Solutions to Assignment #4

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`.

   Solution: The histogram is shown in Figure 1

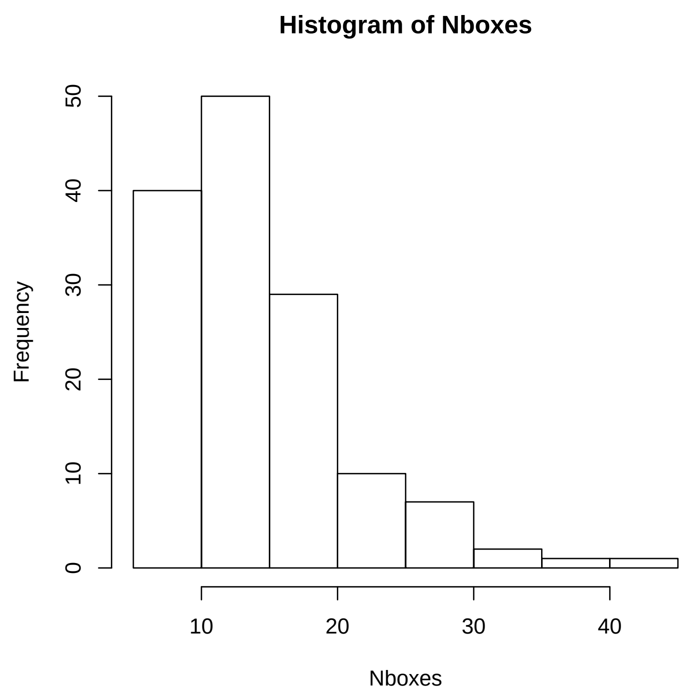

Figure 1: Cumulative Distribution for `Nboxes`

□

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

Solution: We can use the `CumDist()` function to compute

```
CumDist(Nboxes,21)-CumDist(Nboxes,8)
```

to get that $P(8 < Y \leqslant 21) \approx 0.7929$, or about 79.3%. □

ii. $P(Y \geqslant 36)$

Solution: For this calculation we can use the `pHat()` function to get $P(Y \geqslant 36) =$`pHat(Nboxes,36)`≈ 0.0143 or about 1.43%. □

2. (Distributions of Errors[1]) Typographical and spelling errors can be either "nonword 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.

Errors	0	1	2	3
Probability	0.4	0.3	0.2	0.1

Table 2: Distribution of word errors

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

Solution: Let N denote the number of nonword errors. Then

$$E(N) = 0 \cdot (0.1) + 1 \cdot (0.3) + 2 \cdot (0.3) + 3 \cdot (0.2) + 4 \cdot (0.1) = 1.9$$

Similarly, if W denotes the number of word errors, then

$$E(W) = 0 \cdot (0.4) + 1 \cdot (0.3) + 2 \cdot (0.2) + 3 \cdot (0.1) = 1.0$$

□

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

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?

Solution: The distribution for the owner occupied units is shown in Figure 2 while that for renters–occupied units is in Figure 3.

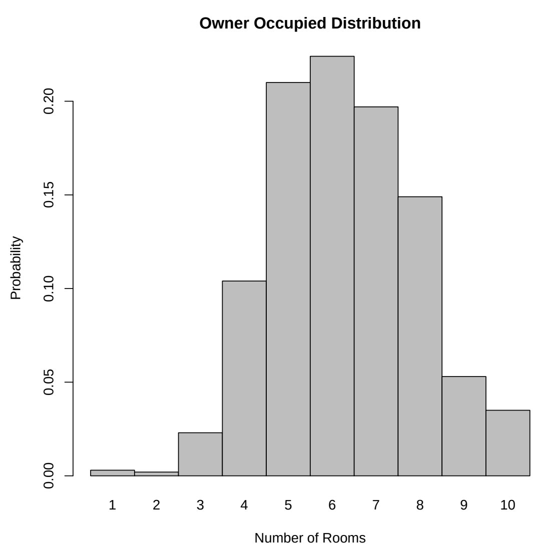

Figure 2: Probability distribution of number of rooms for owner occupied units

The most striking difference is that rented units have typically fewer rooms and their distribution is skewed to the right. □

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

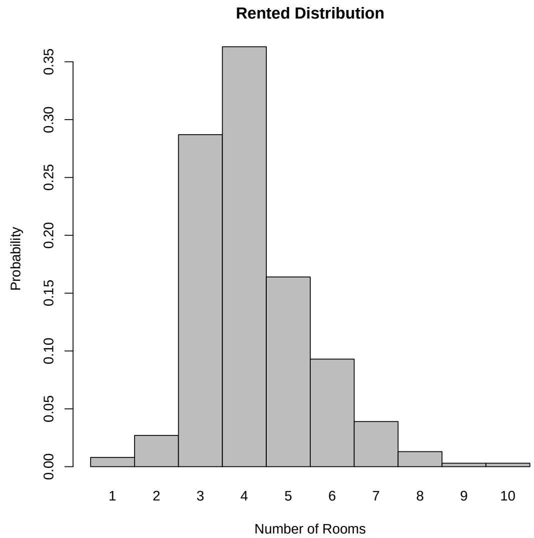

Figure 3: Probability distribution of number of rooms for rented units

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

   Solution: The expected value for the number of rooms in owner–occupied units is

$$1 \cdot (0.003) + 2 \cdot (0.002) + 3 \cdot (0.023) + 4 \cdot (0.104) + 5 \cdot (0.210) \\ +6 \cdot (0.224) + 7 \cdot (0.197) + 8 \cdot (0.149) + 9 \cdot (0.053) + 10 \cdot (0.035)$$

   which is 6.284. A similar calculation for the number of rooms of rented units yields 4.187.

   □

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

   Solution: Let X denote the number of rooms in owner–occupied units. We have seen that $E(X) = 6.284$. To find the standard deviation of of X, we first compute the variance

$$\text{Var}(X) = E(X^2) - [E(X)]^2,$$

where $E(X^2)$ is computed as follows

$$1^2 \cdot (0.003) + 2^2 \cdot (0.002) + 3^2 \cdot (0.023) + 4^2 \cdot (0.104) + 5^2 \cdot (0.210)$$
$$+6^2 \cdot (0.224) + 7^2 \cdot (0.197) + 8^2 \cdot (0.149) + 9^2 \cdot (0.053) + 10^2 \cdot (0.035)$$

which yields 42.178. It then follows that the variance of X is

$$\text{Var}(X) = 42.178 - (6.284)^2 \approx 2.689.$$

The standard deviation of X is then $\sigma_X \approx 1.64$.
A similar calculation for Y, the number of rooms in rented units, yields that $\sigma_Y \approx 1.3$.
The standard deviation for the number of rooms of owner occupied units is larger than that for rented units. Thus, the distribution of number of rooms for owner–occupied units is more spread out than that of rented units. □

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.

   Answer: X and Y are not independent. Having information on the first two cards effects the probability of the outcomes of Y. □

   (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.

   Answer: Outcomes of separate rolls of the dice are independent events. □

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