AP Statistics Assignments
Chapter 1—Exploring Data

****MAKE ALL GRAPHS LARGE, NEAT, CLEARLY LABELED AND EASY TO READ!!**

Assignment #	Due	Section	Assignment	Credit
14		1.1	p7 #1.2 p11 #1.6 p17 #1.10 (make a stemplot & write a paragraph)	
15	Tu 9.29	1.1	p22 #1.12ab only p26 #1.16, 1.17	
16	W	1.1	p31 #1.19, 1.20 p33 #1.22	
17	Th	1.2	p41 #1.32, 1.35 p47 #1.36 (goes to next page)	
18	F	1.2	p52 #1.40, 1.43 p56 #1.45, 1.46	
19	M	1.2	2006 AP #1 2006B AP #1 (both given as handouts)	
20	Tu	Chapter 1 Review	Chapter 1 Practice Problems (handout)	

<u>IMPORTANT DATES:</u>

W 09.23	*John Coltrane's 82nd birthday*
W 09.23	*Les McCann's 73rd birthday*
Sa 09.26	*George Gershwin's 110th birthday*
Sa 09.27	*Bud Powell's 84th birthday*
W 09.30	**Project Proposal Due**
M 10.05	**Chapter 1 Group Test**
W 10.07	**Chapter 1 Test**
Fr 10.16	**Semester 1 Project Due**

AP Statistics
Chapter 1—Displaying, Describing, and Comparing Data

I. Misleading Graphs

Write some observations about the different ways in which graphs can be misleading:

II. Graph #1: Bar Graph

Type of data:________________

Purpose:

Example:

III. Graph #2: Pie Chart

Type of data:________________

Purpose:

Example:

IV. Graph #3: Dotplot

Type of data:__________________

Purpose:

Record the greed game scores in the table below, and draw a dotplot:

V. Graph #4: Stemplot (or stem and leaf plot)

Type of data:__________________

Purpose:

Draw a stemplot (don't forget the key!) of the greed game data here:

VI. Graph #5: Histogram (& frequency table)

Type of data:__________________

Purpose:

Things to remember (for calculator also):

Draw a frequency table and histogram of the greed game data here:

Frequency Table:

Score Range	Frequency

Histogram

VII. Relative Cumulative Frequency, Percentiles, and Graph #6: Ogive

Type of data________________ Purpose:

Greed Game Score Range	Frequency	Relative Frequency (%)	Cumulative Frequency	Cumulative Relative Frequency (Cumulative %)

Now use this data to make a Cumulative Relative Frequency graph (an **Ogive**):

The **p^th^ percentile** of a set of data is defined as the value which has *p* percent of the data below it. Use the graph to find the approximate location of the:

a) 20^{th} percentile:____________ b) 50^{th} percentile (median):________________

c) 75th percentile:___________ d) What percentile is a score of 25? __________

VIII. Graph #7: Time Plot

Type of data:__________________ Example:

Purpose:

Things to remember:

IX. Numeric Descriptions of Data

In this class, when using numbers to summarize a set of data, you'll almost always be using either the **mean** and **standard deviation** or what we call a **five-number summary**. Calculate these for the following set of data (notice that the data have already been arranged in order):

2 2 2 3 5 6 6 6 7 10 10 14 16 19 20 20 20 60

mean =

standard deviation =

five number summary:

low =
Q1 =
median =
Q3 =
high =

What can we determine by comparing the mean and median of a set of data?

What do we mean when we say that the median is a **resistant** measure of center, while the mean is not?

X. Graph #8: Boxplot (box and whisker plot)

Type of data:____________________ Make a boxplot of the greed game data:

Purpose:

Things to remember (for calculator also):

XI. Outliers—An outlier is an "unusually" high or low observation in a set of data. One common mathematical method for determining whether an unusual value is an outlier is based on the **interquartile range (IQR = Q3 – Q1).** Under this rule, an observation is considered an outlier if either of the following are true:

(observation) > Q3 + 1.5*IQR (a high outlier)
(observation) < Q1 – 1.5*IQR (a low outlier)

In the space below, use this test to determine whether any of your Greed Game scores were outliers:

Low outlier boundary =

High outlier boundary =

Any outliers?

Now use your calculator to make a boxplot of this data. Just by looking, you can tell fairly easily whether there are any outliers according to this rule. If there were no outliers, change the highest or lowest value so that it is an outlier and then make a regular boxplot again, then make a **modified boxplot** (that shows outliers). Draw it here:

XII. Variance and Standard Deviation—Choose 7 random people from the class and record their highest bowling scores in the table below. Calculate the mean of these 7 scores, then make a dotplot of the 7 scores in the space below. On the dotplot, draw an arrow at the mean.

Now look at the distance between each observation and the mean. What do you think is the mean distance from the mean (take a guess)?________ The **standard deviation (*s* when we're talking about a sample like this, σ if the data we have are an entire population, like the whole class)**, which is a measure of **spread**, is a mathematical measure of the "average" (remember that "average" does not always indicate the mean) distance, or a typical distance from the mean in a set of data. It, like the **IQR** and the **range**, gives a numeric representation of the degree of spread in a set of data.

Is **s** a resistant measure of spread?

Is the **range** a resistant measure of spread?

Is the **IQR** a resistant measure of spread?

Now for the calculations:

Highest Bowling Score x_i	Deviation From the Mean $x_i - \bar{x}$	Squared Deviation From the Mean $(x_i - \bar{x})^2$
Sum = $\sum x_i$ = Mean = $\bar{x}$ =	Sum = $\sum (x_i - \bar{x})$ =	Sum = $\sum (x_i - \bar{x})^2$ =

Why does it make sense to square the deviations from the mean?

Now take the sum in the 3rd column and divide by (n-1):

Variance = s^2 = $\frac{\sum(x_i - \bar{x})^2}{n-1} =$

This value is called the **variance (s^2 for a sample, σ^2 for a population)**. This represents the average (but not the mean!) squared distance from the mean. Dividing by (n-1) instead of n inflates this value a little bit, which is appropriate because we only have a sample of the population, and we don't want to underestimate the spread in the population and draw conclusions that are false. We'll talk more about this later in the course.

The variance is the square of the standard deviation. So, to get the standard deviation, all we have to do now is take the square root:

Standard Deviation = s = $\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} =$

Again, this number represents an **average, or typical, distance from the mean** in our set of data.

Why does it make more sense to report the standard deviation instead of the variance?

So now we have 2 choices when giving a numeric summary of a set of data:

1. The 5-number summary (Min, Q1, Median, Q3, Max)
2. The mean ($\bar{x}$) and standard deviation (s)

How do we choose which one to use?

XIII. Linear Transformations of Data—A linear transformation of data occurs when we take every observation in a data set and mathematically change it (every observation in the same way) with multiplication (but not squaring), division, addition, and/or subtraction. When would we do this?

Our question is: What happens to our measures of center (mean and median) and spread (s, range, IQR) when we do this?

Make a dotplot of this data (scale the dotplot from 0 to 15): 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 4, 4

Now **multiply each observation by 3** and make a new dotplot (same scale):

Effect on: mean:

median:

range:

s.d.:

IQR:

Now **add 3 to each observation** (the original ones) and make a new dotplot (same scale):

Effect on: mean:

median:

range:

s.d.:

IQR:

Now **multiply by 3 and add 3** (to the original data) and make a new dotplot (same scale):

Effect on: mean:

median:

range:

s.d.:

IQR:

XIV. Comparing Distributions—To compare distributions, you first need to make a graph that makes it easy to see the 2 distributions side-by-side:

a) 2 dotplots or histograms on the same scale
b) Back-to-back stemplot
c) Side-by-side boxplots

You should also calculate numeric summaries of each distribution. Then, using the C.U.S.S. framework, compare each aspect of the distributions. Be sure to use **words of comparison (higher, lower, more, less, similar, etc.)** for each aspect!! Don't just separately list characteristics of each, that's not a comparison! An example (from p57 in your text):

A study in Switzerland examined the number of cesarean sections performed in a year by doctors. Here are the data for 15 male doctors and 10 female doctors:

Male doctors: 27, 50, 33, 25, 86, 25, 85, 31, 37, 44, 20, 36, 59, 34, 28

Female doctors: 5, 7, 10, 14, 18, 19, 25, 29, 31, 33

Compare the number of cesarean sections performed by these male and female doctors. Begin by making a back-to-back stemplot.