Chapter 11

Introduction to Hypothesis Testing

Statistical Inference

Hypothesis testing is the second form of statistical inference. It also has greater applicability.

To understand the concepts we'll start with an example of *nonstatistical hypothesis testing.*

Nonstatistical Hypothesis Testing

A criminal trial is an example of hypothesis testing without the statistics.

In a trial a jury must decide between two hypotheses. The null hypothesis is

H_0: The defendant is innocent

The alternative hypothesis or research hypothesis is

H_1: The defendant is guilty

The jury does not know which hypothesis is true. They must make a decision on the basis of evidence presented.

Nonstatistical 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).

Nonstatistical Hypothesis Testing

If the jury acquits 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.

Nonstatistical Hypothesis Testing

There are two possible errors.

A Type I error occurs when we reject a true null hypothesis. That is, a Type I error occurs when the jury convicts an innocent person.

A Type II error occurs when we don't reject a false null hypothesis. That occurs when a guilty defendant is acquitted.

Nonstatistical Hypothesis Testing

The probability of a Type I error is denoted as α (Greek letter *alpha*). The probability of a type II error is β (Greek letter *beta*).

The two probabilities are inversely related. Decreasing one increases the other.

Nonstatistical Hypothesis Testing

In our judicial system Type I errors are regarded as more serious. We try to avoid convicting innocent people. We are more willing to acquit guilty people.

We arrange to make α small by requiring the prosecution to prove its case and instructing the jury to find the defendant guilty only if there is "evidence beyond a reasonable doubt."

Nonstatistical Hypothesis Testing

The critical concepts are theses:

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.
4. There are two possible decisions:

   Conclude that there is enough evidence to support the alternative hypothesis.

   Conclude that there is ***not*** enough evidence to support the alternative hypothesis.

Nonstatistical Hypothesis Testing

5. Two possible errors can be made.

   Type I error: Reject a true null hypothesis

   Type II error: Do not reject a false null hypothesis.

   P(Type I error) = α

   P(Type II error) = β

Concepts of Hypothesis Testing (1)

There are **two** hypotheses. One is called the ***null hypothesis*** and the other the ***alternative*** or ***research hypothesis***. The usual notation is:

pronounced
H "nought"

H_0: — *the 'null' hypothesis*

H_1: — *the 'alternative' or 'research' hypothesis*

The null hypothesis (H_0) will always state that the ***parameter equals the value*** specified in the alternative hypothesis (H_1)

Concepts of Hypothesis Testing

Consider Example 10.1 (mean demand for computers during assembly lead time) again. Rather than estimate the mean demand, our operations manager wants to know ***whether the mean is different from 350 units***. We can rephrase this request into a test of the hypothesis:

$H_0: \mu = 350$

Thus, our research hypothesis becomes:

$H_1: \mu \neq 350$

This is what we are interested in determining…

Concepts of Hypothesis Testing (2)

The testing procedure begins with the ***assumption that the null hypothesis is true***.

Thus, until we have further statistical evidence, we will **assume**:

H_0: $\mu = 350$ (assumed to be TRUE)

Concepts of Hypothesis Testing (3)

The **goal** of the process is to determine ***whether there is enough evidence*** to infer that the alternative hypothesis is true.

That is, is there sufficient statistical information to determine if this statement is true?

H_1: $\mu \neq 350$

This is what we are interested in determining…

Concepts of Hypothesis Testing (4)

There are **two** possible decisions that can be made:

Conclude that there ***is enough evidence*** to support the alternative hypothesis

(also stated as: rejecting the null hypothesis in favor of the alternative)

Conclude that there ***is not enough evidence*** to support the alternative hypothesis

(also stated as: **not** rejecting the null hypothesis in favor of the alternative)

NOTE: we **do not** say that we **accept** the null hypothesis…

Concepts of Hypothesis Testing

Once the null and alternative hypotheses are stated, the next step is to randomly sample the population and calculate a ***test statistic*** (in this example, the sample mean).

