Social Studies 201
Fall 2006
Answers to Problem Set No. 3
October 31, 2006

1. **Variation in divorce rates**.

   (a) Table 3 contains the calculations for obtaining the statistics. The method of differences from the mean is used for the Eastern provinces and the alternative formula of sums of X and sums of squares of X is used for the Western provinces.

Table 1: Calculations for mean and standard deviation of divorce rates for Eastern and Western provinces

Eastern provinces			Western provinces	
X	$X - \bar{X}$	$(X - \bar{X})^2$	X	X^2
17.1	-13.02	169.5204	37.0	1,369.00
27.3	-2.82	7.9524	30.2	912.04
28.9	-1.22	1.4884	29.0	841.00
27.6	-2.52	6.3504	40.0	1,600.00
49.7	19.58	383.3764	39.8	1,584.04
150.6	0.00	568.6880	176.0	6,306.08

While not requested in the question, the means may be useful when commenting on the results, and the means are as follows.

For the Eastern provinces, the mean is a divorce rate of 30.12.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{150.6}{5} = 30.1$$

For the Western provinces, the mean is a divorce rate of 35.2.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{176.0}{5} = 35.2$$

For the Eastern provinces the variance is

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{568.6880}{4} = 142.172$$

and the standard deviation is a divorce rate of 11.924.

$$s = \sqrt{s^2} = \sqrt{142.172} = 11.924$$

For the Western provinces, the variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right) \\ &= \frac{1}{4}\left(6,306.08 - \frac{176.0^2}{5}\right) \\ &= \frac{6,306.08 - 6,195.20}{4} \\ &= \frac{110.88}{4} \\ &= 27.72 \end{aligned}$$

and the standard deviation is a divorce rate or 5.265.

$$s = \sqrt{s^2} = \sqrt{27.72} = 5.265$$

A summary of the statistics for the three variables is contained in Table 2.

Table 2: Summary of statistics for divorce rates

	Provinces	
Statistic	Eastern	Western
Mean	30.1	35.2
Variance	142.2	27.7
Std. dev.	11.9	5.3

From the statistics in Table 2, divorce rates differ more across the five Eastern provinces than across the five Western provinces.

The standard deviation for the five Eastern provinces (11.9) is a little over double that for the five Western provinces (5.3). The means for the two groups of provinces is similar, with the Western provinces having a mean divorce rate a little greater than in the Eastern provinces. However, the actual divorce rates for the five Western provinces do not differ all that much, ranging from 29.0 in Saskatchewan to 40.0 in Alberta. In contrast, for the Eastern provinces, the rates range from a low of 17.1 in Newfoundland to a hight of 49.7 in Quebec. This larger difference among the divorce rates in the Eastern provinces produces a larger standard deviation for Eastern provinces, as compared with the Western provinces.

(b) If the Territories are added to the Western provinces as a sixth case, with a divorce rate of 27.6, the standard deviation will increase. This is because this extra case has a divorce rate that is below the divorce rates of any of the other five Western provinces. The addition of this case increases the difference among the rates, with the result that the standard deviation will inrease.

Table 3: Calculations for the standard deviation of divorce rates for the Western provinces with the Territories added

X	X^2
37.0	1,369.00
30.2	912.04
29.0	841.00
40.0	1,600.00
39.8	1,584.04
27.6	761.76
203.6	7,067.84

For the Western provinces plus the Territories, the variance is

$$s^2 = \frac{1}{n-1}\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right)$$

$$\begin{aligned} &= \frac{1}{5}\left(7,067.84 - \frac{203.6^2}{6}\right) \\ &= \frac{7,067.84 - 6,908.83}{5} \\ &= \frac{159.01}{5} \\ &= 31.80 \end{aligned}$$

and the standard deviation is a divorce rate or 5.64.

$$s = \sqrt{s^2} = \sqrt{31.80} = 5.639$$

2. **Ages of mature Saskatchewan singles**. From the distributions of Table 2 of the problem set, the calculations for the mean and standard deviation are given in Table 4.

Table 4: Calculations for mean and standard deviation of ages of mature Saskatchewan singles,

	Won't marry			Will marry		
X	f	fX	fX^2	f	fX	fX^2
32	9	288	9,216	24	768	24,576
37	16	592	21,904	15	555	20,535
42	14	588	24,696	14	588	24,696
47	12	564	26,508	12	564	26,508
52	9	468	24,336	4	208	10,816
Total	60	2,500	106,660	69	2,683	107,131

For those who say they won't marry,

$$\begin{aligned} \bar{X} &= \frac{2,500}{60} = 41.667 \\ s^2 &= \frac{1}{n-1}\left(\Sigma fX^2 - \frac{(\Sigma fX)^2}{n}\right) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{59}\left(106,660 - \frac{2,500^2}{60}\right) \\ &= \frac{1}{59}\left(106,660 - 104,166.667\right) \\ &= \frac{1}{59}\left(2,493.333\right) \\ &= 42.260 \\ s &= \sqrt{s^2} = \sqrt{42.260} = 6.501. \end{aligned}$$

The mean age for mature Saskatchewan singles who say they won't marry is 41.7 years and the standard deviation is 6.5 years. The CRV is 156.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{6.501}{41.667} \times 100 = 15.6$$

For those who say they will marry,

$$\begin{aligned} \bar{X} &= \frac{2,683}{69} = 38.884 \\ s^2 &= \frac{1}{n-1}\left(\Sigma f X^2 - \frac{(\Sigma f X)^2}{n}\right) \\ &= \frac{1}{68}\left(107,131 - \frac{2,683^2}{69}\right) \\ &= \frac{1}{68}\left(107,131 - 104,325.928\right) \\ &= \frac{1}{68}\left(2,805.072\right) \\ &= 41.251 \\ s &= \sqrt{s^2} = \sqrt{41.251} = 6.423. \end{aligned}$$

The mean age for mature Saskatchewan singles who say they will marry is 38.9 years and and the standard deviation is 6.4 years. The CRV is 165.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{6.423}{38.884} \times 100 = 16.5$$

Table 5: Summary of statistics of ages of two groups of mature Saskatchewan adults

	Group	
Statistic	Won't marry	Will marry
Mean	41.7	38.9
Variance	42.3	41.2
Std. dev.	6.5	6.4
CRV	15.6	16.5

A summary of the statistics for the two groups is contained in Table 5.

From the summary of the statistics in Table 5, the variation in ages for the two groups is very similar. The variance and standard deviation are almost identical for the two groups. The group that indicates they will marry is approximately three years younger, as measured by the mean. This makes the CRV for the 'will marry' group a little greater. However, the difference is not large – both groups have distributions of ages that are spread out across the years from age 30 to 54 in a similar manner, with the centre of the distribution of the 'won't marry' group at about three years above that for the 'will marry' group.

3. **Incomes of retired Saskatchewan respondents**. The calculations for the standard deviation and coefficient of relative variation are contained in Table 6.

   Using the calculations in Table 6 and the formulae for percentages, for satisfied respondents, the mean, variance, standard deviation, coefficient of relative variation are:

$$\begin{aligned} \bar{X} &= \frac{4,800}{100} = 48.0 \\ s^2 &= \frac{1}{100}\left(\Sigma PX^2 - \frac{(\Sigma PX)^2}{n}\right) \end{aligned}$$

Table 6: Calculations for standard deviation of household incomeyears

	Satisfied			Not satisfied		
X	P	PX	PX^2	P	PX	PX^2
10	11	110	1,100	28	280	2,800
25	18	450	11,250	20	500	12,500
35	16	560	19,600	13	455	15,925
50	24	1,200	60,000	19	950	47,500
80	31	2,480	198,400	20	1,600	128,000
	100	4,800	290,350	100	3,785	206,725

$$\begin{aligned} &= \frac{1}{100}\left(290,350 - \frac{4,800^2}{100}\right) \\ &= \frac{1}{100}\left(290,350 - 230,400\right) \\ &= \frac{1}{100}\left(59,500\right) \\ &= 599.5 \\ s &= \sqrt{s^2} = \sqrt{599.5} = 24.485 \end{aligned}$$

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{24.485}{48.0} \times 100 = 51.0$$

For those respondents who were not satisfied with their retirement experiences, the mean, variance, standard deviation, coefficient of relative variation are:

$$\begin{aligned} \bar{X} &= \frac{3,785}{100} = 37.85 \\ s^2 &= \frac{1}{100}\left(\Sigma PX^2 - \frac{(\Sigma PX)^2}{n}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{100}\left(206,725 - \frac{3,785^2}{100}\right) \\
&= \frac{1}{100}\left(206,725 - 143,262.25\right) \\
&= \frac{1}{100}\left(63,462.75\right) \\
&= 634.6275 \\
s &= \sqrt{s^2} = \sqrt{596} = 25.192
\end{aligned}
$$

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{25.192}{37.85} \times 100 = 66.557$$

Table 7: Summary of statistics of household income of two groups of Saskatchewan retirees

	Group	
Statistic	Satisfied	Not satisfied
Mean	48.0	37.8
Variance	599.5	634.6
Std. dev.	24.5	25.2
CRV	51.0	66.6

The summary statistics are given in Table 7, with each statistic rounded to the nearest tenth of a point. The clearest difference between the two groups is that those who are satisfied with their retirement experiences have higher mean income ($48,000) than those who are not satisfied ($37,800). The difference is approximately ten thousand dollars, lending support to the view that a major factor in terms of satisfaction is the income level. The variation of incomes for the two groups is remarkably similar though, with values of the standard deviations and with variances being very similar for the two groups. So both groups of retirees are of approximately equal variability in terms of income. In relative terms, given that the 'not satisfied' have lower income, this

means that incomes of those who are not satisfied are more variable that those who are satisfied. That is, relative to the mean or average income, the 'not satisfied' group is more varied as compared with the 'satisfied' group.

4. **Explanations of probability**.

   (a) This can be considered to be a **frequency** interpretation of probability, modified by the **judgment** of the meteorologist who made the estimate. That is, the method used to estimate this was likely a close observation of how often there was precipation in the past when weather conditions were of the type observed for October 16. Since weather conditions can never be exactly repeated under uniform conditions though, it is likely that there is some subjective judgment involved, although this is presumably an expert judgment, that of a meteorologist.

   (b) This is a **frequency** or **empirical** use of probability. Statements such as this cannot be reasoned out in theoretical terms and most probably involve careful study of a large number of cases through surveys or administrative data.

   (c) This is a **subjective** or **judgment** interpretation of probability, where the player makes his own estimate of the chances of winning. While it may be based on many previous games, thus suggesting a frequency approach, it is difficult to compare games under uniform conditions, so this is primarily a player's own judgment.

   (d) This is an example of the **classical** or **theoretical** approach. A mathematician who understands the structure of bingo games can calculate the probability of various outcomes, assuming that numbers that pop up and are called by the bingo caller are produced randomly.

   (e) This is an example of the **frequency** or **empirical** approach to obtaining probabilities. The statement refers to a result that must have been obtained from a large sample. There is no way this could be reasoned in theoretical terms and statements of this sort are generally obtained from analyses of survey data. Statistics Canada would not ordinarily release a result such as this if it was based only on someone's judgment.

5. Computer file.

a. (i) Tables for statistics on hours at various activities from *Descriptives* and *Frequencies* procedures.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
STHOURS Study Hours	668	0	100	16.61	11.849
EXTHOURS Extracurricular Hours	597	0	50	2.12	4.472
DEPHOURS Care of Dependents	533	0	140	5.65	16.993
HWHOURS Housework Hours	639	0	100	5.14	6.766
VOLHOURS VOLUNTARY WORK HOURS	546	0	33	1.73	3.869
RELHOURS RELIGIOUS ACTIVITIES HOURS	556	0	30	1.10	2.588
Valid N (listwise)	495				

Statistics

		STHOURS Study Hours	EXTHOURS Extracurricular Hours	DEPHOURS Care of Dependents	HWHOURS Housework Hours	VOLHOURS VOLUNTARY WORK HOURS	RELHOURS RELIGIOUS ACTIVITIES HOURS
N	Valid	668	597	533	639	546	556
	Missing	39	110	174	68	161	151
Mean		16.61	2.12	5.65	5.14	1.73	1.10
Std. Deviation		11.849	4.472	16.993	6.766	3.869	2.588
Percentiles	25	8.00	.00	.00	2.00	.00	.00
	50	15.00	.00	.00	3.00	.00	.00
	75	20.00	2.00	.50	6.00	2.00	1.00

(ii)

Activity	s	CRV	Range	IQR
Study hours	11.849	(11.849/16.61) x 100 = 71.3	100-0 = 100	20-8 = 12
Extracurricular hours	4.472	(4.472/2.12) x 100 = 210.9	50-0 = 50	2-0 = 2
Dependent hours	16.993	(16.993/5.65) x 100 = 300.8	140-0 = 140	0.5-0 = 0.5
Housework hours	6.766	(6.766/5.14) x 100 = 131.6	100-0 = 100	6-2 = 4
Voluntary hours	3.869	(3.869/1.73) x 100 = 223.6	33-0 = 33	2-0 = 2
Religious hours	2.588	(2.588/1.10) x 100 = 235.3	30-0 = 30	1-0 = 1

From the summary table, it appears that time spent at care for dependents is the most varied of the six activities. For this variable, the standard deviation, CRV, and range are the largest of any variable. This may be because relatively few students reported care for dependents (note that the IQR is only 0.5 hours, so that at least three-quarters reported a half hour or less caring for dependents) whereas those who do care for dependents report a large number of hours. This produces a great variation for this variable. Time spent at extra curricular activities, voluntary work, and religious activities show similar levels of variation in both absolute and relative terms (similar CRV and standard deviations that do not differ greatly). The small IQR for these, with the seventy-fifth percentile at only 1 or 2 hours means that most students spend relatively little time at these and, while there is variation in the time at these activities, it is not great in absolute terms (small s and IQR) although larger relative to the average or mean.

Study hours show the greatest overall variation in terms of number of hours, with the standard deviation and IQR each around twelve hours. Housework is the next most variable in terms of number of hours, with a standard deviation of 6.8 hours and an IQR of four hours. Since the means for each of these variables are somewhat greater than for the other four activities, the CRV turns out to be smaller. That is, relative to the average, these are the two least varied activities.

b. (i). Means procedure

Report

JOB Hold a job?		STHOURS Study Hours	EXTHOURS Extracurricular Hours	DEPHOURS Care of Dependents	HWHOURS Housework Hours	VOLHOURS VOLUNTARY WORK HOURS	RELHOURS RELIGIOUS ACTIVITIES HOURS
1 No	Mean	17.85	2.89	8.03	5.21	1.67	1.07
	N	265	242	208	255	216	216
	Std. Deviation	12.753	5.492	21.402	5.268	3.788	2.239
2 Yes	Mean	15.75	1.57	4.06	5.01	1.73	1.10
	N	397	351	320	379	326	335
	Std. Deviation	11.141	3.526	13.294	7.545	3.912	2.776
Total	Mean	16.59	2.11	5.63	5.09	1.71	1.09
	N	662	593	528	634	542	551
	Std. Deviation	11.848	4.478	17.049	6.718	3.860	2.577

Report

YEAR Year of Program		STHOURS Study Hours	EXTHOURS Extracurricular Hours
1 First	Mean	14.13	1.57
	N	191	168
	Std. Deviation	8.986	2.907
2 Second	Mean	14.46	2.77
	N	131	113
	Std. Deviation	10.617	6.081
3 Third	Mean	18.67	2.64
	N	174	156
	Std. Deviation	13.308	5.123
4 Fourth	Mean	18.74	1.31
	N	114	107
	Std. Deviation	13.126	2.735
5 Fifth or more	Mean	19.51	2.65
	N	57	52
	Std. Deviation	13.279	5.005
Total	Mean	16.62	2.12
	N	667	596
	Std. Deviation	11.848	4.475

(ii) **Time spent at activities, classified by having or not having a job**. For the first table, the means and standard deviations are larger for those without a job than for those with a job for the following activities: study hours, extracurricular hours, and hours spent caring for dependents. In contrast, time spent at voluntary and religious activities are similar for those with and those without a job, as measured by both the mean and standard deviation. For housework, the mean time of just over five hours per week is similar for the two groups, although those with a job are more varied in the number of hours spent at these activities than are those without a job. While there are insufficient data here to explain why these differences emerge, it could be speculated that either jobs make less time for studying, extracurricular activities, or care of dependents, or that the time respondents spend on these latter activities make less time for a job.

Study and extracurricular activities by year of program. The most notable difference by year of program is that those are in third and higher years of their program say that they devote approximately four or five more hours to studying (18-19 hours) than do those in first and second year (14 hours). The first and second year students are also less varied in the hours devoted to studying, while the standard deviation is approximately 13 hours for third through fifth year students. In contrast, there is no clear pattern for the means of extracurricular activities – second, third, and fifth year students report about one hour more spent weekly at these activities than do first and fourth year students. There are also major differences in the variability by year, but again with no clear pattern.