Hypothesis Testing

Einas Al-Eisa, MSc, PhD
King Saud University

The US criminal court system

- Assume innocence until “proven” guilty
- Evidence is presented at a trial
- Proof has to be “beyond a reasonable doubt”
- A jury's possible decision:
  - ➢guilty
  - ➢not guilty
  - ➢“innocent”….?

Can juries make mistakes?

- **Type I error:** if a person is really innocent, but the jury decides guilty, then they've sent an innocent person to jail
- **Type II error:** if a person is really guilty, but the jury finds not guilty, a criminal is walking free on the streets
- Type I error is considered more important than a Type II error

Justice System - Trial

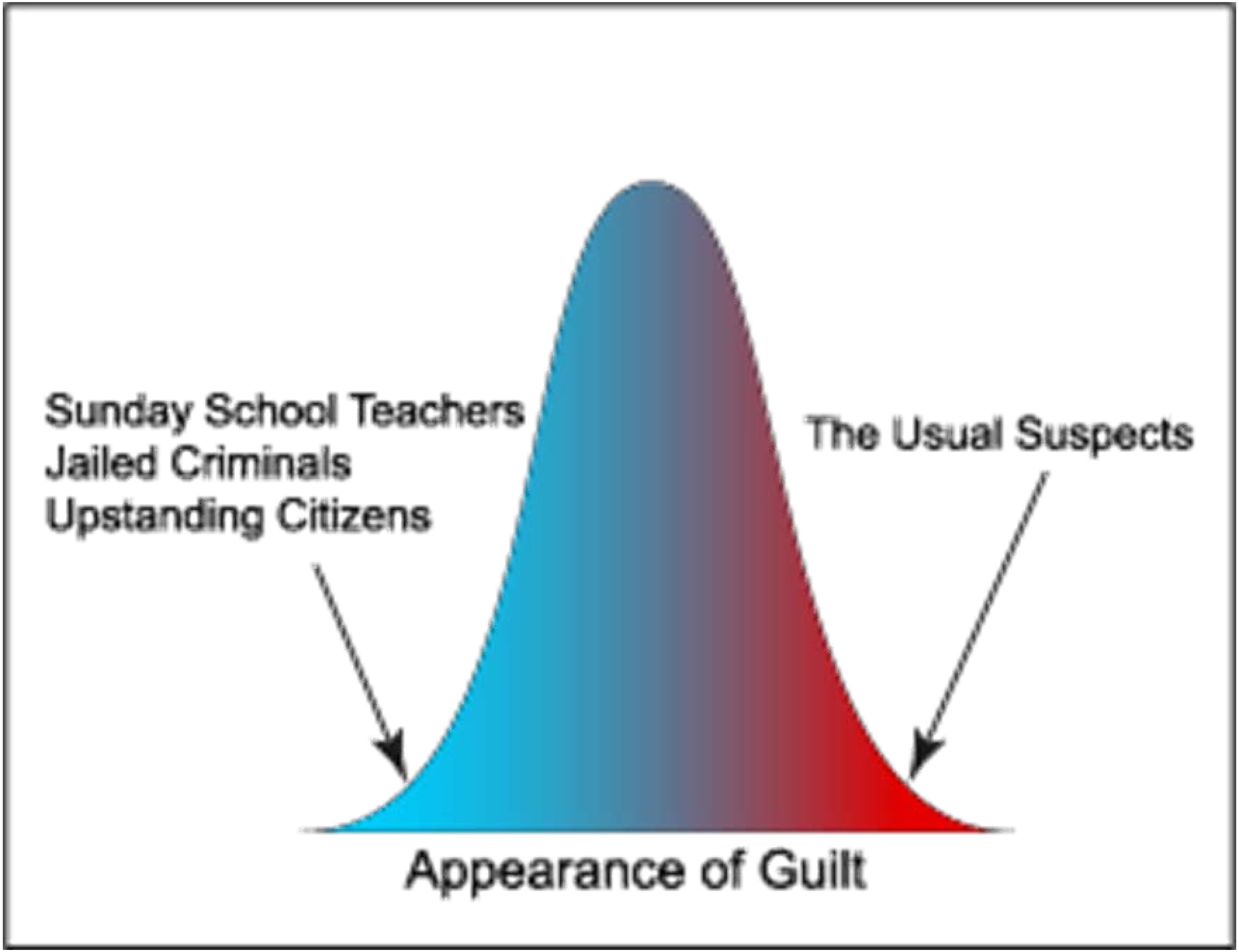
	Defendant Innocent	Defendant Guilty
Guilty Verdict: Reject presumption of innocence	**Type I Error**	**Correct**
Not Guilty Verdict: Fail to reject presumption of innocence	**Correct**	**Type II Error**

