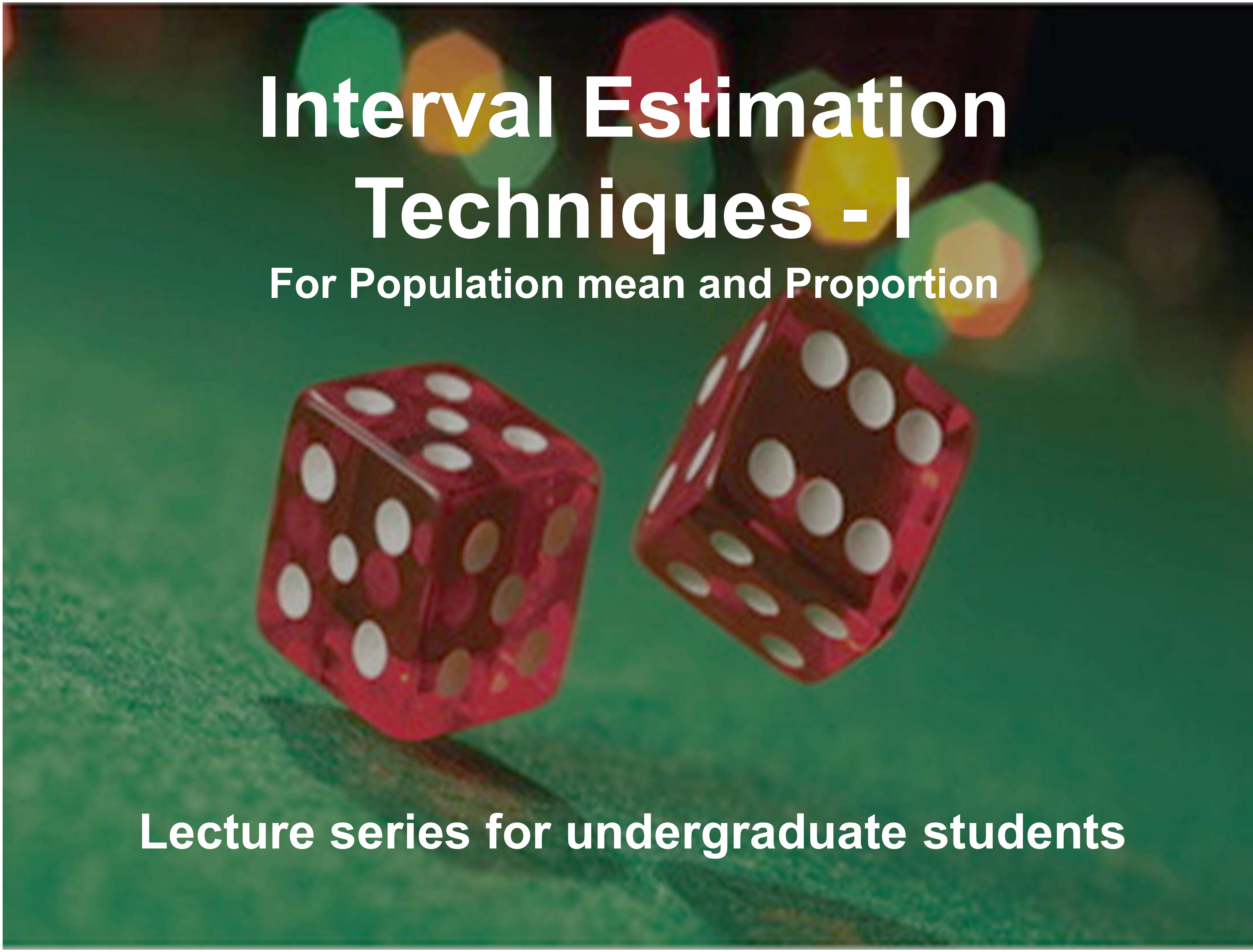

Interval Estimation
Techniques - I
For Population mean and Proportion
Lecture series for undergraduate students

