Social Studies 201

Answers to Problem Set 4

November 15, 2006

1. **Number of close friends by age**

(a) i. The probability of selecting a person of age 45-54 is 0.355.

$$P(\text{age 45-54}) = \frac{N(45-54)}{N} = \frac{173}{487} = 0.355$$

ii. The probability of selecting a person withless than 6 close friends is 0.581.

$$P(\text{less than 6}) = \frac{N(<6)}{N} = \frac{28+83+172}{487} = \frac{283}{487} = 0.581$$

iii. The probability of selecting a person with 11 or more close friends and of age 75 plus is 0.057.

$$P(11+ \text{ and } 75+) = \frac{19+9}{487} = \frac{28}{487} = 0.057$$

iv. This is a conditional probability of selecting a person aged 75 plus, given that the person has more than 25 close friends. The probability is 0.529.

$$P(\text{75 plus/more than 25 close friends}) = \frac{9}{17} = 0.529$$

v. This is another conditional probability, in this case what is given is that the person is age 75 plus. Of these 134 persons of age 75 plus, there are $19 + 9 = 28$ who have 11 or more close friends. So the probability is 0.209.

$$P(11+/75+) = \frac{28}{134} = 0.209$$

(b) i. In order to determine whether the two events of being age 45-54 and having no close friends are independent of each other or not, it is necessary to check whether the conditional probability of 45-54 given event of None, and the overall probability of event 45-54 are equal or not. These probabilities are

$$P(45-54/\text{None}) = \frac{10}{28} = 0.357$$

$$P(45-54) = \frac{173}{487} = 0.355$$

These two probabilities differ by only a small amount and, to two decimal places, are equal at 0.36. For practical purposes these two events are independent of each other.
An alternative way to check this is to see whether the probability of event None given event 45-54, equals the probability of event None. These probabilities are

$$P(\text{None}/45-54) = \frac{10}{173} = 0.058$$

$$P(\text{None}) = \frac{28}{487} = 0.057$$

These two probabilities differ so little that the two events can be considered to be independent of each other.

ii. Again it is necessary to check whether the conditional and overall probabilities are equal or not. This can be done in either of the following ways.

$$P(2 \text{ or less}/75 \text{ plus}) = \frac{(16+22)}{134} = \frac{38}{134} = 0.284$$

$$P(2 \text{ or less}) = \frac{(28+83)}{487} = \frac{111}{487} = 0.228$$

Or

$$P(75 \text{ plus}/2 \text{ or less}) = \frac{(16+22)}{(28+83)} = \frac{38}{111} = 0.342$$

$$P(75 \text{ plus}) = \frac{(134)}{487} = 0.275$$

In each case, the conditional probabilities differ from the overall probability, demonstrating that the events of having 2 or less close friends and the event of being age 75 plus are dependent on each other. From the first pair, if the individual selected is age 75 plus, this means they are more likely to have 2 or less close friends (0.284), as compared with the overall probability of having 2 or less close friends (0.228).

2. **Independence and dependence**

The independent events are those of (A) working outside the home and (B) the amount of time middle-aged women spend providing care. According to the last sentence of the quote in part a., A does not affect B. That is, whether or not the woman works outside the home does not affect the time these women spend providing care. This could be put in the form of conditional and overall probabilities. Let the event of any specific amount of time the women spend providing care be B. Let the event of working outside the home be A. Then

$$P(B/A) = P(B)$$

There are two possible sets of dependent events. In the first quote, the event of being a woman (A) and the event of spending more time at being a caregiver are dependent (B). That is, women spend more time being caregivers than men, so if a woman is selected, the chance is that she spends more time at caregiving. That is, this is greater than a randomly selected individual, who could be either male or female or $P(B/A) > P(B)$]. Similarly, the event of being male and the event of time spent at caregiving are dependent events.

In the second quote, there is also a claim of dependence. In this case, the event of becoming a citizen and the event of being in Canada for any specific length of time are dependent events. The greater the amount of time spent in Canada, the greater the probability that the newcomer becomes a citizen.

3. **Standardized normal distribution**

   (a) For $Z = 2.12$, the A area is 0.4830 and this is the area between $Z = 0$ and $Z = 2.12$. By symmetry, the area between 0 and -2.12 is 0.4830.

