

Többszörös lineáris regresszió.

Confounding és többszörös lineáris regresszió.

készítette: VERES Dániel

2021. november 3.

Az előadásban megvizsgáljuk, hogy milyen bonyodalmak lehetnek egy feltételezett magyarázó változó eredményváltozóra vonatkozó hatásának becslésében – hogyan tudunk „elvonatkoztatni” más változók hatásától.

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

2/30

Az orvostudományban igen gyakoriak az ok-okozati hatásra irányuló kérdések:

pl. rizikófaktor hatása a betegségre. (Hat-e, mennyire hat?)

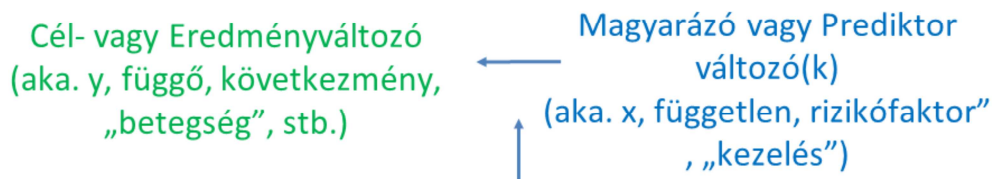
pl. kezelés hatása a betegségre. (Hat-e, mennyire hat?)

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!

3/30

Az ok-okozati hatás feltételezésekor gyakran 2 különböző kérdést feszegetünk:

1. Egyáltalán van-e hatás? 2. Ha van hatás, az mekkora lehet?

(Megjegyzés: a két kérdés együttesen is kezelhető: ha a hatás nem 0, akkor van hatás módon.)

A hatás vizsgálatához szükségünk van megfigyelt eredményekre: változók értékeire.

A nómenklaturában el szoktuk különíteni az „okozatot” (pl. betegséget) leíró változót: ezt leggyakrabban cél- vagy eredményváltozónak hívjuk. Az „ok”-ot („kezelés”) leíró változót pedig magyarázó vagy prediktor változónak nevezzük leggyakrabban.

A hatás nagyságának becslésére sok esetben alkalmas lehet az előző órán is tanult regresszió!

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 **klínikailag** feltételezett, statisztikailag nem vizsgálható.

Kapcsolódó kérdések:

- Van (adott típusú) függvényyszerű 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?

4/30

Ismételjük át a korábbi előadáson elhangzott legfontosabb információkat a regresszióról.

1. A regresszió esetében függvénykapcsolatot feltételezünk. – A kapcsolat „iránya” klínikailag feltételezett.

Ismétlés I. – Lineáris Regresszió

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

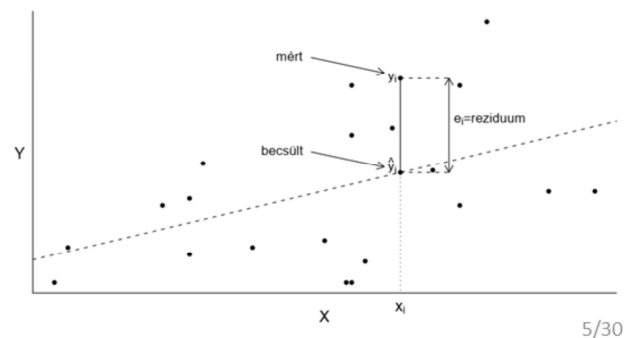
tengelymetszet

Meredekség

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

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

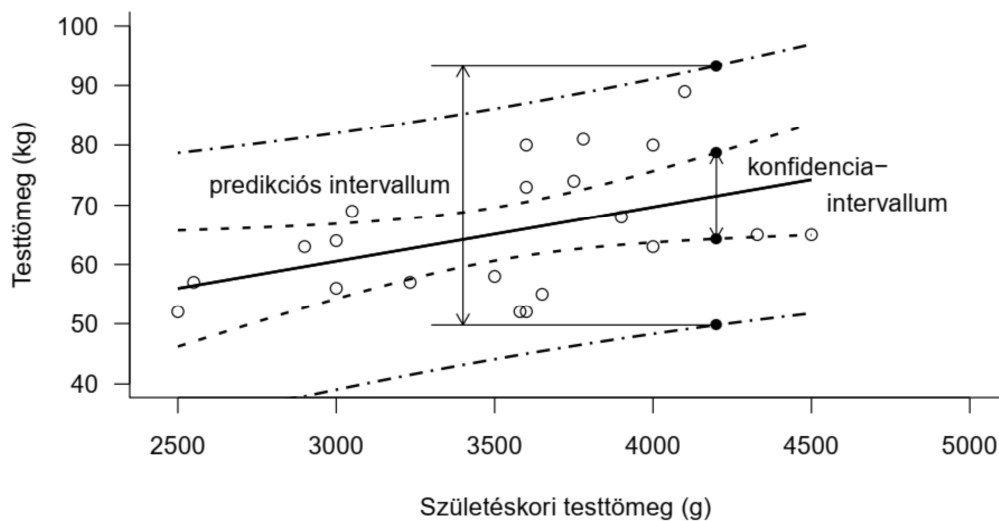
Az OLS szerinti legjobb egyenes az, ahol a legkisebb a pont-egyenes függőleges távolságok négyzetösszege.



