Social Studies 201

Winter 2004

Answers to Problem Set No. 4

March 15, 2004

1. From the table of the standardized normal distribution:

   (a) The area between $Z = 0$ and $Z = 1.25$ is 0.3944 – this is the A area corresponding to $Z = 1.25$.

   (b) The area between $Z = +0.5$ and $Z = +2.5$. This is the A area from the centre to $Z = 2.5$, an area of 0.4938, minus the A area between the centre and $Z = 0.5$, or 0.1915. The area is $0.4938 - 0.1915 = 0.3023$.

   (c) The proportion of cases between $Z = -1.8$ and $Z = +2.5$ is the A area between centre and $Z = -1.8$, 0.4641, plus the A area between centre and $Z = 2.5$, which is 0.4938 from the previous part. The total area is is $0.4641 + 0.4938 = 0.9579$.

   (d) The proportion of cases to the left of $Z = -1.33$ is the B area associated with $Z = -1.33$ This is 0.0918. The percentage of cases is this proportion times 100 or $0.0918 \times 100 = 9.18\%$.

   (e) The area under the normal curve above $Z = -1.83$ is the A area between $Z = -1.83$ and 0, or 0.4664, plus the one-half of the area to the right of centre. The total area is thus $0.4664 + 0.5000 = 0.9644$.

   (f) The percentage of cases that are within one and half standard deviations of the mean is the A area between the centre and $Z = -1.5$ plus the area between the centre and $Z = 1.5$. The latter is 0.4332 and, by symmetry of the distribution, the area for the former is the same. The total area is $0.4332 + 0.4332 = 0.8664$.

   (g) The 75th percentile is at $Z = 0.67$ or 0.68. That is, the 75th percentile is the Z such that 0.75 of the area is to the left of this. There is 0.5000 of the area to the left of $Z = 0$ and another 0.2500 of the area (to reach a total of 0.7500 of the area) when $Z = 0.67$ or $Z = 0.68$, the two Z values with A areas as close as possible to

0.2500. This leaves 0.2500 of the area to the right of this, so the answer is a Z of either 0.67 or 0.68.

(h) The Z-values so that 0.035 of the area is in each tail of the distribution are $Z = -1.81$ and $Z = +1.81$. These are the Z values associated with a B area of 0.0351, very close to 0.0350 in each tail of the distribution.

(i) The fifteenth percentile would be at a Z such that the B area to the left of this is exactly 0.1500. Looking for a B area as close as possible to 0.1500 gives a Z of 1.03 or 1.04. Since this is to be in the left tail of the distribution, $Z = -1.03$ or $Z = -1.04$.

(j) These are the values of Z such that 5% of the cases are in the tail, that is, corresponding to a B area of 0.05. These Z-values are at 1.64 or 1.65 and we usually use 1.645 for this characteristic. So the Z-values for the trim points are at $Z = -1.645$ and $Z = +1.645$.

2. For this question, $\mu = 50$ and $\sigma = 35$ thousand dollars and the question states that the distribution of incomes is assumed to be normal.

(a) The proportion of households below $10,000 is the area under this normal curve to the left of $X = 10$ thousand dollars for males. For this value of income, $Z = (X - \mu)/\sigma = (10 - 50)/35 = -40/35 = -1.14$. The required proportion is the B area to the left of $Z = -1.14$ or 0.1271.

(b) For percentage of households above $100,000, $X = 100$ and $Z = (X - \mu)/\sigma = (100 - 50)/35 = 50/35 = 1.43$. The area to the right of this is the B area, or 0.0764. The percentage of households is this proportion times 100, or $0.0764 \times 100 = 7.64\%$.

(c) The area between 30 and 80 thousand dollars is the area from 30 to 50 (the centre) plus the area from 50 to 80. The Z-values for 30 and 80 are $Z = (X - \mu)/\sigma = (30 - 50)/35 = -20/35 = -0.57$ and $Z = (X - \mu)/\sigma = (80 - 50)/35 = 30/35 = 0.86$. The A areas corresponding to these are 0.2157 for $Z = -0.57$ and 0.3051 for $Z = 0.86$. The area between these Z-values is $0.2157 + 0.3051 = 0.5208$. This is the proportion of households between 30 and 80 thousand dollars.

(d) For ten and thirty thousand dollars, the Z-values are $Z = (X - \mu)/\sigma = (10 - 50)35 = -40/35 = -1.14$ and, from the previous part, $Z = -0.57$. The A areas corresponding to these Z-values are 0.3729 and 0.2157. The proportion of households is the difference between these or $0.3729 - 0.2157 = 0.1572$.

