

Social Studies 201
Winter 2005
Answers to Problem Set No. 2
January 31, 2005

1. “Regina near average”

- (a) In Table 1, the values for the three variables are listed in order, from smallest to largest, with sums for each column in the last row of the table.

Table 1: Municipal ecological footprints

	Distance	Density	Energy use
	4.5	44	3.73
	4.8	57	3.95
	5.4	100	4.05
	6.0	162	4.09
	7.6	187	4.11
	7.6	195	4.18
	7.7	301	4.21
	8.6	690	5.85
	9.2	793	6.03
Total	61.4	2,529	40.20

The median can be obtained from the ordered values in Table 1 as follows. The number of cases is 9, so the middle value is the fifth case, counting from smallest to largest. For commuting, the median is 7.6 km.; for density, the median is 187 people per square kilometre; and for energy use, the median is 4.11 global hectares. The sum of the values of each variable is given in the last row of Table 1, in order to calculate the mean. For the three variables the means are

$$\bar{X} = \frac{\Sigma X}{n} = \frac{61.4}{9} = 6.8$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{2,529}{9} = 281.0$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{40.2}{9} = 4.47$$

For commuting, the range is $9.7 - 4.5 = 5.2$ km.; for density, the range is $793 - 44 = 749$ people per square kilometre; and for energy use, the range is $6.03 - 3.73 = 2.30$ global hectares.

These are summarized in Table 7, along with values of Regina, to help answer part (b).

Table 2: Summary statistics of income of variables and values for Regina

Statistic	Commuting	Density	Energy use
Range	4.7	749	2.30
Mean	6.8	281.0	4.47
Median	7.6	187.0	4.11
Regina	4.5	57	4.18

- (b) Regina is close to average, as measured by the median, for energy use, but not so close to the average for the other variables. As a result, it is difficult to know which concept of average is being used. Given that Regina is among the smaller cities on the list, it seems more likely that the median is the average implied, that is, Regina is somewhere in the middle, between those cities with the highest and lowest values. Larger cities may have more extreme values for some of the variables, thus changing the mean a lot, so Regina may not be close to the mean. It is also possible that the mode is what is meant – Regina may be similar in value to several other cities. In summary, there seems to be no definitive answer and it may be that by “the middle,” the author of the article is referring to something like the median.

2. **Employed and never worked in labour force.** In order to graph this as accurately as possible and obtain the proper interval widths and densities, the real class limits are used for each interval. Table 3

contains this information. The histograms obtained from the frequency distributions and densities are contained in Figures 1 and 2.

Table 3: Frequency distributions and densities for employed individuals and those who never worked in labour force

Age	Real class	Interval	Employed		Never worked	
	limits	Width (w)	f	Density	f	Density
15-19	14.5-19.5	5	2	0.40	13	2.60
20-29	19.5-29.5	10	7	0.70	7	0.70
30-34	29.5-34.5	5	6	1.20	1	0.20
35-39	34.5-39.5	5	13	2.60	2	0.40
40-49	39.5-49.5	10	9	0.90	1	0.10
50-69	49.5-59.5	10	6	0.60	0	0.00
60-69	59.5-69.5	10	6	0.60	7	0.70
70-85	69.5-85.5	16	1	0.06	9	0.56
Total			50		40	

Modes. The mode for each distribution is at the peak of the histogram, or the interval with the greatest density of occurrence. For the employed, the mode is at ages 35-39, or 34.5 to 39.5, where the density is 2.60, by far the highest density for this distribution. The mode is thus ages 35 to 39, or the midpoint of this interval, at age 37.

For those individuals who have never worked in the labour force the greatest density is 2.60, at ages 15 to 19, or 14.5 to 19.5, or the midpoint of this interval, age 17.

Mean. For those who were employed, the mean age is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{2,073.5}{50} = 41.47$$

or, rounded to the nearest tenth of a year, 41.5 years of age.

For those who never worked in the labour force, the mean is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{1,072.5}{40} = 26.81$$

Figure 1: Histogram of age of individuals who were employed

Density

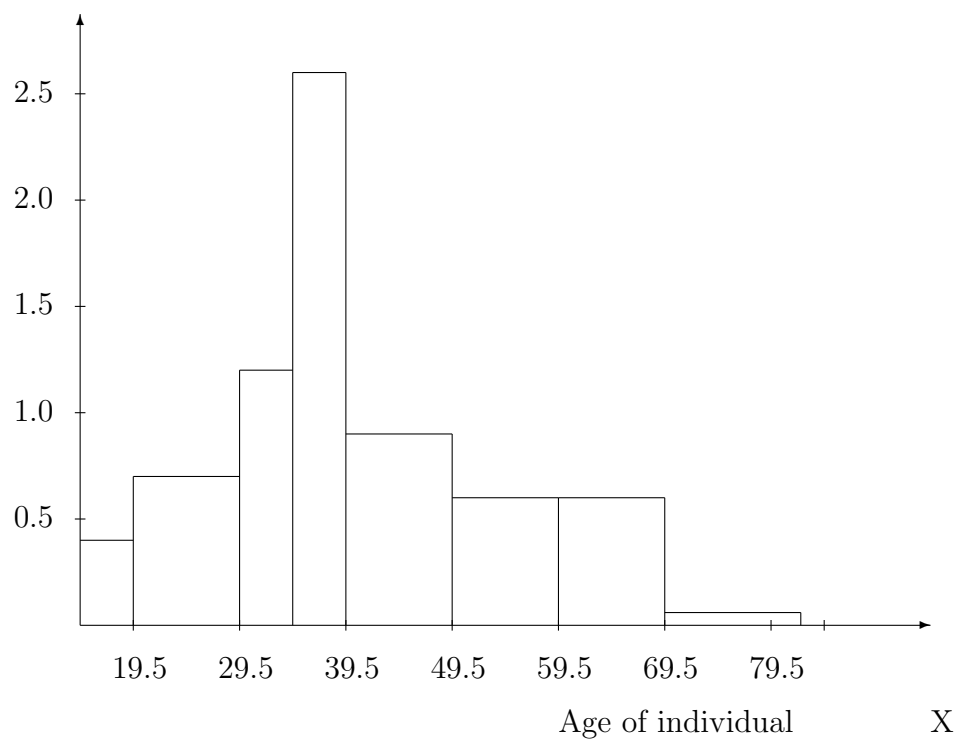
Number of individuals
per year of age

Figure 2: Histogram of age of individuals who never worked in the labour force

Density

Number of individuals
per year of age

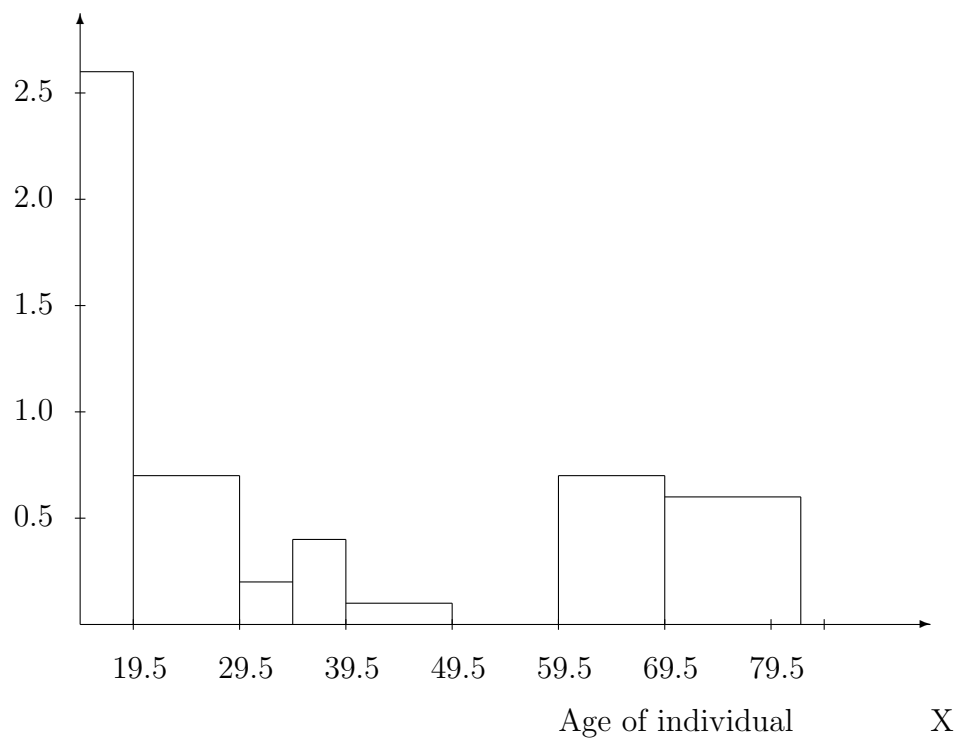


Table 4: Calculations for mean age of those who were employed and who never worked in the labour force

Age in years	Midpoint X	Low income		High income	
		f	fX	f	fX
15-19	17	2	34.0	13	221.0
20-29	24.5	7	171.5	7	171.5
30-34	32.5	6	195.0	1	32.5
35-39	37	13	481.0	2	74.0
40-49	44.5	9	400.5	1	44.5
50-59	54.5	6	327.0	0	0.0
60-69	64.5	6	387.0	7	451.5
70-85	77.5	1	77.5	9	77.5
Total		50	2,073.5	40	1,072.5

or, rounded to the nearest tenth of a year, 26.8 years of age.

3. Alcohol consumption by income

- (a) The calculations for obtaining the mean for each income group are contained in Table 5.

For those with low income, the mean number of alcoholic drinks consumed per week is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{2,861}{762} = 3.755$$

or, rounded to the nearest tenth of a drink, 3.8 drinks per week.
For those with high income, the mean is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{3,133.0}{582} = 5.383$$

or, rounded to the nearest tenth of a drink, 5.4 drinks per week.
The percentages for obtaining the medians and seventieth percentiles are contained in Table 6. Since there is a gap between the

Table 5: Calculations for mean alcohol consumption of low and high income Saskatchewan respondents

Number of drinks	Midpoint X	Low income		High income	
		f	fX	f	fX
None	0	370	0	188	0.0
1-4	2.5	214	535	185	462.5
5-9	7.0	94	658	106	742.0
10-19	14.5	54	783	74	1,073.0
20-39	29.5	30	885	29	855.5
Total		762	2,861	582	3,133.0

end points of the apparent class limits, the real class limits are used to obtain more accurate estimates of the percentiles.

Using the percentage distributions in Table 6, the median consumption for those of low income is in the interval from 1 to 4 or 0.5 to 4.5 drinks per week, where the cumulative percentage first reaches more than fifty per cent. Interpolating in this interval, the median is

$$\begin{aligned}
 P_{50} &= 0.5 + \left(\frac{50 - 48.6}{28.1} \times 4 \right) \\
 &= 0.5 + \left(\frac{1.4}{28.1} \times 4 \right) \\
 &= 0.5 + (0.0498 \times 4) \\
 &= 0.5 + 0.199 \\
 &= 0.699
 \end{aligned}$$

or 0.7 drinks per week.

For the high income individuals, the median is also in the same interval. Interpolating in this interval gives

$$\begin{aligned}
 P_{50} &= 0.5 + \left(\frac{50 - 32.3}{31.8} \times 4 \right) \\
 &= 0.5 + \left(\frac{17.7}{31.8} \times 4 \right)
 \end{aligned}$$

Table 6: Percentages for obtaining percentiles of alcohol consumption, low and high income Saskatchewan respondents

Alcohol consumption	Interval width	Low income		High income	
		<i>P</i>	Cum. <i>P</i>	<i>P</i>	Cum. <i>P</i>
-0.5 - 0.5	1	48.6	48.6	32.3	32.3
0.5 - 4.5	4	28.1	76.7	31.8	64.1
4.5 - 9.5	5	12.3	89.0	18.2	82.3
9.5 - 19.5	10	7.1	96.1	12.7	95.0
19.5 - 39.5	20	3.9	100.0	5.0	100.0
Total		100.0		100.0	

$$= 0.5 + (0.5566 \times 4)$$

$$= 0.5 + 2.226$$

$$= 2.726$$

or 2.7 drinks per week.

For low income individuals, the seventieth percentile is in the interval 1-4, or from 0.5 to 4.5 drinks, since there are only 48.6% of cases with zero drinks per week and another 28.1% of cases in this interval, making a cumulative total of 76.7% once all cases in this interval are included. Interpolating in this interval, the seventieth percentile is

$$\begin{aligned}
 P_{70} &= 0.5 + \left(\frac{70 - 48.6}{28.1} \times 4 \right) \\
 &= 0.5 + \left(\frac{21.4}{28.1} \times 4 \right) \\
 &= 0.5 + (0.7616 \times 4) \\
 &= 0.5 + 3.046 \\
 &= 3.546
 \end{aligned}$$

or 3.5 drinks per week.

For the high income individuals, the median is in the interval from 5 to 9, or 4.5 to 9.5, where the seventy per cent point of the

cumulative percentage distribution is first reached. Interpolating in this interval gives

$$\begin{aligned}
 P_{70} &= 4.5 + \left(\frac{70 - 64.1}{18.2} \times 5 \right) \\
 &= 4.5 + \left(\frac{5.9}{18.2} \times 5 \right) \\
 &= 4.5 + (0.3242 \times 5) \\
 &= 4.5 + 1.621 \\
 &= 6.121
 \end{aligned}$$

or 6.1 drinks per week.

Table 7: Summary statistics of alcohol consumption, low and high income

Statistic	Low income	High income
Mean	3.8	5.4
Median	0.7	2.7
P_{70}	3.5	6.1

- (b) From the statistics obtained in part (a), higher income respondents generally consume more alcohol than those with lower income. Each of the mean and median is approximately 2 drinks per week greater for high than for low income individuals. In the case of the seventieth percentile, the difference is even greater, so that only thirty per cent of low income respondents consume more than 3.5 drinks per week. But in the case of those with higher income, the seventy per cent point is not reached until just over six drinks per week. There is sometimes a perception that lower income people consume more alcohol but surveys consistently find that it is higher income people who consume more alcohol.

4. Education of parents of high school graduates and dropouts

- (a) Since it is difficult to determine an exact interval width associated with categories of income, these are treated here as ordinal, discrete categories. The mode thus occurs in the category with the

greatest percentage of cases. For both groups of parents, those who are parents of dropouts and those who are parents of graduates, the most common category is completed high school (45.2% of parents of dropouts and 34.7% of parents of graduates). Since this is the mode for each of these two groups of parents, the overall mode is also this category, completed high school.

- (b) The calculations for obtaining the mean of each group are in Table 8.

Table 8: Calculations for mean educational level of parents of high school dropouts and graduates

Years of Educ. (X)	Dropouts		Graduates	
	P	PX	P	PX
10	26.9	269.0	8.7	87.0
12	45.2	542.4	34.7	416.4
14	16.9	236.6	26.0	364.0
16	11.0	176.0	30.6	489.6
Total	100.0	1,224.0	100.0	1,357.0

For parents of high school dropouts, the mean educational level is

$$\bar{X} = \frac{\Sigma(PX)}{100} = \frac{1,224.0}{100} = 12.240$$

or, rounded to the nearest tenth of a year, 12.2 years of education. For parents of high school graduates, the mean educational level is

$$\bar{X} = \frac{\Sigma(PX)}{100} = \frac{1,357.0}{100} = 13.570$$

or, rounded to the nearest tenth of a year, 13.6 years of education.

- (c) Parents of graduates generally have higher education levels than parents of dropouts. Proceed from the lowest through the highest levels of education. For the two lowest education levels, the percentage of parents of dropouts exceeds the percentage of parents of graduates. In contrast, for the two highest education levels, the

percentage of parents of dropouts is less than the percentage of parents of graduates. The difference for the two groups is not apparent in the mode, since the two groups have the same mode. But the difference is apparent in the respective means. Using the coding suggested for (b), the parents of graduates have a mean education level approximately 1.5 years greater than the parents of dropouts.

5. **Averages.** Which of the three concepts of average are implied by these quotes is not entirely clear. One explanation is as follows.

- (a) Since any one MRT is equivalent to any other in terms of number of MRTs, the scale of measurement is an interval level scale. The rate of 49 per 100,000 is a mean – the sum of all the MRTs, divided by the population. From the limited information available in the quote, this appears to be the Canadian mean, and the Saskatchewan rate is below this average.
- (b) Exactly what is meant by average in this case is less clear. Since the reference is to mental agility, if this is measured using an interval level scale, then this might refer to the mean. But it could imply the mode or median. If all that is being claimed is that most women who drink a beer or glass of wine have a greater mental agility than those who do not drink alcoholic beverages, this could imply the mode. Or this could represent the median, if the middle value for the two groups is being compared. There does not appear to be sufficient information in this quote to make a definitive conclusion as to which average this might be.