7-2

In this class, a <u>hypothesis</u> is a statement about a population parameter(s) for one or more populations.

In a <u>hypothesis test</u>, we need:

1. a <u>null hypothesis</u> (H_0)
   - a default assumption (<u>Ex</u> A given coin is fair.)
2. an <u>alternative hypothesis</u> (H_1 or H_A)
   (<u>Ex</u> A given coin is not fair.)
3. a <u>significance level</u> (α) - prechosen by user
   0.05 - common

We will either

1. reject H_0 in favor of H_A, or

If OJ is "not guilty," does that mean he's innocent?

2. fail to reject H_0
   (do <u>not</u> say "we accept H_0").
   We don't have sufficient evidence to say that H_0 is wrong.

If α is lower → "harder" to <u>reject H_0</u>.
"publish"

To be continued...

Ⓐ 3 Types of Tests

① Two-tailed

Ex Test the claim that a magician's coin is fair.

Let $p = P$(coin comes up heads).

$H_0: p = 0.5$ (← Claim)

Note 1: H_0 will always involve "=", though some books disagree.

Note 2: The claim could be H_A. Here, H_0.

$H_A: p \neq 0.5$
always for 2-tailed 0.5

$\alpha = 0.01$, say. (We'll discuss this later.)

Gather sample data, and perform a test.

If we reject H_0, our final conclusion is:
"There is sufficient evidence against the claim that the coin is fair." (GUILTY!)

If we do not reject (or "fail to reject") H_0:
"There is insufficient evidence ...."
(NOT GUILTY!)
but don't say "INNOCENT!"

We can "accept" H_A, but never H_0.

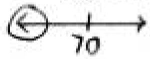
② Left-tailed (one-tailed) test

Ex Test the claim that the average score on a test is less than 70 points.

Let μ = the average test score in the class.

$H_0: \mu = 70$ ← "reference point"

$H_A: \mu < 70$ (← Claim)



$\alpha = 0.05$, say. (What is "sufficient" evidence?)

(Gather sample data, and perform a test.)

If we reject H_0 (in favor of H_A):

"There is sufficient evidence to support the claim that the average test score in the class is less than 70 points."

If we do not reject H_0:

"There is insufficient evidence ...."

③ Right-tailed (one-tailed) test

Ex The test scores in Smith's class have a standard deviation (SD) of 12 points. Jones's class takes the same test. Test the claim that the SD in Jones's class is higher. (Maybe Smith's students had Smith before.)

Let σ = the SD of test scores in Jones's class.

$H_0: \sigma = 12$

$H_A: \sigma > 12$ (← Claim)

12

$\alpha = 0.10$, say.

(Gather sample data, and perform a test.)

If we reject H_0 (in favor of H_A):

"There is sufficient evidence to support the claim that the SD in Jones's class is higher (than in Smith's class)."

If we do not reject H_0:

"There is insufficient evidence...."

If you want to "prove" a claim, make it your H_A (research hypothesis) (if you can), and hope that H_0 is rejected in favor of H_A.

(B) What's α?

α, the significance level of a hypothesis test, must be determined before the test is performed.

α determines when there is "sufficient" evidence against H_0.

See Notes 4.16

6 cards all red

i.e., α determines whether what we see is "unusual" under the assumption that H_0 is true. If we get something unusual under H_0, we have a statistically significant result, and we reject H_0.

What if H_0 is true?

We may get an unusual result by chance!

Ex It is possible (though unlikely) to toss a fair coin 100 times and get 90 heads.

$\alpha = P(\text{we will reject } H_0 \mid H_0 \text{ is true})$

given/assuming that

Think: Convicting an innocent man.
Type I error

$\beta = P(\text{we will not reject } H_0 \mid H_0 \text{ is false})$

beta

1st: Still up to 13
Can do 29, 34, 35

Think: Letting a guilty man go free.
Type II error

7-2 cont., and 7-3 in 1st ed. 7-4 in 2nd ed. } TESTS FOR μ IF σ IS KNOWN

A Assumptions

(σ is known)

CLT applies:

1. $n > 30$, or
2. X roughly Normal

B One Ex., Three Methods

As in 6.09

A machine is designed to make 10.0-lb. widgets. Its widgets' weights are approx. normally distributed with $\sigma = 2.5$ lbs.
We take a sample of 18 widgets, and $\bar{x} = 11.2$ lbs.
Using a 0.10 significance level, test the claim that the mean weight of widgets produced by the machine is 10.0 lbs. ☆

Sol'n

Step 1: State H_0, H_A, α. Claim?

