Social Studies 201
Answers to Problem Set 4
March 11, 2005

1. **Religiosity of Saskatchewan adults**

   (a) Using the cross-classification of Table 1 of Problem Set 4, the probabilities are as follows.

      i. The probability of attending at least once a month when age 15 is 0.672.

$$P(\text{at least once a month, age 15}) = P(M15)$$

$$P(M15) = \frac{N(M15)}{N} = \frac{718}{1,069} = 0.672$$

      ii. The probability of attending several times a year in 2001 is 0.303.

$$P(\text{several times a year, 2001}) = P(S)$$

$$P(S) = \frac{N(S)}{N} = \frac{324}{1,069} = 0.303$$

      iii. The chance of selecting an individual who was or is a regular attender, at least once a month is 0.756.

$$P(M15 \text{ or } M) = P(\text{were regular}) + P(\text{are regular}) - P(\text{were and are})$$

$$P(M15 \text{ or } M) = \frac{718}{1,069} + \frac{510}{1,069} - \frac{420}{1,069} = \frac{808}{1,069} = 0.756$$

      iv. The probability of never attending religious services is the probability of not attending at age 15 and not attending in 2001. This probability is 0.093.

$$P(N15 \text{ and } N) = \frac{\text{not at all, age 15 and not at all, 2001}}{\text{total no. of cases}}$$

$$P(N15 \text{ and } N) = \frac{N \text{ and } N15}{N} = \frac{99}{1,069} = 0.093$$

v. The conditional probability of attending not at all in 2001 (event N) given attendance of at least once a month at age 15 (event M15) is 0.142.

$$P(\text{not at all, 2001/once a month, age 15}) = P(N/M15)$$

$$P(N/M15) = \frac{102}{718} = 0.142$$

vi. The conditional probability of attending several times a year at age 15 (event S15), given attending at least once a month in 2001 (event M) is 0.092.

$$P(S15/M) = \frac{47}{510} = 0.092$$

(b) In order to determine whether the two events are independent of each other or not, it is necessary to check whether the conditional probability of S15 given event N, and the overall probability of event S15 are equal or not. These probabilities are

$$P(S15/N) = \frac{34}{235} = 0.145$$

$$P(S15) = \frac{170}{1,069} = 0.159$$

These two probabilities differ by only a small amount, so while the two events are not entirely independent of each other, they are very close to being independent of each other.

An alternative way to check this is to see whether the probability of even N given event S15, equals the probability of event N. These probabilities are

$$P(N/S15) = \frac{34}{170} = 0.200$$

$$P(N) = \frac{235}{1,069} = 0.220$$

These two probabilities differ so the two events are not independent of each other, but these two probabilities are not all that different from each other, so the two events are close to being independent of each other.

(c) There are 718 respondents who attended at least once a month at age 15 (event M15). Of these, 420 attended at least once a month in 2001. The probability of the same attendance pattern in 2001 for those who attended at least once a month at age 15 is

$$P(M/M15) = \frac{420}{718} = 0.585$$

For those who attended several times a year at age 15 (event S15) the probability that they had the same attendance pattern in 2001 is

$$P(S/S15) = \frac{89}{170} = 0.524$$

Those who attended at least once a month at age fifteen are a little more likely to have the same attendance pattern in 2001 (probability of 0.585) compared with those who attended several times a year at age 15 (probability of 0.524).

2. **Explanations of probability**

(a) The events of being age 30-44 (or over age 60) and the event of correctly identifying King as prime minister during the second world war are depdendent events. For those over age 60, one-half correctly identified King, while for those aged 30-44, only 25 per cent correctly identified King. The probability of correctly identifying King is lower for those age 30-44 (0.25) than for those over age 60 (0.50). This means that the probability of correctly identifying King differs by age group, so that the event of being in any particular age group and the event of correctly identifying King are dependent on each other.

(b) In the first part of the quote, Palmer argues that prejudice and racism increase with the number of immigrants. This means that if there are more immigrants, the probability of prejudice and racism is greater than when there are fewer immigrants. The extent of prejudice and racism is dependent on the number of immigrants.