5/30

Lineáris függvénykapcsolatot feltételezve OLS módszer segítségével becsülhetjük a kapcsolatot leíró egyenest.

Ismétlés I. – Lineáris Regresszió



Merekség - hatásmutató:

Y-ban bekövetkező átlagos változás,

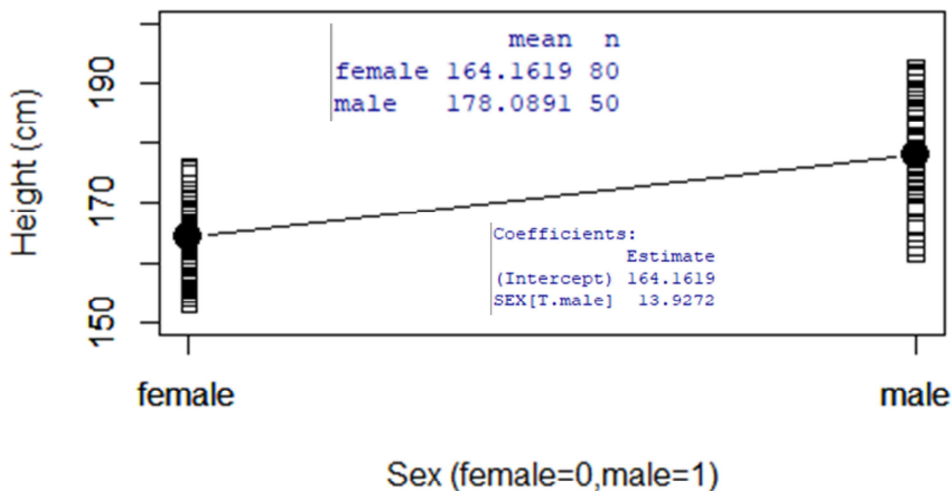
X-ben bekövetkező egységnyi változásra

6/30

Az illesztés alapján becsléseket adhatunk a meredekség és tengelymetszet hibájára, amiből konfidencia-intervallumot is számíthatunk.

A számunkra érdekes hatás nagyságát a meredekség mutatja: a célváltozóban mekkora változást várunk átlagosan, ha a magyarázó változó értéke 1 egységgel változik!

Lineáris regresszió – kategoriális változó?



Merekség:

Y-ban bekövetkező átlagos változás,

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

7/30

Gyakori, hogy a magyarázó változónk kategoriális. Ebben az esetben is használható a lineáris regresszió a hatás becslésére?

IGEN!

Most az egyszerűség kedvéért olyan kategoriális változót feltételezzünk, amelynek 2 kimenete lehetséges.

Illesszünk úgy egyenest, hogy a kategoriális változó egyik értékét 0-nak, másik értékét 1-nek tekintjük és így illesztjük az egyenest!

Ezt követően a matematikája, a hibák számítása, MINDEN megegyezik a korábbiakkal.

A meredekség – az 1 egységnyi változás - jelentése: a magyarázó változó egyik kategóriájából átlépve a másikba mekkora lesz az eredményváltozóban az átlagos változás. Ha megnézzük az illesztett egyenes egyenletét, akkor a tengelymetszet mutatja az egyik csoport átlagát, a meredekség pedig a két csoport átlaga közötti különbséget.

Ennek megfelelően a meredekségre adott becslés

(konfidencia-intervallum, hipotézisvizsgálat) megegyezik az átlagok különbségére vonatkozó becslésekkel.

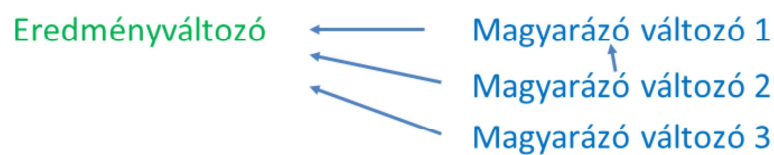
```
Rcmdr code: (using sleep1.xlsx dataset as Sleep)
with(Sleep, plotMeans(HEIGHT, SEX, error.bars="none",
connect=TRUE,ylab="",xlab="",ylim=c(150,200)))
par(new=TRUE)
stripchart(HEIGHT ~ SEX, vertical=TRUE, method="stack",
ylab="Height (cm)",xlab="Sex (female=0,male=1)",
data=Sleep,ylim=c(150,200),xlim=c(1,2))
```