If the test statistic's value is inconsistent with the null hypothesis ***we reject the null hypothesis*** and ***infer that the alternative hypothesis is true***.

Concepts of Hypothesis Testing

For example, if we're trying to decide whether the mean is not equal to 350, a large value of $\bar{x}$ (say, 600) would provide enough evidence.

If $\bar{x}$ is close to 350 (say, 355) we could not say that this provides a great deal of evidence to infer that the population mean is different than 350.

Concepts of Hypothesis Testing (5)

Two possible errors can be made in any test:

A Type I error occurs when we reject a true null hypothesis and

A Type II error occurs when we don't reject a false null hypothesis.

There are probabilities associated with each type of error:

P(Type I error) $= \alpha$

P(Type II error) $= \beta$

α is called the *significance level.*

Types of Errors

A Type I error occurs when we ***reject*** a ***true*** null hypothesis (i.e. Reject H_0 when it is TRUE)

H_0	T	F
Reject	I	
~~Reject~~		II

A Type II error occurs when we ***don't reject*** a ***false*** null hypothesis (i.e. Do NOT reject H_0 when it is FALSE)

Example 11.1

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 11.1

IDENTIFY

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: $\mu = 170$ (this specifies a single value for the parameter of interest)

Example 11.1

IDENTIFY

What we want to show:

H_0: $\mu = 170$ (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 11.1

COMPUTE

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

The ***rejection region*** approach (typically used when computing statistics manually), and

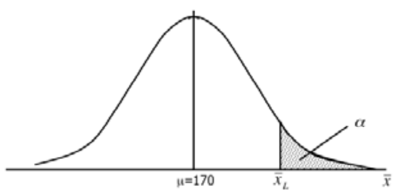
The ***p-value*** approach (which is generally used with a computer and statistical software).

We will explore both in turn…

Example 11.1 Rejection region

COMPUTE

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)

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

Example 11.1

COMPUTE

All that's left to do is calculate $\bar{x}_L$ and compare it to 170.

$$P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > \frac{\bar{x}_L-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > \frac{\bar{x}_L-\mu}{\sigma/\sqrt{n}}\right) = \alpha$$

$$P(Z > z_\alpha) = \alpha$$

$$\therefore \frac{\bar{x}_L-\mu}{\sigma/\sqrt{n}} = z_\alpha$$

we can calculate this based on any level of significance (α) we want…

Example 11.1

COMPUTE

At a 5% significance level (i.e. α =0.05), we get

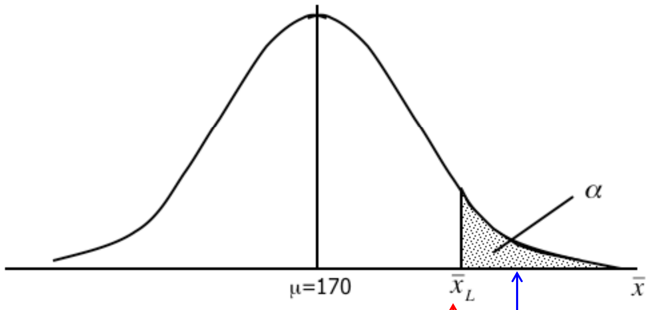
$$\frac{\bar{x}_L-\mu}{\sigma/\sqrt{n}} = z_\alpha \quad \& \quad z_\alpha = z_{.05} = 1.645$$

$$gives: \frac{\bar{x}_L-170}{65/\sqrt{400}} = 1.645$$

Solving we compute $\bar{x}_L$ = 175.34

Since our sample mean (178) is ***greater than*** the critical value we calculated (175.34), we reject the null hypothesis in favor of H_1, i.e. that: $\mu > 170$ and that it is cost effective to install the new billing system

Example 11.1 The Big Picture



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

$H_1: \mu > 170$

$\bar{x}_L = 175.34$

$\bar{x} = 178$

Reject H_0 in favor of

Standardized Test Statistic

An easier method is to use the standardized test statistic:

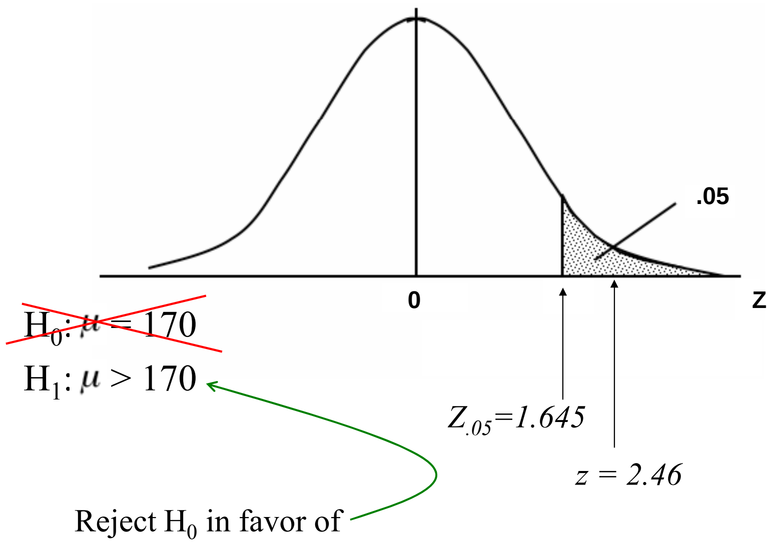
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

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 11.1… The Big Picture Again



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.

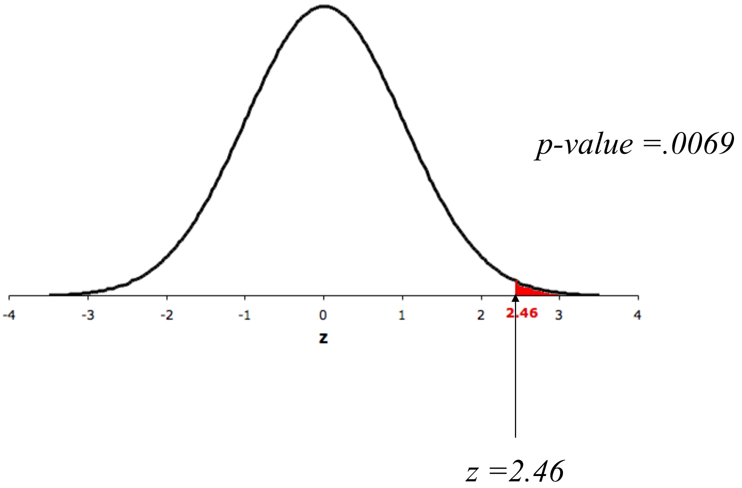
In the case of our department store example, what is the ***probability*** of observing a sample mean ***at least as extreme*** as the one already observed (i.e. $\bar{x} = 178$), given that the null hypothesis (H_0: $\mu = 170$) is true?

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

p-value

P-Value of a Test

p-value = P(Z > 2.46)



Interpreting the p-value

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

If the p-value is less than 1%, there is ***overwhelming evidence*** that supports the alternative hypothesis.

