01RAD - přednáška 10, 12.11.2024

4.4 Transformace

- pokud není splněn některý z předpokladů modelu - linearita, normalita chyb, homoskedasticita
- jednou z možností je pokusit se transformovat proměnné, aby transformovaný model předpoklady alespoň „přibližně“ splňoval

A) Transformace vysvětlované proměnné y

hledáme funkci $h(\cdot)$ tak, aby model $Y_i^* = h(Y_i) = \beta_0 + \sum_{j=1}^m x_{ij}\beta_j + e_j$ předpoklady splňoval

3 hlavní důvody pro transformaci y:

1) transformace škály měření tak, aby pokrývala celé $\boldsymbol{R}$
   (což může odstranit problémy spodmínkami na β)
   - např. studie kapacity plic (FEV data, FEV > 0), chtěli bychom, aby model nepredikoval záporné hodnoty
   - lze obejít modelováním $y^* = \ln FEV$
   - pokud y jsou počty a 0 je možná hodnota, často se používá $y^* = \ln(y+1)$ nebo obecně $y^* = \ln(y+c)$

2) transformace Y, aby její rozdělení bylo „více“ normální

- typicky se snažíme udělat rozdělení hodnot y více symetrické
- často se setkáváme s rozděleními vychýlenými vpravo
- transformace $y^* = \ln y$ nebo $y^* = y^\lambda$, $\lambda < 1$ budou redukovat toto vychýlení
- **typický postup:** začít s hodnotou $\lambda \approx 1$ a pak snižovat, dokud není dosaženo „přibližné“ symetrie reziduí

3) možná nejzásadnější motivace: pokusit se dosáhnout konstantní rozptyl

- např. pro fyzikální veličinu s kladnými hodnotami se často stane, že rozptyl bude malý pro $\mu \approx 0$ a větší pro μ velké (tzv. positive mean-variance relationship)
- nepřesnost měření kladných veličin se často vyjadřuje pomocí koeficientu variace $CV(Y) = \dfrac{s.d.Y}{\mathsf{E}Y}$
- bývá více konstantní než s.d. (variabilitu vyjadřujeme relativně)
- matematicky to znamená, že $\mathrm{Var}\,Y = \varphi(\mathsf{E}Y)^2 = \varphi\mu^2$ pro nějaké φ
- pro odstranění vztahu mezi $\mathsf{E}Y$ a $\mathrm{Var}\,Y$ se často používají transformace $y^* = y^\lambda$ (pro $y > 0$)

Transformace:	$\longleftarrow$	$\cdots$	y^3	y^2	y	$\sqrt{y}$	$\ln y$	$\frac{1}{\sqrt{y}}$	$\frac{1}{y}$	$\frac{1}{y^2}$	$\cdots$	$\longrightarrow$
Box-Cox λ:			3	2	1	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1	-2		

- pokud $\mathrm{Var}\,Y$ klesá s rostoucí $\mathrm{E}\,Y$
- pokud $\mathrm{Var}\,Y$ roste s rostoucí $\mathrm{E}\,Y$

OBECNĚ: předpokládejme vztah $\mathrm{Var}\,Y = \varphi V(\mu)$ a uvažujme transformaci $y^* = h(y)$

Taylorův rozvoj 1. řádu funkce $h(y)$ v bodě μ

$$y^* = h(y) \approx h(\mu) + h'(\mu)(y - \mu)$$

z čehož plyne, že $\mathrm{Var}\,Y^* \approx (h'(\mu))^2 \cdot \mathrm{Var}\,Y$

transformace $y^* = h(y)$ bude přibližně stabilizovat rozptyl, pokud $h'(\mu)$ je úměrné $(\mathrm{Var}\,Y)^{-\frac{1}{2}} = V^{-\frac{1}{2}}(\mu)$

- pokud $V(\mu) = \mu^2$ $\Rightarrow$ transformace stabilizující rozptyl je $h(y) = \ln y$ $\left(h'(\mu) = \frac{1}{\mu}\right)$
- pokud $V(\mu) = \mu$ $\Rightarrow$ transformace stabilizující rozptyl je $h(y) = \sqrt{y}$ $\left(h'(\mu) = \frac{1}{2\sqrt{\mu}}\right)$

