

## QM213 WEEK 13 REVIEW QUESTIONS

---

1. The main purpose of ANOVA in regression of one variable on another is to \_\_\_?
2. A production engineer is interested in the possibility that the mean percentage of flawed products on a production line (F) is a linear function of the number of hours of operation (H) per day for the lines, which are set by policy for each line. She best can analyze this possible relation using a \_\_\_?
3. The F-test in an ANOVA with regression is 2.981 with 1 and 18 degrees of freedom. How should you calculate the probability that the regression coefficient is zero using the usual EXCEL function (assume that each function shown in the answers starts with an equal sign)?
4. The analysts for McDavids fat-food emporium and integrated cardiology clinics discover through intensive study that the higher the fat content of foods, the higher their sales figures for those foods. The results are extremely statistically significant, with  $P(H_0) < 0.001$ . Which of the following is most likely to be a valid statement about the regression coefficient resulting from their analysis?
5. A production engineer is interested in the possibility that the mean percentage of flawed products on a production line (F) is a linear function of the number of hours of operation (H) per day for the lines, which are set by policy for each line. She defines the change in percentage of flaws for each hour increased as  $_b$  and the number of flaws theoretically occurring even with no hours of operation as  $_a$ . She represent this linear relation as \_\_\_?
6. A regression analysis of the amount of wasted coffee/employee produced per day results on the number of employees in a corporation generates the following F test for the  $H_0$ :  $F[1,42]$  equals 12.345. How could you compute  $P(H_0)$  for the null hypothesis using Excel?
7. A regression analysis of the number of trades of a stock in a day on the amount of news coverage (in words) for that day located through the GIGGLE search function for each stock results in a Y-intercept of 10,000 and a regression coefficient of 0.8 which, based on the sample size, turns out to have a probability of only 1 in a million of being the results of chance if the parametric regression coefficient were 0. What is a reasonable implication of this regression coefficient?
8. A regression analysis of the amount of wasted paper/employee produced per day in several divisions of a company results on the number of photocopiers and printers in a corporation generates the following F test for the  $H_0$ :  $F[1,8]$  equals 4.210. Calculate the probability of getting this result (or a more extreme result) by chance alone if  $H_0$  were true and choose the correct answer for  $P(H_0)$ .
9. In the formula for a straight line relating Y to X, the slope is defined as \_\_\_?
10. In regression of one variable on another, the dependent variable is one that is \_\_\_?
11. An economist is studying the annual income figures for a large sample of the population on Jove and comparing them with similar data for Io, Calisto and Phobos. The model tries to find a predictive linear equation between the annual income per person and the population of the moon. What is the appropriate statistical methodology for this analysis?
12. An economist is studying the annual income figures for a large sample of the population on Jove and comparing them with similar data for Io, Calisto and Phobos to compare the means. What is the appropriate statistical methodology for this analysis?
13. In a scattergram, the expected value of the ordinate when the abscissa is zero is called the \_\_\_?

