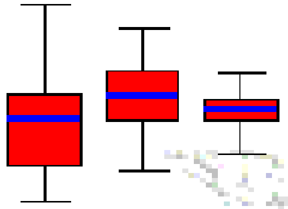
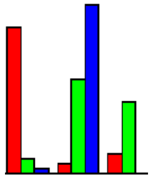
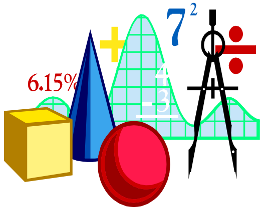




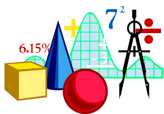
Basics of Inference



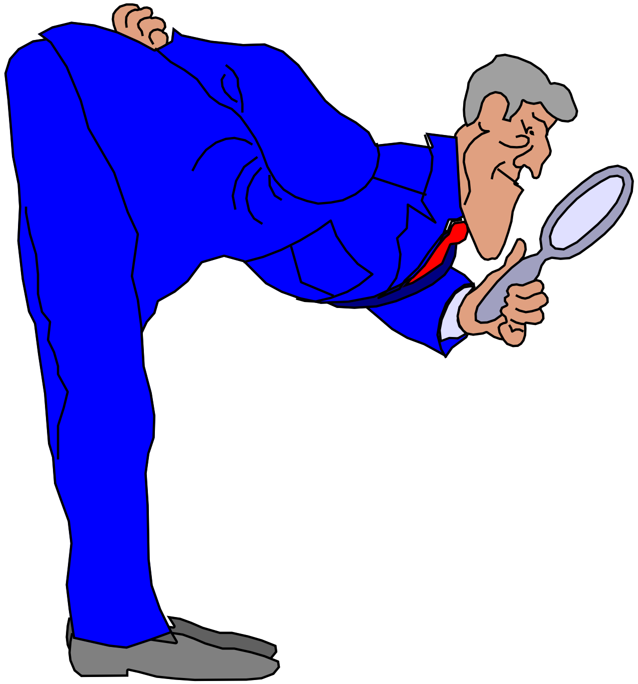
Dr. Vinay Kumar
Department of Statistics
Ram Lal Anand College

Lesson Objectives

- Know what is Inference
- Know what is parameter estimation
- Understand hypothesis testing & the "types of errors" in decision making.
- Know what the α-level means.
- Learn how to use test statistics to examine hypothesis about population mean, proportion



Inference



Use a random sample to learn something about a larger population

Inference

- **Two ways to make inference**
  - Estimation of parameters

    * Point Estimation ($\bar{X}$ or p)

    * Intervals Estimation
  - Hypothesis Testing

Statistic **Parameter**

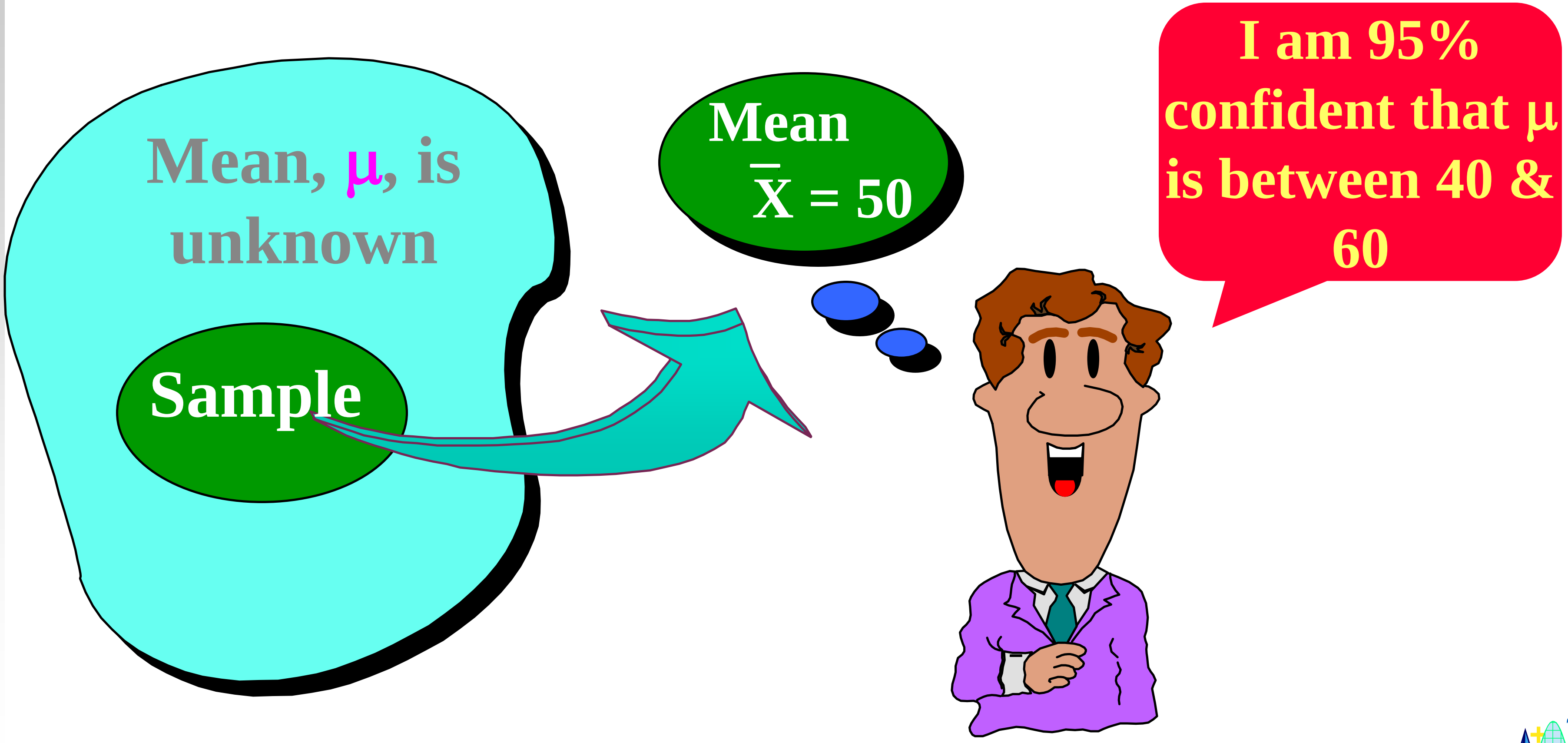
	Statistic		Parameter
Mean:	$\bar{X}$	estimates	_ μ ___
Standard deviation:	s	estimates	_ σ ___
Proportion:	p	estimates	_ π ___

