Ec 122 - HW3

Ben Juarez

10/14/2021

A

1) This article investigates the "absenteeism" on an undergraduate level such that the main focus is regarding how prominent absenteeism is along with its effect on learning as well as how it should be dealt with.
2) The main results of this article deal with the first two points such that absenteeism was quite "rampant" during the time of this article (absentee rate ranging around 30%) and that the evidence strongly suggests that attendance has a substantial impact on learning (essentially a difference between a whole grade letter).
3) With the data samples from various schools, one method running a regression of different types of courses on the absenteeism of the course. Another regression was run using the appropriate parameters that determined that higher enrollment correlates to an increase in absenteeism. Furthermore, another prominent method was the regression of performance on the fraction of lectures attended which gives insight into the significant relation between attendance and performance, as mentioned. While these were the main methods, its worth noting that other measures were taken to address the impact of motivation in various ways.
4) To be honest, I do believe these results are still accurate and perhaps more relevant today. With less and less undergraduate classes having required attendance, especially in a remote environment, there does seem to be a distinction between the performance of students who attend class and those who don't. I think an interesting perspective that is relevant to this article is more about the reasons and explanations behind missing class. From my experience, for classes with optional attendance, there are many times when it becomes too "tempting" to skip class because there are so many other things going on with other classes, etc. that it almost feels more "efficient" to miss certain lectures. However, in these moments, its easy to forget the importance of attending lecture. I also think attending lecture plays a large role in keeping a structured routine, which I beleive positively improves performance.

C

HGL 3.1

Using regression output for food expenditure model shown in Figure 2.9.

Part a

Let us construct a 95% interval estimate for β_1 as follows. We have $b_1 = 83.416, t_c = t_{(1-0.025=0.975, N-2=38)} = 2.024, se(b_1) = 43.41$. Let us construct the estimate as follows:

$$\implies b_1 \pm t_c \times se(b_1) = 83.416 \pm 2.024 \times 43.41 = (-4.44584, 171.2778)$$

So, we are 95% confident that β_1 lies within this range.

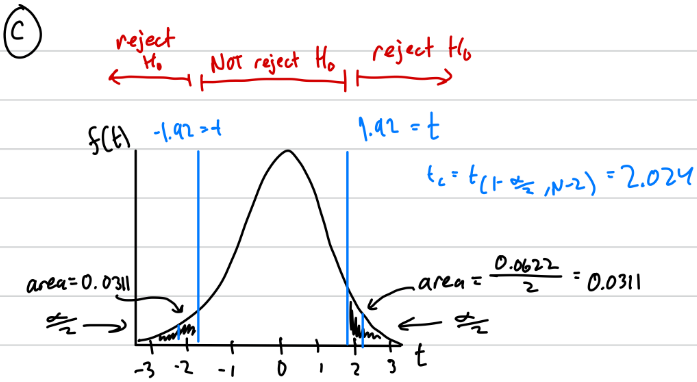
Part b

Let us test $H_0 : \beta_1 = 0 = H_1 : \beta_1 \neq 0$ as follows (5% level of significance) such that we do not reject if $-2.024 < t < 2.024$:

$$\implies t = \frac{b_1 - \beta_1}{se(b_1)} = \frac{83.416}{43.410} = 1.92 < t_c = t_{(0.975,38)} = 2.024 \implies H_0 \text{ not rejected}$$

Part c

c



$\implies$ p-value = 0.0622 represents area under t-distribution to left/right of $t = 1.92$. As shown, each area is $\frac{0.0622}{2}$. We also know that $\frac{\alpha}{2} = 0.025$, so using p-value = 0.0622, we could answer (b) b/c $\frac{\alpha}{2} < \frac{p}{2} = 0.025 < 0.0311$ is equivalent to t falling in the domain between -1.92 and 1.92 $(-t_c < t < t_c)$. Clearly, if $\frac{p}{2} \leq 0.025$, then t would not satisfy $-t_c = 2.024 < t < t_c = 2.024$, meaning H_0 is rejected.

