

# STAT 4352 - Mathematical Statistics Notes

JaimeGoB

March 3, 2021

# 1 Chapter 11 - Interval Estimation

## Point Estimators

$\theta$  is a unknown parameter (feature of a population)

- Ex: population mean  $\mu$
- **Fixed.**

$\hat{\theta}$  is a point estimator of  $\theta$  (it is a numerical value)

- Ex: sample mean  $\bar{x}$
- **Varies from sample to sample.**
- No guarantee of accuracy
- Must be *supplemented by*  $\text{Var}(\theta)$   
Standard Error  $\text{SE}(\hat{\theta})$  measures how much  $\hat{\theta}$  varies from sample to sample.  
small SE  $\implies$  low variance thus a more reliable estimate of  $\theta$

## Interval Estimators

### Def: Interval Estimate

Provides a range of values that best describe the population.

Let  $L = L(x)$  be the Lower Limit

$U = U(x)$  be the Upper Limit

Both  $L, U$  are Random Variables because they are functions of sample data.

### Def: Confidence Level / Confidence Coefficient

Is the probability that the **interval estimate** will include population parameter  $\theta$ .

- Sample means will follow the normal probability distribution for large sample sizes ( $n \geq 30$ )
- For small sample forces us to use the t-distribution probability distribution ( $n < 30$ )
- A confidence level of 95% implies that **95% of all samples would give an interval that includes  $\theta$ , and only 5% of all samples would yield an erroneous interval.**
- The most frequently used confidence levels are 90%, 95%, and 99% with corresponding Z-scores 1.645, 1.96, 2.576.
- The higher the confidence level, the more strongly we believe that the value of the parameter lies within the interval.

**Def: Confidence Interval**

Gives plausible values for the parameter  $\theta$  being estimated where degree of plausibility specified by a confidence level.

To construct an interval estimator of unknown parameter  $\theta$ . We must find two statistics **L** and **U** such that:

$$P\{\mathbf{L} \leq \theta \leq \mathbf{U}\} = 1 - \alpha$$

- $P\{\mathbf{L} \leq \theta \leq \mathbf{U}\}$  **Coverage Probability**, in repeated sampling, what percent of samples or Confidence Intervals capture true  $\theta$ .
- $100(1 - \alpha)$  **Confidence Interval** - for unknown fixed parameter  $\theta$ .
- **L, U - Lower and Upper Bounds** - RVs because they are functions of sample data. Vary from sample to sample.
- $1 - \alpha$  **Confidence Level** (Probability) estimate will include population parameter  $\theta$ .
- $\alpha$  **Level of Significance** Percent chance Confidence Interval will not contain population parameter  $\theta$ .

**Def: Coverage Probability**

$P\{\mathbf{L} \leq \theta \leq \mathbf{U}\}$  Gives what % of samples or Confidence Intervals capture true  $\theta$ .

Ex: Coverage Probability = 95%

Will capture  $\theta$ , 95% of the time.

Will NOT capture  $\theta$ , 5% of the time.

**Properties of Confidence Intervals**

- Confidence Intervals are not unique.
- Desirable to have  $E[\text{Length of CI}]$  to be small.
- A one-sided  $100(1 - \alpha)$  lower-confidence interval on  $\theta$ :  $L = -\infty \implies P\{L \leq \theta\} = 1 - \alpha$
- A one-sided  $100(1 - \alpha)$  upper-confidence interval on  $\theta$ :  $U = \infty \implies P\{\theta \leq U\} = 1 - \alpha$
- If **L, U** are both finite, then we have a two sided interval.

**Correctly Interpreting Confidence Intervals****Not Correct**

There is 90% probability that the true population mean is within the interval.

**Correct**

There is a 90% probability that any given Confidence Interval from a random sample will contain the true population mean.

## How to Construct Confidence Interval Using Pivot Approach:

Suppose we have a random sample  $X_1, X_2, \dots, X_n$  from a population distribution and the parameter of interest is  $\theta$ .

Given value  $\alpha \in (0, 1)$ . We would like to construct a  $1-\alpha$  Confidence Interval using a Pivot Approach:

1. Find a variable  $Y$ , that is function of the parameter  $\theta$  and data  $x$ .
2. The distribution of newly created variable  $Y$  is free of  $\theta$ .