   (b) The area between 1.3 and 2.3 is the area between the centre and $Z = 2.30$ (an A area of 0.4893), minus the area between the centre of the normal distribution and $Z = 1.30$ (A area of 0.4032). The area between $Z = 1.30$ and $Z = 2.30$ is thus $0.4893 - 0.4032 = 0.0861$.

   (c) The proportion of cases between $Z = -0.97$ and $Z = 1.34$ is the sum of the areas between the centre of the distribution and each of these Z values. Between $Z = 0$ and $Z = -0.97$, the area is 0.3340 while the area between $Z = 0$ and $Z = 1.34$ is 0.4099. The sum of these two areas is $0.3340 + 0.4099 = 0.7439$. As a percentage of the total area, this is $0.7439 \times 100 = 74.39\%$.

   (d) The area above $Z = -1.25$ is the A area associated with $Z = -1.25$ (the same as the A area between the centre and $+1.25$) plus the one-half of the area to the right 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 such that $Z < 1.84$ is the A area between a Z of 0 and $Z = 1.84$ plus the one-half of the area to the left of $Z = 0$. The A area for $Z = 1.84$ is 0.4671 so the total required area is $0.4671 + 0.5000 = 0.9671$.

   (f) To determine this, remember that the standard deviation for Z is 1, so that one-half standard deviation is at a Z of 0.5. The required area is the area under the curve between $Z = -0.5$ and $Z = 0.5$. The A area associated with $Z = 0.50$ is 0.1915. Since the required area is that within $Z = \pm 0.50$, this is $0.1915 + 0.1915 = 0.3830$. The percentage of the population within one-half standard deviation is thus $0.3830 \times 100\% = 38.3\%$.

   (g) There is 0.5000 or 50% of the area to the left of centre, so it is necessary to go to the right of centre by $0.6300 - 0.5000 = 0.1300$ or $63 - 50 = 13$ per cent to get to this Z value. Looking for an A area as close as possible to 0.1300 gives a Z of either 0.33 (A area of 0.1293, just below 0.1300) or 0.34 (A area of 0.1331, just above 0.1300). $Z = 0.33$ is associated with an area closest to 0.1300 so the required Z is 0.33 to obtain the 63rd percentile. That is $P_{63} = 0.33$ in the standardized normal distribution.

(h) For an area of 0.0325 in a tail of the distribution, the B area of 0.0325 is associated with a $Z = \pm 1.85$. Outside the Z values of plus and minus 1.85, there is 0.0325 of the distribution in each tail, for a total of 0.065.

(i) The thirty-sixth percentile is to the left of centre, with 36% or 0.3600 of the area below this. This is a B area, and when $B = 0.3600$, $Z = 0.36$ (a little closer than $Z = 0.37$). So the 36th percentile is at $Z = -0.36$, since this is to the left of centre.

(j) The seventy per cent range is the Z-values so that there is a total of 30% or 0.3000 in the two tails of the distribution, or 0.1500 in each tail. For a B area of 0.1500, $Z = 1.04$ or 1.03. The seventy per cent range is thus from $Z = -1.04$ to $Z = +1.04$.

4. **Distribution of income**

(a) For this question, X is the variable representing income in thousands of dollars, the mean $\mu = 33.0$ and the standard deviation $\sigma = 26.4$. The distribution of income is assumed to have a normal distribution for this part of the question.

i. The For $X = 10$, $Z = (X - \mu)/\sigma = (10 - 33.0)/26.4 = -23.0/26.4 = -0.87$. The proportion with less than 10 thousand dollars of income is the B area to the left of $Z = -0.87$ and for this Z, the B area is 0.1921. This represents the proportion of persons with incomes of less than ten thousand dollars.

ii. For the percentage of persons with incomes below \$50,000, find the Z associated with $X = 50$ and the required percentage is the area under the normal curve to the left of this (multiplied by 100%). For $X = 50$,

$$Z = (X - \mu)/\sigma = (50 - 33.0)/26.4 = 17.0/26.4 = 0.64$$

The area between $Z = 0$ and $Z = 0.64$ is an A area of 0.2389. To this must be added the 0.5000 of the area to the left of centre. There are thus $0.5000 + 0.2389 = 0.7389$, or $0.7389 \times 100\% = 73.89\%$ of the persons with incomes of under fifty thousand dollars.

