Social Studies 201
Fall 2004
Answers to Problem Set No. 3
October 27, 2004

1. **Ecological footprints**. Each of the three variables has a ratio scale of measurement so the mean, standard deviation, and CRV can be meaningfully obtained. Table 1 contains the calculations for obtaining these statistics. The method of differences from the mean is used for commuting, and the alternative formula of sums of X and sums of squares of X is used for density and energy use.

Table 1: Calculations for mean and standard deviation of municipal ecological footprints

Commuting			Density		Energy use	
X	$X - \bar{X}$	$(X - \bar{X})^2$	X	X^2	X	X^2
7.6	0.78	0.6084	690	476,100	4.21	17.7241
7.7	0.88	0.7744	187	34,969	6.03	36.3609
7.6	0.78	0.6084	100	10,000	5.85	34.2225
4.5	-2.32	5.3824	57	3,249	4.18	17.4724
4.8	-2.02	4.0804	44	1,936	4.11	16.8921
6.0	-0.82	0.6724	162	26,244	3.73	13.9129
8.6	1.78	3.1684	301	90,601	4.09	16.7281
5.4	-1.42	2.0164	195	38,025	3.95	15.6025
9.2	2.38	5.6644	793	628,849	4.05	16.4025
61.4	0.02	22.9756	2,529	1,309,973	40.20	185.3180

For commuting distance, the mean is 6.82 kilometres.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{61.4}{9} = 6.82$$

For density, the mean is 281 people per square kilometre.

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

For energy use, the mean is 4.47 global hectares.

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

For commuting the variance is

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{22.9756}{8} = 2.8720$$

and the standard deviation is 1.69 kilometres.

$$s = \sqrt{s^2} = \sqrt{2.8710} = 1.69$$

The coefficient of relative variation is 24.8.

$$(s/\bar{X}) \times 100 = 1.69/6.82 \times 100 = 0.2478 \times 100 = 24.8$$

For density, the variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right) \\ &= \frac{1}{8}\left(1,309,973 - \frac{2,529^2}{9}\right) \\ &= \frac{1,309,973 - 710,649}{8} \\ &= \frac{599,324}{8} \\ &= 74,915.5 \end{aligned}$$

and the standard deviation is 273.7 people per square kilometre.

$$s = \sqrt{s^2} = \sqrt{73,665.5} = 273.707$$

The CRV is 97.4.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{273.707}{281.0} \times 100 = 97.4$$

For energy use, the variance is

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right) \\ &= \frac{1}{8}\left(185.318 - \frac{40.2^2}{9}\right) \\ &= \frac{185.318 - 179.56}{8} \\ &= \frac{5.758}{8} \\ &= 0.71975 \end{aligned}$$

and the standard deviation is 0.848 global hectares.

$$s = \sqrt{s^2} = \sqrt{0.71975} = 0.8483$$

The CRV is 19.0.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{0.8483}{4.47} \times 100 = 18.979$$

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

Table 2: Summary of statistics for ecological footprints variables

	Ecological footprint variable		
Statistic	Commuting	Density	Energy use
Mean	6.82	281.0	4.47
Variance	2.87	74,915.5	0.7198
Std. dev.	1.69	273.7	0.848
CRV	24.8	97.4	19.0

From the statistics in Table 2, density has the greatest variability and energy use the least. This conclusion might be drawn from the standard deviations, where the standard deviation for density is 273.7 and for

energy use is 0.848. However, it is hazardous to rely on the standard deviations here, since the three variables are measured using entirely different units.

This is an example where the CRV provides a better comparative measure of variation than the standard deviation. For density, the CRV is 97.4, several times greater than the other two CRVs. From this, it is clear that density has by the greatest variation among the three variables. The CRVs for commuting and energy use do not differ greatly, but the CRV for energy use is 19.0, a little less than the CRV for commuting. As a result, variability is least for energy use.

2. **Hours worked at jobs**. From the distributions of annual hours worked in problem 2 of Problem Set 3, the calculations for the mean and standard deviation are given in Table 3.