(e) In order to compare the two distributions, Table 1 provides the per cent of cases in the normal distribution for each of the intervals used in parts a. through d. From Table 1 of the problem set, the intervals have been grouped together to match the results for the intervals in a. through d.

Table 1: Comparison of areas of household income for Saskatchewan, 2000, and normal distribution

Household Income in Thousands of Dollars	Per cent from normal distribution	Per cent of households
less than 10	12.7	7.1
10 - 30	15.7	27.3
30 - 80	52.2	48.5
80-100	11.8	8.0
100 plus	7.6	9.1
Total	100.0	100.0

From Table 1, it can be seen that the actual distribution for 2000 follows the general shape of a normal curve, with many cases in the middle and fewer cases nearer the tails of the distribution. But there are also some major differences. The largest difference is at 10-30 thousand dollars where there are over one-quarter of households (27.3%) as opposed to only 15.7% that would be in that interval if it were exactly normally distributed. For the lowest end of the distribution (less than 10 thousand dollars) there are only 7.1% of households, whereas there would be 12.7% of households if incomes were exactly normally distributed. At the middle and upper end, thre are fewer households between 30 and 100 thousand dollars than in the normal distribution and more households at 100 thousand plus than would be indicated by a normal distribution.

In summary, such differences are typical of income distributions – there are relatively few individuals or households at the lowest income levels but more than what might be expected at the lower-middle income levels (in this case 10-30 thousand dollars). At the upper end, while there are not all that many individuals or households with really high incomes, these incomes can be very large so the percentage at this highest level may be greater than in the case of a normal distribution.

3. **Television and internet use**

(a) (i) **Watching television**. Let μ be the true mean weekly hours of watching television for all Saskatchewan respondents aged 15-24. The sample mean is $\bar{X}$ and, with large sample size $n = 180$, the distribution of $\bar{X}$ is a normal distribution with mean μ and standard deviation $s/\sqrt{n}$ (by the Central Limit Theorem and with large sample size, s is an acceptable estimate of σ, the standard deviation of the population from which the sample is drawn). For the 92% interval estimate, $Z = 1.75$ – at this value of Z, there is $92/2 = 46\%$ or 0.46 of the area in each half of the centre of the normal distribution and $Z = 1.75$ is associated with an A area of 0.4599, very close to 0.4600. The 92% interval estimate is

$$\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 13.46 \pm 1.75\ \frac{10.18}{\sqrt{180}} \\
&= 13.46 \pm 1.75\ \frac{10.18}{13.416} \\
&= 13.46 \pm 1.75 \times 0.759 \\
&= 13.46 \pm 1.328
\end{aligned}$$

or, to two decimals, (12.13 , 14.79) hours per week.

(ii) **Using internet at home**. Let μ be the true mean weekly hours of use of the internet at home for all Saskatchewan respondents aged 15-24. Since $n = 63$ is large (greater than 30), the method is the same for this interval as for (i) and the 92% interval is

$$\begin{aligned}\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 9.03 \pm 1.75\ \frac{9.61}{\sqrt{63}} \\ &= 9.03 \pm 1.75\ \frac{9.61}{7.937} \\ &= 9.03 \pm 1.75 \times 1.211 \\ &= 9.03 \pm 2.119\end{aligned}$$

or, to two decimals, (6.91 , 11.15) hours per week.

(iii) **Using internet at school**. Let μ be the true mean weekly hours of use of the internet at school for all Saskatchewan respondents aged 15-24. Since $n = 45$ is large (greater than 30), the method is again the same for this interval as for (i) and (ii). The 92% interval is

$$\begin{aligned}\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 3.98 \pm 1.75\ \frac{3.89}{\sqrt{45}} \\ &= 3.98 \pm 1.75\ \frac{3.89}{6.708} \\ &= 3.98 \pm 1.75 \times 0.580 \\ &= 3.98 \pm 1.01\end{aligned}$$

or, to two decimals, (2.97 , 4.99) hours per week.