Part d

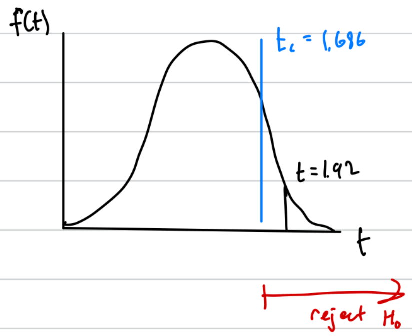
Let us test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$ as follows (5% level of significance) such that we do not reject if $t \geq 1.686$:

$$\implies t = \frac{b_1 - \beta_1}{se(b_1)} = \frac{83.416}{43.410} = 1.92 > t_c = t_{(0.95,38)} = 1.686 \implies H_0 \text{ rejected}$$

See attached sketch. Since we reject H_0, we conclude that $\beta_1 > 0$.

(d) $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0 \implies t = 1.92$ from part (b) * One-tail test w/ alternative (>)

$\implies t_c = t_{(1-\alpha = 1-0.5,\, N-2=38)} = 1.686 \implies t = 1.92 > t_c = 1.686 \implies$ reject H_0, accept $H_1 \implies \beta_1 > 0$



* $t > 1.686 \iff p < 0.05$

$p(t \geq 1.92) = 0.0322$

Part e

The level of significance (α) is essentially the probability of a true null hypothesis being rejected while the level of confidence ($1 - \alpha$) is related to how likely an interval estimator is to contain the true parameter. Both measures are utilized in hypothesis tests and help to determine the rejection/non-rejection region of the null hypothesis.

Part f

This is false. This is an important distinction with confidence intervals and levels of significance. The confidence is in the procedure used to construct the interval estimate as opposed to any single estimate from a sample. With part (d), we used a 5% level of significance, which means there's a 5% chance of a true null hypothesis being rejected. This level of significance does not imply 95% confidence since interval estimates are distinct from hypothesis tests in this sense.

HGL 3.7

```r
setwd("~/Desktop/R/")
capm4 = read.table("dat/capm4.dat")
colnames(capm4) = c("date", "dis", "ge", "gm", "ibm", "msft", "xom", "mkt",
                    "riskfree")
```

Part a

With 5% level of significance, we want to test the hypothesis that each stock's "beta" value is 1 against the alternative that is is not equal to 1: $H_0 : \beta_j = 1$ vs. $H_1 : \beta_j \neq 1$ The economic interpretation of this is essentially testing the relative risk of the company's stock relative to the market portfolio since the beta value measures the stock risk relative to the market portfolio.

We also have $t_c = t_{(1-\frac{\alpha}{2}=0.975, N-2=130)} = 1.978$. So, we reject if t does not fall between -1.978 and 1.978. Let us calculate t as follows using the following formula: $t = \frac{b_j - 1}{se(b_j)}$

```r
library("plotrix")
library("knitr")
companies = c("Disney", "GE", "GM", "IBM", "Microsoft", "Exxon")
```

```r
c = capm4$mkt - capm4$riskfree

dis = capm4$dis - capm4$riskfree
ge = capm4$ge - capm4$riskfree
gm = capm4$gm - capm4$riskfree
ibm = capm4$ibm - capm4$riskfree
msft = capm4$msft - capm4$riskfree
xom = capm4$xom - capm4$riskfree
parta = data.frame(matrix(nrow = 2, ncol = 6, dimnames = list(c("beta","t"), companies)))

dis_lm = lm(dis ~ c)
ge_lm = lm(ge ~ c)
gm_lm = lm(gm ~ c)
ibm_lm = lm(ibm ~ c)
msft_lm = lm(msft ~ c)
xom_lm = lm(xom ~ c)

parta[1,1] = coef(dis_lm)[2]
parta[1,2] = coef(ge_lm)[2]
parta[1,3] = coef(gm_lm)[2]
parta[1,4] = coef(ibm_lm)[2]
parta[1,5] = coef(msft_lm)[2]
parta[1,6] = coef(xom_lm)[2]

se_dis = summary(dis_lm)$coefficients[2,2]
se_ge = summary(ge_lm)$coefficients[2,2]
se_gm = summary(gm_lm)$coefficients[2,2]
se_ibm = summary(ibm_lm)$coefficients[2,2]
se_msft = summary(msft_lm)$coefficients[2,2]
se_xom = summary(xom_lm)$coefficients[2,2]

parta[2, 1] = (parta[1, 1] - 1) / se_dis
parta[2, 2] = (parta[1, 2] - 1) / se_ge
parta[2, 3] = (parta[1, 3] - 1) / se_gm
parta[2, 4] = (parta[1, 4] - 1) / se_ibm
parta[2, 5] = (parta[1, 5] - 1) / se_msft
parta[2, 6] = (parta[1, 6] - 1) / se_xom

kable(parta)
```

