Stat 218 Exam 1 Solutions Winter 2006

Some questions have different versions, so be sure to find the relevant version for your exam.

Frequent flyer question:

a) The value 83% is a statistic, because it pertains only to the sample of people who responded to the poll, not to the population of all adult Americans.

b) The sampling method is almost certainly biased in favor of "yes" responses to using a credit card to accumulate airline miles, because only people interested in the topic would read the article in the first place, and people who use a credit card to accumulate airline miles would be more likely to be interested in the topic.

A common error was not to specify the direction of the bias as it relates to this issue of using a credit card for airline miles.

Short answer questions:

a) Examples of categorical variables are gender, type of injury, whether the patient has health insurance, whether the patient came in an ambulance. Examples of quantitative variables are waiting time before seeing a doctor, total cost of treatment, number of stitches received.

A common error was to give a summary about all patients, such as the number of patients who were treated that day.

b) The midrange is not resistant to outliers, because it depends only on the largest and smallest values in the data, so outliers would affect the midrange greatly.

The midhinge is resistant to outliers, because it depends only on the quartiles, which are not affected by a few extreme values.

c) The sample mean is denoted by $\bar{y}$. It is the midpoint of the confidence interval, because the interval is formed by taking the sample mean plus/minus the margin-of-error. In this case the midpoint, and therefore the sample mean, is (6.5+7.7)/2 = 7.1.

d) The IQR is the difference between the quartiles, which are the 3rd and 8th values (in order). To make the IQR = 0 requires the middle six scores to be identical.

To make the mean greater than the median, include one or two very high values, such as: 60, 60, 60, 60, 60, 60, 60, 60, 100, 100. Here the IQR = 0, median = 60, and mean = 68.

To make the mean less than the median, include one or two very low values, such as: 0, 0, 60, 60, 60, 60, 60, 60, 60, 60. Here the IQR = 0, median = 60, and mean = 48.

Dog height question:

Since we are comparing "apples and oranges," we can do so with z-scores.

A German Shepherd with a height of 28 inches has a z-score of (28-25)/2.5 = 1.2, indicating that his height is 1.2 standard deviations above the mean. A Sheltie with a height of 18 inches has a z-score of (18-15)/1.5 = 2.0, indicating that his height is 2.0 standard deviations above the mean. Therefore, such a Sheltie is more unusual than such a German Shepherd.

A German Shepherd with a height of 22 inches has a z-score of (22-25)/2.5 = -1.2, indicating that his height is 1.2 standard deviations below the mean. A Sheltie with a height of 12 inches has a z-score of (12-15)/1.5 = -2.0, indicating that his height is 2.0 standard deviations below the mean. Therefore, such a Sheltie is more unusual than such a German Shepherd.

Body temperature question:

a) Because we do not know the population standard deviation σ, we need to use the sample standard deviation s and therefore a t-interval rather than a z-interval. With a sample size of 65, the degrees of freedom is 64, but with our table we will use 60, giving t* = 1.990.

The 95% CI for the population mean body temperature of a healthy adult American *female* is: 98.105 +/- 1.990(0.699)/sqrt(65), which is 98.105 +/- 0.173, which is ((97.932, 98.278).

The 95% CI for the population mean body temperature of a healthy adult American *male* is: 98.394 +/- 1.990(0.743)/sqrt(65), which is 98.394 +/- 0.184, which is ((97.210, 98.578).

b) The symbol μ represents the mean body temperature among all healthy adult American males/females.

c) The 95% CI does not include the value 98.6, suggesting that 98.6 is not a plausible value for the population mean body temperature.

d) We are told that the sample was randomly selected, so that condition is met. The second condition requires either a normally distributed population or a large sample size. We can not check normality, but a sample size of 65 is large, so this condition is also met.

e) A 90% confidence interval would have the same midpoint, because the midpoint is the sample mean. The width would be smaller, because with a lower confidence level we do not need as wide an interval to be confident of capturing the population mean.

Sea anemone question:

The Central Limit Theorem tells us that the sample mean diameter will follow a normal distribution with mean 4.2 centimeters and standard deviation 1.4/sqrt(40) = 0.2214 centimeters. The z-scores are therefore (4.0-4.2)/.2214 = -0.90 and (4.5-4.2/.2214) = 1.36. Looking in the normal probability table reveals that the probability that the sample mean diameter is between 4 and 4.5 centimeters is .9131 - .1841 = .7290.

A very common error was to simply find the probability that a single anemone would have a diameter between 4 and 4.5 centimeters, without regard to the sample mean and the sample size of 40.

Highest points question:

a) The observational units are the fifty states.

A very common error was to report the variable: highest point in the state.

b) The distribution of states' highest points is skewed to the right.

c) Because of the right skew, the mean would be larger than the median, as the mean would be pulled up by the few very large values.

d) Because of the right skew, the correct normal probability plot for these data is the one that has a curved, non-linear pattern.