POLI 300 PROBLEM SET #7 due 11/08/10

MEASURES OF DISPERSION AND THE NORMAL DISTRIBUTION

NAME ________________________________

Put all your answers directly on these pages

1. Refer to the continuous frequency density provided with Problem Set #5C (and also used in PS #6), showing the (hypothetical) distribution of income in $000 among U.S. households.

   (a) What is the (approximate) *interquartile range* of incomes? _____________

   Interdecile range? _____________

   (b) Why is the [total] *range* itself not really usable as a measure of dispersion for this data?

   (c) Make a reasonable *guess* of the mean deviation (easier to guess) and standard deviation (somewhat larger than the mean deviation) of income according to this diagram.

   (d) If mean income remains constant but its distribution becomes *more equal,* how would the frequency distribution and its SD change?

   (e) If the income distribution becomes *less equal*, how would the frequency density and its SD change?

   (f) If everyone's income doubled, what would be the impact on the distribution of income and SD?

   (g) If everyone's income increased by $25,000, what would be the impact on the distribution of income and SD?

2. Twenty households have a total of fifty children (see #2 in Problem Set #6). Given only this information, can you estimate the standard deviation (or any other measure of dispersion) for the variable NUMBER OF CHILDREN among these households? If so, what is it? If no such estimate can be made, why not? What is the smallest possible SD in such data? What is the *largest possible* SD in this data?

3. A survey of households measures the number of children in each household. The frequency distribution is shown to the right (see #3 in Problem Set #6):

# of children	*rel. freq.*
0	10%
1	26%
2	16%
3	14%
4	10%
5	8%
6	6%
7	4%
8	3%
9	2%
10	1%
	100%

Make a reasonable *guess* of the *mean deviation* and *standard deviation* of the number of children per household. How did you do this? (Do not try to calculate either the MD or SD by hand unless you believe in self-punishment.)

4. Make a reasonable guess of the *mean deviation* and *standard deviation* of snowfall in Washington DC. (See the chart for Question #6 in PS #6.)

5.

Case #	*X*
1	2
2	7
3	1
4	15
5	4
6	6
7	2
8	3

Compute the *standard deviation* of the variable X for the data pertaining to the set of eight cases given above. ***Show all your work***, preferably using the worksheet format used Handout #7 on *Dispersion*.

6. (a) Add 5 to each value of x above. How does this affect the mean of X ? The SD of X ?

(b) Multiply each value of x by 5. How does this affect the mean of X ? The SD of X ?

7. A group of people has a total of $100. The $100 is then redistributed within the group, some people getting more and others less than before (but the total remains $100). As a result of this redistribution,

(a) *may* the mean of the distribution change? Yes / No
(b) *must* the mean of the distribution change? Yes / No
(c) *may* the SD of the distribution change? Yes / No
(d) *must* the SD of the distribution change? Yes / No
(e) What distribution of the $100 has the *minimum* SD?
(f) What distribution of the $100 has the *maximum* SD

8. We measure the weight in pounds of every person in a large group. The median is 152, the mean is 159, the range is 217, the variance is 576, and the standard deviation is 24. We now "go metric" and remeasure everyone's weights in kilograms. (One pound ≈ 0.453 kilograms.) What are the numerical values of the following summary statistics?

(a) the median _______________
(b) the mean _______________
(c) the range _______________
(d) variance _______________
(e) the standard deviation _______________

9. Consider three possible situations in group consisting of Jim, Bob, and Ann:
(i) Jim earns $10, Bob earns $20, and Ann earns $30;
(ii) Jim earns $50, Bob earns $100, and Ann earns $150; and
(iii) Jim earns $110, Bob earns $120, and Ann earns $130.

(a) Calculate the *Standard Deviation* of earnings in each situation (i), (ii), and (iii).

(i) SD = ____________
(ii) SD = ____________
(iii) SD = ____________

(b) C alculate the *Coefficient of Variation* in each situation (i), (ii), and (iii).

(i) CV = ____________
(ii) CV = ____________
(iii) CV = ____________

(c) Which measure of dispersion best reflect our sense of inequality with respect to Jim's, Bob's, and Alice's earnings?.

