final report

2023-12-19

Abstract

Introduction

Data and Methods

Descriptive Data

Data Cleaning and Preparation

The dataset utilized in this study was derived from a comprehensive breast cancer database. The initial step in data preparation involved the standardization of variable names to ensure consistency. In addition, we converted several categorical variables into factors with defined levels. Specifically, we recoded the survival status variable into a binary format with `Dead` as 1 and `Alive` as 0. Finally, a new variable `node_positive_prop` was created and calculated based on the ratio of `reginol_node_positive` to `regional_node_examined`. This variable represents the proportion of examined nodes that were found to be positive.

Variable and Model Selection Procedures

After preprocessing the dataset, we subdivided our dataset into categorical and numerical variables in order to have a general outline data patterns. In **Table 1**, we summarized the essential statistics of all the categorical variables, which includes variable names, number of missing values, unique and top counts.

For numerical variables, we employed boxplot visualizations to effectively represent their distribution patterns. As illustrated in **Figure 1**, these boxplots serve as a comprehensive visualization, which include potential outliers, important quartiles, and medians of all the numerical variables.

Initially, a comprehensive model incorporating all available predictors was developed, targeting survival status as the response variable. The corresponding estimates, standard errors, and P-values are illustrated in details in **Table 2**. After gaining an initial understanding of the data trends, we employed both stepwise selection and regularization techniques like LASSO and Ridge Regression for selecting the most appropriate model.

In addition, we conducted a series of diagnostic evaluations on this full model. Our first step in this process was to assess multicollinearity, the results of which are presented in the Variance Inflation Factor (VIF) table depicted in **Table 3**. After the implementation of stepwise selection methods, along with LASSO and Ridge Regression, we proceeded to evaluate the classification accuracy of all the models. This was achieved through the generation of Receiver Operating Characteristic (ROC) curves and the analysis of the Area Under the Curve (AUC) statistics. The summary of all AUC statistics are included in **Table 4**. In addition, all ROC curves are illustrated with **Figure 2**, **Figure 3**, **Figure 4**, and **Figure 5**.

Upon completing the process of variable selection, our analysis has yielded four distinct models, each characterized by a unique set of predictor variables. We used internal validation by splitting our dataset into traning and testing data since we couldn't collect new data to perform external validation. In addition, we used 10-fold validation to assess ability of shrinkage models(LASSO and Ridge Regression) to predict the testing data, and we referred to BIC and AIC scores when assessing the fit of backward and forward selection models.

Assumption Checking

Result

Variable Selection

Since our response variable `status` is a binary variable, we decided to implement a logistic regression in order to predict the risk of death based on variables selected. We selected our model variables in four distinct manners: two automatic approaches (both backward and forward selections) and two shrinkage methods(LASSO as well as Ridge Regression). The final models including all the variable selected by different methods are illustrated in **Table 4**, **Table 5**, **Table 6**, and **Table 7** respectively.

Model Selection

First, we closely examined the performance metrics of both backward and forward selection models, with particular emphasis on their AIC and BIC scores, as presented in **Table 8**. Our findings indicate a noticeable distinction between the two models. Specifically, the backward selection model demonstrated superior performance, if we solely assessed based on its lower AIC and BIC scores in comparison to the forward selection model.

To further validate these findings, we employed 10-fold cross-validation techniques on the two automatic research approaches. This procedure yielded insightful results regarding the accuracy of each model. The backward selection model exhibited a high degree of accuracy, achieving a score of 0.858854. The forward selection model, while slightly less accurate, still maintained a high accuracy level with a score of 0.8561146.

In addition, we implemented 10-fold validation on LASSO and Ridge Regression. The results are shown in

Conclusion

Appendix

Table

Table 1: Summary Statistics of Categorical Variables

Variable	Missing	Unique Counts	Top Counts
race	0	3	Whi: 3413, Oth: 320, Bla: 291
marital_status	0	5	Mar: 2643, Sin: 615, Div: 486, Wid: 235
t_stage	0	4	T2: 1786, T1: 1603, T3: 533, T4: 102
n_stage	0	3	N1: 2732, N2: 820, N3: 472
x6th_stage	0	5	IIA: 1305, IIB: 1130, III: 1050, III: 472
differentiate	0	4	Mod: 2351, Poo: 1111, Wel: 543, Und: 19
grade	0	4	2: 2351, 3: 1111, 1: 543, ana: 19
a_stage	0	2	Reg: 3932, Dis: 92
estrogen_status	0	2	Pos: 3755, Neg: 269
progesterone_status	0	2	Pos: 3326, Neg: 698
status	0	2	0: 3408, 1: 616
node_positive_prop	0	NA	NA

Table 2: Full Model Summary

Term	Estimate	Standard Error	P Value
(Intercept)	-3.926	0.461	0.000
age	0.024	0.006	0.000
raceBlack	0.515	0.162	0.002
raceOther	-0.416	0.203	0.040
marital_statusMarried	-0.132	0.135	0.327
marital_statusDivorced	0.082	0.175	0.641
marital_statusSeparated	0.721	0.383	0.060
marital_statusWidowed	0.098	0.219	0.653
t_stageT2	0.279	0.195	0.153
t_stageT3	0.542	0.314	0.084
t_stageT4	0.949	0.450	0.035
n_stageN2	0.562	0.241	0.020
n_stageN3	0.586	0.305	0.055
x6th_stageIIB	0.216	0.232	0.352
x6th_stageIIIA	-0.101	0.295	0.733
x6th_stageIIIB	0.053	0.529	0.921
x6th_stageIIIC	NA	NA	NA
differentiatePoorly differentiated	0.391	0.105	0.000
differentiateUndifferentiated	1.364	0.535	0.011
differentiateWell differentiated	-0.533	0.184	0.004
grade2	NA	NA	NA
grade3	NA	NA	NA
gradeanaplastic; Grade IV	NA	NA	NA
a_stageRegional	-0.060	0.266	0.821
tumor_size	0.000	0.004	0.992
estrogen_statusNegative	0.737	0.178	0.000
progesterone_statusNegative	0.589	0.128	0.000
regional_node_examined	-0.021	0.011	0.053
reginol_node_positive	0.055	0.020	0.007
node_positive_prop	0.590	0.316	0.062

