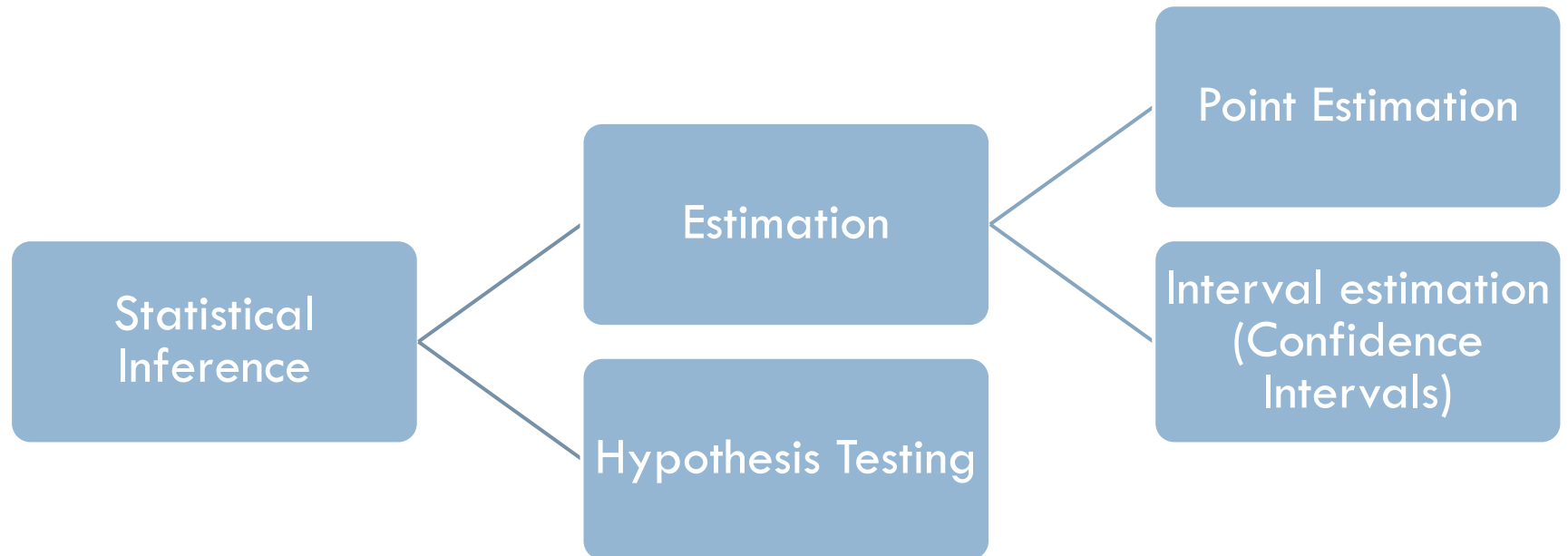


HYPOTHESIS TESTING

STATISTICAL INFERENCE

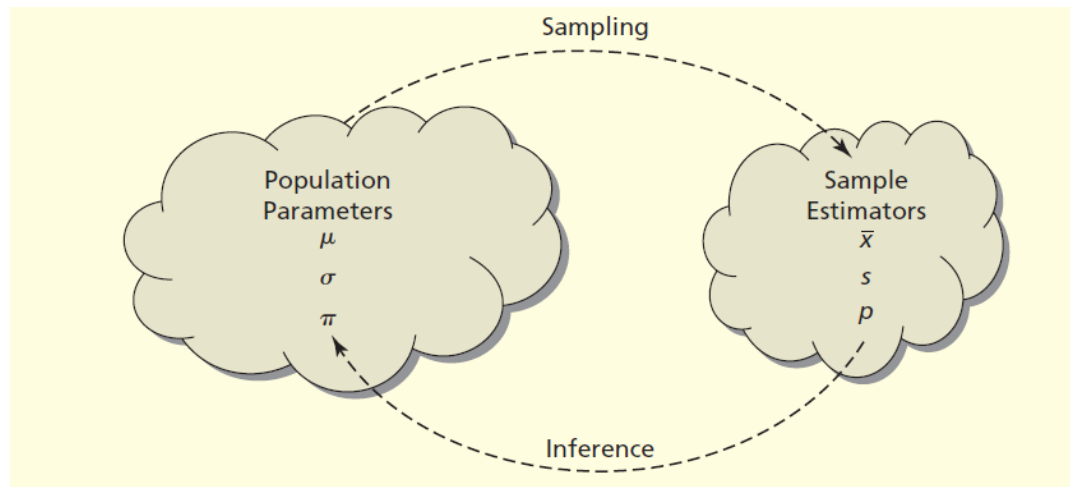
Pau Fonseca, Phd
Nihan Acar-Denizli, Phd

