**Vypočítejte numerické popisné charakteristiky sloupce Z z datové sady data.txt a zobrazte základní grafy. Výsledky řádně okomentujte a interpretujte.**

1. **Základní statistika**

* Průměr 3. Sloupce (z) je:

mean(data1[,3])

[1] 1077.351

* Minimum 3. Sloupce (z) je:

min(data1[,3])

[1] 446.42

* Maximum 3. Sloupce (z) je:

max(data1[,3])

[1] 1658.12

* Variace 3. Sloupce (z) je:

var(data1[,3])

[1] 55481.26

* Lze udělat summary, kde se vypíše základní statistika ze 3. Sloupce

summary(data1[,3])

Min. 1st Qu. MedianMean 3rd Qu. Max.

446.4 914.1 1074.0 1077.0 1233.0 1658.0

* Směrodatná odchylka 3. Sloupce:

sd(data1[,3])

[1] 235.5446

* Pro šikmost a špičatost musím zapnout knihovny e1071

library(e1071)

* Šikmost

skewness(data1[,3])

[1] -0.09617197

* Špičatost

>kurtosis(data1[,3])

[1] -0.3865235

* medián

>median(data1[,3])

[1] 1073.53

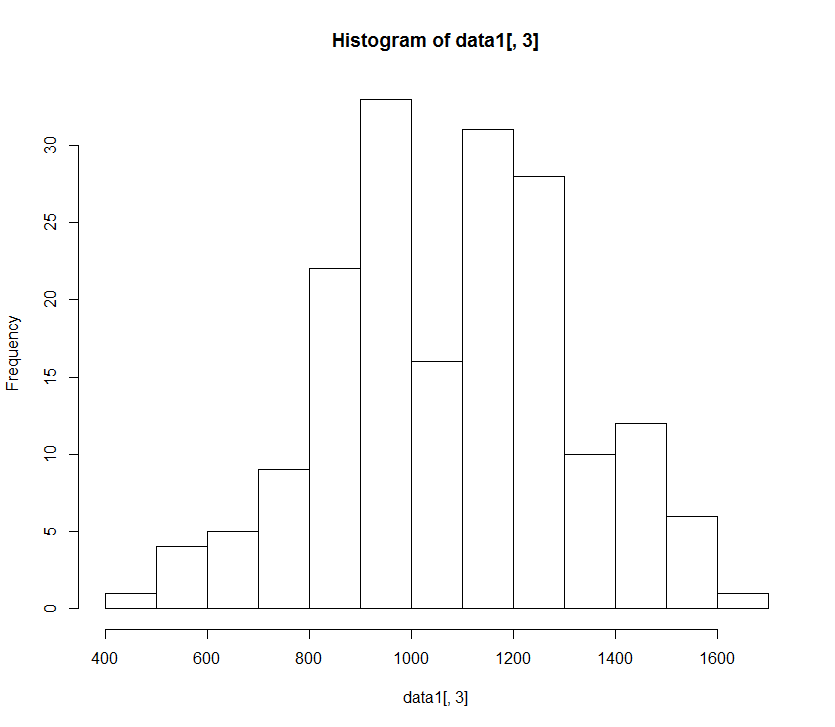
Medián, prostřední hodnota v datech, ve sloupci z

* IQR

Iqrje důležitá součást numerických charakteristik –mír variabilit

* Histogram

hist(data1[,3])



Histogram znázorňuje výskyt dat v intervalech.