Table 3: VIF for Full Model

Term	VIF	CI_low	CI_high	SE_factor	Tolerance	Tolerance_low	Tolerance_high
age	1.1	1.1	1.2	1.1	0.9	0.9	0.9
race	1.1	1.0	1.1	1.0	0.9	0.9	1.0
marital_status	1.1	1.1	1.2	1.1	0.9	0.8	0.9
t_stage	30.6	28.8	32.5	5.5	0.0	0.0	0.0
n_stage	31.8	30.0	33.8	5.6	0.0	0.0	0.0
x6th_stage	61.7	58.0	65.5	7.9	0.0	0.0	0.0
differentiate	1.1	1.1	1.2	1.1	0.9	0.9	0.9
a_stage	1.3	1.2	1.3	1.1	0.8	0.8	0.8
tumor_size	3.7	3.5	3.9	1.9	0.3	0.3	0.3
estrogen_status	1.5	1.4	1.5	1.2	0.7	0.6	0.7
progesterone_status	1.4	1.4	1.5	1.2	0.7	0.7	0.7
regional_node_examined	3.4	3.3	3.6	1.9	0.3	0.3	0.3
reginol_node_positive	7.3	6.9	7.8	2.7	0.1	0.1	0.1
node_positive_prop	4.4	4.2	4.7	2.1	0.2	0.2	0.2

Table 4: Backward Model Summary

Term	Estimate	Standard Error	P Value
(Intercept)	-4.043	0.364	0.000
age	0.024	0.005	0.000
raceBlack	0.571	0.159	0.000
raceOther	-0.436	0.202	0.031
t_stageT2	0.415	0.113	0.000
t_stageT3	0.537	0.149	0.000
t_stageT4	1.081	0.243	0.000
n_stageN2	0.359	0.133	0.007
n_stageN3	0.483	0.239	0.043
differentiatePoorly differentiated	0.390	0.105	0.000
differentiateUndifferentiated	1.343	0.527	0.011
differentiateWell differentiated	-0.514	0.183	0.005
estrogen_statusNegative	0.737	0.177	0.000
progesterone_statusNegative	0.598	0.127	0.000
regional_node_examined	-0.021	0.011	0.053
reginol_node_positive	0.056	0.020	0.005
node_positive_prop	0.603	0.314	0.054

Table 5: Forward Model Summary

Term	Estimate	Standard Error	P Value
(Intercept)	-3.872	0.379	0.000
n_stageN2	0.394	0.134	0.003
n_stageN3	0.529	0.240	0.028
progesterone_statusNegative	0.575	0.128	0.000
differentiatePoorly differentiated	0.407	0.105	0.000
differentiateUndifferentiated	1.347	0.525	0.010
differentiateWell differentiated	-0.539	0.184	0.003
node_positive_prop	0.611	0.313	0.051
raceBlack	0.502	0.162	0.002
raceOther	-0.430	0.202	0.034
age	0.023	0.006	0.000
estrogen_statusNegative	0.749	0.178	0.000
tumor_size	0.007	0.002	0.001
reginol_node_positive	0.054	0.020	0.007
regional_node_examined	-0.020	0.011	0.064
marital_statusMarried	-0.153	0.134	0.254
marital_statusDivorced	0.072	0.175	0.679
marital_statusSeparated	0.712	0.380	0.061
marital_statusWidowed	0.071	0.218	0.744

Table 6: Lasso Model Summary

Term	Estimate	Lambda	Deviation Ratio
(Intercept)	-1.709	0.094	0.000
(Intercept)	-1.760	0.086	0.013
(Intercept)	-1.804	0.078	0.023

Term	Estimate	Lambda	Deviation Ratio
(Intercept)	-1.849	0.071	0.031
(Intercept)	-1.905	0.065	0.039
(Intercept)	-1.956	0.059	0.045
(Intercept)	-2.003	0.054	0.050
(Intercept)	-2.046	0.049	0.055
(Intercept)	-2.086	0.045	0.058
(Intercept)	-2.122	0.041	0.061
(Intercept)	-2.155	0.037	0.063
(Intercept)	-2.185	0.034	0.065
(Intercept)	-2.213	0.031	0.067
(Intercept)	-2.239	0.028	0.068
(Intercept)	-2.270	0.026	0.070
(Intercept)	-2.309	0.023	0.071
(Intercept)	-2.343	0.021	0.073
(Intercept)	-2.375	0.019	0.074
(Intercept)	-2.404	0.018	0.074
(Intercept)	-2.431	0.016	0.075
(Intercept)	-2.455	0.015	0.076
(Intercept)	-2.525	0.013	0.077
(Intercept)	-2.609	0.012	0.077
(Intercept)	-2.686	0.011	0.078
(Intercept)	-2.756	0.010	0.078
(Intercept)	-2.821	0.009	0.079
(Intercept)	-2.880	0.008	0.079
(Intercept)	-2.933	0.008	0.080
(Intercept)	-2.983	0.007	0.080
(Intercept)	-3.028	0.006	0.080
(Intercept)	-3.069	0.006	0.080
(Intercept)	-3.107	0.005	0.080
(Intercept)	-3.141	0.005	0.080
(Intercept)	-3.173	0.004	0.081
(Intercept)	-3.202	0.004	0.081
(Intercept)	-3.228	0.004	0.081
(Intercept)	-3.252	0.003	0.081
(Intercept)	-3.274	0.003	0.081
(Intercept)	-3.294	0.003	0.081
(Intercept)	-3.312	0.003	0.081
(Intercept)	-3.329	0.002	0.081
(Intercept)	-3.344	0.002	0.081
(Intercept)	-3.358	0.002	0.081
(Intercept)	-3.371	0.002	0.081
(Intercept)	-3.382	0.002	0.081
(Intercept)	-3.393	0.001	0.081
(Intercept)	-3.402	0.001	0.081
age	0.001	0.013	0.077
age	0.002	0.012	0.077
age	0.003	0.011	0.078
age	0.004	0.010	0.078
age	0.005	0.009	0.079
age	0.006	0.008	0.079
age	0.007	0.008	0.080
age	0.007	0.007	0.080

