

Többszörös lineáris regresszió.

Confounding és többszörös lineáris regresszió.

előadó: VERES Dániel

2021. november 3.

Kérdések

Orvostudomány igen gyakori kérdései (egyszerűsítve):

- Hogyan hat egy rizikófaktor egy klinikai paraméterre?
 - Életkor hogyan hat a koronavírusok számára?
 - Életkor hogyan hat a COVID tünetek súlyosságára?
 - Testmagasság hogyan hat az alvászavarra?
 - Hormonszint hogyan befolyásolja a CV rizikót?
- Hogyan hat egy kezelés egy klinikai paraméterre?
 - Oltás hogyan hat a koronavírusos betegek számára?
 - Hörgtágító szedése hogyan hat az alvászavarra?
 - Hormonterápia hogyan befolyásolja a CV rizikót?

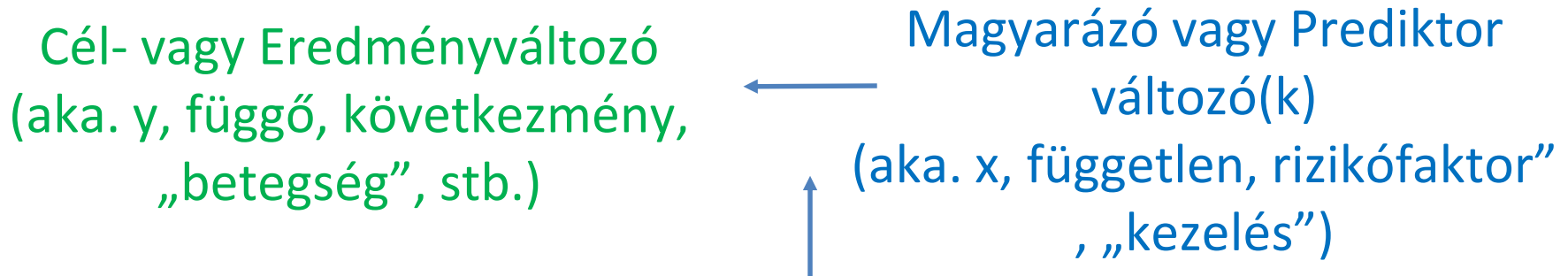
Általánosságban: valamilyen ok-okozat hatásra vagyunk kíváncsiak

Magyarázzuk, illetve következtessünk

Általánosságban: valamilyen ok-okozat hatásra vagyunk kíváncsiak:

- Tényleg magyarázható-e az „okozat az okkal”?
- Mekkora mértékben változik az „ok változására az okozat”?

Mérhető (megfigyelhető) mennyiséggel - azaz változókkal kifejezve:



Mekkora a hatás („van”-e egyáltalán lényeges mértékben)

Mekkora a hatás? – Vizsgáljuk regresszióval!

Ismétlés I. - Regresszió

Függvény kapcsolat (NEM szimmetrikus) egy függő (cél, eredmény, Y) változó és egy független (magyarázó, prediktáló, X) változó(k) között. [Y véletlen változó, X nem feltétlenül]

Y függ X-től – a függőségi viszony iránya **klinikai**lag feltételezett, statisztikaiilag nem vizsgálható.

Kapcsolódó kérdések:

- Van (adott típusú) függvényszerű kapcsolat? (statisztikai kapcsolat! Feltételezve az ok-okozatot!)
- Mekkora Y értéke, ha X:...? (becslés)
- Mekkora X értéke, ha Y:...? (becslés)
- Milyen függvény írja le legjobban Y X-től való függését?

Ismétlés I. – Lineáris Regresszió

Az egyenes becslésére: OLS (Ordinary Least Square method – azaz legkisebb négyzetek módszere)

