

stathelp.hu

Készítette: Soltész-Várhelyi Klára

Egyszerű lineáris regresszió elméleti háttere

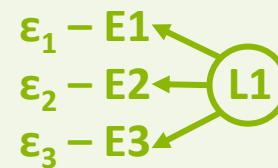
Varianciaanalízis
(ANOVA)



Főkomponens-
elemzés (PCA)



Faktorelemzés
(EFA és CFA)

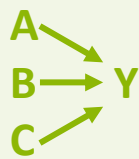


Egyszerű lineáris
regresszió

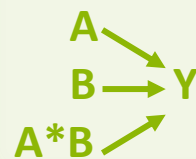


$A \rightarrow Y$

Többszörös lineáris
regresszió



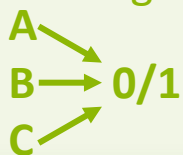
Moderáció



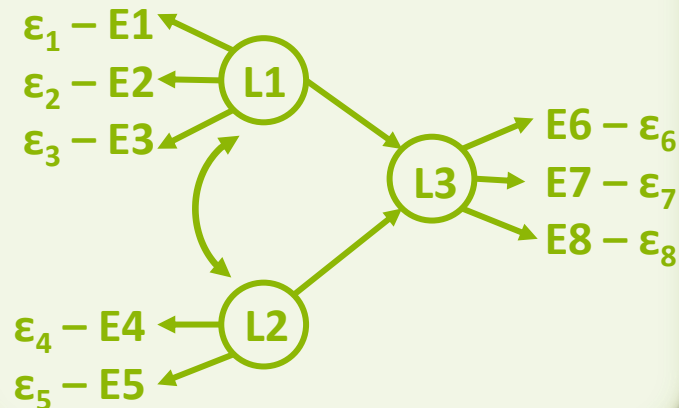
Mediáció



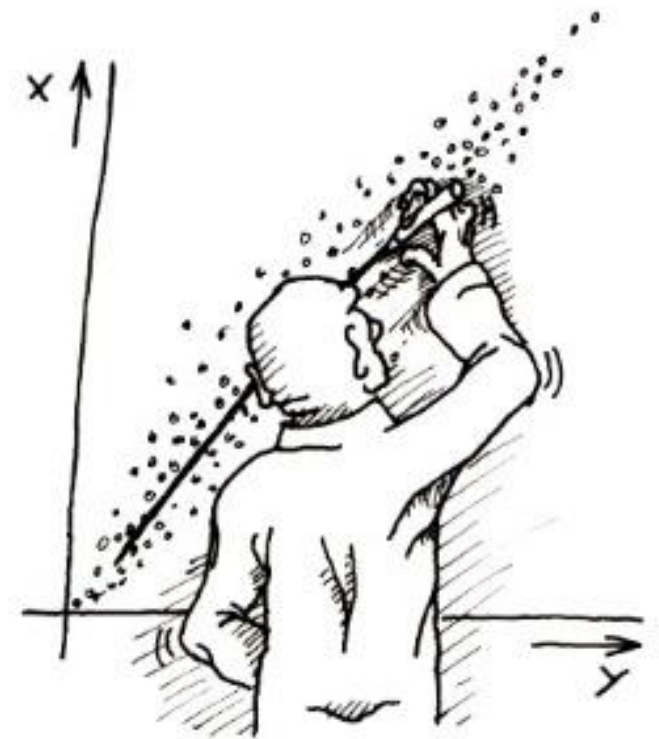
Ordinális és
logisztikus regresszió



Structural equation modeling
(SEM)



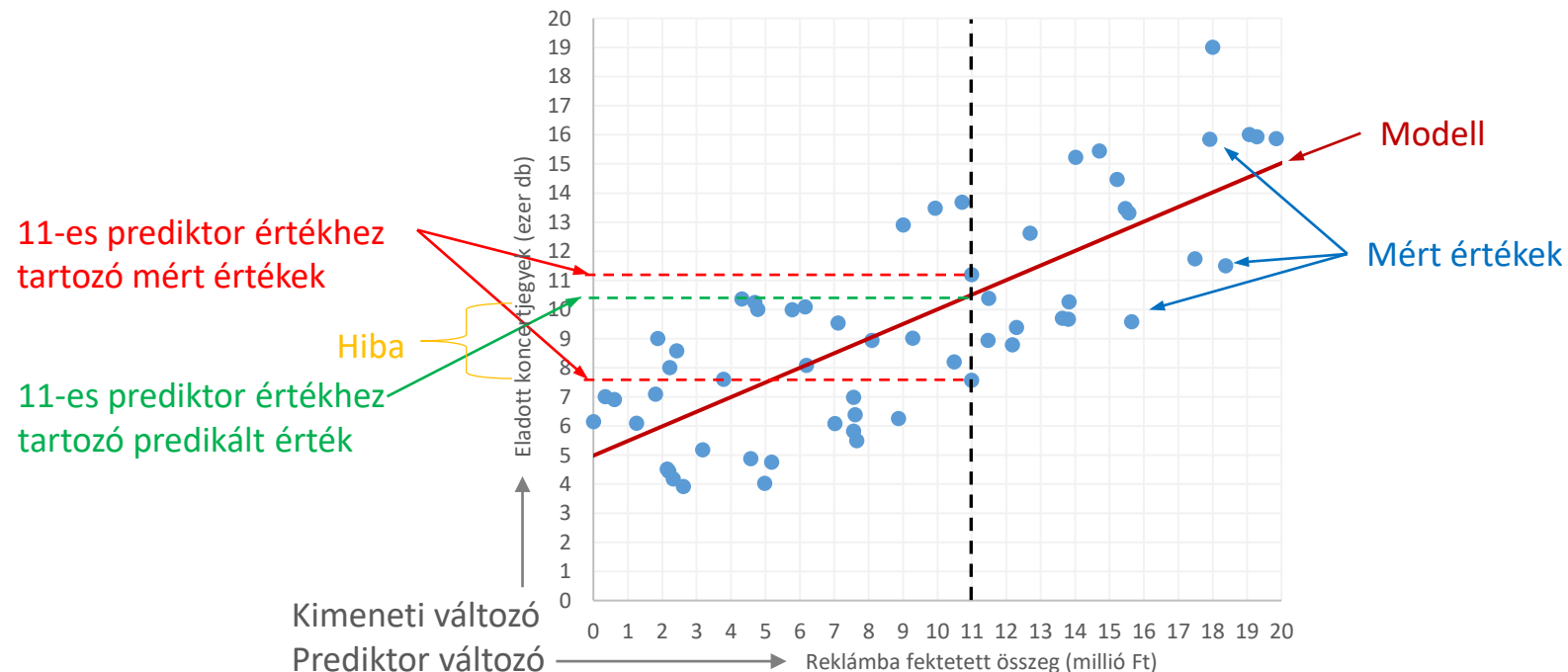
Elmélet

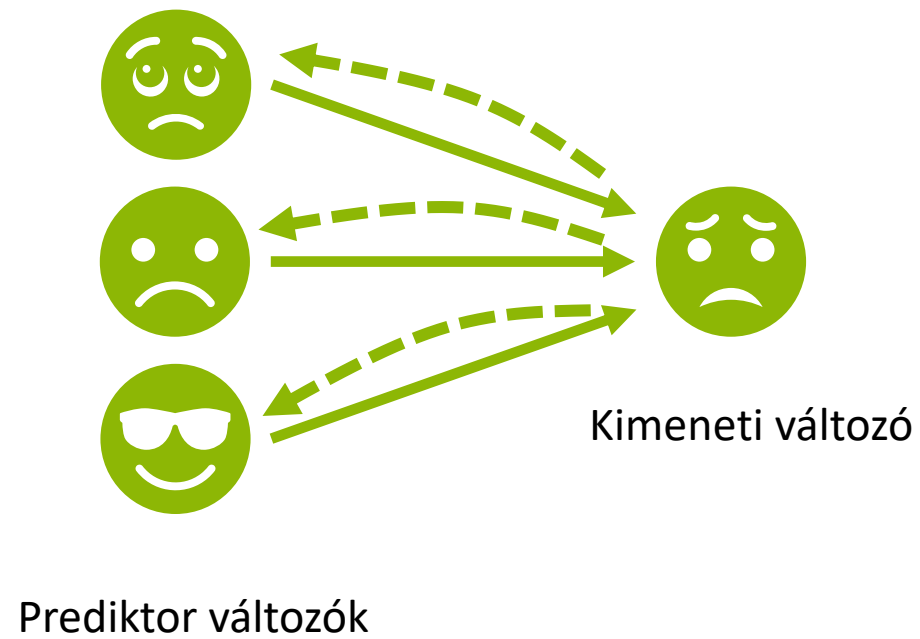
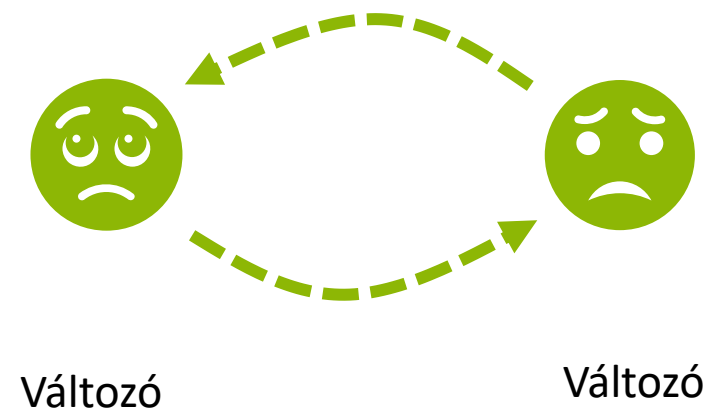


Lineáris Regresszió

Modellt építünk a mért adatainkra, majd a modell segítségével megpróbáljuk a prediktor változókból bejósolni a kimeneti változót.

- **Prediktor változó** és **Kimeneti változó (mért értékek)**
- **Predikált érték** és **Hiba**
- **Kimeneti érték** = **Modell** segítségével a **prediktor érték**ből **predikált érték** + **hiba**

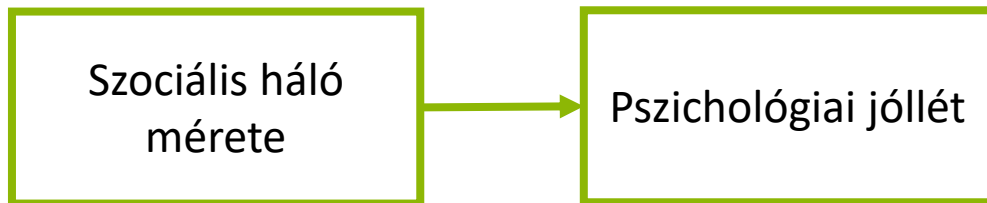




Lineáris regresszió típusai

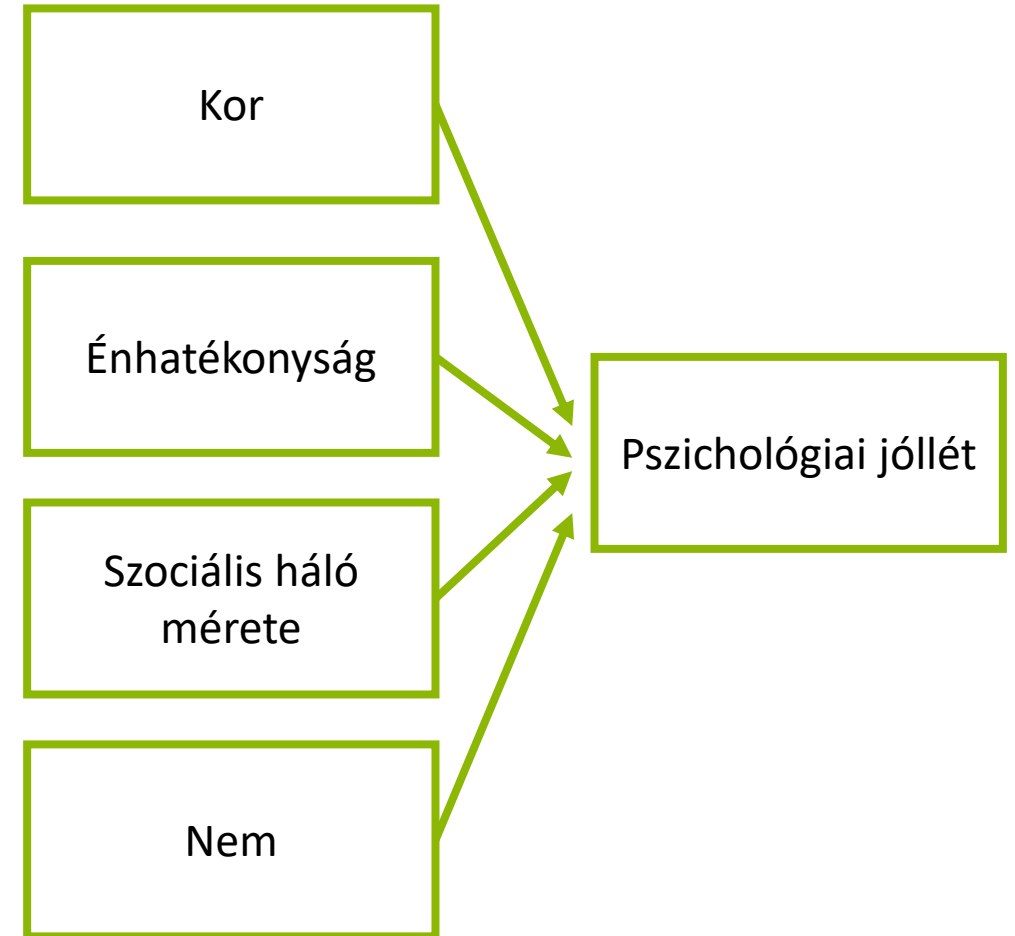
Egyszerű lineáris regresszió (simple regression)

Egyetlen prediktor változóból próbáljuk megjósolni a kimeneti változót.



Többszörös lineáris regresszió (multiple regression)

Több prediktor változóból jósoljuk meg a kimeneti változót.



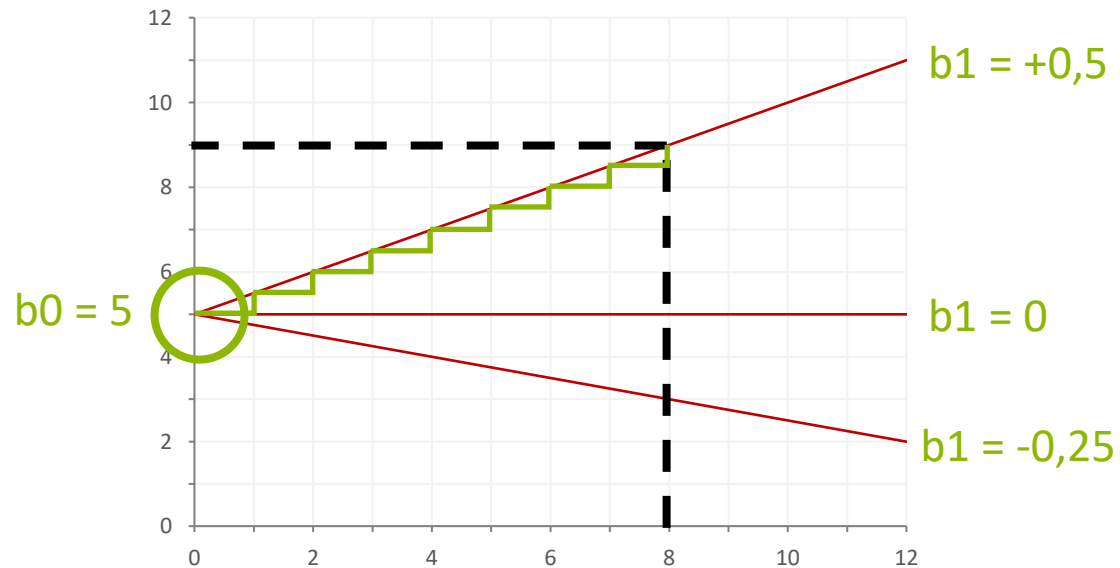
Lineáris regresszió – mert a modellünk alapja egy egyenes.

Egy egyenes két adattal (paraméterrel) írható le – a regressziós elemzésben ezeket nevezzük a két regressziós együtthatónak, azaz **koefficiens**nek (regression coefficient).

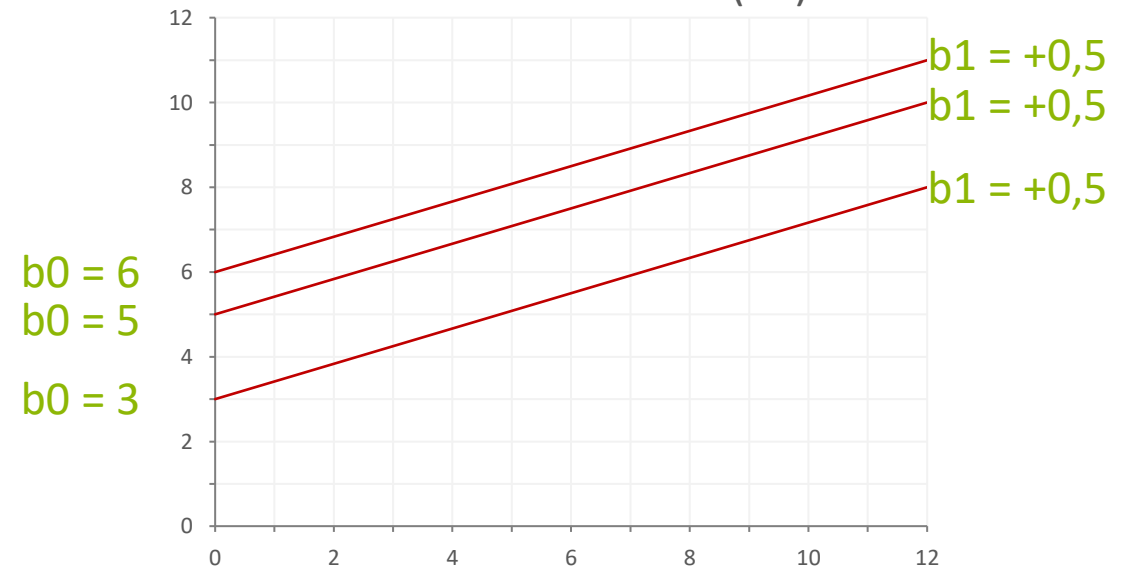
- **Konstans** (**b_0**) – a kezdőpont, ahol az egyenes az y tengelyt metszi.
- **Meredekség** (**b_1**) – egységnyi változás x-ben mekkora változást okoz y-ban

Egyenes egyenlet: **$y = b_0 + b_1 * x$**

Azonos konstans (b_0),
de eltérő meredekség (b_1)



Azonos meredekség (b_1),
de eltérő konstans (b_0)



A lineáris regresszió egyenlete

Predikált értékek kiszámolhatóak a modell segítségével a prediktor értékekből a következő egyenlettel:

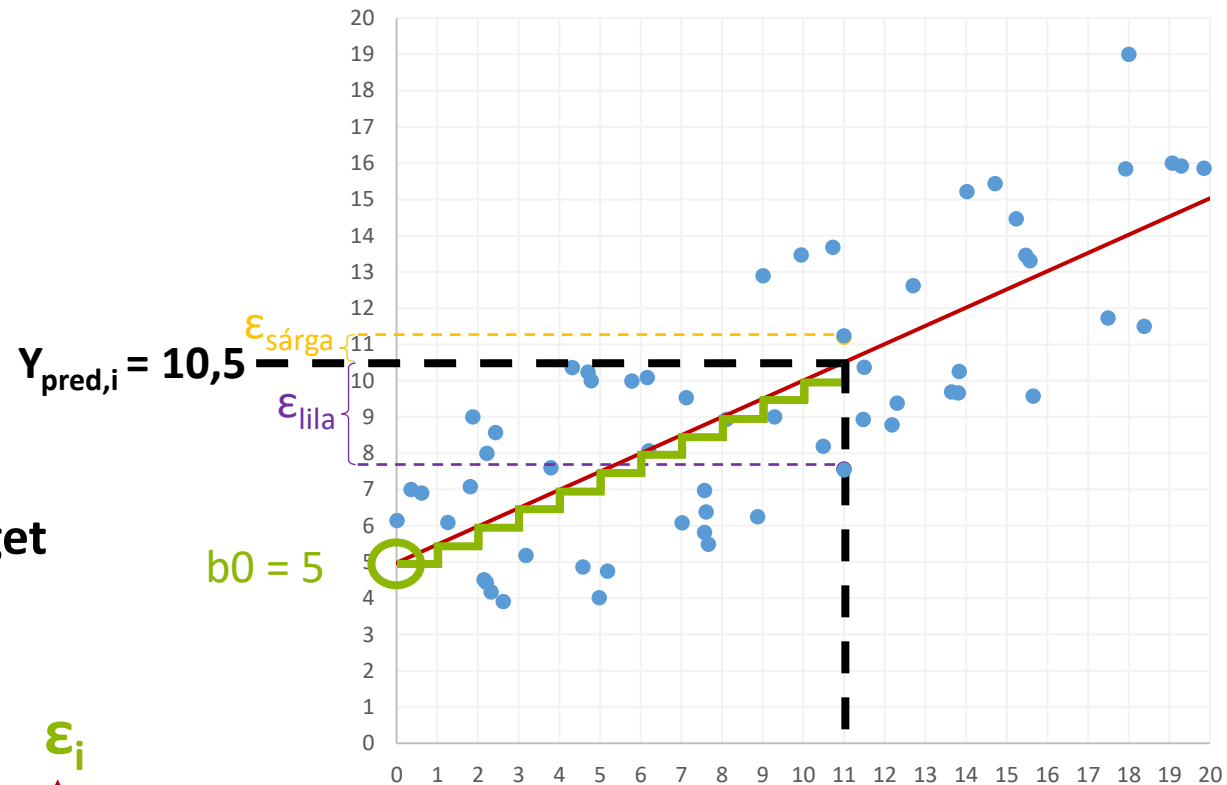
$$Y_{\text{pred},i} = b_0 + b_1 * X_i$$

Predikált érték Konstans Meredekség Prediktor

A **mért értékek** és **predikált értékek** közötti különbséget nevezzük a predikció vagy a **modell hibájának**:

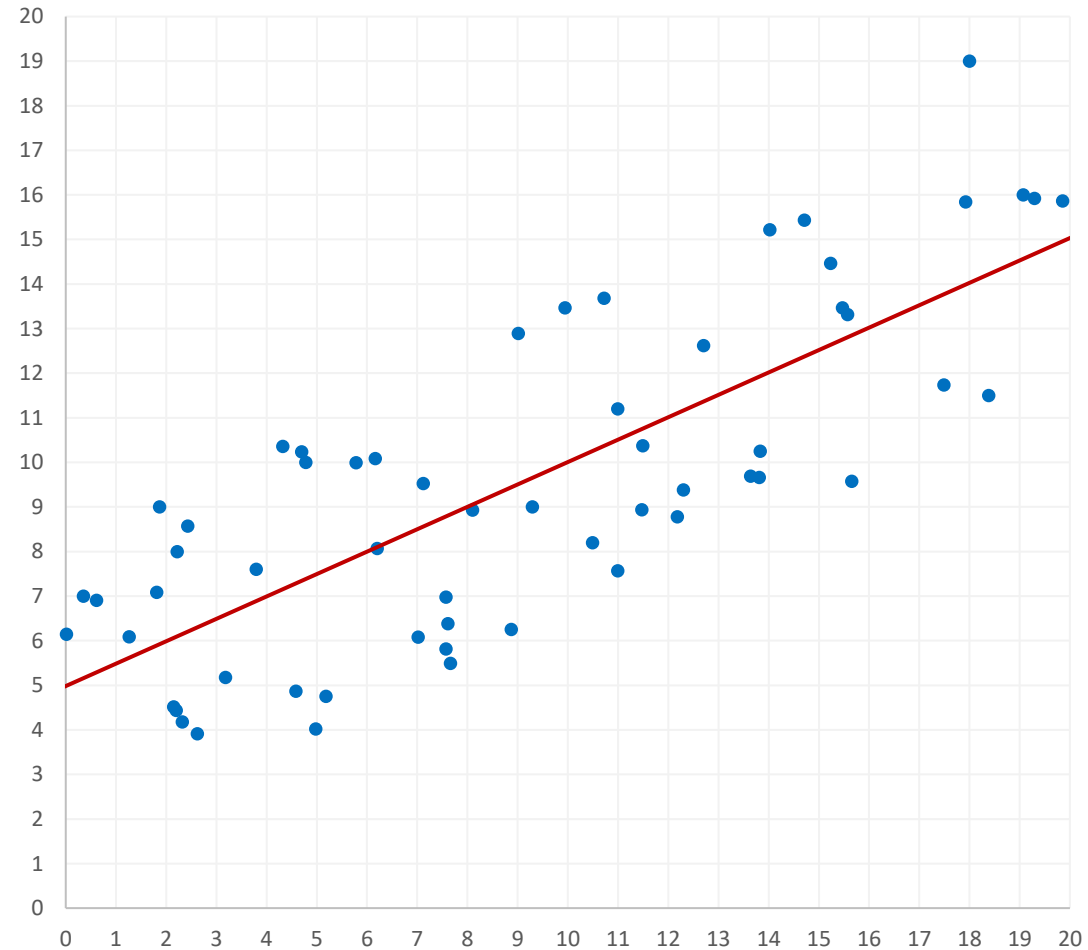
$$Y_i = Y_{\text{pred},i} + \varepsilon_i = b_0 + b_1 * X_i + \varepsilon_i$$

Kimeneti érték Predikált érték Hiba Konstans Meredekség Prediktor Hiba



Legkisebb négyzetek módszere

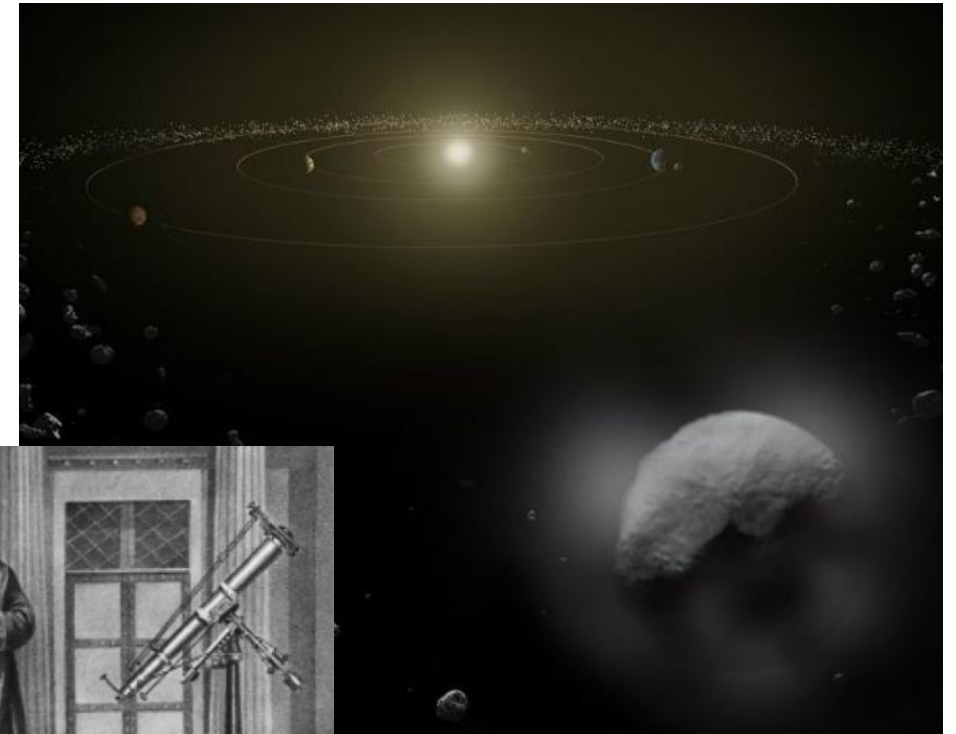
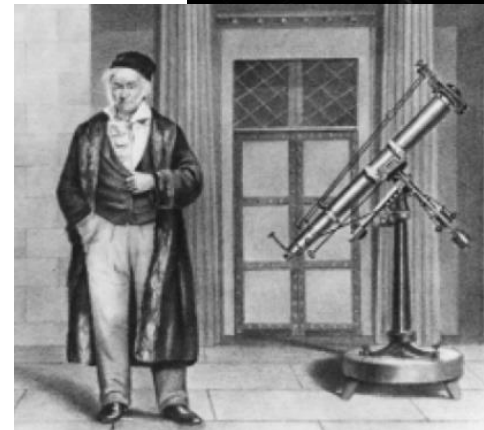
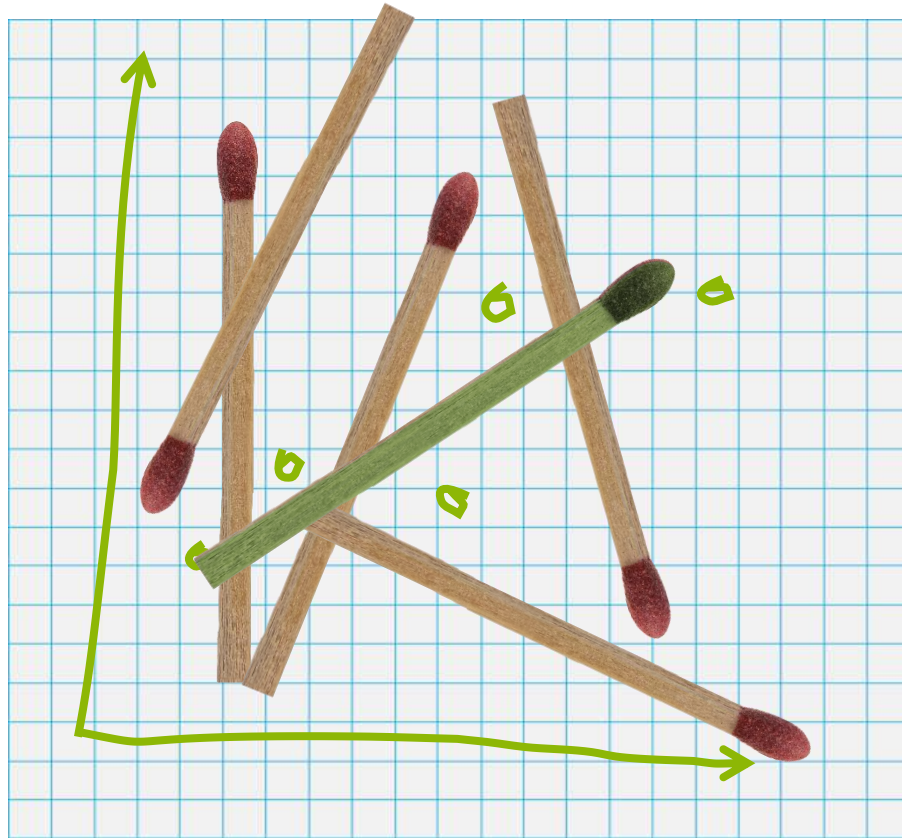
Keressük azt az egyenest, mely a legkisebb hibával jár, azaz a lehető legjobban leírja a pontfelhőt!



Legkisebb négyzetek módszere

Egy adatsorra nagyon sok modell (nagyon sok egyenes) fektethető, ezek közül azt keressük, amelyik legjobban illik az adatokhoz (legjobban leírja azokat).

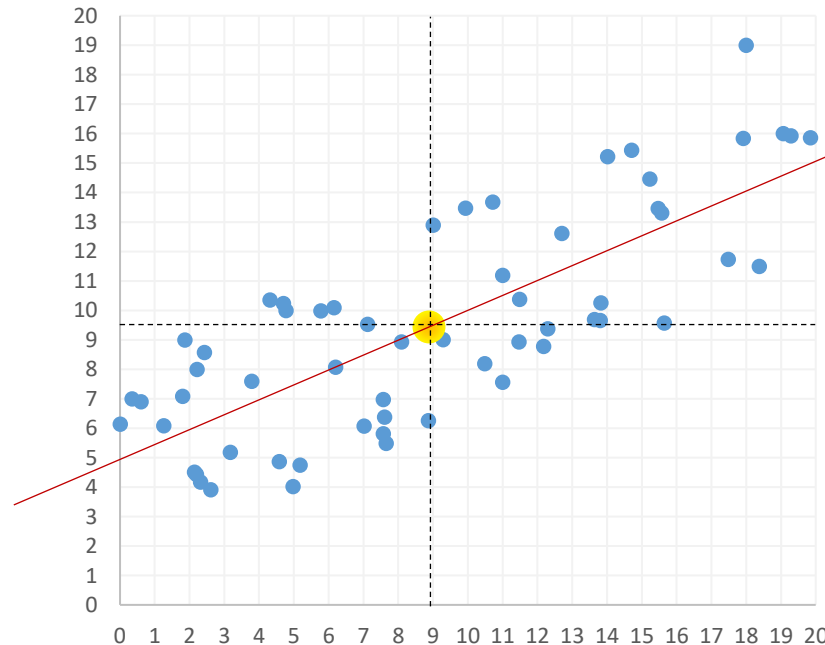
Sok becslési mód létezik, de legelterjedtebb közülük a **legkisebb négyzetek módszere** (ordinary least squares, OLS). A módszert a 24 éves(!) Gauss alkotta meg a Ceres törpebolygó pályájának leírására.



Legkisebb négyzetek módszere

Legkisebb négyzetek módszere

1. Keressük meg a **pontfelhő középpontját**, ezzel fixáljuk az egyenes helyét!
 - X és Y-változó átlaga
 - Ezen biztosan átmegy az egyenes
2. Határozzuk meg az egyenes meredekségét!
 - Válasszuk ki az ezen átmenő egyenesek közül azt, mely a **legkisebb hibával jár**

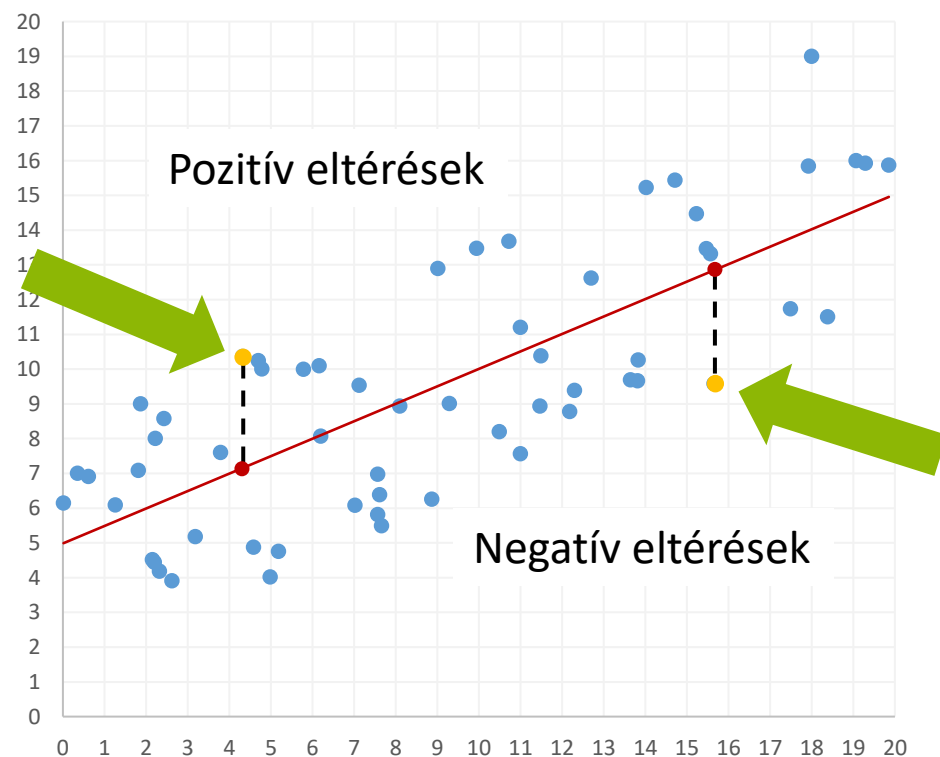


Legkisebb négyzetek módszere

Azt a modellt keressük, amelyik legjobban illik az adatokhoz, másként megfogalmazva, ahol a legkisebb a különbség a ténylegesen mért adatok és predikált értékek (ergo az egyenesünk) között

Reziduális hiba (angolul residuals): Eltérés a predikált értékek (egyenes) és a mért adatok között

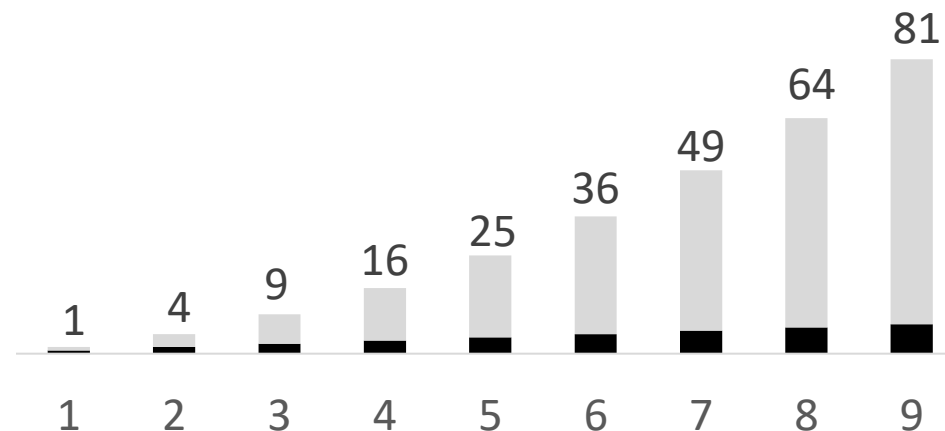
Négyzetes eltérés (squared residuals)



$$(Y_i - Y_{\text{pred},i})^2$$

vagy

$$|Y_i - Y_{\text{pred},i}|$$

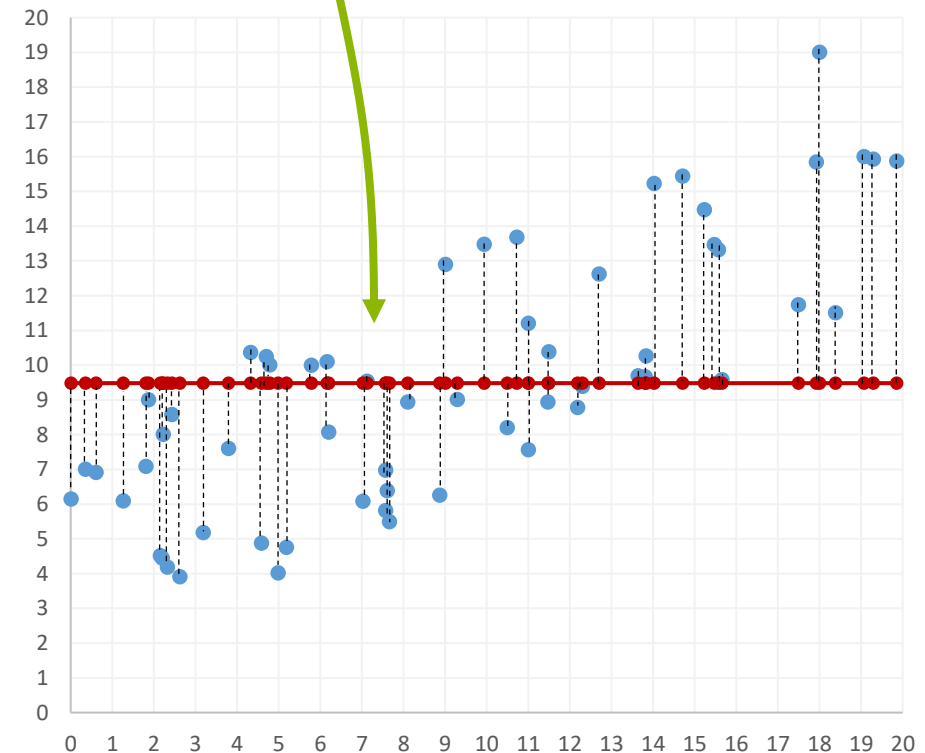
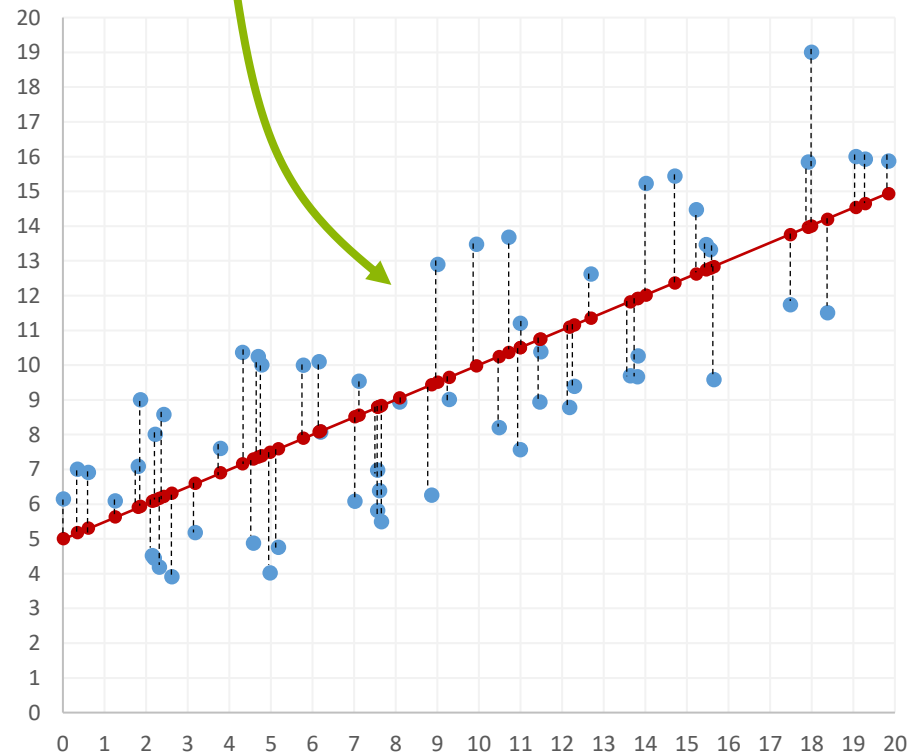


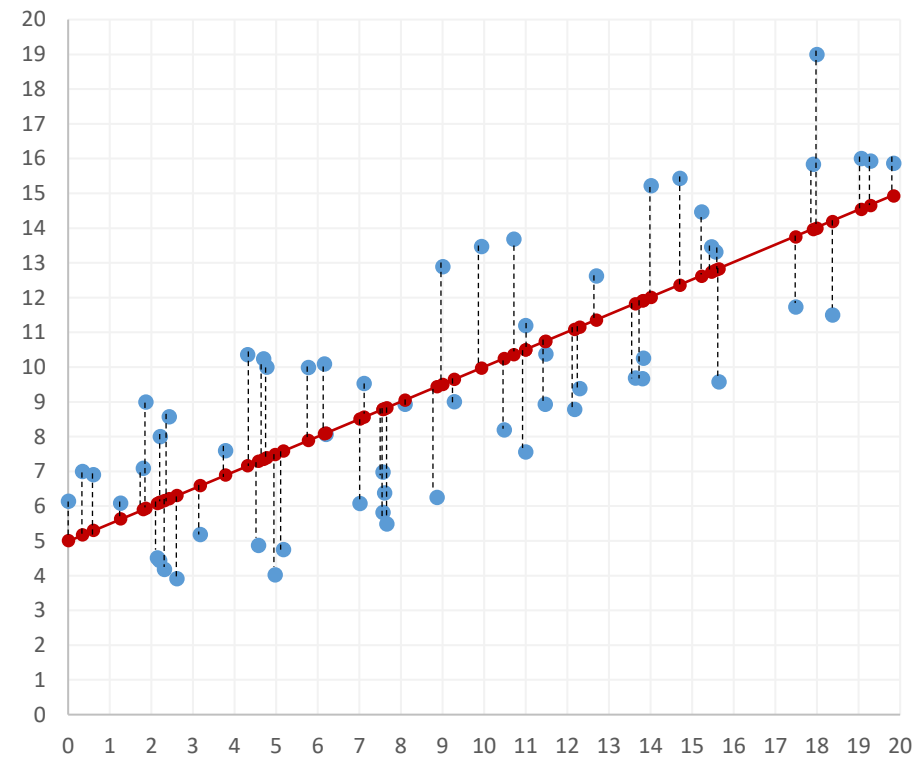
Legkisebb négyzetek módszere

Reziduális hibák négyzetösszege (angolul sum of squared residuals, SS_R)

- a predikált és mért értékek távolsága, az összpontatlanság mértéke
- Azt az egyenest választjuk, melyben a legkisebb a reziduális hibák négyzetösszege

Ezt az egyenest használva kisebb az össztévedés, mint ezt az egyenest használva





Megtaláltuk a lehető legjobb modellt, de mennyire jó az?

