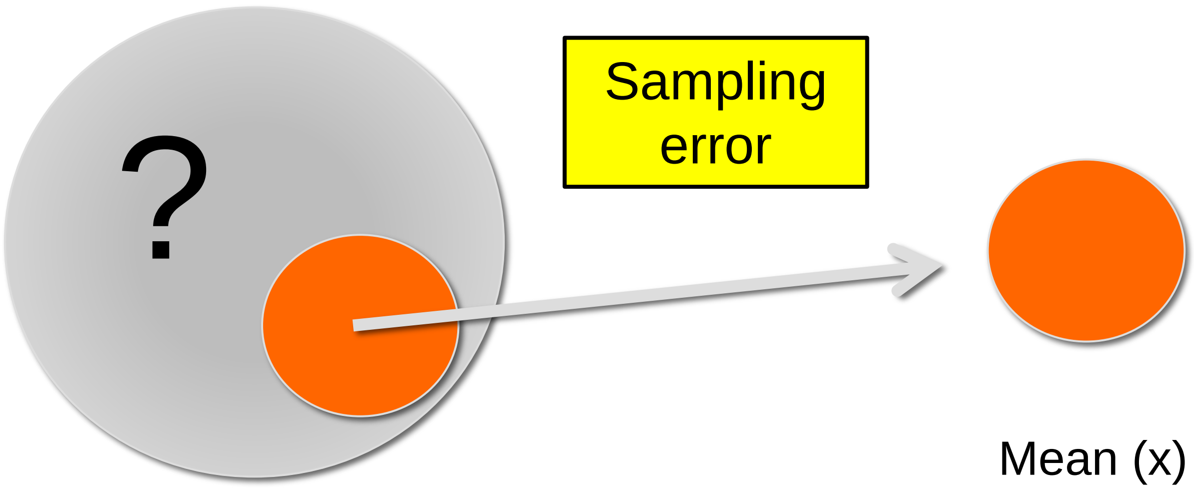
STATISTICAL INFERENCE, ESTIMATION AND HYPOTHESIS TESTING

Jose Andres Calvache MD MSc PhD

Department of Anesthesiology, Universidad del Cauca, Colombia

Department of Anesthesiology, Easmus University Medical Centre, Rotterdam, The Netherlands

Population and sample



Mean (μ)
Standard deviation (δ)
Proportion (P)
Incidence
Prevalence
RR
OR…

PARAMETERS

Mean (x)
Standard deviation (s)
Proportion (p)
Incidence
Prevalence
RR
OR…

ESTIMATES
“ESTIMANDS”

Random sample

- It is a random selection of subjects from a population
  - Technically "it is representative" of the population regardless of its size
  - In this sampling technique, each individual in the population is equally likely to be selected in the sample

Simple sampling

- Extracting a sample n from a population P is no easy task

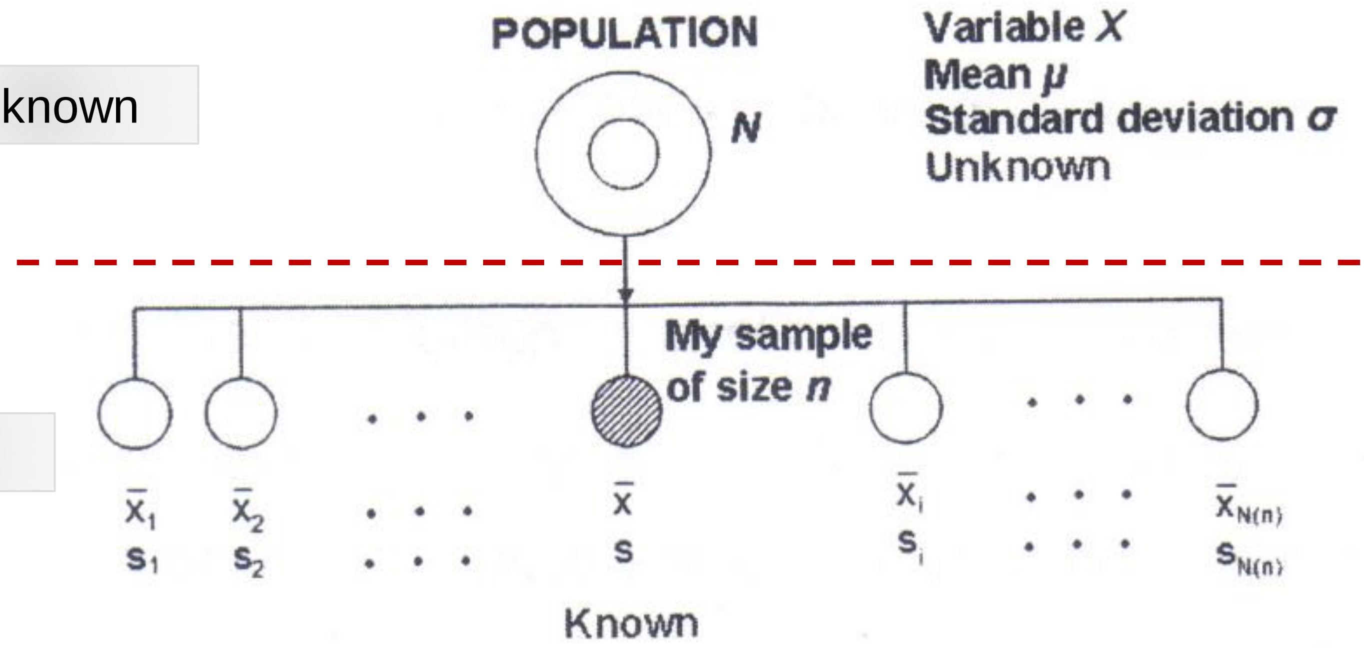
Always avoid human intervention in the selection of subjects, which would affect the inference process

