WS 2021/22 1048 - Econometrics I
Case Study 3

Anja Kulagic (h12100543), Jovana Mileusnic (h12100542), Ema Vargova (h11914081)

Contents

List of Figures

List of Tables

1 Data Description

```
library(knitr)
data <- read.csv("movie.csv", sep = ";", dec = ",")
data$Screens <- as.numeric(data$Screens)
data$Animation <- as.factor(data$Animation)
data$Family <- as.factor(data$Family)
data$Adventure <- as.factor(data$Adventure)
```

```
metric.index <- which(lapply(data, class) == "numeric")
factor.index <- which(lapply(data, class) == "factor")
op <- par(mfrow=c(3, 4), cex=.5, mar = c(1.5, 2.5, 1.5, 1), mgp = c(1.3, .5, 0))
for(i in 1:ncol(data)) {
  if(i %in% metric.index) {
    hist(data[[i]], freq=FALSE, main=names(data)[i], xlab="")
  } else {
    barplot(table(data[[i]]), main=names(data)[i], ylab="Frequency") }}
```

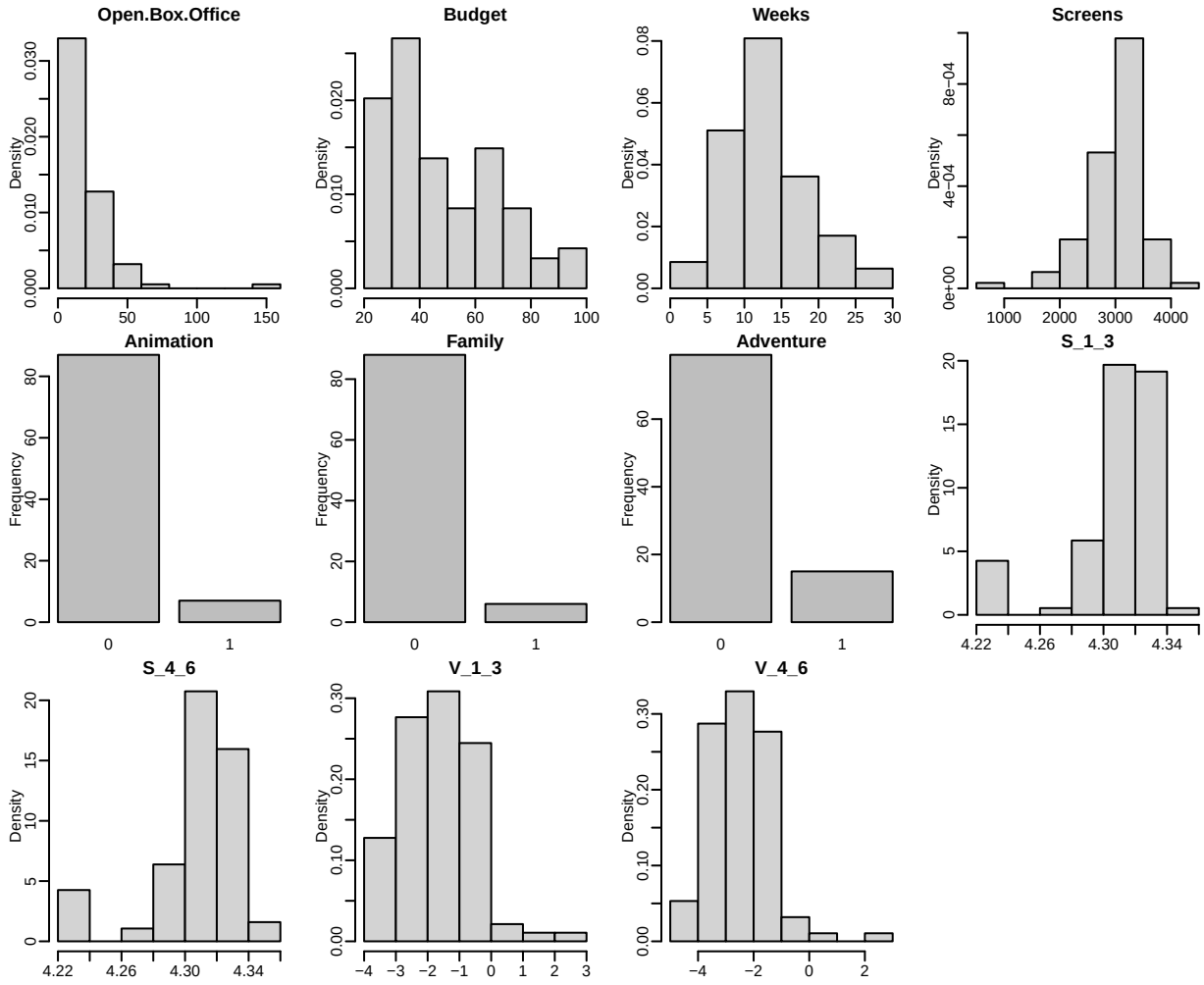

Figure 1: Distribution of variables

2 Data modelling

2.1

```
library(stargazer)
lin.model1 <- lm(Open.Box.Office ~ ., data=data)
stargazer(lin.model1, single.row=TRUE, header = FALSE,
          title="Model 1 - Multiple OLS Regression Model")
```