Term	Estimate	Lambda	Deviation Ratio
age	0.008	0.006	0.080
age	0.008	0.006	0.080
age	0.009	0.005	0.080
age	0.009	0.005	0.080
age	0.010	0.004	0.081
age	0.010	0.004	0.081
age	0.011	0.004	0.081
age	0.011	0.003	0.081
age	0.011	0.003	0.081
age	0.011	0.003	0.081
age	0.012	0.003	0.081
age	0.012	0.002	0.081
age	0.012	0.002	0.081
age	0.012	0.002	0.081
age	0.013	0.002	0.081
age	0.013	0.002	0.081
age	0.013	0.001	0.081
age	0.013	0.001	0.081
tumor_size	0.000	0.026	0.070
tumor_size	0.001	0.023	0.071
tumor_size	0.002	0.021	0.073
tumor_size	0.002	0.019	0.074
tumor_size	0.003	0.018	0.074
tumor_size	0.003	0.016	0.075
tumor_size	0.003	0.015	0.076
tumor_size	0.004	0.013	0.077
tumor_size	0.004	0.012	0.077
tumor_size	0.004	0.011	0.078
tumor_size	0.005	0.010	0.078
tumor_size	0.005	0.009	0.079
tumor_size	0.005	0.008	0.079
tumor_size	0.005	0.008	0.080
tumor_size	0.006	0.007	0.080
tumor_size	0.006	0.006	0.080
tumor_size	0.006	0.006	0.080
tumor_size	0.006	0.005	0.080
tumor_size	0.006	0.005	0.080
tumor_size	0.007	0.004	0.081
tumor_size	0.007	0.004	0.081
tumor_size	0.007	0.004	0.081
tumor_size	0.007	0.003	0.081
tumor_size	0.007	0.003	0.081
tumor_size	0.007	0.003	0.081
tumor_size	0.007	0.003	0.081
tumor_size	0.007	0.002	0.081
tumor_size	0.007	0.002	0.081
tumor_size	0.007	0.002	0.081
tumor_size	0.007	0.002	0.081
tumor_size	0.007	0.002	0.081
tumor_size	0.007	0.001	0.081
tumor_size	0.007	0.001	0.081
reginol_node_positive	0.012	0.086	0.013

Term	Estimate	Lambda	Deviation Ratio
reginol_node_positive	0.022	0.078	0.023
reginol_node_positive	0.029	0.071	0.031
reginol_node_positive	0.033	0.065	0.039
reginol_node_positive	0.035	0.059	0.045
reginol_node_positive	0.038	0.054	0.050
reginol_node_positive	0.040	0.049	0.055
reginol_node_positive	0.042	0.045	0.058
reginol_node_positive	0.044	0.041	0.061
reginol_node_positive	0.046	0.037	0.063
reginol_node_positive	0.048	0.034	0.065
reginol_node_positive	0.049	0.031	0.067
reginol_node_positive	0.050	0.028	0.068
reginol_node_positive	0.051	0.026	0.070
reginol_node_positive	0.052	0.023	0.071
reginol_node_positive	0.053	0.021	0.073
reginol_node_positive	0.053	0.019	0.074
reginol_node_positive	0.054	0.018	0.074
reginol_node_positive	0.054	0.016	0.075
reginol_node_positive	0.055	0.015	0.076
reginol_node_positive	0.055	0.013	0.077
reginol_node_positive	0.056	0.012	0.077
reginol_node_positive	0.056	0.011	0.078
reginol_node_positive	0.056	0.010	0.078
reginol_node_positive	0.057	0.009	0.079
reginol_node_positive	0.057	0.008	0.079
reginol_node_positive	0.057	0.008	0.080
reginol_node_positive	0.058	0.007	0.080
reginol_node_positive	0.058	0.006	0.080
reginol_node_positive	0.058	0.006	0.080
reginol_node_positive	0.058	0.005	0.080
reginol_node_positive	0.059	0.005	0.080
reginol_node_positive	0.059	0.004	0.081
reginol_node_positive	0.059	0.004	0.081
reginol_node_positive	0.059	0.004	0.081
reginol_node_positive	0.059	0.003	0.081
reginol_node_positive	0.059	0.003	0.081
reginol_node_positive	0.059	0.003	0.081
reginol_node_positive	0.060	0.003	0.081
reginol_node_positive	0.060	0.002	0.081
reginol_node_positive	0.060	0.002	0.081
reginol_node_positive	0.060	0.002	0.081
reginol_node_positive	0.060	0.002	0.081
reginol_node_positive	0.060	0.002	0.081
reginol_node_positive	0.060	0.001	0.081
reginol_node_positive	0.060	0.001	0.081
node_positive_prop	0.027	0.071	0.031
node_positive_prop	0.139	0.065	0.039
node_positive_prop	0.239	0.059	0.045
node_positive_prop	0.330	0.054	0.050
node_positive_prop	0.412	0.049	0.055
node_positive_prop	0.486	0.045	0.058
node_positive_prop	0.553	0.041	0.061

Term	Estimate	Lambda	Deviation Ratio
node_positive_prop	0.615	0.037	0.063
node_positive_prop	0.670	0.034	0.065
node_positive_prop	0.721	0.031	0.067
node_positive_prop	0.767	0.028	0.068
node_positive_prop	0.806	0.026	0.070
node_positive_prop	0.839	0.023	0.071
node_positive_prop	0.869	0.021	0.073
node_positive_prop	0.896	0.019	0.074
node_positive_prop	0.921	0.018	0.074
node_positive_prop	0.944	0.016	0.075
node_positive_prop	0.965	0.015	0.076
node_positive_prop	0.982	0.013	0.077
node_positive_prop	0.997	0.012	0.077
node_positive_prop	1.011	0.011	0.078
node_positive_prop	1.024	0.010	0.078
node_positive_prop	1.035	0.009	0.079
node_positive_prop	1.046	0.008	0.079
node_positive_prop	1.055	0.008	0.080
node_positive_prop	1.064	0.007	0.080
node_positive_prop	1.072	0.006	0.080
node_positive_prop	1.080	0.006	0.080
node_positive_prop	1.086	0.005	0.080
node_positive_prop	1.092	0.005	0.080
node_positive_prop	1.098	0.004	0.081
node_positive_prop	1.103	0.004	0.081
node_positive_prop	1.108	0.004	0.081
node_positive_prop	1.112	0.003	0.081
node_positive_prop	1.116	0.003	0.081
node_positive_prop	1.119	0.003	0.081
node_positive_prop	1.123	0.003	0.081
node_positive_prop	1.126	0.002	0.081
node_positive_prop	1.128	0.002	0.081
node_positive_prop	1.131	0.002	0.081
node_positive_prop	1.133	0.002	0.081
node_positive_prop	1.135	0.002	0.081
node_positive_prop	1.137	0.001	0.081
node_positive_prop	1.139	0.001	0.081

