Chapter 4 : Results & Discussion

4.1 Characteristics of woman respondents

First, descriptive analysis using frequency tabulation was conducted. The Lists in Table4.1 indicate that 32,599 women (age 15-49 years) identified in the selected households, 26,923 were successfully interviewed, yielding a response rate of 82.6 percent. It is important to note that while the average response rate for women's was over 90 percent in 11 states, between 80 and 90 percent in five states, between 70 and 80 per cent in two states, between 60 and 70 percent in three states and between 50 and 60 percent in four states, being highest in Gezira at 98.6 per cent and the lowest in Western Bahr El Ghazal at 55.4 per cent. as indicated in Table 4.1, the response rate for women was low. The response rate for women's questionnaire was less than 60 per cent in four states in Southern Sudan. Table4.2 display the characteristics of female respondents 15-49 years of age.
Table 4.2 Women in the age group 25-29 years constituted the largest proportion (21.1 %) of the total number of women followed by women in the age group 20-24 years (18.7 per cent), women in the age group 15-19 years (17.7 per cent), women in the age group 30-34 years (14.9 per cent), and women in the age group 35-39 years (14.1 %). About 8% of the women were in the age group 40-44 years while the lowest proportion of women was in the age group 45-49 years (5.5 per cent). About 65.5 percent were currently married/in union and 28.6 per cent were formerly married/in union while never married/in union women constituted 5.9 percent. Women with no formal education made up 49.8 percent of the total while 41.2 per cent had primary education and 8.9 percent had secondary or higher education. The wealth index quintiles show that about 17.7 percent of women

belong to the poorest households while women from the richest households constitute about 23.5 percent.

Table 4.1: distribution Number of women response rates

state	Completed	Not at home	Refused	Partly completed	Incapacitated	Other	Total	Response rate %
Northern	1290	54	0	0	16	20	1380	93.5%
River Nile	1408	54	2	0	7	1	1472	95.7%
Red Sea	1139	17	3	1	3	12	1175	96.9%
Kassala	1200	14	0	0	7	20	1241	96.7%
Gadarif	1207	44	5	0	8	26	1290	93.6%
Khartoum	1324	183	13	1	1	34	1556	85.1%
Gezira	1533	13	0	0	4	5	1555	98.6%
Sinnar	1347	21	0	0	1	17	1386	97.2%
Blue Nile	1220	101	5	0	5	6	1337	91.2%
White Nile	1500	23	1	0	6	4	1534	97.8%
North kordofan	1258	55	3	0	8	14	1338	94.0%
South kordofan	905	140	3	0	0	12	1060	85.4%
North Darfur	1055	104	4	0	2	32	1197	88.1%
West Darfur	773	97	6	1	1	24	902	85.7%
South Darfur	1027	39	1	0	5	12	1084	94.7%
Jongolei	887	197	33	0	0	339	1456	60.9%
Upper Nile	612	223	17	0	1	101	954	64.2%
Unity	906	274	38	2	1	92	1313	69.0%
Warab	1046	172	24	0	1	114	1357	77.1%
NorthBahrl_Gazal	837	308	31	3	0	319	1498	55.9%
West Bahr Al_Gazal	717	287	18	1	0	272	1295	55.4%
Lakes	899	352	63	0	1	170	1485	60.5%
West Equatoria	825	303	13	0	1	53	1195	69.0%
Central Equatoria	1067	242	43	17	0	47	1416	75.4%
East Equatoria	941	105	11	0	0	66	1123	83.8%
Total	26923	3422	337	26	79	1812	32599	82.6%

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.2 : Women's characteristics

			Number of woman	
			weighted	unweighted
Age	15-19	Count	1529508	4677
		% of Total	17.7%	17.4%
	20-24	Count	1611527	5005
		% of Total	18.6%	18.6%
	25-29	Count	1835955	5847
		% of Total	21.2%	21.7%
	30-34	Count	1291155	4037
		% of Total	14.9%	15.0%
	35-39	Count	1217325	3778
		% of Total	14.1%	14.0%
	40-44	Count	696905	2099
		% of Total	8.0%	7.8%
	45-49	Count	475590	1479
		% of Total	5.5%	5.5%
Total		Count	8657965	8657965
		% of Total	100.0%	100.0%
Marital/Union status	Currently married/in union	Count	5435614	17216
		% of Total	66.1%	67.8%
	Formerly married/in union	Count	2292572	6688
		% of Total	27.9%	26.3%
	Never married/in union	Count	495020	1487
		% of Total	6.0%	5.9%
Total		Count	8223206	25391
		% of Total	100.0%	100.0%
Motherhood status Ever given birth	Yes	Count	5615186	17882
		% of Total	64.9%	66.4%
	No	Count	3041795	9034
		% of Total	33.6%	35.1%
Total		Count	8656981	26916
		% of Total	100.0%	100.0%
Education	None	Count	4353377	14716
		% of Total	50.3%	54.7%
	Primary	Count	3508224	10383
		% of Total	40.5%	38.6%

	Secondary +	Count	784808	1776
		% of Total	9.1%	6.6%
	Missing/DK	Count	11981	48
		% of Total	.1%	.2%
Total		Count	8658390	26923
		% of Total	100.0%	100.0%
Wealth index quintiles	Poorest	Count	1611387	5067
		% of Total	21.4%	21.1%
	Second	Count	1497565	4720
		% of Total	19.9%	19.6%
	Middle	Count	1357048	4329
		% of Total	18.0%	18.0%
	Fourth	Count	1051533	3342
		% of Total	14.0%	13.9%
	Richest	Count	700768	2282
		% of Total	9.3%	9.5%
Total		Count	6218301	19740
		% of Total	82.6%	82.1%

Source : reasercher analysis of SHHS 2006 data by SPSS

4.1.1 Describing the Pattern of Missing Data

Table 4.3 : Univariate Statistics Pattern of Missing Data

	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
Marital/Union status	8210555	1.40	.600	2182830	21.0	0	1487
Wealth index quintiles	7513617	2.66	1.333	2879768	27.7	0	0
Education	8645015	1.60	.707	1748370	16.8	0	48
Ever given birth	8643611	1.35	.477	1749774	16.8	0	0
Age of Woman	8645015			1748370	16.8		

a. Number of cases outside the range (Mean - 2*SD, Mean + 2*SD).

Source : reasercher analysis of SHHS 2006 data by SPSS

Table4.3 Indicate that withl5 (Wealth index quintile) has the greatest number of cases with missing values (27.7%), while age (Age of woman), melevel (level of

education) and cm1 (ever given birth) has the least (16.8%). Marital/Union status has the greatest number of extreme values.

Table 4.4 : Separate Variance t Tests[a] Pattern of Missing Data

		Marital/Union status	Wealth index quintiles	Education	Ever given birth
Marital/Union status	t	.	-74.5-	217.8	432.8
	df	.	2572921.1	483505.4	532098.8
	# Present	8210555	5887772	8210555	8209802
	# Missing	0	1625845	434460	433809
	Mean(Present)	1.40	2.64	1.61	1.36
	Mean(Missing)	.	2.72	1.37	1.13
Wealth index quintiles	t	131.5	.	-182.1-	-129.6-
	df	4420726.6	.	4362590.7	4344345.9
	# Present	5887772	7513617	6208445	6207041
	# Missing	2322783	0	2436570	2436570
	Mean(Present)	1.42	2.66	1.57	1.34
	Mean(Missing)	1.36	.	1.67	1.39
Education	t	.	-88.1-	.	.
	df	.	1873585.8	.	.
	# Present	8210555	6208445	8645015	8643611
	# Missing	0	1305172	0	0
	Mean(Present)	1.40	2.64	1.60	1.35
	Mean(Missing)	.	2.75	.	.
Ever given birth	t	1905.8	-88.0-	23.4	.
	df	8209817.1	1876894.5	1404.1	.
	# Present	8209802	6207041	8643611	8643611
	# Missing	753	1306576	1404	0
	Mean(Present)	1.40	2.64	1.60	1.35
	Mean(Missing)	1.00	2.75	1.31	.
AGE OF WOMAN	t	.	-88.1-	.	.
	df	.	1873585.8	.	.
	# Present	8210555	6208445	8645015	8643611
	# Missing	0	1305172	0	0
	Mean(Present)	1.40	2.64	1.60	1.35
	Mean(Missing)	.	2.75	.	.

For each quantitative variable, pairs of groups are formed by indicator variables (present, missing).

a. Indicator variables with less than 5% missing are not displayed.

Source : reasercher analysis of SHHS 2006 data by SPSS

Table4.4 show that when wealth is missing, the mean education is 1.57, compared to 1.67 when wealth is no missing. In fact, the missingness of wealth seems to affect the means of several of the quantitative (scale) variables. This is one indication that the data may not be missing completely at random.

Table 4.5 : mstatus(Marital status) Pattern of Missing Data

			Total	Currently married/in union	Formerly married/in union	Never married/in union	Missing SysMis
WM9	Present	Count	8645015	5427278	2289044	494233	434460
		Percent	83.2	100.0	100.0	100.0	19.9
	Missing	% SysMis	16.8	.0	.0	.0	80.1
CM1	Present	Count	8643611	5426525	2289044	494233	433809
		Percent	83.2	100.0	100.0	100.0	19.9
	Missing	% SysMis	16.8	.0	.0	.0	80.1
melevel	Present	Count	8645015	5427278	2289044	494233	434460
		Percent	83.2	100.0	100.0	100.0	19.9
	Missing	% SysMis	16.8	.0	.0	.0	80.1
wlthind5	Present	Count	7513617	3804983	1715323	367466	1625845
		Percent	72.3	70.1	74.9	74.4	74.5
	Missing	% SysMis	27.7	29.9	25.1	25.6	25.5

Indicator variables with less than 5% missing are not displayed.

Looking at the Table 4.5 for *melevel (Marital status)*, the number of missing values in the indicator variables does not appear to vary much between *melevel (marital status)* categories. Unmarried people reported *wm9 (Age of woman)* 100.0% of the time, and married people reported the same variable 100.0% of the time. The difference is none.

Table 4.6 : melevel(education) **Pattern of Missing Data**								
			Total	None	Primary	Secondary +	Missing/DK	Missing SysMis
WM9	Present	Count	8645015	4346614	3502454	783983	11964	0
		Percent	83.2	100.0	100.0	100.0	100.0	.0
	Missing	% SysMis	16.8	.0	.0	.0	.0	100.0
CM1	Present	Count	8643611	4345644	3502020	783983	11964	0
		Percent	83.2	100.0	100.0	100.0	100.0	.0
	Missing	% SysMis	16.8	.0	.0	.0	.0	100.0
mstatus	Present	Count	8210555	4048820	3384147	766910	10678	0
		Percent	79.0	93.1	96.6	97.8	89.3	.0
	Missing	% SysMis	21.0	6.9	3.4	2.2	10.7	100.0
wlthind5	Present	Count	7513617	3248990	2432898	517820	8737	1305172
		Percent	72.3	74.7	69.5	66.0	73.0	74.7
	Missing	% SysMis	27.7	25.3	30.5	34.0	27.0	25.3
Indicator variables with less than 5% missing are not displayed.								

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.7 : wlthind5(wealth) **Pattern of Missing Data**									
			Total	Poorest	Second	Middle	Fourth	Richest	Missing SysMis
WM9	Present	Count	8645015	1608916	1495252	699573	1049835	699573	2436570
		Percent	83.2	84.4	82.4	79.2	82.6	79.2	84.6
	Missing	% SysMis	16.8	15.6	17.6	20.8	17.4	20.8	15.4
CM1	Present	Count	8643611	1608916	1494420	699573	1049618	699573	2436570
		Percent	83.2	84.4	82.4	79.2	82.6	79.2	84.6
	Missing	% SysMis	16.8	15.6	17.6	20.8	17.4	20.8	15.4
mstatus	Present	Count	8210555	1529072	1412702	664627	998739	664627	2322783
		Percent	79.0	80.2	77.9	75.3	78.6	75.3	80.7
	Missing	% SysMis	21.0	19.8	22.1	24.7	21.4	24.7	19.3
melevel	Present	Count	8645015	1608916	1495252	699573	1049835	699573	2436570
		Percent	83.2	84.4	82.4	79.2	82.6	79.2	84.6
	Missing	% SysMis	16.8	15.6	17.6	20.8	17.4	20.8	15.4
Indicator variables with less than 5% missing are not displayed.									

