Practice questions for the ALC test

1. A study of iron deficiency among infants compared samples of infants following different feeding regimens. One group contained breast-fed infants, while the infants in another group were fed a standard baby formula without any iron supplements. The data consisted of blood hemoglobin levels at twelve months of age for each infant in the study.

   What is the statistical setting for inference here?

   (A) one sample, inference about the mean μ
   (B) one sample, inference about the proportion p
   (C) two independent groups, inference about the difference of two means
   (D) two independent groups, inference about the difference of two proportions
   (E) inference about the mean of paired differences

2. A study of iron deficiency among infants compared samples of infants following different feeding regimens. One group contained breast-fed infants, while the infants in another group were fed a standard baby formula without any iron supplements. For each infant in the study, the data consisted of an indicator of whether or not the infant was iron deficient at twelve months of age. (An established medical definition of *iron deficiency* for this age group was used.)

   What is the statistical setting for inference here?

   (A) one sample, inference about the mean μ
   (B) one sample, inference about the proportion p
   (C) two independent groups, inference about the difference of two means
   (D) two independent groups, inference about the difference of two proportions
   (E) inference about the mean of paired differences

3. A study of iron deficiency among infants compared samples of infants following different feeding regimens. A large group of participating families with twin infants volunteered for the study. In each of the participating families, mothers had chosen one twin to breast-feed and the other twin to formula-feed. The data consisted of blood hemoglobin levels at twelve months of age for each infant in the study. **Note: You wouldn't really do this.**

   What is the statistical setting for inference here?

   (A) one sample, inference about the mean μ
   (B) one sample, inference about the proportion p
   (C) two independent groups, inference about the difference of two means
   (D) two independent groups, inference about the difference of two proportions
   (E) inference about the mean of paired differences

4. Suppose that at a state college, a random sample of 41 students is drawn, and each of the 41 students in the sample is asked to measure the length of their right foot in centimeters. A 95% confidence interval for the mean foot length for students at this college turns out to be $(21.709, 25.091)$. If instead a 90% confidence interval was calculated, how would it differ from the 95% confidence interval?

   (A) The 90% confidence interval would be narrower.
   (B) The 90% confidence interval would be wider.
   (C) The 90% confidence interval would be the same as the 95% confidence interval.

5. In a survey of likely voters among Democrats in the state of Florida conducted in July, 2022, the percentage reporting they planned to vote for Nikki Fried was 40%, which was close to what she got in the Democratic primary—35%. Which of the following is true regarding the two percentages?

   (A) They are both parameters.
   (B) They are both statistics.
   (C) The first one is a parameter, the second one is a statistic.
   (D) The first one is a statistic, the second one is a parameter.

6. From the previous question, consider the 95% confidence interval for the mean foot length for students at this college, which turned out to be $(21.709, 25.091)$. What does a 95% confidence interval for mean foot length tell us in this case? Select the best answer:

   (A) We are 95% confident that this interval contains the sample mean ($\bar{X}$) foot length.
   (B) About 95% of the state college students will have a foot length within this interval.
   (C) We are 95% confident that most state college students will have foot lengths within this interval.
   (D) If this experiment were repeated independently many times, about 95% of the confidence intervals obtained would contain the population mean foot length.

The next three items refer to the following situation: A research article reports the results of a new drug test. The drug is to be used to decrease vision loss in people with Macular Degeneration. The article gives a P-value of $.04$ in the analysis section. The next three items present three different interpretations of this P-value. Indicate whether each interpretation is valid or invalid.

7. The P-value can be interpreted as the probability of getting results as extreme as or more extreme than the ones in this study if the drug is actually not effective.

   (A) Valid
   (B) Invalid

8. The P-value can be interpreted as the probability that the drug is not effective.

   (A) Valid
   (B) Invalid

9. The P-value can be interpreted as the probability that the drug is effective

   (A) Valid
   (B) Invalid

10. Suppose a researcher is hoping to show that the results of an experiment are statistically significant. In the statistical analysis, which of the following P-values would the researcher prefer?

    (A) .01
    (B) .1
    (C) 1.0
    (D) P-values are not related to statistical significance.

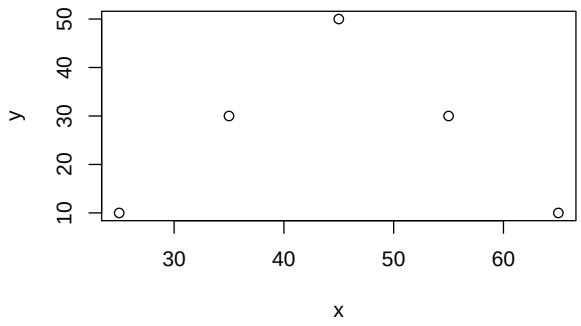
11. For the one-way analysis of variance setting, what does the null hypothesis of the ANOVA F test say?