Table 7: Ridge Regression Summary

Term	Estimate	Lambda	Deviation Ratio
(Intercept)	-1.709	94.435	0.000
(Intercept)	-1.713	86.045	0.001
(Intercept)	-1.713	78.401	0.001
(Intercept)	-1.714	71.436	0.001
(Intercept)	-1.714	65.090	0.001
(Intercept)	-1.715	59.308	0.001
(Intercept)	-1.715	54.039	0.001
(Intercept)	-1.716	49.238	0.001
(Intercept)	-1.717	44.864	0.001

Term	Estimate	Lambda	Deviation Ratio
(Intercept)	-1.718	40.879	0.001
(Intercept)	-1.718	37.247	0.001
(Intercept)	-1.719	33.938	0.001
(Intercept)	-1.720	30.923	0.001
(Intercept)	-1.721	28.176	0.002
(Intercept)	-1.723	25.673	0.002
(Intercept)	-1.724	23.392	0.002
(Intercept)	-1.725	21.314	0.002
(Intercept)	-1.727	19.421	0.002
(Intercept)	-1.729	17.695	0.002
(Intercept)	-1.731	16.123	0.003
(Intercept)	-1.733	14.691	0.003
(Intercept)	-1.735	13.386	0.003
(Intercept)	-1.738	12.197	0.003
(Intercept)	-1.740	11.113	0.004
(Intercept)	-1.743	10.126	0.004
(Intercept)	-1.747	9.226	0.005
(Intercept)	-1.750	8.407	0.005
(Intercept)	-1.754	7.660	0.005
(Intercept)	-1.758	6.979	0.006
(Intercept)	-1.763	6.359	0.007
(Intercept)	-1.768	5.794	0.007
(Intercept)	-1.774	5.280	0.008
(Intercept)	-1.780	4.811	0.008
(Intercept)	-1.786	4.383	0.009
(Intercept)	-1.794	3.994	0.010
(Intercept)	-1.802	3.639	0.011
(Intercept)	-1.810	3.316	0.012
(Intercept)	-1.819	3.021	0.013
(Intercept)	-1.829	2.753	0.014
(Intercept)	-1.840	2.508	0.015
(Intercept)	-1.852	2.285	0.016
(Intercept)	-1.865	2.082	0.018
(Intercept)	-1.879	1.897	0.019
(Intercept)	-1.894	1.729	0.021
(Intercept)	-1.910	1.575	0.022
(Intercept)	-1.927	1.435	0.024
(Intercept)	-1.945	1.308	0.026
(Intercept)	-1.965	1.192	0.028
(Intercept)	-1.986	1.086	0.030
(Intercept)	-2.008	0.989	0.032
(Intercept)	-2.032	0.901	0.034
(Intercept)	-2.057	0.821	0.036
(Intercept)	-2.084	0.748	0.038
(Intercept)	-2.112	0.682	0.040
(Intercept)	-2.141	0.621	0.043
(Intercept)	-2.171	0.566	0.045
(Intercept)	-2.203	0.516	0.047
(Intercept)	-2.236	0.470	0.049
(Intercept)	-2.271	0.428	0.052
(Intercept)	-2.306	0.390	0.054
(Intercept)	-2.342	0.356	0.056

Term	Estimate	Lambda	Deviation Ratio
(Intercept)	-2.379	0.324	0.058
(Intercept)	-2.416	0.295	0.060
(Intercept)	-2.454	0.269	0.062
(Intercept)	-2.493	0.245	0.064
(Intercept)	-2.532	0.223	0.065
(Intercept)	-2.571	0.203	0.067
(Intercept)	-2.610	0.185	0.068
(Intercept)	-2.649	0.169	0.070
(Intercept)	-2.688	0.154	0.071
(Intercept)	-2.726	0.140	0.072
(Intercept)	-2.764	0.128	0.073
(Intercept)	-2.801	0.116	0.074
(Intercept)	-2.837	0.106	0.075
(Intercept)	-2.873	0.097	0.076
(Intercept)	-2.907	0.088	0.076
(Intercept)	-2.941	0.080	0.077
(Intercept)	-2.973	0.073	0.077
(Intercept)	-3.005	0.067	0.078
(Intercept)	-3.035	0.061	0.078
(Intercept)	-3.064	0.055	0.079
(Intercept)	-3.091	0.050	0.079
(Intercept)	-3.118	0.046	0.079
(Intercept)	-3.143	0.042	0.080
(Intercept)	-3.166	0.038	0.080
(Intercept)	-3.189	0.035	0.080
(Intercept)	-3.210	0.032	0.080
(Intercept)	-3.230	0.029	0.080
(Intercept)	-3.248	0.026	0.080
(Intercept)	-3.266	0.024	0.080
(Intercept)	-3.282	0.022	0.081
(Intercept)	-3.297	0.020	0.081
(Intercept)	-3.311	0.018	0.081
(Intercept)	-3.324	0.017	0.081
(Intercept)	-3.336	0.015	0.081
(Intercept)	-3.347	0.014	0.081
(Intercept)	-3.358	0.012	0.081
(Intercept)	-3.367	0.011	0.081
(Intercept)	-3.376	0.010	0.081
(Intercept)	-3.384	0.009	0.081
age	0.000	94.435	0.000
age	0.000	86.045	0.001
age	0.000	78.401	0.001
age	0.000	71.436	0.001
age	0.000	65.090	0.001
age	0.000	59.308	0.001
age	0.000	54.039	0.001
age	0.000	49.238	0.001
age	0.000	44.864	0.001
age	0.000	40.879	0.001
age	0.000	37.247	0.001
age	0.000	33.938	0.001
age	0.000	30.923	0.001

