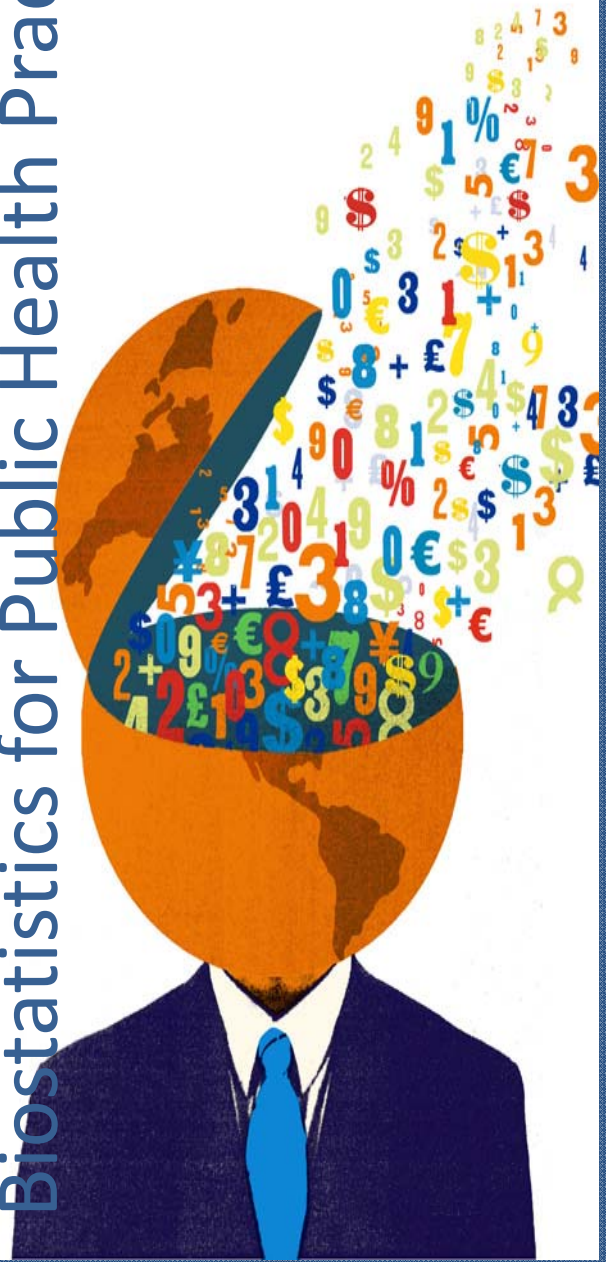
Biostatistics for Public Health Practice