In many cases:

$Y = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})}$  is a pivot and the distribution of  $Y$  is symmetric about 0.

## Using Pivot Approach for Two-Sided Intervals:

Find the critical points denoted  $c_{\alpha/2}$  such that:

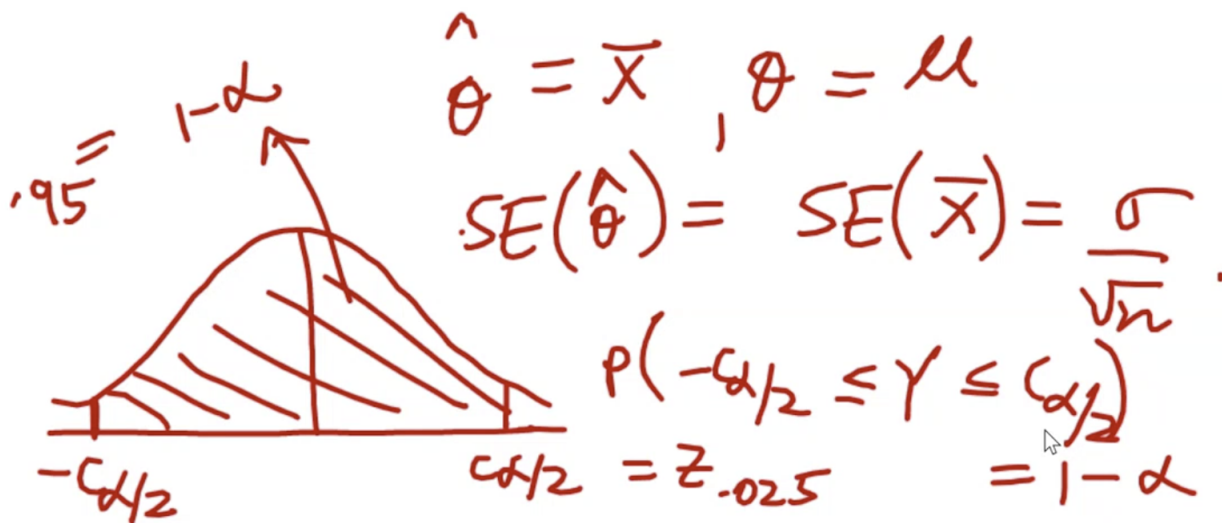
$$P\{-c_{\alpha/2} \leq Y \leq c_{\alpha/2}\} = 1 - \alpha$$

$c_{\alpha/2}$  is the upper  $(\alpha / 2)100$ th percentile.

**Critical points**- give you the area to the right of the point.

## Visualizing elements from Pivot Approach:

Let  $\mu$  be parameter of interest. We can construct CI using pivot approach.



### Symmetric Two-sided CI: Theorem

$\hat{\theta} \pm c_{\alpha/2}(SE(\hat{\theta}))$  is a  $100(1 - \alpha)\%$  confidence interval for  $\theta$

#### Proof:

$$1 - \alpha = P\{-c_{\alpha/2} \leq Y \leq c_{\alpha/2}\}$$

$$= P\{-c_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \leq c_{\alpha/2}\}$$

$$= P\{\hat{\theta} - c_{\alpha/2}(SE(\hat{\theta})) \leq Y \leq \hat{\theta} + c_{\alpha/2}(SE(\hat{\theta}))\}$$

$$\implies \hat{\theta} \text{ is **within** } c_{\alpha/2}(SE(\hat{\theta})) \text{ of } \theta \text{ with **probability** } 1-\alpha$$

$c_{\alpha/2}(SE(\hat{\theta}))$  is known as *Margin of Error* (size of error in estimation)

Ex: In polls you might hear accurate with 0.02 (this is margin of error)

### Asymmetric Two-sided CI(Non-symmetric distributions):

$[\hat{\theta} - c_{\alpha/2}(SE(\hat{\theta})), \hat{\theta} - c_{1-\alpha/2}(SE(\hat{\theta}))]$  is a  $100(1 - \alpha)\%$  confidence interval for  $\theta$

