Probability and Statistics
Lecture 9: 1 and 2-Sample Estimation

to accompany

Probability and Statistics for Engineers and Scientists

Fatih Cavdur

Introduction

A statistic $\bar{\theta}$ is said to be an unbiased estimator of the parameter θ if

$$E(\bar{\theta}) = \theta$$

We can show that

$$E(\bar{X}) = \mu$$

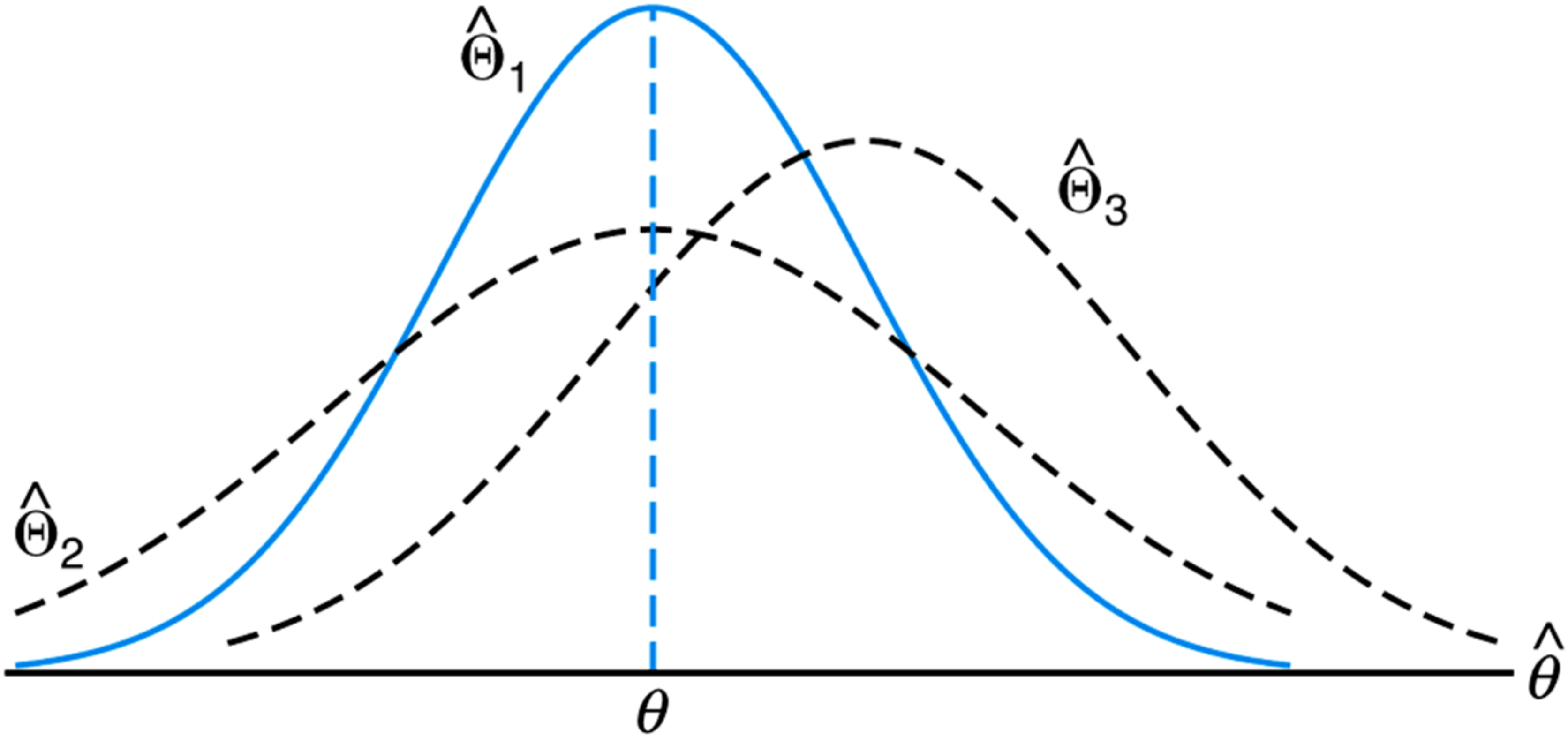
and

$$E(S^2) = \sigma^2$$

Introduction

If we consider all possible unbiased estimators of θ, the one with the smallest variance is said to be the most efficient estimator of θ.