- There are various sampling schemes
- Stratified, systematic, sequential

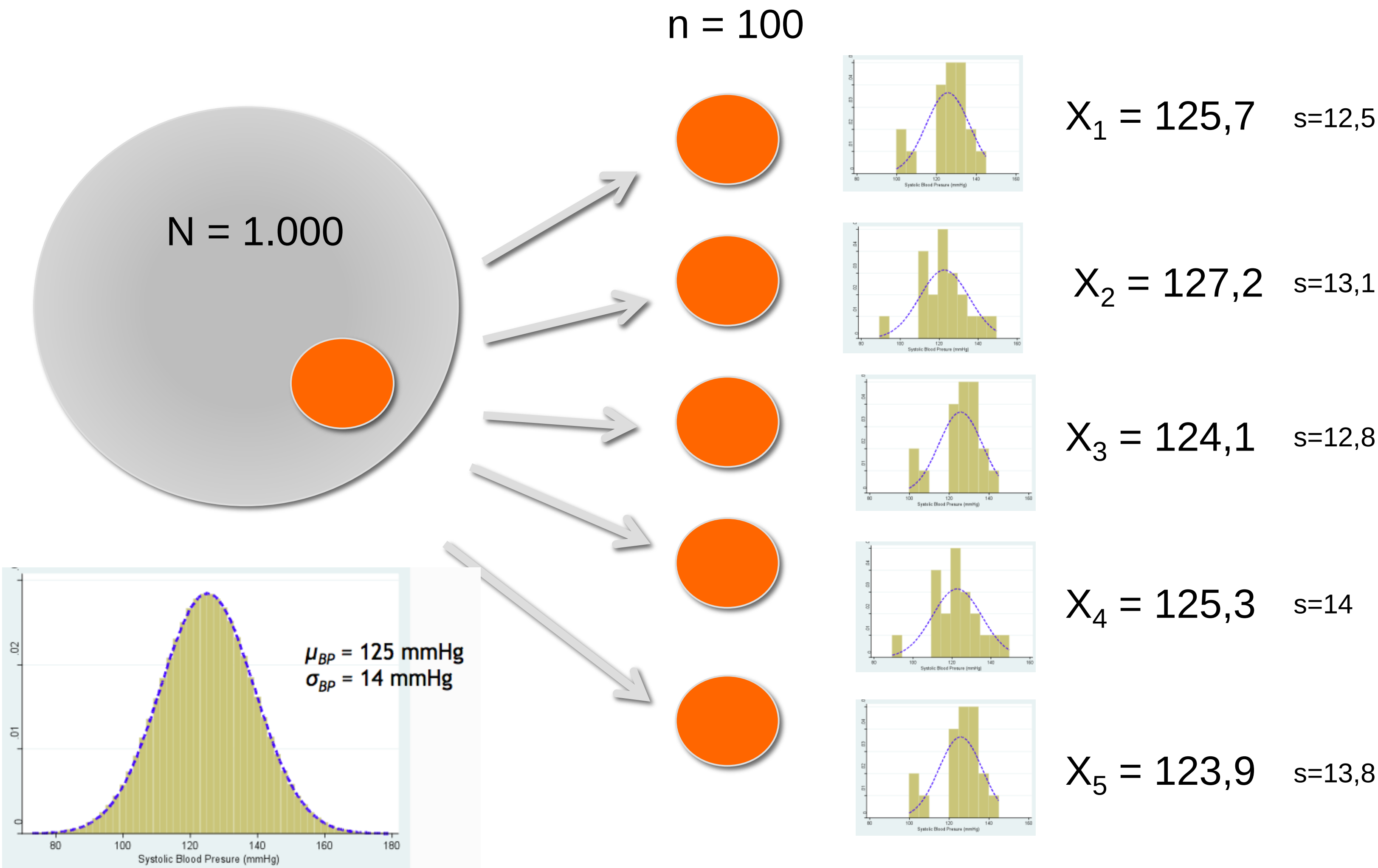
Sampling theory

Unknown

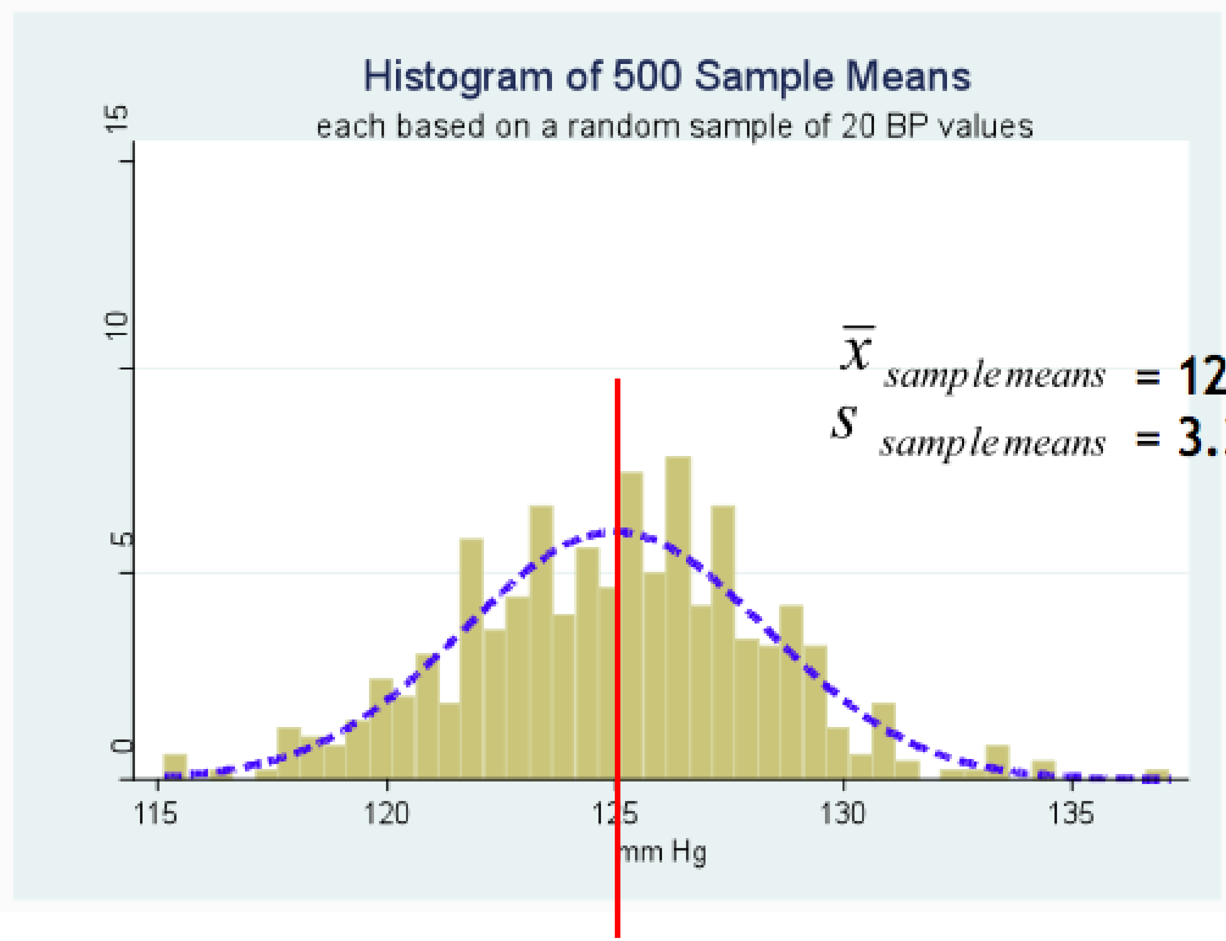
Known



Consider this situation,



"The mean" of the sampling means



$\bar{x}_{sample\,means}$ = 125 mmHg

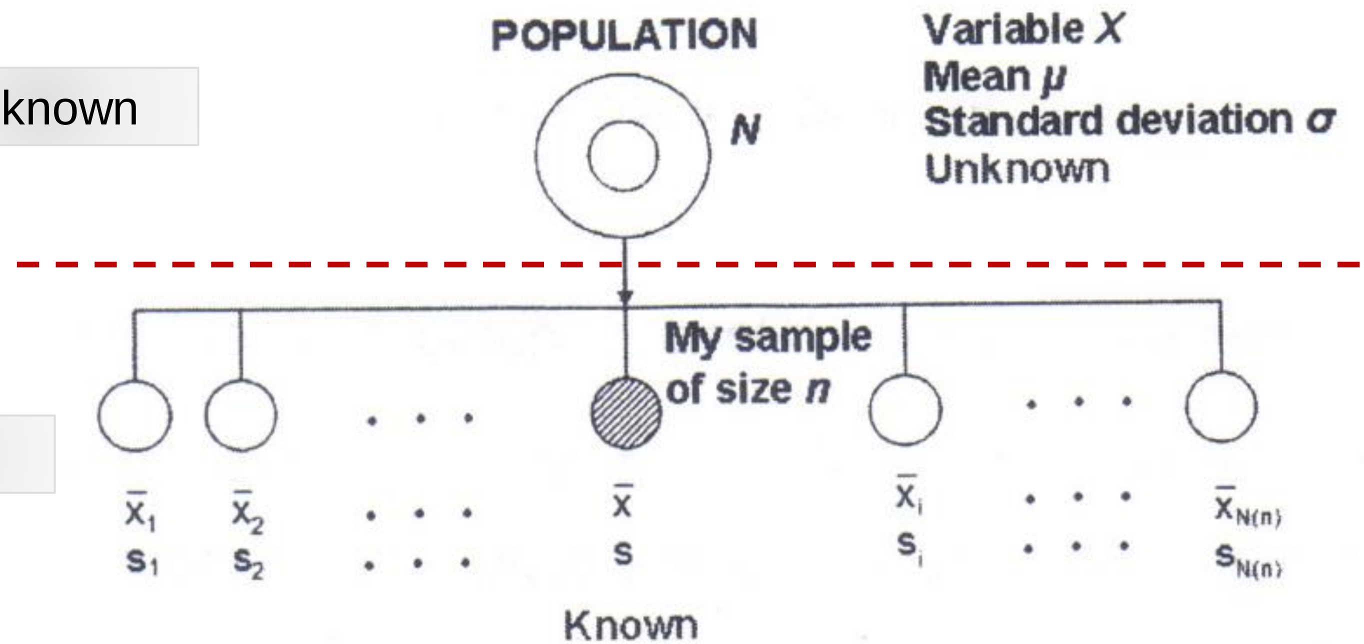
$s_{sample\,means}$ = 3.3 mmHg

Standard Error of the Error (SEM)

Sampling theory

Unknown

Known



1. Average of all estimates $\bar{x}_i = \sum \bar{x}_i / N(n) = \mu$ ($\bar{x}$ is unbiased estimate of μ)
2. The standard deviation of the means is called EE or SEM

$$SE(\bar{x}) = \sqrt{\frac{\sum(\bar{x}_i - \mu)^2}{N(n)}} = \frac{\sigma}{\sqrt{n}}$$

Standard error of the mean SEM

$$SE(\bar{x}) = \sqrt{\frac{\sum(\bar{x}_i - \mu)^2}{N(n)}} = \frac{\sigma}{\sqrt{n}}$$

Very important concept in statistics

- The greater the "σ" the SEM grows
- The higher the "n" the SEM decreases

- Precision depends on variability and n

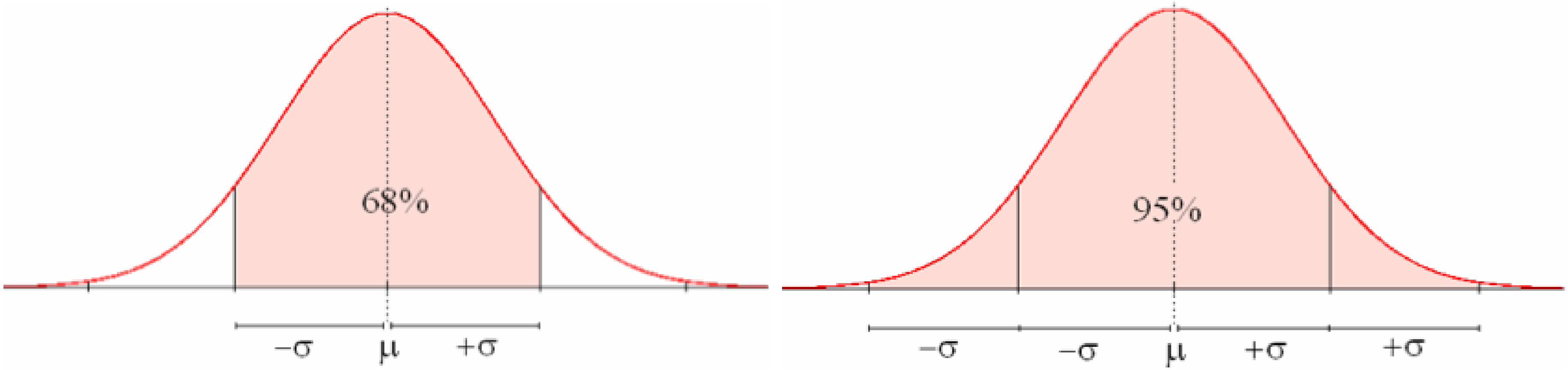
The normal distribution

- Organization of naturally appearing "population" data
- Systolic blood pressure
  - Body temperature
  - Size, weight

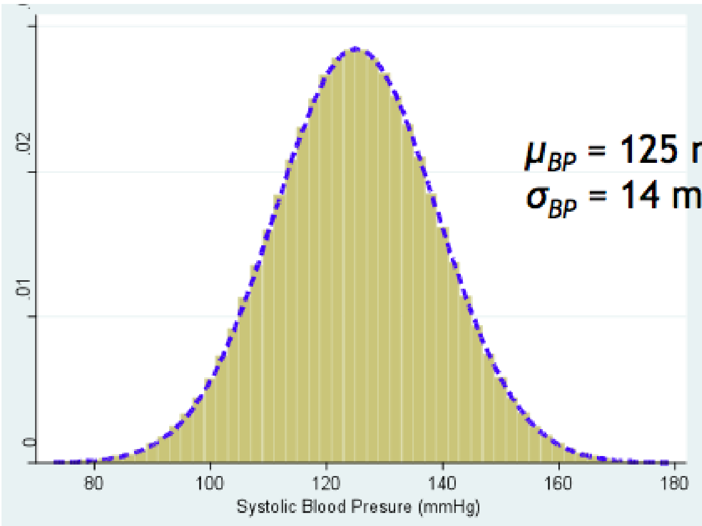
- It is characterized by two parameters: the mean (μ) and the standard deviation (σ)

