Řešení domácího úkolu

4ST416 Regrese

Lubomír Štěpánek

12. května 2017

Obsah

1 Zadání úlohy

Dataset pocházející ze zdroje [1] a dostupný na adrese

http://www1.aucegypt.edu/faculty/hadi/RABE5/Data5/P155b.txt

popisuje automobily různých značek (proměnná `Car`) tak, že pro každou značku je uvedena

- cena automobilu (kvantitativní proměnná `Y`),
- země původu automobilu (kategorická proměnná `Country` s třídami {`USA`, `Japan`, `Germany`, `Other`}),
- koňská síla automobilu (kvantitativní proměnná `Horsepower`).

V datasetu je celkem 90 pozorování. S využitím regresní analýzy se pokusme zodpovědět následující dvě otázky

(i) Je vztah mezi vysvětlovanou proměnnou (`Y`) a vysvětlujícími proměnnými (`Horsepower`, `Country`) aditivní, nebo neaditivní? Pokud je neaditivní, pokusme se popsat charakter této neaditivity.

(ii) Jak prakticky významná je — po očištění o koňskou sílu[1] (`Horsepower`) – závislost ceny automobilu (`Y`) na zemi původu (`Country`)?

2 Řešení úlohy

Celá úloha byla řešena v prostředí R, které je určeno pro statistické výpočty a následné grafické náhledy [2].

Nejdříve proveďme několik grafických náhledů nad daty pro účely jejich intuitivnějšího pochopení, poté zkusme formálnější analýzu pomocí regresních metod.

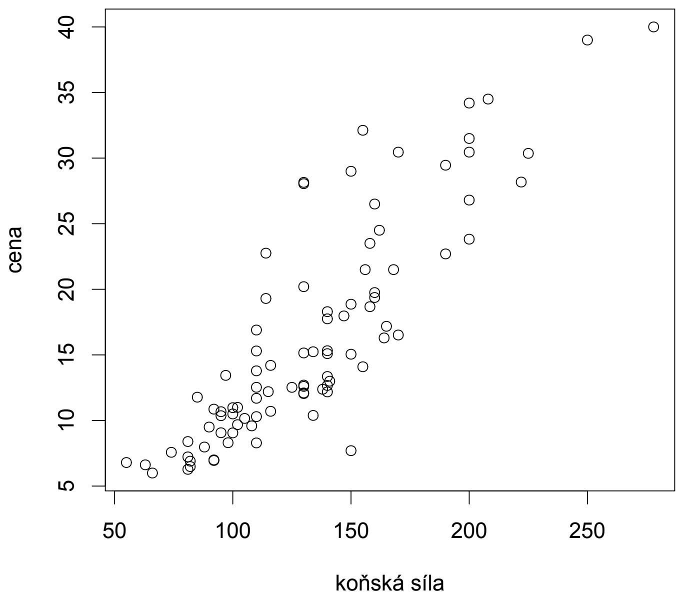
V rámci lepší čitelnosti a jazykové homogenity byly často používané proměnné v datasetu adekvátně přeloženy do češtiny, tedy cena automobilu (proměnná `Y`) je zobrazována jako `cena`, země původu jako (proměnná `Country`) jako `země původu` s kategoriemi {`USA`, `Japonsko`, `Německo`, `Ostatní`} a koňská síla (proměnná `Horsepower`) jako `koňská síla`[2], vždy eventuálně s podtržítkem mezi slovy ve víceslovném názvu.

[1] Pojmem *očištění o koňskou sílu* máme na mysli to, že nás zajímá závislost ceny automobilu na zemi původu při konstantní koňské síle.

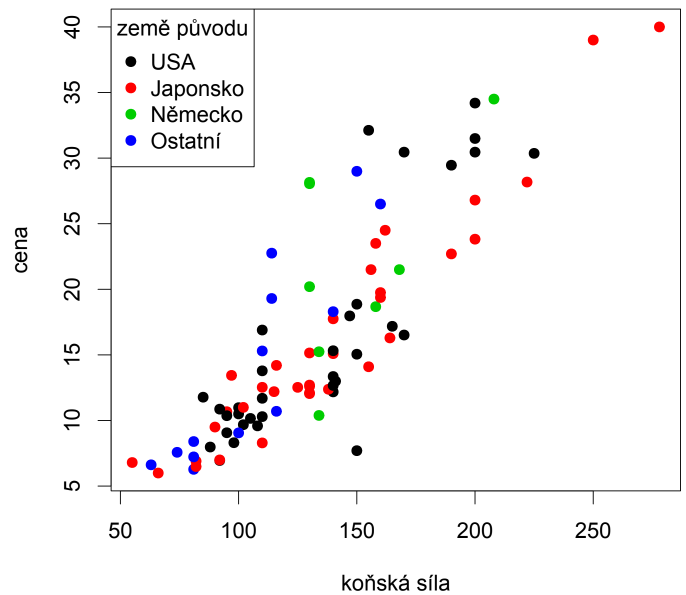
[2] Při dodržení správného kódování skriptů (UTF-8) nečiní diakritika v R problémy; při použití vhodných R-kových enginů (např. `cairo_ps()` podporujícího extended ASCII místo defaultního `setEPS()`) nedělá diakritika obtíže ani v tvorbě českých popisků diagramů.