In the last quote, Palmer argues that negative feelings toward immigrants are much the same, regardless of whether unemployment rates are high or low. This implies that negative feelings

toward immigrants are independent of the events of high or low unemployment.

3. **Standardized normal distribution**

(a) For $Z = 1.43$, the A area is 0.4236 and this is the area between $Z = 0$ and $Z = 1.43$.

(b) The area between 0.4 and 1.6 is the area between the centre and $Z = 1.60$ (an A area), minus the area between the centre of the normal distribution and $Z = 0.40$ (also an A area). The area between the centre of the normal distribution and $Z = 1.60$ is 0.4452. The area between centre and $Z = 0.40$ is 0.1554. The area between $Z = 0.40$ and $Z = 1.60$ is thus $0.4453 - 0.1554 = 0.2899$.

(c) The proportion of cases between $Z = -0.72$ and $Z = 0.87$ is the sum of the areas between the centre of the distribution and each of these Z values. Between $Z = 0$ and $Z = -0.72$, the area is 0.2642 while the area between $Z = 0$ and $Z = 0.87$ is 0.3079. The sum of these two areas is $0.2642 + 0.3079 = 0.5721$. As a percentage of the total area, this is $0.5721 \times 100 = 57.21\%$.

(d) The area to the left of $Z = 1.25$ is the A area associated with $Z = 1.25$ (from centre to 1.25) plus the one-half of the area to the left of centre. For $Z = 1.25$, the A area is 0.3944. The required proportion is $0.3944 + 0.5000 = 0.8944$.

(e) The area under the normal curve with a value larger than $Z = -1.33$ is the area between centre and $Z = -1.33$ plus the one-half of the area to the right of centre. For $Z = -1.33$, the A area is 0.4082 and thus the required area is $0.4082 + 0.5000 = 0.9082$.

(f) To determine this, $Z = \pm 1.50$ and the area from $Z = 1.50$ to the right end of the distribution is is 0.0668. Similarly, the area from $Z = -1.50$ to the left end of the distribution is 0.0668. This means that the total area beyond 1.50 standard deviations of the centre is $0.0668 + 0.0668 = 0.1336$. As a percentage, there are $0.1336 \times 100\% = 13.36\%$ of cases more than one and one-half standard deviations from the mean.

(g) The 35th percentile occurs at a Z so that there is a B area of 0.3500 less than this. For a B area of 0.3500, $Z = 0.38$ or $Z = 0.39$. The Z value so that 0.3500 is to the left of this is $Z = -0.38$ or $Z = -0.39$. That is, $P_{35} = -0.38$ or -0.39.

(h) For an area of 0.04 in a tail of the distribution, the B area of 0.0400 is associated with a $Z = \pm 1.75$. Outside the Z values of plus and minus 1.75, there is 0.04 of the distribution in each tail, for a total of 0.08.

(i) The sixty-second percentile is to the right of centre, at $62-50 = 12$ percentage points above centre. An A area of 0.1200 is at $Z = 0.31$ and this is the position of P_{62}. That is, $P_{62} = 0.31$.

(j) The eighty per cent range is the Z-values so that there is a total of 20% or 0.20 in the two tails of the distribution, or 0.10 in each tail. A B area of 0.1000 is associated with A Z of 1.28. The eighty per cent range is thus from $Z = -1.28$ to $Z = +1.28$, so this range is 2.56 in the standardized normal distribution.

4. **Distribution of years of education**

(a) For this question, X is the variable representing years of education, the mean $\mu = 12.4$ and the standard deviation $\sigma = 2.1$ years. The distribution of years of education is assumed to have a normal distribution.

i. The For $X = 12$ years of education, $Z = (X - \mu)/\sigma = (12 - 12.4)/2.1 = -0.4/2.1 = -0.19$. The proportion with less than 12.4 years of education is the B area to the left of $Z = -0.19$. This area is 0.4246, and this represents the proportion of adults with less than 12 years of education if the distribution of education is normal.

