CHAPTER 22:
Inference about a Population Proportion

Basic Practice of Statistics

7th Edition

Lecture PowerPoint Slides

In Chapter 22, We Cover …

- The sample proportion $\hat{p}$
- Large-sample confidence intervals for a population proportion
- Choosing the sample size
- Significance tests for a proportion
- Plus four confidence intervals for a proportion*

The Sampling Proportion, $\hat{p}$

- The statistic that estimates the population proportion, p, is the **sample proportion:**

$$\hat{p} = \frac{\text{number of successes in the sample}}{\text{total number of individuals in the sample}}$$

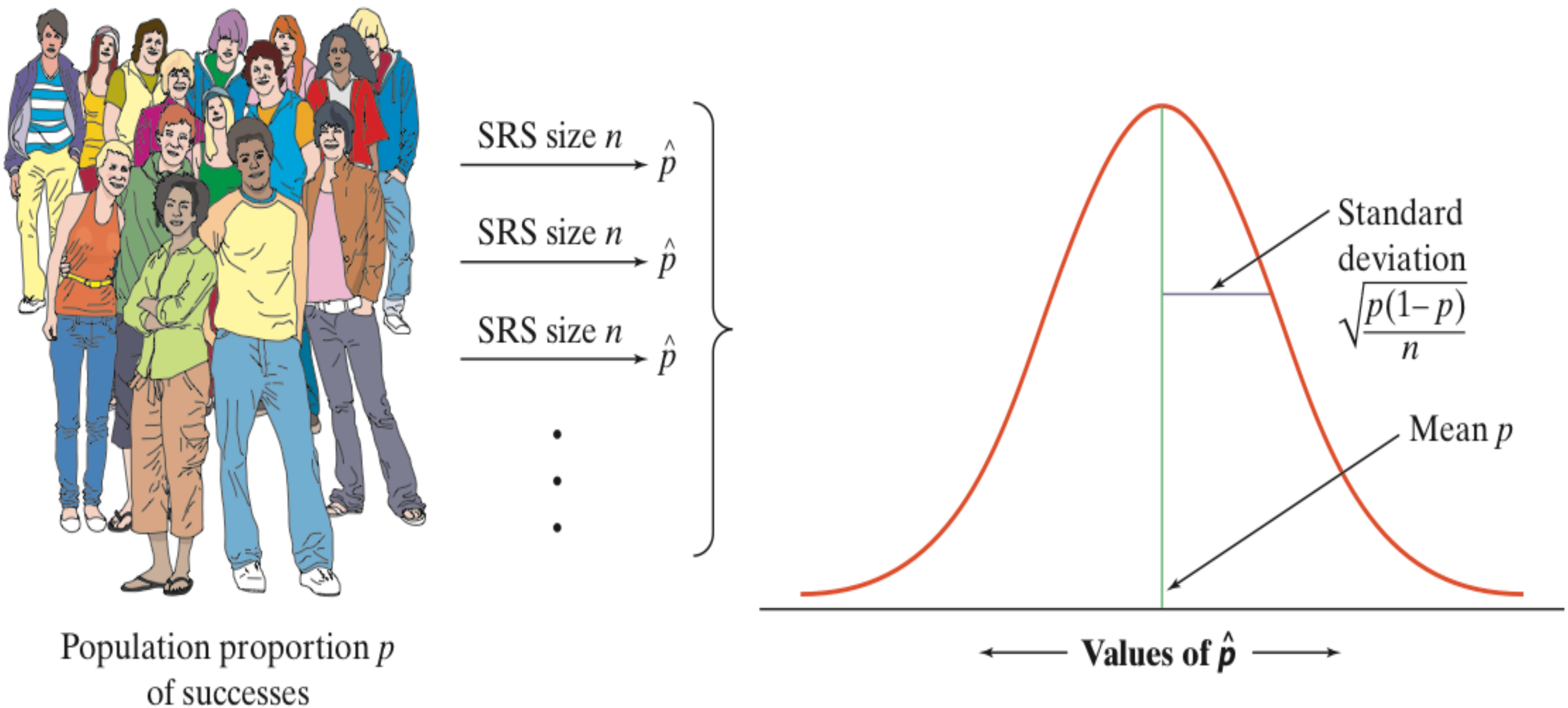
SAMPLING DISTRIBUTION OF A SAMPLE PROPORTION

