UNIT 8: INTRODUCTION TO INTERVAL ESTIMATION

8.1.- Introduction to interval estimation
8.2.- Confidence intervals. Construction and characteristics
8.3.- Confidence intervals for the mean
8.4.- Confidence intervals for the proportion
8.5.- Confidence intervals for the variance

UNIT 8. GOALS

- Describe advantages and disadvantages of point estimates and interval estimates.
- Interpret the characteristics of precision and confidence in estimations.
- Build confidence intervals for the population mean.
- Determine sample size for the mean.
- Build confidence intervals for the population variance and proportion.
- Determine sample size for the proportion.

Types of estimates

GOAL → **Approximate population parameters which are unknown**

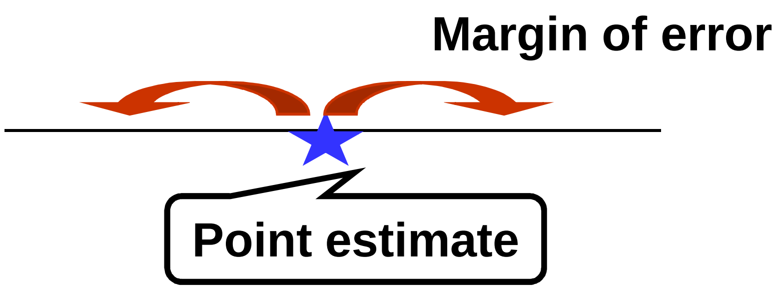
The estimates of the parameters can be expressed as:

A) Point Estimate

Disadvantage: there is no measure of how good the estimate is

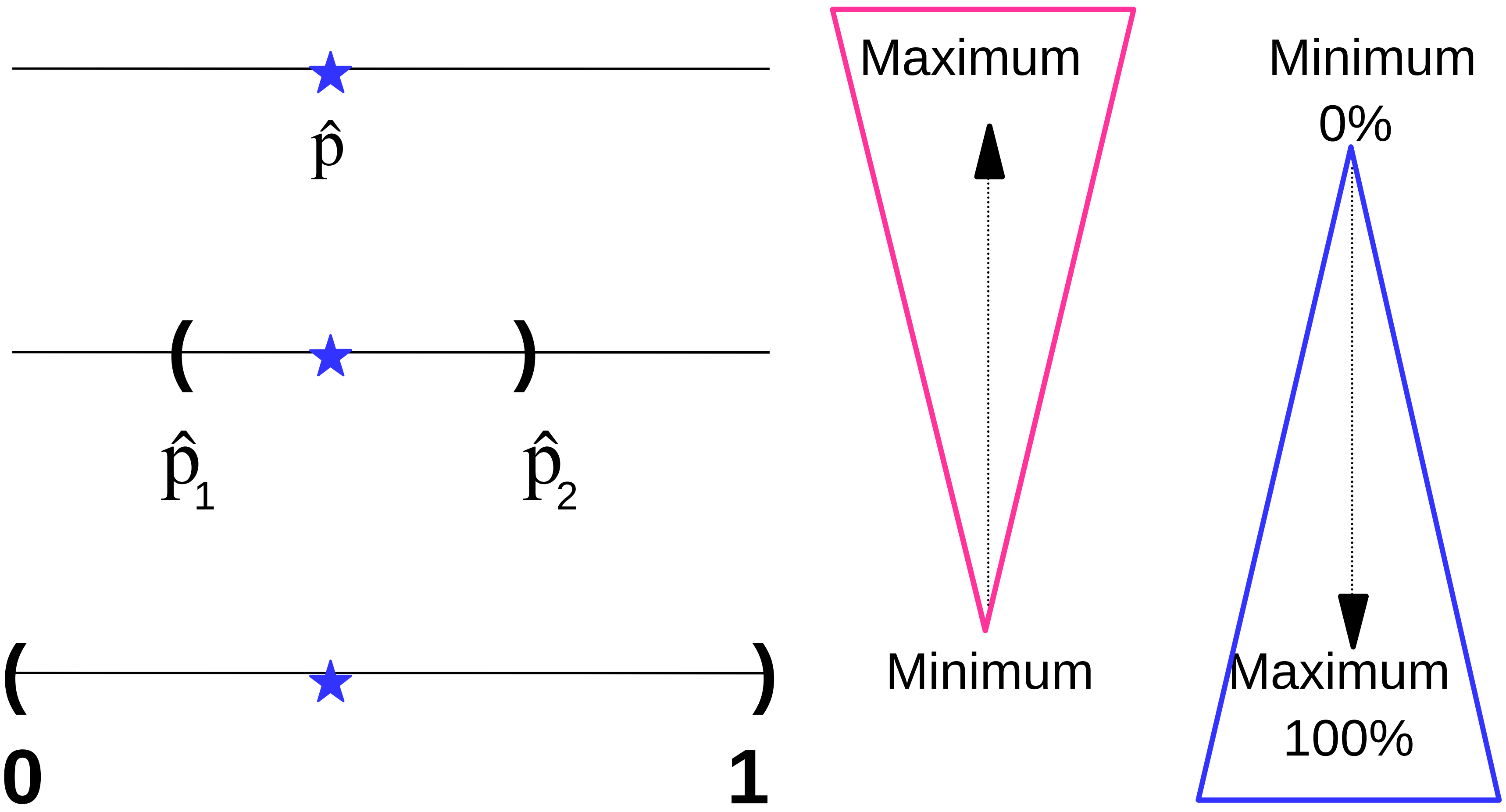
B) Interval Estimate

Specify a range within which the parameter is estimated to lie

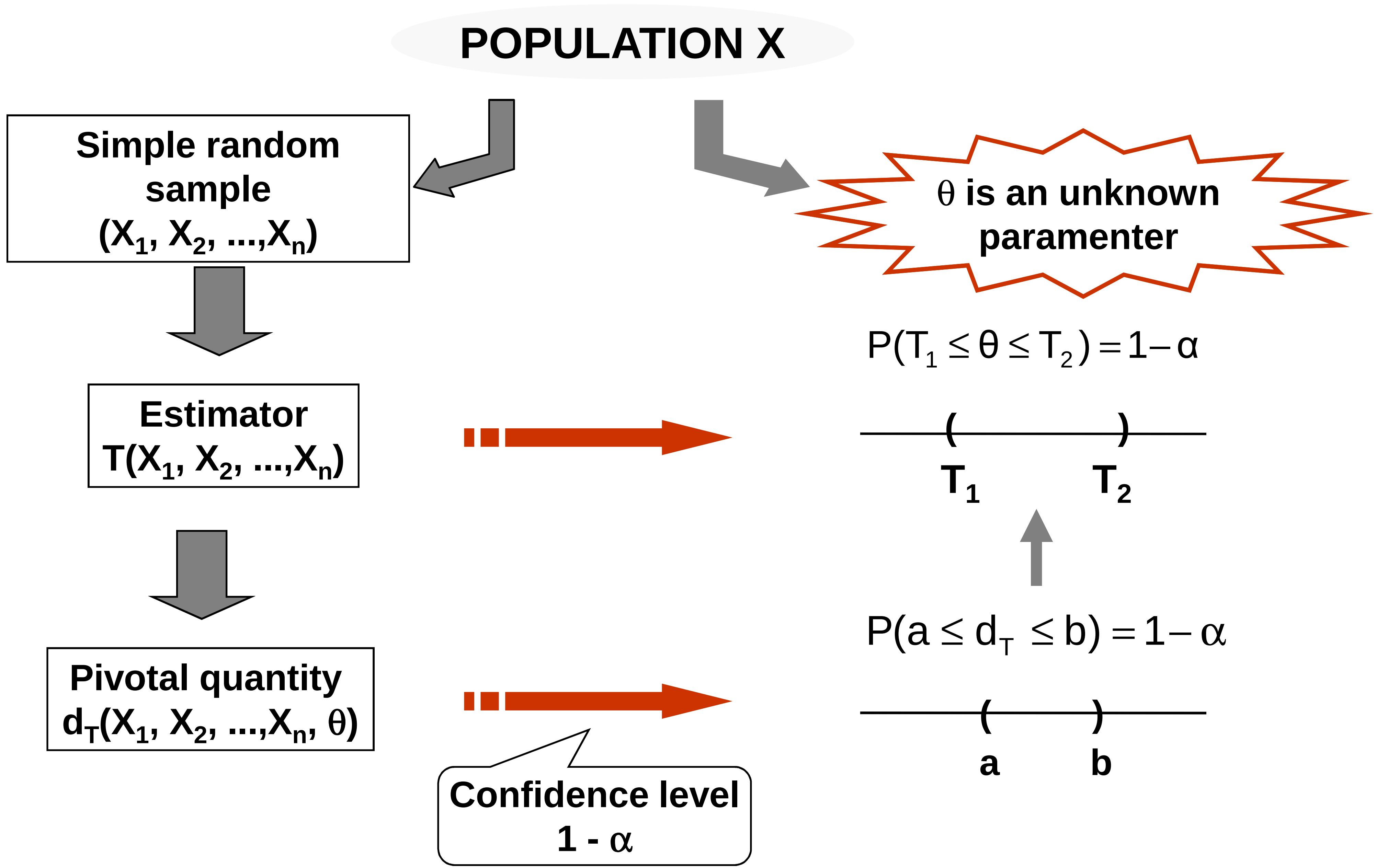


Relationship between precision and confidence level

Estimation of p | **Precision** | **Confidence**



Building Confidence Intervals



Example. Building a confidence interval for the population mean

POPULATION $X \approx N(\mu,\sigma)$
σ Is known

Simple random sample
$(X_1, X_2, \ldots, X_n)$

μ is unknown

Estimator

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

$\bar{X}$

Pivotal quantity

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

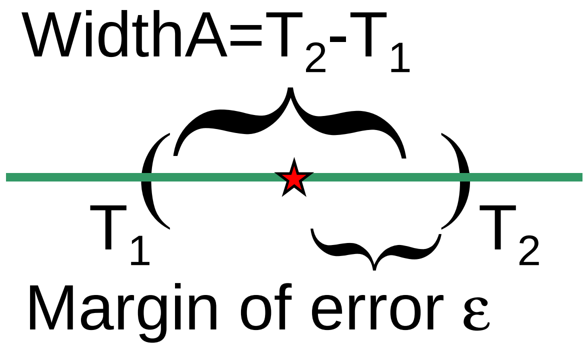
$$P(-1.96 \le d_{N(0,1)} \le 1.96) = 0.95$$

-1.96 **1.96**

Confidence level
$1 - \alpha = 0.95$

Interval Precision

Interval precision is measured through the width of the interval, A. In case of a symmetric interval, precision can be then evaluated through the margin of error, ε=*A*/2