- asi nejvíce užívanou transformací je $y^* = \ln y$ (dobrá interpretovatelnost parametrů β)

Poznámka 4.11 (Interpretace parametrů lineárního modelu)

a) klasický LM: $\mathrm{E}Y = \beta_0 + \beta_1 x_1 + \cdots \beta_m x_m$

jednotková změna proměnné x_j $\Rightarrow$ změnu $\mathrm{E}Y$ o β_j jednotek
(při ostatních proměnných stejných)

b) LM pro $\ln Y$: $\ln Y = \beta_0 + \beta_1 x_1 + \cdots \beta_m x_m + e, \quad e \sim N(0, \sigma^2)$

- pokud je to správný model, $\ln Y \sim N(\mu, \sigma^2)$ $\Rightarrow$ $Y \sim LN(\mu, \sigma^2)$ a $\mathrm{E}Y = e^{\mu + \frac{\sigma^2}{2}}$

  predikce pro $\mathrm{E} \ln Y$ je $\widehat{\mu} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots \widehat{\beta}_m x_m$

  predikce pro $\mathrm{E}Y$ je $e^{\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots \widehat{\beta}_m x_m + \frac{\widehat{\sigma^2}}{2}}$

- uvažujme opět jednotkovou změnu proměnné x_j ($x_j \to x_j + 1$)

$$\frac{\mathrm{E}Y_{new}}{\mathrm{E}Y} = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots \beta_j x_j + \beta_j + \cdots \beta_m x_m + \frac{\sigma^2}{2}}}{e^{\beta_0 + \beta_1 x_1 + \cdots \beta_m x_m + \frac{\sigma^2}{2}}} = e^{\beta_j}$$

  jednotková změna proměnné x_j $\Rightarrow$ multiplikativní změnu $\mathrm{E}Y$ e^{β_j}-krát

  jinak zapsáno: $100(e^{\beta_j} - 1)$ je procentní změna $\mathrm{E}Y$ spojená s jednotkovou změnou x_j

Box-Cox transformace

- pokud chyby nemají normální rozdělení, hledáme transformaci Y, která by nejenom linearizovala model, ale také transformovala chyby, aby byly přibližně normální
- užitečná třída transformací (power family)

$$y^{(\lambda)} = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \text{pokud } \lambda \neq 0 \\ \ln y, & \text{pokud } \lambda = 0 \end{cases}, \qquad \left(\lim_{\lambda \to 0} \frac{y^\lambda - 1}{\lambda} = \ln y \right)$$

  která předpokládá, že data y jsou kladná
- pro nalezení vhodného λ budeme předpokládat, že transformované veličiny $Y_i^{(\lambda)}$, $i = 1, \ldots, n$, splňují podmínky RM, tj.

$$Y_i^{(\lambda)} = \boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{e} \quad \text{kde} \quad \boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n) \qquad (Y_i^{(\lambda)} \sim N(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2),\ Y_i^{(\lambda)} \text{nezávislé})$$

- úkol je odhadnout zároveň $\lambda, \boldsymbol{\beta}, \sigma^2$, použijeme MLE

Dostali jsme tedy tzv. profile log - likelihood

$$\ell_p^{(\lambda)} = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\widehat{\sigma}^2(\lambda) - \frac{n}{2} + \ln \mathrm{J}(\lambda) = C - \frac{n}{2}\ln\widehat{\sigma}^2(\lambda) + (\lambda - 1)\sum_{i=1}^{n}\ln y_i,$$

kde

$$\widehat{\beta}(\lambda) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}^{(\lambda)}, \qquad \widehat{\sigma}^2(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i^{(\lambda)} - \widehat{y}_i^{(\lambda)})^2,$$

a

$$\boldsymbol{y}^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_n^{(\lambda)})^T, \qquad \widehat{y}_i^{(\lambda)} = \boldsymbol{x}_i^T\widehat{\beta}^{(\lambda)}.$$

POZNÁMKA:

- kvůli komplikované závislosti $\ell_p^{(\lambda)}$ na λ bude třeba numerická metoda pro maximalizaci
- lze přepsat do tvaru, kde bude možné využít metody LR

