Social Studies 201
Fall 2006
Answers to Problem Set No. 2
October 3, 2006

1. **Retirement experiences**.

   (a) In order to construct the histograms as accurately as possible and obtain the proper interval widths and densities, the real class limits are used for each interval. Table 1 contains the calculations for the histograms. The histograms are in Figures 1 and 2.

Table 1: Frequency distributions and densities for household income of retired Saskatchewan respondents, satisfied and not satisfied

	Real class	Interval	Satified		Not satisfied	
Income	limits	Width (w)	P	Density (P/w)	P	Density (P/w)
0-19	-0.5-19.5	20	11	0.55	28	1.40
20-29	19.5-29.5	10	18	1.80	20	2.00
30-39	29.5-39.5	10	16	1.60	13	1.30
40-59	39.5-59.5	20	24	1.20	19	0.95
60-99	59.5-99.5	40	31	0.78	20	0.50
Total			100		100	

   (b) The measures of central tendency are as follows.

   Mode. The mode for each distribution is at the peak of the histogram, or the interval with the greatest density of occurrence. For those satisfied with their retirement experiences, the modal interval is 20-29, with a density of 1.80, greater than for any other interval. Or the mode could be the midpoint of this interval, an income of 24.5 thousand dollars.

   For those not satisfied with their retirement experiences, the same 20-29 interval has the greatest density, so the mode is 20-29, or 24.5, thousand dollars.

Figure 1: Histogram of household income of Saskatchewan respondents satisfied with retirement experiences

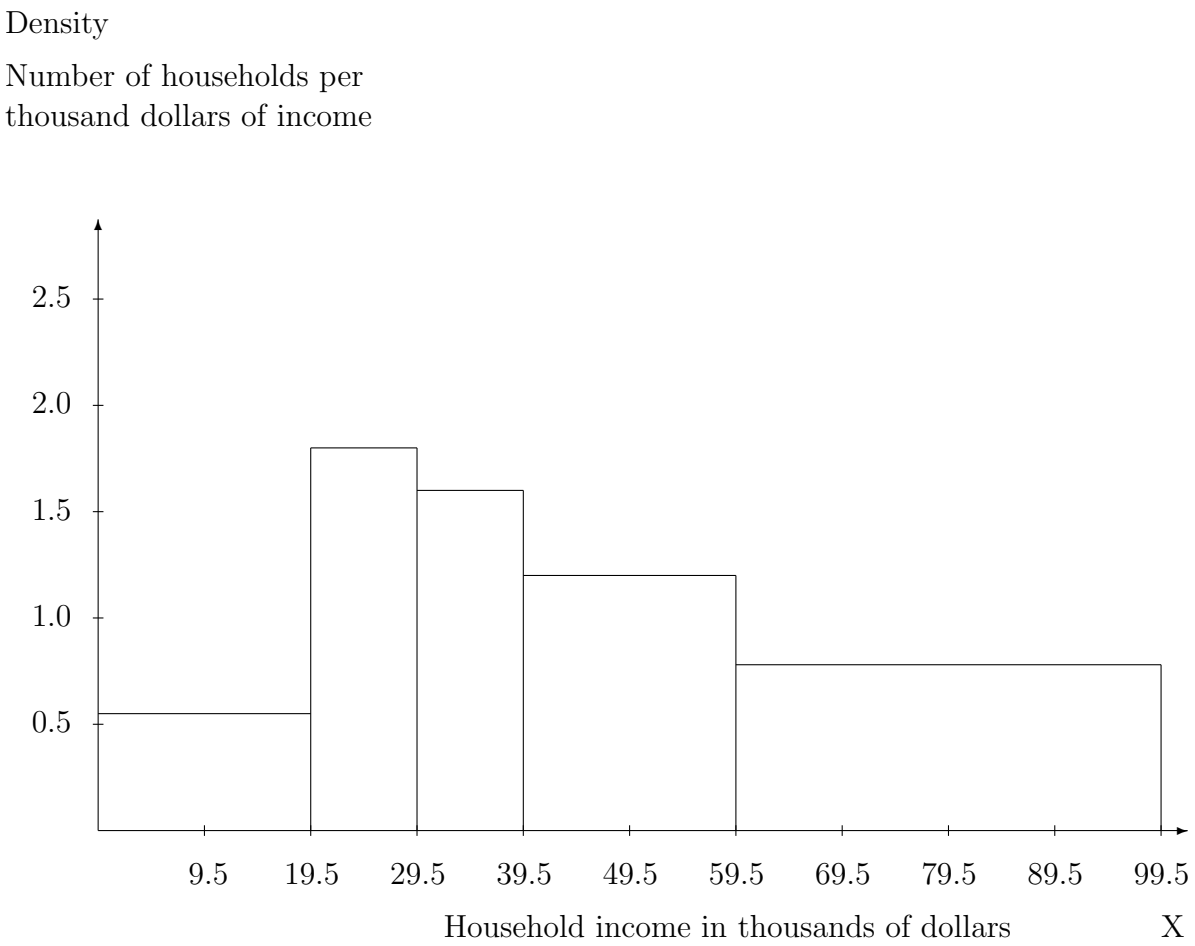

Figure 2: Histogram of household income of Saskatchewan respondents not satisfied with retirement experiences

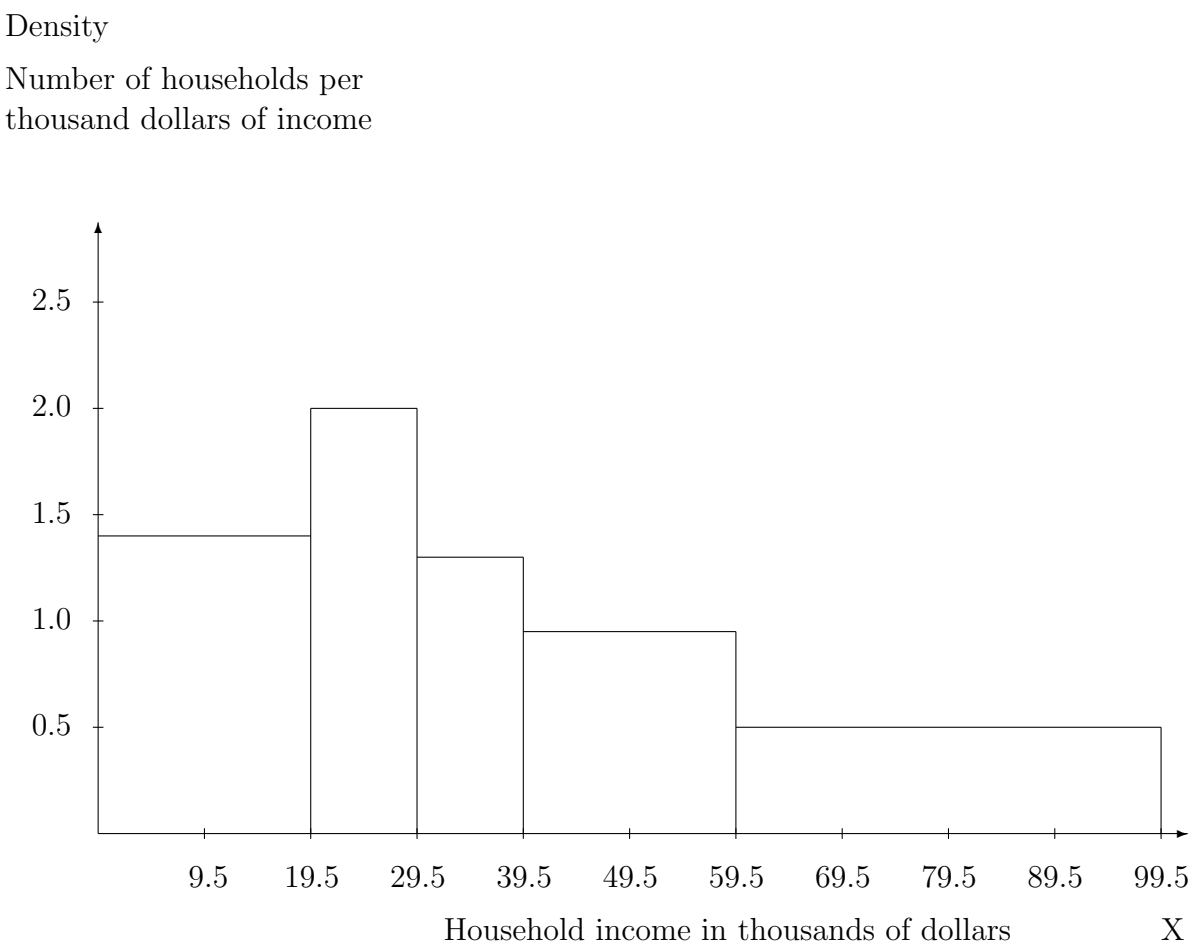

Table 2: Percentages and cumulative percentages for obtaining median income of satisfied and not satisfied Saskatchewan respondents

Real class limits for income	Interval width	Satisfied P	Satisfied Cum. P	Not satisfied P	Not satisfied Cum. P
-0.5 - 19.5	20	11	11	28	28
19.5 - 29.5	10	18	29	20	48
29.5 - 39.5	10	16	45	13	61
39.5 - 59.5	20	24	69	19	80
59.5 - 99.5	40	31	100	20	100
Total		100.0		100	

Median.

Using the percentage distributions in Table 2, the median income of the satisfied group is in the interval from 39.5 to 59.5, where the cumulative percentage first reaches more than fifty per cent. Interpolating in this interval, the median is

$$\begin{aligned} P_{50} &= 39.5 + \left(\frac{50-45}{69-45} \times 20\right) \\ &= 39.5 + \left(\frac{5}{24} \times 20\right) \\ &= 39.5 + (0.2083 \times 20) \\ &= 39.5 + 4.167 \\ &= 43.667 \end{aligned}$$

or an income of $43,700.

For the not satisfied group, the median is in the interval from 29.5 to 39.5. Interpolating in this interval, the median is

$$\begin{aligned} P_{50} &= 29.5 + \left(\frac{50-48}{61-48} \times 10\right) \\ &= 29.5 + \left(\frac{2}{13} \times 10\right) \\ &= 29.5 + (0.1536 \times 10) \end{aligned}$$

$$\begin{aligned} &= 29.5 + 1.538 \\ &= 31.038 \end{aligned}$$

or an income of \$31,000.

Mean. The calculations for obtaining the mean for each of the two groups are contained in Table 3.

Table 3: Calculations for obtaining the mean income of satisfied and not satified Saskatchewan respondents

	Midpoint	Satisfied		Not satisfied	
Income	X	P	PX	P	PX
0-19	9.5	11	104.5	28	266.0
20-29	24.5	18	441.0	20	490.0
30-39	34.5	16	552.0	13	448.5
40-59	49.5	24	1,188.0	19	940.5
60-99	79.5	31	2,464.5	20	1,590.0
Total		100	4,750.0	100	3,735.0

For those who were satisfied, the mean income is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{4,750.0}{100} = 47.5$$

thousand dollars or \$47,500.

For those who were not satisfied, the mean is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{3,735.0}{100} = 37.35$$

thousand dollars, or \$37,350.

(c) Examining the frequency distribution tables and the histograms, it is apparent that those respondents reporting satisfaction with retirement experiences have greater incomes than those who said they were not satisfied with retirement. For example, 48% of respondents who were not satisfied have incomes of under \$30,000

while only 29% of the satisfied were from this group. The histograms also show that the not satisfied group has a larger percentage is with incomes of \$40,000 plus ($24 + 31 = 55\%$) than do the not satisfied group ($19 + 20 = 39\%$). While the modal income for the two groups is the same, the median and mean are both larger for those respondents who are satisfied with their retirement experiences than the not satisfied group. The mean is approximately \$10,000 greater for those who are satisfied than those who are not satisfied and the difference in the medians is even larger, at over \$12,000. While there is insufficient evidence here to conclude that higher incomes lead to greater satisfaction with retirement experiences, the evidence certainly points in this direction.

2. **Undergraduate student survey**

   (a) In Table 4, the values for the two variables are listed in order, from smallest to largest.

Table 4: Undergraduate student grades and weekly hours spent at community service

Grade	Hours
4.6	3.4
4.6	3.8
4.6	4.2
4.7	4.3
4.9	4.4
5.0	4.5
5.2	4.7
5.2	5.4

From Table 4 the mode for grades is 4.6, with three occurrences, the most for any one value. For hours, each value occurs only once, so each value is a mode or there is no mode.

For each variable, there are eight cases, so when the values are ordered the median is the fourth and fifth cases, or the average of

these two. For grades, the median values are 4.7 and 4.9, or an average of 4.8. For hours, the median values are 4.3 and 4.4, or at 4.35.

The range of values is the largest value minus the smallest value. For grade, the range is $5.2 - 4.6 = 0.6$ and for hours the range is $5.4 - 3.4 = 2.0$.

Table 5 contains the calculations for obtaining the mean and standard deviation, using the two different formulae for obtaining the standard deviation.

Table 5: Calculations for mean and standard deviation of grades and hours at community service, undergraduate students

Grades			Community service	
X	$X - \bar{X}$	$(X - \bar{X})^2$	X	X^2
5.0	0.15	0.0225	4.5	20.25
4.9	0.05	0.0025	4.2	17.64
4.6	-0.25	0.0625	4.3	18.49
5.2	0.35	0.1225	4.4	19.36
4.6	-0.25	0.0625	3.8	14.44
4.7	-0.15	0.0225	3.4	11.56
5.2	0.35	0.1225	5.4	29.16
4.6	-0.25	0.0625	4.7	22.09
38.8	0.00	0.4800	34.7	152.99

The sum of the values of each variable is given in the last row of Table 5. For grades, the mean is a grade average of 4.85.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{38.8}{8} = 4.85$$

. For community service, the mean is 4.34 hours per week.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{34.7}{8} = 4.34$$

.

For grades the variance is

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{0.4800}{7} = 0.06857$$

and the standard deviation is a grade point average of 0.262.

$$s = \sqrt{s^2} = \sqrt{0.06857} = 0.262$$

For community service, the variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right) \\ &= \frac{1}{7}\left(152.99 - \frac{34.7^2}{8}\right) \\ &= \frac{152.99 - 150.51}{7} \\ &= \frac{2.48}{7} \\ &= 0.354 \end{aligned}$$

and the standard deviation is 0.595 hours per week.

$$s = \sqrt{s^2} = \sqrt{0.354} = 0.595$$

(b) A summary of the statistics for the two variables is contained in Table 6.

Table 6: Summary of statistics for undergraduate student variables

	Variable	
Statistic	Grades	Hours
Mode	4.6	—
Median	4.8	4.35
Mean	4.85	4.34
Range	0.6	2.0
Std. dev.	0.26	0.60

For these data the median and mean are worth reporting in that they show the differences between the two groups. Each give a similar picture, that the average of grades is around 4.8 and the average hours of community work is around 4.3 hours per week. These measures appear to summarize the distributions in a reasonable manner, given that each is close to the centre of the set of values.

In this example, the mode does not appear to be useful as a measure of centrality. For grades, the mode happens to be at the lower end of the grade distribution. For community work, each value occurs only once, so that there is no mode, or each value is a mode. So for community work, the mode is hardly worth reporting.

3. **Education of adults**.

(a) For each of the two years, the mode is "less than high school education." This is the category that contains the largest percentage of adults for each of 1991 and 2001.

(b) In order to obtain the median and percentiles, it is useful to construct a table of percentages and cumulative percentages from the diagrams in the Problem Set. This is done in Table 7. Given that the percentages in the diagram sum to ony 99% for 1991, I have altered the "Less than high school" percentage for 1991 to 38% (from 37%), in order to produce a sum of 100%.

If education is an ordinal variable, ranked from less than high school as the lowest value of education, to university being the highest value of education, the median for each year is the "High school graduate" category. From Table 7, this is where the cumulative percentage distribution first reaches fifty per cent or more in each case.

The thirtieth percentile for 1991 is less than high school since more than thirty per cent (38%) of adults are at this value. For 2001, the lowest category does not contain thirty per cent and the thirty per cent point is reached only in the next category. As a result, the thirtieth percentile for 2001 is at the "High school graduate" level.

Table 7: Percentages and cumulative percentages of education of adults, 1991 and 2001

	1991		2001	
Education	P	Cum. P	P	Cum. P
Less than high	38	38	29	29
High school grad	23	61	23	52
Trades	12	73	12	64
College	12	85	16	80
University	15	100	20	100
Total	100.0		100	

The seventieth percentile for 1991 occurs at the "Trades" category. At the "High school graduate" level, there are only sixty per cent of adults and the seventy per cent point is not reached until the "Trades" category is reached. By 2001 though, there are only 64% of adults with "Trades" or less so the seventieth percentile is not reached until the "College" category.

(c) The statistics of parts a. and b. are summarized in Table 8.

Table 8: Summary of statistics for education of adults

Statistic	1991	2001
Mode	Less than high	Less than high
Median	High school	High school
P_{30}	Less than high	High school
P_{70}	Trades	College

In this example, the mode and median do not illustrate the change in levels of education. Each of the mode and median are identical for the two years. But the distributions in Figure 1 of the Problem Set clearly illustrate that in 2001 there are fewer adults at the lowest levels of education, as compared with the situation in 1991. The percentiles provide an indication of this increase in education

levels of adults. Each of P_{30} and P_{70} shows that the education level is greater for 2001 than 1991.

4. **Sense of belonging**. The mode for both age groups is category 2, "somewhat stong" since this level of sense of belonging includes more respondents than any other category.

   In order to obtain the median, the frequency distributions in Problem Set 1 have been converted into percentages, and the associated comulative percentages, in Table 9. From this table, the median is value 2, somewhat strong for each of the two age greoups. For each age group, there are less than fifty per cent of the cases at level 1 of sense of belonging. But in each case, when all those at sense of belonging levels 1 and 2 have been accounted for, there are more than fifty per cent. The fifty per cent point of cumulative percentage distribution, the median, is thus at level 2 of sense of belonging for each age group.

Table 9: Frequency, percentage, and cumulative percentage distributions for sense of belonging, Saskatchewan respondents aged 15-24 and 55-64

	15-24			55-64		
	f	P	Cum. P	f	P	Cum. P
1 – very strong	36	20.3	20.3	73	43.5	43.5
2 – somewhat strong	83	46.9	67.2	78	46.4	89.9
3 – somewhat weak	44	24.9	92.1	15	8.9	98.8
4 – very weak	14	7.9	100.0	2	1.2	100.0
Total	177	100.0		168	100.0	

For 15-24 year olds, the mean value of sense of belonging is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{390}{177} = 2.203$$

or, rounded to the nearest tenth of a point, 2.2. For those aged 55-64, the mean is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{282}{168} = 1.679$$

Table 10: Calculations for mean sense of belonging for 15-24 and 55-64 year olds

Midpoint	15-24		55-64	
X	f	fX	f	fX
1	36	36	73	73
2	83	166	78	156
3	44	132	15	45
4	14	56	2	8
Total	177	390	168	282

or, rounded to the nearest tenth of a point, 1.7.

The data from the tables and summary statistics indicate that younger respondents express a lower sense of belonging to Saskatchewan than do older respondents. There are similar numbers of respondents in each of the two age groups but the older group has more respondents stating that they have a very strong sense of belonging as compared with the younger age group. While the mode and median sense of belonging is the same for each of the two age groups, there is a clear difference in the two means. The variable "sense of belonging" has only an ordinal level of measurement, so differences in means should be considerable before concluding the two groups differ. However, the mean sense of belonging for the younger group is 2.2 while it it 1.7 for the older age group, a difference of 0.5 units on a 4-point scale. The mean is thus more useful in demonstrating the differences between the two age groups than are the mode or median.

5. **Averages**.

   By using the word "most," the first quote most likely implies the **mode**. That is, the quote states that there are more Canadians who could lose some fat than those who do not need to lose some fat. The mean and median could also be at a value of fat that is too great, but the evidence presented in the quote does not indicate this. Further, the quote indicates that three-quarters have excess fat, and this implies

the mode.

If the variable is engaging in religious activities on one's own, the quote points to the **median**. There is insufficient data to provide any measure of centrality on the variable "attending religious services." However, the quote actually provides a summary of the distribution of how Canadians engage in religious activities of their own, by providing information about how often Canadians do this. It is reported that over half engage in such activities at least once a month, so the indication is that the median is once a month. Given that over one-half engage in religious activities on their own a least montly, this value of "at least monthly" is also the **mode** of the variable "engage in religious activities of their own."

While this could refer to either the mode, median, or mean age at marriage, given that age has an interval and ratio scale of measurement, it is likely that these averages have been computed as the **mean** age at first marriage.

Last edited October 11, 2006.