

Rekensessie 3

d16030

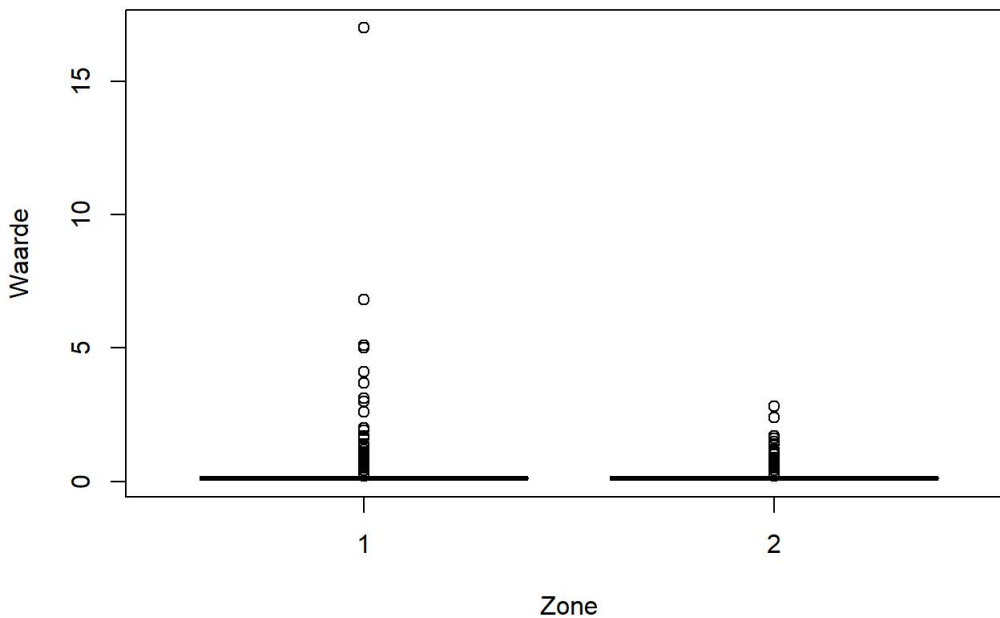
7-8-2019

Check voor verschil tussen Zone 1 en 2; mits er geen stat. sig. verschil is worden zones samengevoegd
"Independent T-test"

Aannames bij het uitvoeren van deze test:

1. Afhankelijke variabele op ratio-schaal en onafhankelijke op nominale schaal:
Waardes zijn naar ratio en zones zijn op nominaal niveau
2. De onafhankelijke variabele heeft twee groepen
Zone 1 en Zone 2
3. Onafhankelijkheid van de observaties
Er zijn geen meetpunten die in beide zones zitten
4. Geen outliers
1 outlier verwijderd

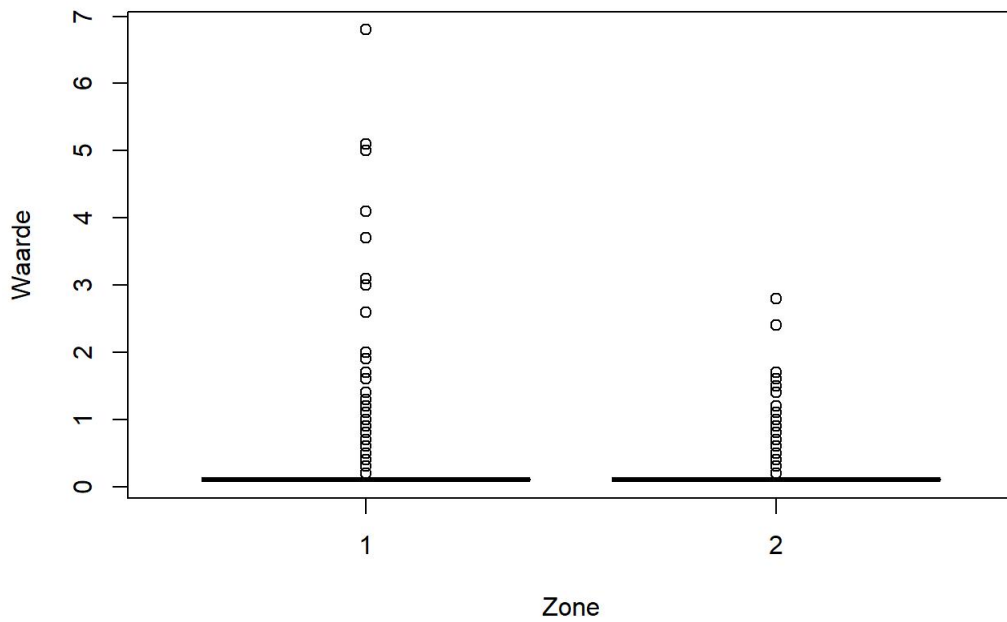
```
g_stat <- g[!(g$Afko == "OS"),] # OS eruit voor outlier analyse, is in procenten en dus niet vergelijkbaar
boxplot(Waarde~Zone, data = g_stat)
```



```
max(g_stat$Waarde)
```

```
## [1] 17
```

```
g <- g[!(g$Waarde == 17),] # Dit meetpunt mocht uit de dataset na overleg met I. Lanting
g_stat <- g[!(g$Afko == "OS"),]
boxplot(Waarde~Zone, data = g_stat) # Verdeling ziet er nu beter uit
```

