1)nastavit set working direktory

setwd("D:/users/Stiskalova/zkouska")

2)nahrát data přes Import Dataset

3)popisné charakteristiky (statistika)

Jedním příkazem- nahrát > library(psych)

Pak > describe(data1)

vars n mean sd median trimmed mad min max range skew kurtosis se

X 1 178 3568.44 161.73 3575.65 3572.49 192.74 3222.05 3874.65 652.60 -0.18 -0.95 12.12

Y 2 178 8622.46 324.66 8577.52 8605.29 358.38 7967.85 9438.00 1470.15 0.41 -0.65 24.33

Z 3 178 1077.35 235.54 1073.53 1080.32 236.38 446.42 1658.12 1211.70 -0.10 -0.39 17.65

a)vypočítat průměr sloupce Z

mean(data1[,3])

[1] 1077.351

b)Minimu a Maximum

min(data1[,3])

[1] 446.42

> max(data1[,3])

[1] 1658.12

c) Rozsah = rozdíl mezi nejvyšší hodnotou a nejnižší

**min <- min(data[,3])**

**> max <- max(data[,3])**

**> max – min**….zavedení promenné

Můžeme rozsah vypočítat rovnou…max(data1[,3]) - min(data1[,3])

d)variace = rozptyl (jak jdou data rozptýlená)

> var(data1[,3])

[1] 55481.26

e)směrodatná odchylka (jak moc se od sebe data liší- je-li směrodatná odchylka malá, jsou si data vzájemně podobná a naopak)

> sd(data1[,3])

[1] 235.5446

f)šikmost (Šikmost zjišťuje, zda jsou hodnoty rozloženy okolo průměru symetricky)

> skewness(data1[,3])

[1] -0.0961719

Problém-zapnutí balíčku e1071

g)špičatost (špičatost pak porovnává koncentraci hodnot blízko průměru a dále od něho)

> kurtosis(data1[,3])

[1] -0.3865235

h)celkové shrnutí

> summary(data1[,3])

Min. 1st Qu. Median Mean 3rd Qu. Max.

446.4 914.1 1074.0 1077.0 1233.0 1658.0

Minimum, 1. Kvartil, median, průměr, 3. Kvartil, maximum

i)IQR (mezikvartilové rozpětí)

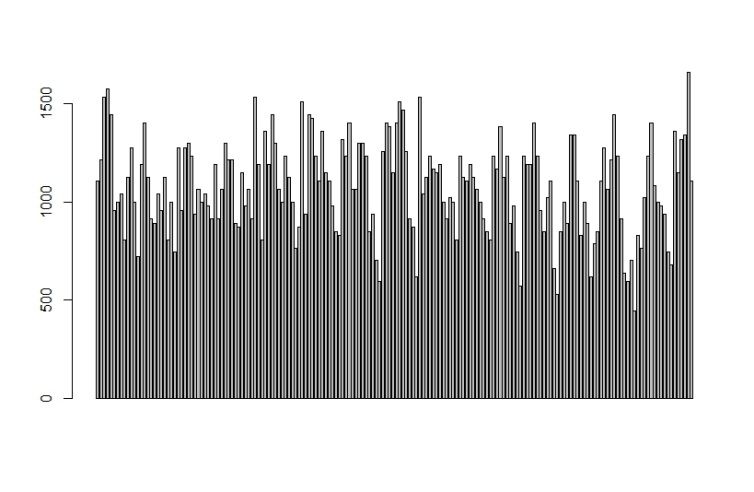
> IQR(data1[,3])