Statistical Inference



Statistical Inference

- An estimator is statistic derived from a sample to infer the value of a population parameter.
- An estimate is the value of the estimator in a particular sample.



Sampling Error

- The **sampling error** is the difference between an estimate and the corresponding population parameter and it is defined by,

$$\text{Sampling Error} = \bar{X} - \mu$$

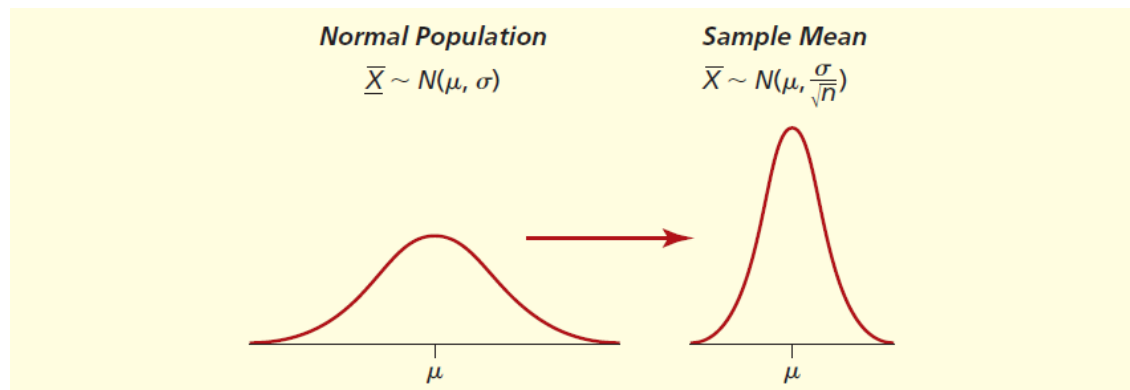
- Usually the parameter is unknown so the sampling error can not be computed.

Central Limit Theorem

- ❑ If the population is exactly normal, the sample mean has exactly a normal distribution centered at μ with the standard deviation,

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

The Central Limit Theorem (CLT) is a statistical theory that posits that the mean and standard deviation derived from a sample, will accurately approximate the mean and standard deviation of the population the sample was taken from as the size of the sample increases.



Confidence Intervals

- **Confidence interval** is an interval estimate by specifying the probability that a given interval will contain μ .

In frequentist statistics, a confidence interval (CI) is a range of estimates for an unknown parameter

- If the sample is large enough that \bar{X} approximately is normal by the Central Limit Theorem then the probability that the interval contains the true mean is $1 - \alpha$.

Confidence Level

- The probability $1 - \alpha$ is usually expressed as a percentage called **the confidence level**.
- The confidence level gives the probability that an interval constructed in this manner will contain μ .

Confidence Intervals for the Mean

- Confidence Interval of a Mean when the Variance is Known:

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

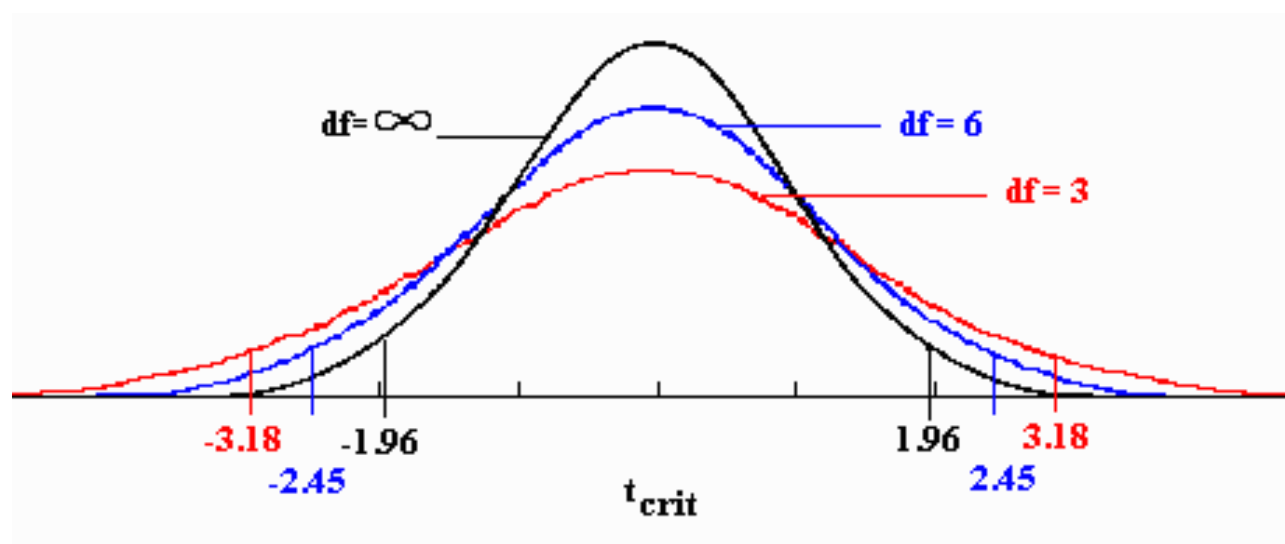
Confidence Intervals for the Mean

- Confidence Interval of a Mean when the Variance is Unknown:

$$P\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

The Standard Normal (z) and t distributions



Degrees of Freedom

- Degrees of freedom indicates the number of sample values that are free to vary.
- If a sample contains n number of observations, then the degrees of freedom is, $df = n - 1$.



Hypotheses

- **Null Hypothesis (H_0):** The statements based on similarity or equality. In scientific research, the null hypothesis (often denoted H_0) is the claim that no relationship exists between two sets of data or variables being analyzed.
- represents a research claim that is liked to be proved.
- always contains an equal sign ($=, \leq, \geq$).
- expresses the belief or assumption about the value of the population parameter.

Hypotheses

- **Alternative Hypothesis (H_a):** The statements based on differences.
 - represents the assumption that the test is attempting to demonstrate.
 - never contains an equal sign regarding the specified value of the population parameter ($\neq, <, >$).
 - expresses the opposite of the belief or assumption in H_0 .

Hypotheses

- The average connection speed is 54 Mbps, as claimed by the internet service provider.
- The average number of users increased by 2000 this year.
- The service times have Normal distribution.

Hypotheses

- The average connection speed is 54 Mbps, as claimed by the internet service provider.

$$H_0: \mu = 54 \qquad H_a: \mu \neq 54$$

- The average number of users exceed 2000 this year.

$$H_0: \mu \leq 2000 \qquad H_a: \mu > 2000$$

- The service times have Normal distribution.

H_0 : The service times follow a Normal distribution.

H_a : The service times does not follow a Normal distribution.

Some Important Definitions

- **Level of Significance (α)** : The probability of rejecting the null hypothesis when it is true. It is also known as Type I error.

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

- **Test statistic** : A statistic value which helps to determine whether a null hypothesis should be rejected.
- **Critical value** : The critical value is the largest value of the test statistic that result in the rejection of the null hypothesis.

Some Important Definitions

- **p value (p)** : It is a probability that provides a measure of evidence against H_0 .
- It answers the question : "If the null hypothesis were true, then what is the probability of observing a test statistic at least as extreme as the one we observed?"
- The smaller p-values indicate more evidence against H_0 .

Type of errors

	H_0 is True	H_0 is False
Reject H_0	Type I error (α) (False Positive)	Correct inference (True Positive) ($1-\beta$)
Fail to reject H_0	Correct inference (True Negative) ($1-\alpha$)	Type II error (β) (False Negative)

An error in the analysis!

