Case Study 4 - Group 4

Annika Janson h11829506
Jan Beck h11814291
Franz Uchatzi h1451890

13.12.2020

2 Model

2.1 Model estimation

2.1.1 and 2.1.2

(See page 2 for model comparison and regression output.)

The R^2 value of model 1 and 2 is **0.348** and **0.828** respectively.

The estimates and standard errors for the non-brand explanatory variables of model 1 and 2 are identical.

The estimates for `rq`, `vo`, `wa`, `ju`/intercept, `education`, `income`, `age` and `price` are significant at the 5%-level.

2.2

Model 1: the estimate for `kr` is **-0.287950**, which means that on average the rating is changing by **-0.2887950** c.p. In other words, we shift the regression line down by 0.2887950.

Model 2: the estimate for `kr` is **20.560087**, this is the intercept for `kr`. On average, if the brand kr and all other variables were 0, the rating would be **20.560087** c.p.

2.3

We can calculate the regression parameter associated with `kr` in Model 1 by subtracting the value of `ju` in Model 2 from the value of `kr` in Model 2.

This is because `ju` was our reference group, so the intercept of Model 1 is equivalent to the intercept of `ju`, which is also shown in Model 2. Model 1 shows us the difference between choosing "kr" or any other group and Model 2 shows us each groups intercept.

Table 1: Model comparison

	Dependent variable:	
	rating	
	(1)	(2)
rq	3.884***	24.732***
	(0.312)	(0.478)
vo	3.557***	24.405***
	(0.312)	(0.478)
wa	0.596*	21.444***
	(0.312)	(0.478)
kr	−0.288	20.560***
	(0.312)	(0.478)
ju		20.848***
		(0.478)
education	−0.257	−0.257
	(0.218)	(0.218)
gender	−0.107	−0.107
	(0.200)	(0.200)
income	−0.641***	−0.641***
	(0.205)	(0.205)
age	0.012**	0.012**
	(0.006)	(0.006)
price	−0.303***	−0.303***
	(0.008)	(0.008)
Constant	20.848***	
	(0.478)	
Observations	3,195	3,195
R^2	0.348	0.828
Adjusted R^2	0.346	0.828
Residual Std. Error (df = 3185)	5.584	5.584
F Statistic	188.881*** (df = 9; 3185)	1,537.900*** (df = 10; 3185)
Note:		*p<0.1; **p<0.05; ***p<0.01

2.4

H0: $\beta_{wa} = 0$
H1: $\beta_{wa} \neq 0$

In model 1, the p-value for β_{wa} is **0.05641**. Therefore, at the $\alpha = 0.05$, we can not reject the null hypothesis. We conclude, that there is no difference in the average rating between the brands `ju` and `wa` c.p.

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 **0.05641**, which is exactly the p-value we expected, as it was the one we could see in the results of `wa` in Model 1.

