NNSE 784
Advanced Analytics Methods

Instructor: F Doyle (CESTM L210)

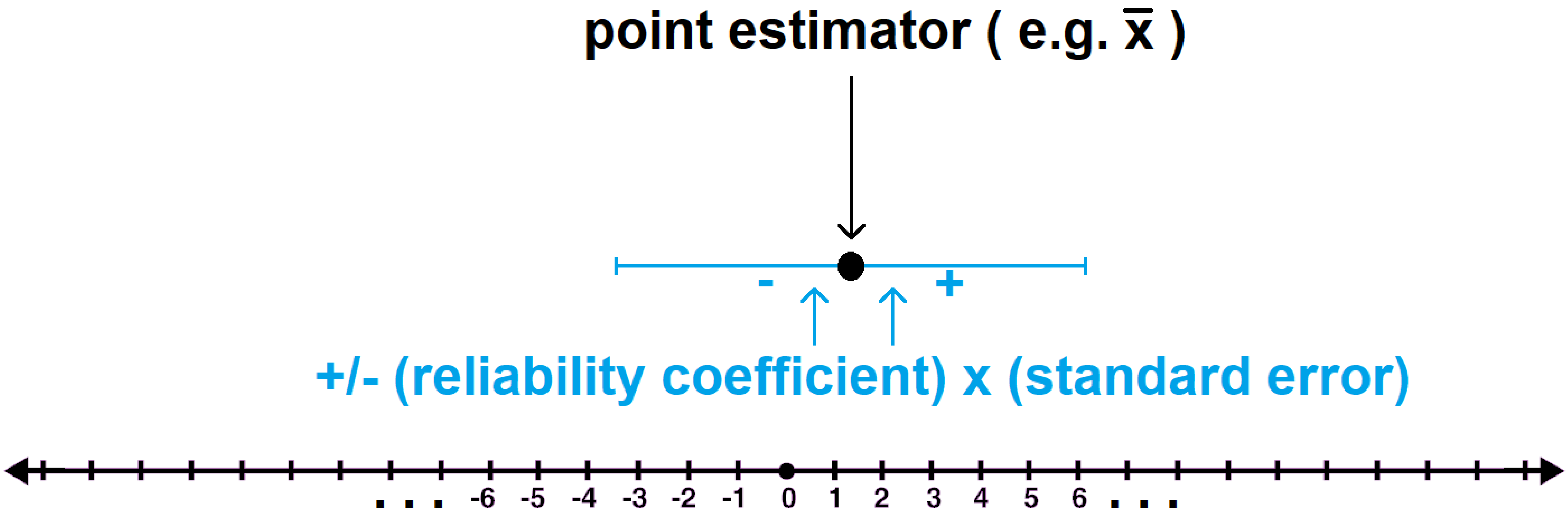
MW 4:30 – 5:50, NFN 203

Slide Set #10
Inferential Statistics: Hypothesis Testing

Lecture Outline

- Interval estimate recap
- Discuss variance estimation
- Hypothesis testing overview

Interval Estimate - recap



Example: the interval estimate for $\mu = \quad \bar{x} \pm t_{(1-\alpha/2)} \frac{s}{\sqrt{n}}$

Confidence Interval for the Variance of a Normally Distributed Population

		Second Draw				
		6	8	10	12	14
	6	0	2	8	18	32
	8	2	0	2	8	18
First Draw	10	8	2	0	2	8
	12	18	8	2	0	2
	14	32	18	$\boxed{8}$	2	0

$$s^2 = \sum (x_i - \bar{x})^2/(n-1) = \frac{(14 - 12)^2 + (10 - 12)^2}{2 - 1} = 8$$

