

# NEX1

*Kajetán Poliak, Adam Novotný*

*18. října 2019*

## NEX - první domácí úkol

### 1. část

- Namerte data: experiment nastavte tak, aby byl znahodnenny !!!!! Poradi mereni si poznamenejte pro dalsi kontrolu pripadne zavislosti na poradi mereni. Jednotlive lidi ve skupine berte jako ruzne operatory experimentu (blokuje). Pocet replikaci u jednoho cloveka a jednoho casu vemte 1. V protokolu krom popisu experimentu diskutujte i jen promenne, ktere mohli mit na mereni vliv.

Experimentu se účastnili dva operátoři. Každý z operátorů se podrobil devíti měření (tečkování do 3 párů kruhů pomocí dominantní, nedominantní a obou rukou), kde každé měření trvalo 10s. Pořadí měření bylo znáhodněno pomocí generátoru náhodné posloupnosti z [www.random.org](http://www.random.org). Jako první se vygenerovala náhodné posloupnost celých čísel od 1 do 9, která určovala pořadí měření. Následně se pomocí náhodné posloupnosti od 1 do 3 přiřadila jedna z úrovní proměnné ruka ke každému ze tří papírů obsahujícím tři různé velikosti kruhu. Na měření mohlo mít vliv jak rozpoložení operátora (různé polohy sedu při experimentu, psychický nátlak okolí,..), tak nepřesnosti měření času nebo různé psací potřeby operátorů.

### 2. část

- Spoctete základní statistiky (mean, median a sd pro jednotlivé faktory - velikost, ruka, operator) Zobrazte namerena data (box plot, interaction plot, effects plot, ...) a okomentujte je co z danych obrazku muzeme pred samotnou analyzou rici o vysledku?

#### # ZÁKLADÍ STATISTIKY

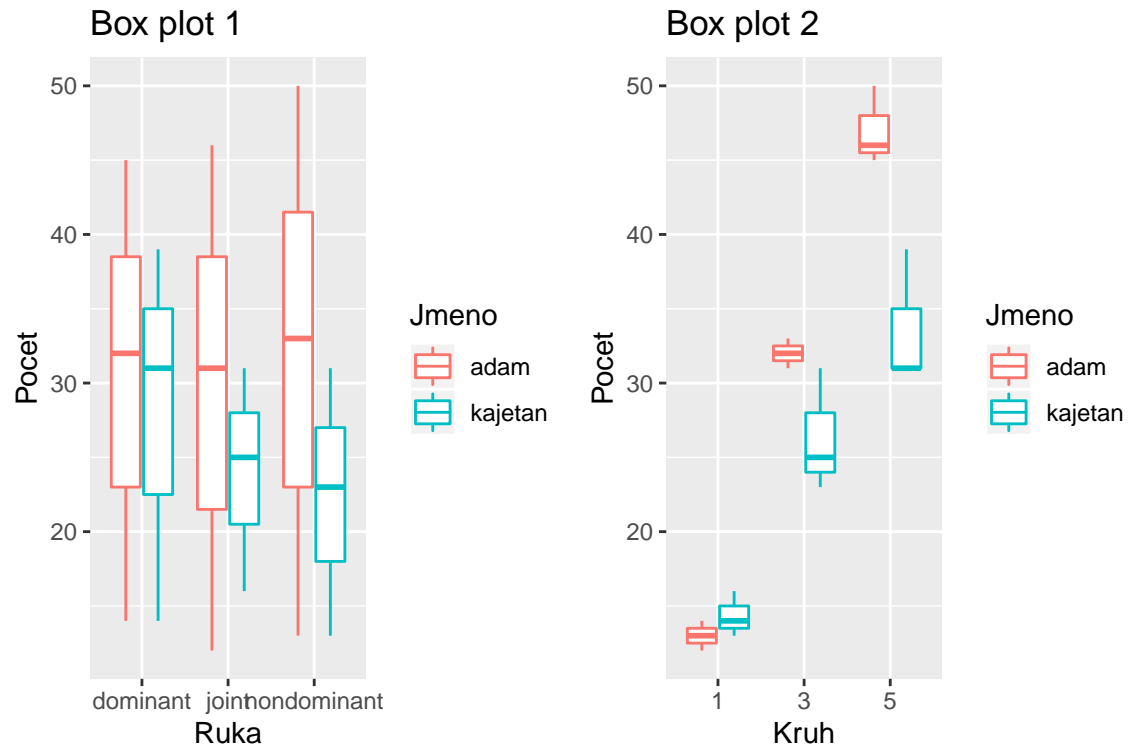
##### characteristics

##	adam	kajetan	dominant	nondominant	joint	1 cm
## Min.	12.00000	13.000000	14.00000	13.00000	12.00000	12.00000
## 1st Qu.	14.00000	16.000000	18.25000	15.50000	18.25000	13.00000
## Median	32.00000	25.000000	31.50000	27.00000	28.00000	13.50000
## Mean	30.66667	24.777778	29.16667	27.16667	26.83333	13.66667
## 3rd Qu.	45.00000	31.000000	37.25000	32.50000	31.00000	14.00000
## Max.	50.00000	39.000000	45.00000	50.00000	46.00000	16.00000
## Sd.	14.83240	9.038498	12.79714	14.06295	12.18879	1.36626
##	3 cm	5 cm				
## Min.	23.000000	31.000000				
## 1st Qu.	26.500000	33.000000				
## Median	31.000000	42.000000				
## Mean	29.166667	40.333333				
## 3rd Qu.	31.750000	45.750000				
## Max.	33.000000	50.000000				
## Sd.	4.119061	8.041559				

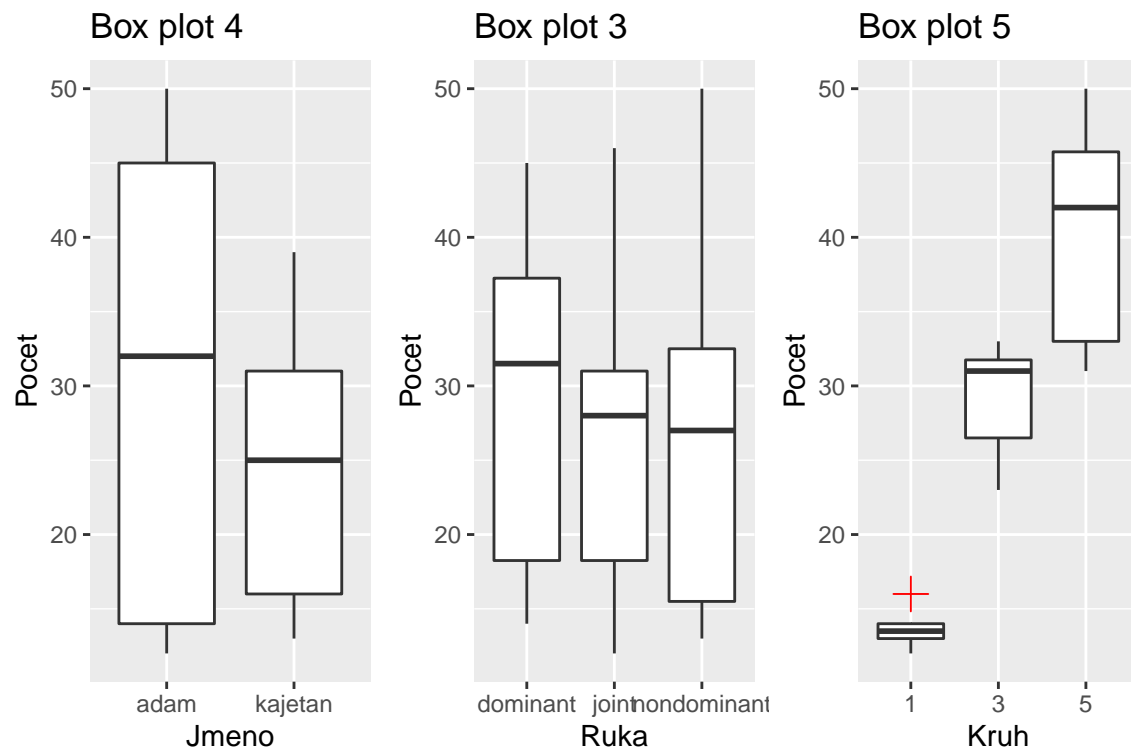
Tabulka zobrazuje základní statistiky experimentu. Vidíme, že operátoři mají podobná minima ale maximum Adama je o celých 11 vyšší. Stejně tak je vyšší střední hodnota počtu teček Adama. Maxima u proměnných dominant, nondominant a joint tedy souvisí s operátorem Adam. Podle očekávání vidíme, že hodnoty maxim i minim jsou vyšší se zvětšujícím se poloměrem kruhu. Směrodatná odchylka u proměnných adam, kajetan, dominant, nondominant a joint dosahují poměrně vysokých hodnot, to je vysvětlitelné tím, že počet teček nabýval pro každý kruh se radikálně lišil. Nejnižší střední hodnotu z různých úrovní proměnné ruka má joint.