iii. The proportion between 50 and 70 thousand can be found by determining the Z for each of these two incomes, finding the A area for each of these Z values, and subtracting them. For $X = 50$, $Z = 0.64$ from ii. For $X = 70$,

$$Z = (X - \mu)/\sigma = (70 - 33.0)/26.4 = 37.0/26.4 = 1.40$$

and the associated A area is 0.4192.
The required area is the A area between the centre and $Z = 1.40$ (or $X = 70$) minus the A area between the centre and $Z = 0.64$ ($X = 50$) from ii. This is $0.4192 - 0.2389 = 0.1803$. There are thus $0.1803 \times 100\% = 18.03\%$ of persons with incomes of fifty to seventy thousand dollars if income is normally distributed.

iv. The fifty fourth percentile of income occurs to the right of centre, so that $0.5400 - 0.5000 = 0.0400$ of the area is to the right of centre. That is, for an A area of 0.0400, $Z = 0.10$. Converting this Z value into income using the transformation from Z to X, this gives the X values

$$X = \mu + Z\sigma = 33.0 + (0.10 \times 26.4) = 33.0 + 2.64 = 35.64$$

thousand dollars. If incomes are normally distributed, then $P_{54} = 35.6$ thousand dollars.

(b) Comparison of the actual and normal distribution. From the diagram of the normal distribution superimposed on the histogram of income, it is apparent that the actual distribution has some similarities to the normal distribution, but also some differences. From Figure 1 of Problem Set 4, there are many more persons with incomes between 10 and 30 thousand dollars (the tall bar) than there would be if the bar exactly fitted the normal distribution. It can also known that the normal distribution extends to infinity in each direction, whereas the actual income distribution stops at zero income on the low side. The normal distribution would extend to negative incomes, so there would be more people at the lowest incomes than there actually are at the lowest incomes, if incomes were normally distributed. At the upper end, there are more people with the highest incomes, above one hundred thousand dollars, than there would be if incomes were normally distributed. The group that appears to come closest to normal is the 70-90 thousand income bracket, where the normal curve crosses the vertical bar of the histogram near the centre of the bar.

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 Problem Set 4.

Years	Actual %	Normal %
Less than 10	16.9%	19.2%
Under 50	78.8%	73.9%
50-70	11.6%	18.0%
54th percentile	\$30,000	\$35,600

From the table, there is a smaller percentage of persons with incomes of under 10 thousand dollars and a greater percentage with income under 50 thousand than in the case of a normal distribution. For the 50-70 thousand income bracket, there is a much smaller percentage of persons than in the case of a normal distribution – the same can be observed in Figure 1, where the bar for this income level falls well below the height of the normal curve. Finally, the actual income distribution reaches the 54th percentile at \$30,000, well before the normal curve reaches the 54th percenitle at \$35,600. This happens because of the very large number of persons in the 30-50 thousand income interval, as compared with the normal distribution.

In summary, there are some parallels between the actual distribution of income and the normal distribution, but the actual distribution differs from the normal at the lower, some of the middle, and upper income levels.

5. **Incomes of employed females**

(a) For each age group, let μ be the true mean income of females in thousands of dollars. The data in Table 1 give the point estimates of μ for each age group.

Table 1: Income statistics, Saskatchewan females, 2003

Age group	$\bar{X}$	s	n
25-34	33.3	13.5	55
35-44	40.3	20.7	57
45-54	45.1	24.1	37
55-64	40.1	25.9	31

Since the sample size for each age group exceed 30, by the Central Limit Theorem the sample mean $\bar{X}$ is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$, where σ is the true standard deviation for each age group and n is the sample size. Since the values of σ are not known for the different age groups, the sample standard deviations s are used as estimates. For each age group, the interval estimates are of the form:

$$\bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

Since the question asks for the 90 per cent interval estimates, the Z value is 1.645. That is, there is 90% or 0.9000 of the area in the centre of the normal distribution, so that there is 0.4500 of the area in each half of the normal curve. Looking through the A areas of the standardized normal distribution, the corresponding Z value is either 1.64 or 1.65. In this case, the A area of 0.4500 is exactly midway between the two areas of 0.4495 and 0.4505, so $Z = 1.645$.

For the 25-34 age group, the 90 per cent interval estimate is

$$\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= \bar{X} \pm 1.645\frac{13.5}{\sqrt{55}} \\
&= \bar{X} \pm 1.645\frac{13.5}{7.416} \\
&= \bar{X} \pm (1.645 \times 1.820) \\
&= \hat{X} \pm 2.994 \\
&= 33.3 \pm 3.0
\end{aligned}$$

or from 30.3 to 36.3 thousand dollars (rounded to the nearest tenth of a thousand dollars). The interval estimate can be written as (30.3 , 36.3) thousand dollars.

