Lecture 19 (ch. 8)

We have built lots of CI's. They are good for 2 things:

1) Convey uncertainty [Reliability]
2) aid in making Yes/No decisions.

→ Suppose a 2-sided 99% CI for μ_x is $[1.1, 2.3]$

Someone claims $\mu_x = 0$, $\mu_x = 1.5$, $\mu = 3$

Reject or [Not Reject ≠ Accept]? Reject cannot reject Reject.

→ Suppose a 2-sided 99% CI for $\mu_1 - \mu_2$ is $[1.1, 2.3]$

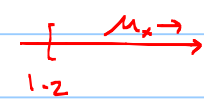
Someone claims $\mu_1 = \mu_2$. Reject

→ Suppose a 2-sided 99% CI for $\mu_1 - \mu_2$ is $[-1.1, 2.3]$

Someone claims $\mu_1 = \mu_2$. Cannot

≠ Accept.

Suppose a 99%, 1 (1-sided) Lower conf. bound is 1.2



Someone claims $\mu_x < 1.5$

Someone claims $\mu_x < 1.1$ Etc. Tricky!

" $\mu_x > 1.5$

" $\mu_x > 1.1$

If Decision-Making (ie. Reject / not-reject) is the final goal of your study, then the machinery of computing C.I. can be massaged to form a more direct response. The revised methodology is called hypothesis testing.

The logic of the methodology is very tricky!

It requires assuming a statement/claim about a pop. parameter, so that we can compute some probabilities.

Then the question one asks is

"Does data provide sufficient evidence contrary to the assumption?"

If yes, then reject the assumption/claim.

If No, then cannot reject the assumption/claim.

i.e., we just don't know!

Notice: cannot reject claim ≠ Accept claim!

One can also ask

"Does the data provide sufficient evidence in support of some claim (ie. the opposite of the assumption)?"

Suppose a sample of size 25 yields $\bar{x}_{obs} = 3$ $s = 1$.
And suppose we want to know how small μ_x can be?

95% lower conf. bound: $3 - 1.711 \frac{1}{\sqrt{25}} = 2.66$ $\quad \underset{2.66}{[} \xrightarrow{\mu_x}$

So, a claim $\mu_x < 1$ can be rejected (with some confidence).

Now, here is a different way of arriving at the last conclusion:

Suppose $\mu_x < 1$: Assumption / Null hypothesis.
Now, let's find evidence to the contrary. Q: what's contrary?
A: Really large $\bar{x}$'s justify rejecting $\mu_x < 1$. So, let's find

$$pr(\bar{x} > \bar{x}_{obs} \mid \text{if } \mu_x < 1).$$

But that prob. already assumes $\mu_x < 1$; so if that prob is small, then that's evidence against $\mu_x < 1$, ie. reject $\mu_x < 1$.

Let's start by computing that prob, if $\mu_x = 1$:

$$\underbrace{prob(\bar{x} > \bar{x}_{obs} \mid \mu_x = 1)}_{\text{one type of "p-value"}} = prob\Big(\underbrace{\frac{\bar{x} - \mu_x}{s/\sqrt{n}}}_{t} > \underbrace{\frac{\bar{x}_{obs} - \mu_x}{s/\sqrt{n}}}_{t_{obs}} \;\Big|\; \mu_x = 1\Big)$$

$$= prob(t > 10 \mid \mu_x = 1) \simeq 0$$

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_x}{s/\sqrt{n}} = \frac{3-1}{1/\sqrt{25}} = 10$$

df = 24, Table VI
(Note: right areas

So, let's think! If we assume μ_x is as big as it can get according to the claim, ie. $\mu_x = 1$, then the prob. of getting $\bar{x}$ larger than $\bar{x}_{obs}$ is nearly zero. That is a lot of evidence contrary to the claim/assumption $\mu_x = 1$
I.e. we can reject the claim $\mu_x = 1$ (See next page)

from data

So, the p-value measures evidence against H_0 in favor of H_1.

And the smaller the p-value is, the more evidence there is.

IMPORTANT:

Blue note

Because prob($\bar{x} > \bar{x}_{obs} | \mu_x = 1$) = small, we can reject $\mu_x < 1$, not just $\mu_x = 1$. } same conclusion as from the CI method!

