Az OLS becslő modellfeltevései és az OLS szolgáltatta becslések statisztikai tulajdonságai

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 12.

Tartalom

Tartalomjegyzék

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 hátráltató 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** ($\boldsymbol{\beta}$)
- Az adatbázisunk alapján megkaptuk a regressziós egyenest ($\widehat{\boldsymbol{\beta}}$)
- Az adatbázis azonban csak egy **minta** a sokaságából, így a $\widehat{\beta}_i$ 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 $\widehat{\beta}_i$ 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 szeretnék 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 $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$)
- 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 – $\widehat{\boldsymbol{\beta}_{\text{OLS}}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$ – 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 $\widehat{\boldsymbol{\beta}_{\text{OLS}}} = \left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{Y}$ fog érdekelni!
- Ahogy az előbbi átlagos példában, így itt is igaz lesz, hogy ekkor a $\widehat{\underline{\boldsymbol{\beta}}_{\text{OLS}}}$ 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. 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

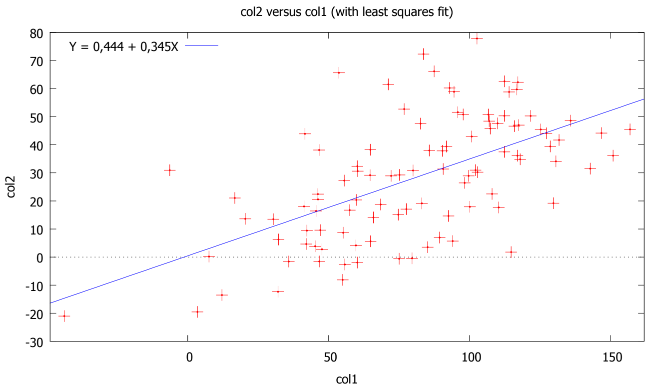
$$(X,Y) \sim \mathcal{N}\left(\begin{pmatrix}77\\26\end{pmatrix}, \begin{pmatrix}42^2 & 0{,}6\cdot 20\cdot 42\\ 0{,}6\cdot 20\cdot 42 & 20^2\end{pmatrix}\right)$$

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

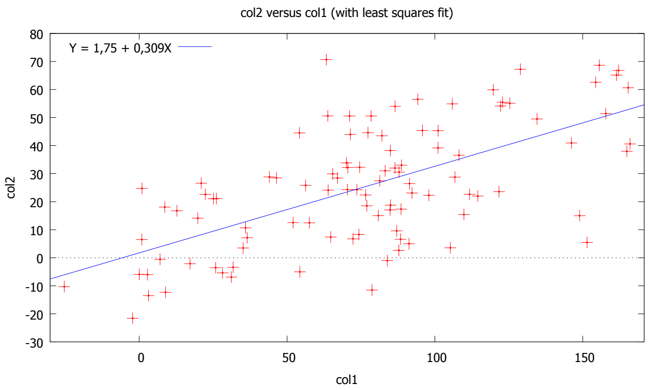
$$\mathbb{E}\left(Y \mid X\right) = 4 + \frac{12}{42}X \approx 4 + 0{,}2857X$$

