MA145 Fall 2006 Final Exam

Name:

Part I: Descriptive Statistics

The following table contains descriptive statistics for a set of data. Questions 1-7 pertain to this table.

Table 2

Measure	**Value**
Minimum	3
Maximum	97
Mean	72
Median	43
Q_1	38
Q_3	48

1) Which of the following terms best describes this data set?

( ) skewed left (X) skewed right
( ) not skewed ( ) symmetric

(Data is skewed right when the mean is larger than the median)

2) What is the value of the interquartile range?
The IQR is $Q_3 - Q_1$=48 − 38=10.

3) What is the value of the **upper** fence?
The upper fence is $Q_3 + 1.5 \cdot IQR$=48 + 1.5 · 10=63

4) What is the value of the **lower** fence?
The lower fence is $Q_1 - 1.5 \cdot IQR$=38 − 1.5 · 10=23

5) For this dataset, which, if any, of the following would be considered outliers?

(X) 65 ( ) 59 ( ) 57
( ) 23 (X) 20 (X) 19

(Any value below the lower fence (23) or above the upper fence (63) is an outlier).

6) How many units wide would the "box" part of the box plot for this data set be?

The box extends from Q_1 to Q_3, so its width is equal to the IQR, which is 10.

7) Write the Five Number Summary for this data set (be sure to identify each number).

The 5-number summary is:

3 (min), 38 ($Q3$), 43, (median), 48 ($Q3$), and 97 (max)

Part II: Definitions Match each entry in the left column with the entry that *best* fits from the right column.

(b) binomial mean μ_X	a) $Q_3 - Q_1$
(l) type I error	b) np
(n) null hypothesis	c) data having the median much smaller than the mean
(m) statistic	d) $Q_1 - 1.5 \cdot IQR$
(f) binomial stdev σ_X	e) (largest data value) − (smallest data value)
(a) interquartile range	f) $\sqrt{np(1-p)}$
(j) parameter	g) $Q_3 + 1.5 \cdot IQR$
(c) skewed right	h) data having the median much greater than the mean
(h) skewed left	i) Do not reject H_0 when H_1 is true
(g) upper fence	j) A descriptive measure of a population
(d) lower fence	k) Five number summary
(i) type II error	l) Reject H_0 when H_0 is true
(k) $\min, Q_1, M, Q_3, \max$	m) A descriptive measure of a sample
(e) range	n) A statement to be tested

Part III: Short Answer Write a few sentences or a short paragraph to answer the following questions:

1 Describe the meaning of the terms "confidence interval" and "level of confidence".

A *confidence interval* consists of an interval of numbers, along with a measure of the likelihood that the interval contains the unknown mean μ of the population. The level of confidence is the percentage of intervals that will contain μ if we repeat the sampling many times and compute a new confidence interval each time.

2 Describe the meaning of the Central Limit Theorem and state why it is important.

The central limit theorem states that if a random sample of size n is drawn from a population with mean μ and standard deviation σ, the sampling distribution of the sample mean $\overline{x}$ will become approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$ as n becomes large. The central limit theorem is important because it allows us to treat the sample mean $\overline{x}$ as if it came from a normal distribution, even if it did not, when n is sufficiently large, say 30 or more.

3 Describe the empirical rule.

The empirical rule states that, for a normal distribution with mean μ and standard deviation σ, approximately 68% of the population will lie within 1 standard deviation on either side of μ. Approximately 95% will lie within 2 standard deviations either side of μ, and approximately 99.7% will lie within 3 standard deviations on either side of μ.

Part IV: Techniques For each of the following problems, indicate which of the following 10 statistical techniques you would use. If the technique involves hypothesis testing, state the null hypothesis.

- Confidence Intervals for Population Means
  - Confidence interval for a population mean - known σ
  - Confidence interval for a population mean - unknown σ
  - Confidence interval for a population proportion
- Testing a Hypothesis about a Population Mean μ
  - Testing a hypothesis about μ with σ known.
  - Testing a hypothesis about μ with σ unknown.
  - Testing a hypothesis about a population proportion p.
- Inference about Two Samples
  - Inference about Two Means with Paired or Dependent Samples
  - Inference about Two Means with Population Standard Deviations Known
  - Inference about Two Means with Population Standard Deviations Unknown
  - Inference about Two Population Proportions

1) A technique for measuring rainfall from radar data indicates that a storm produced 1.25 inches of rain in a certain area. Measurements from 50 rain gauges in the area show an average of 1.18 inches of rain with a standard deviation of 0.3 Does this data indicate that the actual rainfall differs from the radar estimate?

Solution: In this problem, we could consider the radar estimate to be a claim that 1.25 inches of rain actually fell. We will use the sample of 50 rain gauges to test the hypothesis that in fact 1.25 inches of rain did fall. Since we are apparently using the standard deviation of the 50 measurements from the rain gauges, we have estimated the standard deviation from the sample. So, we are using s, the sample standard deviation, and the population standard deviation σ is unknown. We have a large enough sample ($n \geq 30$) to treat the sample mean $\overline{x}$ as if it is approximately normally distributed.

