

Empirical software engineering

Lab 2: ANOVA

Group 9

Anton Lutteman

Daniel Olsson

Gerson Silva Filho

Johan Mejbörn

10 december 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 = 482.3577
##              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
## X11      1      A      1 800 0.1603672 0.1026844 0.1565033 0.1605196 0.10165375 -0.1388049 0.5430277 0.6
## X12      2      B      1 800 0.1597526 0.1029644 0.1586958 0.1596599 0.10462177 -0.1653220 0.5239574 0.6
## X13      3      C      1 800 0.1783154 0.1053693 0.1742952 0.1783728 0.10461222 -0.1222303 0.4953971 0.6
## X14      4      D      1 800 0.1582659 0.1047037 0.1580229 0.1596672 0.09966692 -0.2009586 0.4392143 0.6
## X15      5      E      1 800 0.1698140 0.1022803 0.1682641 0.1699041 0.10205328 -0.1498057 0.4774524 0.6
##              skew      kurtosis      se
## X11 -0.007465712 -0.05591802 0.003630441
## X12 -0.012018272  0.07174217 0.003640342
## X13  0.027066690 -0.11965960 0.003725368
## X14 -0.164163135 -0.11158151 0.003701836
## X15 -0.041525134  0.05807999 0.003616155
```

c) Linear Model

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

