Empirical software engineering

Lab 2: ANOVA
Group 9

Anton Lutteman   Daniel Olsson   Gerson Silva Filho   Johan Mejborn

30 November 2020

Exercise 1 - Time to Develop

a) Minimum amount of users

```
pwr.anova.test(k = 5, n = NULL, f = 0.08, sig.level = 0.05, power = 0.90)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 5
##               n = 482
##               f = 0.08
##       sig.level = 0.05
##           power = 0.9
##
## NOTE: n is number in each group
```

According to the ANOVA test the amount of users in each group is **483** for the experiment have power 0.9. The minimum amount of monthly users that the company must have is 2415, which is 5 (the number of groups) times the amount of users in each group.

The effect size is inversely proportional to the number of users needed for the test. If we accept a larger effect the number of users needed shrinks, with this power and significance levels. On the other hand, if we make the effect smaller we need a higher number of users to this significance level and power.

b) Descriptive Statistics

```
df <- read.csv(file = 'gotaflix-abn.csv',sep = ",")
df$Cover <- as.factor(df$Cover)
df$Engagement <- as.numeric(df$Engagement)

psych::describeBy(df$Engagement,list(df$Cover), mat=T)
```

```
##     item group1 vars   n mean   sd median trimmed mad   min  max range    skew kurtosis     se
## X11    1      A    1 800 0.16 0.10   0.16    0.16 0.1 -0.14 0.54  0.68 -0.0075   -0.056 0.0036
## X12    2      B    1 800 0.16 0.10   0.16    0.16 0.1 -0.17 0.52  0.69 -0.0120    0.072 0.0036
## X13    3      C    1 800 0.18 0.11   0.17    0.18 0.1 -0.12 0.50  0.62  0.0271   -0.120 0.0037
## X14    4      D    1 800 0.16 0.10   0.16    0.16 0.1 -0.20 0.44  0.64 -0.1642   -0.112 0.0037
## X15    5      E    1 800 0.17 0.10   0.17    0.17 0.1 -0.15 0.48  0.63 -0.0415    0.058 0.0036
```

c) Linear Model

```
lm <- lm(Engagement ~ Cover,df)
```

Equation that represents the model:

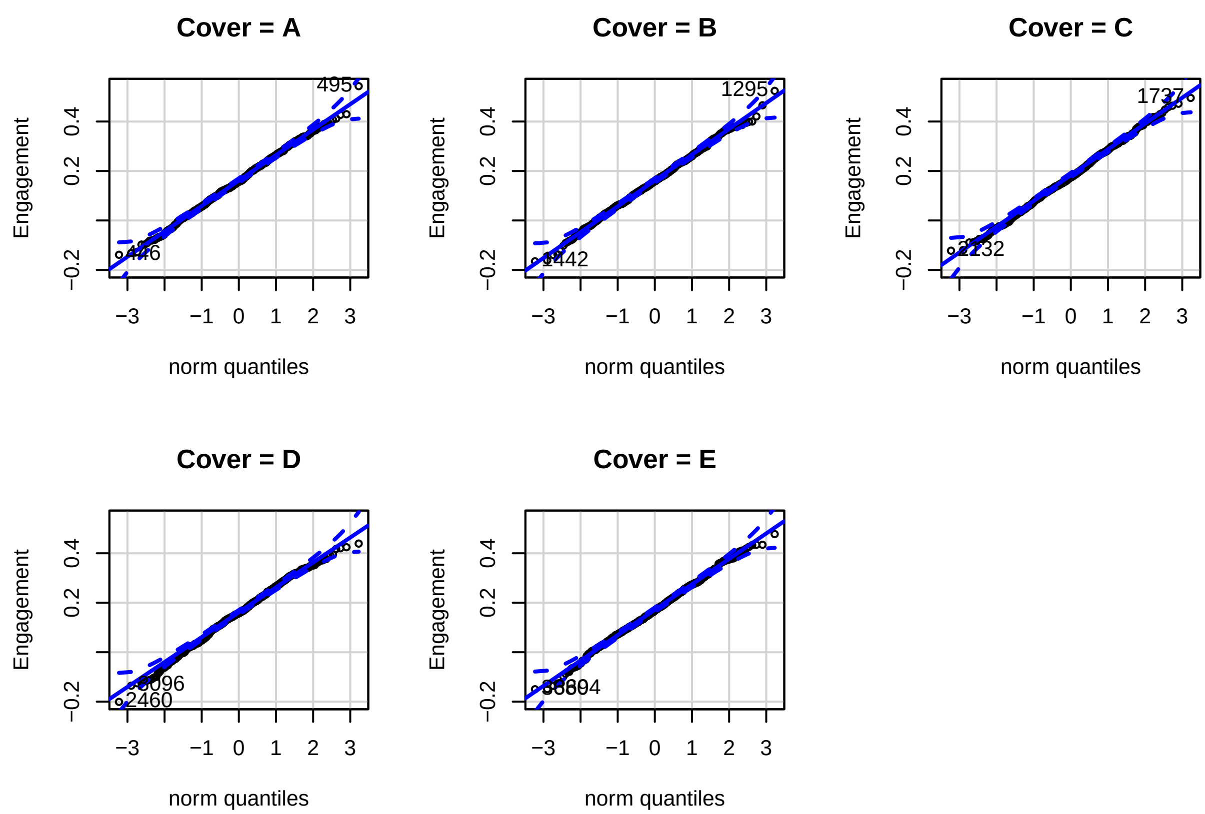
$$C = 1 - > 0.160367 + 0.017948 \tag{1}$$

