

Data Set Location (URL): <https://data.medicare.gov/data/dialysis-facility-compare>

n: 6404

p: ~7

Response Variable: Standardized Mortality Ratio (SMR)

Predictor Variables:

SMR..upper.confidence.limit..95..

Number.of.patients.included.in.survival.summary

Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.

Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.

Lists.the.facility.s.standardized.transfusion.ratio..facility.

Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.

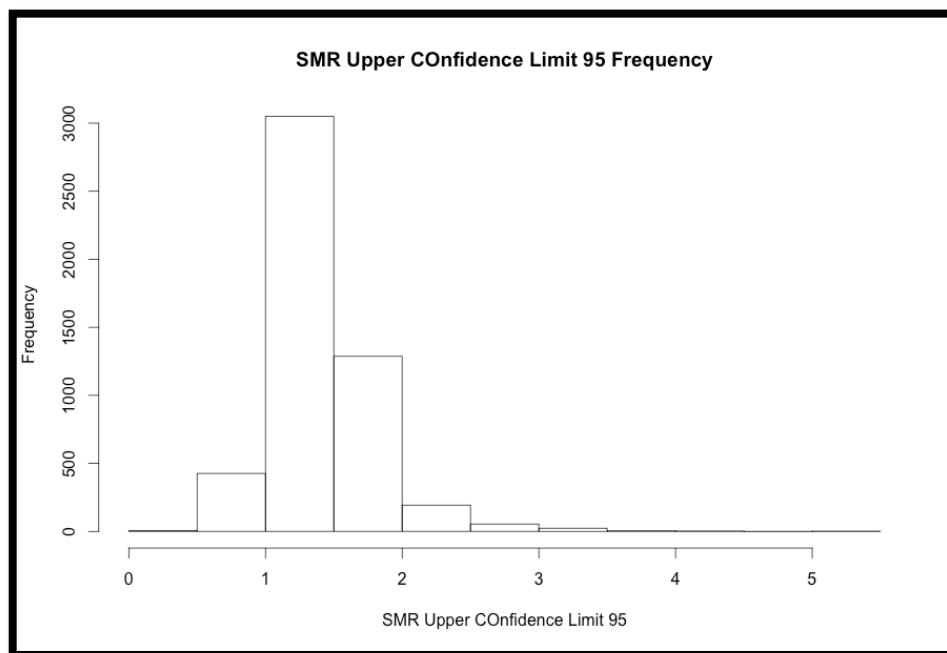
Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.

Descriptive analysis of the predictors

- SMR..upper.confidence.limit..95..

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
0.120	1.170	1.350	1.406	1.560	5.420

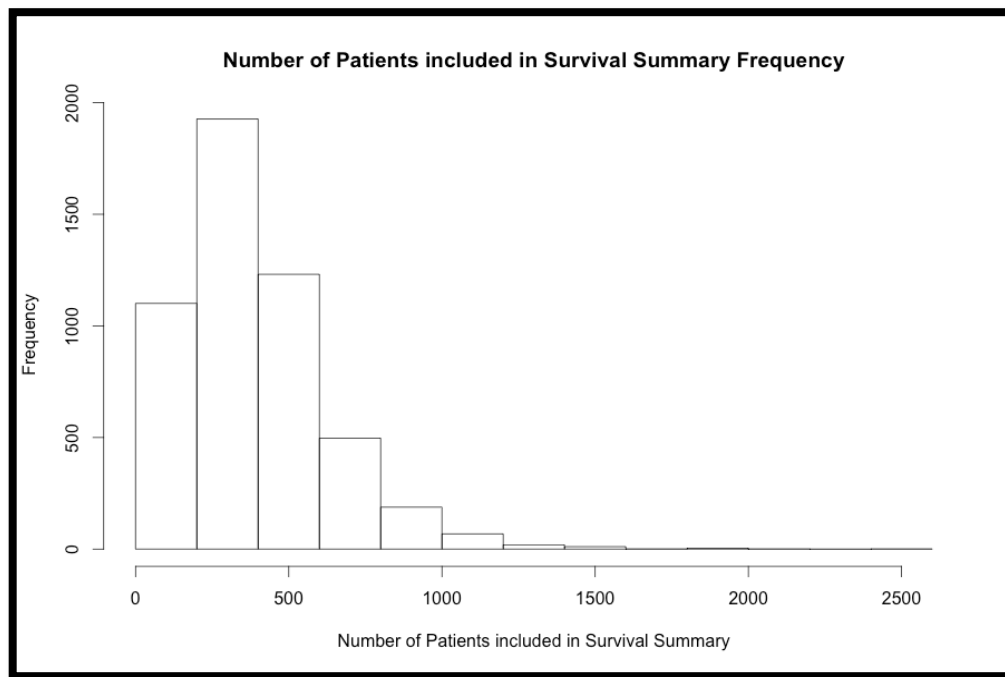
Standard Deviation: 0.3788001



- Number.of.patients.included.in.survival.summary

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
30.0	217.0	343.0	388.6	505.0	2461.0

