

2. zápočtová úloha z 01RAD

Jazmína Kreanová

Popis úlohy

Datový soubor `Boston` je obsažen v balíku `MASS` a lze použít rovnou po načtení příslušné knihovny.

```
knitr::opts_chunk$set(echo = FALSE)
list_of_packages <- c("tidyverse", "MASS", "GGally", "moderndive", "dplyr", "car", "lmtest", "nortest")
missing_packages <- list_of_packages[!(list_of_packages %in% installed.packages()[, "Package"])]
install.packages(missing_packages)
lapply(list_of_packages, library, character.only = TRUE)
#library(tidyverse)
#library(GGally)
#library(knitr)
#library(MASS)

1. 'corrplot' · 'splines' · 'nortest' · 'lmtest' · 'zoo' · 'car' · 'carData' · 'moderndive' · 'GGally' · 'MASS' · 'forcats' · 'stringr' ·
'dplyr' · 'purrr' · 'readr' · 'tidyr' · 'tibble' · 'ggplot2' · 'tidyverse' · 'stats' · 'graphics' · 'grDevices' · 'utils' · 'datasets' ·
'methods' · 'base'
2. 'corrplot' · 'splines' · 'nortest' · 'lmtest' · 'zoo' · 'car' · 'carData' · 'moderndive' · 'GGally' · 'MASS' · 'forcats' · 'stringr' ·
'dplyr' · 'purrr' · 'readr' · 'tidyr' · 'tibble' · 'ggplot2' · 'tidyverse' · 'stats' · 'graphics' · 'grDevices' · 'utils' · 'datasets' ·
'methods' · 'base'
3. 'corrplot' · 'splines' · 'nortest' · 'lmtest' · 'zoo' · 'car' · 'carData' · 'moderndive' · 'GGally' · 'MASS' · 'forcats' · 'stringr' ·
'dplyr' · 'purrr' · 'readr' · 'tidyr' · 'tibble' · 'ggplot2' · 'tidyverse' · 'stats' · 'graphics' · 'grDevices' · 'utils' · 'datasets' ·
'methods' · 'base'
4. 'corrplot' · 'splines' · 'nortest' · 'lmtest' · 'zoo' · 'car' · 'carData' · 'moderndive' · 'GGally' · 'MASS' · 'forcats' · 'stringr' ·
'dplyr' · 'purrr' · 'readr' · 'tidyr' · 'tibble' · 'ggplot2' · 'tidyverse' · 'stats' · 'graphics' · 'grDevices' · 'utils' · 'datasets' ·
'methods' · 'base'
5. 'corrplot' · 'splines' · 'nortest' · 'lmtest' · 'zoo' · 'car' · 'carData' · 'moderndive' · 'GGally' · 'MASS' · 'forcats' · 'stringr' ·
'dplyr' · 'purrr' · 'readr' · 'tidyr' · 'tibble' · 'ggplot2' · 'tidyverse' · 'stats' · 'graphics' · 'grDevices' · 'utils' · 'datasets' ·
'methods' · 'base'
6. 'corrplot' · 'splines' · 'nortest' · 'lmtest' · 'zoo' · 'car' · 'carData' · 'moderndive' · 'GGally' · 'MASS' · 'forcats' · 'stringr' ·
'dplyr' · 'purrr' · 'readr' · 'tidyr' · 'tibble' · 'ggplot2' · 'tidyverse' · 'stats' · 'graphics' · 'grDevices' · 'utils' · 'datasets' ·
'methods' · 'base'

#install.packages("car")
#library(car)
fig <- function(width, height){
  options(repr.plot.width = width, repr.plot.height = height)
}
options(jupyter.plot_scale=2)

head(Boston)
```

A data.frame: 6 × 14														
	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.00632	18	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
2	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
3	0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
4	0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5	0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
6	0.02985	0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7

Obsahuje celkem 506 záznamů z obcí v předměstí města Boston, MA, USA a data pocházejí ze studie v roce 1978. Viz Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102.

Základní charakteristiky ohledně jednotlivých proměnných získáte pomocí funkcí `str(Boston)` a `summary(Boston)`.

Data celkem obsahují 14 proměnných, přičemž naším cílem je prozkoumat vliv 13 z nich na cenu nemovitostí `medv`. Přičemž anglický popis jednotlivých proměnných (sloupců) je následující:

Feature & Description

`crim` per capita crime rate by town

`zn` : proportion of residential land zoned for lots over 25,000 sq.ft

indus : proportion of non-retail business acres per town

chas : Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

nox : nitrogen oxides concentration (parts per 10 million)

rm : average number of rooms per dwelling

age : proportion of owner-occupied units built prior to 1940

dis : weighted mean of distances to five Boston employment centres

rad : index of accessibility to radial highways

tax : full-value property-tax rate per \$10,000

ptratio : pupil-teacher ratio by town

black : $1000(B_k - 0.63)^2$ where B_k is the proportion of blacks by town

lstat : lower status of the population (percent)

medv : median value of owner-occupied homes in \$1000s

▼ Podmínky a body

Úkol i protokol vypracujte samostatně. Pokud na řešení nějaké úlohy budete přesto s někým spolupracovat, radit se, nezapomeňte to u odpovědi na danou otázku uvést. Tato zápočtová úloha obsahuje 10 otázek po 1 bodu. Celkem za 3 zápočtové úlohy bude možné získat 30 bodů, přičemž pro získání zápočtu je potřeba více jak 20. Další dodatečné body mohou případně individuálně udělit za extra práci na mini domácích úkolech nebo za aktivitu v hodině.

Odevzdání

Protokol ve formátu pdf (včetně příslušného Rmd souboru), nebo jak jupyter NB (ideálně s odkazem na Colab) odevzdejte prostřednictvím MS Teams, nejpozději do půlnoci 14. 12. 2022 (tj. za 3 týdny).

Příprava dat:

- Otázka 01

Z dat vyfiltrujte jen pozorování, kde proměnná `chas` je rovna 0, proměnná `rad` je menší než 20 a odevza `medv` neobsahuje opakující se maximální hodnoty vzniklé nejspíše zaokrouhlením. Zkontrolujte, že výsledný datset neobsahuje chybějící hodnoty a vykreslete scatterplot pro proměnné `indus` a `medv`.

```
summary(Boston)
```

crim	zn	indus	chas
Min. : 0.00632	Min. : 0.00	Min. : 0.46	Min. : 0.00000
1st Qu.: 0.08205	1st Qu.: 0.00	1st Qu.: 5.19	1st Qu.: 0.00000
Median : 0.25651	Median : 0.00	Median : 9.69	Median : 0.00000
Mean : 3.61352	Mean : 11.36	Mean : 11.14	Mean : 0.06917
3rd Qu.: 3.67708	3rd Qu.: 12.50	3rd Qu.: 18.10	3rd Qu.: 0.00000
Max. : 88.97620	Max. : 100.00	Max. : 27.74	Max. : 1.00000

nox	rm	age	dis
Min. : 0.3850	Min. : 3.561	Min. : 2.90	Min. : 1.130
1st Qu.: 0.4490	1st Qu.: 5.886	1st Qu.: 45.02	1st Qu.: 2.100
Median : 0.5380	Median : 6.208	Median : 77.50	Median : 3.207
Mean : 0.5547	Mean : 6.285	Mean : 68.57	Mean : 3.795
3rd Qu.: 0.6240	3rd Qu.: 6.623	3rd Qu.: 94.08	3rd Qu.: 5.188
Max. : 0.8710	Max. : 8.780	Max. : 100.00	Max. : 12.127

rad	tax	ptratio	black
Min. : 1.000	Min. : 187.0	Min. : 12.60	Min. : 0.32
1st Qu.: 4.000	1st Qu.: 279.0	1st Qu.: 17.40	1st Qu.: 375.38
Median : 5.000	Median : 330.0	Median : 19.05	Median : 391.44
Mean : 9.549	Mean : 408.2	Mean : 18.46	Mean : 356.67
3rd Qu.: 24.000	3rd Qu.: 666.0	3rd Qu.: 20.20	3rd Qu.: 396.23
Max. : 24.000	Max. : 711.0	Max. : 22.00	Max. : 396.90

lstat	medv
Min. : 1.73	Min. : 5.00
1st Qu.: 6.95	1st Qu.: 17.02
Median : 11.36	Median : 21.20
Mean : 12.65	Mean : 22.53
3rd Qu.: 16.95	3rd Qu.: 25.00
Max. : 37.97	Max. : 50.00

```
#table(factor(Boston$medv))
```

Našou úlohou je vyfiltrovať dataset tak, aby zostali pozorovania s $\$chas=0$, $rad < 20$ a aby dataset neobsahoval opakujúce sa maximálne hodnoty v odozve `medv`.

Vypíšeme si najprv pozorovania s opakujúcou sa hodnotou maxima v kýženej odozve `medv`. To činí 8 rôznych pozorovaní, pričom nepovažujeme za vhodné vybrať jedno pozorovanie, ktoré by sme do datasetu začlenili.

V ďalšom kroku teda vyfiltrujeme dáta tak, že nebudú obsahovať žiadne pozorovanie, ktoré v premennej `medv` nadobúda maximum. a overíme, že v datasete nemáme chýbajúce hodnoty. Nový dataset obsahuje 339 pozorovaní.

```
subset(Boston, Boston$chas==0 & Boston$rad<20 & Boston$medv==max(Boston$medv))
#dim(Boston_mx)
```

A data.frame: 8 × 14

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	me
	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
162	1.46336	0	19.58	0	0.6050	7.489	90.8	1.9709	5	403	14.7	374.43	1.73	
167	2.01019	0	19.58	0	0.6050	7.929	96.2	2.0459	5	403	14.7	369.30	3.70	
187	0.05602	0	2.46	0	0.4880	7.831	53.6	3.1992	3	193	17.8	392.63	4.45	
196	0.01381	80	0.46	0	0.4220	7.875	32.0	5.6484	4	255	14.4	394.23	2.97	
205	0.02009	95	2.68	0	0.4161	8.034	31.9	5.1180	4	224	14.7	390.55	2.88	
226	0.52693	0	6.20	0	0.5040	8.725	83.0	2.8944	8	307	17.4	382.00	4.63	
258	0.61154	20	3.97	0	0.6470	8.704	86.9	1.8010	5	264	13.0	389.70	5.12	
268	0.57834	20	3.97	0	0.5750	8.297	67.0	2.4216	5	264	13.0	384.54	7.44	

```
Boston_data<-subset(Boston, Boston$chas==0 & Boston$rad<20 & Boston$medv != max(Boston$medv))
```

Keďže premenná `chas` obsahuje jedinú hodnotu rovnakú pre všetky pozorovania, pre potreby tohto projektu ju odstránime z nového datasetu.

```
Boston_data<- Boston_data[,!names(Boston_data) %in% c("chas")]
```

```
head(Boston_data)
dim(Boston_data)
```

A data.frame: 6 × 13

	crim	zn	indus	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.00632	18	2.31	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
2	0.02731	0	7.07	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
3	0.02729	0	7.07	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
4	0.03237	0	2.18	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5	0.06905	0	2.18	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
6	0.02985	0	2.18	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7

