STAT 200 – Exam 7 – Practice Exam Solutions

1. D – There is **no** relationship in the population between the treatment and whether or not someone gets malaria. The null hypothesis always states there is no relationship in the population.

2. B – There is a relationship in the population between the treatment and whether or not someone gets malaria. The alternative hypothesis always states there is a relationship in the population.

3. B – In the population, there is a statistically significant relationship between the treatment and getting malaria because the p-value is less than 0.05.

4. C – There is no relationship between illegal drug consumption and gender.

   It is incorrect to use = or ≠ symbols because illegal drug consumption and gender are not values.

5. C – Expected count = (Row total x Column total) / Overall total for the table

   Expected count = (334)(420)/629

6. A – Degrees of freedom (df) takes the number of row categories minus one times the number of column categories minus one. You do not include totals or headings, just the number of categories.

   df = (r – 1)(c – 1)
   df = (2 – 1)(2 – 1)
   df = 1

7. D – Fail to reject the null because the p-value is greater than 0.05. Conclude that there is no statistically significant relationship between gender and illegal drug consumption.

8. A – Since the pattern of OS used is the same for students who live on campus and students who live off campus, there is no support of a relationship between the two variables.

9. B – The primary purpose of a chi-square test is to determine if there is a statistically significant relationship between two **categorical** variables.

10. A result is considered statistically significant if the observed relationship in the **sample** is **large** enough that it is **unlikely** to have occurred by random chance if the **null** hypothesis is true, meaning that the **explanatory** variable **is** important in the **population.**

11. A – The table with the **largest p-value** would be the table that provides the **least** support for the alternative. All of these tables have the **same sample size**, so we are only looking at the strength of the observed relationship. To examine the relationship, we look at the difference in row percentages for males versus females.

    In Table A, there is no difference between the percentage of males and females who say "yes" or between the percentage of males and females who say "no". This means that regardless of gender, the same percentage of people have received a promotion. This would provide the **least** support for a relationship between gender and receiving a promotion, resulting in the **largest** p-value.

12. A – Asking for the table with the **smallest chi-squared statistic** is the same as asking for the table with the **largest p-value.** Therefore, we are looking for the table with the least support for the alternative again.

    In this example, the difference between the percentage of males and females who say they've received a promotion is the same for all tables. However, the tables have different sample sizes.

    The smaller the sample size, the **less support** for the alternative hypothesis.

13. D – All of these tables will have the same chi-squared statistic and p-value.

    All of these tables have the same sample size, so we can only look at the difference in row percentages.

    In each of these tables, there is **no difference** between the percentage of males and females who've received a promotion. Since the difference is the same for all of the tables, they would all provide the same chi-squared statistic and p-value.

    Specifically, the chi-squared statistic would be equal to 0 and the p-value would be 1 because we have no support for the alternative hypothesis.

14. D – 55; Actual counts are the top number printed in each cell.

15. B – 33; Expected counts are the second number printed in each cell.

16. D – Chi-squared contributions are the bottom number printed in each cell. The cell with the **highest** contribution to the chi-squared statistic is in the “male – no” cell.

17. A – Chi-squared contributions are the bottom number printed in each cell. The cell with the **lowest** contribution to the chi-squared statistic is in the “female – yes” cell.

18. B – There is **no** relationship in the **population**. The null hypothesis always states that there is no relationship, and the hypotheses are always about the population.

19. D – There **is** a relationship in the **population**. The alternative hypothesis always states that there is a relationship, and the hypotheses are always about the population.

20. C – For chi-squared tests, the p-value is the probability of getting our chi-squared statistic or any value **larger** if there is **no relationship** between the two variables in the population.

21. B – Because the p-value of 0.022 is **less than** 0.05, we reject the null and statistical significance **has** been found.

22. 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.

23. 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.

24. 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).

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

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

27. 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.

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

29. 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

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

31. 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.

32. 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.

33. 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.

34. 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.

35. 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 **<u>does not</u>** tell us that all of the slopes differ from 0.

36. C – R-squared = 38.5%

37. A – Scatterplots <u>cannot</u> prove causation.

38. B – IQ does the worst job of explaining the variation in cholesterol because it has the highest SSE.

    Although we haven't gone into much detail about SSR, SSE, and SSTO, we know that regression lines aim to minimize SSE (sum of squared errors). The lower SSE, the closer the points fall to a straight line, which means that the explanatory variable is doing a better job of predicting the response variable.

39. D – Weight does the best job of explaining the variation in cholesterol because it has the lowest SSE.

40. C – It is possible to determine both actual and expected counts.

    The best way to determine this is to try to put the data into a two-way table. The numbers in black are given in the problem, and the numbers in red can be solved for.

	High	**Low**	**Total**
Under 25	80 – 65 = 15	65	80
25 or Older	90	120 – 90 = 30	120
Total	15 + 90 = 105	65 + 30 = 95	80 + 120 = 200

    This shows us that we can determine actual (observed) counts. Anytime we have the actual counts, we can use the row totals and column totals to find the expected counts.

    Expected count = (Row total) x (Column total) / Overall n for the table

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

42. B – The statistical quality being represented is the slope because it gives the change in the response variable for each one unit increase in the explanatory variable.

43. D – Correlation can only be obtained for two quantitative variables.

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

45. F – Chi-square test because the researcher is testing to see if there is a relationship between two categorical variables.