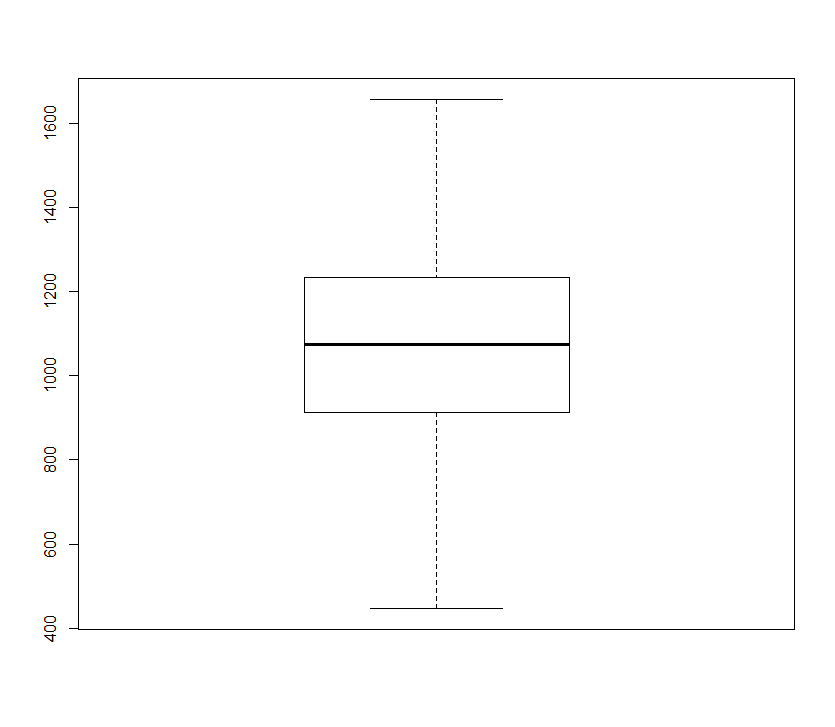
* Boxplot – whiskers plot

boxplot(data1[,3])

Boxplot

Prostřední černá úsečka nám vykresluje střední hodnotu – medián.

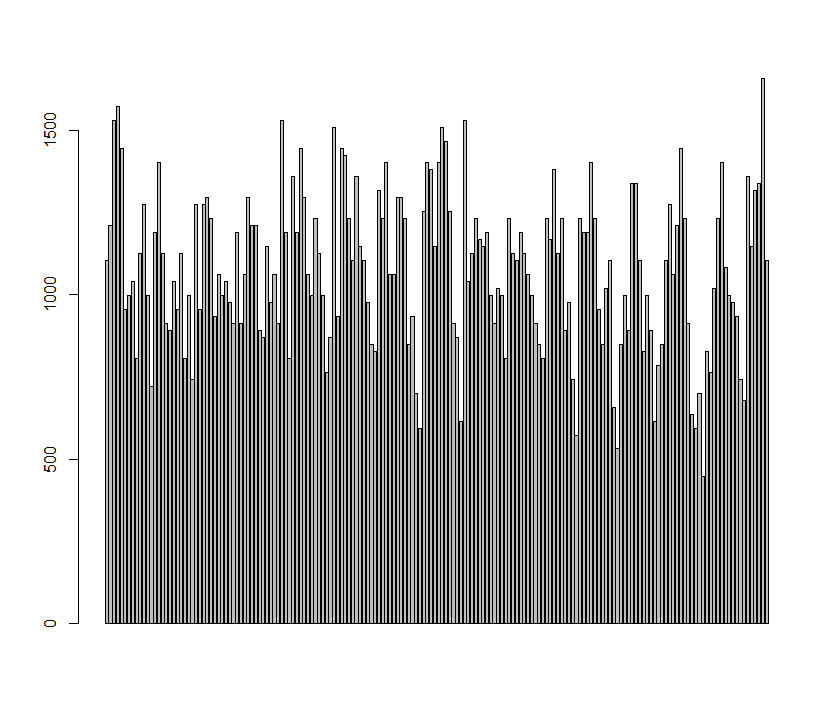
Spodní čára obdélníku je první kvartil a horní čára obdélníku je 3. kvartil



* Barplot

barplot(data1[,3])