339 × 13

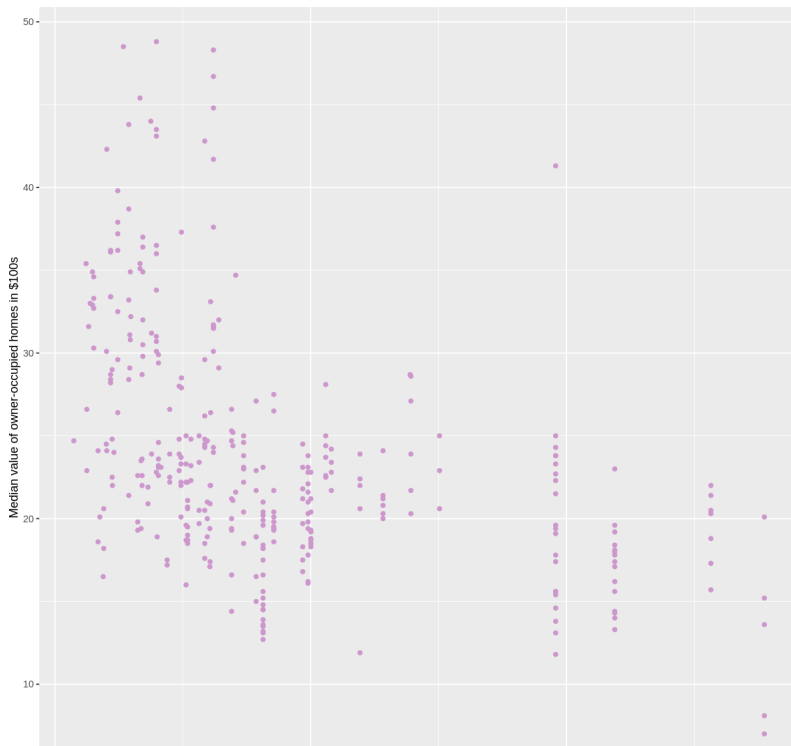
```
Boston_data %>%
  filter(if_any(everything(), is.na))
```

A data.frame: 0 × 13

crim	zn	indus	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>

Vykreslíme scatterplot pre odozvu `medv` a vysvetľujúcu premennú `indus`. Podľa vykresleného scatterplotu nepozorujeme viditeľnú lineárnu závislosť mediánovej ceny nehnuteľností na podieli nemaloobchodného podnikania. Predpokladáme, že cena teda závisí aj na iných premenných, respektíve premenná `indus` môže byť závislá.

```
ggplot(Boston_data,aes(x=indus,y=medv))+
  geom_point(col="plum3")+
  labs(x= "Proportion of non-retail business acres per town", y="Median value of owner-occupied homes in $100s")
```



Regresní model závislosti mediánu ceny nemovitosti na zastoupení ne-maloobchodního podnikání v daném místě:

- Otázka 2

Sestavte jednoduchý regresní model a na jeho základech zjistěte zdali proměnná `indus` ovlivňuje median ceny nemovitostí určených k bydlení `medv`. Pokud ano, o kolik je průměr mediánů cen nemovitostí nižší/vyšší při vzrůstu zastoupení nemaloobchodního podnikání o 5 jednotek?

Zostavíme jednoduchý regresný model závislosti ceny nehnuteľností od podielu podnikania v danej oblasti. Volíme model s interceptom, pretože predpokladáme, že nehnuteľnosť bude mať nenulovú cenu aj v oblasti bez dostatočnej občianskej vybavenosti, navyše z predošlej úvahy v otázke jedna predpokladáme aj iné závislosti.

```
model = lm(medv ~ indus, data = Boston_data)
summary(model)
```

```
Call:
lm(formula = medv ~ indus, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-11.371  -4.313  -1.028   3.385  23.361

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.43376    0.58668   48.466  <2e-16 ***
indus        -0.53598    0.05528   -9.696  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.447 on 337 degrees of freedom
Multiple R-squared:  0.2181,    Adjusted R-squared:  0.2158
F-statistic: 94.01 on 1 and 337 DF,  p-value: < 2.2e-16
```

```
model$coefficients[2]*5
```

```
indus: -2.6799182744655
```

Odhadli sme model: $Y = 28.434 - 0.536 \cdot X$,

kde nám intercept aj premenná `indus` vyšli ako štatisticky významné, *p-hodnota* oboch parametrov vyšla veľmi malá ($2 \cdot 10^{-16}$). Hodnota charakteristiky $R^2 = 0.2181$ je pomerne nízka (čo sme v predošlých úvahách očakávali).

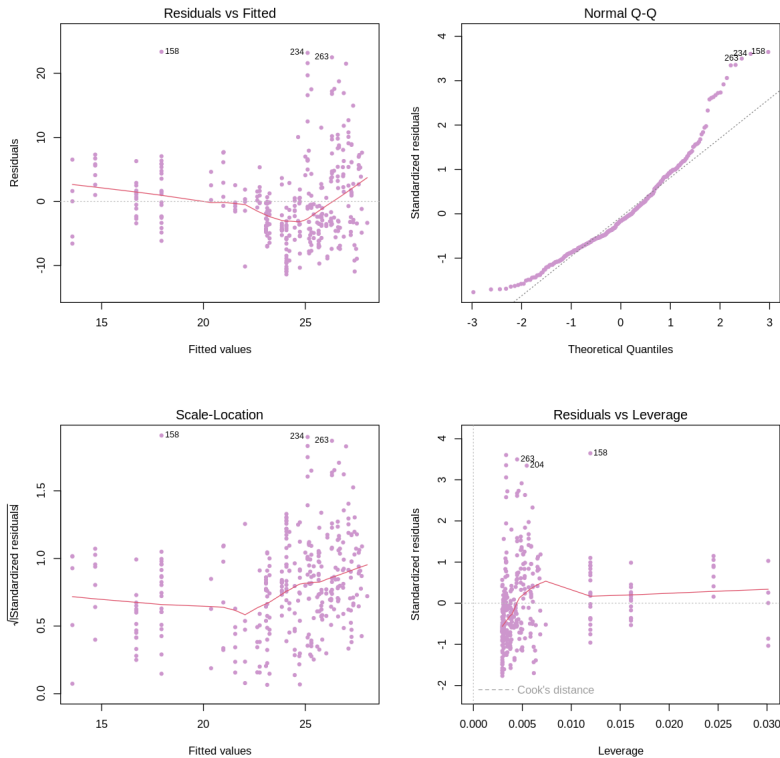
Na základe nášho modelu dôjde pri náraste nemaloobchodného podnikania o 5 jednotiek k poklesu (koeficient má zápornú hodnotu) cien v priemere o 2679.92 \$.

- Otázka 3

Vyzkoušejte model s mocninou a logaritmickou transformací odezvy. Pro výběr mocniné transformace vykreslete optimální log-věrohodnostní profil u Box-Coxovy transformace a porovnejte navrženou transformaci s provedenou logaritmickou.

Vykreslíme ploty reziduí našeho modelu. Z grafů Residuals vs. Fitted a Scale-Location pozorujeme, že reziduá sa v rozptyle rozbiehajú, teda dochádza k porušeniu podmienku homoskedasticity, naviac Q-Q plot indikuje ťažké chvosty, čo nás navádza na transformáciu veličín.

```
#cv8,9
par(mfrow = c(2, 2))
plot(model, col="plum3",pch=20)
```



Mocninná transformácia Box-Coxova transformácia:

$$g(Y + a) = \ln(Y + a) \text{ if } \lambda = 0,$$

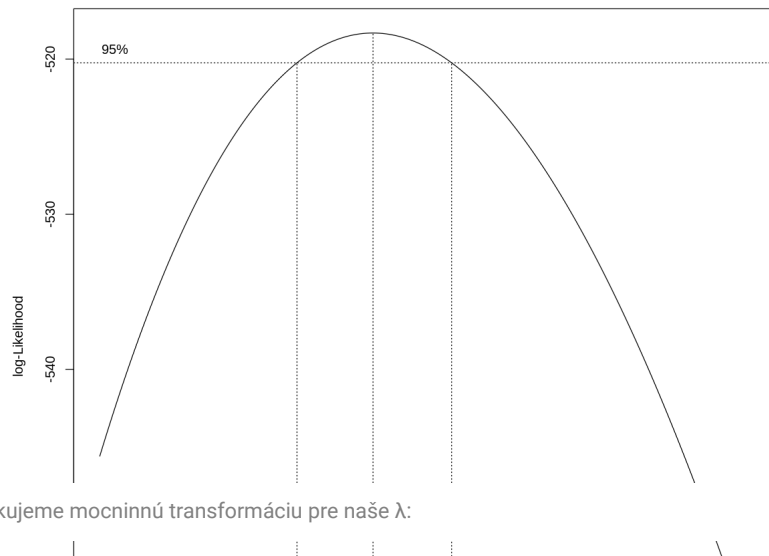
$$g(Y + a) = \frac{(Y + a)^\lambda - 1}{\lambda} \text{ if } \lambda \neq 0,$$

Chceme transformovať dáta pomocou Box-Coxovej transformácie. Ako najvhodnejší exponent nám vyšlo $\lambda = -0.16$, čo spadá do intervalu spoľahlivosti 0 a teda vhodné by bolo použiť logaritmickú transformáciu.

Našou úlohou v tejto časti je porovnať mocninnú a logaritmickú transformáciu, preto nebudeme uvažovať $\lambda = 0$.

```
bc <- boxcox(model, lambda = seq(-1,1 , 1/100))
lambda <- bc$x[which.max(bc$y)]
print("lambda:")
lambda
print("CI:")
c(min(bc$x[bc$y > max(bc$y) - 1/2 * qchisq(.95,1)],max(bc$x[bc$y > max(bc$y) - 1/2 * qchisq(.95,1)]))
```

```
[1] "lambda:"
-0.16
[1] "CI:"
-0.39 · 0.08000000000000001
```



Aplikujeme mocninnú transformáciu pre naše λ :

```
# Power transform
Boston_bcT = Boston_data
Boston_bcT$medv_pwr = (Boston_bcT$medv^lambda-1)/lambda
model_pwr = lm(medv_pwr~indus, data = Boston_bcT)
summary(model_pwr)
```

Call:
lm(formula = medv_pwr ~ indus, data = Boston_bcT)

Residuals:

Min	1Q	Median	3Q	Max
-0.50836	-0.09525	-0.00418	0.09975	0.50572

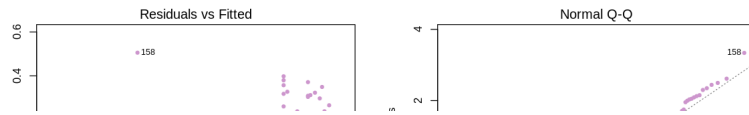
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.580596	0.013861	186.18	<2e-16 ***
indus	-0.014424	0.001306	-11.04	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1523 on 337 degrees of freedom
Multiple R-squared: 0.2657, Adjusted R-squared: 0.2636
F-statistic: 122 on 1 and 337 DF, p-value: < 2.2e-16

```
par(mfrow = c(2, 2))
plot(model_pwr, col="plum3",pch=20)
```



Logaritmická transformácia

$$\ln(y) = \mathbf{X} \cdot \beta + \epsilon$$

$$Y = e^{\mathbf{X} \cdot \beta + \epsilon} = e^{\mathbf{X} \cdot \beta} e^{\epsilon}$$



```
model_ln = lm(log(medv) ~ indus, Boston_data)
summary(model_ln)
```

Call:

```
lm(formula = log(medv) ~ indus, data = Boston_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.73082	-0.15957	-0.01244	0.16211	0.85196

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.33005	0.02293	145.2	<2e-16 ***
indus	-0.02355	0.00216	-10.9	<2e-16 ***

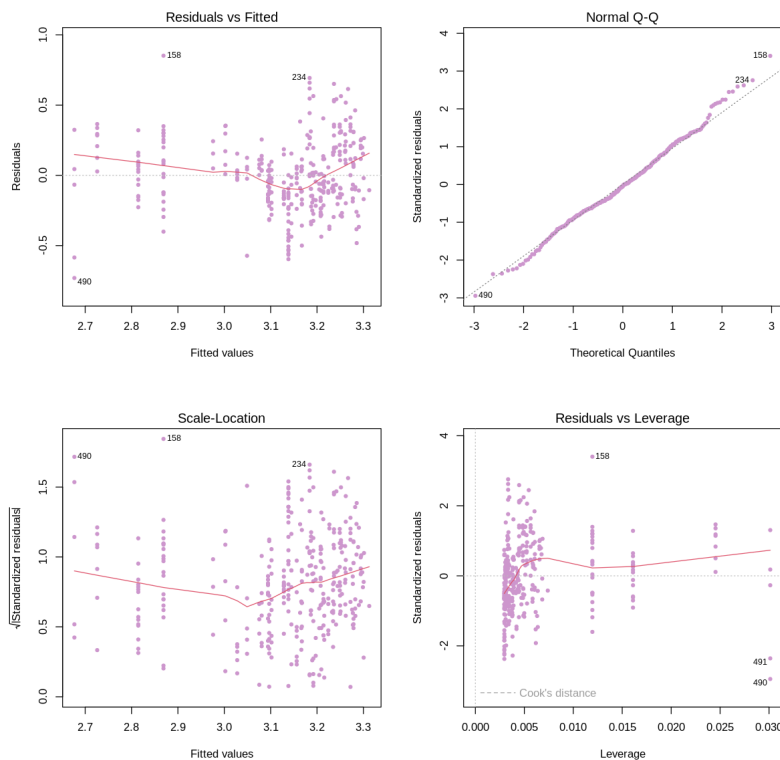
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2519 on 337 degrees of freedom

Multiple R-squared: 0.2607, Adjusted R-squared: 0.2585

F-statistic: 118.9 on 1 and 337 DF, p-value: < 2.2e-16

