NEX1

Kajetán Poliak, Adam Novotný

24. října 2019

NEX - první domácí úkol

1. část

- Namerte data: experiment nastavte tak, aby byl znahodneny !!!!! Poradi mereni si poznamenejte pro dalsi kontrolu pripadne zavislosti na poradi mereni. Jednotlive lidi ve skupine berte jako ruzne operatory experimentu (blokujte). 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 3 technik dominantní, nedominantní a obou rukou), kde každé meření trvalo 10s. Celkově tedy máme 18 pozorování se čtyřmi vysvětlujícími proměnnými, operátor `Jmeno` na dvou úrovních, použití ruky `Ruka` na třech úrovních, velikost kruhu `Kruh` také na třech úrovních a pořadí `Poradi` pro oba operátory.

Pořadí meření bylo znáhodněno pomocí generátoru náhodné posloupnosti z www.random.org. Jako první se vygenerovala náhodná posloupnost celých čísel od 1 do 9, která určovala pořadí meření. Následně se pomocí náhodné posloupnosti od 1 do 3 přiřadila jedna z úrovní proměnné `Ruka` ke každému ze tří papírů obsahujícímu tři různé velikosti kruhu.

Na meření mohlo mít vliv spousta faktorů. Mezi tyto faktory patří rozpoložení operátora, tedy různé polohy sedu při experimentu, psychický nátlak okolí nebo dlouhodobý fyzický a psychický stav operátora. Důvodem také mohlo být, že se experimentu každý z operátorů podrobil jiný den. Dalšími faktory mohly být nepřesnosti meření času druhým člověkem, různé psací potřeby operátorů nebo způsob držení papíru.

2. část

- Spoctete zakladni statistiky (mean, median a sd pro jednotlive 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?

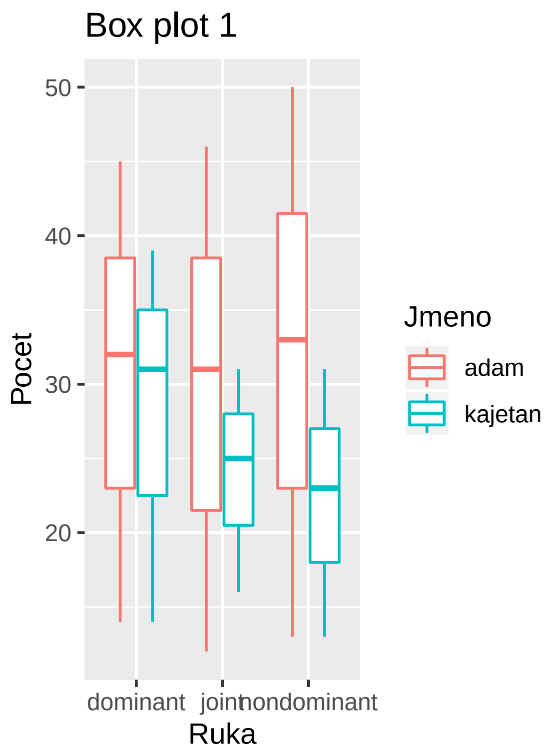
Zobrazme si nejprve základní statistiky, tedy postupně minimum, první kvartil, medián, třetí kvartil, maximum a směrodanou odchylku pro všechny úrovně.