Source : reasercher analysis of SHHS 2006 data by SPSS

consider the cross tabulation Table 4.6 for melevel (Level of education). If a respondent has at least some secondary+ education, a response for marital status is more to be missing. At least 93.1% of the respondents with none education reported marital status. On the other hand only 97.8% of those with a secondary reported marital status, the number is even lower for those with none education.

consider the cross tabulation Table 4.7 for wlthind5 (wealth). If a respondent has at least some wealth, a response for melevel (education level) is more to be missing. At least 84.4% of the respondents with poorest wealth reported melevel (education). On the other hand, only 82.6% of those with Middle reported melevel (education level). The number is even lowering for those with richest.

Table 4.8 : EM Estimated Statistics

EM Means[a]	
WM9(age of woman)	
28.20	
a. Little's MCAR test: Chi-Square = .001, DF = 0, Sig. =.	
EM Covariances[a]	
	WM9
WM9	73.123
a. Little's MCAR test: Chi-Square = .001, DF = 0, Sig. =.	
EM Correlations[a]	
	WM9
WM9	1
a. Little's MCAR test: Chi-Square = .001, DF = 0, Sig. =.	

Table 4.8 describe that the null hypothesis for Little's MCAR test is that the data are missing completely at random (MCAR). Because the significance value is less

than 0.05 in our work, we can conclude that the data are *not* missing completely at random. This confirms the conclusion we drew from the descriptive statistics and tabulated patterns.

4.1.2 Using Multiple Imputations to Complete and Analyze a Dataset

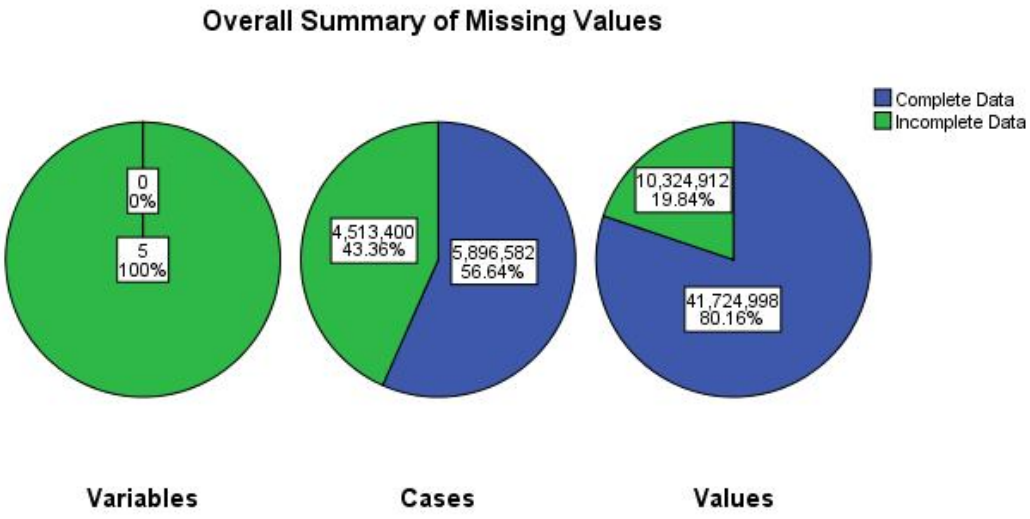

Fig. 4.1

Fig. 4.1 shows that:

- The *Variables* chart shows that each of the 5 analysis variables has at least one missing value on a case.
- The *Cases* chart shows that 4,513,400 of the 10,000,000 cases have at least one missing value on a variable.
- The *Values* chart shows that 10,324,912 of the 50,000,000 values (cases × variables) are missing.
- There are 5896582 (56.64 %) complete cases and 80.16% complete values.

4.1.3 Imputation Models

Table 4.9 : Imputation Specifications	
Imputation Method	Automatic
Number of Imputations	5
Model for Scale Variables	Linear Regression
Interactions Included in Models	(none)
Maximum Percentage of Missing Values	100.0%
Maximum Number of Parameters in Imputation Model	100
Replication Weight Variable	wmweight

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.10 : Imputation Results		
Imputation Method		Fully Conditional Specification
Fully Conditional Specification Method Iterations		10
Dependent Variables	Imputed	WM9,CM1,mstatus,melevel,wlthind5
	Not Imputed(Too Many Missing Values)	
	Not Imputed(No Missing Values)	
Imputation Sequence		WM9,melevel,CM1,mstatus,wlthind5

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.11 : Imputation Models				
	Model		Missing Values	Imputed Values
	Type	Effects		
Age of woman	Linear Regression	melevel,CM1,mstatus,wlthind5	1307138	6535690
Education	Logistic Regression	CM1,mstatus,wlthind5,WM9	1307138	6535690
Ever given birth	Logistic Regression	melevel,mstatus,wlthind5,WM9	1308549	6542745
Marital/Union status	Logistic Regression	melevel,CM1,wlthind5,WM9	1742478	8712390
Wealth index quintiles	Logistic Regression	melevel,CM1,mstatus,WM9	2440309	12201545

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.12 : WM9(age of woman) imputed values						
Data	Imputation	N	Mean	Std. Deviation	Minimum	Maximum
Original Data		8658984	28.37	8.636	15.00	49.00
Imputed Values	1	1307138	28.76	8.528	-4.47-	56.87
	2	1307138	28.39	8.702	-7.61-	57.23
	3	1307138	28.59	8.441	-.81-	57.55
	4	1307138	28.90	8.564	-.41-	57.32
	5	1307138	28.94	8.599	-3.80-	57.55
Complete Data After Imputation	1	9966122	28.42	8.623	-4.47-	56.87
	2	9966122	28.37	8.645	-7.61-	57.23
	3	9966122	28.40	8.611	-.81-	57.55
	4	9966122	28.44	8.628	-.41-	57.32
	5	9966122	28.44	8.633	-3.80-	57.55

Source : reasercher analysis of SHHS 2006 data by SPSS

The descriptive statistics Table 4.12 for wm9 (Age of woman) shows means and standard deviations in each set of imputed values roughly equal to those in the original data; however, an immediate problem presents itself when you look at the

minimum and see that negative values for age have been imputed. We will need to run a custom model with constraints on certain variables. However, age shows other potential problems. The mean values for each imputation are considerably higher than for the original data, and the maximum values for each imputation are considerably lower than for the original data. The distribution of age tends to be highly right-skew, so this could be the source of the problem.

4.1.4 Custom Imputation Model

wm9(age of woman's) is highly right-skew, and further analysis will likely use the logarithm of age, so it seems sensible to impute the log-age directly see Table 4.13.

Table 4.13 : logage

Data	Imputation	N	Mean	Std. Deviation	Minimum	Maximum
Original Data		8658984	3.2982	.30925	2.7081	3.8918
Imputed Values	1	1307138	3.3007	.30916	2.2083	4.3542
	2	1307138	3.3056	.31072	2.1610	4.3468
	3	1307138	3.3097	.30939	2.2814	4.3984
	4	1307138	3.2919	.30790	2.2108	4.3906
	5	1307138	3.2924	.31190	2.2033	4.4032
Complete Data After Imputation	1	9966122	3.2986	.30924	2.2083	4.3542
	2	9966122	3.2992	.30945	2.1610	4.3468
	3	9966122	3.2997	.30929	2.2814	4.3984
	4	9966122	3.2974	.30908	2.2108	4.3906
	5	9966122	3.2975	.30961	2.2033	4.4032

Source : reasercher analysis of SHHS 2006 data by SPSS

The descriptive statistics in Table 4.13 for logage (age of woman) under the custom imputation model with constraints shows that the problem of negative imputed values for tenure has been solved

4.1.5 Nominal Regression

Table 4.14 Case Processing Summary			
		N	Marginal Percentage
Result of women 's interview	Completed	153204	87.7%
	Not at home	12880	7.4%
	Refused	1230	.7%
	Partly completed	85	.0%
	Incapacitated	295	.2%
	Other	7040	4.0%
Ever given birth	Yes	115766	66.3%
	No	58968	33.7%
Marital/Union status	Currently married/in union	117017	67.0%
	Formerly married/in union	47510	27.2%
	Never married/in union	10207	5.8%
Education	None	93235	53.4%
	Primary	66909	38.3%
	Secondary +	11895	6.8%
	Missing/DK	2695	1.5%
Wealth index quintiles	Poorest	44191	25.3%
	Second	41833	23.9%
	Middle	38289	21.9%
	Fourth	29543	16.9%
	Richest	20878	11.9%
Valid		174734	100.0%
Missing		20860	
Total		195594	
Subpopulation		24001[a]	
a. The dependent variable has only one value observed in 24001 (100.0%) subpopulations.			

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.15 Model Fitting Information				
	Model Fitting Criteria	Likelihood Ratio Tests		
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	169938.358			
Final	160575.209	9363.149	55	.000

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.16 Pseudo R-Square

Cox and Snell	.052
Nagelkerke	.084
McFadden	.055

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.17 Likelihood Ratio Tests				
	Model Fitting Criteria	Likelihood Ratio Tests		
Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	160575.209[a]	.000	0	.
logage	160576.806	1.597	5	.902
melevel	169534.249	8959.040	15	.000
CM1	160645.603	70.394	5	.000
mstatus	160618.077	42.868	10	.000
wlthind5	160840.481	265.272	20	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.
a. This reduced model is equivalent to the final mode l because omitting the effect does not increase the degrees of freedom.

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.18 : Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	169938.358			
Final	160575.209	9363.149	55	.000

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.19 Pseudo R-Square

Cox and Snell	.052
Nagelkerke	.084
McFadden	.055

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.20 Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	160575.209[a]	.000	0	.
logage	160576.806	1.597	5	.902
melevel	169534.249	8959.040	15	.000
CM1	160645.603	70.394	5	.000
mstatus	160618.077	42.868	10	.000
wlthind5	160840.481	265.272	20	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Source : reasercher analysis of SHHS 2006 data by SPSS

4.1.6 Two-step Cluster Analysis

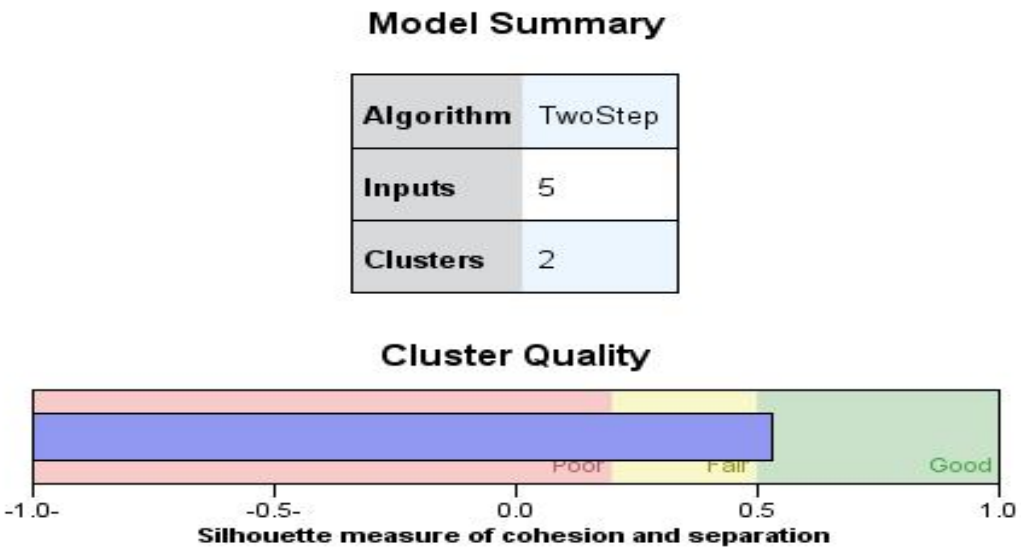

Fig. 4.1 Model Summary

- The model summary table in Fig.4.2 indicates that tow clusters were found based on the seven input features (fields) selected.
- The cluster quality chart in Fig.4. 2, Fig. 4.5 and Fig. 4.10 indicates that the model summary quality is "Good" while quality chart in Fig. 4.6, Fig. 4.8 and Fig. 4.12 indicates that the model summary quality is "Fair".

