STATISTICS
SECTION II
Part B
Question 6

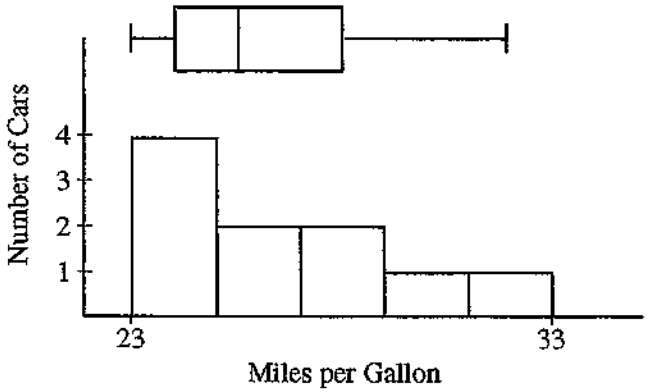
Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A consumer organization was concerned that an automobile manufacturer was misleading customers by overstating the average fuel efficiency (measured in miles per gallon, or mpg) of a particular car model. The model was advertised to get 27 mpg. To investigate, researchers selected a random sample of 10 cars of that model. Each car was then randomly assigned a different driver. Each car was driven for 5,000 miles, and the total fuel consumption was used to compute mpg for that car.

   (a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

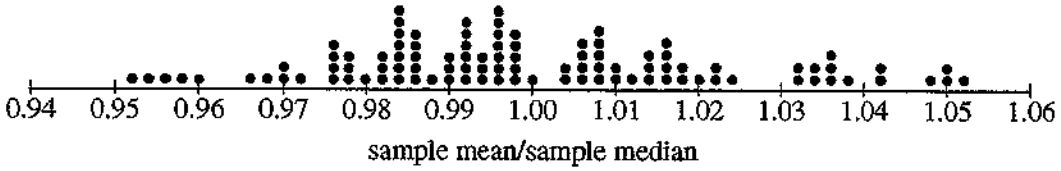
   One condition for conducting a one-sample t-test in this situation is that the mpg measurements for the population of cars of this model should be normally distributed. However, the boxplot and histogram shown below indicate that the distribution of the 10 sample values is skewed to the right.



   (b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

© 2009 The College Board. All rights reserved.
Visit the College Board on the Web: www.collegeboard.com.

(c) Even though the mpg values in the sample were skewed to the right, it is still possible that the population distribution of mpg values is normally distributed and that the skewness was due to sampling variability. To investigate, 100 samples, each of size 10, were taken from a normal distribution with the same mean and standard deviation as the original sample. For each of those 100 samples, the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was calculated. A dotplot of the 100 simulated statistics is shown below.



In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars.

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

STOP

END OF EXAM

© 2009 The College Board. All rights reserved.
Visit the College Board on the Web: www.collegeboard.com.

2009 #6

a] We want to know the TRUE AVERAGE mpg for the car in question.

$H_0: \mu = 27$

$H_A: \mu < 27$

b] Skewed right means the mean is larger than the median so the statistic $\frac{\text{sample mean}}{\text{sample med}}$ will get larger as the distribution gets more skewed right.

c] "this is a wierd problem but watch"

H_0: The distribution is Normal [statistic = 1]

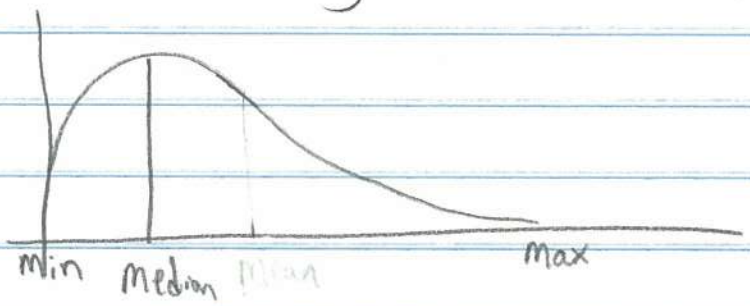
H_A: The distribution is skewed Right [statistic > 1]

p-value = "the probability of getting what you got or something more extreme IF the null is true"

so in this problem there are 14 dots larger than 1.03 out of 100 dots so p-value = .14

Since the p-value is larger than any alpha level I could have picked, there is not enough evidence to reject the null; therefore this could have been a Normal Distribution.

d] Answers will vary but I'm thinking of



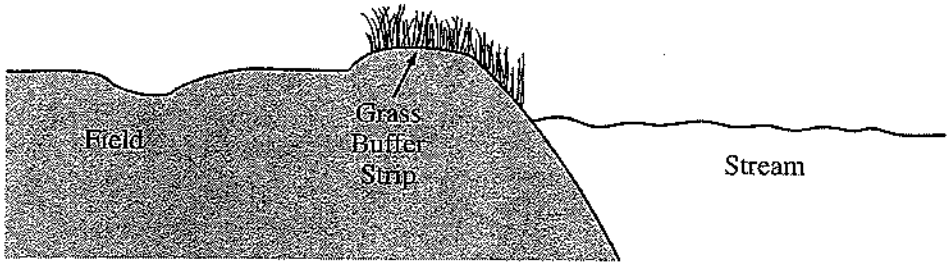
$\frac{\text{max} - \text{median}}{\text{median} - \text{min}}$ = a large # would imply right skewedness.