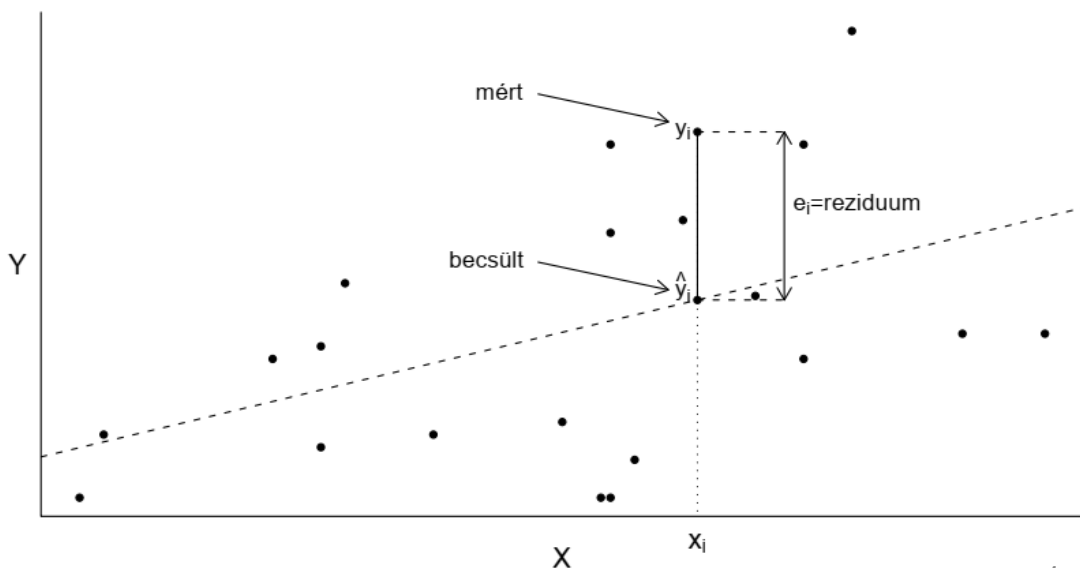
tengelymetszet

Merekség

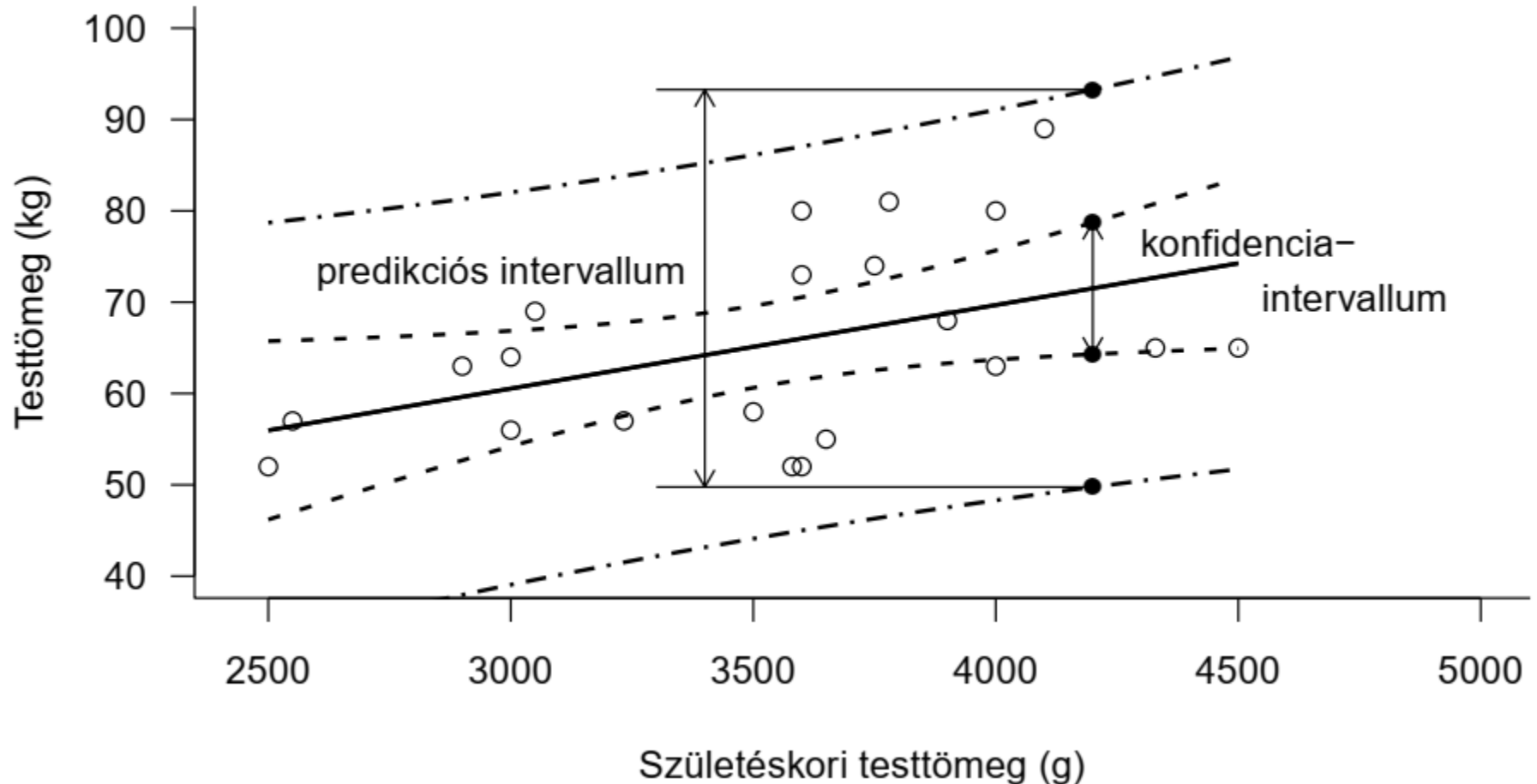
Lineáris függvény: $Y = \beta_0 + \beta_1 * X + \epsilon$

Hibatag; reziduum: pont-egyenes függőleges távolsága
(a becült és mért értékek különbsége)

Az OLS szerinti legjobb egyenes az, ahol a legkisebb a pont-egyenes függőleges távolságok négyzetösszege.