Introduction

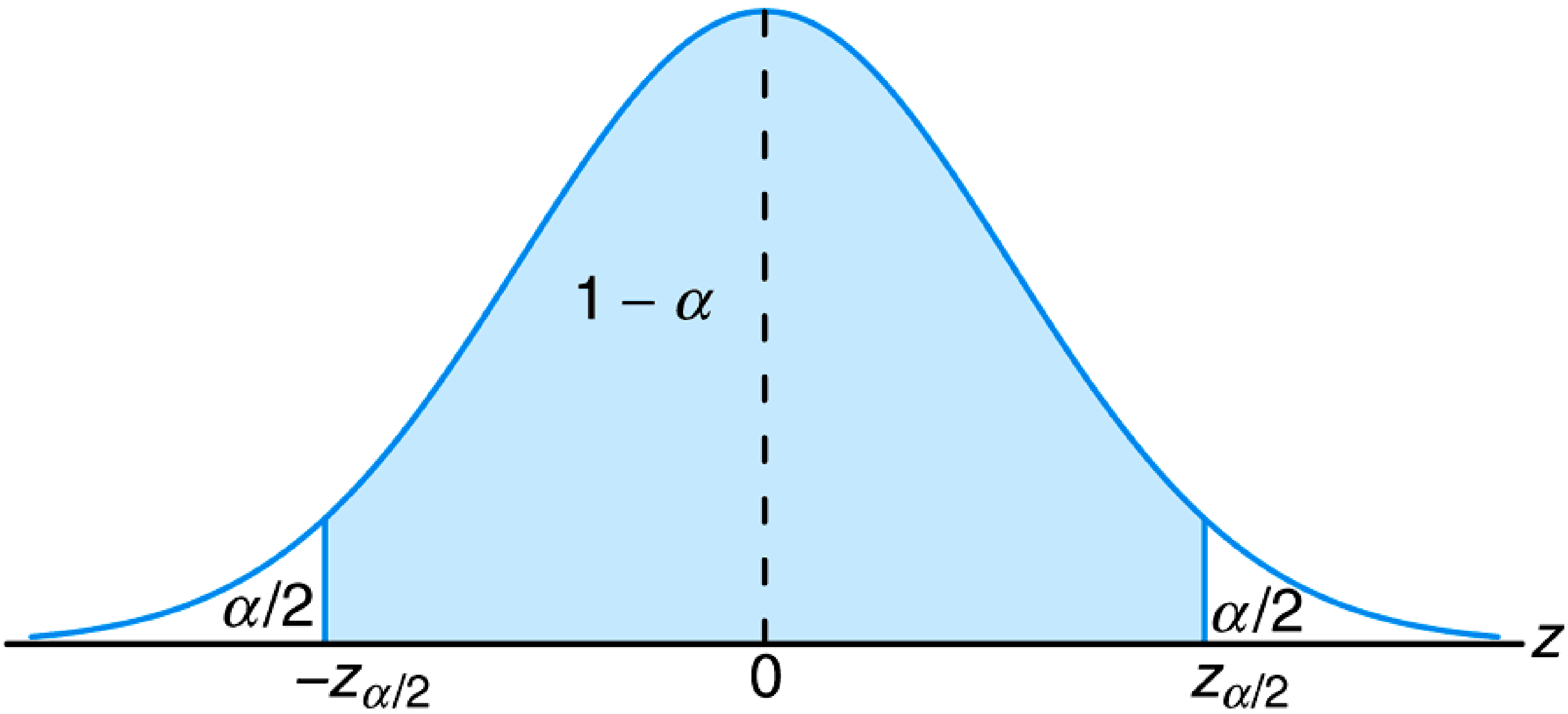


1-Sample: Estimating the Mean