#BOX-PLOTS

grid.arrange(p1,p2,ncol=2)



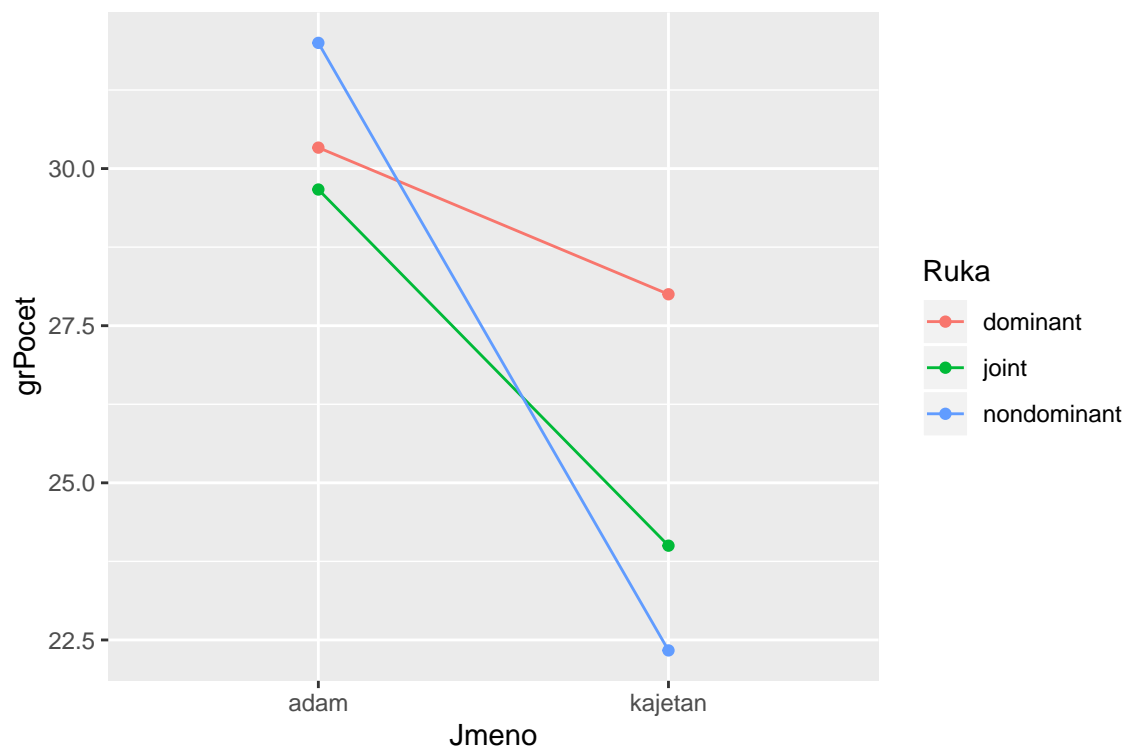
grid.arrange(p4,p3,p5,ncol=3)



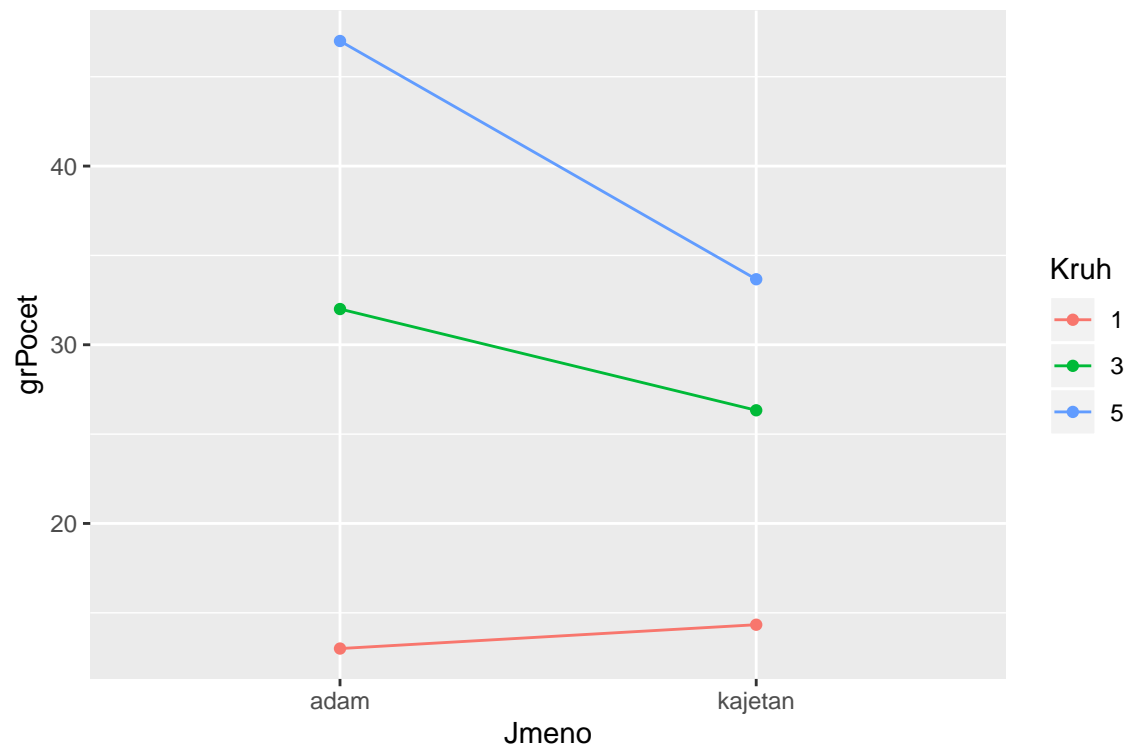
Z boxplotu 1 vidíme, že střední hodnoty u dominant mají operátoři podobné. U nondominant a joint má Adam vyšší střední hodnoty i znatelně větší variance. Boxplot 2 ukazuje, že Adam měl vyšší počet teček všude, kromě nejmenšího kruhu.

#### #INTERACTION PLOTS

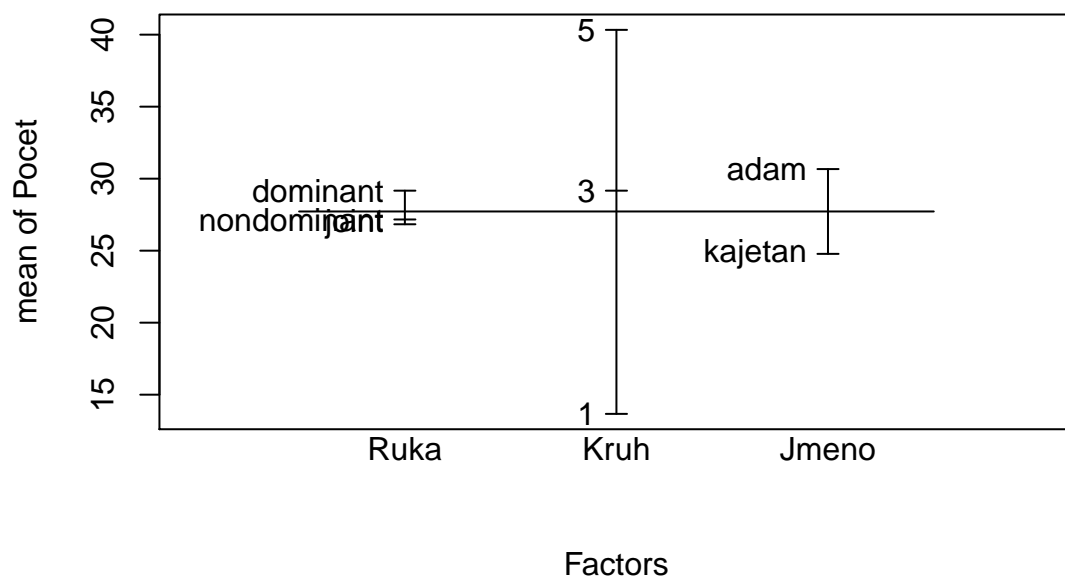
```
dat2 %>%
  ggplot() +
  aes(x = Jmeno, y = grPocet, color = Ruka) +
  geom_line(aes(group = Ruka)) +
  geom_point()
```



```
dat3 %>%
  ggplot() +
  aes(x = Jmeno, y = grPocet, color = Kruh) +
  geom_line(aes(group = Kruh)) +
  geom_point()
```



```
#EFFECT PLOTS
plot.design(Pocet~Ruka+Kruh+Jmeno, data = dat)
```



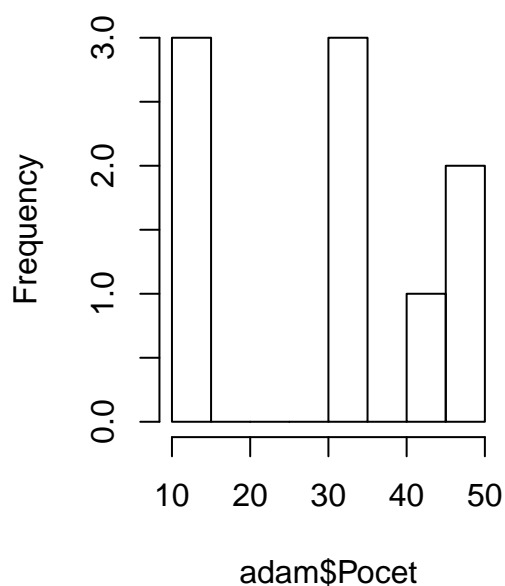
### 3. část

- Zamerte se zvláště na faktory ruční a velikost kola. Otestujte hypotézu o shodnosti rozptylu pro jednotlivé úrovně a vhodným testem ověřte stejnost středních hodnot. Proveďte Tukey HSD a Fisher LSD test pro párové porovnání středních hodnot jednotlivých skupin s vybranou korekcí p-hodnoty.