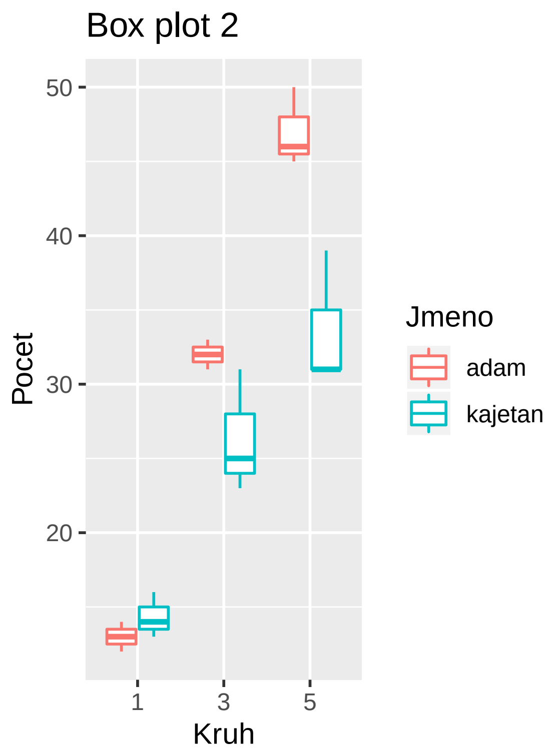
```
##              adam   kajetan dominant nondominant    joint     1 cm
## Min.     12.00000 13.000000 14.00000    13.00000 12.00000 12.00000
## 1st Qu.  14.00000 16.000000 18.25000    15.50000 18.25000 13.00000
## Median   32.00000 25.000000 31.50000    27.00000 28.00000 13.50000
## Mean     30.66667 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 průměr odlišný, kdy Adam má průměr větší. Oba mají podobná minima ale maximum Adama je o celých 11 vyšší. Stejně tak je vyšší střední hodnota počtu teček Adama. Maxima u proměnné `Ruka`, tedy dominant, nondominant a joint souvisí s operátorem Adam.

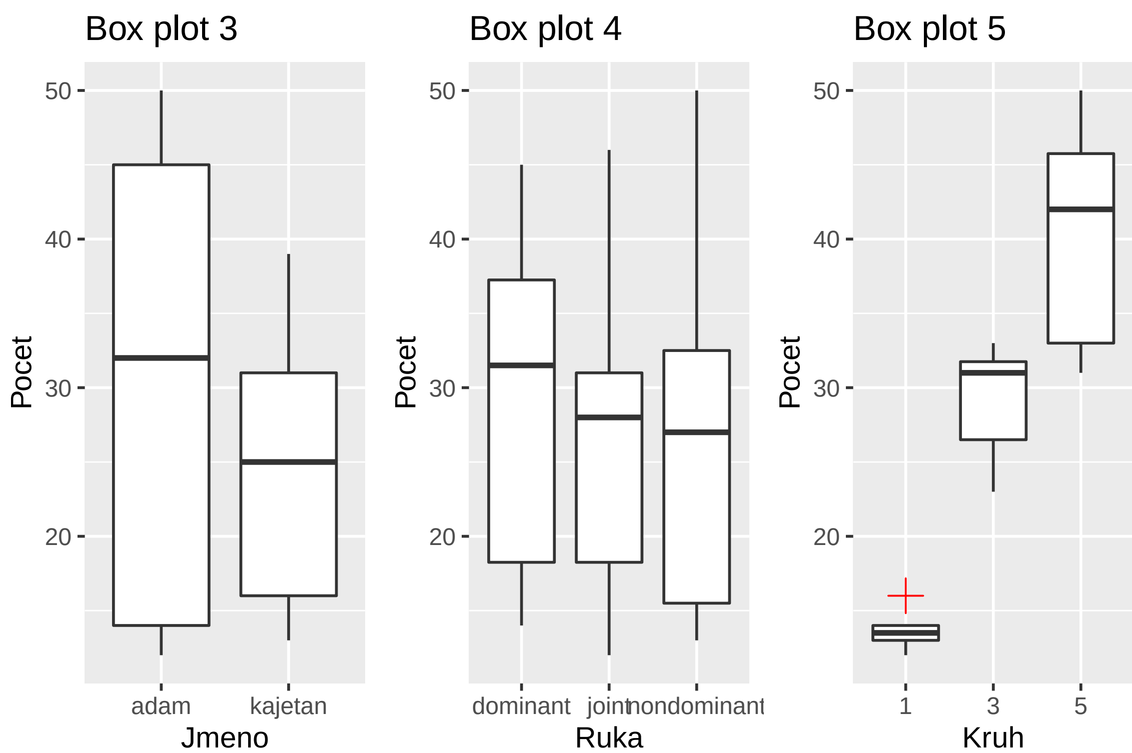
Podle očekávání vidíme, že hodnoty charakteristik proměnné `Kruh` jsou vyšší se zvětšujícím se poloměrem kruhu. Směrodatná odchylka u proměnných adam, kajetán, dominant, nondominant a joint dosahují poměrně vysokých hodnot. To je vysvětlitelné tím, že počet teček se pro každý kruh radikálně lišil. Nejnižší průměrnou hodnotu z různých úrovní proměnné `Ruka` má joint.

Zobrazme si nyní postupně boxploty, interaction ploty a effects ploty.





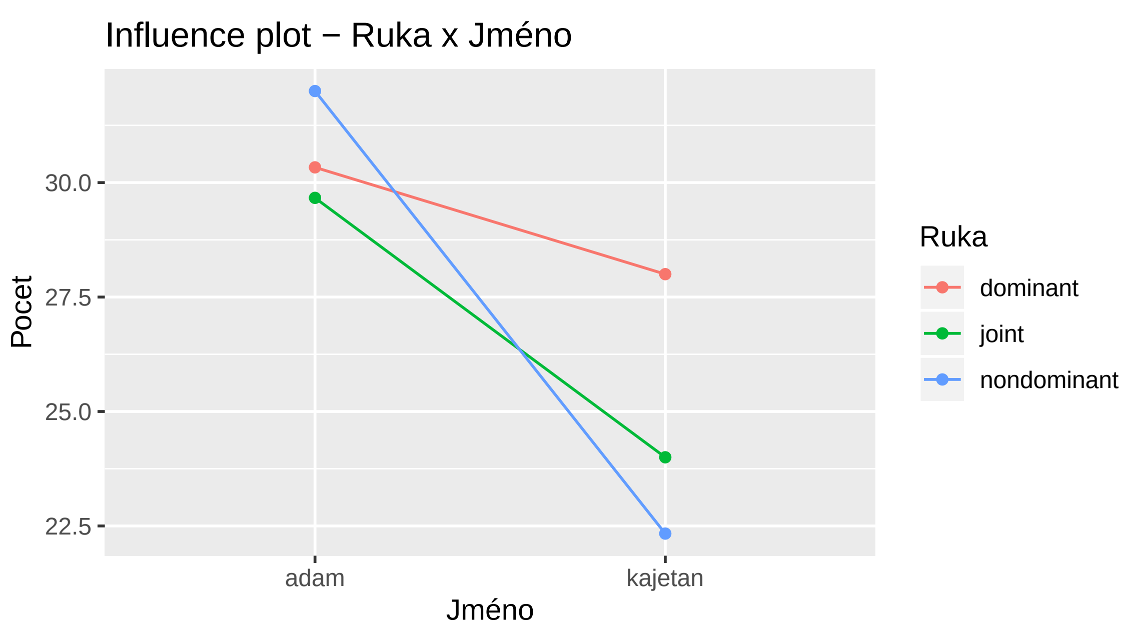
Z boxplotu 1 vidíme, že Adam dosahoval ve všech úrovních `Ruka` větších průměrů, nejméně však u úrovně dominant. U úrovní nondominant a joint má Adam vyšší průměry i znatelně větší variance. V úrovních `Kruh` v boxplotu 2 dosahuje Kajetán většího průměru u nejmenšího kruhu, naopak Adam u ostatních dvou úrovních. Lze vidět, že na úrovních 3 cm a 5 cm má Adam větší počet teček i v rámci statistické odchylky.



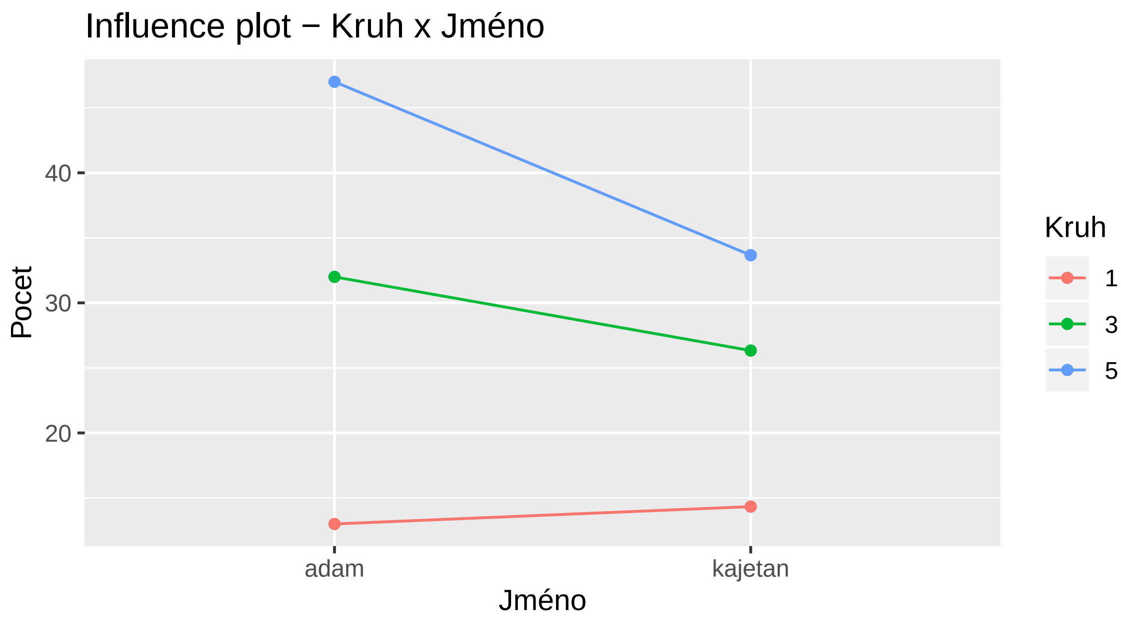
Další boxploty nevypovídají, zdali je významný rozdíl mezi úrovněmi `Jmeno` a `Ruka`, naopak pozorujeme rozdíl mezi úrovněmi `Kruh`. V případě úrovně 1 cm pozorujeme i outlier, který však je dán malým počtem pozorování a zdánlivý velký rozdíl mezi hodnotami 16 a ostatními, jak můžeme vidět na hodnotách 1 cm.