```
##
## Call:
## lm(formula = Engagement ~ Cover, data = df)
##
## Coefficients:
## (Intercept)      CoverB      CoverC      CoverD      CoverE
##  0.1603672    -0.0006146    0.0179482   -0.0021013    0.0094468
```

Equation that represents the model:

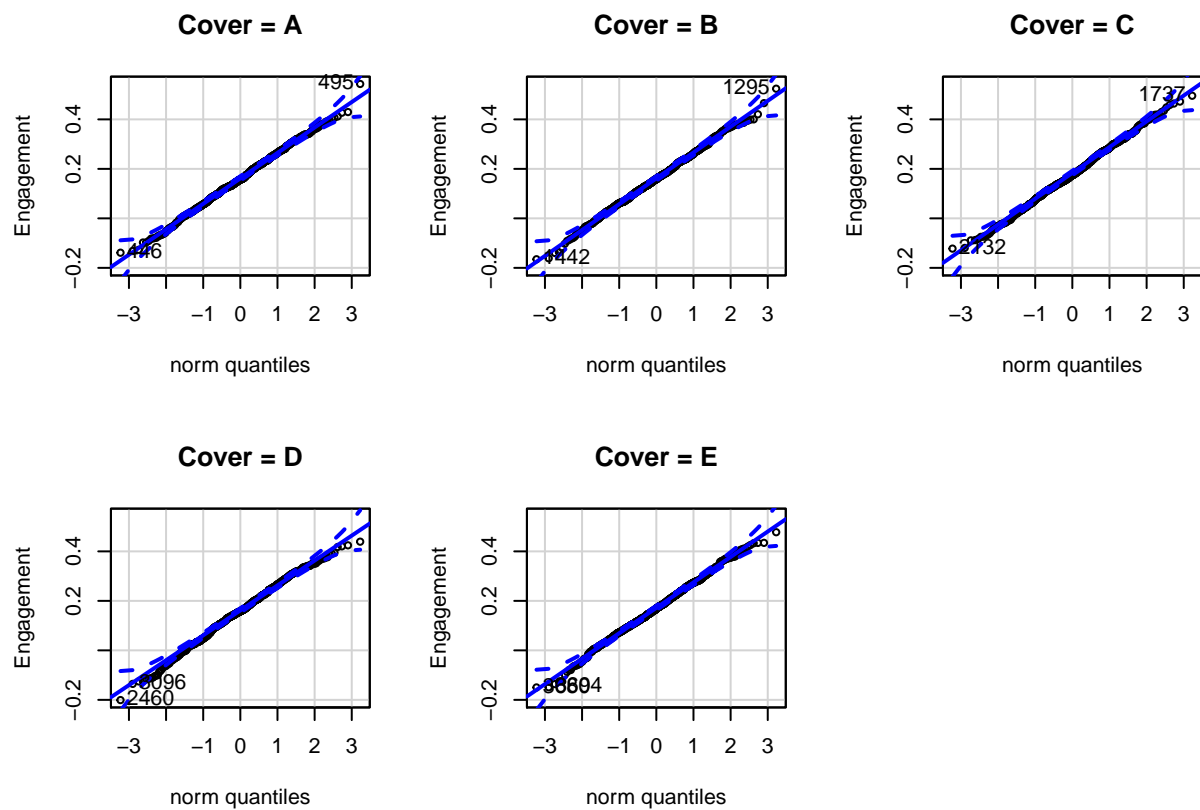
$$C = 1 - > 0.160367 + 0.017948 \quad (1)$$

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

If the model gives only **Cover C** as 1, it means that it's the mean value of Cover C.

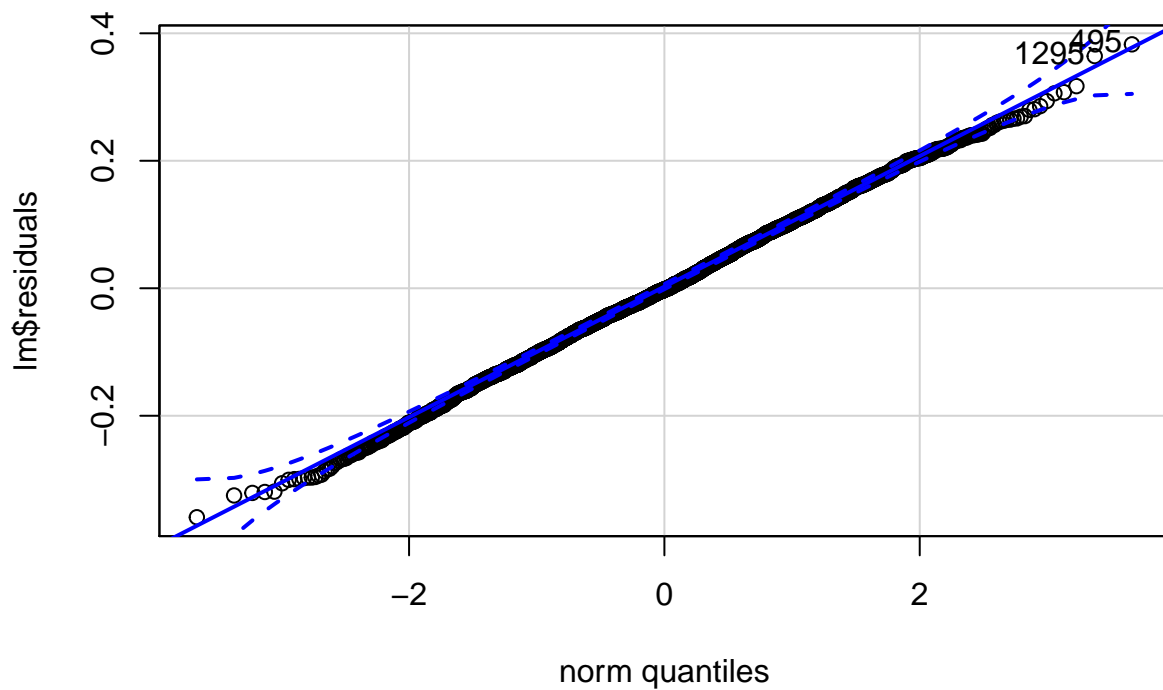
d) Normality

```
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.

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



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

Also the plot of the residuals indicates normality, but the data visualization is a bit hard.

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

```
##
##  Shapiro-Wilk normality test
##
## data:  df$Engagement
## W = 0.99943, p-value = 0.2713
```