Interval Estimation techniques- I

For Population mean and Proportion

Lecture 06

Week-06

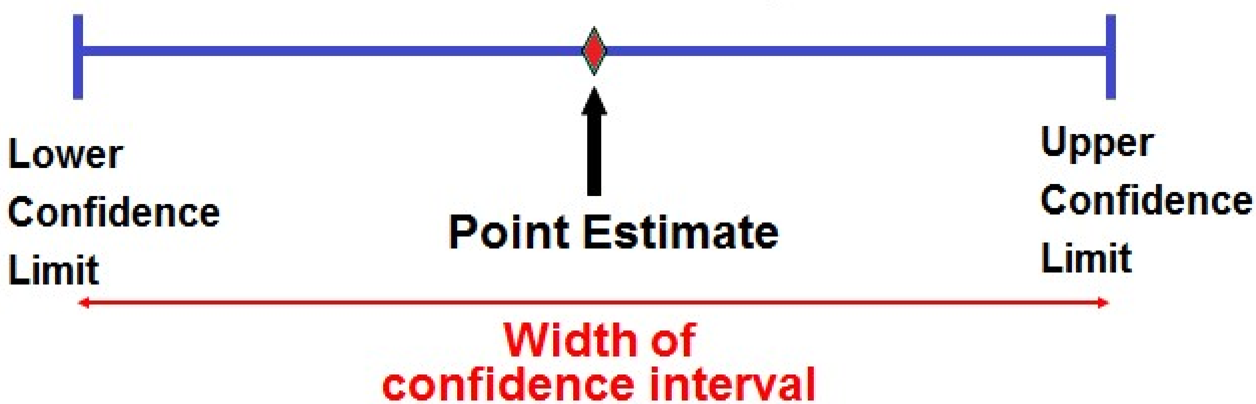
Dr. Tahseen A. Jilani

Recap- Point estimation techniques

- A given point estimate is a single realization of the random variable.
- The actual estimate may or may not be close to the parameter of interest. Therefore, if we only provide a point estimate of the parameter of interest, we are not giving any information about the *accuracy* of the estimation procedure.
- For example, saying that the $\bar{x} = 550$ is giving a point estimate of μ. This estimate ($\hat{\mu}$) does not tell us how close may be to actual unknown μ.
- Main techniques for point estimation:
  - **Least squares estimation (LSE)**
  - **Method of moments (MM)**
  - **Maximum likelihood estimation (MLE)**

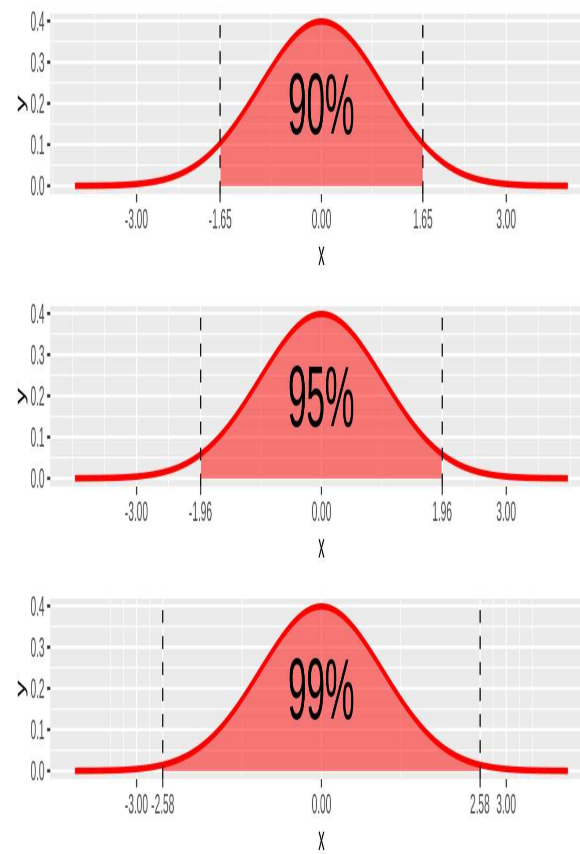
Recap - Point Estimation

Measure	Parameter	Statistic or point estimator	Sampling error
Mean	μ	$\bar{x}$	$\lvert\bar{x}-\mu\rvert$
Standard deviation	σ	s	$\lvert s-\sigma\rvert$
Proportion	p	$\bar{p}$	$\lvert\bar{p}-p\rvert$
No. of elements	N	n	



Interval estimation techniques

- Let, "We are *90% confident* that unknow μ lies in the interval [440, 560]."
- Let, "We are *95% confident* that unknown μ lies in the interval [425, 475]." **(When based on the same information, however, an interval of higher confidence level is wider).**
- **<u>Example</u>** is GPA of students in a class. If N = 80 and n = 10 then we can prepare the sampling distribution for $\bar{x}$ and can find 90%, 95%, 99% confidence intervals about the population mean GPA (μ) of the whole class.