- N (μ, σ) (centralization and variability)

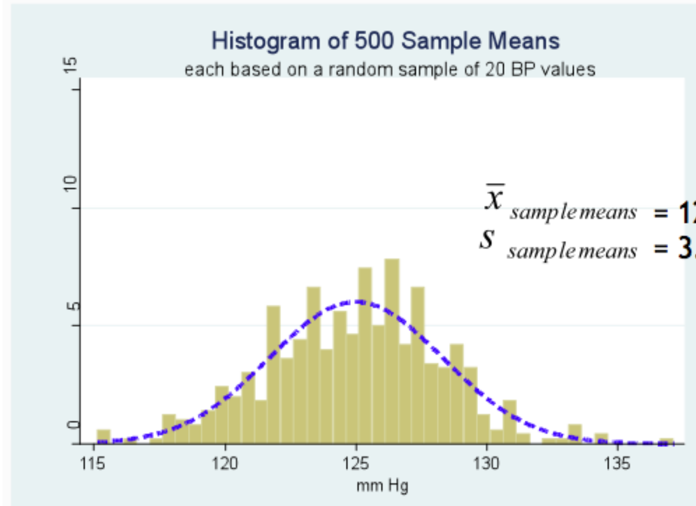
The normal distribution



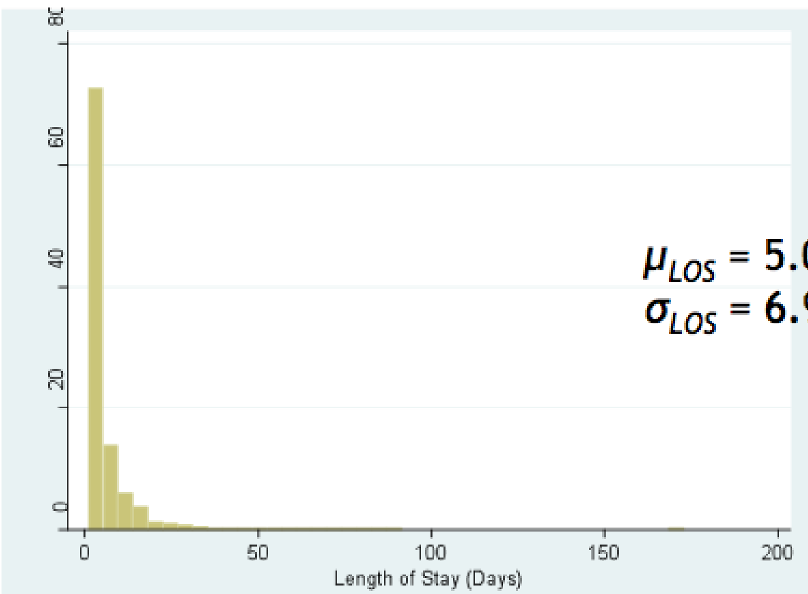
- Symmetrical and mean and median match
- At a distance from,
  - μ ± 1 σ = 68% of the data
  - μ ± 1,96 σ = 95% of the data
  - μ ± 2,5 σ = 99% of the data



μ_{BP} = 125 mmHg
σ_{BP} = 14 mmHg

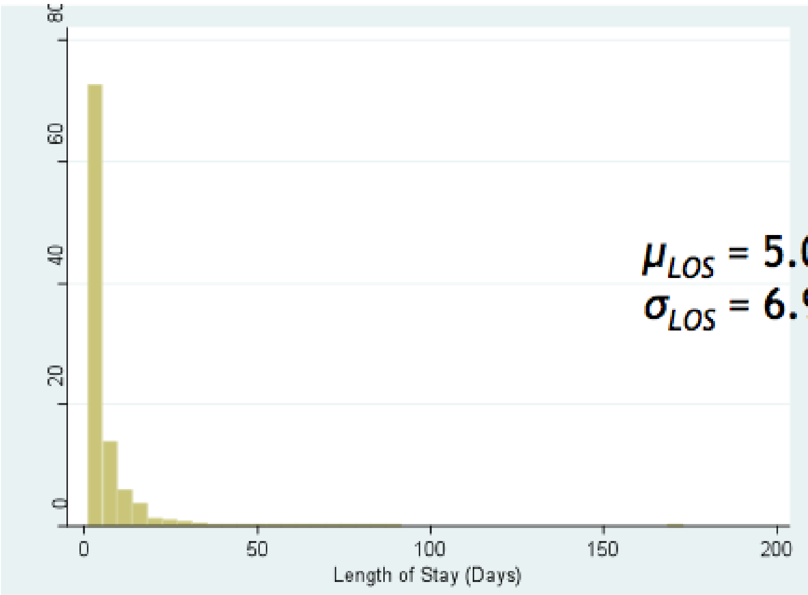


$\bar{x}_{\text{sample means}}$ = 125 mmHg
$s_{\text{sample means}}$ = 3.3 mmHg

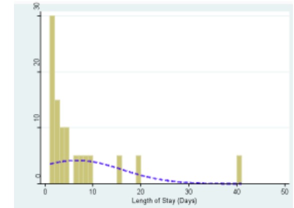


μ_{LOS} = 5.0 days
σ_{LOS} = 6.9 days

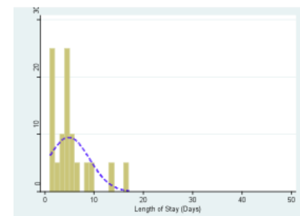
?



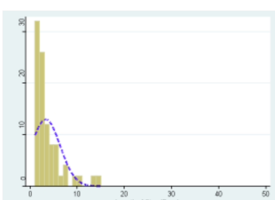
μ_{LOS} = 5.0 days
σ_{LOS} = 6.9 days



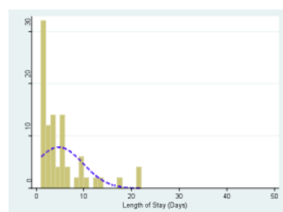
$\bar{x}_{LOS}$ = 6.6 days
s_{LOS} = 9.5 days



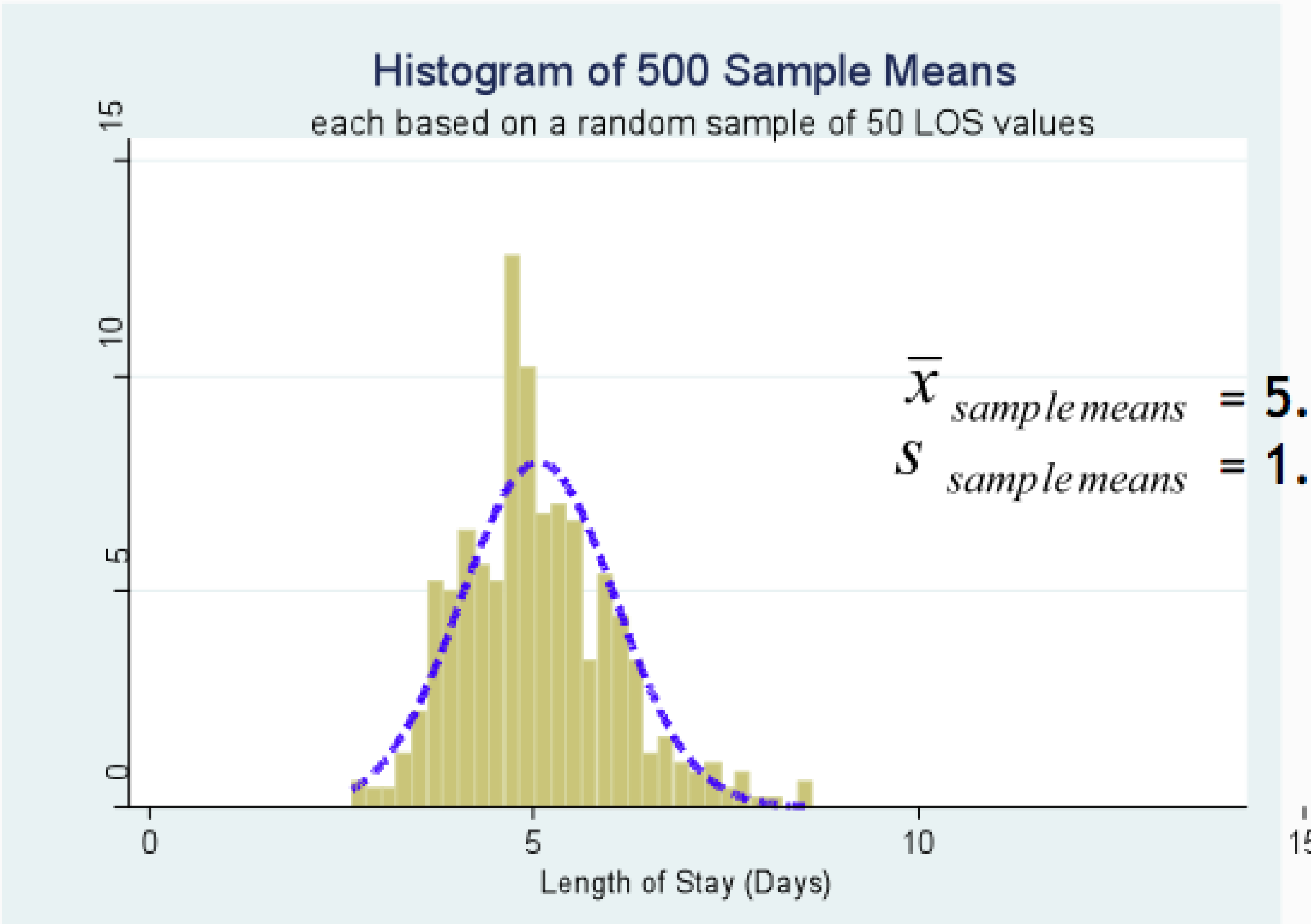
$\bar{x}_{LOS}$ = 4.8 days
s_{LOS} = 4.2 days