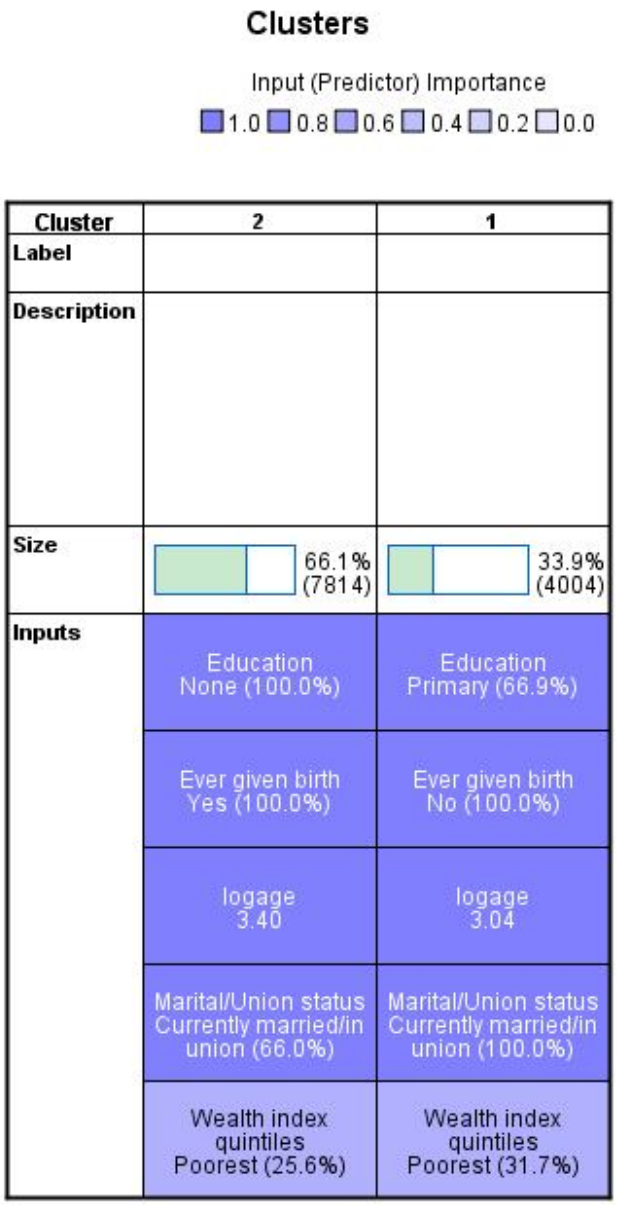

Fig.4. 2 Custer

The Cluster Sizes view in Fig. 3 shows the frequency of each cluster. Hovering over a slice in the pie chart reveals the number of records assigned to the cluster. 33.9% (4004) of the records were assigned to the first cluster and 66.1% (7814) to the second.

