Hypothesis Testing

Chapter-Opening Example 1

- The dean of Business School claims that the average salary of the school's graduates one year after graduation is \$800 per week with a standard deviation of \$100.
- A second-year student would like to check whether the claim about the mean is correct.
  - She conducts a survey of 25 people who graduated one year ago and determines their weekly salary.
  - She discovers the sample mean to be \$750.
- Is this proof in favor or against the dean's claim?

Example 1

- To interpret her finding she needs to calculate the probability that a sample of 25 graduates would have a mean of \$750 or less when the population mean is \$800 and the standard deviation is \$100.

Example 1

- We want to compute:

$$P(\overline{X} < 750)$$

- Assume the dean is correct, what does the Central Limit Theorem tell us about the distribution of $\bar{X}$?
- What if we know that the salary of one-year-graduates is normally distributed?

Chapter-Opening Example 1

$$P(\overline{X} < 750) = P\left(\frac{\overline{X} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} < \frac{750 - 800}{20}\right)$$

$$= P(Z < -2.5) = .0062$$

- Conclusion:
  - The probability of observing a sample mean as low as \$750 when the population mean is \$800 is extremely small
  - Because the event is **quite unlikely**, we would conclude that the dean's claim is not justified
  - Hence, we **reject** the Dean's statement (or Hypothesis)

Example 2: Hypothesis Testing

- **Statement:** 50% of citizens eligible for jury duty in the South between 1960 and 1980 were African American.
- On an 80-person panel of possible jurors, only 4 were African American.
- Could this just be the luck of the draw?
  A result of pure chance?
  Or is the original statement wrong?

Example 2: Hypothesis Testing

- Assume juror selection was random!
- Using CLT, we know that the sample proportion $\hat{p}$ of African American jurors on a panel follows:

$$\hat{p} \sim N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

$$\hat{p} \sim N\left(0.5, \sqrt{\frac{0.5(1-0.5)}{80}}\right)$$

= N(0.5, 0.056)

Example 2: Hypothesis Testing

- What is the probability of drawing a random sample of size 80 in which the sample proportion of African-Americans is smaller than 4/80?

- P($\hat{p}$ <4/80) = **norm.s.dist**((4/80-0.5)/0.056, 1)
  = 0.000000000000000465
- Since this is such a small number, it is evidence **AGAINST** our assumption, that 50% of jurors are black.
- We **reject** the original statement, or hypothesis.

General ideas about hypothesis testing

Hypothesis Testing

- A criminal trial is an example of a hypothesis test.
- In a trial a jury must decide between two hypotheses.
- The **null hypothesis** (common belief):

  H0: The defendant is innocent

- The **alternative hypothesis** (research hypothesis) is:

  H1: The defendant is guilty

- The jury does not know which hypothesis is true.
- They must make a decision on the basis of evidence (i.e. data) presented.

Hypothesis Testing

- In the language of statistics convicting the defendant is called

  rejecting the null hypothesis in favor of the alternative hypothesis.
- That is, the jury is saying that there is enough evidence to conclude that the defendant is guilty
- i.e., there is enough evidence to support the alternative hypothesis.

Hypothesis Testing

- If the jury acquits, then it is stating that

  there is not enough evidence to support the alternative hypothesis.
- Notice that the jury is **not** saying that the defendant is innocent, only that there is not enough evidence to support the alternative hypothesis.
- That is why we **never say** that we accept the null hypothesis, but rather we fail to reject the null.

Steps in Hypothesis Testing

The critical concepts are:

1. There are two hypotheses, the null and the alternative hypotheses.
2. The procedure begins with the assumption that the null hypothesis is true.
3. The goal is to determine whether there is enough evidence to infer that the alternative hypothesis is true.

Steps in Hypothesis Testing

5. There are two possible decisions:
    1. Conclude that there is enough evidence to support the alternative hypothesis.
    2. Conclude that there is ***not*** enough evidence to support the alternative hypothesis.

6. Two possible errors can be made:

    Type I error: Reject a true null hypothesis

    Type II error: Do not reject a false null hypothesis

Errors in Hypothesis Testing

A **Type I** error occurs when we reject a true null hypothesis (i.e. innocent person goes to prison)

A **Type II** error occurs when we don't reject a false null hypothesis (i.e. guilty person goes free)

P (***Type I Error***) = P (Reject H0|H0 is true) = α

P (***Type II Error***) = P (FTR H0|H0 is false) = β

Power of a Test = P (Reject H0|H0 is false)=1- β

Hypothesis Test Error

	H0 is true	H0 is false
Reject H0	**Worse!** **(Type I Error)**	Good! (**Power of Test**)
FTR H0	Good!	**Bad!** **(Type II Error)**

Details of Hypothesis Testing

- Hypotheses are a pair of mutually exclusive, collectively exhaustive statements about unknown (population) PARAMETERS (i.e. μ or π in our course)
- One statement or the other must be true, but they cannot both be true.
- H0: Null Hypothesis
  H1: Alternative Hypothesis

The NULL Hypothesis

- The statement of what would be true if any variation in the process being studied is caused only by chance.

- The current or maintained belief
- The target
- **Always contains the equals sign in some form.**

- Written as
  - H0: $\pi = 0.5$
  - H0: parameter =, ≤, ≥ a value

The ALTERNATIVE Hypothesis

- Something besides chance is operating in the process being studied.
- What we are trying to prove or investigate, hence the name: **research hypothesis**
- **Never contains an equal sign**.
- Written as
  - H1: $\pi \neq 0.5$
  - H1: parameter $\neq$, $<$, $>$ value

Some Language

Simple Null Hypothesis:

- When H0 specifies a single value (=), we have a two-tailed test

Composite Null Hypothesis:

- When H0 specifies a range (≤ or ≥), we have a one-tailed test
- It can be right tailed or left tailed
- The alternative hypothesis tells you which tail, as it “points” to it!

Some Practice Problems

1. The EPA estimated that 20% of auto emission systems are tampered with. You believe this is not true.

- H0:
- H1:

2. Kellogg's claims that breakfast cereal packages contain 32 oz, but customers have been complaining that they are underweight. You are Ralph Nader, the consumer watchdog.

- H0:
- H1:

Some Practice Problems: Answers

1. The EPA estimated that 20% of auto emission systems are tampered with. You believe this is not true.

- H0: $\pi = 20$
- H1: $\pi \neq 20$ Two tailed test

2. Kellogg's claims that breakfast cereal packages contain 32 oz, but customers have been complaining that they are underweight. You are Ralph Nader, the consumer watchdog.

- H0: $\mu \geq 32$
- H1: $\mu < 32$ Left tailed test

Some Practice Problems:

Florida Realtors report that 55% of those who purchase vacation residences want a condominium. Test to see if this statement is correct. A random sample of 400 found 228 who wanted a condominium.