2.1 Exploratory Data Analysis (EDA)

V rámci grafického prozkoumávání dat nejdříve nahlédněme obecně na závislost ceny automobilů (proměnná `cena`) na koňské síle (proměnná `koňská síla`), viz obrázky 1 a 2.



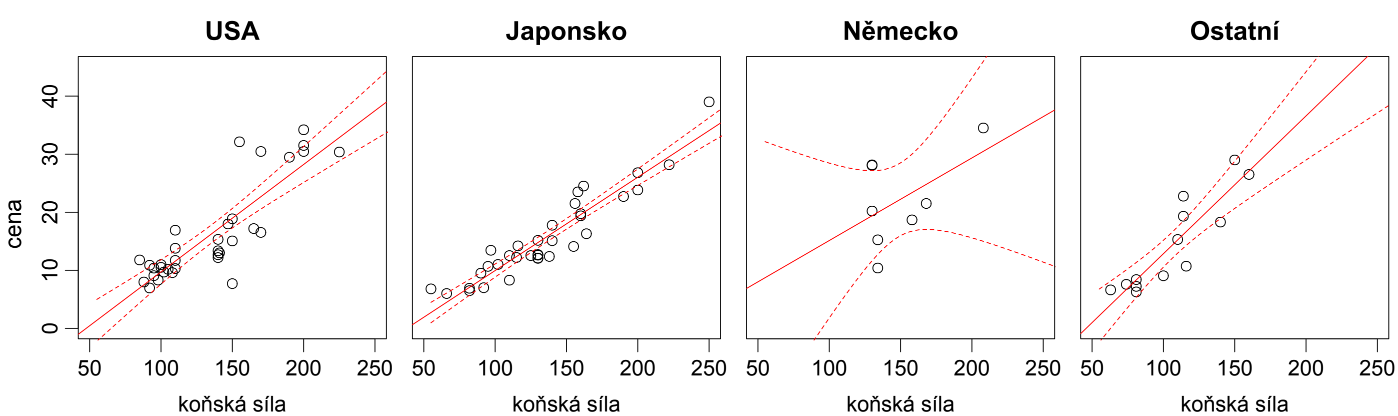
Obrázek 1: Závislost ceny automobilu na jeho koňské síle



Obrázek 2: Závislost ceny automobilu na jeho koňské síle s odlišením země původu automobilu

Na obrázku 1 vidíme, že cena automobilu relativně silně a relativně „jednoznačně“ roste s hodnotou koňské síly, velmi neformálně lze mluvit o *přímé úměrnosti* mezi oběma veličinami. Na obrázku 2 pak vidíme, že body [`koňská síla`, `cena`] nevykazují na první pohled pro žádný z jednotlivých států žádnou nápadnou odlišnost od obecně rostoucího trendu závislosti.

Ještě se podívejme na jednotlivé závislosti ceny automobilů na koňské síle vždy izolovaně pro každou zemi původu. Náhled vidíme na obrázku 3; osy dílčích diagramů jsou sjednoceny pro možná grafická porovnávání. Zdá se, že kvantitativně chápaná rychlost růstu ceny automobilu s růstem koňské síly je u vozů z u všech zemí původů velice podobná. U německých vozů je kvůli jejich malému počtu relativně široký 95% konfidenční pás, rovněž se zdá, že průsečík regresní přímky s vertikálou, kde `koňská síla` = 50, je o „něco“ výše než u automobilů ostatních zemí původu. Nelze tedy primárně říct, že by země původu automobilu nehrála v jeho ceně roli – může to tak být. Dle grafické analýzy se nezdá, že by byla přítomna odlehlá pozorování.



