Linearni model v matrični obliki

Odzivno spremenljivko y modeliramo na podlagi k regresorjev (splošni normalni linearni model)

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

V matrični obliki:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

Linearni model v matrični obliki

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \ddots & \vdots & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- $\mathbf{y}$ vektor odzivne spremenljivke
- $\mathbf{X}$ modelska matrika reda $(n \times k + 1)$
- $\boldsymbol{\beta}$ vektor parametrov modela velikosti $(k + 1) \times 1$
- $\boldsymbol{\varepsilon}$ vektor napak velikosti $(n \times 1)$, $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ in $Var(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, $\mathbf{I}$ je enotska diagonalna matrika reda $n \times n$

Minimiramo vsoto kvadratov napak:

$$S(\beta_0, \beta_1, ..., \beta_k) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}))^2$$

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Parcialno odvajamo po parametrih β_j, $j = 0, ..., k$, in odvode izenačimo z 0. Dobimo **normalni sistem** $k + 1$ **linearnih enačb**:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Rešitev obstaja, če je $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ **nesingularna**.

$\mathbf{X}^{\mathrm{T}}\mathbf{X}$ je nesingularna:

- če je $n \geq k+1$; to pomeni, da je število enot vsaj tako veliko kot število ocenjevanih parametrov;
- če nobena spremenljivka ni linearna kombinacija ostalih spremenljivk, kar pomeni, da ima matrika $\mathbf{X}$ polni rang $k+1$, gre za **linearni model polnega ranga** (*full rank linear model*).

Rešitev je vektor cenilk parametrov $\mathbf{b} = (b_0, b_1, ..., b_k)$:

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

Posamezno cenilko b_j lahko zapišemo

$$b_j = \frac{\sum_{i=1}^{n} x_{ij}^{*} y_i}{\sum_{i=1}^{n} (x_{ij}^{*})^2},$$

x_{ij}^{*} je vrednost spremenljivke x_{ij} po tem, ko je bila prilagojena na vse ostale napovedne spremenljivke $x_1, ..., x_k$ brez spremenljivke x_j.

Vektor prilagojenih vrednosti: $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$

Vektor ostankov: $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$

Izrek 2.1: v linearnem modelu polnega ranga so cenilke parametrov izračunane po metodi najmanših kvadratov $\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ nepristranske:

$$\mathbb{E}(\mathbf{b}) = \boldsymbol{\beta}$$

variančno-kovariančno matrika vektorja cenilk je

$$Var(\mathbf{b}) = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

Dokaz:

$$\mathbb{E}(\mathbf{b}) = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbb{E}(\mathbf{y}) = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta}$$

Edina predpostavka: $\mathbb{E}(\boldsymbol{\varepsilon}) = 0$. Cenilke parametrov modela so nepristranske tudi, če varianca σ^2 ni konstantna ali če so napake korelirane.

$$Var(\mathbf{b}) = ((\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}})Var(\mathbf{y})((\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}})^{\mathrm{T}} = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

Tu smo upoštevali predpostavko konstantne variance $Var(\mathbf{y}) = \sigma^2$.

Gauss-Markov izrek:

Naj bo $\mathbf{b}^*$ nepristranska cenilka za $\boldsymbol{\beta}$ in $\mathbf{b}$ cenilka za $\boldsymbol{\beta}$ po metodi najmanjših kvadratov, potem velja, da je $Var(b_i) \leq Var(b_i^*)$, $i = 1, ..., k+1$.

Pravimo, da je **b najboljša linearna nepristranska cenilka** za $\boldsymbol{\beta}$ (BLUE, *Best Linear Unbiesed Estimator*).

(Brez dokaza.)

Ker so cenilke parametrov linearnega normalnega modela linearne kombinacije odzivne spremenljivke, za katero smo predpostavili normalno porazdelitev, je njihova porazdelitev **večrazsežna normalna porazdelitev**.