1. What is the parameter being tested?

   A. p B. $\bar{x}$ C. μ D. π

2. What sign will be in the null hypothesis?

   A. $\neq$ B. = C. < D. $\geq$

3. What is the value of the sample statistic used as evidence in this situation?

   A. 220 B. 228 C. 0.55 D. 0.57

Some Practice Problems: Answers

Florida Realtors report that 55% of those who purchase vacation residences want a condominium. Test to see if this statement is correct. A random sample of 400 found 228 who wanted a condominium.

1. What is the parameter being tested?

   A. p   B. $\bar{x}$ C. μ   **D. π**

2. What sign will be in the null hypothesis?

   A. $\neq$ **B. =**   C. <   D.$\geq$

3. What is the value of the sample statistic used as evidence in this situation?

   A. 220   B. 228   C. 0.55   **D. 0.57**=228/400

Example 3

- The manager of a department store is thinking about establishing a new billing system for the store's credit customers.
- She determines that the new system will be cost-effective only if the mean monthly account is more than \$170.
- A random sample of 400 monthly accounts is drawn, for which the sample mean is \$178.
- The manager knows that the accounts are approximately normally distributed with a standard deviation of \$65.
- Can the manager conclude from this that the new system will be cost-effective?

Example 3

The system will be cost effective if the mean account balance for all customers is greater than $170.

We express this belief as our research hypothesis, that is:

H_1: $\mu > 170$
(this is what we want to determine)

Thus, our null hypothesis becomes:

H_0: $\mu \leq 170$

Example 3

What we want to show:

H_0: $\mu \leq 170$ (we'll *assume* this is true)

H_1: $\mu > 170$

We know:

$n = 400$,

$\bar{x}$ = **178**, and

$\sigma = 65$

What to do next?!

Example 3

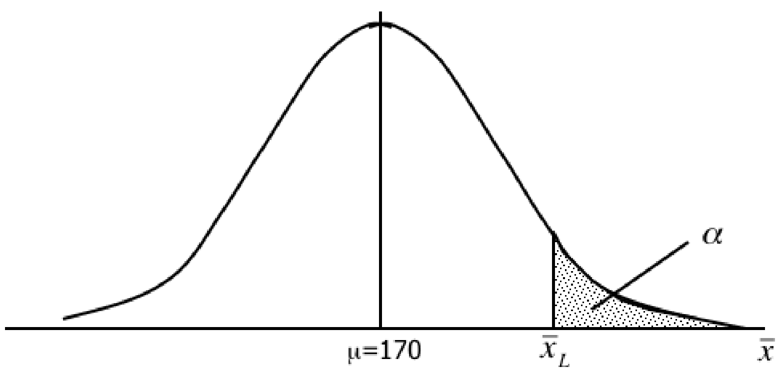
- To test our hypotheses, we can use two different approaches:

1. The ***rejection region*** approach (typically used when computing statistics manually), and
2. The ***p-value*** approach (which is generally used with a computer and statistical software).

- We will explore both in turn…

Example 3 Rejection Region

It seems reasonable to reject the null hypothesis in favor of the alternative if the value of the sample mean is ***large*** relative to 170, that is if $\bar{x} > \bar{x}_L$.



α = P(Type I error)

= P(reject H_0 given that H_0 is true)

$\alpha = P(\bar{x} > \bar{x}_L)$

Standardized Test Statistic

Transform sample mean of **178** into standardized z-score using:

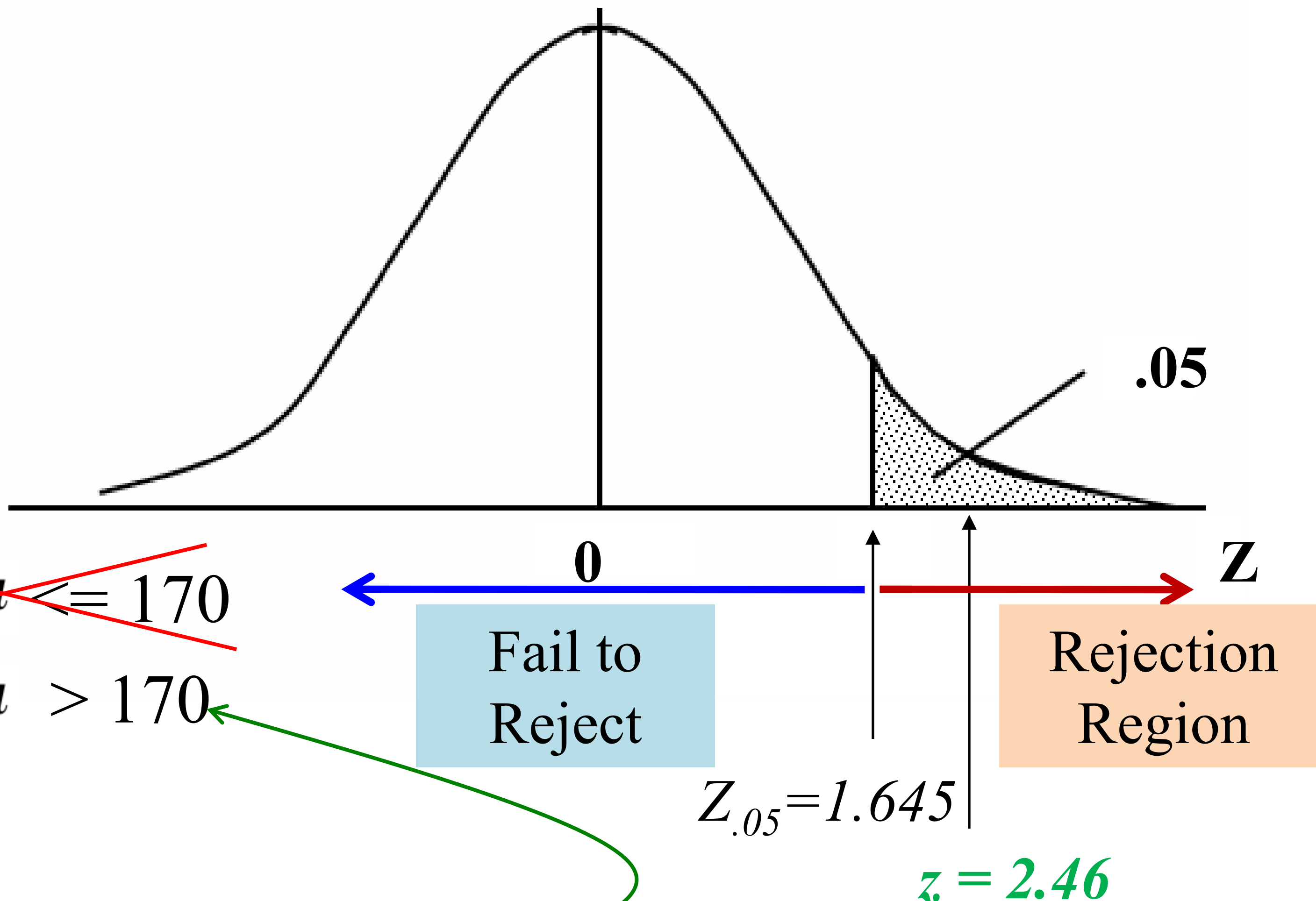
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

and compare its result to z_α: (rejection region: $z > z_\alpha$)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{178 - 170}{65 / \sqrt{400}} = 2.46$$