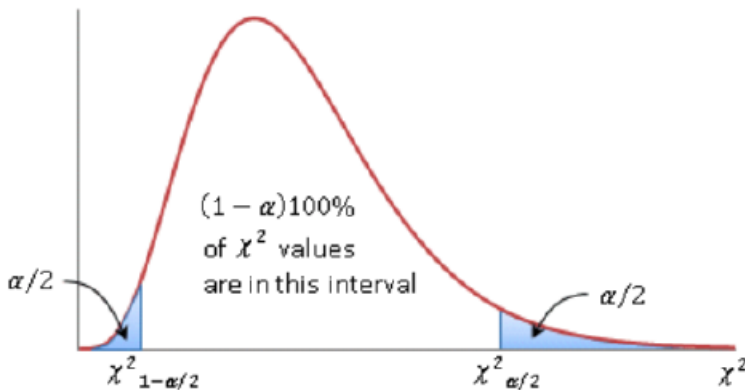
#### Proof:

$$1 - \alpha = P\{c_{1-\alpha/2} \leq Y \leq c_{\alpha/2}\}$$

$$= P\{c_{1-\alpha/2} \leq \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \leq c_{\alpha/2}\}$$

$$= P\{\hat{\theta} - c_{\alpha/2}(SE(\hat{\theta})) \leq \theta \leq \hat{\theta} - c_{1-\alpha/2}(SE(\hat{\theta}))\}$$

Ex: Chi-Square distribution critical points



### One-sided Confidence Bound:

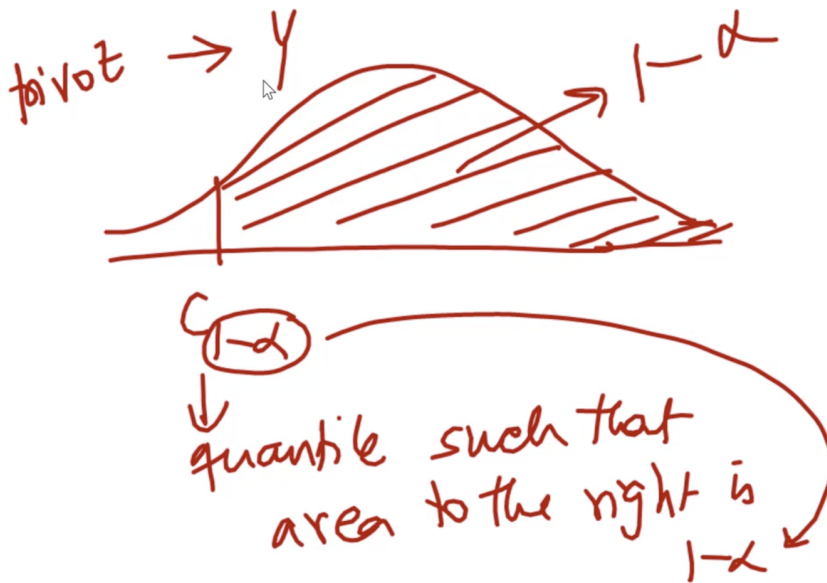
A one-sided confidence bound defines the point where a certain percentage of the population is either higher or lower than the defined point.

Upper Bound:  $U = \hat{\theta} - c_{1-\alpha}(SE(\hat{\theta}))$  when  $L = -\infty$

Lower Bound:  $L = \hat{\theta} - c_{\alpha}(SE(\hat{\theta}))$  when  $U = \infty$

### Proof(Upper Bound):

Coverage probability is  $1 - \alpha$ .



$$\begin{aligned} 1 - \alpha &= P\{Y \geq c_{1-\alpha}\} \\ &= P\left\{\frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \geq c_{1-\alpha}\right\} \\ &= P\{\theta \leq \hat{\theta} - c_{1-\alpha}(SE(\hat{\theta}))\} \end{aligned}$$

$$\implies U = \hat{\theta} - c_{1-\alpha}(SE(\hat{\theta}))$$

The lower bound can be computed in the same manner.