(b) **Internet at work**. Let μ be the true mean weekly hours of using the internet at work for all Saskatchewan respondents aged 15-24. The sample mean is $\bar{X}$ and, with a small sample size of only $n = 15$, the distribution of $\bar{X}$ cannot be regarded as a normal distribution. But if we assume that the distribution of hours of internet use at work is a normally distributed variable, then $\bar{X}$ has a t-distribution with mean mean μ, standard deviation $s/\sqrt{n}$, and $n-1 = 15-14 = 14$ degrees of freedom. For the 90% interval estimate for a t-distribution with 14 degrees of freedom, $t = 1.761$. The 90% interval estimate is

$$\bar{X} \pm t\frac{\sigma}{\sqrt{n}} = 7.60 \pm 1.761\ \frac{10.80}{\sqrt{15}}$$

$$\begin{aligned} &= 7.60 \pm 1.761 \, \frac{10.80}{3.873} \\ &= 7.60 \pm 1.761 \times 2.789 \\ &= 7.60 \pm 4.911 \end{aligned}$$

or, to two decimals, (2.69 , 12.51) hours per week.

(c) The formula for determining the required sample size is

$$n = \left(\frac{Z\sigma}{E} \right)^2 .$$

$Z = 2.33$ for 98% confidence, the accuracy of being correct to within one hour gives $E = 1$, and s can be used as an estimate for σ. Since there are three different values of the standard deviation from the three different samples, it is necessary to decide which of these to use. In order to ensure that a large enough sample size is obtained, it is probably best to use the largest of the three standard deviations, that is, $s = 10.80$ for use of internet at work. Using these values,

$$n = \left(\frac{2.33 \times 10.80}{1} \right)^2 = 25.164^2 = 633.2.$$

A random sample of $n = 634$ undergraduates would provide the required precision.

(d) From Table 2 of the problem set, the mean weekly hours watching television was 13.46 and the mean weekly hours using the internet at work was 7.60, so neither of these is real close to 11 hours per week and the analyst is incorrect concerning the sample values. But the means in Table 2 of the problem set are only from samples and it is possible that the true value could still be 11 hours for each of these two activities. Looking at the interval estimate in part (a), the interval runs from 12.1 to 14.8 hours per week for watching television. Since there is a reasonably high confidence (92%) that the intervals from a sample of size $n = 180$ contain the true mean, and since the hypothesized mean of 11 is not in this interval, the researcher is likely incorrect concerning the statement about television use. The sample of size 180 is reasonably large and if the sample is really a random sample, we can be quite confident that the true mean is somewhat greater than 11.

For internet use at work, by just looking at the mean of 7.60, one might be tempted to conclude that the mean is not 11 hours. But the interval in part (b) is quite wide, running from 2.7 hours to 12.5 hours, which includes 11 hours. The interval is very wide because the sample size of $n = 15$ is quite small. The small sample size does not produce a sample mean that provides a very precise estimate of the true mean. As a result, it is possible that the researcher is correct in stating that the mean weekly hours of internet use at work is 11.

4. Problems using data from t:\students\public\201\ssae98.sav

a. i. Study hours

For study hours, the mean is 16.6 and the standard deviation is 11.8, so the interval within one standard deviation of the mean is from 16.6 – 11.8 = 4.8 to 16.6 + 11.8 = 28.4. Examining the frequency distribution, to see the frequencies from 4.8 to 28.4 gives 1 + 48 + 27 + 16 + 29+ 8 + …. + 3 + 1 = 513 cases, out of a total of 668. This is (513/668) x 100% = 76.8% of the cases.

Two standard deviations is 2 x 11.85 = 23.70, so that the mean plus or minus two standard deviations is 16.6 – 23.70 to 16.6 + 23.70. Since 16.6-23.70 produces a value less than 0, and since study hours cannot be less than zero, such an interval is from 0 to 16.6 + 23.70 = 40.3. This includes all the cases up to at least 40, and from the cumulative per cent, this is 96.7 per cent of all cases.

For the normal distribution there are 0.3413 + 0.3413 = 0.6826, or 68.26% of the cases within one standard deviation of the mean. That is, for a Z of 1, the A area from the centre to one standard deviation to the right of centre is 0.3413. By symmetry, the area between centre and one standard deviation to the left of centre is 0.3413.

For Z=2, or two standard deviations, the A area is 0.4772. Twice this is 0.9544, the area within two standard deviations of the mean.

From the above, the percentage of undergraduates within two standard deviations of the mean is 96.7%, as opposed to the theoretical per cent of 95.44 per cent of a normal distribution. So the actual distribution has a slightly larger per cent of cases within two standard deviations than does the normal distribution. Within one standard deviation of the mean, the normal distribution has 68.26% of cases while the distribution of study hours has 76.8% of cases. So the actual distribution is a little more concentrated closer to the mean than is the normal distribution.