Ismétlés I. – Lineáris Regresszió

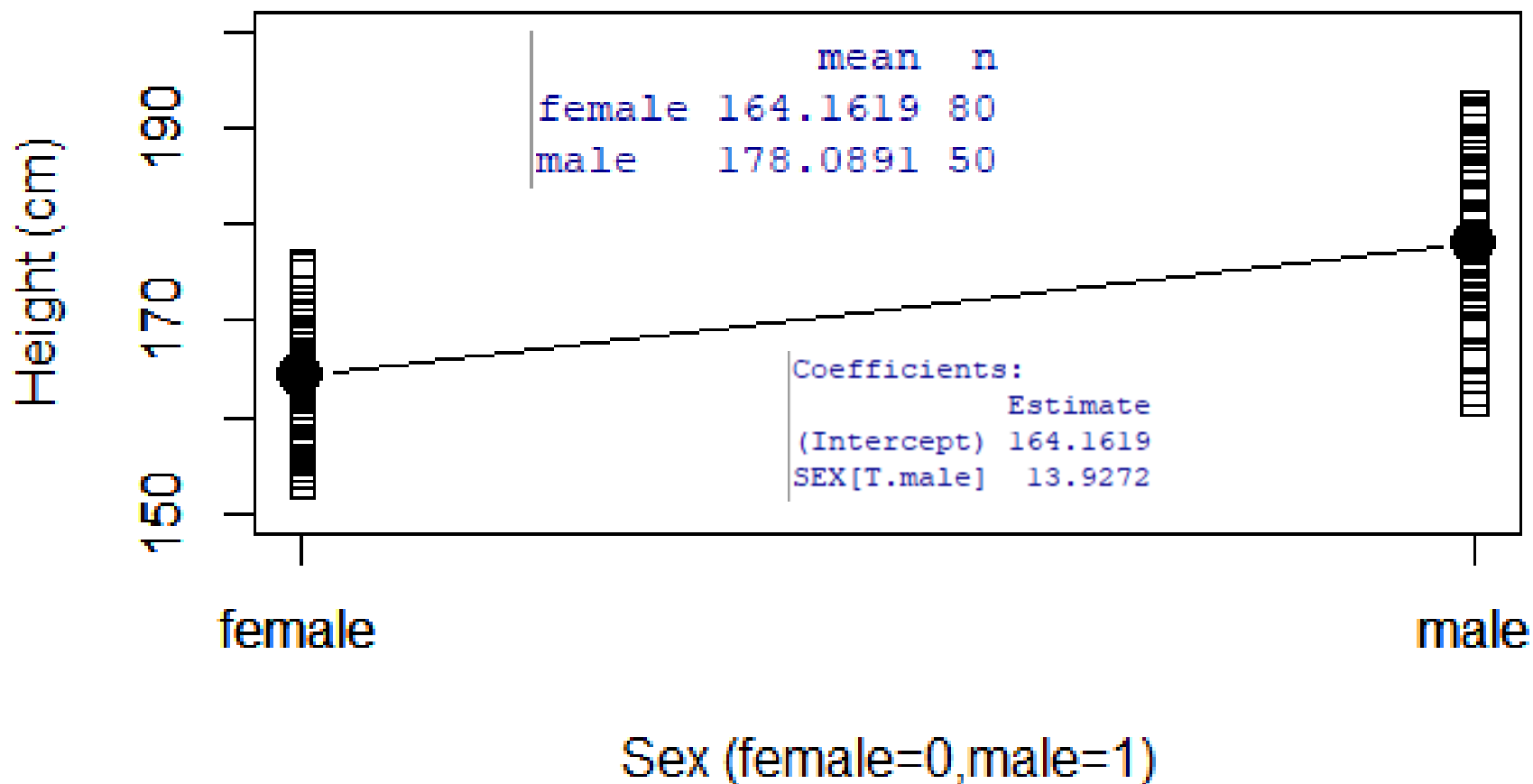


Merekség - hatásmutató:

Y-ban bekövetkező átlagos változás,

X-ben bekövetkező egységnyi változásra

Lineáris regresszió – kategoriális változó?



Merekség:

Y-ban bekövetkező átlagos változás,

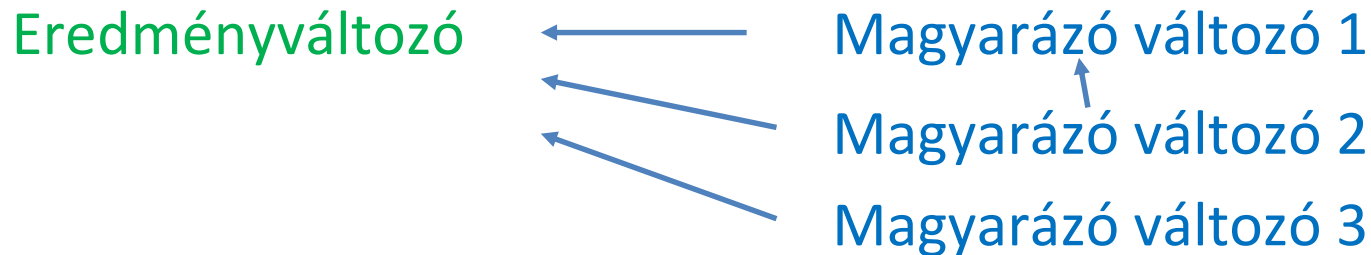
X egyik (0) kategóriájából a másik (1) kategóriába lépéskor

Bonyodalmak!

- **Korreláció nem jelent ok-okozatot!**

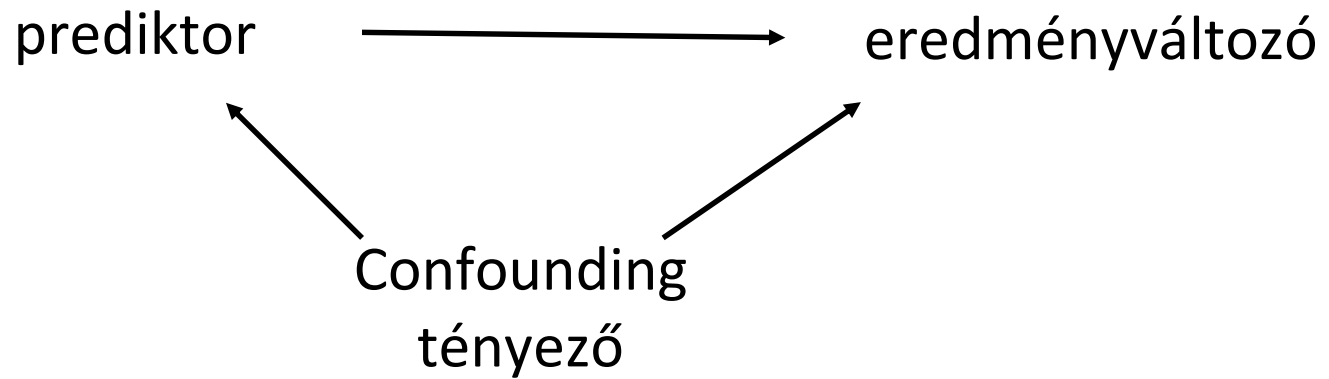
ha a prediktor értéke együtt változik az eredményváltozó értékével, nem jelenti, hogy a prediktor okozza a változást

- **Több prediktor** (általában ez a szituáció)

