Social Studies 201

Fall 2003

Answers to Problem Set No. 4

November 3, 2003

1. (a) The statement that Thatcher is no risk to most people is primarily a **subjective** interpretation of probability. That is, it is the witness's judgment about whether Thatcher really is a risk or not. There may be some aspects of a **frequency** interpretation in that those who have been in prison for a long time may not present much risk to others, but this is mostly a subjective interpretation.

   (b) This probability is calculated by someone who has carefully used principles of probability, considering the drawing of cards from a deck of cards to conform to the ideal games of chance. This is thus a **theoretical** or **classical** interpretation of probability; it is not based on observations of actual poker games or on someone's judgment, but can be calculated by considering drawings from the deck of cards to conform to an ideal game of chance.

   (c) This probability is a **frequency** or **empirical** interpretation of probability, presumably based on surveys or administrative records of causes of infant deaths. There is no way that this can be a theoretical interpretation, since the likelihood of this cause of infant death cannot be calculated using theoretical principles – the actual record of causes of infant deaths would need to be considered. It could be a judgment probability, but the fact that the quote says "40 times more likely" indicates that some actual records were used to obtain this probability.

   (d) This quote may mix **frequency** and **judgment** probabilities. The data comes from a survey, so this means a frequency interpretation, with data obtained from a large sample of cases. But whether or not the individual surveyed is likely to buy a house is a **subjective** probability. That is, the individual reports a likelihood or chance that he or she will purchase a house – this is presumably based on the individual's judgment.

(e) This is a **frequency** interpretation of probability. There is no way that this can be reasoned out theoretically, without considering how many people of each income level go to university. At the same time, it appears to be more than someone's judgment, since the data come from a "study" conducted by Statistics Canada.

2. Using the crosstabulation table of V4 by sex of respondent in Table 1, the probabilities are as follows.

Table 1: Cross-classification of responses to V4 by sex of respondent, number of respondents with each combination of characteristics

Response to question V4	Number of males	Number of females	Total
1 (Strongly disagree)	71	70	141
2	28	58	86
3	70	113	183
4	53	108	161
5 (Strongly agree)	38	86	124
Total	260	435	695

(a) The probability of selecting a male is

$$P(\text{Male}) = \frac{\text{number of males}}{\text{total number}} = \frac{260}{695} = 0.374.$$

(b) The likelihood of selecting someone who strongly disagrees with same-sex marriage is

$$P(\text{Strongly disagree}) = \frac{\text{number strongly disagreeing}}{\text{total number}} = \frac{141}{695} = 0.203.$$

(c) The probability of selecting an individual who agrees (4 or 5) is

$$P(4) + P(5) - P(4 \text{ and } 5) = \frac{161}{695} + \frac{124}{695} - \frac{0}{695} = \frac{285}{695} = 0.410$$

.

(d) The probability of selecting a female or someone with a neutral response is

$$P(\text{female})+P(3)-P(\text{female and }3) = \frac{435}{695}+\frac{183}{695}-\frac{113}{695} = \frac{505}{695} = 0.727.$$

(e) The probability of selecting a male and someone with a neutral response is

$$P(\text{male and }3) = \frac{70}{695} = 0.101.$$

(f) The probability of selecting an individual who strongly disagrees, given that the individual is male is

$$P(\text{strongly disagree/male}) = \frac{71}{260} = 0.273.$$

(g) The probability that the individual selected stongly disagrees, given female, is

$$P(\text{strongly disagree/female}) = \frac{70}{435} = 0.161.$$

(h) If A is the event of strongly agreeing and B is the event of being female, one way to determine whether the two events are independent is to calculate $P(A/B)$ and $P(A)$. If these two probabilities are equal, then the two events are independent, if unequal, A and B are dependent. These probabilities are

$$P(A/B) = \frac{86}{435} = 0.198.$$

$$P(A) = \frac{124}{695} = 0.178.$$

While these two probabilities are close to each other, they differ so A and B are somewhat dependent events.

Alternatively, a check to see whether $P(B/A)$ and $P(B)$ are equal is another way to examine this. These probabilities are

$$P(B/A) = \frac{86}{124} = 0.694.$$

$$P(B) = \frac{435}{695} = 0.626.$$

Again, these probabilities differ, so there is a mild dependence between the events of strongly agreeing with treating gay and lesbian couples as married and being female.

