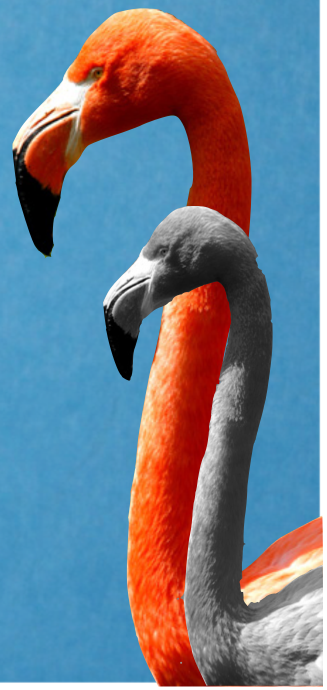
CHAPTER 6
Random Variables

6.3b
Binomial and Geometric Random Variables



The Practice of Statistics, 5th Edition
Starnes, Tabor, Yates, Moore

Bedford Freeman Worth Publishers

Binomial and Geometric Random Variables

Learning Objectives

After this section, you should be able to:

- ✓ DETERMINE whether the conditions for using a binomial random variable are met.
- ✓ COMPUTE and INTERPRET probabilities involving binomial distributions.
- ✓ CALCULATE the mean and standard deviation of a binomial random variable. INTERPRET these values in context.
- ✓ FIND probabilities involving geometric random variables.

Binomial Settings

When the same chance process is repeated several times, we are often interested in whether a particular outcome does or doesn't happen on each repetition. Some random variables count the number of times the outcome of interest occurs in a fixed number of repetitions. They are called **binomial random variables.**

A **binomial setting** arises when we perform several independent trials of the same chance process and record the number of times that a particular outcome occurs. The four conditions for a binomial setting are:

B • **B**inary? The possible outcomes of each trial can be classified as "success" or "failure."

I • **I**ndependent? Trials must be independent; that is, knowing the result of one trial must not tell us anything about the result of any other trial.

N • **N**umber? The number of trials n of the chance process must be fixed in advance.

S • **S**uccess? There is the same probability p of success on each trial.

How to Find Binomial Probabilities

How to Find Binomial Probabilities

Step 1: State the distribution and the values of interest. Specify a binomial distribution with the number of trials n, success probability p, and the values of the variable clearly identified.

Step 2: Perform calculations—show your work!

Do one of the following:

(i) Use the binomial probability formula to find the desired probability; or

(ii) Use `binompdf` or `binomcdf` command and label each of the inputs.

Step 3: Answer the question.

Pop Quiz!

Let's say I give my students a 10-item, multiple-choice quiz. The catch is, students must simply guess an answer (A through D) for each question. I used by calculator to randomly generate the answer key, so each answer has an equal chance to be chosen. Ella is one of the students in the class. Let X = the number of Ella's correct guesses.

(a) Verify that X is a binomial random variable

Binary? success = question answered correctly; failure = question not answered correctly ***Independent?*** The calculator randomly assigned correct answers to the questions, so knowing the result of one trial (question) should not tell you anything about the result of another trial. ***Number?*** 10 ***Success?*** Probability of success is p = 0.25

(b) Find the probability that Ella gets three questions correct.

P(X = 3) = *binompdf*(trials: 10, p: 0.25, X value 3) = 0.2503

There is a 25% chance that Ella will guess exactly 3 correct answers.

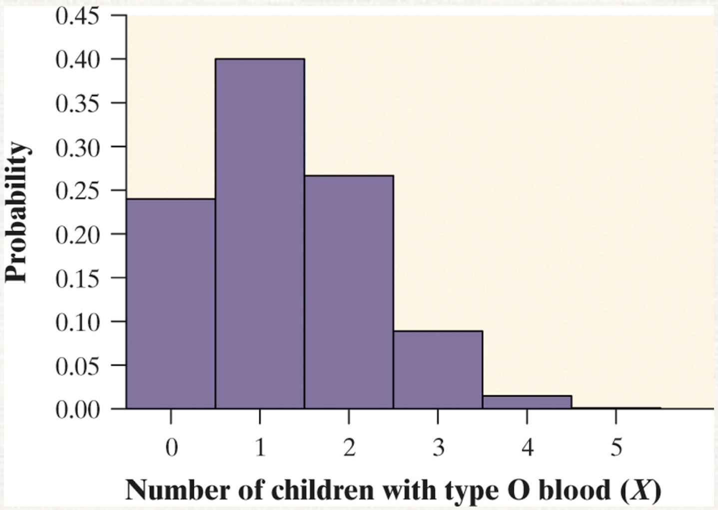
(c) A passing score would be at least 7. Would you be surprised if Ella passed? Compute an appropriate probability to support your answer.

