Unit 5

Formulating Research Problem

Considerations: Relevance, Interest, Data Availability, Choice of data, Analysis of data, Generalization and Interpretation of analysis

Statistics in Research:

1. Measures of Central Tendency
2. Measures of Dispersion
3. Measures of Asymmetry
4. Chi-square Test
5. T-tests

Suppose we have data about heights of the people present in the room or age of the people. So, mean is a kind of representative data that cuts the data set at the middle.

1. ***Measures of Central Tendency***

a. Mean = Average

Data set of exam results:

x1 Mathew 77
x2 Mark 73
x3 Luke 69
x4 John 64
x5 Peter 52

$$= \frac{(x_1 + x_2 + x_3 + x_4 + x_5)}{5}$$

Mean = 69

b. Median = Middle / Mid point

We list the values in either in ↑ and ↓ order and read off the middle value.

Mathew	77
Mark	73
Luke	69 = Median
John	64
Peter	52

If we have even num of values

Mathew	77
Mark	73
Paul	71
Luke	69
John	64
Peter	52

Take avg of 2 middle values

Median = 70

c. Mode

The most common value in data set . can be found easily by finding which value occur most frequently.

For Example: 52, 52, 64, 64, 69, 71, 73, 73, 73, 77 Mode = 70

2. ***Measures of Dispersion***

a. Range (R) It is the difference between highest and lowest value of the series.

$$R = H - L$$

Example : [Individual Series]

For given series 3, 8, 2, 5, 6, 9, 10 find range?

R = 10 – 2

R = 8

b. Mean Deviation

Mean Deviation from Mean [$\bar{x}$] $MD_{\bar{x}}$

Mean Deviation from Median [M] MD_M

Example : [Discrete Series]

$R = H - L$

R = 50 – 10

R = 40

x	Freq
10	6
20	3
30	4
40	7
50	5

Do not see in the frequency.

c. Standard Deviation

Datta | Data

$$\sigma^2 = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}$$

b. Mean Deviation

Mean Deviation from Mean [$\bar{x}$] $MD_{\bar{x}}$

x	$\lvert x - \bar{x} \rvert$
10	10-30=20
20	10
30	0
40	10
50	20

$$MD_{\bar{x}} = \frac{\sum|x - \bar{x}|}{N} = \frac{60}{5} = 12$$

$$\sum x = 150 \quad \bar{x} = \frac{\sum x}{N} = \frac{150}{5} = 30$$

Mean Deviation from Median [M] MD_M

x	$\lvert x - M \rvert$
10	10-30=20
20	10
30	0
40	10
50	20

$$MD_M = \frac{\sum|x - M|}{N}$$

$$MD_M = \frac{60}{5} = 12$$

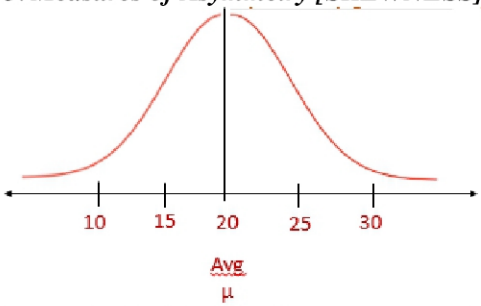
$M = 30$

$$M = \left(\frac{n+1}{2}\right)^{th} item = \frac{5+1}{2} = 3$$

Generally, we need to find out Mean Deviation from Mean and Mean Deviation from Median.

c. Standard Deviation: Suppose we have data about heights of the people present in the room or age of the people. So, mean is a kind of representative data that cuts the data set at the middle. We want to find out the deviation from the mean, that is any data value is how far away from mean. Let the mean is 5 then the diff b/w 10 & 5 is +5, from 7&5 will be +2 but in other side so by using the formula I am going to get ZERO…that is not my goal. I want a number…So, what we will do is, we will square it. Squaring action make it positive.

3. ***Measures of Asymmetry [SKEWNESS]***



When the distribution of item in a series happens to be perfectly symmetrical, such curve called as a **Normal curve**.

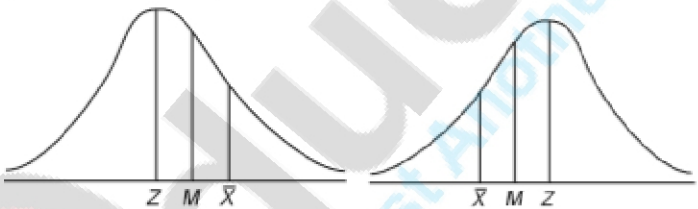
1. As we go farther away from average line the probably falls down.

2. Since the shape of the curve is just like bell that's why it is called as **Bell Curve.** Lager population value is right around the mean.

3. This is a continuous distribution.

4. There is symmetry in this diagram if we fold the paper on this avg line we will get perfect symmetry.

5. This graph is completely defined by mean (μ) and standard deviation (σ).



Curve showing positive skewness
In case of positive skewness we have:
$Z < M < \bar{X}$

Curve showing negative skewness
In case of negative skewness we have:
$\bar{X} < M < Z$

Curve is distorted on the right side Curve is distorted on the left side

IQ distribution of population across the country. We will take an example of watermelon. Now there is whole giant shelf of watermelon in an supermarket. N u randomly pick up mature watermelon and then u take messuere to end to end n u record that no. and u take second and so on and length of that is denoted by x. We know that even they look like same in length but still by fraction of some no they differ in length from each other. There will be some trend most of the watermellon have an avg length of certain value. Some will be bigger than that some will be smaller then that.. U may have too small or too bigger than avg length. If u want to represent the spread of these

watermelon or may be the probably of drawing of water mellon till certain length, we can draw it. It is called Normal Distribution. We say that the length of these watermelon are normally distributed.

1. By looking at the diagram we can say that as we go farther away from average line the probably falls down.
2. Since the shape of the curve is just like bell that's why it is called as **Bell Curve.** Lager population value is right around the mean.
3. This is a continuous distribution. Since watermelon can be any length b/w 10-15 or 25-30.
4. There is symmetry in this diagram if u fold the paper on this avg line u will get perfect symmetry.
5. This graph is completely defined by mean (μ) and standard deviation (σ) mean tells where the peak is and standard deviation tells how fatter the curve will be..

When the distribution of item in a series happens to be perfectly symmetrical, we then have the following type of curve for the distribution:
Such a curve is technically described as a normal curve and the relating distribution as normal
distribution. Such a curve is perfectly bell shaped curve in which case the value of X or M or Z is just the same and skewness is altogether absent. But if the curve is distorted (whether on the right side or on the left side), we have asymmetrical distribution which indicates that there is skewness. If the curve is distorted on the right side, we have positive skewness but when the curve is distorted towards left, we have negative skewness as shown here under:
Skewness is, thus, a measure of asymmetry and shows the manner in which the items are clustered around the average.
In a symmetrical distribution, the items show a perfect balance on either side of the mode, but in a skew distribution the balance is thrown to one side. The amount by which the balance exceeds on one side measures the skewness of the series. The difference between the mean, median or the mode provides an easy way of expressing skewness in a series. In case of
positive skewness, we have $Z < M < X$ and in case of negative skewness we have $X < M < Z$.

4. Chi-square test [Goodness of fit test]

In order to judge the significance of association between two attributes, we make use of Chi square test.

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where :
O_{ij} = observed frequencies
E_{ij} = expected frequencies

1. **Degree of freedom** is (DF)= (r-1) * (c-1)

2. The **P-value** is the probability of obtaining a sample "more extreme" than the ones observed in our data.

Example: 256 people were surveyed to find out their zodiac sign. The results were: Aries (29), Taurus (24), Gemini (22), Cancer (19), Leo (21), Virgo (18), Libra (19), Scorpio (20), Sagittarius (23), Capricorn (18), Aquarius (20), Pisces (23).

Q. **What is Chi-square test? Explain its significance in statistical analysis.**
Hypothesis : Zodiac signs are evenly distributed across visual artists.

Count	Observed	Expected	Obs-Exp	(Obs-Exp)*2	[(Obs-Exp)*2]/Exp
1	29	21.33333	7.666667	58.77777778	2.755208333
2	24	21.33333	2.666667	7.111111111	0.333333333
3	22	21.33333	0.666667	0.444444444	0.020833333
4	19	21.33333	-2.33333	5.444444444	0.255208333
5	21	21.33333	-0.33333	0.111111111	0.005208333
6	18	21.33333	-3.33333	11.11111111	0.520833333
7	19	21.33333	-2.33333	5.444444444	0.255208333
8	20	21.33333	-1.33333	1.777777778	0.083333333
9	23	21.33333	1.666667	2.777777778	0.130208333
10	18	21.33333	-3.33333	11.11111111	0.520833333
11	20	21.33333	-1.33333	1.777777778	0.083333333
12	23	21.33333	1.666667	2.777777778	0.130208333
	256				5.09375

It is mean = 256/12

23-21.3333= 1.6667

In order to judge the significance of association between two attributes, we make use of Chisquare test*

Two basic terminologies used for performing Chi-square test is

1. Degree of freedom
2. P-value: the P-value is the probability of observing a sample statistics as extreme as the test statistic.

4. Chi-square test [Goodness of fit test]

Degree of freedom is (DF)= n-1 = 12-1 = 11

Chi-square table

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801

Hypothesis : Zodiac signs are evenly distributed across visual artists.

0.900 < P-value < 0.950

Level of significance = 0.05

P-value > Level of significance

Null Hypothesis is accepted.

From given Chi Square table we will find out the P-value. Since DF= 11 so we will look into 11th row where statistical value (5.094) of Chi Square lies?

So, 0.900 < P-value < 0.950

5. t test

The sampling theory for large samples is not applicable in small samples because when samples are small, we cannot assume that the sampling distribution is approximately normal. As such we require a new technique for handlng small samples, particularly when population parameters are unknown. Sir William S. Gosset (pen name Student) developed a significance test, known as Student's t-test, based on t distribution and through it made significant contribution in the theory of sampling applicable in case of small samples. Student's t-test is used when two conditions are fulfilled viz., the sample size is 30 or less and the population variance is not known. While using t-test we assume that the population from which sample has been taken is normal or approximately normal, sample is a random sample, observations are independent, there is no measurement error and that in the case of two

samples when cc, we assume that the population variances are equal. For applying t-test, we work out the value of test statistic (i.e., 't') and then compare with the table value of t (based on 't' distribution) at certain level of significance for given degrees of freedom. If the calculated value of 't' is either equal to or exceeds the table value, we infer that the difference is significant, but if calculated value of t is less than the concerning table value of t, the difference is not treated as significant.

Technique for handling small samples. Ie. the sample size is 30 or less and the population variance is not known.

$$t\ value = \frac{Difference\ between\ two\ means}{variability\ in\ two\ groups} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where

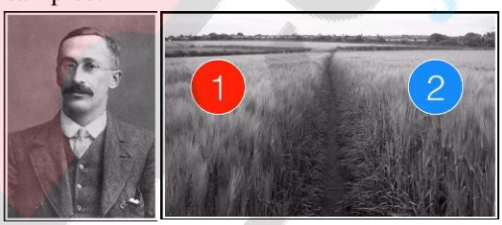
$\bar{x}_1$ & $\bar{x}_2$ means of two different populations

σ_1^2 & σ_2^2 standard deviations of two different populations

n_1 & n_2 are numbers of sample

$$= \frac{|15\,38 - 15\,68|}{\sqrt{\frac{0\,097}{16} + \frac{0\,165}{16}}} = \frac{0.3}{.13} = 2.3$$

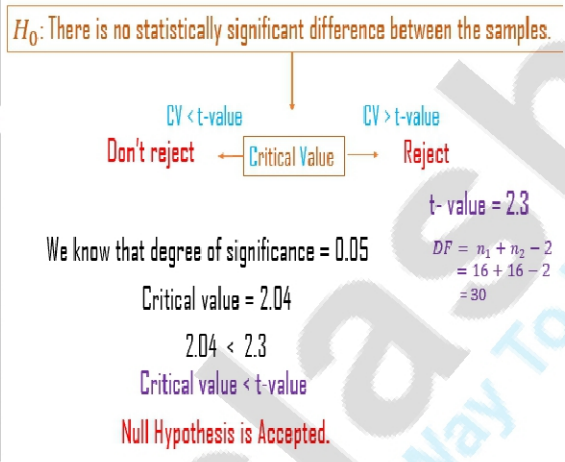
Sir William S. Gosset (pen name Student) developed a significance test, known as Student's t-test, based on t distribution and through it made significant contribution in the theory of sampling applicable in case of small samples.



Equality of the two population means is to be tested

Example:

Degrees of Freedom	p=0.05	p=0.025	p=0.01
1	12.71	25.45	63.66
2	4.30	6.20	9.92
3	3.18	4.17	5.84
4	2.78	3.50	4.60
5	2.57	3.16	4.03
6	2.45	2.97	3.71
7	2.36	2.84	3.50
8	2.31	2.75	3.36
9	2.26	2.68	3.25
10	2.23	2.63	3.17
11	2.20	2.59	3.11
12	2.18	2.56	3.05
13	2.16	2.53	3.01
14	2.14	2.51	2.98
15	2.13	2.49	2.95
16	2.12	2.47	2.92
17	2.11	2.46	2.90
18	2.10	2.44	2.88
19	2.09	2.43	2.86
20	2.09	2.42	2.84
21	2.08	2.41	2.83
22	2.07	2.41	2.82
23	2.07	2.40	2.81
24	2.06	2.39	2.80
25	2.06	2.38	2.79
26	2.06	2.38	2.78
27	2.05	2.37	2.77
28	2.05	2.37	2.76
29	2.04	2.36	2.76
30	2.04	2.36	2.75



We have to check this Null Hypothesis-
H0: There is no statistically significant difference between the samples. To prove that we need a critical value if lower then t- value then we Don't reject the Null Hypothesis or if critical value is higher than t-value then we will reject Null Hypothesis.

Formulating Research Problem:

The certain aim of this chapter is to detail the process of problem formulation. The specific process that you are likely to adopt depends upon:

- Your expertise in research methodology
- Your knowledge of the subject area
- Your understanding of the issues to be examined
- The extent to which the focus of your study is predetermined 2

The research problem A research problem can be any question that you want to answer and any assumption or assertion that you want to challenge or investigate.
'Potential research questions may occur to us on a regular basis, but the process of formulating them in a meaning way is not at all an easy task'.

The importance of formulating a research problem The formulation of a research problem is the first and most important step of the research process. Kerlinger (1986: 17) 'If one wants to solve a problem, one must generally know what the problem is. It can be said that a large part of the problem lies in knowing what one is trying to do.

The ways you formulate a problem determine almost every step follow:

- The type of study design that can be used
- The type of sampling strategy that can be employed
- The research instrument that can be used or developed
- The type of analysis that can be undertaken The importance of formulating a research problem (con) 6

Sources of research problems:

Most research in the humanities revolves around four Ps:

- People
- Problems
- Programs
- Phenomena.

Considerations in selecting a research problem:

When selecting a research problem/topic there are a number of considerations to keep in mind.

There are seven main points such as:

• ***Interest:*** should be the most important consideration in selecting a research problem.

• ***Magnitude:*** should have sufficient knowledge about the research process to be able to visualize the work involved in completing the proposed study. Considerations in selecting a research problem (con)

• ***Measurement of concepts:*** if you are using a concept in your study, make sure you are clear about its indicators and their measurement. For example: if you plan to measure the effectiveness of a health promotion program, you must be clear as to what determines effectiveness and how it will be measured. Do not use concept in your research problem that you are not sure how to measure.

• ***Level of expertise:*** Make sure you have an adequate level of expertise for the task you are proposing.

• ***Relevance:*** select a topic that is of relevance to you as a professional. Ensure that you study adds to the existing body of knowledge, bridges current gaps or useful in policy formulation.

• ***Availability of data:*** if you topic entails collection of information from secondary sources (office record, client , records, census or other already-

polished reports, ect..) before finalizing your topic make sure that these data are available and in the format you want.

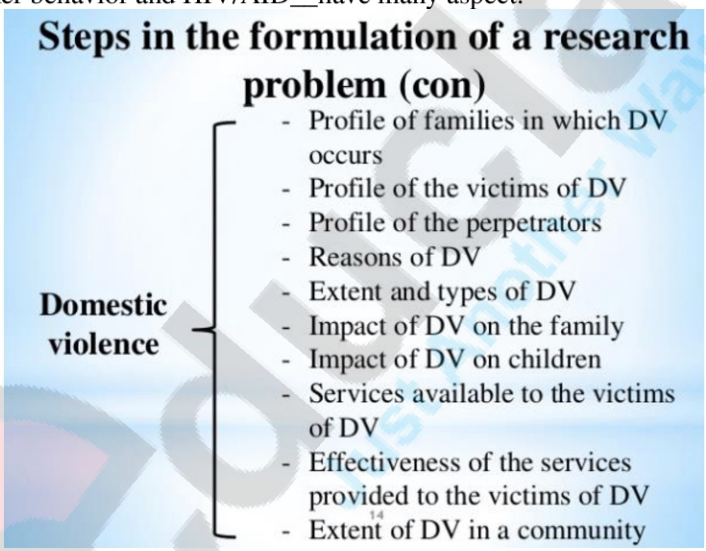
• ***Ethical issues:*** other important consideration in formulating a research problem is the ethical issues involved

Steps in the formulation of a research problem:

A step in the formulation of a research problem is the most crucial part of the research journey on which the quality of the entire project depends. Steps in formulating research problem alcoholism

Step1: Identify a broad field or subject area of interest to you. Asked yourself, what is it that really interest me as a professional?

Step 2: Dessert the broad area into subareas You will relies that all the broad areas mentioned above __youth welfare, refugees, domestic violence, consumer behavior and HIV/AID__have many aspect.



Step 3: Select what is of most interest to you. - It is neither advisable nor feasible to study all subareas. - Select issues or subareas about which you are passionate.

Step 4: Raise research questions

1. What is it that I want to find out about in the subareas?
2. Asked the question what you want to find yourself in a situation. Steps in the formulation of a research problem (con) 15

Step 5: Formulate your main objectives and your sub objectives - The main difference between objectives and research questions is in to behavioural aims by using action –oriented words such as to find out, to determine' , 'to ascertain and 'to examine'

Step 6: Assess your objectives - Now examine your objectives to ascertain the feasibility of achieving them through your research endeavour. - Consider them in the light of the time, resources (financial and human) and technical expertise at your disposal. 16

Step 7: Double-check. - Go back and give final consideration to whether or not you are sufficiently interested in the study, and have adequate resources to undertake it. - Ask yourself, am I really enthusiastic about this study? - Do I really have enough resources to undertake it? - Answer these questions thoughtfully and realistically.

The formulation of objectives

- Objectives are the goals you set out to attain in your study.
- Objectives inform a reader of what you want to achieve through the study, it is extremely important to word them clearly and specifically.
- Objectives should be listed under two headings: - main objectives; - sub objectives

The main objective is an overall statement of the trust of your study. - It is also a statement of the association and relationships that you seek to discover or establish. - The subjective are the specific aspects of the topic that you want to investigate within the main framework of your study.

Clear Complete Specific Identify the main variables to be correlated Identify the direction of the relationship Characteristics of objectives

Establishing Operational Definitions in every study there are two components: the subject area and the study population (discussed in sources research problems' earlier in this chapter) - The main aim of formulating a research problem is to clearly and precisely define the research problem. - In a research study it is important to develop, define or establish a set of rules, indicators or yardsticks in order to clearly establish the meaning of such words/items.

Examples studies help to explain the main objectives: - to find out the number of children living below the poverty line in Australia; - to ascertain the impact of immigration on family roles among immigrants; - to measure the effectiveness if a retraining program designed to help young people Although these objectives clearly state the main thrust of the studies, they are not specific in terms of the main variables to be studied and the study populations.

Generalization and Interpretation of analysis:

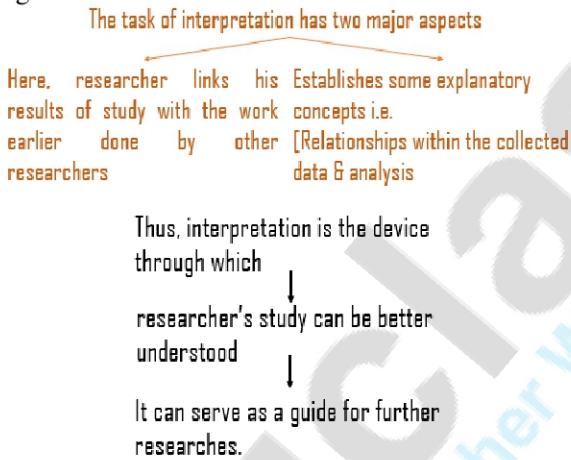
Generalization:

1. Generalization is an essential component of the scientific process. In an ideal world, to test a hypothesis, researcher would sample an entire population.
2. Researcher selects a representative group that reflect the whole population.

3. Researcher must ensure that the sample group is as truly representative of the whole population as possible.

Interpretation:

Refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. It is a search for broader meaning of research findings.



Interpretation is the technique of how generalization should be done and concepts be formulated.

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz.,

(i) The effort to establish continuity in research through linking the results of a given study with those of another, and

(ii) The establishment of some explanatory concepts. "In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses." Thus, interpretation is the device through which the factors that seem to explain what has been observed by researcher in the course of the study can be better understood and it also provides a theoretical conception which can serve as a guide for further researches.

WHY INTERPRETATION?

Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic component of research process because of the following reasons:

(i) It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies, having the same abstract principle, and thereby can predict about the concrete world of events. Fresh inquiries can test these predictions later on. This way the continuity in research can be maintained.

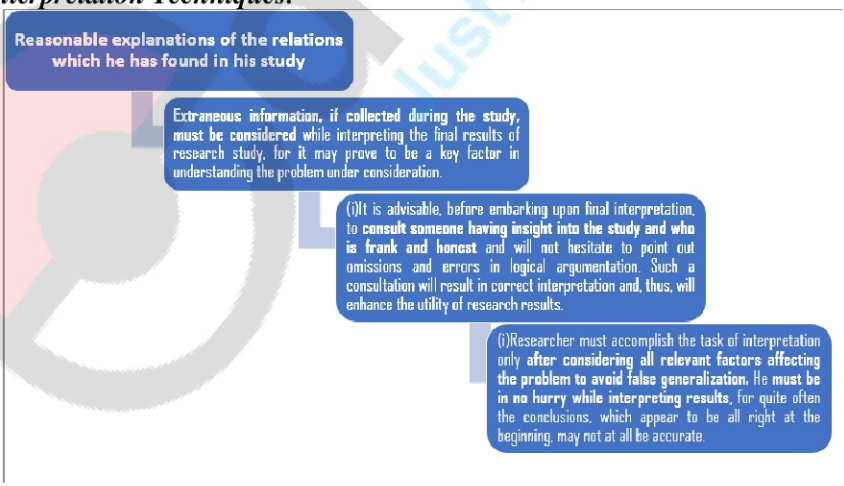
(ii) Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.

(iii) Researcher can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.

(iv) The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such a study have to be interpreted on a post-factum basis in which case the interpretation is technically described as 'post factum' interpretation.

Generalization and Interpretation of analysis:

Interpretation Techniques:



TECHNIQUE OF INTERPRETATION:

The task of interpretation is not an easy job, rather it requires a great skill and dexterity on the part of researcher. Interpretation is an art that one learns through practice and experience. The researcher may, at times, seek the guidance from experts for accomplishing the task of interpretation. The technique of interpretation often involves the following **steps**:

(i) Researcher must give reasonable explanations of the relations which he has found and he must interpret the lines of relationship in terms of the underlying processes and must try to find out the thread of uniformity that lies under the surface layer of his diversified research findings. In fact, this is the technique of how generalization should be done and concepts be formulated.

(ii) Extraneous information, if collected during the study, must be considered while interpreting the final results of research study, for it may prove to be a key factor in understanding the problem under consideration.

(iii) It is advisable, before embarking upon final interpretation, to consult someone having insight into the study and who is frank and honest and will not hesitate to point out omissions and errors in logical argumentation. Such a consultation will result in correct interpretation and, thus, will enhance the utility of research results.

(iv) Researcher must accomplish the task of interpretation only after considering all relevant factors affecting the problem to avoid false generalization. He must be in no hurry while interpreting results, for quite often the conclusions, which appear to be all right at the beginning, may not at all be accurate.