Name:________________________________ Date:__________________

Ch 12 and 13 Assignment 2 Sample Hypothesis Test and Confidence Inverval

The following assignment will contain examples that go through each type of the inference procedures:

2 Sample Z-interval	2 Sample Z – Test	2 sample T-interval
2 Sample T – Test	2 Sample Proportion Interval	2 Sample Proportion Test
Matched Pair T- Test		

1. Studies have shown that moderately drinking red wine may help reduce cholesterol by having polyphenol. In an experiment subjects(male) were randomly assigned either red or white wine each day for 2 weeks. Polyphenol levels were measured in their blood after each two week periods. Percentage changes in their polyphenol levels are given below. Is there good evidence that drinking red wine increases polyphenol more than white wine?

Red	3.4	8.2	7.3	4.1	0.6	4.8	8.5	7.1	5.4
White	3.2	0.6	–3.9	4.2	–0.8	2.8	1.8	–5.8	0.2

a) If we are to construct a confidence interval, will we be using a Z-interval or a T-interval?

b) Construct and interpret a 90% confidence interval for the difference in mean percent change in polyphenol levels for red wine and white wine treatments

c) Indicate which of the conditions for inference are met

d) Does the interval in part (a) suggest that red wine is more effective than white wine? Explain:

2. A sample of 840 men and 1077 women, aged 21 to 25 years old, were assessed on their literacy skills using the NAEP test. The mean and standard deviation of scores are as follow: $\bar{x}_{men} = 272.40$, $s_{men} = 59.2$, $\bar{x}_{women} = 274.73$, and $s_{women} = 57.5$.

a) Construct and interpret a 90% confidence interval for the difference in mean score for male and female young adults.

b) Based only on the interval from part (a), is there convincing evidence of a difference in mean score for male and female young adults?

3. A survey was conducted to ask if the federal government should raise tax on cigarettes to supplement the cost for the health care system. Subjects surveyed were classified as "Smokers" or "Non Smokers". 605 non smokers surveyed and 351 responded yes. 195 smokers surveyed and 41 responded yes.

   a) Is there sufficient evidence at the significance level of 5% to conclude that the two populations – smokers vs non-smokers, differ significantly with respect to their opinions?
   State the Null and Alternative Hypothesis, check if all conditions for inference are met, calculate your test - statistics and P-value. Interpret your results

   b) Construct a 95% confidence interval for the difference in proportions between the two population. Interpret your results:

4. A researcher conducted a medical study to investigate whether taking a low-dose aspirin reduces the chance of developing colon cancer. As part of the study, 10000 adult volunteers were randomly assigned to one of two groups. 6500 volunteers were assigned to the experimental group that took a low-dose aspirin each day. The rest of the group (3500) were assigned to the control group that took a placebo each day. At the end of six years, 191 of the people who took the low-dose aspirin had developed colon cancer and 212 of the people who took the placebo had developed colon cancer. At the significance level of a = 5%, do that data provide convincing statistical evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers

   a) State the Null and Alternative Hypothesis. Indicate what your parameter of interest

   b) Check if all conditions for inference are met

   c) Calculate your test statistics and P-value. Interpret your results

   d) Suppose research has shown that taking a low dose of aspirin would reduce the proportion in difference between the two population by 2.8%. What is the probability of a Type II error?

5. Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation.

The researchers wanted to perform a hypothesis test of

$$H_0 : p_m - p_c = 0$$
$$H_a : p_m - p_c < 0,$$

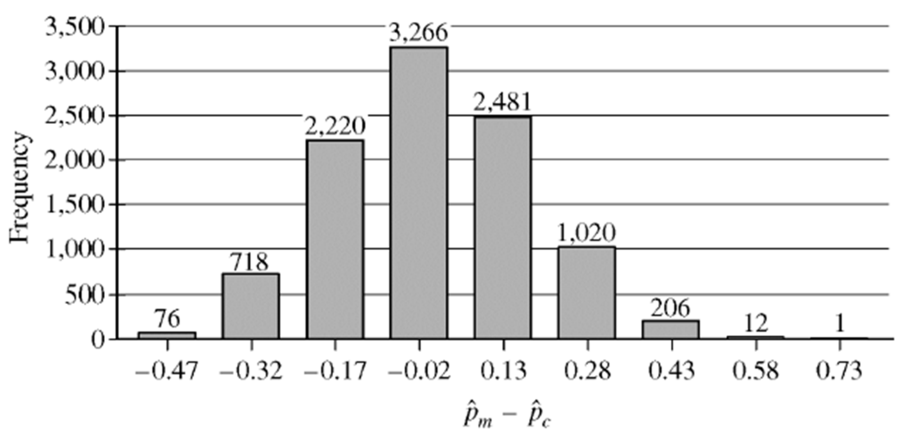
where p_m is the proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation and p_c is the proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation.