Obrázek 3: Ceny automobilů v závislosti na koňské síle izolovaně vždy pro každou zemi původu, zobrazena je i regresní přímka závislosti (plně červeně) a 95% konfidenční pás (přerušovaně červeně)

2.2 Regresní analýza

Předpokládejme, že předchozí grafická explorační analýza dat a některých závislostí nám dala určitou představu, která odpovídá apriorním očekáváním, tudíž že cena automobilu roste nějakým způsobem s rostoucím výkonem motoru (zde měřeným v koňské síle) a že země původu automobilu (uvažujeme-li současně závislost na koňské síle) může hrát v ceně automobilu nějakou roli.

(i) Lineární model budeme považovat za aditivní, pokud pro každou vysvětlující proměnnou je změna vysvětlované proměnné závislá pouze na této vysvětlující proměnné (a konstantách), nikoliv však na jiné či jiných vysvětlujících proměnných, jsou-li tyto fixovány (tzv. *ceteris paribus*); jinak je model neaditivní. Formální definice aditivity je např. v [3]. Pro účely posouzení vhodnosti aditivity vztahu mezi vysvětlovanou proměnnou (`cena`) a vysvětlujícími proměnnými (`koňská síla`, `země původu`) uvažujme následující dva lineární modely. Jednak aditivní model

$$Y_i = \beta_0 + \sum_{j=2}^{4} \gamma_j C_{j,i} + \beta_1 H_i + \varepsilon_i \tag{1}$$

a naopak neaditivní model (pravděpodobně nejjednodušší, který se nabízí)

$$Y_i = \beta_0 + \sum_{j=2}^{4} \gamma_j C_{j,i} + (\beta_1 + \sum_{j=2}^{4} \delta_j C_{j,i}) H_i + \varepsilon_i, \tag{2}$$

kde pro všechna $i \in \{1, 2, \ldots, 90\}$ je Y_i cena i-tého automobilu, H_i je koňská síla i-tého automobilu, β_0, β_1 jsou regresní koeficienty, které je třeba odhadnout, γ_j a δ_j pro $j \in \{2, 3, 4\}$ jsou regresní koeficienty pro j-tou hodnotu z uspořádané množiny $[\mathrm{USA}, \mathrm{Japonsko}, \mathrm{Německo}, \mathrm{Ostatní}]$ (odpovídající si koeficienty nemusí být pro model 1 a 2 shodné!) a $C_{i,j}$ jsou identifikátorové proměnné tak, že

$$C_{i,j} = \begin{cases} 1, & \text{pokud } i\text{-tý automobil pochází } j\text{-té země} \\ 0, & \text{jinak.} \end{cases} \tag{3}$$

při klasickém postupném dummy kódování kontrastů (proti jedné referenční zemi, zde odpovídající USA[3]). Ptáme se, zda interakční člen v druhém modelu, tedy $\sum_{j=2}^{4} \delta_j C_{j,i} H_i$ má pro model statistický význam. Řešením podle minimalizace součtu čtverců reziduí získáme pro oba modely tyto odhady koeficientů uvedené v tabulkách 1 a 2.

	odhad	střední chyba	t-hodnota	p-hodnota
$\hat{\beta}_0$	-7.279	1.570	-4.636	< 0.001
$\hat{\gamma}_2$ (země = Japonsko)	-0.656	0.975	-0.673	0.503
$\hat{\gamma}_3$ (země = Německo)	3.474	1.596	2.176	0.032
$\hat{\gamma}_4$ (země = Ostatní)	3.162	1.351	2.340	0.022
$\hat{\beta}_1$ (koňská síla)	0.174	0.010	16.592	< 0.001

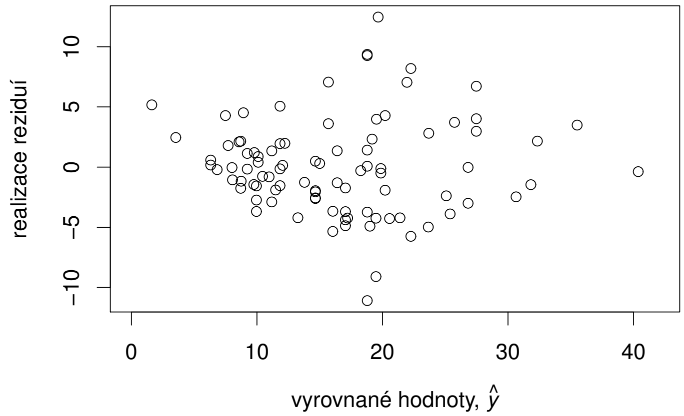
Tabulka 1: Odhady regresní koeficientů pro aditivní model (1)

