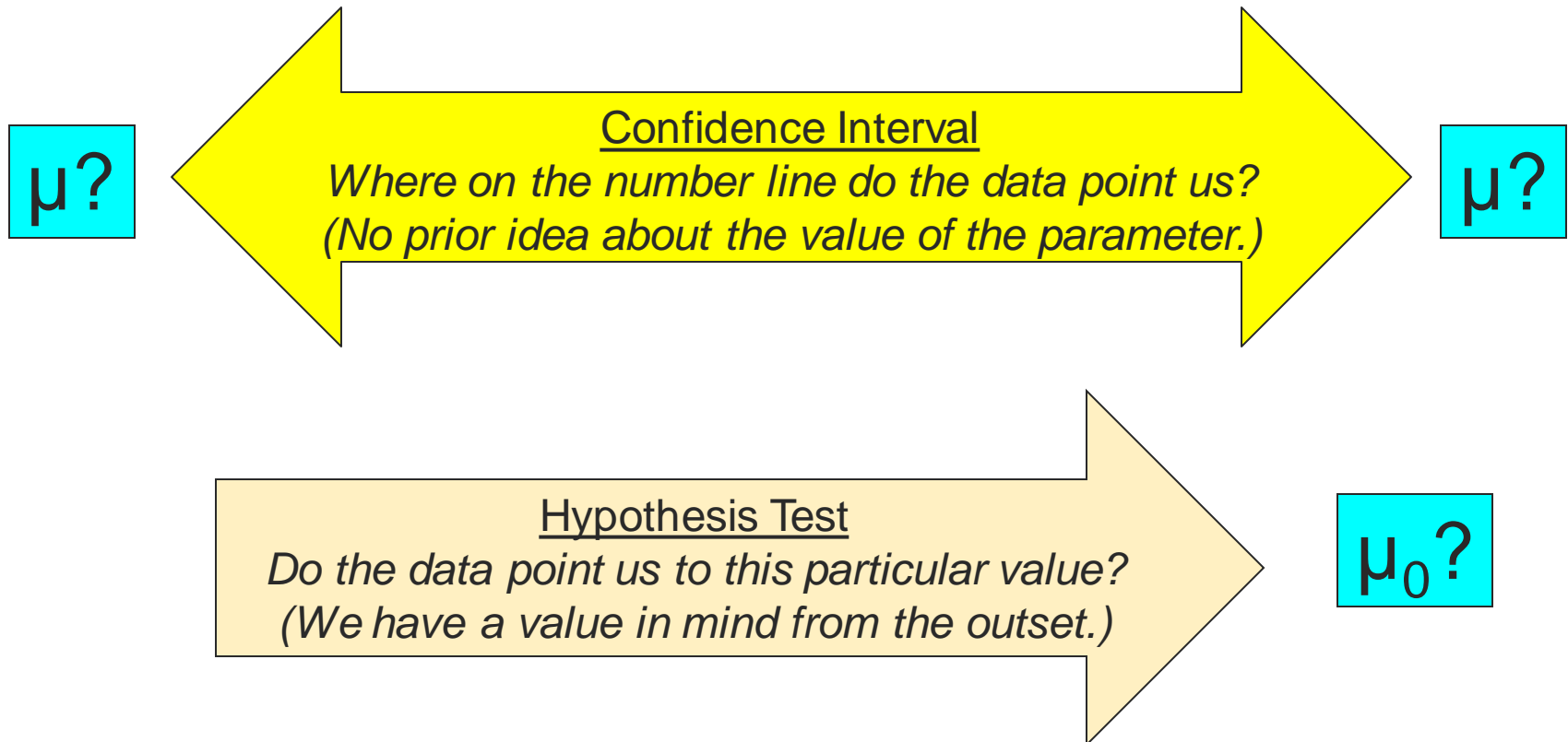# Statistics

Chapter 8: Tests of Hypotheses

Where We've Been

- Calculated *point estimators* of population parameters
- Used the sampling distribution of a statistic to assess the reliability of an estimate through a *confidence interval*

Where We're Going

- Test a specific value of a population parameter
- Measure the reliability of the test

8.1: The Elements of a Test of Hypotheses

μ?

Confidence Interval
Where on the number line do the data point us?
(No prior idea about the value of the parameter.)

μ?