The diagram of the normal distribution superimposed on the histogram shows that there are more respondents reporting study hours of from about 5 to 15 than would be expected in a normal distribution. The other bars fall a bit short of the normal distribution, indicating fewer cases than expected from a normal distribution.

The general shape of the histogram and normal distribution is the same but the bars of the diagram indicate the actual distribution of study hours is a little more concentrated around the mean than is the normal distribution, as confirmed by the earlier calculations. Of course, there are also a few cases of very large study hours, fifty or more, and this is more than what would be expected from the normal distribution. But the total number of respondents who report these high study hours is very small.

STHOURS Study Hours

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2	.3	.3	.3
	1	4	.6	.6	.9
	2	12	1.7	1.8	2.7
	3	19	2.7	2.8	5.5
	4	1	.1	.1	5.7
	4	16	2.3	2.4	8.1
	5	1	.1	.1	8.2
	5	48	6.8	7.2	15.4
	6	27	3.8	4.0	19.5
	7	16	2.3	2.4	21.9
	8	29	4.1	4.3	26.2
	9	8	1.1	1.2	27.4
	10	96	13.6	14.4	41.8
	11	1	.1	.1	41.9
	12	24	3.4	3.6	45.5
	13	2	.3	.3	45.8
	13	7	1.0	1.0	46.9
	14	13	1.8	1.9	48.8
	15	61	8.6	9.1	57.9
	16	1	.1	.1	58.1
	16	7	1.0	1.0	59.1
	17	10	1.4	1.5	60.6
	18	3	.4	.4	61.1
	18	9	1.3	1.3	62.4
	19	1	.1	.1	62.6
	20	90	12.7	13.5	76.0
	21	6	.8	.9	76.9
	22	2	.3	.3	77.2
	23	1	.1	.1	77.4
	24	3	.4	.4	77.8
	25	39	5.5	5.8	83.7
	27	3	.4	.4	84.1
	28	1	.1	.1	84.3
	28	4	.6	.6	84.9
	30	43	6.1	6.4	91.3
	33	1	.1	.1	91.5
	35	13	1.8	1.9	93.4
	38	2	.3	.3	93.7
	40	20	2.8	3.0	96.7
	43	2	.3	.3	97.0
	45	4	.6	.6	97.6
	50	8	1.1	1.2	98.8
	55	1	.1	.1	99.0
	60	3	.4	.4	99.4
	65	2	.3	.3	99.7
	70	1	.1	.1	99.9
	100	1	.1	.1	100.0
	Total	668	94.5	100.0	
Missing	996 Other	1	.1		
	997 Uncertain	4	.6		
	999 No response	24	3.4		
	System	10	1.4		
	Total	39	5.5		
Total		707	100.0		

V3 Affirmative Action

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	59	8.3	8.5	8.5
	2	125	17.7	18.0	26.4
	3	267	37.8	38.4	64.8
	4	174	24.6	25.0	89.8
	5 Strongly Agree	71	10.0	10.2	100.0
	Total	696	98.4	100.0	
Missing	9 No Response	11	1.6		
Total		707	100.0		

V4 Gay and Lesbians Married

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly Disagree	141	19.9	20.3	20.3
	2	86	12.2	12.4	32.7
	3	183	25.9	26.3	59.0
	4	161	22.8	23.2	82.2
	5 Strongly Agree	124	17.5	17.8	100.0
	Total	695	98.3	100.0	
Missing	9 No Response	12	1.7		
Total		707	100.0		

a. ii. V3 and V4

From the statistics by the histogram of V3, affirmative action, the mean is 3.1 and the standard deviation is 1.08. In order to determine the proportion of cases between 2.5 and 3.5, it is necessary to calculate the Z-values for each of these. For

$$X=2.5,\ Z = (2.5 - 3.1) / 1.08 = -0.56$$

$$X=3.5,\ Z = (3.5 - 3.1) / 1.08 = 0.37$$

From the normal table, the A areas associated with $Z = -0.56$ is 0.2123 and for $Z = 0.37$ it is 0.1443. The sum of these two is 0.3566, or 35.66%, the percentage of cases that would be expected to be exactly at three, if this distribution were exactly normally distributed. This compares with 38.4% (the valid per cent for an opinion of 3 in the distribution of V3). As confirmed by the normal distribution superimposed on the histogram, the actual percentage and the percentage expected from the normal distribution are not exactly the same, but fairly close. From the diagram, there is a little larger percentage of cases with a neutral opinion of 3 than what would be expected in the case of a normal distribution, as indicated in the above calculations.

