

model Univariate Analysis Report

2024-02-08

Overview

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, and 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.
mean_concave_points	91.2%	94.8%	91.0%	91.3%	92.9%	81.6%	90.4%
worst_concave_points	89.3%	95.1%	87.4%	89.9%	91.1%	78.1%	88.5%
mean_concavity	88.4%	91.9%	89.4%	88.1%	90.6%	75.5%	87.3%
mean_area	87.0%	88.8%	90.8%	85.7%	89.8%	72.0%	85.7%
area_error	86.8%	84.7%	96.4%	83.6%	90.2%	71.8%	85.6%
worst_concavity	84.7%	90.7%	84.3%	84.8%	87.4%	68.3%	83.4%
perimeter_error	81.9%	84.0%	88.0%	79.8%	85.9%	60.8%	80.1%
radius_error	81.5%	83.7%	87.7%	79.5%	85.6%	60.0%	79.7%
mean_compactness	79.1%	86.5%	79.0%	79.1%	82.6%	56.9%	77.2%
worst_compactness	78.7%	83.9%	81.8%	77.7%	82.8%	54.9%	76.6%
smoothness_error	51.1%	67.9%	42.0%	54.3%	51.9%	8.5%	45.9%
symmetry_error	50.1%	67.3%	39.8%	53.6%	50.0%	7.2%	44.7%
texture_error	49.2%	64.4%	42.6%	51.5%	51.3%	2.9%	43.6%
mean_fractal_dimension	48.3%	63.1%	42.6%	50.3%	50.8%	0.6%	42.6%
fractal_dimension_error	38.3%	51.4%	30.0%	41.2%	37.9%	-1.77e-01	30.2%
mean_symmetry	36.6%	49.2%	36.4%	36.6%	41.9%	-2.60e-01	29.1%
mean_smoothness	35.5%	48.2%	37.8%	34.7%	42.4%	-2.96e-01	28.2%
worst_fractal_dimension	33.7%	45.5%	28.0%	35.7%	34.7%	-2.84e-01	24.9%
worst_smoothness	32.3%	44.9%	34.7%	31.5%	39.2%	-3.58e-01	24.5%
worst_symmetry	32.5%	44.5%	30.8%	33.1%	36.4%	-3.30e-01	24.1%
compactness_error	30.9%	41.7%	25.2%	32.9%	31.4%	-3.41e-01	21.3%
worst_texture	28.1%	40.3%	30.3%	27.4%	34.6%	-4.38e-01	19.5%
concavity_error	29.0%	39.0%	23.2%	31.0%	29.1%	-3.80e-01	18.9%

mean_texture	27.2%	39.2%	28.9%	26.7%	33.2%	-4.52e-01	18.3%
concave_points_error	26.9%	38.1%	26.3%	27.1%	31.1%	-4.47e-01	17.5%

This table shows an overview of the results for the variables in this file, representing those whose average test score are ranked between 2 and 26 of the variables passed to the model.

Univariate Report

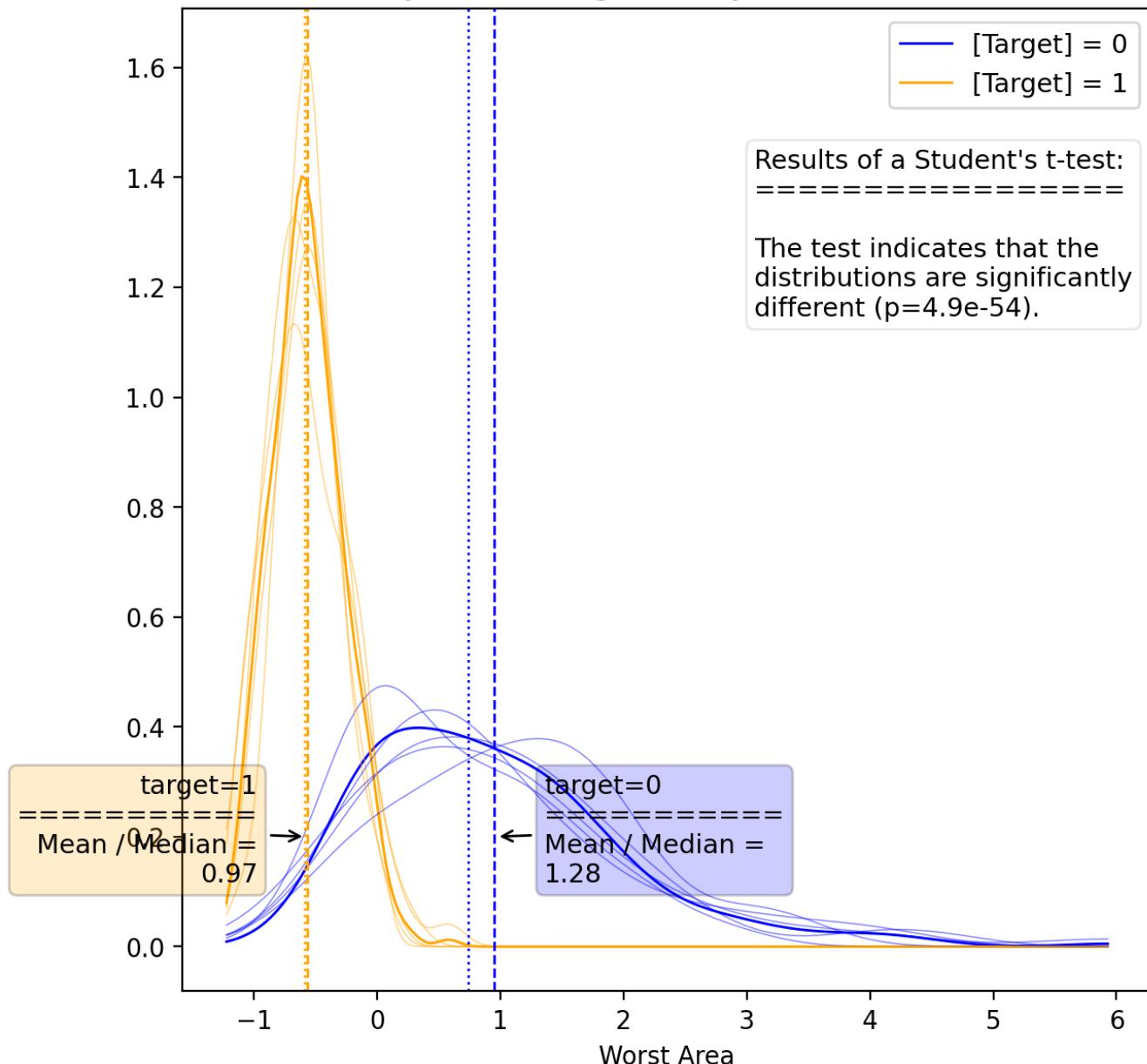
Worst Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-5.2e+00	-5.2e+00	-5.1e+00	-5.1e+00	-5.1e+00	-3.3e-04	4.1e-02
Fitted p-Value	1.0e-27	5.7e-26	8.6e-27	2.6e-25	1.4e-26	8.2e-05	1.1e-25
Fitted Std. Err.	0.472	0.490	0.477	0.494	0.474	0.000	0.010
Conf. Int. Lower	-6.1e+00	-6.1e+00	-6.0e+00	-6.1e+00	-6.0e+00	-4.9e-04	5.3e-02
Conf. Int. Upper	-4.2e+00	-4.2e+00	-4.2e+00	-4.2e+00	-4.1e+00	-1.7e-04	3.7e-02
Train Accuracy	92.7%	91.7%	91.5%	91.8%	92.1%	91.9%	0.4%
Val Accuracy	87.0%	92.6%	94.5%	92.4%	92.2%	91.9%	2.8%
Train AUC	90.9%	89.6%	89.6%	90.1%	90.1%	90.0%	0.5%
Val AUC	83.8%	91.9%	92.5%	90.7%	90.7%	90.0%	3.5%
Train F1	94.4%	93.7%	93.4%	93.7%	93.9%	93.8%	0.4%
Test F1	90.2%	94.1%	95.9%	94.0%	94.0%	93.8%	2.1%
Train Precision	91.3%	89.9%	89.7%	91.2%	90.0%	90.4%	0.8%
Val Precision	83.3%	93.5%	93.4%	89.7%	92.2%	90.4%	4.3%
Train Recall	97.7%	97.9%	97.5%	96.4%	98.2%	97.5%	0.7%
Val Recall	98.2%	94.7%	98.6%	98.7%	95.9%	97.5%	1.8%
Train MCC	84.2%	82.3%	82.1%	82.2%	83.3%	82.7%	0.9%
Val MCC	73.3%	84.2%	87.7%	84.4%	82.8%	82.7%	5.4%
Train Log-Loss	2.64	2.98	3.06	2.96	2.86	2.91	0.16
Val Log-Loss	4.70	2.66	1.98	2.75	2.82	2.91	1.02

Univariate Report

Worst Area - Kernel Density Plot

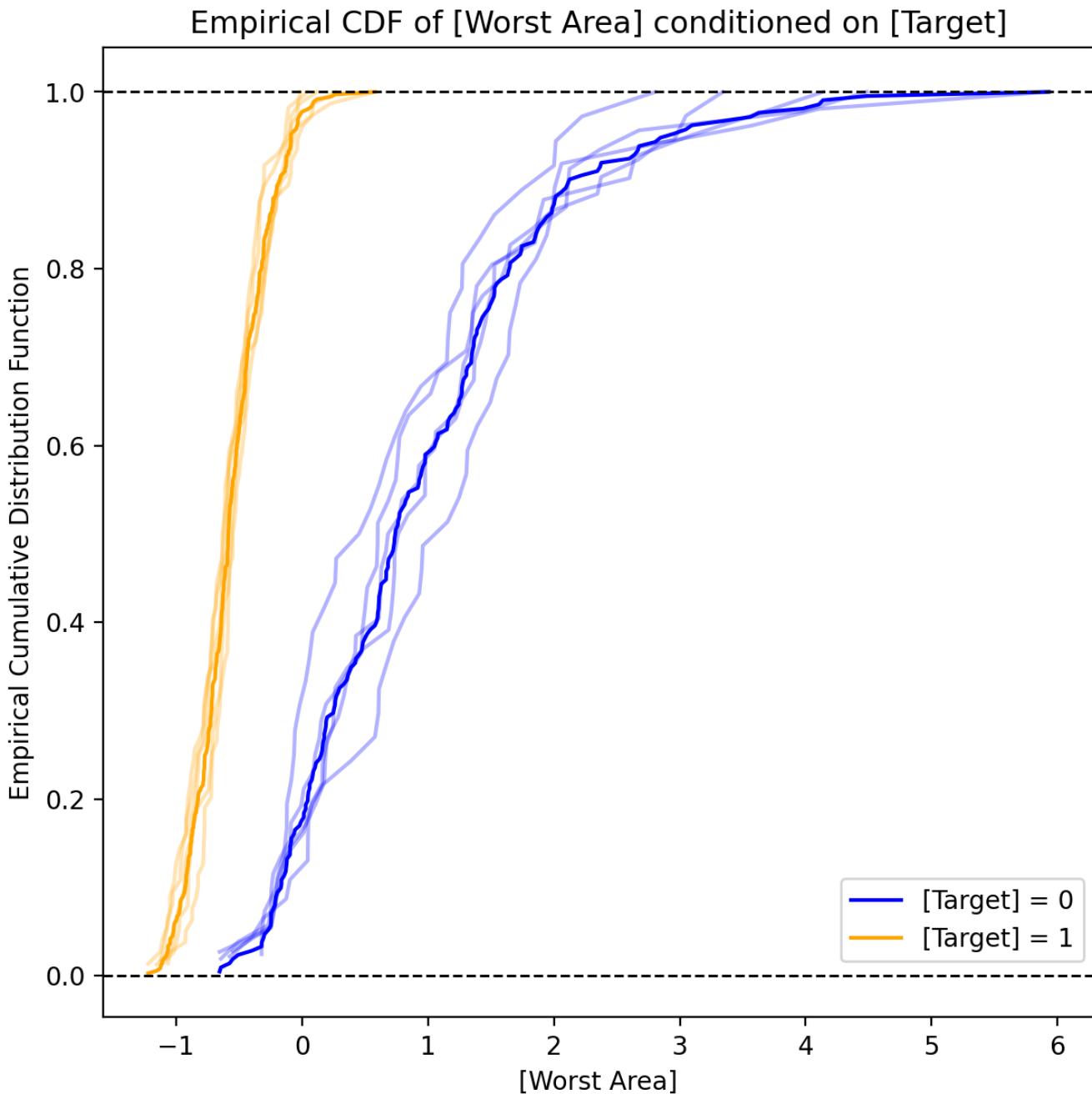
Kernel Density Plot of [Worst Area] by [Target]
Distributions by level are significantly different at the 95% level.



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

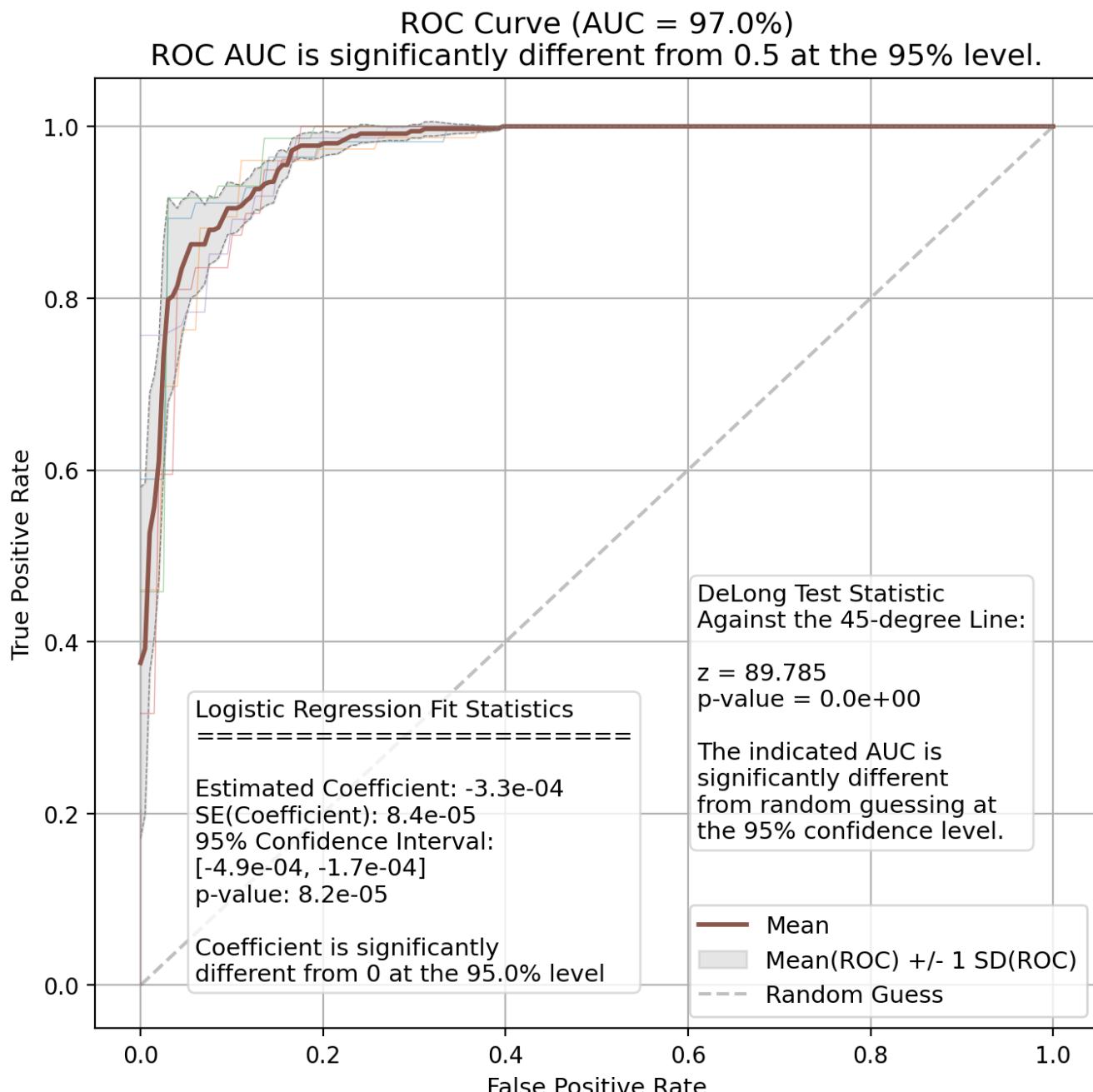
Worst Area - Empirical CDF Plot



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 Area - ROC Curve



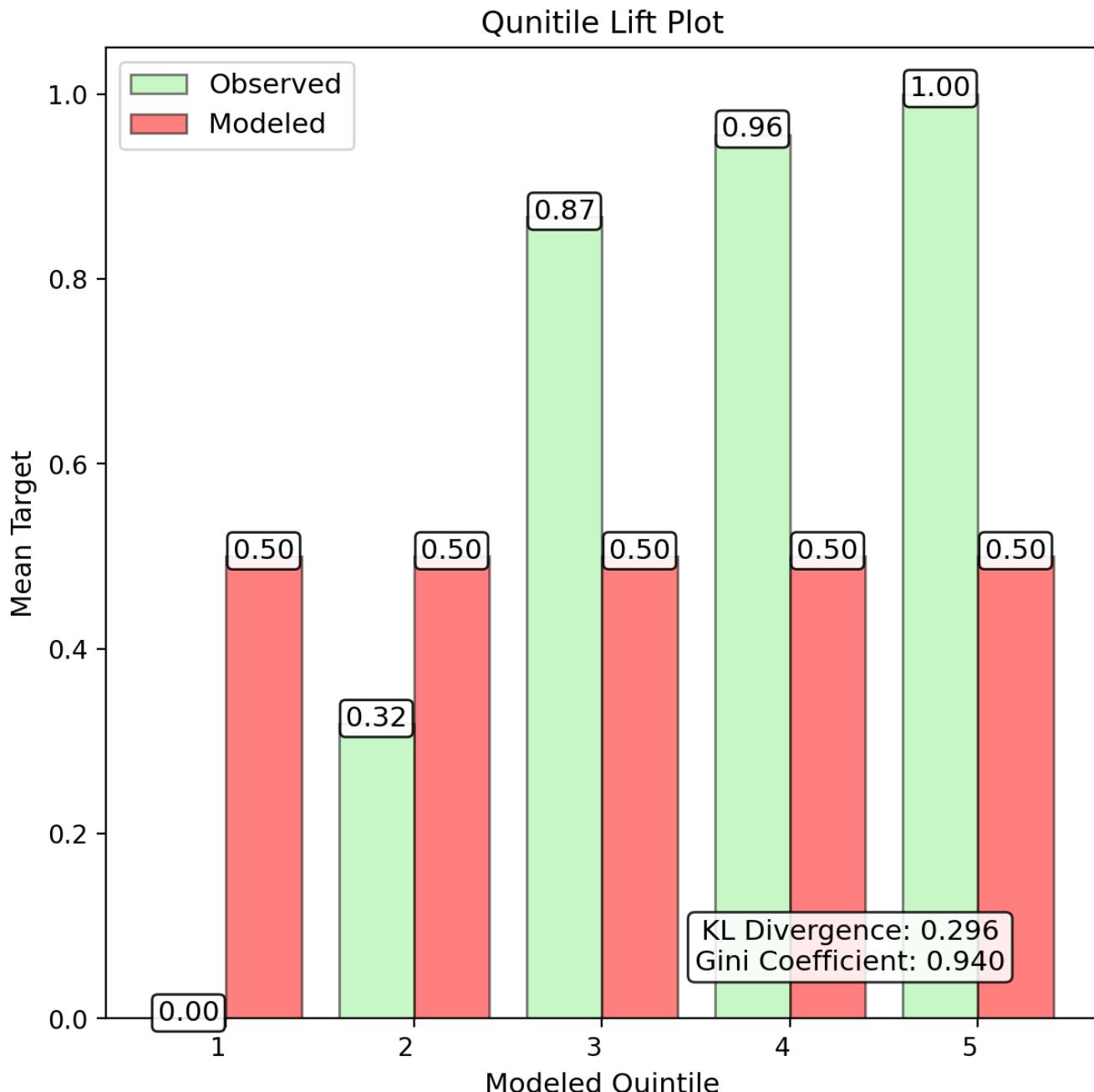
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Area - 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

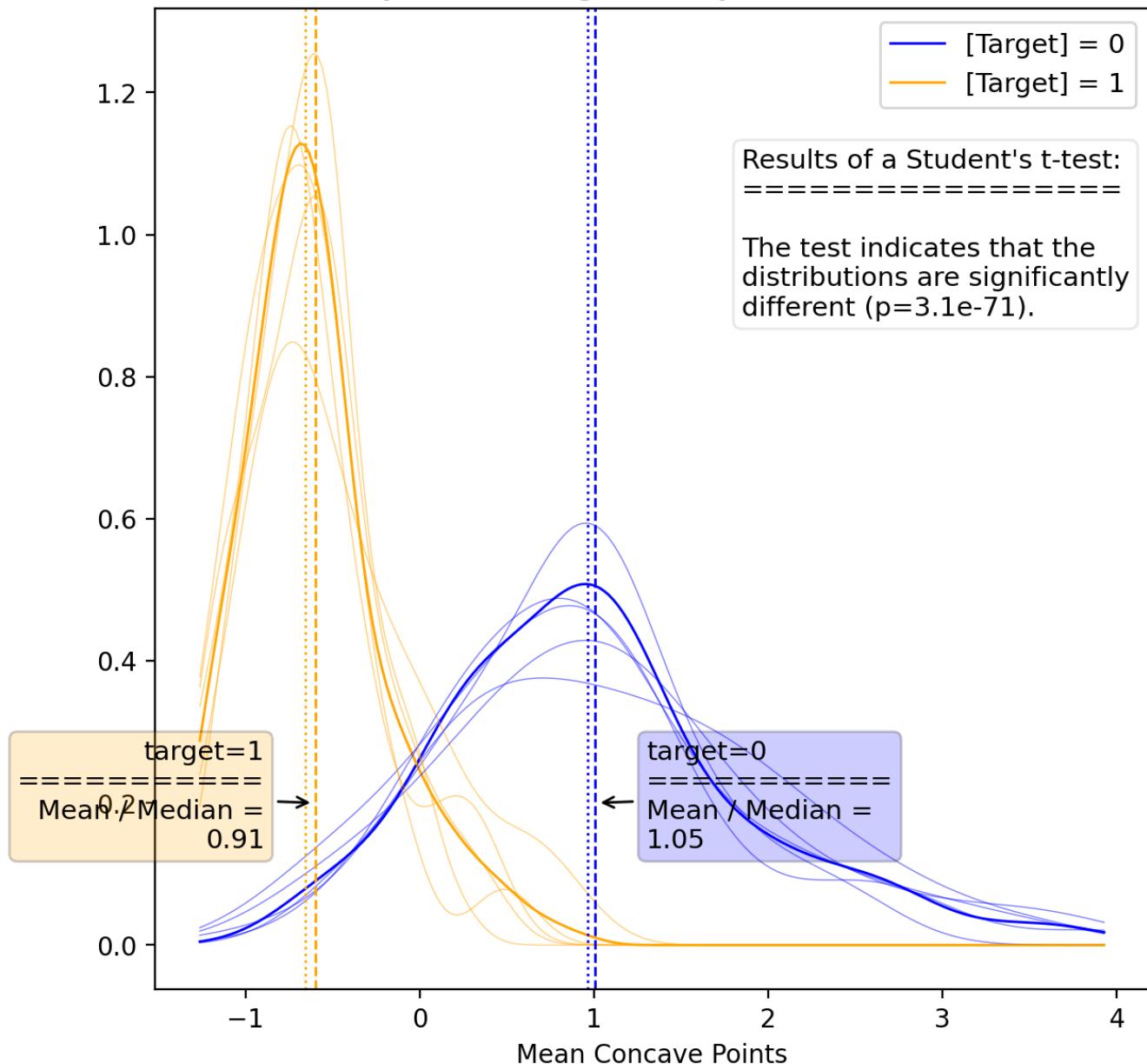
Mean Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.8e+00	-4.5e+00	-3.7e+00	-3.6e+00	-3.9e+00	-9.1e+00	3.5e-01
Fitted p-Value	3.1e-31	3.0e-27	4.6e-31	9.8e-30	1.4e-29	1.9e-09	1.3e-27
Fitted Std. Err.	0.327	0.415	0.323	0.315	0.347	1.507	0.041
Conf. Int. Lower	-4.4e+00	-5.3e+00	-4.4e+00	-4.2e+00	-4.6e+00	-1.2e+01	4.3e-01
Conf. Int. Upper	-3.2e+00	-3.7e+00	-3.1e+00	-3.0e+00	-3.2e+00	-6.1e+00	2.7e-01
Train Accuracy	90.6%	93.3%	90.7%	90.2%	91.6%	91.2%	1.3%
Val Accuracy	94.6%	82.8%	94.5%	94.7%	90.4%	91.2%	5.1%
Train AUC	90.6%	93.2%	90.6%	90.4%	91.7%	91.3%	1.2%
Val AUC	94.0%	83.6%	95.2%	94.3%	90.4%	91.3%	4.8%
Train F1	92.4%	94.6%	92.3%	92.1%	93.2%	92.9%	1.0%
Test F1	95.6%	85.3%	95.7%	95.6%	92.4%	92.9%	4.5%
Train Precision	94.4%	95.6%	93.8%	94.7%	94.9%	94.8%	0.7%
Val Precision	94.7%	91.0%	98.5%	95.0%	94.4%	94.8%	2.7%
Train Recall	90.4%	93.6%	90.9%	89.6%	91.5%	91.0%	1.5%
Val Recall	96.4%	80.3%	93.1%	96.2%	90.5%	91.0%	6.6%
Train MCC	80.2%	85.8%	80.4%	79.5%	82.5%	81.6%	2.6%
Val MCC	88.6%	65.5%	88.3%	88.8%	79.6%	81.6%	10.1%
Train Log-Loss	3.40	2.42	3.37	3.54	3.02	3.17	0.45
Val Log-Loss	1.96	6.20	1.98	1.93	3.45	3.17	1.85

Univariate Report

Mean Concave Points - Kernel Density Plot

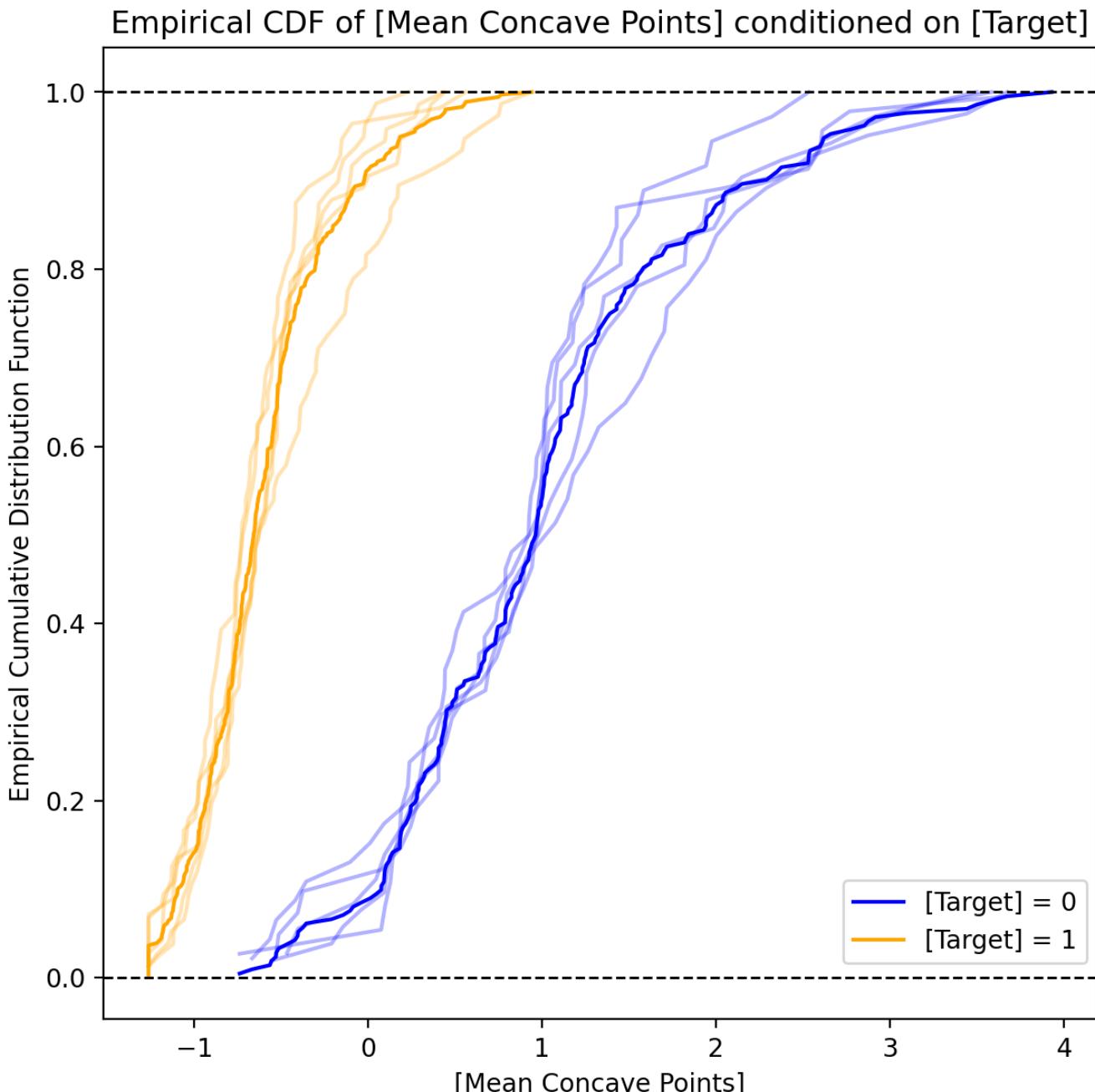
Kernel Density Plot of [Mean Concave Points] by [Target]
Distributions by level are significantly different at the 95% level.



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

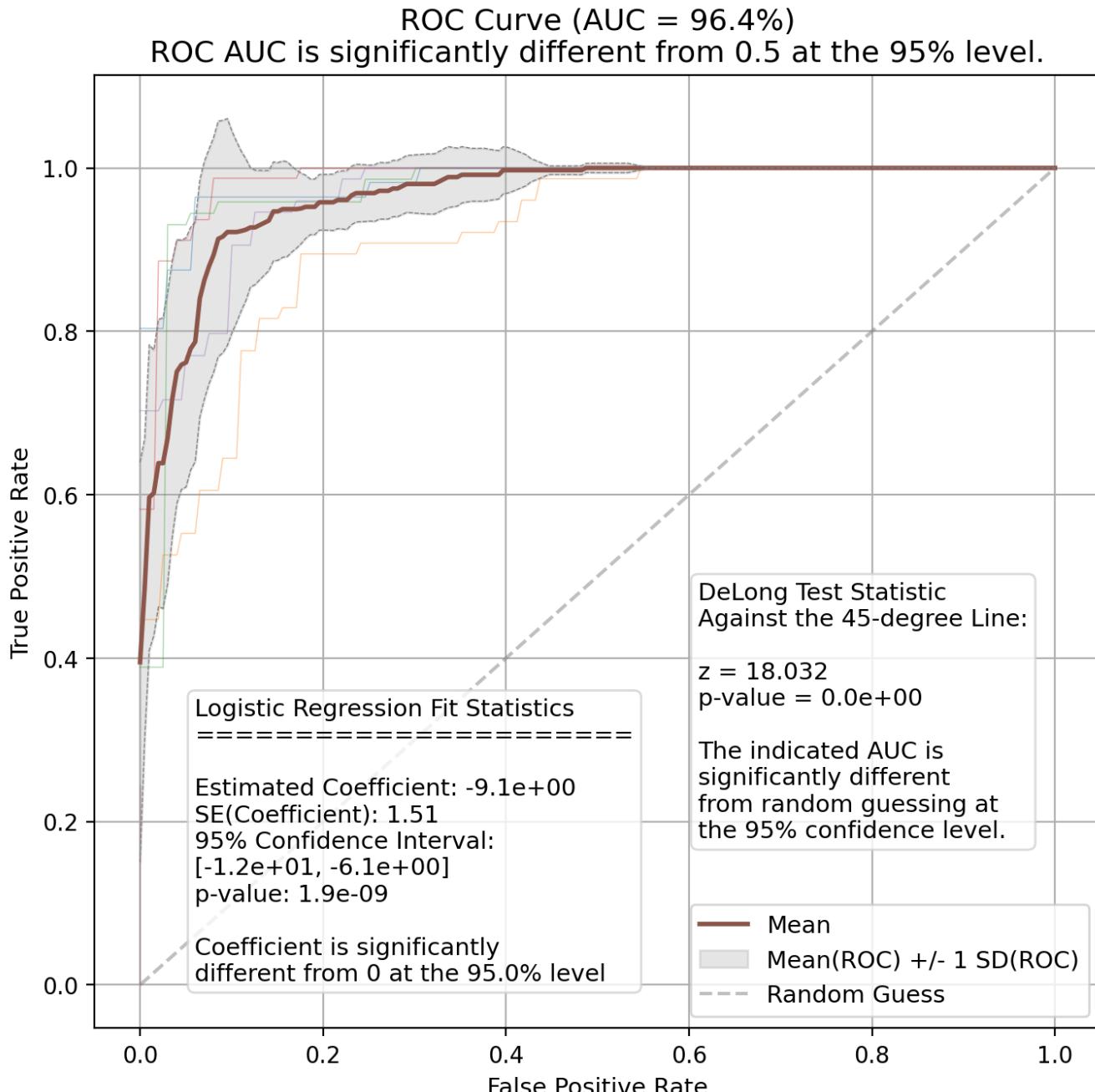
Mean Concave Points - Empirical CDF Plot



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 Concave Points - ROC Curve



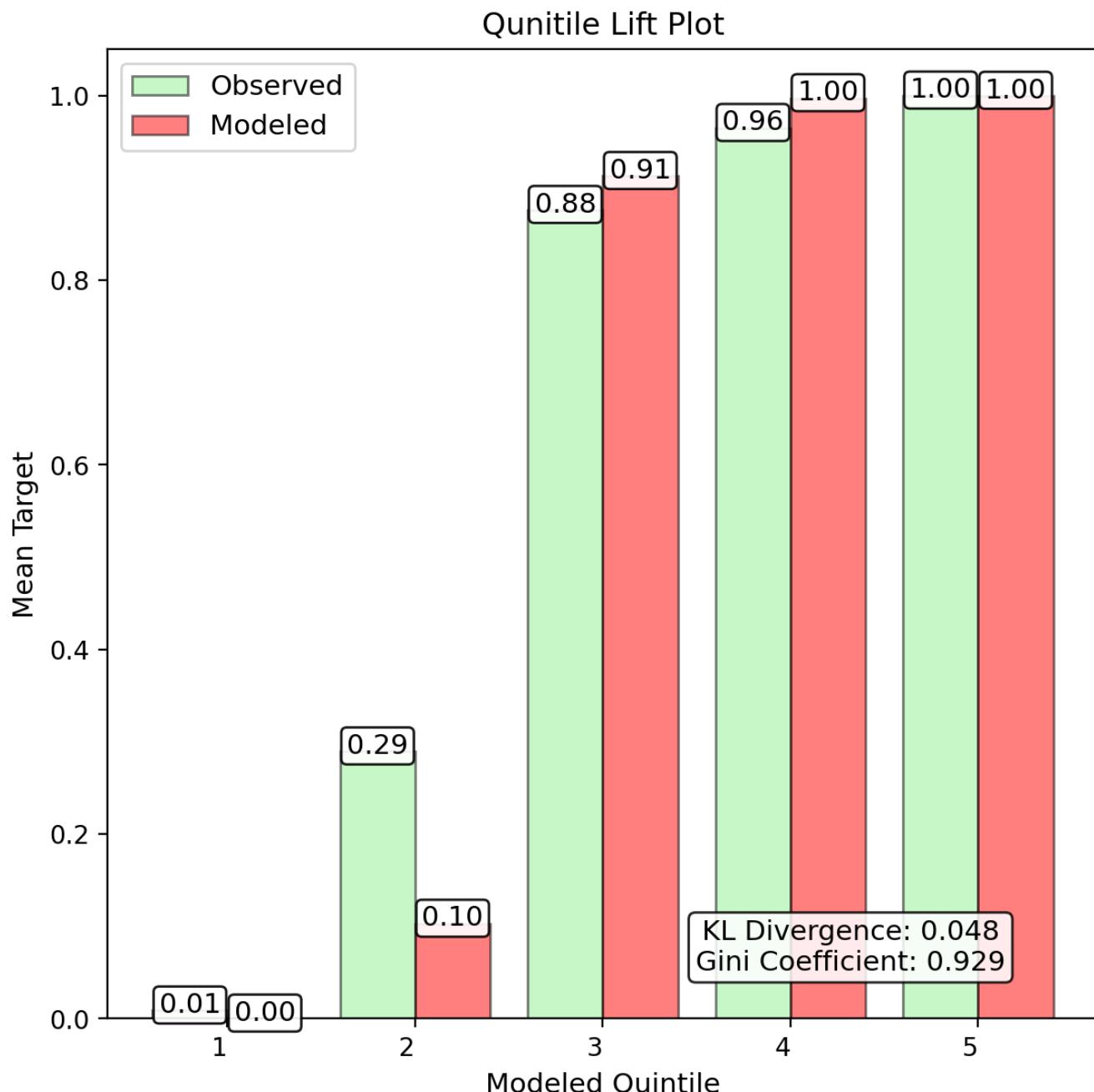
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean Concave Points - 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

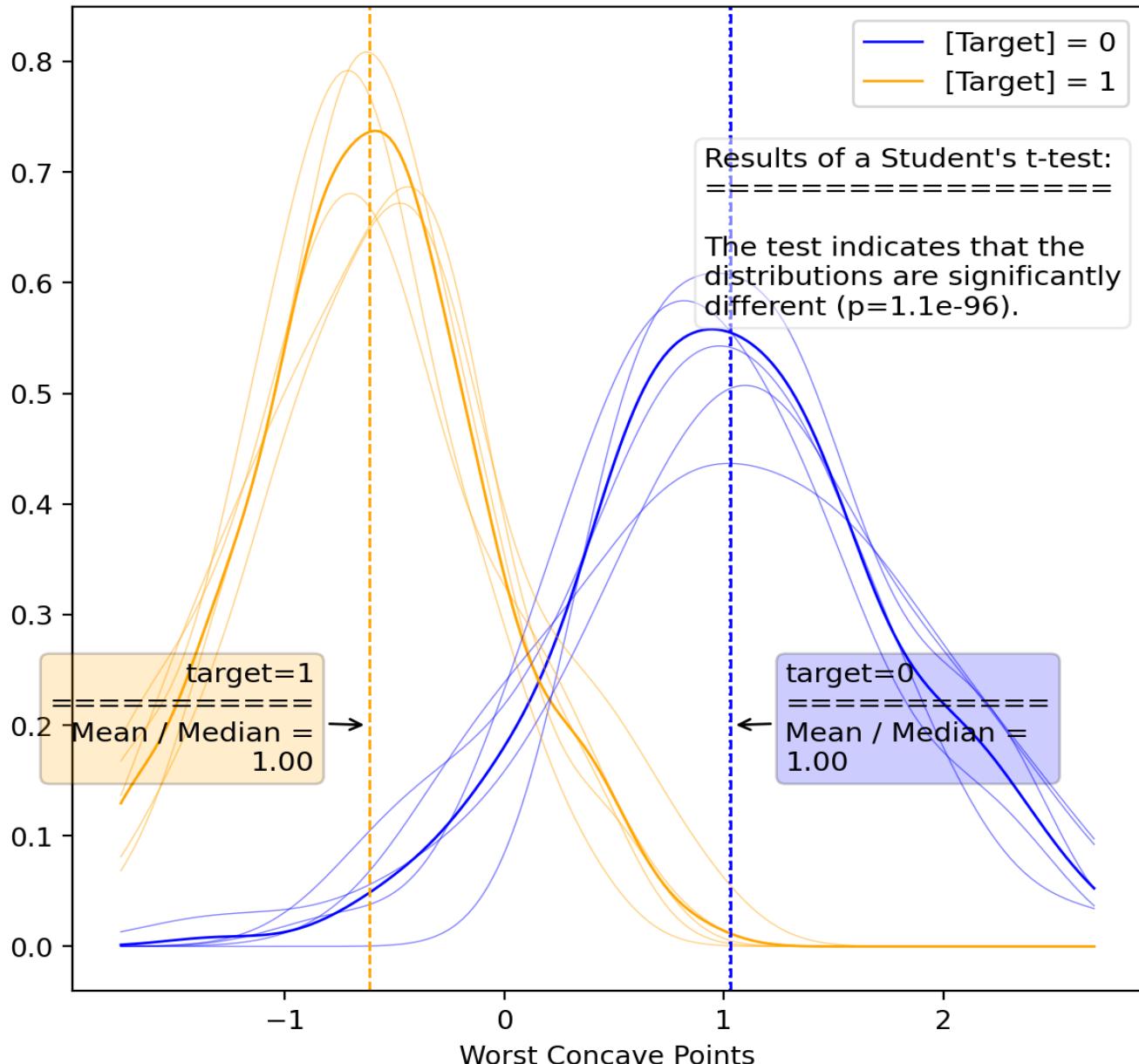
Worst Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.3e+00	-3.9e+00	-3.4e+00	-3.3e+00	-3.6e+00	-2.4e+00	2.4e-01
Fitted p-Value	8.3e-30	2.3e-26	2.4e-29	2.4e-27	5.3e-28	2.0e-04	1.0e-26
Fitted Std. Err.	0.291	0.364	0.301	0.308	0.330	0.651	0.029
Conf. Int. Lower	-3.9e+00	-4.6e+00	-4.0e+00	-3.9e+00	-4.3e+00	-3.7e+00	2.9e-01
Conf. Int. Upper	-2.7e+00	-3.2e+00	-2.8e+00	-2.7e+00	-3.0e+00	-1.1e+00	1.8e-01
Train Accuracy	87.8%	90.2%	89.3%	88.8%	89.9%	89.3%	0.9%
Val Accuracy	95.7%	84.4%	89.9%	90.1%	87.8%	89.3%	4.1%
Train AUC	88.5%	90.8%	90.1%	89.6%	90.4%	89.9%	0.9%
Val AUC	96.4%	85.4%	89.7%	90.5%	88.9%	89.9%	4.0%
Train F1	89.9%	91.9%	91.0%	90.8%	91.6%	91.1%	0.7%
Test F1	96.3%	86.7%	92.2%	91.5%	90.0%	91.1%	3.5%
Train Precision	94.2%	95.8%	95.4%	95.3%	95.1%	95.1%	0.6%
Val Precision	100.0%	92.5%	94.2%	94.6%	95.5%	95.1%	2.8%
Train Recall	86.0%	88.3%	87.0%	86.7%	88.3%	87.4%	1.0%
Val Recall	92.9%	81.6%	90.3%	88.6%	85.1%	87.4%	4.4%
Train MCC	75.2%	79.9%	78.5%	77.2%	79.2%	78.1%	1.9%
Val MCC	91.4%	68.9%	78.1%	79.9%	75.4%	78.1%	8.2%
Train Log-Loss	4.38	3.55	3.84	4.03	3.65	3.86	0.33
Val Log-Loss	1.57	5.61	3.64	3.58	4.39	3.86	1.47

Univariate Report

Worst Concave Points - Kernel Density Plot

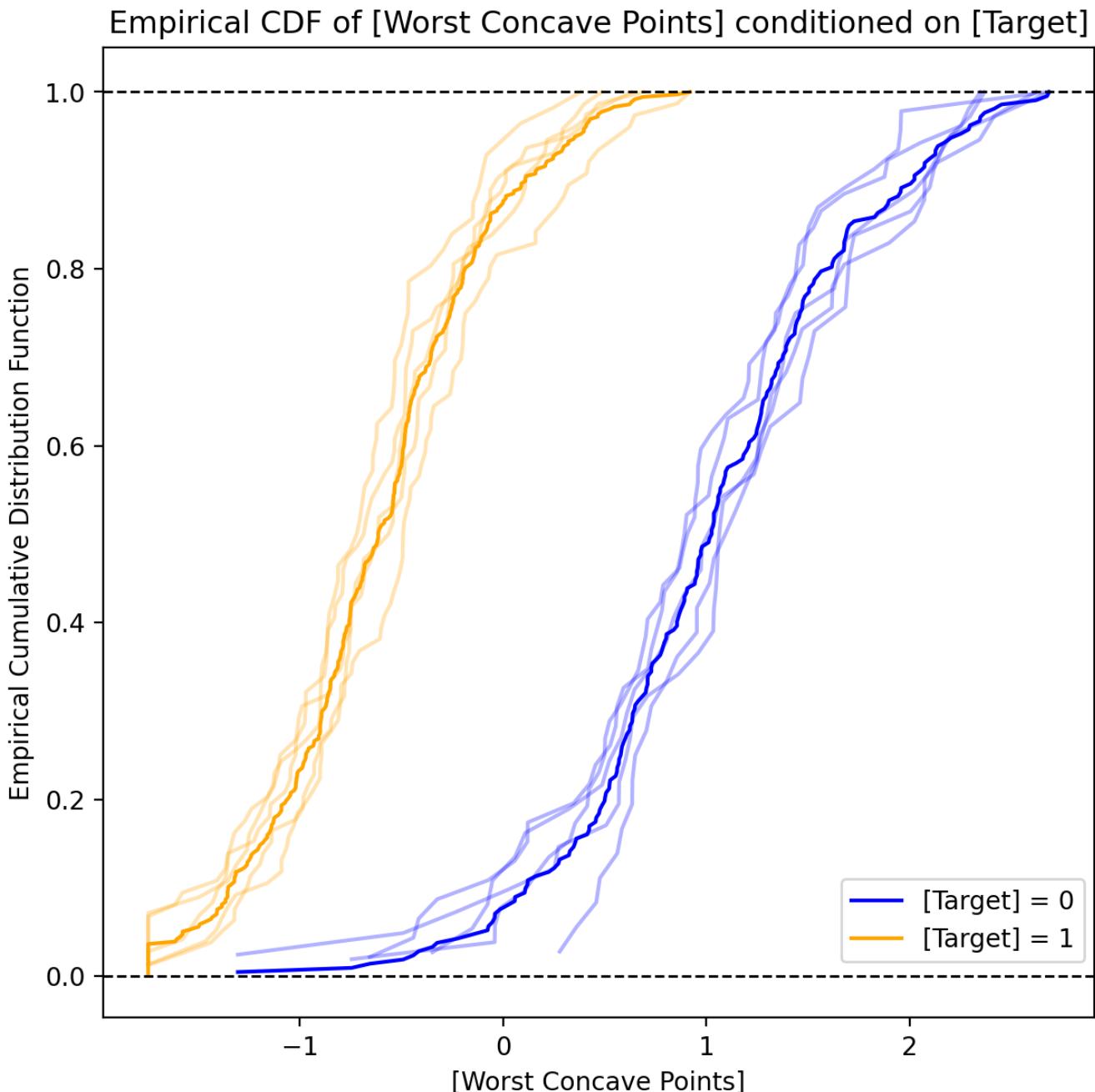
Kernel Density Plot of [Worst Concave Points] by [Target]
Distributions by level are significantly different at the 95% level.