```
par(mfrow = c(2, 2))
plot(model_ln, col="plum3", pch=20)
```



Porovnanie

```
#summary(model)
#summary(model_ln)
#summary(model_pwr)
```

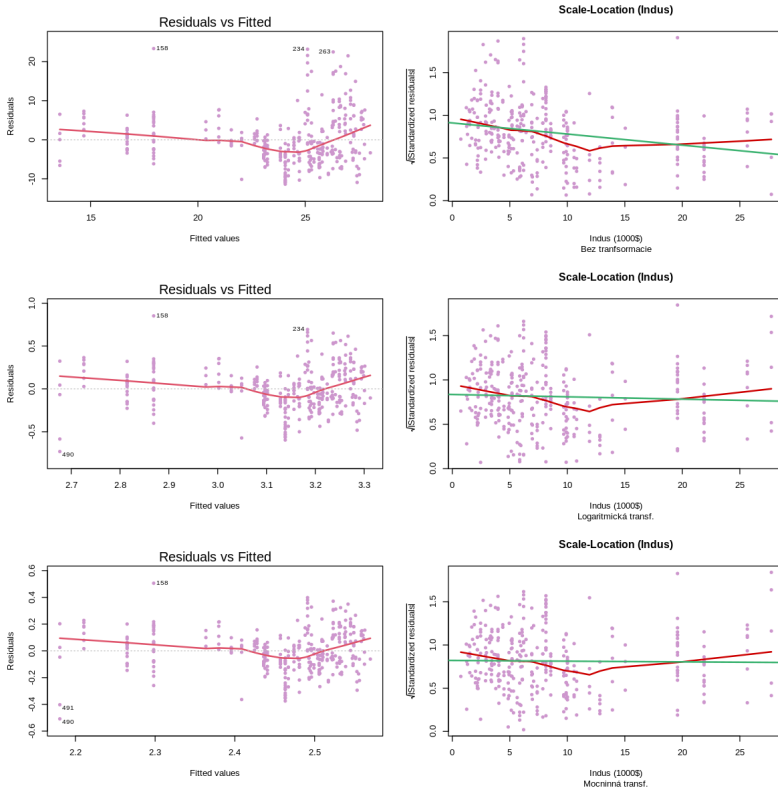
Chceme porovnať naše modely.

Pôvodný lineárny model mal hodnotu charakteristiky $R^2 = 0.2181$ a smerodajnú odchylku reziduí $sd_e = 6.447$. S použitím mocninnnej transformácie sa hodnota zvýšila na $R^2 = 0.2657$, smerodajná odchylka poklesla na $sd_e = 0.1523$ a v rámci logaritmickej transformácie sme obdržali hodnotu $R^2 = 0.2607$, čo je nižšie ako pre mocninnú, ale vyššie než pre netransformovaný model a a $sd_e = 0.2519$ opäť trochu horšie než pre mocninný model a markantne lepšie než pre netransformovaný model.

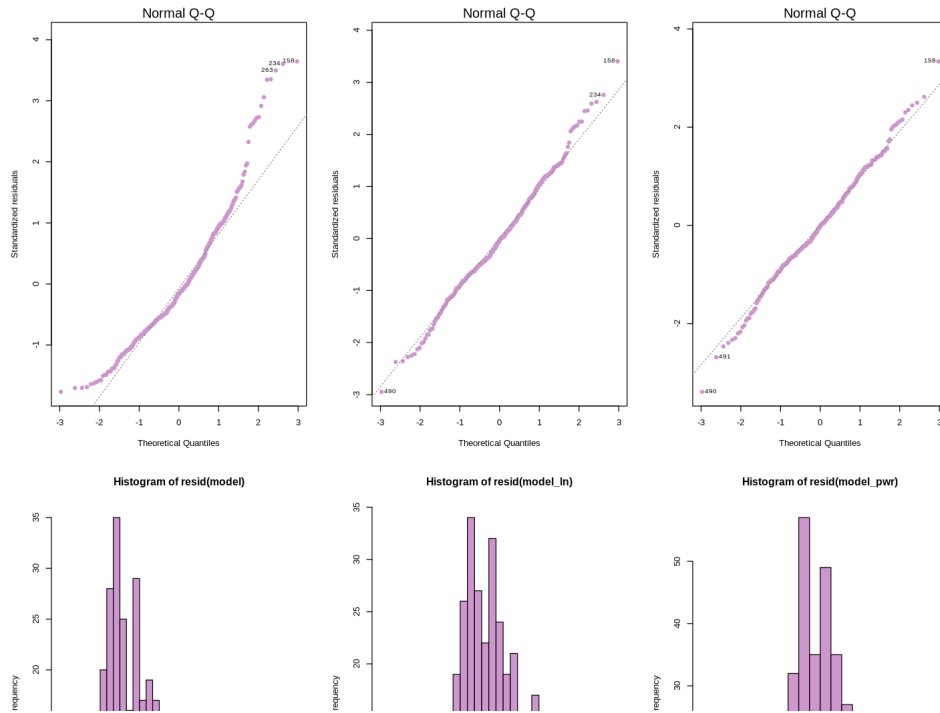
Ďalej vyobrazíme Q-Q ploty pre jednotlivé modely. Môžeme pozorovať, že oba transformované modely grafy indikujú normalitu dát.

```
#fig(12,12)
op <- par(mfrow=c(3,2))
plot(model, sub = "Bez transformácie", which = 1, pch = 20, col = "plum3", lwd = 2)
plot(Boston_data$indus, sqrt(abs(rstandard(model))), sub = "Bez transformácie", pch = 20, col = "plum3", bg = "royalblue4",
      xlab = "Indus (1000$)", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
      main = "Scale-Location (Indus)")
lines(lowess(Boston_data$indus, sqrt(abs(rstandard(model)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(model)))~Boston_data$indus), col = "mediumseagreen", lwd = 2)
plot(model_ln, sub = "Logaritmickej transform.", which = 1, pch = 20, col = "plum3", lwd = 2)
plot(Boston_data$indus, sqrt(abs(rstandard(model_ln))), pch = 20, col = "plum3", bg = "royalblue4",
      xlab = "Indus (1000$)", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
      sub = "Logaritmickej transform.", main = "Scale-Location (Indus)")
lines(lowess(Boston_data$indus, sqrt(abs(rstandard(model_ln)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(model_ln)))~Boston_data$indus), col = "mediumseagreen", lwd = 2)

plot(model_pwr, sub = "Mocninná transform.", which = 1, pch = 20, col = "plum3", lwd = 2)
plot(Boston_data$indus, sqrt(abs(rstandard(model_pwr))), pch = 20, col = "plum3", bg = "royalblue4",
      xlab = "Indus (1000$)", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
      sub = "Mocninná transform.", main = "Scale-Location (Indus)")
lines(lowess(Boston_data$indus, sqrt(abs(rstandard(model_pwr)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(model_pwr)))~Boston_data$indus), col = "mediumseagreen", lwd = 2)
```

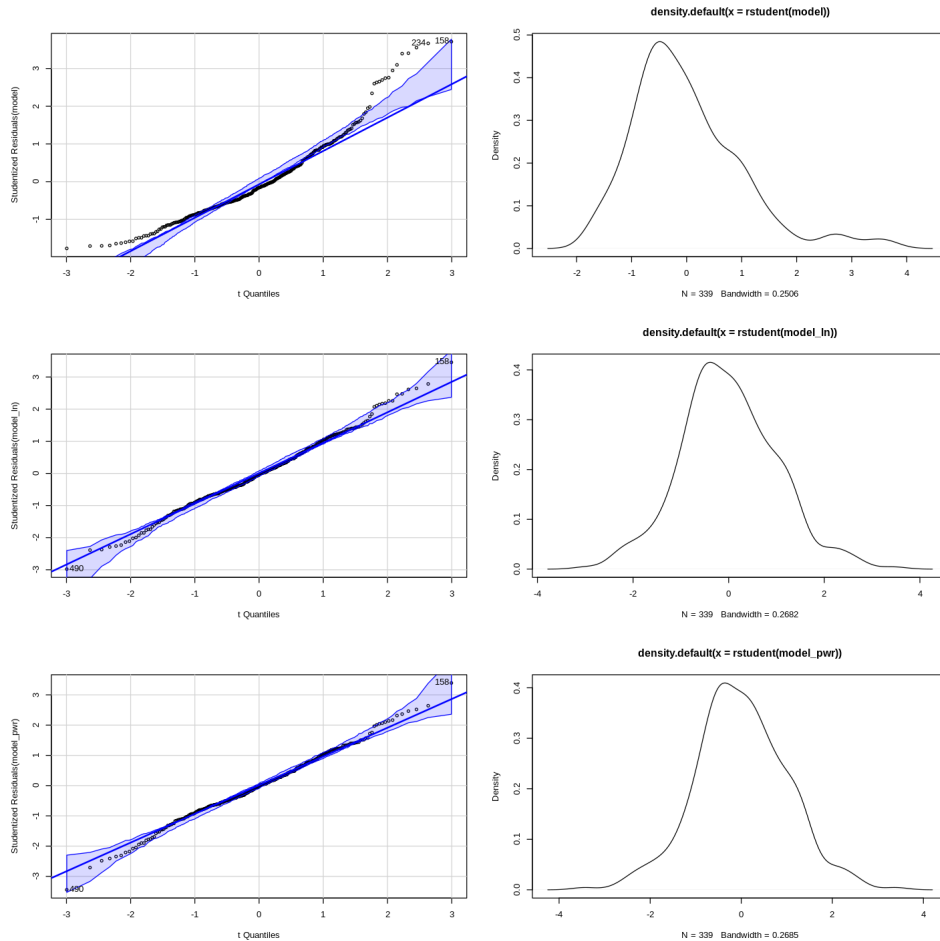


```
fig(12,12)
op <- par(mfrow=c(2,3))
plot(model, which=2, pch = 20, col = "plum3", lwd = 2, sub = "Bez transf.")
plot(model_ln, which=2, pch = 20, col = "plum3", lwd = 2, sub = "Logaritmickej transform.")
plot(model_pwr, which=2, pch = 20, col = "plum3", lwd = 2, sub = "Mocninná transform.")
hist(resid(model), breaks=30, col = "plum3", sub = "Bez transf.")
hist(resid(model_ln), breaks=30, col = "plum3", sub = "Logaritmickej transform.")
hist(resid(model_pwr), breaks=30, col = "plum3", sub = "Mocninná transform.")
```

```
# compare models with and without log transform
op <- par(mfrow=c(3,2))
qqPlot(model, simulate=TRUE, line="quartiles")
plot(density(rstudent(model)))
qqPlot(model_ln, simulate=TRUE, line="quartiles")
plot(density(rstudent(model_ln)))
qqPlot(model_pwr, simulate=TRUE, line="quartiles")
plot(density(rstudent(model_pwr)))
par(op)
```

```
158:      154 234:      210
158:      154 490:      323
158:      154 490:      323
```



Aplikujeme príslušné testy, aby sme prípadne vyvrátili naše domnienky z grafov (podotknime, že v praxi často stačí záver na základe Q-Qplotov, či absencia kónusu reziduálneho scatterplotu).

Pre test normality využijeme Lillieforsov, Shapiro-Wilkov a Anderson-Darlingov test s levelom signifikantnosti $\alpha=5\%$. Pre netransformovaný model dostávame veľmi nízke p-hodnoty (pod 10^{-7}) pre všetky testy a teda zamietame hypotézu o normalite dát. Pre logaritmickej transformácii máme hodnoty nad 0.077, čiže nemáme dostatok dôkazov na zamietnutie hypotézy o normalite dát a pre mocninnú transformáciu sú p-hodnoty o niečo vyššie a opäť nemožno hypotézu normality zamietnuť.