from sample

from entire population

Estimation of parameters

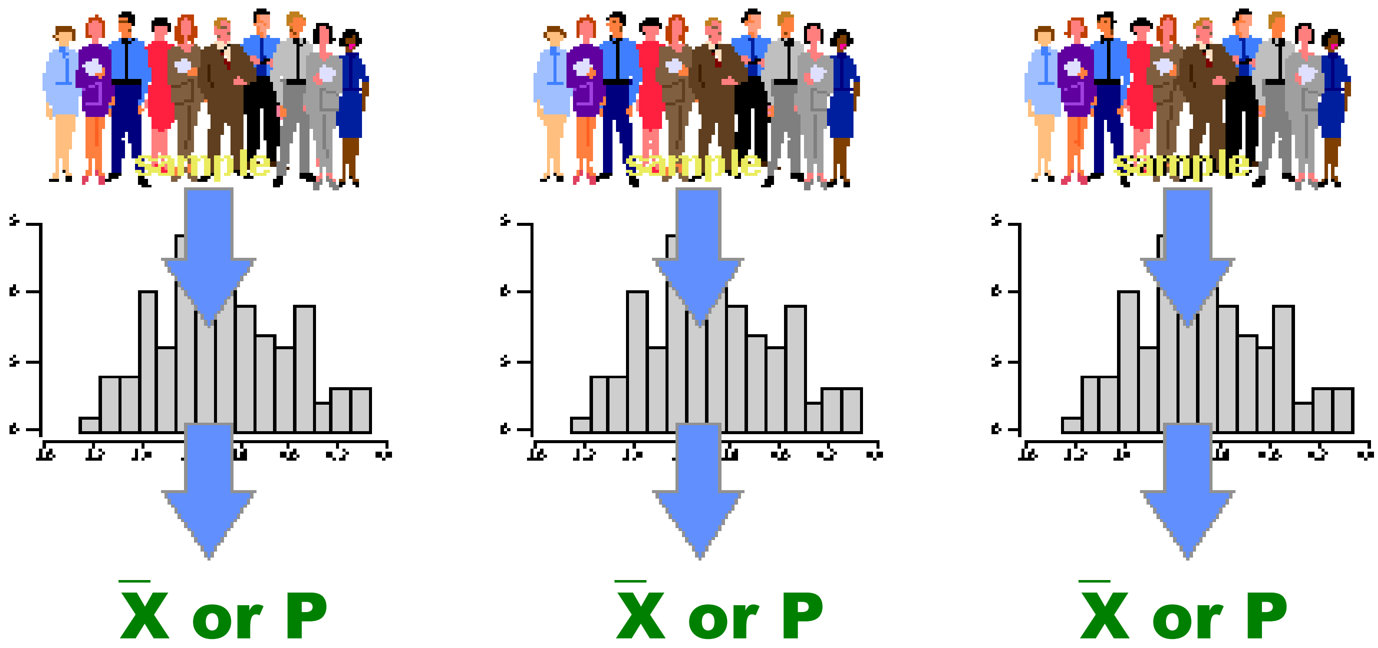
Population **Point estimate** **Interval estimate**



