Social Studies 201 – Fall 2006

Problem Set 4

Due no later than November 15, 2006

Note: If you complete this problem set by Friday, November 10, we will attempt to have it marked prior to the November 17 midterm. Answers will be posted on the web site on Wednesday afternoon, November 15.

1. Probabilities of events. Suppose an individual is randomly selected from the 487 persons in Table 1.

a. What are the following probabilities?

(i)	Selecting a person of age 45-54.
(ii)	Selecting a person with less than six close friends.
(iii)	Selecting a person with 11 or more close friends and of age 75 plus.
(iv)	Selecting a person aged 75 plus given that the person has more than 25 close friends.
(v)	Selecting a person with 11 or more close friends, given that the person is age 75 plus.

b. Explain whether the following events are independent or dependent.

(i)	Age 45-54 and having no close friends.
(ii)	Having 2 or less close friends and being age 75 plus.

Table 1. Number of Saskatchewan respondents with various numbers of close friends, 2003, ages 15-24. 45-54, and 75 plus

Number of close friends	Number of respondents of			Total
	Age 15-24	Age 45-54	Age 75 plus	
None	2	10	16	28
1 or 2	28	33	22	83
3 to 5	76	57	39	172
6 to 10	52	48	29	129
11 to 25	19	20	19	58
More than 25	3	5	9	17
Total	180	173	134	487

Source: Adapted from Statistics Canada. General Social Survey of Canada, 2003. Cycle 17: Social Engagement [machine readable data file]. 1st Edition. Ottawa, ON: Statistics Canada [publisher and distributor] 10/1/2004.

2. Independence and dependence. From the following quotes, identify one pair of independent events and one pair of dependent events. Explain your reasoning.

a. In a study of who cares for seniors, Statistics Canada reported "Although these middle-aged caregivers are just as likely to be men as women, the women dedicate almost twice as much time to their tasks – 29.6 hours per month, compared with 16.1 hours for men. Working outside the home does not significantly reduce the amount of time middle-aged people spend providing care." (*Canadian Social Trends*, Autumn 2004, p. 2).

b. A Statistics Canada report on immigrants to Canada taking Canadian citizenship reports that "the longer newcomers stay in Canada, the more likely they are to become citizens." (*The Daily*, March 8, 2005).

3. Standardized normal distribution. Obtain the following from the standardized normal distribution.

    a.   The area between Z of 0 and Z of -2.12?

    b.   What is the area between Z of +1.3 and Z of +2.3?

    c.   What is the percentage of cases between Z = -0.97 and Z = +0.1.34?

    d.   What proportion of the area under the normal curve above Z = -1.25?

    e.   What is the area under the normal curve such that Z < 1.84?

    f.   In a normally distributed population, what is the percentage of the population is within one-half of a standard deviation from the mean?

    g.   What is the Z-value so that 0.63 of the area is less than this Z? (That is, what is the Z-value of the 63rd percentile?)

    h.   What are the Z-values so that there is 0.0325 of the area in each tail of the distribution lying beyond these Z-values, for a total of 6.5 per cent in the two tails of the distribution?

    i.   In a standardized normal distribution, where is the thirty-sixth percentile?

    j.   The 70 per cent range can be defined as the range so that only the extreme thirty per cent of cases (fifteen per cent in each end of the distribution) are eliminated. In the standardized normal distribution, what are the Z-values for the seventy per cent range?

4. Distribution of income. From the Saskatchewan respondents in the General Social Survey of Canada, Cycle 17, 2003, the mean personal income is 33.0 thousand dollars and the standard deviation of personal income is 26.4 thousand dollars.

a. If personal income is normally distributed, what are the following?

    i.   Proportion of persons with an income of less than ten thousand dollars.

    ii.   Percentage of persons with an income of under fifty thousand dollars.

    iii.   Percentage of persons with incomes of fifty to seventy thousand dollars.

    iv.   The fifty-fourth percentile of income.

b. From part a., Figure 1, and Table 2, comment on how the distribution of personal income differs from that of a normal distribution.

