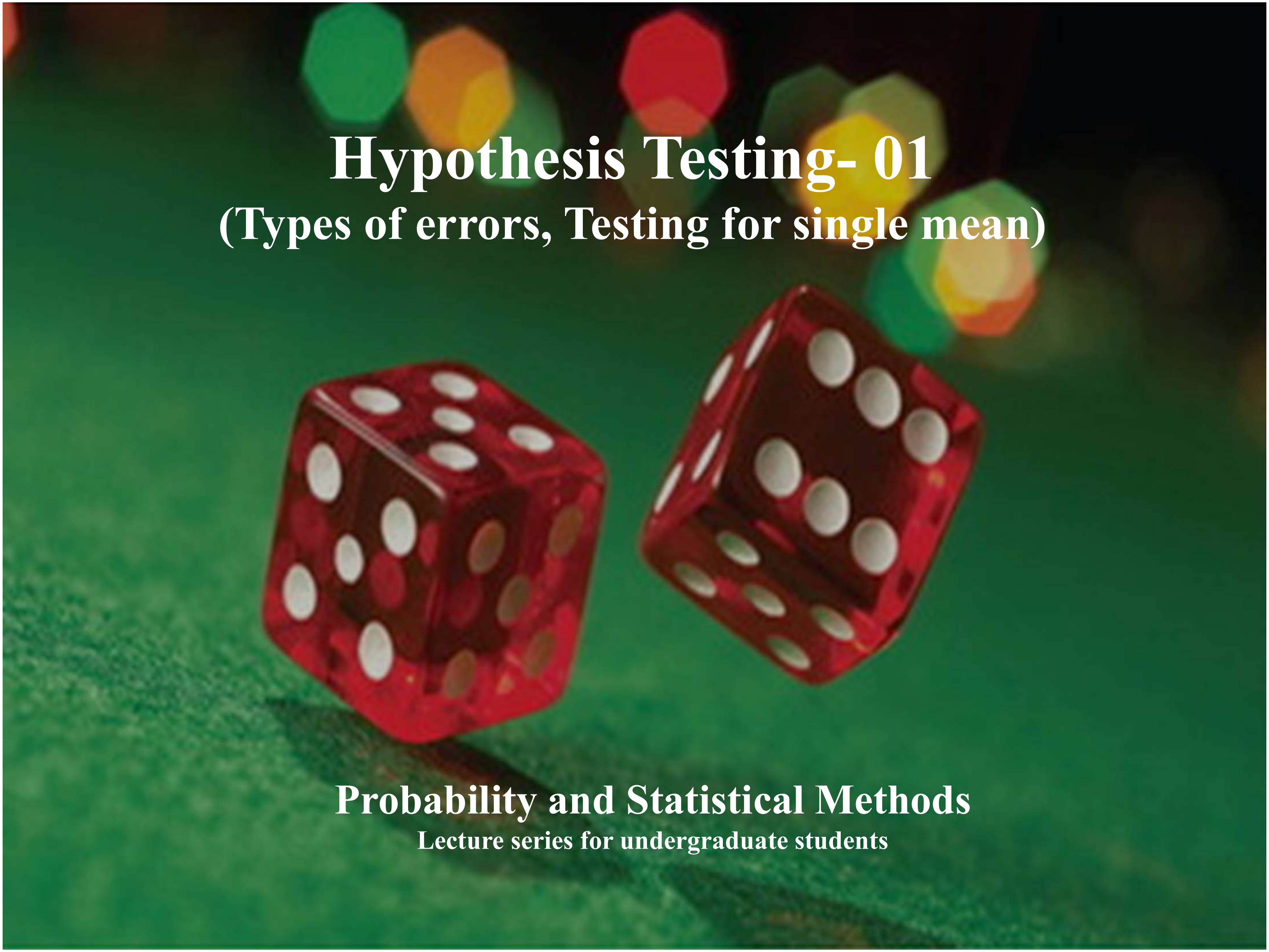


Hypothesis Testing- 01

(Types of errors, Testing for single mean)

Probability and Statistical Methods

Lecture series for undergraduate students

Introduction to Hypothesis Testing

Hypothesis testing is used to assess the plausibility of a hypothesis about some population characteristics by using sample statistic.

Total contents in this topic:

- Testing for one and two means, proportions and Variances.
- Chi-square test for independence, homogeneity and several proportions.
- Chi-square for Goodness-of-fit test.