More work is needed.

$$P(H_1|H_0) = \alpha$$

$$P(H_0|H_1) = \beta$$

$$P(H_1|H_1) = 1 - \beta \quad (\text{Power of Test})$$

$$P(H_0|H_0) = 1 - \alpha \quad (\text{Confidence level})$$

Example: Type I vs. Type II Error

- Consider the following hypothesis of a person's judgement:

H_0 : The defendant is innocent.

H_a : The defendant is guilty.

Reality/Decision	H_0 rejected (Guilty)	H_0 accepted (Innocent)
H_0 True (Innocent)	Type I Error (α)	Right Decision ($1-\alpha$)
H_0 False (Guilty)	Right Decision ($1-\beta$)	Type II Error (β)

- In a legal system it is assumed that a defendant is innocent unless the evidence gathered by the prosecutor is sufficient to reject this assumption.

Type of errors

- Type I error (False Positive): "rejecting the null hypothesis when it is true".
- Type II error (False negative): "accepting the null hypothesis when it is false".
- Type III error: "solving the wrong problem [representation]".
- Type IV error: "the incorrect interpretation of a correctly rejected hypothesis".

Joking?

Dirty Rotten Strategies: How We Trick Ourselves and Others into Solving the Wrong Problems Precisely (High Reliability and Crisis Management), Ian I. Mitroff, Abraham Silvers. ISBN-13: 978-0804759960

Type of errors

- Type 1 errors are more important than Type 2 errors → evidence.
 - ▣ p -values are only correlated with evidence.
 - ▣ Evidence in science is necessarily *relative*. When data is more likely assuming one model is true (e.g., a null model) compared to another model (e.g., the alternative model), we can say the model provides **evidence** for the null compared to the alternative hypothesis. p -values only give you the probability of the data under one model – what you need for evidence is the relative likelihood of two models.

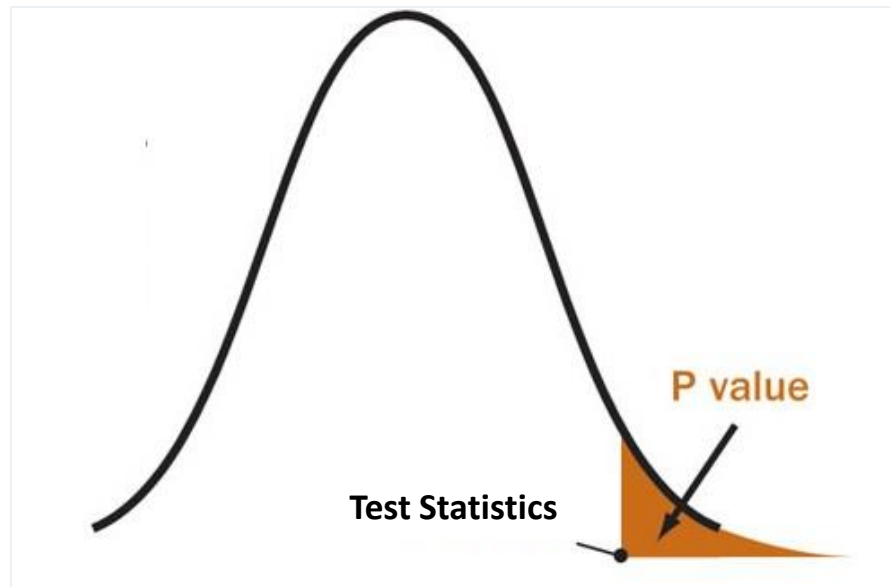
Steps of Hypothesis Testing

1. Determine the null and alternative hypothesis
2. Specify the level of significance and the decision rules
3. Calculate the related test statistic to test the hypothesis
4. Make a decision (by using p value or critical value approach)

Hypothesis Testing

□ p-value approach:

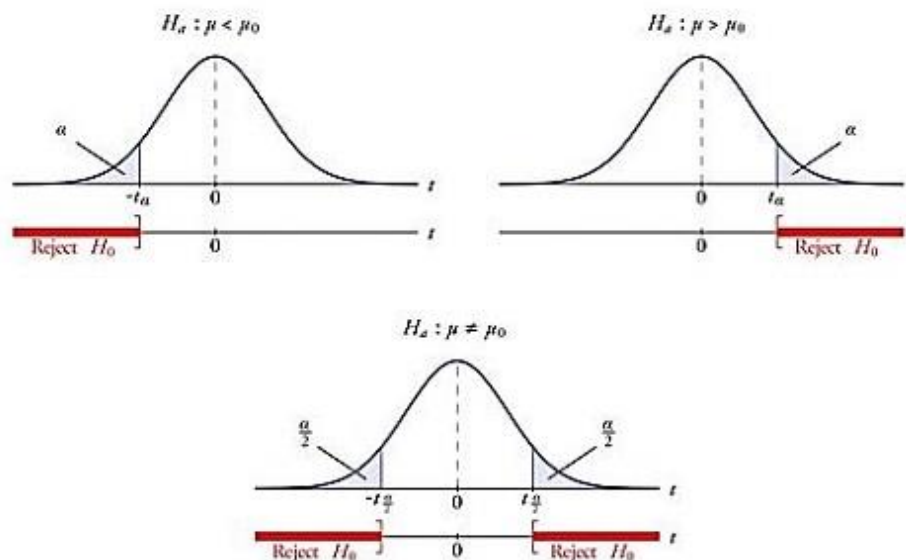
- Use the test statistic to compute the p value.
- Reject H_0 if the p-value $\leq \alpha$.



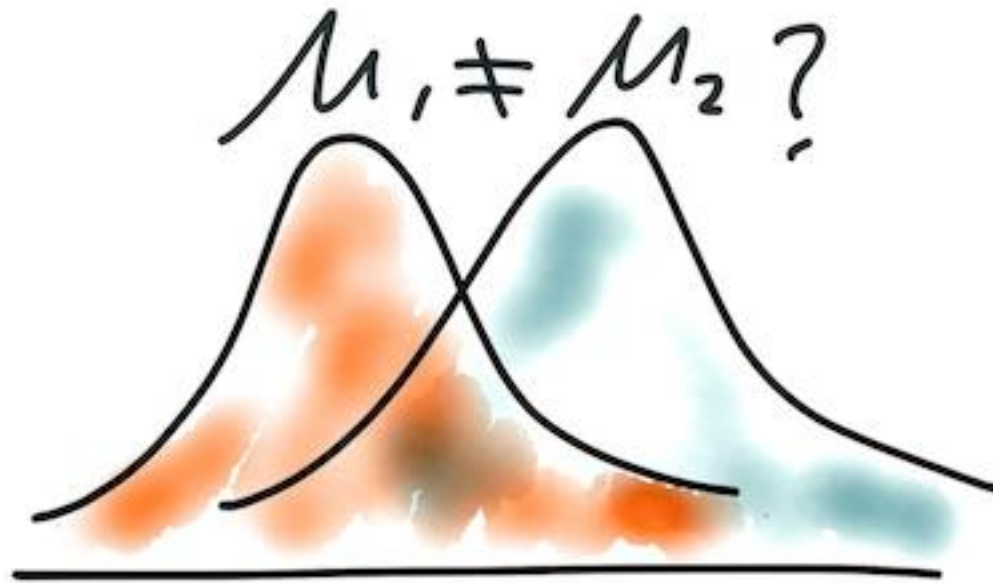
Hypothesis Testing

□ Critical value approach:

- Use the level of significance to determine the critical value and the rejection rule.
- Use the value of the test statistic and the rejection rule to determine whether to reject H_0 .