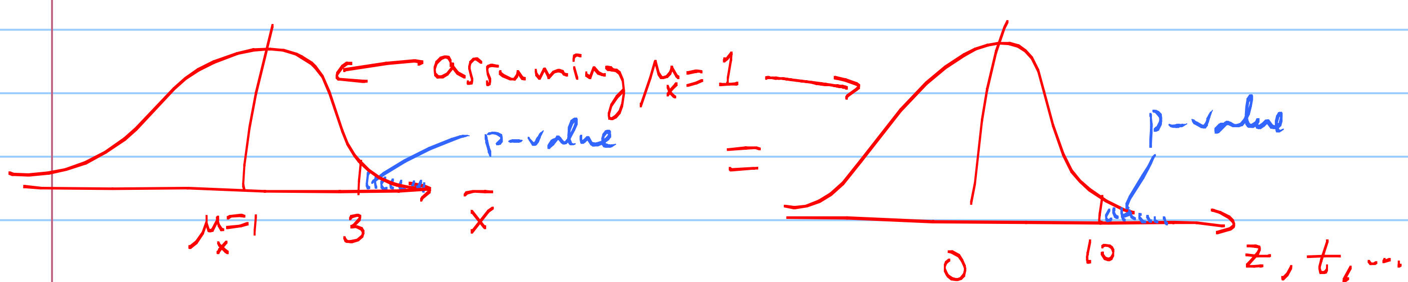
This is because, if $\mu_x < 1$, then t_{obs} is even larger (than 10), which means that the p-value is even smaller.

In short, it's sufficient to test $\mu_x = 1$

FYI

Some skeptical students will want to compute the prob of $\bar{x} > \bar{x}_{obs}$ if $\mu_x = 1$, and add to that the prob. of $\bar{x} > \bar{x}_{obs}$ if $\mu_x = 0.99$, and add to that the prob when $\mu_x = 0.98$, etc. The problem with that logic is that technically $pr(\bar{x} > \bar{x}_{obs} | \mu_x = \text{anything})$ is not well defined, because "$\mu_x = \text{anything}$" is not a random thing. (But see the fig. on the last page below, anyway!)

Pictorially, a p-value is an area:



One element of this process that has not been explicitly addressed is the alternative in whose favor $\mu_x < 1$ is rejected. That alternative hypothesis is, in fact, an important part of the hypothesis testing machinery.

So, let's do things a bit more generally & systematically.

general procedure for hypothesis testing for μ:

Dropping subscript. | In our example

1) Decide the pop. parameter being tested — μ

See Blue Note

sufficient to test $\ominus$

2) Set-up H_0 : $\mu > \mu_0$, $\mu < \mu_0$, $\mu = \mu_0$ — $\mu < 1$ $(\mu_0 = 1)$

null hypothesis. Based on prior (to data) belief.

3) " " H_1 : $\mu < \mu_0$, $\mu > \mu_0$, $\mu \neq \mu_0$ — $\mu > 1$

Alternative hypothesis.

4) Choose appropriate statistic : z, t, ... — t

5) Assume $H_0 = \text{TRUE}$ (ie. set $\mu = \mu_0$) Null param. — $\mu = 1$

6) Compute test statistic for $\underbrace{\text{observed data}}_{\text{Sample}}$ $t_{obs} = \dfrac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$ — 10

7) Find prob of getting a random test statistic more extreme than the observed one,] p-value

e.g. $\text{prob}(\bar{x} > \bar{x}_{obs}) = \text{prob}(t > t_{obs})$ — $\text{prob}(t > 10) \approx 0$

8) Decide if p-value is sufficiently small to reject H_0 in favor of H_1. — ? See below.

2 questions: { more extreme ? / sufficently small ?

!! TYPO on p.360 in book, 3rd and 4th line : $H_0 \to H_a$ / H_1 !!

Q More extreme?