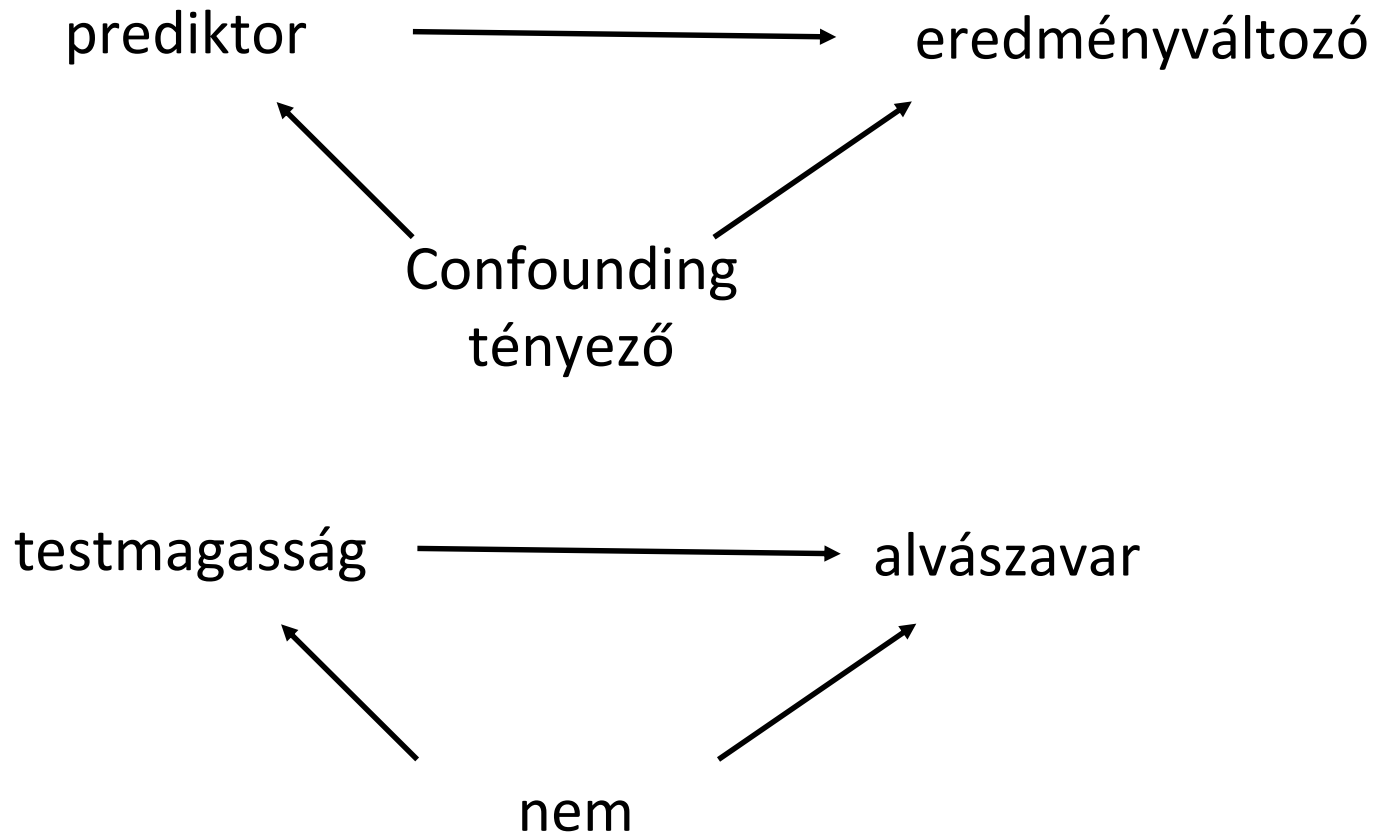


- **Confounding**

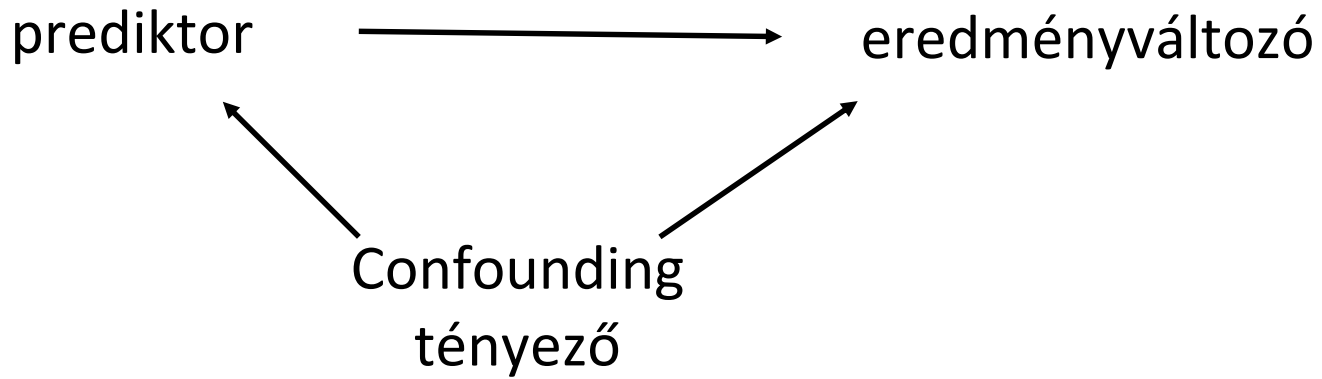
Counfounding



Counfounding



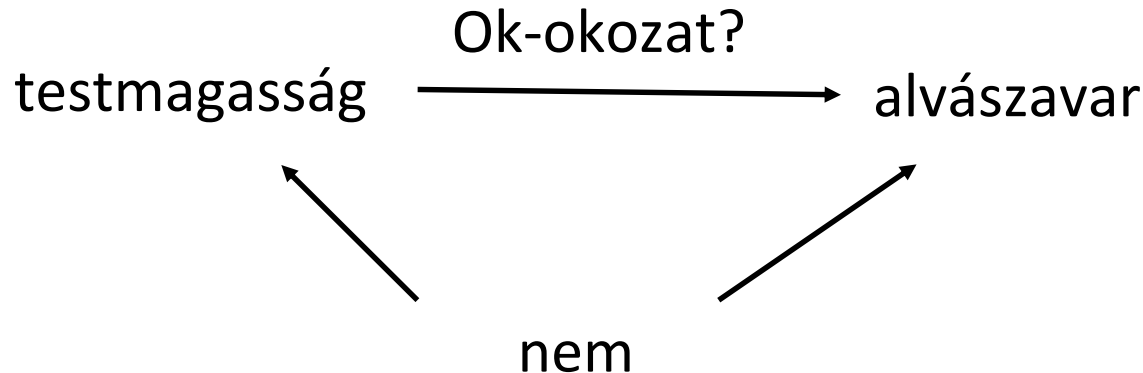
Counfounding



Minket a prediktor közvetlen hatása érdekel

– „függetlenül” a confoundigtól (és a többi prediktortól)!