Let μ = mean weight of widgets produced by the machine.

$H_0: \mu = 10.0$ [lbs.] (Claim)
$H_A: \mu \neq 10.0$ [lbs.] (← Unless otherwise implied, use a two-tailed test.)
$\alpha = 0.10$

Method 1: Use a $1-\alpha$ CI

$1-\alpha = 1-0.10 = 0.90$

In Notes 6.09, we found that a 90% CI for μ was (10.3, 12.1).

This CI does not include 10.0, so we reject H_0.

Final Conclusion

We have sufficient evidence against the claim that the mean weight of widgets produced by the machine is 10.0 lbs. (at the 0.10 level of significance). (A)

Method 2: P-Value Method

Old Ex

$X \overset{\text{approx.}}{\sim}$ Normal $\Rightarrow$ CLT applies $\Rightarrow \bar{X} \sim$ Normal
$\sigma = 2.3$ lbs. $\Rightarrow \sigma$ known, so use z.
$\bar{x} = 11.2$ lbs. $\Rightarrow$ Key sample statistic
$n = 18$

$H_0: \mu = 10.0$
$H_A: \mu \neq 10.0$
$\alpha = 0.10$

For now, assume H_0 is true.
Under H_0, $\mu = 10.0$.

$$\bar{x} \sim N\left(\mu_{\bar{x}} = \mu = 10.0,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.3}{\sqrt{18}} \approx 0.5421\right)$$

Test statistic

People often blow compound fractions!

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \quad \left(\text{or } \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

compute 1st

$$\approx \frac{11.2 - 10.0}{0.5421}$$

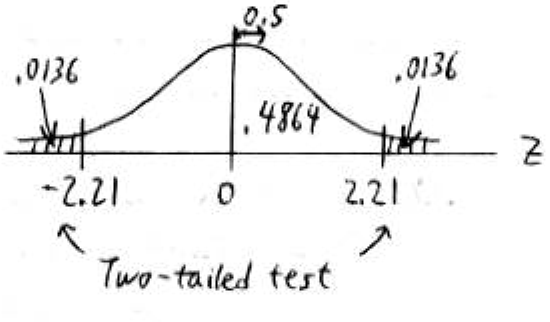
$$\approx 2.21 \qquad \text{Is this unusual?}$$

P-Value

= P(getting a test statistic [value] <u>at least as extreme</u> as what we got | H_0 is true)

see Notes 4.15-4.17

given/assuming that



$$\text{P-Value} \approx 2(.0136) = .0272$$

Decision Rule

If P-value $\le \alpha$, we say that our test stat [value] is "unusual" under H_0. We reject H_0.

If P-value $> \alpha$, we do not reject H_0.

If α were .01, would we reject H_0?

Here, $.0272 \le .10$, so we reject H_0.

Final Concl.

(same as before)

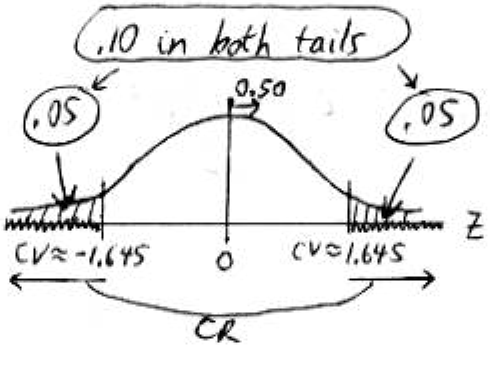
Note: Some articles give P-values and let the reader use his/her own predetermined value of α. (Pfizer vs. Johnson & Johnson?)

Method 3: Traditional / Classical Method

As before, Key sample stat : $\bar{x} = 11.2$
Test stat : $z \approx 2.21$

Find the critical values (CVs) and the critical region (CR).

Based on α, what values of the test stat are "unusual" under H_0?



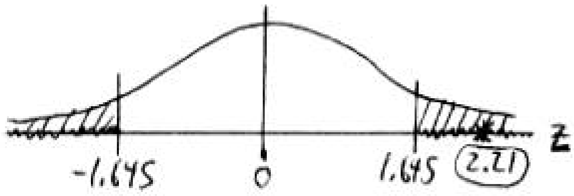
$$CR = \{ \underset{\substack{\uparrow \\ \text{the set} \\ \text{of all}}}{z} \ \underset{\substack{\uparrow \\ \text{such} \\ \text{that}}}{|} \ z < -1.645 \text{ or } z > 1.645 \}$$

<u>Decision Rule</u>

If the test stat [value] falls in the CR,
we reject H_0.

