LionTutors

www.LionTutors.com

STAT 200 – Exam 3 – Practice Exam Solutions

1. C – N (0.310, 0.071) – The normal distribution used to replace a bootstrap distribution should be centered at the original sample statistic (0.310), and the standard deviation is the standard error (0.071) of the bootstrap statistics.

   We do not need to use the formula for standard error if the standard error of the bootstrap statistics is already given.

2. B – 310 ± 2.539 x 4.02

   Confidence interval = Sample statistic ± Multiplier x Standard Error
   CI (μ) = $\bar{x}$ ± t* x SE($\bar{x}$)

   $\bar{x}$ = 310
   t* for 98% confidence = 2.539
   SE = 4.02 (Again, SE is given so we don't need to use the formula to calculate it.)

3. B – n = 20 and the population is skewed – The Central Limit Theorem (CLT) for means applies when the sample size is at least 30 or the population is normally distributed. Since neither of these conditions are met in choice B, the CLT does not apply.

4. B – The CLT for proportions (in other words the z-method conditions) apply when np ≥ 10 and n(1 – p) ≥ 10.

   In choice B, np = 50(0.10) = 5

5. B – z = $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.06 - 0.08}{\sqrt{\frac{0.08(1-0.08)}{100}}}$

6. A – t = $\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(14.6 - 9.4) - 0}{\sqrt{\frac{4.1^2}{72} + \frac{2.3^2}{76}}}$

7. C – Since this is a two-sided test (based on the alternative hypothesis), the p-value is the area above z = 1.75 and below –z = –1.75.

8. B – 0.64; Without actually calculating the p-value, we can eliminate choices A, C, and D.

   We know that the standard normal distribution is centered at 0 with a standard deviation of 1. That means that –0.35 is slightly to the left of the mean.

   The p-value is the area to the right of –0.35 (because of the greater-than alternative hypothesis), so we know that the p-value is more than 50% of the distribution, or more than 0.50. However, the answer cannot be 1 since the area to the left of –0.35 does not cover the entire distribution. That only leaves choice B, 0.64.

9. B – The p-value is the probability of getting the observed sample statistic, 0.09, or something more extreme in the direction of the alternative hypothesis, if the null hypothesis is actually true.

   In this example, "more extreme" means higher because the researcher is trying to prove that the proportion for females is <u>higher</u> than the proportion for males, or that $p_F - p_M > 0$.

10. C – We use the t-distribution because this situation deals with means rather than proportions.

    df = n – 1 = 20 – 1 = 19

11. A – We use the standard normal distribution because this problem deals with proportions.

    With the standard normal (z) distribution, we do not use degrees of freedom.

12. C – We use the t-distribution because this situation deals with means rather than proportions.

    df = n – 1 = 100 – 1 = 99

13. C – Test of two independent means because we are comparing a quantitative variable for two separate groups.

14. A – Test of one mean because there is one quantitative variable.

15. B – Test of paired means because we are comparing a quantitative variable for paired data. We know this is paired because the two lawn services are taking measurements for the same lawns.

16. E – Test of two proportions because we are comparing a categorical variable for two groups.

17. B – Mean because there is one quantitative variable.

18. A – We should build a confidence interval because we want to know how many students attend.

19. A – Proportion because there is one categorical variable.

20. B – We should conduct a hypothesis test because we want to know if the majority of Netflix account holders are content.

21. C – Difference in proportions because there is a categorical variable being compared for two groups.

22. B – We should conduct a hypothesis test because we want to know if there is a difference in approval ratings.

23. E – Mean of paired differences because there is a quantitative variable being measured with paired samples. We know it is paired because we are looking at the runner's race time with both Gatorade and water.

24. B – We should conduct a hypothesis test because we want to know if the times improve when consuming Gatorade.

25. D – Difference in means because there is a quantitative variable being compared for two independent groups. Note that "difference in means" implies that there are two independent means rather than using the mean of paired differences.

26. A – We should build a confidence interval because we want to know how different the salaries are.

27. B – Paired samples because researcher would measure the difference in weight loss for each pair of twins.

28. A – Scenario A uses paired means because each of the 20 participants is timed both before and after the camp. The researcher would be interested in the difference between the two measurements for each individual.

    Scenario B uses two independent means because an observation from the "7:30am class" sample would not be matched to an individual observation in the "8:30am class" sample.

29. Using a significance level of 0.05, the researchers should **not reject** the null hypothesis (because the p-value is greater than 0.05). It is possible that they have committed a **type 2** error, defined as **failing to reject** the null hypothesis when it is in fact **not true**.

30. Top picture – C – The standard normal distribution is a smooth, bell-shaped curve with a center of 0 and a standard deviation of 1.

    Middle picture – B – A normal distribution is a smooth, bell-shaped curve with any center and standard deviation.

