

Statistics 341
Fall 2008 - Assignment #4
Due Wednesday, October 15

1. An individual claims to have extrasensory perception (ESP). As a test, a fair coin is flipped ten times, and he is asked to predict in advance the outcome.
 - (a) If the individual does not have ESP, what is the probability he will correctly guess the outcome on any given coin flip?
 - (b) If the individual does not have ESP, what is the probability that he will correctly guess 7 or more out of 10 flips?
 - (c) If the individual does not have ESP, what is the probability that he will correctly guess 9 or more out of 10 flips?
 - (d) The test is conducted and the individual guesses correctly on 6 out of 10 coin flips. What do you think about the individual's claim that he has ESP? Explain your answer.
 - (e) A new test is devised where the person is asked to predict the outcome in advance for 100 coin flips. The test is conducted and the individual guesses correctly on 60 out of 100 coin flips. What do you think about the individual's claim that he has ESP? Explain your answer.

2. Go to the course webpage and click on the link **Using R to Investigate the Binomial Distribution**. In this problem, you will use the information in this file to investigate the simulated observations from different binomial distributions. In doing this, you will be studying the effect of changes in both n and p on the binomial distribution.
 - (a) Use R to simulate 10,000 observations each from binomial distributions with parameters $n = 2, 10, 20$ and 200 and $p = 0.5$. Use R to calculate the means and variances and to make histograms of these 4 sets of observations. Use your output to answer the following questions.
 - i. What happens to the values of the mean as n increases? What values should these means be close to?
 - ii. What happens to the values of the variance as n increases? What values should these variances be close to?
 - iii. Describe the changes you see in the 4 histograms of the observations as n increases.
 - (b) Use R to simulate 10,000 observations each from binomial distributions with parameters $n = 4, 20, 40$ and 400 and $p = 0.25$. Use R to calculate the means and variances and to make histograms of these 4 sets of observations. Use your output to answer the following questions.
 - i. What happens to the values of the mean as n increases? What values should these means be close to?
 - ii. What happens to the values of the variance as n increases? What values should these variances be close to?
 - iii. Describe the changes you see in the 4 histograms of the observations as n increases.

- (c) Use R to simulate 10,000 observations each from binomial distributions with parameters $n = 10, 50, 100$ and 1000 and $p = 0.1$. Use R to calculate the means and variances and to make histograms of these 4 sets of observations. Use your output to answer the following questions.
- What happens to the values of the mean as n increases? What values should these means be close to?
 - What happens to the values of the variance as n increases? What values should these variances be close to?
 - Describe the changes you see in the 4 histograms of the observations as n increases.
- (d) Use R to simulate 10,000 observations each from binomial distributions with parameters $n = 100, 500, 1000$ and 10000 and $p = 0.01$. Use R to calculate the means and variances and to make histograms of these 4 sets of observations. Use your output to answer the following questions.
- What happens to the values of the mean as n increases? What values should these means be close to?
 - What happens to the values of the variance as n increases? What values should these variances be close to?
 - Describe the changes you see in the 4 histograms of the observations as n increases.
- (e) In each of the 4 problems above, there is the same relationship between the four values of n and p . What is this relationship?
3. In a playlist of 10 songs, 3 songs are from the band Queen, 4 songs are from the band Poisson, and 3 songs are from Jimmy Buffett. The playlist is shuffled and the first three songs are played. Let Y denoted the number of Jimmy Buffett songs played in the first three songs played.
- Find the probability distribution function for Y .
 - Find the mean and variance for Y .
4. An urn contains N marbles, of which 50% are green, 20% are blue and 30% are red. Three marbles are to be drawn from the urn. Let the random variable Y be the number of green marbles selected from the urn.
- Assume $N = 10$ and the marbles are drawn from the urn without replacement.
 - What is the distribution of the random variable Y ?
 - Use R to calculate the probability distribution function values for the random variable Y .
 - What is the expected value of the random variable Y ?
 - What is the variance of the random variable Y ?
 - Still assume $N = 10$ but that the marbles are drawn from the urn with replacement. This means that the first marble is drawn, its color noted and then replaced before the second marble is drawn from the urn.
 - What is the distribution of the random variable Y ?
 - Use R to calculate the probability distribution function values for the random variable Y .

- iii. What is the expected value of the random variable Y ?
 - iv. What is the variance of the random variable Y ?
- (c) Now assume $N = 1000$ and the marbles are drawn from the urn without replacement.
- i. What is the distribution of the random variable Y ?
 - ii. Use R to calculate the probability distribution function values for the random variable Y .
 - iii. What is the expected value of the random variable Y ?
 - iv. What is the variance of the random variable Y ?
- (d) Still assume $N = 1000$ but that the marbles are drawn from the urn with replacement.
- i. What is the distribution of the random variable Y ?
 - ii. Use R to calculate the probability distribution function values for the random variable Y .
 - iii. What is the expected value of the random variable Y ?
 - iv. What is the variance of the random variable Y ?
- (e) Use your answers to above to discuss the effect of the size of N and the selection method (without replacement and with replacement) have on the probability distribution function, the expected value and variance of the random variable Y .
5. In a Newsweek article from January, 2005, Steven Levy reported on the perceived non-random behavior of the random shuffle feature on the iPod. As one example, he reported on the lack of favoritism for a particular song he purchased online. “Months after I bought *Wild Thing* from the iTunes store, I’m still waiting for my iPod to cue it up.” If you listen to an entire shuffled playlist, you will hear all songs on the playlist just once. However, this scenario almost never happens since people will listen only to the first n songs in an N song playlist before reshuffling the same playlist or choosing another.
- (a) For one shuffled playlist where you listen to the first n songs out of N total, what is the probability that one song will be played?
 - (b) Assume that successive shuffles of the playlist are independent. Let Y denote the number of shuffles required in order to hear a particular song once when you listen to the first n songs out of N total in each shuffle. What is the distribution of Y ? What is the mean of Y ? What is the variance of Y ?
 - (c) Assume the playlist has $N = 3000$ songs. What is the mean and variance of Y if you listen to the first $n = 20, 50, 100$ songs before reshuffling the same playlist or choosing another?
6. In a library of 3000 songs, 50 are from a particular group. The library is shuffled and played until the first song from this group appears. Let Y denote the number of the song in the shuffle where the first song from this group is played. For example, if the first song from this group is the 4th song played in the shuffle, $Y = 4$.
- (a) What is the distribution of the random variable Y ? What values can the random variable Y take?
 - (b) Use R to calculate the value of $p(y)$ for each possible value of y . What do you notice about the value of $p(y)$ as y increases?

- (c) Use R to find the theoretical value of the median of Y . Define what each of these values means in words.
7. The number of bacteria colonies of a certain type in sample of polluted water has a Poisson distribution with a mean of 2 per 1-cubic centimeter.
- (a) Find the probability that a 1-cubic centimeter sample taken from this water will have at least two bacteria colonies.
- (b) If 10 independent 1-cubic centimeter samples are taken from this water, find the probability that exactly 3 of these samples will contain two or more bacteria colonies.
8. In class, we learned a Poisson distribution has the same value for the mean and variance. Use R to conduct an investigation of the shape of the Poisson distribution for different values of the parameter λ . What do you notice about the shape of the distribution relative to the parameter λ ?