Term	Estimate	Lambda	Deviation Ratio
age	0.000	28.176	0.002
age	0.000	25.673	0.002
age	0.000	23.392	0.002
age	0.000	21.314	0.002
age	0.000	19.421	0.002
age	0.000	17.695	0.002
age	0.000	16.123	0.003
age	0.000	14.691	0.003
age	0.000	13.386	0.003
age	0.000	12.197	0.003
age	0.000	11.113	0.004
age	0.000	10.126	0.004
age	0.000	9.226	0.005
age	0.000	8.407	0.005
age	0.000	7.660	0.005
age	0.000	6.979	0.006
age	0.000	6.359	0.007
age	0.000	5.794	0.007
age	0.000	5.280	0.008
age	0.000	4.811	0.008
age	0.000	4.383	0.009
age	0.000	3.994	0.010
age	0.000	3.639	0.011
age	0.000	3.316	0.012
age	0.001	3.021	0.013
age	0.001	2.753	0.014
age	0.001	2.508	0.015
age	0.001	2.285	0.016
age	0.001	2.082	0.018
age	0.001	1.897	0.019
age	0.001	1.729	0.021
age	0.001	1.575	0.022
age	0.001	1.435	0.024
age	0.001	1.308	0.026
age	0.001	1.192	0.028
age	0.001	1.086	0.030
age	0.001	0.989	0.032
age	0.002	0.901	0.034
age	0.002	0.821	0.036
age	0.002	0.748	0.038
age	0.002	0.682	0.040
age	0.002	0.621	0.043
age	0.002	0.566	0.045
age	0.003	0.516	0.047
age	0.003	0.470	0.049
age	0.003	0.428	0.052
age	0.003	0.390	0.054
age	0.003	0.356	0.056
age	0.004	0.324	0.058
age	0.004	0.295	0.060
age	0.004	0.269	0.062
age	0.004	0.245	0.064

Term	Estimate	Lambda	Deviation Ratio
age	0.005	0.223	0.065
age	0.005	0.203	0.067
age	0.005	0.185	0.068
age	0.006	0.169	0.070
age	0.006	0.154	0.071
age	0.006	0.140	0.072
age	0.007	0.128	0.073
age	0.007	0.116	0.074
age	0.007	0.106	0.075
age	0.008	0.097	0.076
age	0.008	0.088	0.076
age	0.008	0.080	0.077
age	0.009	0.073	0.077
age	0.009	0.067	0.078
age	0.009	0.061	0.078
age	0.009	0.055	0.079
age	0.010	0.050	0.079
age	0.010	0.046	0.079
age	0.010	0.042	0.080
age	0.011	0.038	0.080
age	0.011	0.035	0.080
age	0.011	0.032	0.080
age	0.011	0.029	0.080
age	0.011	0.026	0.080
age	0.012	0.024	0.080
age	0.012	0.022	0.081
age	0.012	0.020	0.081
age	0.012	0.018	0.081
age	0.012	0.017	0.081
age	0.013	0.015	0.081
age	0.013	0.014	0.081
age	0.013	0.012	0.081
age	0.013	0.011	0.081
age	0.013	0.010	0.081
age	0.013	0.009	0.081
tumor_size	0.000	94.435	0.000
tumor_size	0.000	86.045	0.001
tumor_size	0.000	78.401	0.001
tumor_size	0.000	71.436	0.001
tumor_size	0.000	65.090	0.001
tumor_size	0.000	59.308	0.001
tumor_size	0.000	54.039	0.001
tumor_size	0.000	49.238	0.001
tumor_size	0.000	44.864	0.001
tumor_size	0.000	40.879	0.001
tumor_size	0.000	37.247	0.001
tumor_size	0.000	33.938	0.001
tumor_size	0.000	30.923	0.001
tumor_size	0.000	28.176	0.002
tumor_size	0.000	25.673	0.002
tumor_size	0.000	23.392	0.002
tumor_size	0.000	21.314	0.002

Term	Estimate	Lambda	Deviation Ratio
tumor_size	0.000	19.421	0.002
tumor_size	0.000	17.695	0.002
tumor_size	0.000	16.123	0.003
tumor_size	0.000	14.691	0.003
tumor_size	0.000	13.386	0.003
tumor_size	0.000	12.197	0.003
tumor_size	0.000	11.113	0.004
tumor_size	0.000	10.126	0.004
tumor_size	0.000	9.226	0.005
tumor_size	0.000	8.407	0.005
tumor_size	0.000	7.660	0.005
tumor_size	0.000	6.979	0.006
tumor_size	0.000	6.359	0.007
tumor_size	0.000	5.794	0.007
tumor_size	0.000	5.280	0.008
tumor_size	0.000	4.811	0.008
tumor_size	0.000	4.383	0.009
tumor_size	0.000	3.994	0.010
tumor_size	0.001	3.639	0.011
tumor_size	0.001	3.316	0.012
tumor_size	0.001	3.021	0.013
tumor_size	0.001	2.753	0.014
tumor_size	0.001	2.508	0.015
tumor_size	0.001	2.285	0.016
tumor_size	0.001	2.082	0.018
tumor_size	0.001	1.897	0.019
tumor_size	0.001	1.729	0.021
tumor_size	0.001	1.575	0.022
tumor_size	0.001	1.435	0.024
tumor_size	0.001	1.308	0.026
tumor_size	0.001	1.192	0.028
tumor_size	0.001	1.086	0.030
tumor_size	0.002	0.989	0.032
tumor_size	0.002	0.901	0.034
tumor_size	0.002	0.821	0.036
tumor_size	0.002	0.748	0.038
tumor_size	0.002	0.682	0.040
tumor_size	0.002	0.621	0.043
tumor_size	0.002	0.566	0.045
tumor_size	0.003	0.516	0.047
tumor_size	0.003	0.470	0.049
tumor_size	0.003	0.428	0.052
tumor_size	0.003	0.390	0.054
tumor_size	0.003	0.356	0.056
tumor_size	0.003	0.324	0.058
tumor_size	0.004	0.295	0.060
tumor_size	0.004	0.269	0.062
tumor_size	0.004	0.245	0.064
tumor_size	0.004	0.223	0.065
tumor_size	0.004	0.203	0.067
tumor_size	0.004	0.185	0.068
tumor_size	0.005	0.169	0.070

