Regresszió a mintában: következtetés

Ferenci Tamás
tamas.ferenci@medstat.hu

2018. február 7.

Tartalom

1. Az OLS elv és használata a lineáris regresszió becslésére
   - Az OLS-elv
   - A lineáris regresszió becslése tisztán deskriptíve
   - Modellminősítés tisztán deskriptíve
2. Az OLS mintavételi tulajdonságai
   - Mintavételi helyzet
   - A mintavétel tulajdonságok szemléltetése szimulációval
   - A mintavétel tulajdonságok matematikai levezetése
3. Becslés és hipotézisvizsgálat a lineáris modellben
   - Alkalmazási feltételek
   - Egy paraméter
   - Modell egésze
   - Lineáris megkötés(ek)

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
- Jelesül $Y \sim \mathcal{N}\left(\mu, \sigma_0^2\right)$, amit nem mellesleg úgy is írhatnánk, hogy $Y = \mu + \varepsilon$, ahol $\varepsilon \sim \mathcal{N}\left(0, \sigma_0^2\right)$
- 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:
$$\widehat{y_i} = m$$

Az OLS-elv

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

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

- (Aminek a megoldása természetesen $\widehat{\mu} = \overline{y}$)

A mintavétel a lineáris regressziós feladatban

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

Lineáris regresszió becslése OLS-elven

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

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

- A becsült értékek adott $b_0, b_1, \ldots, b_k$ sokasági paraméterek mellett:

$$\widehat{y_i} = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_k x_{ik}$$

- A feladat tehát ugyanaz:

$$\left(\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}, \ldots, \widehat{\beta_k}\right) = \underset{b_0, b_1, b_2, \ldots, b_k}{\arg\min} \sum_{i=1}^{n} \left(y_i - \widehat{y_i}\right)^2 =$$

$$= \underset{b_0, b_1, b_2, \ldots, b_k}{\arg\min} \sum_{i=1}^{n} \left[y_i - \left(b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_k x_{ik}\right)\right]^2$$

- 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:

$$\arg\min_{\mathbf{b}} \left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)^T \left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)$$

- Az $\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)^T \left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)$ hibanégyzetösszeget *ESS*-sel (error sum of squares) is fogjuk jelölni

Az OLS-becslési feladat megoldása

A megoldás:

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

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

$$\frac{\partial}{\partial \mathbf{b}}\left[\mathbf{y}^T\mathbf{y} - 2\mathbf{b}^T\mathbf{X}^T\mathbf{y} + \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{b}\right] =$$
$$= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\mathbf{b} = 0 \Rightarrow \widehat{\beta_{\mathrm{OLS}}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y},$$

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

Pár további gondolat

- Az ún. reziduumok:
$$\widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}}$$
- Az előrejelzések a mintánkra:
$$\widehat{\mathbf{y}} = \mathbf{X}\widehat{\beta} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$
- Ez alapján vezessük be a
$$\mathbf{P} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$$
mátrixot, ezzel $\widehat{\mathbf{y}} = \mathbf{P}\mathbf{y}$
- 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

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 $\rightarrow \widehat{e_i} = 0 \Rightarrow ESS = 0$
  - 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:

$$TSS = ESS_{\mathrm{null}} = \sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2 .$$

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:

$$RSS = TSS - ESS.$$

Az új mutató bevezetése

Ezzel az alkalmas modelljellemző 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 jóságának legszéleskörűbben 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:

$$\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2=\sum_{i=1}^{n}\left(y_i-\widehat{y}_i\right)^2+\sum_{i=1}^{n}\left(\widehat{y}_i-\overline{y}\right)^2$$

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

$$TSS=ESS+RSS$$

- Összevetve az előző definícióval, kapjuk, hogy

$$RSS=\sum_{i=1}^{n}\left(\widehat{y}_i-\overline{y}\right)^2$$

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!