Using the means and standard deviations in Table 1, the 90% interval estimates are summarized in Table 2.

(b) Table 2 provides a summary of the interval estimates for the mean income of Saskatchewan females in the four age groups. For each interval estimate, the same $Z = 1.645$ is used.

Table 2: 95% interval estimates of incomes of Saskatchewan females of various age 2003

Age group	Sample mean	Interval
25-34	33.3	(30.3 , 36.3)
35-44	40.3	(35.8 , 44.8)
45-54	45.1	(38.6 , 51.6)
55-64	40.1	(32.4 , 47.8)

From Table 1, the pattern of sample mean incomes is that the youngest have te lowest income (33.3) and then incomes increase with age through the 45-54 group (to 45.1), but then income declines for the oldest age group. However, the sample sizes and sample standard deviations also differ among the age groups, so that the interval estimates are wider for some age groups than for others. As a result, several of the intervals overlap, meaning that the true means may not be all that different for the four groups. For example, for the 35-44 group, the sample mean is 40.3 but the true mean could be anywhere between 35.8 and 44.8, at 90% confidence. For the 45-54 group, the sample mean is 45.1, considerably greater than for the 35-44 age group. But the interval estimate for the 45-54 group is from 38.6 to 51.6, meaning that the true mean for this group could be anywhere in this interval. But these two intervals overlap a lot, so the true mean could be very lettle different. The same is true for each of the last three age groups. The problem here is the small sample sizes, which do not lead to very precise estimates of the mean. While the pattern of mean income increasing by age, and then decreasing, noted in Table 1 may hold true for all Saskatchewan females, the wide interval estimates cast some doubt on this.

6. Computer problems

a. (i) Means of respect for government classified by provincial political preference

Report

Respect for Governments

provincial political	Mean	N	Std. Deviation
Liberal	3.13	97	.942
NDP	3.24	174	.844
Conservative	3.09	98	1.006
None	2.60	165	1.046
Total	2.99	534	.992

Crosstabs with column percentages

Respect for Governments * provincial political preference Crosstabulation

			provincial political preference				
			Liberal	NDP	Conservative	None	Total
Respect for Governments	No Respect	Count	3	1	6	33	43
		% within provincial political preference	3.1%	.6%	6.1%	20.0%	8.1%
	2	Count	23	32	20	35	110
		% within provincial political preference	23.7%	18.4%	20.4%	21.2%	20.6%
	3	Count	34	77	38	64	213
		% within provincial political preference	35.1%	44.3%	38.8%	38.8%	39.9%
	4	Count	32	53	27	31	143
		% within provincial political preference	33.0%	30.5%	27.6%	18.8%	26.8%
	Great Respect	Count	5	11	7	2	25
		% within provincial political preference	5.2%	6.3%	7.1%	1.2%	4.7%
Total		Count	97	174	98	165	534
		% within provincial political preference	100.0%	100.0%	100.0%	100.0%	100.0%

(ii) The pattern is most apparent from the Means procedure, where those who support the NDP demonstrate the greatest respect for governments (mean of 3.24) and those who support no political party show the least respect (mean of 2.60). Those who support the Liberals and Conservatives are in the middle (with means of 3.13 and 3.09, respectively). From the crosstabulation table, the picture is the same. For example, many of those who support no party showing no respect (20% at level 1, a much larger percentage than for supporters of any of the parties).

P(Conservative) = 98 / 534 = 0.184

P(4 or 5) = (143 + 25) / 534 = 168 / 534 = 0.315

P(None and no respect) = 33 / 534 = 0.062

P(No respect / Liberal) = 3 / 97 = 0.031

P(No respect / None) = 33 / 165 = 0.200

The latter two conditional probabilities are consistent with the means shown earlier in that those who support none of the parties showed the least respect (mean of 2.6 and a high probability of no respect at 0.200). In contrast, for the Liberals, the mean shows a higher level of respect (3.13), or a lower probability of no respect (0.031).

b. (i)

