Question 5

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.

Question 5 (continued)

Scoring

Parts (a), (b), and (c) were scored as essentially correct (E), partially correct (P), or incorrect (I).

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.

Question 5 (continued)

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).

Part (c) is scored as follows:

Essentially correct (E) if the response provides evidence that the difference in the sample proportions $\hat{p}_m - \hat{p}_c \approx -0.47$ was calculated *AND* clearly uses the results of the simulation *AND* includes the following two components:

- States that values less than or equal to -0.47 were very unlikely, by comparing 0.0076 to a common significance level or saying that a difference of -0.47 or less is very unlikely.
- Draws an appropriate conclusion in context.

Partially correct (P) if the response provides evidence that the difference in the sample proportions was calculated *AND* clearly uses the results of the simulation *AND* includes exactly one of the two components listed above.

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

Note:

- If the response subtracts the sample proportions in the opposite order, calculates the difference to be $+0.47$, and uses the right side of the simulated distribution correctly, then the response is essentially correct if it also includes the two components listed above.

Question 5 (continued)

4 Complete Response

All three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and one part incorrect

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct and two parts incorrect

OR

One or two parts partially correct

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.

No, the researchers cannot make any conclusions about cause-and-effect because they merely conducted an observational study, and not a controlled experiment. Had they performed the latter with randomization, then conclusions about cause-and-effect could be made.

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 $\hat{p}_m - \hat{p}_c$?

In order to use a Normal approximation, the sampling distributions of both $\hat{p}_m$ and $\hat{p}_c$ need to be approximately Normal, meaning that $n_m\hat{p}_m$, $n_m(1-\hat{p}_m)$, $n_c\hat{p}_c$, and $n_c(1-\hat{p}_c)$ need to be at least 10.

Using the numbers given, however,

$n_m\hat{p}_m = (11)(\frac{0}{11}) = 0 \not\geq 10$ $\quad$ $n_c\hat{p}_c = (17)(\frac{8}{17}) = 8 \not\geq 10$

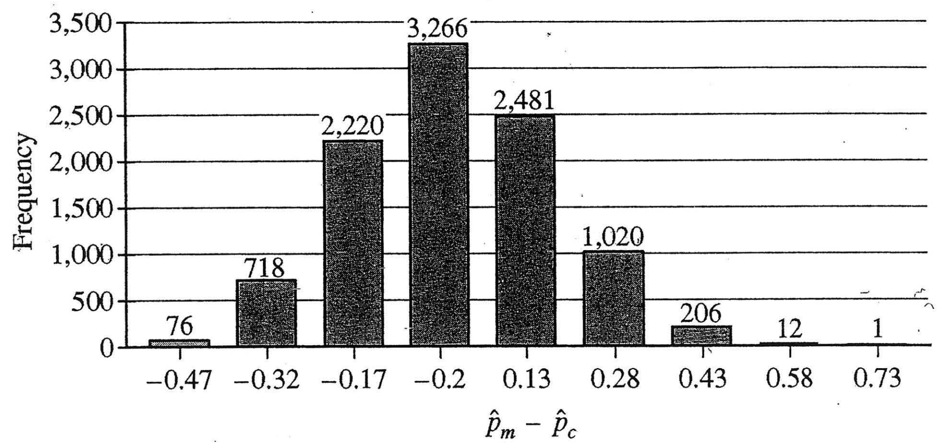
$n_m(1-\hat{p}_m) = (11)(\frac{11}{11}) = 11 \geq 10$ ✓ $\quad$ $n_c(1-\hat{p}_c) = (17)(\frac{9}{17}) = 9 \not\geq 10$

this condition is not met, and thus it is not reasonable to use a Normal approximation for the sampling distribution of $\hat{p}_m - \hat{p}_c$.

GO ON TO THE NEXT PAGE.

 5A2

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?