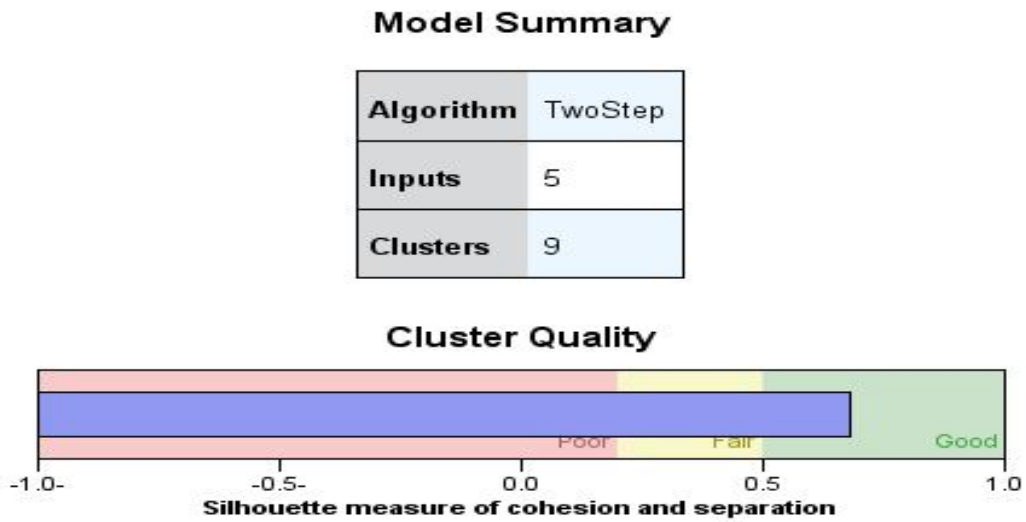

Fig. 4.3 Model Summary

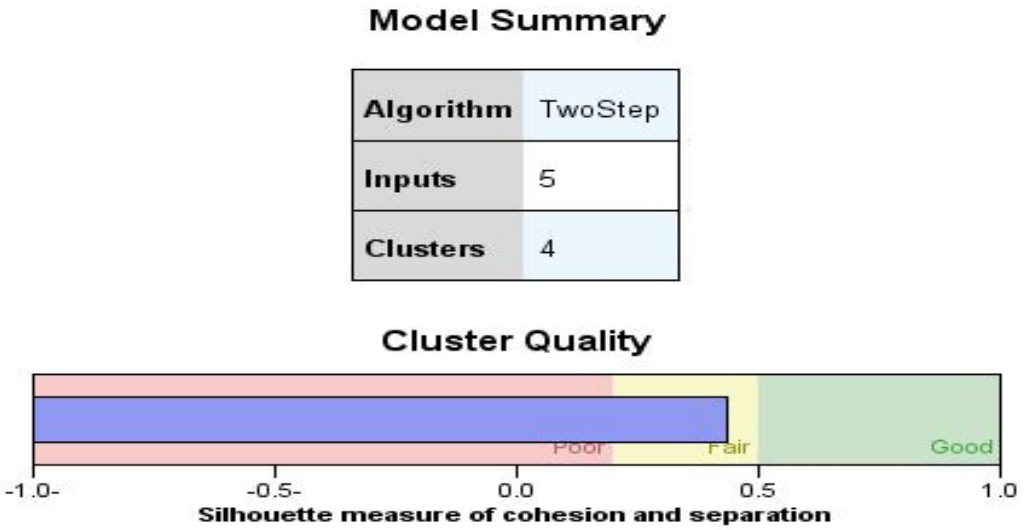

Fig. 4.4 Model Summary

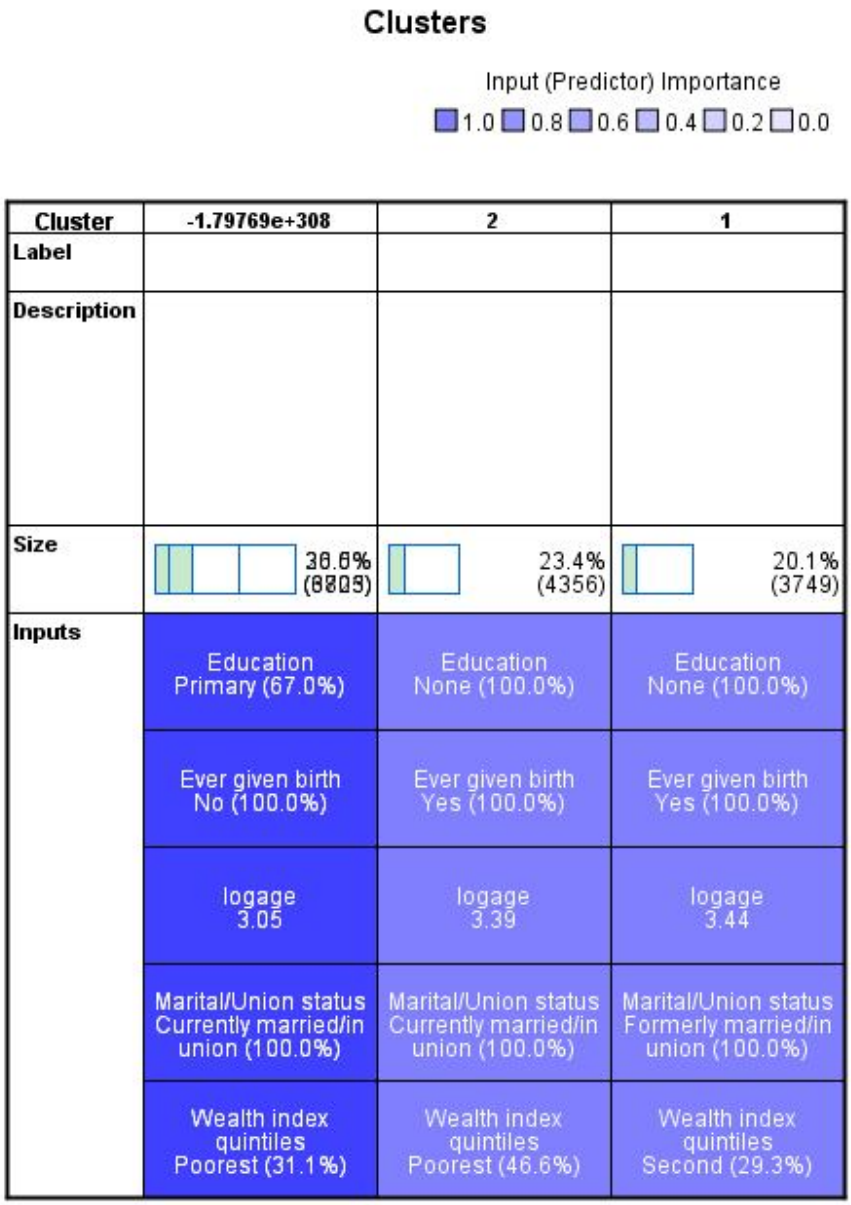

Fig. 4.5 Clusters

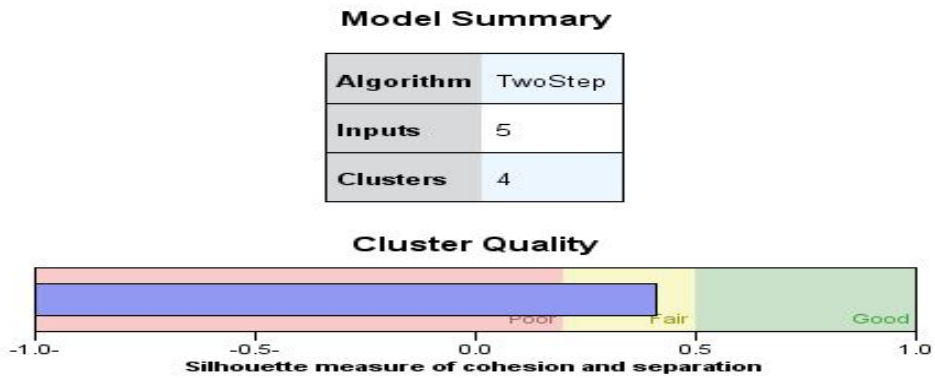

Fig. 4.6 Model Summary Imputation

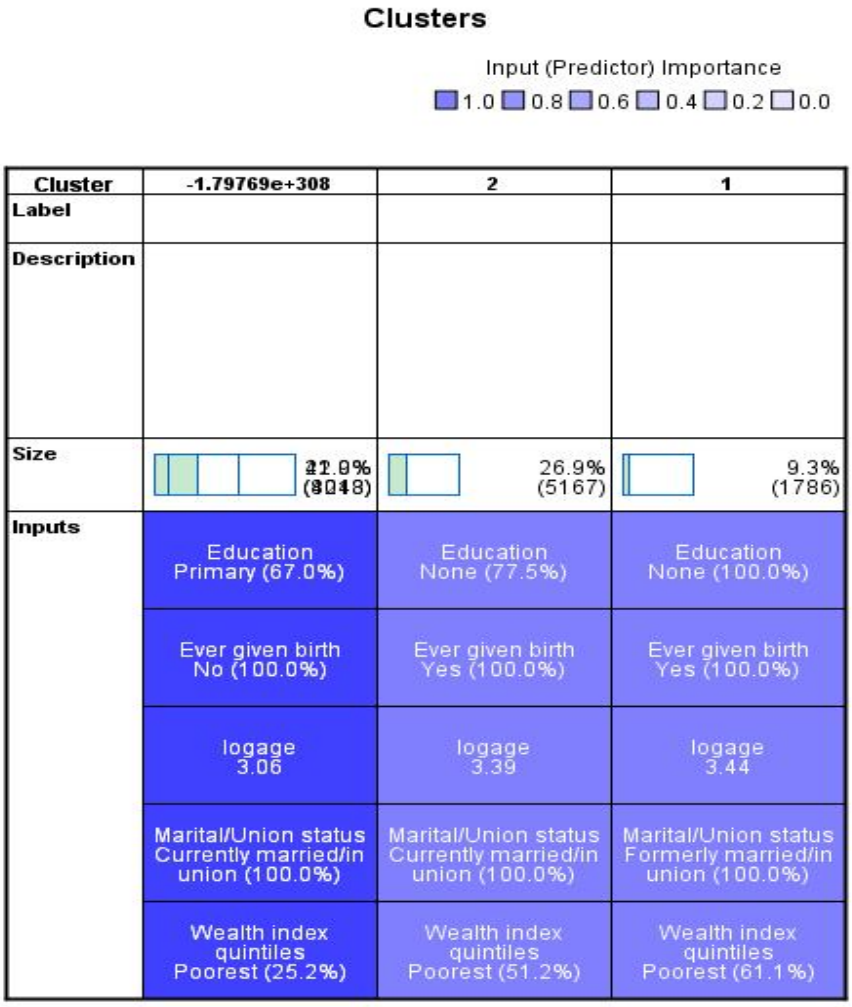

Fig. 4.7 Clusters

Imputation Number = 4

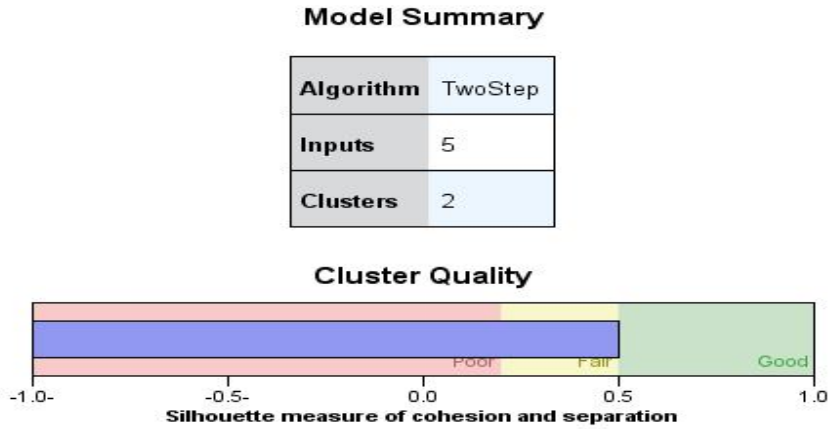

Fig. 4.8 Model Summary Imputation

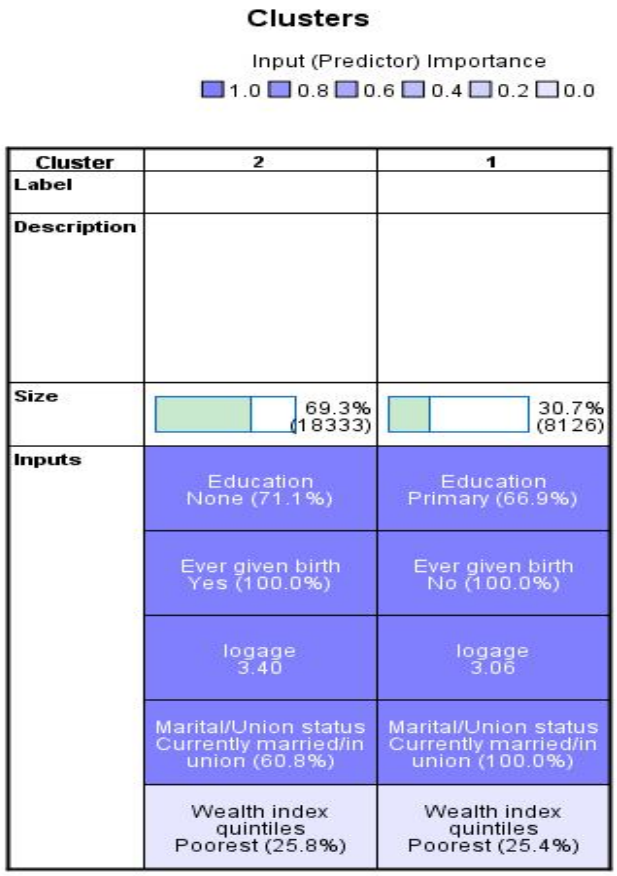

Fig. 4.9 Cluster

Imputation Number = 5

Model Summary

Algorithm	TwoStep
Inputs	5
Clusters	4

Cluster Quality

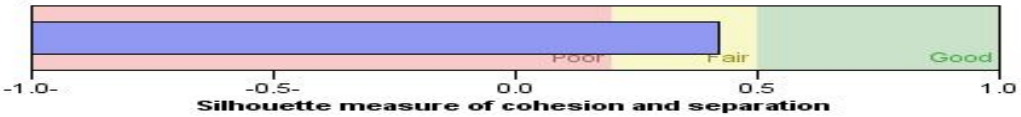

Fig. 4.10 Model Summary Imputation

Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	-1.79769e+308	2
Label			
Description			
Size	27.5% (6361)	39.7% (9184)	17.1% (3964)
Inputs	Education None (61.8%)	Education None (100.0%)	Education None (100.0%)
	Ever given birth Yes (100.0%)	Ever given birth Yes (100.0%)	Ever given birth Yes (100.0%)
	logage 3.42	logage 3.39	logage 3.39
	Marital/Union status Formerly married/in union (100.0%)	Marital/Union status Currently married/in union (100.0%)	Marital/Union status Currently married/in union (100.0%)
	Wealth index quintiles Second (28.5%)	Wealth index quintiles Second (44.7%)	Wealth index quintiles Poorest (53.2%)

Fig. 4.11 Cluster

4.2 Knowledge of means of HIV/AIDS of women

Table 4.21 Knowledge of HIV/AIDS Percentage of woman Year of birth **(1951-1991)**						
		Ever heard of HIV or AIDS				
		Yes	No	DK	Missing	Total
Year of birth of woman	Count	5303224	2052597	227	48775	7404823
	% of Total	71.6%	27.7%	.0%	.7%	100.0%
		Can AIDS be avoided?				
		Yes	No	DK	Missing	Total
	Count	4497187	171670	620316	63057	5352230
	% of Total	84.0%	3.2%	11.6%	1.2%	100.0%
		Healthy-looking person to have AIDS				
		Yes	No	DK	Missing	Total
	Count	2913588	1353238	988175	97223	5352224
	% of Total	54.4%	25.3%	18.5%	1.8%	100.0%
	Count	AIDS from mother to child during pregnancy				
		Yes	No	DK	Missing	Total
		3530084	653207	1065719	103221	5352231
	% of Total	66.0%	12.2%	19.9%	1.9%	100.0%
		AIDS from mother to child at delivery				
		Yes	No	DK	Missing	Total
	Count	3067909	926956	1220824	136540	5352229
	% of Total	57.3%	17.3%	22.8%	2.6%	100.0%
		AIDS from mother to child through breast milk				
		Yes	No	DK	Missing	Total
	Count	2746182	1189134	1306943	109964	5352223
	% of Total	51.3%	22.2%	24.4%	2.1%	100.0%

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.21 present the percentage knowledge of HIV/AIDS women who Ever heard of HIV or AIDS is 71.6%, only about one-half of women (51.3 per cent) knew that AIDS transmitted from mother to child through breast milk. 84.0% of women knew Can AIDS be avoided ,about 54.4% believed that a Healthy-looking person to have AIDS , 66.0% of women knew that AIDS from mother to child during pregnancy, and only 57.3 % of women knew that AIDS from mother to child at delivery .

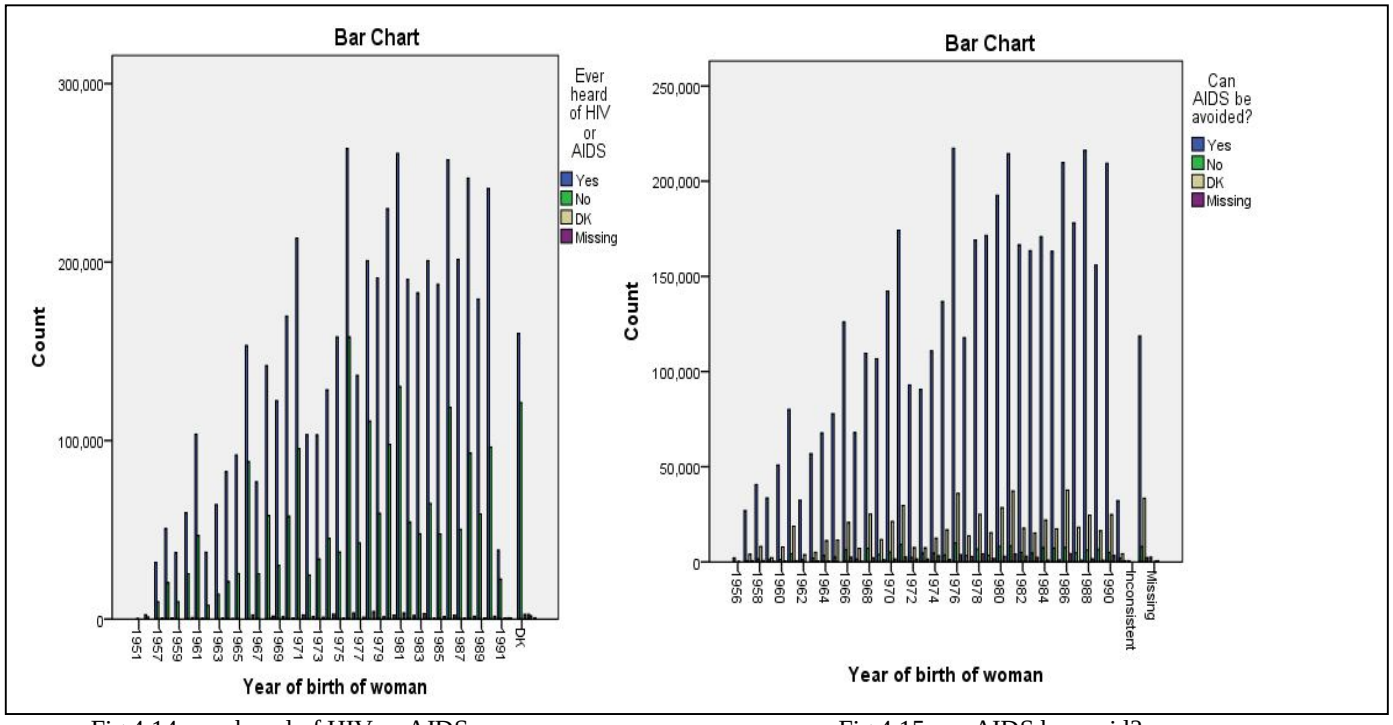

Fig.4.14 ever heard of HIV or AIDS

Fig.4.15 can AIDS be avoid?

