Lesson 13:
Chapter 8 Section 1

Caleb Moxley

BSC Mathematics

19 October 15

§8.1 Inference for a Single Proportion

Often, we collect categorical data, and it might seem difficult to make inferential claims about non-numeric data. We can, however, always convert categorical data to proportions and make and test claims about these proportions!

§8.1 Inference for a Single Proportion

Often, we collect categorical data, and it might seem difficult to make inferential claims about non-numeric data. We can, however, always convert categorical data to proportions and make and test claims about these proportions! Inference about population proportions is very similar to inference about population means, so we're really just extending the knowledge we have about those types of inference tests.

§8.1 Inference for a Single Proportion

Often, we collect categorical data, and it might seem difficult to make inferential claims about non-numeric data. We can, however, always convert categorical data to proportions and make and test claims about these proportions! Inference about population proportions is very similar to inference about population means, so we're really just extending the knowledge we have about those types of inference tests. We use the normal approximation to the sampling distribution of the proportion that we learned in Chapter 5!

§8.1 Inference for a Single Proportion

Often, we collect categorical data, and it might seem difficult to make inferential claims about non-numeric data. We can, however, always convert categorical data to proportions and make and test claims about these proportions! Inference about population proportions is very similar to inference about population means, so we're really just extending the knowledge we have about those types of inference tests. We use the normal approximation to the sampling distribution of the proportion that we learned in Chapter 5! Recall:

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N\left(\mu_{\hat{p}}, \sigma_{\hat{p}}\right).$$

§8.1 Inference for a Single Proportion

With this knowledge of the sampling distribution of $\hat{p}$, we can compute confidence intervals, P-values, and critical values using the standard normal tables/Minitab.

§8.1 Inference for a Single Proportion

With this knowledge of the sampling distribution of $\hat{p}$, we can compute confidence intervals, P-values, and critical values using the standard normal tables/Minitab.

Definition (large n confidence interval for a population proportion)

Choose a SRS of size n from a large population with unknown proportion of success p. The sample proportion is $\hat{p} = \frac{X}{n}$, where X is the number of successes. The standard error of $\hat{p}$ is $\mathrm{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n}$, and the margin of error for confidence C is $m = z^* SE_{\hat{p}}$, where z^* is the value of the $N(0,1)$ RV with area C between $-z^*$ and z^*. The approximate level C confidence interval for p is

$$\hat{p} \pm m.$$

§8.1 Inference for a Single Proportion

With this knowledge of the sampling distribution of $\hat{p}$, we can compute confidence intervals, P-values, and critical values using the standard normal tables/Minitab.

Definition (large n confidence interval for a population proportion)

Choose a SRS of size n from a large population with unknown proportion of success p. The sample proportion is $\hat{p} = \frac{X}{n}$, where X is the number of successes. The standard error of $\hat{p}$ is $\mathrm{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n}$, and the margin of error for confidence C is $m = z^* SE_{\hat{p}}$, where z^* is the value of the $N(0,1)$ RV with area C between $-z^*$ and z^*. The approximate level C confidence interval for p is

$$\hat{p} \pm m.$$

In order to get this confidence interval, the number of successes and failures should be at least ten each.

Steps for constructing CI:

§8.1 Inference for a Single Proportion

Steps for constructing CI:

1. Compute $\hat{p} = \frac{X}{n}$.

§8.1 Inference for a Single Proportion

Steps for constructing CI:

1. Compute $\hat{p} = \frac{X}{n}$.
2. Compute $\text{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

§8.1 Inference for a Single Proportion

Steps for constructing CI:

1. Compute $\hat{p} = \frac{X}{n}$.
2. Compute $\text{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
3. Compute z^* corresponding to your confidence level C.

§8.1 Inference for a Single Proportion

Steps for constructing CI:

1. Compute $\hat{p} = \frac{X}{n}$.
2. Compute $\mathrm{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
3. Compute z^* corresponding to your confidence level C.
4. Create the confidence interval $\hat{p} \pm z^* \mathrm{SE}_{\hat{p}}$.

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Would this sample support the claim that at least 10% of Fortune 500 employees hold advanced degrees?

§8.1 Inference for a Single Proportion

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Would this sample support the claim that at least 10% of Fortune 500 employees hold advanced degrees? Yes.

