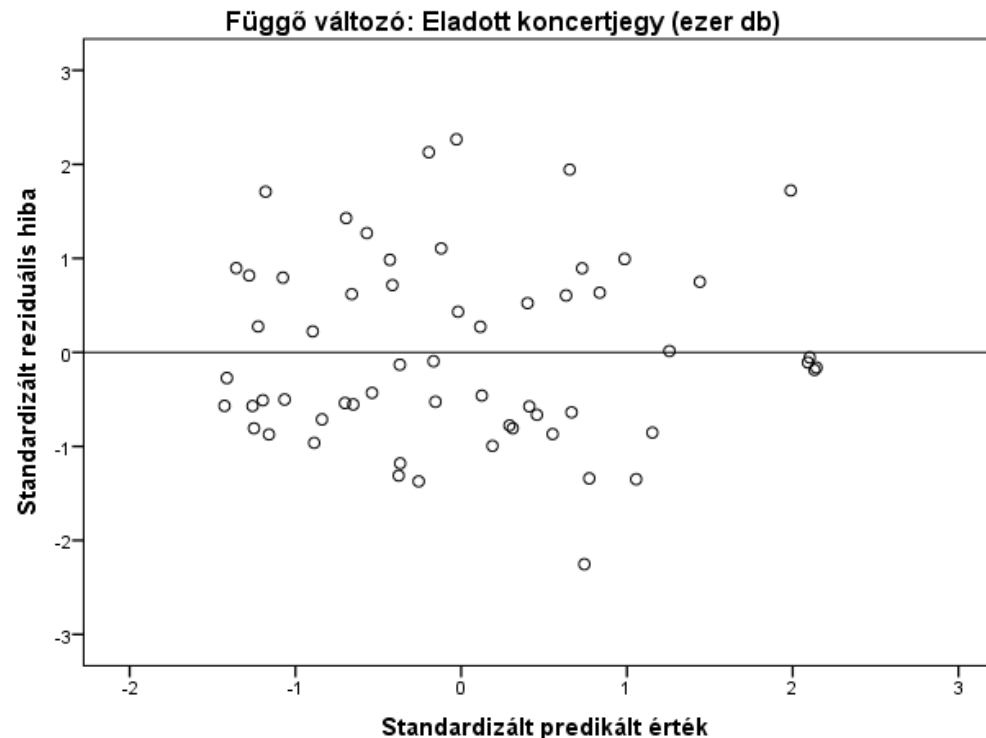
Multikollinearitáshoz használt tolerancia és VIF értékek közül elég az egyiket publikálni. Toleranciát százalékban adjuk meg, a VIF-et általában $VIF = [\text{számérték}]$ formában. Ha valamely prediktor toleranciája nagyon alacsony lenne, már itt fel lehet hívni az olvasó figyelmét, hogy azt a prediktort majd csak óvatosan szabad értelmezni.

A kollinearitás feltételének ellenőrzésére alkalmas tesztek alapján a változók között sem kollinearitás, sem multikollinearitás nem áll fenn. A prediktor változók közötti korreláció $r = .393$, és a tolerancia értékek minden prediktorra magasak: REKLÁM 84.5%, YOUTUBE 84.5%. A prediktorok nagymértékű függetlensége lehetővé teszi, hogy hatásukat tisztán értelmezzük a regresszióban.

Többváltozós regresszió publikálása

A szóráshomogenitás (aka homoszkedaszticitás) és linearitás ellenőrzéséhez használt pontdiagrammot nem feltétlen kell a dolgozatba betenni, elég a mellékletbe. Szövegesen azonban írd le, hogy teljesül a feltétel!

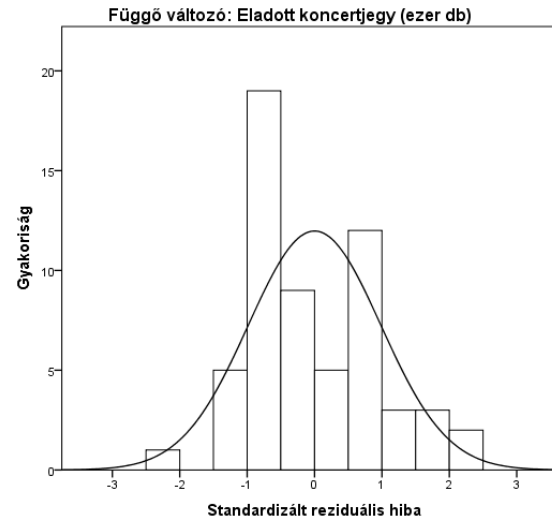
A standardizált predikált értékeket és standardizált reziduális hibákat megjelenítő pontdiagramm alapján megállapítható, hogy a mintán teljesül a homoszkedaszticitás és linearitás feltétele.



Többváltozós regresszió publikálása

A hibák normalitását elég hisztogramon ellenőrizni, és a hisztogramot elég a függelékbe betenni.

A hisztogram alapján a standardizált reziduális hibák normalitásának feltétele teljesül.



Ha viszont bizonytalan voltál a hisztogramban, és kimentetted a standardizált reziduálisokat, és elvégezted rajtuk a S-W tesztet, azt az annak megfelelő módon kell publikálni.

