O LINEARNI REGRESIJI

Martin Raič

September 2014

1 Linearna regresija kot statistični model

Linearna regresija je statistični model, pri katerem opazimo slučajni vektor:

$$\boldsymbol{Y} = \boldsymbol{v} + \boldsymbol{\varepsilon}\,, \tag{1.1}$$

kjer je $\boldsymbol{v}$ neopazljiv deterministično parameter, $\boldsymbol{\varepsilon}$ pa neopazljiv slučajen šum. Pri tem vse skupaj živi v n-razsežnem evklidskem prostoru W. Za $\boldsymbol{v}$ privzamemo, da pripada določenemu k-razsežnemu podprostoru V, za šum $\boldsymbol{\varepsilon}$ pa privzamemo, da ima matematično upanje $\mathbf{0}$ in kovariančno matriko $\sigma^2\mathbf{I}_W$, kjer je σ^2 še en neopazljiv deterministično parameter; le-temu bomo rekli *varianca modela*. Velja torej $\mathbb{E}\,\boldsymbol{\varepsilon} = \mathbf{0}$ in $\mathbb{E}\,\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}} = \mathbf{I}_W$.

V praksi, ko podatke podajamo s številkami, je smiselno privzeti, da je $W = \mathbb{R}^n$, V pa je zaloga vrednosti injektivne linearne preslikave $\mathbf{X}\colon\ \mathbb{R}^k \to \mathbb{R}^n$, ki jo bomo identificirali z matriko velikosti $n \times k$. Tako lahko model parametriziramo v obliki:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{1.2}$$

kjer je $\boldsymbol{\beta} \in \mathbb{R}^k$. Ekvivalentno lahko zapišemo tudi:

$$\begin{aligned} Y_1 &= \beta_1 f_1(X_1) + \beta_2 f_2(X_1) + \cdots + \beta_k f_k(X_1) + \varepsilon_1\,,\\ Y_2 &= \beta_1 f_1(X_2) + \beta_2 f_2(X_2) + \cdots + \beta_k f_k(X_2) + \varepsilon_2\,,\\ &\vdots\\ Y_n &= \beta_1 f_1(X_n) + \beta_2 f_2(X_n) + \cdots + \beta_k f_k(X_n) + \varepsilon_n\,, \end{aligned} \tag{1.3}$$

kjer so $X_1, \ldots X_n$ deterministične in opazljive spremenljivke. Tem pravimo *pojasnjevalne*, medtem ko so $Y_1, \ldots Y_n$ *odvisne* spremenljivke. Tu je seveda:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} f_1(X_1) & f_2(X_1) & \cdots & f_k(X_1) \\ f_1(X_2) & f_2(X_2) & \cdots & f_k(X_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(X_1) & f_2(X_1) & \cdots & f_k(X_1) \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Obravnavali bomo naslednje naloge inferenčne statistike:

- točkasto in intervalsko ocenjevanje parametrov;
- napovedovanje spremenljivk, povezanih z modelom.
- testiranje hipotez v zvezi z modelom.

2 Točkasto ocenjevanje in napovedovanje

Oglejmo si najprej model, podan z zvezo (1.1). Iskali bomo čim boljšo *nepristransko* cenilko $\hat{\boldsymbol{v}} = h(\boldsymbol{Y})$ parametra $\boldsymbol{v}$, t. j. tako, ki ima najmanjšo možno kovariančno matriko $\mathrm{var}(\hat{\boldsymbol{v}}) := \mathbb{E}\big[(\hat{\boldsymbol{v}} - \boldsymbol{v})(\hat{\boldsymbol{v}} - \boldsymbol{v})^{\mathrm{T}}\big]$. Seveda je to simetrična matrika, na teh pa bomo gledali delno urejenost, kjer je $\mathbf{A} \leq \mathbf{B}$, če je $\mathbf{B} - \mathbf{A}$ pozitivno semidefinitna, t. j. $\boldsymbol{a}^{\mathrm{T}}\mathbf{A}\boldsymbol{a} \leq \boldsymbol{a}^{\mathrm{T}}\mathbf{B}\boldsymbol{a}$ za vse $\boldsymbol{a} \in V$. Najprej se bomo omejili na *linearne* cenilke, t. j. take, kjer je $h(\boldsymbol{y}) = \mathbf{A}\boldsymbol{y} + \boldsymbol{v}_0$.

Izrek 2.1 (Gauss, Markov). *Naj velja* (1.1) *skupaj s pripadajočimi predpostavkami. Tedaj je nepristranska linearna cenilka za $\boldsymbol{v}$ z najmanjšo možno kovariančno matriko kar pravokotna projekcija opaženega vektorja $\boldsymbol{Y}$ na podprostor V.*

Preden gremo dokazat izrek, uvedimo nekaj notacije za inkluzijo in pravokotno projekcijo. Naj bosta $G \subseteq H$ evklidska prostora. Z $\iota_{G\to H}$ bomo označevali inkluzijo iz G v H. Pri pravokotni projekciji pa bomo ločili dve niansi: označevali jo bomo tako s $p_{H\to G}$ kot tudi s $\mathbf{P}_{H\to G}$, toda prvo bomo gledali kot preslikavo iz H na G, drugo pa kot preslikavo iz H v H, torej kot operator na H. Velja:

$$p_{H\to G}^{\mathrm{T}} = \iota_{G\to H}\,, \qquad \mathbf{P}_{H\to G}^{\mathrm{T}} = \mathbf{P}_{H\to G}\,.$$

Pri zapisu cenilke lahko seveda uporabimo obe različici:

$$\hat{\boldsymbol{v}} = p_{W\to V}\boldsymbol{Y} = \mathbf{P}_{W\to V}\boldsymbol{Y}\,. \tag{2.1}$$

Dokaz izreka Gaussa in Markova. Najprej preverimo, kdaj je linearna cenilka $\hat{\boldsymbol{v}} = \mathbf{A}\boldsymbol{Y} + \boldsymbol{v}_0$ nepristranska. Ker je $\mathbb{E}\,\varepsilon = \mathbf{0}$, mora veljati:

$$\mathbb{E}\,\hat{\boldsymbol{v}} = \mathbf{A}\boldsymbol{v} + \boldsymbol{v}_0 = \boldsymbol{v}\,,$$

in sicer ne glede na dejansko vrednost neznanega parametra $\boldsymbol{v}$. Torej mora biti $\boldsymbol{v}_0 = \mathbf{0}$, $\mathbf{A}$ pa mora mirovati na V. Varianca cenilke $\hat{\boldsymbol{v}}$ bo najmanjša natanko tedaj, ko bo izraz:

$$\boldsymbol{a}^{\mathrm{T}}\operatorname{var}(\hat{\boldsymbol{v}})\boldsymbol{a} = \sigma^2\boldsymbol{a}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\boldsymbol{a} = \sigma^2\big\|\mathbf{A}^{\mathrm{T}}\boldsymbol{a}\big\|^2$$

minimalen za vse $\boldsymbol{a} \in V$. Ker $\mathbf{A}$ miruje na V, za vse $\boldsymbol{a}, \boldsymbol{b} \in V$ velja:

$$\boldsymbol{b}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{a} = (\mathbf{A}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{a} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{a}\,,$$

kar pomeni, da mora $\mathbf{A}^{\mathrm{T}} - \iota_{V\to W}$ slikati v ortogonalni komplement prostora V. Z drugimi besedami, lahko pišemo $\mathbf{A}^{\mathrm{T}} = \iota_{V\to W} + \mathbf{B}$, kjer $\mathbf{B}$ slika iz V v njegov ortogonalni komplement. Sedaj pa lahko izračunamo:

$$\big\|\mathbf{A}^{\mathrm{T}}\boldsymbol{a}\big\|^2 = \|\boldsymbol{a} + \mathbf{B}\boldsymbol{a}\|^2 = \|\boldsymbol{a}\|^2 + \big\|\mathbf{B}\boldsymbol{a}\big\|^2$$

(upoštevali smo ortogonalnost). Slednji izraz pa je minimalen (za vse $\boldsymbol{a}$), če je $\mathbf{B} = \mathbf{0}$, torej če je $\mathbf{A}^{\mathrm{T}} = \iota_{V\to W}$, torej če je $\mathbf{A} = p_{W\to W}$. ∎

Opomba 2.2. Iz dokaza izreka Gaussa in Markova lahko razberemo, da je, če je $\hat{\boldsymbol{v}}$ tako kot v (2.1), kovariančna matrika te cenilke enaka:

$$\operatorname{var}\big(\hat{\boldsymbol{v}}\big) = \sigma^2\mathbf{I}_V\,. \tag{2.2}$$

Izrek Gaussa in Markova nam da tudi pripadajoči rezultat za linearne funkcionale parametra $\boldsymbol{v}$: za poljuben $\boldsymbol{a} \in V$ tudi $\mathbf{a}^{\mathrm{T}}\hat{\boldsymbol{v}}$ nepristranska cenilka karakteristike $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}$ z najmanjšo možno varianco in slednja je enaka:

$$\operatorname{var}\big(\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{v}}\big) = \sigma^2\|\boldsymbol{a}\|^2\,. \tag{2.3}$$

Privzemimo zdaj, da je model *Gaussov*, t. j. da je šum porazdeljen *večrazsežno normalno*, se pravi zvezno z gostoto:

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{w} \mid \boldsymbol{v}) = e^{-\|\boldsymbol{w}\|^2/2}\,.$$

To pomeni, da je opažanje porazdeljeno zvezno z gostoto:

$$f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{v}) = e^{-\|\boldsymbol{y}-\boldsymbol{v}\|^2/2}\,.$$

Pri fiksnem $\boldsymbol{y}$ je maksimum po $\boldsymbol{v} \in V$ dosežen natanko tedaj, ko je $\boldsymbol{v}$ pravokotna projekcija vektorja $\boldsymbol{y}$ na V. To pomeni, da se cenilka $\hat{\boldsymbol{v}} = p_{W\to V}\boldsymbol{Y}$ ujema s cenilko po metodi največjega verjetja.

