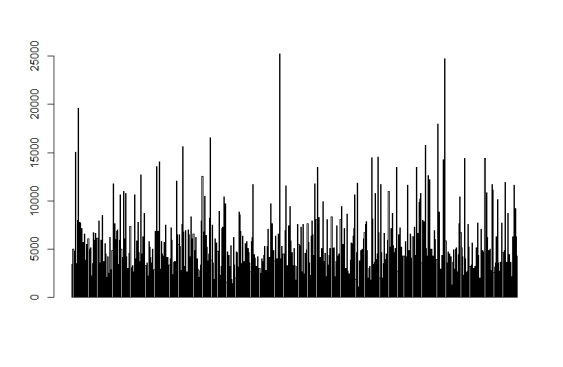
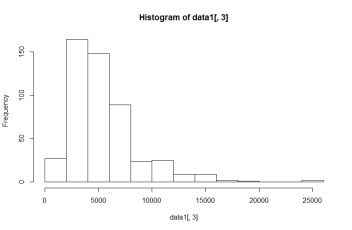
1)ZÁKLADNÍ STATISTIKA

-nastavení setworking directory a nahrání dat

|  |
| --- |
| setwd("D:/users/Stiskalova/zkouska/prakticka")  > data1 <- read.table("D:/users/Stiskalova/zkouska/termin4/data1.txt", header=T, quote="\"")  > View(data1)  a)vypočítat průměr sloupce Z  > mean(data1[,3])  [1] 5466.692  b)minimum a maximum sloupce Z  > min(data1[,3])  [1] 1053.606  > max(data1[,3])  [1] 25210.14  c) Rozsah = rozdíl mezi nejvyšší hodnotou a nejnižší  > max(data1[,3]) - min(data1[,3])  [1] 24156.54  d) variace = rozptyl (jak jdou data rozptýlená)  var(data1[,3])  [1] 10527006  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] 3244.535  f) šikmost (Šikmost zjišťuje, zda jsou hodnoty rozloženy okolo průměru symetricky)  skewness(data1[,3])  [1] 1.97692  g) špičatost (špičatost pak porovnává koncentraci hodnot blízko průměru a dále od něho)  kurtosis(data1[,3])  [1] 6.172273  h)celkové shrnutí  summary(data1[,3])  Min. 1st Qu. Median Mean 3rd Qu. Max.  1054 3342 4647 5467 6685 25210  Minimum, 1. Kvartil, median, průměr, 3. Kvartil, maximum  i)IQR  IQR(data1[,3])  [1] 3342.254  j)základní grafy  barplot (znázorňuje velikost jednotlivých bodů)  barplot(data1[,3]) |
|  |
| |  | | --- | |  | |

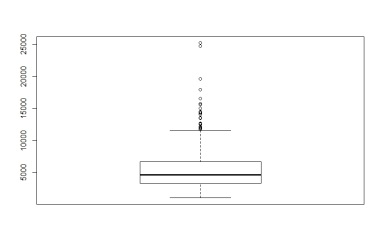


Histogram

hist(data1[,3]) 

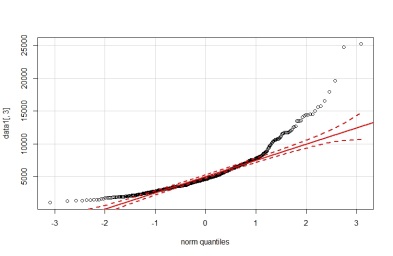
c)boxplot *krabicový graf*- znázorňuje nejnižší hodnotu spodní úsečkou ( Outliery) a nejvyšší hodnotu a graficky vizualizuje numerické data pomocí kvartilů, Prostřední černá úsečka nám vykresluje střední hodnotu (median)

boxplot(data1[,3])



d)qqplot

> qqPlot(data1[,3])



Celkový popis dat

> describe(data1)

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

x 1 500 6033.07 1279.63 6008.50 6018.99 1568.59 3726.00 8613.00 4887.00 0.07 -1.03

y 2 500 22825.46 375.75 22831.50 22818.76 438.11 22119.00 23649.00 1530.00 0.12 -0.92

z 3 500 5466.69 3244.53 4647.32 4970.44 2317.16 1053.61 25210.14 24156.54 1.98 6.17

V4 4 500 1.99 0.78 2.00 1.98 1.48 1.00 3.00 2.00 0.02 -1.38

se

x 57.23

y 16.80

z 145.10

V4 0.04

4)REGRESE

data2 <- read.table("D:/users/Stiskalova/zkouska/termin4/data2.txt", header=T, quote="\"")

> View(data2)

> attach(data2)

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

[1] -0.8475514

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

> summary(lm)

lm

Call:

lm(formula = data2[, 1] ~ data2[, 2] + I(data2[, 2]^2))

Coefficients:

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

35.8286989 -0.1052732 0.0001255

Regresivni rovnice y=ax+b

KVALITA MODELU-R-squared (čím více se blíží k 1 tím lepší model)

Multiple R-squared: 0.7927

2)KRIGING

-interpolační metoda založená na výpočtu vzájemných vztahů mezi body

-zapneme geoR

-nejprve zjistíme, zda mají data normální rozdělení

shapiro.test(data1[,3])

Shapiro-Wilk normality test

data: data1[, 3]