Nepristranska cenilka za σ^2

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-k-1}(\mathbf{y} - \mathbf{X}\mathbf{b})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Poglejmo povezavo med $\hat{\mathbf{y}}$ in $\mathbf{y}$:

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

Matrika $\mathbf{H}$ reda $n \times n$ (*hat matrix*) je ključna pri izračunu napovedi $\hat{\mathbf{y}}$:

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

Matrika $\mathbf{H}$ ima lepe lastnosti, pokažemo lahko, da velja:
$\mathbf{H} = \mathbf{H}^{\mathrm{T}} = \mathbf{H}^2$ (idempotentna matrika).

Vektor ostankov lahko zdaj zapišemo tudi z matriko $\mathbf{H}$:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

Varianca ostankov $Var(\mathbf{e})$ je ob predpostavki $Var(\epsilon) = \sigma^2\mathbf{I}$:

$$Var(\mathbf{e}) = (\mathbf{I} - \mathbf{H})\ (\sigma^2\mathbf{I})\ (\mathbf{I} - \mathbf{H})^{\mathrm{T}} = \sigma^2(\mathbf{I} - \mathbf{H})$$

Glavna predpostavka za statistično sklepanje je:
cenilke parametrov **b** so porazdeljene po **večrazsežnostni normalni porazdelitvi**

$$\mathbf{b} \sim N(\boldsymbol{\beta}, \ \ \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$$

Velja tudi, da so cenilke prametrov **b** neodvisne od cenilke variance napak $\hat{\sigma}^2$.

Porazdelitev za reskalirano varianco napak $(n-k-1)\hat{\sigma}^2/\sigma^2$ je $\boldsymbol{\chi}^{\mathbf{2}}$**-porazdelitev** s stopinjami prostosti $SP = n-k-1$:

$$\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

Interval zaupanja za posamezen parameter modela β_j, $j = 0, ..., k$, **ob upoštevanju ostalih regresorjev v modelu** imenujemo **parcialni interval zaupanja**.
Definiran je na podlagi statistike

$$\frac{b_j - \beta_j}{\hat{\sigma}\sqrt{a_{jj}}}$$

$\sqrt{a_{jj}}$ je diagonalni element matrike $\mathbf{A} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$.

Kakšna je porazdelitev te statistike?

Velja, da je statistika:

$$\frac{b_j - \beta_j}{\sigma\sqrt{a_{jj}}} \sim N(0,1)$$

Zgornji izraz delimo s korenom reskalirane variance napak deljene z $(n-k-1)$, dobimo statistiko, ki je porazdeljena po t-porazdelitvi z $SP = n-k-1$ (Izrek1.4):

$$\frac{b_j - \beta_j}{\sigma\sqrt{a_{jj}}} / \sqrt{\frac{\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2}}{n-k-1}} \sim t_{n-k-1}$$

Ko zgornji izraz poenostavimo, dobimo

$$\frac{b_j - \beta_j}{\hat{\sigma}\sqrt{a_{jj}}} \sim t_{n-k-1}$$

Posledično je $100(1-\alpha)$ % parcialni interval zaupanja za β_j ob upoštevanju ostalih napovednih spremenljivk v modelu:

$$\left(b_j - |t_{\frac{\alpha}{2};n-k-1}|\hat{\sigma}\cdot\sqrt{a_{jj}},\quad b_j + |t_{\frac{\alpha}{2};n-k-1}|\hat{\sigma}\cdot\sqrt{a_{jj}}\right)$$

Funkcija `confint()` vrne parcialne 95 % intervale zaupanja za vse parametre v modelu.

Za j-ti parameter lahko zapišemo H_0 in H_1, $j = 0, ..., k$:

$H_0 : \beta_j = \gamma_j$ ob upoštevanju vseh ostalih členov v modelu
$H_1 : \beta_j \neq \gamma_j$

Testna statistika je

$$t = \frac{b_j - \gamma_j}{\hat{\sigma}\sqrt{a_{jj}}}$$

ki je pod ničelno domnevo porazdeljena t_{n-k-1}.

