Chi-Square Test, Fisher Exact Test, Graph Reading

Dr Komal Patel

1st year resident

Department of Pharmacology

Chi-square test

Introduction

- Chi-square test is a non-parametric test.
- It follows a specific distribution known as Chi-square distribution.
- The chi-square distribution is the most frequently employed statistical technique for the analysis of count or frequency of data.

Calculation of Chi-square value

The three essential requirements for Chi-square test are:

- A random sample
- Qualitative data
- Lowest data frequency not less than 5

The calculation of Chi-square value is as follows

- Make the contingency tables.

- Note the frequencies observed (O) in each class of one event, row-wise and the number in each group of the other event, column-wise.

- Determine the expected number (E) in each group of the sample or the cell of table on the assumption of null hypothesis.

- Find the difference between the observed and the expected frequencies in each cell (O – E).

- Calculate the Chi-square values by the formula.

- Sum up the Chi-square values of all the cells to get the total Chi-square value.

Formula for Chi – Square statistic

$$\chi^2 = \sum \left(\frac{(O - E)^2}{E} \right)$$

Where, O = Observed Frequency

E = Expected Frequency

Calculate the degrees of freedom which are related to the number of categories in both the events.

- The formula adopted in case of contingency table is

$$\textbf{Degrees of freedom (d.f.)} = (c - 1)(r - 1)$$

Where, c = columns

r = rows.

Example

In 1992, the U.S. Public Health Service and the Centers for Disease Control and Prevention recommended that all women of childbearing age consume 400 mg of folic acid daily to reduce the risk of having a pregnancy that is affected by a neural tube defect such as spina bifida or anencephaly. In a study by Stepanuk, 636 pregnant women called a teratology information service about their use of folic acid supplementation. The researchers wished to determine if preconceptional use of folic acid and race are independent. The data appear in Table.

TABLE -1: Race of Pregnant Caller and Use of Folic Acid

Race	Preconceptional Use of Folic Acid		
	Yes	No	Total
White	260	299	559
Black	15	41	56
Other	7	14	21
Total	282	354	636

TABLE -2: Observed and Expected Frequencies

Race	Preconceptional Use of Folic Acid		
	Yes	No	Total
White	260	299	559
	$= \frac{559 \times 282}{636} = 247.86$	$= \frac{559 \times 354}{636} = 311.14$	
Black	15	41	56
	$= \frac{56 \times 282}{636} = 24.83$	$= \frac{56 \times 354}{636} = 31.17$	
Other	7	14	21
	$= \frac{21 \times 282}{636} = 9.31$	$= \frac{21 \times 354}{636} = 11.69$	
Total	282	354	636

Cell	O	E	O - E	$(O - E)^2$	$\frac{(O - E)^2}{E}$
1	260	247.86	12.14	147.38	0.5946
2	299	311.14	-12.14	147.38	0.4737
3	15	24.83	-9.83	96.63	3.8917
4	41	31.17	9.83	96.63	3.1000
5	7	9.31	-2.31	5.34	0.5736
6	14	11.69	2.31	5.34	0.4568
∑					9.090

Calculation of test statistic: $\chi^2 = \sum\left(\frac{(O-E)^2}{E}\right)$

$$= \frac{(260-247.86)^2}{247.86} + \frac{(299-311.14)^2}{311.14} + \frac{(260-247.86)^2}{247.86} + \ldots + \frac{(14-11.69)^2}{11.69}$$

$$= 0.59461 + 0.47368 + \ldots + 0.45647$$

$$= 9.08960$$

$$\chi^2_{cal} = 9.0900$$

Step – 3: Chi – Square tabulated:

Degree of freedom = (c - 1) * (r - 1)

= (2 - 1)* (3 - 1)

= 2

$\alpha = 0.05$

$\chi^2_{tab} = 5.991$

Chi – Square tabulated at 5 % level of significance and with 2 d.f. is 5.991

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528

Step – 4: Decision rule.

Reject if the computed value of χ^2_{cal} is equal to or greater than 5.991.

Step – 5: Statistical decision.

We reject H_0 since $\chi^2_{cal} = 9.08960 > 5.991 = \chi^2_{tab}$

Step – 6: Conclusion.

We conclude that H_0 is false, and that there is a relationship between race and preconceptional use of folic acid.

Formula For 2×2 contingency table

Sometimes each of two criteria of classification may be broken down into only two categories. When data are cross classified in this manner, the result is a contingency table consisting of two rows and two columns. Such a table is commonly referred to as a 2×2 table. In the case of a 2×2 contingency table, however, χ^2 may be calculated by the following shortcut formula:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Example

According to Silver and Aiello, falls are of major concern among polio survivors. Researchers wanted to determine the impact of a fall on lifestyle changes. Table shows the results of a study of 233 polio survivors on whether fear of falling resulted in lifestyle changes.

	Made Lifestyle Changes Because of Fear of Falling		
	Yes	No	Total
Fallers	a =131	b = 52	a + b = 183
Non fallers	c = 14	d = 36	c + d = 50
Total	a + c = 145	b + d = 88	n = 233

Step – 1: Hypothesis

H_0 :Fall status and lifestyle change because of fear of falling are independent

H_1 : The two variables are not independent

Let $\alpha = 0.05$

Step – 2: Test statistic

$$\chi^2 = \frac{n(ad - bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$\chi^2 = \frac{233((131 * 36) - (52 * 14))^2}{145 * 88 * 183 * 50}$$

$$\chi^2 = 31.7191$$

Step – 3: Chi – Square tabulated

Degree of freedom = (r - 1) * (c - 1)

= (2 - 1)* (2 - 1) = 1

$\alpha = 0.05$

$\chi^2_{tab} = 3.841$

Step – 4: Decision rule

Reject H_0 if the computed value of χ^2 is equal to or greater than 3.841

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528

Step – 5: Statistical decision

We reject H_0 since 31.7391 > 3.841.

Step – 6: Conclusion

We conclude that H_0 is false, and that there is a relationship between experiencing a fall and changing the lifestyle because of fear of falling.

Applications of Chi-square

- The important applications of Chi-square
  - Test of proportion
  - Test of association
  - Test of goodness of fit

Test of proportion

- It is an alternate test to find the significance of difference in two or more than two proportions.

- It is applied to find significance in the same type of data with two more advantages,

  - to compare the values of two binomial samples even if they are small.

  - to compare the frequencies of two multinomial samples.

Test of association

- Test of association is the most important application of Chi-square test in statistical methods.

- It is measures the probability of association between two discrete attributes.

Test of Goodness of fit

- Chi-square test is also applied as a test of “goodness of fit”.

- It is used to determine deviation between calculated value and tabulated value.

Fisher exact test

When to use Fisher’s exact test:

- Cell counts are smaller than 20
- A cell has an expected value 5 or less.

Example

- Let’s work through the voting by gender example. Fisher’s exact test will determine whether a statistically significant relationship exists between gender and voting.

- As with any hypothesis test, this analysis has a null and alternative hypothesis. For our example, the hypotheses are the following:

- Null (H_0): There is no association between gender and voting. They are independent.
- Alternative (H_1): A relationship between gender and voting exists in the population.

	Male	Female
Yes	**4**	**9**
No	**10**	**3**

In the table, it appears that females are more likely to vote Yes, while males are more likely to vote No.

- The cell counts are too small for the chi-square analysis.
- We will use fisher’s exact test to determine whether a statistically significant relationship exists between gender and voting.
- We will use a Fisher exact test calculator to obtain the p-value.
- When calculated p-value is below significance level (e.g., 0.05), reject the null hypothesis

- Enter the following values for each letter field in the calculator and choose two-tailed in Test type:

  A: 4

  B: 9

  C: 10

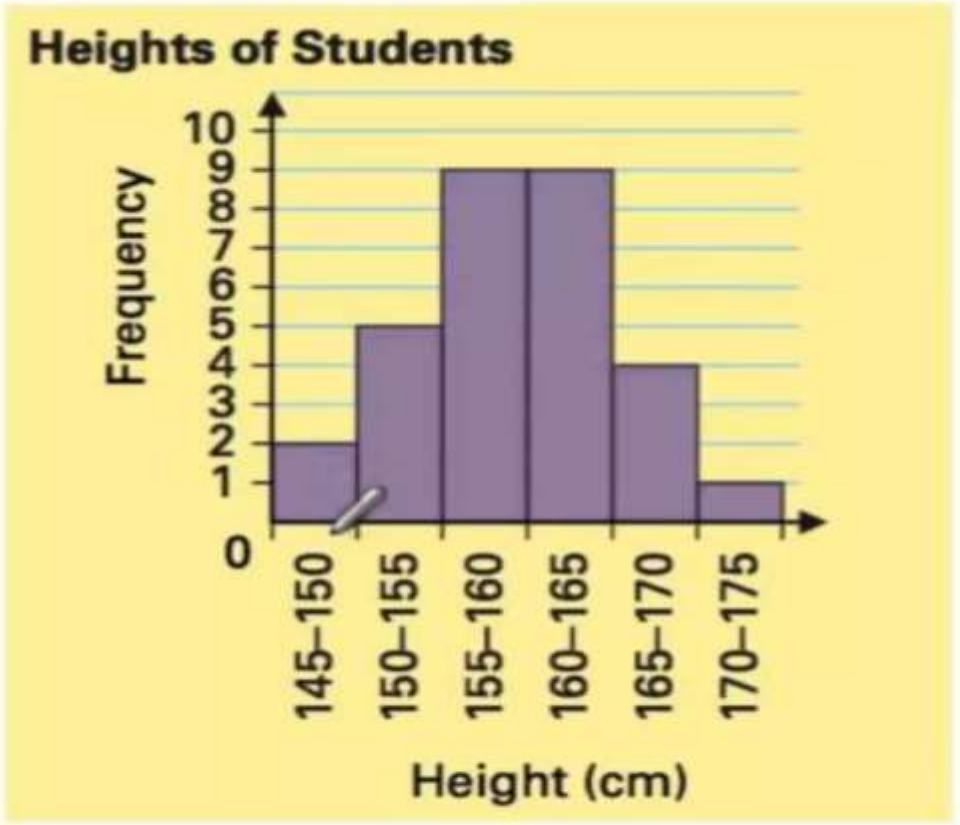
  D: 3

- The calculator calculates a p-value of 0.047 for the Fisher's exact test, which is less than our significance level of 0.05.
- Our results are statistically significant. We can reject the null and conclude that a relationship exists between gender and voting choice.

Graph reading

Histogram

Height (cm)	Frequency
145–150	2
150–155	5
155–160	9
160–165	9
165–170	4
170–175	1



- Represented by a set of rectangular bars

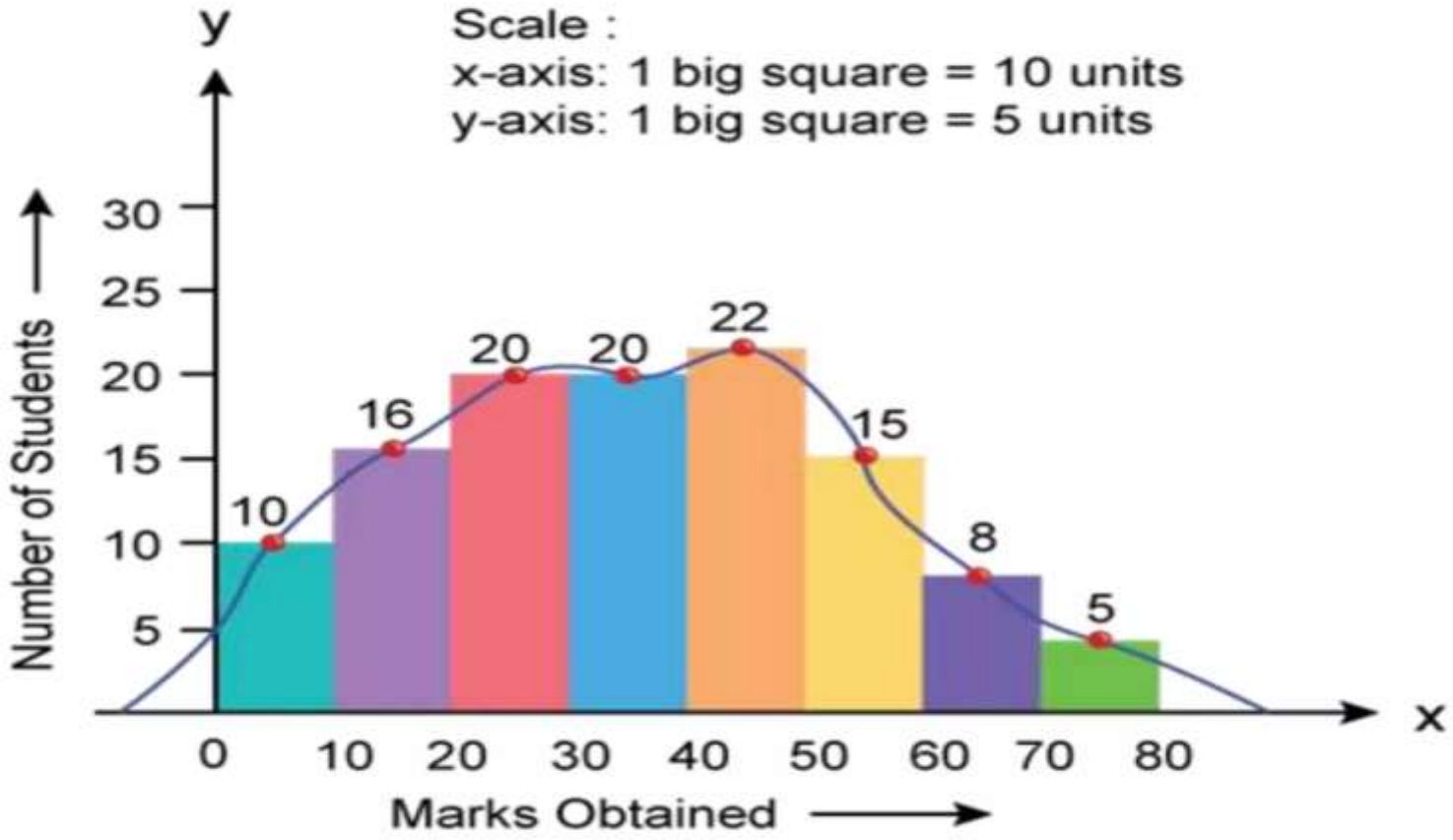
- Variables (Class) is taken along the X-axis & frequency along the Y-axis.

Note :

- The total area of the rectangles in a histogram represent total frequency.

- Mode of the distribution can be obtained from the histogram.

Frequency curve



- Frequencies are plotted against the class mid-values and then, these points are joined by a smooth curve.
- The curve so obtained is the frequency curve.

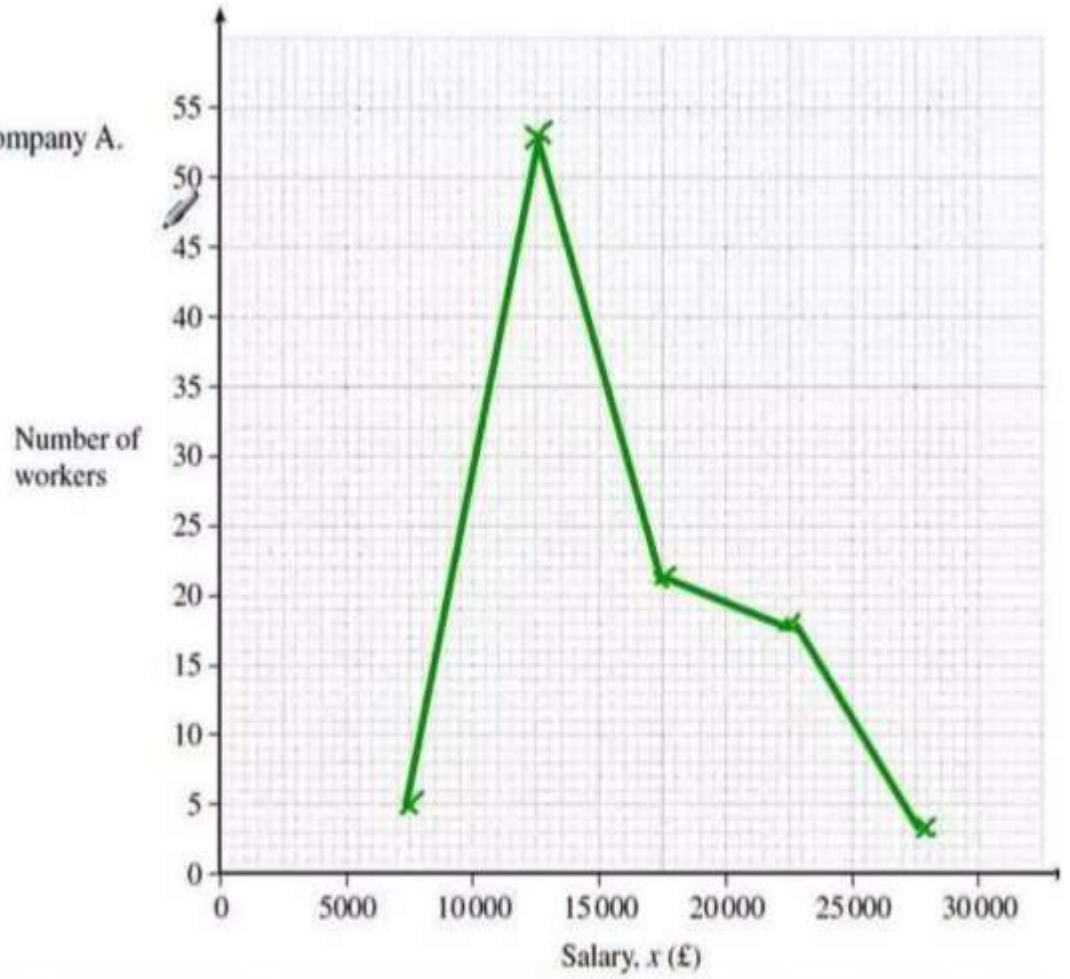
Frequency polygon

Frequency Polygon

The table shows the annual salaries of 100 office workers in Company A.