A standardizált reziduális hibák normalitás-feltételének teszteléséhez Ghasemi és Zahedi-Asl (2012) útmutatását követve az ismertebb Lilliefors korrigált Kolmogorov-Smirnov teszt helyett Shapiro-Wilk normalitás tesztet használtam. A teszt eredményei alapján a hiba eloszlásgörbéje nem tér el szignifikánsan a normál eloszlástól, $W(59) = .968$ $p = .128$.

Többváltozós regresszió publikálása

Hibák korrelálatlanságáról a Durbin-Watson értéket kell betenni.

Az adatokon teljesül a reziduális hibák függetlenségének feltétele, a Durbin-Watson érték = 1.891 (Durbin és Watson Durbin, 1951).

Többváltozós regresszió publikálása

Írd le, hogyan épült fel a modelled. Ez segít értelmezni az olvasónak, hogy a különböző modellek mit takarnak. A *hierarchikus* szó azt jelenti, hogy több blokkod is volt. Ha csak egy blokkod volt, és ebbe az egyetlen blokkba tetted be ENTER-rel az összes változód, akkor *többszörös* regressziót írn!

Hierarchikus regresszió első szintjén a REKLÁM változót léptettük be a modellbe (ide írhatod, hogy ENTER módszerrel, de fölösleges, mert egyetlen változó van csak). A második szinten a YOUTUBE került az elemzésbe.

Többváltozós regresszió publikálása

Az eredményeket inntentől lehet szövegesen vagy táblázatosan is közölni, de a kettő közül általában csak az egyiket használjuk. Ha egyszerű a modell, akkor érdemes szövegesen írni ki az eredményeket, de egy összetettebb modell esetében átláthatóbb lehet a táblázatos leírás.

Példaként először leírom szövegesen, majd megmutatom azt is, hogy néz ki táblázatosan.

Az első modellben a REKLÁM a KONCERTJEGY varianciájának 60,8%-át magyarázza ($r^2_{Adj} = .601$), a modell szignifikáns, $F(1,57) = 88.447$ $p < .001$. A REKLÁM hatása szignifikáns ($\beta = .780$ $t(59) = 9.405$ $p < .001$), egy millió Forinttal több befektetés az eladási mennyiségben 502db növekedést jelent ($B = 0.502$ $SE_B 0.053$).

A második modellben a REKLÁM és YOUTUBE pretiktorok a KONCERTJEGY varianciájának 74,2%-át magyarázzák ($r^2_{Adj} = .732$). Mind a modell, mind a megmagyarázott variancia előző modellhez viszonyított növekedése is szignifikáns, $F(2,56) = 80.357$ $p < .001$ és $F(1,56) = 28.929$ $p < .001$. Mindkét prediktor változó szignifikánsan bejósolja az eladott koncertjegyek számát, és mindkettőnek pozitív hatása van arra. REKLÁM: $\beta = .623$ $t(59) = 8.439$ $p < .001$; YOUTUBE: $\beta = .397$ $t(59) = 5.379$ $p < .001$.

Ez volt a szöveges publikálás. Egy ilyen egyszerű modellnél még lehet a szöveges verziót választani, de ennél bonyolultabbnál már előnyösebb a táblázat, mert sokkal átláthatóbb. Ekkor a táblázatban megjelenő értékeket nem kell külön szövegesen is leírni, minden mást, az értelmezést viszont igen!

Többsváltozós regresszió publikálása

Publikálás táblázatos formában:

Nyers reg. együtthetők B0-hoz és B1-hez tartozó SE Standardizált reg. együtthetők A t-próba df értéke az SS_T df értékével azonos

	B	SE B	β	t
Modell 1				
Konstans	4.983	0.565		$t(58) = 8.817 \text{ } p < .001$
REKLÁM	0.502	0.053	.780	$t(58) = 9.405 \text{ } p < .001$
Modell 2				
Konstans	3.290	0.560		$t(58) = 5.876 \text{ } p < .001$
REKLÁM	0.402	0.048	.623	$t(58) = 8.439 \text{ } p < .001$
YOUTUBE	0.003	0.0005	.397	$t(58) = 5.379 \text{ } p < .001$
Modell1: $R^2 = .608$ $R^2_{Adj} = .601$ $F(1,57) = 88.447 \text{ } p < .001$.				
Modell2: $R^2 = .742$ $R^2_{Adj} = .732$ $F(2,56) = 80.357 \text{ } p < .001$. Változás: $F(1,56) = 28.929 \text{ } p < .001$				
A REKLÁM a KONCERTJEGY varianciáját szignifikánsan, 60.8%-ban magyarázza. A REKLÁM és a YOUTUBE közösen 74.2%-ot magyaráz, mely szintén szignifikáns, és az előző modellhez képest szignifikáns növekedés a magyarázóerőben. E második modellben mind a REKLÁM, mind a YOUTUBE szignifikáns, pozitív hatással van a KONCERTJEGY értékére.				

Többváltozós regresszió publikálása

- Hivatkozott cikkek:
 - Tukey, J. W. (1977). Exploratory data analysis.
 - Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. Journal of the American Statistical Association, 82(400), 1147-1149.
 - Barnett, V., & Lewis, T. (1994). Outliers in statistical data (Vol. 3). New York: Wiley.
 - Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression.
 - Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. The American Statistician, 32(1), 17-22.
 - Stevens, J. P. (2012). Applied multivariate statistics for the social sciences. Routledge.
 - Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. Biometrika, 159-177.
 - Ghasemi, A., & Zahedi-Asl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486.
 - Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. American Psychologist, 60(2), 170.