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

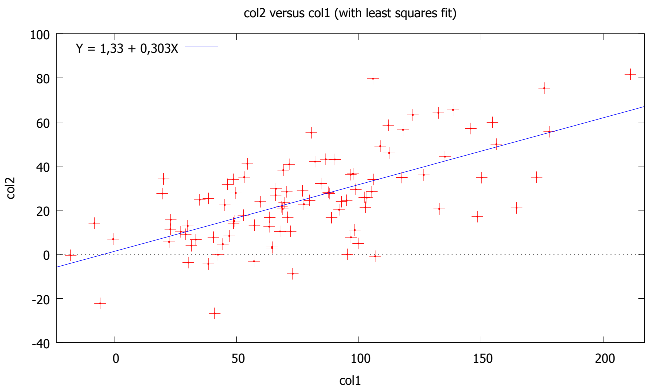
A szimuláció eredményei: 1. futtatás



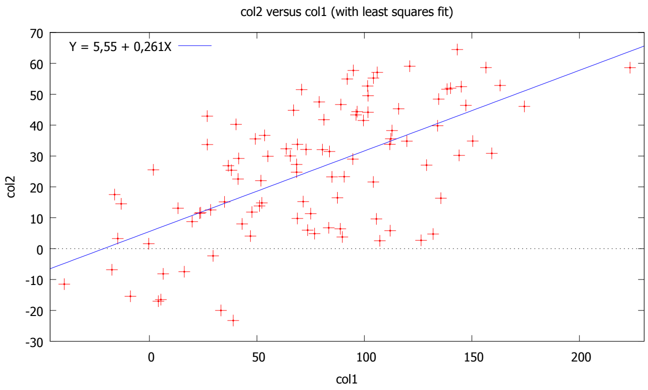
A szimuláció eredményei: 2. futtatás



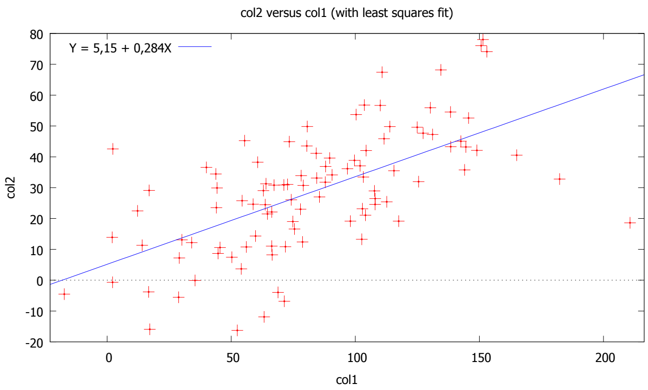
A szimuláció eredményei: 3. futtatás



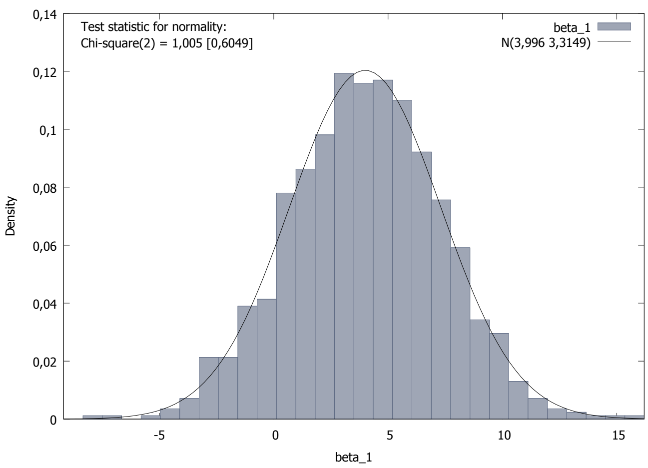
A szimuláció eredményei: 4. futtatás



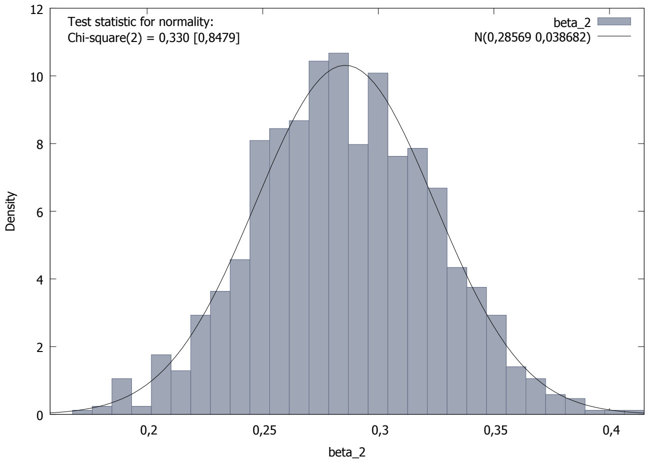
A szimuláció eredményei: 5. futtatás



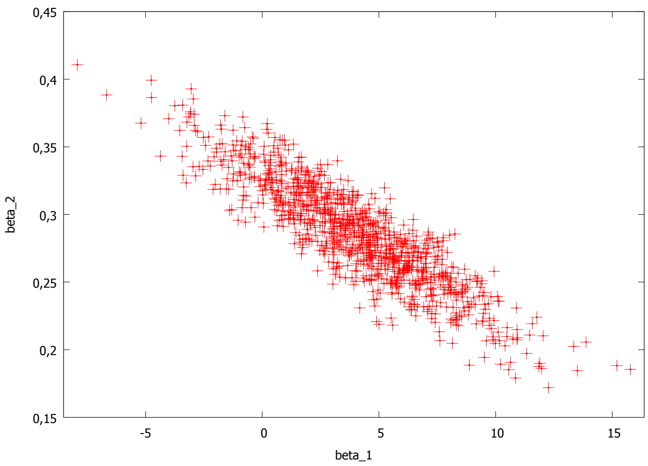
A szimuláció eredményei: konstans



A szimuláció eredményei: meredekség



A szimuláció eredményei: mindkét paraméter együtt



3. A mintavétel tulajdonságok matematikai levezetése

Az OLS-becslő mintavételi eloszlása

- Tudjuk, hogy $\widehat{\boldsymbol{\beta}_{\text{OLS}}} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{Y}$
- Valamint elfogadtuk feltételezésként, hogy a sokasági modell $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon = \underline{X}^T \boldsymbol{\beta} + \varepsilon$
  - És ez van mindegyik megfigyelési egység mögött is, tehát a mintavétel elemzéséhez ezt is írhatjuk:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i$$

- Röviden, értelemszerű vektorokba/mátrixokba fogással: $Y_i = \underline{X}_i^T \boldsymbol{\beta} + \varepsilon$ avagy az egész adatbázisra: $\underline{Y} = \underline{\underline{X}}\boldsymbol{\beta} + \underline{\varepsilon}$

- Na de rakjuk csak össze a kettőt:

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

4. Az OLS modellfeltevései

Az OLS standard modellfeltevési

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

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

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

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

Nincs egzakt multikollinearitás

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

Nincs egzakt multikollinearitás

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