Nonstatistical Hypothesis Testing…

- A criminal trial is an example of hypothesis testing. (A pers on is Innocent until proven guilty).
- 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. Jury 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 means, 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).

- **Or there is not enough evidence to convict the defendant (fail to accept 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.

Type-I error in hypothesis testing

- **<u>Note: H_0: Innocent and H_1: guilty.</u>**
- A Type I error occurs when we **<u>Reject a true null hypothesis (RT)</u>**.
- That is: the jury convicts an innocent person.
- P(defendant is guilty given that he was innocent) = P(Reject H_0 given that H_0 was true). (Say RT for memory)
- **P(Type I error) = α [usually 0.05 or 0.01]**

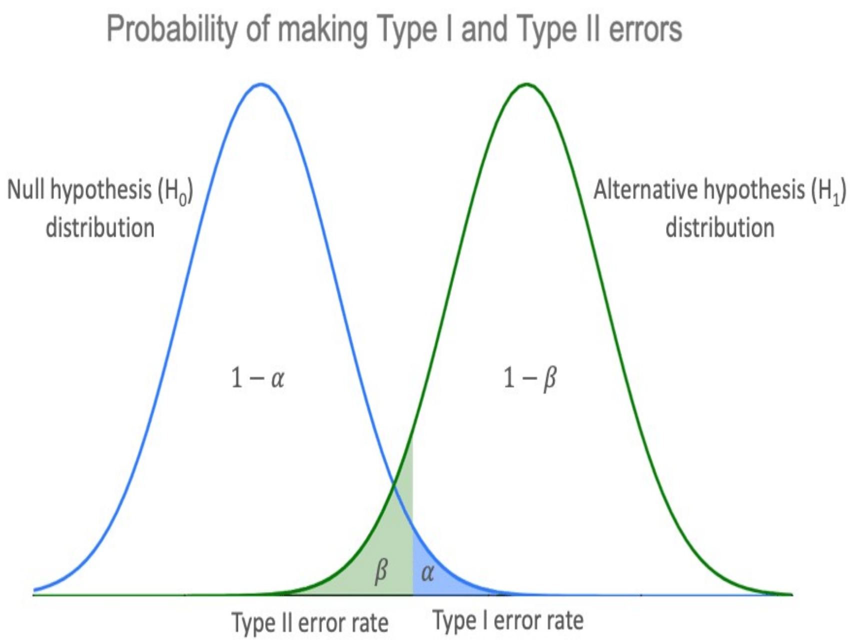
H_0: the defendant is innocent.

H_1: the defendant is guilty.

		Reality	
		Innocent	Guilty
Verdict	Innocent	Correct Verdict	Type II Error
	Guilty	Type I Error	Correct Verdict

Type-II error in hypothesis testing

- The jury fail to convict a guilty person.
- Type-II error occurs when we **accept a false null hypothesis (AF)**.
- P(defendant is innocent given that he was guilty) = P(Accept H_0 given that H_0 was False). (For memory call it AF)

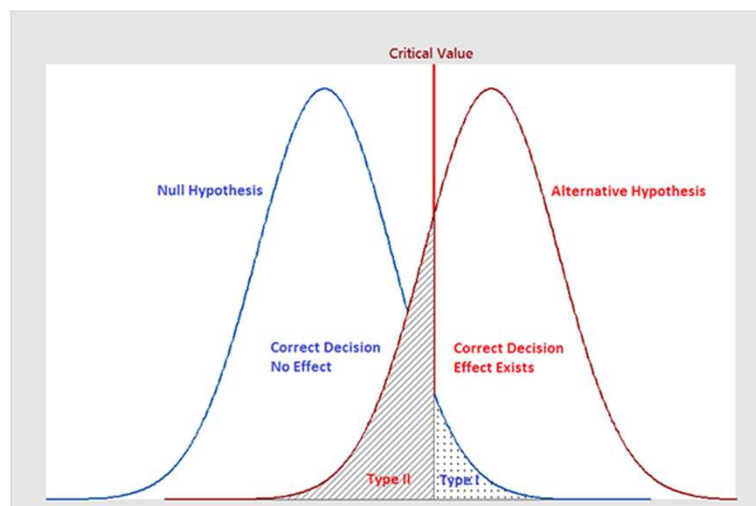


		Truth	
		H_o True	H_o False
Your Findings	H_o True	Correct	Type II Error (β)
	H_o False	**Type I Error (α)**	Correct

