

# Homework 1

## SOCI 303: Statistics for the Social Sciences (Spring 2017)

10 points

### Overview:

In this assignment, you'll be completing 10 problems from the "Chapter Exercises" section of Chapters 1 through 5. This homework assignment may be typed up or hand written or a combination of both.

### Chapter 1 Problems

- 3 (a, b, c, d, e, f, g, h)
  - (a) Interval-ratio
  - (b) Nominal
  - (c) Interval-ratio
  - (d) Ordinal
  - (e) Nominal
  - (f) Interval-ratio
  - (g) Interval-ratio
  - (h) Nominal
- 6 (a, b, d, e)
  - (a) Unemployment records could be used to determine the actual number of unemployed; a descriptive statistic based upon the population.
  - (b) A survey is taken to estimate student opinions about the quality of food; inferential statistic.
  - (d) The ratings will be gathered from a survey, so this is inferential.
  - (e) A university should be able to report GPA by major, so this is a descriptive statistic based upon the population.

### Chapter 2 Problems

- 1 (a, b)
  - (a) Race is a nominal variable. Class is an ordinal variable, since the categories can be ordered from lower to higher status.

	Race	Frequency ( $f$ )
(b)	White	17
	Nonwhite	13
	Total	30

- 6 (a, b)

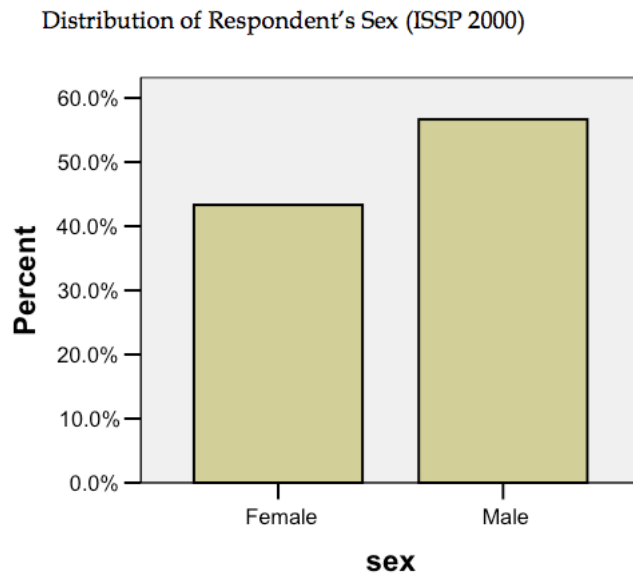
(a)

Email hours per week	Frequency	Cf	%	C%
0	19	19	19	19
1	20	39	20	39
2	13	52	13	52
3	5	57	5	57
4	2	59	2	59
5	6	65	6	65
6	5	70	5	70
7	2	72	2	72
8	3	75	3	75
9	1	76	1	76
10 or more	23	99	23	99
Total	99		99%	

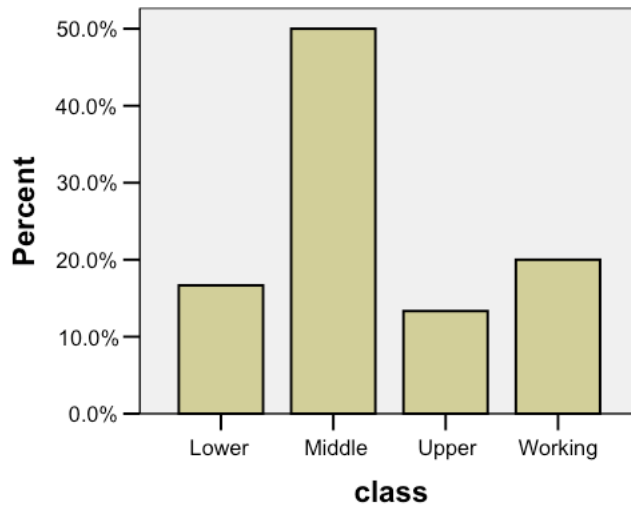
(b) .575 (57/99) spent 3 hours or less on email per week.

