Pledge:

10/2/2009 MATH 171 Name: Solution
Dr. Lunsford Test 1 100 Points Possible

Please show all work on this test (including calculator commands with input) to receive full credit.

I. Short Answer and Multiple Choice. All problems count 2 points unless otherwise indicated. (55 points total)

1. The median survival time for women who have been diagnosed with breast cancer is somewhat smaller than the mean ~~waiting~~ (survival) time. This means that the distribution of survival times is most likely: (circle one)

a) left skewed b) symmetric c) unimodal (d) right skewed) e) none of these

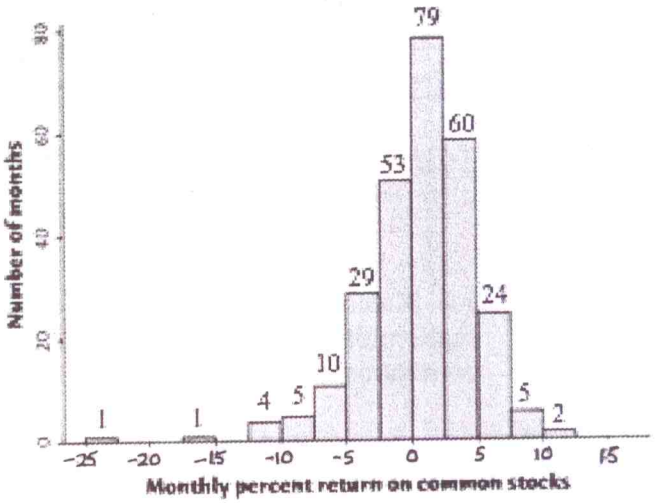
2. A student evaluation survey asks students to rate their professor's effectiveness on a scale from 1 to 5, with 1 meaning "not effective", 2 meaning "marginally effective", 3 meaning "somewhat effective", 4 meaning "effective", and 5 meaning "very effective". The type of data being asked for is (circle one)

a) quantitative b) explanatory (c) categorical) d) response e) none of these

3. Circle **all** of the following numerical descriptors that are not resistant to outliers.

(a) mean) b) median (c) standard deviation) d) IQR e) none of these

The histogram at right shows the distribution of monthly percent returns on U.S. common stocks for the 273 months from January 1985 to September 2007. The class width is 2.5, and you may assume that no data point fell on a class boundary. Use the histogram to answer questions 5 through 6 :



4. Which class (e.g. -25 to -22.5, or -22.5 to -20, etc.) contains the median of the distribution?

0 to 2.5

5. What percentage of months had a positive monthly percent return on common stocks? (3 points)

$\frac{79+60+24+5+2}{273} = \frac{170}{273} = 0.6227$ (62.3%)

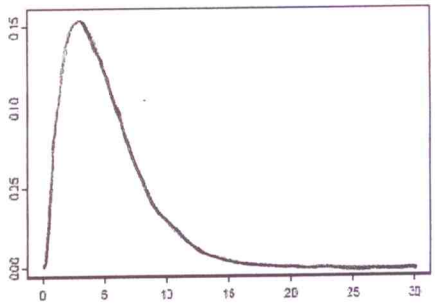
6. The quartiles for this distribution are Q1 = -1.25, and Q3 = 3.63. What is the largest monthly percent return than can be considered a low outlier? (3 points)

$IQR = Q3 - Q1 = 3.63 - (-1.25) = 4.88$

$Q_1 - 1.5\,IQR = -1.25 - 1.5(4.88) = \boxed{-8.57}$

I. Short Answer and Multiple Choice, continued.

7. Please circle *all* words below that describe the shape of the density curve drawn to your right below: (3 points)



Unimodal Normal Bimodal Symmetric

Skewed Right Skewed Left

8. A Senator wants to know what the voters of his state think of proposed legislation on gun control. He mails a questionnaire on the subject to an SRS of 2500 voters in his state. His staff reports that 448 questionnaires have been returned, 343 of which support the legislation. The population for this study is: (circle one)

a. the voters in his state.
b. the 343 letters supporting the legislation.
c. the 448 letters received.
d. the 2500 voters receiving the questionnaire.

9. The sample in the setting of the previous exercise is: (circle one)

a. the voters in his state.
b. the 343 letters supporting the legislation.
c. the 448 letters received.
d. the 2500 voters receiving the questionnaire.

