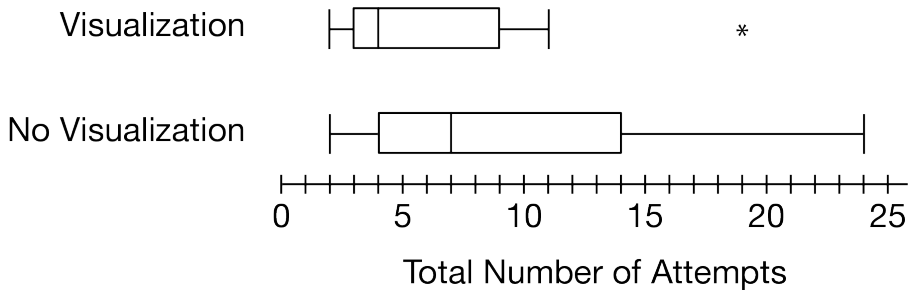
Setting Up a Test for the Difference of Two Population Means Quiz

1. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

A team of psychologists studied the concept of visualization in basketball, where players visualize making a basket before shooting the ball. They conducted an experiment in which 20 basketball players with similar abilities were randomly assigned to two groups. The 10 players in group 1 received visualization training, and the 10 players in group 2 did not.

Each player stood 22 feet from the basket at the same location on the basketball court. Each player was then instructed to attempt to make the basket until two consecutive baskets were made. The players who received visualization training were instructed to use visualization techniques before attempting to make the basket. The total number of attempts, including the last two attempts, were recorded for each player.

The total number of attempts for each of the 20 players are summarized in the following boxplots.

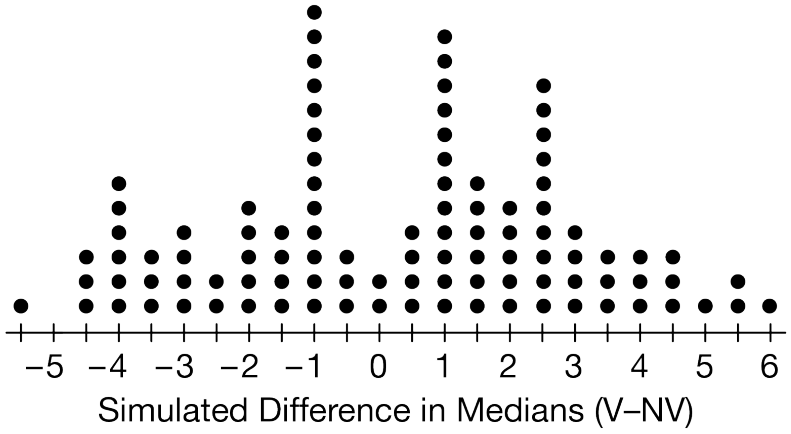


(a) Based on the boxplots, did basketball players who received visualization training tend to need fewer attempts to make two consecutive baskets from a distance of 22 feet than players who did not receive the training? Explain your reasoning.

(b) State and check conditions for conducting a two-sample t-test for a difference in means.

Because both distributions, visualization (V) and no visualization (NV), are skewed, the psychologists conducted a simulation to test for a difference in medians rather than means. For each trial of the simulation, the 20 values of the total number of attempts observed in the experiment were combined into one group and then randomly split into two groups of 10. The difference in the medians $(\mathrm{V} - \mathrm{NV})$ of the groups was calculated for each trial. The following dotplot shows the difference in the medians for 100 trials of the simulation.

Setting Up a Test for the Difference of Two Population Means Quiz



(c) Using the observed difference in medians $(\mathrm{V} - \mathrm{NV})$ and the results of the simulation, estimate a p-value for a test for the difference in medians. Show the work needed to calculate this p-value.

(d) Based on the p-value in part (c), is there convincing statistical evidence that basketball players similar to the ones in this study who receive visualization training need fewer attempts to make two consecutive baskets from a distance of 22 feet than those who do not receive such training? Justify your answer.

Part A, B, C, and D

The primary goals of this question are to assess a student's ability to (1) use information from box plots to draw a preliminary conclusion about the results of an experiment; (2) state and verify the conditions for a two-sample *t* test for a difference in means; (3) use the results of a simulation to estimate a *p*-value; (4) use a *p*-value to draw an inferential conclusion about the results of an experiment.

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as $1/2$ point.