Statistics

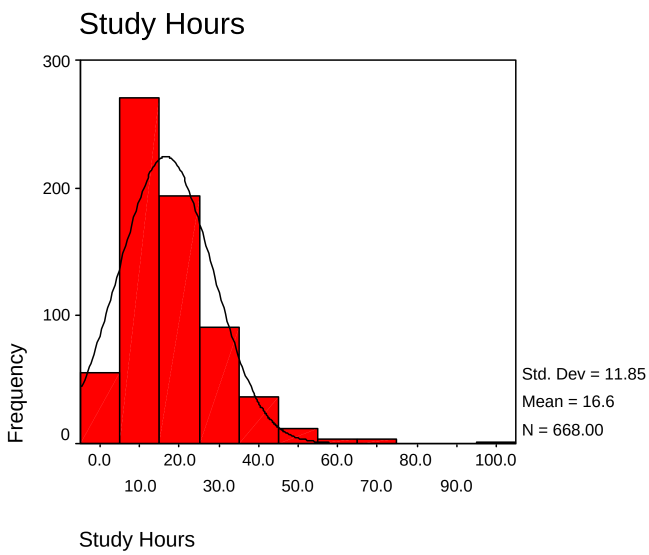
		Study Hours	Housework Hours
N	Valid	668	639
	Missing	39	68
Mean		16.61	5.14
Std. Deviation		11.849	6.766

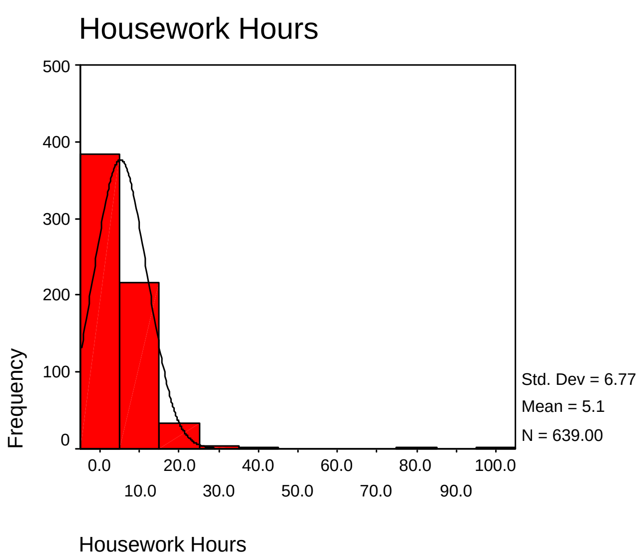
Study Hours

		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	Other	1	.1		
	Uncertain	4	.6		
	No response	24	3.4		
	System	10	1.4		
	Total	39	5.5		
Total		707	100.0		

Housework Hours

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	31	4.4	4.9	4.9
	1	4	.6	.6	5.5
	1	61	8.6	9.5	15.0
	2	3	.4	.5	15.5
	2	139	19.7	21.8	37.2
	3	3	.4	.5	37.7
	3	85	12.0	13.3	51.0
	4	57	8.1	8.9	59.9
	5	96	13.6	15.0	75.0
	6	22	3.1	3.4	78.4
	7	20	2.8	3.1	81.5
	8	4	.6	.6	82.2
	8	17	2.4	2.7	84.8
	9	2	.3	.3	85.1
	10	45	6.4	7.0	92.2
	11	1	.1	.2	92.3
	12	6	.8	.9	93.3
	14	3	.4	.5	93.7
	15	8	1.1	1.3	95.0
	17	1	.1	.2	95.1
	18	2	.3	.3	95.5
	19	1	.1	.2	95.6
	20	20	2.8	3.1	98.7
	24	2	.3	.3	99.1
	25	2	.3	.3	99.4
	30	1	.1	.2	99.5
	35	1	.1	.2	99.7
	80	1	.1	.2	99.8
	100	1	.1	.2	100.0
	Total	639	90.4	100.0	
Missing	Other	1	.1		
	UNCERTAIN	1	.1		
	NO RESPONSE	43	6.1		
	System	23	3.3		
	Total	68	9.6		
Total		707	100.0		





(ii) For study hours, the mean is 16.61 and the standard deviation is 11.849, or 11.85, to two decimals. The interval within one standard deviation of the mean is from 16.61-11.85=4.76 to 16.61+11.85=28.46. Looking at the table of study hours there are 5.7% of cases less than or equal to 4.76 (cumulative percent column) and 84.3% less than or equal to 28.46. This means there are 84.3-5.7=78.6% of cases within one standard deviation of the mean. From the standardized normal distribution, between Z=-1 and Z=+1 there are 0.3413+0.3413=0.6826 or 68.3% of the cases. As a result, in the actual distribution of study hours, there is a larger percentage of cases within one standard deviation of the mean than in the case of the normal distribution.