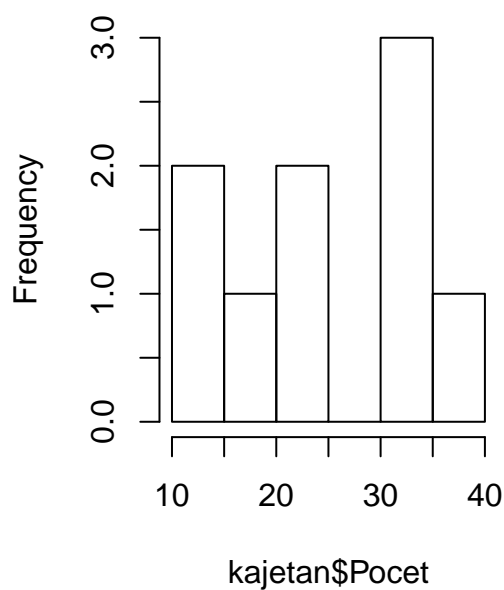
```
#hypoteza shodnosti rozptylu pro jednotlivé úrovně  
alpha = 0.95
```

```
# podmínky f-testu: normalita, tj. histogram a shapiro-wilk test  
par(mfrow = c(1,2))  
hist(adam$Pocet, breaks=6)  
hist(kajetan$Pocet, breaks=6)
```

**Histogram of adam\$Pocet**



**Histogram of kajetan\$Pocet**



```
shapiro.test(adam$Pocet)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: adam$Pocet  
## W = 0.88219, p-value = 0.1656
```

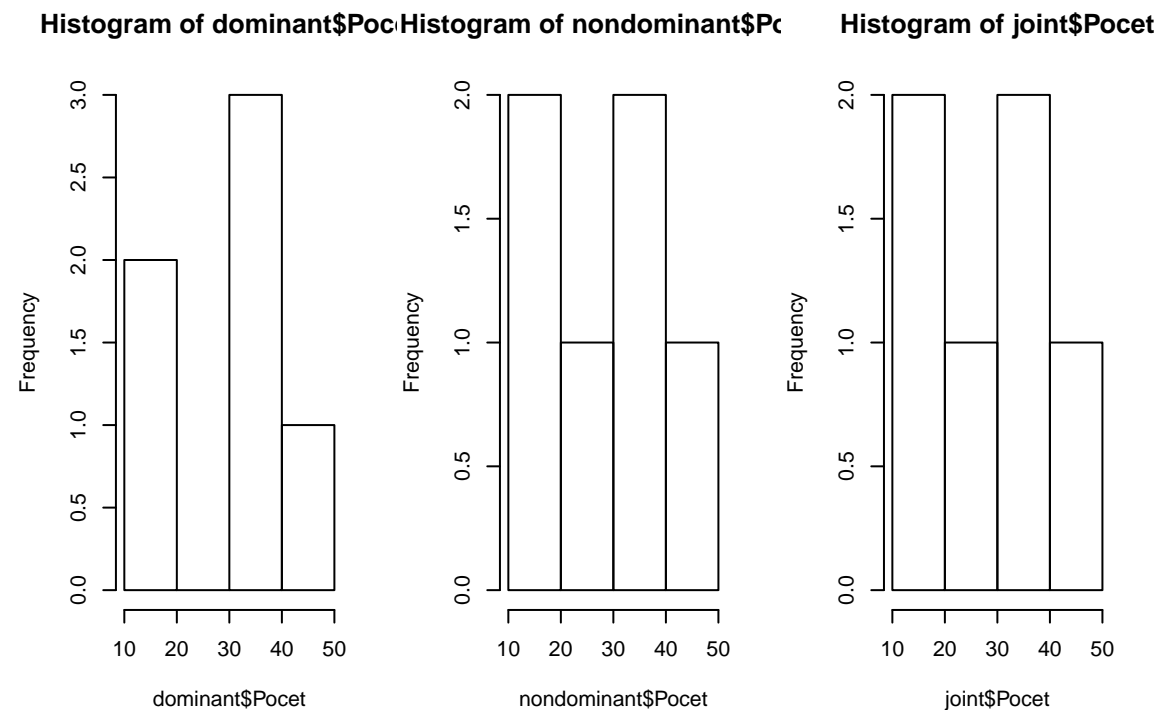
```
shapiro.test(kajetan$Pocet) # p-hodnoty > 0.05, tj. nezamítáme normalitu
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: kajetan$Pocet  
## W = 0.92229, p-value = 0.4116
```

```
var.test(adam$Pocet, kajetan$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha) # p-hodnota
```

```
##
## F test to compare two variances
##
## data: adam$Pocet and kajetan$Pocet
## F = 2.693, num df = 8, denom df = 8, p-value = 0.1827
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.607445 11.938599
## sample estimates:
## ratio of variances
##      2.692962
```

```
## RUKA
par(mfrow = c(1,3))
hist(dominant$Pocet)
hist(nondominant$Pocet)
hist(joint$Pocet)
```



```
shapiro.test(dominant$Pocet)
```

```
##
## Shapiro-Wilk normality test
##
## data: dominant$Pocet
## W = 0.89281, p-value = 0.3332
```

```
shapiro.test(nondominant$Pocet)
```

```
##
## Shapiro-Wilk normality test
##
```

```

## data: nondominant$Pocet
## W = 0.91686, p-value = 0.4831
shapiro.test(joint$Pocet) # p-hodnoty > 0.05, tj. nezamítame normalitu

##
## Shapiro-Wilk normality test
##
## data: joint$Pocet
## W = 0.9526, p-value = 0.7613
var.test(dominant$Pocet, nondominant$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha) #

##
## F test to compare two variances
##
## data: dominant$Pocet and nondominant$Pocet
## F = 0.82808, num df = 5, denom df = 5, p-value = 0.8411
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.1158741 5.9177775
## sample estimates:
## ratio of variances
## 0.8280802
var.test(dominant$Pocet, joint$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha) # p-hodnoty > 0.05, tj. nezamítame normalitu

##
## F test to compare two variances
##
## data: dominant$Pocet and joint$Pocet
## F = 1.1023, num df = 5, denom df = 5, p-value = 0.9175
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.1542474 7.8775351
## sample estimates:
## ratio of variances
## 1.102311
var.test(nondominant$Pocet, joint$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha) # p-hodnoty > 0.05, tj. nezamítame normalitu

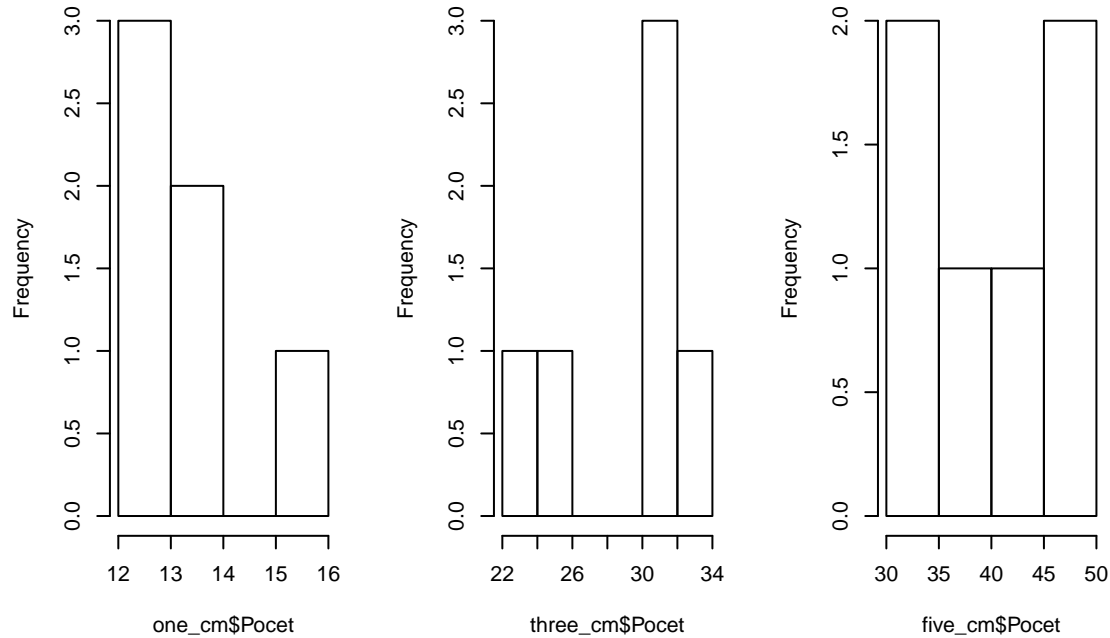
##
## F test to compare two variances
##
## data: nondominant$Pocet and joint$Pocet
## F = 1.3312, num df = 5, denom df = 5, p-value = 0.7613
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.1862711 9.5130095
## sample estimates:
## ratio of variances
## 1.331164

## KRUH
par(mfrow = c(1,3))
hist(one_cm$Pocet)
hist(three_cm$Pocet)

