Math 4501 - Probability and Statistics II

7.6 - Confidence intervals for regression parameters (related with Sec. 7.1)

based mostly on finding confidence intervals for the mean.

Notes:

- The $100(1-\gamma)\%$ prediction interval for Y at $X = x$ is wider than the $100(1-\gamma)\%$ confidence interval for $\mu(x)$:
- The difference between one observation of Y (at a given value of x) and its predictor tends to vary more than the difference between the mean of the entire population of Y values (at the same x) and its estimator $\hat{\mu}(x)$.

$S_Y = \sqrt{\cdots}\sqrt{1+\frac{\sigma}{n}^2 + \cdots}$

$S_\mu = \sqrt{\vdots}\sqrt{\frac{\sigma}{n}^2 + \cdots}$

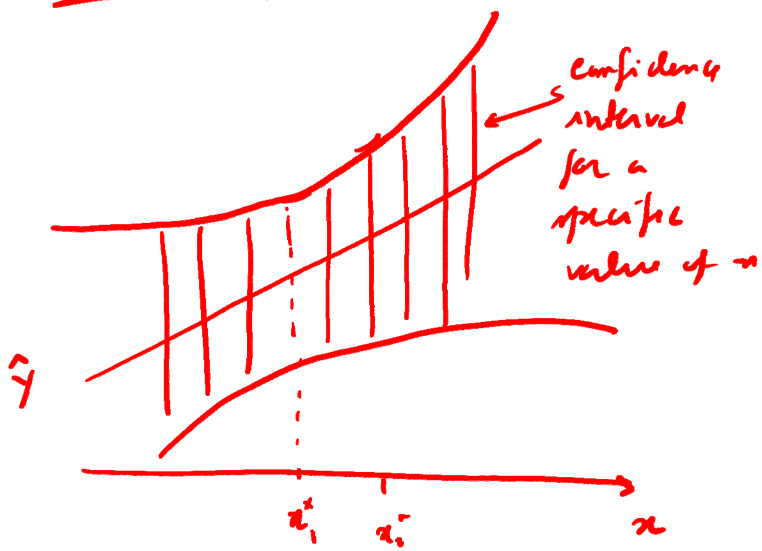
Definition

1) The collection of all $100(1-\gamma)\%$ confidence intervals $I_\mu(x)$ for $\mu(x)$, $\{I_\mu(x) : x \in \mathbb{R}\}$, is called a pointwise $100(1-\gamma)\%$ *confidence band* for $\mu(x)$.
2) The collection of all $100(1-\gamma)\%$ prediction intervals $I_Y(x)$ for Y at $X = x$, $\{I_Y(x) : x \in \mathbb{R}\}$, is called a pointwise $100(1-\gamma)\%$ *prediction band for* Y.

Note: From the expressions for S_μ and S_Y in the confidence and prediction intervals, respectively, these bands are narrowest at $x = \bar{x}$.