Hypothesis Test
Do the data point us to this particular value?
(We have a value in mind from the outset.)

μ_0?

8.1: The Elements of a Test of Hypotheses

Null Hypothesis: H_0
- This will be supported unless the data provide evidence that it is false
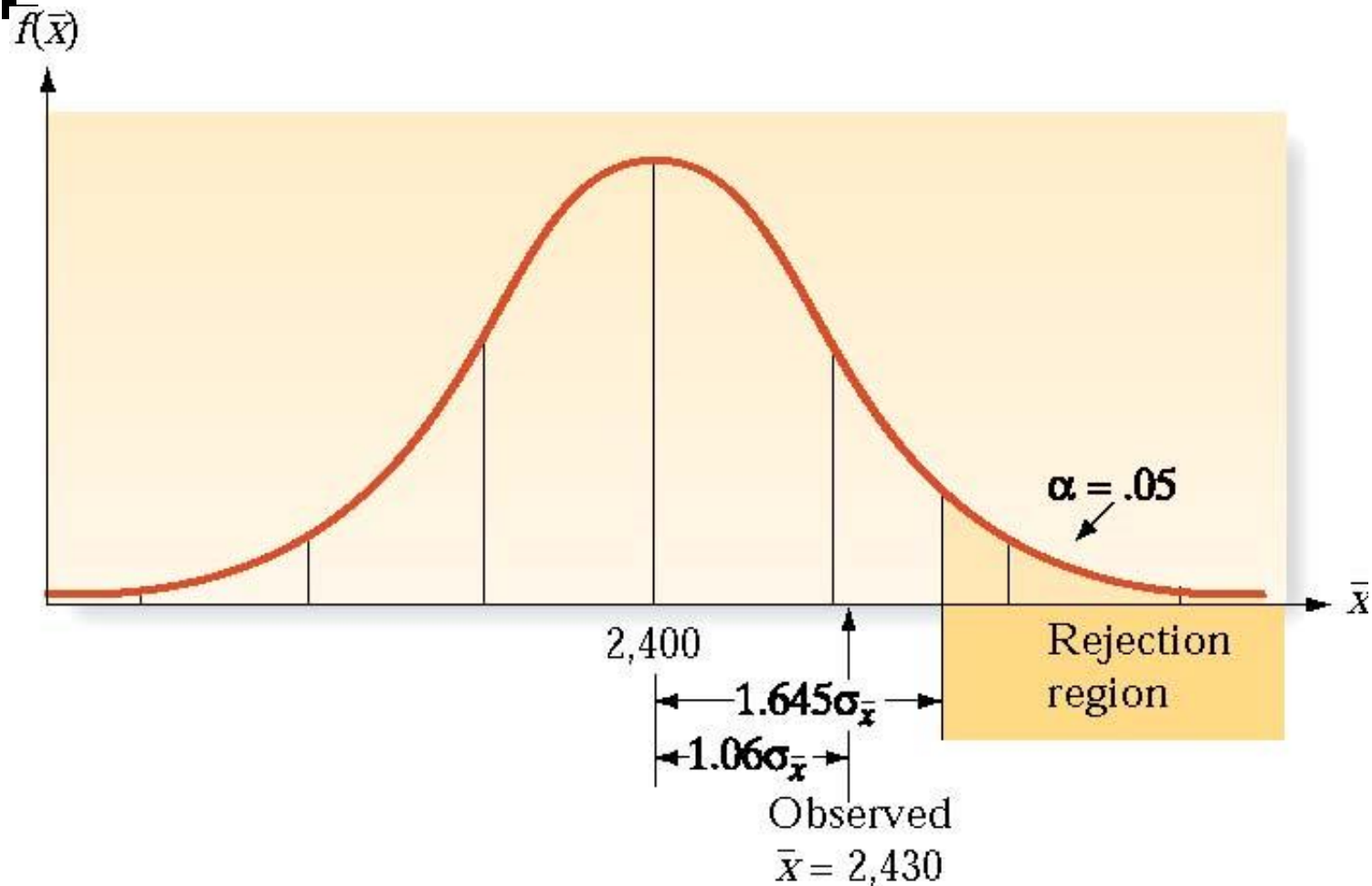- The *status quo*

Alternative Hypothesis: H_a
- This will be supported if the data provide sufficient evidence that it is true
- The *research hypothesis*

