Midterm #2 Answers

Instructions: You have 2-hours to complete the examination. The exam consists of two questions worth 120 possible points. You may use a calculator, but you may not use pre-programmed formulas. The standard normal “Z” table from the second page of your book has been provided to you. If you have questions relating to the wording, feel free to ask for clarification. Good luck!

1. (40 points) True, False, or Uncertain? Pick any 5 of the statements below and indicate whether each is true, false, or uncertain. For each statement provide a short paragraph justifying your answer. (Hints: Pictures might be helpful. None of the statements are intended to be excessively confusing: Do not psych yourself out.)

A) (8 points) By failing to reject the null hypothesis we confirm that it is true with a reasonable degree of certainty.

False. It takes strong evidence (an event that would happen less than 10% of the time depending on \( \alpha \) if the null is true) to reject the null hypothesis (4). Just because we fail to reject the null doesn’t mean that it is true (4), just that we don’t have strong evidence to suggest that it is false.

B) (8 points) The probability of a Type I error is known for any hypothesis test that we have covered.

True. The probability of a Type I error is equal to the probability of rejecting the null when it is true. This probability is predetermined by the significance level chosen for the test.

C) (8 points) There is a tradeoff between the probability of committing a Type I error and the probability of committing a Type II error?

True. For any fixed sample size, the only way the probability of a Type II error can be reduced is to increase the probability of a Type I error (the significance level \( \alpha \)). For any fixed sample size, reducing the probability of a Type I error (the significance level \( \alpha \)) will result in an increased probability of a Type II error.
D) (8 points) When the sample size is not sufficiently large for the Central Limit Theorem to apply the expected value of the sample mean is generally not equal to population mean.

False. Assuming the sampling is random sample, the expected value of the sample mean is equal to the population mean irrespective of the sample size. What changes when the sample size is not sufficiently large to apply the CLT is that the sampling distribution of \( \bar{x} \) is no longer normal in absence of strong assumptions.

E) (8 points) Estimating a 95% confidence interval that doesn’t include the true population parameter only happens when a sample is drawn that yields a value of the sample mean in the top or bottom 2.5% of the sampling distribution.

True. We don’t know how to draw the pictures though.

If you drew a value of \( \bar{x} \) the middle 95% of the sampling distribution and then constructed a 95% confidence interval by adding and subtracting 

\[ 1.96 \left( \frac{\sigma}{\sqrt{n}} \right) \]

the confidence interval will always include the true value of the population mean as the cutoff values that leave 2.5% in the upper and lower tail of the sampling distribution are exactly 1.96 standard deviations of \( \bar{x} \) from the population mean.

Alternatively, if you drew a value of \( \bar{x} \) in the top or bottom 2.5% of the sampling distribution and then constructed a 95% confidence interval, the confidence interval would never include the true value of the population mean as the cutoff values that leave 2.5% in the upper and lower tail of the sampling distribution are more than 1.96 standard deviations of \( \bar{x} \) from the population mean.

F) (8 point) The p-value can best be thought of as the probability that the null hypothesis is true.

False. The p-value is the probability of a sample outcome at least as likely as the one that occurred assuming the null is correct – or in plainer English – the probability of observing a sample outcome like the one that was observed assuming the null is true. This isn’t even close to the probability that the null is true.
G) (8 points) There is no way to increase the statistical power in hypothesis test without increasing the probability of a Type II error.

False. These probabilities are compliments of each other, thus when the power \(1 - P(\text{Type II error})\) increases it must be the case that the probability of a Type II error decreases.

H) (8 points) When the sample size is not sufficiently large to apply the Central Limit Theorem, one must assume that the population is normal if the tools developed in class over the past month are to be used conduct statistical inference.

True. When the sample size is not sufficiently large to apply the CLT, we cannot generally assume the sampling distribution of \(\bar{x}\) is normal. The only way that the normality of the sampling distribution is assured is to assume the population is normal.
2. (80 points) Self Sufficiency Project. The Self Sufficiency Project (SSP) was a Canadian randomized experiment designed to measure the effects of an earnings supplementation program. In the experiment Income Assistance (IA) recipients in New Brunswick and British Columbia were selected at random from IA records. Half were randomly assigned to an experimental group and offered a generous earnings supplement, with the remaining participants assigned to a control group.

