

Zkouška 15. 6. 2021

Popis situace: V rámci rozsáhlého psychologického výzkumu bylo vyšetřeno velké množství dětí školního věku. V datovém souboru `prospech.xls` jsou uloženy tyto informace o žácích druhého stupně základní školy:

`vzdel_m` ... nejvyšší dosažené vzdělání matky (varianty 1 ... ZŠ, 2 ... SŠ, 3 ... VŠ)

`IQ_celk` ... celkový inteligenční kvocient žáka dosažený ve Wechslerově testu inteligence pro děti WISC III

`prospech` ... průměrný školní prospěch žáka vypočítaný ze známek na posledním vysvědčení

`ln_prospech` ... přirozený logaritmus prospěchu (proměnná `prospech` byla transformována logaritmem kvůli přiblížení se normálnímu rozložení)

Cíl výzkumu:

1. Zjistit, zda se liší prospěch žáků ve třech skupinách definovaných nejvyšším dosaženým vzděláním matky.

2. Zjistit, zda je vzdělání matky významným faktorem působícím na prospěch, pokud eliminujeme vliv celkového IQ.

Upozornění: při analýzách používáme proměnnou `ln_prospech`.

(Při testování hypotéz vždy volíme hladinu významnosti 0,05.)

Výstupy ze statistického software k cíli 1

```
> vzdel_m<-factor(prospech$vzdel_m, labels=c('ZS','SS','VS'))
> summary(prospech)
  vzdel_m      IQ_celk      prospesch      ln_prospech
Min.   :1.000   Min.   : 63.0   Min.   :1.000   Min.   :0.0000
1st Qu.:1.000   1st Qu.: 91.0   1st Qu.:1.280   1st Qu.:0.2469
Median :2.000   Median :100.0   Median :1.600   Median :0.4700
Mean   :1.666   Mean   :100.1   Mean   :1.730   Mean   :0.4989
3rd Qu.:2.000   3rd Qu.:110.5   3rd Qu.:2.085   3rd Qu.:0.7347
Max.   :3.000   Max.   :131.0   Max.   :3.900   Max.   :1.3610

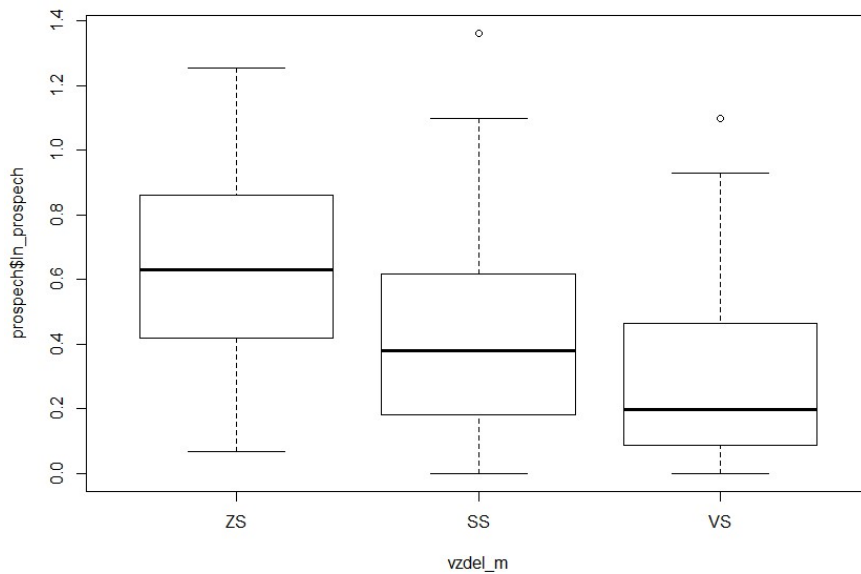
> table(vzdel_m)
vzdel_m
 ZS  SS  VS
136 159  28

> tapply(prospech$ln_prospech,vzdel_m,mean)
      ZS      SS      VS
0.6393329 0.4147648 0.2945521

> tapply(prospech$ln_prospech,vzdel_m,sd)
      ZS      SS      VS
0.2806782 0.2884584 0.2938934

>
tapply(prospech$ln_prospech,vzdel_m,sd)/tapply(prospech$ln_prospech,vzdel_m,mean)
      ZS      SS      VS
0.4390174 0.6954746 0.9977638
```

```
> boxplot(prospech$ln_prospech~vzdel_m)
```



```
> model.anova<-aov(prospech$ln_prospech~prospech$vzdel_m)
> summary(model.anova)
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
prospech$vzdel_m  2  4.977  2.4884   30.49 7.56e-13 ***
Residuals       320 26.114  0.0816
```