(i) If A is the event of having a neutral response and B is the event of being female, these two events can be checked for independence by examining the probabilities of $P(A/B)$ and $P(A)$. These probabilities are

$$P(A/B) = \frac{113}{435} = 0.260.$$

$$P(A) = \frac{183}{695} = 0.263.$$

These two probabilities are very close to equal, so A and B can be considered to be independent of each other in practice.

Alternatively, a check to see whether $P(B/A)$ and $P(B)$ are equal is another way to examine this. These probabilities are

$$P(B/A) = \frac{113}{183} = 0.617.$$

$$P(B) = \frac{435}{695} = 0.626.$$

Again, these two probabilities are very close to equal, so A and B are practically independent. That is, the neutral response to whether gay and lesbian couples should be regarded as married is more or less independent of the event of being female.

(j) From parts f and g, the probability of strongly disagreeing with treating gays and lesbians as married is considerably greater for males (0.273) than for females (0.161). In part h, while the dependence is not great, females are more likely to express strong agreement with treating gays and lesbians as married (0.198) than is the sample as a whole (0.178). As a result of these sets of probabilities, it is reasonably clear that in this sample females are more likely to be in favour of treating gays and lesbians as married, and males less likely to favour this. Part i indicates that a neutral response is very close to independent of being female, so it appears that similar proportions of females and males are in the middle, with a neutral response. But this might be expected, with males more likely on the disgree end of the attitude and females more likely on the agree end, with similar proportions in the middle.

3. Questions from cross-classification table of V8 by V9.

Table 2: Cross-classification of responses to V8 by V9, number of respondents with each combination of characteristics

Response to question V8 (view on user fees)	Response to question V9 (more health care spending) 1	2	3	4	5	Total
1 (Strongly disagree)	21	23	75	94	89	302
2	3	22	81	72	14	192
3	5	9	41	31	11	97
4	0	11	23	21	7	62
5 (Strongly agree)	6	6	10	3	6	31
Total	35	71	230	221	127	684

(a)

P(strongly agree with more \$s/strongly disagree user fees) $= 89/302 = 0.294$.

(b)

P(strongly agree with more \$s/strongly agree user fees) $= 6/31 = 0.194$.

(c) If A is event of somewhat agreeing (4) with V9 (more health care dollars) and B is the event of strongly disagreeing with user fees, check whether $P(A/B) = P(A)$. These probabilities are

$$P(A/B) = \frac{94}{302} = 0.311.$$

$$P(A) = \frac{221}{684} = 0.323.$$

These two probabilities are fairly close to being equal, so A and B can be considered to be close to independent of each other.

Alternatively, a check to see whether $P(B/A)$ and $P(B)$. These probabilities are

$$P(B/A) = \frac{94}{221} = 0.425.$$

$$P(B) = \frac{302}{684} = 0.442.$$

These two probabilities are very close to equal, so A and B can be considered to be practically independent of each other in practice.

(d) If A is the event of agreeing (4 or 5) with V9 and B is the event of somewhat agreeing with user fees (V8 of 4), the probabilities that need to be checked for equality are $P(A/B)$ and $P(A)$. These probabilities are

$$P(A/B) = \frac{21+7}{62} = \frac{28}{62} = 0.452.$$

$$P(A) = \frac{221+127}{684} = \frac{348}{684} = 0.509.$$

These probabilityes differ somewhat, so A and B are mildly dependent on each other.

Alternatively, check whether $P(B/A)$ and $P(B)$ are equal. These probabilities are

$$P(B/A) = \frac{21+7}{221+127} = \frac{28}{348} = 0.080.$$

$$P(B) = \frac{62}{684} = 0.091.$$

Again, a difference between these two probabilities demonstrates a mild dependence between A and B.

(e) From the table and the above probabilities, I conclude that there some evidence that those individuals opposing user fees also express strong support for more health care spending, but it does not appear that the evidence is strong. From a and b, the probability of supporting more health care spending is greater among those who strongly disagree with user fees, as opposed to those who strongly agree with user fees. However, in part c, somewhat agreeing with more health care dollars and strongly disagreeing with user fees are practically independent. A dependence between these events would provide more evidence for the relationship suggested in the question. Again, in part d, while the overall probability of agreeing with more health care spending ($P(A)$) is greater than the conditional probability of agreeing, given agreement with user fees, the difference is not great. While it would take a more detailed analysis of conditional probabilities and independence to provide a complete answer to this question, the probabilities calculated above do not lend strong support to the idea that those opposed to user fees are the same individuals who support more health care spending. It appears that more spending is broadly supported, more or less independently of whether there is opposition to or support of user fees.