(a) If the study were to provide significant evidence against H_0 in favor of H_a, would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

(b) Let $\hat{p}_m$ represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let $\hat{p}_c$ represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to use a normal approximation for the sampling distribution of $\hat{p}_m - \hat{p}_c$?

Although a normal approximation cannot be used, it is possible to simulate the distribution of $\hat{p}_m - \hat{p}_c$. Under the assumption that the null hypothesis is true, 10,000 values of $\hat{p}_m - \hat{p}_c$ were simulated. The histogram below shows the results of the simulation.



(c) Based on the results of the simulation, what can be concluded about the relationship between blood pressure and meditation among men in the retirement community?

6. A large company produces an equal number of brand-name lightbulbs and generic lightbulbs. The director of quality control sets guidelines that production will be stopped if there is evidence that the proportion of <u>all</u> lightbulbs that are defective is greater than 0.10. The director also believes that the proportion of brand-name lightbulbs that are defective is not equal to the proportion of generic lightbulbs that are defective. Therefore, the director wants to estimate the average of the two proportions.

   To estimate the proportion of brand-name lightbulbs that are defective, a simple random sample of 400 brand-name lightbulbs is taken and 44 are found to be defective. Let X represent the number of brand-name lightbulbs that are defective in a sample of 400, and let p_X represent the proportion of all brand-name lightbulbs that are defective. It is reasonable to assume that X is a binomial random variable.

   (a) One condition for obtaining an interval estimate for p_X is that the distribution of $\hat{p}_X$ is approximately normal. Is it reasonable to assume that the condition is met? Justify your answer.

(b) The standard error of $\hat{p}_X$ is approximately 0.0156. Show how the value of the standard error is calculated.

(c) How many standard errors is the observed value of $\hat{p}_X$ from 0.10 ?

To estimate the proportion of generic lightbulbs that are defective, a simple random sample of 400 generic lightbulbs is taken and 104 are found to be defective. Let Y represent the number of generic lightbulbs that are defective in a sample of 400. It is reasonable to assume that Y is a binomial random variable and the distribution of $\hat{p}_Y$ is approximately normal, with an approximate standard error of 0.0219. It is also reasonable to assume that X and Y are independent.

The parameter of interest for the manager of quality control is D, the average proportion of defective lightbulbs for the brand-name and the generic lightbulbs. D is defined as $D = \dfrac{p_X + p_Y}{2}$.

(d) Consider $\hat{D}$, the point estimate of D.

   (i) Calculate $\hat{D}$ using data from the sample of brand-name lightbulbs and the sample of generic lightbulbs.

(ii) Calculate $s_{\hat{D}}$, the standard error of $\hat{D}$.

7. High cholesterol levels in people can be reduced by exercise, diet, and medicateion. Twenty middle -aged males with cholesterol readings between 220 to 240 mg per decileter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20 Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40 Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

SolutionQ1:

State: μ_1 = the true mean level of polyphenols in blood after drinking red wine(%)
μ_2 = the true mean level of polyphenols in blood after drinking white wine (%)
we want to estimate the difference in mean level of polyphenols in blood ($\mu_1 - \mu_2$) after drinking red or white wine.
$\bar{x}_1 = 5.5\%$ $\bar{x}_2 = 0.2\bar{3}\%$

Plan: random: the data come from a well-designed randomized experiment. ✓
10% condition: $n_1 = 9$ 90 < all healthy men ✓
$n_2 = 9$ 90 < all healthy men ✓
Normal/Large: $n_1 = 9 < 30$, $n_2 = 9 < 30$ but a graph of the data shows no strong skewness and no outliers. ✓
Because our conditions are met, we will construct a 2-sample t-interval for difference in means $\mu_1 - \mu_2$.

Do: 2-SampTInt: $\bar{x}_1$: 5.5 $\bar{x}_2$: $0.2\bar{3}$ C-level: 0.9
Sx_1: 2.52 Sx_2: 3.29 pooled: no = (2.8448, 7.6885)
n_1: 9 n_2: 9 df = 14.97 t^* = 1.75

Conclude: we are 90% confident that the interval from 2.8448% to 7.6885% captures the difference in mean level of polyphenols in the blood after drinking red vs. white wine ($\mu_1 - \mu_2$).

b. Does the interval in part (a) suggest that red wine is more effective than white wine? Explain. Yes! The entire interval of plausible values is positive which means red wine is more effective than white wine at increasing polyphenol levels in blood.