10. Indicate whether each of the following statements is true or false by writing "true" or "false" in the blank provided next to each statement. (2 points each, 6 total)

False A) If the association between two variables is strong, then the explanatory variable *causes* the response variable.

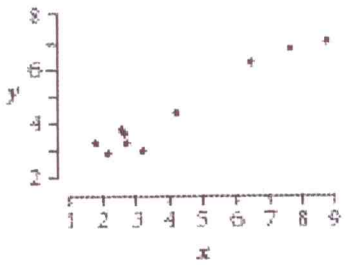
True B) The standard normal distribution (z-score distribution) has a standard deviation equal to 1.

True C) If people with larger feet tend to be more intelligent then we would expect the correlation between foot size and intelligence to be positive.

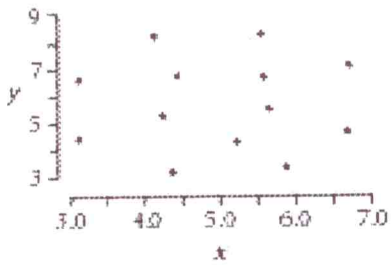
11. Match the correlation coefficients with the scatterplots by writing the name of the scatterplot with the given correlation coefficient in the blank provided next to the correlation coefficient. (2 points each - 6 points total)

r=0.4342 C *r=0.0015* B *r=0.8915* A

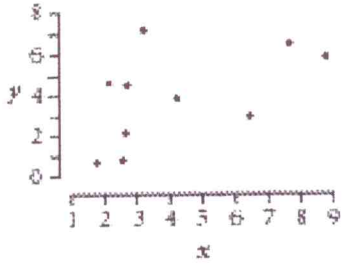
A.



B.

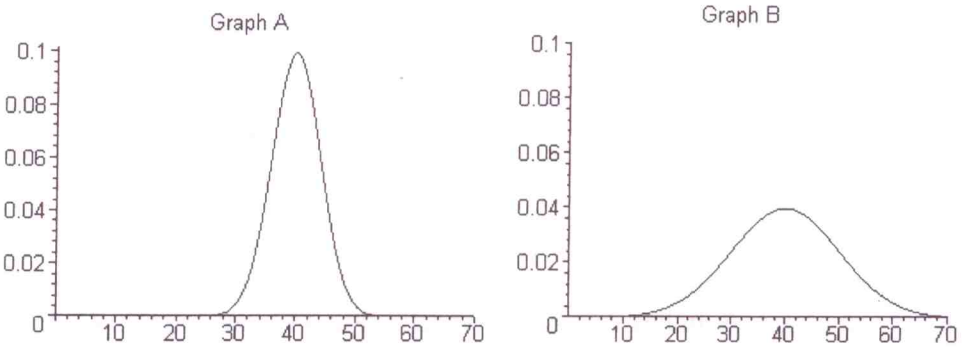


C.



I. Short Answer and Multiple Choice, continued.

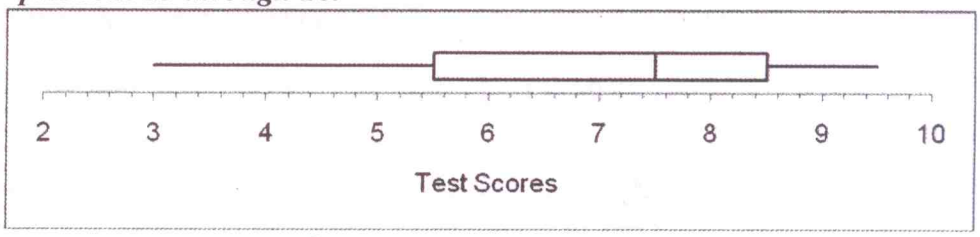
12. Below you are given the graphs of two normal density curves, both with $\mu = 40$. Use these density curves to answer the following questions: (2 points each - 8 points total)



(a) True or False (circle one): The area under each of these curves is equal to 1.
(b) Which curve has the smaller standard deviation? A
(c) Which distribution has a larger percent of its data between 35 and 40 units? A
(d) Give a rough estimate of the standard deviation of the density curve in Graph A:

$\frac{52-28}{6} = 4$ ≈ 4

At right is a box plot of test scores (out of 10 possible) from 36 students taking a certain professor. Use it to answer questions 13 through 16.



13. **True** or **False** (circle one) The distribution of test scores is likely to be left-skewed.