Celkem: $\max_{\lambda} \ell_p(\lambda) \iff \min_{\lambda} s_\lambda^2$,

kde

$$s_\lambda^2 = \frac{\widehat{\sigma}^2(\lambda)}{[(\dot{y})^{\lambda-1}]^2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i^{(\lambda)}}{(\dot{y})^{\lambda-1}} - \frac{\widehat{y}_i^{(\lambda)}}{(\dot{y})^{\lambda-1}}\right)^2 ,$$

tzn. s_λ^2 je reziduální součet čtverců (SSE/n) v modelu $\frac{y_i^{(\lambda)}}{(\dot{y})^{\lambda-1}}$ v závislosti na $\boldsymbol{x}_i^T$
(tzn. s_λ^2 lze snadno získat pomocí funkce lm( ))

Algoritmus:

1) zvolit oblast hodnot λ, $I = \langle \lambda_{min}, \lambda_{max} \rangle$, a body $\lambda \in I$
(typicky $I = \langle -2, 2 \rangle$ a 10-20 rovnoměrně rozdělených bodů)
2) naladit model $\frac{y^{(\lambda)}}{(\dot{y})^{\lambda-1}} \sim x$ a spočítat $\frac{1}{n} SSE = s_\lambda^2$.
3) z grafu (λ, s_λ^2) vybrat $\widehat{\lambda}$, které minimalizuje s_λ^2
4) pro zvolené $\widehat{\lambda}$ naladit model $y^{(\widehat{\lambda})} \sim x$ a pokračovat standardní analýzou

Intervaly spolehlivosti pro λ:

- snadno lze odvodit LRT test pro test $H_0: \ \lambda = \lambda_0$
  ($H_0: \ \lambda = 1$ zda je třeba transformace, pokud zamítneme $H_0 \Rightarrow$ transformace pomocí $\widehat{\lambda}$)
- LRT statistika: $\Lambda = -2\ln\frac{L(\lambda_0)}{L(\widehat{\lambda})} = 2(\ell_p(\widehat{\lambda}) - \ell_p(\lambda_0))$, víme že $\Lambda \xrightarrow{L} \chi^2(1)$
- invertováním přípustné oblasti LRT testu, dostaneme as. $100(1-\alpha)\%$ IS pro λ

$$\left\{\lambda \in \mathbb{R} \,\middle|\, n \cdot \ln\frac{s^2_\lambda}{s^2_{\widehat{\lambda}}} \leq \chi^2_{1-\alpha}(1)\right\}, \qquad \text{kde} \quad \widehat{\lambda} \text{ je MLE } \lambda$$

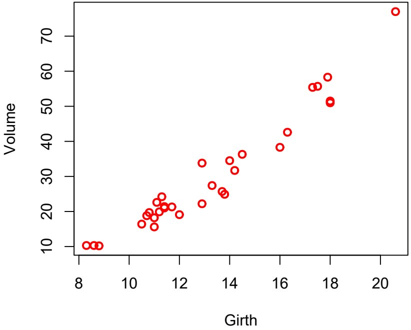
Poznámka 4.12

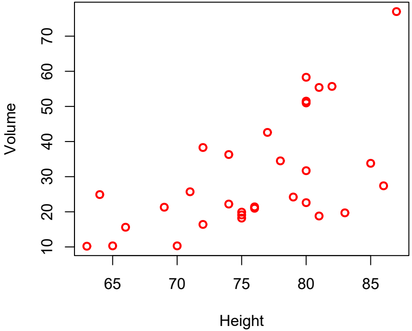
Kvůli jednoduchosti interpretace se často doporučuje zaokrouhlit $\widehat{\lambda}$ na nejbližší $\dfrac{1}{4}$ nebo $\dfrac{1}{3}$.