Solution Q2:

State: μ_1 = true mean NAEP score for men (Quantitative reasoning).
μ_2 = true mean NAEP score for women (Quantitative reasoning).
$\bar{x}_1 = 272.40$ we want to estimate the true difference
$\bar{x}_2 = 274.73$ in means $\mu_1 - \mu_2$ with 90% confidence.

Plan: Random: random sample split into 2 independent groups ✓
10% condition: $n_1 = 840$ 8400 < all men aged 21-25 years ✓
$n_2 = 1077$ 10770 < all women aged 21-25 years ✓
Normal/Large: 840 ≥ 30 ✓ 1077 ≥ 30 ✓
Because our conditions are met, we will construct a 2-sample t interval for difference in 2 means $\mu_1 - \mu_2$.

Do: 2-SampTInt:
$\bar{x}_1$: 272.4 $\bar{x}_2$: 274.73 C-level: 0.90
Sx_1: 59.2 Sx_2: 57.5 pooled: no = (−6.759, 2.0988)
n_1: 840 n_2: 1077 df: 1777.5 t^*: 1.646

Conclude: we are 90% confident that the interval from −6.759 to 2.0988 (points) captures the true difference in mean NAEP Quantitative reasoning scores between men and women $\mu_1 - \mu_2$.

b. Based only on the interval from part (a), is there convincing evidence of a difference in mean score for male and female young adults?

No, because the interval of plausible values contains zero (no difference).

Solution to Q3: https://online.stat.psu.edu/stat415/lesson/9/9.4

Marking Scheme to Q4

Step 3 is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and a *p*-value that is consistent with the stated alternative hypothesis.

Partially correct (P) if the response correctly calculates the test statistic but not the *p*-value;
OR
if the response calculates the test statistic incorrectly but then calculates the correct *p*-value for the computed test statistic;
OR
if the response reports the correct *p*-value but no calculations or test statistic are shown.

Incorrect (I) if the response fails to meet the criteria for E or P.

Note: The *p*-value is considered correct if it is consistent with the alternative stated in the response and the calculated test statistic, even if those are incorrect.

Step 4 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion in context, with justification based on linkage between the *p*-value and the given $\alpha = 0.05$.

Partially correct (P) if the response provides a correct conclusion, with linkage to the *p*-value, but not in context;
OR
if the response provides a correct conclusion in context, but without justification based on linkage to the *p*-value.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:

- The conclusion must be related to the alternative hypothesis.
- If the *p*-value is incorrect, then step 4 is scored as E if the response includes proper linkage and a conclusion in context consistent with that *p*-value.
- If the *p*-value is less than 0.05, wording that states or implies that the alternative hypothesis is *proven* lowers the score one level (that is, from E to P or P to I) in step 4.
- If the *p*-value is incorrect and greater than 0.05, wording that states or implies that the null hypothesis is *accepted* lowers the score one level (that is, from E to P or P to I) in step 4.

Step 1 is scored as follows:

Essentially correct (E) if the response identifies correct parameters *AND* both hypotheses are labeled and state the correct relationship between the parameters.

Partially correct (P) if the response identifies correct parameters *OR* states correct relationships, but not both.

Incorrect (I) if the response does not meet the criteria for E or P.

Note: Either defining the parameters in context, or simply using common parameter notation with subscripts clearly relevant to the context, such as p_{asp} and p_{plac}, is sufficient.

Step 2 is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:

1. Identifies the correct test procedure (by name or by formula).
2. Notes that the use of random assignment satisfies the randomness condition.
3. Checks for approximate normality of the test statistic by citing that all four counts are larger than some standard criterion such as 5 or 10.

Partially correct (P) if the response correctly includes only two of the three components.

Incorrect (I) if the response correctly includes at most one of the three components.

Notes:

- For the randomness component, it is (minimally) acceptable to say "random assignment — check" but not acceptable to say "random — check" or "SRS — check." The important concept here is that it is random assignment, and not random sampling, that is required. If the response implies that the study used a random sample, the randomness component is not satisfied, regardless of whether random assignment is correctly addressed.
- The normality check may use the expected counts under the null hypothesis in place of observed counts.

SOL Q5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) recognize the limited conclusions that can be drawn from an observational study; (2) determine whether a condition for applying a particular inference procedure is satisfied; and (3) draw an inferential conclusion from a simulation analysis.