```

```
hist(five_cm$Pocet)
```

Histogram of one\_cm\$Poce   Histogram of three\_cm\$Pocet   Histogram of five\_cm\$Poce



```
shapiro.test(one_cm$Pocet)
```

```
##
## Shapiro-Wilk normality test
##
## data:  one_cm$Pocet
## W = 0.92664, p-value = 0.5544
```

```
shapiro.test(three_cm$Pocet)
```

```
##
## Shapiro-Wilk normality test
##
## data:  three_cm$Pocet
## W = 0.8307, p-value = 0.109
```

```
shapiro.test(five_cm$Pocet) # p-hodnoty > 0.05, tj. nezamitame normalitu
```

```
##
## Shapiro-Wilk normality test
##
## data:  five_cm$Pocet
## W = 0.88926, p-value = 0.3143
```

```
var.test(one_cm$Pocet, three_cm$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha) # ZAMI
```

```
##
## F test to compare two variances
##
```



```
## data: one_cm$Pocet and three_cm$Pocet
## F = 0.11002, num df = 5, denom df = 5, p-value = 0.0301
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.01539515 0.78624240
## sample estimates:
## ratio of variances
## 0.1100196

var.test(one_cm$Pocet, five_cm$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha) # ZAMIT.

##
## F test to compare two variances
##
## data: one_cm$Pocet and five_cm$Pocet
## F = 0.028866, num df = 5, denom df = 5, p-value = 0.00139
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.004039244 0.206287311
## sample estimates:
## ratio of variances
## 0.02886598

var.test(three_cm$Pocet, five_cm$Pocet, ratio = 1, alternative = "two.sided", conf.level = alpha)

##
## F test to compare two variances
##
## data: three_cm$Pocet and five_cm$Pocet
## F = 0.26237, num df = 5, denom df = 5, p-value = 0.1683
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.03671384 1.87500430
## sample estimates:
## ratio of variances
## 0.2623711
```

Na začátku otestujeme normalitu jakožto podmínku F-testu. Pracujeme na 5% hladině významnosti. Normalitu testujeme pomocí Shapiro-Wilkova testu. Pro operátory nezamítáme nulovou hypotézu o normálním rozdělení. F-test pro operátory vrací p-hodnotu 0.1827, nezamítáme tedy nulovou hypotézu o rovnosti rozptylů. Přecházíme k faktoru Ruka. Pro všechny úrovně nezamítáme, dle Shapiro-wilkova testu, normalitu. F- test vrací pro páry dominant-nondominant, dominant-joint a nondominant-joint respektive p-hodnoty 0.84, 0.92 a 0.76 nezamítáme tedy rovnosti rozptylů. Jako poslední testujeme faktor Kruh. P-hodnoty S-W testu pro úrovně 1cm, 3cm a 5cm jsou postupně 0.55, 0.11 a 0.31, ani pro jednu z úrovní nezamítáme normalitu. F-test vrací pro dvojice 1cm-3cm, 1cm-5cm a 3cm-5cm p-hodnoty 0.030, 0.001 a 0.168 respektive. Pro dvojice 1cm-3cm a 1cm-5cm zamítáme nulovou hypotézu o rovnosti rozptylů.

```
# hypoteza shodnosti strednich hodnot
```

```
pairwise.t.test(dat$Pocet, dat$Ruka, p.adjust.method="bonferroni", var.equal = T)
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: dat$Pocet and dat$Ruka
##
## dominant joint
```

```
## joint      1      -
## nondominant 1      1
##
## P value adjustment method: bonferroni
pairwise.t.test(dat$Pocet,dat$Ruka,p.adjust.method="hochberg", var.equal = T)

##
## Pairwise comparisons using t tests with pooled SD
##
## data:  dat$Pocet and dat$Ruka
##
##          dominant joint
## joint      0.97      -
## nondominant 0.97      0.97
##
## P value adjustment method: hochberg
pairwise.t.test(dat$Pocet,dat$Kruh,p.adjust.method="bonferroni", var.equal = F)

##
## Pairwise comparisons using t tests with pooled SD
##
## data:  dat$Pocet and dat$Kruh
##
##      1      3
## 3 0.0004  -
## 5 8.4e-07 0.0069
##
## P value adjustment method: bonferroni
pairwise.t.test(dat$Pocet,dat$Kruh,p.adjust.method="hochberg", var.equal = F)

##
## Pairwise comparisons using t tests with pooled SD
##
## data:  dat$Pocet and dat$Kruh
##
##      1      3
## 3 0.00027  -
## 5 8.4e-07 0.00229
##
## P value adjustment method: hochberg
```

Dalším bodem je testování hypotézy shodnosti středních hodnot. Jedinou podmínkou použití t-testu je normalita, která je již otestována. Pro faktor Ruka zamítáme nulovou hypotézu shodnosti středních hodnot mezi všemi úrovněmi ( $p\text{-val} = 0.97$ ). U faktoru Kruh nulovou hypotézu nezamítáme.

```
#Tukey HSD + Fisher LSD
#Celkový aov
#Kruh bychom ani testovat, protože jsou různé variance ve skupinách 1cm, 3cm, 5cm
aov_celk = aov(Pocet ~ Jmeno + Ruka + Kruh, data = dat) #ruka není stat. významná
summary(aov_celk)
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Jmeno      1  156.1    156.1    7.728 0.0167 *
## Ruka       2   19.1     9.6    0.473 0.6342
```

```
## Kruh          2 2152.1 1076.1 53.285 1.07e-06 ***
## Residuals    12 242.3   20.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

LSD1 <- LSD.test(aov_celk, "Jmeno"); LSD1

## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
## 20.19444 12 27.72222 16.21018 2.178813 4.615615
##
## $parameters
##      test p.adjusted name.t ntr alpha
## Fisher-LSD      none Jmeno  2 0.05
##
## $means
##      Pocet      std r      LCL      UCL Min Max Q25 Q50 Q75
## adam    30.66667 14.832397 9 27.40293 33.93040 12 50 14 32 45
## kajetan 24.77778  9.038498 9 21.51404 28.04151 13 39 16 25 31
##
## $comparison
## NULL
##
## $groups
##      Pocet groups
## adam    30.66667      a
## kajetan 24.77778      b
##
## attr(,"class")
## [1] "group"

LSD2 <- LSD.test(aov_celk, "Ruka"); LSD2

## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
## 20.19444 12 27.72222 16.21018 2.178813 5.652951
##
## $parameters
##      test p.adjusted name.t ntr alpha
## Fisher-LSD      none  Ruka  3 0.05
##
## $means
##      Pocet      std r      LCL      UCL Min Max Q25 Q50 Q75
## dominant 29.16667 12.79714 6 25.16943 33.16391 14 45 18.25 31.5 37.25
## joint    26.83333 12.18879 6 22.83609 30.83057 12 46 18.25 28.0 31.00
## nondominant 27.16667 14.06295 6 23.16943 31.16391 13 50 15.50 27.0 32.50
##
## $comparison
## NULL
##
## $groups
##      Pocet groups
## dominant 29.16667      a
## nondominant 27.16667      a
## joint    26.83333      a
```

```
##
## attr("class")
## [1] "group"
LSD3 <- LSD.test(aov_celk, "Kruh"); LSD3

## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
##      20.19444 12 27.72222 16.21018 2.178813 5.652951
##
## $parameters
##      test p.adjusted name.t ntr alpha
## Fisher-LSD      none   Kruh   3 0.05
##
## $means
##      Pocet      std r      LCL      UCL Min Max  Q25  Q50  Q75
## 1 13.66667 1.366260 6 9.669426 17.66391 12 16 13.0 13.5 14.00
## 3 29.16667 4.119061 6 25.169426 33.16391 23 33 26.5 31.0 31.75
## 5 40.33333 8.041559 6 36.336093 44.33057 31 50 33.0 42.0 45.75
##
## $comparison
## NULL
##
## $groups
##      Pocet groups
## 5 40.33333      a
## 3 29.16667      b
## 1 13.66667      c
##
## attr("class")
## [1] "group"
TukeyHSD(aov_celk, "Jmeno", ordered = FALSE, conf.level = alpha)

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = Pocet ~ Jmeno + Ruka + Kruh, data = dat)
##
## $Jmeno
##              diff      lwr      upr      p adj
## kajetan-adam -5.888889 -10.5045 -1.273273 0.0166546
TukeyHSD(aov_celk, "Ruka", ordered = FALSE, conf.level = alpha)

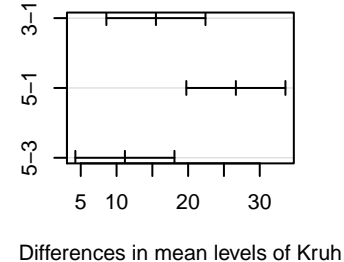
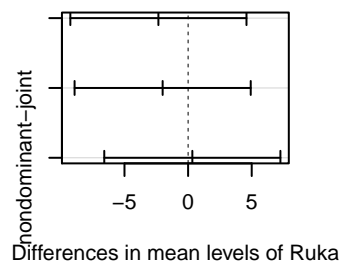
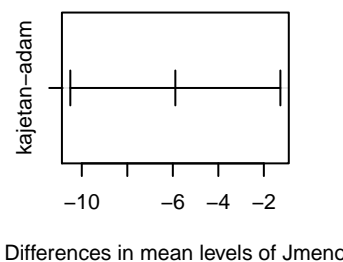
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = Pocet ~ Jmeno + Ruka + Kruh, data = dat)
##
## $Ruka
##              diff      lwr      upr      p adj
## joint-dominant -2.333333 -9.255132 4.588465 0.6509234
## nondominant-dominant -2.000000 -8.921799 4.921799 0.7272003
## nondominant-joint 0.333333 -6.588465 7.255132 0.9909481
```

```
TukeyHSD(aov_celk, "Kruh", ordered = FALSE, conf.level = alpha)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Pocet ~ Jmeno + Ruka + Kruh, data = dat)
##
## $Kruh
##      diff      lwr      upr    p adj
## 3-1 15.50000  8.578201 22.42180 0.0001769
## 5-1 26.66667 19.744868 33.58847 0.0000007
## 5-3 11.16667  4.244868 18.08847 0.0027206
```

```
par(mfrow = c(2,3))
#plot(LSD1)
#plot(LSD2)
#plot(LSD3)
plot(TukeyHSD(aov_celk, "Jmeno", ordered = FALSE, las=1))
plot(TukeyHSD(aov_celk, "Ruka", ordered = FALSE, las=1))
plot(TukeyHSD(aov_celk, "Kruh", ordered = FALSE, las=1))
```

