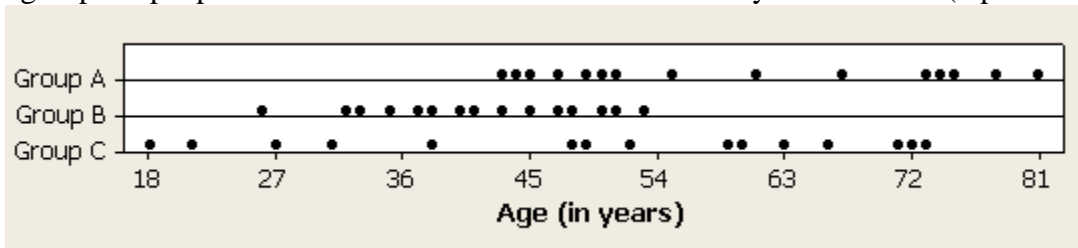


Please show all work on this test. No work shown implies no full credit if (even if the answer is correct!) and no partial credit if the answer is incorrect.

Problem I. Multiple Choice and Short Answer. (30 points total)

- An opinion poll asks respondents whether they approve or disapprove of the job that Barack Obama is doing as President. The possible answers are “strongly approve”, “approve”, “disapprove”, and “strongly disapprove”. The type of data being asked for is (circle one): (3 points)
a) quantitative b) explanatory c) categorical d) response e) none of these
- Circle all of the following numerical descriptors that *are* resistant to outliers: (3 points)
a) mean b) median c) standard deviation d) IQR e) none of these
- A research team wants to determine whether people who are deprived of sleep are more likely to develop depression. The team plans to break their results down by gender, age group (adolescent, young adult, adult, and senior), and ethnicity. The response variable in this study is: (3 points)
a) ethnicity b) age group c) gender d) depression e) sleep deprivation f) none of these
- The distribution of grades in a certain class is strongly left skewed. This means that the median is probably: (3 points)
a) bigger than the mean b) bigger than the IQR c) smaller than the mean
d) smaller than the IQR e) none of these
- To answer questions (a) and (b) below use the following dotplots of ages (graphed on the same scale) for three groups of people. No calculation is needed to determine your answers. (3 points each, 6 total)



- Which group has the youngest mean (average) age?
 - Which group has the largest standard deviation in ages?
- True** or **False** (circle one) If the linear association between two variables is strong, then we would expect the correlation coefficient r to be closer to one or negative one than to zero. (3 points)
 - Suppose the observational units in a study are customers arriving at a post office. For each of the following variables recorded, identify if it is quantitative or categorical. (3 points each, 9 points total)

_____ Total time spent waiting in line and being served
_____ Total amount of money spent
_____ Whether they used a credit card or not.

Problem II. The article “Headset Phones May Still Pose Risks for Drivers” (2004) reported on a University of Utah study that used a driving simulator to see how drivers performed when randomly assigned to talk on a headset cell phone or to talk to another passenger. Part of the simulation instructed the drivers to take a specific exit. Twelve of 24 using a cell phone with a headset missed the exit, 3 of 24 talking to a passenger missed it. Please answer the following questions. (7 points total)

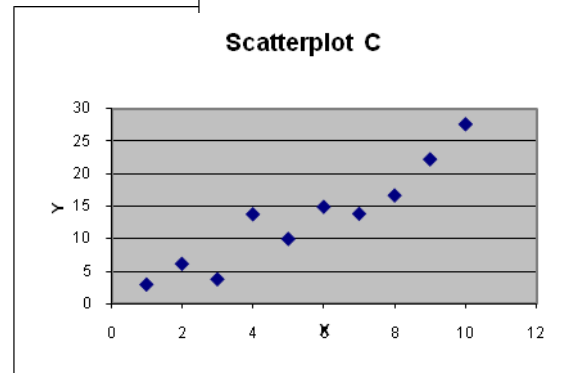
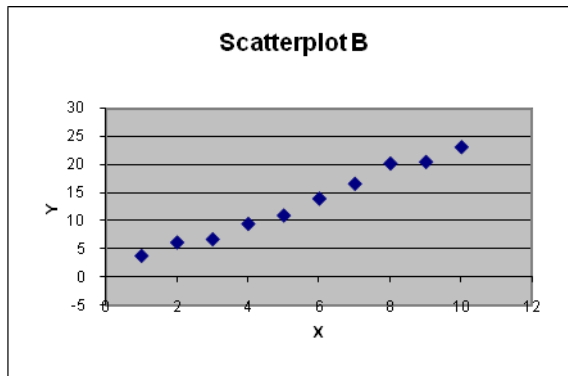
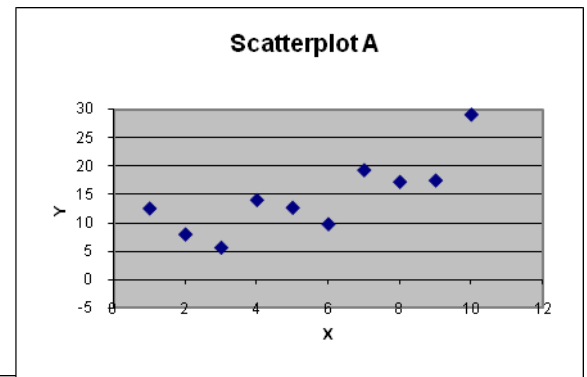
(a) Is this an experimental or observational study? Explain why? (4 points)

(b) TRUE or FALSE. (Circle one) If the researchers find that the difference in percentages of those who missed the exit between the two groups is significant, it would be appropriate to conclude that using the cell phone *caused* the subjects to be more likely to miss the exit than talking to a passenger. (3 points)