The intercept, in this case, represents the **Cover A**.

If the the model gives only **Cover C** as 1, it means that it's the reference for all the other covers.

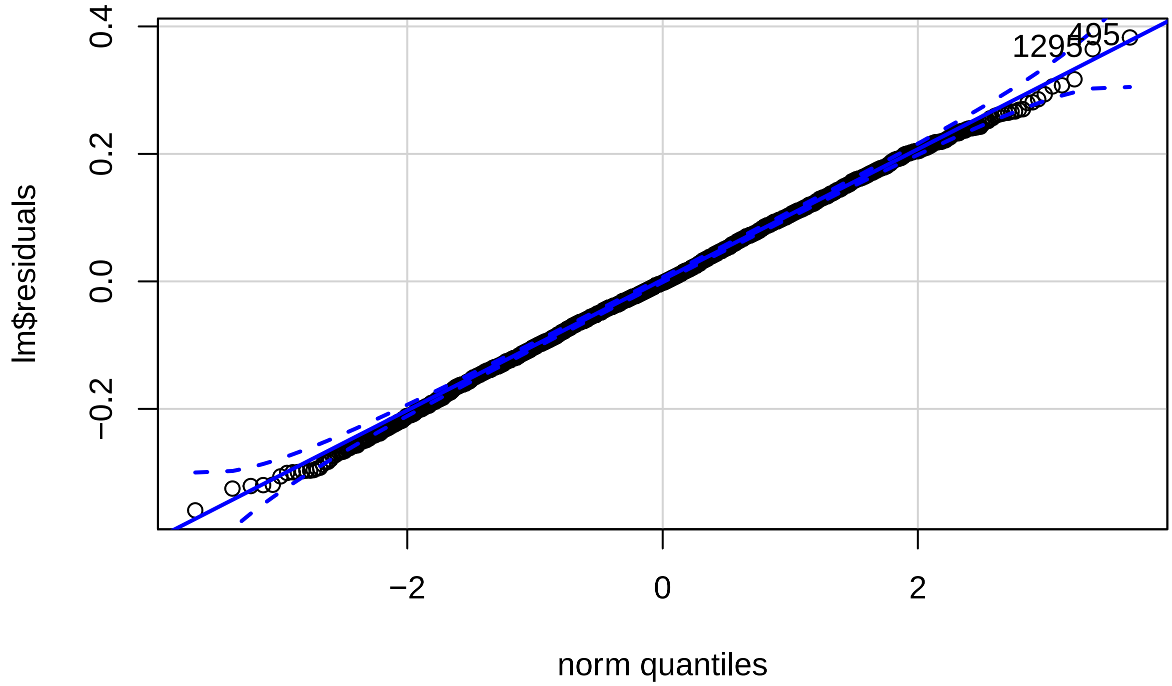
d) Normality

```
car::qqPlot(Engagement ~ Cover,df)
```