Příklad 4.4 (Data TREES)

	Girth	Height	Volume
1	8	70	10
2	9	65	10
3	9	63	10
4	10	72	16
⋮	⋮	⋮	⋮

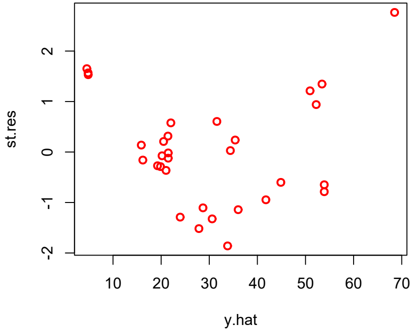
```
mod <- lm(Volume ~ Girth + Height)
summary(mod)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -57.9877     8.6382  -6.713 2.75e-07 ***
## Girth         4.7082     0.2643  17.816  < 2e-16 ***
## Height        0.3393     0.1302   2.607   0.0145 *
## ---
## Residual standard error: 3.882 on 28 degrees of freedom
## Multiple R-squared:  0.948,       Adjusted R-squared:  0.9442
## F-statistic:   255 on 2 and 28 DF,  p-value: < 2.2e-16
```

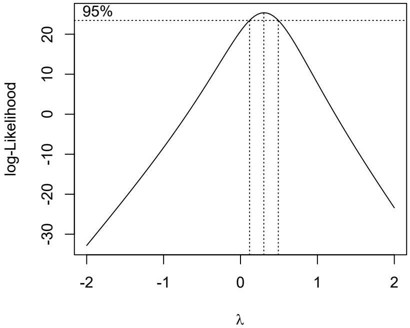




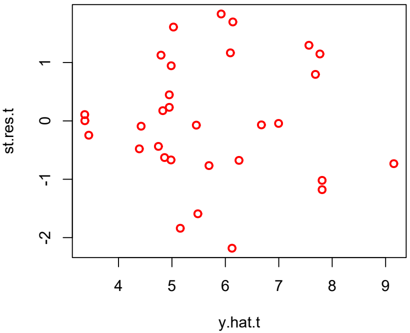
```
y.hat <- predict(mod)
st.res <- rstudent(mod)
plot(y.hat, st.res, pch = c(1), col = c("red"), lwd=2)

bc<- boxcox(mod, lambda = seq(-2,2,1/10))

lambda.hat <- bc$x[which.max(bc$y)]; lambda.hat
##  0.3030303

Vol.lambda <- (Volume^lambda.hat-1)/lambda.hat

mod.t <- lm(Vol.lambda ~ Girth + Height)
summary(mod.t)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.733542   0.500080  -5.466 7.77e-06 ***
## Girth        0.409448   0.015299  26.764  < 2e-16 ***
## Height       0.039685   0.007535   5.267 1.34e-05 ***
## ---
## Residual standard error: 0.2247 on 28 degrees of freedom
## Multiple R-squared: 0.9775, Adjusted R-squared: 0.9759
## F-statistic: 609.6 on 2 and 28 DF,  p-value: < 2.2e-16
```





```
y.hat.t <- predict(mod.t)
st.res.t <- rstudent(mod.t)

plot(y.hat.t, st.res.t, pch = c(1), col = c("red"), lwd=2)
```



B) Transformace vysvětlujících proměnných x

- pokud diagnostika modelu naznačuje, že vztah mezi $\boldsymbol{y}$ a $\boldsymbol{X}$ není lineární pro jeden nebo více regresorů, může být vhodné přeformulovat model pomocí transformací proměnných $\boldsymbol{x}$
- předpokládejme, že v modelu $Y = \beta_0 + \sum_{j=1}^{m} \beta_j x_j + e$ máme podezření na nelinearitu v j-té proměnné x_j
- jednou z možností jak postupovat je nahrazení x_j proměnnou $z_j = f(x_j)$, model tedy bude

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j z_j + \cdots + \beta_m x_m + e$$

- pokud je f známé, jedná se o model LR a lze ho analyzovat standardně,

  pokud je tato transformace vhodná, mělo by se to projevit ve zlepšení statistik $\mathrm{R}^2, \mathrm{t}, \mathrm{F}$ a zlepšení grafu reziduí pro z_j oproti těm pro x_j
- bohužel f většinou známá není, možný přístup je parametrizovat nějak tuto funkci a pak odhadnout tyto parametry společně s $\boldsymbol{\beta}$
  - typická parametrizace: $z_j = x_j^{\lambda}$, kde $\lambda \in \mathbb{R}$ vhodné
  - aproximace f pomocí polynomu vhodného stupně, tzn. $z_j = \sum_{k=1}^{l} r_k x_j^k$, kde r_k musí být odhadnuty
  - další možností je použití trigonometrických funkcí nebo splines (piecewise polynomials)