8.1: The Elements of a Test of Hypotheses

- If the *test statistic* has a high probability when H_0 is true, then H_0 is *not rejected.*

- If the *test statistic* has a (very) low probability when H_0 is true, then H_0 is *rejected.*

8.1: The Elements of a Test of Hypotheses

8.1: The Elements of a Test of Hypotheses

Reality↓ / Test Result →	Do not reject H_0	Reject H_0
H_0 is true	Correct!	Type I Error: rejecting a true null hypothesis P(Type I error) = α
H_0 is false	Type II Error: not rejecting a false null hypothesis P(Type II error) = β	Correct!

8.1: The Elements of a Test of Hypotheses

Note: Null hypotheses are either rejected, or else there is insufficient evidence to reject them. (I.e., we don't *accept* null hypotheses.)

8.1: The Elements of a Test of Hypotheses

- *Null hypothesis (H_0)*: A theory about the values of one or more parameters
  - Ex.:   *H_0*: $\mu = \mu_0$ (a specified value for μ)
- *Alternative hypothesis (H_a)*: Contradicts the null hypothesis
  - Ex.:   *H_0*: $\mu \neq \mu_0$
- *Test Statistic*: The sample statistic to be used to test the hypothesis
- *Rejection region*: The values for the test statistic which lead to rejection of the null hypothesis
- *Assumptions*: Clear statements about any assumptions concerning the target population
- *Experiment and calculation of test statistic*: The appropriate calculation for the test based on the sample data
- *Conclusion*: Reject the null hypothesis (with possible Type I error) or do not reject it (with possible Type II error)

8.1: The Elements of a Test of Hypotheses

Suppose a new interpretation of the rules by soccer referees is expected to increase the number of yellow cards per game. The average number of yellow cards per game had been 4. A sample of 121 matches produced an average of 4.7 yellow cards per game, with a standard deviation of .5 cards. At the 5% significance level, has there been a change in infractions called?

8.1: The Elements of a Test of Hypotheses

$$H_0: \mu = 4$$
$$H_a: \mu \neq 4$$

Sample statistic $\bar{x} = 4.7$

$$\alpha = 0.05$$

Test statistic: $z^* = \dfrac{\bar{x} - \mu_0}{s_{\bar{x}}} = \dfrac{4.7 - 4}{.064} = 10.94$

$$z_{0.025} = 1.96$$

Since |z*|>1.96, we reject the null hypothesis.

Conclusion: There has been a change in infractions called.

8.2: Large-Sample Test of a Hypothesis about a Population Mean

The null hypothesis is usually stated as an equality …
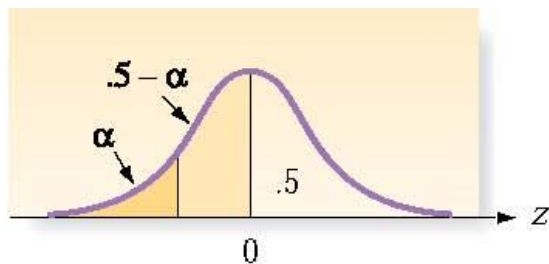
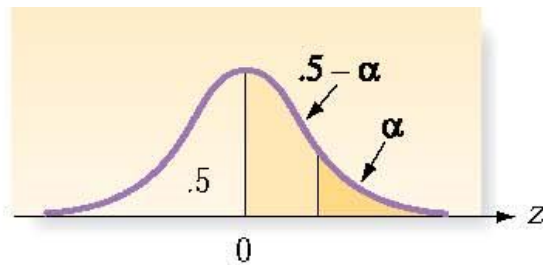$$H_0: \mu = \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu \neq \mu_0$$

