

Cancer Model Univariate Analysis Report

2024-02-16

Overview

Cancer Model Univariate Analysis Report

These sorted results for the features in this report indicate the average cross-validated test scores for each feature, if it were used as the only predictor in a simple linear model. Keep in mind that these results are based on the average, without considering the standard deviation. This means that the results are not necessarily the best predictors, but they are the best on average, and provide a fine starting point for grouping those predictors that are on average better than others. This means that nothing was done to account for possible sampling variability in the sorted results. This is a limitation of the univariate analysis, so it is important to keep this in mind when interpreting the results. It is also important to consider further that depending on the purpose of the model, the most appropriate features may not be the ones with the highest average test scores, if a different metric is more important.

In particular, this should not be taken as an opinion (actuarial or otherwise) regarding the most appropriate features to use in a model, but it rather provides a starting point for further analysis.

	Accuracy	Precision	Recall	AUC	F1	MCC	Ave.
Concave Points Error	67.5%	76.6%	69.0%	67.1%	72.6%	33.3%	64.4%
Worst Texture	66.7%	78.0%	64.8%	67.3%	70.8%	33.5%	63.5%
Compactness Error	65.8%	73.5%	70.4%	64.3%	71.9%	28.2%	62.4%
Worst Symmetry	62.3%	74.1%	60.6%	62.8%	66.7%	24.9%	58.6%
Fractal Dimension Error	62.3%	71.2%	66.2%	61.0%	68.6%	21.6%	58.5%
Worst Fractal Dimension	61.4%	70.8%	64.8%	60.3%	67.6%	20.2%	57.5%
Pc2	57.0%	67.7%	59.2%	56.3%	63.2%	12.3%	52.6%
Texture Error	55.3%	70.0%	49.3%	57.2%	57.9%	14.1%	50.6%
Mean Symmetry	55.3%	70.0%	49.3%	57.2%	57.9%	14.1%	50.6%
Smoothness Error	52.6%	68.1%	45.1%	55.1%	54.2%	10.0%	47.5%
Symmetry Error	45.6%	56.3%	56.3%	42.1%	56.3%	-15.8%	40.2%
Mean Fractal Dimension	44.7%	58.0%	40.8%	46.0%	47.9%	-7.8%	38.3%

This table shows an overview of the results for the variables in this file, representing those whose average test score are ranked between 21 and 32 of the variables passed to the Cancer Model.

Univariate Report

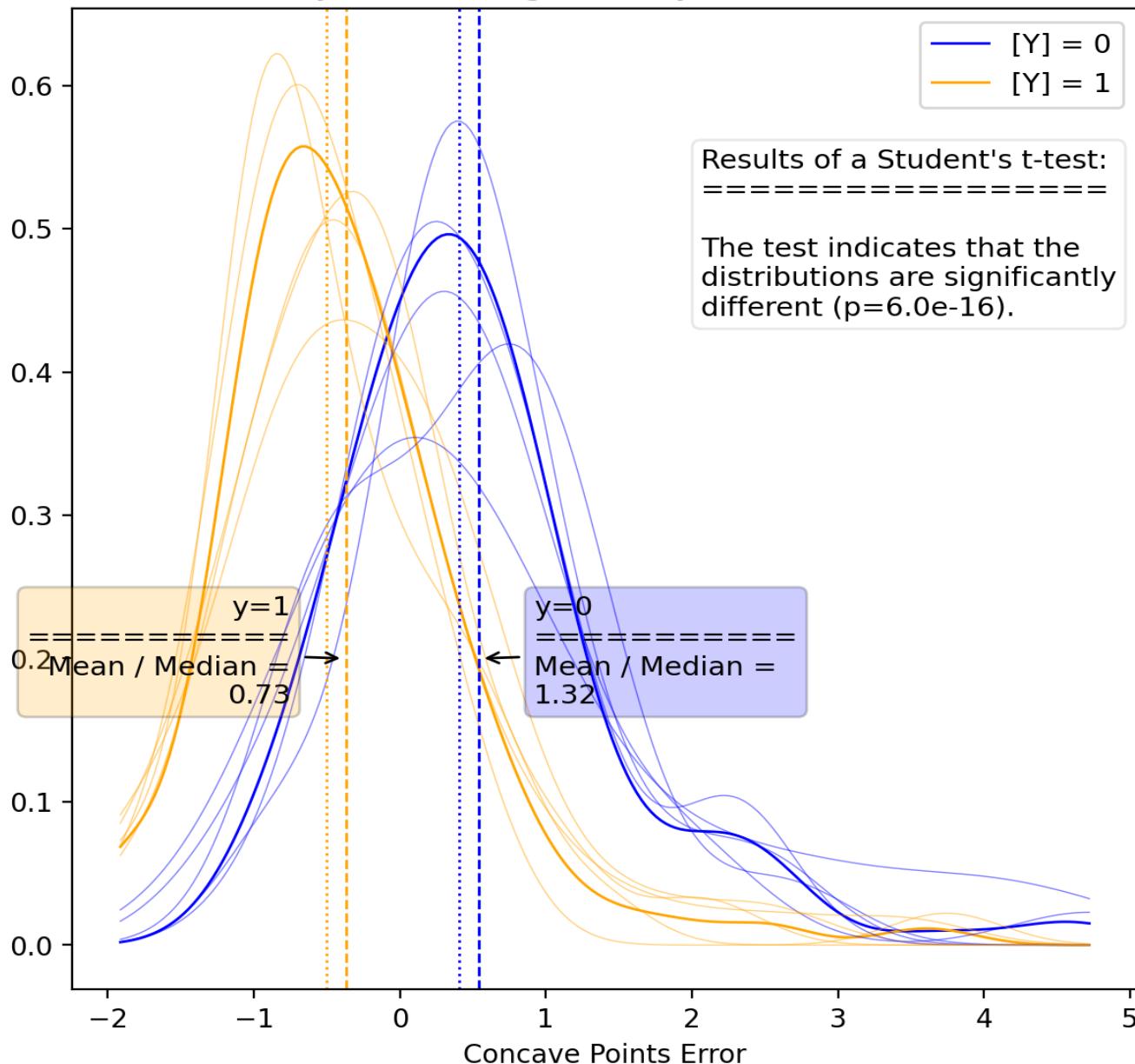
Concave Points Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-1.12	-1.36	-1.43	-1.11	-1.28	-1.25	0.14
Fitted p-Value	8.9e-10	1.2e-11	1.1e-11	2.3e-09	1.3e-11	4.0e-13	1.0e-09
Fitted Std. Err.	0.183	0.200	0.210	0.186	0.188	0.173	0.011
Conf. Int. Lower	-1.48	-1.75	-1.84	-1.47	-1.64	-1.59	0.16
Conf. Int. Upper	-0.764	-0.965	-1.016	-0.745	-0.906	-0.915	0.121
Train Accuracy	70.8%	75.0%	74.3%	73.4%	74.1%	73.3%	1.6%
Val Accuracy	84.3%	66.7%	72.6%	75.7%	69.1%	67.5%	6.9%
Train AUC	70.4%	74.9%	74.2%	73.2%	74.8%	73.5%	1.9%
Val AUC	85.0%	67.0%	71.8%	75.1%	68.2%	67.1%	7.2%
Train F1	75.4%	79.8%	77.7%	77.9%	78.0%	77.4%	1.6%
Test F1	87.6%	64.6%	79.6%	79.5%	73.8%	72.6%	8.5%
Train Precision	78.6%	84.8%	81.1%	82.2%	84.5%	82.5%	2.6%
Val Precision	92.9%	72.4%	86.7%	81.4%	75.0%	76.6%	8.4%
Train Recall	72.5%	75.3%	74.5%	74.0%	72.4%	72.9%	1.3%
Val Recall	83.0%	58.3%	73.6%	77.8%	72.7%	69.0%	9.2%
Train MCC	40.0%	47.9%	47.6%	45.2%	47.9%	45.6%	3.4%
Val MCC	67.1%	34.5%	40.0%	49.7%	36.1%	33.3%	13.4%
Train Log-Loss	10.51	9.01	9.28	9.58	9.33	9.62	0.58
Val Log-Loss	5.66	12.01	9.87	8.77	11.14	11.70	2.47

Univariate Report

Concave Points Error - Kernel Density Plot

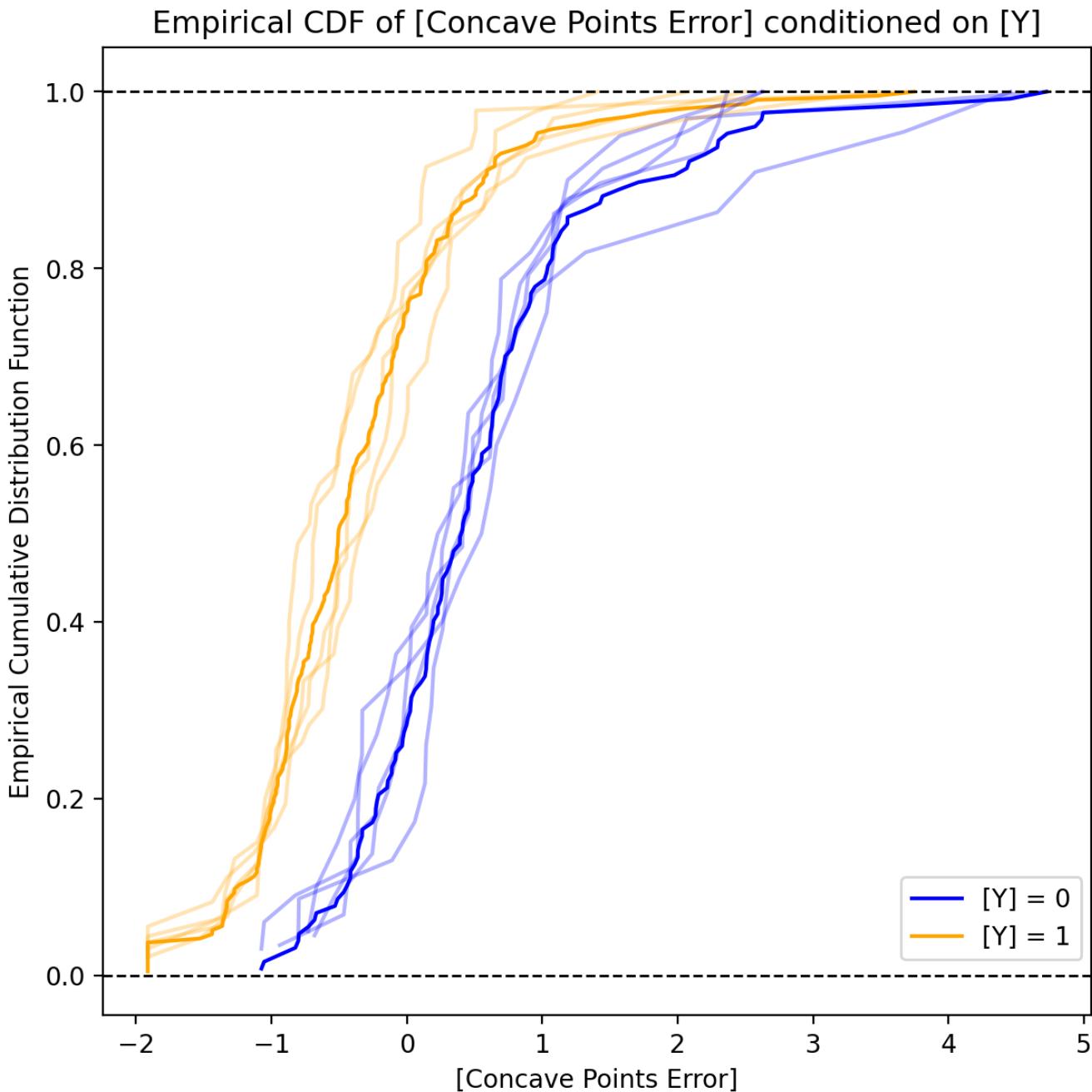
Kernel Density Plot of [Concave Points Error] by [Y].
Distributions by level are significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

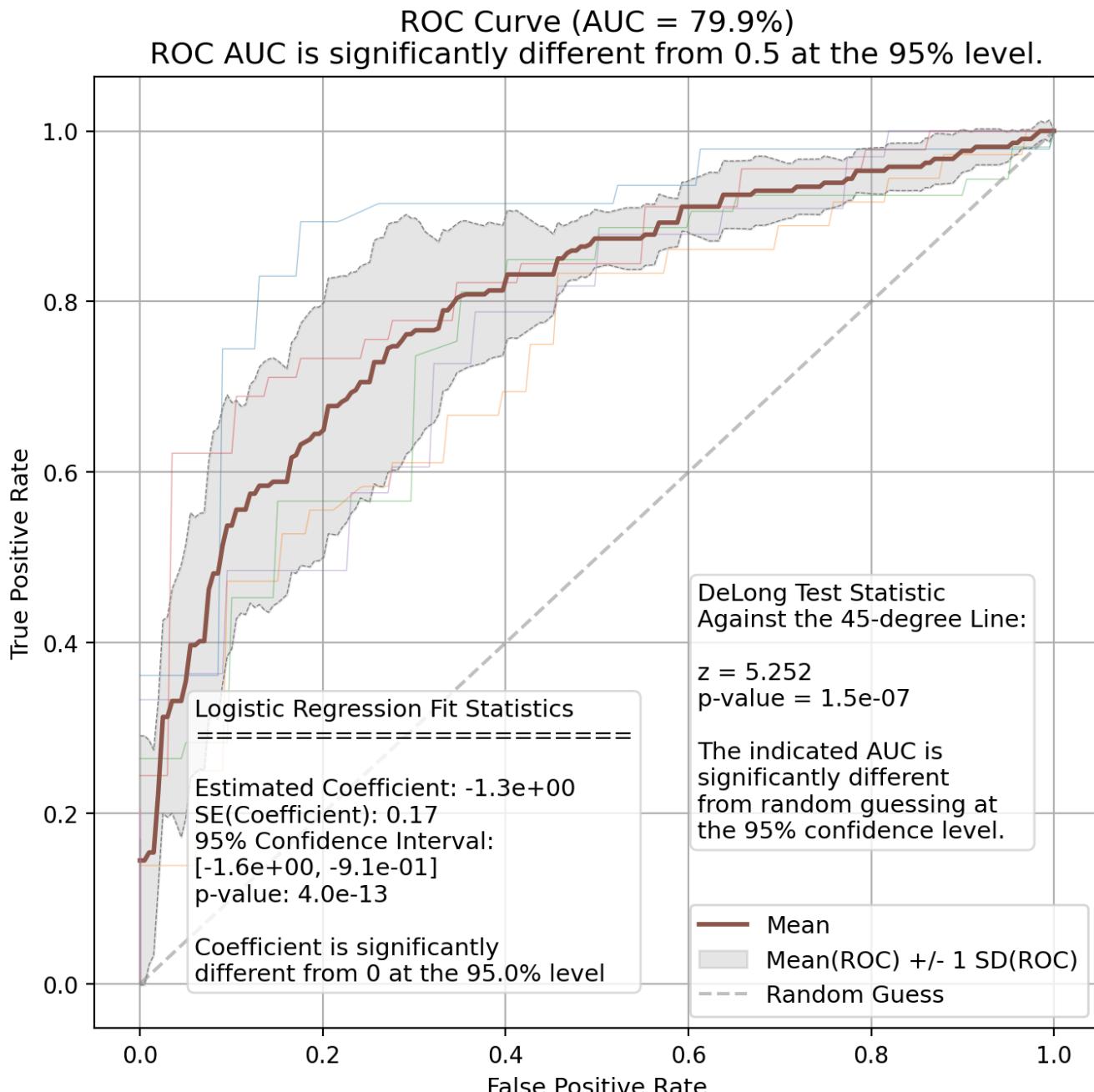
Concave Points Error - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Concave Points Error - ROC Curve

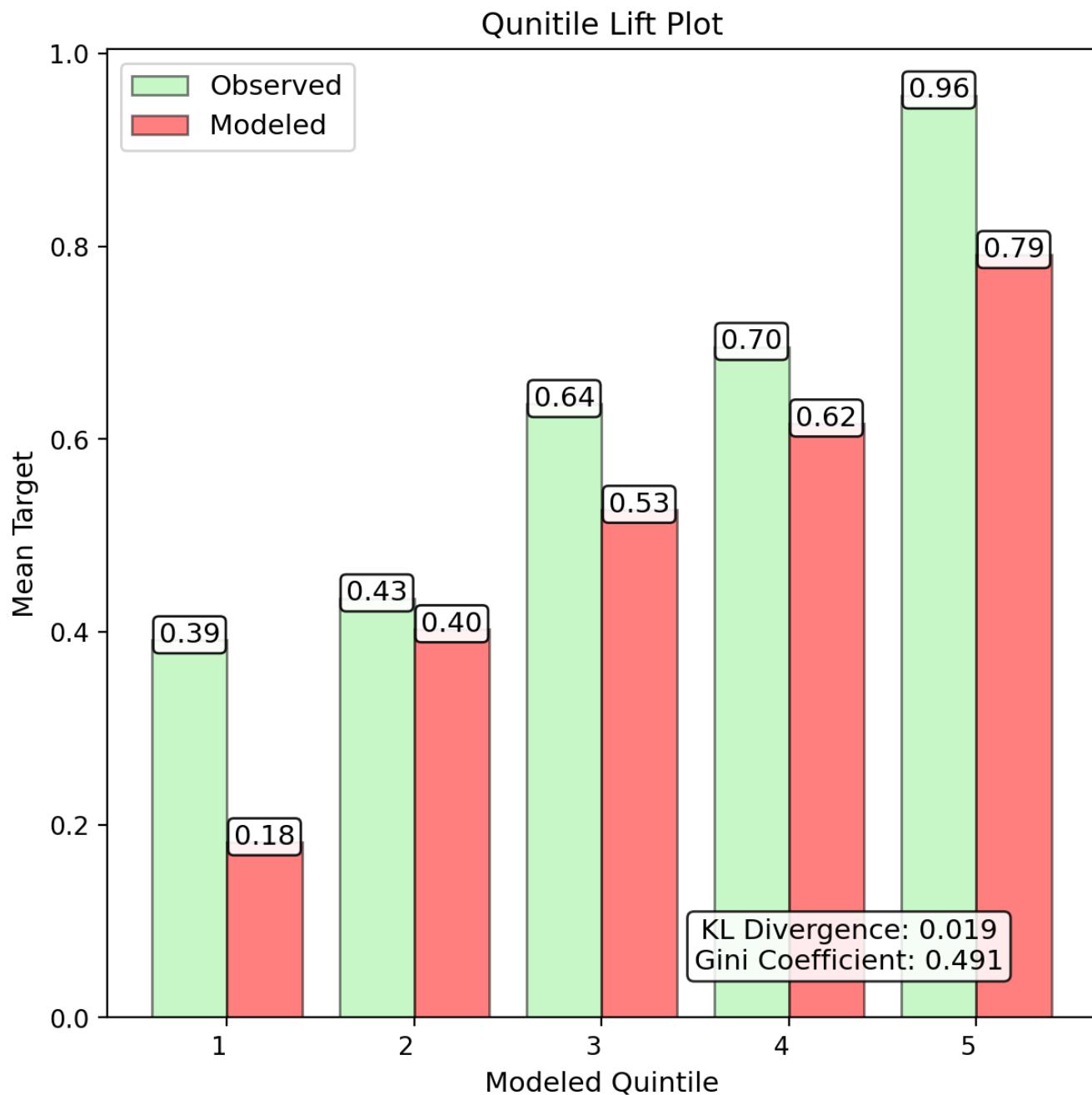


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Concave Points Error - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

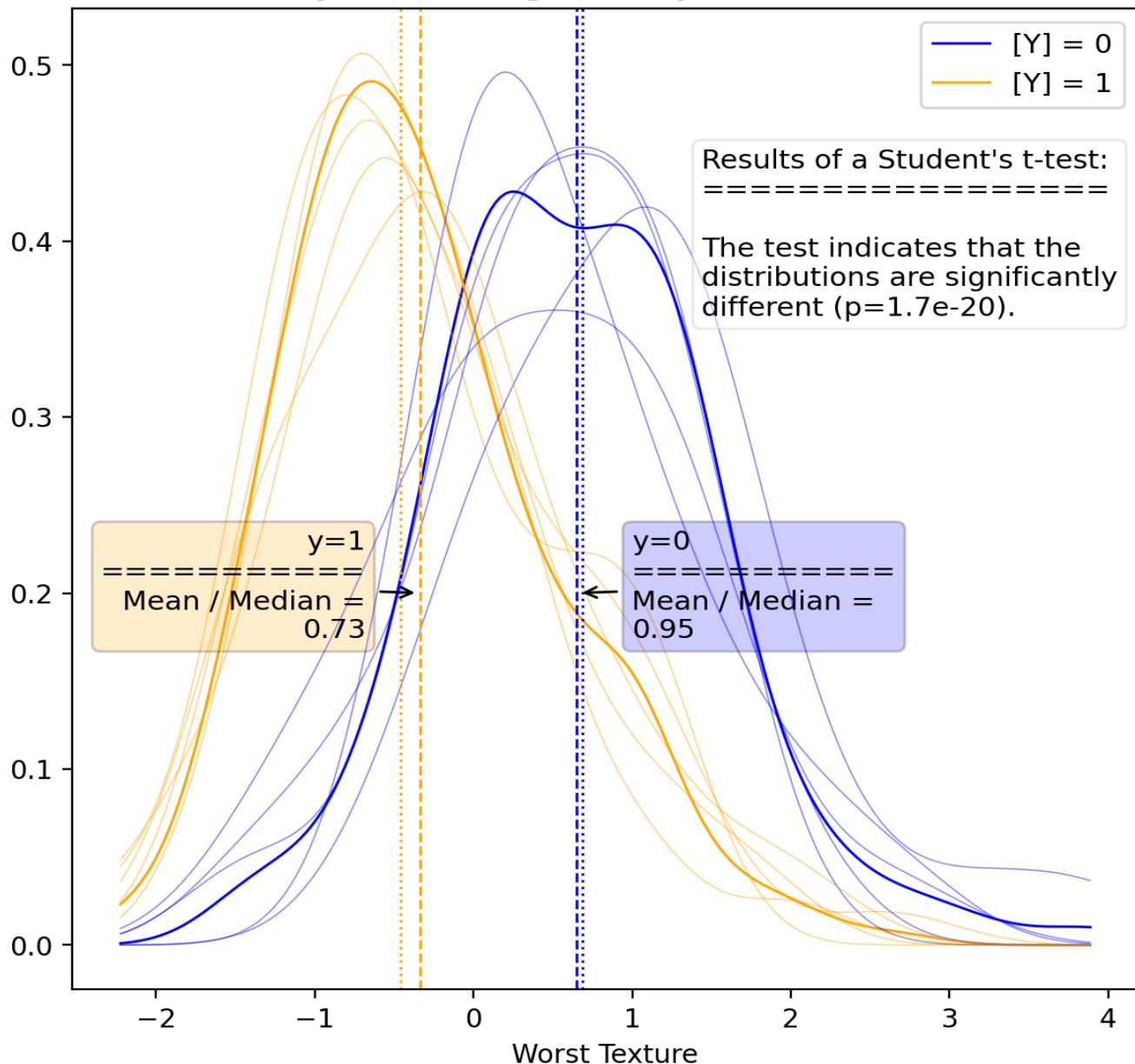
Worst Texture - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-1.24	-1.16	-1.33	-1.03	-1.24	-1.20	0.11
Fitted p-Value	2.0e-12	7.6e-12	5.7e-13	6.6e-10	5.4e-13	1.1e-14	2.9e-10
Fitted Std. Err.	0.176	0.169	0.184	0.167	0.172	0.155	0.007
Conf. Int. Lower	-1.59	-1.49	-1.69	-1.36	-1.58	-1.50	0.12
Conf. Int. Upper	-0.895	-0.828	-0.967	-0.706	-0.902	-0.894	0.099
Train Accuracy	75.6%	73.5%	73.9%	71.2%	74.8%	74.2%	1.7%
Val Accuracy	68.6%	72.5%	75.3%	82.4%	72.7%	66.7%	5.1%
Train AUC	76.2%	74.5%	74.0%	71.6%	75.3%	74.6%	1.7%
Val AUC	66.6%	72.6%	76.8%	83.7%	73.5%	67.3%	6.3%
Train F1	78.8%	77.9%	77.1%	75.4%	78.7%	78.0%	1.4%
Test F1	75.6%	72.5%	81.2%	84.3%	75.4%	70.8%	4.8%
Train Precision	84.8%	85.8%	81.4%	81.9%	84.7%	83.9%	2.0%
Val Precision	79.1%	75.8%	90.7%	92.1%	82.1%	78.0%	7.2%
Train Recall	73.7%	71.3%	73.3%	69.8%	73.5%	72.9%	1.7%
Val Recall	72.3%	69.4%	73.6%	77.8%	69.7%	64.8%	3.4%
Train MCC	51.2%	46.8%	47.2%	41.9%	49.0%	47.8%	3.5%
Val MCC	32.0%	45.2%	48.6%	65.9%	46.0%	33.5%	12.1%
Train Log-Loss	8.78	9.54	9.41	10.39	9.07	9.30	0.61
Val Log-Loss	11.33	9.93	8.89	6.33	9.83	12.01	1.85

