Sampling and Non Sampling Errors

“Sampling error is the error that arises in a data collection process as a result of taking a sample from a population rather than using the whole population.

Sampling error is one of two reasons for the difference between an estimate of a population parameter and the true, but unknown, value of the population parameter.

The sampling error for a given sample is unknown but when the sampling is random, for some estimates (for example, sample mean, sample proportion) theoretical methods may be used to measure the extent of the variation caused by sampling error.” Sampling error is mainly cause due to the reason that sample not whole population.

Non-sampling error is the error that arises in a data collection process as a result of factors other than taking a sample.

Non-sampling errors have the potential to cause bias in polls, surveys or samples.

There are many different types of non-sampling errors and the names used to describe them are not consistent.

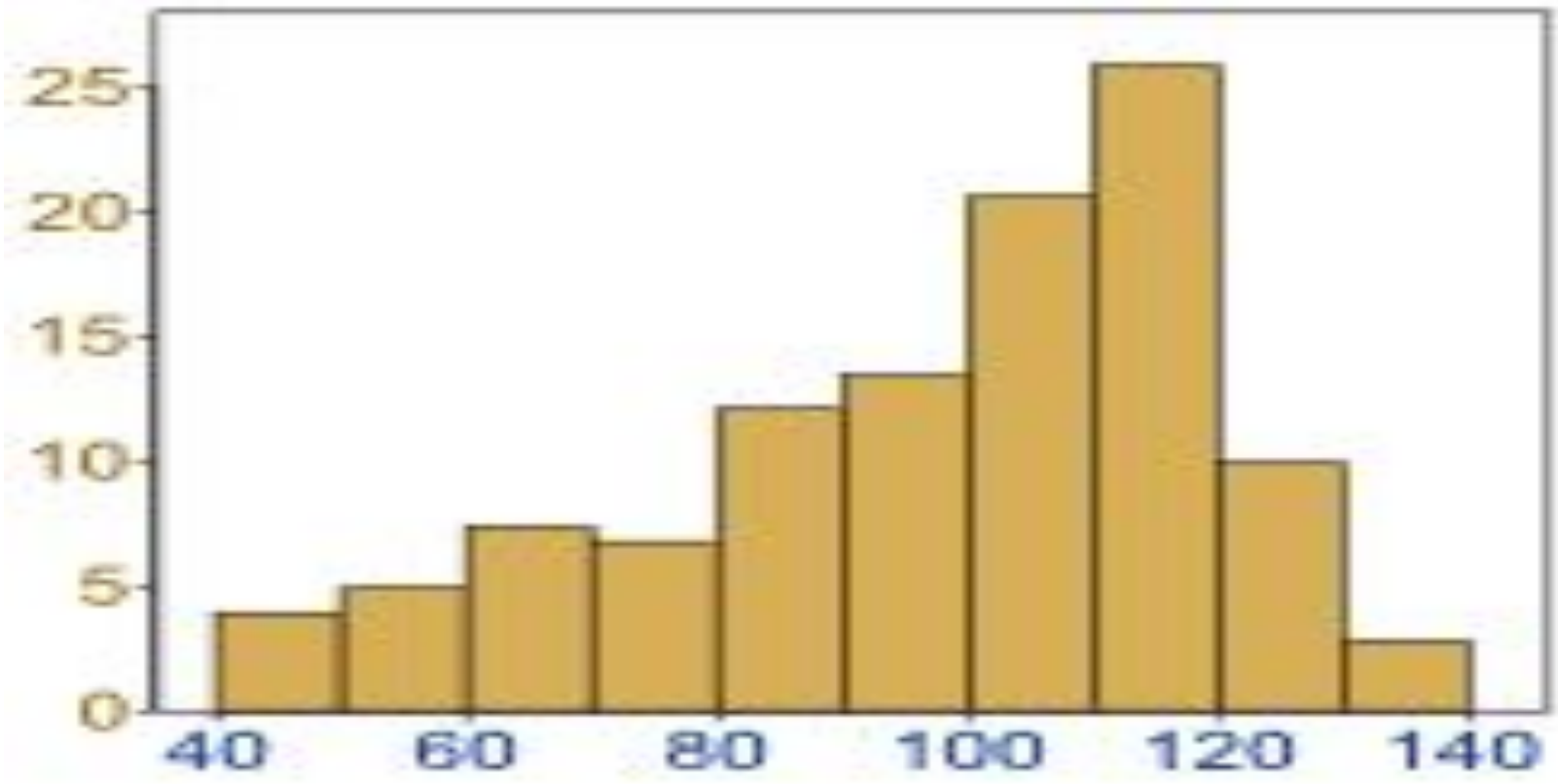
This may be due to poor sampling method, measurement errors, and behavioural effect.

ITEMS	SAMPLING ERROR	NON-SAMPLING ERROR
Meaning	Sampling error is a type of error, occurs due to the sample selected does not perfectly represents the population.	An error occurs due to sources other than sampling, while conducting survey activities is known as non sampling error.
Cause	Deviation between sample mean and population mean	Deficiency and analysis of data
Type	Random	Random or Non-random
Occurs	Only when sample is selected.	Both in sample and census.
Sample size	Possibility of error reduced with the increase in sample size.	It has nothing to do with the sample size.

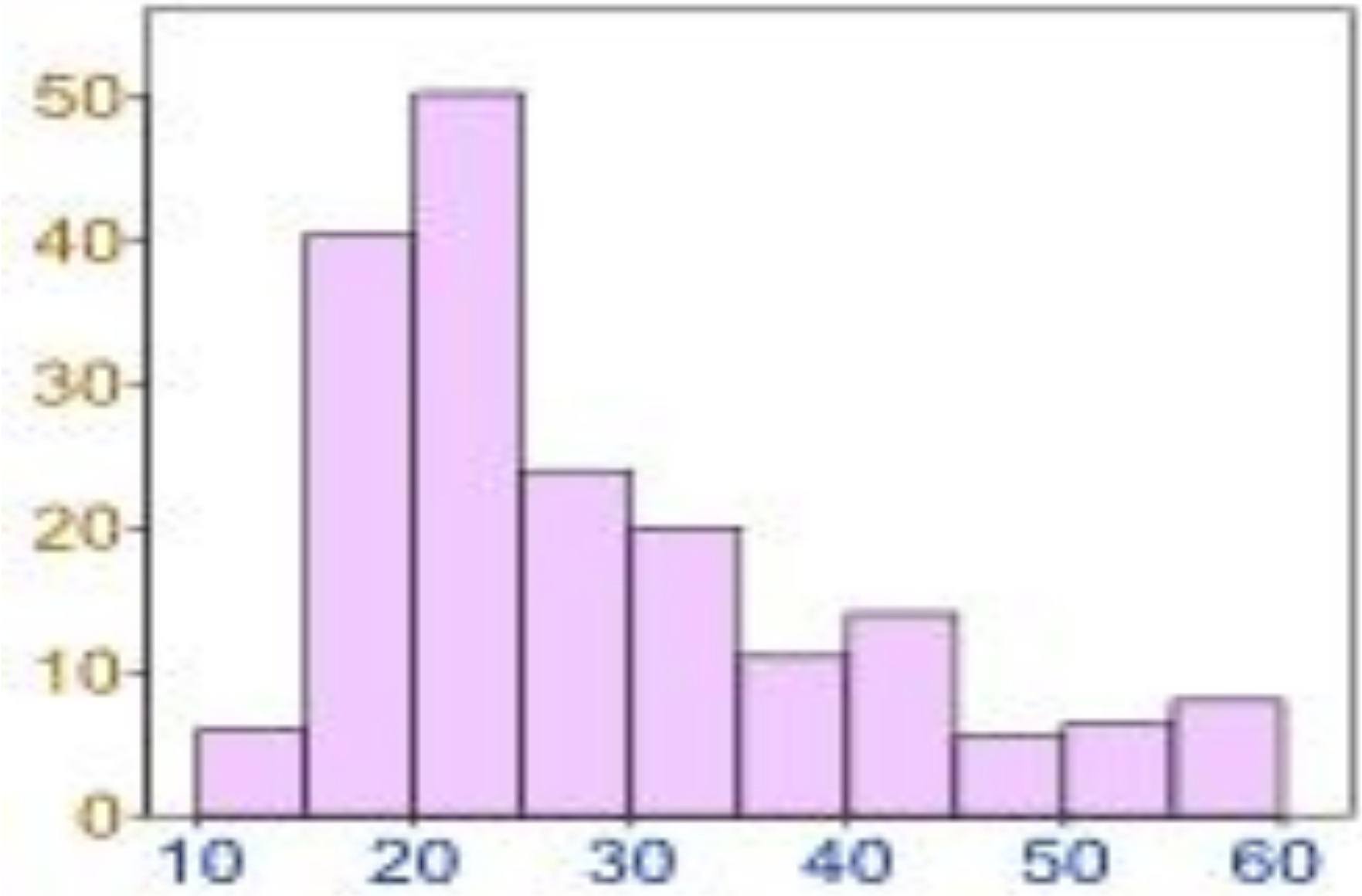
Skewness and Kurtosis

Normal distribution: Data can be "distributed" (spread out) in different ways.

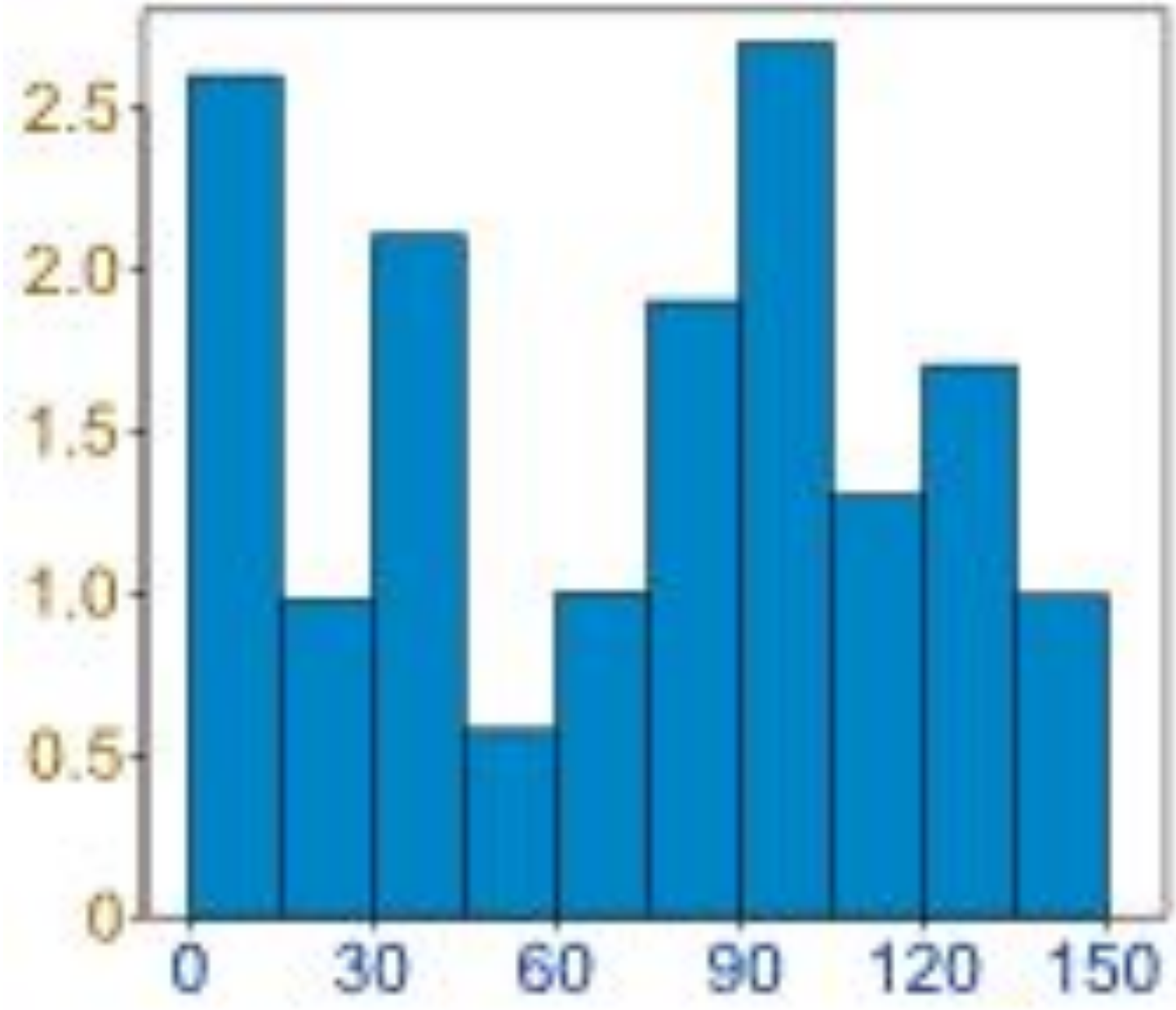
It can be spread out more on the left:



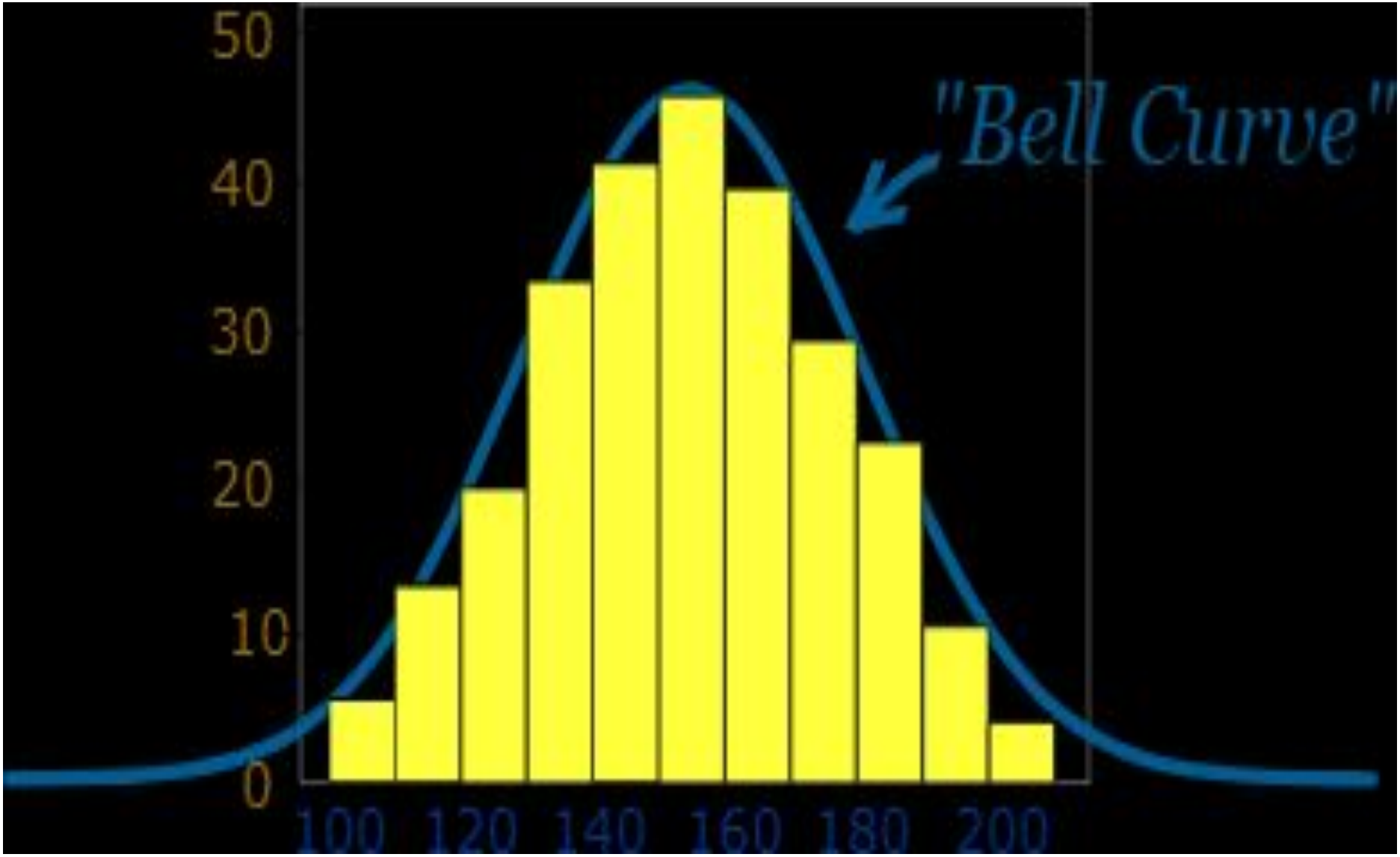
Or more on the right



Or it can be all jumbled up

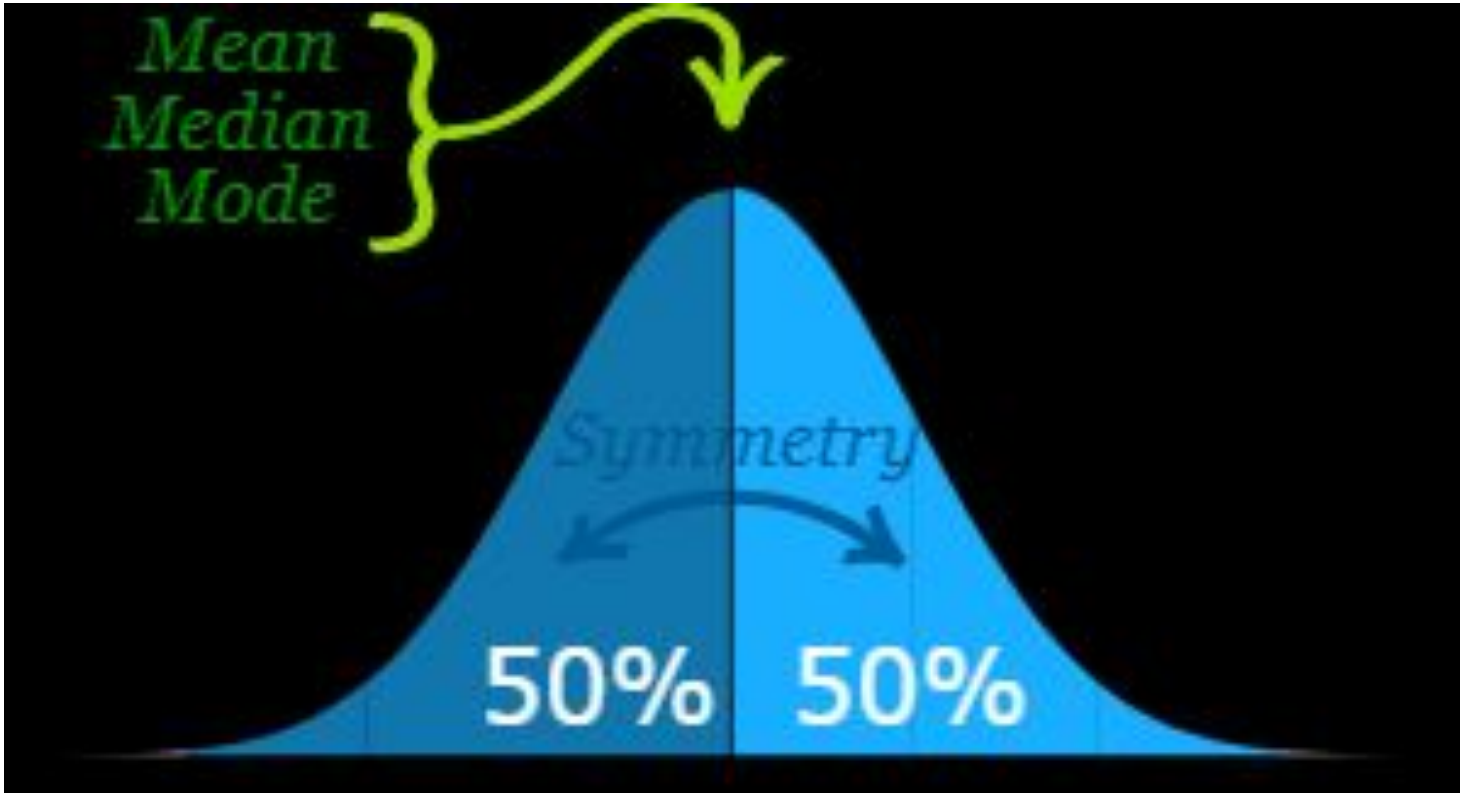


But there are many cases where the data tends to be around a central value with no bias left or right, and it gets close to a "Normal Distribution" like this:



The "Bell Curve" is a Normal Distribution.
Many things closely follow a Normal Distribution:
-heights of people, size of things produced by machines
errors in measurements blood pressure, marks on a test
We say the data is "normally distributed":
The **Normal Distribution** has: Mean=Median=Mode
symmetry about the center 50% of values less than the mean
and 50% greater than the mean

.



Skewness, in basic terms, implies off-centre, so does in statistics, it means lack of symmetry. With the help of skewness, one can identify the shape of the distribution of data.

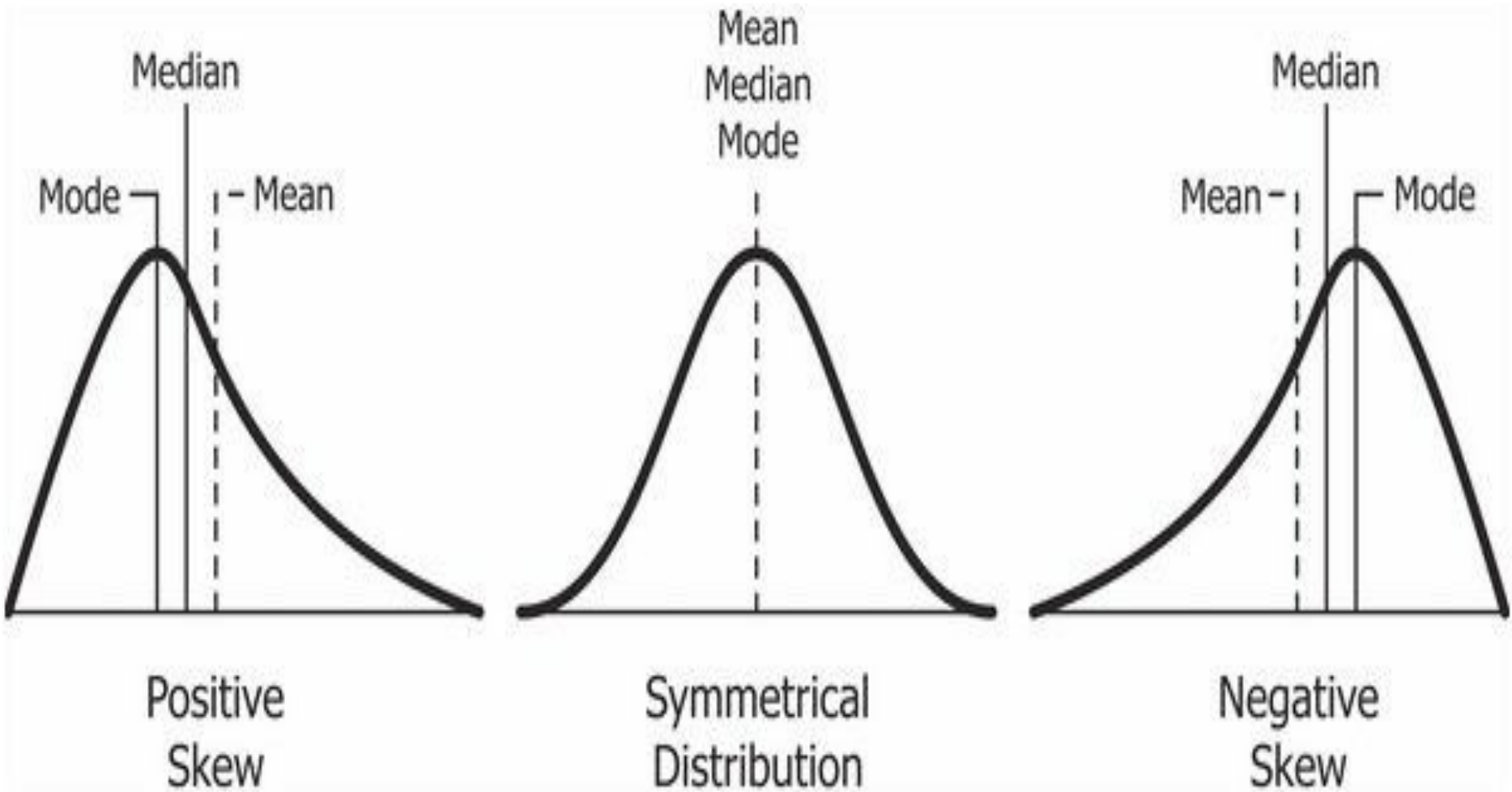
Kurtosis, on the other hand, refers to the pointedness of a peak in the distribution curve. The main difference between skewness and kurtosis is that the former talks of the degree of symmetry, whereas the latter talks of the degree of peakedness, in the frequency distribution.

Data can be distributed in many ways, like spread out more on left or on the right or evenly spread.

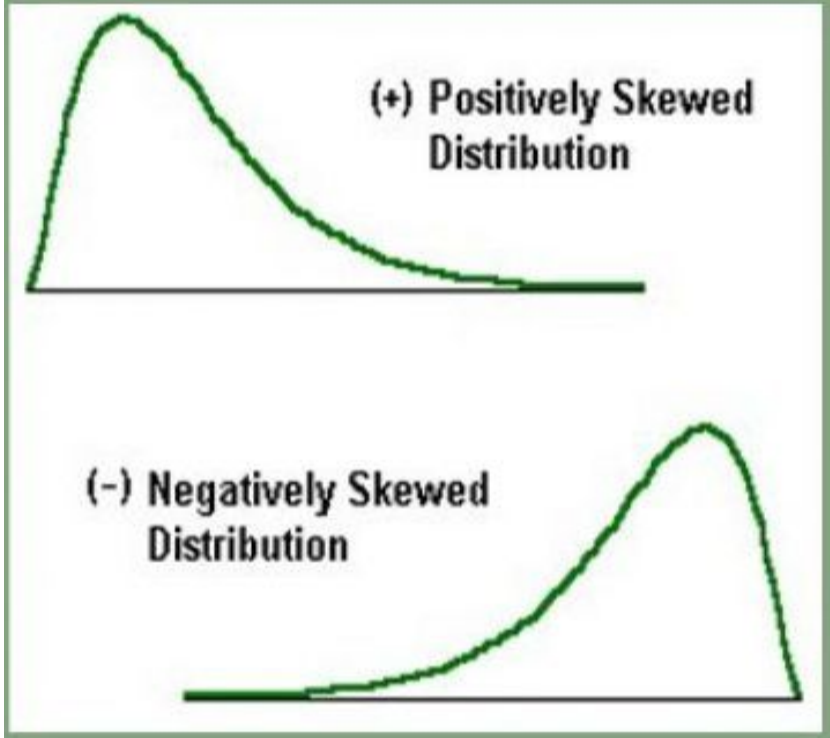
When the data is scattered uniformly at the central point, it called as Normal Distribution (perfectly symmetric) both the sides are equal, and so it is not skewed. Here all the three mean, median and mode lie at one point.

ITEMS	SKEWNESS	KURTOSIS
Meaning	Skewness alludes the tendency of a distribution that determines its symmetry about the mean.	Kurtosis means the measure of the respective sharpness of the curve, in the frequency distribution.
Measure for	Degree of lopsidedness in the distribution.	Degree of tailedness in the distribution.
What is it?	It is an indicator of lack of equivalence in the frequency distribution.	It is the measure of data, which is either peaked or flat in relation to the normal distribution.
Represents	Amount and direction of the skew.	How tall and sharp the central peak is?

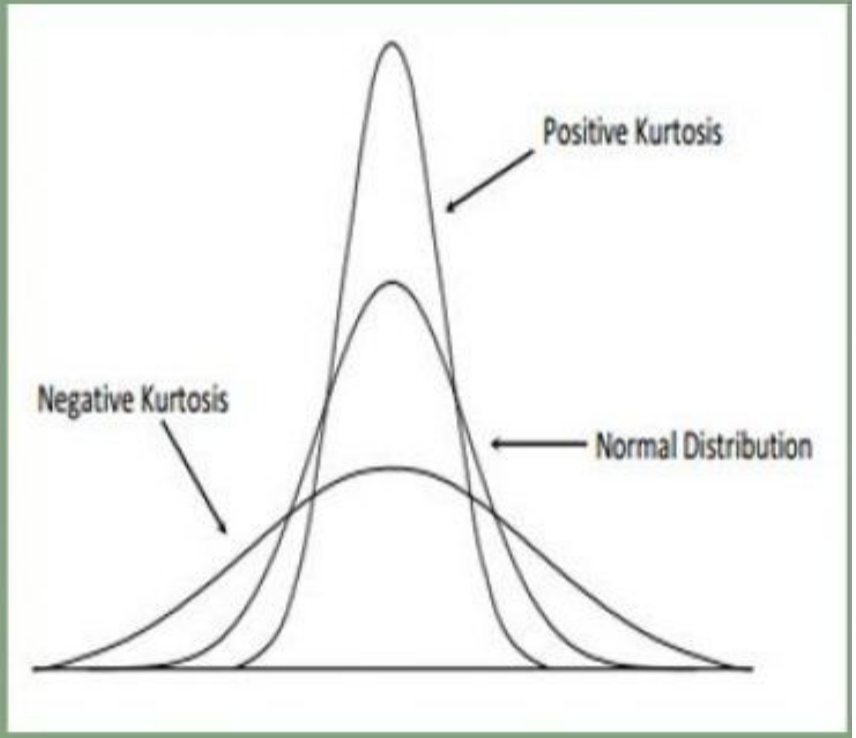
- The term ‘skewness’ is used to mean the absence of symmetry from the mean of the dataset. It is characteristic of the deviation from the mean, to be greater on one side than the other, i.e. attribute of the distribution having one tail heavier than the other. Skewness is used to indicate the shape of the distribution of data. In a skewed distribution, the curve is extended to either left or right side. So, when the plot is extended towards the right side more, it denotes positive skewness, wherein mode < median < mean. On the other hand, when the plot is stretched more towards the left direction, then it is called as negative skewness and so, mean < median < mode.


Median
Mode
Mean
Positive Skew
Mean
Median
Mode
Symmetrical Distribution
Median
Mean
Mode
Negative Skew

Skewness



Kurtosis



- Kurtosis is defined as the parameter of relative sharpness of the peak of the probability distribution curve. It ascertains the way observations are clustered around the centre of the distribution.
- It is used to indicate the flatness or peakedness of the frequency distribution curve and measures the tails or outliers of the distribution.
- Positive kurtosis represents that the distribution is more peaked than the normal distribution, whereas negative kurtosis shows that the distribution is less peaked than the normal distribution. There are three types of distributions:
- **Leptokurtic**: Sharply peaked with fat tails, and less variable.
- **Mesokurtic**: Medium peaked
- **Platykurtic**: Flattest peak and highly dispersed.

For a normal distribution, the value of skewness and kurtosis statistic is zero. The crux of the distribution is that in skewness the plot of the probability distribution is stretched to either side. On the other hand, kurtosis identifies the way; values are grouped around the central point on the frequency distribution.

$$\textit{Sample skewness, } S_k = \{ \Sigma(X_i - X)^3\}/s^3 * 1/n$$

$$\textit{Sample kurtosis, } S_{kr} = \{ \Sigma(X_i - X)^4 \}/s^4 * 1/n$$

Where,

Xi is observation

X is sample mean

s is sample standard deviation

n is number of observations

Example of skewness: 12,13, 54,56,25(Xi)

Calculate mean = 160/5= 32 (X)

Calculate variance (S^2)=374

Calculate standard deviation (s) = 19.34

Calculate $(xi-x)^3$

Xi	Xi-X	$(xi-x)^3$
12	-20	-8000
13	-19	-6859
54	22	10648
56	24	13824
25	-7	-347
160		9266

Sample skewness, $S_k = \{ \Sigma(X_i - X)^3\}/s^3 * 1/n$

Sk = (9262/7233.85) * 1/5 =0.2560

Skewness is positive, Hence, the data has a positively skewed distribution

Interpretation: A positive excess kurtosis indicates a leptokurtic distribution. A zero value indicates a mesokurtic distribution. Lastly, a negative excess kurtosis represents a platykurtic distribution.

Sample Kurtosis

Sample kurtosis, $S_{kr} = \{ \Sigma(X_i - X)^4 \}/s^4 * 1/n$

Skr = (858754/139902.659 * 1/5

= 1.2276 Sample kurtosis is always measured relative to the kurtosis of a normal distribution, which is 3. Therefore, we are always interested in the "excess" kurtosis i.e.*Excess kurtosis = sample kurtosis – 3, Therefore =Excess Kurtosis= 1.2276-3 = -1.7724*

Since the excess kurtosis is negative, we have a platykurtic distribution.

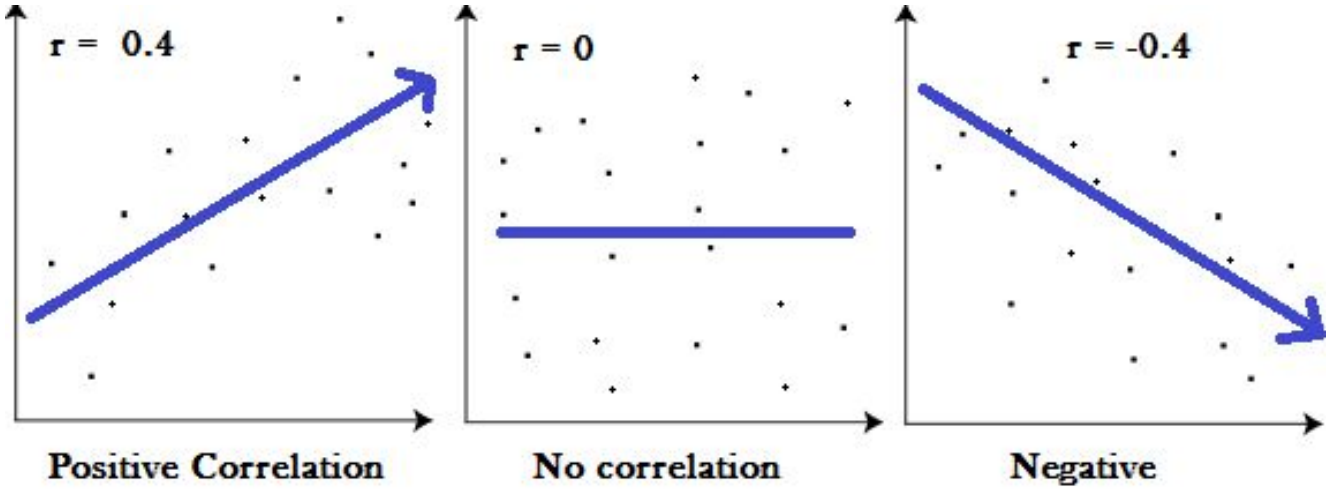
Xi	Xi-X	$(xi-x)^4$
12	-20	160000
13	-19	130321
54	22	234256
56	24	331776
25	-7	2401
160		858754

Simple, Partial and Multiple Correlation and Rank Correlation

The correlation is one of the most common and most useful statistics. A correlation is a single number that describes the degree of relationship between two variables.

Correlation is used to test relationships between two variables (quantitative or categorial). In other words, it's a measure of how things are related. The study of how variables are correlated is called **correlation analysis.**

The Correlation Coefficient is a way to put a value to the relationship. Correlation coefficients have a value of between -1 and 1. A “0” means there is **no relationship** between the variables at all, while -1 or 1 means that there is a **perfect negative or positive correlation.**



The most common correlation coefficient is the Pearson Correlation Coefficient. It’s used to test for linear relationships between data.

Calculating the Correlation

The formula for the correlation is:

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

N	=	number of pairs of scores
Σxy	=	sum of the products of paired scores
Σx	=	sum of x scores
Σy	=	sum of y scores
Σx^2	=	sum of squared x scores
Σy^2	=	sum of squared y scores

Example: Calculate correlation coefficient between age of the farmers and level of knowledge about soybean production practices.

S. NO.	AGE (X)	KNOWLEDGE LEVEL (Y)	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,]\,[\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

Where:

n	=	6
Sum XY	=	20485
Sum X	=	247
Sum Y	=	486
Sum X^2	=	11409
Sum Y^2	=	40022
$(\text{Sum Y})^2$	=	236196
$(\text{Sum X})^2$	=	61009

r = 0.5298

Test of significance for r

We perform a hypothesis **test** of the "significance of the **correlation coefficient**" to decide whether the linear relationship in the sample data is strong enough to use to model the relationship in the population. The hypothesis test lets us decide whether the value of the population correlation coefficient ρ is "close to zero" or "significantly different from zero". We decide this based on the sample correlation coefficient r and the sample size n.

Null Hypothesis H0**:** the population correlation coefficient is not significantly different from zero. There is not a significant linear relationship(correlation) between x and y in the population.

Alternate hypothesis Ha**:** the population correlation coefficient is significantly different from zero. There is a significant linear relationship (correlation) between x and y in the population.

The formula for the test statistic is

$$t = r\sqrt{(n-2)}/\sqrt{(1-r^2)}.$$

The test statistic t has the same sign as the correlation coefficient r . For above example t will be

$$t = 0.5298\sqrt{(4)}/\sqrt{(1-0.280688)}$$

$$= 0.5298\,(2)/0.8481$$

$$= 1.0596\,/0.8481$$

$$= 1.2493$$

(see the t table for 4 degree of freedom)

Partial Correlation Coefficient

Partial correlation is a measure of the strength and direction of a linear relationship between two continuous variables whilst controlling for the effect of one or more other continuous variables (also known as 'covariates' or 'control' variables).

Partial correlation is the measure of association between two variables, while controlling or adjusting the effect of one or more additional variables. Partial correlations can be used in many cases that assess for relationship, like whether or not the yield of a soybean (x1) is related to the expenditure on fertilizer (X2) when the effect of seed rate (X3) is controlled.

The partial correlation coefficient assesses the degree of association between two variables *X*1 and *X*2, when controlling (keeping constant) a third variable *X*3. Mathematically, the partial correlation coefficient r between X1 and X2, when controlling for X3 is written as *r*12.3, and it is computed using the following formula:

$$r12.3 = r12 - r13\ r23 / \sqrt{(1 - r^2 13)}\ \sqrt{(1 - r^2 23)}$$

Multiple Correlation

A **multiple correlation** coefficient (R) yields the maximum degree of liner relationship that can be obtained between two or more independent variables and a single dependent variable.

R^2 represents the proportion of the total variance in the dependent variable that can be accounted for by the independent variables.

Given variables y, x1 and x2, we define the **multiple correlation coefficient** (R).

$$R = \frac{\sqrt{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}(r_{yx_2})(r_{x_1x_2})}}{\sqrt{1 - r_{x_1x_2}^2}}$$

Rank Correlation

A rank correlation coefficient measures the degree of similarity between two rankings, and can be used to assess the significance of the relation between them.

The Spearman rank correlation coefficient, r_s, can take values from +1 to -1.

r_s of +1 indicates a perfect association of ranks,

r_s of zero indicates no association between ranks and r_s of -1 indicates a perfect negative association of ranks. The closer r_s is to zero, the weaker the association between the ranks.

Adoption & knowledge of soybean growers

Adoption	knowledge	Rank of Adoption	Rank of knowledge	d	d^2
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	3	9
62	65	6	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1
61	63	7	6	1	1
total					54

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

rR = 1- 6 (54)/10 (10^2-1)

rR = 1- 324/990

= 1-0.3232

= 0.6768

This indicates a strong positive relationship between the ranks individuals farmers obtained in the adoption and knowledge. That is, the higher you ranked in knowledge, the higher you ranked in adoption also, and vice versa.

Simple & Multiple Regression

Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:

One variable, denoted *x*, is regarded as the **predictor**, **explanatory**,
or **independent** variable.

The other variable, denoted *y*, is regarded as the **response**, **outcome**, or **dependent** variable.

Simple linear regression gets its adjective "simple," because it concerns the study of only one predictor variable.

There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable, or regress. The independent variables can be called exogenous variables, predictor variables, or regressors.

The term ***linear in a simple regression*** model means that it is linear in the parameters; variables in the regression model may or may not be linear.

In a cause and effect relationship, the **independent variable** is the cause, and the **dependent variable** is the effect. **Least squares linear regression** is a method for predicting the value of a dependent variable *Y*, based on the value of an independent variable *X*. *Equation for simple regression analysis*

$$Y = a + bx$$

.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Example: Calculate regression coefficient between age of the farmers and level of knowledge about soybean production practices.

S. NO.	INDEPENDENT VARIABLE AGE (X)	DEPENDENT VARIABLE KNOWLEDGE LEVEL (Y)	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

$b = n (\Sigma xy) - (\Sigma x) (\Sigma y) / n (\Sigma x^2) (\Sigma x)^2$

$b = 6 (20485) – (247) (486) / 6 (11409) -(61009)$

$b = 122910 - 120042 / 68454 - 61009$

$b = 2868/7445$

$b = 0.3852$

Since $Y = a + bx$

Therefore $a = y - bx$ Take the mean of x and y and calculate a

$a = 81 - (0.3852) 41.17$

$a = 81 - 15.85 = 65.15$

Thus fitted equation will be $Y = 65.15 + 0.3852 X$

Multiple Regression

Multiple regression is an extension of simple **linear regression**. It is used when we want to predict the value of a variable based on the value of two or more other variables.

Formally, the model for multiple linear regression, given *n* observations, is

$Y_i = a + b_1 x_{i1} + b_2 x_{i2} + \ldots b_p x_{ip} + e_i$ for $i = 1,2, \ldots n$.

Where,

Y = dependent variable

X_{i1} to x_{ip} are independent variables

a = Intercept value

b1 to bp are regression coefficients

e = error term

Discriminant & Dummy Variable Analysis

Discriminant analysis is a technique that is used by the researcher to **analyze** the research data when the dependent variable is categorical and the independent variable is interval in nature.

e.g. The term categorical variable means that the dependent variable is divided into a number of categories. For example, three size holdings, Small Size, Medium Size and Large size can be the categorical dependent variable.

The objective of discriminant analysis is to develop discriminant functions that are nothing but the linear combination of independent variables that will discriminate between the categories of the dependent variable in a perfect manner. It enables the researcher to examine whether significant differences exist among the groups, in terms of the independent variables. It also evaluates the accuracy of the classification.

Discriminant analysis is a regression based statistical technique used in determining which particular classification or group (such as small, medium and large size of holdings) an item of data belongs to on the basis of its characteristics.

Analysis using dummy variables

- A dummy variable is a numeric variable that represents categorical data, such as gender, race, adopter, etc.
- Their range of values is small; they can take on only two quantitative values. As a practical matter, regression results are easiest to interpret when dummy variables are limited to two specific values, 1 or 0. Typically, 1 represents the presence of a qualitative attribute, and 0 represents the absence.
- The number of dummy variables required to represent a particular categorical variable depends on the number of values that the categorical variable can assume. To represent a categorical variable that can assume *k* different values, a researcher would need to define *k - 1* dummy variables.
- For example, suppose we are interested in gender, a categorical variable that might assume two values male or female (k =2), We could represent gender with one dummy variables:
- X_1 = 1, if male; X_1 = 0 female.

The role of dummy variables in analysis

- Dummy variables are the main way that categorical variables are included as predictors in statistical models.
- If you have a nominal variable that has more than two levels, you need to create multiple dummy variables to "take the place of" the original nominal variable. For example, imagine that you wanted to predict yield of soybean from four categories of districts i.e. yield between 10-12 q, 12-14 q, 14-16 q and more than 16 q.,
- What you need to do is to recode “yield range" into a set of dummy variables, each of which has two levels. The first step in this process is to decide the number of dummy variables. This is easy; it's simply k-1, where k is the number of levels of the original variable (in this case there will be 3 dummy variables).
- We are going to take 3 of the levels of “yield", and create a variable corresponding to each level, which will have the value of yes or no (i.e., 1 or 0). In this instance, we can create a variable called
- “10-12 q," “12-14 q," and “14-16 q." Each instance of “yield range" would then be recoded into a value for “
- If a farmer were a 12-14 q, equal to 1 then “10-12 q " would be equal to 0, if 14-16 q would be equal to 1, and “more than 16" would be equal to 0 and if 10-12 q is 1 than 12-14q would be zero.

Index Number

- An index number is an economic data figure reflecting quantity compared with a standard or base value.
- The base usually equals **100** and the index number is usually expressed as **100** times the ratio to the base value.
- Index number is a technique of measuring changes in a variable or group of variables with respect to time, geographical location or other characteristics. There can be various types of index numbers,

Characteristics of Index Number

- The technique of index numbers is used to measure the relative changes in the level of a phenomenon where the measurement of absolute change is not possible and the series are expressed in different types of items.
- Index numbers are meant to study the changes in the effects of such factors which cannot be measured directly. For example, the general price level is an imaginary concept and is not capable of direct measurement. But, through the technique of index numbers, it is possible to have an idea of relative changes in the general level of prices by measuring relative changes in the price level of different commodities.
- The technique of index numbers measures changes in one variable or group of related variables. For example, one variable can be the price of wheat, and group of variables can be the price of sugar, the price of milk and the price of rice.
- The technique of index numbers is used to compare the levels of a phenomenon on a certain date with its level on some previous date (e.g., the price level in 1980 as compared to that in 1960 taken as the base year) or the levels of a phenomenon at different places on the same date (e.g., the price level in India in 1980 in comparison with that in other countries in 1980).

Simple Aggregative Method

In this method, the index number is equal to the sum of quantities for the year for which index number is to be found divided by the sum of actual quantities for the base year e.g. Price index calculation

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100$$

Where P_{01} Stands for the index number

ΣP_1 Stands for the sum of the prices for the year for which index number is to be found :

ΣP_0 Stands for the sum of prices for the base year.

Commodity	Prices in Base Year 1980 (in Rs.) P_o	Prices in current Year 1988 (in Rs.) P_1
A	10	20
B	15	25
C	40	60
D	25	40
Total	$\Sigma P_0 = 90$	$\Sigma P_1 = 145$

Index Number $(P_{01}) = \frac{\Sigma P_1}{\Sigma P_0} \times 100$; $P_{01} = \frac{145}{90} \times 100$; $P_{01} = 161.11$

Simple Average of quantity Relatives Method:

In this method, the index number is equal to the sum of quantities relatives divided by the number of items and is calculated by using the following formula:

$$P_{01} = \frac{\Sigma R}{N}$$

Where ΣR stands for the sum of price relatives i. e. $R = \frac{P_1}{P_0} \times 100$ and

N stands for the number of items.

Example

Commodity P_0	Base Year Prices (in Rs.) P_1	Current year Prices (in Rs.)	Price Relatives $R = \frac{P_1}{P_0} \times 100$
A	10	20	$\frac{20}{10} \times 100 = 200.0$
B	15	25	$\frac{25}{15} \times 100 = 166.7$
C	40	60	$\frac{60}{40} \times 100 = 150.00$
D	25	40	$\frac{40}{25} \times 100 = 160.0$
N = 4			$\Sigma R = 676.7$

$$\text{Index Number } (p_{01}) = \frac{\Sigma R}{N}$$

$$P_{01} = \frac{676.7}{4}; P_{01} = 169.2$$

Construction of Index Number for Time Series Data

In time series data index number is estimated using following formula

IDN = (Current year data/Base year data) X 100

Year	Area of Wheat (Lakh ha)	Index number
1985-86	22.30	100.00
1986-87	25.60	114.80
1987-88	36.50	163.68
1988-89	45.30	203.14
1989-90	78.89	353.77
1990-91	65.20	292.38
1991-92	82.30	369.06

Hypothesis Testing

Null hypothesis: Null hypothesis is a statistical hypothesis that assumes that the observation is due to a chance factor. Null hypothesis is denoted by; $H0: \mu1 = \mu2$, which shows that there is no difference between the two population means.

Alternative hypothesis: Contrary to the null hypothesis, the alternative hypothesis shows that observations are the result of a real effect it is denoted by HA $\mu1 \neq \mu2$,

Level of significance: Refers to the degree of significance in which we accept or reject the null-hypothesis , so we therefore select a level of significance that is usually 1% or 5%.

Type I error: When we reject the null hypothesis, although that hypothesis was true.

Type II errors: When we accept the null hypothesis but it is false. .

One-tailed test: When the given statistical hypothesis is one value like $H0: \mu1 = \mu2$, it is called the one-tailed test.

Two-tailed test: When the given statistics hypothesis assumes a less than or greater than value, it is called the two-tailed test.

t-Test

- A **t-test** is used as a **hypothesis testing** tool, which allows **testing** of an assumption applicable to a population.
- The test procedure, called the **two-sample t-test**, is appropriate when the following conditions are met:

1. The sampling method for each sample is Simple random sampling.
2. The samples are independent.
3. Each population is at least 20 times larger than its respective samples.
4. The sampling distribution is approximately normal,

At the end of the two trainings, in each training evaluation is made using same standardized test. In first training an average evaluation score of knowledge is 78, with a standard deviation of 10 for 30 farmers; and in the second training average evaluation score is 85, with a standard deviation of 15 for 25 farmers.

Test the hypothesis that both the trainings are equally effective.

State the hypotheses. The first step is to state the null hypothesis and an alternative hypothesis.

Null hypothesis: $\mu_1 - \mu_2 = 0$

Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

The null hypothesis will be rejected if the difference between sample means is too big or if it is too small.

Analyze sample data. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

$$SE = sqrt[(s_1^2/n_1) + (s_2^2/n_2)]$$

$$SE = sqrt[(10^2/30) + (15^2/25] = sqrt(3.33 + 9)$$

$$SE = sqrt(12.33) = 3.51$$

$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$

$$DF = (10^2/30 + 15^2/25)^2 / \{ [(10^2 / 30)^2 / (29)] + [(15^2 / 25)^2 / (24)] \}$$

$$DF = (3.33 + 9)^2 / \{ [(3.33)^2 / (29)] + [(9)^2 / (24)] \} = 152.03 / (0.382 + 3.375) = 152.03/3.757 = 40.47$$

$$t = [(x_1 - x_2) - d] / SE = [(78 - 85) - 0] / 3.51 = -7/3.51 = -1.99$$

where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, n_2 is the size of sample 2, x_1 is the mean of sample 1, x_2 is the mean of sample 2, d is the hypothesized difference between the population means (i.e.zero), and SE is the standard error.

One Sample t test

The **one sample t test** compares the **mean** of your **sample** data to a known value. For **example**, you might want to know how your **sample mean** compares to the population **mean**.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

e. g. Yield of wheat	28	32	26	33	25	28	30	30

Does the actual yield of soybean is deviate significantly from 31 (State average yield)

Compute the mean and standard deviation of your sample. You will find that the mean is equal to 29.00 and the standard deviation is equal to 2.78. Now, compute your *t* statistic:

t = 29-31/(2.78/Squar 8)

t = -2.04

The degrees of freedom (*df*) for your test is equal to n – 1. In this study, *df* = 8 – 1 = 7.

Since your obtained absolute value of *t* (2.04) is less than the critical value (2.365), you would conclude that the average yield of wheat of your sample did <u>not</u> differ significantly from 31.00 q (state average).

ANOVA & F-Test

Analysis of variance (**ANOVA**) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other.

F-Statistic : *The statistic which measures if the means of different samples are significantly different or not is called the F-Ratio*. Lower the F-Ratio, more similar are the sample means. In that case, we cannot reject the null hypothesis.

The data below resulted from measuring the difference in yield of green gram resulting from use of three different doses of N (N0, N10 and N20). The observation in each group was taken for five plants. Thus we have an experiment in which each of three treatments was replicated 5 times.

Observation	N0	N10	N20	Total
1	6.9	8.3	8.0	23.20
2	5.4	6.8	10.5	22.70
3	5.8	7.8	8.1	27.70
4	4.6	9.2	6.9	20.70
5	4.0	6.5	9.3	40.50
Total	26.7	38.6	42.8	108.10
	Mean 5.34	7.72	8.56	

Step1: Compute CM, the correction for the mean

CM= $(TO)^2/N$

TO= Total of all observations (108.10)

N= Number of observations (15)

CM= $(108.10)^2/15$ = **779.041**

Step 2: Compute the total SS. The total SS = Sum of squares of all observations −CM.

$(6.9)^2+(5.4)^2+\ldots+(6.9)^2+(9.3)^2-CM$

829.390−779.041= **45.439**

Step-3: Compute SST, the treatment sum of squares. First we compute the total (sum) for each treatment. SST=$(26.7)^2/5+(38.6)^2/5+(42.8)^2/5-779.041=27.897$

Step- 4 Compute SSE, the error sum of squares. Here we utilize the property that the treatment sum of squares plus the error sum of squares equals the total sum of squares. Hence, SSE= ToSS −TSS=45.349−27.897=17.45.

Step-5 Compute MST, MSE, and their ratio, F.

MST is the mean square of treatments, MST=TSS/k−1=27.8972=13.949

MSE is the mean square of error MSE=ESS/N−k=17.45212=1.454

F=MST/MSE=9.59.

ANOVA-Table

Source	SS	DF	MS	F
Treatments	TSS	k−1	TMS(k−1)	TMS/EMS
Error	ESS	N−k	EMS/(N−k)	
Total	ToSS	N-1		

Source	SS	DF	MS	F
Treatments	27.897	2	13.949	9.59
Error	17.452	12	1.454	
Total	45.349	14		

Where,
K= Number of treatment
N = Number of observations
F= TMS/EMS

Factor Analysis

Factor Analysis is a useful tool for investigating variable relationships for complex concepts such as socioeconomic status, dietary patterns, or psychological scales.

The key concept of factor analysis is that multiple observed variables have similar patterns of responses because they are all associated with a latent (i.e. not directly measured) variable.

For example, people may respond similarly to questions about income, education, and occupation, which are all associated with the latent variable socioeconomic status.

Analytical Methods

Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors. This technique extracts maximum common variance from all variables and puts them into a common score. As an index of all variables, we can use this score for further analysis.

Principal Component analysis: PCA starts extracting the maximum variance and puts them into the first factor. After that, it removes that variance explained by the first factors and then starts extracting maximum variance for the second factor. This process goes to the last factor.

2. **Common factor analysis: I**t extracts the common variance and puts them into factors. This method does not include the unique variance of all variables.

3. **Image factoring:** This method is based on correlation matrix. OLS Regression method is used to predict the factor in image factoring.

4. **Maximum likelihood method:** This method also works on correlation metric but it uses maximum likelihood method to factor.

Cluster Analysis

Cluster: a collection of data objects.

Similar to one another within the same cluster
Dissimilar to the objects in other clusters

- Cluster analysis – Grouping a set of data objects into clusters
- Clustering is unsupervised classification: no predefined classes
- Typical applications – As a stand-alone tool to get insight into data distribution –

As a preprocessing step for other algorithms

Examples of Clustering Applications

Marketing: Help marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs **Land use**: Identification of areas of similar land use in an earth observation database

Insurance: Identifying groups of motor insurance policy holders with a high average claim cost

City-planning: Identifying groups of houses according to their house type, value, and geographical location

Disaster studies: Observed disaster epicenters should be clustered according to nature of disaster events.

Characteristics of Good Clustering

- A good clustering method will produce high quality clusters with – high intra-class similarity – low inter-class similarity.
- The quality of a clustering result depends on both the similarity measure used by the method and its implementation.
- The quality of a clustering method is also measured by its ability to discover some or all of the hidden patterns.

Conjoint Analysis

- Conjoint analysis is an advanced market research analysis method that attempts to understand how people make complex choices.
- Every day, people make choices that require trade-offs—so often that we may not even realize it. Simple decisions such as, “what type of seed should I buy?” or “which market should I prefer for sale of mango?” all contain multiple elements that ultimately formulate a decision.
- Conjoint analysis is one of the most effective models in extracting consumer preferences during the purchasing process into a quantitative measurement. It evaluates products or services in a way no other method can.

- Conjoint analysis is a set of market research techniques that measures the value the market places on each feature of your product and predicts the value of any combination of features.
  Conjoint analysis is, at its essence, all about features and trade-offs.
- By asking respondents to “trade-off” one product feature to obtain another, conjoint unearths the importance of product features to consumers’ purchase decisions, and calculates the particular combination of price and specific product features which maximize a product’s appeal to consumers.

Conjoint Overview

Factors or Features

Levels within Factors

Key Outputs:
-Relative importance of factors
-Preference for levels of a given factor
-Utility structures

Key Analysis:
-Simulation models
- **-current product mix**
- **-new product development**
- **-segmentation**

-Price elasticity

Multidimensional Analysis

- Multidimensional analysis is a class of procedures for representing perceptions and preferences of respondents spatially by means of visual display.
- Perceived psychological relationships among stimuli are represented as geometric relationships among points in multidimensional space.
- These geometric representations are often called spacial maps.

Multidimensional scaling are use for:

- To determine the number and nature of dimensions consumers use to perceive different brands.
- To position brands on these dimensions.
- To identify the position of consumer's ideal brand.

Applications in marketing:

- Brand image measurement
- Market segmentation
- New product development – by looking at the spacial map the empty spaces represent the unexplored by competitors market segments.
- The development of a product or a service to fit in the unused space on the spacial map might have a commercial sense.
- Assessing advertising activities and its effectiveness
- Distribution channel decisions – managers by looking at the multidimensional map of relevant brands can judge whether creating an additional retail outlet near other brand's outlets is compatible with their branding strategy.

Business Decisions & Equations

Example 1 A custom hiring center has 250 tractors to rent. If they rent x tractor then their monthly profit, in rupees, is given by,

$P(x)= -8x^2+3200x -80,000$. How many tractors should they rent in order to maximize their profit?

All that we're really being asked to do here is to maximize the profit subject to the constraint that x must be in the range $0 \leq x \leq 250$.

First, we'll need the derivative and the critical point(s) that fall in the range $0\leq x \leq 250$.

$dp/dx= -16x + 3200 \Rightarrow 3200-16x=0, \Rightarrow x =3200/16=200$

$P(0)=-80,000$ (how$-8x^2+3200x-80,000$ i.e.$-8 (0)^2 +3200 (0)-80000$

$P(250) =22,0000$ (how$-8x^2+3200x-80,000$ i.e.$-8 (250)^2 +3200 (250) -80000$)

$P (200) = 24,0000$ (how$-8x^2+3200x-80,000$ i.e.$-8 (200)^2 +3200 (200) -80000$)

Use of Determinants and Matrices in Business Decisions

A **matrix** is simply a set of numbers arranged in a rectangular table. Example of a 2 × 4 matrix. It has 2 rows and 4 columns. We usually write matrices inside parentheses ( ) or brackets [ ].

$$\begin{bmatrix} 2 & 4 & -1 & 0 \\ 1 & 3 & 7 & 2 \end{bmatrix}$$

We can add, subtract and multiply matrices together, under certain conditions.

Matrices are used to solve business problems in:

- linear programming
- optimisation

A **Determinant**

A **determinant** is a square array of numbers (written within a pair of vertical lines) which represents a certain sum of products.

In general, we find the value of determinant of 2 × 2 matrix with elements *a*, *b*, *c*, *d* as follows:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - cb$$

Example:

$$\begin{vmatrix} 4 & 1 \\ 2 & 3 \end{vmatrix} = 12 - 2 = 10$$

Role and uses of quantitative techniques in business decision making

Quantitative techniques in management study include **methods** or tools, which focus on objective measurement, and analyzing numbers in order to draw a conclusion about given problems. It is a scientific **method** or **technique used** by the **business** world for problem solving and **decision-making**.

Project Management: Quantitative methods have found wide applications in project management. These techniques are used for optimizing the allocation of manpower, machines, materials, money and time. Projects are scheduled with quantitative methods and synchronized with delivery of material and workforce.

Production Planning and Scheduling

Determining the size and location of new production facilities is a complex issue. Quantitative techniques aid in evaluating multiple proposals for costs, timing, location and availability of transportation. Product mix and scheduling get analyzed to meet customer demands and maximize profits

Purchasing and Inventory: Predicting the amount of demand for a product is always risky. Quantitative techniques offer guidance on how much raw material to purchase, levels of inventory to keep and costs to ship and store finished products.

Marketing: Marketers apply quantitative methods to set budgets, allocate media purchases, adjust product mix and adapt to customers' preferences.

Finance : Financial managers rely heavily on quantitative techniques. They evaluate investments with discounted cash flow models and return on capital calculations. Products get analyzed for profit contribution and cost of production. They use probabilities and statistics to prepare annual profit plans.

Research and Development: Managers look to mathematical projections about the probability of success and eventual profitability of products to make investment decisions.

Agriculture: Operations research techniques have long been employed by farm managers. They utilize decision trees and make assumptions about weather forecasts to decide which crops to plant. If forecasters predict cold weather, is it more profitable to plant corn or wheat? What happens if the weather is warm? These are all probabilities that farmers use to plan their crop rotations.

- **Primary and secondary data**
- **Sources of data, instruments of data collection**
- **Mailed questionnaires and interview schedule**
- **Preparation of schedule**
- **Structured and unstructured-open and close ended questions.**
- **Scaling techniques, scales of measurement- Nominal, ordinal, interval, ratio, Likert scale and other scales.**
- **Data editing, classification, coding, validation, tabulation, presentation and analysis.**

The Building Blocks of Research

KNOWLEDGE

Blend of information, experience and in-sights that provides a framework that can be thoughtfully evaluated when assessing new information or evaluating relevant situations (e.g. extent of yield gap that needs to be bridge through adoption)

INFORMATION

Determination of relationship amongst data with a view to facilitating understanding of the phenomena, their relationships and decision-making (e.g. analyzing yield gap in wheat)

DATA

Measurements of phenomena (e.g. yield of wheat of different farmers)

Categories of Data

Something assumed as facts and made the basis of reasoning or calculation.

1. **Qualitative or Categorical:**

Categorical data represents characteristics, it can represent things like a person's gender, language etc. Categorical data can also take on numerical values (e.g. 1 for female & 0 for male). Those numbers don't have mathematical meaning other examples Color, Race, caste etc.

2. **Quantitative or Numerical: these are of two types**

Discrete Data: if its values are distinct and separate. In other words: We speak of discrete data if the data can only take on certain values. This type of data **can't be measured but it can be counted**. It basically represents information that can be categorized into a classification. An example is the number of labour, number of housesholds, number of tractors.

Continuous Data: Continuous Data represents measurements and therefore their values **can't be counted but they can be measured**. An example would be the height of a person, age of the farmer, size of holding, which you can describe by using intervals on the real number line.

Categorical Data

Nominal: Categories of data cannot be ordered one above the other. Sex: Male, Female or Marital Status: Single, Married, Divorced**.** Nominal values represent discrete units and are used to label variables, that have no quantitative value. Just think of them as „labels“. Note that nominal data that has no order. Therefore if you would change the order of its values, the meaning would not change.

What language do you speak (tick

1. **English**
2. **Hindi**
3. **Other…….**

Ordinal: Categories of the data can be ordered one above the other or vice versa. Level of Knowledge: Good, Average, Poor or Opinion: Fully Agree, Agree, Disagree. Ordinal values represent discrete and ordered units. It is therefore nearly the same as nominal data, except that it's ordering matters. You can see an example below:

What is your educational background

1. **Primary**
2. **High school**
3. **Graduate**
4. **Postgraduate**

Types of data

Primary Data – refers to the data that the investigator collects for the very first time. This type of data has not been collected either by this or any other investigator before. A primary data will provide the investigator with the most reliable first-hand information about the respondents e.g. farmers field data, companies survey etc.

Secondary Data – refers to the data that the investigator collects from another source. Past investigators or other agencies collect data required for their study. Moreover, the investigator does not have a clear idea about the intricacies of the data. There may be ambiguity in terms of the sample size and sample technique. There may also be unreliability with respect to the accuracy of the data e.g. published data from statistical abstracts, Agriculture at a Glance, etc.

Sources of Data

Primary data sources

Direct Personal Investigation

Consists of the collection of data by the investigator in a direct manner. The investigator is responsible for personally approaching a respondent and investigating the research and gather appropriate information. Thus, this method of data collection ensures first-hand information. This data is all the more reliable for an intensive research.

- But in an extensive research, this data is inadequate and proves to be unreliable.
- This method of collection of data is time-consuming.
- Greatest demerit is that this method is very subjective in nature and is not suitable for objective based extensive researches.

Indirect Oral Interview

Consists of the collection of data by the investigator in an indirect manner. The investigator approaches an indirect respondent who possesses the appropriate information for the research. Thus, this method of data collection ensures first-hand information because the interviewers can cross-question for the right and appropriate information.

Mailed Questionnaire

Consists of mailing a set or series of questions related to the research. The respondent answers the questionnaire and forwards it back to the investigator after marking his/her responses. This method of collection of data has proven to be time-saving. It is also a very cost-efficient of collecting the required data. An investigator who has the access to the internet and an email account can undertake this method of data collection. The researcher can only investigate those respondents who also have access to the internet and an email account. This remains the only major restriction of this method.

Interview Schedules : Scheduling involves a face to face situation with the respondents. In this method of collecting data, the interviewer questions the respondent according to the questions mentioned in a form. This form is known as a schedule. This is different than a questionnaire. A questionnaire is personally filled by the respondents and the interviewer may or may not be physically present. Whereas, the schedule is filled by the interviewer after asking the respondent his/her answer to a specific question. And in scheduling method of collecting data, the interviewer is physically present.

Local agencies : In this method, the information is not directly or indirectly collected by either the interviewer. Instead, the interviewer hires or employs a local agency to work for him/her and help in gathering appropriate information. These local agents are often known as correspondents as well. Correspondents are only responsible for gathering accurate and reliable information. They work according to their preference and adopt different methods to do so.

Secondary data:

Published Sources

There are many national organizations, international agencies and official publications that collect various statistical data. They collect data related to business, commerce, trade, prices, economy, productions, services, industries, currency and foreign affairs. They also collect information related to various (internal and external) socio-economic phenomena and publish them. These publications contain statistical reports of various kinds. Central Government Official Publication, Publications of Research Institutions, Committee Reports and International Publications are some published sources of secondary data. Many websites are their for gathering secondary published data. e.g. area, production and productivity of different crops in Madhya Pradesh

Unpublished Sources

Some statistical data are not always a part of publications. Such data are stored by institutions an private firms. Researchers often make use of these unpublished data in order to make their researches all the more original. Daily weather data available with weather station of JNKVV, Jabalpur.

Preparation of Schedule

Six major steps for forming a schedule in social Science research, i.e,

1) **Knowledge About the Different Aspects of Problem,**
2) **Knowledge About the Information to be Studied,**
3) **Framing the Actual Questions,**
4) **Content of Schedule,**
5) **The Fifth Step is the General Layout of the Schedule, and**
6) **Testing the Validity of Schedule.**

Step 1- Knowledge About the Different Aspects of Problem: While framing the schedule the first step is to have proper knowledge about the different aspects of the problem. The researcher has to put a great deal of thought into the selected research problem.

(i) The researcher must have interest in the topic of research. (ii) The nature of the problem or topic must have some social reference. (iii) Should understand the problem thoroughly. (iv) The problem must be defined in clear and explicit manner. (v) The problem must be defined unambiguously, so that it will help to differentiate relevant data from irrelevant ones. (vi) Existing literature on the subject should be studied. (vii) The problem under study should be split up into various aspects; the determination of these aspects depends upon the clear understanding of the problem. Exp. study of yield gap of soybean, the researchers must have knowledge about package of practices and different input use in production.

Step 2 -Knowledge About the Information to be Studied: While framing a good schedule the second important step is to decide what information is necessary for a valid generalization on each aspect of the problem. An extensive literature survey usually helps the researcher to get a proper knowledge about different aspects of the research problem. By studying the previous studies in the related field the researcher gets knowledge about relevant information needed for his current study. He can again subdivide each aspect of the topic. e.g. in soybean yield gap study researcher can divide reasons for yield gap as technological, sociological, economic, policy, and other.

Step 3- Framing the Actual Questions: This is the most essential part of the schedule and any error in it may invalidate the whole research study by providing biased, incorrect, incomplete or irrelevant information. While framing the actual questions in a schedule the following are the some of sub-steps one must take into consideration.

(a) Nature of Questions to be Given: For framing a schedule there is no exclusive rules & regulation regarding the selection of nature of questions to be asked. It all depends on the nature of the research topic, skill of the researcher, kinds of respondents and other factors.

(ii) Simple language: In choosing the language for a schedule, the population being studied should be kept in mind. The aim in question wording is to communicate with respondents as nearly as possible in their own language. e.g. What are the different sources of irrigation? Or Area irrigated under different sources (ha)

Tube well	Canal	Open well	River	Tank	Other

(iii) Attention to Questions Involving Memory: Most factual questions, to some extent, involve the respondent in calling information. His degree of success in presenting this accurately is thus a basic determinant of the quality of his response. e.g. Name the varieties of soybean sown during last five years?

(iv) Questions must be within the Intellectual Capacity of the Respondent: The questions included in the schedule should be within the respondents' intellectual capacity to give answer. The researcher should not expect any reply which is beyond his informational scope. For example, an illiterate cannot replay about e-commerce, internet, e.NAM, etc.

(v) Inter-Relation of Questions: Various questions asked by the researcher should be inter-related with each other. They should be asked in a proper order, so that it will be systematic, interesting and continuous one. e.g. Which variety of soybean you prefer?

What is the seed rate of this variety?

(vi) Cross-Checking Questions:

In a schedule the researcher should include certain questions for cross checking. It will provide a scope of verification to the researcher and he can check the incorrect or bias answers of the respondents.

e.g. What is the level of income of the family (a) below 1 lakh/year (b) 1.1 to 2 lakh (c) 2.1 to 3 lakh.

What are source of income of the family (Rs/year)?

Farming	Dairy	Poultry	Wages	Other

(b) Questions to be Avoided:

(i) Ambiguous Questions:

Ambiguous questions are to be avoided. If an ambiguous word creeps in, different people will understand the questions differently and will in effect be answering differently to the some question, e.g. Are loan waiver is beneficial for you? The question was asked to all farmers in the survey, irrespective of whether they were taken loan or not. What then, did a 'No' answer mean? Depending on the respondent, it might have meant—"No, I am not benefited". Such ambiguity has to be avoided in any social research, otherwise it will minimize the objectivity of research.

(ii) Double Barreled Questions: Ambiguity may also arise with double barreled questions, such as the following question on varietal adoption, “Do you like JS 9752 or JS 9560?” Respondents liking one and disliking the other would be in a dilemma in answering this question. Clearly it needs to be divided into two separate questions, each concerned with a single idea, in this case with a single variety.

(iii) Vague Words: Vague words and phrases like ‘kind-of, ‘fairly’, ‘generally’, ‘often’, ‘much the same’, ‘on the whole’, etc. should be avoided. If one asks ”What kind of house do you have”? Without specifying a frame of references, some people will answer that is 2BHK, others that it is sub-urban, others that it is very pleasant and so on. A similar type of vagueness occurs in ‘why’ questions. In answering the question “Why did you go to KVK?”. Some respondents will say that they wanted to see the technology, some that ‘they did not want to stay at home’, others that ‘the friend suggested it’.

(iv) Leading or Suggestive Questions: Leading or suggestive type of questions should be avoided as they result in biased answers. A leading question is one which, by its content, structure or wording, leads the respondent in the direction of a certain answer. For example "You don't think that KVK is best source for information?" as obviously leads to a negative answer and the question form like "Should not something be done about KVK for information dissemination ?" leads to a positive one.

(v) Presuming Question: While forming a schedule, the researcher should not presume anything about the respondent. For example questions like "How did you vote in the last election?" are best asked only after a 'filter question' has revealed that the respondent did vote in the last election. Without knowing this the researcher should not presume anything about the respondent. Otherwise the respondent may feel insulted and become reluctant to provide various information on research topic.

vi) Hypothetical Questions: Hypothetical question like “Would you like to live in a rural area?” one of very limited value. Another kind of hypothetical question is “Would you like a more frequent bus service?” or “Would you like an increase in wages?” Such questions are unlikely to be of any value because the respondent is being asked if he would like something for nothing. It is hard to see how he could possibly say “No”. If he did, it could be because he has taken into account some hidden factors of his own, or because he has failed to understand the question.

(vii) Personal Questions: Questions regarding personal, private or secret things of a respondent should be avoided unless they are relevant for the inquiry. People are usually reluctant to disclose their personal matters regarding savings, investments, various diseases etc. to a stranger.

(viii) Embarrassing Questions: Questions that may put the respondent in an embarrassing position should also be avoided. Subjects which people do not like to discuss in public create a problem to the schedule designer. Respondents are often embarrassed to discuss their personal matters, to give low-prestige answers, and to admit to socially unacceptable behaviour and attitudes. If, for instance, questions on frequency of taking bath, cheating in examinations or attitudes to communism were asked in the usual way, many respondents would probably refuse to reply and other would distort their answer.

(ix) Too Long Questions: Too long questions are boring and respondents do not follow it easily. If a researcher feels the necessity of giving a long question then he should break it into some interrelated parts, so that it will be easier on the part of the respondent to reply it.

(x) Question Causing suspicion: Question which creates suspicion in the mind of the respondent like question about monthly income, accumulation of wealth etc. should be avoided as far as possible unless they are absolutely necessary.

(xi) Question on Sensitive Issue: Question creates ill feeling for others or hurt one's emotion like "Are religious practices unscientific?" "Is Islam religion better than Hinduism?"

(xii) Question Against Universally Accepted Norms: Every society is having its own accepted normative structure. The members of that particular society always show respect to these norms. Asking the chicken liking to vegetarian

(c) Language: While framing a schedule the researcher should be careful about the proper wording or language.

Following types of word should be avoided as far as possible:

(i)Abbreviation: In order to reply a question the respondent should understand it clearly. A particular abbreviation given in schedule may be known to the researcher but the respondents may not understand it. Therefore, the researcher should try to avoid such abbreviations. If such abbreviations are used then it's meaning and full form should be given in the schedule for the better understanding of the respondents e. g. a.i. in case of agro-chemical

(ii) Value–Loaded Words: Words carrying different values viz. good and bad should be avoided as far as possible.

(iii) Native or Unusual Words: The researcher should try to avoid highly localized languages in his schedule. It is always better to use the words which can be understood by everyone e.g. Bigha.

iv) Multi-Meaning Words: Words carrying different, meanings should be avoided e.g. value and cost of the product

(d) Sequence of Questions: In order to get a proper response the sequence of questions in a schedule needs to be planned. A proper sequence of questions, e.g. variety sown followed by seed rate, seed treatment etc.

(e) Types of Questions:

)**Open end Question:** The open form, open end or unrestricted type of questions call for a free response in the respondent's own words. The respondent is having much freedom here to provide his own response. No clues are provided. It probably provides for gathering depth of response. The respondent reveals his mind, provides his frame of reference with the reasons for his responses. They are used mostly in pilot studies to get an idea about the research area and the possible replies.

The following are the some of the examples of open end questions: 1: What is your view about the current budget?

2: Is it beneficial for the poor people?

3.: Enlist the constraints of soybean production?

(ii) Closed form of question: The questions that call for short, limited responses are known as restricted or closed form of questions. They provide for marking a yes or no, a short response, or checking an item out of a list of given responses. It restricts the choice of response for the respondents. He has simply to select a response out of supplied responses and has not to frame responses in his own way. The following are the illustrations of closed form of questions.

1: Are you literate? Yes/No.

2: Are you a housewife? Yes/No.

Many of the questions here are opinion questions, in which the respondents are given choice between 'good' and 'bad', 'very bad', 'important', 'very important' and 'not at all important'. Such questions are very common in opinion Research.

(iii) Multiple Choice Questions: In these questions the reply is not confined to two alternatives only, but to a number of possible alternatives. For example 'What, according to you is an important cause of poverty in India? (a) population growth (6) lack of education (c) lack of governmental initiative for its eradication (d) illness of people (e) lack of industry (f) any other (specify).

Step 4- Content of Schedule:

The fourth step in forming a schedule is to prepare the content of a schedule. It is nothing but the systematic structure of a schedule.

The whole schedule may be divided into three parts viz.

(a) Introductory part,

(b) Main schedule and

(c) Instructions to the interviewer/observer.

(a) Introductory Part: This part includes introductory information about the schedule and its respondents.

In this opening part, the following type of information with regard to inquiry and respondent are sought:

(i) Name of the survey with the name and address of its conducting authority.

(ii) Reference or case number.

(iii) Name of the respondent, his address, age, sex, education, profession etc.

(iv) Place of interview.

(v) Time and date of interview.

(b) Main Schedule: This is the main and vital portion of the schedule. It has to be prepared with great care. This part of schedule contains with different questions, columns, as well as blank tables where information supplied by the respondent has to be filled.

(c) Instruction to Interviewer: In this part the field worker who has to present the schedule and collect data is given elaborate instructions regarding the presenting the schedule and the method of interview. Instructions about the use of various units, technical terms, general method of fulfilling the schedule and the way in which the interview is to be conducted smoothly be given.

Step 5- The Fifth Step is the General Layout of the Schedule: The layout or physical design of the schedule is very important. If it is planned properly the interview will bring high response.

The following are some of the steps which may be taken to achieve this objective:

(i) Size of Schedule: Usually small sized schedule is preferred by the respondents because they can follow a small sized schedule easily.

(ii) Paper: The paper used for printing of schedule should be of high quality.

(iii) Margin: The margin on the left should be about and on the right it should be 1 ½. This makes the schedule an attractive one. Besides the researcher can take some notes in this marginal space. Absence of margin may create problem for punching.

(iv) Spacing: In between the questions, titles, subtitles and columns there should be reasonable space for noting the responses and demarcating one from the other.

(v) Printing: A printed schedule is obviously more desirable as printing makes a schedule more attractive.

(vi) Use of Picture: Sometimes the use of pictures in schedule influences the respondent in right manner and the respondent takes greater interest for giving replies. Therefore, it is desirable to insert suitable pictures whenever possible.

Step 6# Testing the Validity of Schedule: The last step for forming a schedule is testing the validity of schedule. After the schedule has been prepared the investigator should test it on a sample population in order to examine its validity. Thus various mistakes, unsatisfactory or unnecessary things, can be located only when the schedule has been operated on trial basis.

Scaling Techniques

Definition: Scaling is the process of generating the continuum, a continuous sequence of values, upon which the measured objects are placed.

Scaling emerged from the social sciences in an attempt to measure or order attributes with respect to quantitative attributes or traits. Scaling provides a mechanism for measuring abstract concepts. A comparative scale is an ordinal or rank order scale that can also be referred to as a nonmetric scale. Respondents evaluate two or more objects at one time and objects are directly compared with one another as part of the measuring process.

Scaling techniques, scales of measurement- Nominal, ordinal, interval, ratio, Likert scale and other scales.

Variables

Variables are things that we measure, control, or manipulate in research. They differ in many respects, most notably in the role they are given in our research and in the type of measures that can be applied to them.

Dependent vs. Independent Variables

Independent variables are those that are manipulated whereas dependent variables are only measured or registered.

Independent variable: fertilizer doses, seed rate, number of labour

Dependent variable: Yield, production ,area, income

Variables differ in how well they can be measured, i.e., in how much measurable information their measurement scale can provide. There is obviously some measurement error involved in every measurement, which determines the amount of information that we can obtain. Another factor that determines the amount of information that can be provided by a variable is its type of measurement scale. Specifically, variables are classified as (a) nominal, (b) ordinal, (c) interval, or (d) ratio.

Nominal scale measurement

- **Nominal measurement allow for only qualitative classification.**
- **That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories,**
- **but we cannot quantify or even rank order those categories.**
- **For example, all we can say is that two individuals are different in terms of variable A (e.g., they are of different colour), but we cannot say which one "has more" of the quality represented by the variable. Typical examples of nominal variables are *gender*, *race*, *color*, *city*, etc.**

Ordinal Scale

- **Ordinal measurement allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable,**
- **but still they do not allow us to say "how much more."**
- **A typical example of an ordinal variable is the socioeconomic status of families.**
- **For example, we know that upper-middle is higher than middle but we cannot say that it is, for example, 18% higher.**
- **Another example high school education is higher than primary Scholl education but we can not say that it is 50% higher.**

Interval scale

- **Interval scale allow us not only to rank order the items that are measured,**
- **but also to quantify and compare the sizes of differences between them.**
- **For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale.**
- **We can say that a temperature of 40 degrees is higher than a temperature of 30 degrees,**
- **and that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees.**

Ratio scale

- Ratio variables are very similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus, they allow for statements such as x is two times more than y.
- Typical examples of ratio scales are measures of time or space.
- For example, as the Kelvin temperature scale is a ratio scale, not only can we say that a temperature of 200 degrees is higher than one of 100 degrees, we can correctly state that it is twice as high.
- Interval scales do not have the ratio property. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

Rank-Order Scaling

This gives the respondent a set of items and then asks the respondent to put those items in some kind of order. The “order” could be something like preference, liking, importance, effectiveness, etc. This can be a simple ordinal structure such as A is higher than B or be done by relative position (give each letter a numerical value as in A is 10 and B is 7). You could present five items and ask the respondent to order each one A-E in order of preference. In Rank-Order scaling only (n-1) decisions need to be made.

Likert Scale

- Likert (1932) developed the principle of measuring attitudes by asking people to respond to a series of statements about a topic, in terms of the extent to which they agree with them, and so tapping into the cognitive and affective components of attitudes.
- A Likert-type scale assumes that the strength/intensity of experience is linear, i.e. on a continuum from strongly agree to strongly disagree, and makes the assumption that attitudes can be measured. Respondents may be offered a choice of five to seven or even nine pre-coded responses with the neutral point being neither agree nor disagree.
- A Likert scale is a rating scale, often found on survey forms, that **measures how people feel about something**.

Likert Scale Examples

Agreement

1. Strongly Agree
2. Agree
3. Undecided
4. Disagree
5. Strongly Disagree

Frequency

1. Very Frequently
2. Frequently
3. Occasionally
4. Rarely
5. Never

Importance

1. Very Important
2. Important
3. Moderately Important
4. Of Little Importance
5. Unimportant

Likelihood

1. Almost Always True
2. Usually True
3. Occasionally True
4. Usually Not True
5. Almost Never True

Data Editing

Data editing is defined as the process involving the review and adjustment of collected survey **data**. The purpose is to control the quality of the collected **data**.

Data editing can be performed manually, with the assistance of a computer or a combination of both.

Purpose of data editing: With editing the data the researcher makes sure that all responses are now very clear to understand.

Bringing clarity is important otherwise the researcher can draw wrong inferences from the data.

The respondents might not be able to express their opinion in proper wording. The editor can rephrase the response, but he needs to be very careful in doing so. Any bias can be introduced by taking the wrong meanings of the respondents point of view. e.g. name of rice variety pony in place of pony farmers may say bony.,

Care to be taken in data editing

1. **Omission:** By chance or by some mistake some responses are left incomplete, the editor has to see what has been an oversight by the respondent. In an interview you can better assess what they want to tell and what they are trying to hide. **e.g.name of variety is not filled by the interviewer in that case try to get common variety grown in that area.**
2. **Avoid biased editing:** The editor needs to be very objective and should not try to hide or remove any information. He should not add anything in the responses without any sound reason. He should have to be confident in making any changes or corrections in the data. In short, he should make least changes and only logical changes. He should not add anything that shows his opinion on the issue.

3. **Make judgments:** Sometimes the respondents leave something incomplete, to complete the sentence or a phrase the editor has to make a judgment. He should have to have good judgment to do so. He should do it so well that his personal bias do not involve in the responses. e.g. name of soybean variety JS 95. This should be JS 95-60

4. **Logical adjustments:** Logical adjustments must be made or otherwise the data will become faulty. There might be need for some logical corrections, for example, a respondent gives that he is graduate in education status and in age status he has given 15 year.

Classification & organization of Collected

Data classification is the process of sorting and categorizing **data** into various types, forms or any other distinct class.

Data classification enables the separation and **classification** of **data** according to **data** set requirements for various objectives of the study. It is mainly a **data** management process.

In other words, classification of data is the process of organizing data into groups according to various parameters. The most crucial parameter is the similarities that exist among data.

For example, the number of soybean growers who have trained by KVK- Jabalpur can be classified on the following basis:

1. Gender
2. Educational status
3. Size of holdings
4. Proximity to market
5. Socio-economic status

Functions of Classification of Data

Classifying data have the following functions:

1. **Studying relations** – classifying the collected data helps analyse and study the relationships between them. Moreover, the organization of statistical data can enable effective decision making e.g. relationship between size of holding and income.

2. **Condense the data** – sometimes the data collected for statistical manipulations are wide and raw. In order to make decisions based on the data, it is crucial to make the data more comprehensive. This can be done with the help of tabulation.

Size groups	No. of farmers	Knowledge level	Adoption level
Small	30	Moderate	high
Medium	15	high	moderate
Large	10	moderate	moderate
Total	55	moderate	Moderate

3. Treatment of data –It becomes difficult to treat raw and unclassified data and is hence important to classify the data before doing so. Classification of data helps facilitate the statistical treatment of the data e.g for working out correlation between adoption and knowledge of different size group of farmers we required data of all 55 farmers regarding knowledge and adoption and rearrange the same according to size of holding in three

4. Comparisons – wide, raw and unclassified data is impossible to deal with and arrive at any conclusion. Conclusions cannot be arrived at without treating the data and making a statistical analysis. Hence, classified/organized/tabulated data enables analysts to make meaningful comparisons on various criteria e.g. small farmers per ha income is higher than large farmers income.

Rules For Classifying Data:

It is all the more important to remember the rules of classifying the data. These rules form the backbone and act as guiding principles for well-classified data. These rules are mentioned below:

Unambiguous – the classes should be rigid and unambiguous (clear). An unclear classification can have severe consequences and can also impact all further statistical treatments small size of holding upto 2ha

Exhaustive – every classified data must be exhaustive in the sense that they should belong to one of the classes or categories classification according to age as young (up to 30 years), middle (31 to 50 years) and old (above 50 years)

Stability – in order to facilitate effective comparisons of data, it is important that the classified data are stable, in the sense that the same classification pattern must be adopted throughout the analysis. Adopting different classification techniques for the same analysis would lead to ambiguity e.g. irrigated and semi-irrigated.

Suitable for the purpose – it is crucial to remember the objective of the report or analysis while classifying data. Avoid classifying the data in a manner that does not suit the purpose of the inquiry.

Flexibility – it is important to classify data in a manner that allows future modification. Due to changing conditions, there may arise the need to change the statistical methods and data classifications. In such a situation, a flexible classification of data would solve many issues.

Problems With Classifying Data:

Classification of data has many functions and various benefits. But there are also some key issues in organizing data. The most important problems associated with it are mentioned below:

- Organizing data can be a very tedious and complex task for many situations or individuals.
- Classifying data is a purely experience based and a non-experience action that can lead to mis-judgements. These mis-judgements can often cause a lot of inconvenience and errors.
- Redoing the entire process of classification can be very time consuming.
- Classifying data can be done only with the help of a statistical analyst.
- It is impossible to classify data without having moderate knowledge on the same.

Organization of Data:

Data can be organized into four broad bases:

Chronological Classification – The chronological classification of data emphasizes the occurrence of time. Under this type of data classification, data is classified on the bases of differences in time. The time series data (used frequently in economic and business statistics) is an example of data being classified in a chronological manner.

Geographical Classification – The geographical organization of data emphasizes on the geographical representation of data. Dta is classified on the basis of geographical boundaries and location differences. Classifying based on states, cities and districts is a geographical classification. Classifying based on countries and continents are also examples of data being classified in a geographical manner.

Qualitative Classification – The qualitative classification of the data emphasizes on certain qualitative phenomenon of the data. Under this type of data classification, data is classified on the basis of qualitative measurements. Classifying based on qualities like honesty, intelligence and also aptitude are some examples of data being classified in a qualitative manner.

Quantitative Classification – The quantitative classification of the data emphasizes on certain quantitative phenomenon of the data. Under this type of data classification, data is classified on the basis of quantitative measurements. Classifying based on quantities like size, sales, profits, age, height and also weight are some examples of data being classified in a quantitative manner.

DATA

Something assumed as facts and made the basis of reasoning or calculation.

1. **Qualitative or Categorical**

   Sex, Color, Race, caste etc

2. **Quantitative or Numerical**

   Age, Height, land holding etc

Categorical Data

- **Nominal: categories of data cannot be ordered one above the other.**

  Sex: Male, Female

  Marital Status: Single, Married, Divorced.

- **Ordinal: Categories of the data can be ordered one above the other or vice versa.**

  Level of Knowledge: Good, Average, Poor

  Opinion: Fully Agree, Agree, Disagree.

Data Coding

Data coding is the process of driving codes from the observed **data**. In qualitative research the **data** is either obtained from observations, interviews or from questionnaires. The purpose of **data coding** is to bring out the essence and meaning of the **data** that respondents have provided.

Example: Knowledge about biofertilizer

1. Perfect knowledge code 5
2. Moderate knowledge code 4
3. Partial knowledge code 3
4. Poor knowledge code 2
5. No knowledge code 1

The codes are given meaningful names or numbers and they are put in categories. These categories help refine the research a lot. When data is coded again and again, it get refined. The refined data itself leads to patterns and themes. The patterns are the key to find out the true results of the research. These patterns or categories determine where does the large amount of the data inclines.

Coded as name

Jabalpur-JBP

Nagpur -NGP

Validation of data

- Data validation is a method for checking the accuracy and quality of your data, typically performed prior to importing and processing.
- Data validation ensures that your data is complete (no blank or null values), unique (contains distinct values that are not duplicated), and the range of values is consistent with what you expect.
- Data validation helps ensure that when you perform analysis, your results are accurate.

Steps to data validation

Step 1: Determine data sample: Determine the data to sample. If you have a large volume of data, you will probably want to validate a sample of your data rather than the entire set. You'll need to decide what volume of data to sample, and what error rate is acceptable to ensure the success of your project.

Step 2: Validate the database: Before you move your data, you need to ensure that all the required data is present in your existing database. Determine the number of records and unique IDs, and compare the source and target data fields.

Step 3: Validate the data format: Determine the overall health of the data and the changes that will be required of the source data to match the schema in the target. Then search for incomplete counts, duplicate data, incorrect formats, and null field values.

Variable

An item of data that can be observed or measure

Quantitative Variable

A variable that has a numerical value

e.g. Age, Schooling years, No. of Children etc.

Qualitative Variable

A variable that is not characterized by a numerical value.

e.g. Sex, Category of farmer, level of education etc.

Quantitative Variable

Discrete Variable

A quantitative variable, whose possible values are in whole numbers.

Example: Number of farmers, number of students

Continuous Variable

A quantitative variable that has an uninterrupted range of values

Example: size of holding, Weight, height etc

Types of Variables

Independent Variable

A variable, whose effect is being measured. (Cause) e.g. rainfall

Dependent Variable

The variable, on whom the effect is being observed. (Effect) e.g. farming income

Confounding Variable

A variable, which affects both independent as well as dependent variable. (cause as well as Effect) e.g. productivity of crops

Data Collection Techniques

- Using available information
- Observing
- Interviewing
- Administering questionnaire
- Focus Group Discussion (FGD)
- Nominal Group Technique (NGT)
- The Delphi Technique

Hypothesis

Idea/ suggestion put forward as a starting point for reasoning or explanation

Hypothetical

Not based on certain knowledge

Hypothesis is a supposition, which is tested by collecting facts. The analysis of these facts leads to its acceptance or rejection.

Cont……

A researcher wants to compare the farming in two regions.

There can be two hypothesis.

- **There is no difference between farming of the two areas being studies (Null Hypothesis)**
- **There is difference between farming of two areas (Alternative Hypothesis)**

Hypothesis Testing

- Student's "t" Test
- Chi – Square Test
- Correlation
- Regression
- F-test

Presentation of Data

1. Tables
2. Graphs

- Bar
  -Vertical
  -Horizontal
- Histogram
- Line Graph
- Scatter
- Pie Chart
- Column

Types of tables

One way Tables

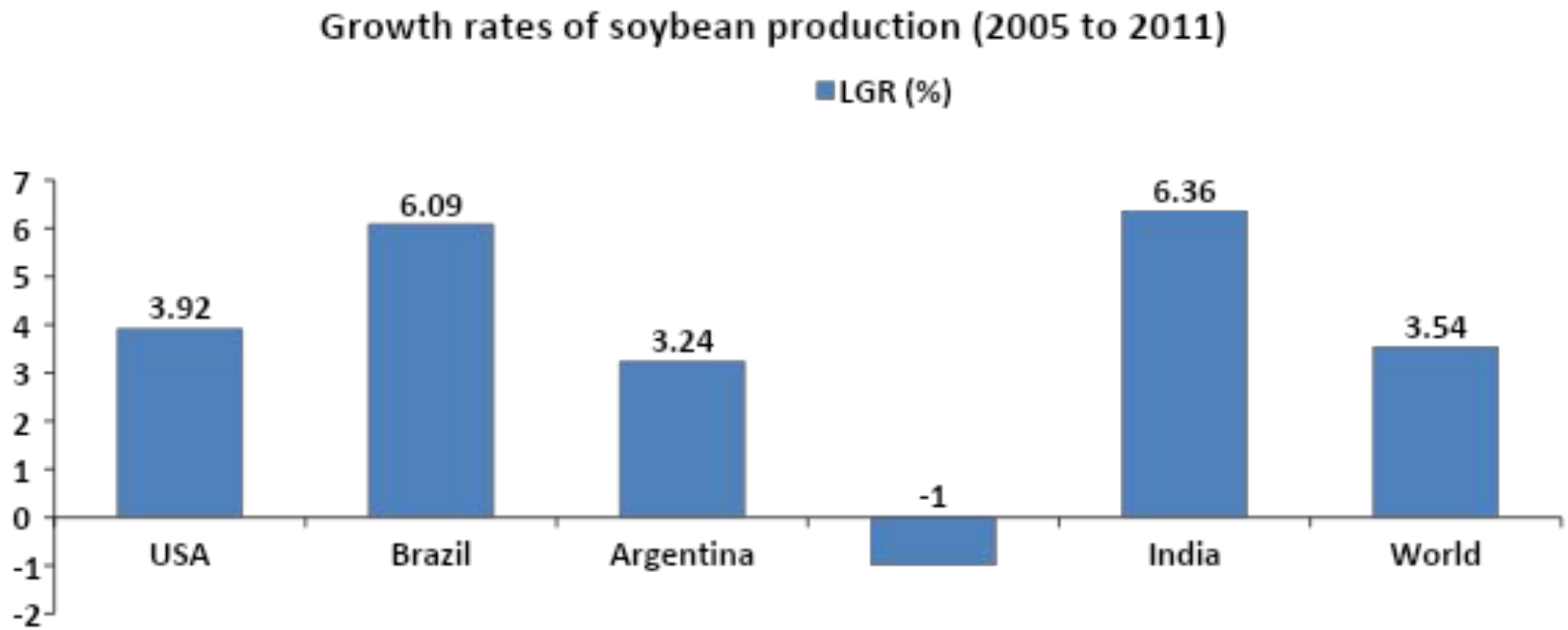
Method of planting	Yield (kg/ha)	Percentage increase
Farmers practice	1349	----
Ridge-furrow planting system	1732	28.4

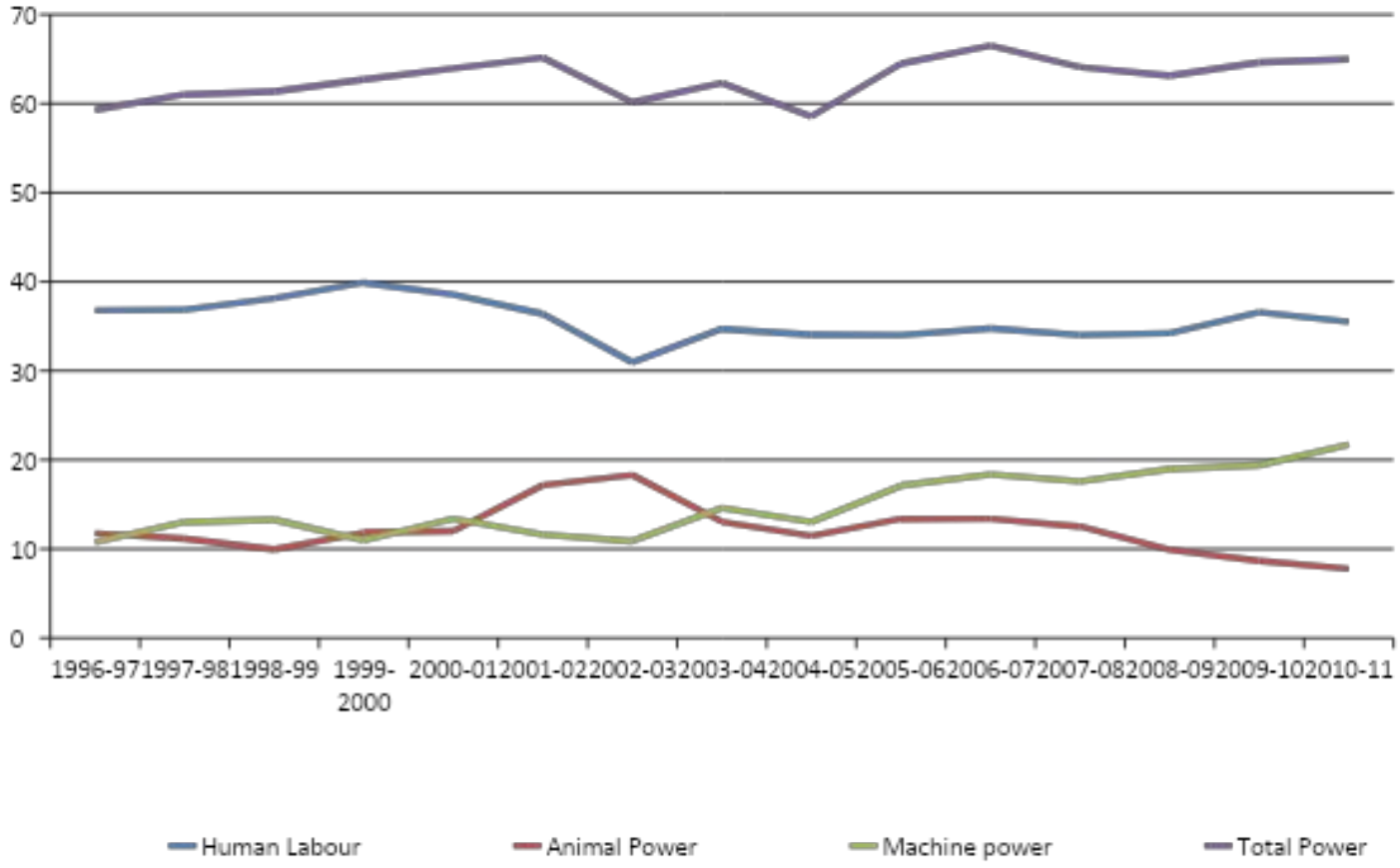
Two way Tables

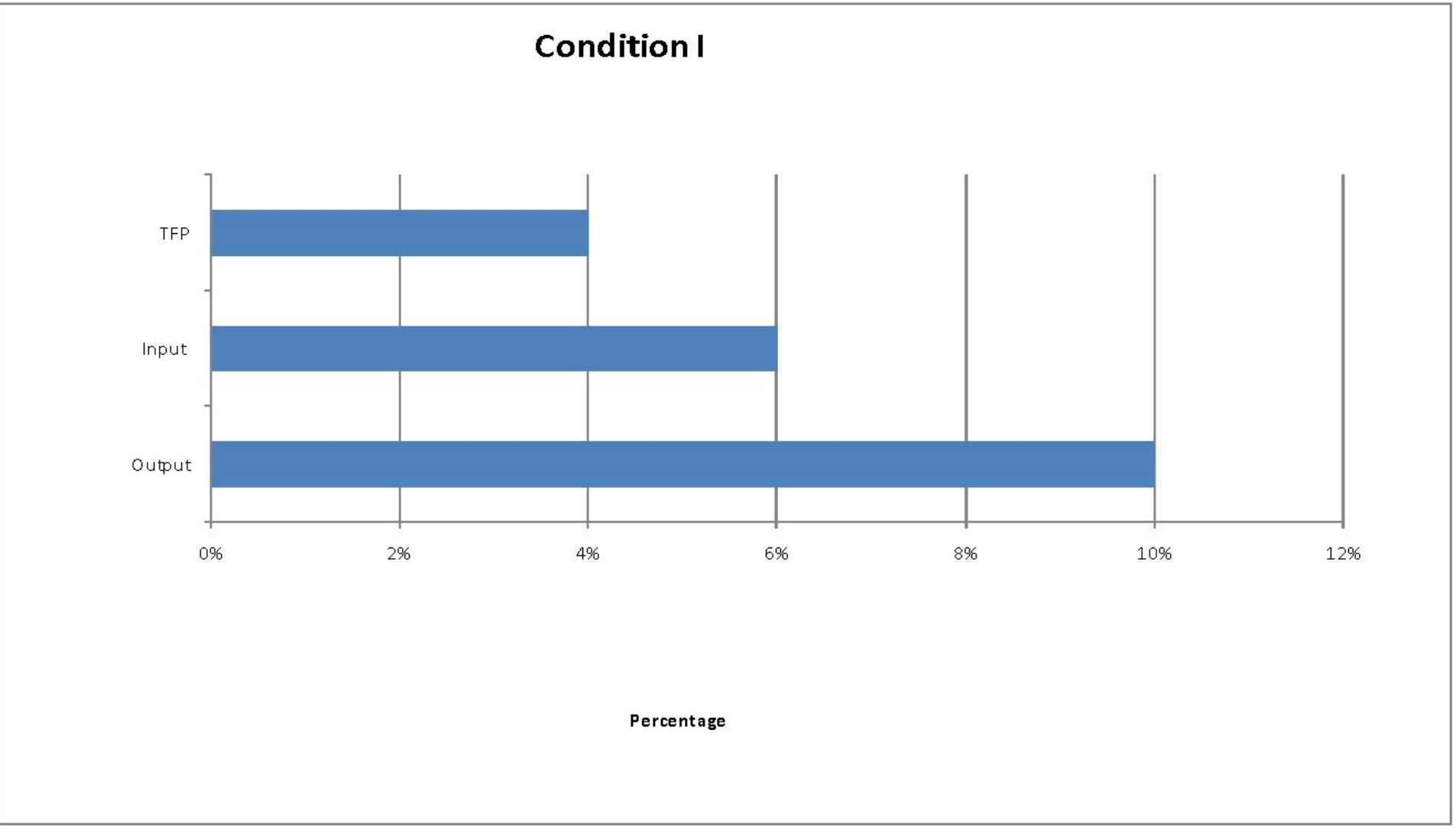
CROPS	2007-08			2008-09			2009-10		
	AREA	PROD.	YIELD	AREA	PROD	YIELD	AREA	PRO	YIELD
TUR	3.04	1.97	643	3.10	2.47	795	3.29	2.57	780
URD	5.70	1.95	341	5.17	1.96	379	5.70	2.14	375

Types of Graphs

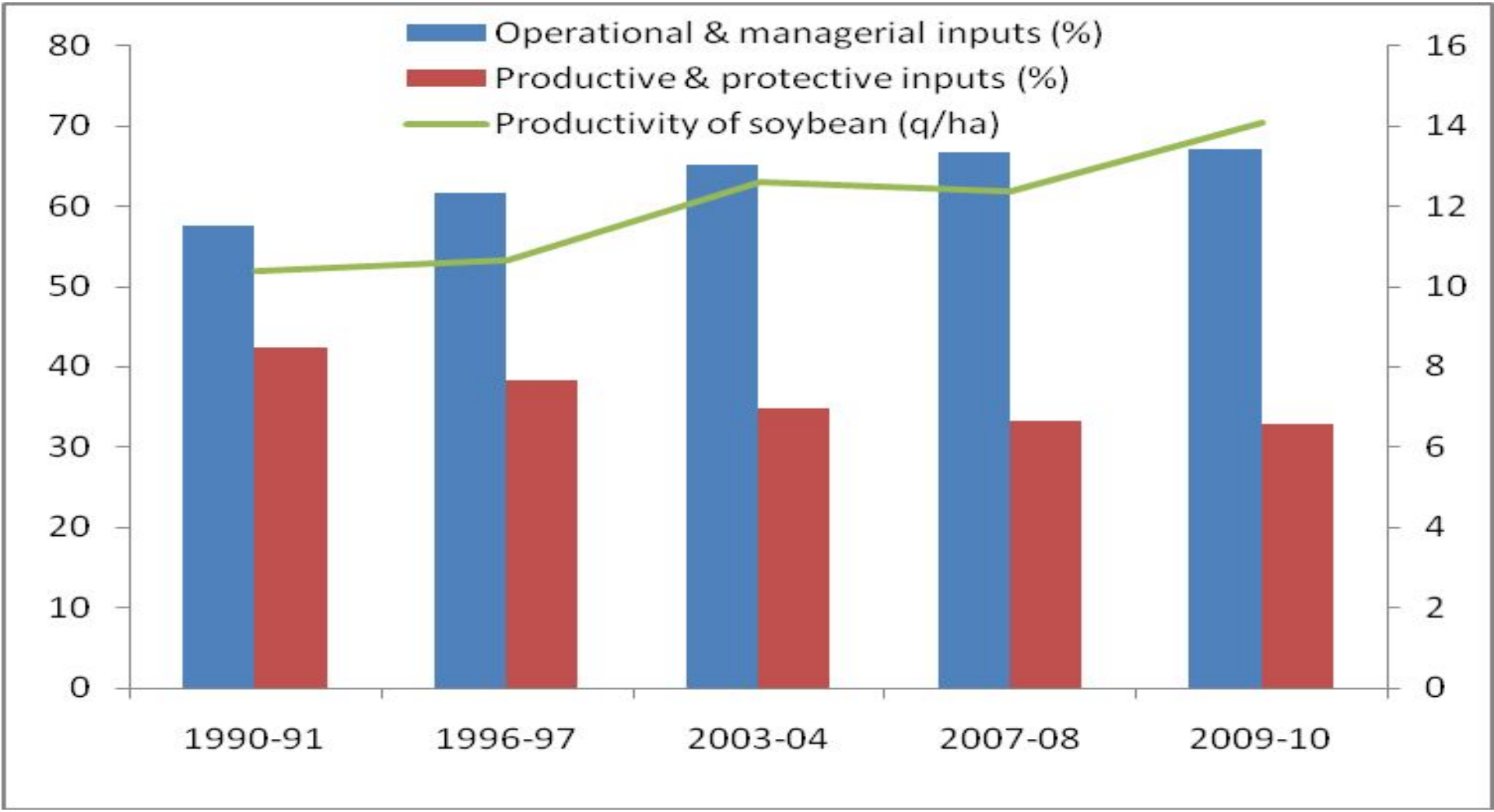
Horizontal bar diagram



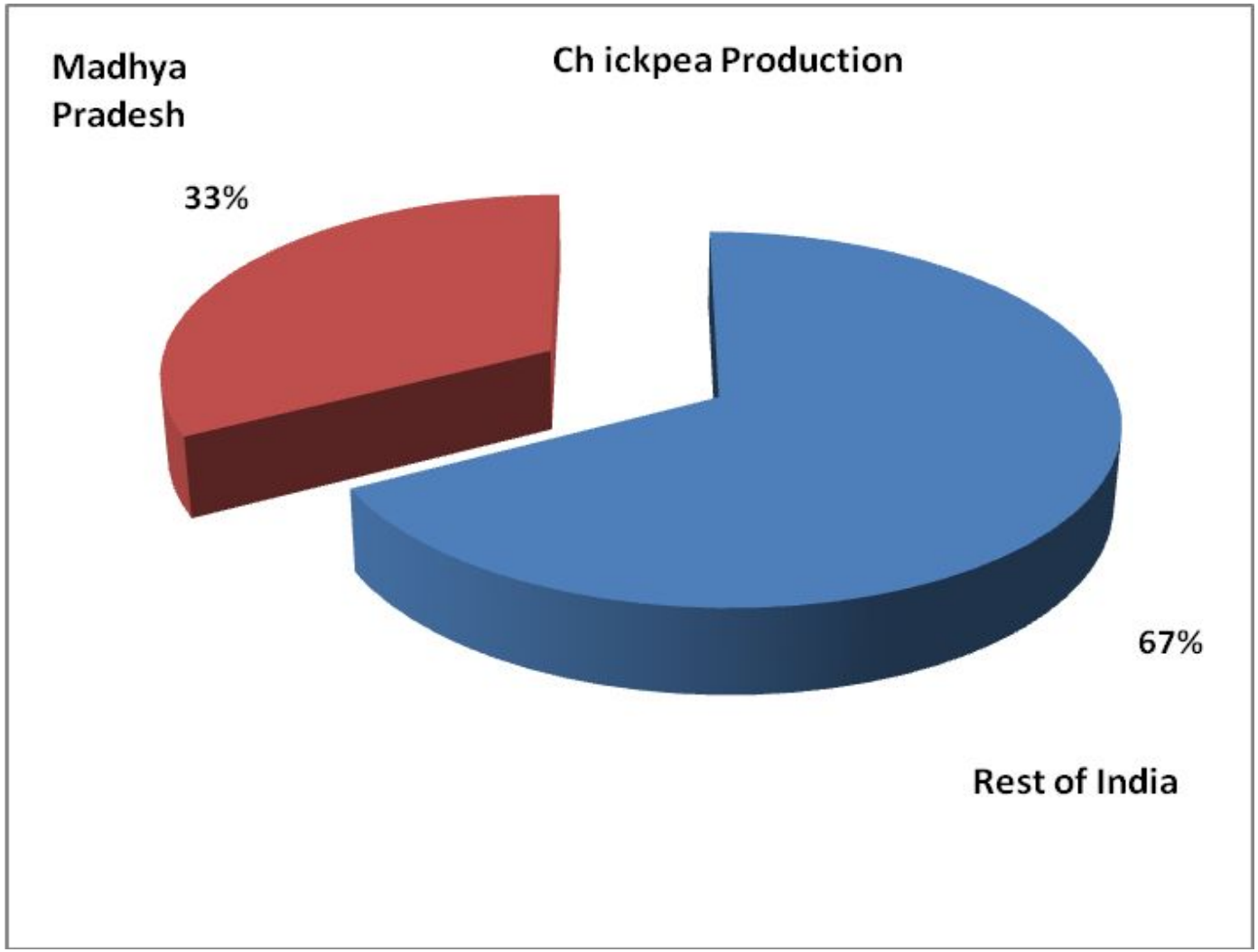

70
60
50
40
30
20
10
0
1996-97
1997-98
1998-99
1999-2000
2000-01
2001-02
2002-03
2003-04
2004-05
2005-06
2006-07
2007-08
2008-09
2009-10
2010-11
Human Labour
Animal Power
Machine power
Total Power


Condition I
TFP
Input
Output
0%
2%
4%
6%
8%
10%
12%
Percentage

Compound graphs



Pie Diagram



Results

- Present your findings in a logical sequence
- Use tables and graphs to summarize data.
- Mention negative results of interest.
- Give statistical significance.
- Estimate accuracy and precision of results.
- Avoid vague statements.

Results

- Present the findings in simple, standard, scientific language.
- Avoid abbreviations where ever possible.
- Tables should be self contained.
- The findings presented in the tables should not differ from those given in the text.
- As far as possible, do not repeat in the text, what has been given in the tables.

Discussion

- How are the findings different from other studies?
- What is the statistical significance of these difference?
- What are the probable reasons for these difference?
- What are the policy implications of these findings?

Research Methodology for Social Sciences & Business Management

(AGECON 506 & ABM 511)

Sunil Bhaskar Rao Nahatkar
DIRECTOR
INSTITUTE OF AGRIBUSINESS MANAGEMENT

(updated on 13.04.2020)

J. N. Krishi Vishwa Vidyalaya, Jabalpur 482004 (M.P.)

Course Content (AGECON 506 (1+1)

Unit-I

- Importance & scope of research in agricultural economics.
- Types of research- Fundamental vs Applied.
- Concept of researchable problems.
- Research prioritization.
- Selection of research problems
- Approach to research-research process

Unit-II

- Hypothesis-meaning, characteristics ,types
- Review of literature
- Setting of course objectives and hypothesis
- Testing of hypothesis

Unit-III

- Sampling theory and sampling design
- Sampling errors
- Method of sampling-probability and non-probability sampling methods and criterions for selection.
- Project proposal-content and scope
- Different types of project to meet different needs
- Trade off between scope and cost of the study
- Research designs and techniques-types of research designs.

Unit-IV

- Data collection-assessment of data needs-sources of data collection- discussion on different situations.
- Mailed questionnaires and interview schedule
- Structured and unstructured-open and close ended questions.
- Scaling techniques
- Preparation of schedule.
- Problems in measurement of variables in agriculture
- Interviewing techniques and field problems.
- Methods of conducting survey.
- Reconnaissance survey and pre-testing

Unit-V

- Coding editing-tabulation-validation of data.
- Tools of analysis-data processing
- Interpretation of results
- Preparing research report/thesis
- Universal procedure for preparation of bibliography.
- Writing of research articles.

Practical

- Exercise on problem identification
- Project proposal-content and scope
- Formulation of objectives and hypothesis
- Assessment of data need-sources of data
- Method of data collection
- Method of sampling & criterion for choosing best method
- Discussion on sampling under different situations
- Scaling techniques-measurement of scales.

Practical

- Preparation of interview schedule-field testing
- Method of conducting survey
- Exercise on coding, editing, tabulation and validation of data
- Preparing for data entry into computer
- Hypothesis testing-parametric and non parametric tests.
- Exercises on format for thesis/report writing.
- Preparation of results

Course Content (ABM 511 (3+0))

Unit-I

- Meaning, course objectives, types and process of research
- Research methodology in management- Exploratory, descriptive, experimental, diagnostic.
- Problem formulation, setting of course objectives, formulation of hypothesis.

Unit-II

- Scales of measurement- Nominal, ordinal, interval, ratio, Likert scale and other scales.
- Primary and secondary data.
- Sources of data, instruments of data collection
- Data editing, classification, coding, validation, tabulation, presentation and analysis.

Unit-III

- Concept of sampling, probability and non-probability sampling techniques.
- Simple random sampling, stratified sampling, multi-stage sampling, systematic sampling, purposive sampling, quota sampling, judgment sampling and convenience sampling.
- Sampling size determination and sampling and non sampling errors.

Unit-IV

- Role and uses of quantitative techniques in business decision making.
- Use of equations, use of determinants and matrices in business decisions.
- Frequency distribution, measures of central tendency, measures of variation,
- Skewness and kurtosis.
- Simple, partial and multiple correlation, rank correlation.
- Simple and multiple regression.
- Discriminate and dummy variable analysis.

Unit-V

- Index number , hypothesis testing, ANOVA, factor analysis, cluster analysis, conjoint analysis, multidimensional analysis etc.
- Report writing, types of reports, essentials and contents of good report writing.

Research designs

A **research design** is the set of methods and procedures used in collecting and analyzing measures of the variables specified in the problem **research**.

A **research design** is a framework that has been created to find answers to **research** questions.

Research design is defined as a framework of methods and techniques chosen by a researcher to combine various components of research in a reasonably logical manner so that the research problem is efficiently handled.

It provides insights about “how” to conduct research using a particular methodology. Every researcher has a list of research questions which need to be assessed – this can be done with research design.

Essential Elements of Research Design

1. Accurate purpose statement of research design
2. Techniques to be implemented for collecting details for research
3. Method applied for analyzing collected details
4. Type of research methodology
5. Probable objections for research
6. Settings for research study
7. Timeline
8. Measurement of analysis

Characteristics Research Design

There are four key characteristics of research design:

Neutrality: The results projected in research design should be free from bias. Understand opinions about the final evaluated scores and conclusion from multiple individuals and consider those who agree with the derived results.

Reliability: If a research is conducted on a regular basis, the researcher involved expects similar results to be calculated every time. Research design should indicate how the research questions can be formed to ensure the standard of obtained results.

Validity: Valid measuring tools are those which help a researcher in obtaining results according to the objective.

Generalization: Should be applicable to a population and not just a restricted sample.

Types of Research Designs.

Exploratory Research: Just as the word implies, it explores, that is to find out about something by answering the question in “what” or “How” manner.

Descriptive Research: This is more in-depth research, that answered the question what and how.

Explanatory Research: This seeks to explain the subject matter being researched and tries to answer the question what, how and why.

Evaluation Research: This is quite extensive as it measures the effectiveness of a program.

Quantitative Research design

A *quantitative research design* is used to examine the relationship between variable by using numbers and statistics to explain and analyze its findings and there are four types of quantitative research design:

Descriptive design research: As the name implies, it is intended to describe the present status. This type of design does not require a hypothesis to begin with. These analyses are generated from existing data educational status of farmers in district Jabalpur.

Correlational design research: This seeks to discover If two variables are associated or related in some way, using statistical analysis, while observing the variable e.g. relationship between yield and irrigated area.

Experimental design research: This is a method used to establish a cause and effect relationship between two variables or among a group of variables. The independent variable is manipulated to observe the effect on the depended variable. For example, a certain group is exposed to a variable and then compared with the group not exposed to the variable. e.g Beneficiaries and non-beneficiaries of Watershed program

Quasi-experimental design research: This experiment is designed just like the true experimental design, except that it does not use randomized sample groups. Also, it is used when a typical research design is not practicable.

Qualitative Research Design

Qualitative research design, on the other hand, is exploratory in nature as it tries to explore not predict the outcome. It seeks to answer the questions what and how. A *qualitative research design* is used to explore the meaning and understanding of complex social environments, like the nature of people's experience, using case studies. Similarities with quantitative research:

- An outline question stating the problem that needs to be solved.
- Has a set order and procedure used to answer these questions?
- Analyses the data generated.
- Draws its conclusion after the data has been collated and analyzed so that the conclusion drawn from the findings are not predetermined.

Besides the similarities identified above, a *qualitative research design* also intends to understand, describe or discover the findings.

- The researcher is usually the primary instrument that formulates the question and interprets the meaning of a data.
- The data used are mostly documented words from interview, newspapers videos etc.
- More than one type of data is collected during this research, from the field, where the participants are. In other words, the research goes beyond the intended scope, so making it emergent because the method of research changes and different types of data might be collected as the research goes on.

Universal Procedure for Preparation of Bibliography.

A **bibliography** is a list of all of the sources you have used in researching your work. B**ibliography** should include: the authors' names. the titles of the works. the names and locations of the companies that published your copies of the sources.

e.g. for **Book chapter**

Nahatkar, S.B. (2014). An Assessment of Economic Feasibility of Agricultural Technologies, in *Participatory Approaches for Transfer of Agricultural Technology* (eds M. M. Patel, S. K. Badodiya and Shobhana Gupta), Biotech Books, New Delhi.

Published Book:

Khare, Direndra, Nahatkar, S.B., Shrivastava, A.K. & Jha, Amit K. (2018). Farm Mechanization for Production, Scientific Publishers (India), New Delhi, pp 238 (ISBN 978-93-87893-80-1).

Paper/Abstract presented in Seminar/Symposium/ Conference

Nahatkar, S.B. (2014). Dynamics of Profitability from soybean in Central India, paper presented in International Soybean Research Conference (SOYCON-2014) on Mitigating Productivity Constraints in Soybean for Sustainable Agriculture, organized by Society for Soybean Research and Development, Indore, Directorate of Soybean Research and Indian Council of Soybean Research, 22-24 February, 2014 at Indore, Madhya Pradesh.

Technical Bulletin

Nahatkar, S.B., Sharma, Hari Om and Rathi, Deepak (2018). Impact of Tejaswani Rural Women Empowerment Program on Empowerment of Rural Women Through SRI in Mandla and Balaghat districts of Madhya Pradesh, Ad-hoc Research Study No.127, Agro-Economic Research Center, JNKVV, Jabalpur, 31p.

Popular articles

Nahatkar, Sunil (1988). Augmenting facilities for marketing of Agricultural Products, *M. P. Chronicle,* page 7, dated Oct 20.

Research Paper

Nahatkar, Sunil (2017). Production Trend and Cost –Profitability structure of Chickpea in Central India: The Dynamic Scenario, *Agricultural Situation in India*, 73 (10):30-39.

How to arrange bibliography

1. Bibliography is always given at the end of research paper/ report/ thesis/ book/ Book chapter.
2. It is arranged alphabetical order according to surname of first author.
3. For same authors if there are two papers it is arranged in chronological order of year of publication.
4. Foe citation without name is known as Anonymous (written buy unknown author)

Example of writing bibliography

1. Gautam, D. S., Nahatkar, S. B. and Rajput, R. L. (1992). Economic Analysis of Major Kharif Oilseed crops in Narmada Valley: Soybean and Sesame, *Bhartiya Krishi Anusandhan Patrika*, 7 (4): 285-290.
2. Gautam, D .S. and Nahatkar, S. B. (1993). Profit Structure of Soybean, Cotton and Sorghum Cultivation in Central Narmada Valley of India, *Gujarat Agriculture University Research Journal*, 19 (1): 98-102.
3. Jain, S .K., Sharma, H. O., and Nahatkar, S. B (1991). Price Spread and Marketing Efficiency of Soybean in Sehore District of M.P., *Indian Journal of Agricultural Marketing*, 5 (1):106-107.
4. Kumar, Santosh, Rathi, Deepak, Nahatkar, S.B. and Masuda, Tadayoshi (2015). Constraints in adoption of soybean production technologies in Northern Hill Region of Chhattisgarh Agro-Climatic zone of Madhya Pradesh, *Economic Affairs*, 60(4): 769-775.
5. Nahatkar, Sunil Bhaskar Rao (2014). Dynamics of profitability from soybean in central India, *Soybean Research* (special issue):12 (1): 144-155.
6. Okamoto, H., Kobayashi, S., Tsuji, K. Tomar, S. S. Nahatkar. S. B. and Sharma, S.K. (2014). What is the most effective technology to improve soybean yield in Madhya Pradesh, India? *Jpn. J. Crop Sci*. 83 (1): 80-81.

Twenty steps of Writing a Research Articles

1. **Determine the authors**. When designing a research paper, we recommend preparing an initial list & order of authors. Such a list authors should be based on established guidelines and should make explicit the estimated contribution of each individual to the project.
2. **Start writing before the experiments are complete.** Start writing while you are still doing the experiments. Writing often evokes new ideas.
3. **Decide about time to publish**. It is time to publish when your findings represent a complete story, one that will make a significant contribution to the scientific literature.

4. Draft a title & abstract. Drafting a working title and an abstract helps define the contents of the paper.

5. Reexamine the list of authors. Reevaluate list based on the contributions that were made to those experiments and the additional contributions that will be made through the preparation of the manuscript. If a list already exists, make adjustments to ensure compliance with your guidelines.

6. Determine the basic format. There are three basic formats for peer-reviewed research articles:

• *Full-length research articles*: These articles contain a comprehensive investigation of the subject matter and are viewed as the standard format. It uses the "IMRAD" format: Introduction, Methods, Results and Discussion.

• *Short communications*: While not as comprehensive in scope as full-length research articles, these papers also make a significant contribution to the literature. Their length will be set by the journal but is usually 3500 words or less and will contain up to 2 tables and figures. Unlike full papers, methods, results, and discussions may be combined into a single section.

7. Select the journal. There are several factors to consider when choosing a journal. It is unlikely that one journal will have all the features you are looking for, so you may have to compromise. However, there is one essential feature you should not compromise on – *manuscripts must be peer reviewed for publication if they are to be considered research articles*.

Language: English has become the dominant form for international scientific communication. Thus, if you are interested in communicating your results widely to the international scientific community, then it is essential to publish in English

Focus: What type of research does the journal publish? Is its focus broad or narrow? Which disciplines are represented? What is the journal's orientation – for example, is it related to agriculture, agri-business or agro-industries?

Indexing: Is the journal indexed in the major electronic databases such as NAAS rating, SJR ranking etc.

Availability: Is the journal broadly available? Is there an online version of the journal? Are papers provided in PDF format?

Reputation: Look at recent articles and judge their importance. Determine the journal's impact factor (an annual measure of the extent to which articles in a given journal are cited. Try to find out the acceptance rate of the journal.

Format: Do you like the appearance of published articles – the format, typeface, and style used in citing references? If relevant, does the journal publish short communications?

Time to Print: Using the "date submitted" and a "date accepted" that are published on the article, along with the date of the issue, you can estimate the length of the review process as well as the time from acceptance to publication in print.

Charges: Some journals bill the author for page charges, a cost per final printed page.

8. Stock the sections of your paper. As you think about your paper, store relevant material in folders marked Introduction, Methods, Results, and Discussion. This will save time and avoid frustration when the writing begins. Stored items might include figures, references, and ideas.

9. Construct the tables, figures, and legends. Create figures and tables before the writing begins. The entire paper should be organized around the data you will present. By preparing the tables and figures (and their legends and appropriate statistical analyses), you will be certain of your results before you worry too much about their interpretation.

10. Outline the paper. An outline is like a road map. And if you have "stocked" your sections, those files should be useful here and in the writing that follows.

11. Write the first draft. Write the first draft of the entire manuscript. If you are writing with coauthors, you may wish to assign different aspects of the manuscript to different authors. At this point, do not worry about quality. That comes later. Some people recommend that you begin your writing with the Introduction and continue through in order each section of the paper. This can help ensure flow.

12. Revise the manuscript. This step involves three major tasks, each to be carried out in the order given:

Make major alterations: Fill in gaps, correct flaws in logic, restructure the document to present the material in the most logical order.

Polish the style: Refine the text, then correct grammar and spelling.

Format the document: Make your manuscript attractive and easy to read. It is important to do the tasks in the stated order. Otherwise, you may find yourself spending a lot of time revising material that you later delete.

13. Check the references. Ensure that the citations are correct and complete. Do one last literature search to make certain that you are up to date.

14. Write the final title and abstract. Many changes are made during the editing process. Make certain that your title and abstract match the final version of your article.

15. Reread the journal's Instructions to Authors. Review the details of how the manuscript is to be formatted and submitted. Revise where necessary.

16. Prepare the final illustrations. Ensure that your tables, figures, and figure legends are complete, clear, self-contained, and in the format required by the journal. Do not allow any chance for misunderstanding.

17. Get feedback on your manuscript and then revise your manuscript again. Getting feedback is one of the most important things that you can do to improve your article. First, be sure your co-authors have had a chance to read and comment on the draft. Then, when it is ready, give the manuscript to some colleagues. Indicate when you would like to receive their comments, and what levels of information you would like. After you get their comments, revise your manuscript to address their concerns.

Do not submit your manuscript until you feel it is ready for publication. Once it is accepted, further changes in your manuscript will be difficult and may also be costly.

18. Submit the manuscript to the editor. Follow the Instructions to Authors to determine what items you need to submit, how to submit them, and to whom you should send them. You may not submit your manuscript to more than one journal at a time!

19. Deal with reviewers' comments. Most manuscripts are not accepted on the first submission. However, you may well be invited to resubmit a revised manuscript. If you choose to do so, you will need to respond to the reviewer comments. Do this with tact. Answer every concern of the reviewers, and indicate where the corresponding changes were made in the manuscript if they were, indeed, made. You do not need to make all of the changes that the reviewer recommended, but you do need to provide a convincing rationale for any changes that you did not make. When you resubmit the manuscript, indicate in your cover letter that this is a revised version.

20. Check the proofs. Once the manuscript is accepted and prepared for print, the publisher will send the corresponding author page proofs of the article. This may be accompanied by a list of queries, such as missing information regarding a reference. The proofs may be sent via e-mail or as hard copy. Carefully correct any typos and factual errors. And read the manuscript for clarity – this is your last chance!

Writing Review of Literature

- A literature review discusses published information in a particular subject area, and sometimes information in a particular subject area within a certain time period.
- A literature review can be just a simple summary of the sources, but it usually has an organizational pattern and combines both summary and synthesis.
- A summary is a recap of the important information of the source, but a synthesis is a re-organization, or a reshuffling, of that information.
- It might give a new interpretation of old material or combine new with old interpretations. Or it might trace the intellectual progression of the field, including major debates.
- And depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant.

Arrangement of Review of Literature

Chronological: Review follows the chronological method, you could write about the materials according to when they were published.

e.g. review of 1995 will come first than 1996.

Within year follow alphabetical order e.g. Nahatkar, et al. (1995) will come first than Sharma, et al (1995) Why et al,

When there are more than two authors use et al while writing review of literature

e.g. Okamoto, H., Kobayashi, S., Tsuji, K. Tomar, S. S. Nahatkar. S. B. and Sharma, S.K. (2014). What is the most effective technology to improve soybean yield in Madhya Pradesh, India? *Jpn. J. Crop Sci*. 83 (1): 80-81.

Okamoto et al. (2014)…….. (while writing review)

Gautam, D .S. and Nahatkar, S. B. (1993). Profit Structure of Soybean, Cotton and Sorghum Cultivation in Central Narmada Valley of India, *Gujarat Agriculture University Research Journal*, 19 (1): 98-102.

Gautam & Nahatkar (1993)…… (while writing review)

Nahatkar, Sunil Bhaskar Rao (2014). Dynamics of profitability from soybean in central India, *Soybean Research* (special issue):12 (1): 144-155.

Nahatkar (2014)…….. (while writing review)

Research Methodology

- Importance & scope of research
- Approach & types to research-research process
- Types of research- Fundamental vs Applied.
- Exploratory, descriptive, experimental, diagnostic.
- Concept of researchable problems
- Selection of research problems
- Problem formulation, setting of course objectives, formulation of hypothesis.
- Research prioritization

Science and the Scientific Method

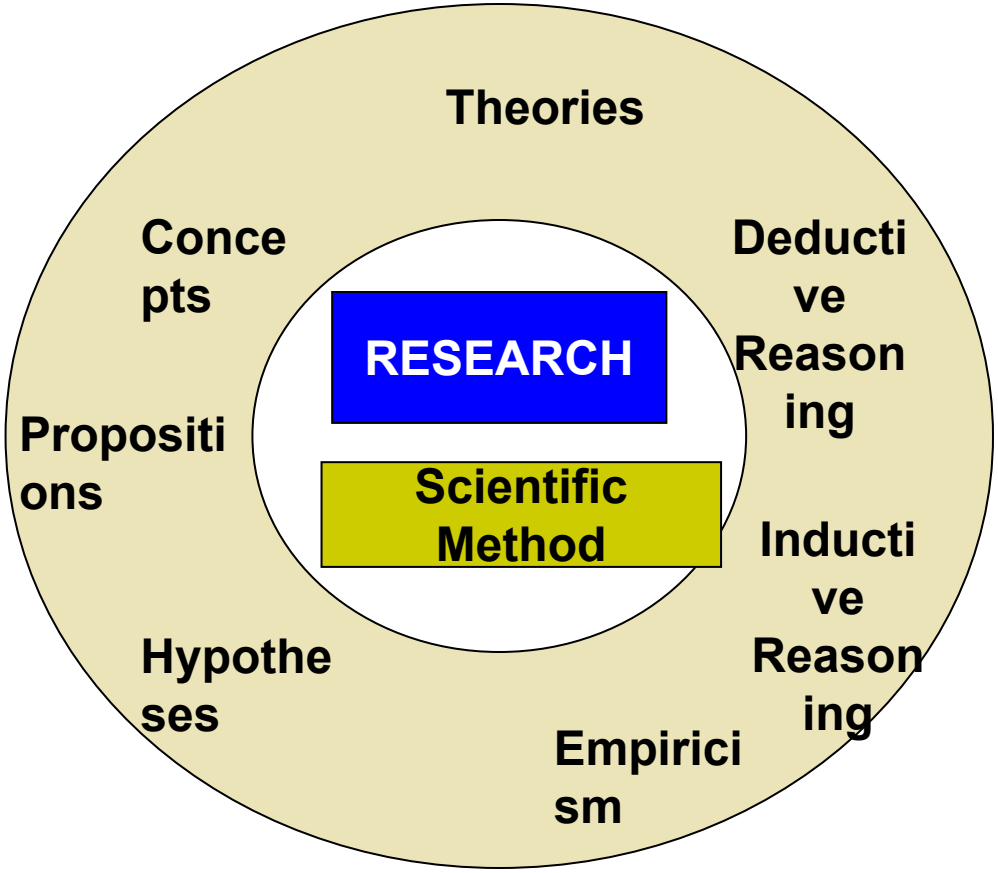
Science has been defined as "the methodological and systematic approach to acquisition of new knowledge" (Geoffrey Marcyzk, David DeMatteo, David Festinger, *Essentials of Research Design and Methodology*, John Wiley & Sons, 2005, p. 4)

The scientific method, which has evolved since the 13th century, concerns the set of tools, techniques and procedures used by basic and applied researchers to analyze and understand phenomena and prove or disprove prior conceptions

Research

- **An investigation undertaken in order to discover new facts or to generate additional information.**
- **Research is systematic collection, analysis and interpretation of data to answer a specific question or solve a problem (e.g. Study on yield gap in soybean).**
- ***Research is the systematic approach to obtaining and confirming new and reliable knowledge"***
  - **Systematic and orderly (following a series of steps)**
  - **Purpose is <u>new</u> knowledge, which must be reliable.**

The Dimensions of Research



Inductive Reasoning is an approach to logical thinking that involves making generalizations based on specific details e.g. Every chicken we've seen has been brown. All chickens in this area must be brown.(specific to generalize or bottom up logic)

Deductive Reasoning relies on a general statement or hypothesis—sometimes called a premise or standard—held to be true. e.g. if/then statement. If A = B and B = C, then deductive reasoning tells us that A = C..(generalize to specific or top-down logic)

Research is…

1. **<u>Searching for explanation of events, phenomena, relationships and causes</u>**
   - **What, how and why things occur**
   - **Are there interactions?**
2. **<u>A process</u>**
   - **Planned and managed – to make the information generated credible**
   - **The process is creative**
   - **It is circular – always leads to more questions**

Research

Research = creation of knowledge

How to create knowledge

- Construct an argument
- Argument must be a position or point of view on a question
- Argument must be supported by evidence

What is Research?

A systematic means of problem solving (Tuckman 1978)

5 key characteristics:

1. Systematic – research process
2. Logical – induction/deduction
3. Empirical – evidence based
4. Reductive – generalisation
5. Replicable – methodology.

What is Research?

There are many accepted definitions for the term "research", for example:

Research is an active, diligent and systematic process of inquiry in order to discover, interpret or revise facts, events, behaviours, or theories, or to make practical applications with the help of such facts, laws or theories. The term 'research' is also used to describe the collection of information about a particular subject

(Encyclopedia Wikipedia)

What is Research?

Other insightful definitions of "research" are:

- Systematic, intensive, patient study and investigation in some field of knowledge, usually employing the techniques of hypothesis and experiment, whose purpose is to reveal new facts, theories, or principles
- Means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge
- Systematic study directed toward more complete scientific knowledge or understanding of the subject studied

Purposes of Research

Exploration

gaining some familiarity with a topic, discovering some of its main dimensions, and possibly planning more structured research

Description

- Census report on population of India
- Political poll predicting who will win an election. Besides this this will help in
- Validate perception
- Improve methods
- Demands of the Job
- For publication

Classification of Research

Basic or fundamental – to determine or establish fundamental facts and relationships within a discipline or field of study. Develop theories … (e.g. law of demand?)

Applied – undertaken specifically for the purpose of obtaining information to help resolve a particular problem (e.g. demand for food in 2050)

- **The distinction between them is in the application**
  - Basic has little application to real world policy and management but could be done to guide applied research

Types of Research

Basic	Applied
Theoretical?	Quick Answers?
More persistent?	Less persistent?
Laboratory /observationBased?	Field Based?
Tightly Controlled?	Loosely Controlled?
Lacks External Validity?	Externally Valid?
Focus on Mechanism	Focus on Effect
More difficult	Less difficult

Internal Validity?

>

Basic Research aims to expand the frontiers of science and knowledge by verifying or disproving the acceptability of a given theory or attempting to discover more about a certain concept (non-specificity): Example: How does motivation affect employee performance?

Applied Research focuses on a real-life problem or situation with a view to helping reach a decision how to deal with it (Specificity):

Example: Should company adopt a paperless office environment?

Exploratory & Descriptive Research

Exploratory Research

- undertaken with the aim of clarifying ambiguous problems
- general problems usually known but not sufficiently understood
- the purpose is to get more information, not to uncover specific courses of action (subsequent research)
- Determining a specific course of action to follow is not a purpose of exploratory research! Example: Child-Care support programme for employees

Descriptive Research

- undertaken with the aim of determining the characteristics of a population or phenomenon
- Previous knowledge of problem exists
- High degree of precision or accuracy required, Examples:

Who are the main consumers of organic foods?

Casual & Problem Solving Research

Causal Research:

- undertaken with the aim of identifying cause and effect relationships amongst variables
- are normally preceeded by exploratory and descriptive research studies
- Often difficult to determine because of the influence of other variables (concommitant Variation and the presence of other hidden variables): e.g., higher agro-chrmical use causes cancer

Problem solving research:

- Designed to solve a specific problem for a specific decision maker
- Often results in recommendations on decisions or actions
- Problem-solving research is holistic – uses all information relevant to the specific problem, problem-solving research the least durable, e.g. No podding syndrome in soybean.

Business Research

- Business Research may be defined as the “systematic and objective process of gathering, recording and analyzing data for making business decisions” (Zikmund, Business Research Methods, 2002)
- **Business research** is a process of acquiring detailed information of all the areas of **business** and using such information in maximizing the sales and profit of the **business**.
- It is an important management activity that helps companies determine which products will be most profitable for companies to produce. Several steps are necessary when conducting business research; each step must be thoroughly reviewed to ensure that the best decision is made for the company.
-

Methodology

Methodology and Method are often (incorrectly) used interchangeable

- Methodology – is the study of the general approach to inquiry in a given field
- Method – the specific techniques, tools or procedures applied to achieve a given objective
  - Research methods in economics/business include regression analysis, mathematical analysis, operations research, surveys, data gathering, etc
  - While these are different they are interdependent (in the same way as science and research are related)

Reasons to Study Methodology

- Methodology is the manner in which we approach and execute functions or activities
  - Consists of approaches or guidelines, not specific details of how we do the task (they are methods)
- Within a discipline, there are accepted rules of evidence and reasoning
- Research methodology provides the principles for organizing, planning, designing and conducting research. (It does not tell you how to do specific research).
- The central reason for studying research methodology is that it provides a time-tested, proven means of producing new, reliable knowledge.

The Value of Business Research for Managers

Reduction of uncertainty and improvement in the quality of decision-making with several consequent advantages (e.g. strategic, operational) and benefits for organizations

Business Research Methods can be employed in each of the following four stages:

(1) **Identification of problems** and/or opportunities:
Useful for strategy planning, analysis of internal and external organizational environment.

(2) **Diagnosing and Assessment of problems** and/or opportunities: Its purpose is to gain insight into the underlying reasons and causes for the situation. If there is a problem, it asks what happened and why? If there is an opportunity, it seeks to explore, clarify and refine the nature of the opportunity and, in the case of multiple opportunities, seeks to set priorities

(3) **Selection and Implementation of Courses of Action**

After alternative courses of action have been determined, selection of the best possible course.

An important consideration is the quality of forecasting which is an essential tool of research

(4) **Evaluating the Course of Action**

Business Research Methods are used after a course of action has been implemented in order to determine whether activities have been properly implemented and have accomplished what they intended to do

Research Methodology in Social Sciences

- Study which integrates the various components of economics to accomplish a defined, goal-directed research
- To expand our knowledge and make that knowledge useful to the study of world problems
- To learn by doing under the supervision of an advisor (shown to be an effective model)
- Pull together various aspects of economic theories, methods, and analysis to present in a coherent, logical, reliable and useful manner.

Recommendation from the Commission on Graduate Education for Economics in the US (1991)

1. More emphasis on ‘real world’ problems and the application of economic research to them;
2. More emphasis on communication skills, especially writing, and the ability to relate economic knowledge to the public.
3. Not determined solely by the understanding of theories or techniques
4. What is needed is the understanding of economic issues, literature, research process and ability to conduct research and communicate results to the stakeholders.

Common Problems in Methodology

Unable to

- Establish the reason for the research
- Provide clear & concise objectives
- Provide complete reference to prior research on the subject and methods
- Lack of understanding for the conceptual and theoretical basis of the research
- Selection of analytical structural model for mere empirical convenience (or familiarity)
- Presenting conclusions that are merely restatements of analytical findings (i.e. results)

Examples of ‘bad’ Methodology

- Unclear about the research problem
- Unclear about the objectives
- Lack thorough awareness of previous work
- Incomplete conceptualization of the problem
- Confusing research means with ends

“Good research … is no accident.”

Two divergent but related aspects of economic research methodology

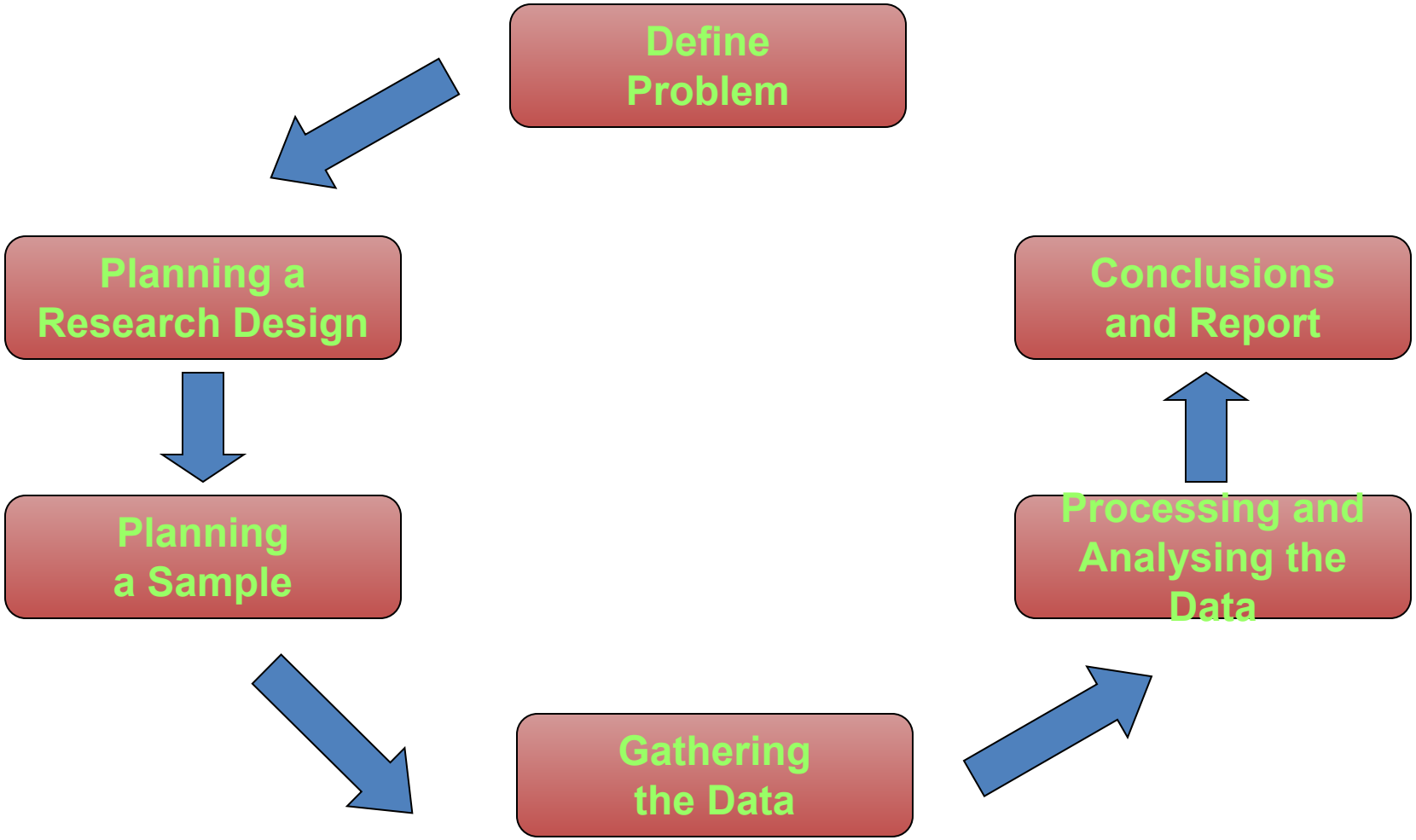
Processes of ***discovery*** *and* ***confirmation***

- Discovery deals with formulating, finding, and creating new knowledge, information
- Confirmation deals with validity or reliability of information
- Discovery is a creative process (art) requiring questioning, probing, pursuing alternative avenues of exploration, etc..
- Confirmation is more highly developed and this is the ‘science’ part of the discipline

Six Steps in Research Process

- Develop an effective research question
- Survey the literature
- Conceptualize the problem
- Test the hypothesis
- Analyze and interpret the results
- Communicate the findings

Stages in the Research Process



The Process of Research

- The process is initiated with a question or problem (**step 1**)
- Next, goals and objectives are formulated to deal with the question or problem (**step 2**)
- Then the research design is developed to achieve the objectives (**step 3**)
- Results are generated by conducting the research (**step 4**)
- Interpretation and analysis of results follow (**step 5)**

Creativity in the Research Process

- Research is a creative process
- *"…research includes far more than mere logic … It includes insight, genius, groping, pondering – 'sense' … The logic we can teach; the art we cannot"*
- Research requires (or at least works best) with imagination, initiative, intuition, and curiosity.
- There are different types of creativity, characteristic of different situations – "applied" and "theoretical" most closely associate with economic research

Fostering Creativity (Ladd 1987)

A. Gather and use previously developed *knowledge*
B. Exchange ideas
C. Apply deductive logic (
D. Look at things alternate ways
E. Question or challenge assumptions
F. Search for patterns or relationships
G. Take risks
H. Cultivate tolerance for uncertainty

Developing an Effective Research Question

Three questions must be answered

1. What is the research topic :General area (unemployment, pollution, poverty)
2. What is the research question: One-sentence question
3. What is the research hypothesis: Proposed answer to your question

Developing an effective research question – How to begin

- Choose a general topic
  - e.g. Indian Journal of Agricultural Economics,
- Start reading the literature
  - What has been done
  - What questions remain
  - Are there contradictions in the literature
- Select a research question from the gaps in the literature
- Introduction to articles
  - Every article will identify the research question and how it differs from other research in the introduction
- Conclusion of articles
  - Every article should list its weaknesses and areas for further study in the conclusion

Choose a Problem

- Based on an idea
- Based on your experience
- Based on your reading
- Originality
- Problem must be clearly recognized
- Determine information already available and what further information is required, as well as the best approach for obtaining it
- Obtain and assess information objectively to help inform the decision

Problem Definition

Describe broader context (background)

After a problem has been “discovered”, it must be “defined” so that the objectives of the research are clear and realizable

State the objectives or purposes

Inform reader about the scope of the study, including defining any terms, limitations, or restrictions

- **Reduces potential criticisms**

State the hypothesis (es)

Problem definition efforts result in statements of research questions and research objectives. These add clarity to the research undertaking and gives managers, researchers and all others concerned an understanding of the approach being used

Research questions should be as specific as possible and be answered with the aid of hypotheses which, by definition, are empirically testable.

Problems in agricultural marketing

- What type of services available and provided to the farmers?
- How to Prioritize services on the basis of its importance?.
- How to minimize the time required for disposal of produce in the market?
- What are the criterions of farmers in selecting the mandi for sale of his produce?
- What are the sources of information on prices?
- What would be the best sources for prices dissemination?
- Feedback on services and facilities provided in mandi and suggestions for their improvement.

- Present level of information about different beneficiary schemes and services available in mandi for farmers and suggestions for easy access of information for farmers.
- Suggestions for improvement in cashless transaction for farmers.
- Additional services required for fruit and vegetable marketing through mandi.

Choosing the title of the study

Characteristics of a Good Research Title

Making a good title involves ensuring that the research title accomplishes four goals

1. A good title predicts the content of the research
2. A good title should be interesting to the reader
3. It should reflect the theme of research
4. Should contain important keywords that will make it easier to be located during a keyword search.
5. If required it should show the location of the research

What should be title for above problems?

1. Study on marketing problems of the farmers.
2. Evaluation of agri-product marketing problems in Madhya Pradesh.
3. An evaluation of krishi upaj mandies' for efficient marketing of agricultural produce in Madhya Pradesh.
4. Study of Krishi Upaj Mandies of Madhya Pradesh

Defining objectives

- Try to keep these simple
- The more variables the more difficult
- However use the opportunity
- The objectives should be stated to the researcher in clear and measurable terms
- Get help at this stage
  - Senior colleagues
  - Experienced researchers
  - Previous studies/literature

Research objectives explain the purpose of the research in measurable terms and define standards of what the research should accomplish. Research objectives indicate the information needed by a researcher to make a decision

Objectives for above problems in markeitng

1. To study the sources of information about prevailing market prices and extent of knowledge about different facilities and services provided by Mandi committee to the farmers.
2. To prioritize the services provided by Mandi Committee for facilitating farmers more effectively.
3. To examine the criterions for decision making in selection of Mandi and time required for final disposal of the produce by the farmers.
4. To assess the accessibility of cashless transactions, its benefits and problems face by farmers.

5. **To examine the services available and to be required for efficient marketing of fruits and vegetables through Mandi.**
6. **To suggest the policy measures for efficient and effective marketing of agricultural produce for betterment of farming community.**

Hypothesis

- It is a testable statement about the relationship between two or more variables or a proposed explanation for some observed phenomenon.
- In a scientific experiment or study, the hypothesis is a brief summation of the researcher's prediction of the study's findings, which may be supported or not by the outcome.
- Hypothesis testing is the core of the scientific method.
- Statistically it is an assumption about certain characteristics of a population. If it specifies values for every parameter of a population, it is called a simple hypothesis; if not, a composite hypothesis.
- If it attempts to nullify the difference between two sample means (by suggesting that the difference is of no statistical significance), it is called a null hypothesis.

- A simple hypothesis might predict a casual relationship between two variables, meaning that one has an effect on the other.
- Example: More irrigation will result in higher yield of sugarcane. irrigation, in this statement, is the independent variable and the yield of sugarcane is dependent variable.
- The independent variable is manipulated and the dependent variable is measured to see how it is affected as the independent variable changes.

Types of hypothesis

A **simple hypothesis** is a prediction of the relationship between two variables: the independent variable and the dependent variable e.g. **Education of the farmers leads to higher adoption**.

A **complex hypothesis** examines the relationship between two or more independent variables and two or more dependent variables e. g. **education and proximity to urban area leads to higher knowledge and adoption.**

A **null hypothesis** (H0) exists when a researcher believes there is no relationship between the two variables, or there is a lack of information to state a scientific hypothesis. This is something to attempt to disprove or discredit e.g. **there is no significant relation between education of the farmers and adoption of technology .**

This is where the **alternative hypothesis** (H1) enters the scene. In an attempt to disprove a null hypothesis, researchers will seek to discover an alternative hypothesis e.g. **there is a significant relation between education of the farmers and adoption of technology**

A **logical hypothesis** is a proposed explanation possessing limited evidence. Generally, you want to turn a logical hypothesis into an empirical hypothesis, putting your theories or postulations to the test, e.g. **tomato plant will grow faster than paml plant on the mars** (Until we're able to test plant growth in Mars' ground for an extended period of time, the evidence for this claim will be limited and the hypothesis will only remain logical.)

An **empirical hypothesis**, or working hypothesis, comes to life when a theory is being put to the test, using observation and experiment. It's no longer just an idea or notion. It's actually going through some trial and error, and perhaps changing around those independent variables e.g. roses watered with liquid biofertilizer will grow faster than roses watered with solid biofertilizers (Here, trial and error is leading to a series of findings.)

A **statistical hypothesis** is an examination of a portion of a population.

If you wanted to conduct a study on the survival of seasonal plants under waste land condition , you would want to examine every single plant. This is not practical. Therefore, you would conduct your research using a statistical hypothesis, or a sample of the planted samples.

Parameters of a Good Hypothesis

1. Define the independent and dependent variables very specifically
2. Be logical and use precise language.
3. Keep your language clean and simple.
4. State your hypothesis as concise, and to the point, as possible.
5. Make sure your hypothesis is testable with research and experimentation.

Hypothesis for above problem

H_1

- Mass media approach is a good source for information about prices.
- Selection of specific mandi is dependent on socio-economic factors of the farmers.
- Marketing of fruits and vegetables required specialized markets.

H_0

- Mass media approach is not a good source for information about prices.
- Selection of specific mandi is not dependent on socio-economic factors of the farmers.
- Marketing of fruits and vegetables is not required specialized markets.

Research Prioritization

Research prioritization is one of the key nodal points in the **research policy planning cycle**, which encompasses

1. Research planning,
2. Research priority setting,
3. Strategies and implementation of research priorities,
4. Research utilization,
5. Research monitoring and evaluation and
6. Overall research policy management.

Prioritization and P Matrix

Prioritize means to rank in order of importance. There are so many research problems you should be sure to **prioritize** the important one. **Prioritize** can also **mean** to set something at the top of a ranking system.

As a principle, it means doing **'first things first**;' as a process, it means evaluating a group of items and ranking them in their **order of importance** or urgency.

A **Prioritization Matrix** is a useful technique to identify which problems **are** the most important to work on solving first. The **Matrix** helps you rank problems or issues generated through brainstorming, using weighted criteria that **are** important to your project and/or organization.

Research Prioritization in Agriculture

The prevailing fund crunch add urgency to the **optimum allocation of scarce resource among competing enterprises**. Hence the ***resources allocated to agricultural*** **research are getting less and the expectations from research are at a higher scale**

Hence, it is an imperative need to **analyze and prioritize the allocation of research resources among the competing research programs and projects**?

With a view to make research prioritization an objective & transparent activity, several methods were developed and are recommended for adoption by the National and international research organization and institutions.

Methods of Research Prioritization

1. **Congruence Method:**

- Research funds are to be allocated to commodities in the same proportion as their existing contribution to Gross Domestic Product (GDP) from the Agricultural Sector.
- This method ensures certain minimum allocation of research resources for each agricultural commodity.
- It implicitly assumes that opportunities for research are equal across commodities, and that the value of new knowledge generated by research is proportional to the value of output.
- It also assumes constancy of relative shares of different commodities as it is based on present values.
- ***Suitable for the initial fund allocation at the national level*** **and** ***applicable to the commodity related research only.***

Example: Research institute XYZ is having 200 crores for research during 2019-20. Important commodities of the region are soybean, paddy, pigeonpea, chickpea, wheat and lentil.

1step to collect the data on percentage contribution of commodities in agricultural GDP of the Nation/state during last three years (mean value).

Commodities	Mean AGDP contribution	Allocation of funds (Rs.Crores)
Soybean	20.62	41.24
Paddy	10.56	21.12
Pigeonpea	5.60	11.20
Chickpea	21.87	43.74
Wheat	25.63	51.26
Lentil	12.30	24.60
Other	3.42	6.84
Total	100	200.00

2. Domestic Resources Cost Ratio Method:

This method ranks the commodity programs on the basis of comparative advantage of domestic production.

DRCR = A/(B – C)

Where

‘A’ is the unit cost of producing a commodity domestically

‘B’ is the unit cost of importing a commodity from other countries

‘C’ is the cost of imported inputs used in domestic production (per unit out output)

When this ratio is less than one, relatively a high priority must be accorded to that commodity. If the ratio is more than one, there is a case of low priority as it will be cheaper to import it from outside rather than produce locally.

Example: For palm oil research

1. Unit cost of producing one ton of palm oil is domestically is Rs. 90,000 **(A)**
2. Unit cost of importing the one ton of palm oil from other countries is Rs. 85,000. **(B)**
3. Cost of imported inputs (fertilizes and Agro-chemicals) use in domestic production of one ton palm oil is Rs 1,5000 **(C)**

Therefore,

DRCR = A/(B – C) will be 90000 /(85000-15000)

That will be 1.28 it is better to import palm oil from other countries rather than spending on research for domestic cultivation of this crop

3. Checklist Method

- This method basically consists of a set of criteria requiring answers to set priorities for research programs.
- Depending on the degree of satisfaction of the set criteria, programs and projects are ranked.
- Due to its *simplicity & non-mathematical approach, it is* most *liked by scientists.*
- Cannot be used for individual lines of work within the
- commodity programs
- Applicable to both commodity as well as non-commodity
- research
- Used for arriving at priorities even for individual lines of research

Example in entomological research of the institute

Criterion	Soybean leaf miner	Chickpea cater pillar	Cotton pink bollworm	Maize stem borer
Stage of loss	01	03	05	02
Extent of losses	02	04	05	03
Importance of the crop for the farmer	04	03	05	02
Cost of plant protection	01	02	03	04
Alternate crop	04	01	05	01
Total	12	13	23	12
Highest expected score	25	25	25	25
% of HES	48	52	92	48

Weight for higher extent 5 and least extent 1

4. Scoring or Weighted Criteria Method

- Relative weights are assigned to the criteria chose to arrive at the set of priorities.
- Information on the research area may be collected on each criterion from primary and secondary sources.

5. Economic Surplus Approach

- This model provides a relatively simple, flexible approach to specifying the value of research, by comparing the situations with and without it.
- It computes gains in production efficiency through reduction in per unit cost.
- It is useful to estimate the benefits

6. Other Methods: Econometrics, Programming, and simulation models

Notice that:

"… truth was not used in the definition of research"

"This concept of truth is outside of the productive realm of thinking by researchers"

"…the beauty of economics rests in its theory, but the power of economics lies in its application to current problems."

<u>Data Collection</u>

- **an intermediate step to gain reliable knowledge**
- **collecting reliable data is part of the research process**

<u>Searching out published research results in libraries (or the internet):</u> This is an important early step of research

- **The research process always includes synthesis and analysis**
- **But, just reviewing of literature is not research**

-Concept of sampling,

-probability and non-probability

-sampling techniques.

-Simple random sampling,

-stratified sampling,

-multi-stage sampling,

-systematic sampling,

-purposive sampling, quota sampling, judgment sampling and convenience sampling.

-Sampling size determination and

- sampling and non sampling errors.

Concept of sampling

- **Sampling** is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population.
- The methodology used to **sample** from a larger population depends on the type of analysis being performed but may include simple random **sampling** or systematic **sampling.**

Sampling

- The procedure of selecting certain number of study units from a defined population is called Sampling.
- A representative sample has all important characteristics of the population from which it is drawn.

Types of Sampling

- **Probability Sampling**

  When each sampling unit has an equal and known chance of being included in sample.

- **Non probability Sampling**

  When sampling units do not have an equal chance of being included in the sample.

Non Probability Sampling

- Convenience Sampling
- Quota Sampling
- Snow-ball Sampling
- Temporal Sampling

Convenience Sampling

In this method, the study units that happen to be available at the time of data collection, are selected in the sample e. g. number of primary students in the school.

Quota Sampling

This method insured that a certain number of study units from different categories having specific characteristics appear in the sample so that all these characteristics are represented e.g. sample to be drawn from different size of holdings Small, medium and large.

Snow Ball Sampling

When individual with certain characters are asked to identify similar individuals for inclusion in the study e.g. farmers using combine harvesters with facility of bhusa making machine.

Temporal Sampling

When all cases occurring in a specific period time are included in the study e.g. farmers sown chickpea during period of 15th November to 15th December.

Probability Sampling

Probability Sampling involves selecting procedure to ensure that all study units have an equal and / or known chance of being included in the study.

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Sampling
- Cluster Sampling
- Multi-stage Sampling

Simple Random Sampling

For selecting a simple random sample we need :

1. **To make a numbered list of all the study units in the population from which you want to draw a sample**
2. **To select the required number of sampling units using a *Lottery method or Random Number Tables.***
3. ***Example selection of 50 chillie growers from the list of chllie growers of the village (1250 farmers) use lottery system by putting all the number cheats 1-1250 in a bowl and draw 50 number cheats.***

Systematic Random Sampling

In systematic random sampling, the sampling units are selected at regular intervals (e.g. every 5th, 15th, 34th) from the sampling frame e.g. from the similar list of chillie growers select every 25th interval so that we can select 50 samples (50 X 25 = 1250)

Stratified Sampling

If it is important that the sample includes the representative groups of study units with specific characteristics e.g. chille growers from different size group of holdings then:

- The sampling frame must be divided into groups or STRATE according to these characteristics (size of holding). In above case divide 1250 farmers according to size group of holdings e.g. 850 small, 230, medium and 170 large.
- Random or systematic random samples of predetermined size will then be obtained from each group (Stratum).

Cluster Sampling

- In Cluster Sampling, a simple random sample is selected not of individual subjects but of groups or clusters of individuals. The clusters may be villages, apartments, classes, housing units, families etc.
- This is often a convenient method, especially when there is no sampling frame showing all individual subjects.
- The cluster are often geographic unit (village) or organizational units (clinics, factories, schools).

Multi–stage Sampling

- In very large and diverse populations, sampling may be done in two or more stages.
- This is often the case in the community based studies, where the people are to be chosen from different areas.
- The sampling procedure is carried out in phases.
- Example multi stage sampling in case of chillie growers as drawing the sample at first stage farmers growing chilles for dry chllies, green chillies and for both purpose than classifying them as small, medium and large farmers

Sampling Frame

- An important issue influencing the choice of the most appropriate sampling method is whether a sampling frame is available or not.
- Sampling Frame is the listing of all the units that from the study population.

When to use probability and non-probability sampling

If a sampling frame is not available, it is not possible to sample the study units in a such a way that the probability for the different units to be selected in such case use non-probability sampling.

If the sampling frame is exist or can be compiled, each study unit has a known probability of being selected in the sample. In such situation probability sampling can be sued.

Sample size

Single mean:

Standard Deviation ^{2} /Standard error ^{2}

We want to determine the mean yield of the soybean in the Ujjain district. The mean yield is expected to be 1200 kg/ha± sd 20 kg/ha and standard error is 5 kg.

$200^2 / 5^2 = 40000/25 = 1600$ samples

Single rate:

Rate / Standard error ^{2}

The rate of adoption of seed treatment technology is supposed to be 55/1000 farmers in Ujjain district in soybean with permissible SE is 5/1000 farmers.

$(55/1000)/(5/1000)^2 = 0.055/.000025$

=2200 soybean growers from the district.

Single proportion:

Proportion – (100-Proportion) / Standard error ^{2}

Proportion of adopter farmers in total farmers is 200:1000 and standard error is 0.50

(200)- (100-200)/ 0.50 ^{2} = 200 – (-100)/0.50 ^{2}

$^=$ 300/2.5 = 120 samples

Some important Concepts

Percentage =

(Units with specific character/total units) X 100

e.g.How much per cent of farmers are small if their number is 79 out of 220

= (79/220) X 100 = 35.90

Proportion

Relationship of one part of the population in total population

= One part of population/ total population

50 men in population of 200 =50/200=1/4

Percentile: Is a difference between highest value and lowest value divided by 100

= (Highest value – lowest value)/100

e.g. yield of soybean of 10 sample farmers

12, 11, 10, 15, 16, 10, 9, 14,12,11

P = (16-9)/100 = 0.07

Ratio: Relationship of two parts of the population e.g. 50 small farmers in 150 farmers i.e 150/50= 3 =3:1

Measures of Central Tendency

Mean, Median and Mode

Mean: Sum (∑) of all values (x) divides by number of observations (n) **Thus Mean X = ∑x/n**

It is the most common and best general purpose measure of the mid-point (around which all other values cluster) of a set of values.

The mean is also known commonly as the average or arithmetic mean and is a basic mathematical function which allows us to better understand populations.

e. g, yield (kg/ha) of soybean of 10 sample farmers

X = 12, 11, 10, 15, 16, 10, 9, 14,12,11= ∑x120

Thus Mean x = 120/10= 12kg/ha

Advantages : 1.Easy to calculate 2. Give an idea of central tendency.

Disadvantages: 1. Influence by extreme values 2. May not convey proper sense e.g mean of labour use in soybean production 120.24.

There are other types such as Geometric Mean and Harmonic Mean.

Median

Median: When the data is arranged in ascending of descending order, the median is the value that divides the data in two equal parts.

- If the number of data points is odd, the median is the middle data point in the list.
- If the number of data points is even, the median is the average of the two middle data points in the list.

The middle number; found by ordering all data points and picking out the one in the middle (or if there are two middle numbers, taking the mean of those two numbers).

Advantages: It is not influence by extreme values

Disadvantages: 1.Not very precise measure 2.Not useful for other calculations

Example: X = 12, 11, 10, 15, 16, 10, 9, 14,12,11

Arrange in ascending order

09,10,10,11,11,12,12,14,15,16

Thus median will be 11 & 12 and mean of these two value will be 11.50

Example: 12,13,14,17,12,15,16,19, 20

Arrange in ascending order

12,12,13,14,15,16,17,19,20

The median will be 15

Mode

- **Mode:** Most frequent value of the series called as mode or The most frequent number- that is, the number that occurs the highest number of times.
- **Example:** X = 12, 11, 10, 15, 16, 17, 9, 14, 18, 11 Mode is 11

Mean of frequency classified data

Adoption index of 21 farmers is given below:

59, 65, 61, 62, 53, 55, 60, 70, 64, 56, 58, 58, 62, 62, 68, 65, 56, 59, 68, 61, 67

Simple mean will be Sum of 21 data/21 = 61.38095..

Median will be

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

And mode will be

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

Adoption index classes	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4
Total	21

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

The groups (51-55, 56-60, etc), also called class intervals, are of width 5. The midpoints are in the middle of each class: 53 ((51+55)/2), 58 ((56+60)/2) and so on, 63 and 68.

Mid point (M)	Frequency (F)	M X F
53	2	106
58	7	406
63	8	504
68	4	272
Totals:	21	1288

Estimated mean = $\sum MF / \sum F$ = 1288/21= 61.333333

Simple mean was 61.38095.