$$H_a: \mu > \mu_0$$

… even though the alternative hypothesis can be either an equality or an inequality.
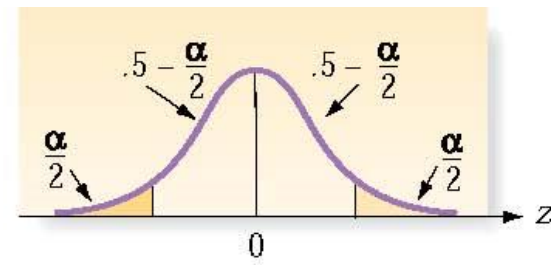
8.2: Large-Sample Test of a Hypothesis about a Population Mean
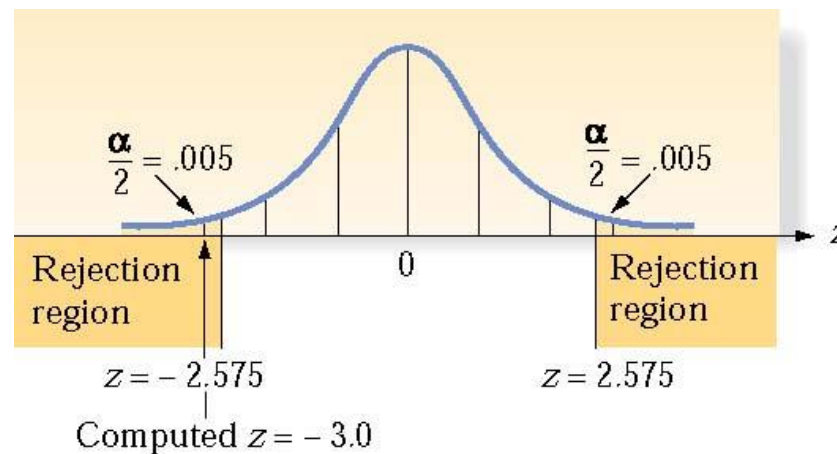

a. Form of H_a: $<$

b. Form of H_a: $>$

c. Form of H_a: $\neq$

8.2: Large-Sample Test of a Hypothesis about a Population Mean

Rejection Regions for Common Values of α

	Lower Tailed	Upper Tailed	Two tailed
$\alpha = .10$	$z < - 1.28$	$z > 1.28$	$\mid z \mid > 1.645$
$\alpha = .05$	$z < - 1.645$	$z > 1.645$	$\mid z \mid > 1.96$
$\alpha = .01$	$z < - 2.33$	$z > 2.33$	$\mid z \mid > 2.575$

8.2: Large-Sample Test of a Hypothesis about a Population Mean

One-Tailed Test

- $H_0 : \mu = \mu_0$

$H_a : \mu < \text{ or } > \mu_0$

Test Statistic: $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$

Rejection Region: $|z| > z_\alpha$

Two-Tailed Test

- $H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

Test Statistic: $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$

Rejection Region: $|z| > z_{\alpha/2}$

Conditions: 1) A random sample is selected from the target population.
2) The sample size n is large.

8.2: Large-Sample Test of a Hypothesis about a Population Mean

- The *Economics of Education Review* (Vol. 21, 2002) reported a mean salary for males with postgraduate degrees of $61,340, with an estimated standard error ($s_{\bar{x}}$) equal to $2,185. We wish to test, at the $\alpha = .05$ level, $H_0: \mu = \$60,000$.

8.2: Large-Sample Test of a Hypothesis about a Population Mean

- The *Economics of Education Review* (Vol. 21, 2002) reported a mean salary for males with postgraduate degrees of $61,340, with an estimated standard error ($s_{\bar{x}}$) equal to $2,185. We wish to test, at the α = .05 level, H_0: μ = $60,000.

- H_0: $\mu = 60{,}000$
  H_a: $\mu \neq 60{,}000$
  Test Statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

$$z = \frac{61{,}340 - 60{,}000}{2{,}185}$$

$$z = .613$$

Rejection Region: | z | > z$_{.025}$ = 1.96

Do not reject H_0

8.3:Observed Significance Levels: p-Values

Suppose z = 2.12.
P(z > 2.12) = .0170.

Reject H_0 at the α= .05 level

Do not reject H_0 at the α= .01 level

But it's pretty close, isn't it?

8.3: Observed Significance Levels: *p* - Values

The **observed significance level**, or ***p-value***, for a test is the probability of observing a test statistic as extreme or more than the one actually observed (*z**) assuming the null hypothesis is true.

$$P(z \geq z^* \mid H_0)$$

The lower this probability, the less likely H_0 is true.

8.3:Observed Significance Levels: *p* - Values

Let's go back to the *Economics of Education Review* report ($\overline{x}$= $61,340, $s_{\overline{x}}$ = $2,185). This time we'll test H_0: μ = $65,000.

