

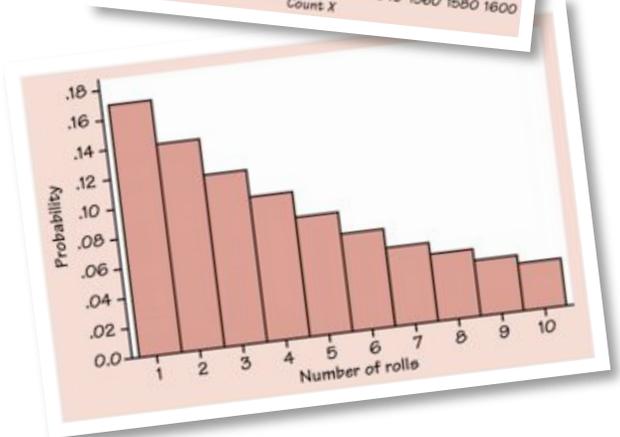
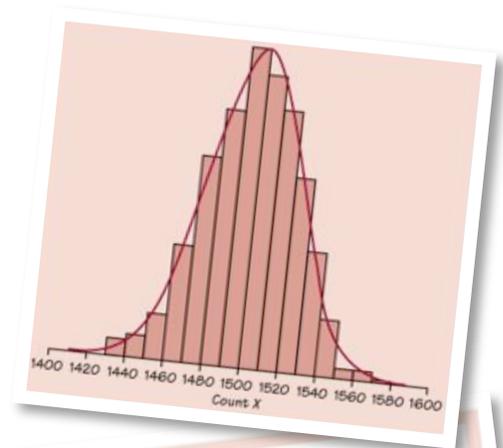
CHAPTER 8

THE BINOMIAL AND GEOMETRIC DISTRIBUTIONS

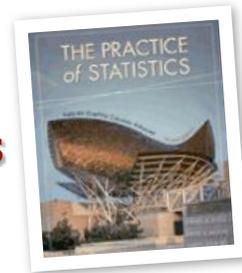
In Chapter 7, we learned that a random variable is one whose values are the numeric outcomes of random phenomena. In practice, we often encounter situations where there are two outcomes of interest. These situations can be described by binomial and geometric distributions. In this chapter, we will use what we have learned about probability and random variables to complete our foundational study of inference.

BINOMIAL AND GEOMETRIC DISTRIBUTIONS:

- 8.1: The Binomial Distributions
- 8.2: The Geometric Distributions



AP STATISTICS CHAPTER 8: BINOMIAL AND GEOMETRIC DISTRIBUTIONS



"I shall persevere until I find something that is certain -- or, at least until I find for certain that nothing is certain."

-RENE DESCARTES

Tentative Lesson Guide					
Date		Stats	Lesson	Assignment	Done
Wed	1/3	8.1	Binomial Distributions	Rd 439-449 Do 1-5, 9-13	
Thu	1/4	8.1	Binomial Distributions	Rd 450-459 Do 27-36	
Fri	1/5	Quiz	Quiz 8.1 (Take-Home?)	Rd 464-473 Do 37-40	
Mon	1/8	Rev	Geometric Distributions	Review Worksheet	
Tues	1/9	Rev	Review Chapter 8	Review Worksheet	
Wed	1/10	Exam	Exam Chapter 8	Online Quiz Due	
Thurs	1/11	9.1	Sampling Distributions	Penny Sampling Activity	
Fri					
Start Studying for your Semester 1 Final Exam!					

Note:

The purpose of this guide is to help you organize your studies for this chapter.

The schedule and assignments may change slightly.

Keep your homework organized and refer to this when you turn in your assignments at the end of the chapter.

Class Website:

Be sure to log on to the class website for notes, worksheets, links to our text companion site, etc.

<http://web.mac.com/statsmonkey>

Don't forget to take your online quiz!. Be sure to enter my email address correctly!

<http://bcs.whfreeman.com/yates2e>

My email address is:

jmmolesky@isd194.k12.mn.us



Chapter 8 Objectives and Skills:

These are the expectations for this chapter. You should be able to answer these questions and perform these tasks accurately and thoroughly. Although this is not an exhaustive review sheet, it gives a good idea of the "big picture" skills that you should have after completing this chapter. The more thoroughly and accurately you can complete these tasks, the better your preparation.

