wtp - deskriptiv statistik

Emil Brohus & Nicolai Haslund

# Deskriptiv statistik af positiv betalingsvillighed

I det følgende vil de statistiske beregninger der er blevet foretaget på baggrund af vores undersøgelse blive dokumenteret.

Scriptet er konstrueret således, at data materiale bliver hentet direkte fra github, så alle der er i besidelse af scriptet kan få adgang til vores data, og tjekke undersøgelses validitet.

## Data indlæses

wtp = read.table(url('https://raw.githubusercontent.com/EmBro/Bachelor/master/wtp.csv',  
 method = 'libcurl'), dec = ',')  
  
wtp = as.matrix(wtp)  
  
colname = 'wtp'  
  
colnames(wtp) = colname

Data har også fået anvist overskriften 'wtp'.

## Shapiro-test

Der anvendes en shapiro test for at tjekke sandsynligheden for at projektets respondenter repræsenterer en normalfordelt population.

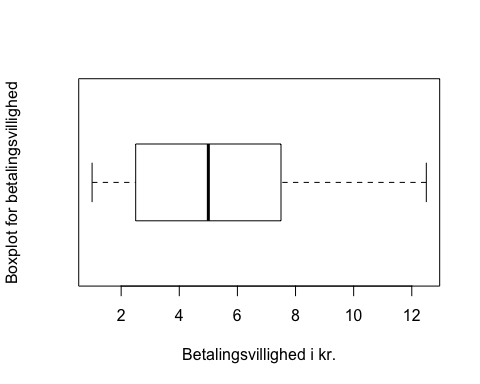
shapiro.test(wtp)

##   
## Shapiro-Wilk normality test  
##   
## data: wtp  
## W = 0.89803, p-value = 1.269e-06

## Boxplot

Der produceres et boxplot for at tjekke data for outliers, og for at få en fornemmelse af hvordan data er distribueret.

boxplot(wtp, ylab = 'Boxplot for betalingsvillighed',  
 xlab = 'Betalingsvillighed i kr.', horizontal = T)



Som det fremgår af boxplottet, findes der en skævhed i datasættet. Skævheden befinder sig til venstre, hvilket fortæller os, at der er flere respondenter med betalingsvillighed til venstre for datasættets median, og et større antal ekstremiteter til højre for medianen.

## Basis statistik

Basis udregninger er udført til videre behandling.

m = mean(wtp)  
md = median(wtp)  
sd = sd(wtp)  
n = length(wtp)  
  
error = qt(0.975,df=n-1)\*sd/sqrt(n)  
lower = m - error  
upper = m + error  
lower

## [1] 4.848189

upper

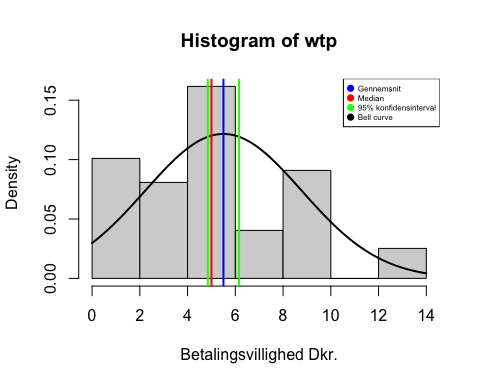
## [1] 6.156861

Derudover konstrueres et 95% konfidensinterval.

## Histogram

Et histogram bliver produceret for at anskueliggøre og visualisere distribution, og statistik af data.

hist(wtp, freq = F, col = 'light grey',  
 xlab = 'Betalingsvillighed Dkr.')  
curve(dnorm(x, m, sd), add=TRUE, col = 'black', lwd = 2)  
abline(v = m, col = 'blue', lwd = 2, lheight = 0.1)  
abline(v = md, col = 'red', lwd = 2)  
abline(v = lower, col = 'green', lwd = 2)  
abline(v = upper, col = 'green', lwd = 2,)  
legend('topright', c('Gennemsnit','Median','95% konfidensinterval', 'Bell curve')  
 ,pch = c(16,16, 16, 16), col = c('blue','red', 'green', 'black'),  
 cex=0.5, pt.cex = 1)



summary(wtp)

## wtp   
## Min. : 1.000   
## 1st Qu.: 2.500   
## Median : 5.000   
## Mean : 5.503   
## 3rd Qu.: 7.500   
## Max. :12.500

Som det kan ses af histogrammet, befinder de fleste respondenters betalingsvillighed omkring 5 kr. Denne betalingsvillighed skal dog læses med meget stort forbehold, da der ikke er taget højde for alle respondenter uden betalingsvillighed.

# Undersøgelsens samlede betalingsvillighed

Data for undersøgelses samlede wtp loades ind, ligeledes direkte fra github.

wtp\_samlet = read.table(url('https://raw.githubusercontent.com/EmBro/Bachelor/master/samlet\_wtp.csv', method = 'libcurl'), dec = ',')  
  
wtp\_samlet = as.matrix(wtp\_samlet)  
  
colname = 'wtp\_samlet'  
  
colnames(wtp\_samlet) = colname  
  
summary(wtp\_samlet)

## wtp\_samlet   
## Min. : 0.00   
## 1st Qu.: 0.00   
## Median : 0.00   
## Mean : 2.51   
## 3rd Qu.: 5.00   
## Max. :12.50

Som det ses, bærer den samlede betalingsvillighed kraftigt præg af de mange respondenter der ingen betalingsvillighed har. Den samlede gennemsnitlige betalingsvillighed beregnes til 2,51 kr. hvilket også er den værdi der er anvendt i projektet til estimering af øget billetindkomst.

# Betalingsvillighed og køn

Følgende chunk loader det samlede datagrundlag der er anvendt i projektet ind, ligeledes direkte fra github. Derudover navgives kollonene der repræsenterer respondenters køn, og der konstruerers et boxplot. Boxplottet er anvendt i opgaven under afsnittet omkring betalingsvillighed og køn.

data = read.table(url('https://raw.githubusercontent.com/EmBro/Bachelor/master/RawData.csv',  
 method = 'libcurl'),sep = ';',  
 dec = ',',)  
library(ggplot2)

for (i in c(1:8,10:20)){  
 data[,i] = factor(data[,i])  
}  
  
names(data)[9] = 'wtp'  
names(data)[10] = 'køn'  
  
levels(data$køn) = c('F','M','andet')  
summary(data$køn)

## F M andet NA's   
## 93 120 4 7

data$betale = factor(NA,levels=c('ja','nej'))  
data$betale[data$wtp==0] = 'nej'  
data$betale[data$wtp>0] = 'ja'  
summary(data$betale)

## ja nej NA's   
## 105 118 1

give.n <- function(x){  
 return(c(y = min(x)-5, label = length(x)))  
 # the -5 adjust the position  
}  
ggplot(subset(data, køn %in% c('F','M') & betale=='ja'),  
 aes (y= wtp, x= køn, fill =køn),  
 geom='boxplot',color=køn) + geom\_boxplot()+theme\_bw() +   
 stat\_summary(fun.data = give.n, geom = "text")