```
## [1] 12 13 14 14 13 16
```

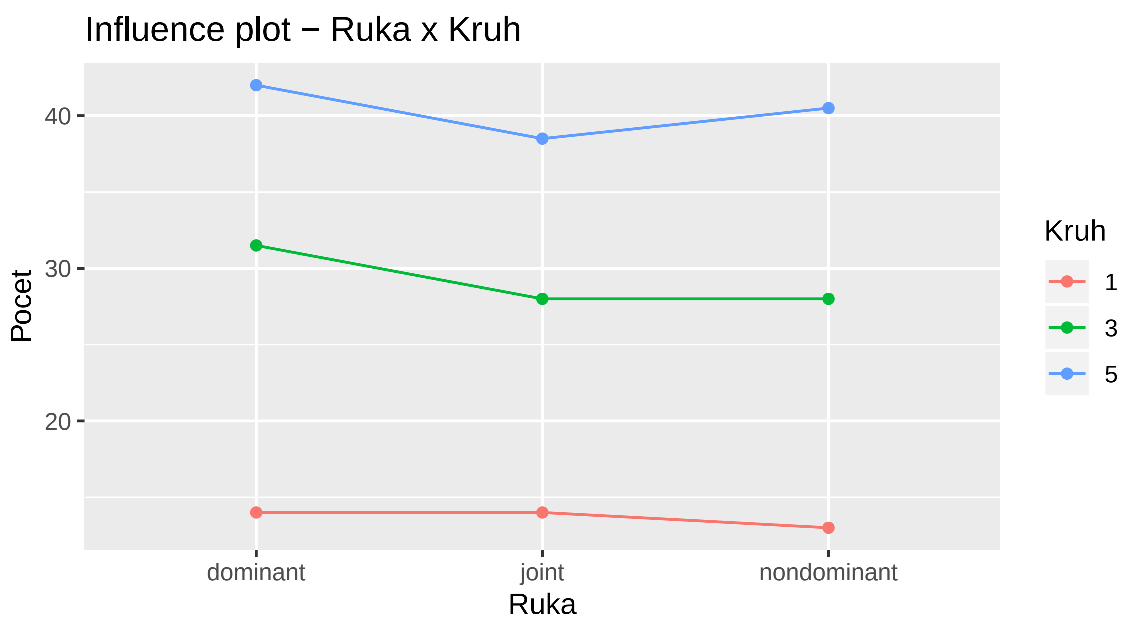
V případě influence plotů v závislosti ruce na jménu pozorujeme, že ve všech úrovních má Adam větší v pořadí, přičemž maxima dosahuje u nondominant, dále dominant a joint. Kajetán maxima dosahuje u dominant, dále joint a nondominant. Interakce mezi proměnnými by mohla být významná.



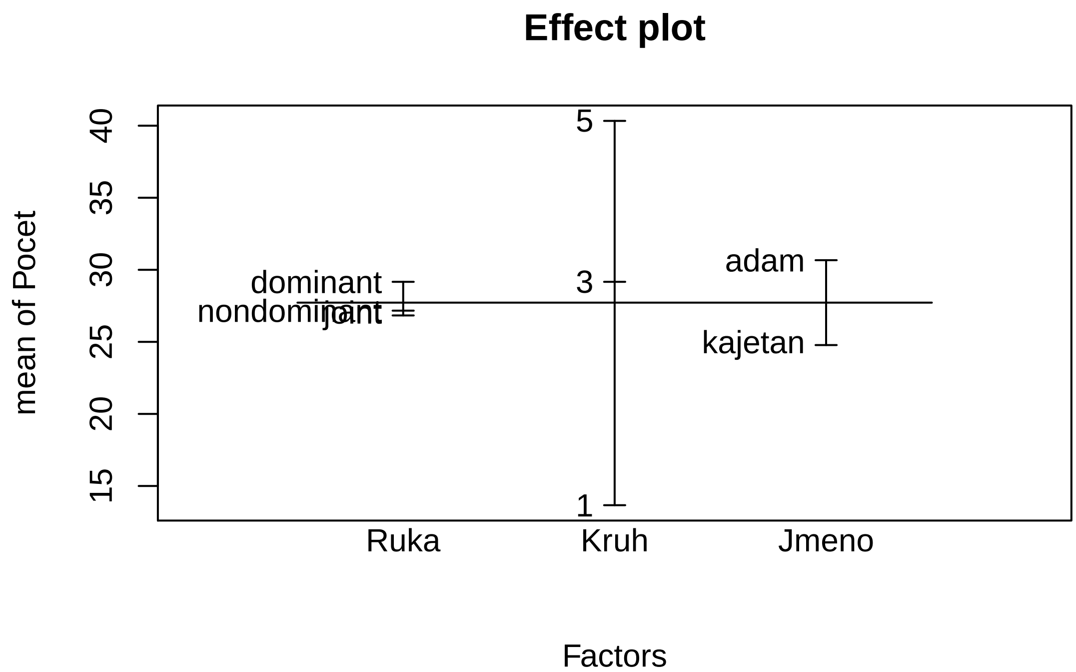
V případě závislosti kruhu na jménu pozorujeme, že v obou případech je největší počet pro 5 cm, dále 3 cm a 1 cm. Jak jsme již pozorovali dříve, Kajetán dosahuje většího počtu na úrovni 1 cm, jinak na všech úrovních Adam. Interakce mezi proměnnými by mohla být významná.



V případě závislosti kruhu na ruce nepozorujeme žádnou interakci.



V případě effect plotu pozorujeme již zmíněné okolnosti.



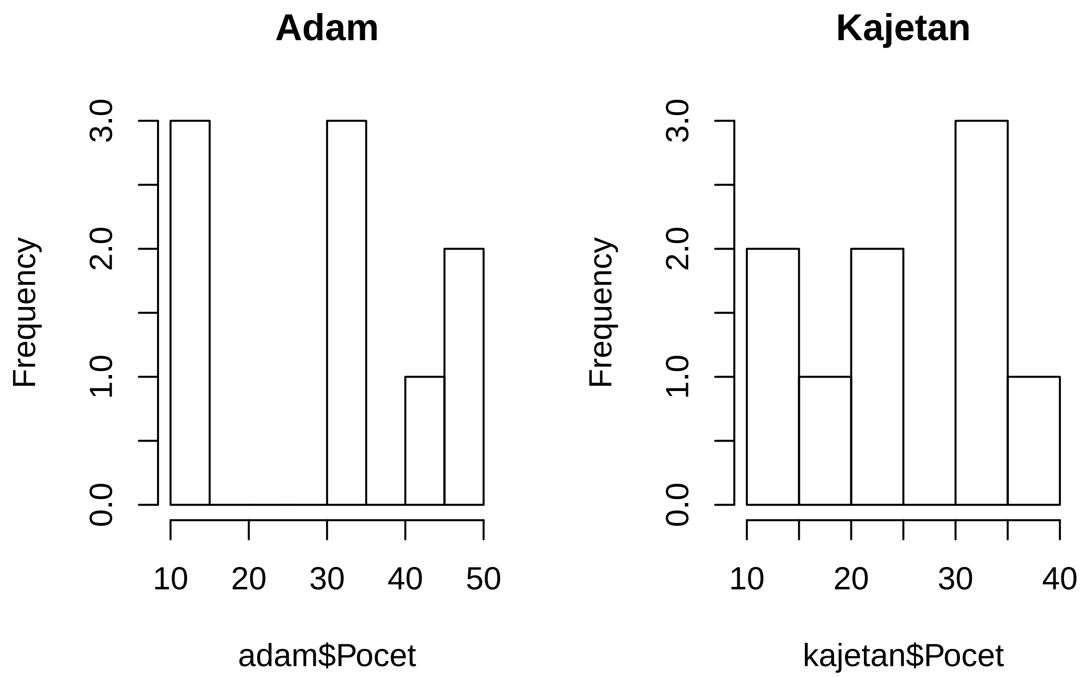
Před samotnou analýzou můžeme vidět, že úrovně proměnné `Kruh` by v analýze rozptylu a lineární regresi měly být signifikantní, u proměnných `Jmeno` a `Ruka` si nemůžeme být jistí. Můžeme však vidět, že rozptyly úrovní `Kruh` nebudou shodné, naopak u úrovní `Jmeno` a `Ruka` očekáváme stejné rozptyly. Střední hodnoty u úrovní `Kruh` taktéž nebudou shodné, ale stejně tak jako dříve u proměnných `Jmeno` a `Ruka` si nemůžeme být jisti.

3. část

- Zamerte se zvlaste na faktory ruka a velikost kola. Otestujte hypotezu o schodnosti rozptylu pro jednotlive urovne a vhodnym testem overte stejnost strednich hodnot. Provedte Tukey HSD a Fisher LSD test pro parove porovnani stednich hodnot jednotlivych skupin s vybranou korekci p-hodnoty.

Pro otestování shodnosti rozptylu a středních hodnot použijeme F-test, resp. t-test. U těchto testů předpokládáme normalitu dat. Nejprve si vykresleme histogramy a následně otestujme hypotézu normality Shapirovým–Wilkovým testem. V celé práci uvažujeme $\alpha = 0{,}05$.

V případě proměnné `Jmeno` histogramy nenaznačují případné porušení normality, navíc testy nezamítáme normalitu. Můžeme tedy přejít o otestování shodnosti rozptylů.



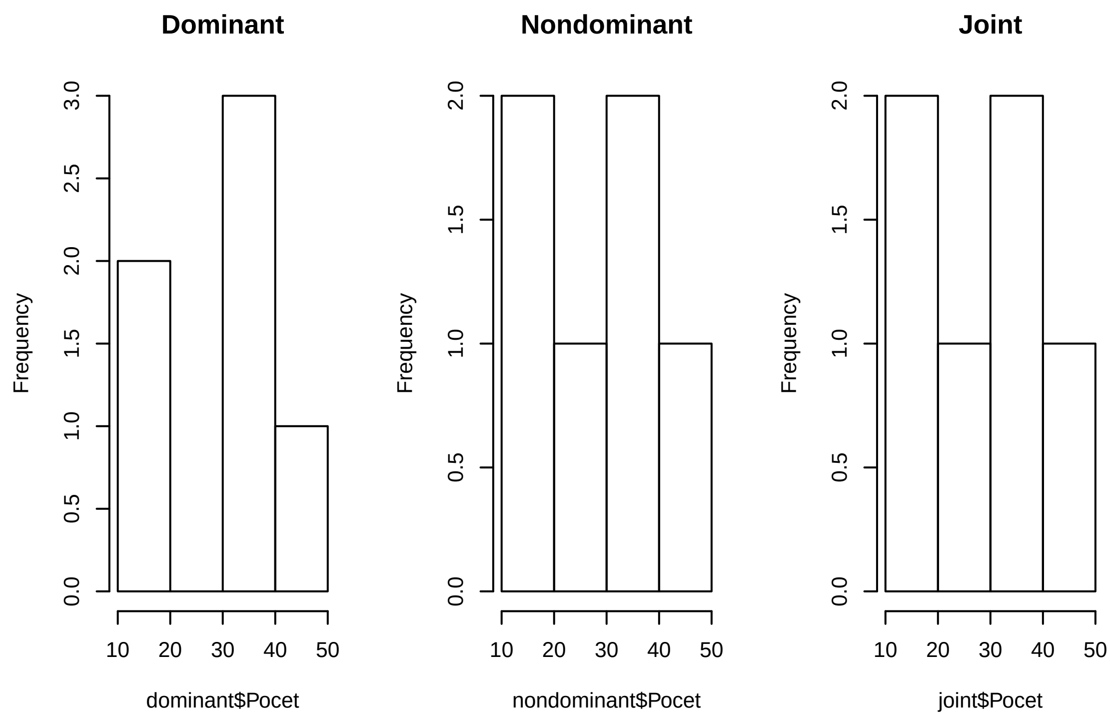