Table 3: Calculations for mean and standard deviation of annual hours worked at jobs, 15-24 and 35-44 year olds

	15-24			35-44		
X	f	fX	fX^2	f	fX	fX^2
0.5	129	64.5	32.25	39	19.5	9.75
1.0	91	91.0	91.00	44	44.0	44.00
1.5	93	139.5	209.25	74	111.0	166.50
2.0	132	264.0	528.00	380	760.0	1,520.00
2.5	34	85.0	212.50	63	157.5	393.75
3.0	23	69.0	207.00	84	252.0	756.00
Total	502	713.0	1,280.00	684	1,344.0	2,890.00

For 15-24 year olds,

$$\bar{X} = \frac{713.0}{502} = 1.420$$

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

$$\begin{aligned} &= \frac{1}{501}\left(1,280.00 - \frac{713.0^2}{502}\right) \\ &= \frac{1}{501}\left(1,280.00 - 1,012.69\right) \\ &= \frac{1}{501}\left(267.313\right) \\ &= 0.534 \\ s &= \sqrt{s^2} = \sqrt{0.534} = 0.730. \end{aligned}$$

The mean annual hours worked at jobs for 15-24 year olds is 1,460 hours and the standard deviation is 730 hours. The CRV is 51.4.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{0.730}{1.420} \times 100 = 51.4$$

For 35-44 year olds,

$$\begin{aligned} \bar{X} &= \frac{1,344.0}{684} = 1.965 \\ s^2 &= \frac{1}{n-1}\left(\Sigma f X^2 - \frac{(\Sigma f X)^2}{n}\right) \\ &= \frac{1}{683}\left(2,890.00 - \frac{1,344.0^2}{684}\right) \\ &= \frac{1}{683}\left(2,890.00 - 2,640.84\right) \\ &= \frac{1}{683}\left(249.1579\right) \\ &= 0.3648 \\ s &= \sqrt{s^2} = \sqrt{0.3648} = 0.604. \end{aligned}$$

The mean annual hours worked at jobs for 35-44 year olds is 1,965 hours and the standard deviation is 604 hours. The CRV is 30.7.

$$\text{CRV} = \frac{s}{\bar{X}} \times 100 = \frac{0.604}{1.965} \times 100 = 30.7$$

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

Table 4: Summary of statistics for annual hours worked for 15-24 and 35-44 year olds

	Group	
Statistic	15-24	35-44
Mean	1,420	1,965
Std. dev.	730	604
CRV	51.4	30.7

From Table 4 and Table 2 of Problem Set 3, it is fairly clear that the variation in annual hours worked is less for 35-44 year olds than for 15-24 year olds. Both the standard deviation and the CRV are lower for the older age group. While the standard deviations do not differ greatly, the larger mean for the older age group leads to a CRV that is considerably less for the older age group. From the frequency distributions in the original table of the Problem Set, it is apparent that 35-44 year olds tend to be concentrated in the 1,750 to 2,250 interval, whereas the 15-24 year olds are more spread out across the different intervals.

3. **Interpretations of probability**.

   (a) This is a frequency interpretation of probability since it is based on "findings" of a study, presumably from a large sample size of parents and children. There is no way this result could be reasoned from the ideal principles of probability, since the combinations of results could not be known in theory. This would not seem to be a subjective probability using someone's best judgment about the chances of this occurring, but is a result obtained by studying children and parents.

   (b) This is a subjective probability since it is based on Graeme Smith's judgment about the chance that Mr. Wall will make the announcement. This could not be reasoned using logical deduction or theoretical considerations and the situation is not repeatable, so this probability is no more than a judgment.

   (c) Like part (a), this appears to be a frequency interpretation of probability in that it comes from a study of workers in a variety of occupations. Evidence for this approach comes form the citing of percentages and averages relating to workers across occupations.

   (d) These odds appear to be derived from theoretical considerations. Hopefully these are more than someone's subjective judgment of the probability of winning. Further, each lottery draw occurs only at one time, so this cannot be a frequency interpretation. These probabilities seem to be obtained by considering the number of tickets sold and using principles of probability to obtain the odds of winning.

4. Variation and patterns

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
FUTURE Economic Future?	700	1	3	1.78	.765
KNOW Knowledge about Computing	701	1	4	2.18	.719
PREP Univ Preparing	692	1	5	3.03	.973
ACCESS Lack of Access	689	1	5	2.76	1.244
Valid N (listwise)	681				

Variable	Range	CRV
FUTURE	**3 – 1 = 2**	**(0.765 / 1.78) X 100 = 43.0**
KNOW	**4 – 1 = 3**	**(0.719 / 2.18) X 100 = 33.0**
PREP	**5 – 1 = 4**	**(0.973 / 3.03) X 100 = 32.1**
ACCESS	**5 – 1 = 4**	**(1.244 / 2.76) X 100 = 45.1**

From the standard deviations, the variables from most to least variable are ACCESS (1,244), PREP (0.973), FUTURE (0.765), and KNOW (0.719). However, the range for the variables differs considerably – FUTURE has a range of only 2, from 1 to 3, whereas PREP and ACCESS each have a range double that – of 4, from 1 to 5. Given the larger possible set of values across which PREP and ACCESS vary, it is no great surprise that these two variables have the largest standard deviation.

