

Social Studies 201
Winter 2001
Answers to Problem Set No. 2
February 2, 2001

1. The Graduating Students Survey 2000, conducted by the Canadian Undergraduate Survey Consortium gives data concerning the anticipated annual earnings of students who attend Canadian universities. From this survey, the frequency distributions for anticipated annual earnings for students in primarily undergraduate universities and in universities with undergraduate and graduate programs are given in Table 2.

Using the data in Table 2

- (a) Construct a histogram for the anticipated earnings of students at each of the two types of universities.
- (b) Calculate the median and interquartile range for each of the two distributions.
- (c) In words, briefly describe the differences in the two histograms.

Answer.

- (a) The intervals into which the data are grouped are of unequal width so that densities must be calculated in order to produce the histograms. The densities are calculated and shown in Table 1 and the histograms are in Figure 1.
- (b) Using the percentage and cumulative percentage distributions in Table 2, the median of anticipated annual earnings for undergrad only universities is in the interval \$25-30 thousand. Interpolating in this interval, the median is

$$P_{50} = 25 + \left(\frac{50 - 46}{62 - 46} \times 5 \right) = 25 + (0.25 \times 5) = 25 + 1.25 = 26.25$$

or 26 thousand dollars per year.

For universities with both undergraduate and graduate programs, the median is \$30,000, since it is exactly at this anticipated income that the 50% point is reached.

Figure 1: Anticipated Earnings (in Thousands of Dollars) of Students in Two Types of Universities

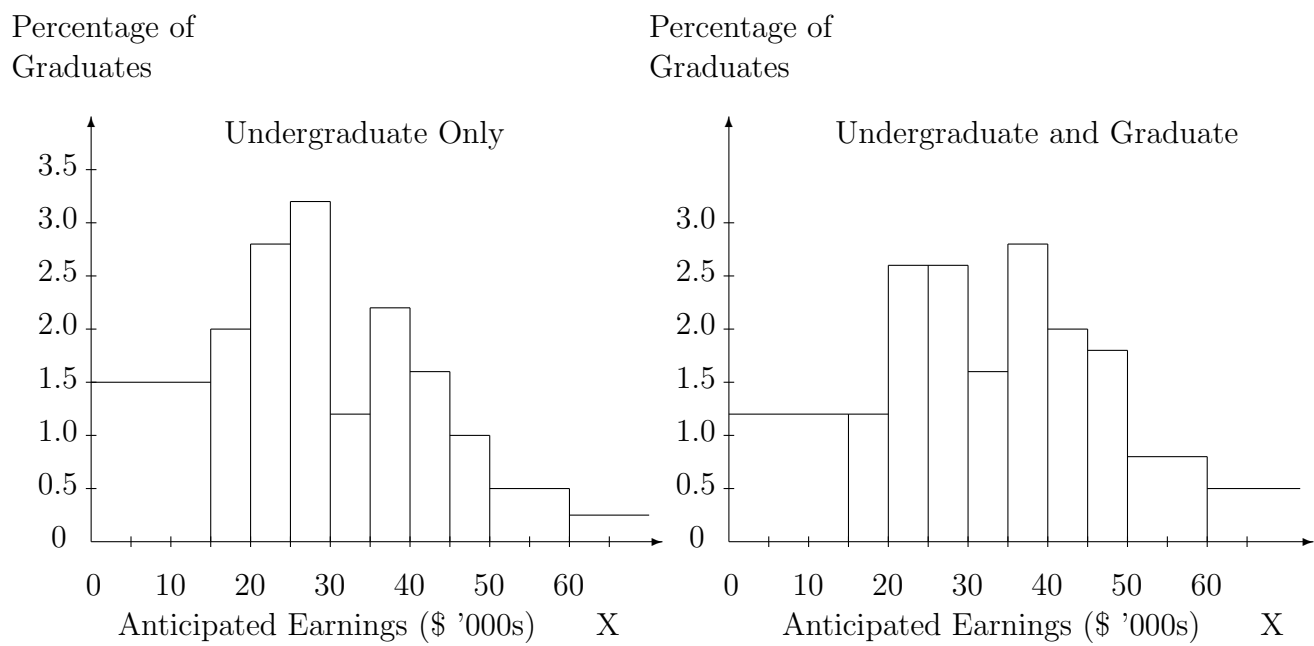


Table 1: Percentage Distributions of Annual Anticipated Earnings of Students in Two Types of Canadian Universities

Anticipated Annual Earnings in thousands of dollars	Interval Width	Per Cent in Universities with:			
		Undergrad Only %	Density	Undergrad and Grad %	Density %
Less than 15	15	22	1.5	18	1.2
15 - 20	5	10	2.0	6	1.2
20 - 25	5	14	2.8	13	2.6
25 - 30	5	16	3.2	13	2.6
30 - 35	5	6	1.2	8	1.6
35 - 40	5	11	2.2	14	2.8
40 - 45	5	8	1.6	10	2.0
45 - 50	5	5	1.0	9	1.8
50 - 60	10	5	0.5	8	0.8
60 plus	—	3	—	1	—
Total		100		100	
Sample Size		664		717	

For universities with undergraduate programs only, the 75th percentile is

$$P_{75} = 35 + \left(\frac{75 - 68}{79 - 68} \times 5 \right) = 35 + (0.636 \times 5) = 35 + 3.2 = 38.2$$

and the 25th percentile is

$$P_{25} = 15 + \left(\frac{25 - 22}{32 - 22} \times 5 \right) = 15 + (0.3 \times 5) = 15 + 1.5 = 16.5.$$

The interquartile range is $38.2 - 16.5 = 21.7$.

For universities with both undergraduate and graduate programs, the 75th percentile is

$$P_{75} = 40 + \left(\frac{75 - 72}{82 - 72} \times 5 \right) = 40 + (0.3 \times 5) = 40 + 1.5 = 41.5$$

Table 2: Percentage Distributions of Annual Anticipated Earnings of Students in Two Types of Canadian Universities

Anticipated Annual Earnings in thousands of dollars	Per Cent in Universities with:			
	Undergrad Only		Undergrad and Grad	
	%	Cum. %	%	Cum. %
Less than 15	22	22	18	18
15 - 20	10	32	6	24
20 - 25	14	46	13	37
25 - 30	16	62	13	50
30 - 35	6	68	8	58
35 - 40	11	79	14	72
40 - 45	8	87	10	82
45 - 50	5	92	9	91
50 - 60	5	97	8	99
60 plus	3	100	1	100
Total	100	100		
Sample Size	664	717		

and the 25th percentile is

$$P_{25} = 20 + \left(\frac{25 - 24}{37 - 24} \times 5 \right) = 20 + (0.077 \times 5) = 20 + 0.4 = 20.4.$$

The interquartile range is $41.5 - 20.4 = 21.1$.

- (c) The two distributions have a similar shape and similar median and interquartile range. That is, each distribution peaks in the \$20-40,000 range, with relatively few students anticipating earnings of under \$20,000 and fewer and fewer students anticipating earnings at each higher level of income above \$40,000. The interquartile range is almost exactly the same for the two distributions, just over \$21,000 in each case. This means that the middle half of each distribution is spread across an interval of just over twenty thousand dollars. The median is about \$4,000 lower for students in undergraduate only institutions, indicating that these students

may have slightly lower expectations of earnings than students in universities with combined undergraduate and graduate programs.

2. The Canadian Undergraduate Survey Consortium, in *Graduating Students Survey 2000: University of Regina*, provides the data in Table 3 and notes that “the **average** student in our sample is a single female, who is 25 years old.” In words, explain the meaning of **average**, noted in bold face in the quote and in Table 3. Note any problems you see with the statement.

Table 3: Profile of Students in Survey

Characteristic	Per Cent with Characteristic
Gender	
Male	34%
Female	66%
Age	
20 or less	1%
21	13%
22	30%
23	10%
24	10%
25 to 29	14%
30 and over	12%
Average age	25 years
Marital Status	
Married	14%
Single	79%
Other	7%

Answer. In this example, **average** appears to have different meanings for different variables. The report that the average student is a single

female must be based on the mode of marital status and sex, respectively, since these two variables have no more than a nominal scale of measurement. For marital status, there are 79% of respondents who are single, and for sex, 66% are female, so these are the averages for these two variables, in the sense that these are the most common values of these variables.

In contrast, age is an interval level scale, and the **average** reported in the table is likely to represent the mean age – it cannot be the mode, since the mode is age 22, with 30% of respondents, more than any other single value of age. Also, 25 cannot be the median age, since the median must be 23 or 24 years, given these data. In order for the mean age to be 25, the midpoint of the open-ended interval would need to be 37 years, and it is possible that this could be the case.

Note a couple of oddities in these data. First, while the average student may be a 25 year old, single female, the data do not really provide sufficient evidence to conclude this. While the mode of sex is female and the mode of marital status is single, this does not necessarily prove that the mode of the two categories together is single female. Also note that the percentages at the various ages only total to 90%. It is not clear whether some data is missing here or whether the other 10 per cent of cases did not report their age.

In summary, the average here is not a single clear cut statistic, and this example shows some of the various ways that the term **average** can be used.

3. The responses of fourteen Saskatchewan respondents from the 1996 General Social Survey, concerning their income and their attitude to life, are given in Table 4. The attitude to life question was “Would you describe yourself as being usually (1) happy and interested in life, (2) somewhat happy, (3) somewhat unhappy, or (4) unhappy with little interest?”
 - (a) For the attitude to life and income questions calculate the mode, median, mean, and standard deviation. (For income, ignore the four respondents who did not provide their income).

- (b) The fourteen respondents in Table 4 represent a sample of all Saskatchewan respondents. When all Saskatchewan respondents were considered, the results were as follows.

Compare your results in (a) with the above statistics and provide possible reasons why the statistics from the sample of size fourteen differ from the statistics for all Saskatchewan respondents.

Table 4: Responses to Income and Attitude to Life Questions, 14 Respondents, 1996 General Social Survey

Respondent Number	Income in Thousands of Dollars	Attitude to Life
1	–	1
2	7	1
3	6	2
4	48	1
5	8	3
6	14	1
7	5	1
8	13	2
9	–	4
10	18	1
11	–	1
12	–	1
13	29	1
14	8	2

Answer.

- (a) For income, the **mode is 8**, with 2 cases, more than any other single value. For attitude to life, the **mode is 1**, with nine cases, whereas 2 occurs three times, 3 once, and 4 once.

For the first sample, the ten values from smallest to largest are 5, 6, 7, 8, 8, 13, 14, 18, 29, 48 and the **median** is the 5th and

6th cases, 8 and 13, or 11.5. For the second sample the fourteen values are 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, and 4. The **median** is the 7th and 8th values, and these are both 1, so the median is 1.

Table 5: Calculations for Mean and Standard Deviation, Income and Attitude to Life

Income			Attitude to Life	
X	$X - \bar{X}$	$(X - \bar{X})^2$	X	X^2
7	-8.6	73.96	1	1
6	-9.6	92.16	1	1
48	32.4	1049.76	2	4
8	-7.6	57.76	1	1
14	-1.6	2.56	3	9
5	-10.6	112.36	1	1
13	-2.6	6.76	1	1
18	2.4	5.76	2	4
29	13.4	179.56	4	16
8	-7.6	57.76	1	1
			1	1
			1	1
			1	1
			2	4
156	0.0	1638.40	22	46

The calculations for the mean and standard deviation are given in Table 5, with the two methods of obtaining the standard deviation shown in this table. For income, the mean is $\Sigma X/n = 156/10 = 15.6$.

The variance is

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{1638.40}{9} = 182.04$$

and the standard deviation is

$$s = \sqrt{s^2} = \sqrt{182.04} = 13.49$$

or 13.5 thousand dollars.

For the responses on the attitude to life question, the mean is $\Sigma X/n = 22/14 = 1.57$ or 1.6 on the attitude scale. Using the alternative formula, the variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left(\Sigma X^2 - \frac{(\Sigma X)^2}{n} \right) \\ &= \frac{1}{13} \left(46 - \frac{22^2}{14} \right) \\ &= \frac{(46 - 34.57)}{13} = 0.879 \end{aligned}$$

and the standard deviation is

$$s = \sqrt{s^2} = \sqrt{0.879} = 0.938$$

or 0.9 on the attitude scale.

- (b) The following table compares the statistics for the sample of fourteen respondents and the whole population. The main reason for the difference in the value of the statistics is likely to be the fact that only a small sample of Saskatchewan residents is selected in the sample. A small sample is unlikely to be exactly representative of the whole population, even if it is selected randomly. A few cases in the sample may be quite unrepresentative of the whole group, and with only a few cases in the whole sample, this often means that the sample is not exactly representative.

Statistic	Value of Statistic	
	Saskatchewan	Sample
Mode of attitude	1	1
Median attitude	1	1
Mean attitude	1.2	1.6
Median income	17.5	11.5
Mean income	23.6	15.6

4. The histograms in Figure 1 are taken from data presented on page 22 of the *Second Graduat Survey* of the University of Regina. The data refer

to the weekly hours worked at a job by 1995 and 1999 graduates during an average semester while they attended university. The numbers in the bars of the histogram are the percentages of respondents represented by each bar. From the two histograms, answer the following.

- Calculate the mean and standard deviation of weekly hours worked at a job for 1995 and 1999.
- What is the median category of hours worked for each of the two years? Explain your reasoning. (You need not interpolate, but in a sentence or two explain which interval contains the median).
- In a sentence or two compare the distributions of weekly hours worked at a job for the two years.

Answer.

- Table 6 contains the data for calculating the mean and standard deviation of each group.

Table 6: Calculations for Mean and Standard Deviation of Weekly Hours Worked, University of Regina Graduates, 1995 and 1999

Past Week				Annual			
X	P	PX	PX^2	X	f	fX	fX^2
4	16	64	256	4	12	48	192
12	29	348	4,176	12	28	336	4,032
20	26	520	10,400	20	29	580	11,600
28	12	336	9,408	28	14	392	10,976
36	10	360	12,960	36	11	396	14,256
48	7	336	16,128	48	6	288	13,824
Total	100	1,964	53,328		100	2,040	54,880

For 1995,

$$\bar{X} = \frac{1,964}{100} = 19.64$$

$$\begin{aligned}
s^2 &= \frac{1}{n-1} \left(\Sigma fX^2 - \frac{(\Sigma fX)^2}{n} \right) \\
&= \frac{1}{100} \left(53,328 - \frac{1,964^2}{100} \right) \\
&= \frac{1}{100} (53,328 - 38,572.96) \\
&= \frac{1}{100} (14,755.04) \\
&= 147.55 \\
s &= \sqrt{s^2} = \sqrt{147.55} = 15.73.
\end{aligned}$$

For 1995, the mean weekly hours worked at a job is 19.6 hours, with a standard deviation of 15.7 hours.

For 1999,

$$\begin{aligned}
\bar{X} &= \frac{2,040}{100} = 20.40 \\
s^2 &= \frac{1}{n-1} \left(\Sigma fX^2 - \frac{(\Sigma fX)^2}{n} \right) \\
&= \frac{1}{100} \left(54,880 - \frac{2,040^2}{100} \right) \\
&= \frac{1}{100} (54,880 - 41,616) \\
&= \frac{1}{100} (13,264) \\
&= 132.64 \\
s &= \sqrt{s^2} = \sqrt{132.64} = 11.52.
\end{aligned}$$

For 1999, the mean weekly hours worked at a job is 20.4 hours, with a standard deviation of 11.5 hours.

- (b) The median category is the interval that contains the cumulative percentage of 50%, that is, the interval such that one-half of the cases have values less than or equal to the values in the interval and the other one-half of the cases have values greater than or equal to values in the interval.

For 1995, the cumulative percentages by interval are 16% up to 8 hours worked, 45% (29+16) up to 24 hours worked, and 71% up to 24 hours worked. The 50% point is in the third interval, from 16 to 24 hours worked.

For 1999, the cumulative percentages for the first three intervals are 12%, 40%, and 69%, so again the third interval, from 16 to 24 hours worked, contains the median.

- (c) The two distributions are quite similar in that the most common hours worked are in the range from 8 to 24 in each of the two years, with fewer students working less than eight hours than 8-16 or 16-24 hours. The mean and median are similar in the two cases, so the two distributions do not differ much in where they are centred – the data show that in 1995, students worked slightly fewer hours than they did in 1999. The main difference between the two distributions is that the 1999 distribution appears more concentrated than the 1995 distribution. The standard deviation is about four hours lower in 1999 than in 1995. Also note that in 1999, there are fewer students working less than 8 hours or more than 40 hours, as compared with 1995. That is, the 1999 distribution tends to be a little more concentrated around the middle than in 1995.

Statistic	Value of Statistic	
	1995	1999
Mean	19.6	20.4
Median	16-24	16-24
Standard Deviation	15.7	11.5

5. Use the percentage distributions in Table 2 for this question. These distributions come from the article “Are children going to religious services?” in *Canadian Social Trends*, Autumn 1999, pp. 13-16.
- (a) What is the mode, median, and 33rd percentile for each of (i) Roman Catholic, (ii) Anglican, and (iii) Baptist?
- (b) Someone suggests that the mean could be calculated by attaching the number 52 to weekly, 12 to monthly, 4 to occasionally, and 0

to not at all. Using these values, calculate the mean for Baptists. In words, explain what this mean represents.

- (c) Explain which religious affiliation is associated with the greatest attendance and which the least.

Table 7: Religious Service Attendance of Children by Religious Affiliation, Canada, 1994-1995

Religious Affiliation	Per Cent who Attend:			
	Weekly	Monthly	Occasionally	Not at All
Roman Catholic	22	18	31	29
Anglican	18	16	30	36
Baptist	61	10	12	17

Answer.

- (a) For Roman Catholic, the mode is **Occasionally**, the most commonly reported category, including 31% of all respondents. The median is also **Occasionally** since the cumulative percentages from least to most frequent attendance are 29% at not at all and 60% at occasionally, so the 50% point is crossed at occasionally. The 33rd percentile is also at **Occasionally**, since there are only 29% at not at all, and the 33% point is crossed at occasionally.

For Anglican, the mode is **not at all**, the median is **Occasionally**, and the 33rd percentile is **not at all**.

For Baptists, the mode is **weekly**, the median is also **Weekly**, and the 33rd percentile is at **monthly**.

- (b) The calculations for the mean are in Table 8. The mean is the values suggested for X approximate the number of times each year that members of each category report attendance at religious services. The mean of 34 thus represents the mean number of times annually that Baptists report attending religious services.

Table 8: Calculations for Mean Religious Attendance

X	P	PX
52	61	3,172
12	10	120
4	12	48
0	17	0
Total	100	3,340

- (c) Baptists report greatest attendance and Anglicans report the least attendance. That is, over 70% of Baptists report attending weekly or monthly, as opposed to only 34% of Anglicans and 40% of Roman Catholics. In contrast, 36% of Anglicans report attending not at all, much more than for Baptists and somewhat more than Roman Catholics.