With this plots is a bit hard to be sure about the normality, but there are no reason to believe the opposite.

```r
car::qqPlot(lm$residuals)
```




```
## [1]  495 1295
```

Also the plot of the residuals look ok, but the data visualization is a bit hard.

```r
shapiro.test(df$Engagement)
```

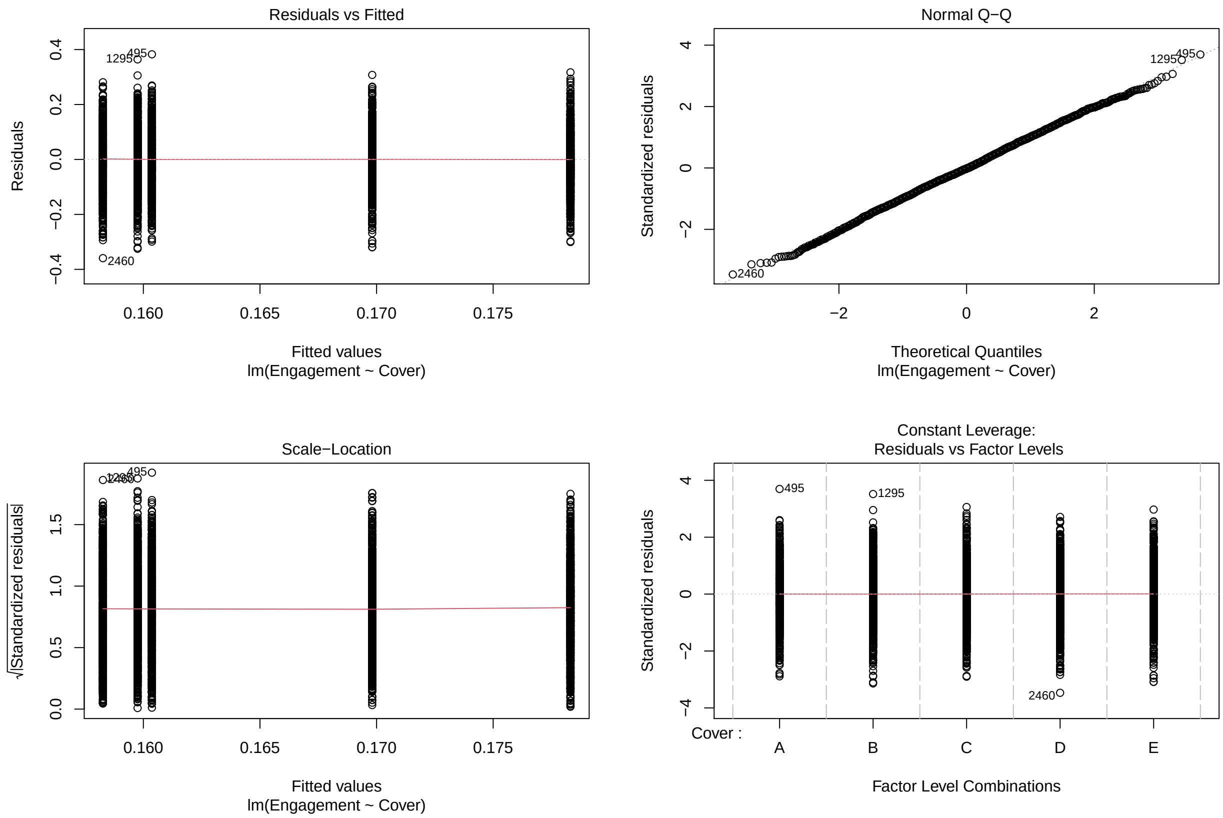
```
##
##  Shapiro-Wilk normality test
##
## data:  df$Engagement
## W = 1, p-value = 0.3
```

Using the Shapiro Wilk test we can believe that the data follows a normal distribution sing the W value is 1 and p-value is bigger than alpha.

e) Scatter plot

```r
plot(lm)
car::leveneTest(lm)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group    4    0.22   0.93
##       3995
```