Term	Estimate	Lambda	Deviation Ratio
tumor_size	0.005	0.154	0.071
tumor_size	0.005	0.140	0.072
tumor_size	0.005	0.128	0.073
tumor_size	0.005	0.116	0.074
tumor_size	0.005	0.106	0.075
tumor_size	0.005	0.097	0.076
tumor_size	0.006	0.088	0.076
tumor_size	0.006	0.080	0.077
tumor_size	0.006	0.073	0.077
tumor_size	0.006	0.067	0.078
tumor_size	0.006	0.061	0.078
tumor_size	0.006	0.055	0.079
tumor_size	0.006	0.050	0.079
tumor_size	0.006	0.046	0.079
tumor_size	0.007	0.042	0.080
tumor_size	0.007	0.038	0.080
tumor_size	0.007	0.035	0.080
tumor_size	0.007	0.032	0.080
tumor_size	0.007	0.029	0.080
tumor_size	0.007	0.026	0.080
tumor_size	0.007	0.024	0.080
tumor_size	0.007	0.022	0.081
tumor_size	0.007	0.020	0.081
tumor_size	0.007	0.018	0.081
tumor_size	0.007	0.017	0.081
tumor_size	0.007	0.015	0.081
tumor_size	0.007	0.014	0.081
tumor_size	0.007	0.012	0.081
tumor_size	0.007	0.011	0.081
tumor_size	0.007	0.010	0.081
tumor_size	0.008	0.009	0.081
regional_node_examined	0.000	94.435	0.000
regional_node_examined	0.000	86.045	0.001
regional_node_examined	0.000	78.401	0.001
regional_node_examined	0.000	71.436	0.001
regional_node_examined	0.000	65.090	0.001
regional_node_examined	0.000	59.308	0.001
regional_node_examined	0.000	54.039	0.001
regional_node_examined	0.000	49.238	0.001
regional_node_examined	0.000	44.864	0.001
regional_node_examined	0.000	40.879	0.001
regional_node_examined	0.000	37.247	0.001
regional_node_examined	0.000	33.938	0.001
regional_node_examined	0.000	30.923	0.001
regional_node_examined	0.000	28.176	0.002
regional_node_examined	0.000	25.673	0.002
regional_node_examined	0.000	23.392	0.002
regional_node_examined	0.000	21.314	0.002
regional_node_examined	0.000	19.421	0.002
regional_node_examined	0.000	17.695	0.002
regional_node_examined	0.000	16.123	0.003
regional_node_examined	0.000	14.691	0.003

Term	Estimate	Lambda	Deviation Ratio
regional_node_examined	0.000	13.386	0.003
regional_node_examined	0.000	12.197	0.003
regional_node_examined	0.000	11.113	0.004
regional_node_examined	0.000	10.126	0.004
regional_node_examined	0.000	9.226	0.005
regional_node_examined	0.000	8.407	0.005
regional_node_examined	0.000	7.660	0.005
regional_node_examined	0.000	6.979	0.006
regional_node_examined	0.000	6.359	0.007
regional_node_examined	0.000	5.794	0.007
regional_node_examined	0.000	5.280	0.008
regional_node_examined	0.000	4.811	0.008
regional_node_examined	0.000	4.383	0.009
regional_node_examined	0.000	3.994	0.010
regional_node_examined	0.001	3.639	0.011
regional_node_examined	0.001	3.316	0.012
regional_node_examined	0.001	3.021	0.013
regional_node_examined	0.001	2.753	0.014
regional_node_examined	0.001	2.508	0.015
regional_node_examined	0.001	2.285	0.016
regional_node_examined	0.001	2.082	0.018
regional_node_examined	0.001	1.897	0.019
regional_node_examined	0.001	1.729	0.021
regional_node_examined	0.001	1.575	0.022
regional_node_examined	0.001	1.435	0.024
regional_node_examined	0.001	1.308	0.026
regional_node_examined	0.001	1.192	0.028
regional_node_examined	0.001	1.086	0.030
regional_node_examined	0.002	0.989	0.032
regional_node_examined	0.002	0.901	0.034
regional_node_examined	0.002	0.821	0.036
regional_node_examined	0.002	0.748	0.038
regional_node_examined	0.002	0.682	0.040
regional_node_examined	0.002	0.621	0.043
regional_node_examined	0.002	0.566	0.045
regional_node_examined	0.002	0.516	0.047
regional_node_examined	0.002	0.470	0.049
regional_node_examined	0.002	0.428	0.052
regional_node_examined	0.002	0.390	0.054
regional_node_examined	0.003	0.356	0.056
regional_node_examined	0.003	0.324	0.058
regional_node_examined	0.003	0.295	0.060
regional_node_examined	0.003	0.269	0.062
regional_node_examined	0.003	0.245	0.064
regional_node_examined	0.003	0.223	0.065
regional_node_examined	0.003	0.203	0.067
regional_node_examined	0.003	0.185	0.068
regional_node_examined	0.003	0.169	0.070
regional_node_examined	0.002	0.154	0.071
regional_node_examined	0.002	0.140	0.072
regional_node_examined	0.002	0.128	0.073
regional_node_examined	0.002	0.116	0.074

Term	Estimate	Lambda	Deviation Ratio
regional_node_examined	0.002	0.106	0.075
regional_node_examined	0.002	0.097	0.076
regional_node_examined	0.002	0.088	0.076
regional_node_examined	0.002	0.080	0.077
regional_node_examined	0.002	0.073	0.077
regional_node_examined	0.002	0.067	0.078
regional_node_examined	0.001	0.061	0.078
regional_node_examined	0.001	0.055	0.079
regional_node_examined	0.001	0.050	0.079
regional_node_examined	0.001	0.046	0.079
regional_node_examined	0.001	0.042	0.080
regional_node_examined	0.001	0.038	0.080
regional_node_examined	0.001	0.035	0.080
regional_node_examined	0.001	0.032	0.080
regional_node_examined	0.000	0.029	0.080
regional_node_examined	0.000	0.026	0.080
regional_node_examined	0.000	0.024	0.080
regional_node_examined	0.000	0.022	0.081
regional_node_examined	0.000	0.020	0.081
regional_node_examined	0.000	0.018	0.081
regional_node_examined	0.000	0.017	0.081
regional_node_examined	0.000	0.015	0.081
regional_node_examined	-0.001	0.014	0.081
regional_node_examined	-0.001	0.012	0.081
regional_node_examined	-0.001	0.011	0.081
regional_node_examined	-0.001	0.010	0.081
regional_node_examined	-0.001	0.009	0.081
reginol_node_positive	0.000	94.435	0.000
reginol_node_positive	0.000	86.045	0.001
reginol_node_positive	0.000	78.401	0.001
reginol_node_positive	0.000	71.436	0.001
reginol_node_positive	0.000	65.090	0.001
reginol_node_positive	0.000	59.308	0.001
reginol_node_positive	0.000	54.039	0.001
reginol_node_positive	0.000	49.238	0.001
reginol_node_positive	0.000	44.864	0.001
reginol_node_positive	0.000	40.879	0.001
reginol_node_positive	0.000	37.247	0.001
reginol_node_positive	0.001	33.938	0.001
reginol_node_positive	0.001	30.923	0.001
reginol_node_positive	0.001	28.176	0.002
reginol_node_positive	0.001	25.673	0.002
reginol_node_positive	0.001	23.392	0.002
reginol_node_positive	0.001	21.314	0.002
reginol_node_positive	0.001	19.421	0.002
reginol_node_positive	0.001	17.695	0.002
reginol_node_positive	0.001	16.123	0.003
reginol_node_positive	0.001	14.691	0.003
reginol_node_positive	0.001	13.386	0.003
reginol_node_positive	0.001	12.197	0.003
reginol_node_positive	0.002	11.113	0.004
reginol_node_positive	0.002	10.126	0.004

