

gl04_4_hoermann

Paul Hörmann

11/26/2019

Aufgabe 4

a)

```
library(foreign)
dmobile = read.dta(file = "./mobil-1.dta")
head(dmobile)

##      dauer    kosten geschlecht tarif    alter
## 1 332.6756 141.89270      Frau  Wenig 18.66894
## 2 347.0924 150.21748      Mann  Wenig 16.15220
## 3 133.3460  73.30383      Mann  Wenig 24.74806
## 4 162.0536 114.47440      Frau   Viel 23.91885
## 5 162.6686  89.74941      Frau   Viel 25.40752
## 6 474.6637 155.90486      Frau Normal 18.38577

dmobile[which(dmobile$dauer == max(dmobile$dauer)),]
```

```
##      dauer    kosten geschlecht tarif    alter
## 194 499.8094 104.6147      Frau   Viel 23.87184
```

Durchschnittliche Telefonat dauer:

```
mean(dmobile$dauer)
```

```
## [1] 276.5569
```

Frauen telefonieren im Schnitt:

```
mean(dmobile[which(dmobile$geschlecht == "Frau"),]$dauer)
```

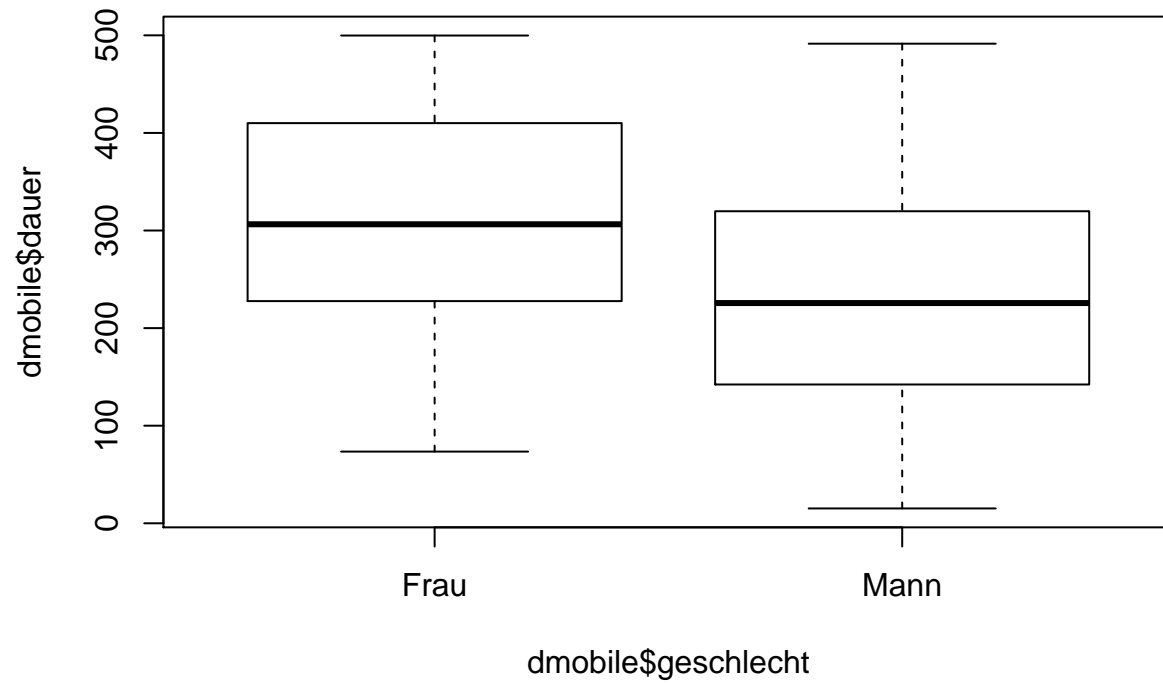
```
## [1] 313.5014
```

Männer telefonieren im Schnitt:

```
mean(dmobile[which(dmobile$geschlecht == "Mann"),]$dauer)
```

```
## [1] 233.41
```

