DANA4830 – Fall 2021

Submitted by:
Mohamed Ghayaas Anjum and Umme Salma

Data Cleaning:

The original dataset that is taken for data cleaning consists of 165 observations (rows) and 164 variables (columns). The data cleaning has been performed in Python. Initially, we check the number of missing values in each variable of dataset by using isnull() function. Then, we check the data for cardinality. It refers to detection of the number of categories present inside each categorical variable. There are 24 categorical variables and 139 numerical variables present in this dataset excluding ID of patient.

Cleaning all the categorical Variables:

Sl No.	Name of Variable	Inaccuracy Identified	Replaced by
1	Diabetes Problem	3	Nan
2	Historical cholecystitis problem	3,5	Nan
3	Clinical symptoms of defecation	T0, T30	t0 and t30
4	clinical symptoms of diarrhoea	T0, T30, to, tn6 and TRV	t0, t30, t0, t6 and nan
5	Clinical symptoms of Abdominal distension	T0, T30, to, t, T6	t0, t30, t0, t0 and t6
6	Blood pressure	Both systolic and diastolic are given as a single entry	Created two different variables for systolic and

			diastolic blood pressure.
7	subclinical examination - (Abdominal fluid) computer tomography	All the observations other than have and no	categorized as not clear.
8	subclinical examination - balthazar score	Case mismatch present	Lowercase letters
9	subclinical examination - CTSI score (with computer tomography)	23 and e	Replaced by nan
10	subclinical examination - bilirubin total	Value is separated by /	Created separate variables called direct and indirect bilirubin. “ ” and none to nan
11	subclinical examination - bilirubin total for t6	Bilirubin represented as ratio	Created separate variables called direct and indirect bilirubin. “ ” and none to nan and “4” by 4
12	subclinical examination - bilirubin total for t30	Bilirubin represented as ratio	Created separate variables called direct and indirect bilirubin. “ ” and none to nan and “4” by 4

13	subclinical examination - bilirubin total for t72	Bilirubin represented as ratio	Created separate variables called direct and indirect bilirubin. “ ” and none to nan
14	AST, ALT (liver funtion)	Value represented as ratio	“ ” and none to nan.
15	subclinical examination AST ALT at 6 hours t6	Value represented as ratio	“ ” and none to nan.
16	subclinical examination AST ALT at 30 hours t30	Value represented as ratio	“ ” and none to nan.
17	subclinical examination - calci total	serum calcium and ionized calcium are separated with a “/”	“ ” and none to nan
18	Result - dead or alive	No redundancy present	Replaced Alive by 1

Then, we drop all the redundant information from the categorical variables present and include the new features. Later, we import the new dataset into SAS for further analysis.

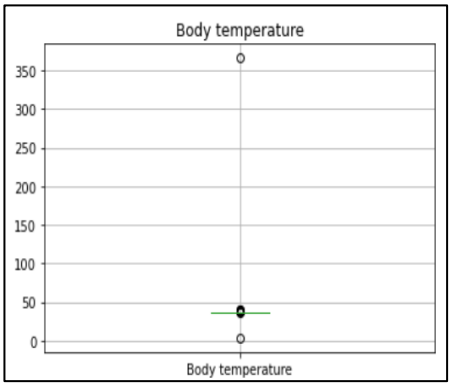
Outlier Detection and Cleaning:

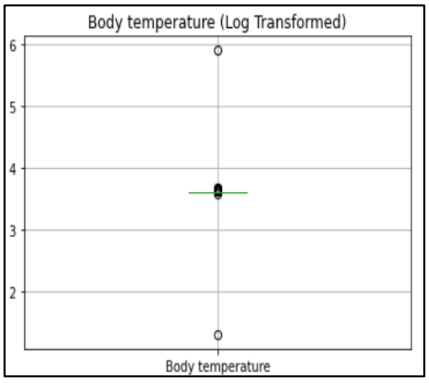
With the help of visualization using boxplot, we determine the outliers for continuous variables.

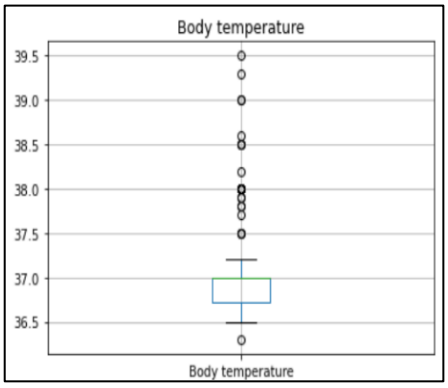
The outliers are observed with the skewed nature of the boxplots. All the continuous variables have also been transformed with their logs in this step to observe their boxplots. These two

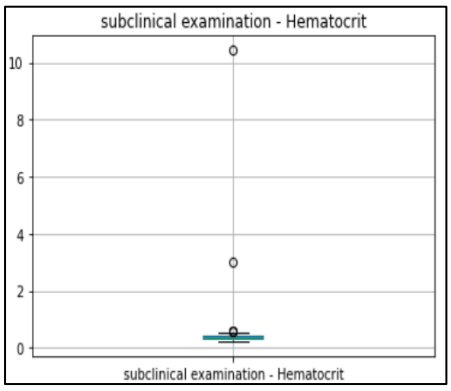
figures indicate which continuous variables have out of range values. Such values are filtered in python and appropriate correction method is applied. For example, The inaccuracies found due to typo error such as 243 value in a variable could possibly represent 24.3 because all the other values in that variable are in the range 20 – 30. Some outliers are left untreated in variables like age, body temperature and other such variables. Though, these values do not fall under the normal distribution of the data but are meaningful values. The details of the cleaning action for outliers in continuous variables is tabulated in the table below.

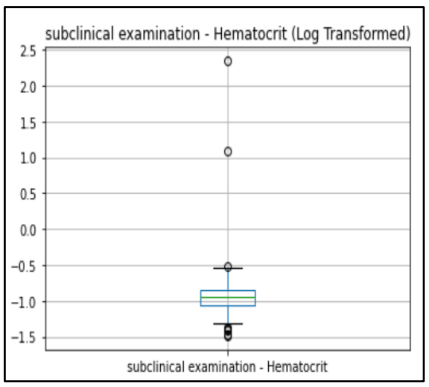
The figures below represents example boxplots for each variable. The first boxplot represents outliers before data cleaning, second boxplot is for outliers after log transformation and third represents outliers after data cleaning.

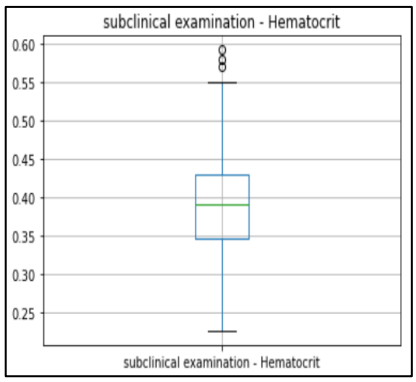


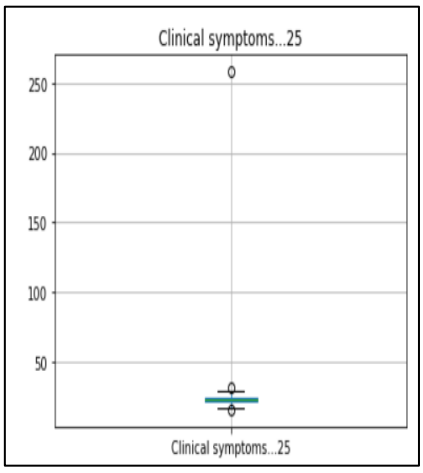


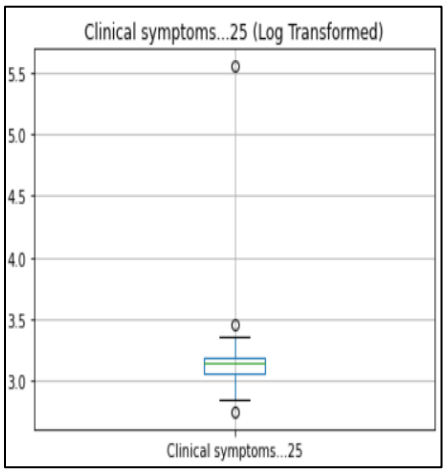


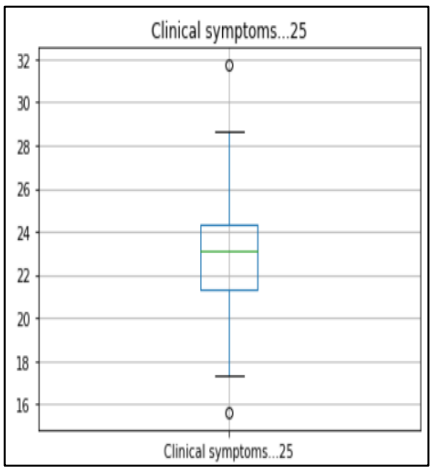












Data Cleaning for continuous variables with outliers

Sl No.	Variable Name	Cleaning action
1	Clinical symptoms...25	258 is replaced by 25.8
2	Body Temperature	366 and 3.7 are replaced by 36.6 and 37
3	subclinical examination - Hematocrit	replace 10.43 to 0.43, and 3 to 0.3
4	subclinical examination -...61	468.0 is replaced by 4.68
5	prothrombin	Indicates outliers for values more than 500. Not imputed as some patients might have a clotting problem.
6	subclinical examination -...67	6638.0 is replaced by 66.38
7	subclinical examination -...68	0.98 is replaced by 98.0