Izrek 2.3. *Če je model Gaussov, ima cenilka $\hat{\boldsymbol{v}} = p_{W\to V}\boldsymbol{Y}$ najmanjšo možno kovariančno matriko izmed* vseh *možnih nepristranskih cenilk.*

Dokaz. Sklicali se bomo na *Lehmann–Schefféjev izrek* (glej npr. [1], izrek 5.5, str. 298). Najprej iz zapisa:

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{v}) &= \exp\left(-\frac{\|\boldsymbol{v}\|^2}{2}\right)\exp\left(-\frac{\|\boldsymbol{y}\|^2}{2}\right)\exp\bigl(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{y}\bigr) = \\ &= \exp\left(-\frac{\|\boldsymbol{v}\|^2}{2}\right)\exp\left(-\frac{\|\boldsymbol{y}\|^2}{2}\right)\exp\bigl(\boldsymbol{v}^{\mathrm{T}}p_{W\to V}\boldsymbol{y}\bigr) \end{aligned}$$

razberemo, da gre za neizrojeno eksponentno družino s pripadajočo zadostno statistiko $p_{W\to V}\boldsymbol{Y}$, ki je posledično kompletna (glej npr. [1], izrek 2.74, str. 108). Naj bo $\boldsymbol{a} \in V$. Ker je cenilka $\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{v}}$ karakteristike $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}$ funkcija pripadajoče zadostne statistike in nepristranska, ima po Lehmann–Schefféjevem izreku najmanjšo možno varianco: za vsako drugo nepristransko cenilko $h_{\boldsymbol{a}}(\boldsymbol{Y})$ velja $\operatorname{var}\bigl(h_{\boldsymbol{a}}(\boldsymbol{Y})\bigr) \geq \sigma^2\|\boldsymbol{a}\|^2$. Brž ko je torej $h(\boldsymbol{Y})$ kakšna druga nepristranska cenilka za $\boldsymbol{v}$, je tudi $\boldsymbol{a}^{\mathrm{T}}h(\boldsymbol{Y})$ nepristranska cenilka za $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}$ in po prejšnjem velja:

$$a^{\mathrm{T}}\operatorname{var}\bigl(h(\boldsymbol{Y})\bigr)\boldsymbol{a} = \operatorname{var}\bigl(\boldsymbol{a}^{\mathrm{T}}h(\boldsymbol{Y})\bigr) \geq \sigma^2\|\boldsymbol{a}\|^2 = a^{\mathrm{T}}\operatorname{var}\bigl(\hat{\boldsymbol{v}}\bigr)\boldsymbol{a}\,,$$

torej je $\operatorname{var}\bigl(h(\boldsymbol{Y})\bigr) \geq \operatorname{var}\bigl(\hat{\boldsymbol{v}}\bigr)$. ∎

Razliki $\hat{\boldsymbol{\varepsilon}} := \boldsymbol{Y} - \hat{\boldsymbol{v}}$ pravimo *rezidual* in je ocena za šum. Rezidual je pravokotna projekcija vektorja $\boldsymbol{Y}$ na ortogonalni komplement prostora V v W. Če torej za evklidska prostora $G \subseteq H$ s $q_{H\to G}$ oziroma s $\mathbf{Q}_{H\to G}$ označimo ortogonalni projektor s prostora G na ortogonalni komplement prostora G v H (pri čemer je $q_{H\to G}$ definirana kot preslikava na ustrezen ortogonalni komplement, $\mathbf{Q}_{H\to G}$ pa kot operator na H, tako da velja $\mathbf{Q}_{H\to G} = \mathbf{I}_H - \mathbf{P}_{H\to G}$ in tudi $\mathbf{P}_{H\to G}\mathbf{Q}_{H\to G} = \mathbf{Q}_{H\to G}\mathbf{P}_{H\to G} = \mathbf{0}$), velja:

$$\hat{\boldsymbol{\varepsilon}} = q_{W\to V}\boldsymbol{Y} = \mathbf{Q}_{W\to V}\boldsymbol{Y}$$

in ker je $q_{W\to V}\boldsymbol{v} = \mathbf{0}$ in $\mathbf{Q}_{W\to V}\boldsymbol{v} = \mathbf{0}$, je tudi:

$$\hat{\boldsymbol{\varepsilon}} = q_{W\to V}\boldsymbol{\varepsilon} = \mathbf{Q}_{W\to V}\boldsymbol{\varepsilon}\,.$$

Pravokotna projekcija vektorja na določen podprostor je vektor v tem podprostoru, ki je izvirnemu vektorju najbližje glede na evklidsko normo. V kontekstu našega modela (1.1) to pomeni, da je norma reziduala najmanjša možna, torej da je vsota kvadratov komponent reziduala minimalna. Zato pravimo, da je cenilka $\hat{\boldsymbol{v}}$ za $\boldsymbol{v}$ dobljena po *metodi najmanjših kvadratov*.

Rezidual $\hat{\boldsymbol{\varepsilon}}$ je nekoreliran z $\hat{\boldsymbol{v}}$, saj je:

$$\begin{aligned}\operatorname{cov}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{\varepsilon}}) &= \operatorname{cov}\big(\mathbf{P}_{W\to V}\boldsymbol{\varepsilon}, \mathbf{Q}_{W\to V}\boldsymbol{\varepsilon}\big) = \\ &= \mathbf{P}_{W\to V}\operatorname{var}(\boldsymbol{\varepsilon})\mathbf{Q}_{W\to V}^{\mathrm{T}} = \\ &= \sigma^2\mathbf{P}_{W\to V}\mathbf{I}_W\mathbf{Q}_{W\to V} = \\ &= \sigma^2\mathbf{P}_{W\to V}\mathbf{Q}_{W\to V} = \\ &= \mathbf{0}\,.\end{aligned}$$

Izračunamo lahko še:

$$\begin{aligned}\mathbb{E}\,\|\hat{\boldsymbol{\varepsilon}}\|^2 &= \operatorname{sl}\mathbb{E}\big[\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\big] = \\ &= \operatorname{sl}\mathbb{E}\big[\mathbf{Q}_{W\to V}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\mathbf{Q}_{W\to V}^{\mathrm{T}}\big] = \\ &= \operatorname{sl}\big(\mathbf{Q}_{W\to V}\,\mathbb{E}\big[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\big]\mathbf{Q}_{W\to V}\big) = \\ &= \sigma^2\operatorname{sl}\big(\mathbf{Q}_{W\to V}\mathbf{I}_W\mathbf{Q}_{W\to V}\big) = \\ &= \sigma^2\operatorname{sl}\mathbf{Q}_{W\to V} = \\ &= (n-k)\sigma^2\,.\end{aligned}$$

Torej je:

$$\hat{\sigma}^2 = \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{n-k} \tag{2.4}$$

nepristranska cenilka za σ^2. Količini $\hat{\sigma}^2$ bomo zato rekli kar *ocenjena varianca modela*.

Zdaj si lahko ogledamo, kolikšno napako v grobem naredimo pri ocenjevanju linearnih funkcionalov parametrov. Če ocenjujemo določen linearni funkcional parametra $\boldsymbol{v}$ iz zveze (1.1), t. j. $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}$, iz opombe 2.2 vemo, da je $\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{v}}$ nepristranska cenilka z varianco $\sigma^2\|\boldsymbol{a}\|^2$. Grobo merilo za napako, ki smo jo naredili pri ocenjevanju, je *ocenjena standardna napaka*:

$$\operatorname{SE}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}) = \hat{\sigma}\|\boldsymbol{a}\|\,. \tag{2.5}$$

Oznaka SE tu pomeni operator, ki je definiran na linearnih funkcionalih na parametričnem prostoru V ob predpostavki, da je definirano tudi opažanje $\boldsymbol{Y}$. Pri tem dani linearni funkcional preslika v kvadratni koren nepristranske cenilke variance taistega funkcionala, uporabljenega na projekciji opažanja $\boldsymbol{Y}$ na V. Pri tem cenilka variance izhaja iz (2.4), kar je vedno možno, saj je varianca funkcionala znan večkratnik variance modela. Pri pisavi $\operatorname{SE}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v})$ je nekaj zlorabe notacije: vektor $\boldsymbol{v}$ smo identificirali z identiteto na V. Tovrstne zlorabe notacije so v matematiki zelo pogoste in se jih navadno niti ne zavedamo: tudi pri diferencialih pišemo spremenljivko, v resnici pa jo uporabimo kot preslikavo.

