1. A sociologist will conduct a two-sample t-test for a difference in means to investigate whether there is a significant difference, on average, between the salaries of people with bachelor's degrees and people with master's degrees. From a random sample of 32 people with a bachelor's degree, the average salary was \$55,000 with standard deviation \$3,500. From a random sample of 28 people with a master's degree, the average salary was \$58,000 with a standard deviation of \$4,000.

With a null hypothesis of no difference in the means, which of the following is the test statistic for the appropriate test to investigate whether there is a difference in population means (master's degree minus bachelor's degree) ?

(A) $t = \frac{(58{,}000 - 55{,}000)}{\sqrt{\frac{4{,}000}{28} + \frac{3{,}500}{32}}}$

(B) $t = \frac{(58{,}000 - 55{,}000)}{\sqrt{\frac{4{,}000^2}{28} + \frac{3{,}500^2}{32}}}$ ✓

(C) $t = \sqrt{\frac{(58{,}000 - 55{,}000)}{\frac{4{,}000^2}{28} + \frac{3{,}500^2}{32}}}$

(D) $t = \frac{(58{,}000 - 55{,}000)}{\sqrt{\frac{4{,}000 + 3{,}500}{28 + 32}}}$

(E) $t = \frac{(58{,}000 - 55{,}000)}{\sqrt{\frac{4{,}000^2 + 3{,}500^2}{28 + 32}}}$

Answer B

Correct. The numerator is the difference of sample means, or $58{,}000 - 55{,}000$. The denominator is the standard error of the difference in sample means, or $\sqrt{\frac{4{,}000^2}{28} + \frac{3{,}500^2}{32}}$.

2. A two-sample t-test for a difference in means was conducted to investigate whether defensive players on a football team can bench-press more weight, on average, than offensive players. The conditions for inference were met, and the test produced a test statistic of $t = 1.083$ and a p-value of 0.15.

Based on the p-value and a significance level of $\alpha = 0.05$, which of the following is the correct conclusion?

(A) Reject the null hypothesis because $0.15 > 0.05$. There is not convincing evidence that defensive players can bench-press more weight, on average, than offensive players.

(B) Reject the null hypothesis because $0.15 > 0.05$. There is convincing evidence that defensive players can bench-press more weight, on average, than offensive players.

(C) Fail to reject the null hypothesis because $0.15 > 0.05$. There is not convincing evidence that defensive players can bench-press more weight, on average, than offensive players. ✓

(D) Fail to reject the null hypothesis because $0.15 > 0.05$. There is convincing evidence that defensive players can bench-press more weight, on average, than offensive players.

(E) Fail to reject the null hypothesis because $0.15 > 0.05$. There is convincing evidence that defensive players can bench-press the same amount of weight, on average, as offensive players.

Answer C

Correct. Because $p > \alpha$, the null hypothesis should not be rejected, indicating there is not convincing evidence to support the alternative hypothesis.

3. Animal researchers studying cows and horses conducted a two-sample t-test for a difference in means to investigate whether grazing cows eat more grass, on average, than grazing horses. All conditions for inference were met, and the test produced a test statistic of $t = 1.664$ and a p-value of 0.0487.

Which of the following is a correct interpretation of the p-value?

(A) The probability that cows eat more grass than horses, on average, is 0.0487.

(B) The probability that cows eat the same amount of grass as horses, on average, is 0.0487.

(C) Assuming that the mean amount of grass eaten by cows is greater than the mean amount of grass eaten by horses, the probability of observing a test statistic of at most 1.664 is 0.0487.

(D) Assuming that the mean amount of grass eaten by cows is equal to the mean amount of grass eaten by horses, the probability of observing a test statistic of at most 1.664 is 0.0487.