Effect-size

- A kimeneti változó értékeit mennyire tudjuk megmagyarázni a prediktor változó(k)kal azaz a kimeneti változó varianciáját hány százalékban tudjuk magyarázni.
- A *hatás nagyságát* írja le

$$\text{effect size} = \frac{\text{megmagyarázott változatosság}}{\text{teljes változatosság}}$$

Statisztikai érték

- A hatás és hiba aránya (a modell segítségével megmagyarázható rész és meg nem magyarázható rész aránya)
- A *hatás valószínűségének* becsléséhez szükséges

$$\text{statisztikai érték} = \frac{\text{hatás}}{\text{hiba}}$$

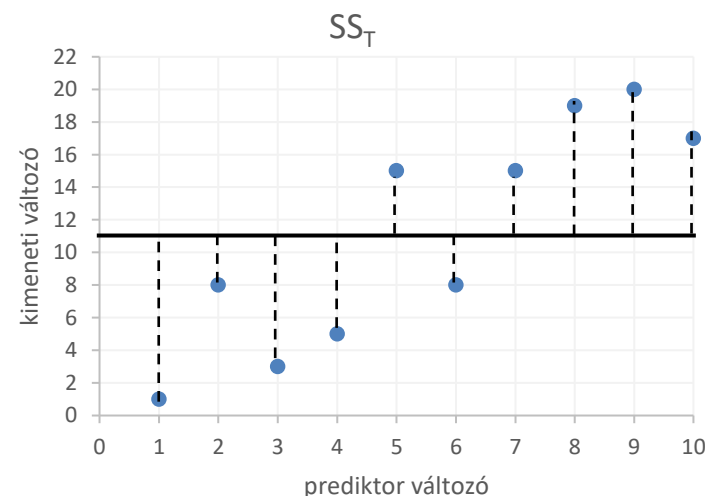
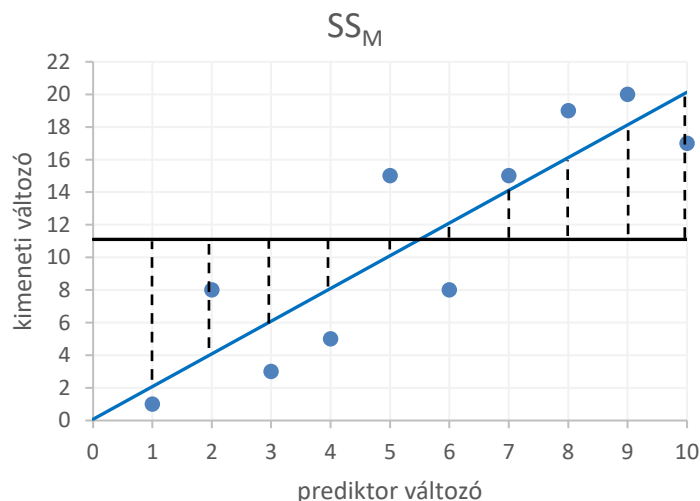
Effect-size:

- A kimeneti változó értékeit milyen mértékben tudjuk megmagyarázni a prediktor változó(k)kal
- a kimeneti változó varianciáját hány százalékban tudjuk magyarázni

$$\text{effect size} = \frac{\text{megmagyarázott változatosság}}{\text{teljes változatosság}} = R^2 = \frac{SS_M}{SS_T}$$

SS_T = sum of squares of total – a mért értékek átlagtól való négyzetes össztvolsága

SS_M = sum of squares of model – a predikált értékek átlagtól való négyzetes össztvolsága



Statisztikai érték - lineáris regressziónál F-érték

A hatás és hiba aránya - a modell segítségével megmagyarázható és nem magyarázható rész aránya

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

SS_M = sum of squares of model – a predikált értékek átlagtól való négyzetes összértéke

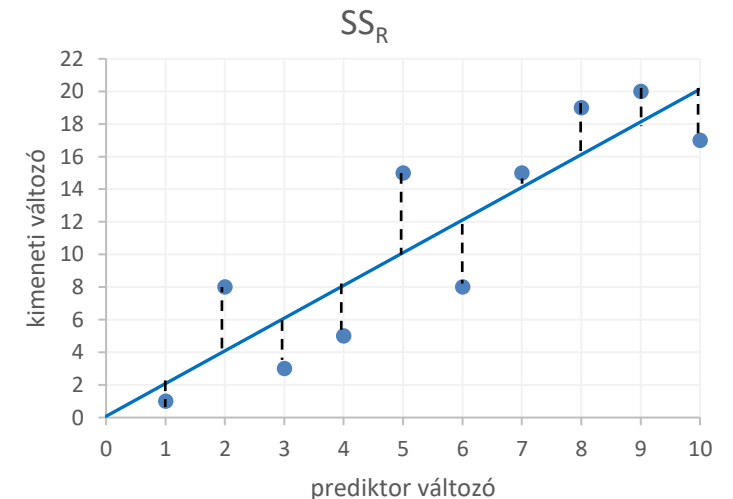
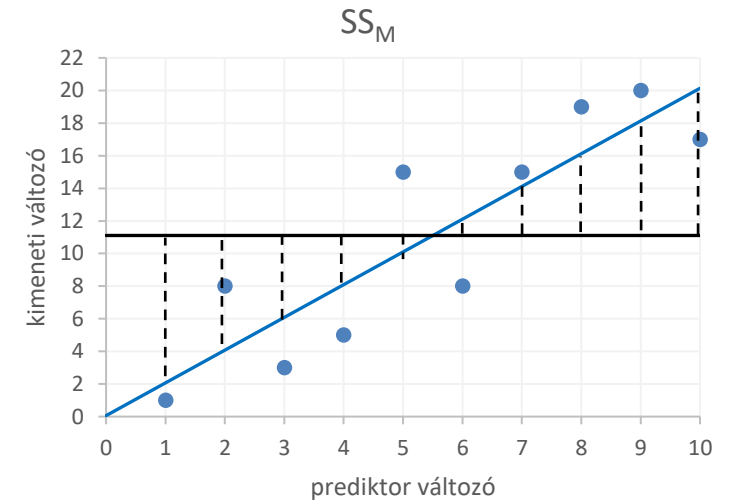
SS_R = sum of squares of residuals – a predikált és mért értékek eltérésének négyzetösszege

df_M = modell szabadságfoka – modell összetettsége, prediktorok száma

df_R = hiba szabadságfoka – elemszám mínusz becsült paraméterek száma

MS_M = egy prediktor változóra jutó átlagos hatás – hatás

MS_R = átlagos pontatlanság a becslés során – hiba



$$F = \frac{\frac{SS_M}{df_M}}{\frac{SS_R}{df_R}}$$

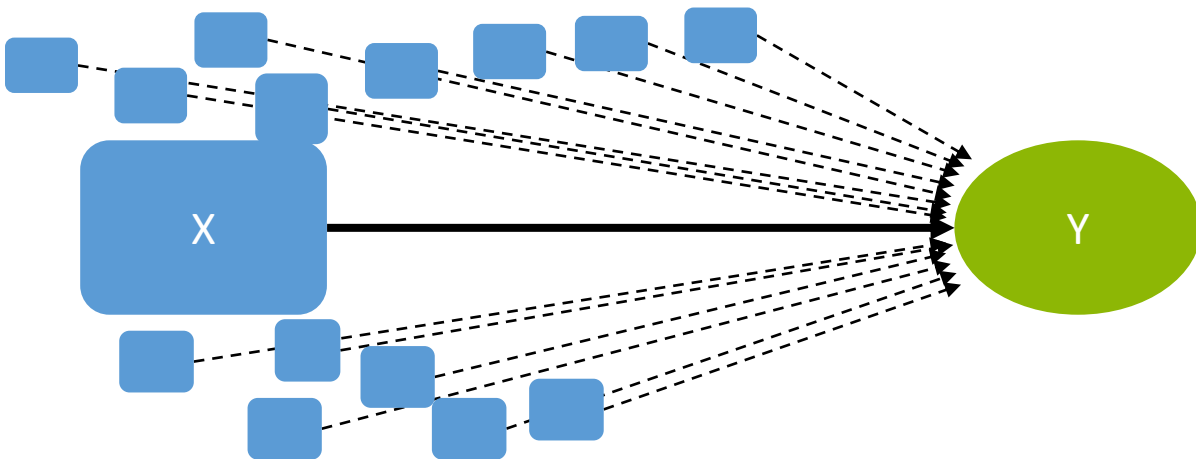
Modellhez tartozó khi-négyzet eloszlás és szabadságfoka

Hibához tartozó khi-négyzet eloszlás és szabadságfoka

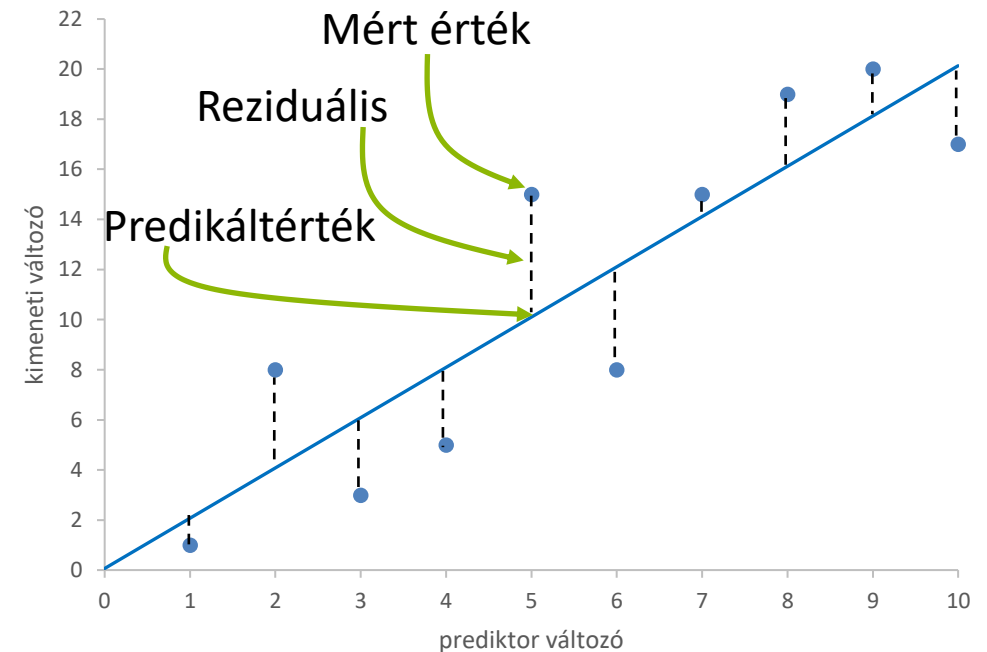
Normalitás feltétele:

A becsült értékek hibájának normáeloszlása

Ha a hiba független a modeltől, akkor a központi határeloszlástétel miatt teljesül



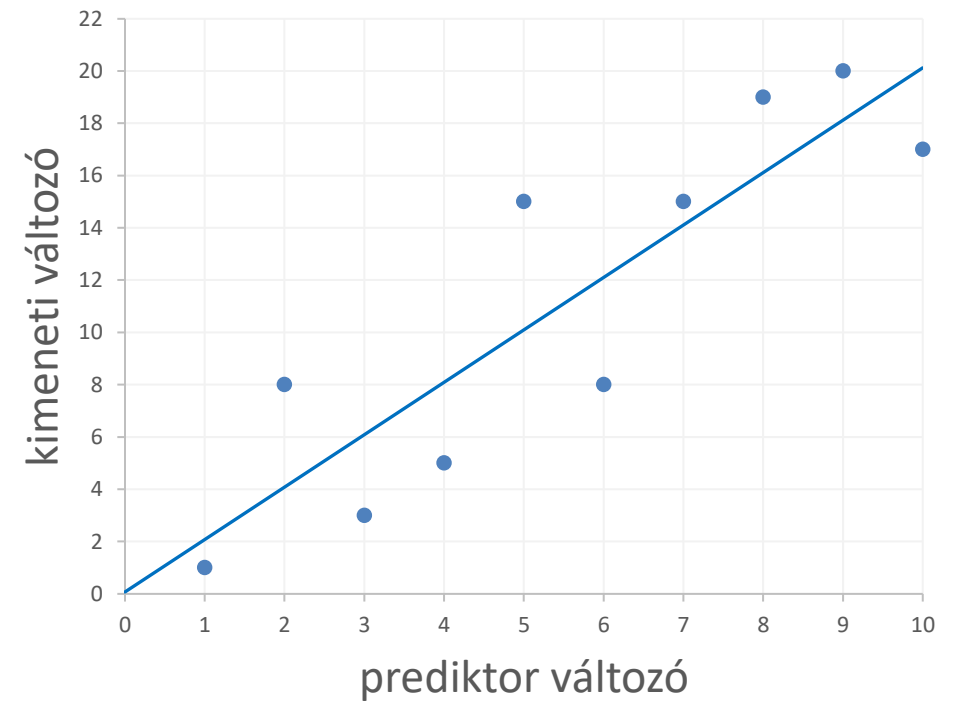
$$Q(df=k) = Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_k^2$$

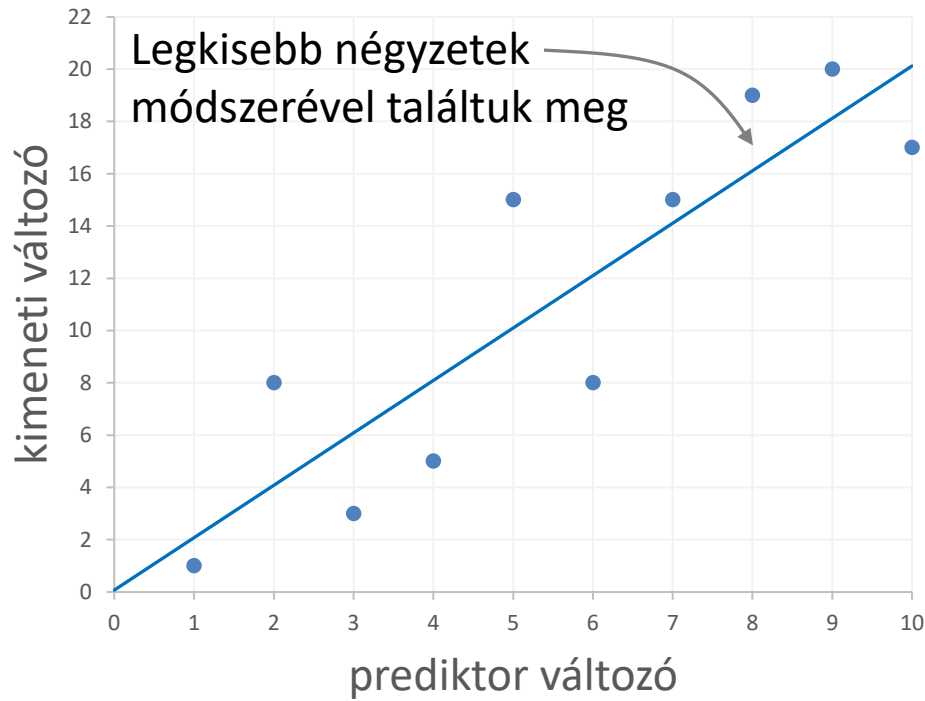


Reziduális: predikált és mért érték távolsága

$$SS_R = Rez_1^2 + Rez_2^2 + Rez_3^2 + \dots + Rez_k^2$$

F-érték és R^2 érték összefoglalása





R^2 – effect-size mutató

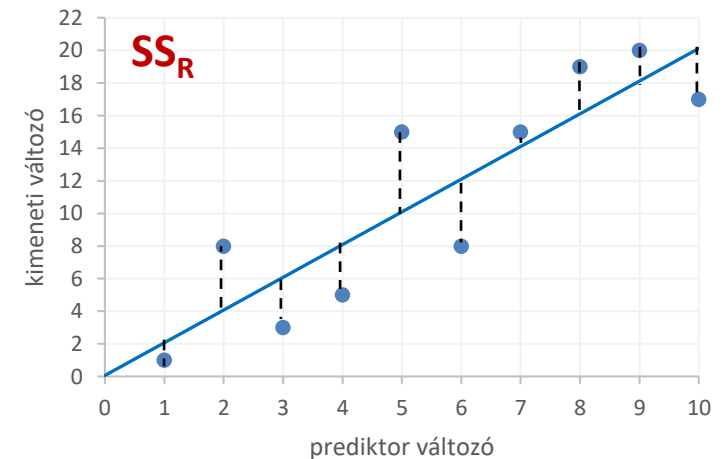
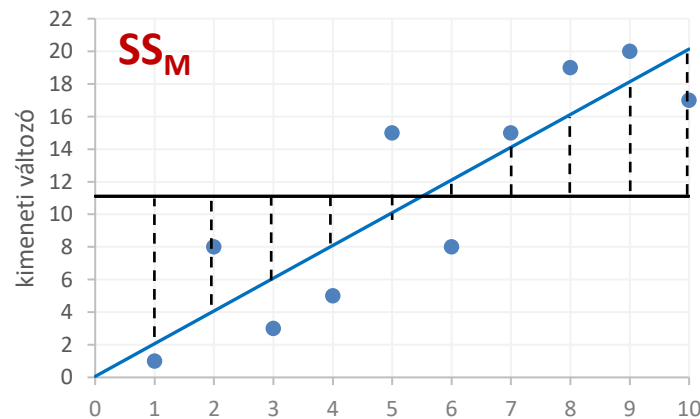
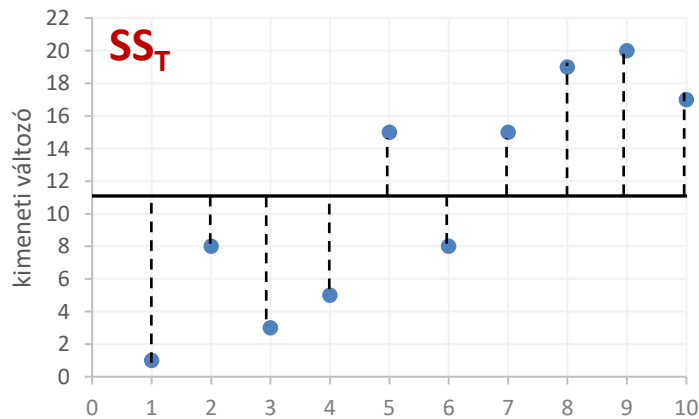
Hatás nagyságát írja le.

