

# Regresszió a mintában: következtetés

Ferenci Tamás  
tamas.ferenci@medstat.hu

2018. február 7.

## Tartalom

## Tartalomjegyzék

<b>1. Az OLS elv és használata a lineáris regresszió becslésére</b>	<b>1</b>
1.1. Az OLS-elv . . . . .	1
1.2. A lineáris regresszió becslése tisztán deskriptíve . . . . .	3
1.3. Modellminősítés tisztán deskriptíve . . . . .	5
<b>2. Az OLS mintavételi tulajdonságai</b>	<b>7</b>
2.1. Mintavételi helyzet . . . . .	7
2.2. A mintavétel tulajdonságok szemléltetése szimulációval . . . . .	8
2.3. A mintavétel tulajdonságok matematikai levezetése . . . . .	11
<b>3. Becslés és hipotézisvizsgálat a lineáris modellben</b>	<b>19</b>
3.1. Alkalmazási feltételek . . . . .	19
3.2. Egy paraméter . . . . .	20
3.3. Modell egésze . . . . .	21
3.4. Lineáris megkötés(ek) . . . . .	22

## 1. Az OLS elv és használata a lineáris regresszió becslésére

### 1.1. Az OLS-elv

#### Előkészületek az OLS-becsléshez

- Nem kell hozzá semmilyen regresszió, a legközönségesebb következtető statisztikai példán is elmondható
- Például: sokasági várható érték becslése normalitás esetén (legyen a szórás is ismert)
- Ami fontos: bár egy alap következtető statisztika kurzuson nem szokták mondani, de lényegében itt is az a helyzet, hogy egy *modellt* feltételezünk a sokaságra
- Jeleül  $Y \sim \mathcal{N}(\mu, \sigma_0^2)$ , amit nem melleleg úgy is írhatnánk, hogy  $Y = \mu + \varepsilon$ , ahol  $\varepsilon \sim \mathcal{N}(0, \sigma_0^2)$
- A másik ami fontos: a modellből következik egy *becsült érték* minden mintabeli elemhez

- Jelen esetben, ha  $m$  egy feltételezett érték az ismeretlen sokasági várható értékre:

$$\hat{y}_i = m$$

### Az OLS-elv

- OLS-elvű becslés: az ismeretlen sokasági paraméterre az a becült érték, amely mellett a tényleges mintabeli értékek, és az adott paraméter melletti, modelltől származó becült értékek közti eltérések négyzetének összege a legkisebb:

$$\hat{\mu} = \arg \min_m \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \arg \min_m \sum_{i=1}^n (y_i - m)^2$$

- (Aminek a megoldása természetesen  $\hat{\mu} = \bar{y}$ )

### A mintavétel a lineáris regressziós feladatban

- Tételezzük fel, hogy az  $(Y, X_1, X_2, \dots, X_k)$  változóinkra veszünk egy  $n$  elemű mintát
- Az  $i$ -edik mintaelem:  $(y_i, x_{i1}, x_{i2}, \dots, x_{ik})$
- Feltételezzük azt is, hogy a mintavétel fae (független, azonos eloszlású)

A fae feltevés keresztmetszeti esetben sokszor lehet elfogadható közelítés (bár ott sem mindig teljesül, erre egy jó példa az ún. térbeli autokorreláció – de ez túlmutat a mostani kereteken), idősoros adatoknál azonban sohasem. Ott nagyon részletesen fogunk ezzel foglalkozni.

### Lineáris regresszió becslése OLS-elven

- *Hajszálpontosan ugyanaz* történik, mint az előbb, csak a sokaságra feltételezett modellünk kicsit bonyolultabb, jelesül:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

- A becült értékek adott  $b_0, b_1, \dots, b_k$  sokasági paraméterek mellett:

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_k x_{ik}$$

- A feladat tehát ugyanaz:

$$\begin{aligned} (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) &= \arg \min_{b_0, b_1, b_2, \dots, b_k} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \\ &= \arg \min_{b_0, b_1, b_2, \dots, b_k} \sum_{i=1}^n [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_k x_{ik})]^2 \end{aligned}$$

- Annyi bonyolódottság van, hogy itt most *több* paramétert kell becsülni, de ez csak a kivitelezést nehezíti, elvileg teljesen ugyanaz a feladat

## Az OLS-becslési feladat vektoros-mátrixos jelölésekkel

- A jelölések egyszerűsítése érdekében fogjuk össze mindent vektorokba és mátrixokba; egyedül a magyarázó változók nem triviálisak, mert kiegészítjük őket egy csupa 1 oszloppal (ún. design mátrix):

$$\mathbf{X}_{n \times (k+1)} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

- Így ugyanis a feladat:

$$\arg \min_{\mathbf{b}} (\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb})$$

- Az  $(\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb})$  hibanégyzetösszeget *ESS*-sel (error sum of squares) is fogjuk jelezni

Sajnos néhány irodalom az általunk használt *ESS*-re inkább az *RSS*-t (residual sum of squares) rövidítést használja, ami a jelölési zavarok legszerencsétlenebb típusa, ugyanis az *RSS*-t majd később mi is fogjuk használni, csak épp másra. Éppen ezért, ha ilyenekről olvasunk, mindig tisztázni kell, hogy a könyv vagy program írói mit értenek alatta.

## 1.2. A lineáris regresszió becslése tisztán deskriptíve

### Az OLS-becslési feladat megoldása

A megoldás:

$$\arg \min_{\mathbf{b}} (\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb}) = \arg \min_{\mathbf{b}} \left[ \mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{Xb} \right]$$

A szélsőérték-keresést oldjuk meg többváltozós deriválással (kvadratikus felület konvex, a stacionárius pont egyértelmű globális szélsőérték hely):

$$\begin{aligned} \frac{\partial}{\partial \mathbf{b}} \left[ \mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{Xb} \right] &= \\ &= -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{Xb} = 0 \Rightarrow \widehat{\beta}_{\text{OLS}} = \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y}, \end{aligned}$$

ha  $\mathbf{X}^T \mathbf{X}$  nem szinguláris

Az első lépésnél lényegében egyszerű algebrai átalakításokat végzünk (és a definíciókat használjuk), hiszen a zárójeleket felbontani, műveleteket elvégezni, mátrixokkal-vektorokkal is hasonlóan kell mint valós számokkal. (A transzponálás tagonként elvégezhető, azaz  $(\mathbf{a} - \mathbf{b})^T = \mathbf{a}^T - \mathbf{b}^T$ .) Egyedül annyit kell észrevenni, hogy a  $\mathbf{y}^T \mathbf{Xb}$  egy egyszerű valós szám, ezért megegyezik a saját transzponáltjával,  $\mathbf{b}^T \mathbf{X}^T \mathbf{y}$ -nal. Ezért írhattunk  $-(\mathbf{Xb})^T \mathbf{y} - \mathbf{y}^T \mathbf{Xb}$  helyett egyszerűen  $-2\mathbf{b}^T \mathbf{X}^T \mathbf{y}$ -t. (Itt mindenhol felhasználtuk, hogy a transzponálás megfordítja a szorzás sorrendjét:  $(\mathbf{AB})^T = \mathbf{A}^T \mathbf{B}^T$ .)

Itt jelentkezik igazán a mátrixos jelölésrendszer előnye. A  $\mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{Xb} + \mathbf{b}^T \mathbf{X}^T \mathbf{Xb}$  lényegében egy „másodfokú kifejezés” többváltozós értelemben (az  $ax^2 + bx + c$  többváltozós megfelelője), és ami igazán szép: pont ahogy az  $ax^2 + bx + c$  lederiválható a változója ( $x$ ) szerint (eredmény  $2ax + b$ ), ugyanúgy ez is lederiválható a változója (azaz  $\mathbf{b}$ ) szerint... és az eredmény az egyváltozóssal teljesen analóg lesz, ahogy fent is látható! (Ez persze bizonyítást igényel! – lásd

többszörös analízisből.) Bár ezzel átléptünk egyváltozóról többszörösre, a többszörös analízisbeli eredmények biztosítanak róla, hogy formálisan ugyanúgy végezhető el a deriválás. (Ezt írja le röviden a „vektor szerinti deriválás” jelölése. Egy  $\mathbf{b}$  vektor szerinti derivált alatt azt a vektort értjük, melyet úgy kapunk, hogy a deriválandó kifejezést lederiváljuk  $\mathbf{b}$  egyes  $b_i$  komponensei szerint (ez ugye egyszerű skalár szerinti deriválás, ami már definiált!), majd ez eredményeket összefoglaljuk egy vektorba. Látható tehát, hogy a vektor szerinti derivált egy ugyanolyan dimenziós vektor, mint ami szerint deriváltunk.) Ami igazán erőteljes ebben az eredményben, az nem is egyszerűen az, hogy „több” változónk van, hanem, hogy nem is kell tudnunk, hogy mennyi – mégis, általában is működik!