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** ($\boldsymbol{\beta}$)
- 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

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 – $\widehat{\beta_{\mathrm{OLS}}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$ – 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 $\widehat{\beta_{\mathrm{OLS}}} = \left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{Y}$ fog érdekelni!
- Ahogy az előbbi átlagos példában, így itt is igaz lesz, hogy ekkor a $\widehat{\underline{\beta}_{\mathrm{OLS}}}$ 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

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

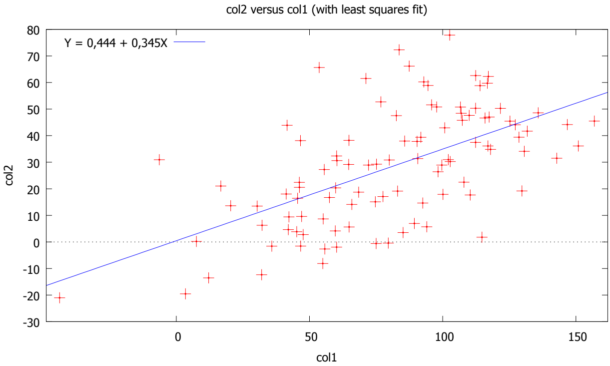
$$(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} 77 \\ 26 \end{pmatrix}, \begin{pmatrix} 42^2 & 0{,}6 \cdot 20 \cdot 42 \\ 0{,}6 \cdot 20 \cdot 42 & 20^2 \end{pmatrix}\right)$$

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

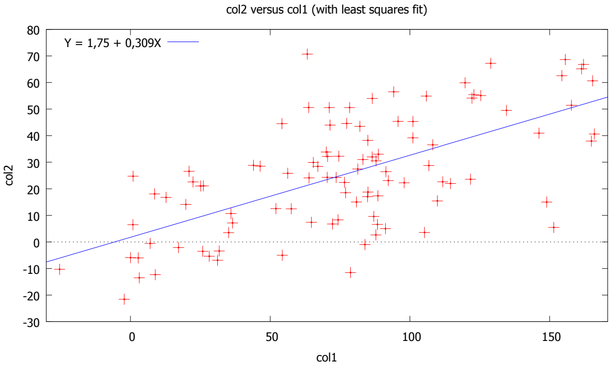
$$\mathbb{E}\left(Y \mid X\right) = 4 + \frac{12}{42} X \approx 4 + 0{,}2857 X$$