H_0: $\mu = 65,000$

H_a: $\mu \neq 65,000$

Test Statistic:

$$z = \frac{\overline{x} - \mu_0}{\sigma_{\overline{x}}}$$
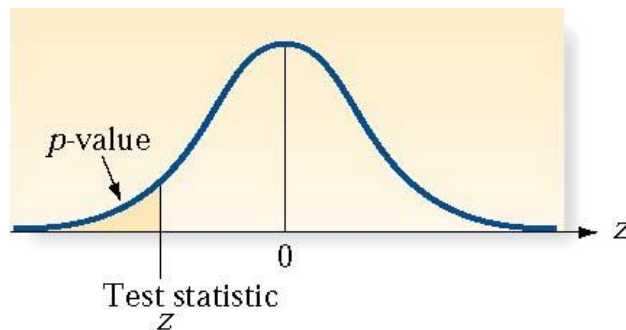
$$z = \frac{61,340 - 65,000}{2,185}$$
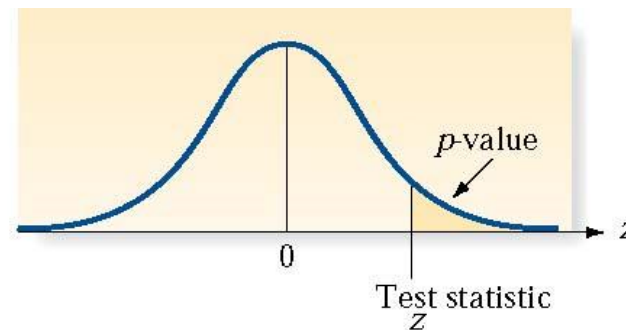
$$z = -1.675$$

p-value: $2P(\overline{x} < 61,340 \,|H_0) = P(|z| > 1.675) = .0475$

8.3:Observed Significance Levels: *p* - Values

- Reporting test results
  - Choose the maximum tolerable value of α
  - If the p-value $< \alpha$, reject H_0
  - If the p-value $> \alpha$, do not reject H_0

a. Lower–tailed test, $H_a: \mu < \mu_0$

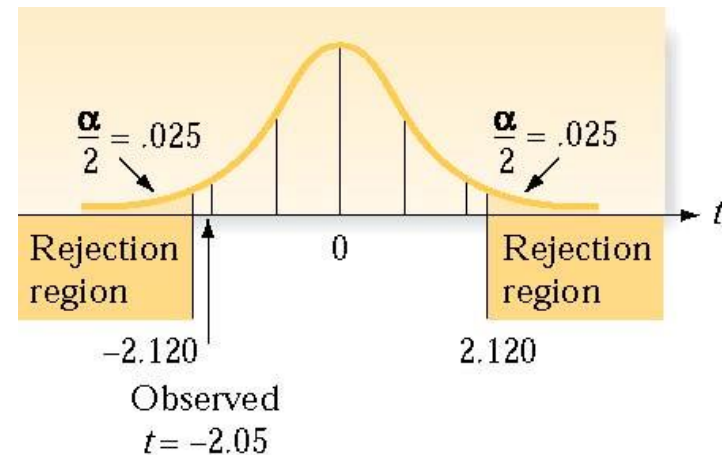b. Upper–tailed test, $H_a: \mu > \mu_0$

8.4: Small-Sample Test of a Hypothesis about a Population Mean

If the sample is small and σ is unknown, testing hypotheses about μ requires the t-distribution instead of the z-distribution.

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$\frac{\alpha}{2} = .025$ $\frac{\alpha}{2} = .025$

Rejection region   0   Rejection region

−2.120   2.120

Observed
$t = -2.05$

8.4: Small-Sample Test of a Hypothesis about a Population Mean

One-Tailed Test

- $H_0 : \mu = \mu_0$
  $H_a : \mu <$ or $> \mu_0$

Test Statistic: $\quad t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$

Rejection Region: $|t| > t_\alpha$

Two-Tailed Test

- $H_0 : \mu = \mu_0$
  $H_a : \mu \neq \mu_0$

Test Statistic: $t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$

Rejection Region: $|t| > t_{\alpha/2}$

Conditions:
1) A random sample is selected from the target population.
2) The population from which the sample is selected is approximately normal.
3) The value of t_α is based on $(n - 1)$ degrees of freedom

8.4: Small-Sample Test of a Hypothesis about a Population Mean

Suppose copiers average 100,000 between paper jams. A salesman claims his are better, and offers to leave 5 units for testing. The average number of copies between jams is 100,987, with a standard deviation of 157. Does his claim seem believable?

