Social Studies 201
Fall 2003
Answers to Problem Set No. 2
September 30, 2003

1. The variable X is "number of times per year felt rushed" and there are samples from two groups, respondents under age 45 and respondents age 45 and over. Since the data are categorized into intervals of different width (1-10, 10-20, 20-80, etc.), in order to construct the histogram for each group it is necessary to calculate the densities. This is done in Table 1

Table 1: Frequency distributions and densities for number of times felt rushed, respondents under age 45 and respondents age 45 plus

Number of times felt rushed (X)	Interval Width	Under age 45		Age 45 plus	
		Number	Density	Number	Density
0	1	2	2.00	69	69.0
1 - 10	9.5	8	0.80	54	5.70
10 - 20	10	15	1.50	23	2.30
20 - 80	60	39	0.65	40	0.67
80 - 300	220	118	0.54	61	0.28
300 - 365	65	173	2.66	105	1.62

The histograms are in Figures 1 and 2. From the histograms or the densities of Table 1, the mode of number of times per year felt rushed for those under age 45 is 300-365, or the midpoint of this, 332.5, or in words, "every day." For those aged 45 and over, the mode is especially distinctive at zero or never. For the category of $X = 0$, the bar is so high it has to be drawn extending off the top of the histogram.

Figure 1: Respondents under age 45

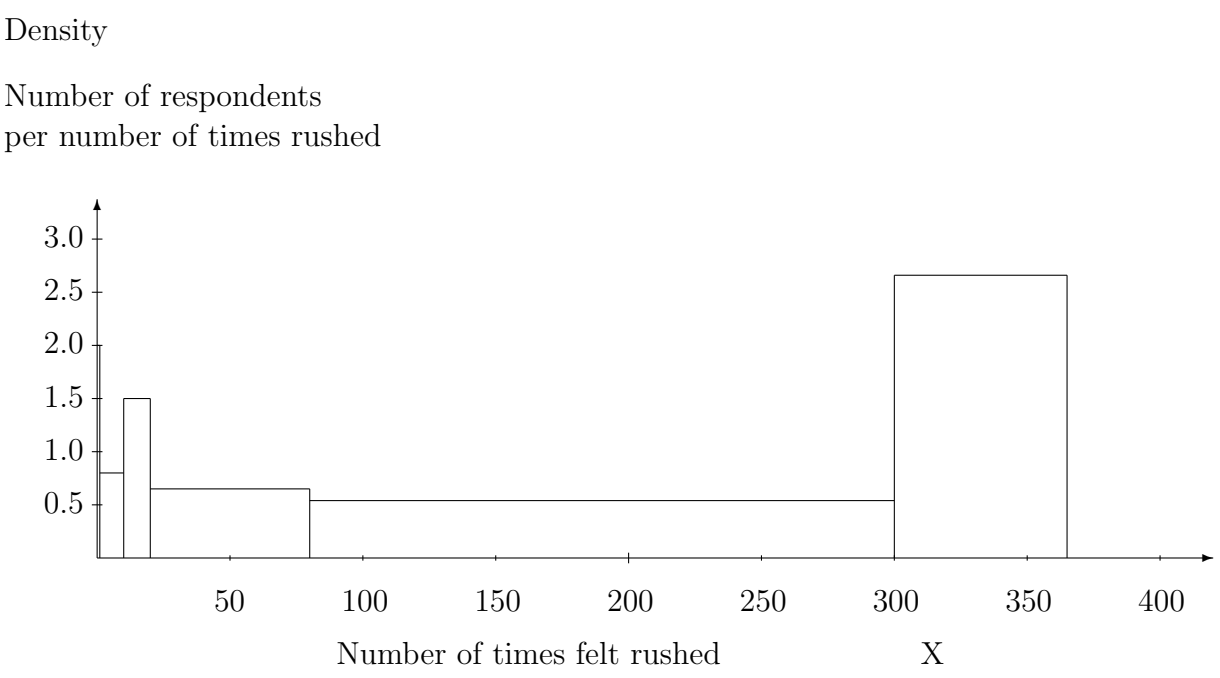

Figure 2: Respondents age 45 plus

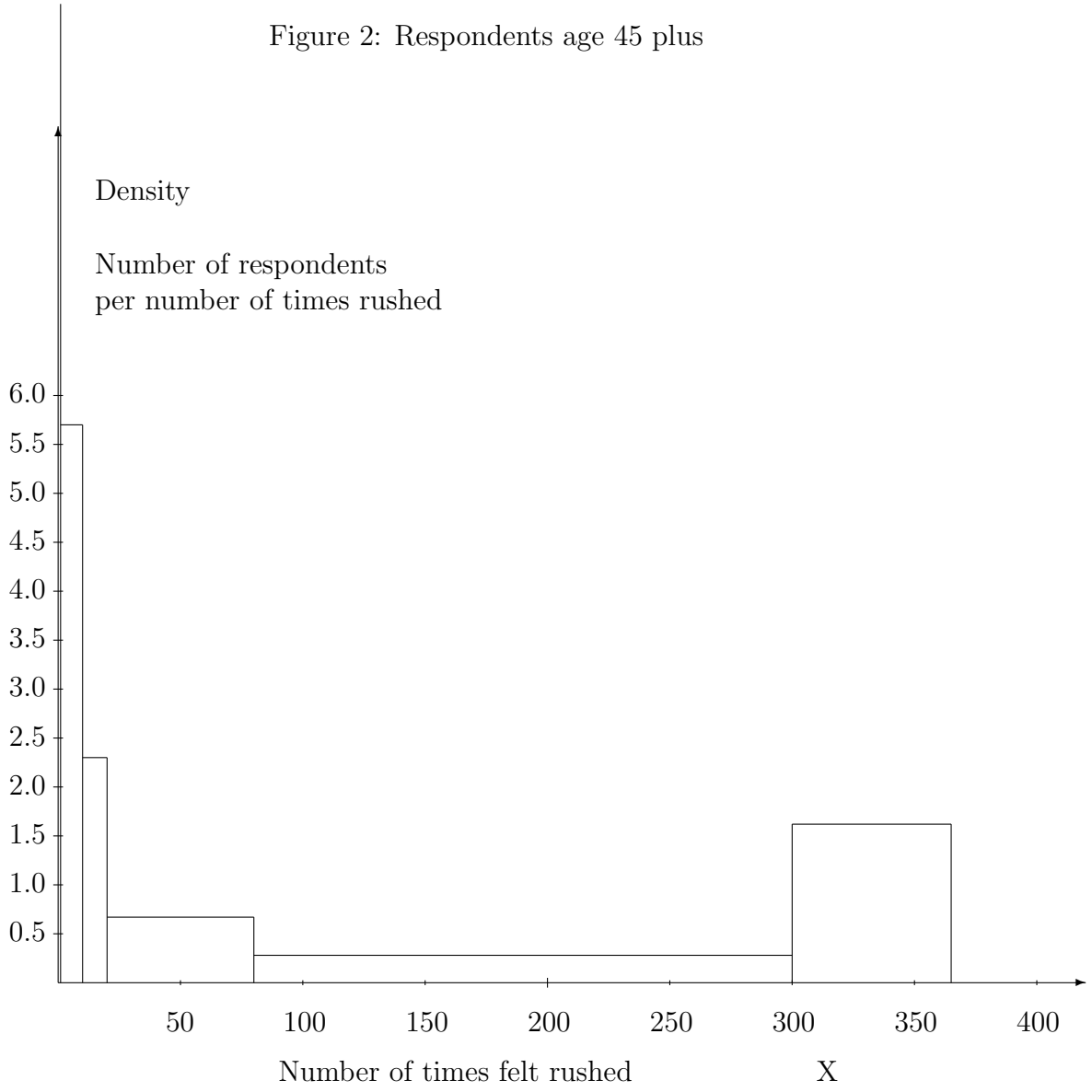

Table 2: Calculations for mean for number of times felt rushed (X), respondents under age 45 and respondents age 45 plus

Number of times per year felt rushed	Midpoint X	Under age 45 f	Under age 45 fX	Age 45 plus f	Age 45 plus fX
0	0	2	0.0	69	0.0
1 - 10	5	8	40.0	54	270.0
10 - 20	15	15	225.0	23	345.0
20 - 80	50	39	1950.0	40	2,000.0
80 - 300	190	118	22,420.0	61	11,590.0
300 - 365	332.5	173	57,522.5	105	34,912.5
Total		355	82,157.5	352	49,117.5

The calculations for obtaining the means are in Table 2 and the means for the two groups are as follows. For those under age 45, the mean number of times rushed is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{82,157.5}{355} = 231.4$$

and for those ages 45 and over, the mean number of times rushed is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{49,117.5}{352} = 139.5$$

. These might be rounded to 231 times per year rushed for those under age 45 and 140 times per year rushed for those aged 45 and over.

For the under age 45 group, the median is the $355/2 = 177.5$ or the 178th case. Beginning from the smallest value of X, ie. 0, and moving toward larger values, there are $2 + 8 + 15 + 39 = 64$ cases by the time all the "once a week" are accounted for. In the next interval, "a few times per week," there are 118 cases, so once all these are accounted for, this means $64 + 118 = 182$ cases, so the 178th case is one of these. As a result, the median is in the category "a few times a week."

For the age 45 and over group, the median value is the $352/2 = 176$th or 177th case. Adding up values from the lowest value of X and proceeding until the 176 or 177th case is encountered yields a median of "once a week." That is, there are $69+54+23 = 146$cases up to "once a month" and another 40 cases at "once a week" so the median is in the latter category, since this contains the 176th and 177th cases.

Similarities and differences. Both distributions range over the values from 0 to 365 and both distributions have relative peaks at each end of the distribution, with less frequent occurrence between the first and last categories. But the absolute peak for the age 45 plus is at 0, while it is at "every day" for the under age 45 group. The peak for the older age group is especially notable, with zero being the category with by far the largest density. Even the category 1-10 or "less than once a month" has density or relative frequency much greater than any of the densities for the under age 45 group. As a result, each of the mode, median, and mean for the under age 45 group is greater for the under age 45 group than for the 45 and older group. These distributions show that while there are considerably numbers in each group that feel rushed and others who do not, the younger age group reports feeling rushed much more than does the older group.

2. (a) The report that average student debt is below the national average most likely refers the the **mean** debt of students. The figure of $12,000 is quoted, so the variable "student debt" is being measured in dollars, an interval and ratio scale. As a result, the mean debt in dollars can be meaningfully calculated and the analyst most likely used the mean.

   (b) Services in First Nations communities are reported as less than what Canadians enjoy as an average. While it is possible that this could imply the mean or median, this most likely refers to the **mode** or the most common level of services. There are many types of services, with no indication as to which is reported here. But some might be availability of running water and sewer, garbage collection, policing, public libraries, etc. Many of these refer to whether or not the service exists or not, although some might be measured at an ordinal or interval or ratio level. Since these are

merely categorizations, it seems likely that it is the mode that is implied here.

(c) If Canada's teenage pregnancy rate is middling among those of industrial nations, this implies the **median** level of pregnancy rate. In order to report a country as middling, there must be some ranking, so the countries have probably been ranked from lowest to highest rates and Canada is somewhere in the middle. It could be that the mean is implied, but the use of a descriptive category like "middling" implies that the comparative rates have not necessarily been measured on an interval or ratio scale, but are more likely ranked.

3. (a) i. Ranked in order of number of homicides, the values are 1, 1, 2, 3, 4, 8, 9, 27, 34, 70, 85, 140, and 170. Since there are 13 values, there is a single middle value, that of the 7th case. This is the value of 9 homicides, with 6 cases less than this and 6 cases greater than 9 homicides. The province with 9 homicides is Nova Scotia.

ii. . For the homicide rate, the values in order are 0.19, 0.95, 1.06, 1.43, 1.44, 1.89, 2.08, 2.28, 2.66, 2.96, 3.35, 9.79, and 10.65. Again the 7th case is the median and this is 2.08, or British Columbia.

The provinces and territories have quite different population sizes, so those with small populations would be expected to have fewer homicides. In terms of the homicide rate, this measures the likelihood of homicide in each region, and ranking on this is not related to population, but to the likelihood of homicide. This seems greater in Western Canada and the territories, and one of these (B.C.) turns out to be have the median value – note that the other western provinces and the territories all have greater rates and the provinces to the east of Manitoba have a lower rate.

(b) i. The sum of the number of homicides (X) is $\Sigma X = 554$ and there are $n = 13$ cases, so the mean number of homicides per province or territory is

$$\bar{X} = \frac{\Sigma X}{n} = \frac{554}{13} = 42.6$$

homicides.

ii. For the homicide rate, the sum of the rates for the thirteen provinces and territories is 40.73. The mean homicide rate is

$$\bar{X} = \frac{\Sigma X}{n} = \frac{40.73}{13} = 3.13$$

. This is considerably larger than the homicide rate of 1.78 for Canada as a whole.

The reason for the difference between 3.13 and 1.78 is that each province and territory has a different population – thus they weight more or less heavily in the Canada-wide average of 1.78. In this case, each of the three territories has a high homicide rate but small population. So their high homicide experience should not weight heavily in constructing the Canada wide total. Yet in the calculation above, each territory is included equally with the more populous provinces. As a result, the high homicide rate of each territory is calculated into the mean on a par with provinces in central and Eastern Canada with much larger population, with the result that the mean is raised by including these regions with small populations but high homicide rates. The mean of 3.13 is thus an artificial number – it is the average of the 13 numbers in Table 2 of Problem Set 2, but it does not represent the average Canadian experience with homicide.

4. (a) The calculations for obtaining the mean are contained in Table 3. Using these calculations, for Saskatchewan, the mean Blishen score is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{3,756.9}{100.0} = 37.569$$

and for Ontario the mean Blishen score is

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{4,428.5}{100.0} = 44.285$$

. It might be best to round these to 37.6 for Saskatchewan and 44.3 for Ontario.

Table 3: Calculations for mean Blishen score

Number of Blishen score	Midpoint (X)	Saskatchewan P	Saskatchewan PX	Ontario P	Ontario PX
15-25	20	12.9	24.4	6.6	132.0
25-35	30	35.3	1,059.0	25.9	777.0
35-45	40	18.7	748.0	24.2	968.0
45-55	50	13.4	670.0	15.9	795.0
55-65	60	14.8	888.0	19.9	1,194.0
65-85	75	4.9	367.5	7.5	562.5
Total		100.0	3,756.9	100.0	4,428.5

(b) Median

For Saskatchewan, the median, or fiftieth percentile P_{50}, and quartiles are

$$P_{50} = 35+\left(\frac{50-48.2}{18.7}\times 10\right) = 35+(0.096\times 10) = 35+0.96 = 35.96$$

$$P_{25} = 25+\left(\frac{25-12.9}{35.3}\times 10\right) = 25+(0.343\times 10) = 25+3.43 = 28.43$$

$$P_{75} = 45+\left(\frac{75-66.9}{13.4}\times 10\right) = 45+(0.604\times 10) = 45+6.04 = 51.04$$

For Ontario, the median, or fiftieth percentile P_{50}, and quartiles are

$$P_{50} = 35+\left(\frac{50-32.5}{24.2}\times 10\right) = 35+(0.723\times 10) = 35+7.23 = 42.23$$

$$P_{25} = 25+\left(\frac{25-6.6}{25.9}\times 10\right) = 25+(0.710\times 10) = 25+7.10 = 32.10$$

$$P_{75} = 55+\left(\frac{75-72.6}{19.9}\times 10\right) = 55+(0.121\times 10) = 55+1.21 = 56.21$$

Table 4: Calculations for median and quartiles of Blishen score

Number of Blishen score (X)	Interval width	Saskatchewan P	Saskatchewan Cum P	Ontario P	Ontario Cum P
15-25	10	12.9	12.9	6.6	6.6
25-35	10	35.3	48.2	25.9	32.5
35-45	10	18.7	66.9	24.2	56.7
45-55	10	13.4	80.3	15.9	72.6
55-65	10	14.8	95.1	19.9	92.5
65-85	20	4.9	100.0	7.5	100.0
Total		100.0		100.0	

For Saskatchewan, the interquartile range is

$$P_{75} - P_{25} = 51.0 - 28.3 = 22.7$$

and for Ontario, the interquartile range is

$$P_{75} - P_{25} = 56.2 - 32.1 = 24.1$$

and these values are each in units of Blishen scores or socioeconomic status.

(c) From these data, the distribution of Blishen scores is more concentrated at larger values in Ontario than in Saskatchewan. As measures by the respective values of the mean, median, and quartiles, Blishen scores are 4-6 points lower than in Ontario. At the same time, the general shape of each distribution in similar, with the distribution peaking at a low to middle values of socioeconomic status. There are relatively few really low Blishen scores, although there are more of these in Saskatchewan (almost 13%) than in Ontario (only about 7%). At the other end of the distribution, there are relatively few respondents with very high Blishen scores, although in this case there are more in Ontario (7.5% above 65) than in Saskatchewan (only 5% at highest level). The variation in the two distribution, as measured by the interquartile range, is quite similar, only slightly less for Saskatchewan than Ontario.

In summary, the middle part and the variation of the two distributions may be reasonably similar, but the Saskatchewan distribution is more concentrated as lower Blishen scores while Ontario has relatively greater numbers of larger Blishen scores.