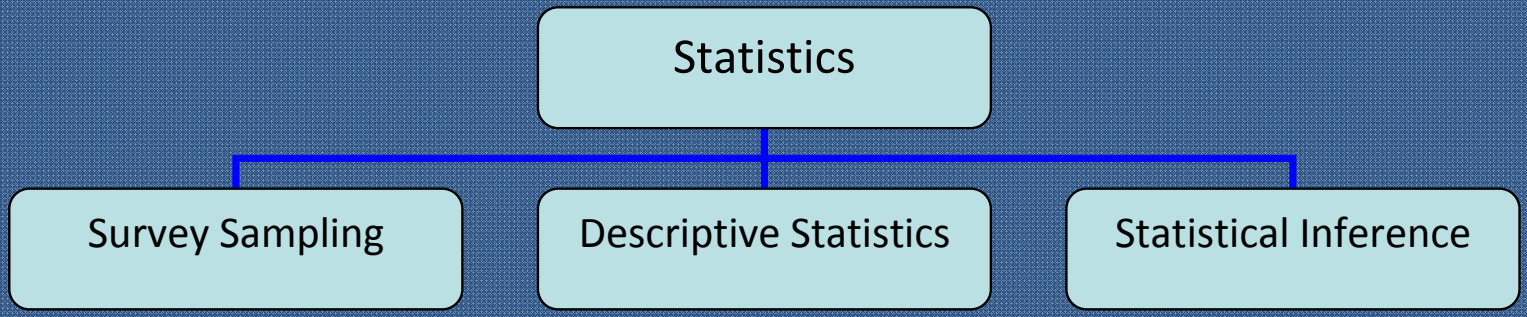


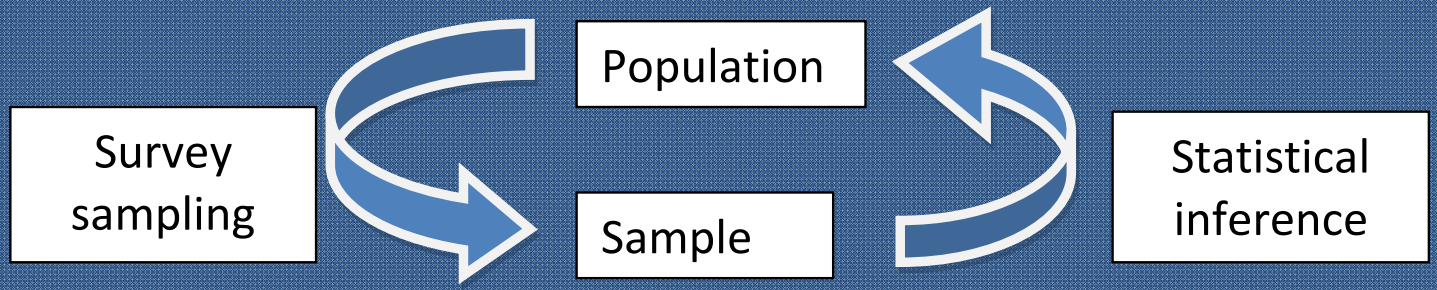
Week 03-3

Concepts of Statistical Inference

Associate Professor Theo Niyonsenga

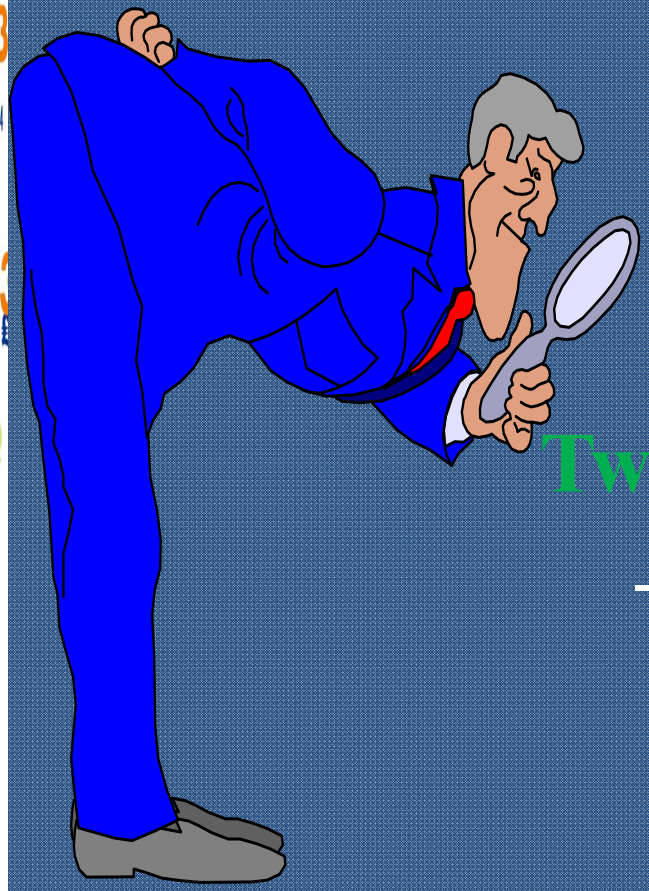
Statistical Inference





Note: Descriptive statistics applies to both sample and population.