The Binomial Distribution

The binomial distribution is useful in situations where there are two outcomes of interest, such as success or failure.

The Binomial Setting:

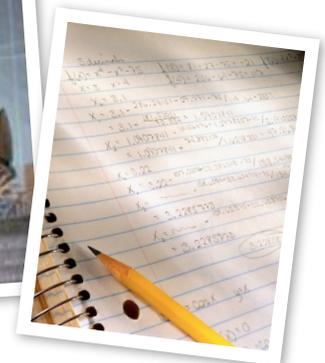
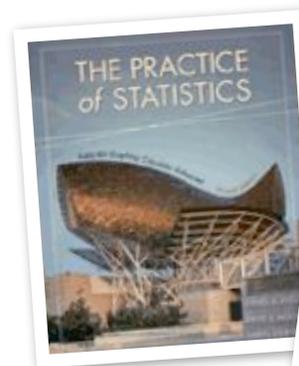
- Each observation is in one of two categories: success or failure.
 - A fixed number, n , of observations.
 - Observations are independent.
 - The probability of success is the same for each observation.
- If a count, X , has a binomial distribution with number of observations, n , and probability of success, p , then:
- Mean(X) = $\mu_X = np$
 - Standard Deviation(X) = $\sqrt{np(1-p)}$
 - $P(X=k) = \binom{n}{k}(p^k)(1-p)^{n-k}$
 - On the TI-83, `binompdf(n,p,k)`
 - $P(X \leq k) = \text{binomcdf}(n,p,k)$
- As the number of observations increases, a normal approximation can be used to find binomial probabilities. Generally, this approximation will be accurate when both np and $n(1-p)$ are greater than 10.

The Geometric Distribution

The Advanced Placement Statistics Syllabus states that students need only know how to obtain geometric probabilities through simulation. The geometric setting is somewhat similar to the binomial setting, the basic difference being that the geometric setting does not have a fixed number of observations.

The Geometric Setting:

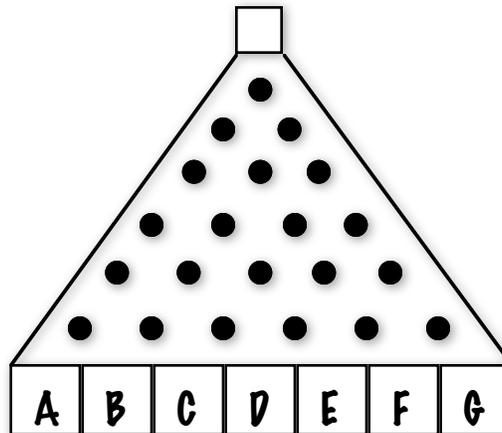
- Each observation is in one of two categories: success or failure.
 - The probability of success is the same for each observation.
 - Observations are independent.
 - The variable of interest is the number of trials required to obtain the first success.
- The mean of a geometric count is $\mu = 1/p$
- The standard deviation of a geometric count is $\sigma = \sqrt{(1-p)/p^2}$
- The probability it will take more than n trials for an event to occur is $(1-p)^n$



8.1: Introduction to Binomial Distributions - "Plinko"

Java version of "Plinko", <http://www.math.psu.edu/dlitttle/java/probability/plinko/index.html>

In the game of Plinko, a ball is dropped down a chute riddled with pegs. When it hits a peg, it has an equal chance of falling to the left or to the right. It then proceeds down to another row of pegs, hitting on of them and falling left or right to the next row. The board below shows the general idea for a Plinko board with 7 bins.



Our goal is to determine the probability that a ball settles in slot "D". To do so, we'll start by using a simulation method to get an estimate. As our ball moves down the game board, it will hit exactly one peg in each row. Each time it hits a peg, it has a 0.5 chance of falling to the right and a 0.5 chance of falling to the left. A ball will land in bin "D" if it falls to the right three times and to the left three times in any order.