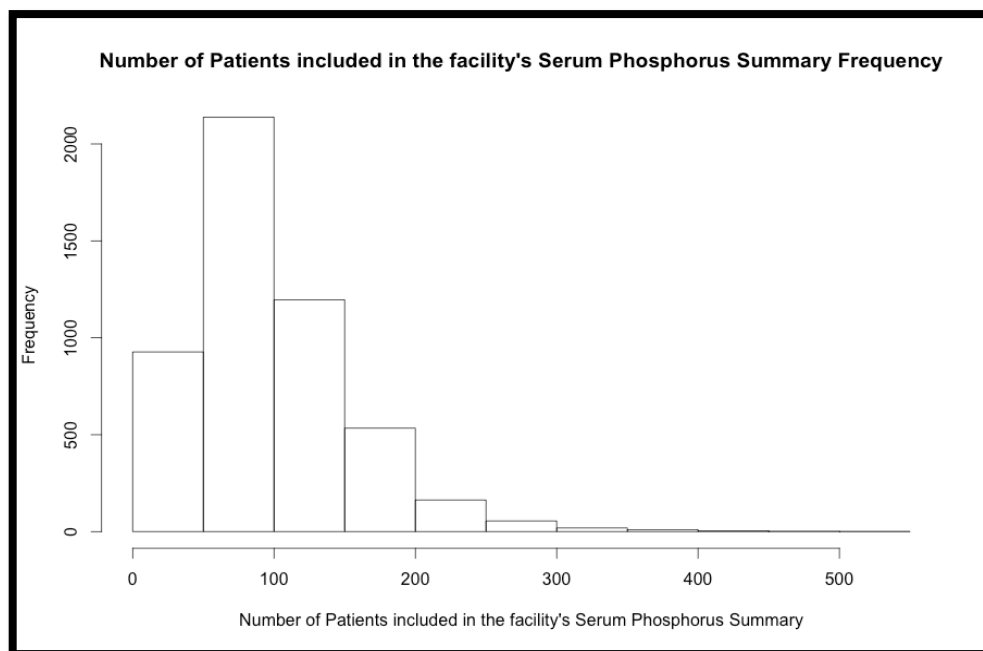
Standard Deviation: 233.9469



- Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
12.00	59.00	87.00	98.25	125.00	501.00

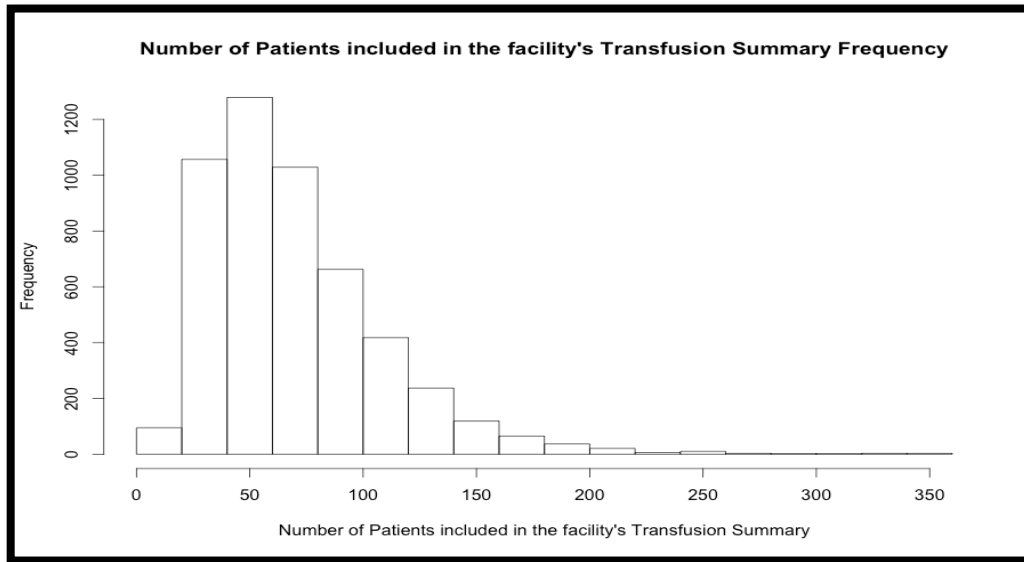
Standard Deviation: 55.04442



- Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..fa
cility.