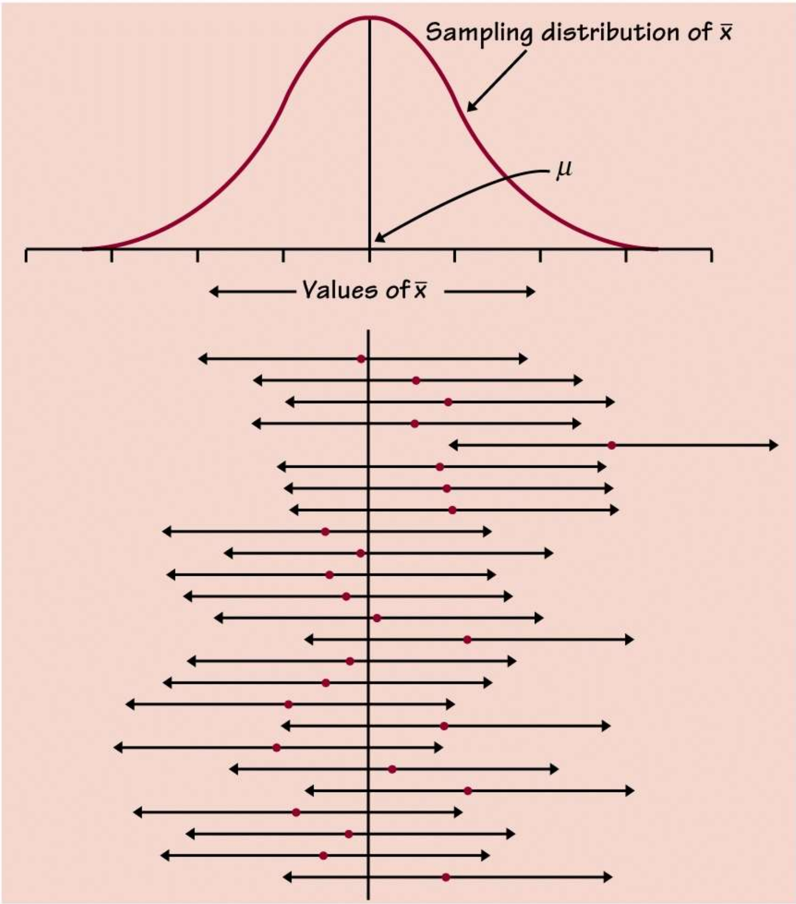


Confidence interval (CI)

- A confidence interval (CI) is a range of numbers believed to include an unknown population parameter $\hat{\theta}_L < \theta < \hat{\theta}_U$. Associated with the interval is a measure of the confidence (1-α). That means we are (1-α)100% confident that the unknown parameter lies inside the interval.
- α is called the level of significance (to be used in hypothesis testing). It can assume values like 0.10, 0.05, 0.01 and even smaller.
- Examples:
  - I am 95% sure that it will rain tomorrow.
  - The average GDP growth will be between 2.5% to 4.5% in 2021-2022.

What is (1-α)100% Confidence Interval

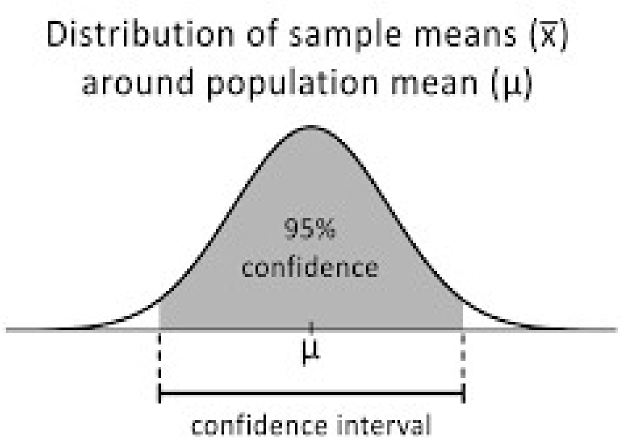
- A 95% confidence interval means out of 100 confidence intervals (using different samples from same population), 95 will contain the UNKNOWN population parameter μ.