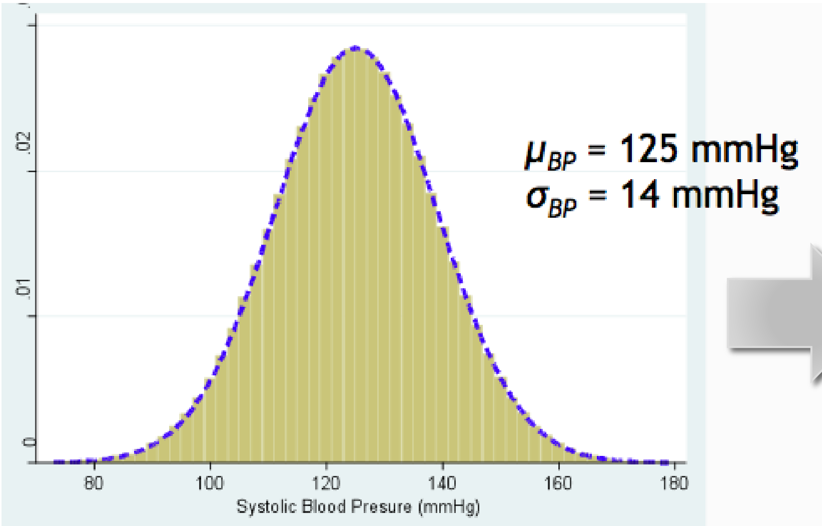
$\bar{x}_{LOS}$ = 3.3 days
s_{LOS} = 3.1 days

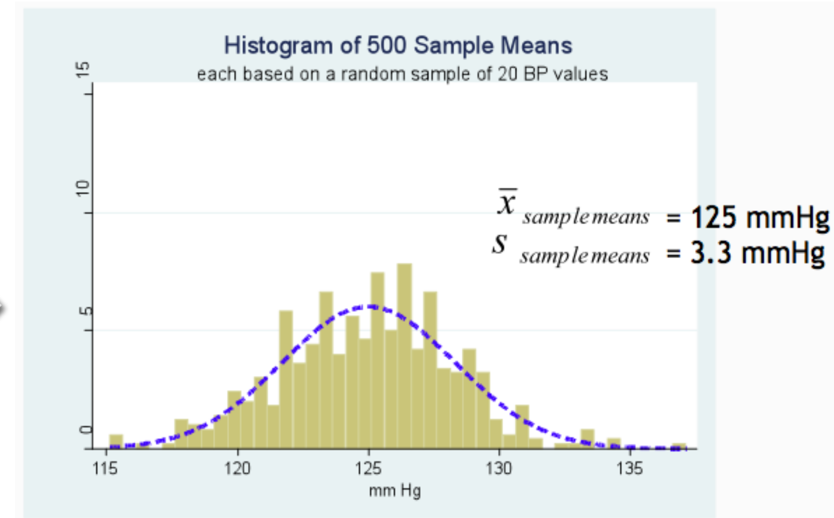


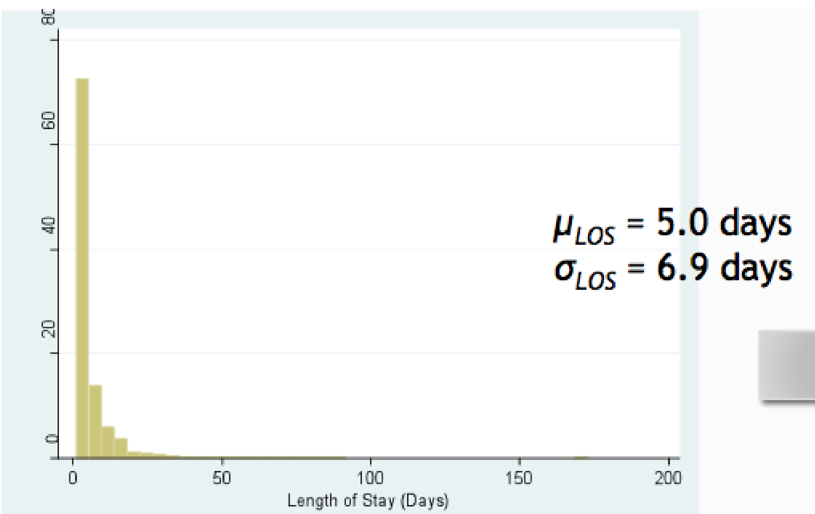
$\bar{x}_{LOS}$ = 4.7 days
s_{LOS} = 5.1 days

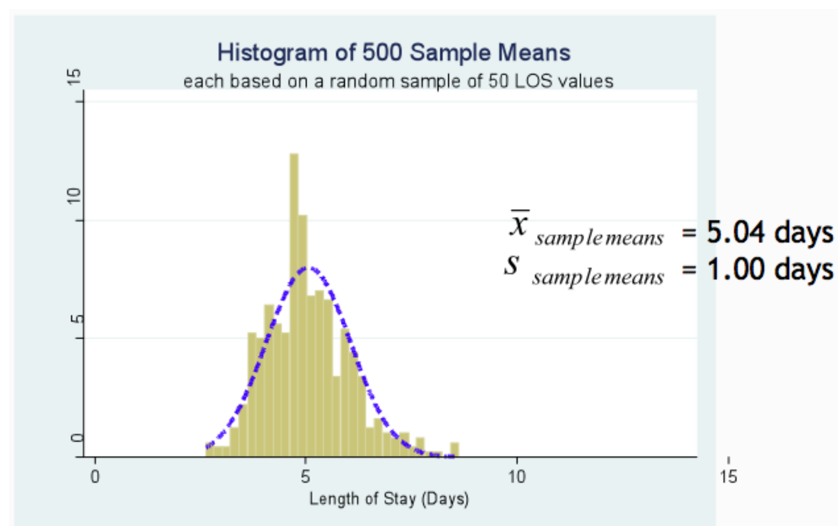


$$\bar{x}_{sample\,means} = 5.04 \text{ days}$$
$$s_{sample\,means} = 1.00 \text{ days}$$


μ_{BP} = 125 mmHg
σ_{BP} = 14 mmHg
Systolic Blood Presure (mmHg)
0
.01
.02
80
100
120
140
160
180


Histogram of 500 Sample Means
each based on a random sample of 20 BP values
$\bar{x}_{sample\,means}$ = 125 mmHg
$s_{sample\,means}$ = 3.3 mmHg
mm Hg
0
5
10
15
115
120
125
130
135


μ_{LOS} = 5.0 days
σ_{LOS} = 6.9 days
Length of Stay (Days)
0
20
40
60
80
0
50
100
150
200

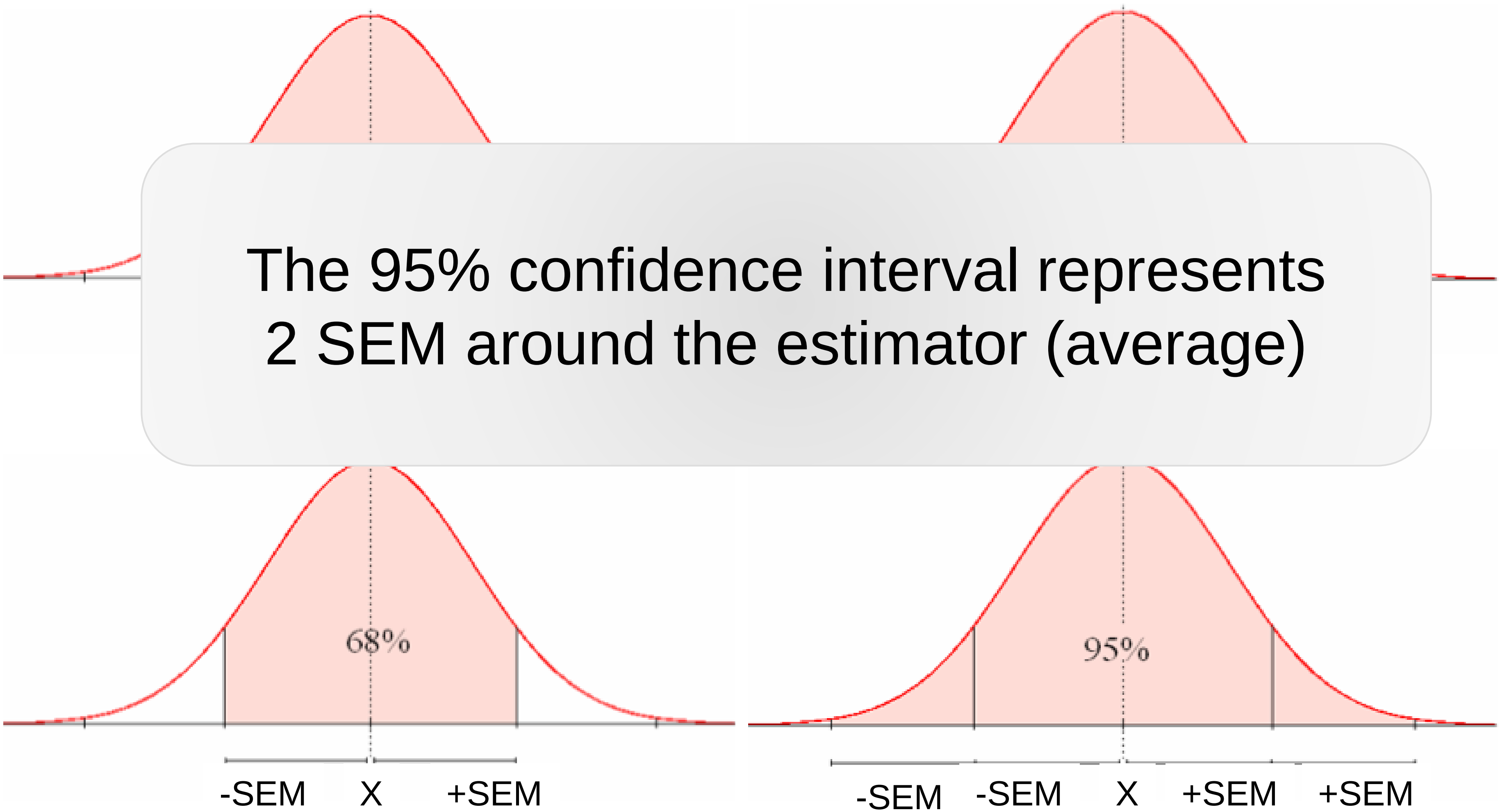

Histogram of 500 Sample Means
each based on a random sample of 50 LOS values
$\bar{x}_{sample\,means}$ = 5.04 days
$s_{sample\,means}$ = 1.00 days
Length of Stay (Days)
0
5
10
15
0
5
10
15

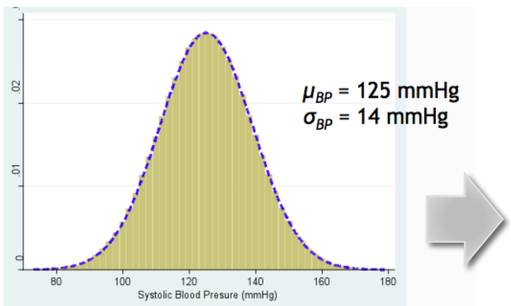
Why is the normal distribution important?