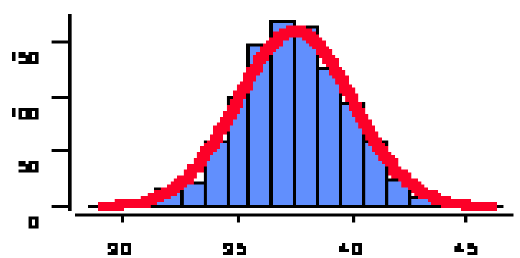
Parameter

= Statistic ± Its Error

Sampling Distribution



The Sampling Distribution…



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

Standard Error

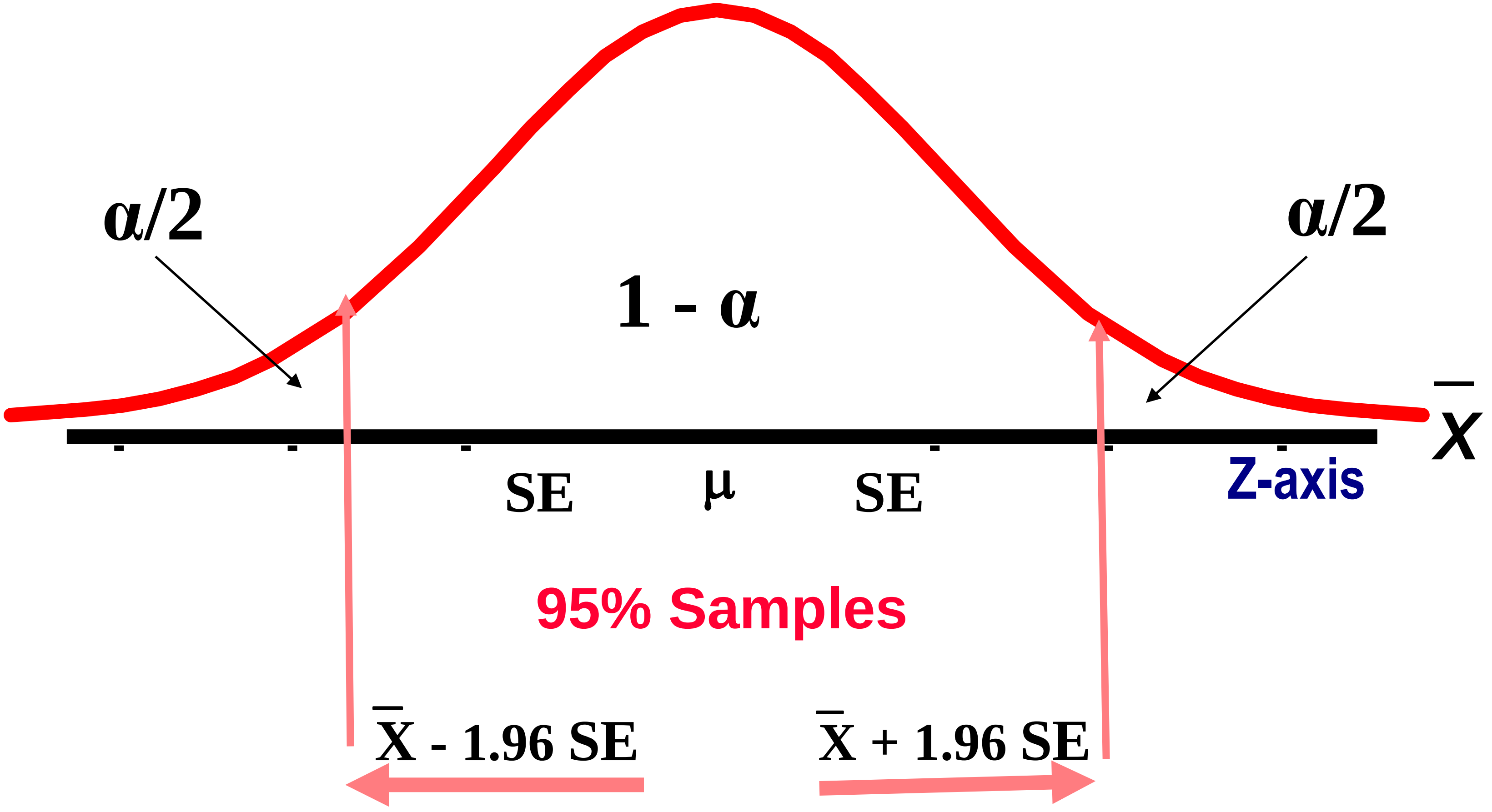
Quantitative Variable

$$SE\ (Mean) = \frac{S}{\sqrt{n}}$$

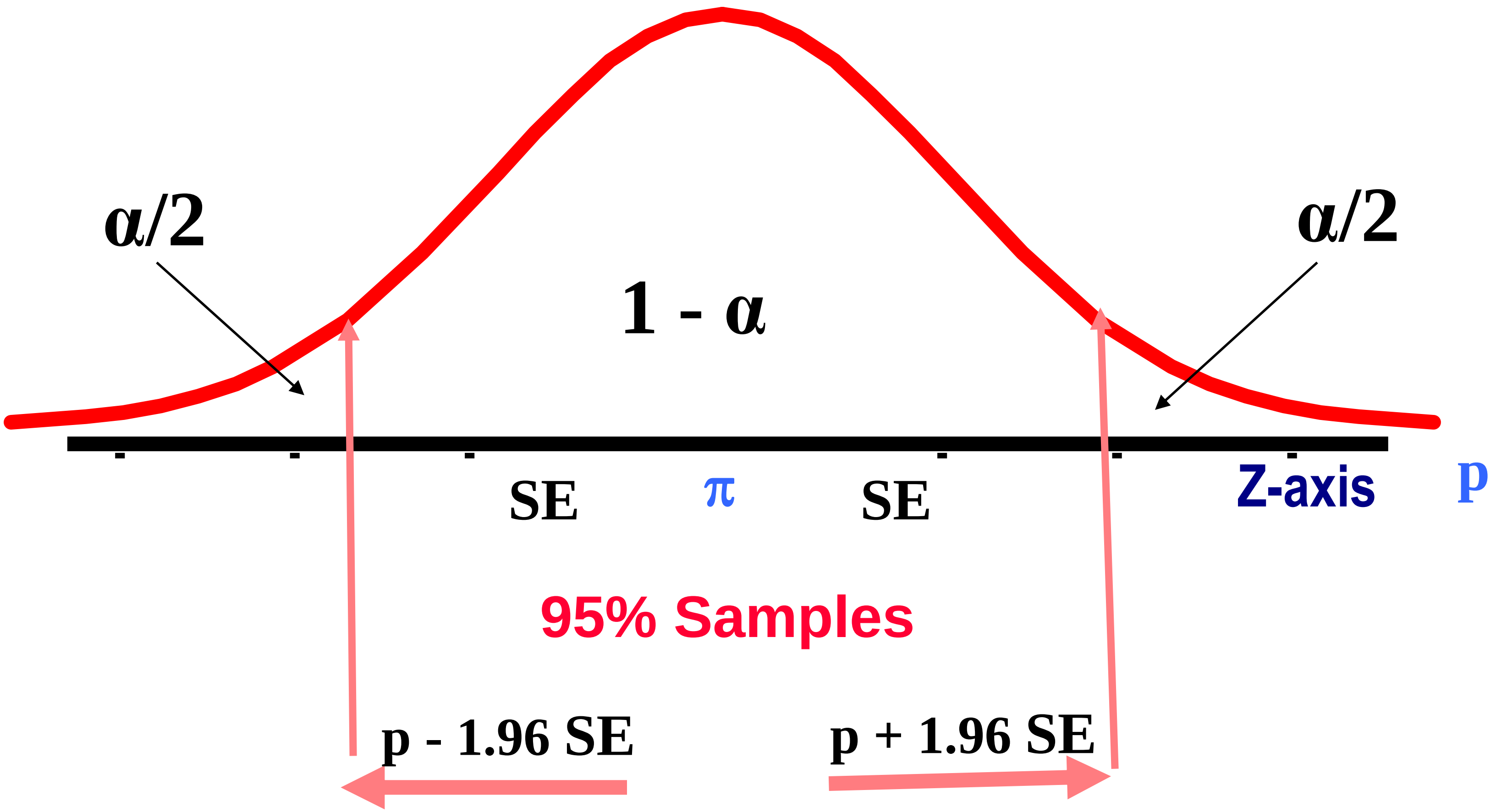
Qualitative Variable

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

Confidence Interval



Confidence Interval



Interpretation of CI

Probabilistic

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

Practical

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

Example (Sample size$\geq$30)

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

$$\mu = \bar{X} \pm Z \times SE$$

SE = 10/10 = 1

95 %

Then CI:

$\mu = 170 \pm 1.96 \times 1$ $168.04 \leq \mu \geq 171.96$

Example (Proportion)

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

$$\pi = p \pm Z\sqrt{\frac{P(1-p)}{n}} \qquad SE = \sqrt{\frac{0.35(1-0.35)}{140}} = 0.04$$