Azt, hogy a megtalált stacionaritási pont tényleg minimumhely, úgy ellenőrizhetjük, hogy megvizsgáljuk a Hesse-mátrixot a pontban. A mátrixos jelölésrendszerben ennek az előállítására is egyszerű, még egyszer deriválni kell a függvényt a változó(vektor) szerint:

$$\frac{\partial^2}{\partial \mathbf{b}^2} [\mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{b}] = \frac{\partial}{\partial \mathbf{b}} [-2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \mathbf{b}] = 2\mathbf{X}^T \mathbf{X}.$$

Az ismert tétel szerint a függvénynek akkor van egy pontban ténylegesen is (lokális, de a konvexitás miatt egyben globális) minimuma, ha ott a Hesse-mátrix pozitív definit. Esetünkben ez minden pontban teljesül. A  $\mathbf{X}^T \mathbf{X}$  ugyanis pozitív szemidefinit (ez egy skalárszorzat-mátrix, más néven Gram-mátrix, amelyek mindig pozitív szemidefinitek), a kérdés tehát csak a határozott definitésg. Belátható azonban, hogy ennek feltétele, hogy  $\mathbf{X}^T \mathbf{X}$  ne legyen szinguláris – azaz itt is ugyanahhoz a feltételhez értünk! Megjegyezzük, hogy ez pontosan akkor valósul meg, ha az  $\mathbf{X}$  teljes oszloprangú. (Erre a kérdésre a modellfeltevések tárgyalásakor még visszatérünk.)

Végül egy számítástechnikai megjegyzés: az együttthatók számításánál a fenti formula direkt követése általában nem a legjobb út, különösen ha sok megfigyelési egység és/vagy változó van. Ekkor nagyméretű mátrixot kéne invertálni, amit numerikus okokból (kerekítési hibák, numerikus instabilitás stb.) általában nem szeretünk. Ehelyett, a különféle programok igyekeznek a direkt mátrixinverziót elkerülni, tipikusan az  $\mathbf{X}$  valamilyen célszerű mátrix dekompozíciójával (QR-dekompozíció, Cholesky-dekompozíció). Extrém esetekben még az is elképzelhető, hogy az egzakt, zárt alakú megoldás előállítás helyett valamilyen iteratív optimalizálási algoritmus (gradiens módszer, Newton–Raphson-módszer) alkalmazása a gyakorlatban járható út, annak ellenére is, hogy elvileg van zárt alakban megoldása.

A kapott eredmény nem más, mintha  $\mathbf{X}$  Moore–Penrose pszeudoinverzével szoroznánk  $\mathbf{y}$ -t.

### Pár további gondolat

- Az ún. reziduumok:

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$$

- Az előrejelzések a mintánkra:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

- Ez alapján vezessük be a

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

mátrixot, ezzel  $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$

- Emiatt szokták „hat” mátrixnak is nevezni

### Az OLS geometriai interpretációja

$\mathbf{P}$  projektormátrix lesz ( $\mathbf{P}^2 = \mathbf{P}$ , azaz idempotens)  $\rightarrow$  út az OLS geometriai interpretációjához

Mindenekelőtt emlékeztetünk rá, hogy az  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$  vektorok által kifeszített alteret azok a pontok alkotják, melyek előállnak e vektorok lineáris kombinációjaként. (E pontok mindig az eredeti vektortér – ami felett a vektorokat értelmeztük – alterét alkotják, ezért jogos az elnevezés.) Ha most vektortérnek az  $\mathbb{R}^n$ -et tekintjük, vektoroknak pedig az  $\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_k$  magyarázóváltozókat (és a konstans), azaz  $\mathbf{X}$  oszlopvektorait, akkor az ezek által kifeszített alter – ezt szokás egyébként az  $\mathbf{X}$  mátrix oszlopterének nevezni – épp azon pontokból áll, melyek előállhatnak becült eredményváltozó(vektor)ként valamilyen regressziós koefficienssekkel! (Hiszen a becült eredményváltozót is e vektorok lineáris kombinációjaként állítjuk elő.) Általánosságban persze nem várható, hogy a tényleges eredményváltozó(vektor) benne legyen ebben az alterben (azaz egzaktan – értsd: minden egyes megfigyelési egységre megvalósulóan – elő lehessen állítani lineáris kombinációként), ezt fejezi ki a reziduum. Mint a tényleges és a becült eredményváltozó(vektor) különbségvektora, a reziduum hossza megmutatja, hogy mennyire messze van a becült és a tényleges eredményváltozó egymástól (az  $\mathbb{R}^n$ -ben). Mi azt szeretnénk, ha ez minimális lenne. Választva a szokásos euklideszi metrikát, visszakapjuk a legkisebb négyzetes értelmezést. A kérdés már csak az, hogy adott ponthoz (tényleges eredményváltozó) hogyan határozható meg az alter (azaz: amit lineáris regresszióval elő tudunk állítani) legközelebbi pontja... de hát ez épp a geometriai vetítés művelete! A megoldás tehát az, hogy a tényleges eredményváltozót merőlegesen rávetítjük (ortogonális projekció) a magyarázóváltozók (és a konstans) által kifeszített alterre! A vetítés eredményeként kapott pont lesz a ténylegeshez legközelebbi előállítható becült eredményváltozó, az előállításában szereplő együtthatók pedig az optimális becült regressziós koefficienssek. Így aztán azt is megállapíthatjuk, hogy a fenti  $\mathbf{P}$  mátrix nem más, mint ami a tényleges eredményváltozót levetíti a magyarázóváltozók (és a konstans) által kifeszített alterre.

### 1.3. Modellminősítés tisztán deskriptíve

#### Modell jóságának viszonyítási pontjai

- A modell minősítése az ESS alapján?  $\rightarrow$  kézenfekvő, de nem önmagában: viszonyítani kell! Két kézenfekvő alap:
  - Tökéletes (v. szaturált, perfekt modell): minden mintaelemre a pontos értéket becsüli  $\rightarrow \hat{\epsilon}_i = 0 \Rightarrow ESS = 0$
  - Nullmodell: semmilyen külső (magyarázó)információt nem használ fel  $\rightarrow$  minden mintaelemet az átlaggal becsül
- Egy adott regressziós modell teljes négyzetösszegének nevezzük, és  $TSS$ -sel jelöljük a hozzá tartozó (tehát ugyanazon eredményváltozóra vonatkozó) nullmodell hibanégyzetösszegét:

$$TSS = ESS_{\text{null}} = \sum_{i=1}^n (y_i - \bar{y})^2.$$

#### Hogyan jellemezzük modellünk jóságát?

- A minősítést képezzük a „hol járunk az úton?” elven: a tökéletesen rossz modelltől a tökéletesen jó modellig vezető út mekkora részét tettük meg
- Az út „hossza”  $TSS$  ( $= TSS - 0$ ), amennyit „megtettünk”:  $TSS - ESS$

- Egy adott regressziós modell regressziós négyzetösszegének nevezzük, és  $RSS$ -sel jelöljük a teljes négyzetösszegének és a hibanégyzetösszegének különbségét:

$$RSS = TSS - ESS.$$

Ahogy már említettük is, sajnos néhány könyv az  $RSS$ -t más néven, hogy még rosszabb legyen a helyzet, néha  $ESS$ -ként, emlegeti. (Az itteni  $ESS$  pedig épp  $RSS$  az ottani terminológiában. . .)