Salary, x (£)	Number of workers
$5000 < x \leqslant 10000$	5
$10000 < x \leqslant 15000$	53
$15000 < x \leqslant 20000$	21
$20000 < x \leqslant 25000$	18
$25000 < x \leqslant 30000$	3

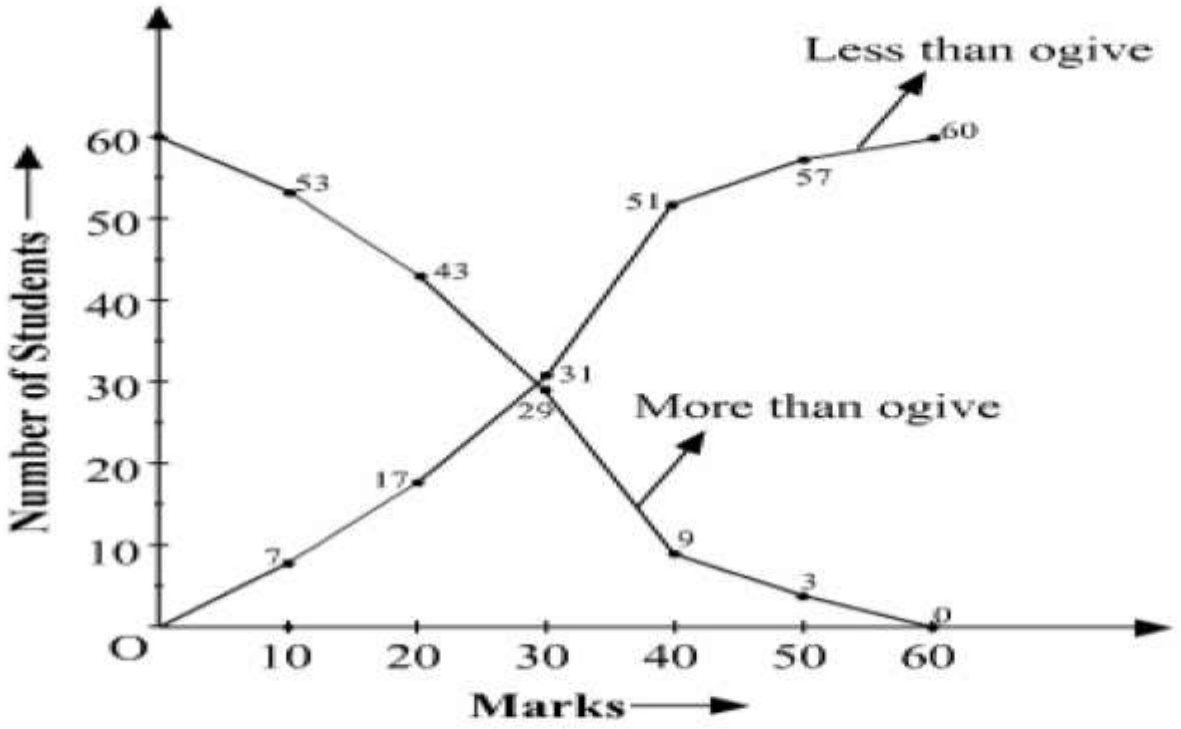
Draw a frequency polygon to represent this data.



- Class frequencies are plotted against the class mid-values and then these points are joined by Straight line.

- The figure so obtained is the frequency polygon.

Ogives (cumulative frequency Curves)

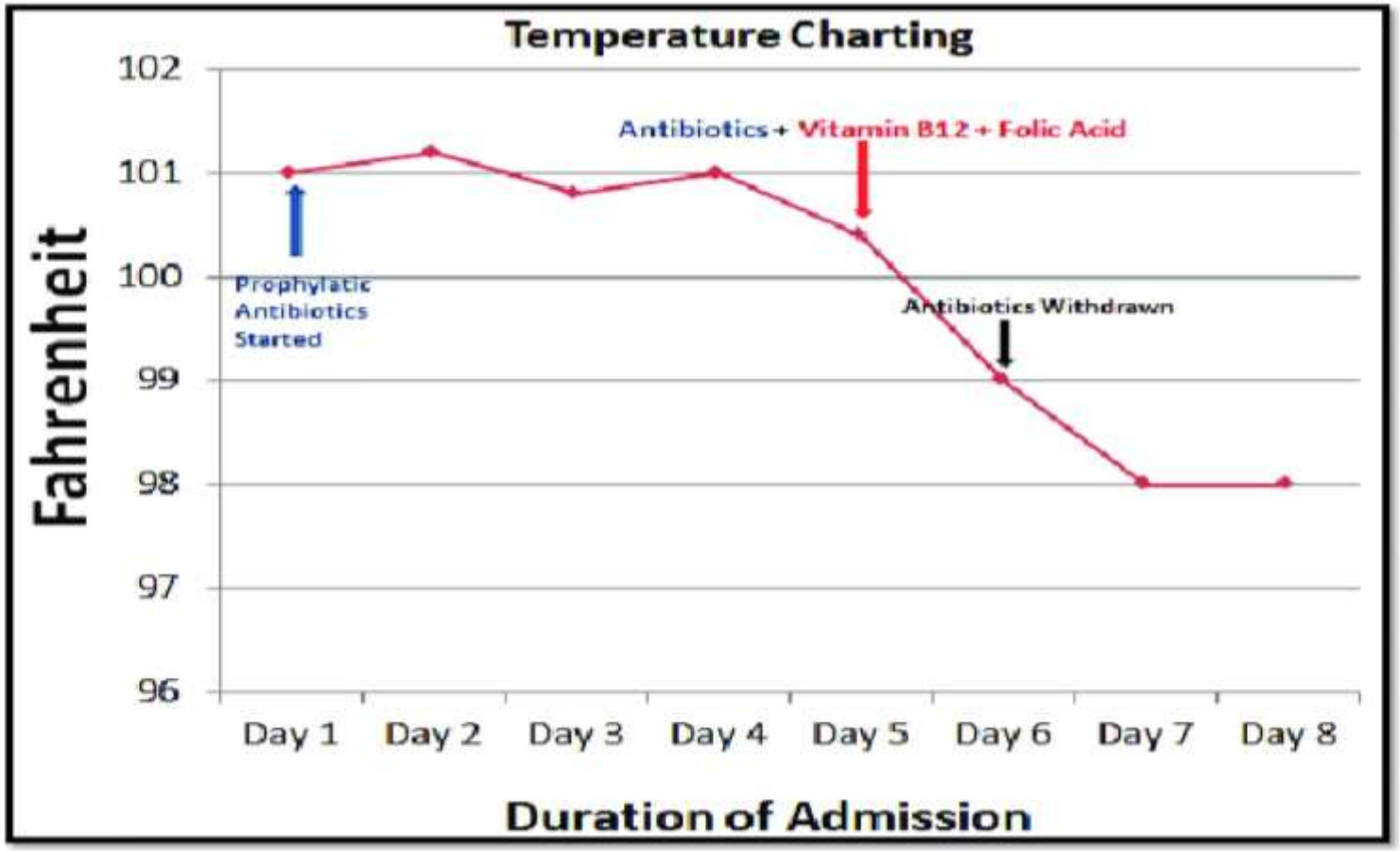


Note:

- The two ogives are drawn together with common axis.

- The points of intersection of the two ogives gives the Median point of the distribution.