	odhad	střední chyba	t-hodnota	p-hodnota
$\hat{\beta}_0$	-8.806	2.530	-3.481	0.001
$\hat{\gamma}_2$ (země = Japonsko)	2.679	3.252	0.824	0.412
$\hat{\gamma}_3$ (země = Německo)	9.699	8.597	1.128	0.263
$\hat{\gamma}_4$ (země = Ostatní)	-2.076	4.916	-0.422	0.674
$\hat{\beta}_1$ (koňská síla)	0.185	0.018	10.229	< 0.001
$\hat{\delta}_2$ (země = Japonsko)	-0.024	0.023	-1.069	0.288
$\hat{\delta}_3$ (země = Německo)	-0.043	0.057	-0.749	0.456
$\hat{\delta}_4$ (země = Ostatní)	0.052	0.042	1.235	0.220

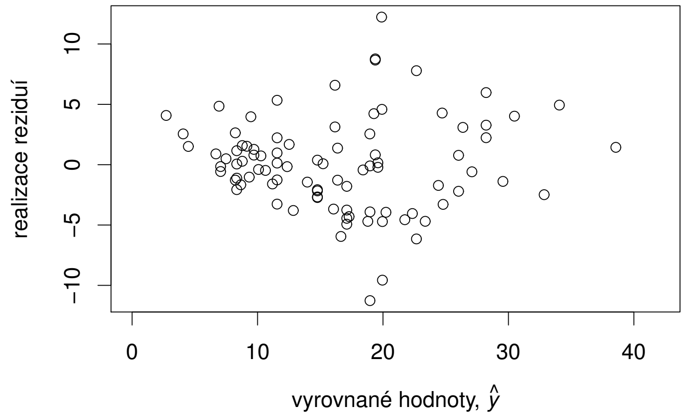
Tabulka 2: Odhady regresní koeficientů pro neaditivní model (2)

[3] Volbu obhájíme např. tím, že se jedná o klasického producenta relativně velkého množství vozů.

V obou modelech se jeví jako statisticky významné pro vysvětlení pozorované variability vždy proměnná *koňská síla*. Uveďme ještě alespoň náhled na regresní diagnostiku – realizace teoretických reziduí obou modelů by měly „co možná nejvíce“ splňovat slabou sadu předpokladů (střední hodnotu rovnou 0, shodný rozptyl a nezávislost). Na obrázcích 4 a 7 nahlédneme, že střední hodnota realizací reziduí víceméně nulová, avšak předpoklad homoskedasticity (a velmi pravděpodobně i non-multikolinearity) je udržitelný jen rámcově (v centrálních částech je větší variabilita realizací reziduí), přesto se nejeví jeho narušení jako příliš významné.



Obrázek 4: Realizace reziduí vs. vyrovnané hodnoty u aditivního modelu



Obrázek 5: Realizace reziduí vs. vyrovnané hodnoty u aditivního modelu

Nejsnazším způsobem, jak zjistit přínos interakčního členu je analýza rozptylu modelu (2), kde je obsažen. V analýze rozptylu záleží na pořadí regresorů, zařazení interakčního členu jako posledního tedy zhodnotí přínos k vysvětlení celkové variability modelu; výsledky jsou v tabulce 3. Regresní diagnostika analýzy rozptylu je závislá na obdobných předpokladech jako diagnostika lineárních modelů 1 a 2, navíc se očekává i normální rozdělení realizací reziduí – předpoklady jsou narušeny ne příliš „silnou“ homoskedasticitou reziduí. Zjevně je aditivní model bez interakčního členu (1) v podstatě obecnější variantou složitějšího neaditivního modelu (2), resp. jsou to obdobné modely; není třeba zamítat nulovou hypotézu o dostatečnosti aditivního modelu. V rámci požadavku na *parsinomitu* modelu tedy akceptujeme aditivní závislost mezi cenou vs. koňskou silou a zemí původu, protože neaditivní závislost nevysvětlí prakticky žádný díl celkové variability závisle proměnné navíc.

	stupně volnosti	součet čtverců	střední čtverce	F-hodnota	p-hodnota
země	3.000	316.237	105.412	6.549	0.001
koňská síla	1.000	4502.600	4502.600	279.740	< 0.001
země $\times$ koňská síla	3.000	70.462	23.487	1.459	0.232
reziduály	82.000	1319.846	16.096		

Tabulka 3: Analýza rozptylu neaditivního modelu (2)

