A Project Report on

SOCIO-ECONOMIC STATUS AND IMPACT OF DRUDGERY ON HEALTH OF FARMING WOMEN

The Maharaja Sayajirao University of Baroda, Gujarat

Faculty of Science

Department of Statistics

GUIDED BY,

Dr. MANISHA PANDYA

&

Mr. SHREY PANDYA

PRESENTED BY,

Vishal Harijan,

Rajdeep Chhasatiya,

Darpan Padaria,

Dipak Bariya

ABSTRACT

The Present study shows the effect on the health of a farming women. Implementation data has been obtained from women farmers of Kareli and Samoj villages of Jambusar taluka of Bharuch district. Data were gathered from 349 women farmers through a well-structured questionnaire, focused group discussion and personal interview. Findings revealed that women farmers use traditional tools and implements since a long time but most of the women farmers felt immense drudgery in their use. In this study the effect of drudgery work on a woman farmer has been shown. Apart from this, it has also been shown which woman farmers are facing which physical problem due to which farming activities. And this study also shows the socio-economic status of women farmers.

This study can provide scope for promotion of technology in gender perspective towards the challenges of farmwomen would help in reducing drudgery and occupational health problems of women workers in agriculture.

CERTIFICATE

This is to certify that Vishal Harijan, Rajdeep Chhasatiya, Darpan Padaria, and Dipak Bariya have successfully and satisfactorily completed the project titled:

"SOCIO-ECONOMIC STATUS AND IMPACT OF DRUDGERY ON HEALTH OF FARMING WOMEN"

as a team for the academic year 2020-21 and have submitted the work to the Department of Statistics in fourth semester as a partial fulfilment for the degree of Master of Science in Statistics and have represented their original work.

I wish them a grand success in future.

Mr. SHREY PANDYA	Dr. MANISHA PANDYA	Dr. V.A. KALAMKAR
Guide	Guide	Head of Department
Assistant Professor	Professor	

Department of Statistics

Faculty of Science

The Maharaja Sayajirao University of Baroda

ACKNOWLEDGEMENT

This work could not have been completed without the help of these pioneers of statistics.

Firstly, we would like to express our gratitude to **Dr. Manisha Pandya and Mr. Shrey Pandya.** As students of Master's program in Statistics, we have had the privilege to work under their excellent guidance. We have got all kinds of support and tremendous experience in collection, exploration and analysis of data.

We are extremely grateful to all the people of both villages for providing us with the data in form of their opinions in our questionnaire. They were highly co-operative and patient in providing us the same.

We are very much thankful to all the Faculties and Staff Members of Statistics Department, Faculty of Science, The M.S. University of Baroda.

The support of these many people will always remain permanently attached in our memory and we are extremely grateful to them. Despite the help of so many able people, we alone accept full responsibility for any deficiencies the project may possess.

CONTENTS

INTRODUCTION

India is a developing country. Here the main occupation is agriculture, because of 70 percent of the population is involved in this occupation. Many women in developing countries are occupied in agriculture. The rural women play a significant role in agriculture and other agro based activities. The daily work schedule of rural women is very demanding and arduous. Also in agriculture women are the major workers assigned the heavy drudgery work, such as seeding, transplanting, weeding, fertilizer & manures application, plant protection, thinning, harvesting, processing, selling, winnowing, storing, etc.

Agriculture is a primary unorganized sector in which women farm workers perform the majority of the drudgery prone work. Such work is often called "drudgery work", which means dull and boring work, but it is also heavy work, back-breaking work, often dangerous work. Women themselves consider this drudgery work as normal, which is in fact part of the problem.

These works which lead to "drudgery" is conceived as physical and mental strain, agony, monotony and hardship experienced by human beings while all of the women in the is regard is alarming as they continue to be constrained by illiteracy, malnutrition, and unemployment. Many believe that women's involvement in agricultural tasks and large is a source of heavy burden of drudgery on them. The farmwomen perform agricultural tasks with the age-old traditional tools since gender friendly appropriate tools are either not available or insufficient in number or unawareness. Unsafe, hazardous, unhealthy and long hours of work with age-old traditional and cumbersome tools accelerate health related problems, especially among women farmers. Farmers/ farmwomen are not always aware of the improvements they could make by using scientific and technological knowledge. Thus, the attention of farmwomen was directed towards the women friendly improved farm tools. Farmwomen from the unorganized sector are vulnerable as new and improved technologies are inaccessible for them. It is imperative that they are exposed to these technologies and motivated to adopt the new technologies, which would help them to improve their quality of life.

Women are the backbone of agricultural work force because they perform more than 79% of farm activities. Farmwomen often lack education and information on the health hazards and habitually view pain as a normal part of work and seek care only when the condition becomes severe or disabling. Usually, they do not understand the association of a health problem with its source. Further, Women

being over burdened with so much workload both on farm and at home; they usually neglect their health. Some of the reasons responsible for health hazards are labour intensive field operations, excessive reliance on human power, low level of adoption of drudgery reducing implement, low productivity of human labour, difficult nature of work and decrease in the labour available for agriculture. Other factors might be poverty, inadequate training and lack of awareness, which delay to deal with the occupational related health problems.

The health of farmwomen is one of the important resources for agricultural development. Therefore, drudgery reduction measures needs to be initiated to avoid occurrence of health hazards among farmwomen. Hence, an urgent need to make women aware about latest drudgery reducing tools, implements and other technologies and motivate them to adopt the same was felt. If appropriate drudgery reduction technologies are made available to the farmwomen at home and farm, it would definitely contribute in reducing their drudgery, increasing their working capability, increasing farm production resulting in improved quality of life. Several types of drudgery reducing technologies are available in market but to what extent these are being used by farmwomen and whether farmwomen are aware about these technologies or not are the questions of investigation. To increase agricultural production, processing, marketing and generate livelihood opportunities, it is necessary to trained women farmers to use traditional and modern ICT tools to improve their working efficiency.

The general trend existing in rural India is limited resources available to women because of low socio-economic status in the society and within that limited access to resources, there exists a strong disparity that, most of the women's earnings goes towards nutritional security of the households.

Most women cannot invest in the technology. Introduction to new technologies in agricultural operation adopted by farm women leading to the mechanization will reduce the drudgery and improve the efficiency.

Aatapi Seva Foundation strives to facilitate the holistic and sustainable development on socio-economic status, health of farming women and impact assessment of drudgery reduction tool through need based intervention voiced by the community.

REVIEW OF LITERATURE

Drudgery Reduction

- Women as farmer or farm workers, participate in several activities such as seeding, transplanting, weeding, fertilizer & manures application, plant protection, thinning, harvesting, processing, selling, winnowing, storing, etc. (Sudharani and Raju, 1991Verma and Sinha, 1991; Sudharani and Raju, 1991; Begam, 2000; Oberoi and Singh, 2001; Rani, 2007; Mukherjee, 2014).

- These works which lead to "drudgery" is conceived as physical and mental strain, agony, monotony and hardship experienced by human beings while all of the women in the is regard is alarming as they continue to be constrained by illiteracy, malnutrition, and unemployment (Armstrong, 1983; Nag and Nag, 2004)

- Many believe that women's involvement in agricultural tasks and large is a source of heavy burden of drudgery on them (Verma and Sinha, 1991)

- The farmwomen perform agricultural tasks with the age-old traditional tools since gender friendly appropriate tools are either not available or insufficient in number or unawareness. Unsafe, hazardous, unhealthy and long hours of work with age-old traditional and cumbersome tools accelerate health related problems, especially among women farmers (Nag and Nag, 2004)

- Farmers/ farmwomen are not always aware of the improvements they could make by using scientific and technological knowledge. Thus, the attention of farmwomen was directed towards the women friendly improved farm tools (Gite and Singh, 2005; Nag and, 2004; AICRP, 2009; Anon, 2010; Patel et al., 2015).

- The study suggests that the workload of women (e.g., time and energy) in the farming activities can be reduced in two ways: (1) making existing tasks easier or increasing the productivity of existing labor, or (2) changing farm practices with new technology (Urmila Aryal and Rishi Ram Kattel)

- While there is a need to promote the use of more innovative techniques and machineries, one also had to be mindful of the fact that an overemphasis on

introducing machines to reduce drudgery among women could adversely impact the demand for labour. Hence, in this situation, the program felt that semi-automated machines were the most feasible option since they reduced drudgery thereby increasing productivity, were more affordable than fully automated machines and kept the demand for women workers high thus, not taking away their livelihoods.(CARE india 2016)

- When women have more time to themselves, without affecting the productivity of the SMEs, they are more likely to participate in activities of women collectives and SHGs (CARE india 2016)

- With the growing feminization of farm labour due to male rural to urban migration, women are forced to carry out work previously done by men. Consequently women are increasing their workload and taking care of a wider scope of agriculture tasks, but the degree to which they have access to improved technologies need special consideration. (Jatinder Kishtwaria and Aruna Rana)

It is true that at national level modernization of agriculture is taking place at a rapid pace, but women continue to perform farm operations which are full of drudgery while mechanized operations are performed by men.

NEED FOR STUDY

Many people in our country are involved in the farm business. Many of these women also take part in farm work. Rural women play a significant role in agriculture and agro-based activities. The work done by these women in different farming activities is very tedious as well as time consuming.

Activities like sowing, weeding, irrigation, harvesting, winnowing, threshing, fodder cutting etc. And they are mostly done by manually or by traditional tool. For Example: When women are engaged in sowing activities in agriculture, their waist is always bent during this sowing activity and they do the same position all day long. So standing in the same position for a long time has a very bad effect on their waist. They also have pain in the lower back as well as in the legs and buttocks. In this way all the farming activities have a very bad effect on different parts of the body of women.

In addition to working on the farm, women also have to do household chores. They also have to do chores like farm work, housework, childcare, moving away to fetch water, chopping wood to make meals, etc. So all these factors result in women suffering from physical and mental fatigue, exploitation, pain etc.

Also traditional tools used by women workers involves operating in bending or squatting posture which cause drudgery and lead to serious health issues such as back pain, knee pain and sometime also causes injury to women operating it. Therefore, there is need to develop tools/equipment considering women suitability to work.

Such work is often called "drudgery work", which means dull and boring work, but it is also heavy work, back-breaking work, often dangerous work. Women themselves consider this drudgery work as normal, which is in fact part of the problem.

Here are is the two images of women doing farm activity. (Refer the images given in annexure)

Inside the first image, when the women farmers are transplanting during the paddy harvest, their body position is curved from the waist down. Working in such a monotonous position for a long time has a very bad effect on their body.

While in another image you can see that when transplanting women use a new drudgery tool their body position looks a bit normal and using this new tool they transplant in a very short time, easily and without bent the waist.

By providing drudgery reduction tool these all problems can be solved or minimize the risk of women's health so it is very necessary to first study the impact of different drudgery activities on women's health. And also this will improve rural women's economic condition in long run.

OBJECTIVES

- To study the relationship between different drudgery activity and their impact on physical injuries / problems.
- To study the socio - economic status of farming and non- farming women.
- To study the socio-economic status of SHG (Self Help Groups) and non-SHG groups.

With the help of the first objective, we will look at the relationship between the various drudgery activities carried out by women and also how much it impacts on their health.

With the help of another objective we will see the socio-economic status of farming and non-farming women.

METHODOLOGY

Source of data:

Study is based on primary data.

Target Population:

All women (farming and non-farming) of Kareli and Samoj village of Jambusar taluka in Bharuch district.

Survey Design:

In survey design we have obtained our primary data using questionnaires and for this we had conducted personal interviews with women.

Data on socio-economic status and for different farm activity have been collected by questionnaires.

SAMPLE SIZE DETERMINATION

Sample is the part of the population that helps us to draw inferences about the population. Collecting research of the complete information about the population is not possible and it is time consuming and expensive. Thus, we need an appropriate sample size so that we can make inferences about the population based on that sample.

Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. In practice, the sample size used in a study is usually determined based on the cost, time, or convenience of collecting the data, and the need for it to offer sufficient statistical power.

We have found sample size in three stages here.

Stage 1: Within stage 1 we got sample size using simple random sampling.

In statistics, a simple random sample is a subset of individuals chosen from a larger set in which each individual is chosen randomly and entirely by chance. More Specifically, each individual has the same probability of being chosen at any stage during the sampling process, and each subset of k individual has the same probability of being chosen for the sample as any subset of k individual. This process and technique is known as simple random sampling.

Simple random sampling is basic type of sampling, since it can be a component of other more complex sampling methods. The principle simple random sampling is that every object has the same probability of being chosen.

Frist we have applied simple random sampling to derive total sample size and then by using stratified sampling necessary sample size for kareli and samoj village is derived.

No. of women in Kareli (N1) = 2712 (By 2011 census data)

No. of women in Samoj (N2) = 717 (By 2011 census data)

By Simple random sampling without replacement,

N = Total Population Size = N1 + N2 = 2712 + 717 = 3429

Alpha $(\alpha) = 0.05$

Z score = 1.96

Epsilon (Ɛ) = Margin of error = 0.05

P = 0.83 = proportion

n = Total Sample size

- Simple Random Sampling

$$n = \frac{NZ_{\alpha/2}^2 pq}{(N-1)\mathcal{E}^2 p^2 + Z_{\alpha/2}^2 pq}$$

By Simple Random Sampling our total sample size is 288.

Note: Here, proportion p comes from the pilot survey, where we had asked 20 respondents about, does drudgery work plays a significant role in their life or not?

Stage 2: Stratified Proportional Allocation

Proportional allocation is a procedure for dividing a sample among the strata in a stratified sample survey. A sample survey collects data from a population in order to estimate population characteristics. A stratified sample selects separate samples from subgroups of the population, which are called "strata" and can often increase the accuracy of survey results. In order to implement stratified sampling, it is necessary to be able to divide the population at least implicitly into strata before sampling.

By Simple Random Sampling our total sample size is 288. This 288 sample size belongs to both kareli and samoj villages but now we have made stratified proportional allocation for how many samples both kareli and samoj have to be collected from the villages.
Now, we are dividing our sample into two strata. i.e. we are considering kareli and samoj villages as strata.

$$\mathbf{n_i = \frac{N_i}{N} * n}$$

Village wise sample size:

Table 1:

Village	**Population**	**Sample**	
Kareli	2712	227.7795	228
Samoj	717	60.22047	60

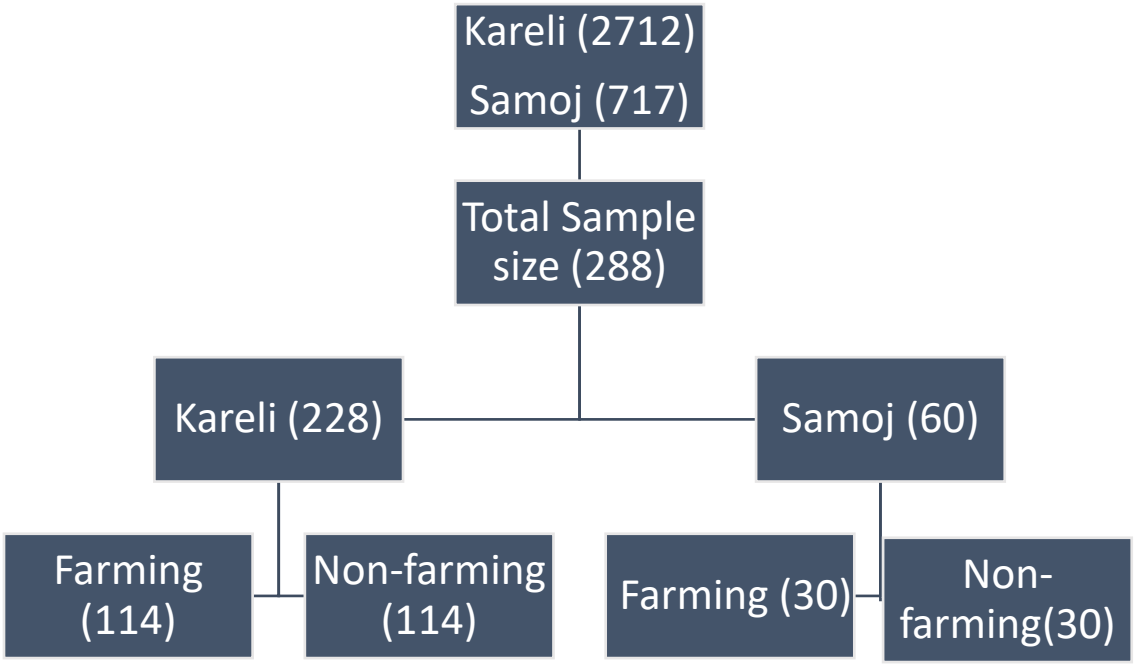
Stage 3: Equal Proportion allocation

To study the drudgery impact on women's health it is necessary to include non-farming women in research, so from both the villages sample size will be further divided into farming and non-farming women. Here we did not know about what is the population of framing and non-farming women of kareli and samoj both villages. So we have collected sample data of framing and non-farming women of both villages by using Equal Proportion allocation.

Table 2:

Village	**Population**	**Sample**		**Farming**	**Non-farming**
Kareli	2712	227.7795	228	114	114
Samoj	717	60.22047	60	30	30

Sampling chart



The above sample size is required theoretical number, but due to sudden rises of covid-19 cases in both the villages we had to stop data collection and we have hardly managed, to get the responses of 125 women.

Therefore, further statistical analysis is based on those 125 respondents.

UDAI-PAREEKH SCALE:

Our first objective is to study the socio-economic status of women involved in farm activity.

Here, socio-economic score has been calculated by the Udai-Pareekh scale. Scoring of the parameters has been stated below.

Table 2: Udai Pareek revised scale

Components	Score	Components	Score
Caste		**Social participation**	
Schedule caste	1	None	0
Lower caste	2	Member of one organization	1
Artisan caste	3	Member of more than one organization	2
Agriculture caste	4	Office holder in such an organization	3
Prestige caste	5	Wide public leader	4
Dominant caste	6	**House**	
Occupation		No house	0
None	0	Hut	1
Labourer	1	Kutcha house	2
Caste occupation	2	Mixed house	3
Business	3	Pucca house	4
Independent profession	4	Mansion	5
Cultivation	5	**Farm power**	
Service	6	No draught animals	1
Education		1-2 draught animals	2
Illiterate	0	3-4 draught animals	4
Can read only	1	5-6 draught animals	6
Can read and write	2	**Material possessions**	
Primary	3	Bullock cart	0
Middle	4	Cycle	1
High school	5	Radio	2
Graduate	6	Chairs	3
And above	7	Mobile phone	4
Land		Television	5
No land	0	Refrigerators	6
<1 acre	1	**Family member**	
1-5 acre	2	Up to 5	2
5-10 acre	3	>5	1
10-15 acre	4		
15-20 acre	5		
≥20	6		

And,
Drudgery Index (DI) is calculated with the following formula,

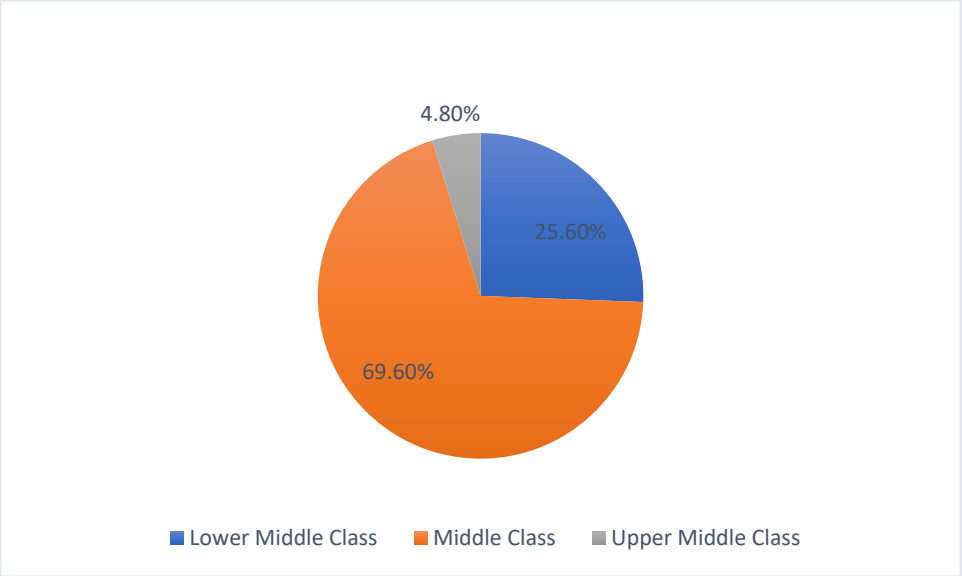
$$Drudgery\ Index = \frac{x+y+z}{3} * 100$$

Where,
x = Co - efficient pertaining to difficulty felt.
y = Co - efficient pertaining time spent in particular farm activity.
z = Co - efficient pertaining to frequency of performance

GRAPHICAL VISUALIZATION

NOTE: Consider DES as the drudgery score and SES as the socio-economic score.

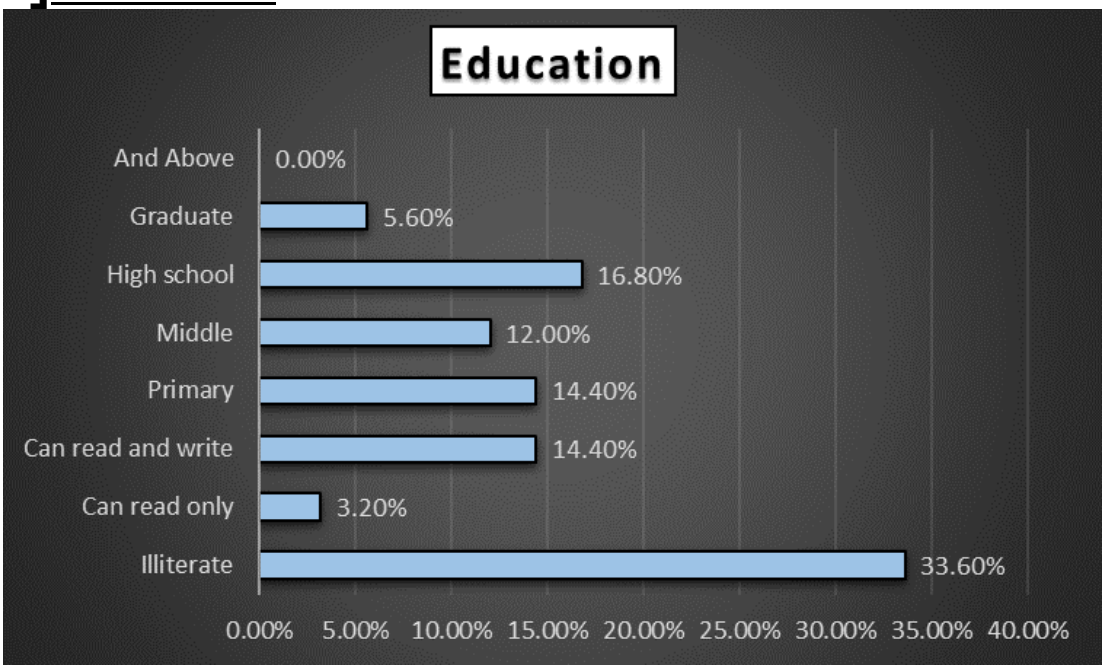
1] Socio- Economic Category



INTERPRETATION:

From the above graph, we can see that 69.60% of women belongs to Middle class family, while after that 25.60% women belongs to Lower middle class family. And 4.80% comes from upper middle class family.

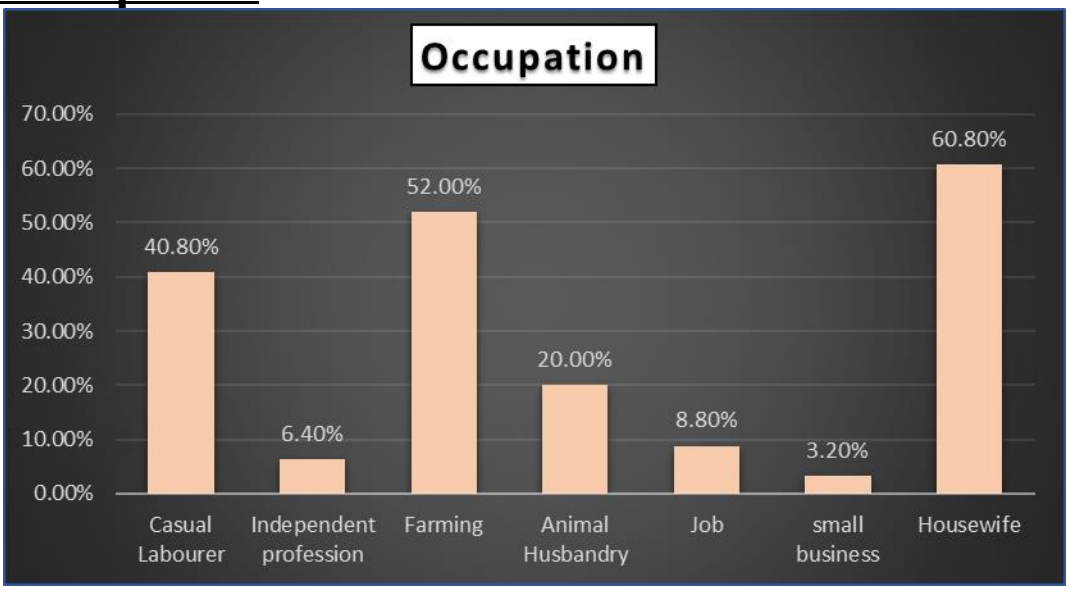
2]Education



INTERPRETATION:

From the graph, we can see that percentage of illiterate women are bigger than any other education category. After that most women completes their primary, middle and high school education.

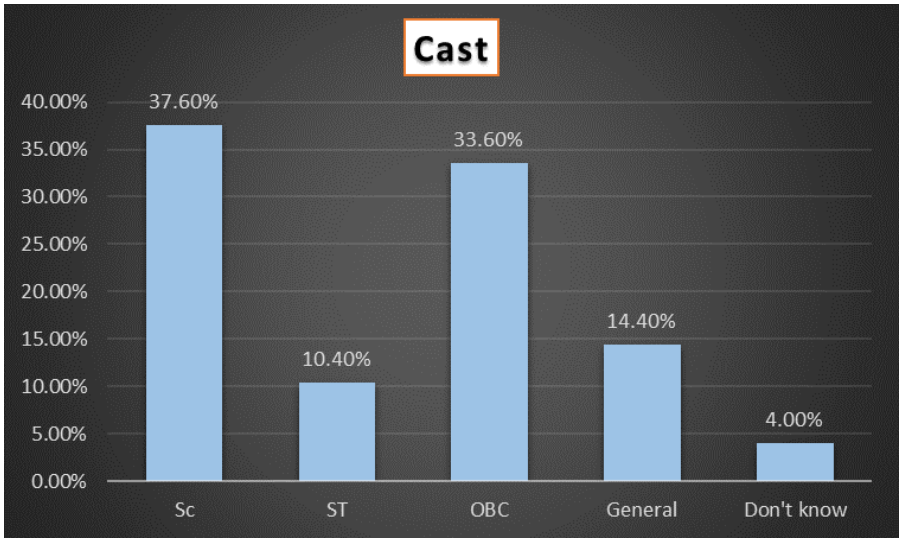
3] Occupation:



INTERPRETATION:

From the graph, one can say that in village, most women generally prefer either to do housework or to work in farm or both.

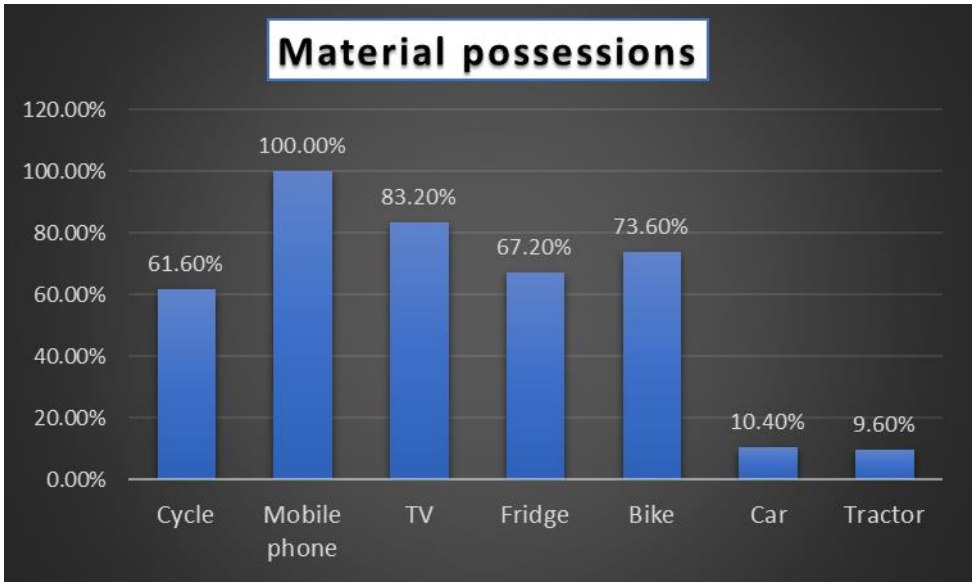
4] Cast:



INTERPRETATION:

From the graph, we can see that, in village there are more population comes from SC and OBC cast.

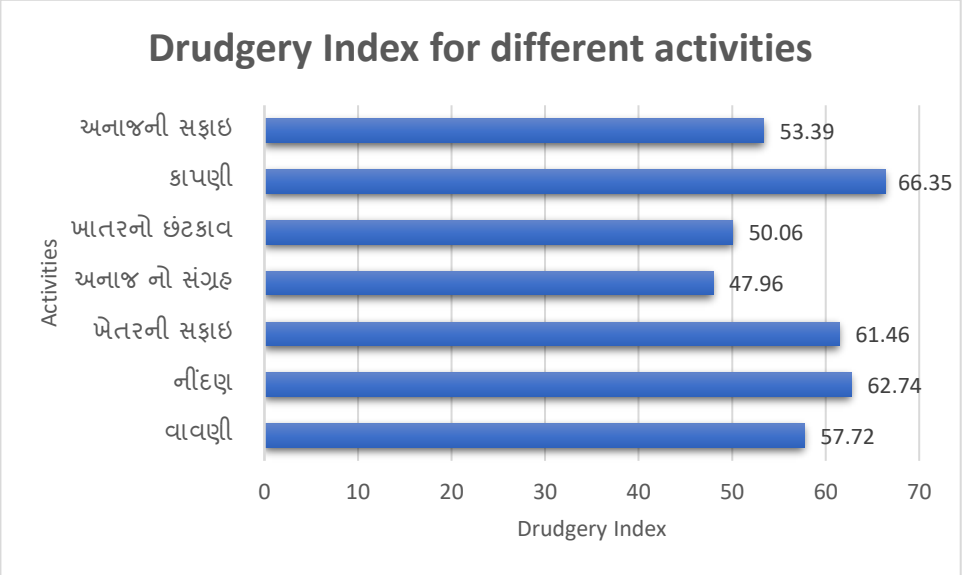
5] Material possession:



INTERPRETATION:

Here 100% women are having mobile phones, 83.20% have TV, 73.60% having a bike, and 67.20% have fridge at home.

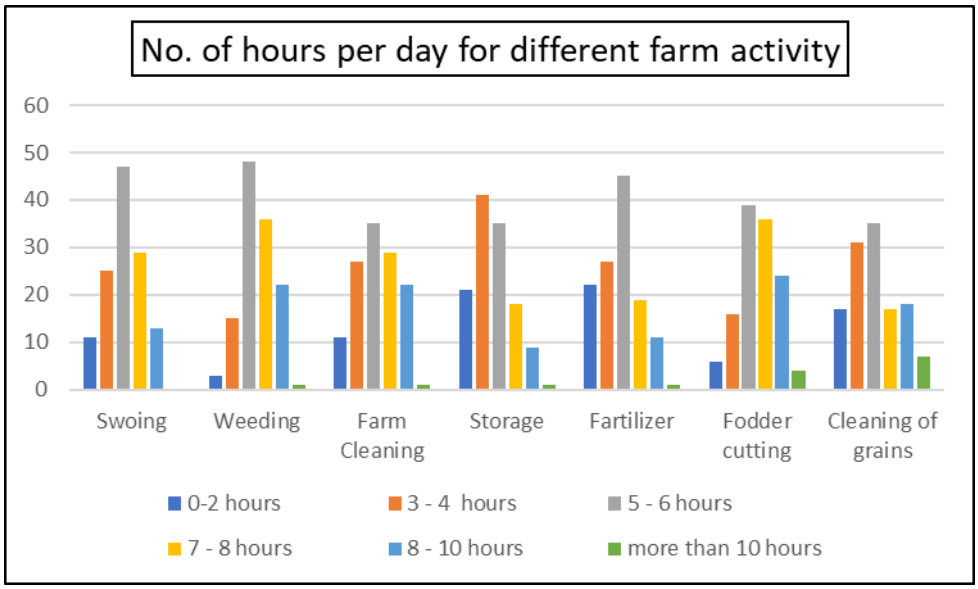
6] Drudgery Index for different Activities:



INTERPRETATION:

Weeding, Farm Cleaning, Sowing, and Fodder cutting are the most drudgery prone farm activities.

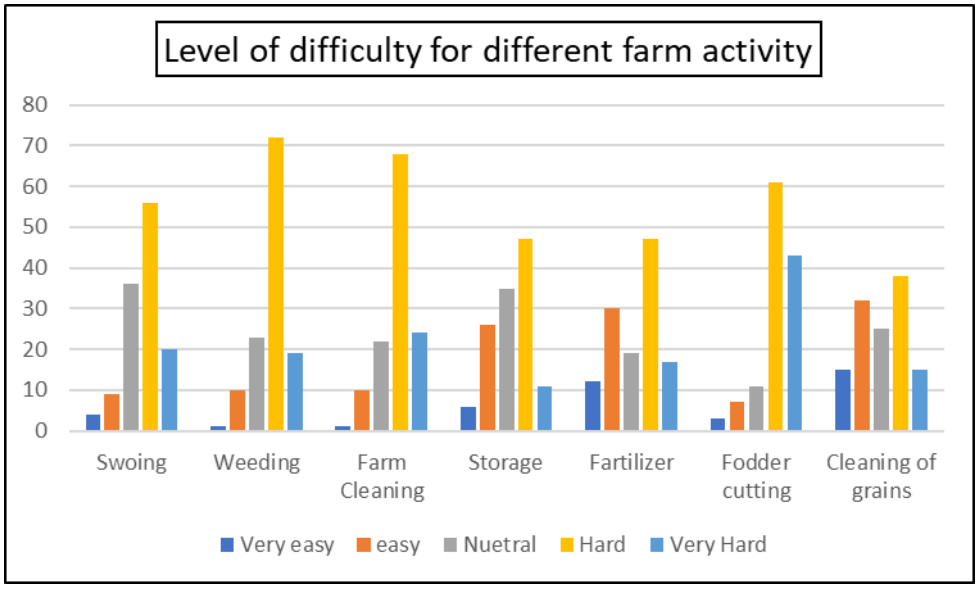
7] No of Hours per day for different activity:



INTERPRETATION:

From the graph, we can observe that they are spending almost 5-8 hours in farm for Storage and Fodder cutting activity.

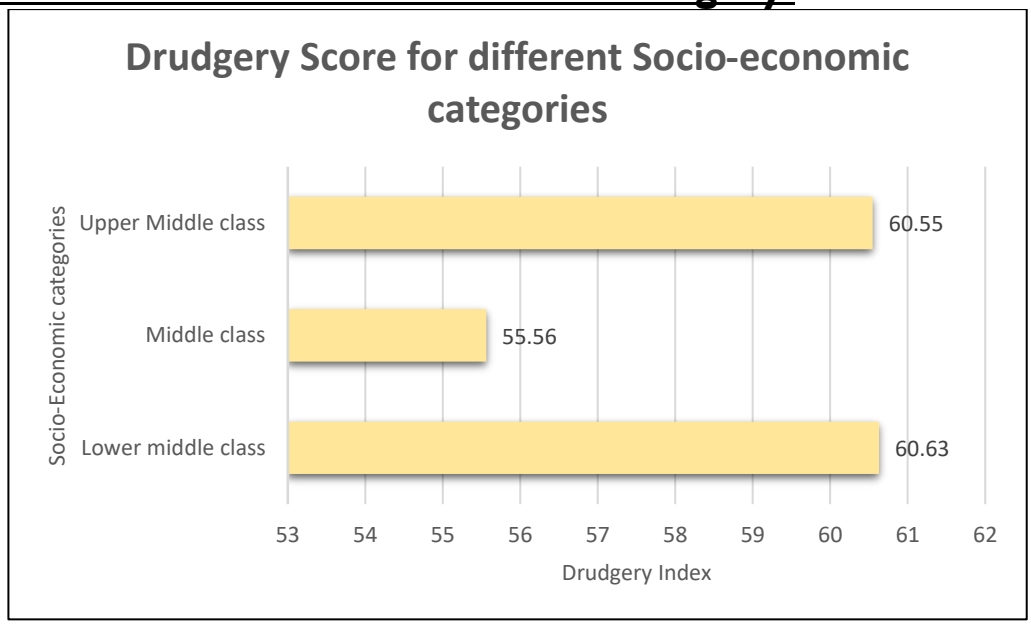
8] LEVEL OF DIFFICULTY FOR DIFFERERNT FARM ACTIVITY:



INTERPRETATION:

From the above graph, most people feels that all activities are hard for them. According to women Fodder cutting activity is the most challenging activity for them.

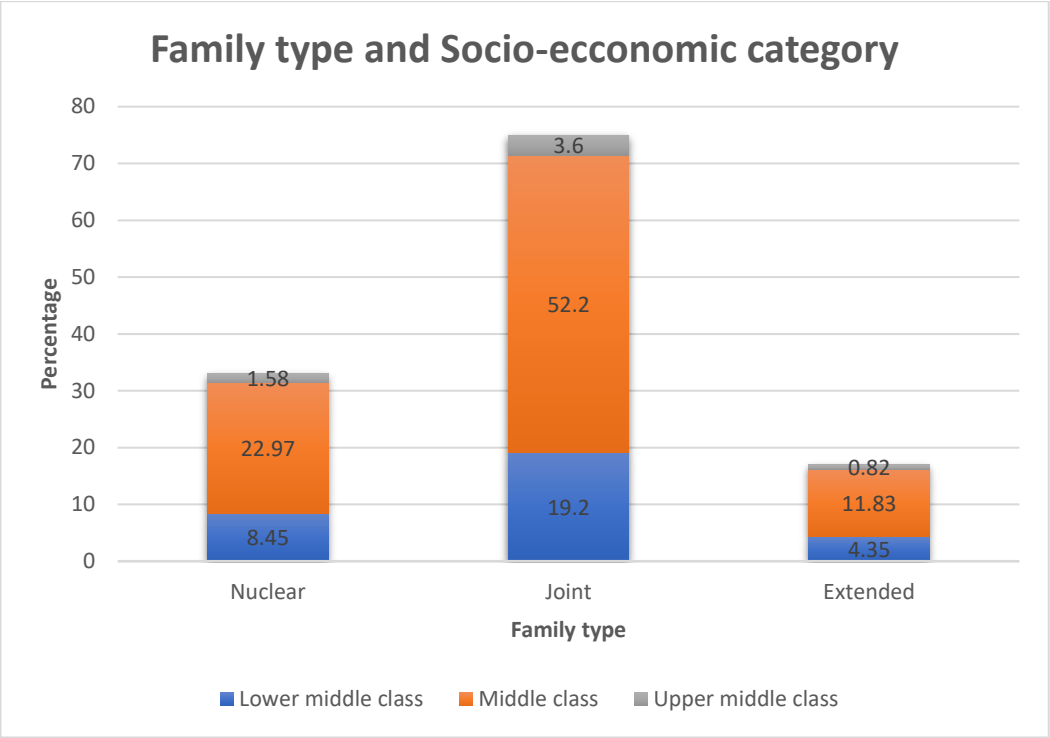
9] Mean of DS and Socio-Economic category:



INTERPRETATION:

The above plot shows that mean drudgery score for lower middle class and upper middle class are almost same.

10] Family type and Socio-Economic Category:



INTERPRETATION:

From the above graph, one can say that women staying in a joint family are belongs to the middle class.

RELIABILITY TEST

Reliability is the overall consistency of a measure. A measure is said to have a high reliability if it produces similar results under consistent conditions. "It is the characteristic of a set of test scores that relates to the amount of random error from the measurement process that might be embedded in the scores. Scores that are highly reliable are precise, reproducible, and consistent from one testing occasion to another. That is, if the testing process were repeated with a group of test takers, essentially the same results would be obtained. Various kinds of reliability coefficients, with values ranging between 0.00 (much error) and 1.00 (no error), are usually used to indicate the amount of error in the scores.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.601	.924	100

Here, we get cronbach's alpha = 0.601, which indicates there are good reliability in our questionnaire.

STATISTICAL ANALYSIS

Normality Test

In statistics, normality tests are used to determine if a data set is well-modeled by a normal distribution and to compute how likely it is for a random variable underlying the data set to be normally distributed.

A normality test is used to determine whether sample data has been drawn from a normally distributed population (within some tolerance). A number of statistical tests, such as the Student's t-test and the one-way and two-way ANOVA require a normally distributed sample population.

An informal approach to testing normality is to compare a histogram of the sample data to a normal probability curve. The empirical distribution of the data should be bell-shaped and resemble the normal distribution. This might be difficult to see if the sample is small. In this case one might proceed by regressing the data against the quantiles of a normal distribution with the same mean and variance as the sample. Lack of fit to the regression line suggests a departure from normality.

Shapiro Wilk Test

The Shapiro–Wilk test is a test of normality in frequentist statistics. It was published in 1965 by Samuel Sanford Shapiro and Martin Wilk.

The Shapiro-Wilk test is a way to tell if a random sample comes from a normal distribution. The test gives you a W value; small values indicate your sample is not normally distributed (you can reject the null hypothesis that your population is normally distributed if your values are under a certain threshold)

Test Statistic:

$$W = \frac{(\sum_{i-1}^{n} aix(i))^2}{\sum_{i=1}^{n}(xi - \bar{x})^2}$$

where

$x_{(i)}$ is the ith order statistic, i.e., the ith-smallest number in the sample;

$\bar{x} = (x_1+x_2+\ldots+x_n)/n$ is the sample mean.

ai are constants generated from the covariances, variances and means of the sample (size n) from a normally distributed sample.

Age

Descriptive statistic for Age

			Statistic	Std. Error
Age	Mean		39.67	1.133
	95% Confidence Interval for Mean	Lower Bound	37.43	
		Upper Bound	41.91	
	5% Trimmed Mean		39.22	
	Median		40.00	
	Variance		160.367	
	Std. Deviation		12.664	
	Minimum		19	
	Maximum		83	
	Range		64	
	Interquartile Range		18	
	Skewness		.459	.217
	Kurtosis		.156	.430

HYPOTHESIS:

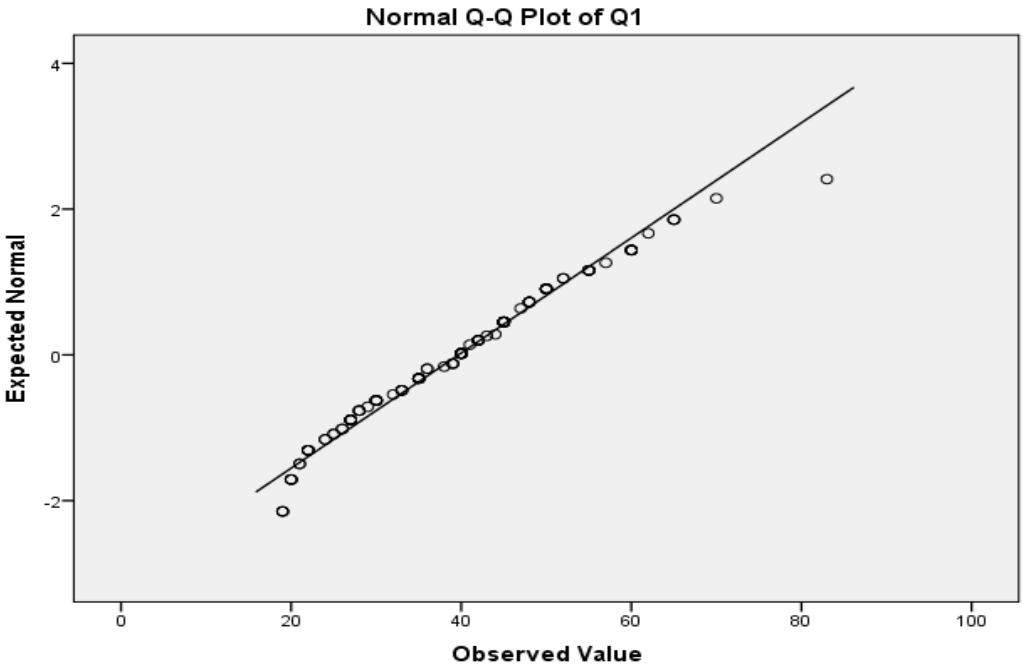
Ho: Age is normally distributed
H1: Age is not-normally distributed.

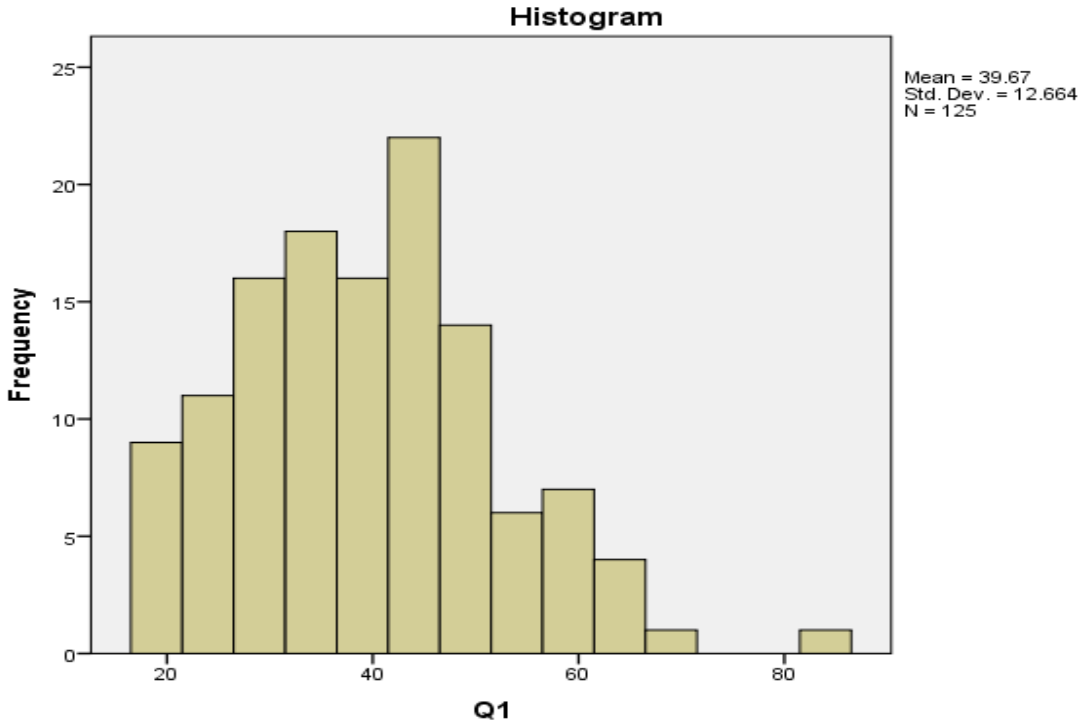
Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Q1	.971	125	.009

a. Lilliefors Significance Correction

CONCLUSION: p value < alpha, we have enough evidence to reject Ho and conclude that the population is non-normally distributed.





INTERPRETATION:

Above both figure indicates that the population does not have normal distribution.

2. **Socio-Economic Score (SES)**

Descriptive statistics for SES

			Statistic	Std. Error
SES	Mean		25.86	.383
	95% Confidence Interval for Mean	Lower Bound	25.10	
		Upper Bound	26.61	
	5% Trimmed Mean		25.84	
	Median		26.00	
	Variance		18.350	
	Std. Deviation		4.284	
	Minimum		16	
	Maximum		38	
	Range		22	
	Interquartile Range		6	
	Skewness		-.058	.217
	Kurtosis		-.072	.430

HYPOTHESIS:

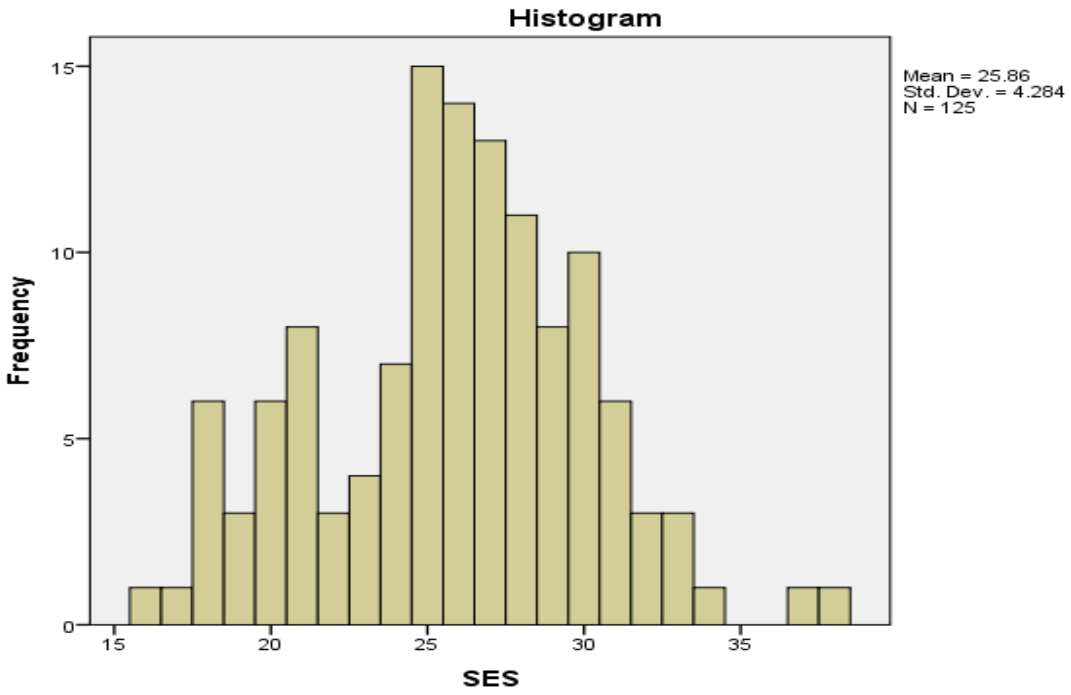
Ho: SES is normally distributed
H1: SES is not-normally distributed.

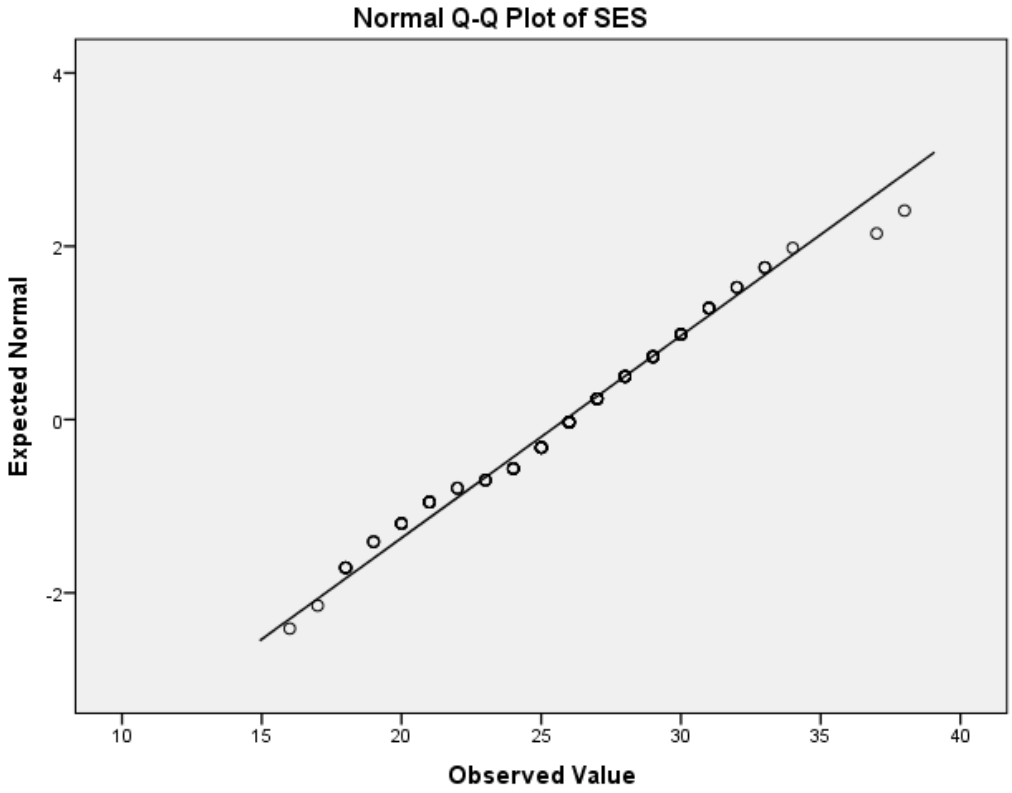
Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
SES	.981	125	.079

a. Lilliefors Significance Correction

CONCLUSION: p value > alpha, we have not enough evidence to reject Ho and conclude that the population is normally distributed.





INTERPRETATION:

Above figure indicates that the SES have normal distribution with mean 25.86 and standard deviation 4.284.

3. Drudgery index:

Descriptive Statistics

		Statistic	Std. Error
Mean		57.10	.774
95% Confidence Interval for Mean	Lower Bound	55.57	
	Upper Bound	58.63	
5% Trimmed Mean		56.78	
Median		56.83	
Variance		74.842	
Std. Deviation		8.651	
Minimum		35	
Maximum		84	
Range		49	
Interquartile Range		12	
Skewness		.435	.217
Kurtosis		.632	.430

HYPOTHEISIS:

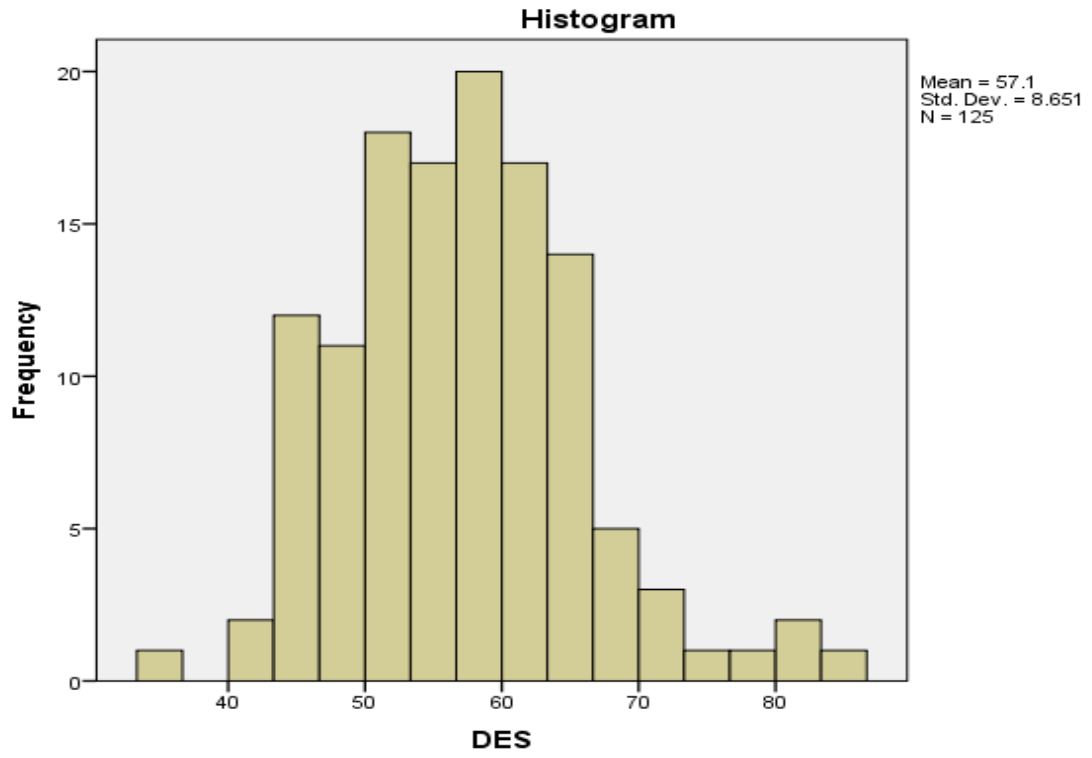
Ho: Drudgery index is normally distributed
H1: Drudgery index is not-normally distributed

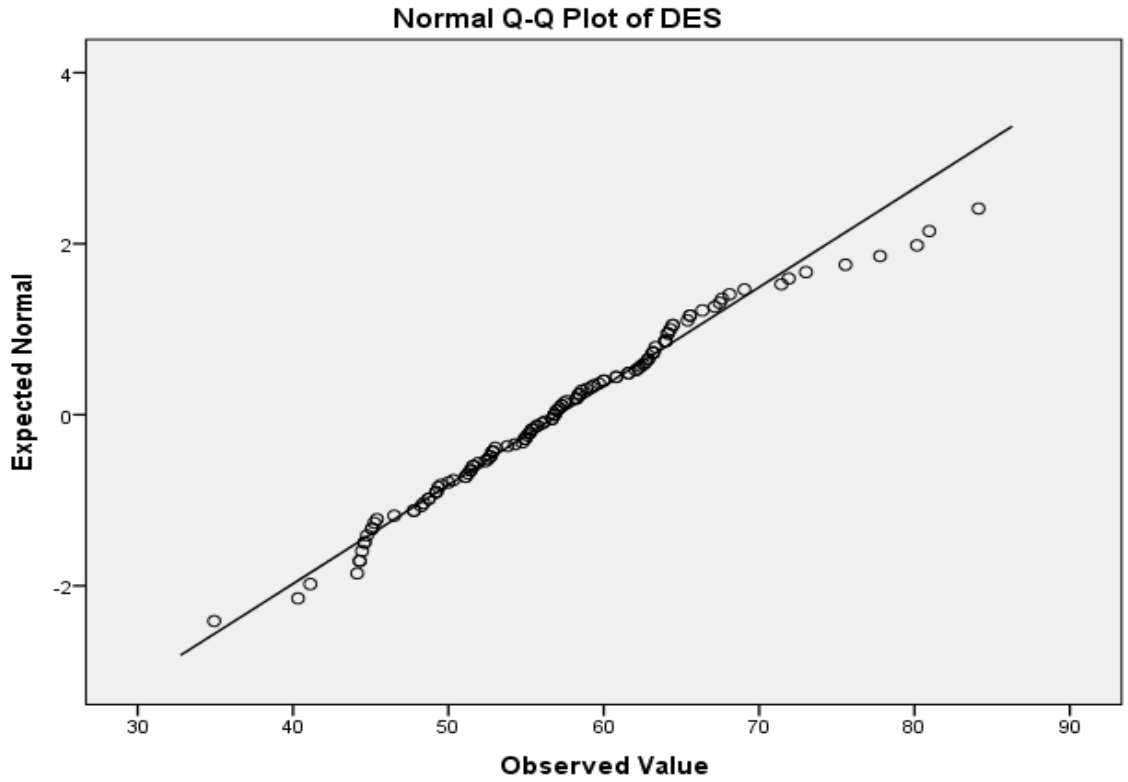
Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
DES	.981	125	.081

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

CONCLUSION: p value > alpha, we have not enough evidence to reject Ho and conclude that the population is normally distributed.





INTERPRETATION:

The above graphs indicate that Drudgery score are normally distributed.

PROPORTION TEST:

1] OBJECTIVE:

To check whether the proportion of farming and non-farming women are same.

HYPOTHESIS:

H0: In a population, proportion of farming and non-farming women are same.
H1: In a population, proportion of farming and non-farming women are not same.

Statistic value = 5.6920

P value = 0.000

CONCLUSION:

Here, p value < alpha, we reject Ho and conclude that proportion of farming and non-farming women are not same.

2] OBJECTIVE:

To check whether the proportion of SHG and non-SHG group are same.

HYPOTHESIS:

H0: In a population, proportion of SHG and non SHG women are same.
H1: In a population, proportion of SHG and non SHG women are not same.

Statistic value = 4.6801

P value = 0.000

CONCLUSION:

Here, p value < alpha, we reject Ho and conclude that proportion of SHG and non SHG women are not same.

F test for Equality of Variance

An F-test is used to test if the variances of two populations are equal. This test can be a two-tailed test or a one-tailed test. The two-tailed version tests against the alternative that the variances are not equal. The one-tailed version only tests in one direction, that is the variance from the first population is either greater than or less than (but not both) the second population variance. The choice is determined by the problem. For example, if we are testing a new process, we may only be interested in knowing if the new process is less variable than the old process.

H0: $\sigma_1^2 = \sigma_2^2$
H1: $\sigma_1^2 \neq \sigma_2^2$

Test Statistics: $F = S_1^2 / S_2^2$

Where S_1^2 and S_2^2 are the sample variances. The more this ratio deviates from 1, the stronger the evidence for unequal population variances.

1] Objective: To check whether the variability of SES score for farming and non-farming groups are equal:

HYPOTHESIS:

H0: Variance of SES for farming and non-farming women are equal
H1: Variance of SES for farming and non-farming women are not equal

F statistics = 3.6791
P value = 0.0607
Alpha = 0.05

CONCLUSION: p value > alpha, so we fail to reject Ho, and consider that the variance of SES for farming and non-farming groups are same.

2] Objective: To check whether the variability of SES for SHG and non-SHG groups are same:

HYPOTHESIS:

H0: Variance of SES for SHG and non-SHG are equal. Against

H1: Variances of SES for SHG and non-SHG are not equal.

OUTPUT:

F statistic = 0.7487
P value = 0.38854
alpha = 0.05

CONCLUSION: p value > alpha, so we fail to reject Ho, and conclude that the variance of SES for SHG and non-SHG are same.

Equality of Mean

The two-sample t-test is used to determine if two population means are equal. A common application is to test if a new process or treatment is superior to a current process or treatment.

Definition The two-sample t-test for unpaired data is defined as:

H0: $\mu 1 = \mu 2$

Ha: $\mu 1 \neq \mu 2$

Test Statistic: $Y = \frac{(\overline{Y1} - \overline{Y2})}{\sqrt{\left(\frac{S1^2}{n1}\right) + \left(\frac{S2^2}{n2}\right)}}$

where N1 and N2 are the sample sizes, $\overline{Y1}$ and $\overline{Y2}$ are the sample means, and $s^2 1$ and $s^2 2$ are the sample variances.

1] Objective: To check whether mean value of SES are same for farming and non-farming group:

Here, as we have same variance for both group, we are using two sample t-test with unknown and equal variance.

HYPOTHESIS:

Ho: Mean values of SES for farming and non – farming are same.

H1: Mean values of SES for farming and non – farming are not same.

OUTPUT:

T statistic = -1.9817

P value = 0.0497

Alpha = 0.05

CONCLUSION:

p value < alpha, so we reject Ho, and conclude that test is significant and we have not enough evidence to reject Ho. i.e SES for farming and non-farming are not same.

2] Objective: To check whether mean value of SES for SHG and non – SHG are same:

HYPOTHEIS:

Ho: Mean values of SES for farming and non – farming are same.

H1: Mean values of SES for farming and non – farming are not same.

OUTPUT:

T statistics = 0.2031

P value = 0.8393

Alpha = 0.05

CONCLUSION:

p value > alpha, so we do not reject Ho, and conclude that test is not sign ificant and we have not enough evidence to reject Ho. i.e DS for SHG and non-SH G are same.

Analysis of Drudgery Load

1. Equality of Variance

1] OBJECTIVE: The variance of drudgery score for all categories of SES are same.

HYPOTHESIS:

Ho: $\sigma 1 = \sigma 2 = \sigma 3$

H1: Inequality at one place

OUTPUT:

	Sum sq	Df	F	P value
SES	675.32	2	4.7872	0.0099
Residual	8605.09	122		

CONCLUSION:

Here, p value < alpha, So, we reject our Hypothesis and conclude that the variance are, not equal for all activity.

2] **OBJECTIVE**: The variance of drudgery score for all farming activity are same.

HYPOTHESIS:

Ho: $\sigma 1 = \sigma 2 = \sigma 3 = \sigma 4 = \sigma 5 = \sigma 6 = \sigma 7$

H1: Inequality at one place

OUTPUT:

F statistic = 2.31172

P value = 0.03204

Alpha = 0.05

CONCLUSION:

Here, p value < alpha, So, we reject our Hypothesis and conclude that the variance are not equal for all activity.

Equality of Mean

1] **OBJECTIVE**: To check whether mean drudgery score is same for all activity.

HYPOTHESIS:

H0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$

H1: Inequality at one place.

OUTPUT:

T statistics = 30.7074

P value = 0.000

Alpha = 0.05

CONCLUSION:

P value < alpha, we reject our null hypothesis, and conclude that mean drudgery score for different activity are not same.

In same way, the two sample t test has been done on all possible combinations. Which will be shown in below table.

Activity1	Activity2	T test	P value	Result
Sowing	Weeding	-3.0212	0.0027	Reject Ho
	Farm cleaning	-2.1941	0.0291	Reject Ho
	Storage	5.7668	0.0000	Reject Ho
	Fertilizer application	4.2380	0.0000	Reject Ho
	Fodder cutting	-4.8378	0.0000	Reject Ho
	Cleaning of grains	2.2792	0.0235	Reject Ho
Weeding	Farm cleaning	0.8056	0.4212	Do not reject Ho
	Storage	9.3558	0.0000	Reject Ho

	Fertilizer application	7.4461	0.0000	Reject Ho
	Fodder cutting	-2.1546	0.0321	Reject Ho
	Cleaning of grains	5.1862	0.0000	Reject Ho
Farm cleaning	Storage	8.3124	0.0000	Reject Ho
	Fertilizer application	6.5364	0.0000	Reject Ho
	Fodder cutting	-2.8478	0.0047	Reject Ho
	Cleaning of grains	4.3819	0.000	Reject Ho
Storage	Fertilizer application	-1.2094	0.2276	Reject Ho
	Fodder cutting	-10.7643	0.0000	Reject Ho
	Cleaning of grains	-2.9572	0.0034	Reject Ho
Fertilizer Application	Fodder cutting	-8.9364	0.0000	Reject Ho
	Cleaning of grains	-1.7130	0.0879	Do not reject Ho
Fodder cutting	Cleaning of grains	6.7615	0.0000	Reject Ho

From the above table, For

-weeding and farm cleaning the null hypothesis is significant. i.e their mean values of drudgery score are same.

-Fertilizer application and cleaning of grains, the null hypothesis is significant. i.e their mean values of drudgery score are same.

CHI-SQUARE TEST

Chi-square is a statistical test commonly used to compare observed data with expected data based on specific hypothesis.

The chi-square test is always use for testing the null hypothesis, which states that there is no significant difference between the expected and observed result.

The Chi-square test can also be used to test for independence between rows and columns of a contingency table.

HYPOTHESES:

Ho: Attributes are independent

v/s

H1: Attributes are not independent

TEST STATISTIC:

$$\chi^2_{cal} = \Sigma \frac{(O-E)^2}{E}$$

$$\chi^2_{tab} = \chi^2_{(\alpha,(r-1)(c-1))}$$

Where:
O = observed frequency
E = expected frequency
α= level of significance
r = no. of rows
c = no. of columns

The conventional rule of thumb is that if all of the expected numbers are greater than 5, it's acceptable to use the chi-square, if an expected number is less than 5, we should use an alternative, such as a Fisher's exact test of independence.

Fisher's exact test is a statistical test used to determine if there are non-random associations between two categorical variables. this method is more powerful, particularly in 2×2 tables.

When the dimensionality exceeds 2x2, Fisher's exact test quickly becomes computationally infeasible. Then, the p-value cannot be calculated exactly anymore but one has to resort to Monte Carlo simulation.

- **GOODMAN AND KRUSKAL'S GAMMA**

The gamma coefficient (also called the gamma statistic, or Goodman and Kruskal's gamma) tells us how closely two pairs of data points "match". Gamma tests for an association between points and also tells us the strength of association. The goal of the test is to be able to predict where new values will rank. For example, if score A scores "LOW" for question 1 and "HiGH" for question 2, will score B also result in a LOW/High response?

Gamma can be calculated for ordinal (ordered) variables that are continuous variables (like height or weight) or discrete variables (like "hot" "hotter" and "hottest"). While there are other coefficients that can calculate relationships for these types of variables, like Somer's D or Kendall's Tau, Goodman and Kruskal's gamma is generally preferred for when you have many tied ranks. It is also particularly useful when your data has outliers, as they don't affect the results much. For some fields of study it may be the preferred method for all ordinal data arranged in a bivariate table. If you have two dichotomous variables (e.g. responses that are yes/no), use Yule's Q instead.

Goodman and Kruskal's gamma uses the following formula,

$$\gamma = \frac{Nc - Nd}{Nc + Nd}$$

Where,

Nc is the number of pairs that rank the same (concordant pairs)

Nd is the number of pairs that don't rank the same (discordant pairs)

The gamma coefficient ranges between -1 and 1.

1 = perfect positive correlation: if one value goes up, so does the other.

-1 = perfect inverse correlation: as one value goes up, the other goes down.

0 = there is no association between the variables

The closer you get to a 1 (or -1), the stronger the relationship. You can reduce the significance of your result by running a significance test for gamma (see below). But how strong these relationships need to be depend upon which field of study you're working in. For example, a .75 might be "strong enough" in one field while another might require over .8.

You can interpret gamma as the proportion of ranked pairs in agreement. For example, if gamma = +1, it means that every single pair in your experiment is in agreement, or that every later has agreed upon which order the items should be ranked.

Gamma treats the variables symmetrically; you don't have to hypothesize which might be dependent and which might be independent variables.

Objective: Independence of SES with different variables:

HYPOTHESIS:

H0: There is no association between variables.
H1: There is association between variables.

1. Age

Hypothesis:

Ho: There is no association between age and SES.

H1: There is association between age and SES.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	73.789[a]	66	.239
Likelihood Ratio	66.905	66	.446
N of Valid Cases	125		

a. 99 cells (97.1%) have expected count less than 5. The minimum expected count is .05.

Here, P value = 0.239 and alpha = 0.05

CONCLUSION: p-value > alpha, therefore the data do not provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is no association between age and SES.

2. Education

	Socio-Economic Category			Total
	Lower middle class	Middle class	Upper middle class	
Uneducated	26	16	0	42
Read Only	1	3	0	4
Read and Wright only	1	17	0	18
Primary	4	14	0	18
Upper Primary	0	14	1	15
Secondary	0	18	3	21
Graduation	0	5	2	7
Total	32	87	6	125

HYPOTHESIS:

Ho: There is no association between Education and SES.

H1: There is association between Education and SES

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	61.536[a]	12	.000
Likelihood Ratio	66.641	12	.000
N of Valid Cases	125		

a. 14 cells (66.7%) have expected count less than 5. The minimum expected count is .19.

Here P value = 0.000 and alpha = 0.05

CONCLUSION: p-value < alpha, therefore the data provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is association between Education and SES.

As there is association between two attributes, i.e. Education and SES. Then how much? For that we use Kruskal and Gamma's association.

There is positively 85% association between SES and Education.

Symmetric Measures[c]

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Gamma	.852	.052	8.323	.000
N of Valid Cases		125			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Correlation statistics are available for numeric data only.

3. Cast

	Socio Economic Category			Total
	Lower middle class	Middle class	Upper middle class	
Don't Know	3	2	0	5
General	0	13	5	18
OBC	11	30	1	42
SC	11	36	0	47
ST	7	6	0	13
Total	32	87	6	125

HYPOTHESIS:

Ho: There is no association between Cast and SES.

H1: There is association between Cast and SES

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	36.521[a]	8	.000
Likelihood Ratio	32.471	8	.000
N of Valid Cases	125		

a. 9 cells (60.0%) have expected count less than 5. The minimum expected count is .24.

CONCLUSION: p-value < alpha, therefore the data provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is association between Cast and SES.

As there is association between two attributes, i.e. cast and SES. Then how much? For that we use Kruskal and Gamma's association. There is negatively 40% association between SES and cast.

Symmetric Measures[c]

	Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal Gamma	-.369	.137	-2.521	.012
N of Valid Cases	125			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Correlation statistics are available for numeric data only.

4. Family Type

Count

	Socio Economic Category			Total
	Lower middle class	Middle class	Upper middle class	
Extended	1	14	2	17
Joint	25	46	4	75
Nuclear	6	27	0	33
Total	32	87	6	125

HYPOTHESIS:

Ho: There is no association between family type and SES.

H1: There is association between family type and SES

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.231[a]	4	.037
Likelihood Ratio	12.392	4	.015
N of Valid Cases	125		

a. 4 cells (44.4%) have expected count less than 5. The minimum expected count is .82.

CONCLUSION: p-value < alpha, therefore the data provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is association between Family type and SES.

Symmetric Measures[c]

	Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal Gamma	-.130	.148	-.859	.390
N of Valid Cases	125			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Correlation statistics are available for numeric data only.

As there is association between two attributes, i.e. family type and SES. Then how much? For that we use Kruskal and Gamma's association. There is negatively **13%** association between SES and family type.

5. Ration Card

	Socio Economic Category			Total
	Lower middle class	Middle class	Upper middle class	
APL	4	26	3	33
BPL	27	55	0	82
Card not	0	6	3	9
Don't Know	1	0	0	1
Total	32	87	6	125

HYPOTHESIS:

Ho: There is no association between ration card type and SES.

H1: There is association between ration card type and SES

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	30.225[a]	6	.000
Likelihood Ratio	27.658	6	.000
N of Valid Cases	125		

a. 7 cells (58.3%) have expected count less than 5. The minimum expected count is .05.

CONCLUSION: p-value < alpha, therefore the data provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is association between ration card type and SES.

Symmetric Measures[c]

	Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal Gamma	-.176	.191	-.937	.349
N of Valid Cases	125			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Correlation statistics are available for numeric data only.

As there is association between two attributes, i.e. ration card type and SES. Then how much? For that we use Kruskal and Gamma's association. There is negatively **17%** association between SES and ration card type.

6. Type of House

	Socio Economic Category			Total
	Lower middle class	Middle class	Upper middle class	
2	20	11	0	31
3	9	34	0	43
4	3	42	6	51
Total	32	87	6	125

HYPOTHEIS:

Ho: There is no association between type of House and SES.

H1: There is association between type of House and SES

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	42.087[a]	4	.000
Likelihood Ratio	43.268	4	.000
N of Valid Cases	125		

a. 3 cells (33.3%) have expected count less than 5. The minimum expected count is 1.49.

CONCLUSION: p-value < alpha, therefore the data provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is association between type of House and SES.

Symmetric Measures[c]

	Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal Gamma	.818	.070	6.760	.000
N of Valid Cases	125			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Correlation statistics are available for numeric data only.

As there is association between two attributes, i.e. House type and SES. Then how much? For that we use Kruskal and Gamma's association. There is positively **82%** association between SES and house type.

19. Cultivable land

Count

	Socio Economic Category			Total
	Lower middle class	Middle class	Upper middle class	
No	8	28	4	40
Yes	24	59	2	85
Total	32	87	6	125

HYPOTHESIS:

Ho: There is no association between type of cultivable land and SES.

H1: There is association between type of cultivable and SES

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.036[a]	2	.133
Likelihood Ratio	3.775	2	.151
N of Valid Cases	125		

a. 2 cells (33.3%) have expected count less than 5. The minimum expected count is 1.92.

CONCLUSION: p-value > alpha, therefore the data do not provide enough evidence to reject Ho at 5%level of significance. Hence, we conclude that there is no association between cultivable land and SES

Multiple Response Analysis:

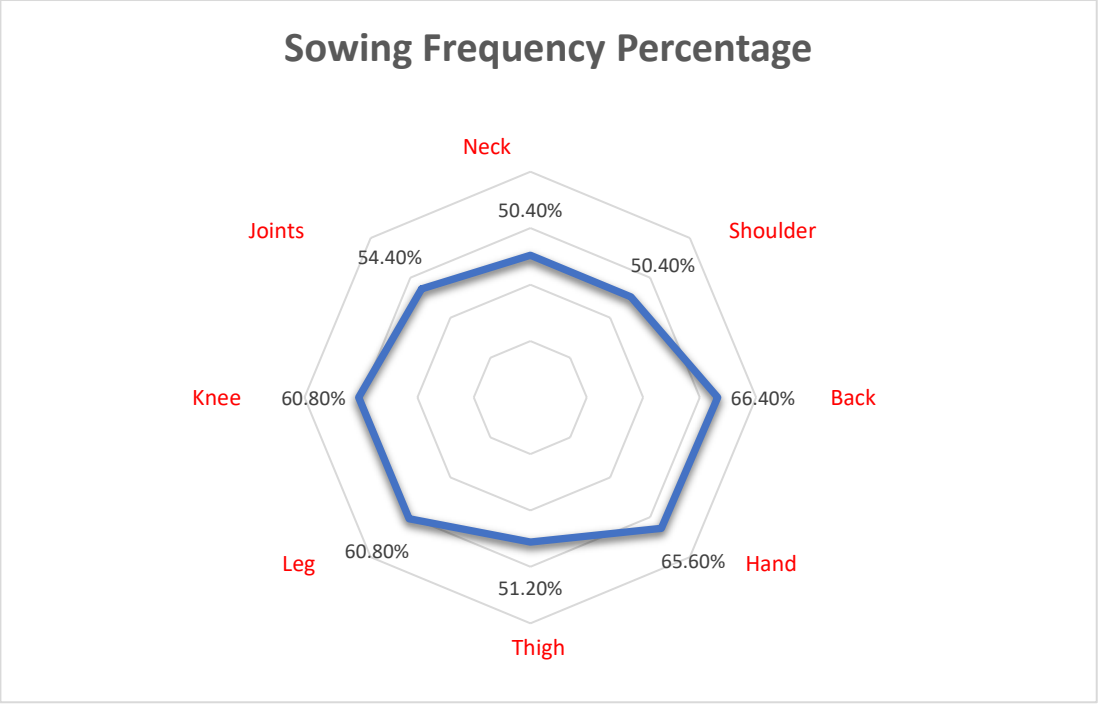
Objective: Analysis of physical injuries/pain happens to women.

1] Sowing Activity:

$Sowing Frequencies

		Responses		Percent of Cases
		N	Percent	
Sowing	Neck	63	11.0%	50.4%
	Shoulder	63	11.0%	50.4%
	Back	83	14.4%	66.4%
	Hand	82	14.3%	65.6%
	Thigh	64	11.1%	51.2%
	Leg	76	13.2%	60.8%
	Knee	76	13.2%	60.8%
	Joints	68	11.8%	54.4%
Total		575	100.0%	460.0%

a. Dichotomy group tabulated at value 1.



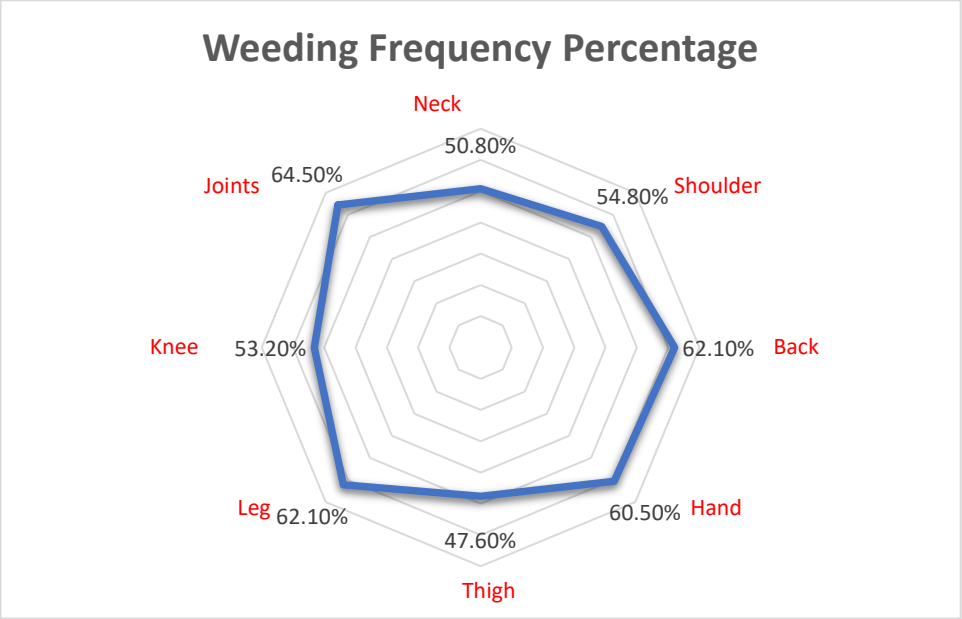
OBSERVATION: From above table we can see that, 66% women have Back and Hand injuries to them. While 60% women says they have Leg and Knee injuries.

2] WEEDING ACTIVITY:

$Weeding Frequencies

		Responses		Percent of Cases
		N	Percent	
Weeding	Neck	63	11.2%	50.8%
	Shoulder	68	12.0%	54.8%
	Back	77	13.6%	62.1%
	Hand	75	13.3%	60.5%
	Thigh	59	10.4%	47.6%
	Leg	77	13.6%	62.1%
	Knee	66	11.7%	53.2%
	Joints	80	14.2%	64.5%
Total		565	100.0%	455.6%

a. Dichotomy group tabulated at value 1.



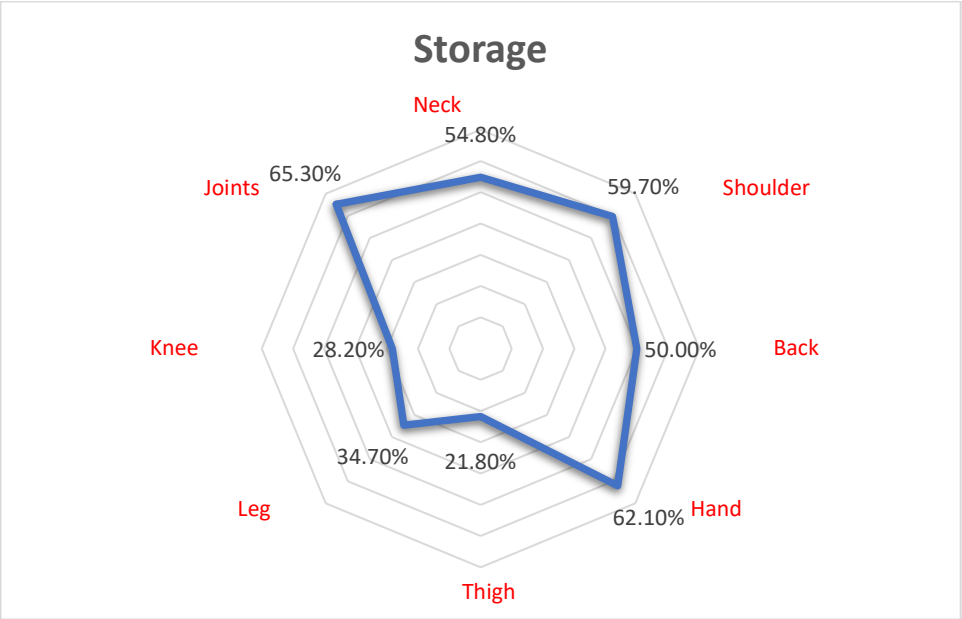
OBESERVATION: In weeding most women have Joints problem, after that 62% women feels they have back and leg injuries.

3] STORAGE ACTIVITY:

$Storage Frequencies

		Responses		Percent of Cases
		N	Percent	
$Storage[a]	Neck	68	14.6%	54.8%
	Shoulder	74	15.8%	59.7%
	Back	62	13.3%	50.0%
	Hand	77	16.5%	62.1%
	Thigh	27	5.8%	21.8%
	Leg	43	9.2%	34.7%
	Knee	35	7.5%	28.2%
	Joints	81	17.3%	65.3%
Total		467	100.0%	376.6%

a. Dichotomy group tabulated at value 1.



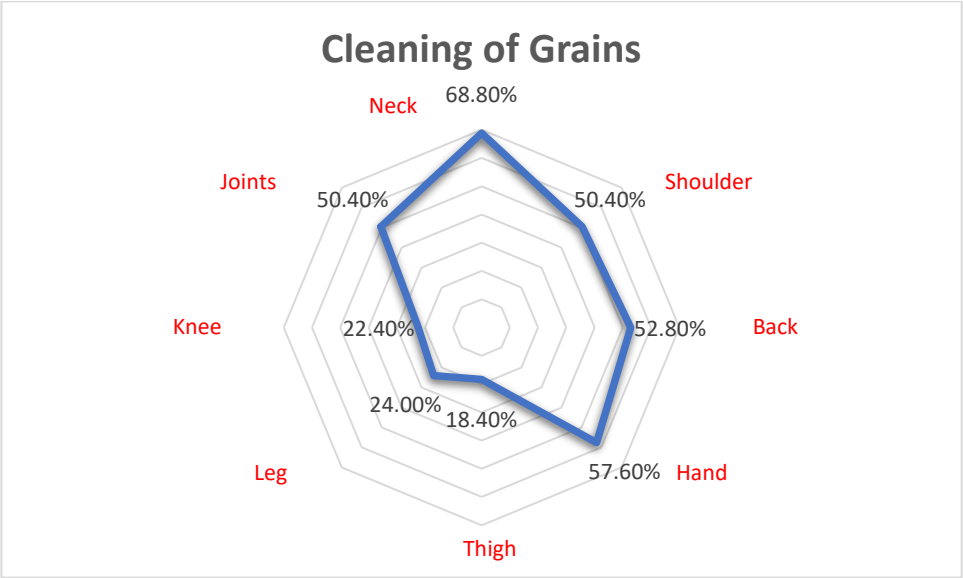
OBSERVATION: While doing storage activity, proportion of women are less who feels they have injuries on thigh and leg, which is 22% and 35% respectively. While they feel most painful body part are Joints.

4] CLEANING OF GRAINS ACTIVITY:

$cleaning_of_grains Frequencies

		Responses		Percent of Cases
		N	Percent	
$cleaning_of_grains[a]	Neck	86	20.0%	68.8%
	Shoulder	63	14.6%	50.4%
	Back	66	15.3%	52.8%
	Hand	72	16.7%	57.6%
	Thigh	23	5.3%	18.4%
	Leg	30	7.0%	24.0%
	Knee	28	6.5%	22.4%
	Joints	63	14.6%	50.4%
Total		431	100.0%	344.8%

a. Dichotomy group tabulated at value 1.



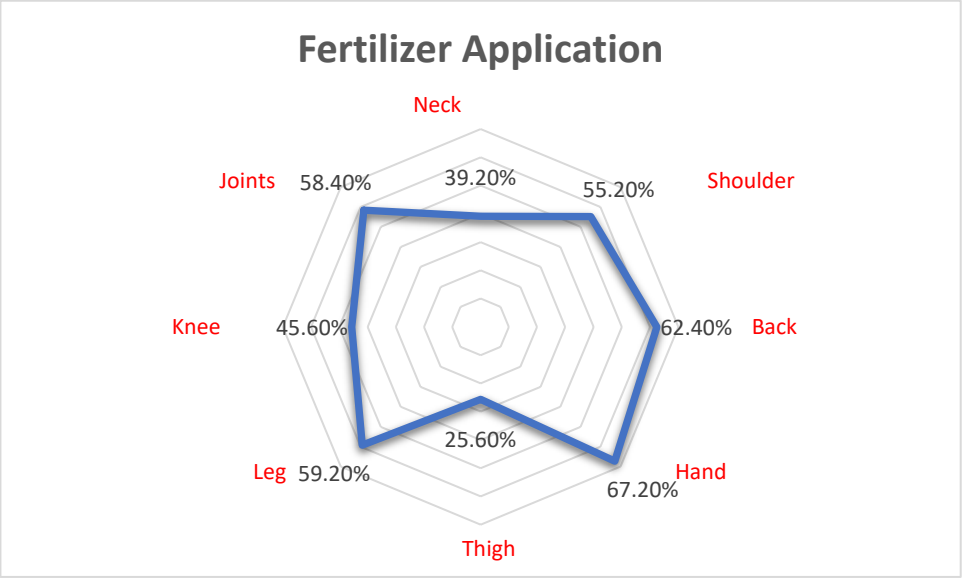
OBSERVATION: In cleaning of grains activity women feels pain in neck is the most painful body part then hands. This activity has least drudgery prone activity comparing to other activity.

5] **FERTILIZER APPLICATION ACTIVITY:**

$FERTLIZER_APPLICATION Frequencies

		Responses		Percent of Cases
		N	Percent	
FERTLIZER_APPLICATION	Neck	49	9.5%	39.2%
	Shoulder	69	13.4%	55.2%
	Back	78	15.1%	62.4%
	Hand	84	16.3%	67.2%
	Thigh	32	6.2%	25.6%
	Leg	74	14.3%	59.2%
	Knee	57	11.0%	45.6%
	Joints	73	14.1%	58.4%
Total		516	100.0%	412.8%

a. Dichotomy group tabulated at value 1.



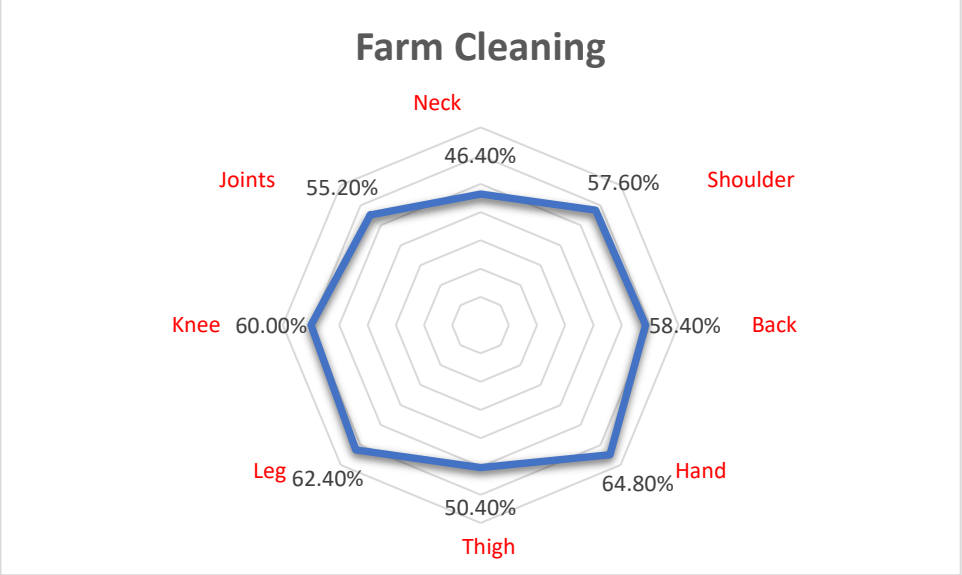
OBSERVATION: Here, most women feels they have back pain after doing these activity.

6] FARM CLEANING:

Farm_cleaning Frequencies

		Responses		Percent of Cases
		N	Percent	
Farm_cleaning	Neck	58	10.2%	46.4%
	Shoulder	72	12.7%	57.6%
	Back	73	12.8%	58.4%
	Hand	81	14.2%	64.8%
	Thigh	63	11.1%	50.4%
	Leg	78	13.7%	62.4%
	Knee	75	13.2%	60.0%
	Joints	69	12.1%	55.2%
Total		569	100.0%	455.2%

a. Dichotomy group tabulated at value 1.



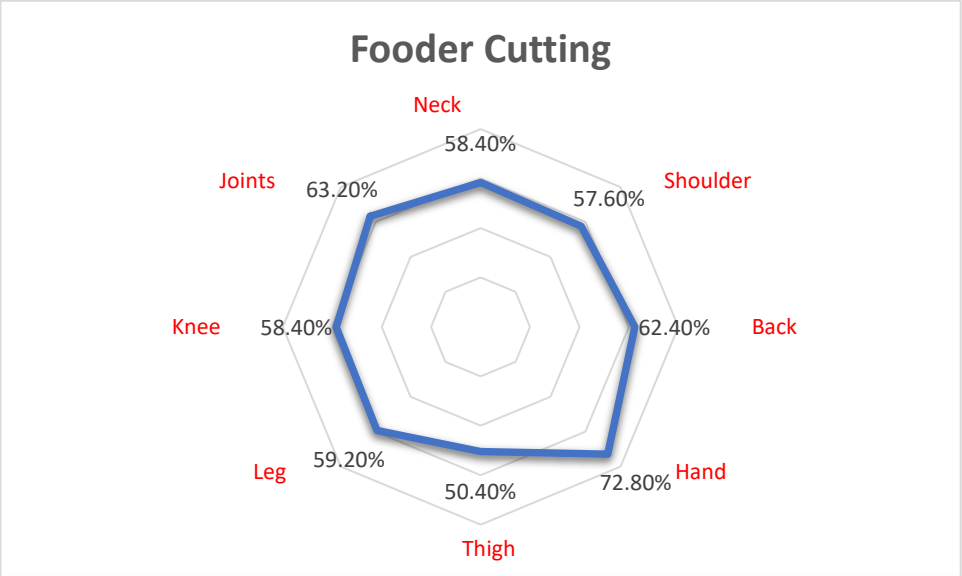
OBSERVATION: According the table, farm cleaning is the second most drudgery prone activity. Here most of the women feels they have pain in hand, back, leg, knee and thigh.

7] FODDER CUTTING ACTIVITY:

Fodder_cutting Frequencies

		Responses		Percent of Cases
		N	Percent	
Fodder_cutting	Neck	73	12.1%	58.4%
	Shoulder	72	11.9%	57.6%
	Back	78	12.9%	62.4%
	Hand	91	15.1%	72.8%
	Thigh	63	10.4%	50.4%
	Leg	74	12.3%	59.2%
	Knee	73	12.1%	58.4%
	Joints	79	13.1%	63.2%
Total		603	100.0%	482.4%

a. Dichotomy group tabulated at value 1.



OBSERVATION: From the table, we can see that Fodder cutting is the most drudgery prone activity.

CORRESPONDANCE ANALYSIS:

CA is multivariate analysis technique. It deals with Categorical Data. It converts this data in certain distance Values. Using those distance values it creates Perceptual Map. Here, by Categorical data we mean nominal / ordinal data.

CA can be used for both Ordinal and Nominal Variables. CA is useful in identifying the categories which are similar. CA is also known by dimension reduction technique because it's reduces the dimensionality.

It is conceptually similar to principal component analysis, but applies to categorical rather than continuous data. In a similar manner to principal component analysis, it provides a means of displaying or summarizing a set of data in two-dimensional graphical form.

All data should be the same scale for CA to be applicable, keeping in mind that the method treats rows and columns equivalently. It is traditionally applied to contingency table –CA decomposes the chi-squared statistic associated with this table into orthogonal factors. Because CA is a descriptive technique, it can be applied to tables whether or not the chi-square statistic is appropriate.

CA create perceptual maps where columns and rows are simultaneously potted in the perceptual map based directly on the association of the variables and objects.

Here, we have done CA on no of days taken for different activity.

No. of day	Different activities							
		Sowing	Weeding	Farm cleaning	Storage	Fertilizer application	Fodder Cutting	Cleaning of grains
	0 - 2 days	20	6	11	33	38	4	15
	3 - 4 days	30	29	26	60	37	32	45
	5 - 6 days	28	37	32	21	24	28	27
	7 - 8 days	31	36	35	10	18	24	20
	9 - 10 days	13	14	19	1	8	31	14
	More than 10 days	3	3	2	0	0	6	4

- Here, in row we can see that there are no. of days are there and in column there are no. of women who are doing farm activities.
- Here, 20 farming women are attributed 0 - 2 days in sowing activity. Similarly 30, 28, 31, 13 and 3 framing woman's attributed in 3 - 4 days, 5 - 6 days, 7 - 8 days, 9 - 10 days and more than 10 days in weeding, farm cleaning, fertilizer application, fodder cutting, cleaning of grains respectively.
- So in this manner we can say that there are some relationship between no. of days and no. of farm activities done by those no. of days.

Test of Independence:

Here, we apply chi-square test because we want to know is there any association between no. of days and no. of farm activities.

When we study this relationship between two qualitative variance, usually we first preform a test of significance of the relationship using a chi-square.

H_0:- Assumes that there is no association between(row and column) the two variable.(there is no association between no. of day in farm activities and no. of farm activities.)

H_A:- Assumes that there is an association between(row and column) the two variable.(there is association between no. of day in farm activities and no. of farm activities.)

$$X^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{\left(n_{ij} - \frac{n_{i.} \, x \, n_{.j}}{n_{..}}\right)^2}{\left(\frac{n_{i.} \, x \, n_{.j}}{n_{..}}\right)}$$

Where,

nij — the element of the cross-tabulation (or contingency table) in the i-th row and j-th column

$$Row\ marginal\ Total:\ n_{i.} = \sum_{j=1}^{b} n_{ij}$$

$$Column\ marginal\ total:\ n_{.j} = \sum_{i=1}^{a} n_{ij}$$

$$Total\ Frequency\ :\ n_{..} = \sum_{ij} n_{ij}$$

Chi-square calculated = 132.97 (=Here chi-square calculated in SPSS)

Chi-square tabulated (or critical) =18.439 (= CHISQ.INV(0.05,30))

Chi-square calculated > Chi-square tabulated (or critical)

There for we reject the null hypothesis

S0, there are difference on Average profile and both (row and column) profile. or in other word assuming that no. of days are homogeneous with respect to their farm activities at the 5% of level of significance.

Now above test we know that there is association between row and column.
Correspondence analysis works with the table of probabilities but says nothing about significance.
It really just aims to visualize the nature of the relationship between the two variables.

step by step CA:-

Step -1

first develop correspondence matrix

$$\mathbf{Z} = (\mathbf{Z_{ij}}) = \left(\frac{\mathbf{n_{ij}}}{\mathbf{n..}} \right)$$

where,

nij — the element of the cross-tabulation (or contingency table) in the i-th row and j-th column
n.. = Total frequency

	Sowing	Weeding	Farm cleaning	Storage	Fertilizer application	Fodder Cutting	Cleaning of grains	**Row Mass**
0 - 2 days	0.02286	0.00686	0.01257	0.03771	0.04343	0.00457	0.01714	**0.14514**
3 - 4 days	0.03429	0.03314	0.02971	0.06857	0.04229	0.03657	0.05143	**0.296**
5 - 6 days	0.032	0.04229	0.03657	0.024	0.02743	0.032	0.03086	**0.22514**
7 - 8 days	0.03543	0.04114	0.04	0.01143	0.02057	0.02743	0.02286	**0.19886**
9 - 10 days	0.01486	0.016	0.02171	0.00114	0.00914	0.03543	0.016	**0.11429**
More than 10 days	0.00343	0.00343	0.00229	0	0	0.00686	0.00457	**0.02057**
column Mass	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**1**

Step -2

	Sowing	Weeding	Farm cleaning	Storage	Fertilizer application	Fodder Cutting	Cleaning of grains	**Row**
0 - 2 days	0.15748	0.04724	0.08661	0.25984	0.29921	0.0315	0.11811	**1**
3 - 4 days	0.11583	0.11197	0.10039	0.23166	0.14286	0.12355	0.17375	**1**
5 - 6 days	0.14213	0.18782	0.16244	0.1066	0.12183	0.14213	0.13706	**1**
7 - 8 days	0.17816	0.2069	0.20115	0.05747	0.10345	0.13793	0.11494	**1**
9 - 10 days	0.13	0.14	0.19	0.01	0.08	0.31	0.14	**1**
More than 10 days	0.16667	0.16667	0.11111	0	0	0.33333	0.22222	**1**
Column	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**0.14286**	**1**

Find Row profile

- here 0 - 2 day's position is 0.15748,0.047244,0.0866142, and so on.
- 0 - 2 days position if we consider in row profile so that means respect to this 7 categories of farm activities.
- So, under 7 dimensions we are basically putting that where is 0 - 2 days?, where is 3 - 4 days?, where is 5 - 6 days? and so on..

Step -3

Find column profile

	Sowing	Weeding	Farm cleaning	Storage	Fertilizer application	Fodder Cutting	Cleaning of grains	**Row**
0 - 2 days	0.16	0.048	0.088	0.264	0.304	0.032	0.12	**0.14514**
3 - 4 days	0.24	0.232	0.208	0.48	0.296	0.256	0.36	**0.296**
5 - 6 days	0.224	0.296	0.256	0.168	0.192	0.224	0.216	**0.22514**
7 - 8 days	0.248	0.288	0.28	0.08	0.144	0.192	0.16	**0.19886**
9 - 10 days	0.104	0.112	0.152	0.008	0.064	0.248	0.112	**0.11429**
More than 10 days	0.024	0.024	0.016	0	0	0.048	0.032	**0.02057**
Column	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

- here sowing position is 0.16, 0.24, 0.224, 0.248 and so on.
- sowing position if we consider in column profile so that means respect to this 6 categories of no. of (6)days.
- So, under 6 dimensions we are basically putting that where is sowing? ,where is weeding?, where is farm cleaning? and so on..

Step -4

Now we find Weighted Chi-square distance :-

$$Z = D_{(r)}^{-\frac{1}{2}} (P - rc') D_{(c)}^{-\frac{1}{2}}$$

Where,

Z is distance matrix,
D(r) and D(c) are diagonal matrices,
P is correspondence matrix,
r and c are row and column messes respectively.

Weighted Chi-square distance (Z) =

0.0578	-0.0958	-0.0562	1.168214	0.1575	-0.1129	-0.025
-0.039	-0.0447	-0.0612	0.12866	0	-0.0276	0.0414
-0.0011	0.0563	0.02445	-0.0463694	-0.02678	-0.0011	-0.007
0.0415	0.0753	0.06884	-0.100899	-0.0462	-0.0061	-0.33
-0.0109	-0.0024	0.042	-0.12452759	-0.0562	0.1493	-0.002
0.0077	0.0077	-0.0132	-0.0557384	-0.055	0.0718	-0.0293

Step -5

Here we Find singular value decomposition:

Decompose the weighted chi-square matrix.

Z = UdV'

Where,

- Column of a x k matrix U are eigenvector of ZZ'.
- Column of b x k matrix V are eigenvector of Z'Z.
- Where d = diag(d1^2, d2^2, … , dk^2) which is similar to d = diag (λ1 , λ2 , ….., λk)

- Where (λ1^2, λ2^2, …, λk^2) are the nonzero eigenvalues of ZZ' and Z'Z.

Z Matrix							
	0.0578	-0.0958	-0.0562	1.168214	0.1575	-0.1129	-0.025
	-0.039	-0.0447	-0.0612	0.12866	0	-0.0276	0.0414
	-0.0011	0.0563	0.02445	-0.04637	-0.02678	-0.0011	-0.007
	0.0415	0.0753	0.06884	-0.1009	-0.0462	-0.0061	-0.33
	-0.0109	-0.0024	0.042	-0.12453	-0.0562	0.1493	-0.002
	0.0077	0.0077	-0.0132	-0.05574	-0.055	0.0718	-0.0293

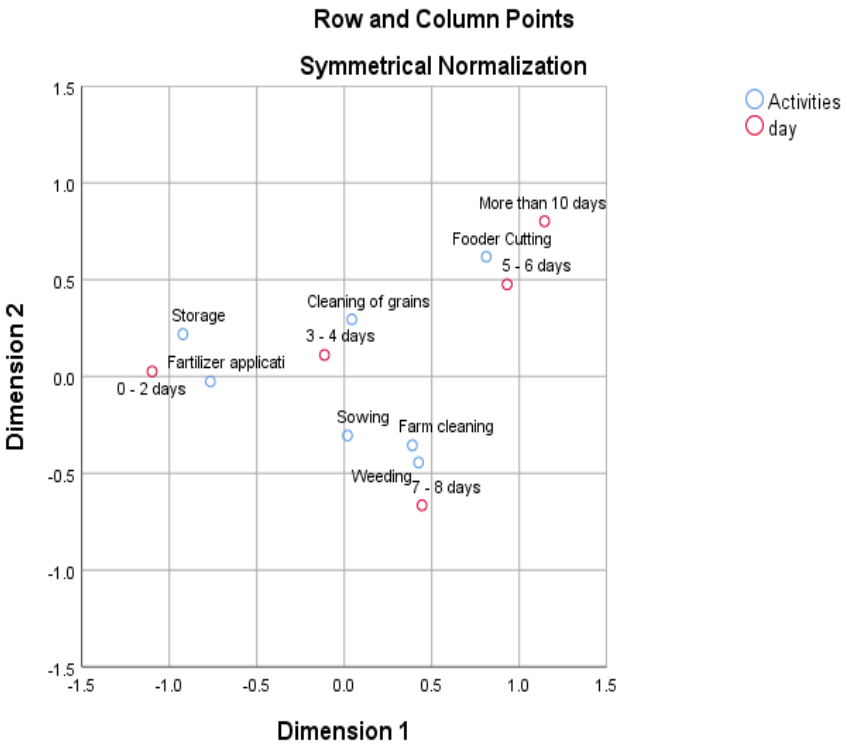
To plot the coordinates for row profile deviations in two dimensions we plot the row of the first two column of X and to plot the coordinates for column profile deviations in two dimensions, we plot the row of the two columns of Y

Row coordinates:-

0 - 2 days	-1.09844	0.025685
3 - 4 days	-0.11301	0.110941
5 - 6 days	0.931026	0.475264
7 - 8 days	0.444394	-0.66553
9 - 10 days	0	0
More than 10 days	1.14484	0.801349

Column coordinates:-

Sowing	0.018225	-0.30539
Weeding	0.424522	-0.44457
Farm cleaning	0.389434	-0.35587
Storage	-0.92261	0.21877
Fertilizer application	-0.76495	-0.02598
Fodder Cutting	0.811849	0.618368
Cleaning of grains	0.043525	0.294675



RESULTS:

Form the above graph we can say,

- In different farm activities Storage and Fertilizer application are mostly done by 0 - 2 days.
- In different farm activities weeding, sowing and farm cleaning are mostly done by 7 - 8 days.
- In different farm activities Cleaning of grains is mostly done by 3-4 days.
- In different farm activities Fodder cutting is mostly done by 5 - 6 days or some time it's takes more than 10 days.

K MODES CLUSTERING:

K-means clustering algorithm cannot cluster categorical data because of the dissimilarity measure it uses.

The K-modes clustering algorithm is based on K-means paradigm but removes the numeric data limitation whilst preserving its efficiency. The K-modes algorithm extends K means paradigm to cluster categorical data by removing the limitation imposed by K-means through following modifications:

• Using a simple matching dissimilarity measure or the hamming distance for categorical data objects

• Replacing means of clusters by their modes

Steps:

1. Randomly select k initial center points.

2. For each sample point in the data set, calculate the distance between the sample point and the k center points (Heming distance is calculated here, for different attributes of the two sample points), The number of values divide the sample points into the category corresponding to the center point closest to it.

3. After the classification is completed, re-determine the center point of the category, and use the mode of all the features of all samples in the category as the value of the corresponding feature of the new center point.

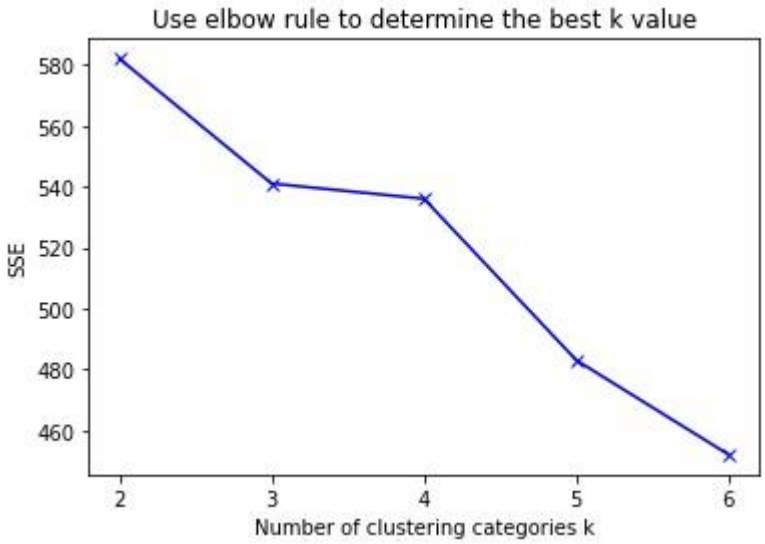
4. Repeat steps 2-3 until the total distance (the sum of the distance between the samples in each cluster and the center of each cluster) no longer decreases, and the final clustering result is returned.

By specifying the number k of different categories, SSE values corresponding to different k values are obtained (SSE value is the sum of the distances from the sample point to the center point in each category), and then different k-modes models are constructed.

Different k values correspond to SSE values, and finally a curve is drawn. This curve is equivalent to the human elbow, and the point corresponding to the elbow is the best k value point, that is, the curve Turning Point.

OBJECTIVE: Classification of Socio-Economic category using the K modes clustering.

While applying these techniques using python, we get following results.



The above figure is drawn for the sum of square error Vs different values of K. Here, we can observe that the first elbow point is at 3. So we chose K=3.

OUTPUT:

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 0, 2, 2, 0, 1, 0, 0, 0, 0, 2, 0, 2, 2, 2, 0, 1, 1, 0, 2, 2,
       0, 2, 0, 0, 0, 0, 2, 2, 2, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2,
       0, 0, 0, 0, 1, 1, 0, 1, 1, 2, 2, 0, 2, 0, 0], dtype = uint16)
```

Here, 0 means Lower middle class category

1 means Middle class category

2 means Upper middle class category

CONCLUSION:

These results are quite similar to udai - pareekh scale, but the difference is that, here the no of parameters are more than earlier approach. So, these results are more reliable then udai-pareekh scale.

SUMMARY:

From the above analysis, we have concluded below points.

- In a population, Socio-Economic score for farming and non-farming women are not same but their variances are same.
- In a population, Socio-Economic score of SHG and Non-SHG groups are same, and their variance also same.

- In a population, variance of drudgery score for farming activities are not equal.
- In a population, drudgery score for
    - Weeding and Farm cleaning activity are same
    - Fertilizer application and cleaning of grains activity are same.

  Except this, for all other combination of activities the drudgery score is different.

- Dependence between different socio-economic variables.

Variable 1	Variable 2	Independency	Association
SES	Age	Independent	
SES	Education	Dependent	Positively 85%
SES	Cast	Dependent	Negatively 40%
SES	Family type	Dependent	Negatively 13%
SES	Ration card	Dependent	Negatively 17%
SES	Type of house	Dependent	Positively 85%

- Using this associations, we have predicted clusters for socio-economic categories. Which shows some quite similar results to udai-pareekh scale.

REFERENCES:

1. IDENTIFYING DRUDGERY PRONE FARM ACTIVITIES PERFORMED BY WOMEN, by Dr Ruplekha Borah and M Kalita,Assam University.

2. Calculate the drudgery involved in the selected drudgery reducing technologies and factors responsible for non adoption modern technology in dairy farming by Amisha kumari, Neelma Kunwar and Rashmee Yadav, Uttar Pradesh (Link)

3. Socioeconomic status scales-modified Kuppuswamy and Udai Pareekh's scale updated for 2019 by Rabbanie Tariq Wani, Jammu and Kashmir.

4. Bank customer clustering by Asish, Kaggle Novice, Karnataka, (Link)

5. Clustering Algorithms: A one Stop-Shop by Ilias Miraoui (Link)

ANNEXURE:

QUESTIONNAIRE:

SOCIO-ECONOMIC STATUS AND HEALTH OF FARMING WOMEN, AND IMPACT ASSESMENT OF DRUDGERY REDUCTION TOOLS

નામ : _____

જન્મ તારીખ : _____ ફોન નંબર : _____

ઉમર : _____ ગામ : _____

1	નીચેના માંથી શાના સદસ્ય છો? 1) કઈ નહીં 2) SHG 3) ઉજાશ 4) FPO 5) કોઈ સંગઠન ના પ્રમુખ	
2	જ્ઞાતિ: 1) SC 2) ST 3) OBC 4) General	
3	વૈવાહીક સ્થિતી 1) પરિણીત 2) અપરીણિત 3) છૂટા છેડા	

4	ઉતરદાતા ની શૈક્ષણિક લાયકાત 1) નિરક્ષર 2) વાંચતાં આવડે 3) વાંચતાં – લખતા આવડે 4) પ્રાથમિક 5) માધ્યમીક 6) ઉચ્ચતર માધ્યમીક 7) સ્નાતક 8) અનુસ્નાતક	
5	તમે શું કામ કરો છો? (બહુવિધ પસંદગી) 1) કઈ નહીં 2) મજૂરી 3) ડેરી 4) ધંધો 5) ખેતીકામ 6) નોકરી	
6	તમારા કુટુંબ માં કુલ કેટેલા લોકો છે? _____	
7	ઘર માં કેટલા લોકો કમાય છે? _____	
8	કૌટુમ્બિક કૂલ આવક (મહિના ની) 1) નથી 2) ૧-૫૦૦૦ 3) ૫૦૦૦ - ૧૦૦૦૦ 4) ૧૦૦૦૦-૧૫૦૦૦ 5) ૧૫૦૦૦ થી 20000 6) 20000 થી વધારે	
9	તમારી દૈનિક આવક 1) નથી 2) ૧૦૦ ઓછું 3) ૧૦૦-૫૦૦ 4) ૫૦૦-૧૦૦૦ 5) ૧૦૦૦ થી વધારે	
10	શું તમારી આવક તમારા કૌટુંબિક ખર્ચ સાથે અનુકૂળ છે ? 1) ના 2) પર્યાપ્ત 3) વધારે	
11	તમે કેવા ઘર માં રહો છો ? 1) ઝુંપડી 2) કાચા 3) અડધા પાક્કા 4) પાક્કા 5) હવેલી / બંગલો	
12	તમારી પાસે કયા સાધનો છે ?(બહુવિધ પસંદગી) 1) સાયકલ 2) મોબાઇલ ફોન 3) ટીવી 4) ફ્રિજ 5) બાઇક 6) કાર 7) ટ્રેક્ટર	
13	આર્થિક સ્તિથિ 1) એપીએલ 2) બીપીએલ 3) અંત્યોદય 4) નથી ખબર	
14	પીવાના પાણીનો સ્ત્રોત 1) ખાનગી નળ 2) હેંડ પમ્પ 3) સાર્વજનિક નળ 4) ટુબવેલ /કુવો ટેંકર 5) તળાવ/નદી/નેહેર/સરોવર	
15	તમારી પાસે કૂલ કેટલા પશુઓ છે ? _____	
	Drudgery Related	

1	તમારી પાસે પોતાની જમીન છે ? 1) હા 2) ના	
2	શું તમે ખેતીની પ્રવૃત્તિમાં કામ કરો છો? 1) હા 2) ના	
3	તમારી પાસે કેટલી જમીન છે ? 1) નથી 2) 1 થી 3 વીઘા 3) 4 થી 6 વીઘા 4) 7 થી 10 વીઘા 5) 10 થી 13 વીઘા 6) 13 વીઘા થી 15 વીઘા 7) 15 થી વધારે વીઘા	

4-5] નીચે જણાવેલ પ્રવુતિ કરતાં તમને કેટલા દિવસ / સમય લાગે છે

	પ્રવુતિ દરમિયાન						દિવસ દરમિયાન			
	0-3 દિવસ	4-6 દિવસ	7-10 દિવસ	11-15 દિવસ	15 થી વધારે દિવસ		0-2 કલાક	2 – 4 કલાક	4-6 કલાક	6-8 કલાક
વાવણી										
નીંદણ										
ખેતર ની સફાઈ										
સંગ્રહ										
ખાતર ના છંટકાવ										
કાપણી										
અનાજ ની સફાઇ										

6] નીચે જણાવેલ પ્રવુતિ દરમિયાન પડતી મુશ્કેલી ના સ્તર ને દર્શાવો

	ખૂબ જ સરળ	સરળ	તટસ્થ	અઘરી	ખૂબ જ અઘરી
વાવણી					
નીંદણ					
ખેતર ની સફાઇ					
સંગ્રહ					
ખાતર ના છંટકાવ					
કાપણી					
અનાજ ની સફાઇ					

7] નીચે જણાવેલ ખેતી કામ કરતી વખતે શરીર ના અંગ માં મુશ્કેલી પડતી હોય તો હા લખો .

	વાવણી	નીંદણ	ખેતર ની સફાઇ	સંગ્રહ	ખાતર ના છંટકાવ	કાપણી	ઊપણવું
ગરદન							
ખભો							
પીઠ							
હાથ							
જાંઘ							
પગ							
ઘૂંટણ							
સાંધાઓ							

IMAGES:

image 1.

image 2

Folder Link: Here is the link to the folder.

https://drive.google.com/drive/folders/1kKVoW_guqRTyYhPdLeSMzqzaIeJfB1QJ?usp=sharing