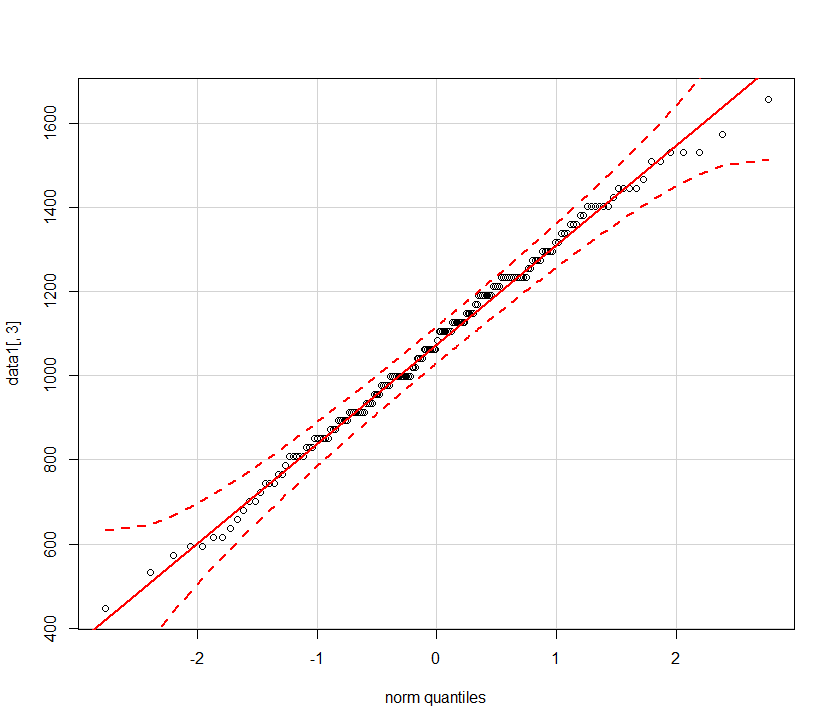
znázorňuje nám velikosti jednotlivých entit



* Qqplot

zapnutí knihovny car – bez toho nelze vytvořit qq plot

qqPlot(data1[,3])



Shapiro test, abych zjistila, zda data mají normální rozdělení

shapiro.test(data1[,3])

Shapiro-Wilk normality test

data: data1[, 3]

W = 0.9947, p-value = 0.7757

Dála mjí normální rozdělení, p-value je vetší jak 0,05.

* **Vhodně interpolujte datovou sadu data1.txt (sloupec Z) pomocí krigingu. V úvahu vezměte jak model bez zahrnutí jakéhokoliv trendu, tak i model, který zohledňuje trend v datech. Porovnejte oba modely. Výsledky interpolace a interpolační chyby spolu s isoliniemi vykreslete a slovně okomentujte.**

1. **KRIGING (krig – důlní inženýr, 1150)**

Instalace knihovny geoR (důležitá knihovna pro kriging

geodata<- as.geodata(data1, coords.col=1:2, data1.col=3)

Vytvoření 4. Sloupce a převedení na geodata

data1[,4] <- log(data1[,3])

hodnoty zlogaritmuji pro snazsipraci s nimi

geodata<- as.geodata(data1, coords.col=1:2, data.col=4)

