Statistics 341

Fall 2008 - Assignment #1 Solutions

Due Friday, September 12, 2008

This assignment is worth a total of 66 points.

1. (20 pts) Go to a computer with the statistical package R. In class, we used R to explore the probability of obtaining heads when flipping a coin. In this problem, you will use R to explore the probability of obtaining a 1 when rolling one dice. You will explore the possible outcomes of rolling one dice 100 times and 10,000 times and looking at both the number and proportion of the number of 1's that appear in the 100 rolls and the 10,000 rolls. Use the R help file **Introduction to the Concept of Probability** on the course webpage to get started and for assistance with some of the R programming. Use your analysis to answer the following questions.

   (a) (1 pt) What is the theoretical probability of obtaining a 1 when rolling one dice?
   The theoretical probability of obtaining a 1 when rolling one dice is 1/6.

   (b) (2 pts) Roll your dice in R 100 times. How many of your rolls were 1's? What proportion of the total number of rolls were 1's?
   R code:

```
dice<- c(0,0,0,0,0,1)
roll100<- sample(dice, 100, replace = T)
sum(roll100) #number of rolls that were 1's
sum(roll100/100) #proportion of 100 rolls that were 1's
```

   Your answers for sum(roll100) and sum(roll100/100) will vary. You probably have a value somewhere between 5 and 25 for sum(roll100) and 0.05 and 0.25 for sum(roll100/100). I got 17 and 0.17 for my 100 rolls.

   (c) (2 pts) Roll your dice in R 10,000 times. How many of your rolls were 1's? What proportion of the total number of rolls were 1's?
   R code:

```
roll10000<- sample(dice, 10000, replace = T)
sum(roll10000) #number of rolls that were 1's
sum(roll10000)/10000 #proportion of 10000 rolls that were 1's
```

   Your answers for sum(roll10000) and sum(roll10000/10000) will vary. You probably have a value somewhere between 1550 and 1800 for a proportion of 0.1550 to 0.1800. I got 1640 and 0.1640 for sum(roll10000) and sum(roll10000/10000).

   (d) (5 pts) Use R to conduct 10,000 trials where each trial consists of rolling a dice 100 times. What is the range of values for the number of 1's obtained out of 100 rolls? What is the range of values for the proportion of 1's obtained out of 100 rolls? How do these values compare to the values that are expected?
   R code:

```
num1s100<- rep(0, 10000)
prop1s100<- rep(0, 10000)
for (i in 1:10000){
rolls100<- sample(dice, 100, replace = T)
num1s100[i]<- sum(rolls100)
prop1s100[i]<- sum(rolls100)/100
}
```

The number of 1's in 100 rolls is between around 5 and 30 making the proportion of 1's in 100 rolls between around 0.05 and 0.30. The majority of the time, the number of 1's in 100 rolls is between around 10 and 23, making the proportion 0.10 and 0.23.

In 100 rolls, we would expect around 16 to 17 1's for a proportion of 0.16 or 0.17. Our values are centered around these expectations.

(e) (5 pts) Use R to conduct 10,000 trials where each trial consists of rolling a dice 10,000 times. What is the range of values for the number of 1's obtained out of 10,000 rolls? What is the range of values for the proportion of 1's obtained out of 10,000 rolls? How do these values compare to the values that are expected?

R code:

```
num1s10000<- rep(0, 10000)
prop1s10000<- rep(0, 10000)
for (i in 1:10000){
rolls10000<- sample(dice, 10000, replace = T)
num1s10000[i]<- sum(rolls10000)
prop1s10000[i]<- sum(rolls10000)/10000
}
```

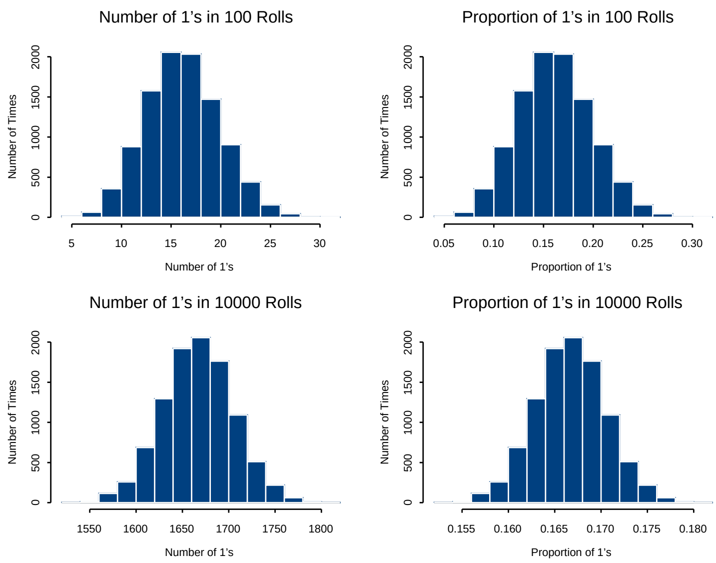
The number of 1's in 10000 rolls is between around 1550 and 1800 making the proportion of 1's in 10000 rolls between around 0.1550 and 0.1800. The majority of the time, the number of 1's in 10000 rolls is between around 1600 and 1730, making the proportion 0.1600 and 0.1730.

