

Statistics 341

Fall 2008 - Assignment #1

Due Friday, September 12, 2008

1. 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) What is the theoretical probability of obtaining a 1 when rolling one dice?
 - (b) 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?
 - (c) 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?
 - (d) 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?
 - (e) 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?
 - (f) 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?
2. 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 assistance 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.
3. 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) List the sample space S . How many events does S contain?
 - (b) 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?
 - (c) Let the event A be the event that both men are selected to attend the conference. What is the probability of this event?
 - (d) 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.
4. 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) List the sample space S . How many events belong to the sample space S ?
 - (b) Assign a probability to each possible outcome in S . Explain how you assigned this probability.
 - (c) 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.
5. 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) Three of the hearts are white.
 - (b) Three are white, 2 are tan, 1 is pink, 1 is yellow, and 2 are green.
 - (c) Why is the probability of the event in part (a) larger than the probability of the event in part (b)?
6. 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) Using R, calculate the total number of possible winning numbers for the Powerball lottery when there are 45 white balls.
- (b) Using R, calculate the total number of possible winning numbers for the Powerball lottery when there are 55 white balls.
- (c) 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.
- (d) 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?