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

$$0.27 \leq \pi \geq 0.43$$

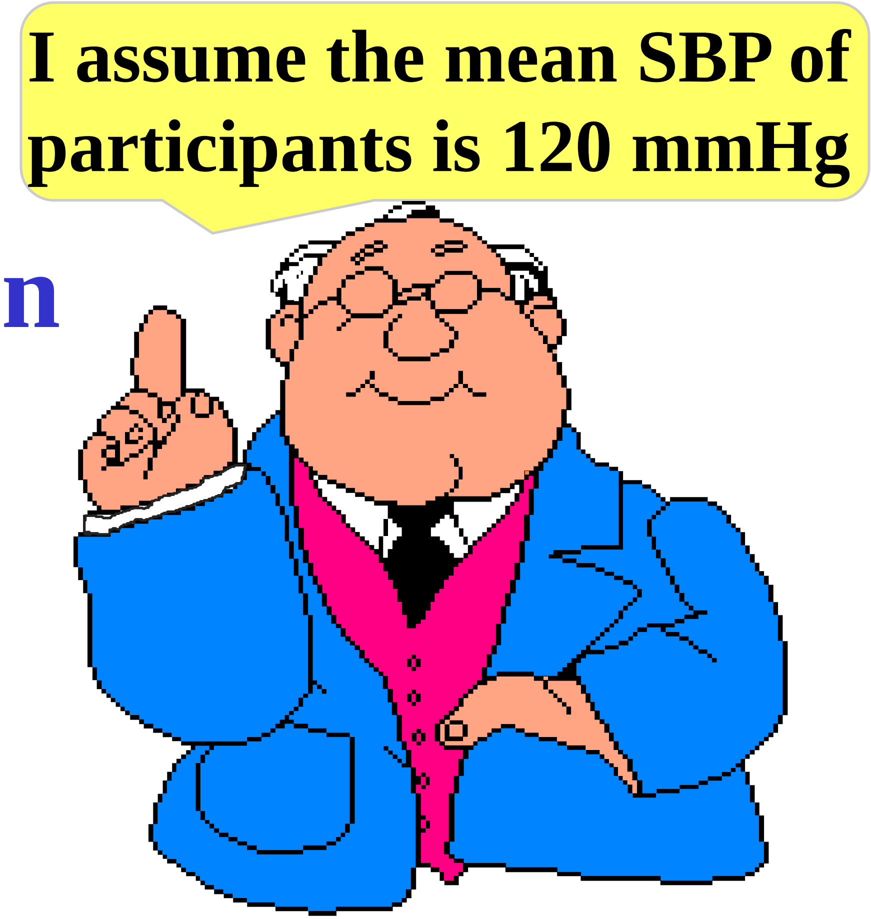
$$27\% \leq \pi \geq 43\%$$

Hypothesis testing

A statistical method that uses sample data to evaluate a hypothesis about a population parameter. It is intended to help researchers differentiate between real and random patterns in the data.

What is a Hypothesis?

An assumption about the population parameter.



Null & Alternative Hypotheses

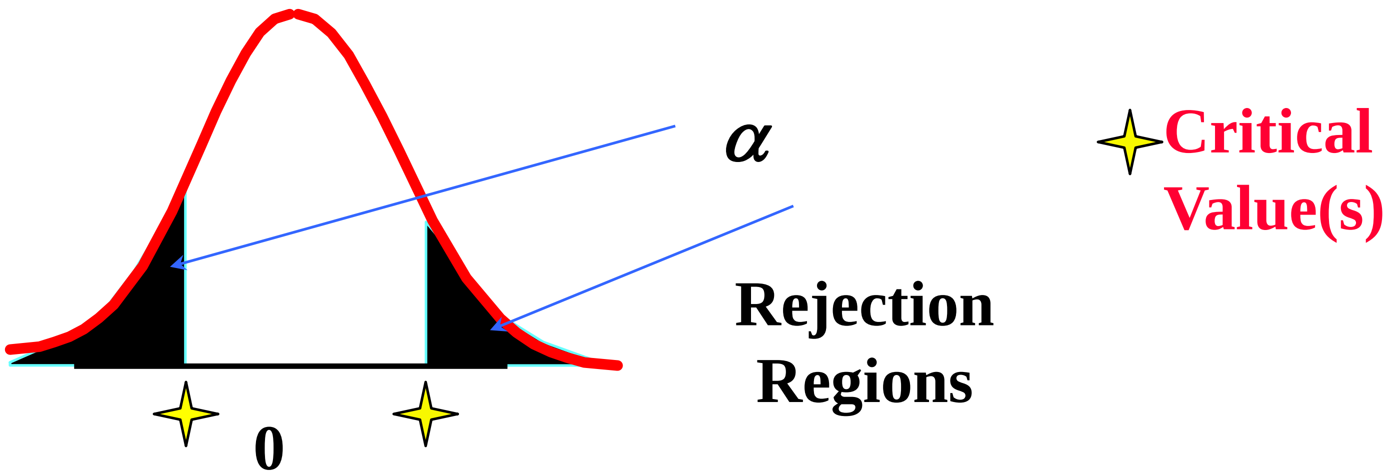
- ***H_0*** **Null Hypothesis states the Assumption to be tested** **e.g. SBP of participants = 120** **(H_0: $\mu = 120$).**

