

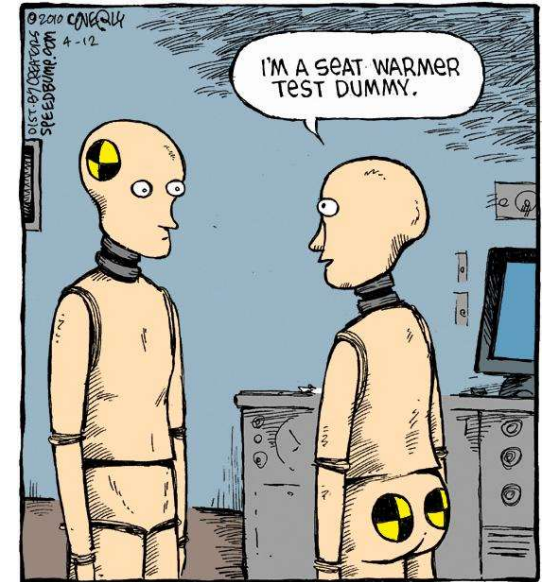
Többváltozós statisztika elméleti alapjai (BMNPS07500M)

Készítette: Soltész-Várhelyi Klára

Összetettebb regressziós modellek

2. Dummy változók, görbe illesztés, moderáló, mediáló hatások

Kategoriális változók regresszióban

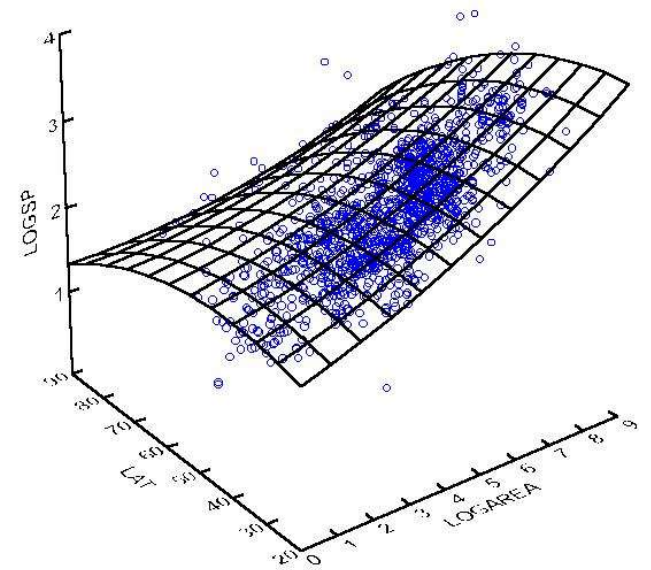


Feltételek

- Lineáris regresszió feltétele a **változók típusára**
 - **Prediktor változó: skála típusú vagy dichotóm**
 - **Kimeneti változó: skála típusú, folytonos**
- Hogy tehető be egy kategoriális prediktort?
 - **Dummy változók** segítségével
 - A kategoriális változó felbontása több dichotóm változóra
 - Egy N kategóriájú nominális változó felbontható N db dichotóm változóra
 - Az utolsó dummy sorsa
 - Redundáns információt tartalmaz, mert ha az összes előző 0, akkor ő 1
 - Az utolsó dummy-t ezért kihagyjuk az elemzésekből (nem is kell kiszámolni)

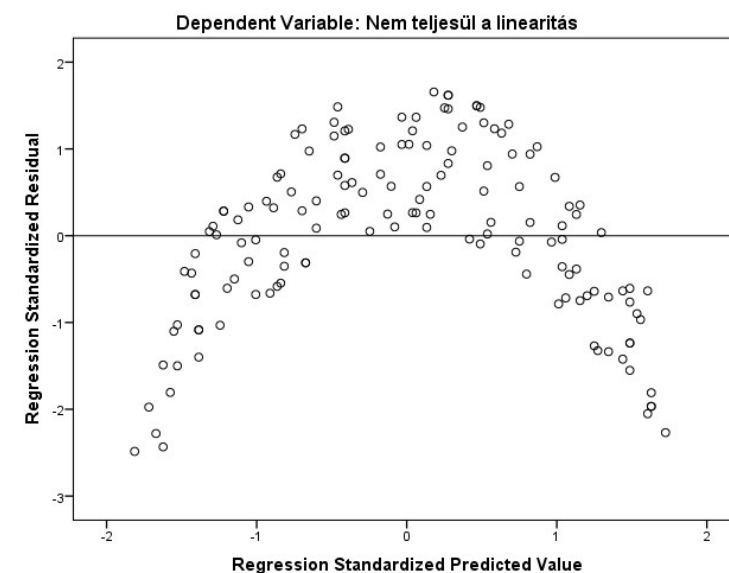
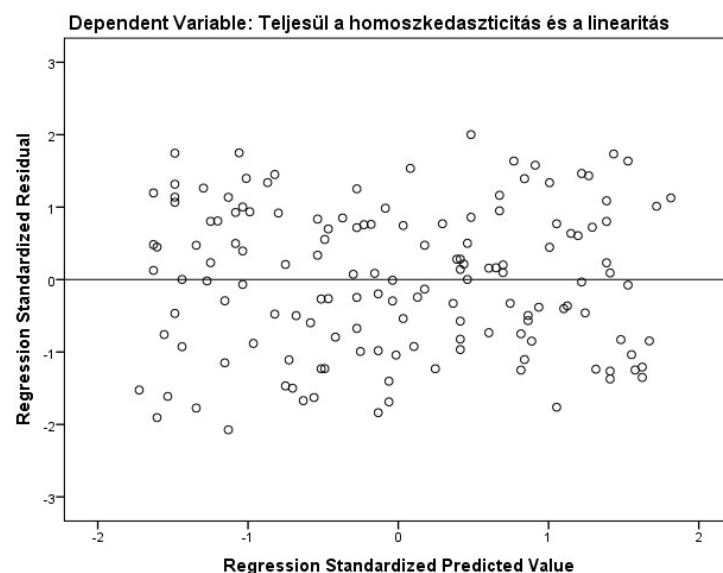
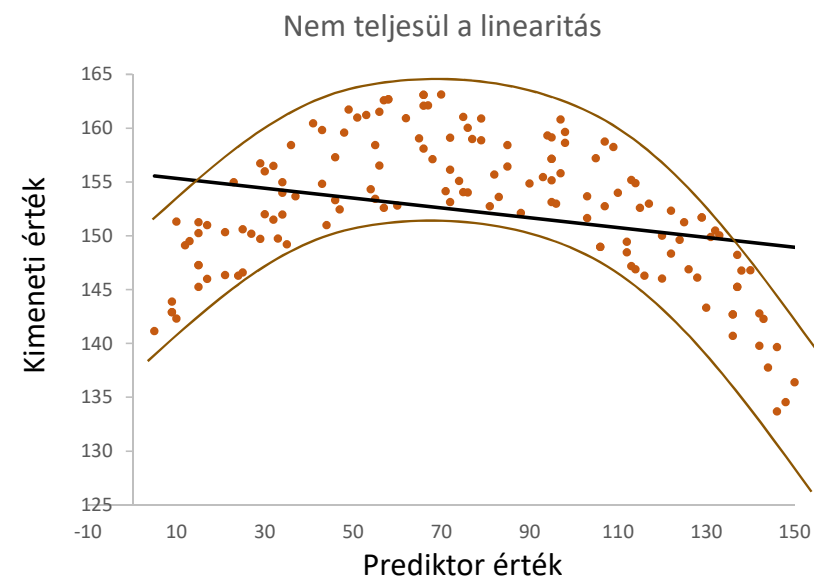
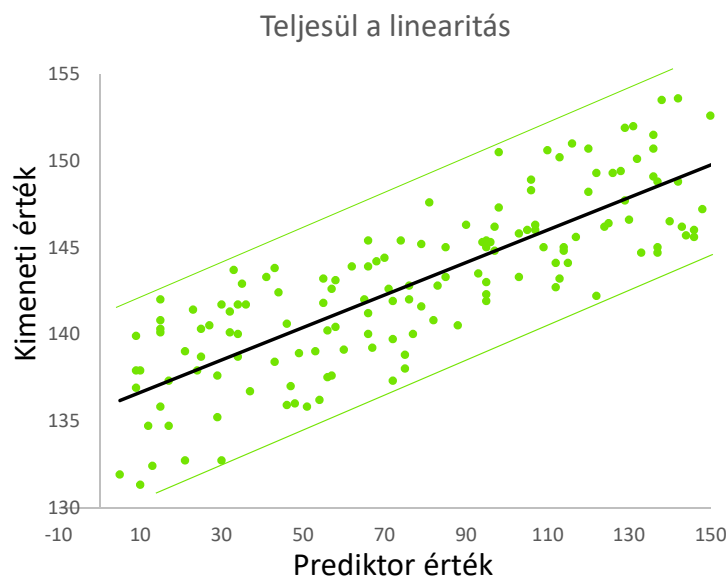
stílus	rock	metal	pop
1 (rock)	1	0	0
2 (metal)	0	1	0
3 (pop)	0	0	1
3	0	0	1
2	0	1	0
1	1	0	0
2	0	1	0
3	0	0	1
3	0	0	1
2	0	1	0
2	0	1	0
1	1	0	0

Görbe illesztés



Lineáris regresszióban feltétel volt a lineáris kapcsolat a prediktor változók és kimeneti változó között. Ezt kétféleképpen ellenőriztük:

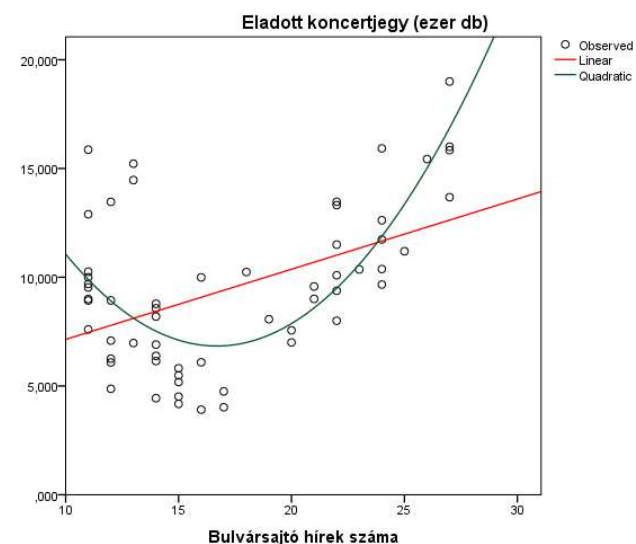
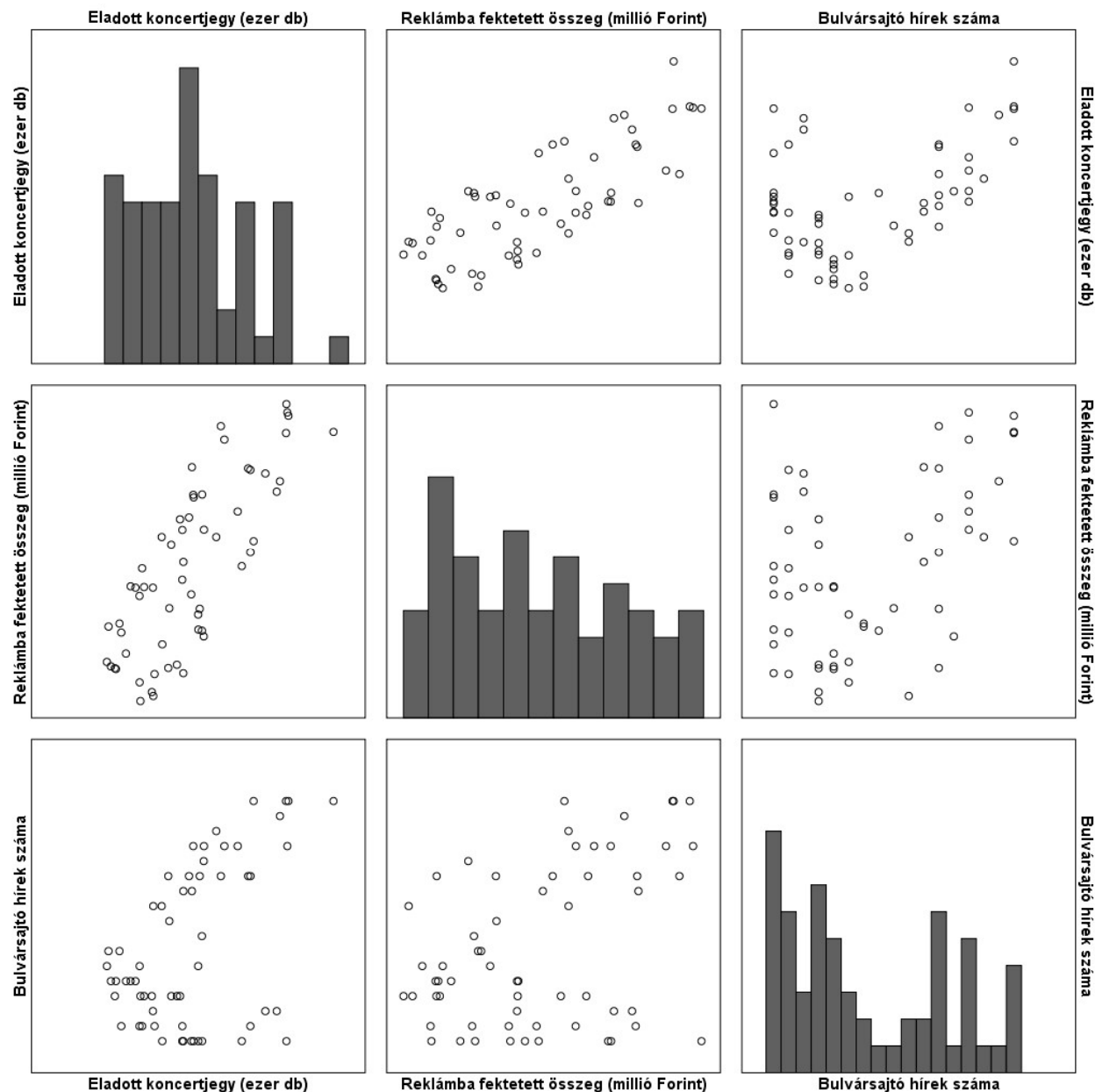
- 1) A prediktor változó és kimeneti változó kapcsolatának plottolásával
- 2) A predikált értékek (regressziós egyenes) és a reziduális hibák (a regressziós egyenestől való eltérések) plottolásával



Linearitás hiánya

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

Görbeillesztéssel látható (most tanuljuk majd meg), hogy a koncertjegy eladás variációját 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)

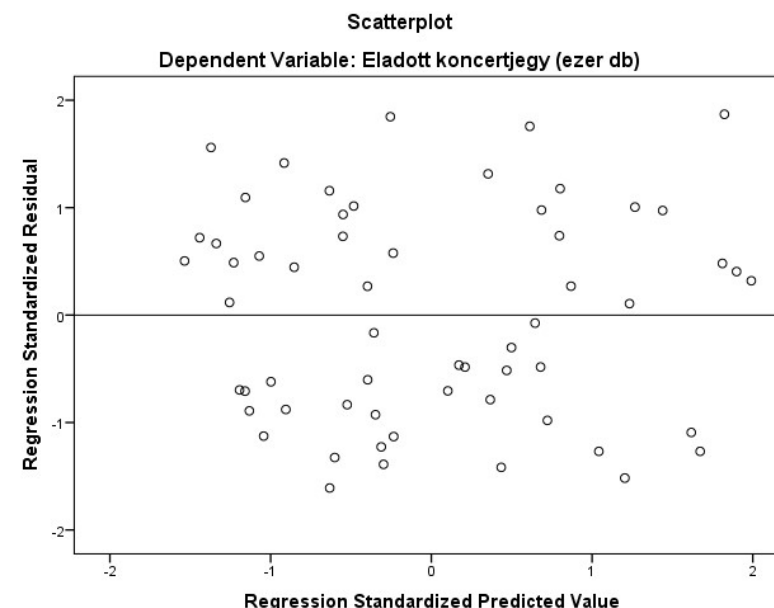
ANOVA^a

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!

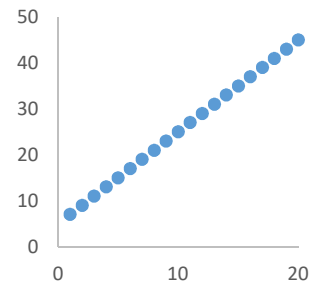


Görbék matematikai háttere

- Ahogy az egyenesnek is ismerjük a paramétereit, és konkrét megvalósulását néhány paraméter (b_0 és b_1 értékek) beállításával meg lehet határozni, ugyanúgy meghatározhatóak paraméterek segítségével a különböző görbületet tartalmazó modellek is.

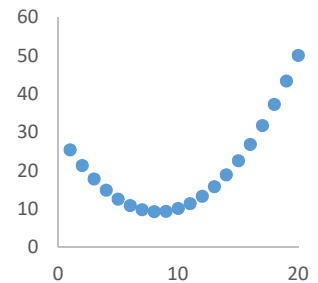
- Lineáris modell**

- Legegyszerűbb, elsőrendű modell
- Alakja az egyenes
- $Y_{\text{pred},i} = b_0 + b_1 * X_i$



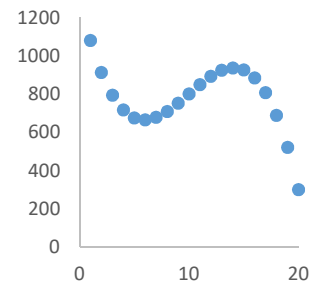
- Kvadratis model**

- Másodrendű modell, a lineáris tagon kívül tartalmaz egy négyzetes tagot is
- Alakja az egy görbülettel rendelkező (U alakú) görbe
- $Y_{\text{pred},i} = b_0 + b_1 * X_i + b_2 * X_i^2$



- Kubikus modell**

- Harmadrendű modell, tartalmaz egy köbös tagot is
- Alakja a két görbülettel rendelkező görbe
- $Y_{\text{pred},i} = b_0 + b_1 * X_i + b_2 * X_i^2 + b_3 * X_i^3$

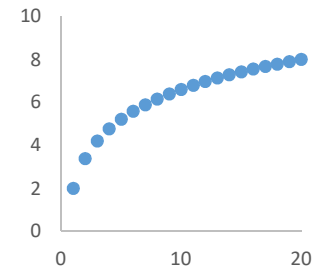


Görbék matematikai háttere

- Ahogy az egyenesnek is ismerjük a paramétereit, és konkrét megvalósulását néhány paraméter (b_0 és b_1 értékek) beállításával meg lehet határozni, ugyanúgy meghatározhatóak paraméterek segítségével a különböző görbületet tartalmazó modellek is.

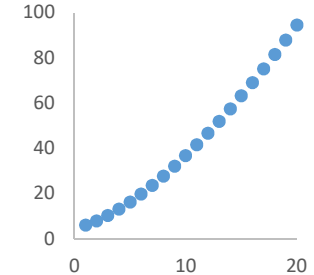
- Logaritmikus modell**

- A prediktor változó tizesalapú logaritmusával számolt regresszió
- Alakja az egyre csillapodón növekedő görbe
- $Y_{\text{pred},i} = b_0 + b_1 * \ln(X_i)$



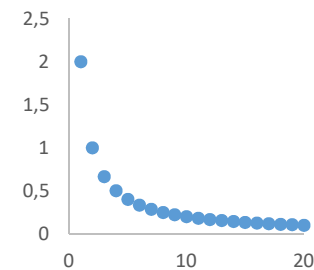
- Hatványozott modell (power)**

- A prediktor változó valamely hatványával számolt regresszió
- Alakja az egyre gyorsabban növekedő görbe
- $Y_{\text{pred},i} = b_0 + X_i^{b_1}$



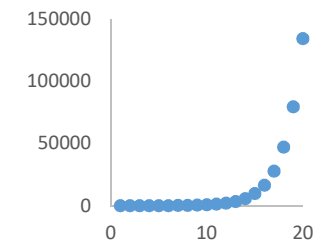
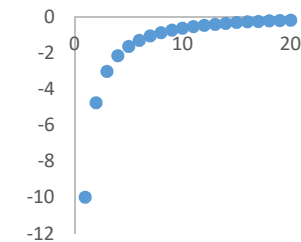
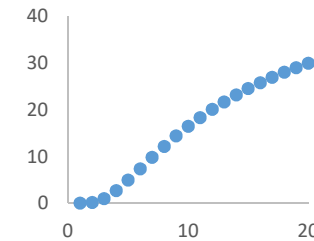
- Inverz modell**

- A prediktor változó inverzével számolt regresszió
- Alakja az eleinte gyorsan, majd egyre csillapodva csökkenő görbe
- $Y_{\text{pred},i} = b_0 + b_1 * 1 / X_i$



Görbék matematikai háttere

- Ahogy az egyenesnek is ismerjük a paramétereit, és konkrét megvalósulását néhány paraméter (b_0 és b_1 értékek) beállításával meg lehet határozni, ugyanúgy meghatározhatóak paraméterek segítségével a különböző görbületet tartalmazó modellek is.
- Sigmoid modell**
 - S-alakú göbe
- Logisztikus modell**
 - Egy plafonértékhez futó görbe
- Összetett modell**
- Bármilyen tetszőleges modell**
 - A prediktor változókon végzett műveletekkel és azok parametrizálásával bármilyen görbe kialakítható



Modellek bonyolultsága

- **Alulillesztés**

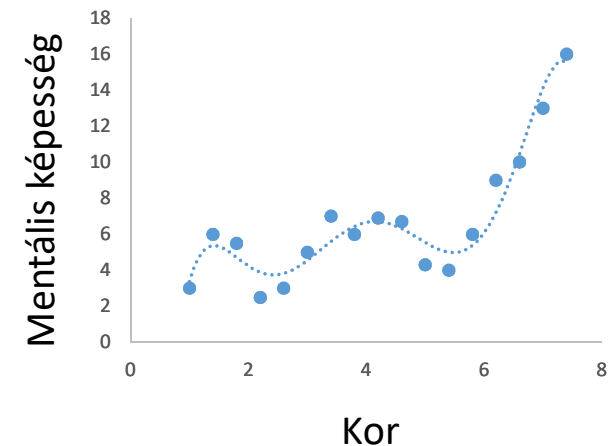
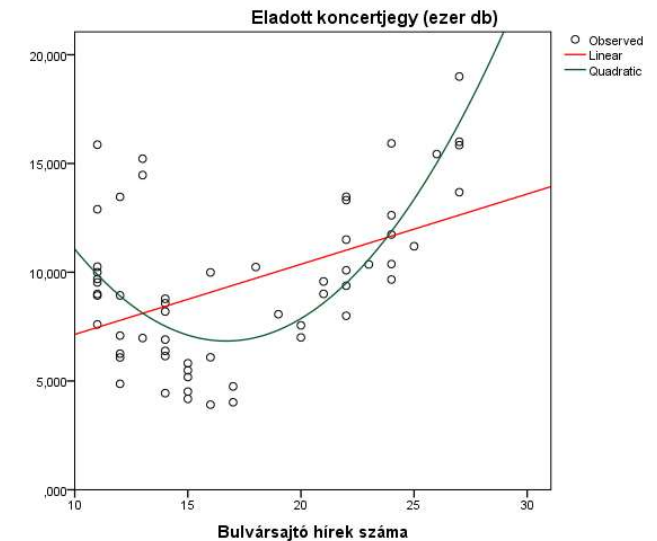
- Az alulillesztett modell nem képes a változók kapcsolatát megragadni