Zaměříme se na $z_j = x_j^\lambda$:

- možnost je opět zvolit jistou množinu hodnot λ, naladit modely pro všechna λ a vybrat model s nejlepší shodou s daty, např. s nejmenší SSE nebo největší R^2 nebo F
- může být časově náročné, můžeme minout vhodnou hodnotu λ, pokud nebyla v původní množině (nevíme jak R^2, F, SSE závisí na λ)

Box-Tidwell metoda

- předpokládejme, že λ se příliš neliší od $\lambda = 1$, Taylorův rozvoj 1. řádu kolem $\lambda = 1$ dává

$$x^\lambda \approx x^1 + (\lambda - 1)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\lambda}\bigg|_{\lambda=1}, \quad \frac{\mathrm{d}x^\lambda}{\mathrm{d}\lambda}\bigg|_{\lambda=1} = x^\lambda \ln x\Big|_{\lambda=1} = x \ln x, \quad \text{tedy} \quad x^\lambda \approx x + (\lambda - 1)x \ln x$$

- dosazením do modelu

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \ldots + \beta_{j-1}x_{j-1} + \beta_j\big(x_j + (\lambda - 1)x_j \ln x_j\big) + \ldots + \beta_m x_m + e \\ &= \beta_0 + \sum_{k=1}^{m} \beta_k x_k + \underbrace{\beta_j(\lambda - 1)}_{\beta_{m+1}(\lambda)} x_j \ln x_j + e \end{aligned}$$

- máme lineární model pro parametry β_k, $0 \le k \le m+1$, protože $\beta_{m+1} = (\lambda - 1)\beta_j$, můžeme (λ, β_j) odhadnout následovně

1) naladíme původní model a spočteme *LSE* $\widehat{\beta}_j$ parametru β_j

2) naladíme rozšířený model s $x_{m+1} = x_j \ln x_j$ a spočteme $\widehat{\beta}_{m+1}$

3) z rovnosti $\widehat{\beta}_{m+1} = (\widehat{\lambda} - 1)\widehat{\beta}_j$ dostaneme $\widehat{\lambda} = \frac{\widehat{\beta}_{m+1}}{\widehat{\beta}_j} + 1$

- tento postup umožňuje testovat potřebu transformace $H_0 : \lambda = 1 \quad \times \quad H_1 : \lambda \neq 1$

  pomocí t-testu pro $H_0 : \ \beta_{m+1} = 0$

Poznámka 4.13

- pokud model s $\widehat{\lambda}$ vypadá neadekvátně, lze postupovat iterativně a získat posloupnost $\widehat{\lambda}(l), \ l \geq 1$
- $\widehat{\lambda}(0) = \widehat{\lambda}$ a rozvineme x_j^λ kolem $\widehat{\lambda}(0)$, tzn. $x_j^\lambda \approx x_j^{\widehat{\lambda}(0)} + \big(\lambda - \widehat{\lambda}(0)\big) x_j^{\widehat{\lambda}(0)} \ln x_j$
- dosazením do rovnice modelu

$$Y = \beta_0 = \sum_{\substack{k=1 \\ k \neq j}}^{m} \beta_k x_k + \beta_j x_j^{\widehat{\lambda}(0)} + \underbrace{\beta_j \big(\lambda - \widehat{\lambda}(0)\big)}_{\beta_{m+1}} x_j^{\widehat{\lambda}(0)} \ln x_j + e$$

- naladíme tento model s a bez přidané proměnné $x_{m+1} = x_j^{\widehat{\lambda}(0)} \ln x_j$, označíme $\widehat{\beta}_j(1)$ a $\widehat{\beta}_{m+1}(1)$ příslušné odhady

- potom

$$\widehat{\lambda}(1) = \widehat{\lambda}(0) + \frac{\widehat{\beta}_{m+1}(1)}{\widehat{\beta}_j(1)}$$