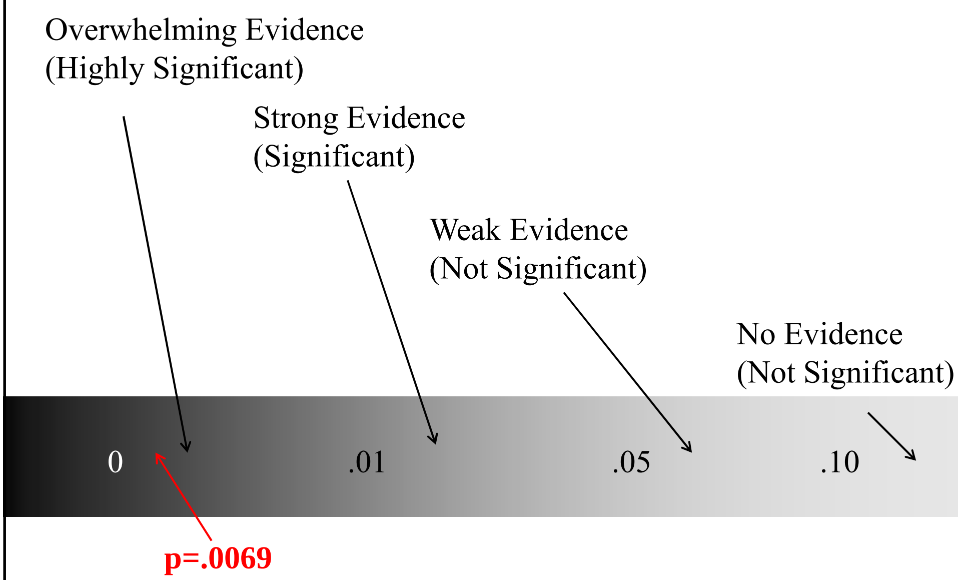
If the p-value is between 1% and 5%, there is a ***strong evidence*** that supports the alternative hypothesis.

If the p-value is between 5% and 10% there is a ***weak evidence*** that supports the alternative hypothesis.

If the p-value exceeds 10%, there is ***no evidence*** that supports the alternative hypothesis.

We observe a p-value of .0069, hence there is overwhelming evidence to support H_1: $\mu > 170$.

Interpreting the p-value



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 = .0069 < α = .05, we reject H_0 in favor of H_1

Example 11.1

COMPUTE

	A	B	C	D
1	**Z-Test: Mean**			
2				
3				*Accounts*
4	Mean			178.00
5	Standard Deviation			68.37
6	Observations			400
7	Hypothesized Mean			170
8	SIGMA			65
9	z Stat			2.46
10	P(Z<=z) one-tail			0.0069
11	z Critical one-tail			1.6449
12	P(Z<=z) two-tail			0.0138
13	z Critical two-tail			1.96

Conclusions of a Test of Hypothesis

If we reject the null hypothesis, we conclude that there is enough evidence to infer that the alternative hypothesis is true.

If we do not reject the null hypothesis, we conclude that there is not enough statistical evidence to infer that the alternative hypothesis is true.

Remember: The alternative hypothesis is the more important one. It represents what we are investigating.

Chapter-Opening Example SSA Envelope Plan

Federal Express (FedEx) sends invoices to customers requesting payment within 30 days.

The bill lists an address and customers are expected to use their own envelopes to return their payments.

Currently the mean and standard deviation of the amount of time taken to pay bills are 24 days and 6 days, respectively.

The chief financial officer (CFO) believes that including a stamped self-addressed (SSA) envelope would decrease the amount of time.

Chapter-Opening Example SSA Envelope Plan

She calculates that the improved cash flow from a 2-day decrease in the payment period would pay for the costs of the envelopes and stamps.

Any further decrease in the payment period would generate a profit.

To test her belief she randomly selects 220 customers and includes a stamped self-addressed envelope with their invoices.

The numbers of days until payment is received were recorded. Can the CFO conclude that the plan will be profitable?

SSA Envelope Plan

IDENTIFY

The objective of the study is to draw a conclusion about the mean payment period. Thus, the parameter to be tested is the population mean.

We want to know whether there is enough statistical evidence to show that the population mean is less than 22 days. Thus, the alternative hypothesis is

$$H_1: \mu < 22$$

The null hypothesis is

$$H_0: \mu = 22$$

SSA Envelope Plan

IDENTIFY

The test statistic is

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

We wish to reject the null hypothesis in favor of the alternative only if the sample mean and hence the value of the test statistic is small enough.

As a result we locate the rejection region in the left tail of the sampling distribution.

We set the significance level at 10%.