Če pa je model podan z zvezo (1.2), lahko to prevedemo na (1.1), tako da postavimo $\boldsymbol{v} = \mathbf{X}\boldsymbol{\beta}$. Tedaj se pravokotni projektor izraža v obliki $\mathbf{P}_{W\to V} = \mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}$. Pišemo lahko $\boldsymbol{\beta} = \mathbf{X}^{-1}\boldsymbol{v}$, kjer je inverz $\mathbf{X}^{-1}$ definiran kot preslikava iz V na $\mathbb{R}^k$ (medtem ko je $\mathbf{X}$ preslikava iz $\mathbb{R}^k$ v $\mathbb{R}^n = W$). Velja:

$$\mathbf{X}^{-1}p_{W\to V} = \big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\,. \tag{2.6}$$

Od tod dobimo še cenilko za $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^{-1}\hat{\boldsymbol{v}} = \mathbf{X}^{-1}p_{W\to V}\hat{\boldsymbol{v}} = \mathbf{X}^{-1}p_{W\to V}\mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} = \big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}\,.$$

Tudi $\hat{\boldsymbol{\beta}}$ je nepristranska cenilka za $\boldsymbol{\beta}$ z najmanjšo možno kovariančno matriko: to sledi iz ustrezne lastnosti cenilke $\hat{\boldsymbol{v}}$ za $\boldsymbol{v}$. Ta kovariančna matrika je enaka:

$$\operatorname{var}(\hat{\boldsymbol{\beta}}) = \sigma^2\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{I}_W\Big(\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\Big)^{\mathrm{T}} = \sigma^2\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\,. \tag{2.7}$$

Spet lahko izpeljemo tudi pripadajoče rezultate za linearne funkcionale oblike $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$, kjer je $\boldsymbol{c} \in \mathbb{R}^k$. S pomočjo zveze (2.7) dobimo varianco:

$$\operatorname{var}\big(\boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}\big) = \boldsymbol{c}^{\mathrm{T}}\operatorname{var}(\hat{\boldsymbol{\beta}})\boldsymbol{c} = \sigma^2\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c} \tag{2.8}$$

To pa lahko izpeljemo tudi tako, da zvezo (1.2) prevedemo na (1.1), tako da postavimo $\boldsymbol{v} := \mathbf{X}\boldsymbol{\beta}$. V skladu z zvezo (2.6) potem dobimo:

$$\boldsymbol{c}^{\mathrm{T}}\beta = \boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{v}\,,$$

torej lahko postavimo $\boldsymbol{a} := \mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c}$ in po (2.3) velja:

$$\operatorname{var}\big(\boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}\big) = \sigma^2\big\|\mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c}\big\|^2$$

in ni se težko prepričati, da je to isto kot v (2.8). Iz variance dobimo ocenjeno standardno napako:

$$\operatorname{SE}\big(\boldsymbol{c}^{\mathrm{T}}\beta\big) = \hat{\sigma}\sqrt{\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c}} = \hat{\sigma}\big\|\mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c}\big\|\,.$$

Ocenjena varianca modela $\hat{\sigma}^2$, definirana v (2.4), se izraža z rezidualom $\hat{\boldsymbol{\varepsilon}}$. Iz prevedbe zapisa (1.2) na zapis (1.1) dobimo, da ima rezidual zdaj obliko:

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\,.$$

Z drugimi besedami, rezidual je razlika med *opaženimi vrednostmi* $\boldsymbol{Y}$ in *teoretičnimi vrednostmi* $\hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

Videli smo, da je, če je model podan z zvezo (1.1), rezidual nekoreliran z $\hat{\boldsymbol{v}}$. No, če je podan z zvezo (1.2), pa je nekoreliran z $\hat{\boldsymbol{\beta}}$, saj je:

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\varepsilon}}) = \operatorname{cov}\big(\mathbf{X}^{-1}p_{W\to V}\hat{\boldsymbol{v}},\ \hat{\boldsymbol{\varepsilon}}\big) = \mathbf{X}^{-1}p_{W\to V}\operatorname{cov}(\hat{\boldsymbol{v}}, \hat{\boldsymbol{\varepsilon}}) = \mathbf{0}\,.$$

Včasih pa nas zanimajo tudi količine, ki nastanejo kot vsota določenega funkcionala na parametričnem prostoru in dodatnega šuma. Aproksimaciji teh količin z opaženimi količinami pravimo *napovedovanje*. Če privzamemo, da je pričakovana vrednost dodatnega šuma enaka nič, lahko dobimo nepristransko napoved v smislu, da je $\mathbb{E}(\hat{S}) = \mathbb{E}(S)$, kjer je S količina, ki jo napovedujemo, $\hat{S}$ pa njena napoved. Če želimo zdaj oceniti napako, ki jo naredimo, potrebujemo *srednjo kvadratično napako* $\mathbb{E}\big((\hat{S} - S)^2\big)$. Kvadratni koren ocene srednje kvadratične napake bo groba ocena za napako, ki jo bomo naredili. Temu bomo prav tako rekli ocenjena standardna napaka.

Ocenjeno standardno napako bo možno določiti pod pogojem, da je dodatni šum nekoreliran s šumom, ki se nanaša na opažanja, t. j. $\boldsymbol{\varepsilon}$, ter da je varianca dodatnega šuma znan večkratnik variance modela. Tako bomo zapisali:

$$S = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{v} + \lambda\eta \qquad \text{oziroma} \qquad S = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta} + \lambda\eta \tag{2.9}$$

kjer je slučajna spremenljivka η nekorelirana z $\boldsymbol{\varepsilon}$ ter velja $\mathbb{E}\,\eta = 0$ in $\mathrm{var}(\eta) = \sigma^2$, λ pa je znan koeficient. Nepristranska napoved za S bo seveda:

$$\hat{S} = \boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{v}} \qquad \text{oziroma} \qquad \hat{S} = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}\,, \tag{2.10}$$

srednja kvadratična napaka pa bo:

$$\mathbb{E}\big((\hat{S} - S)^2\big) = \mathrm{var}(\hat{S} - S) = \mathrm{var}(\boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{v}}) + \lambda^2\sigma^2 = \sigma^2\big(\|\boldsymbol{a}\|^2 + \lambda^2\big)$$

oziroma:

$$\mathbb{E}\big((\hat{S} - S)^2\big) = \mathrm{var}(\hat{S} - S) = \mathrm{var}(\boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}) + \lambda^2\sigma^2 = \sigma^2\Big(\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c} + \lambda^2\Big)\,.$$

Tako bo ocenjena standardna napaka:

$$\mathrm{SE}(S) := \frac{\hat{\sigma}}{\sigma}\sqrt{\mathbb{E}\big((\hat{S} - S)^2\big)} \tag{2.11}$$

vedno opazljiva, in sicer enaka:

$$\mathrm{SE}(S) = \hat{\sigma}\sqrt{\|\boldsymbol{a}\|^2 + \lambda^2} \tag{2.12}$$

oziroma:

$$\mathrm{SE}(S) = \hat{\sigma}\sqrt{\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c} + \lambda^2} = \hat{\sigma}\sqrt{\big\|\mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c}\big\|^2 + \lambda^2}\,. \tag{2.13}$$

3 Intervalsko in območno ocenjevanje in napovedovanje

Pri intervalskem ocenjevanju se bomo omejili na *linearne* funkcionalov vektorjev $\boldsymbol{v}$ oziroma $\boldsymbol{\beta}$, ki jim bo lahko prištet še neodvisen *šum*. V slednjem primeru bomo temu rekli *napovedovanje*. Poleg tega se bomo ukvarjali tudi z linearnimi preslikavami vektorjev $\boldsymbol{v}$ in $\boldsymbol{\beta}$

(v večrazsežne prostore), spet lahko s prištetim šumom. Tovrstne količine bomo ocenjevali oz. napovedovali *območno*.

Pri intervalskem oz. območnem ocenjevanju oz. napovedovanju moramo seveda vedeti nekaj več o porazdelitvah. Privzeli bomo *Gaussov model*, kar pomeni, da je šum porazdeljen *večrazsežno normalno*.