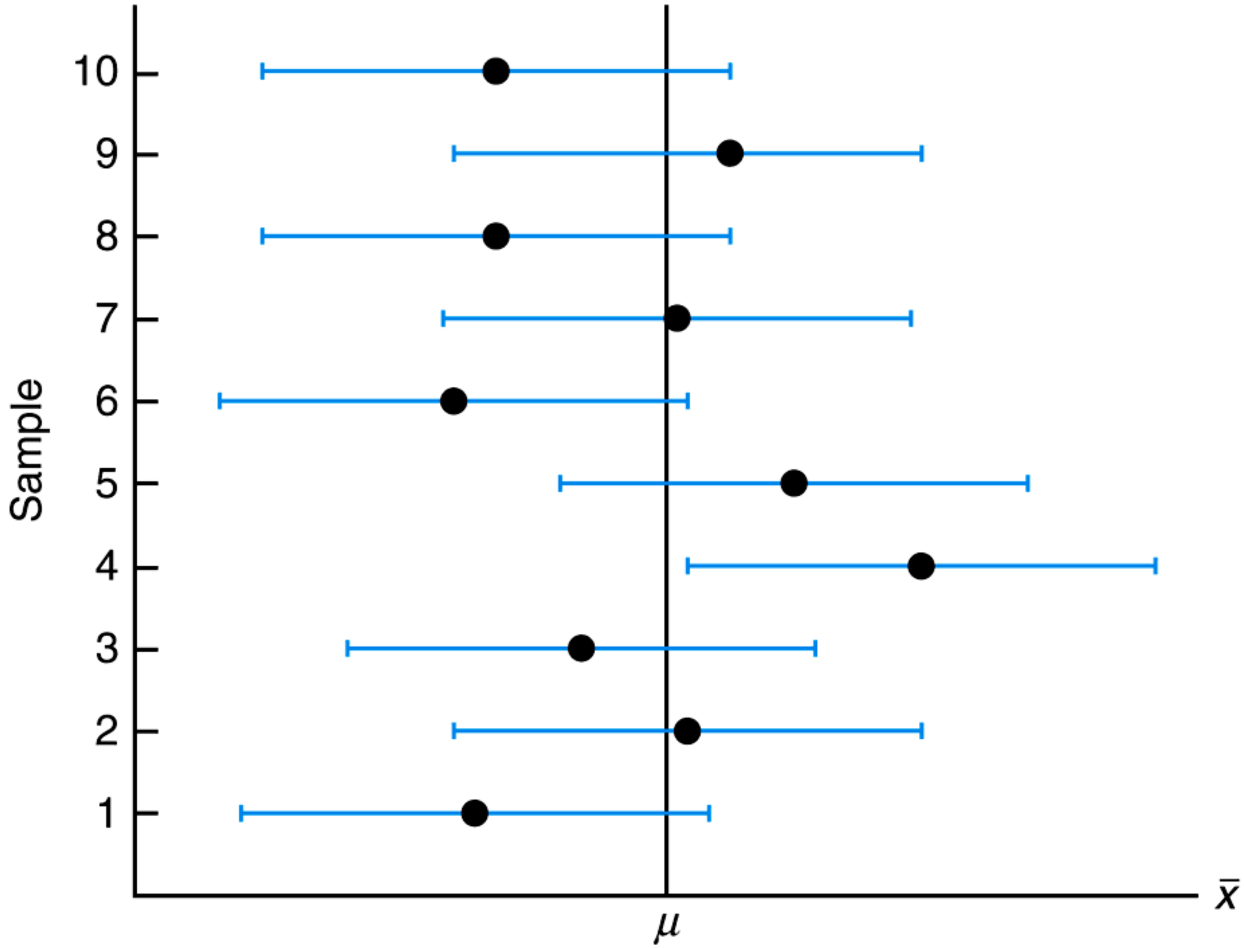
If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

