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.

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

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

  2. **Magnitude**: should have sufficient knowledge about the research process to be able to visualize the work involved in completing the proposed study.

  3. **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.

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

5. **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.

6. **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.

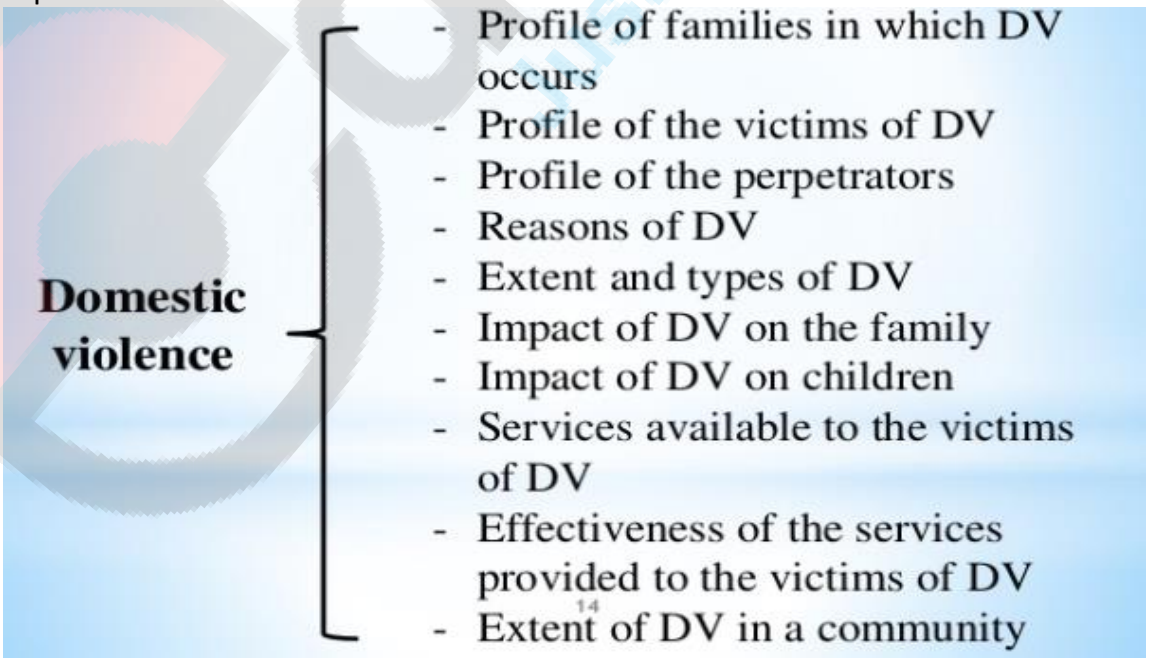
7. **Ethical issues:** other important consideration in formulating a research problem is the ethical issues involved.

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 (Divide) 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.

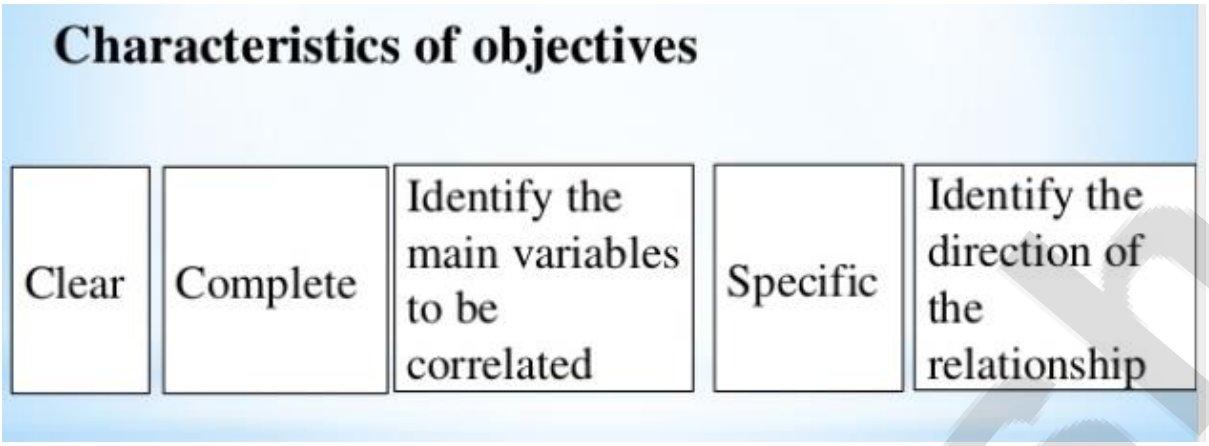
➢ **Step 4:** Raise research questions
- What is it that I want to find out about in the subareas?
- Asked the question what you want to find yourself in a situation.

➢ **Step 5:** Formulate objectives
Formulate your main objectives and your sub objectives
- The main difference between objectives and research questions is in to behavioral 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 o achieving them through your research endeavor.
- 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. 17

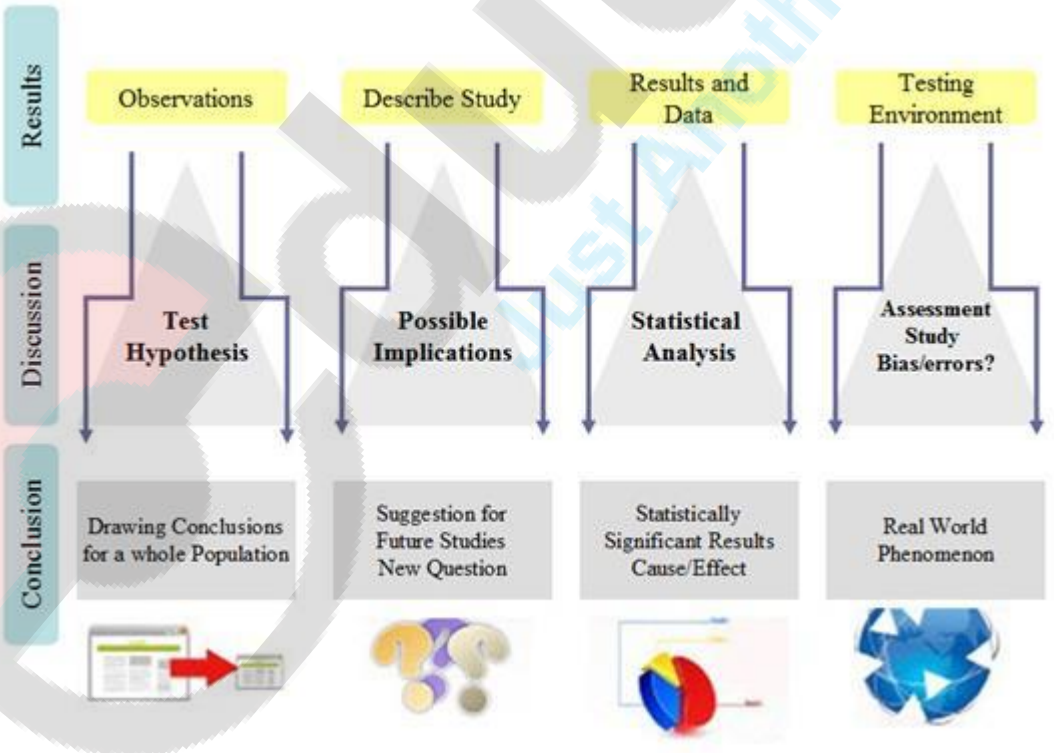
The formulation of objectives
- Objectives are the goals you set out to attain in your study.
- Objectives inform a reader of what you want to achieve through the study, it is extremely important to word them clearly and specifically.
- Objectives should be listed under two headings: -
    1. main objectives;
    2. 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.



Generalization and Interpretation of analysis

Generalization

- Generalization is an essential component of the scientific process. In an ideal world, to test a hypothesis, researcher would sample an entire population.
- Researcher selects a representative group that reflect the whole population.
- 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.
- The task of interpretation has two major aspects
  1- Here, researcher links his results of study with the work earlier done by other researchers
  2- Establishes some explanatory concepts i.e. [Relationships within the collected data & analysis

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

Thus, interpretation is the device through which→researcher's study can be better understood →It 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.

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.

1. Measures of Central Tendency

- **Mean(Average):**

  Find the mean, median, mode, and range for the following list of values:

  13, 18, 13, 14, 13, 16, 14, 21, 13

  The mean is the usual average, so I'll add and then divide:

  (13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13) ÷ 9 = 15

- **Median(Middle/Mid-point)**

  The median is the middle value, so first I'll have to rewrite the list in numerical order: 13, 13, 13, 13, 14, 14, 16, 18, 21

  There are nine numbers in the list, so the middle one will be the

  (9 + 1) ÷ 2 = 10 ÷ 2 = 5th number:

  13, 13, 13, 13, 14, 14, 16, 18, 21

  So the median is 14.

- **Mode:**

  The mode is the number that is repeated more often than any other, so **13** is the mode.

2. Measures of Dispersion

- **Range(R):**

  The largest value in the list is 21, and the smallest is 13, so the range is

  21 – 13 = 8.

- **Mean Deviation :**

$$\text{Mean deviation from mean} = \frac{\sum |X - \bar{X}|}{N}$$

$$\text{Mean deviation from median} = \frac{\sum |X - M|}{N}$$

1. **Mean Deviation from Mean[$\bar{x}$]**

A student took 5 exams in a class and had scores of 93, 65, 85, 90, and 78. Find the mean deviation for her test scores?

Solution:

N = 5

$$\therefore \bar{x} = \frac{\sum X}{N}$$

$$= \frac{93+65+85+90+78}{5}$$

$$= \frac{411}{5}$$

Mean = 82.2

Now lets subtract the mean from each score, take the absolute value of each difference, total the absolute values and then divide by the number of values.

x	$X - \bar{X}$	$\lvert X - \bar{X} \rvert$
93	10.8	10.8
65	-17.2	17.2
85	2.8	2.8
90	7.8	7.8
78	-4.2	4.2
$\bar{X} = 82.2$		$\sum \lvert X - \bar{X} \rvert$ = 42.8

So, Mean deviation = $\sum = \frac{\lvert X - \bar{X} \rvert}{N}$

$$= \frac{42.8}{5}$$

= 8.56

By this we can say that on the average, this student's test scores are deviated by 8.56 points from the mean.

2. **Mean Deviation from Median[M]**

Find the mean deviation about the median for the following data:
3, 9, 5, 3, 12, 10, 18, 4, 7, 19, 21.

Arranging data in ascending order,

3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21

Here, number of observations = n = 11 (odd).

Since **n is odd**,

Median = $(\frac{11+1}{2})^{th}$ observation

M = $(\frac{12}{2})^{th}$ observation

= 6^{th} observation

= 9

Now, we calculate mean deviation about median

3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21

Median (M) = 9

$$\text{Mean deviation about median} = \frac{\sum|x_i - M|}{11}$$

$$\text{M.D.(M)} = \frac{\begin{pmatrix}|3-9| + |3-9| + |4-9| + |5-9| + |7-9| + |9-9| \\ + |10-9| + |12-9| + |18-9| + |19-9| + |21-9|\end{pmatrix}}{11}$$

$$= \frac{\begin{pmatrix}|-6| + |-6| + |-5| + |-4| + |-2| + |0| \\ + |1| + |3| + |9| + |10| + |12|\end{pmatrix}}{11}$$

$$= \frac{(6+6+5+4+2+0+1+3+9+10+12)}{11}$$

$$= \frac{(58)}{11}$$

$$= 5.27$$

- **Standard Deviation:**

Let's say we wanted to calculate the standard deviation for the amounts of gold coins pirates on a pirate ship have.

There are 100 pirates on the ship. In statistical terms this means we have a population of 100. If we know the amount of gold coins each of the 100 pirates have, we use the **standard deviation equation for an entire population:**

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

$where$

$\sigma = the\ standard\ deviation$

$x = each\ value\ in\ the\ population$

$\overline{x} = the\ mean\ of\ the\ values$

$N = the\ number\ of\ values\ (the\ population)$

What if we don't know the amount of gold coins each of the 100 pirates have? For example, we only had enough time to ask 5 pirates how many gold coins they have. In statistical terms this means we have a sample size of 5 and in this case we use the **standard deviation equation for a sample of a population:**

$$s = \sqrt{\frac{\sum(x - \overline{x})^2}{N - 1}}$$

$where$

$s = the\ standard\ deviation$

$x = each\ value\ in\ the\ sample$

$\overline{x} = the\ mean\ of\ the\ values$

$N = the\ number\ of\ values\ (the\ sample\ size)$

The rest of this example will be done in the case where we have a sample size of 5 pirates, therefore we will be using the standard deviation equation for a sample of a population.

Here are the amounts of gold coins the 5 pirates have:

4, 2, 5, 8, 6.

Now, let's calculate the standard deviation:

1. Calculate the mean:

$$\begin{aligned}\overline{x} &= \frac{\sum x}{N} \\ &= \frac{x_1 + x_2 + \cdots + x_N}{N} \\ &= \frac{4+2+5+8+6}{5} \\ &= 5\end{aligned}$$

2. Calculate $x - \overline{x}$ for each value in the sample:

$$x_1 - \overline{x} = 4 - 5 = -1$$

$$x_2 - \overline{x} = 2 - 5 = -3$$

$$x_3 - \overline{x} = 5 - 5 = 0$$

$$x_4 - \overline{x} = 8 - 5 = 3$$

$$x_5 - \overline{x} = 6 - 5 = 1$$

3. Calculate $\sum(x - \overline{x})^2$:

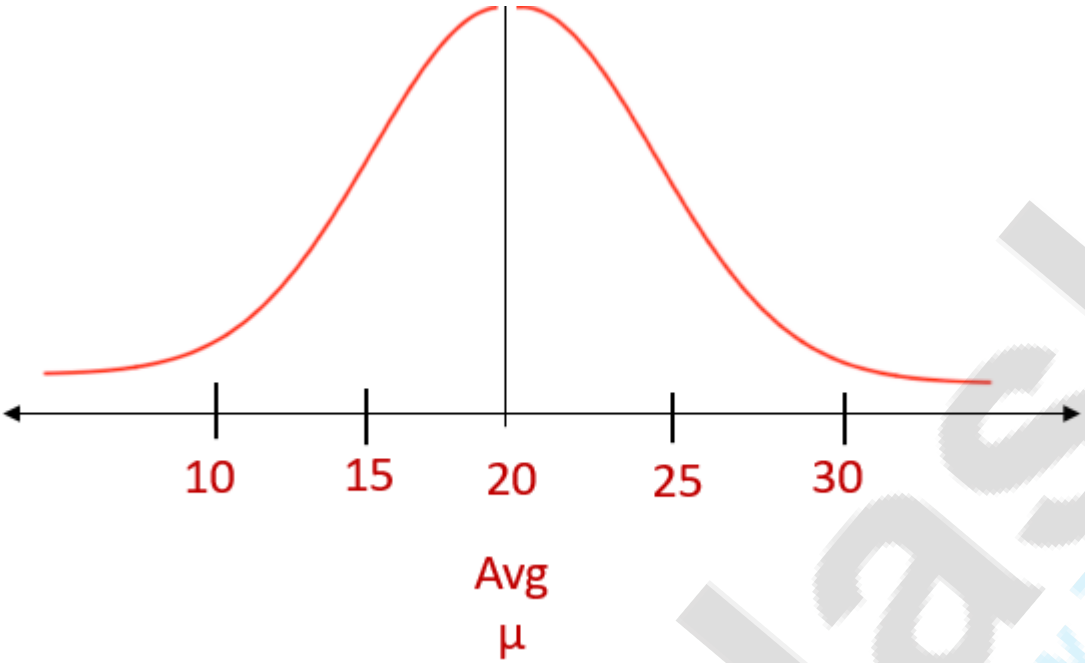
$$\begin{aligned}\sum(x - \overline{x})^2 &= (x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_N - \overline{x})^2 \\ &= (-1)^2 + (-3)^2 + 0^2 + 3^2 + 1^2 \\ &= 20\end{aligned}$$

4. Calculate the standard deviation:

$$\begin{aligned}s &= \sqrt{\frac{\sum(x - \overline{x})^2}{N-1}} \\ &= \sqrt{\frac{20}{5-1}} \\ &= 2.24\end{aligned}$$

The standard deviation for the amounts of gold coins the pirates have is 2.24 gold coins.

3. Measures of Asymmetry



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

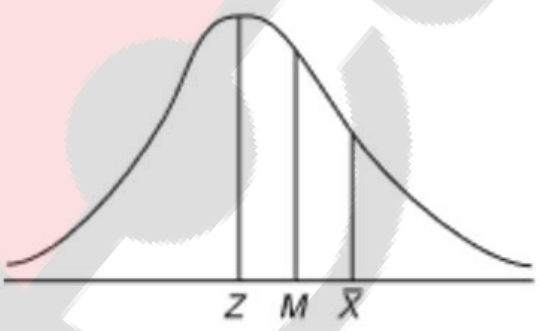
2. Since the shape of the curve is just like bell that's why it is called as **Bell Curve.** Lager population value is right around the mean.

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.

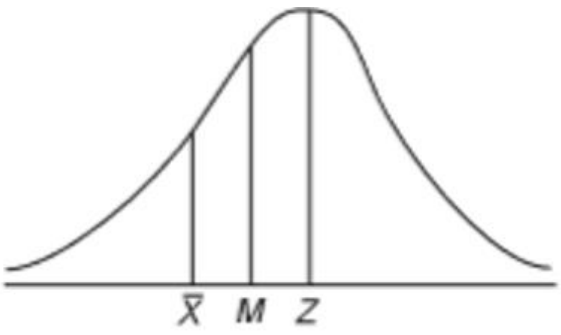
5. This graph is completely defined by mean (μ) and standard deviation (σ).

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



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

Curve is distorted on the right side



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

Curve is distorted on the left side

Skewness is a measure of asymmetry and shows the manner in which the items are clustered around the average.

4. Chi-square Test

There are **two types of chi-square tests**. Both use the chi-square statistic and distribution for different purposes:

- A **chi-square goodness of fit test** determines if a sample data matches a population. For more details on this type, see: *Goodness of Fit Test*.
- A **chi-square test for independence** compares two variables in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of categorical variables differ from each another.
  - A **very small chi square test statistic** means that your observed data fits your expected data extremely well. In other words, there is a relationship.
  - A **very large chi square test statistic** means that the data does not fit very well. In other words, there isn't a relationship.

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

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

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),

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

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

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 We know that

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

0.900 < P-value < 0.950

Level of significance = 0.05

P-value > Level of significance Null Hypothesis is accepted.

5. T-tests

- 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 handling 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.

$$t\ value = \frac{Difference\ between\ two\ means}{variability\ in\ two\ groups}$$

$$= \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

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

${\sigma_1}^2$ & ${\sigma_2}^2$= standard deviations of two different populations

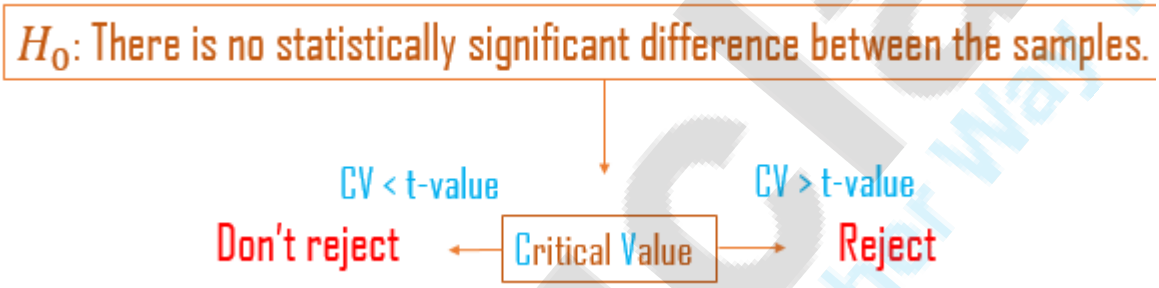
n_1 & n_2 = numbers of sample

$$= \frac{|15.38 \quad -15.68|}{\sqrt{\frac{0.097}{16} + \frac{0.165}{16}}} \quad = \frac{0.3}{.13} \quad = 2.3$$

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.



We know that **degree of significance = 0.05**

t- value = 2.3

$\boldsymbol{DF = n_1 + n_2 - 2 \quad = 16 + 16 - 2 \quad = 30}$

Critical value = 2.04

2.04 < 2.3

Critical value < t-value

Null Hypothesis is Accepted.

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