Potrebovali bomo naslednja dejstva o večrazežni normalni porazdelitvi in njenih izpeljankah:

- Če sta $V \subseteq W$ evklidska prostora in je slučajni vektor $\boldsymbol{Z}$ porazdeljen standardno normalno na W, je $p_{W\to V}\boldsymbol{Z}$ porazdeljen standardno normalno na V.
- Če sta slučajna vektorja $\boldsymbol{Z}_1$ in $\boldsymbol{Z}_2$ nekorelirana in je njuna skupna porazdelitev večrazsežna normalna, sta ta dva slučajna vektorja tudi neodvisna.
- Če je $\boldsymbol{Z}$ porazdeljen standardno normalno na evklidskem prostoru V, ima slučajna spremenljivka $\|\boldsymbol{Z}\|^2$ porazdelitev hi kvadrat s toliko prostostnimi stopnjami, kolikor je dimenzija prostora V; pišemo $\|\boldsymbol{Z}\|^2 \sim \chi^2(\dim V)$.
- Če je slučajna spremenljivka Z porazdeljena standardno enorazsežno normalno in je $U \sim \chi^2(n)$ od nje neodvisna slučajna spremenljivka, ima slučajna spremenljivka $\dfrac{Z}{\sqrt{U}}\sqrt{n}$ Studentovo porazdelitev z n prostostnimi stopnjami; pišemo $\dfrac{Z}{\sqrt{U}}\sqrt{n} \sim \mathrm{Student}(n)$.
- Če sta $U \sim \chi^2(m)$ in $V \sim \chi^2(n)$ neodvisni slučajni spremenljivki, ima $\dfrac{U/m}{V/n}$ Fisherjevo porazdelitev z (m,n) prostostnimi stopnjami; pišemo $\dfrac{U/m}{V/n} \sim \mathrm{Fisher}(m,n)$.

Oglejmo si zdaj, kaj pomenijo ta dejstva pri predpostavkah iz prejšnjega razdelka. Najprej se spomnimo, da sta slučajna vektorja $\hat{\boldsymbol{v}}$ in $\hat{\boldsymbol{\varepsilon}}$ pri predpostavkah, navedenih pod zvezo (1.1), nekorelirana. Če je model Gaussov, sta torej neodvisna. Podobno sta, če privzamemo zvezo (1.2) in vse potrebne predpostavke, neodvisna slučajna vektorja $\hat{\boldsymbol{\beta}}$ in $\hat{\boldsymbol{\varepsilon}}$.

V prejšnjem razdelku smo izračunali matematično upanje slučajne spremenljivke $\|\hat{\boldsymbol{\varepsilon}}\|^2$. Če je model Gaussov, poznamo tudi njeno porazdelitev. Slučajni vektor $\boldsymbol{\varepsilon}/\sigma$ je namreč porazdeljen standardno normalno. Slučajni vektor $\hat{\boldsymbol{\varepsilon}}/\sigma = q_{W\to V}\boldsymbol{\varepsilon}/\sigma$ pa je pravokotna projekcija tega vektorja na ortogonalni komplement prostora V v W, torej je prav tako porazdeljen standardno normalno na tem ortogonalnem komplementu, ki je $(n-k)$-razsežen prostor. To pa pomeni, da je:

$$\frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{\sigma^2} \sim \chi^2(n-k)\,. \tag{3.1}$$

Lotimo se sedaj iskanja napovednega intervala za količino S, definirano tako kot v (2.9), t. j. $S = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{v} + \lambda\eta$, če je model podan z zvezo (1.1), oziroma $S = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta} + \lambda\eta$, če je model podan z zvezo (1.2). Tu je slučajna spremenljivka η porazdeljena normalno $\mathrm{N}(0,\sigma^2)$ in

neodvisna od $\boldsymbol{\varepsilon}$, λ pa je znan koeficient. Naj bo $\hat{S}$ nepristranska napoved za S, definirana v (2.10). Slučajna spremenljivka:

$$\frac{\hat{S}-S}{\sqrt{\mathbb{E}\big((\hat{S}-S)^2\big)}}$$

je porazdeljena standardno normalno in je neodvisna od $\boldsymbol{\varepsilon}$. Zato je:

$$\frac{\dfrac{\hat{S}-S}{\sqrt{\mathbb{E}\big((\hat{S}-S)^2\big)}}}{\dfrac{\|\boldsymbol{\varepsilon}\|}{\sigma}}\sqrt{n-k}=\frac{\sigma}{\hat{\sigma}}\,\frac{\hat{S}-S}{\sqrt{\mathbb{E}\big((\hat{S}-S)^2\big)}}=\frac{\hat{S}-S}{\mathrm{SE}(S)}\sim\mathrm{Student}(n-k) \tag{3.2}$$

(uporabili smo še (2.4) in (2.11)). Če torej s $t_p(m)$ označimo kvantil Studentove porazdelitve z m prostostnimi stopnjami za verjetnost p, je napovedni interval za S pri stopnji zaupanja β možno zapisati v obliki:

$$\hat{S}-\mathrm{SE}(S)\,t_{(1+\beta)/2}(n-k)<S<\hat{S}+\mathrm{SE}(S)\,t_{(1+\beta)/2}(n-k)\,.$$

Oglejmo si še primer, ko napovedujemo slučajni *vektor* $\boldsymbol{S}:=\mathbf{A}\boldsymbol{v}+\boldsymbol{\Lambda\eta}$ z vrednostmi v vektorskem prostoru U, kjer je dodatni šum $\boldsymbol{\eta}\sim\mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}_U)$ neodvisen od $\boldsymbol{\varepsilon}$. V tem primeru lahko določimo *napovedno območje* – množico v U. Velja:

$$\mathbf{A}\hat{\boldsymbol{v}}-\boldsymbol{S}=\mathbf{A}(\hat{\boldsymbol{v}}-\boldsymbol{v})-t\boldsymbol{\eta}\sim\mathrm{N}\big(\mathbf{0},\sigma^2(\mathbf{AA}^{\mathrm{T}}+\boldsymbol{\Lambda\Lambda}^{\mathrm{T}})\big)\,,$$

torej je:

$$\frac{1}{\sigma^2}\Big\|\big(\mathbf{AA}^{\mathrm{T}}+\boldsymbol{\Lambda\Lambda}^{\mathrm{T}}\big)^{-1/2}(\mathbf{A}\hat{\boldsymbol{v}}-\boldsymbol{S})\Big\|^2=\frac{1}{\sigma^2}(\mathbf{A}\hat{\boldsymbol{v}}-\boldsymbol{S})^{\mathrm{T}}\big(\mathbf{AA}^{\mathrm{T}}+\boldsymbol{\Lambda\Lambda}^{\mathrm{T}}\big)^{-1}(\mathbf{A}\hat{\boldsymbol{v}}-\boldsymbol{S})\sim\chi^2(m)\,.$$

kjer je $m=\dim U$. Če se spomnimo še na (3.1) in dejstvo, da sta slučajna vektorja $\hat{\boldsymbol{v}}$ in $\hat{\boldsymbol{\varepsilon}}$ neodvisna, dobimo:

$$\frac{n-k}{m}\,\frac{(\mathbf{A}\hat{\boldsymbol{v}}-\boldsymbol{S})^{\mathrm{T}}\big(\mathbf{AA}^{\mathrm{T}}+\boldsymbol{\Lambda\Lambda}^{\mathrm{T}}\big)^{-1}(\mathbf{A}\hat{\boldsymbol{v}}-\boldsymbol{S})}{\|\hat{\boldsymbol{\varepsilon}}\|^2}\sim\mathrm{Fisher}(m,n-k)\,. \tag{3.3}$$

Če torej z $F_p(m,n)$ označimo kvantil Fisherjeve porazdelitve z (m,n) prostostnimi stopnjami za verjetnost p, je iskano napovedno območje pri stopnji zaupanja β določeno s pogojem:

$$(\boldsymbol{S}-\mathbf{A}\hat{\boldsymbol{v}})^{\mathrm{T}}\big(\mathbf{AA}^{\mathrm{T}}+\boldsymbol{\Lambda\Lambda}^{\mathrm{T}}\big)^{-1}(\boldsymbol{S}-\mathbf{A}\hat{\boldsymbol{v}})<\frac{m}{n-k}\,F_\beta(m,n-k)\,\|\hat{\boldsymbol{\varepsilon}}\|^2\,.$$

Če pa je model podan z zvezo (1.2), lahko poiščemo napovedno območje za slučajni vektor $\mathbf{S}:=\mathbf{C}\boldsymbol{\beta}+\lambda\boldsymbol{\eta}$. V tem primeru spet postavimo $\boldsymbol{v}=\mathbf{X}\boldsymbol{\beta}$ in $\mathbf{A}=\mathbf{C}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}$. Dobimo napovedno območje:

$$(\boldsymbol{S}-\mathbf{C}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\Big(\mathbf{C}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{C}^{\mathrm{T}}+\boldsymbol{\Lambda\Lambda}^{\mathrm{T}}\Big)^{-1}(\boldsymbol{S}-\mathbf{C}\hat{\boldsymbol{\beta}})<\frac{m}{n-k}\,F_\beta(m,n-k)\,\|\hat{\boldsymbol{\varepsilon}}\|^2\,.$$

4 Testiranje hipotez

V celotnem razdelku se bomo omejili na primer, ko je model Gaussov. Testiranje enodimenzionalnih ničelnih hipotez oblike $s = s_0$ kjer je $s = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}$ (če je model podan z zvezo (1.1)) oziroma $s = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$ (če je model podan z zvezo (1.2)) je enostavno in temelji na relaciji (3.2). Karakteristiko s najprej točkasto ocenimo tako kot v (2.10), torej:

$$\hat{s} = \boldsymbol{a}^{\mathrm{T}}\hat{\boldsymbol{v}} \qquad \text{oziroma} \qquad \hat{s} = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}\,.$$

Za testiranje ničelne hipoteze $s = s_0$ izračunamo testno statistiko:

$$T = \frac{\hat{s} - s_0}{\mathrm{SE}(s)}\,,$$

kjer je ocenjena standardna napaka $\mathrm{SE}(s)$ definirana tako kot v (2.5) oziroma (2). Testna statistika je torej *razmerje med opaženo razliko in ocenjeno standardno napako*. V skladu z (3.2) je $T \sim \mathrm{Student}(n-k)$, zato bomo količini T rekli *Studentova statistika*; na njej torej izvedemo T-test z $n-k$ prostostnimi stopnjami (eno- ali dvostranskega).