In this case, the CRVs might be a better means of comparing the variability of the variables. ACCESS still has the largest relative variability, with a CRV of 45.1, while PREP has the least relative variability, with a CRV of only 32.1. FUTURE, with one of the smallest absolute variations, turns out to have a large relative variability of 43.0, almost as large as the relative variability of ACCESS.

By looking at the histograms, the same conclusions emerge. The variables KNOW and PREP each appear more concentrated at or near the centre values, with high bars for value 2 for KNOW and values 2 through 4 for PREP. In contrast, the histograms for FUTURE and ACCESS are more spread out. FUTURE has a small standard deviation only because of its small potential range.

In summary, variability differs across the variables for two reasons – differences in potential values (1-3, 1-4, and 1-5) and inherent variability of student responses. For this question, the CRV may be preferable as a measure when comparing relative variability across the four variables.

FUTURE Economic Future?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Better Off	300	42.4	42.9	42.9
	2 About Same	256	36.2	36.6	79.4
	3 Worse Off	144	20.4	20.6	100.0
	Total	700	99.0	100.0	
Missing	9 No Response	7	1.0		
Total		707	100.0		

KNOW Knowledge about Computing

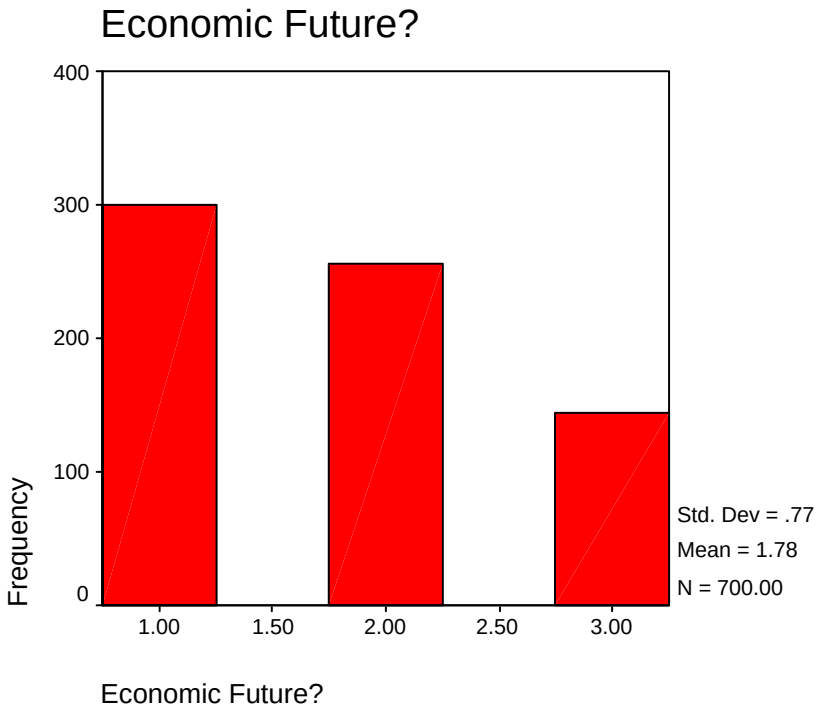
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Very knowledgeable	104	14.7	14.8	14.8
	2 Somewhat Knowledgeable	395	55.9	56.3	71.2
	3 Little Knowledge	176	24.9	25.1	96.3
	4 Not Knowledgeable	26	3.7	3.7	100.0
	Total	701	99.2	100.0	
Missing	7 Uncertain	1	.1		
	9 No Response	4	.6		
	System	1	.1		
	Total	6	.8		
Total		707	100.0		