Table 2. Percentage distribution of personal income, Saskatchewan, 2003

Income in thousands of dollars	Percentage of cases
Less than 10	16.9
10-30	37.2
30-50	24.7
50-70	11.6
70 plus	9.6
Total	100.0

Figure 1. Histogram of personal income distribution, Saskatchewan, 2003

Histogram
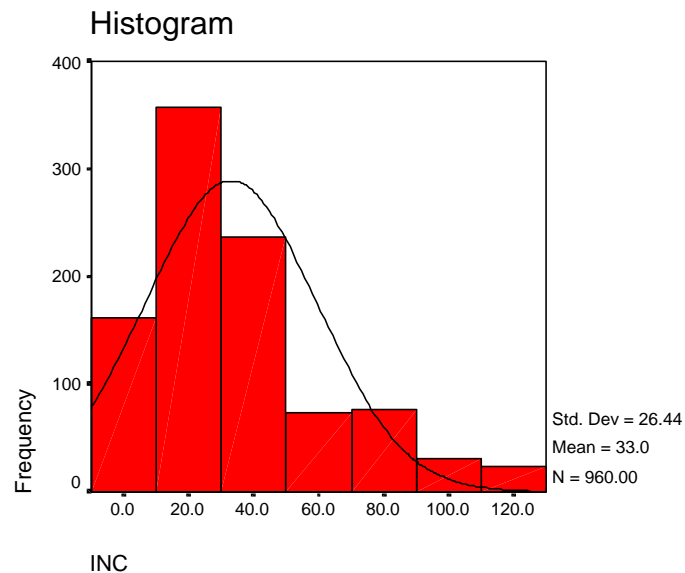
5. Estimation of incomes of employed females.   Table 3 provides the means and standard deviations of income for Saskatchewan females of various ages who were employed full-time and full-year in 2003. Data come from Statistics Canada, General Social Survey, Cycle 17, with sample sizes shown for each of the age groups.

a. Obtain 90 per cent interval estimates for the true mean income for females of each of the four age groups. (Four interval estimates).

b. Briefly comment on the pattern of true mean income by age group, using data from Table 3 and from part a.

Table 3. Income statistics, Saskatchewan females employed full-time and full-year, 2003

Age group	Income in thousands of dollars		Sample size
	Mean	Standard deviation	
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

Source: Data for this question adapted from Statistics Canada. General Social Survey of Canada, 2003. Cycle 17: Social Engagement [machine readable data file]. 1st Edition. Ottawa, ON: Statistics Canada [publisher and distributor] 10/1/2004.

6. Computer problems – use the file ssae98.sav in the folder t:\students\public\201. Hand in the SPSS output form part (i) of each question and provide a written response to the questions in part (ii) of each of the following.

a. (From Problem Set 3). (i) Obtain the mean and standard deviation of the variable "respect for government" (RESPECCTG) classified by provincial political preference (use PV, at the end of the data file).   Then obtain a crosstabulation (*Analyze-Descriptive Statistics-Crosstabs*) to obtain the cross-classification of respect for government by provincial political preference.   When requesting this procedure, you may wish to obtain cell or row percentages to help interpret the table.   (ii) Write a short description of what the two tables illustrate about differences in spect for governments among those of different political preference.   Also, from the cross-classification table, what are the following probabilities:

- Probability of selecting a Conservative supporter.

- Probability of selecting someone with moderate or great respect for government (ie. categories 4 or 5).

- Probability of selecting someone who supports no political party and has no respect for government.

- The conditional probability of no respect given support for Liberals. Given support for none of the parties. Explain whether these respective probabilities seem consistent with the different mean respect for Liberal supporters and those who support no political party?

b. (i) Use *Analyze-Descriptive Statistics-Frequencies*, with options *Charts-Histograms-With Normal Curve* and *Statistics* (mean and standard deviation), to obtain frequency distributions, statistics, and histograms of the hours per week spent studying (*sthours*) and at housework (*hwhours*). The frequency distribution table, statistics, and the histogram, with the normal curve superimposed, should appear on the printout. (ii) Using the table of the frequency and percentage distribution of study hours, obtain the percentage of cases that are within one standard deviation of the mean of study hours; within two standard deviations of the mean of study hours. Compare these results with the percentages of cases within one and two standard deviations from the table of the normal distribution (Appendix H). Use these figures and the diagram on the printout to write a short note comparing the actual distribution of study hours and housework hours with that of a normal distribution.

c. (i) In the first computer problem set, you worked with the four political and social opinion variables *v6* through *v9* from question 15 of the questionnaire. Use Analyze-*Descriptive Statistics-Frequences*, with options *Charts-Histograms-With Normal Curve* to obtain frequency distributions and histograms for these four variables. (ii) Explain which of these variables appears to have a distribution closest to that of a normal distribution and which is least like a normal distribution.