Type I and II errors, Power and Confidence

- P(Type I error) = α = P[Reject H0 | Ho is true] (say RT)
- 1 - α = P[Accept H0 | Ho is true]
- P(Type II error) = β = P[Fail to reject H0 | Ho is false] (say AF)
- Power = 1 – β = P[Reject H0 | Ho is False]

		Reality	
		Positive	Negative
Study Finding	Positive	**True Positive** (Power) (1-β)	False Positive **Type I Error** (α)
	Negative	False Negative **Type II Error** (β)	**True Negative**



Relationship between type-I and type-II errors

- The two probabilities are inversely related. Decreasing one increases the other, for a fixed sample size.
- Both decrease when sample size increases.

Type I and Type II Error Revisited

DECISION	NULL HYPOTHESIS: Actually True	NULL HYPOTHESIS: Actually False
Fail to Reject	☺ $1-\alpha$	Type II error β
Reject	Type I error α	☺ $1-\beta$

Either type error is undesirable and we would like both α and β to be small.

Type-II error is more serious

- In practice, this type of error is by far the most serious mistake we normally make.

- For example, if we test the hypothesis that the amount of medication in a heart pill is equal to a value which will cure your heart problem and “accept the hull hypothesis that the amount is ok”. Later on we find out that the average amount is WAY too large and people die from “too much medication” [I wish we had rejected the hypothesis and threw the pills in the trash can], it’s too late because we shipped the pills to the public.

Main steps in Hypothesis Testing…

The critical concepts are theses:

1. There are two hypotheses, the null (H_0: — *the 'null' hypothesis*) and the alternative hypothesis (H_1: *the 'research' hypothesis*).
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, or the null is not likely to be true.
4. There are two possible decisions:

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

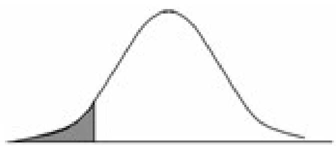
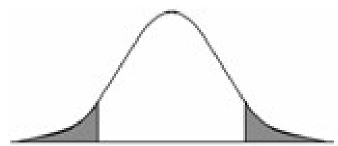
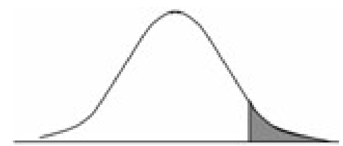
   Conclude that there is *not* enough evidence to support the alternative hypothesis. Fail to reject the null.

Steps in Hypothesis Testing

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

The significance level (α) is the type- I (the test statistic will fall in the critical region when the null hypothesis is actually true). Common choices for α are 0.05, 0.01, and 0.10.

Hypothesis Testing for single Population mean

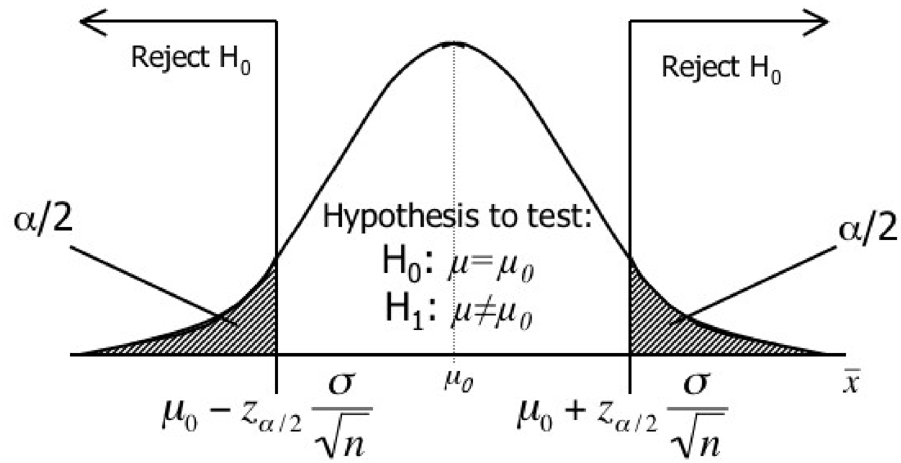
One-Tail Test (left tail)	Two-Tail Test	One-Tail Test (right tail)
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$
		

- **A critical value** is any value separating the critical region (where we reject the H_0) from the region of Accept H_0. For example, the critical value of z = 1.645 corresponds to a α = 0.05.
- The ***rejection region*** is a range of values such that if the test statistic falls into that range, we decide to reject the null hypothesis in favor of the alternative hypothesis.