(ii) Uvažujme nyní model závislosti ceny automobilu pouze na zemi původu (jde o lineární model dobře odpovídající jednorozměrné analýze rozptylu) a model, kde tuto závislost očistíme o vliv koňské síly automobilu (pak půjde o model analýzy kovariance). Neformálně, očištění provedeme pouze přidáním proměnné *koňská síla* do předchozího modelu, čímž tato proměnná vysvětlí svou část variability; při jejím nezařazení může být vliv země původu automobilu na jeho cenu *confoundován* neuvažovanou (byť změřenou) koňskou sílou. Formálně tedy

$$Y_i = \beta_0 + \sum_{j=2}^{4} \gamma_j C_{j,i} + \varepsilon_i \tag{4}$$

a dále

$$Y_i = \beta_0 + \sum_{j=2}^{4} \gamma_j C_{j,i} + \beta_1 H_i + \varepsilon_i, \tag{5}$$

kde pro všechna $i \in \{1, 2, \dots, 90\}$ je Y_i cena i-tého automobilu, H_i je koňská síla i-tého automobilu, β_0, β_1 jsou regresní koeficienty, které je třeba odhadnout, γ_j pro $j \in \{2, 3, 4\}$ jsou regresní koeficienty pro j-tou hodnotu z uspořádané množiny $[\text{USA}, \text{Japonsko}, \text{Německo}, \text{Ostatní}]$ (odpovídající si koeficienty nemusí být pro model 4 a 5 shodné!) a pro $C_{i,j}$ platí vztah (3). Model (5) je vlastně shodný s modelem (1).

Výsledky analýzy rozptylu obou modelů jsou v tabulkách 4 a 5.

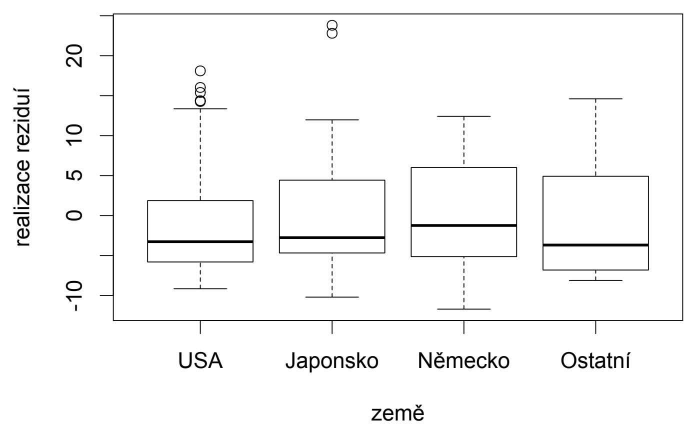
	stupně volnosti	součet čtverců	střední čtverce	F-hodnota	p-hodnota
země	3.000	316.237	105.412	1.538	0.210
reziduály	86.000	5892.908	68.522		

Tabulka 4: Analýza rozptylu modelu závislosti cena vs. země původu

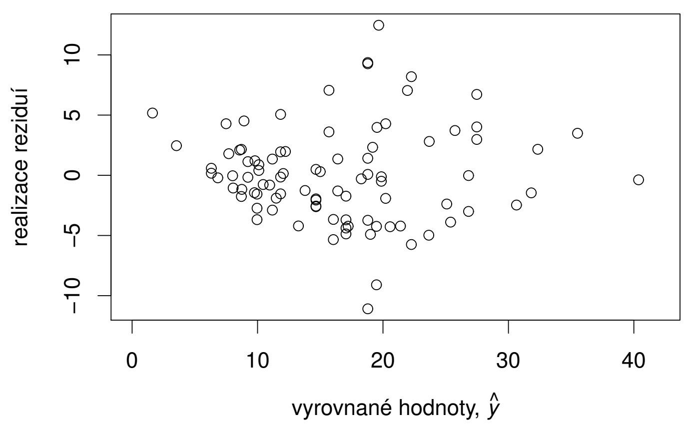
	stupně volnosti	součet čtverců	střední čtverce	F-hodnota	p-hodnota
země	3.000	316.237	105.412	6.445	0.001
koňská síla	1.000	4502.600	4502.600	275.278	< 0.001
reziduály	85.000	1390.308	16.357		

Tabulka 5: Analýza rozptylu modelu závislosti cena vs. země původu a koňská síla

Diagnostika obou modelů je na obrázcích 6 a 7. Snadno nahlédneme, že pro původní model (4) není proměnná *cena* statisticky významná, zatímco po očištění v modelu (5) už svůj statistická význam (ve vysvětlení celkové variability závisle proměnné) má. To potvrzuje i náš objev v rámci explorační grafické analýzy, kde se zdálo, že např. pro německé vozy je cena celkově „posunuta“ výše bez ohledu na koňskou sílu (tj. při očištění modelu o její vliv).