Estimating the Median from Grouped Data

Adoption index classes	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4
Total	21

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

The median is the middle value, which in our case is the 11th one, which is in the 61 - 65 group: We can say "the median group is 61 - 65"

Estimated Median = L + (*(n/2) – B)/***G)**) × w

Estimated Median = L + {*(n/2) – B)/*G)} × W

where:

L is the lower class boundary of the group containing the median

n is the total number of values

B is the cumulative frequency of the groups before the median group

G is the frequency of the median group

W is the group width

In our example

L = 60

n = 21

B = 2 + 7 = 9

G = 8

w = 5

$$\text{Estimated Median} = L + \{(n/2) - B)/G)\} \times W$$

$$\text{Estimated Median} = 60 + \{(21/2) - 9)/8)\} \text{ X } 5$$

$$= 60 + \{ (10.50-9)/8\} \text{ X } 5$$

$$= 60 + 09375$$

$$= 60.9375 = 61.0$$

Estimating mode from Grouped Data

Adoption index classes	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4
Total	21

mode will be

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

We can easily find the modal group (the group with the highest frequency), which is 61 – 65

We can say "the modal group is 61 - 65“

But, we can estimate the Mode using the following formula:

Estimated Mode = L + {($f_m - f_{m-1}$)/ ($f_m - f_{m-1}$) + ($f_m - f_{m+1}$) } × w

Where,

L is the lower class boundary of the modal group

f_{m-1} is the frequency of the group before the modal group

f_m is the frequency of the modal group

f_{m+1} is the frequency of the group after the modal group

w is the group width

Adoption index classes	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4
Total	21

In this example:

$L = 61$

$f_{m-1} = 7$

$f_m = 8$

$f_{m+1} = 4$

$w = 5$

Estimated Mode $= L + \{(f_m - f_{m-1}) / (f_m - f_{m-1}) + (f_m - f_{m+1})\} \times w$

$= 61 + \{\{8-7\}/(8-7) + (8-4)\} \times 5$

$= 61 + \{(1)/(1+4)\} \times 5$

$= 61 + 1 = 62$

Modes of Dispersion

Variance, Standard Deviation and coefficient of variation

Variance: **T**he Variance is defined as: The average of the **squared** differences from the Mean.

Work out the mean (the simple average of the numbers).Then for each number: subtract the Mean and square the result (the *squared difference*).Then work out the average of those squared differences.

S. No.	Yield of wheat on sample farms (q/ha) =X	X^2
1	17	289
2	15	225
3	23	529
4	07	49
5	09	81
6	13	169
Total	84	1342
Mean	14	-

$$\text{Variance} = 1/(n-1)\ \{\sum Xi^2 - (\sum Xi)^2 /n\}$$

$$= 1/(6-1)\{1342- (84)^2 /6$$

$$= 1/5\ \{1342- 1176\}$$

$$= 33.20$$

Standard Deviation is just the square root of Variance, so:

$$SD = \sqrt{V} = \sqrt{33.20} = 5.76$$

Coefficient of variation

The coefficient of variation (relative standard deviation) is a statistical measure of the dispersion of data points around the mean.

CV % = (sd/mean) X 100

Example it will be =(5.76/14) X 100 = 41.14%

DATA

Something assumed as facts and made the basis of reasoning or calculation.

1. **Qualitative or Categorical**

   Sex, Color, Race, caste etc

2. **Quantitative or Numerical**

   Age, Height, land holding etc

Categorical Data

- **Nominal:** categories of data cannot be ordered one above the other.

  Sex: Male, Female

  Marital Status: Single, Married, Divorced.

- **Ordinal:** Categories of the data can be ordered one above the other or vice versa.

  Level of Knowledge: Good, Average, Poor

  Opinion: Fully Agree, Agree, Disagree.

Variable

An item of data that can be observed or measure

Quantitative Variable

A variable that has a numerical value

e.g. Age, Schooling years, No. of Children etc.

Qualitative Variable

A variable that is not characterized by a numerical value.

e.g. Sex, Category of farmer, level of education etc.

Quantitative Variable

Discrete Variable

A quantitative variable, whose possible values are in whole numbers.

Example: Number of farmers, number of students

Continuous Variable

A quantitative variable that has an uninterrupted range of values

Example: size of holding, Weight, height etc

Types of Variables

Independent Variable

A variable, whose effect is being measured. (Cause) e.g. rainfall

Dependent Variable

The variable, on whom the effect is being observed. (Effect) e.g. farming income

Confounding Variable

A variable, which affects both independent as well as dependent variable. (cause as well as Effect) e.g. productivity of crops

Data Collection Techniques

- Using available information
- Observing
- Interviewing
- Administering questionnaire
- Focus Group Discussion (FGD)
- Nominal Group Technique (NGT)
- The Delphi Technique

Types of Questions

- Open-ended
- Close-ended

- **Open – ended Questions**

Permits free response that should be recorded in the respondents own words.

e.g. your opinion about NERAGA for providing employment

- **Close-ended Questions**

Close-ended questions offer a list of possible options or answer from which the respondent must choose.

e.g. NAREGA is providing sizeable employment

1. Yes 2. No

Hypothesis

Idea/ suggestion put forward as a starting point for reasoning or explanation

Hypothetical

Not based on certain knowledge

Hypothesis is a supposition, which is tested by collecting facts. The analysis of these facts leads to its acceptance or rejection.

Cont……

A researcher wants to compare the farming in two regions.

There can be two hypothesis.

- **There is no difference between farming of the two areas being studies (Null Hypothesis)**
- **There is difference between farming of two areas (Alternative Hypothesis)**

Hypothesis Testing

- Student's "t" Test
- Chi – Square Test
- Correlation
- Regression
- F-test

Results

- Present your findings in a logical sequence
- Use tables and graphs to summarize data.
- Mention negative results of interest.
- Give statistical significance.
- Estimate accuracy and precision of results.
- Avoid vague statements.

Results

- Present the findings in simple, standard, scientific language.
- Avoid abbreviations where ever possible.
- Tables should be self contained.
- The findings presented in the tables should not differ from those given in the text.
- As far as possible, do not repeat in the text, what has been given in the tables.

Presentation of Data

1. Tables
2. Graphs

- Bar
  - -Vertical
  - -Horizontal
- Histogram
- Line Graph
- Scatter
- Pie Chart
- Column

Types of tables

One way Tables

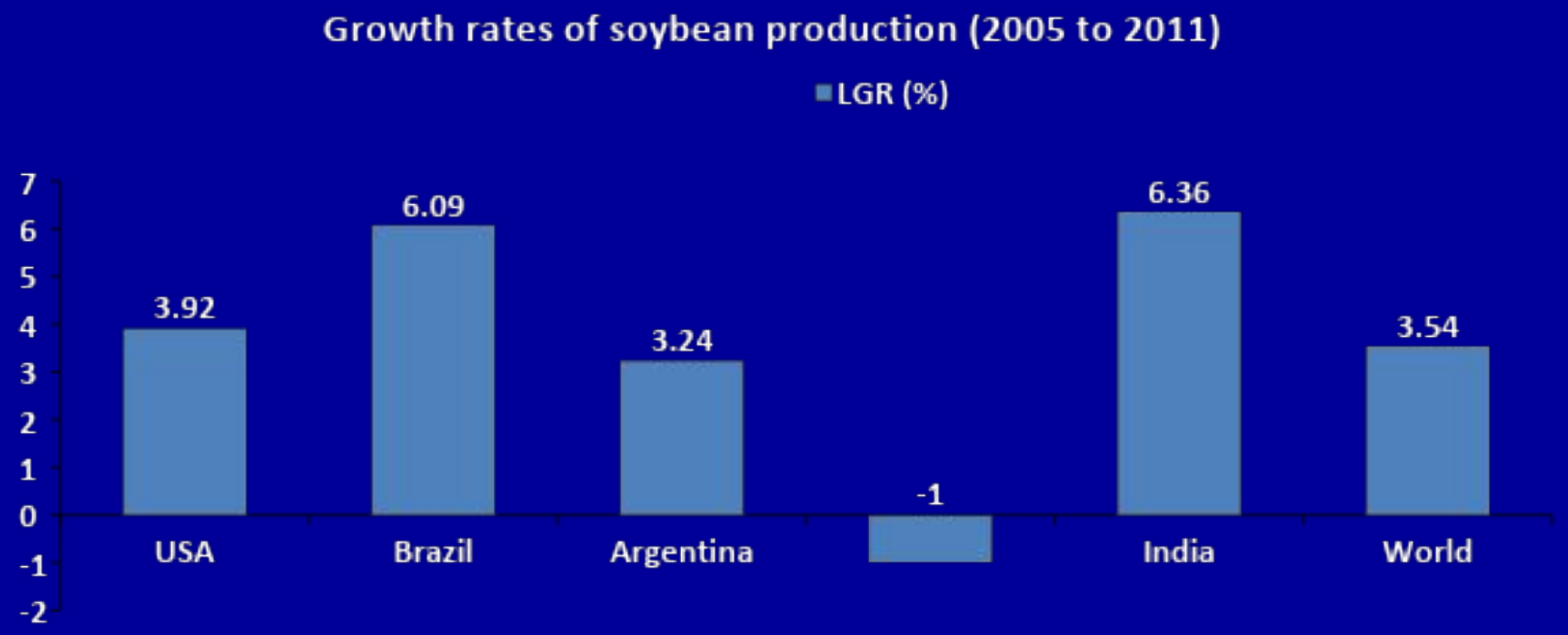
Method of planting	Yield (kg/ha)	Percentage increase
Farmers practice	1349	----
Ridge-furrow planting system	1732	28.4

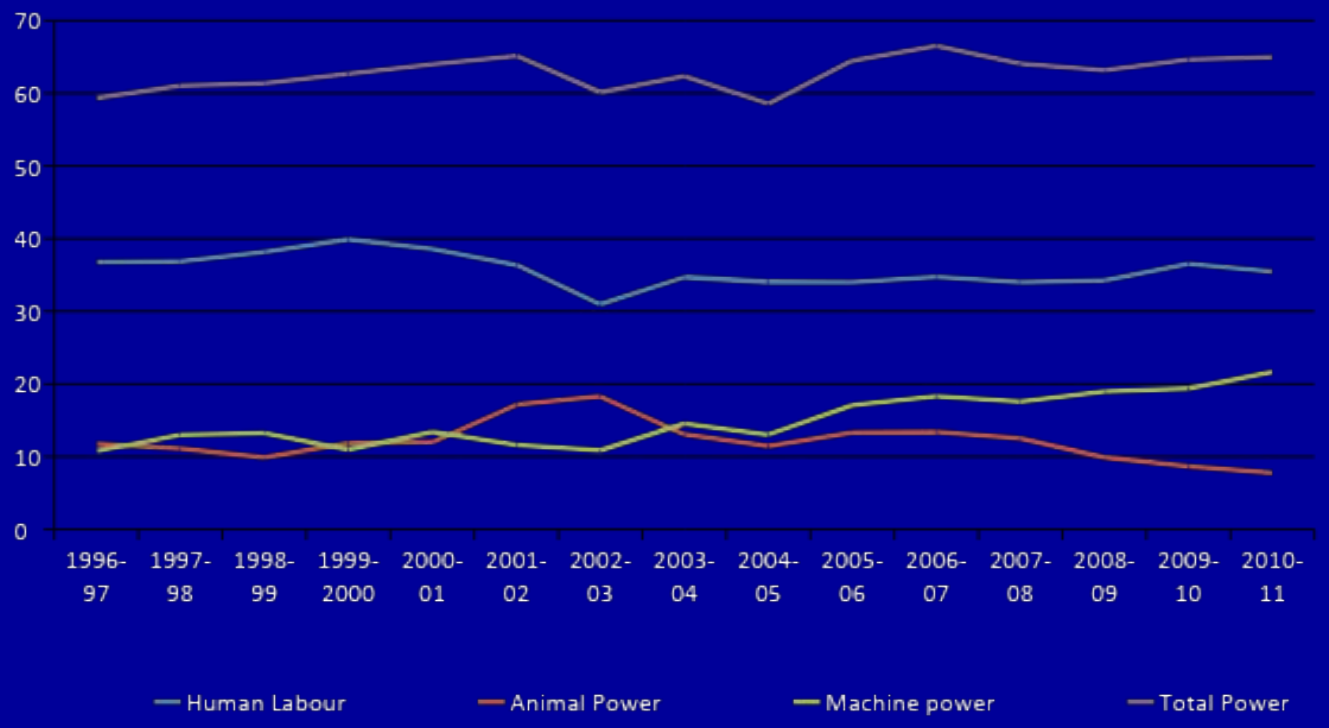
Two way Tables

CROPS	2007-08			2008-09			2009-10		
	AREA	PROD.	YIELD	AREA	PROD	YIELD	AREA	PRO	YIELD
TUR	3.04	1.97	643	3.10	2.47	795	3.29	2.57	780
URD	5.70	1.95	341	5.17	1.96	379	5.70	2.14	375

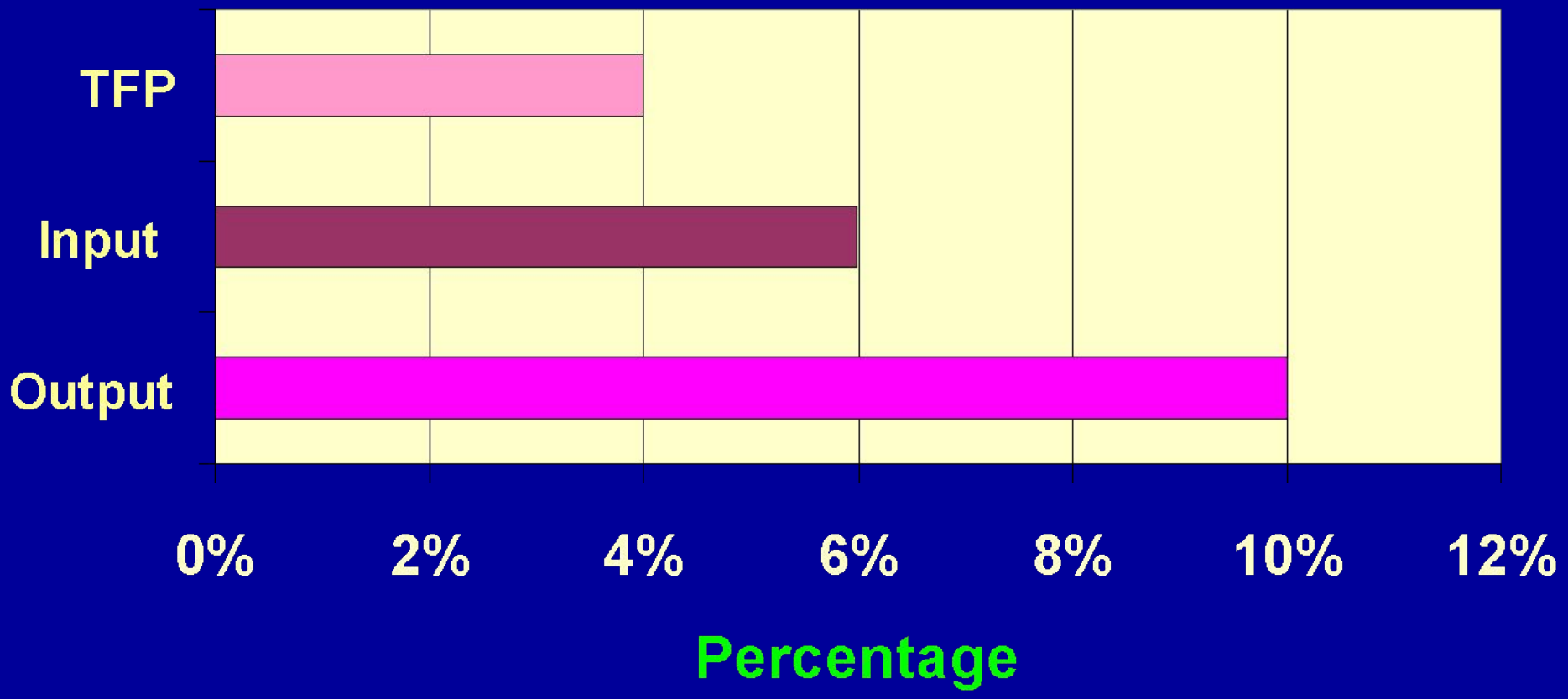
Types of Graphs

Horizontal bar diagram

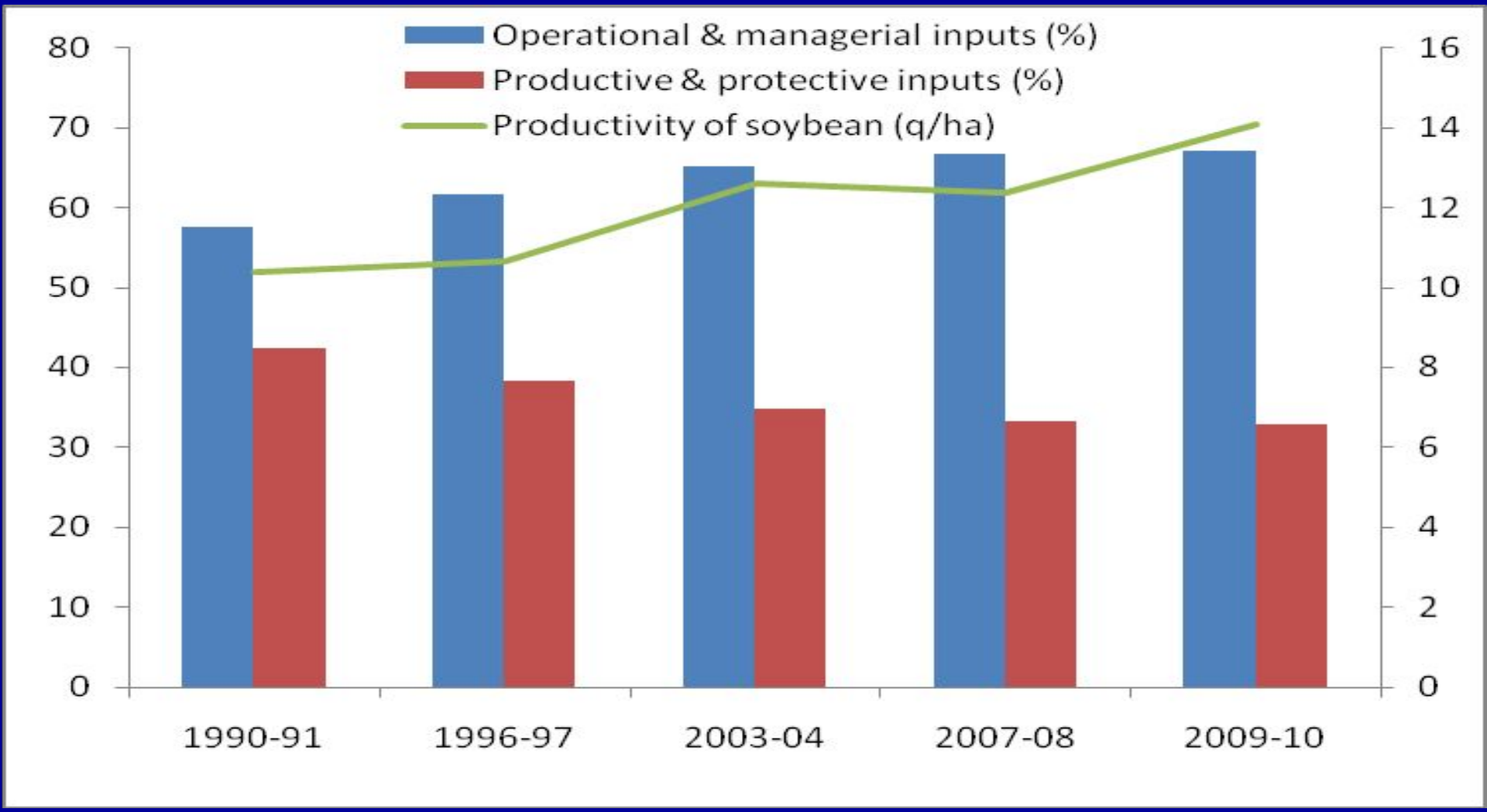



70
60
50
40
30
20
10
0
1996-97
1997-98
1998-99
1999-2000
2000-01
2001-02
2002-03
2003-04
2004-05
2005-06
2006-07
2007-08
2008-09
2009-10
2010-11
Human Labour
Animal Power
Machine power
Total Power

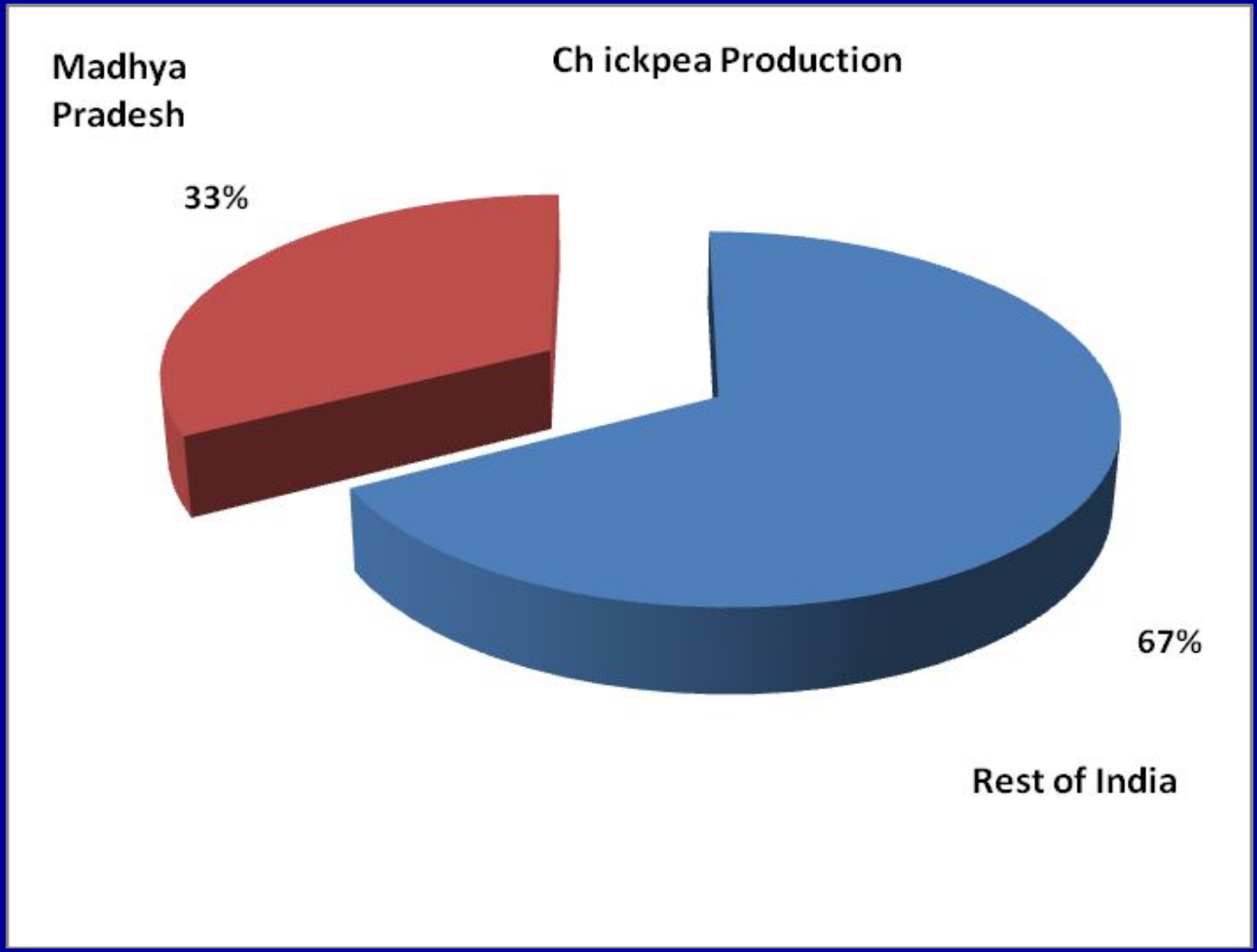
Condition I



Compound graphs



Pie Diagram



Discussion

- How are the findings different from other studies?
- What is the statistical significance of these difference?
- What are the probable reasons for these difference?
- What are the policy implications of these findings?

THANKS