If not, we do not reject H_0.

Here,



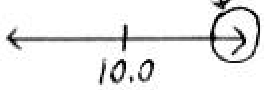
2.21 <u>is</u> in the CR,
so we reject H_0.

<u>Final Concl.</u>
(same as before)

All 3 methods yield same conclusion.

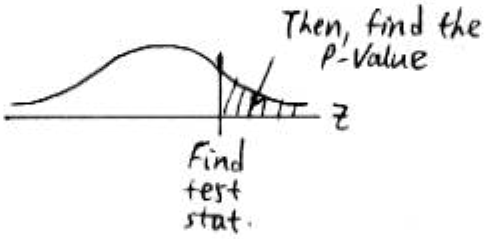
© One-Tailed Tests

Ex $H_0: \mu = 10.0$
$H_A: \mu > 10.0$
Given α

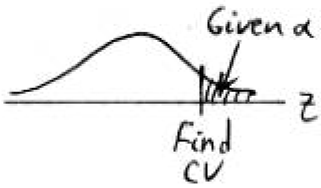


Right-Tailed Test

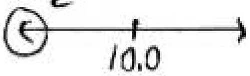
P-Value



Traditional

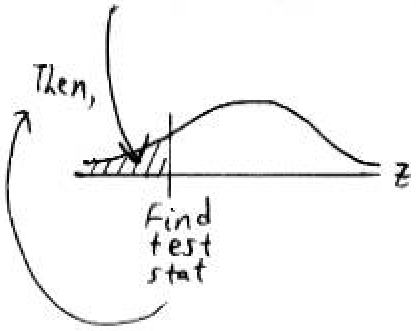


Ex $H_0: \mu = 10.0$
$H_A: \mu < 10.0$
Given α

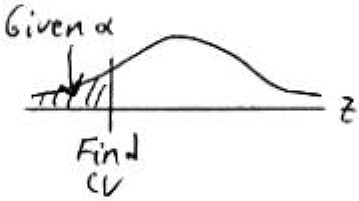


Left-Tailed Test

P-Value



Traditional



7-4 in 1st ed. } TESTS FOR μ
7-5 in 2nd ed. } IF σ IS UNKNOWN

A) Assumptions

(σ is unknown)

CLT applies:

1) $n > 30$, or

2) X roughly Normal

B) Ex

A large class takes a test. Assume the scores are roughly normally distributed. A sample of 7 tests have a mean of 67.5 pts. and a SD of 3.2 pts. Test the claim that the average score for the class is less than 70.0 pts. (A) at the 0.05 significance level.

Sol'n

State H_0, H_A, α. Claim?

Let μ = the average score for the class.

$H_0: \mu = 70.0$ [pts.]
$H_A: \mu < 70.0$ [pts.] (← Claim)
Left-tailed test

$\alpha = 0.05$

Summary

$X \overset{\text{roughly}}{\sim}$ Normal $\Rightarrow$ CLT applies $\Rightarrow \bar{x} \sim$ Normal
$\bar{x} = 67.5$ $\Rightarrow$ Key sample statistic
$s = 3.2$ σ unknown, so use t.
$n = 7$

Under Ho,

$$\bar{x} \sim N(\mu_{\bar{x}} = \mu = 70.0,$$
$$\sigma_{\bar{x}} \approx \frac{s}{\sqrt{n}} = \frac{3.2}{\sqrt{7}} \approx 1.2095)$$

Test statistic

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

$$\approx \frac{67.5 - 70.0}{1.2095}$$

$$\approx -2.07$$

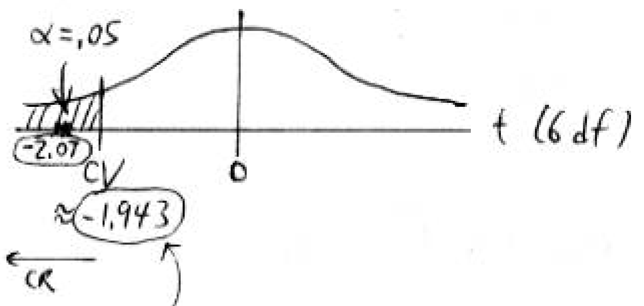
$$n - 1 = 7 - 1 = 6 \text{ df}$$

7.14
7-4 (1st)
7-5 (2nd)

Method 3: Traditional

Find the critical value (CV) and the critical region (CR).