8	APTT	3212.0 is replaced by nan
9	subclinical examination -...70	1102.0 is replaced by 1.102
10	subclinical examination -...72	27.5 is replaced by nan to impute later
11	subclinical examination - Fibrinogen	45.0 is replaced by 4.5
12	subclinical examination - ure	Normal range is between 7 to 20. 193.0,66.0 and 50.0 are replaced by 19.3,6.6 and 5.0 respectively
13	subclinical examination -...78	64.0 is replaced by 6.4
14	subclinical examination -...79, subclinical examination -...80, subclinical examination – creatinin, Cholesterol	Did not alter outliers
15	subclinical examination -...86	64.4 is replaced by nan since most of the values are nan
16	subclinical examination - sodium	4.2 is replaced with nan

17	subclinical examination -...119	3.7 is replaced with nan
18	subclinical examination -...122	137.0 is replaced by 3.7 and 33 by 3.3
19	subclinical examination - pH (in blood air)	741.0 is replaced with 7.41
20	subclinical examination -...128	Two outliers replaced by nan
21	subclinical examination - HCO3-(in blood air)	1708.0 is replaced by 17.08
22	subclinical examination - HCO3-(in blood air)' at different hours	Outliers present due to Decimal error. Hence shifting one decimal point to the left.
23	BE in blood air (in blood air)' at different hours	Outliers present due to Decimal error. Hence shifting one decimal point to the left.
24	Bilirubin_direct_t0	Outliers present due to Decimal error. Hence shifting one decimal point to the left.
25	Bilirubin_direct_t72	Outliers present due to Decimal error. Hence shifting one decimal point to the left.
26	AST (liver funtion)	Outliers present due to Decimal error. Hence shifting one decimal point to the left.

27	AST (liver funtion)_t6	Outliers present due to Decimal error. Hence shifting one decimal point to the left.
28	AST (liver funtion)_t30	Outliers present due to Decimal error. Hence shifting one decimal point to the left.
29	subclinical examination - calci_ionized	Drop this variable as I contains 149 nulls out of 165.

Data Imputation with MICE, Amelia, and MissForest:

After the data cleaning is completed, a threshold of 70% missing values is selected. All variables having more than 70% missing values are dropped from the dataset. Also, all observations (rows) having more than 70% missing values are dropped.

The cleaned dataset is exported from python to R, has less than 70 percent of missing values in each variable and less than 70 percent of missing values in each observation. The resulting clean dataset has 162 variables and 132 Observations. The imputation is performed on this clean dataset using MICE, Amelia, and MissForest in R.

```
Shape of data after removing 70% empty rows and columns: (162, 132)
```

	ID	Age	Gender	Main reason to admit hospital	Duration of staying in hospitals	Hereditary information	Gallbladder problem	Drinking problem	A breakdown of drinking problem...13	Diabetes problem	Historical cholecystitis problem	Vomitting	Clinical symptoms of Abdominal distension	Clinical symptoms...24	sym
0	1	30	Female	stomachache	6	no	no	no	NaN	no	no	NaN	t0	NaN	
1	2	43	Male	stomachache	8	no	no	Have	8.0	no	Have	NaN	t0	21.0	
2	3	50	Male	stomachache	7	NaN	no	Have	20.0	no	Have	NaN	t0	25.0	
3	4	47	Male	stomachache	4	NaN	no	Have	10.0	no	no	NaN	t0	13.0	
4	5	43	Female	stomachache	5	no	no	no	NaN	no	no	0.0	no	NaN	

Most of the out-of-range values are removed or appropriately treated for all the variables as described in the table in the previous question. However, some variables having values outside of the normal distribution, but meaningful, like age of 66, 70 and levels of triglycerides, are

retained in the dataset for further imputation. In the cleaned dataset, the variables retaining some outliers are listed below:

Age
Duration of stay in hospitals
Breakdown of drinking problem
Clinical symptoms….24
Clinical symptoms….25
Clinical symptoms….26
Body Temperature
Saturation of peripheral Oxygen
Apache 2 score at the point of admitting hospitals
CTSI score at the point of admitting hospitals
Subclinical Examination- white blood cell at t0 and t6
Subclinical Examination..47,48,49,50,51,52,53,54,55
Subclinical Examination – hematocrit
Subclinical Examination..57,58,59
Red blood cell
Subclinical Examination..61,62,63,64
Prothrombin
Subclinical Examination..66,67,68
APTT
Subclinical Examination..70,71,72
Subclinical Examination – Fibrinogen
Subclinical Examination..74,75,76
Subclinical Examination – Urea
Subclinical Examination..78,79,80
Subclinical Examination – Creatinine
Subclinical Examination..82,83,84
Subclinical Examination – Glucose

Bi variate Imputation:

The three packages' MICE, Amelia, and MissForest, follows the concept of bivariate and multivariate imputations. Unlike univariate imputation method in which the imputations are performed in a variable based on the mean, median or mode of the same variable, the bivariate imputation considers values of multiple variables or sometimes the entire dataset to impute missing values of a particular variable. For imputation of each variable, the algorithm considers the observations having non null values as the training dataset to train a model. This model considers the missing values of the variable to be imputed as the test set and imputes the values at the null places based on the training information of the training dataset. Thus, Bivariate imputation method results in meaningful imputations than univariate method.

A) Imputation of missing values using Mice Package

Firstly, all the categorical variables are defined as factor variables in R. In mice, ployreg method which refers to multinomial poly regression is used for imputing categorical variables. For all the continuous variables, predictive mean matching (pmm) method is used for imputation. Here, we have specified three iterations which means maxit = 3. The default method is vector of length 4 containing various methods of imputation for various types of data

a) Numeric data
b) Factor data having two levels
c) Factor data with more than two unordered levels
d) Factor data with more than two ordered levels

R code used for imputation is as follows

```
AP_clean_mice <- mice(AP_clean[,-1], m=1, maxit = 3, seed = 88)
```

This results in one imputed dataset.

B) Imputation of missing values using Amelia Package

The Amelia package requires all the nominal variables, ordinal variables, ID variables and time series variables defined explicitly. Therefore, all the nominal variables are grouped into one vector called nominal_vars and all the ordinal variables into ordinal_vars. In the R command, we have specified m = 3 which represents 3 sets of imputed data being created. Since the

dimentionality of the dataset is large and having very less observations when compared to its columns, the **Estimation Maximization EM** algorithm failed to converge and impute results. Hence, the value of empri can be tuned upto 10 times the rows of the dataset. After multiple trials in our imputation, empri is set to 0.8. We also remove the variable -main reason to admit in the hospital as it has only 1 value - stomach ache. The R code used for imputation is as follows

```
AP_clean_amelia <- amelia(AP_clean[,-c(1,4)], m=3, p2s=2, noms = nominal_vars, ords = ordinal_vars, seed=88, parallel = 'snow', empri = 0.8*nrow(AP_clean[,-c(1,4)]))
```

This results in three imputed datasets.

C) Imputation of missing values using MissForest package

Miss forest uses a state-of-the-art algorithm **Random forest** from the family of Decision Tress algorithm to perform data imputations. The major advantages of decision trees over other algorithms are they can handle unscaled variables as well as outliers in variables. The decision tree algorithms are known to perform better with less observations and unscaled and out of range values of variables than conventional regression algorithms and classifiers. Therefore, the complexity of coding and parameter tuning is greatly reduced using Miss forest and the results were obtained faster than Amelia and MICE.
Initially we use the missforest function on our dataset to perform imputation. Later, we calaculate the imputation error. The resulting imputed data is then stored in a separate dataframe. R code used for imputation is as follows

```
AP_clean_missforest <- missForest(AP_clean[,-1])
AP_clean_missforest$ximp
```

This function calculates the imputation error:

```
AP_clean_missforest$OOBerror
```

```
> # checking the imputation error of Miss forest
> AP_clean_missforest$OOBerror
    NRMSE       PFC
0.3674231 0.2767761
```

NRMSE = 0.36742 represents the imputation error for Continuous variables and PFC =0.27677 represents the imputation error for Categorical variables

This results in one imputed dataset.

1) Based on the computation of (3), compare the results from 3 different packages.