- **Type I error:**

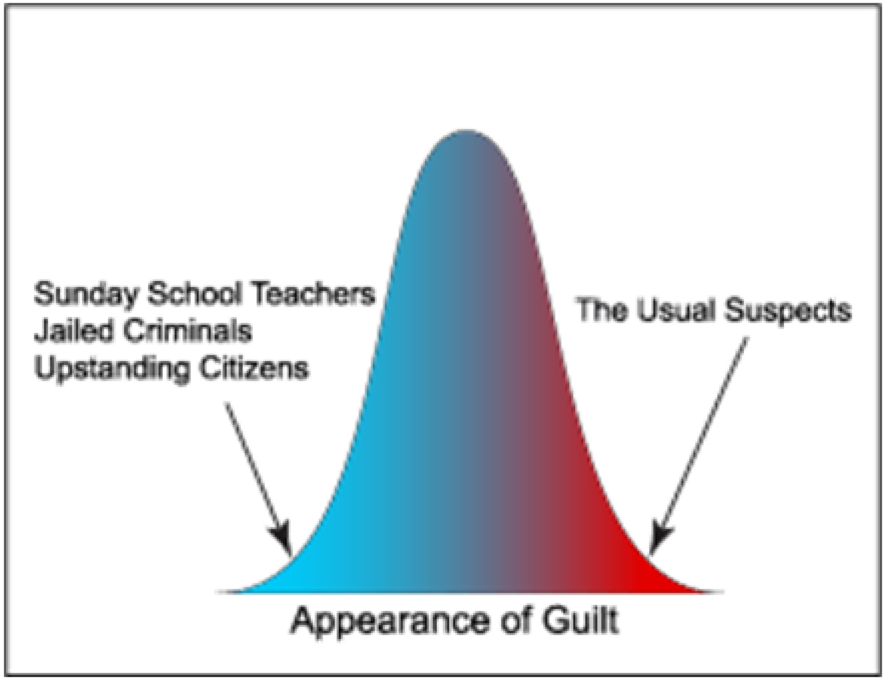
  false alarm

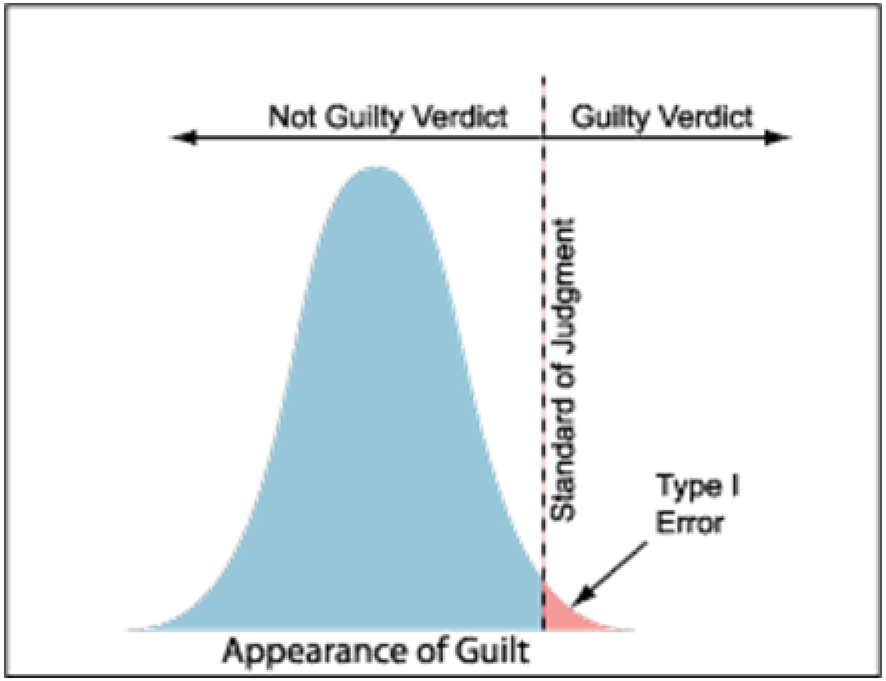
- **Type II error:**

  failed alarm



Sunday School Teachers
Jailed Criminals
Upstanding Citizens
The Usual Suspects
Appearance of Guilt


Sunday School Teachers
Jailed Criminals
Upstanding Citizens
The Usual Suspects
Appearance of Guilt


Not Guilty Verdict
Guilty Verdict
Standard of Judgment
Type I Error
Appearance of Guilt

Science hypotheses

- In science: we disprove unsatisfactory hypotheses → propose & test new hypothesis

- In statistics: we start with a ***null*** hypothesis which we *assume* is correct → our goal is to reject the null in favor of the ***alternative*** hypothesis

Hypotheses

- ***Null* Hypothesis (H_o)** = what we're trying to disprove
- ***Alternate* Hypothesis (H_A)** = *w*hat we think might really be going on

- **Test**:
  - Can we reject H_o in favor of H_A ?
- **Decisions**:
  - Reject
  - Fail to reject
- **Errors**:
  - *Type I:* Reject H_o when H_o is really true.
  - *Type II:* Fail to reject H_o when H_o is really false.

How do we reject or accept H_o?

- Decide the appropriate test statistic to use
- Set up the rejection region
- Calculate the test statistic
- Draw your conclusion: reject or fail to reject
- Interpret your results

Types of Error

- **Type I (alpha error) = p-value:**
  - – Probability of rejecting H_o when it is correct
  - – Probability that your results occurred by chance alone

- **Type II (beta error):**
  - – Probability of accepting H_o when is not correct
  - – Probability of missing a true difference

Reporting the p-value

- $P > 0.05$ → fail to reject H_o
  (no effect / no difference / no relationship?)

- $P < 0.05$ → reject H_o
  (how big or how small is the effect?)

Reporting the p-value

- **null hypothesis** = that there is no effect
- The effect is seldom zero
- Estimate the magnitude of the effect

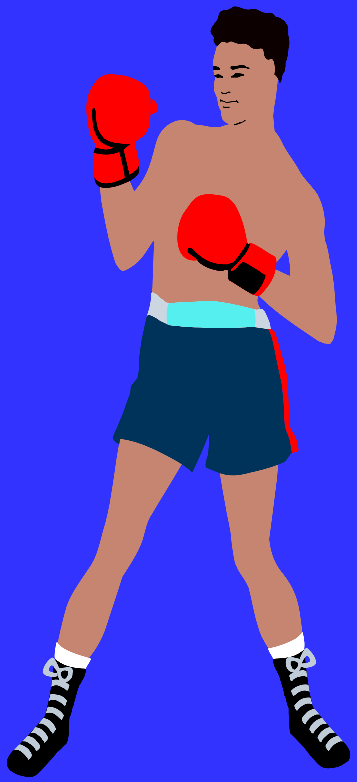
✓**Confidence Intervals**

Test of significance

- Strength of evidence against null hypothesis
- P < .05 Statistically Significant
- P > .05 Statistically not Significant
- Clinical vs statistical significance

Significance

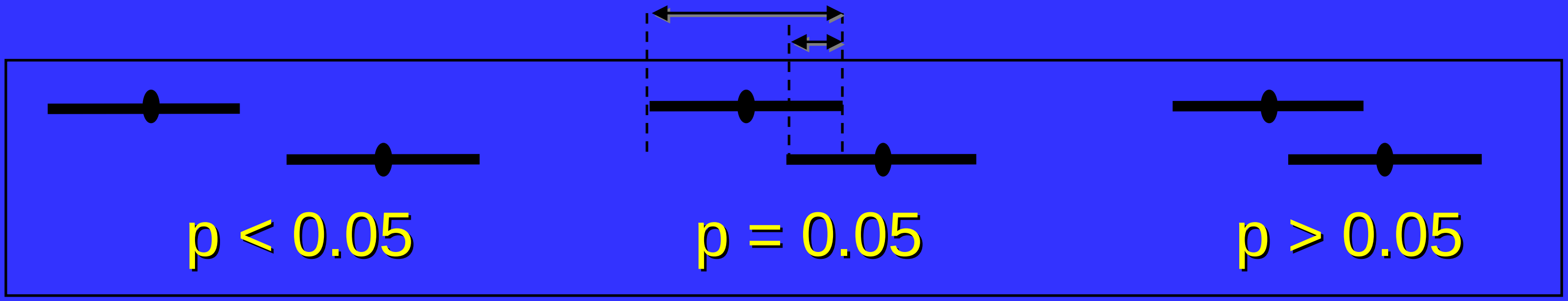
Statistical VS Clinical



Confidence Intervals (CI)

- Quantitative benefit of the intervention
- 95% CI: confident true value lie between point estimate
- statistical precision
- 95% chance the interval includes the true effect size
- 5% fall outside these limit

