

ESTIMATION

CĂLINICI TUDOR

2026

OBJECTIVES

- Understanding the estimation process.
- Learning the terminology of statistical inference.
- Listing unbiased estimators.
- Determining a confidence interval with a specified precision

GENERAL PRINCIPLES

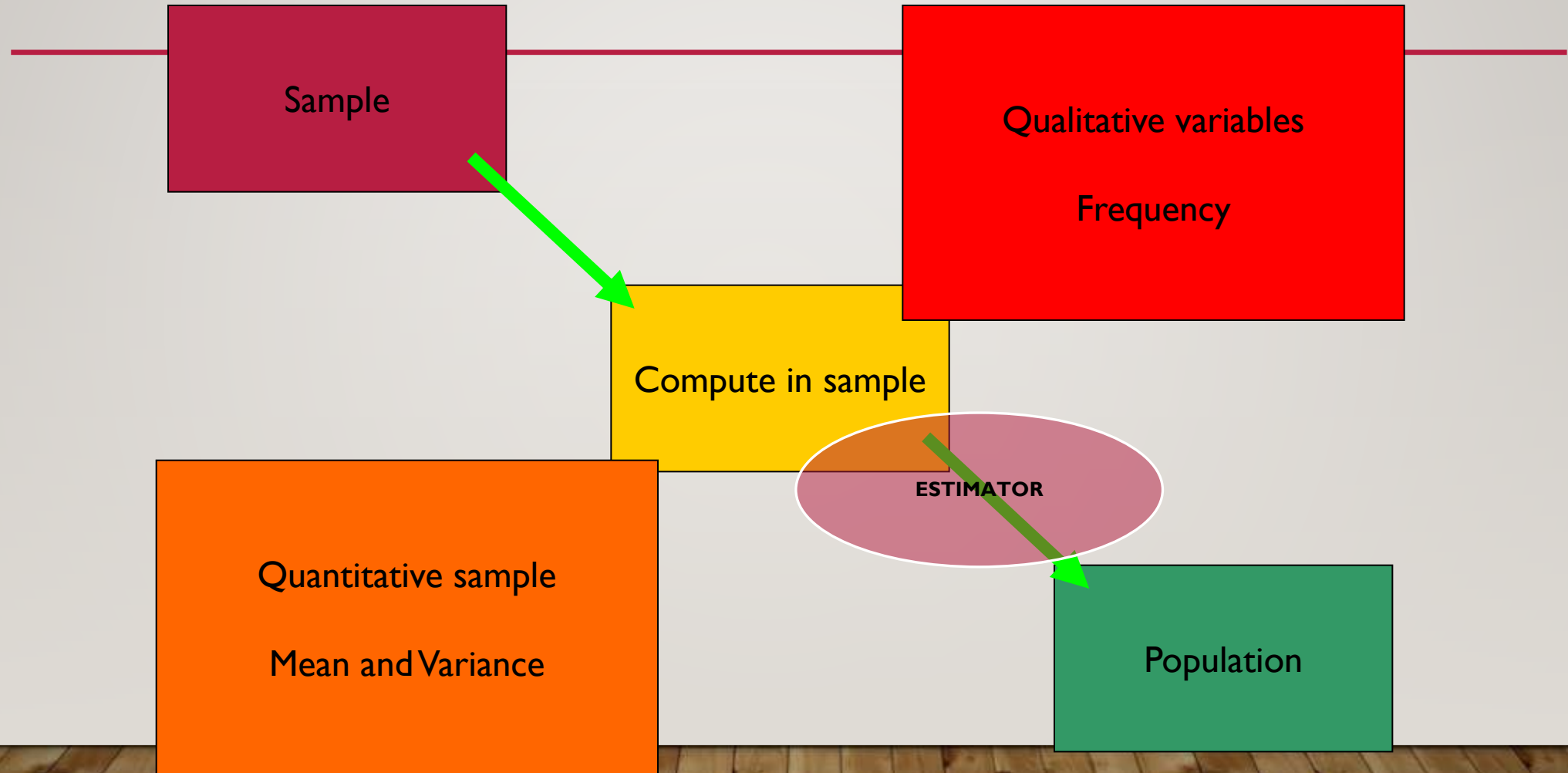
When studying the parameters of a characteristic (quantitative or qualitative) in a population P , it is often necessary to follow this procedure:

- Draw a representative sample from the population.
- Use descriptive statistics to describe the distribution of the characteristic in the sample drawn in step 1. For a qualitative characteristic, determine the observed frequency; for a quantitative characteristic, calculate the mean and variance.
- Use inferential or inductive statistics to generalize the results observed in the sample to the entire population. The sample statistics (frequency, mean, variance, etc.) are used to estimate the “theoretical” parameters of the entire population.

ESTIMATION

- An **estimator** of a parameter is a function of the observations in a randomly selected sample that provides a value called the *local estimate of the parameter*.
- We will compute the local estimation in the sample (the statistic) and we will estimate the value in population (the estimator)

GENERAL PRINCIPLES



ESTIMATION

Qualities for the estimator

- *correct*
- *precise*

LOCAL ESTIMATION

The mean and the frequency are correct estimators!

LOCAL ESTIMATION

- The variance is NOT a correct estimator
- The sampling variance is a correct estimator

$$S^2 = \frac{n}{n-1} s^2$$

ESTIMATION USING CONFIDENCE INTERVALS

- Local estimation of a population parameter provides a single value for that parameter.
- The value of the point estimator is subject to sampling variability and may differ considerably from the true value of the parameter being estimated.
- It is advisable to estimate a population parameter using an interval, called a *confidence interval*, rather than a single value. This interval is constructed using a method that captures the true parameter with a high probability over repeated sampling.

CONFIDENCE INTERVAL

- A confidence interval is a range of values bounded by two endpoints (called confidence limits) that contains the point estimate of the characteristic under study.
- The wider the interval, the more confident we are that it contains the true value of the parameter being estimated.
- The confidence level indicates the proportion of intervals that would contain the true parameter value if the sampling procedure were repeated many times.

WHEN WE USE CONFIDENCE INTERVALS

- Confidence intervals are used when we need to estimate means, proportions, or other parameters in the POPULATION.

-

CONFIDENCE

- How confident am I that my estimate is correct?
- “Critical value” – the higher the confidence level, the higher this value.
 - We will compute a statistical parameter starting from the sample, and according to the level of confidence we will multiply it with the corresponding critical value
 - Example: mean, large intervals the parameter is SE, the critical value will be for 85% - 1.440 ; for 90% - 1.645; for 95% - 1.96; for 97% - 2.17; for 99%- 2.58
- High confidence $\Rightarrow$ wide confidence interval $\Rightarrow$ low precision

ESTIMATION OF THE MEAN IN POPULATION –LARGE SAMPLES

- $N \geq 30$, m - sample mean s – sample standard deviation
 - The confidence interval for mean having the level of significance α is
-

$$[m - Z_{\alpha} * Se, m + Z_{\alpha} * Se]$$

- Se – Standard error

$$Se = \frac{s}{\sqrt{n - 1}}$$

ESTIMATION OF THE MEAN IN POPULATION –LARGE SAMPLES

- The most commonly used significance level is $\alpha = 0.05$. In this case, the critical Z-value is 1.96, so the most commonly used confidence interval for large samples is:

$$\mu \in \left[m - 1,96 \frac{s}{\sqrt{n-1}}, m + 1,96 \frac{s}{\sqrt{n-1}} \right]$$

QUESTION

- We observe a sample of 101 newborns with a mean birth weight of 3,400 g and a standard deviation of 142 g.
- The literature reports a mean birth weight of 3,300 g.
- Is this sample representative of the population described in the literature?
 - **Lower limit = $3400 - 1.96 \times 142/10 = 3372.16$ g**
Upper limit = $3400 + 1.96 \times 142/10 = 3427.83$ g
- We are 95% confident that the mean birth weight in the population from which the sample was drawn lies between 3372.16 and 3427.83 g. The value reported in the literature falls outside this interval; therefore, at the 95% confidence level, this sample is not representative of the population described in the literature.