- **Túlillesztés**

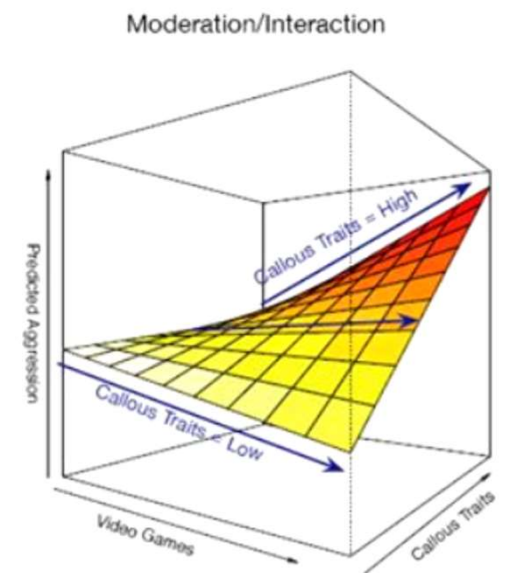
- A túlillesztett modell a zajra is rátanul
- Kellő bonyolultságú modellel minden pontfelhő tökéletesen lekövethető

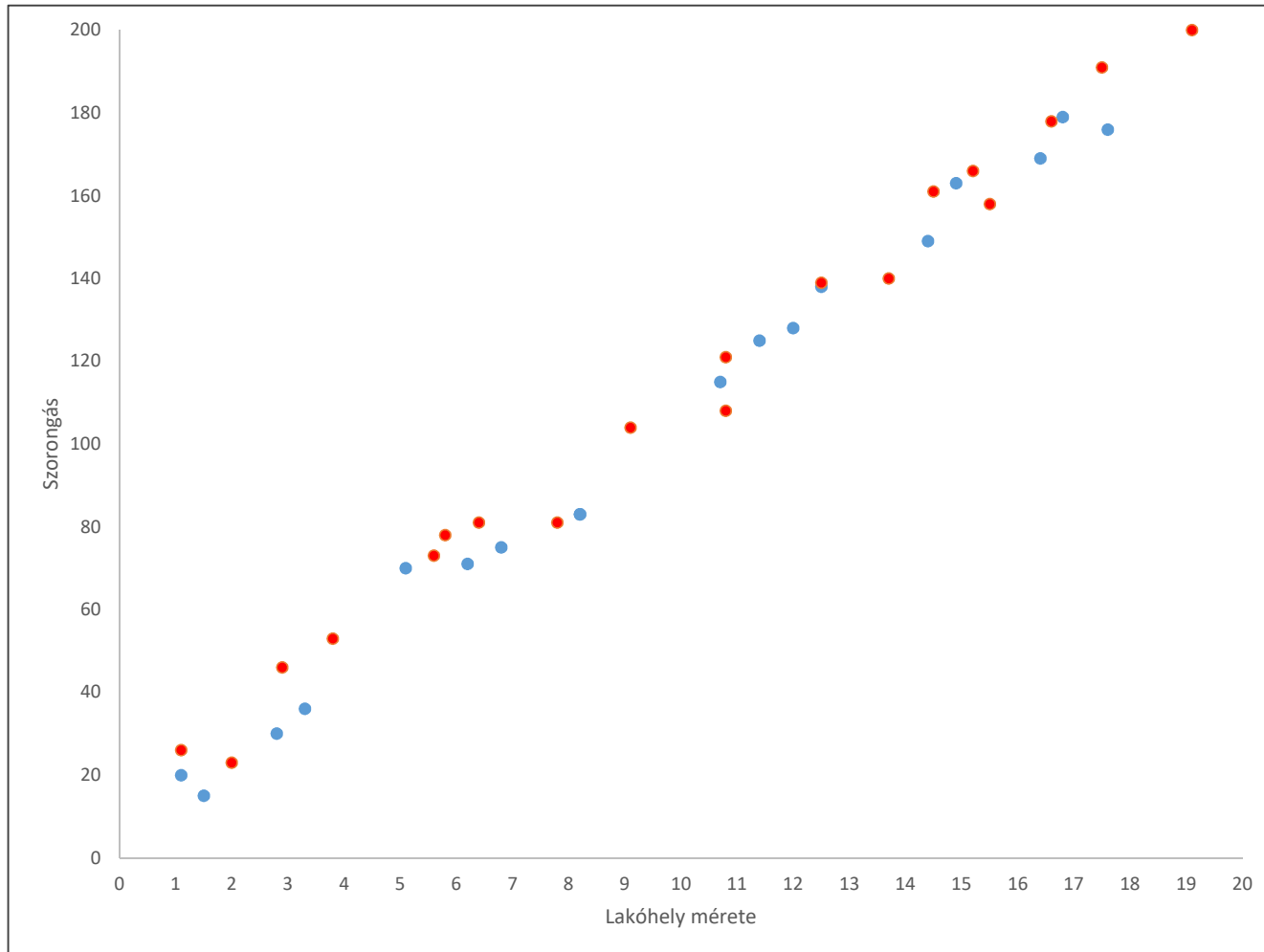
- **Melyik a jó bonyolultságú modell?**

- A legegyszerűbb modell az egyenes
- Occam borotvája: preferáld az egyszerűbb modellt!
- Ellenőrizd, hogy az egyszerű modellel nem szalasztod-e el a hatást!
- Szakmailag magyarázható bonyolultságú modellt válassz!
- Teszteld, hogy az egyszerűbb modellnél a bonyolultabb szignifikánsan jobb magyarázó erővel bír-e. Ha nem, válaszd az egyszerűbbet!

