

Stat 2300 International – Midterm - Answers

1. Correct answer: [2] Simpson's paradox.
2. Correct answer: [1] There were fewer people graduating from college than not graduating from college.
3. Correct answer: [3] 5,600,000
4. Correct answer: [2] about 86.7%
5. Correct answer: [3] Situation II could be exponential after shifting, but not Situation I.
6. Correct answer: [2] 3 and 20 df
7. Correct answer: [2] about 0.056
8. Correct answer: [2] Variability in the sampling distributions has decreased.
9. Correct answer: [1] The two-sample procedure could lead to the wrong conclusion by producing a wider confidence interval than it should and failing to detect a difference in the two means, when it should.
10. Correct answer: [5] about 0.5435 [se = $\sqrt{18.311/(64-2)} = 0.5435$]
11. Correct answer: [5] New Orleans, La.
12. Correct answer: [3] 0.390 thousand dollars
13. Correct answer: [2] The natural variation of the response variable.
14. Correct answer: [2] It is a technique for splitting the variation among responses into its different components.
15. Correct answer: [3] The F-ratio is computed as the ratio of Factor (Exhaust) Mean Square to the Error Mean Square.
16. Correct answer: [2] Are the differences in the population mean volatilities supported by the data, or could the observed differences in sample mean volatilities have occurred by chance?
17. Correct answer: [3] The experimentwise error rate increases as the number of population means decreases.
18. Correct answer: [2] Only common cause variation present.

- 19.** Correct answer: [4] If no processing changes are made, and if no outside sources of variability enter the process, we have reason to believe that an estimate will be valid for the average thickness of metal wires made in the near future.
- 20.** Correct answer: [1] Test 1 does apply; we can not use the sample data to obtain a valid estimate of the process variability until the ranges are within the control limits.
- 21.** Answer: (i) seasonal (ii) cyclical
- 22.** Answer: Use of 4 different fertilizers could possibly result in different yields. Therefore degrees of freedom for (i) fertilizers = 3. Total 20 yield observations were taken. Therefore the degrees of freedom for (iii) total = $20 - 1 = 19$. (ii) Error degrees of freedom = Total df - Factor df = $19 - 3 = 16$.
- 23.** Answer: (i) One crash every three years is $1/36$ per month. The chance that X is more than 0 given $\lambda = 1/36$ is found using the Poisson calculator to be (ii) 0.0274.
- 24.** Answer: The z-score corresponding to the 80th percentile is 0.84. (In the probability calculator, select "to the left of". Then type "0.80" in the space to the immediate left of "Compute!".) Therefore, an applicant must score better than a $70 + 0.84 * 8 = 76.72$.
- 25.** Answer: (i) $1.12 / \sqrt{250} = 0.071$ and (ii) $(0.512 - 0.385) / 0.071 = 1.79$
- 26.** Answer: Jenny: 329.4, Paul: 343.2, Lee: 328.2, Lara: 344.6
- 27.** Answer: H_0 : There are no differences in the population (i.e., all 4 hour shifts of these employees) mean number of customers for these 4 employees. vs. H_a : The population mean number of customers per 4 hour shift is different for two or more of these 4 employees.
- 28.** Answer: $MS\ Error = SS\ Error / df\ Error = 2130 / 16 = 133.125$
- 29.** Answer: The p-value is 0.0687 (which obviously is > 0.05). This means, we fail to reject the null hypothesis, i.e., there are no differences in the population mean number of customers served per 4 hour shift by these 4 employees.
- 30.** Answer: $df = 8$, t-statistic = 2.6604314, p-value = 0.0144. Because the p-value is < 0.05 , the result is statistically significant. We reject the null hypothesis, i.e., Laura indeed does serve more customers per 4 hour shift than Lee (and therefore Laura should be the one who gets promoted).
- 31.** Answer: F-statistic = 21.398153, p-value < 0.0001 . Because the p-value is < 0.05 , the result is highly statistically significant. We reject the null hypothesis that all slopes are 0 at the same time.

32. Answer: The p-value associated with Poverty is <0.0001 (which is < 0.05), the p-value associated with Crime is 0.475 (which is > 0.05), the p-value associated with College is 0.0085 (which is < 0.05), and the p-value associated with Metro is 0.045 (which is < 0.05). Thus, we fail to reject that the slope associated with Crime is equal to 0, meaning that this variable should not be included in the final multiple regression model. We reject that the slopes associated with Poverty, College, and Metro are equal to 0, meaning that these three variables might be included in the final multiple regression model.

33. Answer: While residuals are nicely arranged around 0 for incomes up to 40,000, there seems to be a positive trend afterwards, meaning that higher residuals are associated with higher incomes. The most obvious outlier is AK (= Alaska) with an Income of 52779 and a residual of 11854.657.

34. Answer: The income for AK (= Alaska) is by far the highest income - so we are extremely extrapolating here. An economic reason for this is the major oil industry and tourism that are associated with high average incomes, while standard factors such as Poverty, College, and Metro have less impact for this state.

35. Answer: As previously reported, all variables but Crime are significant. The resulting model (without Crime) is

$$\text{Income} = 50657.273 - 753.8992 * \text{Poverty} - 175.09763 * \text{College} + 71.96461 * \text{Metro}.$$

Because all p-values are now < 0.05 , this should be our final model to predict Income from other socio-economic factors.