## Chapter 3 Problems

- 4



Distribution of Social Class (ISSP 2000)



## Chapter 4 Problems

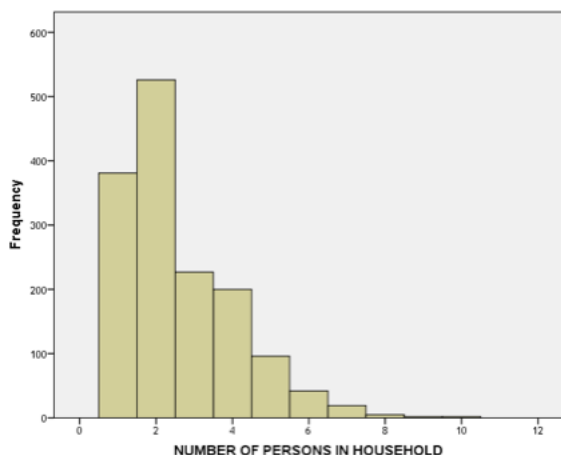
- 1 (a, b, c, d)
  - (a) The mode is “Exciting.”
  - (b)  $\frac{N+1}{2} = \frac{974+1}{2} = 487.5$ th case. Find where the 487th and 488th cases fall. Both of these cases correspond to the category “Routine,” which is the median.
  - (c) The mode is simply the category with the highest frequency (or percentage) in the distribution. The median divides the distribution into two equal parts so that half the cases are below it and half above it.
  - (d) Because this variable is an ordinal-level variable.

• 7

	Household Size	Frequency	Frequency x Y ( <i>fY</i> )
	1	381	381
	2	526	1,052
	3	227	681
	4	200	800
(a)	5	96	480
	6	42	252
	7	19	133
	8	5	40
	9	2	18
	10	2	20
	Total	$N = 1,500$	$\Sigma fY = 3,857$

$$\bar{Y} = \frac{\Sigma fY}{N} = \frac{3857}{1500} = 2.5713$$

- 9 (a, b)
  - (a) There appear to be a few outliers (i.e., extremely high values); this leads us to believe that the distribution is skewed in the positive direction.



- (b) The median can be found two ways: by using either the frequencies column or the cumulative percentages. The data are in frequencies; we'll use those to solve the median. Since the median (2) is less than the mean (2.57), we can conclude that the distribution is skewed in a positive direction. Our answer to question 9a is further supported.

## Chapter 5 Problems

- 5 (a, b, c)

- (a) The range of projected increase in the elderly population for the western states is 36.2%. The range of percent increase for the Midwestern states is 9.8%. The western states have a much larger range.
- (b) The IQR for the western states is 17.3%. The IQR for the Midwestern states is 3.7%. Again, the value for the western states is greater.
- (c) There is great variability in the projected increase in the elderly population in western states, resulting from large increases in Nevada, Arizona, Wyoming, and Alaska, as measured by either the range or the IQR.

- 8 (a, b)

- (a) For those diagnosed with cancer:

$$\text{Variance} = \frac{\Sigma(Y - \bar{Y})^2}{N-1} = \frac{3059.14}{187-1} = 16.45$$

$$SD = S_Y = \sqrt{S_Y^2} = \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{N-1}} = \sqrt{\frac{3059.14}{187-1}} = \sqrt{16.45} = 4.06$$

For those NOT diagnosed with cancer:

$$\text{Variance} = \frac{\Sigma(Y - \bar{Y})^2}{N-1} = \frac{25180.20}{1200-1} = 21$$

$$SD = S_Y = \sqrt{S_Y^2} = \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{N-1}} = \sqrt{\frac{25180.20}{1200-1}} = \sqrt{21} = 4.58$$

- (b) Although there is a slightly more variability in psychological distress score among those not diagnosed with cancer than those diagnosed with cancer, the standard deviations of the two groups are very close (4.58 vs. 4.06). An interesting finding is that the respondents diagnosed with cancer are not only slightly less diverse in terms of the distress level, but their average psychological distress score is also less than those not diagnosed with cancer (3.9 vs. 4.87).