STANDARD ERROR AND THE PRECISION

- An increase in the standard error $\Rightarrow$
Lower precision of the sample mean in estimating the population mean.
- A decrease in the standard error $\Rightarrow$
Higher precision of the sample mean in estimating the population mean.

ESTIMATION BASED ON SMALL-SIZE INTERVALS

- $n < 30$
- Student (T) distribution is used

DEGREES OF FREEDOM

- df
- For a sample size n, df=n-1

$$\left[m - t_{\alpha} \frac{s}{\sqrt{n-1}}, m + t_{\alpha} \frac{s}{\sqrt{n-1}} \right]$$

CALCULUL INTERVALULUI DE ÎNCREDERE

$$\left[m - t_{\alpha} \frac{s}{\sqrt{n-1}}, m + t_{\alpha} \frac{s}{\sqrt{n-1}} \right]$$

- $n=6 \rightarrow df = 5$
 - $\parallel 95\% \rightarrow t = \pm 2,571$
- $n=10 \rightarrow df = 9$
 - $\parallel 95\% \rightarrow t = \pm 2,262$
- $n=30$
 - $\parallel 95\% \rightarrow t = \pm 2,042$

VALUES OF t FOR 90%, 95%, AND 99% CONFIDENCE INTERVALS

df	90%	95%	99%
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

EXAMPLE

- In a sample of 10 people, we observed that, after receiving treatment for hypertension, the mean blood pressure was 115 mmHg, with a standard deviation of 15 mmHg. With 95% probability, what blood pressure value can be expected for a person who receives this treatment?
- Size = 10, df = 9, $T_{95\%} = 2,262$
- Lower bound of 95%CI = $115 - 2,262 * \frac{15}{\sqrt{10-1}} = 103,69$
- Upper bound of 95% CI = $115 + 2,262 * \frac{15}{\sqrt{10-1}} = 126,31$

ESTIMATING THE FREQUENCY IN POPIULATION

- Big samples: $n \cdot f, n(1-f) \geq 10$

$$\left[f - Z_{\alpha} \sqrt{\frac{f(1-f)}{n}}, f + Z_{\alpha} \sqrt{\frac{f(1-f)}{n}} \right]$$

- f = frequency (%)
- For 95% confidence, $Z_{95\%} = 1,96$

EXAMPLE

- We wish to estimate the proportion of military recruits who experience adverse reactions to influenza vaccination. In a sample of 100 recruits, the observed prevalence of adverse reactions was 5%. Given that 10,000 recruits will be vaccinated, determine, with 95% confidence, the expected number of recruits who will develop adverse reactions.
- $f = 5\% = 0.05$ $1 - f = 0.95$ $n = 100$
95% confidence $\Rightarrow Z = 1.96$
- The confidence interval calculated from this sample:
 - Lower bound $0.05 - 1.96 * \sqrt{\frac{0.05 * 0.95}{100}} = 0.007$ - $0.007 * 10000 = 72$
 - Upper bound $0.05 + 1.96 * \sqrt{\frac{0.05 * 0.95}{100}} = 0.092$ - $0.092 * 10000 = 920$
- With 95% confidence, we expect between 70 and 920 soldiers to develop adverse reactions to influenza vaccination.

CONCLUSIONS

- When analyzing samples, we calculate local estimates; when estimating population parameters, we use confidence intervals!
- The precision of an estimate is determined by the width of the confidence interval:
 - Wide interval – low precision
 - Narrow interval – high precision
- In medicine, a 95% confidence level is commonly used – $Z = 1.96$.
- Increasing the confidence level decreases the precision of the estimate.
- Increasing the sample size increases the precision of the estimate.