Univariate Report

Worst Texture - Kernel Density Plot

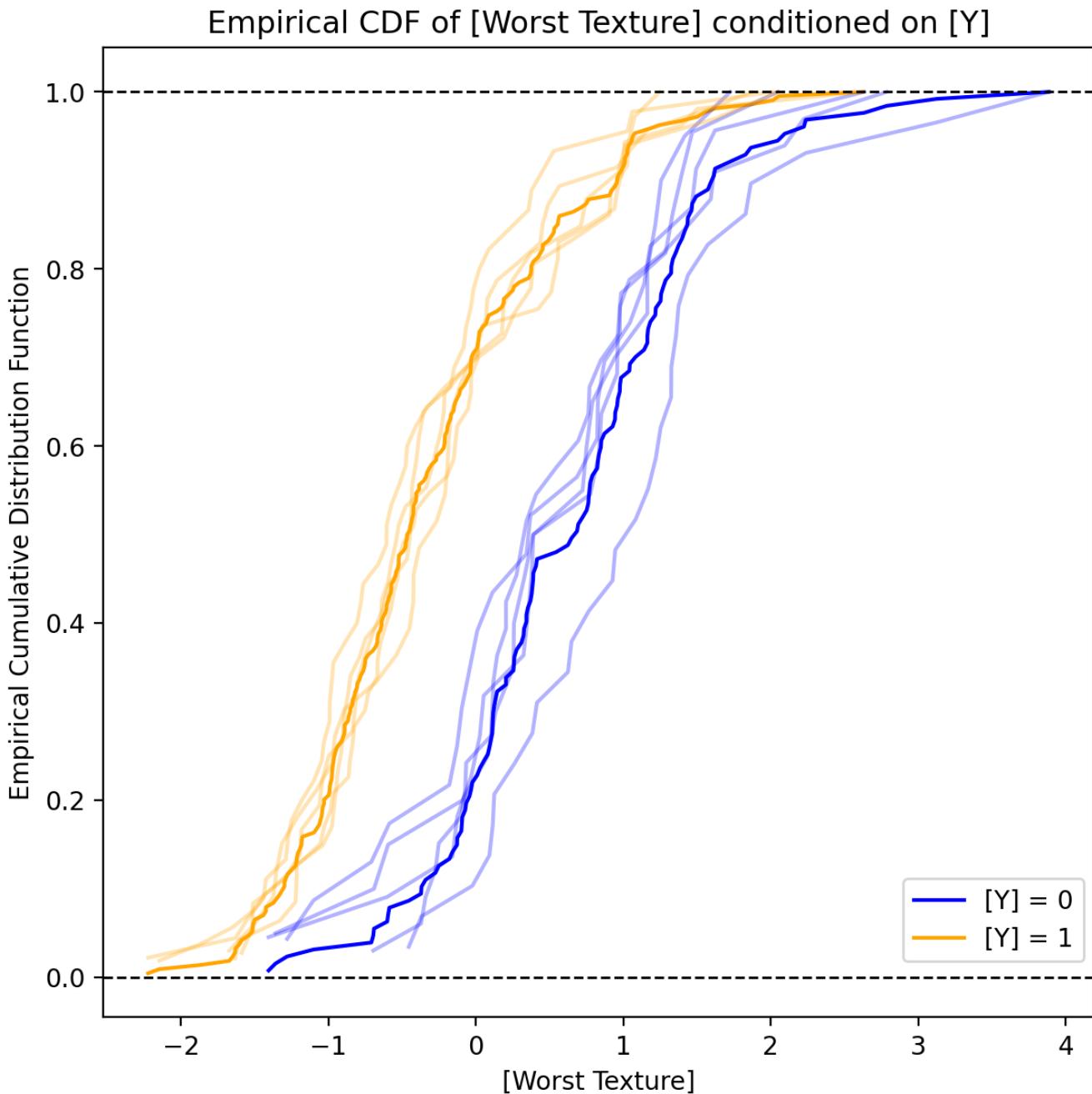
Kernel Density Plot of [Worst Texture] by [Y].
Distributions by level are significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

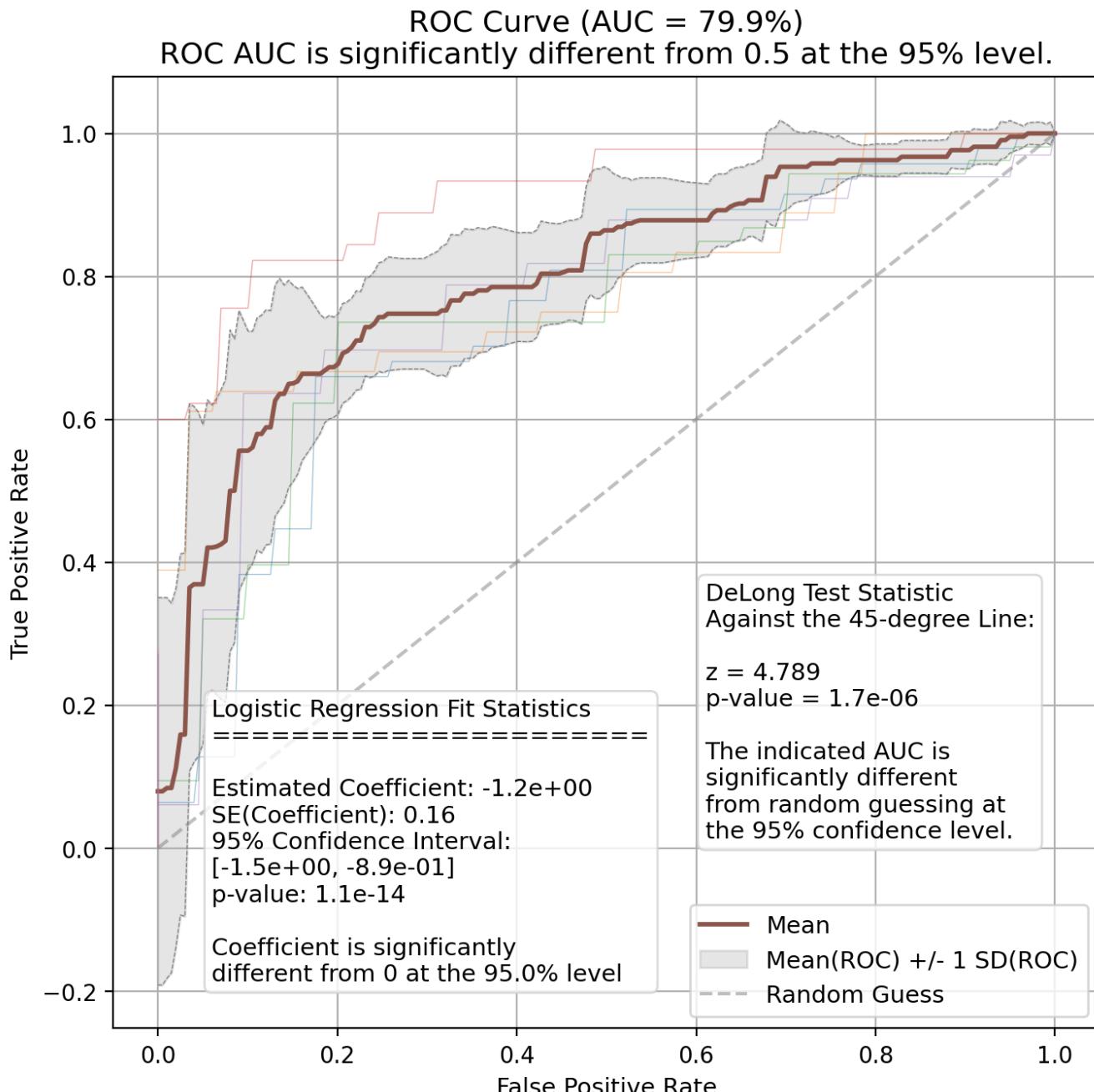
Worst Texture - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Worst Texture - ROC Curve

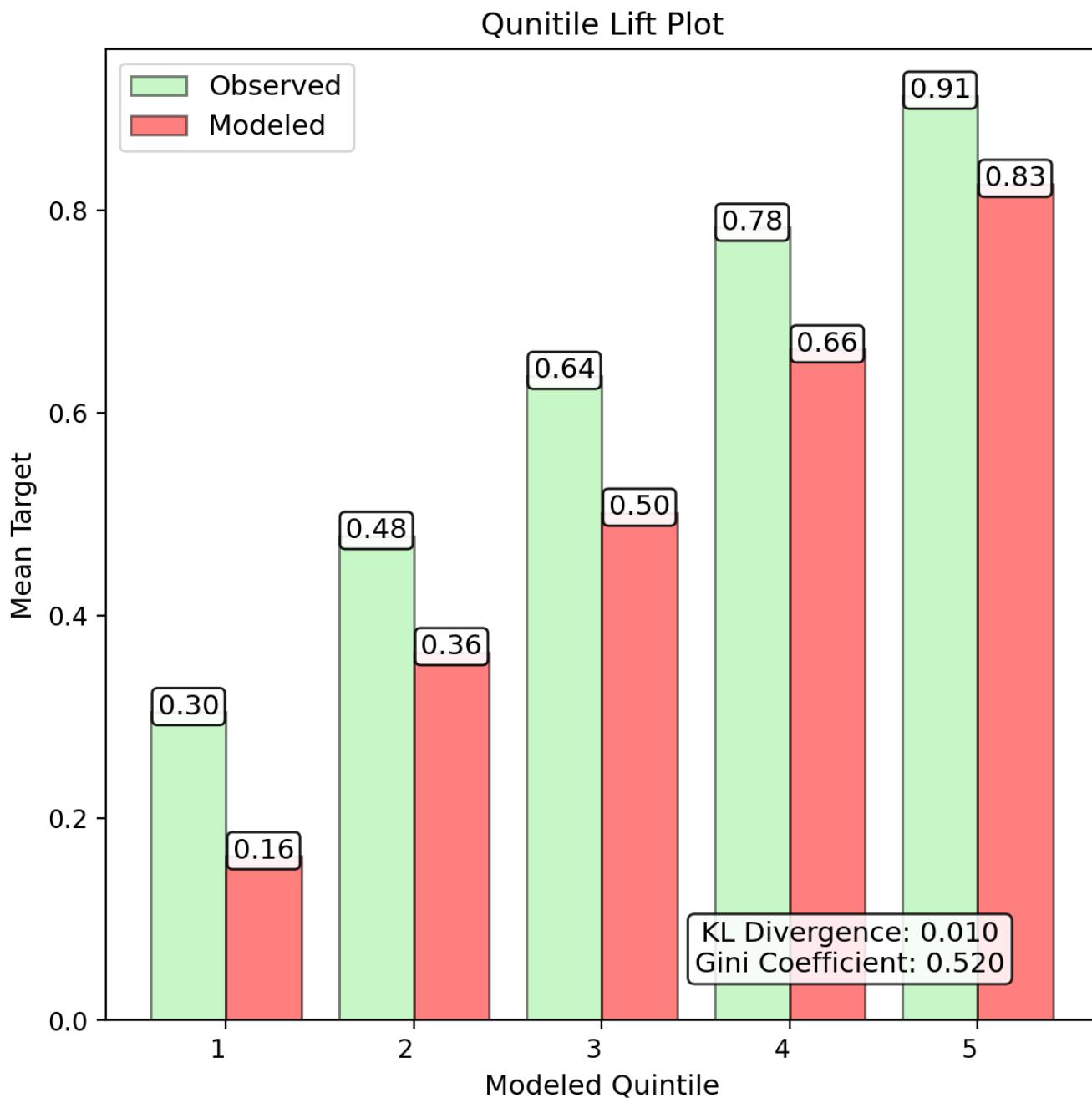


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst Texture - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

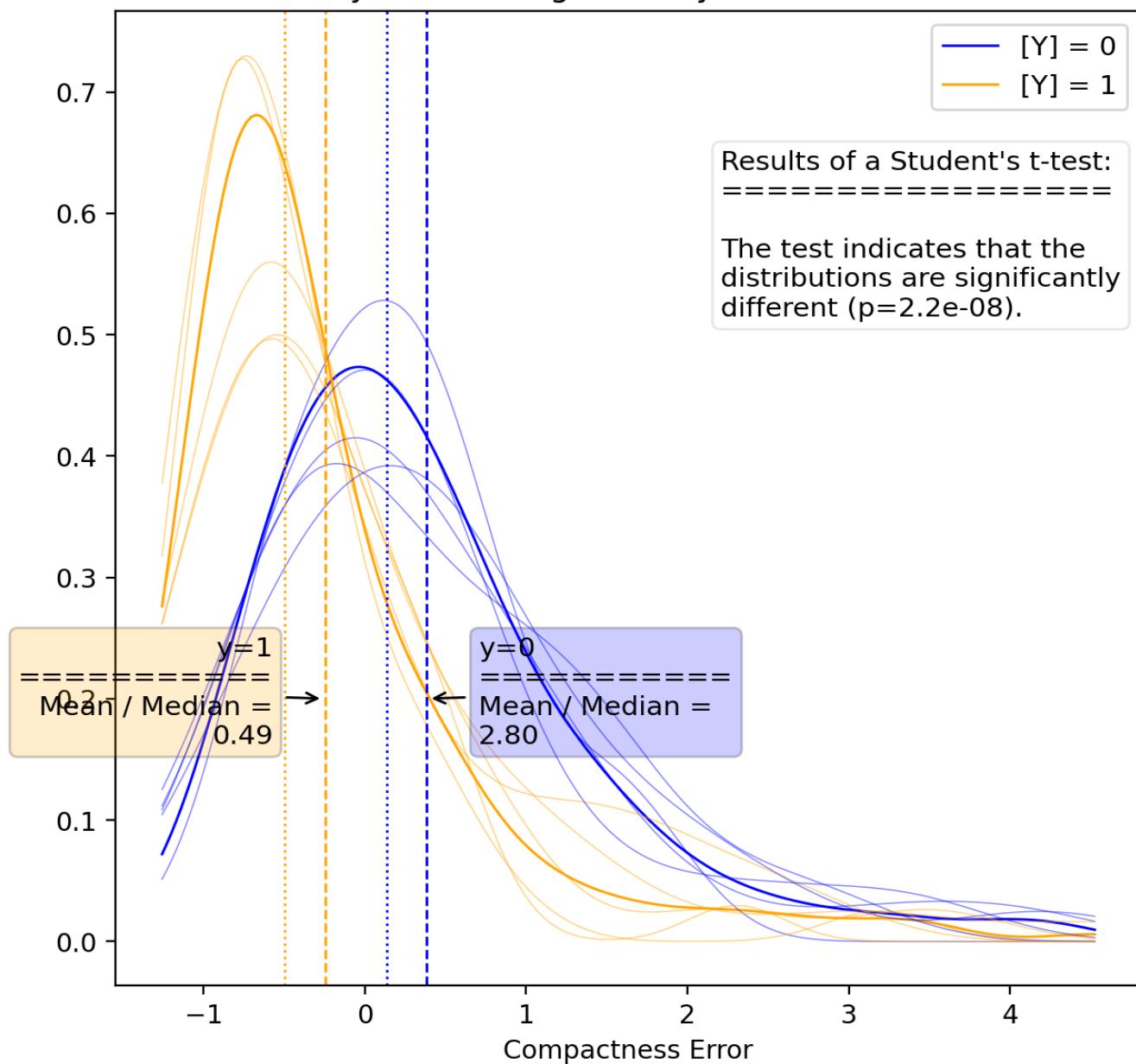
Compactness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.588	-0.870	-0.930	-0.576	-0.761	-0.733	0.161
Fitted p-Value	5.7e-05	4.6e-07	1.6e-07	1.1e-04	1.2e-06	2.8e-07	4.8e-05
Fitted Std. Err.	0.146	0.172	0.177	0.149	0.157	0.143	0.014
Conf. Int. Lower	-0.87	-1.21	-1.28	-0.87	-1.07	-1.01	0.19
Conf. Int. Upper	-0.302	-0.532	-0.582	-0.285	-0.454	-0.453	0.134
Train Accuracy	65.3%	71.0%	70.9%	68.2%	69.9%	69.5%	2.4%
Val Accuracy	78.6%	65.2%	64.4%	73.0%	67.3%	65.8%	6.0%
Train AUC	63.0%	69.0%	69.5%	65.8%	68.0%	67.5%	2.7%
Val AUC	75.2%	65.0%	61.5%	71.0%	65.2%	64.3%	5.5%
Train F1	72.2%	77.2%	75.9%	74.8%	76.0%	75.6%	1.9%
Test F1	84.2%	67.6%	73.5%	78.3%	73.5%	71.9%	6.2%
Train Precision	71.3%	79.3%	75.5%	75.0%	76.8%	75.9%	2.9%
Val Precision	83.3%	65.8%	80.0%	76.6%	71.4%	73.5%	7.0%
Train Recall	73.1%	75.3%	76.4%	74.6%	75.1%	75.2%	1.2%
Val Recall	85.1%	69.4%	67.9%	80.0%	75.8%	70.4%	7.2%
Train MCC	26.1%	37.3%	39.1%	31.6%	35.8%	35.0%	5.2%
Val MCC	50.9%	30.2%	21.0%	42.7%	30.9%	28.2%	11.7%
Train Log-Loss	12.50	10.47	10.49	11.47	10.84	10.99	0.86
Val Log-Loss	7.72	12.54	12.84	9.74	11.80	12.33	2.16

Univariate Report

Compactness Error - Kernel Density Plot

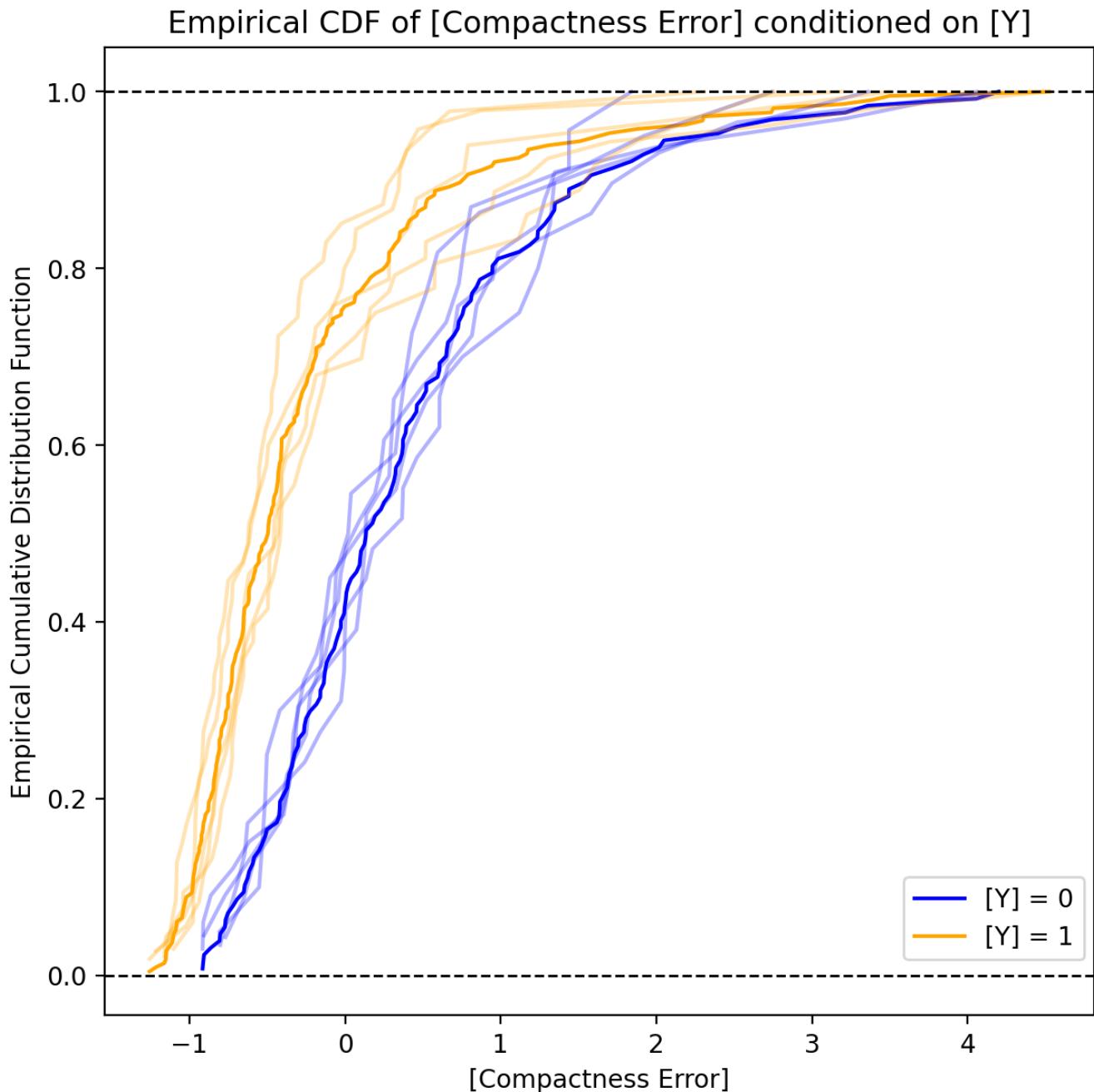
Kernel Density Plot of [Compactness Error] by [Y].
Distributions by level are significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

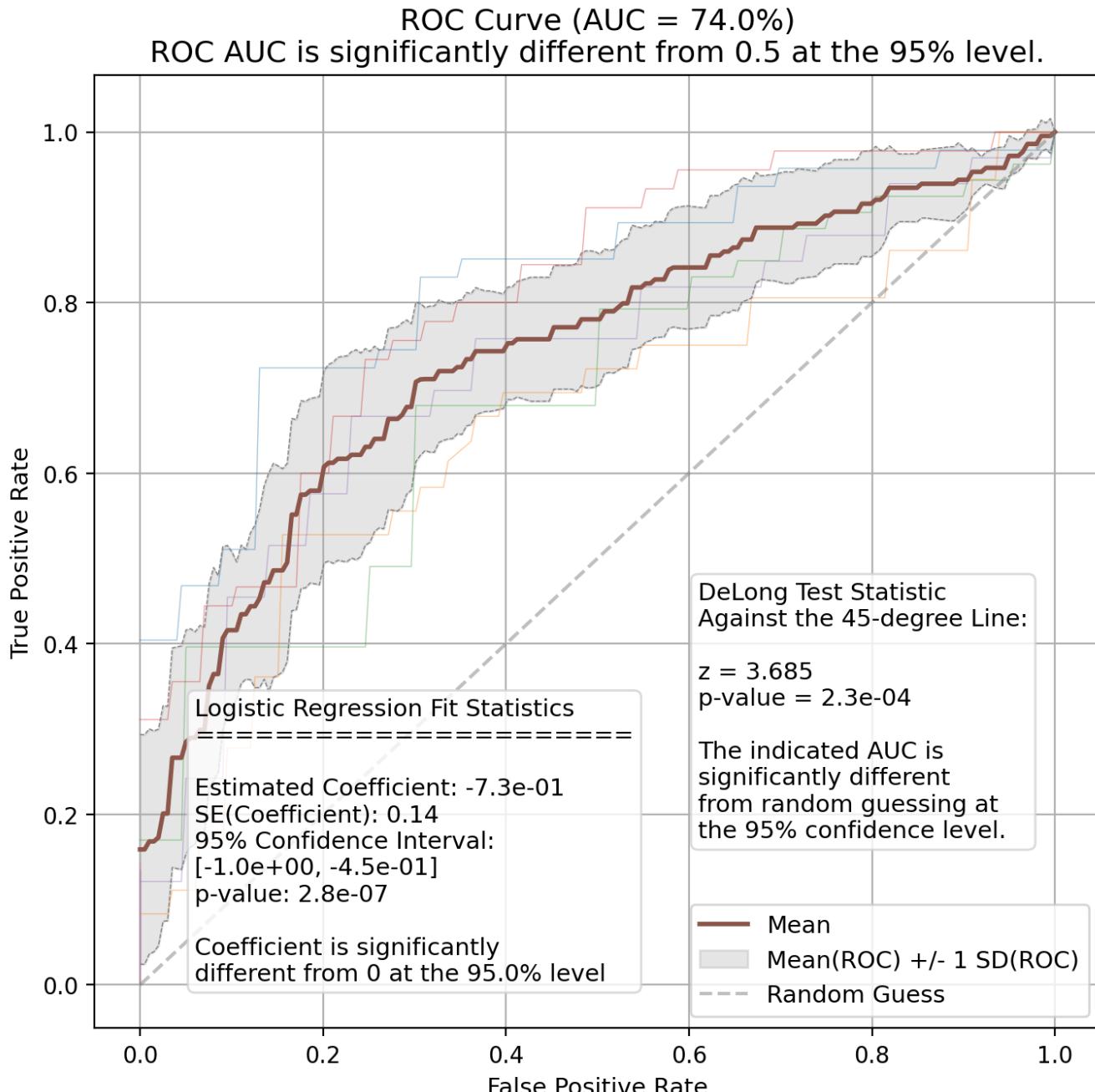
Compactness Error - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Compactness Error - ROC Curve

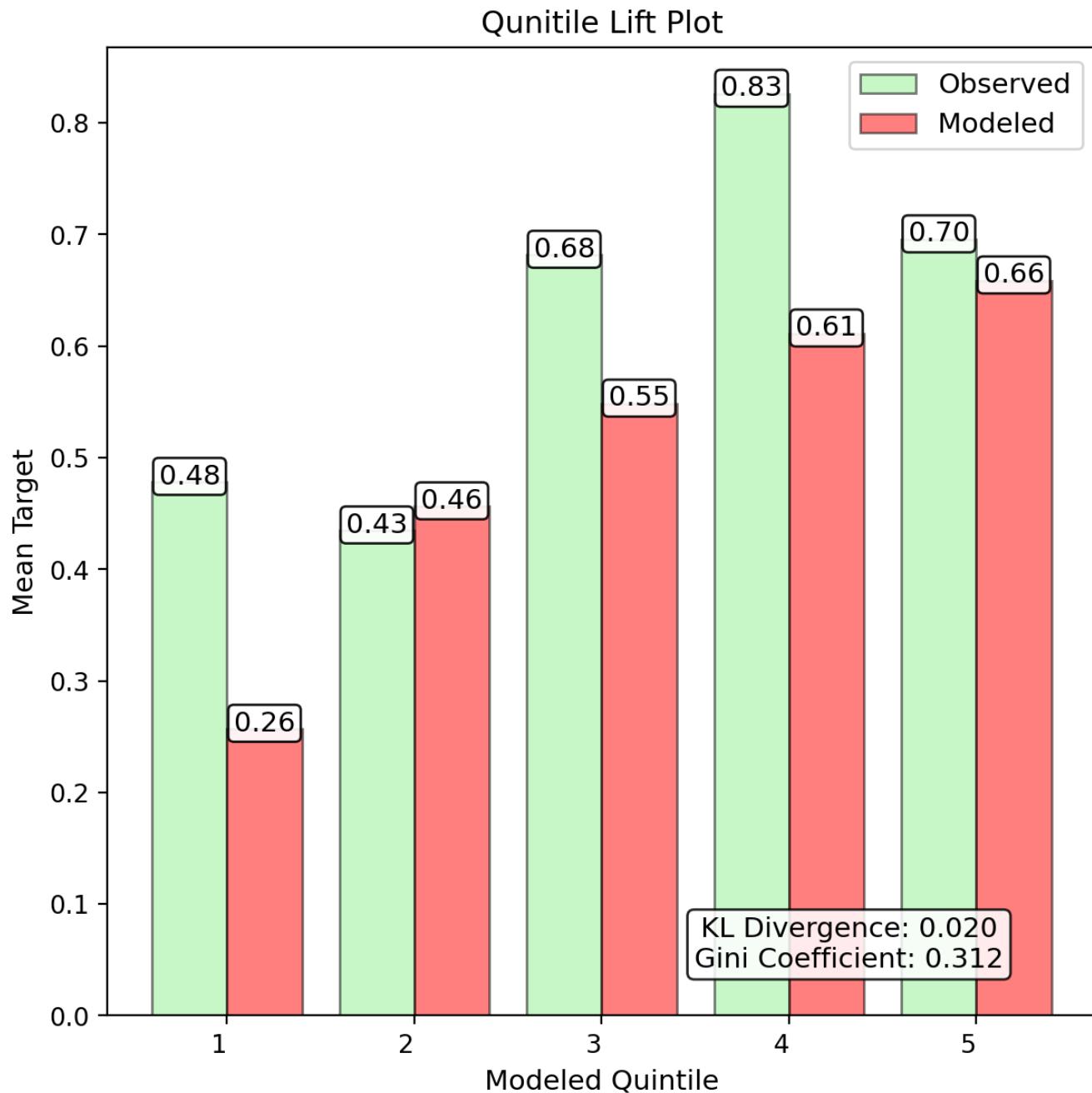


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Compactness Error - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

