Soc 3811 Basic Social Statistics
First Midterm Exam Spring 2010

ANSWERS

INSTRUCTIONS:

(A) Write your name on the line at top front of <u>every sheet</u>.

(B) If you use a page of notes in taking this exam, sign & insert it inside this booklet before turning in your exam.

(C) *Show your calculations for numerical problems in the space provided!*
- -

1. Fill in the blanks **[1 point each]:**

   a. The graph of an continuous variable is called a(n)

   _____**Histogram or polygon**_____.

   b. Descriptive statistics that describe the value most typical of the scores in a frequency distribution are called

   Measures of central tendency
   _____.

   c. Variables like EDUC and AGE, which could have all possible numerical values in a given interval, are called

   _____**Continuous**___.

   d. A variable that classifies persons, objects, or events into two mutually exclusive and exhaustive categories is called a(n)

   _____**Dichotomy**___.

   e. Many card games assign increasing values to the four face cards: Jack, Queen, King, Ace. What type of variable is this?

   Orderable discrete
   _____.

2. This table shows responses to the 2008 GSS variable THEISM, "There is a God who concerns Himself with every human being personally." Complete the last two columns of the table and report the modal category. **[5 points]:**

	CATEGORY	Frequency (f_i)	Percent (%)	Valid Percent (%)
Valid	1. Stongly agree	578	28.6	43.0
	2. Agree	368	18.2	27.4
	3. Neither agree nor disagree	185	9.1	13.8
	4. Disagree	136	6.7	10.1
	5. Strongly disagree	76	3.8	5.7
	Total	1,343	(66.4)	100.0
Missing	0. IAP	658	32.5	
	9. No answer, don't know	22	1.1	
	Total	680	(33.6)	
TOTAL		2,023	100.0	

MODE: _____ 1. Strongly agree _____

Your name: _____

3. This table displays POLVIEWS in the 2008 GSS. Calculate cumulative frequencies, valid percentages, cumulative percentages, and report the median. **[5 points]:**

DO YOU THINK OF YOURSELF AS LIBERAL OR CONSERVATIVE?	Frequency (f)	Cumulative Frequency (cf)	Valid Percent (%)	Cumulative Percent (c%)
1. Extremely liberal	69	69	3.6	3.6
2. Liberal	240	309	12.4	16.0
3. Slightly liberal	221	530	27.4	27.4
4. Moderate	740	1,270	38.3	65.7
5. Slightly conservative	268	1,538	13.9	79.6
6. Conservative	327	1,865	16.9	96.5
7. Extremely conservative	68	1,933	3.5	100.0
TOTAL	1,933	--	100.0	
8. Don't Know	77	--		
9. No Answer	13	--		
TOTAL	2,023	--		

MEDIAN: ___**4. Moderate**_____

4. Calculate the Index of Diversity and Index of Qualitative Variation for this table showing 2008 GSS variable SCISTUDY, which asked how well the respondents understand the term "scientific study." **[5 points]**

	SCISTUDY	Frequency (f)	Percent (%)	Valid Percent (%)
Valid	1. Clear understanding	424	21.0	28.6
	2. General sense	749	37.0	50.5
	3. Little understanding	309	15.3	20.9
	Total	1,482	73.3	100.0
Missing	9. IAP; DK, NA	541	26.7	
TOTAL		2,023	100.0	

$$D = 1 - (p_1^2 + p_2^2 + p_3^2)$$
$$= 1 - (.286^2 + .505^2 + .209^2)$$
$$= 1 - 0.381 = 0.619$$

$$IQV = \left(\frac{K}{K-1}\right) D$$
$$= \left(\frac{3}{3-1}\right) 0.619 = \left(\frac{3}{2}\right) 0.619 = 0.9285$$

D: ___**0.619**___   **IQV:** ___**0.929**___

Your name: _____

5. Calculate the mean, median, mode, and range of an ungrouped variable with these scores. **[5 points]**

$$Y_1 = 3 \qquad Y_4 = 5 \qquad Y_7 = 7$$
$$Y_2 = 5 \qquad Y_5 = 9 \qquad Y_8 = 5$$
$$Y_3 = 7 \qquad Y_6 = 11 \qquad Y_9 = 11$$

