Problem Set 7

How long did the following elements take:

- R portion of assignment:
- Completing the assignment:
- Typing the assignment (LaTeX/RMD):

Exercises

1) (2 points) You want to know whether adults in your country think the ideal number of children is equal to 2, or higher or lower than that.
   (a) Define notation and state the null and alternative hypotheses for studying this. For responses in a recent GSS to the question "What do you think is the ideal number of children to have?" software shows results:

   Test of mu = 2.0 vs. mu not = 2.0

Variable	n	Mean	StDev	SE Mean	T	P-value
Children	302	2.29	0.953	0.055	5.272	0.0000

   Report the test statistic value, and show how it was obtained from other values reported in the table.
   (b) Explain what the P-value represents, and interpret its value.

2) (3 points) Same-sex marriage was legalized across Canada by the Civil Marriage Act enacted in 2005. Is this supported by a majority, or a minority, of the Canadian population? In an Ipsos Global poll conducted for *Reuter News* in May 2013, of 1000 Canadians that asked whether legalization should stand or be repealed 63% supported legalization. Let π denote the population proportion of Canadian adults who support legalization. For testing $H_0 : \pi = 0.50$ against $H_a : \pi \neq 0.50$
   (a) Find the standard error and interpret.
   (b) Find the test statistic and interpret.
   (c) find the P-value, and interpret in context.

3) (3 points) Washington and Smith separately conduct studies to test $H_0 : \pi = 0.50$ against $H_a : \pi \neq 0.50$, each with n = 400. Washington gets $\hat{\pi} = 220/400 = 0.550$. Smith gets $\hat{\pi} = 217/400 = 0.5425$
   (a) (0.5 points) Show that z=2.00 and P-value of 0.046 for Washington.
   (b) (0.5 points) What z-statistic and p-value do you get for Smith?
   (c) (1 point) Using $\alpha = 0.05$, indicate in each case whether the results is statistically significant. Interpret.
   (d) (1 point) Use this example to explain why important information is lost by reporting the result as significant or not without reporting the p-value.

4) (2 points) A report by the Collaborative on Academic Careers in Higher Education indicated that there is a notable gap between female and male academics in their confidence that tenure rules are clear, with men feeling more confident. The 4500 faculty members in the survey were asked to evaluate policies on a scale of 1 to 5 (very unclear to very clear). The mean response about the criteria for tenure was 3.51 for females and 3.55 for males, which was indicated to meet the test for statistical significance. Use this study to explain the distinction between statistical significance and practical significance.

5) (2 points) You are designing a survey to estimate the gender gap: the difference in support for a candidate among men and women. Assuming the respondents are a simple random sample of the voting population, how many people do you need to poll so that the standard error is less than 5 percentage points?

6) (1 point) A research study conducts 75 significance tests. Of these, three are significant at the 0.05 level. The authors write a report stressing only the three "significant" results, not mentioning the other 72 tests that were "not significant." Explain what is misleading about their report.

7) (3 points) Application: Consider the article linked here: https://www.vox.com/policy-and-politics/2017/10/12/16464188/john-roberts-sociological-gobbledygook-eduardo-bonilla-silva-gerrymandering In the article, the author states "The Supreme Court is currently considering a landmark case challenging partisan gerrymandering, specifically Wisconsin Republicans' efforts to draw state assembly districts so as to firmly entrench their majority. At the heart of the case is a concept called the 'efficiency gap,' a simple number that political scientist Eric McGhee and law professor Nicholas Stephanopoulos have devised to measure how much a given district map favors one party over the other. If the number gets too high, it's an indication that one party has rigged the game to ensure they keep getting reelected. . . . Chief Justice John Roberts . . . dismissed the concept as 'sociological gobbledygook' in oral arguments for the case."

   (a) Make your case for whether or not what you've been doing the last seven or so weeks has, in fact, been 'goobbeldygook'. Be sure to reference what statistical analysis does and does NOT permit us to understand.

