AP Review Assignment #1 Name _____ Per. _____

Topic I. Exploratory Analysis of Data

1. The data in the chart below shows the survival times in days for guinea pigs after they were injected with tubercle bacilli in a medical experiment.

43	45	53	56	56	57	58	66	57	73	74	79	80	80	81	81	81	82	82	83
83	84	88	89	91	91	92	97	99	99	100	101	102	102	102	103	104	107	108	109
113	114	118	121	123	126	128	137	138	139	144	147	156	162	174	178	179	184	191	198
211	214	243	249	329	380	403	511	522	508	510	514	520	520	521	530	530	533	540	541

a. Create a histogram for the data in the chart above:

b. Discuss the main features of the histogram. (Remember your SOCS?)

c. Find the following values for the data.

Measures of Center: Median: _____ Mean: _____

Measures of Spread: Range:_____ IQR:_____Standard Deviation:_____ Variance:_____

Measures of Position: Q1:____ Q2:___ Q3:____ Min:_____ Max:_____ The 70th Percentile:_____

c. Find the standardized scores (z-scores) for 80 days and 520 days.

d. List the 5-number summary and create the modified box plot for the data.

e. Identify any outliers by using the IQR method. (Show work.)

f. If the data were changed in the following ways, which one of the summary measures would change and how would they change?

Change the max days to 1000:

Change the unit of measures by dividing every piece of data by 100:

2. The following quiz scores are from 2 different classes for an AP Stats test in chapter 1.

4th Hour	48	76	82	96	92	84	100	98	96	76	92	72	88	82	66	58	78	81	78
78	92	92	78	84	52	70	84	88	92	84									

5th Hour	90	96	78	94	94	88	86	96	86	82	90	87	88	76	92	94	80	82	88
84	86	80	86	72	96	90													

a. Create back-to-back box-plots (on the same scale) and compare them on the following:

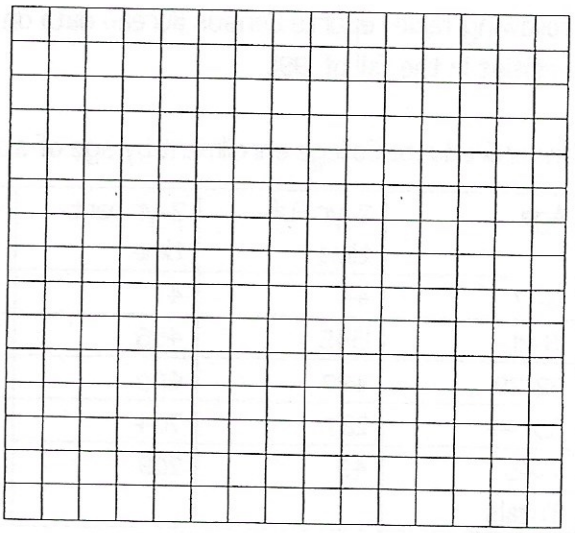
Spread: Center: Outliers: Shape:

3. Is there a correlation between test anxiety and exam score performance? Data on x = score on a measure of test anxiety and y = exam score are given in the table below.

X = test anxiety	23	14	14	0	7	20	20	15	21
Y = score on exam	43	59	48	77	50	52	46	51	51

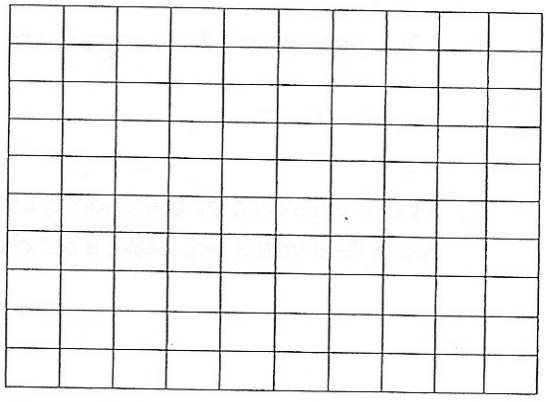
a. Which one of the variables is the explanatory and which is the response variable?

b. Construct a scatter plot and comment on the features of the plot. (Overall pattern, deviations, direction, form, strength)



c. Find the correlation coefficient, the coefficient of determination and the LSRL.

d. Construct a residual plot.



e. Comment on the relationship between test anxiety and test scores based upon the analysis you performed.

f. If we were to add the data point (5, 100) how would it affect the LSRL? What is this point called?

4. The sample correlation coefficient between annual raises and teaching evaluations for a sample of 353 college faculty was found to be r = .11.

   a. Interpret this value.

   b. If a straight line were fit to the data using least squares regression, what proportion of variation in raises could be attributed to the approximate linear relationship between raises and evaluation?

5. The following table reports Census Bureau data on undergraduate students in U.S. colleges and universities in the fall of 1991.

UnderGraduate College enrollment by age of students – Fall 1991 (thousands of students)

Age	2-yr Full-time	2-yr part-time	4-yr full time	4-yr part-time	Totals
15-17	44	4	79	0	
18-21	1345	456	3869	159	
22-29	489	690	1358	494	
30-44	287	704	289	627	
>=45	49	209	62	160	
Totals					GT ( )

a. Fill in the "totals" in the table above. What is the grand total (GT) of students who were enrolled in colleges and universities in the fall of 1991?

b. What percent of all undergraduate students were 18-21 years old in the fall of the 1991?

c. Find the percent of the undergraduates enrolled in each of the four types of programs who were 18-21 years old. Make a bar chart to compare these percents.

d. The 18-21 group is the "traditional" age group for college students. Briefly summarize what you have learned from the data about the extent to which this group predominates in different kinds of college programs.