The chart in Fig.4.14, Fig.4.15, Fig.4.16, and Fig.4.17 *shows the percentage of HIV/AID women year.*

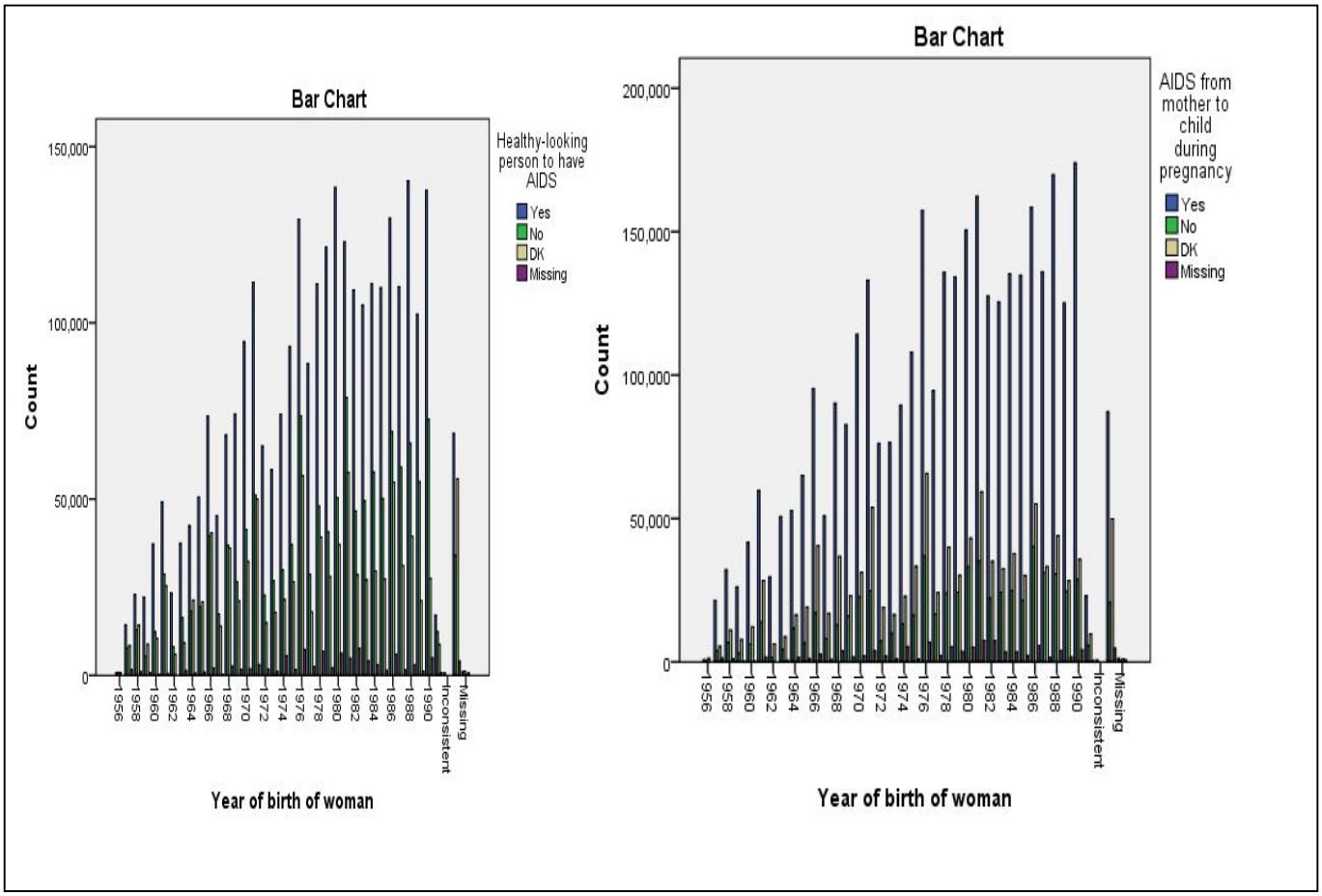

Fig.4.16 ever heard of HIV or AIDS Fig.4.17 can AIDS be avoid?

Table 4.22 displays the percentage of cases with missing values and provides a good measure for comparing the extent of missing data among variables.
Case processing summary of missing 14.5% for year of birth women and 38.2% all other cases of knowledge Hiv/AIDS of women, all cases indicate Number of valid cases is different from the total count in the crosstabulation table because the cell counts have been rounded except year of birth women.

Table 4.22. Case Processing Summary

Year of birth of woman	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Ever heard of HIV or AIDS	7404823	85.5%	1253566.825	14.5%	8658389.825	100.0%
Can AIDS be avoided?	5352230[a]	61.8%	3306159.825	38.2%	8658389.825	100.0%
Healthy-looking person to have AIDS	5352224[a]	61.8%	3306165.825	38.2%	8658389.825	100.0%
AIDS from mother to child during pregnancy	5352231[a]	61.8%	3306158.825	38.2%	8658389.825	100.0%
AIDS from mother to child at delivery	5352229[a]	61.8%	3306160.825	38.2%	8658389.825	100.0%
AIDS from mother to child through breastmilk	5352223[a]	61.8%	3306166.825	38.2%	8658389.825	100.0%

a. Number of valid cases is different from the total count in the crosstabulation table because the cell counts have been rounded.

Source : reasercher analysis of SHHS 2006 data by SPSS

4.2.1 Multiple Imputations

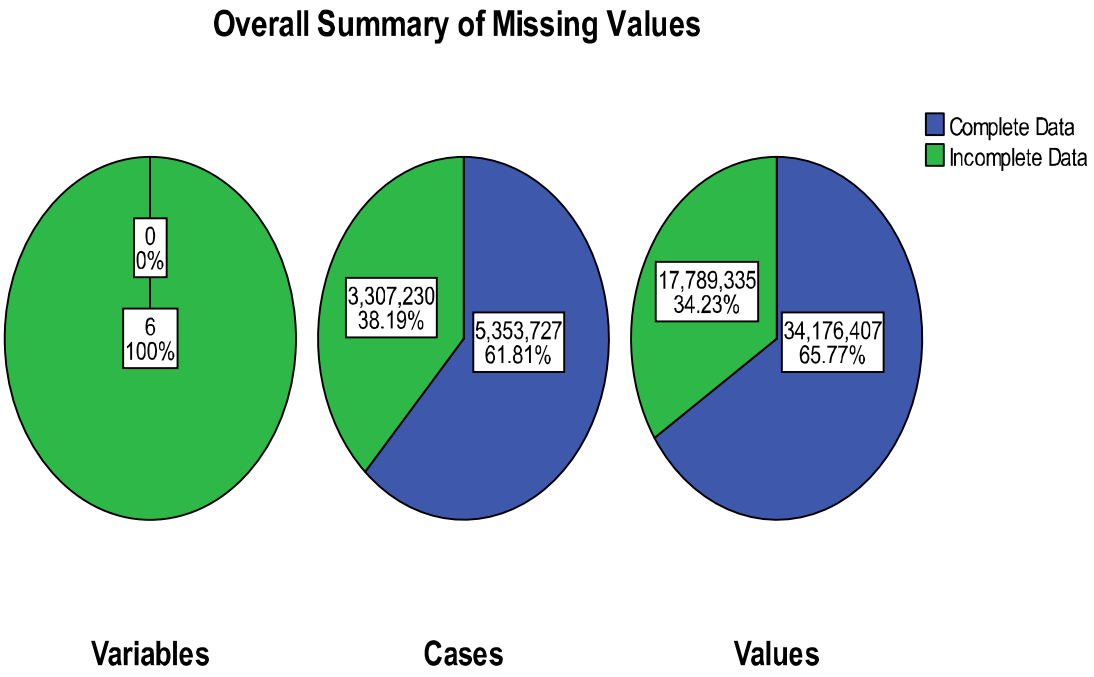

Fig. 4.18 Summary missing values

Fig.4.18 shows that:

- The *Variables* chart shows that each of the six analysis variables has at least one missing value on a case.
- The *Cases* chart shows that 3,307,230 of the 10,000,000 cases have at least one missing value on a variable.
- The *Values* chart shows that 17,789,335 of the 50,000,000 values (cases × variables) are missing.
- There are 5353727 (61.81 %) complete cases and 65.77% complete values.

Table 4.23 Variable Summary					
	Missing		Valid N	Mean	Std. Deviation
	N	Percent			
AIDS from mother to child through breastmilk	3307230	38.2%	5353727		
AIDS from mother to child at delivery	3307230	38.2%	5353727		
AIDS from mother to child during pregnancy	3307230	38.2%	5353727		
Healthy-looking person to have AIDS	3307230	38.2%	5353727		
Can AIDS be avoided?	3307230	38.2%	5353727		
Year of birth of woman	1253185	14.5%	7407772	2291.58	1555.874

Source : reasercher analysis of SHHS 2006 data by SPSS

The variable summary is displayed for variables with at least 10% missing values, and shows the number and percent of missing values for each variable in the table. It also displays the mean and standard deviation for the valid values of scale variables, and the number of valid values for all variables. AIDS from mother to child through breastmilk , AIDS from mother to child at delivery , AIDS from mother to child during pregnancy, Healthy-looking person to have AIDS, and Can AIDS be avoided?, have the most missing values, in that order.

The descriptive statistics in Table 4.23, 14.5% for (Year of birth of woman) shows means and standard deviations in each set of imputed values, 38.2% for all other variables.

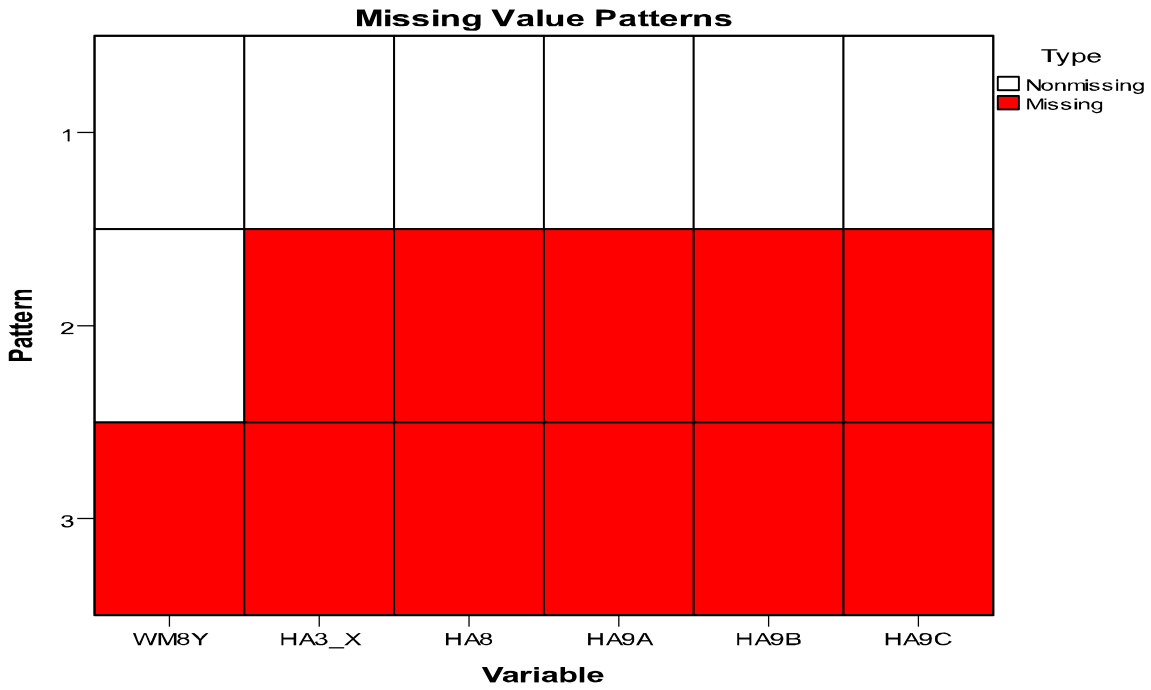

Fig. *4.19 missing value patterns*

The patterns chart Fig.4.19 displays missing value patterns for the analysis variables. Each Pattern corresponds to a group of cases with the same pattern of incomplete and Complete data. Pattern 1 represents cases, which have no missing values, while Pattern 2 represents cases that have missing values on *HA9B,HA8C, HA9A,HAB and HA3_x* and Pattern 3 represents cases which have missing values on *all variables*.

This dataset is nonmonotone and there are many values that would need to be imputed in order to achieve monotonicity.

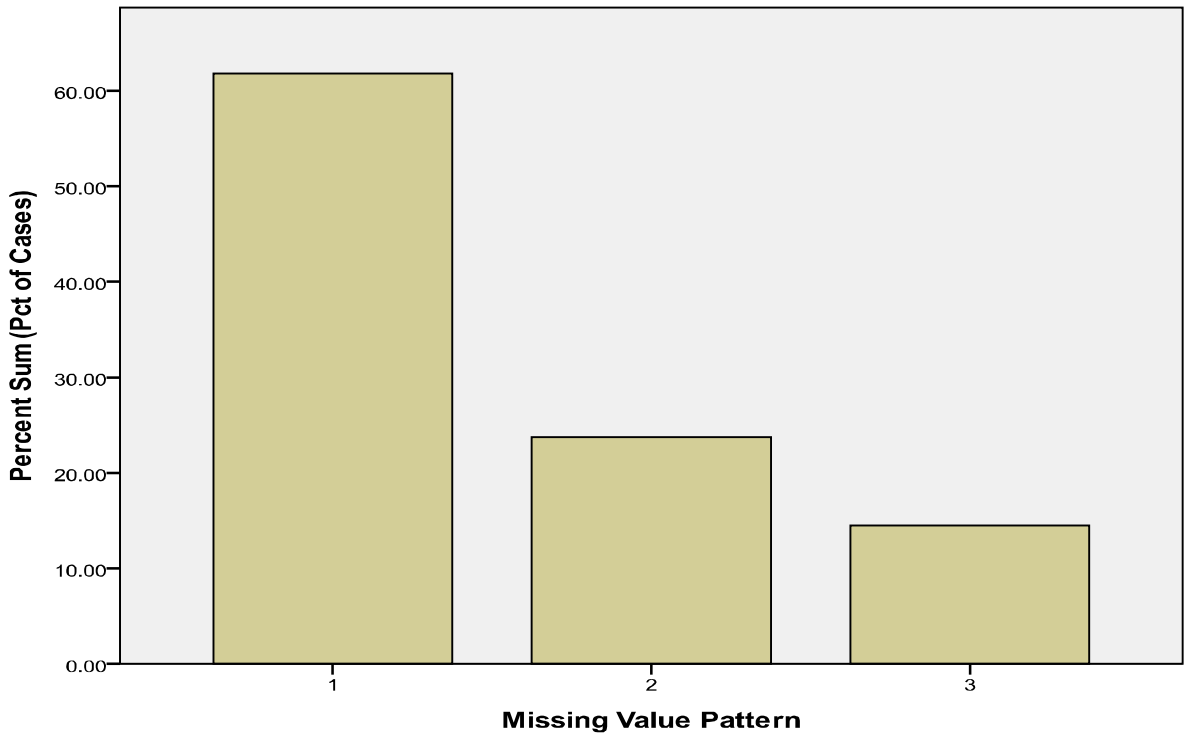

Fig. *4.20 missing value pattern*

The bar chart in Fig.4.20 displays the percentage of cases for each pattern. This shows that over half of the cases in the dataset have Pattern 1, and the missing value patterns chart shows that this is the pattern for cases with no missing values. Pattern 2 represents cases with a missing value on, HA8C, *HA9A, HAB and HA3_x*, Pattern 3, represents cases with a missing value on, HA8C, *HA9A, HAB, HA3_x* and WM8Y.