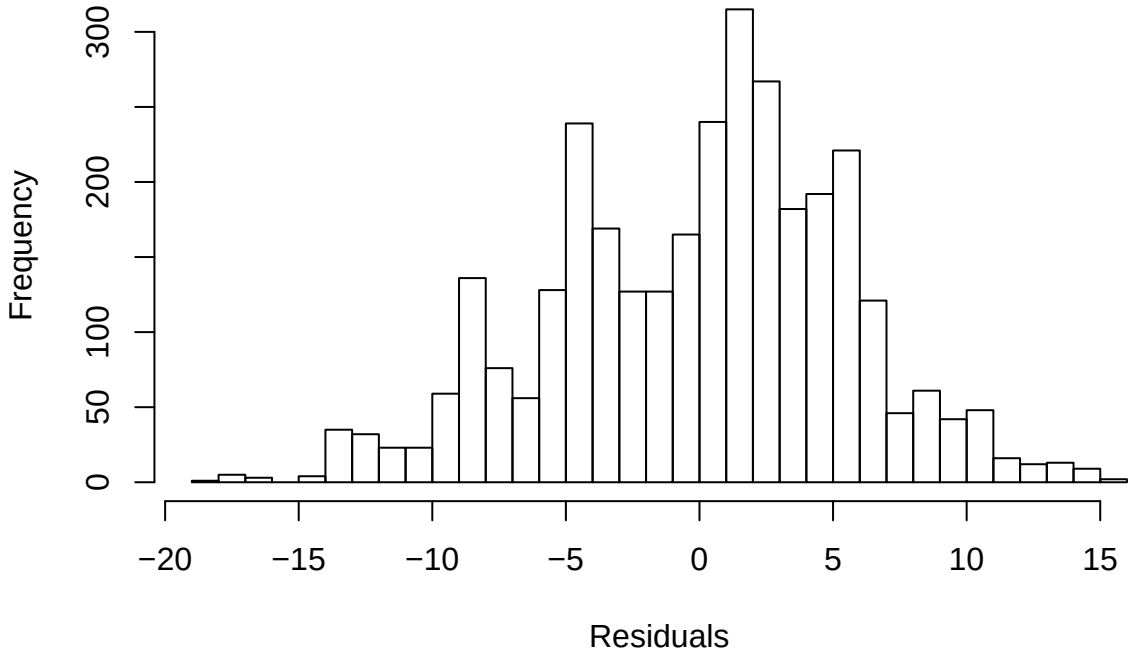
2.5

2.5.1

2.5.2

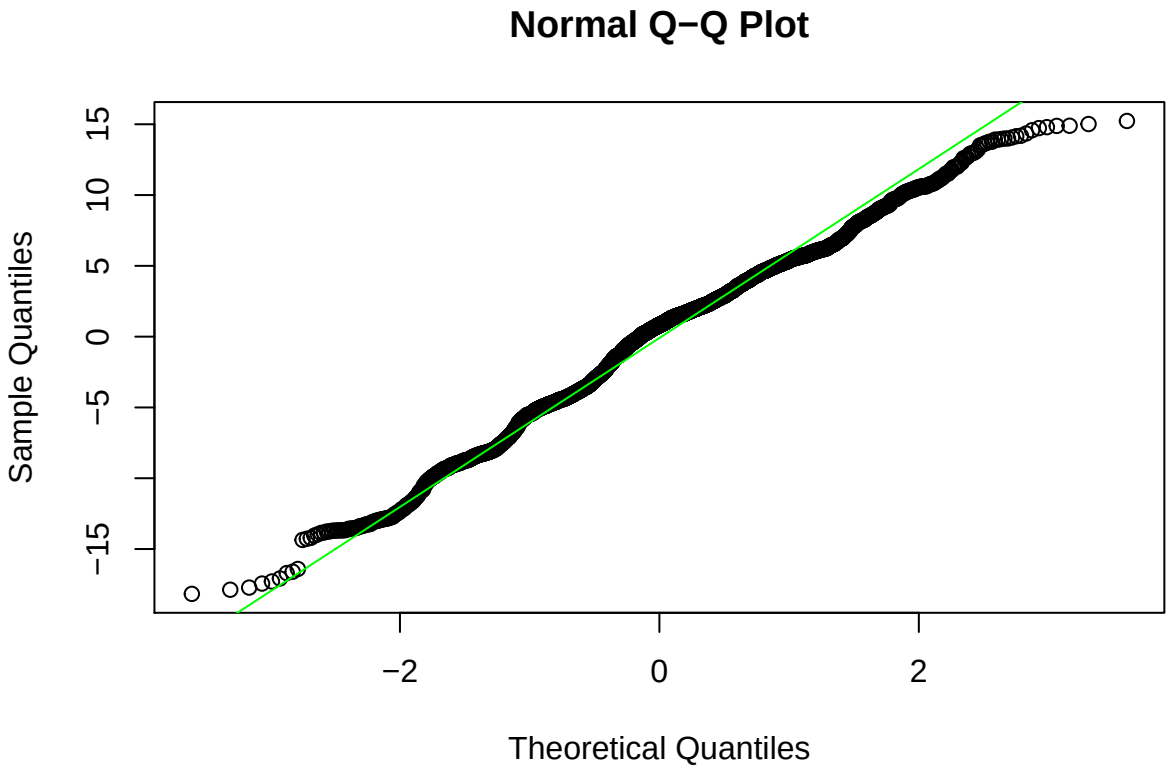
2.6

```
## [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
```




```
##
##  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