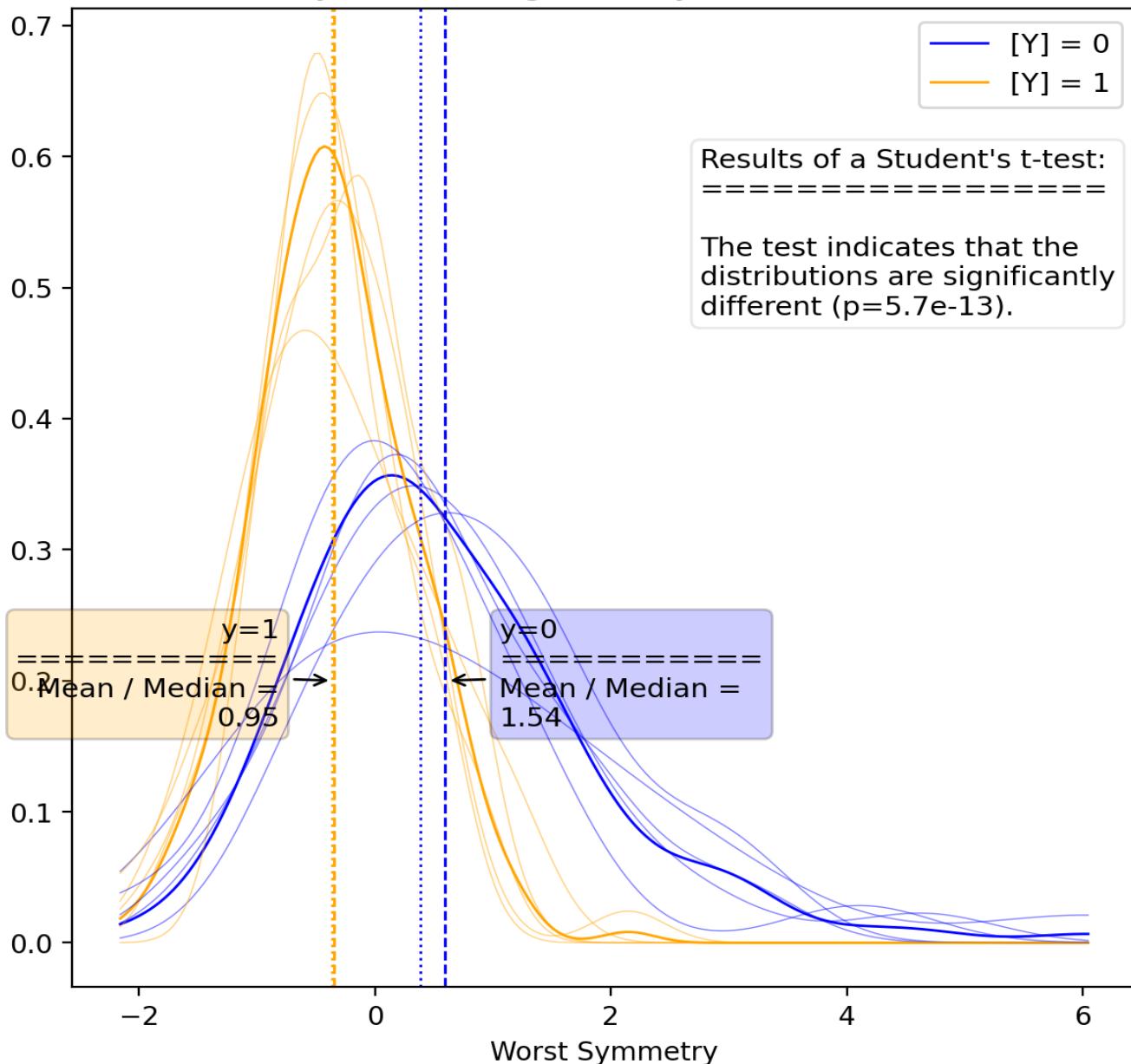
Worst Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-1.10	-1.32	-1.28	-1.29	-1.24	-1.24	0.08
Fitted p-Value	2.0e-09	2.9e-11	1.0e-10	2.0e-10	3.4e-11	6.9e-13	8.4e-10
Fitted Std. Err.	0.184	0.198	0.198	0.202	0.187	0.173	0.008
Conf. Int. Lower	-1.46	-1.71	-1.67	-1.68	-1.61	-1.58	0.10
Conf. Int. Upper	-0.744	-0.929	-0.892	-0.891	-0.874	-0.904	0.071
Train Accuracy	67.2%	70.2%	69.4%	69.3%	69.6%	69.2%	1.2%
Val Accuracy	77.1%	63.8%	67.1%	67.6%	67.3%	62.3%	5.0%
Train AUC	66.1%	69.2%	68.6%	68.2%	68.8%	68.3%	1.2%
Val AUC	77.4%	63.8%	64.9%	67.2%	64.4%	62.8%	5.7%
Train F1	72.6%	76.1%	74.1%	74.8%	74.9%	74.6%	1.3%
Test F1	81.8%	64.8%	75.5%	72.1%	74.3%	66.7%	6.2%
Train Precision	74.7%	80.1%	75.5%	77.7%	78.3%	77.4%	2.2%
Val Precision	87.8%	65.7%	82.2%	75.6%	70.3%	74.1%	8.9%
Train Recall	70.7%	72.5%	72.7%	72.2%	71.8%	72.0%	0.8%
Val Recall	76.6%	63.9%	69.8%	68.9%	78.8%	60.6%	6.1%
Train MCC	31.8%	37.2%	36.9%	35.7%	36.7%	35.8%	2.2%
Val MCC	52.3%	27.5%	27.3%	33.8%	30.1%	24.9%	10.5%
Train Log-Loss	11.84	10.73	11.03	11.07	10.96	11.10	0.42
Val Log-Loss	8.24	13.06	11.85	11.69	11.80	13.60	1.81

Univariate Report

Worst Symmetry - Kernel Density Plot

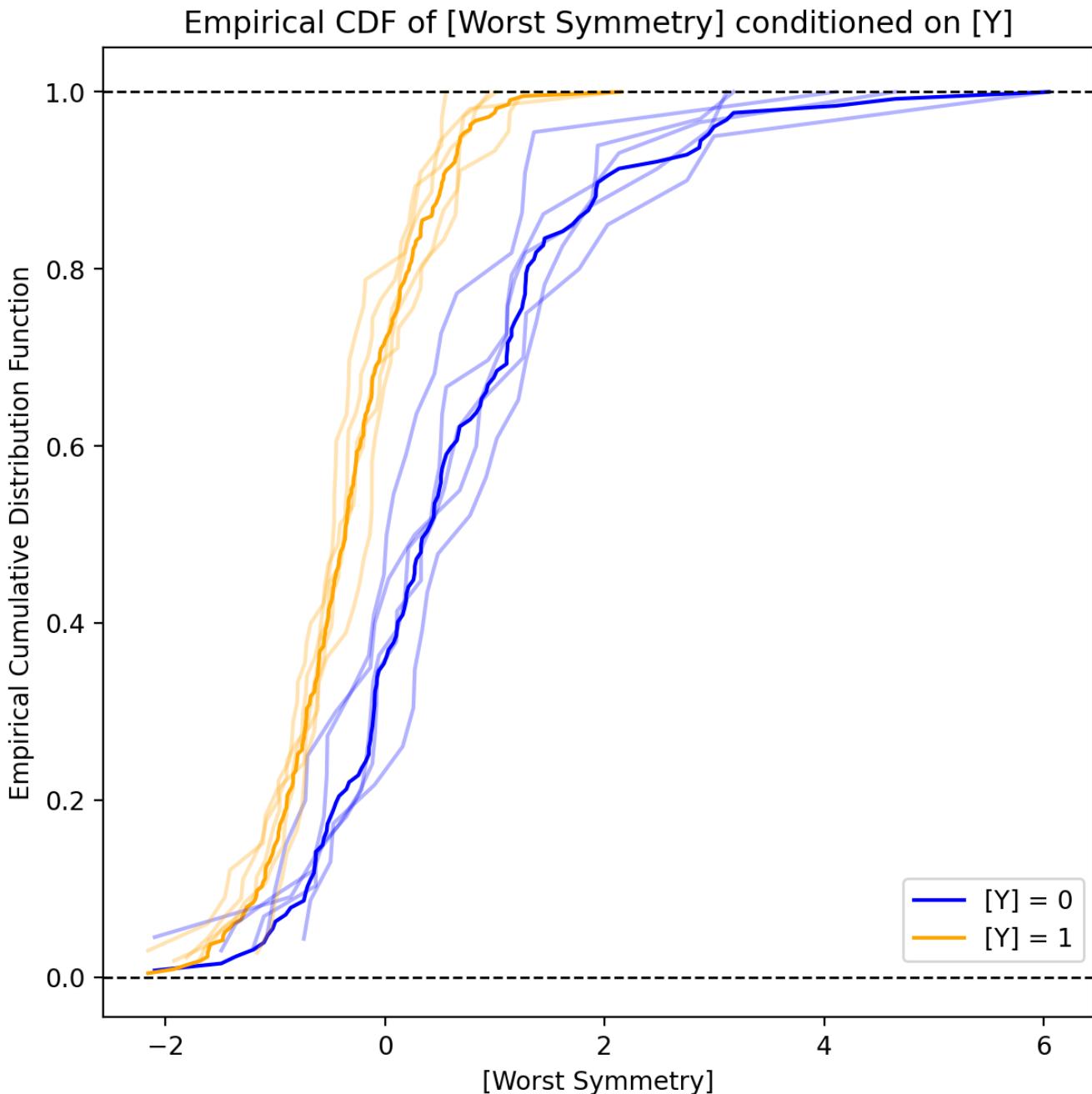
Kernel Density Plot of [Worst Symmetry] by [Y].
Distributions by level are significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

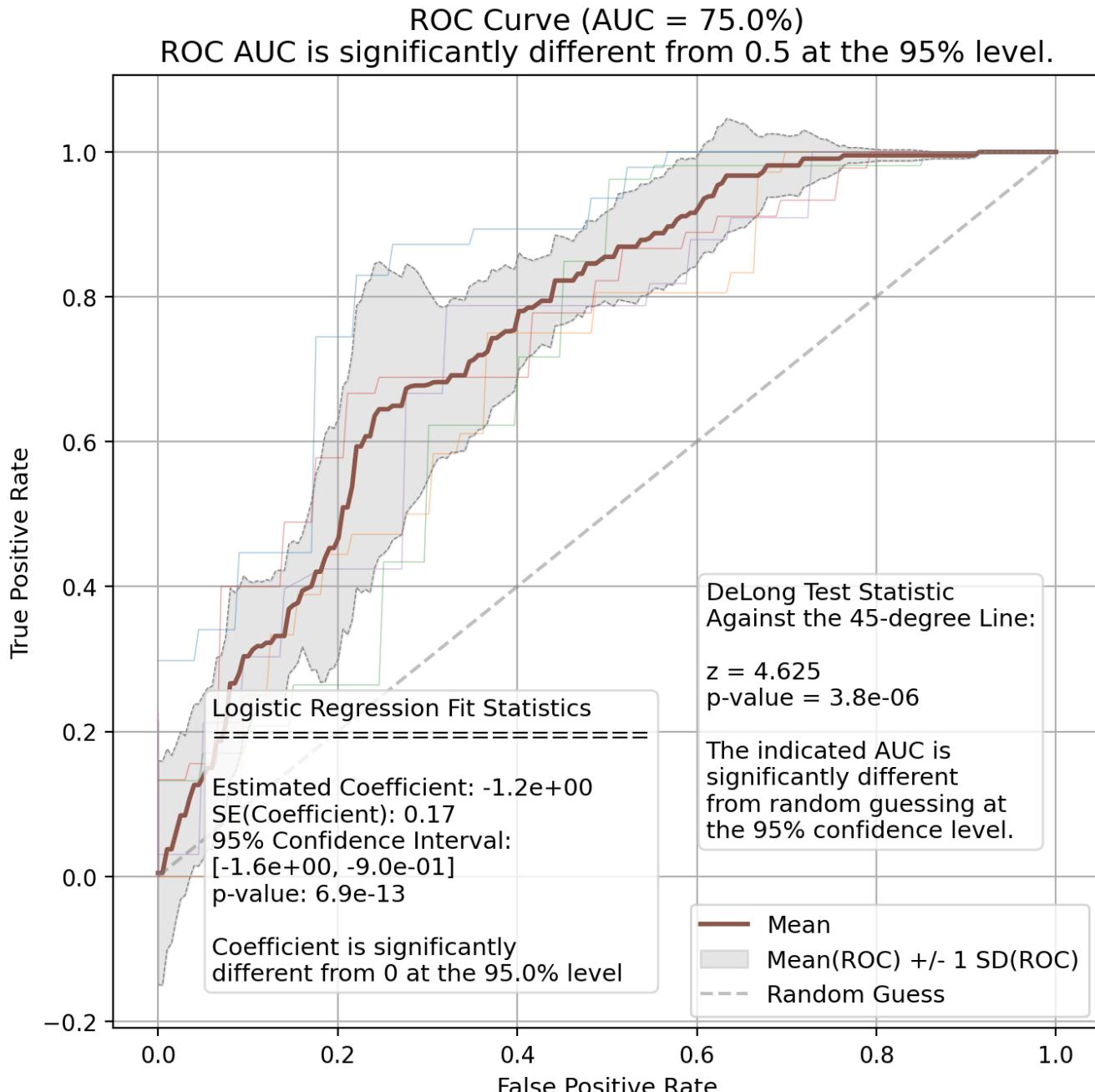
Worst Symmetry - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Worst Symmetry - ROC Curve

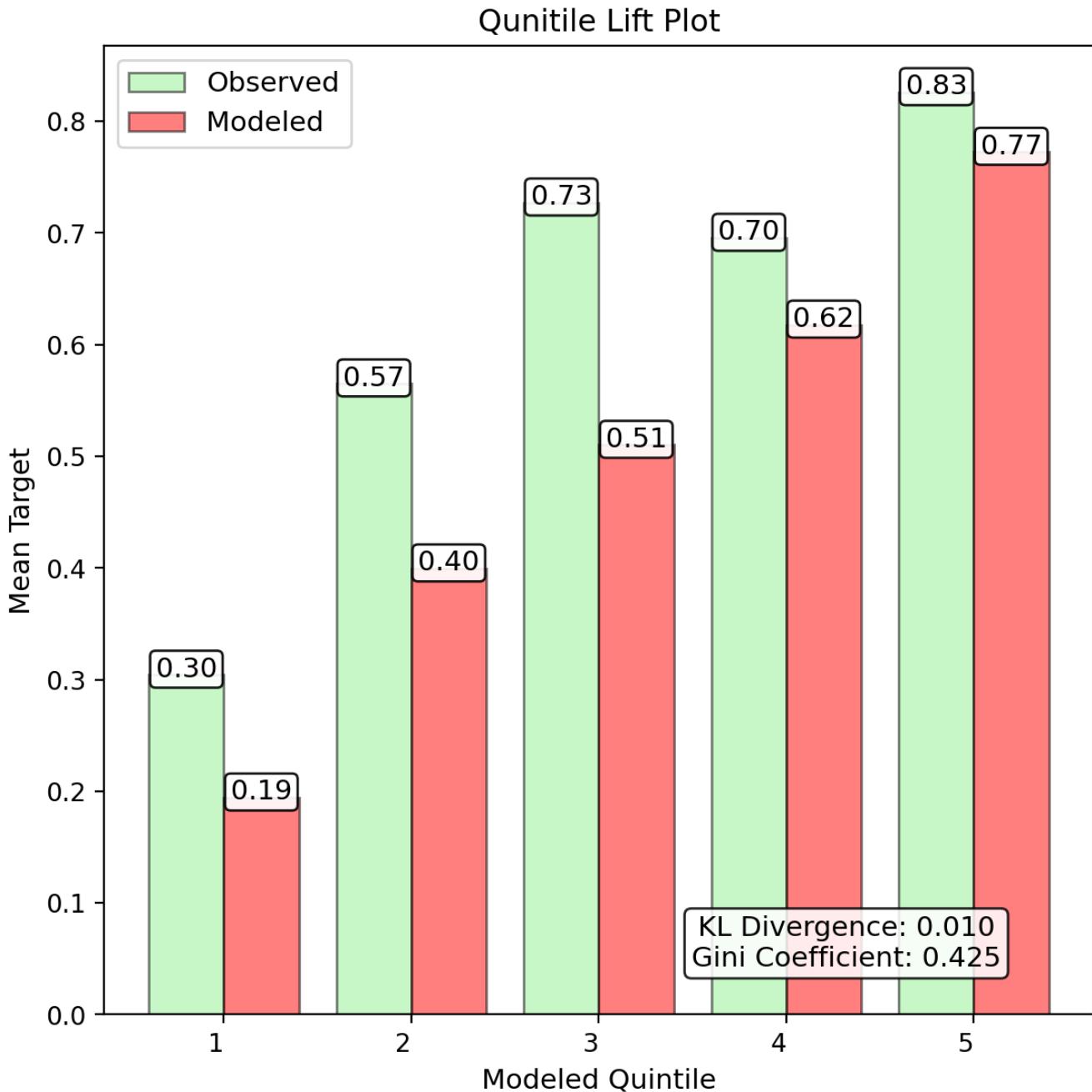


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst Symmetry - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

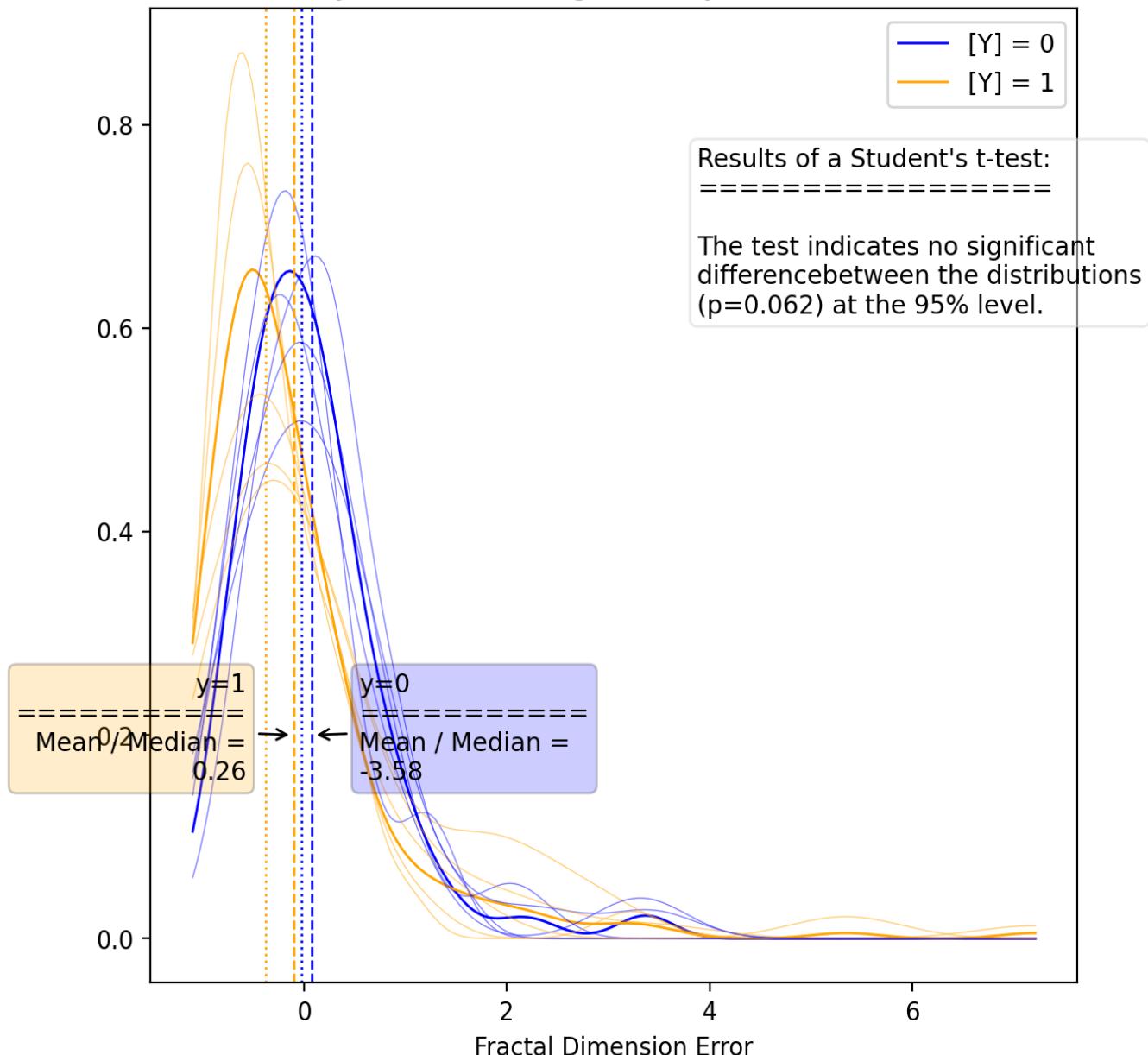
Fractal Dimension Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.125	-0.349	-0.413	-0.129	-0.248	-0.235	0.129
Fitted p-Value	3.3e-01	3.3e-02	1.5e-02	3.3e-01	9.1e-02	7.2e-02	1.6e-01
Fitted Std. Err.	0.129	0.164	0.169	0.132	0.146	0.130	0.018
Conf. Int. Lower	-0.378	-0.672	-0.745	-0.387	-0.534	-0.490	0.165
Conf. Int. Upper	1.3e-01	-2.7e-02	-8.1e-02	1.3e-01	3.9e-02	2.1e-02	9.4e-02
Train Accuracy	62.0%	62.5%	63.1%	59.9%	62.9%	63.0%	1.3%
Val Accuracy	62.9%	53.6%	60.3%	68.9%	65.5%	62.3%	5.8%
Train AUC	59.7%	60.1%	61.1%	56.6%	61.1%	60.8%	1.9%
Val AUC	54.6%	53.7%	60.2%	67.1%	61.4%	61.0%	5.5%
Train F1	69.3%	70.3%	69.7%	68.6%	69.9%	70.3%	0.7%
Test F1	74.0%	54.3%	68.8%	74.7%	74.0%	68.6%	8.6%
Train Precision	69.0%	72.9%	68.7%	68.0%	71.9%	71.0%	2.2%
Val Precision	69.8%	55.9%	80.0%	73.9%	67.5%	71.2%	8.9%
Train Recall	69.5%	68.0%	70.8%	69.2%	68.0%	69.6%	1.2%
Val Recall	78.7%	52.8%	60.4%	75.6%	81.8%	66.2%	12.6%
Train MCC	19.5%	19.6%	22.4%	13.2%	21.9%	21.5%	3.7%
Val MCC	10.0%	7.3%	18.3%	34.4%	25.0%	21.6%	11.1%
Train Log-Loss	13.70	13.52	13.31	14.44	13.36	13.32	0.46
Val Log-Loss	13.39	16.72	14.32	11.20	12.45	13.60	2.08