4. Some of the pairs of dependent and independent events are as follows:

   (a) **Dependent events**. Events of consuming dietary fat (A) and having diabetes (B).

   (b) **Dependent events**. Events of being inactive (A) and having diabetes (B).

   (c) **Dependent events**. Events of being an arts and culture graduate (A) and moonlighting (B).

   (d) **Dependent events**. Events of being an arts and culture graduate (A) and earning lower pay (B).

   (e) **Independent events**. Events of consuming sugar (A) and having diabetes (B).

   Explanation. For the first two sets of dependent events, the probability of B (diabetes) is increased by consuming dietary fat or being inactive. That is, the conditional probability of having diabetes is greater for those who consume dietary fat or are inactive, as compared with the overall probability of contracting diabetes, for the population as a whole. For the second set of dependent events, the event of being an arts or culture graduate increases the probability of moonlighting and reduces the pay level. The set of events that may be independent is consumption of sugar (A) and having diabetes (B). That is, the probability of having diabetes (B) does not appear to be different for those who consume sugar from those who do not, since it is argued that sugar is not a culprit (or cause) of diabetes.

5. Standardized normal distribution.

   (a) For $Z = 1.53$, the A area if 0.4370 and this is the area between $Z = 0$ and $Z = 1.53$.

   (b) The area between the centre of the normal distribution and $Z = 2.12$ is the A area of 0.4830. The area between the centre of the normal distribution and $Z = 1$ is 0.3413. The area between $Z = 1$ and $Z = 2.12$ is thus $0.4830 - 0.3413 = 0.1417$.

   (c) The proportion of cases between $Z = -1.20$ and $Z = 0.70$ is the sum of the areas between the centre of the distribution and each of these Z values. Between $Z = 0$ and $Z = -1.20$, the area is 0.3849 (that is, the same as the area between $Z = 0$ and $Z = +1.20$) while the area between $Z = 0$ and $Z = 0.70$ is 0.2580. The sum of these two areas is $0.3849 + 0.2580 = 0.6429$ and this is the required area.

   (d) The area to the right of $Z = 1.75$ is the B area associated with $Z = 1.75$. This area is 0.0401. As a percentage, this is $0.0401 \times 100 = 4.01$, or 4 per cent.

   (e) The area under the normal curve above $Z = -1.08$ is the area between $Z = -1.08$ and $Z = 0$ plus the 0.5000 of the area to the right of $Z = 0$. For $Z = 1.08$, the A area is 0.3599 and, by symmetry, this is the same area as between $Z = 0$ and $Z = -1.08$. The area above $Z = -1.08$ is thus $0.3599 + 0.5000 = 0.8599$.

   (f) Recall that the standard deviation for Z is 1, so 1.2 standard deviations is equivalent to $Z = 1.20$. For $Z = 1.20$, the A area is 0.3849. The area under the normal curve below 1.2 standard deviations above the mean is this area of 0.3849 plus the 0.5000 to the left of the mean, or $0.3849 + 0.5000 = 0.8849$. This is $0.8849 \times 100 = 88.49$, or eighty-eight per cent of population memebers.

(g) If 0.3 of the area is in the right tail of the distribution, the B area is 0.3000. The Z closest to providing a B area of 0.3000 is $Z = 0.52$ with an area of 0.3015 in the right tail, or possibly 0.53, with an area of 0.2981 in the right tail. This is the required Z value.

(h) If there is 0.04 of the area in one tail of the distribution (B area), the appropriate Z value associated with this is $Z = 1.75$, where the B area is 0.0401, just over 0.0400. The required Z values are thus -1.75 and +1.75. Beyond these Z values, there is very close to 0.0400 of the area in each tail of the distribution.

(i) By definition, the first quartile is the value of the variable such that 0.25 of the area is less than this. Looking for a B area of 0.2500 gives $Z = 0.67$ or 0.68 as the closest Z values. Thus the first quartile, or twenty-fifth percentile occurs at $Z = -0.67$ or $Z = -0.68$.