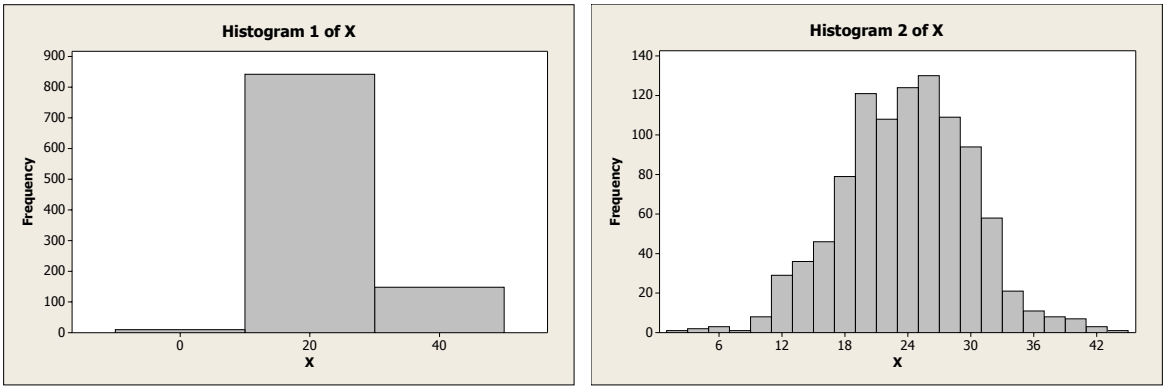
Name _____

MATH 160A Introduction to Applied Statistics Fall 2007 Exam #1

Instructions: This exam is a tool to help me (and you) assess how well you are learning the course material. As such, you should report enough written detail for me to understand how you are thinking about each problem. (130 points total)

---

1. Below are two histograms for the same distribution.



   (a) Which histogram provides a more useful view of the distribution? Why? (6 points)

   (b) What are the key features of this distribution? (6 points)

2. Monthly rental (in dollars) for four studio apartments near the university are

   450 475 520 699

   (a) Compute the mean and standard deviation for this distribution without using statistical features on your calculator. (6 points)

   (b) A larger study gathers rentals for 100 apartments. The mean of the new distribution is \$525 and the standard deviation is \$58. The researcher comes across one more rental value to include and it turns out to be \$525. Is the mean of the new distribution (with 101 values) less than, equal to, or greater than \$525? Is the standard deviation of the new distribution (with 101 values) less than, equal to, or greater than \$58? Explain the reasoning you use to reach each conclusion. (6 points)

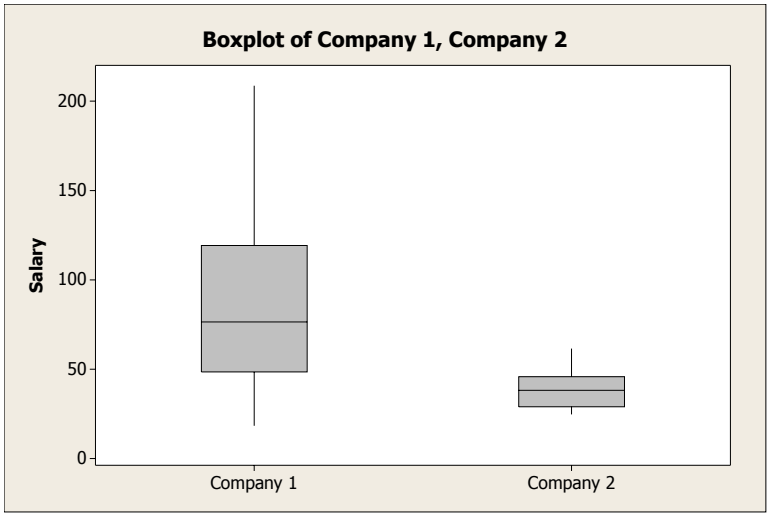
3. Annual salaries (in thousands of dollars) for 14 employees at Company 1 are

137.3 201.2 78.2 71.0 54.2 95.3 113.2 74.6 20.5 44.2 18.3 49.9 208.7 91.8

(a) Give the five-number summary and the IQR for this distribution. (8 points)

(b) Determine if any values are considered potential outliers by the 1.5×IQR rule. (4 points)

(c) Make a boxplot for this distribution in the space provided on the graph below. The graph already shows a boxplot for the distribution of annual salaries at Company 2. (4 points)



(d) Compare the distributions of annual salary for Company 1 and Company 2. As part of this, compare the salaries at Company 2 with the median for Company 1. (4 points)

4. A forester determines that the heights of Douglas fir trees in a certain forest are approximately normally distributed with a mean of 120 feet and a standard deviation of 15 feet.