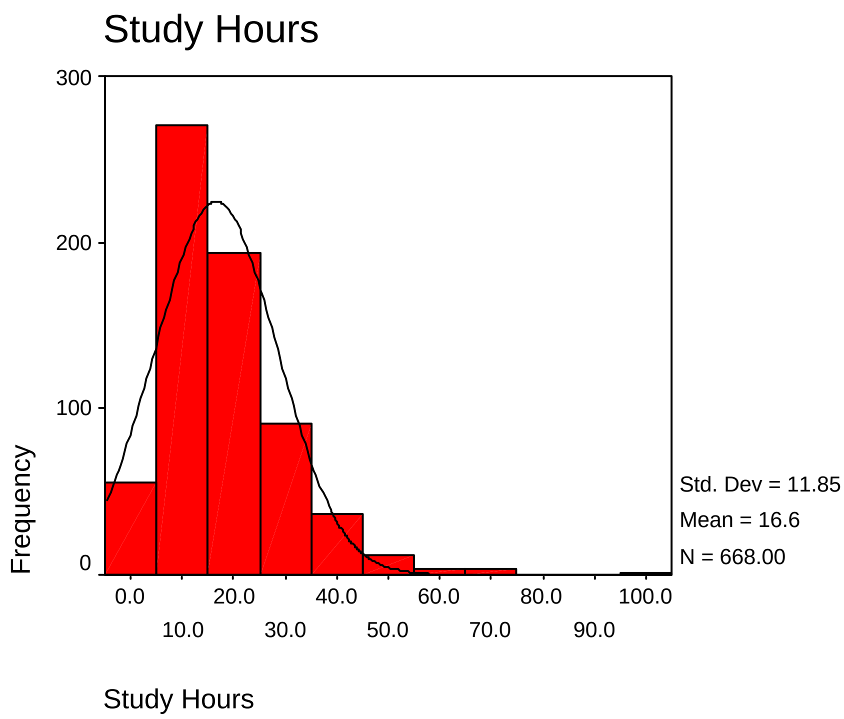
For V4, view with respect to gays and lesbians being recognized as married, the mean is again 3.1 but the standard deviation is larger at s = 1.37. The area from 2.5 to 3.5 in a normal distribution is provided by the following Z-values

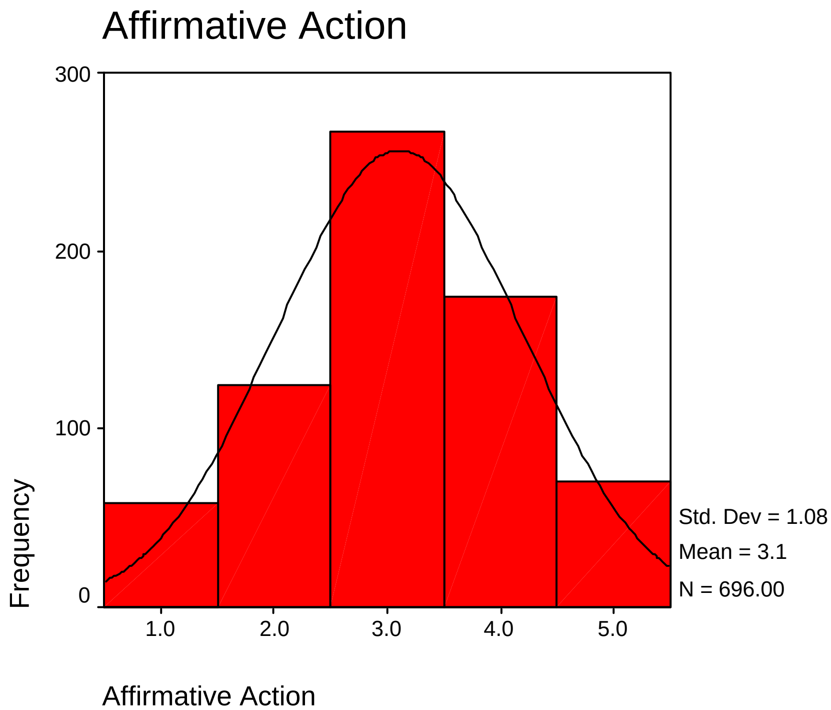
For X=2.5, Z = (2.5 – 3.1) / 1.37 = –0.44 and A area is 0.1700.
For X=3.5, Z = (3.5 – 3.1) / 1.37 = 0.29 and A area is 0.1141.