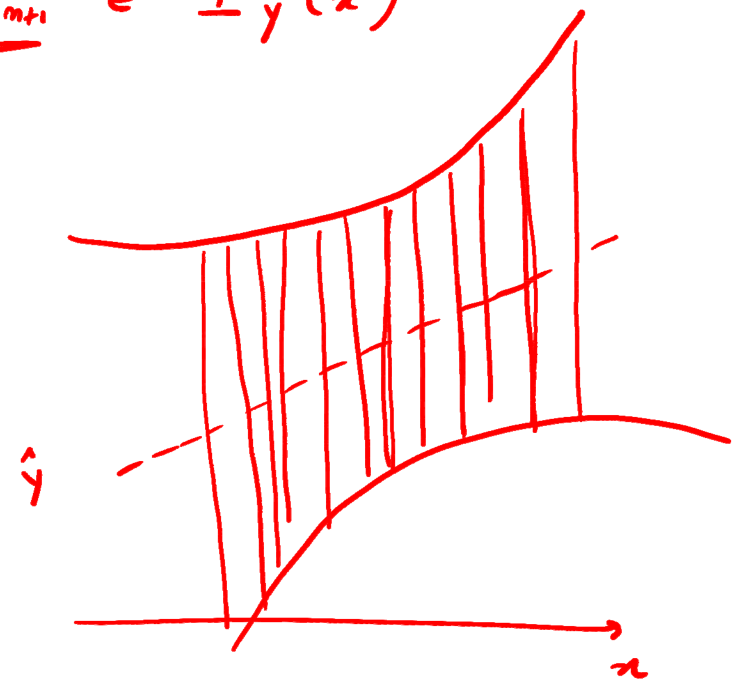
VIZUALIZATION

$\underline{\mu(x)} \in I_{\mu}(x)$ ← confidence band



prediction band

$\underline{Y_{n+1}} \in I_{y}(x)$



Example (previous example from Chp. 6 continued)

The table contains ten pairs of test scores of ten students in a certain class, with x being the score on a preliminary test and y the score on the final examination,

$n=10$

x	y	x^2	xy	y^2	$\widehat{y}$	$y-\widehat{y}$	$(y-\widehat{y})^2$
70	77	4,900	5,390	5,929	82.561566	–5.561566	30.931016
74	94	5,476	6,956	8,836	85.529956	8.470044	71.741645
72	88	5,184	6,336	7,744	84.045761	3.954239	15.636006
68	80	4,624	5,440	6,400	81.077371	–1.077371	1.160728
58	71	3,364	4,118	5,041	73.656395	–2.656395	7.056434
54	76	2,916	4,104	5,776	70.688004	5.311996	28.217302
82	88	6,724	7,216	7,744	91.466737	–3.466737	12.018265
64	80	4,096	5,120	6,400	78.108980	1.891020	3.575957
80	90	6,400	7,200	8,100	89.982542	0.017458	0.000305
61	69	3,721	4,209	4,761	75.882687	–6.882687	47.371380
$\sum x_i$: 683	$\sum y_i$: 813	47,405	56,089	66,731		0.000001	217.709038

Example

Find:

$\alpha = 0.05$

1. a 95% confidence interval for $\mu(x) = \alpha + \beta(x - \bar{x})$ for $x = 60$ and $x = 70$;
2. a 95% prediction interval for Y when $x = 60$.

Using formulas for regression line (Chp. 6)

We have seen previously that

$$\bar{x} = 68.3, \quad \widehat{\alpha} = 81.3, \quad \text{and} \quad \widehat{\beta} = 0.7421 ,$$

obtaining the regression line

$$\widehat{y} = 81.3 + (0.742)(x - 68.3) = 30.6 + 0.742x .$$

We have also found that $\widehat{\sigma^2} = 21.7709$. Moreover, we also need

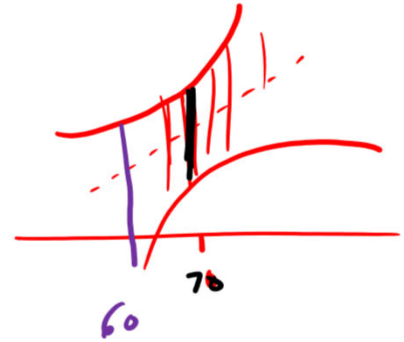
From Chp 6

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} x_i\right)^2$$

$$= 47405 - \frac{683^2}{10} = 756.1 .$$

Recall that a a $100(1-\gamma)\%$ confidence interval for $\mu(x) = \alpha + \beta(x - \bar{x})$ is given by

$$\widehat{\alpha} + \widehat{\beta}(x - \bar{x}) \pm t_0 S_\mu \ ,$$



where $t_0 = t_{\gamma/2}(n-2)$ and

$$S_\mu = \sqrt{\frac{n\widehat{\sigma^2}}{n-2}}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \ .$$

For a 95% confidence, we find $t_{0.025}(8) = 2.306$. When $x = 60$, the endpoints for a 95% confidence interval for $\mu(60)$ are then