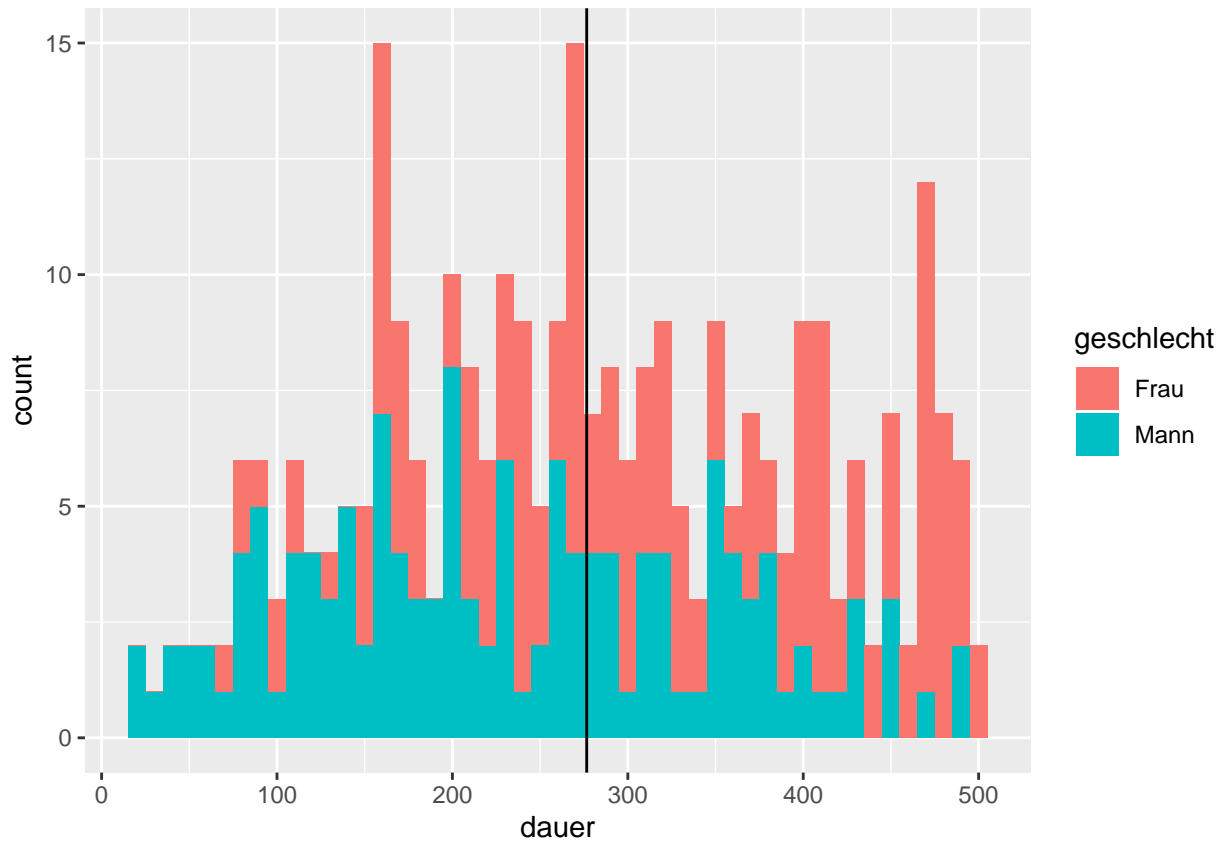
```
boxplot(dmobile$dauer~dmobile$geschlecht)
```



H0: Telefondauer unabhängig vom Geschlecht

H1: Frauen telefonieren länger

```
library(ggplot2)
ggplot(dmobile, aes(x=dauer, fill=geschlecht), col = "red") +
  geom_histogram(binwidth = 10) +
  geom_vline(data = dmobile, aes(xintercept = mean(dauer)))
```



```
wilcox.test(dmobile$dauer~dmobile$geschlecht, alternative = "greater")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: dmobile$dauer by dmobile$geschlecht
## W = 15014, p-value = 1.964e-08
## alternative hypothesis: true location shift is greater than 0
```

Das heißt Frauen telefonieren im Schnitt länger. ### c)