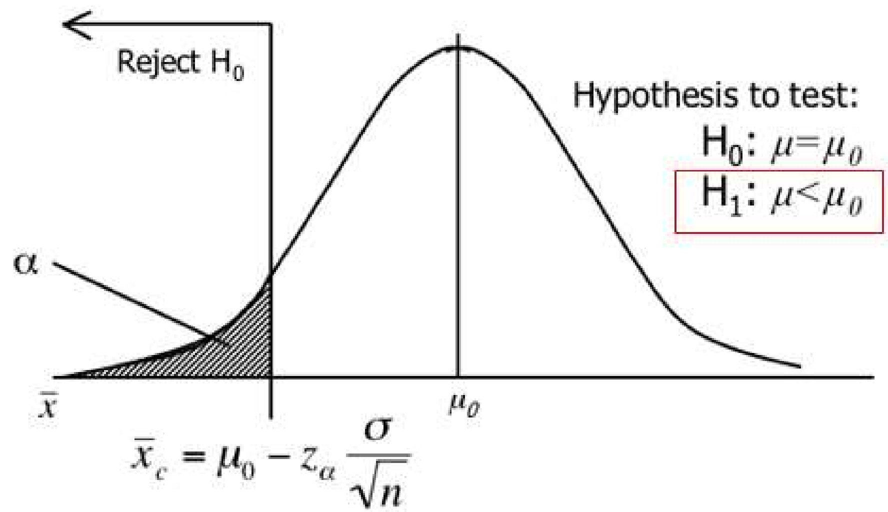
One and two tail tests

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

The rejection region is split equally between the two tails.



$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$



Six step procedure for hypothesis test

- Define $\mathbf{H_0}$: $\boldsymbol{\mu = \mu_0}$
- Define $\mathbf{H_1}$: $\boldsymbol{\mu < \mu_0}$ or $\mathbf{H_1}$: $\boldsymbol{\mu > \mu_0}$ or $\mathbf{H_1}$: $\boldsymbol{\mu \neq \mu_0}$
- Calculate test statistic value:

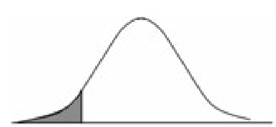
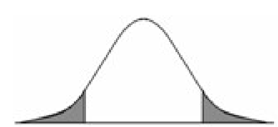
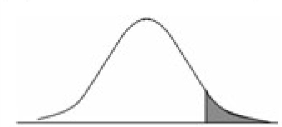
$$\boldsymbol{z_{cal}} = \frac{\bar{\boldsymbol{x}} - \boldsymbol{\mu_0}}{\boldsymbol{\sigma}/\sqrt{\boldsymbol{n}}} \textbf{ or } t_{cal} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \text{ depending on given information.}$$

- Compute critical values:
  a) If z test is applied in step-04: then
     - *a)* $z_{tab} = -z_\alpha$ (for left tailed test)
     - *b)* $z_{tab} = +z_\alpha$ (for right tailed test)
     - *c)* $z_{tab} = -z_{\alpha/2}$ *and* $+z_{\alpha/2}$(for two tailed test)

  b) If t-test is applied in step-04: then
     - *a)* $t_{tab} = -t_{(\alpha,n-1)}$(for left tailed test)
     - *b)* $t_{tab} = +t_{(\alpha,n-1)}$ (for right tailed test)
     - *c)* $t_{tab} = -t_{(\alpha/2,n-1)}$ *and* $+t_{(\alpha/2,n-1)}$(for two tailed test)

Six step procedure for hypothesis test

- **Decision step:** Compare statistic and table critical values.
  - For left tailed test: Reject H0 iff: $\boldsymbol{z_{cal}} < -z_{tab}$
  - For right tailed test: Reject H0 iff: $\boldsymbol{z_{cal}} > z_{tab}$
  - For two tailed test: Reject H0 iff: $\boldsymbol{z_{cal}} < -z_{tab}$ or $\boldsymbol{z_{cal}} > z_{tab}$
- The same is true for t-test conditions:
  - For left tailed test: Reject H0 iff: $\boldsymbol{t_{cal}} < -t_{tab}$
  - For right tailed test: Reject H0 iff: $\boldsymbol{t_{cal}} > t_{tab}$
  - For two tailed test: Reject H0 iff: $\boldsymbol{t_{cal}} < -t_{tab}$ or $\boldsymbol{t_{cal}} > t_{tab}$
- **Write conclusions**

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

Example-01

A departmental store manager determines that a new billing 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 accounts are approximately normally distributed with a standard deviation of $65.