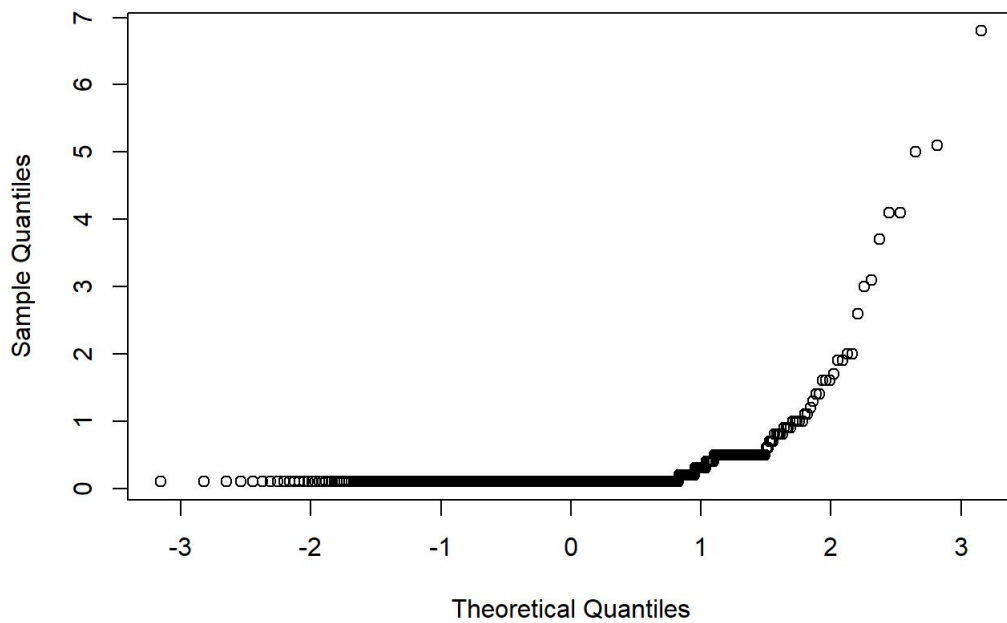


5. De afhankelijke variabele moet een Gaussische curve volgen

```
z1 <- g_stat[(g_stat$Zone == "1"),]
z2 <- g_stat[!(g_stat$Zone == "1"),]

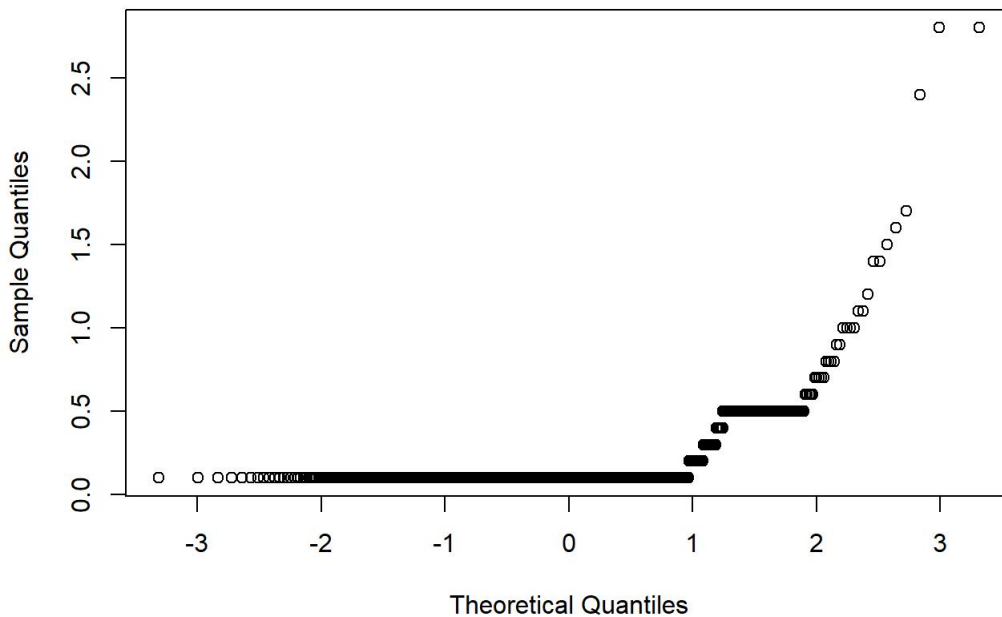
qqnorm(z1$Waarde)
```

Normal Q-Q Plot



```
qqnorm(z2$Waarde)
```

Normal Q-Q Plot



Een rechte diagonale lijn impliceert een normale verdeling, maar dit is hier niet het geval.

6. Homogeniteit in variantie

```
leveneTest(Waarde ~ Zone, data = g_stat)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value    Pr(>F)
## group  1  18.163 2.139e-05 ***
##      1709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De nul-hypothese is dat de waarden van zone 1 en zone 2 dezelfde variantie hebben. De test geeft een stat. sig. resultaat aan. D.w.z. dat de zones niet dezelfde variantie aan meetwaarden hebben.

Dit is echter geen probleem, we kunnen hier een Welsh aanpassing doen om die aanname toch te mogen maken.

De daadwerkelijke onafhankelijke T-test

```
# deze test voert automatisch een Welsh aanpassing toe (d.m.v. een aanpassing in vrijheidsgraden)
t.test(z1$Waarde, z2$Waarde)
```

```
##
## Welch Two Sample t-test
##
## data:  z1$Waarde and z2$Waarde
## t = 3.4891, df = 736.62, p-value = 0.0005135
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03587774 0.12819828
## sample estimates:
## mean of x mean of y
##  0.252480  0.170442
```

Conclusie: De T-test wijst uit dat de gemeten waarden uit Zone 1 en Zone 2 statistisch significant van elkaar verschillen.