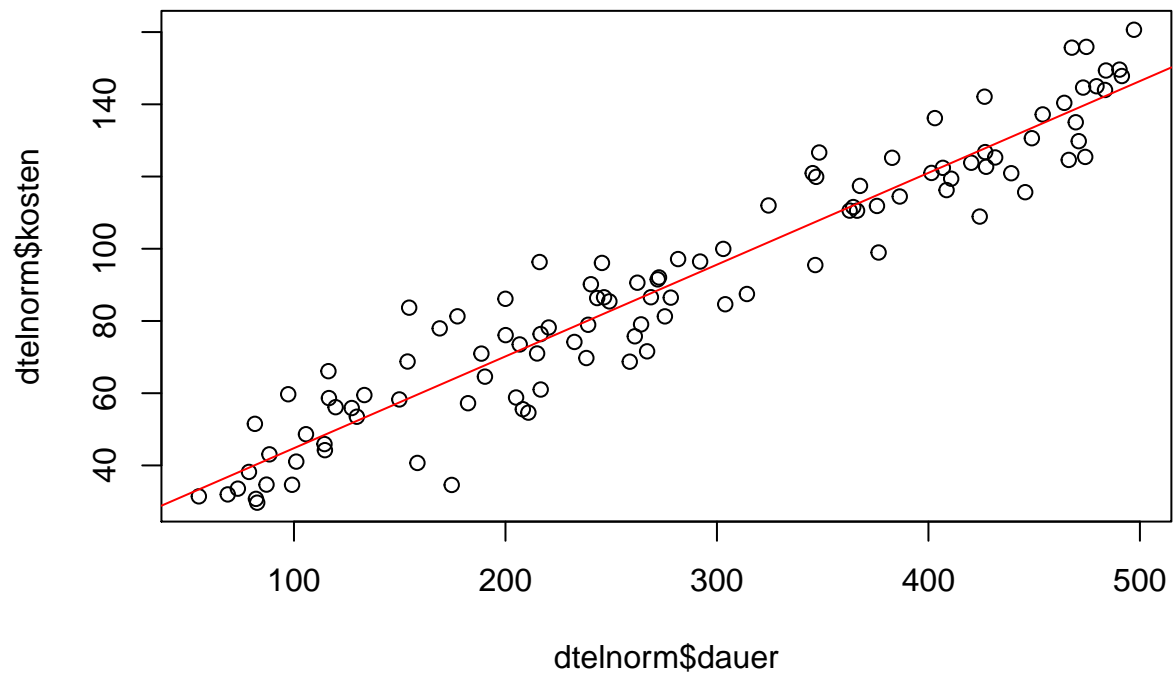
```
dtelnorm = dmobile[dmobile$tarif == "Normal",]
dtelviel = dmobile[dmobile$tarif == "Viel",]
dtelwenig = dmobile[dmobile$tarif == "Wenig",]
ltelnorm = lm(dtelnorm$kosten~dtelnorm$dauer)
ltelviel = lm(dtelviel$kosten~dtelviel$dauer)
ltelwenig = lm(dtelwenig$kosten~dtelwenig$dauer)
```

```

ltnorm

##
## Call:
## lm(formula = dtelnorm$kosten ~ dtelnorm$dauer)
##
## Coefficients:
## (Intercept) dtelnorm$dauer
##      19.3661      0.2541
plot(dtelnorm$dauer, dtelnorm$kosten)
abline(ltnorm, col = "red")

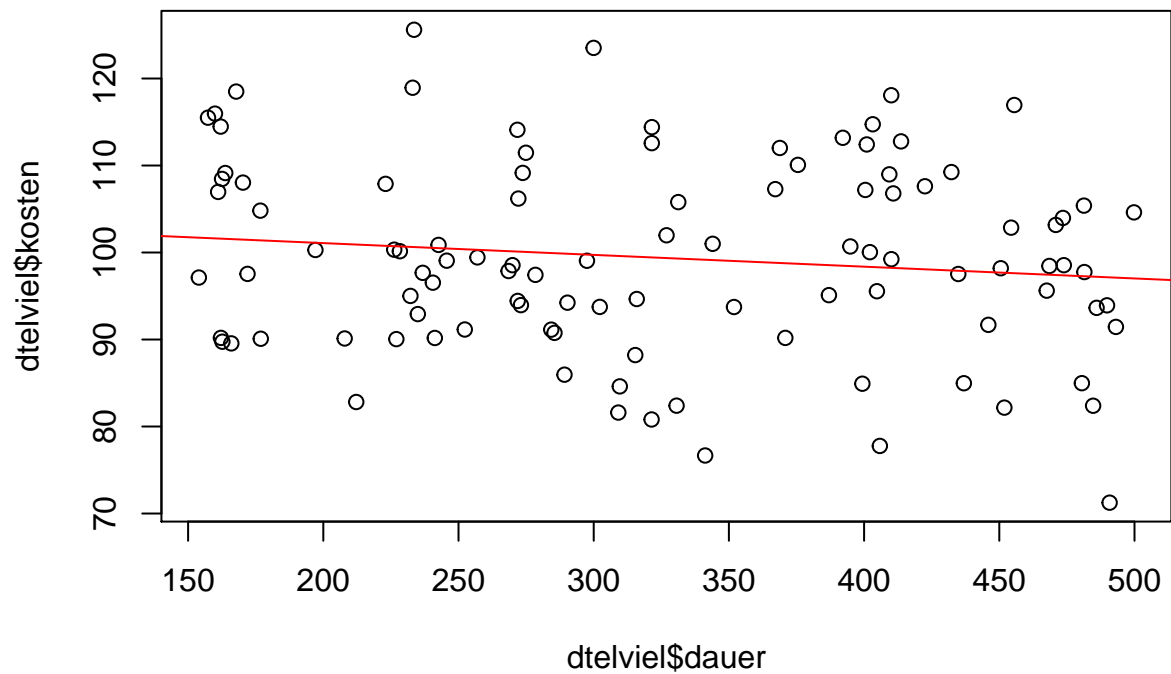
```