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

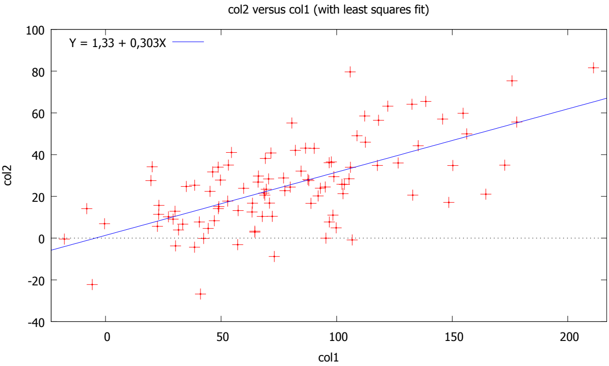
A szimuláció eredményei: 1. futtatás



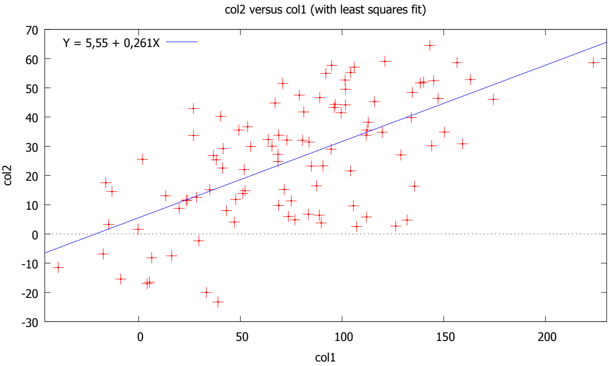
A szimuláció eredményei: 2. futtatás



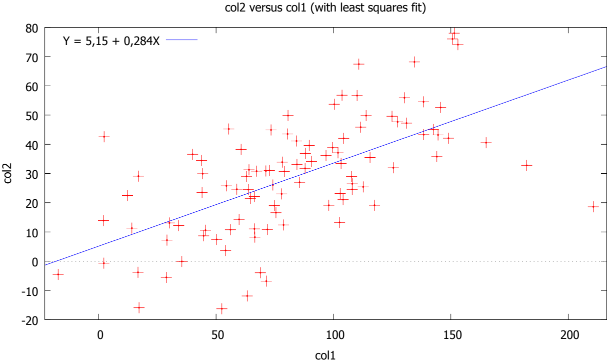
A szimuláció eredményei: 3. futtatás



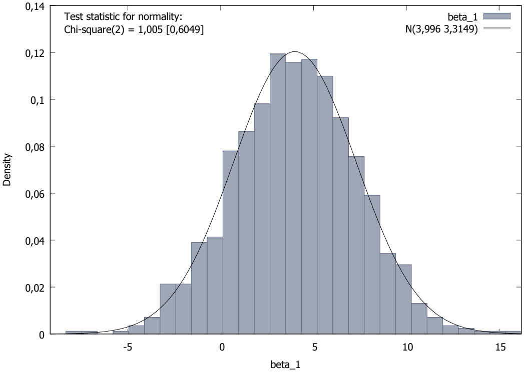
A szimuláció eredményei: 4. futtatás



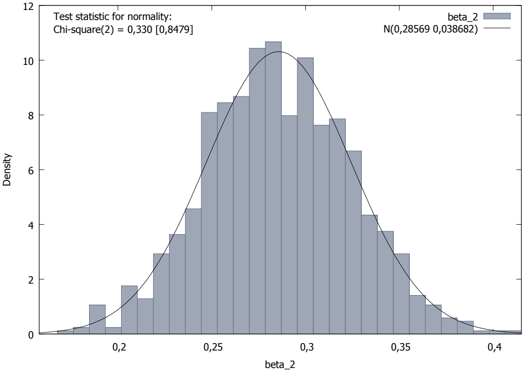
A szimuláció eredményei: 5. futtatás



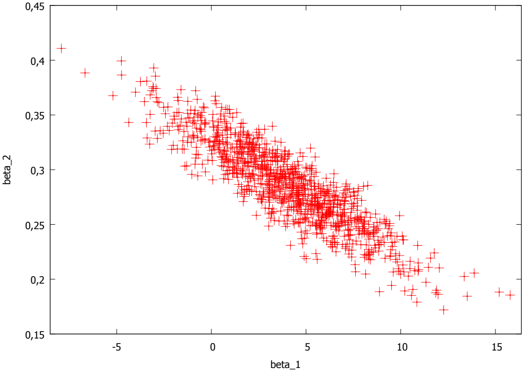
A szimuláció eredményei: konstans



A szimuláció eredményei: meredekség



A szimuláció eredményei: mindkét paraméter együtt



Az OLS-becslő mintavételi eloszlása

- Tudjuk, hogy $\widehat{\boldsymbol{\beta}_{\mathrm{OLS}}} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{Y}$
- 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:

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

$$= \left(X^T X\right)^{-1} \underline{\underline{X}}^T \underline{\underline{X}}\boldsymbol{\beta} + \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{\varepsilon} = \boldsymbol{\beta} + \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{\varepsilon}$$

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 többi 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:
$$\mathbb{P}\left(\operatorname{rank}\underline{\underline{X}} = k\right) = 1$$
- Ez implikálja, hogy $n \geq k$ (kevesebb mint k-dimenziós vektorból nincs k független)

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

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: $\mathrm{cov}\left(X_{ik}, \varepsilon_j\right) = 0$ 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 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 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.