Table 4.24 Imputation Specifications	
Imputation Method	Fully Conditional Specification
Number of Imputations	5
Model for Scale Variables	Linear Regression
Interactions Included in Models	(none)
Maximum Percentage of Missing Values	100.0%
Maximum Number of Parameters in Imputation Model	100
Replication Weight Variable	hhweight

Source : reasercher analysis of SHHS 2006 data by SPSS

The imputation specifications in table 4.24 is a useful review to confirm that the specifications were correct , Imputation Method is Fully Conditional Specification , Number of Imputations *is 5 and* Model for Scale Variables is Linear Regression , table 5 display that Imputation Results of Fully Conditional Specification Method Iterations is 10 and Dependent Variables Imputed HA3_X,HA8,HA9A,HA9B,HA9C.

Table 4.25 Imputation Results		
Imputation Method		Fully Conditional Specification
Fully Conditional Specification Method Iterations		10
Dependent Variables	Imputed	HA3_X,HA8,HA9A,HA9B,HA9C
	Not Imputed(Too Many Missing Values)	
	Not Imputed(No Missing Values)	WM8Y
Imputation Sequence		WM8Y,HA3_X,HA8,HA9A,HA9B,HA9C

Source : reasercher analysis of SHHS 2006 data by SPSS

Table 4.26 Imputation Models				
	Model		Missing Values	Imputed Values
	Type	Effects		
Can AIDS be avoided?	Logistic Regression	HA8,HA9A,HA9B,HA9C,WM8Y	2054045	10270225
Healthy-looking person to have AIDS	Logistic Regression	HA3_X,HA9A,HA9B,HA9C,WM8Y	2054045	10270225
AIDS from mother to child during pregnancy	Logistic Regression	HA3_X,HA8,HA9B,HA9C,WM8Y	2054045	10270225
AIDS from mother to child at delivery	Logistic Regression	HA3_X,HA8,HA9A,HA9C,WM8Y	2054045	10270225
AIDS from mother to child through breastmilk	Logistic Regression	HA3_X,HA8,HA9A,HA9B,WM8Y	2054045	10270225

Source : reasercher analysis of SHHS 2006 data by SPSS

The imputation models in table 4.26 gives details about variable was Imputed. Note in particular that:

- All categorical variables modeled with a logistic regression.
- Each model uses all other variables as main effects.
- The number of missing values for each variable is reported, along with the total number of values imputed for that variable (number missing × number of imputations) for example (2054045×5=10270225).

4.2.2 Descriptive Statistics knowledge HIV/AIDS

Table 4.27 HA3_X (Can AIDS be avoided?)

Data	Imputation	Category	N	Percent
Original Data		1	4498332	84.0
		2	171720	3.2
		8	620571	11.6
		9	63104	1.2
Imputed Values	1	1	1093635	53.2
		2	71756	3.5
		8	432127	21.0
		9	456527	22.2
	2	1	971065	47.3
		2	46581	2.3
		8	411349	20.0
		9	625050	30.4
	3	1	991824	48.3
		2	48031	2.3
		8	428699	20.9
		9	585491	28.5
	4	1	987850	48.1
		2	59633	2.9
		8	412233	20.1
		9	594329	28.9
	5	1	991901	48.3
		2	47845	2.3
		8	422065	20.5
		9	592234	28.8

Source : reasercher analysis of SHHS 2006 data by SPSS

The table 4.27 for **HA3_X (Can AIDS be avoided?)** now has an imputation (5) whose distribution is more in line with the original data, but the majority are still showing a greater proportion of the cases estimated as being avoided than in the

original data. This could be due to random variation, but might require further study of the data to determine whether these values are not missing at random (MAR). We will not pursue this further here.

Table 4.28 HA9A(AIDS from mother to child during pregnancy)				
Data	Imputation	Category	N	Percent
Original Data		1	3530910	66.0
		2	653426	12.2
		8	1066113	19.9
		9	103278	1.9
Imputed Values	1	1	457387	22.3
		2	216996	10.6
		8	757612	36.9
		9	622050	30.3
	2	1	471058	22.9
		2	207205	10.1
		8	732895	35.7
		9	642887	31.3
	3	1	472529	23.0
		2	218044	10.6
		8	761608	37.1
		9	601864	29.3
	4	1	473524	23.1
		2	210061	10.2
		8	761403	37.1
		9	609057	29.7
	5	1	469251	22.8
		2	208048	10.1
		8	755345	36.8
		9	621401	30.3

Source : reasercher analysis of SHHS 2006 data by SPSS

The table 4.28 for HA9A(AIDS from mother to child during pregnancy) has an interesting result in that, for the imputed values, a greater proportion of the cases are estimated as being AIDS during pregnancy than in the original data. This could be due to random variation; alternatively, the chance of being missing may be related to value of this variable.

Table 4.29 HA9B(AIDS from mother to child at delivery)

Data	Imputation	Category	N	Percent
Original Data		1	3068673	57.3
		2	927199	17.3
		8	1221241	22.8
		9	136614	2.6
Imputed Values	1	1	409625	19.9
		2	247480	12.0
		8	788967	38.4
		9	607973	29.6
	2	1	414009	20.2
		2	848814	41.3
		8	751484	36.6
		9	39738	1.9
	3	1	411984	20.1
		2	833175	40.6
		8	776338	37.8
		9	32548	1.6
	4	1	408847	19.9
		2	829529	40.4
		8	781710	38.1
		9	33959	1.7
	5	1	872933	42.5
		2	245323	11.9
		8	783093	38.1
		9	152696	7.4

Source : reasercher analysis of SHHS 2006 data by SPSS

4.2.3 Checking FCS Convergence

When using the conditional specification method, it is a good idea to check plots of the means and standard deviations by iteration and imputation for each scale dependent variable for which values are imputed in order to help assess model convergence.

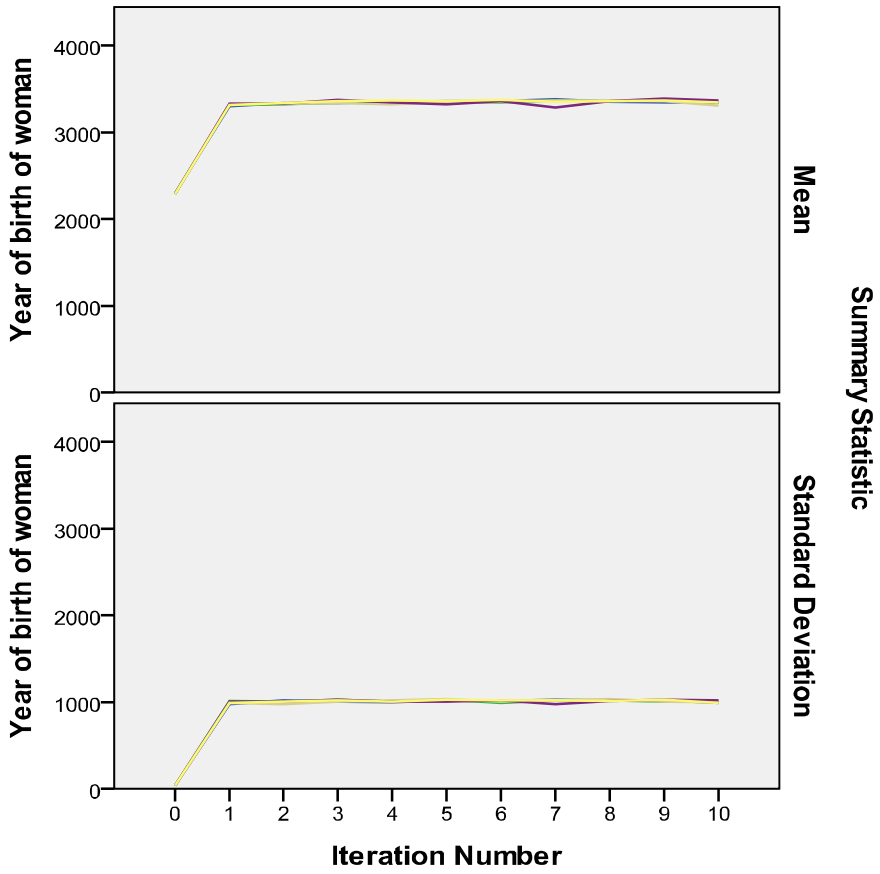

Fig. 4.21 FCS Iteration number

You have created a pair of multiple line charts Fig. 4.21, showing the mean and standard deviation of the imputed values of *year of birth women* at each iteration of the FCS imputation method for each of the five requested imputations. The

purpose of this plot is to look for patterns in the lines. There should not be any, and these look suitably “random”. You can create similar plots for the other scale variables, and note that those plots also show no discernable patterns.

4.2.4 Two-step Cluster Analysis

4.2.4.1 Model Summary and Cluster Quality

- The model summary in Fig.4.22 and Fig.4.37 indicates that tow clusters were found based on the six input features (fields) selected.
- The model summary in Fig.4.25, Fig.4.31 and Fig.4.34 indicates that four clusters were found based on the six input features (fields) selected.
- The model summary in Fig.4.28 indicates that three clusters were found based on the six input features (fields) selected.
- The model summary of cluster quality chart in Fig.4.22, Fig.4.25, Fig.4.28, Fig.4.31, Fig.4.34, Fig.4.36 indicates that the overall model quality is "Fair".

4.2.4.2 Cluster Distribution

The Cluster Sizes view in Fig.4.23 shows the frequency of each cluster. The pie chart assigned to the cluster, 41.3% of the records were assigned to the first cluster and 58.7% to the second. while Fig.4.39 , Fig.4.32 and Fig.4.35 shows 4 cluster size , 11.5% size of smallest cluster and 43.6% size of largest cluster, Fig.28 indicate that 3 cluster size 11.8% size of smallest cluster and 56.9% size of largest cluster, only 2 cluster size in Fig.38 indicate 34.0% for first cluster and 66.0% for the second cluster.

Fig.4.24 clusters are sorted from smallest to largest by cluster size, so they are currently ordered 1, 2.

Fig.4.27 clusters are sorted from smallest to largest by cluster size, so they are currently ordered 2, 1, 3.

Fig.4.30 clusters are sorted from smallest to largest by cluster size, so they are currently ordered 3, 1, 2.

Fig.4.33 clusters are sorted from smallest to largest by cluster size, so they are currently ordered 1, 2, 3.

The cluster means suggest that the clusters are well separated.

The cluster means (for continuous fields) and modes (for categorical fields) are useful, but only give information about the cluster centers, In order to get a visualization of the distribution of values for each field by cluster.