SSA Envelope Plan

COMPUTE

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

From the data in Xm11-00 we compute

and
$$\bar{x} = \frac{\sum x_i}{220} = \frac{4{,}759}{220} = 21.63$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{21.63 - 22}{6 / \sqrt{220}} = -.91$$

p-value = P(Z < -.91) = .5 - .3186 = .1814

SSA Envelope Plan

COMPUTE

	A	B	C	D
1	**Z-Test: Mean**			
2				
3				*Payment*
4	Mean			21.63
5	Standard Deviation			5.84
6	Observations			220
7	Hypothesized Mean			22
8	SIGMA			6
9	z Stat			-0.91
10	P(Z<=z) one-tail			0.1814
11	z Critical one-tail			1.2816
12	P(Z<=z) two-tail			0.3628
13	z Critical two-tail			1.6449

SSA Envelope Plan

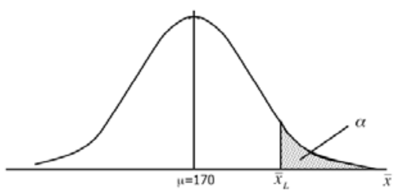
INTERPRET

Conclusion: There is not enough evidence to infer that the mean is less than 22.

There is not enough evidence to infer that the plan will be profitable.

One– and Two–Tail Testing

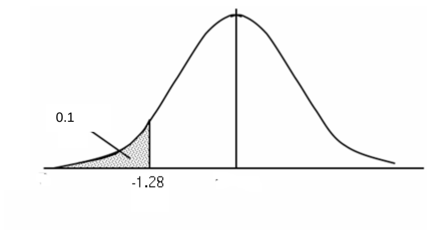
The department store example (Example 11.1) was a ***one tail test***, because the rejection region is located in only one tail of the sampling distribution:



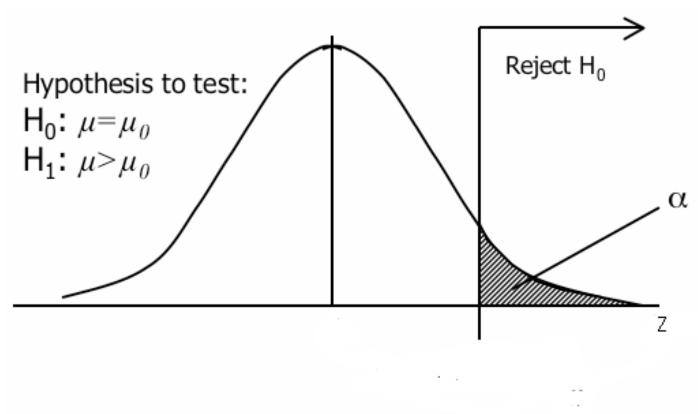
More correctly, this was an example of a ***right tail test***.

One– and Two–Tail Testing

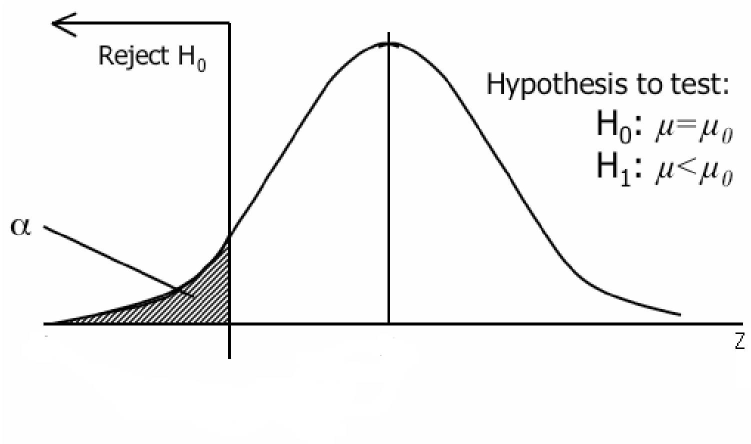
The SSA Envelope example is a left tail test because the rejection region was located in the ***left*** tail of the sampling distribution.