- Draw an SRS of size *n* from a large population that contains proportion p of successes.
- Let $\hat{p}$ be the **sample proportion** of successes:

$$\hat{p} = \frac{\text{number of successes in the sample}}{n}$$

- Then: The **mean** of the sampling distribution is p.
- The **standard deviation** of the sampling distribution is $\sqrt{\frac{p(1-p)}{n}}$.
- As the sample size increases, the sampling distribution of $\hat{p}$ becomes **approximately Normal**. That is, for large *n*, $\hat{p}$ has approximately the $N\left(p, \sqrt{p(1-p)/n}\right)$ distribution.

Sampling Distribution of a Sample Proportion



Confidence Intervals for a Population Proportion

- We note the standard deviation of $\hat{p}$ depends on the parameter, p—a value we don't know. We therefore estimate the standard deviation with the standard error of $\hat{p}$:

$$\mathrm{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

LARGE-SAMPLE CONFIDENCE INTERVAL FOR A POPULATION PROPORTION

- Draw an SRS of size *n* from a population having unknown proportion *p* with some characteristic. An approximate level *C* **confidence interval for $\boldsymbol{p}$** is

$$\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$$

  where *z** is the critical value for the standard Normal density curve with area C between –*z** and *z**.
- Use this interval only when the numbers of successes and failures in the sample are both at least 15.

Example

Your instructor claims 50% of the beads in a container are red. A random sample of 251 beads is selected, of which 107 are red. Calculate and interpret a 90% confidence interval for the proportion of red beads in the container. Use your interval to comment on this claim.

z	**.03**	**.04**	**.05**
– 1.7	.0418	.0409	.0401
– 1.6	.0516	.0505	.0495
– 1.5	.0630	.0618	.0606

✓ sample proportion = 107/251 = 0.426

✓ This is an SRS and there are 107 successes and 144 failures. Both are greater than 15.

✓ For a 90% confidence level, z^* = 1.645.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.426 \pm 1.645\sqrt{\frac{(0.426)(1-0.426)}{251}}$$

$$= 0.426 \pm 0.051$$

$$= (0.375,\ 0.477)$$

We are 90% confident that the interval from 0.375 to 0.477 captures the actual proportion of red beads in the container.

Since this interval gives a range of plausible values for p *and since 0.5 is not contained in the interval, we have reason to doubt the claim.*

Choosing the Sample Size

- The margin of error in the large-sample confidence interval for p is

$$m = z^*\sqrt{\hat{p}(1-\hat{p})/n}$$

- Here z^* is the standard Normal critical value for the level of confidence we want.

SAMPLE SIZE FOR DESIRED MARGIN OF ERROR

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

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

where p^* is a guessed value for the sample proportion. The margin of error will always be less than or equal to *m* if you take the guess p^* to be 0.5.

Significance Test for a Proportion

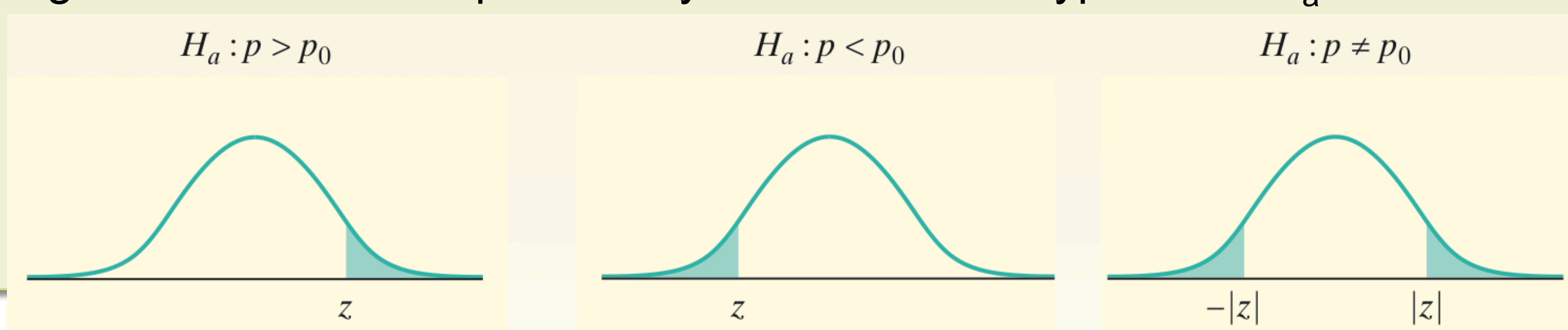
The z statistic has approximately the standard Normal distribution when H_0 is true. P-values therefore come from the standard Normal distribution. Here is a summary of the details for a **z test for a proportion.**

z Test for a Proportion

Choose an SRS of size n from a large population that contains an unknown proportion p of successes. To test the hypothesis H_0: $p = p_0$, compute the z statistic:

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

Find the P-value by calculating the probability of getting a z statistic this large or larger in the direction specified by the alternative hypothesis H_a:



Example

A potato-chip producer has just received a truckload of potatoes from its main supplier. If the producer determines that more than 8% of the potatoes in the shipment have blemishes, the truck will be sent away to get another load from the supplier. A supervisor selects a random sample of 500 potatoes from the truck. An inspection reveals that 47 of the potatoes have blemishes. Carry out a significance test at the $\alpha = 0.10$ significance level. What should the producer conclude?

State: We want to perform a test at the $\alpha = 0.10$ significance level of

$$H_0: p = 0.08$$
$$H_a: p > 0.08$$

where p is the actual proportion of potatoes in this shipment with blemishes.

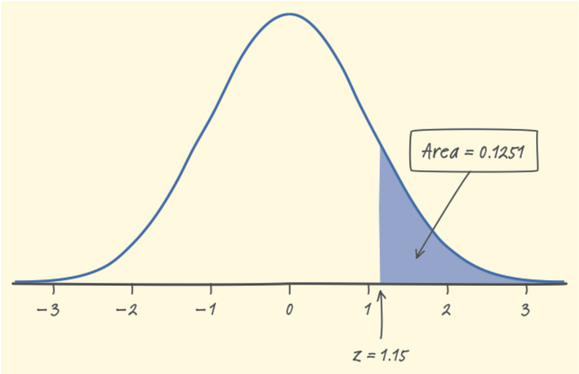
Plan: If conditions are met, we should do a one-sample z test for the population proportion p.

- ✓ *Random* The supervisor took a random sample of 500 potatoes from the shipment.
- ✓ *Normal* Assuming H_0: $p = 0.08$ is true, the expected numbers of blemished and unblemished potatoes are $np_0 = 500(0.08) = 40$ and $n(1 - p_0) = 500(0.92) = 460$, respectively. Because both of these values are at least 10, we should be safe doing Normal calculations.

Example

Do: The sample proportion of blemished potatoes is $\hat{p} = 47/500 = 0.094$.

Test statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0.094 - 0.08}{\sqrt{\dfrac{0.08(0.92)}{500}}} = 1.15$



***P*-value** The desired *P*-value is:

$P(z \geq 1.15) = 1 - 0.8749 = 0.1251$

Conclude: Since our *P*-value, 0.1251, is greater than the chosen significance level of $\alpha = 0.10$, we fail to reject H_0. There is not sufficient evidence to conclude that the shipment contains more than 8% blemished potatoes. The producer will use this truckload of potatoes to make potato chips.

Which of the following symbols represents the sample proportion?

a) p
b) $\hat{p}$
c) $\bar{x}$
d) P-value

The standard error of $\hat{p}$ ______________________________.

a) is the exact standard deviation of the sampling distribution of $\hat{p}$.
b) is an estimate of the standard deviation of the sampling distribution of $\hat{p}$
c) measures the maximum difference expected between p and $\hat{p}$ at a specified level of confidence

Suppose we are interested in testing

$$H_0: p = p_0$$
$$H_a: p \neq p_0$$

What is p_0 in these hypotheses?

a) The sample proportion.
b) The population proportion.
c) The hypothesized value of the population proportion.