Let the random variable $X = \#$ of times the ball falls to the right. Our goal is to find $P(X=3)$

1. Using your random number generator or a random number table, simulate 40 Plinko games. To do so, enter $\text{RandInt}(1,2,6)$ on your calculator. 1=Left, 2=Right. Each selection of 6 digits represents 1 trial. Simulate 20 trials and record the number of times the ball lands in bin "D" below:

Lands in Bin D	
Doesn't Land in Bin D	

Combine your results with those of other students in the class. Calculate the relative frequency of landing in bin "D":

$$\frac{\text{Lands in "D"}}{\text{Total Trials}} = \frac{\quad}{\quad} =$$

2. Determine the total number of possible outcomes for the Plinko game and list them systematically in the sample space below:

LLLLLL	RLLLLL	RRLLLL	RRRLLL	RRRRLL	RRRRRL	RRRRR
--------	--------	--------	--------	--------	--------	-------

3. Complete the probability distribution table for the random variable $X = \#$ times the ball falls to the right

X	0	1	2	3	4	5	6
P(X)							

Did the results of your simulation come close to the theoretical value for $P(X=3)$?

Flipping a coin six times and letting heads represent falling to the right would simulate this situation in much the same way as our random number generator.

The characterizing features of this experiment are as follows:

- ☉ There are two outcomes: Left or Right - heads and tails or evens and odds {success and failure}
- ☉ Each coin is flipped 6 times or 6 digits are drawn for a single trial.
- ☉ The flips or draws are independent - one outcome has no influence on the next outcome.
- ☉ The probability of success (falling to the right) is the same for each trial.

A situation in which these four conditions are satisfied is called a **binomial setting**.

Binomial Setting

-
-
-
-

4. Suppose 5 Plinko balls are dropped down the board in succession. Find the probability that all of them land in slot “D”. Find the probability that exactly two of them land in slot “D”.

5. If a ball landing in slot “A” or “G” pays \$50, a ball landing in slot “B” or “F” pays \$25, a ball landing in slot “C” or “E” pays \$10, and a ball landing in slot “D” pays \$5, find the expected winnings when 5 balls are dropped. What is the standard deviation of the total amount won?

8.1: Binomial Distributions

As we noted in the Plinko Example, a Binomial Setting is one in which four conditions are satisfied--Each observation has two outcomes, Trials are independent, There is a fixed number of observations, and P(success) is the same for all observations.

In practice, when we are studying a random variable that is binomial, we will use the notation “**X is B(n,p)**” where n represents the fixed number of observations and p represents the probability of success.

Example: Suppose you receive a shipment of five monkey-scooters {nothing is funnier than a monkey on a scooter.} Each scooter has a 15% chance of not working. What is the probability 3 or more scooters in your shipment will be defective?

Why does this situation satisfy the binomial setting?

X = # of defective scooters out of a shipment of 5 X is B(5 , 0.15)

We know X can take on any value from 0 to 5 (why?). To find the probability of any individual outcome, we just need to use a little common sense.

Let's consider the case where three scooters are defective. The probability of that happening is:

$$(0.15)(0.15)(0.15)(0.85)(0.85) = 0.002438$$

However, that probability describes the unique situation where the *first three* scooters are defective. What is the probability the *last three* are defective? What about the first, third, and fifth?

Note, each of these probabilities is the same...so, the probability of three defective scooters must take into account the number of ways in which 3 out of 5 can be defective.

The number of possible combinations of 3 defective scooters out of 5 is

$${}_5C_3 = \binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

Therefore, the probability of having three defective scooters is $P(X=3) = 10(.002438) = .02348$

Binomial Probability

If X is $B(n,p)$, the possible values of X are $0, 1, 2, 3, \dots, n$. Let k be one of these values...

$$\begin{aligned} P(X = k) &= (\text{number of combinations of } k \text{ successes})(P(k \text{ successes}))(P(n - k \text{ failures})) \\ &= {}_n C_k (p^k)(1 - p)^{n-k} \end{aligned}$$

For example, to find $P(3 \text{ defective scooters out of } 5)$, $P(X = 3) = ({}_5 C_3)(0.15^3)(0.85)^2 = 0.02438$

Use the formula above to complete the probability distribution for $X = \#$ defective scooters out of 5.

X	0	1	2	3	4	5
P(X)						

Sketch the probability histogram for X :

Find: $P(3 \text{ or more scooters are defective}) = P(X \geq 3) =$
 $P(\text{no more than 1 scooter is defective}) = P(X \leq 1) =$
 $P(\text{odd number of defective scooters}) =$

TI Note: Many binomial calculations can be done on your calculator...look under your DISTR menu:

$$P(X=k) = \text{binompdf}(n, p, k)$$

$$P(X \leq k) = \text{binomcdf}(n, p, k)$$

Example: Suppose I am successful on 75% of my penalty shot attempts. What is the probability I will make 5 or fewer of my next 9 attempts? Define a binomial random variable and calculate the probability.