- můžeme dále iterovat do konvergence nebo skončit po pevném počtu iterací

Poznámka 4.14 (Další užívané transformace v $\boldsymbol{x}$, $\boldsymbol{y}$)

a) centrované proměnné: $\boldsymbol{X}_C$ a $\boldsymbol{y}_C$

$(\boldsymbol{X}_C)_{ij} = x_{ij} - \overline{x}_j, \; i \in \hat{n}, \; j \in \hat{m}$, kde $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad (\boldsymbol{y}_C)_i = y_i - \overline{y}$

1) $\widehat{\beta}_1, \ldots, \widehat{\beta}_m$ je řešením $\boldsymbol{X}_C^T \boldsymbol{X}_C \beta = \boldsymbol{X}_C^T \boldsymbol{y}_C$, 2) $\widehat{\beta}_0 = \overline{y} - \sum_{j=1}^{m} \widehat{\beta}_j \overline{x}_j$

b) centrované a škálované proměnné $\boldsymbol{X}_{SC}$ a $\boldsymbol{y}_C$ ($\boldsymbol{y}_{SC}$):

škálování sloupců tak, aby jejich norma byla 1,

tzn. každý prvek j-tého sloupce matice $\boldsymbol{X}$ podělíme $s_j = \left(\sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2 \right)^{\frac{1}{2}}$

centrovaná a škálovaná matice $\boldsymbol{X}_{SC}$ pak bude $\boldsymbol{X}_{SC} = \boldsymbol{X}_C \boldsymbol{S}, \quad \boldsymbol{S} = \mathrm{diag}\left(\frac{1}{s_1}, \ldots, \frac{1}{s_m}\right)$

model bude $\boldsymbol{Y}_C = \boldsymbol{X}_{SC} \beta_s + e$

(lze použít i $\boldsymbol{Y}_{SC}$, tedy centrované a škálované $\boldsymbol{Y}$)

4.5 Vážené nejmenší čtverce (weighted least squares - WLS)

- budeme předpokládat, že chyby e_i jsou normální, nezávislé, ale $\mathrm{Var}(e_i) = \sigma_i^2$ závisí na i
- konkrétně $\sigma_i^2 = \dfrac{\sigma^2}{w_i}$, kde $w_i > 0$, $i \in \hat{n}$, se nazývají váhy
- uvažujeme tedy model

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}, \quad \text{kde} \quad \boldsymbol{e} \sim N_n(0, \sigma^2 \boldsymbol{W}) \quad \text{a} \quad \boldsymbol{W} = \mathrm{diag}\left(\frac{1}{w_1}, \dots, \frac{1}{w_n}\right) \tag{4.2}$$

- pokud jsou váhy w_i známé, lze MLE odhady parametru β a σ^2 nalézt následovně

$$\boldsymbol{W} = \boldsymbol{K}\boldsymbol{K}^T, \text{ kde } \boldsymbol{K} = \boldsymbol{W}^{\frac{1}{2}} = \mathrm{diag}\left(\frac{1}{\sqrt{w_1}}, \dots, \frac{1}{\sqrt{w_n}}\right)$$

- definujeme $\boldsymbol{Z} = \boldsymbol{K}^{-1}\boldsymbol{Y}$, $\boldsymbol{M} = \boldsymbol{K}^{-1}\boldsymbol{X}$ a $\varepsilon = \boldsymbol{K}^{-1}\boldsymbol{e}$
- dostaneme model

$$\boldsymbol{Z} = \boldsymbol{M}\beta + \varepsilon, \quad \text{kde} \quad \varepsilon \sim N_n(0, \sigma^2 \boldsymbol{I}_n), \tag{4.3}$$

(protože $\mathrm{Cov}(\varepsilon) = \boldsymbol{K}^{-1}\sigma^2\boldsymbol{W}\left(\boldsymbol{K}^{-1}\right)^T = \sigma^2\boldsymbol{K}^{-1}\boldsymbol{K}\boldsymbol{K}^T\left(\boldsymbol{K}^T\right)^{-1} = \sigma^2\boldsymbol{I}_n$)