Problem III. Below you are given three scatterplots and three correlation coefficients. Please write the name of the scatterplot (i.e. either A, B, or C) in the blank next to the correlation coefficient for the scatterplot. (3 points each – 9 points total)

$r = 0.785293$ _____ $r = 0.94104$ _____

$r = 0.993153$ _____



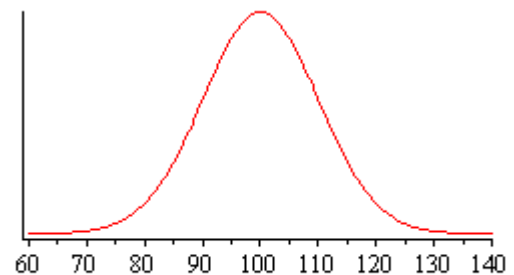
Problem IV. To your right is the graph of the density curve for a normal distribution. Use it to answer the following questions: (3 points each, 6 total)

(1) The mean of the normal distribution is (circle one)

- a) $\mu = 55$ b) $\mu = 100$ c) $\mu = 145$
 d) $\mu = 90$ e) none of these

(2) The standard deviation of the normal distribution is approximately (circle one)

- a) $\sigma = 100$ b) $\sigma = 90$ c) $\sigma = 20$
 d) $\sigma = 10$ e) none of these

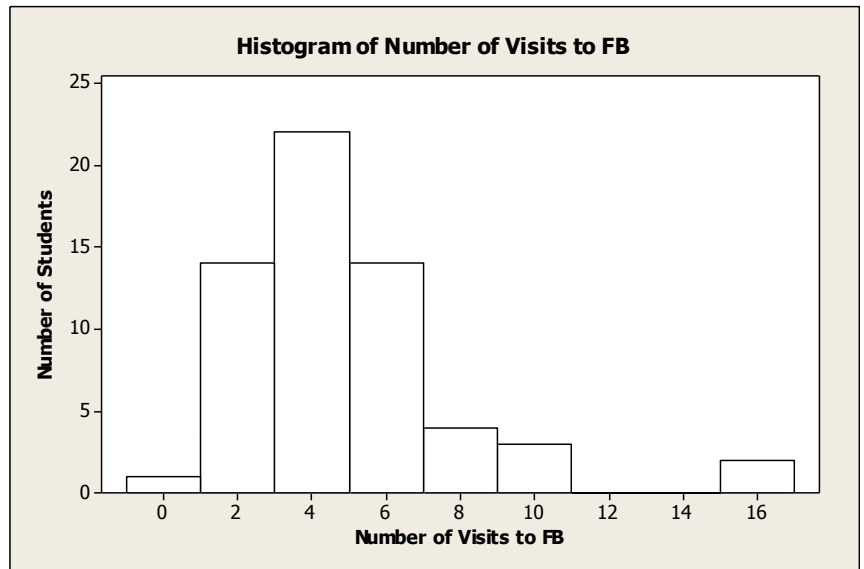


Problem V. A random sample of 60 students at a large university was asked to estimate how many times they visited Facebook on average per day. Below is a histogram of their answers. The five number summary of the data are: Min = 0, 1st Quartile = 2, Median = 4, 3rd Quartile = 8, Max = 16. Please answer the following questions. (21 points total)

(a) Draw a boxplot of the data on the axes provided to your right. (5 points)



(b) Use the 1.5 IQR rule to determine the lowest value at for which a data point could be considered a high outlier. Based on the histogram below, about how many high outliers does the data set contain? (6 points)



(c) Use the histogram to estimate the percent of students who visit Facebook at least 7 times a day? Please shade the region of the histogram that corresponds to this percent. (4 points)

(d) Based on the and box plot, please circle *all words* below that describe this distribution. (3 points)

Symmetric unimodal uniform left skewed right skewed normal

(e) Based on the histogram and boxplot, we believe the mean number of visits to Facebook will be (circle one)

greater smaller

than the median number of visits. (3 points)

Problem VI. The battery in an iPod has a run-time (time until it needs to be recharged) that is normally distributed with a mean of 6 hours and a standard deviation of 30 minutes (i.e. 0.5 hours). Use this information to the following. (4 points each, 12 total)

(a) What percentage of these batteries must be recharged in fewer than 5 hours?

a) 16% b) 5% c) 2.3% d) 0.3% e) none of these

(b) The third quartile for the time to be recharged is

a) 0.67 hours b) 20 minutes c) 6.33 hours d) 6.67 hours e) none of these

(c) The battery in Jane’s iPod had to be recharged after 6 hours and 45 minutes. What is the z-score for this run-time?

a) $z = 1.5$ b) $z = -1.5$ c) $z = 0.5$ d) $z = 2.25$ e) none of these

Problem VII. Below are the data for the second test score and final exam score for eight students in a statistics class. You are also given a scatterplot of the data, including the graph of the least squares regression line. Please use this information to answer the questions below. Please show all calculator input. (15 points total)

second test score	158	162	144	162	136	158	175	153
final exam score	145	140	145	170	145	175	170	160

(a) Find the centroid of the data and plot it on the scatter plot. Clearly indicate the answer below and clearly show the centroid on your graph. (4 points)

(b) Find the equation of the least squares regression line. Clearly indicate your answer below. (4 points)



(c) Use the regression equation model found in part (b) to estimate the final exam score for a person who made 170 on the second test. Please show and work and also show this prediction graphically using the regression line on the graph. (4 points)

(d) Do you think the least squares regression line found in part (b) is a good predictive model of final exam score given the score on test two? Why or why not? (Hint: Use the coefficient of variation, r^2 , to answer this question.) (3 points)