ii. For the proportion with between 9 and 12 years of education it is necessary to calculate Z for each of these. For $X = 9$,

$$Z = (X - \mu)/\sigma = (9 - 12.4)/2.1 = -3.4/2.1 = -1.62$$

From part i., the Z for 12 years of education is -0.19, with an A area of 0.0754 between $Z = -0.19$ and $Z = 0$. The area

between $Z = -1.62$ and $Z = 0$ is an A area of 0.4474. The difference between these two areas is $0.4474 - 0.0754 = 0.3720$. If the distribution is normal, this is the proportion of adults with at least 9 but less than 12 years of education.

iii. For $X = 15$ years of education, $Z = (X - \mu)/\sigma = (15 - 12.4)/2.1 = 2.6/2.1 = 1.24$. The B area to the right of $Z = 1.24$ is 0.1075 and, as a percentage, this is $0.1075 \times 100\% = 10.75\%$. The percentage is thus approximately 10.8%.

iv. The interquartile range is the range from the twenty-fifth percentile to the seventy-fifth percentile, so that there is 0.2500 of the distribution in each of the two tails of the distribution. An B area of 0.2500 is associated with $Z = 0.67$, so the interquartile range is from $Z = -0.67$ to $Z = +0.67$. Converting these Z values into years of education using the transformation from Z to X, this gives the X values

$$X = \mu + Z\sigma = 12.4 + (-0.67 \times 2.1) = 12.40 - 1.41 = 10.99$$

or 11.0 years to

$$X = \mu + Z\sigma = 12.4 + (0.67 \times 2.1) = 12.40 + 1.41 = 13.81$$

or 13.8 years. Thus the interquartile range is $13.8 - 11.0 = 2.8$ years.

(b) Comparison of the actual and normal distribution. From the diagram of the normal distribution superimposed on the histogram of years of education, it is apparent that the actual distribution has some similarities to the normal distribution, and other differences. There are many more persons with 12 years of education than what would be expected if the distribution were exactly normal. At the same time, the distribution of years of education is reasonably symmetrical and peaked in the centre.

The following table demonstrates several places where the normal distribution differs from the actual distribution. The percentages for the normal distribution come from part (a) and the actual percentages come from Table 2 of the problem set.

From the table, there are fewer persons with either 9-12 or less than 12 years of education in the actual than normal distribution.

Years	Normal %	Actual %
Less than 12	42.5%	32.7%
9 to under 12	37.2%	21.4%
More than 15	10.8%	15.8%

This may occur because there are more at exactly 12 years than in the normal distribution (the tall bar in the histogram). At the very upper end of 15 or more years, the situation is reversed, with more persons, 15.8%, than in the case of a normal distribution, 10.8%. The reason for this might be that many people plan or attempt to complete a particular type of schooling to obtain the credential, whereas there are not so many who only partially complete. That is, there are a lot who complete secondary (12 years) or university (15 or 16 years), but relatively few in between.

In summary, the distribution of years of education is not really normal, but has a similar general shape to a normal distribution. But there are more persons at the middle and then again at the upper end, compared with what the distribution would be if it was exactly normal.

5. Computer problems

a. Statistics, frequency distribution, and histogram with normal curve for JOBHOURS.

Statistics

JOBHOURS HOURS PER WEEK AT JOB - F98

N	Valid	396
	Missing	311
Mean		20.12
Median		20.00
Std. Deviation		11.764

For the distribution of JOBHOURS, the mean is 20.1 and the standard deviation is 11.8. In order to determine the percentage of cases within one and two standard deviations of the mean, construct the appropriate intervals and then count the frequency of occurrence of cases within these intervals.

For within one standard deviation of the mean, the interval is

20.1 – 11.8 = 8.3 to 20.1 + 11.8 = 31.9 or from 9 hours to under 32 hours.

Looking at the frequency distribution on the next page, this includes all the cases from 9 to 31, that is, 2 + 24 + 2 + …. + 7 + 24 = 279. So there are 279 cases within one standard deviation of the mean.

In total, there are 396 cases, so within one standard deviation of the mean there are (279/396) x 100% = 70.5% of all cases.

