Lesson 8:
Chapter 5 Section 2

Caleb Moxley

BSC Mathematics

30 September 15

§5.2 Sampling Distribution of a Sample Proportion

Remember this example?

§5.2 Sampling Distribution of a Sample Proportion

Remember this example? We had this probability **model**

Black	Brown	Blue	Green	Other
0.3	0.35	0.15	0.1	0.1

§5.2 Sampling Distribution of a Sample Proportion

Remember this example? We had this probability **model**

Black	Brown	Blue	Green	Other
0.3	0.35	0.15	0.1	0.1

And we wanted to convert it to a probability distribution describing the random variable of the number of brown-eyed people in a sample of five people.

§5.2 Sampling Distribution of a Sample Proportion

Remember this example? We had this probability **model**

Black	Brown	Blue	Green	Other
0.3	0.35	0.15	0.1	0.1

And we wanted to convert it to a probability distribution describing the random variable of the number of brown-eyed people in a sample of five people. Our random variable turned out to be as below.

0	1	2	3	4	5
0.116	0.312	0.337	0.181	0.049	0.005

This was a **binomial random variable**, and in this section we'll discuss the binomial random variable and how it relates to **population proportions**.

§5.2 Sampling Distribution of a Sample Proportion

Definition (proportion)

A proportion (of a sample or a population) is simply the proportions of “successes” in a sample or a population.

§5.2 Sampling Distribution of a Sample Proportion

Definition (proportion)

A proportion (of a sample or a population) is simply the proportions of "successes" in a sample or a population. A population proportion is denoted as p whereas a sample proportion is denoted as $\hat{p}$.

§5.2 Sampling Distribution of a Sample Proportion

Definition (proportion)

A proportion (of a sample or a population) is simply the proportions of "successes" in a sample or a population. A population proportion is denoted as p whereas a sample proportion is denoted as $\hat{p}$.

Example (proportions)

We'd like to estimate the proportion of students with brown eyes, so we sample 50 students and find that 17 have brown eyes. What's the sample proportion?

§5.2 Sampling Distribution of a Sample Proportion

Definition (proportion)

A proportion (of a sample or a population) is simply the proportions of "successes" in a sample or a population. A population proportion is denoted as p whereas a sample proportion is denoted as $\hat{p}$.

Example (proportions)

We'd like to estimate the proportion of students with brown eyes, so we sample 50 students and find that 17 have brown eyes. What's the sample proportion?

$$\hat{p} = \frac{17}{50} = 0.34.$$

§5.2 Sampling Distribution of a Sample Proportion

Definition (proportion)

A proportion (of a sample or a population) is simply the proportions of "successes" in a sample or a population. A population proportion is denoted as p whereas a sample proportion is denoted as $\hat{p}$.

Example (proportions)

We'd like to estimate the proportion of students with brown eyes, so we sample 50 students and find that 17 have brown eyes. What's the sample proportion?

$$\hat{p} = \frac{17}{50} = 0.34.$$

Notice, this is close to the population proportion $p = 0.35$, but some variation might be expected due to sampling.

§5.2 Sampling Distribution of a Sample Proportion

Many probability models can be transformed into probability distributions via proportions/counting the number of successes.

§5.2 Sampling Distribution of a Sample Proportion

Many probability models can be transformed into probability distributions via proportions/counting the number of successes.

Sample proportions and sample counts (which are simply the number of successes observed in a sample) are frequently used statistics, so we should be interested in understanding their sampling distributions. To that end, let's talk about the **binomial random variable/binomial distribution**.

§5.2 Sampling Distribution of a Sample Proportion

For a binomial distribution to occur, we must have the following setting.

1. There must be a fixed number of trails.

§5.2 Sampling Distribution of a Sample Proportion

For a binomial distribution to occur, we must have the following setting.

1. There must be a fixed number of trails.
2. Each trail must be independent.

§5.2 Sampling Distribution of a Sample Proportion

For a binomial distribution to occur, we must have the following setting.