Univariate Report

Fractal Dimension Error - Kernel Density Plot

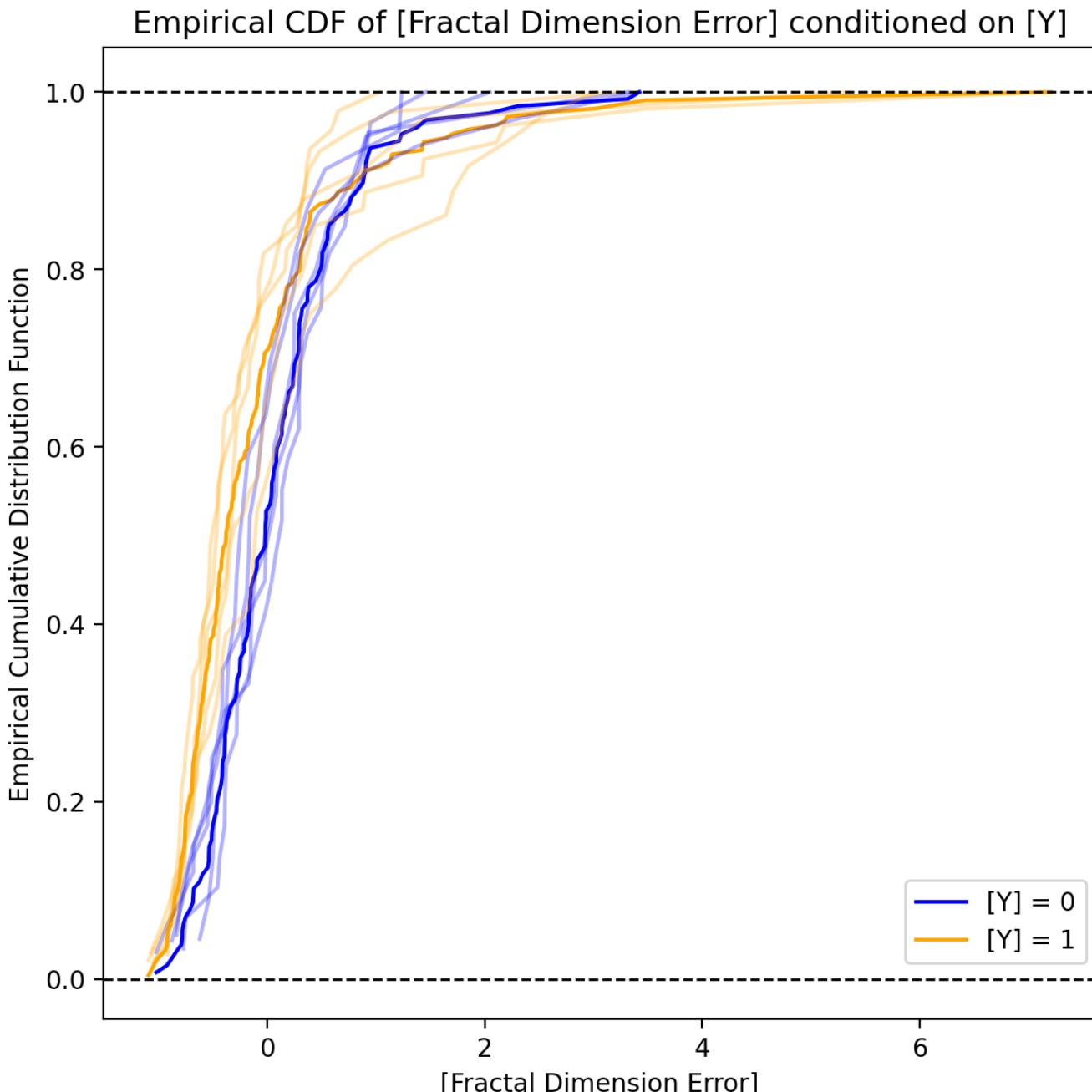
Kernel Density Plot of [Fractal Dimension Error] by [Y].
Distributions by level are not significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

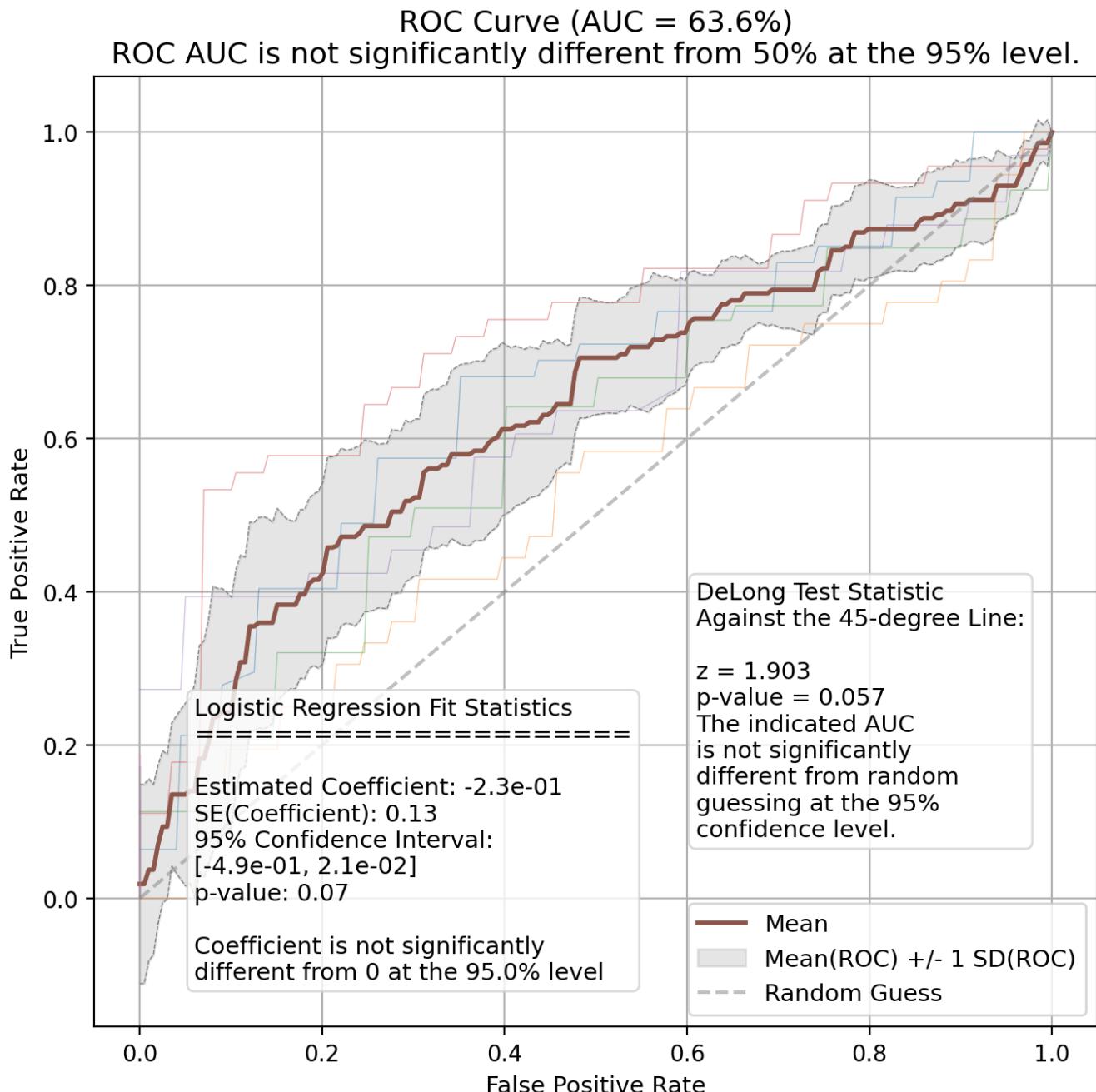
Fractal Dimension Error - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Fractal Dimension Error - ROC Curve

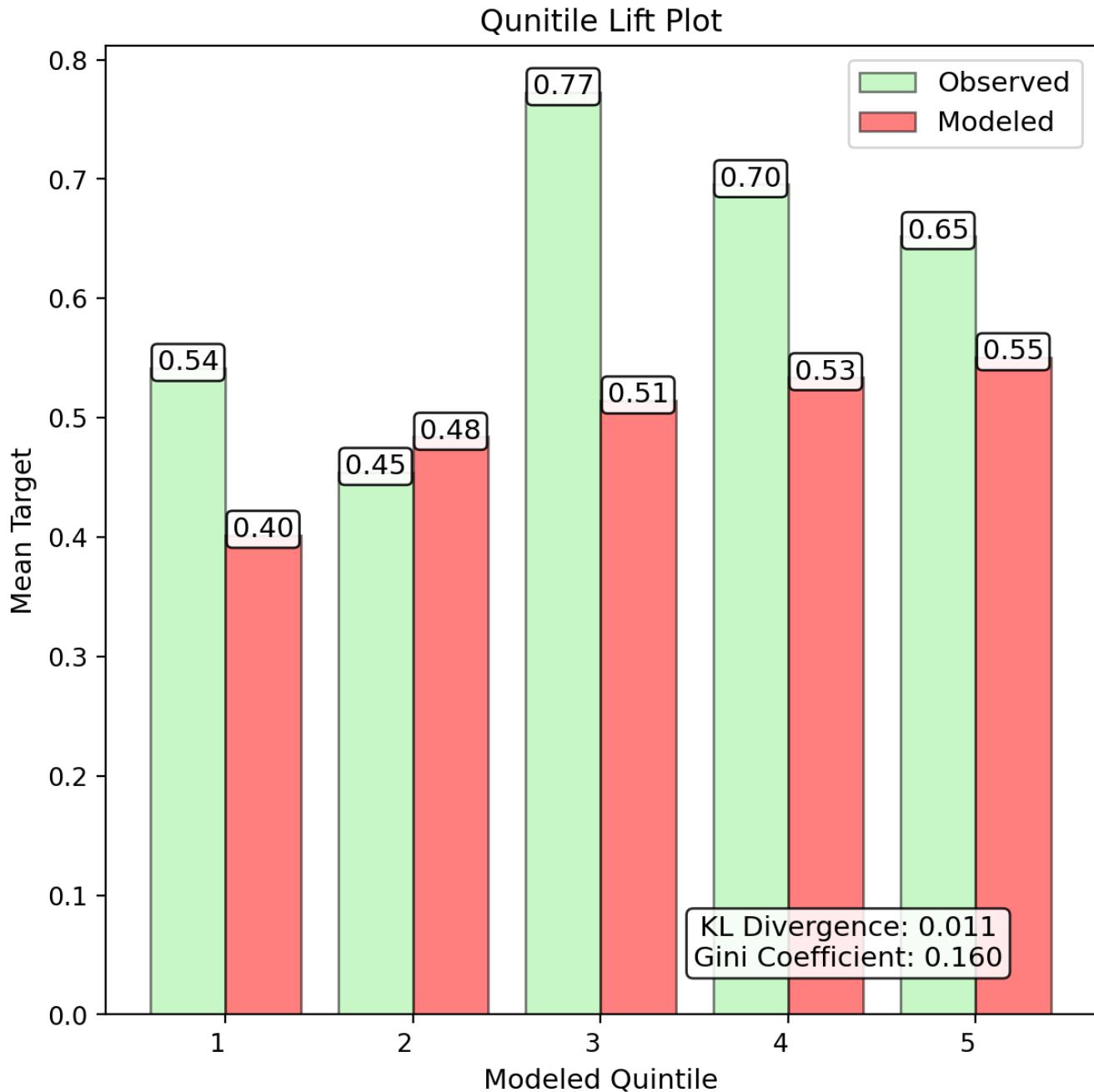


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Fractal Dimension Error - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

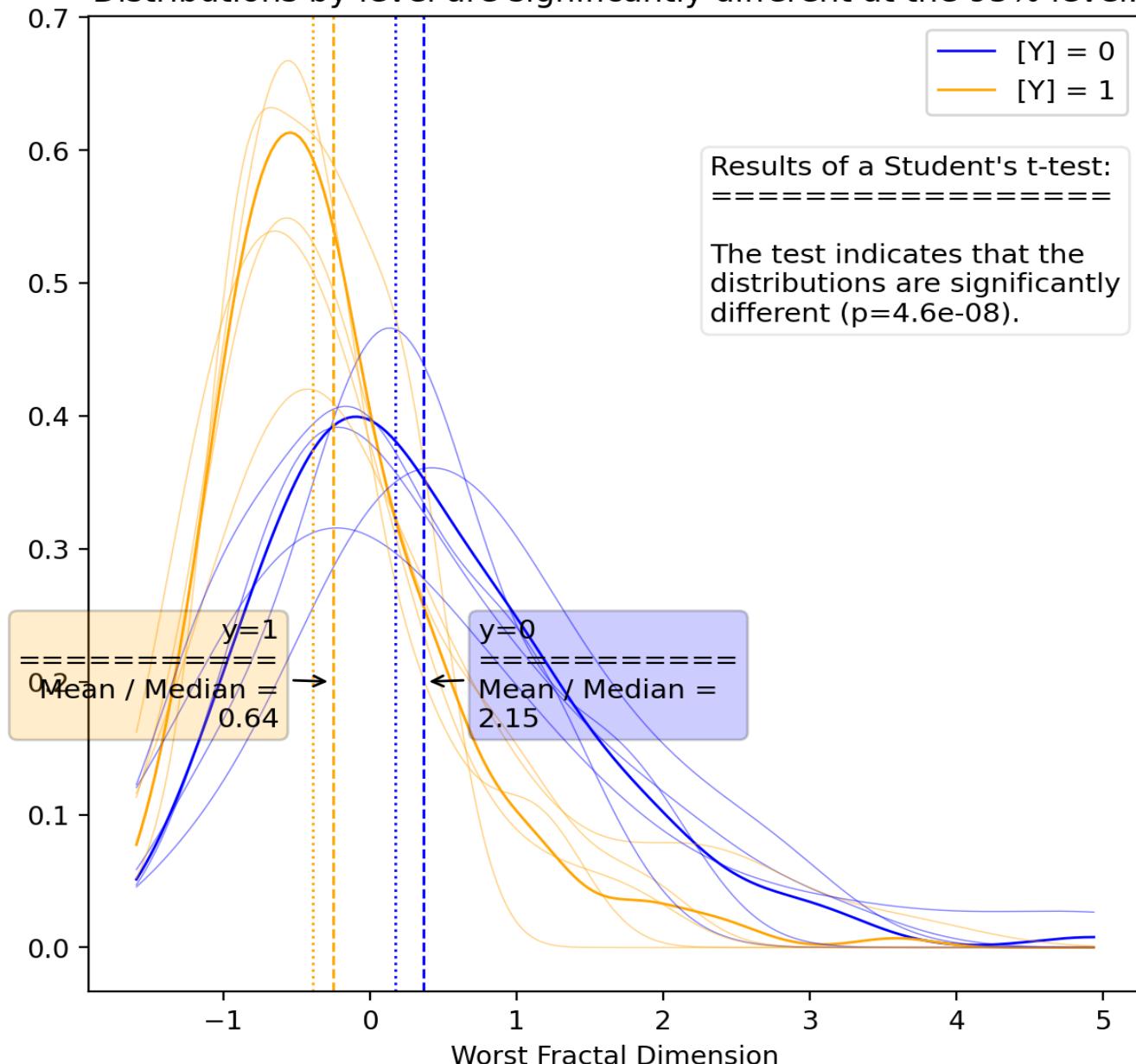
Worst Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.646	-1.051	-0.837	-0.609	-0.798	-0.776	0.176
Fitted p-Value	1.2e-05	6.6e-09	3.9e-07	7.5e-05	4.0e-07	6.1e-08	3.2e-05
Fitted Std. Err.	0.147	0.181	0.165	0.154	0.157	0.143	0.013
Conf. Int. Lower	-0.93	-1.41	-1.16	-0.91	-1.11	-1.06	0.20
Conf. Int. Upper	-0.357	-0.696	-0.513	-0.308	-0.489	-0.496	0.152
Train Accuracy	66.4%	68.0%	68.3%	65.2%	68.2%	67.4%	1.4%
Val Accuracy	71.4%	58.0%	64.4%	77.0%	63.6%	61.4%	7.4%
Train AUC	64.6%	66.5%	67.0%	63.3%	66.7%	65.7%	1.6%
Val AUC	69.8%	57.8%	59.9%	76.2%	60.6%	60.3%	7.8%
Train F1	72.7%	74.5%	73.5%	71.9%	74.2%	73.6%	1.1%
Test F1	77.8%	60.3%	74.0%	80.9%	71.4%	67.6%	7.9%
Train Precision	72.9%	77.9%	73.8%	73.5%	76.2%	74.9%	2.1%
Val Precision	81.4%	59.5%	78.7%	81.8%	67.6%	70.8%	9.9%
Train Recall	72.5%	71.3%	73.3%	70.4%	72.4%	72.4%	1.1%
Val Recall	74.5%	61.1%	69.8%	80.0%	75.8%	64.8%	7.2%
Train MCC	29.1%	32.1%	34.0%	26.2%	32.8%	31.2%	3.2%
Val MCC	38.3%	15.7%	18.5%	52.1%	22.1%	20.2%	15.5%
Train Log-Loss	12.10	11.53	11.43	12.55	11.47	11.73	0.49
Val Log-Loss	10.30	15.15	12.84	8.28	13.11	13.91	2.67

Univariate Report

Worst Fractal Dimension - Kernel Density Plot

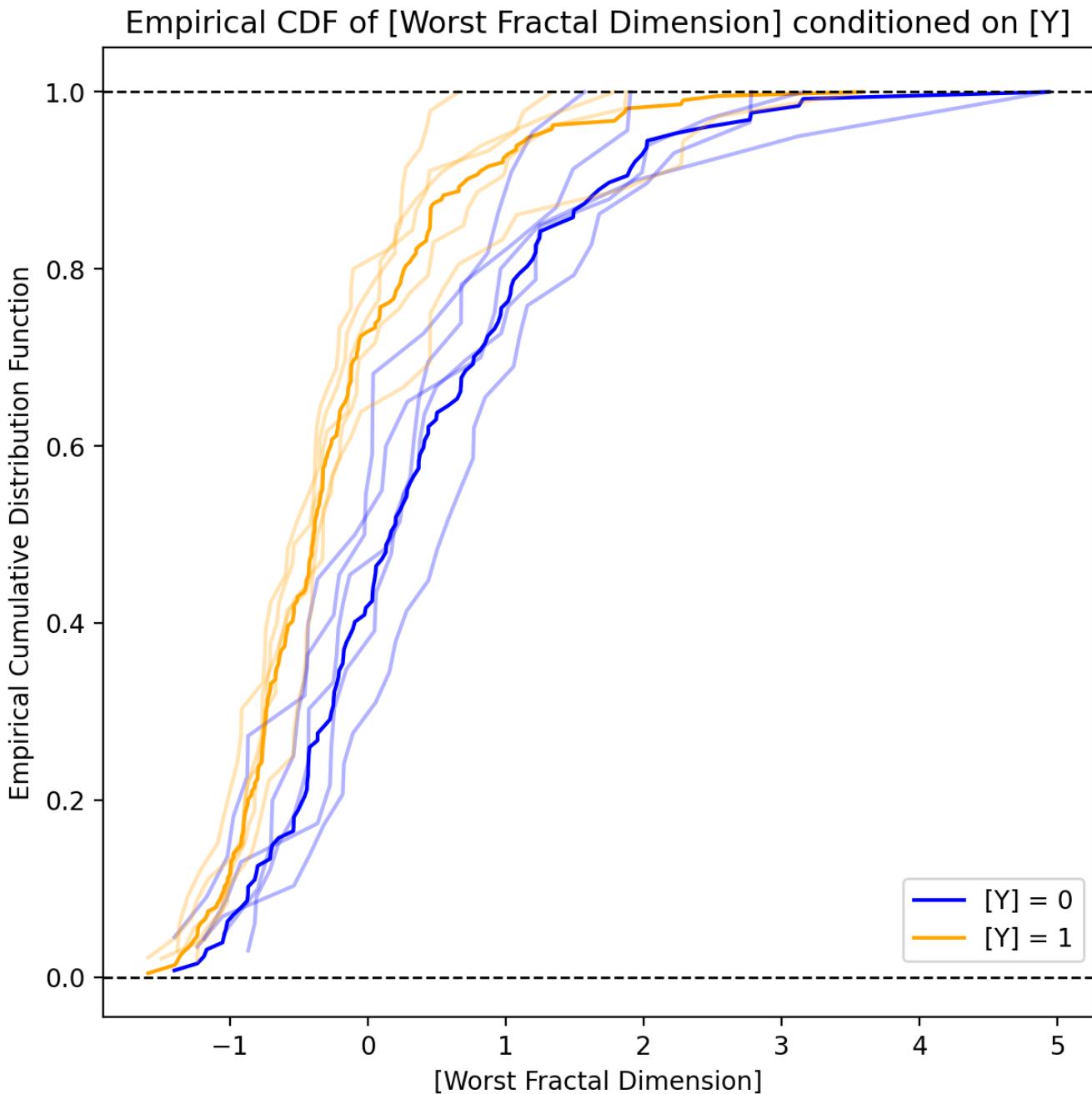
Kernel Density Plot of [Worst Fractal Dimension] by [Y].
Distributions by level are significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

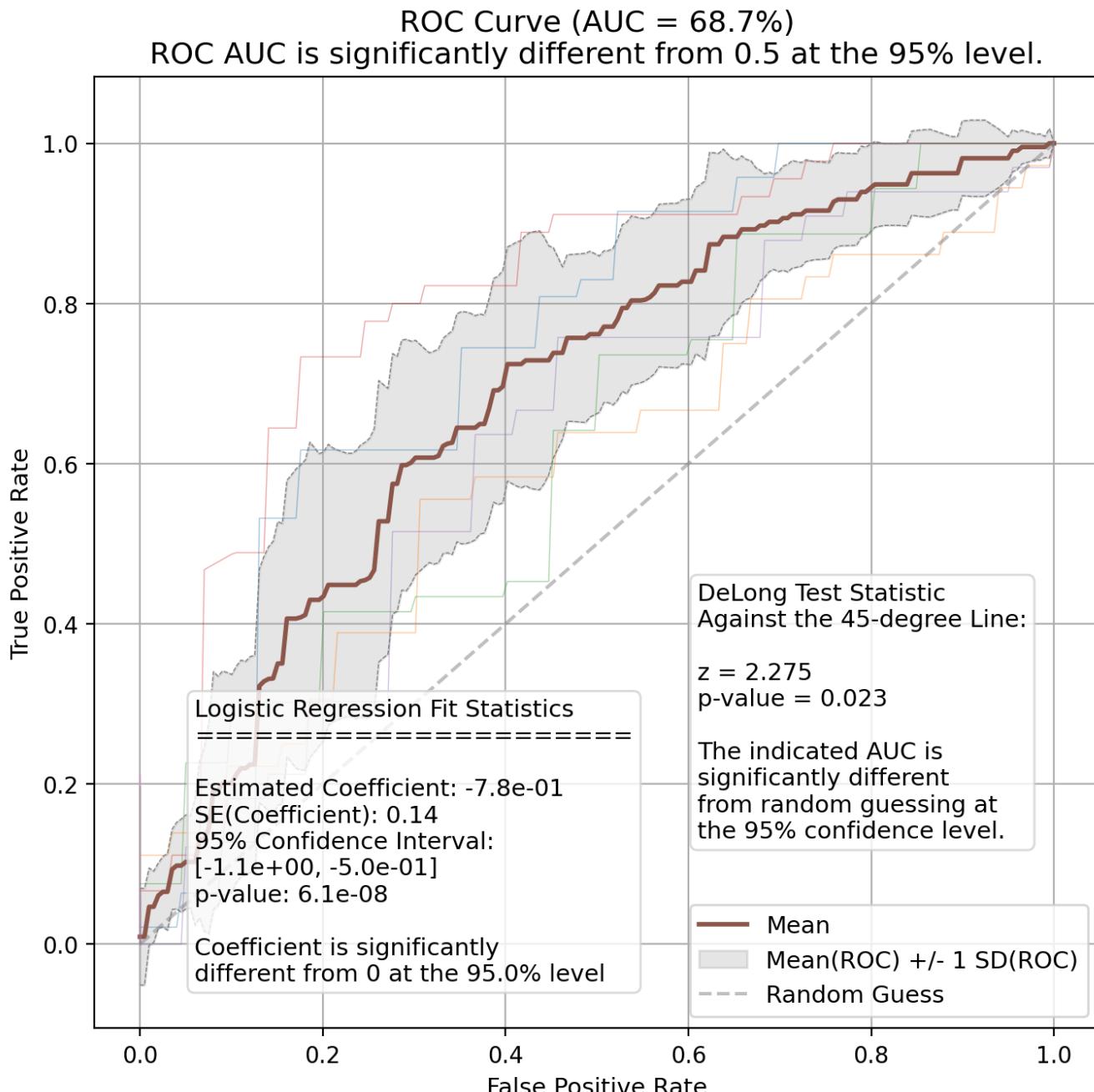
Worst Fractal Dimension - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Worst Fractal Dimension - ROC Curve

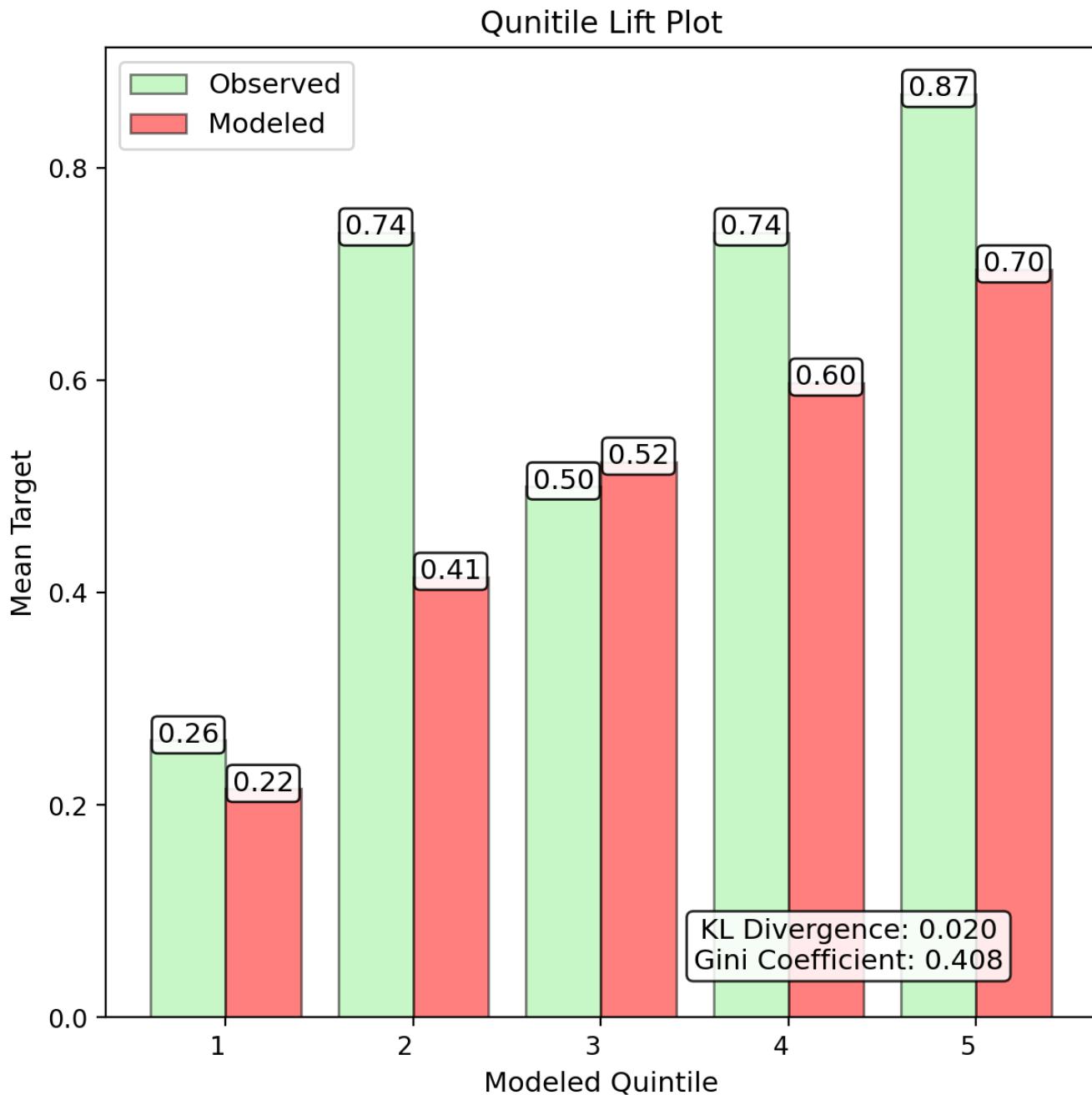


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst Fractal Dimension - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