Right-Tail Testing

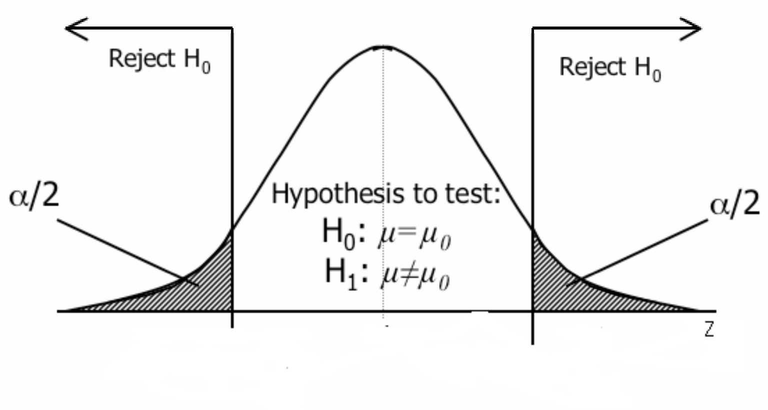


Left-Tail Testing



Two–Tail Testing

Two tail testing is used when we want to test a research hypothesis that a parameter is not equal (≠) to some value



Example 11.2

In recent years, a number of companies have been formed that offer competition to AT&T in long-distance calls.

All advertise that their rates are lower than AT&T's, and as a result their bills will be lower.

AT&T has responded by arguing that for the average consumer there will be no difference in billing.

Suppose that a statistics practitioner working for AT&T determines that the mean and standard deviation of monthly long-distance bills for all its residential customers are $17.09 and $3.87, respectively.

Example 11.2

He then takes a random sample of 100 customers and recalculates their last month's bill using the rates quoted by a leading competitor.

Assuming that the standard deviation of this population is the same as for AT&T, can we conclude at the 5% significance level that there is a difference between AT&T's bills and those of the leading competitor?

Example 11.2

IDENTIFY

The parameter to be tested is the mean of the population of AT&T's customers' bills based on competitor's rates.

What we want to determine whether this mean differs from $17.09. Thus, the alternative hypothesis is

H_1: $\mu \neq 17.09$

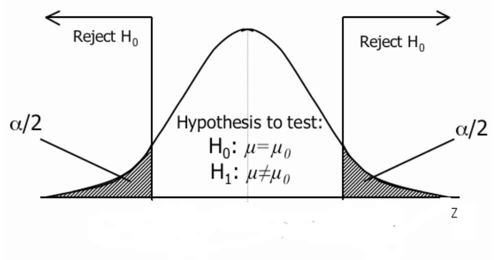
The null hypothesis automatically follows.

H_0: $\mu = 17.09$

Example 11.2

IDENTIFY

The rejection region is set up so we can reject the null hypothesis when the test statistic is large ***or*** when it is small.



stat is "small"

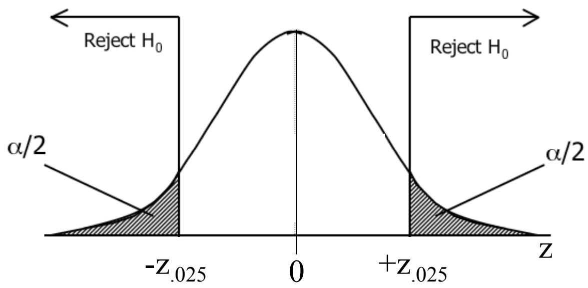
stat is "large"

That is, we set up a two-tail rejection region. The total area in the rejection region must sum to α, so we divide this probability by 2.

Example 11.2 IDENTIFY

At a 5% significance level (i.e. $\alpha = .05$), we have $\alpha/2 = .025$. Thus, $z_{.025} = 1.96$ and our rejection region is:

$z < -1.96$ **-or-** $z > 1.96$



Example 11.2 COMPUTE

From the data (Xm11-02), we calculate $\bar{x} = 17.55$

Using our standardized test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

We find that: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{17.55 - 17.09}{3.87 / \sqrt{100}} = 1.19$

Since $z = 1.19$ is not greater than 1.96, nor less than –1.96 we cannot reject the null hypothesis in favor of H_1. That is ***"there is insufficient evidence to infer that there is a difference between the bills of AT&T and the competitor."***

Two-Tail Test p-value

COMPUTE

In general, the p-value in a two-tail test is determined by

$$\text{p-value} = 2P(Z > |z|)$$

where z is the actual value of the test statistic and |z| is its absolute value.