4.2.5 Imputation Number = Original data

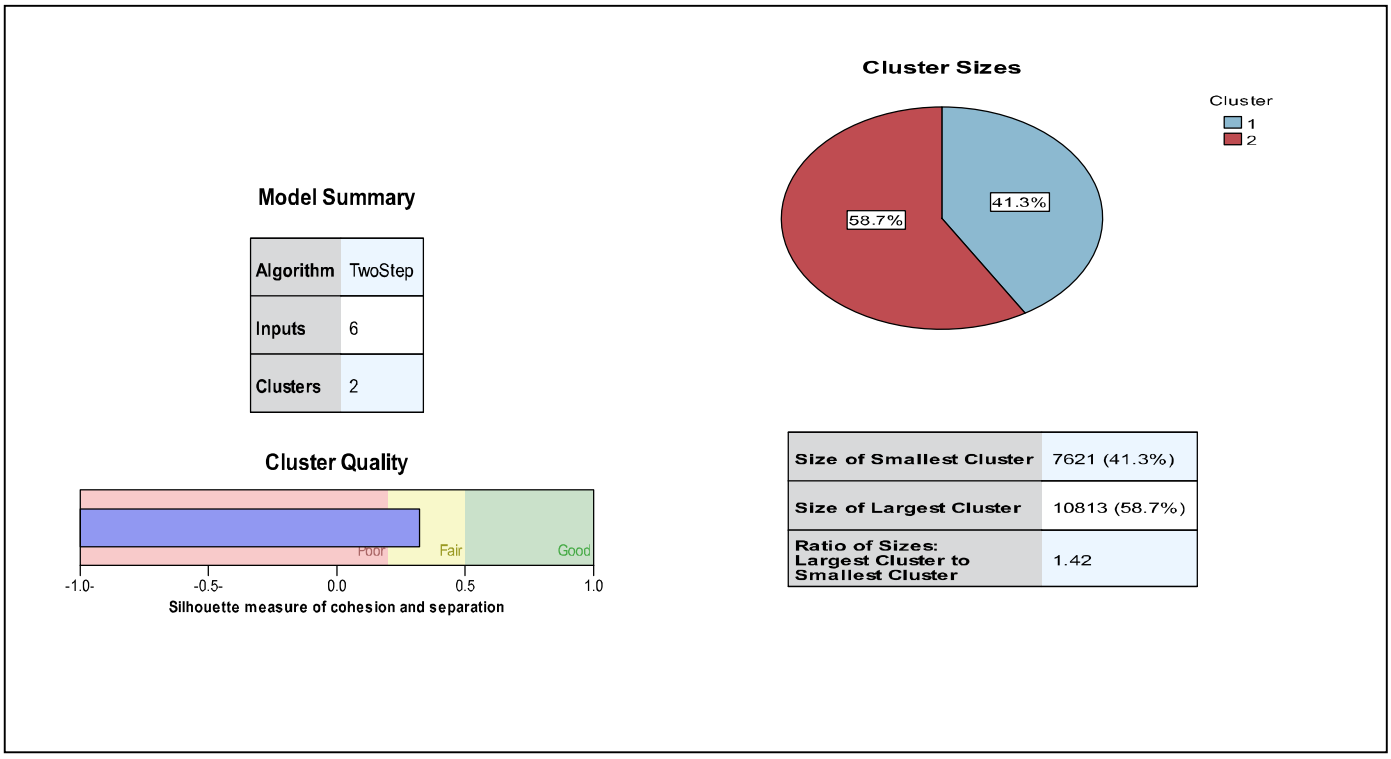

Fig. 4.22 imputation original Model Summary **Fig. 4.23 imputation original cluster size**

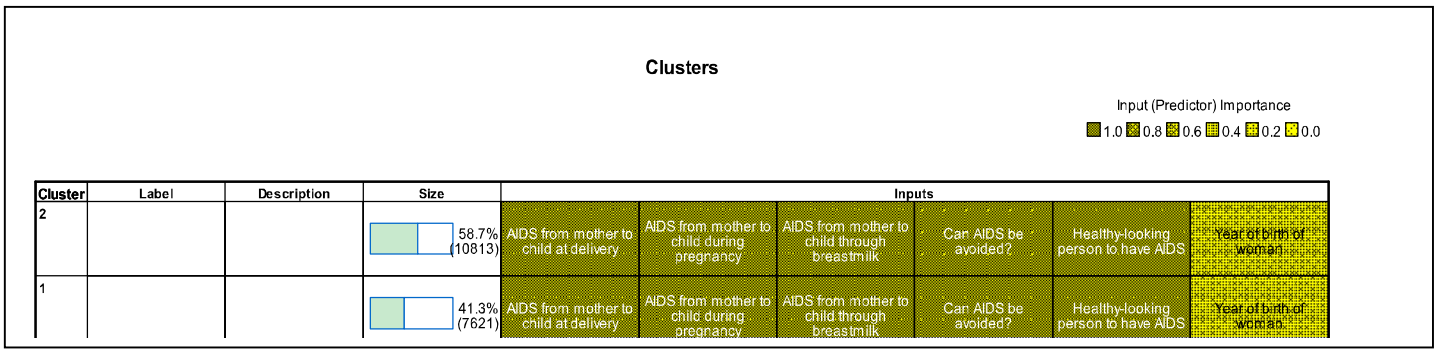

Fig. 4.24 imputation original data Custers

4.2.5.1 Imputation Number = 1

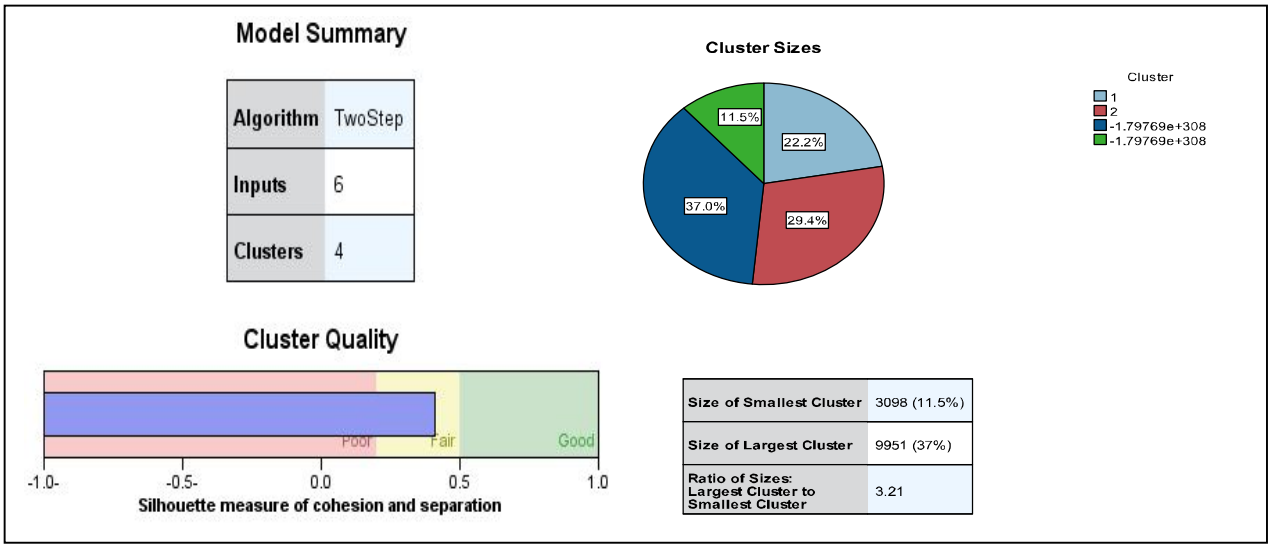

Fig.4. 25 imputation number 1 Model Summary **Fig. 4.26 imputation number 1 cluster size**