Statistical Inference



Use a random sample to learn something about a larger population

Two ways to make inference

- Estimation of parameters
  - * Point Estimation ($\bar{X}$ or p)
  - * Intervals Estimation
- Hypothesis Testing

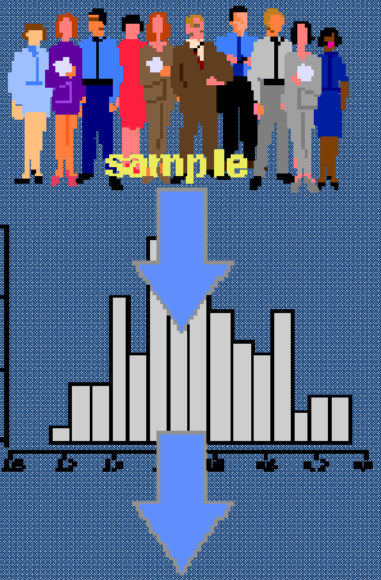
Statistical Inference

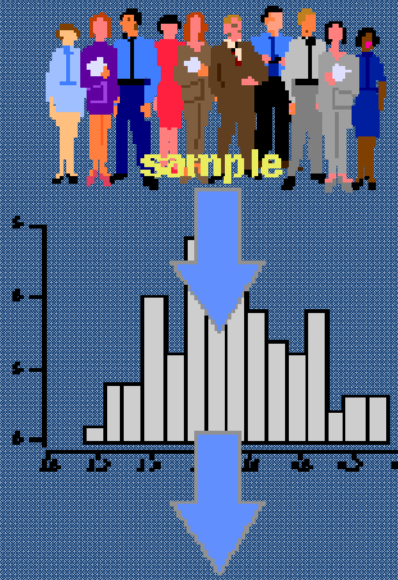
	Statistic		Parameter
Mean:	$\bar{x}$	estimates	μ
Standard deviation:	s	estimates	σ
Proportion:	p	estimates	π

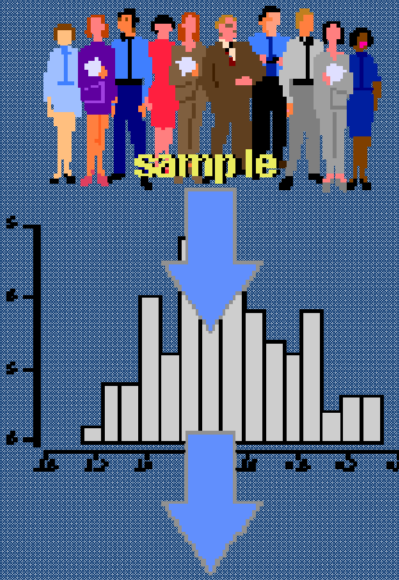
from sample

from entire population

Sampling Distribution







Average

$\bar{X}$ or P

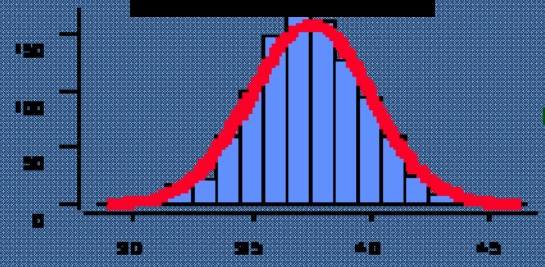
Average

$\bar{X}$ or P

Average

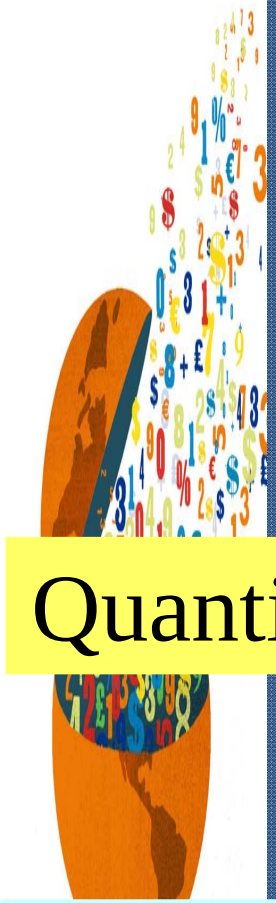
$\bar{X}$ or P

The Sampling Distribution...