In 10000 rolls, we would expect around 1666 1's for a proportion of 0.1666. Our values are centered around these expectations.

(f) (5 pts) Write a statement that compares your results from parts (d) and (e) above. Which experiment gives you a better estimate of the theoretical probability of obtaining a 1 when rolling one dice; the experiment with 100 rolls or the experiment with 10,000 rolls?

Here is a picture of the four histograms from parts (d) and (e) above.

From the histograms for 100 rolls, the number of 1's is off by about plus or minus 11. This means that the proportions are off by about plus or minus 0.11. From the histograms of 10000 rolls, the number of 1's is off by about plus or minus 60 or so. However, since this is out of 10000 rolls, the proportions are off by about plus or minus 0.006 or so. So in absolute numbers, the experiment with 100 rolls is closer to the expected number, but in terms of proportion, the experiment with 10000 rolls is closer to the expected proportion, which is the theoretical probability. So the experiment with 10000 rolls is closer to the theoretical probability than the experiment with 100 rolls.



2. (5 pts) Go to a computer with the statistical package R. In my library of songs on iTunes, I have 49 songs from Shania Twain out of a total of 1,821 songs. I own an iPod Shuffle that can hold around 250 songs. If I randomly select songs from my library to load onto my iPod Shuffle, how many songs from Shania Twain will typically be selected? What numbers of songs from Shania Twain would be atypical? To assist you with this problem, use the R help file **The iPod Shuffle** on the course webpage. Remember, we are interested in looking at a particular artist in the random selection of songs for the iPod Shuffle. Include any output or commands from R that you feel is relevant to your answers.

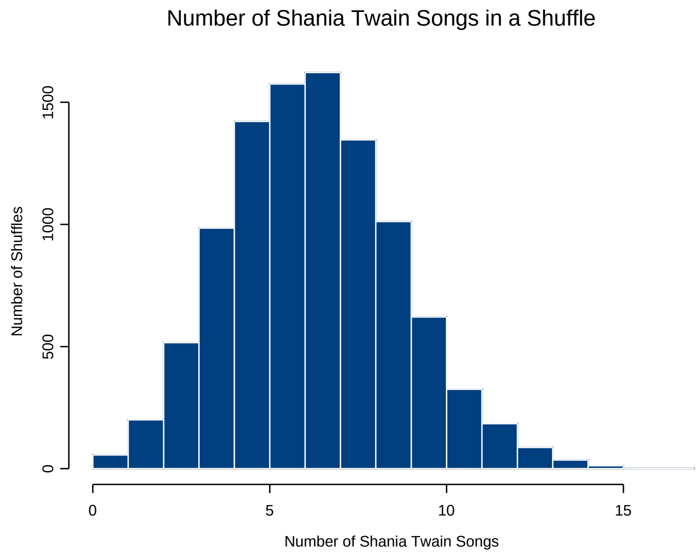
   R code:

```
artist<- c(rep(1,49), rep(0, 1772))
numsongs<- rep(0, 10000)
for (i in 1:10000){
shuffle<- sample(artist, 250, replace = F)
numsongs[i]<- sum(shuffle)
}
```

   Below is a histogram of the number of songs from Shania Twain selected for the shuffle. These data were generated from the code above and are based on 10000 shuffles.

   From the histogram, you can see that is typical to get around 2 to 10 songs and it is possible to get anywhere from 0 to 15 songs or so from Shania Twain in a shuffle. It would be very unusual to see more than around 18 songs or so in one shuffle from Shania Twain.

3. (11 pts) A company has 5 executives applying for the opportunity to attend a business leadership conference. There are two spaces available at the leadership conference. Of the 5 executives, two are men (denoted as M_1 and M_2) and three are women (denoted as W_1, W_2, W_3). Let S denote the set of all possible outcomes for the employer's selection of the two executives to attend the conference. Remember, since the two spaces at the conference are identical, it does not matter which person is chosen first or second.