Mode: 3, 5, 5, 5, 7, 7, 9, 11, 11

Median: 3, 5, 5, 5, 7, 7, 9, 11, 11

$$\overline{Y} = \sum_{i=1}^{N} \frac{Y_i}{N}$$
$$= \frac{3+5+5+5+7+7+9+11+11}{9}$$
$$= \frac{63}{9} = 7.0$$

Range: 11 – 3 = 8

MODE: __5__ **MEDIAN:** __7__ **MEAN:** __7.0__ **RANGE:** __8__

6. Now calculate the variance and standard deviation for the same ungrouped data in question 5. **[5 points]**

$$
\begin{array}{lll}
Y_1 = 3 & Y_4 = 5 & Y_7 = 7 \\
Y_2 = 5 & Y_5 = 9 & Y_8 = 5 \\
Y_3 = 7 & Y_6 = 11 & Y_9 = 11
\end{array}
$$

$$
s_Y^2 = \frac{\displaystyle\sum_{i=1}^{N} (Y_i - \overline{Y})^2}{N-1}
$$

$$
= \frac{1(3-7)^2 + 3(5-7)^2 + 2(7-7)^2 + 1(9-7)^2 + 2(11-7)^2}{9-1}
$$

$$
= \frac{64}{8} = 8.0
$$

$$
s_Y = \sqrt{8} = 2.83
$$

VARIANCE: __3.25__   STAND. DEV. __1.80__

Your name: _____

7. This table shows how often people pray. Find the mode and median. Then write an interpretation of those two central tendency statistics, explaining which measure more accurately describes how often people pray. **[5 points]**

| HOW OFTEN DOES R PRAY				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1. Never	138	6.8	10.2	10.2
2. Less than once a week	28	1.4	2.1	12.3
3. About 1-2 times per year	25	1.2	1.9	14.2
4. Several times a year	68	3.4	5.0	19.2
5. About once a month	38	1.9	2.8	22.0
6. 2-3 times per month	63	3.1	4.7	26.7
7. Nearly every week	27	1.3	2.0	28.7
8. Every week	66	3.3	4.9	33.6
9. Several times a week	237	11.7	17.6	`51.2`
10. Once a day	268	13.2	19.9	71.1
11. Several times a day	`391`	19.3	`29.0`	100.0
Total	1,349	66.7	100.0	
Missing				
9. INAP, DK, NA	674	33.3		
TOTAL	2,023	100.0		

MODE: 11. Several times a day

MEDIAN: 9. Several times a week

The mode is the category with the most cases, but it occurs in the highest category this ordered variable's distribution. The median is the category that exactly divides the distribution in half (where the 50th percentile occurs).

The median of "several times a week" more accurately represents the typical frequency because it occurs closer to the middle of the variable's values than does "several times a day."