We interpret the null hypothesis of the test as being if the data have homoscedasticity. Since the P-value is 0.93, it is larger than alpha and the null hypothesis can't be rejected.

f) Independence Assumption

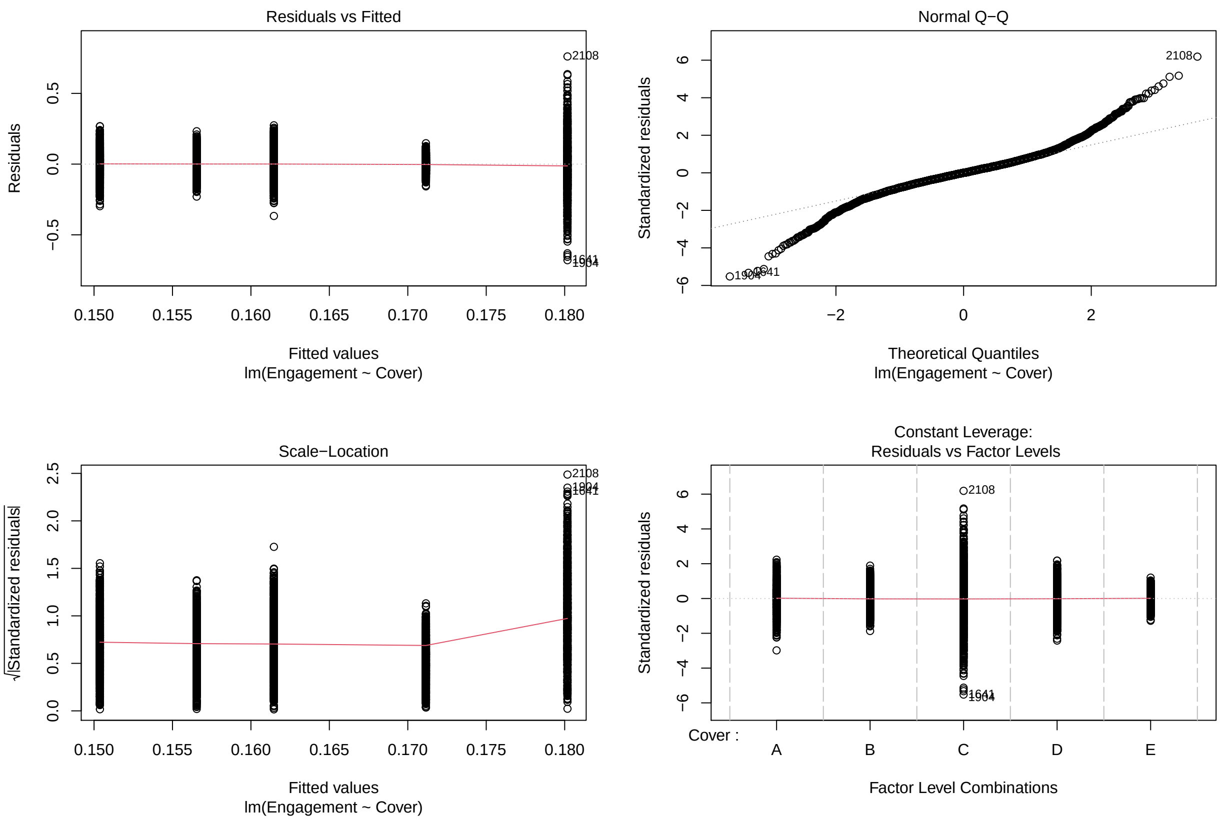
There is no test that can run to verify the independence of the data. It's part of the design of the experiment and should be handled in the collection phase.

g) Homoscedasticity analysis modified data - CHECK PLOT ARGUMENTS ««««««««<

```r
plot(lm2)
car::leveneTest(lm2)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group    4     359 <2e-16 ***
##       3995
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



We interpret the null hypothesis of the test as being if the data have homoscedasticity. Since the P-value is very small, and much smaller than alpha, the null hypothesis can be rejected and we can say with confidence that the data **does not have homoscedasticity.**

h) Which art cover had a better engagement?