1. There must be a fixed number of trails.
2. Each trail must be independent.
3. Each observation should be able to be labeled as either a "success" or "failure" but not both.

§5.2 Sampling Distribution of a Sample Proportion

For a binomial distribution to occur, we must have the following setting.

1. There must be a fixed number of trails.
2. Each trail must be independent.
3. Each observation should be able to be labeled as either a "success" or "failure" but not both.
4. The probability of success p in a single trail must be the same for each observation.

§5.2 Sampling Distribution of a Sample Proportion

For a binomial distribution to occur, we must have the following setting.

1. There must be a fixed number of trails.
2. Each trail must be independent.
3. Each observation should be able to be labeled as either a "success" or "failure" but not both.
4. The probability of success p in a single trail must be the same for each observation.

We call this the **binomial setting**.

§5.2 Sampling Distribution of a Sample Proportion

For a binomial distribution to occur, we must have the following setting.

1. There must be a fixed number of trails.
2. Each trail must be independent.
3. Each observation should be able to be labeled as either a "success" or "failure" but not both.
4. The probability of success p in a single trail must be the same for each observation.

We call this the **binomial setting**.

Example

Does our previous example of counting the number of brown-eyed people in a sample of 5 people qualify as a binomial setting?

§5.2 Sampling Distribution of a Sample Proportion

Definition (binomial distribution)

The distribution of the count X of successes in a binomial setting is called the **binomial distribution** with parameters n and p. The parameter n is the number of observations/trails while p is the probability of success in any single observation/trail. The possible values of X are the whole numbers 0 to n. We often denote this distribution of X as $B(n, p)$.

§5.2 Sampling Distribution of a Sample Proportion

Definition (binomial distribution)

The distribution of the count X of successes in a binomial setting is called the **binomial distribution** with parameters n and p. The parameter n is the number of observations/trails while p is the probability of success in any single observation/trail. The possible values of X are the whole numbers 0 to n. We often denote this distribution of X as $B(n, p)$.

There is a formula that describes how to assign probabilities to each of these values of X. We'll see this later.

§5.2 Sampling Distribution of a Sample Proportion

Theorem (sampling distribution of a count)

A population contains proportion p of successes. If the population is much larger than the sample, the count X of successes in a SRS of size n has approximately the binomial distribution $B(n, p)$.

§5.2 Sampling Distribution of a Sample Proportion

Theorem (sampling distribution of a count)

A population contains proportion p of successes. If the population is much larger than the sample, the count X of successes in a SRS of size n has approximately the binomial distribution $B(n, p)$.

The accuracy of this approximation improves as the size of the population increases relative to the size of the sample. **As a rule, we will use the binomial sampling distribution for counts when the population is at least 20 times as large as the sample.** *You can relax this rule if the sampling is done with replacement.*

§5.2 Sampling Distribution of a Sample Proportion

Theorem (sampling distribution of a count)

A population contains proportion p of successes. If the population is much larger than the sample, the count X of successes in a SRS of size n has approximately the binomial distribution $B(n, p)$.

The accuracy of this approximation improves as the size of the population increases relative to the size of the sample. **As a rule, we will use the binomial sampling distribution for counts when the population is at least 20 times as large as the sample.** *You can relax this rule if the sampling is done with replacement.*

So, now that we know what a binomial distribution is and why it's useful, we need to figure out how to calculate binomial probabilities.

§5.2 Sampling Distribution of a Sample Proportion

If you want to calculate $P(X = k)$ in some $B(n, p)$ distribution, you must calculate

$$\binom{n}{k} p^k q^{n-k}.$$

§5.2 Sampling Distribution of a Sample Proportion

If you want to calculate $P(X = k)$ in some $B(n, p)$ distribution, you must calculate

$$\binom{n}{k} p^k q^{n-k}.$$

Here, $\binom{n}{k}$ (read n choose k) simply denotes $\dfrac{n!}{(n-k)!k!}$.