```
##
##  Shapiro-Wilk normality test
##
## data:  adam$Pocet
## W = 0.88219, p-value = 0.1656
```

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

```
##
##  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
```

F-test pro operátory vrací p-hodnotu 0.1827, nezamítáme tedy nulovou hypotézu o rovnosti rozptylů.

Přecházíme k faktoru `Ruka`.



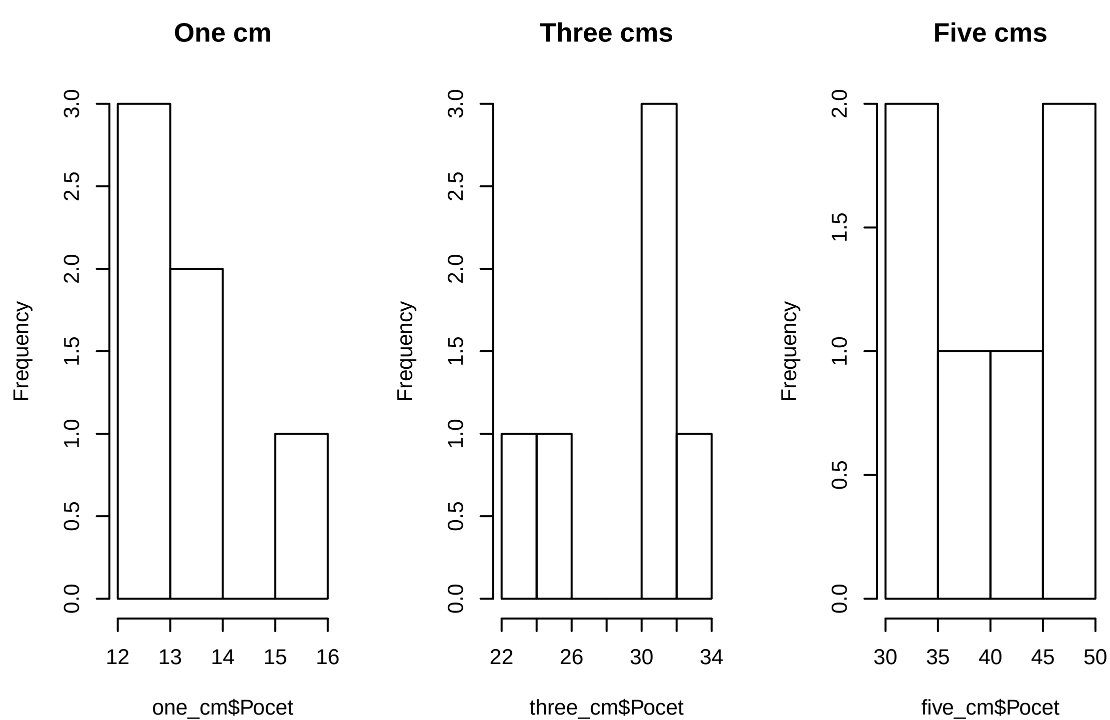