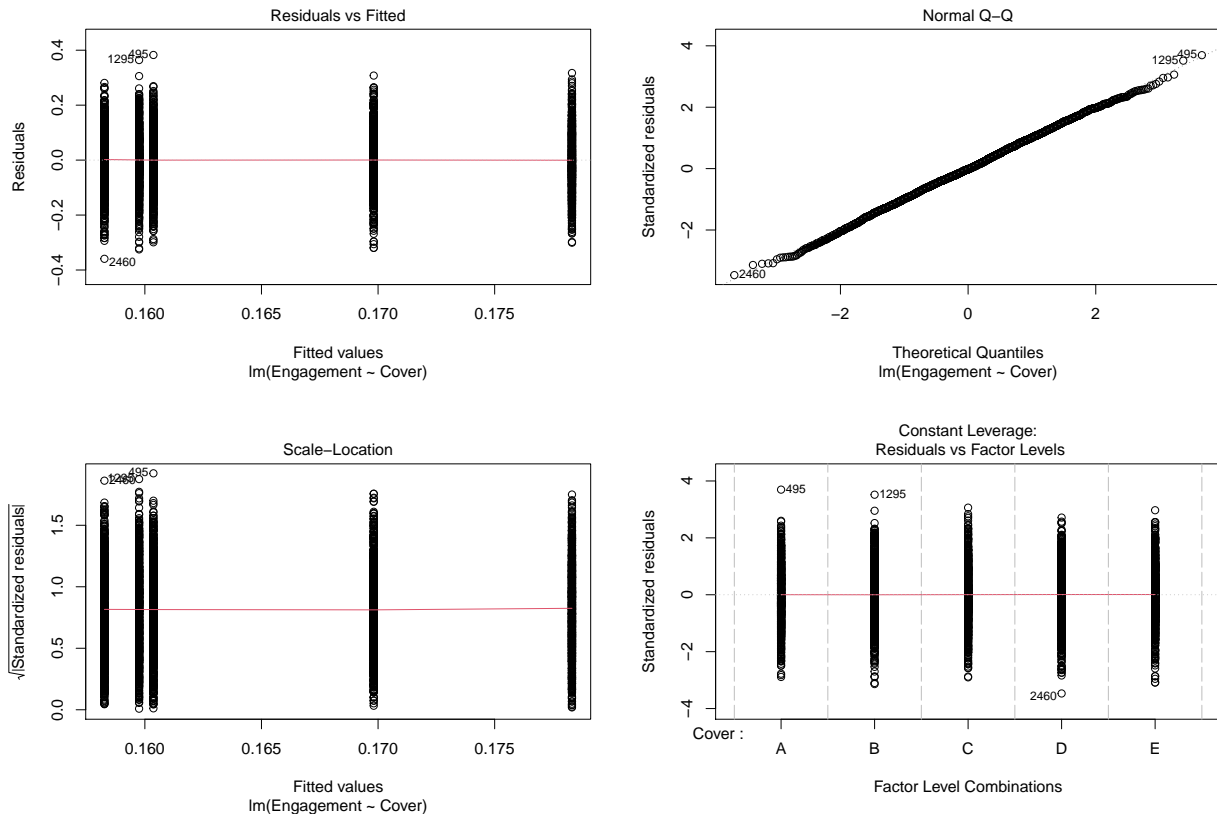
Using the Shapiro Wilk test we can believe that the data follows a normal distribution since the W value is 1 and p-value is bigger than alpha. Both the plot and the Shapiro Wilk test give the same indication of normality.

e) Scatter plot

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

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##           Df F value Pr(>F)
## group      4  0.2158 0.9297
##           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

```
car::leveneTest(lm2)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value    Pr(>F)
## group      4 358.91 < 2.2e-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?