Based on the summary statistics for imputations using three given packages, we can infer that

1. There is reduced negative skewness of data when imputation is performed using Amelia whereas the negative skewness remains the same in case of Mice and Missforest.
2. The highest mean value for subclinical examination – amylase keeps reducing for each package. In mice, it is 530.67, whereas 399.43 in Amelia and 407.73 in MissForest.
3. The variable Prothrombin has the highest range of 843 which remains the same across all the packages.
4. In our opinion, missforest package accepts any type of data upon which imputation can be performed and is convenient and user friendly.
5. Lesser imputation accuracy is obtained with Mice package followed by Missforest. Comparatively, highest accuracy of imputation is obtained with Amelia.
6. A regression model was fit for the response variable – Number of days in hospital to compare all the three imputed datasets (Discussed in detail in Q6). Based on the R2 values of the model fit, it was observed that the dataset obtained after imputation with Ameila was a better fit model, followed by Miss Forest and Mice imputed.

The details and range of values in each variable after imputations can be observed from the SAS outputs below

Summary statistics for MICE imputed Dataset

The MEANS Procedure

Variable	N	N Miss	Mean	Median	Minimum	Maximum	Range	Skewness
Age	162	0	41.1481481	41.0000000	21.0000000	77.0000000	56.0000000	0.5753243
Gender	162	0	0.6851852	1.0000000	0	1.0000000	1.0000000	-0.8049245
Main.reason.to.admit.hospital	158	4	1.0000000	1.0000000	1.0000000	1.0000000	0	.
Duration.of.staying.in.hospitals	162	0	6.7777778	6.0000000	2.0000000	18.0000000	16.0000000	0.8412181
Hereditary.information	162	0	0.3518519	0	0	1.0000000	1.0000000	0.6262671
Gallbladder.problem	162	0	0.0061728	0	0	1.0000000	1.0000000	12.7279221
Drinking.problem	162	0	0.4444444	0	0	1.0000000	1.0000000	0.2257020
A.breakdown.of.drinking.problem	162	0	12.3271605	10.0000000	0	30.0000000	30.0000000	0.5372476
Diabetes.problem	162	0	0.1913580	0	0	1.0000000	1.0000000	1.5839226
Historical.cholecystitis.problem	162	0	0.5000000	0.5000000	0	1.0000000	1.0000000	0
Vomitting	162	0	0.5864198	1.0000000	0	1.0000000	1.0000000	-0.3542495
Clinical.symptoms.of.Abdominal.d	162	0	1.1172840	1.0000000	0	2.0000000	2.0000000	1.2681542
Clinical.symptoms...13	162	0	19.7592593	18.0000000	2.0000000	46.0000000	44.0000000	0.5325487
Clinical.symptoms...14	162	0	22.7175926	23.1100000	15.6300000	31.7200000	16.0900000	-0.1129477
Clinical.symptoms...15	162	0	107.2160494	107.5000000	68.0000000	158.0000000	90.0000000	0.1036862
Body.temperature	162	0	37.1574074	37.0000000	36.3000000	39.5000000	3.2000000	1.8388597
Saturation.of.peripheral.oxygen	162	0	97.0308642	98.0000000	90.0000000	100.0000000	10.0000000	-1.8712328
apache.2.score.at.the.points.of.	162	0	3.9814815	3.0000000	0	16.0000000	16.0000000	0.8793032
ranson.score.at.the.points.of.ad	162	0	1.6358025	2.0000000	0	5.0000000	5.0000000	0.6479153
CTSI.score.at.the.points.of.admi	162	0	4.2716049	4.0000000	0	10.0000000	10.0000000	0.8779325
imre.score.at.the.points.of.admi	162	0	1.3827160	1.0000000	0	4.0000000	4.0000000	0.1971558
sofa.score.at.the.points.of.admi	162	0	1.5864198	1.0000000	0	7.0000000	7.0000000	1.0635489
subclinical.examination....Abdom	162	0	0.9135802	1.0000000	0	2.0000000	2.0000000	0.1367378
subclinical.examination...baltha	162	0	3.0370370	3.0000000	0	4.0000000	4.0000000	-1.1281153
subclinical.examination...CTSI.s	162	0	4.2469136	4.0000000	0	10.0000000	10.0000000	0.8965300
subclinical.examination...white.	162	0	10.5359012	10.0850000	1.6700000	22.5900000	20.9200000	0.4657561
subclinical.examination.....28	162	0	10.4848148	9.8350000	0.9400000	21.5700000	20.6300000	0.1406235
subclinical.examination.....29	162	0	8.9368519	9.0050000	0.7800000	19.8400000	19.0600000	0.1189479
subclinical.examination.....30	162	0	10.0854198	10.1050000	2.5600000	20.3100000	17.7500000	0.1641694
subclinical.examination.....31	162	0	11.4425309	10.3850000	0.5000000	23.6800000	23.1800000	0.3647690
subclinical.examination.....32	162	0	193.1925926	178.0000000	14.6000000	422.0000000	407.4000000	0.7273201
subclinical.examination.....33	162	0	195.8765432	190.0000000	71.0000000	307.0000000	236.0000000	-0.0298230
subclinical.examination.....34	162	0	148.9362963	142.0000000	1.9600000	296.0000000	294.0400000	0.1482749
subclinical.examination.....35	162	0	189.6481481	193.0000000	22.0000000	363.0000000	341.0000000	0.0613375
subclinical.examination.....36	162	0	179.8333333	153.0000000	52.0000000	393.0000000	341.0000000	0.5500216
subclinical.examination...Hemato	162	0	0.3879938	0.3900000	0.2260000	0.5920000	0.3660000	0.2748797
subclinical.examination.....38	162	0	0.3220247	0.3100000	0	0.4760000	0.4760000	-0.7043587
subclinical.examination.....39	162	0	0.3433457	0.3400000	0.1900000	0.5200000	0.3300000	0.1261595
subclinical.examination.....40	162	0	0.3159012	0.3285000	0.2200000	0.4100000	0.1900000	-0.0397939
red.blood.cell	162	0	4.5285741	4.4850000	2.7000000	6.8800000	4.1800000	0.2947060
subclinical.examination.....42	162	0	4.1140741	4.1200000	2.4000000	6.2400000	3.8400000	0.1002017
subclinical.examination.....43	162	0	3.9783333	3.8300000	2.0900000	6.1300000	4.0400000	0.2554940
subclinical.examination.....44	162	0	3.6545062	3.6150000	2.1100000	5.2300000	3.1200000	0.1068171
subclinical.examination.....45	162	0	3.6474074	3.6100000	2.4200000	5.1900000	2.7700000	0.1711648
prothrombin	162	0	104.4888889	87.0000000	36.0000000	879.0000000	843.0000000	6.1249519
subclinical.examination.....47	162	0	69.3685185	61.3000000	45.0000000	114.0000000	69.0000000	0.4465224
subclinical.examination.....48	162	0	82.2825926	81.0000000	50.1000000	115.0000000	64.9000000	0.0334768
subclinical.examination.....49	162	0	81.5802469	78.0500000	51.0000000	134.1000000	83.1000000	0.4493165
APTT	162	0	1.1922840	1.0550000	0.7600000	4.1600000	3.4000000	4.0322127
subclinical.examination.....51	162	0	2.0610309	1.1600000	0.4100000	5.7300000	5.3200000	1.0161135
subclinical.examination.....52	162	0	1.2823457	1.1000000	0.8400000	3.4600000	2.6200000	2.1580434
subclinical.examination.....53	162	0	1.1242593	1.1100000	0.5500000	1.6100000	1.0600000	-0.0620382
subclinical.examination...Fibrin	162	0	5.5134630	5.3400000	1.2540000	12.0000000	10.7460000	0.2821240
subclinical.examination.....55	162	0	4.9580679	4.5455000	1.7790000	8.5540000	6.7750000	0.1254304
subclinical.examination.....56	162	0	6.5034630	6.8360000	1.8400000	11.0100000	9.1700000	-0.1648986