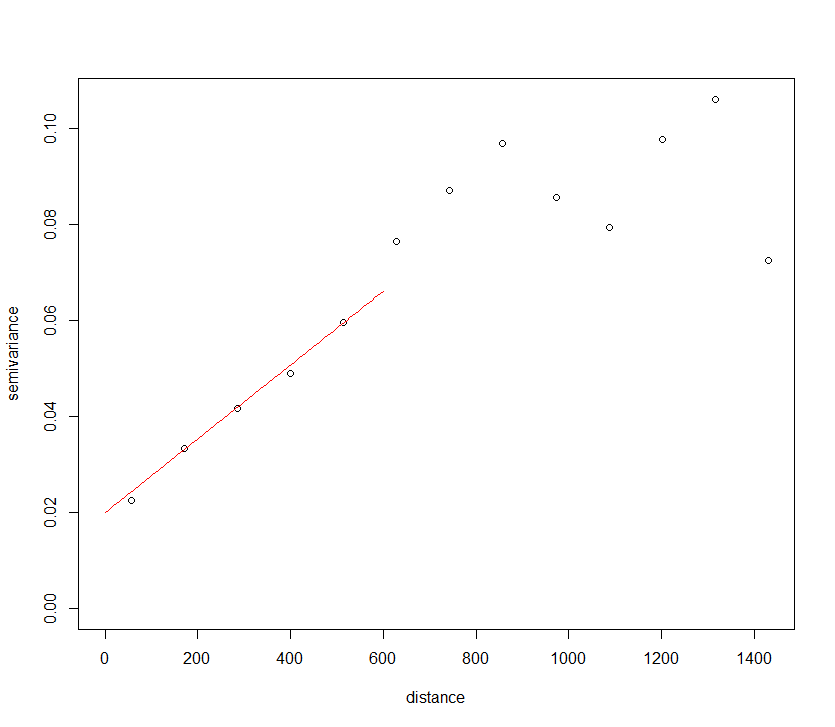
var<-variog(geodata)

plot(var)

>var <- variog(geodata, max.dist=600)

vario.fit<- variofit(var, nugget = 0.02, max.dist = 600, fix.nugget = TRUE)

lines(vario.fit,col="red")



summary(geodata$coords)

X Y

Min. :3222Min. :7968

1st Qu.:3439 1st Qu.:8363

Median :3576Median :8578

Mean :3568 Mean :8622

3rd Qu.:3702 3rd Qu.:8850

Max. :3875Max. :9438

Dle summary zvolíme hodnoty pro vytvoření krigingu, velmi důležité je maximum, minimum a důležité je zvolení b pro tvorbu gridu, do kterého se bude dále vytvářet kriging

>loci <- expand.grid(seq(3200,3900,b=5),seq(7900,9500,b=5))

>kc<- krige.conv(geodata,loc=loci,krige=krige.control(obj.model=vario.fit))

par(mfrow = c(1,2))

vykreslení krigingu

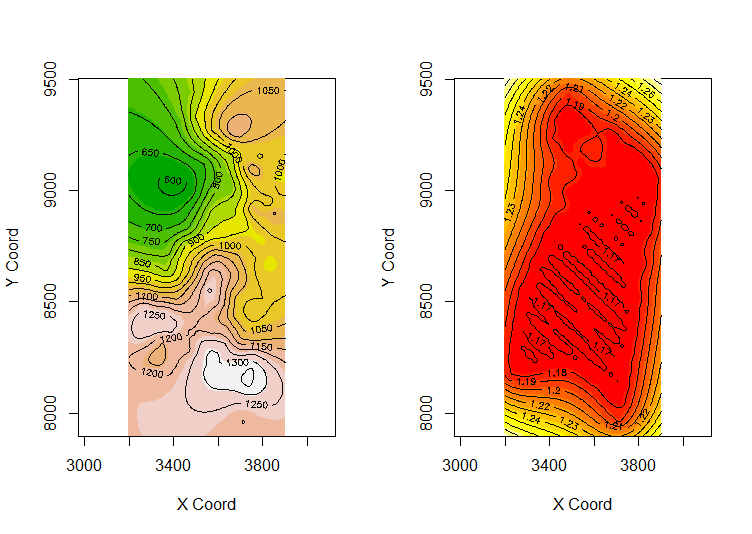
>image(kc, value = exp(kc$predict), col = terrain.colors(12))

>contour(kc,value = exp(kc$predict), nlev = 20, add=T)

Vykreslení chyby(vpravo)

>image(kc, value = exp(sqrt(kc$krige.var)))

>contour(kc,value = exp(sqrt(kc$krige.var)), add=T)