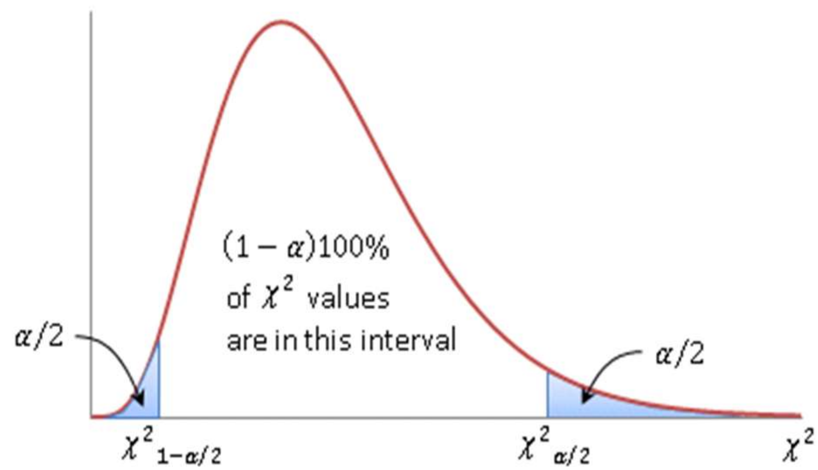


Confidence Interval (CI) estimation techniques

(1-α)100% confidence interval

- Just like point estimation, we have interval estimation formulae for population mean, population proportion and population variance etc.
  - Confidence interval (CI) estimation for single and two population means.
  - Confidence interval (CI) estimation for single and two population proportions.
  - Confidence interval (CI) estimation for one and two variance.





Confidence Interval Estimation for Population mean

There are three cases as we discussed in the central limit theorem.

1. When population variance (σ^2) is known.
2. When Population variance (σ^2) is unknown but $\mathrm{n} \geq 30$.
3. When population variance (σ^2) is unknown AND $\mathrm{n} < 30$.

CASE-I : (1-α)100% Confidence Interval estimation for single population mean (μ)

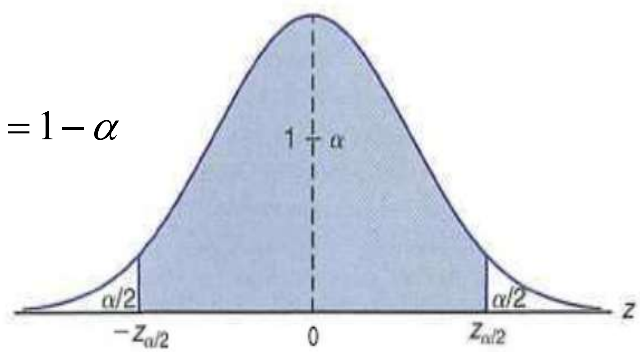
- From Central limit theorem, $\bar{x} \sim N(\mu, \sigma^2/n)$. We have $E(\bar{x}) = \mu$ and

  $SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$. Using $Z = \frac{\bar{x} - mean(\bar{x})}{SD(\bar{x})} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

- Z is symmetrically distributed, thus we split α into tow parts -α/2 (left hand of the mean) and α/2 (on right hand of the mean). Thus the probability that the unknow parameter μ lies between -α/2 and α/2 is given by:

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

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



$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$

(1-α)100% Confidence Interval estimation for μ

Note that

- (Effect of increasing α): When sampling is from the same population, using a fixed sample size, *the higher the confidence level, the wider the interval.*
- (Effect of increasing n): When sampling is from the same population, using a fixed confidence level, *the larger the sample size* n, *the narrower the confidence interval.*

Example: (1-α)100% C.I. for μ

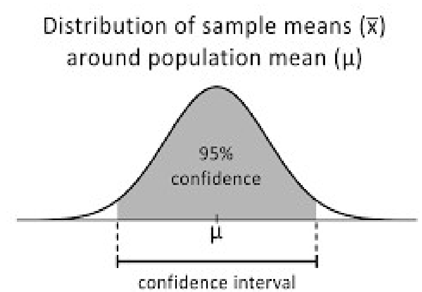
- I randomly select 46 students' Math SAT scores and find $\overline{x}$=600. I know that σ from this population is 50. Find a 95% Confidence Interval and interpret.
- Here α =0.05 and α/2 = 0.025. The value for the 95% CI for mean is 1.960 from the table. Putting values in:

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

$$600 - 1.96.0\ \frac{50}{\sqrt{46}} < \mu < 600 + 1.96.0\ \frac{50}{\sqrt{46}}$$

$$\mathbf{585.6 < \mu < 614.6}$$

Confidence Interval	Z
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291



Practice questions for case-I

- Randomly select 25 students' Math SAT scores and find $\overline{x} = 600$. I know that σ from this population is 50. Find a 90% Confidence Interval and interpret.
- Randomly select 25 students' Math SAT scores and find $\overline{x} = 600$. I know that σ from this population is 50. Find a 99% Confidence Interval and interpret.
- Randomly select 100 students' Math SAT scores and find $\overline{x} = 600$. I know that σ from this population is 50. Find a 99% Confidence Interval and interpret.
- A new sneaker claims that it can make male athletes jump higher. A sample of 25 male athletes is asked to jump once with their own sneakers on and once wearing the new sneakers. The average jump increased by 1.5 inches with the new sneakers! Assume that σ = 3. Find a 95% confidence interval and interpret.