```
##
##  Shapiro-Wilk normality test
##
## data:  dominant$Pocet
## W = 0.89281, p-value = 0.3332

##
##  Shapiro-Wilk normality test
##
## data:  nondominant$Pocet
## W = 0.91686, p-value = 0.4831

##
##  Shapiro-Wilk normality test
##
## data:  joint$Pocet
## W = 0.9526, p-value = 0.7613

##
##  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

##
```

```
##  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
```

```
##
##  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
```

Pro všechny úrovně nezamítám 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. Můžeme dodat, že pokud by nám nešlo o jednotlivé dvě skupiny vůči sobě, mohli bychom použít párový test.

Jako poslední testujeme proměnnou `Kruh`.




```
##
```

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

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

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

##
##  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

##
##  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

##
##  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
```

P-hodnoty S-W testu pro úrovně 1 cm, 3 cm a 5 cm 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.

Můžeme tedy přejít k otestování středních hodnot postupně pro proměnné `Jmeno`, `Ruka` a `Kruh`. Jedinou podmínkou použití t-testu je normalita, která je již otestována. Pro proměnnou `Jmeno` použijeme t-test a pro proměnné `Ruka` a `Kruh` použijeme párový t-test s korekcí "hochberg". Pro proměnné `Jmeno` a `Ruka` uvažujeme pro test stejné rozptyly v rámci úrovní, naopak pro proměnnou `Kruh` ne.

```
##
##  Two Sample t-test
##
## data:  adam$Pocet and kajetan$Pocet
## t = 1.0171, df = 16, p-value = 0.3242
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.384906 18.162684
## sample estimates:
## mean of x mean of y
##  30.66667  24.77778
```

```
##
##  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 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
```

Pro proměnnou `Jmeno` nezamítáme hypotézu shodnosti středních hodnot, pro proměnnou `Ruka` taktéž nezamítáme mezi všemi úrovněmi, naopak u proměnné `Kruh` nulovou hypotézu shodnosti středních hodnot zamítáme.

Můžeme přejít k další podúloze v rámci této části, a to Tukey HSD a Fisher LSD test pro párové porovnání středních hodnot.

```
aov_celk =  aov(Pocet ~ Jmeno + Ruka + Kruh, data = dat) #ruka neni stat. vyznamna
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
```

```r
LSD1 <- LSD.test(aov_celk, "Jmeno"); LSD1$groups
```

```
##            Pocet groups
## adam    30.66667      a
## kajetan 24.77778      b
```

```r
LSD2 <- LSD.test(aov_celk, "Ruka"); LSD2$groups
```

```
##                Pocet groups
## dominant    29.16667      a
## nondominant 27.16667      a
## joint       26.83333      a
```

```r
LSD3 <- LSD.test(aov_celk, "Kruh"); LSD3$groups
```

```
##      Pocet groups
## 5 40.33333      a
## 3 29.16667      b
## 1 13.66667      c
```

Fisherův LSD test radí spojit všechny úrovně v rámci proměnné `Ruka`, tedy tuto proměnnou v analýze nepoužít. Naopak pro proměnné `Jmeno` a `Kruh` radí žádné úrovně nespojovat. Navíc ANOVA pro faktor `Ruka` signifikantní není, proto jej vyhodíme z modelu.

Pokračujme s Tukeyho HSD testem.

```r
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
```

```r
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       -2.3333333 -9.255132 4.588465 0.6509234
## nondominant-dominant -2.0000000 -8.921799 4.921799 0.7272003
## nondominant-joint     0.3333333 -6.588465 7.255132 0.9909481
```

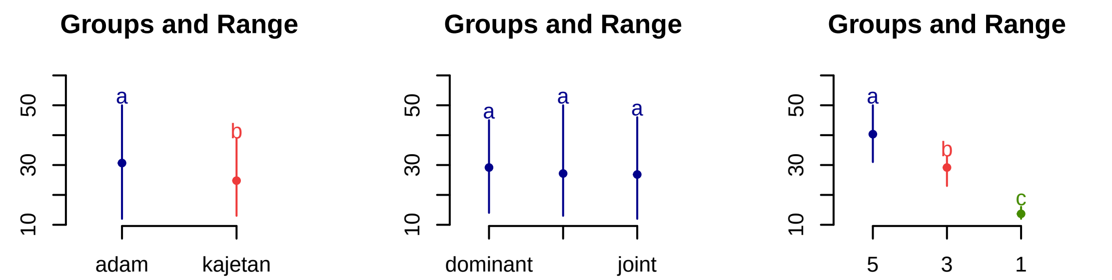
```r
TukeyHSD(aov_celk, "Kruh", ordered = FALSE, conf.level = alpha)
```

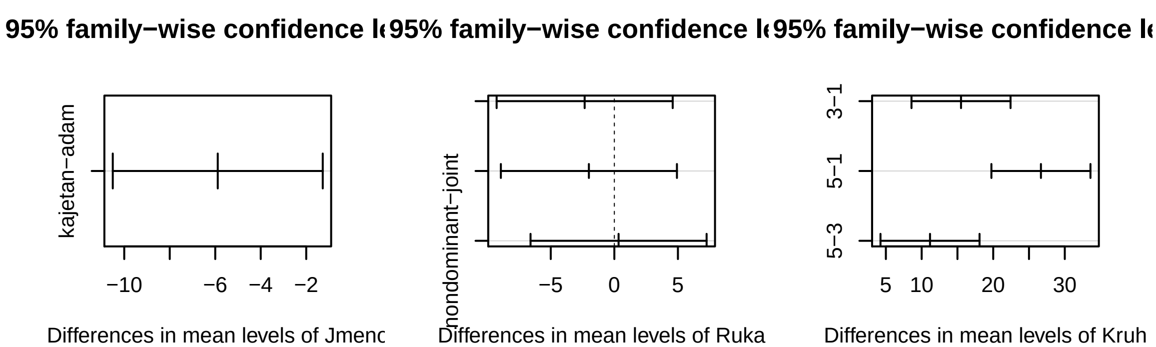
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Pocet ~ Jmeno + Ruka + Kruh, data = dat)
##
## $Kruh
```