Term	Estimate	Lambda	Deviation Ratio
reginol_node_positive	0.002	9.226	0.005
reginol_node_positive	0.002	8.407	0.005
reginol_node_positive	0.002	7.660	0.005
reginol_node_positive	0.003	6.979	0.006
reginol_node_positive	0.003	6.359	0.007
reginol_node_positive	0.003	5.794	0.007
reginol_node_positive	0.003	5.280	0.008
reginol_node_positive	0.004	4.811	0.008
reginol_node_positive	0.004	4.383	0.009
reginol_node_positive	0.004	3.994	0.010
reginol_node_positive	0.005	3.639	0.011
reginol_node_positive	0.005	3.316	0.012
reginol_node_positive	0.006	3.021	0.013
reginol_node_positive	0.006	2.753	0.014
reginol_node_positive	0.007	2.508	0.015
reginol_node_positive	0.007	2.285	0.016
reginol_node_positive	0.008	2.082	0.018
reginol_node_positive	0.009	1.897	0.019
reginol_node_positive	0.009	1.729	0.021
reginol_node_positive	0.010	1.575	0.022
reginol_node_positive	0.011	1.435	0.024
reginol_node_positive	0.012	1.308	0.026
reginol_node_positive	0.013	1.192	0.028
reginol_node_positive	0.014	1.086	0.030
reginol_node_positive	0.015	0.989	0.032
reginol_node_positive	0.016	0.901	0.034
reginol_node_positive	0.017	0.821	0.036
reginol_node_positive	0.018	0.748	0.038
reginol_node_positive	0.019	0.682	0.040
reginol_node_positive	0.021	0.621	0.043
reginol_node_positive	0.022	0.566	0.045
reginol_node_positive	0.023	0.516	0.047
reginol_node_positive	0.025	0.470	0.049
reginol_node_positive	0.026	0.428	0.052
reginol_node_positive	0.028	0.390	0.054
reginol_node_positive	0.029	0.356	0.056
reginol_node_positive	0.031	0.324	0.058
reginol_node_positive	0.032	0.295	0.060
reginol_node_positive	0.033	0.269	0.062
reginol_node_positive	0.035	0.245	0.064
reginol_node_positive	0.036	0.223	0.065
reginol_node_positive	0.038	0.203	0.067
reginol_node_positive	0.039	0.185	0.068
reginol_node_positive	0.040	0.169	0.070
reginol_node_positive	0.041	0.154	0.071
reginol_node_positive	0.043	0.140	0.072
reginol_node_positive	0.044	0.128	0.073
reginol_node_positive	0.045	0.116	0.074
reginol_node_positive	0.046	0.106	0.075
reginol_node_positive	0.047	0.097	0.076
reginol_node_positive	0.048	0.088	0.076
reginol_node_positive	0.049	0.080	0.077

Term	Estimate	Lambda	Deviation Ratio
reginol_node_positive	0.050	0.073	0.077
reginol_node_positive	0.051	0.067	0.078
reginol_node_positive	0.051	0.061	0.078
reginol_node_positive	0.052	0.055	0.079
reginol_node_positive	0.053	0.050	0.079
reginol_node_positive	0.054	0.046	0.079
reginol_node_positive	0.054	0.042	0.080
reginol_node_positive	0.055	0.038	0.080
reginol_node_positive	0.055	0.035	0.080
reginol_node_positive	0.056	0.032	0.080
reginol_node_positive	0.056	0.029	0.080
reginol_node_positive	0.057	0.026	0.080
reginol_node_positive	0.057	0.024	0.080
reginol_node_positive	0.058	0.022	0.081
reginol_node_positive	0.058	0.020	0.081
reginol_node_positive	0.058	0.018	0.081
reginol_node_positive	0.059	0.017	0.081
reginol_node_positive	0.059	0.015	0.081
reginol_node_positive	0.059	0.014	0.081
reginol_node_positive	0.060	0.012	0.081
reginol_node_positive	0.060	0.011	0.081
reginol_node_positive	0.060	0.010	0.081
reginol_node_positive	0.060	0.009	0.081
node_positive_prop	0.000	94.435	0.000
node_positive_prop	0.003	86.045	0.001
node_positive_prop	0.004	78.401	0.001
node_positive_prop	0.004	71.436	0.001
node_positive_prop	0.005	65.090	0.001
node_positive_prop	0.005	59.308	0.001
node_positive_prop	0.006	54.039	0.001
node_positive_prop	0.006	49.238	0.001
node_positive_prop	0.007	44.864	0.001
node_positive_prop	0.007	40.879	0.001
node_positive_prop	0.008	37.247	0.001
node_positive_prop	0.009	33.938	0.001
node_positive_prop	0.010	30.923	0.001
node_positive_prop	0.011	28.176	0.002
node_positive_prop	0.012	25.673	0.002
node_positive_prop	0.013	23.392	0.002
node_positive_prop	0.014	21.314	0.002
node_positive_prop	0.015	19.421	0.002
node_positive_prop	0.017	17.695	0.002
node_positive_prop	0.018	16.123	0.003
node_positive_prop	0.020	14.691	0.003
node_positive_prop	0.022	13.386	0.003
node_positive_prop	0.024	12.197	0.003
node_positive_prop	0.027	11.113	0.004
node_positive_prop	0.029	10.126	0.004
node_positive_prop	0.032	9.226	0.005
node_positive_prop	0.035	8.407	0.005
node_positive_prop	0.038	7.660	0.005
node_positive_prop	0.042	6.979	0.006

