Probability	Name:
Ch.4, sections 4.2 & 4.3 - Binomial and Geometric Distributions	Date: Pd:

4.2. - What is a binomial distribution? How do we find the probability of success?

Suppose you have three daughters. If we assume that girls and boys are equally likely, how unusual is it to have three children that are all girls? We will use simulation to approximate this value. We will consider each number a **success** if it is a zero (girl) and a **failure** if it is a 1 (boy).

What is the probability of success? Failure? How could we do this simulation using a random number table?

1) Use your calculator to randomly produce three integers from 0 - 1. Run the simulation 40 times. Complete the table below using tally marks in the appropriate row.

3 Girls		Total:
Not 3 Girls		Total:

2) Combine your results with that of your group members. What is the relative frequency of the event {3 girls}?

3) You just calculated the **Empirical** (experimental) probability of having exactly 3 girls. How does that value compare to the **Theoretical** probability?

In the example above, there were only 2 possible outcomes of interest (3 girls, not 3 girls). Give three examples where this is also the case (only 2 outcomes).

1) 2) 3)

Binomial Setting: The following 4 conditions must be satisfied (it is **binomial** only if it **FITS**):

1) There is a **f**ixed number *n* of observations.
2) The *n* observations are ALL **i**ndependent.
3) Each observation falls into one of just **t**wo categories (success or failure).
4) The probability for success, *p*, is the **s**ame for each observation.

Binomial Random Variable: the number (X) of successes

Ex: The number of times there were 3 girls

Binomial Distribution: The distribution of the count X of successes in the binomial setting with parameters *n* and *p*. The parameter *n* is the number of observations, and *p* is the probability of a success on any one observation. The possible values of X are whole numbers from 0 to *n*. As an abbreviation, we say that X is

B(n,p).

Ex: Decide if the following are binomial experiments. If it is binomial, specify *n, p, q* (*q* = 1-*p*), and all possible values of the random variable *x.* If it is not binomial, explain why it does not FITS.

a) A certain surgical procedure has a 75% chance of success. A doctor performs the procedure on 5 patients. The random variable represents the number of successful surgeries.

b) A jar contains five red marbles, nine blue marbles, and six green marbles. You randomly select three marbles from the jar, without replacement. The random variable represents the number of red marbles.

Ex: Suppose microfracture knee surgery has a 75% chance of success on patients with degenerative knees. The surgery is performed on six patients. Find the probability of the surgery being successful on exactly two patients. Exactly four patients. More than four patients.

a) Is this a binomial distribution (FITS)?

b) Identify the type of distribution, *p*, *q*, and *n.*

c) Formula? $P(x) = {}_nC_x p^x q^{n-x}$ or $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ or on the calculator: **binompdf(*n,p,x*)**
Learn to use the calculator method, it will make things much easier. There is also a table.
*pdf stands for "probability distribution function". $\binom{n}{k}$ *is known as the "binomial coefficient"*

d) Find *P*(X=2)

e) Find *P*(X=4)

f) Find *P*(X>4)

Verify the problems (d) and (e) above using **binompdf(*n,p,x*)**. 2nd → Vars on calc.

*For letter (f), we wanted to know P(X>4). We went through the process of adding up P(5)+P(6). There is a shortcut. If we use **binomcdf(n,p,x)**, c is for *continuous*, we can enter **binomcdf(6,0.75,4)**. This **ONLY** sums probabilities up to, and including, the value of x. If we subtract this probability from 1, we will get P(X>4).

HW: 1) Identify if the following are binomial experiments. If it is binomial, explain why it FITS. If it is not, explain why.

a) Cyanosis is the condition of having bluish skin due to insufficient oxygen in the blood. About 80% of babies born with cyanosis recover fully. A hospital is caring for five babies born with cyanosis. The random variable represents the number of babies that recover fully.

b) From past records, a clothing store finds that 26% of the people who enter the store will make a purchase. During a one-hour period, 18 people enter the store. The random variable represents the number of people who do not make a purchase.

c) A survey asks 1000 adults, "Do tax cuts help or hurt the economy?" Twenty-one percent of those surveyed said tax cuts hurt the economy. Fifteen adults who participated in the survey are randomly selected. The random variable represents the number of adults who think that tax cuts hurt the economy.

d) A state lottery randomly chooses 6 balls numbered from 1 to 40. You choose six numbers and purchase a lottery ticket. The random variable represents the number of matches on your ticket to the numbers drawn in the lottery.

2) Fifty-nine percent of men consider themselves professional baseball fans. You randomly select 10 men and ask each if he considers himself a professional baseball fan. Find the probability that the number who consider themselves baseball fans is... ***Make sure to check if it FITS.**

a) exactly 8 b) at least 8 c) less than 8

3) Suppose that Randy randomly guesses on each question of a 50-item true-false test. Find the probability that Randy passes if... ***Make sure to check if it FITS.**

a) a score of 25 or more correct is needed to pass.

b) a score of 30 or more correct is needed to pass. c) a score of more than 30 correct is needed to pass.