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**

8/30

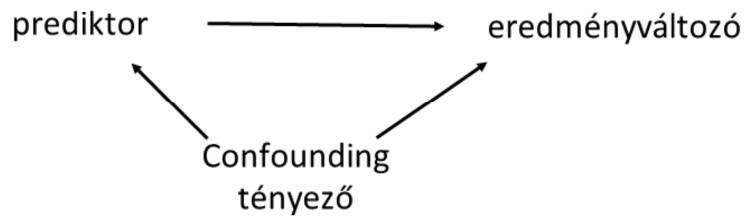
Mint korábban említettük, célunk, hogy ok-okozati viszony hatásnagyságára adjunk becslést. Milyen problémák, bonyodalmak adódnak?

1. Korábban is tanultunk arról, hogy a korreláció nem jelent ok-okozatot – csak feltételezzük.

2. Az orvosi közegben általában azt feltételezzük, hogy egy eredményváltozó értékét számos (nemcsak 1) magyarázó változó befolyásolja.

Mi vannak akkor, ha ezek a magyarázó változók egymásra is hatással vannak? – ehhez kapcsolódik a confounding fogalma.

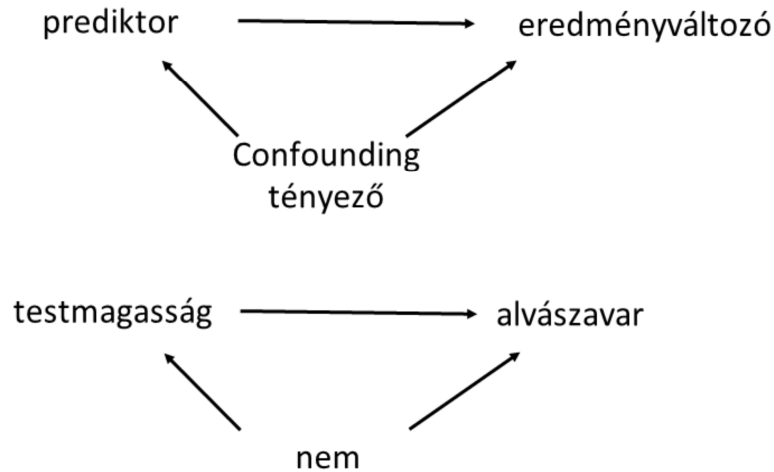
Counfounding



9/30

Röviden összemosó (megzavaró) hibáról, confoundingról akkor beszélünk, ha egy eredményváltozóban megjelenő hatást egy adott magyarázó változó hatásának tekintjük, pedig nem az – vagy a hatás nem akkora. A „megfigyelt” hatás nem az igazi hatás.

Counfounding



10/30

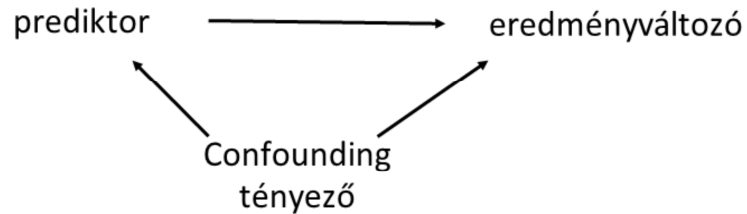
Jobban érthető ez egy példán keresztül: egy vizsgálatban azt tapasztaltuk, hogy a nagyobb testmagasság gyakrabban jár alvászavarral – ezt úgy magyarázhatnánk, hogy a magasság hat az alvászavarra (rövid az ágy, a takaró, ezen emberek görbábban alszanak... ötleteink támadhatnak.)

Azonban az alvászavar valójában nem függ a testmagasságtól: a férfiaknál gyakoribb az alvászavar és ők magasabbak is! Ha külön-külön vizsgálánk a férfiakat és nőket, akkor már nem látnánk a testmagasság hatását. Az összemosó tényező miatt a mért (első ránézésre úgy tűnő) hatás nagyobb, de akár kisebbnek is látszhatnak a valóságosnál.

Confounding változó az lehet, amely egyaránt „hat” a magyarázó változó értékére és az eredményváltozó értékére is.

Leggyakoribb confounding tényezők: nem, életkor – erre mindig gondoljunk!!!!!!!!!!!!

Counfounding

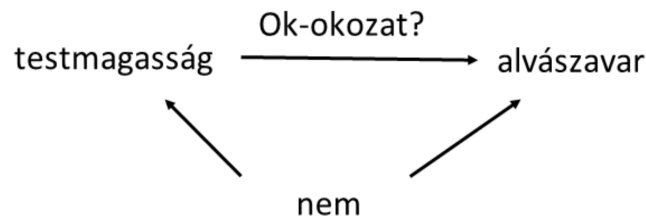


Minket a prediktor közvetlen hatása érdekel
– „függetlenül” a confoundigtól (és a többi prediktortól)!