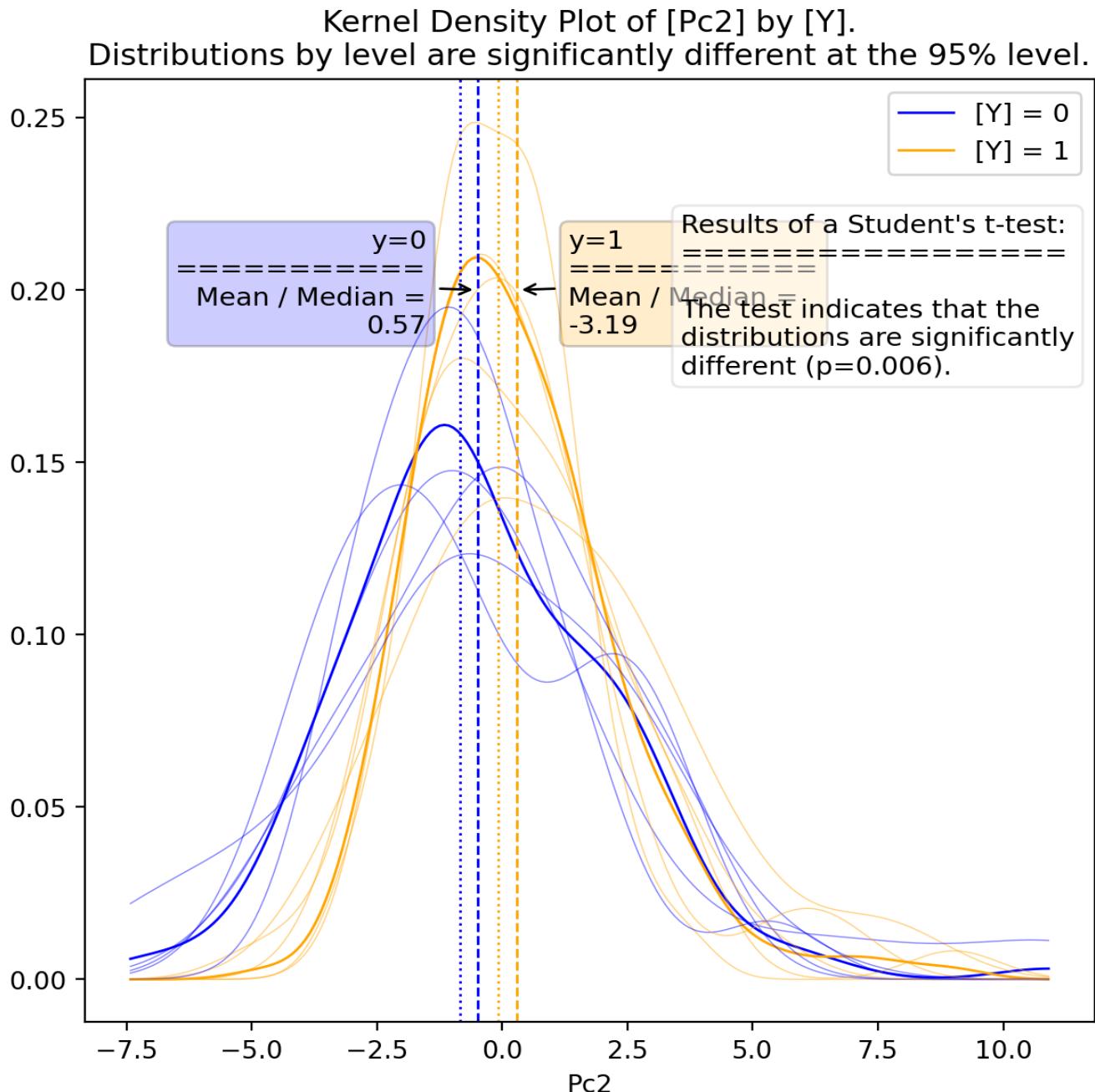
Univariate Report

Pc2 - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	0.160	0.098	0.155	0.149	0.134	0.140	0.025
Fitted p-Value	3.0e-03	8.3e-02	6.3e-03	1.0e-02	1.1e-02	4.9e-03	3.4e-02
Fitted Std. Err.	5.4e-02	5.7e-02	5.7e-02	5.8e-02	5.3e-02	5.0e-02	2.0e-03
Conf. Int. Lower	5.4e-02	-1.3e-02	4.4e-02	3.5e-02	3.0e-02	4.2e-02	2.6e-02
Conf. Int. Upper	0.266	0.209	0.266	0.262	0.238	0.237	0.025
Train Accuracy	55.4%	53.3%	55.2%	53.6%	53.5%	54.0%	1.0%
Val Accuracy	48.6%	58.0%	52.1%	55.4%	56.4%	57.0%	3.8%
Train AUC	57.4%	55.3%	56.3%	55.6%	55.1%	55.6%	1.0%
Val AUC	48.4%	58.2%	56.1%	56.6%	59.1%	56.3%	4.3%
Train F1	57.2%	57.8%	57.7%	56.6%	57.2%	57.2%	0.5%
Test F1	56.1%	56.7%	58.8%	58.2%	55.6%	63.2%	1.4%
Train Precision	69.8%	70.7%	66.7%	69.2%	68.5%	68.6%	1.5%
Val Precision	65.7%	61.3%	78.1%	67.6%	71.4%	67.7%	6.3%
Train Recall	48.5%	48.9%	50.9%	47.9%	49.2%	49.1%	1.1%
Val Recall	48.9%	52.8%	47.2%	51.1%	45.5%	59.2%	2.9%
Train MCC	14.6%	10.1%	12.4%	10.9%	9.8%	11.0%	2.0%
Val MCC	-3.0%	16.5%	10.9%	12.9%	18.3%	12.3%	8.4%
Train Log-Loss	16.09	16.83	16.14	16.74	16.76	16.59	0.36
Val Log-Loss	18.54	15.15	17.28	16.07	15.73	15.49	1.36

Univariate Report

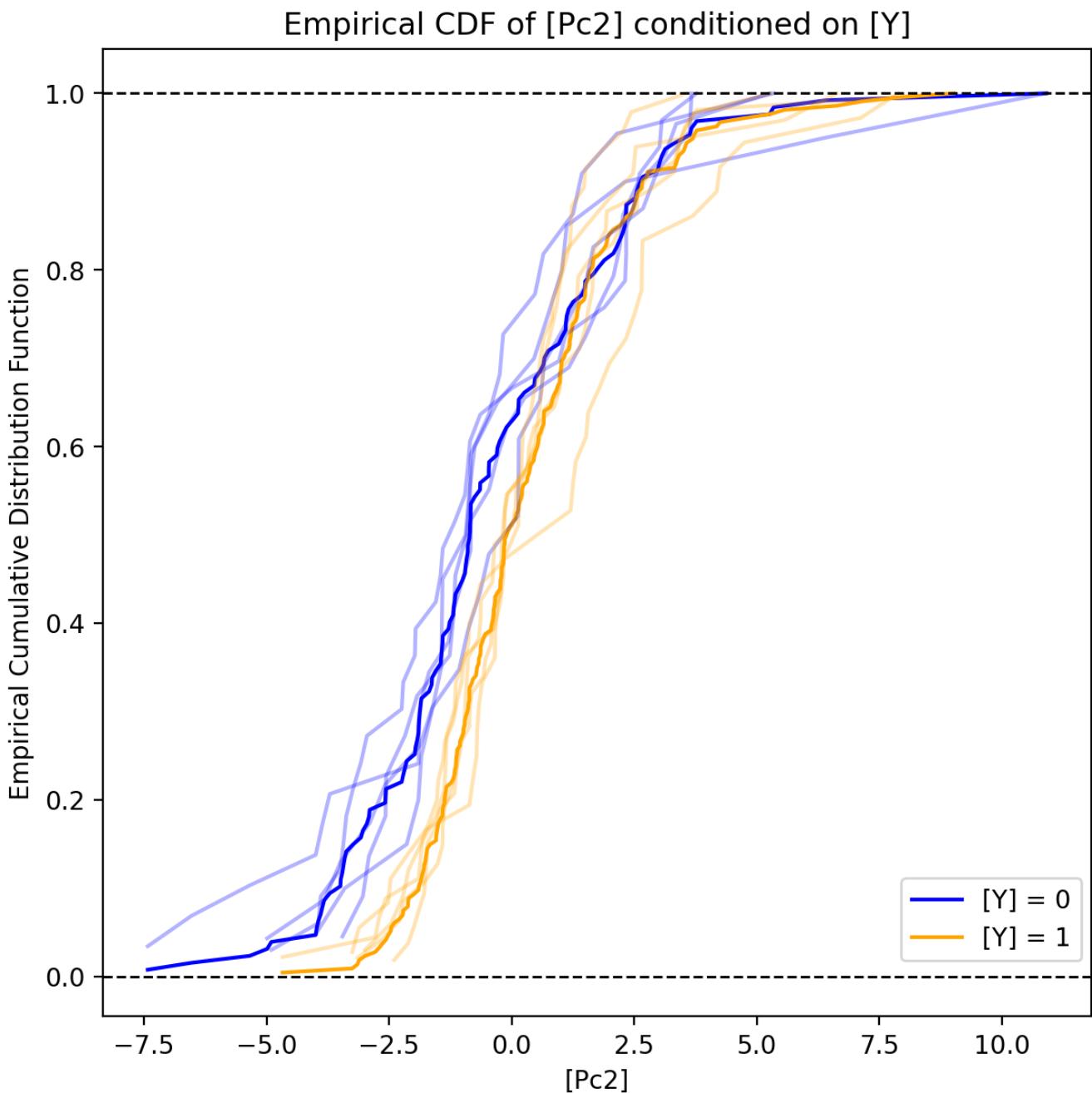
Pc2 - Kernel Density Plot



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

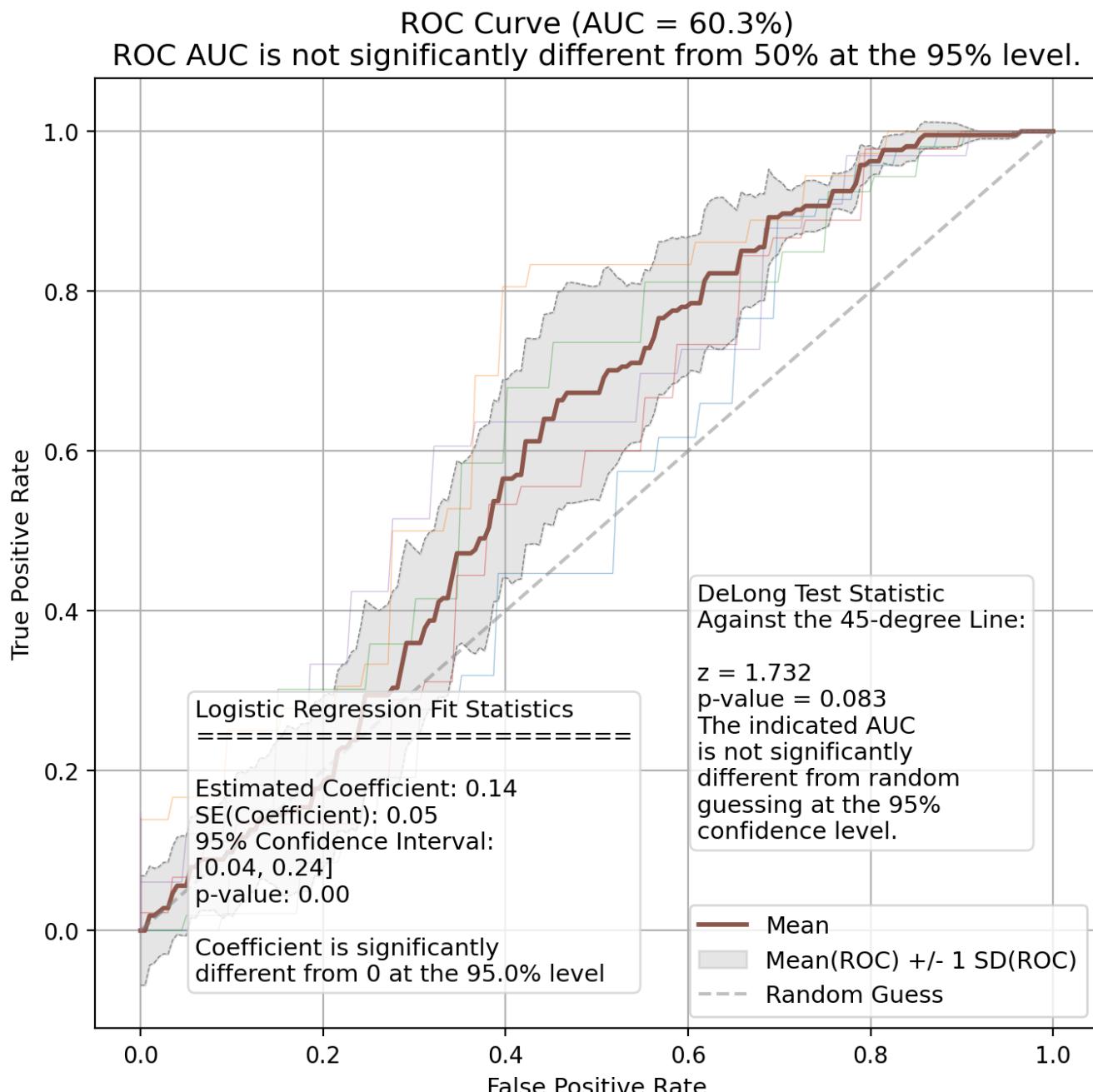
Pc2 - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Pc2 - ROC Curve

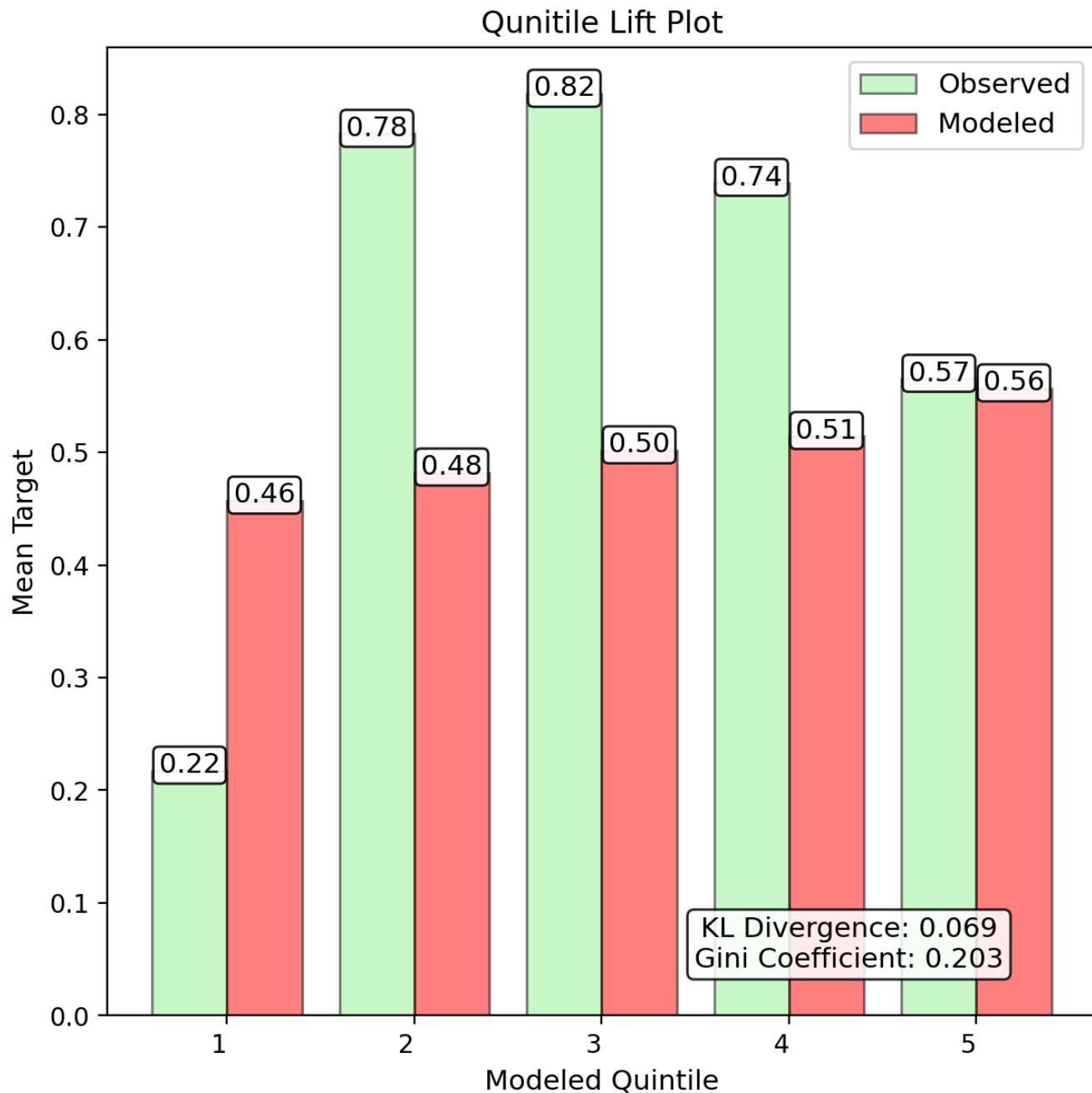


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Pc2 - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

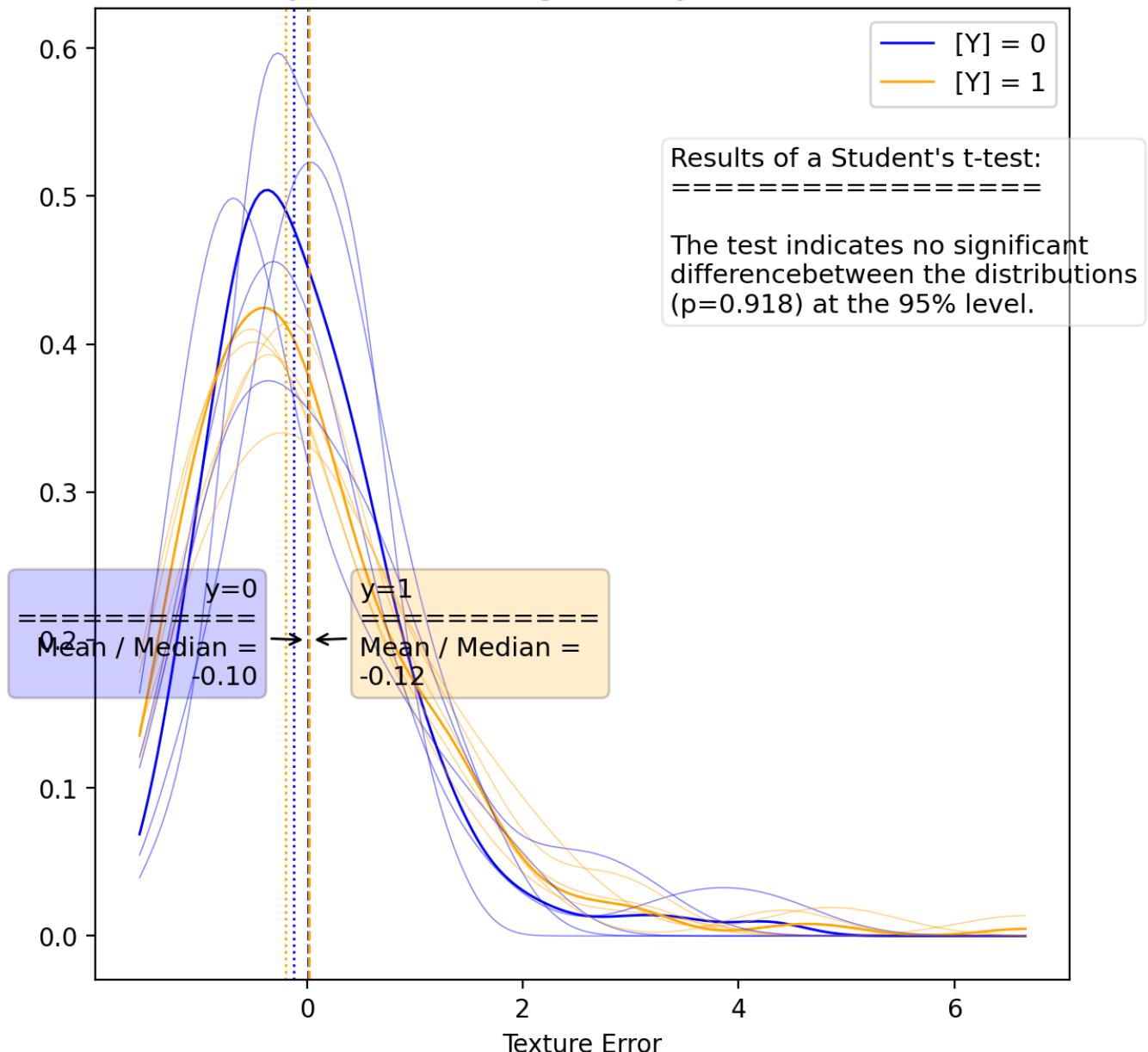
Texture Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	2.7e-03	9.0e-03	3.6e-02	4.4e-02	-2.8e-03	1.9e-02	2.1e-02
Fitted p-Value	0.982	0.938	0.747	0.695	0.981	0.857	0.137
Fitted Std. Err.	0.119	0.116	0.112	0.113	0.115	0.103	0.003
Conf. Int. Lower	-0.231	-0.218	-0.183	-0.177	-0.228	-0.183	0.026
Conf. Int. Upper	0.237	0.236	0.255	0.265	0.222	0.220	0.017
Train Accuracy	45.8%	45.6%	45.9%	46.8%	53.8%	46.0%	3.5%
Val Accuracy	44.3%	50.7%	46.6%	40.5%	49.1%	55.3%	4.0%
Train AUC	47.8%	48.1%	48.2%	49.6%	51.3%	48.5%	1.5%
Val AUC	50.7%	51.3%	47.6%	43.1%	48.5%	57.2%	3.2%
Train F1	46.9%	49.0%	44.9%	48.2%	62.5%	47.4%	7.0%
Test F1	43.5%	45.2%	55.2%	38.9%	54.8%	57.9%	7.2%
Train Precision	59.1%	63.4%	57.8%	62.9%	64.3%	61.0%	2.9%
Val Precision	68.2%	53.8%	70.6%	51.9%	58.6%	70.0%	8.4%
Train Recall	38.9%	39.9%	36.6%	39.1%	60.8%	38.8%	10.0%
Val Recall	31.9%	38.9%	45.3%	31.1%	51.5%	49.3%	8.7%
Train MCC	-4.3%	-3.6%	-3.6%	-0.7%	2.6%	-2.9%	2.9%
Val MCC	1.5%	2.6%	-4.2%	-13.9%	-3.0%	14.1%	6.5%
Train Log-Loss	19.55	19.61	19.50	19.17	16.64	19.45	1.27
Val Log-Loss	20.08	17.76	19.26	21.43	18.35	16.12	1.45

Univariate Report

Texture Error - Kernel Density Plot

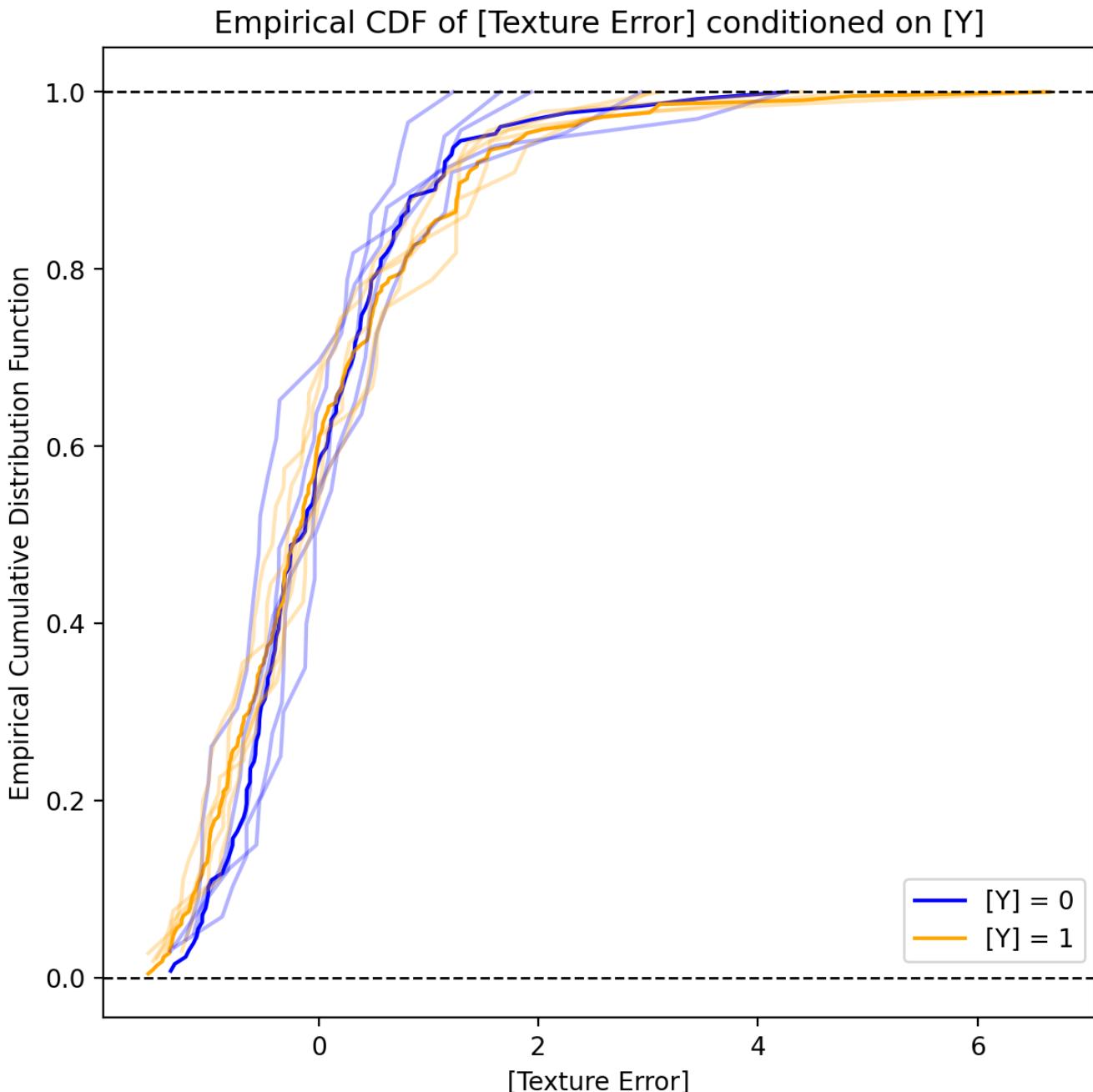
Kernel Density Plot of [Texture Error] by [Y].
Distributions by level are not significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