[1] 318.87

4)základní grafy

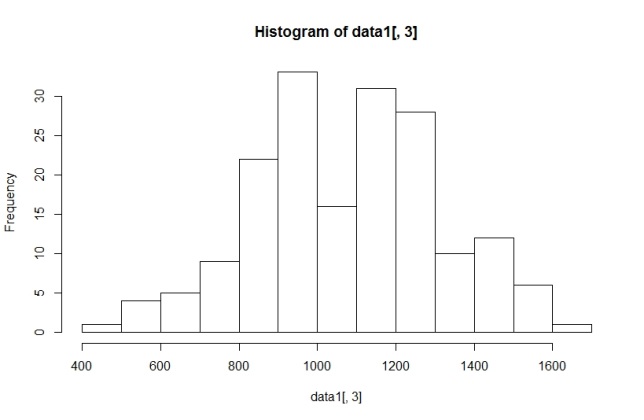
a)barplot (znázorňuje velikost jednotlivých bodů)

> barplot(data1[,3])



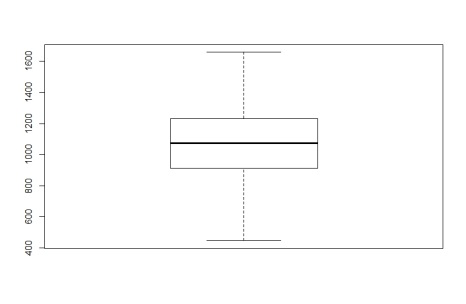
b)histogram

> hist(data1[,3])



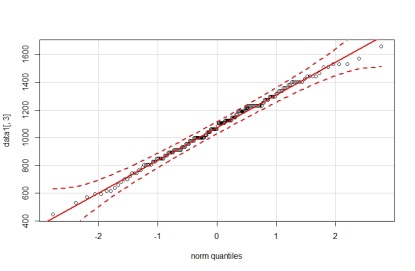
c)boxplot *krabicový graf*- znázorňuje nejnižší hodnotu spodní úsečkou ( Outliery) a nejvyšší hodnotu a graficky vizualizuje numerické data pomocí kvartilů, Prostřední černá úsečka nám vykresluje střední hodnotu (median)

boxplot(data1[,3])



d)qqplot (když nefunguje zkusit zapnout geoR nebo car)

> qqPlot(data1[,3])



5) KRIGING

-zapneme geoR

-nejprve zjistíme, zda mají data normální rozdělení

> shapiro.test(data1[,3])

Shapiro-Wilk normality test

data: data1[, 3]

W = 0.9947, p-value = 0.7757

shapiro test - mensi nez 5 setin -> nemaji normalni rozdeleni

vetsi nez 0,05 -> maji normaln rozdeleni

-když nemají normální rozdělení, musíme hodnoty zlogaritmovat

shapiro.test(log(data1[,3]))

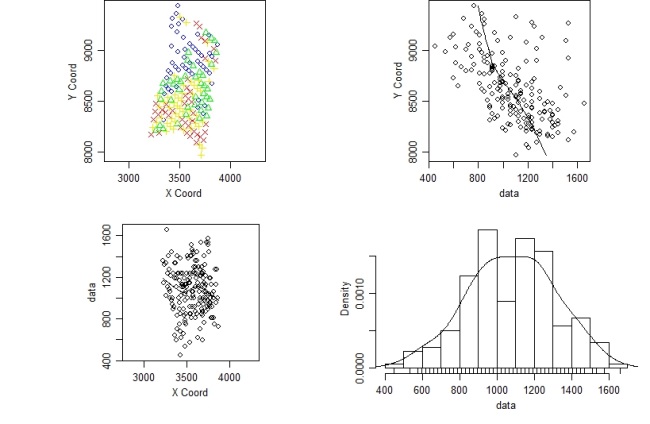
data1[,4] <- log(data1[,3])

-z dat musíme udělat geodata

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

pokud logaritmujeme tak musíme u data.col změnit na ten sloupec, kde máme zlogaritmovaná data

plot(geodata, lowess = T)…

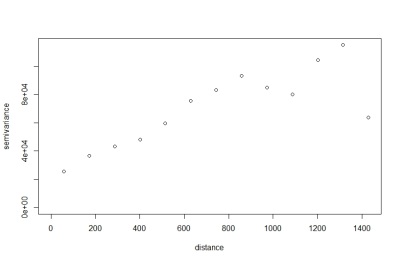


Rozdělení dat

Variogram

var<-variog(geodata)