11/30

Az orvosi gyakorlatban az esetek jelentős részében minket az adott magyarázó változó „közvetlen” hatása érdekel – azaz ha csak azt változtatjuk (változik), hogy változik az eredményváltozó.

Bonyodalmak



A kérdés NEM:

ha magasabb, akkor gyakoribb az alvászavar

HANEM:

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

12/30

A számunkra lényeges kérdés tehát nem az, ha magasabb valaki, - azaz más a prediktor értéke - akkor gyakoribb az alvászavar, hanem ha csak a testmagasságban – az adott prediktorban – van különbség, akkor hogyan változik az alvászavar valószínűsége.

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
elfogódottság

13/30

Látjuk, tehát, hogy ha eltérnek valami másban is a pacienseink, mint a vizsgálandó magyarázó változó, *akkor lehetséges*, hogy nem a valóságos, adott változóra vonatkozó közvetlen hatás értékét, hanem attól eltérőt – **torzítottat – kapunk még akkor is, ha mindenkit megvizsgálunk, azaz ha nincs mintavételi hiba.**

Ilyen torzított – hibás – értéket becslünk akkor, ha nem megfelelően járunk el a mintavétel során. Ezen hibáknak különböző csoportosításai vannak, most a leggyakrabban használt csoportosítás szerinti itt két leggyakoribb hibát – hiba típust emelek ki.

Az egyik a kiválasztási hiba. Ezt akkor követjük el, ha különbség van a felmérésbe beválasztottak és nem beválasztottak vagy a beválasztottak csoportba sorolása között (egy kimenetelt befolyásoló paraméter tekintetében).

Például:

1. Tudva, hogy a *nem* befolyásolja az alvászavart, az alvászavar ellen új gyógyszert kapók, illetve a placebót kapó csoportok között eltérő a nők aránya.
2. Utánkövetés megszűnésével nem számolunk: IV droghasználóknál vagy homoszexuálisoknál gyakoribb az AIDS: akik AIDS-esek lesznek, gyakrabban „kilépnek” az utánkövetésből, mint akik nem kapnak AIDS-t, valamint az IDU használók is gyakrabban „eltűnnek”, mint homoszexuálisok.
3. Alappopuláció más: csonttörés nőknél és táplálkozás összefüggése: csonttöröttet trauma osztályról választjuk, kontrollt az adott kórház belgyógyászatáról (De a belgyógyászatban levőknek más betegsége is van! Pl. diabetes gyakoribb!)

A másik nagy hibatípus az információs hiba. Erről akkor beszélünk, ha téves az alanyoktól kapott, alanyokról gyűjtött információ (, amely befolyásolja a kimenetet).

Például:

1. Visszaemlékezésből adódó hiba: azok a szülők, akiknek a gyermeke daganatos, jobban visszaemlékeznek gyermekük korábbi fertőzőes megbetegedéseire, mint azok, akiknek gyermeke egészséges.
2. Elfogódottság miatti hiba: a gyermekeket, időseket, bizonyos populációkat jobban megvizsgáljuk, mint nem ilyeneket.

További:

Catalogue of Bias Collaboration, Spencer EA, Brasseley J, Mahtani K., 2017. <https://catalogofbias.org/>

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)

14/30

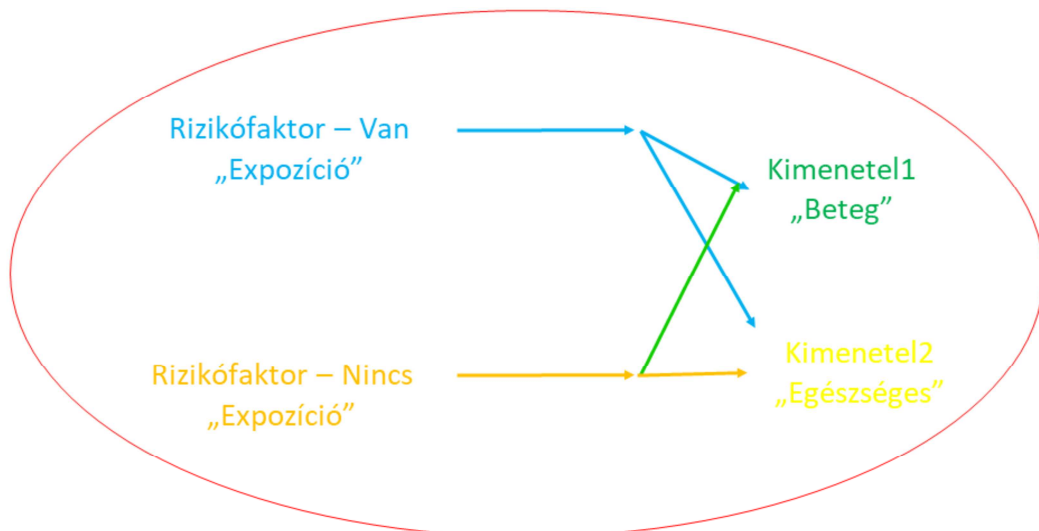
A torzítás megjelenése, természetesen függ attól is, hogy milyen „klinikai vizsgálati elrendezést használunk”.