Because the value of $\hat{p}_m - \hat{p}_c$ obtained from the sample yields $\left(\frac{0}{11}\right) - \left(\frac{8}{17}\right)$

$= -0.47$

Because only 76 of the 10000 simulated values have a value of −0.47 or less (more extreme), or .76%, as this is very unlikely, we can reject Ho and conclude that the true proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily medidation is greater than the true proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation.

GO ON TO THE NEXT PAGE.

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.

No, because this was not a controlled experiment and association does not imply causation. There may have been confounding variables associated with participating in meditation that were the real cause of a difference in blood pressure. For example, people who participate in meditation may be more health-conscious and exercise more, decreasing blood pressure.

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.

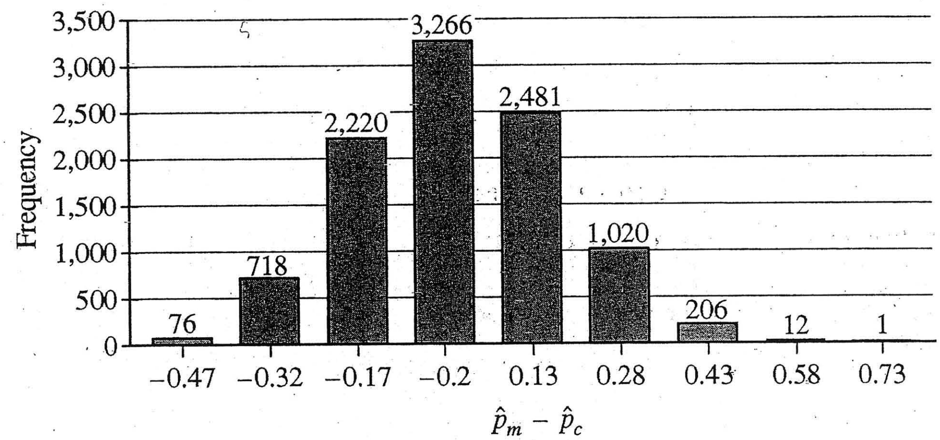
(b) Let $\hat{p}_m$ represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let $\hat{p}_c$ represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to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

the Sample is not large enough because there are less than 10 successes among men who meditate daily. A normal approximation can be used if there are at least 10 successes and 10 failures for each group.

GO ON TO THE NEXT PAGE.

5B2 5B2

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?

the psychologists had a value of $\hat{p}_m - \hat{p}_c = \frac{0}{11} - \frac{8}{17} = -.47$.

from the simulation, we see that this is the most extreme negative value achieved, with only 76 times out of 10,000 giving a value of −.47. Given that the null hypothesis is true and there is no difference in the proportions for men who meditate and men who don't, there is an approximately 0.76% chance that $\hat{p}_m - \hat{p}_c$ is this extreme or more extreme.

Because of this very low probability (less than any reasonable α), we can reject the null hypothesis and conclude that the proportion of men with high blood pressure is less for men in this community who meditate than for men in the retirement community who don't meditate.

GO ON TO THE NEXT PAGE.

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.

No, the samples of 11 and 17 are too small to draw any assumptions from.

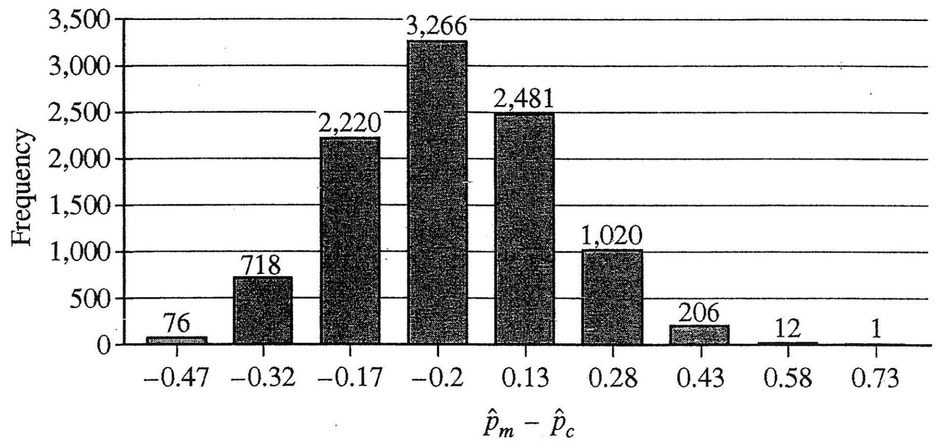
The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

(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 $\hat{p}_m - \hat{p}_c$?

