

1. (15 points) Please answer the following; you might find it useful to make a sketch.

- a. For a Normal Distribution that has mean 1 and standard deviation 6.5, what is the area to the left of 1.65?

Area to the left of 1.65 is 0.5398.

- b. For a Normal Distribution that has mean 8 and standard deviation 2.7, what is the area in both tails farther from the mean than 13.67?

The area in both tails farther from the mean than 13.67 is less than 0.0001.

- c. For a Normal Distribution that has mean -11 and standard deviation 4, what is the area in both tails farther from the mean than -5.4?

The area in both tails farther from the mean than -5.4 is 0.1770

- d. For a Normal Distribution that has mean 14 and standard deviation 7.4 what two values leave probability 0.158 in both tails?

The two values are 21.4201 and 6.57995

- e. A regression coefficient is estimated to be equal to 6.56 with standard error 4.1; there are 24 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?

The P-value is 0.1227

- f. A regression coefficient is estimated to be equal to -0.24 with standard error 0.4; there are 4 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?

The P-value is 0.5808

2. (10 points) As we consider, “did everything change after March 2020?” look at crude oil prices. The average daily return of crude oil was 0.000145 with standard deviation of 0.0213 in 289 days before March 1. Average daily return after that date was -0.0210 with standard deviation of 0.271 in 174 days after. Is there a statistically significant difference

in the mean? Calculate t-stat and p-value for the test against no difference in daily returns.

3. (10 points) In good news, there was information about vaccine trials. Consider (these are not quite the actual data but a simplified version) looking at 2 groups, each with 10,000 people. In the control group who did not get the vaccine there were 90 infections. In the test group that did get the vaccine there were 15 infections. Calculate the t-stat and p-value for the test against no difference in infection rates between groups.

The next questions will use the PUMS ACS 2017 NY data that we've been using so often in class. We'll consider people's decisions about usual hours worked, given by the variable UHRSWORK. There are 3 broad categories: people who are not in the labor force, those working part time (UHRSWORK > 0 and < 35) and those working full time (>= 35).

4. (10 points) Create a subgroup of the sample, that makes sense as we focus on the decision of whether to work full time or part time. Explain your rationale.

```
attach(acs2017_ny)
use_varb <- (AGE >= 25) & (AGE <= 55) & (FAMSIZE>1) & (LABFORCE == 2) &
(UHRSWORK>0 & UHRSWORK<35) & (UHRSWORK >= 35)
dat_use <- subset(acs2017_ny,use_varb)
detach()
```

I created this subset with people in the labor force who are above or at the age of 35 up to people who are under or 55 who have a family size of greater than 1 because those people probably have to earn more in order to support their families.

5. (10 points) Within this subgroup, test if there is a difference between men and women. Calculate a t-stat and p-value for the test of no difference between the groups. Provide simple statistics for a few other relevant categories.
When I changed the subset to female, there was a slight difference in the tstat and p value as shown in the r data.
6. (20 points) Estimate a simple OLS model for hours worked, within your subsample.
 - a. Explain what variables you choose to use as predictors. Do they seem exogenous? Consider whether polynomials in Age are important or interactions with dummy variables.
 - b. Do your estimates seem plausible? Are the estimates each statistically significant?
 - c. Construct a joint test of whether a reasonable set of coefficients (such as age polynomials, or education dummies) are all zero.
 - d. What are the predicted probabilities for a few particular groups?
 - e. How many Type I and Type II errors are made by the model?
7. (20 points) Estimate a simple logit model, for the outcome variable Work_Fulltime <- (UHRSWORK >= 35), within your subsample.

- a. Explain what variables you choose to use as predictors. Do they seem exogenous? Consider whether polynomials in Age are important or interactions with dummy variables.
 - b. Do your estimates seem plausible? Are the estimates each statistically significant?
 - c. Construct a joint test of whether a reasonable set of coefficients (such as age polynomials, or education dummies) are all zero.
 - d. What are the predicted probabilities for a few particular groups?
 - e. How many Type I and Type II errors are made by the model?
8. (25 points) Estimate one or more additional models with other methods (not OLS or logit) to predict hours or if a person works full time. Explain as in previous. Compare the different models and make judgements about strengths and weaknesses of each.

All of the work on this exam is my own, answered honestly as rules state.

Name: Fareha Sameen

Date: November 23, 2020