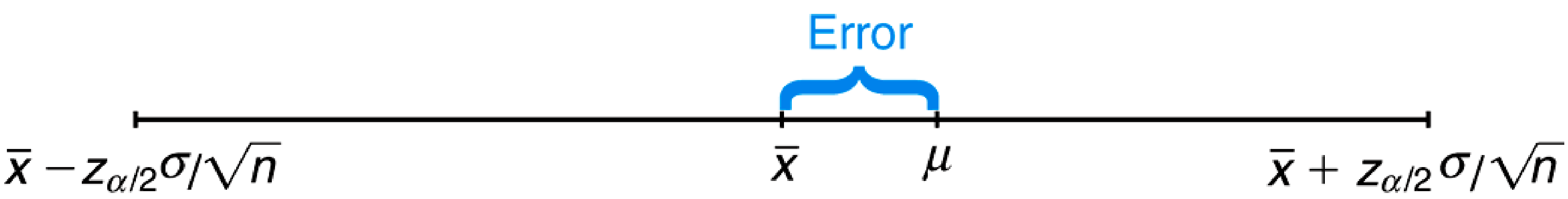
1-Sample: Estimating the Mean



1-Sample: Estimating the Mean



1-Sample: Estimating the Mean



1-Sample: Estimating the Mean

Example 9.2:

The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 grams per milliliter.

1-Sample: Estimating the Mean

The 95% confidence interval is

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$2.6 - z_{0.025}\frac{0.3}{\sqrt{36}} < \mu < 2.6 + z_{0.025}\frac{0.3}{\sqrt{36}}$$

$$2.6 - (1.96)\frac{0.3}{\sqrt{36}} < \mu < 2.6 + (1.96)\frac{0.3}{\sqrt{36}}$$

$$2.50 < \mu < 2.70$$

1-Sample: Estimating the Mean

The 99% confidence interval is

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$2.6 - z_{0.005}\frac{0.3}{\sqrt{36}} < \mu < 2.6 + z_{0.005}\frac{0.3}{\sqrt{36}}$$

$$2.6 - (2.575)\frac{0.3}{\sqrt{36}} < \mu < 2.6 + (2.575)\frac{0.3}{\sqrt{36}}$$

$$2.47 < \mu < 2.73$$

1-Sample: Estimating the Mean

If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a 1-sided $100(1-\alpha)\%$ confidence interval on μ:

$$\mu \leq \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$
$$\mu \geq \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

1-Sample: Estimating the Mean

Example 9.4:

In an experiment, 25 subjects are selected randomly and their reaction time to a particular stimulus is measured. Past experience suggests that the standard deviation of these types of stimuli is 2 seconds and that the distribution of reaction times is approximately normal. The average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