4.2 - How do we find the mean and standard deviation of a binomial distribution?

Rv: What does FITS stand for? When do we use it? What does SOCKS stand for? When do we use it?

Rv: Define standard deviation.

Rv: What is the probability he makes all 3 shots?

Population Parameters of a Binomial Distribution:

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard Deviation: $\sigma = \sqrt{npq}$

Ex: Ten percent of adults say oatmeal raisin is their favorite cookie. You randomly select 12 adults and ask each if his or her favorite cookie is oatmeal raisin. Find the probability that the number who say oatmeal raisin is their favorite cookie is...(don't forget to check to see if it FITS")

a) exactly four

b) at least four

c) less than 4

d) Find the mean and standard deviation of the example above. Interpret the results.

Ex: What if for the cookie problem, we randomly select 3000 people and want to find the probability that more than 265 people like oatmeal raisin cookies.
Do we really want to add the probabilities from X = 265 up to X = 3000? That would be ridiculous!

Compute the mean and standard deviation. Calculate the z-score, draw a picture, and find the area under the curve. Compare this value to the value you would find when using the calculator **(binomcdf(n,p,x)).**

Normal Approximation for Binomial Distributions: Suppose that a count X has the binomial dist. with *n* trials and success probability *p*. When *n* is **large**, the distribution of X is approximately normal, $N(np, \sqrt{npq})$

*As a rule of thumb, we will use the normal approximation when *n* and *p* satisfy $np \geq 10 \; and \; nq \geq 10$.

Ex: Check that the problem above satisfies the rule of thumb conditions for using the normal approximation to the binomial distribution.

Why > 10?

HW:

1) Your mail-order company advertises that it ships 90% of its orders within three working days. You select an SRS of 100 of the 5000 orders received in the past week for an audit. The audit reveals that 86 of these orders were shipped on time.

a) If the company really ships 90% of its orders on time, what is the probability that 86 or fewer in an SRS of 100 orders are shipped on time?

b) A critic says, "Aha! You claim 90%, but in your sample the on-time percentage is only 86%. So the 90% claim is wrong." Explain why the result of the sample does not refute the 90% claim. What % would we consider unlikely? (*Hint: Are *np* and *nq* both at least 10?)

2) You are planning a sample survey of small businesses in your area. You will choose an SRS of businesses listed in the telephone book's Yellow Pages. Experience shows that only about half the businesses you contact will respond.

a) If you contact 150 businesses, it is reasonable to use the normal distribution with $n = 150$ and $p = 0.5$ for the number X who respond. Explain why.

b) What is the expected number who will respond?

c) What is the probability that 70 or fewer will respond? (Use the normal approximation)

d) How large of a sample must you take to increase the mean number of respondents to 100?

3) Three in four adults says he or she has trouble sleeping at night. You randomly select five adults and ask each if he or she has trouble sleeping at night (*binompdf*(n,p)$\rightarrow L_1$).

a) Construct a binomial distribution

b) Graph the binomial distribution using a histogram.

c) Find the mean, variance, and standard deviation.

d) Describe the histogram (SOCKS).

4.3 - What is a geometric distribution?

Many actions in life are repeated until a success occurs. For instance, when you get your driver license you take the test until you pass. Can you think of any other examples?

The situation above can be represented by a **geometric distribution**.

Geometric Setting: The following 4 conditions must be satisfied (it is **geometric** only if it **FITS**):
1) The variable of interest is the number of trials required to obtain the **f**irst success.
2) The *n* observations are ALL **i**ndependent.
3) Each observation falls into one of just **t**wo categories (success or failure).
4) The probability for success, *p*, is the **s**ame for each observation.

Ex: Are the following binomial or geometric? Explain why.
1) The probability that a student passes the written test for a private pilot's license is 0.75. What is the probability that a student will fail the test on the first attempt and pass it on the second?

2) Forty percent of adults in the United States exercise at least 30 minutes a week. In a survey of 120 randomly chosen adults, people were asked, "Do you exercise at least 30 minutes a week?" What is the probability that exactly 50 of the people answer yes?

3) About 21% of Americans say they could not go one week without eating meat. You select, at random, 30 Americans. What is the probability that the first person who says he or she cannot go one week without eating meat is the fifth person selected?

Rule for Calculating Geometric Probabilities: If *X* has a geometric distribution with probability *p* of success and (1-*p*) of failure on each observation, the possible values of *X* are 1, 2, 3, …. If *n* is any one of these values, the probability that the first success occurs on the *n*th trial is

$P(X = n) = (1-p)^{n-1}p$ or on the calculator **geometpdf(*p,n*)**

Ex: Suppose that for a certain couple, the probability for a male birth is 0.47 and the probability for a female birth is 0.53. If the couple decides to continue having children until they have one son, what is the probability that they will have four children?

a) 0.05 b) 0.07 c) 0.19 d) 0.25 e) 0.47

HW:

1) For each of the following, determine if the experiment describes a geometric distribution (FITS). If not, explain why.

a) Flip a coin until you observe a tail.

b) Record the number of times a player makes both shots in a one-and-one foul-shooting situation. (In this situation, you get to attempt a second shot only if you make your first shot.)

c) There are 10 red marbles and 5 blue marbles in a jar. You reach in and, without looking, select a marble. You want to know how many marbles you will have to draw (without replacement), on average, in order to be sure that you have 3 red marbles.