§5.2 Sampling Distribution of a Sample Proportion

If you want to calculate $P(X = k)$ in some $B(n, p)$ distribution, you must calculate

$$\binom{n}{k} p^k q^{n-k}.$$

Here, $\binom{n}{k}$ (read n choose k) simply denotes $\dfrac{n!}{(n-k)!k!}$. If you want to calculate $P(X \leq k)$, you must calculate

$$\sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i}.$$

§5.2 Sampling Distribution of a Sample Proportion

Example (binomial probability)

Calculate $P(X < 2)$ for $B(10, 0.25)$.

§5.2 Sampling Distribution of a Sample Proportion

Example (binomial probability)

Calculate $P(X < 2)$ for $B(10, 0.25)$.

$$\sum_{i=0}^{1} \binom{10}{i} (0.25)^i (0.75)^{10-i} =$$

§5.2 Sampling Distribution of a Sample Proportion

Example (binomial probability)

Calculate $P(X < 2)$ for $B(10, 0.25)$.

$$\sum_{i=0}^{1} \binom{10}{i} (0.25)^i (0.75)^{10-i} =$$

$$\binom{10}{0} (0.25)^0 (0.75)^{10} + \binom{10}{1} (0.25)^1 (0.75)^9 =$$

§5.2 Sampling Distribution of a Sample Proportion

Example (binomial probability)

Calculate $P(X < 2)$ for $B(10, 0.25)$.

$$\sum_{i=0}^{1} \binom{10}{i} (0.25)^i (0.75)^{10-i} =$$

$$\binom{10}{0} (0.25)^0 (0.75)^{10} + \binom{10}{1} (0.25)^1 (0.75)^9 =$$

$$(1)(1)(0.75)^{10} + (10)(0.25)(0.75)^9 =$$

Example (binomial probability)

Calculate $P(X < 2)$ for $B(10, 0.25)$.

$$\sum_{i=0}^{1} \binom{10}{i} (0.25)^i (0.75)^{10-i} =$$

$$\binom{10}{0} (0.25)^0 (0.75)^{10} + \binom{10}{1} (0.25)^1 (0.75)^9 =$$

$$(1)(1)(0.75)^{10} + (10)(0.25)(0.75)^9 = 0.244.$$

§5.2 Sampling Distribution of a Sample Proportion

Example (binomial probability)

Calculate $P(X < 2)$ for $B(10, 0.25)$.

$$\sum_{i=0}^{1} \binom{10}{i} (0.25)^i (0.75)^{10-i} =$$

$$\binom{10}{0} (0.25)^0 (0.75)^{10} + \binom{10}{1} (0.25)^1 (0.75)^9 =$$

$$(1)(1)(0.75)^{10} + (10)(0.25)(0.75)^9 = 0.244.$$

TI-83/84, Minitab, and tables can also be used to compute these probabilities! Feel free to use these in addition to the formula.

§5.2 Sampling Distribution of a Sample Proportion

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws?

§5.2 Sampling Distribution of a Sample Proportion

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws? We need to calculate $1 - P(X \leq 6)$ where X is the number of free throws **missed**!

§5.2 Sampling Distribution of a Sample Proportion

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws? We need to calculate $1 - P(X \leq 6)$ where X is the number of free throws **missed**!

$$1 - \sum_{i=0}^{6} \binom{12}{i} (0.25)^i (0.75)^{12-i} =$$

§5.2 Sampling Distribution of a Sample Proportion

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws? We need to calculate $1 - P(X \leq 6)$ where X is the number of free throws **missed**!

$$1 - \sum_{i=0}^{6} \binom{12}{i} (0.25)^i (0.75)^{12-i} = 0.0143.$$

§5.2 Sampling Distribution of a Sample Proportion

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws? We need to calculate $1 - P(X \leq 6)$ where X is the number of free throws **missed**!

$$1 - \sum_{i=0}^{6} \binom{12}{i} (0.25)^i (0.75)^{12-i} = 0.0143.$$

So, it would be very unusual for her to have missed 7 or more free throws, which means that 7 is an unusually high number of missed three throws (out of 12).