Two standard deviations is 2x11.849=23.70. The interval of two standard deviations around the mean is from 16.61-23.70=-7.09 hours to

16.61+23.70=40.31 hours. Since study hours cannot be less than zero, this means the interval is from 0 to 40.31 hours. The percentage of cases up to 40.31 hours is 96.7% (from the cumulative percentage distribution). For the normal distribution, the area between Z=-2 and Z=+2 is 0.4772+0.4772=0.9544 or 95.44%, slightly less than in the case of the actual distribution of study hours.

From the histogram of study hours, it can be seen that the distribution is not all that different from a normal distribution, except for the very tall bar centred at 10 hours. While the normal curve does not always match the height of the bars, for the other bars it comes close. But for housework hours, the normal curve would extend to the left, well below 0. The actual distribution of housework hours is very different from a normal curve, in that most students report very few housework hours, with the distribution concentrated at close to 0. This is far from being a symmetrical distribution, as the normal curve is. For the bars centred at 10 and 20 hours of housework though, the normal curve actually comes fairly close to matching the height of the bars.

c. (i)

Statistics

		Government Helps Business	Power to Affect Future	User Fees	More Health Care Dollars
N	Valid	688	692	693	686
	Missing	19	15	14	21
Mean		3.58	3.27	2.03	3.49
Std. Deviation		1.028	1.141	1.176	1.067

Government Helps Business

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	15	2.1	2.2	2.2
	2	89	12.6	12.9	15.1
	3	210	29.7	30.5	45.6
	4	229	32.4	33.3	78.9
	Strongly Agree	145	20.5	21.1	100.0
	Total	688	97.3	100.0	
Missing	No Response	19	2.7		
Total		707	100.0		

Power to Affect Future

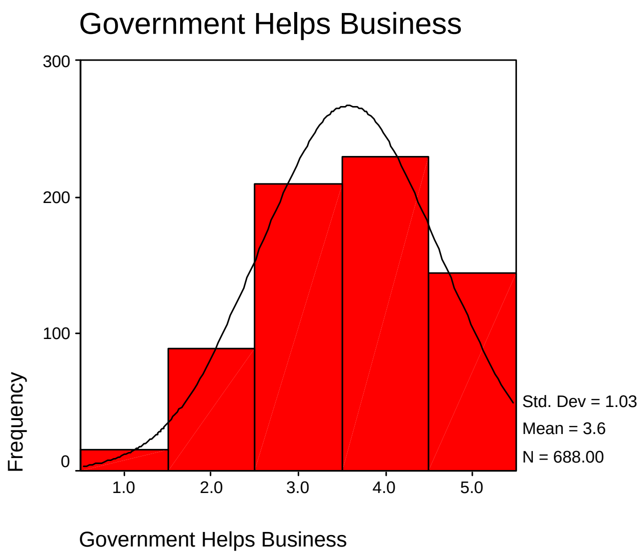
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	48	6.8	6.9	6.9
	2	127	18.0	18.4	25.3
	3	220	31.1	31.8	57.1
	4	186	26.3	26.9	84.0
	Strongly Agree	111	15.7	16.0	100.0
	Total	692	97.9	100.0	
Missing	No Response	15	2.1		
Total		707	100.0		