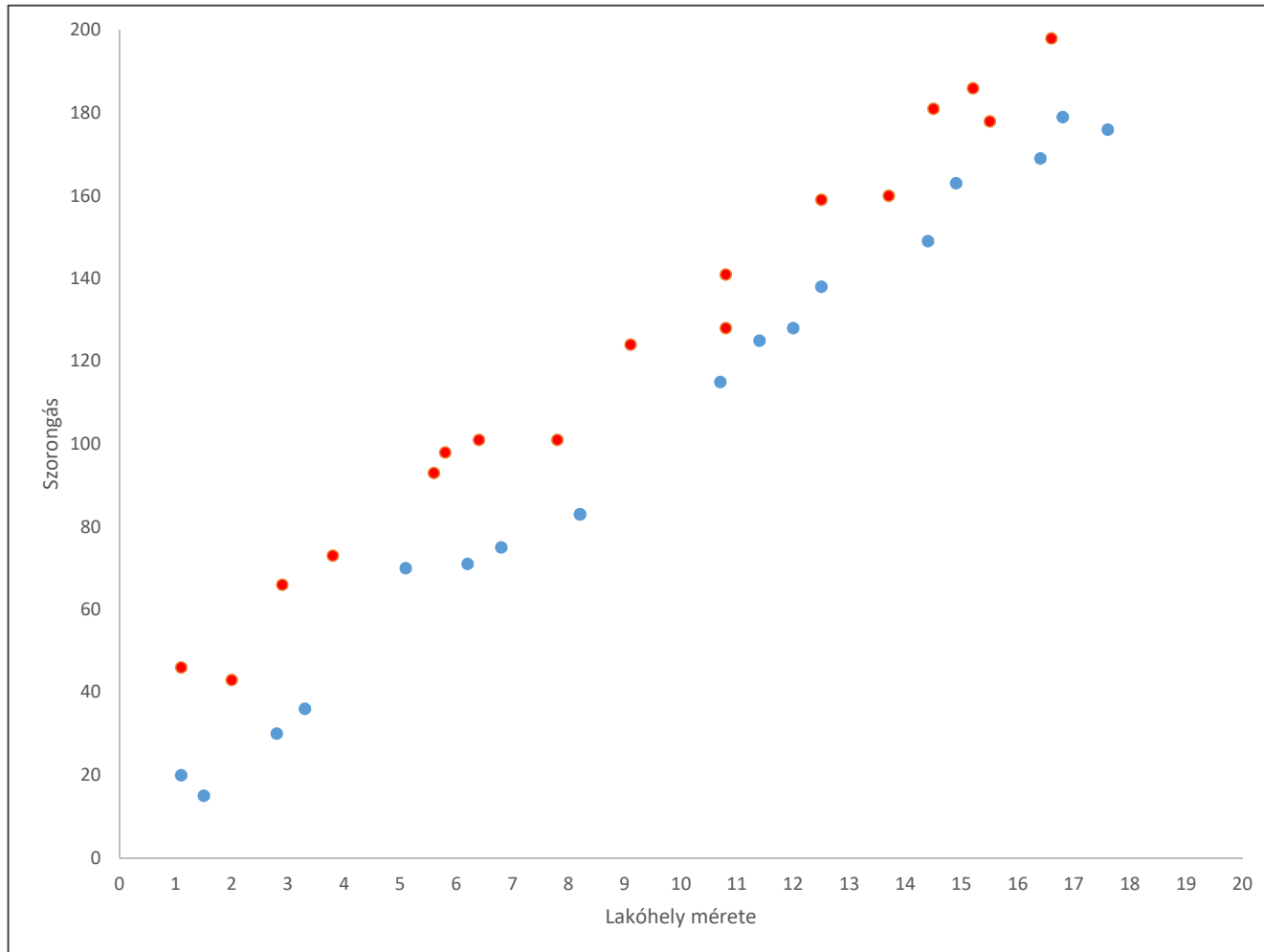


Moderáló hatás

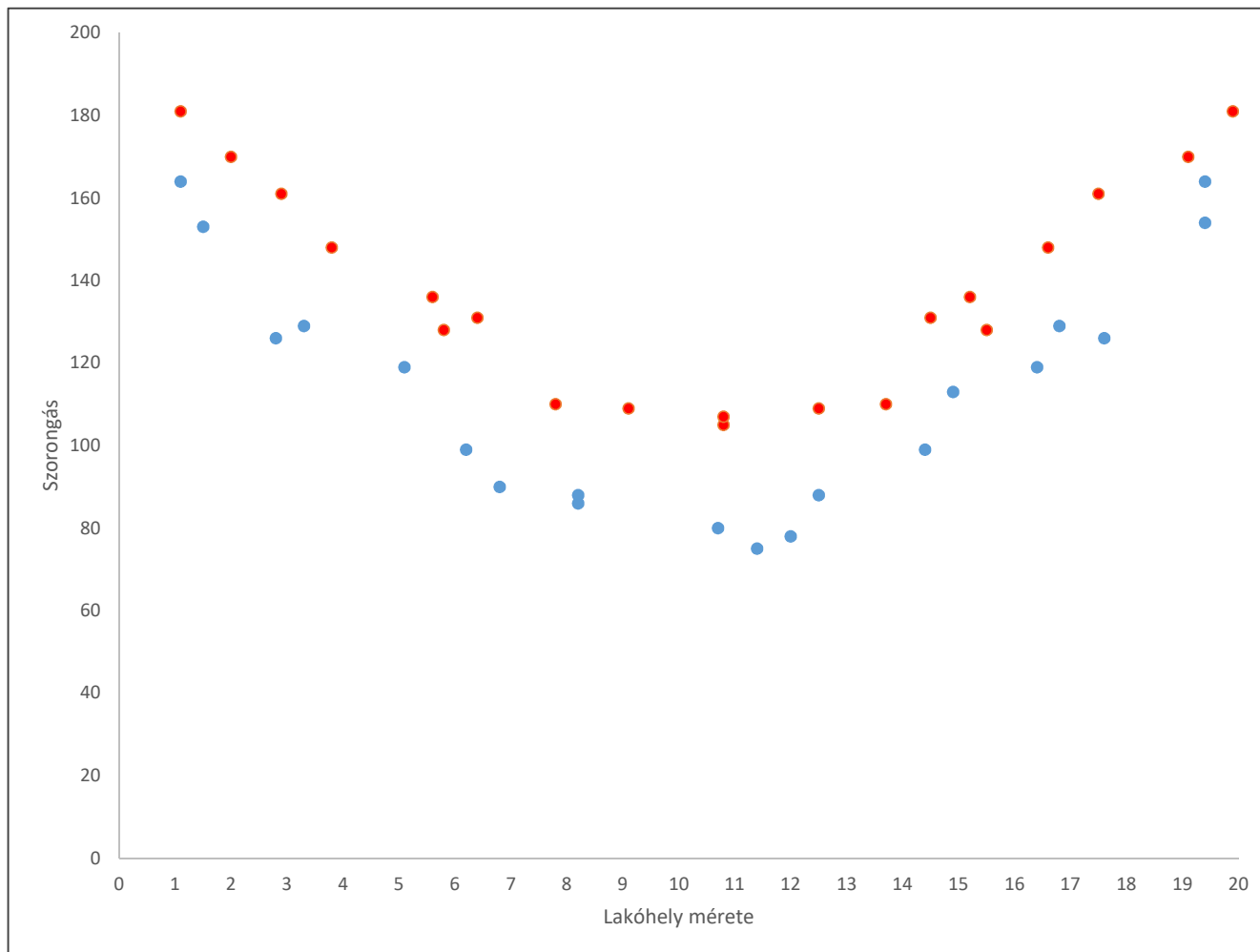




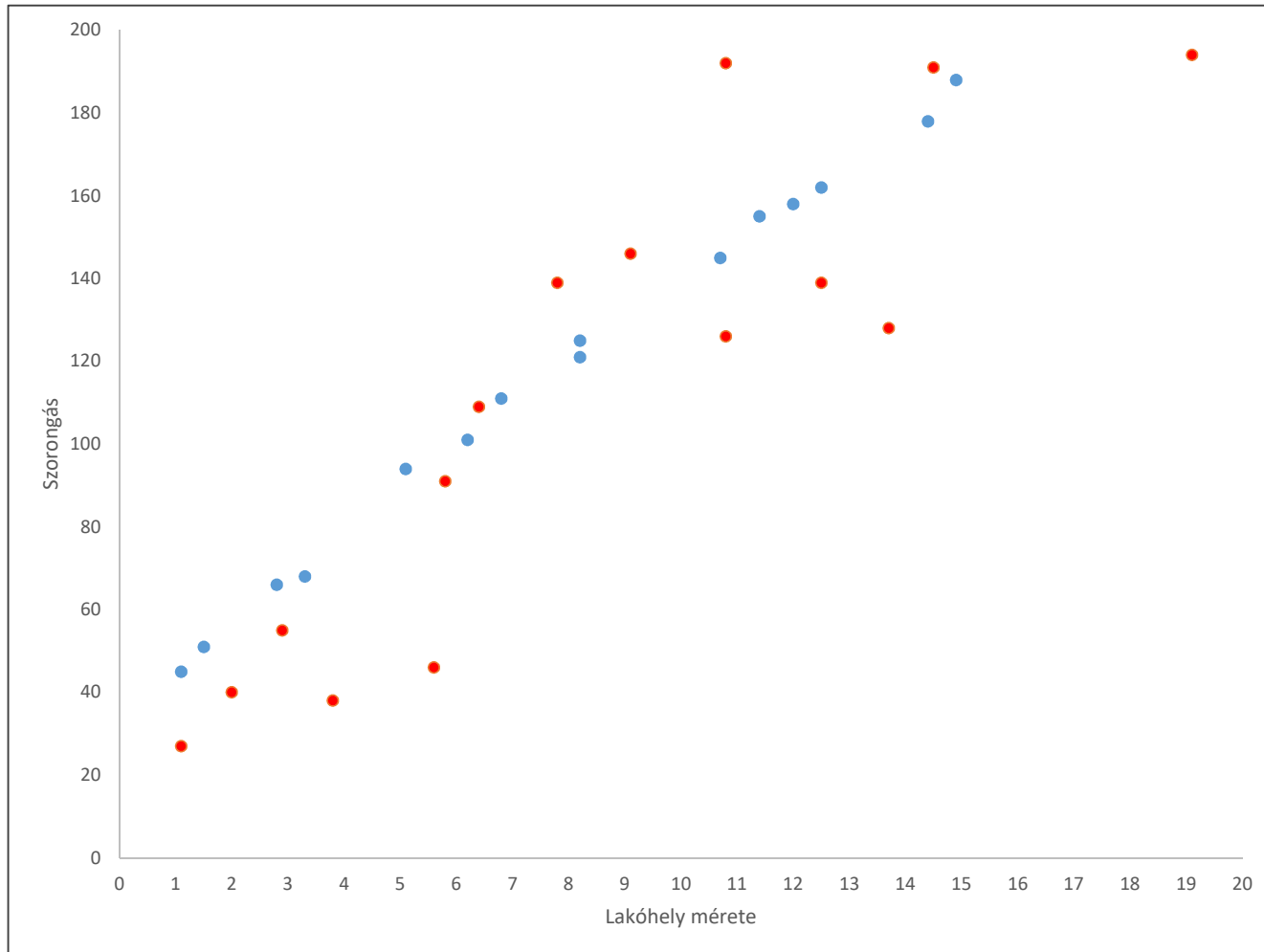
A férfi és női pontfelhő együtt, egymástól elkülöníthetetlenül fut, tehát a nemnek nincs hatása a szorongásra, és nem befolyásolja a lakóhely mérete és szorongás kapcsolatát sem.



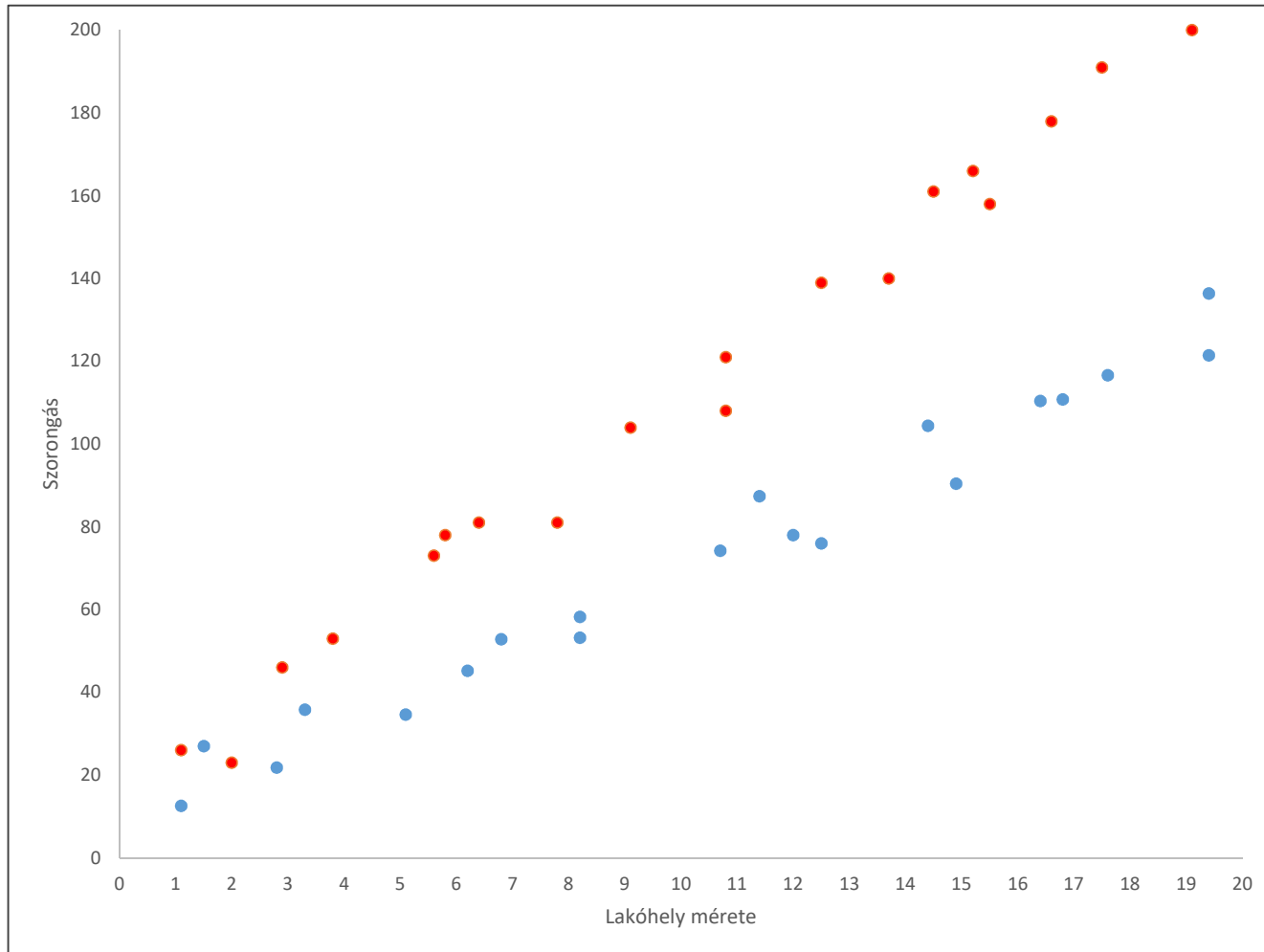
A férfi és női pontfelhő egymástól elkülönülten, de párhuzamosan fut, tehát a nemnek van hatása a szorongásra, a nők magasabb szorongásértéket vesznek fel, mint a férfiak, de nincs hatása a lakóhely mérete és szorongás közötti kapcsolatra, a lakóhely méretének növekedése ugyanannyira növeli férfiak és nők szorongását.



