

Stat 2000 International – Final - Answers

1. Correct answer: [1] Students who participated in the survey.
2. Correct answer: [4] The responses you receive should not be considered representative of all shoppers in the chain because you spoke only with shoppers who were willing to participate in the survey.
3. Correct answer: [2] cholesterol level
4. Correct answer: [4] The set of 40 answers you receive when you ask the students how many hours they spend on the Internet per day.
5. Correct answer: [3] Both preceding statements are true.
6. Correct answer: [4] The outlier is a person who has about 8 brothers and weighs about 160 pounds.
7. Correct answer: [2] about 0.65
8. Correct answer: [3] about 0.1061
9. Correct answer: [4] 0.7
10. Correct answer: [3] 0.75
11. Correct answer: [2] about 0.0297
12. Correct answer: [3] For each sample size, the distribution is centered at about the same point, which also is approximately the population mean. For $n = 25$, there is less variation among the different sample means than when $n = 9$.
13. Correct answer: [3] \$0.56
14. Correct answer: [2] about 0.5
15. Correct answer: [4] The manager can be 95% confident that the true average travel time between customers is between 21.63 and 22.91 minutes.
16. Correct answer: [2] $H_a: p < 1/2$
17. Correct answer: [3] about 0.1538
18. Correct answer: [4] $H_0: \text{Population mean} = 400$; $H_a: \text{Population mean} < 400$

19. Correct answer: [2] Since the two confidence intervals overlap, we can be confident in concluding that the average stress level for females does not differ from the average stress levels for males.
20. Correct answer: [ANY] No correct answer provided - full credit for everyone. [Possible correct answers might be "If the population mean is really 23, the probability is 0.50 (and not 0.0003) that the mean of the 36 ages could be more than 23." or "If the population mean is really 23, the probability is 0.0003 that the mean of the 36 ages could be more than 25.16."]
21. Correct answer: [4] Sample No. 4
22. Correct answer: [1] A significance test of the null hypothesis that the means are equal.
23. Correct answer: [1] The p-value would increase.
24. Correct answer: [5] 10 pounds
25. Correct answer: [1] 2.37
26. Answer: Randomly select and survey a sample of students from the list of all students who attend the school.
27. Answer: (a) Median weight about 250 pounds. (b) Heaviest person about 520 pounds. (c) About 25% above 330 pounds.
28. Answer: 4. None of the above
29. Answer: Test results are independent so we use:
 $P(\text{at least one of them positive}) = 1 - (0.15) * (0.15) = 0.9775$
30. Answer: You can start with the "Graphing Normal z-Score/Probability" calculator for example. Enter 64 for the mean and 2.6 for the standard deviation. Calculate the "Area left of" 62 (0.2209) and "Area right of" 70 (0.0105) and then subtract these two numbers from 1.0. This will tell us that the proportion of women between 62 and 70 inches tall is $1.0 - (0.2209 + 0.0105) = 0.7686$.
31. Answer: You can start with the "Graphing Normal z-Score/Probability" calculator for example. Enter 64 for the mean and 2.6 for the standard deviation. Choose "Area left of" and type in the area (the red number) as 0.70. Then click on "Compute!" and the answer will be that 65.36 is the 70th percentile.
32. Answer: Assign the numbers 1 to 30 to the thirty students. In some way, randomly select 15 numbers between 1 and 30. The students with those numbers are assigned to the Web-based approach. The other students are assigned to the textbook approach.

33. Answer: The hypothesized value 183.5 is not in the confidence interval. Reject the null hypothesis - the mean is not 183.5.
34. Answer: The mean blood pressure in the population of women is higher than 115. The p-value is small (smaller than 0.05) so we reject the null hypothesis. The result is statistically significant.
35. Answer: The confidence interval suggests that the difference in the means could be equal to 0. Therefore the first mean and the second mean may in fact be identical.
36. Answer: treat-A = 6.4625, treat-B = 12.3375, treat-C = 27.35
37. Answer: treat-A = 6.35, treat-B = 12.05, treat-C = 27.9
38. Answer: The three means are very similar to the corresponding three medians. This is a first indicator that the distributions of weight losses are symmetric around the corresponding means.
39. Answer: According to the boxplots, there is one outlier: The smallest value 4.1 for treat-B is far below the lower fence. The boxplots for treat-A and treat-C indicate that the data is fairly symmetric. The boxplot for treat-B indicates that the data is skewed towards the higher values.
40. Answer: We have to use a t-test (with 7 degrees of freedom) here since the sample size is relatively small (< 30) but the data seems to approximately follow a normal curve. The p-value is 0.1047 which is obviously more than 0.05. We fail to reject H_0 , concluding that the mean weight loss under treatment C may indeed be 35 pounds.
41. Answer:
 99% Confidence interval results (treat-A): (2.4812884, 10.443711)
 99% Confidence interval results (treat-C): (7.978941, 46.721058)

Since the two confidence intervals overlap, there is indeed a (slight) chance that the two treatments are equally good, i.e., that the mean weight loss under treatment A is the same as the mean weight loss under treatment C.

42. Answer: Use a Graphing t-Distribution calculator (with 7 degrees of freedom) and determine a value such that an area of 0.995 is to the left of this value and an area of 0.005 is to the right. The value 3.499 satisfies this condition and is the multiplier (for a 99% CI) we need.

The standard error can be calculated as follows:
 $\text{standard deviation} / \sqrt{n} = 3.2177799 / \sqrt{8} = 1.1376569$

43. Answer: We are using a chi-Squared test with 7 degrees of freedom. The p-value is 0.3279 which is greater than 0.05. We fail to reject the null hypothesis that the variance of all weight losses under treatment A of all customers under this treatment is 9.

44. Answer: There is a linear trend, with higher SAT scores associated with higher GPA scores. There seems to be no outlier.
45. Answer: The correlation is 0.70355475. This indicates that there is a quite strong positive association and also that the points are quite close to a line with positive slope.
46. Answer: $GPA = -9.480994 + 0.011256661 * SAT$

The p-value associated with SAT is <0.0001 (which obviously is < 0.05), i.e., the slope is significantly different from 0. Therefore, we can use the regression equation to predict GPA scores from SAT scores.

47. Answer: For 1000: around 1.7756671; for 1100: around 2.901333; for 1200: around 4.0269995.

The first and the last value (for SAT scores of 1000 and 1200) are least reliable. We are extremely extrapolating for the first value (a SAT score of 1000 is more than 40 below the smallest observed value - this person may not be admitted to the university at all). The predicted value for 1200 is above the maximum possible value of 4.00 for GPA scores and therefore also is not reliable.

48. Answer: The histogram is skewed towards the lower values. There are no outliers. The modal bar is the one from 3.0 to 4.0 (or, more precisely, from 3.001 to 4.001).
49. Answer: When assuming independence, this relates to a $\text{Bin}(20,000, 650/2,300,000)$ distribution. The expected value is: $n * p = 20,000 * (650/2,300,000) = 5.65$.
50. Answer: No - this is not possible. As stated before, the maximum SAT score is 1600 while the maximum GPA score is 4.00. The regression equation already predicts a GPA score higher than 4.00 for a SAT score of 1200. Clearly, it is not possible to use this equation to predict GPA scores for all applicants to Utah State University, in particular not for applicants with SAT scores above 1200. Even for the range of SAT scores where data is available, i.e., from about 1040 to somewhat less than 1200, the linear regression equation may not be fully justified (correlation and regression are most relevant for scatter plots that show an elliptical point cloud - but the point cloud here has been cut off at 4.00 and therefore is not fully elliptical).