Rezultate testiranja posamičnih $k + 1$ ničelnih domnev za parametre modela dobimo v povzetku `lm` modela. Ti testi so medsebojno odvisni. Hkratnost testiranja odvisnih ničelnih domnev tu ni upoštevana.

Linearni model v matrični obliki
Tabela analize variance

Spomnimo se

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

$$SS_{yy} = SS_{model} + SS_{residual}$$

$$= (\mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} - C) + (\mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y})$$

kjer je $C = (\sum_{i=1}^{n} y_i)^2/n$ je t. i. korekcijski člen.

Tabela: Shema tabele ANOVA za spošni linearni model s k regresorji

Vir variabilnosti	df	SS	$MS = SS/df$	F
Model	k	SS_{model}	MS_{model}	$MS_{model}/MS_{residual}$
Ostanek (*Residual*)	$n-k-1$	$SS_{residual}$	$MS_{residual}$	
Skupaj	$n-1$	SS_{yy}		

Za linearni model z več napovednimi spremenljivkami na podlagi F-statistike testiramo ničelno domnevo, da so parametri $(\beta_1, \beta_2, ..., \beta_k)$ hkrati enaki nič:

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

$$H_1 : \text{vsaj en parameter} \quad \beta_j, \quad j = 1, ..., k, \text{ je različen od nič}$$

Ničelno domnevo testiramo na podlagi F-statistike

$$F = \frac{SS_{model}/k}{SS_{residual}/(n-k-1)}$$

Ob predpostavki $\varepsilon \sim iid\ N(0, \sigma^2\mathbf{I})$ je F-statistika porazdeljena $F_{k,n-k-1}$.

V izpisu povzetka `lm` modela najdemo poleg koeficienta determinacije

$$R^2 = SS_{model}/SS_{yy} = 1 - SS_{residual}/SS_{yy}$$

tudi **prilagojeni koeficient determinacije** (*Adjusted R-squared*), ki vsebuje tudi informacijo o stopinjah prostosti:

$$R_a^2 = 1 - \frac{\frac{SS_{residual}}{(n-k-1)}}{\frac{SS_{yy}}{(n-1)}} = 1 - \frac{(n-1)\hat{\sigma}^2}{SS_{yy}}$$

R_a^2 je bolj primeren za primerjavo dveh modelov z različnimi napovednimi spremenljivkami kot R^2. V primerjavi z ostalimi kriteriji za izbiro ustreznega modela (jih še ne poznamo), je njegova uporaba zastarela.

Za vsak y_i, $i = 1, ...n$, imamo vrednosti k napovednih spremenljivk $(x_{i1}, x_{i2}, ..., x_{ik})$. Označimo z $\mathbf{x}_i$ vektor $\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ik})^{\mathrm{T}}$ in zapišimo

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i$$

Zapišimo še napovedano vrednost za odzivno spremenljivko y_* pri vrednostih napovednih spremenljivk $\mathbf{x}_* = (1, x_{*1}, x_{*2}, ..., x_{*k})^{\mathrm{T}}$

$$y_* = \mathbf{x}_*^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_*$$

napaka ε_* ima pričakovano vrednost 0, varianco σ^2 in je neodvisna od ε_i, $i = 1, 2, , ..., n$. Zanimata nas dva intervala zaupanja, najprej za **povprečno napoved** $\mathbf{x}_*^{\mathrm{T}} \boldsymbol{\beta}$ in nato še za **posamično napoved** y_*.

Linearni model v matrični obliki

Varianca povprečne napovedi in matrika $\mathbf{H}$

Diagonalnim elementom matrike $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ pravimo **vzvodi**. Označimo jih h_{ii}.

Pokazali smo že, da velja $\hat{y} = \mathbf{H}y$, oziroma $\hat{y}_i = h_{ii}y_i$.

Torej vzvod predstavlja neko mero vpliva y_i na $\hat{y}_i$.