Bonyodalmak



A kérdés NEM:

ha magasabb, akkor gyakoribb az alvászavar

HANEM:

ha **csak** a testmagasságban van különbség (nő), **akkor** gyakoribb az alvászavar

Nem *csak abban* van eltérés - *TORZÍTÁSOK*

Kiválasztási hiba:

Különbség van a felmérésbe beválasztottak és nem beválasztottak vagy a beválasztottak csoportba sorolása között (egy kimenetelt befolyásoló paraméter tekintetében)

tipikus hibák: életkor, nem eltérő a csoportokban
alappopuláció eltérő
utánkövetés eltérő

Információs hiba:

Téves az alanyoktól kapott, alanyokról gyűjtött információ
(, amely befolyásolja a kimenetelt)

tipikus hibák: visszaemlékezés rossz
elfogultság

Módszerek és hibák

Először vegyük ismét sorra, mik a klinikai vizsgálatok alapvető típusai (study designs):

Megfigyeléses: nem avatkozok be, csak megfigyelek

Keresztmetszeti – adott időpillanatban

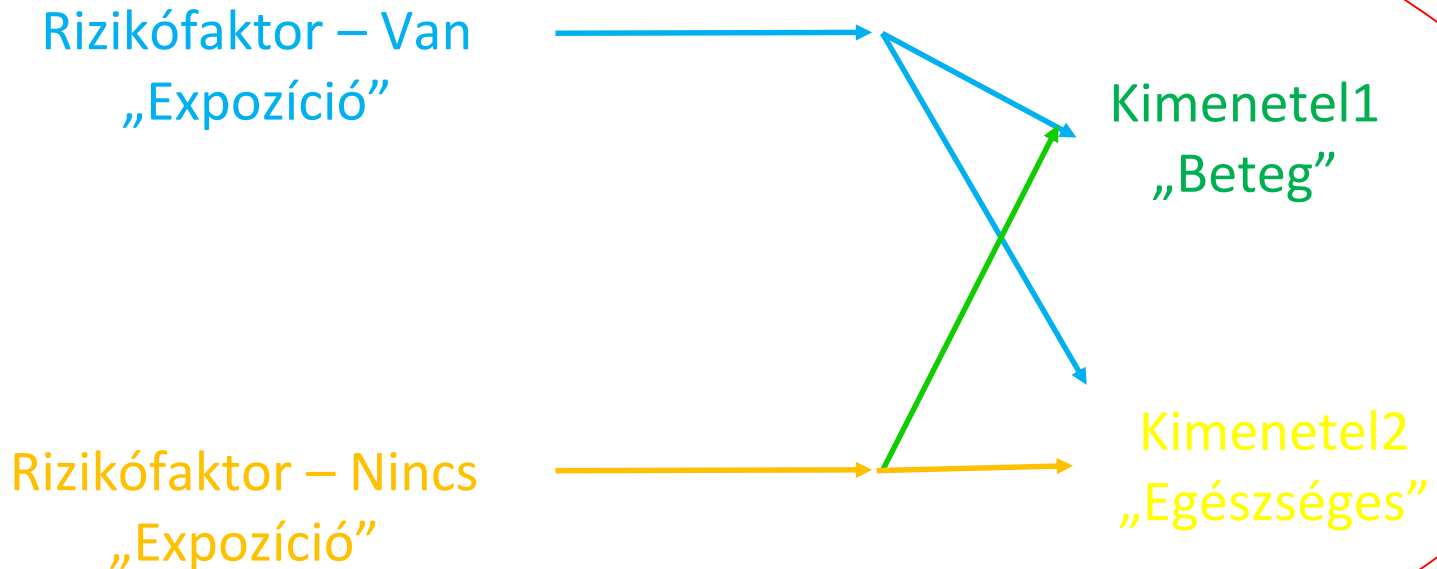
Eset-kontroll – visszatekintő (retrospektív)

Kohort/Kohorsz – előre/visszatekintő (pro-/retrospektív)

Kísérletes (experimentális): aktívan beavatkozok („kezelek”)

pl. randomizált kontrollált/klinikai vizsgálat (RCT)