8.4: Small-Sample Test of a Hypothesis about a Population Mean

Suppose copiers average 100,000 between paper jams. A salesman claims his are better, and offers to leave 5 units for testing. The average number of copies between jams is 100,987, with a standard deviation of 157. Does his claim seem believable?

$H_0: \mu = 100{,}000$

$H_a: \mu > 100{,}000$

Test Statistic:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$$t = \frac{100{,}987 - 100{,}000}{157 / \sqrt{5}}$$

$$t = 14.06$$

p-value: $P(\bar{x} > 100{,}987 | H_0) = P(t_{df=4} > 14.06) < .001$

8.4: Small-Sample Test of a Hypothesis about a Population Mean

Suppose copiers average 100,000 between paper jams. A salesman claims his are better, and offers to leave 5 units for testing. The average number of copies between jams is 100,987, with a standard deviation of 157. Does his claim seem believable?

Reject the null hypothesis based on the very low probability of seeing the observed results if the null were true.
So, the claim does seem plausible.

$$t = \frac{100,\!\ldots - \ldots 00,\!000}{\ldots / \sqrt{5}}$$

$$t = 14.0\ldots$$

p-value: $P(\overline{x} > 100{,}987 | H_0) = P(t_{df=4} > 14.06) < .001$

One-Tailed Test

- $H_0 : p = p_0$

$H_a : p <$ or $> p_0$

Test Statistic: $z = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}}$

Rejection Region: $|z| > z_\alpha$

Two-Tailed Test

- $H_0 : p = p_0$

$H_a : p \neq p_0$

Test Statistic: $z = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}}$

Rejection Region: $|z| > z_{\frac{\alpha}{2}}$

p_0 = hypothesized value of p, $\sigma_{\hat{p}} = \sqrt{\dfrac{p_0 q_0}{n}}$, and $q_0 = 1 - p_0$

Conditions: 1) A random sample is selected from a binomial population.
　　　　　　2) The sample size n is large (i.e., np_0 and nq_0 are both > 15).

8.5: Large-Sample Test of a Hypothesis about a Population Proportion

Rope designed for use in the theatre must withstand unusual stresses. Assume a brand of 3" three-strand rope is expected to have a breaking strength of 1400 lbs. A vendor receives a shipment of rope and needs to (destructively) test it.

- The vendor will reject any shipment which cannot pass a 1% defect test (that's harsh, but so is falling scenery during an aria). 1500 sections of rope are tested, with 20 pieces failing the test. At the α = .01 level, should the shipment be rejected?

8.5: Large-Sample Test of a Hypothesis about a Population Proportion

The vendor will reject any shipment that cannot pass a 1% defects test . 1500 sections of rope are tested, with 20 pieces failing the test. At the α = .01 level, should the shipment be rejected?

H_0: p = .01

H_a: p > .01

Rejection region: z > 2.236

Test statistic:

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

$$z = \frac{.013 - .01}{\sqrt{(.013)(.987)/1500}}$$

$$z = 1.14$$

■ The vendor will reject any shipment that cannot pass a 1% defects test . 1500 sections of rope are tested, with 20 pieces failing the test. At the α = .01 level, should the shipment be rejected?

H_0: p = .01

H_a: p

There is insufficient evidence to reject the null hypothesis based on the sample results.

$z = \dfrac{}{\sigma_{\hat{p}}}$

$$z = \frac{.013 - .01}{\sqrt{(.013)(.987)/1500}}$$

$$z = 1.14$$

31

8.6: A Nonparametric Test About a Population Median

- The **sign test** provides inferences about population *medians*, or *central tendencies*, when skewed data or an outlier would invalidate tests based on normal distributions.

