5. előadás - Regressziószámítás

2016. október 3.

Kétváltozós eset

A modell:

$$Y_i = \alpha + \beta X_i + u_i, \quad i = 1, \ldots, T,$$

ahol X_i független u_i-től minden i esetén, (u_i) pedig i.i.d. sorozat 0 várható értékkel és σ szórással. (OLS modell)
Paraméterbecslés: legkisebb négyzetes módszerrel történik, azaz a

$$V(a,b) = \sum_{i=1}^{T} e_i^2 = \sum_{i=1}^{T} (Y_i - a - bX_i)^2$$

veszteségfüggvényt kell minimalizálni az a, b paraméterek függvényében.

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \quad \text{és} \quad \hat{\beta} = \frac{\sum_{i=1}^{T} X_i Y_i - \bar{X}\bar{Y}}{\sum_{i=1}^{T} X_i^2 - T \cdot \bar{X}^2}.$$

σ^2 becslése: $s^2 = \frac{\sum_{i=1}^{T} e_i^2}{T-2}$.

Determinációs együttható

- TSS: $\sum_{i=1}^{T}(Y_i - \bar{Y})^2$ az átlagtól való teljes szóródás,
- RSS: $\sum_{i=1}^{T} e_i^2$ a hibák négyzetösszege, azaz a nem magyarázható szóródás
- ESS: $\hat{\beta} \cdot \sum_{i=1}^{T}(X_i - \bar{X})(Y_i - \bar{Y})$ a magyarázott szóródás.

Világos, hogy $TSS = ESS + RSS$, így definiálható az

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \in [0,1]$$

determinációs együttható, mely az illeszkedés jóságát méri. Könnyen látható, hogy $R =$Corr$(Y_i, \hat{Y})$.
Könnyen igazolható, hogy α, β és σ becslései torzítatlanok, és ezek a legjobb lináris torzítatlan becslései (BLUE) a paramétereknek.

Alapmodell

A modell:

$$Y_t = \beta_1 + \beta_2 X_{2,t} + \ldots + \beta_k X_{k,t} + u_t, \quad t = 1, \ldots, T,$$

ahol $T \geq k$,

- $X_2, \ldots, X_k$ $(k-1)$ darab független regresszor, avagy magyarázó változó, Y a függő, avagy magyarázott változó,
- u_t azonos eloszlású, korrelálatlan sorozat, azaz $\mathrm{E}u_t = 0$, $\mathrm{E}u_t^2 = \sigma^2$ és $\mathrm{E}u_t u_s = 0$, ha $t \neq s$. (OLS)

Legyen $y = (Y_1, \ldots, Y_T)^\intercal$, $\beta = (\beta_1, \ldots, \beta_k)^\intercal$, $u = (u_1, \ldots, u_T)^\intercal$,

$$X = \begin{pmatrix} 1 & X_{2,1} & \ldots & X_{k,1} \\ 1 & X_{2,2} & \ldots & X_{k,2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{2,T} & \ldots & X_{k,T} \end{pmatrix}.$$

Ekkor a modell az $y = X\beta + u$ kompakt formában is felírható.

Paraméterbecslés

Legyen $b \in \mathbb{R}^k$ egy tetszőleges futó paramétervektor. Ekkor a hibavektor

$$e = e(b) = y - Xb,$$

tehát a költségfüggvény, azaz a hibák négyzetösszege a

$$V(b) = e^\intercal e = (y - Xb)^\intercal (y - Xb) = y^\intercal y - 2b^\intercal X^\intercal y + b^\intercal X^\intercal X b$$

alakot ölti. Ennek a minimumát kellene megtalálni. Mivel

$$\begin{aligned} \frac{\partial V(b)}{\partial b} &= -2X^\intercal y + 2X^\intercal X b \\ \frac{\partial^2 V(b)}{\partial b^2} &= 2X^\intercal X \geq 0, \\ \hat{\beta} &= (X^\intercal X)^{-1} X^\intercal y, \end{aligned}$$

ha X teljes rangú, mert ekkor létezik az inverz és egyértelmű a minimum.