$$\mathbb{P}\left(\operatorname{rank} \underline{\underline{X}} = k+1\right) = 1$$

- Ez implikálja, hogy $n \geq k+1$ (kevesebb mint $k+1$-dimenziós vektorból nincs $k+1$ 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 $\underline{\underline{X}}$-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 $i = 1, 2, \ldots, n$-re
$$\mathbb{E}\left(\varepsilon_i \mid \underline{X}_i\right) = 0$$

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

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

Igazából elég azt kikötni, hogy $\mathbb{E}\left(\varepsilon_i \mid \underline{X}_i\right) = \text{konst}$, 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:

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

- A várható érték függetlenség implikálja a korrelálatlanságot: $\operatorname{cov}\left(X_{ik}, \varepsilon_j\right) = 0$ avagy – ezzel egyenértékűen, hiszen $\mathbb{E}\varepsilon_i = 0$ – $\mathbb{E}\left(X_{ik}\varepsilon_j\right) = 0$

- Szokás a korrelálatlanság helyett azt is mondani, hogy a hibák ortogonálisak a magyarázóváltozókra

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

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

így ε és X korrelálatlanok (hiszen $\mathbb{E}\varepsilon = 0$); ezzel bebizonyítottuk, hogy a várható érték függetlenség implikálja a korrelálatlanságot. Másik oldalról, tekintsük példaként az $X \sim \mathcal{N}(0,1)$ változót és egy olyan ε változót melyre $\mathbb{E}\left(\varepsilon|X\right) = X^2$. Ekkor $\mathbb{E}X = 0$, $\mathbb{E}\varepsilon = \mathbb{E}\left[\mathbb{E}\left(\varepsilon \mid X\right)\right] = \mathbb{E}\left(X^2\right) = \mathbb{D}^2 X = 1$ és $\mathbb{E}(\varepsilon X) = \mathbb{E}\left[\mathbb{E}\left(\varepsilon X \mid X\right)\right] = \mathbb{E}\left[X\mathbb{E}\left(\varepsilon \mid X\right)\right] = \mathbb{E}\left(X^3\right) = 0$, így mutattunk egy példát, amikor a változók korrelálatlanok, de mégsem várható érték függetlenek.