### Az új mutató bevezetése

Ezzel az alkalmas modelljellemező mutató: a többszörös determinációs együttható (jele  $R^2$ ):

$$R^2 = \frac{TSS - ESS}{TSS} = \frac{RSS}{TSS}.$$

### Az $R^2$ -ről bővebben

- Ha van konstans a modellben, akkor nyilván  $ESS < TSS$ , így minden regressziós modellre, amiben van konstans:  $0 \leq R^2 \leq 1$ .
- Az  $R^2$  egy modell jószágának legszéleskörűben használt mutatója
- Értelmezhető %-ként: a magyarázó változók ismerete mennyiben csökkentette az eredményváltozó tippelésekor a bizonytalanságunkat (ahhoz képest, mintha nem ismertünk volna egyetlen magyarázó változót sem)
- De vigyázat: nagyságának megítélése, változók száma stb.
- A belőle vont négyzetgyököt többszörös korrelációs együtthatónak szokás nevezni
- Mondani sem kell, ez az  $R^2$  a korábban bevezetett (sokasági)  $R^2$  mintabeli analógja

### Az $R^2$ -ről bővebben

- Ha van konstans a modellben, akkor érvényes a következő felbontás:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

- (Négyzetek nélkül nyilvánvaló, de négyzetekkel is!)
- Röviden tehát:

$$TSS = ESS + RSS$$

- Összevetve az előző definícióval, kapjuk, hogy

$$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

### Egy megjegyzés a konstans szerepéről

- Az előzőek is motiválják, hogy megállapítsuk: konstanst *mindenképp* szerepeltetünk a regresszióban, ha inszignifikáns, ha nem látszik különösebb értelme stb. *akkor is!* – csak és kizárólag akkor hagyhatjuk el, ha az a modell tartalmából adódóan elméleti követelmény (erre látni fogunk nemsokára egy példát is, a standardizált regressziót)
- Ellenkező esetben (ún. konstans nélküli regresszió), a fenti felbontás nem teljesül, így a „hol járunk az úton” elven konstruált  $R^2$  akár negatív is lehet!

Néhány könyv, az  $R^2$  alternatív definiálása révén, a negatív esetet kizárja.

## 2. Az OLS mintavételi tulajdonságai

### 2.1. Mintavételi helyzet

#### Deskriptív és következtető statisztika

- Az előbbi tárgyalás pusztán deskriptív volt: egy darab mintát tekintett, amire meghatározott egy darab regressziós függvényt és kész
- Mintha a feladat csak annyi lenne, hogy pontokra húzzunk egy rájuk jól illeszkedő görbét
- Ez a „görbeillesztési” szemlélet első ránézésre könnyen megérthető, és látszólag egyszerűsíti a helyzetet, valójában azonban rendkívül mérgező a valódi megértésre nézve
- Nem teszi lehetővé ugyanis annak megértését, hogy a háttérben van egy sokaság, és a görbe nem univerzálisan jellemzi azt, hanem csak az adott, konkrét mintára illeszkedik legjobban, és *másik mintából másik görbét kaptunk volna*
- Azaz: figyelmen kívül hagyja a mintavételi helyzetet

#### A mintavételi helyzet hatásai

- Van egy elméleti regresszió a **sokaságban** ( $\beta$ )
- Az adatbázisunk alapján megkaptuk a regressziós egyenest ( $\hat{\beta}$ )
- Az adatbázis azonban csak egy **minta** a sokaságából, így a  $\hat{\beta}_i$  paraméterek annak hatását *is* tükrözik, hogy konkrétan milyen mintát választottunk
- *Mintavételi ingadozás* lép fel (még akkor is, ha tökéletesen véletlen a mintavétel, ennek tehát semmi köze pl. a reprezentativitáshoz)
- Tehát: az egyes  $\hat{\beta}_i$  paraméterek „mintáról-mintára ingadoznak”: minden mintából más paramétereket kapnánk
- (Természetesen reméljük, hogy az ingadozás „kellemes” tulajdonságokkal bír, például a valós érték körül történik, szorosan körülötte stb., erről később)
- Ez tehát egy becslési feladat; az OLS-nek, mint becslőfüggvénynek vizsgálhatóak a tulajdonságai

Végeredményben a mintából számolt jellemzőkben nem csak az fog tükröződni, amit szeretnénk vizsgálni (azaz a tényleges – értsd: elméleti, sokasági – regresszió), hanem az is, hogy a sokaságból konkrétan hogyan vettük a mintát (azaz megjelenik a mintavételi ingadozás hatása is).

### Még egy fontos megjegyzés

- Nem elég annyit mondani, hogy „jó, hát akkor a háttérben van egy sokaság is”, mintha ezzel el lenne intézve ez a kérdés
- Azt is világosan látni kell, hogy az egész tárgyalás *kiindulópontja*, hogy erre *feltételezünk* egy modellt (pl. azt, hogy  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$ )
- Ez egy feltételezés, mivel a sokaságot nem ismerjük, így biztosan nem tudhatjuk, hogy igaz-e (csak következtethetünk rá)
- De minden további levezetés mögött ott lesz, hogy mi mit gondoltunk, hogyan viselkedik a sokaság, mi a *sokasági modell*

### Előkészület a mintavétel vizsgálatához

- Ahhoz, hogy a mintavétel hatását matematikailag tudjuk vizsgálni, az OLS-becslőt val. változókra kell ráereszteni (szemben az eddigi képlettel –  $\widehat{\beta}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$  – ahol konkrét számokra futtattuk)
- Pontosan ugyanúgy, ahogy az  $\frac{1}{n} \sum_{i=1}^n x_i$ -t sem tudjuk következtető statisztikailag vizsgálni (az egy szám), hanem az  $\frac{1}{n} \sum_{i=1}^n X_i$ -t nézzük
- Minket tehát  $\widehat{\beta}_{OLS} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{\underline{Y}}$  fog érdekelni!
- Ahogy az előbbi átlagos példában, így itt is igaz lesz, hogy ekkor a  $\widehat{\beta}_{OLS}$  már nem egy konkrét érték (vektor), hanem egy – vektor értékű – val. változó, tehát eloszlása van!
- Ez a mintavételi eloszlás, mi erre, ennek tulajdonságaira, a jó tulajdonságok feltételeire stb. leszünk kíváncsiak
- Előbb szimulációval nyerünk képet, aztán matematikailag is levezetjük

## 2.2. A mintavétel tulajdonságok szemléltetése szimulációval

### Monte Carlo szimuláció használata

- Számos konkrét véletlen mintát veszünk egy előre specifikált populációból (véletlenszám-generátort használunk)
- Lényegében: empirikusan vizsgálunk egy elméleti kérdést
- Valszámos embert játszunk: ugye azt mondtuk, hogy a valszámosok úgy dolgoznak mintha valahonnan ismernék a sokasági eloszlást – hát most tényleg ismerjük!
- Példának okáért, legyen a valódi sokasági eloszlás

$$(X, Y) \sim \mathcal{N} \left( \begin{pmatrix} 77 \\ 26 \end{pmatrix}, \begin{pmatrix} 42^2 & 0,6 \cdot 20 \cdot 42 \\ 0,6 \cdot 20 \cdot 42 & 20^2 \end{pmatrix} \right)$$

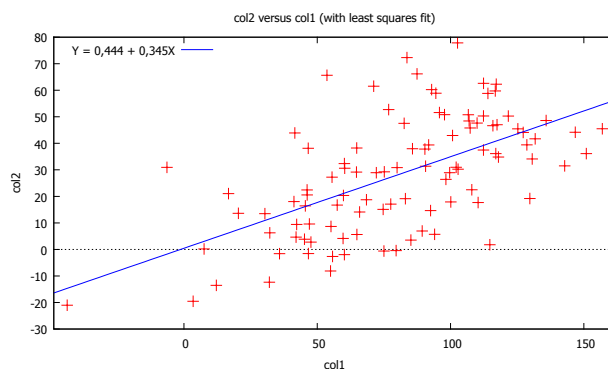
- Ezért a *valódi* regressziós egyenes, a már látottak szerint:

$$\mathbb{E}(Y | X) = 4 + \frac{12}{42} X \approx 4 + 0,2857 X$$

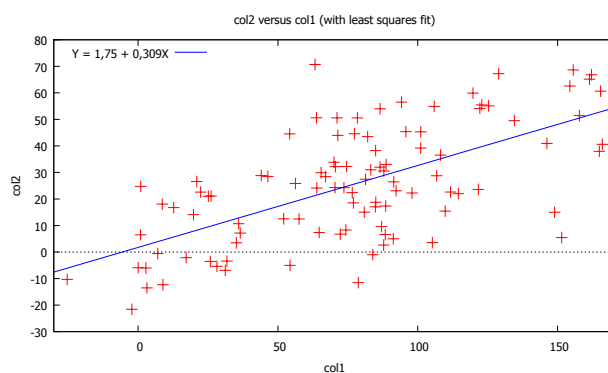
- Szimulációs paraméterek:  $n = 100$  elemű minta a fenti sokaságból, 1000 ismétlés