Vlevo kriging, vpravo krigovacíchyba

**Využijte opět datovou sadu data1.txt. Datovou sadu rozdělte do skupin podle příslušnosti ke kvadrantům, které jsou definovány pomocí průměrné hodnoty osy X i osy Y (viz schéma). Zjistěte, zda existuje rozdíl mezi hodnotami sloupce Z v závislosti na vytvořených skupinách. Následně vykreslete graf vícenásobného porovnávání. Vaše výsledky (test i graf) vhodně interpretujte.**

1. **ANOVA**

a <- quantile(Y, probs = c(1:4/4))

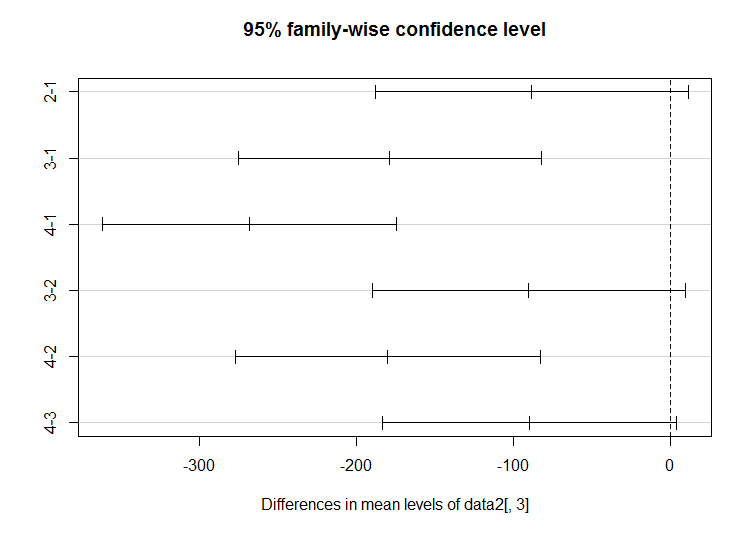
vytvořím 3. Sloupec pro rozdělení

data2[,3] <- ifelse(Y < a[1], 1, ifelse(Y < a[2], 2, ifelse(Y < a[3], 3,4)))

data2[,3] <- as.factor(data2[,3])

anova<- aov(X~data2[,3], data = data2)

>plot(TukeyHSD(anova))



1. **KVADRATICKÁ REGRESE**

View(data2)

Zjištění korelace – vzájemného vztahu

Dle korelace mají hodnoty velký záporný vztah

cor(data2[,1],data2[,2])

[1] -0.8475514

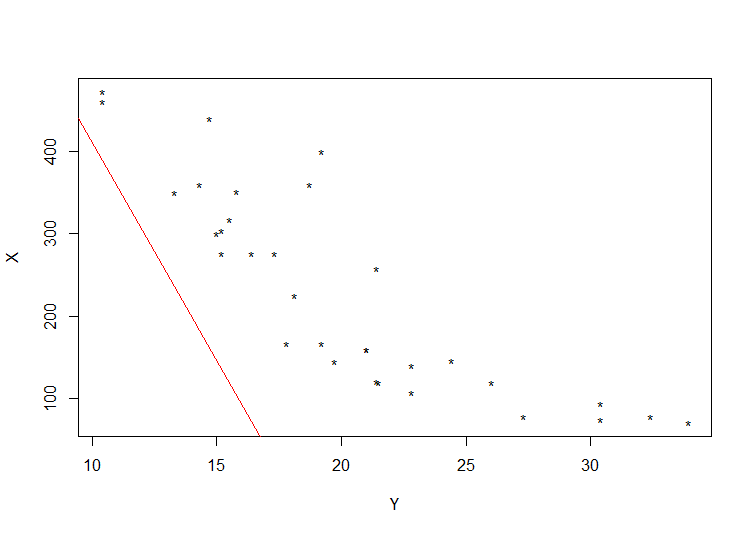
kvm<- lm(X ~ Y + I(Y^2), data = data2)

plot(X ~ Y, pch = '\*')

abline(kvm, col = 'red')

Regresní rovnice

X ~ Y + I(Y^2)



Andrea HOHNOVÁ