A Depends on (H_0, H_1):

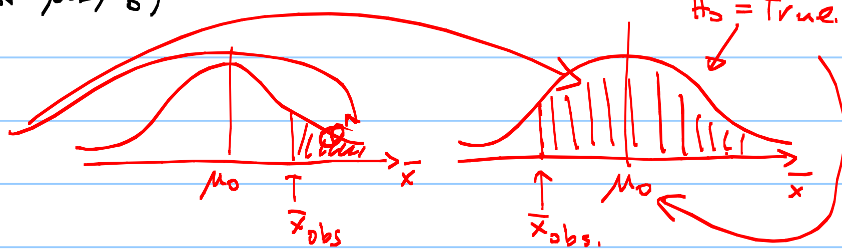
Because of the blue note, above, it is sufficient to test $\mu=\mu_0$

1-sided

If $H_0: \mu \le \mu_0$ (or $\mu=\mu_0$)

$H_1: \mu > \mu_0$

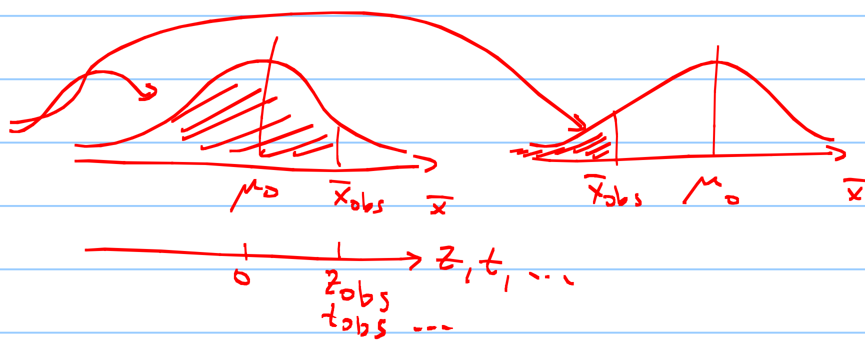
Then p-value =



If $H_0: \mu \ge \mu_0$ (or $\mu=\mu_0$)

$H_1: \mu < \mu_0$

Then p-value =

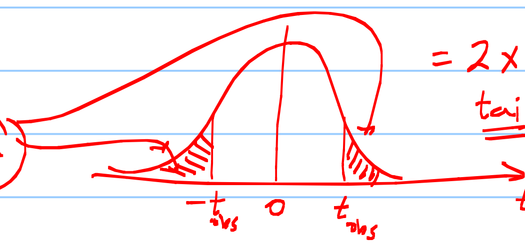


2-sided

If $H_0: \mu = \mu_0$

$H_1: \mu \ne \mu_0$

Then p-value = Sum



= 2 × 1-sided tail area.

mnemonic

In short: $H_1: \mu > \mu_0$: p-value = right-area

$H_1: \mu < \mu_0$: = left-area

$H_1: \mu \ne \mu_0$: = left + right = 2 × (1 tail)

Q = Who decides what's "sufficiently small"?

A = You do! (or The book does)

This "threshold probability" is labeled α.
The same α That showed-up in C.I.

It's called significance level. (= 1 − Conf. level)
.05 sign. level = 95% Conf. level.

Some common values are .05, .01, .001, but The choice depends on The cost of making the wrong decision (ie. of rejecting H_0 when it's True).

In short, to make a decision:

1) You choose The value of α.

2) Compute p-value from The above procedure.

3) If p-value $< \alpha$, Then Reject H_0 in favor of H_1.
Else, cannot reject " " " " "

Example: Data says: $n = 64$, $\bar{x}_{obs} = 34.4$, $S = 1.1$.

Does the data provide evidence to support $\mu > 34$?

1) The param. of interest : μ (dropping subscript x).

2,3) H_0 : $\mu < 34$ (or $\mu = 34$) $\mu_0 = 34$

H_1 : $\mu > 34$

Setting up H_0, H_1 is the hardest part of these problems.
The next page offers 2 way of deciding.

4) Appropriate test statistic : z, t

z is appropriate, because n = large. But I'll use t for illustration.

5) Assume $H_0 = T$. (I.e. set $\mu = 34$)

6) Compute statistic assuming $H_0 = $ True (ie. $\mu = \mu_0$) $t_{obs} = \frac{34.4 - 34}{1.1/\sqrt{64}} = 2.91$

See prev. page.