Sample variance

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = 8$$

$$S^2 = \frac{\sum (x_i - \mu)^2}{N-1} = 10$$

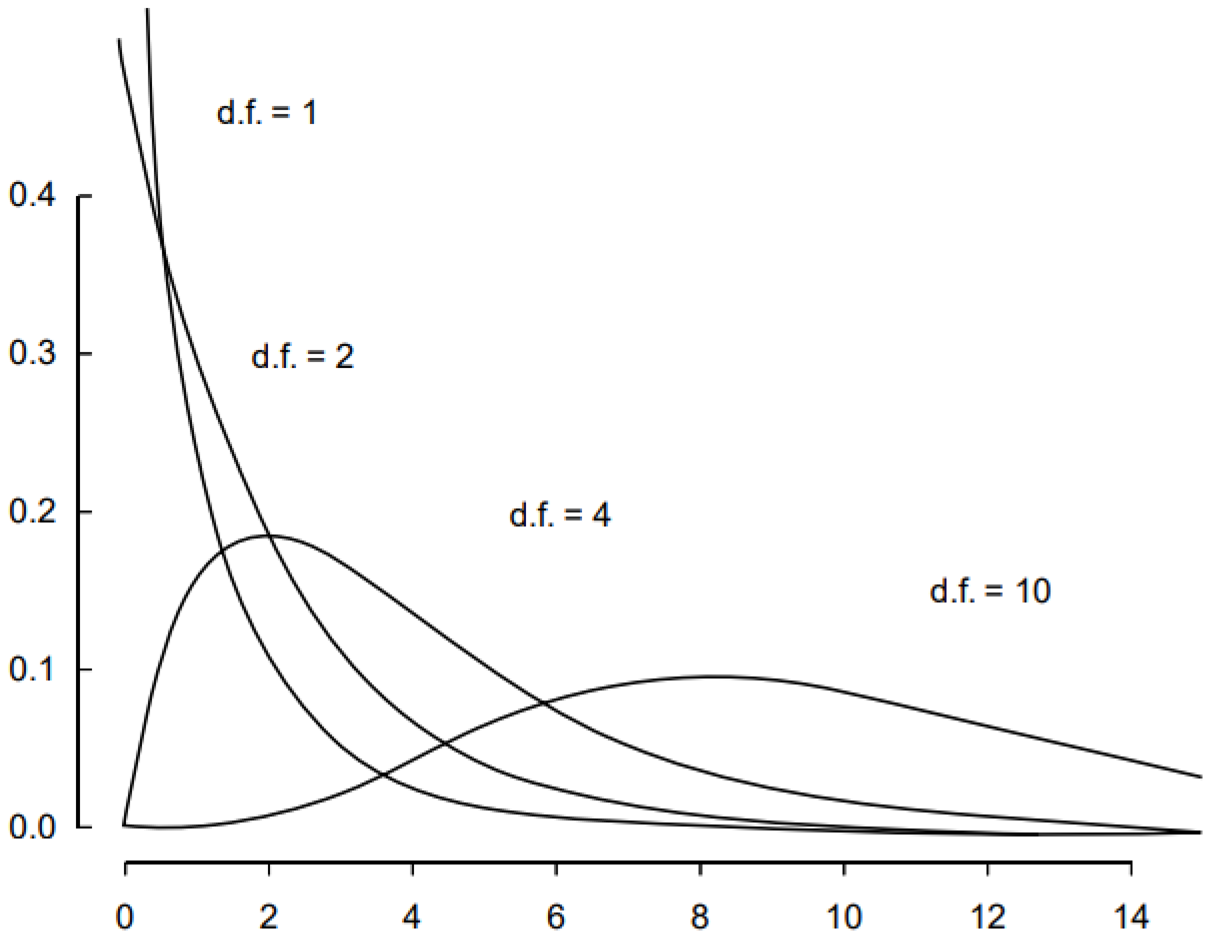
S^2 is a different <u>parameter</u> for dispersion.