## QM213 WEEK 13 REVIEW QUESTIONS

---

14. A regression analysis between number of votes (Y) for a candidate and amount of money (X) in credits spent on political ads on Pluto results in the following equation  $\hat{Y}$  equals  $643,229 + 0.05X$  which turns out to have a highly statistically significant regression. The above equation implies that \_\_\_?
15. The MS<sub>regression</sub> in an ANOVA with regression is 1293.33 with 1 degree of freedom; the MS<sub>residual</sub> with 75 degrees of freedom is 306.2. What is the F-test for significance of this regression?
16. When using regression analysis where  $Y = a + bX + e$ , the confidence intervals for a predicted Y at a specific value of X \_\_\_?
17. In the regression model M equals  $rS + n$ , where S is the independent variable, which variable represents the Y-intercept?
18. In the regression model R equals  $i + wT$ , where w is the slope, which variable will be chosen for the abscissa on a graph of the data?
19. The analysts for McDiarmid's slow-food chain of restaurants and in-house dental clinics discover through intensive study that the higher the sugar content, the higher their overall profits. The results are highly statistically significant, with  $P(H_0)$  equals 0.007. Which of the following could reasonably turn out to be the regression coefficient in their analysis?
20. A regression analysis between number of birds seen flying over an observation point in a day and the temperature in Celsius at the observation point results in a regression coefficient  $_b_$  which equals 10 which, based on the sample size, turns out to have a probability of 1 in a thousand of being the results of chance if the parametric regression coefficient were 0. The Y-intercept is 23. What is a reasonable interpretation of this regression analysis?
21. The change in Y when X increases by one unit is measured by the \_\_\_?
22. The MS<sub>regression</sub> in a regression ANOVA is 42.7 with 1 degree of freedom; the MS<sub>residual</sub> with 52 degrees of freedom is 12.8. What is the F-test for significance of this regression?
23. In regression of one variable on another, if the F value is significant [e.g.,  $P(H_0)$  is .05 or less], then we \_\_\_?
24. The F-test for regression with 1 and 112 degrees of freedom is 3.871 Calculate the probability that the regression coefficient is zero using the usual EXCEL function and choose the correct answer.
25. The F-test for regression with 1 and 17 degrees of freedom is 2.562. Calculate the probability that the regression coefficient is zero using the usual EXCEL function and choose the correct answer.
26. The average squared difference between observed and predicted values in a regression analysis can be called the \_\_\_?
27. Another name for the unexplained error in a linear regression model is the \_\_\_?
28. In the regression model G equals  $t + mU$ , where m is the regression coefficient, which variable will usually be shown on the Y-axis?
29. In the formula for a straight line relating Y to X, the Y-intercept is known as \_\_\_?
30. A regression analysis between number of votes (Y) for a candidate and amount of money (X) in dollars spent on political ads on Mars Colony results in the following equation:  $\hat{Y}$  equals  $325,000 + 0.05X$ . The regression coefficient is highly statistically significant. The above equation implies that \_\_\_?
31. In the regression model N equals  $qR + p$ , where R is the independent variable, which variable represents the slope?
32. In regression of one variable on another, the independent variable is one that is \_\_\_?

## QM213 WEEK 13 REVIEW QUESTIONS

---

33. In the regression model  $P$  equals  $q + rS$ , what will the value of  $P$  be when  $S$  equals 0?
34. A regression analysis between sales ( $Y$ ) and price ( $X$ ) in dollars results in the following equation:  $\hat{Y}$  equals  $50,000 - 8,000X$ . The regression analysis has a highly statistically significant regression coefficient. The above equation implies that \_\_\_\_?
35. An atmospheric scientist wants to see if there is any effect of carbon dioxide concentration in the atmosphere on mean air temperatures over the South Pole in Antarctica. The scattergram does not seem to follow any sort of straight line relationship. What is the best initial analytical technique she should use in testing the idea that there might be some effect (not necessarily linear) of  $CO_2$  on temperature?
36. In the regression model  $P$  equals  $f + nQ$ , where  $n$  is the slope, which variable will be chosen for the  $X$ -axis in a scatterplot?
37. In the regression model  $R$  equals  $i + wT$ , where  $w$  is the slope, which variable will be chosen for the ordinate on a graph of the data?
38. A regression analysis of the incidence of violent headaches on dosage of a new drug, Morestaticol, results in a regression coefficient 0.9 and a  $Y$ -intercept of 12. Based on the sample size, the results of the analysis turn out to have a probability of 0.0002 of being the results of chance if the parametric regression coefficient were 0. Which of the following could have been the statistic that resulted in this  $P(H_0)$ ?
39. An economist is studying the industrial productivity figures for a large sample of the population on Jove and seeing if productivity is linearly related to the ages of the sample participants. What is the most appropriate statistical methodology for this analysis among the options shown?

