

Applied Stats I: Exam 2

Ariana Alves Antunes

Due: December 9, 2022

Question 1

- (a) Interpret the coefficients for GDP and Democracy.

```
1 # a) It seems that the estimate is high correlated between GDP and
    democracy, due to the small difference. And, the variation in FDI when
    a country GDP is increasing in one unit (1$) the FDI investment is -3
    increase in GDP and countries democracy -2.99.
2 # the different is very small, 0.1, but the
```

- (b) The author claims that she 'cannot reject the null hypothesis that GDP has no effect on FDI ($H_0 : \beta_{GDP} = 0$)'. Using the coefficient estimate and the standard error for GDP construct a 95% confidence interval for the effect of GDP on FDI. Based on the confidence interval, do you agree with the author? Explain your answer.

```
1 -3 + 1.96*(0.00013)
2 -3 - 1.96*(0.00013)
3 [-2.999745, -3.000255]
```

Negative association between the FDI of a countries investment and the GDP number.

- (c) Calculate the difference in predicted FDI between low and high values of Education for non-democratic countries holding GDP constant at its sample mean. Use 24860.42 as the mean of GDP and use +/- one standard deviation around the mean of Education (from 10.96 to 13.02) for low and high values of Education respectively.

```
1 Sample.mean <- -3
2 gdp.mean <- 24860.42
3 sd.education <- 10.96 ~ 13.02
4
5 x <- rbinom(10.96, 13.02)
6
7 var(x)
```

Question 2

It could cause bias to different countries, cultures and factors associated with inequalities, for example GDP and standards of living varying between what wealth is and how are those counted. We could access a bigger study with divided countries and it's estimated, and change the monetary value by GDP.

Question 3

- (a) Correlation between the estimated coefficients as intercept is 0.25(.090), relates to increase of every 0.73 in well depth, and -1.01 distance of 100k to the closest commercial factory.
- (b) Yes, it vary. Additive affects, since the same effect for each well depth. The appropriate test should be reporting the effect sizes and confidence intervals, since these would convey the relative importance and magnitude of the effect.

(c)

```
2 # c) Compute the average difference
3 deep_well <- 1
4 dist100 <- c(0.37, 2.11)
5
6 avg <- list(deep_well, dist100, deep_well * 2, dist100 * 3)
7 avg
8
9 cat("After finding the average of each element of the list", "\n")
10 sapply(avg, mean)
11 # > sapply(avg, mean)
12 # [1] 1.00 1.24 2.00 3.72
```

Question 4

- (a) first question $E(y) = -18.1368 + 2.2980 - 4.0716$
- (b) second question
 $E(y) = 2.2980x + 10x$
- (c) third question Ram lambs has the highest Fatness index for every weight, with -4.0716 every 2.29 weight.

Question 5

- (a) a) Partial F-test : Describes how statistic significant is the regression/model, and test if "variable" is useful in the model at all. The number further from 1 is actually better,

but we can check the p-value on the side to check it as well. Larger F values give stronger evidence against H_0 .

- (b) Residuals: Also known as the prediction errors, in an observation, the difference between an observed value and the predicted value of the response variable, $y - \hat{y}$, is called the residual.
- (c) Categorical data/dummy variables: Data that can take only a specific set of values representing a set of possible categories, and dummy variables are used to avoid multicollinearity when doing a regression model. For example we can re code factor data to be used in the regression.
- (d) Constituent term:

Question 6

- (a) 2. Additive (salary = age + education) regression model
- (b) 1. True
- (c) 3. minimize the residual sum of squares
- (d) 4. QQ plot of residuals