(a) (5 pts) List the sample space S. How many events does S contain?

The sample space S contains 10 events. Here are the events.

(M_1, M_2)	(M_1, W_1)	(M_1, W_2)	(M_1, W_3)
	(M_2, W_1)	(M_2, W_2)	(M_2, W_3)
		(W_1, W_2)	(W_1, W_3)
			(W_2, W_3)

(b) (2 pts) The company feels that all 5 executives are equally deserving of attending the conference. So the company selects the two executives to attend the conference randomly from the 5 executives. What is the probability of each of the events in S? Why?

Since there are 10 events in S and the selection is done randomly, each of the 10 events should be equally likely. Since the probability of all 10 events must add to 1 and probability is a value between 0 and 1, the probability of each event in S is 1/10.

(c) (1 pt) Let the event A be the event that both men are selected to attend the conference. What is the probability of this event?

Since the event $A = \{(M_1, M_2)\}$; the probability of this event is 1/10.

(d) (3 pts) The company announces that the two men have been selected to attend the business leadership conference. The three women executives claim gender discrimination. Do you think their claim has any merits? Explain your answer.

The probability of the event that occurred, the two men were selected, is 1/10. This probability assumes that the selection process was random. At a probability of 1/10, some people would think this probability is too low to occur, and so would question the randomness of the selection. However, other people would argue that this probability isn't that low, and so would not question the randomness of the selection. Obviously, the lower this probability gets, the more we would question whether the selection was random. At a probability of 1/10, you are really in a somewhat of a grey area.

4. (10 pts) Two additional jurors are needed to complete a jury for a criminal trial. The trial involves a case of domestic violence and attempted murder. There are six prospective jurors,

two women and four men. The two jurors will be chosen randomly from the six prospective jurors. The spots on the jury are identical so it does not matter which juror is chosen first or second.

(a) (5 pts) List the sample space S. How many events belong to the sample space S?

The sample space S contains 15 events. Here are the events.

$$\begin{array}{ccccc} (M_1, M_2) & (M_1, M_3) & (M_1, M_4) & (M_1, W_1) & (M_1, W_2) \\ & (M_2, M_3) & (M_2, M_4) & (M_2, W_1) & (M_2, W_2) \\ & & (M_3, M_4) & (M_3, W_1) & (M_3, W_2) \\ & & & (M_4, W_1) & (M_4, W_2) \\ & & & & (W_1, W_2) \end{array}$$

(b) (2 pts) Assign a probability to each possible outcome in S. Explain how you assigned this probability.

Since the two jurors are to be selected randomly from the pool of 6 jurors, each of the 15 events in S should be equally likely. Since all the events must sum to 1, and probability is between 0 and 1, each event has a probability of 1/15.

(c) (3 pts) When the random selection process is complete, the two women are selected for the jury. The defendant questions the randomness of the selection procedure. Do you think his claim has any merits? Explain your answer.

The probability the two women were chosen is 1/15. This probability assumes that the jurors were selected randomly. At a probability of 1/15, many people would think that probability is too low to occur, and so would question the randomness of the selection. One important piece of information is given in the problem; the defendant is on trial for domestic violence and attempted murder. It is more likely than not this defendant would not want the selection of two women, and would therefore, question the randomness of the selection process.

5. (11 pts) A box of candy hearts contains 52 hearts of which 19 are white, 10 are tan, 7 are pink, 3 are purple, 5 are yellow, 2 are orange, and 6 are green. If you select and eat 9 pieces of candy randomly from the box (no replacement possible), give the probability that

(a) (4 pts) Three of the hearts are white.

The number of outcomes in S is $N = \binom{52}{9}$.

You must select 3 hearts from the 19 white hearts. The other 6 hearts must be selected from the 33 remaining pieces. This makes the number of events in A $\binom{19}{3}\binom{33}{6}$. The probability is then

$$\frac{\binom{19}{3}\binom{33}{6}}{\binom{52}{9}}$$

(b) (5 pts) Three are white, 2 are tan, 1 is pink, 1 is yellow, and 2 are green.

You must select 3 hearts from the 19 white, 2 hearts from the 10 tan, 1 heart from the 7 pink, 1 heart from the 5 yellow and 2 hearts from the 6 green. You do not need to select any hearts from the other colors, since you have selected the total of 9 hearts. This makes the number of events in A $\binom{19}{3}\binom{10}{2}\binom{7}{1}\binom{5}{1}\binom{6}{2}$. The probability is then

$$\frac{\binom{19}{3}\binom{10}{2}\binom{7}{1}\binom{5}{1}\binom{6}{2}}{\binom{52}{9}}$$

(c) (2 pts) Why is the probability of the event in part (a) larger than the probability of the event in part (b)?