$E(s^2) = \sigma^2$ *when sampling is with replacement*

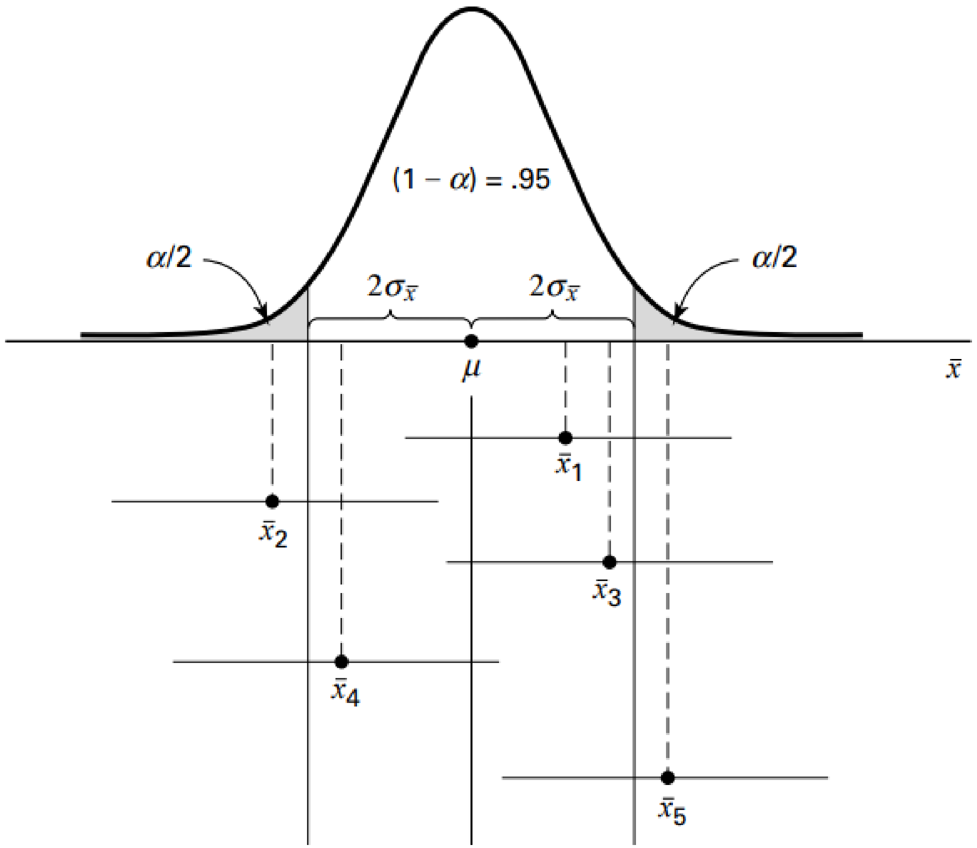
$E(s^2) = S^2$ *when sampling is without replacement*

