

Test 3 SHORT ANSWER QUESTIONS

1. Explain why Colorado is probably doing better than Alabama in education despite the fact Alabama has a higher SAT average than Colorado?
2. Our scatter plot of states with percent taking SAT and SAT average would probably show what if we colored the southern states' dots a different color 70 years ago?
3. Our scatter plot of states with percent taking SAT and SAT average would probably show what if we colored the southern states' dots a different color today?
4. What is the notation for the sample linear correlation coefficient?
5. What is the notation for the population correlation coefficient?
6. What does the least squares line minimize?
7. What happens if you switch x and y when finding correlation?
8. What happens if you switch x and y when finding the regression line?
9. What happens to r if the units of measurement on x and/or y are changed?
10. Why does not r change if the units of measurement on x and/or y are changed?
11. If you are 1.43 standard deviations taller than the mean when measured in inches, how many standard deviations above the mean will you be when measured in centimeters?
12. The linear correlation coefficient is always between what two numbers?
13. If there is a negative relationship, then r will be negative in part because bigger than average x 's will correspond to _____ than average y 's making $\left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$ the product of a + and a - which is -.
14. If there is a negative relationship, then r will be negative in part because bigger than average x 's will correspond to smaller than average y 's making $\left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$ the product of a _____ and a - which is -.
15. If there is a negative relationship, then r will be negative in part because bigger than average x 's will correspond to smaller than average y 's making $\left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$ the product of a + and a _____ which is -.
16. If there is a negative relationship, then r will be negative in part because bigger than average x 's will correspond to smaller than average y 's making $\left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$ the product of a + and a - which is _____.
17. Does r measure the strength of the relationship between x and y ?
18. The only kind of relationship r measures is what?
19. Name three other relationships besides linear. Such as exponential, quadratic & _____
20. Name three other relationships besides linear. Such as exponential, logistic & _____
21. Name three other relationships besides linear. Such as logistic, quadratic, & _____
22. Is r sensitive to outliers?

23. Is the regression line sensitive to outliers?
24. A change in one standard deviation of x results in a change of ____ standard deviations in y .
25. What is the meaning of r^2 ?
26. If a scatter plot does not show a linear pattern can you still find the line of best fit?
27. If a scatter plot does not show a linear pattern should you still find the line of best fit?
28. If r is close to 1 or -1 is that enough of a reason to find the line of best fit?
29. Are predictions for y based on an x far beyond the range of x 's you have data for are reliable?
30. Predicting a y based on an x far beyond the range of x 's you have data for is called what?



31. Which scatter plot shows a stronger relationship?
32. Which scatter plot will have a higher value of r ?
33. If there is a strong correlation between x and y does that mean that changing x will most likely bring about a change in y ?
34. Give an example in which there is a strong association between x and y , but there is no cause and effect.
35. There is a strong relationship between elementary kids' grades and involvement in soccer, explain how this could be true even if there is no cause and effect.
36. Give an example in which there is a fairly strong linear correlation between x and y but there is another variable contributing to the differences in y besides x . Name the two variables.
37. People are often interested in how one variable affects another, give an example in which there are many variables involved and it is basically impossible to do so.
38. Do you think that people with an agenda will still try to show x affects y even if the setting is too complex with many variables interacting?
39. What was a possible lurking variable that would explain why it appears that smoking causes lung cancer despite a high correlation between smoking and lung cancer?
40. What is some really good evidence that there is not some gene that both causes lung cancer and nicotine addiction?
41. There is a strong correlation between education and wealth. Give a possible lurking variable that could explain this without having education have a cause and effect on wealth.
42. If a person is motivated they are likely to become wealthy and also become educated. Do you think that motivation explains all the association between education and wealth, so in fact there is no cause and effect?
43. If a person is motivated they are likely to become wealthy and also educated. Do you think that motivation explains part of the association between education and wealth, so in fact the cause and effect still exists, but it not a strong as many might think?
44. Give an example in which a lurking variable makes a cause and effect look weaker than it actually is.

45. Which scatter plot will have more scatter, or will they be about the same? A) SAT math vs SAT verbal for individual students, B) SAT math vs SAT verbal for state averages.
46. If you try to predict an individual student's SAT verbal from their SAT math using the regression line for state averages instead of individuals will the prediction be too high, too low, or about right?
47. If you try to predict an individual student's SAT verbal from their SAT math using the regression line for state averages instead of individuals will the prediction be more reliable, less reliable, or have about the right amount of reliability?
48. With categorical data a what takes the place of a scatter plot?
49. What is Simpson's Paradox?
50. Give an example of Simpson's Paradox.
51. With the O and E stuff, why is the rejection region is always to the right? It is because if H_0 is wrong then O and E will _____, making $\sum \frac{(O - E)^2}{E}$ big, which is to the right.
52. With the O and E stuff, why is the rejection region is always to the right? It is because if H_0 is wrong then O and E will differ a lot, making $\sum \frac{(O - E)^2}{E}$ _____ which is to the right.
53. The E's in the O and E stuff are found assuming what?
54. With the O and E stuff we want all the E's to be at least what to get good results?
55. The 4 assumptions for ANOVA are _____, equal variances, independent samples and SRS's.
56. The 4 assumptions for ANOVA are normal populations, _____, independent samples and SRS's.
57. The 4 assumptions for ANOVA are normal populations, equal variances, _____ and SRS's.
58. The 4 assumptions for ANOVA are normal populations, equal variances, independent samples and _____.
59. If you do a good job of collected data from different sources, the data will vary for only two reasons, those are _____ and luck.
60. If you do a good job of collected data from different sources, the data will vary for only two reasons, those are source and _____.
61. In ANOVA to reject H_0 : "all means equal" you hope the variance due to _____ is high and the variance due to error is low.
62. In ANOVA to reject H_0 : "all means equal" you hope the variance due to factor is high and the variance due to _____ is low.
63. Variance due to factor is a weighted _____ of the sample means.
64. Variance due to factor is a weighted variance of the sample _____.
65. Variance due to error is a weighted _____ of the sample variances.
66. Variance due to error is a weighted mean of the sample _____.
67. There are four things should you check in addition to graphing a scatter plot before calculating the least-squares line, those are that _____, the data are normally distributed about the line, the standard deviation is similar throughout the line, and the data is independent.
68. There are four things should you check in addition to graphing a scatter plot before calculating the least-squares line, those are that the data are linear, _____, the standard deviation is similar throughout the line, and the data is independent.

69. There are four things should you check in addition to graphing a scatter plot before calculating the least-squares line, those are that the data are linear, the data are normally distributed about the line, _____, and the data is independent.
70. There are four things should you check in addition to graphing a scatter plot before calculating the least-squares line, those are that the data are linear, the data are normally distributed about the line, the standard deviation is similar throughout the line, and _____.