**95% family-wise confidence interval**



Fish-erův LSD test vrací pro proměnnou jména least significant difference rovný 4.6 úrovně jsou tedy dvě rozdílné skupiny (jejich rozdíl je 5.889). LSD pro Ruka je roven 5.65, všechny úrovně jsou ve stejné skupině. Tukeyho honest significant difference test vrací pro proměnnou Jména p-hodnotu 0.017, nezamítáme tedy hypotézu, že úrovně mají různou střední hodnotu. Tento test na proměnné Ruka vrací p-hodnoty indikující, že všechny úrovně patří do stejné skupiny. Výsledek je tedy shodný s předchozími testy.

## 4.část

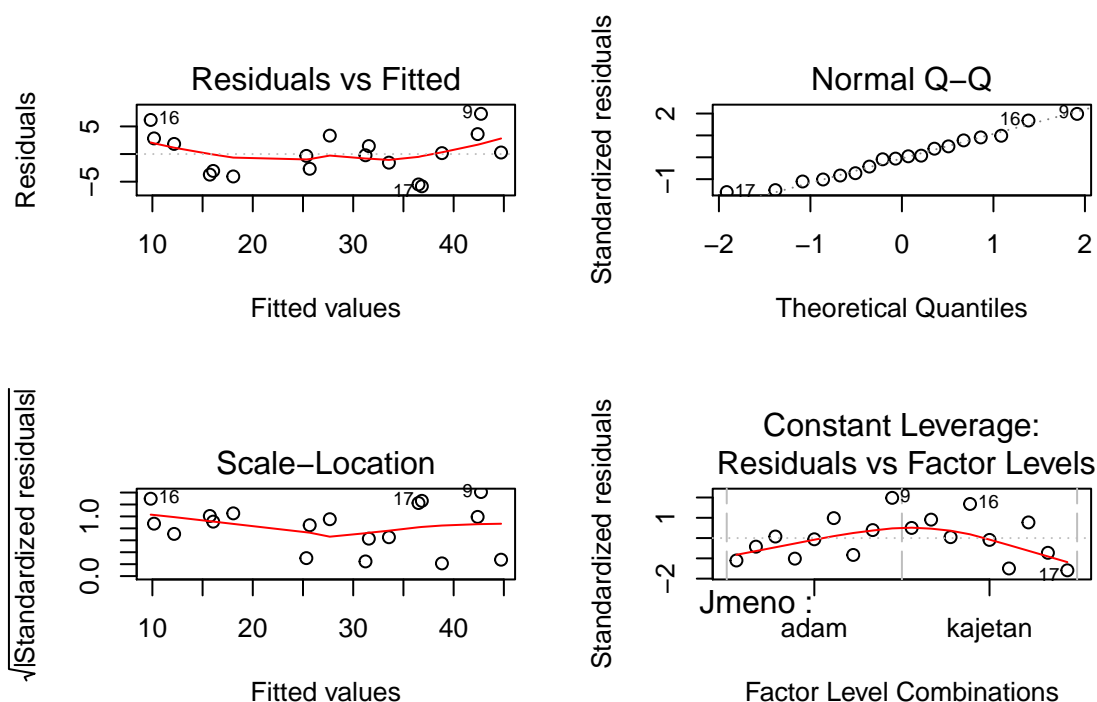
- Analyzujte data pomocí ANOVA (všechny proměnné berete jako faktor) - s předpokladem, že každý z účastníků experimentu tvoří jeden blok. - diskutujte výsledky, overte předpoklady, vykreslete QQ-plot, residua x fitted values, residua x čas, ... - diskutujte vliv znahodnění experimentu a vývoj reziduí v case (číslo měření)

```
summary(aov_celk)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Jmeno         1  156.1   156.1    7.728  0.0167 *
## Ruka          2   19.1    9.6     0.473  0.6342
## Kruh          2 2152.1 1076.1   53.285 1.07e-06 ***
## Residuals    12  242.3    20.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
plot(aov_celk)
```



ANOVA vrací pro faktory Jmeno a Kruh p-hodnoty  $< 0.05$ , tzn, mají vliv na proměnnou Poče, faktor Ruka signifikantní není, proto jej vyhodíme z modelu. Residuals vs. Fitted plot ukazuje několik možných outlierů, které ovlivňují lineární trend u obou konců křivky. Bod č.17 je značně vychýlený i u Normálního Q-Q plotu, kde máme zcela zřetelný lineární trend. Scale-Location plot má body víceméně rovnoměrně rozdělené a křivka zaujímá horizontální polohu, vypadá to tedy na rovnoměrné rozdělení reziduí přes prediktory a homoskedasticitu.

## 5.část

- Porovnejte a diskutujte výsledky z bodu 3 a 4.

Ve třetím bodu jsme několika testy (Fisher LSD, Tukey HSD, pairwise t-test) určili, že faktory Ruka a Jména

mají úrovně s různými středními hodnotami (pocházejí z různých skupin). Jinak to bylo s faktorem Kruh, kde všechny úrovně patří do stejné. Model ANOVA má u faktorů Jmeno a Kruh p-hodnoty  $< 0.05$ , je zde signifikantní rozdíl mezi úrovněmi. Faktor Kruh, má velmi vysokou p = hodnotu, není signifikantní. Závěry se tedy shodují nezávisle na volbě nástroje.

## 6. část

- Pokud data nesplňují předpoklady pro použití ANOVA, diskutujte možné transformace (logaritmická, Box-Cox, ...) a důvod proč data předpoklady nesplňují? Vyskyt outlieru, způsob měření, divný operator, ...

```
#log trafo
aov1 <- aov(Pocet ~ Jmeno + Ruka + Kruh, data = dat); summary(aov1)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## Jmeno         1  156.1   156.1    7.728  0.0167 *
## Ruka          2   19.1    9.6     0.473  0.6342
## Kruh          2 2152.1  1076.1   53.285 1.07e-06 ***
## Residuals    12   242.3    20.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

aov2 <- aov(Pocet ~ Jmeno + Kruh + Jmeno*Kruh, data = dat)
aov_trafo <- aov(log(Pocet) ~ Jmeno + Kruh + Jmeno*Kruh, data = dat) #ruka není stat. významna
summary(aov_trafo)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## Jmeno         1  0.099   0.0992    9.539  0.00939 **
## Kruh          2   3.619   1.8095  174.068 1.37e-09 ***
## Jmeno:Kruh     2   0.148   0.0741    7.128  0.00911 **
## Residuals    12   0.125   0.0104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

shapiro.test(aov_trafo$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  aov_trafo$residuals
## W = 0.94021, p-value = 0.2922

bptest(aov_trafo)

##
##  studentized Breusch-Pagan test
##
## data:  aov_trafo
## BP = 7.7308, df = 5, p-value = 0.1717

#B-C trafo
modelBC <- boxcox(aov_celk, objective.name = "PPCC", lambda = seq(0,5, by=0.5)); modelBC

##
## Results of Box-Cox Transformation
## -----
##
## Objective Name:                PPCC
```

```
##
## Linear Model:                                aov_celk
##
## Sample Size:                                18
##
## lambda      PPCC
##    0.0 0.9851693
##    0.5 0.9813793
##    1.0 0.9909187
##    1.5 0.9942694
##    2.0 0.9945821
##    2.5 0.9916795
##    3.0 0.9877888
##    3.5 0.9829078
##    4.0 0.9779361
##    4.5 0.9722093
##    5.0 0.9674904

lambda <- cbind(modelBC$lambda,modelBC$objective)
lam <- lambda[which.max(lambda[,2]),1];lam

## [1] 2

modelBC_final <- aov((Pocet)^lam ~ Jmeno + Kruh + Jmeno*Kruh , data = dat);
summary(modelBC_final)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Jmeno	1	909001	909001	24.10	0.00036 ***
Kruh	2	6700590	3350295	88.83	6.42e-08 ***
Jmeno:Kruh	2	950903	475452	12.61	0.00112 **
Residuals	12	452611	37718		

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

## 7. část

- Vyberte nejvhodnější model a i kdyby nesplňoval předpoklady pro použití ANOVA reste následující:
  - Spočítejte sílu testu v ANOVA (pro max. dvoufaktorialní analýzu - jeden z faktorů zanedbejte, nebo vezmete měření pro jednu jeho konkrétní úroveň)
  - Předpokládejme, že standardní odchylka disturbance bude pro prováděný experiment 4 a maximální (pro nás signifikantní) rozdíl, který chceme detekovat je 5 bodů v kruhu. Spočítejte počet potřebných replikací, aby síla výsledného testu byla větší než 0.9