...is the distribution of a statistic across an infinite number of samples

Sampling distribution



SE (Standard Error):

- Measure of dispersion of sampling distribution
- Standard deviation of sampling distribution

Quantitative Variable

$$SE(\text{Mean}) = s/\sqrt{n}$$

Qualitative Variable

$$SE(\text{p}) = \sqrt{p(1-p)/n}$$

Confidence Interval

Confidence Interval of a Parameter

= Statistic ± Its Error

What is in "Its Error"?

- SE(Statistic)
- A number associated to the Level of confidence
- Derived from z-score, or t-score, or chi-square, or Fisher F

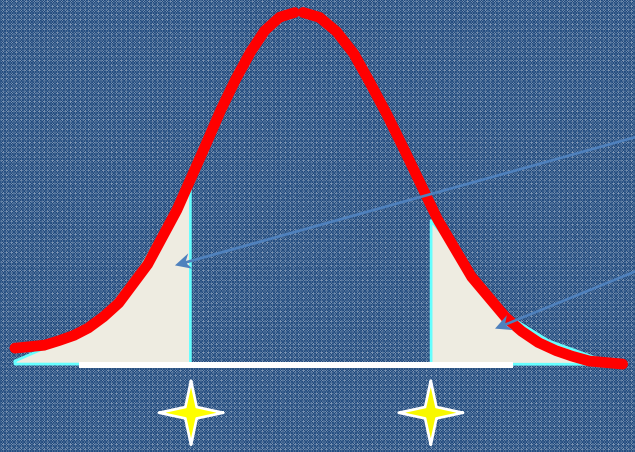
Confidence Interval

Level of significance, α & Level of confidence, $1-\alpha$

- Typical values are 1% (99%), 5% (95%)
- Selected by the Researcher at the Start
- Provides Critical Value(s) of the statistic
- Critical Values define regions of unlikely (and likely) values of the sample statistic within the sampling distribution