```r
summary(lm)
```

```
##
## Call:
## lm(formula = Engagement ~ Cover, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3592 -0.0665 -0.0021  0.0716  0.3827
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.160367   0.003663   43.78  < 2e-16 ***
## CoverB      -0.000615   0.005180   -0.12  0.90556
```

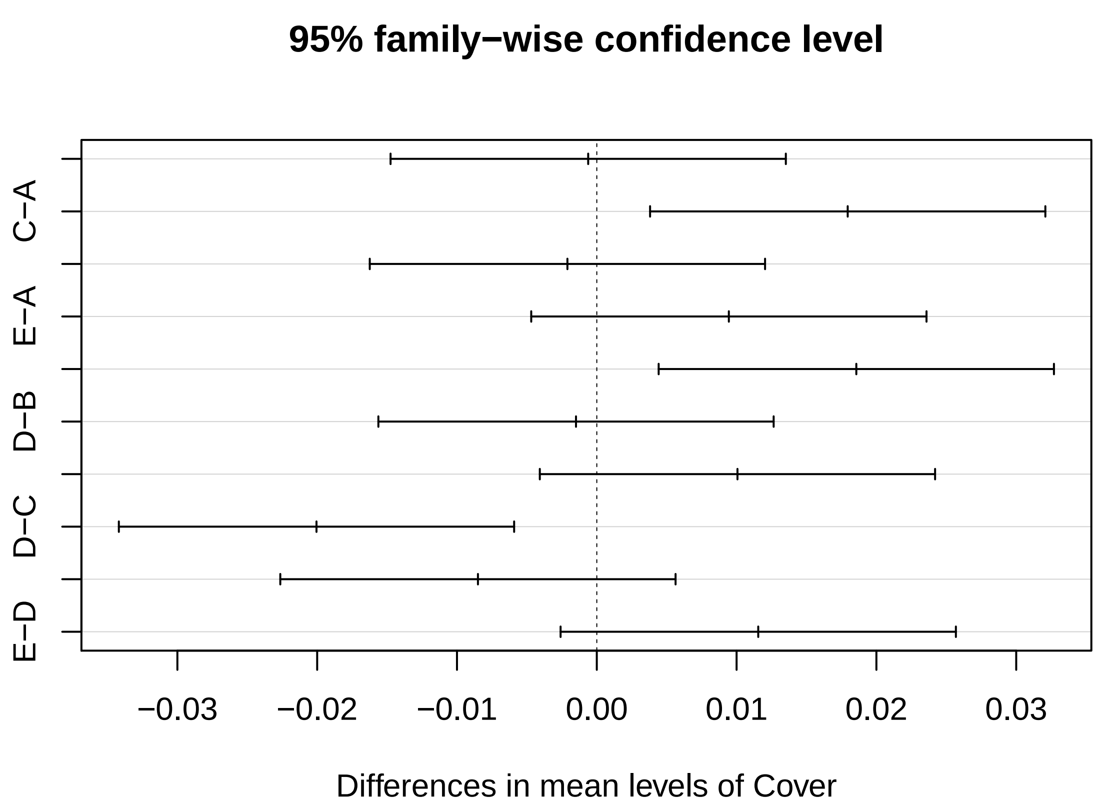
```
## CoverC       0.017948   0.005180    3.46  0.00054 ***
## CoverD      -0.002101   0.005180   -0.41  0.68504
## CoverE       0.009447   0.005180    1.82  0.06829 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1 on 3995 degrees of freedom
## Multiple R-squared:  0.00546,    Adjusted R-squared:  0.00447
## F-statistic: 5.48 on 4 and 3995 DF,  p-value: 0.000211
```

```r
car::Anova(lm)
```

```
## Anova Table (Type II tests)
##
## Response: Engagement
##             Sum Sq   Df F value  Pr(>F)
## Cover          0.2    4    5.48 0.00021 ***
## Residuals   42.9 3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model is statistically significant and we reject the hypothesis that the mean is equal for all groups.

```r
tuk <- TukeyHSD(aov(lm))
plot(tuk)
```



We are confident that Cover C is better than A, B and D by looking at the plot of the Tukey test. However, we can't say with confidence that C is better than E.

Exercise 2 - Full Factorial Experiment

a) Experimental Groups

b) Linear model equation

c) ANOVA assumptions

d) ANOVA table