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
14.00	42.00	62.00	70.31	89.00	349.00

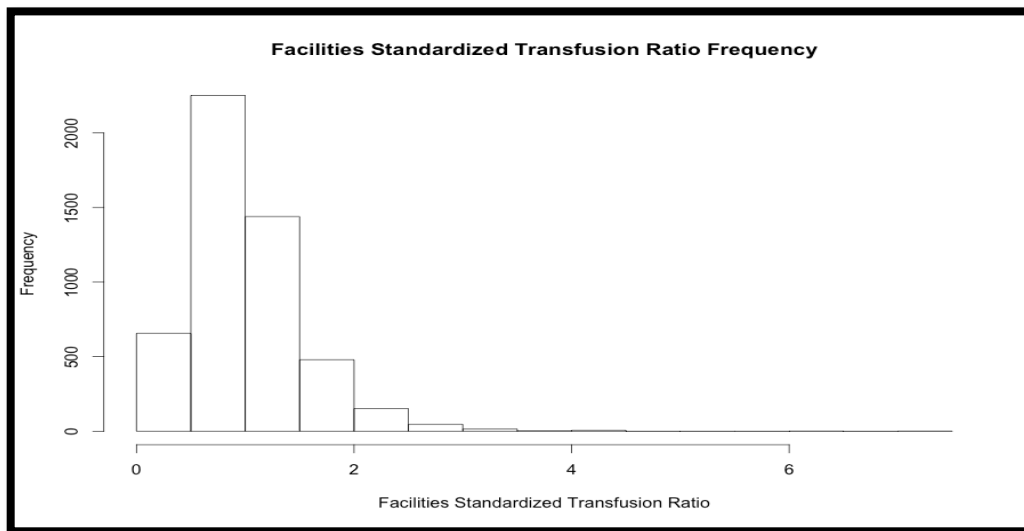
Standard Deviation: 39.38742



- Lists.the.facility.s.standardized.transfusion.ratio..facility.

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
0.000	0.650	0.910	1.002	1.250	7.150

Standard Deviation: 0.5226825

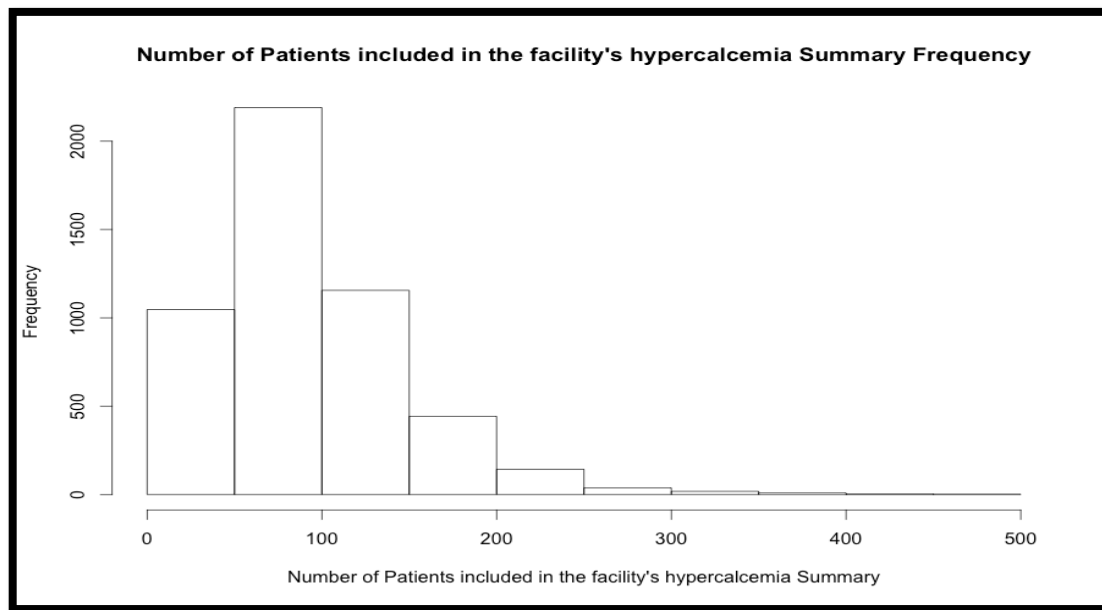


- Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary
..facility.

