Practice Exam #2 (Sta-209, S25)

Ryan Miller

Formula Sheet

Definitions:

- Type I error - rejection of a true null hypothesis
- Type II error - failure to reject a false null hypothesis

Central Limit theorem results:

Statistic	Standard Error	Conditions
$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$np \geq 10$ and $n(1-p) \geq 10$
$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	normal population or $n \geq 30$
$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$n_i p_i \geq 10$ and $n_i(1-p_i) \geq 10$ for $i \in \{1,2\}$
$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	normal populations or $n_1 \geq 30$ and $n_2 \geq 30$

Question #1 (conceptual questions)

Part A: In your own words, explain the meaning of "95% confidence" in the statistical term "95% confidence interval". You should avoid using the words "confidence" or "confident" as core components of your explanation and you should limit yourself to around 2-sentences.

- 95% confidence describes the long-run success rate of the procedure used to make the confidence interval. If this procedure were used many times, such as for many different random samples, 95% of the intervals produced would contain the true population parameter.

Part B: Suppose your friend is planning to perform a hypothesis test and is considering using a decision threshold of $\alpha = 0.01$ rather the conventional threshold of $\alpha = 0.05$. How will their proposed threshold influence the likelihood of a Type I and a Type II error relative the likelihood of these errors using the conventional threshold.

- Their threshold ($\alpha = 0.01$) would decrease the chances of making a Type I error and increase the chances of making a Type II error. This is because the threshold makes it more difficult to reject a true null hypothesis (decreased Type I error rate) but also more difficult to reject a false null hypothesis (increased Type II error rate).

Part C: In your own words, briefly explain the difference between a "sample" and a "population" in the context of statistical inference. That is, how to the concepts of a "sample" and a "population" relate to statistical procedures like confidence interval estimation and hypothesis testing. You should limit your answer to around 3-sentences.

- A population is all of the cases you would like to draw a conclusion about, and a sample is a subset of those cases whose data you have available for analysis. The idea behind the statistical procedures of hypothesis testing and confidence interval estimation is generalize the relationships that exist within a sample to make conclusions about the population that consider the uncertainty involved in only having a subset of cases.

Part D: Suppose you and your friend are each planning to collect a sample of Grinnell students to estimate the proportion of students who have taken Sta-209. You each plan to present a confidence interval estimate of this proportion based upon the data you collect in your sample. Before collecting any data, and assuming *everything else is the same*, which of the changes would lead your confidence interval estimate to be more likely to contain the true proportion than your friend's? You may simple circle or mark the options that will make your interval more likely to contain the truth.

i) You use a 95% confidence level when your friend only uses a 90% confidence level
ii) You use a procedure that is free of sampling bias, while your friend uses a biased sampling protocol
iii) You sample 50 Grinnell students when your friend only samples 30 Grinnell students

- Only i and ii are correct (they will increase the chances of "success") and iii is not. The confidence level and sampling bias directly influence the success rate of the procedure. Sample size will influence the interval's width, but if the confidence level of both intervals is the same it will not influence their long-run success rates.

Part E: Suppose a researcher is planning to perform several thousand hypothesis tests on data from an experiment involving a large number of genes. The researcher is primarily concerned with identifying as many important genes as possible, but they also want to have some control over the number of false positive findings. Would you recommend the researcher use a *false discovery rate control* procedure or the *Bonferroni correction*? State your answer and provide a brief (1-2 sentence) explanation

- False discovery rate control should be recommended because it is powerful than the Bonferroni correction. It will allow a certain fraction of the significant findings to be Type I errors rather than trying to limit the chances of any Type I errors.

Question #2

An experiment conducted at the University of Sydney in Australia investigated whether electrical stimulation to the brain could help participants successfully solve problems that required non-routine approaches.

The experiment trained 40 participants to solve problems in a particular way and then asked them solve an unfamiliar problem that required a creative solution. Of the study's participants, 20 were randomly assigned to receive electrical stimulation to the brain prior to working on the unfamiliar problem, and the other 20 received a placebo condition (the same apparatus without any electricity).

In the electrical stimulation group, 8 participants (40%) successfully solved the problem, while only 4 participants (20%) in the placebo group solved the problem.