$x=60$ $x=60$

$$81.3 + 0.7421(60 - 68.3) \pm \left[\sqrt{\frac{10(21.7709)}{8}}\sqrt{\frac{1}{10} + \frac{(60-68.3)^2}{756.1}}\right](2.306)$$

or

$$75.1406 \pm 5.2589$$ $\mu(60)$

When $x = 70$, the endpoints for a 95% confidence interval for $\mu(70)$ are

$$82.5616 \pm 3.8761$$

Recall that a a $100(1-\gamma)\%$ prediction interval for Y is given by

$$\widehat{\alpha}+\widehat{\beta}(x-\bar{x}) \pm t_0 S_Y,$$

where $t_0 = t_{\gamma/2}(n-2)$ and

$$S_Y = \sqrt{\frac{n\widehat{\sigma^2}}{n-2}}\sqrt{1+\frac{1}{n}+\boxed{\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}.$$

For the same $\underline{\underline{x}}$

$S_y > S_\mu$

due to the $+1$ under 2nd square root

For a 95% confidence, we find $t_{0.025}(8) = 2.306$. When $x = 60$, the endpoints for a 95% confidence interval for Y are then

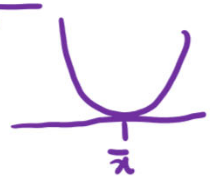
$$81.3+0.7421(60-68.3) \pm \left[\sqrt{\frac{10(21.7709)}{8}}\sqrt{1+\frac{1}{10}+\frac{(60-68.3)^2}{756.1}}\right](2.306)$$

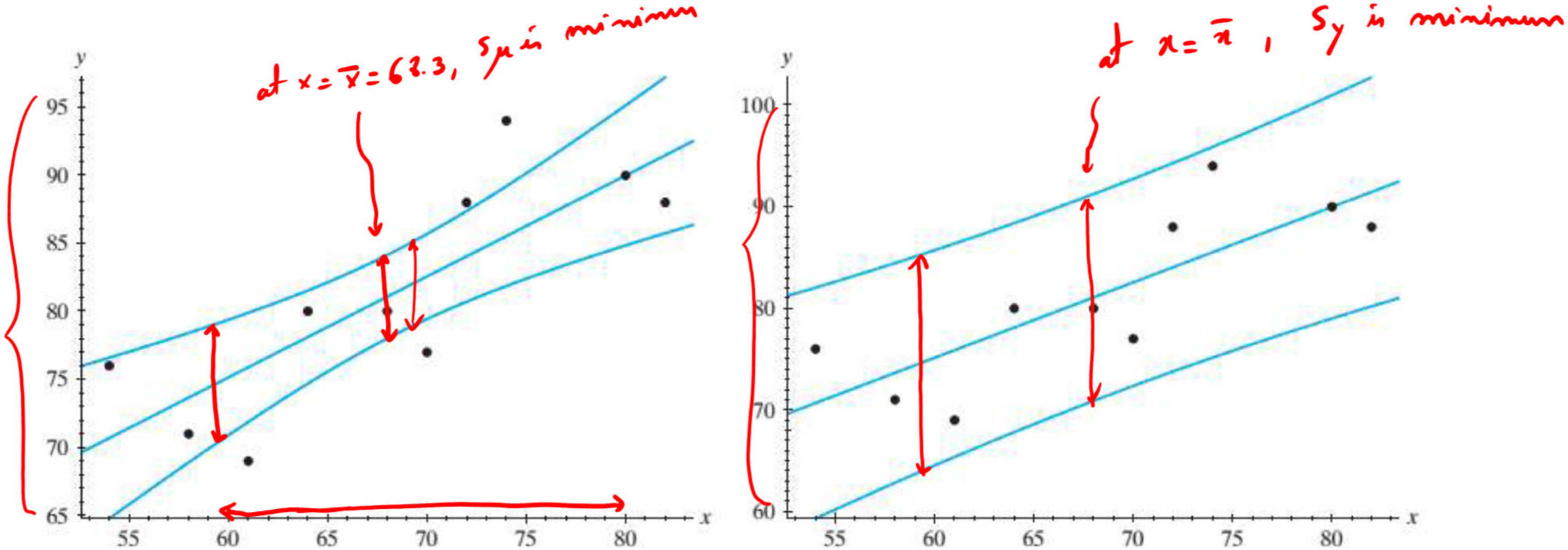
or

$$75.1406 \pm \underline{13.1289},$$

a wider interval than the confidence interval for obtained for $\mu(60)$.

