1) collection of data
2) Analysis of data to get an acceptable conclusion.

Why we collect data - Data is collected to perform a statistical experiment.

How we collect data - Data is collected through sample.

Data is collected from the sample and conclusion based on the observation of the sample drawn for the population.
Acceptable at a certain level of confidence
at 90% level of confidence.

__Random Sampling__ :-
Sample is collected without Biasedness.

If every sample has equal chance (equal probability) to be included then the sample is random.

Stratified sampling :-

Population is devided into diffe... strata.

sample should be the represantative of the whole population.

One data is collected

The data has some central values → Measure of central tendency

central values → Mean, → Median, → mode

Mean :-

If $x_1, x_2 --- x_n$ are 'n' observation the mean.

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + -- + x_n)$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i$$

Mean is affected by extreme values

45, 50, 52, 55, 48, 300

Trimmed mean :-

10% trimmed mean then 10% of higher values and 10% of lower values are deleted and the mean of rest are taken.

4th Jan 16

Median :—

Median is the middle most value when the data is arranged in order of magnitude.

If the arranged value are

$x_1, x_2, \ldots x_n$

Then median = Mean of two middle value in 'n' is even.

→ If the no of observation is even then there are two middle values.

" " " is odd then there is one middle value.

$$\text{Median} = \frac{x_{n/2} + x_{n/2+1}}{2}$$

Median = $x_{\frac{n+1}{2}}$ if n is odd.

Variability :— (Measures of dispersion)

Range

variance and standard deviation

Range = difference between the highest and lowest

0, 50, 52, 48, 55, 45, 300

Variance:—

Variance of a population

$x_1, x_2, \ldots, x_n$

$$\boxed{Var = \frac{1}{n} \sum (x_i - \bar{x})^2}$$

Variance of the sample

$x_1, x_2 \ldots x_n$ are n observation n the sample.

then $$\boxed{Var = \frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

$$\boxed{\bar{x} = \frac{1}{n} \sum x_i}$$

Standard deviation is the +ve square root of variance.

Ex.

1, 2, 4, 8, 11

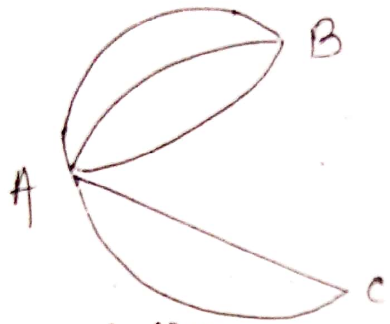
Fundamental Rule of Counting:—

Product Rule:—

If one Operation can be performed in m map and after the completion of the first operation, a second operation can be performed in n ways. Then total operation can be performed in $m \times n$ ways.

Sum Rule :—

If one operation can be performed in m ways and another operation can be performed in n ways then either of the operation can be performed in m+n ways.



Permutation

No of arrangement of n different things.

1) order is maintained
2) Repetition is not allowed.

^nP_r or $P(n,r)$

$$= \frac{n!}{(n-r)!}$$

$n! = 1 * 2 \times 3 \times \cdots \times n$

$0! = 1$

$$\boxed{^nP_n = \frac{n!}{(n-n)!} = n!}$$

Combination.

No of arrangement of n different things taking r at a time.

1) order is not maintained
2) Repetition is not allowed.

$$\boxed{^nC_r \text{ or } C(n,r) = \frac{n!}{r!(n-r)!}}$$

Lecture 1
Chapter 1: Introduction to Statistics and Data Analysis

1.3 Measures of Location: The Sample Mean and Median

The aim of this lecture is to explain the following concepts :

- Measures of Location.
- The Sample Mean and Median.
- The Sample Range and Sample Standard Deviation.
- Histogram.

Definition 1 *Suppose that the observations in a sample are $x_1, x_2, ..., x_n$. The sample mean, denoted by $\bar{x}$, is*

$$\bar{x} = \sum_{i=1}^{n} \frac{x_1 + x_2 + ... + x_n}{n}$$

Definition 2 *Given that the observations in a sample are $x_1, x_2, ..., x_n$, arranged in increasing order of magnitude, the sample median is*

$$\bar{x} = \begin{cases} x_{(n+1)/2}, & if\ n\ is\ odd \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & if\ n\ is\ even. \end{cases}$$

Definition 3 *The sample variance, denoted by s^2, is given by*

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

The sample standard deviation, denoted by s, is the positive square root of s^2, that is,

$$s = \sqrt{s^2}$$

Example 1 *An engineer is interested in testing the "bias" in a pH meter. Data are collected on the meter by measuring the pH of a neutral substance (pH = 7.0). A sample of size 10 is taken, with results given by*
7.07 7.00 7.10 6.97 7.00 7.03 7.01 7.01 6.98 7.08.
Find sample variance and standard deviation.

Solution : *The sample mean $\bar{x}$ is given by*

$$\bar{x} = \frac{7.07 + 7.00 + 7.10 + \ + 7.08}{10} = 7.0250.$$

The sample variance s^2 is given by

$$s^2 = \frac{1}{9}[(7.07-7.025)^2+(7.00-7.025)^2+(7.10-7.025)^2+....+(7.08-7.025)^2] = 0.001939.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.001939} = 0.044.$$

So the sample standard deviation is 0.0440 with n-1 = 9 degrees of freedom.

Exercises:

1. The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.
3.4 2.5 4.8 2.9 3.6
2.8 3.3 5.6 3.7 2.8
4.4 4.0 5.2 3.0 4.8
Assume that the measurements are a simple random sample.
(a) What is the sample size for the above sample?
(b) Calculate the sample mean for these data.
(c) Calculate the sample median.
(d) Compute the 20 trimmed mean for the above data set.
Solution :

(a) sample size = 15.

(b) $\bar{x} = \frac{1}{15}(3.4 + 2.5 + 4.8 + \ + 4.8) = 3.787$

(c) Sample median is the 8th value, after the data is sorted from smallest to largest = 3.6.

(d) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:
2.9 3.0 3.3 3.4 3.6
3.7 4.0 4.4 4.8.
So. the trimmed mean is

$$\bar{x}_{tr20} = \frac{1}{9}(2.9 + 3.0 + \;\; + 4.8) = 3.678.$$

2. According to the journal Chemical Engineering, an important property of a fiber is its water absorbency. A random sample of 20 pieces of cotton fiber was taken and the absorbency on each piece was measured. The following are the absorbency values:
18.71 21.41 20.72 21.81 19.29 22.43 20.17
23.71 19.44 20.50 18.92 20.33 23.00 22.85
19.25 21.77 22.11 19.77 18.04 21.12
(a) Calculate the sample mean and median for the above sample values.
(b) Compute the 10% trimmed mean.
Solution :
Given sample size = 15.
(a) Mean=20.768 and Median=20.610.
(b) $\bar{x}_{tr10} = 20.743$.

7. Consider the drying time data for Exercise 1.1 on page 13. Compute the sample variance and sample standard deviation.
Solution : The sample variance s^2 is given by

$$s^2 = \frac{1}{15-1}[(3.4-3.787)^2+(2.5-3.787)^2+(4.8-3.787)^2+.....+(4.8-3.787)^2] = 0.94284.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.9428} = 0.971.$$

8. Compute the sample variance and standard deviation for the water absorbency data of Exercise 1.2 on page 13.
Solution : The sample variance s^2 is given by

$$s^2 = \frac{1}{20-1}[(18.71-20.768)^2+(21.41-20.768)^2+....+(21.12-20.768)^2] = 0.94284.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{2.5345} = 1.592.$$

Histogram:

A table listing relative frequencies is called a **relative frequency distribution**.

The information provided by a relative frequency distribution in tabular form is easier to grasp if presented **graphically**.

Using the midpoint of each interval and the corresponding relative frequency, we construct a **relative frequency histogram**.

Class Interval	Class Midpoint	Frequency, f	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

Figure 1: Relative Frequency Distribution of Battery Life

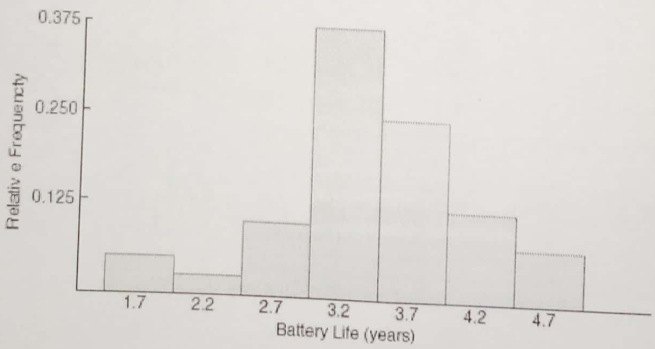


Figure 2: Relative frequency histogram