### How to interpret a one-sided CI?

For **Lower Bound** critical region  $\in [c_{\alpha}, \infty]$ : We are sure parameter  $\theta$  is below  $c_{\alpha}$

For **Upper Bound** critical region  $\in [-\infty, c_{1-\alpha}]$ : We are sure parameter  $\theta$  is above  $c_{1-\alpha}$

## Choice of Sample Size for CI for Mean when Variance Known

General form of a CI:

100(1- $\alpha$ )% CI for  $\theta$ :  $\hat{\theta} \pm c_{\alpha/2} SE(\hat{\theta})$

Mean  $\mu$  of a Normal population:  $\bar{x} \pm Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$

**Margin Of Error:**  $MOE = Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$

**Width of CI:**  $2 \times MOE = 2 \cdot Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$

### Properties of MOE:

- **As (1- $\alpha$ ) increases MOE increases:**

The larger the CI, the larger critical points are needed.

As critical points increase, Margin Of Error:  $MOE = Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$  increase

$\implies$  Increase in CI will make width of CI wider

- **As  $\sigma$  increases MOE increases:**

Margin Of Error:  $MOE = Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$   $\sigma$  is part of numerator, if it increases it will make MOE increase too.

- **As  $n$  increases MOE decreases:**

Margin Of Error:  $MOE = Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$   $n$  is part of denominator, if it increases it will make MOE decrease.

### Getting $n$ from predefined MOE:

Let  $E_0$  be a pre-specified MOE. We can find a value for  $n$  to make the following equation true.

$$\frac{Z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \leq E_0 \text{ solving for } n \text{ we get: } n \geq \left( \frac{Z_{\alpha/2} \cdot \sigma}{E_0} \right)^2 \text{ (round up to nearest } n.)$$

We do this when we want to know how many observations ( $n$ ) will give pre-specified margin of error -  $E_0$

# Theorem 11.1: Confidence Interval on the Mean of a Normal Distribution with known Variance

Let  $X$  be normal random variable with:

Unknown mean  $\mu$

Known variance  $\sigma^2$

Suppose a random sample  $n$ ,  $(X_1, X_2, \dots, X_n)$  is taken.

A  $100(1-\alpha)\%$  confidence interval on  $\mu$  can be obtained by considering sampling distribution of the sample mean  $\bar{X}$ .

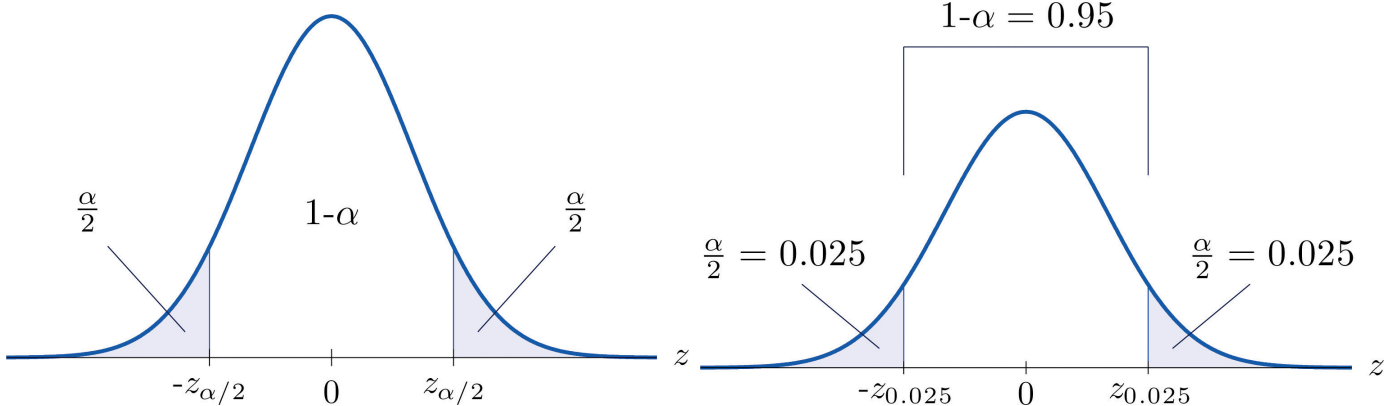
## Central Limit Theorem:

$$E(\bar{X}) = \mu \text{ and } SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}, \text{ so } \bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \text{ as } n \rightarrow \infty$$