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Would this sample support the claim that at least 10% of Fortune 500 employees hold advanced degrees? Yes. Would it support the claim that 12% of Fortune 500 employees hold advanced degrees?

§8.1 Inference for a Single Proportion

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Would this sample support the claim that at least 10% of Fortune 500 employees hold advanced degrees? Yes. Would it support the claim that 12% of Fortune 500 employees hold advanced degrees? Yes.

§8.1 Inference for a Single Proportion

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Would this sample support the claim that at least 10% of Fortune 500 employees hold advanced degrees? Yes. Would it support the claim that 12% of Fortune 500 employees hold advanced degrees? Yes. More than 12%?

Example

In a SRS of 400 employees at Fortune 500 companies, 60 had advanced degrees. Create a 95% confidence interval for the proportion of employees at Fortune 500 companies who hold advanced degrees.

Answer: (0.115,0.185)

Would this sample support the claim that at least 10% of Fortune 500 employees hold advanced degrees? Yes. Would it support the claim that 12% of Fortune 500 employees hold advanced degrees? Yes. More than 12%? No.

§8.1 Inference for a Single Proportion

When the sample does not contain a least ten failures and successes, we may use the following rule.

§8.1 Inference for a Single Proportion

When the sample does not contain a least ten failures and successes, we may use the following rule.

Definition (plus four confidence interval for a single proportion)

When there are not at least four successes and failures in a sample for a population proportion, we may construct the confidence interval using $\tilde{p}$ instead of $\hat{p}$ where

$$\tilde{p} = \frac{X+2}{n+4}.$$

§8.1 Inference for a Single Proportion

When the sample does not contain a least ten failures and successes, we may use the following rule.

Definition (plus four confidence interval for a single proportion)

When there are not at least four successes and failures in a sample for a population proportion, we may construct the confidence interval using $\tilde{p}$ instead of $\hat{p}$ where

$$\tilde{p} = \frac{X + 2}{n + 4}.$$

Basically, we assume that the sample contains four more observations, half of which are successes.

When the sample does not contain a least ten failures and successes, we may use the following rule.

Definition (plus four confidence interval for a single proportion)

When there are not at least four successes and failures in a sample for a population proportion, we may construct the confidence interval using $\tilde{p}$ instead of $\hat{p}$ where

$$\tilde{p} = \frac{X+2}{n+4}.$$

Basically, we assume that the sample contains four more observations, half of which are successes.

We then repeat the process for creating a confidence interval using $\tilde{p}$ rather than $\hat{p}$.

§8.1 Inference for a Single Proportion

When the sample does not contain a least ten failures and successes, we may use the following rule.

Definition (plus four confidence interval for a single proportion)

When there are not at least four successes and failures in a sample for a population proportion, we may construct the confidence interval using $\tilde{p}$ instead of $\hat{p}$ where

$$\tilde{p} = \frac{X+2}{n+4}.$$

Basically, we assume that the sample contains four more observations, half of which are successes.

We then repeat the process for creating a confidence interval using $\tilde{p}$ rather than $\hat{p}$. See page 494 for an example of this.

§8.1 Inference for a Single Proportion

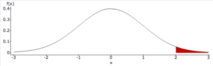
Definition (large sample z test for population proportion)

To test the hypothesis $H_0 : p = p_0$ based on a SRS of size n from a population with unknown p, compute the **test statistic**

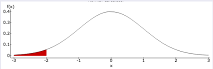
$$z = \frac{\hat{p} - p_0}{\sqrt{(p_0(1 - p_0))/n}}.$$

For a $N(0, 1)$ RV, the P-value for a test of H_0 against

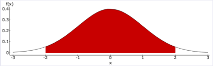
$H_a : p > p_0$ is $P(Z > z)$



$H_a : p < p_0$ is $P(Z < z)$



$H_a : p \neq p_0$ is $2P(Z > |z|)$



§8.1 Inference for a Single Proportion

Notice: The P-value test given in the previous slide follows essentially the same process as our previous P-value tests! The only difference is the way that we calculated the test statistic and P value.

Notice: The P-value test given in the previous slide follows essentially the same process as our previous P-value tests! The only difference is the way that we calculated the test statistic and P value. We also have the following set of requirements to use the above test.