Idézzük fel az 1. előadást és az ott bemutatott klinikai vizsgálatok típusait, hogy a hozzájuk kapcsolódó hibaforrásokat jobban megértsük.

Két nagy alcsoportját vizsgáljuk most meg a klinikai vizsgálatoknak: a megfigyeléses és kísérletes vizsgálatokat.

A megfigyeléses vizsgálatok fő típusai a keresztmetszeti, az eset-kontroll és a kohort vizsgálat.

Keresztmetszeti vizsgálatok



A rizikófaktor és kimenetel meglétét egyszerre (ugyanakkor) vizsgáljuk

15/30

Keresztmetszeti vizsgálatoknál a rizikófaktor (magyarázó változót) és kimenetel meglétét (eredmény változó értéke) egyszerre (ugyanakkor) vizsgáljuk: egyszerre kapunk információt a változókról.

Például:

Ivott-e reggel kávét (1. változó), mennyi érzi ébernek magát (2. változó)?

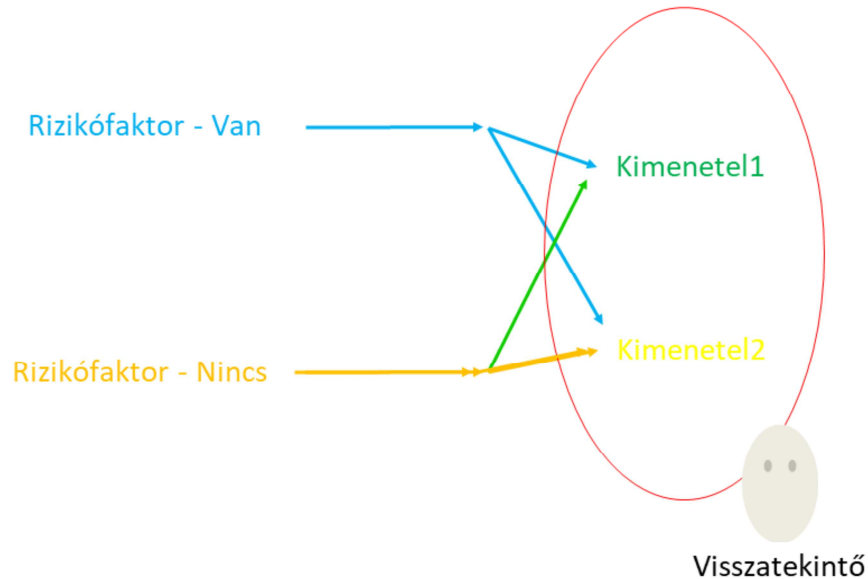
Családfő olvas-e újságot (1.) – tudja-e hogy a HIV terjedhet anyáról gyerekére (2.).

Oksági viszonyt sokkal nehezebben (pontosbban NEM) tud vizsgálni, hiszen időbeli információt nem ad.

Pl. 1. előadásban: Vakság (1.)– relatíve (életkorához, neméhez, stb megfelelő) kis testtömeg (2.) : melyik lehet az ok és okozat? (kicsi testtömeg miatt fertőzés -> vakság?)

Eset-kontroll vizsgálatok

A kiválasztás a kimenetel alapján történik.