Left-tailed test



What's CV?

See Table A-3 { p. 553 - 1st ed.
p. 568 - 2nd ed. }

df	α .05 (one tail)
⋮	
6	1.943

Left-tailed test ⇒ Use -1.943.

Decision

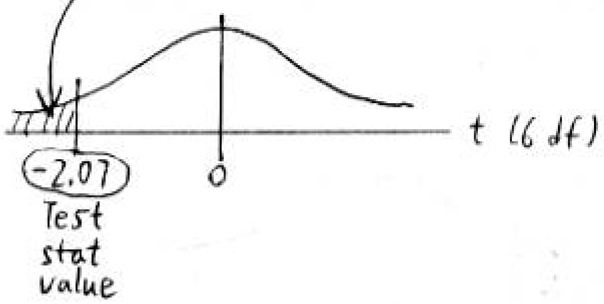
Our test stat value, -2.07, is in the CR, so we reject H_0.

Final Conclusion

There is sufficient evidence in support of the claim that ☆.

Method 2: P-Value

Left-tailed test



What's the P-Value?

Table A-3

Probs. in col. header, not in body, as in z.

df	α = .025	α = .05
	← (one tail)	
⋮		
6	2.447	1.943

-2.07 is between -2.447 and -1.943, so

$$.025 < \text{P-Value} < .05$$

Decision

P-Value $\le \underset{(\alpha)}{.05}$, so we reject H_0.

Final Concl. (same as before)

Method 1: CI

We'll ignore for one-tailed tests.

7-5 in 1st ed. / 7-3 in 2nd ed. } TESTS FOR p

Ex In a poll of 450 likely U.S. voters, 240 intend to vote for Bush. Use a significance level of 0.01 to test the claim that a majority of likely U.S. voters intend to vote for Bush.

Sol'n

Binomial experiment ✓

Let p = the proportion of likely U.S. voters who intend to vote for Bush.

$H_0: p = 0.5 \quad (\Rightarrow q = 0.5)$
$H_A: p > 0.5 \quad$ (Claim)

$\alpha = 0.01$

Summary

$n = 450$
$x = 240$

$$\hat{p} = \frac{x}{n} = \frac{240}{450} \approx 0.533$$

Key sample stat

Can we use the normal approx. to the binomial?

Under H_0, $np = (450)(0.5) = 225 \geq 5$ ✓
$nq = (450)(0.5) = 225 \geq 5$ ✓

Test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Under H_0

Given on tests

1st ed. - p. 374
2nd ed. - pp. 360-1

Why? $z = \frac{x-\mu}{\sigma} = \frac{x - np}{\sqrt{npq}} \begin{matrix} \leftarrow \div n \\ \leftarrow \div n \end{matrix} = \frac{\overbrace{\left(\frac{x}{n}\right)}^{\hat{p}} - p}{\sqrt{\frac{pq}{n}}}$

Note We ignore continuity corrections.

$$\approx \frac{0.533 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{450}}}$$

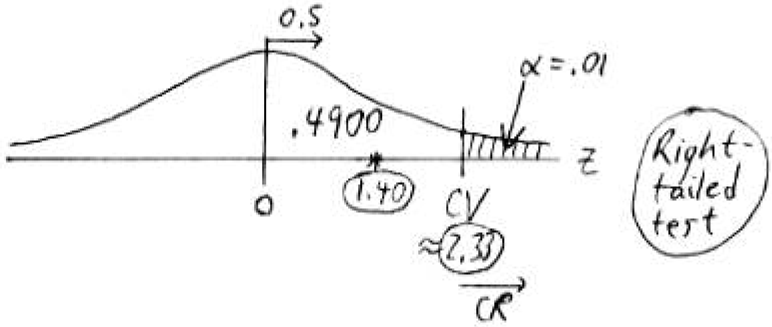
Work this out 1st.

$$\approx \frac{0.033}{0.0236}$$

$$\approx 1.40$$

Method 3: Traditional

Find CVs, CR



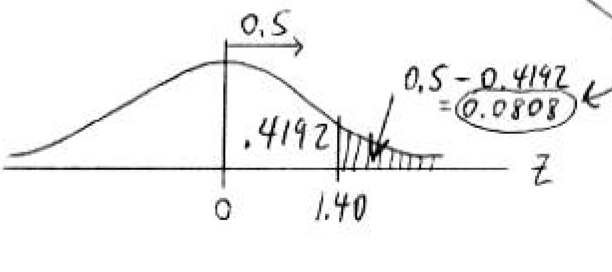
Decision

1.40 is not in the CR, so we do not reject H_0.