$X =$

X is $B(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$

$P(X \leq 5) =$

8.1: Mean and Standard Deviation of Binomial Random Variables

If a Random Variable X is $B(n,p)$, what is the expected number of successes? How much variability will there be from trial to trial? We can guess the answer to the first question in most situations, but with a little common sense, we can derive formulas to calculate the mean and standard deviation of X .

Let $X = \#$ of successes in a single trial of a binomial setting

X	0	1		$E(X)$	$= \mu_X = 0(1-p) + 1(p) = p$
$P(X)$	$(1-p)$	p		$VAR(X)$	$= \sigma^2_X = (0-p)^2(1-p) + (1-p)^2(p)$
					$= p^2(1-p) + p(1-p)^2$
					$\sigma^2_X = p(1-p)(p+(1-p)) = p(1-p)$
					$\sigma_X = \sqrt{p(1-p)}$

If we consider n independent trials of X , then $\mu = np$ and $\sigma = \sqrt{np(1-p)}$... Why?

Mean and Standard Deviation of a Binomial Random Variable

If a Random Variable X is $B(n,p)$, the mean and standard deviation of X are:

$$\mu_X =$$

$$\sigma_X =$$

The Normal Approximation For Binomial Distributions

As the number of trials increases, the formulas for binomial probabilities become quite cumbersome. Thankfully, the following fact can help us out...

As the number of trials n gets larger, the binomial distribution of X gets closer to a normal distribution.

$$B(n,p) = N(np, \sqrt{np(1-p)})$$

When n is large, we can use normal probability calculations to approximate hard-to-calculate binomial probabilities. As a rule of thumb, we can use the normal approximation when $np \geq 10$ and $n(1-p) \geq 10$.

Example: Suppose 60% of adults have credit card debt. If we survey 2500 adults, what is the probability more than 1520 would have credit card debt?

$$X = \# \text{ of adults who have credit card debt out of } 2500$$

$$X \text{ is } B(2500, 0.6)$$

We want $P(X > 1520)$. What would the probability histogram look like? Can we use a normal approximation? What would the mean and standard deviation of the normal approximation be? Use this to calculate the probability.

8.2: Geometric Distributions

When dealing with Binomial settings, we have a fixed number of trials and the random variable X counts the number of successes within that fixed number. However, a number of situations of interest involve counting how long it takes for something to happen. For example, how long does it take to roll 2 dice until a sum of 7 comes up? How many flips of a coin does it take until you get a head? How many licks does it take to get to the center of a Tootsie Pop®? Ok, maybe not that one, but you get the idea.

Geometric Setting:

A random variable X is *Geometric* if



Generally, geometric random variables are studied through simulation, but we can calculate probabilities using a little common sense.

Example: On the leeward side of the island of Oahu in the small village of Nanakuli, about 80% of the residents are of Hawai'ian ancestry (The Honolulu Advertiser). Suppose you fly to Hawai'i and visit Nanakuli. What is the $P(\text{first villager you meet is Hawai'ian})$? What is the $P(\text{you don't meet a Hawai'ian until the second villager})$? Etc?

Let $X = \#$ of villagers you must meet until you meet an actual Hawai'ian

$$P(X=1) =$$

$$P(X=2) =$$

$$P(X=3) =$$

$$P(X=4) =$$

$$P(X=5) =$$

Sketch the probability distribution of X and note its shape. All geometric random variables will exhibit this general shape...why?

Handy Dandy Formulas:

If X is a Geometric random variable with $P(\text{success}) = p$

$$P(X = n) = \qquad \mu_X =$$

$$P(X > n) = \qquad \sigma_X =$$

Find the probability it will take more than 4 villagers to meet a Hawai'ian.

Find the average number of villagers it will take to meet a Hawai'ian. How much variability is there in the number of villagers required to meet a Hawai'ian?