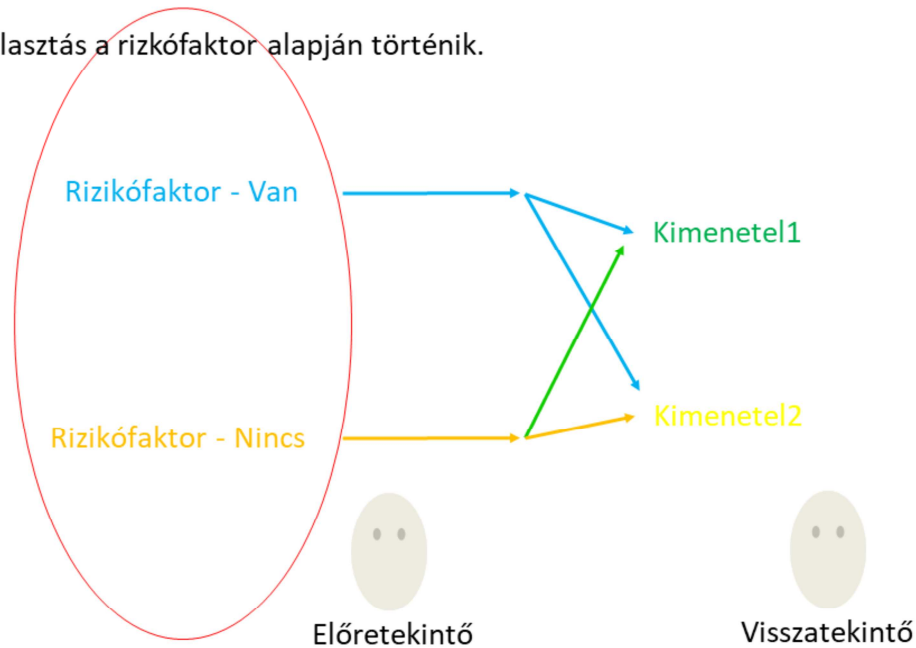
16/30

Az eset kontroll vizsgálatok esetében a beteg-kiválasztás alapja a kimenetel (betegség megléte vagy hiánya), azaz az eredményváltozó értéke alapján gyűjtjük össze a vizsgálandókat, majd utána „keressük vissza” – azaz *visszatekintünk* –, hogy mi történt velük korábban, azaz kinél mi volt a prediktor változó értéke. Az eset-kontroll vizsgálat, tehát egy visszatekintő módszer.

Akkor érdemes ezt a vizsgálati elrendezést használnunk, - azaz akkor hatékony -, ha a betegség ritka.

Kohort vizsgálatok

A kiválasztás a rizikófaktor alapján történik.



17/30

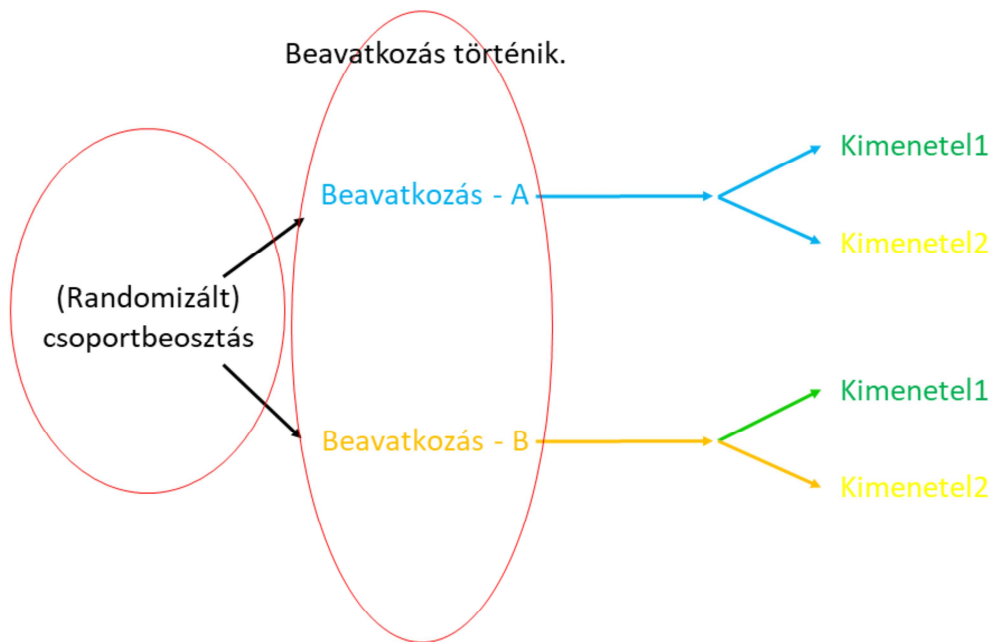
A kohort (vagy kohorsz) vizsgálatoknál a kiválasztás alapja a rizikófaktor megléte, illetve hiánya. Ha a rizikófaktor alapján való kiválasztás még a kimenetel (az eredményváltozó értéke: beteg lesz, nem lesz beteg) előtt megtörténik, akkor előretekintő (prospektív) vizsgálatról, ha a kiválasztáskor már az „eredmény megszületett” (csak mi nem tudjuk még), akkor visszatekintő (retrospektív) vizsgálatról beszélünk.

Például:

Sorozatos fejsérülése volt a labdarúgónak (hiszen fejelget) (prediktor: labdarúgó) – lesz-e demenciája (eredményváltozó: demencia)?

Dohányzik (prediktor) vs. Emlőtumor (eredményváltozó)?

Kísérletes vizsgálatok



18/30

A kísérletes vizsgálatnál:

Beavatkozunk – pl: egyik csoport gyógyszert, másik placebót kap.

Emiatt általában randomizált, hogy ki melyik csoportba kerüljön.

A randomizációval csökkenthetők a torzítások, hiszen a csoportjaink csak a beavatkozásban térnek el (ha elég nagy a minta...)!!

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?