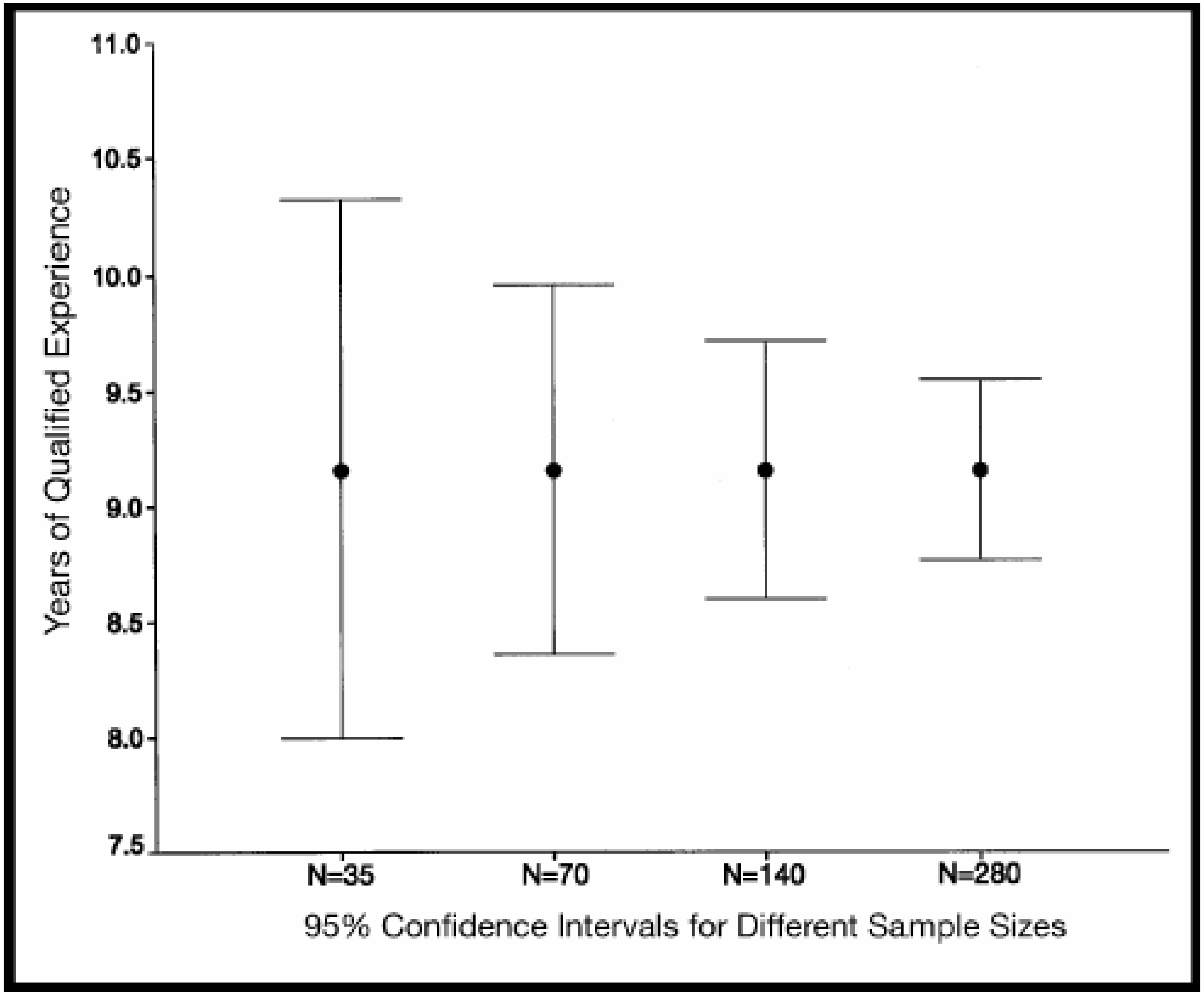
Statistical significance & confidence intervals