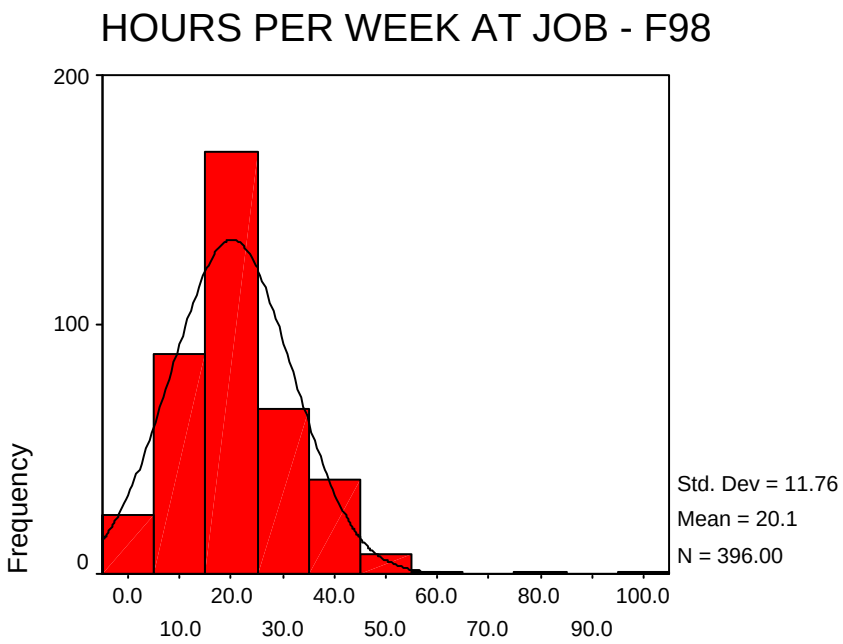
One standard deviation is 11.764 hours so two standard deviations is 23.528, or 23.5 hours. The interval within two standard deviations of the mean is

20.1 – 23.5 = -3.4 (but since hours cannot be less than 0, let this be 0).

To 20.1 + 23.5 = 43.6 hours .

That is, this interval contains all the cases from 0 to 43.6, but since there are none that are at 43.6, this includes all the cases up to and including 40 hours. This is all but the 11 cases of more than 40 hours, or 385 of the 396 cases. This is 97.2% of all the cases.

From the table of the normal distribution, there are 68.26% of cases within one standard deviation (2 x 0.3413 = 0.6826) and 95.44% within two standard deviations (2 x 0.4772=0.9544). This compares to 70.5% and 97.2%, respectively, so the actual percentage of cases is close to the percentage of cases in a normal distribution.



From the diagram it can be seen that the distribution of hours worked at jobs has a similar shape to the normal distribution, in that there is a peak at the centre, with relatively fewer cases the further one goes from the centre of the distribution. But the actual distribution is more concentrated at the centre (tall bar in the centre) than is the normal distribution.

While the percentage of cases within one and standard deviations is very similar to that of the normal distribution, this does not necessarily mean that the distribution of hours worked is normal, since there may be other differences. But in terms of the percentage of cases within one and two standard deviations, the actual percentages are very close to those of the normal distribution.