(E) Assuming that the mean amount of grass eaten by cows is equal to the mean amount of grass eaten by horses, the probability of observing a test statistic of at least 1.664 is 0.0487. ✓

Answer E

Correct. The test is right-tailed, since the investigator is looking at whether cows eat more grass than horses. The p-value is the area under the t-curve to the right of $t = 1.664$. The area of 0.0487 represents the probability of obtaining a test statistic equal to or greater than 1.644 if the null hypothesis is true; that is, if the mean amounts eaten by cows and horses are equal.

4. City health officials will conduct a two-sample t-test for a difference in means to investigate whether stray dogs in the city have more fleas, on average, than do stray cats. A random sample of 22 stray dogs had a mean of 120 fleas and a standard deviation of 20 fleas. A random sample of 15 stray cats had a mean of 100 fleas and a standard deviation of 30 fleas.

Assuming a null hypothesis of no difference, which of the following is a correct test statistic for the test?

(A) $t = \frac{120-100}{\sqrt{\frac{20+30}{22+15}}}$

(B) $t = \frac{120-100}{\sqrt{\frac{20^2+30^2}{22+15}}}$

(C) $t = \frac{120-100}{\sqrt{\frac{20^2}{22}+\frac{30^2}{15}}}$

(D) $t = \frac{120-100}{\sqrt{\frac{20}{22}+\frac{30}{15}}}$

(E) $t = \sqrt{\frac{120-100}{\frac{30^2}{22}+\frac{20^2}{15}}}$

Answer C

Correct. The numerator is the difference in the sample means, or $120 - 100$. The denominator is the standard error of the difference in sample means, or $\sqrt{\frac{20^2}{22} + \frac{30^2}{15}}$.

5. A two-sample t-test for a difference in means was conducted to investigate whether the average time to swim a lap with the freestyle stroke is different from the average time to swim a lap with the butterfly stroke. With all conditions for inference met, the test produced a test statistic of $t = -2.073$ and a p-value of 0.042.

Based on the p-value and a significance level of $\alpha = 0.05$, which of the following is a correct conclusion?

(A) There is convincing statistical evidence that the average time to swim a lap with the freestyle stroke is less than the average time to swim a lap with the butterfly stroke.

(B) There is convincing statistical evidence that the average time to swim a lap with the freestyle stroke is different from the average time to swim a lap with the butterfly stroke.

(C) There is not convincing statistical evidence that the average time to swim a lap with the freestyle stroke is greater than the average time to swim a lap with the butterfly stroke.

(D) There is not convincing statistical evidence that the average time to swim a lap with the freestyle stroke is different from the average time to swim a lap with the butterfly stroke.

(E) There is not convincing statistical evidence that the average time to swim a lap with the freestyle stroke is less than the average time to swim a lap with the butterfly stroke.

Answer B

Correct. Since the p-value is less than the value of α $(0.042 < 0.05)$, the null hypothesis would be rejected. There is convincing statistical evidence to conclude the alternative hypothesis is correct; that is, the average swim times are different for a lap with the freestyle stroke and a lap with the butterfly stroke.

6. Researchers on car safety studied driver reaction time and cell phone use while driving. Participants in the study talked on either a hands-free phone or a handheld phone while driving in a car simulator. A two-sample t-test for a difference in means was conducted to investigate whether the mean driver reaction time between the two groups of participants was different. All conditions for inference were met, and the test produced a test statistic of $t = -2.763$ and a p-value of 0.03.

Which of the following is a correct interpretation of the p-value?

(A) Assuming that the mean reaction times for hands-free and handheld phones are equal, the probability of obtaining a test statistic less than -2.763 is 0.03.

(B) Assuming that the mean reaction times for hands-free and handheld phones are equal, the probability of obtaining a test statistic greater than 2.763 or less than -2.763 is 0.03. ✓

(C) Assuming that the mean reaction times for hands-free and handheld phones are not equal, the probability of obtaining a test statistic less than -2.763 is 0.03.

(D) Assuming that the mean reaction times for hands-free and handheld phones are not equal, the probability of obtaining a test statistic greater than 2.763 or less than -2.763 is 0.03.

(E) The probability that the mean reaction time for hands-free phones will be less than that for handheld phones is 0.03.