```
# vyberame asi Pocet ~ Jmeno*Kruh
anova_fin <- aov(Pocet ~ Jmeno*Kruh, data = dat)
summary(anova_fin)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Jmeno	1	156.1	156.1	18.727	0.000983 ***
Kruh	2	2152.1	1076.1	129.127	7.66e-09 ***
Jmeno:Kruh	2	161.4	80.7	9.687	0.003131 **
Residuals	12	100.0	8.3		

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

shapiro.test(anova_fin$residuals) # OK... zamítame
```



```
##
## Shapiro-Wilk normality test
##
## data: anova_fin$residuals
## W = 0.92219, p-value = 0.1413

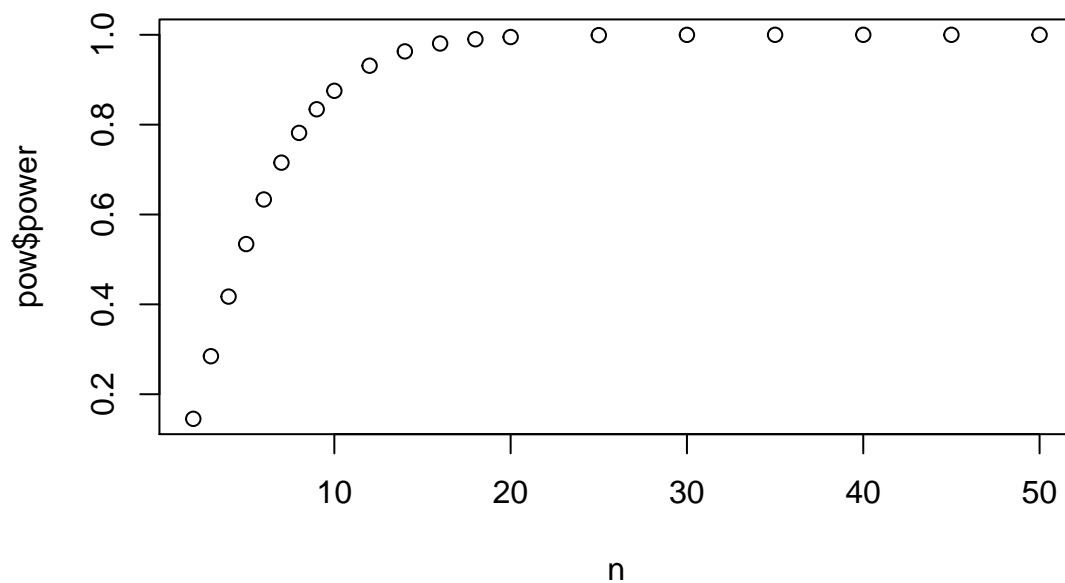
#leveneTest(anova_fin) #OK...zazitame
var1 <- var(dat$Pocet[which(dat$Jmeno=="kajetan")])
var2 <- var(dat$Pocet[which(dat$Jmeno=="adam")])
mean1 <- mean(dat$Pocet[which(dat$Jmeno=="kajetan")])
mean2 <- mean(dat$Pocet[which(dat$Jmeno=="adam")])

variances <- c(var1,var2)
means <- c(mean1, mean2)

power.anova.test(groups = 2, between.var = var(means), n = length(dat$Pocet[which(dat$Jmeno=="kajetan")]))

##
##      Balanced one-way analysis of variance power calculation
##
##      groups = 2
##      n = 9
##      between.var = 17.33951
##      within.var = 16
##      sig.level = 0.05
##      power = 0.8342625
##
## NOTE: n is number in each group
# kolik potrebných replikací pro power > 0,9
power.anova.test(groups = 2, between.var = var(means), n = NULL , sig.level = 0.05, within.var = 16, power = 0.9)