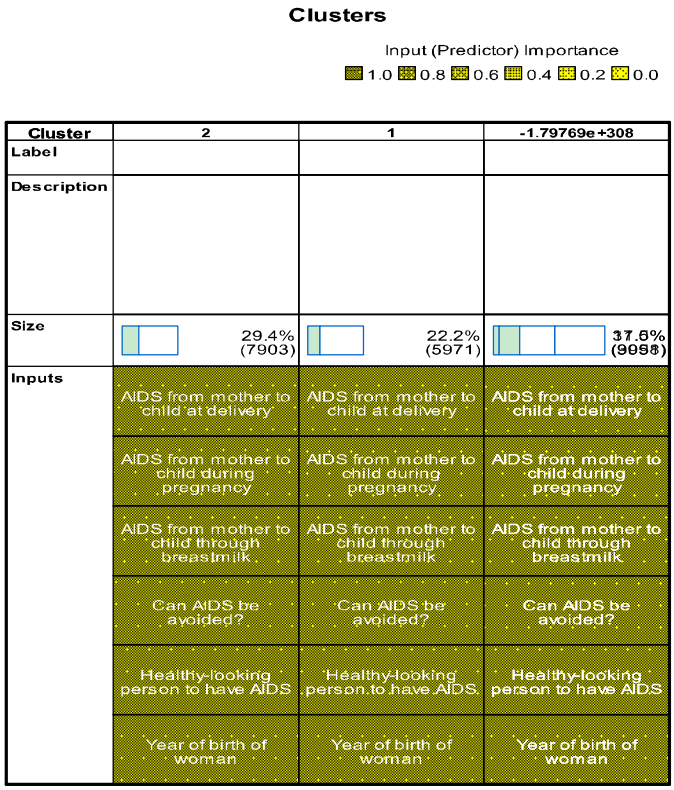

Fig. 4.27 imputation number 1 clusters

4.2.5.2 Imputation Number = 2

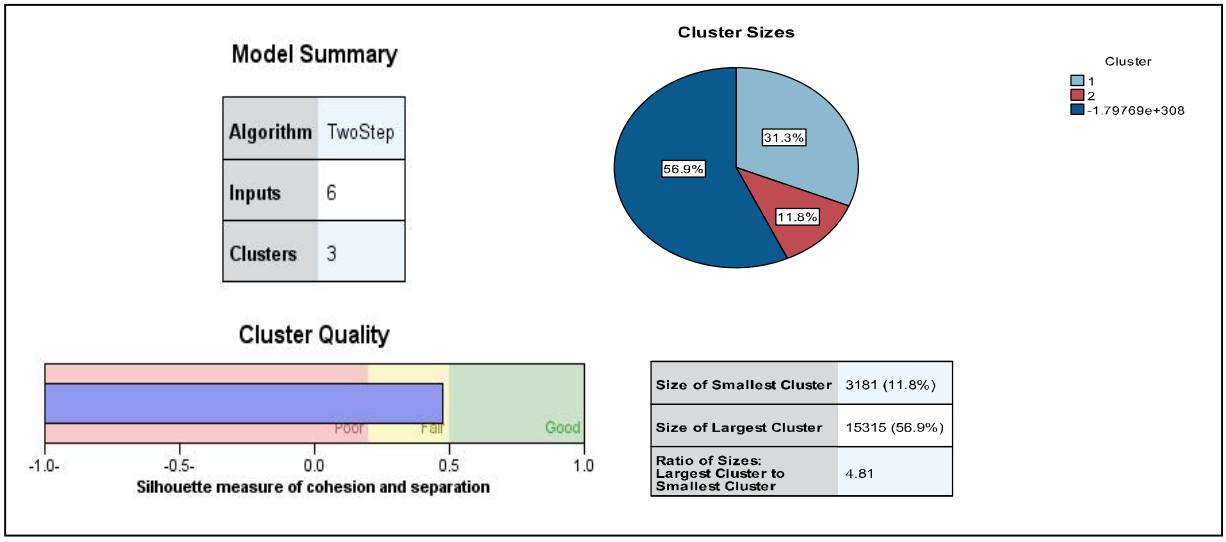

Fig. 4.28 imputation number 2 Model Summary

Fig. 4.29 imputation number 2 cluster size

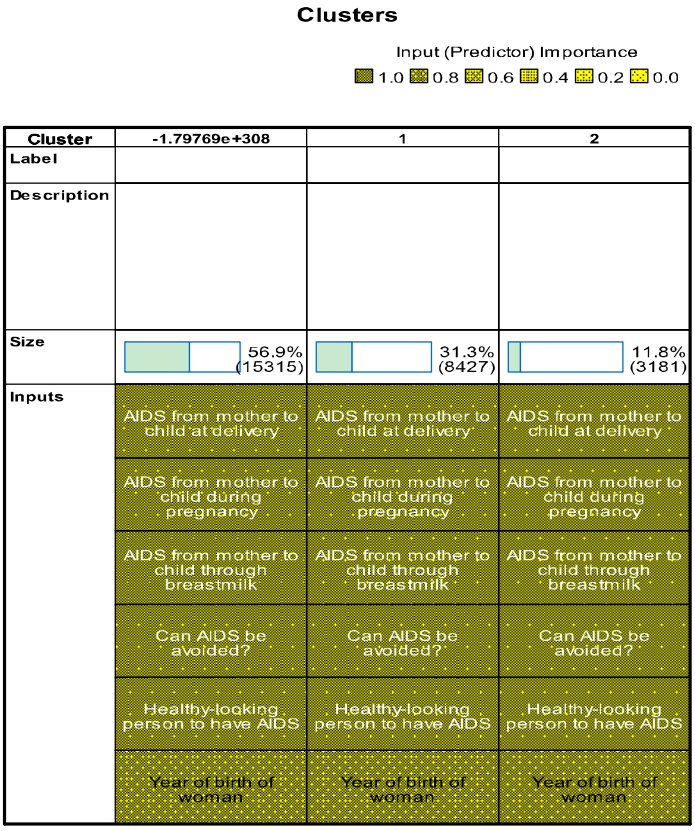

Fig. 4.30 imputation number 2 clusters

4.2.5.3 Imputation Number = 3

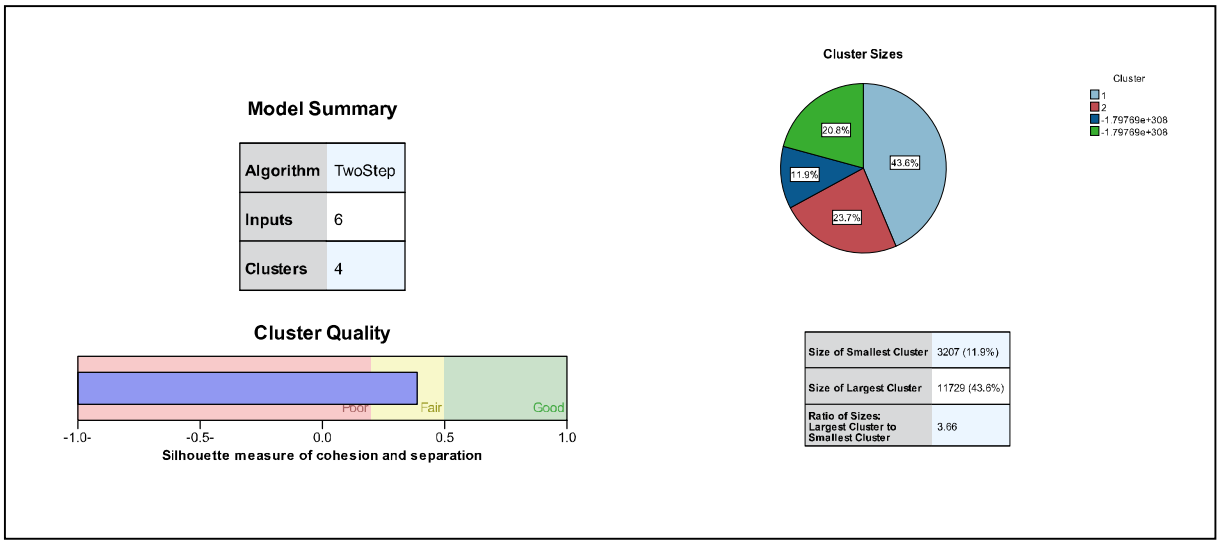

Fig. 4.31 imputation number 3 Model Summary

Fig. 4.32 imputation number 3 cluster size

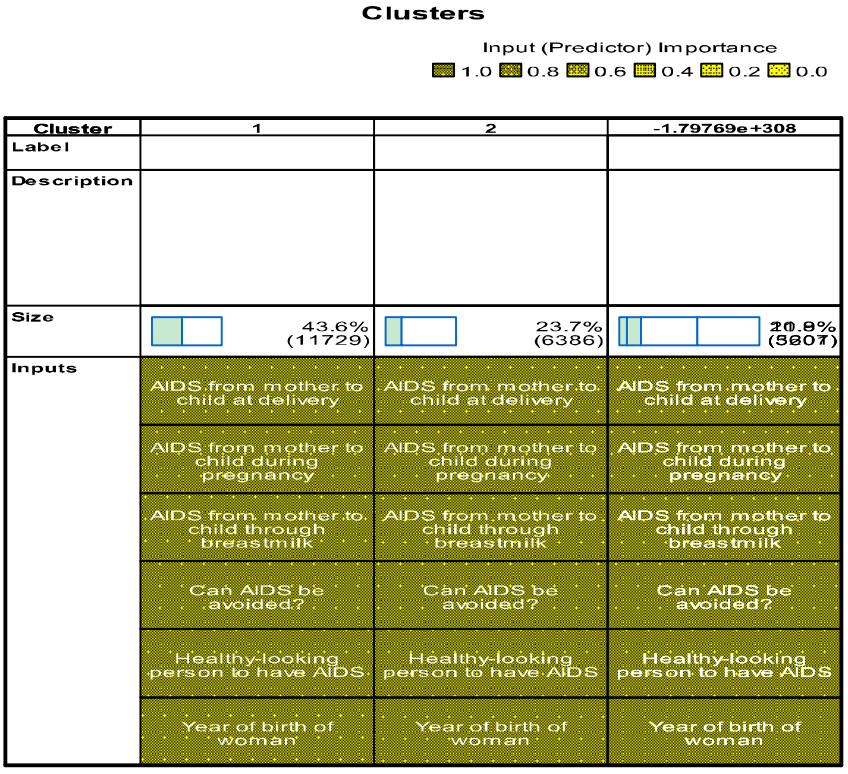

Fig. 4.33 imputation number 3 clusters

4.2.5.4 Imputation Number = 4

Fig. 4.34 imputation number 4 Model Summary

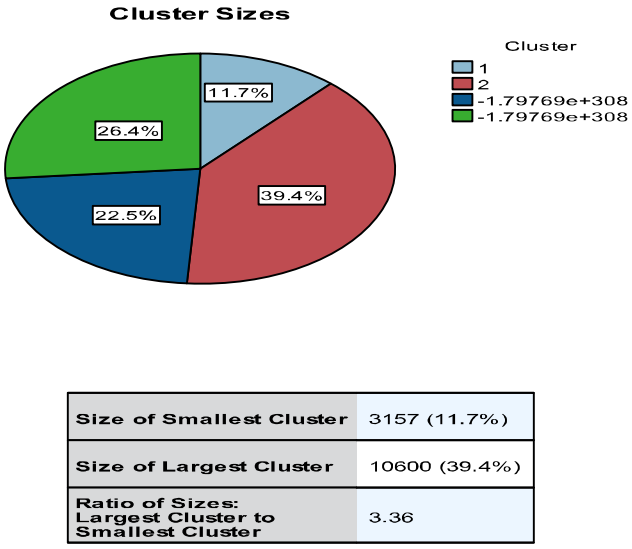

Fig. 4.35 imputation number 4 cluster size

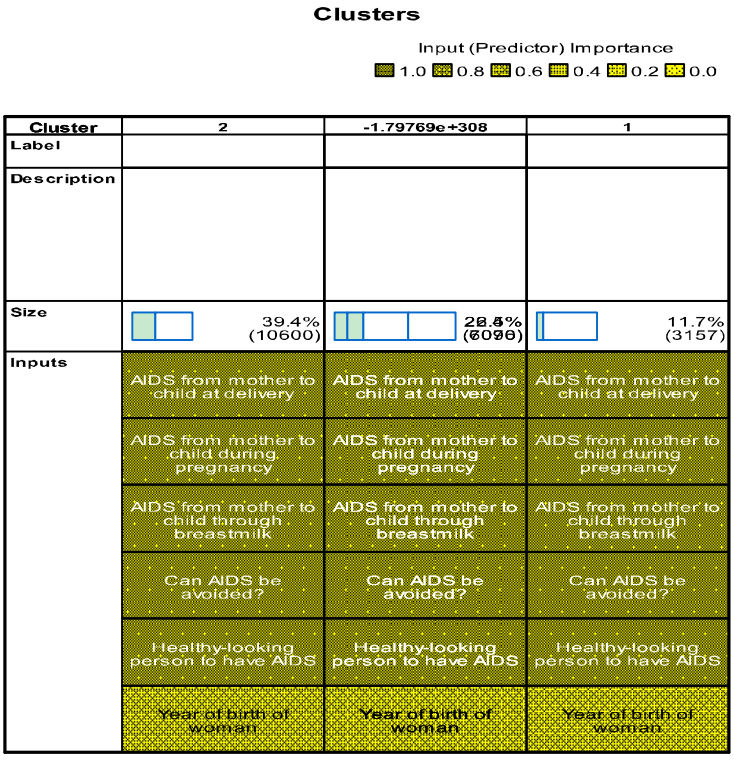

Fig. 4.36 imputation number 4 clusters

4.2.5.5 Imputation Number = 5

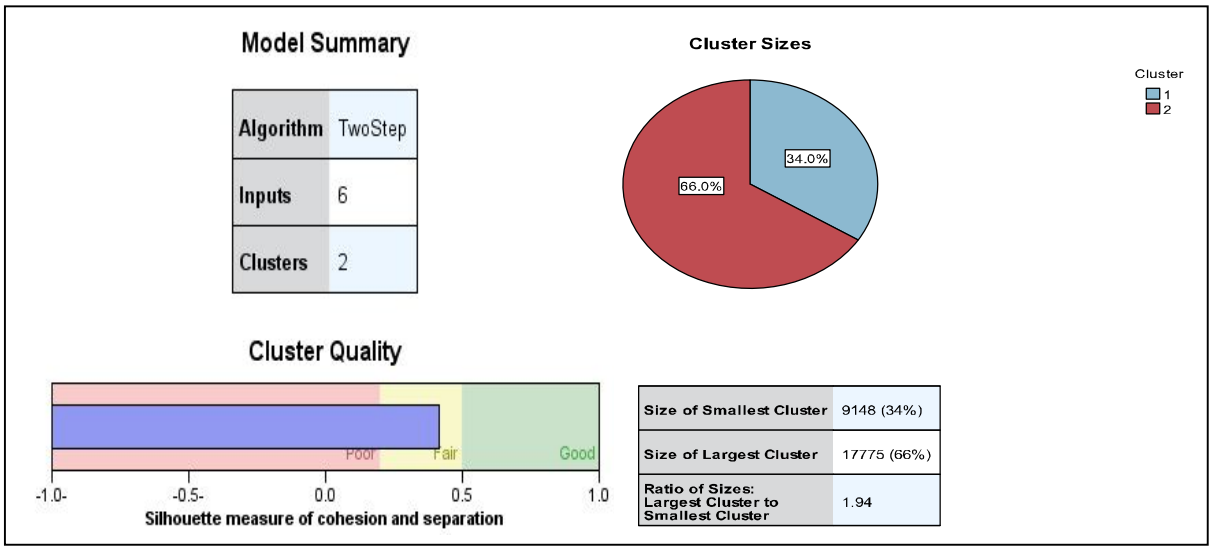

Fig. 4.37 imputation number 5 Model Summary

Fig. 4.38 imputation number 5 cluster size

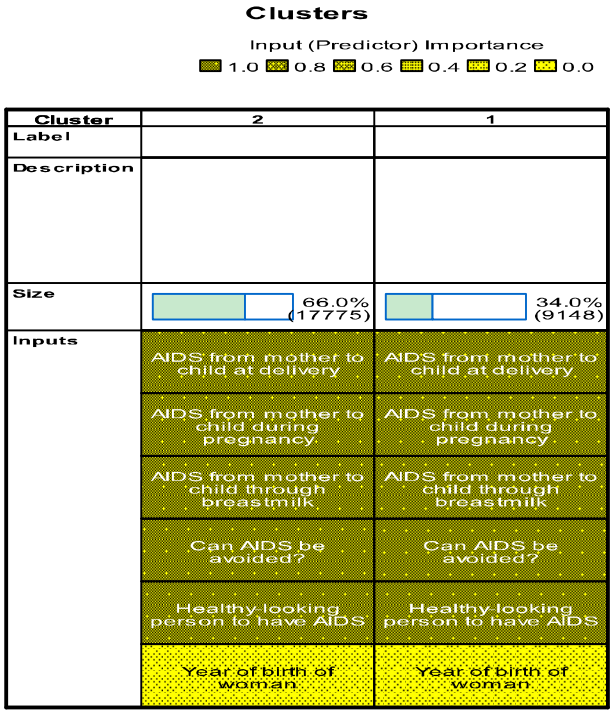

Fig. 4.39 Fig. 15 imputation number 5 clusters