```
##          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
```

Tukeyho HSD test vrací pro proměnnou `Jmeno` p-hodnotu rovnou 0.017, nezamítáme tedy hypotézu, že úrovně mají různou střední hodnotu. Tento test na proměnnou `Ruka` vrací p-hodnoty indikující, že všechny úrovně patří do stejné skupiny. Výsledek je tedy shodný s předchozími LSD testy.

Výsledky LSD a HSD testů můžeme přehledně vykreslit. Vrchní tři zobrazují výsledky LSD testu a příslušnost do jednotlivých skupin, spodní tři testy pak HSD test, kdy pokud interval prochází nulou, tak je vhodné úrovně spojit.





4.část

- Analyzute data pomoci ANOVA (vsechny promenne berte jako faktor) - s predpokladem, ze kazdy z ucastniku experimentu tvori jeden blok. - diskutujte vysledky, overte predpoklady, vykreslete QQ-plot, residua x fitted values, resida x cas, … - diskutujte vliv znahodneni experimentu a vyvoj rezidui v case (cislo mereni)

Ve třetí části jsme z analýzy rozptylu vyřadili proměnnou `Ruka`, která nebyla statisticky významná. V této části nejprve zkusme model, který bude obsahovat proměnné `Jmeno`, `Ruka`, jejich interakci a `Poradi`.

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Jmeno         1  156.1   156.1  24.546 0.000433 ***
## Kruh          2 2152.1  1076.1 169.252 5.48e-09 ***
## Poradi        1   22.9    22.9   3.601 0.084297 .
## Jmeno:Kruh    2  168.6    84.3  13.261 0.001173 **
## Residuals    11   69.9     6.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ukazuje se, že všechny zmíněné proměnné jsou významné na hladině významnosti $\alpha = 0.05$ až na proměnnou `Poradi`, jejíž p-hodnota avšak není tak vysoká. V dalším modelu ji ale uvažovat nebudeme.

```
##             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
```

Ověříme podmínky analýzy rozptylu, tedy normalita dat pomocí Shapirova–Wilkova testu a Lillieforsova testu na rezidua, navíc homoskedasticitu otestujeme Breutschovým–Paganovým testem a Levenovým testem.

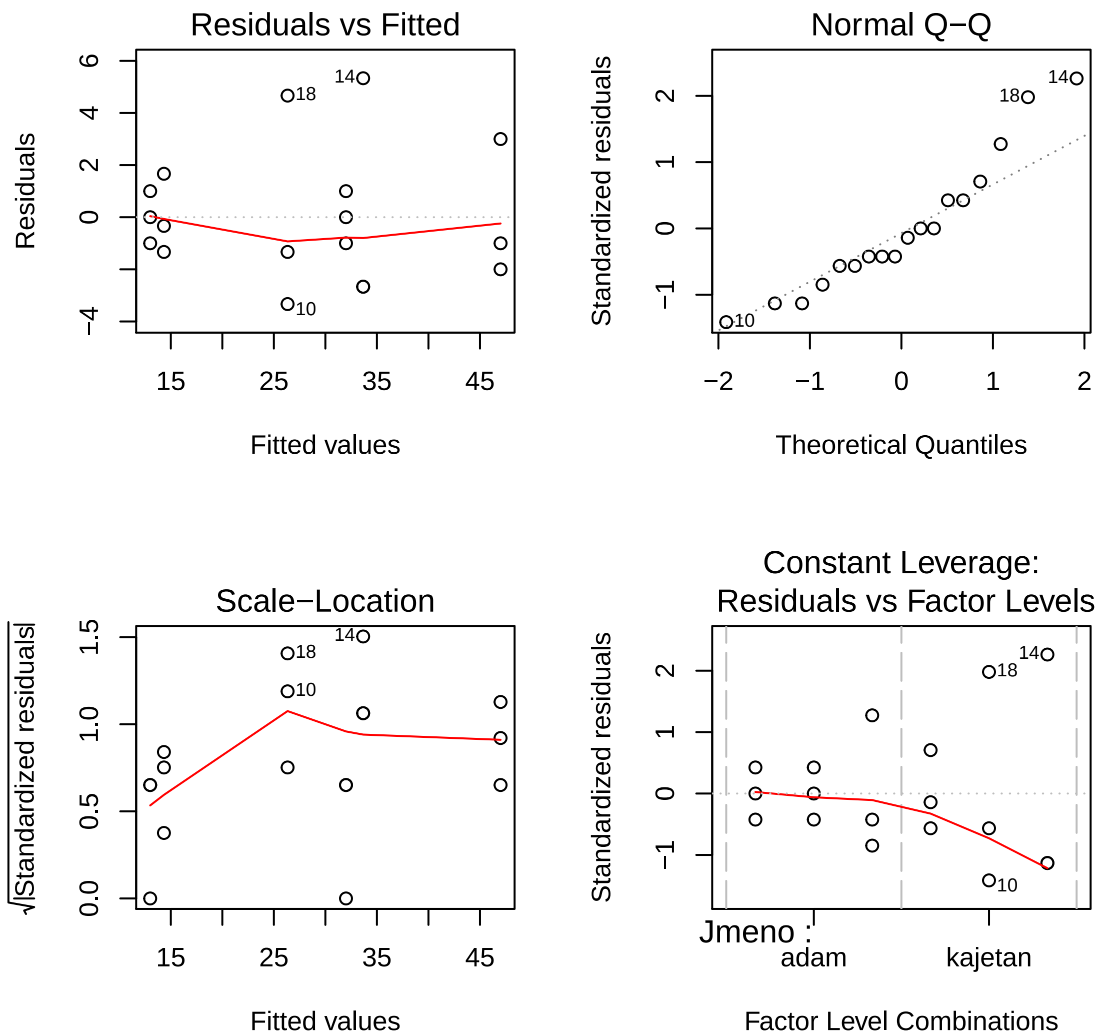
```
## 
##  Shapiro-Wilk normality test
## 
## data:  aov1$residuals
## W = 0.92219, p-value = 0.1413
```

```
## 
##  Lilliefors (Kolmogorov-Smirnov) normality test
## 
## data:  aov1$residuals
## D = 0.16667, p-value = 0.2062
```

```
## 
##  studentized Breusch-Pagan test
## 
## data:  aov1
## BP = 8.8802, df = 5, p-value = 0.1139
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.4291 0.8199
##       12
```

Vykresleme si ještě postupně rezidua vůči fitted values, Q-Q plot pro grafickou vizualizaci normality; scale location plot pro homoskedasticitu a rezidua vůči proměnné `Jmeno`.



Grafy mají rezidua víceméně rovnoměrně rozdělené a body zaujímají horizontální polohu. Jako outlier by být označen bod 14. V případě všech testů zamítáme a navíc graficky jsme ověřili, že podmínky analýzy rozptylu jsou splněny.

Diskutujme významnost proměnné `Poradi`. Vypišme si rezidua, zda se jejich hodnota nemění v čase. Jak můžeme vidět, tak ne razantně (výjimkou může být bod 14, což potvrzuje předchozí analýzu). To potvrzuje statistická nevýznamnost proměnné `Poradi`.