- ***H_1*** **Alternative Hypothesis is the opposite of the null hypothesis (SBP of participants $\neq$ 120** **(H_1: $\mu \neq 120$).** **It may or may not be accepted and it is the hypothesis that is believed to be true by the researcher**

Level of Significance, α

- Defines unlikely values of sample statistic if null hypothesis is true. Called rejection region of sampling distribution
- Typical values are 0.01, 0.05
- Selected by the Researcher at the Start
- Provides the Critical Value(s) of the Test

Level of Significance, α and the Rejection Region



Result Possibilities

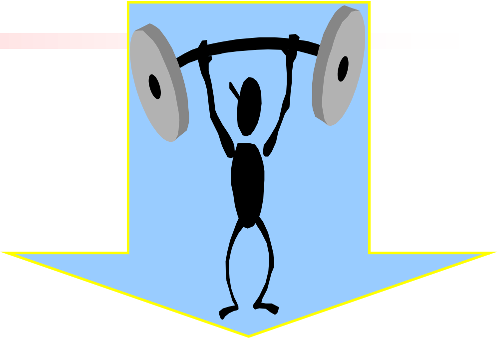
H_0: Innocent

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

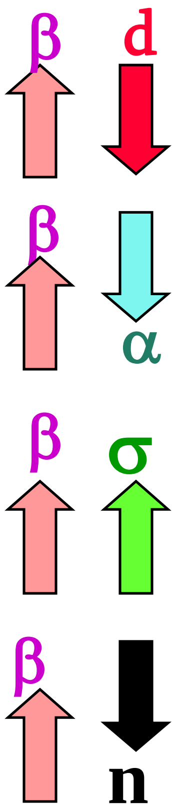
False Positive

False Negative

Factors Increasing Type II Error

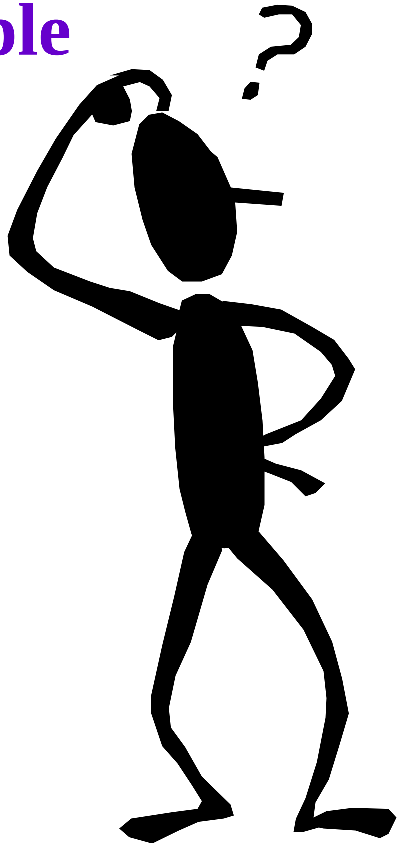


- True Value of Population Parameter
  - Increases When Difference Between Hypothesized Parameter & True Value Decreases
- Significance Level α
  - Increases When α Decreases
- Population Standard Deviation σ
  - Increases When σ Increases
- Sample Size *n*
  - Increases When *n* Decreases



p Value Test

- Probability of Obtaining a Test Statistic More Extreme ($\leq$ or $\geq$) than Actual Sample Value Given H_0 Is True
- Called Observed Level of Significance
- Used to Make Rejection Decision
  - If p value $\geq \alpha$, Do Not Reject H_0
  - If p value $< \alpha$, Reject H_0



Hypothesis Testing: Steps

Test the Assumption that the true mean SBP of participants is 120 mmHg.

State H_0	$H_0: \mu = 120$
State H_1	$H_1: \mu \neq 120$
Choose α	$\alpha = 0.05$
Choose n	n = 100
Choose Test:	*Z, t, X^2 Test (or p Value)*

Hypothesis Testing: Steps

Compute Test Statistic ***(or compute P value)***

Search for Critical Value

Make Statistical Decision rule

Express Decision

One sample-mean Test

- **Assumptions**
  - **Population is normally distributed**

- **t test statistic**

$$t = \frac{\text{sample mean} - \text{null value}}{\text{standard error}} = \frac{\bar{x} - \mu_0}{s \big/ \sqrt{n}}$$

Example Normal Body Temperature

What is **normal body temperature**? Is it actually 37.6^{o}C (on average)?