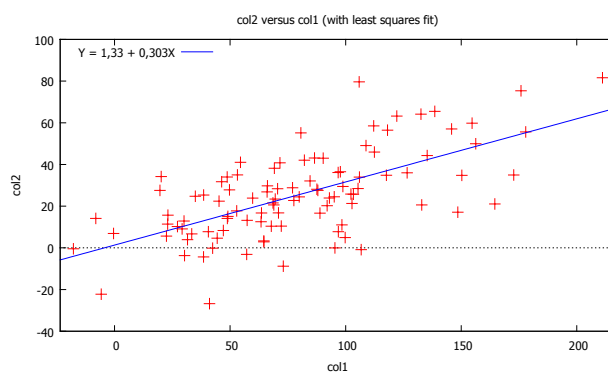
### A szimuláció eredményei: 1. futtatás



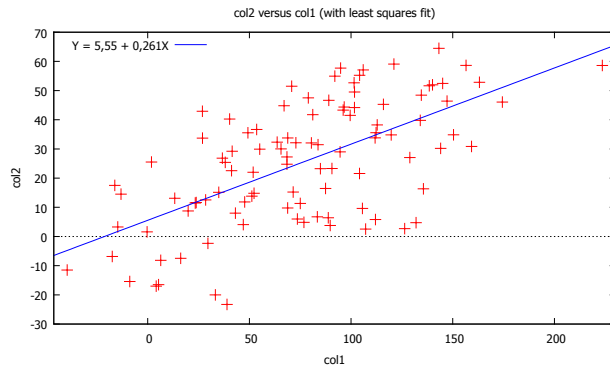
### A szimuláció eredményei: 2. futtatás



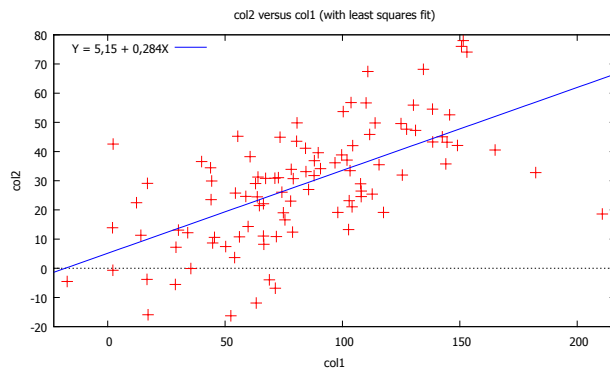
### A szimuláció eredményei: 3. futtatás



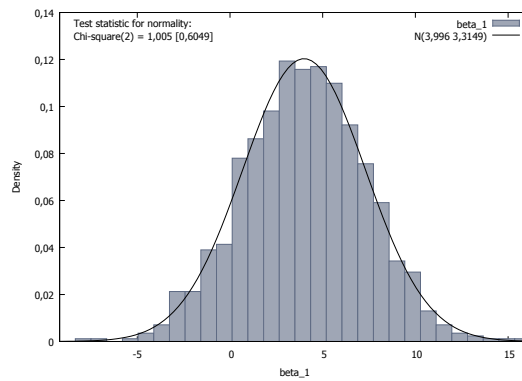
### A szimuláció eredményei: 4. futtatás



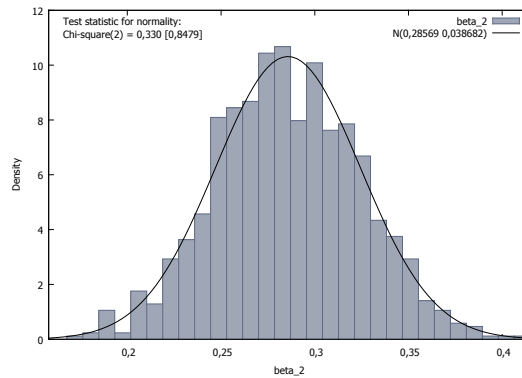
A szimuláció eredményei: 5. futtatás



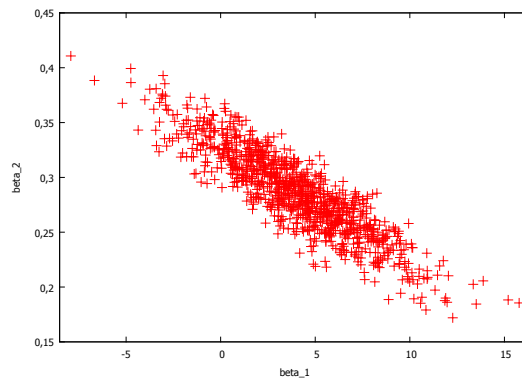
A szimuláció eredményei: konstans



A szimuláció eredményei: meredekség



A szimuláció eredményei: mindkét paraméter együtt



### 2.3. A mintavétel tulajdonságok matematikai levezetése

Az OLS-becslő mintavételi eloszlása

- Tudjuk, hogy  $\widehat{\beta}_{OLS} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$
- Valamint elfogadtuk feltételezéseként, hogy a sokasági modell  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon = \underline{X}^T \beta + \varepsilon$ 
  - És ez van mindegyik megfigyelési egység mögött is, tehát a mintavétel elemzéséhez ezt is írhatjuk:
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$
  - Röviden, értelemszerű vektorokba/mátrixokba fogással:  $Y_i = \underline{X}_i^T \beta + \varepsilon$  avagy az egész adatbázisra:  $\underline{Y} = \underline{X} \beta + \underline{\varepsilon}$
- Na de rakjuk össze a kettőt:

$$\begin{aligned} \widehat{\beta}_{OLS} &= (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T (\underline{X} \beta + \underline{\varepsilon}) = \\ &= (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{X} \beta + (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} = \beta + (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \end{aligned}$$

### Az OLS standard modellfeltevései

Ahhoz, hogy az OLS-nek fennálljanak bizonyos előnyös tulajdonságai, meghatározott feltevéseknek teljesülniük kell. Az ún. standard lineáris modell feltevései:

1. Linearitás
2. Nincs egzakt multikollinearitás
3. Erős (vagy szigorú) exogenitás
4. Homoszkedaszticitás
5. Autokorrelálatlanság

### Linearitás

A sokaságot *valójában* leíró modell tényleg a feltételezett, azaz fennáll, hogy

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

és ez igaz mindegyik megfigyelési egységre, és így az egész mintára is:

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

### Nincs egzakt multikollinearitás

- Egzakt multikollinearitásnak nevezzük, ha az adatmátrix nem teljes oszloprangú
- Tehát: az oszlopok között lineáris kapcsolat van
- Azaz valamelyik változó előállítható a többi lineáris kombinációjaként
- Érezhető, hogy nem túl szerencsés: minek használjuk egyáltalán azt a változót...? (Úgyis lineáris kombinációt képezünk a többiből is!) → a hatások nem lesznek szétválaszthatóak
- Sőt: az OLS becslőfüggvényéből az is látszik, hogy ilyenkor teljesen elakadunk:  $\mathbf{X}^T \mathbf{X}$  szinguláris ( $\underline{X}^T \underline{X}$  1 valószínűséggel szinguláris)
- Ennek feltétele:  $\mathbf{X}$  ( $\underline{X}$ ) nem teljes oszloprangú

### Nincs egzakt multikollinearitás

- A feltétel tehát: az adatmátrix 1 valószínűséggel legyen teljes oszloprangú:

$$\mathbb{P}(\text{rank } \underline{X} = k) = 1$$

- Ez implikálja, hogy  $n \geq k$  (kevesebb mint  $k$ -dimenziós vektorból nincs  $k$  független)

Érdemes megfigyelni, hogy a multikollinearitás egy mintában is elképzelhető jelenség (ezért is hivatkozhattunk rá úgy, hogy  $\mathbf{X}$  nem teljes oszloprangú), de mi a tulajdonságot a sokaságban akarjuk kikötni, ezért az állítás az  $\underline{X}$ -re kell vonatkozzon. Itt viszont csak annak van értelme, hogy „majdnem biztosan” (azaz 1 valószínűséggel) követeljük meg.