Table 1: Model 1 - Multiple OLS Regression Model

	Dependent variable:
	Open.Box.Office
Budget	0.133* (0.075)
Weeks	0.699** (0.293)
Screens	0.012*** (0.003)
Animation1	−17.625*** (6.344)
Family1	9.566 (6.387)
Adventure1	8.924** (4.084)
S_1_3	−8.182 (228.450)
S_4_6	101.375 (225.491)
V_1_3	12.692*** (3.938)
V_4_6	−7.020* (3.853)
Constant	−430.199* (218.660)
Observations	94
R^2	0.568
Adjusted R^2	0.516
Residual Std. Error	12.898 (df = 83)
F Statistic	10.901*** (df = 10; 83)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 1 was created using stargazer v.5.2.2 by Hlavac (2018).

Based on Model 1, we see that additional million of movie budget increases the the revenue of the movie by 133,000 dollars and one more week in theaters increases the revenue by 699,000 dollars while the additional screening of the movie increases the revenue by 12,000 dollars. When looking at the genres, when the movie is a family or adventure movie, it increases its revenue by 9,566,000 and 8,924,000 dollars respectively while when the movie is an animation, it decreases its revenue by 17,625,000 dollars. When it comes to the sentiment score of the tweets posted about a movie, and additional point in the sentiment score increases the revenue of the movie by 101,375,000 dollars in the four to six weeks before the release while an additional point decreases the revenue by 8,182,000 dollars when the tweet is posted one to three weeks before the movie. Regarding the volume of the tweets, one additional point in the volume score decreases the revenue by 7,020,000 dollars four to six weeks before the release while the score increases it by 12,692,000 dollars one to three before the movie release. The reported coefficients of the variables are their average effects on the movie's revenue while holding other variables constant.

Dummy variables included in data set are not complementary, as its sum does not have necessarily to be one, so that the potential problem with perfect multicollinearity can be excluded in advance.

2.2

The coefficient of determination is 56.8% which is the proportion of variance in the revenue generated by the movies which can be explained by the regressors included in Model 1.

2.3

```
op <- par(mar = c(2.5, 2.5, 2.5, 1), mgp = c(1.3, .5, 0), cex=0.7)
plot(1:nrow(data), lin.model1$residuals, pch=19, ps=0.6, main="Residual plot",
     xlab="Row number", ylab="Residuals")
abline(h=mean(lin.model1$residuals), lty=2)
legend("topleft", legend=c("Mean"), lty=2)
```