20/30

Visszatérve az alapproblémára: a confounding elkerülésére nekünk az kell, hogy az eredményváltozó tekintetében csak a prediktorban legyen eltérés. Kísérletes vizsgálatban ez teljesül a randomizálás miatt.

Sajnálatos módon azonban megfigyeléses vizsgálatban ez nem teljesül – pl. más az alvászavarosok és nem alvászavarosok csoportjaiban a nemek megoszlási aránya a mintánkban, és mi a testmagasság hatását vizsgálánk az alvászavarra (amit szintén befolyásol a nem).

Tehetünk valamit?

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.

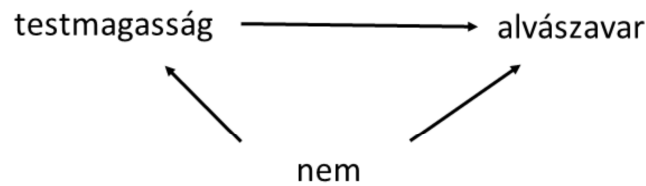
Feltétel: legyen adatunk a lehetséges confounderről (rögzítve legyen)

21/30

Igen, van megoldás.

DE! Csak akkor, ha a lehetséges confounder (pl. a nem) is rögzítve van.

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?**

22/30

Az egyik megoldás az úgynevezett *rétegzés*: felbontjuk a vizsgálatunkat a lehetséges confounder szerint csoportokra, és ezekben a csoportokban külön-külön vizsgáljuk a hatást.

De mi van, ha a confoundernek sok kategóriája van? Vagy akár számszerű, folytonos változó? És ha több lehetséges confounderünk van? Ezekben az esetekben nagyon sok rétegünk lesz! – nehézkes, sok elemzés kell és a kisebb elemszám miatt nő a becslés bizonytalansága is.

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)

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

Y: az eredményváltozó értékei ()

X: egyes prediktor változók (számszerű) értékei

β : meredekségek (, illetve tengelymetszet)

ϵ : reziduum

23/30

Egy másik megoldás: készítsünk regressziót nem egy, hanem több magyarázó változóval! Ezt nevezzük Többszörös regressziónak.

Az egyszerűség kedvéért tegyük fel:

1. A számszerű prediktorok és az eredményváltozó lineáris viszonyban van.
2. Nincs interakció a változók között.

Interakció: a prediktor „saját” (más prediktoroktól független, „valós”) hatása nem függ attól, hogy mi a másik prediktor értéke.

Megjegyzés: ezektől a feltételektől is „meg lehet szabadulni”, de ezekről ebben a kurzusban nem beszélünk.

Ekkor a dián látható egyenlettel írhatjuk le a változók viszonyát.

A számunkra érdekes paraméterek a meredekségek.

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		

Meredekség:

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

24/30

Végezzünk el egy lineáris regressziót a korábbi előadásban látható példára, ahol a testtömeget prediktáltuk a születéskori testtömeggel.

Tegyük fel, hogy hipotézisvizsgálatot is végeztünk 5%-os szignifikancia szinten – ehhez tudjuk viszonyítani a p-értékeinket.

A dia bal oldalán csak a születéskori testtömeget tartalmazó modell eredménye láthat, amíg jobb oldalon a nem-et is magába foglaló modellé.

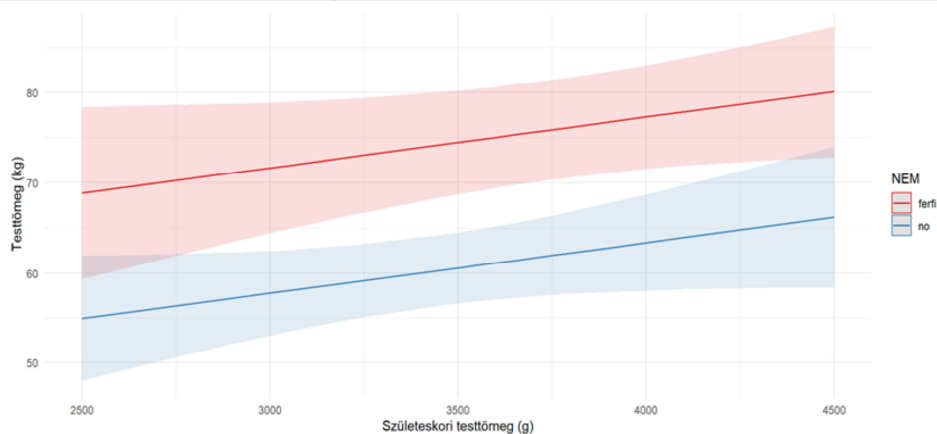
A bal oldalon azt látjuk, hogy úgy becsüljük, hogy 1g-nyi növekmány a születéskori testtömegben 0.0091 kg-nyi növekményt jelent a felnőttkori testtömegben.