$\dfrac{(x-\bar{x})^2}{\Sigma}$





Left: pointwise 95% confidence band for $\mu(x)$;
Right: pointwise 95% prediction band for Y

using data from this example

NOTE: The bands are narrower when $x = \bar{x}$, and as x gets further away from $\bar{x}$, the bands become wider

due to term $\dfrac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ in both s_μ and s_y

Math 4501 - Probability and Statistics II

8.1 - Hypothesis tests: one mean

Overview

A *statistical hypothesis* is a statement about the parameters describing a population or distribution (not a sample).

A *statistical hypothesis test* is a method of statistical inference under which two hypothesis – null and alternative – are compared:

- procedure whose inputs are samples and whose result is a statement regarding a hypothesis.

We will:

- introduce key notions regarding statistical hypothesis tests.
- study hypothesis tests for one mean.

Simple and composite hypothesis

A *simple hypothesis* is any hypothesis which specifies the population distribution completely.

A *composite hypothesis* is any hypothesis which does not specify the population distribution completely.

Example

Assume that X is $N(\mu, 36)$ and let μ_0 be a given constant.

A statement of the form

$$\mu = \mu_0$$

is a simple hypothesis.

All the following statements

$$\mu \neq \mu_0 \qquad \mu < \mu_0 \qquad \mu > \mu_0$$

are a composite hypotheses.

Null and alternative hypothesis

A *null hypothesis*, usually denoted H_0, is a hypothesis associated with a contradiction to a theory one would like to prove.

H_0 is often simple

An *alternative hypothesis*, usually denoted H_1, is a hypothesis (often composite) associated with a theory one would like to prove.

Example

Assume that X is $N(\mu, 36)$.

A test of a null hypothesis

$$H_0 : \mu = 50$$

against the alternative hypothesis

$$H_1 : \mu > 50$$

aims to access whether the mean of the distribution of X is larger than 50.

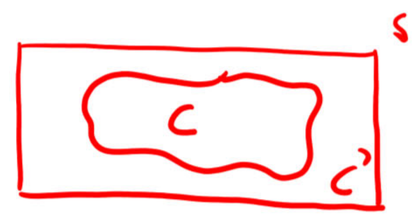
Critical region

To test a null hypothesis H_0 against an alternative hypothesis H_1, we set up a rule which leads to one of two decisions:

- reject H_0
- fail to reject H_0

Procedure: partition the sample space into two parts, denoted C and C', so that if a random sample yields the observations $(x_1, x_2, \ldots, x_n)$, then:

- if $(x_1, x_2, \ldots, x_n) \in C$, H_0 is rejected
- if $(x_1, x_2, \ldots, x_n) \in C'$, H_0 not rejected.



The rejection region C for H_0 is called the *critical region* for the test.

The partitioning of the sample space is often specified in terms of the value of a *test statistic*.

Once we determine the critical region C, we can apply test

procedure becomes fully specified

Example

Assume that X is $N(\mu, 36)$.

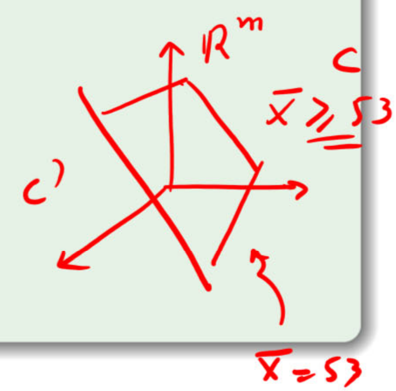
We could test the hypotheses

related with point estimate
for the unknown parameter

$$H_0 : \mu = 50$$
$$H_1 : \mu > 50$$

using the test statistic $\bar{X}$ to specify the critical region of the test as, for instance

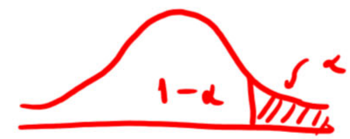
$$C = \{(x_1, x_2, \ldots, x_n) : \bar{x} \geq 53\} \ .$$



With this choice:

- we reject H_0 if $\bar{x} \geq 53$
- we fail to reject H_0 if $\bar{x} < 53$

Type I error and significance level



Type I error:

- to reject H_0 when H_0 is true
- occurs if $(x_1, x_2, \ldots, x_n) \in C$ even if H_0 is true

The *significance level* of a statistical test is defined as:

$$
\begin{aligned}
\alpha &= P(\text{Type I error}) \\
&= P(\text{reject } H_0 | H_0 \text{ true}) \\
&= P(\underbrace{(X_1, \ldots, X_n) \in C}_{} | \underbrace{H_0 \text{ true}}_{}) \,,
\end{aligned}
$$

def of α (significance)

where C denotes the test critical region.