The event in part (a) includes the event in part (b) and many more possible events with exactly three white hearts. There are more possible simple events in part (a) therefore in part (b), making the probability larger for part (a).

6. (9 pts) Go to a computer with the statistical package R. The Powerball lottery game is played in many states across the country, including Iowa. In the game, five white balls and one red ball (called the Powerball) are selected. The grand prize is won if the player matches each of the five white balls and the one red ball. (The order in which the balls are selected does not matter). The probability of winning the grand prize is thus dependent upon the number of white balls and the number of Powerballs in use. In recent years, the Powerball lottery group has increased the number of white balls from 45 to 55. The number of Powerballs in use has stayed constant at 42 for the last 8 years. What effect does increasing the number of white balls in use have on the probability of winning the Powerball grand prize?

To do this problem, you will need to use a couple of commands in R. To calculate the number of combinations of r objects taken from n objects, the command is **choose(n,r)**. To multiply two numbers a and b, the command is **a*b**. To create a variable that is a sequence of integer values from a to b, the command is **a:b**. To plot two variables x and y, the command is **plot(x,y)**.

(a) (1 pt) Using R, calculate the total number of possible winning numbers for the Powerball lottery when there are 45 white balls.

R code:

```
choose(45,5)*choose(42,1)
```

The answer is 51,313,878.

(b) (1 pt) Using R, calculate the total number of possible winning numbers for the Powerball lottery when there are 55 white balls.

R code:

```
choose(55,5)*choose(42,1)
```

The answer is 146,107,962.

(c) (2 pts) Using R, create a variable called **n**. Set the variable **n** equal to a sequence of integer values from 40 to 60. Use the variable **n** to calculate the total number of possible winning numbers for the Powerball lottery when there are **n** white balls. (Hint: simply replace the number of white balls in the calculation for parts (a) and (b) with the variable **n**). Save this calculation as the variable **winnums**. Make sure to print out this variable on the R output.

R code:

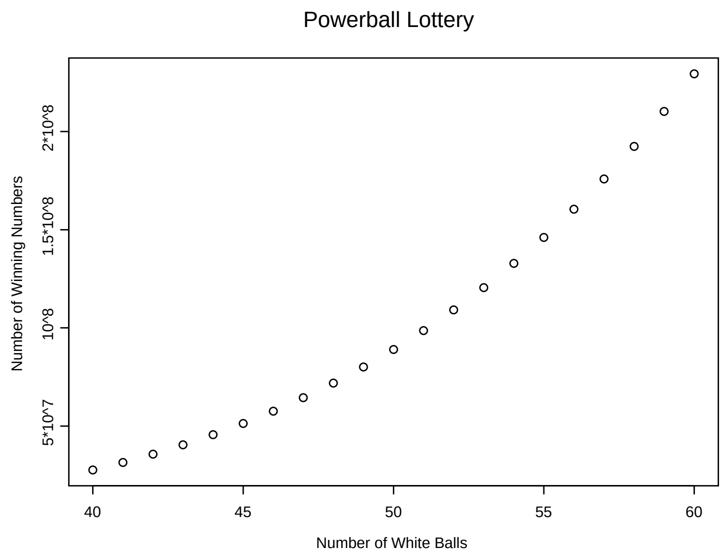
```
n<- c(40:60)
winnums<- choose(n,5)*choose(42,1)
```

Here are the values of the variable **winnums**.

```
 27636336  31474716  35728056  40429116  45612336
 51313878  57571668  64425438  71916768  80089128
 88987920  98660520 109156320 120526770 132825420
146107962 160432272 175858452 192448872 210268212
229383504
```

(d) (5 pts) Using R, plot the variable **n** against the total number of possible winning numbers (the variable **winnums**. Hint: The variable **n** is the x variable and the variable **winnums** is the y variable. Include the plot with your assignment. Describe the plot (what shape do the points resemble, etc). What happens to the number of possible winning numbers as the number of white balls in use increases? As a consequence, what happens to the probability of winning the grand prize as the number of white balls in use increases?

The plot is given below.



The plot is increasing in n. As the number of balls increases, the number of possible winning numbers increases as well. The shape of the plot looks much like an exponential type function. The probability of winning the grand prize is the inverse function of the number of possible winning numbers. So as the number of balls increases, the probability of winning the grand prize will decrease.