```
summary(lm)
```

```
##
## Call:
## lm(formula = Engagement ~ Cover, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35922 -0.06652 -0.00211  0.07163  0.38266
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1603672   0.0036631  43.779  < 2e-16 ***
## CoverB       -0.0006146   0.0051804  -0.119  0.905564
## CoverC        0.0179482   0.0051804   3.465  0.000536 ***
## CoverD       -0.0021013   0.0051804  -0.406  0.685038
## CoverE        0.0094468   0.0051804   1.824  0.068290 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1036 on 3995 degrees of freedom
## Multiple R-squared:  0.005461, Adjusted R-squared:  0.004466
## F-statistic: 5.484 on 4 and 3995 DF, p-value: 0.0002114
```

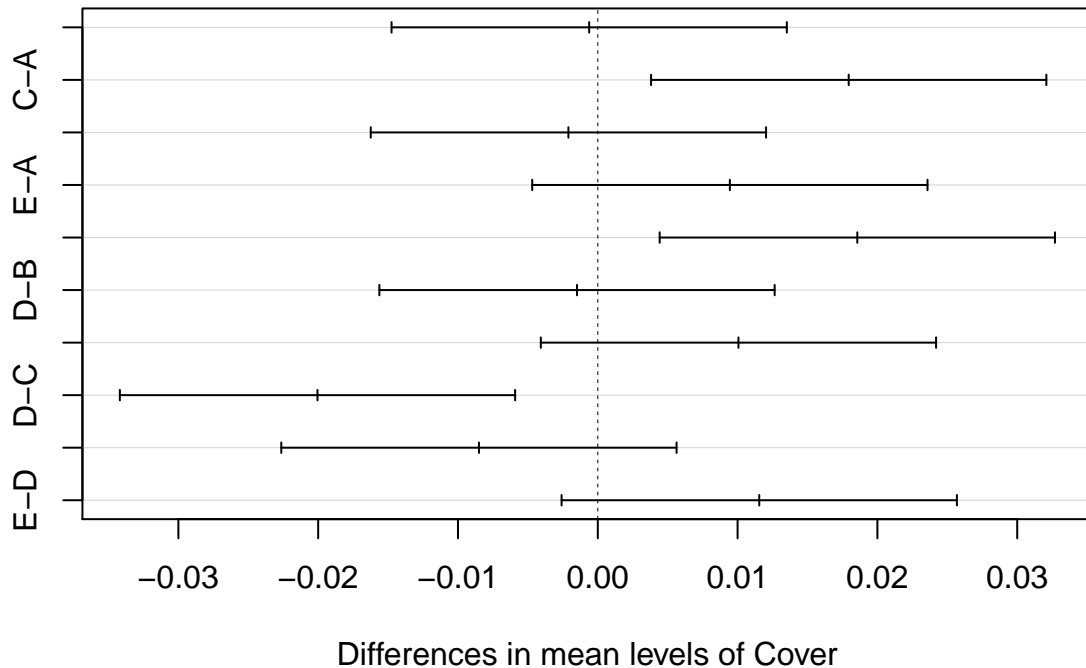
```
car::Anova(lm)
```

```
## Anova Table (Type II tests)
##
## Response: Engagement
##           Sum Sq   Df F value    Pr(>F)
## Cover         0.235    4  5.4844 0.0002114 ***
## Residuals 42.884 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.

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

95% family-wise confidence level



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

There are four experimental groups assigned to each of the combinations of treatments. The groups consist of the randomly assigned users.

b) Linear model equation

```
lm(Engagement ~ Cover + Summary, df3)
```

```
##
## Call:
## lm(formula = Engagement ~ Cover + Summary, data = df3)
##
## Coefficients:
## (Intercept)    CoverGenre  SummaryGenre
##      0.169028      -0.008812      -0.006087
```

Equation that represents the model:

$$\text{Engagement} = 0.169028 - 0.008812 * \text{Cover} - 0.006087 * \text{Summary} \quad (2)$$

```
lm(Engagement ~ Cover*Summary, df3)
```

```
##
## Call:
## lm(formula = Engagement ~ Cover * Summary, data = df3)
##
## Coefficients:
##              (Intercept)              CoverGenre              SummaryGenre  CoverGenre:SummaryGenre
##              0.17527              -0.02129              -0.01856              0.02495
```

Equation that represents the model:

$$\text{Engagement} = 0.17527 - 0.02129 * \text{Cover} - 0.01856 * \text{Summary} + (0.02495 * \text{Cover} * \text{Summary}) \quad (3)$$

The intercept, in both cases, is the engagement for factors Cover = Character and Summary = Character

c) ANOVA assumptions

The samples were taken randomly so the assumption of independence holds.

Test for normality:

```
shapiro.test(df3$Engagement)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df3$Engagement
## W = 0.99938, p-value = 0.3843
```

The Shapiro Wilks test with W=0.999 with a p value of 0.3843, therefore we cannot reject the null hypothesis.

Test for homoscedasticity:

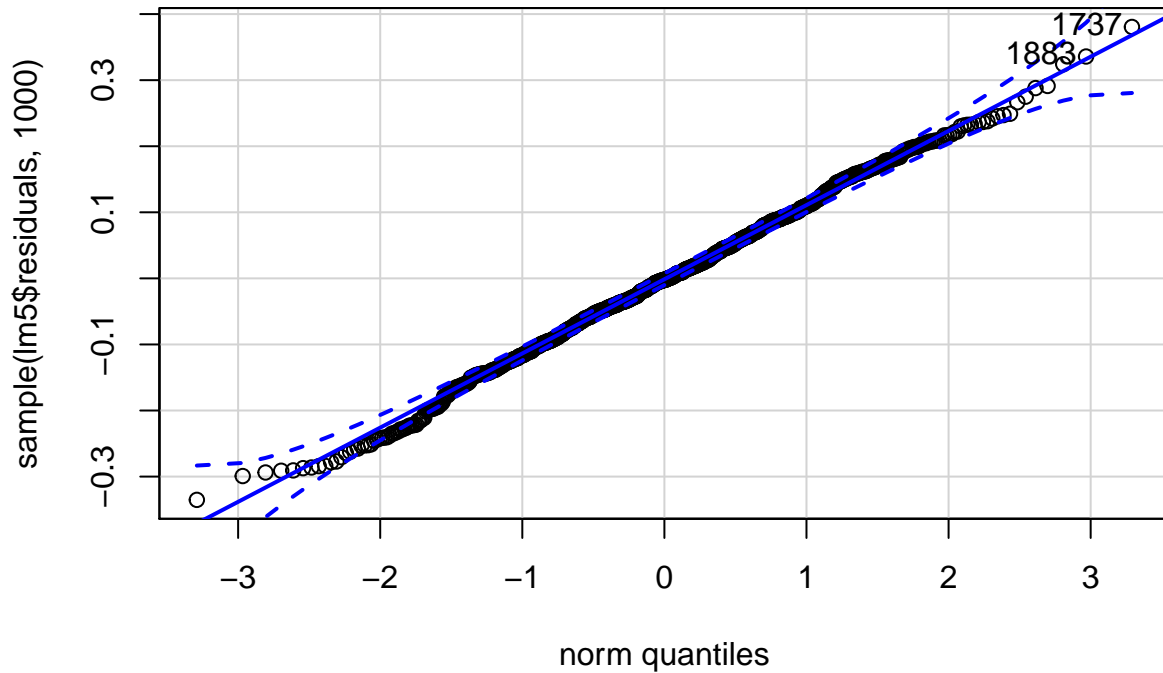
```
car::leveneTest(lm5)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##              Df F value    Pr(>F)
## group      3  15.873 3.04e-10 ***
##              3196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of the levene test tells us that we can reject the null hypothesis that the data is homoscedastic, however with large sample sizes the test is likely to report small p values for very small deviations from homoscedasticity.

We plot the data to observe the homoscedasticity, also we sampled it in order to get a better visualization.


```
car::qqPlot(sample(lm5$residuals,1000))
```



```
## 1737 1883
## 999 313
```

Our conclusion is that we cannot reject the hypothesis that the data is homoscedastic.
Given the analysis we assume that the assumptions hold.

d) ANOVA table

```
stargazer::stargazer(car::Anova(lm5),summary = F, header = F)
```

Table 1:

	Sum Sq	Df	F value	Pr(>F)
Cover	0.062	1	4.709	0.030
Summary	0.030	1	2.247	0.134
Cover:Summary	0.125	1	9.440	0.002
Residuals	42.156	3,196		

The conclusion that we draw from the ANOVA table is that the Cover and the Interaction were statistically significant. The effect sizes of each value can be seen in Table 1 F Value column.

```
TukeyHSD(aov(lm5))
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lm5)
##
## $Cover
##              diff              lwr              upr              p adj
## Genre-Character -0.008811645 -0.01677315 -0.000850142 0.0300747
##
## $Summary
##              diff              lwr              upr              p adj
## Genre-Character -0.006087055 -0.01404856 0.001874448 0.1339517
##
## $'Cover:Summary'
##              diff              lwr              upr              p adj
## Genre:Character-Character:Character -0.02128739 -0.036047699 -0.0065270815 0.0012203
## Character:Genre-Character:Character -0.01856280 -0.033323108 -0.0038024911 0.0067932
## Genre:Genre-Character:Character -0.01489870 -0.029659009 -0.0001383913 0.0468911
## Character:Genre-Genre:Character 0.00272459 -0.012035718 0.0174848992 0.9647097
## Genre:Genre-Genre:Character 0.00638869 -0.008371618 0.0211489989 0.6818224
## Genre:Genre-Character:Genre 0.00366410 -0.011096209 0.0184244085 0.9196953
```

According to the Tukey test we can say that the Character is better than Genre for Cover to, get a higher engagement. For the interaction, we see that the combination Character - Character for both factors is the one with higher engagement and also the ones statistically significant.