7) p-value = prob$(\bar{x} > \bar{x}_{obs})$ = prob. of getting an $\bar{x}$ as large as (or larger) than the obs. $\bar{x}$.

= prob$(t > 2.91) \simeq .0025$ df $= n-1 = 63$

1 − pt(2.91, 63) in R.

Conclusion: At $\alpha = .05$, p-value $< \alpha$.

Therefore,

Data provides sufficient evidence to reject H_0 ($\mu<34$) in favor of H_1 ($\mu>34$).

"In English": Data provides suff. evidence in favor of $\mu > 34$.

At $\alpha = 0.001$, p-value $> \alpha$

Therefore, we cannot reject H_0 ($\mu<34$) in favor of H_1 ($\mu>34$).

"In English": There is no support for $\mu > 34$

Note: This conclusion is NOT the same thing as "There is support for $\mu < 34$." All we can say is that we cannot reject $\mu < 34$.

ways I go about deciding what H_0, H_1 should be:

1) The question asks "does data provide evidence." Meanwhile, the hyp. testing procedure begins by assuming H_0 is True. So, it makes no sense to assume the claim is true even before data!

→ Don't assume what the data is supposed to test.

2) The data provides evidence for H_1 (against H_0), because of the way the whole procedure is set-up. Recall that the procedure requries assuming H_0 = True. Then, if the evidence is weak (e.g. when there is no data at all), then the procedure leaves you with H_0. So ask yourself this: what conclusion should the procedure yield if the evidence is really really weak, e.g. No data at all? The answer is your H_0. In this example, if there is no data, then we should stay with $\mu < 34$. That tells us $H_0: \mu < 34$.

→ Ask yourself what statement you should be left with if there is no data at all! The answer to that question tells you what H_0 should be. Then, the "opposite" of H_0 is H_1.

3) Another way of deciding on H_0, H_1 will be discussed later, when we learn the meaning of α.

The hardest part of hyp. testing is setting up H_0, H_1.
In doing so, keep the following in mind:

The whole procedure is set-up so that

- H_0 is assumed to be true.
- The data provide evidence for H_1 (against H_0).

So, you should not assume what you are trying to see if the data is supporting. Otherwise, you are assuming what you want to test.