```
#Normality
print("*****Orig*****")
lillie.test(resid(model)) # Lilliefors test
shapiro.test(resid(model)) # Shapiro-Wilk test
ad.test(resid(model))
print("*****Log*****")
lillie.test(resid(model_ln)) # Lilliefors test
shapiro.test(resid(model_ln)) # Shapiro-Wilk test
ad.test(resid(model_ln))
print("*****Power*****")
lillie.test(resid(model_pwr)) # Lilliefors test
shapiro.test(resid(model_pwr)) # Shapiro-Wilk test
ad.test(resid(model_pwr))

[1] "*****Orig*****"

      Lilliefors (Kolmogorov-Smirnov) normality test

data: resid(model)
D = 0.088726, p-value = 9.4e-07

      Shapiro-Wilk normality test

data: resid(model)
W = 0.93378, p-value = 3.868e-11

      Anderson-Darling normality test

data: resid(model)
A = 5.0044, p-value = 2.136e-12
[1] "*****Log*****"

      Lilliefors (Kolmogorov-Smirnov) normality test

data: resid(model_ln)
D = 0.037909, p-value = 0.2774

      Shapiro-Wilk normality test

data: resid(model_ln)
W = 0.99375, p-value = 0.1755

      Anderson-Darling normality test

data: resid(model_ln)
A = 0.67475, p-value = 0.07742
[1] "*****Power*****"

      Lilliefors (Kolmogorov-Smirnov) normality test

data: resid(model_pwr)
D = 0.035125, p-value = 0.3921

      Shapiro-Wilk normality test

data: resid(model_pwr)
W = 0.99577, p-value = 0.4948

      Anderson-Darling normality test

data: resid(model_pwr)
A = 0.49608, p-value = 0.2121
```

Rovnako potrebujeme otestovať homoskedasticitu dát. Využijeme Breusch-Pagan test, s nutnosťou normálne rozdelených rezidií. Testujeme len transformované modely. Dostávame hodnoty 0.79 pre logaritmickej a 0.24 pre mocninnú transformáciu v dôsledku čoho pri $\alpha=5\%$ nezamietame homoskedasticitu.

Test homoskedasticity vyšiel významnejšie pre logaritmickej transformáciu, čo spolu s úzusom z praxe pre λ z konfidenčného intervalu 0 a ľahšou interpretovateľnosťou modelu s touto transformáciou považujeme za vhodné ďalej pracovať práve s logaritmickej transformovaným modelom

```
# for dependence on mean value (residual variance on the response expectation)
# Breusch-Pagan test statistic - library lmtest
```

```
print("*****Log*****")
bptest(model_ln)
print("*****Power*****")
bptest(model_pwr)

[1] "*****Log*****"

studentized Breusch-Pagan test

data: model_ln
BP = 0.069204, df = 1, p-value = 0.7925
[1] "*****Power*****"

studentized Breusch-Pagan test

data: model_pwr
BP = 1.3695, df = 1, p-value = 0.2419
```

cv_8 help

[] ↪ 1 skrytá bunka

.

[] ↪ 34 skrytých buniek

▼ Vícerozměrný regresní model

• Otázka 7

Zkonstruujte lineární model s logaritmicke transformovanou odezvou `medv` a zkuste najít vztah mezi cenou a dalšími nezávislými proměnnými, které máte k dispozici (stačí aditivní model bez interakcí). Na základě kritérií jako jsou AIC, BIC, R^2 , F, atd. vyberte podle vás nejvhodnější model. Lze vztah mezi `indus` a `medv`, pokud existuje, vysvětlit pomocí jiných proměnných? Tj, že například v oblastech s větším zastoupením velkoobchodu a průmyslu bydlí chudší lidé, je tam větší znečištění, nebo větší kriminalita atd.?

```
head(Boston_data)
```

A data.frame: 6 × 13

	crim	zn	indus	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.00632	18	2.31	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
2	0.02731	0	7.07	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
3	0.02729	0	7.07	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
4	0.03237	0	2.18	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5	0.06905	0	2.18	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
6	0.02985	0	2.18	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7

V prvom kroku vytvoríme model logaritmickej odozvy zo všetkých premenných. Tento model nadobúda $R^2 = 0.8553$. Štatisticky významnými premennými sú okrem interceptu `rm`, `age`, `dis`, `rad`, `tax`, `ptratio`, `black`, `lstat`. Podotkneme, že premenná `indus`, ktorú sme predtým používali pre modelovanie ceny je v modeli popri ostatných premenných najnevýznamnejšia.

```
#lm(Y ~ .,data=Z)
m_all=lm(log(medv)~.,data = Boston_data)
summary(m_all)
```

```
Call:
lm(formula = log(medv) ~ ., data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.50582 -0.06369 -0.00784  0.06203  0.37730

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.195e+00  1.811e-01  12.120 < 2e-16 ***
crim         1.069e-02  1.711e-02   0.625 0.532657
zn          6.503e-04  3.579e-04   1.817 0.070098 .
indus       -3.146e-04  1.591e-03  -0.198 0.843324
```

Skontrolujeme multikolinearitu pomocou funkcie `vif()`. Hodnoty sú nízke (< 10) a teda môžeme prejsť k modelu s redukovanými premennými na významné.

```
---
1  4.001  0.00000000  4.001  0.00000000  0.000  0.00000000 ***
vif(m_all)

crim: 2.63767178830982 zn: 2.27357678979342 indus: 2.68027851651607 nox: 4.83322630619225
rm: 2.40934398190735 age: 2.74633818593881 dis: 3.36440609708597 rad: 1.15345051016351
tax: 1.51645012888732 ptratio: 1.48377710681079 black: 1.3501237238269 lstat: 0.00000000000000000000
```

Vytvoríme model s významnými premennými. hodnota R^2 zostáva rovnaká.

```
Multiple R-squared:  0.8553    Adjusted R-squared:  0.85
m_more=lm(log(medv)~rm+ age+ dis+ rad+ tax+ ptratio+ black+ lstat,data=Boston_data)
summary(m_more)
```

```
Call:
lm(formula = log(medv) ~ rm + age + dis + rad + tax + ptratio +
    black + lstat, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.50965 -0.06750 -0.01092  0.06310  0.38334

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.096e+00  1.451e-01  14.446 < 2e-16 ***
rm           2.787e-01  1.519e-02  18.345 < 2e-16 ***
age          -2.409e-03  3.398e-04  -7.088 8.28e-12 ***
dis          -2.715e-02  3.891e-03  -6.976 1.67e-11 ***
rad           1.346e-02  3.890e-03   3.460 0.000611 ***
tax          -5.883e-04  8.234e-05  -7.145 5.81e-12 ***
ptratio      -2.666e-02  3.005e-03  -8.872 < 2e-16 ***
black         6.257e-04  1.736e-04   3.604 0.000362 ***
lstat        -8.807e-03  1.822e-03  -4.833 2.06e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1134 on 330 degrees of freedom
Multiple R-squared:  0.8533,    Adjusted R-squared:  0.8498
F-statistic: 240 on 8 and 330 DF, p-value: < 2.2e-16
```

Vytvorili sme model s 8 premennými (regresormi) s R^2 o hodnote 0.8533 a smerodajnou odchylkou 0.1134. Tento model je pomerne zložitý, keďže pozostáva z 8 regresorov.

Pokusíme sa preto vybrať podmnožinu regresorov, ktorá bude dobre popisovať model. "Ideálny model" by mal mať najmenší možný počet regresorov, ktorý umožňuje adekvátnu interpretáciu (alebo predikciu) --prednáška 01RAD

Použijeme teda metódu postupnej (stepwise) regresie. V každom kroku pridám premennú, a potom skontrolujeme či možno nejakú odobrať. Vyskúšame si kritérium na základe F štatistiky a BIC štatistiky.

```
intercept_only<-lm(log(medv)~1 ,data = Boston_data)
#summary(intercept_only)
AIC(intercept_only)

131.775971587473
```

Stepwise regression with F stat criterion

Vo výpise v bunke nižšie za základe F-value vidíme, že v každom kroku je pridanie ďalšej premennej štatisticky významné. V procese nie je navrhnuté akúkoľvek premennú odobrať.

Pripomíname, že nechceme príliš komplikovaný model. Po konzultácii úlohy s Bc. Eliškou Pečenkovou sa zameriame na modely s približne polovicou regresorov.

Ak vezmeme 3 regresory: rm, lstat, ptratio, dostávame $R^2 = 0.7928$ a $s_n = 0.1339$. V ďalšom kroku nám metóda radí pridať premennú tax (4 regresory: rm, lstat, ptratio, tax), hodnota R^2 stúpa na 0.8141 a s_n klesá na 0.1269.

Pozreli sme sa na prípad, kedy by sme miesto `tax` pridali premennú `age` a dostali by sme $R^2 = 0.8028$ a $s_n = 0.1307$.

Ak by sme chceli 5 regresorov, dostávame sa na $R^2 = 0.8216$ a $s_n = 0.1245$ pre `rm`, `lstat`, `ptratio`, `tax`, `age`. Ak by sme `age` nahradili premennou `black`, dostávame sa na $R^2 = 0.8202$ a $s_n = 0.125$

```
bothF <- step(intercept_only, direction='both', scope=list(lower=intercept_only, upper=m_more),test="F")
```