Since $z = 2.46 > 1.645$ ($= z_{.05}$), we reject H_0 in favor of H_1…

Example 3… The Big Picture Again



~~$H_0: \mu <= 170$~~

$H_1\ \mu > 170$

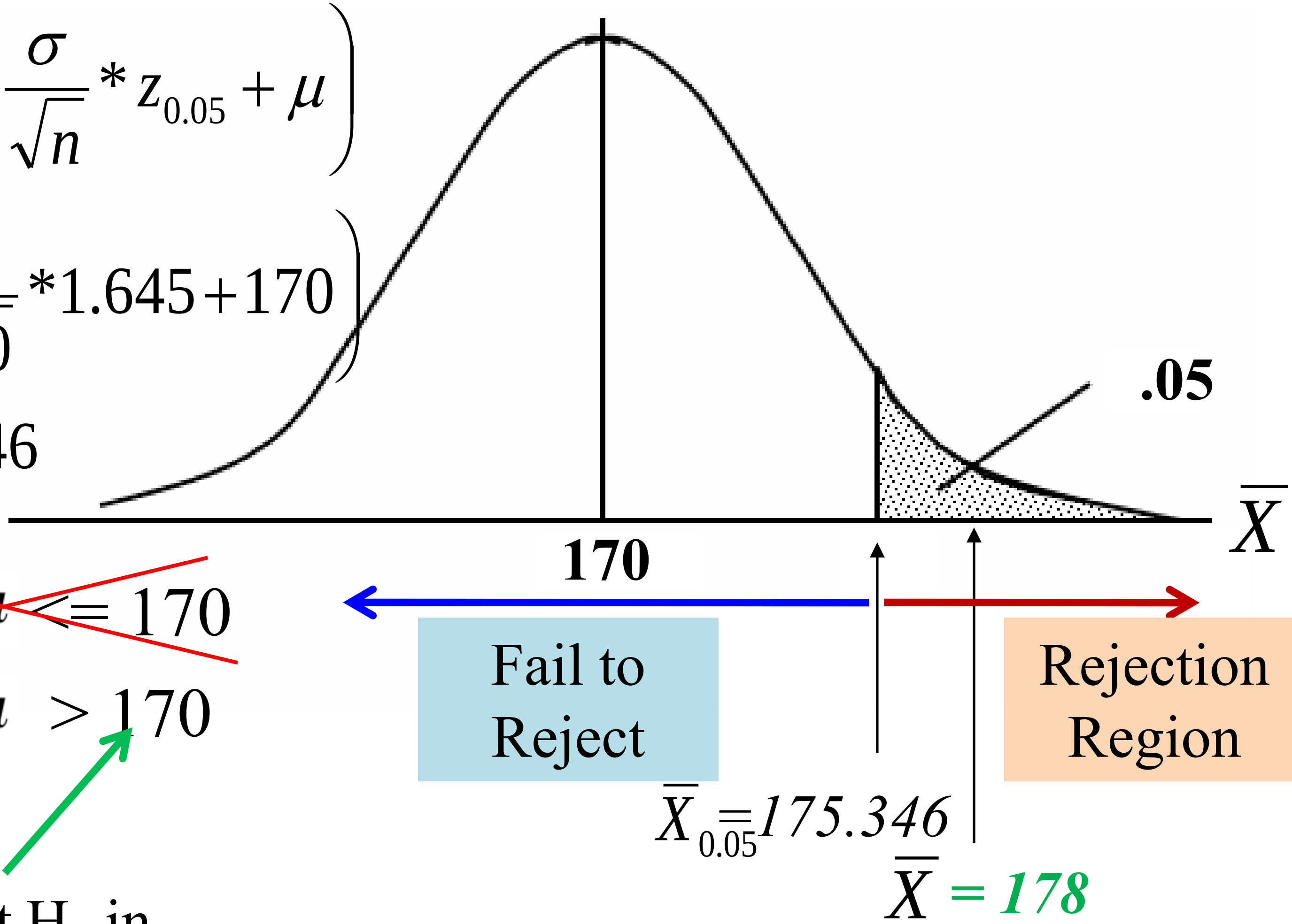
Reject H_0 in favor of H_1

Example 11.1… The Big Picture Again

$$\bar{X}_{0.05} = \left(\frac{\sigma}{\sqrt{n}} * z_{0.05} + \mu \right)$$

$$= \left(\frac{65}{\sqrt{400}} * 1.645 + 170 \right)$$

$$= 175.346$$



$\cancel{H_0: \mu <= 170}$

$H_1\ \mu > 170$

Reject H_0 in favor of H_1

p-Value of a Test

- The ***p-value*** of a test is the probability of observing a test statistic at least as extreme as the one computed given that the null hypothesis is true.
- In the case of our department store example, what is the ***probability*** of observing a sample mean ***at least as extreme*** as the one already observed (i.e. $\bar{x}$ = **178**), given that the null hypothesis (H_0: $\mu = 170$) is true?

$$P(\bar{x} > 178) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{178 - 170}{65/\sqrt{400}}\right) = P(Z > 2.46) = \boxed{.0069}$$