subclinical.examination.....57	162	0	6.0941481	6.5700000	2.5600000	9.2920000	6.7320000	-0.1833850
subclinical.examination...ure	162	0	4.7166667	3.8000000	1.2000000	25.6000000	24.4000000	3.0706073
subclinical.examination.....59	162	0	4.8790123	4.2500000	1.0000000	9.4000000	8.4000000	0.1123205
subclinical.examination.....60	162	0	4.5524691	3.4000000	1.2000000	14.3000000	13.1000000	1.1842035
subclinical.examination.....61	162	0	8.2425926	4.7000000	1.3000000	41.9000000	40.6000000	2.2968944
subclinical.examination...creati	162	0	83.1338272	68.0000000	1.8000000	727.0000000	725.2000000	4.9879213
subclinical.examination.....63	162	0	67.8209877	68.0000000	13.5000000	138.0000000	124.5000000	0.3881699
subclinical.examination.....64	162	0	104.4938272	66.5000000	25.0000000	406.0000000	381.0000000	2.0069615
subclinical.examination.....65	162	0	98.9938272	66.0000000	27.0000000	328.0000000	301.0000000	1.5515353
subclinical.examination...glucos	162	0	13.1099383	9.7500000	3.2000000	66.0000000	62.8000000	2.8394307
subclinical.examination.....67	162	0	771.4790123	309.5000000	30.0000000	3546.00	3516.00	1.3497304
cholesterol	162	0	15.8255556	12.6300000	3.9100000	99.0000000	95.0900000	3.7378527
triglycerid	162	0	27.9497531	18.5250000	11.2100000	131.5500000	120.3400000	2.5562127
subclinical.examination.....70	162	0	18.4569136	7.4700000	1.0100000	76.9400000	75.9300000	1.2725312
subclinical.examination.....71	162	0	13.8192593	7.7250000	0.7200000	84.4200000	83.7000000	2.6886035
subclinical.examination...amylas	162	0	530.6700617	351.0000000	6.6700000	1519.80	1513.13	0.8013602
subclinical.examination...lipase	162	0	529.8339506	366.0000000	6.0000000	1728.10	1722.10	0.9021498
subclinical.examination...protei	162	0	58.9148148	59.5000000	32.1000000	84.3000000	52.2000000	0.1264814
subclinical.examination...albumi	162	0	31.4543210	30.7000000	13.6000000	56.2000000	42.6000000	0.3148844
subclinical.examination...sodium	162	0	131.3827160	132.0000000	108.0000000	141.0000000	33.0000000	-1.0152308
subclinical.examination.....77	162	0	137.5820988	138.0000000	122.0000000	150.0000000	28.0000000	-0.1808373
subclinical.examination.....78	162	0	134.6790123	135.0000000	126.0000000	144.0000000	18.0000000	-0.0366735
subclinical.examination...potasi	162	0	3.7196296	3.6000000	2.6000000	5.3000000	2.7000000	0.7219203
subclinical.examination.....80	162	0	3.7890123	3.5000000	2.4000000	9.5000000	7.1000000	2.9922735
subclinical.examination.....81	162	0	3.5054321	3.3000000	2.8000000	4.9000000	2.1000000	0.8546488
subclinical.examination...pH..in	162	0	7.3914321	7.3900000	7.1000000	7.5700000	0.4700000	-0.6487461
subclinical.examination.....83	162	0	7.3679012	7.4100000	7.1000000	7.5000000	0.4000000	-0.6872264
subclinical.examination.....84	162	0	7.4066605	7.4100000	7.2000000	7.6300000	0.4300000	-0.0328029
subclinical.examination...paco2.	162	0	32.5882716	31.7500000	9.0000000	97.0000000	88.0000000	2.8564515
subclinical.examination.....86	162	0	34.4962963	34.0000000	14.0000000	53.0000000	39.0000000	-0.0996473
subclinical.examination.....87	162	0	33.7506173	35.0000000	18.0000000	53.0000000	35.0000000	-0.0876901
subclinical.examination...pa.Oxy	162	0	94.2055556	88.3500000	32.0000000	251.0000000	219.0000000	1.7521995
subclinical.examination.....89	162	0	81.2000000	68.0000000	45.0000000	165.0000000	120.0000000	0.8399303
subclinical.examination.....90	162	0	96.6981481	86.0000000	2.0000000	201.0000000	199.0000000	0.2996536
subclinical.examination...HCO3..	162	0	19.7506790	19.5500000	-18.6000000	77.6000000	96.2000000	2.3274564
...92	162	0	20.1251235	22.0000000	5.7000000	29.8000000	24.1000000	-0.5234400
...93	162	0	19.6567901	22.6000000	-18.9000000	33.6000000	52.5000000	-1.4671748
BE..in.blood.air.	162	0	-4.9641975	-4.7000000	-24.7000000	16.0000000	40.7000000	-0.0714324
...95	162	0	-4.5222222	-2.9000000	-20.2000000	5.6000000	25.8000000	-0.3540184
...96	162	0	-3.1290123	-2.6500000	-17.9000000	10.0000000	27.9000000	-0.1825023
p.f..paO2..O2.	162	0	317.5086420	333.0000000	3.8000000	562.0000000	558.2000000	-0.2741276
lactatr..in.blood.air.	162	0	2.2172222	1.6000000	0.4000000	9.0000000	8.6000000	1.6781102
...99	162	0	1.5351852	0.9000000	0.4000000	4.7000000	4.3000000	0.9142265
treatment...fluide.intake...100	162	0	4725.62	4300.00	60.0000000	9650.00	9590.00	0.5398841
treatment...fluide.intake...101	162	0	4451.30	4000.00	1000.00	9500.00	8500.00	0.8065474
treatment...fluide.intake...102	162	0	3839.02	3500.00	1200.00	8500.00	7300.00	0.6830151
treatment...fluide.output...103	162	0	2563.21	2350.00	950.0000000	6900.00	5950.00	1.7013966
treatment...fluide.output...104	162	0	3318.52	2650.00	620.0000000	10760.00	10140.00	1.6438767
treatment...fluide.output...105	162	0	3181.27	2630.00	270.0000000	8020.00	7750.00	0.9801471
treatment...balance.fluid.in.and	162	0	456.0925926	800.0000000	-3780.00	2650.00	6430.00	-0.7846999
treatment...day.without.food.int	162	0	1.8765432	1.0000000	0	12.0000000	12.0000000	3.1874307
treatment...PEX.treatment.of.whi	162	0	3.2037037	3.0000000	1.0000000	7.0000000	6.0000000	1.0440151
treatment...number.of.PEX.treatm	162	0	1.6604938	1.0000000	1.0000000	3.0000000	2.0000000	0.6987924
treatment...triglycerid.before.f	162	0	45.1106790	21.3350000	2.4100000	131.5500000	129.1400000	0.8914635
treatment...triglycerid.after.fi	162	0	17.2768519	6.2250000	1.0100000	76.9400000	75.9300000	1.5847462
treatment...APAche.2.sHavere.bef	162	0	4.6804938	4.0000000	0	16.0000000	16.0000000	1.1493156
treatment...APAche.2.sHavere.aft	162	0	2.6851852	2.0000000	0	9.0000000	9.0000000	0.9473550
treatment...Imre.sHavere.before.	162	0	2.0061728	2.0000000	0	4.0000000	4.0000000	0.1456552
treatment...Imre.sHavere.after.f	162	0	1.1296296	1.0000000	0	3.0000000	3.0000000	0.4602510
Result...dead.or.alive	162	0	0.7160494	1.0000000	0	1.0000000	1.0000000	-0.9672541
Potential.Complication	162	0	0.4876543	0	0	1.0000000	1.0000000	0.0498606
Patient.with.PEX.or.without.PEX	159	3	0.4968553	0	0	1.0000000	1.0000000	0.0126990
Blood.pressure.systolic	162	0	119.8148148	120.0000000	90.0000000	140.0000000	50.0000000	-0.2222517
Blood.pressure.diastolic	162	0	76.0493827	80.0000000	60.0000000	100.0000000	40.0000000	0.2479682
Bilirubin_direct_t0	162	0	52.8979012	17.0000000	2.1000000	1401.00	1398.90	7.4794586
Bilirubin_indirect_t0	162	0	10.7296914	5.9000000	0.6000000	81.0000000	80.4000000	2.6344636
Bilirubin_direct_t30	162	0	28.6709877	18.1500000	4.8000000	132.0000000	127.2000000	2.0382742
Bilirubin_indirect_t30	162	0	17.2302469	6.0000000	1.8000000	84.5000000	82.7000000	1.8041200
Bilirubin_direct_t72	162	0	46.6104938	17.9000000	5.0000000	213.0000000	208.0000000	1.6996715
Bilirubin_indirect_t72	162	0	29.1339506	18.3000000	2.1000000	100.0000000	97.9000000	0.9631917
AST..liver.funtion.	162	0	114.3596296	45.2500000	7.9000000	2315.00	2307.10	7.0464007
ALT..liver.funtion.	162	0	61.2395062	25.0000000	4.2000000	326.0000000	321.8000000	2.1895990
subclinical.examination...calci_	162	0	4.9559259	1.5350000	0.6100000	35.0000000	34.3900000	2.5445111