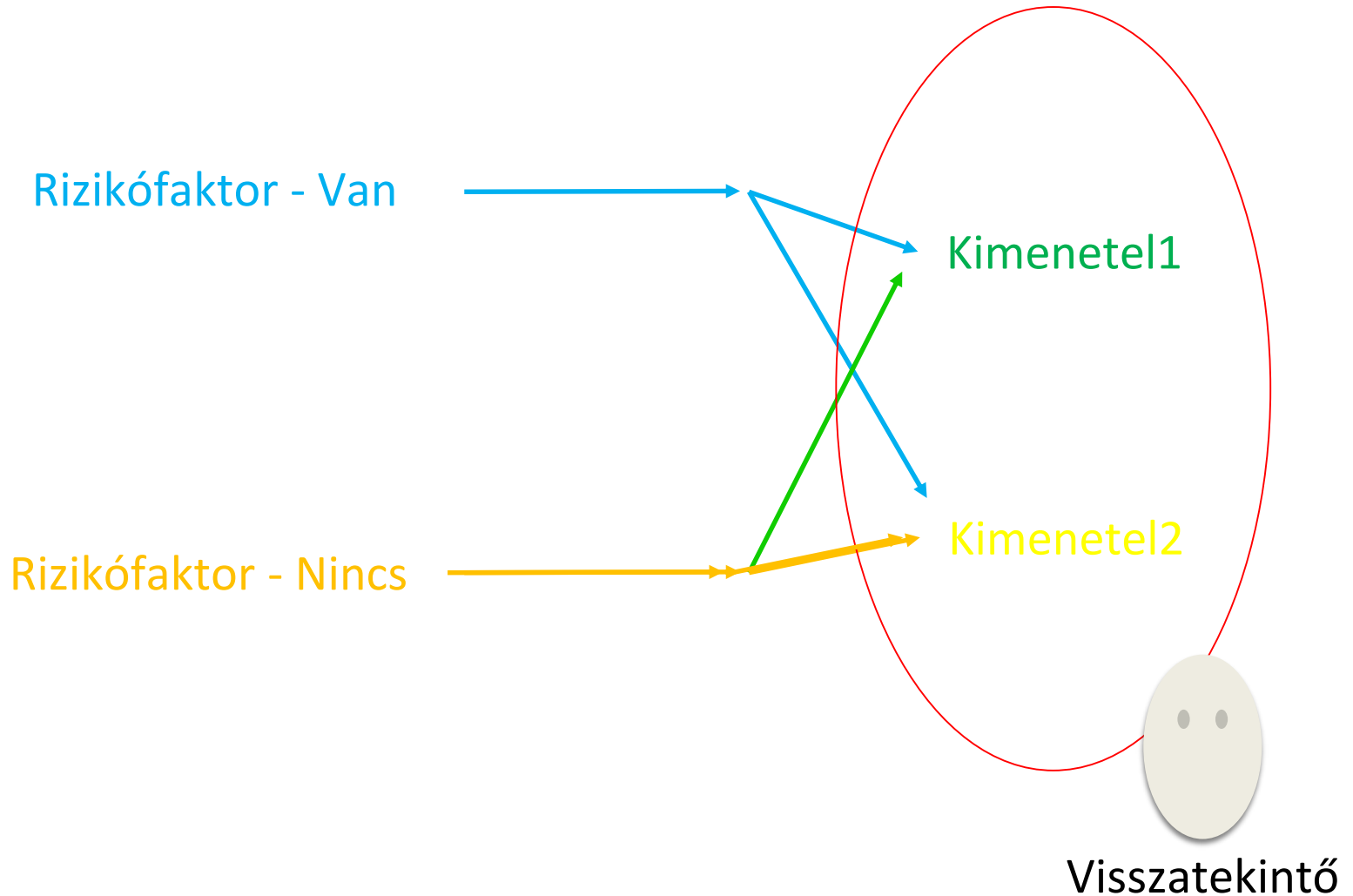
Keresztszeti vizsgálatok



A rizikófaktor és kimenetel meglétét egyszerre (ugyanakkor) vizsgáljuk

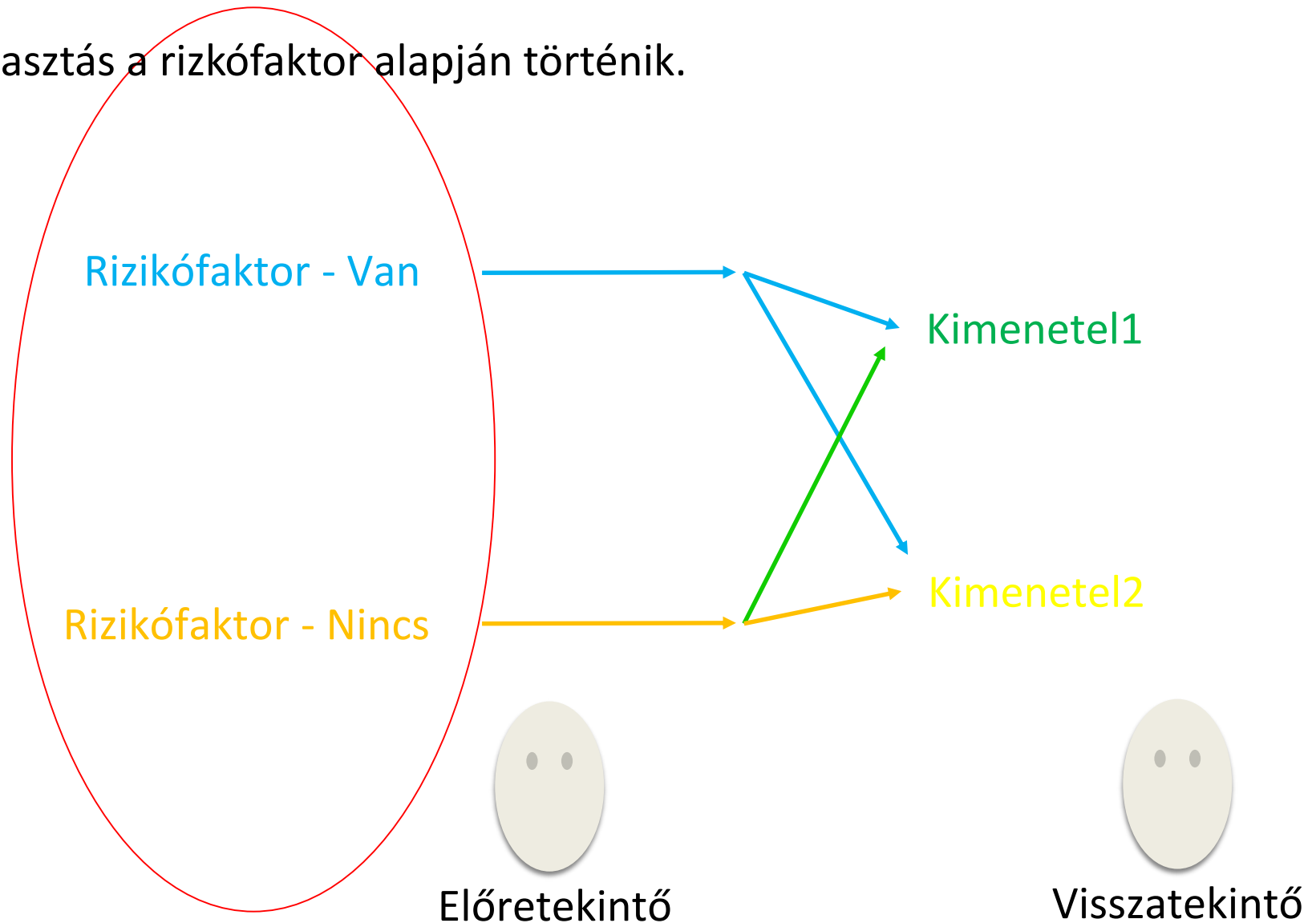
Eset-kontroll vizsgálatok

A kiválasztás a kimenetel alapján történik.



Kohort vizsgálatok

A kiválasztás a rizikófaktor alapján történik.



Kísérletes vizsgálatok

Beavatkozás történik.