p-value

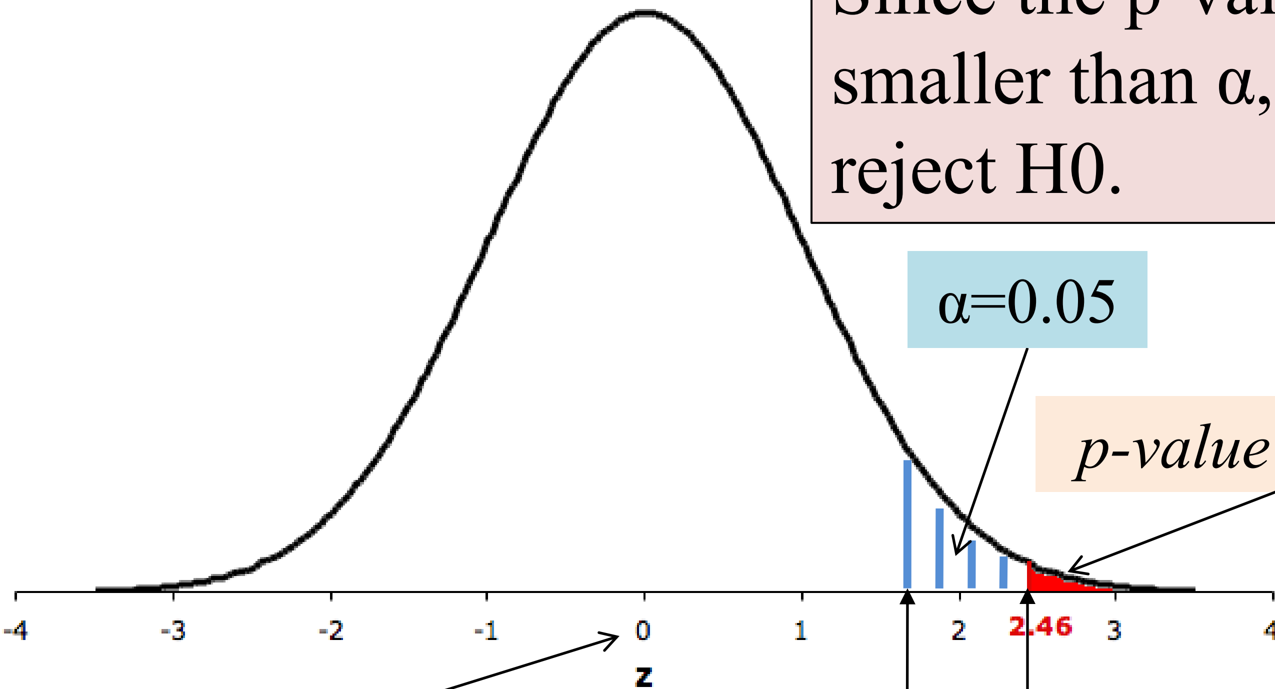
P-Value of a Test

p-value = $P(Z > 2.46)$ **= 0.069**

Since the p-value is smaller than α, we reject H0.

α=0.05

p-value =.0069



from H0: μ <= 170

$Z_{.05}=1.645$

$z = 2.46$

Interpreting the p-value

The smaller the p-value, the more statistical evidence exists to support the alternative hypothesis.

1. If the p-value is less than 1%, there is ***overwhelming evidence*** that supports the alternative hypothesis.
2. If the p-value is between 1% and 5%, there is a ***strong evidence*** that supports the alternative hypothesis.
3. If the p-value is between 5% and 10% there is a ***weak evidence*** that supports the alternative hypothesis.
4. If the p-value exceeds 10%, there is ***no evidence*** that supports the alternative hypothesis.

We observe a p-value of 0.0069, hence there is overwhelming evidence to support H_1: $\mu > 170$.

Interpreting the p-value

- Compare the p-value with the selected value of the significance level:
- If the p-value is less than α , we judge the p-value to be small enough to reject the null hypothesis.
- If the p-value is greater than α , we do not reject the null hypothesis.

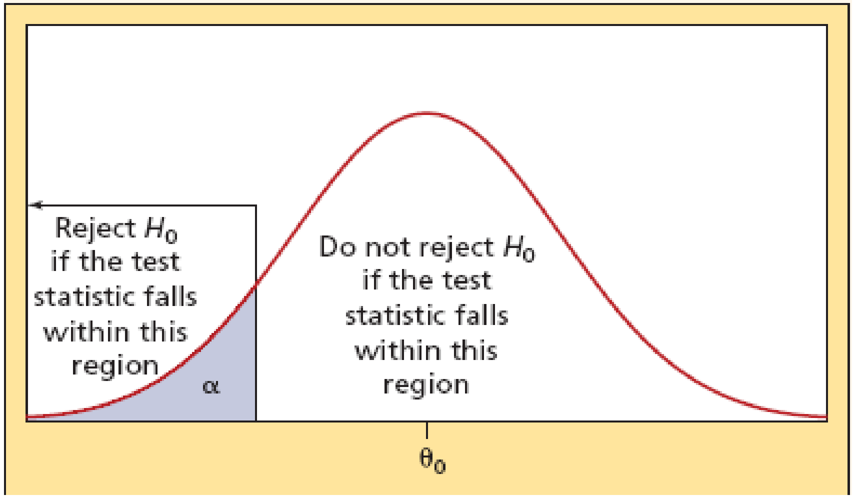
Since p-value = 0.0069 $\leq \alpha = 0.05$,

we reject H_0 in favor of H_1

Steps of Hypothesis Testing

1. State the hypothesis.
2. Select a level of significance (α),
   - our willingness to be WRONG,
   - the probability that we will reject a true null hypothesis!
   - Level of Confidence=1- α
3. Identify the test statistic (z or t score).
   - Depending on whether we know the population standard deviation (sigma) we use a z or t-score.
   - We call this observation z_{obs}, or t_{obs}.
4. Write a decision rule.
   - A statement of what you will do with your null when the sample information is close to the null and when it is far away from the null.
   - e.g. If $|z_{obs}| > z_{\alpha}$, reject H0, otherwise fail to reject H0.
   - If the probability of getting the test statistic farther from the null is smaller than α, reject H0, otherwise fail to reject (FTR) H0.
5. Select sample, calculate test statistic and make decision
   - Reject H0/FRT H0, according to the decision rule.

One Sided and Two Sided Tests



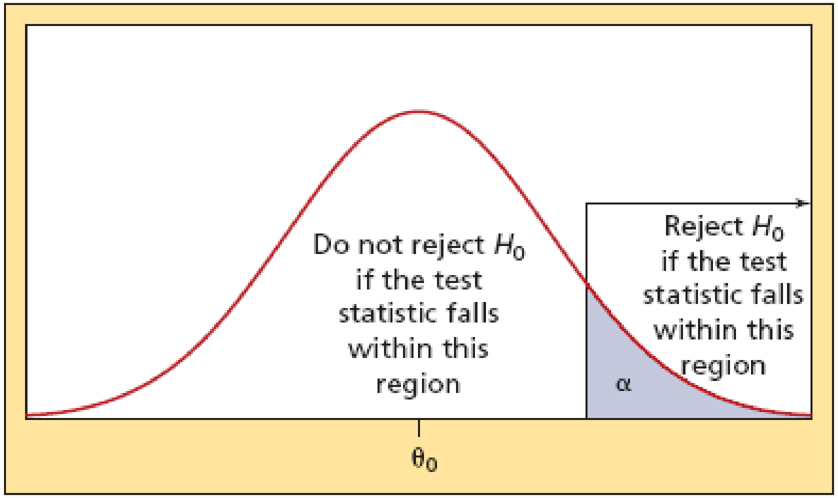
$$H_0 : \mu \geq 0$$

$$H_1 : \mu < 0$$

$$Pval = P(\overline{X} < \overline{x}_{obs})$$

$$where\ \overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$= norm.s.dist\left(\frac{\overline{x}_{obs} - \mu}{\sigma / \sqrt{n}}, 1\right)$$