A másik alapvető szemléletes példa a multikollinearitásra az, ha van konstans a modellben és valamelyik magyarázó változónak nincs szórása (az összes megfigyelés ugyanaz rá). Könnyű elképzelni (pl. két dimenzióban), hogy nem lehet semmilyen regressziós egyenest húzni akkor, ha minden pontunk egymás fölött van.

Végezetül megjegyezzük, hogy gyakorlati jelentősége annak is lesz, ha ugyan nincs egzakt multikollinearitás, de van változó ami „elég jól” előállítható a többi lineáris kombinációjaként. Ezzel a kérdéskörrel fontossága miatt külön fogunk foglalkozni.

## Erős exogenitás

- Minden  $i = 1, 2, \dots, n$ -re

$$\mathbb{E}(\varepsilon_i | \underline{X}_i) = 0$$

- Tartalma: a hibák – az ún. várható érték függetlenség értelemben – függetlenek a magyarázó változóktól

Ha nem lenne a mintavétel, akkor azt kellene kikötni, hogy  $\mathbb{E}(\underline{\varepsilon} | \underline{X}) = 0$ . Ha azonban a mintavétel van, akkor az automatikusan teljesül, hogy  $\varepsilon_i$  minden  $\underline{X}_j$ -től független – és így várható érték független is – ha  $j \neq i$ .

Igazából elég azt kikötni, hogy  $\mathbb{E}(\varepsilon_i | \underline{X}_i) = \text{konst}$ , belátható, hogy ha van konstans a modellben, akkor ez egyenértékű a fentivel.

## Az erős exogenitás következményei

- Toronyszabály miatt a feltétel *nélküli* várható érték is nulla:

$$\mathbb{E}[\mathbb{E}(\varepsilon_i | \underline{X}_i)] = \mathbb{E}\varepsilon_i = \mathbb{E}(0) = 0$$

- A várható érték függetlenség implikálja a korrelátlanságot:  $\text{cov}(X_{ik}, \varepsilon_j) = 0$  avagy – ezzel egyenértékűen, hiszen  $\mathbb{E}\varepsilon_i = 0$  –  $\mathbb{E}(X_{ik}\varepsilon_j) = 0$
- Szokás a korrelátlanság helyett azt is mondani, hogy a hibák ortogonálisak a magyarázó-változókra

Az  $\mathbb{E}(\varepsilon_i | \underline{X}_i) = 0$  *szigorúan erősebb* követelmény mint a korrelátlanság. Ezt a fogalmat szokás várható érték függetlenségnek (mean independence) nevezni. Lássuk is ezt be, a jelölés megkönnyítése végett hívjuk egyszerűen  $\varepsilon$ -nak és  $X$ -nek a két változónkat. Emlékeztetőül  $\text{cov}(\varepsilon, X) = \mathbb{E}(\varepsilon X) - \mathbb{E}\varepsilon\mathbb{E}X$ . Egyik oldalról

$$\mathbb{E}(\varepsilon X) = \mathbb{E}[\mathbb{E}(\varepsilon X | X)] = \mathbb{E}[X\mathbb{E}(\varepsilon | X)] = \mathbb{E}[X \cdot 0] = \mathbb{E}0 = 0,$$

így  $\varepsilon$  és  $X$  korrelátlanak (hiszen  $\mathbb{E}\varepsilon = 0$ ); ezzel bebizonyítottuk, hogy a várható érték függetlenség implikálja a korrelátlanságot. Másik oldalról, tekintsük példaként az  $X \sim \mathcal{N}(0, 1)$  változót és egy olyan  $\varepsilon$  változót melyre  $\mathbb{E}(\varepsilon | X) = X^2$ . Ekkor  $\mathbb{E}X = 0$ ,  $\mathbb{E}\varepsilon = \mathbb{E}[\mathbb{E}(\varepsilon | X)] =$

$\mathbb{E}(X^2) = \mathbb{D}^2 X = 1$  és  $\mathbb{E}(\varepsilon X) = \mathbb{E}[\mathbb{E}(\varepsilon X | X)] = \mathbb{E}[X \mathbb{E}(\varepsilon | X)] = \mathbb{E}(X^3) = 0$ , így mutattunk egy példát, amikor a változók korrelálatlanok, de mégsem várható érték függetlenek.

A várható érték függetlenség tehát *szigorúan erősebb* fogalom, mint a korrelálatlanság. Érdekes azt is megjegyezni, hogy viszont *szigorúan gyengébb* mint az igazi függetlenség! (Ez könnyen belátható. Függetlenség esetén minden feltételes eloszlás ugyanaz, márpedig ekkor nyilván a várható értékük is ugyanaz. Másik oldalról tekintsünk egy origó körüli  $0 < r < R$  körgyűrűre koncentrált egyenletes eloszlást. Ez nyilván várható érték független – a feltételes várható érték konstans nulla –, viszont természetesen nem független.) Lényegében arról van szó, hogy a függetlenség a feltételes eloszlások *teljes* egyezőségét követeli meg, míg a várható érték függetlenség csak annyit, hogy a feltételes eloszlások várható értéke legyen egyező. (A feltételes szórás már kapásból lehet eltérő.)

A másik fontos megjegyzés itt, hogy az OLS-nek megvan az a tulajdonsága, hogy a *reziduumok* mindig korrelálatlanok a magyarázó változókkal. Valaki megkérdezheti, hogy akkor minek ezt külön is kikötni? Vigyázat! A reziduumok (amik a mintában vannak) tényleg mindig korrelálatlanok, csak hogy itt a hibákról beszélünk, amik a sokaságban vannak! A feltétel a *sokaságról* mond valamit (amit mi nem ismerhetünk soha biztosan – épp ettől feltétel!), nem a mintáról. A mintában a reziduumok természetesen mindig korrelálatlanra lesznek állítva, de ezzel nem sokra megyünk, ha a sokaságban nem teljesül ez a feltétel, hiszen ez esetben a reziduumoknak nem sok közük lesz a hibákhoz, pont ez a probléma...

### Az erős exogenitás sérülésének tipikus esetei

- Van olyan változó, ami lényeges magyarázó változó lenne (tehát valódi – sokasági –  $\beta$ -ja nem nulla), de mégsem szerepel a modellben, miközben legalább egy magyarázó változóval korrelál (kihagyott változó esete, „omitted variable bias”) – ez épp a confounding!
- Mérési hiba magyarázó változónál (tehát a mérési változók valódi értékét nem, csak valamilyen zajjal terhelve tudjuk mérni)
- Szimultaneitás (többegyenletes modelleknél)

Az első eset szolgáltatja a talán legjellemzőbb példákat a 'korreláció nem implikál kauzalitást' statisztikai alapelvére: ez a confounding, amit már részletesen tárgyaltunk. Szemléltessük ezt egy korábban már említett példán: emberek fizetését regresszáljuk ki az oktatásban töltött éveik számával (tehát az előbbi az eredmény-, az utóbbi az – egyetlen – magyarázó változó). Ekkor a hibába vélhetően olyan tényezők fognak beleszámítani, mint a nem-oktatással összefüggő munkaalkalmasság, a munkamorál, a szakmai tapasztalat stb. Az egyszerűség kedvéért mondjuk, hogy csak a legelső adja a hibát. Ekkor a szigorú exogenitás feltétele, a várható érték függetlenség azt fogalmazza meg, hogy a munkaalkalmasság feltételes várható értéke minden képzettség, mint feltétel mellett legyen ugyanakkora, tehát, hogy ne függjön a képzettségtől. (Amint mondtuk, konstans jelenléte esetén ez mellesleg azt jelenti, hogy nulla is legyen ez az állandó feltételes várható érték.) Baj akkor van, ha a képzettség különböző szintjei mellett a várható munkaalkalmasság *nem* állandó – tehát például a magasabb képzettségűeknek a munkaalkalmasságuk is nagyobb, azaz a nagyobb képzettséggel *együttal* a munkaalkalmasság is emelkedik. Ekkor megsérül a szigorú exogenitási feltétel. Épp innen kapta a feltétel a nevét: olyasmit fejez ki, hogy a magyarázó változókhoz képest exogén információ az, ami a hibákban össze van fogva. (Ez nyilván nem teljesül a fenti esetben.)