Term	Estimate	Lambda	Deviation Ratio
node_positive_prop	0.046	6.359	0.007
node_positive_prop	0.050	5.794	0.007
node_positive_prop	0.055	5.280	0.008
node_positive_prop	0.060	4.811	0.008
node_positive_prop	0.065	4.383	0.009
node_positive_prop	0.071	3.994	0.010
node_positive_prop	0.078	3.639	0.011
node_positive_prop	0.085	3.316	0.012
node_positive_prop	0.092	3.021	0.013
node_positive_prop	0.101	2.753	0.014
node_positive_prop	0.110	2.508	0.015
node_positive_prop	0.119	2.285	0.016
node_positive_prop	0.130	2.082	0.018
node_positive_prop	0.141	1.897	0.019
node_positive_prop	0.153	1.729	0.021
node_positive_prop	0.166	1.575	0.022
node_positive_prop	0.180	1.435	0.024
node_positive_prop	0.194	1.308	0.026
node_positive_prop	0.210	1.192	0.028
node_positive_prop	0.227	1.086	0.030
node_positive_prop	0.244	0.989	0.032
node_positive_prop	0.263	0.901	0.034
node_positive_prop	0.282	0.821	0.036
node_positive_prop	0.302	0.748	0.038
node_positive_prop	0.324	0.682	0.040
node_positive_prop	0.346	0.621	0.043
node_positive_prop	0.369	0.566	0.045
node_positive_prop	0.393	0.516	0.047
node_positive_prop	0.417	0.470	0.049
node_positive_prop	0.442	0.428	0.052
node_positive_prop	0.467	0.390	0.054
node_positive_prop	0.493	0.356	0.056
node_positive_prop	0.519	0.324	0.058
node_positive_prop	0.546	0.295	0.060
node_positive_prop	0.572	0.269	0.062
node_positive_prop	0.599	0.245	0.064
node_positive_prop	0.625	0.223	0.065
node_positive_prop	0.651	0.203	0.067
node_positive_prop	0.677	0.185	0.068
node_positive_prop	0.702	0.169	0.070
node_positive_prop	0.727	0.154	0.071
node_positive_prop	0.751	0.140	0.072
node_positive_prop	0.774	0.128	0.073
node_positive_prop	0.797	0.116	0.074
node_positive_prop	0.819	0.106	0.075
node_positive_prop	0.840	0.097	0.076
node_positive_prop	0.860	0.088	0.076
node_positive_prop	0.879	0.080	0.077
node_positive_prop	0.898	0.073	0.077
node_positive_prop	0.915	0.067	0.078
node_positive_prop	0.931	0.061	0.078
node_positive_prop	0.947	0.055	0.079

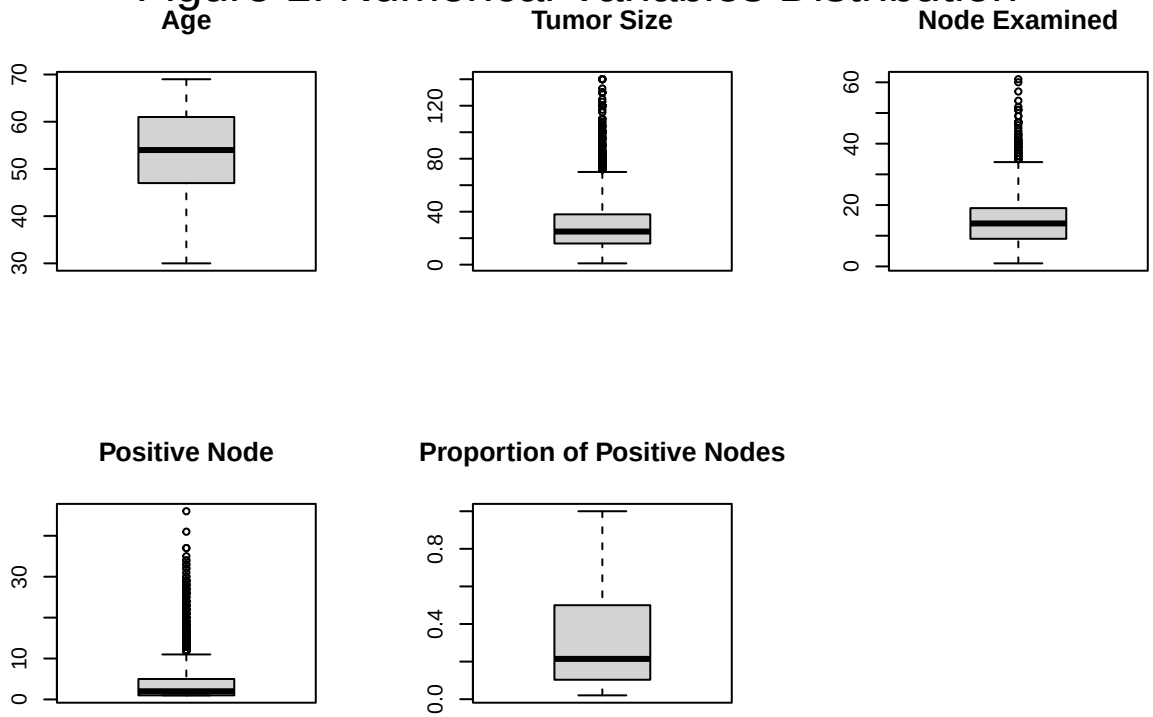
Term	Estimate	Lambda	Deviation Ratio
node_positive_prop	0.961	0.050	0.079
node_positive_prop	0.975	0.046	0.079
node_positive_prop	0.987	0.042	0.080
node_positive_prop	0.999	0.038	0.080
node_positive_prop	1.010	0.035	0.080
node_positive_prop	1.020	0.032	0.080
node_positive_prop	1.029	0.029	0.080
node_positive_prop	1.037	0.026	0.080
node_positive_prop	1.045	0.024	0.080
node_positive_prop	1.052	0.022	0.081
node_positive_prop	1.058	0.020	0.081
node_positive_prop	1.064	0.018	0.081
node_positive_prop	1.069	0.017	0.081
node_positive_prop	1.074	0.015	0.081
node_positive_prop	1.078	0.014	0.081
node_positive_prop	1.081	0.012	0.081
node_positive_prop	1.085	0.011	0.081
node_positive_prop	1.088	0.010	0.081
node_positive_prop	1.090	0.009	0.081

Table 8: AIC & BIC Scores for Backward and Forward Selection Model

aic_back	aic_forward	bic_back	bic_forward
2992.229	3004.342	3099.33	3124.043

Figure

Figure 1: Numerical Variables Distribution



Contribution