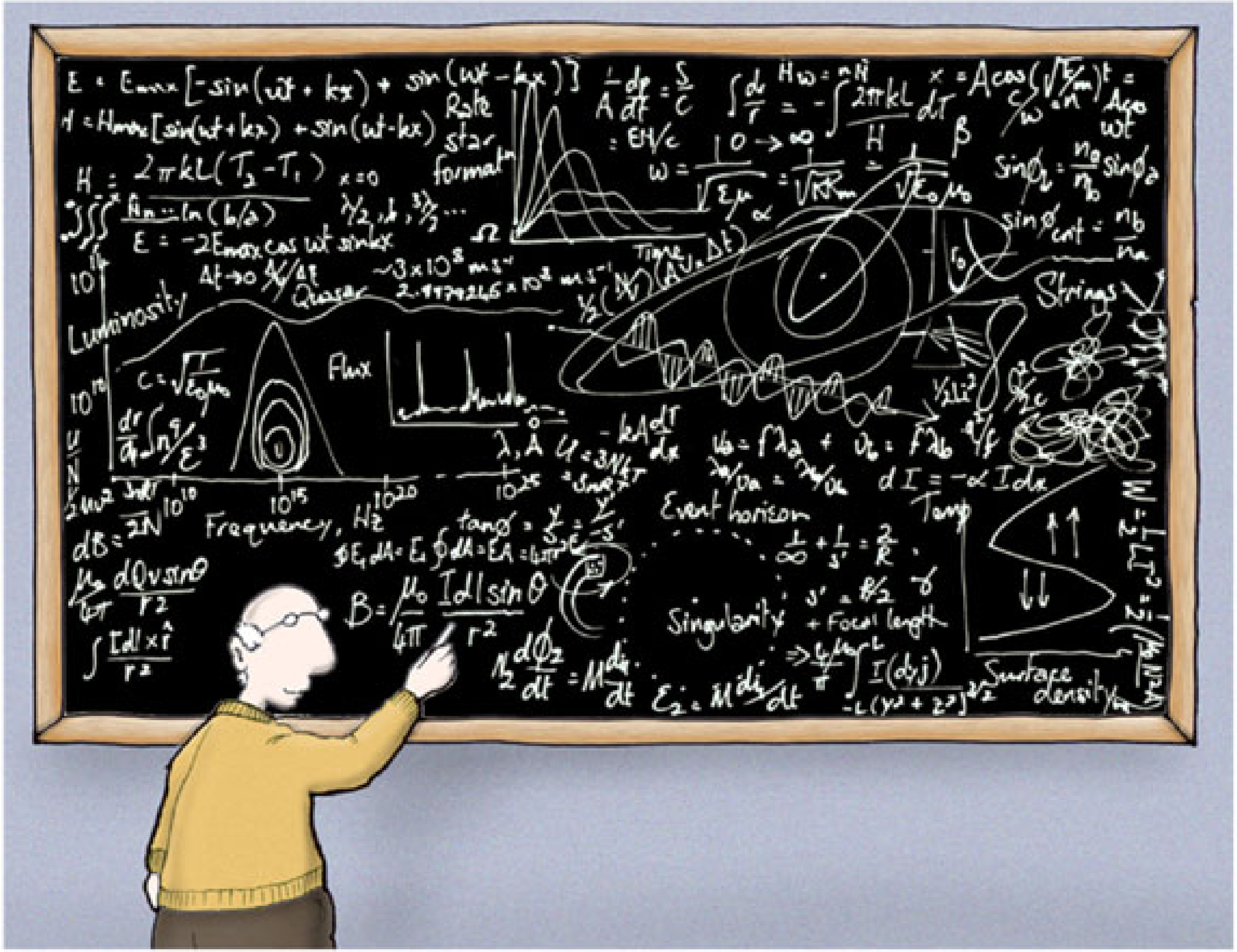
p < 0.05

p = 0.05

p > 0.05



As the sample size increase, the CI decrease



Astrophysics made simple

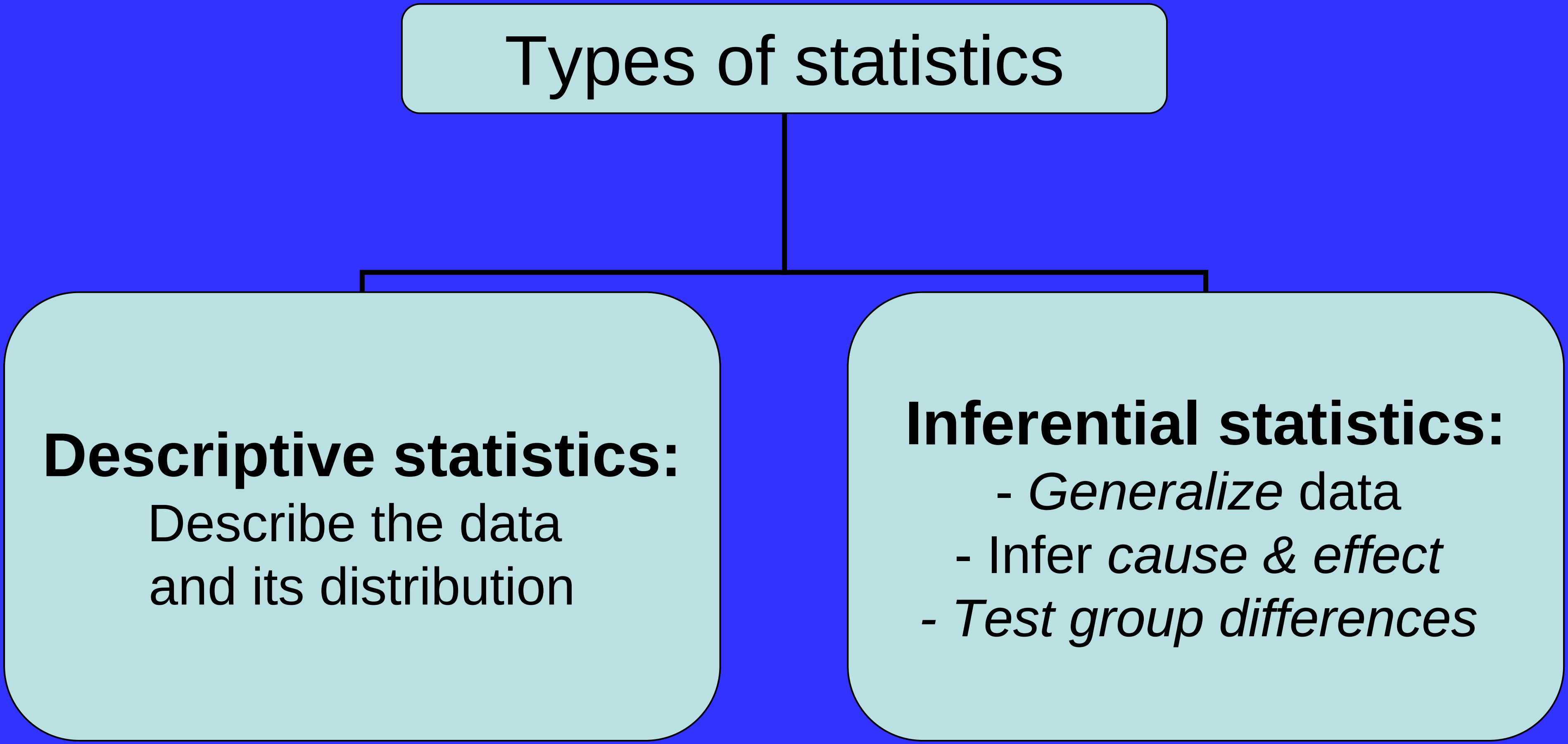
Statistics

- We muddle through life making choices based on incomplete information

- *Statistics* help us quantify *uncertainty*

Statistics

- = applied mathematics and rules of probability which allow researchers to make sense of their data



Types of statistics

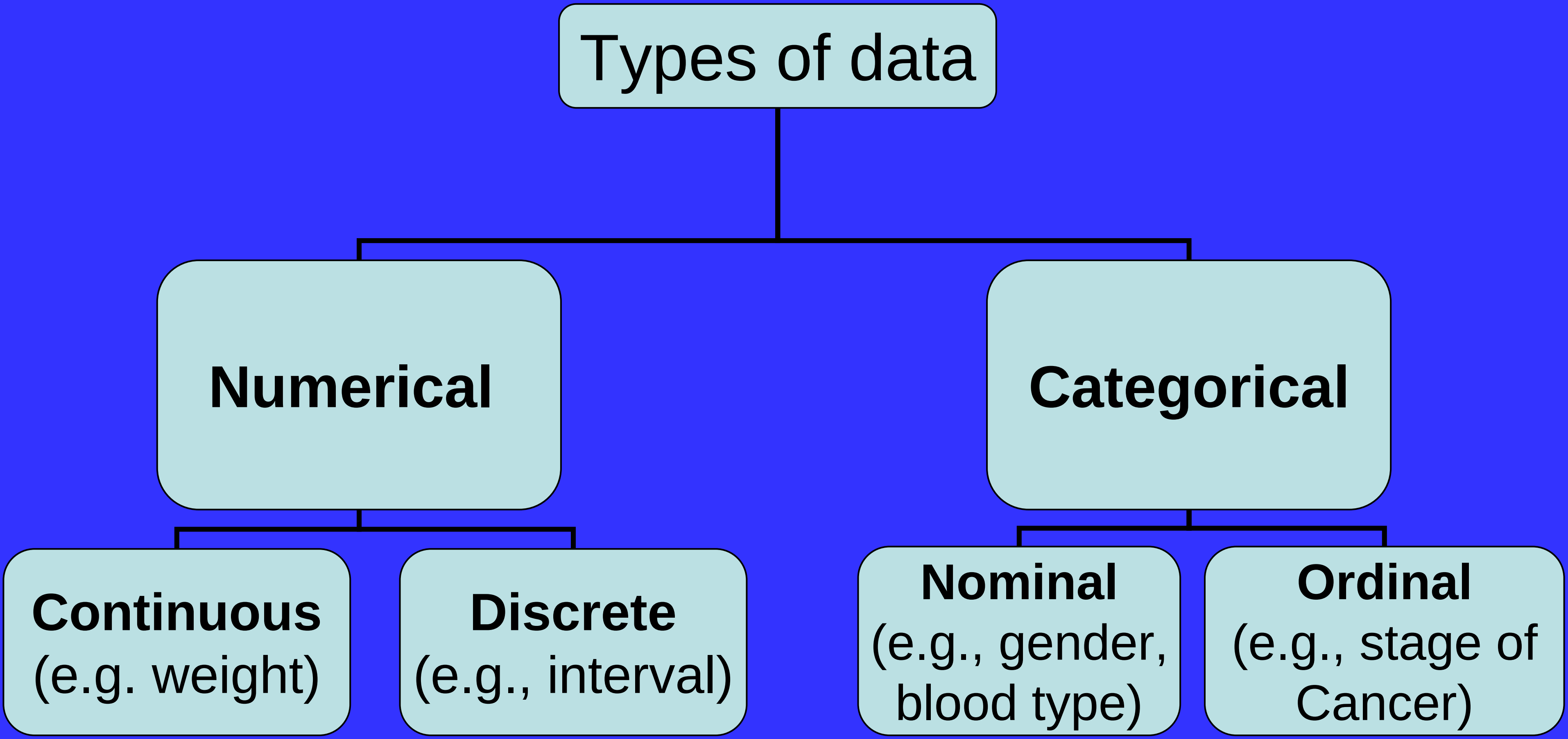
Descriptive statistics:
Describe the data
and its distribution

Inferential statistics:
- *Generalize* data
- Infer *cause & effect*
- *Test group differences*