Can we conclude that the new system will be cost-effective?

Given: Sample size = n = 400, sample mean = $\bar{x}$ =178 and $\sigma = 65$

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

Step-01: H_0: $\mu = 170$ (this specifies a single value for the parameter of interest) – Actually H_0: $\mu \leq 170$

Step-02: H_A: $\mu > 170$.

Example-01 (Conti…)

Step-03: As σ is know, we can use z –statistic:

$$z_{cal} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{178 - 170}{65/\sqrt{400}} = 8 \times \frac{20}{65} = 2.46$$

Step-04: Critical value: This is right tailed test: Thus,

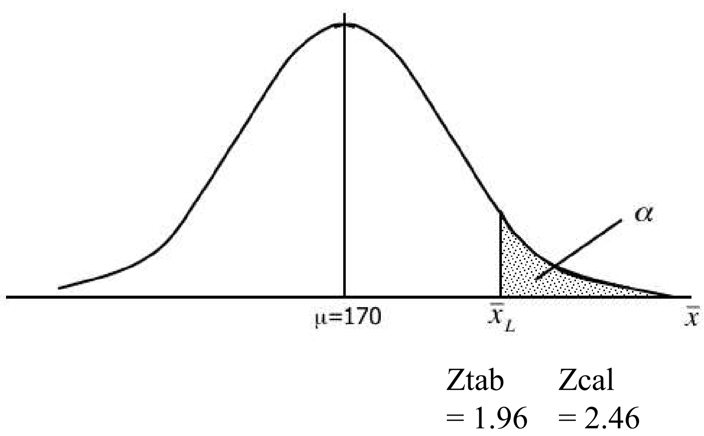
$$z_{tab} = z_{\alpha} = z_{0.05} = +1.96.$$

Step-05: Comparison: As $z_{cal} > z_{tab}$, z_{cal} lies inside critical region.

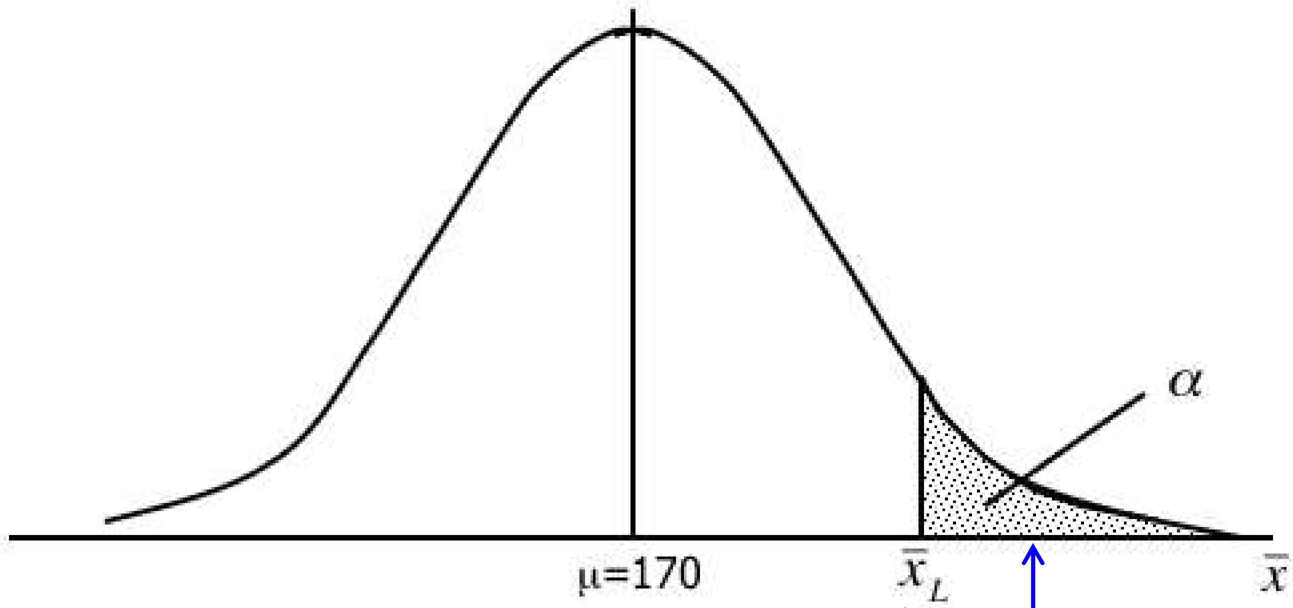
Step-06: We cannot reject the null hypothesis. The z score of 2.46 is within the nonrejection area.