2) 65 participants in a weight loss program are weighed at the start and end of the program. The average difference in their before and after weights is -2.1 pounds with a standard deviation of 3.2. Does this data support the claim that the program is effective?

Solution: The data in this problem represents before and after measures on the same individual, so this is a case of inference on two means with paired or dependent samples. The null hypothesis would be that there is no difference in the mean weights of people before and after the weight loss program, that is to say, that the program is not effective. Rejecting the null hypothesis is equivalent to affirming that the program is effective.

3) An electronics manufacturer draws a sample of 250 100-ohm resistors from a large production run and tests them. The average resistance for the sample is 100.3 ohms with a standard deviation of 0.03. Construct a 95% confidence interval for the mean resistance in the production run.

Solution: It is clear from the statement that this problem involves a confidence interval for a population mean. We are given the mean of a sample of 250 resistors (100.3 ohms) and the sample standard deviation, 0.03. Since we have only the sample standard deviation to work with, this would be a case of determining a confidence interval for a population mean with population standard deviation σ unknown. The population in this case is the large production run.

4) A commonly used intelligence test for children has a mean of 100 and a standard deviation of 15. A sample of 120 students entering a public elementary school are tested and score 107.3 on average. A sample of 114 students entering a private school are tested and average 112.8. Does this data indicate a difference in the average scores on this intelligence test for students at the two schools?

Solution: In this problem, we are given the means of two samples, 120 from a public school and 114 from a private school. We are not given the *sample* standard deviations, but rather the mean and standard deviation for the test. For this type of standardized test, it is usually assumed that the quoted standard deviation of 15 applies to any population to which we administer the test. This would be considered a reliable estimate of the population standard deviation σ, so we would analyze this data as inference on two means with σ known. The null hypothesis would be that there is no difference between the mean scores on the intelligence test at the two schools.

5) A car manufacturer advertises that 85% of its cars sold in the last 10 years are still on the road. A sample of 65 cars made by this company in the last 10 years finds that 42 are still on the road. Does this data support the manufacturer's claim?

Solution: This problem presents a claim by the manufacturer regarding the proportion of cars it sold in the past 10 years, namely that that figure is 85%. Here we are testing a hypothesis on a population proportion. The null hypothesis is that the proportion of cars on the road is actually 85%. If the sample proportion $\hat{p} = 42/65$ differs significantly from 0.85, we would reject the manufacturer's claim.

6) A sample of 50 home sold in a certain area has an average selling price of 217.5 (in thousands) and a standard deviation of 43.1. In another area, a sample of 47 homes had an average selling price of 286.2 with a standard deviation of 51.1. Does this data support the claim that the average selling price in the two areas is the same?

Solution: In this problem we have two samples of homes from different areas. We are given the means $\overline{x}_1$ and $\overline{x}_2$ and the *sample* standard deviations s_1 and s_2. This is a case of inference about two means with population standard deviation σ unknown, because we have only the sample standard deviations to work with. The null hypothesis is that there is no difference between the average home price in the two areas.

7) A large survey by the census bureau finds that average household income has a standard deviation of 34.5. A small city claims that the average household income for residents is 73.2. A sample of 98 households has a mean of 68.9. Does this data support the city's claim?

Solution: In this problem we are given a claim regarding the average household income in a certain city. We are also given the mean of the household incomes of a sample of 98 households (but not the sample standard deviation). The problem states that a large survey by the census bureau reports that the standard deviation of household income is 34.5. This would be considered a reliable estimate of the population standard deviation σ, so we are testing a hypothesis on a population mean with σ known. The null hypothesis would be that the claim is correct, that the mean household income is in fact 73.2.

8) A sample of 180 male employees in a large corporation finds that 82 were promoted in the last 5 years. A sample of 160 female employees finds that 59 were promoted in the last 5 years. Does this data support the claim of equal promotion opportunity for men and women by this corporation?

Solution: This is a case of inference on two population proportions. Recall that inference on two means or two proportions is a special case of hypothesis testing, and the null hypothesis is usually that the population means or proportions are identical. That is true here, because the employer's claim is that males and females have equal opportunity for promotion in the company. We would reject the claim if the sample proportions $\hat{p}_1 = 82/180$ and $\hat{p}_2 = 59/160$ are significantly different.

9) In a survey of 500 likely voters, 273 indicate that they plan to vote for a certain candidate. Find a 99% confidence interval for the proportion of likely voters who plan to vote for this candidate.

Solution: Clearly from the statement of the problem, this is a case of constructing a confidence interval for a proportion. The midpoint of the interval will be $\hat{p} = 273/500$.

10) Nationwide data indicates that the number of times a family dines out per month has a standard deviation of 3.2. A survey in a certain market indicates that on average, families dine out 5.6 times per month. Construct a 99% confidence interval for the average number of times a family dines out in this market.

Solution: It is clear from the statement that this problem involves a confidence interval for a population mean. The standard deviation of 3.2 based on nationwide data would be considered a reliable estimate of the population standard deviation, so we would consider the population standard deviation σ to be known in this case. Note that the problem as stated does *not* contain enough information to construct the confidence interval, because we need to know the size of the sample in the survey.