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

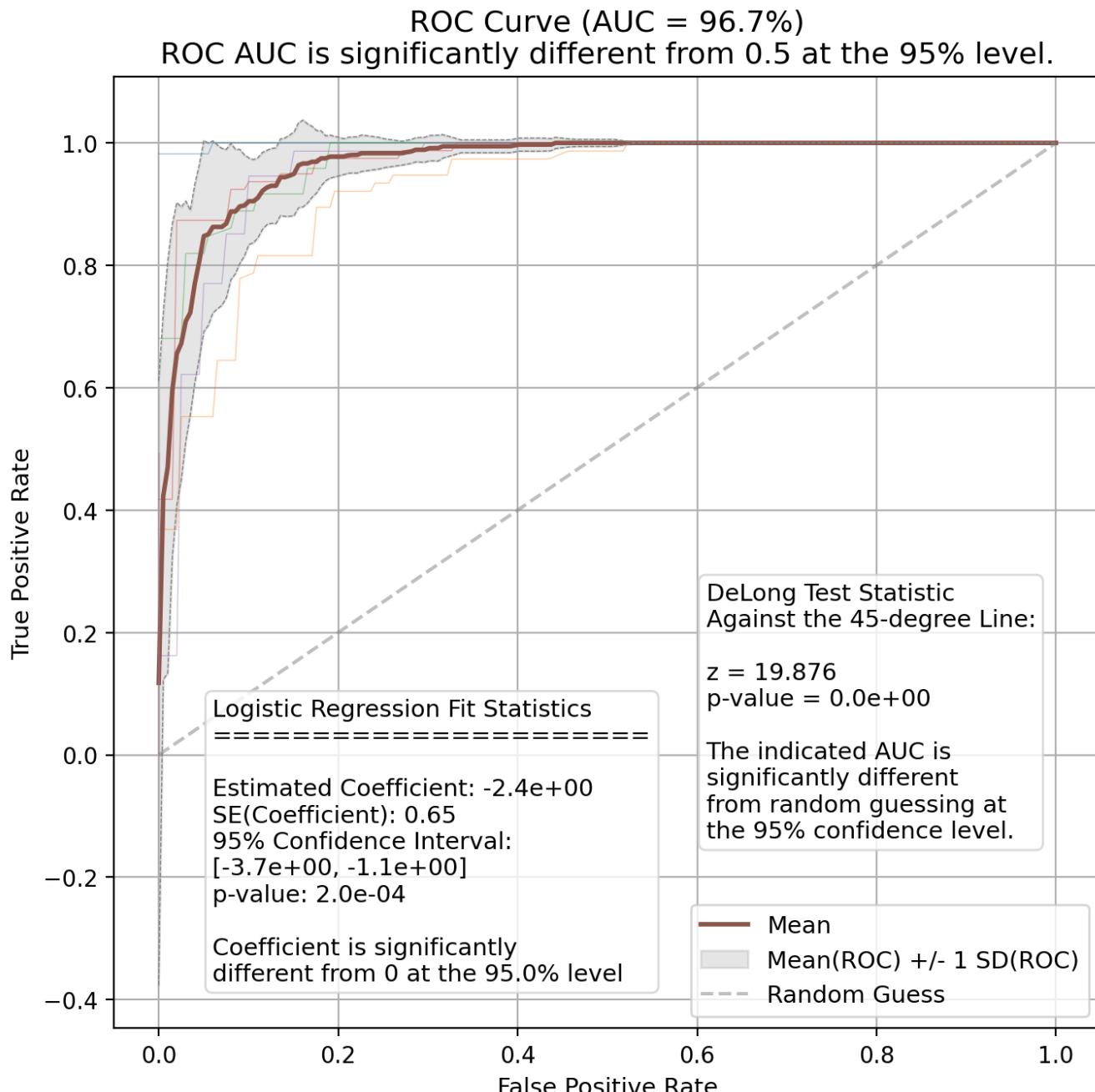
Worst Concave Points - Empirical CDF Plot



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 Concave Points - ROC Curve



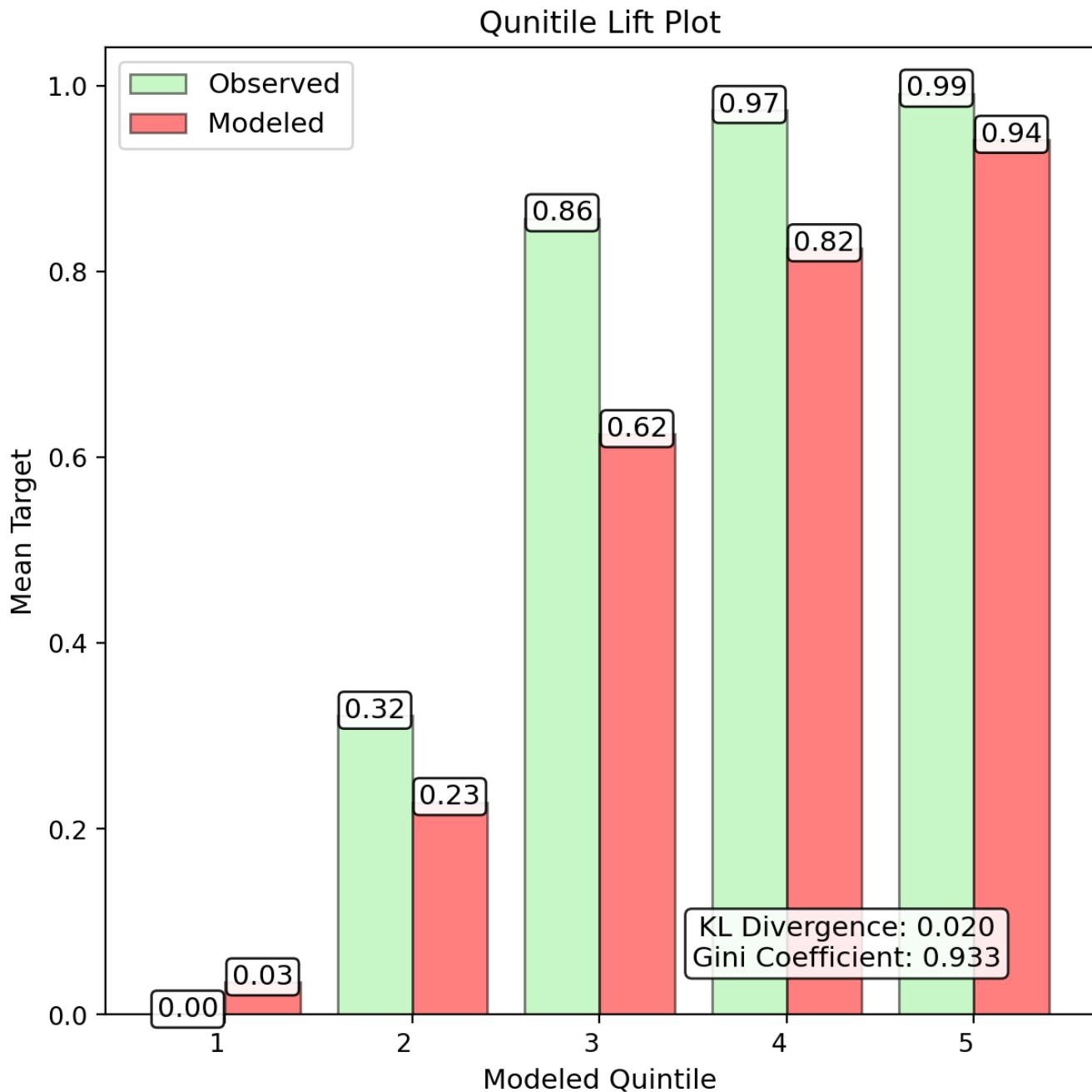
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Concave Points - 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

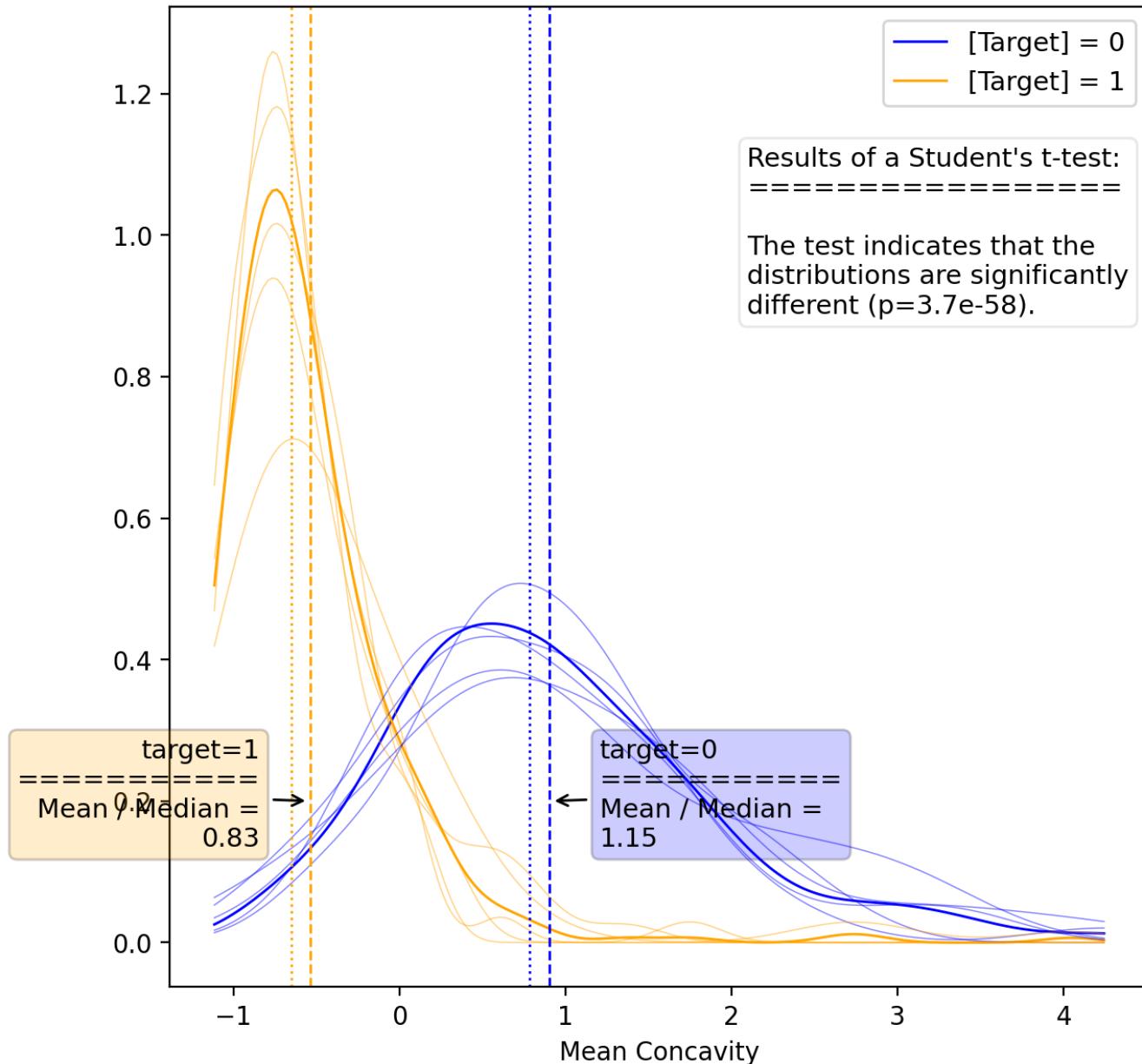
Mean Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.7e+00	-3.6e+00	-2.8e+00	-2.6e+00	-3.0e+00	-4.7e+00	4.0e-01
Fitted p-Value	4.6e-31	9.1e-30	8.4e-31	4.6e-29	3.2e-30	5.6e-09	1.9e-29
Fitted Std. Err.	0.233	0.319	0.242	0.235	0.266	0.802	0.036
Conf. Int. Lower	-3.2e+00	-4.2e+00	-3.3e+00	-3.1e+00	-3.6e+00	-6.2e+00	4.7e-01
Conf. Int. Upper	-2.2e+00	-3.0e+00	-2.3e+00	-2.2e+00	-2.5e+00	-3.1e+00	3.3e-01
Train Accuracy	87.6%	89.5%	88.0%	87.4%	89.0%	88.4%	0.9%
Val Accuracy	92.4%	82.0%	89.9%	91.6%	86.1%	88.4%	4.3%
Train AUC	87.4%	89.0%	87.8%	87.3%	88.5%	88.1%	0.7%
Val AUC	91.8%	82.5%	89.1%	90.7%	86.5%	88.1%	3.7%
Train F1	90.0%	91.6%	90.2%	89.9%	91.1%	90.6%	0.7%
Test F1	93.8%	84.7%	92.3%	93.2%	88.7%	90.6%	3.8%
Train Precision	91.7%	92.4%	91.7%	92.1%	91.8%	91.9%	0.3%
Val Precision	93.0%	89.7%	93.0%	91.5%	92.6%	91.9%	1.4%
Train Recall	88.4%	90.7%	88.8%	87.8%	90.5%	89.4%	1.3%
Val Recall	94.6%	80.3%	91.7%	94.9%	85.1%	89.4%	6.4%
Train MCC	73.9%	77.6%	74.9%	73.5%	76.7%	75.5%	1.8%
Val MCC	84.0%	63.5%	77.7%	82.4%	71.1%	75.5%	8.5%
Train Log-Loss	4.46	3.79	4.31	4.53	3.97	4.18	0.32
Val Log-Loss	2.74	6.50	3.64	3.03	5.01	4.18	1.56

Univariate Report

Mean Concavity - Kernel Density Plot

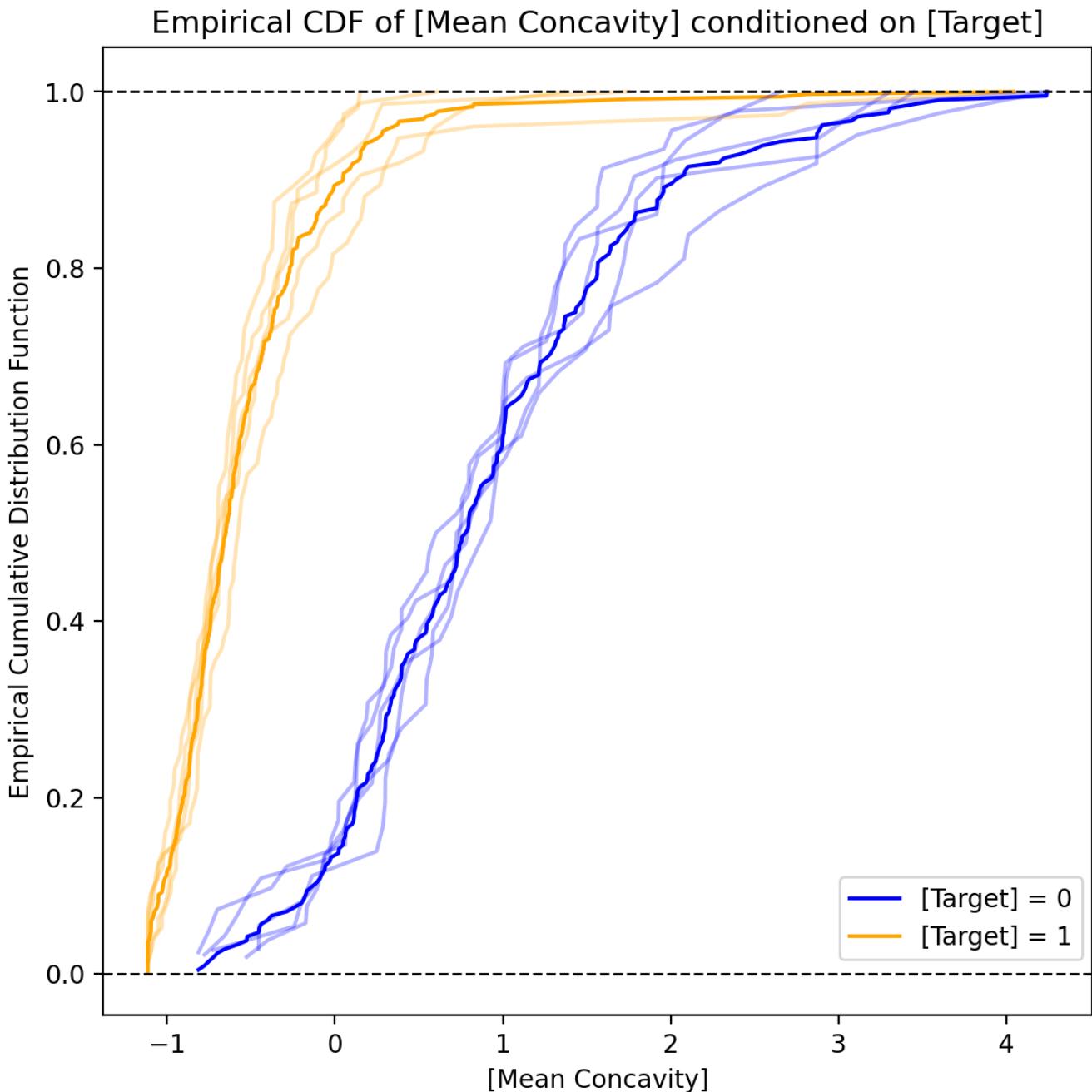
Kernel Density Plot of [Mean Concavity] by [Target]
Distributions by level are significantly different at the 95% level.



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

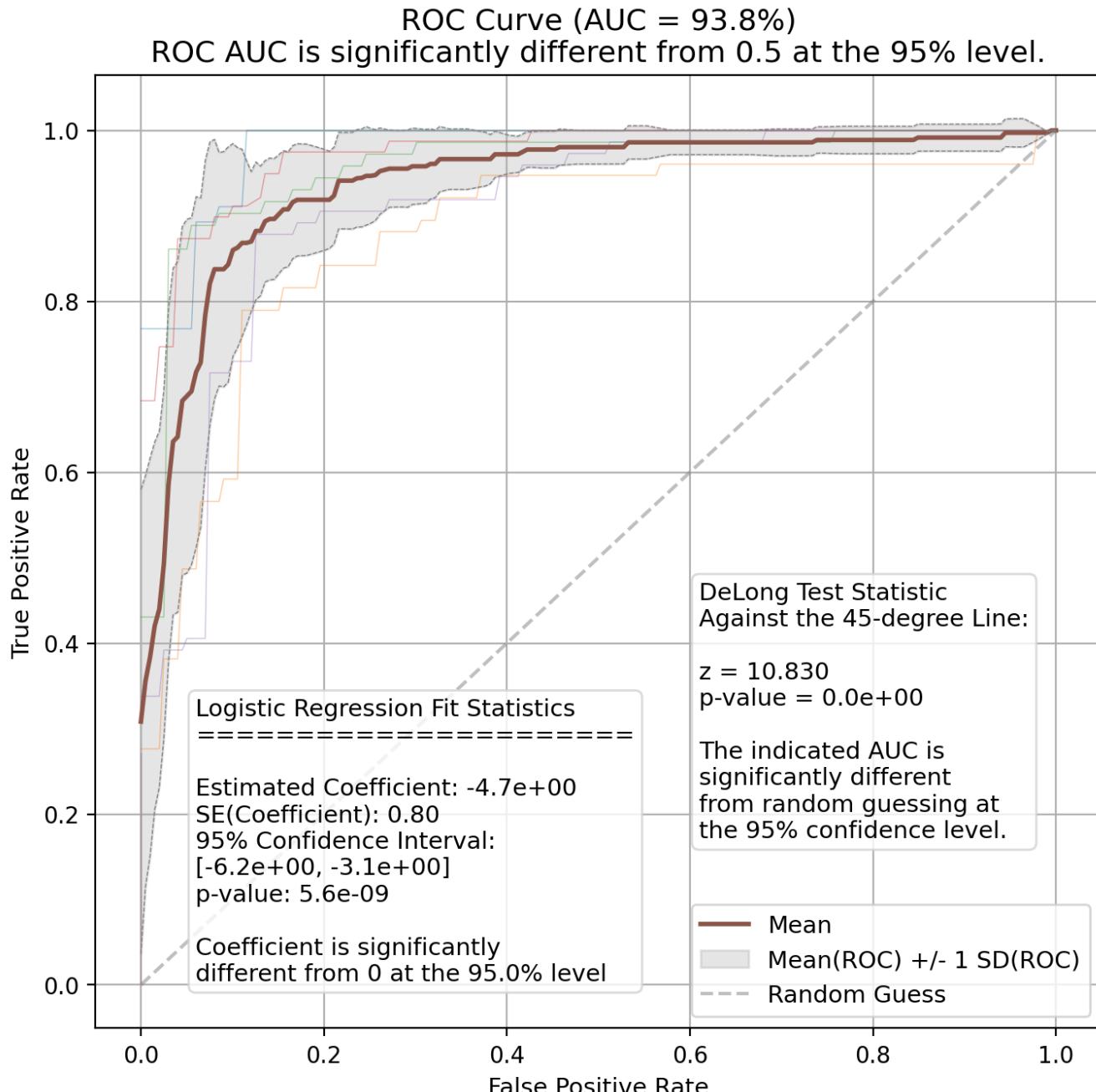
Mean Concavity - Empirical CDF Plot



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 Concavity - ROC Curve



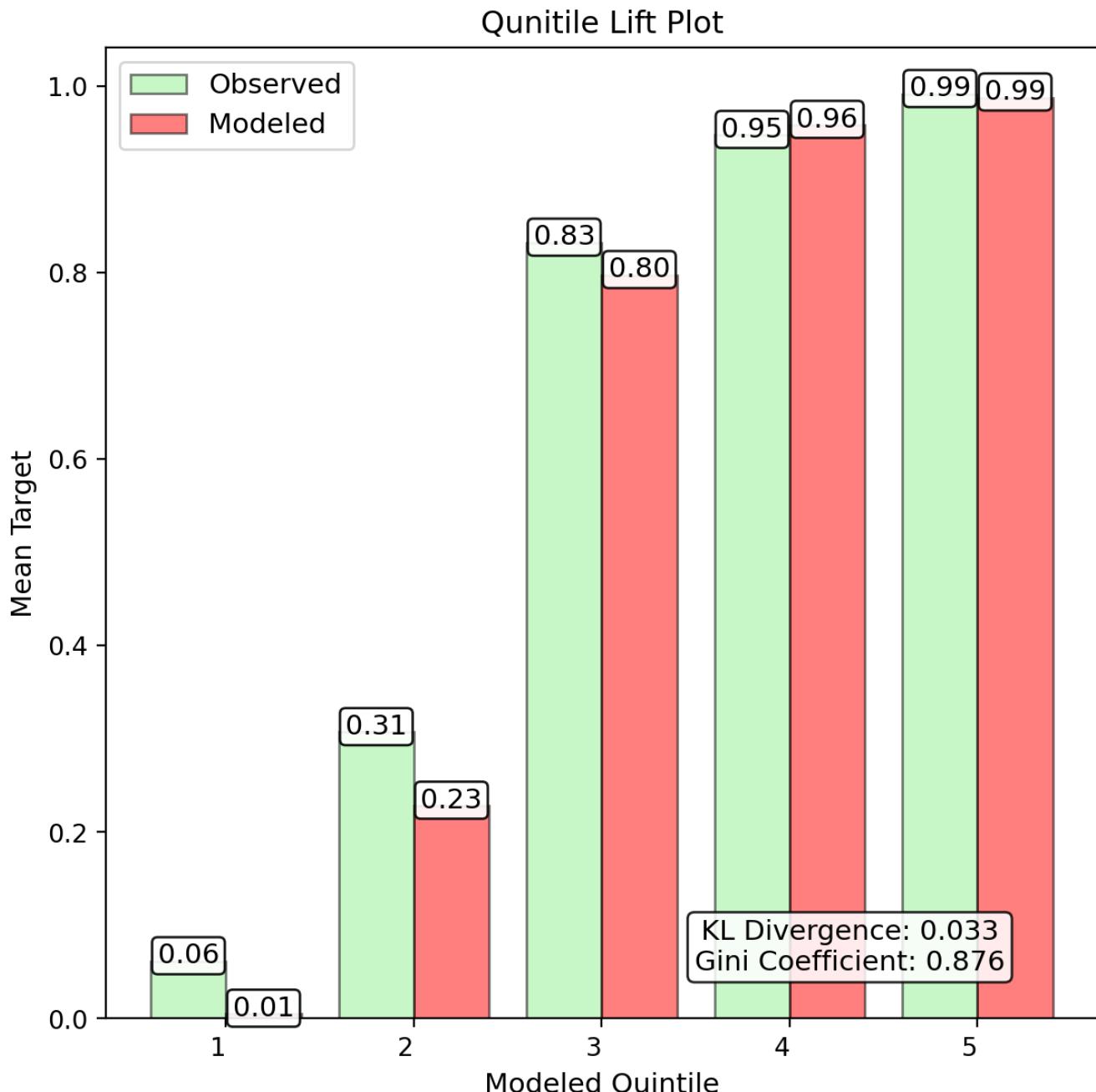
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean Concavity - 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

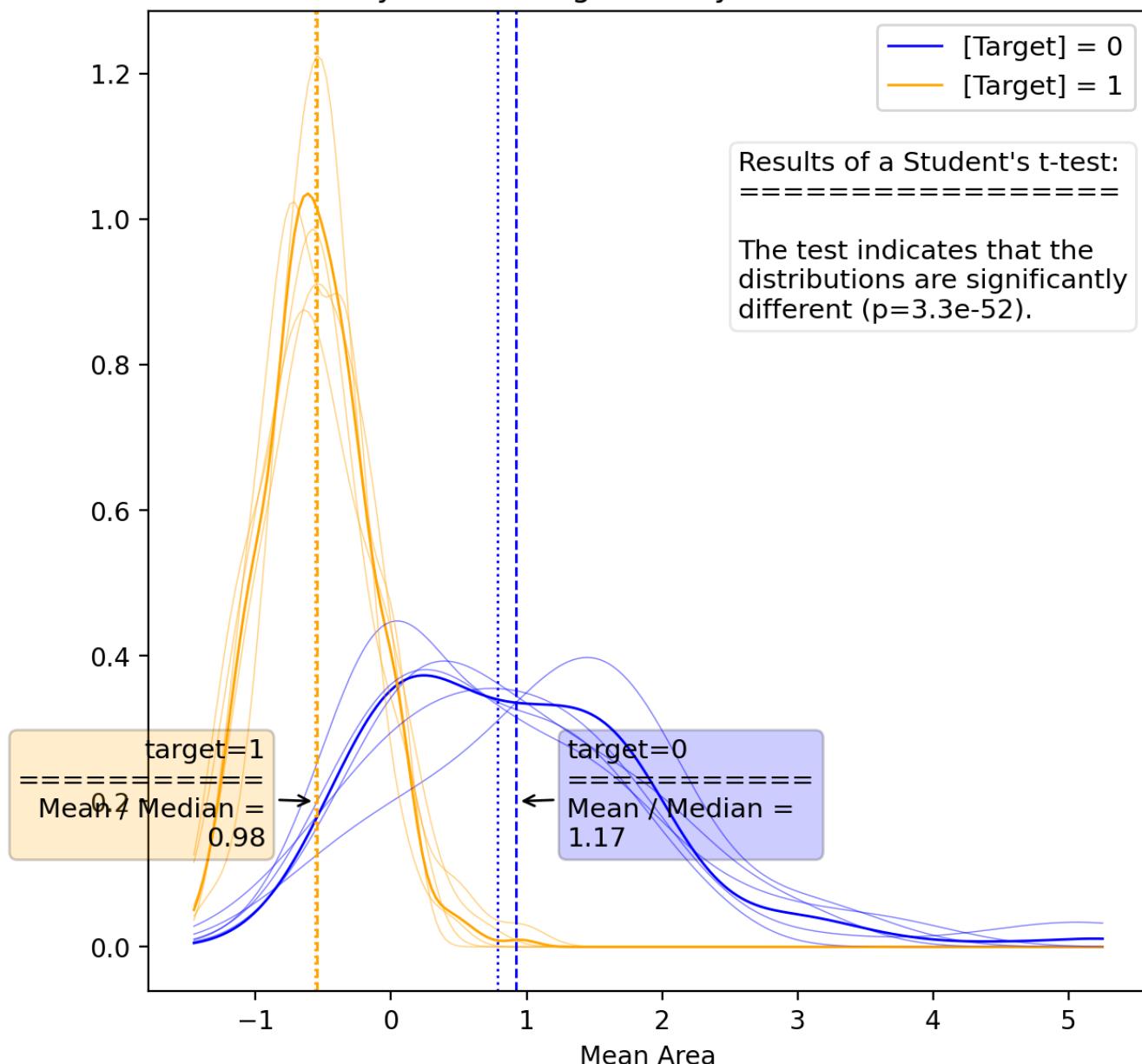
Mean Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.0e+00	-4.0e+00	-3.9e+00	-3.9e+00	-3.9e+00	-2.7e-04	6.3e-02
Fitted p-Value	2.7e-28	1.5e-26	2.8e-27	4.8e-26	2.8e-27	2.0e-02	2.0e-26
Fitted Std. Err.	0.359	0.376	0.358	0.365	0.363	0.000	0.007
Conf. Int. Lower	-4.7e+00	-4.7e+00	-4.6e+00	-4.6e+00	-4.6e+00	-4.9e-04	7.3e-02
Conf. Int. Upper	-3.3e+00	-3.3e+00	-3.2e+00	-3.1e+00	-3.2e+00	-4.1e-05	5.6e-02
Train Accuracy	88.5%	86.8%	85.9%	86.3%	87.9%	87.0%	1.1%
Val Accuracy	83.7%	86.9%	91.7%	87.8%	85.2%	87.0%	3.0%
Train AUC	87.1%	85.3%	84.6%	85.1%	86.6%	85.7%	1.0%
Val AUC	81.2%	86.5%	90.5%	86.9%	83.1%	85.7%	3.6%
Train F1	91.0%	89.7%	88.7%	89.2%	90.4%	89.8%	0.9%
Test F1	87.4%	89.3%	93.8%	90.0%	88.7%	89.8%	2.4%
Train Precision	89.7%	88.3%	87.7%	88.9%	89.0%	88.8%	0.8%
Val Precision	82.5%	90.5%	93.2%	88.9%	87.0%	88.8%	4.0%
Train Recall	92.4%	91.1%	89.8%	89.6%	91.9%	90.8%	1.2%
Val Recall	92.9%	88.2%	94.4%	91.1%	90.5%	90.8%	2.4%
Train MCC	75.0%	71.5%	69.8%	70.4%	74.0%	72.0%	2.3%
Val MCC	65.4%	72.4%	81.5%	74.4%	67.4%	72.0%	6.3%
Train Log-Loss	4.16	4.76	5.09	4.94	4.37	4.69	0.39
Val Log-Loss	5.88	4.73	2.98	4.40	5.33	4.69	1.10

Univariate Report

Mean Area - Kernel Density Plot

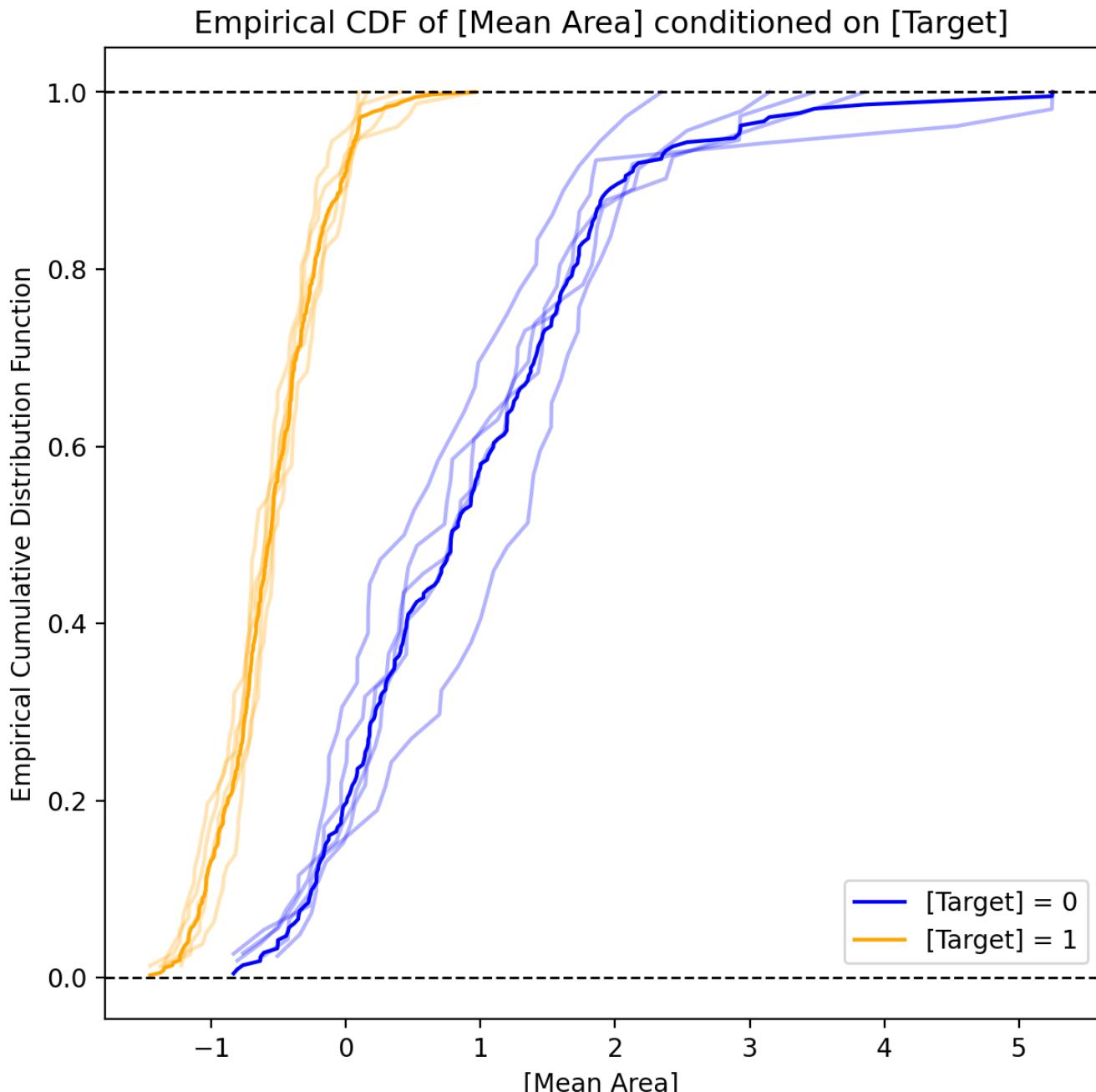
Kernel Density Plot of [Mean Area] by [Target]
Distributions by level are significantly different at the 95% level.



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

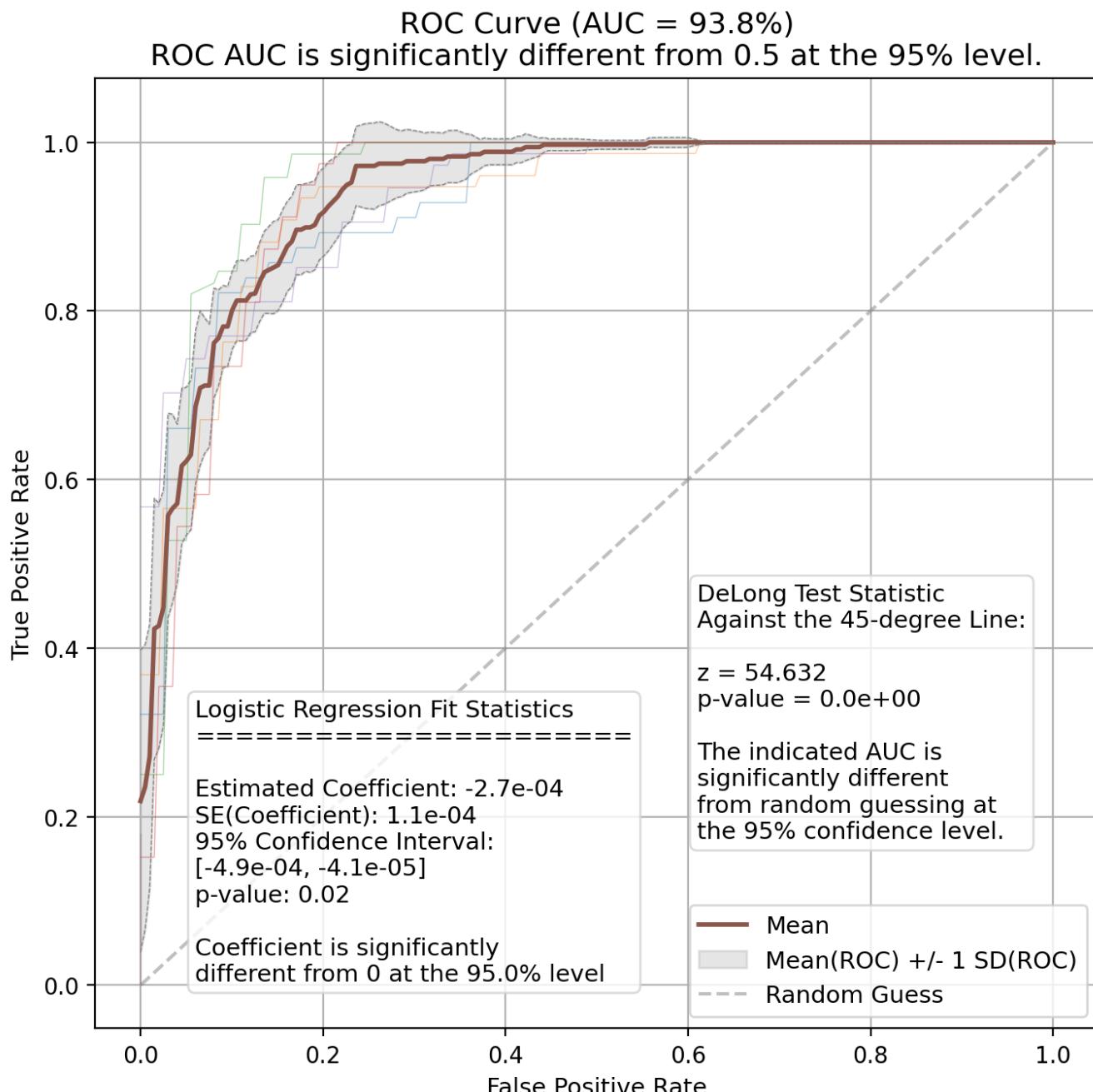
Mean Area - Empirical CDF Plot



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 Area - ROC Curve



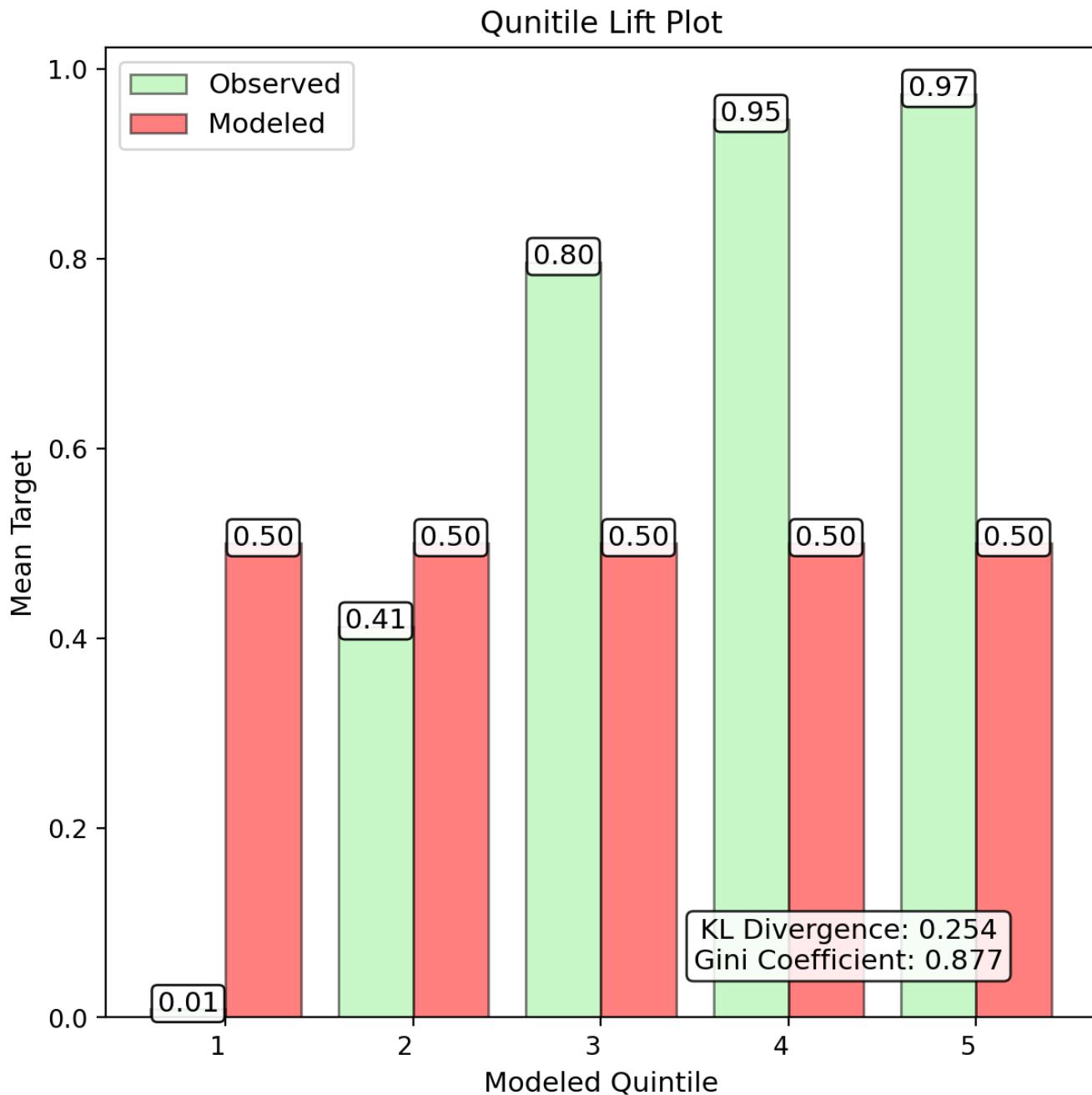
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean Area - 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

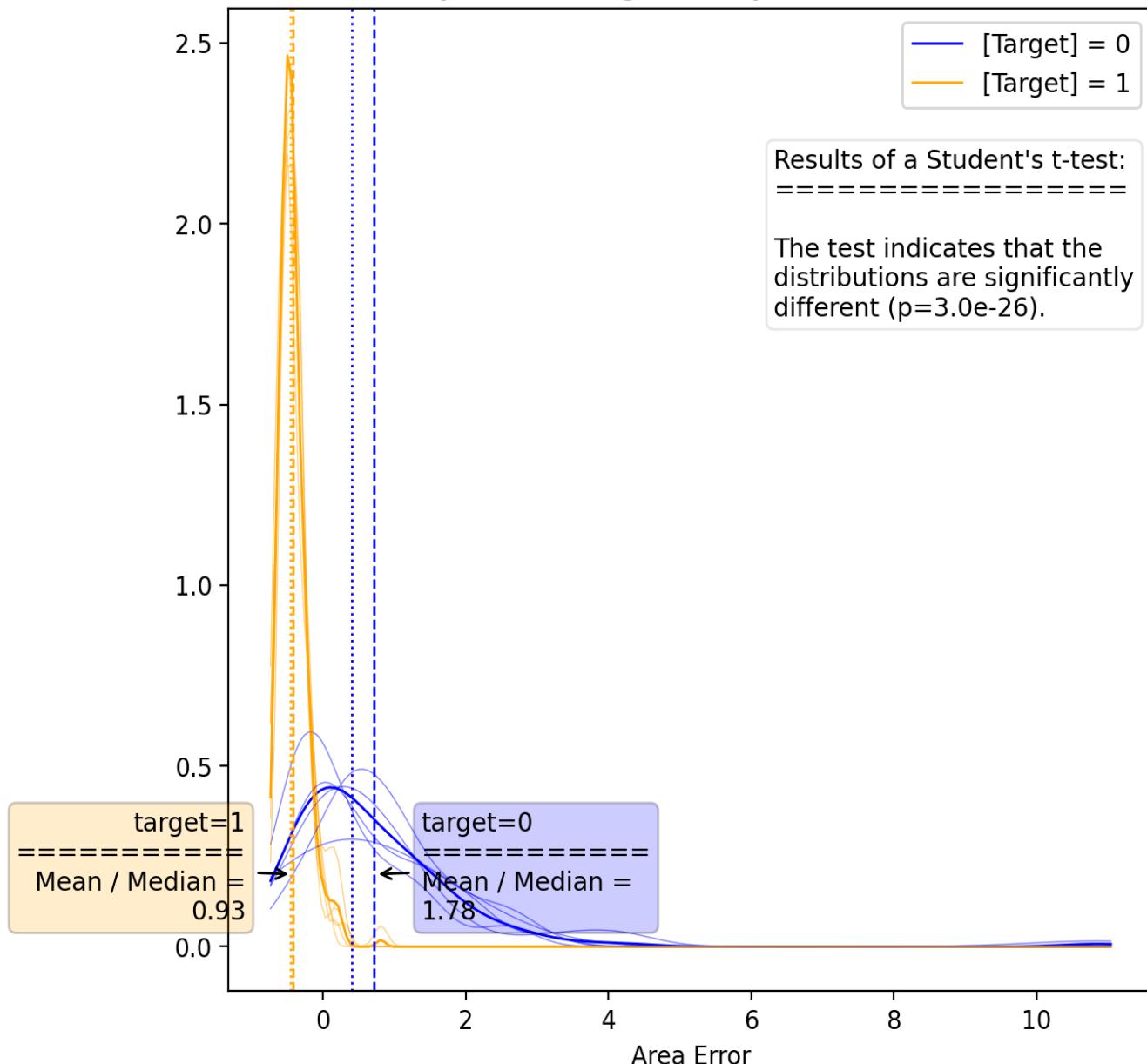
Area Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.9e+00	-4.6e+00	-4.5e+00	-4.7e+00	-4.7e+00	-9.6e-03	1.7e-01
Fitted p-Value	7.0e-33	4.0e-31	2.7e-31	6.5e-31	1.2e-31	2.2e-07	2.5e-31
Fitted Std. Err.	0.414	0.396	0.384	0.404	0.399	0.002	0.011
Conf. Int. Lower	-5.8e+00	-5.4e+00	-5.2e+00	-5.5e+00	-5.4e+00	-1.3e-02	1.9e-01
Conf. Int. Upper	-4.1e+00	-3.8e+00	-3.7e+00	-3.9e+00	-3.9e+00	-5.9e-03	1.5e-01
Train Accuracy	87.4%	87.2%	84.8%	87.0%	86.6%	86.8%	1.1%
Val Accuracy	78.3%	85.2%	93.6%	87.0%	87.0%	86.8%	5.5%
Train AUC	84.3%	83.8%	81.2%	84.0%	83.3%	83.6%	1.2%
Val AUC	72.7%	83.0%	91.9%	84.3%	83.3%	83.6%	6.8%
Train F1	90.6%	90.5%	88.7%	90.3%	90.0%	90.2%	0.8%
Test F1	84.6%	88.6%	95.2%	90.1%	90.4%	90.2%	3.8%
Train Precision	85.5%	84.8%	82.3%	86.0%	84.3%	84.7%	1.4%
Val Precision	74.3%	85.4%	93.3%	83.7%	85.5%	84.7%	6.8%
Train Recall	96.3%	97.2%	96.1%	95.0%	96.5%	96.4%	0.8%
Val Recall	98.2%	92.1%	97.2%	97.5%	95.9%	96.4%	2.4%
Train MCC	72.9%	72.8%	67.8%	71.6%	71.4%	71.8%	2.1%
Val MCC	55.9%	68.2%	85.6%	73.4%	71.3%	71.8%	10.7%
Train Log-Loss	4.53	4.60	5.48	4.69	4.84	4.75	0.38
Val Log-Loss	7.84	5.32	2.31	4.68	4.70	4.75	1.97

Univariate Report

Area Error - Kernel Density Plot

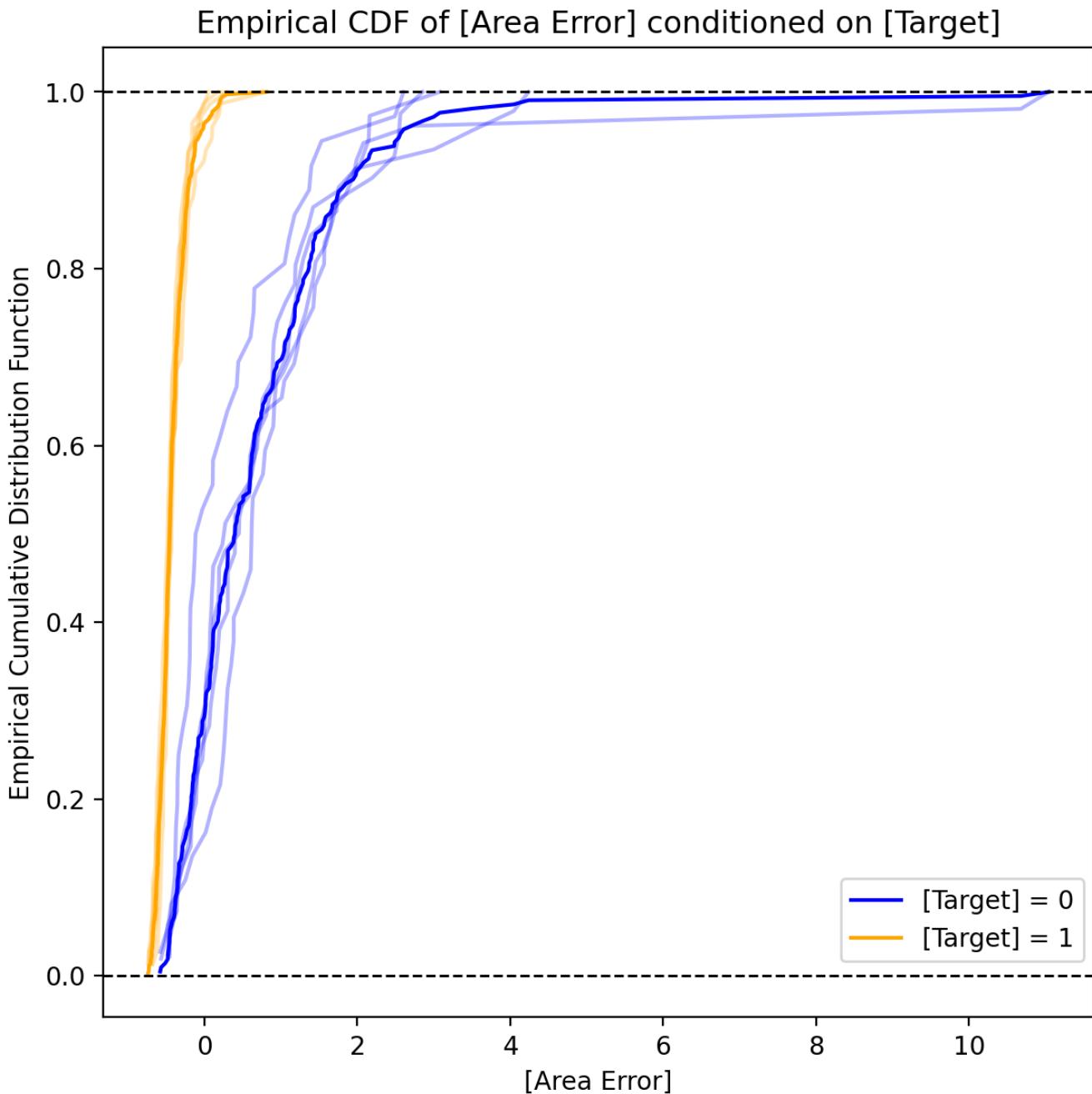
Kernel Density Plot of [Area Error] by [Target]
Distributions by level are significantly different at the 95% level.



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

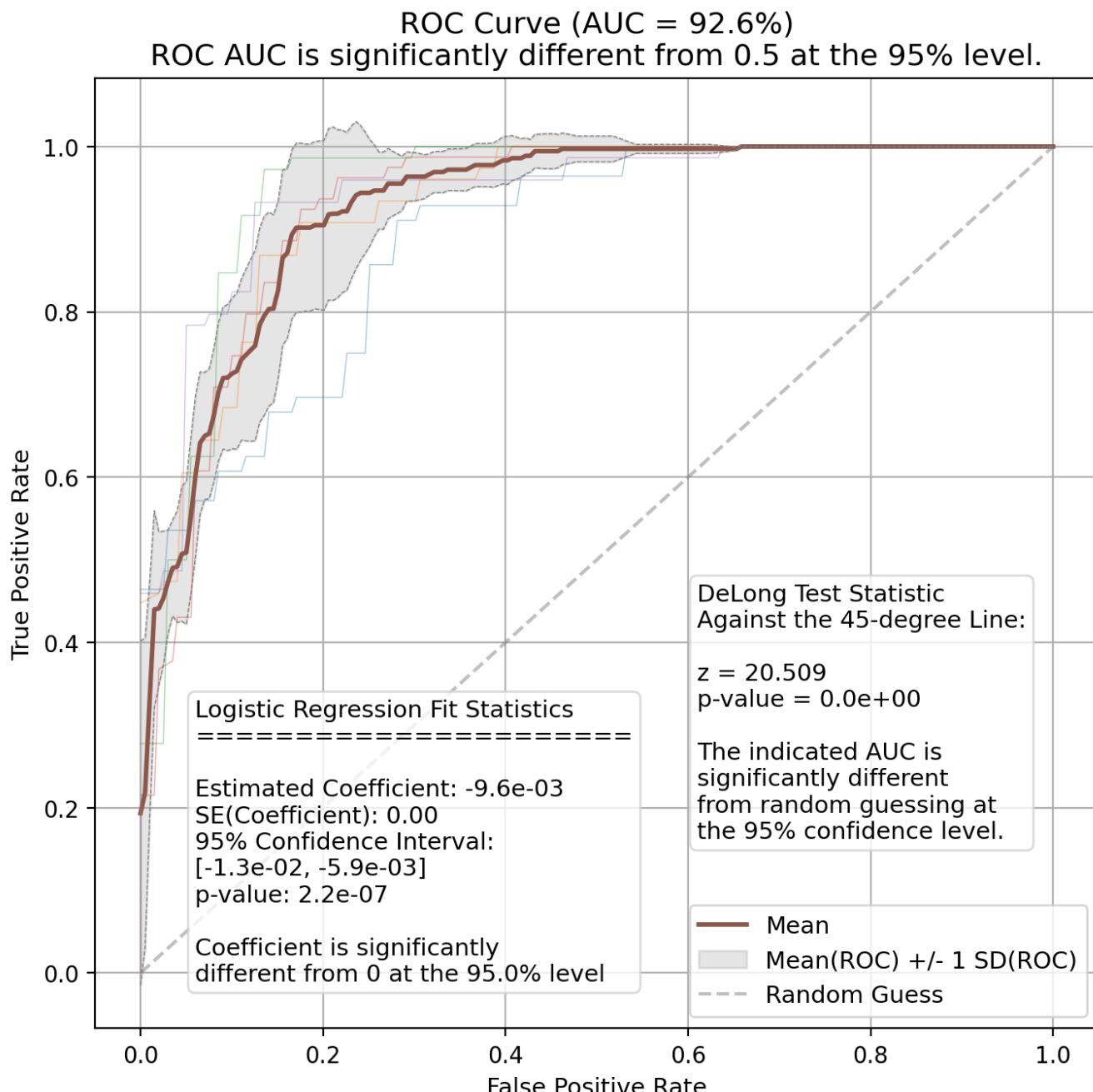
Area Error - Empirical CDF Plot



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

Area Error - ROC Curve



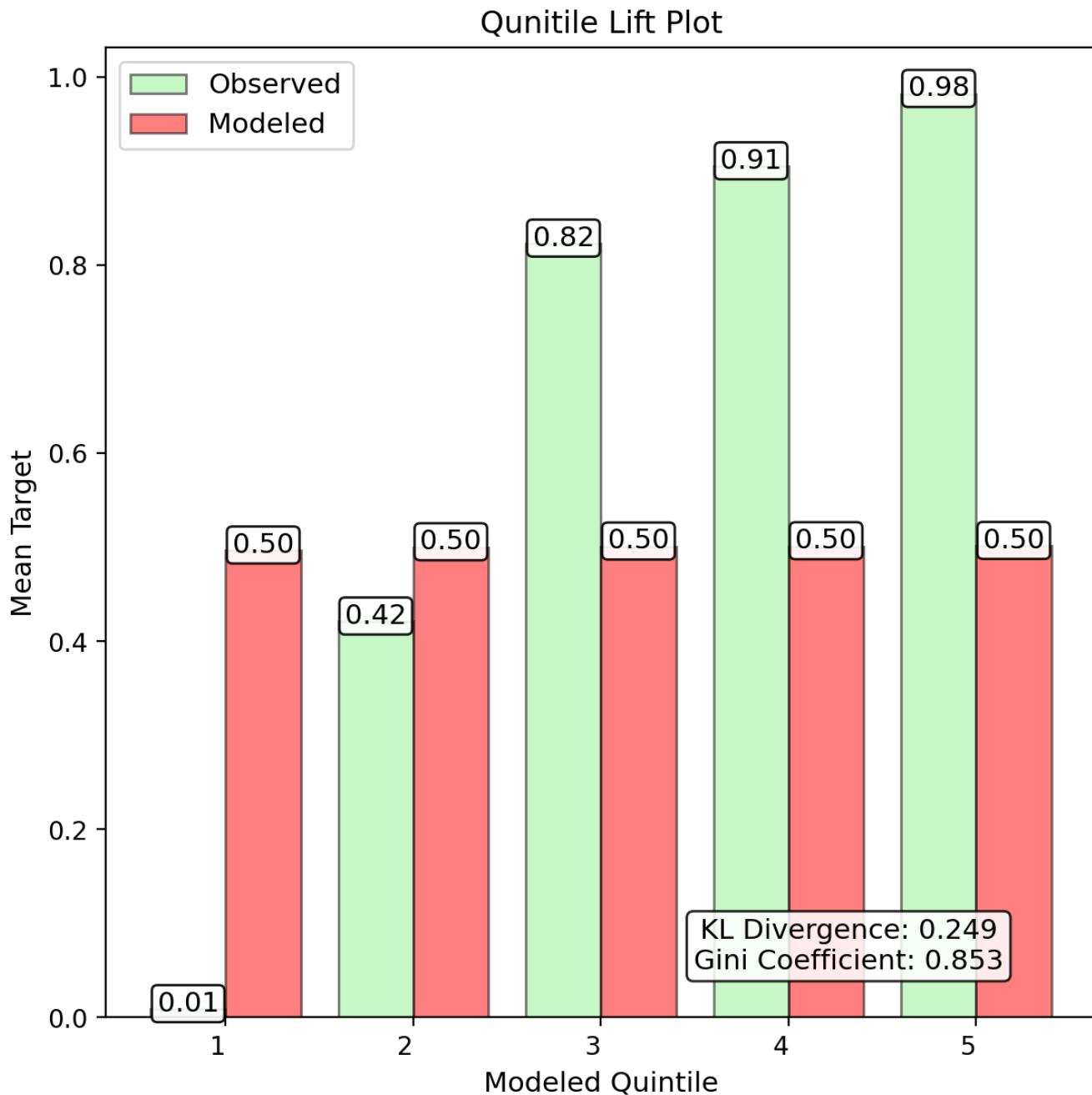
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Area 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

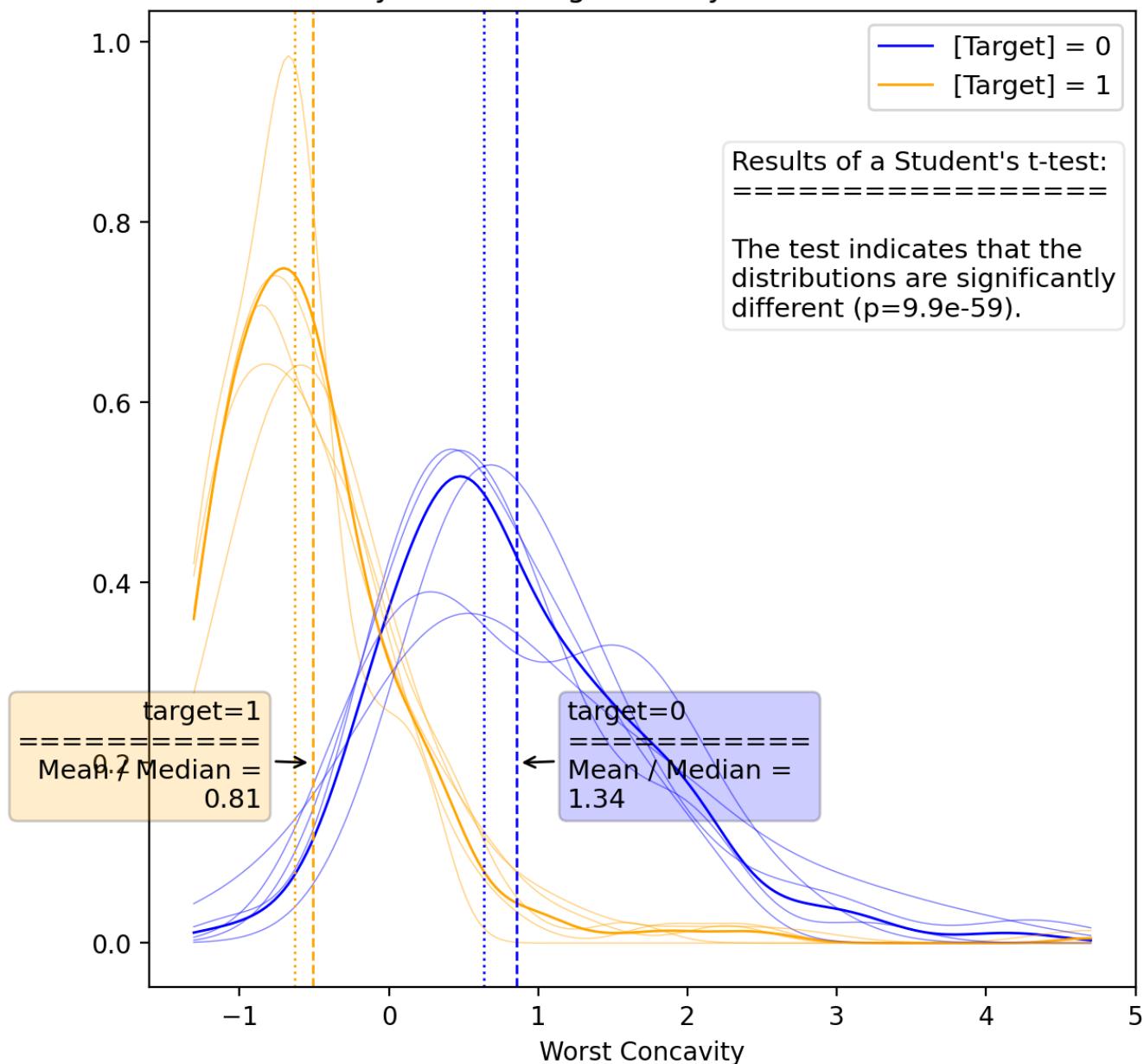
Worst Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.1e+00	-2.7e+00	-2.3e+00	-2.3e+00	-2.5e+00	-1.1e+00	2.3e-01
Fitted p-Value	6.1e-28	1.3e-28	2.1e-28	4.9e-27	2.0e-28	2.0e-05	2.1e-27
Fitted Std. Err.	0.193	0.244	0.208	0.212	0.224	0.258	0.019
Conf. Int. Lower	-2.5e+00	-3.2e+00	-2.7e+00	-2.7e+00	-2.9e+00	-1.6e+00	2.6e-01
Conf. Int. Upper	-1.7e+00	-2.2e+00	-1.9e+00	-1.9e+00	-2.0e+00	-5.9e-01	1.9e-01
Train Accuracy	83.0%	86.1%	85.2%	84.5%	85.2%	84.7%	1.2%
Val Accuracy	93.5%	81.1%	82.6%	85.5%	82.6%	84.7%	5.0%
Train AUC	83.0%	86.3%	85.5%	84.7%	85.4%	84.8%	1.2%
Val AUC	94.1%	81.9%	81.6%	85.3%	82.7%	84.8%	5.3%
Train F1	86.1%	88.6%	87.6%	87.3%	87.8%	87.4%	0.9%
Test F1	94.4%	83.9%	86.5%	87.7%	85.9%	87.4%	4.0%
Train Precision	89.3%	91.6%	91.3%	91.0%	90.9%	90.7%	0.9%
Val Precision	98.1%	89.6%	88.4%	89.5%	89.7%	90.7%	4.0%
Train Recall	83.1%	85.8%	84.2%	83.8%	84.8%	84.3%	1.0%
Val Recall	91.1%	78.9%	84.7%	86.1%	82.4%	84.3%	4.5%
Train MCC	64.7%	71.2%	69.7%	67.8%	69.5%	68.3%	2.5%
Val MCC	86.9%	62.1%	62.0%	70.1%	63.7%	68.3%	10.6%
Train Log-Loss	6.12	5.00	5.33	5.60	5.32	5.51	0.42
Val Log-Loss	2.35	6.80	6.28	5.23	6.27	5.51	1.79

Univariate Report

Worst Concavity - Kernel Density Plot

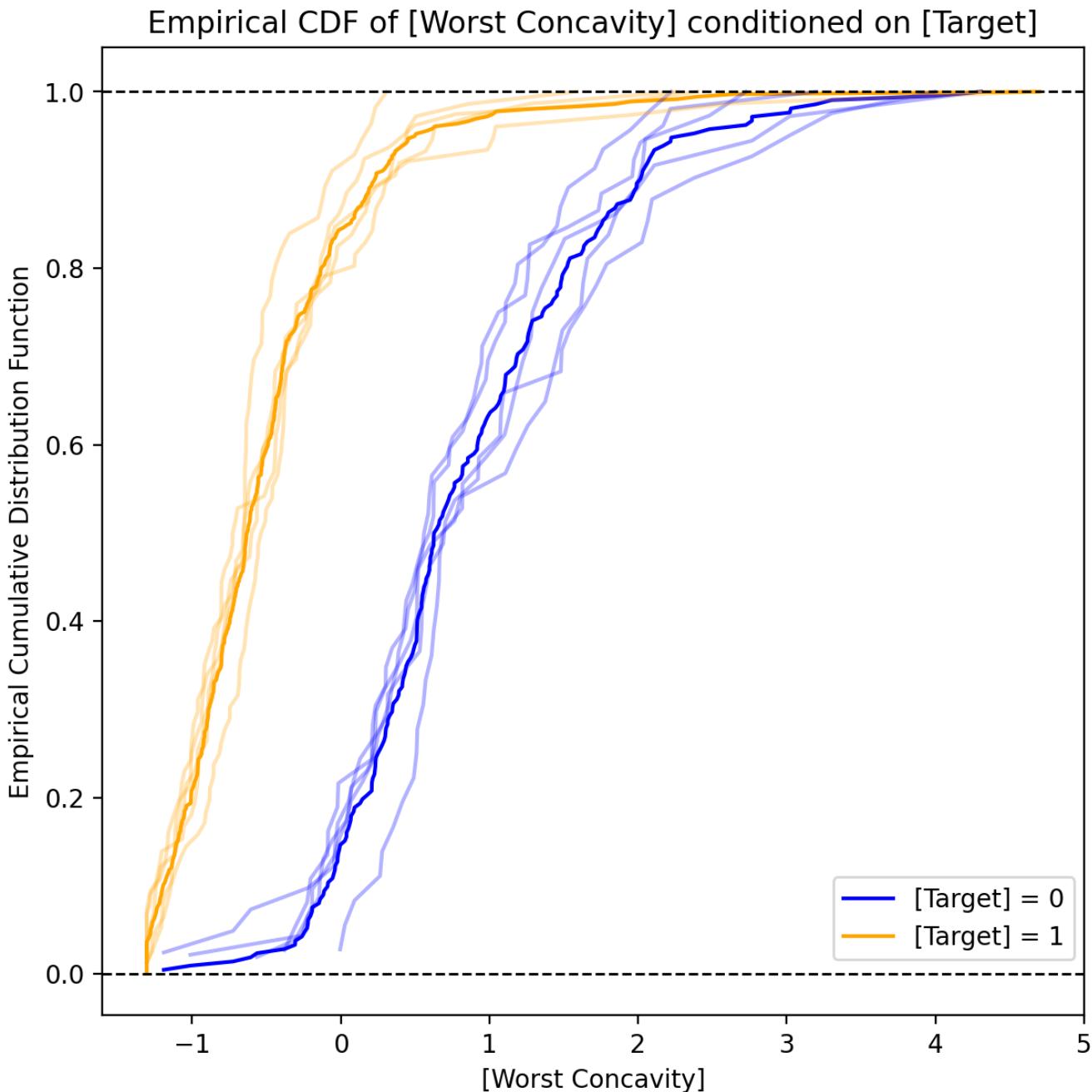
Kernel Density Plot of [Worst Concavity] by [Target]
Distributions by level are significantly different at the 95% level.



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

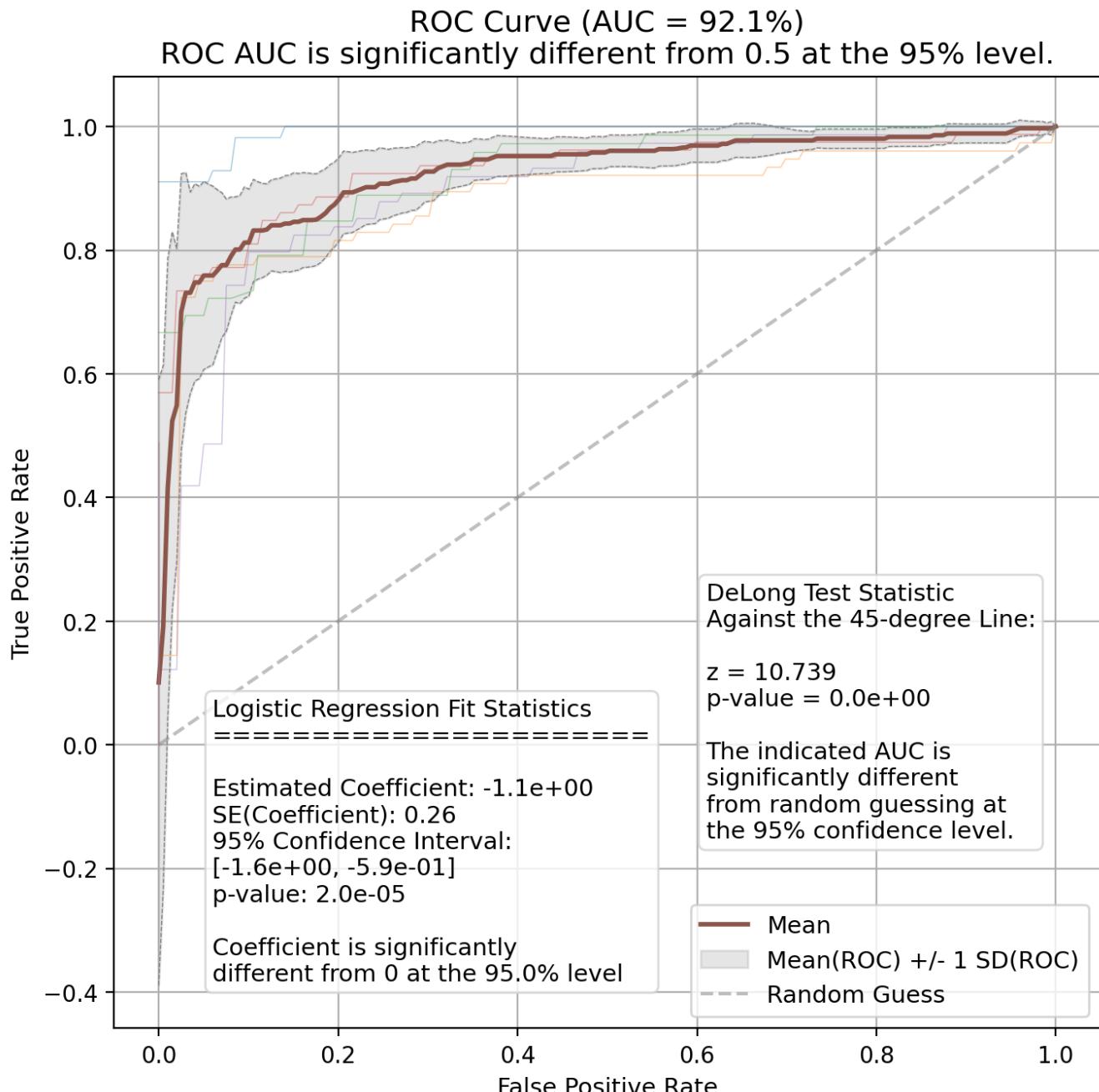
Worst Concavity - Empirical CDF Plot



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 Concavity - ROC Curve



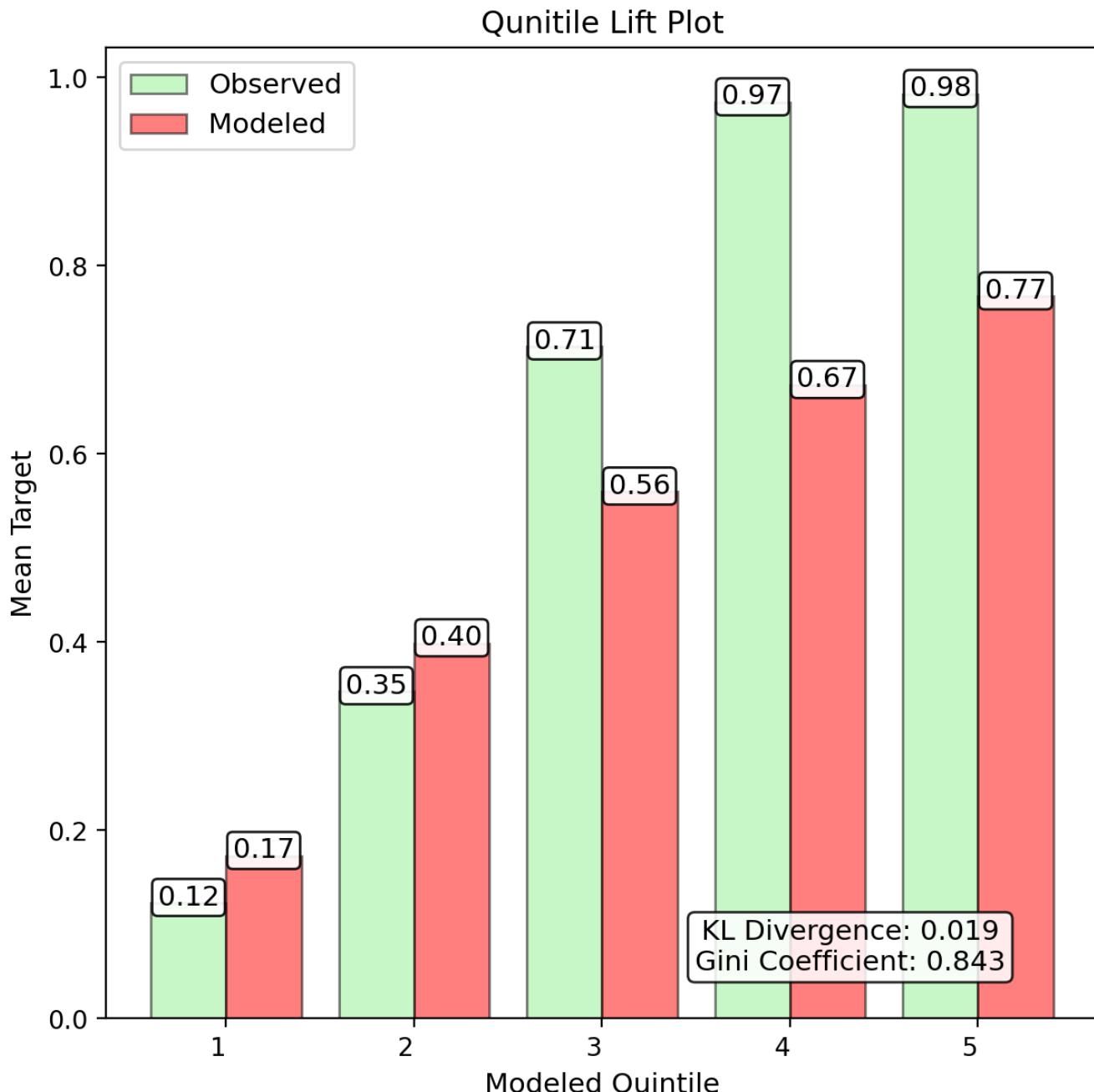
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a \pm two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Concavity - 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

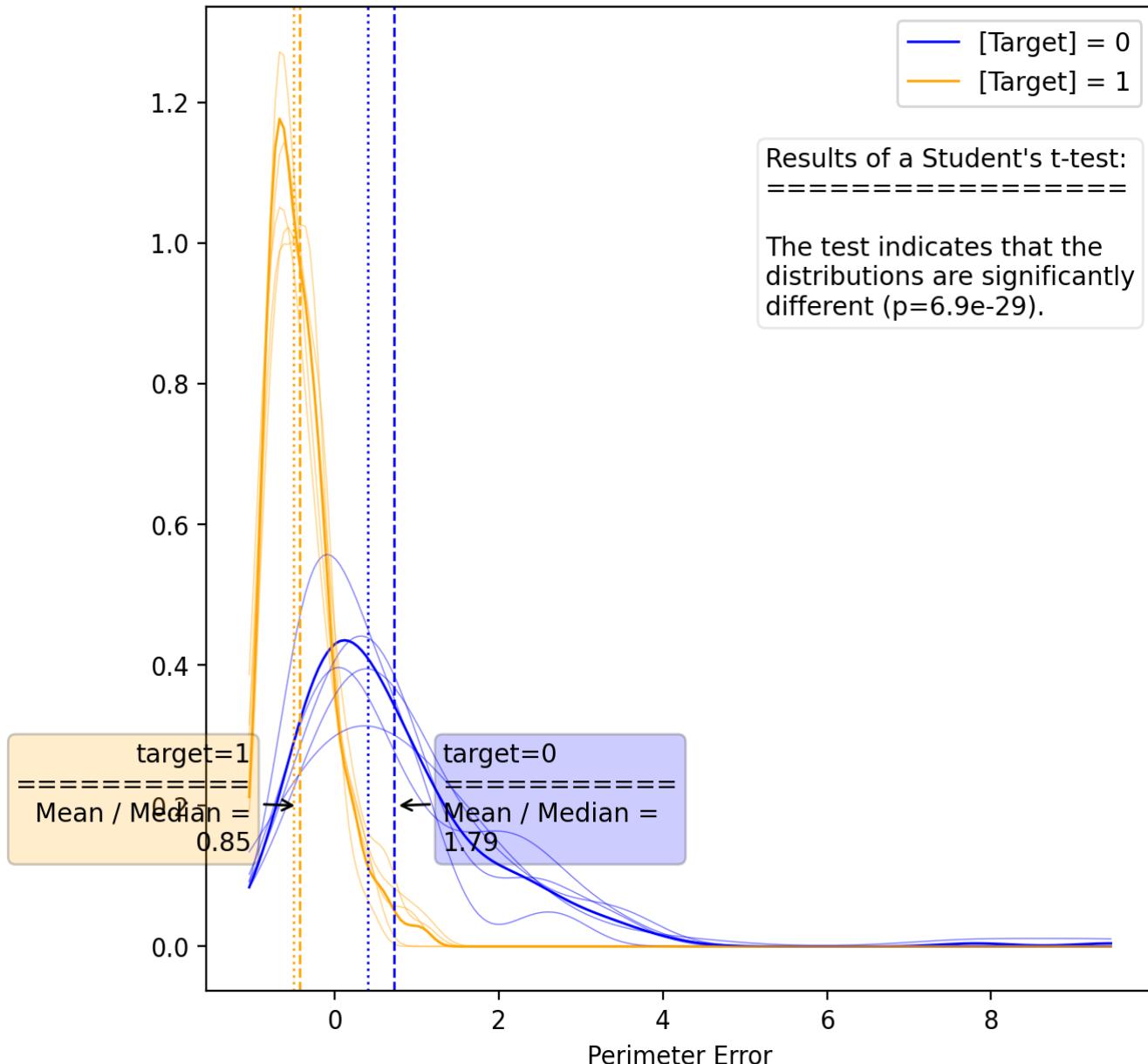
Perimeter Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.0e+00	-3.1e+00	-3.0e+00	-2.8e+00	-3.0e+00	-5.8e-02	9.5e-02
Fitted p-Value	1.5e-28	4.4e-27	3.4e-27	7.5e-26	2.7e-27	1.9e-02	3.2e-26
Fitted Std. Err.	0.269	0.285	0.280	0.268	0.275	0.025	0.007
Conf. Int. Lower	-3.5e+00	-3.6e+00	-3.6e+00	-3.3e+00	-3.5e+00	-1.1e-01	1.1e-01
Conf. Int. Upper	-2.5e+00	-2.5e+00	-2.5e+00	-2.3e+00	-2.4e+00	-9.6e-03	8.4e-02
Train Accuracy	82.2%	82.3%	81.3%	81.1%	82.4%	81.9%	0.6%
Val Accuracy	78.3%	81.1%	84.4%	85.5%	80.0%	81.9%	3.0%
Train AUC	79.9%	80.0%	79.1%	79.2%	80.4%	79.8%	0.6%
Val AUC	74.7%	80.1%	83.6%	83.7%	77.4%	79.8%	3.9%
Train F1	86.3%	86.4%	85.4%	85.2%	86.2%	85.9%	0.5%
Test F1	83.6%	84.8%	87.9%	88.5%	84.8%	85.9%	2.2%
Train Precision	84.0%	83.9%	82.6%	84.5%	84.2%	84.0%	0.7%
Val Precision	77.3%	85.3%	89.9%	84.9%	83.1%	84.0%	4.6%
Train Recall	88.7%	89.0%	88.4%	86.0%	88.3%	88.0%	1.2%
Val Recall	91.1%	84.2%	86.1%	92.4%	86.5%	88.0%	3.5%
Train MCC	61.1%	61.6%	59.7%	58.9%	62.0%	60.8%	1.3%
Val MCC	53.5%	60.0%	66.0%	69.4%	55.8%	60.8%	6.7%
Train Log-Loss	6.42	6.37	6.74	6.83	6.35	6.52	0.22
Val Log-Loss	7.84	6.80	5.62	5.23	7.21	6.52	1.09

Univariate Report

Perimeter Error - Kernel Density Plot

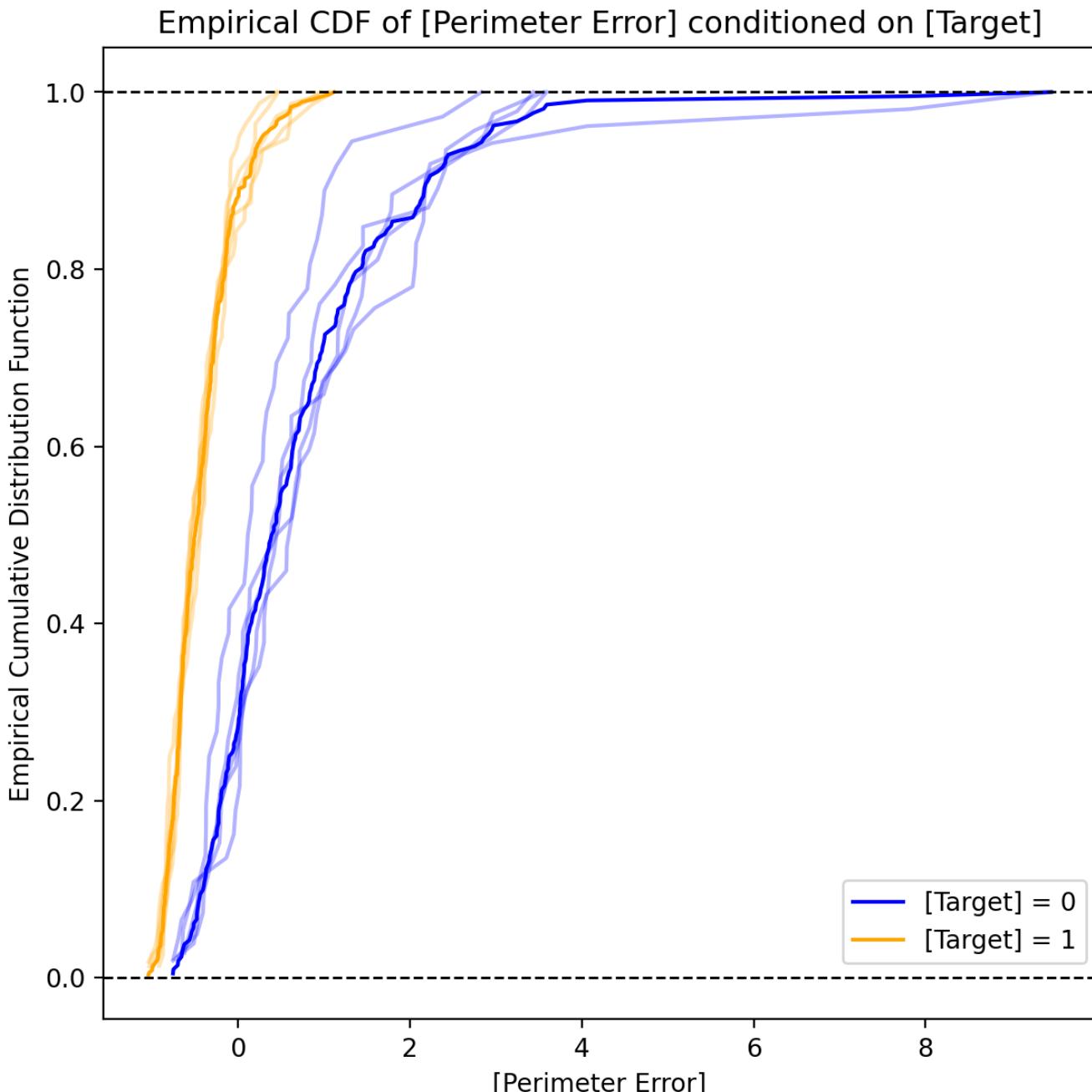
Kernel Density Plot of [Perimeter Error] by [Target]
Distributions by level are significantly different at the 95% level.



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

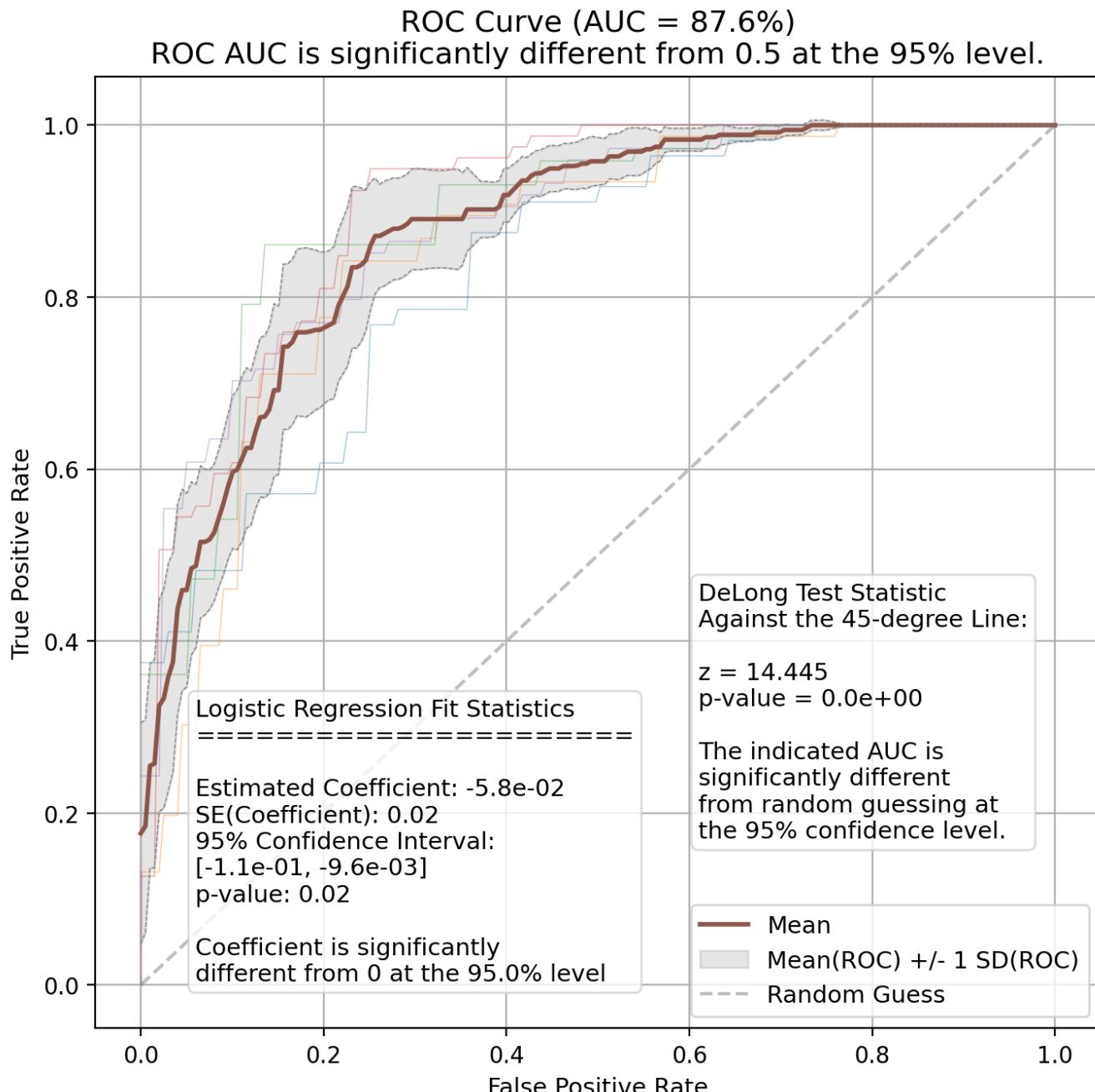
Perimeter Error - Empirical CDF Plot



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

Perimeter Error - ROC Curve



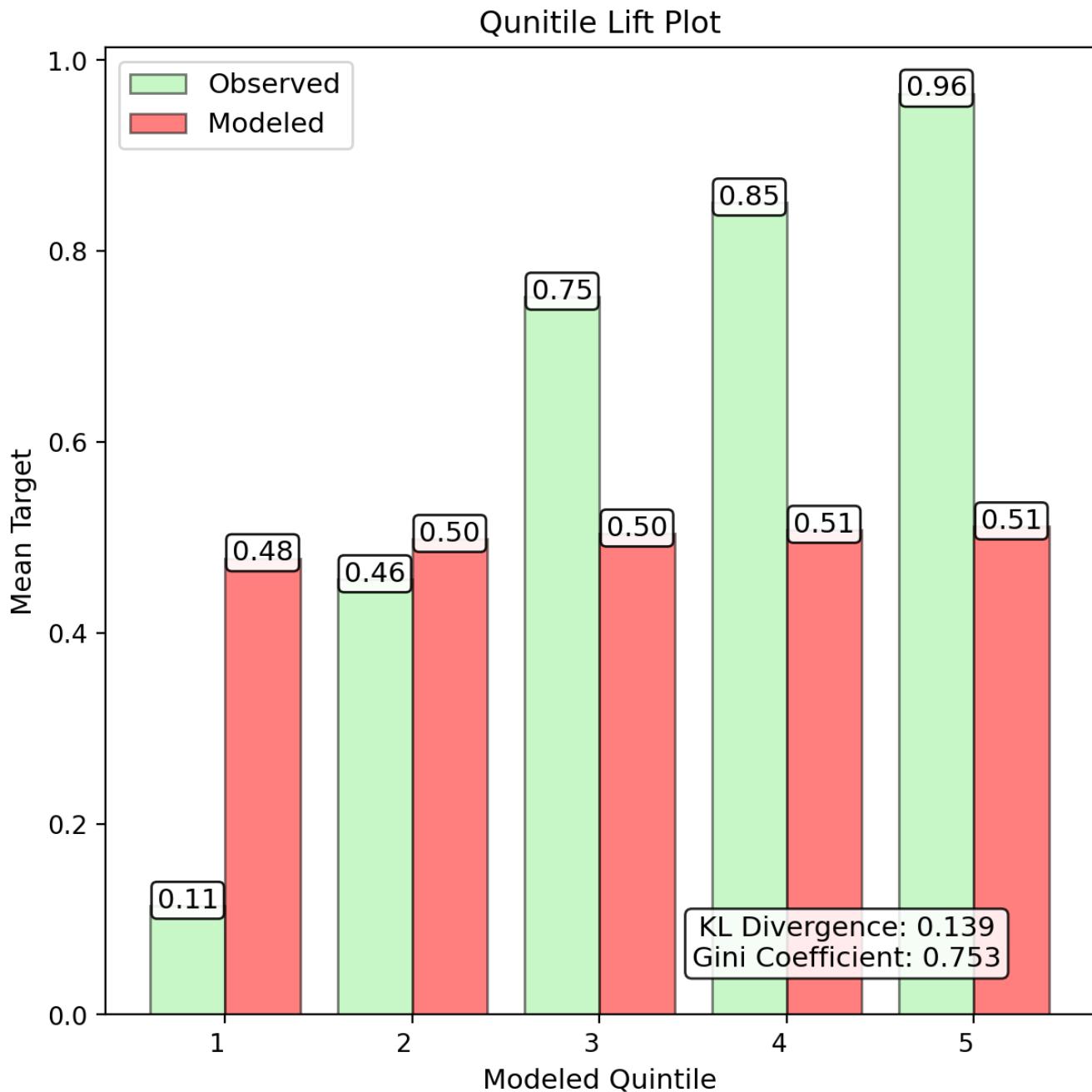
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Perimeter 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

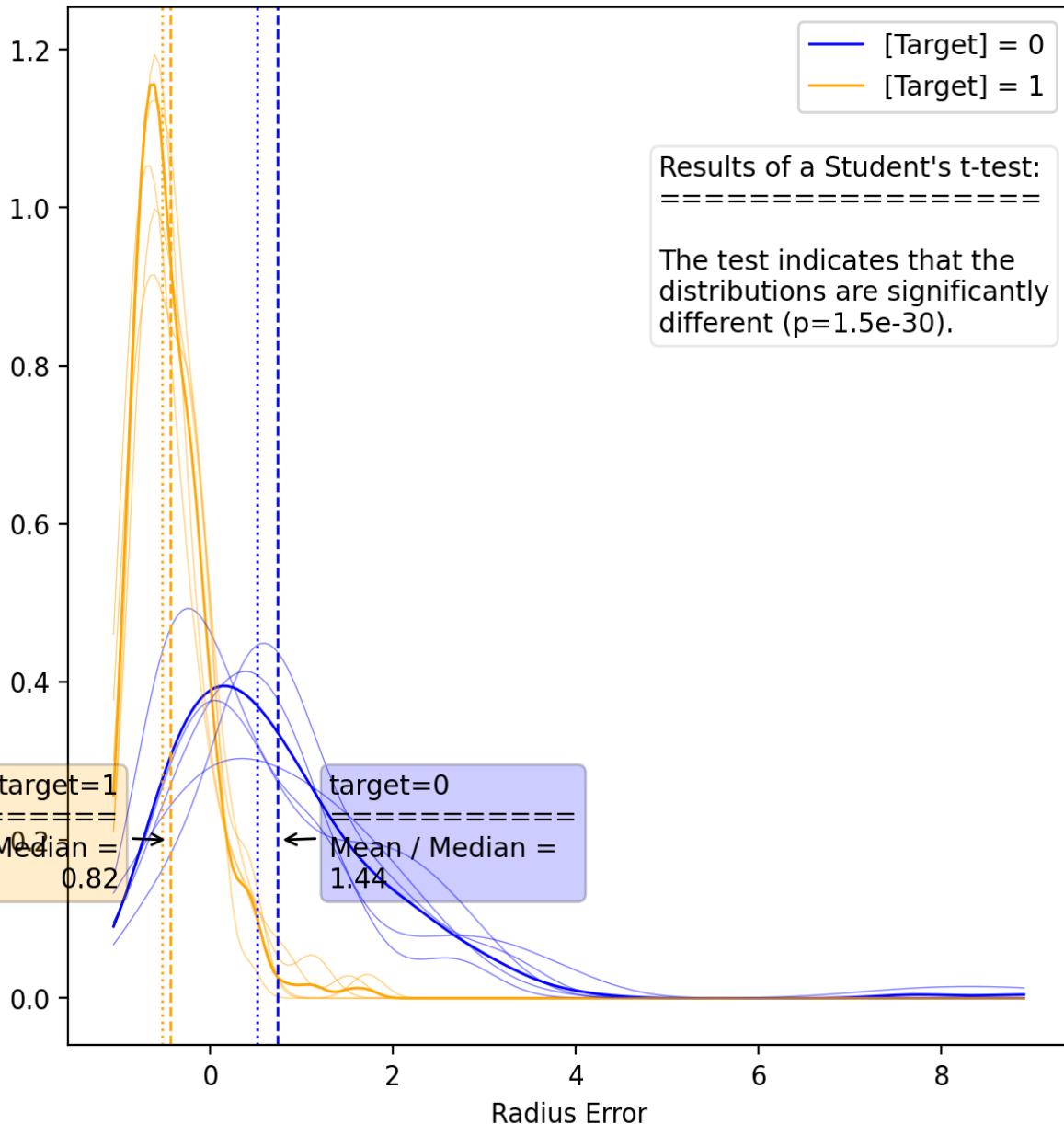
Radius Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.8e+00	-2.8e+00	-2.7e+00	-2.6e+00	-2.7e+00	-4.0e-01	9.7e-02
Fitted p-Value	1.0e-28	5.5e-27	4.1e-26	9.3e-26	6.6e-27	2.1e-02	3.9e-26
Fitted Std. Err.	0.256	0.259	0.255	0.247	0.253	0.175	0.004
Conf. Int. Lower	-3.4e+00	-3.3e+00	-3.2e+00	-3.1e+00	-3.2e+00	-7.5e-01	1.0e-01
Conf. Int. Upper	-2.3e+00	-2.3e+00	-2.2e+00	-2.1e+00	-2.2e+00	-6.0e-02	8.9e-02
Train Accuracy	83.9%	82.3%	80.7%	81.3%	81.7%	81.5%	1.2%
Val Accuracy	75.0%	80.3%	86.2%	84.0%	81.7%	81.5%	4.2%
Train AUC	81.9%	80.3%	78.1%	79.5%	79.7%	79.5%	1.4%
Val AUC	71.0%	79.5%	86.3%	82.1%	78.7%	79.5%	5.6%
Train F1	87.5%	86.3%	85.0%	85.4%	85.7%	85.6%	1.0%
Test F1	81.3%	84.0%	89.2%	87.3%	86.3%	85.6%	3.1%
Train Precision	85.7%	84.4%	81.6%	84.8%	83.6%	83.7%	1.5%
Val Precision	74.6%	85.1%	92.5%	83.7%	83.5%	83.7%	6.4%
Train Recall	89.4%	88.3%	88.8%	86.0%	88.0%	87.7%	1.3%
Val Recall	89.3%	82.9%	86.1%	91.1%	89.2%	87.7%	3.2%
Train MCC	64.9%	61.7%	58.2%	59.4%	60.5%	60.0%	2.6%
Val MCC	46.1%	58.5%	70.6%	66.2%	59.4%	60.0%	9.3%
Train Log-Loss	5.82	6.37	6.97	6.75	6.59	6.65	0.44
Val Log-Loss	9.01	7.09	4.96	5.78	6.58	6.65	1.53

Univariate Report

Radius Error - Kernel Density Plot

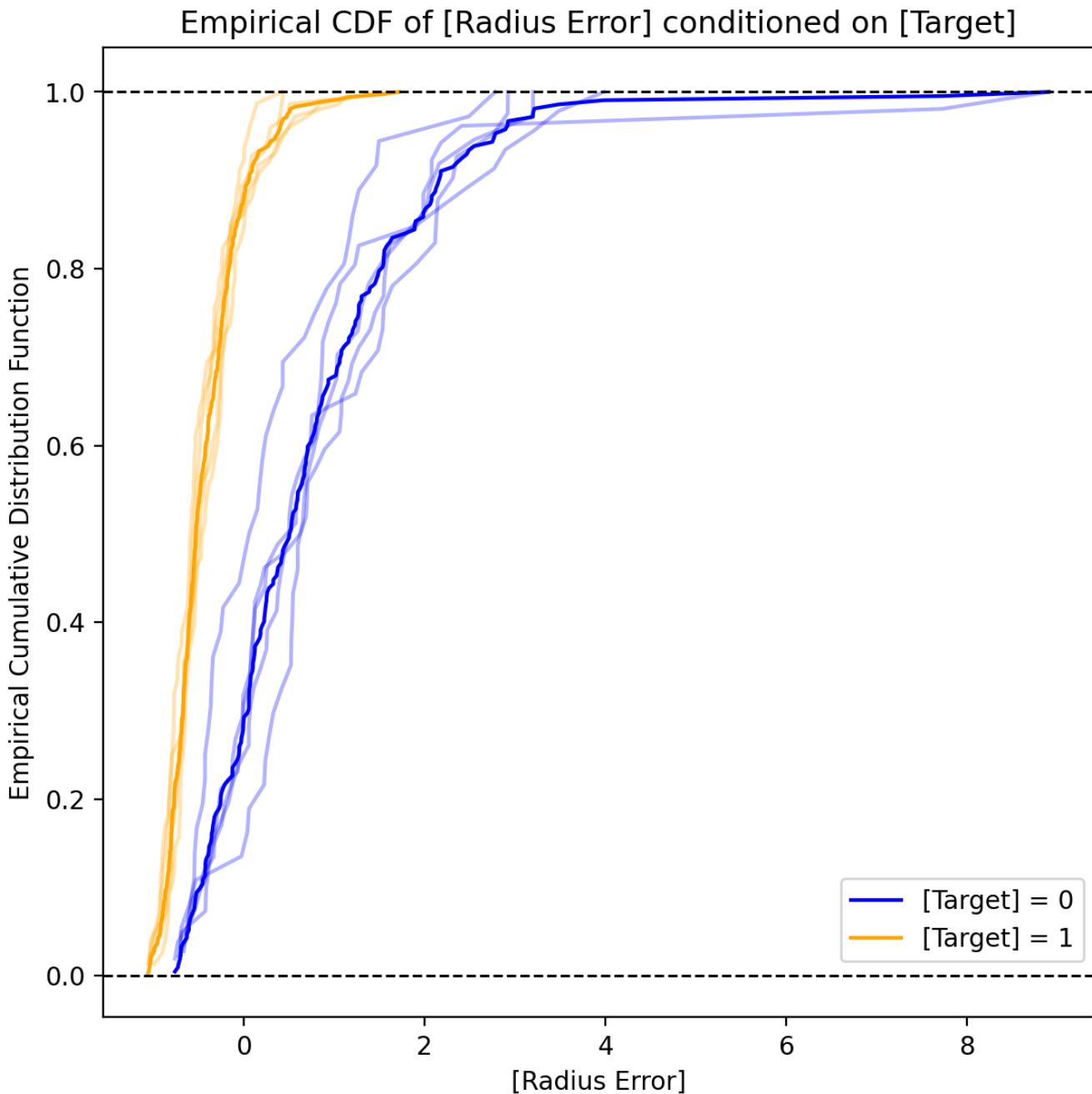
Kernel Density Plot of [Radius Error] by [Target]
Distributions by level are significantly different at the 95% level.



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

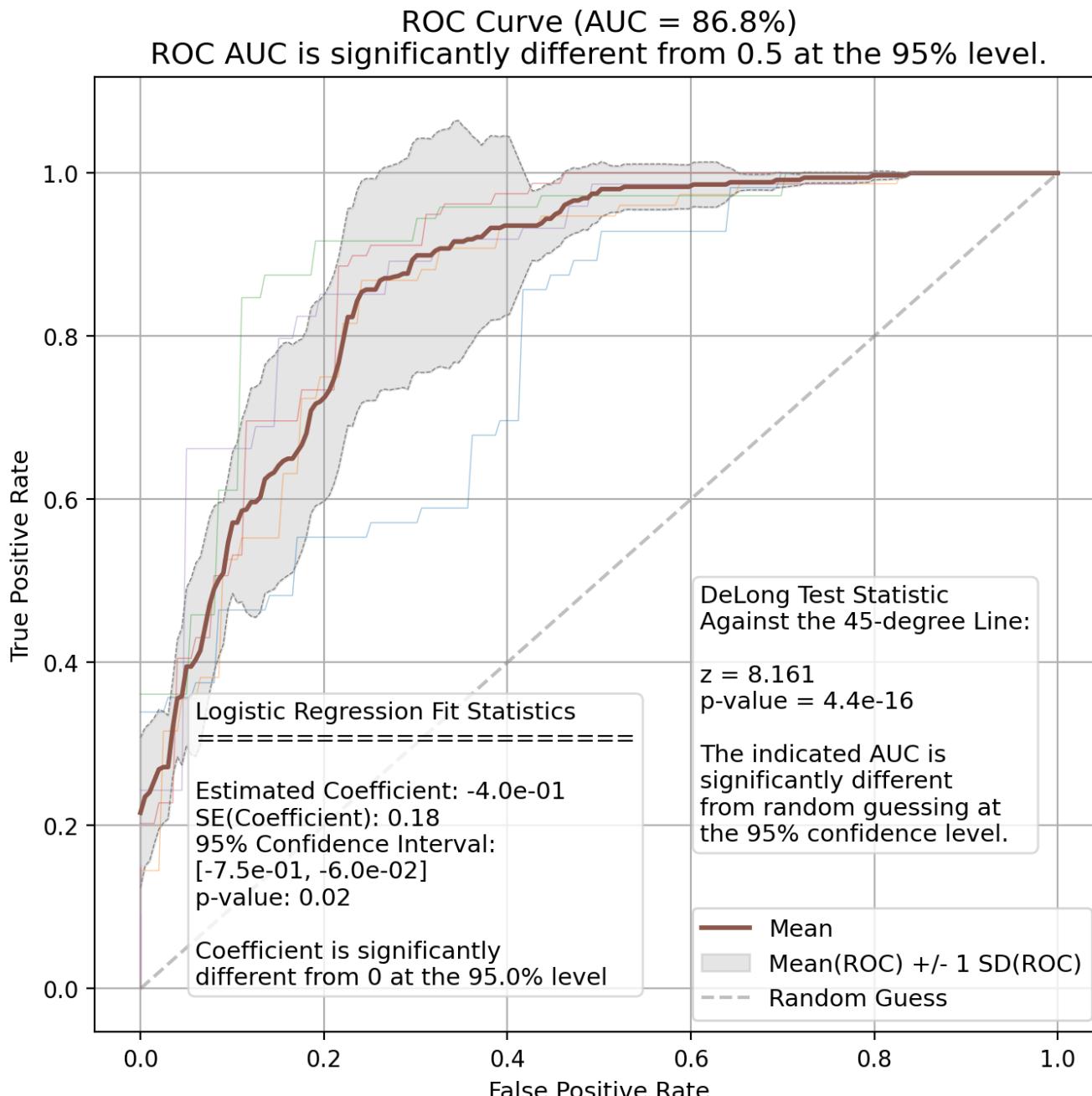
Radius Error - Empirical CDF Plot



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

Radius Error - ROC Curve



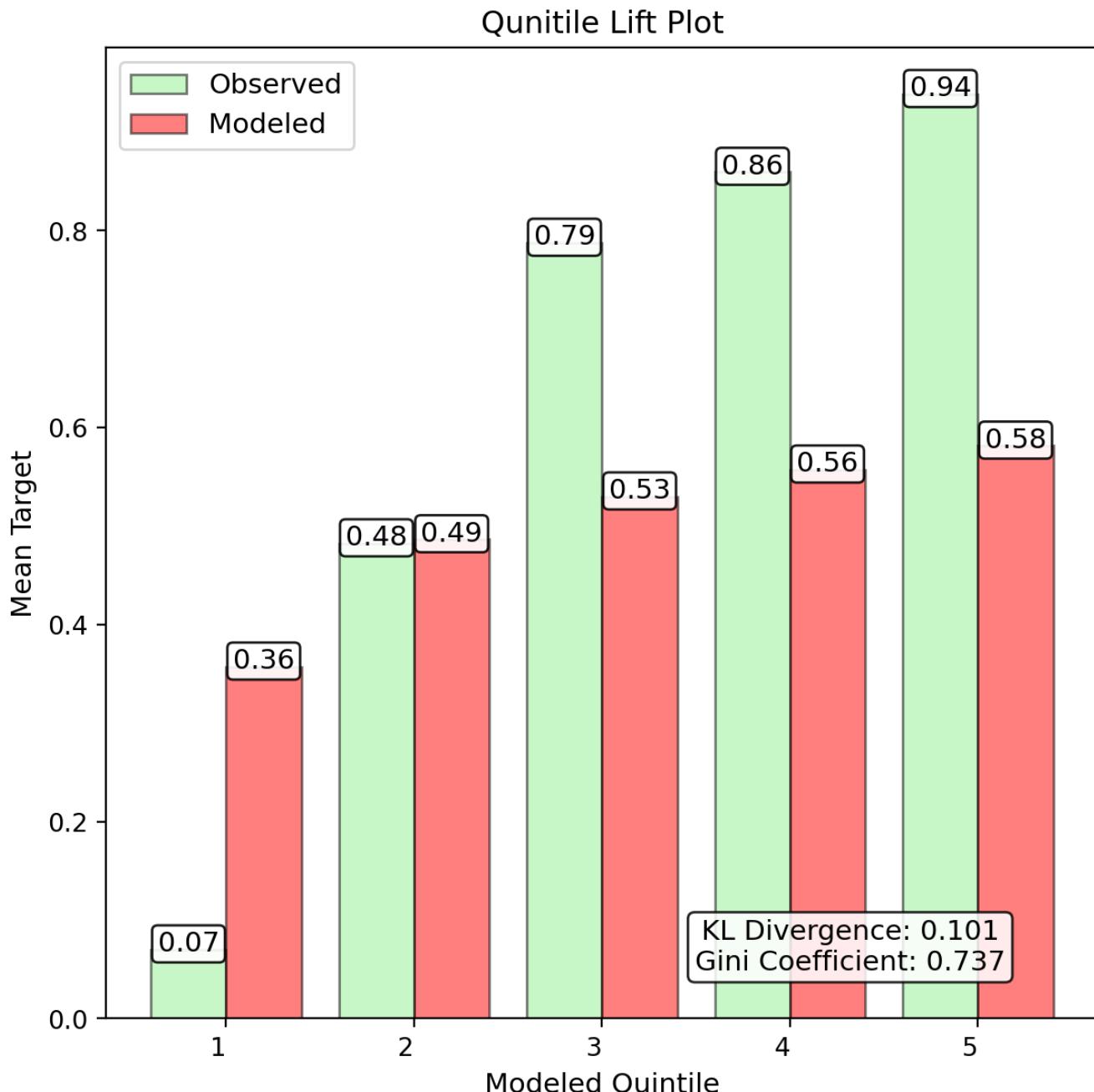
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Radius 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

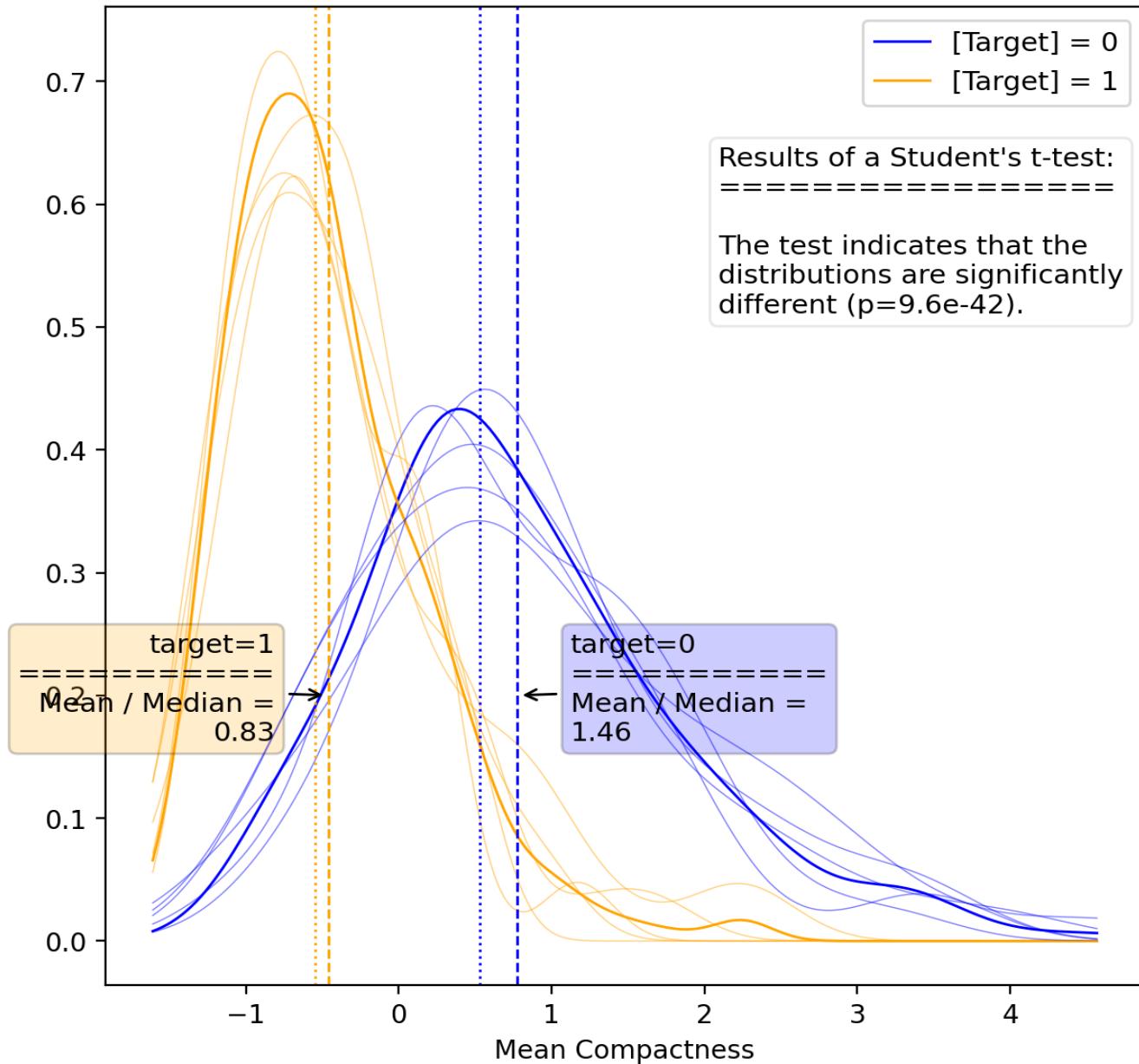
Mean Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.8e+00	-2.3e+00	-1.8e+00	-1.7e+00	-2.0e+00	-5.5e-01	2.1e-01
Fitted p-Value	9.5e-26	1.8e-26	6.4e-25	1.4e-23	3.2e-25	4.4e-01	6.0e-24
Fitted Std. Err.	0.173	0.214	0.178	0.174	0.188	0.718	0.017
Conf. Int. Lower	-2.2e+00	-2.7e+00	-2.2e+00	-2.1e+00	-2.3e+00	-2.0e+00	2.4e-01
Conf. Int. Upper	-1.5e+00	-1.9e+00	-1.5e+00	-1.4e+00	-1.6e+00	8.5e-01	1.8e-01
Train Accuracy	78.2%	80.8%	78.7%	78.8%	78.6%	79.1%	1.0%
Val Accuracy	83.7%	73.0%	78.0%	80.2%	81.7%	79.1%	4.1%
Train AUC	78.0%	80.9%	78.7%	79.0%	78.6%	79.1%	1.1%
Val AUC	85.1%	73.6%	76.1%	79.3%	82.5%	79.1%	4.7%
Train F1	82.0%	84.0%	82.1%	82.4%	82.1%	82.6%	0.9%
Test F1	85.4%	76.6%	83.1%	83.5%	84.9%	82.6%	3.6%
Train Precision	85.6%	87.9%	85.8%	87.1%	85.8%	86.5%	1.0%
Val Precision	93.6%	83.1%	84.3%	83.5%	90.8%	86.5%	4.8%
Train Recall	78.7%	80.4%	78.6%	78.1%	78.8%	79.0%	0.9%
Val Recall	78.6%	71.1%	81.9%	83.5%	79.7%	79.0%	4.8%
Train MCC	54.8%	60.4%	56.3%	56.4%	56.0%	56.9%	2.1%
Val MCC	68.6%	45.8%	51.6%	58.5%	62.9%	56.9%	9.0%
Train Log-Loss	7.86	6.93	7.68	7.65	7.70	7.54	0.36
Val Log-Loss	5.88	9.75	7.94	7.15	6.58	7.54	1.49

Univariate Report

Mean Compactness - Kernel Density Plot

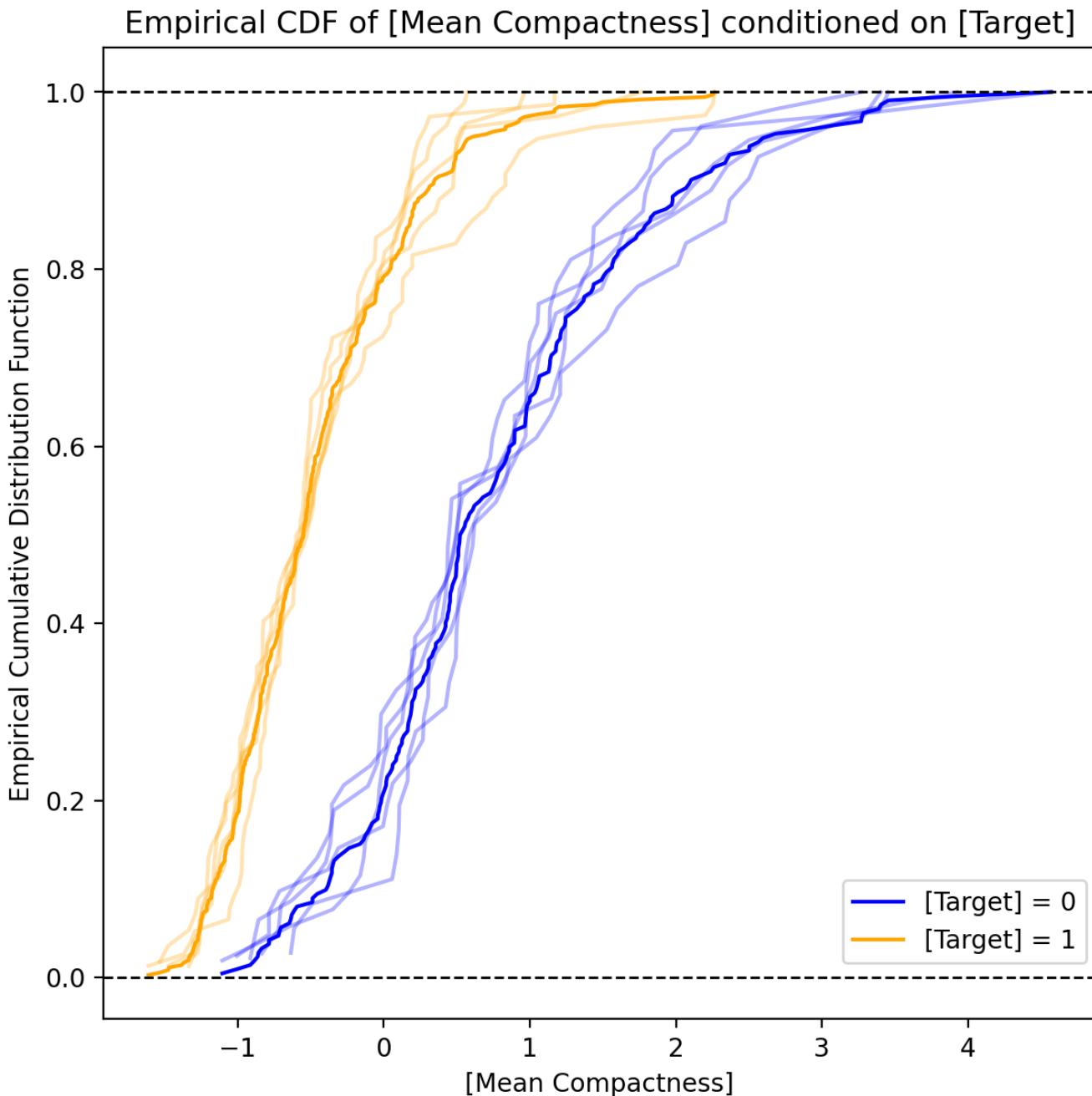
Kernel Density Plot of [Mean Compactness] by [Target]
Distributions by level are significantly different at the 95% level.



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

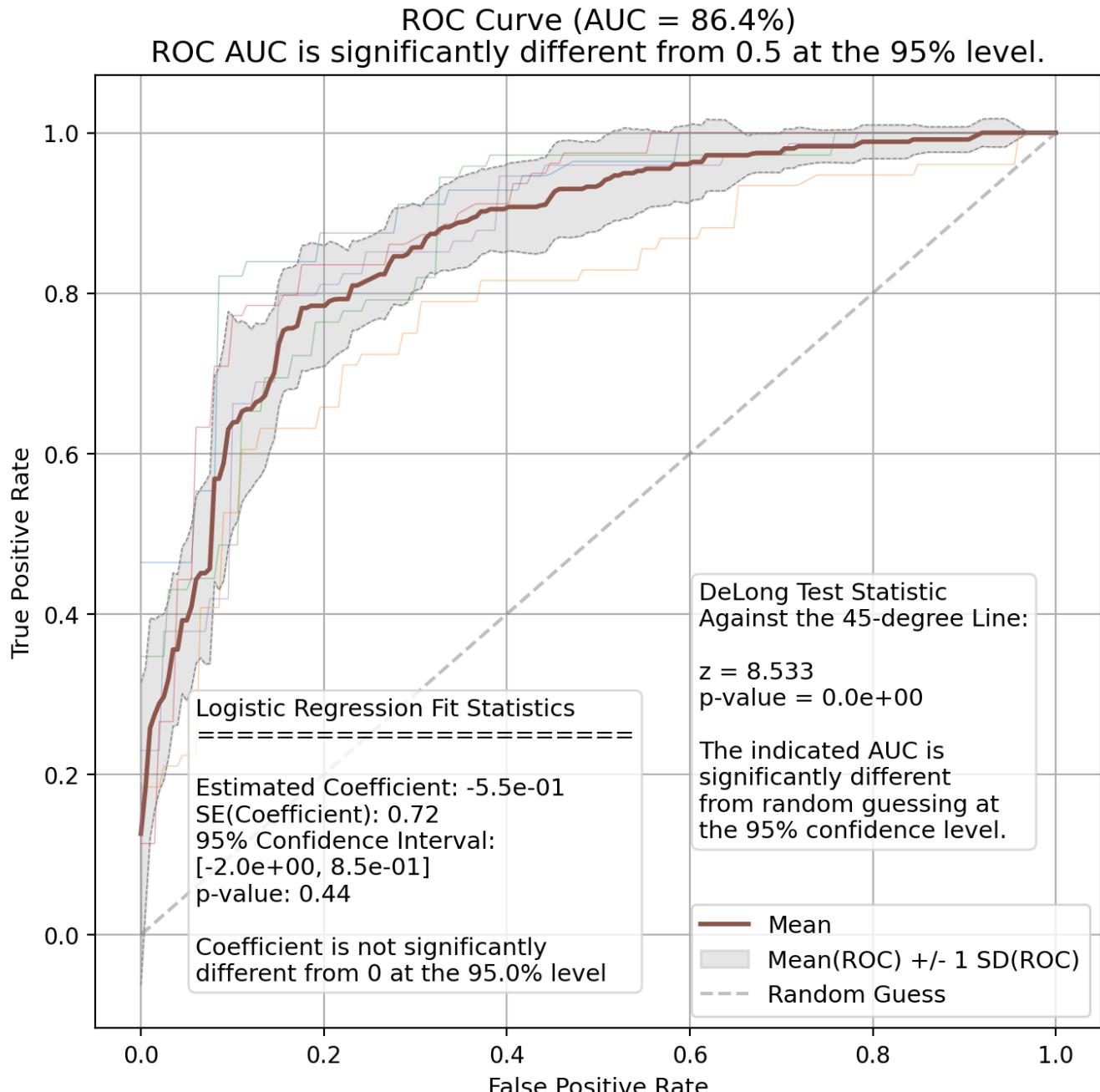
Mean Compactness - Empirical CDF Plot



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 Compactness - ROC Curve



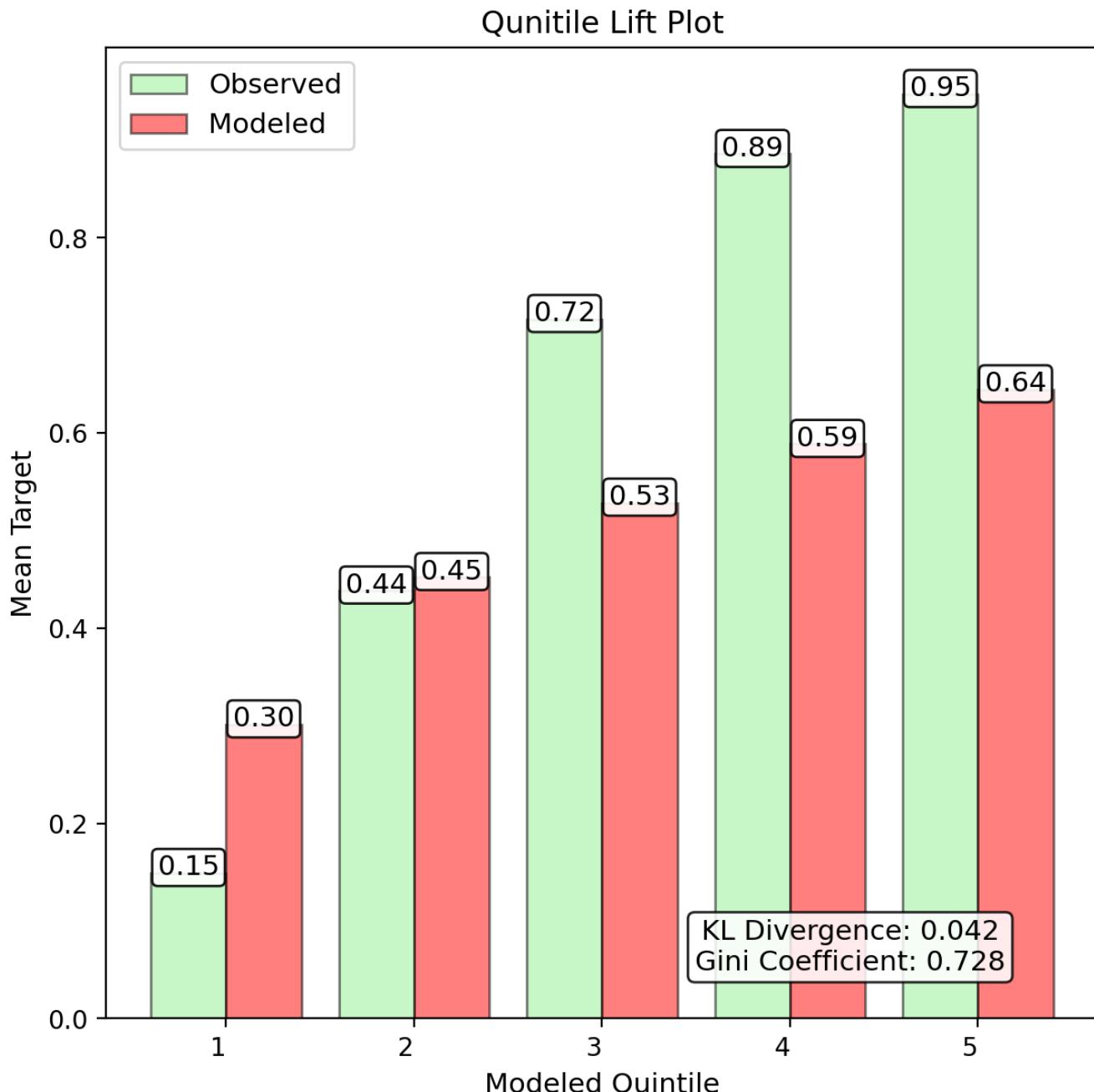
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean Compactness - 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

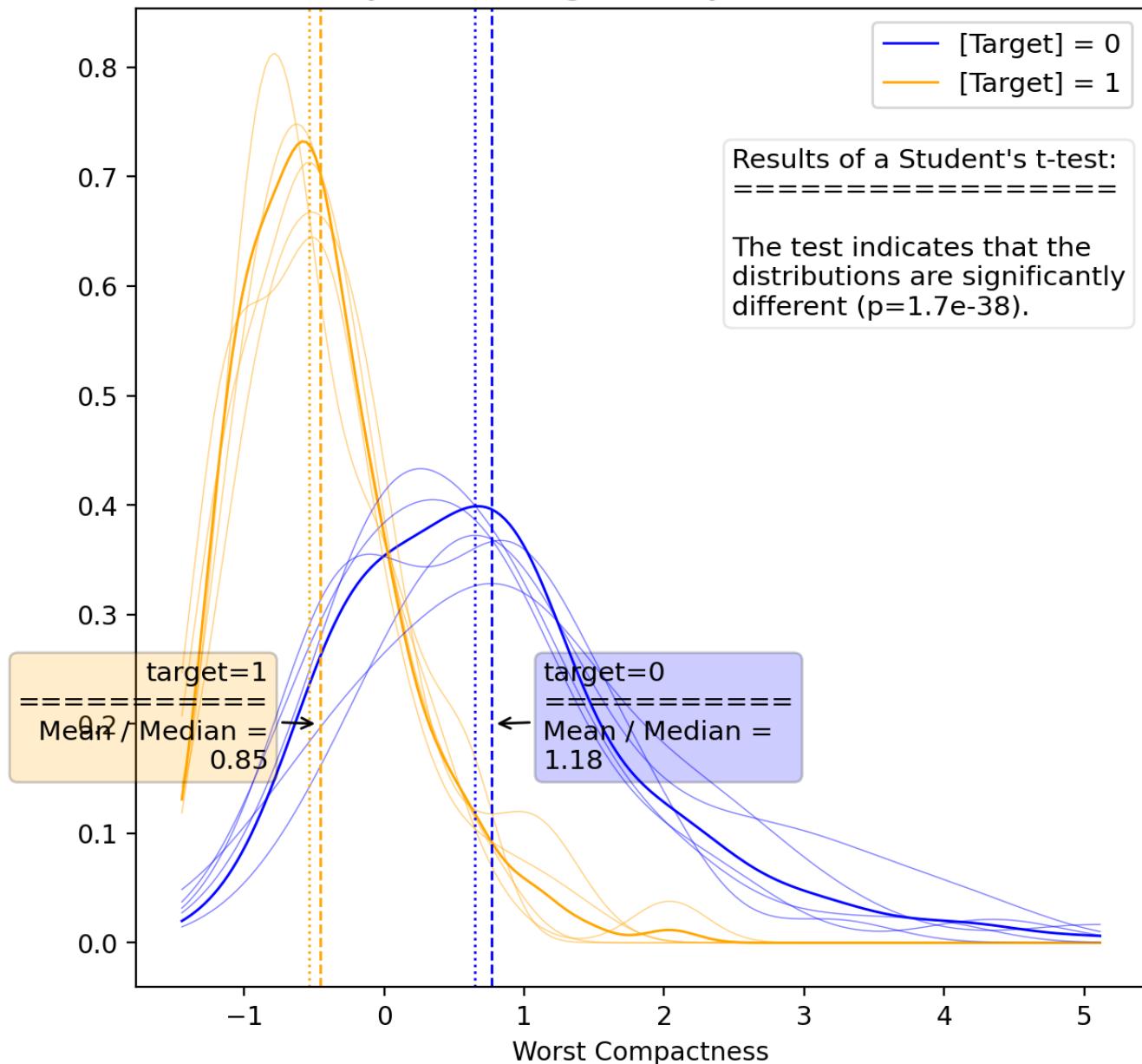
Worst Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.9e+00	-2.2e+00	-2.0e+00	-2.1e+00	-2.1e+00	-5.5e-01	1.1e-01
Fitted p-Value	7.7e-26	3.5e-26	1.8e-25	3.4e-25	2.7e-25	5.1e-02	1.3e-25
Fitted Std. Err.	0.180	0.208	0.189	0.198	0.198	0.283	0.010
Conf. Int. Lower	-2.2e+00	-2.6e+00	-2.3e+00	-2.4e+00	-2.4e+00	-1.1e+00	1.3e-01
Conf. Int. Upper	-1.5e+00	-1.8e+00	-1.6e+00	-1.7e+00	-1.7e+00	2.3e-03	9.3e-02
Train Accuracy	76.1%	79.9%	80.0%	79.5%	78.4%	78.7%	1.6%
Val Accuracy	85.9%	74.6%	77.1%	78.6%	80.9%	78.7%	4.3%
Train AUC	75.2%	78.9%	78.8%	78.4%	77.5%	77.7%	1.6%
Val AUC	86.4%	73.2%	74.8%	77.3%	79.7%	77.7%	5.2%
Train F1	80.6%	83.8%	83.9%	83.6%	82.4%	82.8%	1.4%
Test F1	87.9%	79.5%	82.5%	82.5%	84.9%	82.8%	3.1%
Train Precision	82.6%	85.0%	83.9%	84.8%	83.6%	83.9%	1.0%
Val Precision	92.2%	80.0%	83.1%	81.5%	86.1%	83.9%	4.8%
Train Recall	78.7%	82.6%	83.9%	82.4%	81.3%	81.8%	1.9%
Val Recall	83.9%	78.9%	81.9%	83.5%	83.8%	81.8%	2.1%
Train MCC	49.6%	57.3%	57.6%	56.2%	54.5%	54.9%	3.3%
Val MCC	71.5%	46.1%	49.2%	55.1%	58.8%	54.9%	9.9%
Train Log-Loss	8.61	7.26	7.21	7.41	7.78	7.66	0.58
Val Log-Loss	5.09	9.16	8.27	7.70	6.90	7.66	1.54

Univariate Report

Worst Compactness - Kernel Density Plot

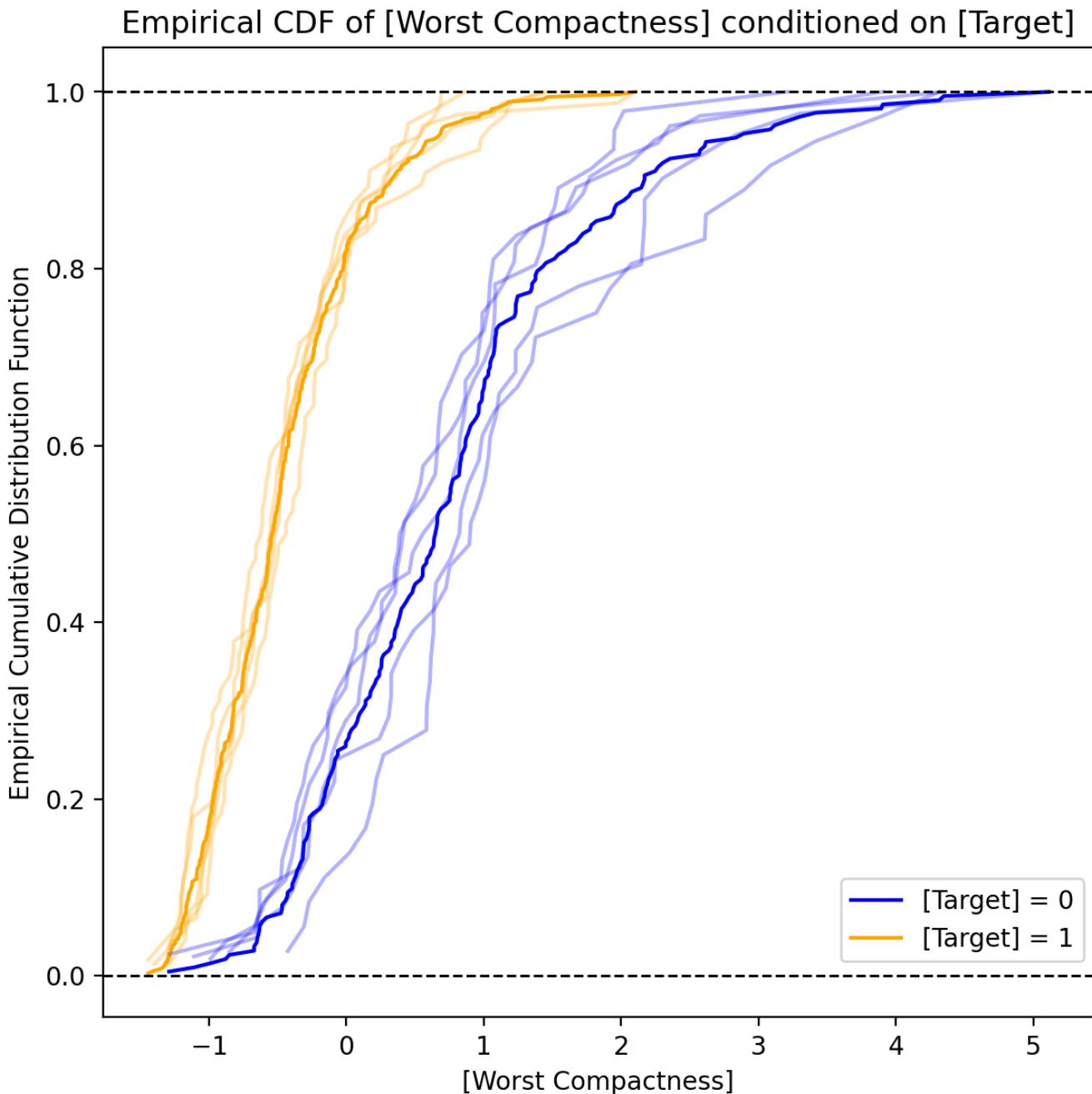
Kernel Density Plot of [Worst Compactness] by [Target]
Distributions by level are significantly different at the 95% level.



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

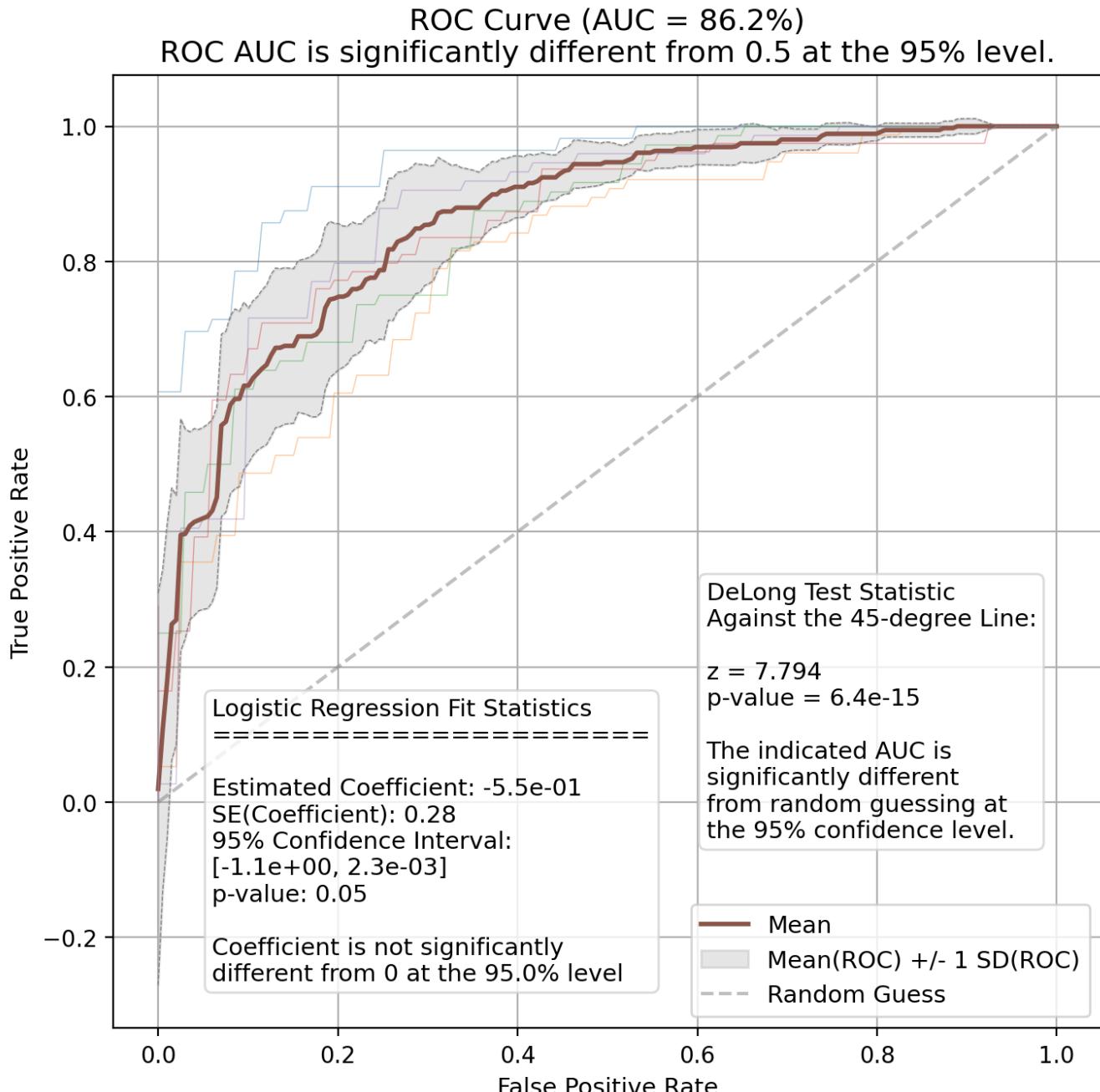
Worst Compactness - Empirical CDF Plot



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 Compactness - ROC Curve



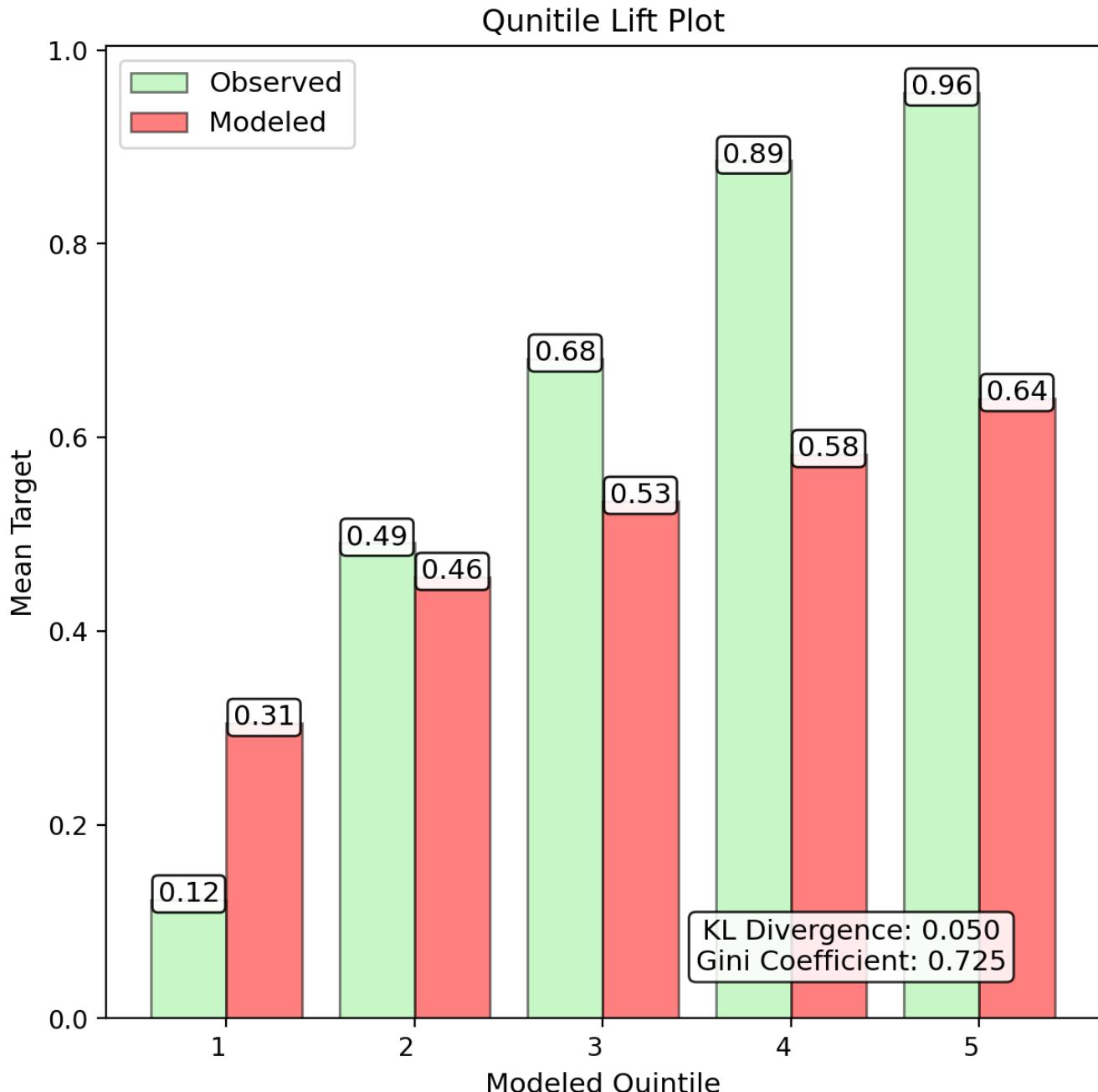
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a \pm two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Compactness - 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

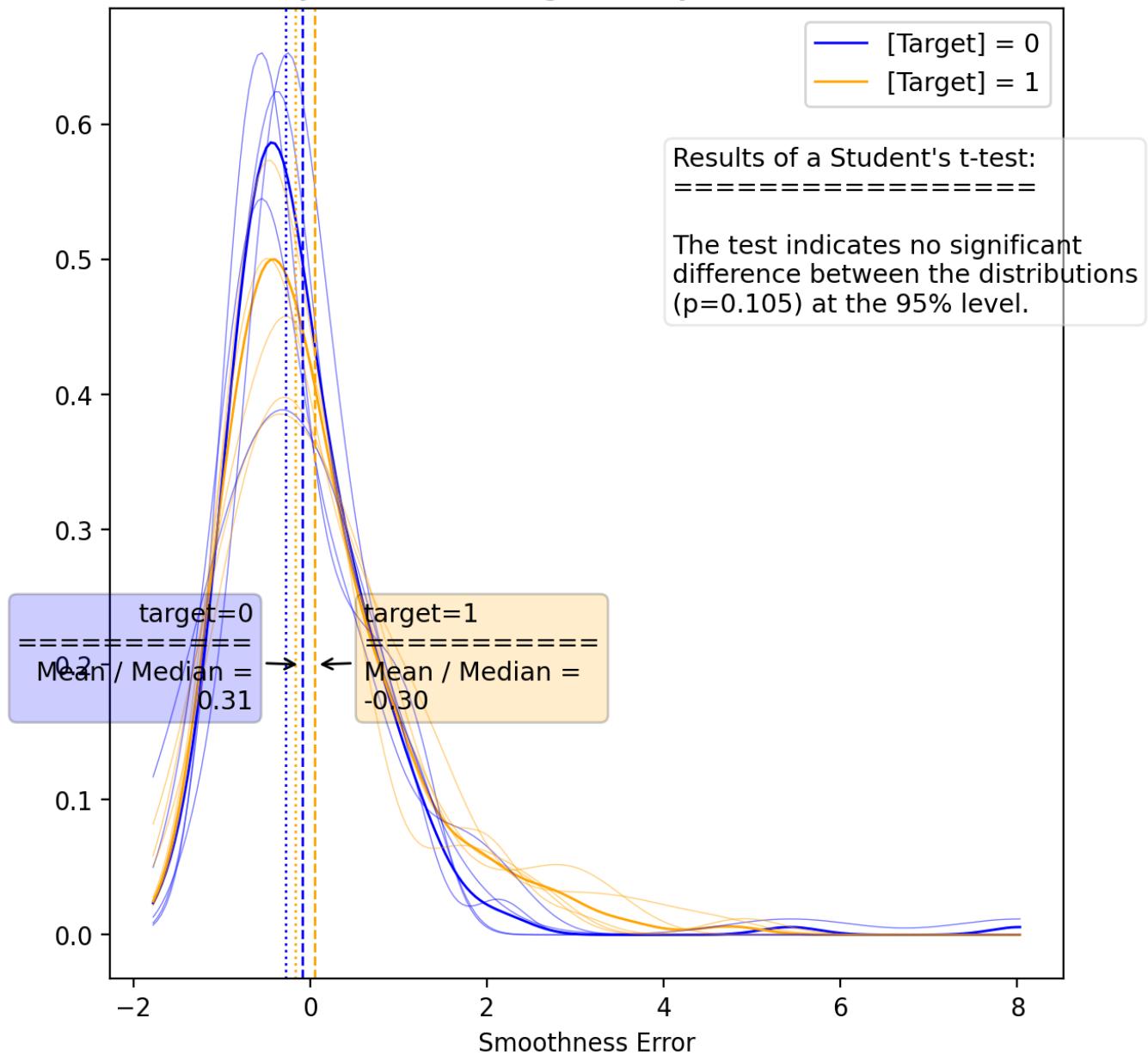
Smoothness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	0.105	0.201	0.136	0.165	0.051	69.880	0.057
Fitted p-Value	0.256	0.058	0.148	0.082	0.594	0.000	0.219
Fitted Std. Err.	9.3e-02	1.1e-01	9.4e-02	9.5e-02	9.5e-02	1.2e+01	5.5e-03
Conf. Int. Lower	-7.6e-02	-6.9e-03	-4.8e-02	-2.1e-02	-1.4e-01	4.7e+01	5.1e-02
Conf. Int. Upper	0.287	0.409	0.319	0.350	0.238	92.918	0.065
Train Accuracy	50.1%	51.9%	51.7%	52.3%	49.3%	51.1%	1.3%
Val Accuracy	56.5%	48.4%	48.6%	45.8%	56.5%	51.1%	5.0%
Train AUC	53.4%	55.0%	54.5%	55.5%	52.2%	54.3%	1.3%
Val AUC	58.8%	51.3%	53.2%	48.8%	59.1%	54.3%	4.6%
Train F1	50.8%	53.0%	52.4%	53.7%	50.0%	51.9%	1.5%
Test F1	57.4%	48.8%	50.0%	43.2%	59.7%	51.9%	6.7%
Train Precision	67.2%	68.8%	67.4%	69.9%	65.0%	67.9%	1.9%
Val Precision	71.1%	63.8%	70.0%	58.7%	74.0%	67.9%	6.2%
Train Recall	40.9%	43.1%	42.8%	43.5%	40.6%	42.0%	1.3%
Val Recall	48.2%	39.5%	38.9%	34.2%	50.0%	42.0%	6.7%
Train MCC	6.7%	9.8%	9.0%	10.9%	4.4%	8.5%	2.6%
Val MCC	17.5%	2.5%	6.3%	-2.4%	17.7%	8.5%	9.0%
Train Log-Loss	17.98	17.34	17.39	17.20	18.26	17.61	0.46
Val Log-Loss	15.67	18.61	18.52	19.54	15.67	17.61	1.81

Univariate Report

Smoothness Error - Kernel Density Plot

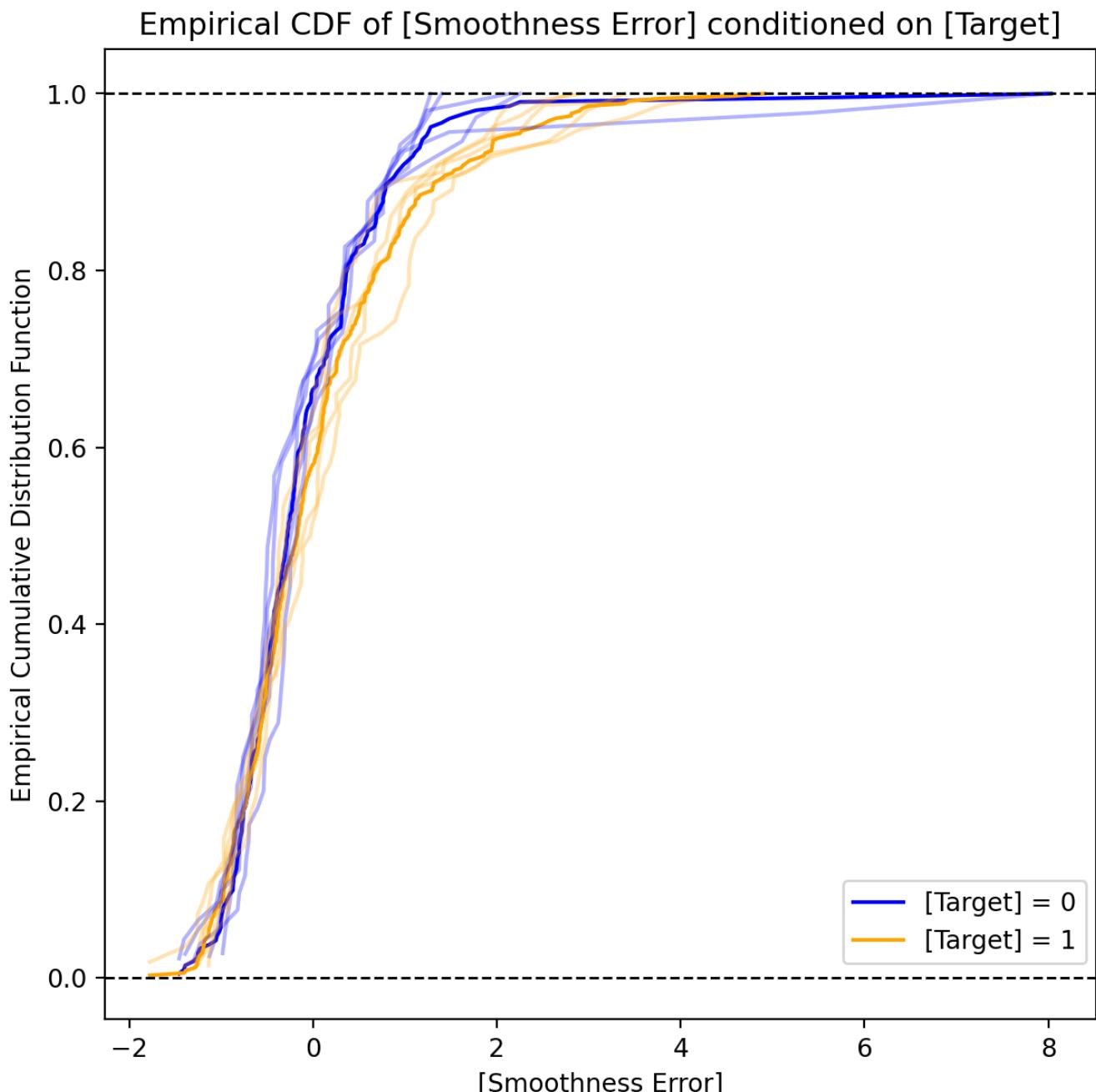
Kernel Density Plot of [Smoothness Error] by [Target]
Distributions by level are not significantly different at the 95% level.



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

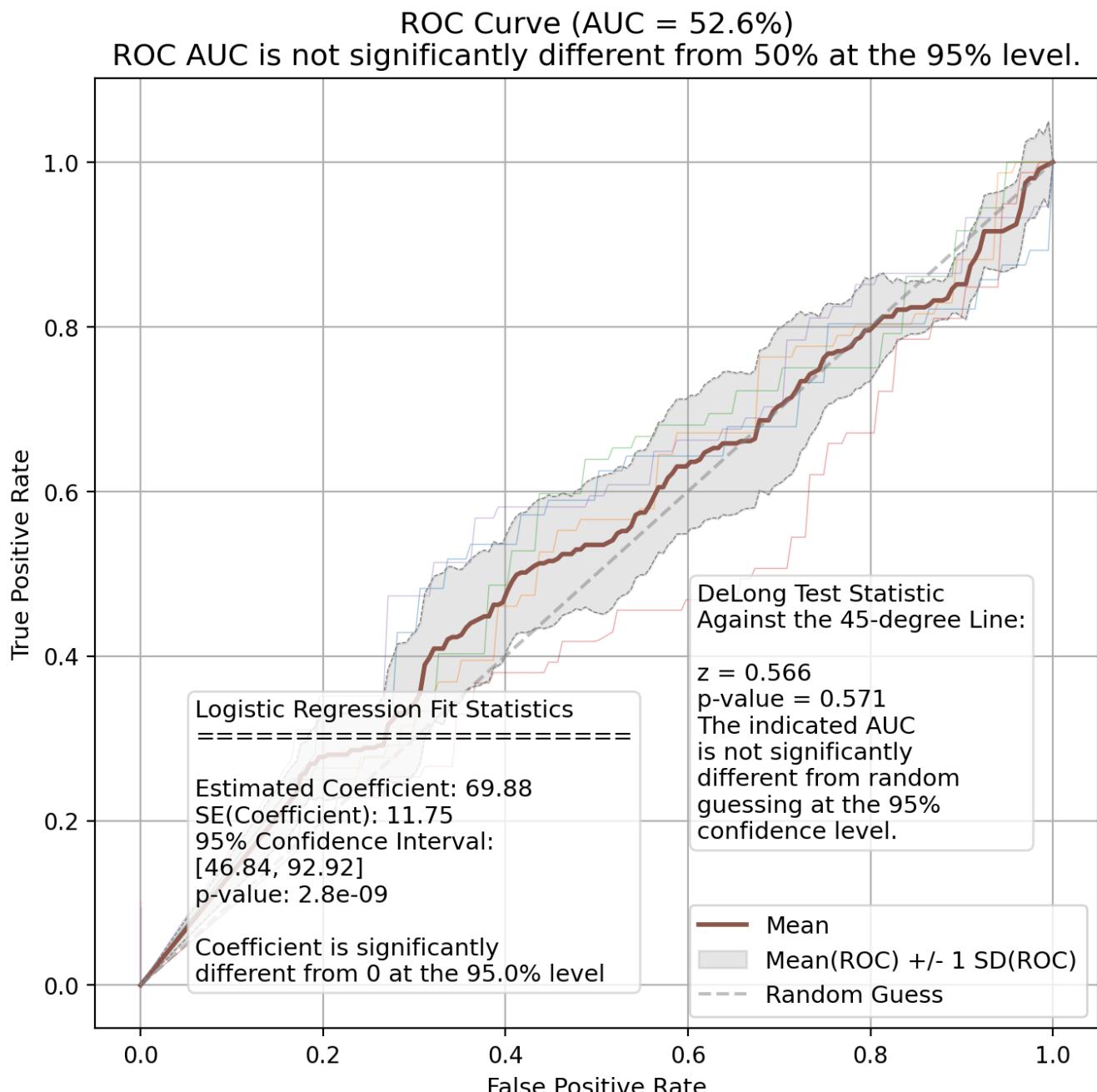
Smoothness Error - Empirical CDF Plot



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



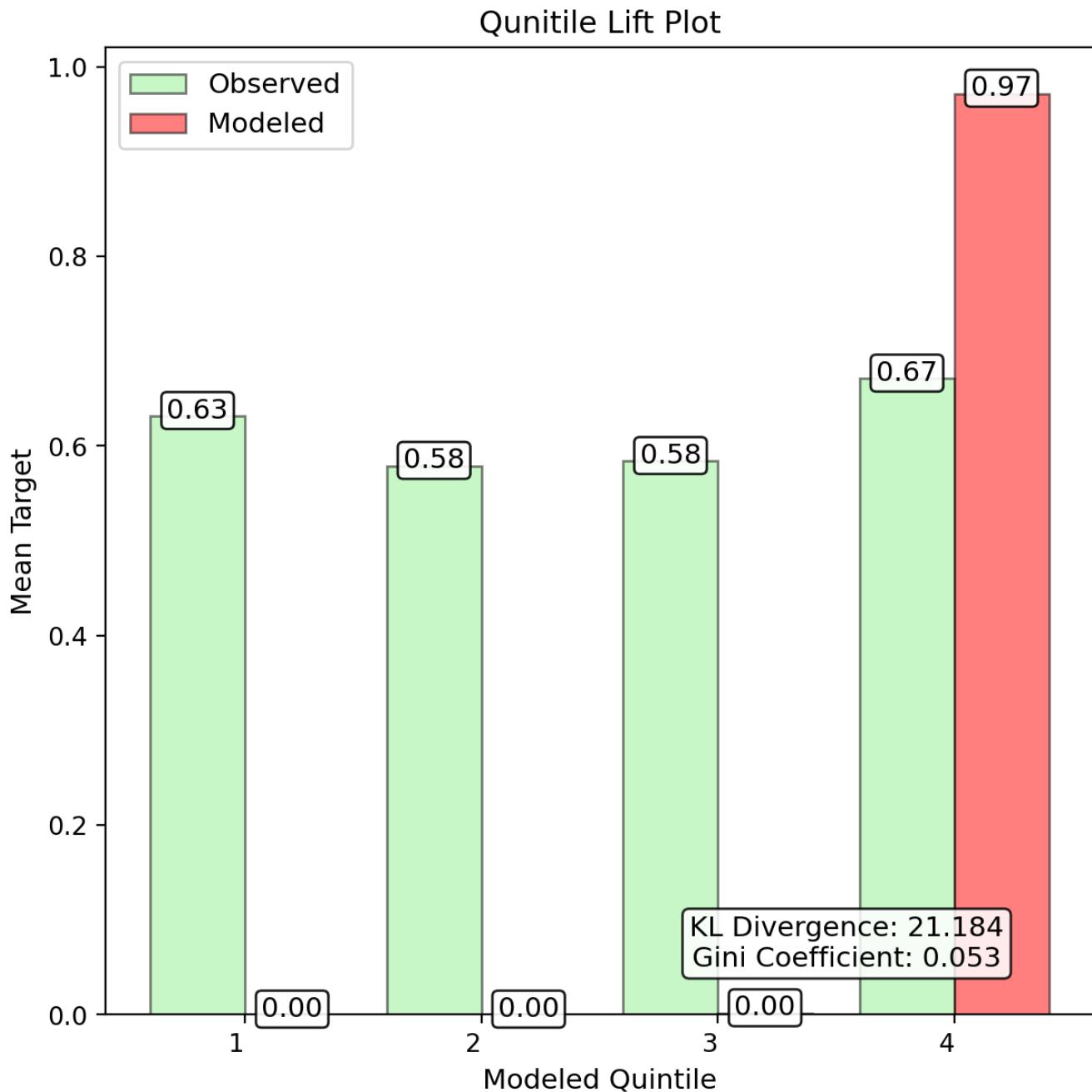
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a \pm two standard deviations from a standard normal

distribution to get the p-value.

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

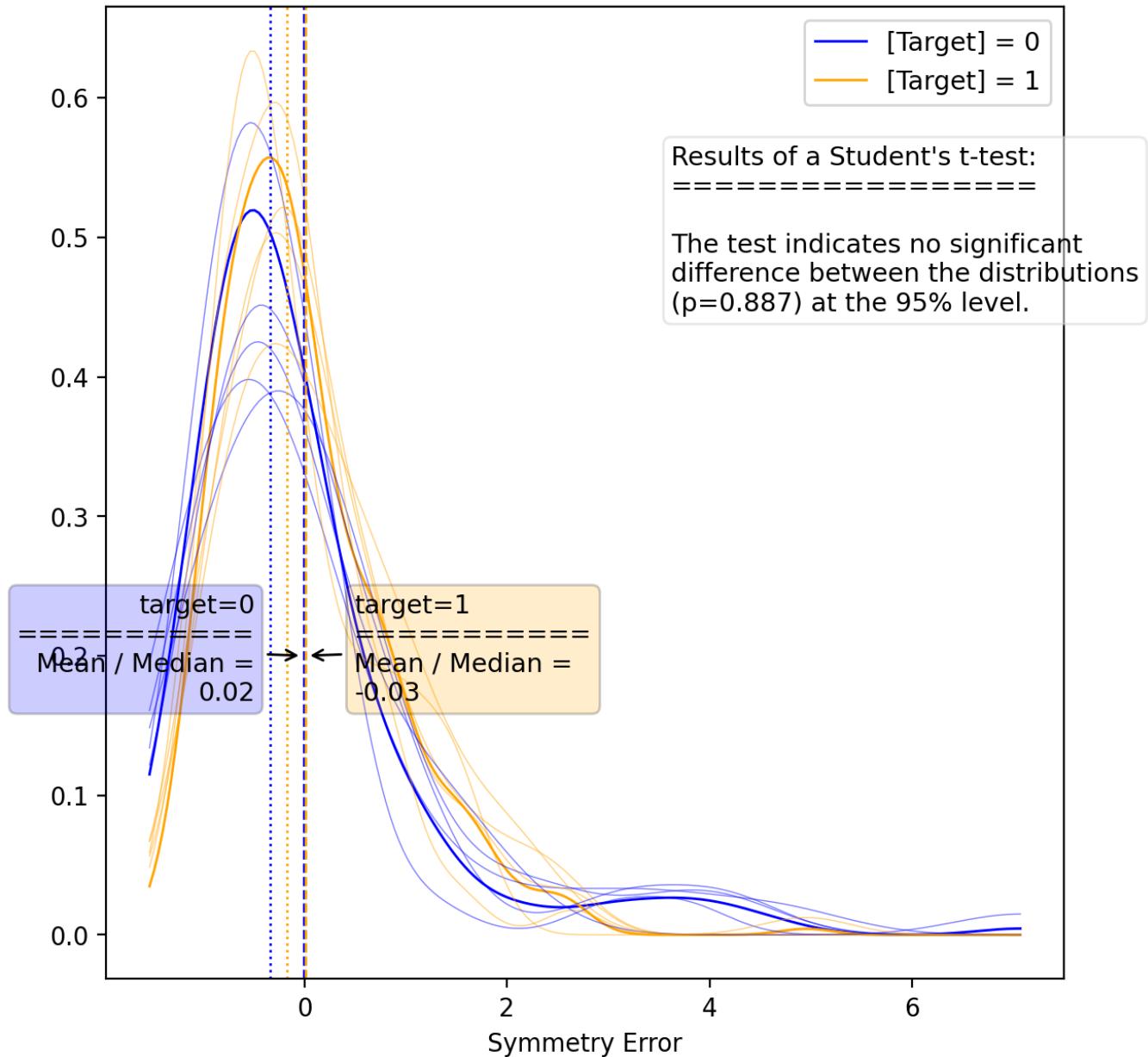
Symmetry Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.1e-03	-4.3e-02	6.4e-02	9.9e-02	-6.0e-02	2.2e+01	6.9e-02
Fitted p-Value	0.965	0.643	0.486	0.298	0.542	0.000	0.246
Fitted Std. Err.	9.3e-02	9.2e-02	9.3e-02	9.5e-02	9.9e-02	4.0e+00	2.7e-03
Conf. Int. Lower	-1.9e-01	-2.2e-01	-1.2e-01	-8.7e-02	-2.5e-01	1.4e+01	7.0e-02
Conf. Int. Upper	0.179	0.138	0.246	0.285	0.133	29.806	0.067
Train Accuracy	50.9%	50.8%	52.2%	51.4%	51.5%	50.1%	0.5%
Val Accuracy	46.7%	46.7%	41.3%	45.8%	42.6%	50.1%	2.5%
Train AUC	47.4%	47.4%	55.7%	55.3%	48.6%	53.6%	4.2%
Val AUC	44.3%	41.8%	45.0%	48.8%	39.1%	53.6%	3.6%
Train F1	61.0%	60.7%	51.5%	51.5%	60.9%	50.0%	5.1%
Test F1	55.9%	59.1%	42.9%	43.2%	53.5%	50.0%	7.5%
Train Precision	61.2%	60.9%	69.2%	70.2%	61.3%	67.3%	4.7%
Val Precision	56.4%	56.6%	60.0%	58.7%	55.9%	67.3%	1.8%
Train Recall	60.8%	60.5%	41.1%	40.6%	60.4%	39.8%	10.8%
Val Recall	55.4%	61.8%	33.3%	34.2%	51.4%	39.8%	12.8%
Train MCC	-5.1%	-5.2%	11.4%	10.6%	-2.7%	7.2%	8.5%
Val MCC	-11.3%	-17.1%	-9.7%	-2.4%	-21.3%	7.2%	7.2%
Train Log-Loss	17.68	17.74	17.24	17.53	17.47	17.99	0.20
Val Log-Loss	19.20	19.20	21.16	19.54	20.69	17.99	0.91

Univariate Report

Symmetry Error - Kernel Density Plot

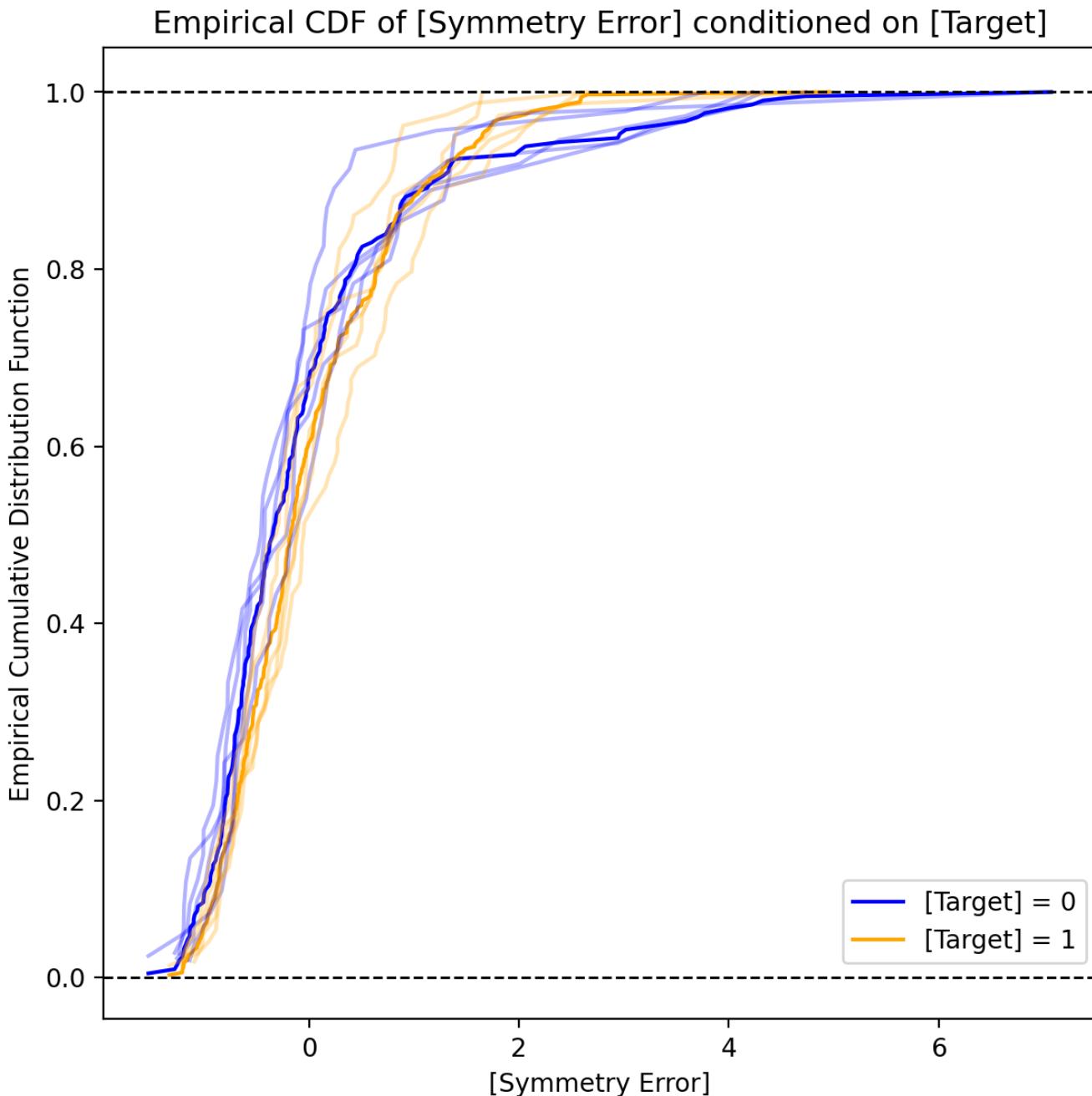
Kernel Density Plot of [Symmetry Error] by [Target]
Distributions by level are not significantly different at the 95% level.



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

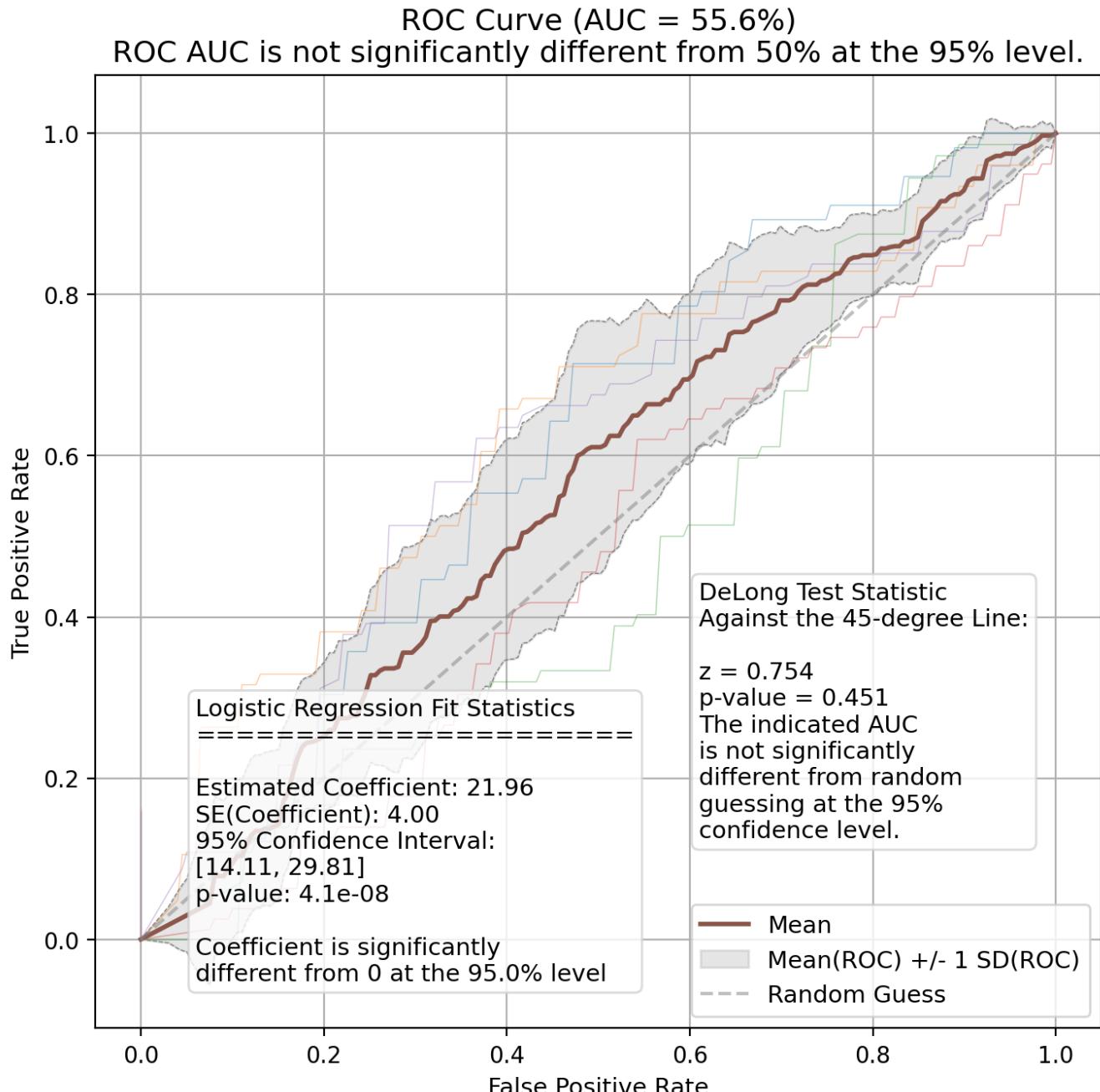
Symmetry Error - Empirical CDF Plot



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



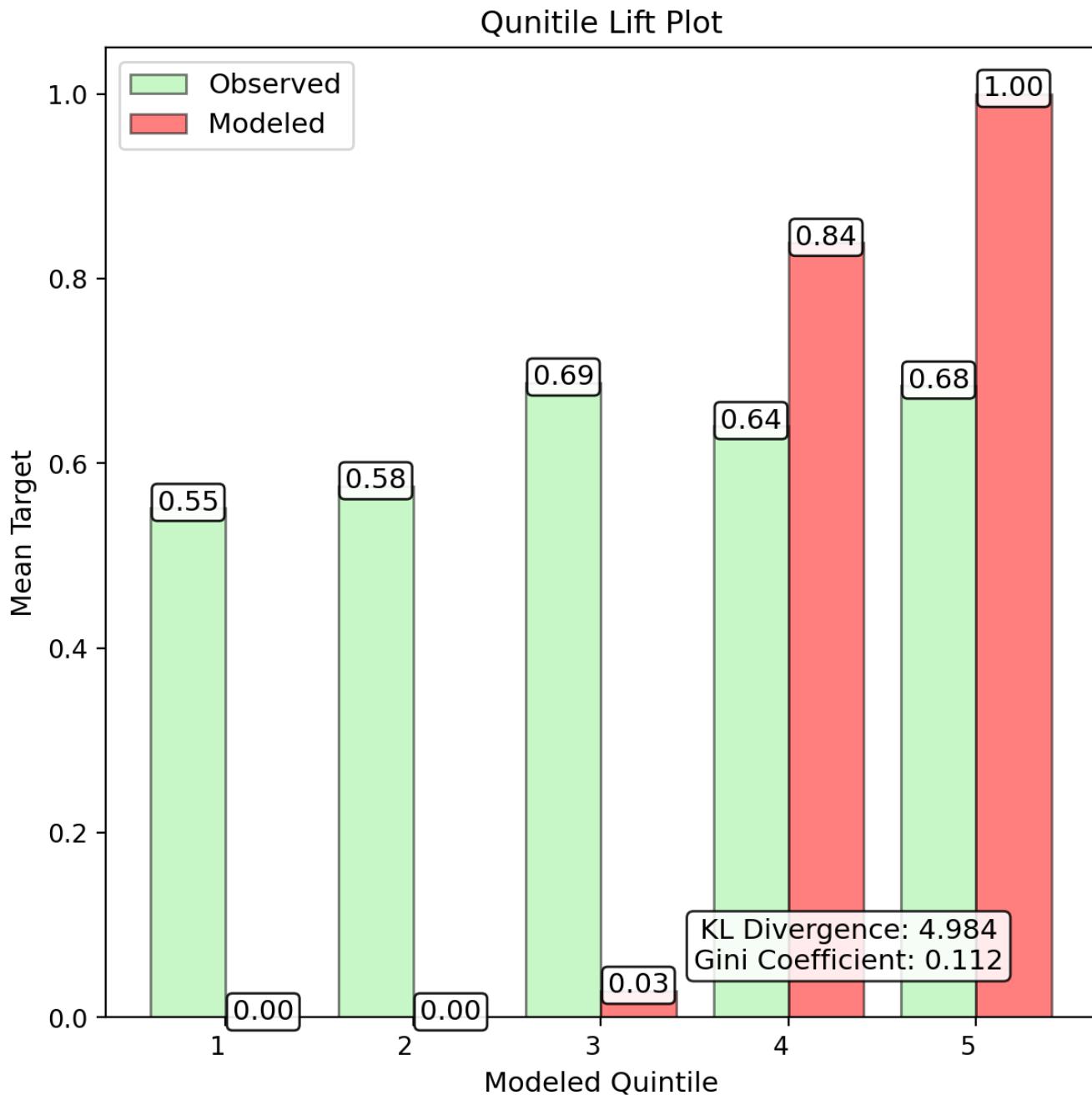
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

Texture Error - Results

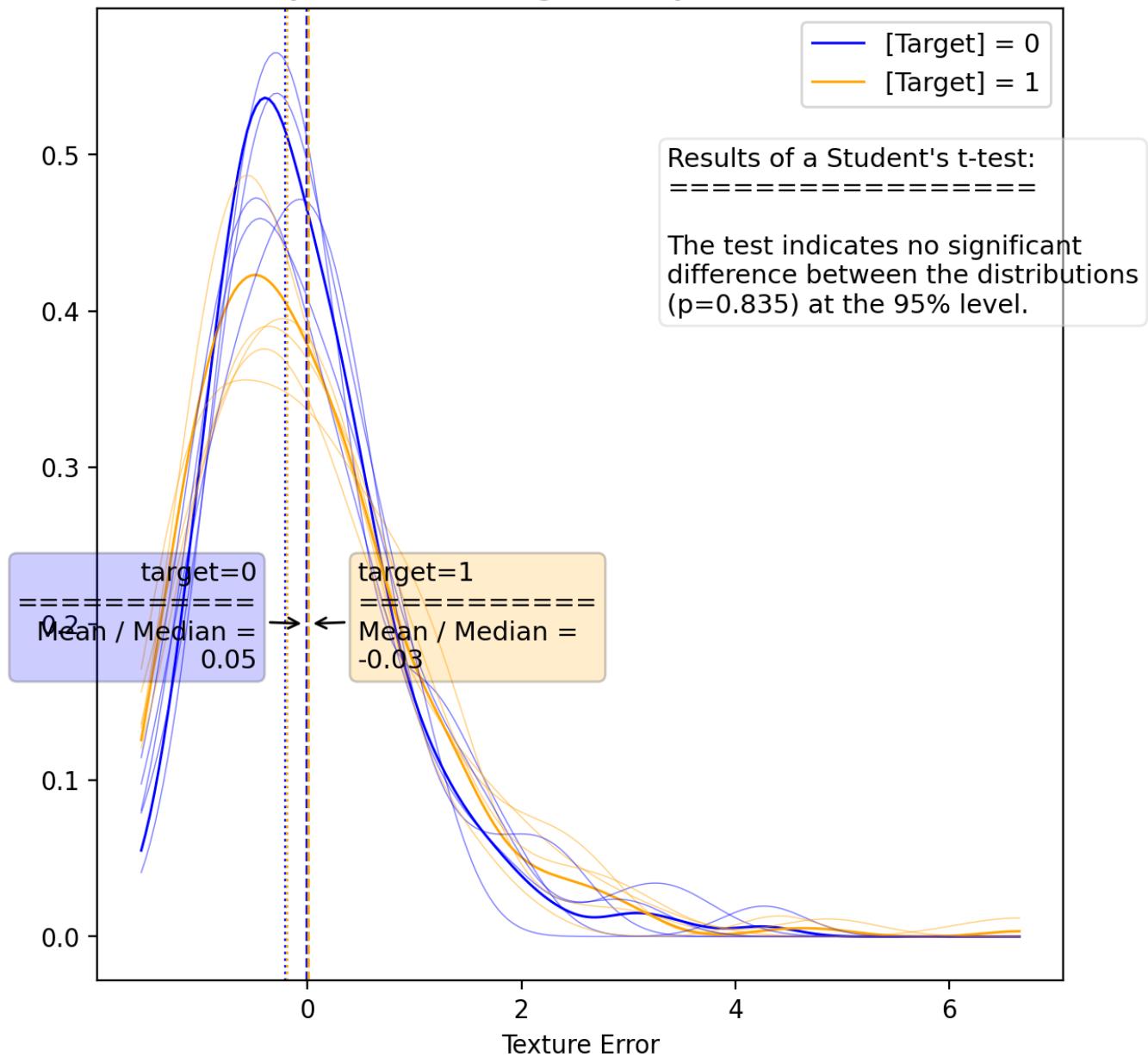
	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.6e-03	-5.3e-04	3.3e-02	6.4e-02	-1.5e-02	3.6e-01	3.3e-02
Fitted p-Value	0.961	0.996	0.721	0.495	0.871	0.000	0.205
Fitted Std. Err.	9.4e-02	9.5e-02	9.2e-02	9.3e-02	9.5e-02	6.6e-02	1.3e-03
Conf. Int. Lower	-1.9e-01	-1.9e-01	-1.5e-01	-1.2e-01	-2.0e-01	2.3e-01	3.4e-02
Conf. Int. Upper	0.180	0.185	0.213	0.247	0.171	0.489	0.031
Train Accuracy	50.5%	51.5%	48.5%	49.1%	51.5%	49.2%	1.4%
Val Accuracy	53.3%	48.4%	49.5%	45.8%	47.8%	49.2%	2.8%
Train AUC	48.5%	49.4%	50.9%	51.3%	49.6%	51.5%	1.1%
Val AUC	51.2%	46.1%	51.3%	49.1%	46.4%	51.5%	2.5%
Train F1	58.9%	59.7%	49.5%	51.8%	59.7%	51.3%	4.9%
Test F1	61.3%	57.1%	54.5%	42.3%	55.9%	51.3%	7.1%
Train Precision	61.9%	62.4%	63.0%	64.9%	62.0%	64.4%	1.2%
Val Precision	61.8%	59.2%	67.3%	59.1%	61.3%	64.4%	3.4%
Train Recall	56.1%	57.3%	40.7%	43.2%	57.6%	42.6%	8.3%
Val Recall	60.7%	55.3%	45.8%	32.9%	51.4%	42.6%	10.6%
Train MCC	-2.9%	-1.1%	1.8%	2.5%	-0.9%	2.9%	2.2%
Val MCC	2.4%	-7.6%	2.5%	-1.8%	-6.9%	2.9%	4.9%
Train Log-Loss	17.83	17.50	18.57	18.35	17.47	18.31	0.50
Val Log-Loss	16.85	18.61	18.19	19.54	18.81	18.31	0.99

Univariate Report

Texture Error - Kernel Density Plot

Kernel Density Plot of [Texture Error] by [Target]

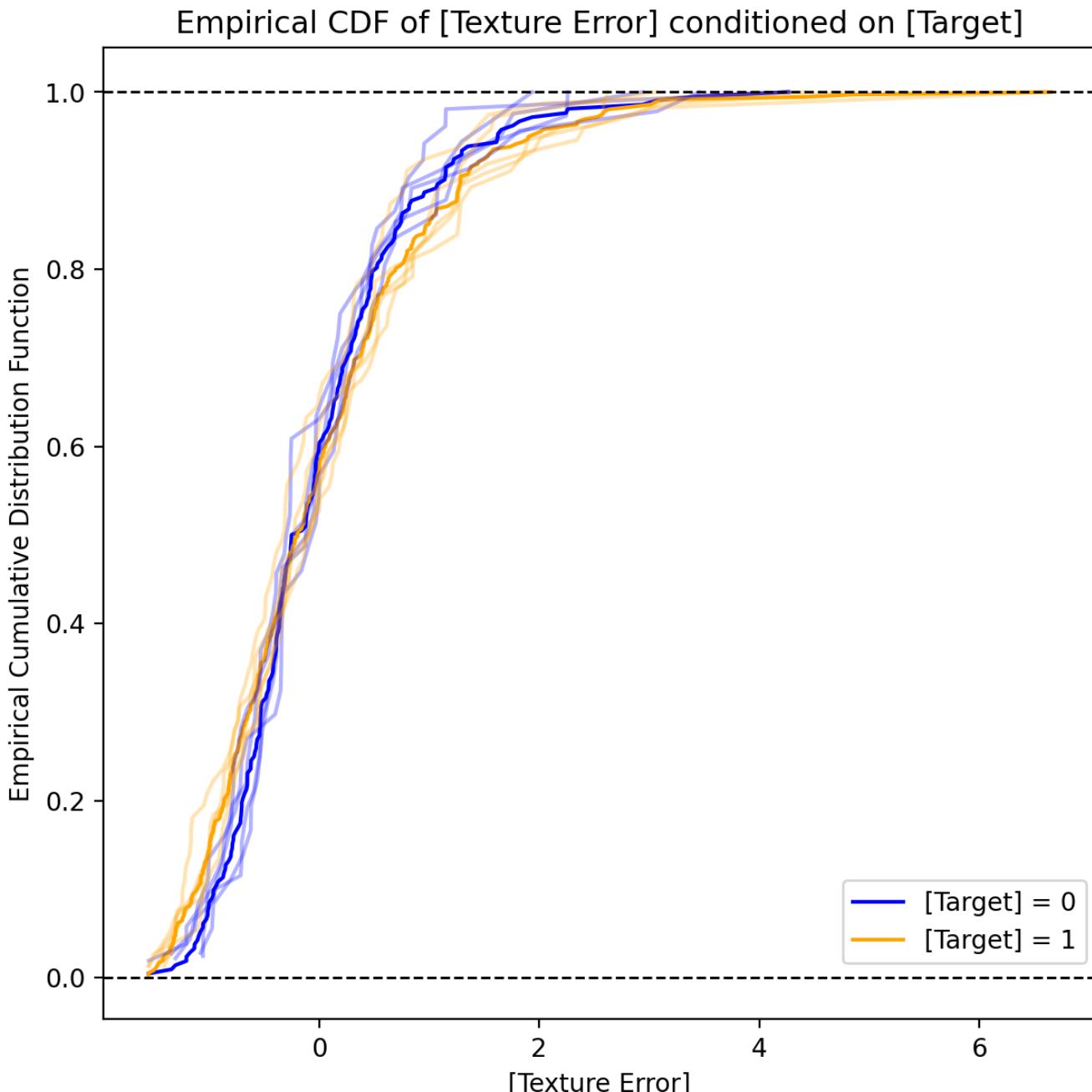
Distributions by level are not significantly different at the 95% level.



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

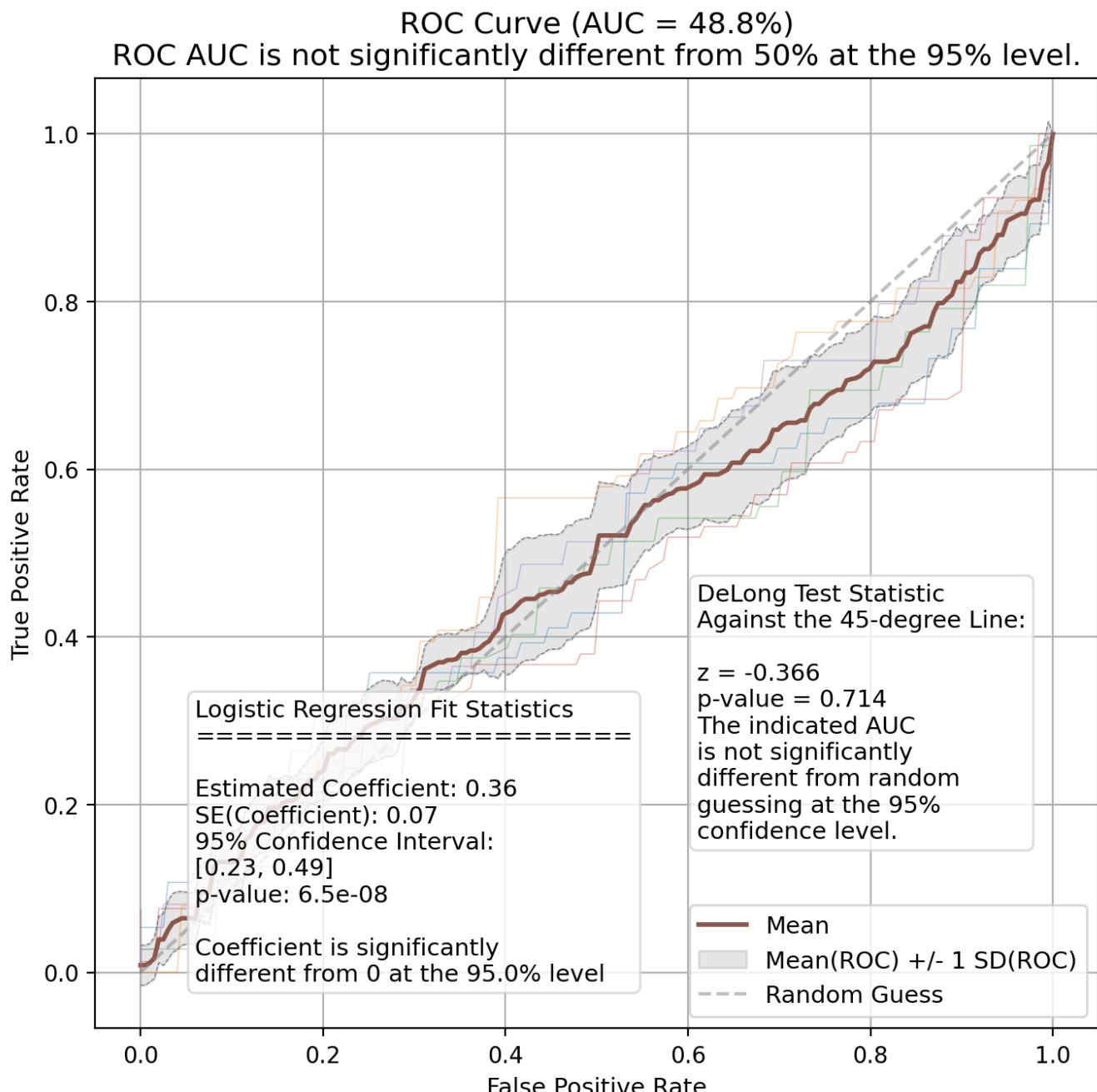
Texture Error - Empirical CDF Plot



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



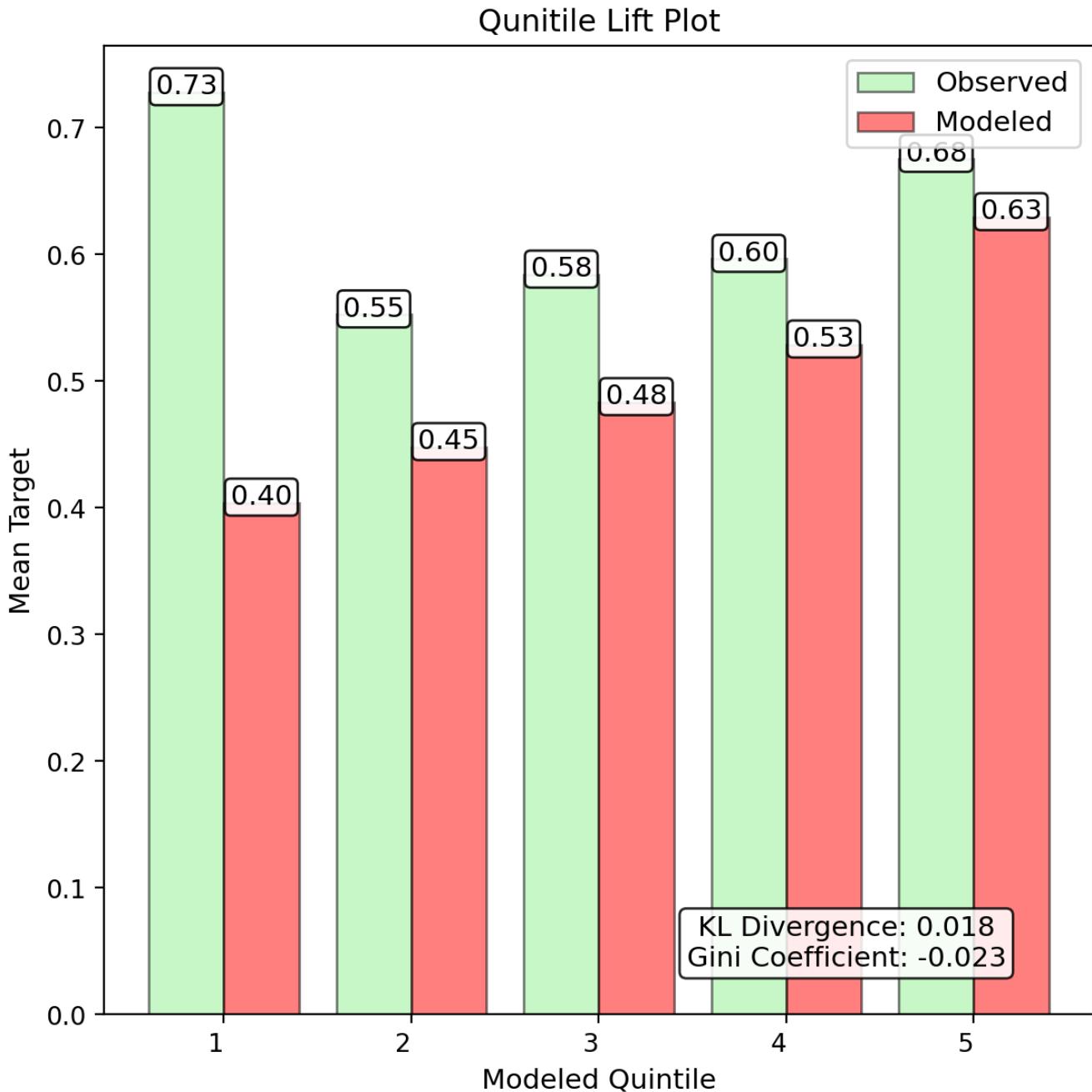
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a \pm two standard deviations from a standard normal

distribution to get the p-value.

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

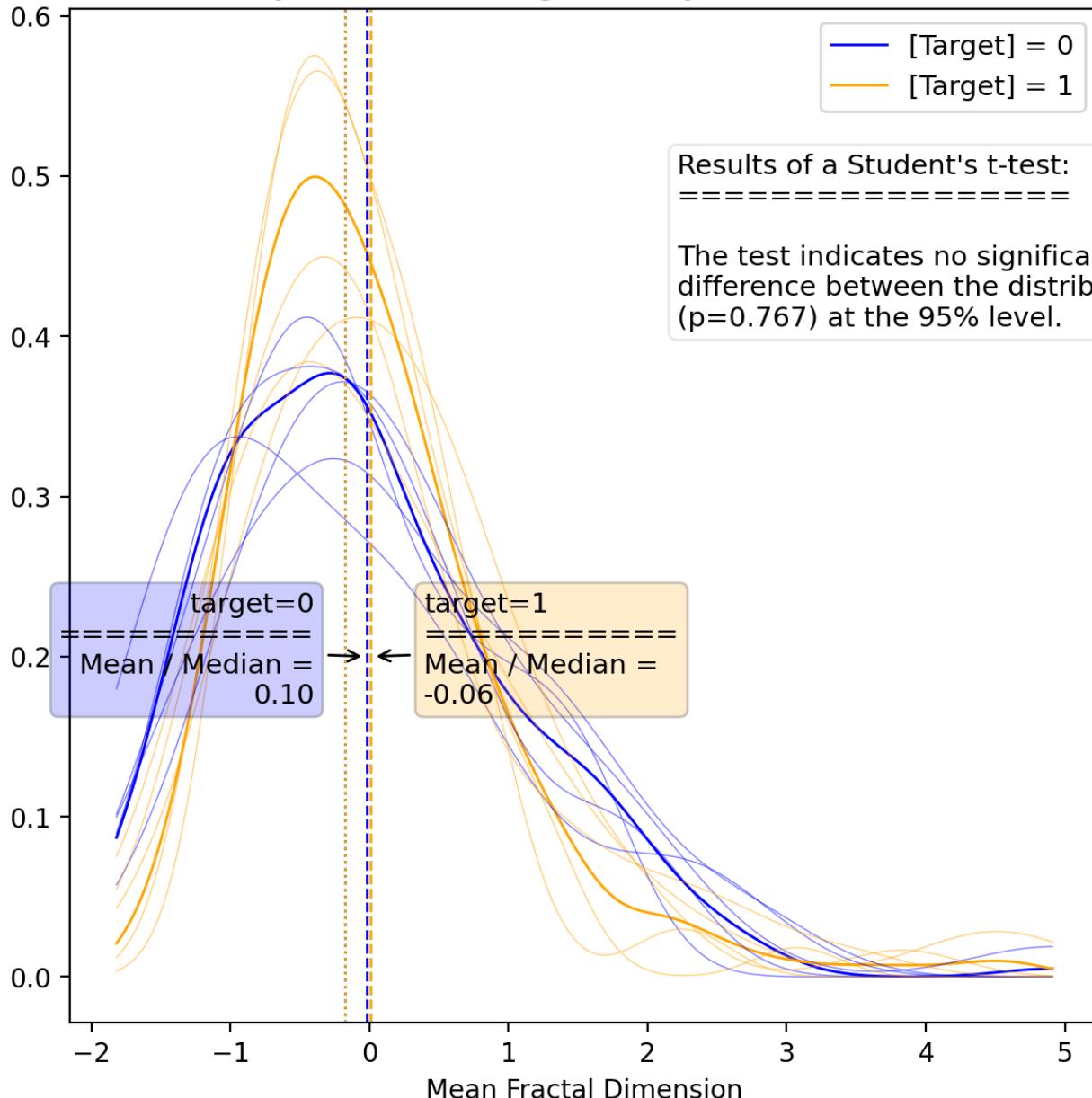
Mean Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	4.0e-02	-4.4e-02	1.7e-02	5.4e-02	4.7e-02	8.1e+00	4.0e-02
Fitted p-Value	0.668	0.658	0.852	0.552	0.625	0.000	0.111
Fitted Std. Err.	9.3e-02	9.9e-02	9.1e-02	9.1e-02	9.6e-02	1.4e+00	3.3e-03
Conf. Int. Lower	-1.4e-01	-2.4e-01	-1.6e-01	-1.2e-01	-1.4e-01	5.4e+00	4.5e-02
Conf. Int. Upper	0.222	0.150	0.196	0.234	0.234	10.768	0.035
Train Accuracy	48.0%	52.3%	48.0%	48.2%	49.8%	48.3%	1.9%
Val Accuracy	52.2%	49.2%	48.6%	48.1%	43.5%	48.3%	3.1%
Train AUC	50.2%	50.5%	50.0%	50.4%	51.6%	50.3%	0.6%
Val AUC	51.8%	47.2%	52.6%	50.7%	45.2%	50.3%	3.2%
Train F1	50.4%	60.3%	49.9%	50.8%	52.3%	50.8%	4.3%
Test F1	57.7%	57.5%	50.9%	46.9%	47.2%	50.8%	5.3%
Train Precision	63.3%	63.3%	62.0%	63.9%	64.1%	63.1%	0.8%
Val Precision	62.5%	60.0%	69.0%	61.2%	59.2%	63.1%	3.9%
Train Recall	41.9%	57.7%	41.8%	42.1%	44.2%	42.6%	6.9%
Val Recall	53.6%	55.3%	40.3%	38.0%	39.2%	42.6%	8.4%
Train MCC	0.4%	1.0%	0.0%	0.8%	3.2%	0.6%	1.2%
Val MCC	3.5%	-5.5%	5.0%	1.5%	-9.3%	0.6%	6.1%
Train Log-Loss	18.74	17.18	18.73	18.68	18.10	18.62	0.68
Val Log-Loss	17.24	18.32	18.52	18.71	20.37	18.62	1.13

Univariate Report

Mean Fractal Dimension - Kernel Density Plot

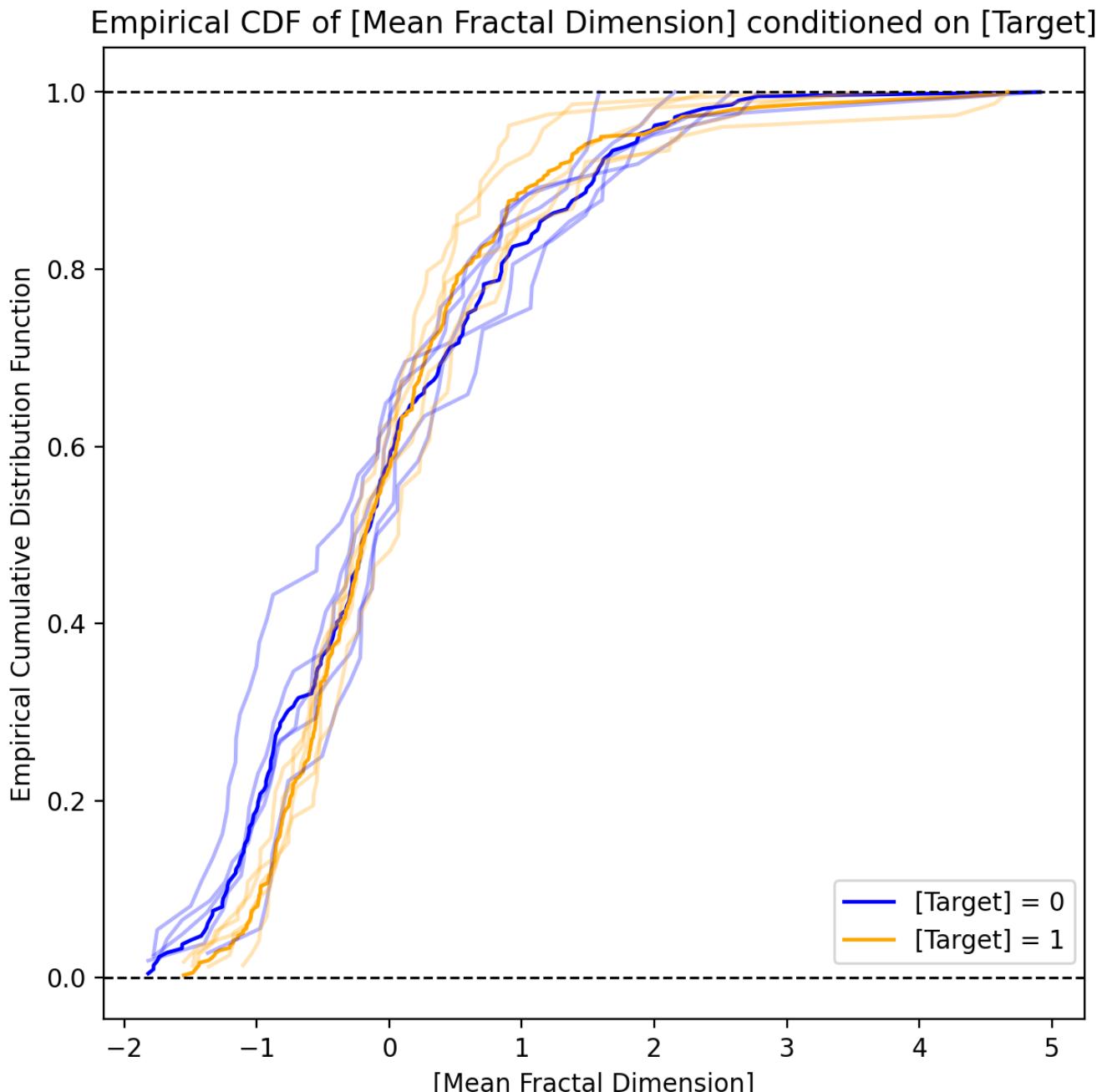
Kernel Density Plot of [Mean Fractal Dimension] by [Target]
Distributions by level are not significantly different at the 95% level.



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

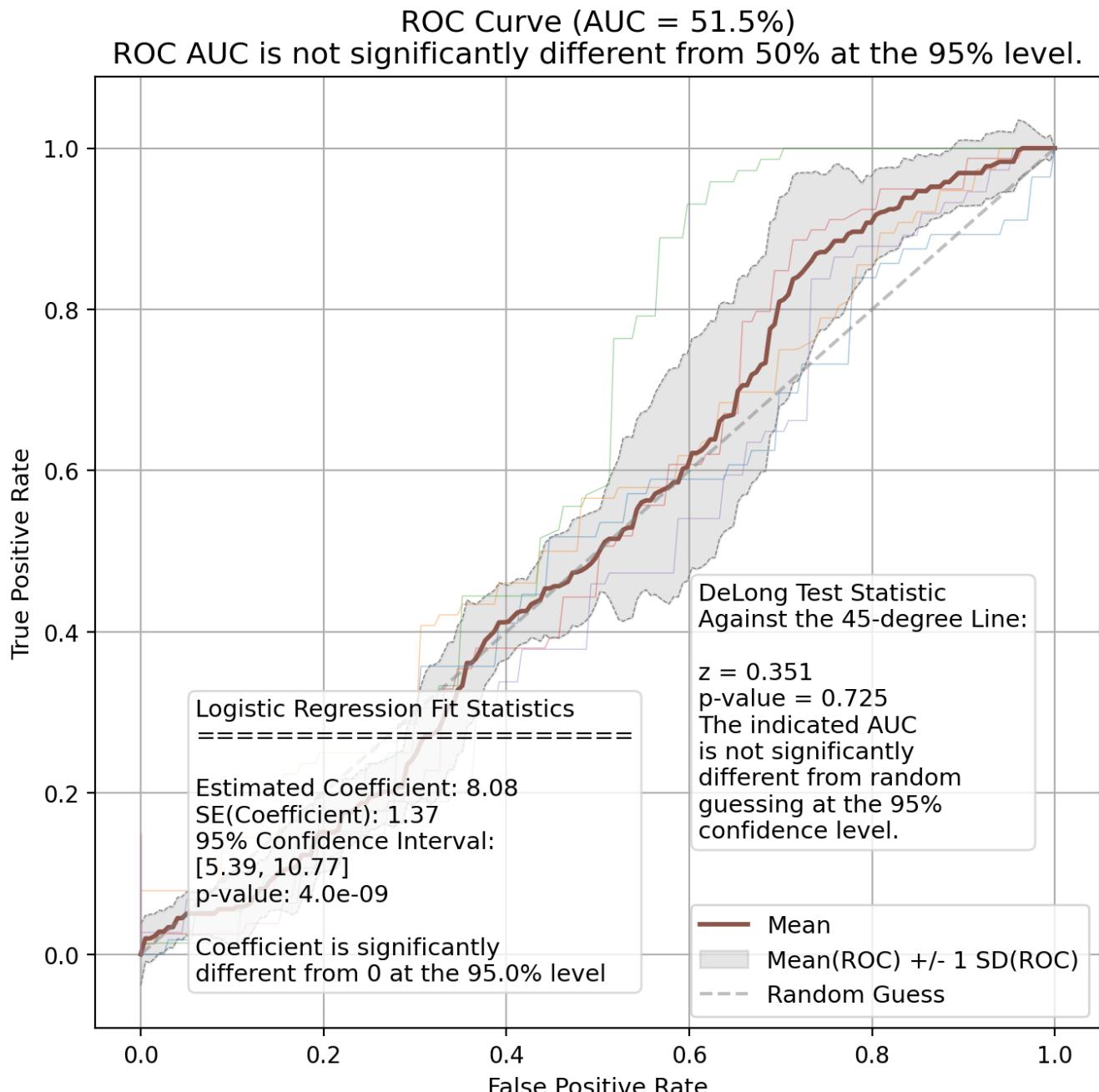
Mean Fractal Dimension - Empirical CDF Plot



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



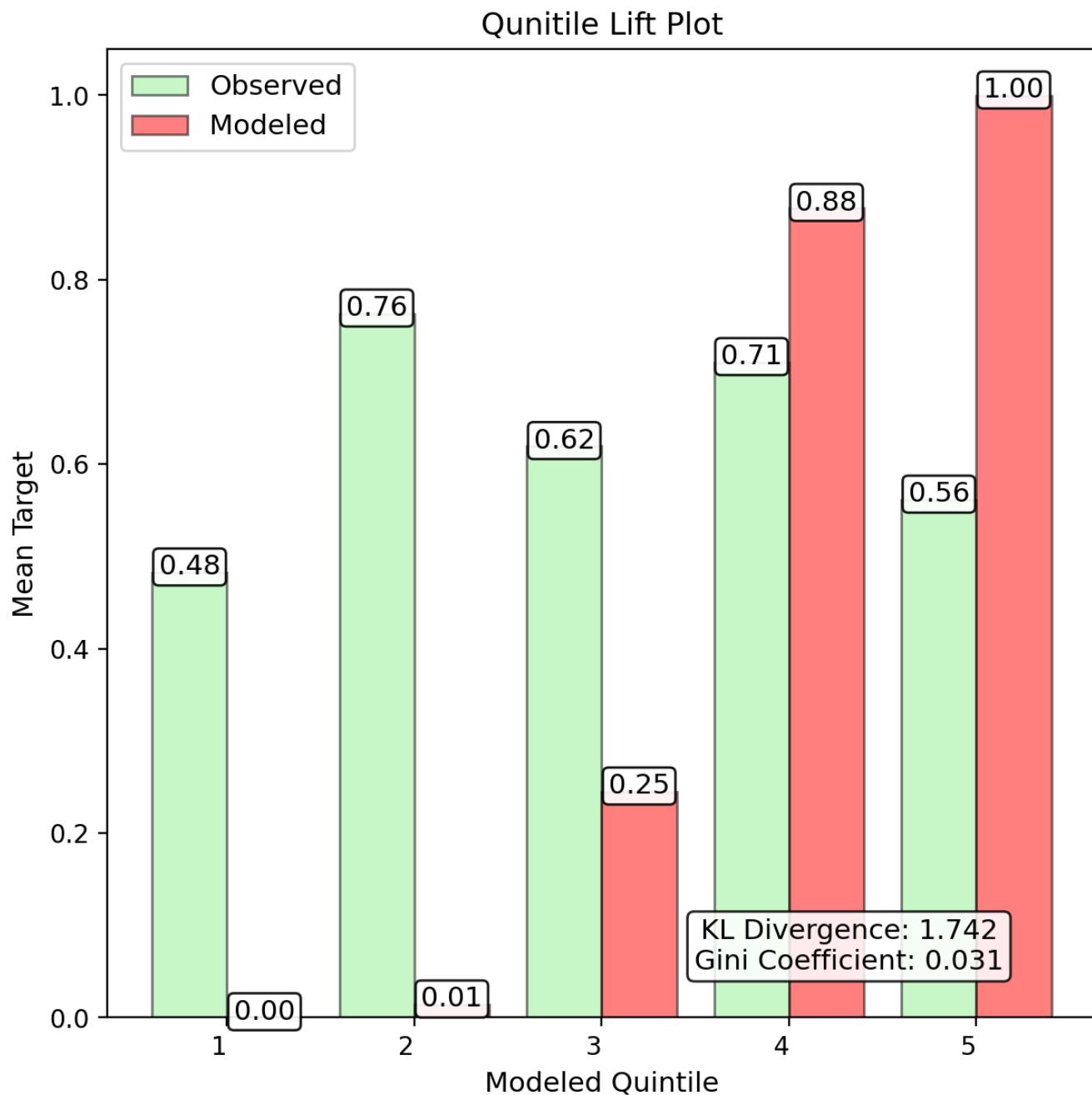
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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.

Univariate Report

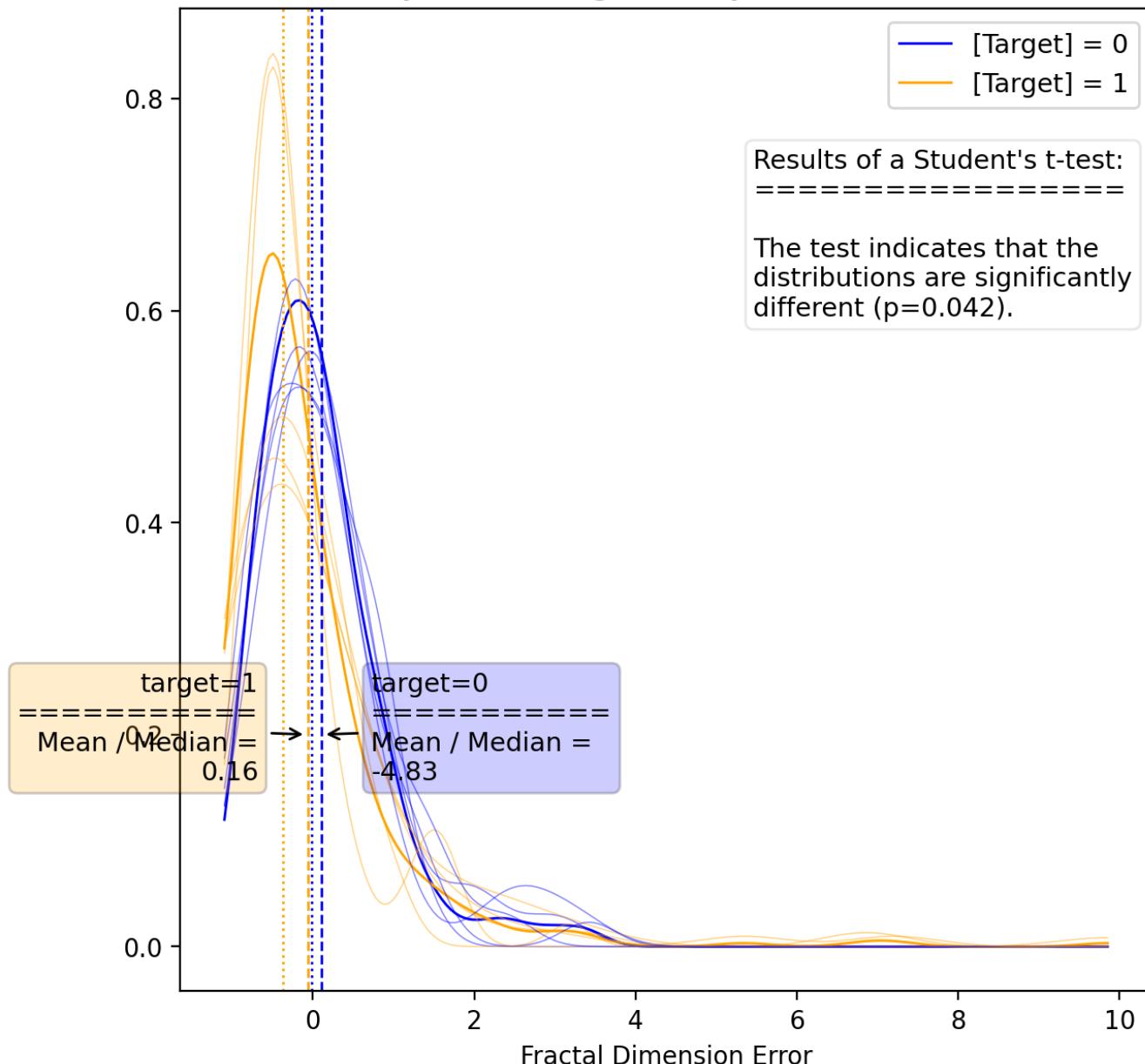
Fractal Dimension Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.6e-01	-2.7e-01	-1.1e-01	-7.5e-02	-2.1e-01	7.5e+01	7.6e-02
Fitted p-Value	0.104	0.023	0.209	0.413	0.060	0.000	0.157
Fitted Std. Err.	0.101	0.118	0.091	0.091	0.110	19.625	0.012
Conf. Int. Lower	-3.6e-01	-5.0e-01	-2.9e-01	-2.5e-01	-4.2e-01	3.6e+01	9.9e-02
Conf. Int. Upper	3.3e-02	-3.8e-02	6.4e-02	1.0e-01	9.0e-03	1.1e+02	5.4e-02
Train Accuracy	61.6%	64.0%	61.7%	60.3%	63.2%	38.3%	1.5%
Val Accuracy	63.0%	56.6%	62.4%	65.6%	57.4%	38.3%	3.9%
Train AUC	58.6%	61.6%	58.8%	57.3%	60.4%	41.2%	1.7%
Val AUC	61.7%	55.3%	58.4%	61.3%	55.5%	41.2%	3.1%
Train F1	69.8%	71.2%	69.8%	68.6%	70.9%	37.9%	1.0%
Test F1	69.1%	63.4%	71.3%	74.3%	65.2%	37.9%	4.4%
Train Precision	69.4%	71.6%	68.4%	68.8%	70.0%	51.4%	1.3%
Val Precision	70.4%	66.7%	71.8%	67.7%	68.7%	51.4%	2.1%
Train Recall	70.1%	70.8%	71.2%	68.3%	71.7%	30.0%	1.3%
Val Recall	67.9%	60.5%	70.8%	82.3%	62.2%	30.0%	8.6%
Train MCC	17.3%	23.1%	17.8%	14.6%	21.0%	-17.7%	3.4%
Val MCC	23.2%	10.3%	16.7%	25.1%	10.6%	-17.7%	6.9%
Train Log-Loss	13.83	12.98	13.79	14.32	13.26	22.23	0.52
Val Log-Loss	13.32	15.66	13.56	12.38	15.36	22.23	1.40

Univariate Report

Fractal Dimension Error - Kernel Density Plot

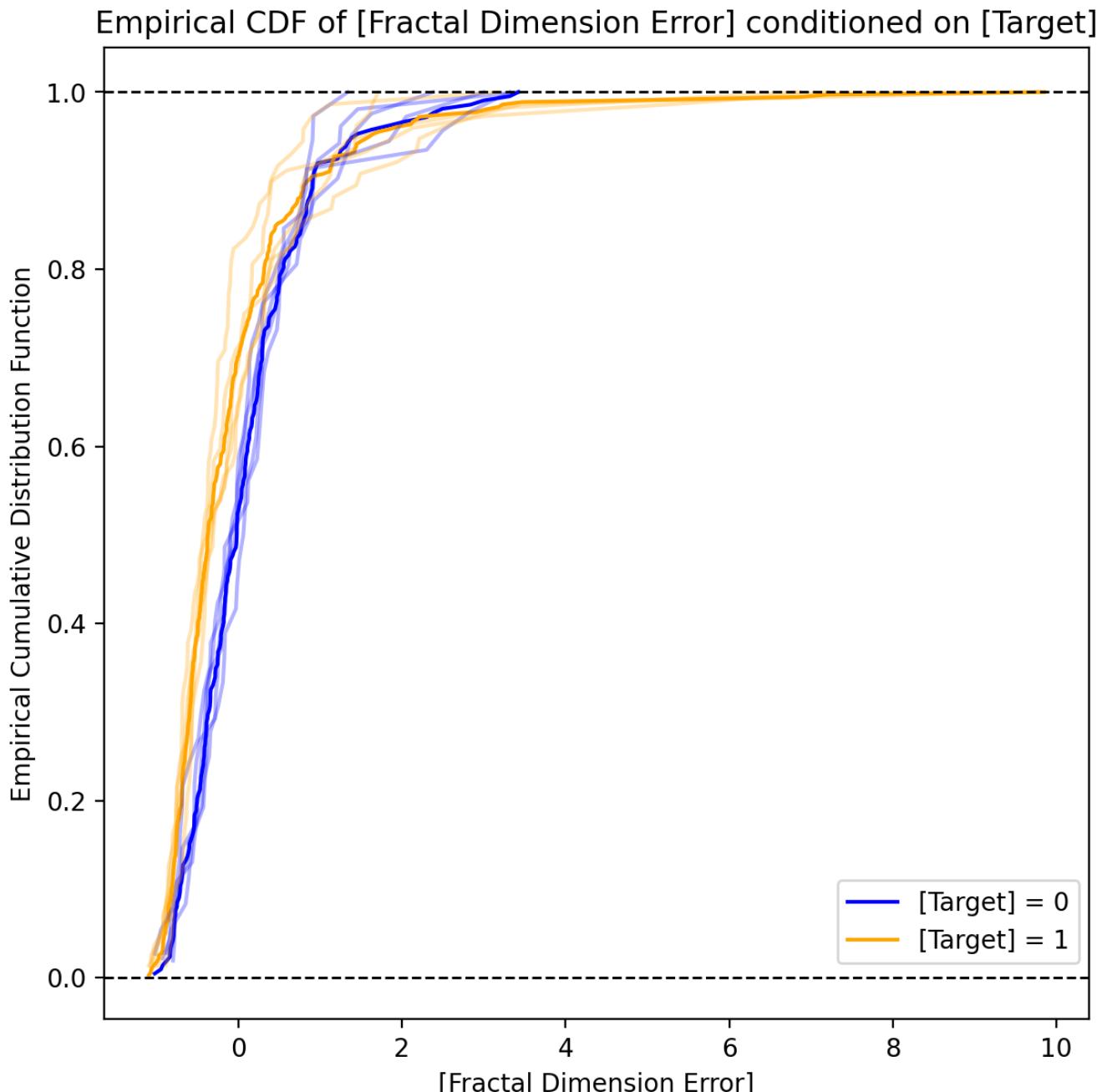
Kernel Density Plot of [Fractal Dimension Error] by [Target]
Distributions by level are significantly different at the 95% level.



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

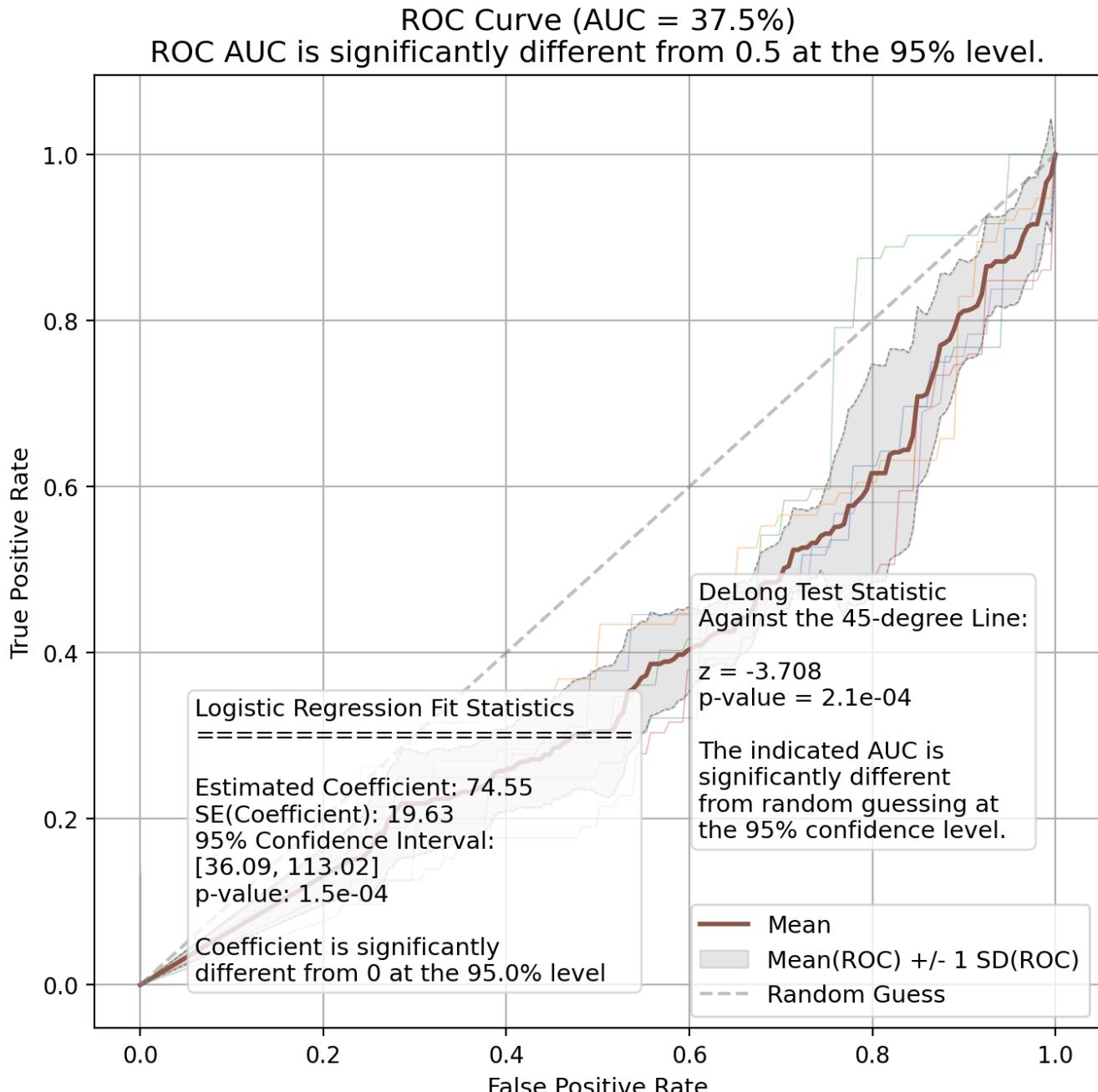
Fractal Dimension Error - Empirical CDF Plot



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



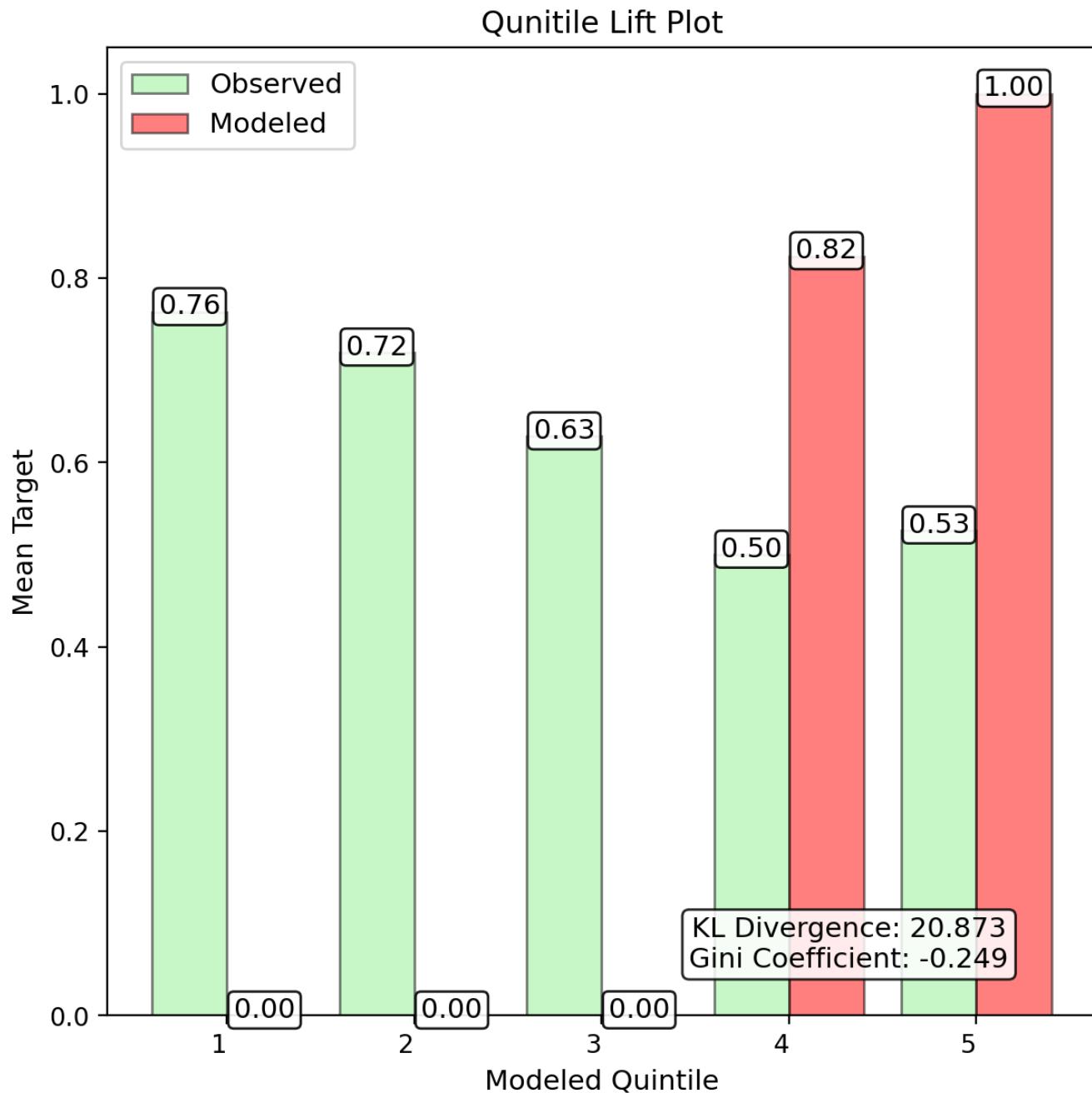
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

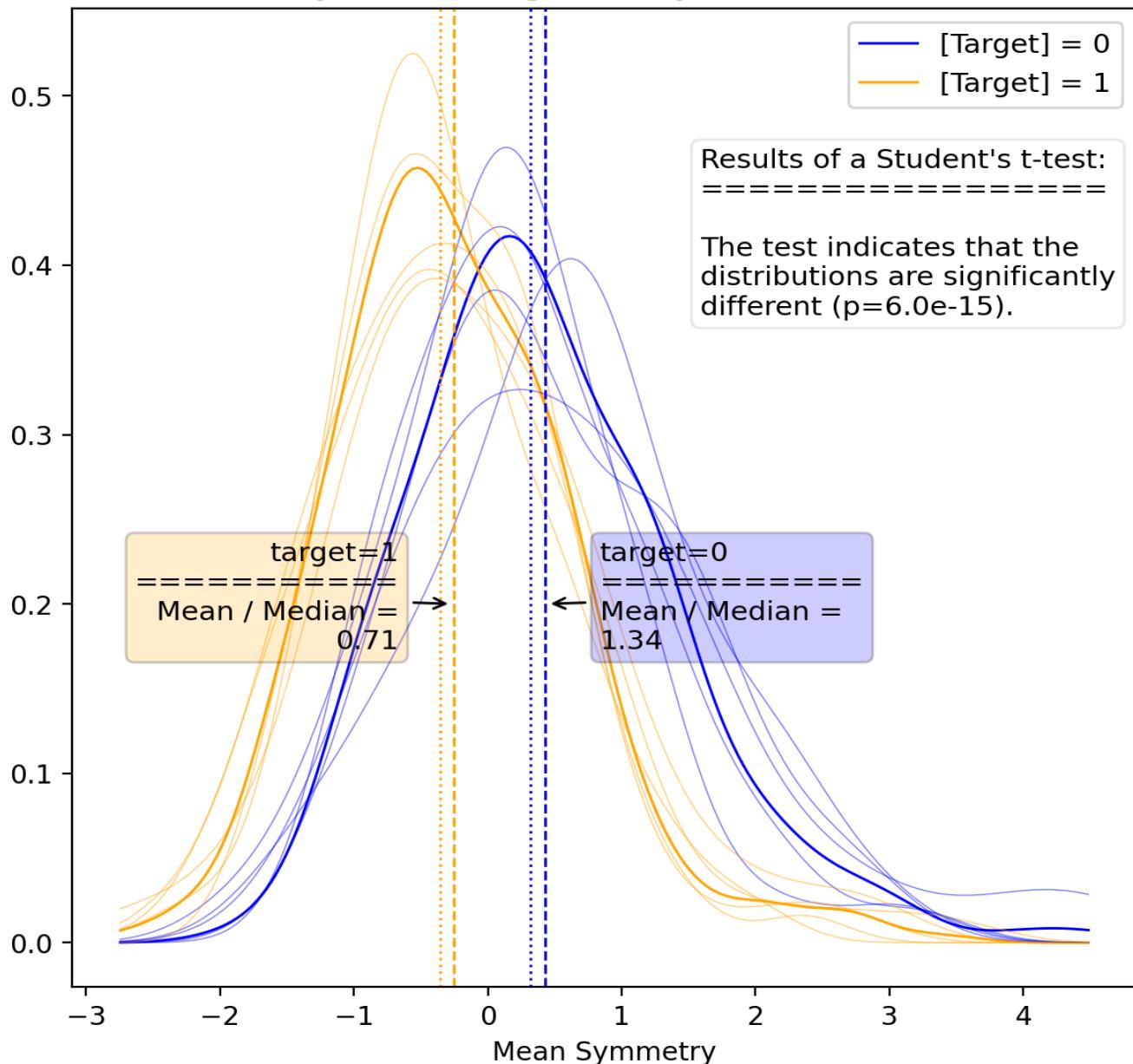
Mean Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-6.6e-01	-8.0e-01	-6.8e-01	-6.9e-01	-8.0e-01	2.2e+00	7.0e-02
Fitted p-Value	1.6e-09	1.4e-11	1.7e-09	8.2e-10	2.8e-11	2.2e-06	8.2e-10
Fitted Std. Err.	0.109	0.119	0.112	0.112	0.119	0.468	0.005
Conf. Int. Lower	-8.7e-01	-1.0e+00	-9.0e-01	-9.1e-01	-1.0e+00	1.3e+00	7.9e-02
Conf. Int. Upper	-4.4e-01	-5.7e-01	-4.6e-01	-4.7e-01	-5.6e-01	3.1e+00	6.1e-02
Train Accuracy	61.8%	64.2%	62.8%	64.4%	64.8%	36.6%	1.2%
Val Accuracy	69.6%	61.5%	66.1%	60.3%	59.1%	36.6%	4.4%
Train AUC	61.7%	64.0%	62.5%	64.1%	65.0%	36.6%	1.3%
Val AUC	70.0%	60.5%	65.8%	59.9%	60.6%	36.6%	4.4%
Train F1	67.3%	69.5%	68.0%	69.9%	69.3%	41.9%	1.1%
Test F1	73.1%	67.6%	72.2%	65.3%	63.6%	41.9%	4.2%
Train Precision	73.3%	74.9%	72.8%	75.4%	75.7%	49.2%	1.3%
Val Precision	79.2%	71.0%	78.7%	69.0%	74.5%	49.2%	4.5%
Train Recall	62.1%	64.8%	63.9%	65.1%	64.0%	36.4%	1.2%
Val Recall	67.9%	64.5%	66.7%	62.0%	55.4%	36.4%	4.9%
Train MCC	22.7%	27.2%	24.4%	27.3%	29.2%	-26.0%	2.6%
Val MCC	39.2%	20.5%	30.1%	19.4%	20.4%	-26.0%	8.6%
Train Log-Loss	13.75	12.90	13.40	12.84	12.70	22.87	0.44
Val Log-Loss	10.97	13.89	12.24	14.31	14.73	22.87	1.58

Univariate Report

Mean Symmetry - Kernel Density Plot

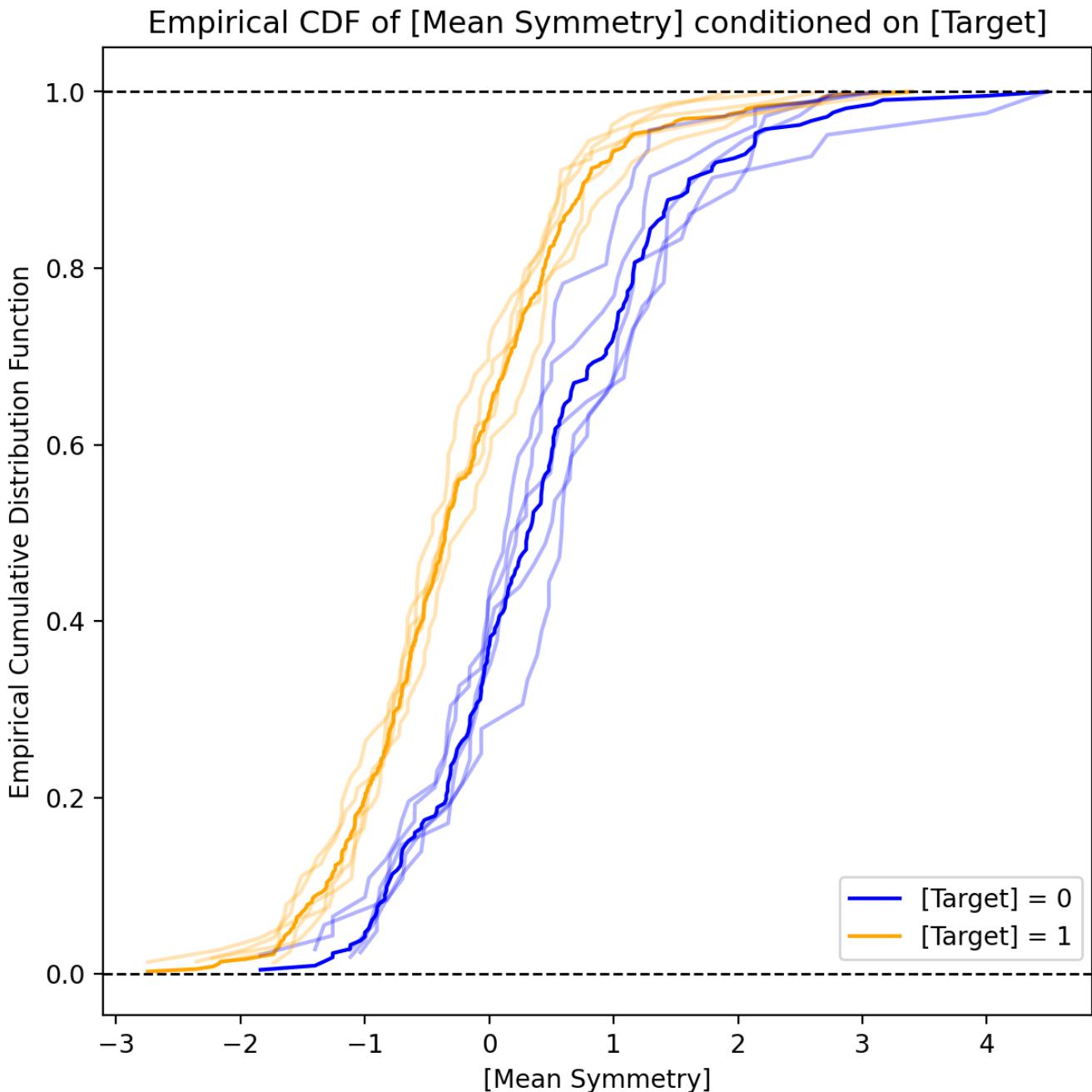
Kernel Density Plot of [Mean Symmetry] by [Target]
Distributions by level are significantly different at the 95% level.



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

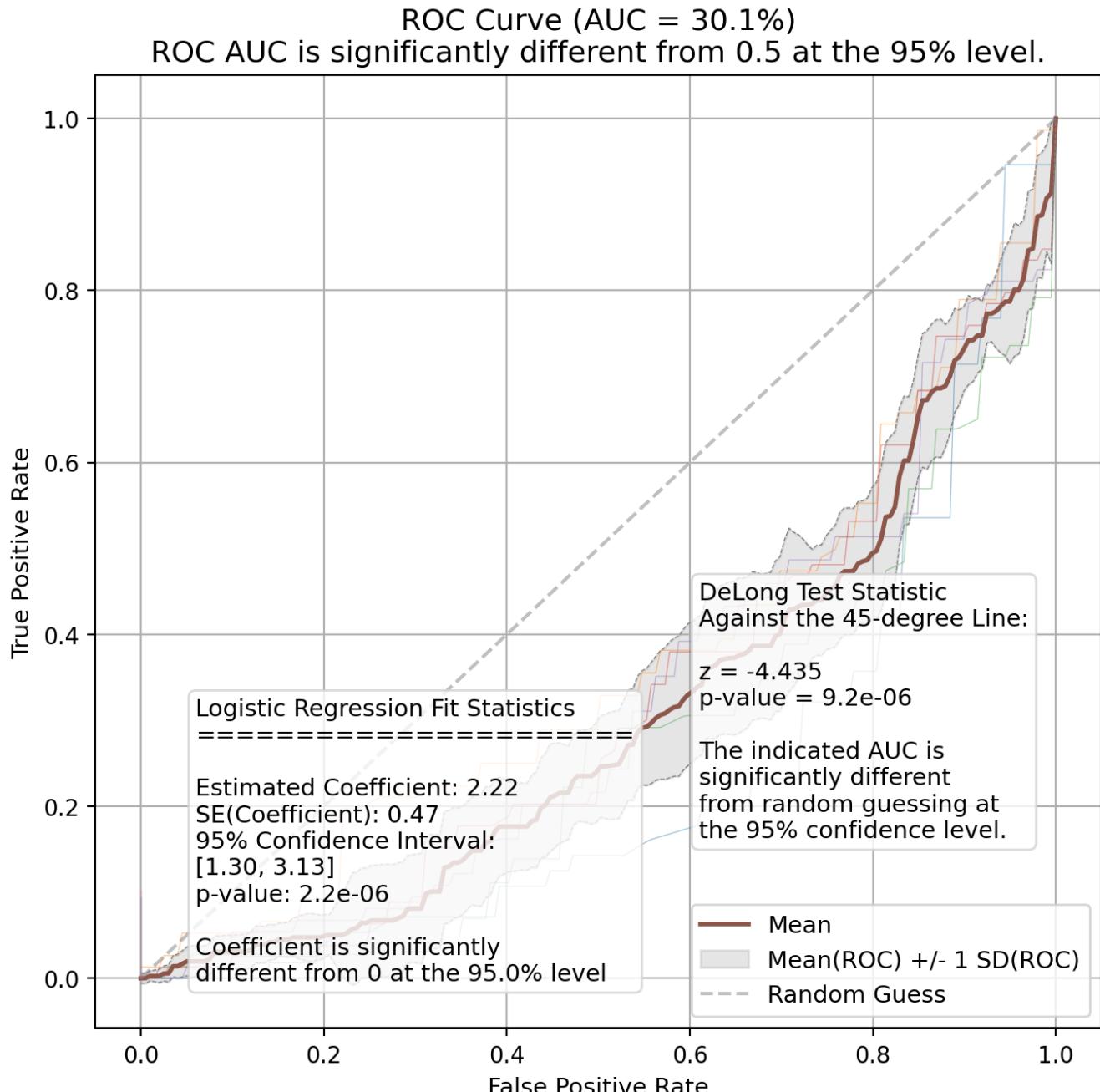
Mean Symmetry - Empirical CDF Plot



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



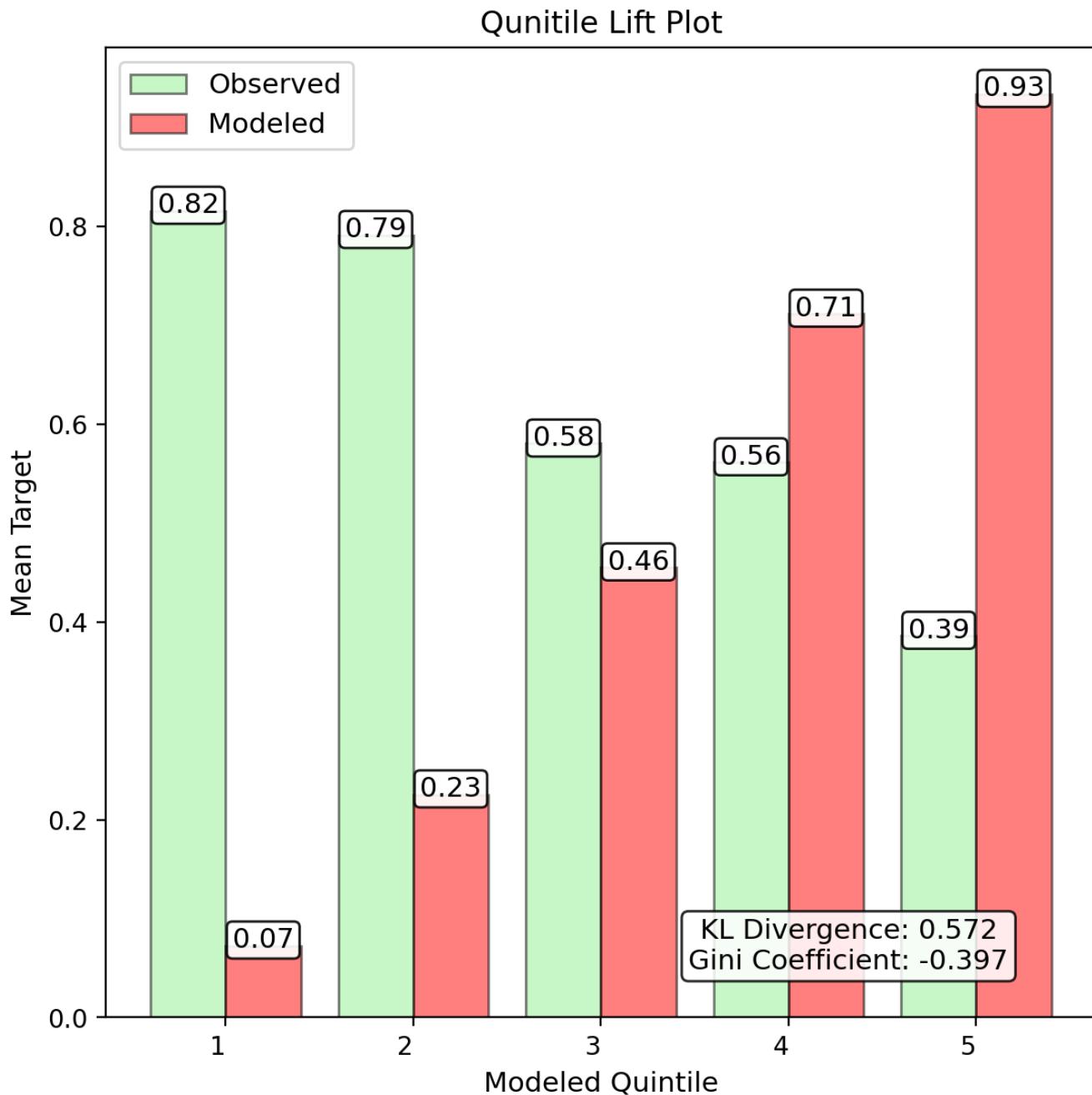
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

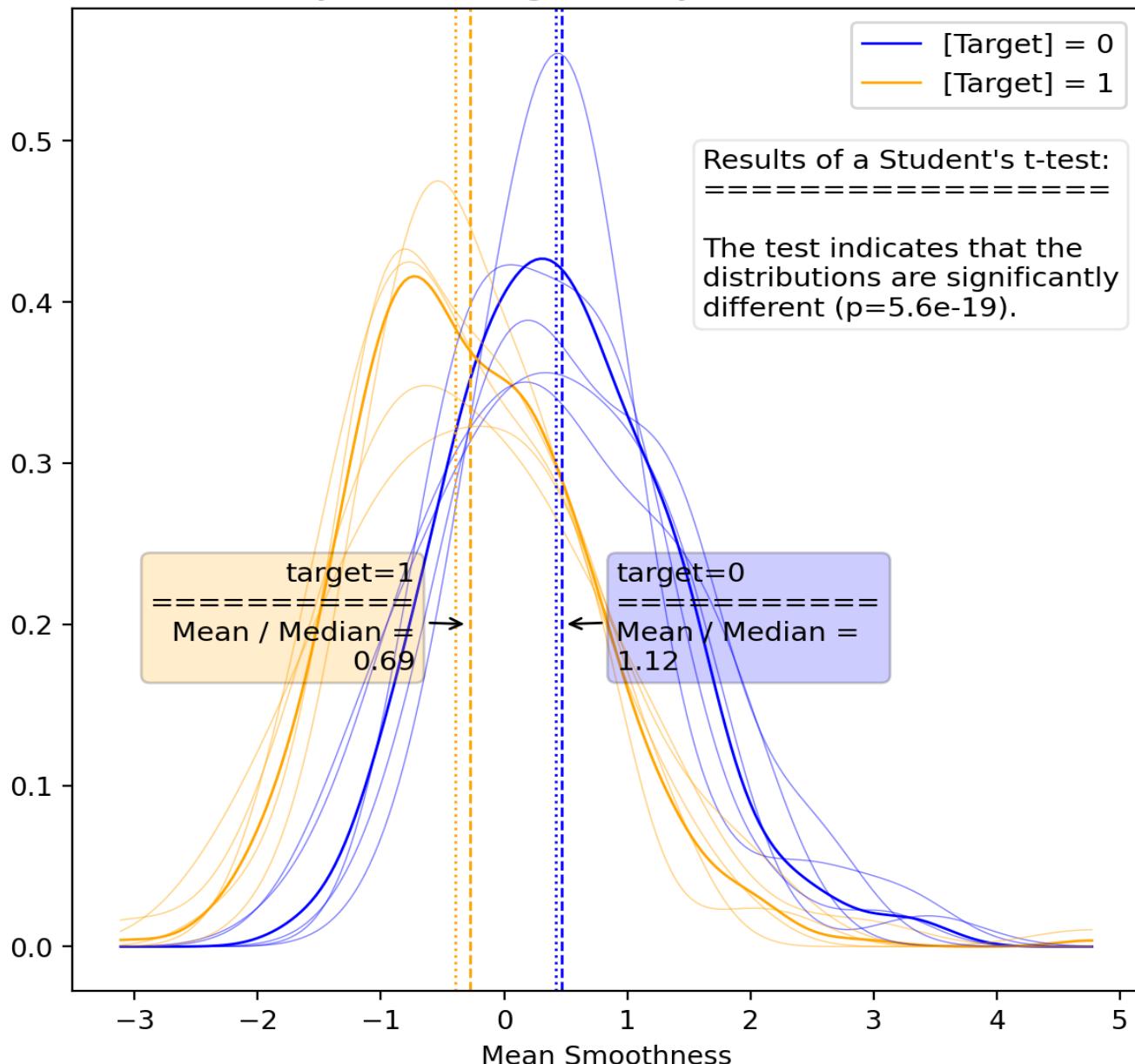
Mean Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-7.8e-01	-9.4e-01	-7.5e-01	-7.7e-01	-7.1e-01	4.1e+00	9.0e-02
Fitted p-Value	3.0e-12	5.3e-14	3.3e-11	1.1e-11	2.1e-10	2.7e-06	9.0e-11
Fitted Std. Err.	0.112	0.125	0.113	0.113	0.111	0.880	0.006
Conf. Int. Lower	-1.0e+00	-1.2e+00	-9.7e-01	-9.9e-01	-9.2e-01	2.4e+00	1.0e-01
Conf. Int. Upper	-5.6e-01	-7.0e-01	-5.3e-01	-5.5e-01	-4.9e-01	5.9e+00	7.9e-02
Train Accuracy	63.9%	66.4%	65.0%	64.6%	63.7%	35.5%	1.1%
Val Accuracy	67.4%	57.4%	67.0%	64.1%	67.8%	35.5%	4.4%
Train AUC	64.5%	67.1%	65.8%	65.5%	64.5%	34.7%	1.1%
Val AUC	69.2%	58.5%	66.5%	64.7%	68.5%	34.7%	4.3%
Train F1	68.6%	70.7%	68.9%	69.1%	67.7%	42.4%	1.1%
Test F1	69.4%	61.2%	73.1%	67.6%	72.6%	42.4%	4.8%
Train Precision	76.1%	78.4%	76.7%	77.6%	75.9%	48.2%	1.0%
Val Precision	81.0%	70.7%	79.0%	74.2%	80.3%	48.2%	4.4%
Train Recall	62.5%	64.4%	62.5%	62.2%	61.1%	37.8%	1.2%
Val Recall	60.7%	53.9%	68.1%	62.0%	66.2%	37.8%	5.5%
Train MCC	27.9%	33.2%	30.7%	29.8%	28.1%	-29.6%	2.1%
Val MCC	37.7%	16.5%	31.5%	28.7%	35.5%	-29.6%	8.3%
Train Log-Loss	13.00	12.10	12.62	12.76	13.10	23.25	0.39
Val Log-Loss	11.75	15.36	11.90	12.93	11.60	23.25	1.57

Univariate Report

Mean Smoothness - Kernel Density Plot

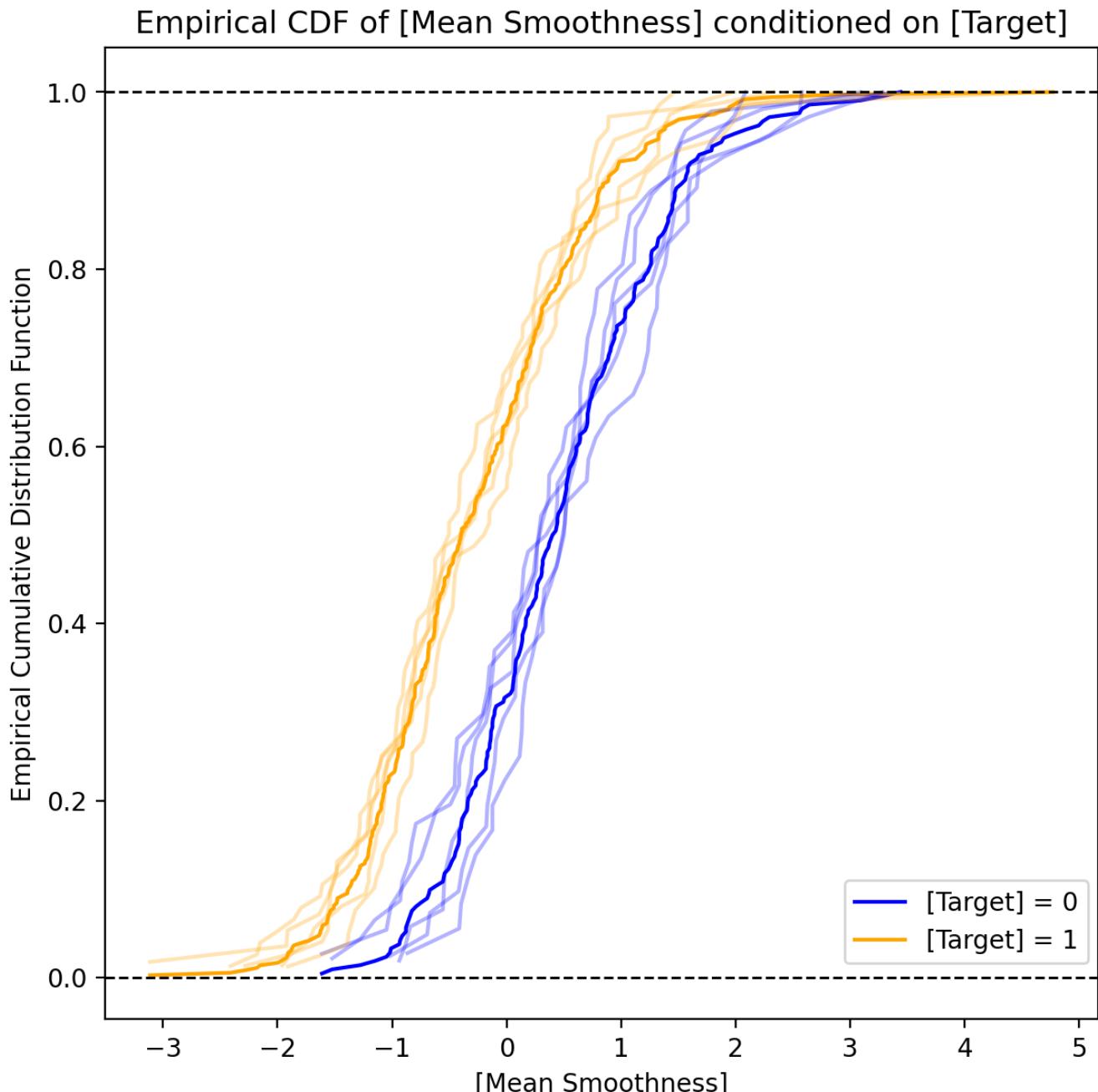
Kernel Density Plot of [Mean Smoothness] by [Target]
Distributions by level are significantly different at the 95% level.



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

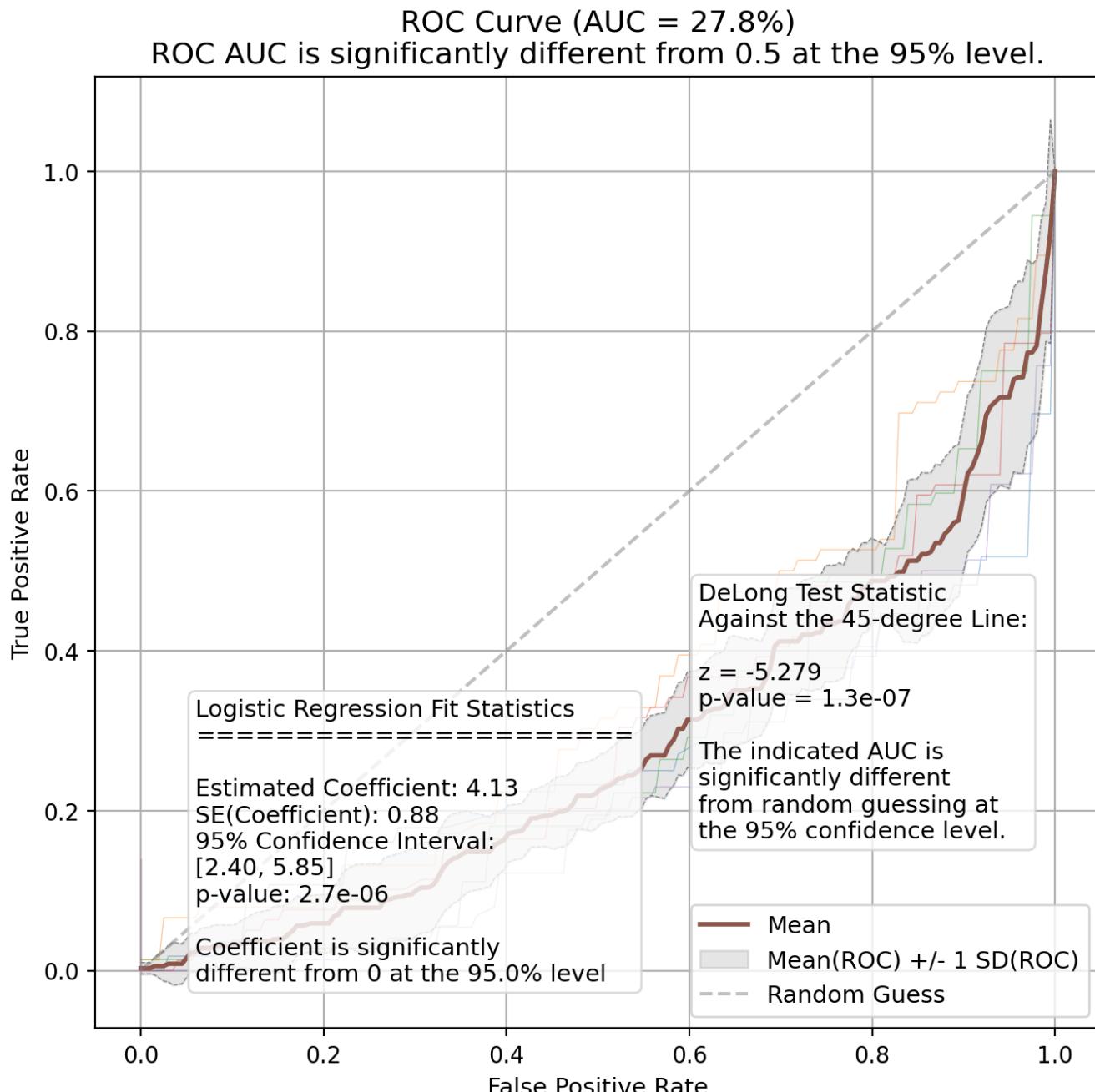
Mean Smoothness - Empirical CDF Plot



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 Smoothness - ROC Curve



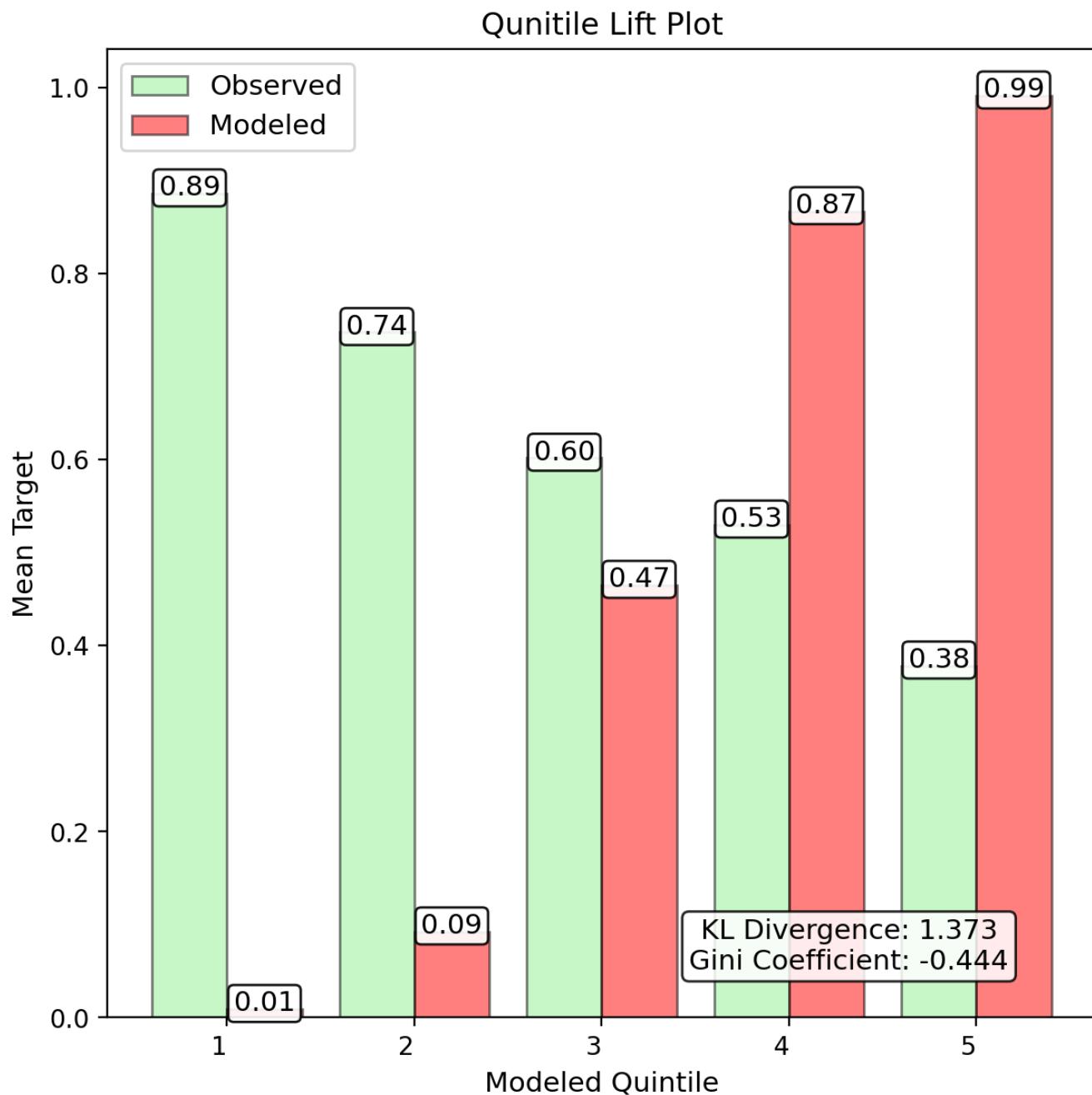
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean Smoothness - 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

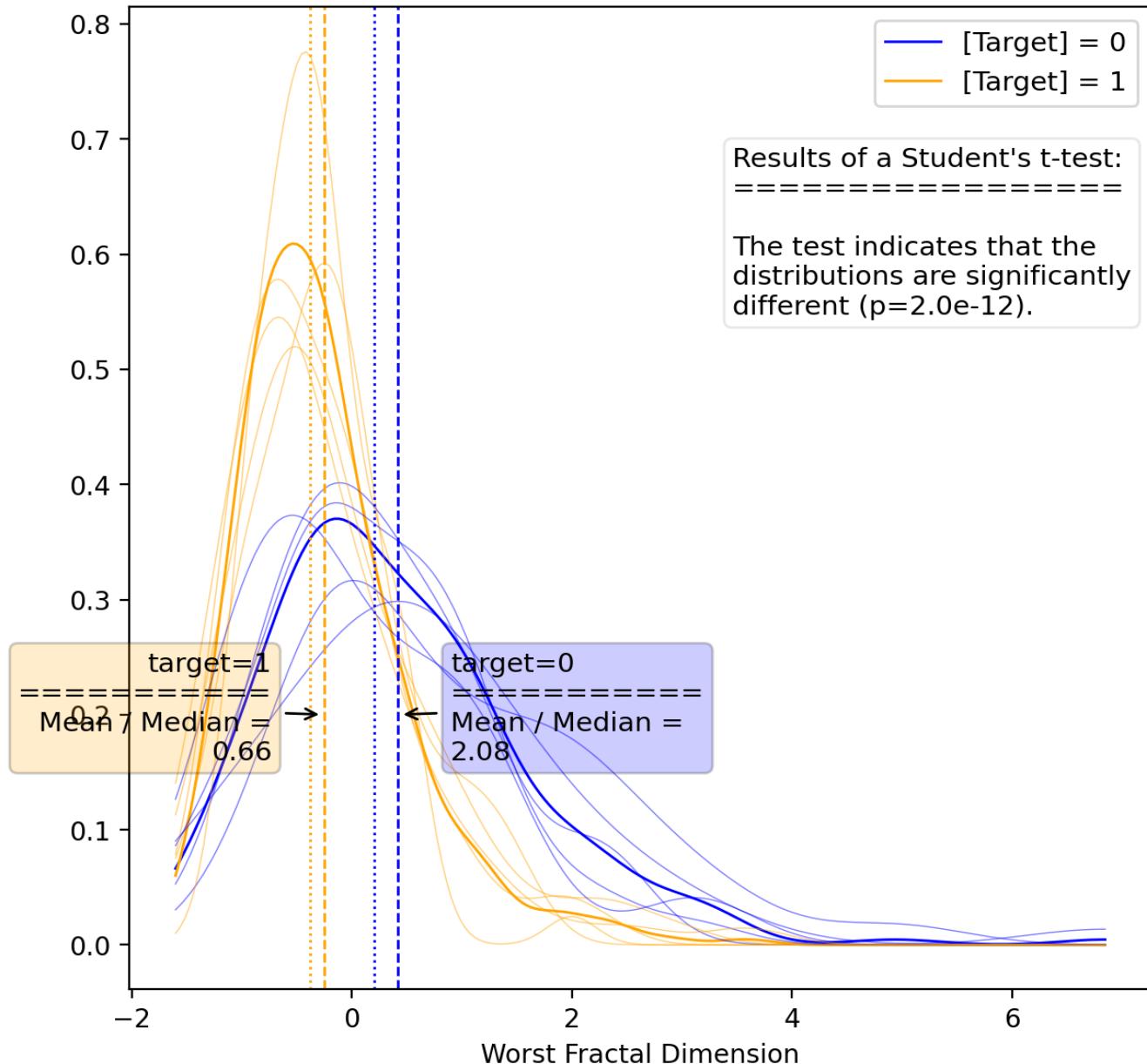
Worst Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-7.0e-01	-8.3e-01	-7.4e-01	-7.6e-01	-7.5e-01	4.2e+00	4.8e-02
Fitted p-Value	2.7e-09	5.9e-11	2.4e-10	6.1e-10	1.7e-09	2.1e-05	1.1e-09
Fitted Std. Err.	0.117	0.127	0.117	0.123	0.124	0.996	0.004
Conf. Int. Lower	-9.3e-01	-1.1e+00	-9.7e-01	-1.0e+00	-9.9e-01	2.3e+00	5.5e-02
Conf. Int. Upper	-4.7e-01	-5.8e-01	-5.1e-01	-5.2e-01	-5.0e-01	6.2e+00	4.1e-02
Train Accuracy	65.4%	67.3%	65.7%	67.1%	64.5%	33.7%	1.2%
Val Accuracy	70.7%	62.3%	66.1%	63.4%	72.2%	33.7%	4.4%
Train AUC	63.4%	65.3%	64.1%	65.3%	62.3%	35.7%	1.3%
Val AUC	71.4%	60.7%	60.5%	61.1%	71.9%	35.7%	6.0%
Train F1	72.2%	73.8%	71.8%	73.5%	71.5%	34.7%	1.1%
Test F1	73.8%	68.9%	75.2%	70.4%	77.1%	34.7%	3.4%
Train Precision	73.3%	74.4%	73.1%	75.2%	71.6%	45.5%	1.4%
Val Precision	80.9%	70.8%	72.7%	68.7%	81.8%	45.5%	6.0%
Train Recall	71.1%	73.3%	70.5%	71.9%	71.4%	28.0%	1.1%
Val Recall	67.9%	67.1%	77.8%	72.2%	73.0%	28.0%	4.3%
Train MCC	26.5%	30.4%	28.0%	30.3%	24.6%	-28.4%	2.5%
Val MCC	41.8%	21.1%	21.9%	22.5%	42.3%	-28.4%	11.1%
Train Log-Loss	12.47	11.77	12.38	11.85	12.78	23.88	0.43
Val Log-Loss	10.58	13.59	12.24	13.21	10.03	23.88	1.57

Univariate Report

Worst Fractal Dimension - Kernel Density Plot

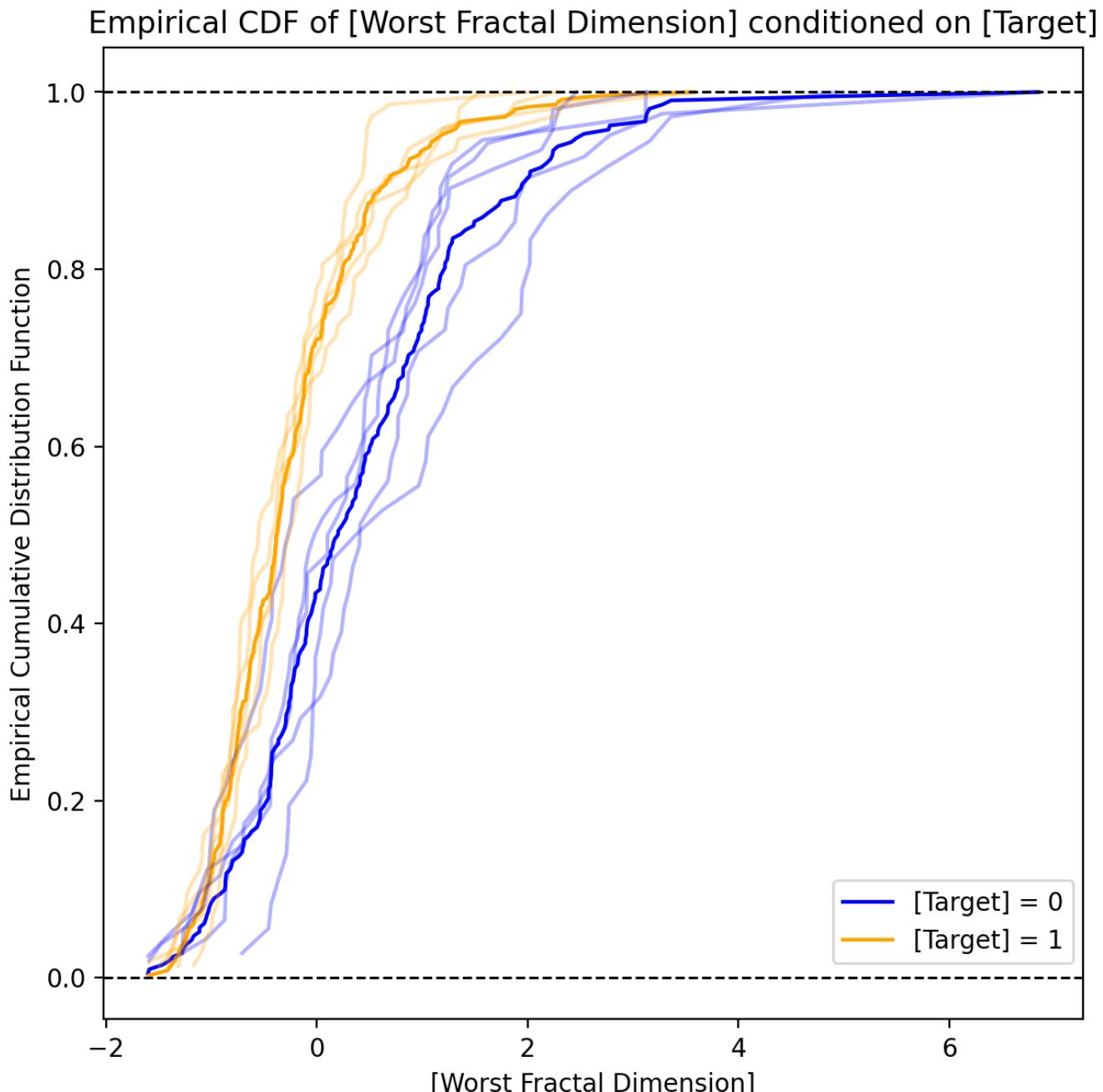
Kernel Density Plot of [Worst Fractal Dimension] by [Target]
Distributions by level are significantly different at the 95% level.



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



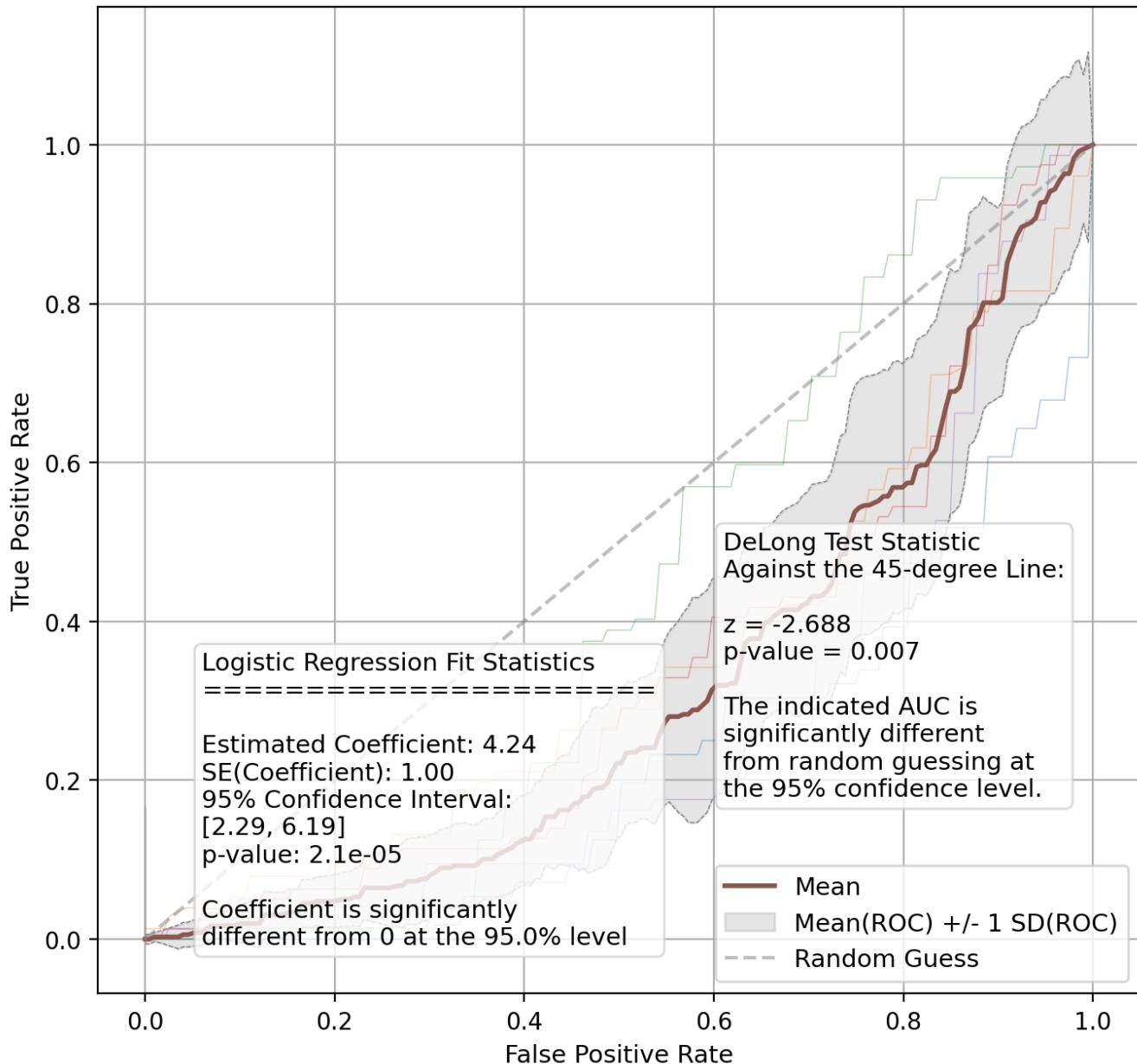
This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly 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

ROC Curve (AUC = 31.4%)

ROC AUC is significantly different from 0.5 at the 95% level.