P-Value approach:

p-value = P($\bar{x} \geq 178$) = P($Z \geq 2.46$)
= 0.0069 < 0.05   Reject null.



Example… The Big Picture…



$H_1: \mu > 170$

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

$\bar{x}_L = 175.34$

$\bar{x} = 178$

Reject H_0 in favor of

P-value of a hypothesis

P-Value

The **P-value** of a hypothesis test is the probability of getting sample data at least as inconsistent with the null hypothesis (and supportive of the alternative hypothesis) as the sample data actually obtained.[†] We use the letter ***P*** to denote the *P*-value.

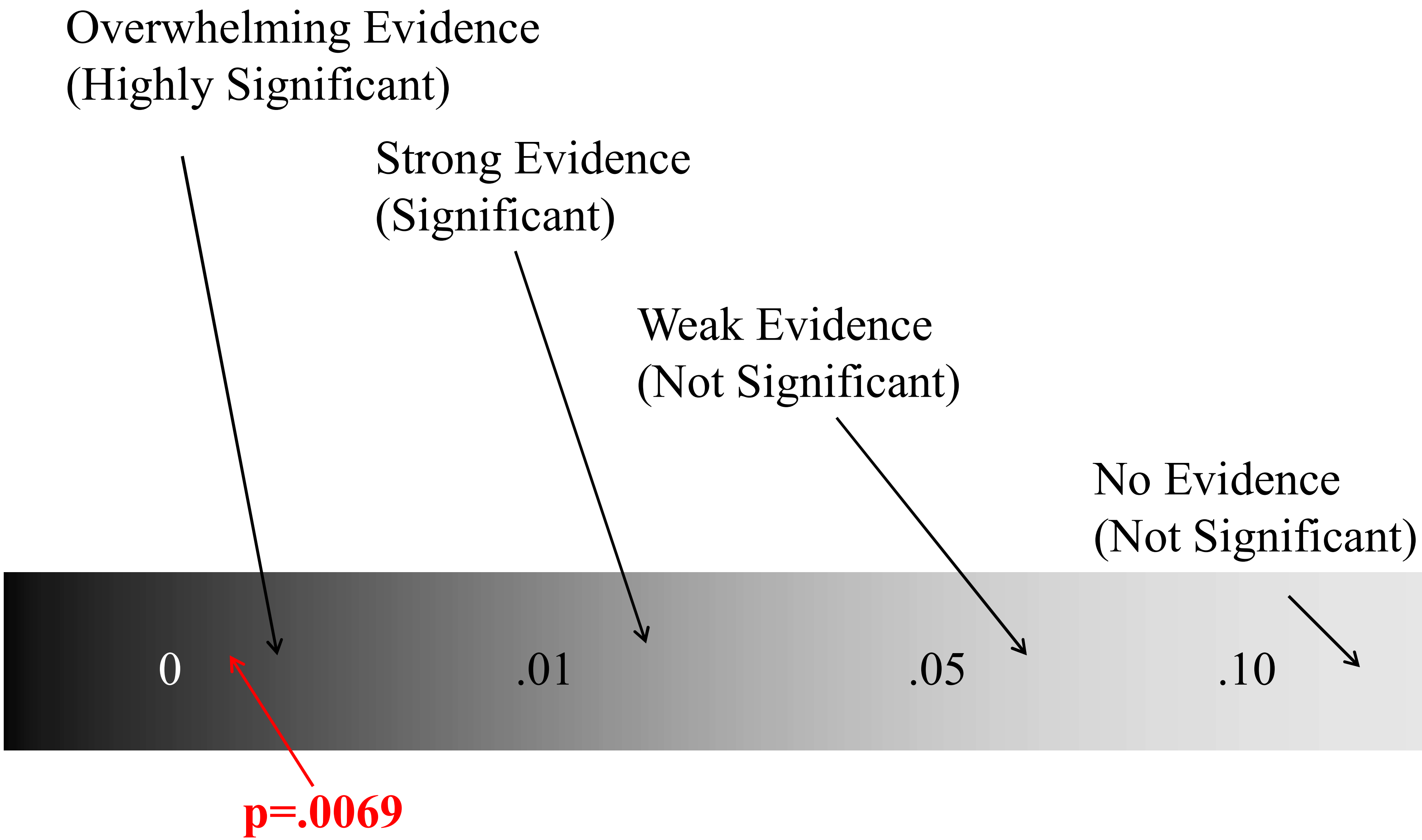
- The smaller (closer to 0) the P-value, the stronger is the evidence against the H_0 and, hence, in favor of the H_A.
- Stated simply, an outcome that would rarely occur if the H_0 were true provides evidence against the H_0 and, hence, in favor of the H_A.
- We can calculate P-value by finding the value of α in a Z or t table using the calculated Z or t value.