$$R^2 = \frac{SS_M}{SS_T}$$

F-érték – statisztikai érték

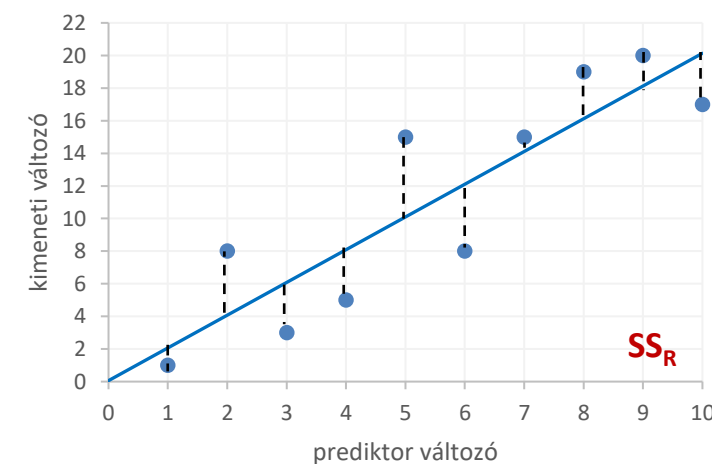
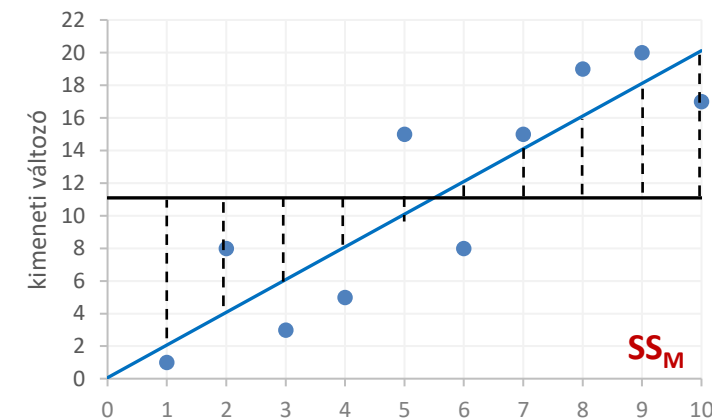
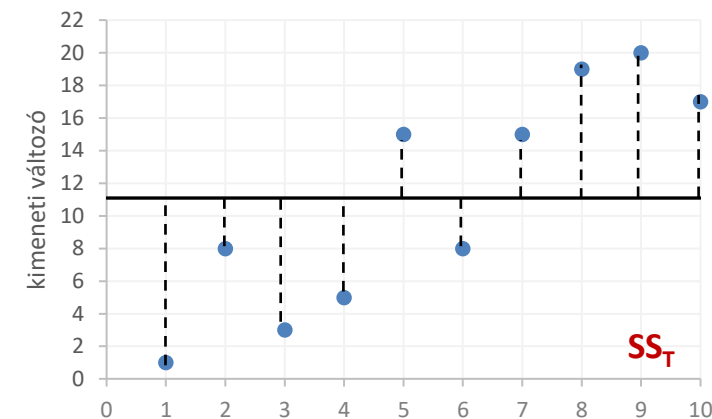
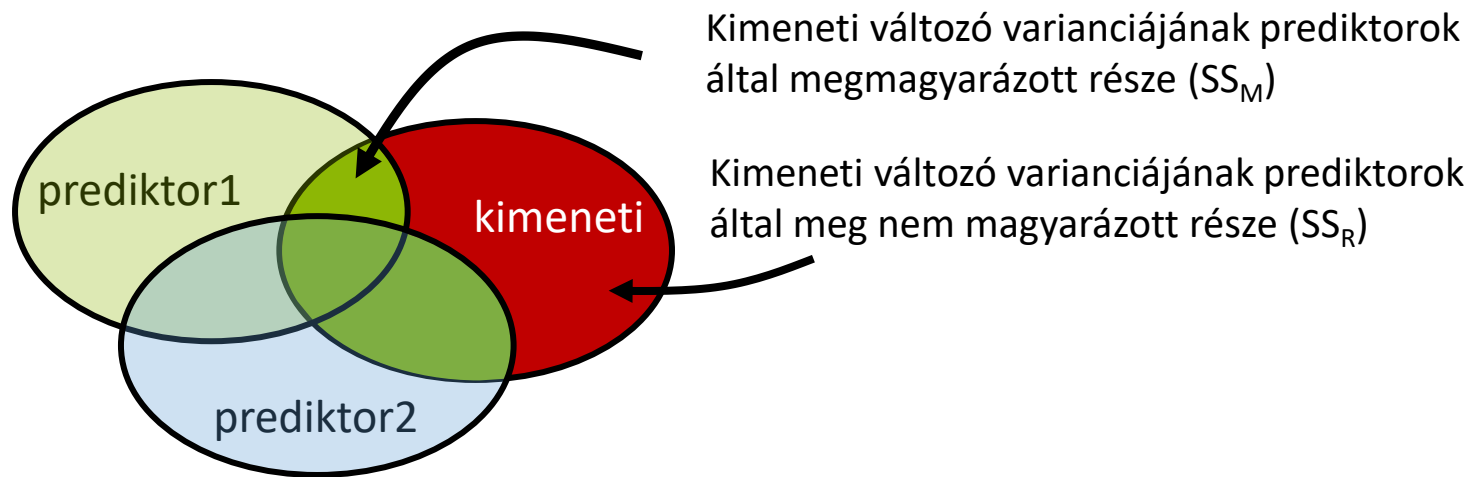
A szignifikancia értéket rendeljük majd hozzá, a modell általánosíthatóságát írja le.

$$F = \frac{MS_M}{MS_R} = \frac{\frac{SS_M}{df_M}}{\frac{SS_R}{df_R}}$$



$$R^2 = \frac{SS_M}{SS_T} \quad \text{és} \quad F = \frac{MS_M}{MS_R} = \frac{\frac{SS_M}{df_M}}{\frac{SS_R}{df_R}}$$

$SS_M + SS_R = SS_T$ – evidens, a teljes variancia felbontható megmagyarázott és meg nem magyarázott részre



$$R^2 = \frac{SS_M}{SS_T} \quad \text{és} \quad F = \frac{MS_M}{MS_R} = \frac{\frac{SS_M}{df_M}}{\frac{SS_R}{df_R}}$$

$df_T = N - 1$, ahol az elemszámból kivonjuk az egyetlen összefüggést az átlagot. Az MS_T a függő változó varianciájával lenne egyenlő – ezt a kifejezést nem szokás használni.

df_M - a modell bonyolultságát tükrözi, a prediktorok számával egyenlő.

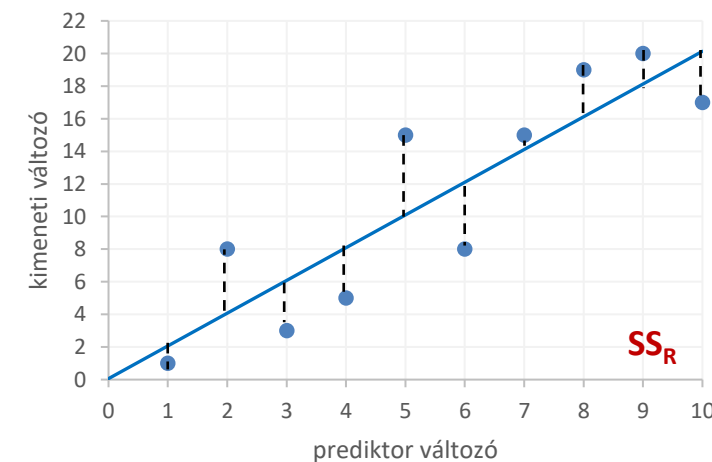
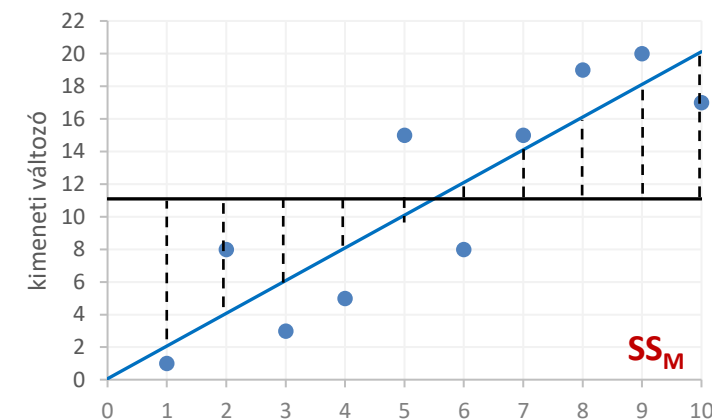
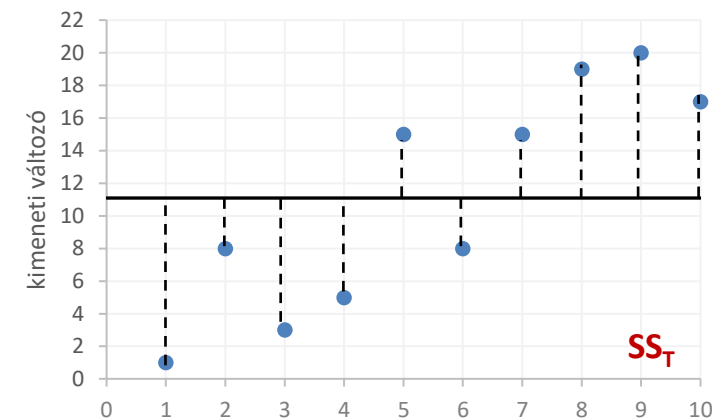
df_R - a minta nagyságából kell kivonni a becsült paraméterek számát.

Egy prediktor esetén

- $df_M = 1$, mert egy prediktor van,
- és $df_R = N - 2$, mert a két becsült paraméter a konstans és a meredekség.

Két prediktor esetén

- $df_M = 2$, mert két prediktor van,
- és $df_R = N - 3$, ahol a három paraméter a konstans és a két prediktorhoz tartozó dimenzióban a regressziós egyenes meredeksége



Az elnevezésekből több verzió is van, és szerencsétlenül félreérthetőek:

SS_M

(sum of squares of model)

=

ESS

(explained sum of squares)

=

SSR

(sum of squares of regression)

SS_R

(sum of squared residuals)

=

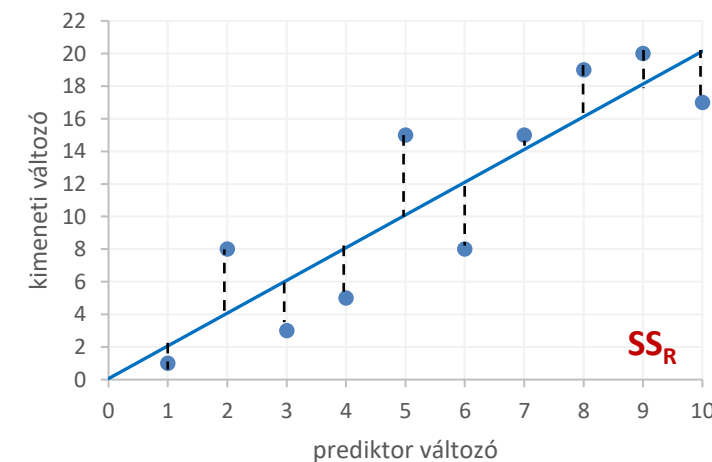
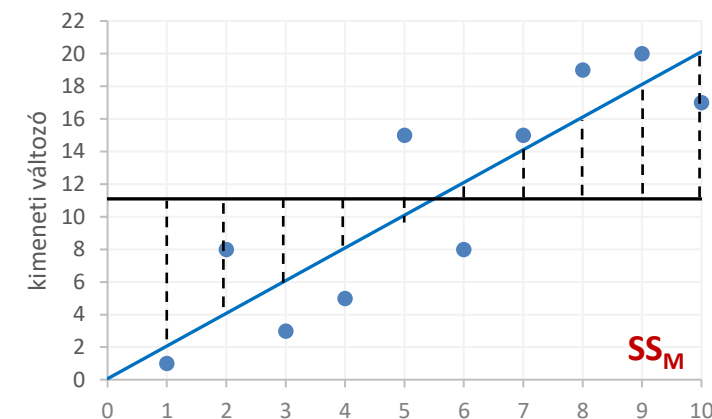
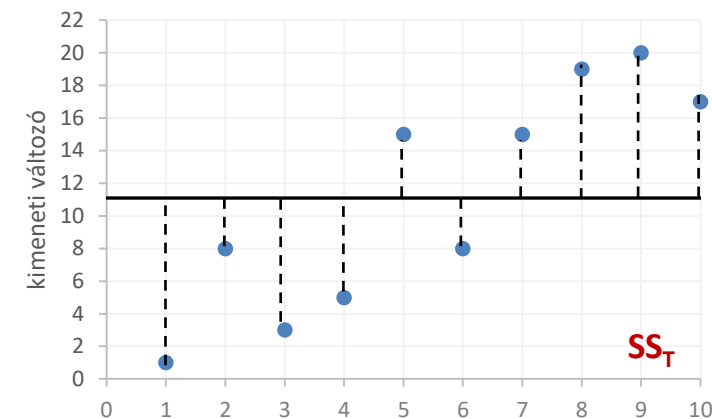
RSS

(residual sum of squares)

=

SSE

(sum of squares of errors)



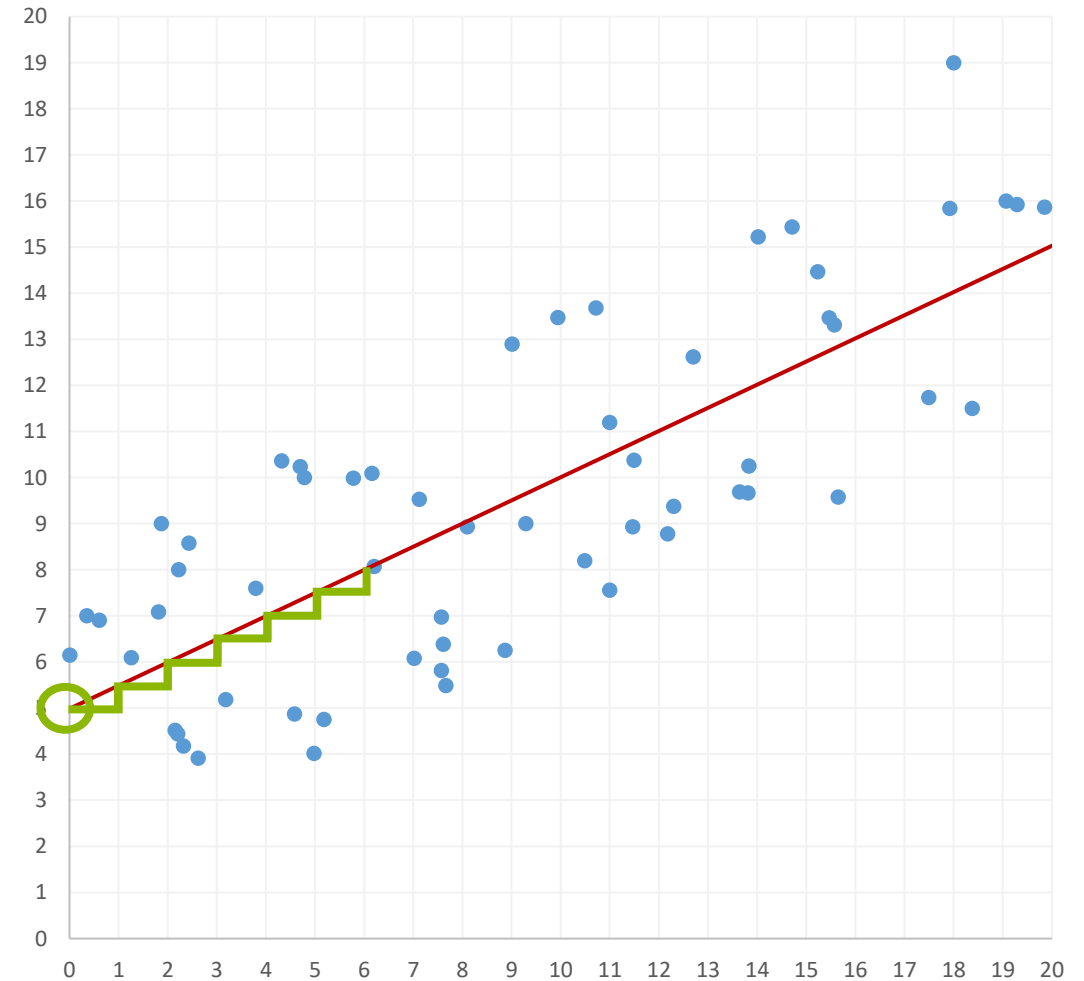
Regressziós modell „jóságának” leírása

- F-érték (statisztikai érték)
- R-négyzet (effect-size mutató)

Regressziós egyenes leírása

Az egyenes paraméterei (együttható, **koefficiensei**) segítségével

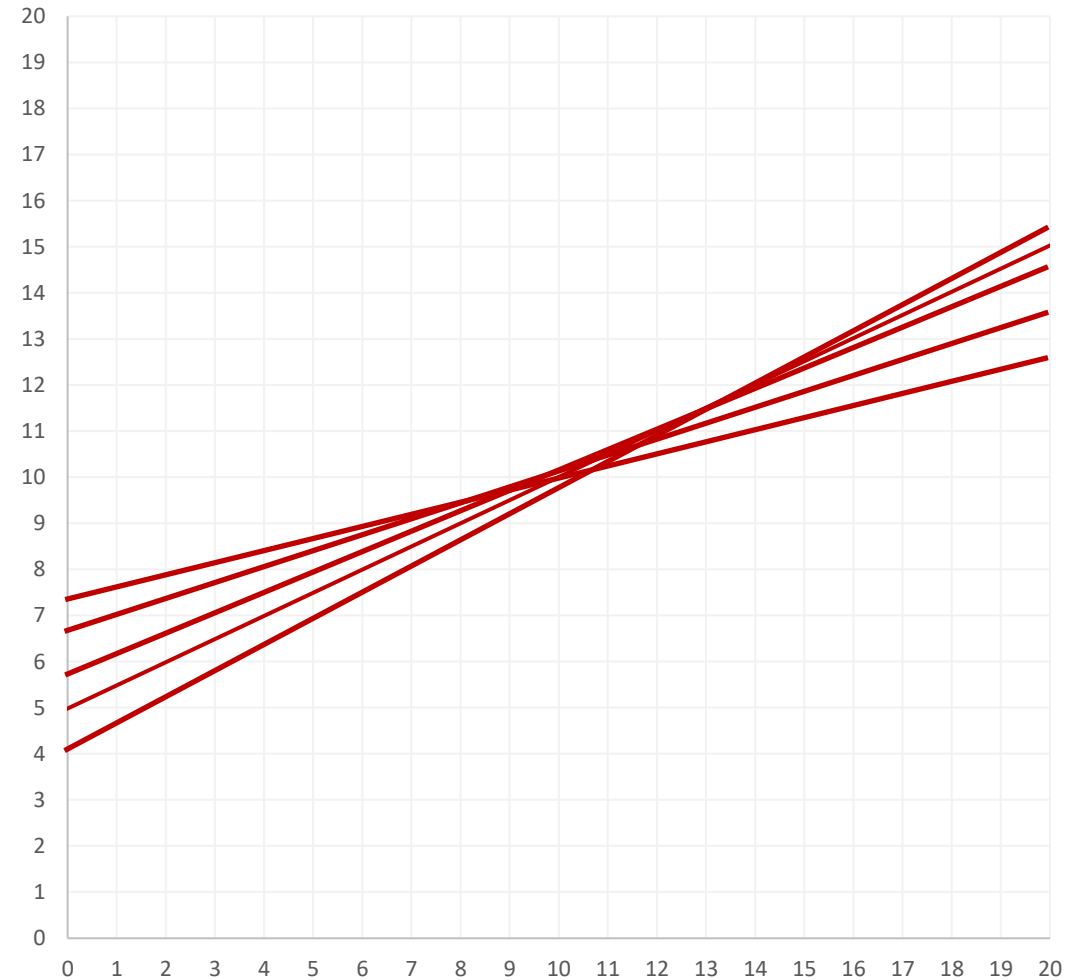
- A **b0 regressziós együttható** megadja, hogy mennyi lenne a kimeneti változó értéke, ha (minden) prediktor értéke 0 lenne, ez az y tengely metszéspontja
- A **b1 regressziós együttható** megadja, hogy a prediktor változó egységnyi változása mekkora változást hoz létre a kimeneti változóban



Regressziós együtthatók standard errorja

A regressziós együtthatókhoz (a konstanshoz is és a meredekséghez is) tartozik egy **standard error** (SE) érték is, mely megadja a paraméter bizonytalanságát.

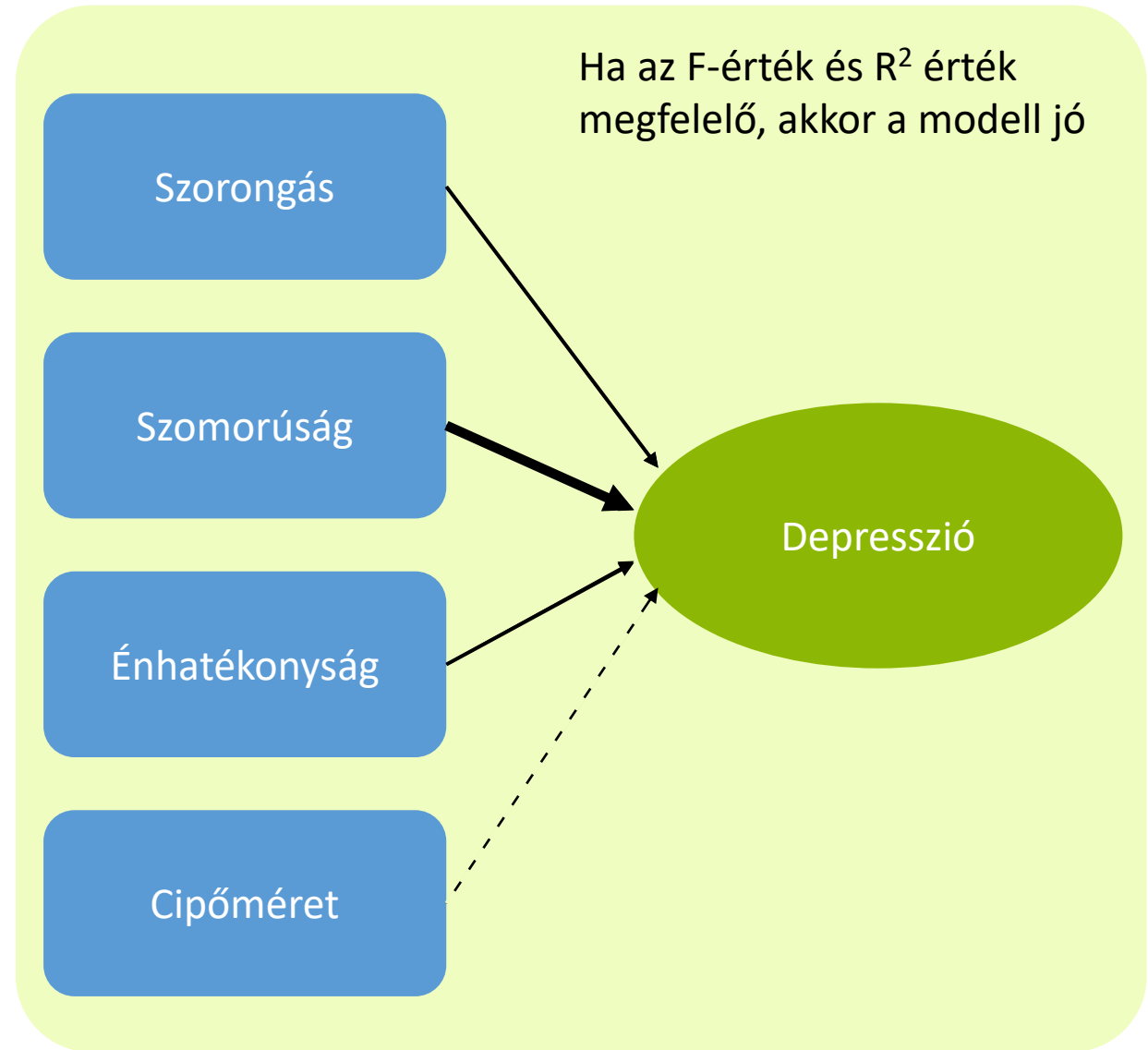
Azt adja meg, hogy, ha többször vennék mintát a populációból, és többször elvégezném a regressziós, akkor a kapott együtthatóknak mekkora lenne a szórása.