$P(X \geq 7) = 1 - P(X \leq 6)$ = 1 – binomcdf(trials: 10, p: 0.25, X value 6) = 0.0035

There is only a 0.35% of guessing 7 or more questions correct, so we would be quite surprised if Ella was able to pass.

Mean and Standard Deviation of a Binomial Distribution

We describe the probability distribution of a binomial random variable just like any other distribution – by looking at the shape, center, and spread. Consider the probability distribution of X = number of children with type O blood in a family with 5 children.



x_i	0	1	2	3	4	5
p_i	0.2373	0.3955	0.2637	0.0879	0.0147	0.00098

Shape: The probability distribution of X is skewed to the right. It is more likely to have 0, 1, or 2 children with type O blood than a larger value.

Center: The median number of children with type O blood is 1. Based on our formula for the mean:

$$\mu_X = \sum x_i p_i = (0)(0.2373) + 1(0.39551) + ... + (5)(0.00098)$$
$$= 1.25$$

Spread: The variance of X is

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i = (0 - 1.25)^2(0.2373) + (1 - 1.25)^2(0.3955) + ... +$$
$$(5 - 1.25)^2(0.00098) = 0.9375$$

The standard deviation of X is $\sigma_X = \sqrt{0.9375} = 0.968$

Mean and Standard Deviation of a Binomial Distribution

Mean and Standard Deviation of a Binomial Random Variable

If a count X has the binomial distribution with number of trials n and probability of success p, the **mean** and **standard deviation** of X are

$$\mu_X = np$$

$$\sigma_X = \sqrt{np(1-p)}$$

<u>Note</u>: These formulas work ONLY for binomial distributions. They can't be used for other distributions!

Example: Mean and Standard Deviation

Mr. Bullard's 21 AP Statistics students did the bottled water vs. tap water experiment. If we assume the students in his class *cannot* tell tap water from bottled water, then each has a 1/3 chance of correctly identifying the different type of water by guessing. Let *X* = the number of students who correctly identify the cup containing the different type of water.

Find the mean and standard deviation of *X*.

Since *X* is a binomial random variable with parameters *n* = 21 and *p* = 1/3, we can use the formulas for the mean and standard deviation of a binomial random variable.

$$\mu_X = np$$
$$= 21(1/3) = 7$$

$$\sigma_X = \sqrt{np(1-p)}$$
$$= \sqrt{21(1/3)(2/3)} = 2.16$$

We'd expect about one-third of his 21 students, about 7, to guess correctly.

If the activity were repeated many times with groups of 21 students who were just guessing, the number of correct identifications would differ from 7 by an average of 2.16.

Tastes as good as the real thing?

The makers of a diet cola claim that its taste is indistinguishable from the full-calorie version of the same cola. To investigate, an AP® Statistics student named Emily prepared small samples of each type of soda in identical cups. Then she had volunteers taste each cola in a random order and try to identify which was the diet cola and which was the regular cola. Overall, 23 of the 30 subjects made the correct identification.

If we assume that the volunteers really couldn't tell the difference, then each one was guessing with a 1/2 chance of being correct. Let X = the number of volunteers who correctly identify the colas.

(a) Explain why X is a binomial random variable.

(b) Find the mean and the standard deviation of X. Interpret each value in context.

(c) Of the 30 volunteers, 23 made correct identifications. Does this give convincing evidence that the volunteers can taste the difference between the diet and regular colas?

Tastes as good as the real thing? results

(a) The chance process is each volunteer guessing which sample is the diet cola. *Binary?* Yes; guesses are either correct or incorrect. *Independent?* Yes; the results of one volunteer's guess tells us nothing about the results of other volunteers' guesses. *Number?* Yes; there are 30 trials. *Success?* Yes; the probability of guessing correctly is always 0.5. Because *X* is counting the number of successful guesses, *X* is a binomial random variable with $n = 30$ and $p = 0.50$.