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

```
> shapiro.test(model.anova$residuals)
```

Shapiro-wilk normality test

```
data:  model.anova$residuals
W = 0.97598, p-value = 3.16e-05
```

```
> leveneTest(prospech$ln_prospech~vzdel_m, center=mean)
Levene's Test for Homogeneity of Variance (center = mean)
      Df F value Pr(>F)
group  2  0.0196 0.9806
      320
```

```
> durbinwatsonTest(model.anova)
lag Autocorrelation D-W Statistic p-value
  1      0.9270828      0.1085627      0
Alternative hypothesis: rho != 0
```

```
> PostHocTest(model.anova,method=c('hsd'))
```

Posthoc multiple comparisons of means : Tukey HSD
95% family-wise confidence level

```
$`prospech$vzdel_m`
      diff      lwr.ci      upr.ci    pval
2-1 -0.2245681 -0.3031346 -0.14600150 2.4e-10 ***
3-1 -0.3447808 -0.4843753 -0.20518637 4.4e-08 ***
3-2 -0.1202128 -0.2580727  0.01764713 0.1014
```

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

```
> model.lm<-lm(prospech$ln_prospech~prospech$vzdel_m)
> summary(model.lm)
```

```
Call:
lm(formula = prospech$ln_prospech ~ prospech$vzdel_m)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.57167 -0.23244 -0.03633  0.20654  0.94621
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.63933    0.02450  26.099 < 2e-16 ***
prospech$vzdel_m2 -0.22457    0.03337  -6.730 7.83e-11 ***
prospech$vzdel_m3 -0.34478    0.05928  -5.816 1.46e-08 ***
---
```

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

```
Residual standard error: 0.2857 on 320 degrees of freedom
Multiple R-squared:  0.1601, Adjusted R-squared:  0.1548
F-statistic: 30.49 on 2 and 320 DF, p-value: 7.562e-13
```

Výstupy ze statistického software k cíli 2

```
> model.interakce=lm(prospech$ln_prospech~vzdel_m*prospech$IQ_celk)
> summary(model.interakce)
```

```
Call:
lm(formula = prospech$ln_prospech ~ vzdel_m * prospech$IQ_celk)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.53882 -0.17953 -0.02112  0.18271  0.75561
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.839047    0.174042  10.567 < 2e-16 ***
vzdel_mSS    -0.389084    0.236270  -1.647  0.101
vzdel_mVS    -0.524134    0.471650  -1.111  0.267
prospech$IQ_celk -0.012597    0.001813  -6.947 2.13e-11 ***
vzdel_mSS:prospech$IQ_celk  0.002543    0.002379  1.069  0.286
vzdel_mVS:prospech$IQ_celk  0.003131    0.004431  0.707  0.480
---
```

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

```
Residual standard error: 0.2513 on 317 degrees of freedom
Multiple R-squared:  0.3559, Adjusted R-squared:  0.3457
F-statistic: 35.03 on 5 and 317 DF, p-value: < 2.2e-16
```

```
> model.bez.interakce=lm(prospech$ln_prospech~vzdel_m+prospech$IQ_celk)
> summary(model.bez.interakce)
```

```
Call:
lm(formula = prospech$ln_prospech ~ vzdel_m + prospech$IQ_celk)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.55376 -0.17821 -0.02116  0.18865  0.73786
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.686113    0.109375  15.416 < 2e-16 ***
vzdel_mSS    -0.139706    0.030586  -4.568 7.06e-06 ***
vzdel_mVS    -0.206833    0.053985  -3.831 0.000154 ***
prospech$IQ_celk -0.010992    0.001126  -9.761 < 2e-16 ***
---
```