> plot(var)



vario.fit <- variofit(var, nugget = 0.18, max.dist = 200, fix.nugget = TRUE)

NUGGET= KDE ZAČÍNAJÍ KOLEČKA NA OSE y

MAX.DISTANCE=KDE SE KOLEČKA ZAČÍNAJÍ MĚNIT NA OSE x

lines(vario.fit,col="red")….vykreslení červené čáry

Provedeme základní charakteristiku k vytvoření rastru

> summary(geodata$coords)

X Y

Min. :3222 Min. :7968

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

Median :3576 Median :8578

Mean :3568 Mean :8622

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

Max. :3875 Max. :9438

Podle charakteristik vytvoříme rastr

ČÍM MENŠÍ ROZDÍL TÍM VĚTŠÍ RASTR (ASI)

Min a max X a Y, velikost pixelu 100x100

loci <- expand.grid(seq(3222,3875,b=100),seq(7968,9438,b=100))

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

krige.conv: model with constant mean

krige.conv: Kriging performed using global neighbourhood

> par(mfrow = c(1,2))

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

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

DÁT SI POZOR NA ZLOGARITMOVANÁ A NEZLOGARITMOVANÁ DATA!!! KDYŽ NEJSOU LOGARITMOVANÁ NEMUSÍM DÁVAT EXP image(kc, value =(kc$predict), col = terrain.colors(12))..

REGRESE

data2 <- read.table("D:/Skola/I-STGST/Cviceni/cvXII/data2.txt", header=T, quote="\"")

KUBICKÁ REGRESE

attach(data2)

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

lm <- lm(data2[,1] ~ data2[,2] + I(data2[,2]^2) + I(data2[,2]^3))

summary(lm)

KVADRTICKÁ REGRESE

data2 <- read.table("D:/Skola/I-STGST/Cviceni/cvXII/data2.txt", header=T, quote="\"")

attach(data2)

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

lm <- lm(data2[,1] ~ data2[,2] + I(data2[,2]^2) )

summary(lm)

LINEÁRNÍ REGRESE

data2 <- read.table("D:/Skola/I-STGST/Cviceni/cvXII/data2.txt", header=T, quote="\"")

attach(data2)

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

lm <- lm(data2[,1] ~ data2[,2])

summary(lm)

napsat jenom lm->vypsat čísla Coefficients:

(Intercept) data2[, 2] I(data2[, 2]^2) I(data2[, 2]^3)

5.070e+01 -3.372e-01 1.109e-03 -1.217e-06

REGRESIVNÍ ROVNICE y=ax+b(na druhou) a u kubické regrese přidáme c (na třetí)

KVALITA MODELU-R-squared (ČÍM VÍC SE BLÍŽÍ K 1 TÍM LEPŠÍ MODEL)

ANOVA

ROZDĚLENÍ NA KVARTILY NA OSE Y

data2[,3]<-0

attach(data2)

a <- quantile(data2[,1], probs=c(1:4/4))

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))

ROZDĚLENÍ PODLE MICHALA

data[,5] <- 0

attach(data)

Přiřadíme podle kvadrantu

a <- quantile(x, probs = c(0.5))

b <- quantile(y, probs = c(0.6))

data[,5] <- ifelse(x < a[1], 1, ifelse(y < b[1], 2, 3))

data[,5] <- as.factor(data[,5])

anova <- aov(z~data[,5], data = data)

plot(TukeyHSD(anova))

ROZDĚLENÍ PODLE NÁS CVIČENÍ 3

data2[,3] <- ifelse(X<a[1]&Y<b[1],1,ifelse(X<a[1]&Y>b[1],2,ifelse(X>a[1]&Y<b[1],3,4)))