Line graph: (time series graph)

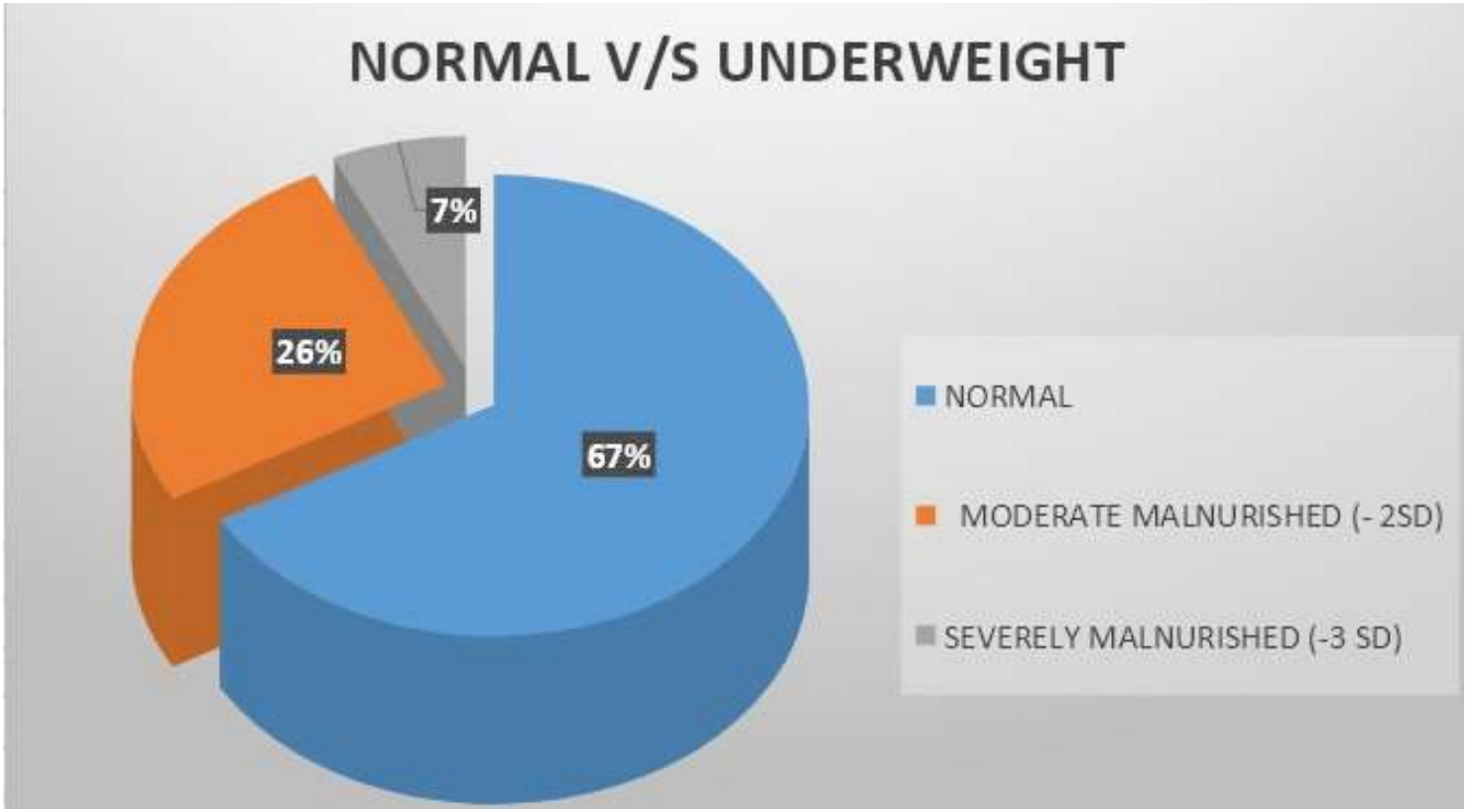


- Line graphs are used to display the comparison between two variables which are plotted on the X-axis and Y-axis.

- Line graphs displays a change in direction.

- It shows trend of an event occurring over a period of time to know whether it is increased or decreased.

Pie diagram (Sector diagram)



- Presenting discrete data of qualitative characteristics such as blood groups, RH factor, Age group, causes of mortality, etc.
- The frequencies of the groups are shown in a circle.
- Size of each angle is calculating by multiplying the frequency/total frequency by 360°
- Size of each angle (degree measure) = $\frac{\text{frequency}}{\text{total frequency}} \times 360°$