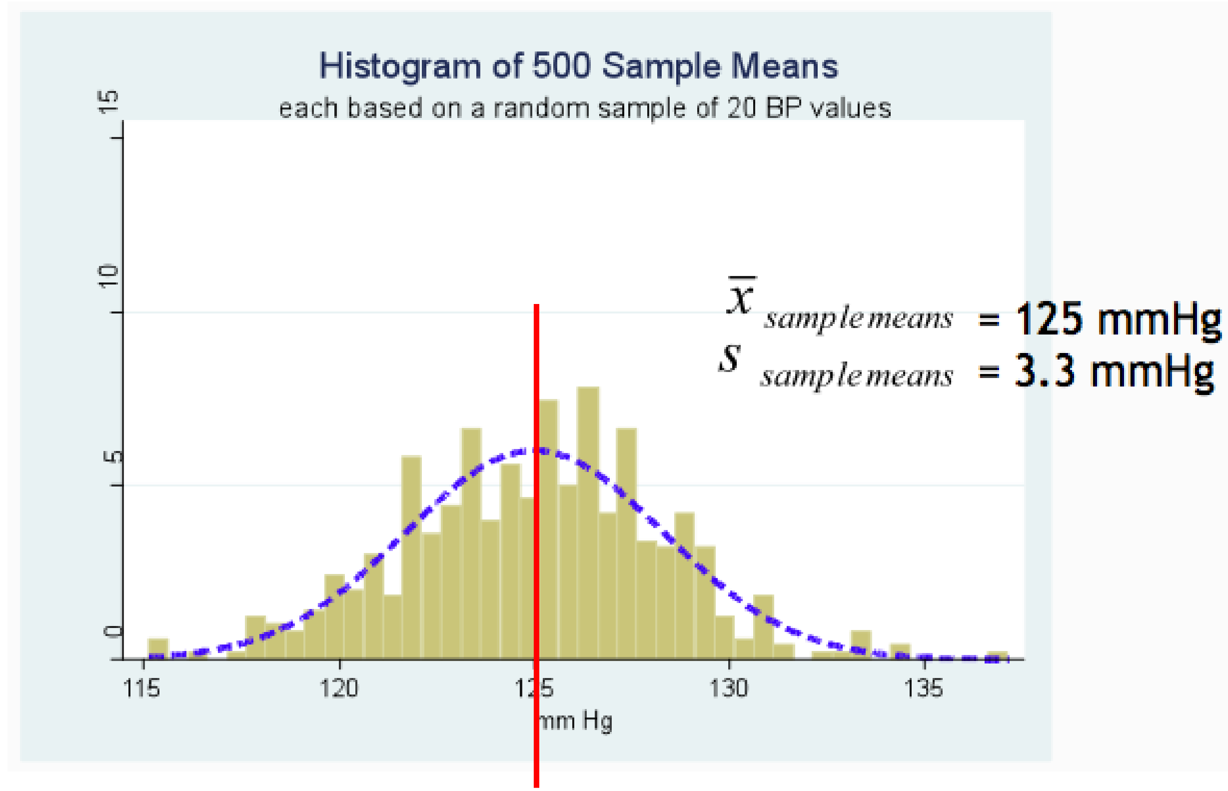
- Although a random variable does not have a "normal" population distribution, the distribution of certain data extracted from random samples DOES have it!
- Distributions of the sampling mean tend to be normally distributed even though the population data are not.
- Distribution of the sample mean is the best estimator of the population mean

Standard error of the mean (SEM)

The 95% confidence interval represents
2 SEM around the estimator (average)







Lower limit = X – 1,96 (SEM) = 125 – (6,4) = 118,6 mmHg
Upper limit = X + 1,96 (SEM) = 125 + (6,4) = 131,4 mmHg

Mean 125 mmHg, CI 95% (118 - 131)

95% confidence interval

- It is a measure of the dispersion of the estimate under study
- Represents a range in which the actual population parameter is found 95% of the time (frequentist)
- It is obtained from the standard error and can be calculated for any estimand: mean, proportions, OR, RR, r2, etc.
- More useful than the same estimator because it evaluates "uncertainty"

95% confidence interval

- Provides a range of possible values for the parameter
- The researcher will never be able to observe the real parameter
- "X" is the best estimate you can have from your sample
- The 95% CI starts with this estimate and also provides the degree of uncertainty in this amount.

Por ejemplo, (p)

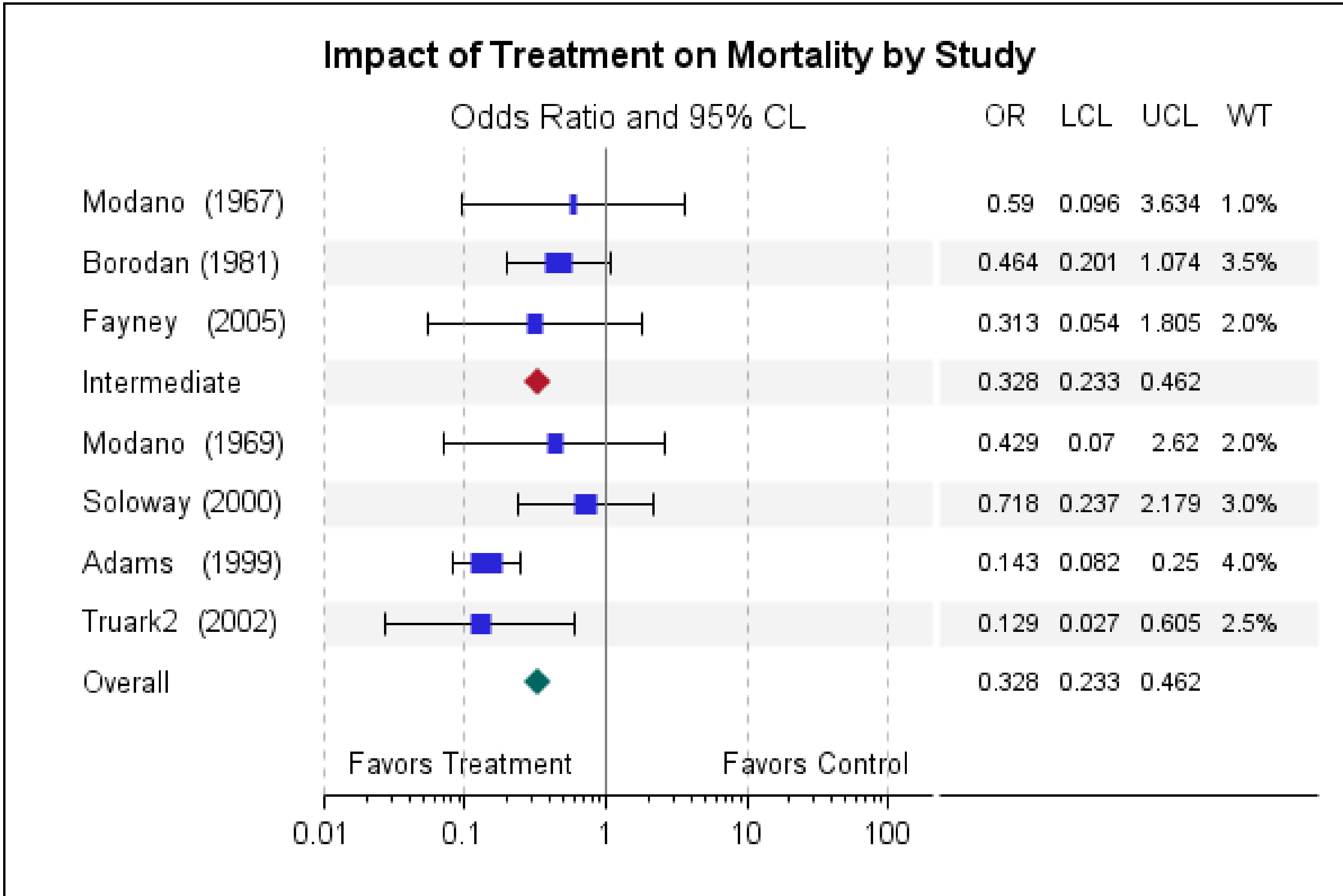
Table 2 Frequency of the incidence of inadequate postoperative analgesia.

Frequency	*n* (to 798, %)	95% CI of percentage
Never	457 (57.3)	48.3–68.6
Rarely	241 (30.2)	22.5–37.1
Occasionally	85 (10.6)	5.2–16.4
Frequently	15 (1.9)	0.8–3.7

Rarely: 1–2 times 1 month; occasionally: 2–5 times 1 month; frequently: >5 times 1 month.