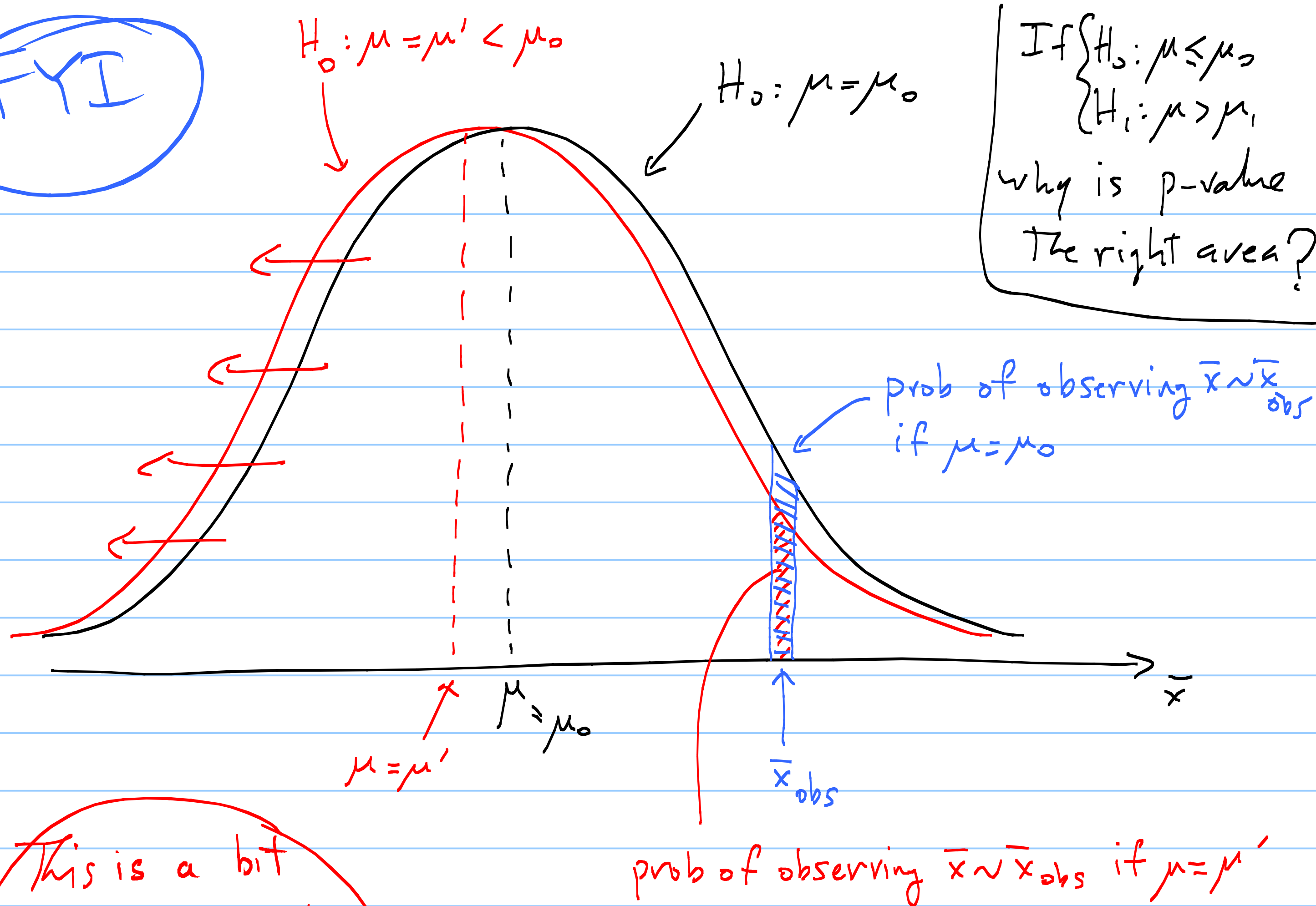
- H_0 and H_1 are statements about some/any pop. param.
- Reject H_0 in favor of H_1, if data provide sufficient evidence against (the assumed true) H_0, in favor of H_1.
- p-value is the quantity that represents the evidence provided by the data, in favor of H_1.
- But note that smaller p-value means more evidence.
- Some problems ask you to test some prior belief (ie. some claim based on something other than data). Then, that belief should be H_0.

→ If you cannot reject H_0 in favor of H_1, then we don't know anything! Not rejecting H_0 is not the same thing as accepting it. Making the mistake of interpreting the lack of evidence for H_1 as support for H_0 is the source of many contradictory findings in the literature.