Obrázek 6: Realizace reziduí vs. vyrovnané hodnoty u modelu (4)



Obrázek 7: Realizace reziduí vs. vyrovnané hodnoty u modelu (5)

2.3 Implementace v R

Zde je uveden kód v jazyce R, ve kterém byly zpracovávány veškeré výpočty a rovněž generovány diagramy.

```
###############################################################################
###############################################################################
###############################################################################

## instaluji a loaduji balíčky ---------------------------------------------

invisible(
    lapply(c(
            "xtable",
            "openxlsx"
        ),
```

```r
        function(package){

            if(!(package %in% rownames(installed.packages()))){

                install.packages(
                    package,
                    dependencies = TRUE,
                    repos = "http://cran.us.r-project.org"
                )

            }

            library(package, character.only = TRUE)

        }
    )
)


## -----------------------------------------------------------------------------

################################################################################

## nastavuji handling se zipováním v R ----------------------------------------

Sys.setenv(R_ZIPCMD = "C:/Rtools/bin/zip")


## -----------------------------------------------------------------------------

################################################################################

## nastavuji pracovní složku --------------------------------------------------

while(!"script.R" %in% dir()){
    setwd(choose.dir())
}

mother_working_directory <- getwd()


## -----------------------------------------------------------------------------

################################################################################

## vytvářím posložky pracovní složky ------------------------------------------

setwd(mother_working_directory)

for(my_subdirectory in c("vstupy", "vystupy")){

    if(!file.exists(my_subdirectory)){

        dir.create(file.path(

            mother_working_directory, my_subdirectory

        ))

    }

}
```

```
## -----------------------------------------------------------------------------

################################################################################

## loaduji data -----------------------------------------------------------------

my_data <- read.table(

    file = "http://www1.aucegypt.edu/faculty/hadi/RABE5/Data5/P155b.txt",
    header = TRUE,
    sep = "\t"

)


## -----------------------------------------------------------------------------

################################################################################

## vytvářím pro jistotu offline kopii dat a ukládám ji ------------------------

setwd(paste(mother_working_directory, "vstupy", sep = "/"))

#### nejdříve plaintextová kopie dat -------------------------------------------

write.table(
    x = my_data,
    file = "P155b.txt",
    quote = FALSE,
    sep = "\t",
    row.names = FALSE,
    col.names = TRUE
)


#### a nyní Excelovou kopii dat ------------------------------------------------

###### vytvářím sešit ----------------------------------------------------------

addWorksheet(
    wb = data_backup <- createWorkbook(),
    sheetName = "data"
)


###### ukládám do sešitu data --------------------------------------------------

writeData(
    wb = data_backup,
    sheet = "data",
    rowNames = FALSE,
    colNames = TRUE,
    x = my_data
)


###### vytvářím dva své styly - jednak tučné písmo, jednak písmo zarovnané
###### doprava v rámci buňky ---------------------------------------------------

my_bold_style <- createStyle(textDecoration = "bold")
right_halign_cells <- createStyle(halign = "right")

addStyle(
```

```r
    wb = data_backup,
    sheet = "data",
    style = my_bold_style,
    rows = rep(1, dim(my_data)[2]),
    cols = c(1:dim(my_data)[2])
)

addStyle(
    wb = data_backup,
    sheet = "data",
    style = right_halign_cells,
    rows = 2:(dim(my_data)[1] + 1),
    cols = 2:dim(my_data)[2],
    gridExpand = TRUE
)


###### nastavuji automatickou šířku sloupce ----------------------------------

setColWidths(
    wb = data_backup,
    sheet = "data",
    cols = 1:dim(my_data)[2],
    widths = "auto"
)


###### ukládám workbook -------------------------------------------------------

saveWorkbook(
    wb = data_backup,
    file = "P155b.xlsx",
    overwrite = TRUE
)


setwd(mother_working_directory)


## ----------------------------------------------------------------------------

###############################################################################