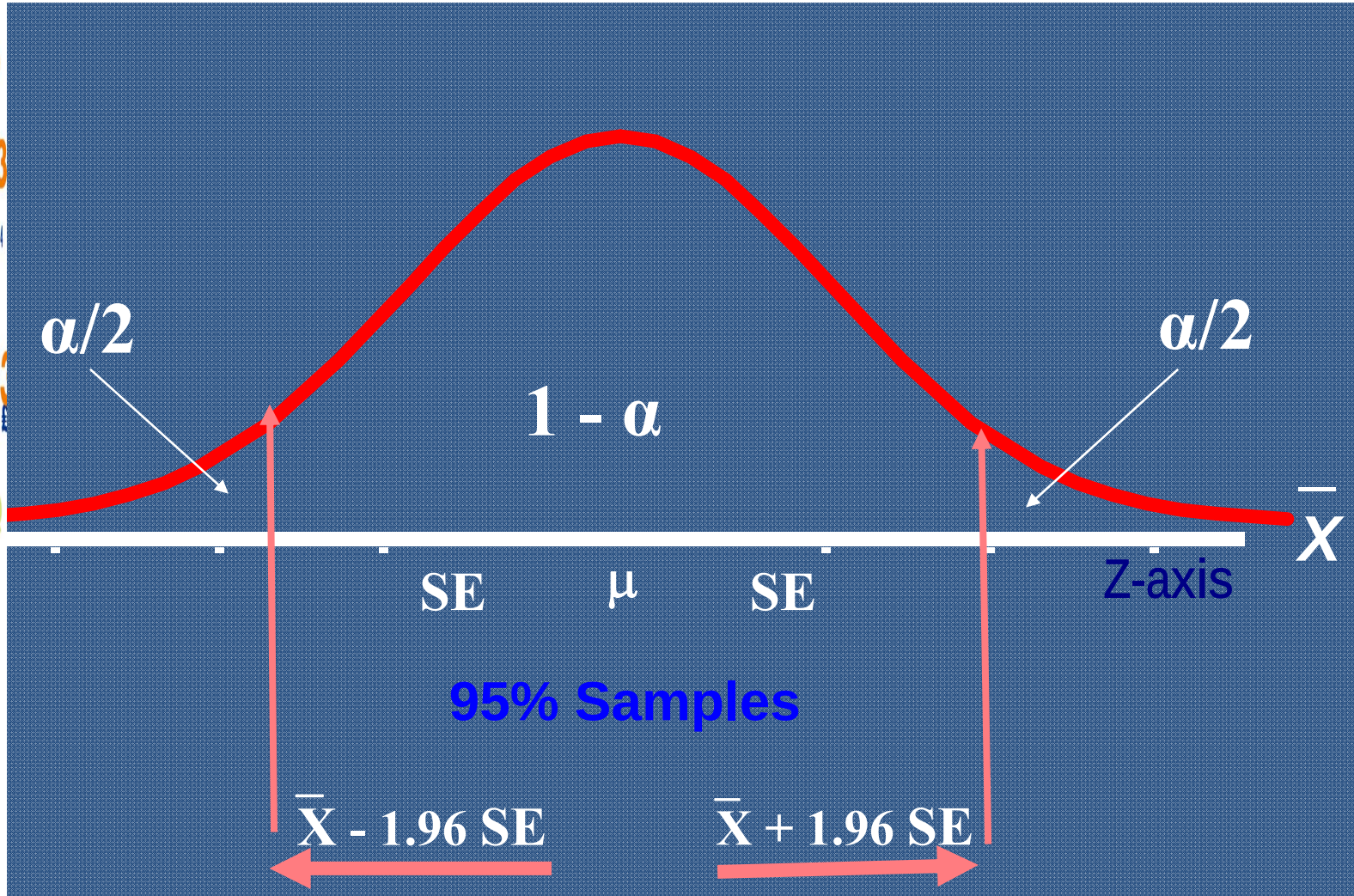
Confidence Interval

Level of Significance, *a* and Rejection Region

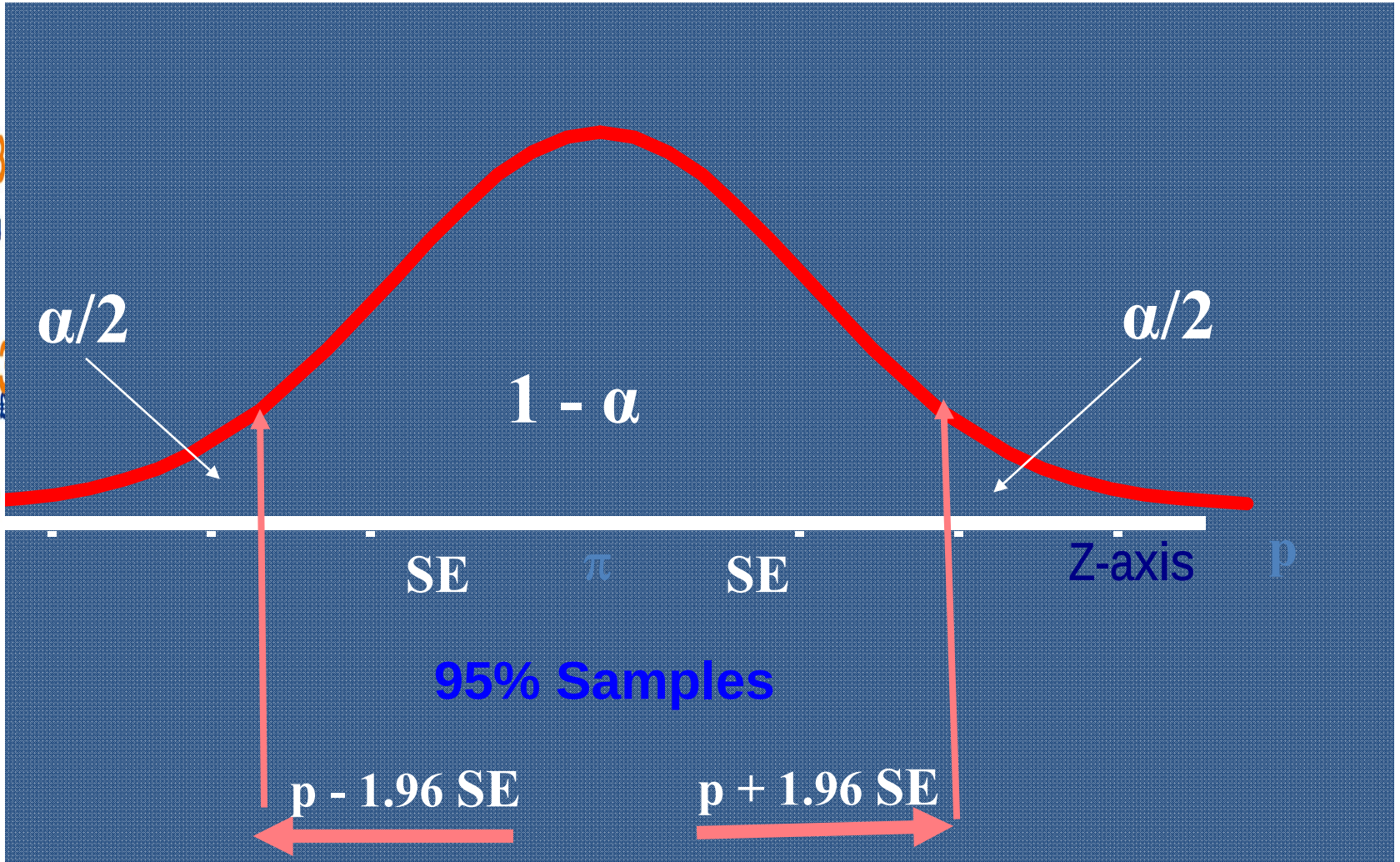


Level of Confidence, 1-*a* and Acceptance Region

Confidence Interval



Confidence Interval



Confidence Interval

Interpretation of CI

Probabilistic

In repeated sampling 100(1-α)% of <u>all intervals</u> around sample means will in the long run include μ

Practical

We are 100(1-α)% confident that the <u>single</u> computed CI contains μ

Confidence Interval

In a survey of 140 asthmatics, 35% had allergy to house dust. Construct the 95% CI for the population proportion.

95% CI of $\pi = p \pm Z * SE(p)$;

SE= $\sqrt{0.35*0.65/140} = 0.04$

$0.35 - 1.96 \times 0.04 \leq \pi \leq 0.35 + 1.96 \times 0.04$

$0.27 \leq \pi \leq 0.4$, or $27\% \leq \pi \leq 43\%$

Confidence Interval

An epidemiologist studied the blood glucose level of a random sample of 100 patients. The mean was 170, with a SD of 10.

95% CI of μ = mean $\pm$ Z * SE(mean);

SE=10/10=1 ($SE(\text{Mean}) = 10/\sqrt{100}$)