Aus der Grafik lässt sich erkennen dass die Regressionsgerade eine gute Näherung ist.

```
ltelviel

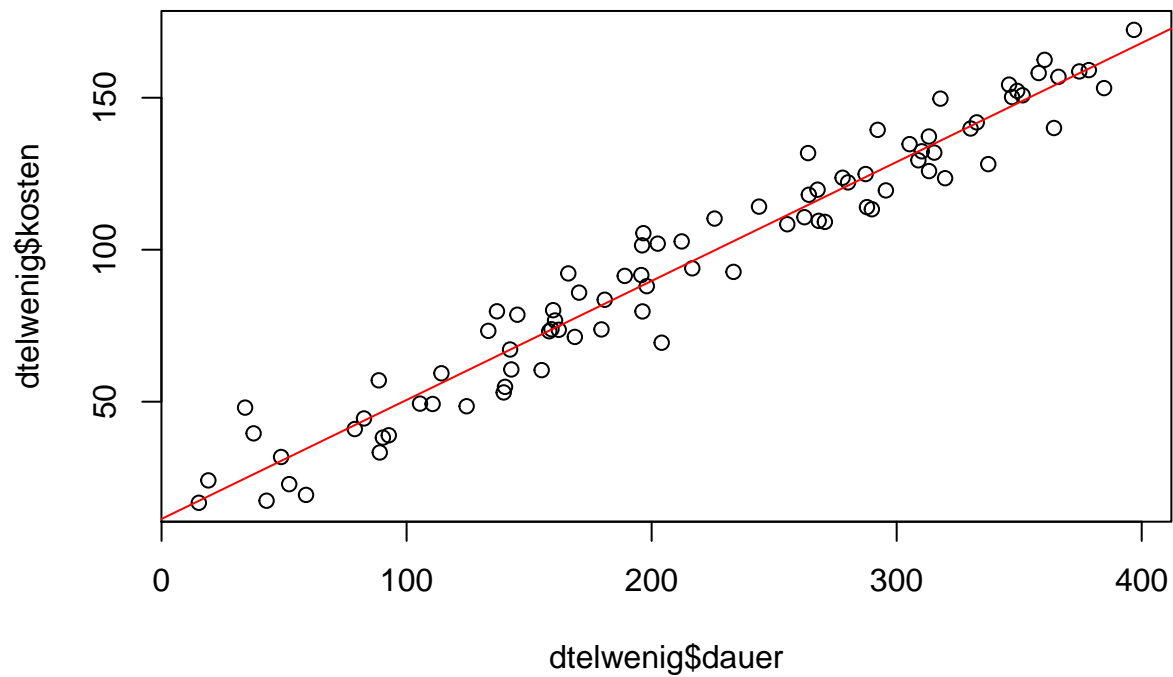
##
## Call:
## lm(formula = dtelviel$kosten ~ dtelviel$dauer)
##
## Coefficients:
## (Intercept) dtelviel$dauer
## 103.78868 -0.01354
plot(dtelviel$dauer, dtelviel$kosten)
abline(ltelviel, col = "red")
```