State the null and alternative hypotheses

$$H_0: \mu = 37.6^{o}\text{C}$$

$$H_a: \mu \neq 37.6^{o}\text{C}$$

Example Normal Body Temp (cont)

Data: random sample of n = 18 normal body temps

37.2	**36.8**	**38.0**	**37.6**	**37.2**	**36.8**	**37.4**	**38.7**	**37.2**
36.4	**36.6**	**37.4**	**37.0**	**38.2**	**37.6**	**36.1**	**36.2**	**37.5**

Summarize data with a test statistic

Variable	n	Mean	SD	SE	t	P
Temperature	18	37.22	0.68	0.161	2.38	0.029

$$t = \frac{\text{sample mean} - \text{null value}}{\text{standard error}} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

STUDENT’S t DISTRIBUTION TABLE

Degrees of freedom	Probability (p value)		
	0.10	0.05	0.01
1	6.314	12.706	63.657
5	2.015	2.571	4.032
10	1.813	2.228	3.169
17	1.740	2.110	2.898
20	1.725	2.086	2.845
24	1.711	2.064	2.797
25	1.708	2.060	2.787
∞	1.645	1.960	2.576

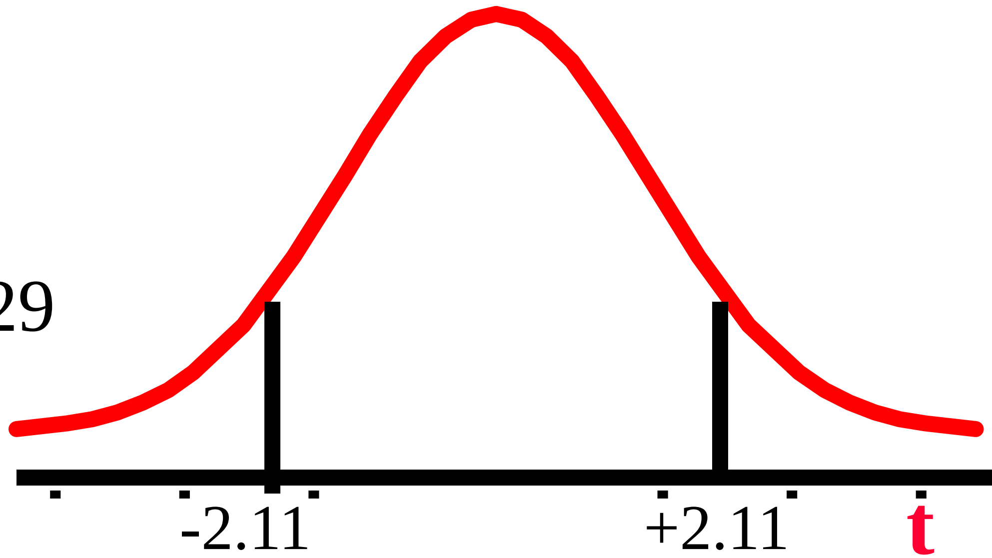
Example Normal Body Temp (cont)

Find the *p*-value

Df = n – 1 = 18 – 1 = 17

From SPSS: *p*-value = 0.029

From t Table: *p*-value is between 0.05 and 0.01.



Area to left of t = -2.11 equals area to right of t = +2.11.

The value t = 2.38 is between column headings 2.110& 2.898 in table, and for df =17, the *p*-values are 0.05 and 0.01.

Example Normal Body Temp (cont)

Decide whether or not the result is statistically significant based on the *p*-value

Using $\alpha = 0.05$ as the level of significance criterion, the results are **statistically significant** because 0.029 is less than 0.05. In other words, we can reject the null hypothesis.

Report the Conclusion

We can conclude, based on these data, that the mean temperature in the human population does not equal 37.6.

One-sample test for proportion

- Involves categorical variables
- Fraction or % of population in a category
- Sample proportion (p)

$$p = \frac{X}{n} = \frac{number\ of\ successes}{sample\ size}$$

- Test is called Z test
  where:
- Z is computed value
- π is proportion in population
  (null hypothesis value)

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Critical Values: 1.96 at α=0.05

2.58 at α=0.01

Example

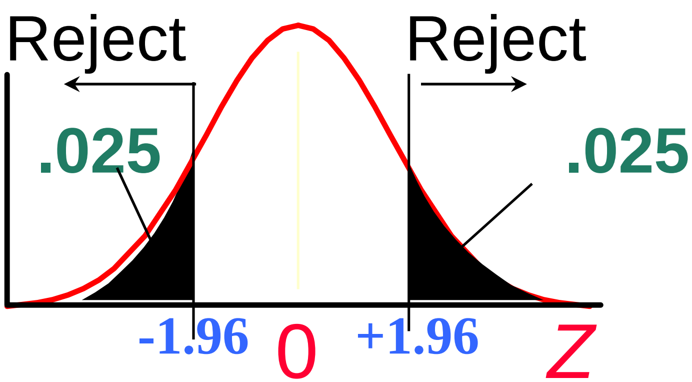
- **In a survey of diabetics in a large city, it was found that 100 out of 400 have diabetic foot. Can we conclude that 20 percent of diabetics in the sampled population have diabetic foot.**
- **Test at the α =0.05 significance level.**