14. Fill in the blank: the median of this distribution is $m =$ 7.5.

15. Fill in the blank: The interquartile range of this distribution is 3.

16. Fill in the blank: The percentage of employees who scored 8.5 or better is 25.

17. The equation $\hat{y} = 102 - 4.5x$ predicts final grade (response) based on number of absences (explanatory). What final grade is predicted for a student with 4 absences? (3 points) $102 - 4.5(4) = 84$

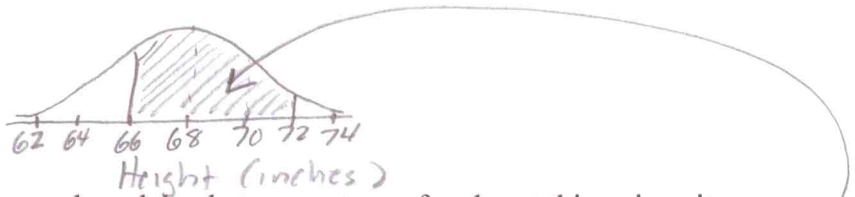
18. Continuing with the regression model in the previous problem, how much does final grade change on average with each additional absence? (3 points)

It decreases by 4.5 points

(i.e. -4.5)

Problem II. Suppose that the heights of male students at a large state university can be modeled with a normal distributed with mean 68 inches and standard deviation 2 inches. Use this information to answer the following questions. (18 points total)

(a) Draw a nice graph of the distribution of male heights. On the graph you should show values ± 1, ± 2, and ± 3 standard deviations from the mean. Clearly label your axis. (5 points)



(b) According to this normal model, what percentage of males at this university are between between 5 feet 6 inches tall and 6 feet tall? Please represent this probability on your graph above. You should clearly indicate any calculator functions (including inputs) you use to make this computation. (4 points)

82%

5'6" = 66", 6' = 72" normalcdf(66, 72, 68, 2) = .8186

(c) Juan has a z-score of -1.45, how tall is he? (4 points)

$68 - 1.45(2) =$ 65.1 inches

(d) What is the cutoff point (i.e. the shortest a male student can be) to be considered among the tallest 8 percent of male students at this university? (5 points)

invNorm (.92, 68, 2) = 70.81 inches

Problem III. Many studies have found that people who drink alcohol in moderation have lower risk of heart attacks than either nondrinkers or heavy drinkers. Does alcohol consumption also improve survival after a heart attack? One study followed 1913 people who were hospitalized after severe heart attacks. In the year before their heart attacks, 47% of these people did not drink, 36% drank moderately, and 17% drank heavily. After four years, fewer of the moderate drinkers had died. (11 points total)

(a) Is this an observational study or an experiment? Why? (3 points) Observational.

No treatment was imposed on the subjects

(b) What are the explanatory and response variables? Please indicate whether each of these variables is quantitative or categorical. (4 points)

Explanatory: Alcohol consumption level — categorical

Response: whether person survived after 4 years — categorical

(c) Suggest a lurking variable that may be confounded with the drinking habits of the subjects. Explain how this lurking variable could give the same response as the one observed. You should tie the lurking variable to both the explanatory and response variables. (4 points)

Those who drink moderately may also take better care of themselves (i.e. a better diet and/or exercise) and thus would be more likely to survive fours years after having a heart attack.

Problem IV. Below are the sample statistics for data on the number of absences and final grades for students in a statistics class and a scatterplot of the actual data. (16 points total)

	mean	std. deviation	correlation
num. absences	8.86	6.35	-0.9098
final grade	65.4	39.1	

(a) Please describe the scatterplot being sure to address its form, direction, and strength. (3 points)

Linear with a relatively strong negative association.
↑ form ↓ strength ↑ direction

(b) Accurately plot the centroid of the data on the scatterplot. (2 points)

centroid = (8.86, 65.4)

(c) By using the appropriate formulas and the statistics above, show that the equation of the least-squares regression line for predicting final grade based on number of absences is $\hat{y} = 115 - 5.6x$. (5 points)

$m = -0.9098\left(\frac{39.1}{6.35}\right) = -5.60$ → so $\hat{y} = 115 - 5.6x$

$y\text{-intercept} = \bar{y} - m\bar{x} = 65.4 - (-5.60)(8.86) = 115.02$

(d) Accurately plot the regression line on the scatterplot. (4 points)

(e) What percent of the variation in Final Grade is explained by the regression on Number of Absences? (2 points)

$r^2 = (-.9098)^2 = 0.8277$ → 82.8%