Solution

Part (a):

No, it would not be reasonable to conclude that meditation causes a reduction in blood pressure for men in the retirement community. Because this is an observational study and not an experiment, no cause-and-effect relationship between meditation and lower blood pressure can be inferred. It is quite possible that men who choose to meditate could differ from men who do not choose to meditate in other ways that were also associated with blood pressure.

Part (b):

The sample sizes were too small, relative to the overall sample proportion of successes, to justify using a normal approximation. One way to check this is to note that the combined sample proportion of successes is $\hat{p} = \frac{0+8}{11+17} = \frac{8}{28} \approx 0.286$, so neither $n_m\hat{p} = 11 \times \frac{8}{28} \approx 3.143$ nor $n_c\hat{p} = 17 \times \frac{8}{28} \approx 4.857$ is at least 10.

Part (c):

The observed value of the sample statistic $\hat{p}_m - \hat{p}_c$ is $\frac{0}{11} - \frac{8}{17} \approx -0.47$. The graph of simulation results reveals that a difference of -0.47 or more extreme was very rare. In fact, the value -0.47 was the smallest possible outcome and occurred in only 76 of the 10,000 repetitions in the simulation. Thus, assuming that all men in the retirement community were equally likely to have high blood pressure whether they meditate or not, there is an approximate probability of 0.0076 of getting a difference of -0.47 or smaller by chance alone. Because this approximate *p*-value is very small, there is convincing evidence that men in this retirement community who meditate were less likely to have high blood pressure than men in this retirement community who do not meditate. However, because this is an observational study, even though we can conclude that meditation is *associated* with a lower chance of having high blood pressure, we cannot conclude that meditation *causes* a reduction in the likeliness of having high blood pressure.

Part (a) is scored as follows:

Essentially correct (E) if the response correctly claims that a cause-and-effect conclusion cannot be justified *AND*

- Provides an explanation based on the study design (for example, noting that this study was not an experiment, or was just an observational study, or that treatments weren't randomly assigned, or that no variables were controlled)
  OR
- Provides a complete explanation of confounding in the context of this question by describing that men who choose to meditate could differ from men who do not choose to meditate in other ways that were also associated with blood pressure.

Partially correct (P) if the response correctly claims that a cause-and-effect conclusion cannot be justified *AND* provides a weak or incomplete explanation (for example, only citing that association is not causation, only noting that there could be confounding/lurking variables, or only stating that other variables such as diet might affect blood pressure).

Incorrect (I) if the response claims that a cause-and-effect conclusion can be drawn *OR* answers that no cause-and-effect conclusion can be drawn but provides an incorrect explanation or does not provide an explanation (for example, only saying "We cannot conclude causation, we can only conclude association" without providing a reason).

Notes

1. A response that says a cause-and-effect conclusion cannot be justified and provides a correct explanation based on the study design (bullet 1) and *also* mentions confounding/lurking variables without a complete explanation of confounding is scored essentially correct.
2. A response that provides an additional *incorrect* explanation (for example, that the sample size is too small, or that the conditions for inference weren't met, or that $n < 30$), lowers the score one level (that is, from E to P, or from P to I) in part (a).
3. A response that makes an incorrect claim about a significance test (for example, "we cannot conclude cause-and-effect from a significance test" or "significance tests can only show association") lowers the score one level (that is, from E to P, or from P to I) in part (a). However, a correct statement such as "a significance test *alone* isn't sufficient to justify cause-and-effect" is not penalized.

Part (b) is scored as follows:

Essentially correct (E) if the response indicates that at least one observed or expected count is too small *AND* includes the following three components:

- States the numerical value of at least one of the relevant observed or expected counts of successes or failures for one of the two groups
- Clearly labels/identifies the count using words (for example, number of meditators who have high blood pressure), symbols with at least one subscript (for example, $n_m\hat{p}_m$, $n\hat{p}_m$, np_m), or evidence of calculation (for example, $11 \times \frac{0}{11}$).
- Correctly compares this count to a reasonable boundary (for example, 5 or 10, but not 30)

Partially correct (P) if the response indicates that at least one observed or expected count is too small *AND* includes exactly two of the three components listed above.

Incorrect (I) if the response does not satisfy the criteria for E or P

Notes

- If the response correctly discusses other conditions for a two-sample *z* test for a difference in proportions, these should be ignored. However, if the response makes an *incorrect* statement about the conditions (for example, the sample size should be greater than 30, the population is/should be Normal, the sample is/should be Normal), then the response lowers the score one level (that is, from E to P, or from P to I) in part (b). Summary statements about the sample size (for example, "the sample size is too small") were not penalized because they were not proposing an additional condition.
- Any statement about conditions for performing inference in part (a) should not be considered in part (b).