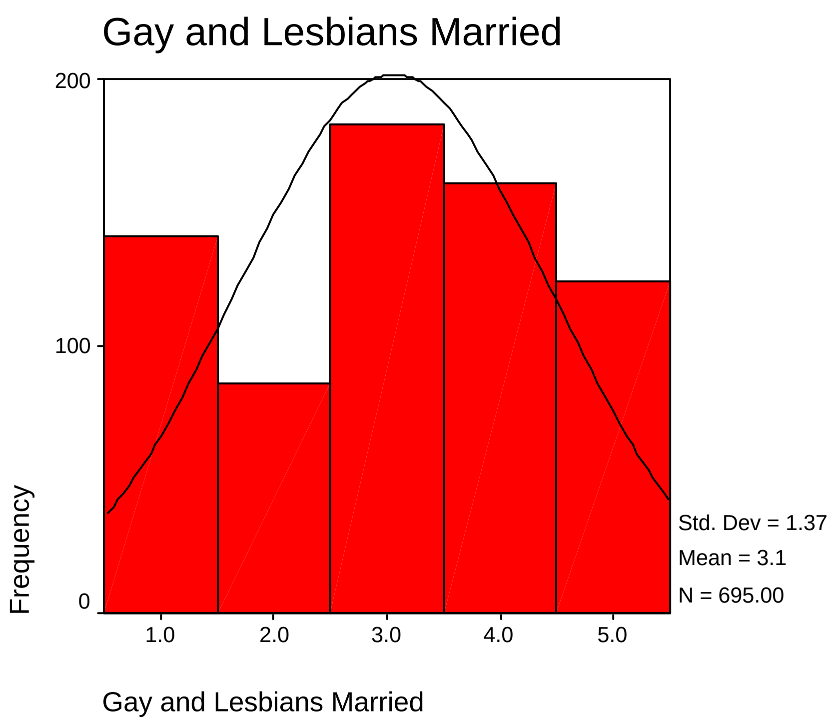
The sum of the two A areas is 0.1700 + 0.1141 = 0.2841 or 28.41%. This compares with 26.3% of cases that expressed the neutral response of 3 in the actual distribution. While the two percentages are close in this distribution as well, for V4 there are fewer respondents who gave the neutral response of 3 than what would be expected had the distribution been exactly normal.

In summary, as compared with a normal distribution of responses, there are more respondents who express the neutral response of 3 for variable V3, but fewer in the case of variable V4. The histograms with normal curves superimposed on them confirm these results.

Histogram




Affirmative Action
300
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Std. Dev = 1.08
Mean = 3.1
N = 696.00
Affirmative Action


Gay and Lesbians Married
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Std. Dev = 1.37
Mean = 3.1
N = 695.00
Gay and Lesbians Married

4.b. Confidence intervals

For the variable measuring weekly study hours, the mean is 16.61, the standard deviation is 11.849, and the sample size is n = 668. For the 90% confidence interval, the Z-value is 1.645. The general formula for an interval estimate is

Mean plus or minus Z times (s / square root of n)

The 90% interval estimate for the true mean study hours for all undergraduates is

16.61 plus or minus 1.645 (11.849/25.846) = 16.61 plus or minus 0.75

The resulting interval is (15.86 , 17.36) and this is almost the same as the 90% interval estimate of 15.85 to 17.36 given in the computer printout. The small difference is likely due to rounding differences.

The standard error of 0.458 for the sample mean is the standard deviation divided by the square root of n, the amount in brackets above. This is 11.849/26.846 = 0.458, the same as on the printout.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
STHOURS Study Hours	668	94.5%	39	5.5%	707	100.0%

Descriptives

			Statistic	Std. Error
STHOURS Study Hours	Mean		16.61	.458
	90% Confidence Interval for Mean	Lower Bound	15.85	
		Upper Bound	17.36	
	5% Trimmed Mean		15.57	
	Median		15.00	
	Variance		140.409	
	Std. Deviation		11.849	
	Minimum		0	
	Maximum		100	
	Range		100	
	Interquartile Range		12.00	
	Skewness		1.689	.095
	Kurtosis		5.212	.189

Descriptives

			Statistic	Std. Error
STHOURS Study Hours	Mean		16.61	.458
	95% Confidence Interval for Mean	Lower Bound	15.71	
		Upper Bound	17.51	
	5% Trimmed Mean		15.57	
	Median		15.00	
	Variance		140.409	
	Std. Deviation		11.849	
	Minimum		0	
	Maximum		100	
	Range		100	
	Interquartile Range		12.00	
	Skewness		1.689	.095
	Kurtosis		5.212	.189