Usual procedure: Fix significance level α and use it to determine C

Type II error and power of a test

Type II error:

- to not reject H_0 when H_1 is true
- occurs if $(x_1, x_2, \ldots, x_n) \notin C$ even if H_1 is true

The probability of occurrence of an error of type II is denoted as β:

$$\left.\begin{aligned} \beta &= P(\text{Type II error}) \\ &= P(\text{do not reject } H_0 | H_1 \text{ true}) \\ &= P((X_1, \ldots, X_n) \notin C | H_1 \text{ true}) \,, \end{aligned}\right\} \text{def of } \beta.$$

where C denotes the test critical region.

The *power of a test* is the quantity $1 - \beta$, the probability of correctly rejecting the null hypothesis H_0 when the alternative hypothesis H_1 is true.

Ideally: we want $1-\beta$ to be as large as possible with $\underbrace{\alpha \text{ fixed}}_{\text{small}}$.

Example

$X_1, \dots, X_n \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Assume that X is $N(\mu, 36)$.

A sample of size $n = 16$ was taken to test the hypotheses

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu = 55 \end{cases}$$

using the critical region

$$\boxed{C = \{(x_1, x_2, \dots, x_n) : \bar{x} \geq 53\} .}$$

$\bar{X} \sim N\left(50, \frac{36}{16}\right)$
$\Uparrow$
$X \sim N(50, 36)$
$\Uparrow$
$\mu = 50$

Find the probabilities of type I and II errors, α and β.

Note that:

- $\bar{X}$ is $N(50, 36/16)$ when H_0 is true
- $\bar{X}$ is $N(55, 36/16)$ when H_1 is true.

→ α = P(Error type I) = P(Rej H_0 | H_0 true)

→ β = P(Error type II) = P(Not rej H_0 | H_1 true)

$\bar{X} \sim N\left(55, \frac{36}{16}\right) \Leftarrow X \sim N(55, 36) \Leftarrow \mu = 55$

Thus,

significance

$$
\begin{aligned}
\alpha &= P(\text{Type I error}) \\
&= P(\text{reject } H_0 \mid H_0 \text{ true}) \\
&= P((X_1, \ldots, X_n) \in C \mid H_0 \text{ true}) \\
&= P\left(\bar{X} \geq 53 \mid \mu = 50\right) \\
&= P\left(\underbrace{\frac{\bar{X} - 50}{6/4}}_{Z} \geq \underbrace{\frac{53 - 50}{6/4}}_{2}\right) = 1 - \Phi(2) = 0.0228
\end{aligned}
$$

standardization

$\bar{X} \sim N\left(50, \frac{36}{16}\right) \Rightarrow Z = \dfrac{\bar{X} - 50}{6/4} \sim N(0,1)$

$P(Z \geq 2) = 1 - P(Z < 2)$

and

$$
\begin{aligned}
\beta &= P(\text{Type II error}) \\
&= P(\text{do not reject } H_0 \mid H_1 \text{ true}) \\
&= P((X_1, \ldots, X_n) \notin C \mid H_1 \text{ true}) \\
&= P\left(\bar{X} < 53 \mid \mu = 55\right) \\
&= P\left(\underbrace{\frac{\bar{X} - 55}{6/4}}_{Z} < \underbrace{\frac{53 - 55}{6/4}}_{-4/3}\right) = \Phi\left(-\frac{4}{3}\right) = 1 - 0.9087 = 0.0913
\end{aligned}
$$

standardization

$\bar{X} \sim N\left(55, \frac{36}{16}\right) \Rightarrow Z = \dfrac{\bar{X} - 55}{6/4} \sim N(0,1)$