Answer B

Correct. The test is two-sided. The p-value is the combined area under the t-curve to the left of -2.763 and to the right of 2.763. The combined area of 0.03 represents the probability of obtaining a test statistic of at most -2.763 or at least 2.763 if the null hypothesis of no difference in mean reaction time is true.

7. A two-sample t-test for a difference in means will be conducted to investigate whether the average length of a cell phone call is shorter this year compared with 5 years ago. From a random sample of 35 phone call records this year, the average length was 25 minutes with a standard deviation of 4 minutes. From a random sample of 32 phone call records from 5 years ago, the average length was 27 minutes with a standard deviation of 5 minutes. The difference (this year minus five years ago) in means will be calculated.

With a null hypothesis of no difference in length, which of the following is a correct test statistic for the test?

(A) $t = \frac{25-27}{\sqrt{\frac{4^2}{35}+\frac{5^2}{32}}}$ ✓

(B) $t = \frac{25-27}{\sqrt{\frac{4}{35}+\frac{5}{32}}}$

(C) $t = \frac{25-27}{\sqrt{\frac{4^2}{35}}+\sqrt{\frac{5^2}{32}}}$

(D) $t = \frac{27-25}{\sqrt{\frac{4^2}{35}+\frac{5^2}{32}}}$

(E) $t = \frac{27-25}{\sqrt{\frac{4}{35}+\frac{5}{32}}}$

Answer A

Correct. The numerator is the difference (this year minus five years ago) in the sample means, or $25 - 27$. The denominator is the standard error of the difference of sample means, or $\sqrt{\frac{4^2}{35} + \frac{5^2}{32}}$.

8. To prepare for a certification exam, candidates can use one of two exam preparation books, book J or book K. A two-sample t-test for a difference in means was conducted to investigate whether the average score on the exam for candidates using book J is less than the average score for candidates using book K. With all conditions for inference met, the test produced a test statistic of $t = -1.356$ and a p-value of 0.101.

Based on the p-value and a significance level of $\alpha = 0.05$, which of the following is a correct conclusion?

(A) There is convincing statistical evidence that the average score of candidates using book J is equal to the average score of candidates using book K.

(B) There is convincing statistical evidence that the average score of candidates using book J is less than the average score of candidates using book K.

(C) There is not convincing statistical evidence that the average score of candidates using book J is less than the average score of candidates using book K.

(D) There is not convincing statistical evidence that the average score of candidates using book J is different from the average score of candidates using book K.

(E) There is not convincing statistical evidence that the average score of candidates using book J is greater than the average score of candidates using book K.

Answer C

Correct. Because $p > \alpha$, the null hypothesis is not rejected. There is not convincing evidence to support the alternative hypothesis.

9. Zoologists studying two populations of tigers conducted a two-sample t-test for the difference in means to investigate whether the tigers in population X weigh more, on average, than the tigers in population Y. Two independent random samples were taken, and the difference between the sample means was calculated. All conditions for inference were met, and the test produced a p-value of 0.02.

Which of the following is a correct interpretation of the p-value?

(A) The probability that the mean weight for tigers in population X is greater than that for population Y is 0.02.

(B) The probability that the mean weight for tigers in population X is equal to that for population Y is 0.02.

(C) Assuming that the mean weights for populations X and Y are equal, the probability of observing a difference equal to the sample difference is 0.02.

(D) Assuming that the mean weights for populations X and Y are equal, the probability of observing a difference as great or greater than the sample difference is 0.02.

(E) Assuming that the mean weight for population X is greater than the mean weight for population Y, the probability of observing a difference as great as or greater than the sample difference is 0.02.

Answer D

Correct. The test is right-tailed because in the investigation, the zoologists believe that one population weighs more than the other. The p-value is the area under the t-curve to the right of the test statistic created from the sample difference. In this case, the area of 0.02 is the probability of observing a difference equal to or greater than the sample difference if there is no difference in the mean tiger weights of the two populations.