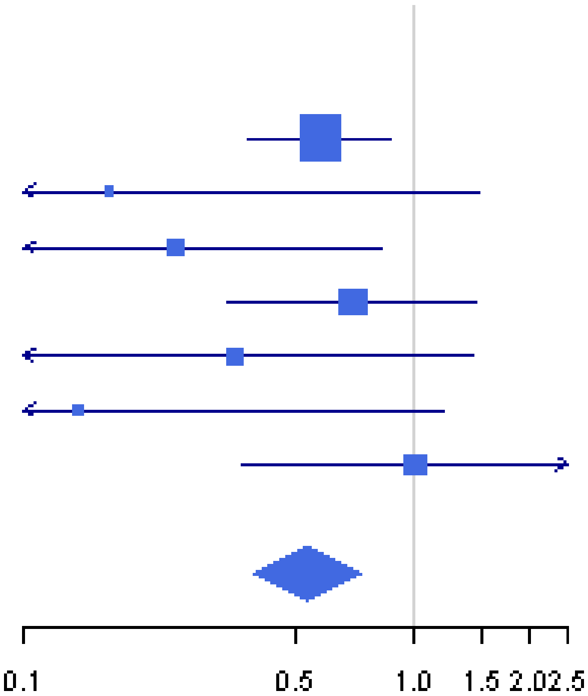
Prevalence of moderate–severe pain

The calculated prevalence rate of moderate–severe pain in this population was 27% (95% CI, 23% to 32%).



Impact of Treatment on Mortality by Study
Odds Ratio and 95% CL
OR LCL UCL WT
Modano (1967) 0.59 0.096 3.634 1.0%
Borodan (1981) 0.464 0.201 1.074 3.5%
Fayney (2005) 0.313 0.054 1.805 2.0%
Intermediate 0.328 0.233 0.462
Modano (1969) 0.429 0.07 2.62 2.0%
Soloway (2000) 0.718 0.237 2.179 3.0%
Adams (1999) 0.143 0.082 0.25 4.0%
Truark2 (2002) 0.129 0.027 0.605 2.5%
Overall 0.328 0.233 0.462
Favors Treatment
Favors Control
0.01
0.1
1
10
100

Study	Deaths (steroid)	Deaths (placebo)	OR
Auckland	36	60	0.58
Block	1	5	0.16
Doran	4	11	0.25
Gamsu	14	20	0.70
Morrison	3	7	0.35
Papageorgiou	1	7	0.14
Tauesch	8	10	1.02
Summary			**0.53**



Finally:
Standard deviation "versus" standard error

- The term "standard deviation" refers to the variability of observations in a population or sample. (Reference ranges).

- The "standard error" is also a measure of variability, but it applies to the distribution of multiple sample means taken from a population (the basis of the 95% CI).

HYPOTHESIS TESTING

Jose Andres Calvache MD MSc PhD
Department of Anesthesiology, Universidad del Cauca, Colombia
Department of Anesthesiology, Easmus University Medical Centre, Rotterdam, The Netherlands

Test or hypothesis contrasts

- A statistical hypothesis is a statement regarding some characteristic of a population.
- This hypothesis has not yet been tested.
  - The average age of patients in the emergency room is 30 years
  - The proportion of hospitalized male patients is 40%
  - Gender is associated with having type 2 diabetes

- To contrast a hypothesis is to compare the statements with the reality we observe –in a sample-.

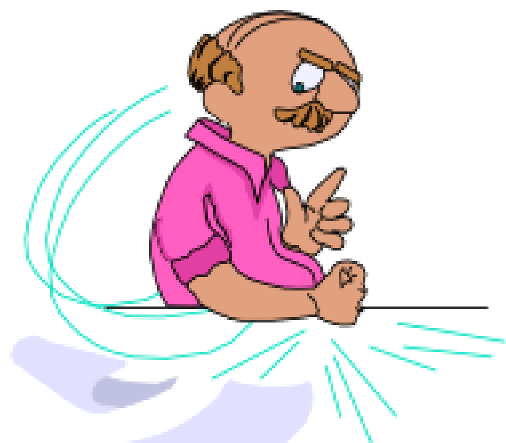
According to such a comparison,
we will “accept or reject” our hypotheses.

**Too much binary, too much black/white

I think the average age is 30 years !

But there are too many !

X_{age} **= 68 years old**



The difference is very large

I reject my hypothesis!!

Hypothesis

- It is a statement about parameters of a population
  - Means
  - Proportions
  - Standard deviation
  - Association, correlation, etc

If we want to contrast it, you must set up BEFORE analysis

Contrasts of hypotheses – Steps,

1. Identify and construct the hypothesis
2. Collect our data (sample)
3. Define a level of significance
4. Make a decision regarding contrast

1. Identify and construct the hypothesis

Null hypothesis H_0

- The one we contrast
- Non-association
- Our data can refute it
- It should not be rejected without good reason

- Sign =

Alternative hypothesis H_1

- Denies or opposes H0
- There is an association
- Data can show evidence on its favor
- It should not be accepted without great evidence on its favor

- Sign ≠, <, >

The null and alternative hypothesis ($H_0 – H_1$)

- Gender is associated with the incidence of type 2 diabetes!

Null hypothesis H_0	**Alternative hypothesis H_1**
$p_m = 50\%$	$p_m \neq 50\%$

The null and alternative hypothesis ($H_0 – H_1$)

- The average age of patients is 30 years

$$\mu = 30 \text{ years}$$

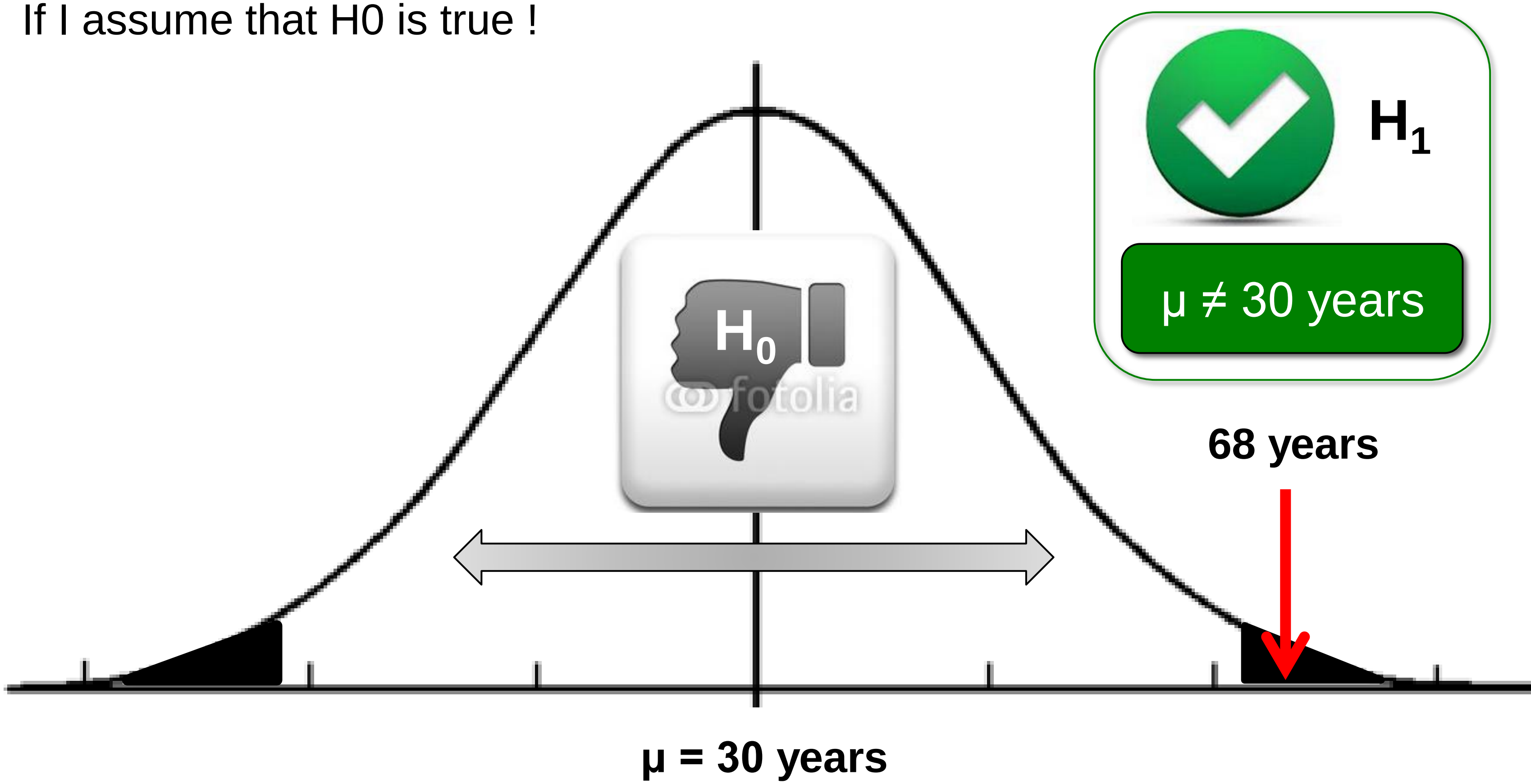
- The average age of patients **is not** 30 years