	Disney	GE	GM	IBM	Microsoft	Exxon
beta	0.8978381	0.8992599	1.261411	1.188208	1.318947	0.413969
t	-0.8263722	-1.0198258	1.292683	1.488605	1.983622	-6.532269

As we can see, we fail to reject H_0 for Disney, GE, GM, IBM, and we reject H_0 for Microsoft and Exxon. This implies that the volatility of the stocks for Microsoft and Exxon does not closely match the volatility of the market portfolio.

Part b

We want to test $H_0 : \beta_j \geq 1$ vs $H_1 : \beta_j < 1$ at 5% level of significance for Exxon. We have $t_c = t_{(0.05, N-2=130)} = -1.657$ so we reject H_0 if $t \leq t_c$. Let us calculate t:

```
(parta[1, 6] - 1) / se_xom
```

```
## [1] -6.532269
```

Since $t = -6.53 < t_c$, we reject H_0. This implies that the beta estimate for Exxon is less than 1, meaning that their stock is less "risky" relative to the market portfolio. This could be interpreted as being defensive.

Part c

We want to test $H_0 : \beta_j \leq 1$ vs $H_1 : \beta_j > 1$ at 5% level of significance for Microsoft. We have $t_c = t_{(0.95, N-2=130)} = 1.657$ so we reject H_0 if $t \geq t_c$. Let us calculate t:

```
(parta[1, 5] - 1) / se_msft
```

```
## [1] 1.983622
```

Since $t = 1.98 > t_c$, we reject H_0. This implies that the beta estimate for Microsoft is greater than 1, meaning that their stock is more "risky" relative to the market portfolio. This could be interpreted as being aggressive.

Part d

Let us construct a 95% interval estimate for Microsoft's β_j using

$$b_j \pm t_{(0.975,130)} \times se(b_j)$$

```
parta[1, 5] + 1.9784 * se_msft
```

```
## [1] 1.637054
```

```
parta[1, 5] - 1.9784 * se_msft
```

```
## [1] 1.00084
```

Hello (insert investor name), I hear you want the inside scoop on Microsoft? My boy Bill Gates just gave me all the details. Let me break it down for you. Looking at the numbers, we are 95% confident that β_j for Microsoft falls between 1.001 and 1.637. Again, there's a slight chance that this value falls outside of this interval, but the odds are greatly against it. We know this isn't a super narrow window for determining β_j precisely with this type of analysis, but one thing's for sure: Microsoft's stock volatility sure is higher than the market with these numbers, making it more aggressive than some of the other companies you might want to invest in.

