Assignment #4

Due on Wednesday October 8, 2008

Read Chapter 3 in the class notes, *Introduction to Probability,* in the course webpage at `http://pages.pomona.edu/~ajr04747`

Read Section 4.4 on *Means and Variances of Random Variable* in Moore, McCabe and Craig.

Do the following problems

1. The results of the simulations for Activity #3, *The Cereal Box Problem,* are contained in the MS Excel file `CerealBoxProblemAllClassSimulations.xls`, which may be downloaded from `http://pages.pomona.edu/~ajr04747`.

   The column labeled `Nboxes` denotes the number of trials that it took to get all six numbers (i.e., all six prizes in the cereal box) in each run.

   (a) Use R to plot a histogram on `Nboxes`.

   (b) Let Y denote the random variable that gives the number of boxes of cereal that need to be bough in order to obtain all six prizes. Estimate the following probabilities:

      i. $P(8 < Y \leqslant 21)$

      ii. $P(Y \geqslant 36)$

2. (Distributions of Errors[1]) Typographical and spelling errors can be either "non-word errors" or "word errors." A nonword error is not a real word, as when "the" is typed as "teh." A word error is a real word, but not the right word, as when "lose" is typed as "loose." A group of students where asked to write a 250–word essay (without spell–checking). Based on the data collected, the following probability distribution for nonword errors was obtained:

Errors	0	1	2	3	4
Probability	0.1	0.3	0.3	0.2	0.1

   Table 1: Distribution of nonword errors

   For word errors, the distribution was obtained is shown in Table 2.

   Compute the mean number of nonword errors and word errors in an essay.

[1] Adapted from Exercise 4.74 in Moore, McCabe abd Graig, *Introduction to the Practice of Statistics,* Sixth Edition, p. 286

Errors	0	1	2	3
Probability	0.4	0.3	0.2	0.1

Table 2: Distribution of word errors

3. (Owner–occupied versus rented housing units[2]) How do rented housing units differ from units occupied by their owners? According to the Census Bureau's 1998 American Housing Survey, the distribution of the number of rooms of owner–occupied and renter–occupied units in San Jose, California, is given by the following table

Rooms	1	2	3	4	5	6	7	8	9	10
Owner	0.003	0.002	0.023	0.104	0.210	0.224	0.197	0.149	0.053	0.035
Rented	0.008	0.027	0.287	0.363	0.164	0.093	0.039	0.013	0.003	0.003

What are the most striking differences between the two distributions?

4. (Refer to the owner–occupied versus renter–occupied data given in the previous problem).

   (a) Compute the mean number of rooms for each kind of housing unit described in the previous problem

   (b) Compute the standard deviations for the two distributions displayed in the previous problem. How do they differ?

5. (Will you Assume Independence?[3]) In which of the following games of chance would you be willing to assume independence of X and Y in making a probability model? Explain your answer in each case.

   (a) In blackjack, you are dealt two cards and examine the total points, X, on the cards (face cards count 10 points). You choose to be dealt another card and compete based on the total points, Y, on all three cards.

   (b) In craps, the betting is based on successive rolls of two dice. X is the sum of the faces on the first roll, and Y is the sum of the faces in the next roll.

[2]Adapted from Exercise 4.53 in Moore, McCabe abd Graig, *Introduction to the Practice of Statistics,* Sixth Edition, p. 268

[3]Adapted from Exercise 4.83 in Moore, McCabe abd Graig, *Introduction to the Practice of Statistics,* Sixth Edition, p. 288