Podobno bi lahko na podlagi relacije (3.3) (pri $\boldsymbol{\Lambda} = 0$) testirali hipoteze o vektorjih $\mathbf{A}\boldsymbol{v}$ ali $\mathbf{C}\boldsymbol{\beta}$: le-te bi trdile, da se ustrezni vektor nahaja na nekem afinem prostoru. Vendar pa se izkaže priročneje hipoteze formulirati nekoliko drugače.

Če delamo z zvezo (1.1), problem formuliramo tako, da v prototipu $\boldsymbol{Y} = \boldsymbol{v} + \boldsymbol{\varepsilon}$ testiramo ožji model $\boldsymbol{v} \in U$ proti širšemu modelu $\boldsymbol{v} \in V$, kjer je U podprostor prostora V. Lahko bi vzeli, da je U afin podprostor, a problem lahko brez težav premaknemo, tako da bomo privzeli, da je U kar vektorski podprostor. Spomnimo se oznak $n = \dim W$ in $k = \dim V$ in označimo še $l := \dim U$. Test temelji na razmerju projekcij $q_{V\to U}\boldsymbol{Y}$ in $q_{W\to V}\boldsymbol{Y} = \hat{\boldsymbol{\varepsilon}}$. Vemo, da sta ti dve projekciji nekorelirani, če je model Gaussov, pa tudi neodvisni. Ker je:

$$\frac{1}{\sigma^2}\|q_{V\to U}\boldsymbol{Y}\|^2 \sim \chi^2(k-l) \quad \text{in} \quad \frac{1}{\sigma^2}\|q_{W\to V}\boldsymbol{Y}\|^2 \sim \chi^2(n-k)\,,$$

je:

$$F := \frac{n-k}{k-l}\,\frac{\|q_{V\to U}\boldsymbol{Y}\|^2}{\|q_{W\to V}\boldsymbol{Y}\|^2} = \frac{n-k}{k-l}\,\frac{\|\mathbf{Q}_{V\to U}\boldsymbol{Y}\|^2}{\|\mathbf{Q}_{W\to V}\boldsymbol{Y}\|^2} \sim \mathrm{Fisher}(k-l, n-k)\,.$$

Glede na porazdelitev bomo količini F rekli *Fisherjeva statistika*; na njej izvedemo enostranski F-test. Na to Fisherjevo statistiko pa lahko pogledamo še malo drugače. Glede na to, da je:

$$\mathbf{Q}_{W\to U} = \mathbf{Q}_{W\to V} + \mathbf{Q}_{V\to U}\,,$$

rezidual v ožjem modelu $\mathbf{Q}_{W\to U}\boldsymbol{Y}$ razpade na *pojasnjeni rezidual* $\mathbf{Q}_{V\to U}\boldsymbol{Y}$ in *nepojasnjeni rezidual* $\mathbf{Q}_{W\to V}\boldsymbol{Y}$. Izraz 'pojasnjen' tu pomeni 'pojasnjen z razširitvijo modela': ta rezidual izgine, ko ožji model nadomestimo s širšim, nepojasnjeni rezidual pa je tisti, ki ostane tudi potem, ko model že razširimo.

Pojasnjeni in nepojasnjeni rezidual sta pravokotna, zato se kvadrata njunih norm seštejeta v kvadrat norme celotnega reziduala. Tako v skladu s formulo (2.4) razpade tudi

varianca, ocenjena na podlagi ožjega modela:

$$\hat{\sigma}^2 = \frac{\|\mathbf{Q}_{W\to U}\boldsymbol{Y}\|^2}{n-l} = \frac{\|\mathbf{Q}_{V\to U}\boldsymbol{Y}\|^2}{n-l} + \frac{\|\mathbf{Q}_{W\to V}\boldsymbol{Y}\|^2}{n-l}\,.$$

Prvemu členu bomo rekli *pojasnjena ocenjena varianca modela*, drugemu pa *nepojasnjena ocenjena varianca modela*.

Formula (2.4) pa pove še nekaj drugega: če v splošnem za vektorska prostora $G \subseteq H$ s:

$$\hat{\sigma}^2_{H\to G} = \frac{\|\mathbf{Q}_{H\to G}\boldsymbol{Y}\|^2}{\dim H - \dim G} \tag{4.1}$$

označimo varianco, ocenjeno na podlagi razkoraka med G in H, dobimo, da je Fisherjeva statistika enaka:

$$F = \frac{\hat{\sigma}^2_{V\to U}}{\hat{\sigma}^2_{W\to V}}\,. \tag{4.2}$$

To je torej razmerje med varianco, ocenjeno s pomočjo pojasnjenega reziduala, in varianco, ocenjeno s pomočjo nepojasnjenega reziduala.

Če pa delamo z zvezo (1.2), zadevo zastavimo kot test ožjega modela:

$$\boldsymbol{Y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\eta}\,; \quad \boldsymbol{\gamma} \in \mathbb{R}^l \tag{4.3}$$

proti širšemu modelu:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,; \quad \boldsymbol{\beta} \in \mathbb{R}^k\,; \tag{4.4}$$

veljati mora $\operatorname{im}\mathbf{Z} \subset \operatorname{im}\mathbf{X}$, torej $l < k$. Če postavimo $U := \operatorname{im}\mathbf{Z}$ in $V := \operatorname{im}\mathbf{X}$ (in se spomnimo, da je $W = \mathbb{R}^n$), velja:

$$\mathbf{Q}_{W\to V} = \mathbf{I}_W - \mathbf{P}_{W\to V} = \mathbf{I}_W - \mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\,, \tag{4.5}$$

$$\mathbf{Q}_{V\to U} = \mathbf{P}_{W\to V} - \mathbf{P}_{W\to U} = \mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}} - \mathbf{Z}\big(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\big)^{-1}\mathbf{Z}^{\mathrm{T}}\,. \tag{4.6}$$

Glede na to, da za vsak ortogonalni projektor $\mathbf{Q}$ velja $\|\mathbf{Q}\boldsymbol{Y}\|^2 = \boldsymbol{Y}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{T}}\mathbf{Q}\boldsymbol{Y}$, se ocenjeni varianci, potrebni za izračun Fisherjeve statistike, izražata na naslednji način:

$$\hat{\sigma}^2_{W\to V} = \frac{\|\mathbf{Q}_{W\to V}\boldsymbol{Y}\|^2}{n-k} = \frac{\boldsymbol{Y}\Big(\mathbf{I}_W - \mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}}\Big)\boldsymbol{Y}^{\mathrm{T}}}{n-k}\,,$$

$$\hat{\sigma}^2_{V\to U} = \frac{\|\mathbf{Q}_{V\to U}\boldsymbol{Y}\|^2}{k-l} = \frac{\boldsymbol{Y}\Big(\mathbf{X}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{X}^{\mathrm{T}} - \mathbf{Z}\big(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\big)^{-1}\mathbf{Z}^{\mathrm{T}}\Big)\boldsymbol{Y}^{\mathrm{T}}}{k-l}\,.$$

5 Posebni primeri

5.1 Pričakovana vrednost

Denimo, da opazimo vrednosti nekoreliranih slučajnih spremenljivk $Y_1, Y_2, \ldots Y_n$, ki imajo vse matematično upanje μ in varianco σ^2. Želeli bi oceniti μ in σ. To lahko predstavimo

kot $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, kjer je:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \boldsymbol{\beta} = \mu\,, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} Y_1 - \mu \\ Y_2 - \mu \\ \vdots \\ Y_n - \mu \end{bmatrix}.$$

Z drugimi besedami, to je zveza (1.3) za $k = 1$ in $f_1(x) = 1$. Ni težko videti, da so izpolnjene vse predpostavke, navedene k (1.2).