§8.1 Inference for a Single Proportion

Notice: The P-value test given in the previous slide follows essentially the same process as our previous P-value tests! The only difference is the way that we calculated the test statistic and P value. We also have the following set of requirements to use the above test.

- The population from which the sample is drawn should be large or the sample should be done with replacement.

§8.1 Inference for a Single Proportion

Notice: The P-value test given in the previous slide follows essentially the same process as our previous P-value tests! The only difference is the way that we calculated the test statistic and P value. We also have the following set of requirements to use the above test.

- The population from which the sample is drawn should be large or the sample should be done with replacement.
- The sample itself should be large, but not more than 5% of the population.

§8.1 Inference for a Single Proportion

Notice: The P-value test given in the previous slide follows essentially the same process as our previous P-value tests! The only difference is the way that we calculated the test statistic and P value. We also have the following set of requirements to use the above test.

- The population from which the sample is drawn should be large or the sample should be done with replacement.
- The sample itself should be large, but not more than 5% of the population.
- There should be at least ten failures and ten successes in the sample.

§8.1 Inference for a Single Proportion

If the last requirement on the previous slide isn't met, then we could again use $\tilde{p}$ rather than $\hat{p}$ to compute the z test statistic and perform the same test.

If the last requirement on the previous slide isn't met, then we could again use $\tilde{p}$ rather than $\hat{p}$ to compute the z test statistic and perform the same test. However, in practice, we really only use $\tilde{p}$ for confidence intervals and not for P-value tests.

§8.1 Inference for a Single Proportion

If the last requirement on the previous slide isn't met, then we could again use $\tilde{p}$ rather than $\hat{p}$ to compute the z test statistic and perform the same test. However, in practice, we really only use $\tilde{p}$ for confidence intervals and not for P-value tests. Let's see an example!

Example (pizzas delivered on time)

A pizza delivery chain wants to test the claim that at least 95% of pizzas ordered for delivery at its restaurants are delivered on time. It wants to test this claim with 99% confidence. After taking a SRS of 300 deliveries, it determined that 284 were delivered on time. Does this sample support the claim?

Example (pizzas delivered on time)

A pizza delivery chain wants to test the claim that at least 95% of pizzas ordered for delivery at its restaurants are delivered on time. It wants to test this claim with 99% confidence. After taking a SRS of 300 deliveries, it determined that 284 were delivered on time. Does this sample support the claim?

Answer: Well, $z = -0.265$.

Example (pizzas delivered on time)

A pizza delivery chain wants to test the claim that at least 95% of pizzas ordered for delivery at its restaurants are delivered on time. It wants to test this claim with 99% confidence. After taking a SRS of 300 deliveries, it determined that 284 were delivered on time. Does this sample support the claim?

Answer: Well, $z = -0.265$. And $P(Z < -0.265) = 0.3955$. So we fail to reject H_0 and support our claim.

Example (voters close to polling stations)

A SRS of 2000 voters in Alabama revealed that 1729 lived within 15 minutes of their polling stations. Does this sample support the claim that the proportion of voters in Alabama who live within 15 minutes of their polling stations is 84%? Use $\alpha = 0.02$ and a P-value test.

Example (voters close to polling stations)

A SRS of 2000 voters in Alabama revealed that 1729 lived within 15 minutes of their polling stations. Does this sample support the claim that the proportion of voters in Alabama who live within 15 minutes of their polling stations is 84%? Use $\alpha = 0.02$ and a P-value test.

Answer: Well, $z = 2.989$.

§8.1 Inference for a Single Proportion

Example (voters close to polling stations)

A SRS of 2000 voters in Alabama revealed that 1729 lived within 15 minutes of their polling stations. Does this sample support the claim that the proportion of voters in Alabama who live within 15 minutes of their polling stations is 84%? Use $\alpha = 0.02$ and a P-value test.

Answer: Well, $z = 2.989$. And $2(P(Z > |2.989|) = 2(1-0.9986) = 0.0028$.

Example (voters close to polling stations)

A SRS of 2000 voters in Alabama revealed that 1729 lived within 15 minutes of their polling stations. Does this sample support the claim that the proportion of voters in Alabama who live within 15 minutes of their polling stations is 84%? Use $\alpha = 0.02$ and a P-value test.