1-Sample: Estimating the Mean

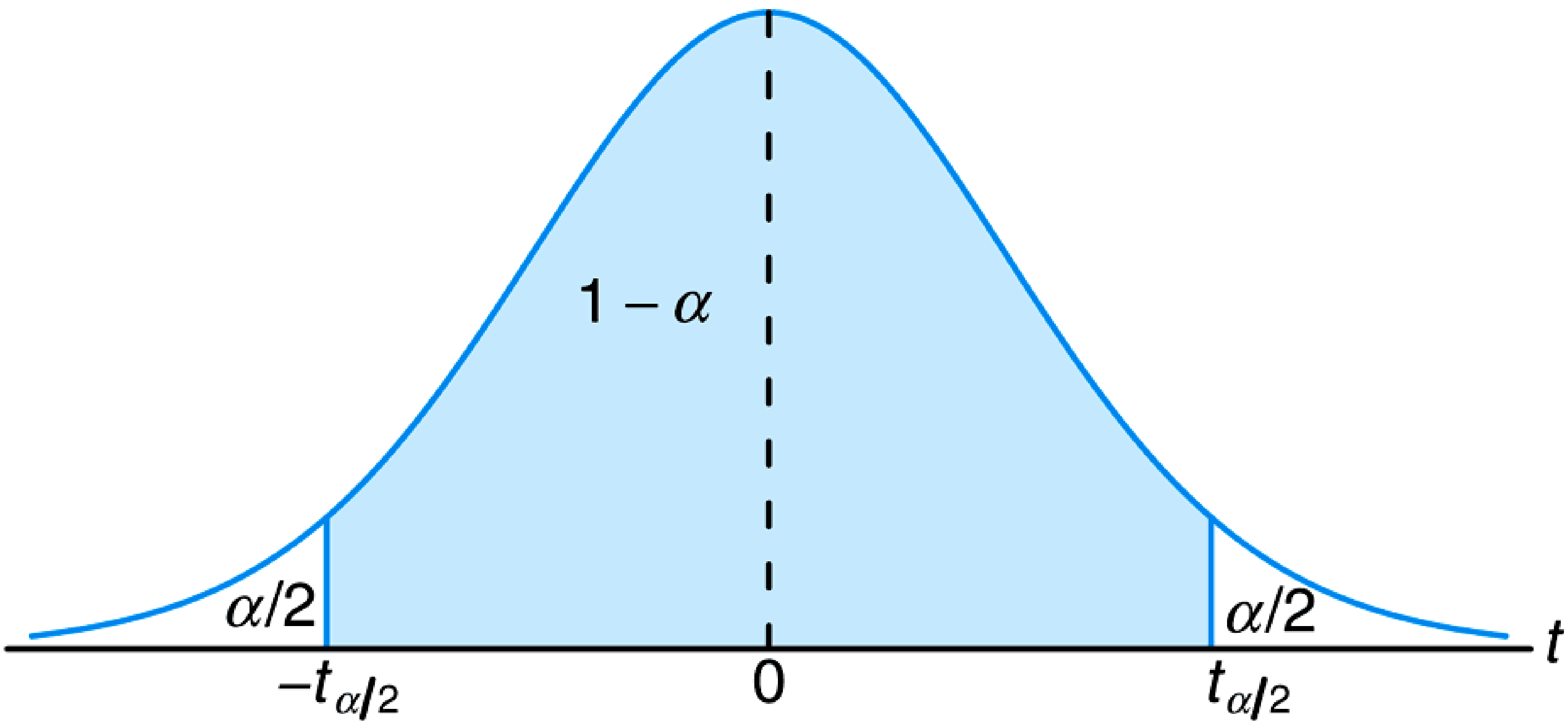
$$\mu \leq \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}} = 6.2 + z_{0.05}\frac{2}{\sqrt{25}} = 6.2 + (1.645)\frac{2}{\sqrt{25}}6.858$$

1-Sample: Estimating the Mean

If $\bar{x}$ and s are the mean and standard deviation of a random sample of size n from a population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ:

$$\bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}$$

1-Sample: Estimating the Mean



1-Sample: Estimating the Mean

Example 9.2:

The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

1-Sample: Estimating the Mean

We have

$\bar{x} = 10.0$, $s = 0.283$ and $t_{0.025;6} = 2.447$. Hence, the 95% confidence interval is

$$\bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}$$

$$10.0 - t_{0.025;6}\frac{0.283}{\sqrt{7}} < \mu < 10.0 + t_{0.025;6}\frac{0.283}{\sqrt{7}}$$

$$10.0 - (2.447)\frac{0.283}{\sqrt{7}} < \mu < 10.0 + (2.447)\frac{0.283}{\sqrt{7}}$$

$$9.74 < \mu < 10.26$$

2-Samples: Estimating the Difference of Means

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random sample of size n_1 and n_2 from a population with known variances σ_1^2 and σ_2^2, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2-Samples: Estimating the Difference of Means

Example 9.10: A study was conducted to compare two types of engines, A and B. Gas mileages, in miles per gallon, was measured. 50 experiments were conducted for engine type A and 75 experiments were conducted for engine type B. All conditions were held constant. The average gas mileages were 36 miles per gallon for type A and 42 miles per gallon for type B. Find a 96% confidence interval on $\mu_B - \mu_A$. Assume that the population standard deviations are 6 and 8 for type A and B, respectively.