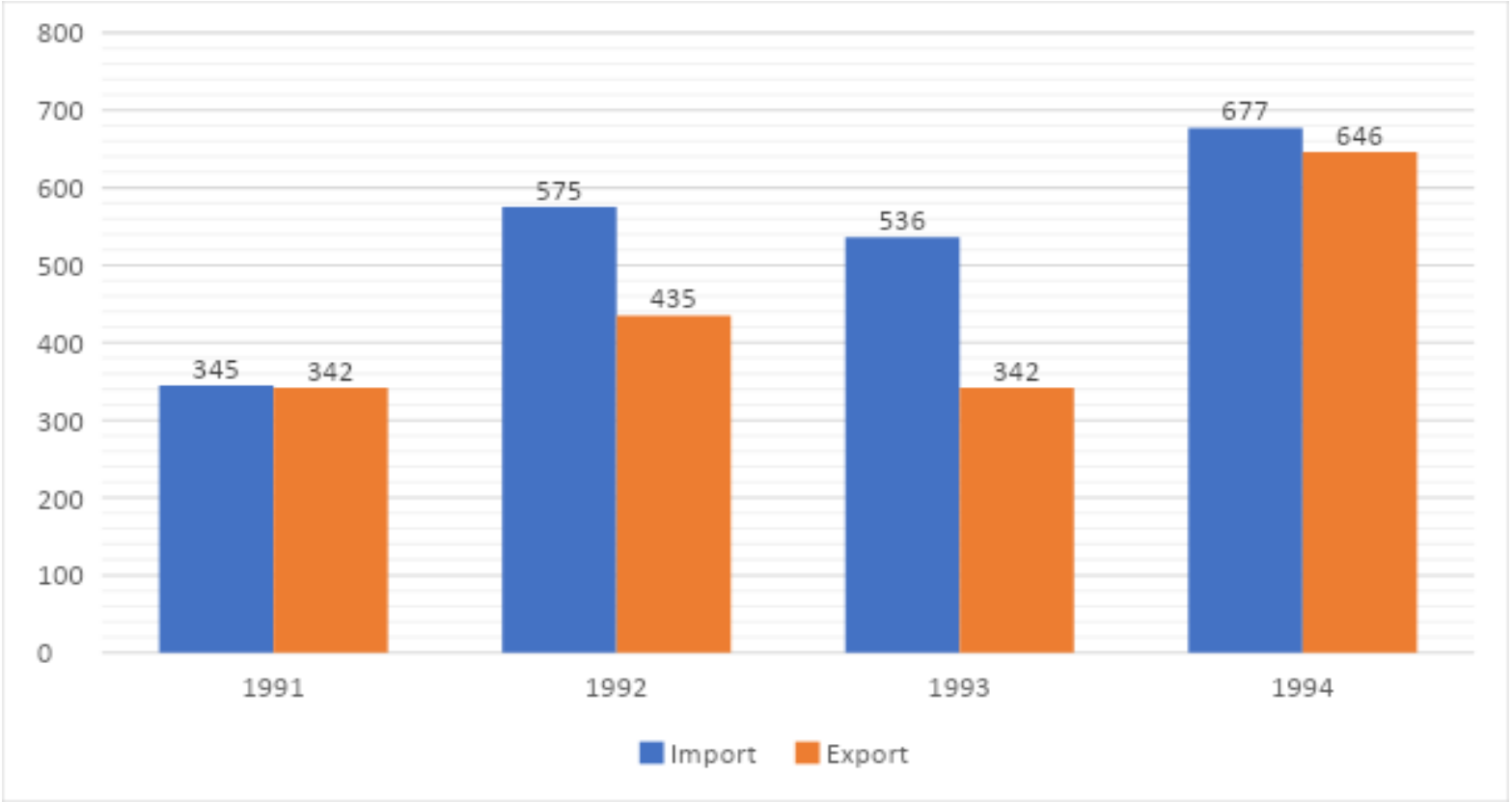
Bar diagram/ Barchart

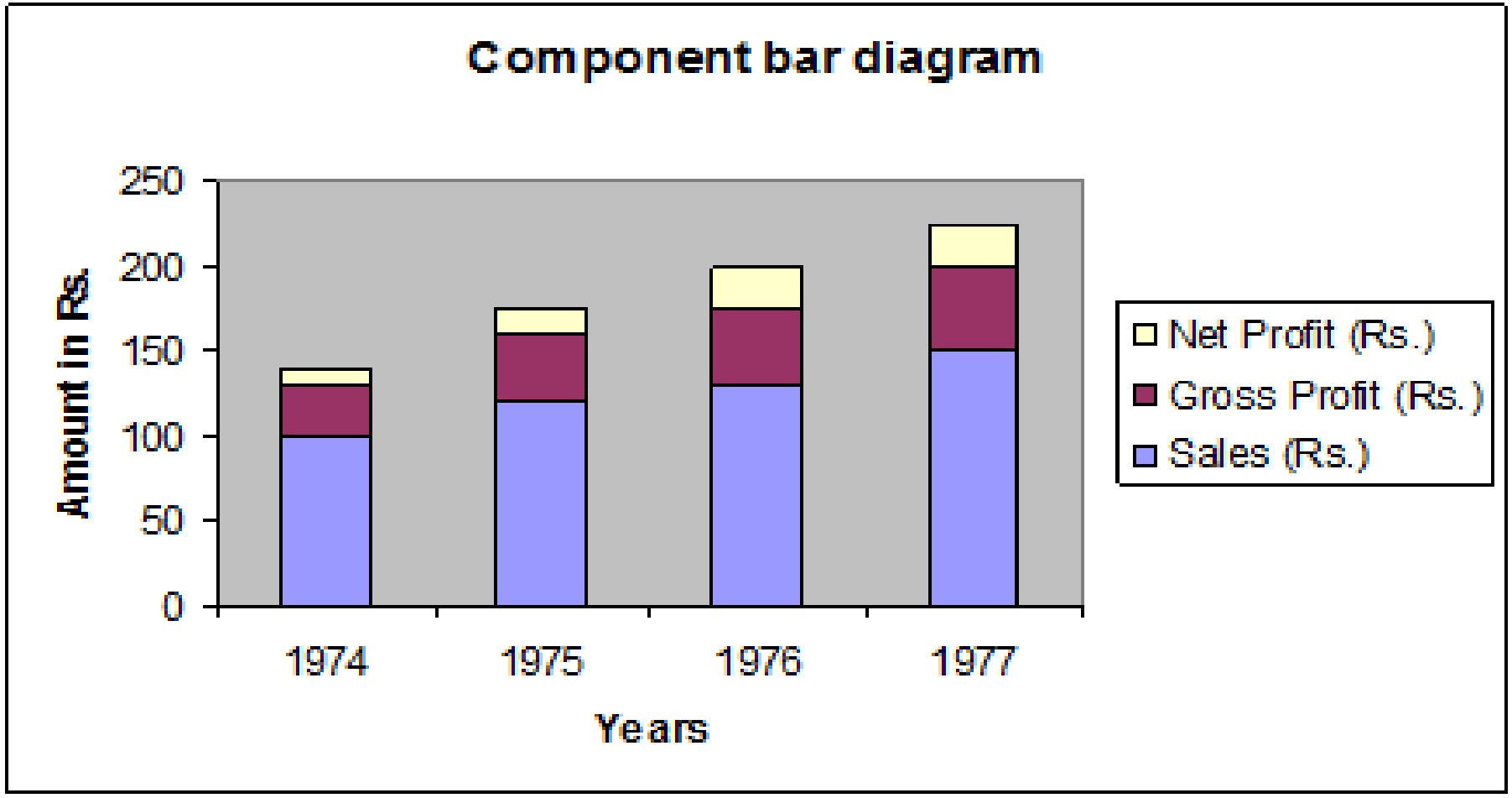
- Bar diagram consists of a series of rectangular bars of equal width.
- The bars stand on common base line with equal gap between one bar and another.
- The bars may be either horizontal or vertical.

Note:

- Space between consecutive bars are equal
- The bars are of equal width.

Multiple or compound bar diagram

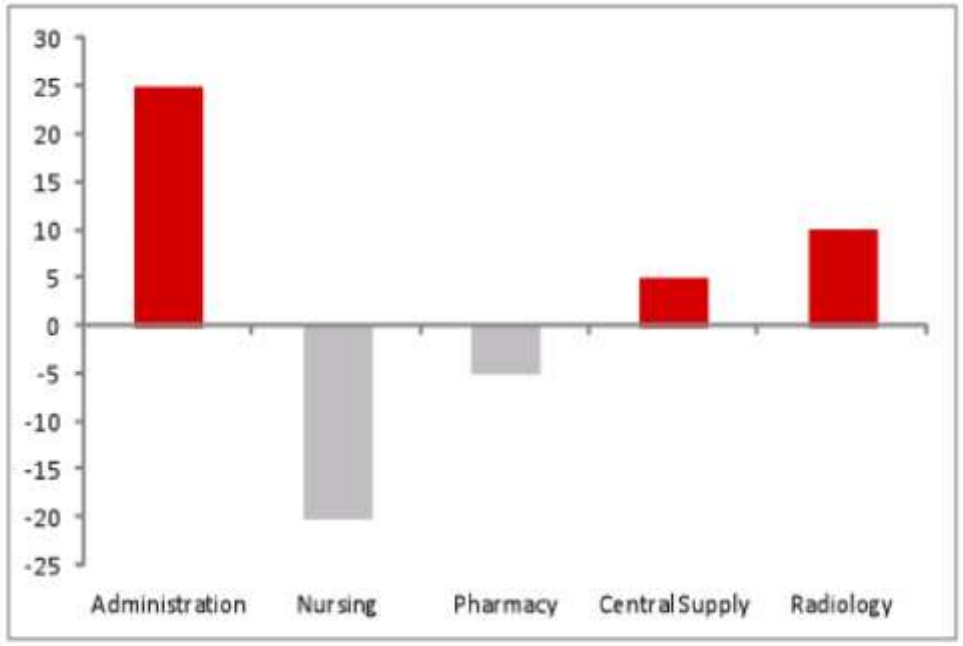



Component bar diagram
Amount in Rs.
250
200
150
100
50
0
1974
1975
1976
1977
Years
Net Profit (Rs.)
Gross Profit (Rs.)
Sales (Rs.)