$$\mu \neq 30 \text{ years}$$

Reasoning of the hypothesis contrast,

$H_0 = \mu = 30$ years

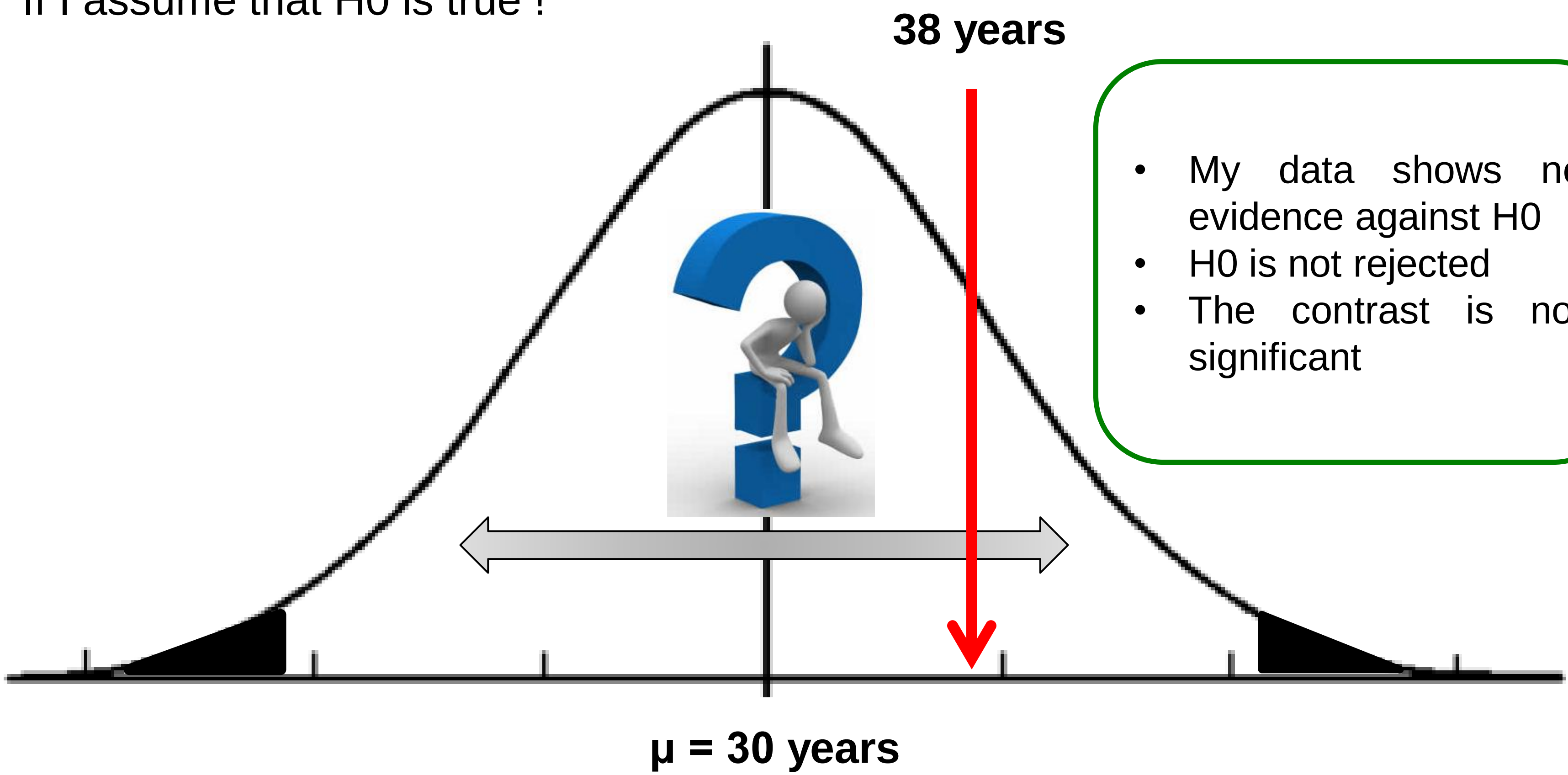
If I assume that H0 is true !



Reasoning of the hypothesis contrast,

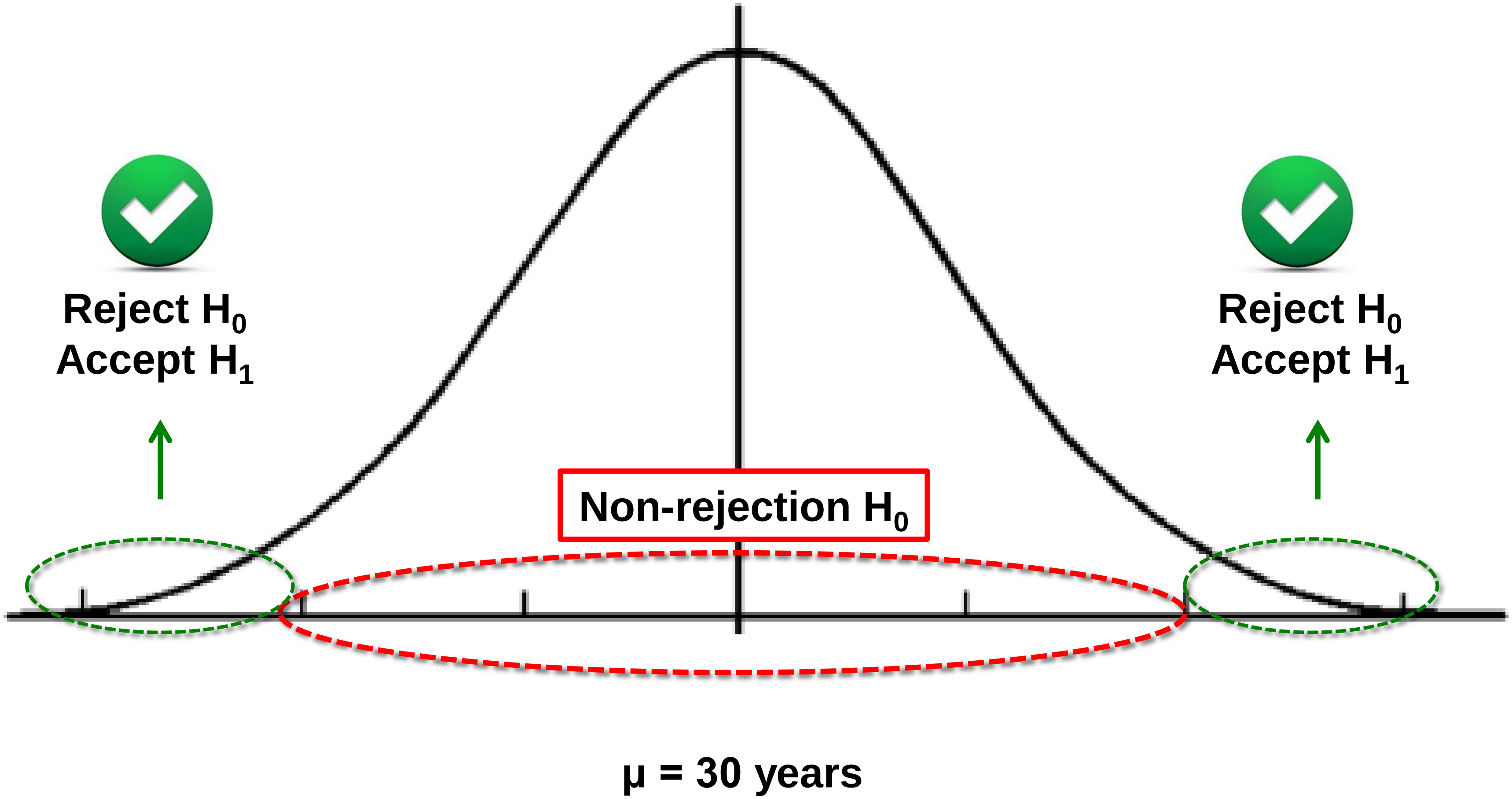
$H_0 = \mu = 30$ years

If I assume that H0 is true !

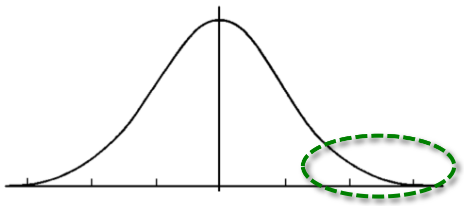


- My data shows no evidence against H0
- H0 is not rejected
- The contrast is not significant

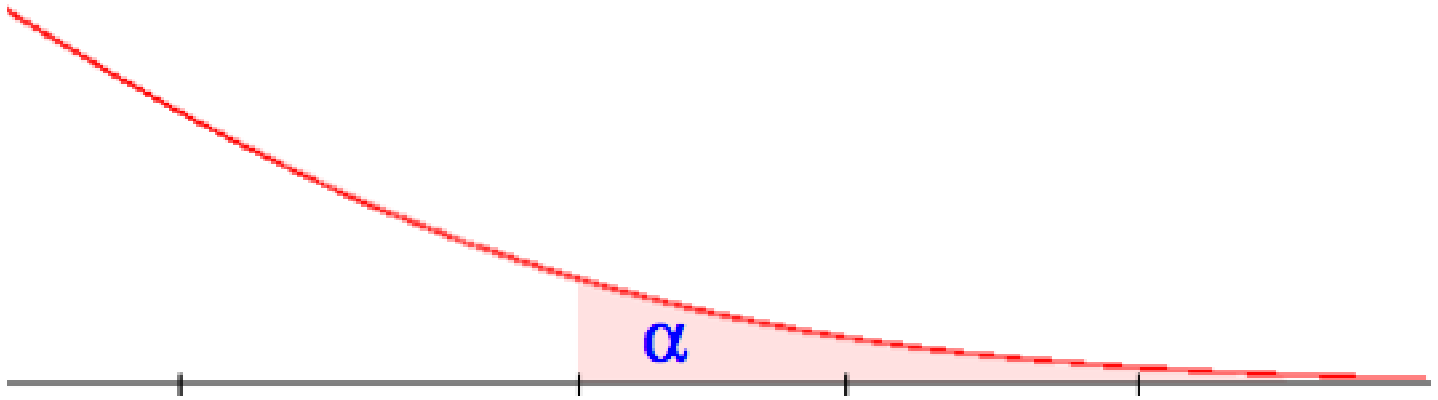
Level of significance



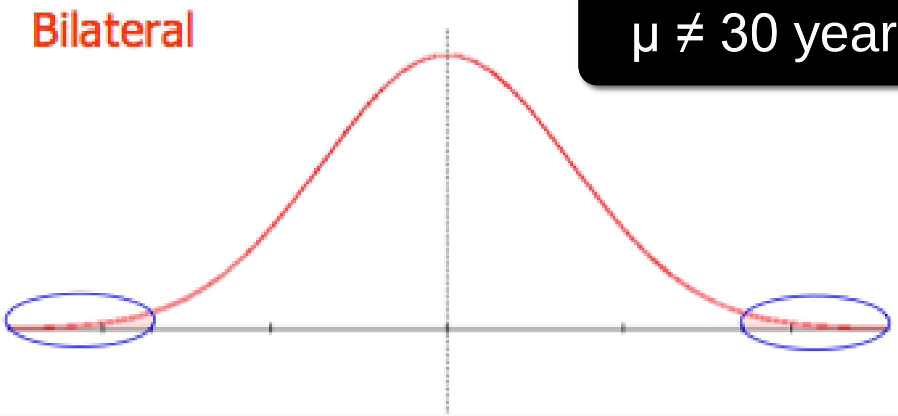
Nivel de significación



- Identified with the Greek letter α
- Corresponds to a small value (1%, 5%)
- Fixed by the researcher before the contrast
- It is the probability of rejecting H0, when it is true