Po drugi strani je vzvod h_{ii} odvisen samo od napovednih spremenljivk:

$$h_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i$$

$\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ik})^T$ vsebuje komponente i-te vrstice modelske matrike $\mathbf{X}$.

Za varianco prilagojene vrednosti/povprečne napovedi $\hat{y}$ se pokaže, da je sorazmerna s h_{ii}:

$$\begin{aligned} Var(\hat{y}_i) &= Var(\mathbf{x}_i^T \mathbf{b}) \\ &= \mathbf{x}_i^T Var(\mathbf{b}) \mathbf{x}_i = \\ &= \sigma^2 \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \\ &= \sigma^2 h_{ii} \end{aligned}$$

Vzvod ima vrednost med $\frac{1}{n}$ in 1, kar pomeni, da je varianca povprečne napovedi vedno manjša od variance napak σ^2.

Za enostavno linearno regresijo že vemo, da varianco prilagojene vrednosti pri x_i izrazimo

$$Var(\hat{y}(x_i)) = \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

kar pomeni, da je vzvod

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$$

Napoved v točki x_* je $\hat{y}_* = \mathbf{x}_*^{\mathrm{T}}\mathbf{b}$, njena pričakovana vrednost je

$$\mathbb{E}(\mathbf{x}_*^{\mathrm{T}}\mathbf{b}) = \mathbf{x}_*^{\mathrm{T}}\boldsymbol{\beta}$$

in njena varianca

$$Var(\mathbf{x}_*^{\mathrm{T}}\mathbf{b}) = \mathbf{x}_*^{\mathrm{T}}Var(\mathbf{b})\mathbf{x}_* = \sigma^2\mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*$$

$\mathbf{A} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$.

Ker je napoved $\mathbf{x}_*^{\mathrm{T}}\mathbf{b}$ linearna kombinacija normalno porazdeljenih spremenljivk, velja, da je porazdeljena normalno

$$\mathbf{x}_*^{\mathrm{T}}\mathbf{b} \sim N(\mathbf{x}_*^{\mathrm{T}}\boldsymbol{\beta}, \quad \sigma^2\mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*)$$

Velja

$$\frac{\mathbf{x}_*^{\mathrm{T}}\mathbf{b} - \mathbf{x}_*^{\mathrm{T}}\boldsymbol{\beta}}{\hat{\sigma}\sqrt{\mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*}} \sim t_{n-k-1}$$

in $(1-\alpha)100$ % interval zaupanja za povprečno napoved je

$$\left(\mathbf{x}_*^{\mathrm{T}}\mathbf{b} - |t_{\frac{\alpha}{2};n-k-1}|\hat{\sigma}\cdot\sqrt{\mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*}, \quad \mathbf{x}_*^{\mathrm{T}}\mathbf{b} + |t_{\frac{\alpha}{2};n-k-1}|\hat{\sigma}\cdot\sqrt{\mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*}\right)$$

Izrazimo razliko med pravo napovedjo in njeno oceno ter varianco te razlike:

$$y_* - \hat{y}_* = \mathbf{x}_*^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_* - \mathbf{x}_*^{\mathrm{T}} \mathbf{b}$$

Velja $\mathbb{E}(y_* - \hat{y}_*) = 0$ in ε_* in $\mathbf{b}$ sta neodvisna.

$$\begin{aligned} Var(y_* - \hat{y}_*) &= Var(\mathbf{x}_*^{\mathrm{T}} \mathbf{b}) + Var(\varepsilon_*) \\ &= \sigma^2 \mathbf{x}_*^{\mathrm{T}} \mathbf{A} \mathbf{x}_* + \sigma^2 \\ &= \sigma^2 (1 + \mathbf{x}_*^{\mathrm{T}} \mathbf{A} \mathbf{x}_*) \end{aligned}$$