Statistics

- **Descriptive:** summarization, organization, classification and tabulation

- **Analytic:** making estimate, conclusion and decision



Types of data

Numerical

Categorical

Continuous
(e.g. weight)

Discrete
(e.g., interval)

Nominal
(e.g., gender, blood type)

Ordinal
(e.g., stage of Cancer)

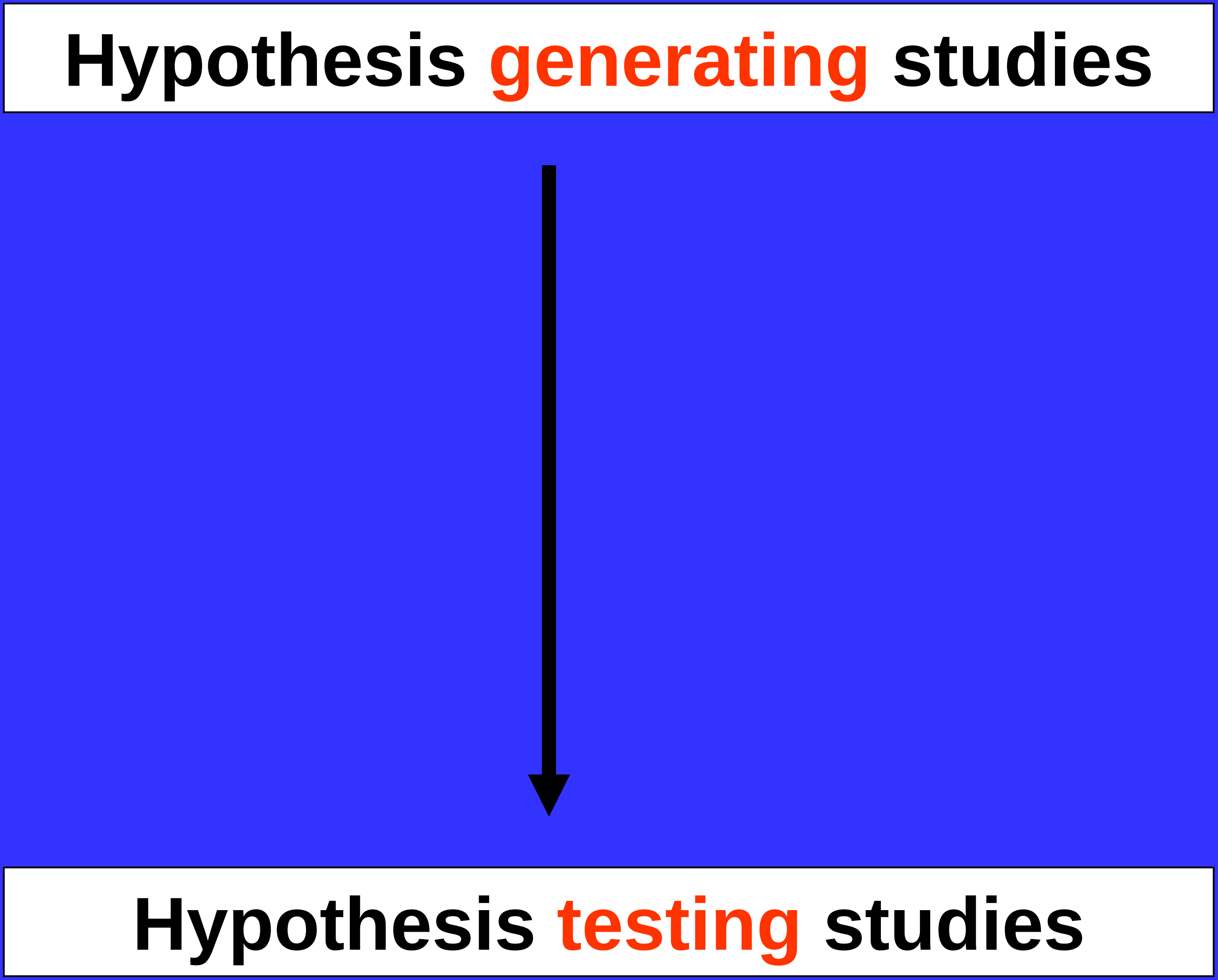
Where do hypotheses come from?

- Casual observation in the clinical setting
  - Not all techniques used in the clinics are based on facts

- Theory testing
  - ➢Theory = guess
  - ➢Most therapeutic approaches are based on theory, and therefore must be tested scientifically

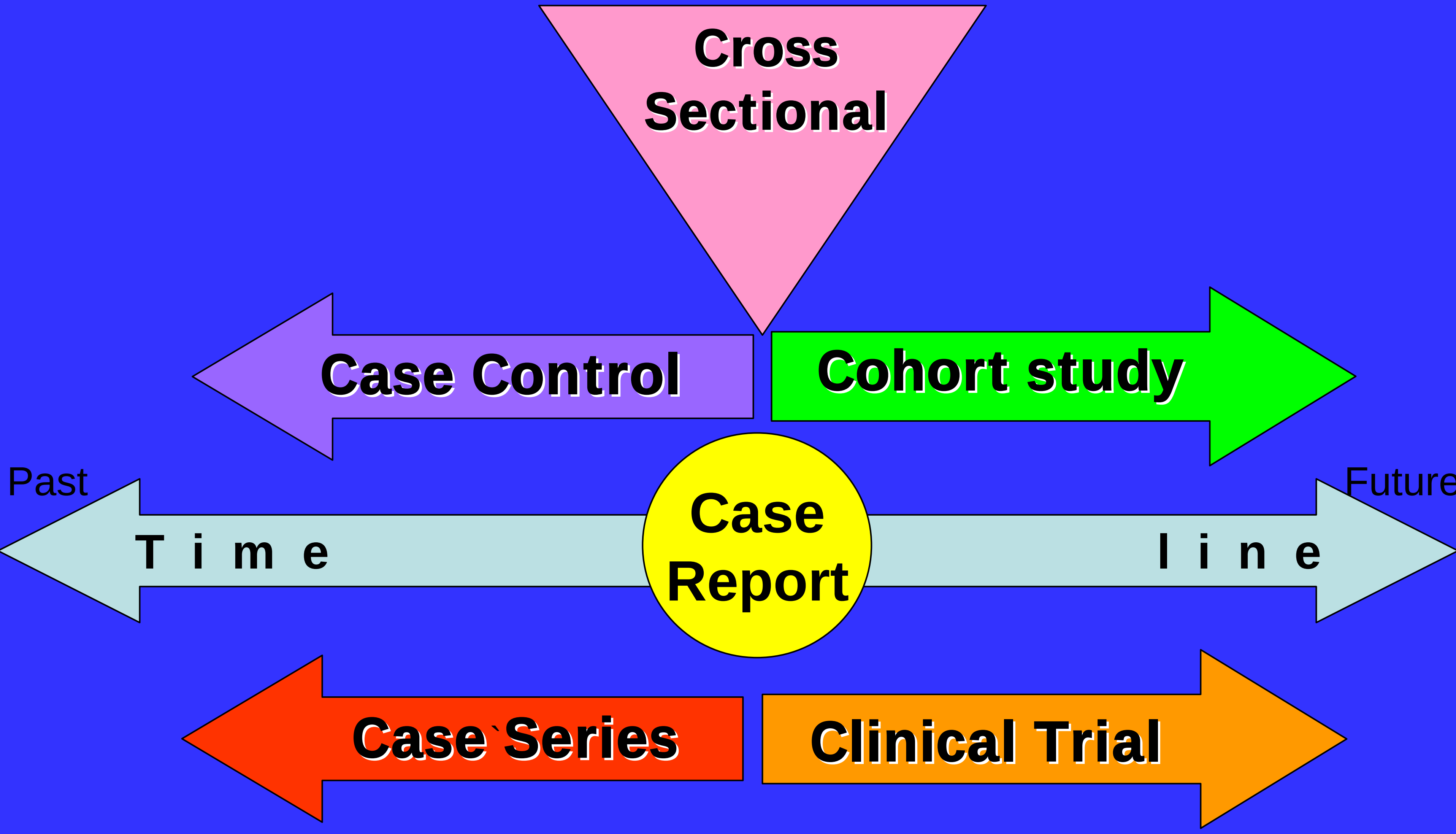
Where do hypotheses come from?