```
Start: AIC=-832.26
log(medv) ~ 1
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ rm	1	20.3201	8.6147	-1240.99	794.9043	< 2.2e-16 ***
+ lstat	1	15.9149	13.0199	-1100.98	411.9334	< 2.2e-16 ***
+ ptratio	1	5.4113	23.5235	-900.45	77.5229	< 2.2e-16 ***
+ tax	1	4.9755	23.9592	-894.23	69.9832	1.607e-15 ***
+ age	1	4.7150	24.2198	-890.56	65.6050	1.021e-14 ***
+ black	1	1.9622	26.9725	-854.07	24.5168	1.167e-06 ***
+ dis	1	0.9623	27.9725	-841.73	11.5932	0.0007418 ***
+ rad	1	0.1970	28.7377	-832.58	2.3107	0.1294268
<none>			28.9348	-832.26		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=-1240.99
log(medv) ~ rm
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ lstat	1	1.5648	7.0499	-1306.94	74.5763	2.388e-16 ***
+ age	1	1.5309	7.0838	-1305.32	72.6157	5.394e-16 ***
+ tax	1	1.0861	7.5286	-1284.67	48.4731	1.767e-11 ***
+ ptratio	1	1.0262	7.5885	-1281.99	45.4376	6.875e-11 ***
+ black	1	0.3421	8.2726	-1252.72	13.8934	0.000227 ***
+ dis	1	0.0672	8.5475	-1241.64	2.6416	0.105035
<none>			8.6147	-1240.99		
+ rad	1	0.0308	8.5839	-1240.20	1.2056	0.272996
- rm	1	20.3201	28.9348	-832.26	794.9043	< 2.2e-16 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=-1306.94
log(medv) ~ rm + lstat
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ ptratio	1	1.0443	6.0056	-1359.3	58.2545	2.420e-13 ***
+ tax	1	0.6337	6.4163	-1336.9	33.0840	1.990e-08 ***
+ age	1	0.3712	6.6788	-1323.3	18.6176	2.105e-05 ***
+ black	1	0.1300	6.9200	-1311.2	6.2928	0.01260 *
+ dis	1	0.0634	6.9866	-1308.0	3.0384	0.08223 .
+ rad	1	0.0536	6.9964	-1307.5	2.5648	0.11021
<none>			7.0499	-1306.9		
- lstat	1	1.5648	8.6147	-1241.0	74.5763	2.388e-16 ***
- rm	1	5.9699	13.0199	-1101.0	284.5265	< 2.2e-16 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=-1359.29
log(medv) ~ rm + lstat + ptratio
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ tax	1	0.6274	5.3782	-1394.7	38.9625	1.310e-09 ***
+ age	1	0.3002	5.7054	-1374.7	17.5765	3.539e-05 ***
+ black	1	0.2779	5.7277	-1373.3	16.2039	7.042e-05 ***
+ dis	1	0.0688	5.9368	-1361.2	3.8690	0.05001 .
+ rad	1	0.0409	5.9647	-1359.6	2.2892	0.13122
<none>			6.0056	-1359.3		

```
m_3=lm(log(medv) ~ rm + lstat + ptratio, data = Boston_data)
summary(m_3)
m_4=lm(log(medv) ~ rm + lstat + ptratio+ tax, data = Boston_data)
summary(m_4)
m_5=lm(log(medv) ~ rm + lstat + ptratio+ tax + age, data = Boston_data)
summary(m_5)
```

```

Call:
lm(formula = log(medv) ~ rm + lstat + ptratio, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.66583 -0.07442  0.00392  0.07128  0.60381

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.063605   0.146692   14.068 < 2e-16 ***
rm           0.271148   0.016943   16.004 < 2e-16 ***
lstat       -0.016038   0.001707   -9.397 < 2e-16 ***
ptratio     -0.026638   0.003490   -7.632 2.42e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1339 on 335 degrees of freedom
Multiple R-squared:  0.7924,    Adjusted R-squared:  0.7906
F-statistic: 426.3 on 3 and 335 DF,  p-value: < 2.2e-16

Call:
lm(formula = log(medv) ~ rm + lstat + ptratio + tax, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.55794 -0.07651  0.00066  0.06573  0.53350

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.255e+00  1.424e-01  15.839 < 2e-16 ***
rm           2.649e-01  1.609e-02  16.469 < 2e-16 ***
lstat       -1.386e-02  1.655e-03  -8.373 1.56e-15 ***
ptratio     -2.656e-02  3.308e-03  -8.029 1.69e-14 ***
tax         -5.542e-04  8.878e-05  -6.242 1.31e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1269 on 334 degrees of freedom
Multiple R-squared:  0.8141,    Adjusted R-squared:  0.8119
F-statistic: 365.7 on 4 and 334 DF,  p-value: < 2.2e-16

Call:
lm(formula = log(medv) ~ rm + lstat + ptratio + tax + age, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.51713 -0.06778 -0.00273  0.06701  0.43257

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.255e+00  1.424e-01  15.839 < 2e-16 ***
rm           2.649e-01  1.609e-02  16.469 < 2e-16 ***
lstat       -1.386e-02  1.655e-03  -8.373 1.56e-15 ***
ptratio     -2.656e-02  3.308e-03  -8.029 1.69e-14 ***
tax         -5.542e-04  8.878e-05  -6.242 1.31e-09 ***
age          0.000000e+00  0.000000e+00  0.000 1.00e+00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(lm(log(medv) ~ rm + lstat + ptratio + age, data = Boston_data))
summary(lm(log(medv) ~ rm + lstat + ptratio + tax + black, data = Boston_data))

```

```
Call:
lm(formula = log(medv) ~ rm + lstat + ptratio + age, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.66620 -0.06827  0.00297  0.07446  0.49858

Coefficients:
```

Stepwise regression with BIC

Využili sme výber na základe Bayesovského informačného kritéria, ktorého vlastnosťou je pokutovať so zvyšujúcim sa počtom regresorov.

Tento postup nám dal úplne rovnaké riešenie ako použitie F štatistiky, preto riešenie nebudeme popisovať a odkazujeme sa o bunku s popisom pre F value

```
bothBIC <- step(intercept_only, direction='both', scope=list(lower=intercept_only, upper=m_more),k=log(nobs(m_more)))
```

```
Start: AIC=-828.44
log(medv) ~ 1

      Df Sum of Sq  RSS    AIC
+ rm    1   20.3201  8.6147 -1233.34
+ lstat  1   15.9149 13.0199 -1093.33
+ ptratio 1    5.4113 23.5235  -892.80
+ tax    1    4.9755 23.9592  -886.58
+ age    1    4.7150 24.2198  -882.91
+ black  1    1.9622 26.9725  -846.42
+ dis    1    0.9623 27.9725  -834.08
<none>                 28.9348  -828.44
+ rad    1    0.1970 28.7377  -824.93
```

```
Step: AIC=-1233.34
log(medv) ~ rm

      Df Sum of Sq  RSS    AIC
+ lstat  1    1.5648  7.0499 -1295.46
+ age    1    1.5309  7.0838 -1293.84
+ tax    1    1.0861  7.5286 -1273.19
+ ptratio 1    1.0262  7.5885 -1270.51
+ black  1    0.3421  8.2726 -1241.25
<none>                 8.6147 -1233.34
+ dis    1    0.0672  8.5475 -1230.16
+ rad    1    0.0308  8.5839 -1228.72
- rm     1   20.3201 28.9348  -828.44
```

```
Step: AIC=-1295.46
log(medv) ~ rm + lstat

      Df Sum of Sq  RSS    AIC
+ ptratio 1    1.0443  6.0056 -1344.0
+ tax     1    0.6337  6.4163 -1321.6
+ age     1    0.3712  6.6788 -1308.0
+ black   1    0.1300  6.9200 -1296.0
<none>                 7.0499 -1295.5
+ dis     1    0.0634  6.9866 -1292.7
+ rad     1    0.0536  6.9964 -1292.2
- lstat   1    1.5648  8.6147 -1233.3
- rm      1    5.9699 13.0199 -1093.3
```

```
Step: AIC=-1343.99
log(medv) ~ rm + lstat + ptratio

      Df Sum of Sq  RSS    AIC
+ tax     1    0.6274  5.3782 -1375.6
+ age     1    0.3002  5.7054 -1355.5
+ black   1    0.2779  5.7277 -1354.2
<none>                 6.0056 -1344.0
+ dis     1    0.0688  5.9368 -1342.1
+ rad     1    0.0409  5.9647 -1340.5
- ptratio 1    1.0443  7.0499 -1295.5
- lstat   1    1.5829  7.5885 -1270.5
- rm      1    4.5915 10.5971 -1157.3
```

```
Step: AIC=-1375.57
log(medv) ~ rm + lstat + ptratio + tax
```

Model

Našou úlohou je vybrať čo najvhodnejší model. Rozhodli sme sa vybrať model so 4 regresormi (tj polovicu zo všetkých významných). Do modelu zakomponujeme premenné `rm`, `lstat`, `ptratio`, `tax` a teda model má $R^2 = 0.8141$ a $s_n = 0.1269$.

Rozhodli sme sa uprednostniť premennú `tax` pred premennou `black`, hoci by mal model o niečo vyššiu R^2 a to z dôvodu, že nechceme do modelu zahrňať prvky rasizmu.

Na záver zostáva preskúmať premennú `indus`.

Skúsime vytvoriť model prostredníctvom stepwise regresie s východným modelom postavenom na premennej `indus`, ktorý má $R^2 = 0.2607$ a premenná je veľmi významná.

Keď k modelu pridáme premennú `rm`, obe premenné sú významné a kvalita modelu sa zvýši razantne, teda neuvažujeme vzťah týchto dvoch premenných.

Pridaním premennej `lstat` sa významnosť `indus` zníži, ak navyše volíme interakciu `lstat:indus`, potom premenná `indus` je nevýznamná. Z tohto pozorovania možno usúdiť, že v oblastiach fabrík a skladov býva viac ľudí s nižším sociálnym postavením. Navyše vidíme aj vyšší korelačný koeficient medzi premennými.

Kombinácia `indus`, `lstat` a `tax` zaistiť nevýznamnosť premennej `indus` v modeli. Povšimneme si, že všetky koeficienty sú záporné a teda ak sa zvýši nemaloobchodné podnikanie (továrne, sklady), daň z nehnuteľnosti sa zníži.

Ak vytvoríme model len z premennej `crim`, táto premenná je štatisticky významná, avšak ak pridáme premennú `indus` tak kriminalita už nemá štatistický význam pre náš model a teda platí, že v oblasti s vyšším počtom neobchodných území klesá kriminalita (oba koeficienty majú záporné znamienka, ak jednému uberieme, druhému sa pridá). Navyše vidíme aj vyšší korelačný koeficient medzi premennými.