Part e

With 5% level of significance, we want to test the hypothesis that each stock's "alpha" value is 0 against the alternative that is is not equal to 0: $H_0 : \alpha_j = 0$ vs. $H_1 : \alpha_j \neq 0$

We also have $t_c = t_{(1-\frac{\alpha}{2}=0.975, N-2=130)} = 1.978$. So, we reject if t does not fall between -1.978 and 1.978. Let us calculate t as follows using the following formula: $t = \frac{a_j}{se(a_j)}$

```
parte = data.frame(matrix(nrow = 2, ncol = 6, dimnames = list(c("alpha","t"), companies)))

parte[1,1] = coef(dis_lm)[1]
parte[1,2] = coef(ge_lm)[1]
parte[1,3] = coef(gm_lm)[1]
parte[1,4] = coef(ibm_lm)[1]
parte[1,5] = coef(msft_lm)[1]
```

```r
parte[1,6] = coef(xom_lm)[1]

se_dis_a = summary(dis_lm)$coefficients[1,2]
se_ge_a = summary(ge_lm)$coefficients[1,2]
se_gm_a = summary(gm_lm)$coefficients[1,2]
se_ibm_a = summary(ibm_lm)$coefficients[1,2]
se_msft_a = summary(msft_lm)$coefficients[1,2]
se_xom_a = summary(xom_lm)$coefficients[1,2]

parte[2, 1] = (parte[1, 1]) / se_dis_a
parte[2, 2] = (parte[1, 2]) / se_ge_a
parte[2, 3] = (parte[1, 3]) / se_gm_a
parte[2, 4] = (parte[1, 4]) / se_ibm_a
parte[2, 5] = (parte[1, 5]) / se_msft_a
parte[2, 6] = (parte[1, 6]) / se_xom_a


for (i in 1:6) {
  for (j in 1:2) {
    parte[j,i] = signif(parte[j,i],4)
  }
}

kable(parte)
```

	Disney	GE	GM	IBM	Microsoft	Exxon
alpha	-0.001149	-0.001167	-0.01155	0.005851	0.006098	0.00788
t	-0.193000	-0.245200	-1.18500	0.960600	0.787100	1.82300

Since each t value is within the interval from -1.978 to 1.978, we do not reject H_0 for any of the company's stocks. This implies that the data is in line with the theory such that the intercepts should be 0.

HGL 3.9

```r
setwd("~/Desktop/R/")
fair4 = read.table("dat/fair4.dat")
colnames(fair4) = c("year", "vote", "party", "person", "duration", "war", "growth",
                    "inflation", "goodnews")
```

Part a

Using the regression model $VOTE = \beta_1 + \beta_2 GROWTH + e$ at a 5% significance level, we will test the null hypothesis that economic growth has no effect on the percentage vote earned by the incumbent part. Let us select the alternative $\beta_2 > 0$ since it makes sense that economic growth would positively affect vote in this case.

Let us test $H_0 : \beta_2 = 0$ vs $H_2 : \beta_2 > 0$. We have $t_c = t_{(1-\alpha=0.95, N-2=22)} = 1.717$. So we reject if $t \geq 1.717$. Let us calculate $t = \frac{b_2}{se(b_2)}$:

```r
growth = subset(fair4, year > 1915)$growth
vote = subset(fair4, year > 1915)$vote
```

```
growth_lm = lm(vote ~ growth)
b2 = coef(growth_lm)[2]
se_b2 = summary(growth_lm)$coefficients[2,2]
t = b2/se_b2
print(t)
```

```
##   growth
## 4.871265
```

So, we have $t = 4.87 > t_c = 1.717$, meaning we reject H_0. This implies we accept $\beta_2 > 0$ which tells us that economic growth does positively affect the vote percentage.

Part b

Using the model from part (a), let us construct a 95% interval estimate for β_2 as follows:

$$b_2 \pm t_{(0.975,22)} se(b_2)$$

```
i1 = b2 + 2.0739 * se_b2
i2 = b2 - 2.0739 * se_b2
unname(i1)
```

```
## [1] 1.263131
```

```
unname(i2)
```

```
## [1] 0.5087621
```

With 95% confidence, we can say that β_2 falls between 0.51 and 1.26 which implies that the percentage vote will increase by an amount in this range with a 1% increase in the rate of growth.