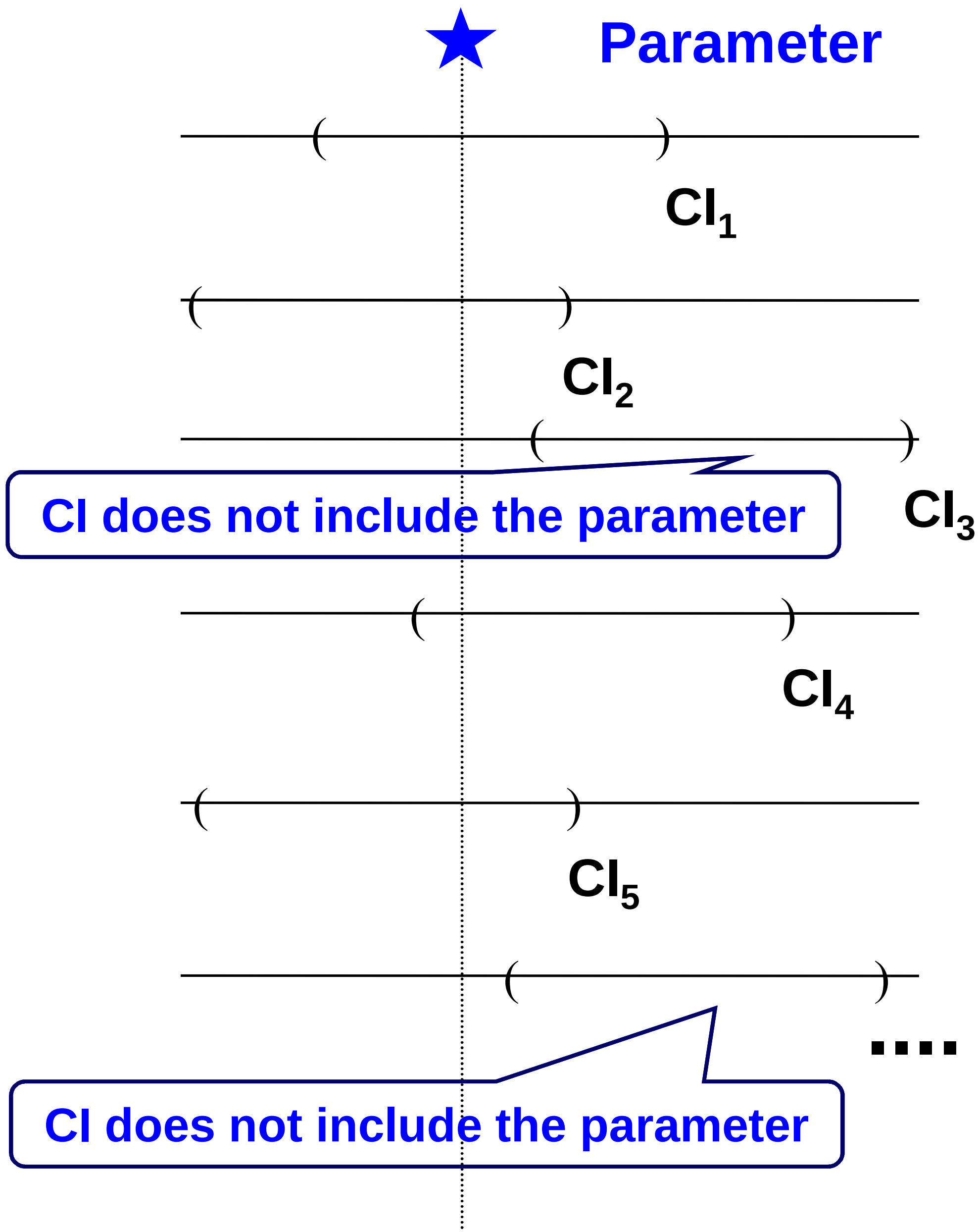
Factors influencing precision

- **Confidence level**
- **Information about the population (probability model and parameters)**
- **Sample information (size, sampling methods, estimator)**

Interpretation of confidence level (1-α)

"If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, THEN **a (1-α)% of the intervals will include the unknown population parameter".**

Statistics Glossary
http://www.stats.gla.ac.uk/



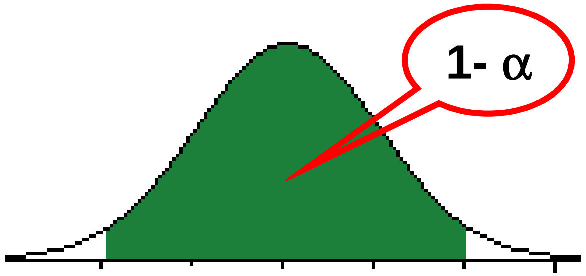
Confidence interval for population mean

NORMAL POPULATION
σ is known

POPULATION IS NOT NORMAL
σ is known and n is large

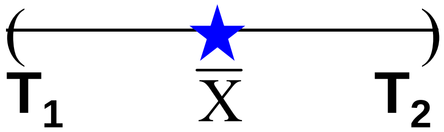
$$d_{\bar{X}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

Standard Normal Distribution N(0,1)