Because np and nq are not greater than 10 for both samples.

5C2 5C2

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?

Meditation does tend to be related to lower blood pressure among men in the retirement community.

GO ON TO THE NEXT PAGE.

Question 5

Overview

The primary goals of this question were to (1) assess a student's ability to 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.

Sample: 5A
Score: 4

In part (a) the response clearly indicates that a cause-and-effect conclusion is not reasonable because of the design of the study ("merely conducted an observational study, and not a controlled experiment"). Because the response provides a correct explanation based on the study design, part (a) was scored as essentially correct. In part (b) the response calculates the observed number of successes and failures in the meditation and non-meditation groups, identifies the counts using symbols (for example, $n_m\hat{p}_m$), and states that three of these counts are less than a reasonable boundary (10). Because the response indicates that at least one observed count was too small and includes all three components, part (b) was scored as essentially correct. In part (c) the response calculates the difference in the sample proportions (-0.47), uses the results of the simulation to determine that obtaining a difference at least this extreme is "very unlikely," and makes an appropriate conclusion in context ("the true proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation is greater than the true proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation.") Because the response calculates the difference in proportions, uses the results of the simulation, and includes both additional components, part (c) was scored as essentially correct. Because all three parts were scored as essentially correct, the response earned a score of 4.

Sample: 5B
Score: 3

In part (a) the response clearly indicates that a cause-and-effect conclusion is not reasonable because of the design of the study ("not a controlled experiment"). The response then gives a very nice description of confounding in the context of the study, by providing an additional difference between men who meditate and men who don't meditate ("more health-conscious and exercise more") that might also decrease blood pressure. Because either response (not an experiment, complete explanation of confounding) would be enough to justify that a cause-and-effect conclusion is not appropriate, part (a) was scored as essentially correct. In part (b) the response states that the sample is not large enough, identifies an observed count in words ("successes among men who meditate"), and provides a reasonable boundary for the counts of success ("less than 10"). However, the response fails to include the numerical value of the count (0). Because the response indicates that an observed count was too small and includes two of the three components, part (b) was scored as partially correct. In part (c) the response calculates the difference in the sample proportions (-0.47), uses the results of the simulation to determine that obtaining a difference this extreme is unlikely ("very low probability"), and makes an appropriate conclusion in context ("the proportion of men with high blood pressure is less for men in this community who meditate than for men in the retirement community who don't meditate.") Because the response calculates the difference in proportions, uses the results of the simulation, and includes both additional components, part (c) was scored as essentially correct. Because two parts were scored as essentially correct and one part was scored as partially correct, the response earned a score of 3.

Question 5 (continued)

Sample: 5C
Score: 1

In part (a) the response says "no" but does not provide an appropriate explanation for why it is not reasonable to conclude a cause-and-effect relationship. Comments about small sample size are inappropriate because the stem of the question suggests that the results are significant. Because the explanation is not correct, part (a) was scored as incorrect. In part (b) the response states that the observed counts are too small by comparing them to a reasonable boundary (10). The counts are identified in symbols and words (*np* and *nq* for *both* samples), but the values of these counts are not provided. Because the response indicates that at least one observed count was too small and includes two of the three components, part (b) was scored as partially correct. In part (c) the response does not provide evidence that the results of the simulation were used to determine that a difference in sample proportions at least as extreme as -0.47 would be unlikely to occur by chance alone. Because there is no evidence that the difference in sample proportions was calculated, part (c) was scored as incorrect. Because one part was scored as partially correct and two parts were scored as incorrect, the response earned a score of 1.