Wie zu erkennen und im Modell zu sehen sind die Punkte sehr verstreut, also wird vermutlich keine lineare Funktion hinter den Kosten stecken. Wohl eher ein personalisiertes Angebot.

```
ltelwenig
```

```
##  
## Call:  
## lm(formula = dtelwenig$kosten ~ dtelwenig$dauer)  
##  
## Coefficients:  
##      (Intercept)  dtelwenig$dauer  
##          11.4255         0.3915  
plot(dtelwenig$dauer, dtelwenig$kosten)  
abline(ltelwenig, col = "red")
```

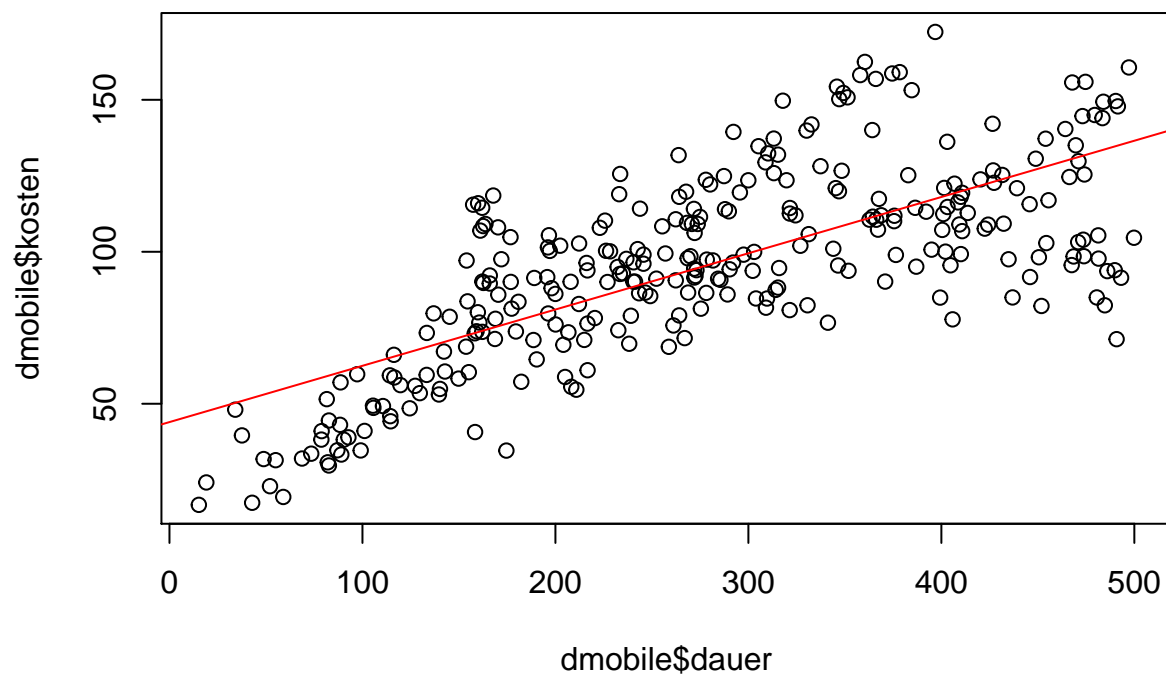


Die Regressionsgerade scheint auch hier eine gute Näherung zu sein.

```
attach(dmobile)
ltelmultivar = lm(kosten~dauer)
detach(dmobile)
ltelmultivar
```

```
##
## Call:
## lm(formula = kosten ~ dauer)
##
## Coefficients:
## (Intercept)      dauer
##    43.9676      0.1851
```

```
plot(dmobile$dauer, dmobile$kosten)
abline(ltelmultivar, col = "red")
```



Die Regressionsgerade scheint auch hier eine gute Näherung zu sein.