$170 - 1.96 \times 1 \leq \mu \leq 170 + 1.96 \times 1$

$168.04 \leq \mu \leq 171.96$

Hypothesis Testing

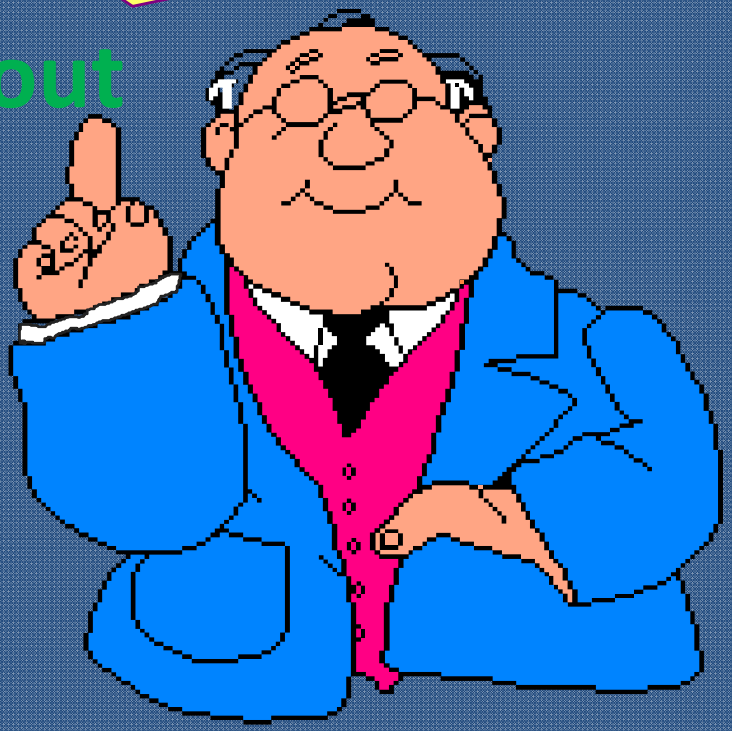
- A statistical procedure that uses a sample data to evaluate a hypothesis about a population parameter
- Intended to help researchers differentiate between real and random patterns in the observed data

Hypothesis Testing

What is a Hypothesis?

I assume the mean SBP of participants is 120 mmHg

An assumption about the population parameter.



Hypothesis Testing

The steps in Hypothesis Testing:

- A claim is made (researcher's hypothesis)
- Evidence (sample data) is collected in order to test the claim
- The data are analyzed in order to support or refute the claim

Null and Alternative Hypotheses:

- Null Hypothesis: Opposite of the researcher's claim about the population parameter; the simplest state of the parameter
- Alternative Hypothesis: Researcher's claim;

Hypothesis Testing

Result Possibilities

H_0: Innocent (*Negative*); H_1: Guilty (*Positive*)

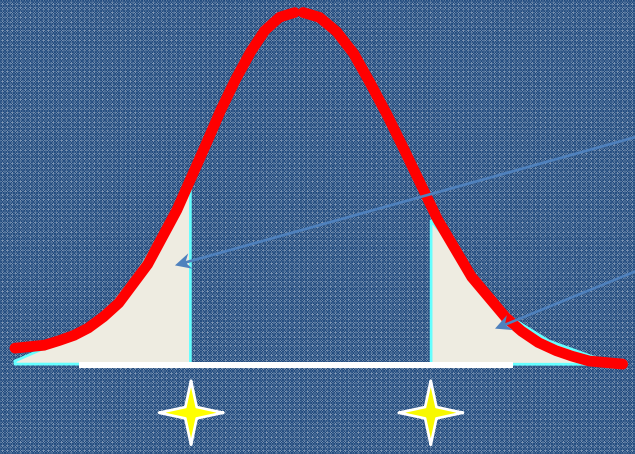
Jury Trial			Hypothesis test		
	Actual Situation			Actual Situation	
Verdict	Innocent	Guilty	Decision	H_0 True	H_0 False
Innocent	Correct	Error	Accept H_0	1 - α	Type II Error (β)
Guilty	Error	Correct	Reject H_0	Type I Error (α)	Power (1 - β)

False Positive

False Negative

Hypothesis Testing

Level of Significance, *a* and Rejection Region



Level of Confidence, 1-*a* and Acceptance Region

Hypothesis Testing

The p value of a test:

- A probability of obtaining a test statistic value as extreme or more than the actual sample value given that the null hypothesis (H_0) is true
- Observed level of significance
- Used to make the decision about H_0 :
  - **If *p* value $\geq \alpha$, Do Not Reject H_0**
  - **If *p* value $< \alpha$, Reject H_0**