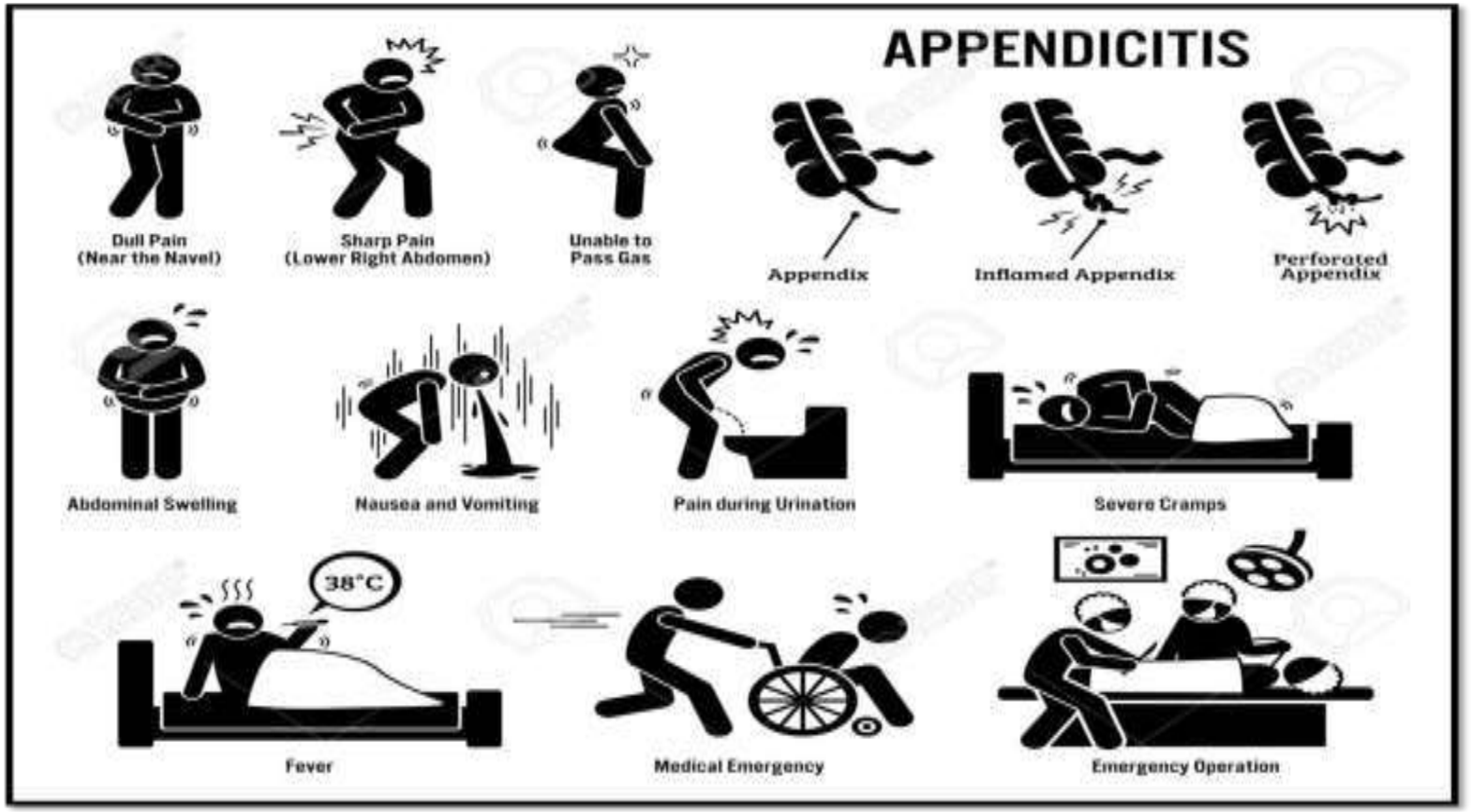
Deviation bar diagrams

Variation from budget in different areas

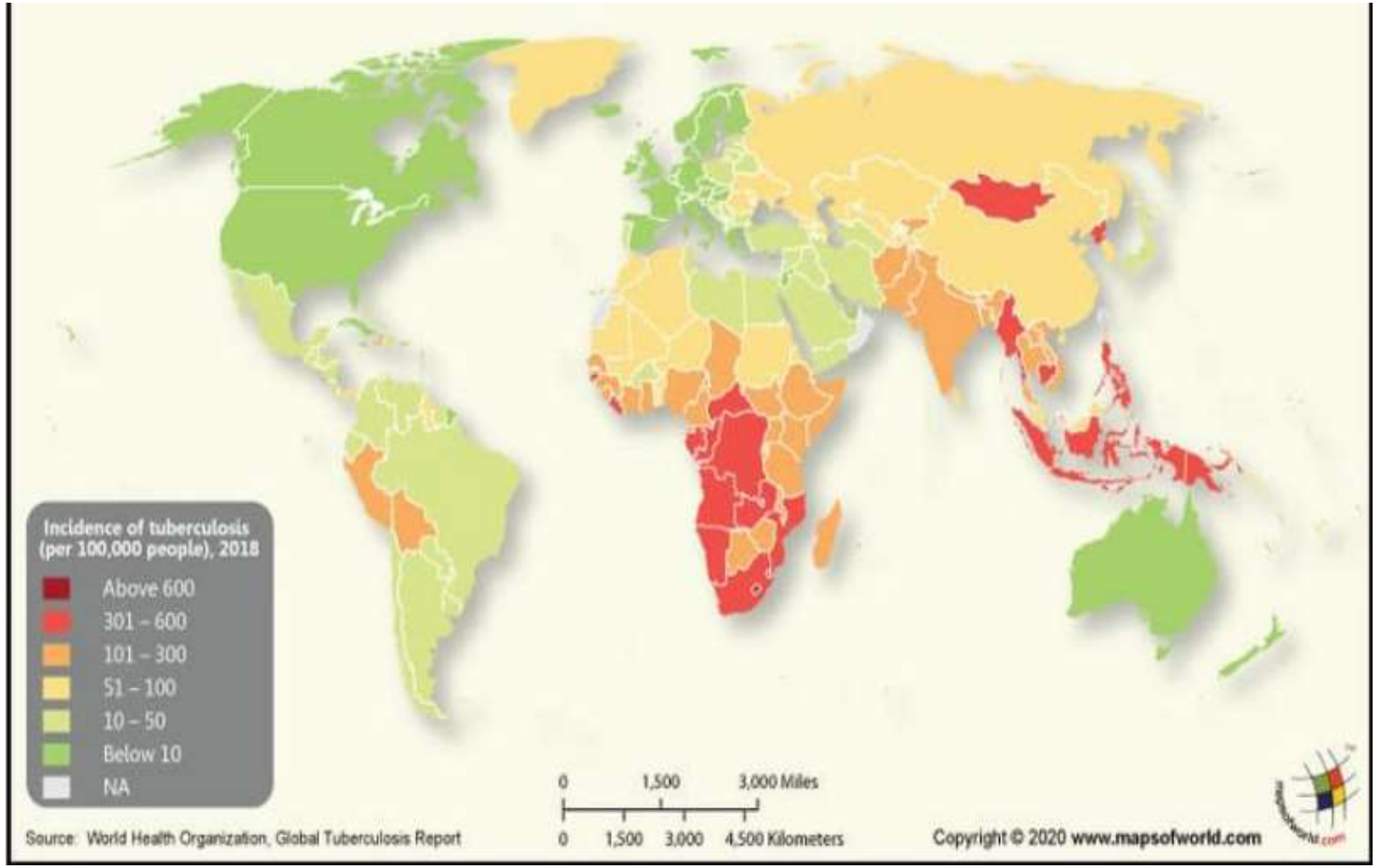
	Actual	Budget	Variance
Administration	200	175	25
Nursing	325	345	-20
Pharmacy	60	65	-5
Central Supply	45	40	5
Radiology	60	50	10



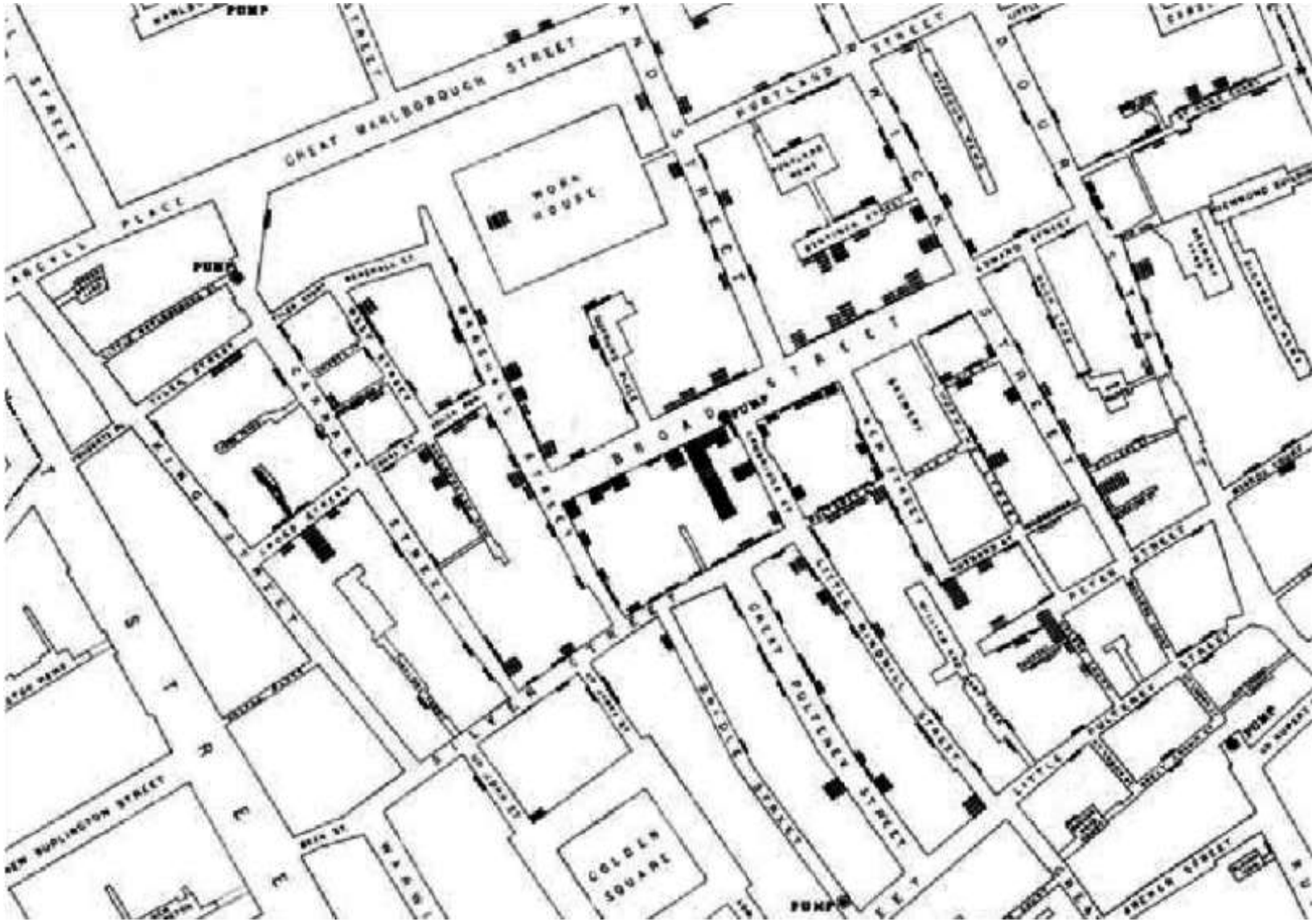
Pictogram


APPENDICITIS
Dull Pain
(Near the Navel)
Sharp Pain
(Lower Right Abdomen)
Unable to
Pass Gas
Appendix
Inflamed Appendix
Perforated
Appendix
Abdominal Swelling
Nausea and Vomiting
Pain during Urination
Severe Cramps
38°C
Fever
Medical Emergency
Emergency Operation

Rate maps


Incidence of tuberculosis
(per 100,000 people), 2018
Above 600
301 – 600
101 – 300
51 – 100
10 – 50
Below 10
NA
0 1,500 3,000 Miles
0 1,500 3,000 4,500 Kilometers
Source: World Health Organization, Global Tuberculosis Report
Copyright © 2020 www.mapsofworld.com

Map diagram


GREAT MARLBOROUGH STREET
POLAND STREET
BROAD STREET
GOLDEN SQUARE
PUMP

Stem and leaf plot

Only place the last digit of each number in the 'leaf'