## preprocessing dat ----------------------------------------------------------

#### překódovávám kategorickou proměnnou "Country" ----------------------------

my_data$Country <- factor(
    my_data$Country,
    levels = c("USA", "Japan", "Germany", "Other"),
    labels = c("USA", "Japonsko", "Německo", "Ostatní")
)


## ----------------------------------------------------------------------------

###############################################################################

## Exploratory Data Analysis --------------------------------------------------

#### nejdříve vytvářím diagram závislosti ceny automobilů na jejich
#### koňské síle --------------------------------------------------------------
```

```r
setwd(paste(mother_working_directory, "vystupy", sep = "/"))

cairo_ps(
    file = "cena_vs_konska_sila.eps",
    width = 8,
    height = 7,
    pointsize = 18
)

par(mar = c(4, 4, 1, 1))


plot(
    Y ~ Horsepower,
    my_data,
    xlab = "koňská síla",
    ylab = "cena"
)

dev.off()


#### -------------------------------------------------------------------------

cairo_ps(
    file = "cena_vs_konska_sila_barevne.eps",
    width = 8,
    height = 7,
    pointsize = 18
)

par(mar = c(4, 4, 1, 1))

plot(
    Y ~ Horsepower,
    my_data,
    xlab = "koňská síla",
    ylab = "cena",
    col = my_data$Country,
    pch = 19
)

legend(
    x = "topleft",
    legend = levels(my_data$Country),
    col = 1:length(my_data$Country),
    pch = 19,
    title = "země původu",
    #cex = 0.85
)

dev.off()


#### -------------------------------------------------------------------------

#### nyní vytvářím matici závislosti cena vs. koňská síla izolovaně pro
#### země původu ------------------------------------------------------------

cairo_ps(
    file = "cena_vs_konska_sila_dle_zeme_puvodu.eps",
    width = 16,
    height = 5.0,
```

```
    pointsize = 27
)

par(
    mfrow = c(1, 4),
    mar = c(0, 0, 0, 0),
    oma = c(0.0, 2.0, 0.0, 0.5),
    #tcl = -0.25,
    mgp = c(2, 0.6, 0)
)

for(my_country in levels(my_data$Country)){

    par(mar = c(4, 1, 2, 0))

    plot(
        Y ~ Horsepower,
        data = my_data[my_data$Country == my_country, ],
        xlab = "koňská síla",
        ylab = "cena",
        main = my_country,
        xlim = c(50, 250),
        ylim = c(0, 45),
        yaxt = "n"
    )

    if(which(levels(my_data$Country) == my_country) == 1){
        axis(2, at = seq(0, 45, 10), labels = seq(0, 45, 10))
    }

    abline(
        lm(
            Y ~ Horsepower,
            data = my_data[my_data$Country == my_country, ]
        ),
        col = "red"
    )

    lines(
        seq(
            min(my_data$Horsepower),
            max(my_data$Horsepower),
            length.out = 1000
        ),
        predict(
            lm(
                Y ~ Horsepower,
                data = my_data[my_data$Country == my_country, ]
            ),
            newdata = data.frame("Horsepower" = seq(
                min(my_data$Horsepower),
                max(my_data$Horsepower),
                length.out = 1000
            )),
            interval = "confidence"
        )[, 2],
        lty = "dashed",
        col = "red"
    )

    lines(
        seq(
            min(my_data$Horsepower),
```

```
            max(my_data$Horsepower),
            length.out = 1000
        ),
        predict(
            lm(
                Y ~ Horsepower,
                data = my_data[my_data$Country == my_country, ]
            ),
            newdata = data.frame("Horsepower" = seq(
                min(my_data$Horsepower),
                max(my_data$Horsepower),
                length.out = 1000
            )),
            interval = "confidence"
        )[, 3],
        lty = "dashed",
        col = "red"
    )

}

mtext("cena", side = 2, outer = TRUE, cex = 0.7, line = 1.0)

dev.off()


setwd(mother_working_directory)


## -----------------------------------------------------------------------------

################################################################################

## zkoumám aditivitu vs. neaditivitu vztahu mezi cenou vs. koňskou silou
## a zemí původu --------------------------------------------------------------

additive_model <- lm(
    Y ~ Country + Horsepower,
    data = my_data
)

non_additive_model <- lm(
    Y ~ Country + Horsepower + Country : Horsepower,
    data = my_data
)

#### regresní diagnostika -----------------------------------------------------

plot(additive_model)
plot(non_additive_model)


setwd(paste(mother_working_directory, "vystupy", sep = "/"))

for(my_model_name in c("additive_model", "non_additive_model")){

    my_model <- get(my_model_name)

    setEPS()
    postscript(
        file = paste(
            "residua_vs_vyrovnane_hodnoty_",
            my_model_name,
```

```r