   (b) The issue at the heart of the case pertains to whether one party has been benefitting from unfairly-drawn district boundaries. There is a proposal to use an 'efficiency gap' to determine if one party is causing the other to waste votes beyond an 'acceptable' threshold (more detail here, if curious: https://www.nytimes.com/interactive/2017/10/03/upshot/how-the-new-math-of-gerrymandering-works-supreme-court.html). Explain how the skills from the last two weeks (significance testing and confidence intervals) might be applied to evaluate whether or not states violate an efficiency gap threshold of 0.07. You can either do this hypothetically (for any given state) or you can choose to discuss the case of Maryland (efficiency gap is 0.105).

R

1) (6 points) For this component, you'll practice writing a function and then you will download a new-to-you dataset as a means to practice and enhance your skills. First, download your dataset: https://qog.pol.gu.se/data/datadownloads/qogstandarddata. You can choose either series (cross-section or time). *This section is graded based on completion.*

   (a) Write a function that will enable you to perform a t test and check your response for question 1.

   (b) Use the codebook to find a continuous or categorical variable you are interested in learning more about. Install the package 'psych' and use the command 'describe' to describe your variable. Compare this output to 'summarize'.

   (c) Create a graph of your chosen variable.

   (d) Create a graph comparing your variable to a similar variable (use the codebook and your best judgement) to identify said variable.

   (e) Provide a summary table of your recoded variable to the original.

   (f) Create a recoded version of one of these two variables. Explain the previous coding, your new coding and a justification for why you made the choice you did. For example, continuous to

dichotomous because the distribution appears, ...

(g) (Optional) Use a package that allows you to export tables into a LaTeX-friendly output, like outreg (install package 'outreg') to export two things: (1) a tabulation of your new variable and (2) your previous plot. This is handy for inserting graphics and tables into a document without needing to recompile! Include which package you used here.

Additional (BONUS/NOT GRADED) Questions

1) Explain the difference between one-sided and two-sided alternative hypotheses, and explain how this affects calculation of the P-value.

2) Explain why the terminology "do not reject H_0" is preferable to "accept H_0."

3) The P-value for a test about a mean with n = 25 is P = 0.05.
   (a) Find the t test statistic value that has this P-value for
      i. $H_0 : \mu \neq 0$,
      ii. $H_a : \mu > 0$,
      iii. $H_a : \mu < 0$.

4) (4 points) The authorship of an old document is in doubt. A historian hypothesizes that the author was a journalist named Jacalyn Levine. Upon a thorough investigation of Levine's known works, it is observed that one unusual feature of her writing was that she consistently began 6% of her sentences with the word *whereas*. To test the historian's hypothesis, it is decided to count the number of sentences in the disputed document that begin with *whereas*. Out of 300 sentences, none do. Let π denote the probability that any one sentence written by the unknown author of the document begins with *whereas*.
   (a) What is the null hypothesis?
   (b) What is the alternative hypothesis?
   (c) What assumptions are needed for your conclusion to be valid?
   (d) What would Type I error be here?

5) A multiple choice test has four options. Assume that a student taking this question either knows the answer or does a pure guess. A random sample of 100 students take the item and 60% Give an estimate and 95% confidence interval for the percentage in the population who knows the answer.

6) The 49 students in a class at the University of Florida made blinded evaluations of pairs of cola drinks. For the 49 comparisons of Coke and Pepsi, Coke was preferred 29 times. In the population that this sample represents, is this strong evidence that majority prefers one of the drinks? Refer to the following software output:

Test of parameter = 0.60 vs. not = 0.50

n	Sample prop	95% CI	z-value	P-value
49	0.5918	(0.454, 0.729)	1.286	0.1985

7) Your friend plans to survey students in your college to study whether a majority feel that the legal age for drinking alcohol should be reduced. He has never studied statistics. How would you explain the concepts of:

   (a) Null and alternative hypotheses

   (b) P-value

   (c) α-level

   (d) Type II error

8) The World Values Survey asked, “Indicate the importance in your life of religion.” Of the people sampled in the Netherlands, Stata reports the following output for the proportion who answered, “very important” or “rather important” (instead of “not very important” or “not at all important”):