Ak vytvoríme model v závislosti na `indus`, `nox` potom premenná `nox` už je automaticky štatisticky nevýznamná a medzi premennými vidíme spojitosť. Navyše vidíme aj vyšší korelačný koeficient medzi premennými.

```
m_indus<- step(model_ln, direction='both', scope=list(lower=model_ln, upper=m_all),test="F")
```

```
Start: AIC=-932.67
log(medv) ~ indus
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ rm	1	13.5059	7.8847	-1269.00	575.5417	< 2.2e-16 ***
+ lstat	1	8.6772	12.7134	-1107.05	229.3290	< 2.2e-16 ***
+ ptratio	1	3.3781	18.0125	-988.94	63.0140	3.102e-14 ***
+ tax	1	0.9878	20.4028	-946.70	16.2674	6.812e-05 ***
+ age	1	0.7223	20.6683	-942.32	11.7431	0.0006865 ***
+ dis	1	0.6664	20.7242	-941.40	10.8051	0.0011192 **
+ zn	1	0.4911	20.8995	-938.55	7.8959	0.0052450 **
+ black	1	0.4314	20.9592	-937.58	6.9156	0.0089377 **
+ crim	1	0.1697	21.2209	-933.37	2.6866	0.1021299
+ rad	1	0.1532	21.2374	-933.11	2.4235	0.1204687
<none>			21.3906	-932.67		
+ nox	1	0.0491	21.3415	-931.45	0.7732	0.3798632

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Step: AIC=-1269
log(medv) ~ indus + rm
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ lstat	1	0.9668	6.9179	-1311.35	46.8167	3.718e-11 ***
+ ptratio	1	0.9175	6.9672	-1308.94	44.1161	1.252e-10 ***
+ age	1	0.8789	7.0058	-1307.07	42.0256	3.226e-10 ***
+ tax	1	0.5519	7.3328	-1291.60	25.2135	8.357e-07 ***
+ black	1	0.1614	7.7233	-1274.02	7.0027	0.008523 **
+ dis	1	0.0938	7.7909	-1271.06	4.0348	0.045373 *
<none>			7.8847	-1269.00		
+ rad	1	0.0332	7.8515	-1268.43	1.4163	0.234848
+ zn	1	0.0241	7.8606	-1268.04	1.0257	0.311902
+ nox	1	0.0044	7.8803	-1267.19	0.1856	0.666869
+ crim	1	0.0006	7.8841	-1267.03	0.0254	0.873499
- rm	1	13.5059	21.3906	-932.67	575.5417	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Step: AIC=-1311.35
log(medv) ~ indus + rm + lstat
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ ptratio	1	0.9937	5.9242	-1361.9	56.0260	6.408e-13 ***
+ tax	1	0.5046	6.4134	-1335.0	26.2771	5.018e-07 ***
+ age	1	0.2678	6.6501	-1322.7	13.4507	0.0002849 ***
+ dis	1	0.2599	6.6580	-1322.3	13.0395	0.0003518 ***
+ black	1	0.0925	6.8255	-1313.9	4.5241	0.0341539 *
+ nox	1	0.0535	6.8644	-1312.0	2.6044	0.1075102
+ rad	1	0.0516	6.8663	-1311.9	2.5091	0.1141390
+ crim	1	0.0515	6.8664	-1311.9	2.5067	0.1143091
<none>			6.9179	-1311.3		
+ zn	1	0.0027	6.9153	-1309.5	0.1287	0.7200027
- lstat	1	0.9668	7.8847	-1269.0	46.8167	3.718e-11 ***
- rm	1	5.7955	12.7134	-1107.0	280.6448	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Step: AIC=-1361.92
```



```
summary(model_1n)
print("*****")
print("*****")
summary(lm(log(medv) ~ indus + rm,data=Boston_data))
print("*****")
print("*****")
summary(lm(log(medv) ~ indus + lstat,data=Boston_data))
print("*****")
print("*****")
summary(lm(log(medv) ~ indus +lstat:indus,data=Boston_data))
print("*****")
print("*****")
summary(lm(log(medv) ~ indus + lstat + tax , data = Boston_data))
print("*****")
print("*****")
summary(lm(log(medv) ~ indus + crim,data=Boston_data))
print("*****")
print("*****")
summary(lm(log(medv) ~ crim,data=Boston_data))
```



```

Call:
lm(formula = log(medv) ~ indus, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.73082 -0.15957 -0.01244  0.16211  0.85196

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.33005     0.02293   145.2  <2e-16 ***
indus       -0.02355     0.00216   -10.9  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2519 on 337 degrees of freedom
Multiple R-squared:  0.2607,    Adjusted R-squared:  0.2585
F-statistic: 118.9 on 1 and 337 DF,  p-value: < 2.2e-16
[1] "*****"
[1] "*****"

Call:
lm(formula = log(medv) ~ indus + rm, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.69044 -0.07901  0.00234  0.09999  0.47767

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.868217     0.103560    8.384 1.43e-15 ***
indus       -0.008154     0.001462   -5.577 5.03e-08 ***
rm          0.368364     0.015355   23.990 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1532 on 336 degrees of freedom
Multiple R-squared:  0.7275,    Adjusted R-squared:  0.7259
F-statistic: 448.5 on 2 and 336 DF,  p-value: < 2.2e-16
[1] "*****"
[1] "*****"

Call:
lm(formula = log(medv) ~ indus + lstat, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.72496 -0.13295 -0.02117  0.11262  0.68351

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.535184     0.022290  158.600 <2e-16 ***
indus       -0.005802     0.002039   -2.846  0.0047 **
lstat      -0.033563     0.002216  -15.144 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1945 on 336 degrees of freedom
Multiple R-squared:  0.5606,    Adjusted R-squared:  0.558
F-statistic: 214.4 on 2 and 336 DF,  p-value: < 2.2e-16
[1] "*****"
[1] "*****"

Call:
lm(formula = log(medv) ~ indus + lstat:indus, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.7042 -0.1421 -0.0377  0.1446  0.6491

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.2489359  0.0230533  140.932 < 2e-16 ***
indus        0.0068005  0.0041387   1.643  0.101
indus:lstat -0.0015890  0.0001906  -8.338 1.97e-15 ***
---

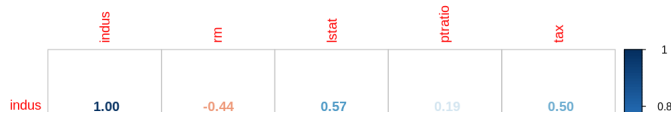
```

summary(lm(log(medv)~indus+nox,data=Boston_data))

```
Call:
lm(formula = log(medv) ~ indus + nox, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.74617 -0.15744 -0.01244  0.15809  0.84944

Coefficients:
            1            2            3            4            5            6            7            8            9            10            11            12            13            14            15            16            17            18            19            20            21            22            23            24            25            26            27            28            29            30            31            32            33            34            35            36            37            38            39            40            41            42            43            44            45            46            47            48            49            50            51            52            53            54            55            56            57            58            59            60            61            62            63            64            65            66            67            68            69            70            71            72            73            74            75            76            77            78            79            80            81            82            83            84            85            86            87            88            89            90            91            92            93            94            95            96            97            98            99            100            101            102            103            104            105            106            107            108            109            110            111            112            113            114            115            116            117            118            119            120            121            122            123            124            125            126            127            128            129            130            131            132            133            134            135            136            137            138            139            140            141            142            143            144            145            146            147            148            149            150            151            152            153            154            155            156            157            158            159            160            161            162            163            164            165            166            167            168            169            170            171            172            173            174            175            176            177            178            179            180            181            182            183            184            185            186            187            188            189            190            191            192            193            194            195            196            197            198            199            200            201            202            203            204            205            206            207            208            209            210            211            212            213            214            215            216            217            218            219            220            221            222            223            224            225            226            227            228            229            230            231            232            233            234            235            236            237            238            239            240            241            242            243            244            245            246            247            248            249            250            251            252            253            254            255            256            257            258            259            260            261            262            263            264            265            266            267            268            269            270            271            272            273            274            275            276            277            278            279            280            281            282            283            284            285            286            287            288            289            290            291            292            293            294            295            296            297            298            299            300            301            302            303            304            305            306            307            308            309            310            311            312            313            314            315            316            317            318            319            320            321            322            323            324            325            326            327            328            329            330            331            332            333            334            335            336            337            338            339            340            341            342            343            344            345            346            347            348            349            350            351            352            353            354            355            356            357            358            359            360            361            362            363            364            365            366            367            368            369            370            371            372            373            374            375            376            377            378            379            380            381            382            383            384            385            386            387            388            389            390            391            392            393            394            395            396            397            398            399            400            401            402            403            404            405            406            407            408            409            410            411            412            413            414            415            416            417            418            419            420            421            422            423            424            425            426            427            428            429            430            431            432            433            434            435            436            437            438            439            440            441            442            443            444            445            446            447            448            449            450            451            452            453            454            455            456            457            458            459            460            461            462            463            464            465            466            467            468            469            470            471            472            473            474            475            476            477            478            479            480            481            482            483            484            485            486            487            488            489            490            491            492            493            494            495            496            497            498            499            500            501            502            503            504            505            506            507            508            509            510            511            512            513            514            515            516            517            518            519            520            521            522            523            524            525            526            527            528            529            530            531            532            533            534            535            536            537            538            539            540            541            542            543            544            545            546            547            548            549            550            551            552            553            554            555            556            557            558            559            560            561            562            563            564            565            566            567            568            569            570            571            572            573            574            575            576            577            578            579            580            581            582            583            584            585            586            587            588            589            590            591            592            593            594            595            596            597            598            599            600            601            602            603            604            605            606            607            608            609            610            611            612            613            614            615            616            617            618            619            620            621            622            623            624            625            626            627            628            629            630            631            632            633            634            635            636            637            638            639            640            641            642            643            644            645            646            647            648            649            650            651            652            653            654            655            656            657            658            659            660            661            662            663            664            665            666            667            668            669            670            671            672            673            674            675            676            677            678            679            680            681            682            683            684            685            686            687            688            689            690            691            692            693            694            695            696            697            698            699            700            701            702            703            704            705            706            707            708            709            710            711            712            713            714            715            716            717            718            719            720            721            722            723            724            725            726            727            728            729            730            731            732            733            734            735            736            737            738            739            740            741            742            743            744            745            746            747            748            749            750            751            752            753            754            755            756            757            758            759            760            761            762            763            764            765            766            767            768            769            770            771            772            773            774            775            776            777            778            779            780            781            782            783            784            785            786            787            788            789            790            791            792            793            794            795            796            797            798            799            800            801            802            803            804            805            806            807            808            809            810            811            812            813            814            815            816            817            818            819            820            821            822            823            824            825            826            827            828            829            830            831            832            833            834            835            836            837            838            839            840            841            842            843            844            845            846            847            848            849            850            851            852            853            854            855            856            857            858            859            860            861            862            863            864            865            866            867            868            869            870            871            872            873            874            875            876            877            878            879            880            881            882            883            884            885            886            887            888            889            890            891            892            893            894            895            896            897            898            899            900            901            902            903            904            905            906            907            908            909            910            911            912            913            914            915            916            917            918            919            920            921            922            923            924            925            926            927            928            929            930            931            932            933            934            935            936            937            938            939            940            941            942            943            944            945            946            947            948            949            950            951            952            953            954            955            956            957            958            959            960            961            962            963            964            965            966            967            968            969            970            971            972            973            974            975            976            977            978            979            980            981            982            983            984            985            986            987            988            989            990            991            992            993            994            995            996            997            998            999            1000            1001            1002            1003            1004            1005            1006            1007            1008            1009            1010            1011            1012            1013            1014            1015            1016            1017            1018            1019            1020            1021            1022            1023            1024            1025            1026            1027            1028            1029            1030            1031            1032            1033            1034            1035            1036            1037            1038            1039            1040            1041            1042            1043            1044            1045            1046            1047            1048            1049            1050            1051            1052            1053            1054            1055            1056            1057            1058            1059            1060            1061            1062            1063            1064            1065            1066            1067            1068            1069            1070            1071            1072            1073            1074            1075            1076            1077            1078            1079            1080            1081            1082            1083            1084            1085            1086            1087            1088            1089            1090            1091            1092            1093            1094            1095            1096            1097            1098            1099            1100            1101            1102            1103            1104            1105            1106            1107            1108            1109            1110            1111            1112            1113            1114            1115            1116            1117            1118            1119            1120            1121            1122            1123            1124            1125            1126            1127            1128            1129            1130            1131            1132            1133            1134            1135            1136            1137            1138            1139            1140            1141            1142            1143            1144            1145            1146            1147            1148            1149            1150            1151            1152            1153            1154            1155            1156            1157            1158            1159            1160            1161            1162            1163            1164            1165            1166            1167            1168            1169            1170            1171            1172            1173            1174            1175            1176            1177            1178            1179            1180            1181            1182            1183            1184            1185            1186            1187            1188            1189            1190            1191            1192            1193            1194            1195            1196            1197            1198            1199            1200            1201            1202            1203            1204            1205            1206            1207            1208            1209            1210            1211            1212            1213            1214            1215            1216            1217            1218            1219            1220            1221            1222            1223            1224            1225            1226            1227            1228            1229            1230            1231            1232            1233            1234            1235            1236            1237            1238            1239            1240            1241            1242            1243            1244            1245            1246            1247            1248            1249            1250            1251            1252            1253            1254            1255            1256            1257            1258            1259            1260            1261            1262            1263            1264            1265            1266            1267            1268            1269            1270            1271            1272            1273            1274            1275            1276            1277            1278            1279            1280            1281            1282            1283            1284            1285            1286            1287            1288            1289            1290            1291            1292            1293            1294            1295            1296            1297            1298            1299            1300            1301            1302            1303            1304            1305            1306            1307            1308            1309            1310            1311            1312            1313            1314            1315            1316            1317            1318            1319            1320            1321            1322            1323            1324            1325            1326            1327            1328            1329            1330            1331            1332            1333            1334            1335            1336            1337            1338            1339            1340            1341            1342            1343            1344            1345            1346            1347            1348            1349            1350            1351            1352            1353            1354            1355            1356            1357            1358            1359            1360            1361            1362            1363            1364            1365            1366            1367            1368            1369            1370            1371            1372            1373            1374            1375            1376            1377            1378            1379            1380            1381            1382            1383            1384            1385            1386            1387            1388            1389            1390            1391            1392            1393            1394            1395            1396            1397            1398            1399            1400            1401            1402            1403            1404            1405            1406            1407            1408            1409            1410            1411            1412            1413            1414            1415            1416            1417            1418            1419            1420            1421            1422            1423            1424            1425            1426            1427            1428            1429            1430            1431            1432            1433            1434            1435            1436            1437            1438            1439            1440            1441            1442            1443            1444            1445            1446            1447            1448            1449            1450            1451            1452            1453            1454            1455            1456            1457            1458            1459            1460            1461            1462            1463            1464            1465            1466            1467            1468            1469            1470            1471            1472            1473            1474            1475            1476            1477            1478            1479            1480            1481            1482            1483            1484            1485            1486            1487            1488            1489            1490            1491            1492            1493            1494            1495            1496            1497            1498            1499            1500            1501            1502            1503            1504            1505            1506            1507            1508            1509            1510            1511            1512            1513            1514            1515            1516            1517            1518            1519            1520            1521            1522            1523            1524            1525            1526            1527            1528            1529            1530            1531            1532            1533            1534            1535            1536            1537            1538            1539            1540            1541            1542            1543            1544            1545            1546            1547            1548            1549            1550            1551            1552            1553            1554            1555            1556            1557            1558            1559            1560            1561            1562            1563            1564            1565            1566            1567            1568            1569            1570            1571            1572            1573            1574            1575            1576            1577            1578            1579            1580            1581            1582            1583            1584            1585            1586            1587            1588            1589            1590            1591            1592            1593            1594            1595            1596            1597            1598            1599            1600            1601            1602            1603            1604            1605            1606            1607            1608            1609            1610            1611            1612            1613            1614            1615            1616            1617            1618            1619            1620            1621            1622            1623            1624            1625            1626            1627            1628            1629            1630            1631            1632            1633            1634            1635            1636            1637            1638            1639            1640            1641            1642            1643            1644            1645            1646            1647            1648            1649            1650            1651            1652            1653            1654            1655            1656            1657            1658            1659            1660            1661            1662            1663            1664            1665            1666            1667            1668            1669            1670            1671            1672            1673            1674            1675            1676            1677            1678            1679            1680            1681            1682            1683            1684            1685            1686            1687            1688            1689            1690            1691            1692            1693            1694            1695            1696            1697            1698            1699            1700            1701            1702            1703            1704            1705            1706            1707            1708            1709            1710            1711            1712            1713            1714            1715            1716            1717            1718            1719            1720            1721            1722            1723            1724            1725            1726            1727            1728            1729            1730            1731            1732            1733            1734            1735            1736            1737            1738            1739            1740            1741            1742            1743            1744            1745            1746            1747            1748            1749            1750            1751            1752            1753            1754            1755            1756            1757            1758            1759            1760            1761            1762            1763            1764            1765            1766            1767            1768            1769            1770            1771            1772            1773            1774            1775            1776            1777            1778            1779            1780            1781            1782            1783            1784            1785            1786            1787            1788            1789            1790            1791            1792            1793            1794            1795            1796            1797            1798            1799            1800            1801            1802            1803            1804            1805            1806            1807            1808            1809            1810            1811            1812            1813            1814            1815            1816            1817            1818            1819            1820            1821            1822            1823            1824            1825            1826            1827            1828            1829            1830            1831            1832            1833            1834            1835            1836            1837            1838            1839            1840            1841            1842            1843            1844            1845            1846            1847            1848            1849            1850            1851            1852            1853            1854            1855            1856            1857            1858            1859            1860            1861            1862            1863            1864            1865            1866            1867            1868            1869            1870            1871            1872            1873            1874            1875            1876            1877            1878            1879            1880            1881            1882            1883            1884            1885            1886            1887            1888            1889            1890            1891            1892            1893            1894            1895            1896            1897            1898            1899            1900            1901            1902            1903            1904            1905            1906            1907            1908            1909            1910            1911            1912            1913            1914            1915            1916            1917            1918            1919            1920            1921            1922            1923            1924            1925            1926            1927            1928            1929            1930            1931            1932            1933            1934            1935            1936            1937            1938            1939            1940            1941            1942            1943            1944            1945            1946            1947            1948            1949            1950            1951            1952            1953            1954            1955            1956            1957            1958            1959            1960            1961            1962            1963            1964            1965            1966            1967            1968            1969            1970            1971            1972            1973            1974            1975            1976            1977            1978            1979            1980            1981            1982            1983            1984            1985            1986            1987            1988            1989            1990            1991            1992            1993            1994            1995            1996            1997            1998            1999            2000            2001            2002            2003            2004            2005            2006            2007            2008            2009            2010            2011            2012            2013            2014            2015            2016            2017            2018            2019            2020            2021            2022            2023            2024            2025            2026            2027            2028            2029            2030            2031            2032            2033            2034            2035            2036            2037            2038            2039            2040            2041            2042            2043            2044            2045            2046            2047            2048            2049            2050            2051            2052            2053            2054            2055            2056            2057            2058            2059            2060            2061            2062            2063            2064            2065            2066            2067            2068            2069            2070            2071            2072            2073            2074            2075            2076            2077            2078            2079            2080            2081            2082            2083            2084            2085            2086            2087            2088            2089            2090            2091            2092            2093            2094            2095            2096            2097            2098            2099            2100            2101            2102            2103            2104            2105            2106            2107            2108            2109            2110            2111            2112            2113            2114            2115            2116            2117            2118            2119            2120            2121            2122            2123            2124            2125            2126            2127            2128            2129            2130            2131            2132            2133            2134            2135            2136            2137            2138            2139            2140            2141            2142            2143            2144            2145            2146            2147            2148            2149            2150            2151            2152            2153            2154            2155            2156            2157            2158            2159            2160            2161            2162            2163            2164            2165            2166            2167            2168            2169            2170            2171            2172            2173            2174            2175            2176            2177            2178            2179            2180            2181            2182            2183            2184            2185            2186            2187            2188            2189            2190            2191            2192            2193            2194            2195            2196            2197            2198            2199            2200            2201            2202            2203            2204            2205            2206            2207            2208            2209            2210            2211            2212            2
```