##
##      Balanced one-way analysis of variance power calculation
##
##      groups = 2
##      n = 10.75368
##      between.var = 17.33951
##      within.var = 16
##      sig.level = 0.05
##      power = 0.9
##
## NOTE: n is number in each group
n <- c(seq(2,10,by=1),seq(12,20,by=2),seq(25,50,by=5))
pow <- power.anova.test(groups = 2, between.var = var(means), n = n , sig.level = 0.05, within.var = 16)
plot(n, pow$power)
```



## 8. část

- Vytvorte regresni model, kde nebudete uvazovat bloky a velikost kruhu bude kvantitativni promenna. - zkuste pridat do modelu i druhou mocninu a porovnejte dva regresni modely mezi sebou a vyberte vhodnejsi - overte predpoklady pro pouziti vybraného modelu a vykreslete QQ-plot, residua x fitted values, residua x prumer , ...

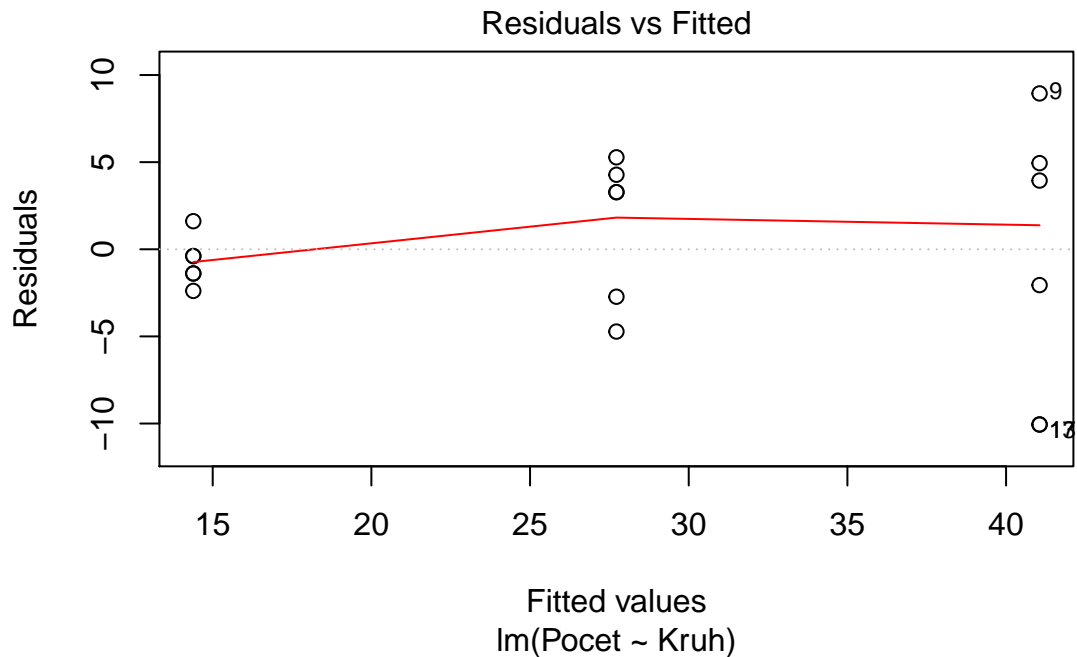
```
dat$Kruh <- as.numeric(dat$Kruh)
```

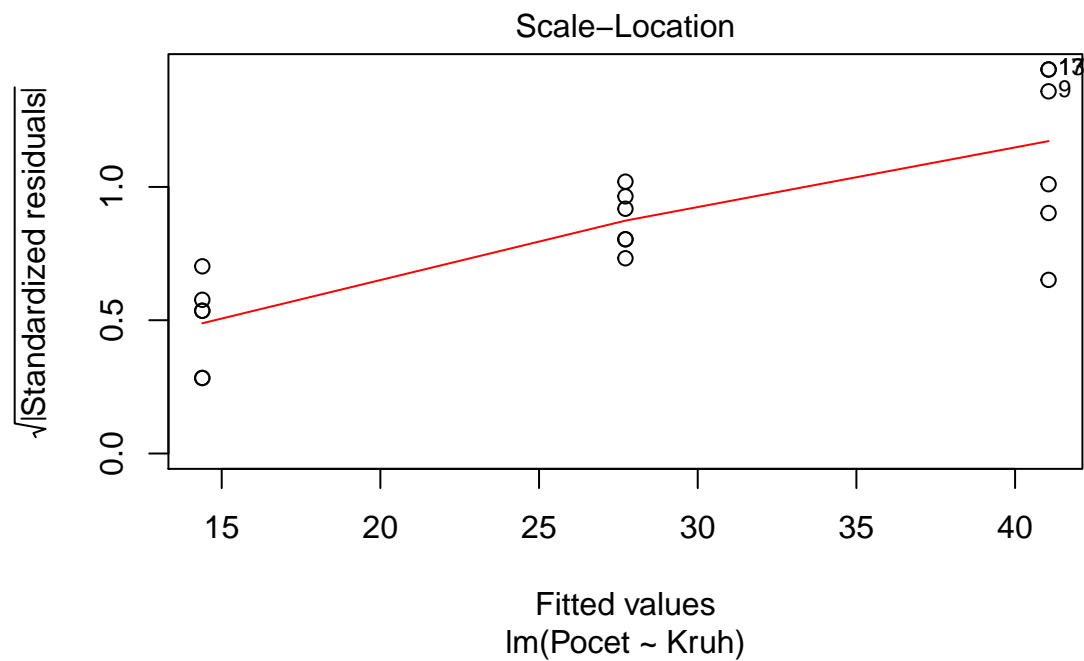
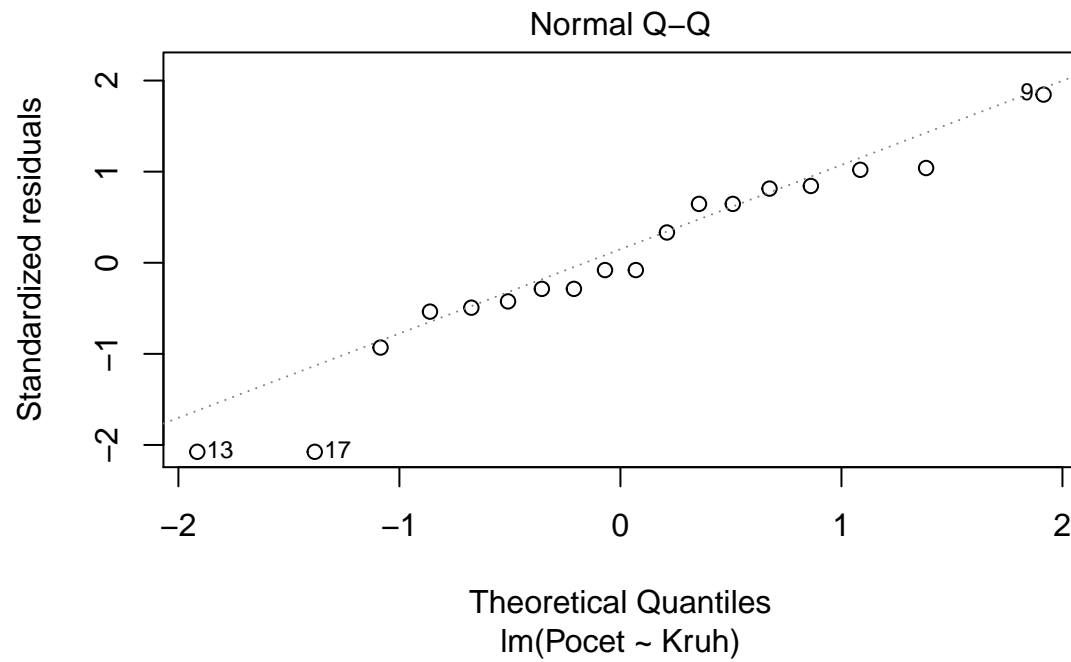
```
linmod <- lm(Pocet ~ Kruh, data = dat)
summary(linmod)
```

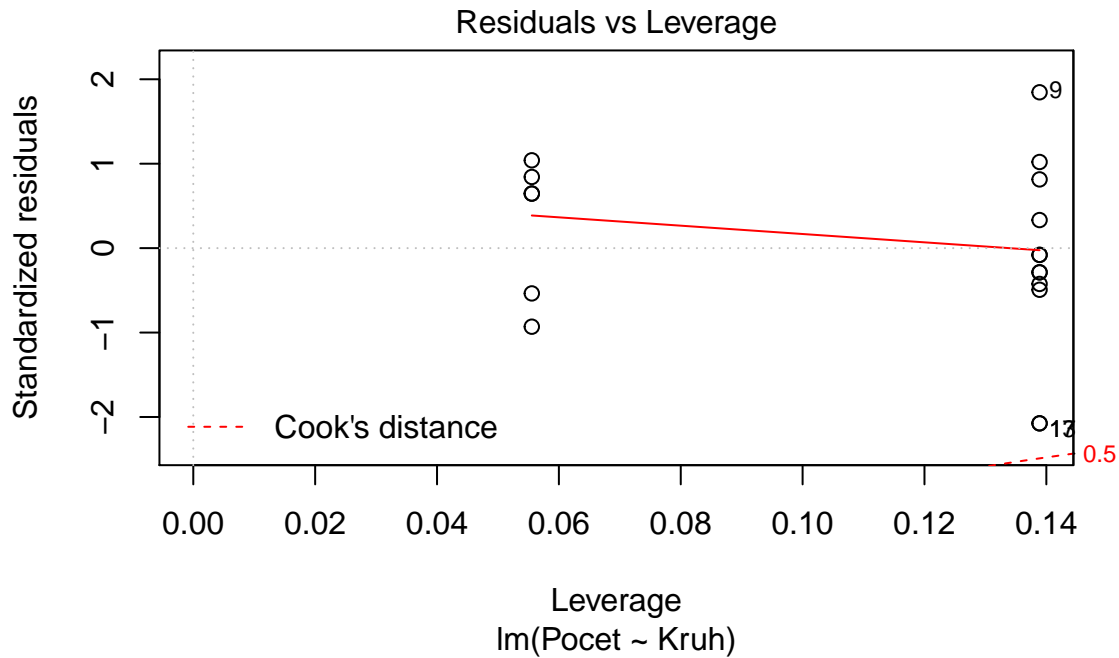
```
##
## Call:
## lm(formula = Pocet ~ Kruh, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.0556  -2.3056  -0.3889   3.7778   8.9444
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.056     3.256   0.324   0.75
## Kruh          13.333     1.507   8.845 1.47e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.222 on 16 degrees of freedom
## Multiple R-squared:  0.8302, Adjusted R-squared:  0.8196
## F-statistic: 78.24 on 1 and 16 DF, p-value: 1.472e-07
```

```
linmod2 <- lm(Pocet ~ Kruh + I(Kruh^2), data = dat)
summary(linmod2) #druha mocnina ani neni vyznamna, takze nema smysl dal tento model uvazovat

##
## Call:
## lm(formula = Pocet ~ Kruh + I(Kruh^2), data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3333 -1.5833  0.3333  2.7083  9.6667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.167      9.388  -0.657  0.5212
## Kruh          22.000     10.661   2.064  0.0568 .
## I(Kruh^2)     -2.167      2.638  -0.821  0.4243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.276 on 15 degrees of freedom
## Multiple R-squared:  0.8375, Adjusted R-squared:  0.8159
## F-statistic: 38.66 on 2 and 15 DF, p-value: 1.205e-06
plot(linmod)
```