A második és harmadik kérdéskör boncolgatása meghaladja jelen kurzus kereteit.

Végül megjegyezzük, hogy idősoros esetben ez nagyon erős feltétel (hiszen például azt jelenti, hogy a magyarázó változóknak a múltbeli, a jelenbeli és a jövőbeli hibákra is ortogonálisnak kell

lenniük!), ami sokszor nem teljesül. (Példaként gondoljunk egy egyszerű késleltetett eredmény-változós modellre.)

### Az erős exogenitás sérülésének kezelése

- A problémát orvosolhatjuk a megfelelő(bb) modellspecifikációval, függően attól, hogy pontosan mi a baj oka. . .
- . . . illetve bizonyos statisztikai eszközök is a rendelkezésünkre állnak, ilyen az instrumentális változós (IV) becslés, a kétfázisú legkisebb négyzetek módszere (TSLS) stb.

E kérdések meghaladják jelen kurzus kereteit.

### Homoszkedaszticitás

- A feltétel azt köti ki, hogy  $\sigma_i^2 := \mathbb{D}^2(\varepsilon_i | \underline{X}) = \sigma^2$   $i$ -től függetlenül minden  $i = 1, 2, \dots, n$ -re
- Tartalma: a hibák különböző megfigyelésekhez tartozó szórása állandó (nem függ attól, hogy melyik megfigyelésről van szó) avagy – másként megfogalmazva ugyanez – a becslési értékek szóródása a tényleges körül állandó
- Jellemzően keresztmetszeti adatoknál felmerülő kérdés (hamarosan foglalkozunk is vele bővebben)

Nem fae mintavételezésnél azt kellene írunk, hogy  $\mathbb{D}^2(\varepsilon_i | \underline{X}) = \sigma^2$   $i$ -től függetlenül minden  $i = 1, 2, \dots, n$ -re.

A feltétellel egyenértékű, hogy  $\mathbb{E}(\varepsilon_i^2 | \underline{X}_i) = \sigma^2$  (hiszen  $\mathbb{E}(\varepsilon_i | \underline{X}_i) = 0$ , így a szórásnégyzet a négyzet várható értéke, azaz a második momentum).

Megjegyzendő, hogy fae mintavételezésnél az mindenképp teljesül, hogy  $\mathbb{D}^2\varepsilon_i$  konstans, de ez kevés: nekünk a *feltételes* szórás állandósága is kell a standard modellfeltevések között.

### Autokorrelálatlanság

- Tartalma: a különböző megfigyelésekhez tartozó hibák korrelálatlanok egymással
- Fae mintavételezésnél ez tehát *automatikusan* teljesül!
- Nem fae esetben a feltétel azt köti ki, hogy  $\text{cov}(\varepsilon_i, \varepsilon_j | \underline{X}) = 0$  minden  $i, j = 1, 2, \dots, n$ ,  $i \neq j$ -re
- Ezzel egyenértékű  $\mathbb{E}(\varepsilon_i \varepsilon_j | \underline{X}) = 0$  (hiszen  $\mathbb{E}\varepsilon_i = 0$ , így a kovariancia a két változó szorzatának várható értéke)
- Elsősorban idősoros adatok kérdésköre, most nem is foglalkozunk vele bővebben

## A homoszkedaszticitás és az autokorrelálatlanság együtt

- Mindkettő felfogható úgy, mint az  $\varepsilon_i$  hibák (feltételes) kovarianciamátrixára vonatkozó megkötés
  - Homoszkedaszticitás: a kovarianciamátrix főátlójában ugyanazok az elemek ( $\sigma^2$ ) vannak (ugye itt vannak a szórásnégyzetek)
  - Autokorrelálatlanság: a kovarianciamátrix főátlóján kívüli elemek nullák (a mátrix diagonális)
- A kettő *együtt*: a kovarianciamátrix  $\sigma^2 \mathbf{I}$  alakú (szokás az ilyet skalármátrixnak is nevezni)

Fae esetben  $\mathbb{D}^2 \underline{\varepsilon} = \mathbb{E}(\underline{\varepsilon} \underline{\varepsilon}^T) = \sigma^2 \mathbf{I}$ , nem fae esetben ki kell írni, hogy  $\mathbb{D}^2(\underline{\varepsilon} | \underline{X}) = \mathbb{E}(\underline{\varepsilon} \underline{\varepsilon}^T | \underline{X}) = \sigma^2 \mathbf{I}$ . (Az első egyenlőségek azért állnak fenn, mert  $\mathbb{E} \underline{\varepsilon} = \mathbf{0}$ ).

Szokás úgy is fogalmazni, hogy a hibavarianciák szferikálisak.

### $\sigma^2$ becslése

Nem részletezzük, de belátható, hogy ez esetben a  $\sigma^2$ -re adható OLS-becslés:

$$\widehat{\sigma^2} = \frac{ESS}{n - (k + 1)} = \frac{\widehat{\mathbf{e}}^T \widehat{\mathbf{e}}}{n - (k + 1)}$$

## Mintavételileg rögzített magyarázó változók

- Egyszerűbb tárgyalások azt feltételezik, hogy a magyarázó változók mintavételileg rögzítettek (mintha determinisztikusan megszabhatnánk az értéküket:  $\underline{X}_i$  igazából  $\mathbf{x}_i$ )
- Ennek sok baja van:
  1. Nem annyira szép és elegáns (nyilván ez speciális esete a mi tárgyalásunknak!)
  2. Nem teszi lehetővé egy sor kérdés mélyebb tárgyalását
  3. Alapjában megkérdőjelezhető az alkalmazása nem-experimentális tudományokban (mint a közgazdaságtan...)
- Az előnye, hogy egyszerűsít: ekkor a hiba feltételes és feltétel nélküli eloszlása ugyanaz lesz, a ' $\underline{X}_i$ ' jellegű feltételek elhagyhatóak...
- ...emiatt a modellfeltevések a következőkre egyszerűsödnek:
  - Erős exogenitás:  $\mathbb{E} \varepsilon_i = 0$  minden  $i = 1, 2, \dots, n$ -re
  - Homoszkedaszticitás:  $\mathbb{D}^2 \varepsilon_i = \sigma^2$  minden  $i = 1, 2, \dots, n$ -re
  - Autokorrelálatlanság:  $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$  minden  $i \neq j = 1, 2, \dots, n$

## A mintavételi tulajdonságok

- Ezek lesznek a standard modellfeltevések...
- ...most nekiállunk megvizsgálni, hogy a teljesülésük esetén milyen tulajdonságokkal bír az OLS-becselő



### Várható érték

- Tudjuk, hogy

$$\widehat{\beta}_{\text{OLS}} = \beta + (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon}$$

- Ez alapján mi  $\widehat{\beta}_{\text{OLS}}$  várható értéke (várható érték-vektora)?

$$\begin{aligned} \mathbb{E} \widehat{\beta}_{\text{OLS}} &= \beta + \mathbb{E} \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \right] = \\ &= \beta + \mathbb{E} \left\{ \mathbb{E} \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \mid \underline{X} \right] \right\} = \\ &= \beta + \mathbb{E} \left\{ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \mathbb{E} [\underline{\varepsilon} \mid \underline{X}] \right\} = \beta \end{aligned}$$

- Az erős exogenitás fennállása esetén tehát az OLS szolgáltatta becslések *torzítatlanok*
- Nem bizonyítjuk, de az is igaz, hogy *konzisztensek*

Az első lépésnél kihasználtuk, hogy a várható érték lineáris (összeg várható értéke a tagok várható értékeinek az összege) és hogy konstans várható értéke saját maga (ne feledjük, hogy  $\beta$  egy konstans! – nem tudjuk ugyan, hogy mennyi az értéke, de ettől még egy konstans). A második lépésnél a toronyszabályt használtuk:  $\mathbb{E} X = \mathbb{E} [\mathbb{E}(X \mid Y)]$ . A harmadik lépésnél pedig az erős exogenitást használtuk ki.