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Grass buffer strips are grassy areas that are planted between bodies of water and agricultural fields. These strips are designed to filter out sediment, organic material, nutrients, and chemicals carried in runoff water. The figure below shows a cross-sectional view of a grass buffer strip that has been planted along the side of a stream.



A study in Nebraska investigated the use of buffer strips of several widths between 5 feet and 15 feet. The study results indicated a linear relationship between the width of the grass strip (x), in feet, and the amount of nitrogen removed from the runoff water (y), in parts per hundred. The following model was estimated.

$$\hat{y} = 33.8 + 3.6x$$

(a) Interpret the slope of the regression line in the context of this question.

(b) Would you be willing to use this model to predict the amount of nitrogen removed for grass buffer strips with widths between 0 feet and 30 feet? Explain why or why not.

A scientist in California wants to know if there is a similar relationship in her area. To investigate this, she will place a grass buffer strip between a field and a nearby stream at each of eight different locations and measure the amount of nitrogen that the grass buffer strip removes, in parts per hundred, from runoff water at each location. Each of the eight locations can accommodate a buffer strip between 6 feet and 13 feet in width. The scientist wants to investigate which combination of widths will provide the best estimate of the slope of the regression line.

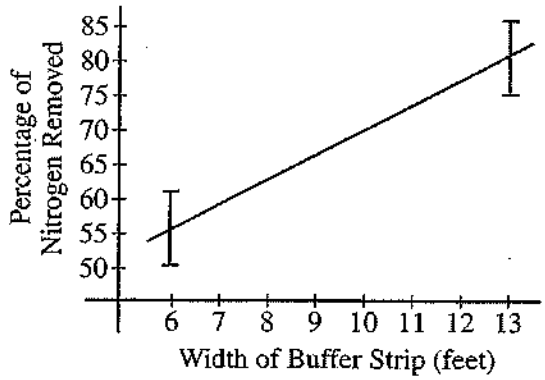
© 2011 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

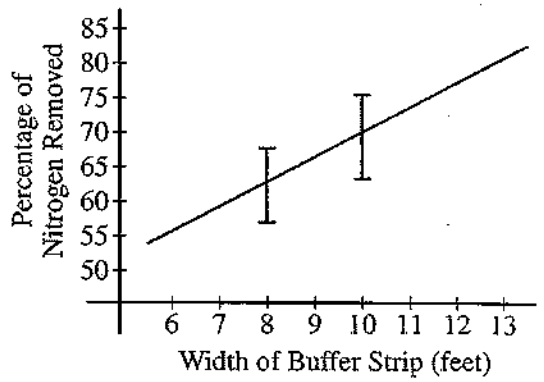
Suppose the scientist decides to use buffer strips of width 6 feet at each of four locations and buffer strips of width 13 feet at each of the other four locations. Assume the model, $\hat{y} = 33.8 + 3.6x$, estimated from the Nebraska study is the true regression line in California and the observations at the different locations are normally distributed with standard deviation of 5 parts per hundred.

(c) Describe the sampling distribution of the sample mean of the observations on the amount of nitrogen removed by the four buffer strips with widths of 6 feet.

(d) Using your result from part (c), show how to construct an interval that has probability 0.95 of containing the sample mean of the observations from four buffer strips with widths of 6 feet.

For the study plan being implemented by the scientist in California, the graph on the left below displays intervals that each have probability 0.95 of containing the sample mean of the four observations for buffer strips of width 6 feet and for buffer strips of width 13 feet. A second possible study plan would use buffer strips of width 8 feet at four of the eight locations and buffer strips of width 10 feet at the other four locations. Intervals that each have probability 0.95 of containing the mean of the four observations for buffer strips of width 8 feet and for buffer strips of width 10 feet, respectively, are shown in the graph on the right below.





If data are collected for the first study plan, a sample mean will be computed for the four observations from buffer strips of width 6 feet and a second sample mean will be computed for the four observations from buffer strips of width 13 feet. The estimated regression line for those eight observations will pass through the two sample means. If data are collected for the second study plan, a similar method will be used.

(e) Use the plots above to determine which study plan, the first or the second, would provide a better estimator of the slope of the regression line. Explain your reasoning.

(f) The previous parts of this question used the assumption of a straight-line relationship between the width of the buffer strip and the amount of nitrogen that is removed, in parts per hundred. Although this assumption was motivated by prior experience, it may not be correct. Describe another way of choosing the widths of the buffer strips at eight locations that would enable the researchers to check the assumption of a straight-line relationship.

STOP

END OF EXAM

© 2011 The College Board.
Visit the College Board on the Web: www.collegeboard.org.

2011 Form B #6

a] As the width of the buffer strip increases by ONE foot, we would predict the Amount of Nitrogen that is removed to increase by 3.6 pph.

b] Since the model given was only produced using strips between 5 and 15 feet, I would NOT be willing to use the model for 0 through 30 Feet. That would be an Extrapolation which could cause errors in analysis.

c] "Shape Center Spread"

I would expect the sampling distribution to be approximately normal with a center at $\bar{y} = 33.8 + 3.6(6) = 55.4$ and a standard deviation at $\frac{5}{\sqrt{4}} = 2.5$

d] $55.4 \pm (2.5)(\text{Invnorm}(.025))$

e] Plan A because if the Sample means were on the extremes of the possibilities, the line would still be closer in slope to the other Plan.

f] Instead of 4 measurements at two widths, Try less Measurements using more variety of widths.