   (a) Sketch a graph of this normal distribution. (4 points)

   (b) What range contains the middle 95% of heights in this distribution? For this problem, do not use statistical features on your calculator or a table of standard normal proportions. (4 points)

   (c) What proportion of heights in this distribution are bigger than 105 feet? For this problem, do not use statistical features on your calculator or a table of standard normal proportions. (4 points)

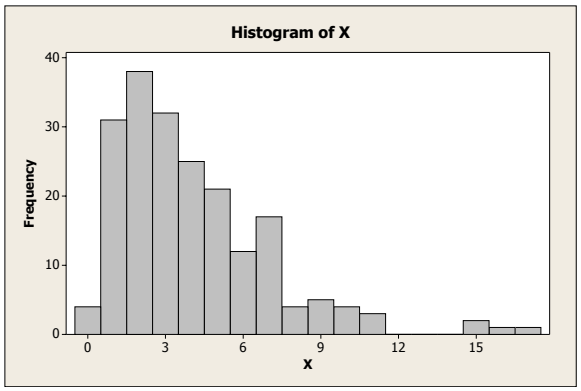
   (d) What proportion of heights in this distribution are larger than 140 feet? (4 point)

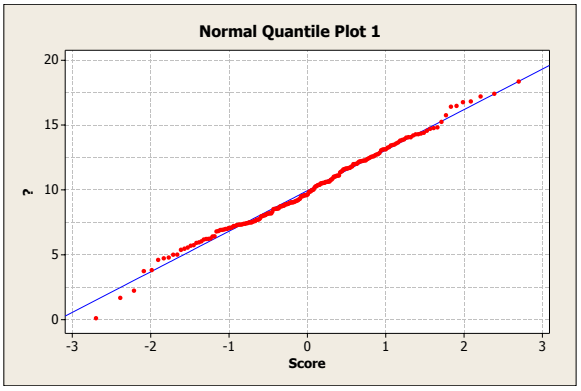
5. Heights of humans are approximately normally distributed with a mean of 65 inches and a standard deviation of 4 inches. Aliens from a faraway planet also have heights that are approximately normally distributed. The aliens measure lengths in a unit called a *blarg*. The distribution of alien heights has a mean of 4.6 blargs and a standard deviation of 1.4 blargs.

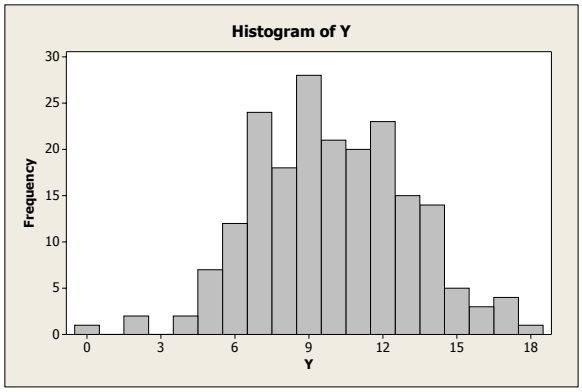
   (a) Which distribution has a wider spread? (6 points)

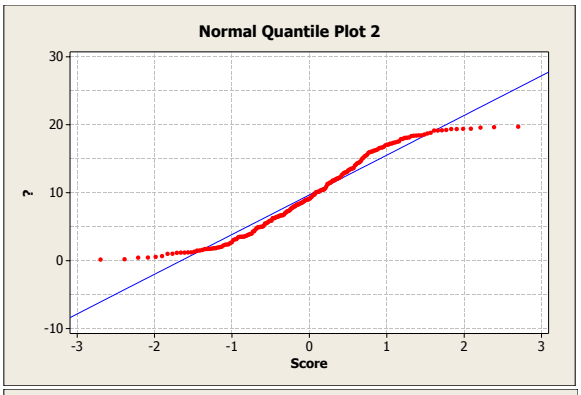
   (b) Who would stand out more in a crowd of their own kind between a human of height 72 inches and an alien of height 6.2 blargs? (6 points)

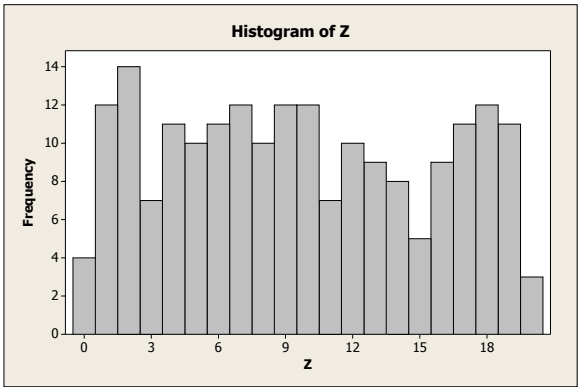
6. Below on the left are histograms for three different distributions, each with 200 values. Below on the right are normal quantile plots for the same distributions but not necessarily in the same order.