Part c

Using the regression model $VOTE = \beta_1 + \beta_2 INFLATION + e$, we will test the null hypothesis that inflation has no effect on percentage vote earned by the incumbent part. Let us select the alternative $\beta_2 < 0$ since it makes sense that inflation would affect vote in this case.

Let us test $H_0 : \beta_2 = 0$ vs $H_2 : \beta_2 > 0$. We have $t_c = t_{(\alpha=0.05, N-2=22)} = -1.717$. So we reject if $t \leq -1.717$. Let us calculate $t = \frac{b_2}{se(b_2)}$:

```
inflation = subset(fair4, year > 1915)$inflation
inflation_lm = lm(vote ~ inflation)
b2_i = coef(inflation_lm)[2]
se_b2_i = summary(inflation_lm)$coefficients[2,2]
t = b2_i/se_b2_i
print(t)
```

```
##  inflation
## -0.7406087
```

So, we have $t = -0.74 > -1.717$, meaning we do not reject H_0. Thus, the data does not imply that inflation negatively affects the vote percentage.

Part d

Using the model from part (c), let us construct a 95% interval estimate for β_2 as follows:

$$b_2 \pm t_{(0.975,22)} se(b_2)$$

```
i11 = b2_i + 2.0739 * se_b2_i
i22 = b2_i - 2.0739 * se_b2_i
unname(i11)
```

```
## [1] 0.7998793
```

```
unname(i22)
```

```
## [1] -1.688504
```

With 95% confidence, we can say that β_2 falls between -1.69 and 0.80 which implies that the percentage vote would increase/decrease by an amount in this range with a 1% increase in the rate of inflation.

Part e

We will test the null hypothesis (5% level of significance) that if $INFLATION = 0$ the expected vote in favor of the incumbent party is 50% or more. The alternative being less than 50% logically.

Let us test $H_0 : \beta_1 \geq 50$ vs $H_1 : \beta_1 < 50$. We have $t_c = t_{(\alpha=0.05,N-2=22)} = -1.717$. So we reject if $t \leq -1.717$. Let us calculate $t = \frac{b_2-50}{se(b_2)}$:

```
b1 = coef(inflation_lm)[1]
se_b1 = summary(inflation_lm)$coefficients[1,2]
t = (b1 - 50) / se_b1
print(unname(t))
```

```
## [1] 1.514558
```

So, we have $t = 1.51 > t_c = -1.717$, mean we do not reject H_0. Thus, the data implies that if inflation is 0, then the expected incumbent vote is greater than or equal to 50%. This is interesting because it makes sense that low inflation generally would reflect positively on the incumbent, so much so that half of the vote goes to the incumbent. It is clear that zero inflation is a very strong factor.

Part f

Let us construct a 95% interval estimate of the expected vote in favor of the incumbent party if $INFLATION = 2\%$ as follows ($E(VOTE|INFLATION = 0.02)$):

$$(b_1 + b_2 \cdot 0.02) \pm t_{(0.975,22)} \cdot se(b_1 + b_2 \cdot 0.02)$$

```
cov_matrix = vcov(inflation_lm)
c_2 = 0.02
low = (b1 + b2 * c_2) - 2.0739 * sqrt(cov_matrix[1,1] + c_2^2 * cov_matrix[2,2] + 2 *
                                      c_2 * cov_matrix[1,2])
high = (b1 + b2 * c_2) + 2.0739 * sqrt(cov_matrix[1,1] + c_2^2 * cov_matrix[2,2] + 2 * c_2
                                     * cov_matrix[1,2])
print(unname(low))
```

```
## [1] 48.77871
```

```r
print(unname(high))
```

```
## [1] 58.07221
```

Thus, the 95% interval is between 48.78 and 58.07, meaning that if the inflation rate is 2%, the expected incumbent vote percentage should fall between this range.