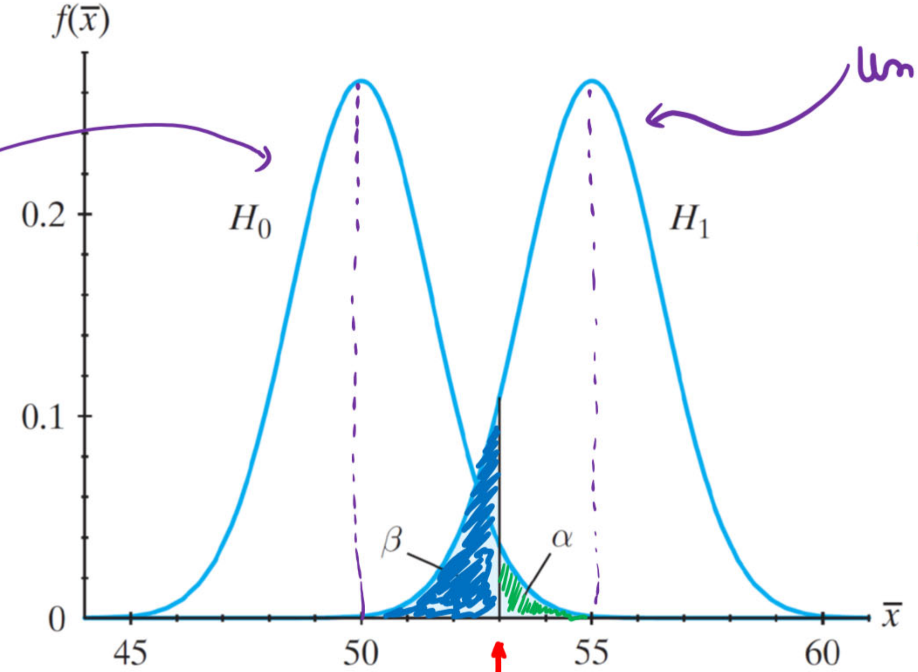
$1 - \phi(4/3)$

$P(Z < -4/3)$

INTERPRETATION / VIZUALIZATION

under H_0: $\bar{X} \sim N\left(50, \frac{36}{16}\right)$

$\alpha = P(\bar{x} \geq 53 \mid H_0 \text{ true})$



$\bar{x} < 53$ 53 $\bar{x} \geq 53$

under H_1: $\bar{X} \sim N\left(55, \frac{36}{16}\right)$

$\beta = P(\bar{x} < 53 \mid H_1 \text{ true})$

p-value

The p-value is the probability, assuming the null hypothesis H_0 is true, of observing a result at least as extreme as the test statistic.

- tail-end probability, under H_0, of the distribution of the statistic beyond the observed value of the statistic.
- we reject H_0 at significance level α if $p < \alpha$.

$p = P$ (more extreme values than those observed | H_0 true)

Procedure   Rej H_0 with significance α if $p < \alpha$

Alternative to specifying the critical region:

Example

Assume that X is $N(\mu, 100)$.

We will test the hypotheses

$$H_0 : \mu = 60 \quad \leftarrow \text{ simple null hypothesis}$$
$$H_1 : \mu > 60 \quad \leftarrow \text{ composite one-sided alternative hypothesis}$$

using a sample mean $\bar{X}$ based on $n = 52$ observations.

Suppose that we obtain the observed sample mean of $\bar{x} = 62.75$.

Compute the p-value, and decide whether the null hypothesis should be rejected at the significance level $\alpha = 0.05$.

Decision rule:

Reject H_0 if $p < \alpha = 0.05$

All we have to do is evaluate $p = P\Big(\underbrace{\bar{X} > \underbrace{62.75}_{\text{observed value}}}_{\bar{X}\text{ more extreme than observed value}} \,\Big|\, \overbrace{H_0 \text{ true}}^{\mu=60}\Big)$

Note that $\bar{X}$ is $N(60, 100/52)$ when H_0 is true.