Obtaning value k

$$\boxed{P(-k \leq N(0,1) \leq k) = 1-\alpha}$$



$$\boxed{P(T_1 \leq \mu \leq T_2) = 1-\alpha}$$

Confidence interval for population mean with a confidence level 1 - α

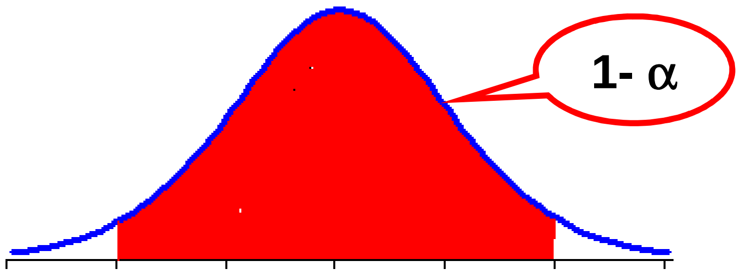
$$\left[\bar{X}-k\frac{\sigma}{\sqrt{n}}, \bar{X}+k\frac{\sigma}{\sqrt{n}}\right]$$

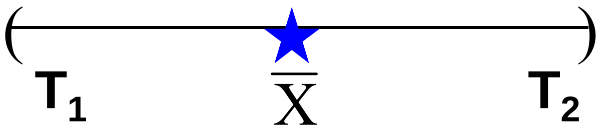
Confidence interval for population mean

NORMAL POPULATION
σ is unknown

$$d_{\bar{X}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \approx t_{n-1}$$

Student's t-distribution





Obtaining value k

$$\boxed{P(-k \leq t_{n-1} \leq k) = 1-\alpha}$$

$$\boxed{P(T_1 \leq \mu \leq T_2) = 1-\alpha}$$

Confidence interval for population mean with a confidence level 1 - α

$$\left[\bar{X}-k\frac{S}{\sqrt{n}}, \bar{X}+k\frac{S}{\sqrt{n}}\right]$$

CI for μ when the distribution of the population is unknown

Tchebysheff's inequality allows to build confidence intervals when the distribution of the population is not known, σ is known and n is small

$$P\left(\left|d_{\bar{X}} - E\left(d_{\bar{X}}\right)\right| < k\sigma_{d_{\bar{X}}}\right) \geq 1 - \frac{1}{k^2} = 1 - \alpha \Rightarrow \boxed{k = \frac{1}{\sqrt{\alpha}}}$$

$$d_{\bar{X}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad E\left(d_{\bar{X}}\right) = 0 \qquad Var\left(d_{\bar{X}}\right) = 1$$

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

Confidence interval for population mean with a confidence level 1 - α

$$\left[\bar{X} - \frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}},\ \bar{X} + \frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}}\right]$$

Confidence intervals for population mean

Situation	Confidence Interval	Confidence Level
X is Normal, σ is known X is not Normal, σ is known and n is large	$\left[\bar{X}-k\frac{\sigma}{\sqrt{n}},\bar{X}+k\frac{\sigma}{\sqrt{n}}\right]$ k is found in N(0,1) tables	$1-\alpha$
X is Normal, σ is unknown	$\left[\bar{X}-k\frac{S}{\sqrt{n}},\bar{X}+k\frac{S}{\sqrt{n}}\right]$ k is found is Student's t-distribution (d.f.= n-1)	$1-\alpha$
X follows an unknown distribution, σ is known and n is small	$\left[\bar{X}-\frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}},\bar{X}+\frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}}\right]$	$\geq 1-\alpha$

Sample size for estimating μ

GOAL

To estimate μ, given a certain confidence level (1-α) and certain margin of error ε

$X \approx N(\mu, \sigma)$

Sample size

$$\left[\bar{X} - k\frac{\sigma}{\sqrt{n}}, \bar{X} + k\frac{\sigma}{\sqrt{n}}\right] \Rightarrow \varepsilon = k\frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{n = \left(\frac{k\sigma}{\varepsilon}\right)^2}$$

X is not known

$$\left[\bar{X} - \frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}}, \bar{X} + \frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}}\right] \Rightarrow \varepsilon = \frac{1}{\sqrt{\alpha}}\frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{n = \frac{\sigma^2}{\alpha\varepsilon^2}}$$