Modell általános értékelése a statisztikai értékkel és effectsize mutatóval.

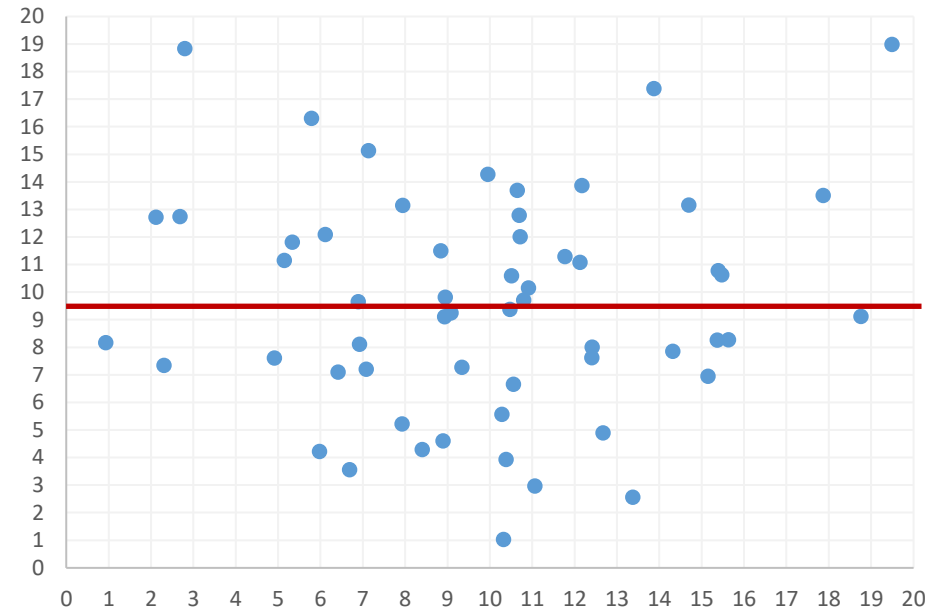
+

Prediktorok egyéni értékelése a hozzájuk tartozó meredekséggel (illetve annak standard errorjával és a t-teszttel)



Ha egy prediktornak **nincs hatása** a kimeneti változóra,

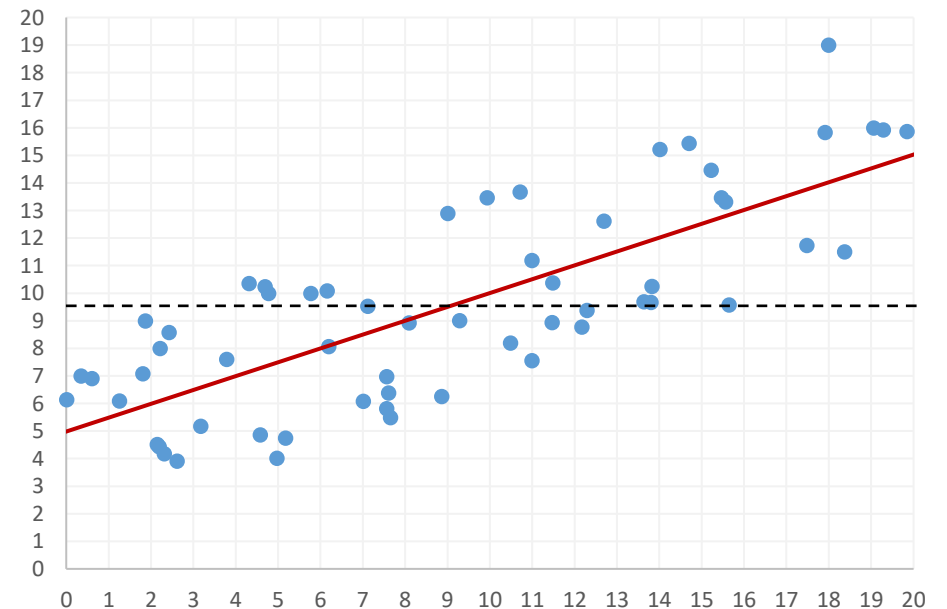
- akkor a regressziós egyenes vízszintes lesz, tehát $b_1 = 0$),
- mert bármilyen prediktor változó értékhez bármilyen kimeneti változó érték tartozhat,
- ergo a legjobb predikciónk sem lesz jobb, mint a kimeneti változó átlaga.



Ha egy prediktornak **hatása van** a kimeneti változóra,

- akkor a regressziós egyenes meredeksége szignifikánsan eltér a nullától.
- Ezt t-teszttel ellenőrizzük

$$t = \frac{b_{\text{mért}} - b_{\text{nullhip}}}{SE_b} = \frac{b_{\text{mért}}}{SE_b}$$



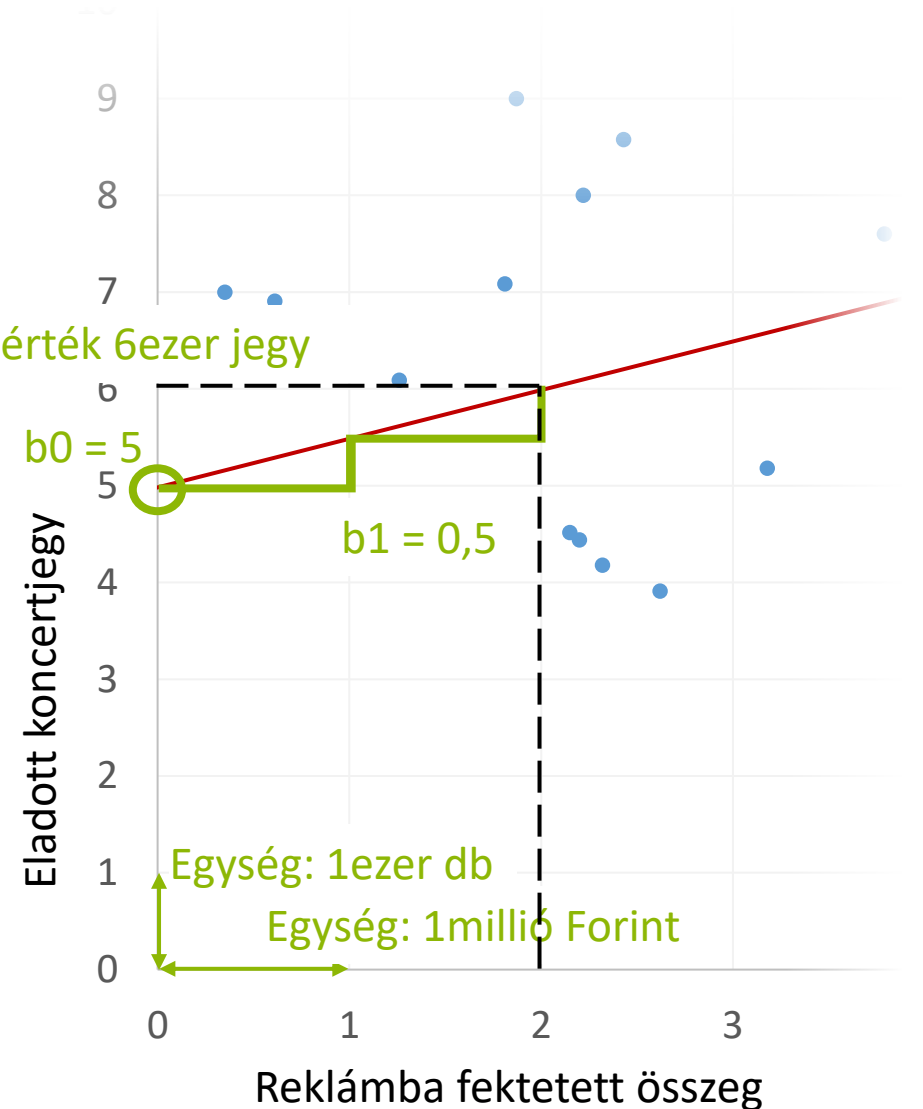
b1 – standardizálatlan regressziós koefficiens

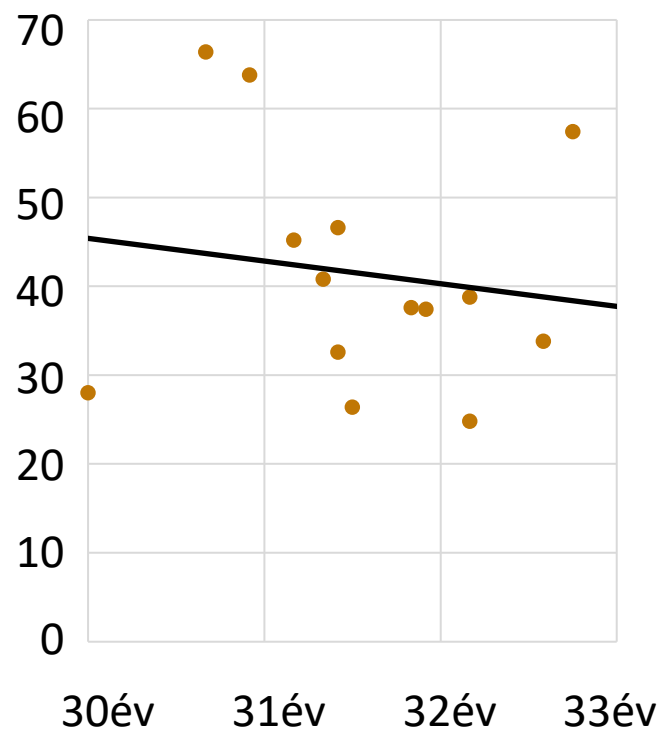
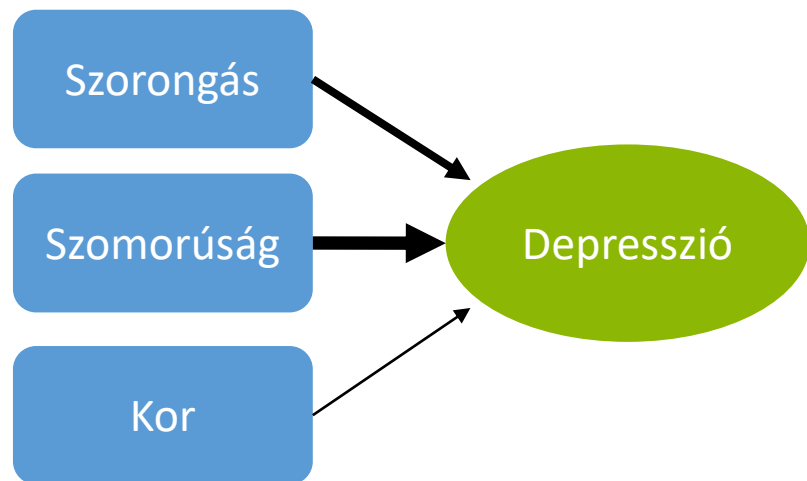
Ennek segítségével lehet a predikált értéket kiszámolni.

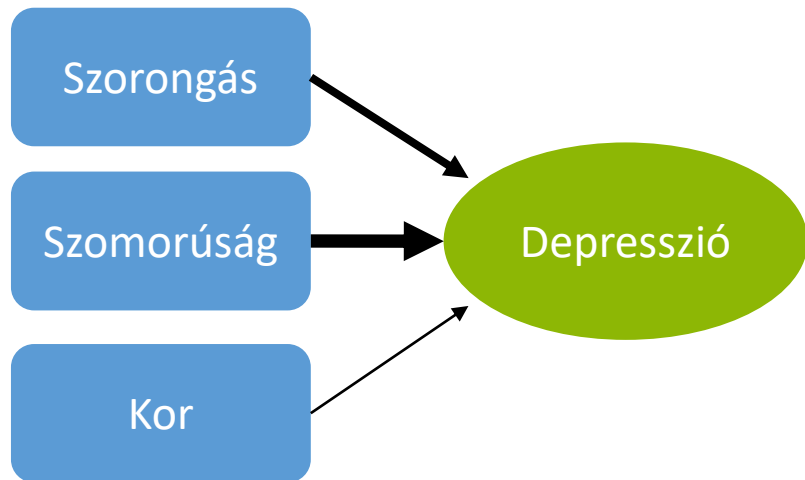
Például

- Reklámba fektetett összeg egysége 1millió Forint, és az eladott koncertjegyek egysége ezer darab.
- Ha a modell konstansa $b_0 = 5$, azaz reklám nélkül várhatóan 5000 jegyet tudok majd eladni, és
- és a reklámba fektetett összeghez tartozó meredekség $b_1 = 0,5$, azaz a reklámba fektetett plusz egy millió Forint 500 darabbal növeli meg az eladott koncertjegyek számát,
- Akkor 2millió Forint reklám esetén $5000 + 2 \cdot 500 = 6000$ db eladott jegyre számíthatok.

Predikált érték 6ezer jegy





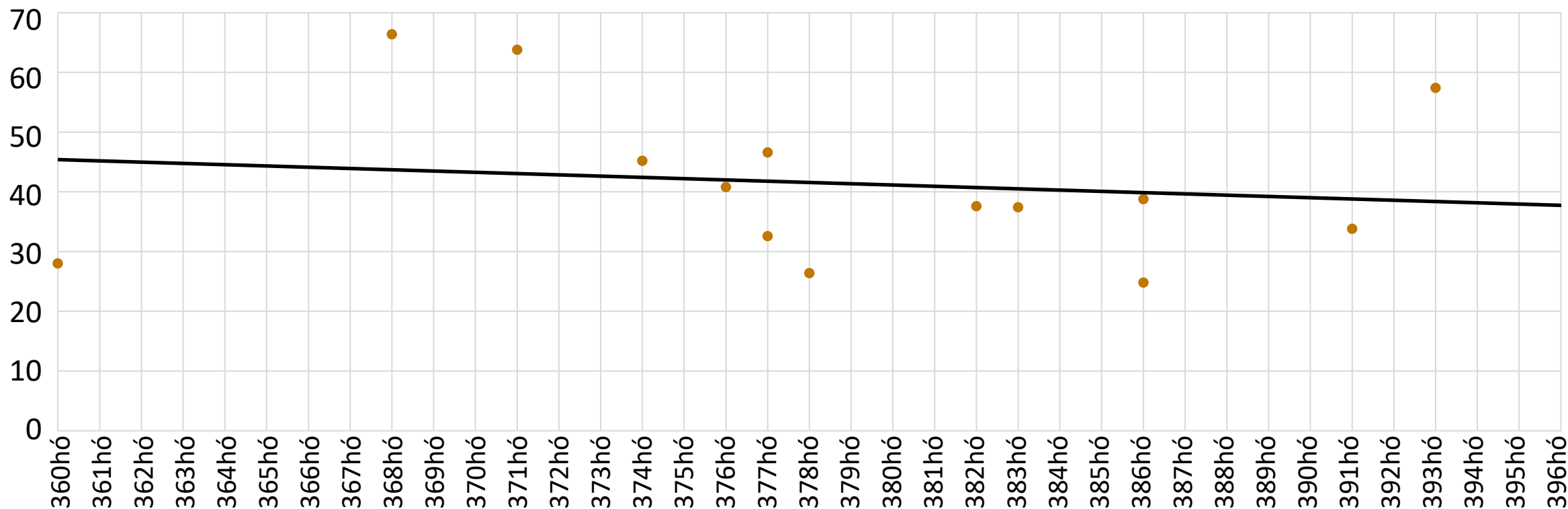


b1 – standardizálatlan regressziós koefficiens

- Skálafüggő,
- ezért különböző skálájú változók hatása nem összehasonlítható

β_1 – standardizált regressziós koefficiens

- A meredekség **skála független** mérőszáma,
- ezek alapján már összehasonlíthatóak a prediktorok.
- Sőt! A meredekségekhez kiszámolt konfidencia intervallumok segítségével az is megállapítható, két prediktor hatása között szignifikáns-e a különbség.



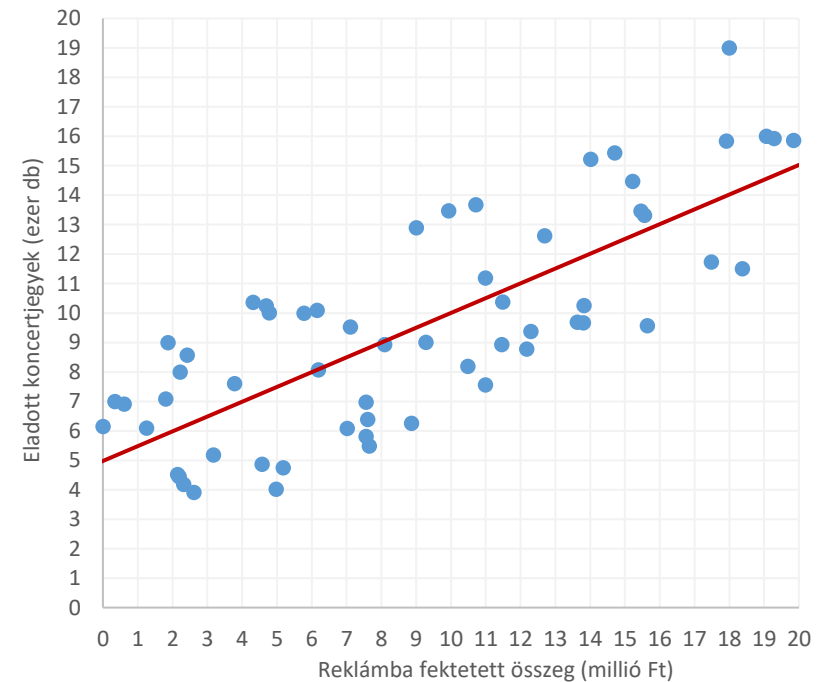
Regressziós egyenes egyenlete

- A regressziós egyenes képlete: $Y_{\text{pred},i} = b_0 + b_1 * X_i$ és a kimeneti változó értéke: $Y_i = Y_{\text{pred},i} + \epsilon_i$
- A regressziós elemzésből tudjuk, hogy $b_0 = 5$ és $b_1 = 0,5$

Azonosító	Prediktor (X)	Kimeneti (Y)	Predikált (Y_{pred})	Hiba (ϵ)
1	4,78	10	7,39	2,61
2	6,16	10,09	8,08	2,01
3	1,87	9	5,935	3,065
4	15,65	9,575	12,825	-3,25
5	13,81	9,665	11,905	-2,24
6	14,02	15,22	12,01	3,21
7	15,47	13,465	12,735	0,73
8	12,7	12,62	11,35	1,27
9	3,79	7,6	6,895	0,705
10	7,61	6,385	8,805	-2,42
11	7,12	9,53	8,56	0,97
12	0,35	7	5,175	1,825
13	15,57	13,315	12,785	0,53
14	2,62	3,91	6,31	-2,4
15	2,15	4,515	6,075	-1,56
16	...	...	...	...

$$Y_{\text{pred},i} = 5 + 0,5 * X_i$$

$$\epsilon_i = Y - Y_{\text{pred},i}$$



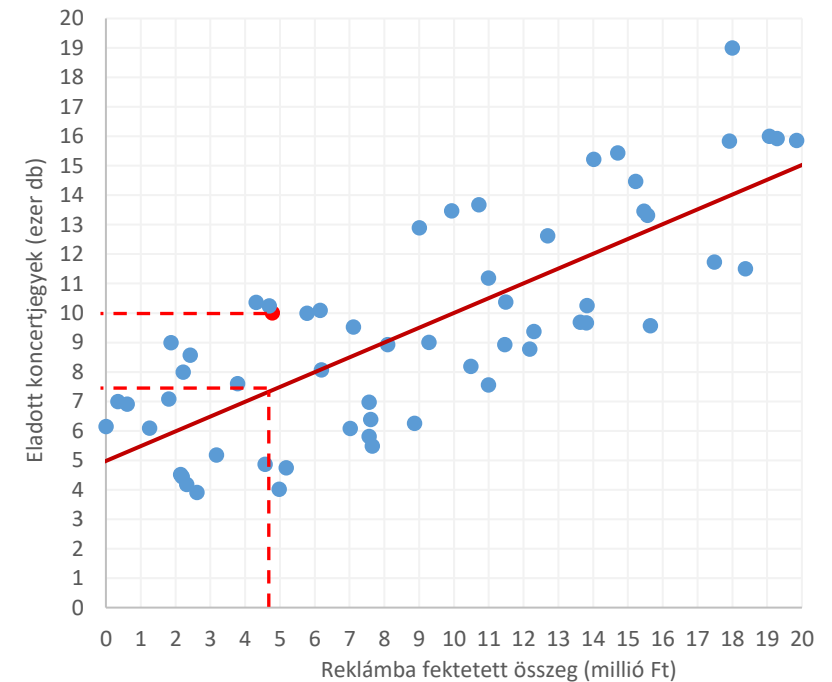
Regressziós egyenes egyenlete

- A regressziós egyenes képlete: $Y_{\text{pred},i} = b_0 + b_1 * X_i$ és a kimeneti változó értéke: $Y_i = Y_{\text{pred},i} + \epsilon_i$
- A regressziós elemzésből tudjuk, hogy $b_0 = 5$ és $b_1 = 0,5$