Tudi tu lahko pokažemo, da je

$$\frac{y_* - \hat{y}_*}{\hat{\sigma}\sqrt{1 + \mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*}} \sim t_{n-k-1}$$

in $(1-\alpha)100\%$ interval zaupanja za posamično napoved je

$$\left(\hat{y}_* - |t_{\frac{\alpha}{2};n-k-1}|\hat{\sigma}\sqrt{1 + \mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*},\ \ \hat{y}_* + |t_{\frac{\alpha}{2};n-k-1}|\hat{\sigma}\sqrt{1 + \mathbf{x}_*^{\mathrm{T}}\mathbf{A}\mathbf{x}_*}\right)$$

Interpretacijo ocen parametrov linearnega modela z več regresorji si poglejmo najprej na primeru dveh številskih regresorjev x_1 in x_2. Zamislimo si pričakovano vrednost tega modela v točki (x_{01}, x_{02}).

$$\mathbb{E}(y|x_{01}, x_{02}) = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02}$$

in v točki $(x_{01}, x_{02} + 1)$

$$\mathbb{E}(y|x_{01}, x_{02}+1) = \beta_0+\beta_1 x_{01}+\beta_2(x_{02}+1) = \beta_0+\beta_1 x_{01}+\beta_2 x_{02}+\beta_2,$$

sledi

$$\mathbb{E}(y|x_{01}, x_{02} + 1) - \mathbb{E}(y|x_{01}, x_{02}) = \beta_2.$$

Torej velja, če x_{02} povečamo za eno enoto in ostane izbrana vrednost x_{01} nespremenjena, se pričakovana vrednost y poveča za β_2.

V linearnem modelu z več regresorji **ima vsak regresor "pogojni vpliv"**: če regresor x_j povečamo za eno enoto, se pogojno na konstantne vrednosti vseh ostalih regresorjev v modelu pričakovana vrednost odzivne spremenljivke poveča za β_j enot.

Pogojni vpliv regresorja x_j v modelu z več regresorji je lahko zelo drugačen, kot je njegov "robni" vpliv na odzivno spremenljivko, ko je x_j edini regresor v modelu. Prisotnost ostalih regresorjev lahko povzroči spremembo velikosti, lahko pa tudi spremembo predznaka parametra β_j.

Geometrijska predstavitev enostavne linearne regresije je **premica** v dvodimenzionalnem prostoru, za model z dvema regresorjema je **ravnina** v tridimenzionalnem prostoru, za model s k regresorji pa je to **hiper ravnina** v $k+1$ dimenzionalnem prostoru.

V regresijski analizi pogosto modeliramo vpliv izbrane spremenljivke na odzivno spremenljivko ob upoštevanju (*controlling for*) določenih t. i. **motečih spremenljivk** (*confounding variables*) v modelu. Zanima nas vpliv te izbrane napovedne spremenljivke, vendar vemo, da je odzivna spremenljivka odvisna tudi od nekaterih drugih spremenljivk, ki pa niso predmet naše raziskave.

`Model.1` je **gnezden znotraj** `Model.2` če velja:

`Model.1`

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i,$$

`Model.2`

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \beta_{k+1} x_{i(k+1)} + ... + \beta_{k+r} x_{i(k+r)} + \varepsilon_i.$$

Zanima nas, ali sta taka modela ekvivalentna oziroma ali je model z več členi v statističnem smislu boljši.

H_0: `Model.1` in `Model.2` sta ekvivalentna:

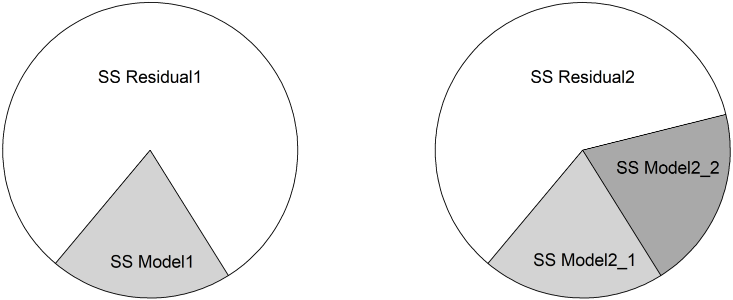
$$H_0 : \beta_{k+1} = \beta_{k+2} = ... = \beta_{k+r} = 0.$$