Többszörös determinációs együttható

Mivel

$$X^\intercal X\hat{\beta} = X^\intercal y = X^\intercal(X\hat{\beta} + e(\hat{\beta})),$$

így $X^\intercal e(\hat{\beta}) = 0$, azaz a hibavektor mindegyik regressziós vektorral korrelálatlan! Ezt felhasználva adódik, hogy a teljes négyzetösszeg (TSS) felírható az alábbi felbontásban:

$$\sum_{i=1}^{T}(Y_i - \bar{Y})^2 = y^\intercal y - T\bar{Y}^2 = (\hat{y} + e)^\intercal(\hat{y} + e) - T\bar{Y}^2 =$$

$$= \hat{y}^\intercal\hat{y} + 2\underbrace{e^\intercal\hat{y}}_{0} + e^\intercal e - T\bar{Y}^2 = \underbrace{(\hat{y}^\intercal\hat{y} - T\bar{Y}^2)}_{ESS} + \underbrace{e^\intercal e}_{RSS}$$

Így, hasonlóan a korábbi esethez, definiálható az

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \in [0,1]$$

többszörös determinációs együttható. Ha $\beta_1 \neq 0$, akkor $R =$Corr$(Y_i, \hat{Y})$.

Probléma

Kérdés: új változó felvétele a modellbe változtatja-e R^2 értékét?

- Az világos, hogy ekkor R^2 értéke csökkeni biztosan nem fog, hiszen "jobban" magyarázzuk Y-t, azaz egy nagyobb változókészleten minimalizáljuk a veszteségfüggvényt. Ezzel együtt viszont nő a modell bonyolultága, ami nem mindig jó!
- Tehát, ha R^2-tel jellemezzük a modellünket, akkor mindig az összes potenciális magyarázó változó felhasználása lesz a legjobb döntés.
- A valóságban azonban ez korántsem biztos!
- Fontos lesz számunkra a modell ún. általánosító-képessége, azaz hogy mennyire jól tud a mintán kívüli világról is számot adni.
- Erre a feladatra viszont az R^2 nem a legszerencsésebb mutató, hiszen ez a minta jó "megjegyzését" adja, de nekünk ennél több kell!

Egy lehetséges megoldás

Modellszelekció: olyan módosítás, mely figyelembe veszi a modell változóinak számát is, és meghatározható vele az optimális magyarázó-változók köre. Ennek egy lehetséges eszköze a **korrigált determinációs együttható:**

$$\bar{R}^2 = 1 - \frac{RSS(T-2)}{TSS(T-k)}$$

- Büntetjük a magyarázó változók számának növelését.
- Könnyen látható , hogy $\bar{R}^2 \leq R^2$, azaz 1-nél biztosan kisebb ez is, de vigyázat, lehet negatív is!
- A gyakorlatban heurisztikus stratégiákat használunk (forward, backward és stepwise szelekciós módszerek), hogy ne kelljen az összes 2^k kombinációt tesztelni.

Egy másik lehetséges megoldás

Módosított veszteségfüggvény használata (avagy információs kritériumok), melyek egyszerre büntetik a magyarázó változók nagy számát és a nagy hibát, a kettő közt egyensúlyt keresve:

$$V_{SC}(b) = \ln \frac{e(b)^\intercal e(b)}{T} + \frac{k}{T} \ln T$$

$$V_{AIC}(b) = \ln \frac{e(b)^\intercal e(b)}{T} + \frac{2k}{T}$$

Itt k optimális értéke is keresendő!

Modellszelekciós tesztek: a változók elhagyására vonatkozó Wald- és Lagrange Multiplikátor-teszt. (Ezeket most az idő rövidsége miatt nem tárgyaljuk részletesebben.)

Statisztikai tulajdonságok

Tétel

A β paraméter fenti $\hat{\beta}$ becslése torzítatlan becslés, továbbá $D^2(\hat{\beta}) = (X^\intercal X)^{-1}\sigma^2$.

Tétel

A zaj σ^2 szórásnégyzetének torzítatlan becslése $s^2 = \frac{e(\hat{\beta})^\intercal e(\hat{\beta})}{T-k}$.