A várható érték függetlenség tehát *szigorúan erősebb* fogalom, mint a korrelálatlanság. Érdemes 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, csakhogy 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 – β-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 *egyúttal* 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 eredményváltozós 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 $\sigma_i^2 := \mathbb{D}^2\left(\varepsilon_i \mid \underline{\underline{X}}\right) = \sigma^2$ i-től függetlenül minden $i = 1, 2, \ldots, n$-re
- Tartalma: a hibák különböző megfigyelésekhez tartozó szórása állandó (nem függ attól, hogy melyik megfigyelésről van szó) avagy – másként megfogalmazva ugyanez – a becsült értékek szóródása a tényleges körül állandó

- Jellemzően keresztmetszeti adatoknál felmerülő kérdés (hamarosan foglalkozunk is vele bővebben)

Nem fae mintavételezésnél azt kellene írnunk, hogy $\mathbb{D}^2\left(\varepsilon_i \mid \underline{\underline{X}}\right) = \sigma^2$ i-től függetlenül minden $i = 1, 2, \ldots, n$-re.

A feltétellel egyenértékű, hogy $\mathbb{E}\left(\varepsilon_i^2 \mid \underline{X}_i\right) = \sigma^2$ (hiszen $\mathbb{E}\left(\varepsilon_i \mid \underline{X}_i\right) = 0$, 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 $\mathrm{cov}\left(\varepsilon_i, \varepsilon_j \mid \underline{\underline{X}}\right) = 0$ minden $i, j = 1, 2, \ldots, n$, $i \neq j$-re
- Ezzel egyenértékű $\mathbb{E}\left(\varepsilon_i\varepsilon_j \mid \underline{\underline{X}}\right) = 0$ (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 ε_i hibák (feltételes) kovarianciamátrixára vonatkozó megkötés
  - Homoszkedaszticitás: a kovarianciamátrix főátlójában ugyanazok az elemek (σ^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 $\mathbb{D}^2\underline{\varepsilon} = \mathbb{E}\left(\underline{\varepsilon}\underline{\varepsilon}^T\right) = \sigma^2\mathbf{I}$, nem fae esetben ki kell írni, hogy $\mathbb{D}^2\left(\underline{\varepsilon} \mid \underline{\underline{X}}\right) = \mathbb{E}\left(\underline{\varepsilon}\underline{\varepsilon}^T \mid \underline{\underline{X}}\right) = \sigma^2\mathbf{I}$. (Az első egyenlőségek azért állnak fenn, mert $\mathbb{E}\underline{\varepsilon} = \mathbf{0}$).

Szokás úgy is fogalmazni, hogy a hibavarianciák szferikálisak.

σ^2 becslése

Nem részletezzük, de belátható, hogy ez esetben a σ^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 '$\mid \underline{X}_i$' jellegű feltételek elhagyhatóak...
- ...emiatt a modellfeltevések a következőkre egyszerűsödnek:
  - Erős exogenitás: $\mathbb{E}\varepsilon_i = 0$ minden $i = 1, 2, \ldots, n$-re
  - Homoszkedaszticitás: $\mathbb{D}^2\varepsilon_i = \sigma^2$ minden $i = 1, 2, \ldots, n$-re
  - Autokorrelálatlanság: $\mathbb{E}\left(\varepsilon_i\varepsilon_j\right) = 0$ minden $i \neq j = 1, 2, \ldots, n$

A mintavételi tulajdonságok

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

Várható érték

- Tudjuk, hogy
$$\widehat{\boldsymbol{\beta}_{\text{OLS}}} = \boldsymbol{\beta} + \left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{\varepsilon}$$
- Ez alapján mi $\widehat{\boldsymbol{\beta}_{\text{OLS}}}$ várható értéke (várható érték-vektora)?
$$\begin{aligned}\mathbb{E}\widehat{\boldsymbol{\beta}_{\text{OLS}}} &= \boldsymbol{\beta} + \mathbb{E}\left[\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{\varepsilon}\right] = \\ &= \boldsymbol{\beta} + \mathbb{E}\left\{\mathbb{E}\left[\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{\varepsilon} \mid \underline{\underline{X}}\right]\right\} = \\ &= \boldsymbol{\beta} + \mathbb{E}\left\{\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\mathbb{E}\left[\underline{\varepsilon} \mid \underline{\underline{X}}\right]\right\} = \boldsymbol{\beta}\end{aligned}$$
- 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 $\boldsymbol{\beta}$ egy konstans! – nem tudjuk ugyan, hogy mennyi az értéke, de ettől még egy konstans). A második lépésnél a toronyszabályt használtuk: $\mathbb{E}X = \mathbb{E}\left[\mathbb{E}\left(X \mid Y\right)\right]$. A harmadik lépésnél pedig az erős exogenitást használtuk ki.