Azonosító	Prediktor (X)	Kimeneti (Y)	Predikált (Y_{pred})	Hiba (ϵ)
1	4,78	10	7,39	2,61
2	6,16	10,09	8,08	2,01
3	1,87	9	5,935	3,065
4	15,65	9,575	12,825	-3,25
5	13,81	9,665	11,905	-2,24
6	14,02	15,22	12,01	3,21
7	15,47	13,465	12,735	0,73
8	12,7	12,62	11,35	1,27
9	3,79	7,6	6,895	0,705
10	7,61	6,385	8,805	-2,42
11	7,12	9,53	8,56	0,97
12	0,35	7	5,175	1,825
13	15,57	13,315	12,785	0,53
14	2,62	3,91	6,31	-2,4
15	2,15	4,515	6,075	-1,56
16	...	...	...	...

$$Y_{\text{pred},1} = 5 + 0,5 * 4,78 = 7,39$$

$$\epsilon_1 = Y - Y_{\text{pred},1} = 10 - 7,39 = 2,61$$



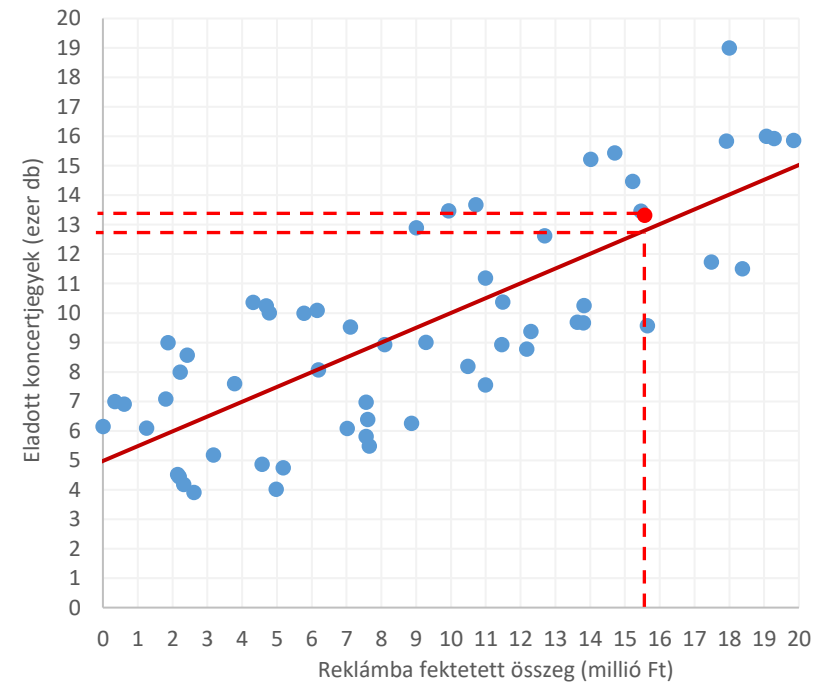
Regressziós egyenes egyenlete

- A regressziós egyenes képlete: $Y_{\text{pred},i} = b_0 + b_1 * X_i$ és a kimeneti változó értéke: $Y_i = Y_{\text{pred},i} + \epsilon_i$
- A regressziós elemzésből tudjuk, hogy $b_0 = 5$ és $b_1 = 0,5$

Azonosító	Prediktor (X)	Kimeneti (Y)	Predikált (Y_{pred})	Hiba (ϵ)
1	4,78	10	7,39	2,61
2	6,16	10,09	8,08	2,01
3	1,87	9	5,935	3,065
4	15,65	9,575	12,825	-3,25
5	13,81	9,665	11,905	-2,24
6	14,02	15,22	12,01	3,21
7	15,47	13,465	12,735	0,73
8	12,7	12,62	11,35	1,27
9	3,79	7,6	6,895	0,705
10	7,61	6,385	8,805	-2,42
11	7,12	9,53	8,56	0,97
12	0,35	7	5,175	1,825
13	15,57	13,315	12,785	0,53
14	2,62	3,91	6,31	-2,4
15	2,15	4,515	6,075	-1,56
16	...	...	...	...

$$Y_{\text{pred},7} = 5 + 0,5 * 15,47 = 12,735$$

$$\epsilon_7 = Y - Y_{\text{pred},7} = 13,465 - 12,735 = 0,73$$



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 együtthatói (koefficiensei)

- b_0 az egyenes kezdőpontja, és b_1 az egyenes meredeksége
- ha a prediktor 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.
- Standardizált együtthatók – b_1 meredekség skálafüggetlen verziói, ezek segítségével hasonlíthatók össze a prediktorok.

Regressziós mutatók az életben

Standardizálatlan regressziós együtthatók

Ezek adja meg a regressziós egyenesünk kezdőpontját és meredekségét, tehát ez alapján tudjuk kiszámolni a predikcióinkat

R^2 , F érték és a hozzá tartozó szignifikanca

Ez alapján tudod eldönteni, mennyire megbízható az általad felépített modell, mennyire hihetsz a modell alapján tett predikcióknak

Standardizált regressziós együttható és t-próba

Több prediktor változó esetén ezek alapján döntesz, hogy melyik prediktor változók a leghasznosabbak számodra, melyek befolyásolják a kimeneti változót jelentős mértékben, melyekből lehet a kimeneti változó értékét leginkább bejósolni.

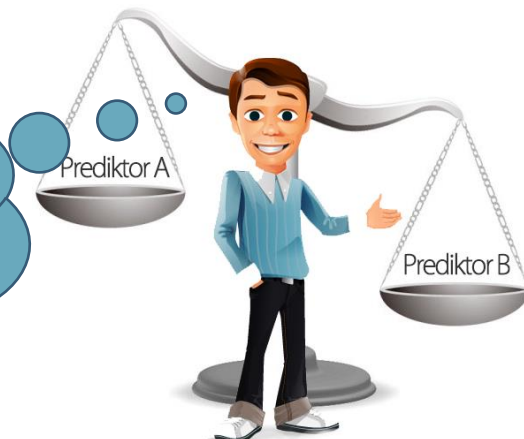


Két milliót akarok a reklámra fordítani, akkor várhatóan $5 + 0,5 * 2 = 6$ ezer jegyet fogok eladni. Hmm... akkor a Copper Box a megfelelő méretű helyszín!



A modell elég megbízhatónak tűnik, ezért bátran lefoglalom a Copper Box-ot. Bízok benne, hogy meg fog telni!

A modell túl megbízhatatlan. Ha lefoglalom a CB-t, elég nagyot bukhatok. Kellene néhány jobb prediktor!



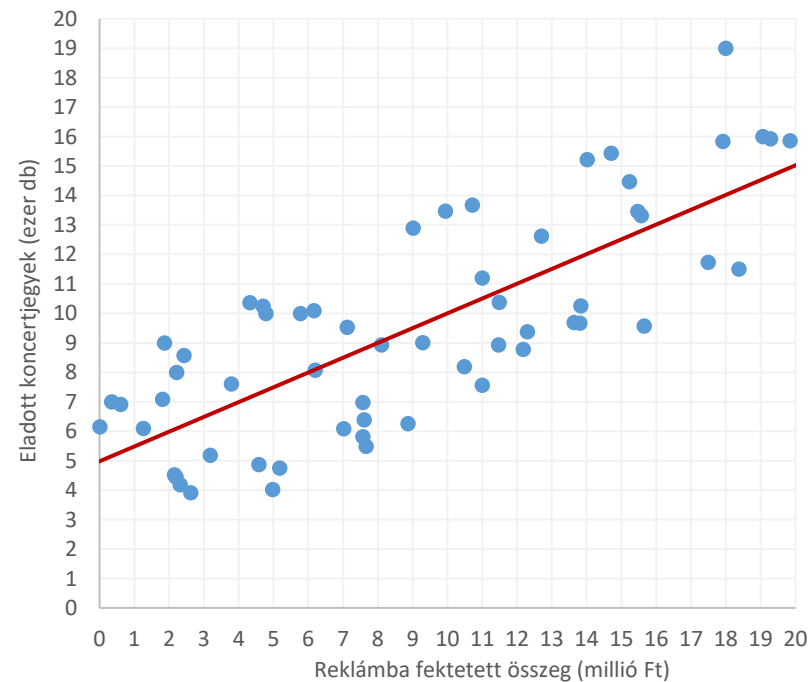
Úgy tűnik, ma már nem számít annyira, hányszor említi a zenekart a Metal Hammer, sokkal inkább, hogy hányan nézik meg Youtube-on. Akkor az alapján számolom a predikciókat!

Egyszerű lineáris regresszió



Egyszerű lineáris regresszió

- Egyszerű lineáris regresszióról beszélünk, ha *egy* prediktor és *egy* kimeneti változó közötti lineáris összefüggést keressük
 - Az egyszerű lineáris regresszió megegyezik a Pearson korrelációs eljárással
- Próbáljuk meg bejósolni a reklámba fektetett összegből azt, hogy hány koncertjegyet sikerült eladni!
 - Feltételek: a lineáris regresszió feltételeit majd többszempontos esetben nézzük át, most elegendő a Pearson korreláció öt feltétele: skála típusú adatok, független kitöltők, normál eloszlás, homoszkedaszticitás (=szóráshomogenitás) és linearitás.



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)

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	471,583	1	471,583	88,447	,0000000000000339 ^b
Residual	303,913	57	5,332		
Total	775,496	58			

a. Dependent Variable: Eladott koncertjegy (1000 db)

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

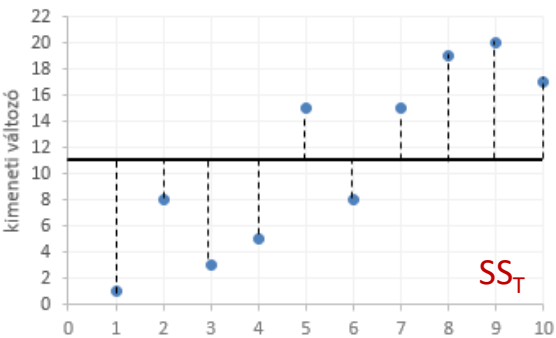
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,983	,565		8,817	,000000000000307
	Reklámba fektetett összeg (millió Forint)	,502	,053	,780	9,405	,000000000000339

a. Dependent Variable: Eladott koncertjegy (1000 db)

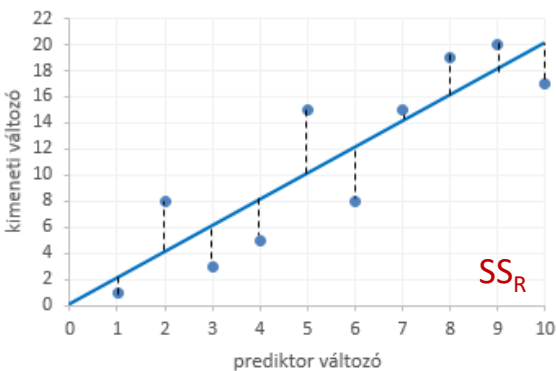
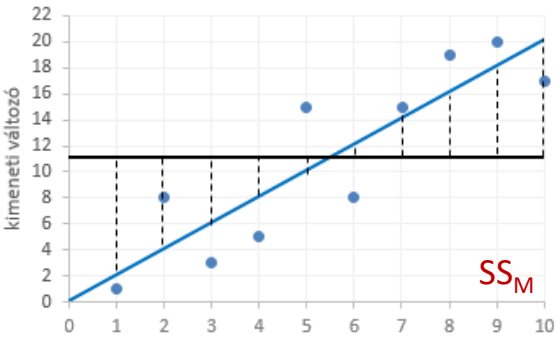
Egyszerű lineáris regresszió

A *Model Summary* és *ANOVA* tábla tartalmazza a modell „jóságának” mutatóit.



$$R^2 = \frac{SS_M}{SS_T} \quad \text{és} \quad F = \frac{MS_M}{MS_R}$$

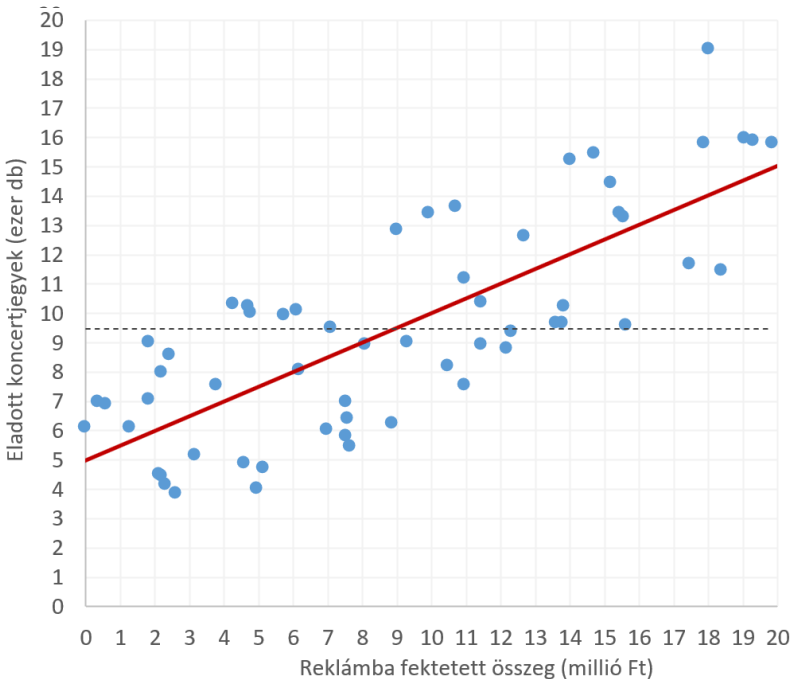
N = 59



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,780 ^a	R² ,608	,601	2,309072

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



ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	SS_M 471,583	df_M 1	MS_M 471,583	F 88,447	,0000000000000339^b
Residual	SS_R 303,913	df_R 57	MS_R 5,332		
Total	SS_T 775,496	df_T 58			

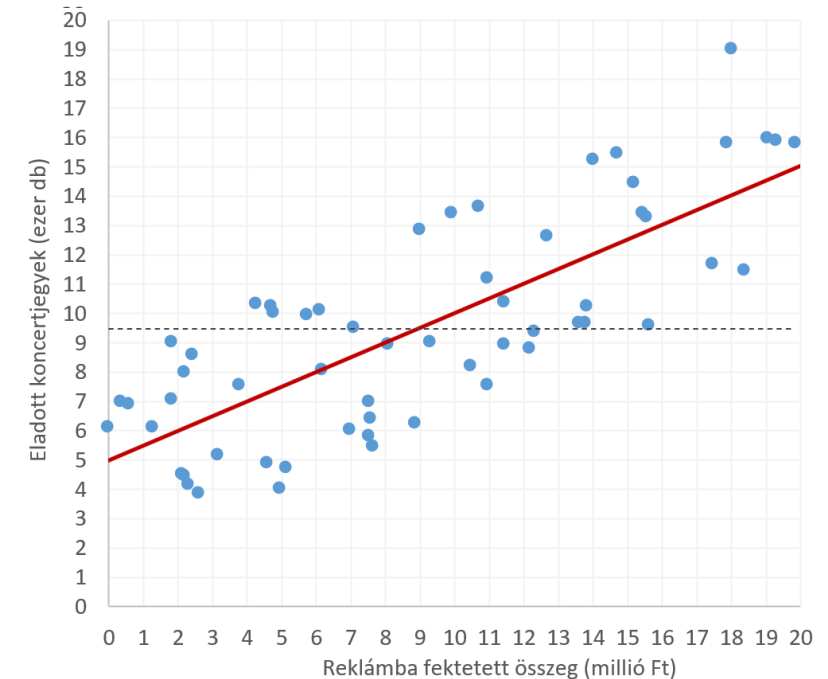
a. Dependent Variable: Eladott koncertjegy (1000 db)

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

azaz $\frac{339}{\text{billiárd}}$

Egyszerű lineáris regresszió

- A *Koefficiens* tábla tartalmazza a regressziós egyenes együtthatóit – ezáltal a prediktor változó(k) egyenkénti hatását a kimeneti változóra.
- A regressziós egyenes egyenlete: $Y_{\text{predikált}} = b_0 + b_1 * x$
- A **b0** az $x=0$ -hoz tartozó értéket adja meg (ahol az y tengelyt metszi az egyenes).
- A **b1** az egyenes meredekségét adja meg standardizálatlan formában (a prediktor egységnyi változására kimeneti változóban bekövetkező változást).
- Az **SE** az együtthatók bizonytalanságát jelzi.

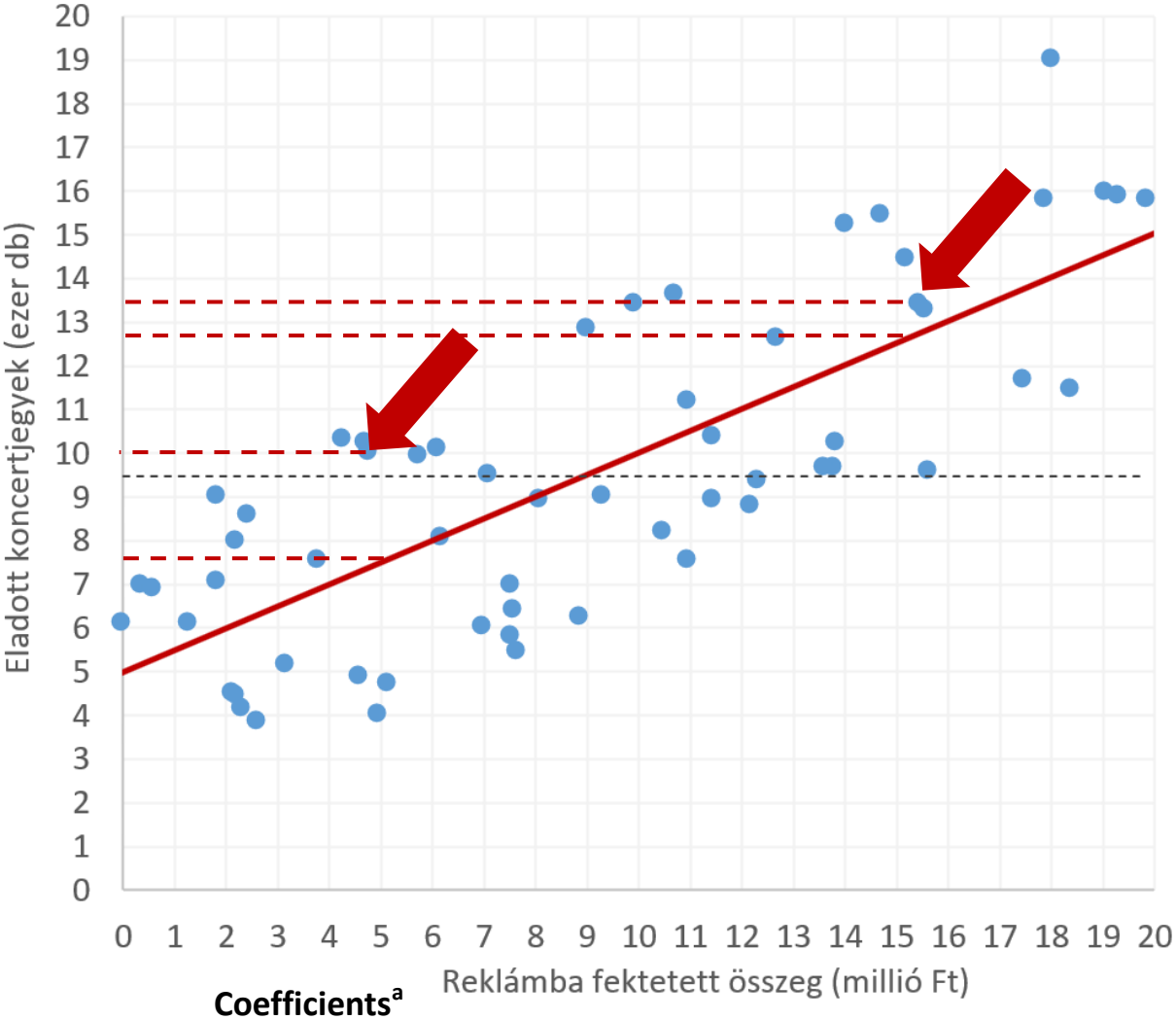


Többszempon-
tós elemzésnél minden
prediktor itt lesz felsorolva

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	b0 4,983	,565		8,817	,000000000000307
	Reklámba fektetett összeg (millió Forint)	b1 ,502	,053	,780	9,405	,000000000000339

a. Dependent Variable: Eladott koncertjegy (1000 db)

Azonosító	Prediktor (X)	Kimeneti (Y)	Predikált (Y_{pred})
1	4,78	10	7,39
2	6,16	10,09	8,08
3	1,87	9	5,935
4	15,65	9,575	12,825
5	13,81	9,665	11,905
6	14,02	15,22	12,01
7	15,47	13,465	12,735
8	12,7	12,62	11,35
9	3,79	7,6	6,895
10	7,61	6,385	8,805
11	7,12	9,53	8,56
12	0,35	7	5,175
13	15,57	13,315	12,785
14	2,62	3,91	6,31
15	2,15	4,515	6,075
16	...	...	...