Za ocenjevanje parametra μ izračunamo $\mathbf{X}^{\mathrm{T}}\mathbf{X} = n$, torej je točkasta ocena:

$$\hat{\mu} = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n} =: \bar{Y}\,.$$

Rezidual je enak:

$$\hat{\varepsilon} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}\bar{Y} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{bmatrix}$$

in nepristranska cenilka za varianco je:

$$\hat{\sigma}^2 = \frac{\|\hat{\varepsilon}\|^2}{n-1} = \frac{1}{n-1}\Big[(Y_1 - \bar{Y})^2 + (Y_2 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2\Big]\,.$$

To je *popravljena vzorčna varianca*. Izračunajmo še ocenjeno standardno napako za μ. Pišemo lahko $\mu = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$, kjer je, kot že rečeno, $\boldsymbol{\beta} = \mu$ in $\boldsymbol{c} = 1$. Torej je $\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c} = 1/n$ in iz (2) dobimo ocenjeno standardno napako:

$$\mathrm{SE}(\mu) = \frac{\hat{\sigma}}{\sqrt{n}}\,.$$

5.2 Enostavna linearna regresija

Enostavna linearna regresija je model *linearnega trenda*, ki pomeni, da je odvisna spremenljivka linearna funkcija pojasnjevalne spremenljivke z določenimi napakami. Naredimo več meritev: pri i-ti meritvi naj ima pojasnjevalna spremenljivka vrednost X_i, odvisna pa Y_i. Iščemo premico $y = a + bx$, ki se tem meritvam najbolje prilega, merilo za prileganje pa naj bo vsota kvadratov odstopanj *v smeri* y. Iskani premici bomo rekli *regresijska premica*. Glede na kriterije, ki smo jih postavili, je vrednosti pojasnjevalne spremenljivke X_i smiselno proglasiti za deterministične, vrednosti odvisne spremenljivke Y_i pa za slučajne,

tako da so odstopanja od premice v smeri y šumi. Tako lahko model predstavimo z zvezo (1.3), kjer je:

$$k = 2\,, \qquad f_1(x) = 1\,, \qquad f_2(x) = x\,, \qquad \beta_1 = a\,, \qquad \beta_2 = b\,.$$

Ekvivalentno lahko model predstavimo z zvezo (1.2), kjer je:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Vektor $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \varepsilon_n \end{bmatrix}$ je vektor odstopanj od premice v smeri y. Seveda privzamemo, da so $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$ nekorelirane z $\mathbb{E}\,\varepsilon_i = 0$ in $\operatorname{var}(\varepsilon_i) = \sigma^2$.

Za točkasto oceno parametrov a in b izračunamo:

$$\begin{aligned} \mathbf{X}^{\mathrm{T}}\mathbf{X} &= \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix}, \\ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} &= \frac{1}{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix}, \\ \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} &= \frac{1}{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n X_i Y_i \\ n\sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i \end{bmatrix}. \end{aligned}$$

Vektor rezidualov pa ima pomembno lastnost, da je vsota njegovih komponent enaka nič. Za vsak vektor $\boldsymbol{a}$ je namreč vsoto komponent mogoče izraziti kot $\begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{X}^{\mathrm{T}}\boldsymbol{a}$. Vsota komponent reziduala je tako enaka:

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{X}^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}) = \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} - \mathbf{X}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}\right) = 0\,.$$

Sledi $\sum_{i=1}^n (Y_i - \hat{a} - \hat{b}X_i) = 0$. Če z $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ in $\bar{Y} := \frac{1}{n}\sum_{i=1}^n Y_i$ označimo povprečji, velja tudi $\bar{Y} = \hat{a} + \hat{b}\bar{X}$. *Točka $(\bar{X}, \bar{Y})$ torej vedno leži na regresijski premici.*

S pomočjo povprečij pa se dasta tudi alternativno izraziti cenilki $\hat{a}$ in $\hat{b}$ regresijskih koeficientov. Velja namreč:

$$\begin{aligned} \hat{b} &= \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}\,, \\ \hat{a} &= \bar{Y} - \hat{b}\bar{X} = \sum_{i=1}^n \left(\frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) Y_i\,. \end{aligned}$$

Iz reziduala dobimo še nepristransko cenilko variance modela:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}X_i)^2\,.$$

V nadaljevanju bomo izračunali ocenjene standardne napake za a, b in $aX_0+b+\lambda\varepsilon_0$. Slednje je mišljeno kot nova izmerjena vrednost odvisne spremenljivke pri vrednosti pojasnjevalne spremenljivke X_0, pri čemer dopuščamo možnost, da je šum drugačne velikosti kot pri že znanih meritvah. Seveda privzamemo, da je novi šum ε_0 nekoreliran z $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$ ter da velja $\mathbb{E}\,\varepsilon_0 = 0$ in $\mathrm{var}(\varepsilon_0) = \sigma^2$.

Ocenjeno standardno napako bomo najprej izračunali za strmino b. Za ta namen postavimo $\boldsymbol{c} := \begin{bmatrix}1\\0\end{bmatrix}$: tako bo $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta} = b$. Izračunajmo:

$$\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c} = \frac{n}{n\sum_{i=1}^{n}X_i^2 - \big(\sum_{i=1}^{n}X_i\big)^2} = \frac{1}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\,.$$

Iz (2) zdaj dobimo:

$$\mathrm{SE}(b) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}\,.$$

Za statistično sklepanje o koeficientu a moramo postaviti $\boldsymbol{c} := \begin{bmatrix}0\\1\end{bmatrix}$. Dobimo:

$$\begin{aligned}\boldsymbol{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\boldsymbol{c} &= \frac{\sum_{i=1}^{n}X_i^2}{n\sum_{i=1}^{n}X_i^2 - \big(\sum_{i=1}^{n}X_i\big)^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2 + n\bar{X}^2}{n\sum_{i=1}^{n}(X_i-\bar{X})^2} = \\ &= \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\,.\end{aligned}$$

Sledi:

$$\mathrm{SE}(a) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}}\,.$$

Zdaj se pa spomnimo, da je a koordinata y na regresijski premici pri $x = 0$. Novo meritev odvisne spremenljivke pri $X_0 = 0$ torej predstavimo kot $a+\lambda\varepsilon_0$. Podobno kot prej iz (2.11) dobimo:

$$\mathrm{SE}(a+\lambda\varepsilon_0) = \hat{\sigma}\sqrt{\lambda^2 + \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}}\,.$$

Za izračun ocenjene standardne napake za novo izmerjeno vrednost pri splošni vrednosti pojasnjevalne spremenljivke X_0 pa lahko model preprosto premaknemo. Naš izvirni model:

$$Y_i = a + bX_i + \varepsilon_i$$

prepišimo v obliki:

$$Y_i = \tilde{a} + b\tilde{X}_i + \varepsilon_i\,,$$

kjer je $\tilde{X}_i = X_i - X_0$ in $\tilde{a} = a + bX_0$. Slednje je spet enostavna linearna regresija. Pomudimo se najprej pri oceni variance modela. To lahko ocenimo v obeh modelih. Toda zaradi linearnosti cenilk je $\hat{\tilde{a}} = \hat{a} + \hat{b}X_0$ in zato:

$$Y_i - \hat{\tilde{a}} - \hat{b}\tilde{X}_i = Y_i - \hat{a} - \hat{b}X_0 - \hat{b}(X_i - X_0) = Y_i - \hat{a} - \hat{b}X_i\,.$$

Torej je ocena za σ^2 v obeh modelih enaka:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\tilde{a}} - \hat{b}\tilde{X}_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}X_i)^2\,.$$

Zdaj lahko izračunamo želeno ocenjeno standardno napako:

$$\mathrm{SE}(a + bX_0 + \lambda\varepsilon_0) = \mathrm{SE}(\tilde{a} + \lambda\varepsilon_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{\tilde{X}}^2}{\sum_{i=1}^{n}(\tilde{X}_i - \bar{\tilde{X}})^2}} = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(\bar{X} - X_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}\,.$$

5.3 Primerjava po parih

Denimo, da vsakič isto stvar izmerimo v dvojih okoliščinah, recimo pred določeno akcijo in po njej. To napravimo na m enotah. Meritev na posamezni enoti v prvih okoliščinah (recimo pred akcijo) označimo z Y_i', v drugih okoliščinah (recimo po akciji) pa z Y_i''. Zanima nas, ali spremenjene okoliščine (npr. izvedena akcija) kaj vplivajo na rezultate naših meritev. Privzamemo, da so meritve nekorelirane ter da je:

$$\mathbb{E}(Y_i') = \mu_i\,, \quad \mathbb{E}(Y_i'') = \mu_i + \delta\,, \quad \mathrm{var}(Y_i') = (\sigma')^2\,, \quad \mathrm{var}(Y_i'') = (\sigma'')^2\,.$$

Pričakovana vrednost se torej lahko spreminja od enote do enote, za pričakovani vpliv spremembe okoliščin (akcije) pa privzamemo, da je ves čas enak. Varianca je lahko odvisna od okoliščin, ne pa tudi od enote.