```
##          1          2          3          4          5          6
##  1.3983740 -0.1260163 -0.6016260 -0.1260163 -0.6504065 -0.1747967
##          7          8          9         10         11         12
##  0.3008130 -1.2723577  1.2520325 -1.5853659  1.2398374 -0.6341463
##         13         14         15         16         17         18
## -1.7926829  5.6829268 -1.8577236  0.6178862 -3.8902439  2.2195122
```

Za finální ANOVA model jsme na základě signifikance proměnných vybrali model Pocet ~ Jmeno*Kruh. Tento model splňuje podmínky ANOVA. Konkrétně, na základě p-hodnoty testů normality nezamítáme normalitu reziduí a stejně tak na základě testů homoskedasticity nezamítáme homoskedasticitu.

5.část

- Porovnejte a diskutujte vysledky z bodu 3 a 4.

Ve třetím bodu jsme několika testy (Fisher LSD, Tukey HSD) určili, že faktory `Ruka` a `Jmeno` mají úrovně pocházejíci z různých skupin.Jinak to bylo s faktorem `Kruh`, kde všechny úrovně patří do stejné. Pairwise t-test u faktoru `Jmeno` zamítá shodnost středních hodnot, u faktoru `Kruh` nikoliv. Model ANOVA má u faktorů `Jmeno` a `Kruh` p-hodnoty < 0.05, je zde signifikantní vliv na `Pocet` Závěry LSD, HSD testů a ANOVA si tedy odpovídají.

6. část

- Pokud data nesplnuji predpoklady pro pouziti ANOVA, diskutujte mozne transformace (logaritmicka, Box-Cox, ...) a duvod proc data predpoklady nesplnuji? Vyskyt outlieru, zpusob mereni, divny operator, ...

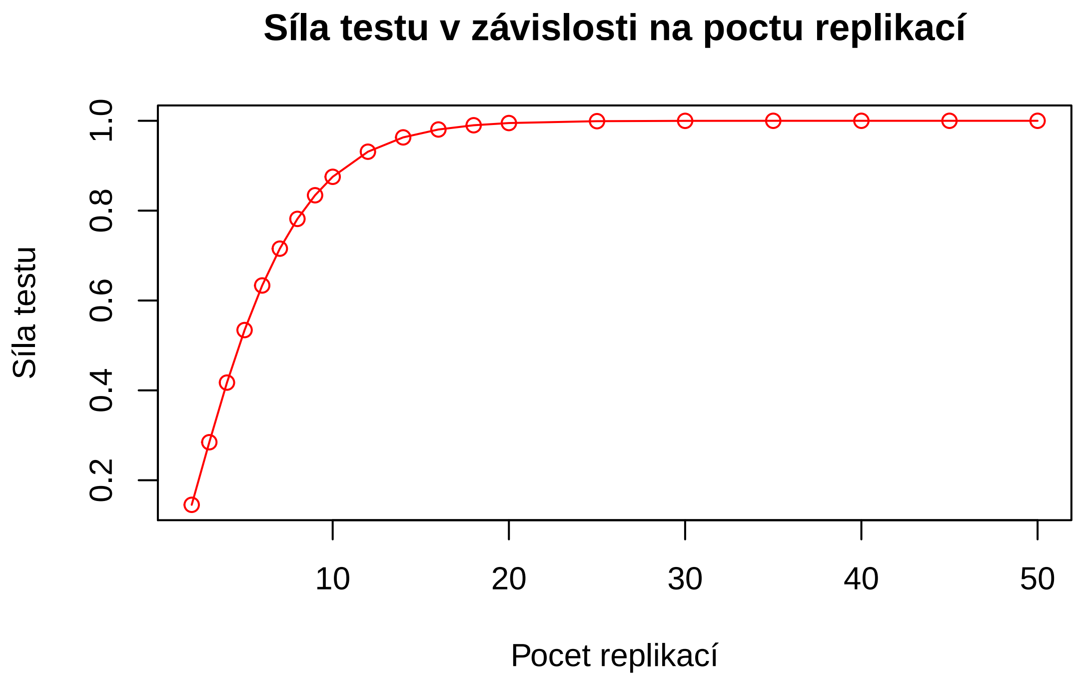
Data splňují předpoklady, tudíž nebudeme používat jakoukoli transformaci. V průběhu analýzy jsme identifikovali nějaké outliery, které ale neměly vliv na model.

7. část

- Vyberte nejvhodnejsi model a i kdyby nesplnoval predpoklady pro pouziti ANOVA reste nasledujici:
  - Spoctete silu testu v ANOVA (pro max. dvoufaktorovou analyzu - jeden z faktoru zanedbejte, nebo vezmete mereni pro jednu jeho konkretni uroven)
  - Predpokladejme, ze standartni odchylka disturbanci bude pro provadeny experiment 4 a maximalni (pro nas signifikantni) rozdil, ktery chceme detekovat je 5 bodu v kruhu. Spoctete pocet potrebnych replikaci, aby sila vysledneho testu byla vetsi nez 0.9

```
##
##      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
```

```
##
##      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
```



Pro výpočet síly testu bereme 5% hladinu významnosti. Předpokládáme, že statistická odchylka disturbancí pro experiment je 4 (rozptyl = 16). Síla testu pro dvouúrovňový faktor `Jmeno`, kde máme 9 pozorování, je 0.834. Pro sílu testu větší než 0.9 bychom potřebovali alespoň 11 replikací. Tato skutečnost je viditelná i z grafu závislosti síly testu na počtu replikací.

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 s sebou a vyberte vhodnejsi - overte predpoklady pro pouziti vybraneho modelu a vykreslete QQ-plot, residua x fitted values, resida x prumer , ...

Sestrojíme lineární regresní model, kde neuvažujeme bloky, jedinou významnou proměnnou je tedy proměnná `Kruh`, která je kvantitativní.

```
##
## 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
```

V lineárním regresním modelu Pocet ~ Kruh, dostáváme $R^2_{adj} = 82$ a proměnnou kruh jako signifikantní na 1% hladině významnosti.

Zkusme model s kvadratickou proměnnou.