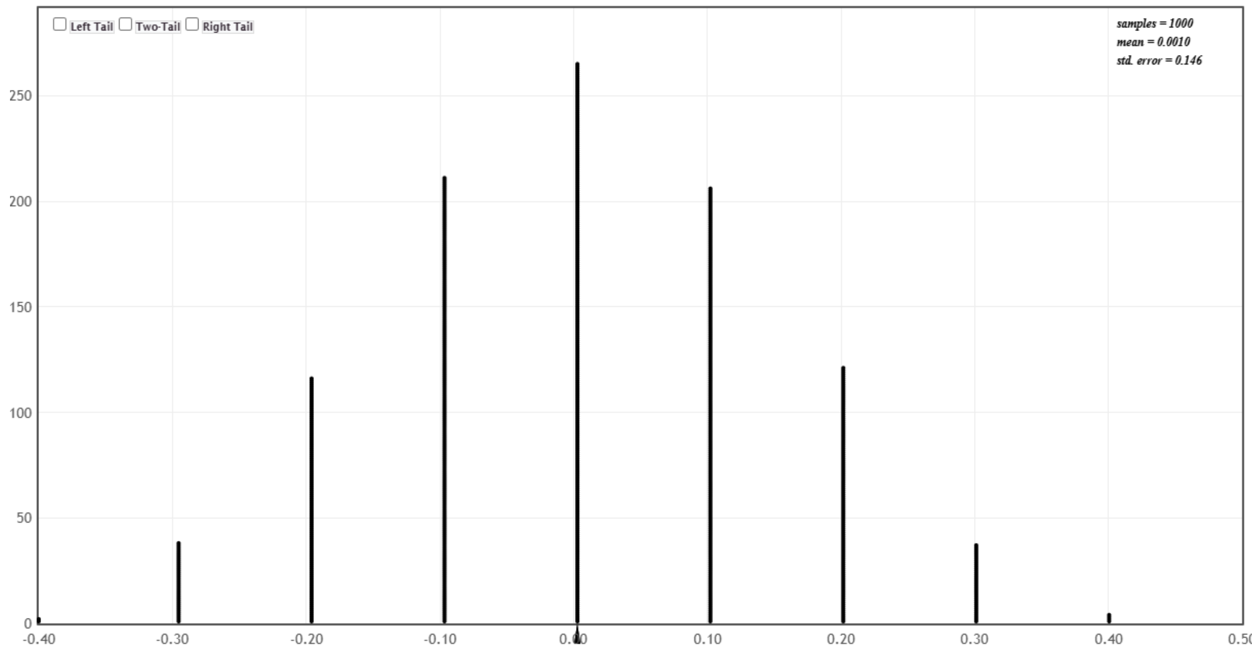
Part A: Notice that a larger proportion of the electrical stimulation group correctly solved the unfamiliar problem. With this in mind, why is a statistical test necessary for the researchers to conclude that electrical stimulation increases the ability of individuals to solve unfamiliar problems? You should limit your explanation to at most 2-sentences.

- Hypothesis testing is necessary to generalize the results from the 40 people who were involved in the study to a broader population while accounting for variability due to the limited sample size.

Part B: Using *both* words and statistical symbols, state the hypotheses that these researchers should be interested in evaluating.

- The null hypothesis is $H_0 : p_1 - p_2 = 0$, which suggests that the proportion of people who correctly solve an unfamiliar problem is the same whether electrical stimulation is used or not. The alternative is $H_a : p_1 - p_2 \neq 0$ which suggests that electrical stimulation has an effect on the ability of people to solve unfamiliar problems.

Part C: The figure below shows a simulated null distribution created using StatKey. Using this distribution, provide an estimate of the two-sided p-value. Your estimate doesn't need to be exact, but it should be reasonably close. You may annotate the figure or show your work to help me assess whether you're estimating your p-value in the proper manner.



- The p-value can be estimated by counting up the outcomes that are at least 0.2 away from zero. There are approximately 125 + 40 + 10 = 175 of these on the right side of the distribution and similar numbers on the other side of the distribution. Thus, there are 350/1000 simulated outcomes that are at least as extreme as the observed difference, so the two-sided p-value is estimated to be around 0.35.

Part D: Using the p-value you estimated in Part C. Provide a 1-2 sentence conclusion that summarizes the results of this study.

- The study finds insufficient evidence of a relationship between the use of electrical stimulation and the ability of people to solve unfamiliar problems.