H_1: `Model.2` je boljši kot `Model.1`:

$$H_1 : \text{vsaj en} \quad \beta_{k+j} \neq 0, \quad j = 1, ..., r.$$

Linearni model v matrični obliki

F-test za primerjavo gnezdenih modelov



Statistično sklepanje temelji na F-statistiki:

$$F = \frac{\frac{SS_{residual1} - SS_{residual2}}{df_{residual1} - df_{residual2}}}{\frac{SS_{residual2}}{df_{residual2}}} \sim F_{df_{residual1} - df_{residual2}, df_{residual2}}$$

Če se modela razlikujeta za r parametrov

$$F = \frac{\frac{SS_{residual1} - SS_{residual2}}{r}}{\frac{SS_{residual2}}{n - k - r - 1}}.$$

R: funkcija `anova(model1, model2)`, prvi model je gnezdeni model. Oba modela morata biti narejena na istih podatkih.

Funkcija `anova` na `lm` modelu z več napovednimi spremenljivkami, vrne **sekvenčne vsote kvadratov ostankov modela** in rezultate **sekvenčnih F-testov**.

Sekvenčni F-test testira vpliv posamezne spremenljivke ob upoštevanju predhodnih spremenljivk v modelu.

Kaj se testira v posamezni vrstici izpisa?

- Prva vrstica: vsota kvadratov ostankov za model $y_i = \beta_0 + \beta_1 x_{1i}$, označimo jo $SS_{\beta_1|\beta_0}$.

  Z F-testom testiramo:

  $H_0 : \beta_1 = 0$

  H_0 : modela $y_i = \beta_0$ in $y_i = \beta_0 + \beta_1 x_{1i}$ sta ekvivalentna.

- Druga vrstica: razlika vsot kvadratov ostankov za model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ in za model $y_i = \beta_0 + \beta_1 x_{1i}$, označimo jo $SS_{\beta_2|\beta_0,\beta_1}$.

  $H_0 : \beta_2 = 0$ ob upoštevanju β_1 v modelu.

  H_0 : modela $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ in $y_i = \beta_0 + \beta_1 x_{1i}$ sta ekvivalentna.

- Če je v modelu k napovednih spremenljivk, se izpiše k vrstic z razlikami vsot kvadratov ostankov:

$$SS_{\beta_1|\beta_0}$$
$$SS_{\beta_2|\beta_0,\beta_1}$$
$$\ldots$$
$$SS_{\beta_k|\beta_0,\ldots,\beta_{k-1}}$$

- v i-ti vrstici se izvede F-test na podlagi F-statistike, $i = 1, \ldots, k$:

$$\frac{SS_{\beta_i|\beta_0,\ldots,\beta_{i-1}}}{SS_{\beta_{i-1}|\beta_0,\ldots,\beta_{i-2}}/(n-i-1)} \sim F_{1,n-i-1}.$$

  Testira se ničelna domneva $H_0 : \beta_i = 0$ ob upoštevanju $i-1$ napovednih spremenljivk v modelu.

- Če je napovedna spremenljivka opisna z d različnimi vrednostmi, se v modelu ocenjuje $d-1$ parametrov in z F-testom testiramo ničelno domnevo, da je vseh $d-1$ parametrov enakih 0:

$$\frac{SS_{\beta_i,\ldots,\beta_{i+d-1}|\beta_0,\ldots,\beta_{i-1}}}{SS_{\beta_{i-1}|\beta_0,\ldots,\beta_{i-2}}/(n-i-d-1)} \sim F_{d-1,n-i-d-1}.$$

Izpis funkcije `anova()` za linearni model je odvisen od vrstnega reda napovednih spremenljivk v modelu.