Többszempon-
tanal minden
prediktor itt lesz felsorolva

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	b0 4,983	,565		8,817	,00000000000307
	Reklámba fektetett összeg (millió Forint)	b1 ,502	,053	,780	9,405	,000000000000339

a. Dependent Variable: Eladott koncertjegy (1000 db)

Egyszerű lineáris regresszió

R – korreláció a mért kimeneti értékek és predikált értékek között

R^2_{adj} - az R2 korrekciója az elemszámot és a prediktorok számát figyelembe véve. Megadja, hogy mekkora magyarázóerőre számíthatunk a populációban

Kimeneti (Y)	Predikált (Y_{pred})
10	7,39
10,09	8,08
9	5,935
9,575	12,825
9,665	11,905
15,22	12,01
13,465	12,735
12,62	11,35
7,6	6,895
6,385	8,805
9,53	8,56
7	5,175
13,315	12,785
3,91	6,31
4,515	6,075
...	...

Model Summary

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

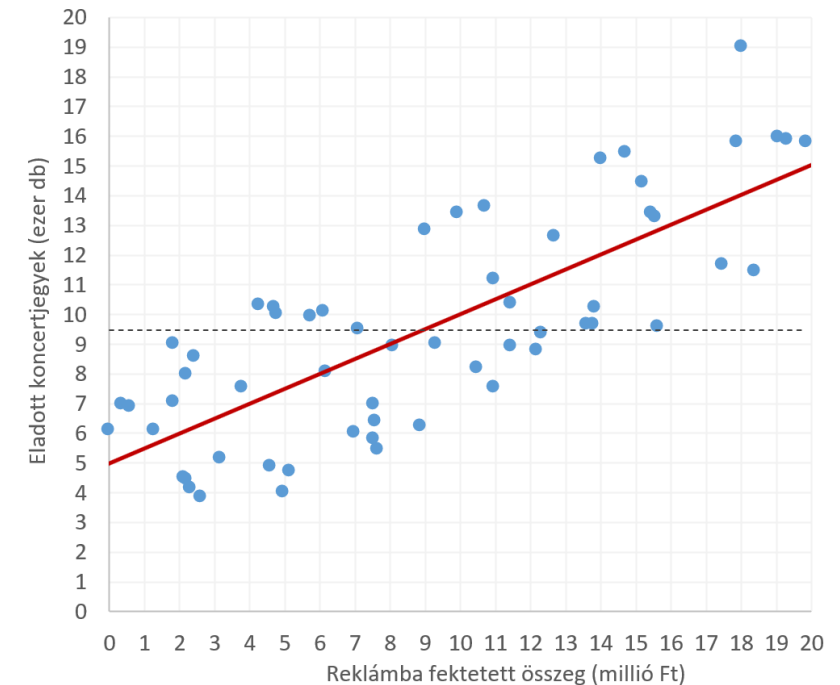
a. Predictors: (Constant), Reklámba fektetett összeg (millió Forint)

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	SS_M 471,583	df_M 1	MS_M 471,583	F 88,447	,000000000000339 ^b
Residual	SS_R 303,913	df_R 57	MS_R 5,332		
Total	SS_T 775,496	df_T 58			

a. Dependent Variable: Eladott koncertjegy (1000 db)

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

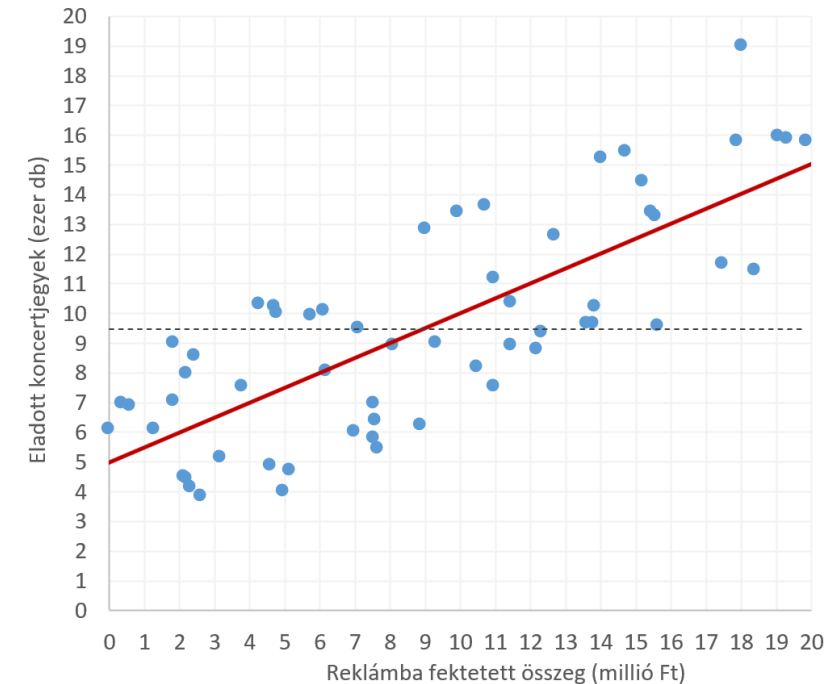


azaz $\frac{339}{\text{billiárd}}$

Egyszerű lineáris regresszió

A β (táblázatban Beta) a standardizált meredekség, mely alapján több prediktor hatása összehasonlítható.

A **t-próbával** és a hozzá tartozó szignifikanciával ellenőrizhetjük, hogy a meredekség szignifikánsan eltér-e nullától, azaz a prediktornak van-e szignifikáns hatása a kimeneti változóra



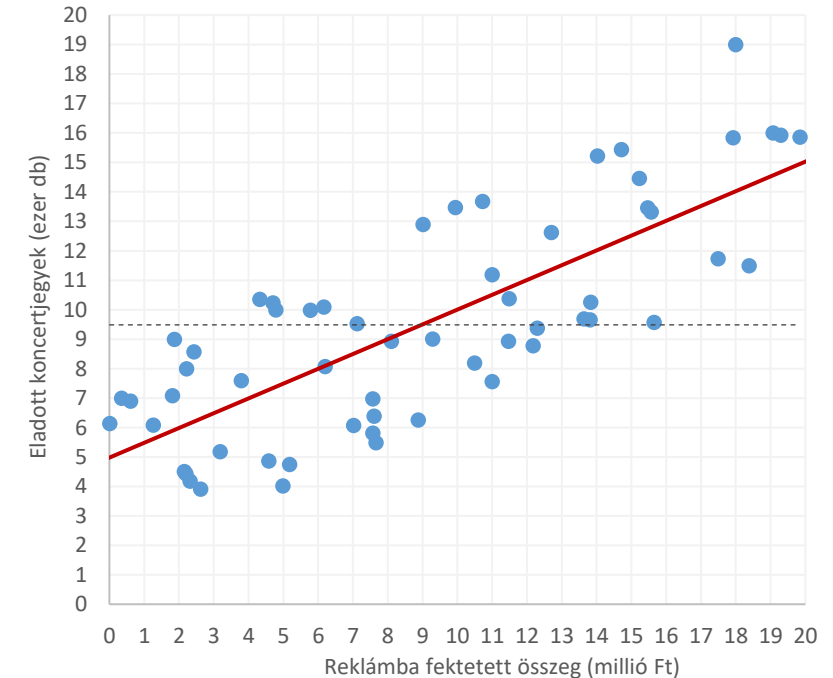
Többszempon-
tós elemzésnél minden
prediktor itt lesz felsorolva

Coefficients ^a							
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
		B	Std. Error	Beta			
1	(Constant)	b0	4,983		8,817	,000000000000307	
	Reklámba fektetett összeg (millió Forint)	b1	,502				
				β1	,780	9,405	,0000000000000339

a. Dependent Variable: Eladott koncertjegy (1000 db)

Egyszerű lineáris regresszió

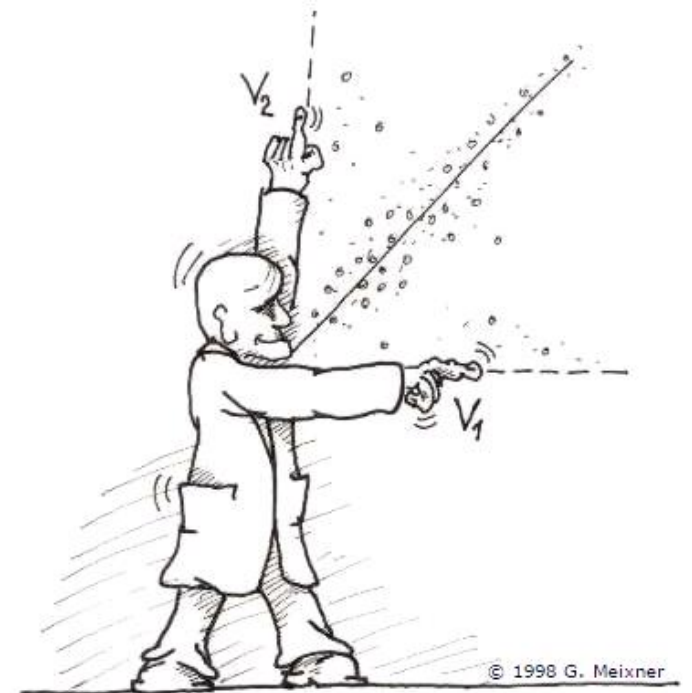
- Néhány megjegyzés:
 - Kapcsolat Pearson korrelációval
 - A *Model Summary* tábla R értéke megegyezik a Pearson-korrelációból kapható r-értékkel, hiszen az egyprediktoros lineáris regresszió „csupán” egy Pearson korreláció.
 - Az *ANOVA* tábla szignifikancia értéke is azonos a Pearson korreláció 2-tailed szignifikanciájával.
 - Kapcsolat F-érték és a b1 koefficiens között
 - Az *ANOVA* táblában a prediktorhoz tartozó t-próba t-értékének négyzete megegyezik az F értékkel, hiszen a $df_M = 1$.
 - Ennek gyakorlati jelentése az, hogy a kimeneti változó varianciájának modellel megmagyarázható részének egészéért az egy darab prediktor változó felelős.
 - Ez többszemponos elemzésben már nem lesz igaz – ott több prediktor hatásának eredője lesz a regressziós egyenes, melyhez az F-érték tartozik.



Correlations		
		Eladott koncertjegy (1000 db)
Reklámba fektetett összeg (millió Forint)	Pearson Correlation	,780**
	Sig. (2-tailed)	,0000000000000339
	N	59

** . Correlation is significant at the 0.01 level (2-tailed).

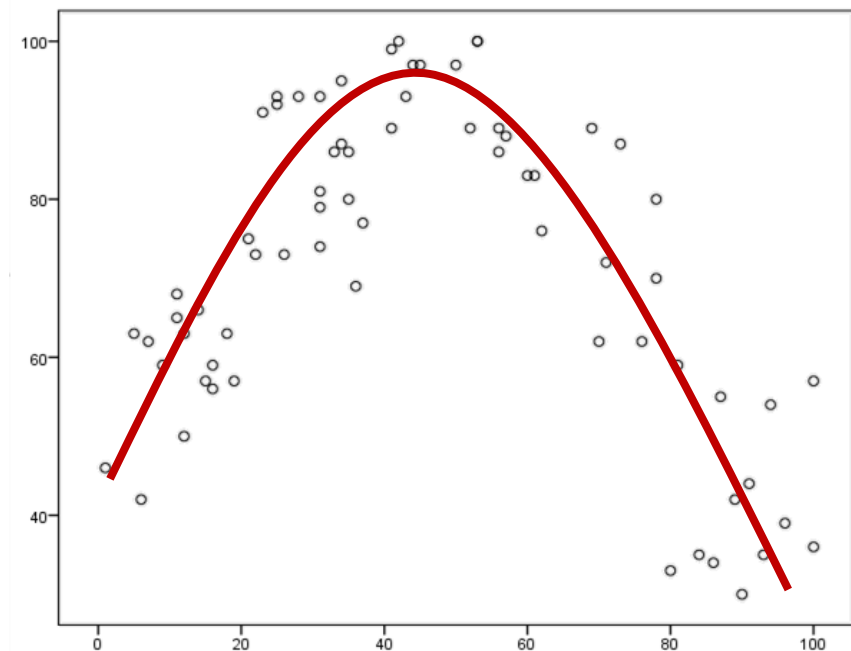
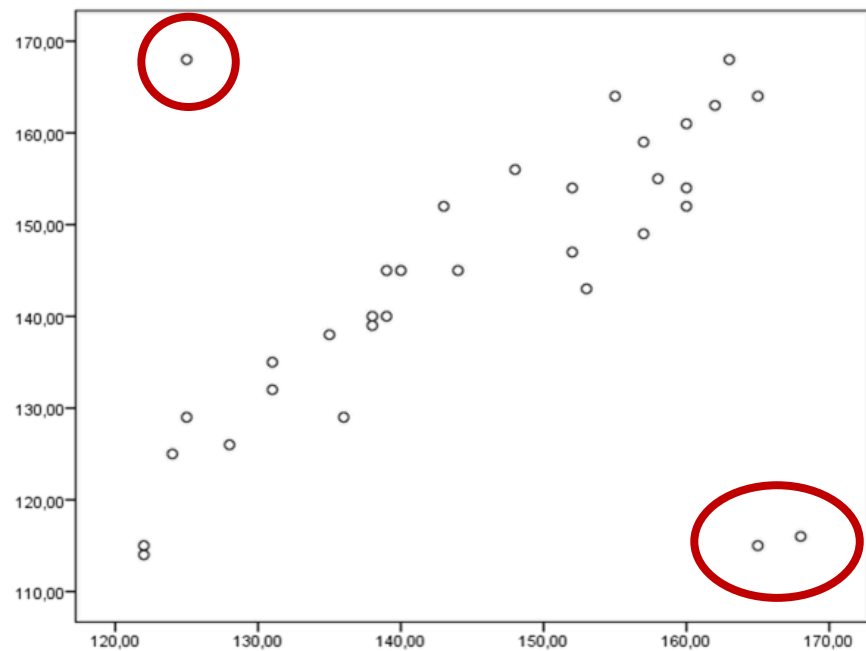
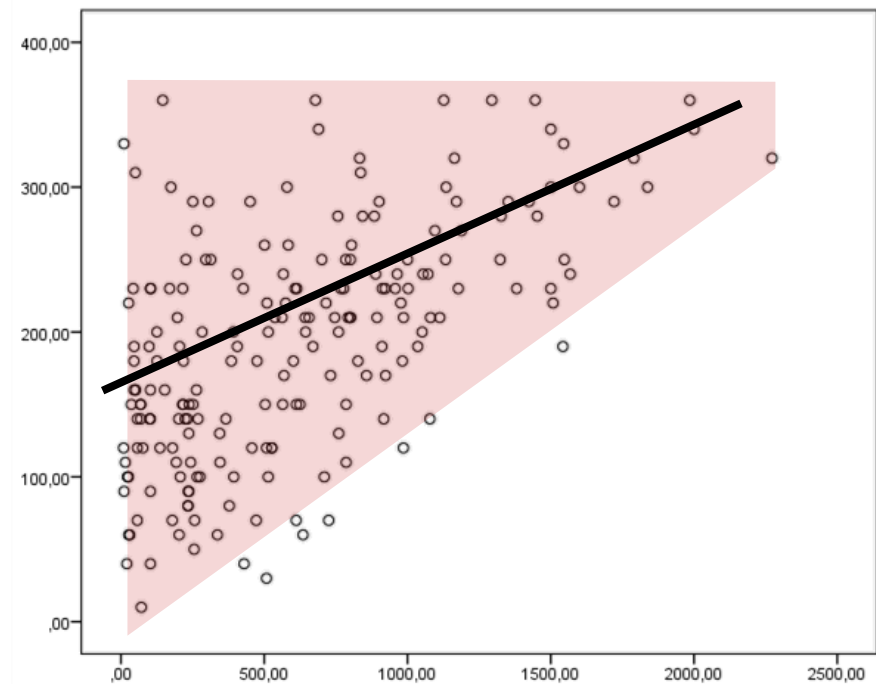
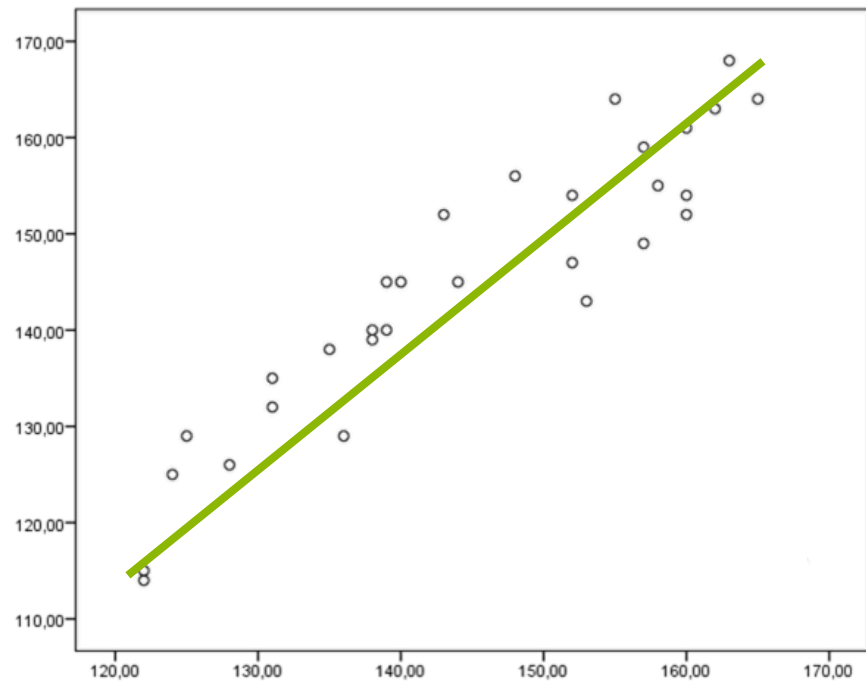
Egyszerű lineáris regresszió JASP-ben



A reklámba fektetett összeg pozitív hatással van az eladott koncertjegyek számára.

Egyszerű regresszióban a feltételek megegyeznek a Pearson korreláció feltételeivel!

- Függelenség
- Skála mérési szint
- Normalitás
- Szóráshomogenitás
- Linearitás



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	0.780	0.608	0.601	2.309

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H ₁	Regression	471.583	1	471.583	88.447	< .001
	Residual	303.913	57	5.332		
	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)	4.983	0.565		8.817	< .001
	Reklam	0.502	0.053	0.780	9.405	< .001

Gyakorló példák

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,469 ^a	,220	,186	7,32620

a. Predictors: (Constant), pred1

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	347,355	1	347,355	6,472	,018 ^b
	Residual	1234,485	23	53,673		
	Total	1581,840	24			

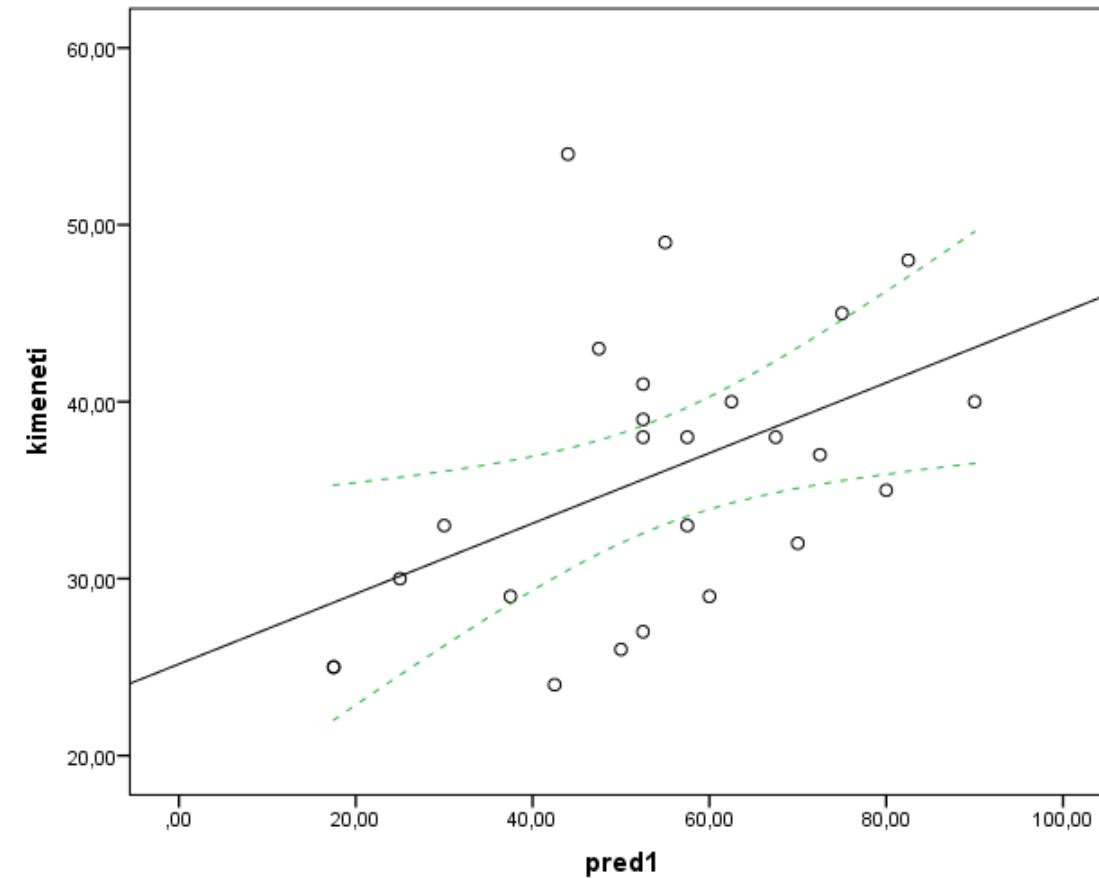
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred1