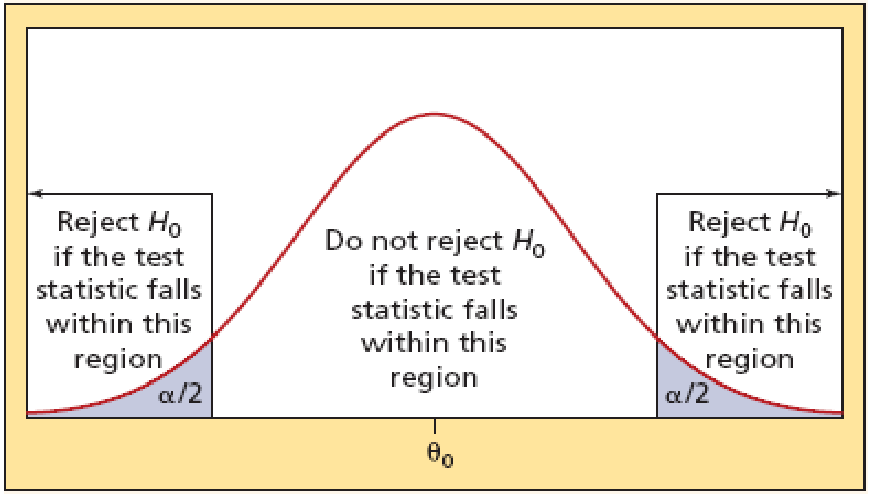
$$H_0 : \mu \leq 0$$

$$H_1 : \mu > 0$$

$$Pval = P(\overline{X} > \overline{x}_{obs})$$

$$where\ \overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$= 1 - norm.s.dist\left(\frac{\overline{x}_{obs} - \mu}{\sigma / \sqrt{n}}, 1\right)$$



$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

$$Pvalue = 2 * \left(1 - norm.s.dist\left(\left|\frac{\overline{x}_{obs} - \mu}{\sigma / \sqrt{n}}\right|, 1\right)\right)$$

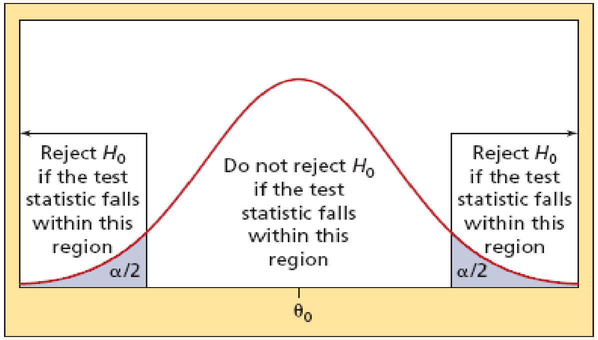
Example 1: Using *Critical Value* Method

$X \sim N(\mu, 5)$, a sample of 9 reveals a mean of 51. At $\alpha=0.05$,

1. Test whether population mean **is** 50.
2. Test whether population mean **is larger** than 50.

- Step1. State hypotheses
- Step2. Select a level of significance.
- Step3. Identify the test statistic
- Step4. Write a decision rule (and draw a picture).
- Step5. Select the sample, calculate the test statistic and apply the decision rule.

Test whether μ = 50



Step1.

H0: $\mu = 50$

H1: $\mu \neq 50$

(**TWO** tails test)

Step2. $\alpha=0.05$

Step3. $\sigma=5$ is known, use Z score, called $Z_{obs} = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

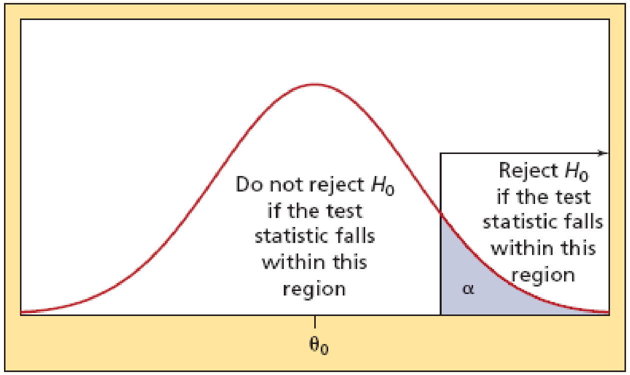
Step4. If $|Z_{obs}| > Z_{\alpha/2} = Z_{0.025} = 1.96$, reject H0; if not, FTR H0.

Where $Z_{\alpha/2}$ is called ***critical value*** ($Z_{critical}$). The region less than -1.96 and more than 1.96 is called ***rejection region***.

Step5. Select the sample, calculate the test statistic and apply the decision rule.

$$Z_{obs} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{51-50}{5/\sqrt{9}} = 0.6, |Z_{obs}| < 1.96, FTR, H0.$$

Test whether μ > 50



Step1. H0: $\mu \leq 50$

H1: $\mu > 50$

(**ONE** tailed test)

Step2. α=0.05

Step3. σ=5 is known, use Z score, called $$Z_{obs} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Step4. If $Z_{obs} > Z_{\alpha} = Z_{0.05} = 1.645$, reject H0; if not, FTR H0.

In this case, Z_{α} is ***critical value*** ($Z_{\mathbf{critical}}$). The region more than 1.645 is called ***rejection region***.

Step5. Select the sample, calculate the test statistic and apply the decision rule.

$$Z_{obs} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{51 - 50}{5 / \sqrt{9}} = 0.6, |Z_{obs}| < 1.645,\ FTR,\ H0$$

The “Trash Bag Advertisement” Case

- A leading manufacturer of trash bags has developed a new bag which they claim is stronger than their current 30 gallon trash bag.
- They wish to advertise this claim, but the TV network wants evidence of its truth. The current bag has a mean breaking strength of 50 pounds.
- A sample of 40 new bags was tested. The sample had a mean breaking strength of 50.575 pounds, with a standard deviation of 1.6438.
- Hint: When the population standard deviation is not known, we cannot calculate $\mathbf{Z}_{obs}$. Instead, we calculate $\mathbf{t}_{obs}$ and use $\mathbf{t}_{critical}$, while all other things remain the same.

Test whether $\mu > 50$ when σ unknown

Step1. H0: $\mu \le 50$
H1: $\mu > 50$
(**ONE** tail test)

Step2. $\alpha=0.05$

Step3. σ is **unknown**, use t score, called $$t_{obs} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Step4. If $t_{obs} > t_{\alpha}(39) = t_{0.05}(39) = \text{TINV}(0.1, 39) = 1.685$, reject H0; if not, FTR H0.
In this case, $t_{\alpha}(39)$ is ***critical value*** ($t_{\textbf{critical}}$). The region more than 1.685 is called ***rejection region***.

Step5. Select the sample, calculate the test statistic and apply the decision rule.

$$t_{obs} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{50.575 - 50}{1.6438/\sqrt{40}} = 2.212, t_{obs} > 1.685,$$ so reject H0.

Conclude the new bag is stronger than the current bag.

Hypothesis about Population Proportion

- When dealing with hypothesis about population proportion,
- first check whether the sample proportion ***p*** is normally distributed: $n\pi \geq 5$ and $n(1-\pi) \geq 5$.
- If ***p*** is normal, use $\mathbf{Z_{obs}}$ and $Z_{\mathbf{critical}}$. If not, cannot do test.

$$Z_{obs} = \frac{p - \pi}{\sqrt{\pi(1-\pi)/n}}$$

- All other things remain same.

Practice

1. Which of the following significance levels will lead to the most frequent rejection of the null in repeated testing?

   A. 0.10 B. 0.001 C. 0.05 D. 0.01

2. H0: μ $\geq$ 17 vs. H1: μ < 17

   a. What is the sample statistic used in this test?