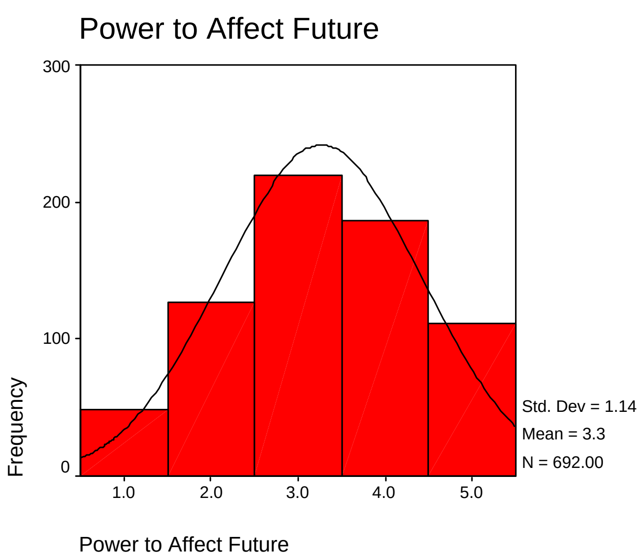
User Fees

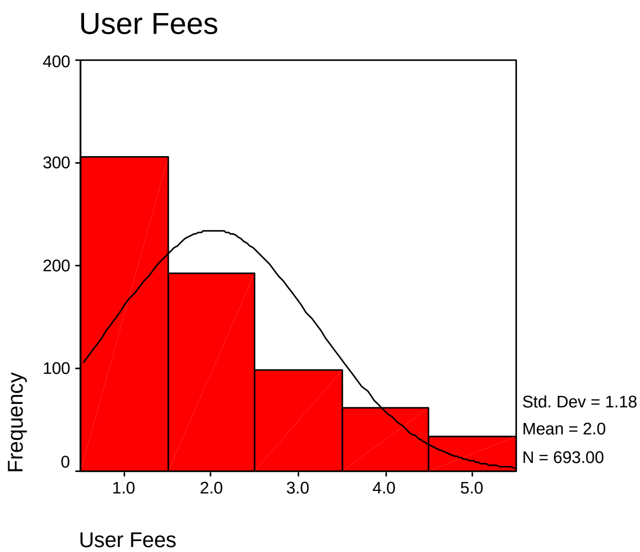
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	306	43.3	44.2	44.2
	2	193	27.3	27.8	72.0
	3	98	13.9	14.1	86.1
	4	62	8.8	8.9	95.1
	Strongly Agree	34	4.8	4.9	100.0
	Total	693	98.0	100.0	
Missing	No Response	14	2.0		
Total		707	100.0		

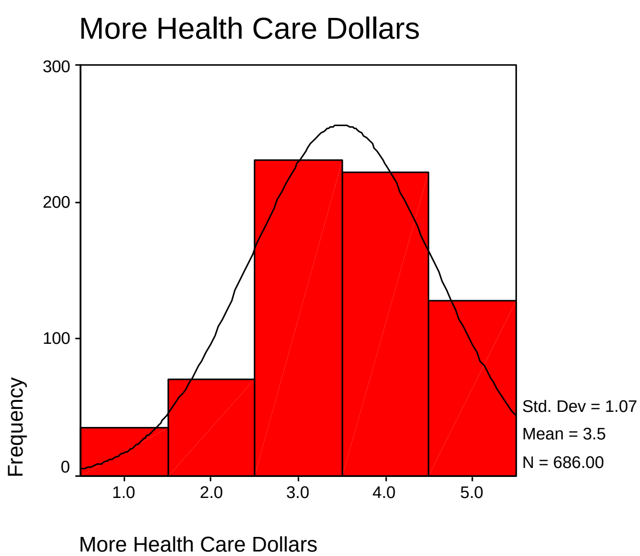
More Health Care Dollars

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	35	5.0	5.1	5.1
	2	71	10.0	10.3	15.5
	3	230	32.5	33.5	49.0
	4	222	31.4	32.4	81.3
	Strongly Agree	128	18.1	18.7	100.0
	Total	686	97.0	100.0	
Missing	Other	1	.1		
	Uncertain	1	.1		
	No Response	18	2.5		
	System	1	.1		
	Total	21	3.0		
Total		707	100.0		


Government Helps Business
300
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Std. Dev = 1.03
Mean = 3.6
N = 688.00
Government Helps Business


Power to Affect Future
300
200
100
0
Frequency
1.0
2.0
3.0
4.0
5.0
Std. Dev = 1.14
Mean = 3.3
N = 692.00
Power to Affect Future





(ii) The last of the four opinion questions is least like the normal curve in that responses to this question are concentrated at a response of 1 and 2, with successively fewer responses at each larger value. That is, the distribution of opinion is skewed toward the disagree end of the scale on the question of user fees. In contrast, the normal curve is symmetric.

The distribution most like the normal is probably the power to affect the future. For that distribution opinions are centred just above 3 (at 3.3) and opinion level 3 this is where the largest number of responses occur. Then there are successively fewer responses as one moves toward both the disagree and agree end of the opinion scale. While these responses are not perfectly symmetric around the centre, the heights of the bars come close to matching the normal curve. The distribution of more health care dollars is also close to normally distributed.