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	25,175	4,471		5,631	,000
	pred1	,199	,078	,469	2,544	,018

a. Dependent Variable: kimeneti



Gyakorló példák

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,673 ^a	,453	,429	6,13531

a. Predictors: (Constant), pred2

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	716,075	1	716,075	19,023	,000 ^b
	Residual	865,765	23	37,642		
	Total	1581,840	24			

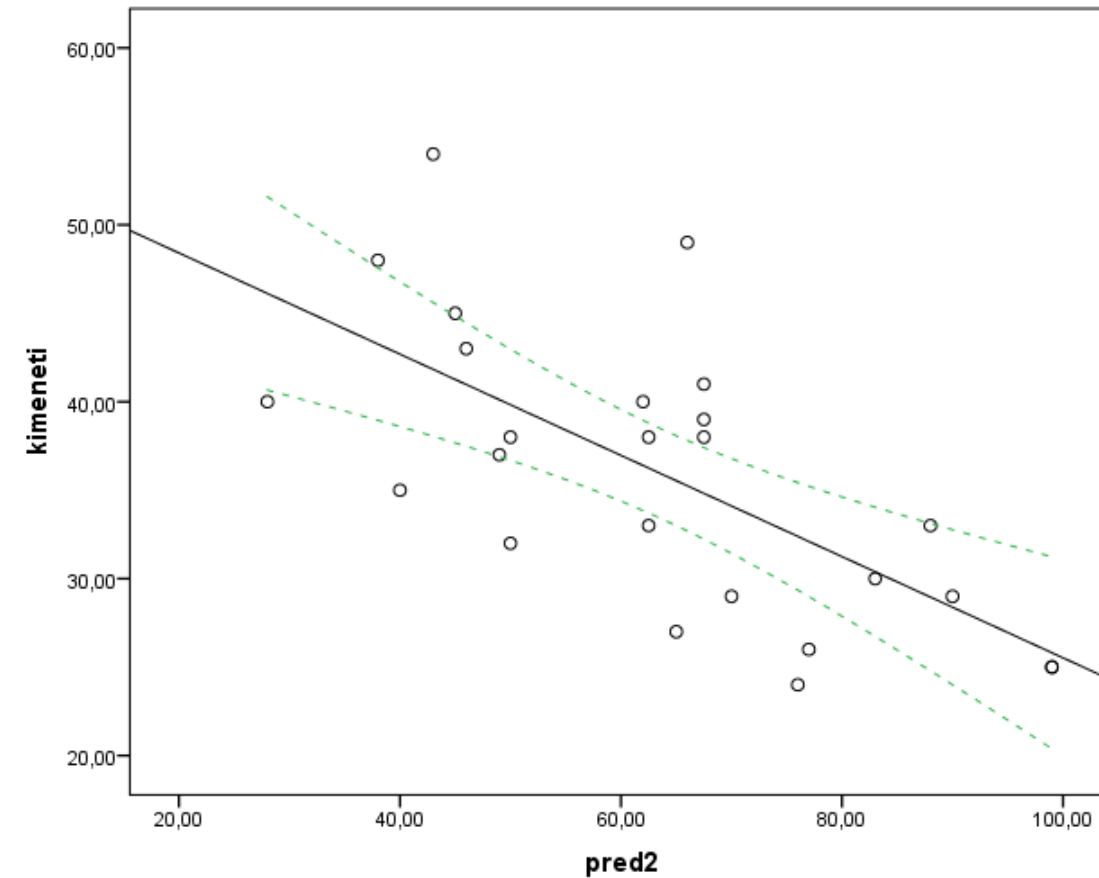
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred2

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	54,131	4,352		12,438	,000
	pred2	-,286	,066	-,673	-4,362	,000

a. Dependent Variable: kimeneti



Gyakorló példa

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,580 ^a	,337	,226	8,70704

a. Predictors: (Constant), pred3

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	231,125	1	231,125	3,049	,131 ^b
	Residual	454,875	6	75,813		
	Total	686,000	7			

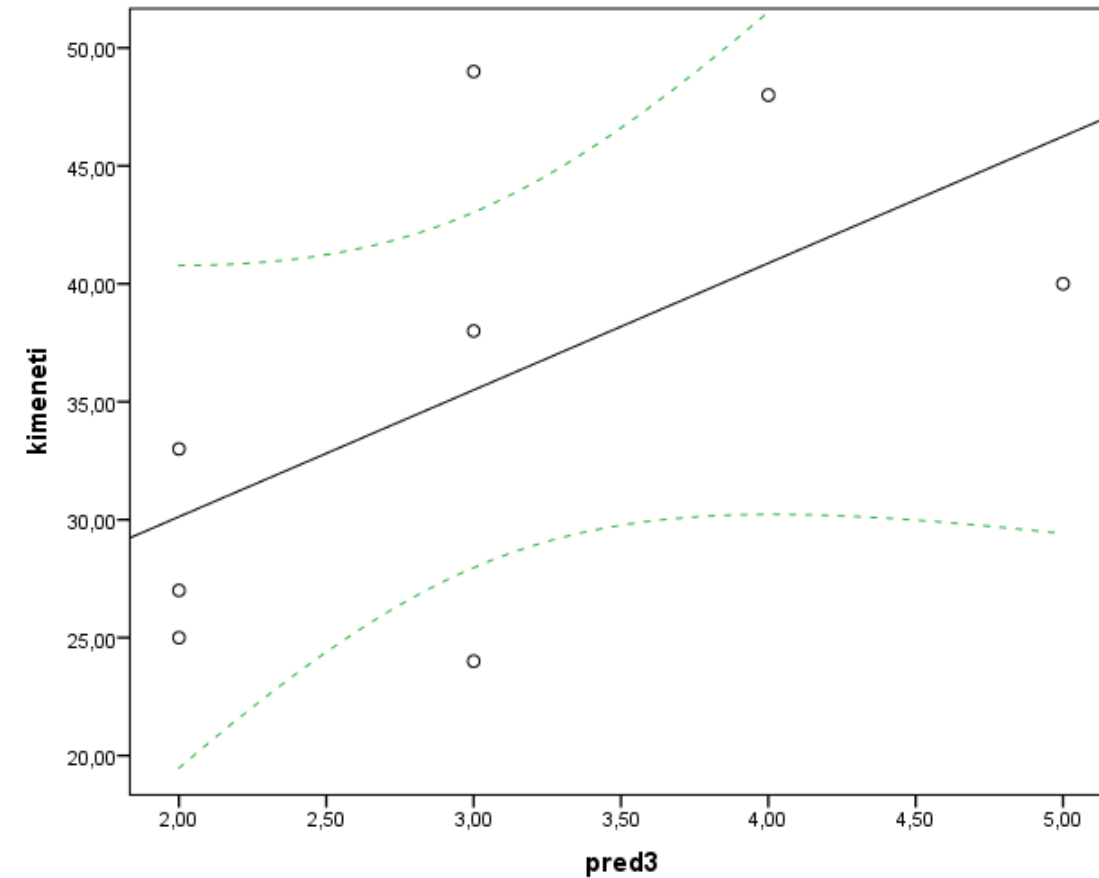
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred3

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19,375	9,735		1,990	,094
	pred3	5,375	3,078	,580	1,746	,131

a. Dependent Variable: kimeneti



Gyakorló példák

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,229 ^a	,052	,043	8,21597

a. Predictors: (Constant), pred4

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	366,180	1	366,180	5,425	,022 ^b
	Residual	6615,210	98	67,502		
	Total	6981,390	99			

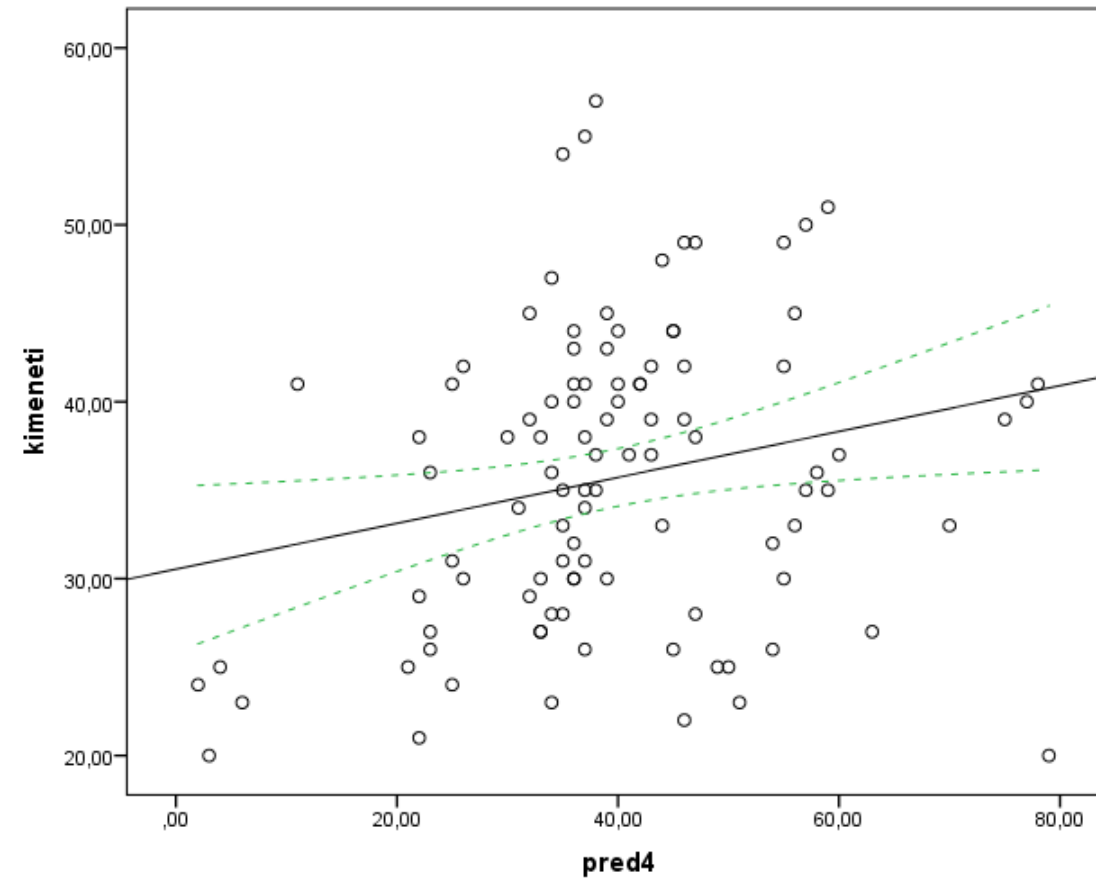
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred4

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	30,536	2,361		12,936	,000
	pred4	,130	,056	,229	2,329	,022

a. Dependent Variable: kimeneti



Gyakorló példák

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,673 ^a	,453	,430	6,13142

a. Predictors: (Constant), pred5

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	717,170	1	717,170	19,077	,000 ^b
	Residual	864,670	23	37,594		
	Total	1581,840	24			

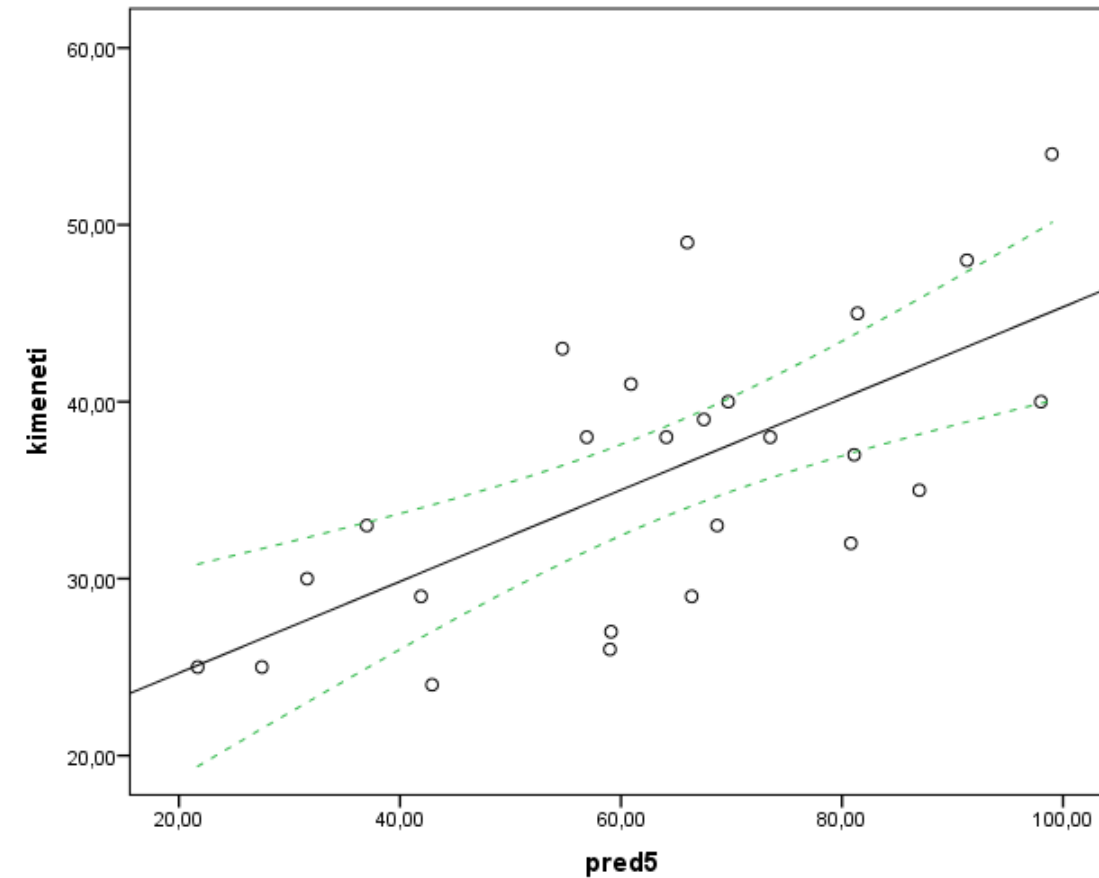
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred5

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19,496	3,955		4,929	,000
	pred5	,259	,059	,673	4,368	,000

a. Dependent Variable: kimeneti



Gyakorló példák

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,745 ^a	,555	,547	5,71458

a. Predictors: (Constant), pred6

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2158,739	1	2158,739	66,105	,000 ^b
	Residual	1730,789	53	32,656		
	Total	3889,527	54			

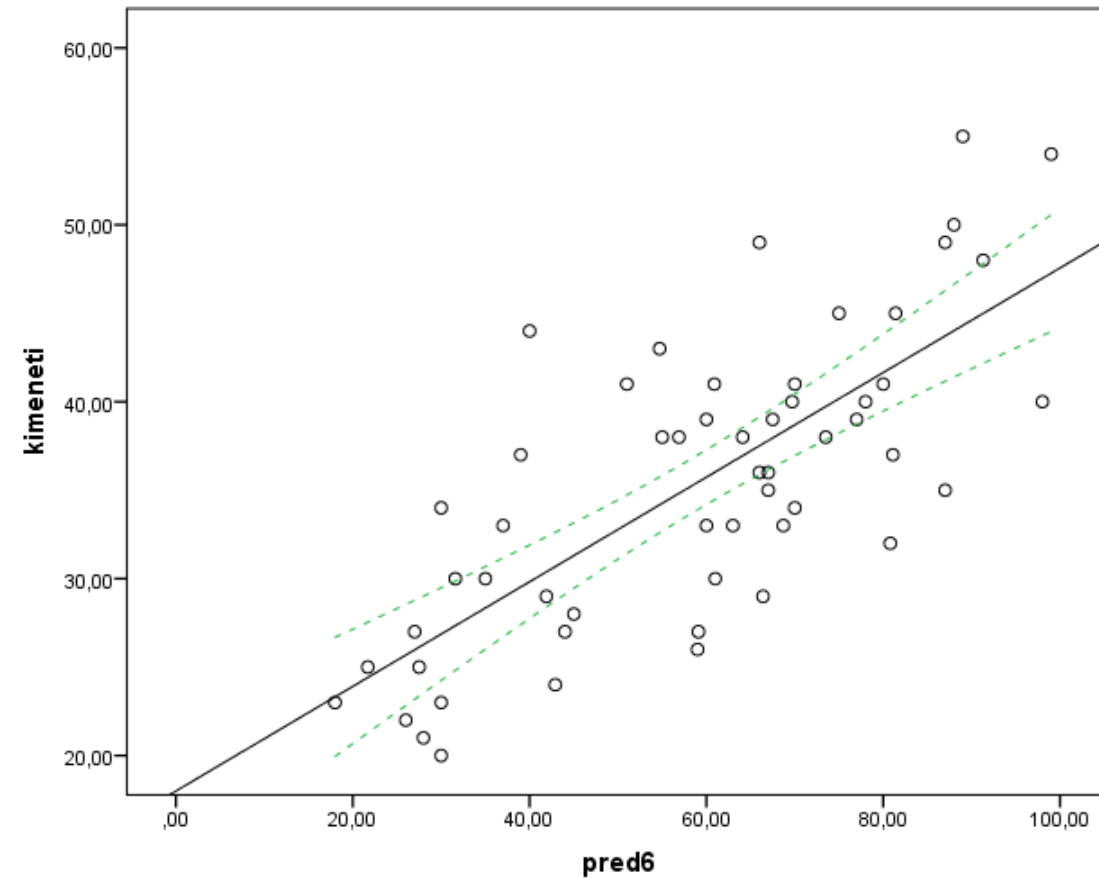
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred6

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17,999	2,279		7,898	,000
	pred6	,296	,036	,745	8,130	,000

a. Dependent Variable: kimeneti



Gyakorló példák

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,285 ^a	,081	,041	7,95030

a. Predictors: (Constant), pred7

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	128,074	1	128,074	2,026	,168 ^b
	Residual	1453,766	23	63,207		
	Total	1581,840	24			

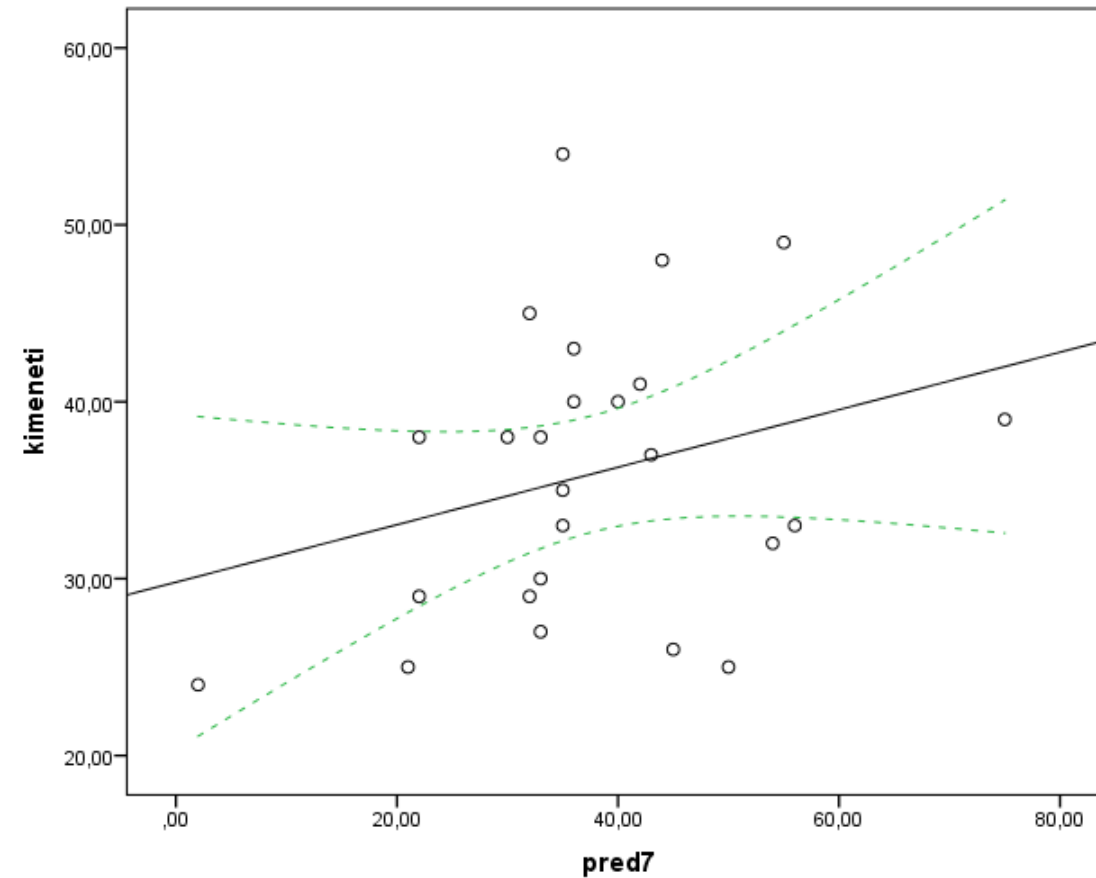
a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred7

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	29,805	4,581		6,506	,000
	pred7	,162	,114	,285	1,423	,168

a. Dependent Variable: kimeneti



Gyakorló példa

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,749 ^a	,560	,552	5,68013

a. Predictors: (Constant), pred8

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2179,539	1	2179,539	67,553	,000 ^b
	Residual	1709,988	53	32,264		
	Total	3889,527	54			

a. Dependent Variable: kimeneti

b. Predictors: (Constant), pred8

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	59,591	3,037		19,622	,000
	pred8	-,285	,035	-,749	-8,219	,000

a. Dependent Variable: kimeneti