      A. μ B. $\bar{x}$ C. p D. 17

   b. How many tails does the test have?

      A. 1 B. 2 C. 3 D. 17

   c. If this test was a right-tailed test, what sign would be in the null hypothesis?

      A. < B. $\geq$ C. > D. $\leq$

3. If you wished to perform a hypothesis test with a simple null hypothesis, at a significance level of 5%, what value of alpha will be found in the right tail?

   A. 0.000 B.0.025 C.0.050 D. Insufficient information

4. Which of the following is a standardized test statistic?

   A. t_{obs} B. $\hat{p}_{obs}$ C. $\bar{x}$ D. μ

Practice

1. Which of the following significance levels will lead to the most frequent rejection of the null in repeated testing?

   A. **0.10** B. 0.001 C. 0.05 D. 0.01

2. H0: μ >= 17 v.s. H1: μ < 17

   a. What is the sample statistic used in this test?

   A. μ B. $\bar{x}$ C. p D. 17

   b. How many tails does the test have?

   A. **1** B. 2 C. 3 D. 17

   c. If this test was a right-tailed test, what sign would be in the null hypothesis?

   A. < B. $\geq$ C. > D. $\leq$

3. If you wished to perform a hypothesis test with a simple null hypothesis, at a significance level of 5%, what value of alpha will be found in the right tail?

   A. 0.000 B.**0.025** C.0.050 D. Insufficient information

4. Which of the following is a standardized test statistic?

   A. t_{obs} B. $\hat{p}_{obs}$ C. $\bar{x}$ D. μ

Example 1: Using *P-value* method:

$X \sim N(\mu, 5)$, a sample of 9 reveals a mean of 51. At $\alpha=0.05$,

1. Test whether population mean **is bigger** than 50.
2. Test whether population mean **is** 50.

Step 1. State hypotheses:

Step 2. Select a level of significance.

Step 3. Identify the test statistic

Step 4. Write a decision rule (and draw a picture).

Step 5. Select the sample, calculate the test statistic and apply the decision rule.

Test whether μ > 50

Step1. H0: $\mu \leq 50$

H1: $\mu > 50$

(**ONE** tail test)

Step2. α=0.05

Step3. σ=5 is known, use Z score, called $Z_{obs} = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$.

Step4. If ***P-value***(Z_{obs})<**α**, reject H0; if not, FTR H0

Step5. Select the sample, calculate the test statistic and apply the decision rule.

$$Z_{obs} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{51-50}{5/\sqrt{9}} = 0.6,$$

$$P - value = P(Z > Z_{obs}) = NORMSDIST(-0.6) = 0.274$$

0.274>0.05, FTR H0.

Test whether μ = 50

Step1. H0: $\mu = 50$
H1: $\mu \neq 50$
(**TWO** tails test)

Step2. **α**=0.05

Step3. **σ**=5 is known, use Z score, called $Z_{obs} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$.

Step4. If ***P-value***(Z_{obs})<**α,** reject H0; if not, FTR H0.

Step5. Select the sample, calculate the test statistic and apply the decision rule.

$$Z_{obs} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{51 - 50}{5/\sqrt{9}} = 0.6,$$

$$P - value = P(|Z| > |Z_{obs}|) = NORMSDIST(-0.6) \times 2 = 0.55$$

0.55> **α**=0.05, FTR H0.

Practice

1. A television warranty claims that not more than 10% of its sets need repair during the first 2 years of operation. A sample of 100 sets found 14 that required repair. Is the claim valid?
2. A plastics manufacturer wants to evaluate the durability of some blocks used in furniture production. Is there evidence that the average hardness of the blocks exceeds 260 Brinell units? A sample of 50 blocks has a mean of 267.6 units, $s = 24.4$.
3. A new machine at a clothing factory is supposed to be producing cloth that has a mean breaking strength of 70 lbs with $\sigma = 3.5$ lbs. Management is concerned about lawsuits if the breaking strength is lower. A sample of 49 pieces had a mean of 69.1 lbs.
4. The purchase of a coin-operated laundry is being considered by a potential entrepreneur. The present owner claims that over the past 5 years the average daily revenue has been \$675, with $\sigma = \$75$. A sample of 30 days has a mean of \$625.

Optional

- Type II Error
- Power of a Test

Probability of a Type II Error

It is important that that we understand the relationship between Type I and Type II errors; that is, how the probability of a Type II error is calculated and its interpretation.

Recall Example 11.1:

H_0: μ <= 170

H_1: μ > 170

At a significance level of 5% we rejected H_0 in favor of H_1 since our sample mean (178) was greater than the critical value of $\bar{x}$ (175.34).

Probability of a Type II Error β

A Type II error occurs when a false null hypothesis is not rejected.

In example 11.1, this means that if $\bar{x}$ is less than 175.34 (our critical value) we will **not reject** our null hypothesis, which means that we will not install the new billing system.

Thus, we can see that:

$$\beta = P(\bar{x} < 175.34 \text{ given that the null hypothesis is false})$$

Example 11.1 (revisited)

β = P($\bar{x}$ < 175.34 given that the null hypothesis is false)