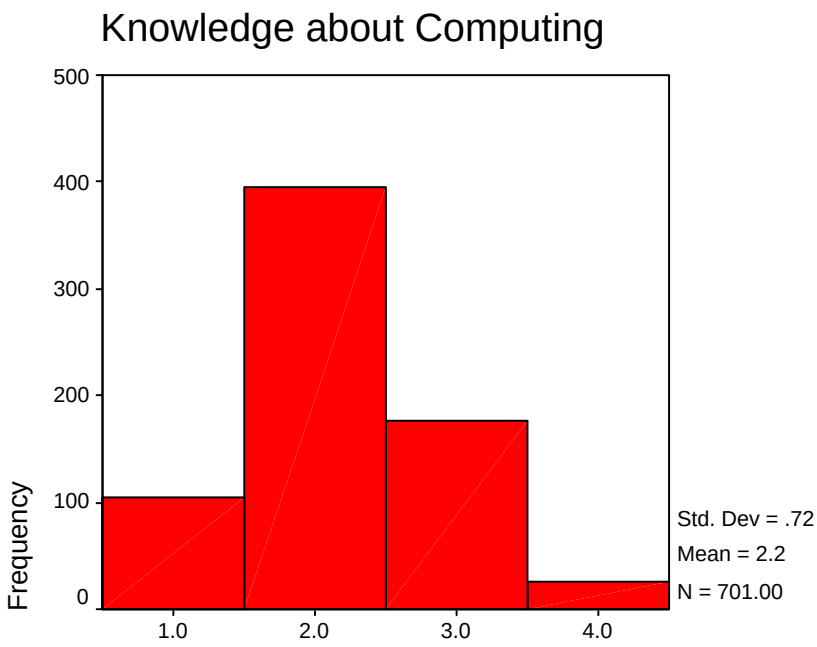
PREP Univ Preparing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	42	5.9	6.1	6.1
	2	151	21.4	21.8	27.9
	3	280	39.6	40.5	68.4
	4	180	25.5	26.0	94.4
	5 Strongly Agree	39	5.5	5.6	100.0
	Total	692	97.9	100.0	
Missing	8 Not applicable	1	.1		
	9 No response	13	1.8		
	System	1	.1		
	Total	15	2.1		
Total		707	100.0		

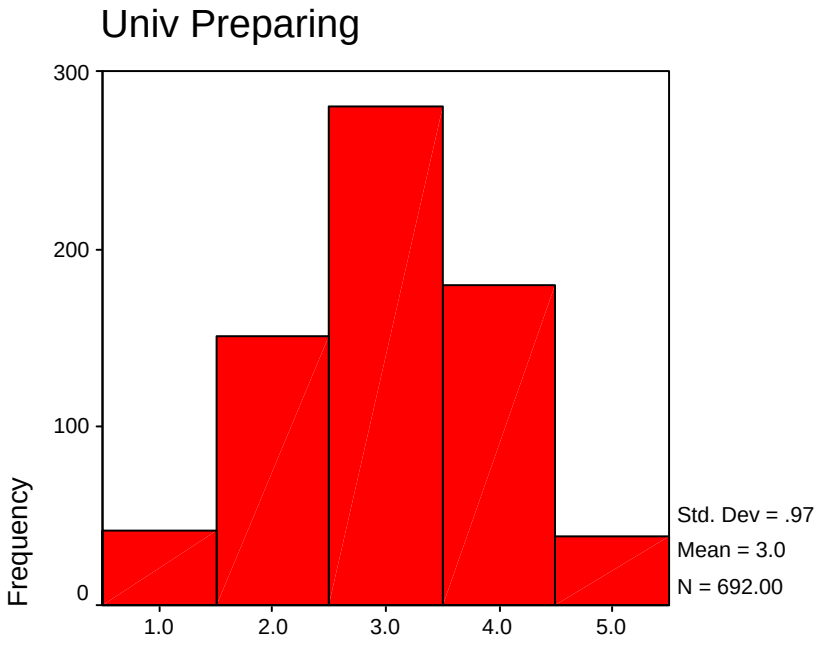
ACCESS Lack of Access

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	134	19.0	19.4	19.4
	2	157	22.2	22.8	42.2
	3	212	30.0	30.8	73.0
	4	111	15.7	16.1	89.1
	5 Strongly Agree	75	10.6	10.9	100.0
	Total	689	97.5	100.0	
Missing	8 NOT APPLICABLE	2	.3		
	9 NO RESPONSE	15	2.1		
	System	1	.1		
	Total	18	2.5		
Total		707	100.0		