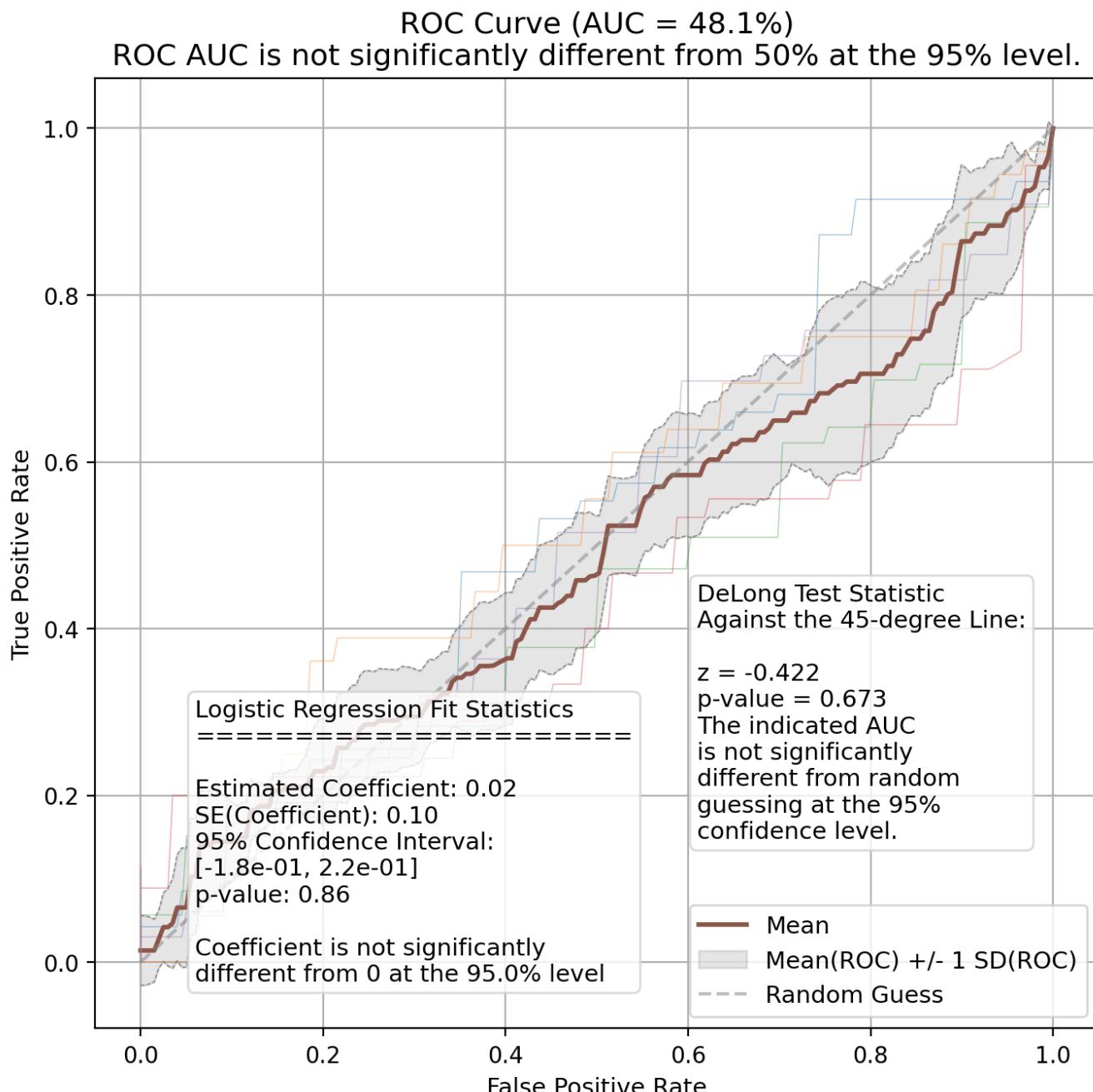
Texture Error - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Texture Error - ROC Curve

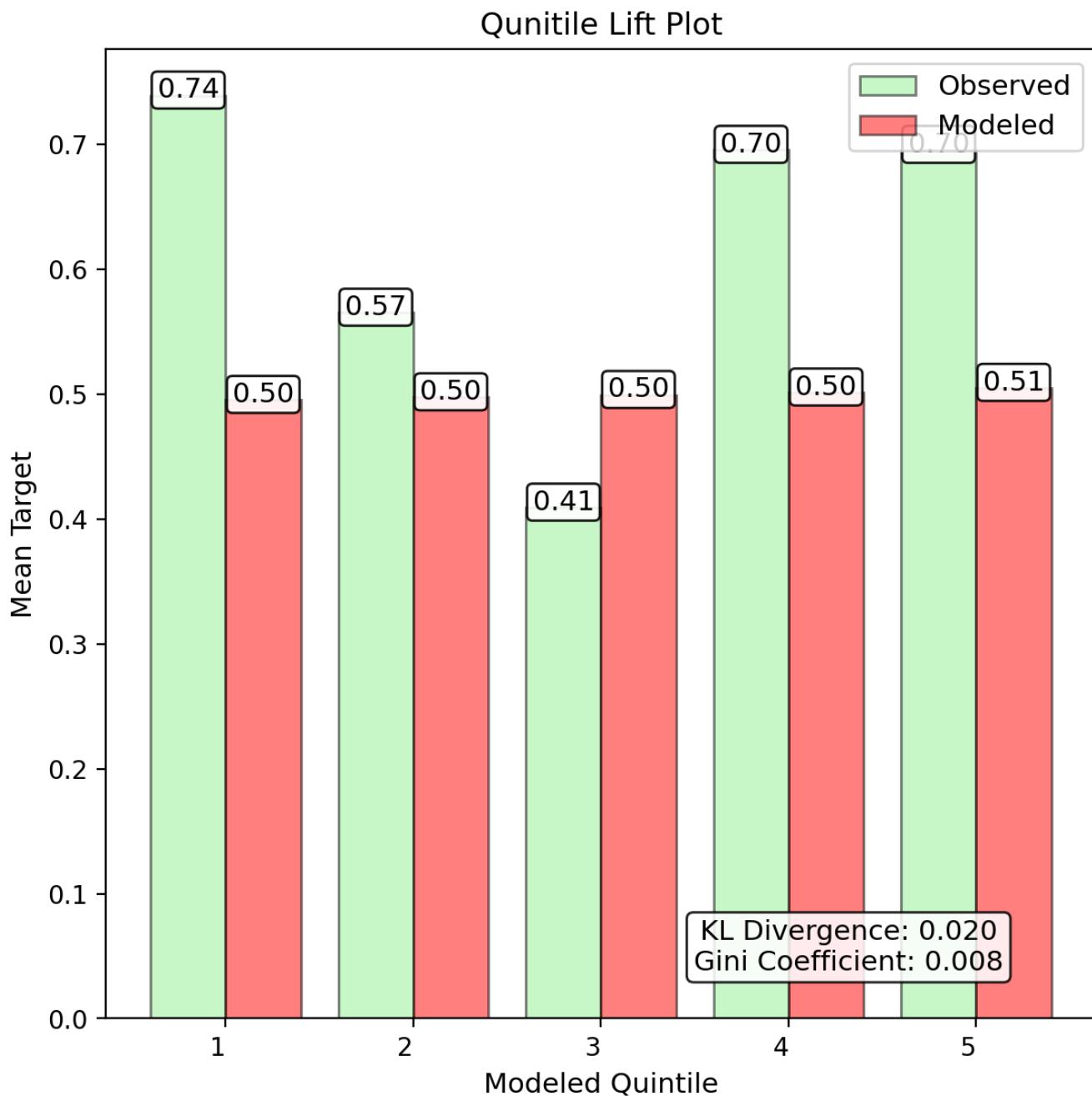


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Texture Error - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

Univariate Report

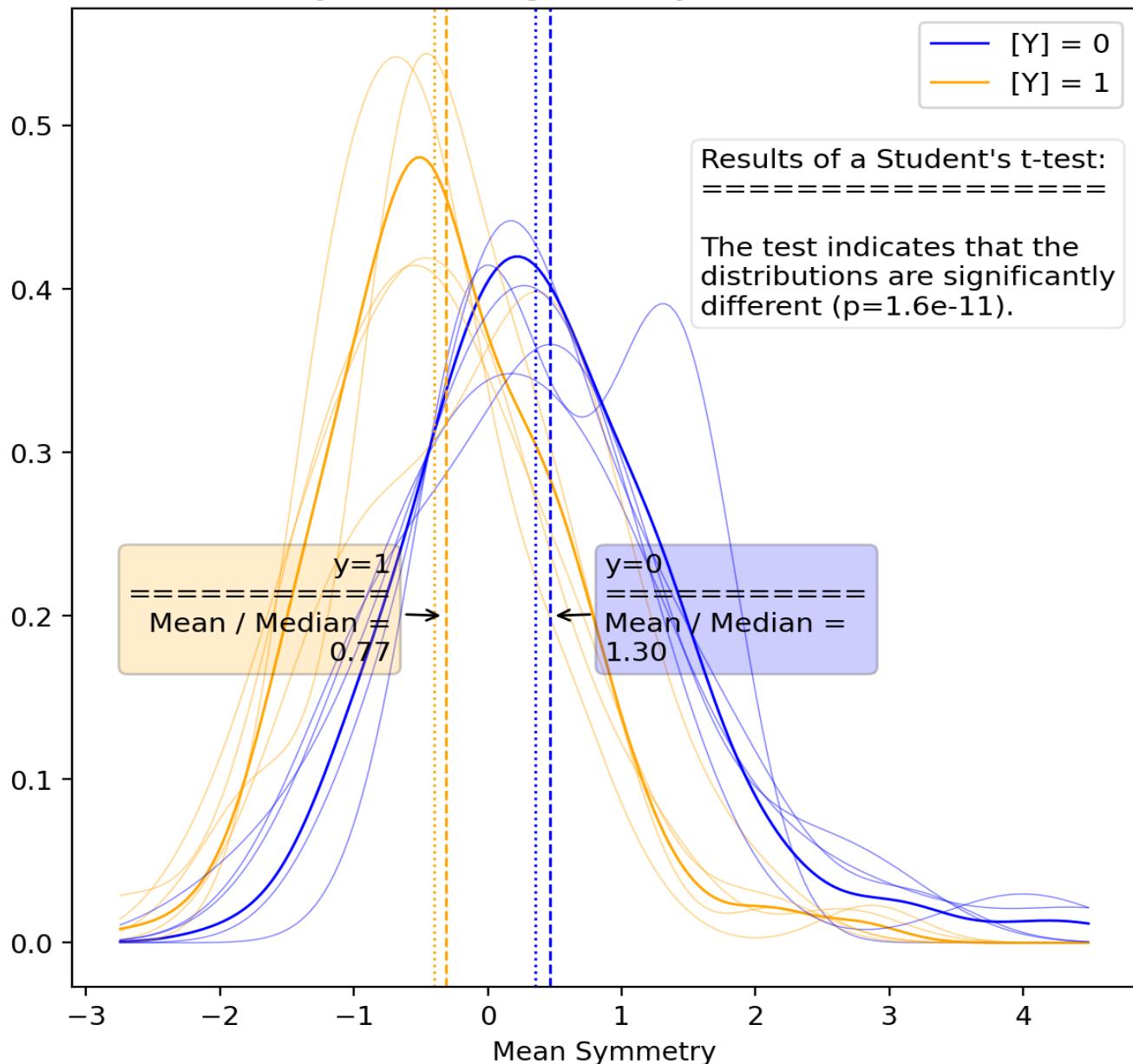
Mean Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.758	-1.084	-0.897	-0.849	-0.914	-0.896	0.119
Fitted p-Value	5.1e-07	3.9e-10	1.5e-08	1.0e-07	5.7e-09	3.3e-10	2.2e-07
Fitted Std. Err.	0.151	0.173	0.159	0.160	0.157	0.143	0.008
Conf. Int. Lower	-1.05	-1.42	-1.21	-1.16	-1.22	-1.18	0.13
Conf. Int. Upper	-0.462	-0.744	-0.586	-0.537	-0.607	-0.617	0.104
Train Accuracy	63.1%	69.1%	65.7%	64.0%	65.4%	65.1%	2.3%
Val Accuracy	75.7%	53.6%	65.8%	68.9%	67.3%	55.3%	8.0%
Train AUC	63.2%	69.6%	65.5%	63.9%	65.5%	65.2%	2.5%
Val AUC	74.1%	54.2%	65.5%	69.5%	65.9%	57.2%	7.4%
Train F1	67.7%	74.2%	69.9%	69.4%	70.4%	70.0%	2.4%
Test F1	81.3%	48.4%	73.7%	72.3%	72.7%	57.9%	12.5%
Train Precision	73.4%	81.8%	73.8%	75.2%	76.6%	76.0%	3.4%
Val Precision	84.1%	57.7%	83.3%	78.9%	72.7%	70.0%	10.9%
Train Recall	62.9%	68.0%	66.5%	64.5%	65.2%	65.0%	1.9%
Val Recall	78.7%	41.7%	66.0%	66.7%	72.7%	49.3%	14.1%
Train MCC	25.7%	37.5%	30.4%	26.9%	29.9%	29.4%	4.6%
Val MCC	46.9%	8.6%	28.0%	38.2%	31.8%	14.1%	14.3%
Train Log-Loss	13.30	11.13	12.37	12.96	12.48	12.58	0.83
Val Log-Loss	8.75	16.72	12.34	11.20	11.80	16.12	2.89

Univariate Report

Mean Symmetry - Kernel Density Plot

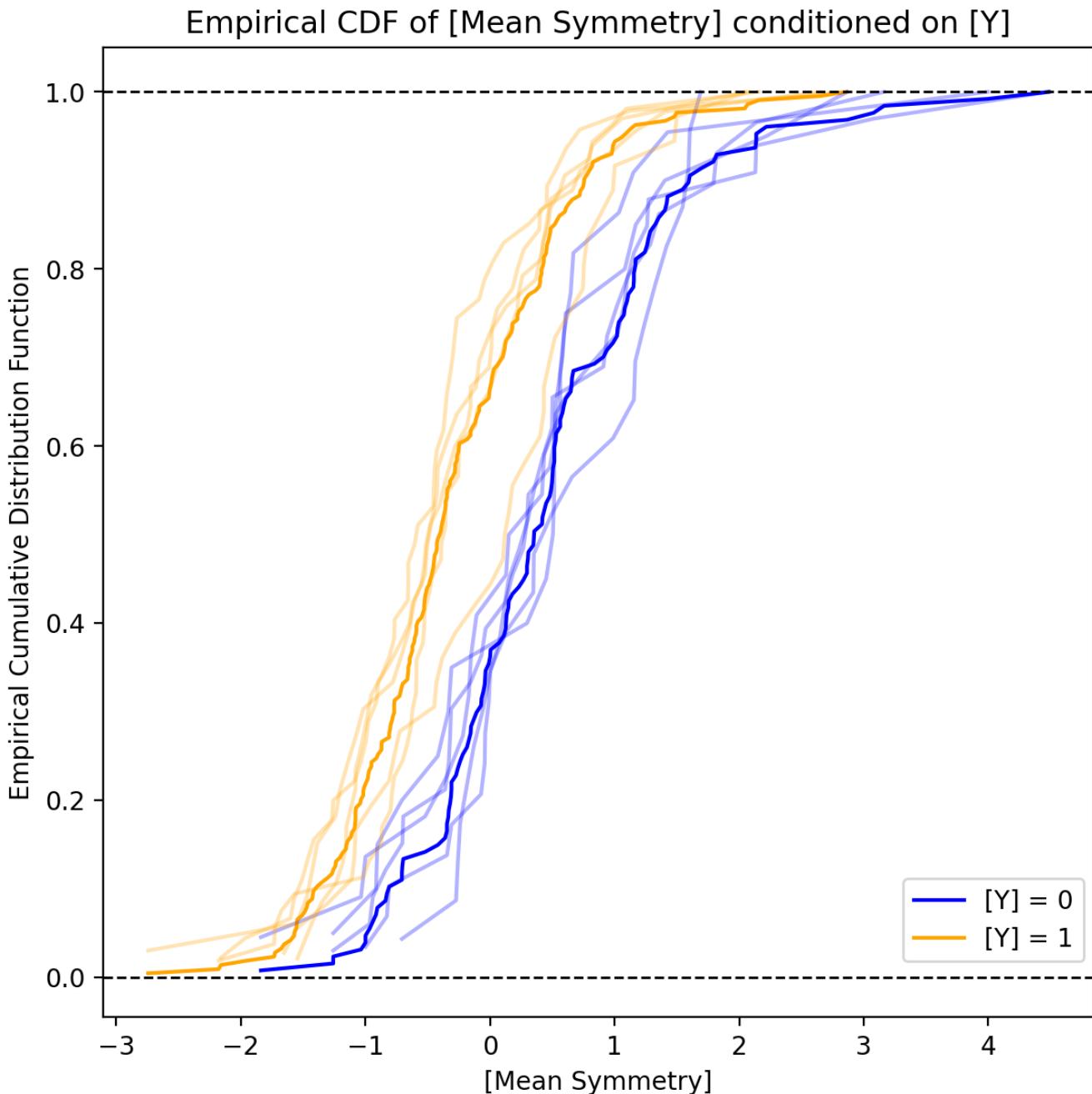
Kernel Density Plot of [Mean Symmetry] by [Y].
Distributions by level are significantly different at the 95% level.



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

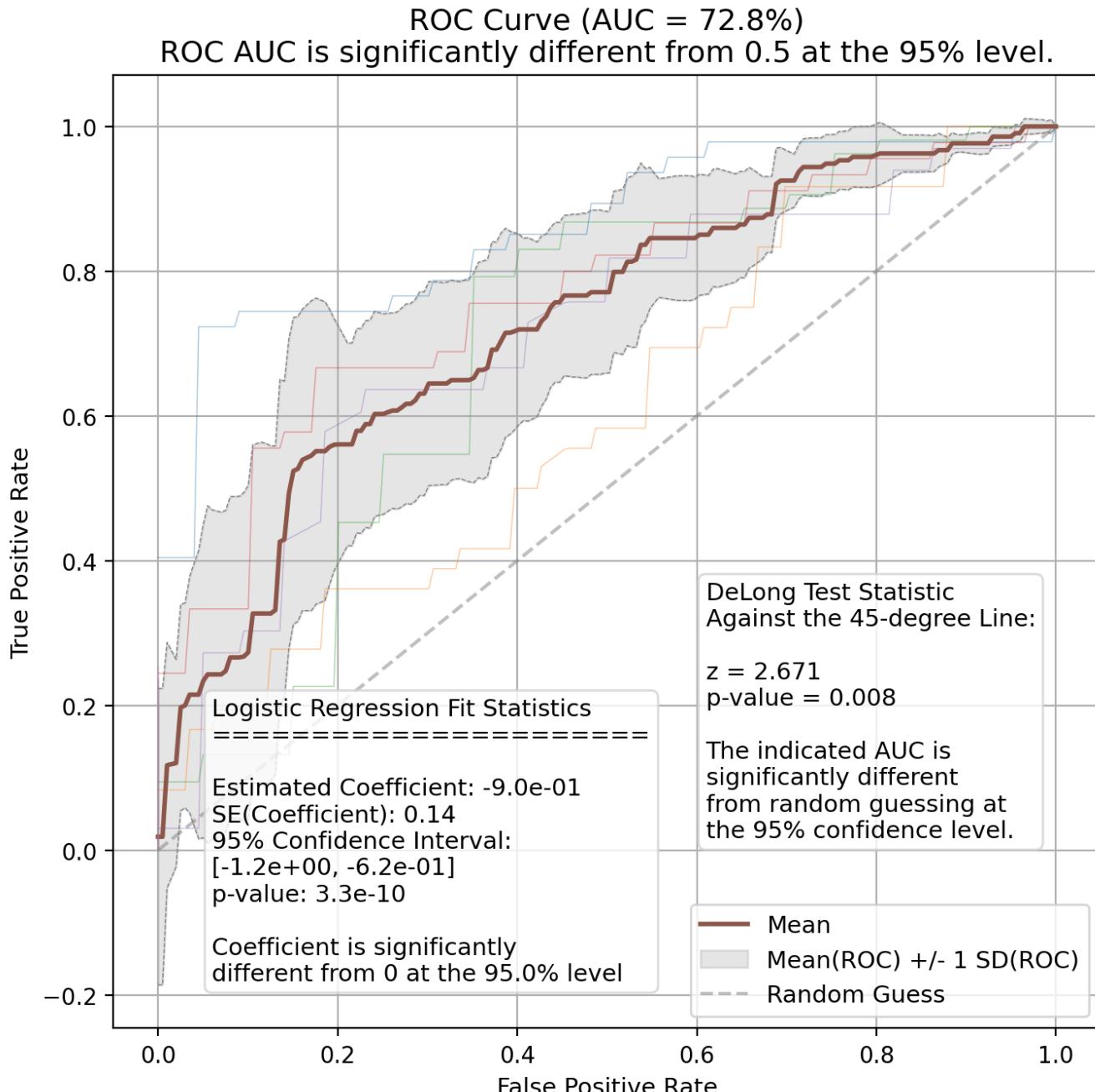
Mean Symmetry - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Mean Symmetry - ROC Curve

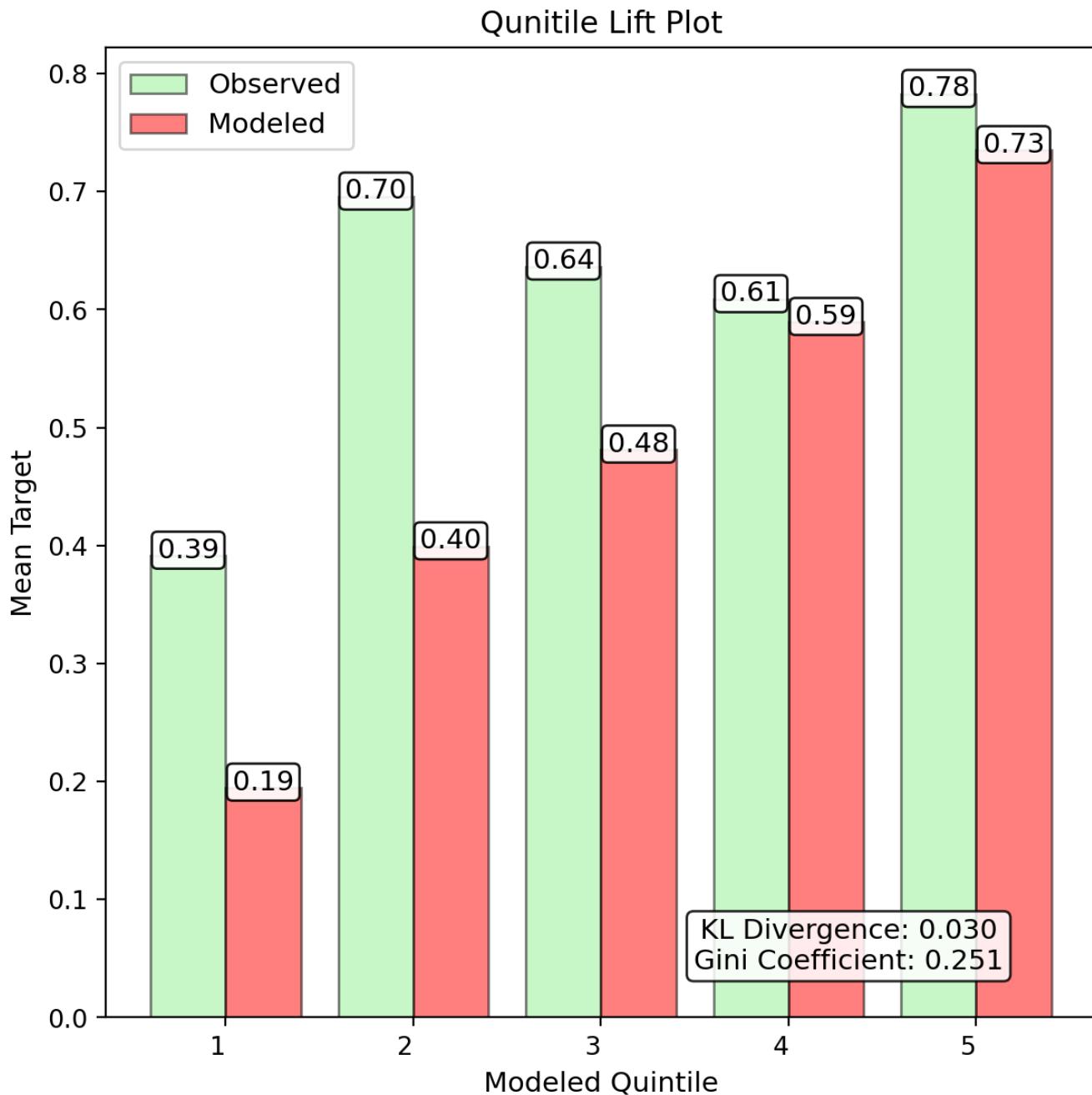


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean Symmetry - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