For Example 11.2 we find

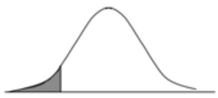
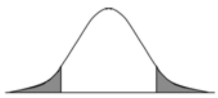
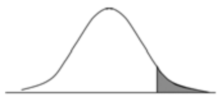
$$\begin{aligned}\text{p-value} &= 2P(Z > 1.19)\\ &= 2(.1170)\\ &= .2340\end{aligned}$$

Example 11.2

COMPUTE

	A	B	C	D
1	**Z-Test: Mean**			
2				
3				*Bills*
4	Mean			17.55
5	Standard Deviation			3.94
6	Observations			100
7	Hypothesized Mean			17.09
8	SIGMA			3.87
9	z Stat			1.19
10	P(Z<=z) one-tail			0.1173
11	z Critical one-tail			1.6449
12	P(Z<=z) two-tail			0.2346
13	z Critical two-tail			1.96

Summary of One- and Two-Tail Tests…

One-Tail Test (left tail)	Two-Tail Test	One-Tail Test (right tail)
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
		

Developing an Understanding of Statistical Concepts

As is the case with the confidence interval estimator, the test of hypothesis is based on the sampling distribution of the sample statistic.

The result of a test of hypothesis is a probability statement about the sample statistic.

We assume that the population mean is specified by the null hypothesis.

Developing an Understanding of Statistical Concepts

We then compute the test statistic and determine how likely it is to observe this large (or small) a value when the null hypothesis is true.

If the probability is small we conclude that the assumption that the null hypothesis is true is unfounded and we reject it.

Developing an Understanding of Statistical Concepts

When we (or the computer) calculate the value of the test statistic

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

we're also measuring the difference between the sample statistic and the hypothesized value of the parameter.

The unit of measurement of the difference is the standard error.

Developing an Understanding of Statistical Concepts

In Example 11.2 we found that the value of the test statistic was z = 1.19. This means that the sample mean was 1.19 standard errors above the hypothesized value of.

The standard normal probability table told us that this value is not considered unlikely. As a result we did not reject the null hypothesis.

The concept of measuring the difference between the sample statistic and the hypothesized value of the parameter in terms of the standard errors is one that will be used frequently throughout this book

Probability of a Type II Error β

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: $\mu = 170$

H_1: $\mu > 170$

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:

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

Example 11.1 (revisited)

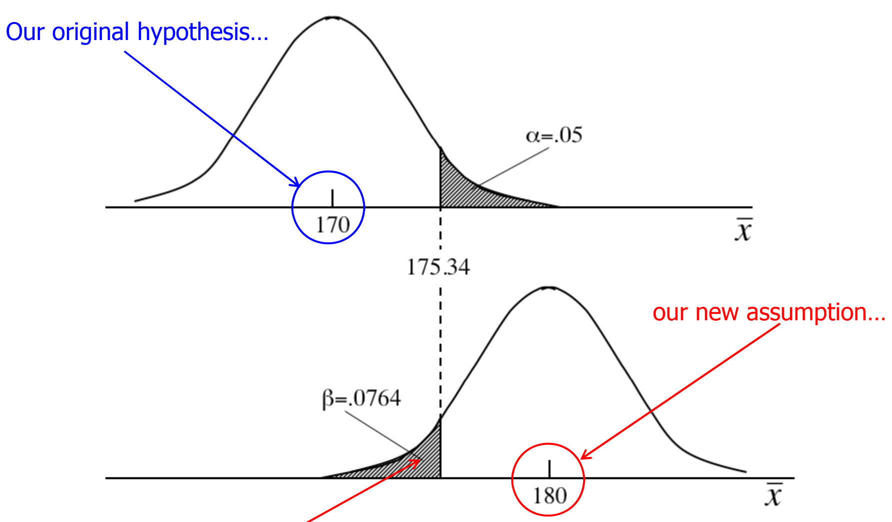
β = P($\bar{x}$ < 175.34 given that the null hypothesis is false)