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
------	---------	--------	------	---------------------	-----

15.00	55.00	82.00	93.38	119.00	488.00
-------	-------	-------	-------	--------	--------

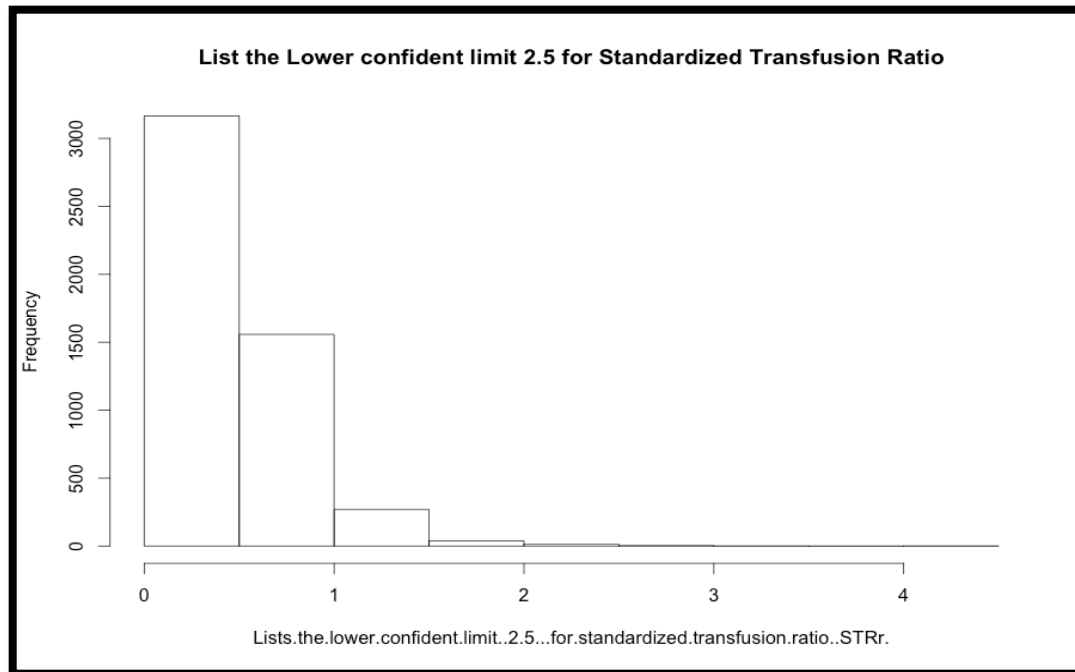
Standard Deviation: 53.14374



- Lists the lower confident limit..2.5...for standardized transfusion ratio..STRr.

Min.	1sr Qu.	Median	Mean	3 rd Qu.	Max
0.0000	0.2300	0.4100	0.4675	0.6200	4.1400

Standard Deviation: 0.3320067



Analysis Plan:

We cleaned our dataset by removing observations that were inconsistent. By inconsistent we mean the observations where not enough data were able to be collected for some facilities. After removing those observations, we proceeded by removing observations that had missing values. After our dataset was clean and reviewed we performed a forward, backward and stepwise variables selection that gave us around 30 variables that were significant to build a multiple linear regression model.

As we had a clean dataset and the list of the significant variables, we continued to build our regression model.

Correlation Matrices and Pair Plots

	SMR..upper.confidence.limit..95.
SMR..upper.confidence.limit..95..	1.0000000
Number.of.patients.included.in.survival.summary	-0.3274663
Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.	-0.2902843
Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.	-0.3025692
Lists.the.facility.s.standardized.transfusion.ratio..facility.	0.2236860
Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.	-0.2912476
Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.	0.1075098

Number.of.patients.included.in.survival.summary
 SMR..upper.confidence.limit..95..
 -0.32746631
 Number.of.patients.included.in.survival.summary
 1.00000000
 Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
 0.91588155
 Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
 0.93533456
 Lists.the.facility.s.standardized.transfusion.ratio..facility.
 0.04466869
 Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
 0.93093565
 Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
 0.18314514

Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
 SMR..upper.confidence.limit..95..
 -0.29028434
 Number.of.patients.included.in.survival.summary
 0.91588155
 Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
 1.00000000
 Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
 0.93414316
 Lists.the.facility.s.standardized.transfusion.ratio..facility.
 -0.04865927
 Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
 0.98977235
 Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
 0.17614319

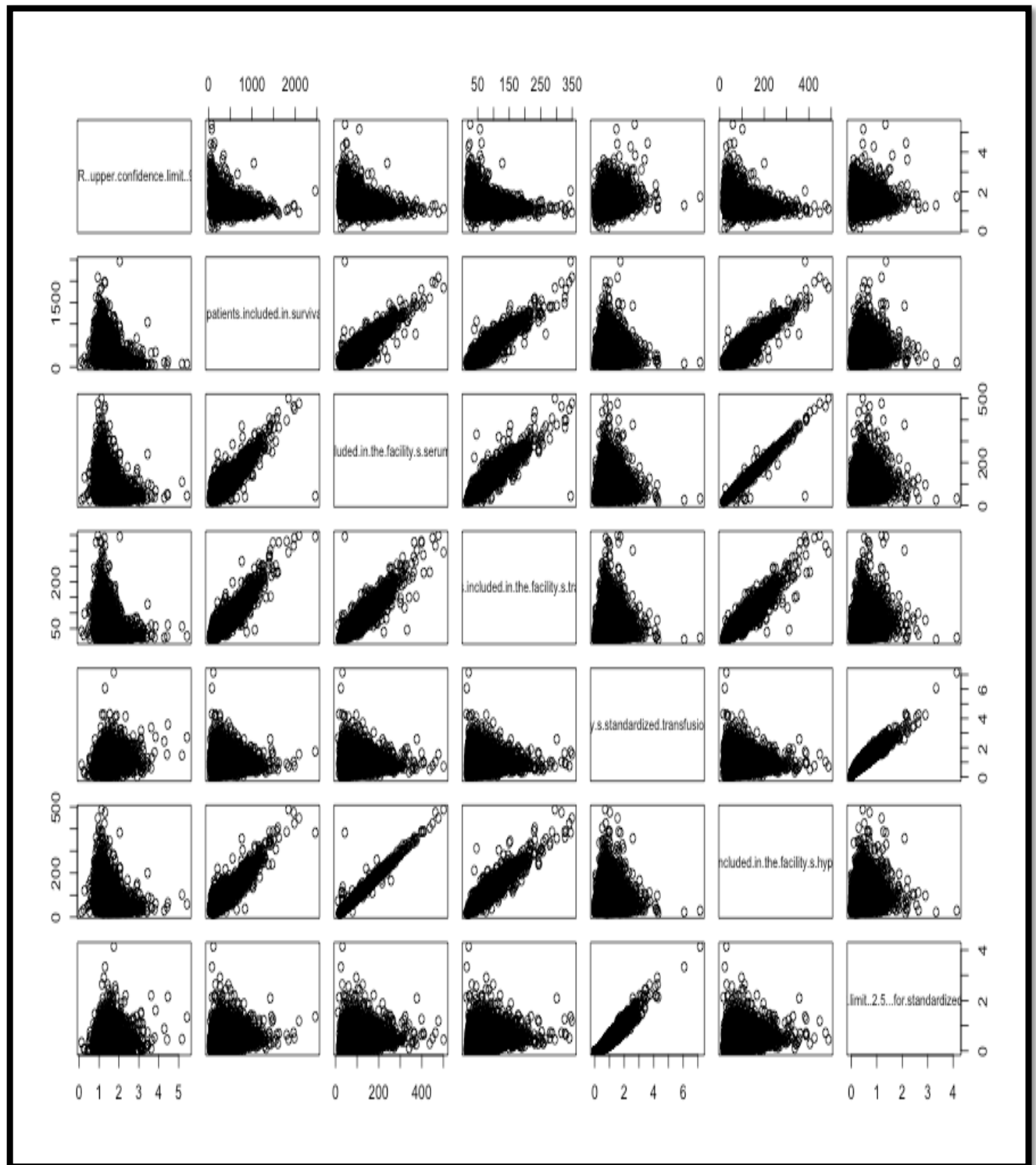
Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
 SMR..upper.confidence.limit..95..
 -0.30256919
 Number.of.patients.included.in.survival.summary
 0.93533456
 Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
 0.93414316
 Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
 1.00000000
 Lists.the.facility.s.standardized.transfusion.ratio..facility.
 -0.03342247
 Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
 0.94664248
 Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
 0.21075855