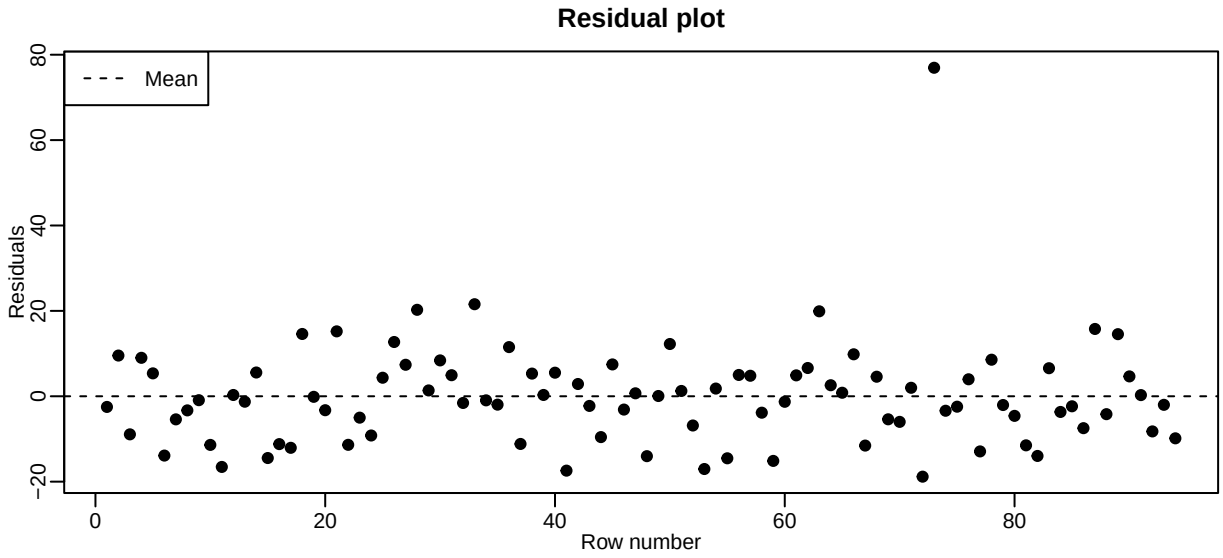

Figure 2: Model 1 - Residual plot

Based on the residual plot, the model specification assumption seems to be correct as the residuals are centered around zero regardless of the what the values of the different regressors are. Furthermore, the homoskedacity assumption is also met as the variance of the residuals does not seem to be dependent on the values of different regressors when we disregard one outlier in the residuals.

Although we may notice the presence of the outlier which can potentially affect the normal distribution of our residues, the histogram plot in some way esemble the shape of the density of a normal distribution, whereas the dots on the Q-Q lie on the line, at least approximately.

```
resids <- residuals(lin.model1)
```

```
op <- par(mfrow = c(1,2), mar = c(2.2, 2, 1.5, 1), mgp = c(1.2, .5, 0),
          cex.lab=0.7, cex.axis=0.7, cex.main=0.8)
hist(resids, breaks=20, xlab="Residuals", main="Histogram of Residuals")
qqnorm(resids,pch=19,cex=0.5)
qqline(resids, col="red")
```

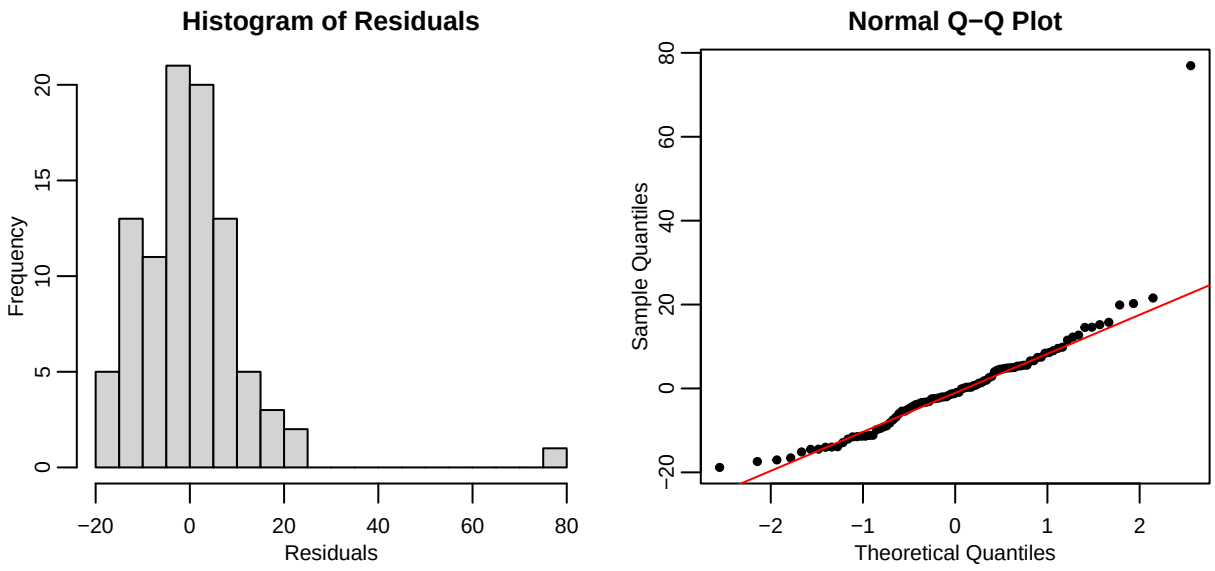

Figure 3: Model 1 - Histogram of Residuals and Normal Q-Q Plot

Jarque-Bera-Statistics:

$$J = \frac{N-K}{6}(m_3^2 + \frac{1}{4}(m_4 - 3)^2)$$

H_0 : The errors follow a normal distribution.

We are rejecting H_0 if $J > \chi^2_{2,0.95}$ (or p-value of J smaller than 0.05).

```
tseries::jarque.bera.test(resids)
```

```
##
##  Jarque Bera Test
##
## data:  resids
## X-squared = 1008.5, df = 2, p-value < 2.2e-16
```

After performing the Jarque-Bera-test and obtaining p-value at value of 2.2e-16, we are rejecting the null hypothesis and concluding that the data does not come from a normal distribution. If your residuals are severely non-normal, your t-statistics, p-values, and hypothesis tests will be meaningless. Note that Gauss-Markov assumptions still holds even though they do not follow normal distribution, but we are losing the practical usefulness.

However, violation of the normality assumption only becomes an issue with small sample sizes. For large sample sizes (assuming that this is our case), the assumption is less important due to the central limit theorem, and the fact that the F- and t-tests used for hypothesis tests and forming confidence intervals are quite robust to modest departures from normality.

2.4

a)

$$H_0 : \beta_7 = 0; \beta_8 - \beta_9 = 0$$
$$H_1 : \beta_7 \neq \text{or } \beta_8 - \beta_9 \neq 0$$

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.2
```

```
kable(linearHypothesis(lin.model1, c("V_1_3=S_4_6", "S_1_3=0")), caption = "Testing the null hypothesis
```

Table 2: Testing the null hypothesis: The effects of the tweet volume in weeks 1-3 and the sentiment score in weeks 4-6 are the same while the sentiment score in weeks 1-3 has no effect on Open.Box.Office (jointly), ceteris paribus.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
85	14236.77	NA	NA	NA	NA
83	13807.37	2	429.3974	1.290614	0.2805643

where the first row stands for the restricted, and the second one for unrestricted model.

The output reveals that the F-statistic for this joint hypothesis test is about 1.290614 and the corresponding p-value is 0.2805643. Thus, we cannot reject the null hypothesis that the effects of the tweet volume in weeks 1-3 and the sentiment score in weeks 4-6 are the same while the sentiment volume in weeks 1-3 has no effect on Open.Box.Office at any level of significance commonly used in practice.

b)

$$H_0 : \beta_7 = \beta_8 = 0$$
$$H_1 : \beta_i \neq 0 \text{ for at least one } i = 7, 8$$

```
kable(linearHypothesis(lin.model1, c("S_1_3=0", "S_4_6=0")),
      caption="Testing the null hypothesis: The sentiment score variables together
      have no impact on the outcome, ceteris paribus")
```

Table 3: Testing the null hypothesis: The sentiment score variables together have no impact on the outcome, ceteris paribus

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
85	14384.04	NA	NA	NA	NA
83	13807.37	2	576.6698	1.733262	0.1830411

where the first row stands for the restricted, and the second one for unrestricted model.

Restricted (true) model: $\text{Open.Box.Office} = \beta_0 + \beta_1 Xbudget + \beta_2 Xweeks + \beta_3 Xscreens + \beta_4 Xanimation + \beta_5 Xfamily + \beta_6 Xadventure + \beta_9 X_{V:1,3} + \beta_{10} X_{V:4,6}$

The homoskedasticity-only F-Statistic is given by

$$F = \frac{(SSR_{restricted} - SSR_{unrestricted})/q}{SSR_{unrestricted}/(n-k-1)}$$

with $SSR_{restricted}$ being the sum of squared residuals from the restricted regression, i.e., the regression where we impose the restriction. $SSR_{unrestricted}$ is the sum of squared residuals from the full model, q is the number of restrictions under the null and k is the number of regressors in the unrestricted regression.

$$F = \frac{(14384 - 13807)/2}{13807/83} \approx 1.734$$

Both results lead us to the same conclusion - the F-statistic for this joint hypothesis test is about 1.173 and the corresponding p-value is 0.183.

After performing a hypothesis test, we can notice that our p-value is greater than any level of significance commonly used in practice, hence our sample did not provide sufficient evidence to conclude that the effects exist.

Table 4: ANOVA table

Sorce of Variance	df	Sum of square
Residual of restricted model	(n-p-1)=85	14384.04
Additional amount of residuals	(k-p)=2	576.67
Residual of Full model	(n-k-1)=83	13807.37

```
op <- par(mar = c(2.3, 2, 1, 1), mgp = c(1.1, .5, 0), cex.lab=0.8, cex.axis=0.6, cex.main=0.8)
confidenceEllipse(lin.model1, which.coef = c("S_1_3", "S_4_6"), fill = T, lwd = 0,
                  ylab=bquote(paste(beta)[8]), xlab=bquote(paste(beta)[7]))
```

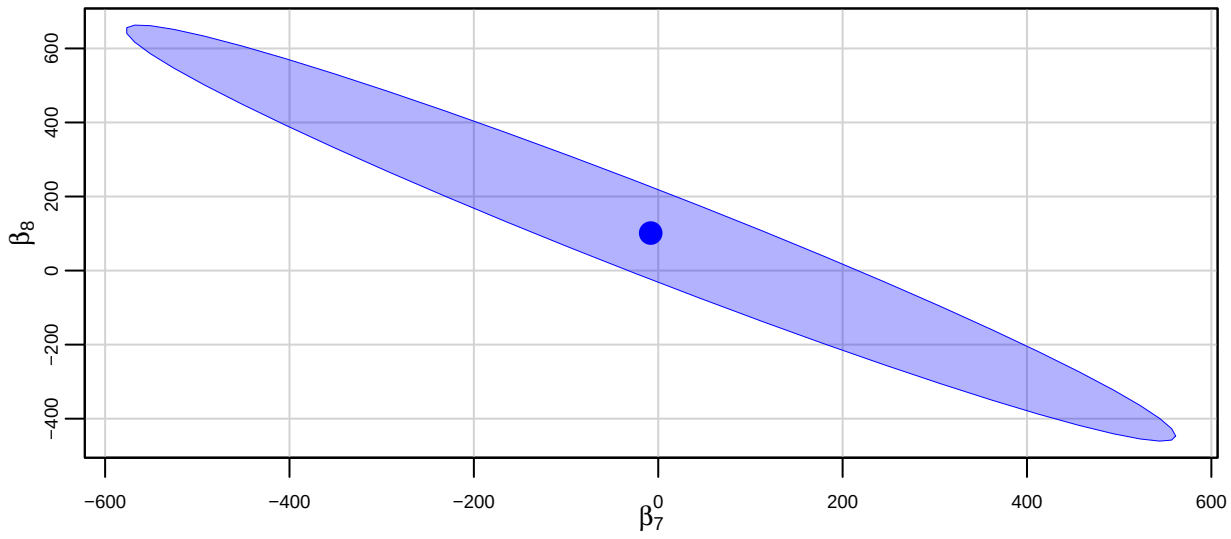

Figure 4: Model 1 - 95% Confidence Set

Note that (0,0) is element of the 95% confidence set so that we cannot reject $H_0 : \beta_7 = \beta_8 = 0$.

c)

$$H_0 : \beta_9 = \beta_{10} = 0$$

$$H_1 : \beta_i \neq 0 \text{ for at least one } i = 9, 10$$

```
kable(linearHypothesis(lin.model1, c("V_1_3=0", "V_4_6=0")),
      caption = "Testing the null hypothesis: The volume score variables
      together have no impact on the outcome, ceteris paribus")
```

Table 5: Testing the null hypothesis: The volume score variables together have no impact on the outcome, ceteris paribus

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
85	17469.20	NA	NA	NA	NA
83	13807.37	2	3661.832	11.00615	5.76e-05

where the first row stands for the restricted, and the second one for unrestricted model.

The output reveals that the F-statistic for this joint hypothesis test is about 11.006 and the corresponding p-value is 0.00006. Thus, we can reject the null hypothesis that both observed coefficients are zero at any level of significance commonly used in practice.

d)

$$H_0 : \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : \beta_i \neq 0 \text{ for at least one } i = 4, 5, 6$$

```
kable(linearHypothesis(lin.model1, c("Animation1=0", "Family1=0", "Adventure1=0")), caption="Testing th
```

Table 6: Testing the null hypothesis: The genre variables have jointly no effect on Open.Box.Office.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
86	15518.49	NA	NA	NA	NA
83	13807.37	3	1711.121	3.428677	0.0207732

where the first row stands for the restricted, and the second one for unrestricted model.

The output reveals that the F-statistic for this joint hypothesis test is about 3.428677 and the corresponding p-value is 0.0208. Thus, we can reject the null hypothesis that genre variable coefficients are zero at 0.1, 0.05 levels of significance which are commonly used in practice, whereas when testing $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$ with significance level $\alpha = 0.01$, we are failing to reject the null.

2.5

a) The effects of 'weeks,' denoted β_1, equals 0.699, meaning that if one more week in theaters increases the revenue by 699,000 dollars, while holding all other variables fixed.

b)

```
kable(cbind("The effects of 'weeks'" = summary(lin.model1)$coef[3,1],
            "The effects of 'screens'" = summary(lin.model1)$coef[4,1]), caption="Effect of weeks vs. e
```

Table 7: Effect of weeks vs. effect of screens

The effects of 'weeks'	The effects of 'screens'
0.6988837	0.0118087

Regarding the effect of variable 'weeks' along with the effect of variable 'screens,' if the predetermined number of weeks is increased by 10 weeks, but the number of screens is decreased by 15 screens, the expected overall increase in revenue is 6,812,869.5 dollars, ceteris paribus.

c) The effects of 'Animation,' denoted β_4, equals -17.625. As a binary variable, the only possible change in variable 'Animation' is from zeto to one ($\Delta x = 1$), meaning that a movie is additionally assigned to the genre Animation. If this were to happen, it decreases its expected revenue by 17,625,000 dollars on average, while holding all other variables fixed. This change is equal to change of the intercept by -17.625.

d)

```
kable(round(confint(lin.model1, level = 0.99), 4), caption="Confidence interval of regression coefficie
```

Table 8: Confidence interval of regression coefficients

	0.5 %	99.5 %
(Intercept)	-1006.6662	146.2687
Budget	-0.0649	0.3303
Weeks	-0.0727	1.4705
Screens	0.0039	0.0197
Animation1	-34.3501	-0.9009
Family1	-7.2721	26.4044
Adventure1	-1.8431	19.6906
S_1_3	-610.4604	594.0963
S_4_6	-493.1013	695.8520
V_1_3	2.3108	23.0738
V_4_6	-17.1773	3.1375

Computing this interval in R manually:

If σ^2 is unknown, then $\text{sd}(\hat{\beta}|\mathbf{X})$ is substituted by $\text{se}(\hat{\beta})$.

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{df}$$

If $t_{df,p}$ is the p-quantiles of the $t_{df} - distribution$, this yields:

(1) β_j lies in $[\hat{\beta}_j - t_{df,1-\alpha/2} se(\hat{\beta}_j), \hat{\beta}_j + t_{df,1-\alpha/2} se(\hat{\beta}_j)]$

(2) $\hat{\beta}_j + t_{df,1-\alpha/2} se(\hat{\beta}_j)$ is an upper bound for β_j

(3) $\hat{\beta}_j - t_{df,1-\alpha/2} se(\hat{\beta}_j)$ is a lower bound for β_j

```
lm_sum1 <- summary(lin.model1)
kable(cbind("Lower" = lm_sum1$coef[3,1] - qt(0.995, df = lm_sum1$df[2])*lm_sum1$coef[3, 2],
            "Upper" = lm_sum1$coef[3,1] + qt(0.995, df = lm_sum1$df[2])*lm_sum1$coef[3, 2]),
      caption="99% confidence interval for the effect of 'Weeks")
```

Table 9: 99% confidence interval for the effect of 'Weeks

Lower	Upper
-0.0727424	1.47051

The upper and the lower bounds coincide. We have used the 0.995-quantile of the $t_8 3$ distribution to get the exact result reported by confint().The interval that contains the true value β_j in 99% of all samples is given by [-0.0727424, 1.47051]. Equivalently, this interval can be seen as the set of null hypotheses for which a 1% two-sided hypothesis test does not reject. Notice that this interval does contain the value zero, thus when testing $H_0 : \beta_{weeks} = 0$ with significance level $\alpha = 0.01$, we are failing to reject the null indicates as our sample did not provide sufficient evidence to conclude that the effect exists. However, at the same time, that lack of evidence does not prove that the effect does not exist.

e)

```
lin.model2 <- lm(Open.Box.Office ~ data$Budget + data$Weeks + data$Screens
                 + data$Animation + data$Family + data$Adventure, data=data)
stargazer(lin.model2, single.row=TRUE, header = FALSE,
          title="Model 2 - Multiple OLS Regression Model (excluding social media data)")
```

Table 10: Model 2 - Multiple OLS Regression Model (excluding social media data)

	Dependent variable:
	Open.Box.Office
Budget	0.122 (0.081)
Weeks	0.928*** (0.318)
Screens	0.015*** (0.003)
Animation1	−19.246*** (6.857)
Family1	5.084 (6.957)
Adventure1	10.271** (4.472)
Constant	−46.042*** (9.172)
Observations	94
R^2	0.446
Adjusted R^2	0.408
Residual Std. Error	14.263 (df = 87)
F Statistic	11.667*** (df = 6; 87)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10 was created using stargazer v.5.2.2 by Hlavac (2018).

```
lm_sum2 <- summary(lin.model2)
kable(cbind("Adjusted R-squared: Model 1" = round(lm_sum2$adj.r.squared,3),
  "Adjusted R-squared: Model 2" = round(lm_sum2$adj.r.squared,3)),
  caption="Comparison of Adjusted R-squared")
```

Table 11: Comparison of Adjusted R-squared

Adjusted R-squared: Model 1	Adjusted R-squared: Model 2
0.408	0.408

Based on the comparison between the Adjusted R^2 for Model 1 and Model 2 (afted exlucidng social media data), we could notice that the additional 10.8% of proportion of variance in the revenue generated by the movies can be explained after adding the social media data (sentimant and volume scores), thus we could anticipate that the social media data improves the forecast of Open.Box.Office. However, we cannot say that each of the social media data variables included has equal significance. For instance, the effect of variables related to sentiment score, both in weeks 1-3 before movie realase and in weeks 4-6 before movie release, as well as the effect of variable related to volume score in weeks 4-6 before release are not statistically significant. This can be shown through the confidence interval (0.99 and 0.95 level of significance) which includes both negative and positive values, as well as zero. Furthermore, along with the wide range of confidence interval goes the high level of p-value, hence our sample did not provide sufficient evidence to conclude that the effect exists.

Although the estimated effects of social media data variables are relatively high, especially for the sentiment score in weeks 4-6 before movie release, the overall analysis leads us to the conclusion that the greatest significant and prognostic power has the volume score, focusing on the period of one to three weeks before release. Its effect is positive with estimated value $\hat{\beta}_9 = 12.692$ and 95% confidence interval [4.8602, 20.5244], meaning that the expected increase in revenue due to one additional point in the volume score is in range of 4,860,200 to 20,524,400 dollars. Also, it is likely for the variable related to volume score in weeks 4-6 before release to have negative effect on the expected revenue, as the null hypothesis would be rejected at 0.1 level of significance. However, when using 0.05 and 0.01 level of significance, our sample would not provide sufficient evidence to conclude that the effects exist, thus its effect stays questionable.

```
kable(round(confint(lin.model1, level = 0.95), 4),
      caption="95% confidence interval of regression coefficients")
```

Table 12: 95% confidence interval of regression coefficients

	2.5 %	97.5 %
(Intercept)	-865.1040	4.7065
Budget	-0.0163	0.2818
Weeks	0.1167	1.2810
Screens	0.0058	0.0178
Animation1	-30.2431	-5.0079
Family1	-3.1371	22.2695
Adventure1	0.8009	17.0466
S_1_3	-462.5598	446.1958
S_4_6	-347.1166	549.8673
V_1_3	4.8602	20.5244
V_4_6	-14.6829	0.6432

```
kable(cbind("S_1_3" = round(lm_sum1$coef[8,4],4), "S_4_6" = round(lm_sum1$coef[9,4],4),
  "V_1_3" = round(lm_sum1$coef[10,4],4), "V_4_6" = round(lm_sum1$coef[11,4],3)),
  caption="P-values: Social Media Data")
```

Table 13: P-values: Social Media Data

S_1_3	S_4_6	V_1_3	V_4_6
0.9715	0.6542	0.0018	0.072

3 Simulation Study

3.1

```
set.seed(1)
b0 <- 2
b1 <- -1.5
iters <- c(10,100,1000)

models <- list()
for (i in seq_along(iters)) {
  x <- runif(iters[i], -3, 3)
  u <- rnorm(iters[i], 0, 2^2)
  y <- b0 + b1*x + u
  models[[i]] <- lm(y ~ x)
}
```

```
stargazer(models[[1]], models[[2]], models[[3]], header = FALSE,
          title="Simulated Simple OLS Regression Models")
```

Table 14: Simulated Simple OLS Regression Models

	Dependent variable:		
	y		
	(1)	(2)	(3)
x	−1.231	−1.670***	−1.449***
	(0.812)	(0.237)	(0.074)
Constant	2.263	1.751***	1.947***
	(1.481)	(0.371)	(0.130)
Observations	10	100	1,000
R^2	0.223	0.337	0.277
Adjusted R^2	0.126	0.330	0.276
Residual Std. Error	4.616 (df = 8)	3.708 (df = 98)	4.122 (df = 998)
F Statistic	2.295 (df = 1; 8)	49.785*** (df = 1; 98)	381.934*** (df = 1; 998)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 14 was created using stargazer v.5.2.2 by Hlavac (2018).

3.2

```
kable(round(confint(models[[1]]), 4),
      caption="Confidence interval of regression coefficients: N=10")
```

Table 15: Confidence interval of regression coefficients: N=10

	2.5 %	97.5 %
(Intercept)	-1.1521	5.6787
x	-3.1041	0.6427

```
kable(round(confint(models[[2]]), 4),
      caption="Confidence interval of regression coefficients: N=100")
```

Table 16: Confidence interval of regression coefficients: N=100

	2.5 %	97.5 %
(Intercept)	1.0143	2.4877
x	-2.1395	-1.2002

```
kable(round(confint(models[[3]]), 4),
      caption="Confidence interval of regression coefficients: N=1000")
```

Table 17: Confidence interval of regression coefficients: N=1000

	2.5 %	97.5 %
(Intercept)	1.6909	2.2025
x	-1.5944	-1.3035

Based on the tables 3-5, we see that the confidence interval is getting narrower as the sample size increases since $\hat{\beta}$ is a consistent estimator of β because the variance of $\hat{\beta}$ decreases as the sample size N increases which can be shown by:

$$\sigma_{\hat{\beta}} = \lim_{N \to \infty} \frac{\sigma}{\sqrt{N s_x^2}} \to 0$$

This implies that the confidence interval gets narrower as the sample size increases since the bounds of the confidence interval approach the $\hat{\beta}$ which means that $\hat{\beta}$ converges to the real β "in probability":

$$c_{lower} = \lim_{\sigma_{\hat{\beta}} \to 0} \hat{\beta} - c_{-\frac{\alpha}{2}} \sigma_{\hat{\beta}} \to \beta$$

$$c_{upper} = \lim_{\sigma_{\hat{\beta}} \to 0} \hat{\beta} + c_{-\frac{\alpha}{2}} \sigma_{\hat{\beta}} \to \beta$$

3.3

a) σ^2 **varies**:

```
set.seed(1)
u <- c()
models.u <- list()
for (i in seq_along(iters)) {
  x <- runif(iters[i], -3, 3)
  u.var <- runif(iters[i], 0, 4)
  for (j in seq_along(u.var)) {
    u[j] <- rnorm(1, 0, u.var[j])
  }
  y <- b0 + b1*x + u
  models.u[[i]] <- lm(y ~ x)
}
```

```
stargazer(models.u[[1]], models.u[[2]], models.u[[3]], header = FALSE,
          title="Simulated Simple OLS Regression Models")
```

Table 18: Simulated Simple OLS Regression Models

	Dependent variable:		
	y		
	(1)	(2)	(3)
x	−2.116***	−1.501***	−1.542***
	(0.344)	(0.151)	(0.043)
Constant	2.848***	1.936***	1.895***
	(0.627)	(0.243)	(0.076)
Observations	10	100	1,000
R^2	0.826	0.501	0.559
Adjusted R^2	0.804	0.495	0.558
Residual Std. Error	1.954 (df = 8)	2.427 (df = 98)	2.415 (df = 998)
F Statistic	37.872*** (df = 1; 8)	98.210*** (df = 1; 98)	1,263.252*** (df = 1; 998)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 18 was created using stargazer v.5.2.2 by Hlavac (2018).

The confidence interval reads: $[\hat{\beta}_j - c_{1-\alpha/2} sd(\hat{\beta}_j|X), \hat{\beta}_j + c_{1-\alpha/2} sd(\hat{\beta}_j|X)$. Furthermore, the standard deviation $sd(\hat{\beta}_j|X)$ can be computed using the following formula: $sd(\hat{\beta}_j|X) = \frac{\sigma}{\sqrt{N s_{xj}^2 (1-R_j^2)}}$. Hence, we can notice that the higher value of the variance of the error terms σ^2 will result in a wider confidence interval.

```
kable(round(confint(models.u[[1]]), 4),
      caption="Confidence interval of regression coefficients when the variance of the error term varies
```

Table 19: Confidence interval of regression coefficients when the variance of the error term varies: N=10

	2.5 %	97.5 %
(Intercept)	1.4027	4.2940
x	-2.9091	-1.3232

```
kable(round(confint(models.u[[2]]), 4),
      caption="Confidence interval of regression coefficients when the variance of the error term varies
```

Table 20: Confidence interval of regression coefficients when the variance of the error term varies: N=100

	2.5 %	97.5 %
(Intercept)	1.4542	2.4186
x	-1.8020	-1.2007

```
kable(round(confint(models.u[[3]]), 4),
      caption="Confidence interval of regression coefficients when the variance of the error term varies
```

Table 21: Confidence interval of regression coefficients when the variance of the error term varies: N=1000

	2.5 %	97.5 %
(Intercept)	1.7448	2.0446
x	-1.6268	-1.4565

b) **The level of confidence varies** ($CL = 1 - \alpha$):

```
kable(round(confint(models[[3]], level=0.9), 4),
      caption="Confidence interval of regression coefficients when the level of confidence varies: N=10
```

Table 22: Confidence interval of regression coefficients when the level of confidence varies: N=1000, CL=90%

	5 %	95 %
(Intercept)	1.7321	2.1613
x	-1.5710	-1.3269

```
kable(round(confint(models[[3]], level=0.95), 4),
      caption="Confidence interval of regression coefficients when the level of confidence varies: N=10
```

Table 23: Confidence interval of regression coefficients when the level of confidence varies: N=1000, CL=95%

	2.5 %	97.5 %
(Intercept)	1.6909	2.2025
x	-1.5944	-1.3035

```
kable(round(confint(models[[3]], level=0.99), 4),
      caption="Confidence interval of regression coefficients when the level of confidence varies: N=10
```

Table 24: Confidence interval of regression coefficients when the level of confidence varies: N=1000, CL=99%

	0.5 %	99.5 %
(Intercept)	1.6103	2.2831
x	-1.6403	-1.2576

Based on the tables 22-24, we can see that as the confidence level increases, the confidence intervals gets narrower.

References

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.