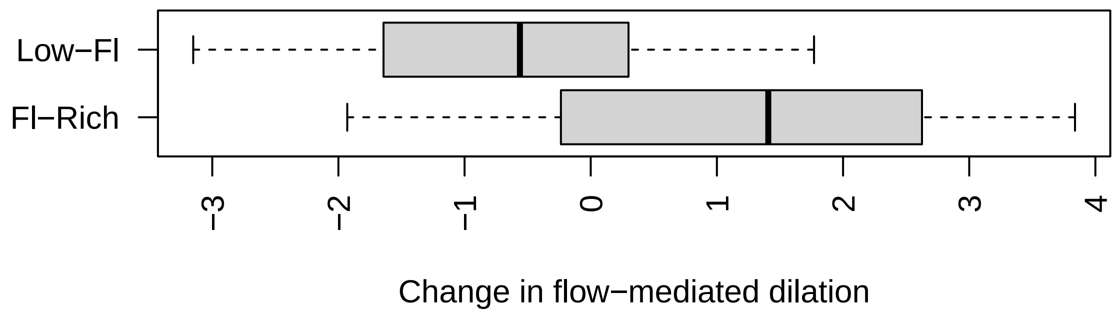
Part E: Suppose these data were used to form a 95% confidence interval estimate for the difference in the proportion of individuals who solve unfamiliar problems with and without electrical stimulation. Would you expect your confidence interval to suggest that a difference of zero is plausible? You should limit your explanation to at most 2-sentences.

- It would contain zero, since the hypothesis test did not find compelling statistical evidence of a difference, which suggests that a difference in proportions of zero is plausible.

Question #3

In a 2004 study, 21 participants were randomly assigned to consume either 46g (1.6 oz) of flavonoid-rich dark chocolate or 46g of chocolate with low flavonoid content every day for 2 weeks. Both chocolates are indistinguishable in taste. Participants had their flow-mediated dilation measured before and after the experiment. Larger increases in flow-mediated dilation indicate a greater improvement in vascular health.

For the 11 participants randomized to consume flavonoid-rich chocolate, the mean increase in flow-mediated dilation was 1.3 units (standard deviation of 2.3). For the 10 participants randomized to consume low-flavonoid chocolate, flow-mediated dilation decreased by an average of -0.9 units (standard deviation of 1.5). The figure below summarizes the observed changes in each group:



Part A: Recall that 95% of the Standard Normal distribution is within 1.96 units of the distribution's center. Additionally, recall that Central Limit theorem suggests the standard error of a single mean can be estimated by $s/\sqrt{n}$. Assuming that a Normal model is appropriate, calculate a 95% confidence interval for the mean change in flow-mediated dilation for individuals who consume flavonoid-rich chocolate.

- The interval is found via point estimate $\pm$ Margin of Error $= 1.3 \pm 1.96 * (2.3/\sqrt{11})$
    - This produces the interval estimate: (-0.06, 2.66)

Part B: Based upon the interval you calculated in Part A, can you confidently conclude that the consumption of flavonoid-rich chocolate, on average, increases flow-mediated dilation? Briefly explain.

- No, the interval suggests that it is plausible for the flavonoid-rich chocolate to produce an average increase of 0, or to decrease flow-mediated dilation by a small amount.

Part C: Your friend looks at the procedure you used to calculate your confidence interval in Part A and suggests you should use the t-distribution rather than the Standard Normal distribution to calibrate the interval's margin of error. If you implemented your friend's suggestion, would the resulting interval be *wider* or *narrower* than the one you previously had calculated? State your answer and provide a brief (1-2 sentence) explanation.

- The interval will be *wider* because the t-distribution has thicker tails than the Normal distribution to account for the additional uncertainty introduced when using the sample standard deviation as an estimate of the population's standard deviation when calculating the margin of error.

Part D: Another friend of yours looks at the procedure being considered in Part C (using the t-distribution) and states that the t-distribution is inappropriate for this application be the sample size, $n = 11$, is too small. Do you agree with this friend? State your answer and provide a brief (1-2 sentence) explanation.

- This friend is wrong and you should disagree with them. The t-distribution was specifically designed for small sample settings where the data are Normally distributed, which seems to apply to this situation.

Part E: This study included a comparison group in addition to the group receiving high-flavanoid chocolate. Suppose we'd like to perform a statistical test to evaluate whether these two groups experienced different changes in flow-mediated dilation. State the appropriate hypotheses for this test using *both* words and statistical symbols.

- The null hypothesis is $H_0 : \mu_1 - \mu_2 = 0$, which suggests that both types of chocolate produce the same average change in flow-mediated dilation. The alternative is $H_0 : \mu_1 - \mu_2 \neq 0$ which suggests the flavanoid content of chocolate influences flow-mediated dilation.

Part F: The p-value for the hypothesis test described in Part E is 0.0204. In your own words, explain what this means.

- This p-value indicates there is a roughly 2% chance of observing a difference in means as large as the difference of 2.2 that was observed in the study if it were the case that the type of chocolate consumed is unrelated to flow-mediated dilation.