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)

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)

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

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

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

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

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

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 lineár 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)

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:
$$\mathbb{D}^2\left[\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\underline{\varepsilon}\right] = \left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\cdot\mathbb{D}^2\underline{\varepsilon}\cdot\underline{\underline{X}}\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1} =$$
$$\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\underline{\underline{X}}^T\cdot\sigma^2\mathbf{I}\cdot\underline{\underline{X}}\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1} = \sigma^2\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}$$
- Összefoglalva: $\widehat{\boldsymbol{\beta}_{\mathrm{OLS}}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2\left(\underline{\underline{X}}^T\underline{\underline{X}}\right)^{-1}\right)$

Emlékeztetőül

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

Becsült regressziós koefficiensek mintavételi eloszlása

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

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

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

- Sajnos ezzel a gyakorlatban nem sokra megyünk, mert σ^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 $t_{\mathrm{emp},i} = \frac{\widehat{\beta_i}}{\mathrm{se}\left(\widehat{\beta_i}\right)}$ kifejezés $n-(k+1)$ szabadságfokú t-eloszlást követ (nulleloszlás)
3. Számítsuk ki a konkrét $t_{\mathrm{emp},i}$-t a mintánkból és döntsük el, hogy hihető-e, hogy $t_{n-(k+1)}$-ból 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):

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

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

Konfidenciaintervallum a paraméterekre

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

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

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:
$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
- 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 nullmodelltől
- Implikálja, hogy minden magyarázó változó külön-külön is irreleváns (tartalmazza ezeket a hipotéziseket) $\rightarrow$ előbb teszteljük a modell egészének irrelevanciáját, és csak ennek elvetése utána teszteljük a változókat parciálisan
- A próba konkrét alakja:

$$F_{\mathrm{emp}} = \frac{RSS/k}{ESS/\left(n-k-1\right)} \overset{H_0}{\sim} F_{k,n-k-1}$$

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 nullmodelltől" 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

Lineáris kombináció tesztelése

- A séma:
$$\lambda_{\beta_1}\beta_1 + \lambda_{\beta_2}\beta_2 + \ldots + \lambda_{\beta_k}\beta_k = \Lambda$$
- Több koefficienst 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 λ 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 λ nulla és $\Lambda = 0$)
    - Az összes koefficiens összege épp nulla (ekkor mindegyik λ 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 $\lambda_{\beta_1}\widehat{\beta}_1 + \lambda_{\beta_2}\widehat{\beta}_2 + \ldots + \lambda_{\beta_k}\widehat{\beta}_k = \widehat{\Lambda}$, ekkor

$$\frac{\widehat{\Lambda} - \Lambda}{\mathrm{se}\left(\widehat{\Lambda}\right)} \overset{H_0}{\sim} t_{n-k-1}$$

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

Több megkötés egyidejű tesztelése

- Az eddigiek kombinálhatóak is: több megkötés (mindegyikük lineáris kombináció), melyeknek egyszerre kell teljesülniük
- Célszerű felírás:
$$H_0 : \mathbf{R}\beta = \mathbf{r},$$
ahol $\mathbf{R}$ $m \times k$ típusú (tehát m a megszorítások száma)
- Az erre adható teszt:
$$F_{\mathrm{emp}} = \frac{\left(\mathbf{R}\widehat{\beta} - \mathbf{r}\right)^T \left[\mathbf{R}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{R}^T\right]^{-1}\left(\mathbf{R}\widehat{\beta} - \mathbf{r}\right)/m}{\mathrm{ESS}/\left(n-k-1\right)} \overset{H_0}{\sim} \mathcal{F}\left(m, n-k-1\right)$$

Konkrét példák a fenti sémára

- Ellenőrizhető, hogy ha például...
  - ... $\mathbf{R} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 & 0 \ldots 0 \end{pmatrix}$ és $r = 0$, akkor a t-tesztet ...
  - ... $\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$ és $\mathbf{r} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ akkor az ANOVA-t...
  - ... $\mathbf{R} = \begin{pmatrix} \lambda_{\beta_1} & \lambda_{\beta_2} & \ldots & \lambda_{\beta_k} \end{pmatrix}$ és $r = \Lambda$, akkor a lineáris kombináció tesztelését...
- ... kapjuk vissza.