Model, ktorý sme zvolili v otázke 7) je založený na premenných `rm`, `lstat`, `ptratio` a `tax`. Premenná `indus` najviac koreluje s premennou `lstat` ($p=0.57$), preto vložíme do modelu na jej úkor.

```
summary(model_ln)
```

```
Call:
lm(formula = log(medv) ~ indus, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.73082 -0.15957 -0.01244  0.16211  0.85196

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.33005    0.02293   145.2   <2e-16 ***
indus       -0.02355    0.00216   -10.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2519 on 337 degrees of freedom
Multiple R-squared:  0.2607,    Adjusted R-squared:  0.2585
F-statistic: 118.9 on 1 and 337 DF,  p-value: < 2.2e-16
```

```
m_i_add <- lm(log(medv) ~ rm + indus + ptratio + tax, data = Boston_data)
summary(m_i_add)
```

```
Call:
lm(formula = log(medv) ~ rm + indus + ptratio + tax, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.51729 -0.07809  0.01042  0.08197  0.41848

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.6739569  0.1325167   12.632   < 2e-16 ***
rm           0.3378917  0.0143277   23.583   < 2e-16 ***
indus       -0.0039302  0.0014703   -2.673   0.00789 **
ptratio     -0.0256338  0.0036097   -7.101  7.48e-12 ***
tax         -0.0005877  0.0001051   -5.591  4.69e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1381 on 334 degrees of freedom
Multiple R-squared:  0.7798,    Adjusted R-squared:  0.7772
F-statistic: 295.7 on 4 and 334 DF,  p-value: < 2.2e-16
```

```
m_i_add$coefficients
```

```
(Intercept): 1.6739569053096 rm: 0.337891747423145 indus: -0.00393017913857012 ptratio:
             0.0005877478490002026
```

V pôvodnom jednoduchom modeli pri náraste podielu nemaloobchodných zón o jednotku dôjde k poklesu cien nehnuteľností o 2.33%. V našom novom modeli, ktorý má 4 vysvetľujúce premenné dôjde pri rovnakej jednotkovej zmene premennej `indus` k poklesu o 0.39%.

Teda premenná `indus` má v rozsiahlejšom modeli nižší vplyv na vývoj ceny, než keby bola jedinou vysvetľujúcou premennou.

```
unnname(100*(exp(model_ln$coefficients[2])-1))
unnname(100*(exp(m_i_add$coefficients[3])-1))
```

```
-2.3276413163198
-0.39224660923991
```

• Otázka 9

Prezentujte váš výsledný model pro predikci `medv`, diskutujte výsledné parametry R^2 , σ , F a porovnejte je s jednoduchým lin. modelom z otázky 6. Jak se změnila a dala se tato změna očekávat? Validujte model (jak graficky, tak pomocí příslušných testů hypotéz). Pomocí Partial regression plots a Partial residual plots diskutujte linearitu použitých proměnných.

V predošlej otázke sme vystavali model na základe premenných `indus`, `rm`, `ptratio`, `tax`. Pozrieme sa teraz na základné štatistiky tohto modelu, do zátvorky vždy uvedieme hodnoty pre jednoduchý lineárny model založený len na premennej `indus`.

- $R^2 = 0.7798$ (0.2519), nárast štatistiky sme očakávali, pretože s pridaním premenných sme boli schopný vysvetliť väčšiu časť variability modelu než tomu bolo pri jednoduchom modeli.
- $\sigma = s_n = 0.1381$ (0.2607), pokles smerodajnej odchylky reziduí sa dal opäť očakávať, pretože s pridaním významných premenných sa dáta dajú lepšie modelovať a teda sú bližšie rozptýlené a linear regression fit sa zlepšuje.
- $F - stat = 295.7$ s $p - val = 2.2 \cdot 10^{-16}$ (118.9; $2.2 \cdot 10^{-16}$), zvýšenie sme opäť predpokladali a p-value < 0.05 nám naznačuje že aspoň 1 nezávislá premenná vysvetľuje Y. (If the p-value associated with the F-statistic < 0.05: Then, AT LEAST 1 independent variable is related to Y.)

```
summary(model_ln)
```

```
Call:
lm(formula = log(medv) ~ indus, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.73082 -0.15957 -0.01244  0.16211  0.85196

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.33005     0.02293   145.2  <2e-16 ***
indus       -0.02355     0.00216   -10.9  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2519 on 337 degrees of freedom
Multiple R-squared:  0.2607,    Adjusted R-squared:  0.2585
F-statistic: 118.9 on 1 and 337 DF,  p-value: < 2.2e-16
```

```
summary(m_i_add)
```

```
Call:
lm(formula = log(medv) ~ rm + indus + ptratio + tax, data = Boston_data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.51729 -0.07809  0.01042  0.08197  0.41848

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.6739569  0.1325167   12.632  < 2e-16 ***
rm           0.3378917  0.0143277   23.583  < 2e-16 ***
indus       -0.0039302  0.0014703   -2.673  0.00789 **
ptratio     -0.0256338  0.0036097   -7.101  7.48e-12 ***
tax         -0.0005877  0.0001051   -5.591  4.69e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1381 on 334 degrees of freedom
Multiple R-squared:  0.7798,    Adjusted R-squared:  0.7772
F-statistic: 295.7 on 4 and 334 DF,  p-value: < 2.2e-16
```

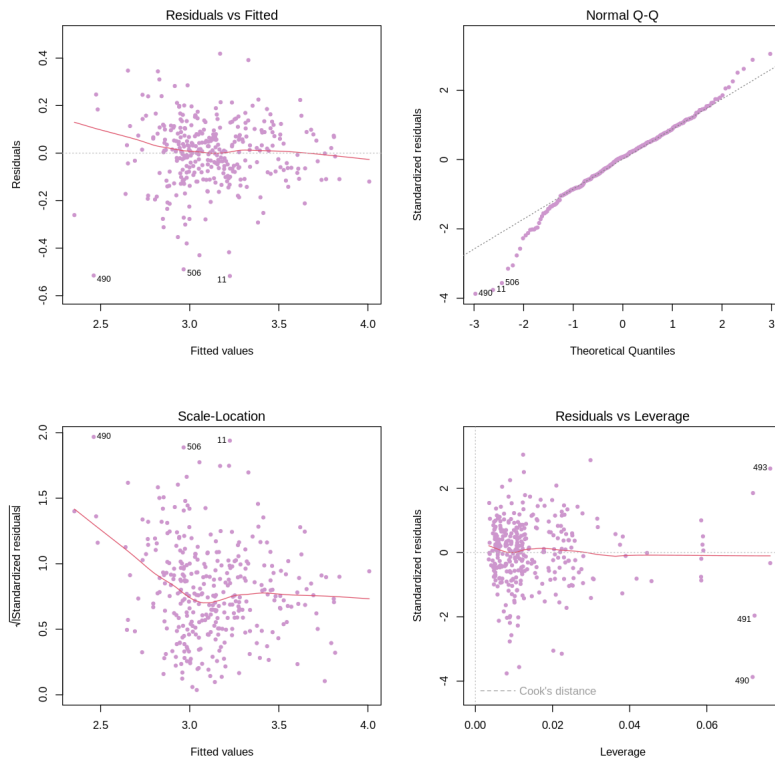
Najprv vybraný model znázorníme graficky.

```
crPlots(m_i_add, col="plum3",pch=20)
```