Summary statistics for Amelia imputed Dataset

The MEANS Procedure

Variable	N	N Miss	Mean	Median	Minimum	Maximum	Range	Skewness
Age	162	0	41.1481481	41.0000000	21.0000000	77.0000000	56.0000000	0.5753243
Duration.of.staying.in.hospital	162	0	6.7777778	6.0000000	2.0000000	18.0000000	16.0000000	0.8412181
A.breakdown.of.drinking.problem	162	0	9.3333333	9.0000000	0	30.0000000	30.0000000	0.8021180
Diabetes.problem	162	0	0.1913580	0	0	1.0000000	1.0000000	1.5839226
Historical.cholecystitis.proble	162	0	0.5061728	1.0000000	0	1.0000000	1.0000000	-0.0249246
Clinical.symptoms...24	162	0	19.0447832	18.4319667	2.0000000	46.0000000	44.0000000	0.4962185
Clinical.symptoms...25	162	0	22.6575404	23.0950000	15.6300000	31.7200000	16.0900000	-0.1250960
Clinical.symptoms...26	162	0	107.4411000	107.8010742	68.0000000	158.0000000	90.0000000	0.1188548
Body.temperature	162	0	37.1397282	37.0000000	36.3000000	39.5000000	3.2000000	1.7912240
Saturation.of.peripheral.oxygen	162	0	97.0965081	98.0000000	90.0000000	100.0000000	10.0000000	-1.8631745
apache.2.score.at.the.points.of	162	0	3.4197531	3.0000000	0	16.0000000	16.0000000	0.9011523
subclinical.examination...white	162	0	10.5659013	10.2150000	1.6700000	22.5900000	20.9200000	0.4509828
subclinical.examination.....47	162	0	9.7926588	9.8116642	-0.0480232	21.5700000	21.6180232	0.2431308
subclinical.examination.....48	162	0	9.0093632	9.1145241	0.7800000	19.8400000	19.0600000	0.0195481
subclinical.examination.....49	162	0	9.2063707	9.3023156	2.5600000	20.3100000	17.7500000	0.3019177
subclinical.examination.....50	162	0	10.4151105	10.2214163	0.5000000	23.6800000	23.1800000	0.5086357
subclinical.examination.....51	162	0	193.3680108	178.0000000	14.6000000	422.0000000	407.4000000	0.7204683
subclinical.examination.....52	162	0	180.2182930	181.1617476	60.9035616	307.0000000	246.0964384	0.2674961
subclinical.examination.....53	162	0	153.2650441	150.2590326	1.9600000	296.0000000	294.0400000	0.1609543
subclinical.examination.....54	162	0	170.4579088	172.6374796	22.0000000	363.0000000	341.0000000	0.1775455
subclinical.examination.....55	162	0	170.8983022	169.8311495	21.8988518	393.0000000	371.1011482	0.3851183
subclinical.examination...Hemat	162	0	0.3900124	0.3900000	0.2260000	0.5920000	0.3660000	0.2993229
subclinical.examination.....57	162	0	0.3370791	0.3409603	0	0.4760000	0.4760000	-0.8925450
subclinical.examination.....58	162	0	0.3498705	0.3463229	0.1900000	0.5200000	0.3300000	0.0745768
subclinical.examination.....59	162	0	0.3187136	0.3211998	0.2200000	0.4100000	0.1900000	-0.1593091
red.blood.cell	162	0	4.4767554	4.4850000	2.7000000	6.8800000	4.1800000	0.2391116
subclinical.examination.....61	162	0	3.8753877	3.8993777	1.8204072	6.2400000	4.4195928	0.1541343
subclinical.examination.....62	162	0	3.9506764	3.9199596	2.0900000	6.1300000	4.0400000	0.1198501
subclinical.examination.....63	162	0	3.7192528	3.7752814	2.1100000	5.2300000	3.1200000	-0.0670000
subclinical.examination.....64	162	0	3.5635122	3.5511169	2.4200000	5.1900000	2.7700000	0.1686968
prothrombin	162	0	96.9335201	87.0000000	36.0000000	879.0000000	843.0000000	8.2377389
subclinical.examination.....66	162	0	74.1005443	75.2500000	32.3775293	114.0000000	81.6224707	-0.1425960
subclinical.examination.....67	162	0	79.6730620	78.9997041	44.5157436	115.0000000	70.4842564	0.2046565
subclinical.examination.....68	162	0	79.9032843	79.0759557	36.1317338	134.1000000	97.9682662	0.2500115
APTT	162	0	1.1430890	1.0450000	0.7527965	4.1600000	3.4072035	5.0226392
subclinical.examination.....70	162	0	1.4332614	1.2703346	-0.0703303	5.7300000	5.8003303	1.9895847
subclinical.examination.....71	162	0	1.1938931	1.1700000	0.5645228	3.4600000	2.8954772	2.8763914
subclinical.examination.....72	162	0	1.0930018	1.0929729	0.5500000	1.6100000	1.0600000	0.0880380
subclinical.examination...Fibri	162	0	5.6285109	5.4795000	1.2540000	12.0000000	10.7460000	0.3466177
subclinical.examination.....74	162	0	5.4385603	5.3130294	1.7790000	9.3738699	7.5948699	0.0684643
subclinical.examination.....75	162	0	6.3130718	6.4217609	1.3578502	11.0100000	9.6521498	-0.1776497
subclinical.examination.....76	162	0	6.3479518	6.4665000	2.5600000	9.2920000	6.7320000	-0.2652858
subclinical.examination...ure	162	0	4.7335076	3.8000000	1.2000000	25.6000000	24.4000000	3.0719104
subclinical.examination.....78	162	0	3.3429056	3.1212547	-0.2115394	9.4000000	9.6115394	0.5926713
subclinical.examination.....79	162	0	4.5862538	4.1000000	-0.3874856	14.3000000	14.6874856	1.0850866
subclinical.examination.....80	162	0	5.3420678	4.4051045	-2.9887012	41.9000000	44.8887012	3.8043246
subclinical.examination...creat	162	0	83.1834748	68.5000000	-2.6362599	727.0000000	729.6362599	5.0898380
subclinical.examination.....82	162	0	64.1961283	64.0958910	13.5000000	138.0000000	124.5000000	0.5364768
subclinical.examination.....83	162	0	70.7422133	63.0000000	-17.2420010	406.0000000	423.2420010	2.9178501
subclinical.examination.....84	162	0	72.0880162	63.9250735	1.7896559	328.0000000	326.2103441	2.5920307
subclinical.examination...gluco	162	0	11.1757638	10.0750000	1.5530408	66.0000000	64.4469592	3.6545316
subclinical.examination.....97	162	0	407.1838051	296.0000000	-823.0590406	3546.00	4369.06	2.1504188
cholesterol	162	0	15.4225292	14.0483600	-9.7139878	99.0000000	108.7139878	3.7986852
triglycerid	162	0	27.6047771	19.1350000	7.8414536	131.5500000	123.7085464	2.5834845
subclinical.examination.....103	162	0	10.3064117	8.2222897	-18.1391882	76.9400000	95.0791882	1.6961530
subclinical.examination.....104	162	0	11.4559253	9.1750000	-10.0817512	84.4200000	94.5017512	2.0221463