§5.2 Sampling Distribution of a Sample Proportion

Remember the laborious work of calculating means and standard deviations of probability distributions before?

§5.2 Sampling Distribution of a Sample Proportion

Remember the laborious work of calculating means and standard deviations of probability distributions before? Well, for the binomial distribution, we can save a bunch of time!

§5.2 Sampling Distribution of a Sample Proportion

Remember the laborious work of calculating means and standard deviations of probability distributions before? Well, for the binomial distribution, we can save a bunch of time!

Theorem (mean, variance, and standard deviation of $B(n, p)$)

For a binomial $X \sim B(n, p)$ probability distribution, we have

$$\mu_X = np \quad \text{and} \quad \sigma_X^2 = npq,$$

where $q = 1 - p$ is the probability of failure. To find the standard deviation, just take the square root of the variance.

§5.2 Sampling Distribution of a Sample Proportion

Remember the laborious work of calculating means and standard deviations of probability distributions before? Well, for the binomial distribution, we can save a bunch of time!

Theorem (mean, variance, and standard deviation of $B(n, p)$)

For a binomial $X \sim B(n, p)$ probability distribution, we have

$$\mu_X = np \quad \textit{and} \quad \sigma_X^2 = npq,$$

where $q = 1 - p$ is the probability of failure. To find the standard deviation, just take the square root of the variance.

You can use this to figure out if a value is unusually high/low as well! Let's revisit our previous example.

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws?

§5.2 Sampling Distribution of a Sample Proportion

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws?

Well, $\mu_X = 12(0.25) = 3$ and $\sigma_X = \sqrt{12(0.25)(0.75)} = 1.5$. Thus, our range of usual values is $[3 - 2(1.5), 3 + 2(1.5)] = [0, 6]$. Notice, 7 is usually high in this setting as well!

Example (missing free throws)

A basketball player makes 75% of her free throws. In a recent game, she had 12 free throw attempts and missed 7 of them. Is this an unusually high number of missed free throws?

Well, $\mu_X = 12(0.25) = 3$ and $\sigma_X = \sqrt{12(0.25)(0.75)} = 1.5$. Thus, our range of usual values is $[3 - 2(1.5), 3 + 2(1.5)] = [0, 6]$. Notice, 7 is usually high in this setting as well! This works because the binomial distribution is roughly symmetric, bell-shaped, and nearly normal!

§5.2 Sampling Distribution of a Sample Proportion

Now that we've talked about sampling distributions for population counts, lets discuss sampling distributions of population proportions! These distributions are not binomial like population counts, though you can often restate questions involving proportions as questions involving counts. Often, however, we simply approximate these sampling distributions with a normal distribution!

§5.2 Sampling Distribution of a Sample Proportion

Theorem (normal approximation for counts and proportions)

A SRS of size n from a large population, having proportion p successes, has sampling distributions for counts X and proportions $\hat{p} = \frac{X}{n}$ given by

$$N\left(np, \sqrt{np(1-p)}\right) \quad \text{and} \quad N\left(p, \sqrt{\frac{p(1-p)}{n}}\right),$$

respectively. **As a rule, we will only use this approximation for values of n and p that satisfy $np \geq 10$ and $n(1-p) \geq 10$.**

§5.2 Sampling Distribution of a Sample Proportion

Theorem (normal approximation for counts and proportions)

A SRS of size n from a large population, having proportion p successes, has sampling distributions for counts X and proportions $\hat{p} = \frac{X}{n}$ given by

$$N\left(np, \sqrt{np(1-p)}\right) \quad \text{and} \quad N\left(p, \sqrt{\frac{p(1-p)}{n}}\right),$$

respectively. **As a rule, we will only use this approximation for values of n and p that satisfy $np \geq 10$ and $n(1-p) \geq 10$.**

We use the rule above because the binomial distribution is not approximately normal unless it has at least 10 successes and failures.