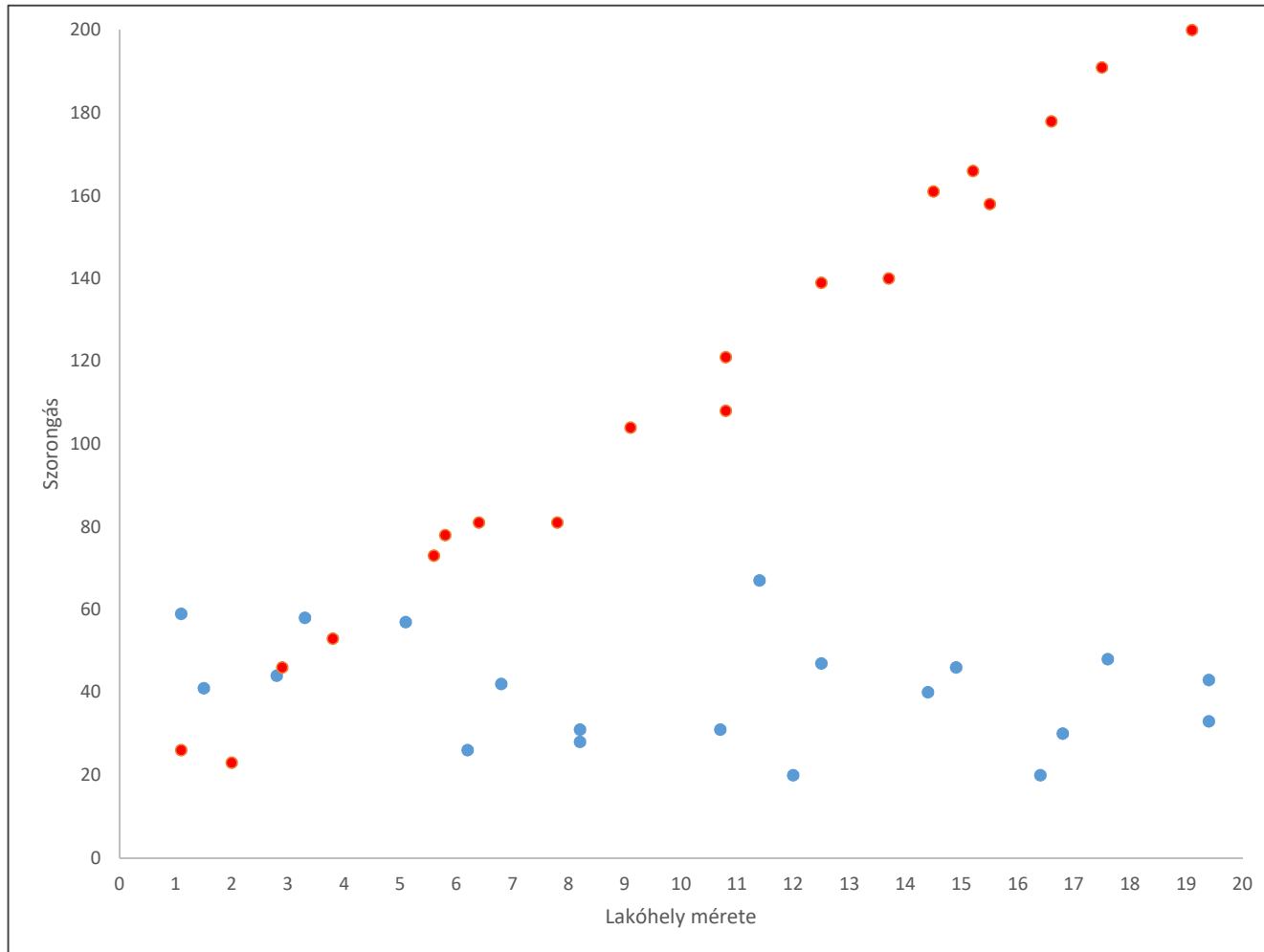
Mint az előbb is, a férfi és női pontfelhő egymástól elkülönülten, de egymással párhuzamosan fut, tehát a nemnek van hatása a szorongásra, de nem befolyásolja a lakóhely mérete és szorongás kapcsolatát. Annyi a különbség az előző ábrához képest, hogy itt a lakóhely és szorongás között nem lineáris a kapcsolat, de ez a moderáló hatás szempontjából nem lényeges.



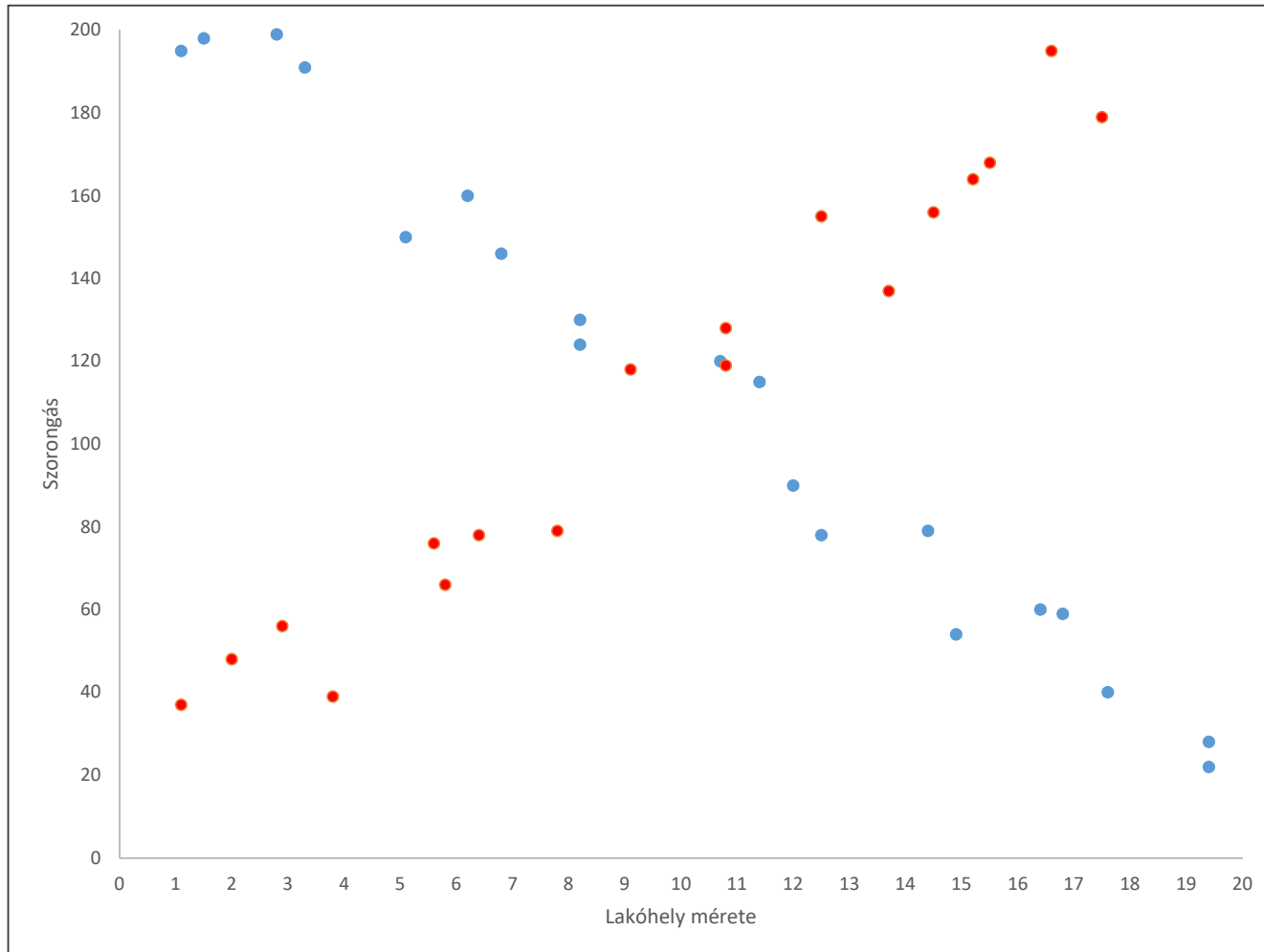
A nem befolyásolja a lakóhely mérete és szorongás kapcsolatát: a férfiak esetében sokkal szorosabb az összefüggés e két változó között, mint nők esetében. Ezt a fajta moderáló hatást nagyon nehéz megragadni, inkább csak a magyarázóerők összehasonlításával szokás.



Tipikus példa a moderáló hatásra. A nem befolyásolja a lakóhely mérete és szorongás közötti összefüggést, nők esetében a lakóhely méretének növekedése nagyobb mértékben növeli a szorongást, mint férfiak esetében. ANOVÁban ezt majd *ordinális interakciónak* nevezzük majd, mert a lakóhely azonos irányban, de eltérő mértékben hat mind a két nem szorongására.



Súlyosabb példa a moderáló hatásra. A nem befolyásolja a lakóhely mérete és szorongás közötti összefüggést: míg nők esetében a lakóhely méretének növekedésével nő a szorongás is, addig férfiak esetében a szorongás független tőle. Az ilyen moderáló hatás ANOVÁban majd valahol az ordinális és diszordinális interakció között fog elhelyezkedni.



A moderáló hatások
legsúlyosabb verziója, nem hogy
csak más mértékben, de eltérő
irányban is hat a lakóhely
méretének változása a
szorongásra férfiak és nők
esetében: Nők esetében a
lakóhely méretének növekedése
növeli, férfiak esetében
csökkenti a szorongást. Tipikus
példa lesz ANOVÁban a
*diszordinális interakció*ra.

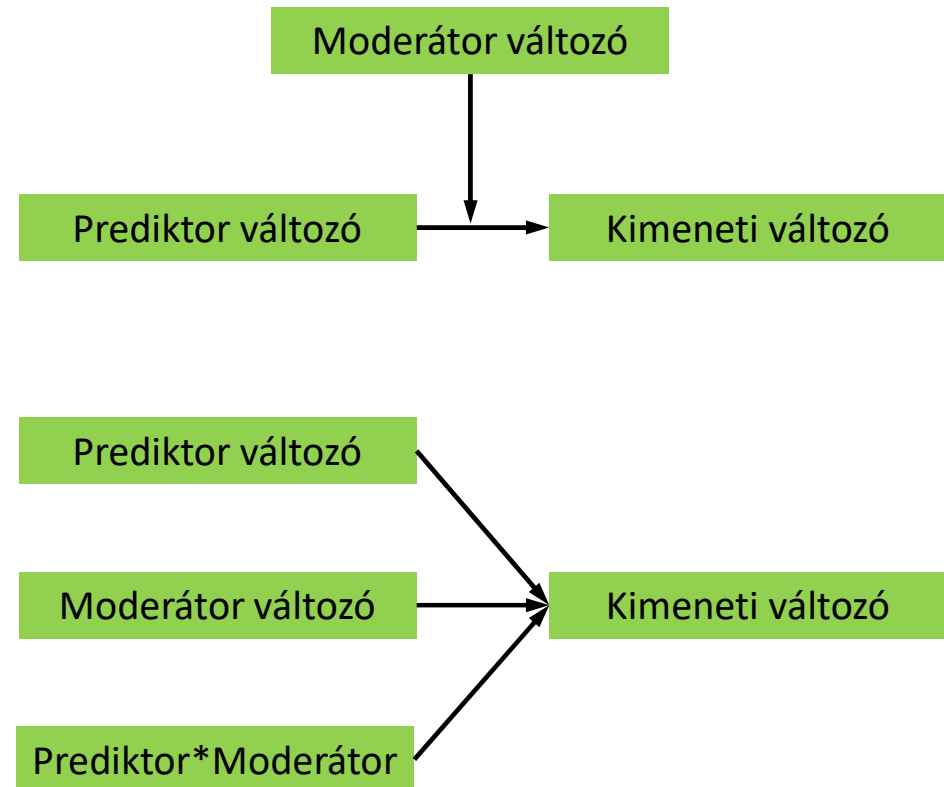
Moderáló hatás

- **Moderáló hatás:**

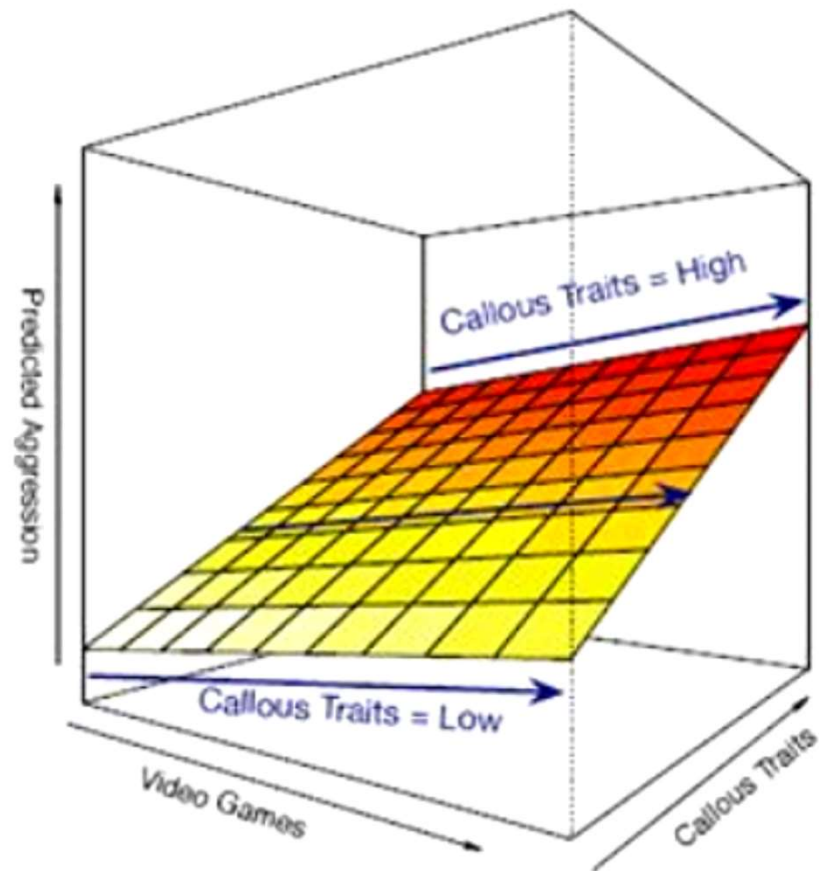
- A prediktor változó kimeneti változóra tett hatását befolyásolja egy harmadik, ún. moderáló változó
- A prediktor változó és a moderáló változó között interakció van