subclinical.examination...amyla	162	0	399.4318533	351.0000000	-282.9702967	1519.80	1802.77	1.1733408
subclinical.examination...lipas	162	0	502.6819400	466.1307142	-258.0071125	1728.10	1986.11	0.6441149
subclinical.examination...prote	162	0	59.0758067	59.5000000	32.1000000	84.3000000	52.2000000	0.1344887
subclinical.examination...album	162	0	31.1246600	30.5398540	13.6000000	56.2000000	42.6000000	0.2545754
subclinical.examination...sodiu	162	0	131.7333637	132.0000000	108.0000000	141.0000000	33.0000000	-0.9159358
subclinical.examination.....118	162	0	137.5253024	137.9995750	122.0000000	150.0000000	28.0000000	-0.2123199
subclinical.examination.....119	162	0	135.7870713	136.0000000	126.0000000	146.0224757	20.0224757	-0.1019311
subclinical.examination...potas	162	0	3.6755614	3.6000000	2.6000000	5.3000000	2.7000000	0.6783735
subclinical.examination.....121	162	0	3.4278448	3.4279278	1.9897720	9.5000000	7.5102280	3.3282198
subclinical.examination.....122	162	0	3.4469781	3.4000000	2.7027245	4.9000000	2.1972755	0.8474530
subclinical.examination...pH..i	162	0	7.3923816	7.3941982	7.1000000	7.5700000	0.4700000	-0.7882964
subclinical.examination.....126	162	0	7.3900649	7.3975956	7.1000000	7.5973273	0.4973273	-0.4195976
subclinical.examination.....127	162	0	7.3884582	7.3909962	7.2000000	7.6300000	0.4300000	-0.000652605
subclinical.examination...paco2	162	0	32.2343349	32.0000000	9.0000000	97.0000000	88.0000000	2.2210386
subclinical.examination.....130	162	0	34.4596356	34.0000000	14.0000000	54.1388884	40.1388884	0.0675553
subclinical.examination.....131	162	0	33.5633276	33.3501936	13.9692007	54.8302397	40.8610390	0.1797766
subclinical.examination...pa.Ox	162	0	91.2619129	88.0000000	32.0000000	251.0000000	219.0000000	1.8749733
subclinical.examination.....135	162	0	77.3823063	76.7764495	20.9191633	165.0000000	144.0808367	0.4643834
subclinical.examination.....136	162	0	89.1635743	88.1649057	2.0000000	201.0000000	199.0000000	0.4575434
subclinical.examination...HCO3.	162	0	19.8357967	19.9000000	-18.6000000	77.6000000	96.2000000	1.9266416
...140	162	0	21.6429455	21.7650006	5.7000000	34.1340953	28.4340953	-0.2631656
...141	162	0	20.8252154	21.2337929	-18.9000000	33.6000000	52.5000000	-1.8241979
BE..in.blood.air.	162	0	-4.8546537	-4.4000000	-24.7000000	16.0000000	40.7000000	-0.0491868
...145	162	0	-2.9874434	-2.9506036	-20.2000000	10.4343970	30.6343970	-0.4013444
...146	162	0	-3.0854841	-3.0356285	-17.9000000	10.0000000	27.9000000	-0.0572508
p.f..paO2..O2.	162	0	341.6524829	350.2699872	3.8000000	571.6842983	567.8842983	-0.3432357
lactatr..in.blood.air.	162	0	2.0539508	1.6000000	-0.5536192	9.0000000	9.5536192	1.7685871
...156	162	0	1.2598636	1.1098531	-0.2028574	4.7000000	4.9028574	1.3860314
treatment...fluide.intake...159	162	0	4584.02	4250.00	60.0000000	9650.00	9590.00	0.5101447
treatment...fluide.intake...160	162	0	4221.42	4000.00	1000.00	9500.00	8500.00	0.7208102
treatment...fluide.intake...161	162	0	3790.88	3737.08	1200.00	8500.00	7300.00	0.6688686
treatment...fluide.output...162	162	0	2529.23	2350.00	349.1673788	6900.00	6550.83	1.7322061
treatment...fluide.output...163	162	0	3151.04	2710.00	620.0000000	10760.00	10140.00	1.5263320
treatment...fluide.output...164	162	0	3036.84	2650.00	-126.9175970	8020.00	8146.92	1.0099186
treatment...balance.fluid.in.an	162	0	727.3466465	843.1385950	-3780.00	3077.85	6857.85	-0.9013379
treatment...triglycerid.before.	162	0	25.2207595	19.5486826	-30.9182209	131.5500000	162.4682209	1.5212097
treatment...triglycerid.after.f	162	0	8.6144705	6.3755024	-24.3913341	76.9400000	101.3313341	2.1382791
Blood.pressure.systolic	162	0	117.8386692	119.9687757	87.5752557	140.7856608	53.2104051	-0.0570742
Blood.pressure.diastolic	162	0	71.7744763	70.3577077	54.3819218	100.0000000	45.6180782	0.5104329
Bilirubin_direct_t0	162	0	41.4197277	17.5500000	-68.5005125	1401.00	1469.50	9.3411214
Bilirubin_indirect_t0	162	0	9.0550258	5.9000000	-16.0763589	81.0000000	97.0763589	3.1032991
Bilirubin_direct_t30	162	0	19.7979845	18.0541107	-14.8104934	132.0000000	146.8104934	2.4768094
Bilirubin_indirect_t30	162	0	8.3671583	6.1997100	-20.9450216	84.5000000	105.4450216	2.0673375
Bilirubin_direct_t72	162	0	25.7137318	21.5515484	-58.0683801	213.0000000	271.0683801	2.4834872
Bilirubin_indirect_t72	162	0	13.1775487	9.7335638	-20.7059844	100.0000000	120.7059844	1.2989892
AST..liver.funtion.	162	0	116.3697981	49.0000000	-564.1606612	2315.00	2879.16	3.9730990
ALT..liver.funtion.	162	0	37.3899747	25.0000000	-53.0998551	326.0000000	379.0998551	3.3000017
subclinical.examination...calci	162	0	2.0813869	1.8200000	-8.8855564	35.0000000	43.8855564	4.8390888

Summary statistics for Miss Forest imputed Dataset

The MEANS Procedure

Variable	N	N Miss	Mean	Median	Minimum	Maximum	Range	Skewness
Age	162	0	41.1481481	41.0000000	21.0000000	77.0000000	56.0000000	0.5753243
Duration.of.staying.in.hospital	162	0	6.7777778	6.0000000	2.0000000	18.0000000	16.0000000	0.8412181
A.breakdown.of.drinking.problem	162	0	10.3168519	9.9200000	0	30.0000000	30.0000000	1.6213122
Diabetes.problem	162	0	0.1920988	0	0	1.0000000	1.0000000	1.5818994
Historical.cholecystitis.proble	162	0	0.5034568	0.7800000	0	1.0000000	1.0000000	-0.0147006
Clinical.symptoms...24	162	0	18.7115123	18.0000000	2.0000000	46.0000000	44.0000000	0.8713453
Clinical.symptoms...25	162	0	22.6871975	23.0950000	15.6300000	31.7200000	16.0900000	-0.1186225
Clinical.symptoms...26	162	0	107.3168519	106.9400000	68.0000000	158.0000000	90.0000000	0.1342973
Body.temperature	162	0	37.1389444	37.0000000	36.3000000	39.5000000	3.2000000	1.8439512
Saturation.of.peripheral.oxygen	162	0	97.0788272	98.0000000	90.0000000	100.0000000	10.0000000	-1.8571899
apache.2.score.at.the.points.of	162	0	3.4169753	3.0000000	0	16.0000000	16.0000000	1.0163181
subclinical.examination...white	162	0	10.5637037	10.2150000	1.6700000	22.5900000	20.9200000	0.4539746
subclinical.examination.....47	162	0	9.8742833	9.6364500	0.9400000	21.5700000	20.6300000	0.5339280
subclinical.examination.....48	162	0	8.9290932	8.9600000	0.7800000	19.8400000	19.0600000	0.1592172
subclinical.examination.....49	162	0	9.5600459	9.9000900	2.5600000	20.3100000	17.7500000	0.2086682
subclinical.examination.....50	162	0	10.7043623	10.7968000	0.5000000	23.6800000	23.1800000	0.6348201
subclinical.examination.....51	162	0	193.1006173	176.5000000	14.6000000	422.0000000	407.4000000	0.7299156
subclinical.examination.....52	162	0	174.2198765	168.9150000	71.0000000	307.0000000	236.0000000	0.5502602
subclinical.examination.....53	162	0	153.1442963	148.7900000	1.9600000	296.0000000	294.0400000	0.2228171
subclinical.examination.....54	162	0	167.7409259	161.7550000	22.0000000	363.0000000	341.0000000	0.4554218
subclinical.examination.....55	162	0	182.4490741	173.6000000	52.0000000	393.0000000	341.0000000	0.6866310
subclinical.examination...Hemat	162	0	0.3894756	0.3900000	0.2260000	0.5920000	0.3660000	0.3214800
subclinical.examination.....57	162	0	0.3434396	0.3554750	0	0.4760000	0.4760000	-1.6321263
subclinical.examination.....58	162	0	0.3465054	0.3428150	0.1900000	0.5200000	0.3300000	0.1767080
subclinical.examination.....59	162	0	0.3238901	0.3359650	0.2200000	0.4100000	0.1900000	-0.5623585
red.blood.cell	162	0	4.4732322	4.4600000	2.7000000	6.8800000	4.1800000	0.2767037
subclinical.examination.....61	162	0	4.0106852	4.1134000	2.4000000	6.2400000	3.8400000	0.0774765
subclinical.examination.....62	162	0	3.9039198	3.8130000	2.0900000	6.1300000	4.0400000	0.3804050
subclinical.examination.....63	162	0	3.7579852	3.8118500	2.1100000	5.2300000	3.1200000	-0.2234163
subclinical.examination.....64	162	0	3.7010389	3.8000000	2.4200000	5.1900000	2.7700000	-0.3051446
prothrombin	162	0	95.5844568	87.0000000	36.0000000	879.0000000	843.0000000	8.6255928
subclinical.examination.....66	162	0	72.7720926	72.7420000	45.0000000	114.0000000	69.0000000	0.3693743
subclinical.examination.....67	162	0	80.8070506	80.9722000	50.1000000	115.0000000	64.9000000	0.3287356
subclinical.examination.....68	162	0	82.5802654	82.7590000	51.0000000	134.1000000	83.1000000	0.6826660
APTT	162	0	1.1452512	1.0600000	0.7600000	4.1600000	3.4000000	5.1866326
subclinical.examination.....70	162	0	1.5643247	1.4878700	0.4100000	5.7300000	5.3200000	3.7462669
subclinical.examination.....71	162	0	1.2096025	1.1946500	0.8400000	3.4600000	2.6200000	4.8791223
subclinical.examination.....72	162	0	1.0726093	1.0600000	0.5500000	1.6100000	1.0600000	0.7059704
subclinical.examination...Fibri	162	0	5.6138295	5.5454100	1.2540000	12.0000000	10.7460000	0.3185568
subclinical.examination.....74	162	0	5.2218472	5.2956900	1.7790000	8.5540000	6.7750000	-0.0664436
subclinical.examination.....75	162	0	6.2400952	6.2511350	1.8400000	11.0100000	9.1700000	-0.1971081
subclinical.examination.....76	162	0	6.2301720	6.2401600	2.5600000	9.2920000	6.7320000	-0.4466046
subclinical.examination...ure	162	0	4.7335432	3.8000000	1.2000000	25.6000000	24.4000000	3.0901815
subclinical.examination.....78	162	0	3.4175679	3.1400000	1.0000000	9.4000000	8.4000000	1.7505673
subclinical.examination.....79	162	0	4.3177778	3.8655000	1.2000000	14.3000000	13.1000000	1.8341458
subclinical.examination.....80	162	0	5.1539877	4.4720000	1.3000000	41.9000000	40.6000000	6.5283790
subclinical.examination...creat	162	0	83.0592123	68.0000000	1.8000000	727.0000000	725.2000000	5.1611439
subclinical.examination.....82	162	0	66.4741975	63.8500000	13.5000000	138.0000000	124.5000000	0.7285442
subclinical.examination.....83	162	0	71.2522222	63.4650000	25.0000000	406.0000000	381.0000000	5.2938282
subclinical.examination.....84	162	0	67.3489506	61.4900000	27.0000000	328.0000000	301.0000000	4.8664168
subclinical.examination...gluco	162	0	11.1120802	10.4028500	3.2000000	66.0000000	62.8000000	4.4397565
subclinical.examination.....97	162	0	454.1109877	346.8400000	30.0000000	3546.00	3516.00	3.2507526
cholesterol	162	0	15.4978679	14.1459000	3.9100000	99.0000000	95.0900000	4.7026045
triglycerid	162	0	27.5456549	19.1350000	11.2100000	131.5500000	120.3400000	2.6196616
subclinical.examination.....103	162	0	11.5796790	10.4259500	1.0100000	76.9400000	75.9300000	4.0067940
subclinical.examination.....104	162	0	9.1486302	8.2570000	0.7200000	84.4200000	83.7000000	7.3531076