2) Suppose we have data that suggest that 3% of a company's hard disk drives are defective. You have been asked to determine the probability that the first defective hard drive is the fifth unit tested.

a) Verify that this is a geometric setting (FITS).

b) Identify the random variable. What constitutes a success in this situation?

c) Answer the original question. What is the probability that the first defective hard drive is the fifth unit tested?

3) One in four people in the United States owns individual stocks. In a random sample of 12 people, what is the probability that the number owning individual stocks is

a) Exactly two

b) At least two

c) More than two

4) The probability you pass your driving test is 78%. What is the probability you pass your driving test on the first try? Third try?

4.3 - How do we calculate the mean and standard deviation of a geometric distribution?

Rv: In a recent year, four percent of trucks sold by a company had diesel engines. Consider a random sample of four trucks sold by the company.

a) Construct a binomial distribution.

b) Graph the distribution using a histogram

c) Describe the distribution.

If you are flipping a coin, you would **expect** to flip a heads every ________ times.
If you are rolling a die, you would **expect** to roll a 3 every ________ times.
If you are examining hard drives, you would **expect** to observe a defective one every ________ times.

The Mean and Standard Deviation of a Geometric Random Variable: If *X* is a geometric random variable with probability of success *p* on each trial, then the **mean** (expected value), of the random variable, that is, the expected number of trials required to get the first success, is $\mu = 1/p$. The variance of *X* is $\sigma^2 = (1-p)/p^2$.

What is the formula to find the standard deviation of a geometric random variable?

Find the mean and standard deviation for the 3 questions listed above.

1) Coin	**2) Dice**	**3) Hard Drives**

Ex: Suppose you are running the ring toss game at the fair. You win if the ring lands around the empty Coke bottle. Since you run the game, you get to play many times for free. After countless hours, you come to the conclusion that you win, on average, 1 out of every 12 games. Assume ring tosses are independent. Let *X* be the number of tosses until a win. Does this game satisfy a geometric distribution? Identify the expected number of tosses and standard deviation of winning this game.

Ex: What if we wanted to find the probability that it takes more than 15 tosses before you win the ring toss game?

HW:

1) The State Department is trying to identify an individual who speaks Farsi to fill a foreign embassy position. They have determined that 4% of the applicant pool are fluent in Farsi.

a) If applicants are contacted randomly, how many individuals can they expect to interview in order to find one who is fluent in Farsi?

b) What is the probability that they will have to interview more than 25 until they find one who speaks Farsi? More than 40?

2) A basketball player makes 80% of her free throws. We put her on the free-throw line and ask her to shoot free throws until she misses one. Let X = the number of free throws the player takes until she misses.

a) What assumptions do you need to make in order for the geometric model to apply? Verify that this is a geometric distribution.

b) What is the probability that the player will make 5 shots before she misses?

c) What is the probability that she will make at most 5 shots before she misses?

*How many shots can she take and still fall under the condition “at most 5 shots”?

3) Carla makes random guesses on her AP Stat multiple-choice test that has five choices for each question. We want to know how many questions Carla answers until she gets one correct.

a) What is the probability that Carla’s first correct answer occurs on problem 5?

b) What is the probability that it takes more than 4 questions before Carla answers one correctly?

4) Describe the difference between a binomial and geometric setting.

5) Dentists are increasingly concerned about the growing trend of local school districts to grant soft drink companies exclusive rights to install soda machines in schools in return for money - usually millions - that goes directly into school coffers (a strongbox or small chest for holding valuables). According to a recent study of the National Soft Drink Association, 62% of schools nationally already have such contracts. This comes at a time when dentists are seeing an alarming increase in horribly decayed teeth and eroded enamel in the mouths of teenagers and young adults. With ready access to soft drinks, children tend to drink them all day. That, combined with no opportunity to brush, leads to disaster, dentists say. Suppose that 20 schools around the country are randomly selected and asked if they have a soft drink contract. Find the probability that the number of "Yes" answers is

a) Exactly 8

b) At most 8

c) At least 4

d) Between 4 and 12, inclusive

6) A handful of years ago, Barry Bonds hit 73 home runs in the 153 games he played. Assume that his home run production stayed at that level the following season. What is the probability that he would hit his first home run, assuming home runs are independent

a) on the first game of the season

b) on the second game

c) within the first three games of the season

d) In your opinion, are hitting home runs independent or dependent events?

7) Describe the difference between **pdf** and **cdf** on the calculator.

8) Describe the distribution for the game Greed.

Greed Analysis