Descriptives

			Statistic	Std. Error
STHOURS Study Hours	Mean		16.61	.458
	99% Confidence Interval for Mean	Lower Bound	15.42	
		Upper Bound	17.79	
	5% Trimmed Mean		15.57	
	Median		15.00	
	Variance		140.409	
	Std. Deviation		11.849	
	Minimum		0	
	Maximum		100	
	Range		100	
	Interquartile Range		12.00	
	Skewness		1.689	.095
	Kurtosis		5.212	.189

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	YEAR Year of Program	N	Percent	N	Percent	N	Percent
DEBT1 Debt before Fall 1998	1 First	177	86.3%	28	13.7%	205	100.0%
	2 Second	118	86.1%	19	13.9%	137	100.0%
	3 Third	158	84.5%	29	15.5%	187	100.0%
	4 Fourth	102	85.7%	17	14.3%	119	100.0%
	5 Fifth or more	50	86.2%	8	13.8%	58	100.0%

Descriptives

	YEAR Year of Program			Statistic	Std. Error
DEBT1 Debt before Fall 1998	1 First	Mean		1061.32	228.671
		80% Confidence Interval for Mean	Lower Bound	767.16	
			Upper Bound	1355.48	
		5% Trimmed Mean		527.71	
		Median		.00	
		Variance		9255439	
		Std. Deviation		3042.275	
		Minimum		0	
		Maximum		25000	
		Range		25000	
		Interquartile Range		.00	
		Skewness		4.387	.183
		Kurtosis		25.073	.363
	2 Second	Mean		2647.46	371.508
		80% Confidence Interval for Mean	Lower Bound	2168.65	
			Upper Bound	3126.27	
		5% Trimmed Mean		2143.13	
		Median		.00	
		Variance		1.6E+07	
		Std. Deviation		4035.605	
		Minimum		0	
		Maximum		17000	
		Range		17000	
		Interquartile Range		5000.00	
		Skewness		1.598	.223
		Kurtosis		1.988	.442
	3 Third	Mean		3193.67	523.277
		80% Confidence Interval for Mean	Lower Bound	2520.23	
			Upper Bound	3867.11	
		5% Trimmed Mean		2203.94	
		Median		.00	
		Variance		4.3E+07	
		Std. Deviation		6577.492	
		Minimum		0	
		Maximum		35000	
		Range		35000	
		Interquartile Range		3000.00	
		Skewness		2.523	.193
		Kurtosis		6.625	.384
	4 Fourth	Mean		5370.59	931.175
		80% Confidence Interval for Mean	Lower Bound	4169.38	
			Upper Bound	6571.79	
		5% Trimmed Mean		4191.72	
		Median		.00	
		Variance		8.8E+07	
		Std. Deviation		9404.408	
		Minimum		0	
		Maximum		35000	
		Range		35000	
		Interquartile Range		7250.00	
		Skewness		1.779	.239
		Kurtosis		2.110	.474
	5 Fifth or more	Mean		6942.48	1476.447
		80% Confidence Interval for Mean	Lower Bound	5024.47	
			Upper Bound	8860.49	
		5% Trimmed Mean		5658.31	
		Median		.00	
		Variance		1.1E+08	
		Std. Deviation		10440.058	
		Minimum		0	
		Maximum		40000	
		Range		40000	
		Interquartile Range		13350.00	
		Skewness		1.648	.337
		Kurtosis		2.362	.662

4. b. ii. Debt level. The results are summarized in the following table.

Sample size, mean, 80% and 99% confidence intervals for mean debt, in dollars

Year	n	Mean	80% interval	99% interval
First	177	1,061	(767, 1,355)	(466, 1,657)
Second	118	2,647	(2,169, 3,126)	(1,675, 3,620)
Third	158	3,194	(2,520, 3,867)	(1,829, 4,558)
Fourth	102	5,371	(4,169, 6,572)	(2,925, 7,815)

From the table, it is apparent that, on average, debt increases by year, from a mean debt of just a little over $1,000 debt in first year to a mean of over $5,000 by fourth year. But these data come from samples of students. The interval estimates provide estimates of the possible range for the true mean debt level for all University of Regina undergraduate students in each of the four undergraduate years, assuming these data are equivalent to random samples from students. The intervals for first year do not overlap much with intervals for the later years, so a researcher could conclude that true mean debt levels are definitely lower for first year students than for students at second or higher years.