Če želimo prevesti v jezik linearne regresije, najprej zapišimo:

$$Y_i' = \mu_i + \varepsilon_i'\,, \qquad Y_i'' = \mu_i + \delta + \lambda\varepsilon_i''\,,$$

kjer so ε_i' in ε_i'' nekorelirane z $\mathbb{E}\,\varepsilon_i' = \mathbb{E}\,\varepsilon_i'' = 0$ in $\mathrm{var}(\varepsilon_i') = \mathrm{var}(\varepsilon_i'') = (\sigma')^2$. Torej je $\lambda = \sigma''/\sigma$. To je neznan deterministično parameter, ki pa ne bo konstitutivni del regresijskega modela. Model bo deloval, če λ v statističnem sklepanju ne bo zajet.

Če označimo:

$$\boldsymbol{Y}' = \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_m' \end{bmatrix}\,, \qquad \boldsymbol{Y}'' = \begin{bmatrix} Y_1'' \\ Y_2'' \\ \vdots \\ Y_m'' \end{bmatrix}\,, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}\,, \qquad \boldsymbol{\varepsilon}' = \begin{bmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \vdots \\ \varepsilon_m' \end{bmatrix}\,, \qquad \boldsymbol{\varepsilon}'' = \begin{bmatrix} \varepsilon_1'' \\ \varepsilon_2'' \\ \vdots \\ \varepsilon_m'' \end{bmatrix}$$

ter če še z $\mathbf{0}_m$ označimo stolpec iz m ničel, z $\mathbf{1}_m$ stolpec iz m enic, z $\mathbf{I}_m$ pa identiteto na $\mathbb{R}^m$, lahko model zapišemo v obliki (1.2) z:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}' \\ \frac{1}{\lambda}\boldsymbol{Y}'' \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_m \\ \frac{1}{\lambda}\mathbf{I}_m & \frac{1}{\lambda}\mathbf{1}_m \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\mu} \\ \delta \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}' \\ \boldsymbol{\varepsilon}'' \end{bmatrix}.$$

Velja še $n = 2m$ in $k = m + 1$. Izračunajmo:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} (1 + \frac{1}{\lambda^2})\mathbf{I}_m & \frac{1}{\lambda^2}\mathbf{1}_m \\ \mathbf{1}_m^{\mathrm{T}} & n\lambda^{-2} \end{bmatrix},$$
$$\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} = \begin{bmatrix} \frac{\lambda^2}{1+\lambda^2}\mathbf{I}_m + \frac{1}{n(1+\lambda^2)}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}} & -\frac{1}{n}\mathbf{1}_m \\ -\frac{1}{n}\mathbf{1}_m^{\mathrm{T}} & \frac{1}{n}(1+\lambda^2) \end{bmatrix},$$

Od tod po krajšem računu dobimo cenilko za $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \frac{\lambda^2}{1+\lambda^2}\boldsymbol{Y}' + \frac{1}{1+\lambda^2}\boldsymbol{Y}'' - \frac{1}{n(1+\lambda^2)}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}(\boldsymbol{Y}'' - \boldsymbol{Y}) \\ \frac{1}{n}\mathbf{1}_m^{\mathrm{T}}(\boldsymbol{Y}'' - \boldsymbol{Y}) \end{bmatrix}.$$

Po komponentah lahko to zapišemo tudi takole:

$$\hat{\mu}_i = \frac{\lambda^2}{1+\lambda^2}\,Y_i' + \frac{1}{1+\lambda^2}\,Y_i'' + \frac{1}{(1+\lambda^2)}\,(\bar{Y}'' - \bar{Y})\,,$$
$$\hat{\delta} = \bar{Y}'' - \bar{Y}'\,.$$

Posebej opazimo, da je cenilka za δ neodvisna od dodatnega parametra λ (in zelo intuitivna). Izračunajmo še ocenjeno standardno napako. Najprej opazimo:

$$\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \frac{1+\lambda^2}{n}\,.$$

Zdaj moramo dobiti še cenilko variance modela, ta pa temelji na rezidualu. Najprej izračunajmo:

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} \frac{\lambda^2}{1+\lambda^2}\boldsymbol{Y}' + \frac{1}{1+\lambda^2}\boldsymbol{Y}'' - \frac{1}{n(1+\lambda^2)}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}(\boldsymbol{Y}'' - \boldsymbol{Y}') \\ \frac{\lambda}{1+\lambda^2}\boldsymbol{Y}' + \frac{1}{\lambda(1+\lambda^2)}\boldsymbol{Y}'' + \frac{\lambda}{n(1+\lambda^2)}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}(\boldsymbol{Y}'' - \boldsymbol{Y}') \end{bmatrix}.$$

Rezidual je tako enak:

$$\hat{\boldsymbol{\varepsilon}} = \frac{1}{1+\lambda^2}\begin{bmatrix} \mathbf{I}_m \\ \lambda\mathbf{I}_m \end{bmatrix}\left(\mathbf{I}_m - \tfrac{1}{n}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}\right)(\boldsymbol{Y}'' - \boldsymbol{Y}')\,,$$

kvadrat njegove norme pa je enak:

$$\|\hat{\boldsymbol{\varepsilon}}\|^2 = \frac{1}{1+\lambda^2}\left\|\left(\mathbf{I}_m - \tfrac{1}{n}\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}\right)(\boldsymbol{Y}'' - \boldsymbol{Y}')\right\|^2 = \frac{1}{1+\lambda^2}\sum_{i=1}^{n}\left((Y_i'' - Y_i' - (\bar{Y}'' - \bar{Y}')\right)^2.$$

Iz formule (2.4) dobimo:

$$\hat{\sigma}^2 = \frac{1}{(n-1)(1+\lambda^2)} \sum_{i=1}^{n} \big((Y_i'' - Y_i' - (\bar{Y}'' - \bar{Y}')\big)^2$$

in končno:

$$\mathrm{SE}(\delta) = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} \big((Y_i'' - Y_i' - (\bar{Y}'' - \bar{Y}')\big)^2}\,.$$

Tudi ocenjena standardna napaka ni odvisna od λ. Opazimo pa še nekaj: natanko isto cenilko in ocenjeno standardno napako bi dobili, če bi uporabili rezultate iz podrazdelka 5.1 na razlikah $Y_i'' - Y_i$, ki seveda ustrezajo tamkajšnjemu regresijskemu modelu. To ne velja le pod predpostavkami iz tega razdelka. Lahko jih namreč oslabimo tako, da ne privzamemo več nekoreliranosti slučajnih spremenljivk $Y_1', \ldots Y_m', Y_1'', \ldots Y_m''$, marveč le nekoreliranost parov $(Y_1', Y_1''), \ldots (Y_m', Y_m'')$: znotraj para (Y_i', Y_i'') je dopustna tudi koreliranost.

5.4 Primerjava dveh skupin

Privzemimo, da isto količino izmerimo na dveh skupinah. Meritve na prvi skupini so $Y_1', Y_2', \ldots Y_{m'}'$, na drugi skupini pa $Y_1'', Y_2'', \ldots Y_{m''}''$. Za vse meritve privzamemo, da so neodvisne. Nadalje naj bo $\mathbb{E}(Y_i') = \mu'$, $\mathbb{E}(Y_i'') = \mu''$ in $\mathrm{var}(Y_i') = \mathrm{var}(Y_i'') = \sigma^2$. V nasprotju s prejšnjim podrazdelkom sta torej lahko skupini različno veliki, zato pa privzamemo popolno neodvisnost in homoskedastičnost, poleg tega pa se pričakovana vrednost ne sme spreminjati od enote do enote. Podobno kot prej pa nas bo zanimalo predvsem, ali so razlike med skupinama plod njunih različnih ustrojev. Zanimala nas bo torej razlika $\mu'' - \mu'$.

Regresijski model zapišimo v obliki:

$$Y_i' = \mu' + \varepsilon_i'\,, \qquad Y_i'' = \mu'' + \varepsilon_i''\,,$$

kjer so ε_i' in ε_i'' nekorelirane z $\mathbb{E}\,\varepsilon_i' = \mathbb{E}\,\varepsilon_i'' = 0$ in $\mathrm{var}(\varepsilon_i') = \mathrm{var}(\varepsilon_i'') = \sigma^2$. Če označimo:

$$\boldsymbol{Y}' = \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_{m'}' \end{bmatrix}, \qquad \boldsymbol{Y}'' = \begin{bmatrix} Y_1'' \\ Y_2'' \\ \vdots \\ Y_{m''}'' \end{bmatrix}, \qquad \boldsymbol{\varepsilon}' = \begin{bmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \vdots \\ \varepsilon_{m'}' \end{bmatrix}, \qquad \boldsymbol{\varepsilon}'' = \begin{bmatrix} \varepsilon_1'' \\ \varepsilon_2'' \\ \vdots \\ \varepsilon_{m''}'' \end{bmatrix},$$

lahko model zapišemo v obliki (1.2) z:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}' \\ \boldsymbol{Y}'' \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \mathbf{1}_{m'} & \mathbf{0}_{m'} \\ \mathbf{0}_{m''} & \mathbf{1}_{m''} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu' \\ \mu'' \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}' \\ \boldsymbol{\varepsilon}'' \end{bmatrix}.$$