Chi-square(χ^2) Distribution

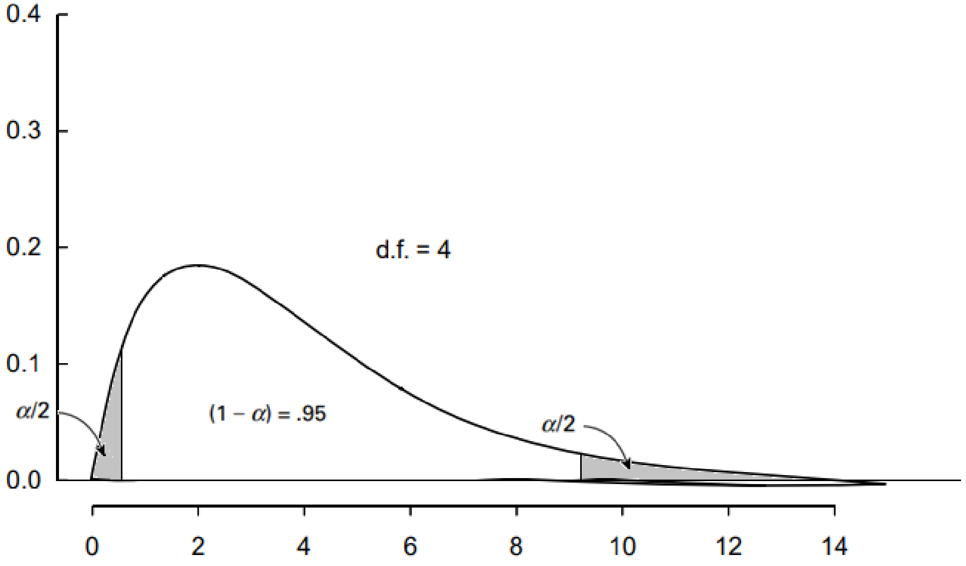
$$(n-1)s^2/\sigma^2$$



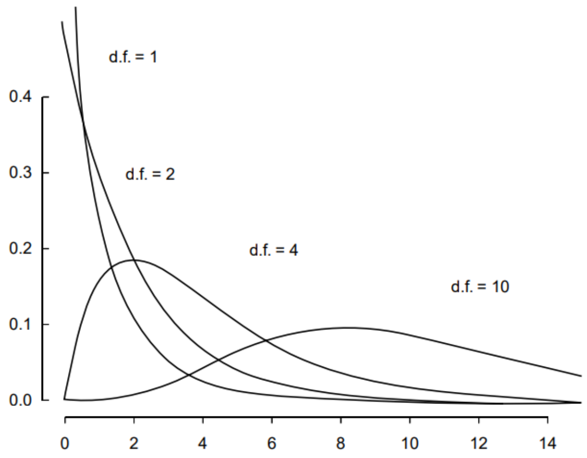
We're Still Talking About Probability Distributions!



vs



100(1-α)Confidence Interval for σ^2 Based on χ^2



Not symmetrical!

$$\chi^2_{\alpha/2} < \frac{(n-1)\,s^2}{\sigma^2} < \chi^2_{1-(\alpha/2)}$$

$$\Downarrow$$

$$\frac{\chi^2_{\alpha/2}}{(n-1)s^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{1-(\alpha/2)}}{(n-1)s^2}$$

$$\Downarrow$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} > \sigma^2 > \frac{(n-1)s^2}{\chi^2_{1-(\alpha/2)}}$$

$$\Downarrow$$

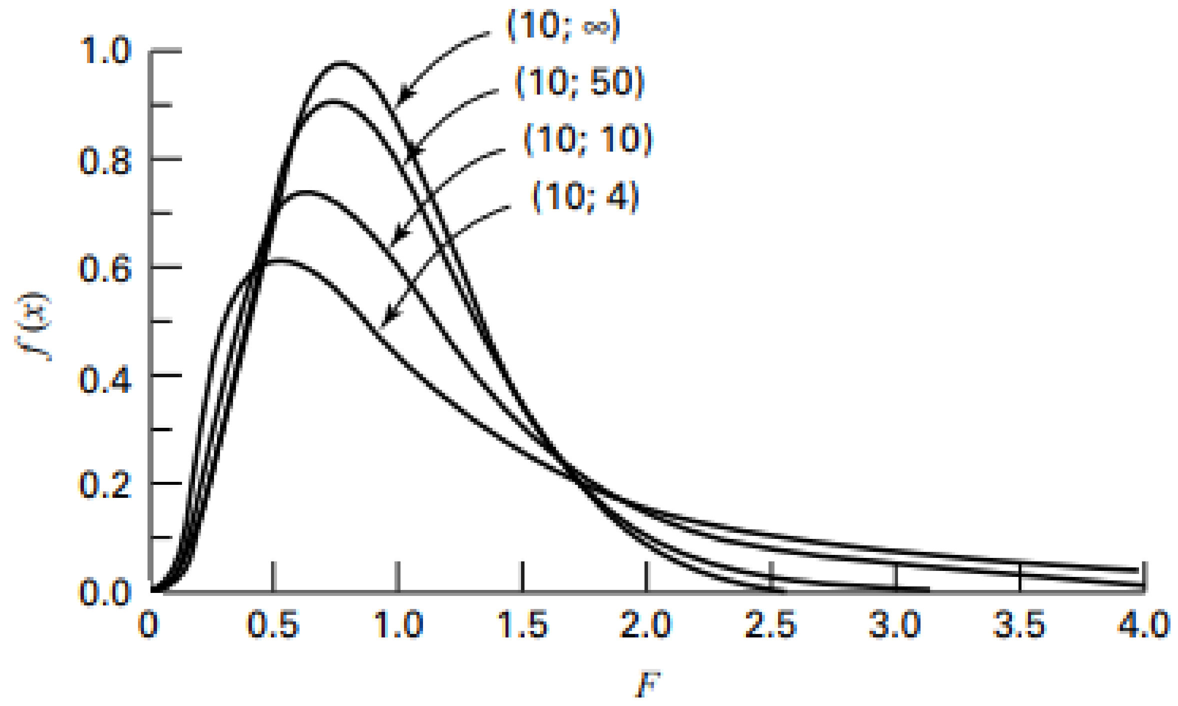
$$\frac{(n-1)s^2}{\chi^2_{1-(\alpha/2)}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