Tétel (Gauss-Markov)

Legyen $c \in \mathbb{R}^k$ tetszőleges és $\mu = c^\intercal \beta$. Ennek legkisebb szórású, lineáris, torzítatlan (BLUE) becslése az y minta alapján $\hat{\mu} = c^\intercal \hat{\beta}$, ahol $\hat{\beta}$ a lineáris regressziós együtthatók fenti LS-becslése.

Parciális korrelációs együttható

Cél: két kvantitatív változó kapcsolatából ki akarjuk szűrni egy vagy több kvantitatív változó hatását.

Kérdés: milyen lenne a vizsgált két változó kapcsolata, ha a kiszűrt változókat állandó szinten tartanánk?

Válasz: parciális korrelációs együttható

$$\rho_{XY,Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1-\rho_{XZ}^2} \cdot \sqrt{1-\rho_{YZ}^2}}$$

Feltételek:

- kvantitatív (skálás) változóink vannak;
- csak lineáris összefüggés létezik köztük;
- X és Y között ugyanolyan jellegű és szintű kapcsolat van a Z változó teljes értéktartományában.

Parciális korrelációs együttható

- Ha a vizsgált változók együttes eloszlása többdimenziós normális, akkor ezek a feltételek szükségképpen fennállnak, azaz csak lineáris típusú összefüggések léphetnek fel, és a két változó közti összefüggés a harmadik változó bármely rögzített értéke esetén ugyanakkora lesz.
- Mérlegelés nélküli, automatikus használata esetén komoly bajok forrása lehet!

Példa: $k = 3$ eset

A háromváltozós modell becslése az alábbi alakban írható fel:

$$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2,t} + \hat{\beta}_3 X_{3,t} + e_t, \quad t = 1, \ldots, T.$$

Átlagolva a fenti egyenletet adódik, hogy

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3.$$

Kivonva egymásból a kettőt kapjuk, hogy

$$\underbrace{Y_t - \bar{Y}}_{=:\xi_t} = \hat{\beta}_2 \underbrace{(X_{2,t} - \bar{X}_2)}_{=:\eta_{2,t}} + \hat{\beta}_3 \underbrace{(X_{3,t} - \bar{X}_3)}_{=:\eta_{3,t}} + e_t, \quad t = 1, \ldots, T,$$

azaz a konstans tagot kiküszöböltük a modellből. Ekkor

$$\det(X^\intercal X) = (\sum_{i=1}^{T} \eta_{2,t}^2)(\sum_{i=1}^{T} \eta_{3,t}^2) - (\sum_{i=1}^{T} \eta_{2,t}\eta_{3,t})^2 =$$

$$= (\sum_{i=1}^{T} \eta_{2,t}^2)(\sum_{i=1}^{T} \eta_{3,t}^2)(1 - r_{23}^2), \quad \text{ahol } r_{23} = \mathrm{corr}(X_2, X_3).$$

Példa: $k = 3$ eset

Mivel $D^2(\hat{\beta}) = \sigma^2(X^\intercal X)^{-1}$, így adódik, hogy

$$D^2(\hat{\beta}_2) = \frac{\sigma^2 \sum_{i=1}^T \eta_{3,t}^2}{(\sum_{i=1}^T \eta_{2,t}^2)(\sum_{i=1}^T \eta_{3,t}^2)(1 - r_{23}^2)} = \frac{\sigma^2}{(\sum_{i=1}^T \eta_{2,t}^2)(1 - r_{23}^2)}$$

$$D^2(\hat{\beta}_3) = \frac{\sigma^2 \sum_{i=1}^T \eta_{2,t}^2}{(\sum_{i=1}^T \eta_{2,t}^2)(\sum_{i=1}^T \eta_{3,t}^2)(1 - r_{23}^2)} = \frac{\sigma^2}{(\sum_{i=1}^T \eta_{3,t}^2)(1 - r_{23}^2)}$$

Azaz ha $|r_{23}| \simeq 1$, azaz X_2 és X_3 között szoros a lineáris kapcsolat, akkor a paraméterek szórása nagy lesz. Ezt a jelenséget hívják multikollinearitásnak. Határesetben X oszlopai lineárisan összefüggők, így ilyenkor β nem is becsülhető. Tehát vigyázni kell az új változók felvételével, mert ekkor nő a multikollinearitás esélye, ami rontja a becslések hibáját!