Velja še $n = m' + m''$ in $k = 2$. Izračunajmo:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} m' & 0 \\ 0 & m'' \end{bmatrix}, \big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1} = \begin{bmatrix} \frac{1}{m'} & 0 \\ 0 & \frac{1}{m''} \end{bmatrix}.$$

Od tod po krajšem računu dobimo cenilko za $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \frac{1}{m'}\,\mathbf{1}_{m'}^{\mathrm{T}}\boldsymbol{Y}' \\ \frac{1}{m''}\,\mathbf{1}_{m''}^{\mathrm{T}}\boldsymbol{Y}'' \end{bmatrix}.$$

Po komponentah je to kar:

$$\hat{\mu}' = \bar{Y}'\,, \qquad \hat{\mu}'' = \bar{Y}''\,,$$

kar je spet zelo intuitivno. Izračunajmo zdaj standardno napako za $\mu'' - \mu' = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$, kjer je $\boldsymbol{c} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Velja:

$$\mathbf{c}^{\mathrm{T}}\big(\mathbf{X}^{\mathrm{T}}\mathbf{X}\big)^{-1}\mathbf{c} = \frac{1}{m'} + \frac{1}{m''}\,.$$

Lotimo se sedaj reziduala. Velja:

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} \frac{1}{m'}\,\mathbf{1}_{m'}\mathbf{1}_{m'}^{\mathrm{T}}\boldsymbol{Y}' \\ \frac{1}{m''}\,\mathbf{1}_{m''}\mathbf{1}_{m''}^{\mathrm{T}}\boldsymbol{Y}'' \end{bmatrix},$$

torej je rezidual enak:

$$\hat{\boldsymbol{\varepsilon}} = \begin{bmatrix} \big(\mathbf{I}_{m'} - \frac{1}{m'}\,\mathbf{1}_{m'}\mathbf{1}_{m'}^{\mathrm{T}}\big)\boldsymbol{Y}' \\ \big(\mathbf{I}_{m''} - \frac{1}{m''}\,\mathbf{1}_{m''}\mathbf{1}_{m''}^{\mathrm{T}}\big)\boldsymbol{Y}'' \end{bmatrix}$$

kvadrat njegove norme pa je:

$$\|\hat{\boldsymbol{\varepsilon}}\|^2 = \sum_{i=1}^{m'}(Y_i' - \bar{Y}')^2 + \sum_{i=1}^{m''}(Y_i'' - \bar{Y}'')^2$$

in iz formule (2.4) dobimo:

$$\hat{\sigma}^2 = \frac{1}{m' + m'' - 2}\left(\sum_{i=1}^{m'}(Y_i' - \bar{Y}')^2 + \sum_{i=1}^{m''}(Y_i'' - \bar{Y}'')^2\right),$$

tako da je končno:

$$\mathrm{SE}(\mu'' - \mu') = \sqrt{\frac{1}{m' + m'' - 2}\left(\frac{1}{m'} + \frac{1}{m''}\right)\left(\sum_{i=1}^{m'}(Y_i' - \bar{Y}')^2 + \sum_{i=1}^{m''}(Y_i'' - \bar{Y}'')^2\right)}\,.$$

5.5 Primerjava več skupin: analiza variance z enojno klasifikacijo

Recimo sedaj, da isto količino večkrat opazimo na več skupinah in želimo ovrednotiti razlike med skupinami. Naj Y_{ij}, kjer je $1 = 1, 2, \ldots r$ in $j = 1, 2, \ldots m_i$, označuje j-to opažanje na i-ti skupini. Testiramo ožji model:

$$Y_{ij} = \mu + \eta_{ij}$$

proti širšemu modelu:

$$Y_{ij} = \mu_i + \epsilon_{ij}\,,$$

kjer so tako šumi η_{ij} kot tudi ε_{ij} neodvisni in porazdeljeni normalno $\mathrm{N}(0,\sigma^2)$. Označimo:

$$\boldsymbol{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{im_i} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_r \end{bmatrix}$$

Označimo spet z $\mathbf{1}_s$ stolpec iz s enic in še $m := m_1 + m_2 + \cdots + m_r$. To je tudi dimenzija opažanja, torej $n = m$. Ožji model je isti kot v podrazdelku 5.1, torej ustreza (4.3) z:

$$\mathbf{Z} = \mathbf{1}_m\,, \qquad \boldsymbol{\gamma} = \mu\,, \qquad l = 1\,.$$

Če označimo skupno povprečje:

$$\bar{Y} := \frac{1}{m}\,\mathbf{1}_m^{\mathrm{T}}\boldsymbol{Y} = \frac{1}{m}\sum_{i=1}^{r}\sum_{j=1}^{m_i} Y_{ij}\,,$$

neposredno iz podrazdelka 5.1 dobimo $\hat{\mu} = \bar{Y}$ in:

$$\mathbf{P}_{W\to U} = \bar{Y}\mathbf{1}_m\,.$$

Širši model pa ustreza (4.4) z:

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{m_1} & \mathbf{0}_{m_1} & \cdots & \mathbf{0}_{m_1} \\ \mathbf{0}_{m_2} & \mathbf{1}_{m_2} & \cdots & \mathbf{0}_{m_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m_r} & \mathbf{0}_{m_r} & \cdots & \mathbf{1}_{m_r} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \end{bmatrix}, \qquad k = r\,.$$

Izračunajmo:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_r \end{bmatrix}, \qquad \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} = \begin{bmatrix} \frac{1}{m_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{m_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{m_r} \end{bmatrix}.$$

Če označimo še povprečja posameznih skupin:

$$\bar{Y_i} = \frac{1}{m_i}\mathbf{1}_{m_i}\boldsymbol{Y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} Y_{ij}\,,$$

velja:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} = \begin{bmatrix}\bar{Y}_1\\ \bar{Y}_2\\ \vdots\\ \bar{Y}_r\end{bmatrix}, \qquad \mathbf{P}_{W\to V}\boldsymbol{Y} = \begin{bmatrix}\bar{Y}_1\mathbf{1}_{m_1}\\ \bar{Y}_2\mathbf{1}_{m_2}\\ \vdots\\ \bar{Y}_r\mathbf{1}_{m_r}\end{bmatrix}.$$

Pojasnjeni rezidual je torej v skladu s formulo (4.6) enak:

$$\mathbf{Q}_{V\to U}\boldsymbol{Y} = \mathbf{P}_{W\to V}\boldsymbol{Y} - \mathbf{P}_{W\to U}\boldsymbol{Y} = \begin{bmatrix}(\bar{Y}_1-\bar{Y})\mathbf{1}_{m_1}\\ (\bar{Y}_2-\bar{Y})\mathbf{1}_{m_2}\\ \vdots\\ (\bar{Y}_r-\bar{Y})\mathbf{1}_{m_r}\end{bmatrix}$$

ocena variance na njegovi podlagi pa je:

$$\hat{\sigma}^2_{W\to V} = \frac{1}{r-1}\sum_{i=1}^{r} m_i(\bar{Y}_i-\bar{Y})^2\,.$$

Nepojasnjeni rezidual pa je v skladu s formulo (4.5) enak:

$$\mathbf{Q}_{W\to V}\boldsymbol{Y} = \boldsymbol{Y} - \mathbf{P}_{W\to V}\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y}_1-\bar{Y}_1\mathbf{1}_{m_1}\\ \boldsymbol{Y}_2-\bar{Y}_2\mathbf{1}_{m_2}\\ \vdots\\ \boldsymbol{Y}_r-\bar{Y}_r\mathbf{1}_{m_r}\end{bmatrix}$$

in ocena variance na njegovi podlagi je:

$$\hat{\sigma}^2_{W\to V} = \frac{1}{m-r}\sum_{i=1}^{r}\sum_{j=1}^{m_i}(Y_{ij}-\bar{Y}_i)^2\,.$$

Ožji model testiramo proti širšemu z F-testom na testni statistiki $F = \hat{\sigma}^2_{V\to U}/\hat{\sigma}^2_{W\to V}$ z $(m-k, k-1)$ prostostnimi stopnjami.

Za konec si to oglejmo še v kontekstu varianc. Pojasnjena varianca:

$$\frac{1}{m-1}\|\mathbf{Q}_{V\to U}\boldsymbol{Y}\|^2 = \sum_{i=1}^{r}\frac{m_i}{m-1}\,(\bar{Y}_i-\bar{Y})^2$$

je *popravljena varianca med skupinami*, medtem ko je nepojasnjena varianca:

$$\frac{1}{m-1}\|\mathbf{Q}_{W\to V}\boldsymbol{Y}\|^2 = \frac{1}{m-1}\sum_{i=1}^{r}\sum_{j=1}^{m_i}(Y_{ij}-\bar{Y}_i)^2$$

popravljena varianca znotraj skupin.

Literatura

[1] M. J. Schervish: *Theory of Statistics*. Druga izdaja. Springer-Verlag, New York, 1997.