MAS 5402

Analysis of quality of life in renal failure patients in Sri Lanka

November 2021

Table of Contents

1. Introduction

A kidney renal failure is a serious disease, which has major impact on life and can be accidentally fatal; several studies have demonstrated the high incidence of renal failure, which are of two types i.e. acute and chronic renal failures. Kidney disease is an important public health issue. It is common and the prevalence increases with age, which means that the disease burden will increase with our aging population. Chronic kidney disease is an independent risk factor for other diseases, particularly cardiovascular disease. It often coexists with other cardiovascular conditions meaning that it needs to be managed alongside other diseases and risk factors such as diabetes and hypertension as well as the social needs that come with frailty and multiple conditions. In a minority of cases, chronic kidney disease progresses to end stage renal disease, which may require renal replacement therapy. This progression and the risks of other vascular events, such as stroke and heart failure can be reduced if chronic kidney disease is identified and managed, early diagnosis is therefore essential.

The acute renal failure (ARF) is characterized usually reversible deterioration of renal function, which develops over a period of days or week. It occurs suddenly, by causing bacterial infection, injuries, shock, congestive heart failure, drug poisoning and severed bleeding which results in uremia. A marked reduction in urine volume is usual and the clinical features, while the rapid problems of diagnosis and management arises. Many of the disorder giving rise to acute renal failure carry high rate of mortality in human beings, but if the patients survives, then the renal function usually returns to normal or near normal.

Chronic kidney disease (CKD) describes abnormal kidney function and/or structure. It is common, frequently unrecognized and often exists together with other conditions (for example, cardiovascular disease and diabetes). CKD can progress to end stage renal disease in a small but significant percentage of people. CKD is usually asymptomatic until the late stages, but it is detectable usually by measurement of serum creatinine or urine testing for protein. There is evidence that treatment can prevent or delay the progression of CKD, reduce or prevent the development of complications and reduce the risk of cardiovascular disease. Statistical simulations techniques and sampling tests are widely used to explore significant empirical results and implications in different allied fields of biology, natural science, and life sciences. In this project we hope to get a better understanding about quality of life of renal failure patients.

Currently, Sri Lanka is facing an increase in kidney disease; of which chronic kidney disease is the most prominent. According to the current statistics, there are five million kidney patients in Sri Lanka alone and it is unfortunate that nearly two million of them are suffering from ESRD. There are many scientific reasons for the occurrence and spread of this kidney disease. But the main purpose of this Statistical analysis of kidney disease is to find out to what extent the behavior pattern of these patients has contributed to the development of these conditions. This analysis is expected to be analyzed mainly under the following basic points.

Objectives of the study

- Identify the variables associate with the quality of life of renal failure patients.
- Test whether if quality of life of renal failure patients, vary with “Gender”.
- Test whether if quality of life of renal failure patients, vary with their “Level of Education”.
- Test whether if quality of life of renal failure patients, vary with “number of children they have”.

2. Presentation of Data

In initial dataset it had 853 records and after removing missing values using SPSS software it has 537 records and based on later data set descriptive and advanced analysis would be carried out. In the initial data set it has 20 variables, 36 variables under SF Score and 30 variables under DSI score. SF -36 is a measure of health-related quality of life which has 36-item short form health survey. Dialysis Symptoms Index (DSI) is a self- reported index to assess symptoms and their severity in patients with end-stage renal disease.

But in this study, we only considered below mentioned variables.

Categorical Variables	Numerical Variables
Gender	Age
Living District	SF- 36 Score
Marital Status	DSI Score
Number of Children	
Education Level	
Occupation	
Average Monthly Income Before Having the Disease	
Current Average Monthly Income	
SF – 36 Descriptors (36 Variables)	
DSI Descriptors (30 Variables)	
Have you changed/ resigned from the job due to your illness	
Have you had difficulties in performing current/ previous job due to illness	

Table 2.1 – Considered variables from the data set

Variables of interest:

Response variable
Quality of life in renal patients; Average of SF-Score

Explanatory variables
District, age in years, Gender, Occupation, Have you changed/ resigned from the job due to your illness, Have you had difficulties in performing current/ previous job due to illness, From all the sources of income what was your average monthly income before your had this disease?, Form all of yours sources of income, what is your current average monthly income?, Average family income, Education, Highest occupation of the household, Highest education of the household, Does your household own or rent the accommodation?, is your marital status?, Number of children, DSI score

3. Methodology and data processing

This study was carried out based on secondary dataset which was collected from the research unit at National Institute of Canine Service and Training (NICST) Sri Lanka and that primary survey was done in year 2016. It covers data from personal, educational, financial, Physical & mental health details of the renal patients in Sri Lanka.

Data set was analyzed & visualized by using SPSS and MS Excel software packages. Data analysis section was divided into two major parts. Those are descriptive analysis and advanced analysis. Under descriptive analysis, analysis section was further divided into univariate analysis and bivariate analysis. Using graphical and numerical summaries univariate analysis was conducted to get primary understanding of the behaviors of the variables considered. Using graphical, numerical and other statistical tests, bivariate analysis was conducted to understand behaviors of relevant variables. Furthermore, advanced analysis was carried out to set a multiple regression model to identify how certain variables under consideration will affect our variables of interest. And some statistical tests used for investigate other objectives.

3.1 Calculating SF-36 Score and DSI Score

Calculation of SF-36 score

SF-36 is a questionnaire which consists of 36 items that describe the quality of life of patients. Each item has multiple choices out of which patients have to select the choice that best describes their health condition. Response sets belonging to each item were scored so that a high score defines a more favourable health state. Each item was scored on a 0 to 100 range so that the lowest and highest possible scores are 0 and 100 respectively. Respective scores for entire 36 items were added together and divided by 36 to obtain the SF-36 score for an individual. SF-36 Score was considered as the response variable in the analysis.

Calculation of DSI score

Dialysis Symptom Index (DSI) was used to assess physical and emotional symptoms as well as their severity. The DSI contains 30 items, each of which describes a specific physical or emotional symptom. Each item has a five-choice response set out of which patients have to choose the choice that best describes the severity of the symptom. Items were scored according to a Five-point Likert scale where value 1 indicates that the symptom is not at all bothersome and value 5 indicates that the symptom bothers very much. Scale values for entire 30 items were added together to obtain the DSI score for an individual. Calculated DSI score value is between 0-150, 0 indicating no symptoms present and 150 indicating worst burden of symptoms.

4. Descriptive Analysis

(a)Univariate Analysis

4.a.1 <u>Distribution of Renal patients across the country by district.</u>

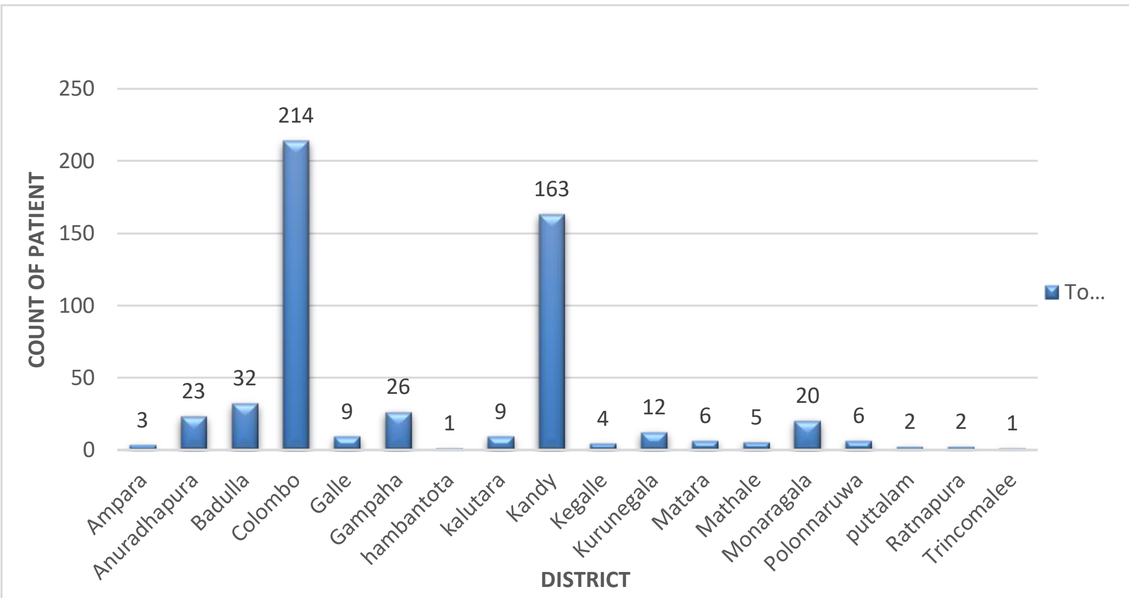


Figure 4.a.1.1: bar graph – District distribution of renal patient

This graph shows how the prevalence of kidney disease in the country by 2016 has been divided in to district. The Highest number of renal patient is reported from Colombo (214) and Kandy (163) district.

However, this distribution is based on data from patient reported from different parts of the country, and the population density of those districts should be considered in order to obtain more accurate results.

4.a.2 <u>Dividing patients by gender</u>

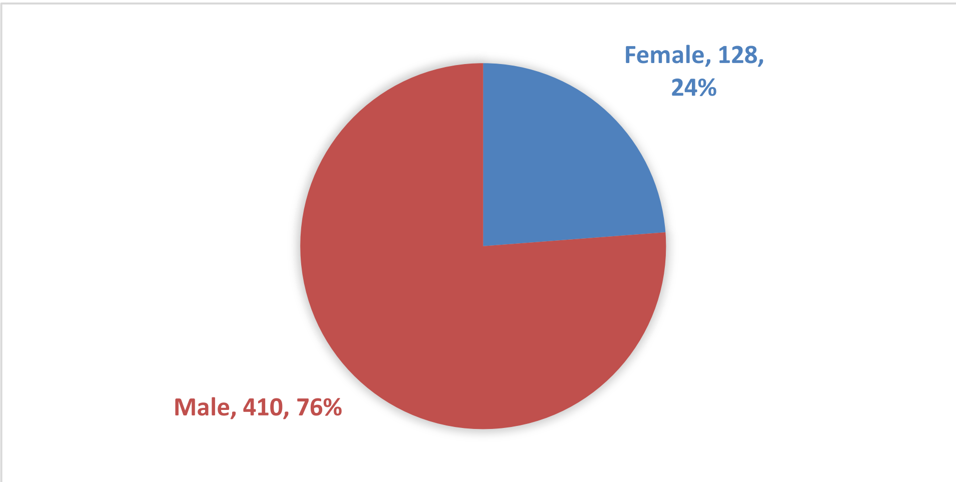


Figure 4.a.2: pie chart -Gender

This pie chart above illustrates the percentage of male and female patients from data set with ages from 7 to 83 years old who took the survey for the renal patient.76% of the survey participants are males while the remaining 24% are females. From a total of 538 patients, 410 patients were male and other 128 patients were females