In order to qualify for the earnings subsidy experimental group members were required to find full-time work within 1-year of assignment. Those participants who qualified for the subsidy received it for 3-years as long as they were working full-time. Because the two groups were similar in all other respects, the “impact” or effect of SSP can be measured by the difference between the program and comparison groups’ subsequent experiences. Based on data generated as part of the SSP Card, Michalopolis, and Robbins (2001) report the attached results:

(A) (5 points) Use the results shown in the table to estimate the impact of the SSP subsidy on the mean log hourly wages of IA recipients (at 33-35 months after assignment).

Estimated Treatment Effect = 2.05 - 2.10 = -0.05

(B) (10 points) Estimate a 90% confidence interval for the “true effect” of the SSP program on log hourly wages 33-35 months after assignment.

$$-0.05 \pm 1.645 \sqrt{\frac{0.03^2}{500} + \frac{0.03^2}{500}} = -0.05 \pm 0.003$$

(C) (5 points) List 5 hypothesis testing results (conclusions) you can draw on the basis of the above interval estimate. These results may involve one or two-tailed test and different levels of significance. Be surely to clearly state the null hypothesis, the alternative hypothesis, and the significance level for each result.

There are infinitely many of these. Below are 5.

$$H_0 : (\mu_T - \mu_C) = 0, \quad \alpha = 0.10, \text{ Reject } H_0$$

---

\section{D) (5 points) Use the results shown above to estimate the impact of the SSP on the employment rate 33-35 months after assignment. Be sure to clearly state what units your estimate is in.}

\textit{Estimated Treatment Effect} = 67 - 65 = 2 \text{ percentage points (or 0.02)}

\section{E) (5 points) What hypothesis would you need to test to conclude that the SSP had a positive impact on employment 33-35 months after assignment? (\textbf{Hint:} For full credit you must correctly state the null and alternative hypotheses).}

\( H_0 : (\mu_T - \mu_C) \leq 0 \)
\[ H_A : (\mu_T - \mu_C) > 0 \]

\section{F) (5 points) Describe the distribution of the sample estimate of the impact of the SSP program on employment 33-35 months after assignment under the null hypothesis that you specified in part E)? (\textbf{Hint:} Make sure you provide all known information about population parameters and sample sizes).}

\( \left( \bar{p}_T - \bar{p}_C \right) \sim N \left( 0, \frac{p(1-p)}{500} + \frac{p(1-p)}{500} \right) \text{ where } p_T = p_c = p \)
G) (10 points) Conduct the hypothesis test you specified in part E) using a significance level of 0.01. What conclusions can you draw from this hypothesis test?

Note that

\[ Z = \frac{0.02}{\sqrt{\frac{\tilde{p}(1-\tilde{p})}{500} + \frac{\tilde{p}(1-\tilde{p})}{500}}} \]

and that the sample sizes are equal so that \( \tilde{p} = 0.66 \). Thus,

\[ Z = \frac{0.02}{\sqrt{\frac{2(0.66)(0.34)}{500}}} = 0.66. \]

Now we can compute the p-value

\[ p-value = P(\Delta \tilde{p} > 0.02 | \tilde{p}_T = \tilde{p}_c, n = 500) = P(Z > 0.66) > 0.01 \]

\( \Rightarrow \) fail to reject \( H_0 \)

We cannot statistically reject the hypothesis that the treatment group had poorer results than the control group. Or in other words, we do not have strong statistical evidence that the treatment group had better employment outcomes than the control group.

H) (10 points) Provide definitions for Type I error, Type II error, and statistical power in the context of the hypothesis that you specified test in part E).

*Type I Error* – rejecting the null when the null is true

*Type II Error* – failing to reject the null when it is not true.

*Statistical power* – the probability of correctly rejecting the null when it is not true.
Suppose you know that similar earnings subsidies offered to IA recipients in other Canadian provinces increased the employment rate (measures as the percent of the population employed) by 5 percentage points 33-35 months after assignment.

I) (10 points) Assume that the subsidy offered to the SSP Experimental Group is as effective at increasing employment in months 33-35 as similar earnings subsidies provided to IA recipients in other Canadian provinces. Calculate the power of the test you specified in part E) with a significance level of 0.01. What does this power calculation tell you? (Hint: Drawing a picture might be useful).