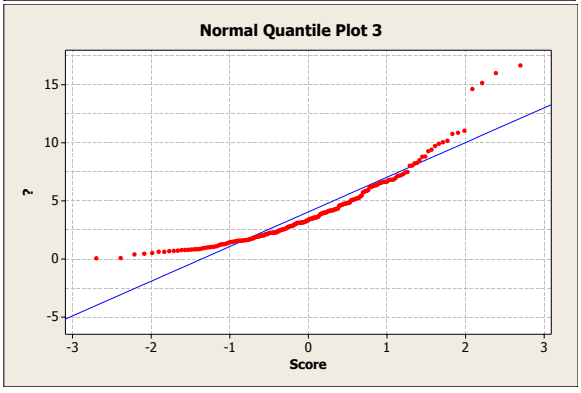










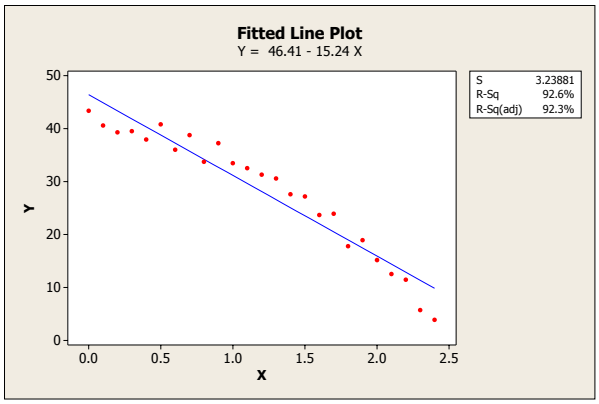


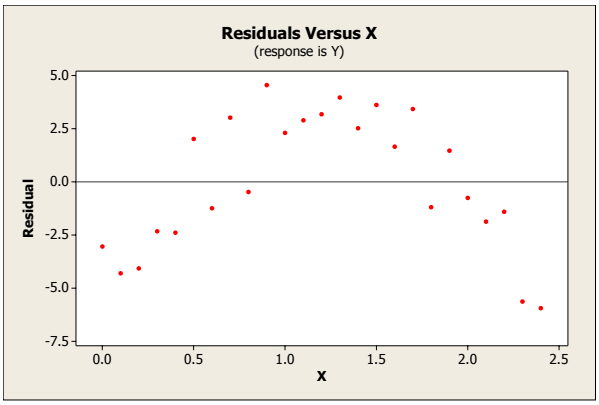
(a) Match each normal quantile plot with the corresponding histogram. (6 points)

(b) Which distribution is best appoximated as normal? (4 points)

(c) For each of the two other distributions, describe what features are not approximated well by a normal distribution. (6 points)

7. A scatterplot for two variables X and Y is shown below on the left. The scatterplot includes the least-square regression line and other values from Minitab. The residual plot is shown below on the right.





(a) Describe the linear association between X and Y that is revealed by this data. (6 points)

(b) What do we learn about the association between X and Y from the residual plot? (4 points)

(c) One point is missing from the residual plot. It corresponds to the data point $(0.8, 33.74)$ on the scatterplot. Find the coordinates of the missing point on the residiual plot and graph the point. (4 points)

8. A medical researcher is interested in the relationship between weight (W) and blood pressure (BP). In an experiment, the researcher measures weight and blood pressure (specifically, diastolic blood pressure) on 1200 individuals. Weight is meaured in pounds and diastolic blood pressure in mmHG. The distribution for W has mean $\overline{W} = 156$ pounds and standard deviation $s_W = 21$ pounds. The distribution for BP has mean $\overline{BP} = 83$ mmHG and standard deviation $s_{BP} = 14$ mmHG. The researcher makes a scatterplot and decides there is evidence of a linear associate between the two variables. The correlation for W and BP is found to be $r = 0.261$.

   (a) Which variable is best thought of as an explanatory variable and which as a response variable? (2 points)

   (b) Describe the strength of the linear association between W and BP. (4 points)

   (c) Find the formula for the least-squares regression line. Use the formula to predict the blood pressure for a person who weighs 167 pounds. (8 points)

   (d) How much of the variance in blood pressure is explained by the least-squares regression on weight? (4 points)

   (e) A journalist reads the researcher's report and writes a story stating that increased weight causes increased blood pressure. Critique this claim. (4 points)

   (f) List three other variables that might reasonably be relevant in understanding blood pressure and give a brief explanation of why each might be relevant. (6 points)