Solution

$H_o: \pi = 0.20$

$H_1: \pi \neq 0.20$

$$Z = \frac{0.25 - 0.20}{\sqrt{\dfrac{0.20\,(1 - 0.20)}{400}}} = 2.50$$

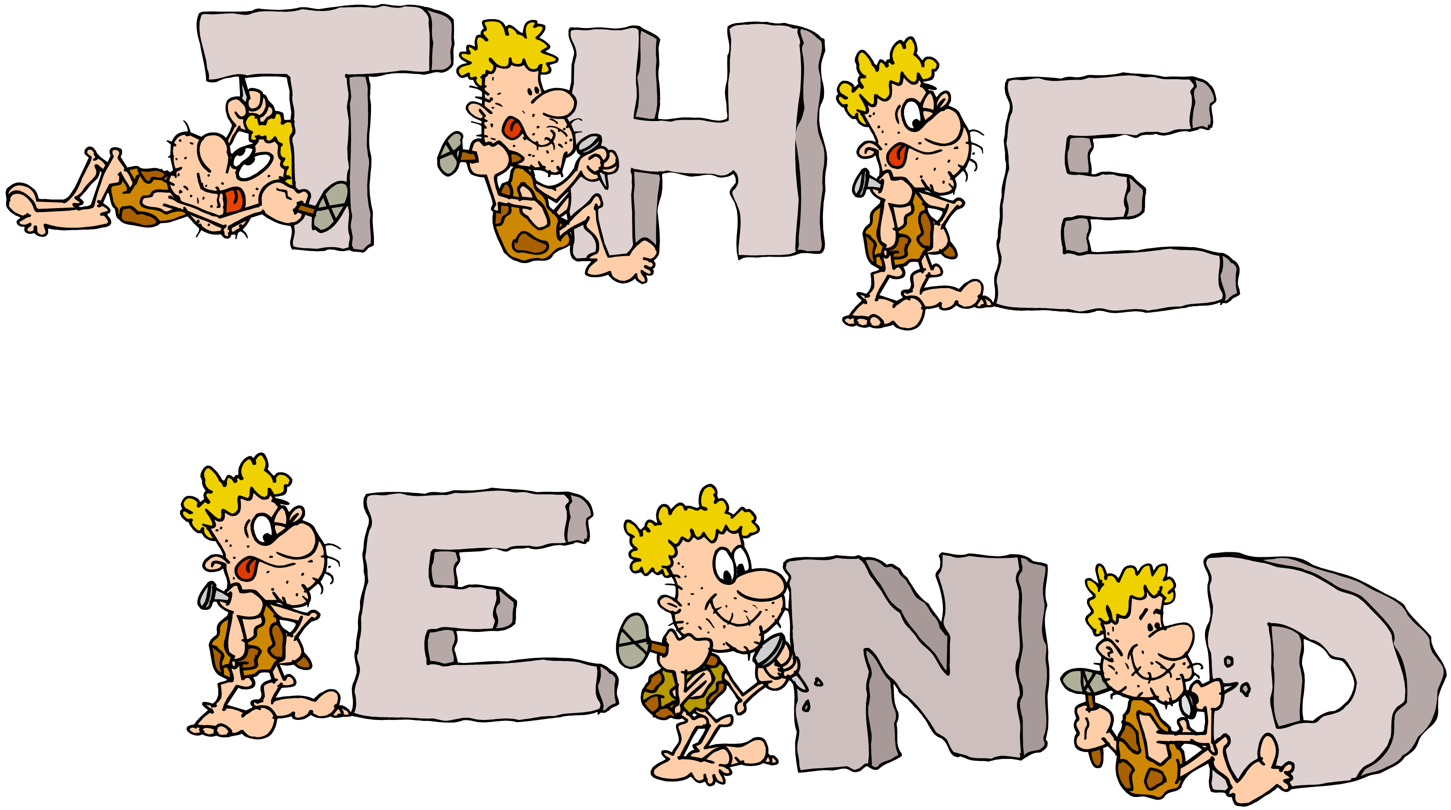


Critical Value: 1.96

Decision:

We have sufficient evidence to reject the Ho value of 20%

We conclude that in the population of diabetic the proportion who have diabetic foot does not equal 0.20


THE
END