            ".eps",
            sep = ""
        ),
        width = 8,
        height = 5,
        pointsize = 18
    )

    par(mar = c(4, 4, 1, 1))

    plot(
        resid(my_model) ~ fitted(my_model),
        xlab = expression(
            paste("vyrovnané hodnoty, ", hat(italic(y)), sep = "")
        ),
        ylab = "realizace reziduí",
        xlim = c(0, 42)
    )

    dev.off()

}

setwd(mother_working_directory)


#### printable verze odhadů koeficientů obou modelů ---------------------------

print(
    xtable(
        summary(additive_model)$coefficients,
        align = rep("", ncol(summary(additive_model)$coefficients) + 1),
        digits = 3
    ),
    floating = FALSE,
    tabular.environment = "tabular",
    hline.after = NULL,
    include.rownames = TRUE,
    include.colnames = TRUE
)

print(
    xtable(
        summary(non_additive_model)$coefficients,
        align = rep("", ncol(summary(non_additive_model)$coefficients) + 1),
        digits = 3
    ),
    floating = FALSE,
    tabular.environment = "tabular",
    hline.after = NULL,
    include.rownames = TRUE,
    include.colnames = TRUE
)

#### je lepší pouze aditivní model ? ------------------------------------------

anova(non_additive_model)    ## zdá se, že interakční člen je "zbytečný"

anova(
    additive_model,
    non_additive_model
)                            ## dostáváme dle očekávání zcela shodný výsledek
                             ## jako výše -- aditivní model bez interakčního
```

```r
                             ## členu je v postatě podmínkovou variantou
                             ## složitejšího modelu, resp. jsou si podobné
                             ## a není třeba zamítat nulovou hypotézu
                             ## o dostatečnosti aditivního modelu

print(
    xtable(
        anova(non_additive_model),
        align = rep("", ncol(anova(non_additive_model)) + 1),
        digits = 3
    ),
    floating = FALSE,
    tabular.environment = "tabular",
    hline.after = NULL,
    include.rownames = TRUE,
    include.colnames = TRUE
)


## ---------------------------------------------------------------------------

###############################################################################

## je závislost ceny na zemi významná po očištění o koňskou sílu? ------------

only_country_model <- lm(
    Y ~ Country,
    data = my_data
)

summary(only_country_model)

anova(
    only_country_model
)

anova(
    additive_model
)

anova(
    only_country_model,
    additive_model
)

print(
    xtable(
        anova(only_country_model),
        align = rep("", ncol(anova(only_country_model)) + 1),
        digits = 3
    ),
    floating = FALSE,
    tabular.environment = "tabular",
    hline.after = NULL,
    include.rownames = TRUE,
    include.colnames = TRUE
)

print(
    xtable(
        anova(additive_model),
        align = rep("", ncol(anova(additive_model)) + 1),
        digits = 3
```

```r
    ),
    floating = FALSE,
    tabular.environment = "tabular",
    hline.after = NULL,
    include.rownames = TRUE,
    include.colnames = TRUE
)


#### tisknu regresní diagnostiku předchozích dvou modelů ---------------------

setwd(paste(mother_working_directory, "vystupy", sep = "/"))

setEPS()
postscript(
    file = "residua_vs_vyrovnane_hodnoty_only_country_model.eps",
    width = 8,
    height = 5,
    pointsize = 18
)

par(mar = c(4, 4, 1, 1))

boxplot(
    list(
        "USA" = resid(
            only_country_model
        )[which(my_data$Country == "USA")],
        "Japonsko" = resid(
            only_country_model
        )[which(my_data$Country == "Japonsko")],
        "Německo" = resid(
            only_country_model
        )[which(my_data$Country == "Německo")],
        "Ostatní" = resid(
            only_country_model
        )[which(my_data$Country == "Ostatní")]
    ),
    xlab = "země",
    ylab = "realizace reziduí"
)

dev.off()


setwd(mother_working_directory)


## ---------------------------------------------------------------------------

###############################################################################
###############################################################################
###############################################################################
```

3 Reference

[1] CHATTERJEE, Samprit a Ali S. HADI. *Regression Analysis by Example (Wiley Series in Probability and Statistics)*. B.m.: Wiley, 2013. ISBN 978-0-470-90584-5.

[2] R CORE TEAM. *R: A Language and Environment for Statistical Computing* [online]. Vienna, Austria: R Foundation for Statistical Computing, 2016. Dostupné z: https://www.R-project.org/

[3] PETR, Hebák. *Vícerozměrné statistické metody*. Praha: Informatorium, 2004. ISBN 80-7333-036-9.