CHAPTER 1

# Classification of Data:

1. Qualitative [Inherently Discrete]:
   1. Normal: No natural order between the categories. (eg. eye color)
   2. Ordinal: If there is a natural order between the categories. (eg. socio-economic status)
2. Quantitative:
   1. Discrete: Measurements are integers.
   2. Continuous: Measurements can take up any value, usually within a range.

# Presenting Discrete Data:

1. Frequency Distribution Tables:

DATA

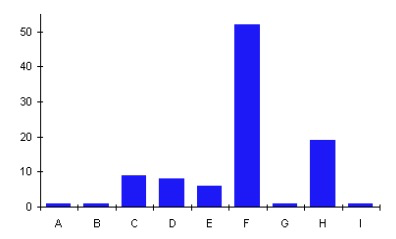
FREQUENCY

REL. FREQUENCY (Freq/Total Freq)

%age FREQUENCY (Rel. Freq \* 100%)

1. Bar Charts:

* Gaps between rectangles for discrete data.



# Presenting Continuous Data:

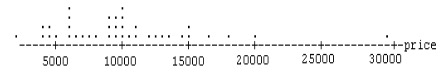
1. Frequency Distribution Tables:

* Choose a set of continuous, non-overlapping class intervals.
  + [a, b) = Inclusive of a but not b.
* Make a table similar to the one for discrete data.

1. Histograms:

* Bar charts without gaps between rectangles.
* Area of Rectangle Frequency i.e. Height Frequency if Class Widths are same.

1. Dot Plots:



# Measures of Central Location:

1. Mode
   1. Data value with the highest frequency. 1 Mode = Unimodal, 2 Modes = Bimodal.
   2. For grouped data, the mode is the midpoint of the class with the highest frequency.
2. Median
   1. Positional Middle Value
   2. For grouped data:

where = Lower Boundary of Median Class

= No. of Items needed to reach Median Value.

= Frequency of Median Class

= Class Width of Median Class

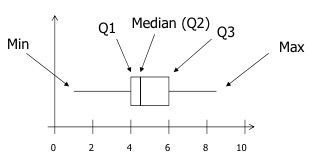
1. Mean
   1. Arithmetic average of all the data values.
   2. For grouped data:

where = Midpoint of Class

* Symmetric Unimodal Distribution 🡪 Mode = Median = Mean

# Measures of Variation

1. Range = Maximum Value – Minimum Value
2. Quartiles:
   1. = Median
   2. If n is odd:
   3. If n is even:
   4. Box Plots:



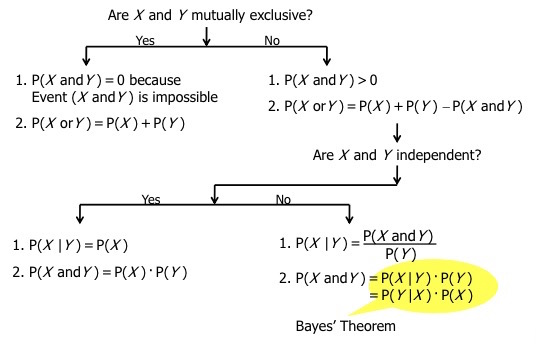
1. Standard Deviation & Variance
   1. S.D. = Average Distance of Data Values from Mean

is an unbiased estimator of .

For grouped data:

CHAPTER 2

* A sample space is the set of all the possible outcomes of an experiment.
* An event is a set of one or more outcomes from the sample space.
* Two events are **mutually exclusive** if they have no outcomes in common.
* Events are **exhaustive** if they cover all possible outcomes.
* Mutually Exclusive Events, and :
* Non-Mutually Exclusive Events:
* Independent Events:
* Dependent Events:



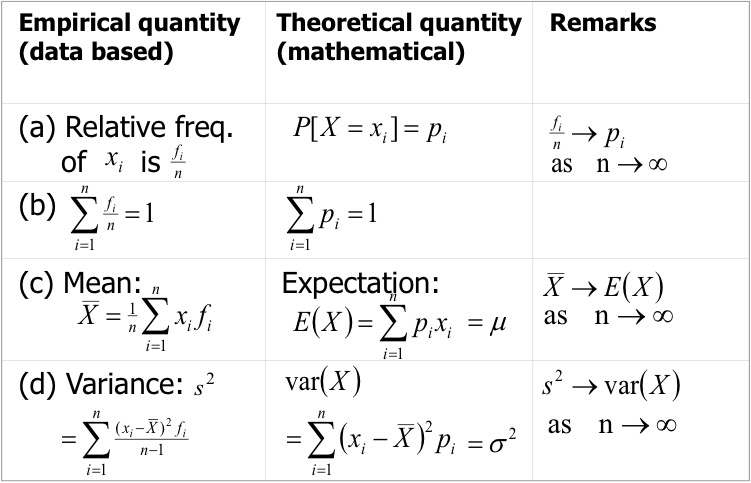
CHAPTER 3

# Random Variables

* A random variable is a variable that has a numerical value determined by the outcomes or events of an experiment.
* A random variable can be discrete or continuous.

**For an R.V :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  | … |  |
|  |  |  |  | … |  |



**For :**

**For :**

when if and are independent.

# Binomial Distribution

1. n Bernoulli Trials – where each trial has exactly two possible outcomes.
2. Probability of success/failure is same for all trials.
3. All trials are independent. Probability of any given combination of successes & failures can be obtained by multiplying the probabilities of every outcome.

**Combining 5 Trials:**

where

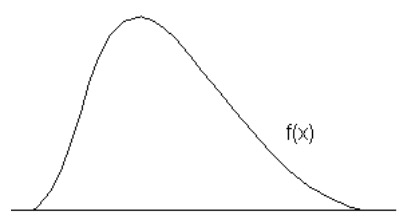
* However, there are many ways to get three Ss & two Fs in 5 trials. For example, SSSFF, SSFFS etc.
* The number of distinct arrangements is given by the binomial coefficient where *n* is the no. of trials & *r* is the no. of required successes.

**Random Variable:**

Let be the R.V equal to the total no. of successes in *n* trials. To calculate the probability of obtaining successes,

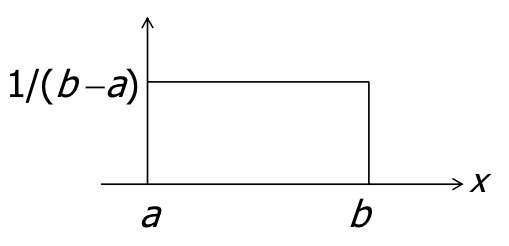
# Continuous Distribution

**Probability Density Function:**

1.  for all
2. Total Area Under Curve 1

* is known as the CDF of .
* for a continuous RV.

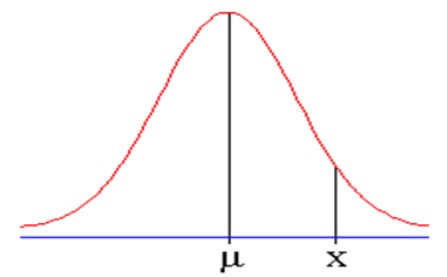
# Uniform Distribution



* can take any real value between & with uniform probability
* Thus, the PDF is for .
* For any value & between & :

**Normal Distribution**

**Standard Normal Distribution**



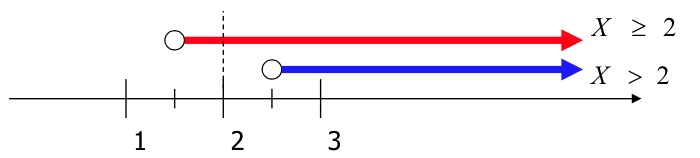
**Normal Approximation to Binomial Distribution**

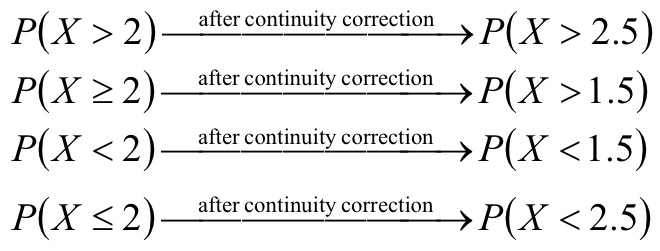
* For large *n* or *p* 0.5 or *np* > 5 & *n(1 – p)* > 5, binomial distribution approximately follows normal distribution.
* For when the conditions above are followed,
  + So,

**Continuity Correction & Accuracy**

Since it’s a discrete R.V approximated to a continuous R.V, continuity correction is needed.

For example, :





CHAPTER 4

**Population VS Sample**

* A population is a collection of all the subjects or objects of interest whereas a sample is a subset of the population used to estimate the characteristics of the population.
* A population parameter is a numerical characteristic of a population such as the mean, standard deviation, variance etc.
* A sample statistic is a numerical characteristic of a sample that can be used to estimate the corresponding population parameter.

**Sampling Distribution**

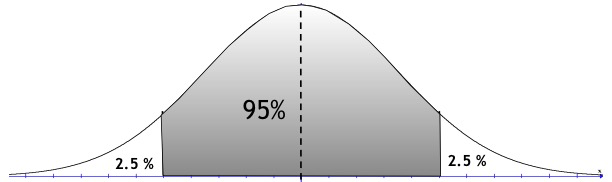
* Suppose we take many different samples and calculate the sample statistic for each (eg. Mean: , , ……), we can then draw a histogram for these sample statistics. Such sample statistics can be regarded as a R.V and the histogram is an approximation to its sampling distribution.
* Central Limit Theorem: As the number of samples, *n* , the distribution of the sample mean approaches a normal distribution with the following characteristics:

**Confidence Intervals**

* The sample mean provides a point estimate for whereas confidence intervals provide an interval estimate together with a degree of confidence that the parameter is in that interval.

**Example – 95% Confidence Interval:**

Now, we need to create a symmetric interval such that .



* Increasing the sample size increases the precision of the estimate by since the width of the interval decreases.

**Hypothesis Testing**

* To test a hypothesis that a population parameter has some specified value, two approaches can be used.

**Confidence Interval Approach**

1. Calculate the confidence interval.
2. Check if H0 (Null Hypothesis) is consistent with the interval.
3. If it is not, H0 is rejected and HA (Alternate Hypothesis) is accepted.

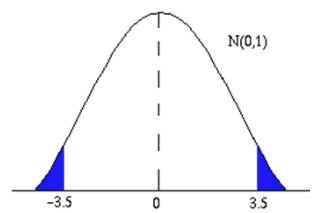
**p-Value Approach**

**Q:** A manufacturer produces bolts with a said mean length of 15 cm and 0.3 cm. A sample of 50 bolts has a mean length 14.85 cm. Does this suggest, with 0.05 significance level, that the average length of all the bolts is not 15 cm?

1. **Define H0 and HA.**
2. **Specify the significance level.**
3. **Calculate the probability of getting a sample as far away from the assumed population mean as was observed.**

Observation:

Thus, needs to be found. This is the *p-value*.



Hence,

1. **Compare the p-value with in order to accept or reject H0. If p-value , there is sufficient evidence to reject H0. Else, there is insufficient evidence to reject H0, thus, it is accepted.**

**Null Hypothesis VS Alternate Hypothesis**

* H0 (Null Hypothesis) can be either:
* In contrast, HA refers to the other possible values of the population parameter. It is true if H0 is not.
* HA (Alternate Hypothesis) can be either:

**Errors**

* Type I Error: H0 is true but it is rejected.
* Type II Error: HA is true but H0 is accepted.

**Tests Concerning Means**

* If the population variance is known and sample size () is anything:
* If the population variance is unknown and :
* If the population variance is unknown and :

i.e. We use the tr-distribution with degrees of freedom .

As , tr-distribution .

**Tests Concerning Proportions**

* Suppose we are interested in a population where each individual can be classified into one of two categories. For example: Pass/Fail, True/False, OK/Defective etc.
* In general, they are called successes and failures.
* We want to arrive at conclusions about , the proportion of successes in the population, using information from a sample.

Let a R.V be the number of successes in a random sample of ­ individuals:

If , we have and . Therefore,

If and , can be approximated to:

Thus, we can use to make inferences about the population parameter, .

**Small Samples**

* When is small, we cannot approximate the distribution of to a normal distribution. Thus, we will base our test directly on the evaluation of the binomial probabilities.

**Q:** In USA, six of the seven first IVF babies were girls. Did IVF somehow affect the sex of the babies?

Since cannot be rejected. Hence, .