Scoring

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

If a response is between two scores (for example, $2\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Reasons to score up:

· *All notation is correct and clearly marked*

· *All explanations are clear*

· *No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)*

· *No minor calculation errors are made, if they are not part of the scoring*

Setting Up a Test for the Difference of Two Population Means Quiz

· Interpretation parts are especially strong

Reasons to score down:

· Notation is not wrong, but is spotty and not clearly marked

· Explanations are not wrong, but are hard to follow

· Wrong or extraneous information is included but not part of scoring

· Minor calculation errors that are not part of the scoring are made

· Interpretation parts are scored an E but are considered a weak E

0	1	2	3	4 ✓

Parts (a) through (d) sum to 4 points

OR

Parts (a) through (d) sum to $3\frac{1}{2}$ points AND a holistic approach is used to decide to score up

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect
- ☐ Part (d) essentially correct
- ☐ Part (d) partially correct
- ☐ Part (d) incorrect

Solution

Part (a)

Because the median number of attempts for players who received visualization training (4) is less than the median number of attempts for players who did not receive training (7), those who received visualization training tend to need fewer attempts to make two consecutive baskets.

Setting Up a Test for the Difference of Two Population Means Quiz

Scoring

Part (a) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

· An indication that players who received visualization training take fewer attempts before making two consecutive baskets than those who did not receive the training

· A justification based on location, such as the median number of attempts is less for the group that received visualization training or that the number of attempts required was typically less for the group that received visualization training

Partially correct (P) if the response satisfies only one of the two components

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

Note:

· To satisfy the second component, a response can compare the quartiles from the boxplots but should not compare the means since the means cannot be determined from the boxplots.

· Comments about variability are considered extraneous but may be used for holistic scoring.

Solution

Part (b)

One condition is that the treatments are randomly assigned. This condition is met because the psychologists randomly assigned the players to one of two groups. Another condition is that the values of the response variable for each treatment come from a normal distribution or, if not, that the group sizes are sufficiently large. Because the group sizes are small $(10 < 30)$ and because both sample distributions are strongly skewed to the right and the visualization boxplot has an outlier, this condition is not met.

Scoring

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following four components:

· Correctly states and verifies the random assignment conditionhas been met

· Correctly states that the sample sizes are not large (e.g., less than 30)

· Correctly states that the sample distributions are skewedor refers to the outlier

· Correctly states that the normal condition has not been met

Partially correct (P) if the response includes only two or three of the four components

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

Solution

Setting Up a Test for the Difference of Two Population Means Quiz

Part (c)

The observed difference in medians $(\mathrm{V} - \mathrm{NV})$ from the boxplots is $4 - 7 = -3$ attempts. Because 17 of the 100 trials in the simulation produced a difference in medians of -3 or less, the estimated *p*-value based on this simulation is approximately 0.17.

Scoring

Part (c) is scored as follows:

Essentially correct (E) if the response indicates that the observed difference in medians is -3 and that there are 17 dots at -3 or lower, resulting in an approximate *p*-value of $\frac{17}{100} = 0.17.$

Partially correct (P) if the response uses -3 , but not less than -3 , leading to a *p*-value of $\frac{4}{100} = 0.04$ because there are four dots at -3

OR

Partially correct (P) if the response indicates that the observed difference in medians is 3 and counts 17 dots at 3 or higher

OR

Partially correct (P) if the response indicates that the observed difference in medians is -3 and counts 17 dots at -3 or less, but does not give an estimate of the *p*-value

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

Notes*:*

· If the response miscounts the dots, the response can still be considered essentially correct if it is clear that the response is attempting to count the number of dots in the correct region.

· If a response indicates that the *p*-value is exact and not an estimate, this should be considered a negative for the purpose of holistic scoring.

Solution

Part (d)

Because the *p*-value of approximately 0.17 is greater than 0.05, there is not convincing evidence that players who receive visualization training take fewer attempts before making two consecutive baskets than those who do not receive the training, similar to the ones in the study.

Scoring

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

· Compares the *p*-value from part (c) to a significance level (e.g., $\alpha = 0.05$) or states that the *p*-value is not small

· States that there is not convincing evidence

· Provides context

Setting Up a Test for the Difference of Two Population Means Quiz

Partially correct (P) if the response satisfies only two of the three components

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

Note:

If the *p*-value is incorrect in part (c), the response in part (d) should be scored based on the incorrect *p*-value from part (c).

2. A reporter responsible for the food section of a magazine investigated the belief that grocery stores sell beef at a higher price in the fall than in the spring. The reporter selected independent random samples of grocery-store beef prices in November and April and computed the mean and standard deviation for the samples. Which of the following are the correct null and alternative hypotheses for the reporter's investigation, where μ_{F} represents the mean price of beef in the fall and μ_{S} represents the mean price of beef in the spring?