Final Concl.

There is insufficient evidence in support of the claim that ☆.

Method 2: P-Value



Decision

$0.0808 > 0.01$, so we do not reject H_0.

Final Concl. (same as before)

7-6 TESTS FOR σ, σ^2

$X \xrightarrow{\text{very close}} \text{Normal}$

We'll only do Method 3 (Traditional).

Ex A women's college goes coed.
We know women's heights have a SD of 2.5 in. at the college.
A sample of 20 men's heights has a SD of 2.8 in.

① Use $\alpha = 0.05$ to test the claim that men's heights and women's heights at the college have different variances. (A)

Let σ^2 = the variance of men's heights at the college.

$H_0: \sigma^2 = \underbrace{(2.5)^2}_{\text{Careful! SD vs. VAR.}} = \underbrace{6.25}_{\text{the variance for women}}$

$H_A: \sigma^2 \neq 6.25$

Two-tailed test

$\alpha = 0.05$

Key sample stat: $s^2 = (2.8)^2 = 7.84$

$n = 20$

Test stat

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$ ← under H_0 (Know this)

Given on tests

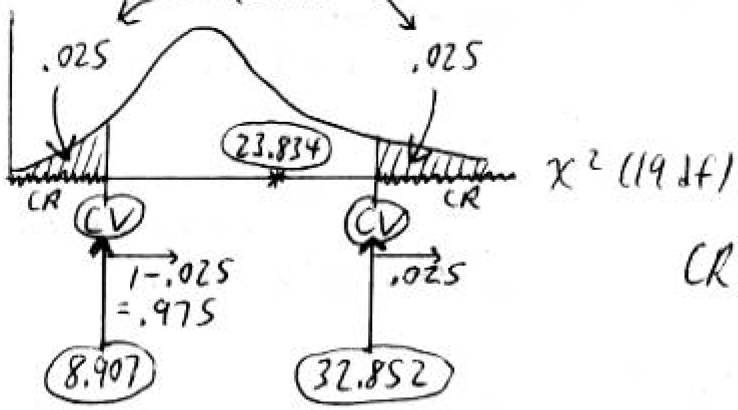
$$= \frac{(20-1)(7.84)}{6.25}$$

$$\approx 23.834$$

$\#df = n-1 = 20-1 = 19\ df$

CVs and CR

(two-tailed)
$\alpha = .05$



CR: "Danger Zone"

Decision

23.834 is not in the CR, so we do not reject H_0.

Final Conclusion (F.C.)

There is insufficient evidence to support the claim that (A).

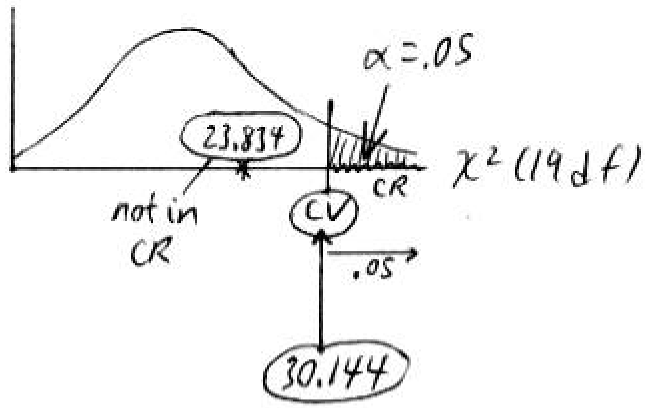
② Use $\alpha = 0.05$ to test the claim that men's heights vary more than women's heights at the college. AA

What's the only diff from ①?

$H_0: \sigma^2 = 6.25$
$H_A: \sigma^2 > 6.25$
$\alpha = 0.05$

As before, $s^2 = 7.84$
$n = 20$
Test $\chi^2 \approx 23.834$, 19 df.

CV and CR



We fail to reject H_0.
F.C.: ....

Note: For a given α, you may reject H_0 in a one-tailed test but not in a two-tailed test!

③ Left-tailed test (What's new?)

Claim
$H_A: \sigma^2 < 6.25$

Do not reject H_0.