Let  $Z = \text{Standardizing } \bar{X}$ ,  $Z$  will follow a Standard Normal Distribution

$$\text{Let } Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

We can see from the image to the left: **Distribution of  $Z$**  and the image to the right: **Confidence Interval of 95%**



We can see that:

$$P\{-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\} = 1 - \alpha$$

substituting  $Z$  into equation:

$$P\{-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\} = 1 - \alpha$$

isolating  $\mu$ :

$$P\{\bar{X} - Z_{\alpha/2}(\sigma/\sqrt{n}) \leq \mu \leq \bar{X} + Z_{\alpha/2}(\sigma/\sqrt{n})\} = 1 - \alpha$$

Conclusion  $\bar{X} \pm Z_{\alpha/2}(\sigma/\sqrt{n})$  is a  $100(1-\alpha)\%$  CI for  $\mu$



## Confidence Interval on the Mean of a Normal Distribution Variance Unknown and/or Small Sample:

$\bar{X} \pm t_{n-1,\alpha/2} \left( \frac{s}{\sqrt{n}} \right)$  is a  $100(1 - \alpha)\%$  CI for  $\mu$

**Proof:**

We know that  $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

$$1 - \alpha = P\{-t_{n-1,\alpha/2} \leq T \leq t_{n-1,\alpha/2}\}$$

$$= P\{-t_{n-1,\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1,\alpha/2}\}$$

$$= P\{\bar{X} - t_{n-1,\alpha/2}(S\sqrt{n}) \leq \mu \leq \bar{X} + t_{n-1,\alpha/2}(S\sqrt{n})\}$$

## Confidence Interval on the Mean no Specific Distribution Variance Known

$\bar{X} \pm Z_{\alpha/2}(\sigma/\sqrt{n})$  is a  $100(1-\alpha)\%$  CI for  $\mu$

**Proof:**

Assuming the sample size is large ( $n \geq 30$ ) then by CLT:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

The mean of any distribution **provided**  $n$  is large ( $n \geq 30$ ) can be approximated using a Normal Distribution.

## Confidence Interval on the Mean no Specific Distribution Variance Unknown

$\bar{X} \pm Z_{\alpha/2}(S/\sqrt{n})$  is a  $100(1-\alpha)\%$  CI for  $\mu$

**Proof:**

Given the fact that  $S^2$  is an unbiased estimator of  $\sigma^2$  we can use sample variance in lieu of population variance. Also sample size is large ( $n \geq 30$ ) then by CLT and LLN:

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \left( \frac{\sigma}{S} \right) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

## Confidence Interval on the Proportion of a Binomial Distribution

$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  is a  $100(1-\alpha)\%$  CI for  $p$

$n$  is random sample of size  $n$  has been taken from a large population and  $X(\leq n)$  observations in this sample belong to a class of interest.

$\hat{p} = X/n$  is the point estimator of the proportion of the population that belongs to this class.

$n$  and  $p$  are the parameters of a binomial distribution.

### Proof:

The sampling distribution of  $\hat{p}$  is approximately normal with mean  $p$  and variance  $p(1-p)/n$ , if  $p$  is not too close to 0 or 1 and  $n$  is large.

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

To construct CI on  $p$ , note that:

$$\begin{aligned} 1 - \alpha &\approx P\{-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\} \\ &\approx P\{-Z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq Z_{\alpha/2}\} \\ &\approx P\{\hat{p} - Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}\} \end{aligned}$$

Since the square root is the SE of estimator  $\hat{p}$  and also contains  $p$  in lower and upper bound.