For the second and third years, the intervals overlap considerably, so there is not strong evidence that the debt levels are greater for third than second year students. While the larger mean for third, than for second year, students points in the direction of greater debt for third than second year students, the overlap of the intervals, even the 80% intervals, is great, so the true mean debt levels for these two years may not be all that different.

For fourth year students, the mean debt level is considerably higher than for second or third year students. The 80% interval for fourth year students does not overlap with those from other years, so mean debt for all fourth year students is very likely to be greater for fourth year students. However, the 99% interval does overlap with those for second and third year students, so if a researcher wants to be very sure that mean debt for fourth year students is greater, this conclusion cannot really be made here, but fourth year mean debt definitely seems greater than the mean level for all first year students.

In terms of the interval widths, there are three factors that govern this – confidence level, sample size, and standard deviation of the group. For the 99% interval for fourth year students, the interval is from $2,925 to $7,815, a width of almost $5,000. In contrast, the 80% interval for first year students is from $767 to $1,355, a width of approximately $600. The reasons are as follows:

- The 80% interval has a much smaller Z-value of 1.28 or 1.29, as compared with a Z-value of 2.575 for 99% confidence. This helps produce a narrower interval for the 80% level.

- The sample size is also larger for the first year students, at n=177 as opposed to only n=102 for fourth year. This again helps produce a narrower interval for first year students.

- Finally, the standard deviation is larger for fourth year students, at $9,404, much larger than the standard deviation of $3,042 for first year. This also helps produce a wider interval for fourth year.

From consideration of all three of the sample size, confidence level, and standard deviation, the combined effect is to produce a wider interval for the true mean debt level of all fourth year students as compared with that for all first year students.

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	YEAR Year of Program	N	Percent	N	Percent	N	Percent
DEBT1 Debt before Fall 1998	1 First	177	86.3%	28	13.7%	205	100.0%
	2 Second	118	86.1%	19	13.9%	137	100.0%
	3 Third	158	84.5%	29	15.5%	187	100.0%
	4 Fourth	102	85.7%	17	14.3%	119	100.0%
	5 Fifth or more	50	86.2%	8	13.8%	58	100.0%

Descriptives

	YEAR Year of Program			Statistic	Std. Error
DEBT1 Debt before Fall 1998	1 First	Mean		1061.32	228.671
		99% Confidence Interval for Mean	Lower Bound	465.85	
			Upper Bound	1656.79	
		5% Trimmed Mean		527.71	
		Median		.00	
		Variance		9255439	
		Std. Deviation		3042.275	
		Minimum		0	
		Maximum		25000	
		Range		25000	
		Interquartile Range		.00	
		Skewness		4.387	.183
		Kurtosis		25.073	.363
	2 Second	Mean		2647.46	371.508
		99% Confidence Interval for Mean	Lower Bound	1674.66	
			Upper Bound	3620.25	
		5% Trimmed Mean		2143.13	
		Median		.00	
		Variance		1.6E+07	
		Std. Deviation		4035.605	
		Minimum		0	
		Maximum		17000	
		Range		17000	
		Interquartile Range		5000.00	
		Skewness		1.598	.223
		Kurtosis		1.988	.442
	3 Third	Mean		3193.67	523.277
		99% Confidence Interval for Mean	Lower Bound	1829.22	
			Upper Bound	4558.12	
		5% Trimmed Mean		2203.94	
		Median		.00	
		Variance		4.3E+07	
		Std. Deviation		6577.492	
		Minimum		0	
		Maximum		35000	
		Range		35000	
		Interquartile Range		3000.00	
		Skewness		2.523	.193
		Kurtosis		6.625	.384
	4 Fourth	Mean		5370.59	931.175
		99% Confidence Interval for Mean	Lower Bound	2925.89	
			Upper Bound	7815.28	
		5% Trimmed Mean		4191.72	
		Median		.00	
		Variance		8.8E+07	
		Std. Deviation		9404.408	
		Minimum		0	
		Maximum		35000	
		Range		35000	
		Interquartile Range		7250.00	
		Skewness		1.779	.239
		Kurtosis		2.110	.474
	5 Fifth or more	Mean		6942.48	1476.447
		99% Confidence Interval for Mean	Lower Bound	2985.67	
			Upper Bound	10899.29	
		5% Trimmed Mean		5658.31	
		Median		.00	
		Variance		1.1E+08	
		Std. Deviation		10440.058	
		Minimum		0	
		Maximum		40000	
		Range		40000	
		Interquartile Range		13350.00	
		Skewness		1.648	.337
		Kurtosis		2.362	.662