

Math 1431
Summer 2003 – Test #1 – Answers

NAME _____

You are allowed to use your calculator. Explain all answers – answers with no explanation will receive only partial credit. Use complete sentences.

1. (10 points)

a. Construct a scatterplot with 20 observations (points) such that the correlation coefficient has a value $r = -0.8$. Explain.

ANS: The scatterplot has a strong negative (i.e., down and to the right) association.



b. Construct a scatterplot with 20 observations (points) such that the correlation coefficient has a value $r = 0.1$. Explain.

ANS: The scatterplot has a weak or very little positive association.



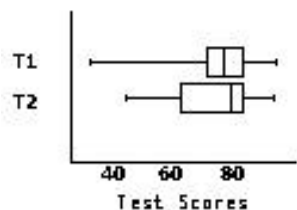
2. (20 points) Here are grades from a previous class that I taught:

Test1	77	85	77	31	78	57	79	89	91	72	73	80	83	74	66	84	97	88	68
Test2	94	47	53	51	81	75	93	81	96	79	84	79	44	81	76	86	83	85	63

a. Did the class perform better on the first or second test? Explain.

ANS:

Here is a side-by-side stemplot of the test scores:



One possible answer is as follows: The medians and means of both tests are nearly the same. In general students performed better on Test 1 because this distribution has smaller spread and test 1 has a possible outlier (31). If this outlier is removed, then the mean and median of Test 1 are closer to those of Test 2. Since the spread is smaller for Test 1, this means that a greater majority of the students performed closer to the median than those in Test 2.

b. On which test is there a larger spread of test scores? Explain.

ANS: The spread of the Test 1 data is 72 – 85 and the spread of the Test 2 data is 63 – 85 (using quartiles). Thus, the Test 2 data has smaller spread. Note that for skewed distributions, the quartiles are a better measure of the spread of the data.

c. Sketch a stemplot of the scores for Test 1 and describe the distribution.

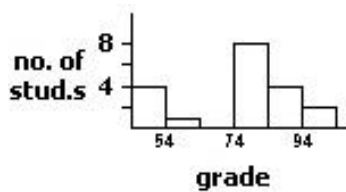
ANS:



There is a possible outlier (31), thus the distribution is left-skewed. (If you don't consider the outlier, then the distribution looks symmetric). There is a peak in the 70s-80s where the center is located (median is 78). The mean is 76 and the spread is 72 – 85 (using quartiles).

d. Sketch a histogram (with class width of 10) of the scores for Test 2 and describe the distribution.

ANS:



There does not seem to be possible outliers, even so, the distribution is left-skewed. There is a peak in the high 70s-low 80s where the center is located (median is 81). The mean is 75 and the spread is 63 – 85 (using quartiles).

3. (16 points) Consider the following data:

33 74 24 57 78 20 46 70 69 16 57 63 76 74 34

ANS: Here is a scatterplot of the data:



The distribution is skewed to the left. There are possible outliers at 16, 20.

a. What is the appropriate measure of center for this data? Explain. What is its value?

ANS: Since the distribution is skewed, the appropriate measure of center is the median (Median is 57).

b. What is the appropriate measure of spread for this data? Explain. What is the value of the spread?

ANS: Since the distribution is skewed, the appropriate measure of spread is $Q_1 - Q_3$ (33 – 74).

c. Find the five number summary of the data. What does the five number summary tell us about the data?

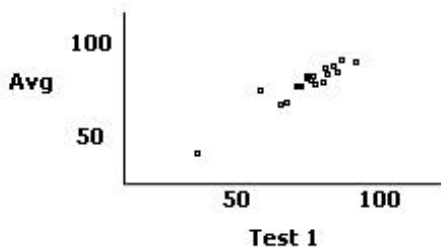
ANS: The five number summary of the data is: 16 33 57 74 78. The five number summary provides a reasonably complete description of center and spread of a distribution.

4. (20 points) In the table below, the test 1 score and the final class average is given for a class of students:

Test 1	77	85	77	31	78	57	79	89	91	72	73	80	83	74	66	84	97	88	68
Average	83	85	84	40	81	76	84	86	93	78	78	79	80	78	68	88	92	89	69

a. Describe the overall pattern of the scatterplot. Compute the correlation coefficient.

ANS: Here is the scatterplot:



The correlation coefficient is $r = 0.94056$. There is a possible outlier at (31, 40). The data has a positive and linear association with most of the data in one cluster.

b. Find the least-squares regression line.

ANS: The least squares regression line is: $\hat{y} = 0.757x + 21.78$.

c. What percent of the observed variation in the final class average is explained by a straight-line relationship with the first test score?

ANS: 88.46% (i.e., r^2) of the observed variation in the final class average is explained by a straight-line relationship with the first test score.

d. Predict the final class average of a student who scored a 60 on the first test.

ANS: $\hat{y} = 0.757(60) + 21.78 \approx 67$. I predict that a student who scored a 60 on the first test would receive a 67 for the final class average.

5. (17 points) The length of human pregnancies from conception to birth is approximately normal with mean 266 days and standard deviation of 16 days.

a. What percent of pregnancies last longer than 280 days? Sketch the area.

ANS: Using NDAREA:

AREA=.19078691

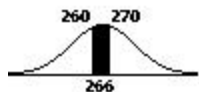


19% of pregnancies last longer than 280 days.

b. What percent of pregnancies last between 260 days and 270 days? Sketch the area.

ANS: Using NDAREA:

AREA=.24487599



24.4% of pregnancies last between 260 days and 270 days.

c. How long do the shortest 10% of pregnancies last?

ANS: Using INVNORM(0.1, 266, 16): The shortest 10% of pregnancies last less than 245.5 days.

6. (17 points) The distribution of SAT scores is approximately normal with mean 500 and standard deviation 100. Use the 68–95–99.7 rule to find the following:

a. the percent of scores between 200 and 800.

ANS: $200 = 500 - 3(100)$ and $800 = 500 + 3(100)$. So, 200 and 800 are three (3) standard deviations from the mean, 500. According to the 68–95–99.7, 99.7% of all scores are between 200 and 800.

b. the percent of scores are lower than 300.

ANS: $300 = 500 - 2(100)$ and $700 = 500 + 2(100)$. So, 300 and 700 are two (2) standard deviations from the mean, 500. According to the 68–95–99.7, 95% of all scores are between 300 and 700. So, 5% of all scores are below 300 and above 700. Since the normal curve is symmetric, there is an equal amount of area below 300 and above 700. So, 2.5% (i.e., one-half of 5%) of scores are below 300.

c. the score above which 16% of all the scores lie?

ANS: $400 = 500 - 1(100)$ and $600 = 500 + 1(100)$. So, 400 and 600 are one (1) standard deviation from the mean, 500. According to the 68–95–99.7, 68% of all scores are between 400 and 600. So, 32% of all scores are below 400 and above 600. Since the normal curve is symmetric, there is an equal amount of area below 400 and above 600. So, 16% (i.e., one-half of 32%) of scores are above 600.