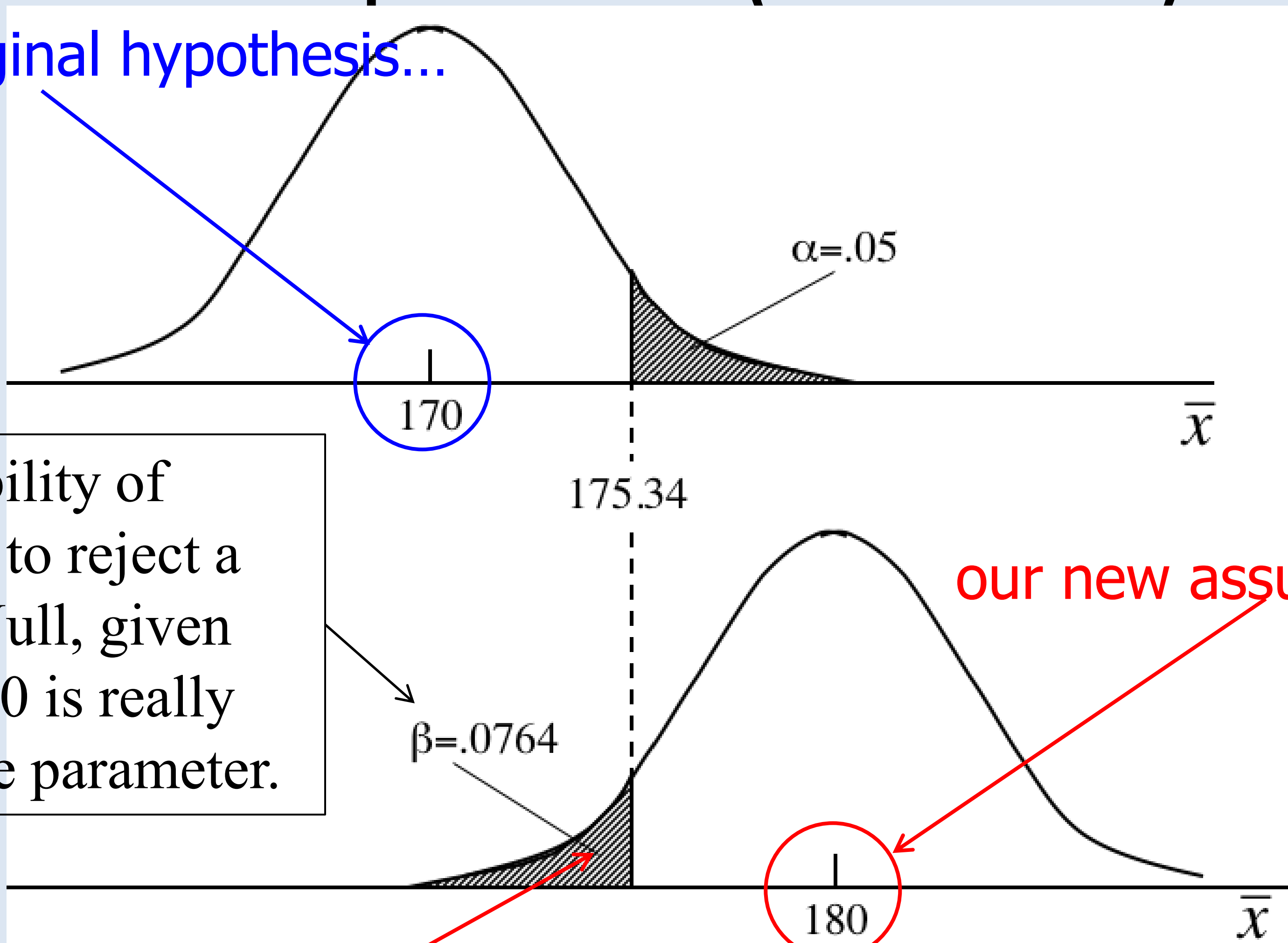
The condition only tells us that the mean ≠ 170. We need to compute β for some new value of μ. For example, suppose that if the mean account balance is $180 the new billing system will be so profitable that we would hate to lose the opportunity to install it.

β = P($\bar{x}$ < 175.34, given that μ = 180), thus…

$$\beta = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{175.34-180}{65/\sqrt{400}}\right) = P(Z < -1.43) = .0764$$

Example 11.1 (revisited)

Our original hypothesis...



our new assumption...

Probability of failing to reject a false Null, given that 180 is really the true parameter.

$$\beta = P(\bar{x} < 175.34, \text{ given that } \mu = 180)$$

Effects on β of Changing α

Decreasing the significance level α, increases the value of β and vice versa. Change α to .01 in Example 11.1.

Stage 1: Rejection region

$$z > z_{\alpha} = z_{.01} = 2.33$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{x} - 170}{65 / \sqrt{400}} > 2.33$$

$$\bar{x} > 177.57$$

Effects on β of Changing α

Stage 2 Probability of a Type II error

$$\beta = P(\bar{x} < 177.57 \mid \mu = 180)$$

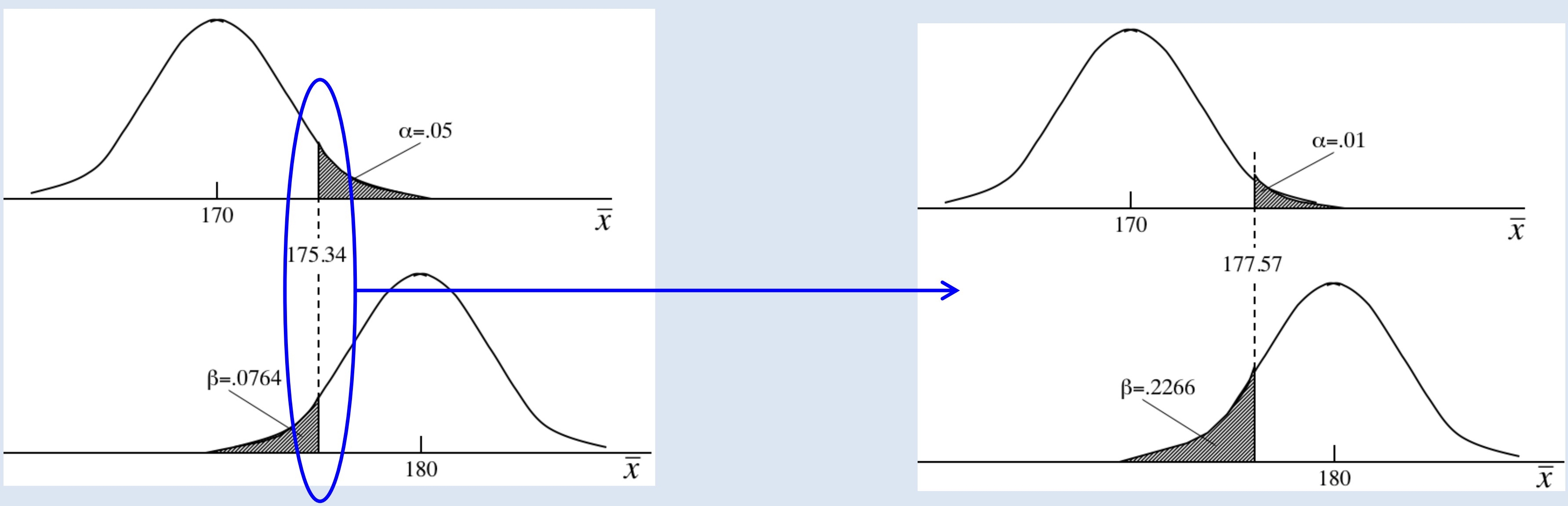
$$= P\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < \frac{177.57 - 180}{65 / \sqrt{400}} \right)$$

$$= P(z < -.75)$$

$$= .2266$$

Effects on β of Changing α

- Decreasing the significance level α, increases the value of β and vice versa.
- Consider this diagram again. Shifting the critical value line to the right (to decrease α) will mean a larger area under the lower curve for β… (and vice versa).



Judging the Test

A statistical test of hypothesis is effectively defined by the significance level (α) and the sample size (n), ***both of which are selected*** by the statistics practitioner.

Therefore, if the probability of a Type II error (β) is judged to be too large, we can reduce it by

Increasing α,

and/or

increasing the sample size, n.

Judging the Test

For example, suppose we increased n from a sample size of 400 account balances to 1,000 in Example 11.1.