We can replace  $p$  with  $\hat{p}$  and use Estimated SE instead of SE.

$$\approx P\{\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\}$$

## Confidence Interval on the Variance or Standard Deviation of a Normal Distribution - Mean is Unknown

$$\left[ \frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2} \right] \text{ is a } 100(1-\alpha)\% \text{ CI for } \sigma^2$$

**Proof:**

According to theorem 8.11:

$$Y = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

The critical points are:  $\chi_{n-1,1-\alpha/2}^2$  and  $\chi_{n-1,\alpha/2}^2$

$$\begin{aligned} 1 - \alpha &= P\{\chi_{n-1,1-\alpha/2}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{n-1,\alpha/2}^2\} \\ &= P\left[ \frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2} \right] \end{aligned}$$

## Confidence Interval on the Variance or Standard Deviation of a Normal Distribution - Mean is Known

$$\left[ \frac{(n)S^2}{\chi_{n,\alpha/2}^2}, \frac{(n)S^2}{\chi_{n,1-\alpha/2}^2} \right] \text{ is a } 100(1-\alpha)\% \text{ CI for } \sigma^2$$

**Proof:**

Since  $\mu$  is known then:

Sum of n, squared standard normal distributions

$\implies$  Sum of n Chi-Square distributions with one df

$\implies \chi_n^2$

$$Y = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} = \left( \frac{x_1 - \mu}{\sigma} \right)^2 + \dots + \left( \frac{x_n - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

The critical points are:  $\chi_{n,1-\alpha/2}^2$  and  $\chi_{n,\alpha/2}^2$

$$\begin{aligned} 1 - \alpha &= P\{\chi_{n,1-\alpha/2}^2 \leq \frac{(n)S^2}{\sigma^2} \leq \chi_{n,\alpha/2}^2\} \\ &= P\left[ \frac{(n)S^2}{\chi_{n,\alpha/2}^2} \leq \sigma^2 \leq \frac{(n)S^2}{\chi_{n,1-\alpha/2}^2} \right] \end{aligned}$$

## Two-Sample Confidence Interval Estimation

### Confidence Interval on the Difference between Means of Two Normal Distributions, Variances Known

In this case both means are unknown but variances are known.

$$\left[ \bar{X}_1 - \bar{X}_2 \pm (Z_{\alpha/2}) \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] \text{ is a } 100(1-\alpha)\% \text{ CI for } \mu_1 - \mu_2$$

#### Proof:

Let  $X_1$  and  $X_2$  be two normally distributed independent random variables.

$X_1 \sim N(\mu_1, \sigma_1^2)$  and  $X_2 \sim N(\mu_2, \sigma_2^2)$

So,  $\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2 \implies \text{SE}(\bar{X}_1 - \bar{X}_2) = \text{SD}(\bar{X}_1 - \bar{X}_2) = \sqrt{\text{Var}(\bar{X}_1 - \bar{X}_2)}$

Because  $\text{Var}(A - B) = \text{Var}(A) + \text{Var}(B)$  when A,B are independent

$$\text{SE}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \implies \boxed{\bar{X}_1 - \bar{X}_2 \sim N(\bar{X}_1 - \bar{X}_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}$$

Independent random samples from normal populations:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Now, to construct a CI:

$$\begin{aligned} 1 - \alpha &= P\{-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\} \\ &= P\left\{-Z_{\alpha/2} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z_{\alpha/2}\right\} \\ &= P\left[(-Z_{\alpha/2})\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) \leq (Z_{\alpha/2})\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \\ &= P\left[\bar{X}_1 - \bar{X}_2 - (Z_{\alpha/2})\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + (Z_{\alpha/2})\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \end{aligned}$$

## Confidence Interval on the Difference between Means of Two Normal Distributions, Variances Unknown or Small Samples

Both means and variances are unknown. However, we can assume  $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\left[ \bar{X}_1 - \bar{X}_2 \pm (t_{n_1+n_2-2, \alpha/2}) S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right] \text{ is a } 100(1-\alpha)\% \text{ CI for } \mu_1 - \mu_2$$

### Proof:

Let  $S_1^2$  and  $S_2^2$  be sample variances of random variables  $X_1$  and  $X_2$ . Since both sample variances are estimates of common variance  $\sigma^2$  we can obtain a pooled estimator of  $\sigma^2$ .