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.

Interpreting the p-value…



Example-02 Using P-value approach

A prescription allergy medicine is supposed to contain an average of 245 parts per million (ppm) of active ingredient. The manufacturer periodically collects data to determine if the production process is working properly. A random sample of 64 pills has a mean of 250 ppm with a standard deviation of 12 ppm.

- Let μ denote the average amount of the active ingredient in pills of this allergy medicine. The null and alternative hypotheses are H_0: $\mu = 245$, H_a: $\mu \neq 245$. The level of significance is 1%.
- The z-test statistic is 3.33 with a P-value of 0.0014. What is the correct conclusion?

- As the P-value of the test is lesser than the level of significance, therefore, we can conclude to reject H_0 in favor of H_A.

t- table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
					Confidence Level						

Exercise-:

The Federal Pell Grant Program gives grants to low-income undergraduate students. According to the Study by the U.S. Department of Education in 2008, the average Pell grant award for 2007-2008 was $2,600. We wonder if the mean amount is different this year for Pell grant recipients at San Jose State University. Suppose that we randomly select 50 Pell grant recipients from this University. For these 50 students, the mean Pell grant award is $2,450 with a standard deviation of $600.

- Let μ = the mean amount of Pell grant awards received by San Jose State University Pell grant recipients this year.
- H_0: $\mu = 2{,}600$, H_a: $\mu \neq 2{,}600$.
- Here n > 30, so a z-model is a good fit for the sampling distribution. Sample mean= 2450 and sample SD= 600.

Question: Compute z-test statistic value and z-table (critical value). What will be your decision based on given sample evidence? What is the P-value? (Find probability using Z_{cal} value).

Question: P-value computation

Suppose we're conducting a t-test for a population mean with the hypotheses: H_0: $\mu = 50$, H_a: $\mu \neq 50$. A sample of size 28 has a t-statistic of -2.43. Find the corresponding P-value.

- Here, df= n - 1= 27. Corresponding to the row of 27 (or nearest one), find where the t_{cal} value exists.
- We can find nearest P-value. For accurate answer, we need to use software.

Example 2… Students work

AT&T's argues that its rates are such that customers won't see a difference in their phone bills between them and their competitors. They calculate the mean and standard deviation for all their customers at $17.09 and $3.87 (respectively).

Note: Don't know the true value for σ, so we estimate σ from the data [$\sigma \sim s = 3.87$] – large sample so don't worry.

They then sample 100 customers at random and recalculate a monthly phone bill based on competitor's rates.