- to už je standardní model LR, ve kterém platí

$$\begin{aligned}
\widehat{\beta}_w &= (\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{z} = \left(\boldsymbol{X}^T(\boldsymbol{K}^{-1})^T\boldsymbol{K}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T(\boldsymbol{K}^{-1})^T\boldsymbol{K}^{-1}\boldsymbol{y} \\
&= (\boldsymbol{X}^T\boldsymbol{W}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}^{-1}\boldsymbol{y} \\
\widehat{\sigma}_w^2 &= \frac{1}{n}\sum_{i=1}^{n}(z_i - \widehat{z}_i)^2 = \frac{1}{n}\sum_{i=1}^{n} w_i(y_i - \widehat{y}_i)^2 = \frac{1}{n}SSE_w,
\end{aligned}$$

kde SSE_w je vážený součet čtverců, $\quad z_i = \sqrt{w_i}\, y_i$ a $\widehat{z}_i = \sqrt{w_i}\,\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}_w = \sqrt{w_i}\,\widehat{y}_i$

- dále platí

a) $\mathrm{E}\widehat{\boldsymbol{\beta}}_w = (\boldsymbol{X}^T\boldsymbol{W}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}^{-1}\mathrm{E}\boldsymbol{Y} = \boldsymbol{\beta}$, tzn. $\widehat{\boldsymbol{\beta}}_w$ je nestranný odhad $\boldsymbol{\beta}$

b) $\mathrm{E}\left(\frac{SSE_w}{n-m-1}\right) = \sigma^2$, tedy $s_w^2 = \dfrac{SSE_w}{n-m-1}$ je nestranný odhad σ^2

Věta 4.4

Nechť $\widehat{\boldsymbol{\beta}}_w$ je WLS odhad $\boldsymbol{\beta}$, jestliže $\operatorname{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{W} = \sigma^2 \operatorname{diag}\left(\frac{1}{w_1}, \dots, \frac{1}{w_n}\right)$. Potom platí:

1) $\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_w) = \sigma^2 (\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1}$
2) nechť δ_i je i-tý diagonální prvek $(\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1}$, jestliže $e_i \sim N\left(0, \frac{\sigma^2}{w_i}\right)$, $i \in \hat{n}$, potom

$$T_i = \frac{\widehat{\beta}_{w,i} - \beta_i}{s_w \sqrt{\delta_i}} \sim t(n-m-1)$$

3) pro $\widehat{\boldsymbol{Y}}_w = \boldsymbol{X} \widehat{\boldsymbol{\beta}}_w$ platí $\mathrm{E} \widehat{\boldsymbol{Y}}_w = \boldsymbol{X} \boldsymbol{\beta}$ a $\operatorname{Cov}(\widehat{\boldsymbol{Y}}_w) = \sigma^2 \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T$
4) Nechť $\widehat{\boldsymbol{e}}_w = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}_w$ jsou rezidua v modelu (4.2) a $\widehat{\boldsymbol{\varepsilon}}_w = \boldsymbol{Z} - \widehat{\boldsymbol{Z}} = \boldsymbol{Z} - \boldsymbol{M} \widehat{\boldsymbol{\beta}}_w$ jsou rezidua v transformovaném modelu (4.3). Potom

$$\widehat{\boldsymbol{\varepsilon}}_w = \sqrt{\boldsymbol{W}^{-1}} \widehat{\boldsymbol{e}}_w = \boldsymbol{W}^{-\frac{1}{2}} \widehat{\boldsymbol{e}}_w \quad \text{a} \quad \mathrm{E}(\widehat{\boldsymbol{e}}_w) = \mathrm{E}(\widehat{\boldsymbol{\varepsilon}}_w) = \boldsymbol{0}.$$

5) nechť $\boldsymbol{H}_w = \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}^{-1}$ je vážená projekční matice, potom

$$\widehat{\boldsymbol{e}}_w = (\boldsymbol{I} - \boldsymbol{H}_w) \boldsymbol{e}, \quad \operatorname{Cov}(\widehat{\boldsymbol{e}}_w) = \sigma^2 (\boldsymbol{I} - \boldsymbol{H}_w) \boldsymbol{W}, \quad \text{tzn.} \quad \operatorname{Cov}(\widehat{\boldsymbol{\varepsilon}}_w) = \sigma^2 \boldsymbol{W}^{-\frac{1}{2}} (\boldsymbol{I} - \boldsymbol{H}_w) \boldsymbol{W}^{\frac{1}{2}}$$

Důkaz.