    Bottom picture – A – A randomization distribution is a dotplot of the expected sample statistics that would occur if the null hypothesis is true. It is centered at the null value, and the standard deviation is the standard error of the sample statistics.

31. C – The 90% confidence level is the central area under the standard normal distribution. The x-axis shows $-z^*$ and z^*, the z-values with 90% of the curve between them.

32. D – 2 x (1 – 0.977)

    If this were a two-tailed test, we would look at the area in both extremes of the distribution. In this case, we would take the area below –2 and above 2.

    The area below –2 is equal to 1 minus the area above –2. Since the curve is symmetrical, we can multiply this area by 2 to include the area in both tails.

33. A – Ho: p = 0.50 can be rejected because 0.50 is not within the confidence interval. When the null hypothesis is rejected, statistical significance has been found.

    Since 0.60 and 0.70 are both within the confidence interval, B and C would not find statistical significance.

34. D – 0.55 to 0.61 is the narrowest interval, so it would correspond to the lowest confidence level.

35. C – n = 50
    A – n = 100
    B – n = 1,000

    The higher the sample size, the narrower the confidence interval.

36. B – False – When the p-value is greater than or equal to 0.05, the sample statistic is somewhat likely to occur even if the null is true.

37. A – True – The p-value is calculated assuming that the null is true.

38. B – False – 15 would be included in the interval because the null (Ho: mu = 15) cannot be rejected.

39. A – True – This is a two-sided test because the researchers just want to know whether the average <u>differs</u> from 15.

40. B – False

    95% confidence interval = 52% ± 5%
    95% confidence interval = 47% to 57%

    Based on this confidence interval, we cannot conclude that a majority of employed U.S. adults are expecting to retire by age 65 because the confidence interval is not entirely above 50%.

41. B – False – We do not know the exact population proportion with certainty. We can only use the confidence interval to estimate the proportion with 95% certainty.

42. A – True – This is an observational study because researchers are not assigning treatments; they are simply asking a question.

43. B – False – The confidence level relates only to the long-run confidence in the procedure. It <u>is not</u> the probability that a specific interval contains the parameter.

44. B – Your score on exam 1 is used to predict your score on exam 2. The predictor variable will be the variable listed in the predictor column of the Minitab output under the row with the information about the constant.

45. B – Positive square root of 0.439. You are given R-Sq (r^2) in the regression output. You are asked to find correlation (r). This means you need to take the square root of r^2.

    You know the answer is the positive square root because the coefficient for the predictor variable (Exam 1) is a positive number. This means that the regression line has a positive slope. When a regression line has a positive slope, correlation (r) will be positive.

46. A – Exam 2 = 45.1 + 0.469(Exam 1)

    The constant value (y-intercept) in the equation is the coefficient for the "constant" row in the regression output. The slope of the regression equation is the coefficient for the predictor variable (Exam 1).

47. C – 43.9% because it is the value of R-Sq.

48. B – Ho: $B_1 = 0$ Ha: $B_1 \neq 0$

49. E – Since the test statistic is 12.46 and the p-value is 0.000 you **reject** Ho and conclude that Exam 1 is a significant linear predictor of Exam 2.

50. B – For each one-unit increase in the x-variable, y-hat changes by the slope of 2.

51. B – If the x-variable increases by 11, y-hat increases by 11(2), or 22.

    You can also plug in numbers for the x-variable that are 11 inches apart:

    Weight = 40 + 2(Height)

    Weight for someone who is 60 inches tall:
    Weight = 40 + 2(60)
    Weight = 160

    Weight for someone who is 71 inches tall:
    Weight = 40 + 2(71)
    Weight = 182

    Difference:
    182 – 160 = 22

52. D – For each one unit increase in the explanatory variable, the response variable changes by the slope.

53. D – Correlation can only be obtained for two <u>quantitative</u> variables.

54. E – Regression (slope) test because the researcher is testing to see if there is a relationship between two <u>quantitative</u> variables. A correlation test would also be appropriate, but that isn't an answer choice.

55. C – The multiple linear regression equation uses the constant plus the coefficient times each explanatory variable.

56. A – The predicted final score will increase by the slope for each 1 point increase in the midterm score, assuming all other variables are held constant.

57. B – A dummy variable will either be equal to 1 or 0. Since the slope is -1.158, the predicted final score will decrease by 1.158 when the variable is present.

58. B – Quiz average is a significant linear predictor of final exam score because it has a p-value less than 0.05. This only holds true when all explanatory variables are present in the model.

59. D – DummyGender has a p-value greater than 0.05, so it is not a significant linear predictor when combined with the other explanatory variables in this model.

60. C – In the ANOVA table, we see that the p-value is less than 0.05. That tells us that at least one of the slopes in the model differs from 0, or is a significant linear predictor of final exam score. It **does not** tell us that all of the slopes differ from 0.

61. C – R-squared = 38.5%