NORTHWESTERN UNIVERSITY

SCHOOL OF PROFESSIONAL STUDIES

R Lesson 4 - Solutions
MSPA 401 – Introduction to Statistical Analysis

1) Assume the fifty shoppers exit the store individually in random order.

   a) If one shopper is picked at random, what is the probability of picking a shopper who spent $40 or more dollars? What is the probability of picking a shopper who spent less than $10?

      With single shoppers, the probability is the number meeting the condition divided by the total number of shoppers.
      Probability of picking a shopper who spent >= $40 is 0.16.
      Probability of picking a shopper who spent < $10 is 0.1.

   For parts b), c) and d) it will be necessary to assume *sampling without replacement*. The way to do this is to determine the number of pairs of shoppers who meet the condition and divide by the total number of shopper pairs. The latter is found by the combinations rule of page 176 of Triola. The former is determined by the specified condition.

   b) If two shoppers are picked at random, what is the probability the pair will include a shopper who spent $40 or more dollars and one who spent less than $10?

      In this case, multiply the number of shoppers spending less than $10 times the number spending $40 or more dollars. This gives the number of pairs meeting the condition. Then divide by the total number of possible pairs.

      The probability of this joint event is 0.0327.

   c) If two shoppers are picked at random, what is the probability the pair will include two shoppers who spent no less than $10 and no more than $40?

      It is necessary to count the number of pairs of shoppers that satisfy the condition. This requires use of the combinations rule applied to the total number of shoppers spending between $10 and $40 dollars. Once this number of pairs is determined, divide that number by the total number of possible pairs of shoppers. This gives a probability of 0.5437.

   d) If four shoppers are picked at random, what is the probability one shopper will have spent less than $10, one shopper will have spent $40 or more dollars and two shoppers will have spent no less than $10 and no more than $40?

      Similarly to c) above, count the subsets of shoppers that satisfy the condition and divide that number by the total number of possible subsets of shoppers. The subsets must include 4 shoppers. Division gives a probability of 0.1157.

   e) If we know a randomly picked shopper has spent more than $30, what is the probability that shopper has spent more than $40?

This is a conditional probability. First count the number of shoppers that have spent more than $40 and divide by the number spending more than $30. The resulting probability is 0.4706.

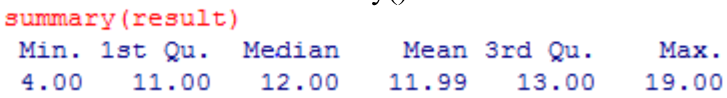
2) Use R to answer the following questions.

a) Draw 100 samples with replacement of size 22 from the 365 integers (i.e. 1,2,...,365). Count the number of samples in which one or more of the numbers sampled is duplicated. Divide by 100 to estimate the probability of such duplication occurring. (If 22 people are selected at random, what is the probability of two or more matching birthdays?)

Simulation results should provide estimated probabilities in the neighborhood of 0.5. With a set.seed(1234) and 100 samples the estimate was 0.53. With 10,000 samples, the result was 0.4817 which is closer to the truth. With a sample of 23 people the probability would be just greater than 0.5.

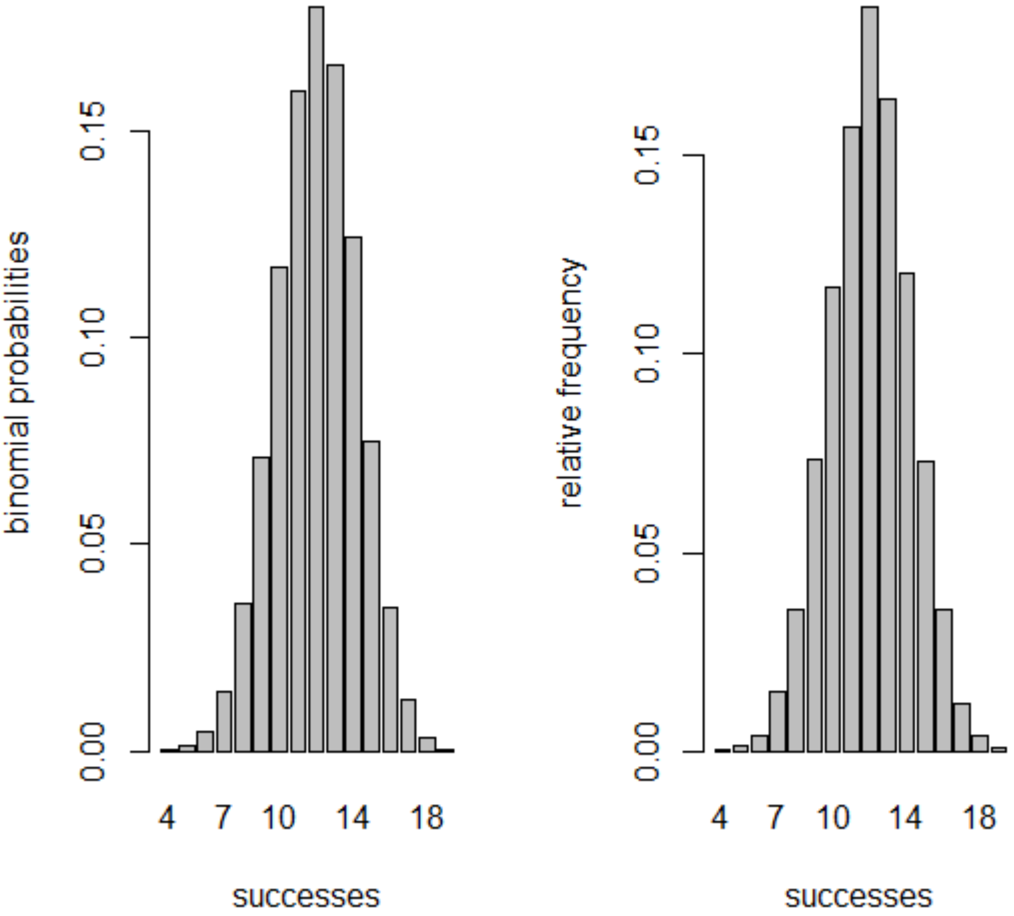
b) Suppose that 60% of marbles in a bag are black and 40% are white. Generate a random sample of size 20 with replacement using uniform random numbers. For the numbers in each sample, if a random number is 0.6 or less, code it as a 1. If it is not 0.6 or less code it a zero. Add the twenty coded numbers. Do this 50 times and calculate the proportion of times the sum is 11 or greater. What have you estimated? Expand the number of trials to 10,000. The exact binomial probability is 0.755 and the expectation is 12.

With a set.seed(1234) the calculated proportion was 0.66. Expanding the number of trials to 10,000 resulted in an estimate of 0.754. The estimated expectation was 11.99. Results from summary() are:



```
summary(result)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   4.00   11.00   12.00   11.99   13.00   19.00 
```

The simulation estimate is approximating values obtained from the binomial distribution with n=20 and p=0.6. Note the two distributions shown below. The one with the relative frequency label is from the simulation, and the other is an exact binomial calculation.


binomial probabilities
0.00
0.05
0.10
0.15
4
7
10
14
18
successes
relative frequency
0.00
0.05
0.10
0.15
4
7
10
14
18
successes