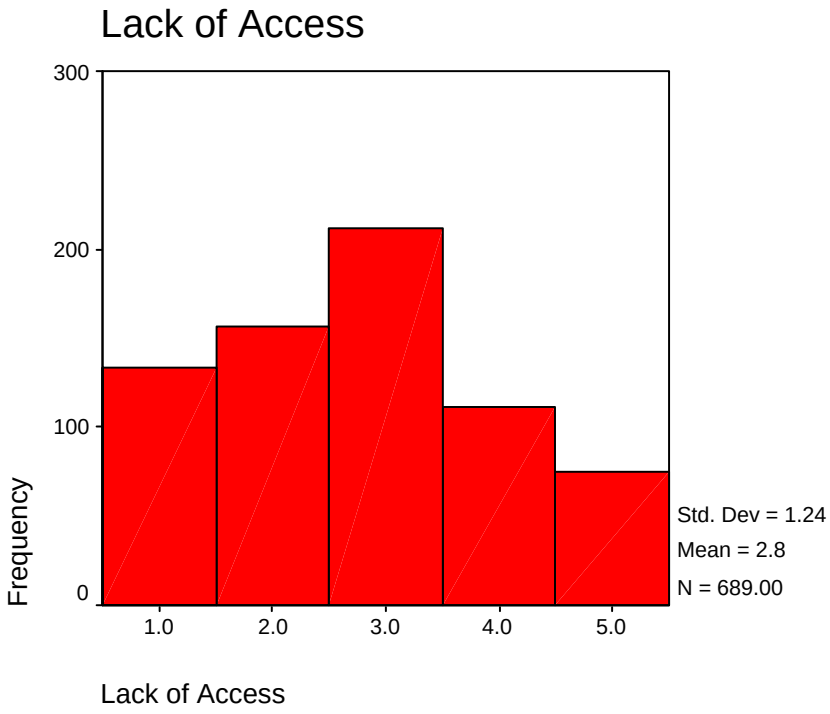




Knowledge about Computing



Univ Preparing


Lack of Access
300
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Lack of Access
Std. Dev = 1.24
Mean = 2.8
N = 689.00

4. Views about multiculturalism by political preference

Report

PV provincial political preference		M2 Equal Access	M4 Eliminate Barriers	M5 Fund Festivals	M6 Canadian Society Enriched
1 Liberal	Mean	4.59	4.07	3.08	4.34
	N	95	94	95	95
	Std. Deviation	.707	1.070	1.277	.883
2 NDP	Mean	4.60	4.19	2.99	4.45
	N	174	170	174	174
	Std. Deviation	.635	.823	1.219	.726
3 Conservative	Mean	4.39	3.97	2.67	4.05
	N	99	98	99	99
	Std. Deviation	.913	1.069	1.262	1.014
4 None	Mean	4.49	4.09	3.16	4.30
	N	166	162	165	165
	Std. Deviation	.822	.948	1.256	.830
Total	Mean	4.52	4.10	3.00	4.31
	N	534	524	533	533
	Std. Deviation	.767	.958	1.257	.854

Each of these variables states a principle of multiculturalism and asks students their view about the statement or principle, asking them to respond on a five-point scale from strongly disagree, indicated by 1, to strongly agree, indicated by 5. A larger mean indicates greater agreement with the principle of multiculturalism.

The first point that might be noted is that mean responses do not differ greatly across political preference. In no case do the means differ by more than about 0.4 points on the five-point scale.

Second, those of Conservative political preference consistently have a lower mean than those of other political preferences. While the means for Conservative are not a lot lower than the other political preferences, they are from 0.1 to 0.3 points lower. This can be interpreted as less support for multiculturalism by Conservatives than by others.

Third, Liberals have neither the largest nor smallest means, indicating they tend to be in the centre. Those supporting the NDP are most strongly in support of three of the four statements (M2, M4, and M6) but not for funding festivals, where None are most supportive. Those of no political preference are also in between the extremes.

In terms of specific questions, equal access is strongly supported by all groups, with means from 4.39 to 4.60 – these are large means, given the scale goes only from 1 to 5. Second greatest support is given to Canadian society enriched, with means only a little lower than for equal access. For both of these, NDP is most supportive, Conservative least, with Liberal and None in between.

Responses to the fund festivals variable is most varied, with larger standard deviations and a somewhat different pattern of means. In general, responses on funding festivals indicate a split between agree and disagree, since the mean response is exactly 3, in the

middle, with approximately equal numbers of respondents on each of the agree and disagree sides of centre.

Finally, the patterns for eliminating barriers are similar to those of equal access and enriched, except that support for eliminating barriers is not as strong as for the other two variables. That is, the mean is around 4, rather than 4.5.

In summary, there is generally strong support for equal access and enriched, middle level of support for eliminate barriers, and not strong support for funding festivals. Conservatives tend to be least supportive, NDP most supportive, and Liberals and None in the middle.