### Kovarianciamátrix

Az előbbi ismeretében:

$$\begin{aligned} \mathbb{D}^2 \widehat{\beta}_{\text{OLS}} &= \mathbb{E} \left[ (\widehat{\beta}_{\text{OLS}} - \mathbb{E} \widehat{\beta}_{\text{OLS}}) \cdot (\widehat{\beta}_{\text{OLS}} - \mathbb{E} \widehat{\beta}_{\text{OLS}})^T \right] = \\ &= \mathbb{E} \left[ (\widehat{\beta}_{\text{OLS}} - \beta) \cdot (\widehat{\beta}_{\text{OLS}} - \beta)^T \right] = \\ &= \mathbb{E} \left\{ \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \right] \cdot \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \right]^T \right\} = \\ &= \mathbb{E} \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \underline{\varepsilon}^T \underline{X} (\underline{X}^T \underline{X})^{-1} \right] = \\ &= (\underline{X}^T \underline{X})^{-1} \underline{X}^T \mathbb{E} (\underline{\varepsilon} \underline{\varepsilon}^T) \underline{X} (\underline{X}^T \underline{X})^{-1} = \\ &= (\underline{X}^T \underline{X})^{-1} \underline{X}^T \cdot \sigma^2 \mathbf{I} \cdot \underline{X} (\underline{X}^T \underline{X})^{-1} = \sigma^2 (\underline{X}^T \underline{X})^{-1} \end{aligned}$$

### A Gauss–Markov tétel

- Ha mindegyik feltevés teljesül, akkor lineár torzítatlan becslők körében az OLS-becslő minimális varianciájú (azaz hatásos)

- Tehát:  $\mathbb{D}^2(\widehat{\beta}_{OLS}) \leq \mathbb{D}^2(\widehat{\beta}')$  bármely más  $\widehat{\beta}'$  lineáris becslőre, amire  $\mathbb{E}(\widehat{\beta}') = \beta$  (azaz torzítatlan)

Emlékeztetőül, ha  $\mathbf{A}$  és  $\mathbf{B}$  négyzetes mátrixok, akkor abban az esetben mondjuk, hogy  $\mathbf{A} \geq \mathbf{B}$  ha  $\mathbf{A} - \mathbf{B}$  pozitív szemidefinit

### Összefoglalva

- Amennyiben a standard modellfeltevések közül teljesül a:
  - Linearitás
  - Nincs egzakt multikollinearitás
  - Erős exogenitás

akkor az OLS szolgáltatatta becslések *torzítatlanok és konzisztensek*

- Ha ezen felül teljesül a:
  - Homoszkedaszticitás
  - Autokorrelálatlanság

akkor az OLS szolgáltatatta becslések *hatásosak* (minimális varianciájuk) is

### BLUE-tulajdonság

Ezt röviden úgy szokták megfogalmazni, hogy ha valamennyi standard modellfeltétel teljesül, akkor az OLS szolgáltatatta becslések BLUE-k:

- Best (minimális varianciájú)
- Linear (lineáris a mintaelemekben)
- Unbiased (torzítatlan)

### A $\sigma^2$ és a koefficiensek kovarianciamátrixának becslői

- A  $\sigma^2$ -nek a  $\widehat{\sigma}^2 = \frac{ESS}{n-(k+1)}$  becslője torzítatlan, ha mindegyik feltétel fennáll
- A  $\beta_i$  koefficiensek kovarianciamátrixának  $\widehat{\sigma}^2 (\underline{X}^T \underline{X})^{-1}$  becslője szintén
- Tehát vigyázat: itt *már* a torzítatlansághoz *is* kell mindegyik feltétel (a homoszkedaszticitás és az autokorrelálatlanság is)!

### A $\widehat{\beta}_i$ koefficiensek eloszlása

- Az eddigi eredmények ugyan nagyon biztatóak, de még mindig nem mondanak semmit arról, hogy konkrétan mi a becslt koefficiensek (mintavételi) eloszlása
- A  $\widehat{\beta}_{OLS} = \beta + (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon}$  nem sok jót sejtet: ebből úgy tűnik, hogy ez  $\underline{X}$ -től és  $\varepsilon$ -től is függ, ráadásul egy elég komplexnek kinéző módon...
- Szerencsére nem ennyire rossz a helyzet!

- Van egy nevezetes speciális eset, amikor a becsült koeficiens eloszlása egyszerű alakú, és *nem is függ*  $\underline{X}$  eloszlásától, ez pedig az, ha a hibák feltételes eloszlása normális
- Vigyázat: a hibák normalitása *nem* része a standard modellfeltevéseknek, azaz a BLUE-ság akkor is megvalósul, ha a hibák eloszlása nem normális!
- Ráadásul, még ha nem is tudjuk, hogy a normalitás teljesül, de nagy a mintánk, akkor a centrális határeloszlás-tétel miatt aszimptotikus közelítésként akkor is használhatjuk az így nyert eredményeket

### Hibák normalitása

- $\underline{\varepsilon}$  feltételes eloszlása feltéve  $\underline{X}$ -et többváltozós normális
- A standard modellfeltevéseket is felhasználva ez azt jelenti, hogy

$$\underline{\varepsilon} \mid \underline{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

- Ez láthatóan nem függ  $\underline{X}$ -től, így persze a hibák feltétel nélküli eloszlása is  $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

### Hibanormalitás és a becsült koeficiens eloszlása

- Ha  $\underline{\varepsilon}$  eloszlása normális, akkor  $(\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon}$ -é is az
- Ez azért nagyon jó hír, mert a normális eloszláshoz csak két dolgot kell tudnunk: várható érték-vektort és kovarianciamátrixot!
- Az viszont könnyen meghatározható (az egyszerűség kedvéért a  $\mid \underline{X}$  feltételt nem írjuk ki a következőkben)
- Várható érték:  $\mathbb{E} \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \right] = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \mathbb{E} \underline{\varepsilon} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \mathbf{0} = \mathbf{0}$
- Kovarianciamátrix:  $\mathbb{D}^2 \left[ (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\varepsilon} \right] = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \cdot \mathbb{D}^2 \underline{\varepsilon} \cdot \underline{X} (\underline{X}^T \underline{X})^{-1} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \cdot \sigma^2 \mathbf{I} \cdot \underline{X} (\underline{X}^T \underline{X})^{-1} = \sigma^2 (\underline{X}^T \underline{X})^{-1}$
- Összefoglalva:  $\widehat{\beta}_{OLS} \sim \mathcal{N}(\beta, \sigma^2 (\underline{X}^T \underline{X})^{-1})$

Emlékezzünk rá, hogy  $\mathbb{E}(\mathbf{A}\mathbf{X}) = \mathbf{A}\mathbb{E}\mathbf{X}$  és  $\mathbb{D}^2(\mathbf{A}\mathbf{X}) = \mathbf{A}\mathbb{D}^2\mathbf{X}\mathbf{A}^T$ . (Itt az  $\underline{X}$ -ket tartalmazó kifejezések azért viselkednek úgy, mint a konstans mátrixok, mert rá feltételeztünk! Csak a rövidség kedvéért nem írtuk ki.)