Arrange the numbers from smallest to largest

101, 131, 114, 102, 125, 101, 115, 103, 120, 122

stem	leaf
10	1 1 2 3
11	4 5
12	0 2 5
13	1

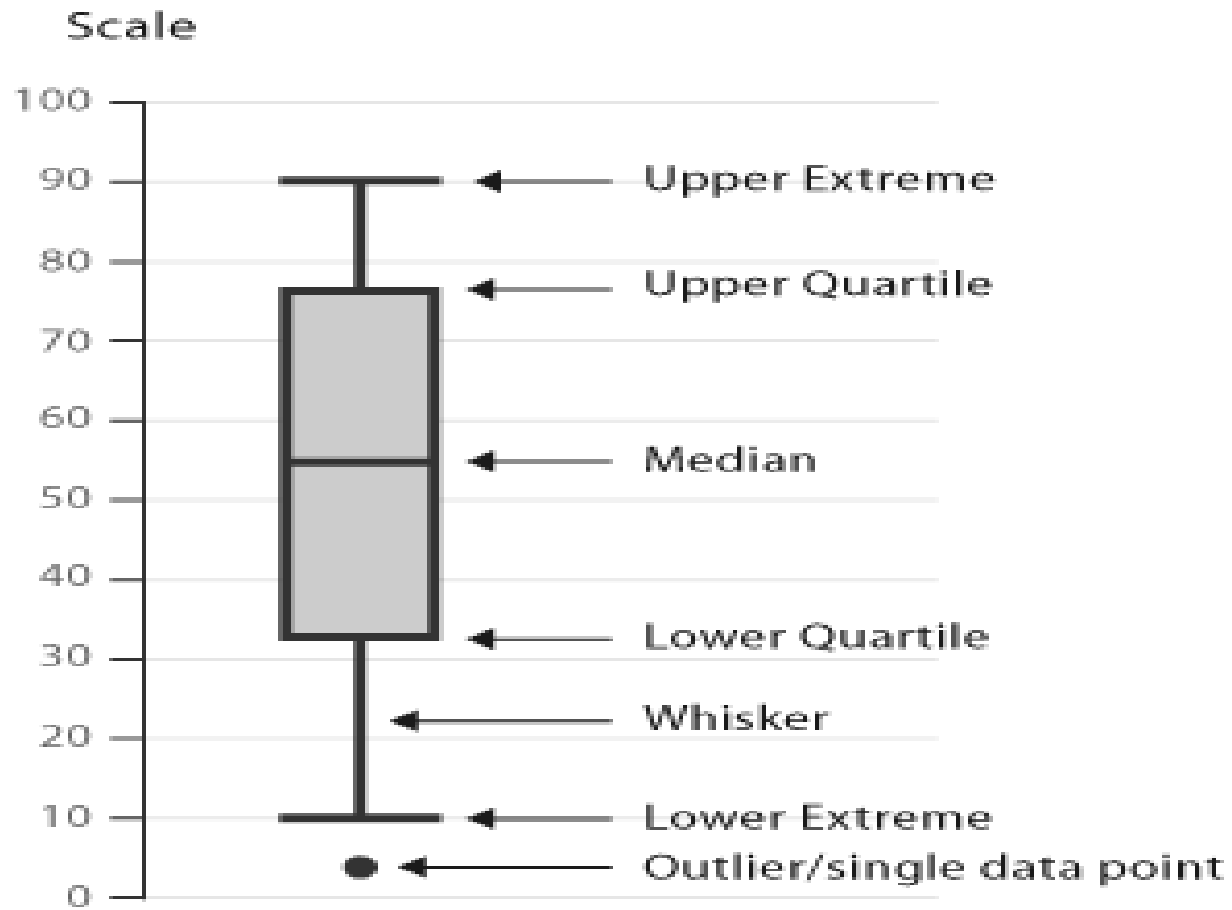
103 is represented with a '3' in the '10' stem

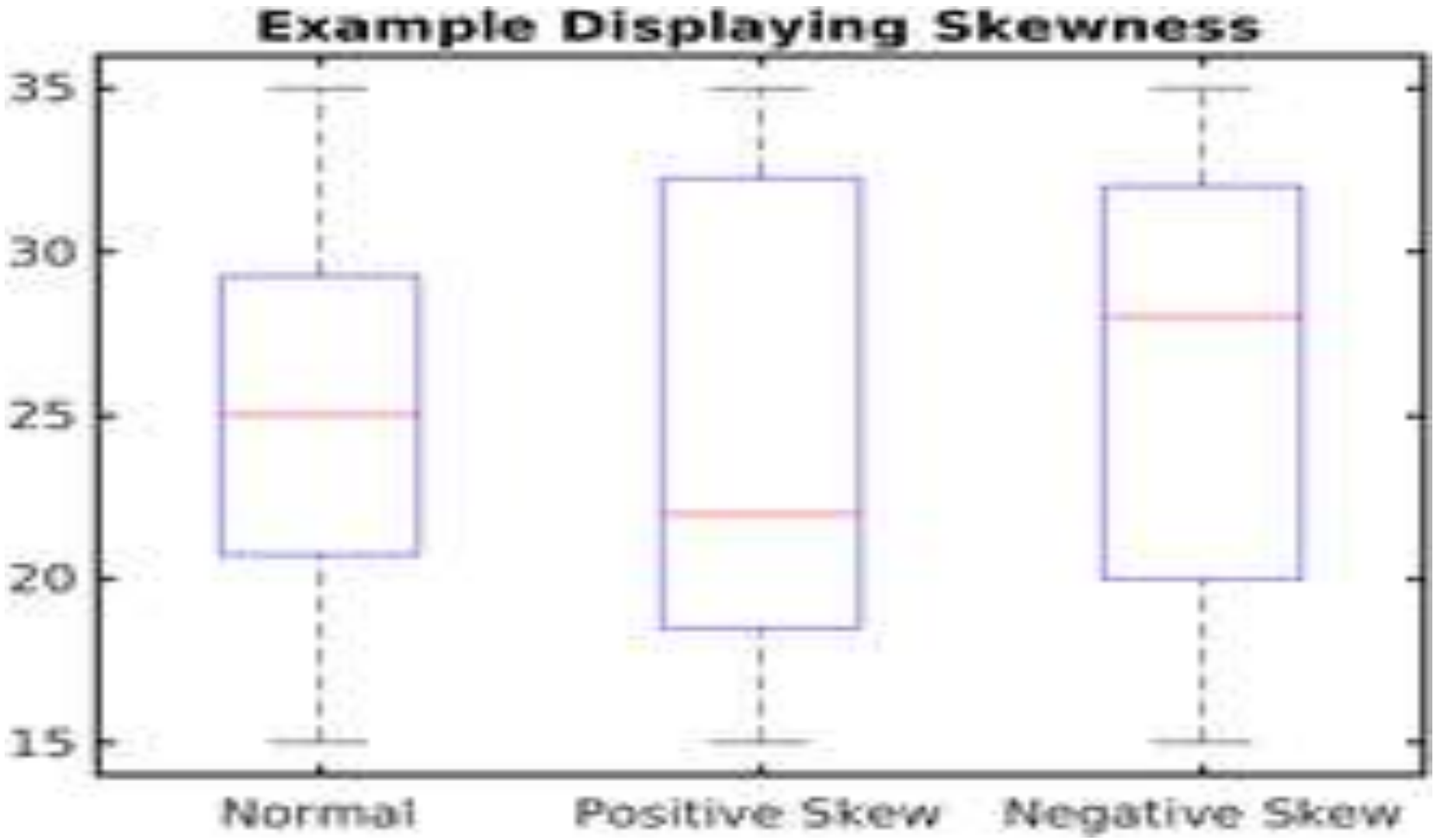
Key:

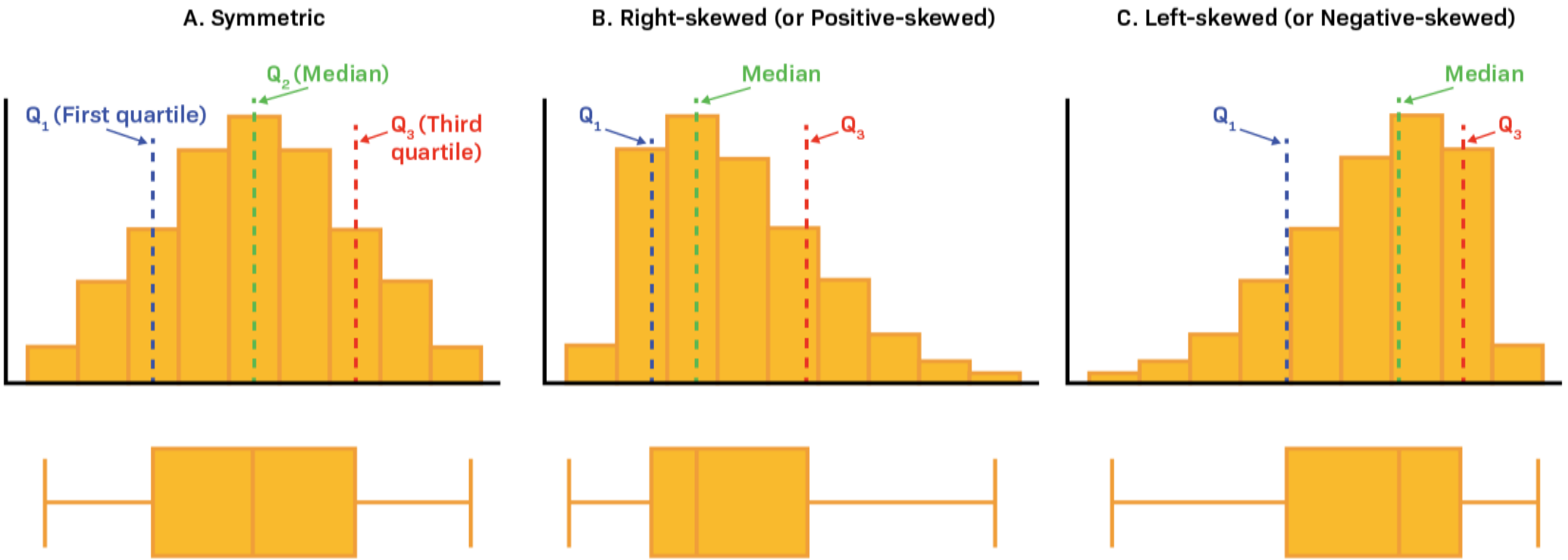
10 | 3 = 103

Place all other digits of the number in the 'stem'