- Reading and analyzing the literature in a specific area of interest
- Contradictory research findings



Hypothesis generating studies

Hypothesis testing studies



Cross Sectional

Case Control

Cohort study

Past

Future

Time

Case Report

line

Case`Series

Clinical Trial

Research purposes

1. Description of a phenomenon (**descriptive** research)

2. Analysis of **relationships**

3. Analysis of **difference** between groups or treatments

Example

- **Topic**: functional recovery after total knee replacement (TKR)

Descriptive (observational) study

Purpose:

1. To ***describe*** the functional status of patients at various intervals after TKR

(case-report, case-series, cross-sectional)

Descriptive (observational) study

Purpose:

- To examine the ***relationship*** between preoperative factors (gait velocity, quadriceps strength) and functional status at intervals after TKR

Experimental (intervention) study

Purpose:

3. To examine the ***differences*** in functional recovery between a group of patients who received individualized postoperative exercise program versus another group who participated in a group exercise program

Descriptive statistics

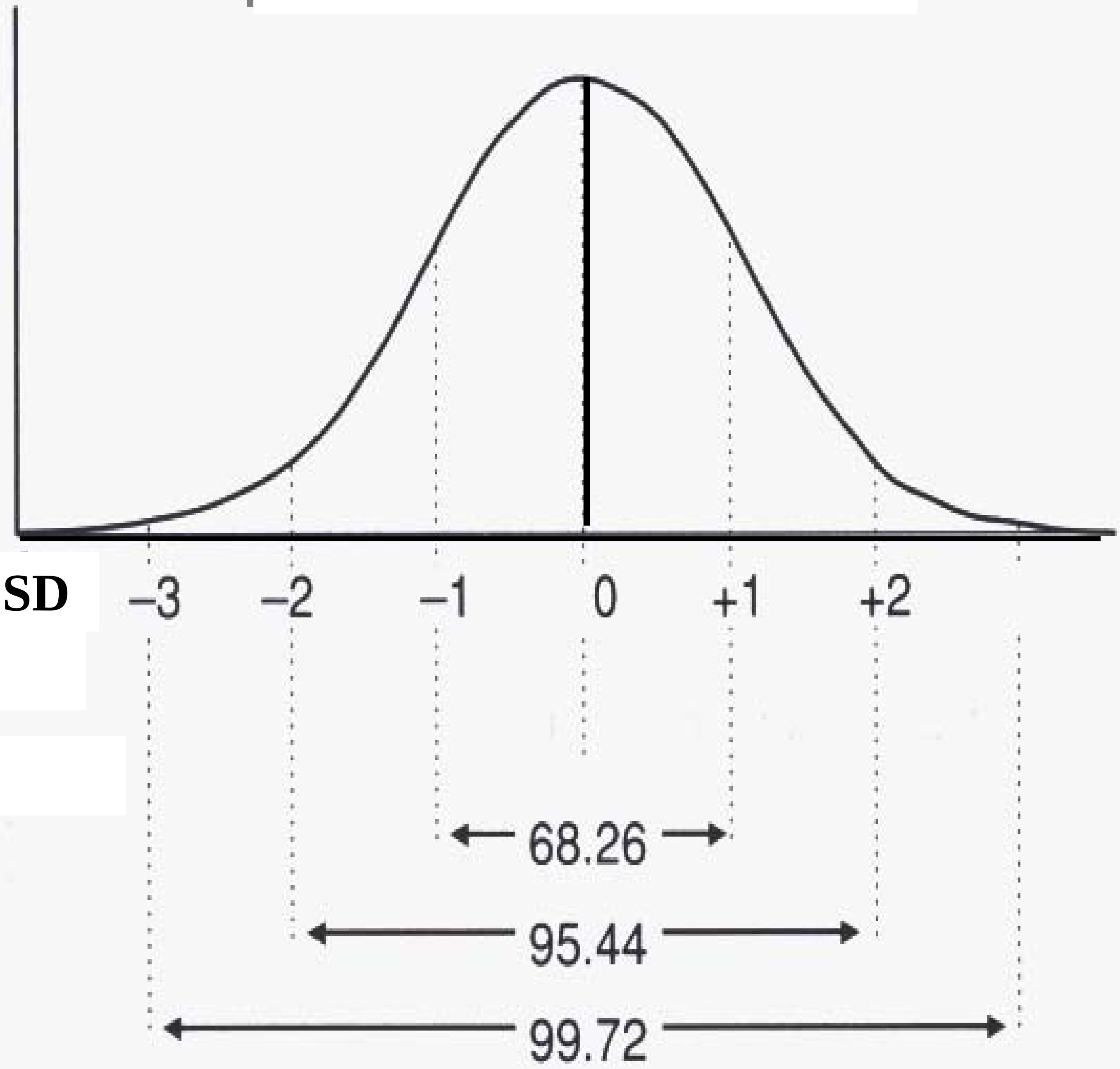
- **Mean:** sum of variables / number of variables
- **Median:** average (50% above & 50% below)
- **Mode:** most frequent occurring value
- **Range:** from lowest to highest
- **SD:** average difference from the mean

The normal distribution

- A symmetric frequency distribution (bell-shaped curve) that can be defined by the mean and standard deviation

- The distribution is symmetric around the mean

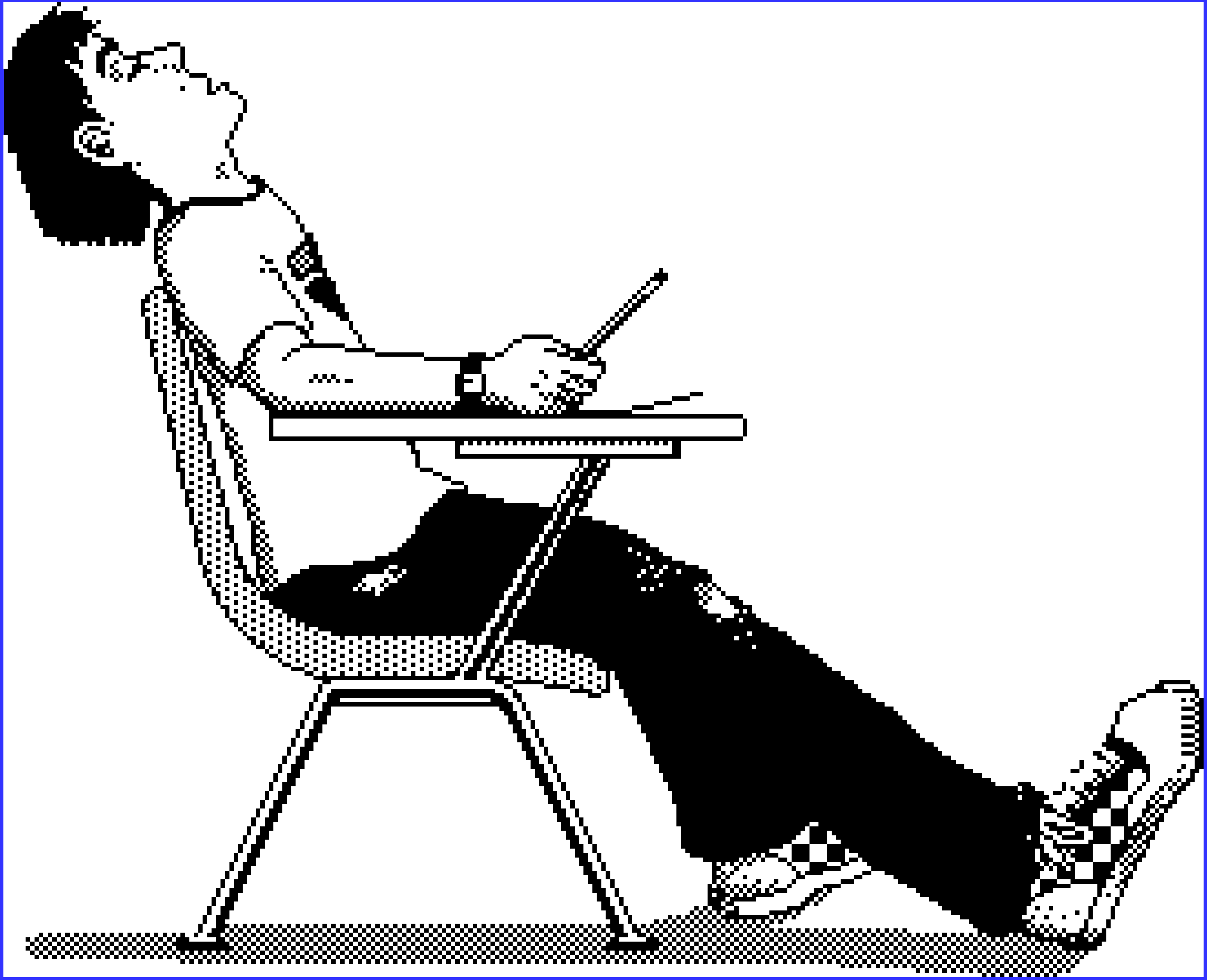
Normal distribution



Types of statistical tests

- **Parametric statistics:**
  - – To describe normally distributed data
  - – For continuous variable

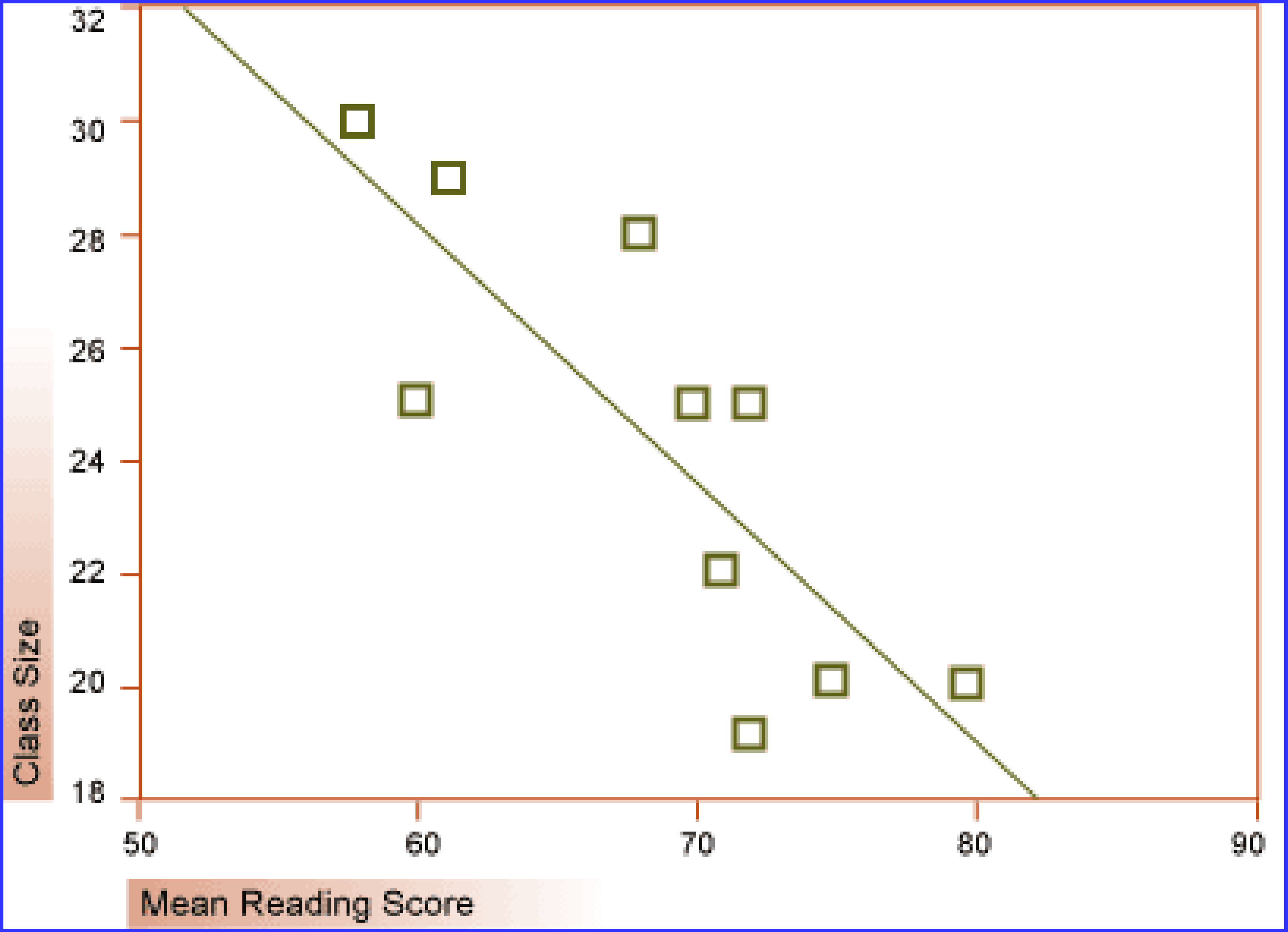
- **Non-parametric statistics:**
  - – When the distribution is not-symmetrical or unknown
  - – For nominal or ordinal data



Correlation

- How closely do two factors follow each other? (e.g., height and weight)

- Does not assume cause-and-effect relationship


32
30
28
26
24
22
20
18
Class Size
50
60
70
80
90
Mean Reading Score

Linear Regression

- Can height predict weight?

  weight = a + **b** (height)

- We can calculate the significance of **b** (is **b** significantly different from zero)

Multiple Linear Regression

- Weight = **a** + **b** (height) + **c** (calories)

- Can calculate the significance of any of **a**, **b**, **c**, ….etc.