## 3. Becslés és hipotézisvizsgálat a lineáris modellben

### 3.1. Alkalmazási feltételek

#### Emlékeztetőül

- A most következő eredmények csak akkor egzakta, ha a hibanormalitás is fennáll
- Ám aszimptotikusak, így közelítőleg akkor is fennállnak, ha elég nagy a mintanagyság (minél nagyobb, annál inkább)

## 3.2. Egy paraméter

### Becsült regressziós koefficiensek mintavételi eloszlása

- A  $\widehat{\beta}_i$  becsült regressziós koefficiens mintavételi ingadozását tehát a következő összefüggés írja le:

$$\frac{\widehat{\beta}_i - \beta_i}{\text{se}(\widehat{\beta}_i)} \sim \mathcal{N}(0, 1),$$

$$\text{ahol } \text{se}(\widehat{\beta}_i) = \sqrt{\sigma^2 \left[ (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \right]_{kk}}$$

- Sajnos ezzel a gyakorlatban nem sokra megyünk, mert  $\sigma^2$ -et általában nem ismerjük
- Helyettesítsük a jó tulajdonságú becslőjével,  $\widehat{\sigma}^2$ -tel!
- Így persze már más lesz az eloszlás, de szerencsére meghatározható, hogy mi, és nem bonyolult:  $n - (k + 1)$  szabadságfokú  $t$ -eloszlás

### Változó relevanciája

Egy változót relevánsnak nevezünk, ha a sokasági paramétere nem nulla:  $\beta_i \neq 0$ .

### Hipotézisvizsgálat változó relevanciájára

Ez alapján már konstruálhatunk próbát változó relevanciájának vizsgálatára:

1.  $H_0 : \beta_i = 0$
2. Ekkor (azaz ha ez fennáll!) a  $t_{\text{emp},i} = \frac{\widehat{\beta}_i}{\text{se}(\widehat{\beta}_i)}$  kifejezés  $n - (k + 1)$  szabadságfokú  $t$ -eloszlást követ (nulleloszlás)
3. Számítsuk ki a konkrét  $t_{\text{emp},i}$ -t a mintánkból és döntsük el, hogy hihető-e, hogy  $t_{n-(k+1)}$ -ből származik

### Hipotézisvizsgálat változó relevanciájára

A hipotézisvizsgálat elvégzéséhez szükséges minden tudnivalót – a nullhipotézisen kívül – összefoglal tehát a következő kifejezés (a későbbiekben is ezt a sémát fogjuk használni hipotézisvizsgálatok megadására):

$$t_{\text{emp},i} = \frac{\widehat{\beta}_i}{\text{se}(\widehat{\beta}_i)} \stackrel{H_0}{\sim} t_{n-(k+1)}.$$

E próba precíz neve: változó relevanciájára irányuló (parciális)  $t$ -próba

### Konfidenciintervallum a paraméterekre

Ez alapján könnyen szerkeszthető CI is,  $1 - \alpha$  megbízhatósági szinten:

$$\widehat{\beta}_i \pm t_{n-(k+1)}^{(1-\alpha/2)} \cdot \text{se}(\widehat{\beta}_i)$$

### 3.3. Modell egésze

#### Modell egészének relevanciája

- A korábban látott  $t$ -próba azért volt „parciális”, mert egy változó irrelevanciáját vizsgálta
- Felmerül a kérdés, hogy definiálható-e a modell *egészének* irrelevanciája
- Igen, mégpedig úgy, hogy *valamennyi* magyarázó változó *együttesen is* irreleváns:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

- Rövid jelölés arra, hogy  $\beta_1 = 0$  és  $\beta_2 = 0$  stb. és  $\beta_k = 0$  (semmilyen más eset jelölésére *ne* használjuk az egyenlőségláncot!)
- Figyelem: az „egyszerre nulla mindegyik” *több* mint, hogy „külön-külön nulla mindegyik”!

#### Modell egészének relevanciája

- A modell egészének irrelevanciájára magyarul azt jelenti, hogy a modell nem tér el lényegesen a nullmodellről
- Implikálja, hogy minden magyarázó változó külön-külön is irreleváns (tartalmazza ezeket a hipotéziseket)  $\rightarrow$  előbb teszteljük a modell egészének irrelevanciáját, és csak ennek elvetése után teszteljük a változókat parciálisan
- A próba konkrét alakja:

$$F_{\text{emp}} = \frac{RSS/k}{ESS/(n-k-1)} \stackrel{H_0}{\sim} F_{k, n-k-1}$$

#### Modell egészének relevanciája

- A tesztstatisztika átírható mint

$$\frac{RSS/k}{ESS/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

- Persze: a „nem tér el lényegesen a nullmodellről” úgy is megfogalmazható, hogy az „ $R^2$  nem tér el lényegesen a nullától” ( $H_0 : R^2 = 0$  is mondható lett volna)

#### Modell egészének relevanciája

- A próba neve: a modell egészének relevanciájára irányuló (globális)  $F$ -próba
- Szokás ANOVA-próbának is nevezni (a  $TSS = ESS + RSS$  variancia-felbontáson alapszik; számlálóban és nevezőben a fokszámmal normált szórásnégyzetek vannak)
- Tipikus eredményközlés az ún. ANOVA-táblában

### 3.4. Lineáris megkötés(ek)

#### Lineáris kombináció tesztelése

- A séma:

$$\lambda_{\beta_1}\beta_1 + \lambda_{\beta_2}\beta_2 + \dots + \lambda_{\beta_k}\beta_k = \Lambda$$

- Több koefficiens is érinthet, de csak egy egyenletet tartalmazhat
- Például:
  - Két koefficiens egyezik,  $\beta_l = \beta_m$  (ekkor  $\lambda_{\beta_l} = +1$ ,  $\lambda_{\beta_m} = -1$ , a többi  $\lambda$  nulla és  $\Lambda = 0$ )
  - Egyik koefficiens  $c$ -szerese a másiknak,  $\beta_l = c\beta_m$  (ekkor  $\lambda_{\beta_l} = +1$ ,  $\lambda_{\beta_m} = -c$ , a többi  $\lambda$  nulla és  $\Lambda = 0$ )
  - Az összes koefficiens összege épp nulla (ekkor mindegyik  $\lambda$  1 és  $\Lambda = 0$ )

#### Lineáris kombináció tesztelése

- A normális lineáris modellben erre teszt szerkeszthető
- Megvalósítás: egyik lehetőség, hogy a  $t$ -próbához hasonló alakra vezetjük vissza
- Legyen  $\lambda_{\beta_1}\hat{\beta}_1 + \lambda_{\beta_2}\hat{\beta}_2 + \dots + \lambda_{\beta_k}\hat{\beta}_k = \hat{\Lambda}$ , ekkor

$$\frac{\hat{\Lambda} - \Lambda}{\text{se}(\hat{\Lambda})} \stackrel{H_0}{\sim} t_{n-k-1}$$

- Ez az ún. *közvetlen  $t$ -próba*
- Vizsgálható Wald-jellegű próbával is (most nem foglalkozunk vele bővebben, de a `gret1` ezt használja)

#### Több megkötés egyidejű tesztelése

- Az eddigiek kombinálhatóak is: több megkötés (mindegyikük lineáris kombináció), melyeknek egyszerre kell teljesülniük
- Célszerű felírás:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r},$$

ahol  $\mathbf{R}$   $m \times k$  típusú (tehát  $m$  a megszorítások száma)

- Az erre adható teszt:

$$F_{\text{emp}} = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T [\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / m}{\text{ESS} / (n - k - 1)} \stackrel{H_0}{\sim} \mathcal{F}(m, n - k - 1)$$

Feltétel még, hogy  $\mathbf{R}$  teljes sorrangú legyen (rank  $\mathbf{R} = m$ ), ami azt a kézenfekvő követelményt fogalmazza meg, hogy a megszorítások ne legyenek (lineáris értelemben) redundánsak.

### Konkrét példák a fenti sémára

- Ellenőrizhető, hogy ha például...

– ... $\mathbf{R} = (0 \ 0 \ \dots \ 0 \ 1 \ 0 \dots 0)$  és  $r = 0$ , akkor a  $t$ -tesztet ...

$$- \dots \mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \text{ és } \mathbf{r} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ akkor az ANOVA-t...}$$

– ... $\mathbf{R} = (\lambda_{\beta_1} \ \lambda_{\beta_2} \ \dots \ \lambda_{\beta_k})$  és  $r = \Lambda$ , akkor a lineáris kombináció tesztelését...

- ...kapjuk vissza.