$\mu = 60 \Rightarrow X \sim N(60, 100)$

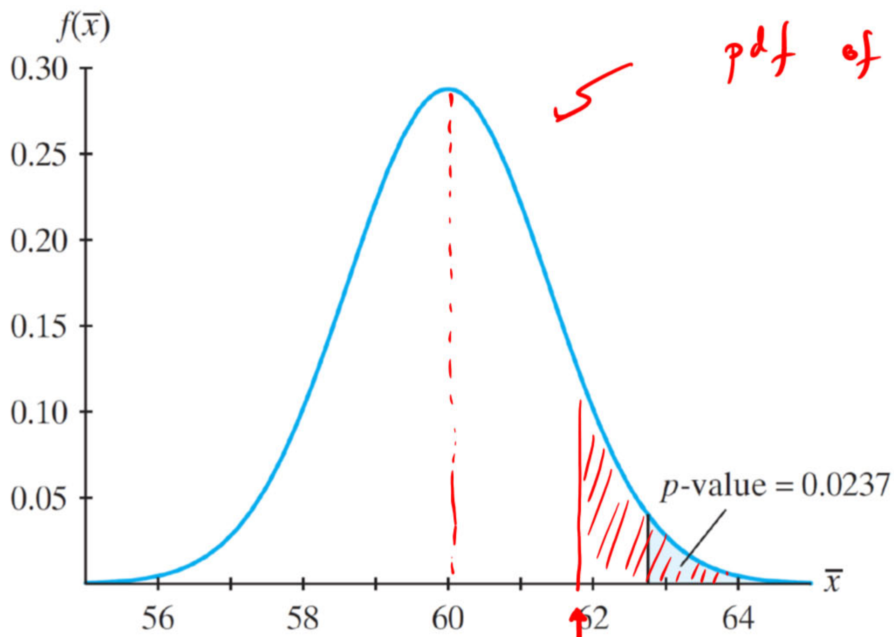
$\Rightarrow \bar{X} \sim N\left(60, \frac{100}{52}\right)$

$\Rightarrow Z = \frac{\bar{X} - 60}{10/\sqrt{52}} \sim N(0,1)$

We find that

$$
\begin{aligned}
p &= P(\bar{X} \geq 62.75 | H_0 \text{ true}) \\
&= P(\bar{X} \geq 62.75 | \mu = 60) \\
&= P\left(\underbrace{\frac{\bar{X} - 60}{10/\sqrt{52}}}_{Z} \geq \frac{62.75 - 60}{10/\sqrt{52}}\right) \\
&= 1 - \Phi\left(\frac{62.75 - 60}{10/\sqrt{52}}\right) = 1 - \Phi(1.983) = 0.0237
\end{aligned}
$$

Since $p = 0.0237 < 0.05 = \alpha$, we reject H_0 at the significance level $\alpha = 0.05$.



pdf of $\bar{X}$ under $H_0: \mu = 60$

$\hookrightarrow \bar{X} \sim N\left(60, \frac{100}{52}\right)$

since p-value $< \alpha$
$\Rightarrow$ observed value is more extreme than the value specifying the critical region with significance α

If, instead, the alternative hypothesis were the two-sided $H_1 : \mu \neq 60$, then the p-value would be double 0.0237, because we would need to include both tails.

see next slide $\rightsquigarrow$ $2\cdot(0.0237) < 0.05 \Rightarrow$ still rej H_0!!

Explanation for bottom comment on previous slide

Under same setup, if we were testing $\boxed{H_0 : \mu = 60}$ vs $H_1 : \mu \neq 60$

composite and two-sided alternative hypothesis!

In this case, the p value (given a sample yielding $\bar{x} = 62.75$) would be:

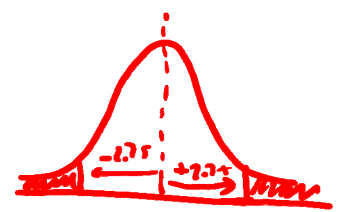
$$p = P(\text{observing a more extreme value than } 62.75 \mid H_0 \text{ true})$$

$$= P(\bar{X} > \underbrace{60 + 2.75}_{62.75} \text{ or } \bar{X} < 60 - 2.75 \mid H_0 \text{ true})$$

$$= P(|\bar{X} - 60| > 2.75 \mid H_0 \text{ true})$$

$$= \text{Area of right tail} + \text{Area of left tail}$$

$$= 2 \text{ Area of right tail} = 2\,(0.0237)$$



$\underbrace{\mu = 60}_{\text{under } H_0}$