- Sample size increases with confidence level
- Sample size increases with the value of the standard deviation
- Sample size increases as the margin of error ε decreases

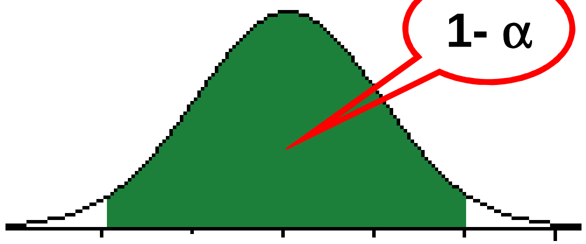
Confidence interval for population proportions

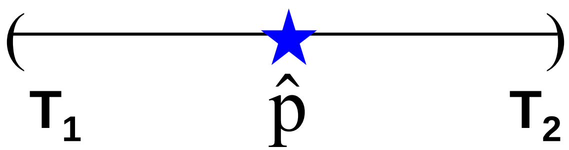
n is large

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

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$$

Standard Normal Distribution
N(0,1)





Obtaning value k

$$\boxed{P(-k \le N(0,1) \le k) = 1 - \alpha}$$

$$\boxed{P(T_1 \le p \le T_2) = 1 - \alpha}$$

Confidence interval for population proportion with a confidence level 1 - α

$$\left[\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}, \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\right]$$

Sample size for estimating p

Goal

To estimate p, given a certain confidence level (1-α) and certain margin of error ε

$$\left[\hat{p} - k\sqrt{\frac{p(1-p)}{n}},\ \hat{p} + k\sqrt{\frac{p(1-p)}{n}}\right]$$

Sample size

$$\varepsilon = k\sqrt{\frac{p(1-p)}{n}} \Rightarrow n = k^2\frac{p(1-p)}{\varepsilon^2}$$

Since p is unknown, there are to basic options to approximate p(1-p):

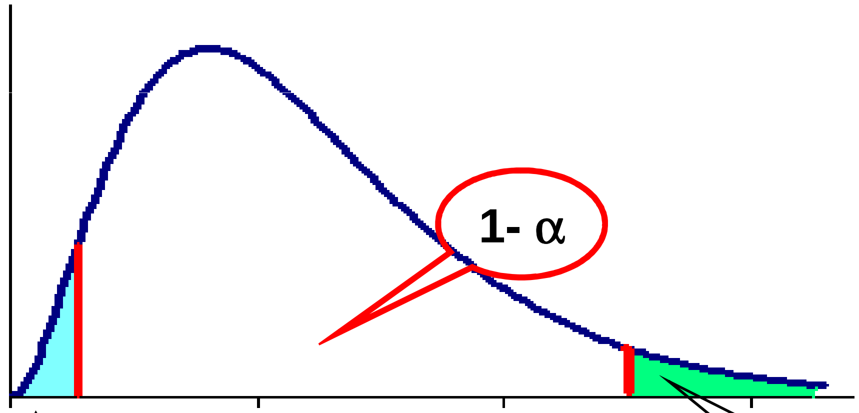
- **Carry out a pilot survey in order to a get a preliminar estimation of p**
- **The upper limit of p(1-p) is assumed, p(1-p)= 0.25**

Confidence interval for population variance

NORMAL POPULATION

$$d_{S^2} = \frac{(n-1)S^2}{\sigma^2} \approx \chi^2_{n-1}$$

Chi-square distribution



k_1 k_2

$$P(\chi^2_{n-1} \leq k_1) = \frac{\alpha}{2}$$

$$P(\chi^2_{n-1} \geq k_2) = \frac{\alpha}{2}$$

Obtaining values k_1 and k_2

$$\boxed{P(k_1 \leq \chi^2_{n-1} \leq k_2) = 1-\alpha}$$

Confidence interval for population variance with a confidence level 1 - α

$$\left[\frac{(n-1)S^2}{k_2}, \frac{(n-1)S^2}{k_1}\right]$$