(b) The mean of X is $\mu_X = np = 30(0.5) = 15$, and the standard deviation of X is $\sigma_X = \sqrt{np(1-p)} = \sqrt{30(0.5)(1-0.5)} = 2.74$. If this experiment were repeated many times and the volunteers were randomly guessing, the average number of correct guesses would be about 15. Also, the number of correct guesses would typically vary by about 2.74 from the mean.

(c) $P(X \geq 23) = 1 - P(X \leq 22) = 1 -$ *binomcdf*(trials: 30, p: 0.5, X value: 22) = 1 – 0.9974 = 0.0026. There is a very small chance that there would be 23 or more correct guesses if the volunteers couldn't tell the difference in the colas. Therefore, we have convincing evidence that the volunteers can taste the difference.

NASCAR cards and cereal boxes

In the "NASCAR Cards and Cereal Boxes" example from 5.1, we read about a cereal company that put 1 of 5 different cards into each box of cereal. Each card featured a different driver: Jeff Gordon, Dale Earnhardt, Jr., Tony Stewart, Danica Patrick, or Jimmie Johnson. Suppose that the company printed 20,000 of each card, so there were 100,000 total boxes of cereal with a card inside. If a person bought 6 boxes at random, what is the probability of getting no Danica Patrick cards?

Let *X* be the number of Danica Patrick cards obtained from 6 different boxes of cereal. Because we are sampling without replacement, the trials are not independent. The distribution of *X* is not quite binomial—but it is close. If we assume *X* is binomial with $n = 6$ and $p = 0.2$, then

$$P(X = 0) = \binom{6}{0}(0.2)^0(0.8)^6 = 0.262144$$

The actual probability, using the general multiplication rule is

P(no Danica Patrick cards) =

$$\left(\frac{80{,}000}{100{,}000}\right)\left(\frac{79{,}999}{99{,}999}\right)\left(\frac{79{,}998}{99{,}998}\right)\left(\frac{79{,}997}{99{,}997}\right)\left(\frac{79{,}996}{99{,}996}\right)\left(\frac{79{,}995}{99{,}995}\right) = 0.262134$$

These two probabilities are quite close!

Binomial Distributions in Statistical Sampling

The binomial distributions are important in statistics when we wish to make inferences about the proportion *p* of successes in a population.

Almost all real-world sampling, such as taking an SRS from a population of interest, is done without replacement. However, sampling without replacement leads to a violation of the independence condition.

When the population is much larger than the sample, a count of successes in an SRS of size *n* has approximately the binomial distribution with *n* equal to the sample size and *p* equal to the proportion of successes in the population.

10% Condition

When taking an SRS of size *n* from a population of size *N*, we can use a binomial distribution to model the count of successes in the sample as long as

$$n \le \frac{1}{10} N$$

Dead batteries

Almost everyone has one—a drawer that holds miscellaneous batteries of all sizes. Suppose that your drawer contains 8 AAA batteries but only 6 of them are good. You need to choose 4 for your graphing calculator. If you randomly select 4 batteries, what is the probability that all 4 of them will work?

Explain why the answer isn't

P(X = 4) = $\binom{4}{4}(0.75)^4(0.25)^0 = 0.3163$

Because we are sampling without replacement, the selections of batteries aren't independent. We can ignore this problem if the sample we are selecting is less than 10% of the population. However, in this case we are sampling 50% of the population (4/8), so it is not reasonable to ignore the lack of independence and use the binomial distribution. This explains why the binomial probability is so different from the actual probability.

P(4 working batteries) = $\left(\frac{6}{8}\right)\left(\frac{5}{7}\right)\left(\frac{4}{6}\right)\left(\frac{3}{5}\right) = .2143$

Binomial and Geometric Random Variables

Section Summary

In this section, we learned how to…

- ✓ DETERMINE whether the conditions for using a binomial random variable are met.
- ✓ COMPUTE and INTERPRET probabilities involving binomial distributions.
- ✓ CALCULATE the mean and standard deviation of a binomial random variable. INTERPRET these values in context.
- ✓ FIND probabilities involving geometric random variables.
- ✓ Read p. 397-404 ccc 79, 81, 83, 85, 87, 89