8.6: A Nonparametric Test About a Population Median

- One-tailed test for a Population Median, η

$$H_0 : \eta \leq (or \geq)\eta_0$$

$$H_a : \eta > (or <)\eta_0$$

Test Statistic:

S = number of sample measurements greater than (less than) η_0

- Two-tailed test for a Population Median, η

$$H_0 : \eta = \eta_0$$

$$H_a : \eta \neq \eta_0$$

Test Statistic:

S = larger of S_1 and S_2 where S_1 is the number of measurements less than η_0 and S_2 the number greater than η_0

8.6: A Nonparametric Test About a Population Median

- One-tailed test for a Population Median

  Observed significance level:

  p-value = P($x>S$)

- Two-tailed test for a Population Median

  Observed significance level:

  p-value = 2P($x>S$)

where x has a binomial distribution with parameters

n and p = .5.

Reject H_0 if p-value $< \alpha$.

8.6: A Nonparametric Test About a Population Median

Median time to failure for a band of compact disc players is 5,250 hours. Twenty players from a competitor are tested, with failure times from 5 hours to 6,575 hours. Fourteen of the players exceed 5,250 hours.

Do the competitor's machines perform differently?

8.6: A Nonparametric Test About a Population Median

Median time to failure for a band of compact disc players is 5,250 hours. Twenty players from a competitor are tested, with failure times from 5 hours to 6,575 hours. Fourteen of the players exceed 5,250 hours.

Do the competitor's machines perform differently?

$$H_0 : \eta = 5,250$$

$$H_a : \eta \neq 5,250$$

$$\alpha = .10$$

$$z^*_{\alpha/2} = z^*_{.05} = 1.645$$

$$z = \frac{(s - .5) - .5n}{.5\sqrt{n}} = \frac{13.5 - 10}{.5\sqrt{20}} = 1.565 < 1.645$$

8.6: A Nonparametric Test About a Population Median

Median time to failure for a band of compact disc players is 5,250 hours. Twenty players from a competitor are tested, with failure times from 5 hours to 6,575 hours. Fourteen of the players exceed 5,250 hours.

Do the competitor's machines perform differently?

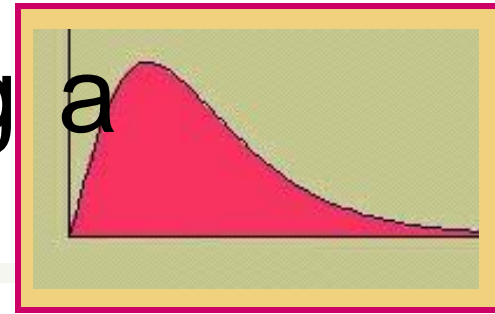$$H_0 : \eta = 5,250$$

$$H_a : \eta \neq 5,250$$

$$\alpha = .10$$

$$z^*_{\alpha/2} = z^*_{.05} = 1.645$$

$$z = \frac{(s-.5) - .5n}{.5\sqrt{n}} = \frac{13.5 - 10}{.5\sqrt{20}} = 1.565 < 1.645$$
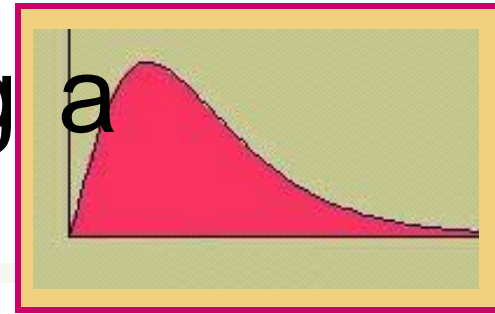
Do not reject H_0

8.7: Inference concerning a Population Variance

Sometimes the primary parameter of interest is not the population mean μ but rather the population variance σ^2.

We choose a random sample of size n from a normal distribution.

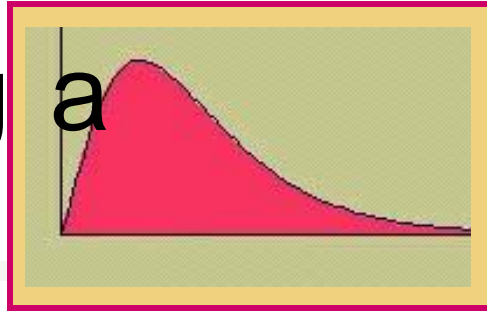# 8.7: Inference concerning a Population Variance

The sample variance s^2 can be used in its standardized form:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

which has a Chi-Square distribution with n - 1 degrees of freedom.

8.7: Inference concerning a Population Variance

To test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_a :$ one or two tailed

we use the test statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \text{ with a rejection region based on}$$

a chi-square distribution with $df = n - 1$.

Confidence interval:

$$\frac{(n-1)s^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{(1-\alpha/2)}^2}$$

8.7: Inference concerning a Population Variance

- A cement manufacturer claims that his cement has a compressive strength with a standard deviation of 10 kg/cm^2 or less. A sample of $n = 10$ measurements produced a mean and standard deviation of 312 and 13.96, respectively.

8.7: Inference concerning a Population Variance

•A cement manufacturer claims that his cement has a compressive strength with a standard deviation of 10 kg/cm^2 or less. A sample of $n = 10$ measurements produced a mean and standard deviation of 312 and 13.96, respectively.

A test of hypothesis:

H_0: $\sigma = 10$ (claim is correct)

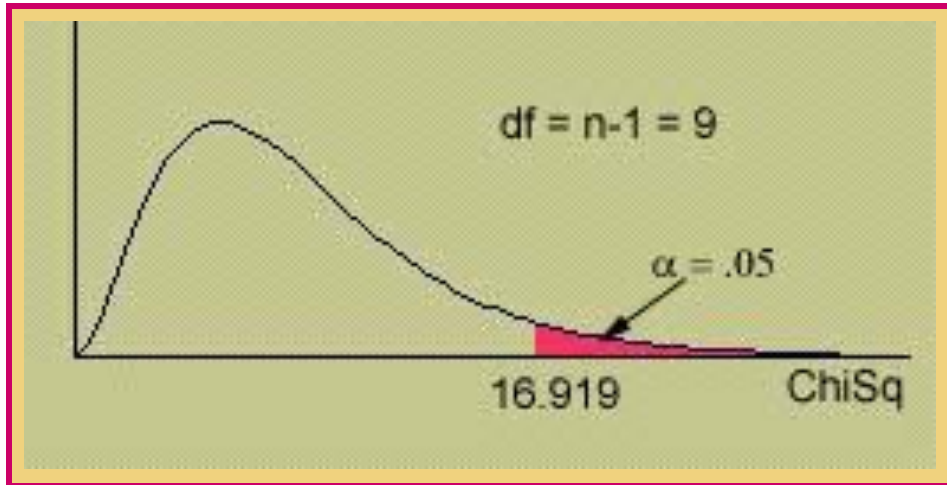H_a: $\sigma > 10$ (claim is wrong)

uses the test statistic:

$$\chi^2 = \frac{(n-1)s^2}{10^2} = \frac{9(13.96^2)}{100} = 17.5$$

8.7: Inference concerning a Population Variance

• Do these data produce sufficient evidence to reject the manufacturer's claim? Use $\alpha = .05$.

df = n-1 = 9

$\alpha = .05$

16.919   ChiSq

Rejection region: Reject H_0 if $\chi^2 > 16.919$ ($\alpha = .05$).

Conclusion: Since $\chi^2 = 17.5$, H_0 is rejected. The standard deviation of the cement strengths is more than 10.

Equivalence of Confidence Interval & Testing

- Testing H_0: $\mu = \mu_0$ vs H_a: $\mu \neq \mu_0$ at .05 level: Reject for $|z| > 1.96$.

- (1-.05) Confidence interval for μ: $|z| < 1.96$

- Thus rejection region is the region outside the confidence interval.

- For one-tailed tests, we can build one sided confidence intervals of the form (a, ∞) or $(-\infty, b)$.

Exercises

The average number of minutes of a television commercial is 4.8. Write down the null and alternative hypothesis for testing the above statement. Assuming the commercial time is normally distributed, give the appropriate rejection region for each of the following sample sizes and significance levels.

a. n = 6,   α = 0.01

b. n = 12, α = 0.05

c. n = 20,   α = 0.1

Exercises

The average number of minutes of a television commercial is 4.8. Assuming the commercial time is normally distributed, give the p-value for each of the following sample sizes, sample statistics and alternative hypotheses.

a. n = 6, $\bar{x} = 4.5, s = 1.5, H_a: \mu < 4.8$

b. n = 12, $\bar{x} = 4.5, s = 1.5, H_a: \mu > 4.8$

c. n = 20, $\bar{x} = 4.5, s = 1.5, H_a: \mu \neq 4.8$