$$\Downarrow$$

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{1-(\alpha/2)}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}}$$

Comparing Variance of Two Populations
The F Distribution

$(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$



Hypothesis Testing Concepts

Hypothesis – a statement about one or more populations

Research Hypothesis – conjecture or supposition that motivates the research

Statistical Hypotheses – hypotheses that are stated in a manner that they may be evaluated via appropriate statistical techniques

Hypothesis Testing Steps

1. Data –
2. Assumptions – for example, do we assume the underlying data is normally distributed?
3. Hypotheses – there are two statistical hypotheses involved in hypothesis testing
    1. The **null hypothesis H_O** – hypothesis to be tested. Presumed to be true. Set up for the express purpose of being discredited and is consequently the complement of the conclusion the researcher is pursuing.
    2. The **alternative hypothesis H_A** - If the testing procedure leads to rejection of the null hypothesis, we say that the data are not compatible with the nll hypothesis but are supportive of some alternative hypothesis. Usually the *alternative hypothesis* and the *research hypothesis* are the same and the terms are typically used interchangeably

Rules For Stating Statistical Hypotheses

- For the sake of simplicity we will describe those types that use an indication of equality (e.g., value of the mean, etc.)

For example:

For a null hypothesis of:	The alternative hypothesis would be:
H_O: $\mu = 60$	H_A: $\mu \neq 60$
H_O: $\mu \leq 60$	H_A: $\mu > 60$
H_O: $\mu \geq 60$	H_A: $\mu < 60$

Hypothesis Testing Steps - continued

4. Test statistic – some statistic that may be computed from the data of the sample. For example:

$$z = \frac{\bar{x} - \boxed{\mu_0}}{\sigma/\sqrt{n}}$$

Hypothesized value of a population mean (e.g., 60 in previous slide hypotheses)

which is related to the familiar:

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

General Formula for Test Statistic:

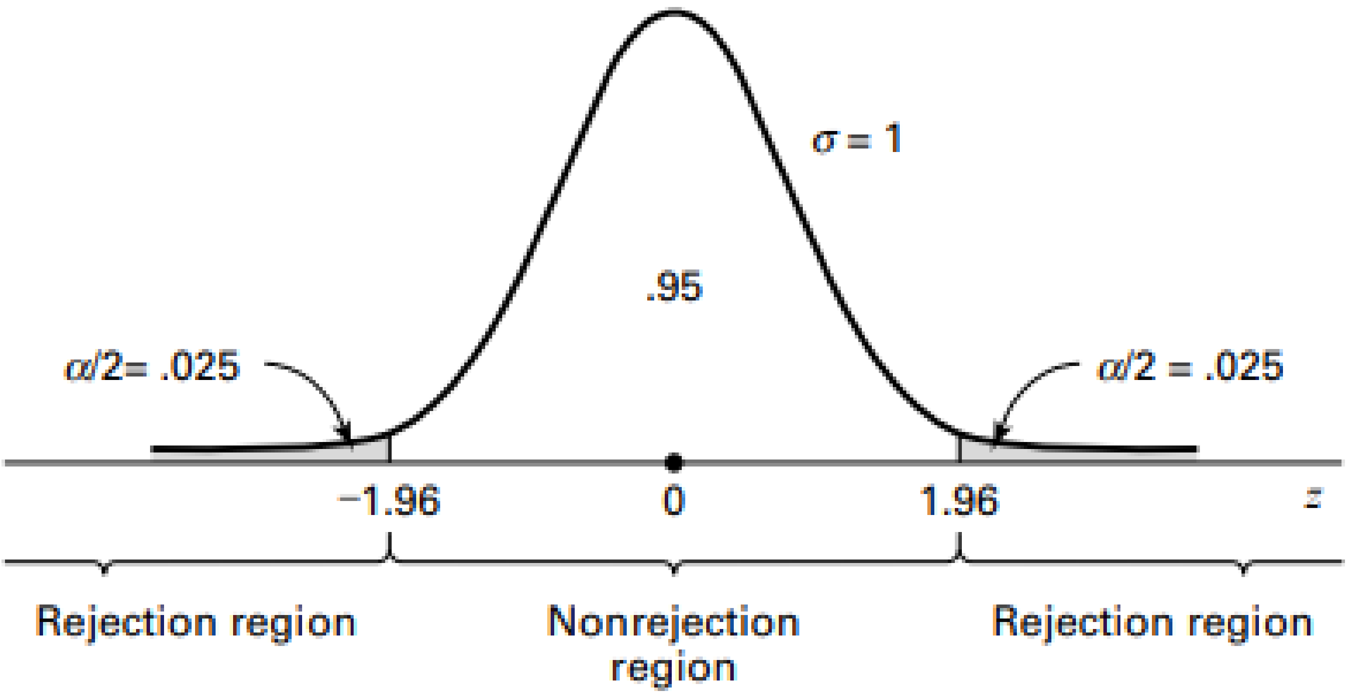
$$\text{test statistic} = \frac{\text{relevant statistic} - \text{hypothesized parameter}}{\text{standard error of the relevant statistic}}$$

Hypothesis Testing Steps - continued

5. Distribution of test statistic – understand what the distribution of the test statistic you are using looks like, particularly given the null hypothesis. For example, the distribution of $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ will follow the standard normal distribution if the null hypothesis is true and the assumptions are met.

Hypothesis Testing Steps - continued

6. Decision rule – all possible values that the test statistic can assume are points on the horizontal axis of the distribution of the test statistic and are divided into two groups:
   - Rejection region
   - Nonrejection region



Hypothesis Testing Steps - continued

7. Calculation of test statistic – from the data collected in sample (e.g., set of experiment replicates), compute the value of the test statistic and compare it with the rejection and non-rejection regions we have specified.

Hypothesis Testing Steps - continued

8. Statistical decision – consists of rejecting or not rejecting the null hypothesis
9. Conclusion – if H_O is rejected, we conclude that H_A is true. If H_O is not rejected, we conclude that H_O may be true.
10. p-values – this is a number that tells us how the probability of having obtained our sample results, if the null hypothesis is true. An extremely low number provides justification for doubting the truth of the null hypothesis.

Type of Errors

- **Type I Error -** rejecting a true null hypothesis. Probability of this is equal to α.
- **Type II Error** - fail to reject a false null hypothesis. Probability of this is designated as β. Generally, we do not know the value of β, but it is usually larger than α.

		Condition of Null Hypothesis	
		True	False
Possible Action	Fail to reject H_0	Correct action	Type II error
	Reject H_0	Type I error	Correct action

Hypothesis Testing Flowchart