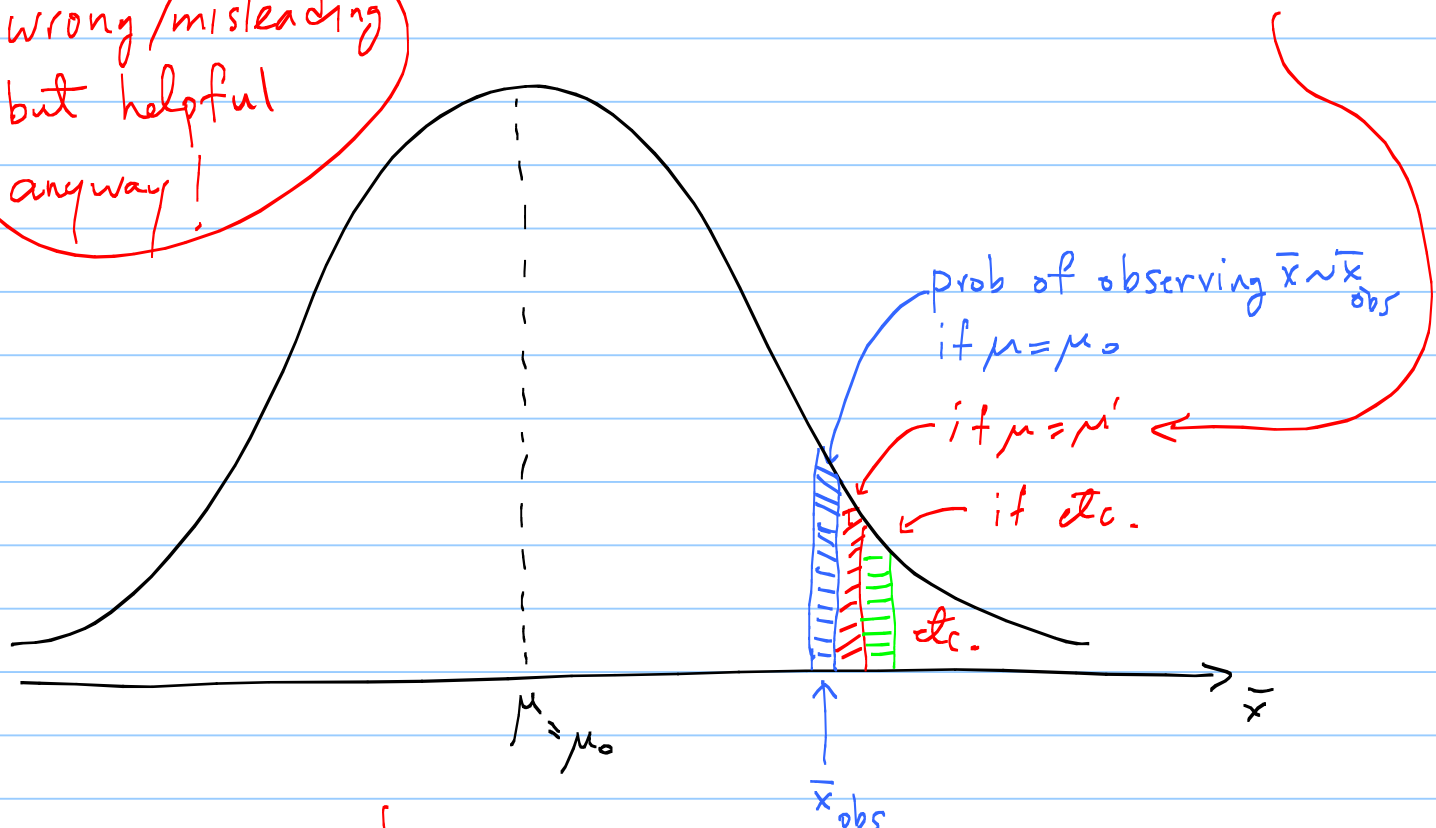
→ In general, we cannot accept a claim about an unknown pop. parameter (e.g. $H_0: \mu \leq 1$). All we can do is either reject it, or not, based on evidence from data (through t_{obs}, or p-value). The mathematical way to see this is to note that the p-value is a "conditional prob", ie. it assumes the claim H_0 is true.

FYI

If $\begin{cases} H_0: \mu \le \mu_0 \\ H_1: \mu > \mu_1 \end{cases}$ why is p-value the right area?



This is a bit wrong/misleading but helpful anyway!



So, if $H_0: \mu \le \mu_0$, then the p-value $= \text{prob}(\bar{x} > \bar{x}_{obs})$

I.e. $H_1: \mu > \mu_0$ mnemonic $=$ right area hw-lect 19-1

Consider the following sample observations: 2781, 2900, 3013, 2856, and 2888.
Suppose we want to test whether there is evidence that mu < 3000.
a) Write the appropriate hypotheses, compute the p-value, and state the conclusion "In English" (i.e., is there evidence that mu is less than 3000?) using alpha = 0.05.
b) Compute the appropriate confidence interval (CI). Is the conclusion the same as in part a? Explain.

One can also arrive at the same conclusion, without the p-value and CI, by what is called the rejection method. I'll walk you through it:
c) If H0 is true, compute the value of x_bar that has an area of alpha=0.05 to the left. This value of x_bar is called the critical value, and the region to its left is called the rejection region. So, in this part of the problem you are computing the rejection region.
d) Is the observed value of x_bar in the rejection region? If so, one can reject H0 in favor of H1; otherwise, one cannot say anything.