(A) $\text{H}_0 : \bar{x}_{\text{F}} - \bar{x}_{\text{S}} = 0$
$\text{H}_{\text{a}} : \bar{x}_{\text{F}} - \bar{x}_{\text{S}} < 0$

(B) $\text{H}_0 : \bar{x}_{\text{F}} - \bar{x}_{\text{S}} = 0$
$\text{H}_{\text{a}} : \bar{x}_{\text{F}} - \bar{x}_{\text{S}} > 0$

(C) $\text{H}_0 : \mu_{\text{F}} - \mu_{\text{S}} = 0$
$\text{H}_{\text{a}} : \mu_{\text{F}} - \mu_{\text{S}} \neq 0$

(D) $\text{H}_0 : \mu_{\text{F}} - \mu_{\text{S}} = 0$
$\text{H}_{\text{a}} : \mu_{\text{F}} - \mu_{\text{S}} < 0$

(E) $\text{H}_0 : \mu_{\text{F}} - \mu_{\text{S}} = 0$
$\text{H}_{\text{a}} : \mu_{\text{F}} - \mu_{\text{S}} > 0$ ✓

Answer E

Correct. The null hypothesis states that there is no difference between the population mean price of beef in the fall and the population mean price of beef in the spring, and the alternative hypothesis states that the population mean price of beef is greater in the fall than in the spring.

3. Hannah wanted to investigate whether there was a difference in the time spent in the checkout line between two grocery stores in a large city. She went to Grocery Store J on a Monday morning and recorded the time, in minutes, it took 30 customers to go through a checkout line. Then she went to Grocery Store K on Monday afternoon and recorded the time it took 30 customers to go through a checkout line. Hannah calculated the mean number of minutes for the customers in each line. She intends to conduct a two-sample t-test for a difference in means between the two stores.

Have all conditions for inference been met?

Setting Up a Test for the Difference of Two Population Means Quiz

(A) Yes, all conditions have been met.

(B) No, the data were not collected using a random method. ✓

(C) No, the sample sizes are greater than 10 percent of the population.

(D) No, the sample sizes are not large enough to assume normality of the sampling distribution.

(E) No, the distributions of the sample data are not approximately normal.

Answer B

Correct. There is no indication that the data were collected in a random manner. Hannah could have randomly selected the days, the times, and the checkout lines for collecting data from each store.

4. A study was conducted to investigate whether the mean price of a dozen eggs was different for two different grocery stores, Store A and Store B, in a large city. A carton of one dozen eggs from each store was randomly selected for each of 35 weeks, for a total sample size of 35 cartons from each store. The mean price of the 35 cartons was recorded for each store. The difference in the mean carton price for the stores will be calculated.

Which of the following is the appropriate test for the study?

(A) A one-sample z-test for a population proportion

(B) A one-sample t-test for a sample mean

(C) A matched-pairs t-test for a mean difference

(D) A two-sample t-test for a difference between population means ✓

(E) A two-sample z-test for a difference between population proportions

Answer D

Correct. Two random samples are selected on a quantitative variable, and the difference in the sample means will be calculated. The appropriate test is the two-sample t-test for a difference in population means.

5. A two-sample t-test for a difference in means will be conducted to investigate mean gasoline prices in two states. From each state, 45 gasoline stations will be selected at random. On the same day, the price of regular gasoline will be recorded for each selected station and the sample mean price for each state will be calculated.

Have all conditions for inference been met?

Setting Up a Test for the Difference of Two Population Means Quiz

(A) Yes, all conditions have been met. ✓

(B) No, the data are not collected using a random method.

(C) No, the sample sizes are greater than 10 percent of the population.

(D) No, the sample sizes are not large enough to assume the sampling distribution is approximately normal.

(E) No, the distributions of the sample data are not approximately normal.

Answer A

Correct. The data are being collected using a random method. The sample sizes are large enough (both 45) to support the assumption of normality for the sampling distribution of the difference in sample means. Also, it is reasonable to assume that 45 gasoline stations is less than 10 percent of all the gas stations in each state.

6. An experiment was conducted to determine whether the price of a golf club affected the distance a golfer could hit a golf ball. A sample of 60 golfers were randomly assigned to one of two groups, C or E. The 30 golfers in group C were given a club and told the price of the club was cheap; the 30 golfers in group E were given the same club and told the price of the club was expensive. In reality, there was no difference in price. The golfers used their assigned clubs to hit a golf ball as far as they could. The distance, in yards, that each golfer hit the golf ball was recorded, and the mean distance calculated for each group. A two-sample t-test for a difference in means will be conducted.

Which of the following statements are true?