- **Interakció:**

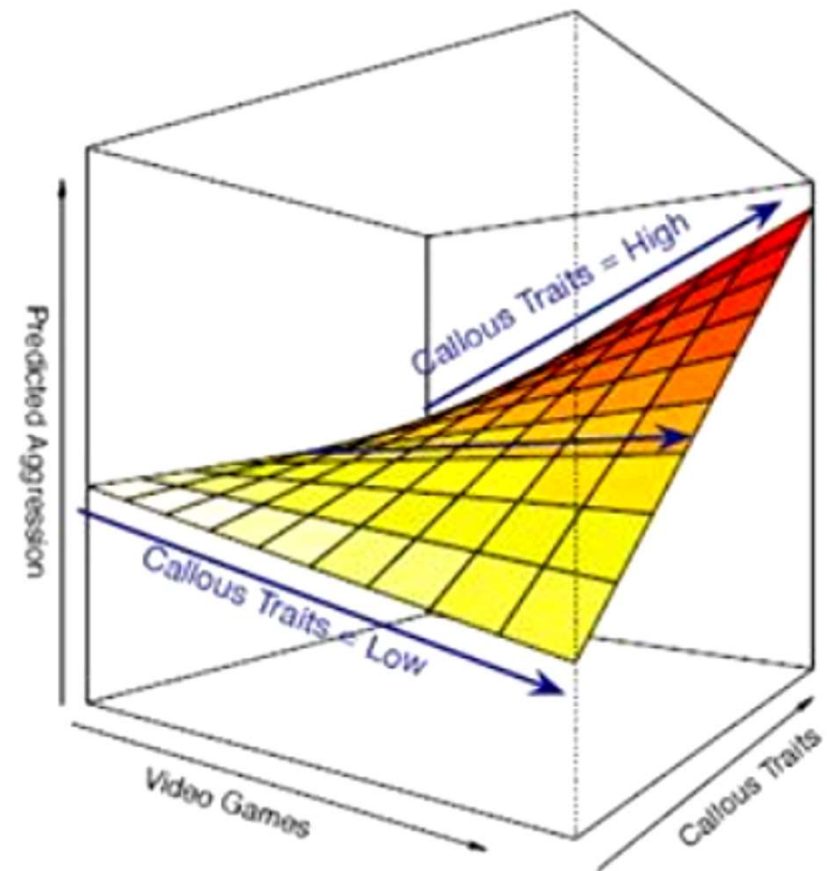
- Az egyik változó különböző szintjein nem ugyanolyan a másik változó hatása a kimeneti változóra
- Matematikailag az interakció a két interakcióban résztvevő prediktor változó szorzatával egyenlő
- $Y_{pred,i} = b_0 + b_1 * X1_i + b_2 * X2_i + b_3 * X1_i * X2_i$
- Megszokott jelölés $X1 \times X2$

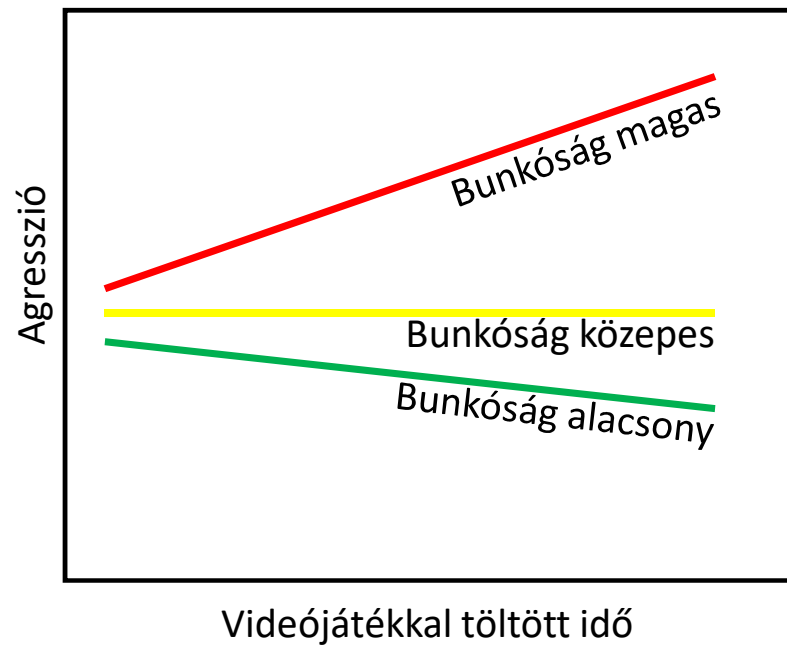


No Moderation/Interaction

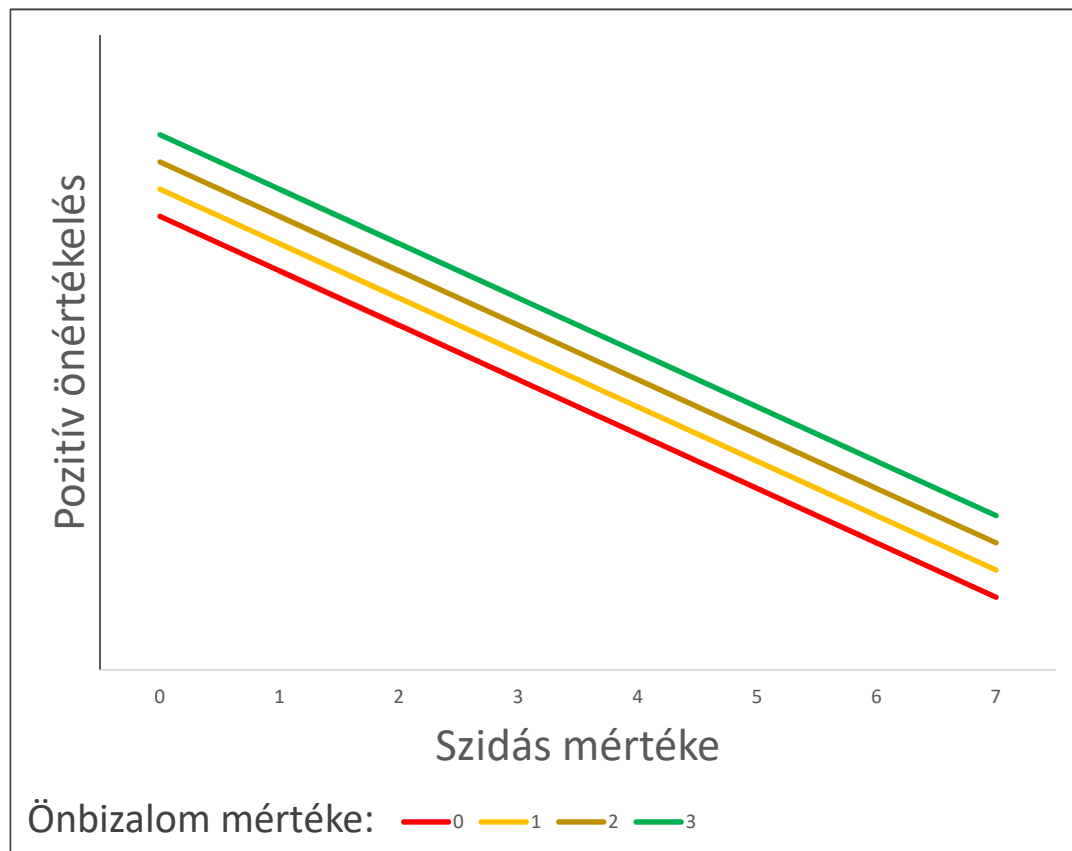


Moderation/Interaction





Az előző dián látható felület diagram látványos, azonban csak az egyszerűbb interakciók jeleníthetők meg rajta jól értelmezhető módon. Ezért sokszor a folytonos moderáló változót tartalmazó interakciót úgy jelenítjük meg (és fontos, hogy ez csak a megjelenítést érinti, nem magát az elemzést), hogy a folytonos moderáló változóból 3-4 csoportot képezünk (pl. itt nagyon alacsony, inkább alacsony, inkább magas, nagyon magas önbizalommal rendelkezőkre bontjuk a mintát), és a különböző csoportokat különböző színű regressziós egyenessel jelöljük.



$$Y_{\text{pred}} = -6 * SZID + 3 * \ddot{O}NB + 0 * SZID \times \ddot{O}NB$$

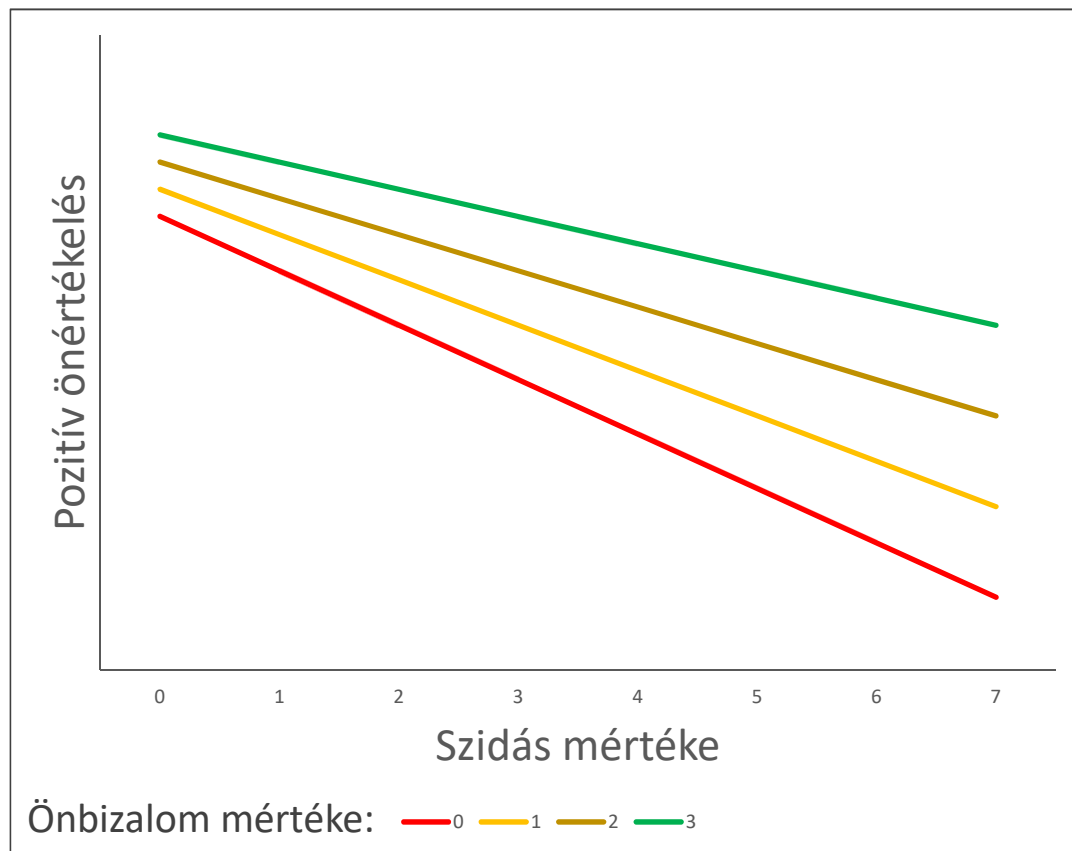
A következő grafikonokon a szidás mértéke és a pozitív önértékelés kapcsolatát vizsgáljuk, az önbizalmat moderáló változóként használva. Tehát azt nézzük meg, hogy az önbizalom befolyásolja-e, hogyan hat a szidás az önértékelésre.