JOBHOURS HOURS PER WEEK AT JOB - F98

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	.3	.5	.5
	2	4	.6	1.0	1.5
	3	6	.8	1.5	3.0
	4	11	1.6	2.8	5.8
	5	1	.1	.3	6.1
	5	11	1.6	2.8	8.8
	6	11	1.6	2.8	11.6
	7	1	.1	.3	11.9
	8	1	.1	.3	12.1
	8	12	1.7	3.0	15.2
	9	2	.3	.5	15.7
	10	24	3.4	6.1	21.7
	11	2	.3	.5	22.2
	12	13	1.8	3.3	25.5
	13	1	.1	.3	25.8
	13	4	.6	1.0	26.8
	14	6	.8	1.5	28.3
	15	35	5.0	8.8	37.1
	16	21	3.0	5.3	42.4
	17	7	1.0	1.8	44.2
	18	2	.3	.5	44.7
	18	15	2.1	3.8	48.5
	19	1	.1	.3	48.7
	20	61	8.6	15.4	64.1
	21	4	.6	1.0	65.2
	22	7	1.0	1.8	66.9
	23	2	.3	.5	67.4
	23	7	1.0	1.8	69.2
	24	7	1.0	1.8	71.0
	25	21	3.0	5.3	76.3
	26	2	.3	.5	76.8
	27	4	.6	1.0	77.8
	28	7	1.0	1.8	79.5
	30	24	3.4	6.1	85.6
	32	6	.8	1.5	87.1
	33	1	.1	.3	87.4
	34	1	.1	.3	87.6
	35	7	1.0	1.8	89.4
	36	2	.3	.5	89.9
	37	2	.3	.5	90.4
	38	2	.3	.5	90.9
	38	1	.1	.3	91.2
	40	24	3.4	6.1	97.2
	45	3	.4	.8	98.0
	46	1	.1	.3	98.2
	50	4	.6	1.0	99.2
	56	1	.1	.3	99.5
	80	1	.1	.3	99.7
	100	1	.1	.3	100.0
	Total	396	56.0	100.0	
Missing	97 UNCERTAIN	2	.3		
	98 NOT APPLICABLE	1	.1		
	99 NO RESPONSE	6	.8		
	System	302	42.7		
	Total	311	44.0		
Total		707	100.0		

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
Median		3.00	4.00	1.00	5.00
Std. Deviation		.957	.994	.805	1.249