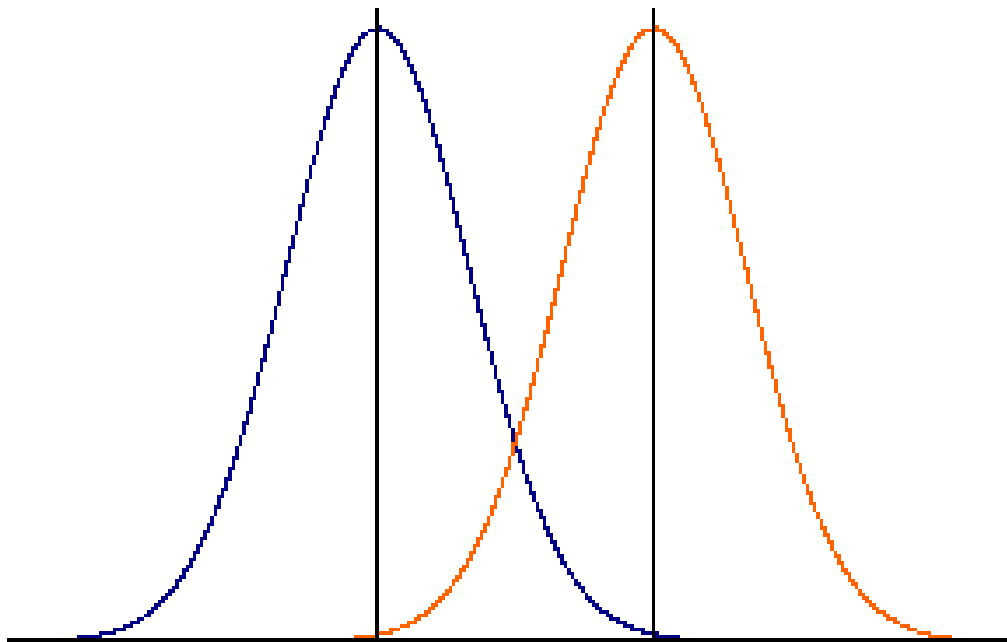
Comparison of Means





Comparison of Two Distributions with Equal Variances

Comparison of two configurations with equal variances.



Comparison of two configurations with equal variances.

- We define the hypothesis test:
 - $H_0: \mu_A \leq \mu_B$
 - $H_1: \mu_A > \mu_B$
- Thanks the central limit theorem we obtain that:

$$\bar{y}_A \approx N(\mu_A, \frac{\sigma_A}{\sqrt{n_A}})$$

$$\bar{y}_B \approx N(\mu_B, \frac{\sigma_B}{\sqrt{n_B}})$$

Comparison of two configurations with equal variances.

□ We can deduce that:

$$\bar{y}_A - \bar{y}_B \approx N(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}})$$

$$\frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \approx N(0,1)$$

Comparison of two configurations with equal variances.

- We define the test, and calculate s , the common sample variance:

$$\frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \approx t_n$$

- Where $n = n_A + n_B - 2$

Confidence Interval for the Difference of Means (Equal Variances)

□ For the case of equal variances:

$$\bar{y}_A - \bar{y}_B \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

where s_p is a pooled variance,

$$s_p = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

Comparison of two configurations with equal variances.

- The test is defined as is shown:

$$\frac{\bar{y}_A - \bar{y}_B}{s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} > t_{\alpha, n}$$

- We reject H_0 if this is true.

Example

Equal variances

Example

Observation	Values for pop A	Values for pop B
1	24.3	24.4
2	25.6	21.5
3	26.7	25.1
4	22.7	22.8
5	24.8	25.2
6	23.8	23.5
7	25.9	22.2
8	26.4	23.5
9	25.8	23.3
10	25.4	24.7

Example

- Mean of the sample.

- ▣ $A=25.14$; $B=23.62$

- ▣ $H_0: \mu_A \leq \mu_B$

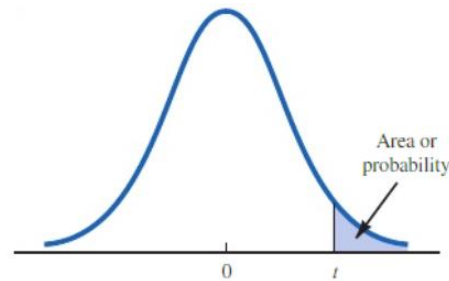
- ▣ $H_1: \mu_A > \mu_B$

Example

- Equal variances:

$$\frac{\bar{y}_A - \bar{y}_B}{s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} > t_{\alpha, n}$$

t-table



Entries in the table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861

Example

- The standard deviation is:

- $S_A = 1.242; S_B = 1.237$

$$\frac{25.14 - 23.62}{1.24 \sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.74 > t_{0.05, 18} = 1.734$$

- Reject H_0



Comparison of Two Distributions with Unequal Variances

Comparison of two configurations with unequal variances

- If we cannot assume equal variances.

$$t' = \frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

Two configurations comparison.