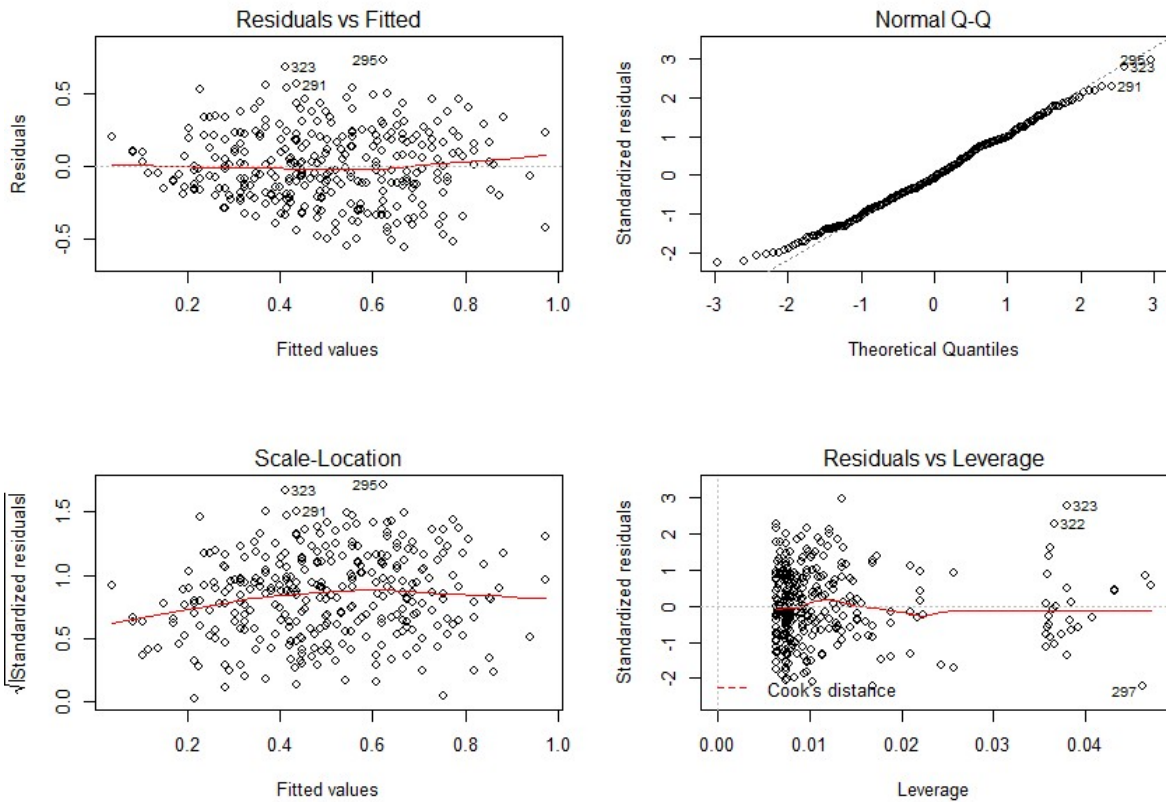
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.2511 on 319 degrees of freedom
 Multiple R-squared: 0.3533, Adjusted R-squared: 0.3472
 F-statistic: 58.08 on 3 and 319 DF, p-value: < 2.2e-16

```
> anova(model.bez.interakce,model.interakce)
Analysis of Variance Table
```

```
Model 1: prospech$ln_prospech ~ vzdel_m + prospech$IQ_celk
Model 2: prospech$ln_prospech ~ vzdel_m * prospech$IQ_celk
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     319 20.108
2     317 20.026  2  0.081904 0.6482 0.5237
```

```
> par(mfrow=c(2,2))
> plot(model.bez.interakce)
```



```
> shapiro.test(model.bez.interakce$residuals)
```

Shapiro-wilk normality test

```
data: model.bez.interakce$residuals
W = 0.99284, p-value = 0.125
```

```
> durbinwatsonTest(lm(prospech$ln_prospech~vzdel_m+prospech$IQ_celk))
```

```
lag Autocorrelation D-W Statistic p-value
1 0.7105037 0.5433228 0
Alternative hypothesis: rho != 0
```

```
> leveneTest(prospech$ln_prospech~vzdel_m, center=mean)
```

```
Levene's Test for Homogeneity of Variance (center = mean)
  group Df F value Pr(>F)
1     2  0.0196 0.9806
320
```

```
> anova(model.bez.interakce)
```

```
Analysis of Variance Table
```

```
Response: prospech$ln_prospech
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
vzdel_m	2	4.9769	2.4884	39.477	4.799e-16	***
prospech\$IQ_celk	1	6.0063	6.0063	95.287	< 2.2e-16	***
Residuals	319	20.1080	0.0630			

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

```
> summary(model.bez.interakce)
```

```
Call:
```

```
lm(formula = prospech$ln_prospech ~ vzdel_m + prospech$IQ_celk)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.55376	-0.17821	-0.02116	0.18865	0.73786

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.686113	0.109375	15.416	< 2e-16	***
vzdel_mSS	-0.139706	0.030586	-4.568	7.06e-06	***
vzdel_mVS	-0.206833	0.053985	-3.831	0.000154	***
prospech\$IQ_celk	-0.010992	0.001126	-9.761	< 2e-16	***

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

```
Residual standard error: 0.2511 on 319 degrees of freedom
```

```
Multiple R-squared:  0.3533, Adjusted R-squared:  0.3472
```

```
F-statistic: 58.08 on 3 and 319 DF, p-value: < 2.2e-16
```