Stat 2000 International – Sample Final - Answers

1. Correct answer: **b.** H_0: p = 1/2

2. Correct answer: **a.** 7.8

3. Correct answer: **c.** The population mean of the population is 0

4. Correct answer: **c.** A t-interval is a confidence interval estimate of the population mean

5. Correct answer: **d.** If the population mean is really 12, the probability is 0.006 that the mean of 23 values could be different from 12.

6. Correct answer: **b.** Sample #2 is large enough to use the z-test.

7. Correct answer: **d.** The means do not differ

8. Correct answer: **d.** There are likely to be confounding variables related to both alcohol consumption and depression

9. Correct answer: **a.** No, there is no linear relationship

10. Correct answer: **b.** 1.70

11. Answers:

- **a.** H_a*: Population correlation coefficient* $\neq 0$

- **b.** The probability of a Type I error is 0.05. We do not have enough information to decide what the probability of a Type II error is.

12. Answers:

- **a.** Minitab tells us that the meteorologist's output is "Test of mu = 80.00 vs mu < 80.00." That is, the null hypothesis, $H_0: \mu = 80$, is competing with the alternative hypothesis, $H_a: \mu < 80$.

- **b.** The question asked: "How likely is it that a sample of 17 people would have an average ideal temperature of 78.6 degrees or less if in fact the population average ideal temperature were 80 degrees? " is the P-value for this hypothesis test. The Minitab output tells us that the calculated P-value is 0.16.

- **c.** The p-value, 0.16, tells us that there is a fair chance of getting a sample average as small as 78.6, if the population average is 80. That is, we shouldn't be surprised at such a sample average. That is, if we consider a p-value below 0.05 small, then our p-value is not small enough, and therefore we cannot reject the null hypothesis.

d. There is not enough evidence to conclude that the ideal mean temperature for everyone in the population is below 80 degrees Fahrenheit.

13. The alternative hypothesis (H_A) is that the proportion of the population who can identify the unique beverage is greater than 0.33.

14. Linear Model B, that is, the points are generally farther from the line in model B than model A.

15. Answers:

a. **Null hypothesis:** The proportion of cars with at least 5 defects is constant across days of the week.
Alternative hypothesis: The proportion of cars with at least 5 defects is not the same for all days of the week.

b. **Null hypothesis:** Whether children beginning kindergarten can spell their names is independent of whether or not they attended preschool.
Alternative hypothesis: Whether children beginning kindergarten can spell their names depends on whether or not they attended preschool.

36. mean = 6.525, median = 6.5, standard deviation = 0.16931233

37. There are no outliers. But the boxplot indicates that there is some skewness towards the higher values.

38. We have to use a t-statistic (with 15 degrees of freedom) since the population standard deviation is not known and the sample size is relatively small (< 30). The p-value is 0.0009 which is less than 0.05. We reject H0, concluding that the mean milk-pH is clearly different from 6.7. The result is statistically significant.

39. The correlation is -0.9773. This seems to indicate that the points are very close to a line with negative slope.

40. The regression equation is:
milk-pH = 6.842 - 0.00749 * temperature

41. The predicted milk-pH value for 30 degrees is 6.617, the predicted milk-pH value for 100 degrees is 6.094. The first value (for 30 degrees) is more reliable since 100 degrees is 22 degrees above the highest observed temperature. We are extremely extrapolating here.

42. The Simple Linear Regression window already provides the answer. The t-stat value for temperature is -17.24 which relates to a p-value < 0.0001. We reject the null hypothesis. The result is statistically significant. We can conclude that milk-pH is indeed related to temperature.

43. No - for a temperature of 40 degrees, the predicted milk-pH value would be 6.54. The sample standard deviation of residuals is 0.037, meaning that about 95% of the data (when looking at a temperature of 40 degrees) should be in the interval 6.54 +/- 2 * 0.037, i.e., in the interval 6.466 to 6.614. So, an observed value of 6.60 would not be surprising.

44. The histogram shows two distinct modes - one in the 20-25 class, and one in the 45-50 class.

45. There seem to be 2 main clusters. Mark the lower cluster that relates to BMI values from 20 to 30 and Height values from 60 to 85 with your mouse and you will see that these points relate entirely to sport 2. The remaining points belong to sport 1.

46. It makes more sense for sport 1. The points related to sport 1 seem to scatter around a line that shows some positive trend. For sport 2, BMI (y) values seem to show no upward or downward trend with increasing Height (x) values.

47. The histogram of Height for sport 1 looks fairly symmetric, with a mode in the 70-75 class. The histogram of Height for sport 2 seems to be skewed towards the lower values with a mode in the 80-85 class and possibly an outlier in the 60-65 class.

48. Sport 1: BMI = 3.78 + 0.5779 * Height
Sport 2: BMI = 15.98 + 0.1133 * Height

49. For sport 1, the p-value associated with Height is 0.115 > 0.05,i.e., the slope is not significantly different from 0.For sport 2, the p-value associated with Height is 0.0064 < 0.05, i.e., the slope is significantly different from 0. Just based on our graphical interpretation, we expected the slope for sport 1 to be different from 0 and the slope for sport 2 to be equal to 0. The reason for our graphical misinterpretation is that in the full scatterplot that shows Height versus BMI for both sports, the data for sport 2 appears very condensed at the bottom of the graph, misleading our visual judgement.

50. around 24.48