The first step in any power calculation is to find the realization of the sample statistic (in this case $\Delta \bar{p}$) that is on the boundary between rejecting the null or not. Note that

$$Z_{critical} = \frac{\Delta \bar{p}_{critical} - \Delta p_0}{\sqrt{(\bar{p})(1-\bar{p})/500 + (\bar{\tilde{p}})(1-\bar{\tilde{p}})/500}} = 2.33$$

In order to calculate the value of $\Delta \bar{p}_{critical}$ we are going to need to make some assumptions about the value of $\bar{\tilde{p}}$ which would normally be computed as the average of the $\bar{p}_c$ and $\bar{p}_e$. I will assume the midpoint of the sample means is going to be 0.66. Thus,

$$Z_{critical} = \frac{\Delta \bar{p}_{critical} - 0}{\sqrt{(0.66)(0.34)/500 + (0.66)(0.34)/500}} = 2.33 \Rightarrow$$

$$\Delta \bar{p}_{critical} = 0.070$$

Now we need to compute the probability of drawing a realization of $\Delta \bar{p} > 0.070$ if we are drawing from the true distribution where $\Delta p = 0.05$. To compute this probably you need to make some assumptions about the true values of $p_T$ and $p_e$. The constraints on these assumptions are (i) that you should pick values that are around the sample means, and (ii) the difference of the values should be 0.05. I assume $p_T = 0.72$ and $p_e = 0.67$. 
This power calculation tells me that there is a 25% chance of correctly rejecting the null assuming the SSP treatment as effective as similar treatments in other provinces.

J) (15 points) In your position as an analyst in Saskatchewan’s Ministry of Social Services, you have been asked to set up a randomized evaluation to estimate the effect of an earnings subsidy very similar to the one offered in BC as part of SSP and in other Canadian provinces. Assuming the effect of such a subsidy is in Saskatchewan to increase the probability of employment at 33-35 months by 5 percentage points, how large would your samples need to be to ensure that you correctly reject the null hypothesis that you specified in part E) 95 percent of the time. You may assume equally sized control and experimental groups. (Hint: This calculation requires that you make some assumptions. For this reason, two students who do the calculations 100% correct may well end up with different answers. Please clearly state any assumptions that you make. To ensure you get full credit please clearly state any assumptions that you make).

This problem looks a lot like the last ones so I’m going to maintain all the same assumptions. The first step is to compute \( \Delta p_{critical} \) which will now be a function of the sample size \( n \). Borrowing from Part I)

\[
Z_{critical} = \frac{\overline{\Delta p}_{critical} - 0}{\sqrt{\frac{(0.66)(0.34)}{n} + \frac{(0.66)(0.34)}{n}}} = 2.33 \Rightarrow \\
\Delta p_{critical} = \frac{1.56}{\sqrt{n}}
\]

Now figure out the sample sizes need to ensure that the probability of drawing a realization of \( \overline{\Delta p} > \Delta p_{critical} \) if we are drawing from the true distribution where \( \Delta p = 0.05 \) is at least 0.95.

The cutoff in the Z-distribution that leaves 0.05 in the lower tail is -2.33. Thus we need
\[ Z_{p.c.} = \frac{1.56 - 0.05}{\sqrt{n}} < -1.645 \Rightarrow \]
\[ \frac{1.56 - 0.05\sqrt{n}}{0.65} < -1.645 \Rightarrow \]
\[ \sqrt{n} > 52.59 \Rightarrow \]
\[ n > 2,766 \]
Comparison of Mean Labor Market Outcomes of Program and Control Group Members who were Working at Assignment (Sample Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Control Group (N=500)</th>
<th>Experimental Group (N=500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Working</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In months 12-14</td>
<td>73.0</td>
<td>78.0</td>
</tr>
<tr>
<td>Percent working</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In months 33-35</td>
<td>65.0</td>
<td>67.0</td>
</tr>
<tr>
<td>Cumulative Months Worked (Months 12-33)</td>
<td>15.3 (0.4)</td>
<td>15.5 (0.4)</td>
</tr>
<tr>
<td>Average Monthly Hours (Months 12-33)</td>
<td>21.1 (0.7)</td>
<td>22.0 (0.7)</td>
</tr>
<tr>
<td>Average Number of Months with SSP (Months 12-33)</td>
<td>0.0 (0.4)</td>
<td>8.7 (0.4)</td>
</tr>
<tr>
<td>Mean Log Hourly Wage (Months 12-14)</td>
<td>1.98 (0.03)</td>
<td>1.94 (0.02)</td>
</tr>
<tr>
<td>Mean Log Hourly Wage (Months 33-35)</td>
<td>2.10 (0.03)</td>
<td>2.05 (0.03)</td>
</tr>
</tbody>
</table>