```
#lillie.test(residuals(linmod))
shapiro.test(resid(linmod))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(linmod)
## W = 0.94847, p-value = 0.4015
```

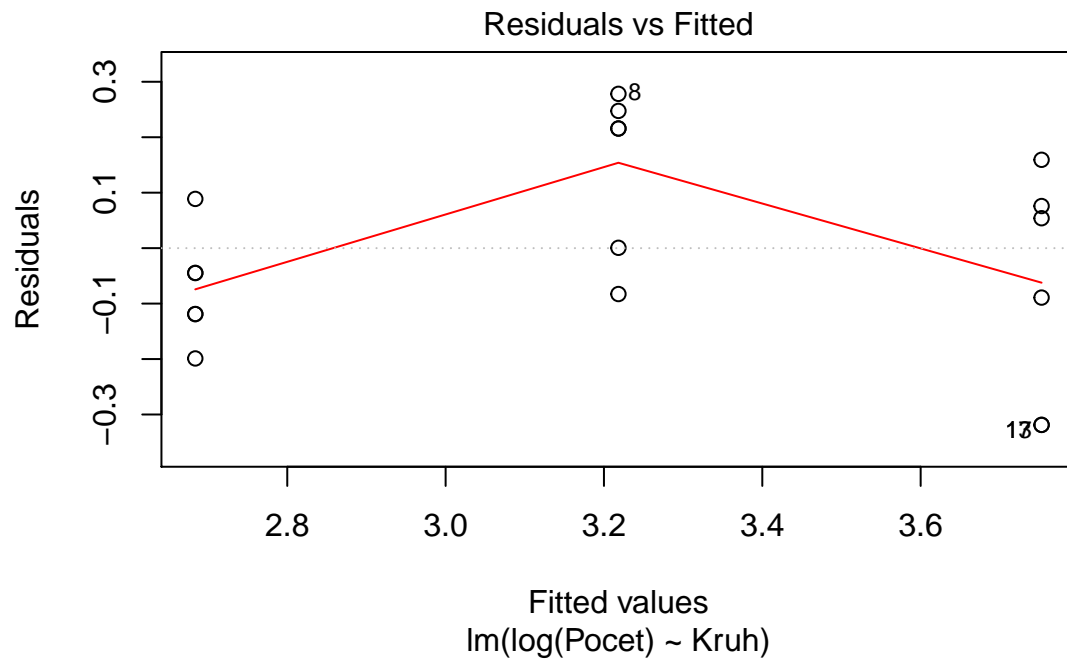
```
bptest(linmod) #bp zamita, mame heteroskedasticitu
```

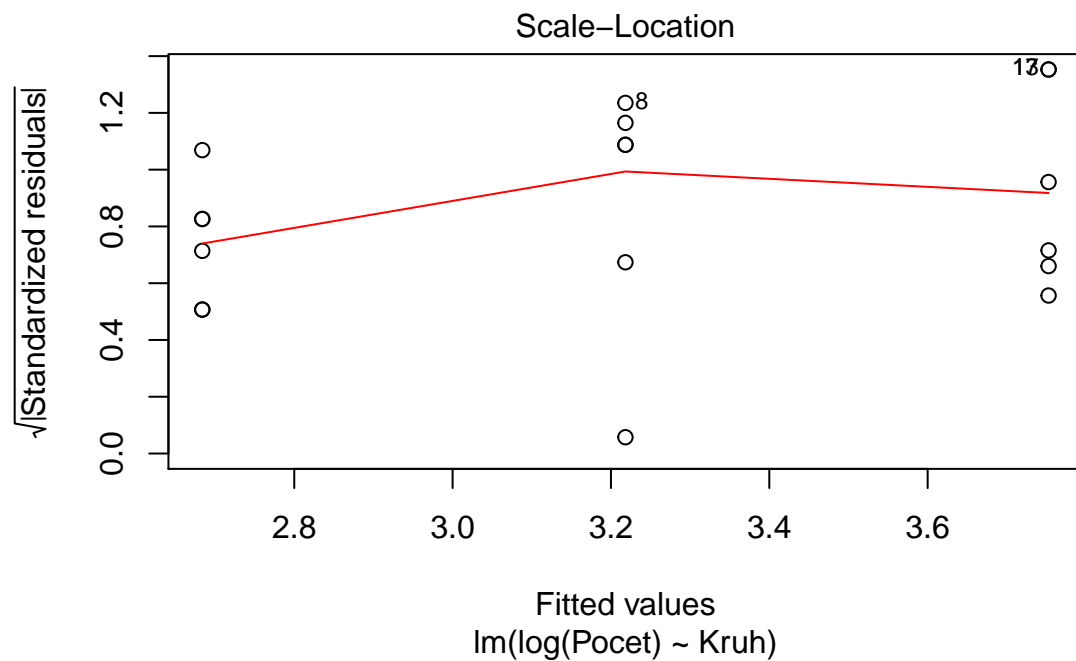
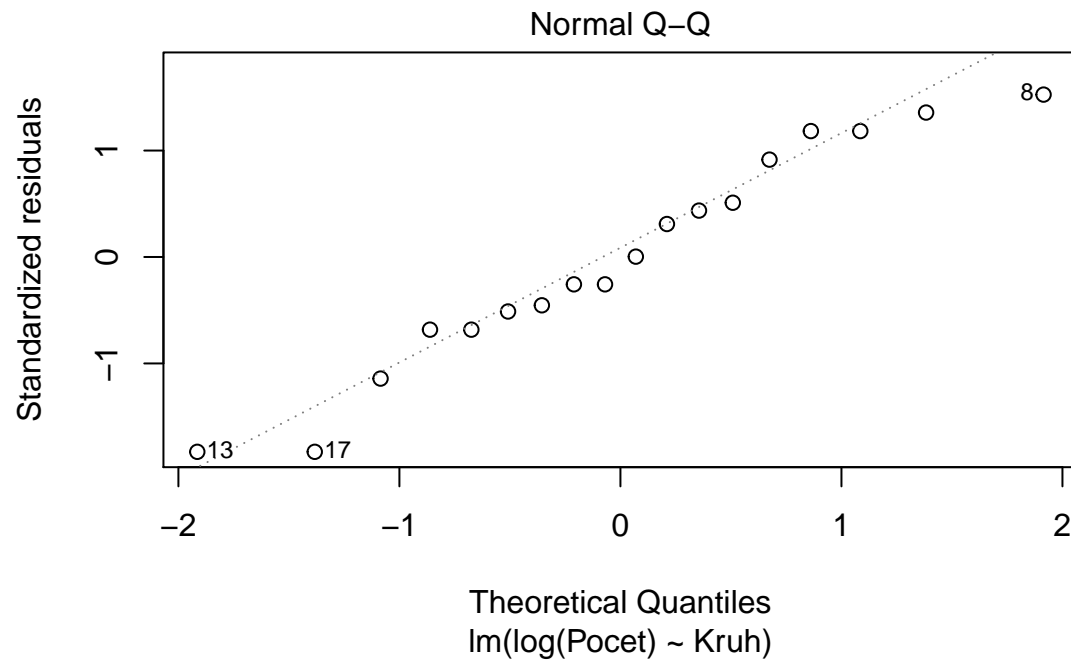
```
##
##  studentized Breusch-Pagan test
##
## data:  linmod
## BP = 7.7543, df = 1, p-value = 0.005358
```

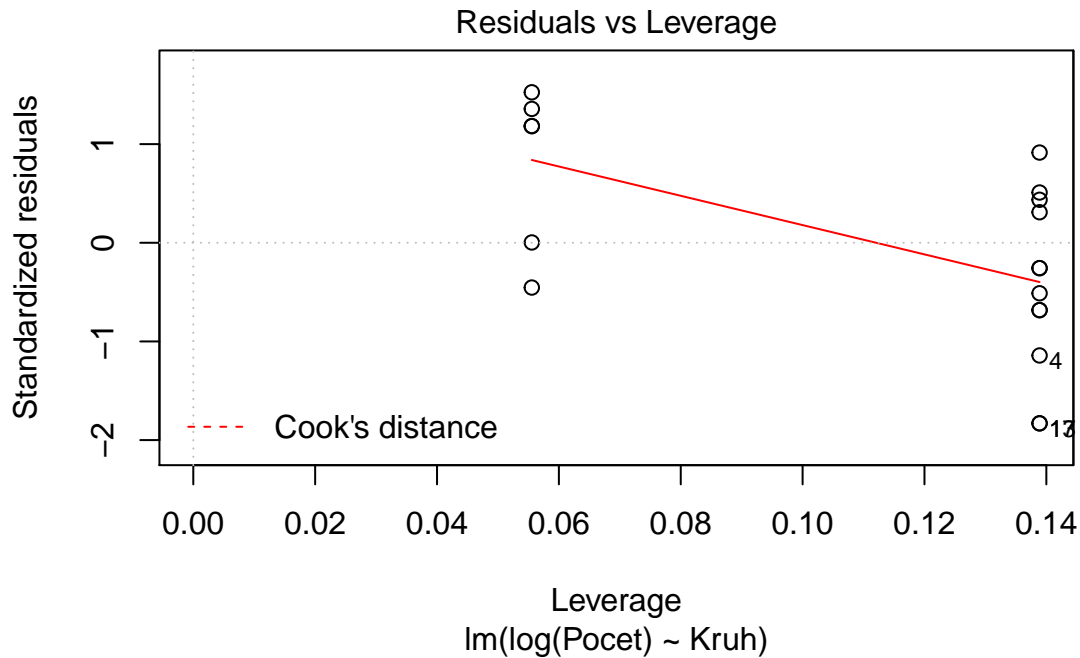
```
#Final model => linmod_log
linmod_log <- lm(log(Pocet) ~ Kruh, data = dat)
summary(linmod_log)
```

```
##
## Call:
## lm(formula = log(Pocet) ~ Kruh, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31873 -0.11145 -0.02209  0.14167  0.27823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  2.14938    0.11703   18.366 3.54e-12 ***
## Kruh         0.53445    0.05417    9.865 3.32e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1877 on 16 degrees of freedom
## Multiple R-squared:  0.8588, Adjusted R-squared:  0.85
## F-statistic: 97.33 on 1 and 16 DF,  p-value: 3.315e-08
plot(linmod_log)
```







```
lillie.test(residuals(linmod_log))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  residuals(linmod_log)
## D = 0.10418, p-value = 0.8718
```

```
shapiro.test(resid(linmod_log))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(linmod_log)
## W = 0.95586, p-value = 0.5242
```

```
bptest(linmod_log)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  linmod_log
## BP = 2.0629, df = 1, p-value = 0.1509
```

```
#jiz je vse v poradku; testy vychazeji i graficky by se to dalo ukecat
#navic i intercept je vyznamny, takze bonus navic; R^2_adj 0.85, to vypada fajn
```