```
##
## 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
```

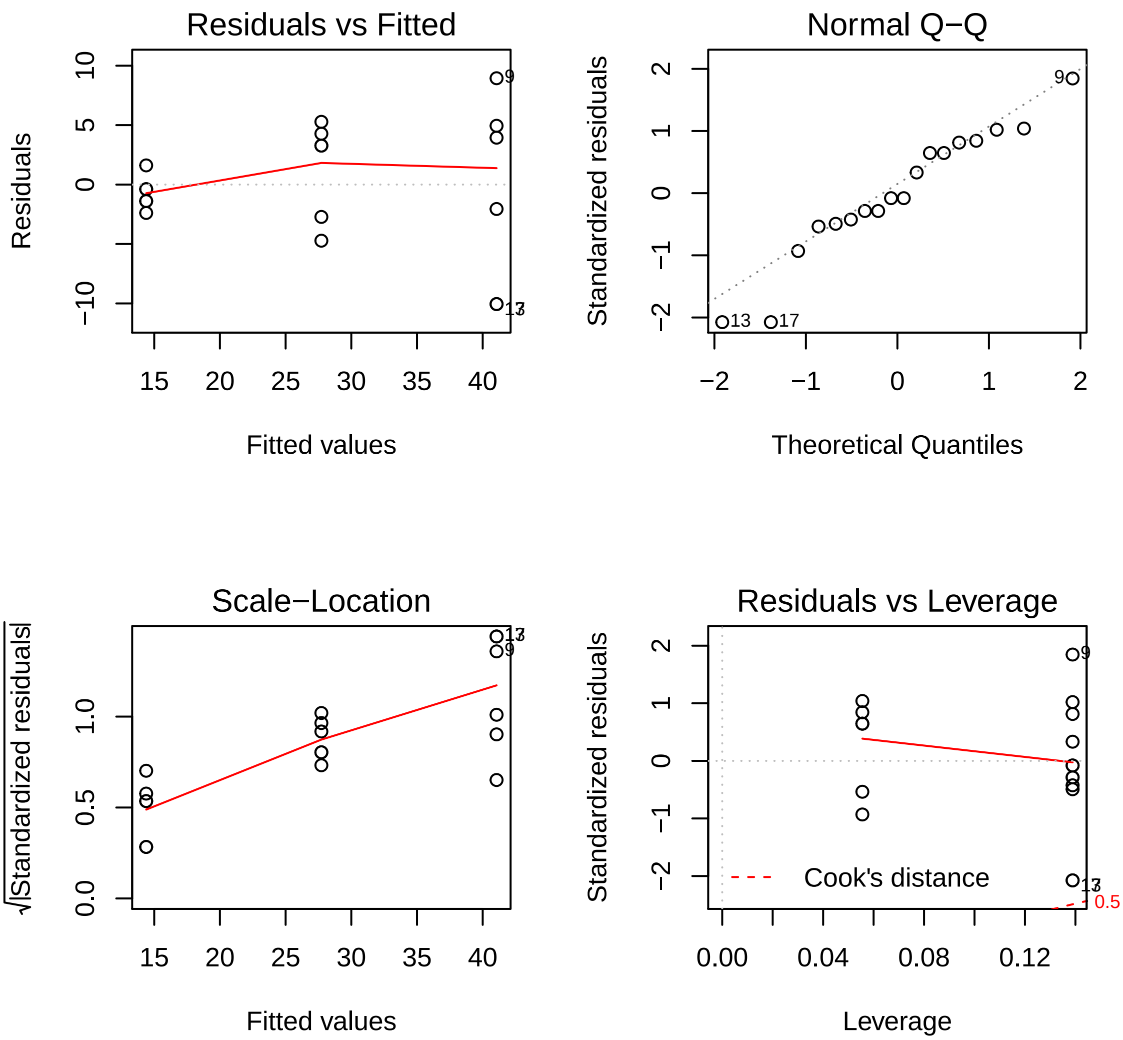
Pokud do modelu přidáme i druhou mocninu, získáváme nesignifikantní proměnné a R^2 se mírně snížil. Tento model dále neuvažujeme.

Ověřme nyní podmínky modelu jak testováním, tak graficky.

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  residuals(linmod)
## D = 0.13008, p-value = 0.5816
```

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

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



Lillieforsův test ani Shapireův-Wilkeův test nezamítají hypotézu normality reziduí. Breuschův-Paganův test zamítá homoskedasticitu. Problém heteroskedasticity je však zřetelný i z grafu Scale-Location.

Z důvodu výskytu heteroskedasticity použijeme logaritmickou transformaci prvního z modelů.

```
##
## 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 ***
## ---
```

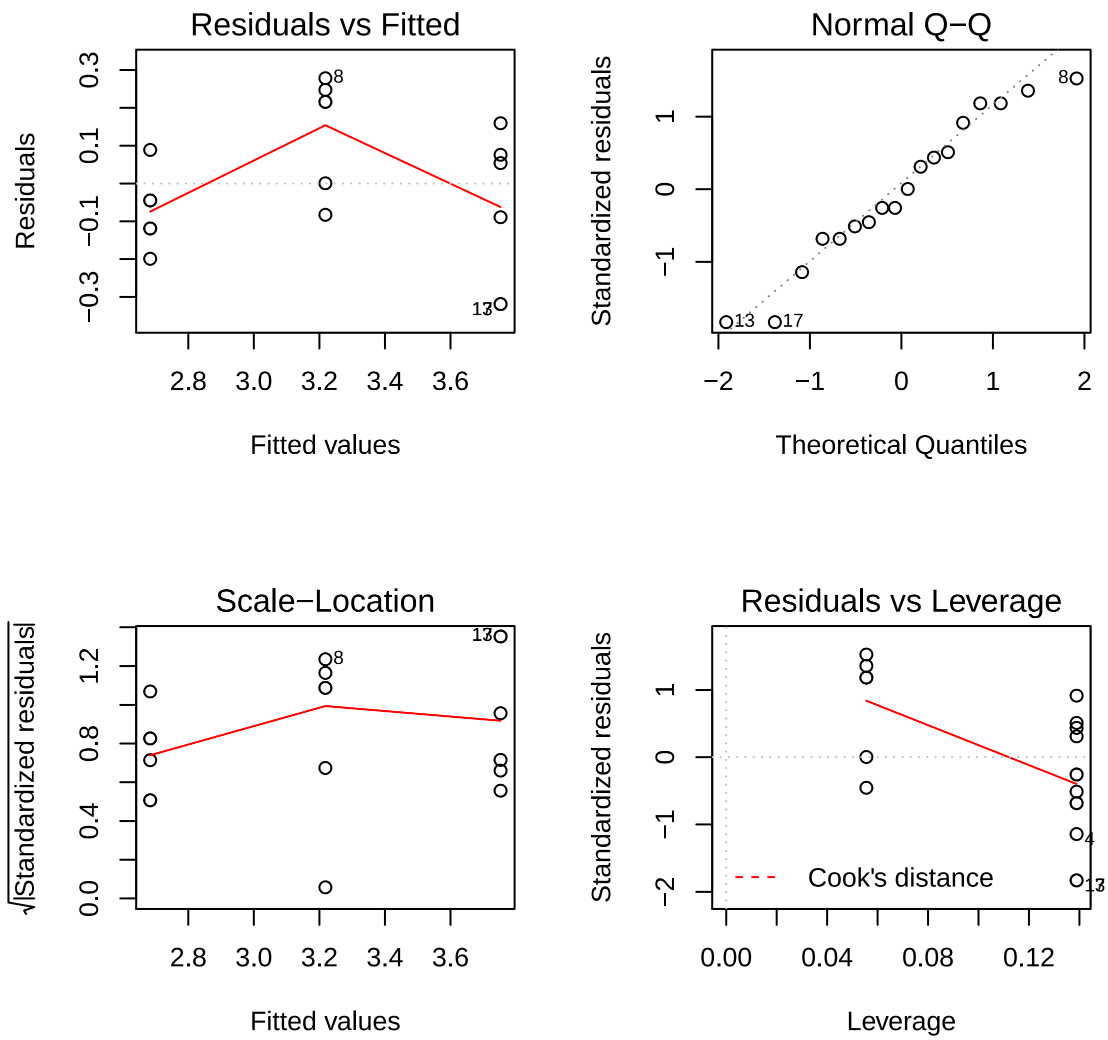
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1877 on 16 degrees of freedom
## Multiple R-squared:  0.8588, Adjusted R-squared:   0.85
## F-statistic: 97.33 on 1 and 16 DF,  p-value: 3.315e-08
```

Jak Intercept, tak Kruh jsou signifikantní, hodnota $R^2 = 85$ je postačující. Nezamítáme jak normalitu reziduí, tak homoskedasticitu. Vidíme zlepšení na Scale-Location grafu. Máme tedy finální model.

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

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

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


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage