Social Studies 201
Winter 2005
Answers to Problem Set No. 3
February 28, 2005

1. **Undergraduate students**. Table 1 contains the calculations for obtaining the statistics. The method of differences from the mean is used for grades, and the alternative formula of sums of X and sums of squares of X is used for hours of community service.

Table 1: 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

While not directly requested in the question, the CRV requires the calculation of the means. These are as follows.

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.

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

The coefficient of relative variation is 5.40.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = 0.262/4.85 \times 100 = 0.0540 \times 100 = 5.40$$

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.

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

The CRV is 97.4.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{0.595}{4.34} \times 100 = 13.7$$

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

From the statistics in Table 2, the distribution of grades appears more concentrated than the distribution of hours of community work. Given that grades are measured in different units (grade point average) than

Table 2: Summary of statistics for undergraduate student variables

	Variable	
Statistic	Grades	Community work
Mean	4.85	4.34
Variance	0.069	0.354
Std. dev.	0.262	0.595
CRV	5.4	13.7

community work (hours per week), the standard deviation and variance cannot be meaningfully compared for the two variables. Given that the coefficient of relative variation (CRV) does not have units, the two CRVs can be meaningfully compared. The CRV for grades (5.4) is less than one-half the CRV for community work (13.7). As a result, for these two distributions, grades are more concentrated and hours per week of community service are less concentrated or more dispersed.

2. **Alcohol consumption**. From the distributions of Table 2 of the problem set, the calculations for the mean and standard deviation are given in Table 3.

Table 3: Calculations for mean and standard deviation of alcohol consumption, low and high income individuals

	Low income			High income		
X	f	fX	fX^2	f	fX	fX^2
0	370	0.0	0.0	188	0.0	0.00
2.5	214	535.0	1,337.5	185	462.5	1,156.25
7.0	94	658.0	4,606.0	106	742.0	5,194.00
14.5	54	783.0	11,353.5	74	1,073.0	15,558.50
29.5	30	885.0	26,550.0	29	855.5	25,237.25
Total	762	2,861.0	43,847.0	582	3,133.0	47,146.00

For those with low income,

$$\begin{aligned}
\bar{X} &= \frac{2,861.0}{762} = 3.755 \\
s^2 &= \frac{1}{n-1}\left(\Sigma fX^2 - \frac{(\Sigma fX)^2}{n}\right) \\
&= \frac{1}{761}\left(43,847.00 - \frac{2,861.0^2}{762}\right) \\
&= \frac{1}{761}\left(43,847.00 - 10,741.89\right) \\
&= \frac{1}{761}\left(33,105.11\right) \\
&= 43.502 \\
s &= \sqrt{s^2} = \sqrt{43.502} = 6.596.
\end{aligned}$$

The mean number of alcoholic drinks consumed per week for those with low income is 3.8 drinks and the standard deviation is 6.6 drinks. The CRV is 175.6.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{6.596}{3.755} \times 100 = 175.6$$

For those with higher incomes,

$$\begin{aligned}
\bar{X} &= \frac{3,133.0}{582} = 5.383 \\
s^2 &= \frac{1}{n-1}\left(\Sigma fX^2 - \frac{(\Sigma fX)^2}{n}\right) \\
&= \frac{1}{581}\left(47,146.00 - \frac{3,133.0^2}{582}\right) \\
&= \frac{1}{581}\left(47,146.00 - 16,865.45\right) \\
&= \frac{1}{581}\left(30,280.55\right) \\
&= 52.118 \\
s &= \sqrt{s^2} = \sqrt{52.118} = 7.219.
\end{aligned}$$

The mean number of alcoholic drinks consumed per week for those with high income is 5.4 drinks and the standard deviation is 7.2 drinks. The CRV is 134.1.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{7.219}{5.383} \times 100 = 134.1$$

A summary of the statistics for the two groups is contained in Table 4.

Table 4: Summary of statistics for alcohol consumption of low and high income individuals

	Group	
Statistic	Low income	High income
Mean	3.8	5.4
Variance	43.5	52.1
Std. dev.	6.6	7.2
CRV	175.6	134.1

From Table 4, the answer to this question is not entirely clear cut. In terms of actual amount of alcohol consumed per week, those in the higher income category have greater variation in that the variance and standard deviation for the high income group (52.1 and 7.1) exceed these same statistics for those of lower income (43.5 and 6.6). But the standard deviations are little different, so perhaps the CRV provides a better comparison. The CRV for those of higher incomes (175.6) is considerably greater than the CRV for those with lower incomes (134.1). This is because the mean is considerably lower for those with low income. For those of lower income, the mean is lower and and the standard deviations for the two groups are similar, so this produces a larger CRV for low income and a smaller CRV for those with high incomes.

3. **Education level of Saskatchewan urban population**. The cumulative percentage distributions required for obtaining the percentiles and interquartile range are given in Table 5.

Table 5: Calculations for standard deviation of years of education

	Regina/Saskatoon		Other cities	
X	P	Cum P	P	Cum P
10	20	20	34	34
12	23	43	21	55
13	14	57	17	72
14	17	74	15	87
16	26	100	13	100
	100		100	

Since the levels of education are discrete, there is no need for interpolation in this example. The required percentiles are the years of education where the cumulative percentages first reach twenty-five and seventy-five per cent.

For those in Regina and Saskatoon, the seventy-fifth percentile is $P_{75} = 16$ and the twenty-fifth percentile is $P_{25} = 12$, so the interquartile range is $IQR = 16 - 12 = 4$.

For those in other Saskatchewan cities, the seventy-fifth percentile is $P_{75} = 14$ and the twenty-fifth percentile is $P_{25} = 10$, so the interquartile range is $IQR = 14 - 10 = 4$.

Using the calculations in Table 6 and the formulae for percentages, for those in Regina and Saskatoon, the variance and standard deviation are

$$\begin{aligned} s^2 &= \frac{1}{100}\left(\Sigma PX^2 - \frac{(\Sigma PX)^2}{n}\right) \\ &= \frac{1}{100}\left(17,666 - \frac{1,312^2}{100}\right) \end{aligned}$$

Table 6: Calculations for standard deviation of years of education

	Regina/Saskatoon			Other cities		
X	P	PX	PX^2	P	PX	PX^2
10	20	200	2,000	34	340	3,400
12	23	276	3,312	21	252	3,024
13	14	182	2,366	17	221	2,873
14	17	238	3,332	15	210	2,940
16	26	416	6,656	13	208	3,328
	100	1,312	17,666	100	1,231	15,565

$$\begin{aligned} &= \frac{1}{100}\left(17,666 - 17,213.44\right) \\ &= \frac{1}{100}\left(452.56\right) \\ &= 4.5256 \\ s &= \sqrt{s^2} = \sqrt{4.5256} = 2.13. \end{aligned}$$

For those in other cities in Saskatchewan, the variance and standard deviation are

$$\begin{aligned} s^2 &= \frac{1}{100}\left(\Sigma PX^2 - \frac{(\Sigma PX)^2}{n}\right) \\ &= \frac{1}{100}\left(15,565 - \frac{1,231^2}{100}\right) \\ &= \frac{1}{100}\left(15,565 - 15,153.61\right) \\ &= \frac{1}{100}\left(411.39\right) \\ &= 4.1139 \\ s &= \sqrt{s^2} = \sqrt{4.1139} = 2.03. \end{aligned}$$

The two distributions have very similar variability, perhaps almost identical. The IQR is identical for the adults in the two groups of cities,

although this might be a bit misleading given that the levels of education are grouped into only a discrete set of categories. If had been a more precise measure of years of education, the two IQRs might have been a little different.

In terms of the standard deviations though, they are also very similar, at 2.13 and 2.03 years. Given that the two sets of measures of variation are so similar, there appears to be little difference in the variability of these distributions.

4. **Explanations of probability**.

   (a) This is a **subjective** or **judgment** interpretation of probability. "Likely" cannot be considered a theoretical interpretation since this is not a repeatable situation that can be reproduced under the same circumstances. It may have some aspect of an empirical approach, in that the author uses some data, but again it is not the type of data that applies to a large number of cases. Rather, the likelihood is someone's judgment, presumably based on some information, but still a judgment.

   (b) This is an example of the **classical** or **theoretical** approach. If a coin is flipped four times, with head (H) and tail (T) as the only possible outcomes, there are four ways of obtaining exactly one head: TTTH, TTHT, THTT, and HTTT. In all there are sixteen possible combinations of heads and tails, so the probability of exactly one head is $4/16 = 0.25$. For the answer to this question, you need not compute this, but should explain that this probability can be deduced from this sort of theoretical reasoning.

   (c) 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 problem. For part a., the tables and calculation of statistics are below, with the comments following the tables.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
UED1 Admission Standards too Lax	644	1	5	2.63	.957
UED2 Some Better at Technical School	668	1	5	3.47	.994
UED3 Governments Should Not Finance	672	1	5	1.40	.805
UED4 Lower Tuition	676	1	5	4.09	1.249
Valid N (listwise)	635				

Statistics

		UED1 Admission Standards too Lax	UED2 Some Better at Technical School	UED3 Governments Should Not Finance	UED4 Lower Tuition
N	Valid	644	668	672	676
	Missing	63	39	35	31
Mean		2.63	3.47	1.40	4.09
Std. Deviation		.957	.994	.805	1.249
Percentiles	25	2.00	3.00	1.00	4.00
	50	3.00	4.00	1.00	5.00
	75	3.00	4.00	2.00	5.00

The required statistics are in the table below in bold print.

	Variable			
	UED1	**UED2**	**UED3**	**UED4**
Maximum	5	5	5	5
Minimum	1	1	1	1
Range	**4**	**4**	**4**	**4**
P75	3	4	2	5
P25	2	3	1	4
IQR	**1**	**1**	**1**	**1**
S	0.957	0.994	0.805	1.249
X bar	2.63	3.47	1.40	4.09
CRV	**36.4**	**28.6**	**57.5**	**30.5**

Since all variables are opinion variables, ranging from strongly disagree to strongly agree, they have the same range. They also have the same IQR of 1. As in problem 3, it might be argued that the IQR is not the best measure here, since there are only discrete categories, and they are few in number, so the IQR does not distinguish variation well.

The standard deviations for each of UED1 through UED3 are fairly similar, and it is only UED4 that has a larger standard deviation – over 0.25 points above the others.

It may be that the CRVs are not the best guide to the variability either, since one of the variables (UED3) has such a small mean. Dividing the standard deviation by the mean gives a large value for the CRV for this variable. But the manner in which UED3 is constructed is somewhat different than the other variables – it is a negative statement about funding, so most students

UED1 Admission Standards too Lax

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	82	11.6	12.7	12.7
	2	192	27.2	29.8	42.5
	3	277	39.2	43.0	85.6
	4	71	10.0	11.0	96.6
	5 Strongly Agree	22	3.1	3.4	100.0
	Total	644	91.1	100.0	
Missing	7 Uncetain	1	.1		
	9 No Response	54	7.6		
	System	8	1.1		
	Total	63	8.9		
Total		707	100.0		

UED2 Some Better at Technical School

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	31	4.4	4.6	4.6
	2	64	9.1	9.6	14.2
	3	220	31.1	32.9	47.2
	4	264	37.3	39.5	86.7
	5 Strongly Agree	89	12.6	13.3	100.0
	Total	668	94.5	100.0	
Missing	9 No Response	30	4.2		
	System	9	1.3		
	Total	39	5.5		
Total		707	100.0		

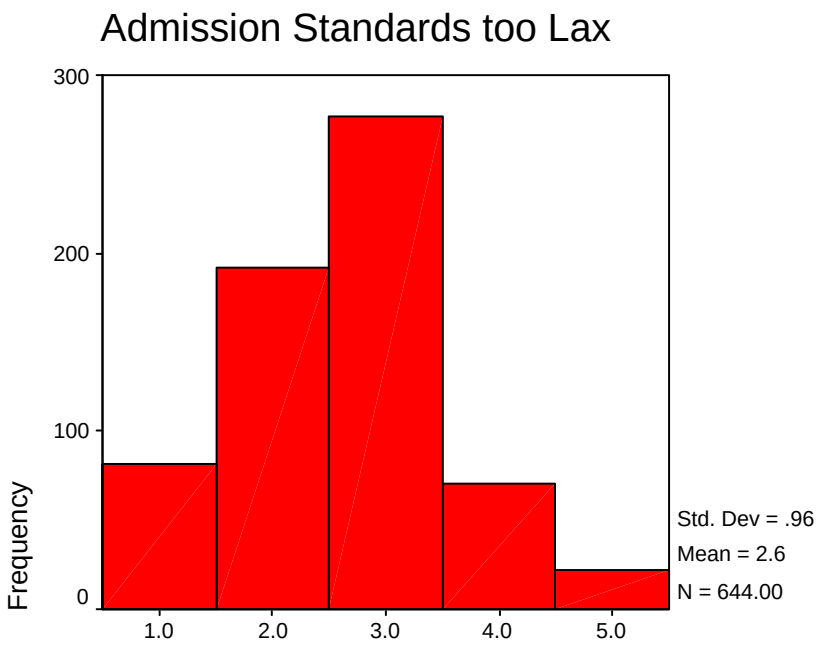
UED3 Governments Should Not Finance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	500	70.7	74.4	74.4
	2	106	15.0	15.8	90.2
	3	44	6.2	6.5	96.7
	4	13	1.8	1.9	98.7
	5 Strongly Agree	9	1.3	1.3	100.0
	Total	672	95.0	100.0	
Missing	9 No Response	27	3.8		
	System	8	1.1		
	Total	35	5.0		
Total		707	100.0		

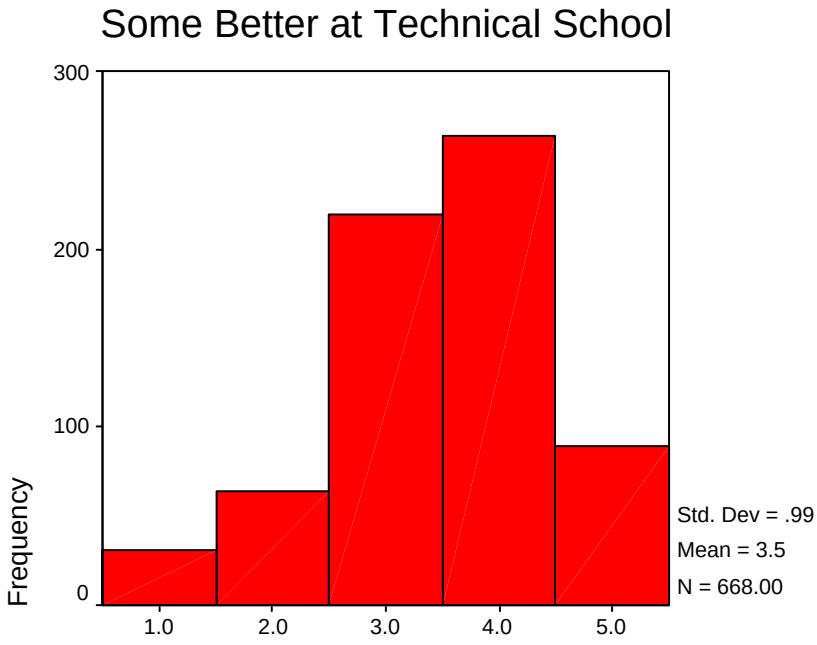
UED4 Lower Tuition

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	50	7.1	7.4	7.4
	2	42	5.9	6.2	13.6
	3	74	10.5	10.9	24.6
	4	144	20.4	21.3	45.9
	5 Strongly Agree	366	51.8	54.1	100.0
	Total	676	95.6	100.0	
Missing	9 No Response	23	3.3		
	System	8	1.1		
	Total	31	4.4		
Total		707	100.0		

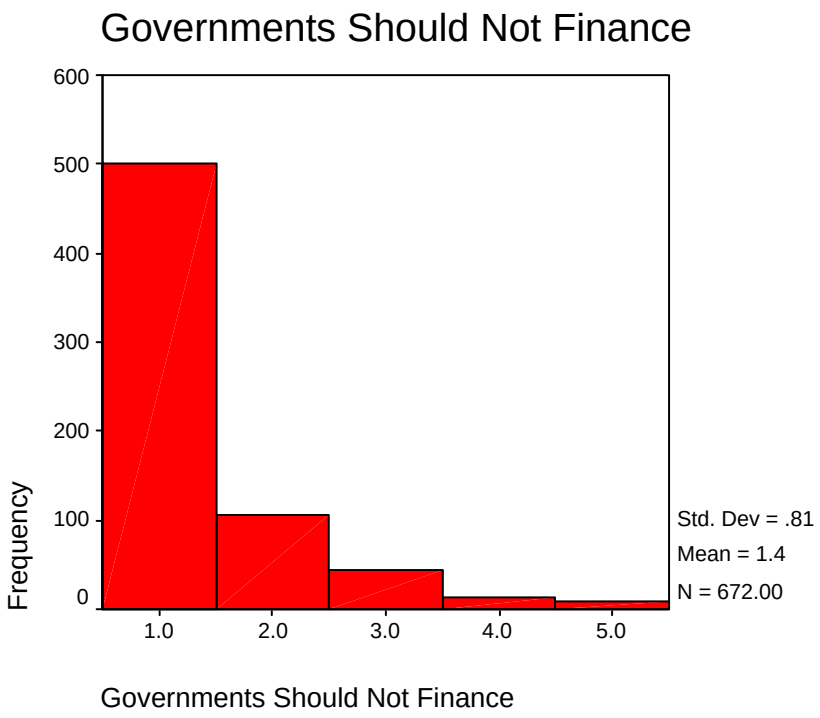
Histogram

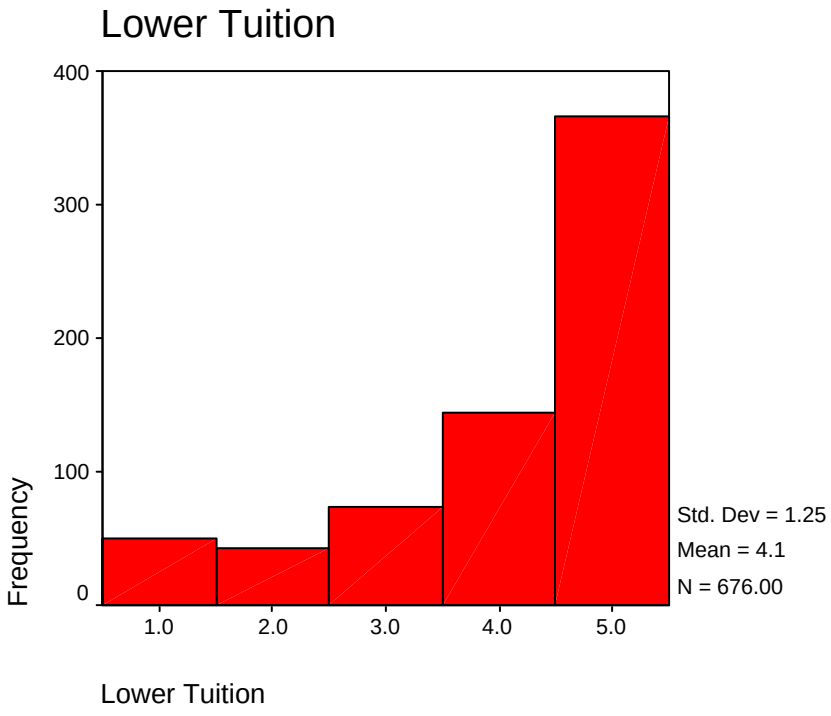


Admission Standards too Lax



Some Better at Technical School





Comparing the last histogram with the three earlier ones illustrates why the standard deviation is greater for UED4 than the other three variables. For UED1 and UED2, the respondents tend to be clustered at responses of 2 and 3 or 3 and 4. For UED3, the respondents are heavily concentrated at response 1. While the responses for UED4 are concentrated at response 4 and 5, there are also a fair number of

responses at response 1. It is this spread across the values that produces a larger standard deviation for UED4.

Question 5. b. Means procedure

JED1 Admission Standards too Lax UED2 Some Better at Technical School UED3 Governments Should Not Finance UED4 Lower Tuition * SEX SEX OF RESPONDENT

SEX SEX OF RESPONDENT		UED1 Admission Standards too Lax	UED2 Some Better at Technical School	UED3 Governments Should Not Finance	UED4 Lower Tuition
1 MALE	Mean	2.78	3.62	1.49	3.92
	N	244	251	252	253
	Std. Deviation	.989	.919	.904	1.293
2 FEMALE	Mean	2.53	3.38	1.35	4.18
	N	400	417	420	423
	Std. Deviation	.925	1.027	.736	1.213
Total	Mean	2.63	3.47	1.40	4.09
	N	644	668	672	676
	Std. Deviation	.957	.994	.805	1.249

Males and females do not have dramatically different views on these issues but there are some consistent differences. Males have larger means than do females for UED1 and UED2, indicating that males are more likely to agree that admission standards are too lax and some at university would be better to go to technical school. Females generally support government funding more than do males. They have greater disagreement with the statement that government should not finance university education (1.35 as opposed to 1.49 for males). They also express more support for reducing tuition than do males (mean of 4.18 as opposed to only 3.92 for females). But the differences between male and female views are not all that great.

UED1 Admission Standards too Lax UED2 Some Better at Technical School UED3 Governments Should Not Finance UED4 Lower Tuition * FV federal political preference

FV federal political preference		UED1 Admission Standards too Lax	UED2 Some Better at Technical School	UED3 Governments Should Not Finance	UED4 Lower Tuition
1 Liberal	Mean	2.68	3.49	1.32	4.24
	N	180	183	185	185
	Std. Deviation	.931	.999	.684	1.160
2 NDP	Mean	2.63	3.29	1.37	4.09
	N	84	89	89	90
	Std. Deviation	1.084	.991	.803	1.233
3 Conservative	Mean	2.63	3.79	1.62	3.79
	N	83	86	87	87
	Std. Deviation	.984	.883	1.037	1.313
4 None	Mean	2.66	3.41	1.32	4.15
	N	140	145	146	147
	Std. Deviation	.958	1.011	.643	1.241
Total	Mean	2.66	3.48	1.38	4.11
	N	487	503	507	509
	Std. Deviation	.973	.992	.772	1.229

In terms of adminission standards, all four groups have much the same view, in that the means are very similar. For the statement about some better off at technical school, those who support the Conservative party are most in agreement (3.79) with NDPers being least in agreement (3.29), while Liberals and those supporting no party are in the middle.

As might be expected, the greatest support for reducing government financing of education comes from the Conservatives (mean of 1.62). The other groups are similar to each other and more in support of government funding, or at least more opposed to ending government funding.

For UED4 the pattern is again different with Conservatives least in support of reducing tuition, but with Liberals and None most in support of reducing tuition. For this variable, the NDPers are between these two groups.