A grafikon értelmezése:

A szidásnak hatása van az önértékelésre, minél több a szidás, annál alacsonyabb az önértékelés (a vonalak csökkenő tendenciát mutatnak)

Bár számunkra most nem ez érdekes, de az önbizalomnak is hatása van az önértékelésre, a magasabb önértékelésű csoportok (barna és zöld) vonalai magasabban futnak

Nincs moderáló hatás, a különböző önértékelésű csoportokban ugyanúgy hat a szidás az önértékelésre (mind a négy vonal párhuzamos, azaz azonos mértékben csökken)



A grafikon értelmezése:

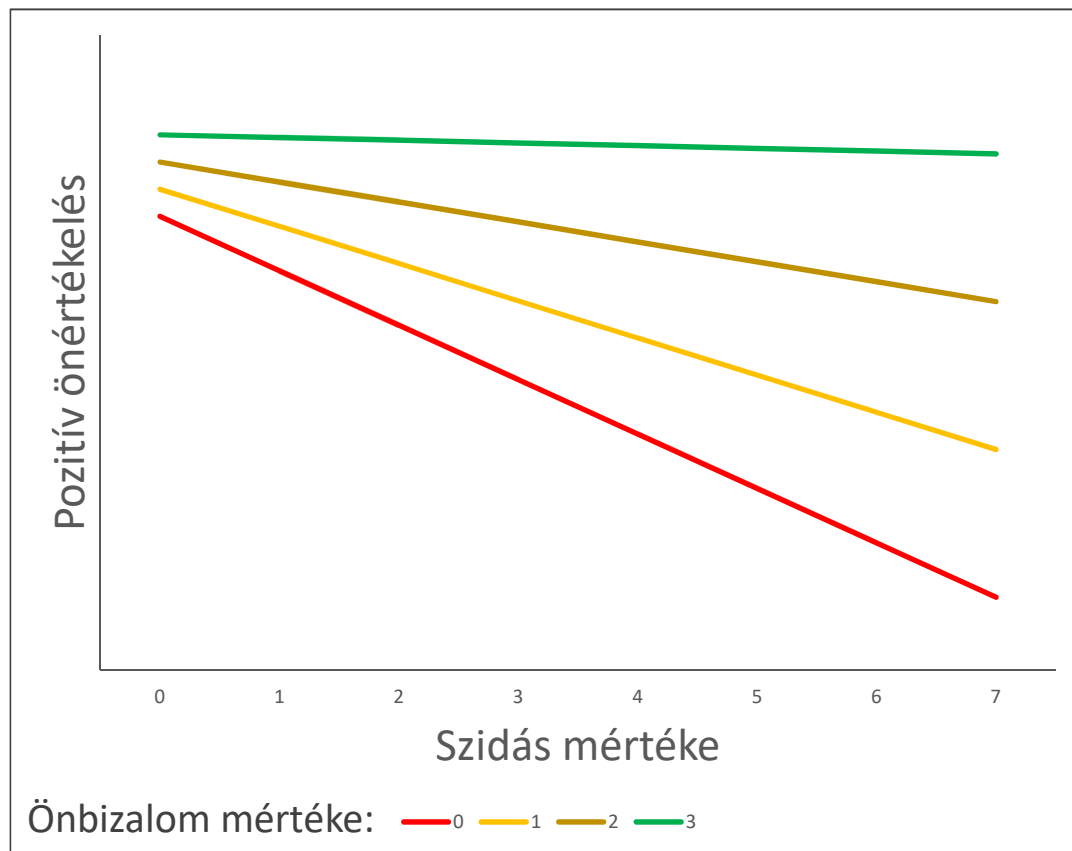
A szidásnak hatása van az önértékelésre, minél több a szidás, annál alacsonyabb az önértékelés (a vonalak csökkenő tendenciát mutatnak)

Bár számunkra most nem ez érdekes, de az önbizalomnak is hatása van az önértékelésre, a magasabb önértékelésű csoportok (barna és zöld) vonalai magasabban futnak

Van moderáló hatás: minél magasabb az önértékelés, annál kevésbé rontja a szidás az önértékelést (például a zöld, magas önbizalmú csoportban kevésbé meredeken csökken az egyenes, mint a piros, azaz alacsony önbizalmú csoportban).

ANOVA-ban ezt nevezzük majd ordinális interakciónak.

$$Y_{\text{pred}} = -6 * SZID + 3 * \ddot{O}NB + 1 * SZID \times \ddot{O}NB$$



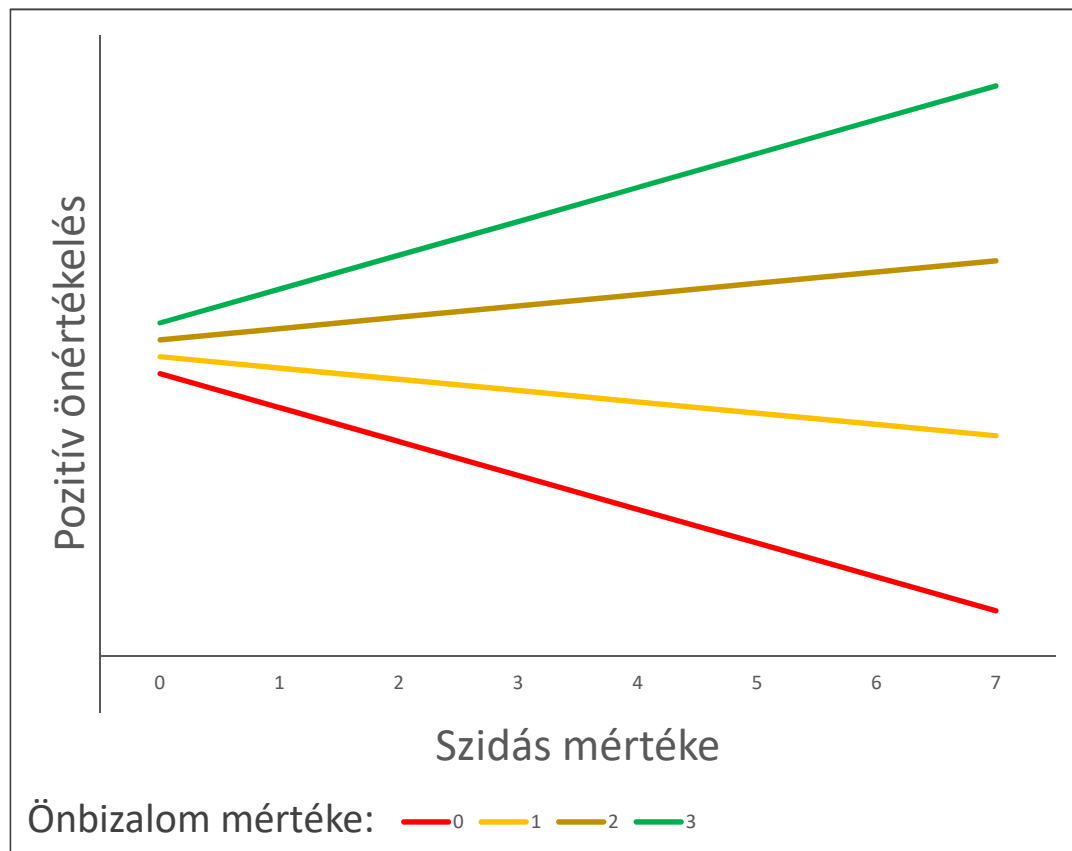
$$Y_{\text{pred}} = -6 * SZID + 3 * \ddot{O}NB + 1,9 * SZID \times \ddot{O}NB$$

A grafikon értelmezése:

Itt már nem jelenthető ki egyértelműen, hogy van-e hatása a szidásnak az önértékelésre, mert a zöld, azaz nagyon magas önbizalmú csoportban nincs.

Bár számunkra most nem ez érdekes, de az önbizalomnak is hatása van az önértékelésre, a magasabb önértékelésű csoportok (barna és zöld) vonalai magasabban futnak

Az előzőnél erősebb moderáló hatást látunk: a magas önbizalom nem hogy csak mérsékeli, de akár meg is szüntetheti a szidás önértékelésre tett negatív hatását.



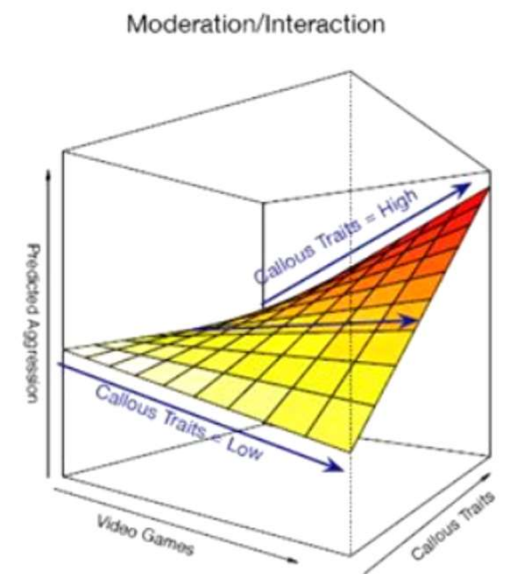
A grafikon értelmezése:

Ezek a legizgalmasabb interakciók! Az önbizalom teljesen megváltoztatja a szidás önértékelésre gyakorolt hatásának természetét. Bár e jelen példa nem tűnik túl életszerűnek, a fejlődépszichológiában gyakran találkozunk ilyen típusú moderációkkal, amikor azt látjuk, hogy a gyerek valamely tulajdonságától (például temperamentumától) függ, hogy a szülői mintával vagy annak ellentétével azonosul.