subclinical.examination...amyla	162	0	407.7386549	363.3716000	6.6700000	1519.80	1513.13	1.6879405
subclinical.examination...lipas	162	0	526.8546605	502.0290000	6.0000000	1728.10	1722.10	1.1715047
subclinical.examination...prote	162	0	59.4488395	60.0000000	32.1000000	84.3000000	52.2000000	0.0917668
subclinical.examination...album	162	0	31.0734136	30.7500000	13.6000000	56.2000000	42.6000000	0.3385767
subclinical.examination...sodiu	162	0	131.6519753	132.0000000	108.0000000	141.0000000	33.0000000	-0.8972013
subclinical.examination.....118	162	0	137.2682778	137.5600000	122.0000000	150.0000000	28.0000000	-0.4507570
subclinical.examination.....119	162	0	135.9195679	136.1950000	126.0000000	144.0000000	18.0000000	-0.5278288
subclinical.examination...potas	162	0	3.6720568	3.6000000	2.6000000	5.3000000	2.7000000	0.7029143
subclinical.examination.....121	162	0	3.6507074	3.6000000	2.4000000	9.5000000	7.1000000	6.7461901
subclinical.examination.....122	162	0	3.4329204	3.4390000	2.8000000	4.9000000	2.1000000	1.5341670
subclinical.examination...pH..i	162	0	7.3918441	7.4000000	7.1000000	7.5700000	0.4700000	-0.8405046
subclinical.examination.....126	162	0	7.3850006	7.3946500	7.1000000	7.5000000	0.4000000	-1.7456649
subclinical.examination.....127	162	0	7.4091757	7.4129750	7.2000000	7.6300000	0.4300000	-0.6037636
subclinical.examination...paco2	162	0	32.0150185	32.0000000	9.0000000	97.0000000	88.0000000	2.4476085
subclinical.examination.....130	162	0	35.0719321	35.2790000	14.0000000	53.0000000	39.0000000	-0.3310962
subclinical.examination.....131	162	0	34.1074938	34.7955000	18.0000000	53.0000000	35.0000000	-0.3152392
subclinical.examination...pa.Ox	162	0	92.2827716	88.3500000	32.0000000	251.0000000	219.0000000	1.9366774
subclinical.examination.....135	162	0	81.8374877	80.8875000	45.0000000	165.0000000	120.0000000	2.1699501
subclinical.examination.....136	162	0	86.4144012	84.9220000	2.0000000	201.0000000	199.0000000	1.3261469
subclinical.examination...HCO3.	162	0	19.8033877	20.0000000	-18.6000000	77.6000000	96.2000000	1.9873778
...140	162	0	21.1347358	21.4171000	5.7000000	29.8000000	24.1000000	-1.2135883
...141	162	0	21.4341358	22.0040000	-18.9000000	33.6000000	52.5000000	-3.0921781
BE..in.blood.air.	162	0	-4.7341111	-4.4000000	-24.7000000	16.0000000	40.7000000	-0.1375404
...145	162	0	-3.4214136	-2.5785000	-20.2000000	5.6000000	25.8000000	-1.0573066
...146	162	0	-2.3411420	-2.0870000	-17.9000000	10.0000000	27.9000000	-0.5738091
p.f..paO2..O2.	162	0	336.0035679	343.3730000	3.8000000	562.0000000	558.2000000	-0.4602158
lactatr..in.blood.air.	162	0	2.1241327	1.7000000	0.4000000	9.0000000	8.6000000	1.9207916
...156	162	0	1.2637654	1.1920000	0.4000000	4.7000000	4.3000000	2.8049380
treatment...fluide.intake...159	162	0	4658.70	4427.65	60.0000000	9650.00	9590.00	0.4823851
treatment...fluide.intake...160	162	0	4198.32	3984.15	1000.00	9500.00	8500.00	0.8224321
treatment...fluide.intake...161	162	0	3707.70	3500.00	1200.00	8500.00	7300.00	1.0129579
treatment...fluide.output...162	162	0	2575.82	2396.80	950.0000000	6900.00	5950.00	1.9222386
treatment...fluide.output...163	162	0	3147.11	2790.00	620.0000000	10760.00	10140.00	1.6265314
treatment...fluide.output...164	162	0	2992.11	2650.00	270.0000000	8020.00	7750.00	1.5842374
treatment...balance.fluid.in.an	162	0	739.6906173	786.7500000	-3780.00	2650.00	6430.00	-1.1539207
treatment...triglycerid.before.	162	0	31.0109833	26.5550000	2.4100000	131.5500000	129.1400000	2.6788851
treatment...triglycerid.after.f	162	0	9.7708988	8.5358500	1.0100000	76.9400000	75.9300000	4.8691673
Blood.pressure.systolic	162	0	120.1864198	120.0000000	90.0000000	140.0000000	50.0000000	-0.3038958
Blood.pressure.diastolic	162	0	74.6382716	75.5000000	60.0000000	100.0000000	40.0000000	-0.0079166
Bilirubin_direct_t0	162	0	40.4757086	18.9000000	2.1000000	1401.00	1398.90	10.1731353
Bilirubin_indirect_t0	162	0	9.2415154	6.7208000	0.6000000	81.0000000	80.4000000	4.0536168
Bilirubin_direct_t30	162	0	20.0773086	18.2445000	4.8000000	132.0000000	127.2000000	5.4174559
Bilirubin_indirect_t30	162	0	9.8748580	8.7040000	1.8000000	84.5000000	82.7000000	5.4151350
Bilirubin_direct_t72	162	0	28.1790617	22.7520000	5.0000000	213.0000000	208.0000000	5.0097744
Bilirubin_indirect_t72	162	0	14.2727901	12.3450000	2.1000000	100.0000000	97.9000000	3.8148610
AST..liver.funtion.	162	0	80.2771630	49.2000000	7.9000000	2315.00	2307.10	11.1265252
ALT..liver.funtion.	162	0	39.8468951	30.6175000	4.2000000	326.0000000	321.8000000	4.6673444
subclinical.examination...calci	162	0	2.6582179	2.1109000	0.6100000	35.0000000	34.3900000	8.6126996

Predictive Modelling – Regression Analysis:

There are 2 major objectives of this study. One to determine the ultimate outcome of the patient. That is for a comparative treatment of AP with or without PEX, does a patient survives or not? and secondly, the number of days a patient was hospitalised for the treatment of AP. To predict the number of days of hospitalization, using predictive modelling, a linear regression or a Poisson regression could be fit to study the response. However, to determine the patient outcome, a classification algorithm like logistic regression, or classifiers could be used.
For both response variables, we ran feature selection techniques in Python to determine the significant variables or predictors that are most important in the study.