Logistic Regression

- Used to determine the effect of a variable on a binominal outcome

  (e.g., dead or alive)

Compare means

- **Unpaired T-test:**
  - To compare two independent groups

- **Paired T-test:**
  - Uses before and after data
  - Less variability → easier to achieve significance

Compare means

- **If > 2 groups:**

Analysis of Variance (ANOVA):

- Tells you if more than 2 groups are different
- H_o: all the means are equal

  H_1: not all the means are equal
- Compares variances within groups to variances between groups (F-value)
- It does not tell you which group is different!

Compare means

- **Multiple Analysis of Variance (MANOVA):**
  - Used to determine not only that there are differences between the means, but what differences are significant

Differences between groups

Normal distribution:

Compare the means

Not-normal distribution:

Compare the median

Non-parametric statistics

- **Ordinal data:**
  - Wilcoxon signed rank test
- **Proportions:**
  - Chi-square test
  - Fisher’s exact test

Sample size

- Number of participants needed to detect difference between the groups

•How large? Every body? 100 - 200 patients?

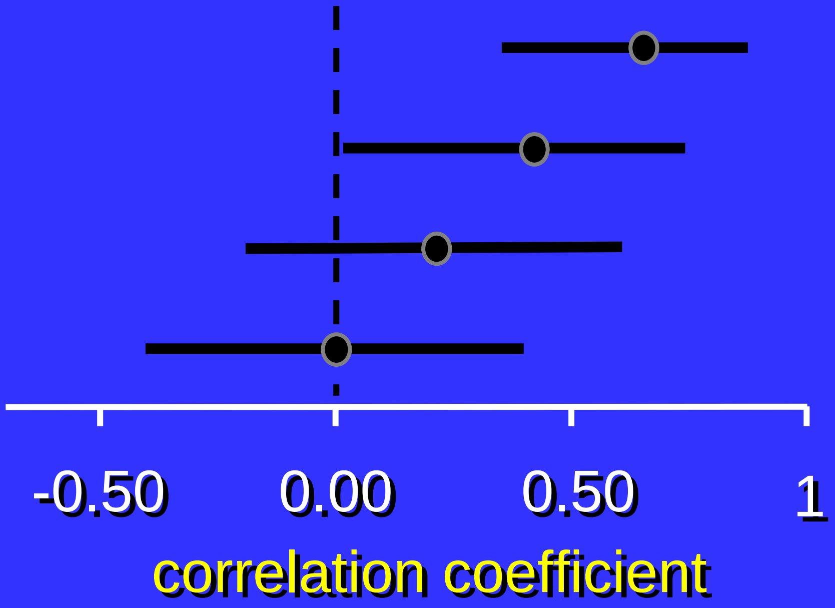
- Too large: costly, longer time, unnecessary patients.

Power

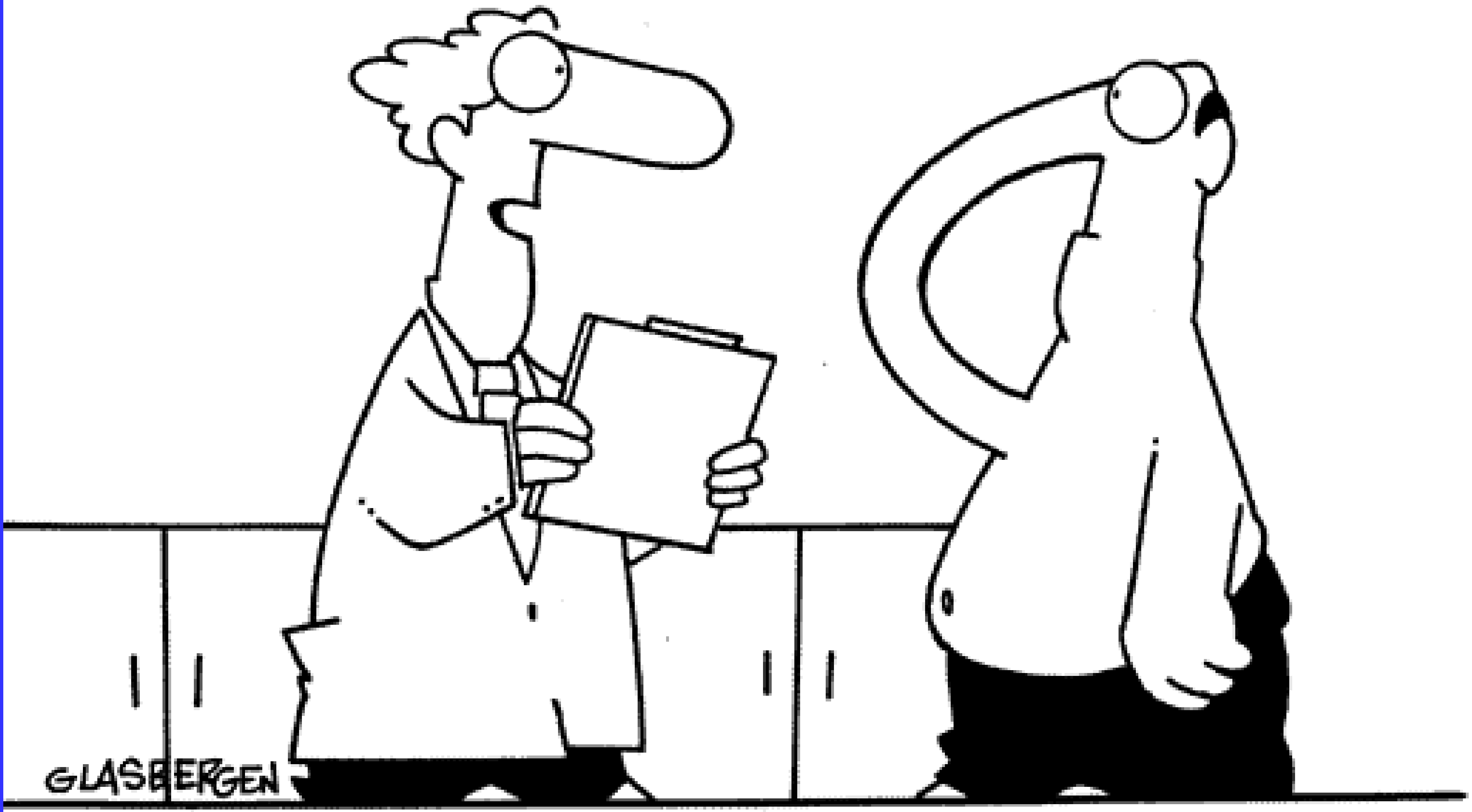
- = probability of finding a true difference

  (1-Beta)

- "Statistically Significant"
  - – P < 0.05
  - – Zero lies outside the confidence interval.
    - Examples: four correlations for samples of size 20.



r	likely range	P
0.70	0.37 to 0.87	0.007
0.44	0.00 to 0.74	0.05
0.25	-0.22 to 0.62	0.29
0.00	-0.44 to 0.44	1.00



"It's an experimental procedure. Every time you blow your nose, you'll clear out your arteries!"