Component + Residual Plots

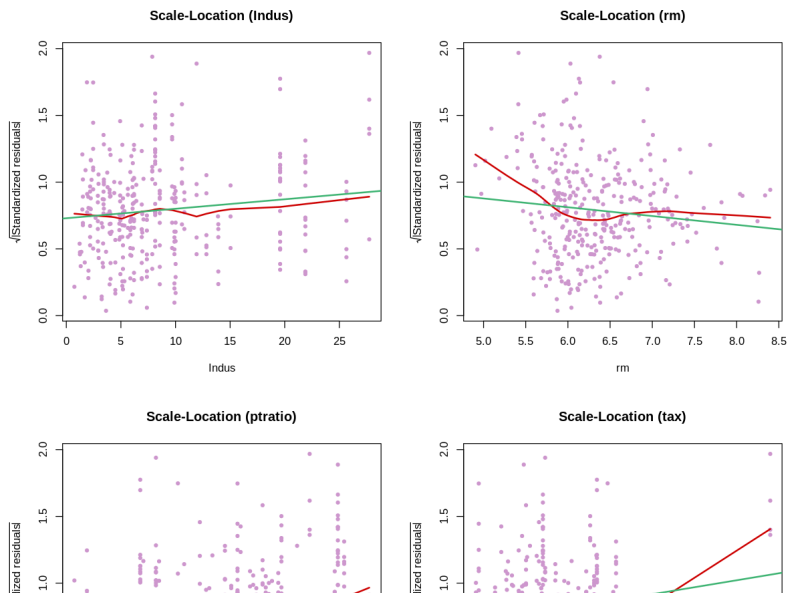
Na obrázku *Residuals vs Fitted* si všimneme, že dáta majú v oblasti 3.0-3.25 x-ovej osi väčší rozptul reziduú ako v oblasti 3.8-4.0. Ďalej na Q-Q plote pozorujeme problém s chvostami rozdelenia, ťažké chvosty nie sú znakom normálneho rozdelenia. Podozrenie na porušenie homoskedasticity a normality otestujeme príslušnými testami.

```
fig(10,10)
par(mfrow=c(2,2))
plot(m_i_add, col="plum3", pch=20)
```



Vykreslíme reziduá v závislosti na jednotlivých premenných. Nepozorujeme však šiadne podozrivé rozloženie reziduú, naopak všimame si rovnomernosť.

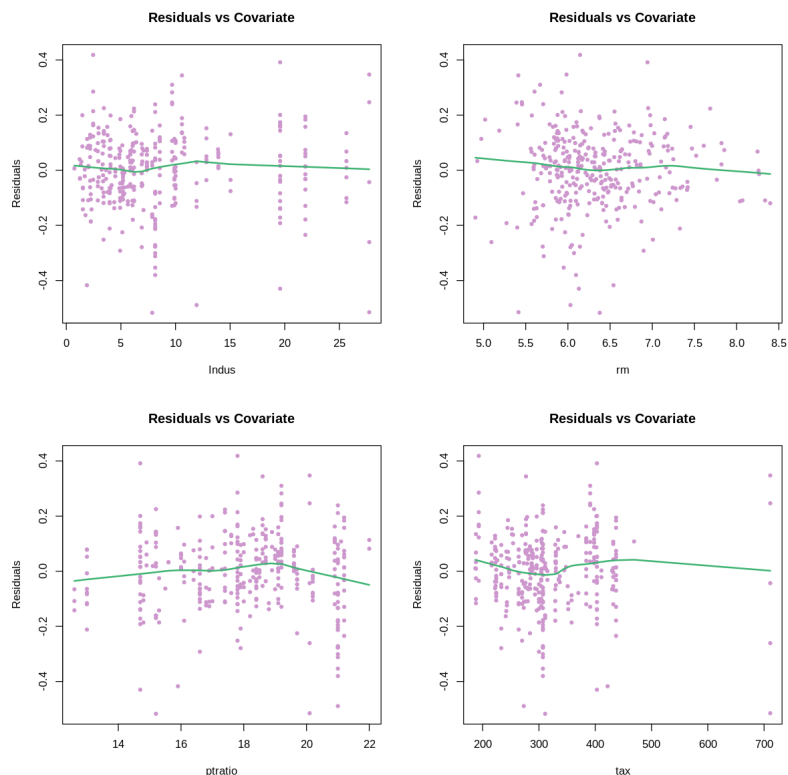
```
fig(10,10)
par(mfrow=c(2,2))
plot(Boston_data$indus, sqrt(abs(rstandard(m_i_add))), pch = 20, col = "plum3", bg = "royalblue4",
     xlab = "Indus", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
     main = "Scale-Location (Indus)")
lines(lowess(Boston_data$indus, sqrt(abs(rstandard(m_i_add)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(m_i_add)))~Boston_data$indus), col = "mediumseagreen", lwd = 2)
plot(Boston_data$rm, sqrt(abs(rstandard(m_i_add))), pch = 20, col = "plum3", bg = "royalblue4",
     xlab = "rm", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
     main = "Scale-Location (rm)")
lines(lowess(Boston_data$rm, sqrt(abs(rstandard(m_i_add)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(m_i_add)))~Boston_data$rm), col = "mediumseagreen", lwd = 2)
plot(Boston_data$ptratio, sqrt(abs(rstandard(m_i_add))), pch = 20, col = "plum3", bg = "royalblue4",
     xlab = "ptratio", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
     main = "Scale-Location (ptratio)")
lines(lowess(Boston_data$ptratio, sqrt(abs(rstandard(m_i_add)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(m_i_add)))~Boston_data$ptratio), col = "mediumseagreen", lwd = 2)
plot(Boston_data$tax, sqrt(abs(rstandard(m_i_add))), pch = 20, col = "plum3", bg = "royalblue4",
     xlab = "tax", ylab = as.expression(substitute(sqrt(abs(yL))), list(yL = as.name("Standardized residuals"))),
     main = "Scale-Location (tax)")
lines(lowess(Boston_data$tax, sqrt(abs(rstandard(m_i_add)))), col = "red3", lwd = 2)
abline(lm(sqrt(abs(rstandard(m_i_add)))~Boston_data$tax), col = "mediumseagreen", lwd = 2)
```



```
fig(10,10)
```

```
par(mfrow=c(2,2))
```

```
plot(Boston_data$indus, residuals(m_i_add), pch = 20, col = "plum3", xlab = "Indus", ylab = "Residuals", main = "Residuals vs Covariate")
lines(lowess(Boston_data$indus, residuals(m_i_add)), col = "mediumseagreen", lwd = 2)
plot(Boston_data$rm, residuals(m_i_add), pch = 20, col = "plum3", xlab = "rm", ylab = "Residuals", main = "Residuals vs Covariate")
lines(lowess(Boston_data$rm, residuals(m_i_add)), col = "mediumseagreen", lwd = 2)
plot(Boston_data$ptratio, residuals(m_i_add), pch = 20, col = "plum3", xlab = "ptratio", ylab = "Residuals", main = "Residuals vs Covariate")
lines(lowess(Boston_data$ptratio, residuals(m_i_add)), col = "mediumseagreen", lwd = 2)
plot(Boston_data$tax, residuals(m_i_add), pch = 20, col = "plum3", xlab = "tax", ylab = "Residuals", main = "Residuals vs Covariate")
lines(lowess(Boston_data$tax, residuals(m_i_add)), col = "mediumseagreen", lwd = 2)
```



Prejdeme k jednotlivým testom.

Normalita

Overíme si, či naše podozrenie porušenia normality reziduí je oprávnené a teda využijeme Lilieforsov, Shapiro-Wilkov a Anderson-Darlingov test normality na hladine $\alpha=0.05$.

Lilieforsov test tesne nezamietajú hypotézu o normalite, ale Shapiro-Wilkov a Anderson Darlingov test normalitu rozdelenia zamietajú. Naše obavy sa potvrdili.

```
#Normality
lillie.test(resid(m_i_add)) # Lilliefors test
shapiro.test(resid(m_i_add)) # Shapiro-Wilk test
ad.test(resid(m_i_add))

Lilliefors (Kolmogorov-Smirnov) normality test

data: resid(m_i_add)
D = 0.047342, p-value = 0.06478

Shapiro-Wilk normality test

data: resid(m_i_add)
W = 0.97415, p-value = 9.161e-06

Anderson-Darling normality test

data: resid(m_i_add)
A = 1.7081, p-value = 0.0002193
```

Homoskedasticita

Chceme otestovať homoskedasticitu. Bohužiaľ nemáme normálne rozdelené reziduá, čo je podmienka Breusch-Paganovho test. Avšak, na hodinách sme si takýto prípad neukazovali a na internete som našla len [Goldfeld-Quandt Tests](#), ktorého neparametrická varianta by teoreticky šla použiť, ale daná knižnica/balíček nejde nainštalovať.

Napriek okolnostiam použijeme `bptest()` na úrovni $\alpha=0.05$.

Testom dostávame $p\text{-val}=0.001838 < \alpha$ a teda zamietame nulovú hypotézu o tom, že by šlo o homoskedasticitu.

```
#NOT WORKING
#install.packages("skedastic")
#library(skedastic)
#goldfeld_quandt(mtcars_lm, deflator = "qsec", method = "nonparametric", restype = "blus", alternative = "two.sided")

bptest(m_i_add)
```

```
studentized Breusch-Pagan test

data: m_i_add
BP = 17.113, df = 4, p-value = 0.001838
```

Nezávislosť

Nezávislosť reziduí otestujeme Durbin-Watsonovým testom, pričom nezávislosť zamietame s $p\text{-hodnotou } 2.2 \cdot 10^{-16}$.

```
dwtest(m_i_add, alternative = "two.sided") # Two-sided
```

```
Durbin-Watson test

data: m_i_add
DW = 1.1017, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

• Otázka 10

Na základě vašeho modelu odpovězte, zdali si myslíte, že pokud bychom dokázali snížit/zvýšit podíl maloobchodu v dané lokalitě, vedlo by to ke zvýšení cen nemovitostí určených k bydlení v dané lokalitě?

Výsledný model má tvar:

$$\ln(Y) = 1.674 + 0.338 \cdot rm - 0.004 \cdot indus - 0.026 \cdot ptratio - 0.001 \cdot tax,$$

čo možno prepísať:

$$Y = e^{1.674} e^{0.338 \cdot rm} e^{-0.004 \cdot indus} e^{-0.026 \cdot ptratio} e^{-0.001 \cdot tax}$$

a Y predstavuje `medv`.

V našom modeli sa premenná `indus` vyskytuje so záporným znamienkom. A teda ak znížime podiel nemaloobchodných zón v danej lokalite, viedlo by to k zvýšeniu cien nehnuteľností. Poznamenajme, že koeficient je pomerne nízky (oproti hodnote interceptu či koeficientu premennej `rm`) a táto zmena v cene bude nízka pre malé zmeny - 0.39% pri jednotkovej zmene podielu neobchodných zón.

Medziročný nárast cien domov v ČR je 17,6% (k dňu 16.11.22 [vid'](#)). K takejto zmene by došlo pri poklese spomínaného podielu o 45 jednotiek (takmer ako dvojnásobok maximálneho non-retail business podielu v našej databáze).


```
summary(Boston_data$indus)
```

45.1282051282051

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.740	3.985	6.410	8.515	10.010	27.740