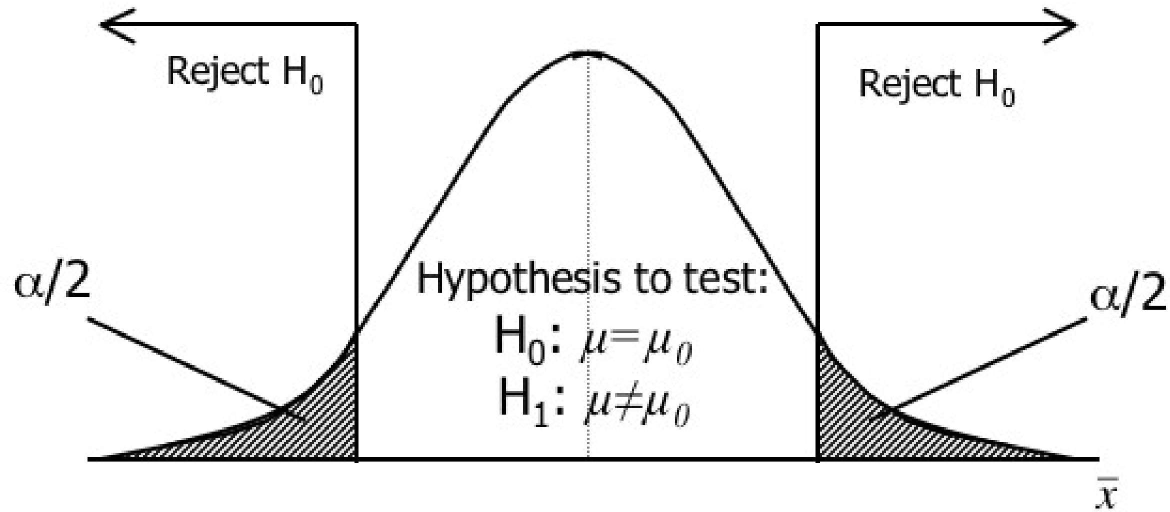
Our null and alternative hypotheses are

$H_1: \mu \neq 17.09$. We do this by assuming that:

$H_0: \mu = 17.09$

Example 2…

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



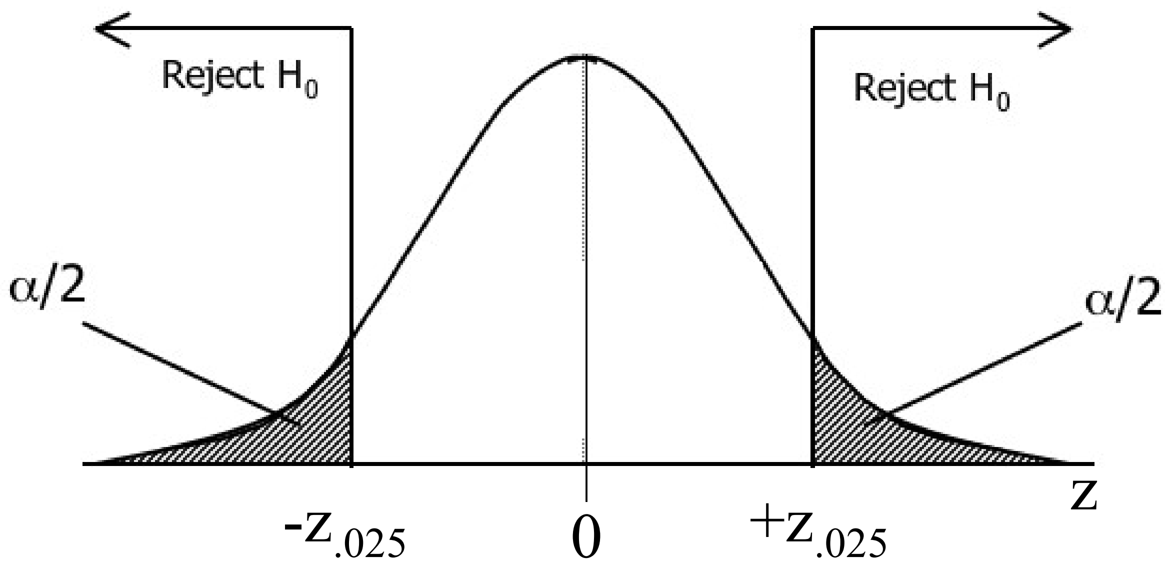
stat is "small" stat is "large"

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

Example 2…

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 …

From the given data, 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."***

Probability of a Type II Error – β

A Type II error occurs when a false null hypothesis is not rejected or "you accept the null when it is not true" but don't say it this way if a statistician is around.

In practice, this is by far the most serious error you can make in most cases, especially in the "quality field".

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 too large [we have insufficient power], we can reduce it by

increasing α , and/or

increasing the sample size, n.

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 and the true mean is something other than the null value for the mean.

If we are testing the hypothesis that the average amount of medication in blood pressure pills is equal to 6 mg (which is good), and we "fail to reject" the null hypothesis, ship the pills to patients worldwide, only to find out later that the "true" average amount of medication is really 8 mg and people die, we get in trouble. This occurred because the P(reject the null / true mean = 7 mg) = 0.32 which would mean that we have a 68% chance on not rejecting the null for these BAD pills and shipping to patients worldwide.

Probability you ship pills whose mean amount of medication is 7 mg approximately 67%