Stat 2300 International – Sample Midterm - Answers

1. Correct answer: **[b]** there is no linear relationship between the two variables.

2. Correct answer: **[d]** The percent of students not having an algebra background that received a grade of "A, B, C," was more than 10% higher than the percent of students having an algebra background that received a grade of "A, B, C,".

3. Correct answer: **[b]** 0.25

4. Correct answer: **[c]** 0.8664

5. Correct answer: **[d]** Neither hand spans of students, nor land areas in square miles. [In this case, neither could be exponential. The factors that tend to yield exponential data are a constant rate of incidents and independence of incidents. Neither is present.]

6. Correct answer: **[b]** 2 and 22

7. Correct answer: **[c]** 0.05

8. Correct answer: **[a]** A standard deviation of $100

9. Correct answer: **[b]** This is not a paired data design, because two different groups of cars, unpaired or unmatched in any way, receive the treatments - a gasoline additive or no additive.

10. Correct answer: **[a]** 0.01

11. Correct answer: **[c]** This residual plot indicates an inequality of standard deviations of the residuals. The residuals fan out as X gets larger, indicating an increasing standard deviation of the residuals.

12. Correct answer: **[c]** $421.4

13. Correct answer: **[a]** The comparability of treatment groups is controlled and results in improved precision.

14. Correct answer: **[b]** 79.13 divided by 20

15. Correct answer: **[b]** 11.7%

16. Correct answer: **[c]** Both of the options presented

17. Correct answer: **[b]** the experimentwise error rate

18. Correct answer: **[b]** indicate the presence of special causes of variation

19. Correct answer: **[d]** The process mean is not in control, we can not use the center line as a valid estimate of process centering.

20. Correct answer: **[c]** approximately 3.87 grams [Since UCL = CL + 3(SE), SE = (9.84 - 4.65)/3 = 1.73. The true standard deviation of the sample mean is approximately 1.73 grams. Let s denote the true standard deviation of the weights. Since samples of size 5 were used, the standard deviation of the sample mean is s/sqrt(5). Thus, s is approximately 1.73*sqrt(5), or 3.87 grams.]

21. Answer: (i) zero (ii) spectral methods.

22. Answer: The use of helicopters with 5 different wing-lengths could possibly result in different flight times. Therefore degrees of freedom for (i) wing-lengths = 5-1 = 4. The flying times were recorded for a total of 150 flights. Therefore the degrees of freedom for (iii) Total = 150 - 1 = 149. (ii) Error degrees of freedom = Total df - Factor df = 149 - 4 = 145.

23. Answer: P(no customer | lambda = 5.5) is found using the Poisson calculator to be 0.0041.

24. Answer: About 99.7% of the vehicle speeds at this location are between 61-3(6) and 61+3(6), that is between 43 and 79.

25. Answer: (i) 0.5 / sqrt(100) = 0.05 and (ii) (11.78 - 12) / 0.05 = -4.4

26. Answer: civil-eng: 230.4, chem-eng: 194.75, elect-eng: 205.25

27. Answer: H0: There are no differences in the population mean expenditure on textbooks by students of three engineering majors. vs. Ha: The population mean expenditure on textbooks is different for two or more engineering majors.

28. Answer: MS Treatments = SS Treatments / df Treatments = 3064.0693 / 2 = 1532.0347

29. Answer: The p-value is 0.3297 (which obviously is > 0.05). This means, we fail to reject the null hypothesis, i.e., there are no differences in the population mean expenditure on textbooks by students of these three engineering majors.

30. Answer: In CyberStats Unit E-3, Uses 2, the p-value is 0.002 (which is < 0.05), i.e., we have a significant result and reject the null hypothesis there. Apparently, the addition of comp-eng dramatically changes the result. One might hypothize (but still needs to formally test) that it is comp-eng where students have higher textbook expenditures than in any of the other three majors combined.

31. Answer: The p-value is < 0.0001 (which is < 0.05). This means the result is highly statistically significant, i.e., we reject the null hypothesis that all slopes are 0 at the same time.

32. Answer: The p-value associated with AirPressure is 0.0916 (which is > 0.05) and the p-value associated with log(AirPressure) is 0.1598 (which is > 0.05). Thus, we fail to reject that the slope associated with AirPressure is equal to 0 and we also fail to reject that the slope associated with log(AirPressure) is equal to 0.

33. Answer: A scatterplot between AirPressure and log(AirPressure) shows that these two variables are highly correlated. In fact, the correlation coefficient between these two variables is 0.99812347. This is the problem reported in Unit D-4, Warnings 2: Given either one of the two variables, the other does not explain a significant amount of the BoilingPoint.

34. Answer:
(i) BoilingPoint = 155.29648 + 1.9017835 * AirPressure
(ii) BoilingPoint = 49.033447 + 47.897335 * log(AirPressure)

35. Answer: The two models are very similar. In both, the parameters (intercept and slope) are highly statistically significant (p-values < 0.0001). In model (i), we have a slightly bigger correlation (0.9972 vs. 0.997) and a slightly smaller estimate of sigma (0.4440299 vs. 0.4585313). Also, model (i) does not require to do a log-transformation of the explanatory variable (which easily can lead to rounding errors). Therefore, model (i) should be used to predict BoilingPoint when both AirPressure and log(AirPressure) are given.