- If $n_A = n_B = n$, the signification level is determined using as a reference distribution a **t of Student** with $n-1$ degrees of freedom.
- If $n_A \neq n_B$, with the value calculated of t' we can find different signification values p_A and p_B in the student distributions, with $n_A - 1$ and $n_B - 1$ degrees of freedom respectively.

Two configurations comparison

- The signification level of the test:

$$\alpha = \frac{\omega_A p_A + \omega_B p_B}{\omega_A + \omega_B}$$

- with:

$$\omega_A = \frac{S_A^2}{n_A} \qquad \omega_B = \frac{S_B^2}{n_B}$$

Confidence Interval for the Difference of Means (Unequal Variances)

- For the case of unequal variances:

$$\bar{y}_A - \bar{y}_B \pm t_{\alpha/2} \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

Equal variance test

□ Hypothesis test:

- $H_0: \sigma_A^2 = \sigma_B^2$

- $H_1: \sigma_A^2 \neq \sigma_B^2$

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}}$$

$$\frac{S_A^2}{S_B^2} \approx F_{n,m}$$

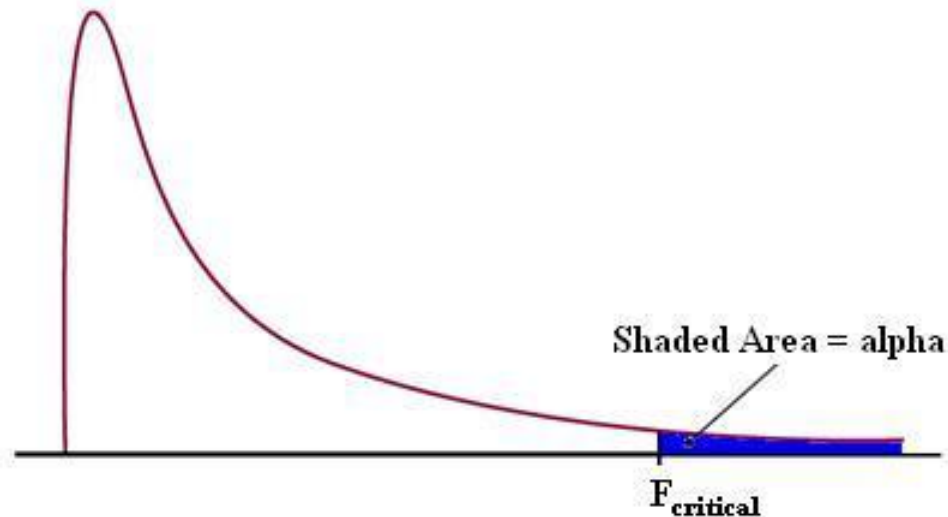
□ F of Fisher-Snedecor

- $n = n_A - 1$

- $m = n_B - 1$

F distribution and F table

- F table: <http://www.z-table.com/f-distribution-table.html>



Example

- $S_A^2 = 1.54$
- $S_B^2 = 2.18$

$$\frac{S_B^2}{S_A^2} = \frac{2.18}{1.54} = 1.42 < F_{0.05,9,9} = 3.18$$

- Accept H_0

Example

- $S_A^2 = 1.54$
- $S_B^2 = 16.3$

$$\frac{S_B^2}{S_A^2} = \frac{16.3}{1.54} = 10.58 > F_{0.05,9,9} = 3.18$$

- Reject H_0

To know more

- Part III: Statistics and Chapter 9.4: Statistical Inference I of **Probability and Statistics for Computer Scientists** (2014 Ed.)