W = 0.8415, p-value < 2.2e-16

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

vetsi nez 0,05 -> maji normaln rozdeleni

-když nemají normální rozdělení, musíme hodnoty zlogaritmovat

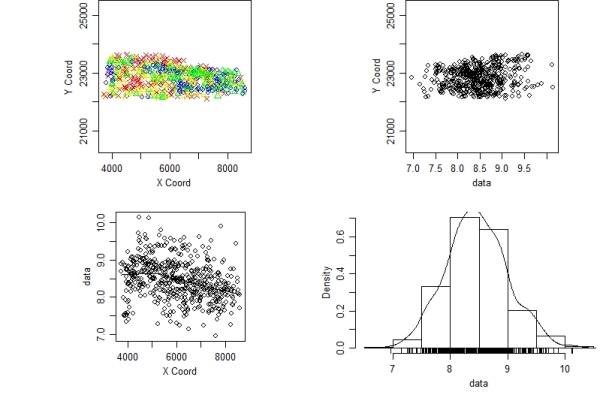
data1[,4] <- log(data1[,3])

> View(data1)

-z dat musíme udělat geodata

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

plot(geodata, lowess = T)

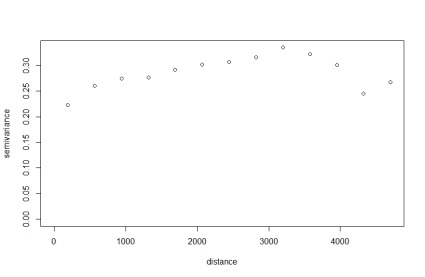


Abychom viděli, jak data vypadají, vytvoříme grafy, které nám zobrazí rozložení dat a jejich soustředění

-variogram-vyjadřuje, jak se mění proměnná mezi místem U a místem U+H, mezi nimiž je vzdálenost H

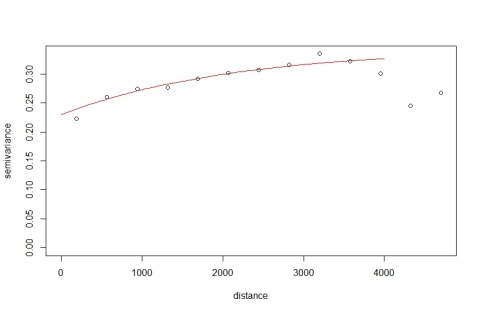
var<-variog(geodata)

> plot(var)



vario.fit <- variofit(var, nugget = 0.23, max.dist = 4000, fix.nugget = TRUE)

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



- Provedeme základní charakteristiku k vytvoření rastru

> summary(geodata$coords)

x y

Min. :3726 Min. :22119

1st Qu.:4978 1st Qu.:22526

Median :6008 Median :22832

Mean :6033 Mean :22825

3rd Qu.:7067 3rd Qu.:23115

Max. :8613 Max. :23649

Podle charakteristiky min a max vytvoříme rastr

loci <- expand.grid(seq(3726,8613,b=100),seq(22119,23649,b=100))

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

par(mfrow = c(1,2))

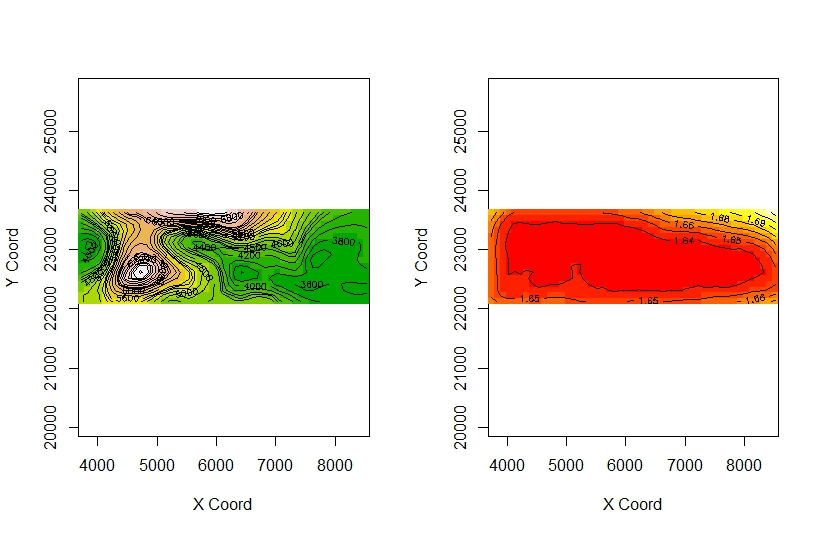
> image(kc, value = exp(kc$predict), col = terrain.colors(12))

contour(kc,value = exp(kc$predict), nlev = 10, add=T)

vykreslení chyby

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

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



-na obrázku vlevo vidíme vykreslený kriging, který je rozdělený na 10 úrovní (fce image+ terénní barvy)

-na obrázku vpravo vidíme vykreslenou krigovací chybu

3)ANOVA

> data1 <- read.table("D:/users/Stiskalova/zkouska/termin4/data1.txt", header=T, quote="\"")

> View(data1)

> data[,4] <- 0

> attach(data1)

-rozdělíme na skupiny podle pomocí sedmého decilu osy X a mediánu osy Y

a <- quantile(x, probs = c(0.7))

b <-median(y)

data1[,4] <-ifelse(x<a[1]&y>b[1],1,ifelse(x<a[1]&y>b[1],2,3))

data1[,4] <- as.factor(data1[,4])

anova<- aov(z~data1[,4], data = data1)

> plot(TukeyHSD(anova))