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_2 + n_1 - 2} \sim \chi_{n_1-1}^2 + \chi_{n_2-1}^2$$

Now pivot T

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Now, to construct a CI:

$$\begin{aligned} 1 - \alpha &= P\{-t_{n_1+n_2-2, \alpha/2} \leq T \leq t_{n_1+n_2-2, \alpha/2}\} \\ &= P\left\{-t_{n_1+n_2-2, \alpha/2} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{n_1+n_2-2, \alpha/2}\right\} \end{aligned}$$

$$P \left[ -t_{n_1+n_2-2, \alpha/2} \left( S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \leq (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) \leq t_{n_1+n_2-2, \alpha/2} \left( S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \right]$$

After solving for both population means:

$$P \left[ \bar{X}_1 - \bar{X}_2 - (t_{n_1+n_2-2, \alpha/2}) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + (t_{n_1+n_2-2, \alpha/2}) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

## Confidence Interval on the Difference between Means of Two Normal Distributions, Variances Unknown and Variances Differ

Both means and variances are unknown and the variances are not equal  $\sigma_1^2 \neq \sigma_2^2$ . In this case we make the assumption that variances are different.

$$\left[ \bar{X}_1 - \bar{X}_2 \pm (t_{v,\alpha/2}) \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right] \text{ is a } 100(1-\alpha)\% \text{ CI for } \mu_1 - \mu_2$$

### Proof:

From previous cases we know that:

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\bar{X}_1 - \bar{X}_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Now let Y be the approximate pivot:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t_v \text{ where df } v = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1 + 1} + \frac{(S_2^2/n_2)^2}{n_2 + 1}}$$

$$1 - \alpha = P\{-t_{v,\alpha/2} \leq T \leq t_{v,\alpha/2}\}$$

$$= P\left\{-t_{v,\alpha/2} \leq \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \leq t_{v,\alpha/2}\right\}$$

$$P\left[-t_{v,\alpha/2} \left(\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) \leq (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) \leq t_{v,\alpha/2} \left(\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)\right]$$

After solving for both population means:

$$P\left[\bar{X}_1 - \bar{X}_2 - (t_{v,\alpha/2})\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + (t_{v,\alpha/2})\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$$

## Confidence Interval on the Ratio of Variances of Two Normal Distributions

► CI for  $\sigma_1^2/\sigma_2^2$ .

$$\text{Recall: } \frac{\chi_{n_1-1}^2/(n_1-1)}{\chi_{n_2-1}^2/(n_2-1)} \sim F_{n_1-1, n_2-1}; \quad \frac{1}{F_{n_1-1, n_2-1}} = \frac{\chi_{n_2-1}^2/(n_2-1)}{\chi_{n_1-1}^2/(n_1-1)} \sim F_{n_2-1, n_1-1}$$

$$\text{Pivot: } Y = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2} \cdot \frac{1}{(n_1-1)}}{\frac{(n_2-1)S_2^2}{\sigma_2^2} \cdot \frac{1}{(n_2-1)}} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1, n_2-1}$$

$$\begin{aligned} 1 - \alpha &= P \left[ F_{1-\alpha/2, n_1-1, n_2-1} \leq \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq F_{\alpha/2, n_1-1, n_2-1} \right] \\ &= P \left[ F_{1-\alpha/2, n_1-1, n_2-1} \cdot \frac{S_2^2}{S_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_{\alpha/2, n_1-1, n_2-1} \cdot \frac{S_2^2}{S_1^2} \right] \\ &= P \left[ \frac{1}{F_{\alpha/2, n_1-1, n_2-1}} \cdot \frac{S_1^2}{S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{F_{1-\alpha/2, n_1-1, n_2-1}} \cdot \frac{S_1^2}{S_2^2} \right] \end{aligned}$$

$$100(1 - \alpha)\% \text{ CI for } \frac{\sigma_1^2}{\sigma_2^2} : \left[ \frac{1}{F_{\alpha/2, n_1-1, n_2-1}} \cdot \frac{S_1^2}{S_2^2}, \frac{1}{F_{1-\alpha/2, n_1-1, n_2-1}} \cdot \frac{S_1^2}{S_2^2} \right]$$

$$100(1 - \alpha)\% \text{ CI for } \frac{\sigma_2^2}{\sigma_1^2} : \left[ F_{1-\alpha/2, n_1-1, n_2-1} \cdot \frac{S_2^2}{S_1^2}, F_{\alpha/2, n_1-1, n_2-1} \cdot \frac{S_2^2}{S_1^2} \right]$$