I. The data were collected using random assignment.
II. The data were collected using random selection.
III. The distribution of the difference in sample means will be approximately normal.

(A) I only

(B) II only

(C) III only

(D) I and III only ✓

(E) I, II, and III

Answer D

Correct. The data were collected using random assignment since the 60 golfers were randomly assigned to treatment groups, so statement I is true. It was not indicated that the 60 golfers in the sample were randomly selected, so statement II is not true. The distribution will be approximately normal since each group size is large enough, so statement III is true.

Setting Up a Test for the Difference of Two Population Means Quiz

7. A study will be conducted to investigate whether there is a difference in pain relief for two brands of headache pills, N and P. Participants will be randomly assigned to one of two groups. One group will take pill N when they experience a headache, and the other group will take pill P when they experience a headache. Each participant will record the number of minutes it takes until relief from the headache is felt. The mean number of minutes will be calculated for each group.

Which of the following is the appropriate test for the study?

(A) A two-sample z-test for a difference between population proportions

(B) A two-sample t-test for a difference between population means ✓

(C) A matched-pairs t-test for a mean difference

(D) A one-sample z-test for a population proportion

(E) A one-sample t-test for a population mean

Answer B

Correct. Two random samples will be selected on a quantitative variable, and the difference in the sample means will be calculated. The appropriate test is the two-sample t-test for a difference in means.

8. A study will be conducted to investigate whether there is a difference in mean tail lengths between two populations of snow leopards. Random samples of leopards will be selected from both populations, and the mean sample tail length will be calculated for each sample.

Which of the following is the appropriate test for the study?

(A) A two-sample t-test for a difference between population means ✓

(B) A two-sample z-test for a difference between population proportions

(C) A one-sample z-test for a population proportion

(D) A one-sample t-test for a sample mean

(E) A one-sample t-test for a population mean

Answer A

Correct. Two random samples will be selected on a quantitative variable, and the difference in the sample means will be calculated. The appropriate test is the two-sample t-test for a difference in means.

Setting Up a Test for the Difference of Two Population Means Quiz

9. A company that packages salted and unsalted mixed nuts received a complaint that claimed that the company's salted packages contain more whole cashews than their unsalted packages do. The quality control department investigated the claim by randomly selecting a sample of 45 of each type of package, counting the number of cashews in each package, and finding the mean and standard deviation for both types of packages. Which of the following are the correct null and alternative hypotheses to test the complaint's claim, where μ_S is the mean number of cashews per package of salted nuts and μ_U is the mean number of cashews per package of unsalted nuts?

(A) $H_0 : \mu_S - \mu_U = 0$
$H_a : \mu_S - \mu_U > 0$ ✓

(B) $H_0 : \mu_S - \mu_U < 0$
$H_a : \mu_S - \mu_U > 0$

(C) $H_0 : \mu_S - \mu_U = 0$
$H_a : \mu_S - \mu_U \neq 0$

(D) $H_0 : \bar{x}_S - \bar{x}_U = 0$
$H_a : \bar{x}_S - \bar{x}_U > 0$

(E) $H_0 : \bar{x}_S - \bar{x}_U = 0$
$H_a : \bar{x}_S - \bar{x}_U < 0$

Answer A

Correct. The null hypothesis states that there is no difference in the mean number of whole cashews in the two types of packages, and the alternative hypothesis states that the mean number of cashews in the salted packages is greater than the mean for the unsalted packages.

10. Two siblings, Alice and Sean, are both convinced that they are faster than the other at solving a puzzle cube. They recorded the length of time it took them to solve the cube 18 times each during a one-month period. Then each calculated the mean amount of time and standard deviation, in minutes, for their times. Let μ_A equal the mean time it took Alice to solve the puzzle cube and μ_S equal the mean time it took Sean. Which of the following are the appropriate null and alternative hypotheses to test for a difference in time for the siblings to solve the cube?

(A) $H_0 : \mu_A - \mu_S = 0$
$H_a : \mu_A - \mu_S > 0$

(B) $H_0 : \mu_A - \mu_S < 0$
$H_a : \mu_A - \mu_S > 0$

(C) $H_0 : \mu_A - \mu_S = 0$
$H_a : \mu_A - \mu_S \neq 0$ ✓

(D) $H_0 : \bar{x}_A - \bar{x}_S = 0$
$H_a : \bar{x}_A - \bar{x}_S > 0$

(E) $H_0 : \mu_S - \mu_A = 0$
$H_a : \mu_S - \mu_A > 0$

Setting Up a Test for the Difference of Two Population Means Quiz

Answer C

Correct. The null hypothesis states that the two population means are equal, and the alternative hypothesis states that they are not equal.