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)



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

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

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

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

$$\overline{x} > 177.57$$

Effects on β of Changing α

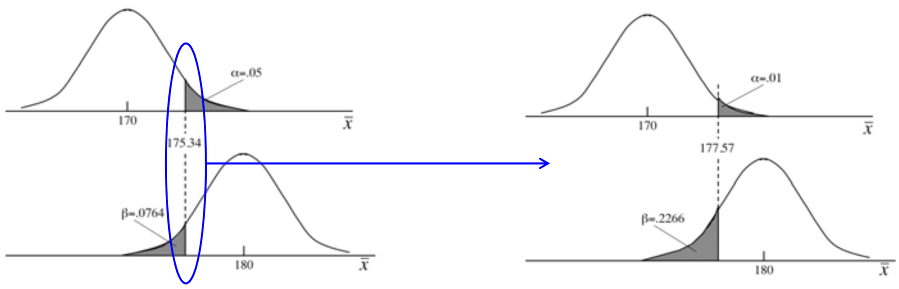
Stage 2 Probability of a Type II error

$$\beta = P(\bar{x} < 177.57 \mid \mu = 180)$$
$$= P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{177.57-180}{65/\sqrt{400}}\right)$$
$$= P(z < -.75)$$
$$= .2266$$

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

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

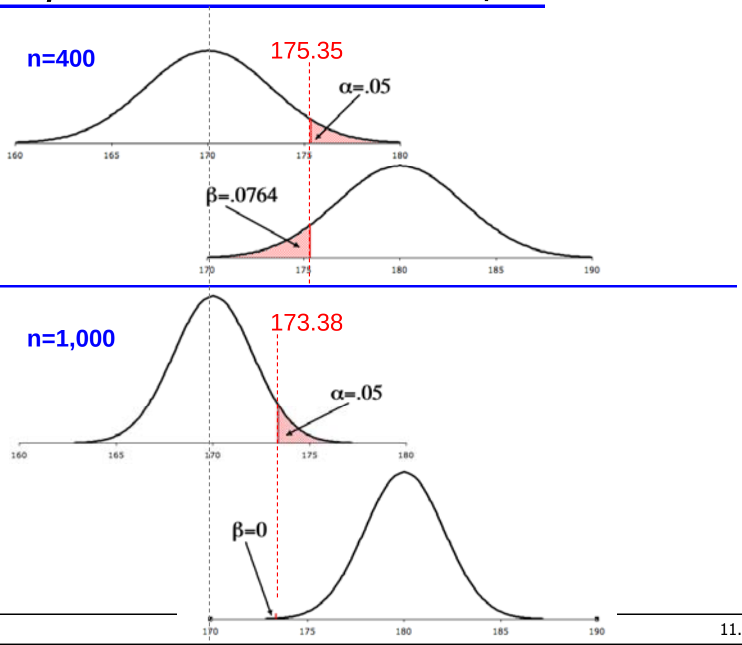
$$= P\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < \frac{173.38 - 180}{65 / \sqrt{1{,}000}}\right)$$

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

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

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."

Judging the Test

The ***power of a test*** is defined as $1-\beta$.

It represents the probability of rejecting the null hypothesis when it is false.

I.e. when more than one test can be performed in a given situation, its is preferable to use the test that is correct more often. If one test has a higher power than a second test, the first test is said to be more powerful and the preferred test.

SSA Example Calculating β

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

Recall that

$H_0{:}\mu = 22$

$H_1{:}\mu < 22$

$n = 220$

$\sigma = 6$

$\alpha = .10$

SSA Example Calculating β

Stage 1: Rejection region

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

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

$$\bar{x} < 21.48$$

SSA 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

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

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

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

$$= .8835$$