2-Samples: Estimating the Difference of Means

We have,

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(42 - 36) - z_{0.02}\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < (42 - 36) + z_{0.02}\sqrt{\frac{64}{75} + \frac{36}{50}}$$

$$3.43 < \mu_B - \mu_A < 8.57$$

2-Samples: Estimating the Difference of Means

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random sample of size n_1 and n_2 from a population with unknown variances $\sigma_1^2 = \sigma_2^2$, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm \left(t_{\alpha/2;n_1+n_2-2}\right)\left(s_p\right)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where s_p is the pooled estimate of the population standard deviation

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

2-Samples: Estimating the Difference of Means

Example 9.11:

Assume that we have two samples with $n_1 = 12$, $\bar{x}_1 = 3.11$, $s_1 = 0.771$ and $n_2 = 10$, $\bar{x}_2 = 2.04$, $s_2 = 0.448$. Find a 99% confidence interval for the difference between the population means assuming that the populations are approximately normally distributed with equal variances.

2-Samples: Estimating the Difference of Means

We have

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(11)(0.771)^2 + (9)(0.448)^2}{12 + 10 - 2}}$$

$$= 0.646$$

2-Samples: Estimating the Difference of Means

and

$$(\bar{x}_1 - \bar{x}_2) \pm \left(t_{\alpha/2;n_1+n_2-2}\right)\left(s_p\right)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(3.11 - 2.04) \pm \left(t_{0.05;20}\right)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}} \Rightarrow 0.59 < \mu_1 - \mu_2 < 1.55$$

2-Samples: Estimating the Difference of Means

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random sample of size n_1 and n_2 from an approximately population with unknown variances σ_1^2 and σ_2^2, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm \left(t_{\alpha/2;v}\right)\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where v is the degrees of freedom of the t distribution and defined as

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

2-Samples: Estimating the Difference of Means

We have,

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_1^2/n_1)^2}{n_2 - 1}} = \frac{\left[\frac{(3.07)^2}{15} + \frac{(0.80)^2}{12}\right]^2}{\frac{[(3.07)^2/15]^2}{14} + \frac{[(0.80)^2/12]^2}{11}} = 16.3$$

$$(\bar{x}_1 - \bar{x}_2) \pm \left(t_{\alpha/2;v}\right)\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(3.84 - 1.49) \pm \left(t_{0.025;16}\right)\sqrt{\frac{(3.07)^2}{15} + \frac{(0.80)^2}{12}} \Rightarrow 0.6 < \mu_1 - \mu_2 < 4.1$$

2-Samples: Estimating the Difference of Means

If $\bar{d}$ and s_d are the mean and standard deviation of the normally distributed differences of n random pairs of measurements, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$\bar{d} - t_{\alpha/2;n-1}\frac{s_d}{\sqrt{n}} < \mu_1 - \mu_2 < \bar{d} + t_{\alpha/2;n-1}\frac{s_d}{\sqrt{n}}$$

2-Samples: Estimating the Difference of Means

Example 9.13:

Assume that we have $n = 20$, $\bar{d} = -0.87$ and $s_d = 2.9773$. We also assume that the distribution of the differences is approximately normal. Find a 95% confidence interval for the difference.

We have

$$\bar{d} - t_{\alpha/2;n-1}\frac{s_d}{\sqrt{n}} < \mu_1 - \mu_2 < \bar{d} + t_{\alpha/2;n-1}\frac{s_d}{\sqrt{n}}$$

$$-0.8700 - t_{0.025;19}\frac{2.9773}{\sqrt{20}} < \mu_1 - \mu_2 < -0.8700 + t_{0.025;19}\frac{2.9773}{\sqrt{20}}$$

$$-2.2634 < \mu_1 - \mu_2 < 0.5234$$

1-Sample: Estimating a Proportion

If $\hat{p}$ is the proportion of successes in a random sample of size n and $\hat{q} = 1 - \hat{p}$, an approximate $100(1-\alpha)\%$ confidence interval for p:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

or

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} \pm \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}$$

2-Samples: Estimating the Difference of Proportions

If $\hat{p}_1$ and $\hat{p}_2$ are the proportions of successes in random samples of sizes n_1 and n_2, respectively, and $\hat{q}_1 = 1 - \hat{p}_1$ and $\hat{q}_2 = 1 - \hat{p}_2$, an approximate $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$:

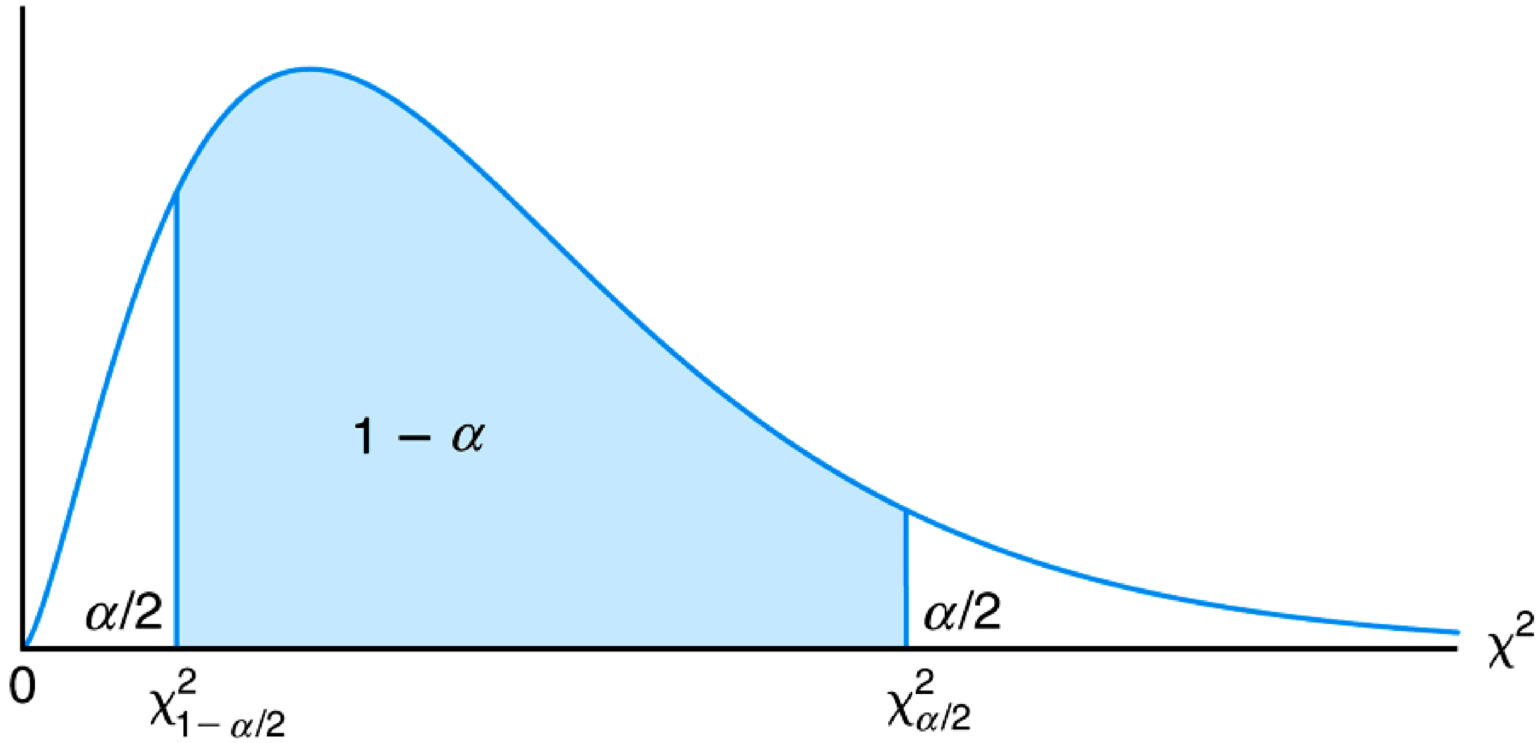
$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

1-Sample: Estimating the Variance

If s^2 is the variance of a random sample of size n from a normal population, a $100(1-\alpha)\%$ confidence interval for σ^2:

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

1-Sample: Estimating the Variance

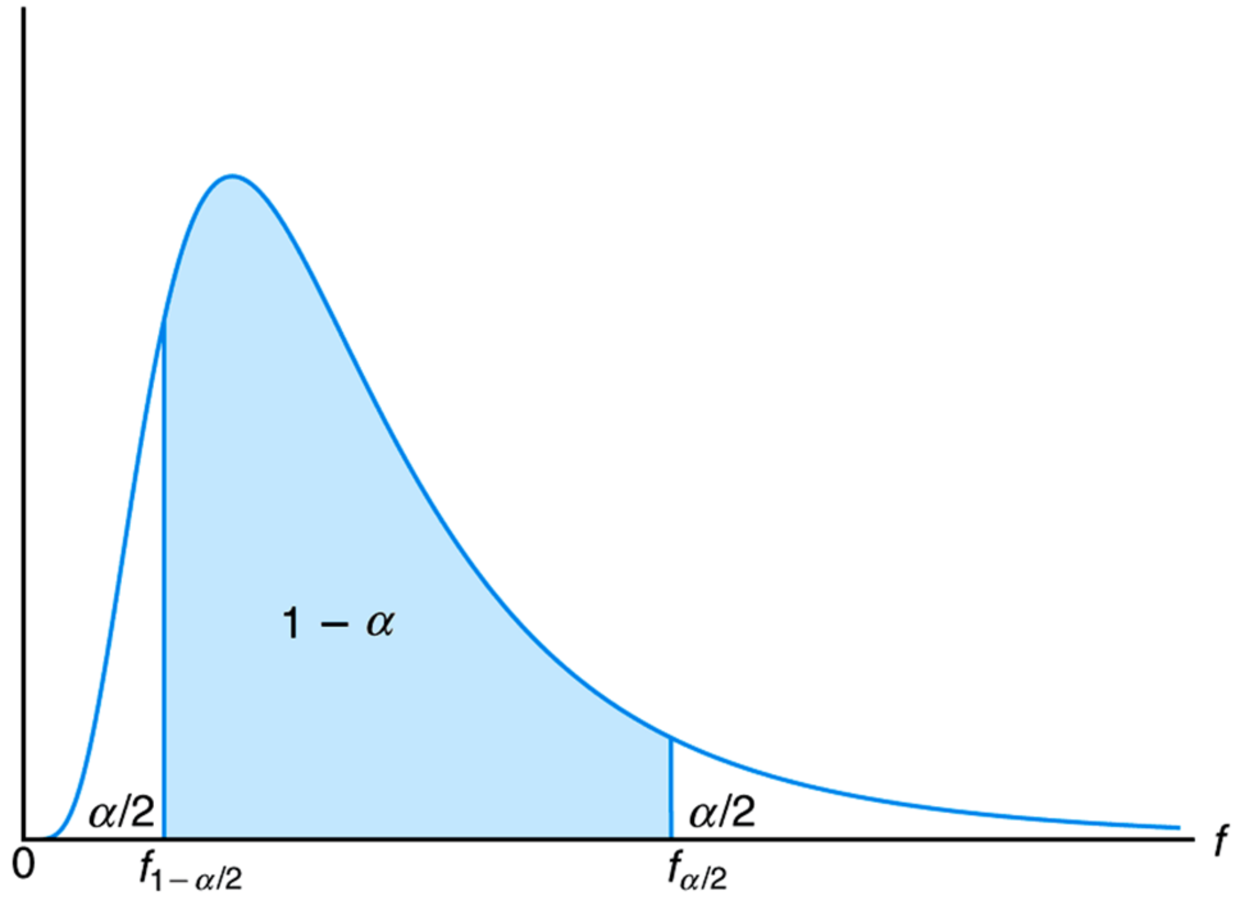


2-Samples: Estimating the Ratio of Variances

If s_1^2 and s_2^2 are the variances of independent samples of sizes n_1 and n_2 from normal populations, respectively, a $100(1-\alpha)\%$ confidence interval for σ_1^2/σ_2^2:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(n_1-1,n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(n_1-1,n_2-1)$$

2-Samples: Estimating the Ratio of Variances



End of Lecture

Thank you! Questions?