Frequency Table

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		

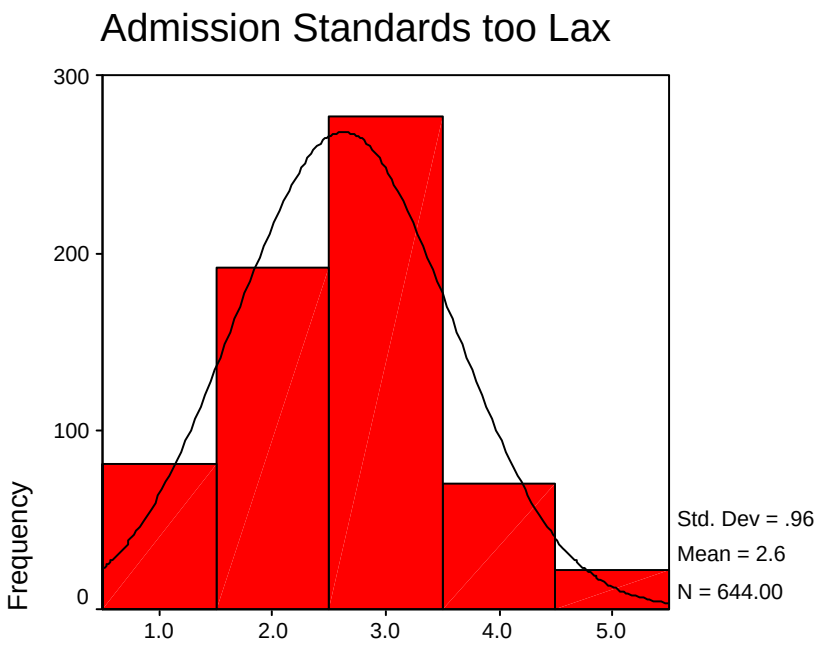
UED3 Governments Should Not Finance

		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		

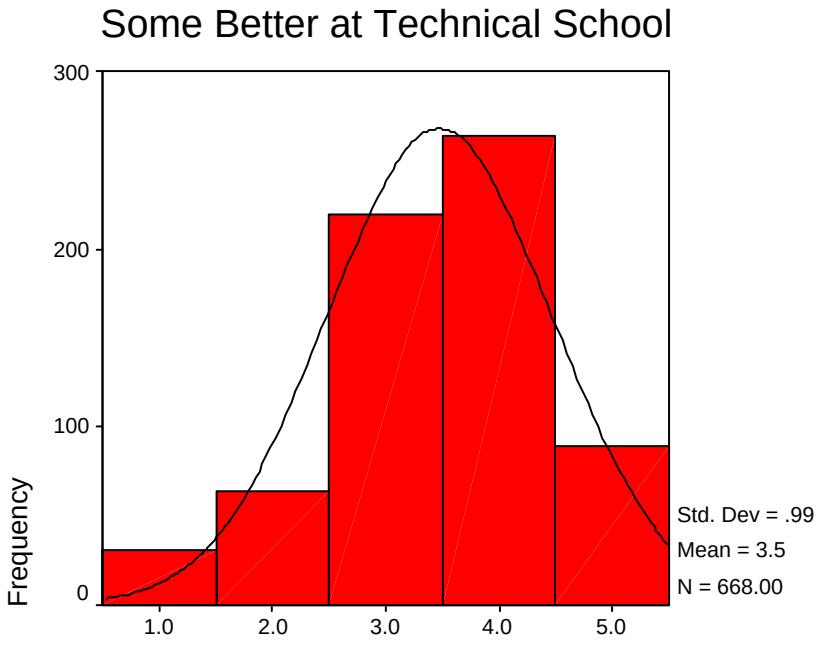
UED4 Lower Tuition

		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		

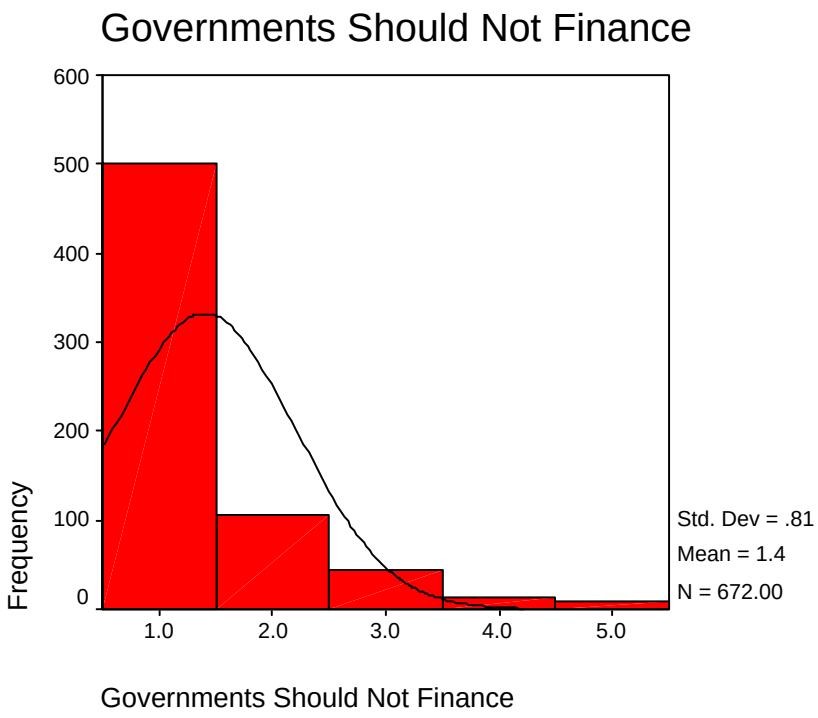
Histogram

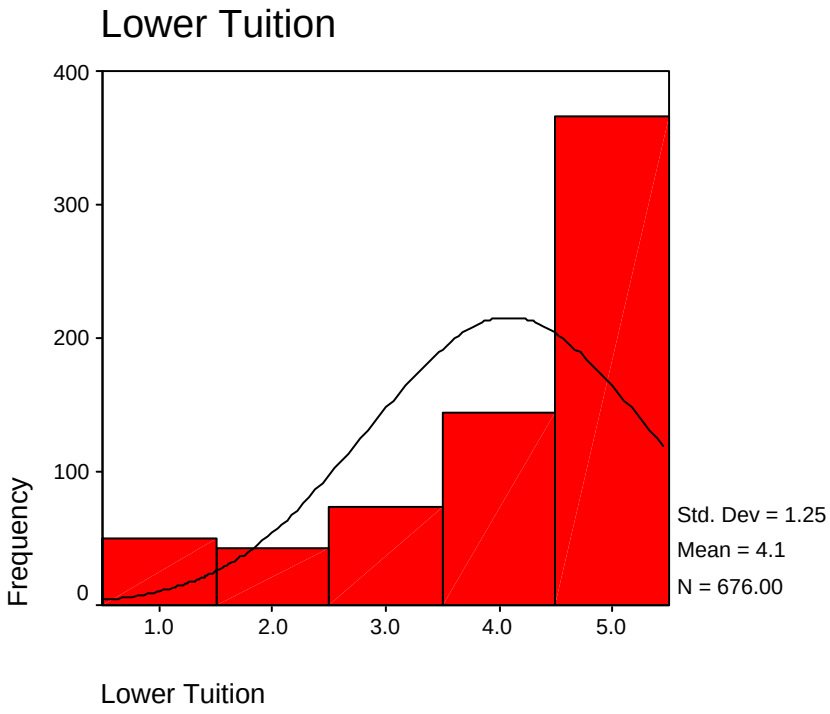


Admission Standards too Lax



Some Better at Technical School


Governments Should Not Finance
600
500
400
300
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Std. Dev = .81
Mean = 1.4
N = 672.00
Governments Should Not Finance


Lower Tuition
400
300
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Std. Dev = 1.25
Mean = 4.1
N = 676.00
Lower Tuition

b. UED1 through UED4 – similarities and differences from the normal distribution.

The two variables UED1 and UED2 have distributions that are fairly similar to the normal distribution. Probably UED1, admission standards are too lax, has a distribution that is closest to the normal, in that UED1 has a peak in the centre. The two variables UED3 and UED4 are very different from the normal distribution, since each of these variables is concentrated at one end of the distribution. Neither is very close to normal – while UED3 is concentrated at the lower end and UED4 at the upper end, both of these distributions are so different from the normal that there is little to choose between them as being most different from the normal.

c. Crosstabulations and probability.

UED4 Lower Tuition * SEX SEX OF RESPONDENT Crosstabulation

Count

		SEX SEX OF RESPONDENT		
		1 MALE	2 FEMALE	Total
UED4 Lower Tuition	1 Strongly Disagree	21	29	50
	2	21	21	42