4.a.3 Prevalence of renal disease by occupation

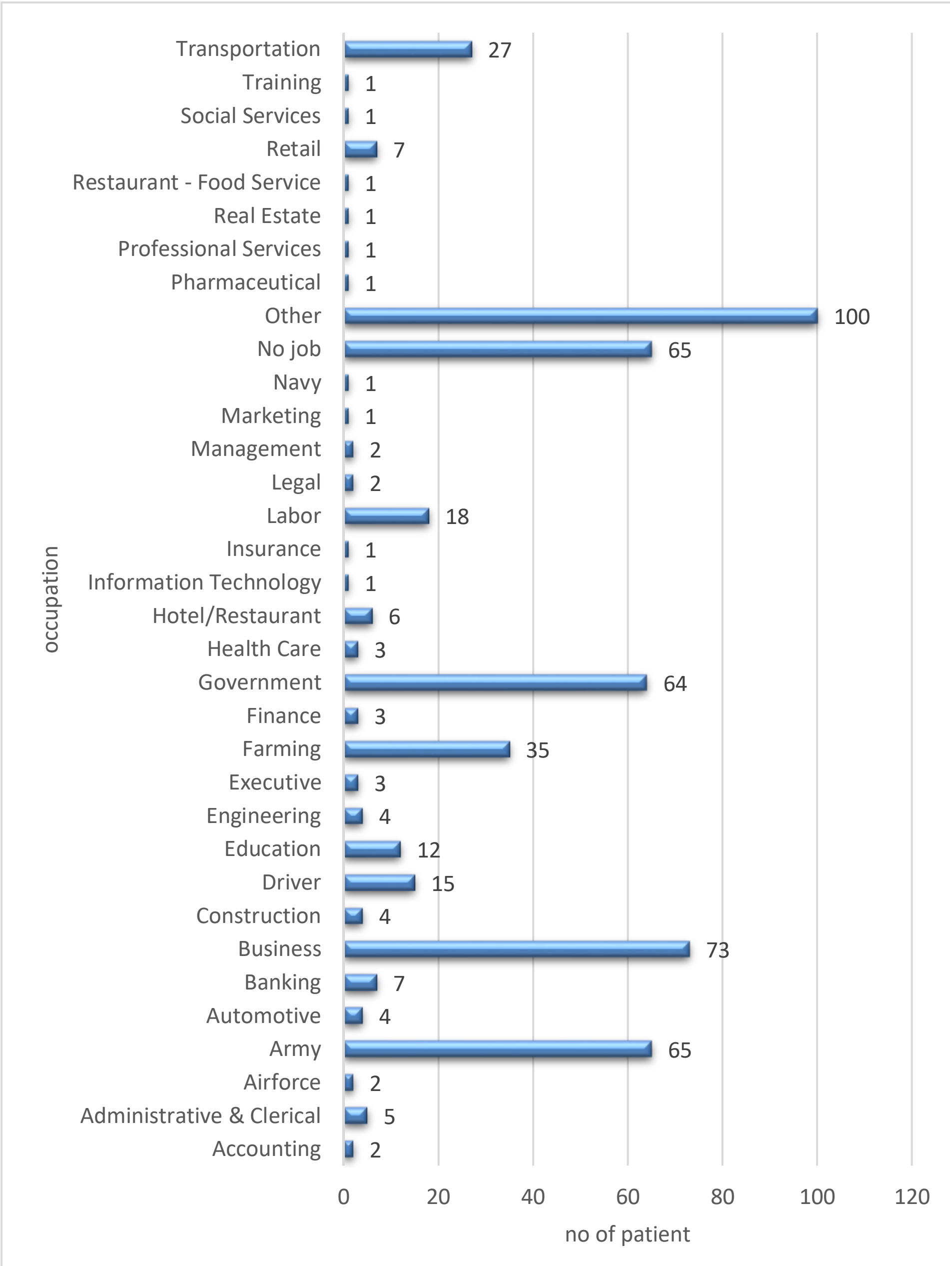


Figure 4.a.3.1: Bar graph – no of patients Vs Occupation

.

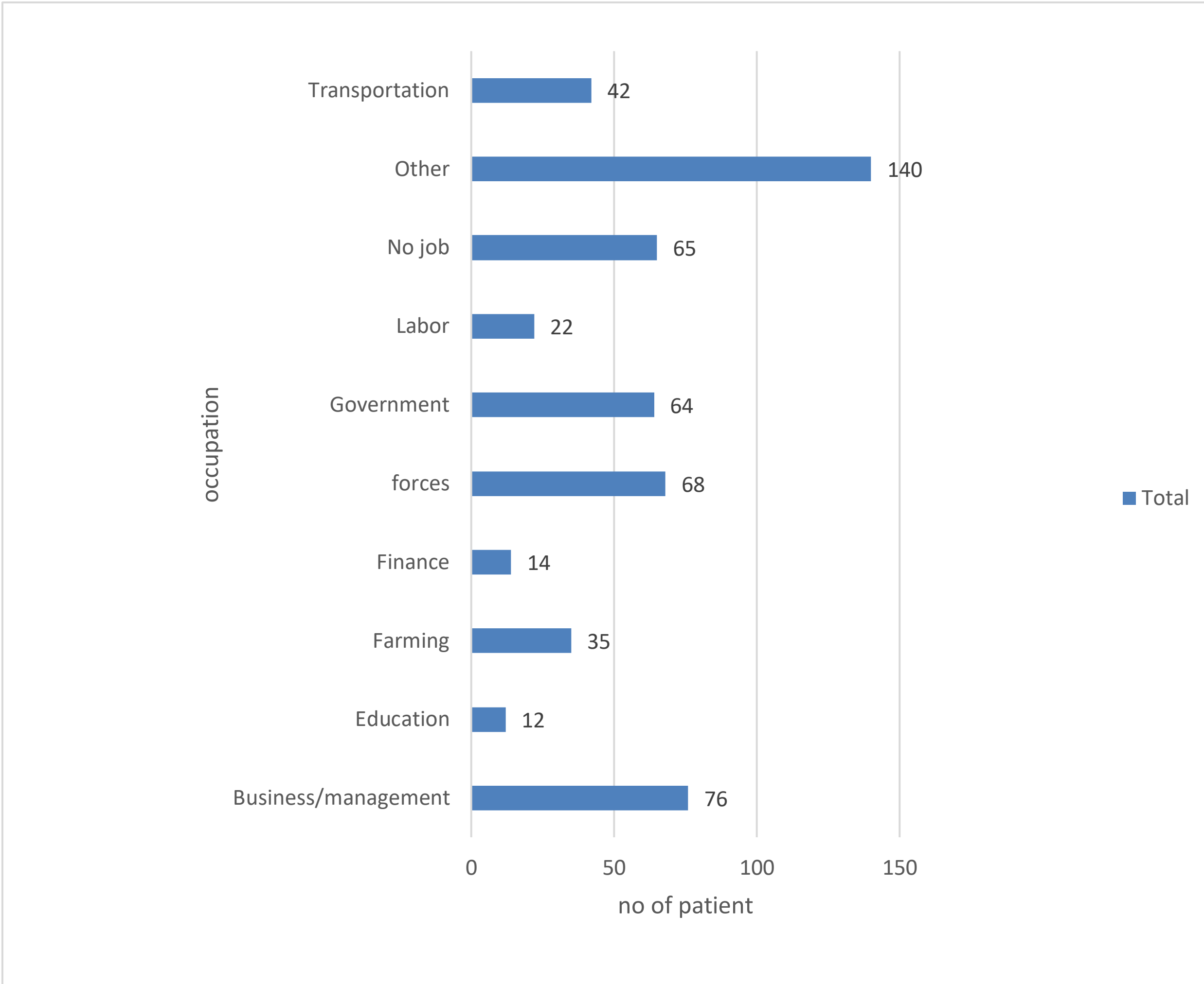


Figure 4.a.3.2: bar graph – no of patients Vs occupations

According to the survey, all the conditions of the occupations of the patients are shown in the **fig-4.a.3.1** graph and as shown in the **fig-4.a.3.2** chart, the job categories have been simplified to several categories for ease of analysis. The other job categories show a high of 140, but more attention should be paid to business, military and other categories. Accordingly, there are 76 patients in the business/management sector and 68 patients in the military sector

4.a.4 The Effect of Age

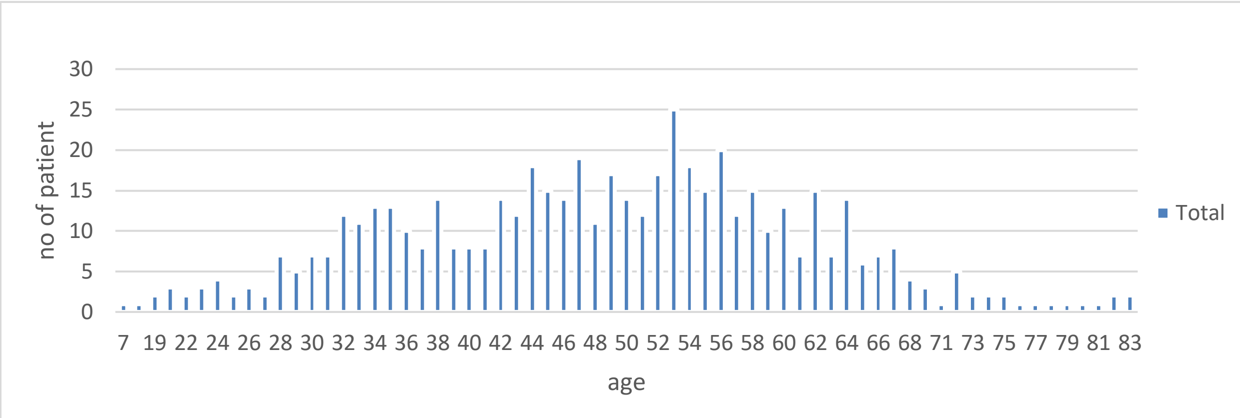


Figure 4.a.4.1: Bar graph – age Vs no of patient

The graph above shows the age levels of all patients.