Abban a modellben, ahol a nemet is „figyelembe vettük” (jobb oldal), látható, hogy becsléseink alapján a nem-től független, „saját” (úgy is mondjuk, hogy a nem-re kontrollált) hatása az 1 egységnyi növekményű születéskori testtömegnek csak 0.0056 kg (ha egyáltalán ezt mondanánk – lásd p-érték).

Általánosságban a meredekség azt jelenti, hogy mekkora az Y-ban bekövetkező átlagos változás, X-beli egységnyi változásra **ha a többi X értéke változatlan! (Kontrollálva a többire.)**

Lineáris regresszió – „eredménye” Értelmezzük!

	Testtömeg (kg)		
Predictors	Estimates	CI	p
(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		



25/30

Az eredményeket ábrán is könnyen megejelíthetjük. Itt a 95%-os konfidencia-sávokat is feltüntettük.

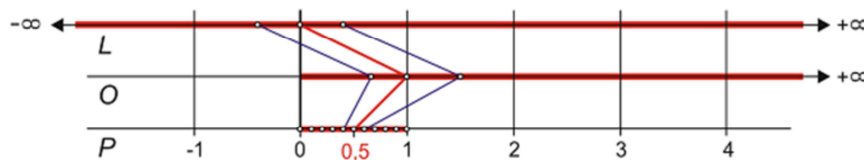
Eredményváltozó kategoriális Esélyhányados és logit

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



26/30

Ha az eredményváltozónk kategoriális és 2 kimenetelű, akkor a fentebbi meredekség-értelmezés ebben a formában nem működik, kicsit változtatnunk kell. Elevenítsük fel, amit tanultunk. Egy esemény oddsza azt mutatja, hogy hányszor akkora a valószínűsége annak, hogy az esemény bekövetkezik, mint annak, hogy nem következik be.

Ennek természetes alapú logaritmusa a logit, más néven log oddsz.

Az ábrán láthatjuk, hogy hogyan viszonyul a valószínűség-oddsz-logit értéke egymáshoz.

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!

27/30

Nagyon gyakori az orvosi gyakorlatban, hogy az eredményváltozó 2 kimenetelű kategoriális változó - például: élő/halott, beteg/nem-beteg, stb. Bővebben itt nem részletezendő módon ekkor is lehet regressziót számítani, aminek a leggyakoribb formáját logisztikus regressziónak nevezzük. Ebben az esetben a hatás nagyságát esélyhányadosként (OR, odds ratio) adjuk meg leggyakrabban. Nézzük meg mit is mond nekünk a fenti eredménytábla. Általánosságban a feltüntetett esélyhányados érték azt mutatja, hogy mekkora az eredményváltozóban bekövetkező átlagos változás, ha a prediktor változó értéke 1-gyel változik – ha a többi prediktor értéke változatlan. Példánkban: a testmagasság 1 cm-rel való növelése 1.27-szeresére növeli az alvászavar odds-át (ahhoz képest, hogy nincs alvászavar), kontrollálva a nemre (azaz adott nemre vonatkozóan). Figyelem!: itt nem igaz, hogy ha a prediktor 1 egységnyi változásra k az OR, akkor 10 egységnyi változására

10*OR!

Mennyire „hihető” és „jó” a MODELL

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

28/30

Fel kell, hogy merüljön bennünk, hogy mennyire „hihető” és „jó” a modell?

– Mennyire elfogadhatóak a feltételeink?

- Mennyire változtathatja meg az eredményt egy-egy kiszóró érték?

- Mennyire „ragaszkodik” a minta értékekhez: mennyire térne el az eredmény, ha hasonlóan vennénk egy mintát?

Ezekről ebben a kurzusban nem tudunk bővebben szót ejteni.

És HA...

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

Ezekre is léteznek megoldások regressziós módon!

- További kurzusok
- Kérdezze statisztikusát....

29/30

Mit tegyünk/tehetünk, ha kiderül, vagy tudjuk, hogy a feltételezéseink nem igazak, nem tűnnek elfogadhatónak? Ezekre is létezik regressziós megoldás – tehát az alapelvek, eredmények értékelése hasonló!

A regressziós témát bővebben taglalják más egyetemi kurzusok is. Például:

Kellermayer Miklós: Bevezetés a klinikai biostatistikába [Biofizikai és Sugárbiológiai Intézet]

Miklós Zsuzsanna: Klinikai vizsgálatok módszertana [Transzlációs Medicina Intézet]

Ferenci Tamás: Az orvosi megismerés módszertana (és az orvosi kutatások kritikus értékelése) [Népegészségtani Intézet, de Ferenci Tamás az Óbudai Egyetem oktatója]

Továbbá ajánlom Ferenci Tamás medstat.hu weboldalát is a téma iránt érdeklődőknek.