(Randomizált)
csoportbeosztás

```
graph LR; A(( )) --> B(Beavatkozás - A); A --> C(Beavatkozás - B); B --> D1(Kimenetel1); B --> D2(Kimenetel2); C --> E1(Kimenetel1); C --> E2(Kimenetel2);
```

The diagram illustrates a randomized controlled trial. It begins with a circle labeled '(Randomizált) csoportbeosztás' (Randomized group assignment). Two arrows lead from this circle to two overlapping ovals. The top oval is labeled 'Beavatkozás - A' (Intervention - A) and the bottom oval is labeled 'Beavatkozás - B' (Intervention - B). Above the ovals, the text 'Beavatkozás történik.' (Intervention occurs.) is written. From the 'Beavatkozás - A' oval, two blue arrows point to 'Kimenetel1' (Outcome 1) and 'Kimenetel2' (Outcome 2). From the 'Beavatkozás - B' oval, two arrows (one green and one yellow) point to 'Kimenetel1' and 'Kimenetel2' respectively.

Beavatkozás - A

Kimenetel1

Kimenetel2

Beavatkozás - B

Kimenetel1

Kimenetel2

Klinikai vizsgálatok hibái és erényei

	Kereszt-metszeti	Eset-kontroll	Kohort	Kísérletes
Tulajdonság	Kiválasztás egy adott időpontban	Kiválasztás alapja: kimenetel (eset/kontroll)	Kiválasztás alapja: rizikófaktor	Beavatkozás történik
Előny	Olcsó, könnyű kivitelezni, azonnali	Ritka betegségeknel jó, viszonylag olcsóbb, gyorsabb	Ritka rizikófaktoroknál jó	Hibák könnyebb kiszűrése
Hátrány	Ok vagy okozat nem megállapítható	Kontrollok kiválasztása nehéz – kiválasztási , információs (pl. visszaemlékezés) hiba gyakori	Hosszú követési idő gyakori információs (pl. követési, visszaemlékezés) hiba	Drága, nehéz kivitelezni, gyakran etikai gondok

Confounding, torzítás

Megfigyeléses vs. Kísérletes elrendezés

CÉL: PONTOSAN (abban és CSAK abban) egy prediktorban legyen különbség:

KÍSÉRLETES vizsgálatban:

Beavatkozunk („kezelés”) és **randomizálunk** – TELJESÜL

MEGFIGYELÉSES vizsgálatban – többségünk ezzel találkozik:

NEM TELJESÜL

Van megoldás?

Confounding, torzítás

Megfigyeléses vs. Kísérletes elrendezés

CÉL: PONTOSAN (abban és CSAK abban) egy prediktorban legyen különbség:

KÍSÉRLETES vizsgálatban:

Beavatkozunk („kezelés”) és **randomizálunk** – TELJESÜL

MEGFIGYELÉSES vizsgálatban – többségünk ezzel találkozik:

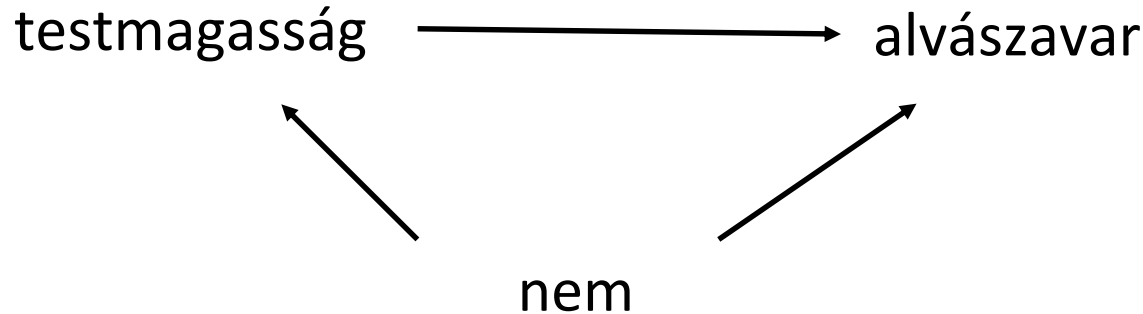
NEM TELJESÜL

Van megoldás?

IGEN, VAN.

Feltétel: legyen adatunk a lehetséges confounderről (rögzítve legyen)

Megoldás I.



Megoldás I: felbontjuk „rétegekre”: RÉTEGZÉS

nemek szerint vizsgáljuk külön-külön a testmagasság-alvászavar viszonyát

Hatás megadható: bármely counfounding értékre (nőre, férfira)

- **Mi van, ha sok réteg/folytonos változó?**
- **És ha több lehetséges confounder?**

Megoldás II.

Lineáris regresszió – több magyarázó változó

Feltételezzük (...):

1. számszerű prediktor – eredményváltozó lineáris viszonyban
2. nincs interakció (egy adott prediktor „saját” hatása független egy más prediktor értékétől)

Lineáris függvény: $Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \epsilon$

Y: az eredményváltozó értékei ()

X: egyes prediktor változók (számszerű) értékei

β : meredekségek (, illetve tengelymetszet)

ϵ : reziduum

Lineáris regresszió – „eredménye”

Értelmezzük!

	Testtömeg (kg)			Testtömeg (kg)		
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	33.0991	2.9998 – 63.1985	0.033	54.8088	29.0056 – 80.6120	<0.001
Születéskori testtömeg (g)	0.0091	0.0007 – 0.0176	0.035	0.0056	-0.0011 – 0.0123	0.095
NEM: nő				-13.9053	-21.4639 – -6.3468	0.001
Observations	21			21		

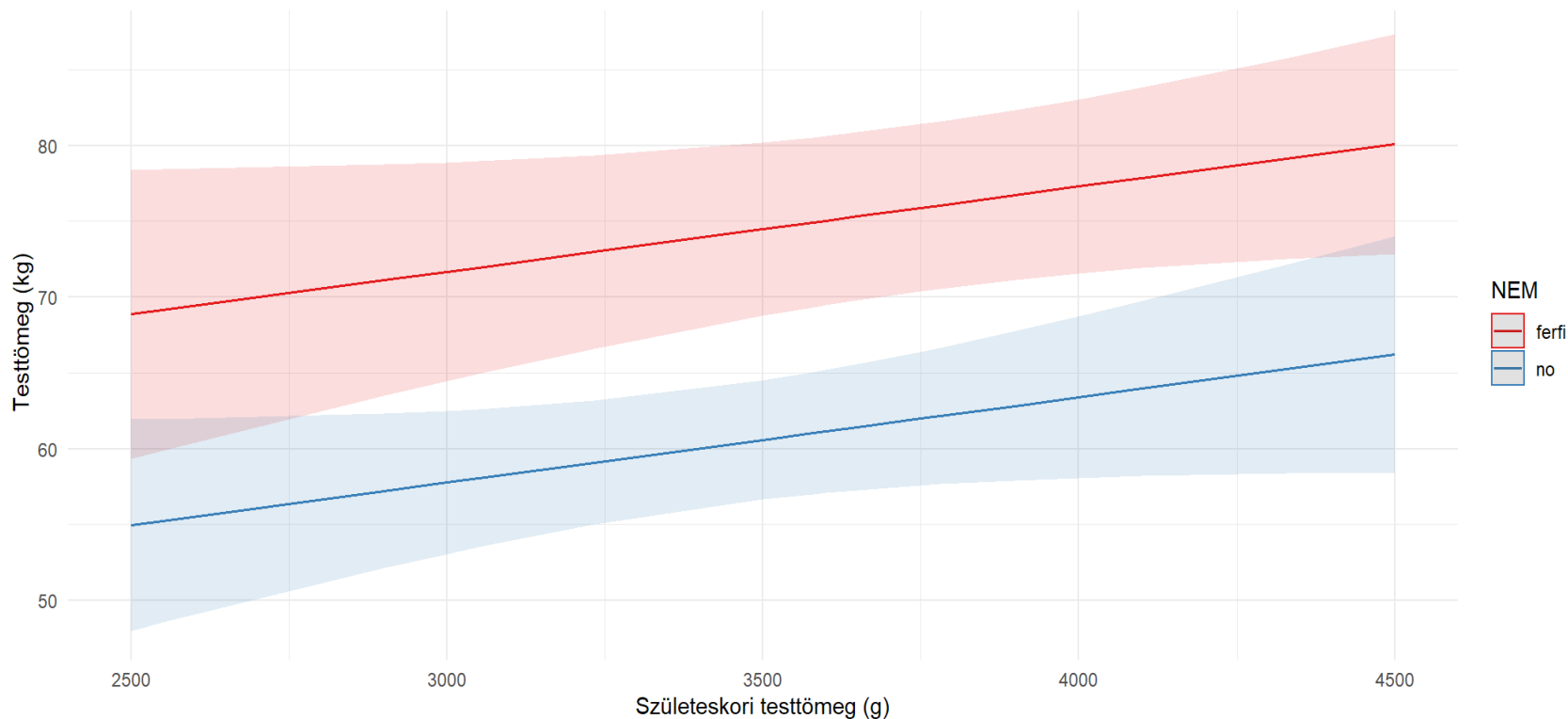
Merekség:

Y-ban bekövetkező átlagos változás, X-beli egységnyi változásra
ha a többi X értéke változatlan! (Kontrollálva a többire.)

Lineáris regresszió – „eredménye”

Értelmezzük!

	Testtömeg (kg)		
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	54.8088	29.0056 – 80.6120	<0.001
Születéskori testtömeg (g)	0.0056	-0.0011 – 0.0123	0.095
NEM: nő	-13.9053	-21.4639 – -6.3468	0.001
Observations	21		



Eredményváltozó kategoriális

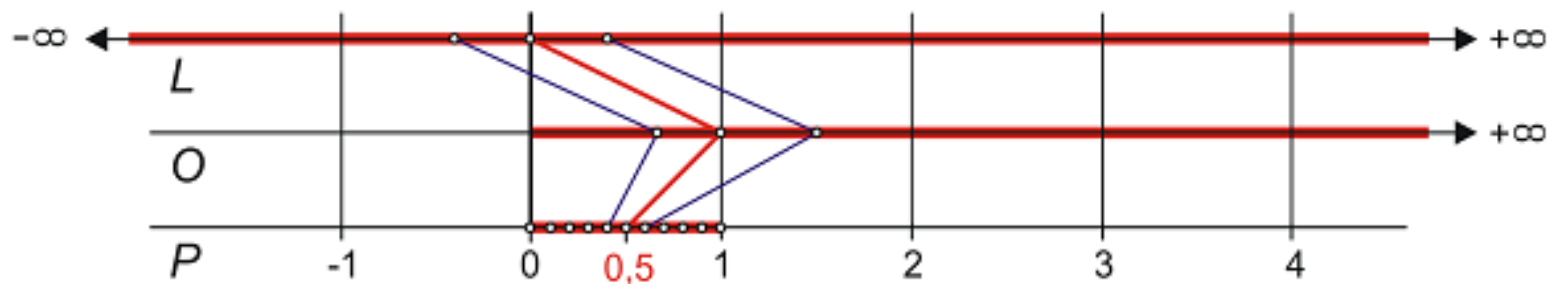
Esélyhányados és logit

Esély (esélyérték; *O* - odds): „hányszor akkora a valószínűsége annak, hogy az esemény bekövetkezik, mint annak, hogy nem következik be”

$$O = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1 - P(A)}$$

Logit (*L*): esély természetes alapú logaritmusa

Logit – Esély – Valószínűség



Logisztikus regresszió „eredménye”

Értelmezzük!

	Alvászavar
<i>Predictors</i>	<i>Odds Ratios</i>
(Intercept)	0.00
Testmagasság (cm)	1.27
NEM [nő]	1.68

Y esélyhányadosában bekövetkező átlagos változás
X-beli egységnyi változásra
ha a többi X értéke változatlan!

FIGYELEM: 10 egységnyi változásra nem 10x-re nő az OR!

Mennyire „hihető” és „jó” a MODELL

- Modelldiagnosztika („feltételek ellenőrzése”),
 - modell validáció
-
- További kurzusok
 - Kérdezze statisztikusát....

És HA...

- nem független,
- nem lineáris,
- nem normál,
- több kategóriájú magyarázó változó,
- interakció
- Stb

Ezekre is léteznek megoldások regressziós módon!

- További kurzusok
- Kérdezze statisztikusát....