8. This graph, using using different frequencies than in problem #7, has these descriptive statistics for praying: mean = 8.2; median = 10.0; variance = 11.2. Calculate skewness and interpret the graph's shape, including the type of skew. **[5 points]**
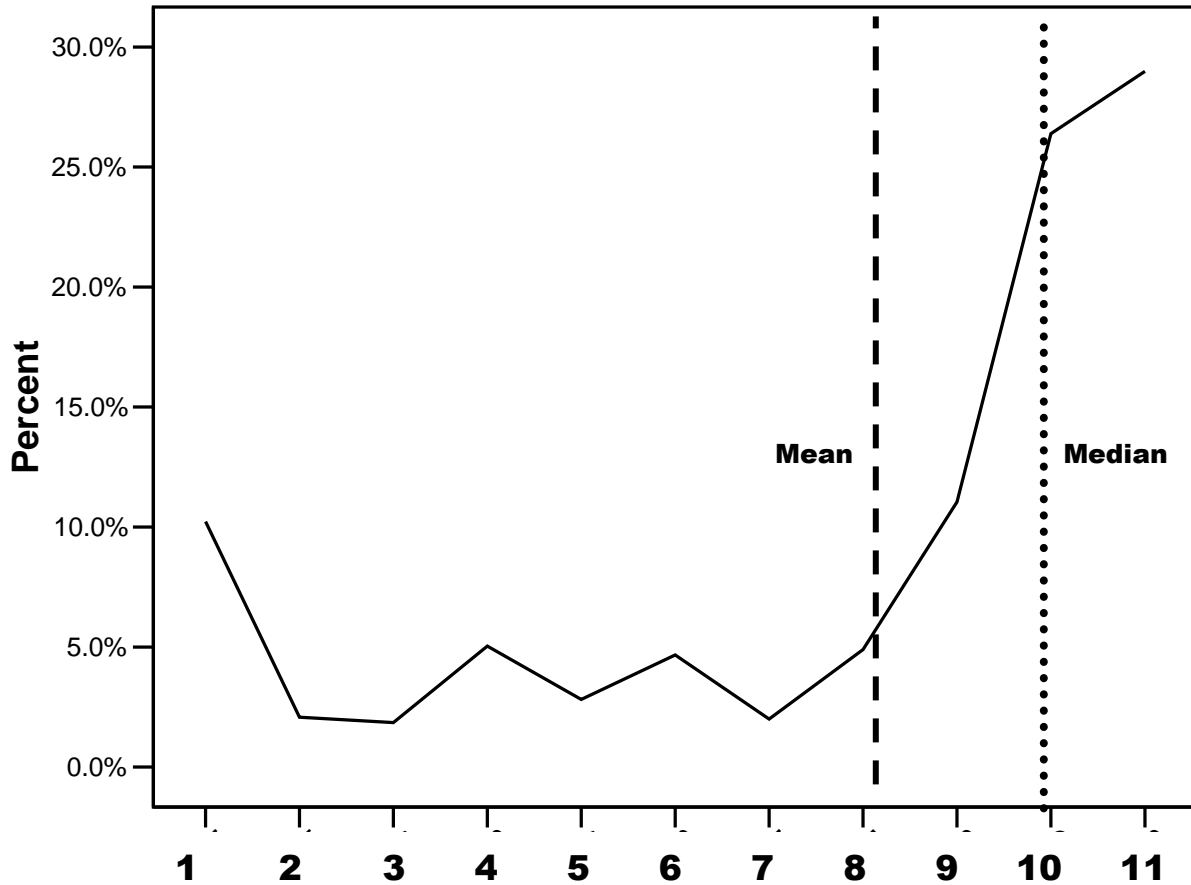
$$Skew = \frac{3(\bar{Y} - Mdn)}{s_Y}$$

$$= \frac{3(8.2 - 10.0)}{\sqrt{11.2}} = \frac{-5.4}{3.35} = -1.61$$

INTERPRETATION:

The distribution has a negative skew, with its tail to the left, because the mean praying is lower than the median. Although the average praying was 8, half the sample prayed 10 or more times.

9. This table displays the number of books bought by 60 students last month. The mean is 3.00. Calculate the variance and standard deviation for this grouped frequency distribution. **[5 points]**

NUMBER OF BOOKS	
Score (Y_i)	Frequency (f_i)
1	4
2	13
3	30
4	5
5	8
TOTAL	60

$$s_Y^2 = \frac{\sum\limits_{i=1}^{K} (Y_i - \overline{Y})^2 (f_i)}{N-1}$$

$$= \frac{\left(\begin{array}{l} (1-3)^2(4) + (2-3)^2(13) + (3-3)^2(30) + \\ (4-3)^2(5) + (5-3)^2(8) \end{array} \right)}{60-1}$$

$$= \frac{66}{59} = 1.112$$

$$s_Y = \sqrt{1.112} = 1.058$$

VARIANCE: **1.112**   STAND. DEV. **1.058**

10. The 2008 GSS variable SIBS ("How many brothers and sisters did you have?") has these descriptive statistics for 2,021 respondents: mode = 2; median = 3; mean = 3.6; range = 55; variance = 10.2. Calculate the standardized scores (Z_i scores) for three respondents with these numbers of siblings (Y_i). **[5 points]**

Y_i Z_i

1 $Z_i = \dfrac{Y_i - \overline{Y}}{s_Y} = \dfrac{1 - 3.6}{3.19} = -0.81$

5 $Z_i = \dfrac{Y_i - \overline{Y}}{s_Y} = \dfrac{5 - 3.6}{3.19} = +0.44$

12 $Z_i = \dfrac{Y_i - \overline{Y}}{s_Y} = \dfrac{12 - 3.6}{3.19} = +2.63$

Calculate the expeced number of siblings (Y_i) for two respondents with these Z_i scores.

Z_i Y_i

-0.4 $Y_i = Z_i(s_Y) + \overline{Y} = (-0.4)(3.19) + 3.6 = 2.32$

+2.2 $Y_i = Z_i(s_Y) + \overline{Y} = (+2.2)(3.196) + 3.62 = 10.62$