Based on our imputation, we perform **Lasso Regression** on Missforest dataset, using

(a) Duration of Hospital Stay as response variable.

```
# printing the number of total and selected features

selected_feat = X_train.columns[(feature_sel_model.get_support())]

print('Dataset: MISS FOREST Imputed')
print()
print('Total features:{}'.format((X_train.shape[1])))
print('Selected Features:{}'.format(len(selected_feat)))
print('features with coefficients shrank to zero:{}'.format(np.sum(feature_sel_model.estimator_.coef_==0)))
```

```
Dataset: MISS FOREST Imputed

Total features:130
Selected Features:108
features with coefficients shrank to zero:22
```

As per the regression analysis, 108 out of 130 features are selected and dropped remaining 22 as they are insignificant. For these 22 variables, α is almost equal to zero.

The list of the 108 selected predictors by the lasso regression feature selection is as follows :

```
Age
Hereditary.information
Drinking.problem
A.breakdown.of.drinking.problem...13
Historical.cholecystitis.problem
Clinical.symptoms.of.Abdominal.distension
Clinical.symptoms...24
Clinical.symptoms...25
Clinical.symptoms...26
Body.temperature
Saturation.of.peripheral.oxygen
apache.2.score.at.the.points.of.admitting.hospitals
ranson.score.at.the.points.of.admitting.hospitals
CTSI.score.at.the.points.of.admitting.hospitals
subclinical.examination....Abdominal.fluid..computer.tomography
subclinical.examination...balthazar.sHavere..with.computer.tomography.
subclinical.examination...CTSI.score..with.computer.tomography.
subclinical.examination...white.blood.cell..t0..at.the.points.of.admitt
ing.hospitals..t6..after.6h.of.admitting.hospitals...
subclinical.examination.....47
subclinical.examination.....49
subclinical.examination.....50
subclinical.examination.....51
subclinical.examination.....52
subclinical.examination.....53
subclinical.examination.....54
```

subclinical.examination.....55
red.blood.cell
subclinical.examination.....61
subclinical.examination.....62
subclinical.examination.....64
prothrombin
subclinical.examination.....66
subclinical.examination.....67
subclinical.examination.....68
APTT
subclinical.examination.....70
subclinical.examination...Fibrinogen
subclinical.examination.....74
subclinical.examination.....76
subclinical.examination...ure
subclinical.examination.....78
subclinical.examination.....79
subclinical.examination.....80
subclinical.examination...creatinin
subclinical.examination.....82
subclinical.examination.....83
subclinical.examination.....84
subclinical.examination...glucose
subclinical.examination.....97
cholesterol
triglycerid
subclinical.examination.....103
subclinical.examination.....104
subclinical.examination...amylase
subclinical.examination...lipase
subclinical.examination...protein
subclinical.examination...albumin
subclinical.examination...sodium
subclinical.examination.....118
subclinical.examination.....119
subclinical.examination...potasium
subclinical.examination.....121
subclinical.examination...paco2.in.blood.air.
subclinical.examination.....130

subclinical.examination.....131
subclinical.examination...pa.Oxy..in.blood.air.
subclinical.examination.....135
subclinical.examination.....136
subclinical.examination...HCO3..in.blood.air.
...140
...141
BE..in.blood.air.
...145
...146
p.f..paO2..O2.
lactatr..in.blood.air.
...156
treatment...fluide.intake...159
treatment...fluide.intake...160
treatment...fluide.intake...161
treatment...fluide.output...162
treatment...fluide.output...163
treatment...fluide.output...164
treatment...balance.fluid.in.and.out...165
treatment...balance.fluid.in.and.out...166
treatment...balance.fluid.in.and.out...167
treatment...day.without.food.intake
treatment...PEX.treatment.of.which.day.of.the.diagnosis
treatment...number.of.PEX.treatment
treatment...triglycerid.before.first.time.of.PEX
treatment...triglycerid.after.first.time.of.PEX
treatment...APAche.2.sHavere.before.first.time.PEX
treatment...APAche.2.sHavere.after.first.time.PEX
treatment...Imre.sHavere.before.first.time.of.PEX
treatment...Imre.sHavere.after.first.time.of.PEX
Result...dead.or.alive
Patient.with.PEX.or.without.PEX
Blood.pressure.systolic
Blood.pressure.diastolic
Bilirubin_direct_t0
Bilirubin_indirect_t0
Bilirubin_direct_t30
Bilirubin_indirect_t30

```
Bilirubin_direct_t72
Bilirubin_indirect_t72
AST..liver.funtion.
ALT..liver.funtion.
subclinical.examination...calci_serum
```

Based on our imputation, we perform **Feature selection with ExtraTree classifier** on Missforest dataset, using

Using Result: Dead or Alive as response variable.

```
# printing the number of total and selected features

selected_feat2 = X_train.columns[(feature_sel_model.get_support())]

print('Dataset: MISS FOREST Imputed')
print()
print('Total features:{}'.format((X_train.shape[1])))
print('Selected Features:{}'.format(len(selected_feat2)))
```

```
Dataset: MISS FOREST Imputed

Total features:130
Selected Features:51
```

Based on the Extra Tree classifier algorithm, it is clear that only 51 significant features have been selected out of 130 for the response: "Result: Dead or Alive".

The 51 Significant features for response "Result: Dead or Alive" using "ExtraTree Classifier" on "Miss Forest" imputed dataset are as following:

```
Gender
Drinking.problem
A.breakdown.of.drinking.problem...13
Diabetes.problem
Vomitting
Clinical.symptoms...25
apache.2.score.at.the.points.of.admitting.hospitals
subclinical.examination...balthazar.sHavere..with.computer.tomography.
subclinical.examination...CTSI.score..with.computer.tomography.
```

subclinical.examination...white.blood.cell..t0..at.the.points.of.admitting.hospitals..t6..after.6h.of.admitting.hospitals...
subclinical.examination.....49
subclinical.examination.....55
subclinical.examination.....59
subclinical.examination.....64
prothrombin
subclinical.examination.....67
APTT
subclinical.examination.....70
subclinical.examination.....71
subclinical.examination.....72
subclinical.examination...Fibrinogen
subclinical.examination.....76
subclinical.examination.....78
subclinical.examination...creatinin
subclinical.examination.....83
subclinical.examination.....84
subclinical.examination.....103
subclinical.examination...protein
subclinical.examination...albumin
subclinical.examination...potasium
subclinical.examination.....121
subclinical.examination.....122
subclinical.examination.....126
subclinical.examination...pa.Oxy..in.blood.air.
subclinical.examination.....135
subclinical.examination...HCO3..in.blood.air.
...145
...146
p.f..paO2..O2.
lactatr..in.blood.air.
...156
treatment...fluide.intake...160
treatment...fluide.intake...161
treatment...fluide.output...164
treatment...APAche.2.sHavere.before.first.time.PEX
treatment...APAche.2.sHavere.after.first.time.PEX
Patient.with.PEX.or.without.PEX

Blood.pressure.systolic

Bilirubin_direct_t0

Bilirubin_indirect_t0

Bilirubin_indirect_t30

Linear regression model is built based on

Mice Imputed dataset:

```
model_mice = LinearRegression(n_jobs=-1)
model_mice.fit(X_train[selected_feat_mice],y_train)
```

```
LinearRegression(n_jobs=-1)
```

```
y_predict = model_mice.predict(X_test[selected_feat_mice])
print('R2 Score:', r2_score(y_test, y_predict))
```

```
R2 Score: -4.4752202159160515
```

Amelia Imputed dataset :

```
model_amelia = LinearRegression(n_jobs=-1)
model_amelia.fit(X_train[selected_feat],y_train)
```

```
LinearRegression(n_jobs=-1)
```

```
y_predict = model_amelia.predict(X_test[selected_feat])
print('R2 Score:', r2_score(y_test, y_predict))
```

```
R2 Score: -4.052677232836444
```

Missforest Imputed dataset :

```
model_missforest = LinearRegression(n_jobs=-1)
model_missforest.fit(X_train[selected_feat],y_train)
```

```
LinearRegression(n_jobs=-1)
```

```
y_predict = model_amelia.predict(X_test[selected_feat])
print('R2 Score:', r2_score(y_test, y_predict))
```

```
R2 Score: -4.223664255885263
```

Conclusions:

- The regression analysis shows the results are poor than the mean model as the R2 score obtained is negative.
- The best fit line of this model is found to be performing less than the horizontal fitted line for the respective models. This pattern is often observed when dimensionality of data set is large and has less no of observations to fit the model.
- The models also need appropriate treatment of variables like standardization, log transformation or normalization. With these transformations, we can hope to get a better fit model
- However, out of 3 datasets, we found better R2 score of amelia with R2 = - 4.05 and R2 of mice is -4.47 The working of the algorithms backing the imputation methods play a major role in obtaining cleaner dataset.
- Amelia uses expectation maximization algorithm which is based on maximum likelihood estimate. These algorithms perform better than the regression algorithms like pmm, regression, poly regression and classification algorithm which back mice imputation.
- Missforest is backed by random forest algorithm which uses decision trees to get imputations. the bagging technique used in dec trees also produces better results in model than Mice imputation, as we can see with R2 score of -4.22