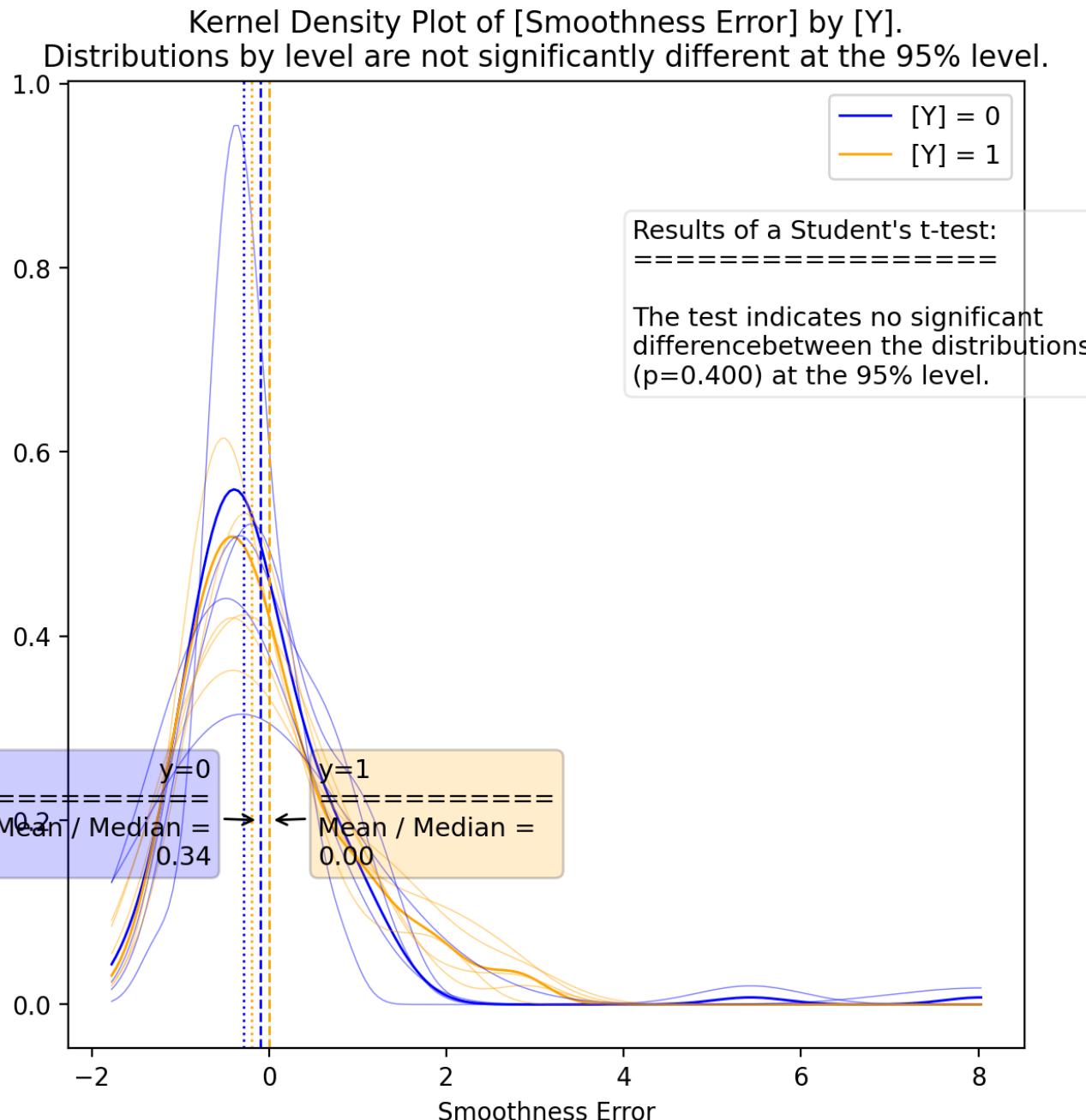
Univariate Report

Smoothness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	1.0e-01	8.1e-03	1.4e-01	-1.4e-02	1.3e-01	7.4e-02	7.2e-02
Fitted p-Value	0.391	0.950	0.254	0.910	0.310	0.503	0.339
Fitted Std. Err.	0.121	0.128	0.121	0.120	0.131	0.110	0.005
Conf. Int. Lower	-0.133	-0.242	-0.099	-0.249	-0.124	-0.142	0.070
Conf. Int. Upper	0.341	0.258	0.376	0.222	0.391	0.290	0.074
Train Accuracy	50.9%	48.9%	53.4%	50.9%	50.7%	50.4%	1.6%
Val Accuracy	48.6%	56.5%	39.7%	44.6%	52.7%	52.6%	6.6%
Train AUC	54.6%	53.2%	56.3%	47.5%	54.3%	54.0%	3.4%
Val AUC	51.7%	57.1%	46.0%	41.0%	55.3%	55.1%	6.7%
Train F1	49.4%	50.2%	51.7%	60.9%	51.2%	50.4%	4.7%
Test F1	52.6%	51.6%	43.6%	55.9%	51.9%	54.2%	4.5%
Train Precision	67.7%	69.3%	68.4%	61.4%	68.5%	67.7%	3.2%
Val Precision	69.0%	61.5%	68.0%	54.2%	66.7%	68.1%	6.1%
Train Recall	38.9%	39.3%	41.6%	60.4%	40.9%	40.2%	9.1%
Val Recall	42.6%	44.4%	32.1%	57.8%	42.4%	45.1%	9.2%
Train MCC	9.3%	6.2%	12.9%	-4.9%	8.5%	7.9%	6.8%
Val MCC	3.3%	14.6%	-7.4%	-18.5%	10.7%	10.0%	13.5%
Train Log-Loss	17.69	18.42	16.81	17.68	17.77	17.86	0.57
Val Log-Loss	18.54	15.67	21.72	19.97	17.04	17.07	2.38

Univariate Report

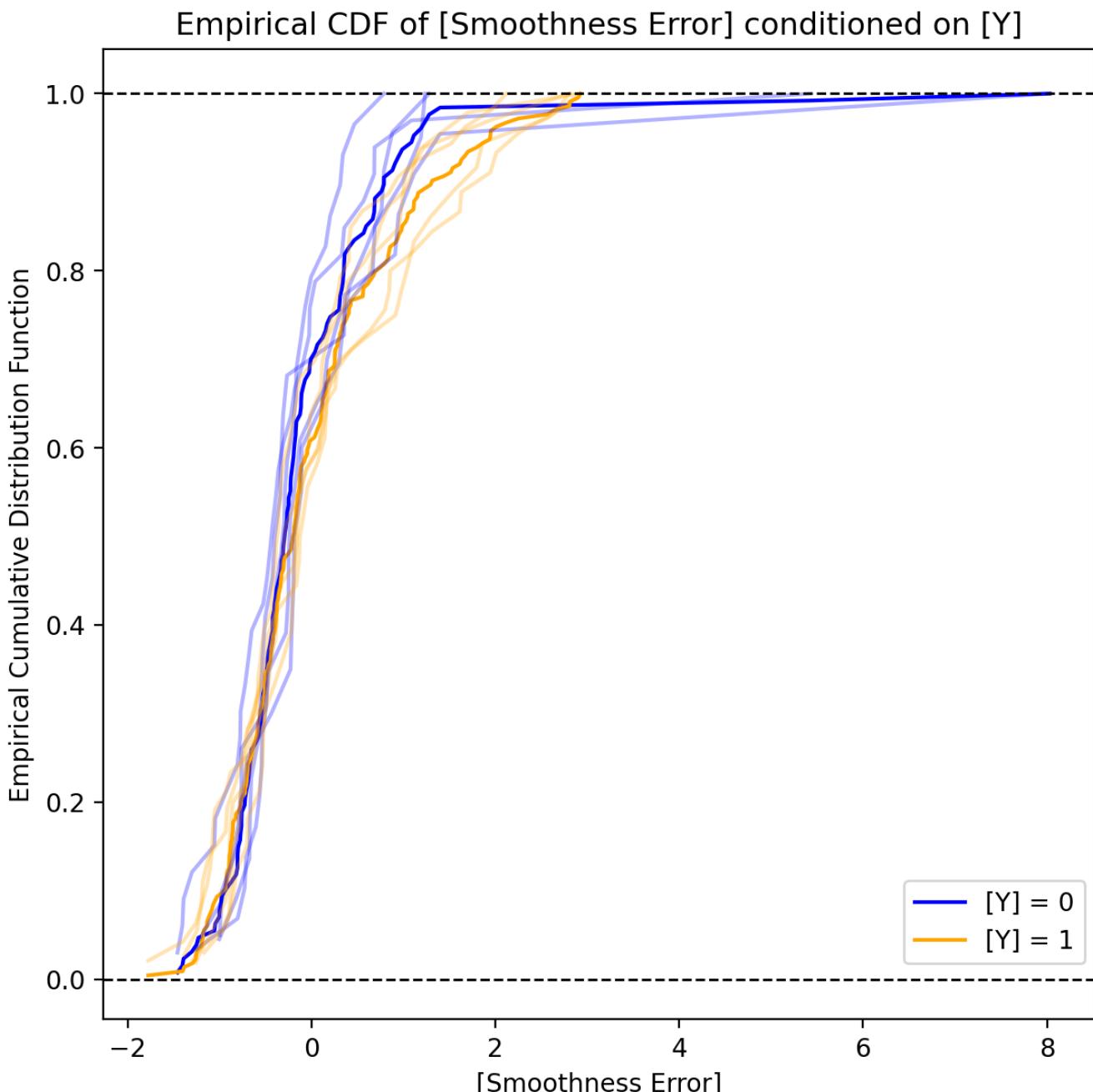
Smoothness Error - Kernel Density Plot



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

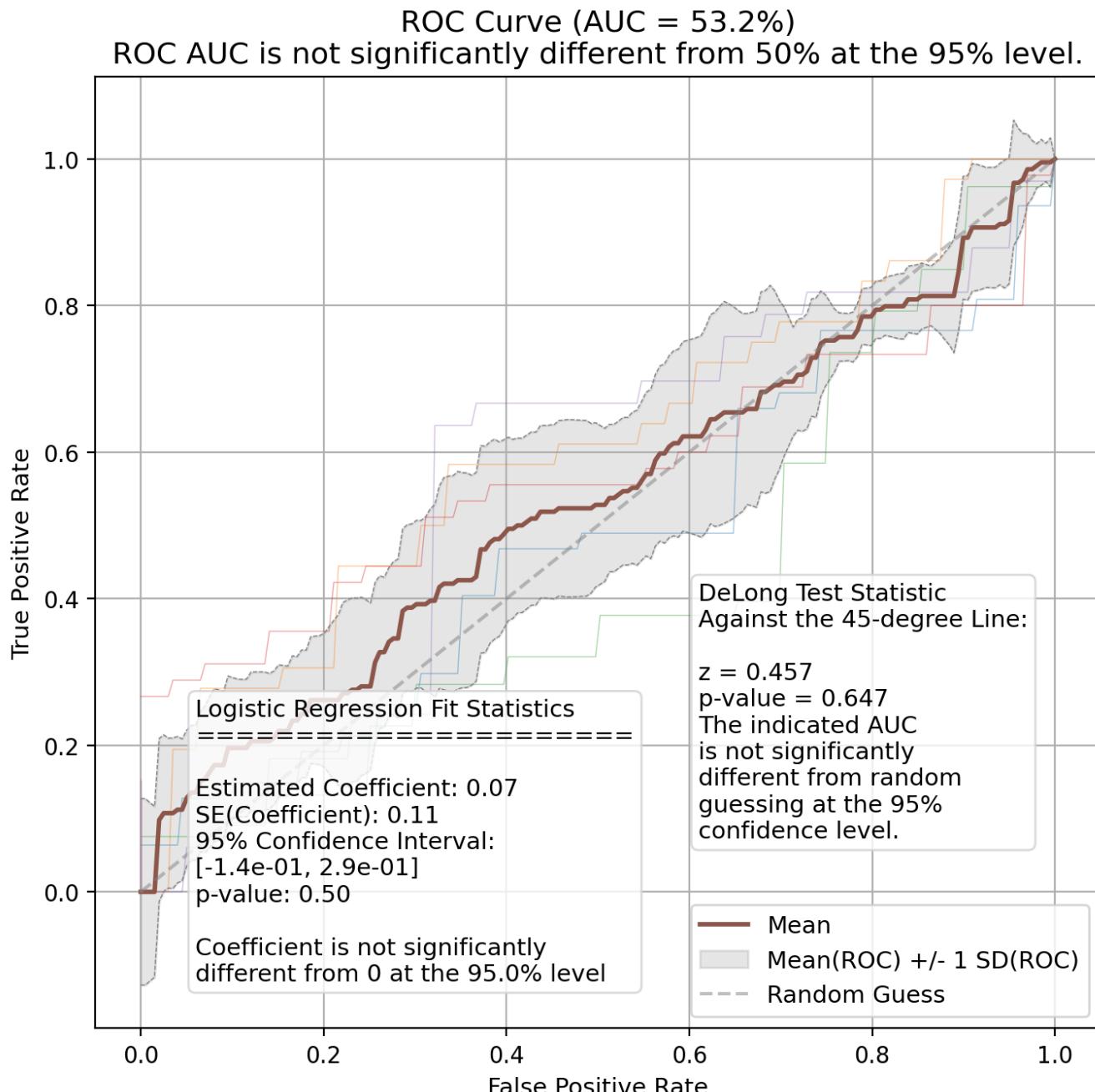
Smoothness Error - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Smoothness Error - ROC Curve

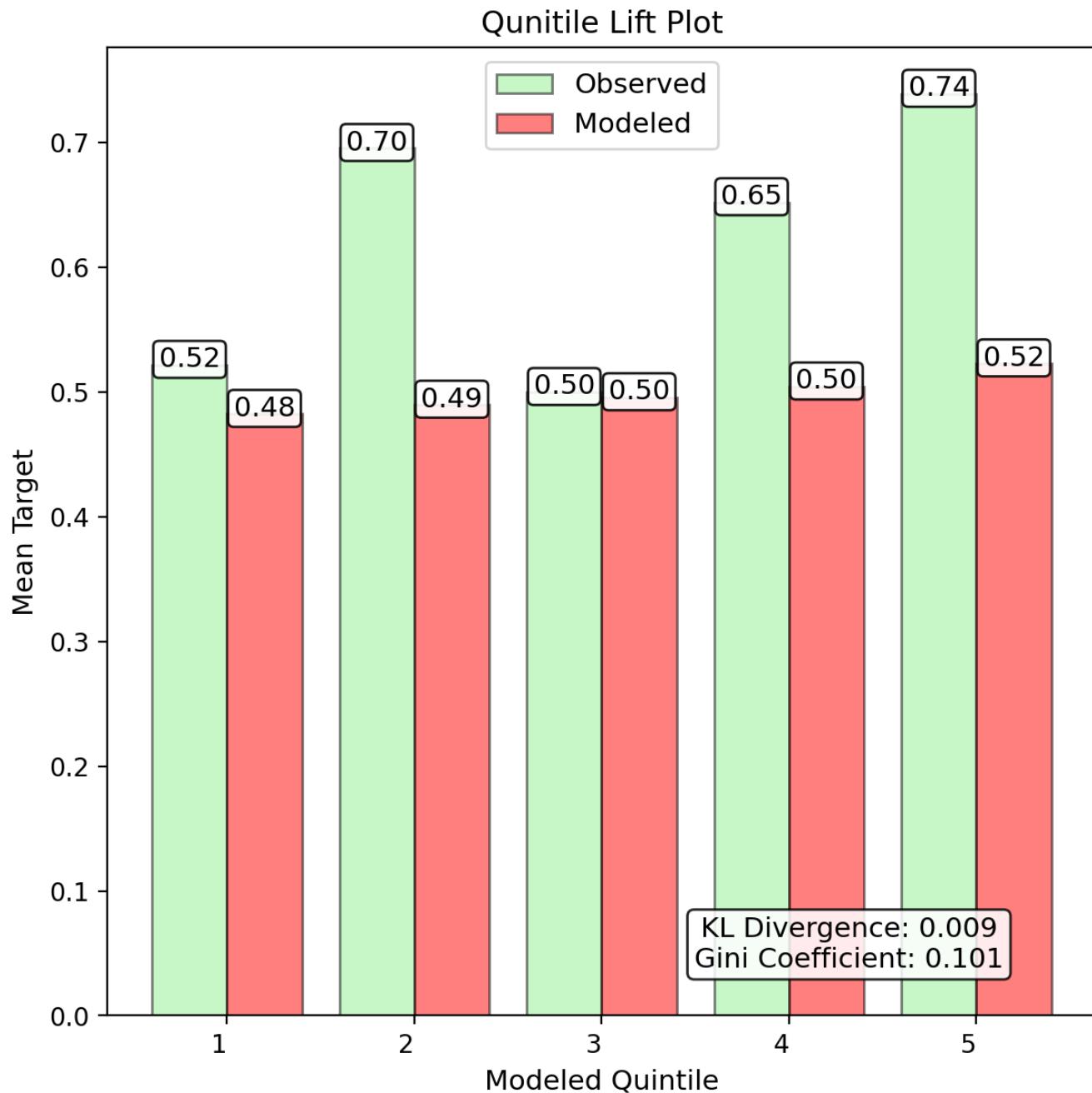


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Smoothness Error - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

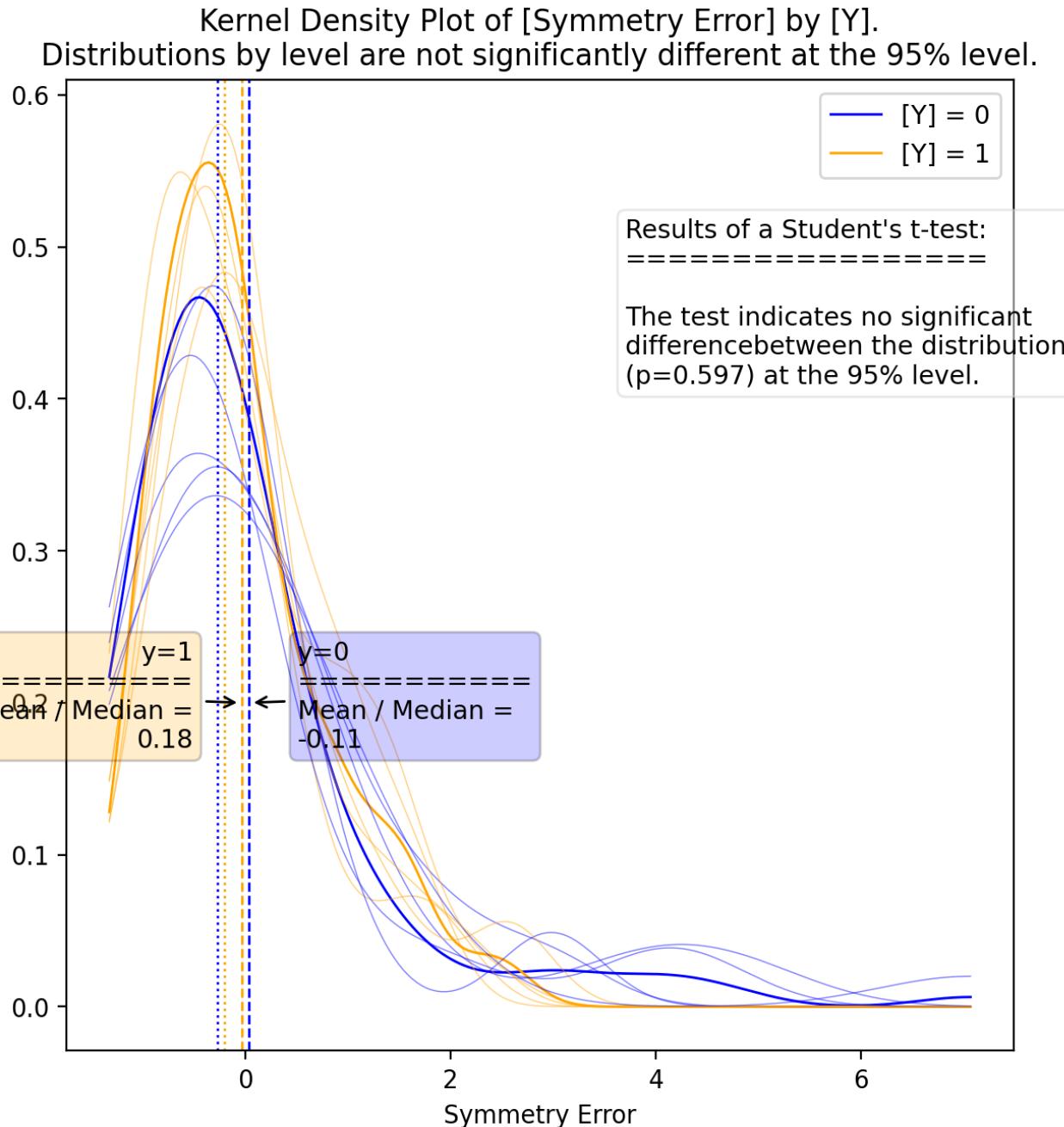
Univariate Report

Symmetry Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	5.9e-03	-1.3e-01	-4.9e-02	-1.2e-01	-4.7e-02	-6.7e-02	5.6e-02
Fitted p-Value	0.959	0.274	0.685	0.335	0.693	0.529	0.283
Fitted Std. Err.	0.115	0.118	0.122	0.124	0.120	0.106	0.003
Conf. Int. Lower	-0.220	-0.359	-0.288	-0.361	-0.282	-0.275	0.060
Conf. Int. Upper	0.231	0.102	0.189	0.123	0.188	0.142	0.053
Train Accuracy	50.9%	54.4%	50.7%	53.6%	51.0%	51.9%	1.7%
Val Accuracy	38.6%	40.6%	56.2%	44.6%	54.5%	45.6%	8.0%
Train AUC	54.6%	49.9%	48.2%	50.5%	47.3%	48.4%	2.8%
Val AUC	43.2%	40.0%	51.1%	40.3%	52.3%	42.1%	5.9%
Train F1	49.4%	65.0%	59.8%	62.9%	61.3%	61.9%	6.0%
Test F1	39.4%	48.1%	67.3%	56.8%	62.7%	56.3%	11.2%
Train Precision	67.7%	65.3%	58.7%	63.6%	61.3%	61.6%	3.5%
Val Precision	58.3%	44.2%	73.3%	54.0%	61.8%	56.3%	10.7%
Train Recall	38.9%	64.6%	60.9%	62.1%	61.3%	62.1%	10.5%
Val Recall	29.8%	52.8%	62.3%	60.0%	63.6%	56.3%	14.0%
Train MCC	9.3%	-0.3%	-3.7%	0.9%	-5.3%	-3.2%	5.7%
Val MCC	-13.5%	-20.6%	2.1%	-20.1%	4.6%	-15.8%	12.1%
Train Log-Loss	17.69	16.43	17.75	16.74	17.64	17.33	0.62
Val Log-Loss	22.14	21.42	15.80	19.97	16.38	19.60	2.90

