+

Chapter 8: Binomial and Geometric Distributions

Section 8.1
Binomial Distributions

The Practice of Statistics, 4th edition – For AP*
STARNES, YATES, MOORE

Section 8.1
Binomial Distribution

Learning Objectives

After this section, you should be able to…

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

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. In some cases, the number of repeated trials is fixed in advance and we are interested in the number of times a particular event (called a "success") occurs. If the trials in these cases are *independent* and each success has an equal chance of occurring, we have a **binomial setting.**

Definition:

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 have any effect on 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? On each trial, the probability p of success must be the same.

Binomial Random Variable

Consider tossing a coin *n* times. Each toss gives either heads or tails. Knowing the outcome of one toss does not change the probability of an outcome on any other toss. If we define heads as a success, then *p* is the probability of a head and is 0.5 on any toss.

The number of heads in *n* tosses is a **binomial random variable *X*.** The probability distribution of *X* is called a **binomial distribution.**

Definition:

The count *X* of successes in a binomial setting is a **binomial random variable.** The probability distribution of *X* is a **binomial distribution** with parameters *n* and *p*, where *n* is the number of trials of the chance process and *p* is the probability of a success on any one trial. The possible values of *X* are the whole numbers from 0 to *n*.

***Note*: When checking the Binomial condition, be sure to check the BINS and make sure you're being asked to count the number of successes in a certain number of trials!**

Binomial Probabilities

In a binomial setting, we can define a random variable (say, X) as the number of successes in n independent trials. We are interested in finding the probability distribution of X.

Example

Each child of a particular pair of parents has probability 0.25 of having type O blood. Genetics says that children receive genes from each of their parents independently. If these parents have 5 children, the count X of children with type O blood is a binomial random variable with $n = 5$ trials and probability $p = 0.25$ of a success on each trial. In this setting, a child with type O blood is a "success" (S) and a child with another blood type is a "failure" (F).
What's $P(X = 2)$?

$$P(SSFFF) = (0.25)(0.25)(0.75)(0.75)(0.75) = (0.25)^2(0.75)^3 = 0.02637$$

However, there are a number of different arrangements in which 2 out of the 5 children have type O blood:

SSFFF	**SFSFF**	**SFFSF**	**SFFFS**	**FSSFF**
FSFSF	**FSFFS**	**FFSSF**	**FFSFS**	**FFFSS**

Verify that in each arrangement, $P(X = 2) = (0.25)^2(0.75)^3 = 0.02637$

Therefore, $\mathbf{P(X = 2) = 10(0.25)^2(0.75)^3 = 0.2637}$

Binomial Coefficient

Note, in the previous example, *any one arrangement* of 2 S's and 3 F's had the same probability. This is true because no matter what arrangement, we'd multiply together 0.25 twice and 0.75 three times.

We can generalize this for any setting in which we are interested in *k* successes in *n* trials. That is,

$$P(X = k) = P(\text{exactly } k \text{ successes in } n \text{ trials})$$
$$= \text{number of arrangements} \cdot p^k (1 - p)^{n-k}$$

Definition:

The number of ways of arranging *k* successes among *n* observations is given by the **binomial coefficient**

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

for $k = 0, 1, 2, \ldots, n$ where

$$n! = n(n-1)(n-2)\cdot\ldots\cdot(3)(2)(1)$$

and $0! = 1$.

Binomial Probability

The binomial coefficient counts the number of different ways in which k successes can be arranged among n trials. The binomial probability $P(X = k)$ is this count multiplied by the probability of any one specific arrangement of the k successes.

Binomial Probability

If X has the binomial distribution with n trials and probability p of success on each trial, the possible values of X are 0, 1, 2, …, n. If k is any one of these values,

$$P(X = k) = \binom{n}{k} p^k \underbrace{(1-p)^{n-k}}$$

Number of arrangements of k successes

Probability of k successes

Probability of n-k failures

Example: Inheriting Blood Type

Each child of a particular pair of parents has probability 0.25 of having blood type O. Suppose the parents have 5 children

(a) Find the probability that exactly 3 of the children have type O blood.