For a hypothesis test for p, the P-value is________________________.

a) the probability that p equals p_0.
b) the probability of getting a $\hat{p}$ equal to or more extreme than the observed value of $\hat{p}$ assuming that H_0 is true
c) the probability of getting a $\hat{p}$ equal to or more extreme than the observed value of $\hat{p}$ assuming that H_a is true

Signif. Tests for a Proportion

I don't believe the claim that the probability of heads when spinning a 1961 penny is 0.1. To test the claim, I spin such a penny 200 times and get 'head' 26 times. What is the test statistic?

a) $$z = \frac{0.13 - 0.1}{\sqrt{\frac{(0.13)(0.87)}{200}}}$$

b) $$z = \frac{0.13 - 0.1}{\sqrt{\frac{(0.1)(0.9)}{200}}}$$

c) $$z = \frac{0.1 - 0.5}{\sqrt{\frac{(0.1)(0.9)}{200}}}$$

d) $$z = \frac{0.13 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{200}}}$$

Choose the Sample Size

We want to estimate *p* using a confidence interval that has a 3% margin of error, and we want to have 99% confidence that our interval captures *p*. We think that *p* is between .3 and .7. What sample size n should we use?

a) $(1.96/.03)^2(.5)(1-.5)=1067.11$, round up to 1068
b) $(2.576/.03)^2(.5)(1-.5)=1843.27$, round down to 1843
c) $(1.96/.03)^2(.5)(1-.5)=1067.11$, round down to 1067
d) $(2.576/.03)^2(.5)(1-.5)=1843.27$, round up to 1844
e) $(2.576/.03)^2\ (.3)(1-.3)=1548.35$

Choose the Sample Size

How many American adults aged 18 and over must be interviewed to estimate the proportion who own MP3 players within ±0.02 with 95% confidence? We think that *p* is between .4 and .6. What sample size n should we use?

a) $(1.96/.02)^2(.4)(1-.4)=2304.96$, round up to 2305
b) $(2.576/.02)^2(.5)(1-.5)=4147.36$, round down to 4147
c) $(1.96/.02)^2(.5)(1-.5)=2401$
d) $(2.576/.02)^2(.6)(1-.6)=3981.47$, round up to 3982

Test about a Population Proportion

State DMV records indicate that of all vehicles undergoing emissions testing during the previous year, 70% passed on the first try. A random sample of 200 cars tested in a particular county during the current year yields 124 that passed on the initial test. At the 1 % level of significance, is this evidence the true proportion for this county during the current year has decreased since the previous statewide proportion?

The alternative hypothesis is	The sample proportion is
a) $H_a : p_0 > .7$ b) $H_a : p < .7$ c) $H_a : p < 124$ d) $H_a : p_0 < .7$	a) $\hat{p} = \frac{200}{124} = 1.61$ b) $\hat{p} = \frac{124}{200} = 0.62$ c) $\hat{p} = \frac{70}{100} = 0.7$ d) $p_0 = \frac{124}{200} = 0.62$

Test about a Population Proportion

The test statistic is

a) $$z = \frac{0.70 - 0.62}{\sqrt{\frac{(0.7)(1-0.7)}{200}}} = 2.47$$

b) $$z = \frac{0.62 - 0.70}{\sqrt{\frac{(0.7)(1-0.7)}{200}}} = -2.47$$

c) $$z = \frac{0.62 - 0.70}{\sqrt{\frac{(0.62)(1-0.62)}{200}}} = -2.33$$

d) $$z = \frac{0.70 - 0.62}{\sqrt{\frac{(0.62)(1-0.62)}{200}}} = 2.33$$

Test about a Population Proportion

Using table C, the p-value is:

a) $0.005 < p - value < 0.01$

b) $0.01 < p - value < 0.02$

c) $p - value < 0.005$

Using table A, the p-value is:

a) $p - value = 0.0068$

b) $p - value = 1 - 0.0068 = 0.9932$

c) $p - value = 0.099$

Test about a Population Proportion

The conclusion is:

a) Significant (p-value > 0.01)
b) Significant (p-value < 0.01)
c) Not significant (p-value > 0.01)
d) Not significant (p-value < 0.01)

The conclusion in terms of the problem is:

a) There is evidence the true proportion for this county during the current year has decreased since the previous statewide proportion.

b) There is no evidence the true proportion for this county during the current year has decreased since the previous statewide proportion

Test about a Population Proportion

The conditions for inference for this test of significance are:

Check for an SRS from the population and

a) $np_0 = (200)(0.7) = 140 \geq 10$ and $n(1 - p_0) = (200)(1 - 0.7) = 60 \geq 10$

b) normality in the population or the sample size is fairly large

c) no more than 20% of expected cell counts are < 5

d) the population must be much larger than the sample