	3	32	42	74
	4	61	83	144
	5 Strongly Agree	118	248	366
Total		253	423	676

i. P (female) = 423 / 676 = 0.626.

ii. P (somewhat agree or strongly agree) = (144 + 366) / 676 = 510/676 = 0.754.

iii. P (female and strongly disagree) = 29 / 676 = 0.043.

iv. P (strongly or somewhat disagree / female) = (29 + 21) / 423 = 50 / 423 = 0.118.

v. P (strongly or somewhat disagree / male) = (21 + 21) / 253 = 42 / 253 = 0.166.

UED4 Lower Tuition * PV provincial political preference Crosstabulation

Count

		PV provincial political preference				
		1 Liberal	2 NDP	3 Conservative	4 None	Total
UED4 Lower Tuition	1 Strongly Disagree	5	12	7	9	33
	2	6	14	8	9	37
	3	5	12	15	19	51
	4	17	48	22	25	112
	5 Strongly Agree	60	84	44	98	286
Total		93	170	96	160	519

i. P (none) = 160 / 519 = 0.308

ii. P (neutral) = 51 / 519 = 0.098

iii. P (strongly agree / Liberal) = 60 / 93 = 0.645

P (strongly agree / NDP) = 84 / 170 = 0.494

P (strongly agree / Conservative) = 44 / 96 = 0.458

P (strongly agree / None) = 98 / 160 = 0.612

From the probabilities in the first table, females express less disagreement with UED4 than do males. The question is whether university education should be made more accessible by lowering tuition, so females are less likely to disagree with this than are males. Although it would be necessary to check the agree probabilities for each group, the evidence here is that females more strongly support lowering tuition.

From the second table, where there is a direct examination of the extent of strong agreement with lowering tuition, it is the Liberal supporters who are most strongly in favour, followed by those who support no political party. It is the NDP and Conservative supporters who are less likely to strongly agree.