    (A) That the sample means for all treatments in the study are equal.
    (B) That the population means for all treatments in the study are equal.
    (C) That the sample means for all treatments in the study are different.
    (D) That the population means for all treatments in the study are different.

12. A 95% confidence interval for the mean reading achievement score for a population of third grade students is $(43, 49)$. The margin of error of this interval is:

    (A) 5
    (B) 3
    (C) 6

13. Consider the scatterplot below. What would the correlation be?



    (A) -1
    (B) $-.5$
    (C) 0
    (D) $.5$
    (E) 1

14. Suppose the sample correlation between ages (in years) of husbands and wives in a certain community is $r = .8$. A researcher decides to convert ages to months by multiplying each age by 12. How does change the value of r?

    (A) The value of r remains $r = .8$.
    (B) The value of r is changed to $r = 0.0$
    (C) The value of r is changed to $r = 1.0$.
    (D) The value of r is changed to $r = -1.0$.

15. Suppose we find that the correlation between educational level attained and yearly income is $r = 0.72$. This finding means that (choose the single correct response):

    (A) Higher income is one result of higher educational level.
    (B) Lower income is associated with higher educational level.
    (C) People with lower educational levels tend to have lower incomes.
    (D) People with higher educational levels tend to have lower incomes.

16. There is a close relationship between the correlation r and the slope $\hat{\beta}_1$ of the least-squares regression line. In particular, it is true that

    (A) r and $\hat{\beta}_1$ always have the same sign, which shows whether the variables are positively or negatively associated.
    (B) r and $\hat{\beta}_1$ both always take values between -1 and 1.
    (C) the slope $\hat{\beta}_1$ is always at least as large as the correlation r.
    (D) the slope $\hat{\beta}_1$ is always equal to r^2, the square of the correlation.
    (E) Both A and B are true.

17. Consider the problem where we wish to estimate the population mean of Y at a particular value of x, and we also wish to predict an individual value of Y at that same particular value of x. Choose the best response among the following options.

    (A) The 95% confidence interval for μ_Y is wider than the 95% prediction interval for Y.
    (B) The 95% confidence interval for μ_Y is the same width as the 95% prediction interval for Y.
    (C) The 95% confidence interval for μ_Y is less wide than the 95% prediction interval for Y.

18. A regression line for predicting Internet usage ($\%$) for 39 countries is $\hat{Y} = -3.61 + 1.55x$, where x is the per capita Gross Domestic Product (GDP), in thousands of dollars, and Y is Internet usage. For one of the 39 countries, the per capita GDP is $\$15{,}000$, and the actual Internet use is 20 percent. For this country, the fitted (i.e. predicted) value is

    (A) 19.64%
    (B) 23246.39%
    (C) 27.39%

19. In simple linear regression, we talk about two equations: $\mu_Y = \beta_0 + \beta_1 x$, and $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$. Which of the following statements is true?

    (A) β_1 is a parameter; we know its value.
    (B) β_1 is a statistic; we do not know its value.
    (C) β_0 is a parameter; we do not know its value.
    (D) β_0 is a statistic; we know its value.

20. In simple linear regression, which of the following is **false** regarding the coefficient of determination R^2? Choose the single **incorrect** statement.

    (A) R^2 depends on the unit of measurement of the response variable Y.
    (B) $R^2 = r^2$, the square of Pearson's correlation coefficient.
    (C) If $R^2 = 1$, then each predicted $\hat{Y}$ value is equal to the corresponding observed Y value.
    (D) The closer R^2 is to 1 (i.e. 100%), the better the explanatory variable predicts Y.

Solutions

1. (C)
2. (D)
3. (E)
4. (A) The higher the confidence level, the wider the interval needs to be. And the lower the confidence level, the narrower the interval will be.
5. (D) The parameter: the percentage who voted for Fried, out of the actual voters. The statistic: the percentage who say they will vote for Fried, out of the sample.
6. (D) Part (A) is wrong because the interval certainly contains the sample mean; the sample mean is in the center of the interval.
7. (A) Valid
8. (B) Invalid. A P-value is not interpreted as a probabililty that the null hypothesis is true or false.
9. (B) Invalid. Same reason as for the previous question.
10. (A) Small P-values indicate statistical significance.
11. (B)
12. (B)
13. (C) In the formula for the correlation r, the contribution of the points sloping upward (left side of plot) would be positive, and the contribution of the points sloped downward (right side of plot) would be negative, and would exactly cancel each other out, because of the symmetry in the plot. So the correlation would be zero.
14. (A) The correlation r is not affected by linear transformations of either variable, or both.
15. (C)
16. (A)
17. (C)
18. (A)
19. (C)
20. (A)