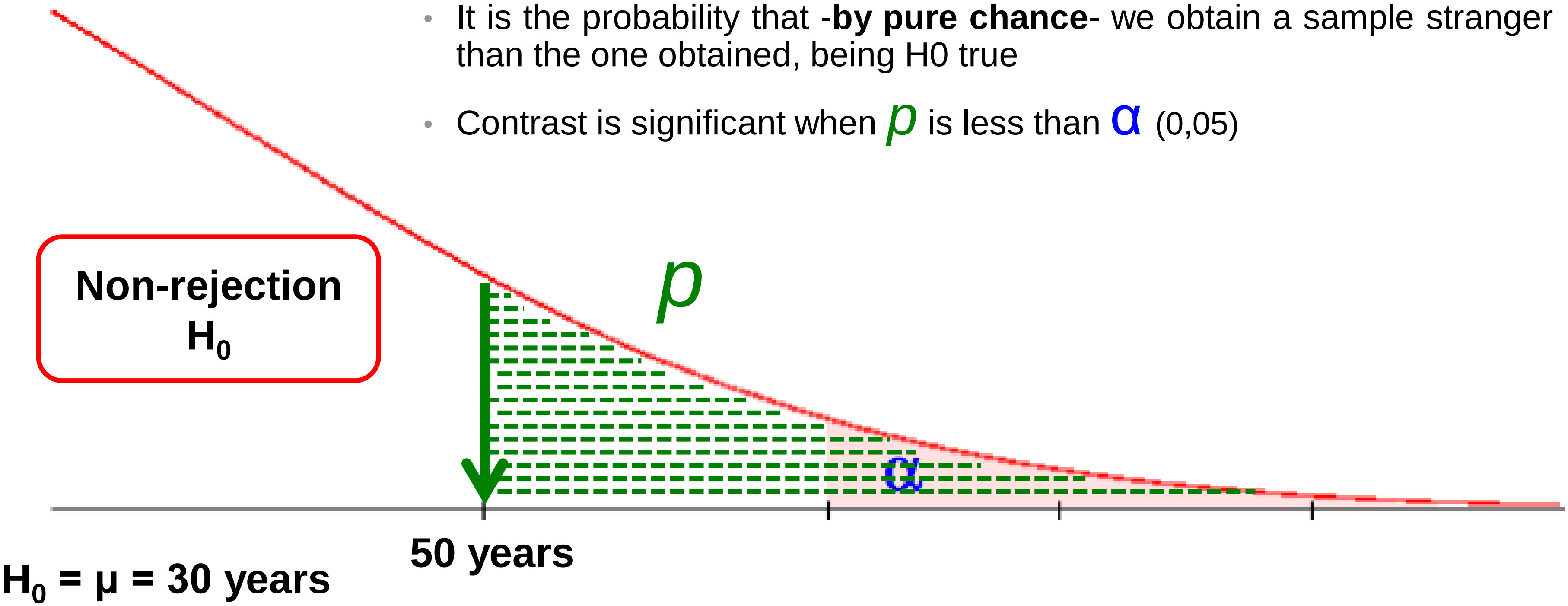
Unilateral or bilateral contrast (H_1)



$\mu \neq 30$ years

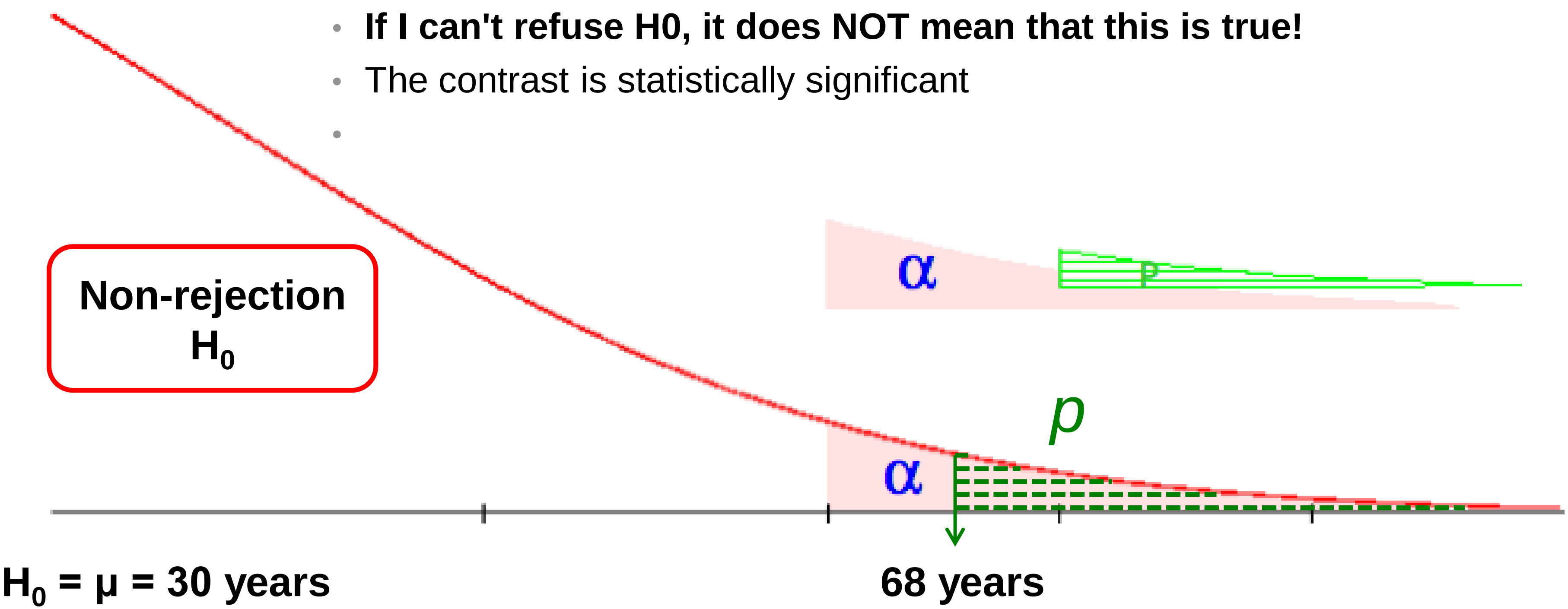
Signification (*p* value)

- It is the probability that an area that started right at the value of the statistic obtained from the sample would have.
- It is the probability of having a sample that disagrees even more than ours of H0
- It is the probability that **-by pure chance-** we obtain a sample stranger than the one obtained, being H0 true
- Contrast is significant when p is less than α (0,05)



Signification (*p* value)

- The area of *p* is lower than that of α (p<0,05)
- The result obtained in my study disagrees more than tolerated a priori
- I have enough evidence in my sample to reject H0 and accept H1
- **If I can't refuse H0, it does NOT mean that this is true!**
- The contrast is statistically significant



In short, α and *p*

- Level of significance α
  - It is known before starting the study. It's a small value.

- Signification *p*
  - It is obtained after performing the study. If it is less than α the contrast is significant.

Statistical analysis was performed using the MedCalc software package (Mariakerke, Belgium). For all experiments, α-level was set at 0.05.

Descriptive statistics, including mean values, standard deviations, and distributions, were used to present the data. The data were analyzed using Statistical Package for the Social Sciences (SPSS) version 16.0 (SPSS, Inc., Chicago, IL, USA). For categorical data, group comparisons were analyzed using chi-square tests for normal distribution. Statistical significance was defined at a *P*-value of ≤ 0.05.

and barriers to resident research education. Data are presented as counts and percentages of responses. All reported *P* values are 2-sided. $P < 0.01$ was required to reject the null hypothesis and to account for multiple comparisons. Confidence intervals (CIs) for differences in pro-

Risks in the contrast of hypotheses

- Presumption of innocence
- Evidence will help me reject it
- If rejection wrongly has serious consequences

- It should not be accepted if the evidence is not overwhelming.
- If rejection wrongly has "less" serious consequences

- **H_0**
  - He is innocent.
  - Until proven otherwise

- **H_1**
  - He is guilty
  - Must be penalized

Judicial legal process

		In reality	
		Innocent	**Guilty**
Judge's verdict	**Innocent**	OK	Error Less serious
	Guilty	Error **“Very serious”**	OK

Other examples,

- Imagine a clinical trial testing a new antibiotic treatment for the management of severe central nervous system infections – very expensive
- H0: Treatment in clinical trial has no effect on reality
- H1: The treatment has a real effect

	Realidad	
	H_0 true	H_0 false
Result of the study **No Rejection H_0**	Correct The treatment has no effect and so it is decided.	
Reject H_0 **Accept H_1**		Correct The treatment has an effect and the study confirms it.

Types I and II errors

- **Type I : (α)** P (reject H_0 | H_0 is true)

  Level of significance = 0,05 (5%)

- **Type II : (β)** P (accept H_0 | H_0 is false) (less bad)

  1 – β : P (reject H_0 | H_0 is false) – **STATISTICAL POWER**

Balance between α and β

- For a fixed sample size, the two types of error cannot be reduced at the same time
- To reduce β (or increase statistical power) you have to increase the sample size

Final summary,

- Hypothesis contrast
  - Strategy for creating a decision (about a parameter)
  - Evaluates the probability of finding samples that differ even more than the one obtained
  - Provide a p value to be able to discriminate
  - **Too much binary and recently very criticized**

- Confidence intervals
  - Quantifies the effect of interest (parameter) and the uncertainty of the estimate
  - Provides a range in which the parameter could be located
  - Allows to evaluate the clinical interest of the estimate
  - It should always be used -when possible-

JAMA Guide to Statistics and Methods JAMA November 23/30, 2021 Volume 326, Number 20

Use of Confidence Intervals in Interpreting Nonstatistically Significant Results

Alexander T. Hawkins, MD, MPH; Lauren R. Samuels, PhD

- The goal of much of medical research is to determine which of 2 or more therapeutic approaches is most effective in a given situation
- The power of a study is the probability of detecting a true treatment effect of a given magnitude and is highly dependent on the number of patients studied
- When a “retrospective” observational study design is used, researchers have little or no control over the sample size, and thus little control over the power to detect a particular treatment effect