- odhady parametrů $\boldsymbol{\beta}$ a σ^2 lze získat použitím transformovaného modelu (4.3)
- protože ale transformovaný model neobsahuje intercept (první sloupec $\boldsymbol{M}$ je $(\sqrt{w_1}, \ldots, \sqrt{w_n})^T$)), nefunguje klasický rozklad součtu čtverců a F a R^2 statistiku nelze definovat obvyklým způsobem (viz. regrese skrz počátek)
- nicméně princip „extra sum of squares" funguje, ať má model intercept nebo ne
  - např. celkový F-test lze provést pomocí statistiky

$$F_w = \frac{\frac{SSE_R - SSE_F}{m}}{s_w^2},$$

  kde SSE_F je reziduální součet čtverců s_w^2 plného modelu a SSE_R je reziduální součet čtverců redukovaného transformovaného modelu $\boldsymbol{Z} = \boldsymbol{M}_0\beta_0 + \boldsymbol{e}$, $\boldsymbol{M}_0 = (\sqrt{w_1}, \ldots, \sqrt{w_n})^T$
  - pokud mají chyby normální rozdělení, platí za $H_0 : \beta_1 = \cdots = \beta_m = 0$, že $F_w \sim F(m, n-m-1)$ a H_0 zamítáme, pokud $F_w > F_{1-\alpha}(m, n-m-1)$
- přirozené je definovat $R^2 = \varrho^2(\widehat{\boldsymbol{z}}, \boldsymbol{z})$, kde $\varrho(\widehat{\boldsymbol{z}}, \boldsymbol{z})$ je výběrový korelační koeficient (pro $\boldsymbol{W} = \boldsymbol{I}$ dostaneme standardní R^2)

Pro analýzu reziduí je třeba uvažovat vhodné grafy reziduí:

- máme dva vektory reziduí:

$$\widehat{e}_i \text{ v původním modelu (4.2)} \qquad \widehat{\varepsilon}_i \text{ v transformovaném modelu (4.3)}$$

  a tedy dvě možnosti
- pro kontrolu konstantního rozptylu lze uvažovat i standardizovaná nebo studentizovaná rezidua (pomocí bodu 4) a 5) věty lze ukázat, že jsou v obou modelech stejná)
- je třeba být opatrný oproti jakým hodnotám budeme rezidua zobrazovat
- grafy $\widehat{\varepsilon}_i$ proti sloupcům $\boldsymbol{M}$ a predikovaným hodnotám $\widehat{\boldsymbol{z}}$ jsou OK, neboť např.

$$\sum_{i=1}^{n} \widehat{z}_i \widehat{\varepsilon}_i = 0$$

  (jsou OG, měl by být vidět roztýlený oblak kolem osy x)
- dosazením $\widehat{\varepsilon}_i = \sqrt{w_i} \cdot \widehat{e}_i$ a $\widehat{z}_i = \sqrt{w_i} \cdot \widehat{y}_i$ dostaneme $\sum_{i=1}^{n} w_i \widehat{e}_i \widehat{y}_i = 0$, tzn. graf $\widehat{e}_i$ proti $\widehat{y}_i$ bude zavádějící
- graf $\sqrt{w_i} \cdot \widehat{e}_i$ proti $\sqrt{w_i} \cdot \widehat{y}_i$ je ale v pořádku
- podobné závěry platí i pro grafy $\widehat{e}_i$ proti $\boldsymbol{x}_j^c$, $j = 1, \ldots, m$

Poznámka 4.15

- pokud jsou váhy neznámé, bylo by třeba je odhadnou společně s β a σ^2 z dat
- to ale není obecně možné, protože máme více parametrů než dat
- někdy to možné je, pokud máme další informace o rozdělení chyb (tvar kovarianční matice atd.)

Poznámka 4.16

- celý postup WLS lze použít i na případ $\boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{W})$, kde $\boldsymbol{W}$ je známá, ale není diagonální
- protože $\boldsymbol{W}$ je PD, ex. regulární $\boldsymbol{K}$ tak, že $\boldsymbol{W} = \boldsymbol{K}\boldsymbol{K}^T$
- stejná transformace jako u WLS opět vede na transformovaný model, kde $\boldsymbol{\varepsilon} \sim N_m(\mathbf{0}, \sigma^2 \boldsymbol{I}_m)$