Bonus Topic: Introduction to Poisson Distributions

If we examine the Binomial distribution as the number of trials n gets larger and larger while the $P(\text{success})$ p gets smaller and smaller, we get the Poisson Distribution. Siméon Denis Poisson (1781-1840) was a French mathematician who studied probabilities of rare events that occur infrequently. His distribution applies to such topics as accident rates, arrival times, defect rates, etc. As with the Binomial and Geometric distributions, we assume only two outcomes, independence, and equal probabilities of success for all events. We are interested in computing the probability of r occurrences in a specified interval.

Poisson Distribution:

Let λ (Greek letter lambda) represent the mean number of successes of an event over a specified time. Let r be the number of successes ($r = 0, 1, 2, 3, \dots$) in a corresponding interval of time.

The probability of r successes within a specified interval is:

$$P(r) = \frac{e^{-\lambda} \lambda^r}{r!}$$

The Poisson Distribution can be used to model the probability of a number of different events including the arrival of cars to a gas station, planes at an airport, calls to a fire station, births of babies, and even fish arriving on a fisherman's line...

Example: Pyramid Lake is located in Nevada on the Paiute Indian Reservation. In addition to its natural beauty, the lake contains some of the world's largest cutthroat trout. The Paiute Nation uses highly trained marine biologists to study and maintain their famous fishery. *Creel Chronicle* (Vol 3, No 2) noted the fact that the November catch averaged 0.667 fish per hour.

Suppose you fish Pyramid Lake for 7 hours during the month of November. Use the information above to find the probability distribution for the number of fish you catch in the 7 hour period.

$r = \#$ of fish caught in a 7 hour period

Since you can expect an average of 0.667 fish per hour, you should be able to calculate $\lambda =$ the average number of fish you would expect over the 7 hour period. $\lambda =$ _____

$$P(r = 0) = \frac{e^{-4.7} (4.7)^0}{0!} =$$

$$P(r = 1) =$$

$$P(r = 2) =$$

$$P(r = 3) =$$

$$P(r = 4) =$$

Use a little common sense to find the probability you will catch four or more fish in the seven hour period.

Practice: Binomial and Geometric Distributions Problems

1. The Los Angeles Times (Dec. 13, 1992) reported that 80% of airline passengers prefer to sleep on long flights rather than watch movies, read, etc. Consider randomly selecting 25 passengers from a particular long flight. Define a random variable X and answer the following questions.
 - a) Calculate and interpret $P(X=12)$
 - b) Calculate and interpret $P(X=25)$
 - c) Calculate and interpret $P(X\geq 20)$
 - d) Calculate and interpret the mean and standard deviation of X

2. Sophie is a dog who loves to play catch. Unfortunately, she isn't very good, and the $P(\text{catches a ball})=0.1$. Define X =# tosses required for Sophie to catch the ball.
 - a) Calculate and interpret $P(X=2)$
 - b) Calculate and interpret $P(X\geq 3)$
 - c) Calculate and interpret the mean and standard deviation of X

3. You are to take a multiple choice exam consisting of 100 questions with five possible responses to each. Suppose you have not studied and decide to guess randomly on each question. Let X =# correct responses on the exam.
 - a) What is your expected score on the exam?
 - b) Compute and interpret the standard deviation of X .
 - c) If 50% is passing, what is the probability you will pass the exam by guessing?

4. In his book, *To Engineer is Human: The Role of Failure in Successful Design*, Henri Petroski indicates a large number of structural failures are the result of hairline cracks. As these cracks increase in number, the strength of the structure decreases. Suppose a retaining wall has an average of 4.2 cracks per 30-foot section. What is the probability of having 3 cracks over a 50 ft section of the wall? Suppose the strength is severely compromised if there are more than 8 cracks over a 50 foot section. Provide a statistical argument as to whether or not we should be concerned about the retaining wall.

5. Suppose 5% of cereal boxes contain a prize. You are determined to buy cereal boxes until you win a prize.
 - a) What is the probability you will have to buy at most 2 boxes?
 - b) What is the probability you will have to buy exactly 4 boxes?
 - c) What is the probability you will have to buy more than 4 boxes?

6. If the temperature in Florida falls below 32°F there is a chance the citrus crop will be damaged. Suppose the probability an orange tree will be damaged in cold weather is .1. Use a normal approximation to determine the approximate probability more than 225 trees will be damaged in an orchard of 2000 trees. Be sure to provide evidence that a normal approximation is appropriate. Compare your probability to the actual binomial calculation.