Case Study 4 - Group 4

Annika Janson h11829506

Jan Beck h11814291

Franz Uchatzi h1451890

29.11.2020

# 2 Model

### 2.1 Model estimation

#### 2.1.1

##   
## Call:  
## lm(formula = rating ~ rq + vo + wa + kr + education + gender +   
## income + age + price, data = marketing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.167 -4.118 0.827 3.931 15.232   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20.848037 0.477726 43.640 < 2e-16 \*\*\*  
## rq 3.884194 0.312412 12.433 < 2e-16 \*\*\*  
## vo 3.557121 0.312412 11.386 < 2e-16 \*\*\*  
## wa 0.596244 0.312412 1.909 0.05641 .   
## kr -0.287950 0.312412 -0.922 0.35675   
## education -0.256875 0.218121 -1.178 0.23902   
## gender -0.106798 0.199892 -0.534 0.59319   
## income -0.641044 0.204691 -3.132 0.00175 \*\*   
## age 0.012078 0.006017 2.007 0.04483 \*   
## price -0.302541 0.008232 -36.750 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.584 on 3185 degrees of freedom  
## Multiple R-squared: 0.348, Adjusted R-squared: 0.3462   
## F-statistic: 188.9 on 9 and 3185 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = rating ~ 0 + rq + vo + wa + kr + ju + education +   
## gender + income + age + price, data = marketing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.167 -4.118 0.827 3.931 15.232   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## rq 24.732231 0.477726 51.771 < 2e-16 \*\*\*  
## vo 24.405157 0.477726 51.086 < 2e-16 \*\*\*  
## wa 21.444281 0.477726 44.888 < 2e-16 \*\*\*  
## kr 20.560087 0.477726 43.037 < 2e-16 \*\*\*  
## ju 20.848037 0.477726 43.640 < 2e-16 \*\*\*  
## education -0.256875 0.218121 -1.178 0.23902   
## gender -0.106798 0.199892 -0.534 0.59319   
## income -0.641044 0.204691 -3.132 0.00175 \*\*   
## age 0.012078 0.006017 2.007 0.04483 \*   
## price -0.302541 0.008232 -36.750 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.584 on 3185 degrees of freedom  
## Multiple R-squared: 0.8284, Adjusted R-squared: 0.8279   
## F-statistic: 1538 on 10 and 3185 DF, p-value: < 2.2e-16

Compare the OLS estimates for all explanatory variables except for the brands??

### 2.2

Model 1: the estimate for kr = -0.287950, which means that the rating is changing by -0.2887950 for a given price. In other words we could say, we move the regression line by -0.2887950.

Model 2: the estimate for kr = 20.560087, this is the intercerpt for kr.So if kr is chosen and all other varialbes would be 0, the rating would be 20.560087.

### 2.3

If we substrate the intercerpt of “kr” from the intercerpt of ju, then we get the estimate of “kr” in Model 1. This is because “ju” was our reference group, so the intercerpt of Model 1 is equivalent to the intercerpt of “ju”, which is also shown in Model 2. Model 1 shows us the difference between choosing “kr” or “ju” and Model two shows us each groups intercerpt.

### 2.4

Welchen test?

H0: There is no significant difference on rating by choosing either “ju” or “wa” H1: There is a significant difference on rating by choosing either “ju” or “wa” In model 1, which shows the effect of change of “wa” in comparison to the effect of change in rating of “ju”. We can observe that the p-value from “wa”, in Model is 0.05641, and for that slightly higher than 0.05. At a significance level a= 0.05 the effect of change on rating by choosing “wa” is not significant. Therefore we keep H0.

Bonus question:

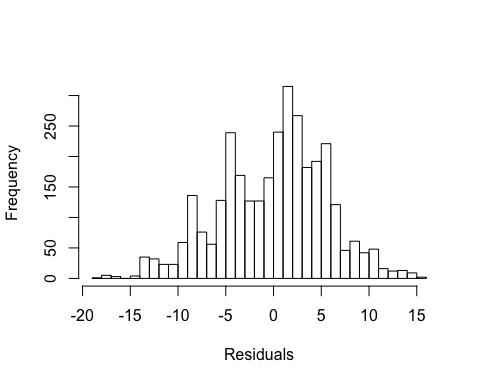
## Linear hypothesis test  
##   
## Hypothesis:  
## wa - ju = 0  
##   
## Model 1: restricted model  
## Model 2: rating ~ 0 + rq + vo + wa + kr + ju + education + gender + income +   
## age + price  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 3186 99433   
## 2 3185 99320 1 113.58 3.6425 0.05641 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The linear hypothesis shows that the p-value again is 0.0561, which is exactly the p-value we expected, as it was the one we could see in the results of “wa” in Model 1. ### 2.5 #### 2.5.1

#### 2.5.2

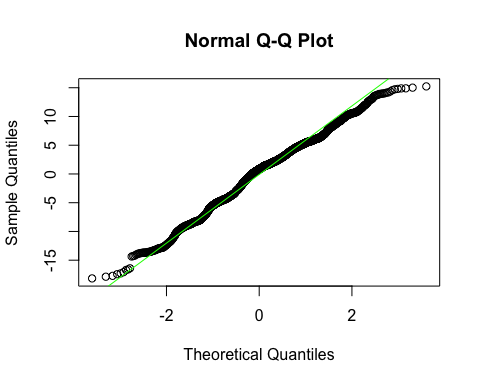
### 2.6

## [1] "Numeric: lengths (3195, 31950) differ"



##   
## Call:  
## lm(formula = rating ~ rq + vo + wa + kr + education + gender +   
## income + age + price, data = marketing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.167 -4.118 0.827 3.931 15.232   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20.848037 0.477726 43.640 < 2e-16 \*\*\*  
## rq 3.884194 0.312412 12.433 < 2e-16 \*\*\*  
## vo 3.557121 0.312412 11.386 < 2e-16 \*\*\*  
## wa 0.596244 0.312412 1.909 0.05641 .   
## kr -0.287950 0.312412 -0.922 0.35675   
## education -0.256875 0.218121 -1.178 0.23902   
## gender -0.106798 0.199892 -0.534 0.59319   
## income -0.641044 0.204691 -3.132 0.00175 \*\*   
## age 0.012078 0.006017 2.007 0.04483 \*   
## price -0.302541 0.008232 -36.750 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.584 on 3185 degrees of freedom  
## Multiple R-squared: 0.348, Adjusted R-squared: 0.3462   
## F-statistic: 188.9 on 9 and 3185 DF, p-value: < 2.2e-16

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo



##   
## Jarque Bera Test  
##   
## data: resids  
## X-squared = 36.524, df = 2, p-value = 1.172e-08

Histogramm: Looking at the Histogramm, it seems that the residuals are not normally distributet, as they are located around