Box and whisker plot



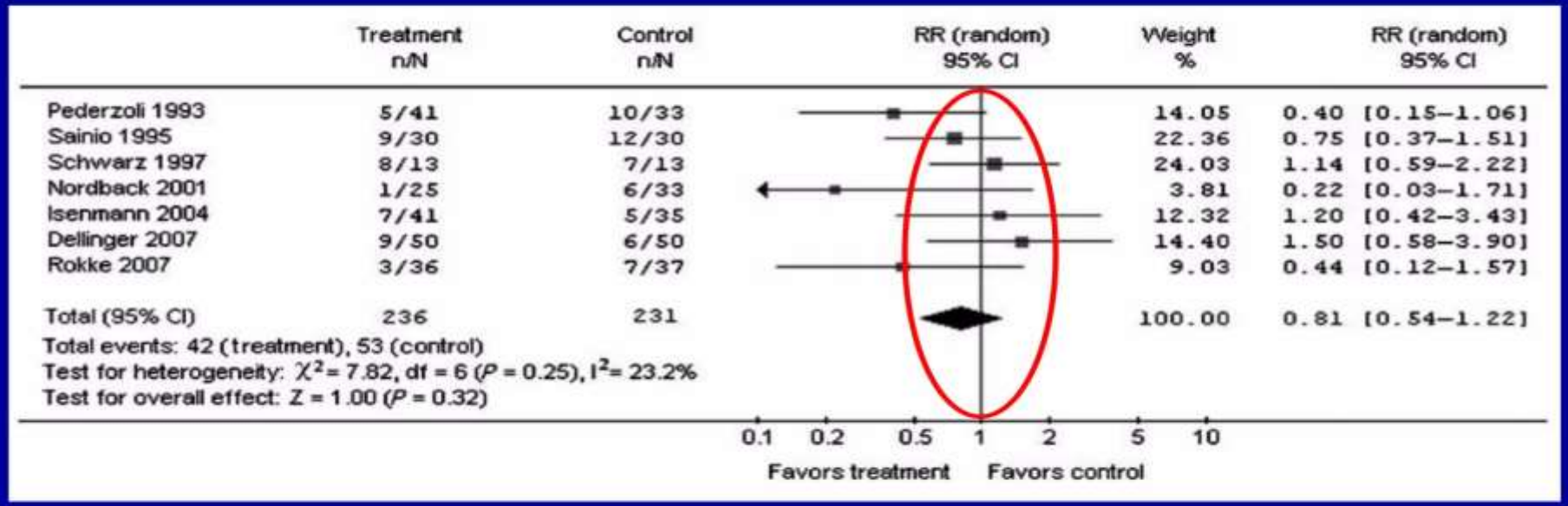

Example Displaying Skewness
35
30
25
20
15
Normal
Positive Skew
Negative Skew


A. Symmetric
Q2 (Median)
Q1 (First quartile)
Q3 (Third quartile)
B. Right-skewed (or Positive-skewed)
Median
Q1
Q3
C. Left-skewed (or Negative-skewed)
Median
Q1
Q3

Forest plot

Antibiotic prophylaxis & infected necrosis

Vertical line or line of no effect



Treatment & control groups have the same effect

Bai Y et al. Am J Gastroenterol 2008 ; 103 : 104 – 110.

Diamond in meta-analysis

- **Diamond on left of the line of no effect**

  Less episodes of outcome in treatment group

- **Diamond on right of the line of no effect**

  More episodes of outcome in treatment group

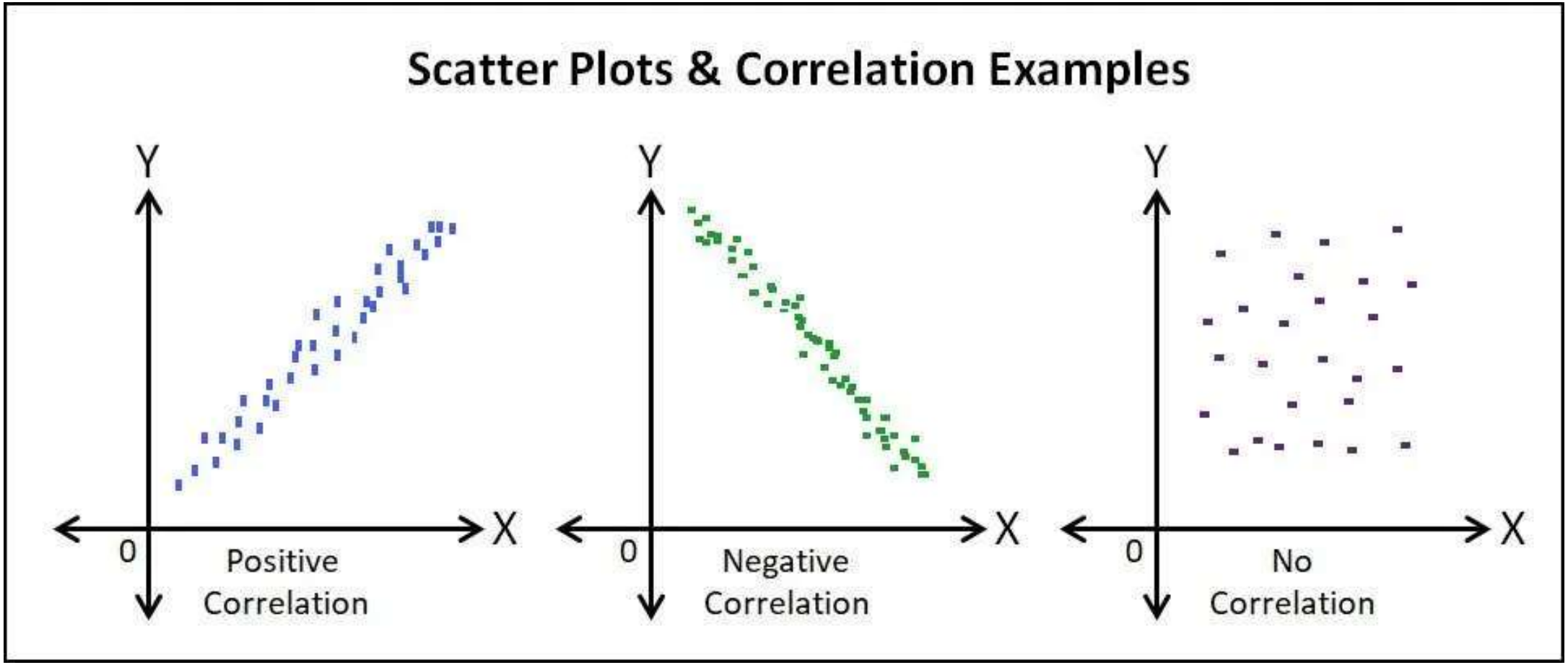
- **Diamond touches the line of no effect**

  No statistically significant difference between groups

- **Diamond does not touch the line of no effect**

  Difference between two groups statistically significant

Scatter plot



Graph for quantitative data

- Histogram
- Frequency polygon, curve
- Scatter plot
- Ogive graph
- Line graph
- Box and whisker plot
- Stem and leaf plot
- Forest plot graph

Graph for qualitative data

- Bar graph
- Pie chart
- Pictogram
- Spot maps
- Rate maps

References

- Daniel WW, Chad Lee Cross. Biostatistics : a foundation for analysis in the health sciences. Hoboken, Nj: Wiley; 2019.
- Khanal arun bhadra, jaypee brothers (jaypee digital. Mahajan's methods in biostatistics for medical students and research workers. Jaypee brothers medical publisher (P) ltd; 8th edition
- K park. Park's textbook of preventive and social medicine. India: bhanot publishers; 23rd edition
- Postgraduate Pharmacology, Sougata Sarkar, V. Srivastava, M. Mohanty

“Thank You”