Let X = the number of children with type O blood. We know X has a binomial distribution with n = 5 and p = 0.25.

$$P(X=3)=\binom{5}{3}(0.25)^3(0.75)^2=10(0.25)^3(0.75)^2=0.0878$$

(b) Should the parents be surprised if more than 3 of their children have type O blood?

To answer this, we need to find $P(X > 3)$.

$$P(X>3)=P(X=4)+P(X=5)$$

$$=\binom{5}{4}(0.25)^4(0.75)^1+\binom{5}{5}(0.25)^5(0.75)^0$$

$$=5(0.25)^4(0.75)^1+1(0.25)^5(0.75)^0$$

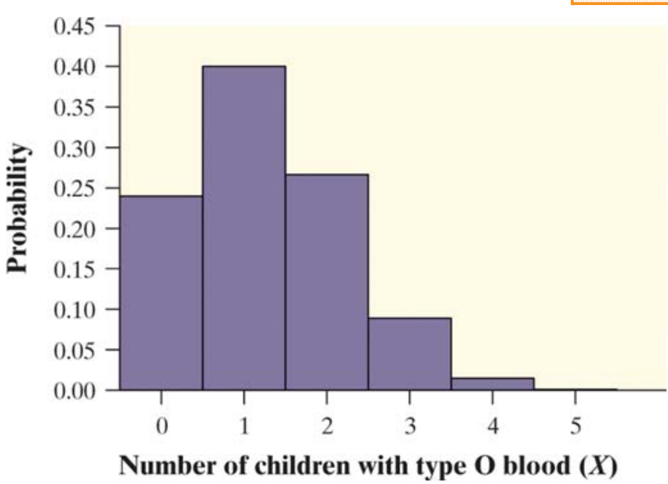
$$=0.01465+0.00098=0.01563$$

Since there is only a 1.5% chance that more than 3 children out of 5 would have Type O blood, the parents should be surprised!

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:

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

Spread: The variance of X is $\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i = (0-1.25)^2(0.2373) + (1-1.25)^2(0.3955) + \ldots + (5-1.25)^2(0.00098) = 0.9375$

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

Mean and Standard Deviation of a Binomial Distribution

Notice, the mean μ_X = 1.25 can be found another way. Since each child has a 0.25 chance of inheriting type O blood, we'd expect one-fourth of the 5 children to have this blood type. That is, $\mu_X = 5(0.25) = 1.25$. This method can be used to find the mean of any binomial random variable with parameters *n* and *p*.

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: Bottled Water versus Tap Water

Mr. Bullard's 21 AP Statistics students did the Activity on page 340. 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.

Binomial Distributions in Statistical Sampling

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

Suppose 10% of CDs have defective copy-protection schemes that can harm computers. A music distributor inspects an SRS of 10 CDs from a shipment of 10,000. Let *X* = number of defective CDs. **What is *P*(*X* = 0)?** *Note, this is not quite a binomial setting. Why?*

The actual probability is $P(\text{no defectives}) = \frac{9000}{10000} \cdot \frac{8999}{9999} \cdot \frac{8998}{9998} \cdot \ldots \cdot \frac{8991}{9991} = 0.3485$

Using the binomial distribution, $P(X = 0) = \binom{10}{0}(0.10)^0(0.90)^{10} = 0.3487$

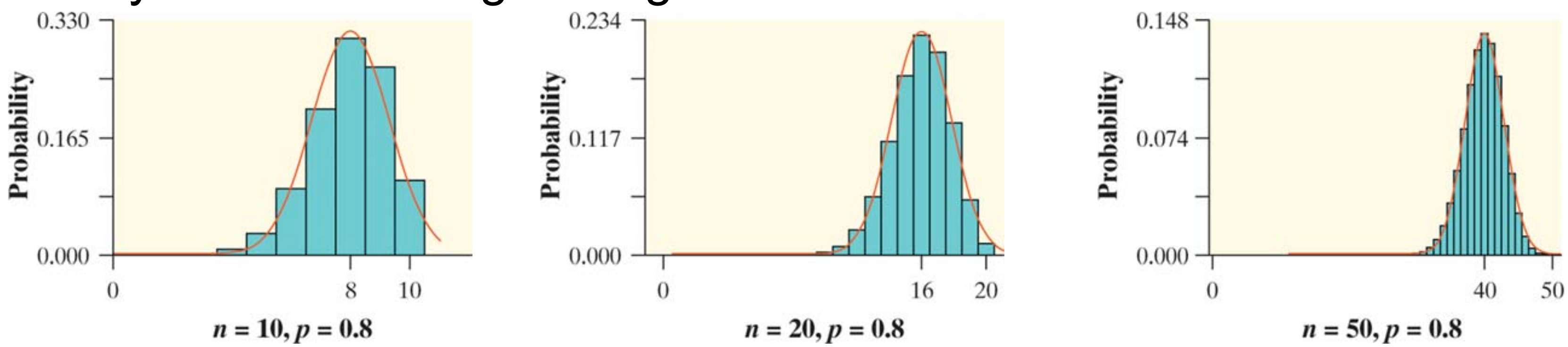
In practice, the binomial distribution gives a good approximation as long as we don't sample more than 10% of the population.

Sampling Without Replacement 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 $n \le \frac{1}{10}N$

Normal Approximation for Binomial Distributions

As *n* gets larger, something interesting happens to the shape of a binomial distribution. The figures below show histograms of binomial distributions for different values of *n* and *p*. What do you notice as *n* gets larger?



Normal Approximation for Binomial Distributions

Suppose that *X* has the binomial distribution with *n* trials and success probability *p*. When *n* is large, the distribution of *X* is approximately Normal with mean and standard deviation

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

As a rule of thumb, we will use the Normal approximation when *n* is so large that $np \geq 10$ and $n(1-p) \geq 10$. That is, the expected number of successes and failures are both at least 10.

Example: Attitudes Toward Shopping

Sample surveys show that fewer people enjoy shopping than in the past. A survey asked a nationwide random sample of 2500 adults if they agreed or disagreed that “I like buying new clothes, but shopping is often frustrating and time-consuming.” Suppose that exactly 60% of all adult US residents would say “Agree” if asked the same question. Let *X* = the number in the sample who agree. **Estimate the probability that 1520 or more of the sample agree.**

1) Verify that X is approximately a binomial random variable.

B: Success = agree, Failure = don’t agree
I: Because the population of U.S. adults is greater than 25,000, it is reasonable to assume the sampling without replacement condition is met.
N: n = 2500 trials of the chance process
S: The probability of selecting an adult who agrees is p = 0.60

2) Check the conditions for using a Normal approximation.

Since $np = 2500(0.60) = 1500$ and $n(1 - p) = 2500(0.40) = 1000$ are both at least 10, we may use the Normal approximation.

3) Calculate $P(X \geq 1520)$ using a Normal approximation.

$$\mu = np = 2500(0.60) = 1500$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{2500(0.60)(0.40)} = 24.49$$

$$z = \frac{1520 - 1500}{24.49} = 0.82$$

$$P(X \geq 1520) = P(Z \geq 0.82) = 1 - 0.7939 = 0.2061$$

Section 8.1
Binomial Distributions

Summary

In this section, we learned that…

- ✓ A **binomial setting** consists of *n* independent trials of the same chance process, each resulting in a success or a failure, with probability of success *p* on each trial. The count *X* of successes is a **binomial random variable**. Its probability distribution is a **binomial distribution**.

- ✓ The **binomial coefficient** counts the number of ways *k* successes can be arranged among *n* trials.

- ✓ If *X* has the binomial distribution with parameters *n* and *p*, the possible values of *X* are the whole numbers 0, 1, 2, . . . , *n*. The binomial probability of observing *k* successes in *n* trials is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Section 8.1
Binomial Distributions

Summary

In this section, we learned that…

- ✓ The **mean** and **standard deviation** of a binomial random variable *X* are

$$\mu_X = np$$

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

- ✓ The **Normal approximation** to the binomial distribution says that if *X* is a count having the binomial distribution with parameters *n* and *p*, then when *n* is large, *X* is approximately Normally distributed. We will use this approximation when $np \geq 10$ and $n(1 - p) \geq 10$.

Looking Ahead…

Homework…

Chapter 8: #'s 1, 4, 5, 7–10, 13, 19, 27, 28, 54, 56, 59