One sample test of proportion:
Number of obs = 1902

Variable	Mean	Std. err.
religion	.25	.0099288

p = proportion(religion)

z = 21.8060

Ho: p = 0.5

Ha: p<0.5

Pr(Z < z) = 0.0000

Ha p! = 0.5

Pr(Z > z) = 1.0000

1. Explain how to interpret all results shown.
2. By contrast, the results for the United States had a sample proportion of 0.68 who answered "very important" or "rather important." Here are the test results:

   Ha: p < 0.5

   Pr(Z < z) = 1.0000

   Ha: p!= 0.5

   Pr(|Z| > |z|) = 0.0000

   Ha: p > 0.5

   Pr(Z > z) = 0.0000

   In non-technical terms, explain the difference between what you conclude for the Netherlands and for the United States from the one-sided test results.

9) Each year in Liverpool, New York, a public librarian estimates the mean number of times the books in that library have been checked out in the previous year. To do this, the librarian randomly samples computer records for 100 books and forms a 95% confidence interval for the mean. This has been done for 20 years. Find the probability that:
   (a) All the confidence intervals contain the true means.
   (b) At least one confidence interval does not contain the true mean.

10) A jury list contains the names of all individuals who may be called for jury duty. The proportion of women on the list is 0.53. A jury of size 12 is selected at random from the list. None selected are women.
   (a) Find the probability of selecting 0 women.
   (b) Test the hypothesis that the selections are random against the alternative of bias against women. Report the P-value, and interpret.

11) An experiment with 26 students in an Israeli classroom consisted of giving everyone a lottery ticket and then alter asking if the would be willing to exchange their ticket for another one, plus a small monetary incentive. Only 7 students agreed to the exchange. In a separate experiment, 31 students were given a new pen and then later asked to exchange it for another pen and a small monetary incentive. All 31 agreed. Conduct inferential statistical methods to analyze the data. Summarize your analyses and interpretations in a short writeup (3-5 sentences).

12) (great exam question!)
Example 6.4 on page 148 tested a therapy for anorexia, using $H_0 : \mu = 0$ and $H_a \mu > 0$ about the population mean and weight change.

(a) In the words of that example, what is a Type I error? What is a Type II error?

(b) The P-value was 0.018. If the decision for $\alpha = 0.05$ were in error, what type of error is it?

(c) Suppose instead $\alpha = 0.01$. What decision would you make? If it is in error, what type of error is it?

13) A decision is planned in a test of $H_0 : \mu = 0$ against $H_a : \mu > 0$ using $\alpha = 0.05$. P(type II error) = 0.17

(a) Explain the meaning of this last sentence.

(b) If the test used $\alpha = 0.01$, would P(type II error) be less than, equal to, or greater than 0.17? Explain.

(c) If $\mu = 10$, would P(type II error) be less than, equal to, or greater than 0.17? Explain

14) You can use the *Errors and Power* applet at www.artofstat.com/wepapps.html to investigate the performance of significance tests, to illustrate their long-run behavior when used for many samples. For significance tests, set the null hypothesis as $H_0 : \pi = 0.50$ for a one-sided test with $H_a : \pi > 0.50$ and sample size 50, and set P(Type I error) = $\alpha = 0.50$. The applet shows the null sampling distribution of $\hat{\pi}$ and the actual sampling distribution of $\hat{\pi}$ for various true values of π. Click on *Show Type II error*, and it also displays P(Type II error), which is the probability of failing to reject H_o even though it is false.

(a) Set the true value of the proportion to be

i. 0.60,

ii. 0.70,

iii. 0.80.

What happens to P(Type II error) as π gets farther from the H_0 value?

(b) Set the true value of the proportion to be

i. 0.53,

ii. 0.52,

iii. 0.51.

What value does P(Type II error) approach as π gets close from the H_0 value?

(c) Fix the true value of the proportion to be 0.60. Show how P(Type II error) changes for n equal to

i. 0.53,

ii. 0.52,

iii. 0.51.

(d) Summarize how the probability of Type II error depends on π (for fixed n) and on n (for fixed π)