Multikollinearitás

Az a jelenség, amikor a magyarázó változók lineáris kapcsolatban vannak egymással. Bár nem tökéletesen precíz, de a gyakorlatban azzal jellemezzük, hogy mennyire magyarázzák egymást. Az ennek megfelelő mérőszám az ún. tolerancia:

$$\text{Tol}_j = 1 - R_j^2 = 1 - R^2_{X_j|X_2,X_3,\ldots,X_{j-1},X_{j+1},\ldots,X_k},$$

azaz hogy a vizsgált magyarázó változót mennyire magyarázza a többi magyarázó változó.
Ekkor

$$D^2(\hat{\beta}_j) = \frac{RSS/(T-k)}{(T-1)D^2(X_j)} \cdot \frac{1}{\text{Tol}_j}$$

azaz, ha a tolerancia romlik (csökken), akkor a becsült paraméter szórása nő!

GLS-becslés

A modell most is $y = X\beta + u$ alakú, és tegyük fel, hogy

$$u \sim \mathcal{N}(\underline{0}, \sigma \cdot \Omega^{1/2}),$$

ahol

- $\Omega > 0$ ismert kovariancia mátrix,
- nem feltétlenül diagonális, és ha az, a diagonális elemek akkor sem feltétlenül egyenlők,
- azaz a modell heteroszkedasztikus, ami annyit jelent, hogy a zaj nem lesz azonos eloszlású folyamat, és a függetlenségi feltételt sem őrizzük meg minden esetben.

Mi történik ekkor az OLS becsléssel? A torzítatlanság és a konzisztencia nem romlik el, de már nem lesz hatásos a becslés, azaz nem ez lesz a legkisebb szórású becslése a paramétereknek. A becsült standard hibák is torzítottak lesznek, így a tesztek érvényüket vesztik!

GLS-becslés

Ω szimmetria tulajdonsága és pozitív definitsége miatt Ω^{-1} is szimmetrikus pozitív definit mátrix, így létezik olyan P nemszinguláris mátrix, melyre $\Omega^{-1} = P^\intercal P$. Szorozzuk végig ezzel a P mátrixszal balról a modell-egyenletet. Ekkor

$$Py = PX\beta + Pu,$$

és legyen $Py = \tilde{y}$, $PX = \tilde{X}$ és $Pu = \tilde{u}$. Könnyen látszik, hogy ekkor

$$\mathrm{E}(\tilde{u}\tilde{u}^\intercal) = P\mathrm{E}(uu^\intercal)P^\intercal = P(\sigma^2\Omega)P^\intercal = \sigma^2 P(P^\intercal P)^{-1}P^\intercal = \sigma^2 I,$$

tehát a transzformált modell már homoszkedasztikus, így működnek a korábbi becsléseink. Azaz

$$\hat{\beta} = (\tilde{X}^\intercal\tilde{X})^{-1}\tilde{X}^\intercal\tilde{y} = (X^\intercal P^\intercal P X)^{-1}X^\intercal P^\intercal P y =$$

$$= (X^\intercal\Omega^{-1}X)^{-1}X^\intercal\Omega^{-1}y.$$

GLS-becslés

Ha Ω nem ismert, akkor az esetek nagy részében a becslése reménytelen. Arra van módszer, hogy diagonális, de nem homoszkedasztikus esetben az $\sigma^2(X^\intercal \Omega X)$ mátrixot becsüljük, ekkor ugyanis

$$\sigma^2(X^\intercal \Omega X) = \sum_{i=1}^{T} \sigma_i^2 X_i X_i^\intercal,$$

ahol σ_i ismeretlen ugyan, de becsülhető az

$$e_i = (y - X_i^\intercal \hat{\beta})^2$$

kifejezéssel, azaz $\sigma^2 \hat{\Omega} =$diag(e_i^2).