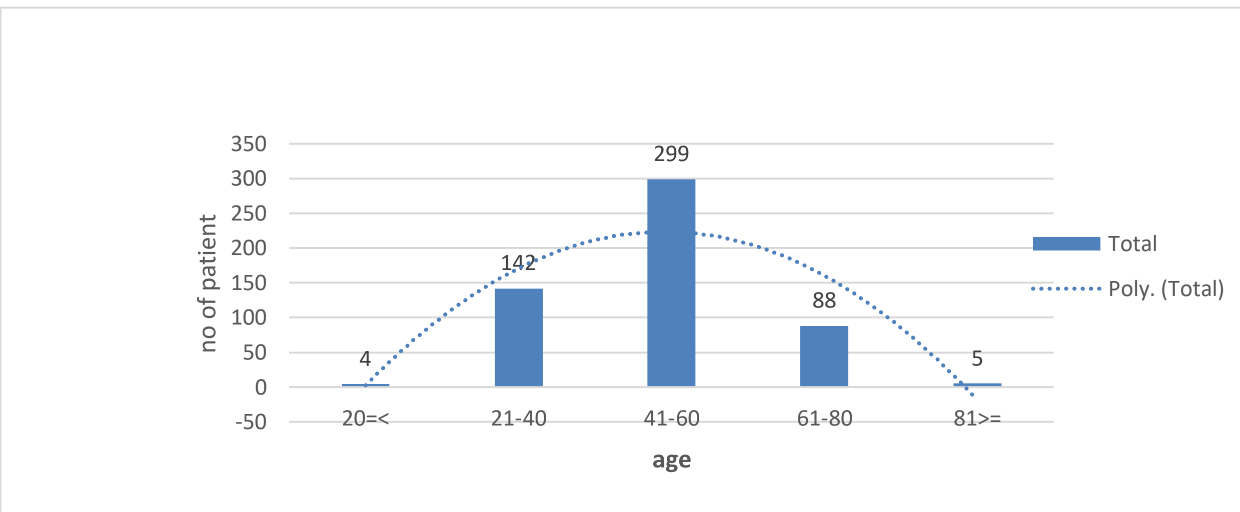


Figure 4.a.4.2: Sorted Bar graph – age Vs no of patients

The age levels represented in **fig-4.a.4.1**above are divided in to five main categories by **fig-4.a.4.2**. The highest number of patients is between the ages of 41 and 60 and that is 299. A further, 142 patients in the 21-40 age group were reported, compared to 88 in the 61-80 age group

4.a.5 Monthly income before and after illness

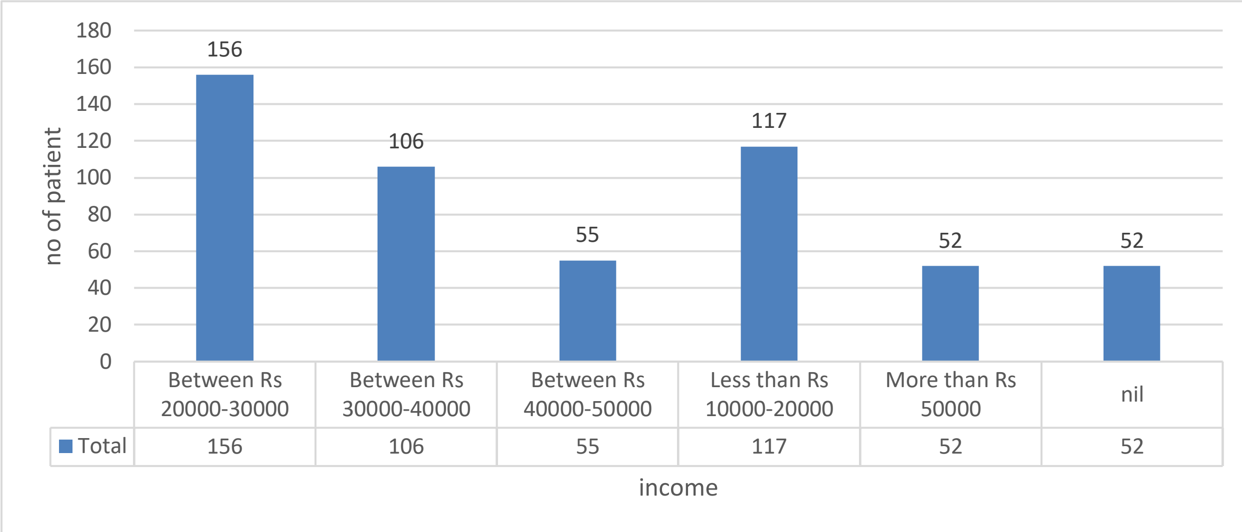


Figure 4.a.5.1: Bar graph of monthly income before illness

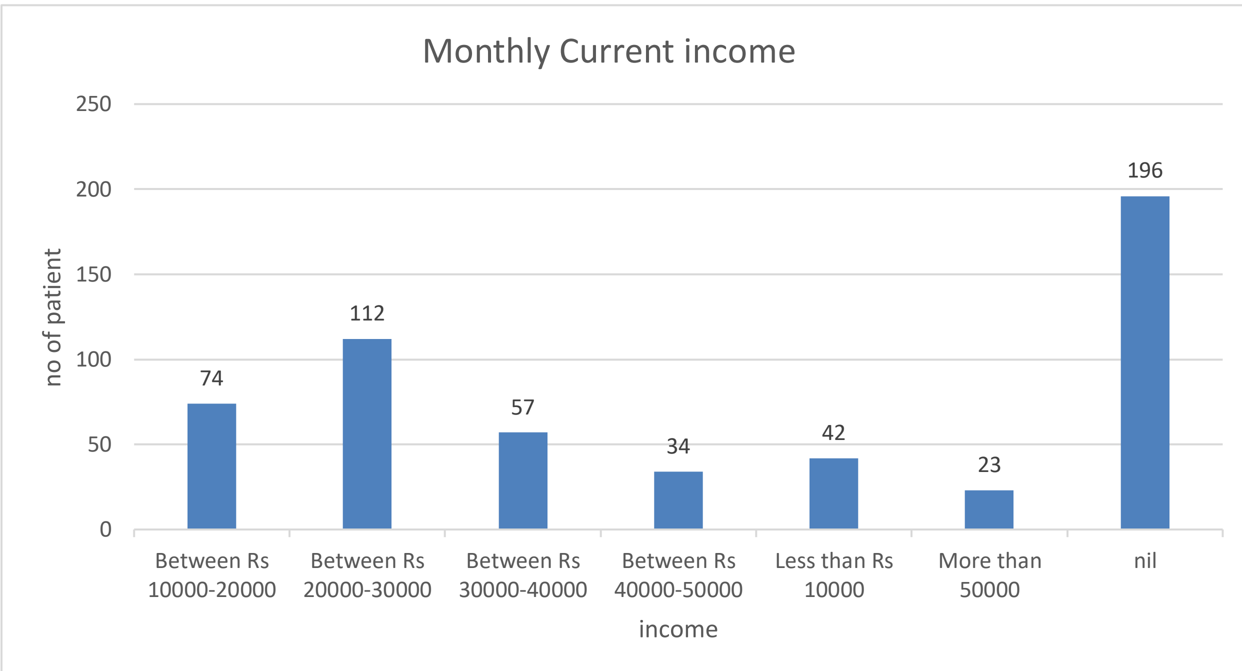
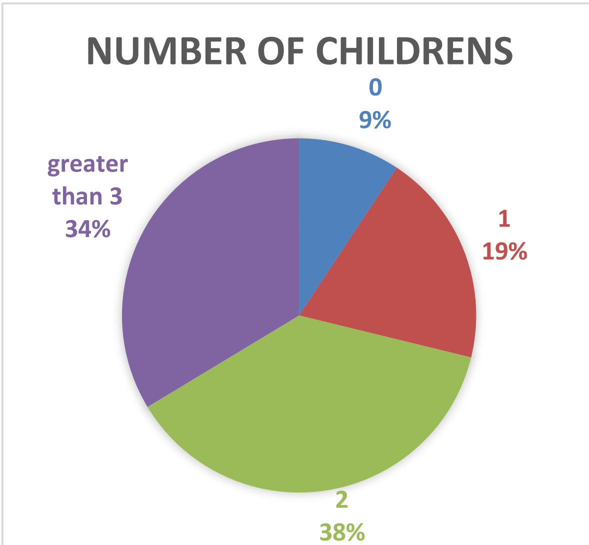
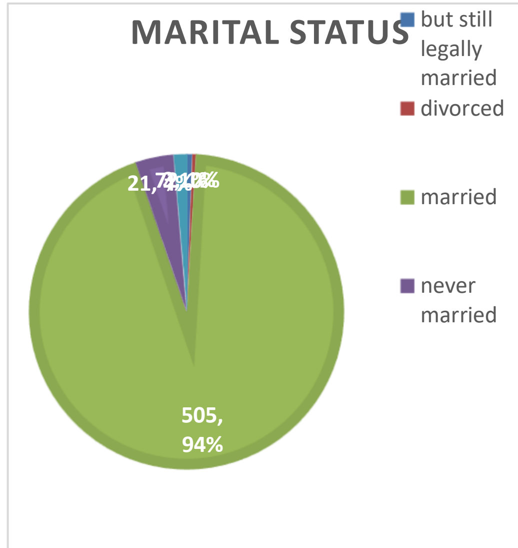


Figure 4.a.5.2: Bar graph of monthly current income

The above two bar graphs show how the patient's monthly income varies before and after the illness. Accordingly, the number of unemployed after the illness has gone up from 52 to 196.

4.a.6 Marital status & no of Children





The two pie charts above shows the number of children and marital status of those kidney patients. Of these 538 patients, 94% are married and most have at least two children.

4.a.7 Education level of patients &their household

of

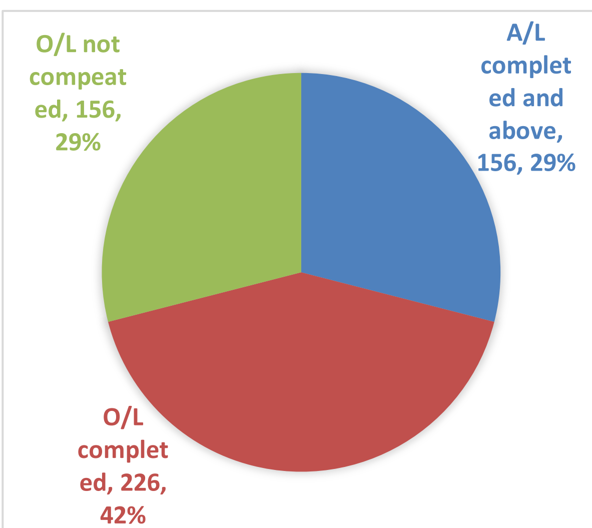


Figure 4.a.7.1: Education level of patient

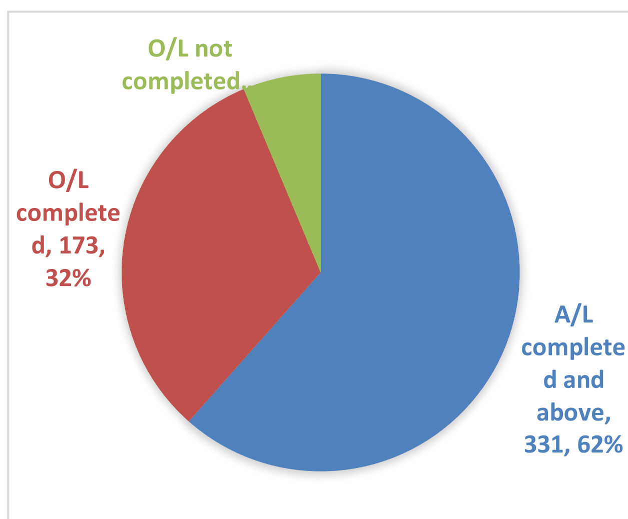


Figure 4.a.7.2: Education level of household

These graphs illustrate the highest education level of patients and their households. According to **fig-4.a.7.1**, O/L not completed percentage is 29% and according to **fig-4.a.7.2** it is 6%.

4.a.8 SF Score

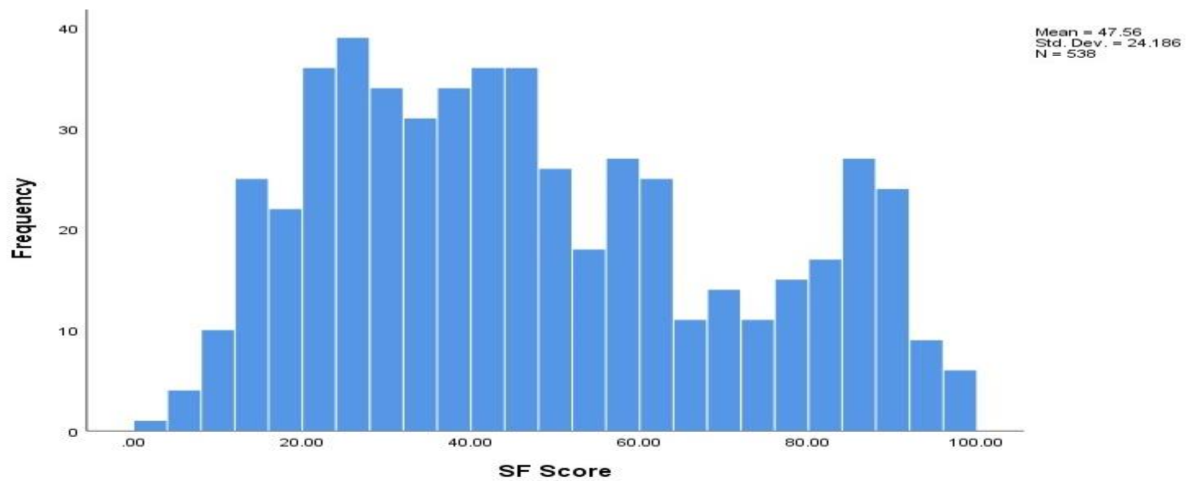


Figure 4.a.8.1: Histogram – SF score Vs Frequency

The above histogram shows how the SF score of the patients changing and is fluctuating. This SF score represents the quality of life and mental level of those patients.

4.a.9 DSI Score

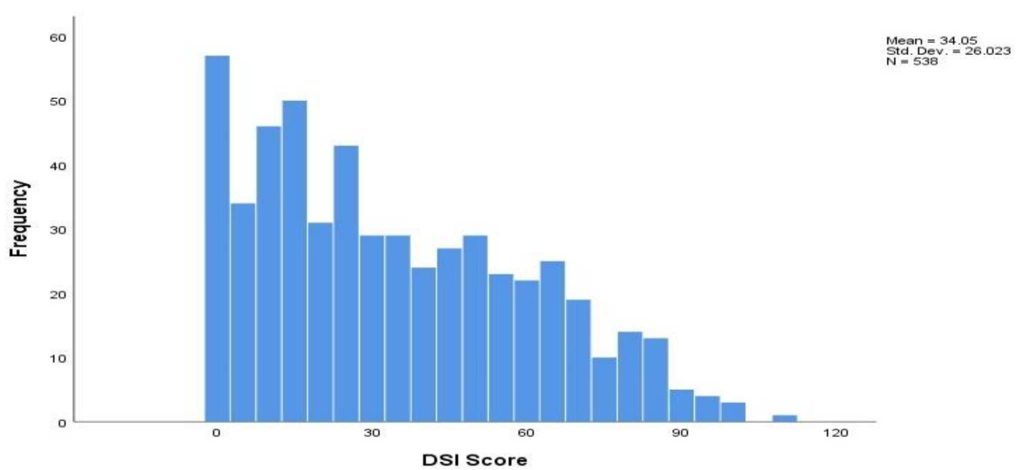


Figure 4.a.9.1: Histogram – distribution of DSI Score

The above graph illustrates how the patient's DSI score changes and some of its inclination. DSI score is a measure of how a patient feels about a disease in medicine.

(b) Bivariate Analysis

4.b.1.Association with living district

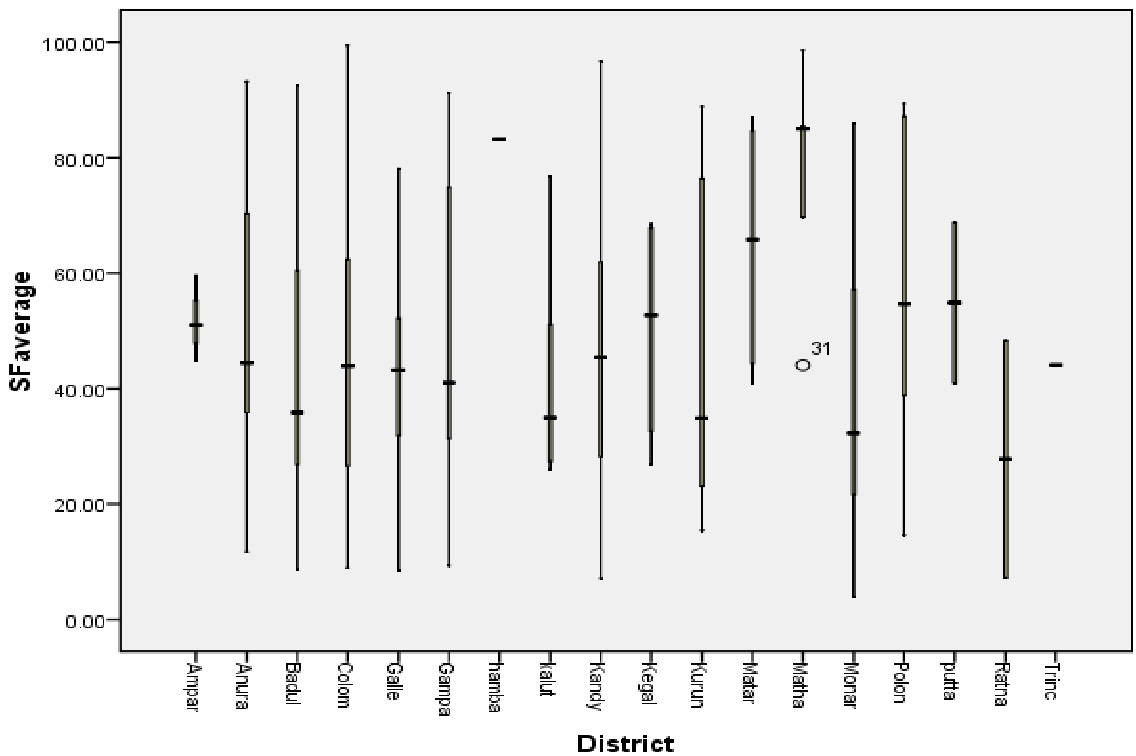


Figure 4.b. 1 : SF Score Vs District

The figure above shows that association of SF value with the living district of each patient. The patients of Ratnapura have least SF value. But numerical value of data is smallest in Ratnapura district and also there no whiskers in that boxplot. Therefore we can decide the patients of Monaragala have least SF value as well as highest SF value.

Median SF value which had the patients of Matara district is greater than median SF value which had other district.

4.b.2 Association with age of patient

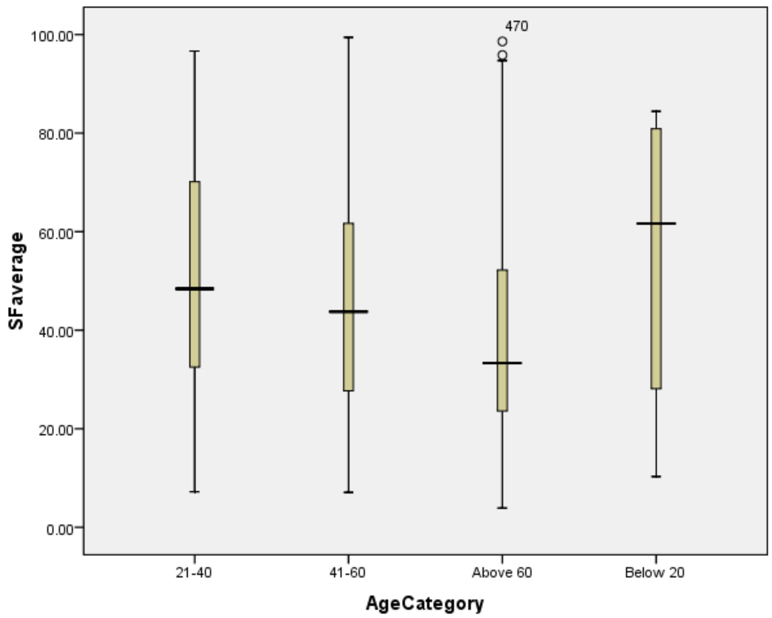


Figure 4.b.2 :SF Score Vs Age

The figure above shows that association of SF value with age of patients. Patients of age above 60 shows the least SF value as well as the highest SF value. Median SF value which had the patients of age below 20 is greater than the median SF value which had other age categories of patients.

4.b.3 **Association with gender of patient**

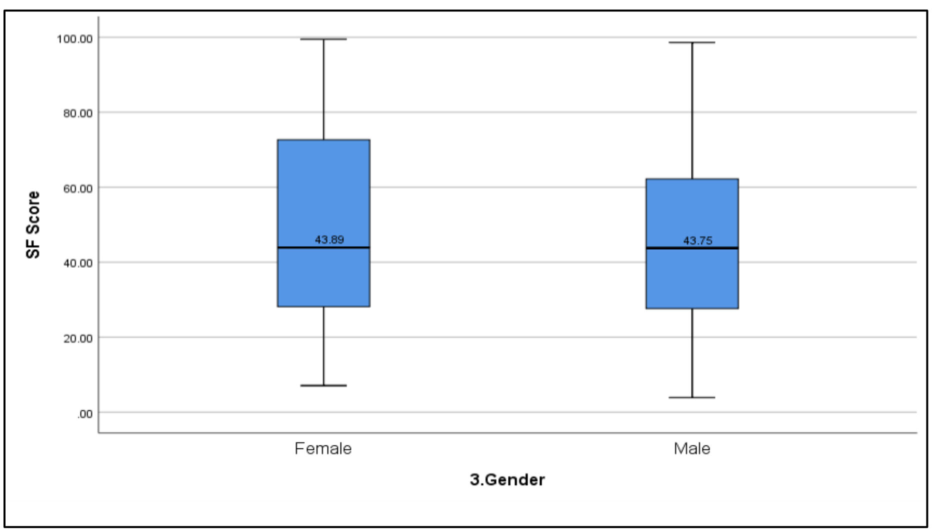


Figure 4.b.3 :SF Score Vs Gender

From the above figure it can be seen that the median SF- Score of patients nearly similar for both Female and male. And also spared of SF- Score is same for both genders.

4.b.4 Association with occupation of patient

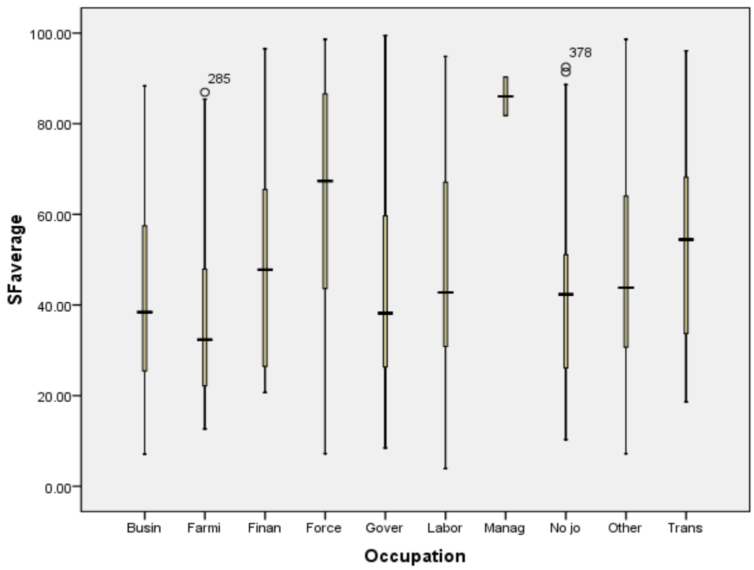


Figure 4.b.4 : SF Score Vs Occupation

The figure above shows that association of patient's occupation with the SF value. The patients who are farmers shows the least SF value as well as the highest SF value. Median SF value which had the patients who are in business and management section is greater than the median SF value which had the patients who have other jobs.

4.b.5 Association with job status (resign or leave) due to illness

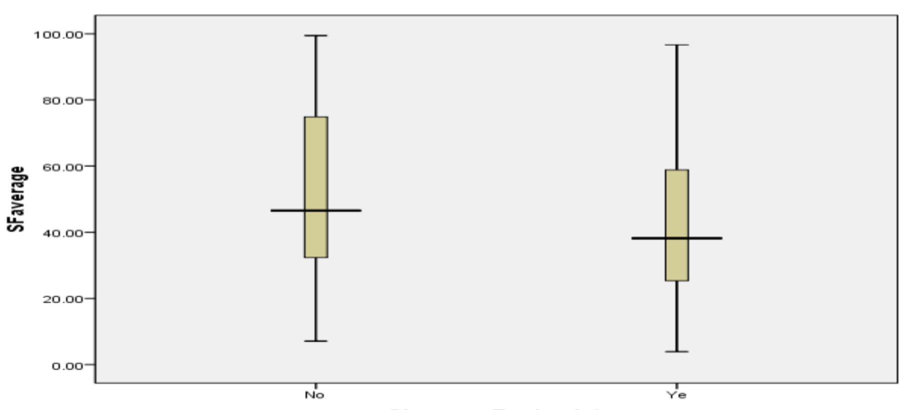


Figure 4.b.5 : SF Score Vs Job status due to illness

The figure above shows that association of job status (resign or leave with illness) with SF value. The patients of resign or leave job shows the least SF value as well as the highest SF value. Median SF value which had the patients of not resign or not leave job is greater than the median SF value which had the patients of resign or leave job due to their illness.

<u>4.b.6 Association with education level of the patient</u>

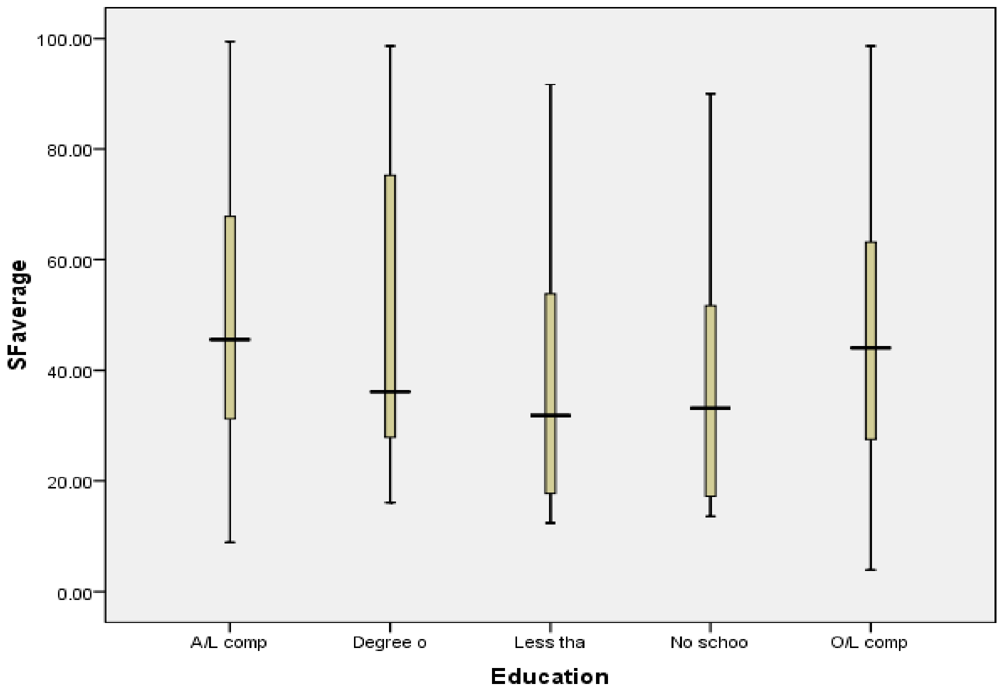


Figure 4.b.6 : SF Score Vs Education Level of patient

The figure above shows that association of education level of patient with the SF value. The patients who have less than grade 5 education level shows the least SF value as well as the highest SF value. Median SF value which had the patients who have completed A/L is greater than the median SF value which had the patients who have other education levels.

4.b.7 Association with highest education level of the household

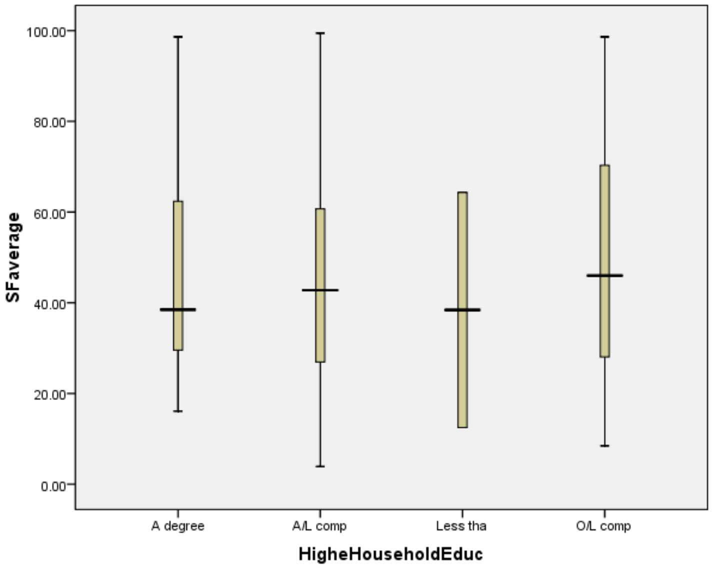


Figure 4.b.7 : SF Score Vs Highest Education Level of Household

The figure above shows that association of highest education level of household with the SF value. The patients of household who have less than grade 5 education level shows the least SF value as well as the highest SF value. Median SF value which had the patient of household who have completed O/L is greater than the median SF value which had the patients of household who have other highest education levels

4.b.8 Association with patient's accommodation status

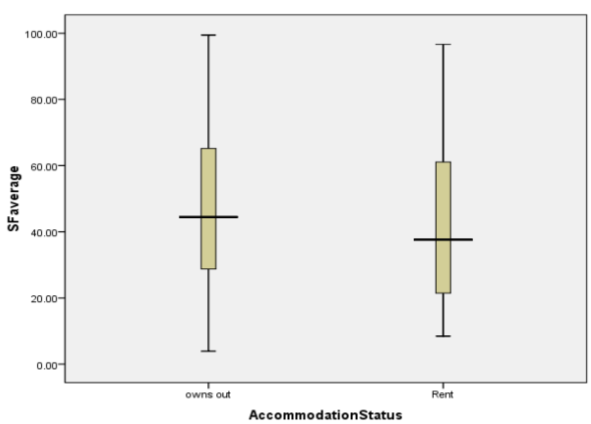


Figure 4.b.8 : SF Score Vs Accommodation Status

The figure above shows that association of patient's accommodation status with SF value. The patients who live in rent home shows the least SF value as well as the highest SF value. Median SF value which had the patient who live in own accommodation is greater than the median SF value which had the patient who live in rent home.

4.b.9 Association with number of children of the patient

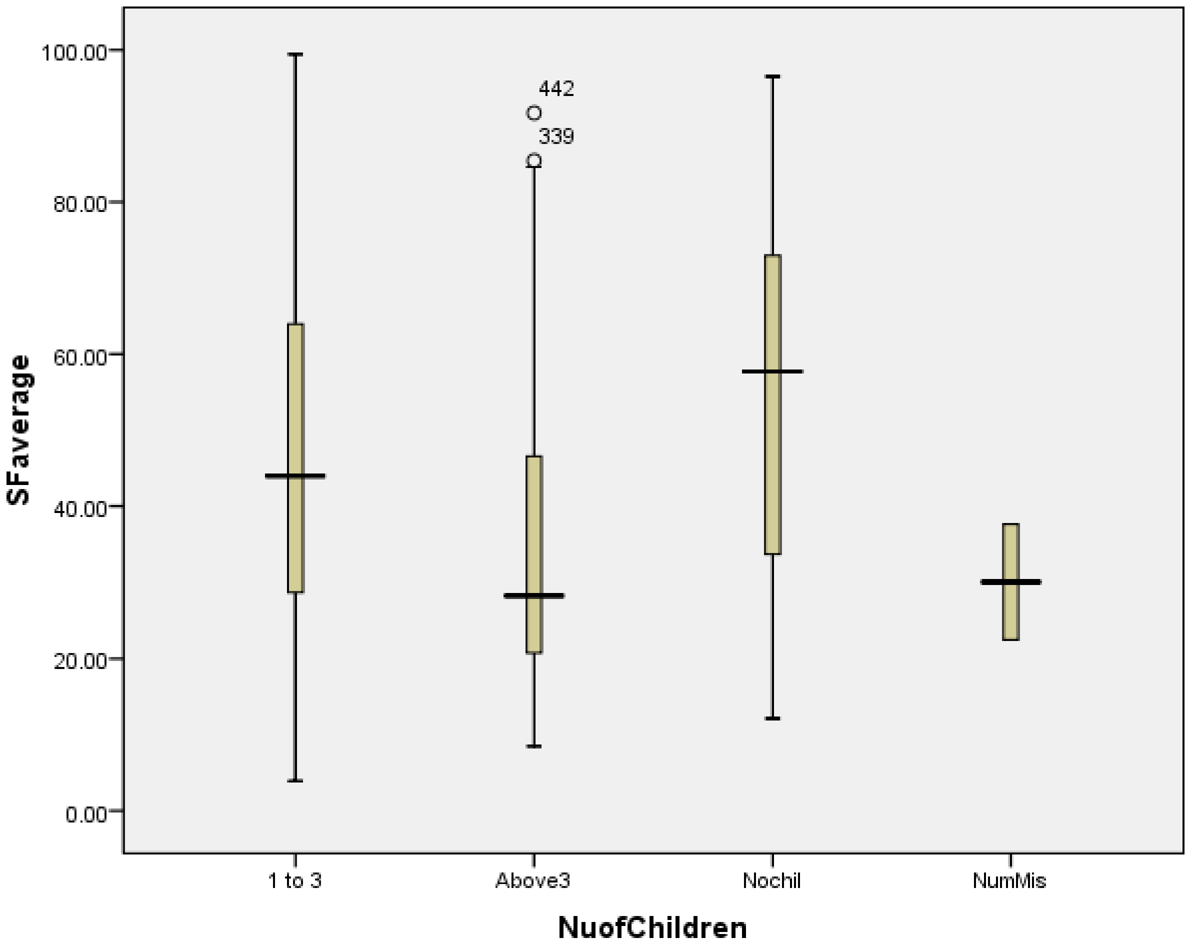


Figure 4.b.9 : SF Score Vs Number of Children of Patient

The figure above shows that association number of children with SF value. The patients who have more than 3 children shows the least SF value as well as the highest SF value. Median SF value which had the patient who don't have children is greater than the median SF value which had the patient who have children.

4.b.10 Association between SF – score and DSI - score

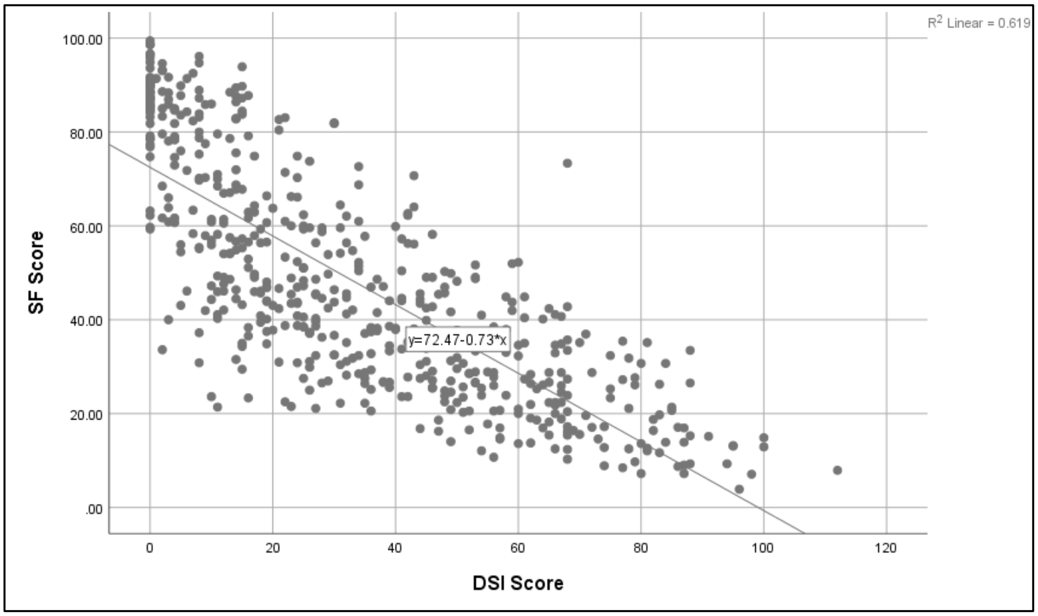


Figure 4.b. 10 : SF Score Vs DSI score

According to above graph there is a linear negative relationship between the SF- score and DSI score of patients. Correlation coefficient is 0.619 as shown in above graph and it can be observed that when the DSI score of a patient is high, the SF – score of patient is low.

4.b.11 Association between SF – score and Age in years

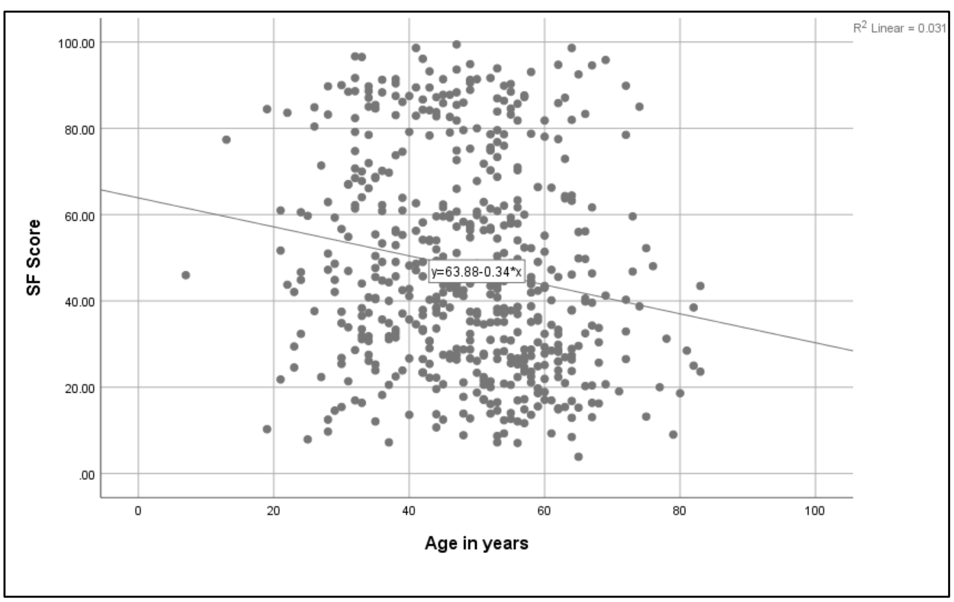


Figure 4.b. 11 : SF Score Vs DSI score

According to above graph there is a linear negative relationship between the SF- score and Age in years of patients. Correlation coefficient is 0.031 as shown in above graph and it can be observed that when the Age of a patient is high, the SF – score of patient is low.

4.c.1 Results from association analysis

Variable	Test used for Measure of association	Nature of the association	Value	P-value
Age in years	Pearson correlation coefficient	Negative	-0.177	0.000*
DSI Score	Pearson correlation coefficient	Negative	-0.787	0.000*
Average monthly income from all of the sources of income, before having this disease	Kendall's rank correlation	Positive	0.117	0.000*
Current average monthly income from all of the sources of income	Kendall's rank correlation	Positive	0.258	0.000*
Number of children	Kendall's rank correlation	Negative	-0.109	0.001*
Education	Kendall's rank correlation	Positive	0.073	0.028*
Have you had difficulties in performing current/ previous job due to illness	Kendall's rank correlation		-0.152	0.000*
Have you changed/ resigned from the job due to your illness	Kruskal-Wallis test			0.000*
Gender	Kruskal-Wallis test			0.280
District	Kruskal-Wallis test			0.293
Occupation	Kruskal-Wallis test			0.000*

Table 4.c.1 – Measures of association of predictors with SF-36 Score

Results from Table 4.c.1 suggest that patients with lower burden of symptoms reflected by DSI score had better quality of life ($p<0.0001$). Further, quality of life was higher for the patients who had relatively less difficulties in performing their current/previous job ($p<0.0001$). Patients who had higher level of average monthly income before and after having this disease ($p<0.0001$ and $p<0.0001$ respectively), and who had higher level of average family income ($p<0.0001$) had better quality of life. Quality of life was found to be better for patients with higher level of education ($p<0.0001$). Patients who had lower number of children ($p<0.0001$) also had better quality of life.

Further according to the association measures, quality of life differ significantly between patients who had changed/ resigned from the job and had not changed/ resigned from the job due to the illness ($p<0.0001$) and their occupation categories ($p<0.0001$).

5. Advanced Analysis

5.1 Advanced Analysis on test whether if quality of life of renal failure patients, vary with "Gender"

Normally the independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable. Under this objective we could use an independent t-test to understand whether SF- Score (Quality of life index of renal failure patients) differed based on gender. But part of the process involves checking to make sure that the data, it is required to analyze can actually be analyzed using an independent t-test. We need to do this because it is only appropriate to use an independent t-test if the **SF - score** should be **approximately normally distributed for each group of the gender,** that are required for an independent t-test to give a valid result.

In here Shapiro – Wilk test and Kolmogorov-Smirnov test used to assess normality. The resulted table given below.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	3.Gender	Statistic	df	Sig.	Statistic	df	Sig.
SF Score	Female	.102	128	.002	.942	128	.000
	Male	.077	410	.000	.953	410	.000

a. Lilliefors Significance Correction

Figure 5.1.1: Shapiro–Wilk test – SF score Vs Gender

Both Shapiro – Wilk test and Kolmogorov-Smirnov test suggest that SF- sore is **not normally distributed** on each group of the gender. Because significance values are less than the alpha value (we use 0.05 as our alpha value).

That implies SF-score not approximately normally distributed for each category of gender, therefore, it is required to go for non-parametric test named "Mann-Whitney U test". Before applying this test there needs to be satisfy the assumption "homogeneity of variance of SF-score on each group of the gender". Below is the resulted table of homogeneity of variance Levine's test.

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
SF Score	Based on Mean	3.620	1	536	.058
	Based on Median	2.470	1	536	.117
	Based on Median and with adjusted df	2.470	1	532.972	.117
	Based on trimmed mean	3.546	1	536	.060

Figure 5.1.2 : Test of Homogenety of variance

From Levine's test we can observe that base on the mean and median, P-value (Sig. Value) is greater than 0.05. It implies that the assumption, "homogeneity of variance of SF-score on each group of the gender" is satisfied. Then the Mann-Whitney U test can apply on this scenario.

Hypothesis

H_0 : There is no difference in Mean Rank in the two populations Male and Female

($M_{Male} = M_{Female}$)

H_1 : Mean Rank of population Male is different from mean rank of population Female

($M_{Male} \neq M_{Female}$)

Ranks

	Gender New	N	Mean Rank	Sum of Ranks
SF Score	Male	410	265.45	108835.00
	Female	128	282.47	36156.00
	Total	538		

Test Statistics[a]

	SF Score
Mann-Whitney U	24580.000
Wilcoxon W	108835.000
Z	-1.081
Asymp. Sig. (2-tailed)	$\boxed{.280}$

a. Grouping Variable: Gender New

Figure 5.1.3: Test of Mann-Whitney U test

Here it can be seen that Female category have 128 observations whose total sum of ranks is 36156.00. This results in a mean rank of 282.47. By contrast for Male category, it has 410 observations whose total sum of ranks is 108835.00. This results in a mean rank of 265.45. So gender category Female has a larger mean rank than gender category Male.

Then using test statistic table we can decide on whether this difference in mean ranks is significant or not. Here we see that the p value, quoted next to Asymp. Sig. (2-tailed), is 0.280, which is greater than 0.05. Therefore there is no significant evidence to reject the null hypothesis that, there is no difference in Mean Rank in the two populations Male and Female. (However, in this case the sample sizes are different for categories Male and Female). In conclusion, It can be say that "Quality of life of renal failure patients do not vary with gender" at 5% level of significance.

5.2 Advanced Analysis on test whether if quality of life of renal failure patients, vary with their "Level of Education"

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of more than two independent (unrelated) groups. Under this objective could use a one-way ANOVA to understand whether SF- Score (Quality of life index of renal failure patients) differed based on Education level of renal failure patients. Before doing the test there also need to do a Shapiro – Wilk test or Kolmogorov-Smirnov test used to assess normality.

The resulted table given below.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	10.New Education	Statistic	df	Sig.	Statistic	df	Sig.
SF Score	O/L not completed	.111	156	.000	.942	156	.000
	A/L not completed	.084	226	.001	.953	226	.000
	A/L completed and above	.108	156	.000	.946	156	.000

a. Lilliefors Significance Correction

Figure 5.2.1: Shapiro–Wilk test – SF score Vs level of Education

Both Shapiro – Wilk test and Kolmogorov-Smirnov test suggest that SF- sore is **not normally distributed** on each level of education. Because significance values are less than the alpha value (we use 0.05 as our alpha value).

That implies SF-score not approximately normally distributed for each category of education, therefore, it is required to go for non-parametric test named "Kruskal-Wallis test". Before applying this test there needs to be satisfy the assumption "homogeneity of variance of SF-score on each group of education". Below is the resulted table of homogeneity of variance Levine's test.

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
SF Score	Based on Mean	.414	2	535	.661
	Based on Median	.230	2	535	.795
	Based on Median and with adjusted df	.230	2	530.848	.795
	Based on trimmed mean	.404	2	535	.668

Figure 5.2.2 : Test of Homogenety of variance

From Levine's test we can observe that base on the mean and median, P-value (Sig. Value) is greater than 0.05. It implies that the assumption, "homogeneity of variance of SF-score on each group of the education" is satisfied. Then the Kruskal-Wallis test can apply on this scenario.

<u>Hypothesis</u>

H_0 : There is no difference in mean rank between the education groups of patients.

H_1 : There is a difference in mean rank in one or more education group.

Kruskal-Wallis Test

Ranks

	10.New Education	N	Mean Rank
SF Score	O/L not completed	156	245.52
	A/L not completed	226	275.60
	A/L not completed and above	156	284.65
	Total	538	

Test Statistics[a,b]

	SF Score
Kruskal-Wallis H	5.542
df	2
Asymp. Sig.	.063

a. Kruskal Wallis Test

b. Grouping Variable: 10.New Education

Figure 5.2.3: Kruskal Wallis test

Then using test statistic table, we can decide on whether this difference in mean ranks is significant or not. Here we see that the p value, quoted next to Asymp. Sig. , is 0.063, which is greater than 0.05. Therefore, there is no significant evidence to reject the null hypothesis that, there is a difference in mean rank in one or more education group. In conclusion, it can be said that "Quality of life of renal failure patients do not vary with their level of education" at 5% level of significance.

5.3 Advanced Analysis on test whether if quality of life of renal failure patients, vary with "number of children they have"

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of more than two independent (unrelated) groups. Under this objective could use a one-way ANOVA to understand whether SF- Score (Quality of life index of renal failure patients) differed based on number of children category they are in. Before doing the test there also need to do a Shapiro – Wilk test or Kolmogorov-Smirnov test used to assess normality.

The resulted table given below.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	New Number of children	Statistic	df	Sig.	Statistic	df	Sig.
SF Score	0	.119	50	.074	.962	50	.109
	1	.109	105	.004	.934	105	.000
	2	.075	202	.008	.966	202	.000
	>2	.110	181	.000	.922	181	.000

a. Lilliefors Significance Correction

Figure 5.3.1: Shapiro–Wilk test – SF score Vs Number of children they have

Both Shapiro – Wilk test and Kolmogorov-Smirnov test suggest that SF- sore is **not normally distributed** on each level of children count group except 1st category (number of children count =0). Because significance values are less than the alpha value except 1st category.

That implies SF-score not approximately normally distributed for each category of children count, therefore, it is required to go for non-parametric test named "Kruskal-Wallis test". Before applying this test there needs to be satisfy the assumption "homogeneity of variance of SF-score on each group of education". Below is the resulted table of homogeneity of variance Levine's test.

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
SF Score	Based on Mean	.282	3	534	.838
	Based on Median	.179	3	534	.911
	Based on Median and with adjusted df	.179	3	527.605	.911
	Based on trimmed mean	.251	3	534	.861

Figure 5.3.2 : Test of Homogenety of variance

From Levine's test we can observe that base on the mean and median, P-value (Sig. Value) is greater than 0.05. It implies that the assumption, "homogeneity of variance of SF-score on each group of the children count" is satisfied. Then the Kruskal-Wallis test can apply on this scenario.

Hypothesis

H_0 : There is no difference in mean rank between the groups of "Number of children they have".

H_1 : There is a difference in mean rank in one or more group.

Kruskal-Wallis Test

Ranks

	New Number of children	N	Mean Rank
SF Score	0	50	307.96
	1	105	275.66
	2	202	277.00
	>2	181	246.93
	Total	538	

Test Statistics[a,b]

	SF Score
Kruskal-Wallis H	7.513
df	3
Asymp. Sig.	.057

a. Kruskal Wallis Test

b. Grouping Variable: New Number of children

Figure 5.3.3: Kruskal Wallis test

Then using test statistic table, we can decide on whether this difference in mean ranks is significant or not. Here we see that the p value, quoted next to Asymp. Sig., is 0.057, which is greater than 0.05. Therefore, there is no significant evidence to reject the null hypothesis that, there is no difference in mean rank between the number of children groups. In conclusion, It can be say that Quality of life of renal failure patients do not vary with, "Number of children they have" at 5% level of significance.

5.4 Advanced analysis on the response variable – SF score (quality of life of renal failure patients)

Model fitting for SF- Score

A multiple linear regression model was fitted to the data. In model fitting, square root transformation was applied to the response variable (SF-36 Score) in order to satisfy the model assumptions. The observation of a U- shaped pattern in the residuals plot which notifies the violation of the assumptions. The residual plot achieved prior to square root transformation is as follows.

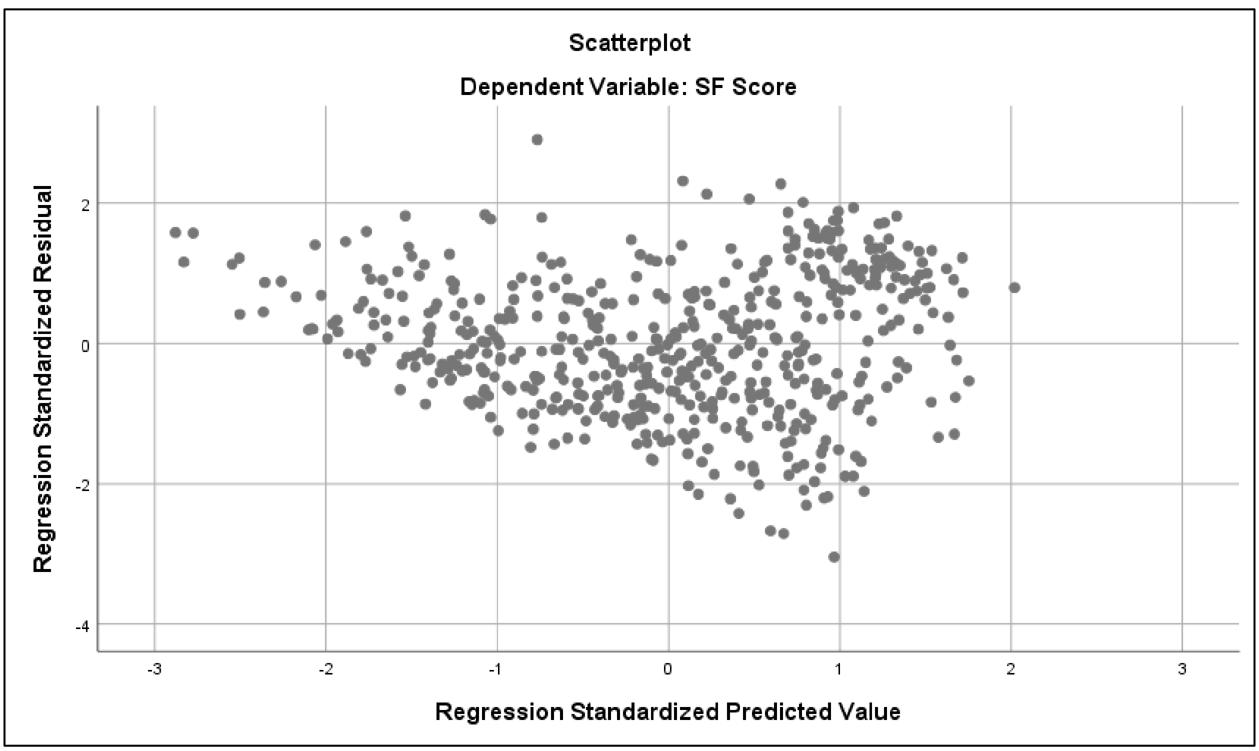


Figure 5.4.1: Residual's plot

Model summary

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.837[a]	.700	.694	1.001075484

a. Predictors: (Constant), Recorded Occupation, New Have you changed/ resigned from the job due to your illness, District New, Gender New, 2.age in years, 10.New Education, DSI Score, Average MONTHLY income BEFORE had diesease (Recorded), New Number of children , CURRENT average MONTHLY income (Recorded), New Have you had difficulties in performing current/ previous job due to illness

b. Dependent Variable: Squreroot of SF score

Figure 5.4.2: Model Summary

From the above adjusted R_Squared value it can be stated that: 69.4% of the variability of the Square root (SF-Score) variable can be explained by the above regression model. Thus the model can be stated to be a quite a good model.

Statistical significance

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1218.230	11	110.748	110.510	.000[b]
	Residual	522.121	521	1.002		
	Total	1740.351	532			

a. Dependent Variable: Squreroot of SF score

b. Predictors: (Constant), Recorded Occupation, New Have you changed/ resigned from the job due to your illness, District New, Gender New, 2.age in years, 10.New Education, DSI Score, Average MONTHLY income BEFORE had diesease (Recorded), New Number of children , CURRENT average MONTHLY income (Recorded), New Have you had difficulties in performing current/ previous job due to illness

Figure 5.4.3: ANOVA Table

The F-ratio in the **ANOVA** table tests whether the overall regression model is a good fit for the data. The above table shows that the independent variables statistically significantly predict the dependent variable, $F(11, 521) = 110.510$, $p < .0005$ (i.e., the regression model is a good fit of the data).

Estimated model coefficients

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	8.242	.342		24.086	.000		
	2.age in years	-.012	.004	-.083	-3.308	.001	.920	1.087
	DSI Score	-.052	.002	-.754	-29.516	.000	.882	1.134
	Average MONTHLY income BEFORE had diesease (Recorded)	9.030E-6	.000	.075	2.308	.021	.543	1.842
	CURRENT average MONTHLY income (Recorded)	1.825E-5	.000	.168	4.718	.000	.452	2.212
	New Number of children	-.023	.049	-.012	-.477	.634	.881	1.135
	10.New Education	-.090	.062	-.038	-1.456	.146	.854	1.170
	New Have you changed/ resigned from the job due to your illness	.124	.138	.034	.897	.370	.394	2.540
	New Have you had difficulties in performing current/ previous job due to illness	-.149	.130	-.041	-1.143	.253	.449	2.229
	Gender New	.129	.104	.030	1.235	.217	.948	1.055
	District New	.006	.016	.009	.386	.700	.967	1.034
	Recorded Occupation	.044	.015	.072	2.858	.004	.920	1.087

a. Dependent Variable: Squreroot of SF score

Figure 5.4.4: Coefficient Table

<u>Thus, the model can be represented as follows:</u>

Predicted (SF – Score) = 8.242 - (0.012 *age in years) – (0.052* DSI Score) + (0.000009*Average MONTHLY income BEFORE had disease) + (0.000018*CURRENT average MONTHLY income) + (0.044* Occupation)

According to the model, following variables have some relationship with SF- Score (Quality of life of renal failure patients).

- **Age in years**
- **Occupation**
- **Average monthly income before had disease**
- **Current average monthly income**
- **DSI Score**

Goodness of Fit of the model

The fit of the above model is then assessed using the **normal probability plot** and the **residuals plot** as shown below.

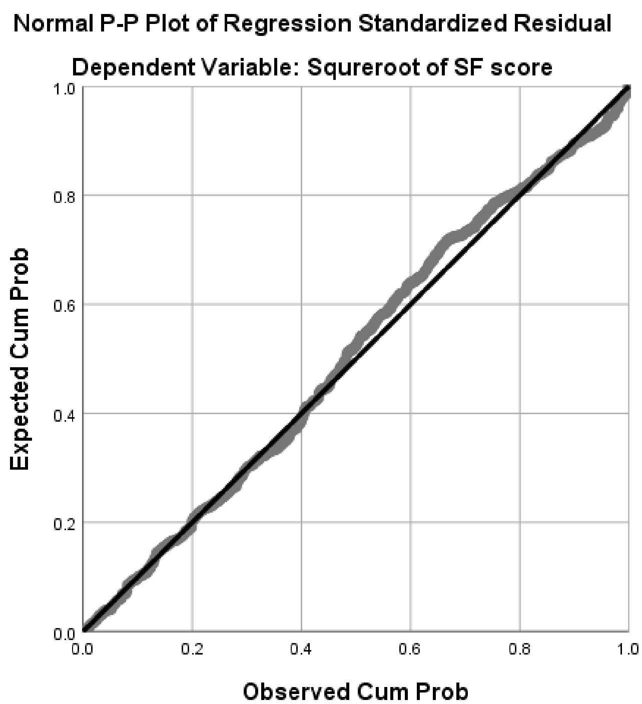


Figure 5.4.5: **Normal Probability Plot for the above model**

The above graph **normal probability plot** can be used to get an idea of how closely the data points follow the fitted distribution line by plotting each residual against its expected value when the distribution is normal. According to the above graph the points falls closely along the fitted line and it can be observed to be nearly linear. The model can thus be said to be in agreement with normality and of a good fit.

Residual Plot

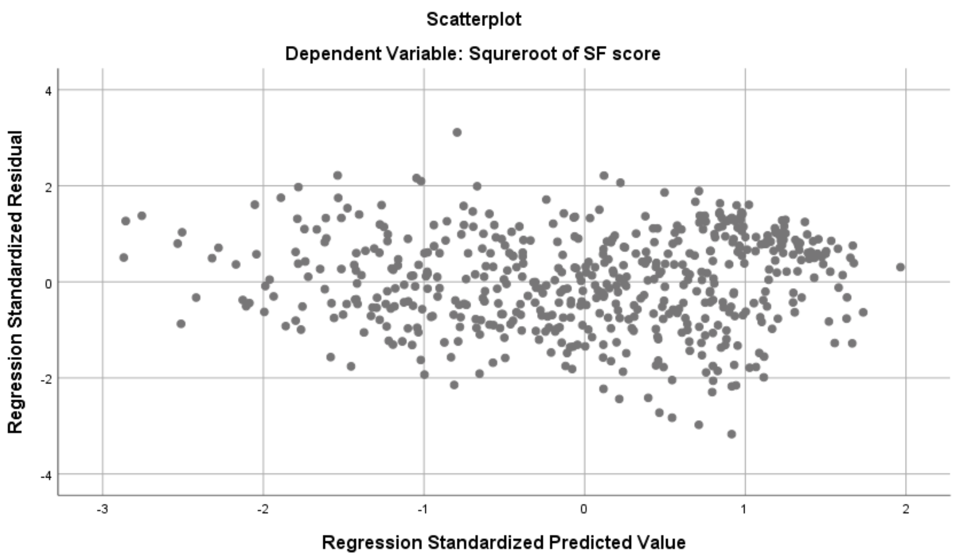


Figure 5.4.6: Residuals Plot for the above model

From the above residual plot it can be seen that almost all of the residuals are nearly randomly scattered around 0-axis within a horizontal band. This denotes that all the assumptions of linearity, independence and constant variance of residuals are nearly satisfied.

6. Conclusion

- According to the regression model **Age in years, Occupation, Average monthly income before had disease**, **Current average monthly income** and **DSI Score** are the variables, which are having some association with quality of life of renal failure patients.
- Also looking at Figure xxx , Figure xxxx and Table xxx obtained above, it was also observed that above mentioned variables have a relationship with SF – Score once considered individually as well.
- Age in years and DSI Score are negatively associated with the quality of life of Renal failure patients.
- Occupation, Average monthly income before had disease and Current average monthly income are positively associated with the quality of life of Renal failure patients.
- When looking at the dummy coded variables included in the model, Occupation, Average monthly income before had disease and Current average monthly income also have association with a SF- Score
- According to the table xxx in descriptive analysis, **Number of children**, **Education, Have you had difficulties in performing current/ previous job due to illness** and **Have you changed/ resigned from the job due to your illness** variables are significant but they are not in the model.
- But according to the advanced analysis, we observed that Quality of life of renal failure patients do not vary with variables, Number of children, Education and Gender at 5% level of significance.

7. Discussion

The above analysis is primarily done to identify the variables associated with quality of life of renal failure patients.

The features relating to the quality of life are identified above. This was concluded depending mainly on the descriptive analysis and the regression model obtained above.

When looking at the goodness of the model the R_Squared value amounts to 69.4% which demonstrates a quite good model. This could also be verified using the respective normal probability plot and the residual plot that was drawn in the above analysis. According to these measurements the regression model derived above can be stated to be of a moderate fit.

Occupation of patients, Education and number of children variables were re-categorized prior to model fitting. Occupation was re-categorized due to the presence of large number of occupation categories among patients and number of children was re-categorized as it is more meaningful.

An ANOVA test could have also been done as an alternative for the above procedure of fitting a multiple linear regression model.

To investigate the other objectives, non-parametric tests have been done, due to violation of normality assumption of respective variables.

References

- https://nicst.com/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4639348/\
- https://jpgim.sljol.info/articles/abstract/10.4038/jpgim.8126/
- Gerasimoula K., Lefkothea L., Maria L., Victoria A., Paraskevi T., Maria P. *quality of life in hemodialysis patients* (2015).
- Zamanian H., Kharameh Z. T. Translation and Psychometric Properties of the Persian Version of the Dialysis Symptom Index in Hemodialysis Patients. *Nephro Urol Mon*. (2015).
- Weisbord S. D., Fried L. F., Arnold R. M., Fine M. J.,Levenson D. J., Peterson R. A., Switzer G. E. Prevalence, Severity, and Importance of Physical and Emotional Symptoms in Chronic Hemodialysis Patients. *American Society of Nephrology.* (2005)