Answer: Well, $z = 2.989$. And $2(P(Z > |2.989|) = 2(1 - 0.9986) = 0.0028$. Thus we reject H_0 and fail to support our claim.

Example (loaded coin)

A coin is tossed 200 times, and it lands on heads 115 times. Test the claim that the coin is loaded so that it comes up heads more than 50% of the time. Use a P-value test and a 10% significance.

Example (loaded coin)

A coin is tossed 200 times, and it lands on heads 115 times. Test the claim that the coin is loaded so that it comes up heads more than 50% of the time. Use a P-value test and a 10% significance.

Answer: Well, $z = 2.121$.

§8.1 Inference for a Single Proportion

Example (loaded coin)

A coin is tossed 200 times, and it lands on heads 115 times. Test the claim that the coin is loaded so that it comes up heads more than 50% of the time. Use a P-value test and a 10% significance.

Answer: Well, $z = 2.121$. And $P(Z > 2.989) = 0.0169$.

Example (loaded coin)

A coin is tossed 200 times, and it lands on heads 115 times. Test the claim that the coin is loaded so that it comes up heads more than 50% of the time. Use a P-value test and a 10% significance.

Answer: Well, $z = 2.121$. And $P(Z > 2.989) = 0.0169$. Thus we reject H_0 and support our claim.

§8.1 Inference for a Single Proportion

You might consider conducting critical value tests for the three previous examples.

You might consider conducting critical value tests for the three previous examples. In order, the three critical values are

- $z^* = -2.33$ for the pizza problem,

§8.1 Inference for a Single Proportion

You might consider conducting critical value tests for the three previous examples. In order, the three critical values are

- $z^* = -2.33$ for the pizza problem,
- $z^* = 2.33$ for the voters problem, and

You might consider conducting critical value tests for the three previous examples. In order, the three critical values are

- $z^* = -2.33$ for the pizza problem,
- $z^* = 2.33$ for the voters problem, and
- $z^* = 1.28$ for the loaded coin problem.

You might consider conducting critical value tests for the three previous examples. In order, the three critical values are

- $z^* = -2.33$ for the pizza problem,
- $z^* = 2.33$ for the voters problem, and
- $z^* = 1.28$ for the loaded coin problem.

Make sure you reach the same conclusion using these critical value tests — and remember to compare z and z^* in the appropriate way.

Finally, we'll determine how large we need to make our samples in order to get a confidence interval for a proportion with a particular level of confidence and a particular margin of error!

Definition (sample size n for a desired margin of error)

The level C confidence interval for a proportion p will have a margin of error approximately equal to a specified value m when the sample size satisfies

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*),$$

where p^* is some guessed proportion of successes for the future sample.

Definition (sample size n for a desired margin of error)

The level C confidence interval for a proportion p will have a margin of error approximately equal to a specified value m when the sample size satisfies

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*),$$

where p^* is some guessed proportion of successes for the future sample. If you have no good guess, let p^* be 50%. This maximizes the above function over all p^*, so you can't go wrong with that guess. The formula then becomes

$$n = \left(\frac{z^*}{m}\right)^2 \frac{1}{4}.$$

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*) =$$

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.75)(0.25) = 288.12.$$

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.75)(0.25) = 288.12.$$

Thus, we should flip the coin 289 times.

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.75)(0.25) = 288.12.$$

Thus, we should flip the coin 289 times. But if we didn't expect it to land on heads 75% of the time, we would have flipped it more!

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*) =$$

§8.1 Inference for a Single Proportion

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.75)(0.25) = 288.12.$$

Thus, we should flip the coin 289 times. But if we didn't expect it to land on heads 75% of the time, we would have flipped it more!

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.5)(0.5) = 384.16.$$

Example (loaded coin)

We think a coin has been loaded so that it will turn up heads 75% of the time. How large should our sample be in order to create a 95% CI for the proportion of times a coin lands on heads if we want the margin of error to be less than 0.05?

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.75)(0.25) = 288.12.$$

Thus, we should flip the coin 289 times. But if we didn't expect it to land on heads 75% of the time, we would have flipped it more!

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.05}\right)^2 (0.5)(0.5) = 384.16.$$

We would have needed to flip it 385 times.