Lists.the.facility.s.standardized.transfusion.ratio..facility.
SMR..upper.confidence.limit..95..
0.22368595
Number.of.patients.included.in.survival.summary
-0.04466869
Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
-0.04865927
Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
-0.03342247
Lists.the.facility.s.standardized.transfusion.ratio..facility.
1.00000000
Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
-0.04902308
Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
0.94959074

Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
SMR..upper.confidence.limit..95..
-0.29124755
Number.of.patients.included.in.survival.summary
0.93093565
Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
0.98977235
Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
0.94664248
Lists.the.facility.s.standardized.transfusion.ratio..facility.
-0.04902308
Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
1.00000000
Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
0.17793688

Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
SMR..upper.confidence.limit..95..
0.1075098
Number.of.patients.included.in.survival.summary
0.1831451
Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.
0.1761432
Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.
0.2107586
Lists.the.facility.s.standardized.transfusion.ratio..facility.
0.9495907
Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.
0.1779369
Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.
1.0000000

Pairs Plots



Multiple Linear Regression:

```
lm.fit = lm(Standardized.mortality.ratio ~  
  SMR..upper.confidence.limit..95.. +Number.of.patients.included.in.survival.summary +  
  Lists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility. +  
  Lists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility. +  
  Number.of.patients.included.in.hospitalization.summary+  
  Lists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STRr.+  
  Lists.the.facility.s.standardized.transfusion.ratio..facility. +  
  Lists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.  
  , data = CleanData)  
  
summary(lm.fit)  
  
(see below)
```

coefficients:	Estimate	Std. Error	t value	Pr(> t)
Intercept)	2.35E-02	8.29E-03	2.829	0.00468
.MR..upper.confidence.limit..95..	6.88E-01	3.87E-03	177.731	< 2e-16
number.of.patients.included.in.survival.summary	7.56E-04	1.77E-05	42.608	< 2e-16
ists.the.number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.	1.06E-03	1.70E-04	6.236	4.86E-10
ists.the.number.of.patients.included.in.the.facility.s.transfusion.summary..facility.	-1.39E-03	1.28E-04	-10.92	< 2e-16
ists.the.lower.confident.limit..2.5...for.standardized.transfusion.ratio..STR.	4.12E-01	2.06E-02	20.043	< 2e-16
ists.the.facility.s.standardized.transfusion.ratio..facility.	-2.26E-01	1.29E-02	-17.585	< 2e-16
ists.the.number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.	-2.54E-03	2.01E-04	-12.606	< 2e-16

Residual standard error: 0.0943 on 5041 degrees of freedom

Multiple R-squared: 0.8731, Adjusted R-squared: 0.8729

F-statistic: 4955 on 7 and 5041 DF, p-value: < 2.2e-16

According to the linear regression model:

$$\begin{aligned} \text{Estimated Mortality ratio} = & 2.345e-02 + 6.884e-01(\text{SMR..upper.confidence.limit..95..}) + 7.556e-04(\text{Number.of.patients.included.in.survival.summary}) + \\ & 1.061e-03(\text{Lists the number.of.patients.included.in.the.facility.s.serum.phosphorus.summary..facility.}) - \\ & 1.392e-03(\text{Lists the number.of.patients.included.in.the.facility.s.transfusion.summary..facility.}) + \\ & 4.119e-01(\text{Lists the lower.confident.limit..2.5...for.standardized.transfusion.ratio..STR.}) - \\ & 2.262e-01(\text{Lists the facility.s.standardized.transfusion.ratio..facility.}) \\ & -2.535e-03(\text{Lists the number.of.patients.included.in.the.facility.s.hypercalcemia.summary..facility.}) \end{aligned}$$

Check the estimated value:

$$\text{predicted_2} <- 2.345e-02 + (6.884e-01*1.38) + (7.556e-04*416) + (1.061e-03*94) + (-1.392e-03*75) + (4.119e-01*0.39) + (-2.262e-01*0.83) + (-2.535e-03*88)$$

$$\text{Predicted_2} = 1.032921$$

$$\text{Actual Value} = 1.08$$

The predicted model is quite close the actual values of mortality ratio.

Plots of the Linear Model:

