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

1. Základní charakteristiky
2. Nejprve jsem nastavil working direktory

**setwd("C:/Users/nerami00.PRFAD/Desktop/termin1")**

1. Nahrál data přes Import Dataset- data.txt
2. Vypočítal průměr sloupce Z

**mean(data[,3])**

[1] 5466.692=průměr

1. Minimum a maximum

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

[1] 1053.606=minimum

**max(data[,3])**

25210.14=maximum

1. Rozsah = rozdíl mezi nejvyšší hodnotou a nejnižší

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

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

**> max - min**

**[1] 24156.54**

1. Variaci=rozptyl

**var(data[,3],)**

**[1] 10527006**

1. Směrodatná odchylka

**sd(data[,3])**

**[1] 3244.535**

1. Šikmost

**skewness(data[,3])**

**[1] 1.97692**

1. Špičatost

**kurtosis(data[,3])**

**[1] 6.172273**

1. Celkové zhodnocení

**summary(data[,3])**

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

**1054 3342 4647 5467 6685 25210**

**Kde je Min= minimum, první kvartil=3342, Střední hodnota=4647, 3 kvartil=6685**

**IQR – 3342.254**

![](data:None;base64,)

Barplot- znazornuje nám velikosti jednotlivých entit(bodů- o velikosti Z, která je promítnutá na Y ose)

barplot(data[,3])

![](data:None;base64,)

Histogram- znázorňuje výskyt dat v intervalech. Můžeme vidět, že největší výskyt je okolo hodnoty 2500 a poté klesá.

hist(z)

![](data:None;base64,)

Boxplot-*krabicový graf*- znázorňuje nejnižší hodnotu spodní úsečkou ( Outliery) a nejvyšší hodnotu a graficky vizualizuje numerické data pomocí kvartilů, mezikvartilové rozpětí je **3700.**

Prostřední černá úsečka nám vykresluje střední hodnotu (median)

boxplot(z)

1. Vhodně interpolujte datovou sadu data.txt (sloupec Z) pomocí **krigingu**. Výsledky spolu s isoliniemi a také krigovací chybu vykreslete a slovně okomentujte. Kriging vykreslete v barevnosti určené pro terén a isolinie rozdělte kvůli přehlednosti pouze do 5 úrovní.

# Kriging

Nejprve zjistíme rozdělení dat a p-value

shapiro.test(data[,3])

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

data: log(data[, 3])

W = 0.9967, p-value = 0.4148

Data musíme upravit a to tak, že ej zlogaritmujeme

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

Uděláme z dat Geodata

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

cpomocí coords.col stanovíme , že sloupec 1 a 2 jsou souřadnice X a Y a data.col=4 –sloupec 4 je Z souřadnice

![](data:None;base64,)

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

> plot(geodata, lowess = T)

Veškerá statistika geodat- density (dole vpravo), Grafy (nahoře vpravo a dole vlevo) ukazují, že data nejsou nijak soustředěná na žádnou světovou stranu

Vytvoříme variogram

var<-variog(geodata)

![](data:None;base64,)

Zde můžeme vidíme několik charakteristik, které nám pomohou zpřesnit Kriging. Nugget efect=0,2 prahová hodnota je v distance=4000

var <- variog(geodata, max.dist=4000)

vario.fit <- variofit(var, cov.model = "spherical", nugget = 0.2, max.dist = 4000, fix.nugget = TRUE)

![](data:None;base64,)

Provedeme základní charakteristiku k vytvoření rastru

summary(geodata$coords)

Min. :3727 Min. :73731

1st Qu.:4979 1st Qu.:75084

Median :6008 Median :76105

Mean :6034 Mean :76085

3rd Qu.:7067 3rd Qu.:77054

Max. :8614 Max. :78834

Podle charakteristik vytvříme rastr

Min a max X a Y, velikost pixelu 100x100

loci <- expand.grid(seq(3800,7100,b=100),seq(73800,78900,b=100))

par(mfrow = c(1,2))

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

krige.conv: model with constant mean

krige.conv: Kriging performed using global neighbourhood

Data musíme odlogaritmovat a zobrazit je zprávně

par(mfrow = c(1,2))

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

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

>

![](data:None;base64,)

Výsledek interpolace s vrstevnicemi

![](data:None;base64,)

## REGRESE

1. Pro data2.txt zjistěte nejdříve korelaci a následně zjistěte regresní vztah s využitím kvadratické regrese. 1. sloupec je závislá proměnná, druhý sloupec je nezávislou proměnnou. Výsledek regrese zapište ve formě **regresní rovnice** a interpretujte ho. Určete z něj **kvalitu** modelu.

Nahrajeme data2.txt

Zjistíme složení (attach(data2))

View(data2)

attach(data2)

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

kvm <- lm(Var1 ~ Var2 + I(Var2^2), data = data2)

kvm

summary(kvm)

plot(kvm)

plot(Var1 ~ Var2, pch = '\*')

abline(kvm, col = 'red')

Var1 ~ Var2 + I(Var2^2) – regresní rovnice

Coefficients:

(Intercept) Var2 I(Var2^2)

-2.177e+00 -5.845e-01 -2.616e-05

Pi-value nám udavává že model je pravdivý <0.05. Z grafu lze vyčíst, že je zde málo extrémů, regresivní přímka nám ukazuje na vztah mezi proměnnými

![D:\users\vesely\zk\termin1\Rplot06.png](data:None;base64,)

## ANOVA

Vytvoříme nový sloupec a vložíme do něj číslo ve kterém se nachází

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])

table(data[,5])

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

anova

plot(TukeyHSD(anova))

![](data:None;base64,)

Anova nám dokázala, že je rozdíl mezi hodnotami sloupce Z v závisloti na skupinách.

Pomocí grafu Tukey HSD vykreslíme rozdíly mezi průměry v hodnotách 1-3 ve skupinách