ANOVÁban ezt nevezzük majd diszordinális interakciónak.

$$Y_{\text{pred}} = -6 * SZID + 3 * \ddot{O}NB + 4 * SZID \times \ddot{O}NB$$

Mediáló hatás

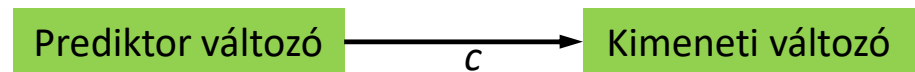


Mediáló hatás

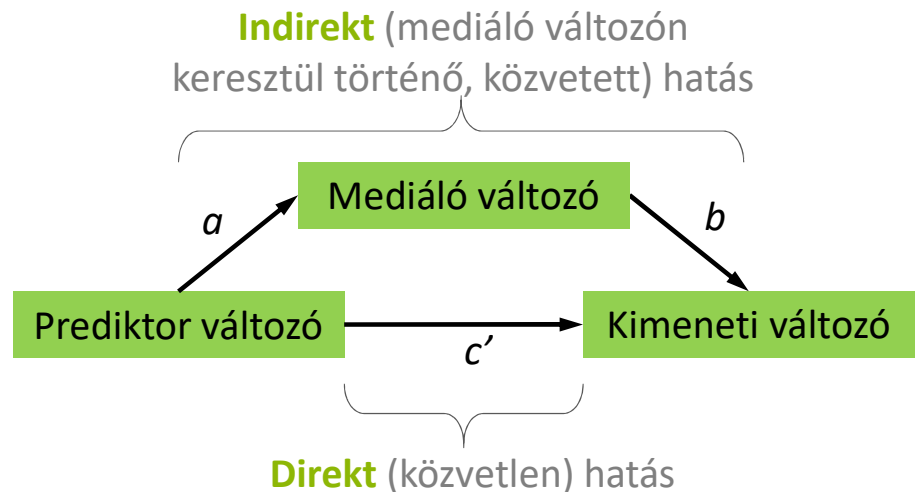
- **Mediáló hatás:**

- Mediáló hatásról akkor beszélünk, ha a prediktor és kimeneti változó közötti összefüggést meg lehet magyarázni egy harmadik változóval való kapcsolatukkal.
- A munkahelyi hangulat és kiégés kapcsolatát mediálja az átélt stressz.
- Ekkor a prediktor és kimeneti változó közötti összefüggés két részre bomlik:
 - **Direkt hatás**
 - A prediktor kimeneti változóra tett hatása a mediáló tényezőre kontrollálva
 - **Indirekt hatás**
 - A prediktor változónak a mediáló változón keresztül történő hatása a kimeneti változóra

Egyszerű lineáris regresszió prediktor és kimeneti változó között



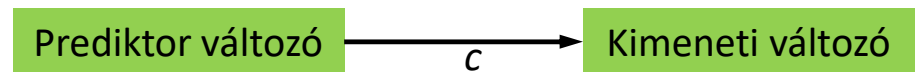
Mediációt tartalmazó regressziós modell



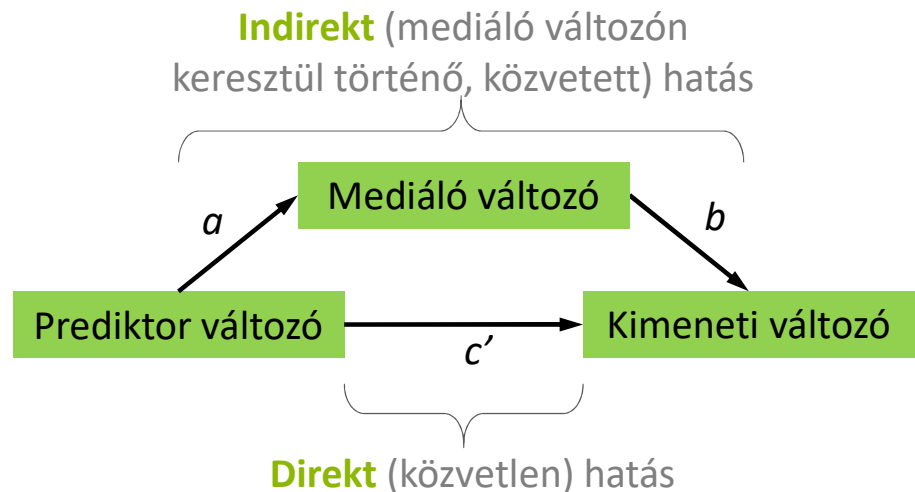
Mediáló hatás

- **A mediáló hatás feltételei** (Baron & Kenny, 1986) alapján:
 - A prediktor változó magyarázza a kimeneti változót (ábrán c)
 - A prediktor változó magyarázza a mediáló változót (ábrán a)
 - A prediktor és mediáló változó magyarázza a kimeneti változót (prediktor útja a c' és mediáló változó útja a b)
 - A prediktor változó magyarázóereje a második modellben alacsonyabb, mint az első modellben. Tehát a prediktor változó kapcsolata a kimeneti változóval részben az indirekt, mediáló változón keresztül úton valósul meg. Ha az összefüggés teljes egészében a mediáló változónak köszönhető, akkor a direkt hatás β -értéke (standardizált meredeksége) 0-ra csökkenne.

Egyszerű lineáris regresszió prediktor és kimeneti változó között



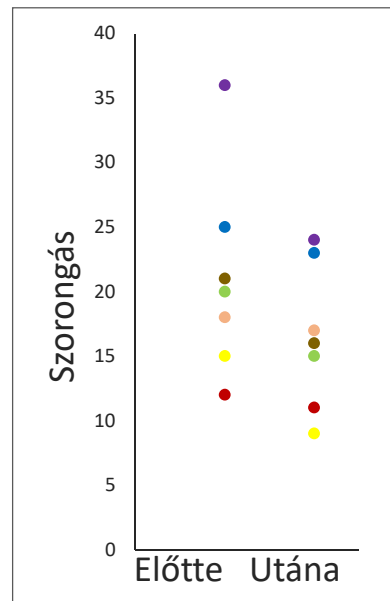
Mediációt tartalmazó regressziós modell



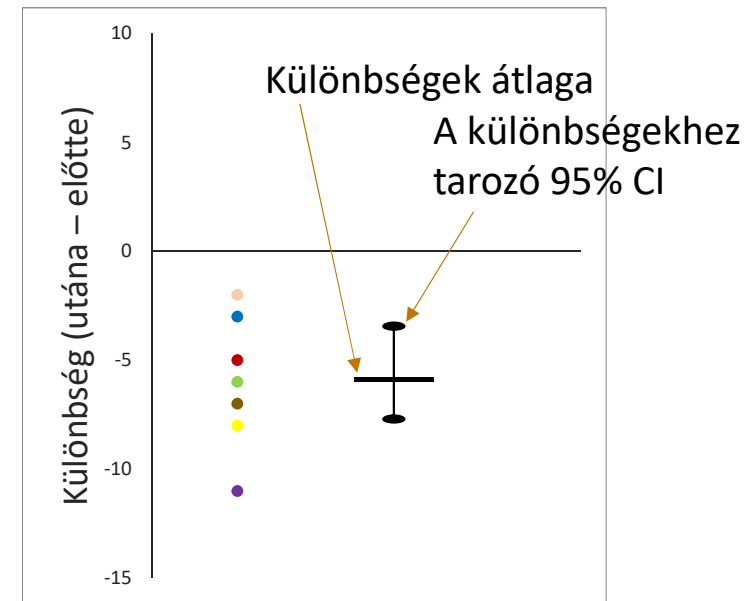
Konfidencia intervallumok olvasása

- Az eredmények értelmezésekor szignifikancia érték helyett csak a konfidencia intervallumokat kapjuk meg – ezért át kell néznünk az értelmezésüket. Példa: *Hatásos-e a szorongást csökkentő terápia?*
- Ha a terápiának volt hatása, akkor az emberek szorongása csökken, tehát a két mérés különbsége negatív szám lesz a legtöbb ember esetében. Ha különbségek átlaga negatív szám lesz, és a különbségekhez tartozó 95%-os konfidencia intervallum alsó és felső hatása a negatív tartományban lesz, akkor legalább 95%-os valószínűséggel kijelenthetjük, hogy a különbségek populáció átlaga egy negatív szám, tehát a két szorongásérték szignifikánsan különbözik egymástól. Ellenőrizve: $t(6) = 5.196$ $p = .002$

Szorongás terápia előtt	Szorongás terápia után
19	13
11	6
36	25
18	16
15	7
21	14
25	22



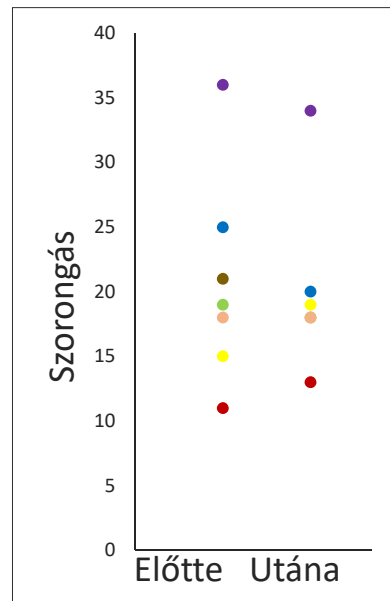
Különbség (utána – előtte)
-6
-5
-11
-2
-8
-7
-3



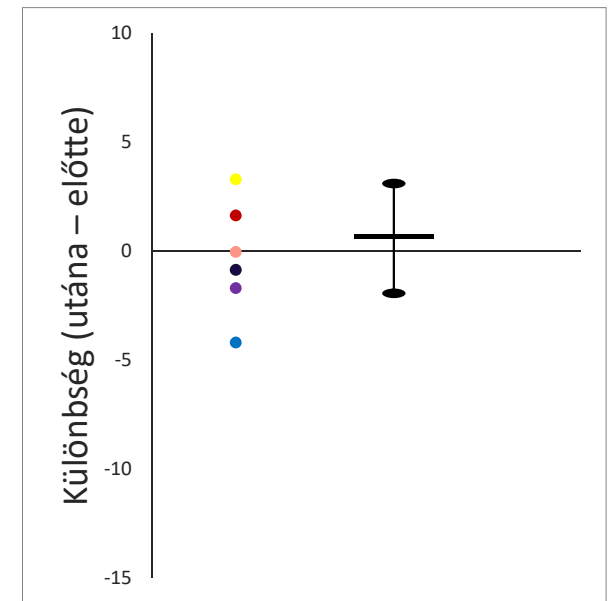
Konfidencia intervallumok olvasása

- Ha a terápiának nem volt hatása, akkor az emberek szorongása nem csökken számottevően. A (sikertelen) terápia előtti és utána mért szorongásértékek különbsége várhatóan valahol nulla körül lesz. A mintavételezésben lévő zaj miatt, lesz akinél a különbség pozitív lesz, és lesz olyan, akinél negatív.
- Ha a különbségekhez tartozó 95% konfidencia intervallum alsó határa a negatív tartományban van, a felső határa a pozitív tartományban (ergo a CI magában foglalja a nullát), akkor nem tudjuk kellő bizonyossággal kijelenteni, hogy a különbségekhez tartozó populációátlag pozitív vagy negatív, a szorongás előtti és utáni szorongás között nincs szignifikáns különbség. Ellenőrizve: $t(6) = 0.394$ $p = .707$

Szorongás terápia előtt	Szorongás terápia után
19	18
11	13
36	34
18	18
15	19
21	20
25	20



Különbség (utána – előtte)
-1
2
-2
0
4
-1
-5



β -értékekhez tartozó intervallum

- Ugyanez történik a β -értékek esetén is. Itt a β -érték 0 meredekségtől való eltérését nézzük (emlékeztető, ha a meredekség nulla, akkor a prediktornak nincs hatása a kimenetre).
- Az előzőekhez hasonlóan ábrázolható a β -érték és a 0 eltérése, és köré rajzolható a 95%-os konfidencia intervallum.
- **Ha az intervallum magában foglalja a nullát**, az azt jeleneti, hogy ha több mintán kiszámolnánk a meredekséget, várhatóan néha pozitív, néha negatív értéket kapnánk, tehát nem tudjuk kijelenteni, hogy a meredekség szignifikánsan eltér a nullától.
- **Ha a konfidencia intervallum alsó és felső határa is a pozitív tartományban van**, akkor a meredekség 95%-os valószínűséggel pozitív, tehát szignifikáns, pozitív hatása van a prediktornak.
- **Ha az intervallum alsó és felső határa is a negatív tartományban van**, akkor szignifikáns negatív hatása van a prediktornak.

