

Stat 2000 International – Final - Answers

1. Correct answer: [3] the use of self-selected volunteers
2. Correct answer: [4] only the students who participated in the study
3. Correct answer: [4] diet plan
4. Correct answer: [3] The sample statistics would change if a new random sample were drawn.
5. Correct answer: [1] Average is larger
6. Correct answer: [3] Plot C
7. Correct answer: [4] Green = $3/4$, Yellow = $1/4$
8. Correct answer: [2] 0.0314
9. Correct answer: [3] 0.3
10. Correct answer: [2] 0.866
11. Correct answer: [2] about 0.049
12. Correct answer: [3] The sample means are spread from about 26 inches to about 32 inches.
13. Correct answer: [3] \$0.28
14. Correct answer: [2] about 0.8414
15. Correct answer: [3] The manager can be 90% confident that the true average travel time between customers is between 21.73 and 22.81 minutes.
16. Correct answer: [3] $H_a: p > 1/2$
17. Correct answer: [2] about 0.1667
18. Correct answer: [5] H_0 : Population mean = 400; H_a : Population mean \neq 400
19. Correct answer: [2] A t-test is a significance testing procedure used to evaluate hypothesis statements about means.
20. Correct answer: [3] If the population mean is really 23, the probability is 0.0003 that the mean of the 36 ages could be less than 23.

21. Correct answer: [1] Sample No. 1

22. Correct answer: [4] There's not enough evidence to say the means are different.

23. Correct answer: [2] Two possibilities are occupation and seniority.

24. Correct answer: [4] 10 pounds

25. Correct answer: [2] 850

26. Answer: We need to know more about possible confounding factors. In particular, we need to know whether the only difference between the two groups is that one group smokes and the other group does not smoke. Do both groups exercise similarly? Do both groups eat healthily? Do both groups have the same genetic predisposition to heart disease?

27. Answer: (a) Median height about 74 inches. (b) Tallest person about 84 inches. (c) About 75% above 72 inches.

28. Answer: 2. This is linear with an outlier. (The points are exactly on a straight line, except for one.)

29. Answer: Test results are independent so we use the special case rule for AND:

$$P(\text{test husband positive and test wife positive}) = (0.85) * (0.85) = 0.7225$$

30. Answer: You can start with the "Graphing Normal z-Score/Probability" calculator for example. Enter 65 for the mean and 2.7 for the standard deviation. Calculate the "Area left of" 62 (0.1333) and "Area right of" 70 (0.032) 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 $1.0 - (0.1333 + 0.032) = 0.8347$.

31. Answer: You can start with the "Graphing Normal z-Score/Probability" calculator for example. Enter 65 for the mean and 2.7 for the standard deviation. Choose "Area left of" and type in the area (the red number) as 0.60. Then click on "Compute!" and the answer will be that 65.68 is the 60th percentile.

32. Answer: No! Generally, surveys of this type performed on the web do not involve obtaining a random sample from the general population. Those who visit the web site and choose to respond are likely to differ from the general population. Therefore, the sample results should not be used to estimate the true percentage of Americans who believe in life "out there."

33. Answer: The null hypothesis is that there is no relationship: H_0 : Population correlation coefficient = 0. Because we are interested in negative correlation, the alternative hypothesis is H_a : Population correlation coefficient < 0.

34. Answer: Mean blood pressure in the population of women is 115. The p-value is large (larger than 0.05) so we cannot reject the null hypothesis.

35. Answer: The confidence interval suggests that the difference in the means is less than 0. Therefore the first mean is smaller than the second mean.

36. Answer: mean = 98.30161, median = 98.25, and standard deviation = 0.6489568

37. Answer: mean (in C) = $(98.30161 - 32) / 1.8 = 36.834$, median (in C) = $(98.25 - 32) / 1.8 = 36.806$, standard deviation (in C) = $0.6489568 / 1.8 = 0.36053$

38. Answer: Overall, the histogram looks fairly symmetric. There do not seem to be any outliers. The modal bar is the one from 98.0 to 98.5.

39. Answer: According to the boxplot, there is one outlier: the largest value 100.1 is slightly beyond the upper fence. The boxplot indicates that the data is fairly symmetric (although a very careful inspection might reveal that there is some very minor skewness towards the higher values).

40. Answer: We can use a z-test here since the sample size is relatively big (> 30) and the data approximately follows a normal curve. The p-value is < 0.0001 which is obviously less than 0.05. We reject H_0 , concluding that the mean temperature is clearly different from 99. The result is statistically significant.

41. Answer:

90% Confidence interval results (based on z-interval): (98.166046, 98.43718)

90% Confidence interval results (based on t-interval): (98.163956, 98.43927)

As expected, there is hardly any difference among the two intervals (the first four decimal digits for the lower and upper limits are identical).

42. Answer: Use a Normal z-Score/Probability calculator and determine a value such that an area of 0.95 is to the left of this value and an area of 0.05 is to the right. The value 1.645 satisfies this condition and is the multiplier that is needed (for a 90% CI).

The standard error can be calculated as follows:

$$\text{standard deviation} / \sqrt{n} = 0.6489568 / \sqrt{62} = 0.08241759$$

43. Answer: We are using a chi-Squared test with 61 degrees of freedom. The P-value is 0.3643 which is greater than 0.05. We do not reject the null hypothesis that the variance of all temperatures is 0.4.

44. Answer: There is a slight linear trend, with higher HeartRate associated with higher Temperature (although only a test will answer whether the slope is really significantly different from 0). There seems to be one outlier - the point with Temperature = 100.8 and HeartRate = 77. Based on this plot, there are no clusters for the two genders.

45. Answer: The correlation is 0.2536499. This indicates that there is some positive association but also that in general the points are not too close to a line with positive slope.

46. Answer: $\text{HeartRate} = -166.27835 + 2.4431736 * \text{Temperature}$

The p-value associated with Temperature is $0.0036 < 0.05$, i.e., the slope is significantly different from 0. Therefore, we can use the regression equation to predict HeartRate from Temperature.

47. Answer: Both histograms look somewhat similar, with a range from about 55 to about 90. Both shapes look somewhat similar to the normal curve, although not perfect. Both could be slightly skewed towards the lower values. For Gender = 1, the mode is in the 70-75 class and for Gender = 2, the mode is in the 75-80 class.

48. Answer:

Gender 1: $\text{HeartRate} = -87.96204 + 1.6444824 * \text{Temperature}$

Gender 2: $\text{HeartRate} = -233.61661 + 3.1279447 * \text{Temperature}$

49. Answer: For Gender 1, the p-value associated with Temperature is $0.1184 > 0.05$, i.e., the slope is not significantly different from 0. For Gender 2, the p-value associated with Temperature is $0.0205 < 0.05$, i.e., the slope is significantly different from 0. Therefore, we can use the regression equation only for Gender 2 to predict HeartRate from Temperature.

50. Answer: For 97: around 69.79; for 99: around 76.05; for 104: around 91.69. The last value (for 104 degrees Fahrenheit = 40 degrees Celsius) is least reliable since 104 degrees is more than 3 degrees above the highest observed Temperature. We are extremely extrapolating here. This person may have a very critical medical condition and it may be very difficult to correctly predict the HeartRate.