Univariate Report

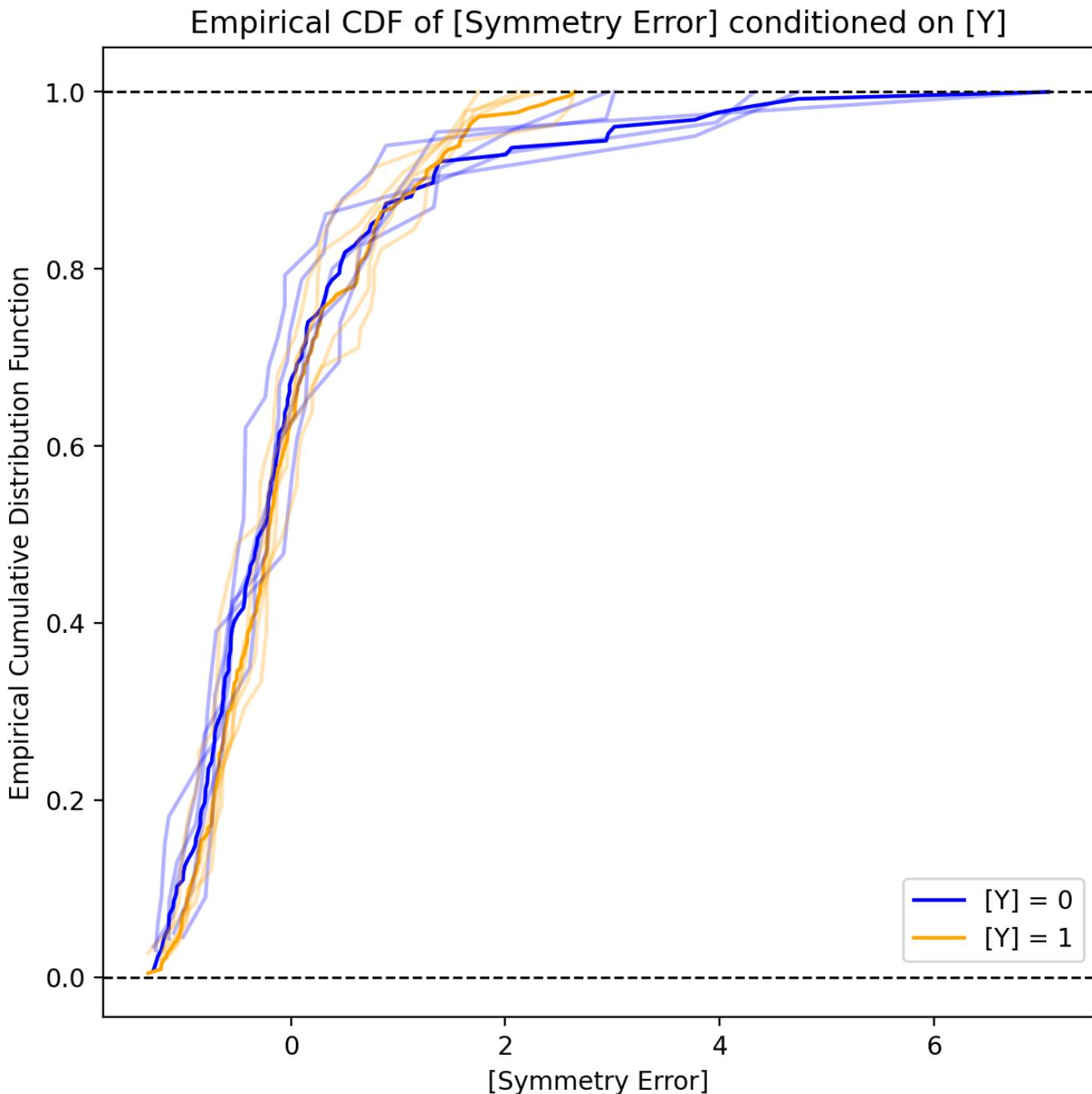
Symmetry Error - Kernel Density Plot



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

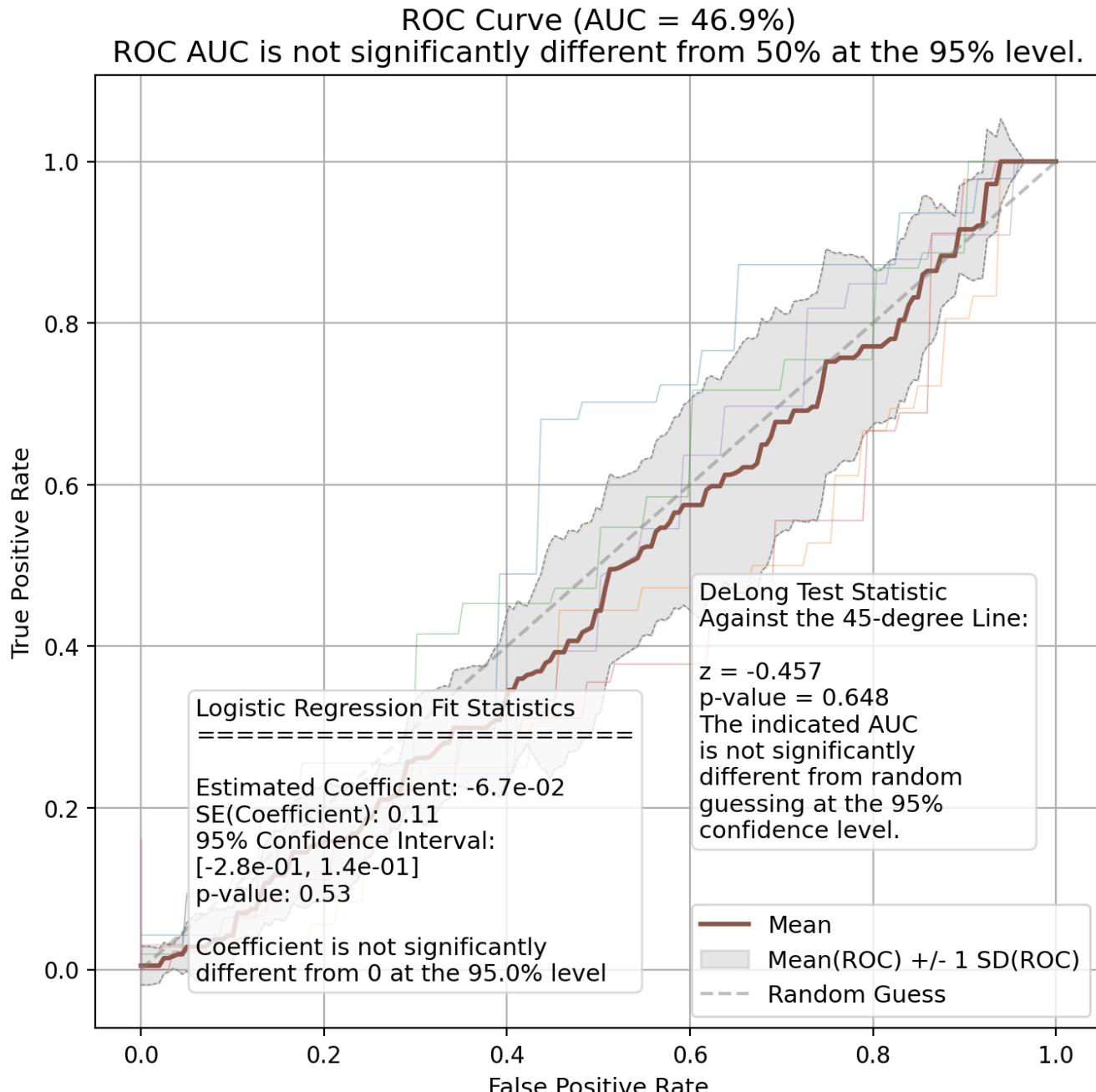
Symmetry Error - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Symmetry Error - ROC Curve

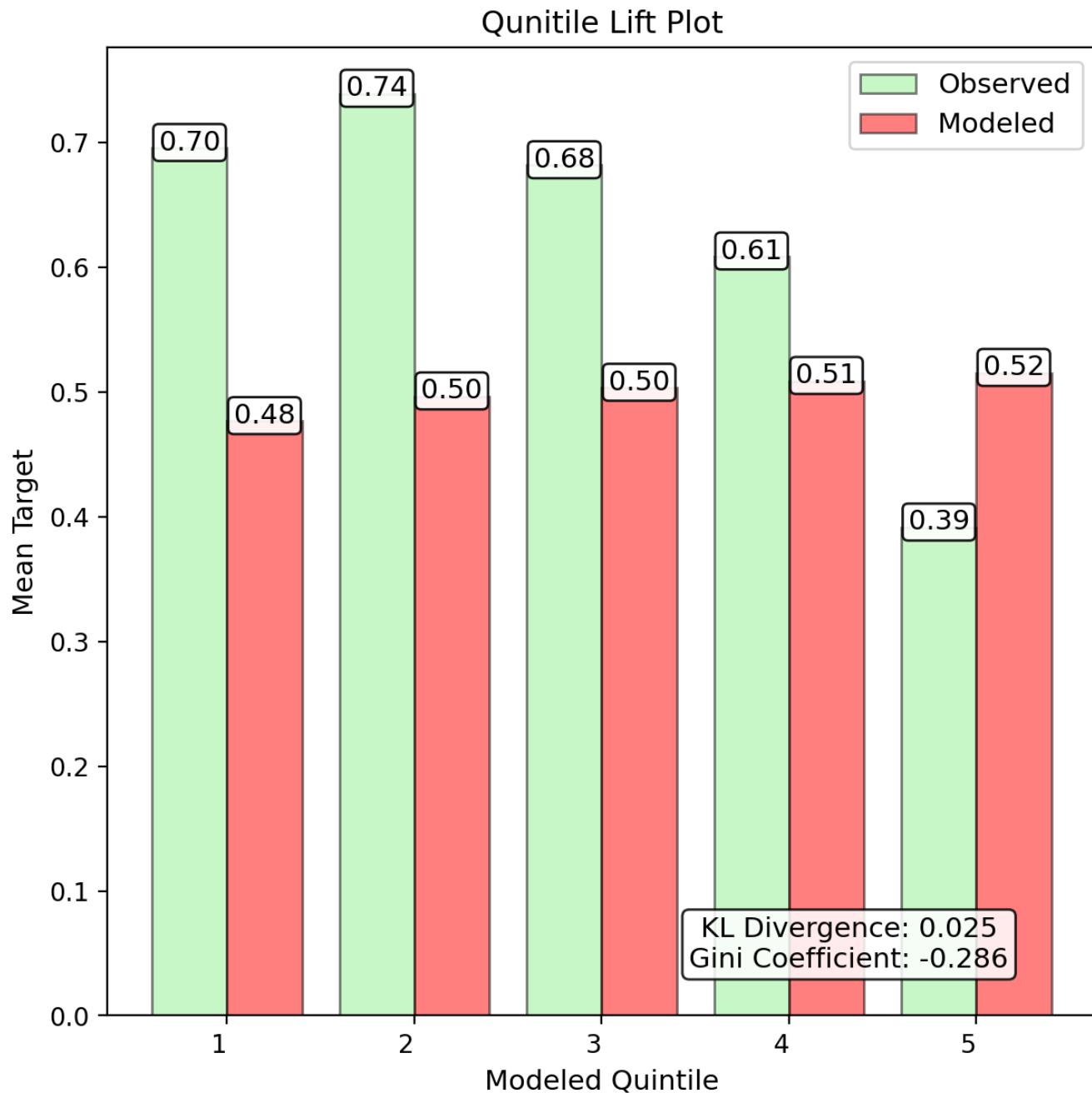


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Symmetry Error - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.

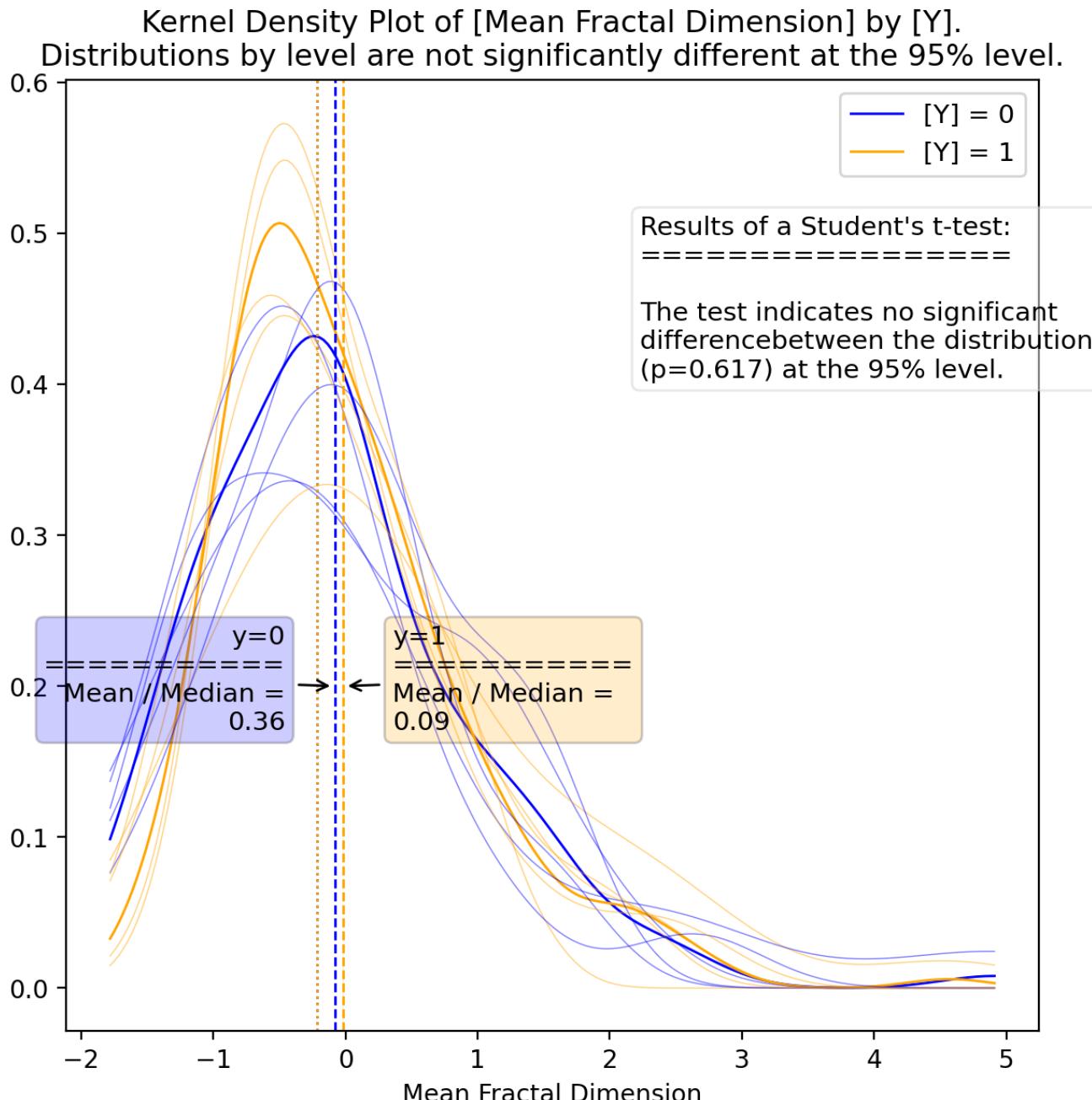
Univariate Report

Mean Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	6.6e-02	-7.6e-02	5.9e-02	8.1e-02	2.2e-02	3.4e-02	6.3e-02
Fitted p-Value	0.581	0.573	0.649	0.517	0.856	0.765	0.132
Fitted Std. Err.	0.120	0.135	0.130	0.125	0.122	0.113	0.006
Conf. Int. Lower	-0.169	-0.340	-0.196	-0.164	-0.218	-0.187	0.072
Conf. Int. Upper	0.301	0.188	0.315	0.325	0.262	0.254	0.056
Train Accuracy	50.6%	52.9%	49.6%	50.2%	48.3%	49.3%	1.7%
Val Accuracy	44.3%	44.9%	49.3%	45.9%	54.5%	44.7%	4.2%
Train AUC	52.4%	50.7%	50.9%	52.9%	50.3%	51.4%	1.2%
Val AUC	49.6%	44.9%	55.8%	45.7%	58.3%	46.0%	6.0%
Train F1	52.5%	61.7%	51.6%	52.0%	51.0%	51.5%	4.5%
Test F1	45.1%	45.7%	54.3%	51.2%	51.0%	47.9%	3.9%
Train Precision	64.3%	66.0%	61.0%	66.7%	63.6%	64.3%	2.2%
Val Precision	66.7%	47.1%	78.6%	56.8%	72.2%	58.0%	12.5%
Train Recall	44.3%	57.9%	44.7%	42.6%	42.5%	43.0%	6.5%
Val Recall	34.0%	44.4%	41.5%	46.7%	39.4%	40.8%	4.9%
Train MCC	4.8%	1.4%	1.7%	5.8%	0.6%	2.8%	2.3%
Val MCC	-0.7%	-10.1%	10.6%	-8.3%	17.4%	-7.8%	11.9%
Train Log-Loss	17.82	16.96	18.16	17.95	18.65	18.29	0.62
Val Log-Loss	20.08	19.85	18.27	19.48	16.38	19.92	1.53

Univariate Report

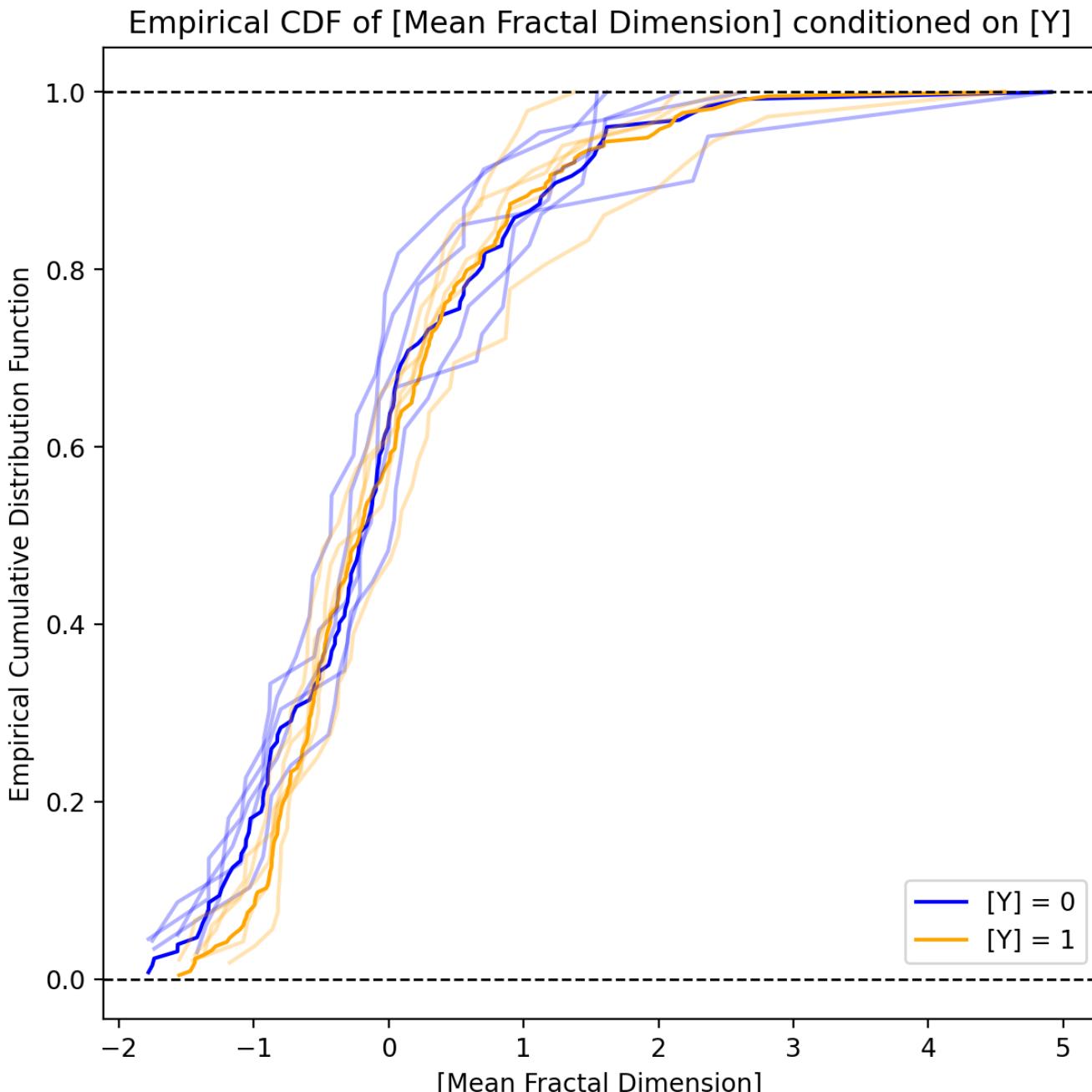
Mean Fractal Dimension - Kernel Density Plot



This plot shows the Gaussian kernel density for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the density of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data. There are annotations with the results of a t-test for the difference in means between the feature variable at each level of the target variable. The annotations corresponding to the color of the target variable level show the mean/median ratio to help understand differences in skewness between the levels of the target variable.

Univariate Report

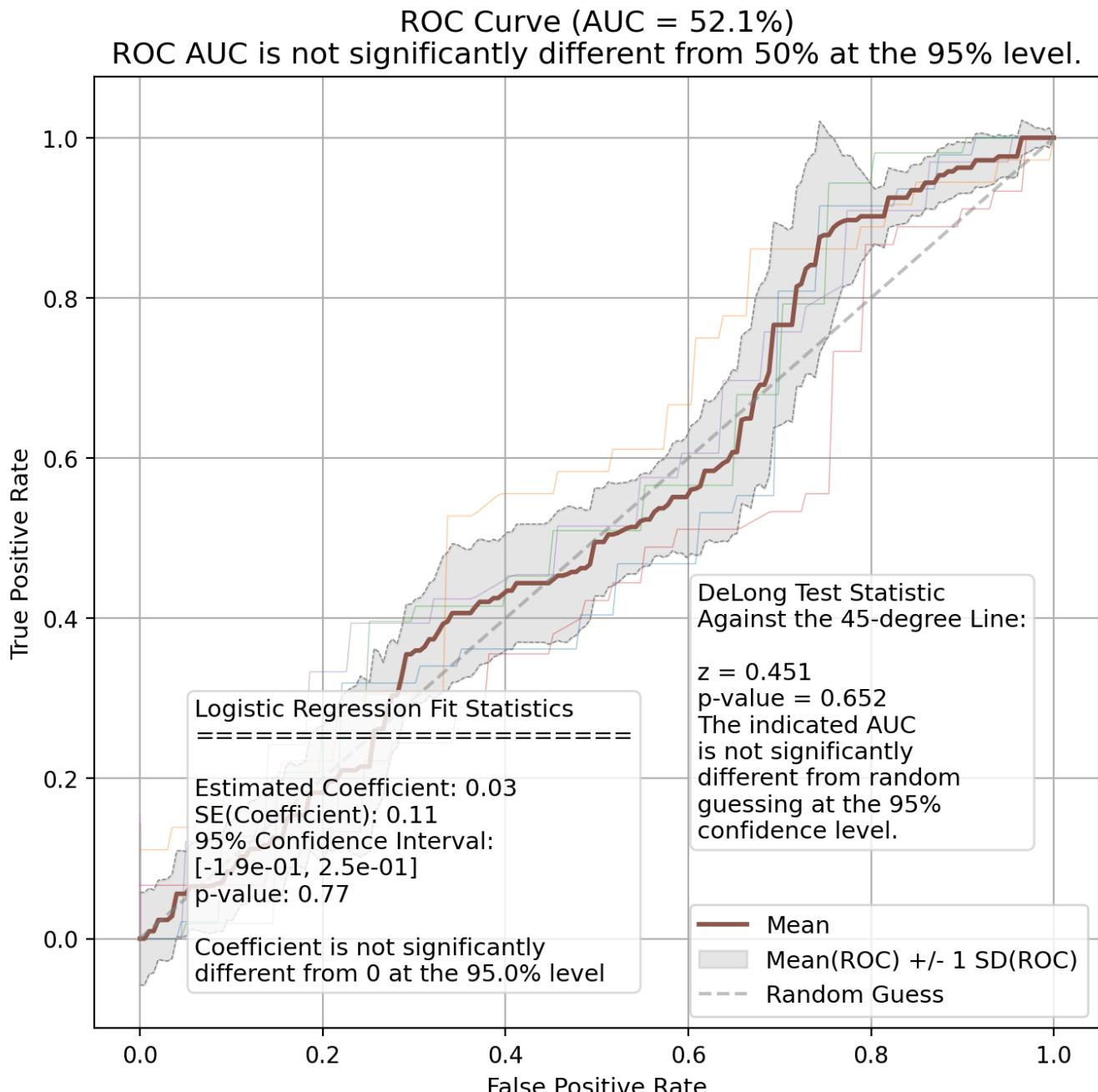
Mean Fractal Dimension - Empirical CDF Plot



This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly washed-out colors to help understand the variability of the data, and whether or not it is reasonable to assume that the data is drawn from different distributions.

Univariate Report

Mean Fractal Dimension - ROC Curve

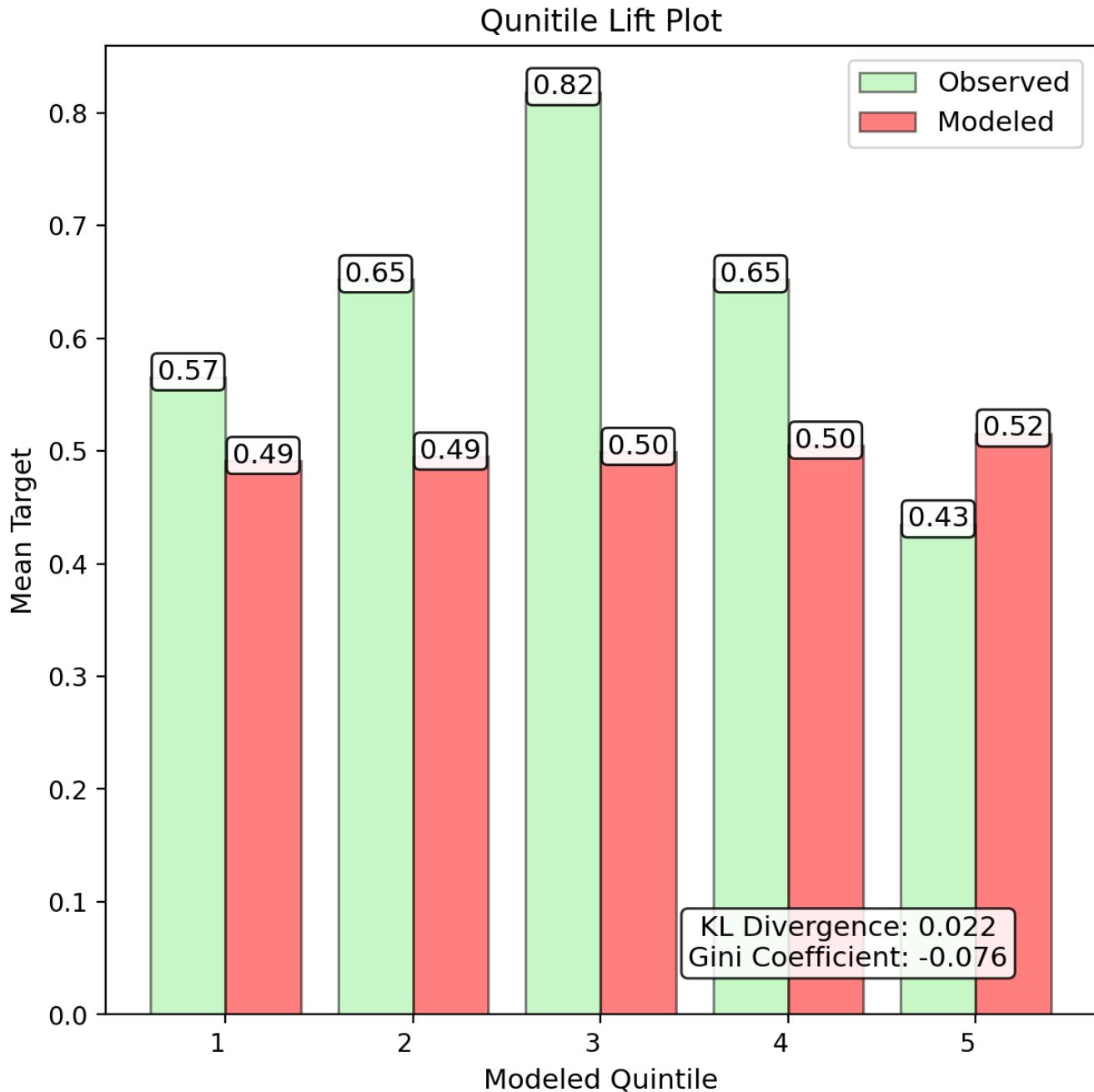


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data.

Significance of the ROC curve is determined based on a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean Fractal Dimension - Quintile Lift



The quintile lift plot is meant to show the power of the single feature to discriminate between the highest and lowest quintiles of the target variable.