10. On the enclosed page, you will find bar charts showing the frequency distributions of NES respondents' perceptions of the ideological positions of the 1992 Presidential candidates. (This data comes from SETUPS/NES 1992 data, not the SETUPS/NES1972-2000 data, which has no measures pertaining to Ross Perot. However, the questions are similar to Questions 25-35 in the Student Survey and variables V35-38 in the SETUPS/NESS Codebook.) In answering the following questions, *treat the code values* (i.e., 1 = Liberal, 2 = Slightly Liberal, etc.) *as if they form an interval scale*.

    (a) What is the *range* of data in each chart?

    (b) Why is the range an almost totally unhelpful measure of dispersion for such data?

    (c) Which chart/variable appears to have the *smallest* SD?

    (d) Which chart/variable appears to have the *largest* SD?

    (e) What is the minimum SD you could possibly find for such a variable (whose potential values run from 1 to 5 in the manner of ideology)? Draw a bar graph of the distribution that produces this minimum SD.

    (f) What is the maximum SD you could possibly find for such variables? Draw a bar graph of the distribution that produces this maximum SD.

11. As you (should) know, the number of electoral votes each state has is equal to its number of House seats plus two. Comparing (i) the distribution of House seats and (ii) the distribution of electoral votes,

(a) which has the greater *range* (or are they equal)?

(b) which has the greater *mean deviation* (or are they equal)?

(c) which has the greater *standard deviation* (or are they equal)?

(d) which has the greater *coefficient of variation* (or are they equal)?

Note. The remaining questions are taken or adapted from David S. Moore, *Statistics: Concepts and Controversies*, a text book previously used in this course.

12. This is a variance (or SD) contest. You must choose four numbers from the set of whole numbers 0 through10, with repeat choices either allowed or not allowed.

(a) Choose four numbers that have the smallest possible standard deviation with repeats allowed.

Are two or more sets of four numbers equally good?

(b) Choose four numbers that have the smallest possible standard deviation with repeats not allowed.

Are two or more sets of four numbers equally good?

(c) Choose four numbers that have the largest possible standard deviation with repeats allowed.

Are two or more sets of four numbers equally good?

(d) Choose four numbers that have the largest possible standard deviation with repeats not allowed.

Are two or more sets of four numbers equally good?

13. A school system employs teachers at salaries between $30,000 and $60,000. The teachers' union and the school board are negotiating the form of next year's increase in the salary schedule. Suppose that every teacher is given a flat $1000 raise.

(a) How much will the mean salary increase? The median salary?

(b) Will a flat $1000 raise increase the spread as measured by the distance between the quartiles (i.e., between those at the 75th percentile and the 25th percentile)?

(c) Will a flat $1000 raise increase the spread as measured by the standard deviation of the salaries?

14. Suppose that the teachers in the previous exercise each receive a 5% raise. The amount of the raise will vary from $1500 to $3000, depending on present salary. Will a 5% across-the-board raise increase the spread of the distribution as measured by the distance between the quartiles? Do you think it will increase the standard deviation?

OVER =>

Note. Answering the two remaining questions focus on concepts (*standard scores* and the *normal distribution*) introduced in Handout #8.

15. Three landmarks of baseball achievement are Ty Cobb's batting average of .420 in 1911, Ted Williams's .406 in 1941, and George Brett's .390 in 1980. These batting averages cannot be compared directly because the distribution of major-league batting averages has changed over the years. The distributions are quite symmetric and (except for outliers such as Cobb, Williams, and Brett) reasonably "normal." While the mean batting average has been held roughly constant by rule changes and the balance between hitting and pitching, the standard deviation has dropped over time. Here are the facts:

Decade	*Mean Batting Average*	*SD of Batting Averages*
1910s	.266	.0371
1940s	.267	.0326
1970s	.261	.0317

Compute the standard scores for the batting averages for Cobb, Williams, and Brett to compare how far each stood above his peers.

16. The distribution of heights of young men is approximately normal with a mean 70 inches and a standard deviation of 2.5 inches. Use the 68-95-99.7% rule to answer the following questions.

(a) What percent of men are taller than 75 inches?

(b) Between what heights do the middle 95% of men fall?

(c) What percent of men are shorter than 67.5 inches?