Kovarianciamátrix

Az előbbi ismeretében:

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

A Gauss–Markov tétel

- Ha mindegyik feltevés teljesül, akkor lineáris torzítatlan becslők körében az OLS-becslő minimális varianciájú (azaz hatásos)
- Tehát: $\mathbb{D}^2\left(\widehat{\boldsymbol{\beta}_{\mathrm{OLS}}}\right) \leq \mathbb{D}^2\left(\widehat{\boldsymbol{\beta}'}\right)$ bármely más $\widehat{\boldsymbol{\beta}'}$ lineáris becslőre, amire $\mathbb{E}\left(\widehat{\boldsymbol{\beta}'}\right) = \boldsymbol{\beta}$ (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 szolgáltatta becslések *hatásosak* (minimális varianciájuk) is

BLUE-tulajdonság

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

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

A σ^2 és a koefficiensek kovarianciamátrixának becslői

- A σ^2-nek a $\widehat{\sigma^2} = \frac{ESS}{n-(k+1)}$ becslője torzítatlan, ha mindegyik feltétel fennáll
- A β_i koefficiensek kovarianciamátrixának $\widehat{\sigma^2}\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}$ 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 becsült koefficiensek (mintavételi) eloszlása
- A $\widehat{\boldsymbol{\beta}_{\mathrm{OLS}}} = \boldsymbol{\beta} + \left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{\varepsilon}$ nem sok jót sejtet: ebből úgy tűnik, hogy ez $\underline{\underline{X}}$-től és ε-tól is függ, ráadásul egy elég komplexnek kinéző módon…
- Szerencsére nem ennyire rossz a helyzet!
- Van egy nevezetes speciális eset, amikor a becsült koefficiensek eloszlása egyszerű alakú, és *nem is függ* $\underline{\underline{X}}$ 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 $\underline{\underline{X}}$-et többváltozós normális
- A standard modellfeltevéseket is felhasználva ez azt jelenti, hogy

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

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

Hibanormalitás és a becsült koefficiensek eloszlása

- Ha $\underline{\varepsilon}$ eloszlása normális, akkor $\left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{\varepsilon}$-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 $\mid \underline{\underline{X}}$ feltételt nem írjuk ki a következőkben)
- Várható érték: $\mathbb{E}\left[\left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{\varepsilon}\right] = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \mathbb{E}\underline{\varepsilon} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \mathbf{0} = \mathbf{0}$
- Kovarianciamátrix: $\mathbb{D}^2\left[\left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{\varepsilon}\right] = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \cdot \mathbb{D}^2 \varepsilon \cdot \underline{\underline{X}} \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \cdot \sigma^2 \mathbf{I} \cdot \underline{\underline{X}} \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} = \sigma^2 \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1}$
- Összefoglalva: $\widehat{\boldsymbol{\beta}_{\text{OLS}}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1}\right)$

Emlékezzünk rá, hogy $\mathbb{E}\left(\mathbf{A}\underline{X}\right) = \mathbf{A} \cdot \mathbb{E}\underline{X}$ és $\mathbb{D}^2\left(\mathbf{A}\underline{X}\right) = \mathbf{A} \cdot \mathbb{D}^2 \underline{X} \cdot \mathbf{A}^T$.

A fenti levezetésekben az $\underline{\underline{X}}$-ket tartalmazó kifejezések azért viselkednek úgy, mint a konstans mátrixok, mert rá feltételeztünk! Csak a rövidség kedvéért nem írtuk ki.

Konfidenciaintervallum a paraméterekre

Hibanormalitás esetén, vagy aszimptotikusan könnyen szerkeszthető konfidenciaintervallum is, $1-\alpha$ megbízhatósági szinten:

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