Stage 1: Rejection region

$$z > z_{\alpha} = z_{.05} = 1.645$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{x} - 170}{65 / \sqrt{1{,}000}} > 1.645$$

$$\bar{x} > 173.38$$

Judging the Test

Stage 2: Probability of a Type II error

$$\beta = P(\bar{x} < 173.38 \mid \mu = 180)$$

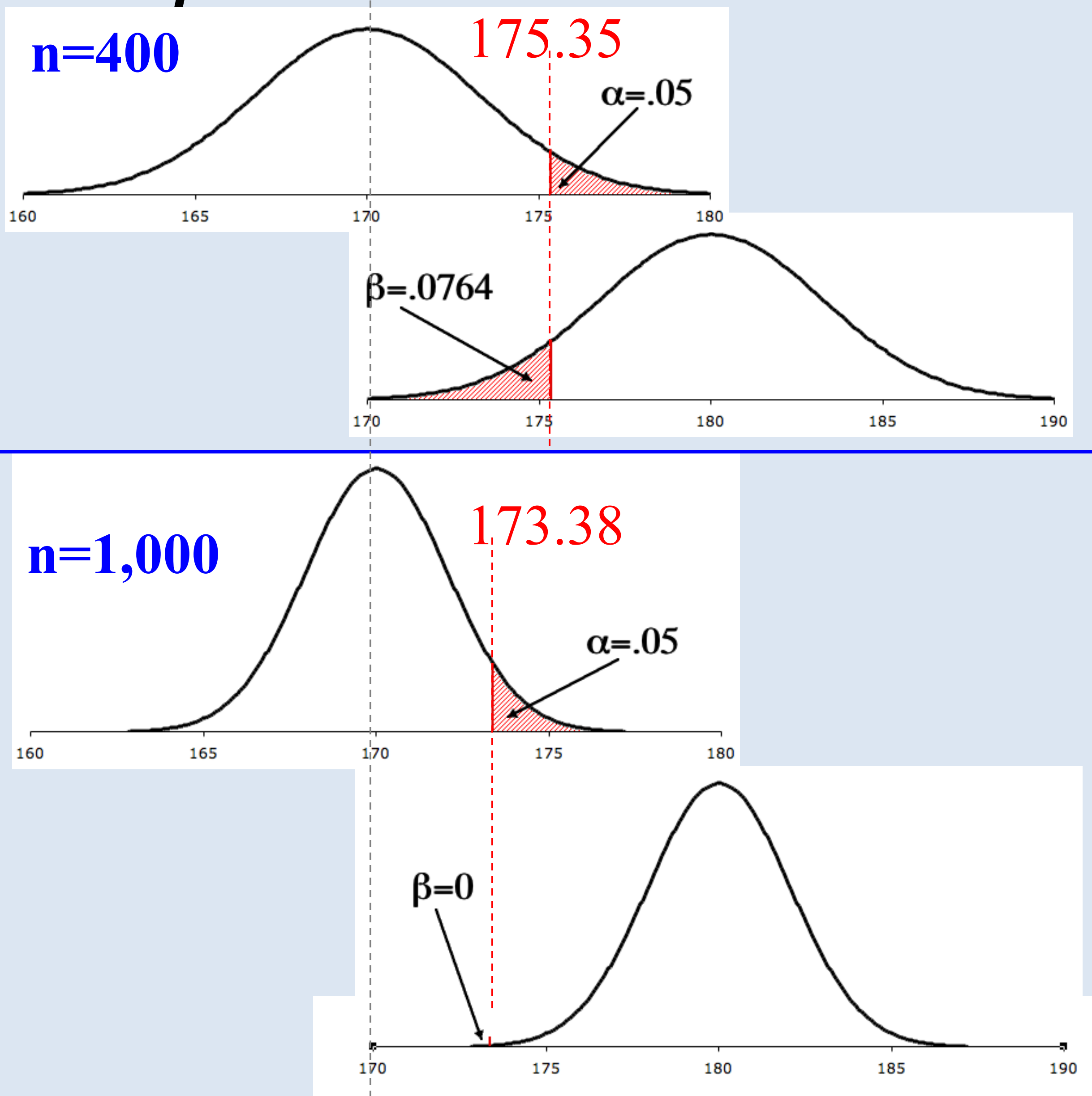
$$= P\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < \frac{173.38 - 180}{65 / \sqrt{1{,}000}} \right)$$

$$= P(z < -3.22)$$

$$= 0 \text{ (approximately)}$$

Compare β at n=400 and n=1,000…

By increasing the sample size we reduce the probability of a Type II error:



Developing an Understanding of Statistical Concepts

The calculation of the probability of a Type II error for n = 400 and for n = 1,000 illustrates a concept whose importance cannot be overstated.

By increasing the sample size we **reduce** the probability of a Type II error. By reducing the probability of a Type II error we make this type of error less frequently.

And hence, we make better decisions in the long run. This finding lies at the heart of applied statistical analysis and reinforces the book's first sentence, "Statistics is a way to get information from data."

Power of a Test

Another way of expressing how well a test performs is to report its *power*: **the probability of its leading us to reject the null hypothesis when it is false**. Thus, the power of a test is: 1- β

When more than one test can be performed in a given situation, we would naturally prefer to use the test that is correct more frequently.

If (given the same alternative hypothesis, sample size, and significance level) one test has a higher power than a second test, the first test is said to be more powerful.

Example Calculating β

Calculate the probability of a Type II error when the actual mean is 21.

Recall that:

$H_0: \mu = 22$

$H_1: \mu < 22$

$n = 220$

$\sigma = 6$

$\alpha = .10$

Example Calculating β

Stage 1: Rejection region

$$z < -z_{\alpha} = -z_{.10} = -1.28$$

$$\frac{\bar{x} - 22}{6\sqrt{220}} < -1.28$$

$$\bar{x} < 21.48$$

Example Calculating β

Stage 2: Probability of a Type II error:

$$\beta = P(\bar{x} > 21.48 \mid \mu = 21)$$

$$= P\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} > \frac{21.48 - 21}{6 / \sqrt{220}} \right)$$

$$= P(z > 1.19)$$

$$= .1170$$

Example 11.2 Calculating β

Calculate the probability of a Type II error when the actual mean is 16.80.

Recall that

$H_0: \mu = 17.09$

$H_1: \mu \neq 17.09$

$n = 100$

$\sigma = 3.87$

$\alpha = .05$

Example 11.2 Calculating β

Stage 1: Rejection region (two-tailed test)

$$z > z_{\alpha/2} \text{ or } z < -z_{\alpha/2}$$

$$z > z_{.025} = 1.96 \text{ or } z < -z_{.025} = -1.96$$

$$\frac{\bar{x} - 17.09}{3.87\sqrt{100}} > 1.96 \quad \Rightarrow \quad \bar{x} > 17.85$$

$$\frac{\bar{x} - 17.09}{3.87/\sqrt{100}} < -1.96 \Rightarrow \quad \bar{x} < 16.33$$

Example 11.2 Calculating β

Stage 2: Probability of a Type II error

$$\beta = P(16.33 < \bar{x} < 17.85 \mid \mu = 16.80)$$

$$= P\left(\frac{16.33 - 16.80}{3.87/\sqrt{100}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{17.85 - 16.80}{3.87/\sqrt{100}}\right)$$

$$= P(-1.21 < z < 2.71)$$

$$= .8835$$