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

§5.2 Sampling Distribution of a Sample Proportion

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(\hat{p} < 0.041)$, where $\hat{p}$ has approximately a normal distribution $N(0.04, 0.004382)$.

§5.2 Sampling Distribution of a Sample Proportion

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(\hat{p} < 0.041)$, where $\hat{p}$ has approximately a normal distribution $N(0.04, 0.004382)$. We convert this to a standard normal or use Minitab to see that

$$P(\hat{p} < 0.041) \approx$$

§5.2 Sampling Distribution of a Sample Proportion

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(\hat{p} < 0.041)$, where $\hat{p}$ has approximately a normal distribution $N(0.04, 0.004382)$. We convert this to a standard normal or use Minitab to see that

$$P(\hat{p} < 0.041) \approx P(Z < 0.23) = 0.5910.$$

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, where X has approximately a normal distribution $N(80, 8.76356)$.

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, where X has approximately a normal distribution $N(80, 8.76356)$. We convert this to a standard normal or use Minitab to see that

$$P(X < 82) \approx$$

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, where X has approximately a normal distribution $N(80, 8.76356)$. We convert this to a standard normal or use Minitab to see that

$$P(X < 82) \approx P(Z < 0.23) = 0.5910.$$

§5.2 Sampling Distribution of a Sample Proportion

In the previous example, we used a continuous normal distribution to estimate the probability described by a **discrete** binomial distribution! There is always an error associated with approximating something discrete by something continuous, and we often correct for this issue.

§5.2 Sampling Distribution of a Sample Proportion

In the previous example, we used a continuous normal distribution to estimate the probability described by a **discrete** binomial distribution! There is always an error associated with approximating something discrete by something continuous, and we often correct for this issue.

Definition (continuity correction)

When calculating the following probabilities of a binomial random variable X, you can improve the approximation by a normal random variable in the following way

§5.2 Sampling Distribution of a Sample Proportion

In the previous example, we used a continuous normal distribution to estimate the probability described by a **discrete** binomial distribution! There is always an error associated with approximating something discrete by something continuous, and we often correct for this issue.

Definition (continuity correction)

When calculating the following probabilities of a binomial random variable X, you can improve the approximation by a normal random variable in the following way

- $P(X \leq K) \approx P(X \leq K + 0.5)$
- $P(X < K) \approx P(X \leq K - 0.5)$

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, which is better approximated by $P(X < 81.5)$, where X has approximately a normal distribution $N(80, 8.76356)$.

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, which is better approximated by $P(X < 81.5)$, where X has approximately a normal distribution $N(80, 8.76356)$. We convert this to a standard normal or use Minitab to see that

$$P(X < 82) = P(X < 81.5) \approx$$

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, which is better approximated by $P(X < 81.5)$, where X has approximately a normal distribution $N(80, 8.76356)$. We convert this to a standard normal or use Minitab to see that

$$P(X < 82) = P(X < 81.5) \approx P(Z < 0.17) = 0.5675.$$

The real probability $P(X < 82)$ is 0.5746.

§5.2 Sampling Distribution of a Sample Proportion

We could investigate the same question from a counts perspective!

Example (bad apples)

An orchard's harvest typically has a proportion of bad apples of about 4%. An applesauce company will only accept a shipment of apples if the shipment contains less than 4.1% bad apples. It orders a shipment of 2000 apples. What's the chance that this shipment will be accepted?

We must calculate $P(X < 82)$, which is better approximated by $P(X < 81.5)$, where X has approximately a normal distribution $N(80, 8.76356)$. We convert this to a standard normal or use Minitab to see that

$$P(X < 82) = P(X < 81.5) \approx P(Z < 0.17) = 0.5675.$$

The real probability $P(X < 82)$ is 0.5746.

§5.2 Sampling Distribution of a Sample Proportion

Chapter 5 discusses the Poisson distribution - which is essentially the distribution which arises from the binomial setting when we assume that the number of trails in unlimited. We won't discuss this distribution in this course. Don't expect to see anything about it on the test.