CASE-II: (1-α)100% CI for μ when σ is unknown but $n \geq 30$

- From Central limit theorem, $\bar{x} \sim \mathrm{N}(\mu, {s^2}/{n})$.
- We have $E(\bar{x}) = \mu$ and $\mathrm{SD}(\bar{x}) = \frac{s}{\sqrt{n}}$. Using $z = \frac{\bar{x} - mean(\bar{x})}{SD(\bar{x})} = \frac{\bar{x} - \mu}{{s}/{\sqrt{n}}}$
- **The only change is to replace σ with sample standard deviation s.**

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

CASE-III: (1-α)100% CI for μ when σ is unknown and $\mathbf{n < 30}$

- As population variance is unknown and sample size is lesser than 30, thus we need to apply t-test rather the z-test for this situation.
- We have $E(\bar{x}) = \mu$ and SD($\bar{x}$) $= \frac{s}{\sqrt{n}}$. Using $t = \frac{\bar{x} - mean(\bar{x})}{SD(\bar{x})} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$$\bar{x} - t_{(n-1,\, \alpha/2)} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1,\, \alpha/2)} \cdot \frac{s}{\sqrt{n}}$$

- For t-table, we required the degrees of freedom n-1 and also level of significance α.

Confidence Interval Estimation for Population Proportion

Interval estimation for population proportion

- A population proportion (P) is defined as selected portion of a population based on some characteristic(s).
  - e.g. proportion of student living more than 10km away from the university.
  - e.g. proportion of people having brown eyes.
  - e.g. proportion of people who don't like potato crips.
- A sample proportion (p or $\widehat{P}$) on the other hand is the random sample from population with defined characteristic(s).
- To estimate population proportion, we use sample proportion.

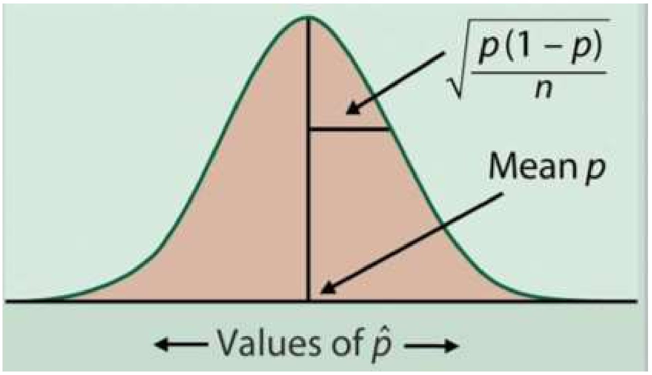
$$\boldsymbol{P = \frac{X}{N}} \text{ and } \boldsymbol{p = \widehat{P} = x/n}$$

- We can get many sample proportion (by taking different samples from same population) and can form sampling distribution for p.

Confidence Interval estimation for Population Proportion

- We can get many sample proportion (by taking different samples from same population) and can form sampling distribution for p same like we did for sample mean.
- Sampling distribution for sample proportion is normal distribution with mean $\mathbf{E}(\boldsymbol{p}) = \boldsymbol{P}$ and $\mathbf{Var}(\boldsymbol{p}) = \frac{\boldsymbol{P}(\mathbf{1}-\boldsymbol{P})}{\boldsymbol{n}}$ (take care for p and P)
- The confidence interval for population proportion P is given by:

$$\left(p - Z^* \sqrt{\frac{p \cdot (1-p)}{n}}, p + Z^* \sqrt{\frac{p \cdot (1-p)}{n}}\right)$$



- Not useful as it include P to estimate P !

Confidence Interval estimation for Population Proportion

- One approach is to replace P with the unbiased estimator $\hat{P}$ or p.

$$\left(\hat{p} - Z^* \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}, \hat{p} + Z^* \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right)$$

- Example: Sample of n= 2000 people asked if they will vote for McCain. Had a sample proportion of 0.51 for "yes." What is a 95% confidence interval for the true population proportion *P* of McCain voters?

- Here n = 2000, $\hat{P}$ = 0.51 and we use Z^*=1.96 so our 95% confidence interval for *P* is:

$$\left(.51 - 1.96\sqrt{\frac{.51 \cdot .49}{2000}}, .51 + 1.96\sqrt{\frac{.51 \cdot .49}{2000}}\right) = (.488, .532)$$

End of Lecture

Week-06