Basics of Statistical Analysis

Maj Gen (R) Suhaib Ahmed, HI(M)
MBBS; MCPS; FCPS; PhD (London)

Genetics Resource Centre (GRC)

www.grcpk.com

Statistics: The Science of

- Data collection
- Data presentation
- Data interpretation
- Statistical analysis

"Data don't make any sense,
we will have to resort to statistics."

“I can prove it or disprove it! What do you want me to do?”

Statistics ~~=~~ *p* value

Statistics

- Why ?
- When ?
- How ?

Statistical Analysis

- Descriptive
- Comparison
- Correlation
- Probability and risk analysis
- Survival analysis

Types of Data

- Categorical data:
  - Two categories: Male /Female, True/False
  - >Two categories: Ethnic groups, Blood groups
    - Ordered categories: Mild Moderate Severe
- Numerical data:
  - Discrete: Number of children
  - Continuous: Blood glucose, Hb
- Other types:
  - Rates
  - Ratios

- Mean
  - Arithmetic average
- Median
  - Middle value when ranked in order
- Mode
  - Value that occurs most often

Why do we need statistical analysis?

- To use the information gained from a "sample population" to make inferences about the "actual population"

Population

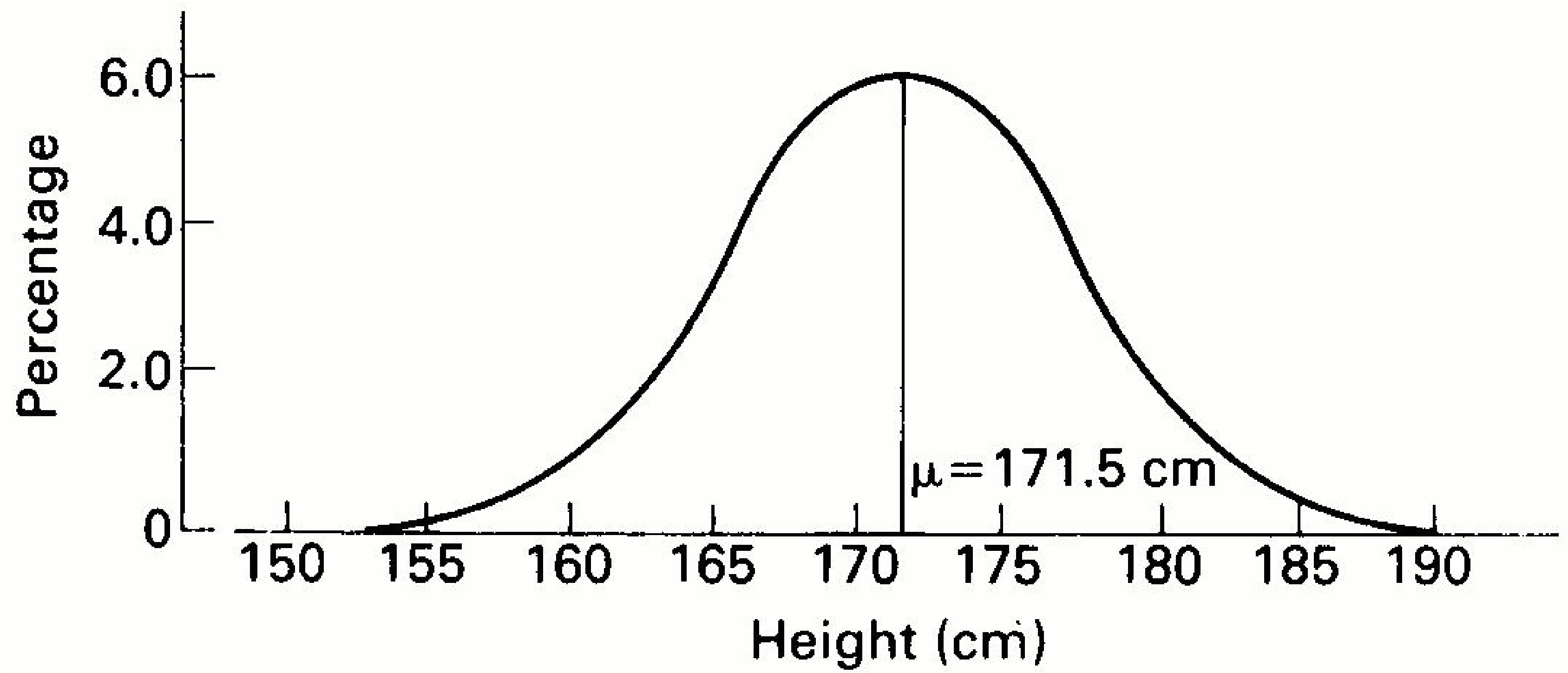
- Inhabitants of a place
- Patient population
- Sample population

Distribution of Population

- Normal
- Abnormal

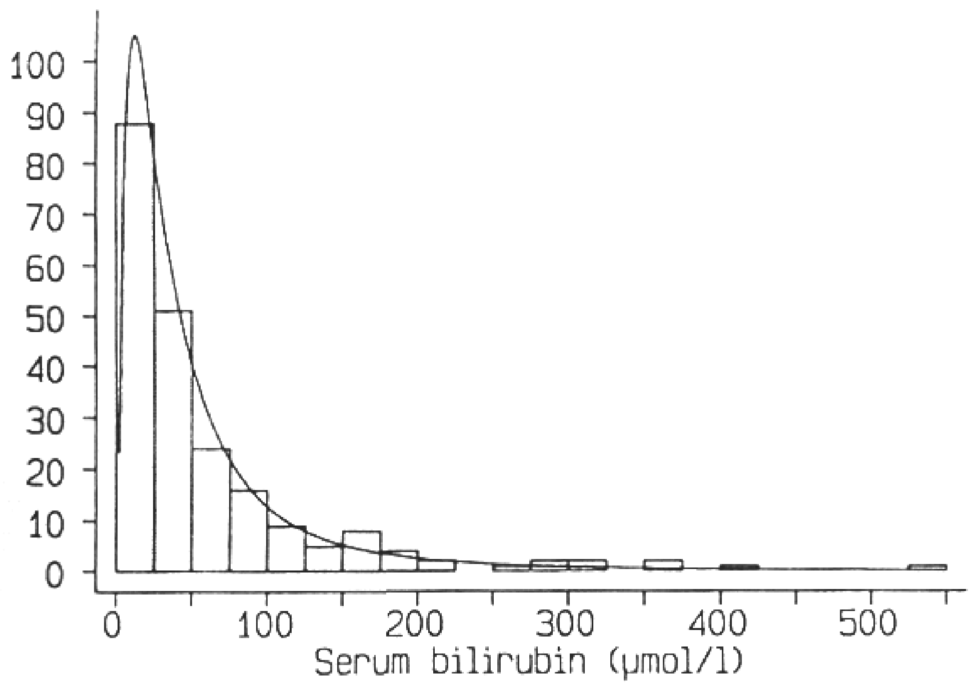
Normal Distribution:

Large sample

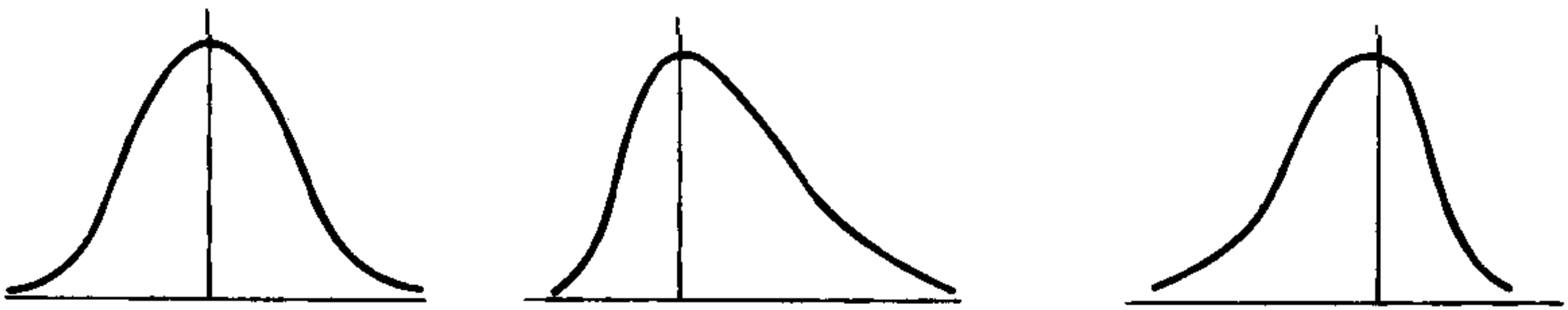


Abnormal Distribution

(Selected sample)



Abnormal Distribution

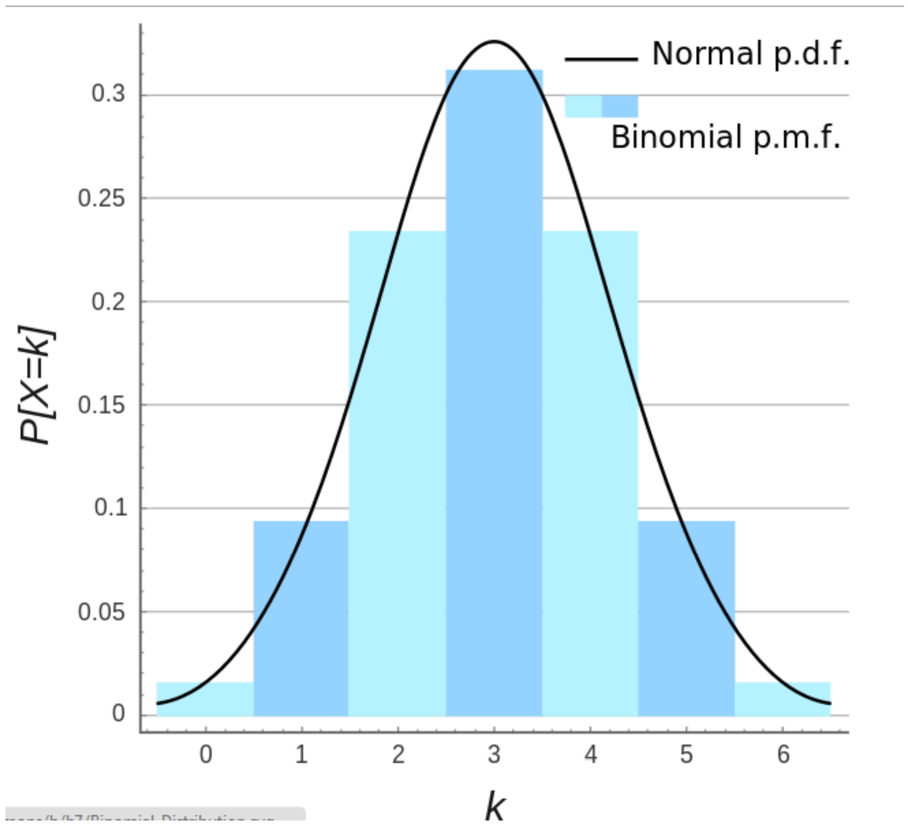


(a) Symmetrical and bell-shaped

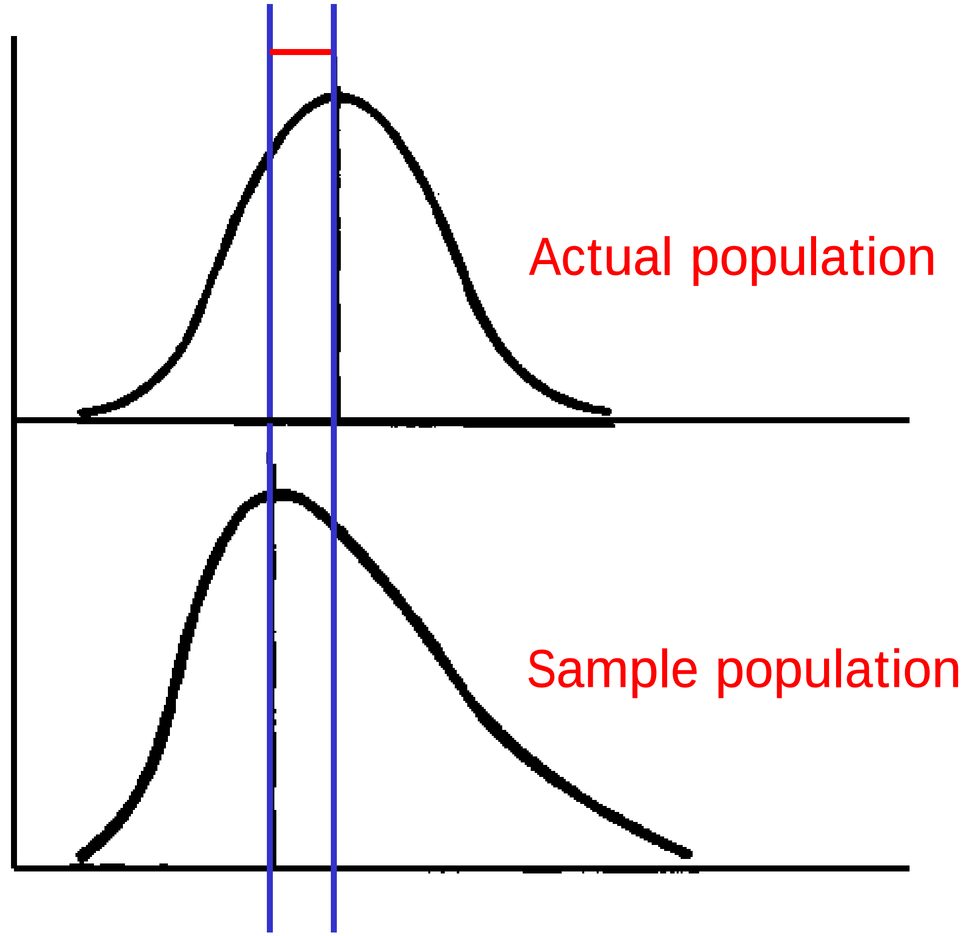
(b) Positively skewed or skewed to the right

(c) Negatively skewed or skewed to the left

Binomial Distribution



Standard Error of Mean (SEM)



Mean Haemoglobin

- 2.5
- 4.5
- 7.2
- 7.9
- 8.7
- 9.6
- 9.8
- 10.5
- 11.7
- 13.2
- 14.1

Mean: 9.1

Range: 2.5-14.5

SD: 3.32

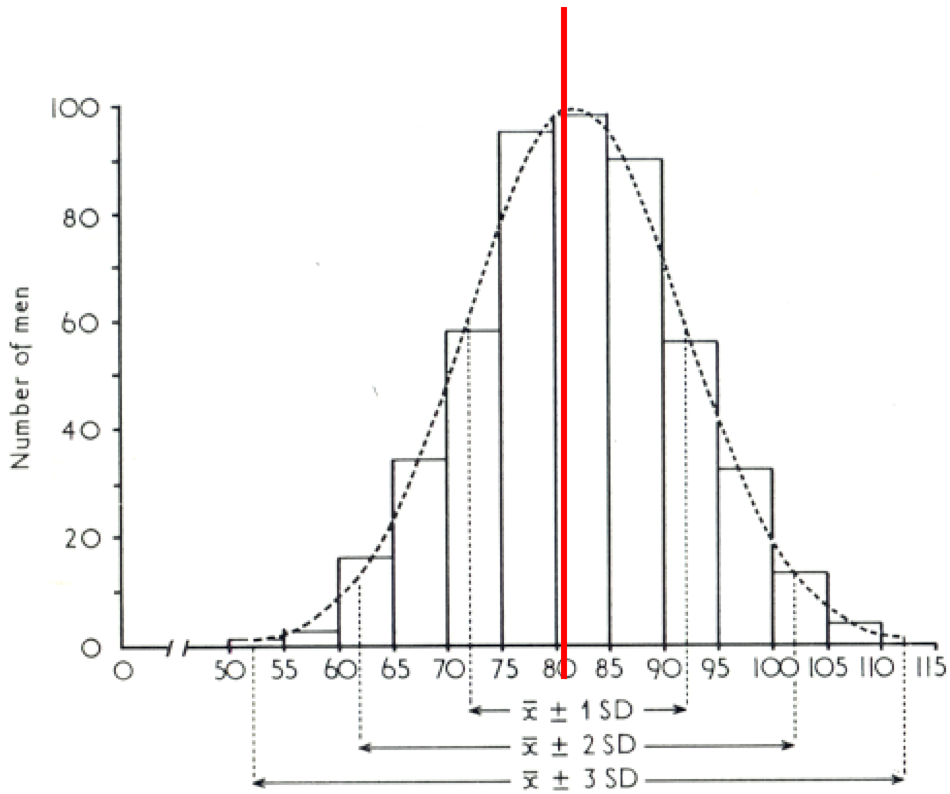
Mean Haemoglobin in Two Groups

Group-I	Group-II
2.5	7.3
4.5	8.4
7.2	8.6
7.9	8.7
8.7	9.0
9.6	9.1
9.8	9.6
10.5	9.7
11.7	10.1
13.2	10.2
14.1	10.3
Mean: 9.1	Mean: 9.2
Range: 2.5-14.5	Range: 7.3-10.3
SD: 3.32	SD: 0.87

Dispersion around the Mean
(Expression of Variability)

- Range
- Standard Deviation (SD)
- Coefficient of Variation (CV)
- Standard Error of the Mean (SEM)

Standard Deviation (SD)



- +1 SD: 70%
- +2 SD: 95%
- +3 SD: 99%

Statistical Analysis

- Parametric Tests
  - Student's t-test
- Non Parametric Tests
  - Mann-Whitney U test

Statistical Analysis:
Comparison between groups

- Hypothesis testing (*p* value)
- Confidence interval (CI)

Comparison between groups:
Hypothesis testing

- Null hypothesis
  - There is no difference between the groups
- Alternate hypothesis
  - There is a difference between the groups

Comparison between groups:

Hypothesis testing

- Null hypothesis
  - There is no difference between the groups
- Alternate hypothesis
  - There is a difference between the groups

- P value

- Probability that the null hypothesis is correct

Interpretation of P value

- 1.0 = 100%
- 0.5 = 50%
- 0.29 = 29%
- 0.12 = 12%
- 0.07 = 7%
- 0.05 = 5%
- 0.01 = 1%
- 0.001 = 0.1%

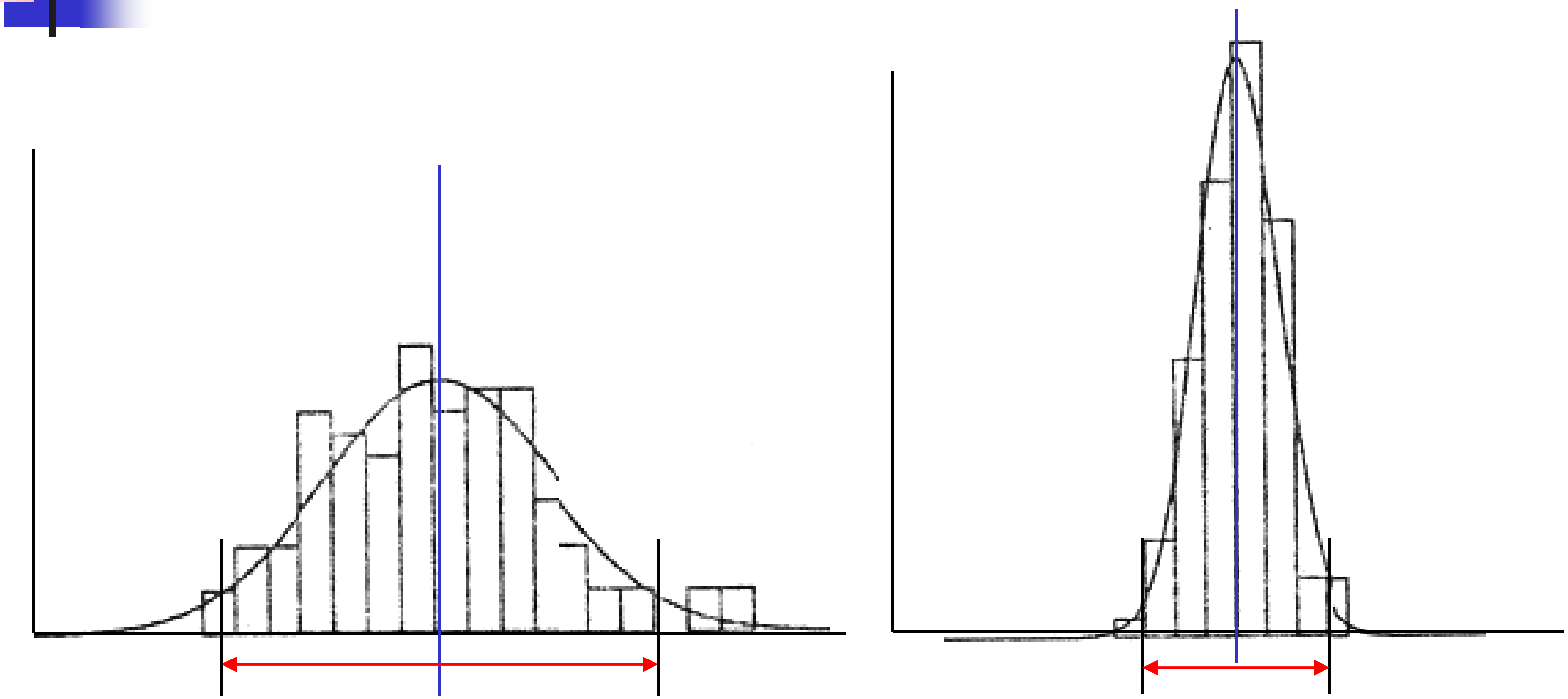
Interpretation of P value

Significant and not significant

- P = 0.03
- P = 0.05
- P = 0.09

Comparison between groups:

Hypothesis testing



Mean:5.1 g/dl
Group-I
n = 39

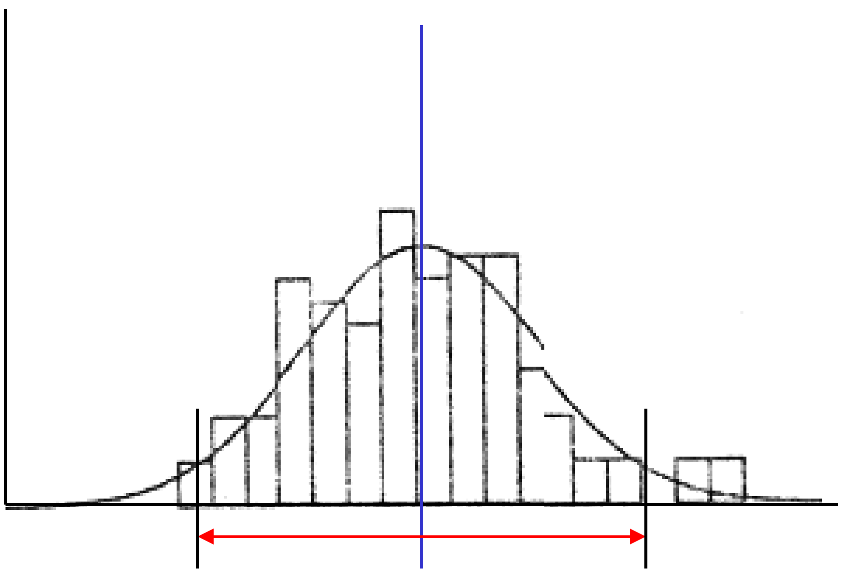
Mean: 12.2 g/dl
Group-III
n = 458

P = 0.03 (P < 0.05)

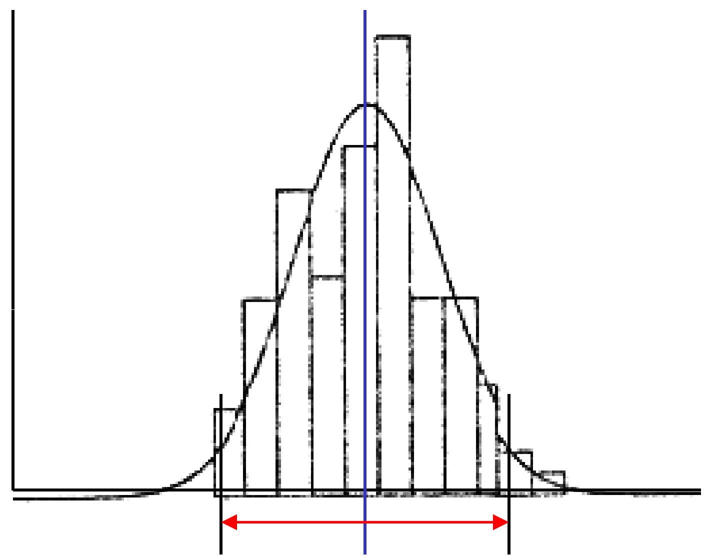
There is a statistically significant difference between the two groups

Comparison between groups:

Hypothesis testing (*P* value)



Mean:5.1 g/dl
Group-I
n = 39



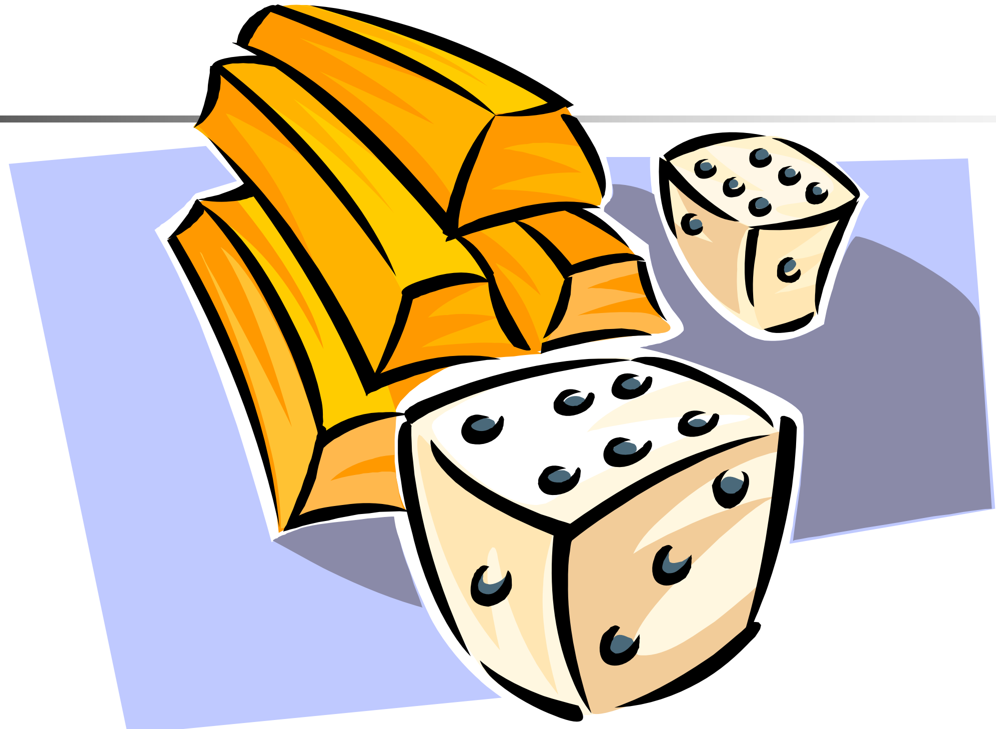
Mean: 7.4 g/dl
Group-II
n = 158

P = 0.09 (P > 0.05)

There is no statistically significant difference between the two groups

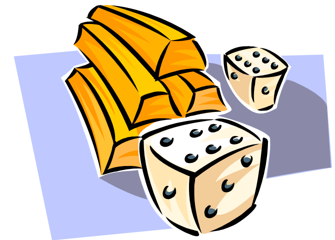
Confidence Interval (CI)

- A range (interval) in which one is confident that it contains the actual population mean
- <u>Example:</u> (95% CI= 8.4-10.5 g/dl)
- % CI ?
  - 90%
  - 95%
  - 99%



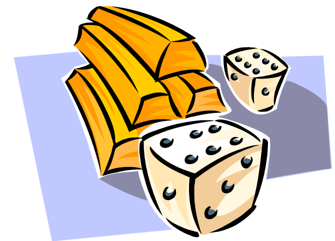
- Six sides
- Chance of each side: 1/6 (16.6%)

Rolling the Dice



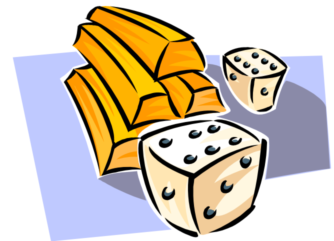
No:	Expected Frequency:	Observed Frequency:			
		I	II	III	IV
12 times	2/12 (17%)	3/12 (25%)	1/12 (8%)	4/12 (33%)	2/12 (17%)

Rolling the Dice



No:	Expected Frequency:	Observed Frequency:			
		I	II	III	IV
12 times	2/12 (17%)	3/12 (25%)	1/12 (8%)	4/12 (33%)	2/12 (17%)
48 times	8/48 (17%)	7/48 (15%)	6/48 (13%)	10/48 (21%)	8/48 (17%)

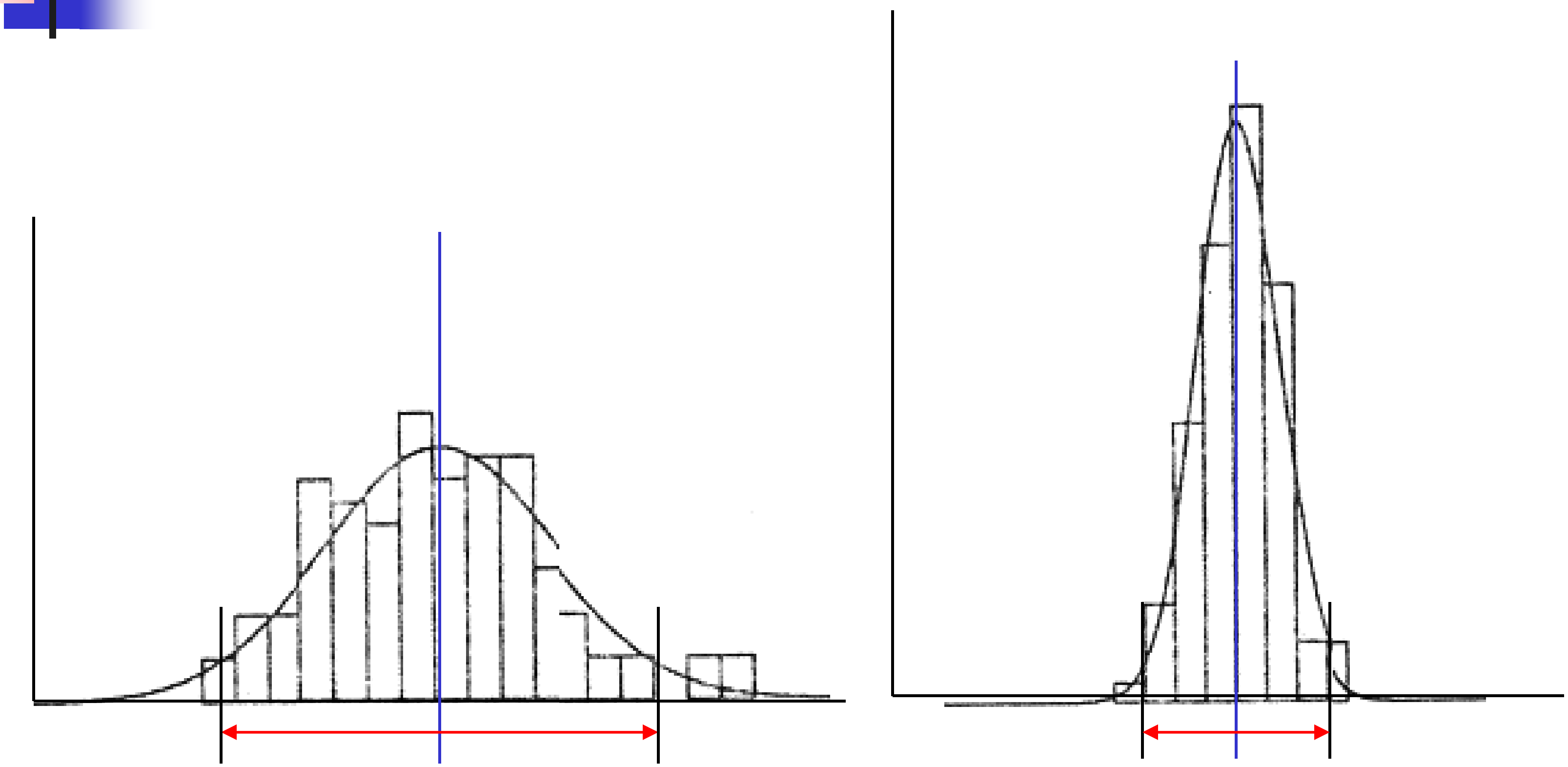
Rolling the Dice



No:	Expected Frequency:	Observed Frequency:			
		I	II	III	IV
12 times (8-33%)	2/12 (17%)	3/12 (25%)	1/12 (8%)	4/12 (33%)	2/12 (17%)
48 times (13-21%)	8/48 (17%)	7/48 (15%)	6/48 (13%)	10/48 (21%)	8/48 (17%)
192 times (16-19%)	32/192 (17%)	31/192 (16%)	32/192 (17%)	36/192 (19%)	35/192 (18%)

Comparison between groups:

Confidence Intervals: CI = $x \pm (t' \times SEM)$

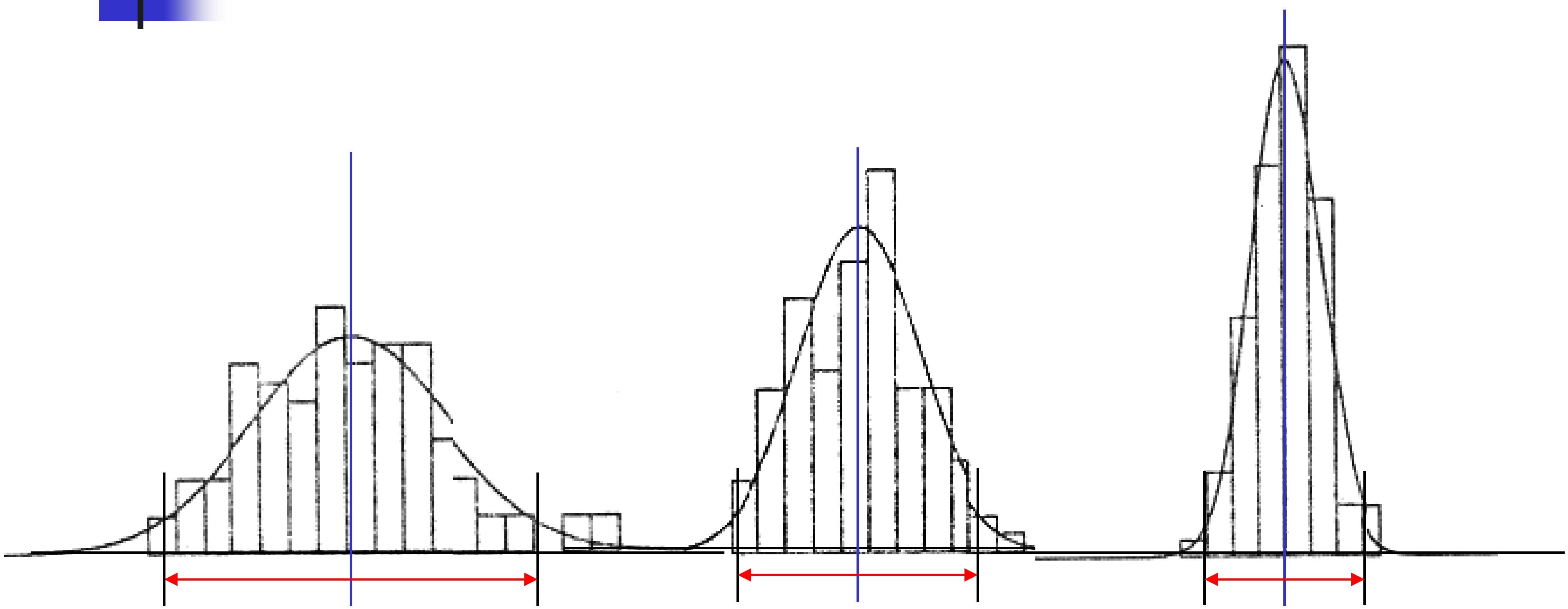


Mean:5.1 g/dl
95% CI: 2.5-8.1 g/dl
Group-I
n = 39

Mean: 12.2 g/dl
95% CI: 10.5-14.1 g/dl
Group-III
n = 458

Comparison between groups:

Confidence Intervals: CI = $x \pm (z' \times SEM)$

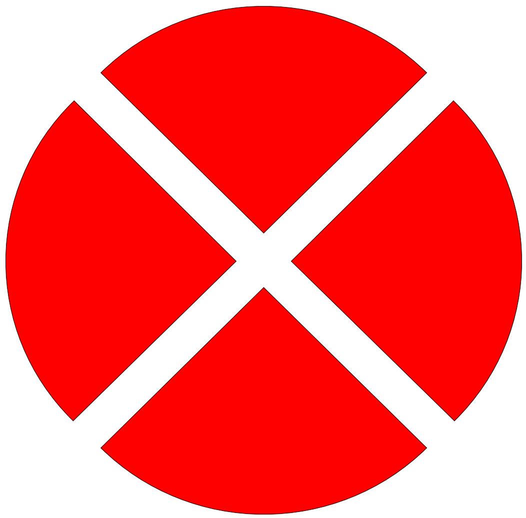


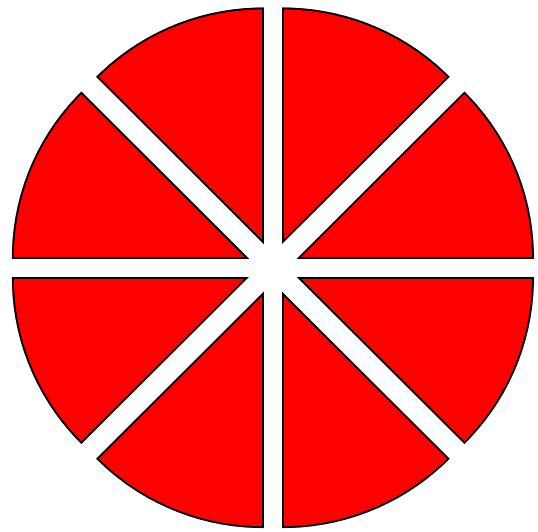
Mean:5.1 g/dl
95% CI: 2.5-8.1 g/d
Group-I
n = 39

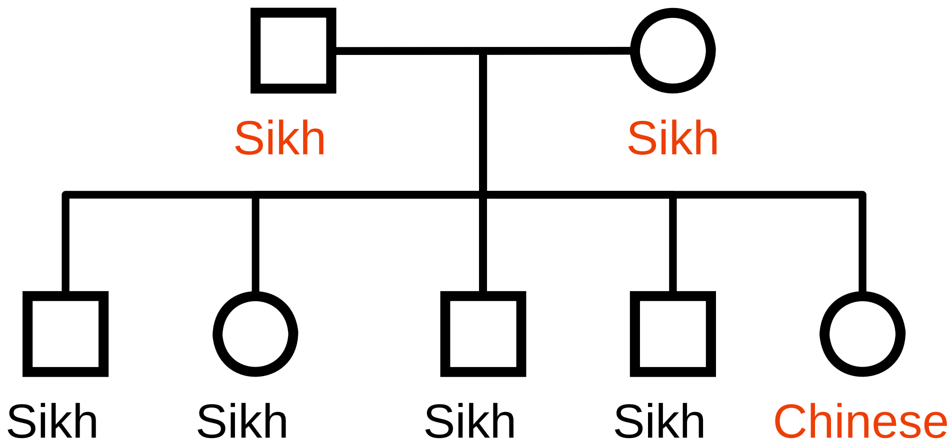
Mean: 7.4 g/dl
95% CI: 5.8-8.9 g/dl
Group-II
n = 158

Mean: 12.2 g/dl
95% CI: 10.5-14.1 g/dl
Group-III
n = 458

How many pieces would you like your pizza to be cut into? 4 or 8?






Sikh
Sikh
Sikh
Sikh
Sikh
Sikh
Chinese