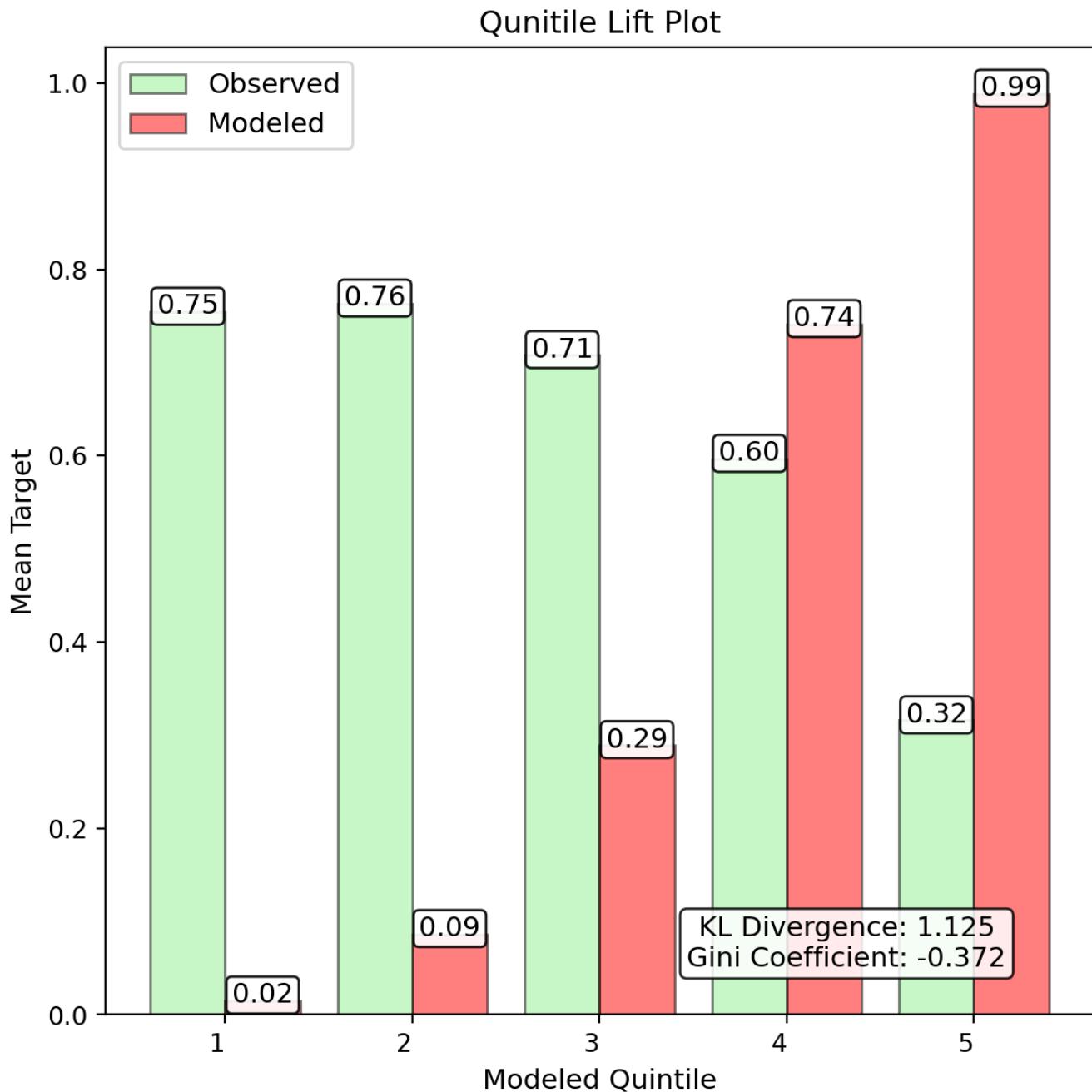
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

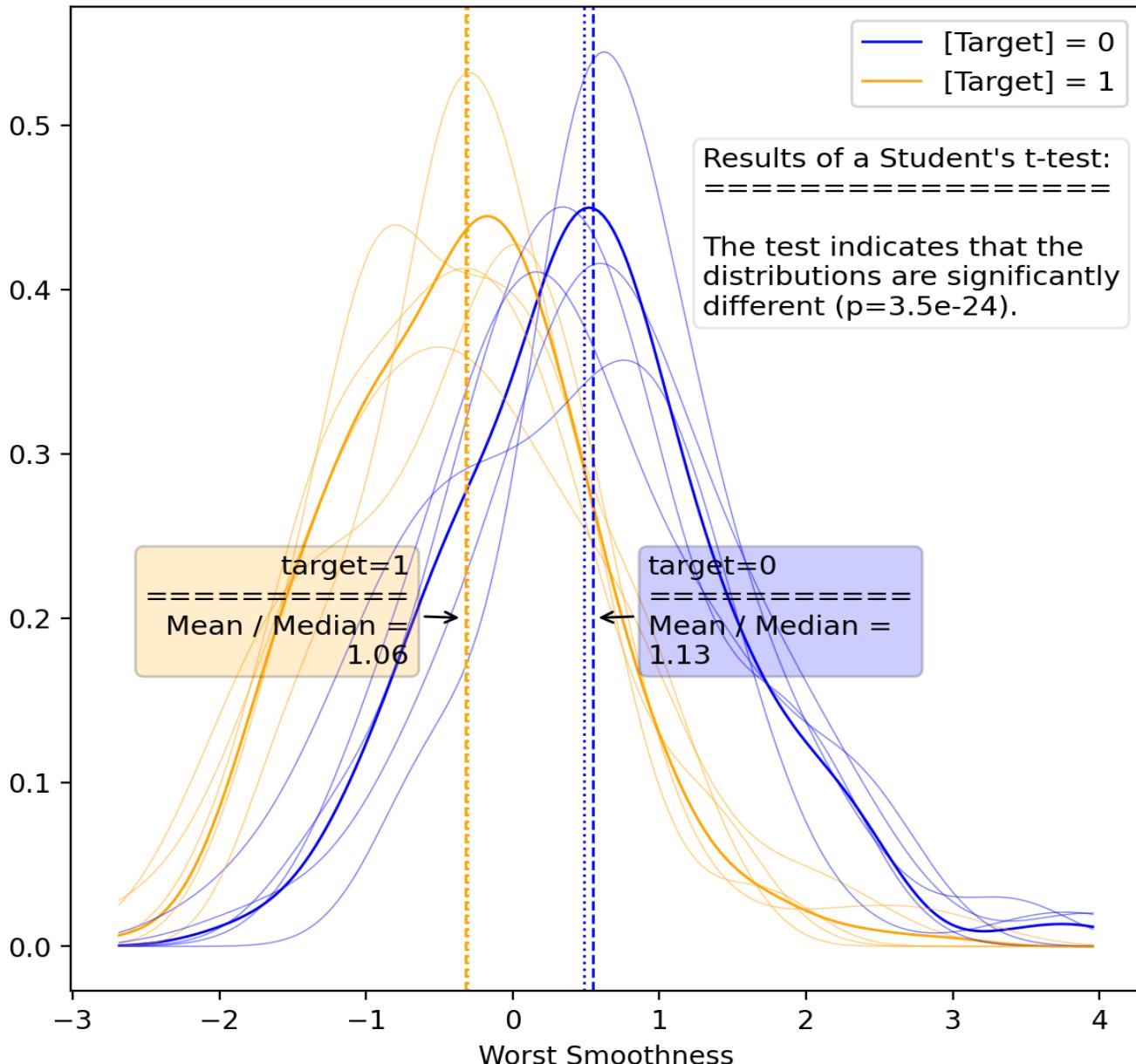
Worst Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-9.5e-01	-1.0e+00	-9.8e-01	-9.9e-01	-9.2e-01	2.7e+00	4.2e-02
Fitted p-Value	4.9e-15	9.8e-16	1.5e-15	1.9e-15	6.5e-14	2.4e-05	2.8e-14
Fitted Std. Err.	0.122	0.128	0.123	0.125	0.122	0.636	0.002
Conf. Int. Lower	-1.2e+00	-1.3e+00	-1.2e+00	-1.2e+00	-1.2e+00	1.4e+00	4.6e-02
Conf. Int. Upper	-7.1e-01	-7.8e-01	-7.4e-01	-7.5e-01	-6.8e-01	3.9e+00	3.7e-02
Train Accuracy	66.2%	68.9%	68.9%	68.3%	66.3%	32.3%	1.4%
Val Accuracy	75.0%	63.9%	65.1%	65.6%	72.2%	32.3%	4.9%
Train AUC	66.8%	69.7%	69.9%	69.4%	67.0%	31.5%	1.6%
Val AUC	77.5%	64.6%	63.1%	65.9%	74.0%	31.5%	6.3%
Train F1	70.8%	72.9%	72.3%	72.3%	70.4%	39.2%	1.1%
Test F1	76.3%	68.1%	72.5%	69.4%	75.8%	39.2%	3.7%
Train Precision	78.0%	80.6%	80.6%	81.2%	77.8%	44.9%	1.6%
Val Precision	90.2%	75.8%	75.8%	75.0%	86.2%	44.9%	7.1%
Train Recall	64.8%	66.5%	65.6%	65.1%	64.3%	34.7%	0.9%
Val Recall	66.1%	61.8%	69.4%	64.6%	67.6%	34.7%	2.9%
Train MCC	32.4%	38.1%	38.7%	37.4%	32.9%	-35.8%	3.0%
Val MCC	54.0%	28.3%	25.4%	31.2%	46.0%	-35.8%	12.4%
Train Log-Loss	12.17	11.21	11.20	11.44	12.15	24.39	0.49
Val Log-Loss	9.01	13.00	12.57	12.38	10.03	24.39	1.77

Univariate Report

Worst Smoothness - Kernel Density Plot

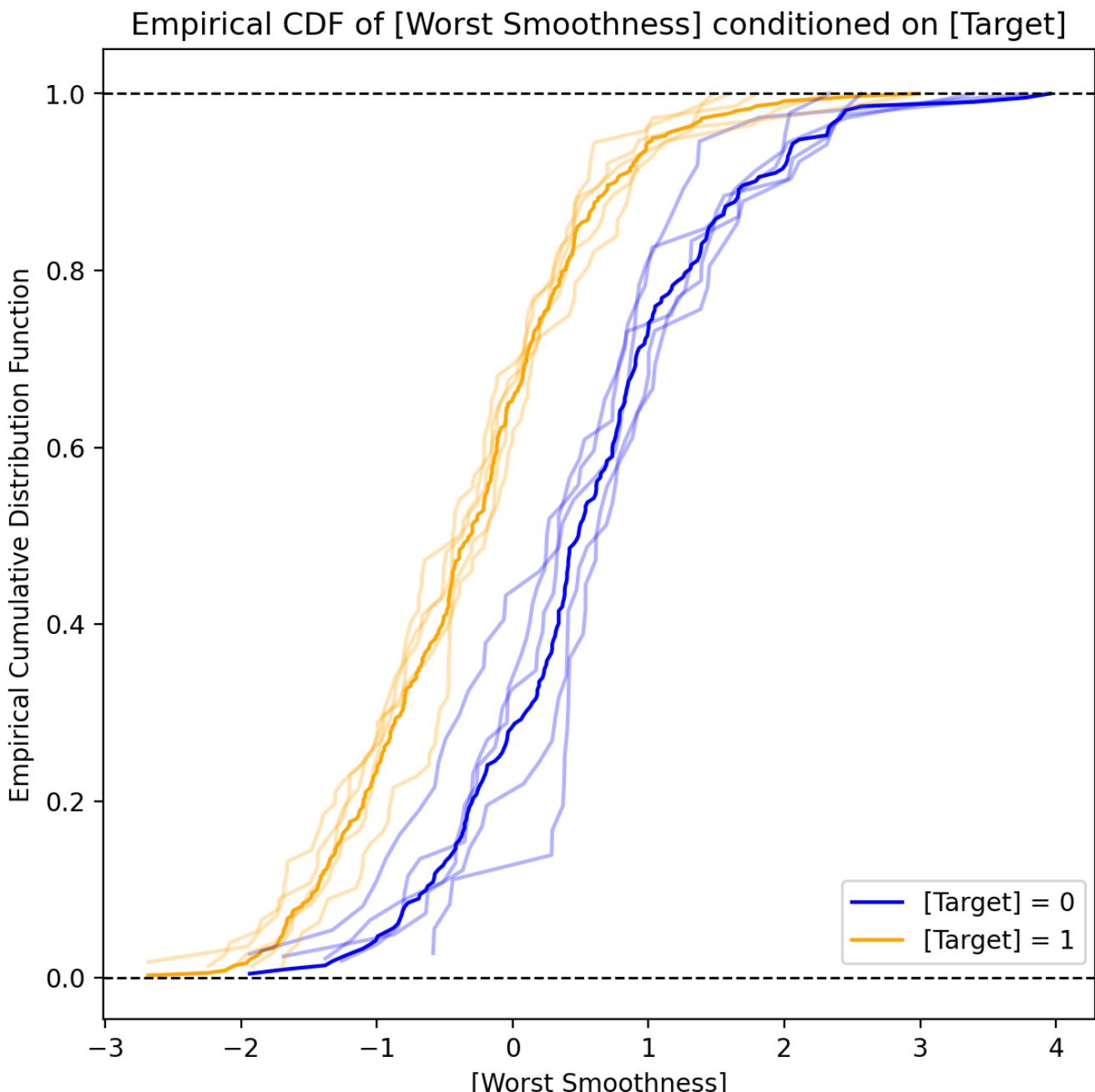
Kernel Density Plot of [Worst Smoothness] by [Target]
Distributions by level are significantly different at the 95% level.



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

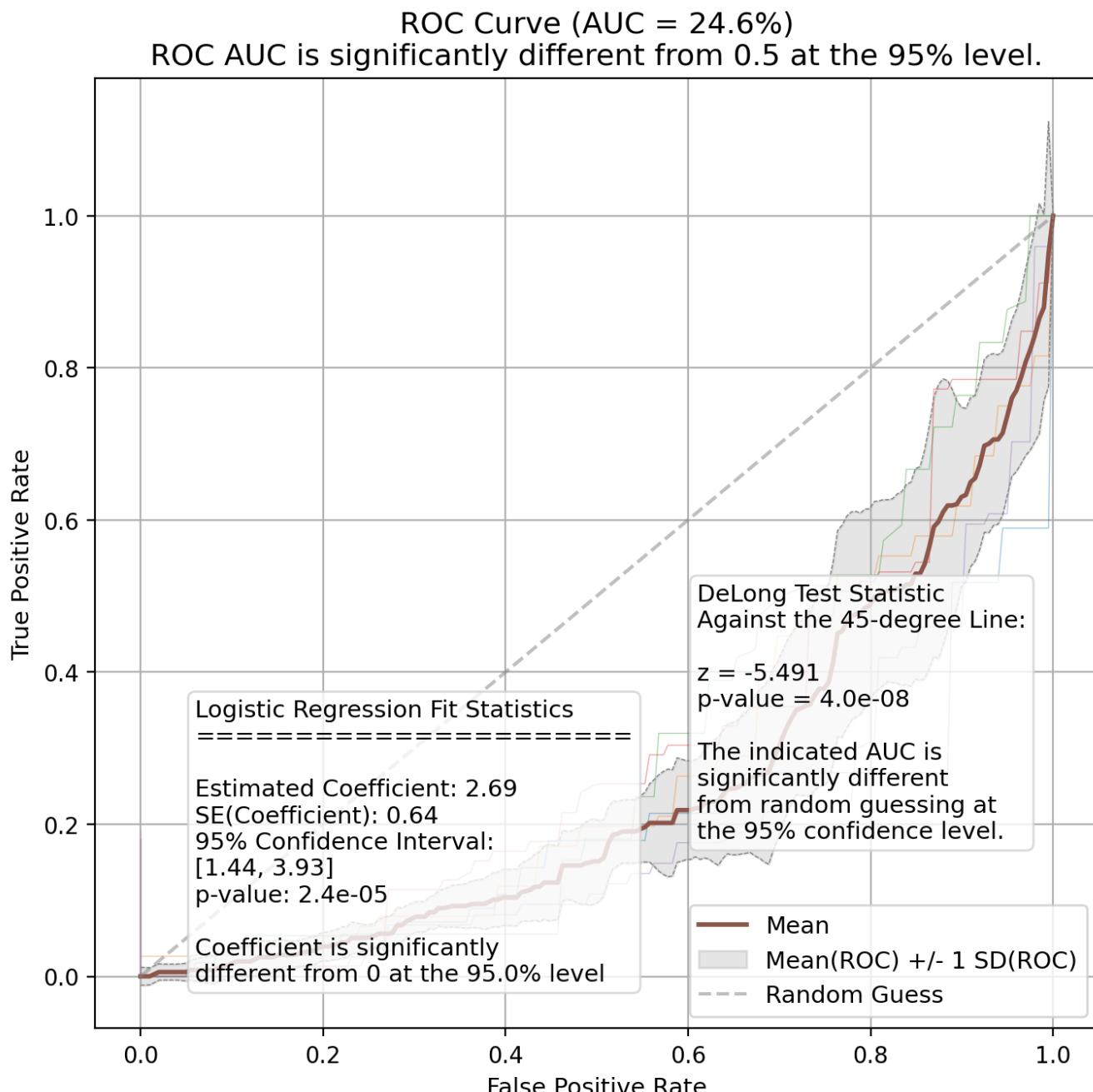
Worst Smoothness - Empirical CDF Plot



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 Smoothness - ROC Curve



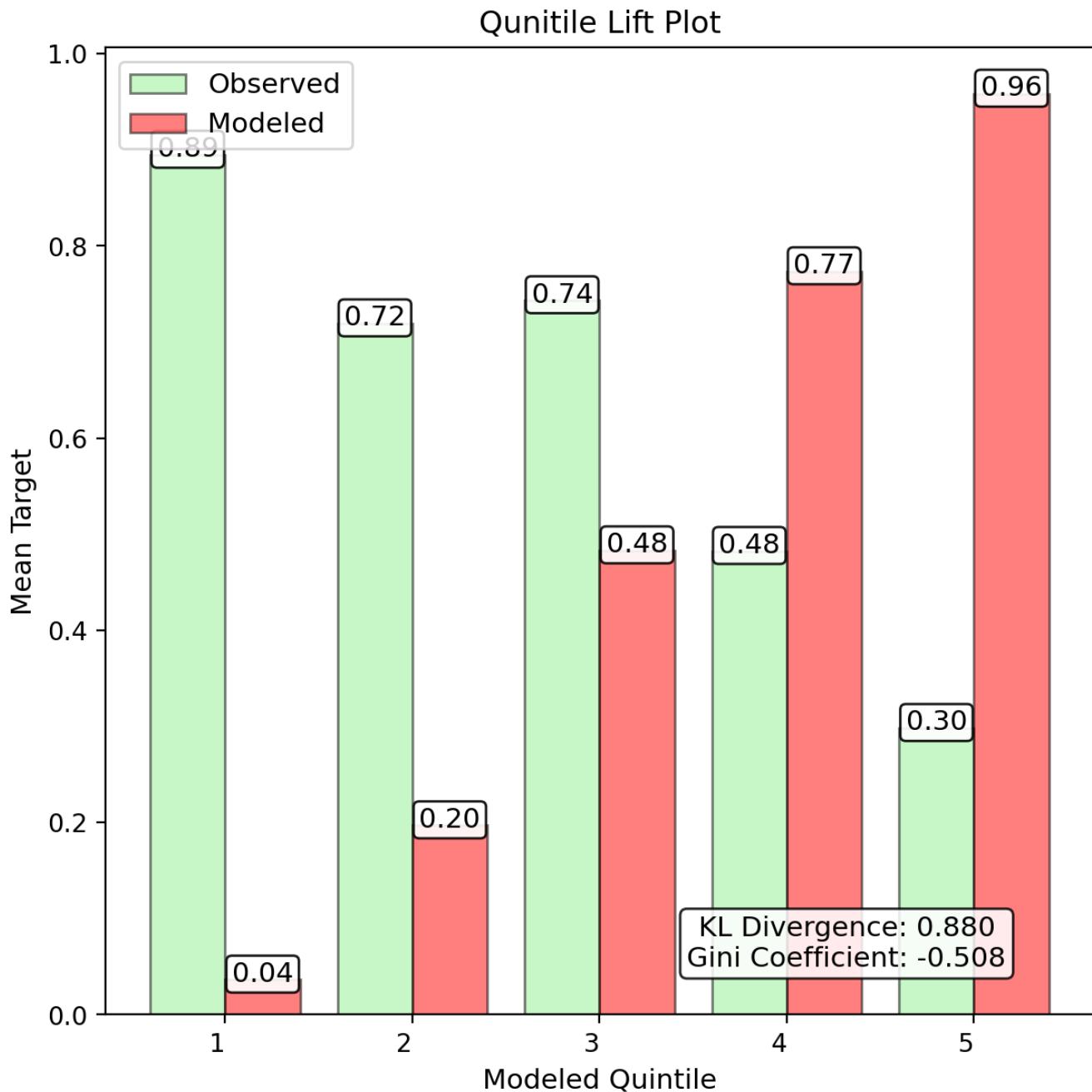
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Smoothness - 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

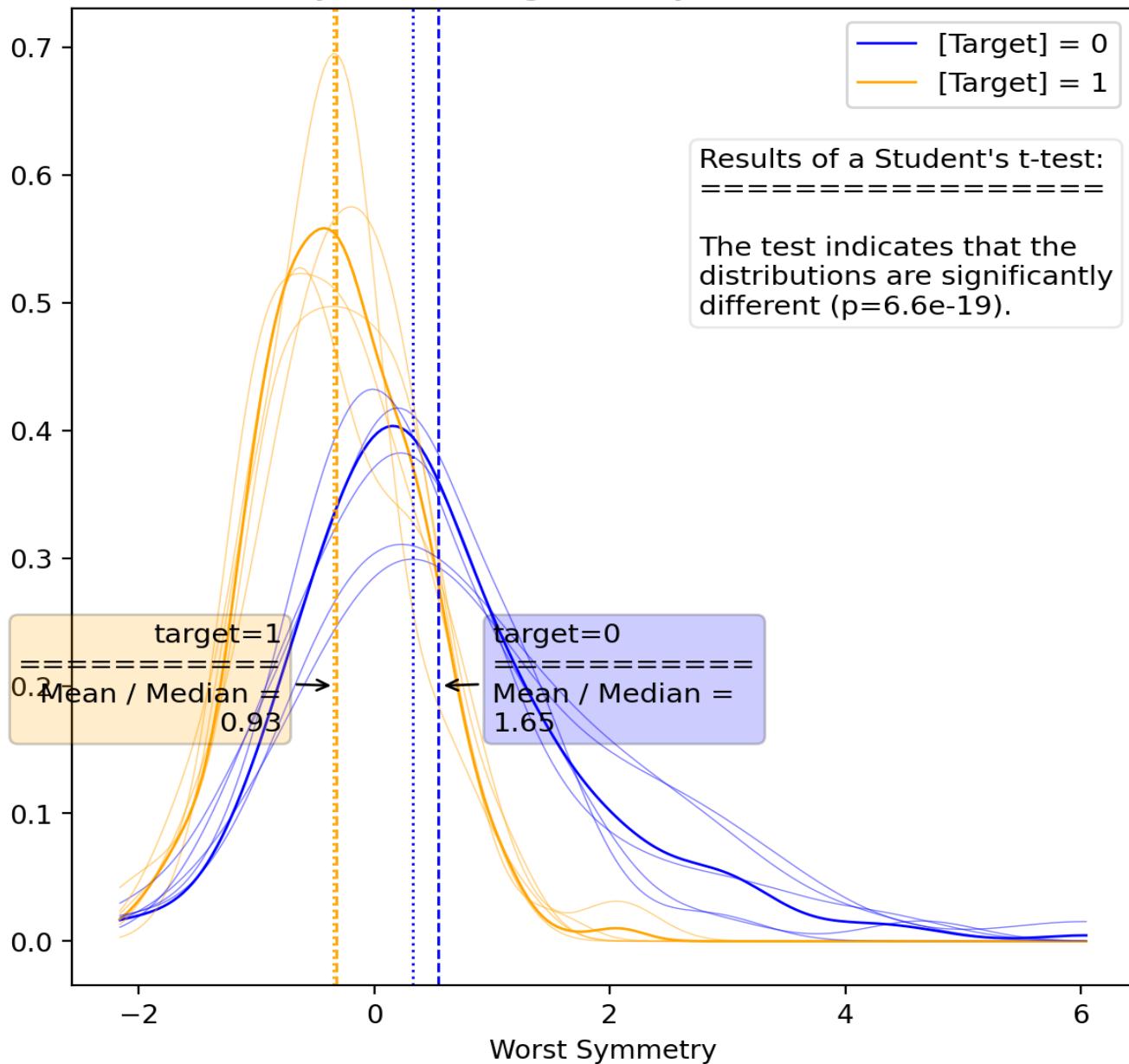
Worst Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.0e+00	-1.2e+00	-1.1e+00	-1.1e+00	-1.1e+00	1.1e+00	7.9e-02
Fitted p-Value	3.6e-14	1.5e-15	3.1e-14	4.2e-15	3.4e-14	1.2e-04	1.7e-14
Fitted Std. Err.	0.137	0.155	0.140	0.146	0.145	0.287	0.007
Conf. Int. Lower	-1.3e+00	-1.5e+00	-1.3e+00	-1.4e+00	-1.4e+00	5.4e-01	9.3e-02
Conf. Int. Upper	-7.7e-01	-9.3e-01	-7.9e-01	-8.6e-01	-8.1e-01	1.7e+00	6.6e-02
Train Accuracy	65.0%	68.5%	67.0%	69.4%	67.6%	32.5%	1.7%
Val Accuracy	78.3%	63.9%	69.7%	61.8%	66.1%	32.5%	6.5%
Train AUC	64.6%	67.8%	66.3%	69.0%	66.9%	33.1%	1.6%
Val AUC	78.2%	62.9%	68.5%	60.8%	67.1%	33.1%	6.7%
Train F1	70.4%	73.7%	72.2%	74.5%	72.9%	36.4%	1.6%
Test F1	81.5%	69.9%	75.9%	67.5%	70.7%	36.4%	5.6%
Train Precision	75.4%	77.3%	75.5%	79.0%	76.2%	44.5%	1.5%
Val Precision	84.6%	72.9%	80.0%	69.3%	79.7%	44.5%	6.1%
Train Recall	66.1%	70.5%	69.1%	70.5%	70.0%	30.8%	1.8%
Val Recall	78.6%	67.1%	72.2%	65.8%	63.5%	30.8%	6.0%
Train MCC	28.3%	34.7%	31.9%	36.9%	33.0%	-33.0%	3.2%
Val MCC	55.5%	25.3%	35.8%	21.4%	32.8%	-33.0%	13.2%
Train Log-Loss	12.62	11.37	11.91	11.03	11.67	24.32	0.60
Val Log-Loss	7.84	13.00	10.91	13.76	12.22	24.32	2.33

Univariate Report

Worst Symmetry - Kernel Density Plot

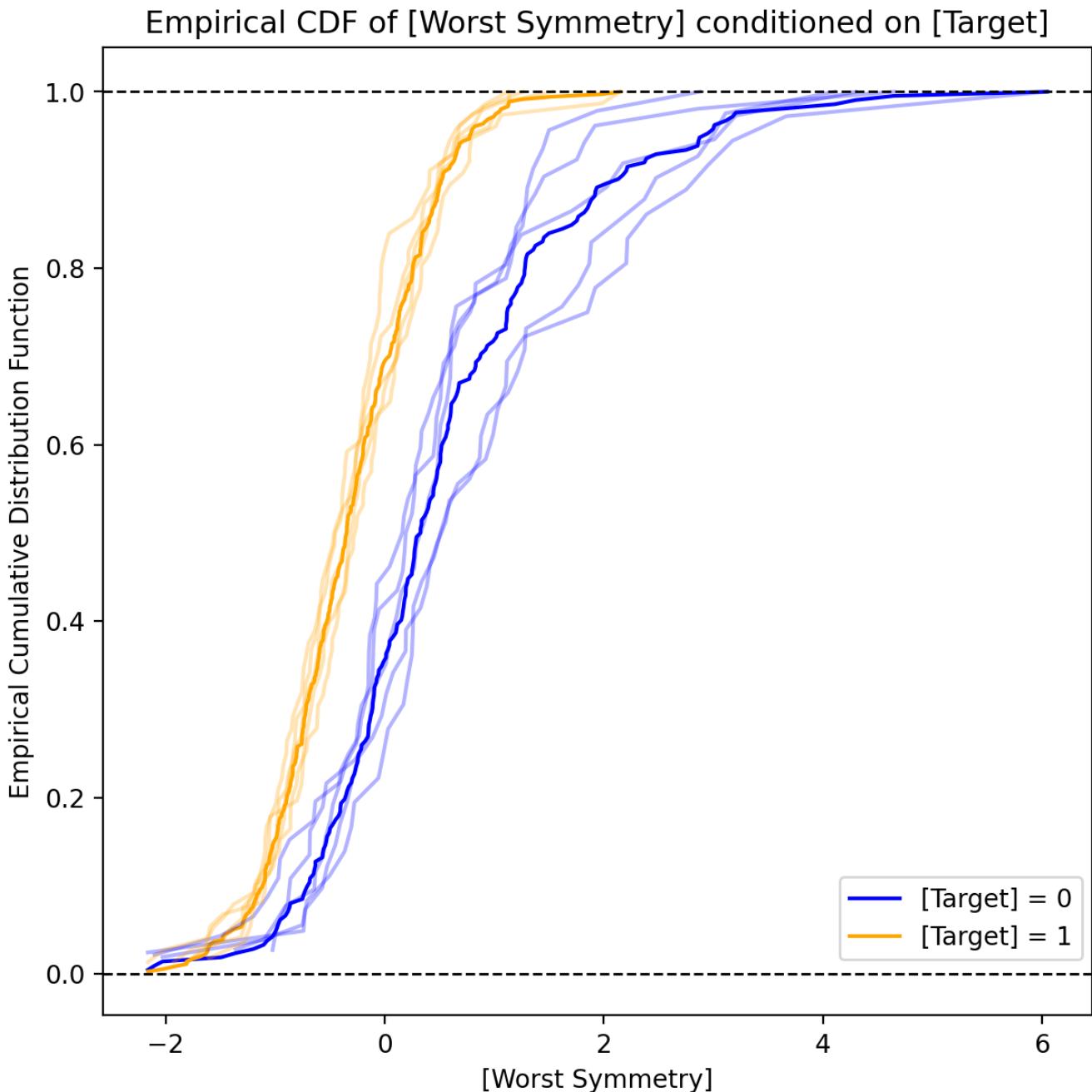
Kernel Density Plot of [Worst Symmetry] by [Target]
Distributions by level are significantly different at the 95% level.



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

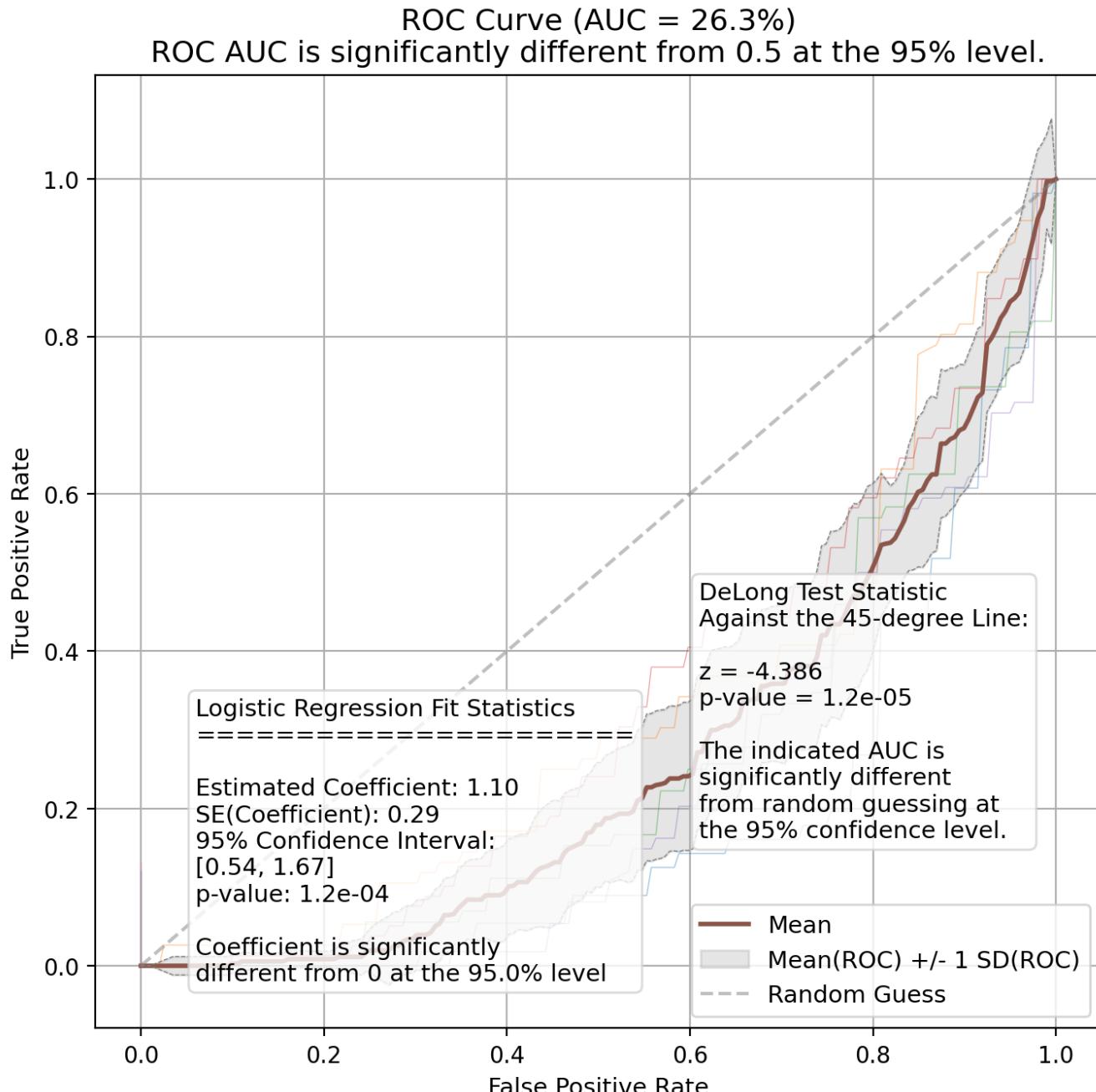
Worst Symmetry - Empirical CDF Plot



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



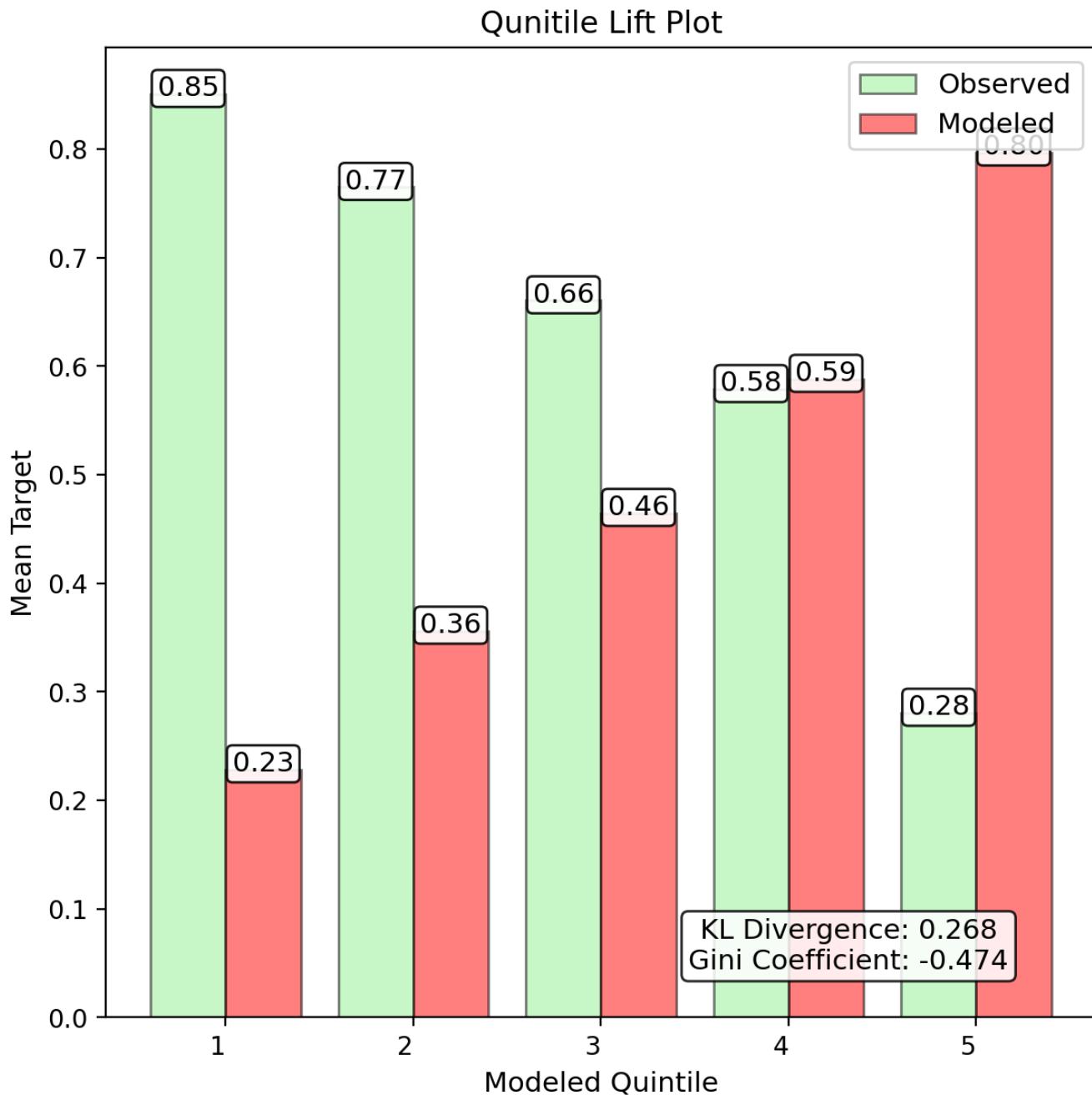
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

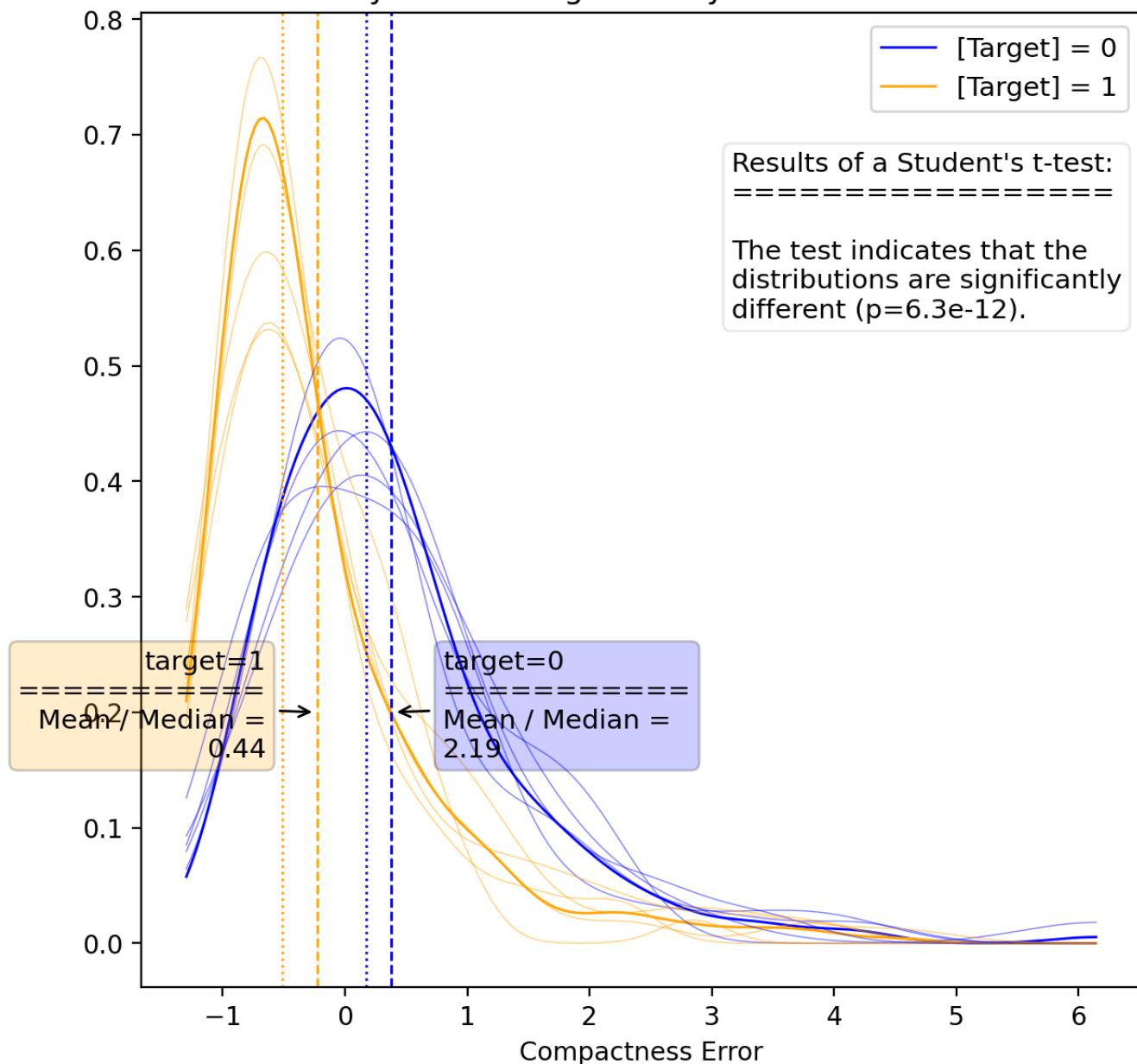
Compactness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-6.3e-01	-8.3e-01	-5.9e-01	-5.8e-01	-7.8e-01	2.9e+00	1.1e-01
Fitted p-Value	4.4e-08	3.4e-10	1.6e-07	6.3e-07	1.6e-09	2.9e-01	2.7e-07
Fitted Std. Err.	0.115	0.132	0.113	0.116	0.129	2.703	0.009
Conf. Int. Lower	-8.6e-01	-1.1e+00	-8.1e-01	-8.0e-01	-1.0e+00	-2.4e+00	1.3e-01
Conf. Int. Upper	-4.1e-01	-5.7e-01	-3.7e-01	-3.5e-01	-5.3e-01	8.2e+00	9.7e-02
Train Accuracy	68.6%	70.9%	68.7%	67.8%	69.2%	30.9%	1.2%
Val Accuracy	72.8%	63.1%	67.9%	71.0%	69.6%	30.9%	3.7%
Train AUC	66.6%	69.1%	66.6%	65.9%	67.3%	32.9%	1.2%
Val AUC	71.7%	61.4%	65.2%	68.4%	68.2%	32.9%	3.9%
Train F1	74.8%	76.7%	74.9%	74.2%	75.2%	31.4%	0.9%
Test F1	77.5%	69.8%	75.2%	77.1%	75.5%	31.4%	3.1%
Train Precision	75.6%	77.3%	74.4%	75.5%	75.4%	41.7%	1.0%
Val Precision	78.2%	71.2%	76.8%	73.6%	78.3%	41.7%	3.1%
Train Recall	74.1%	76.2%	75.4%	73.0%	74.9%	25.2%	1.2%
Val Recall	76.8%	68.4%	73.6%	81.0%	73.0%	25.2%	4.7%
Train MCC	33.0%	38.0%	33.3%	31.4%	34.5%	-34.1%	2.5%
Val MCC	43.2%	22.5%	29.8%	38.1%	35.6%	-34.1%	8.0%
Train Log-Loss	11.33	10.48	11.28	11.60	11.11	24.89	0.42
Val Log-Loss	9.79	13.29	11.57	10.46	10.97	24.89	1.33

Univariate Report

Compactness Error - Kernel Density Plot

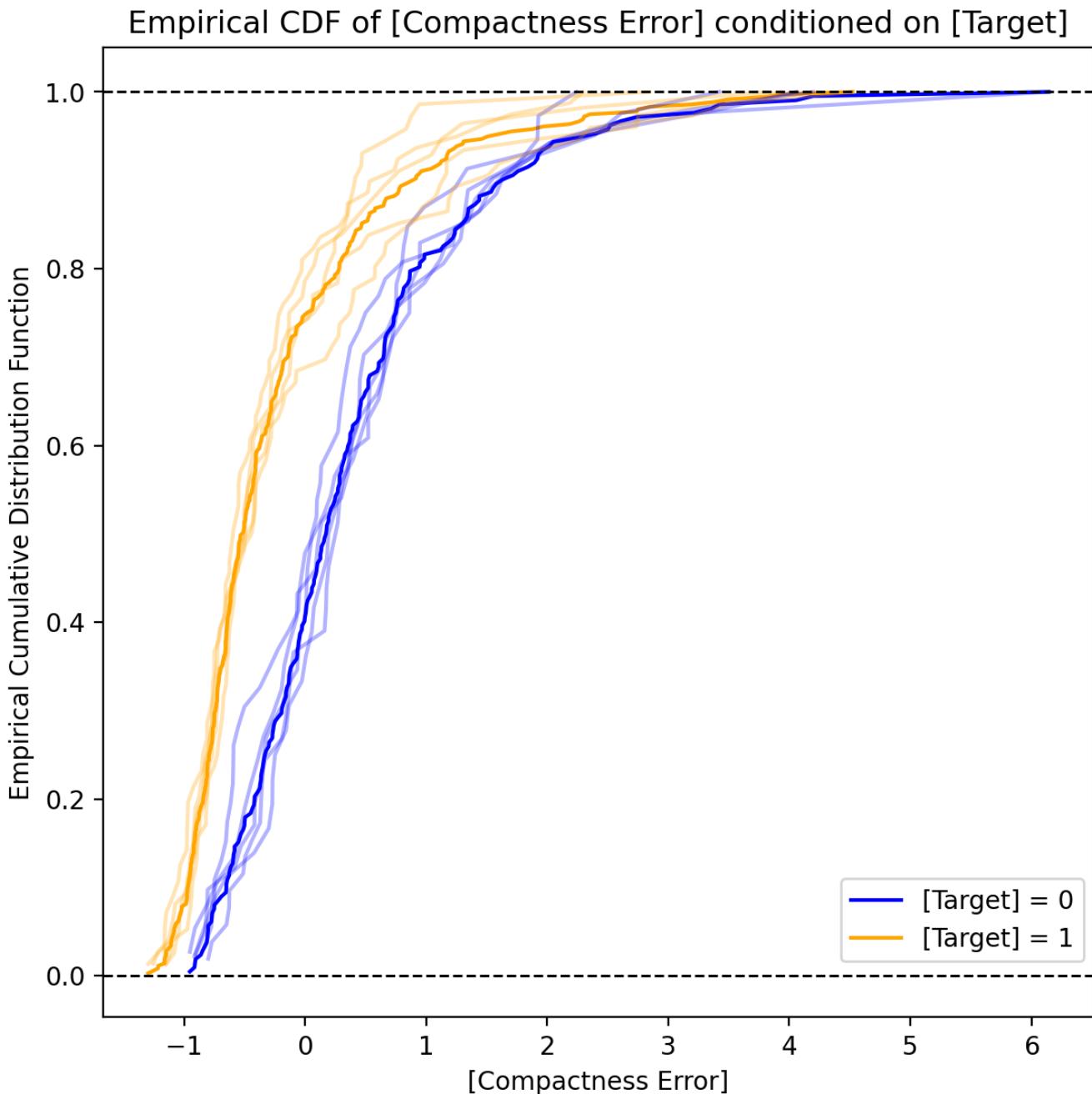
Kernel Density Plot of [Compactness Error] by [Target]
Distributions by level are significantly different at the 95% level.



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

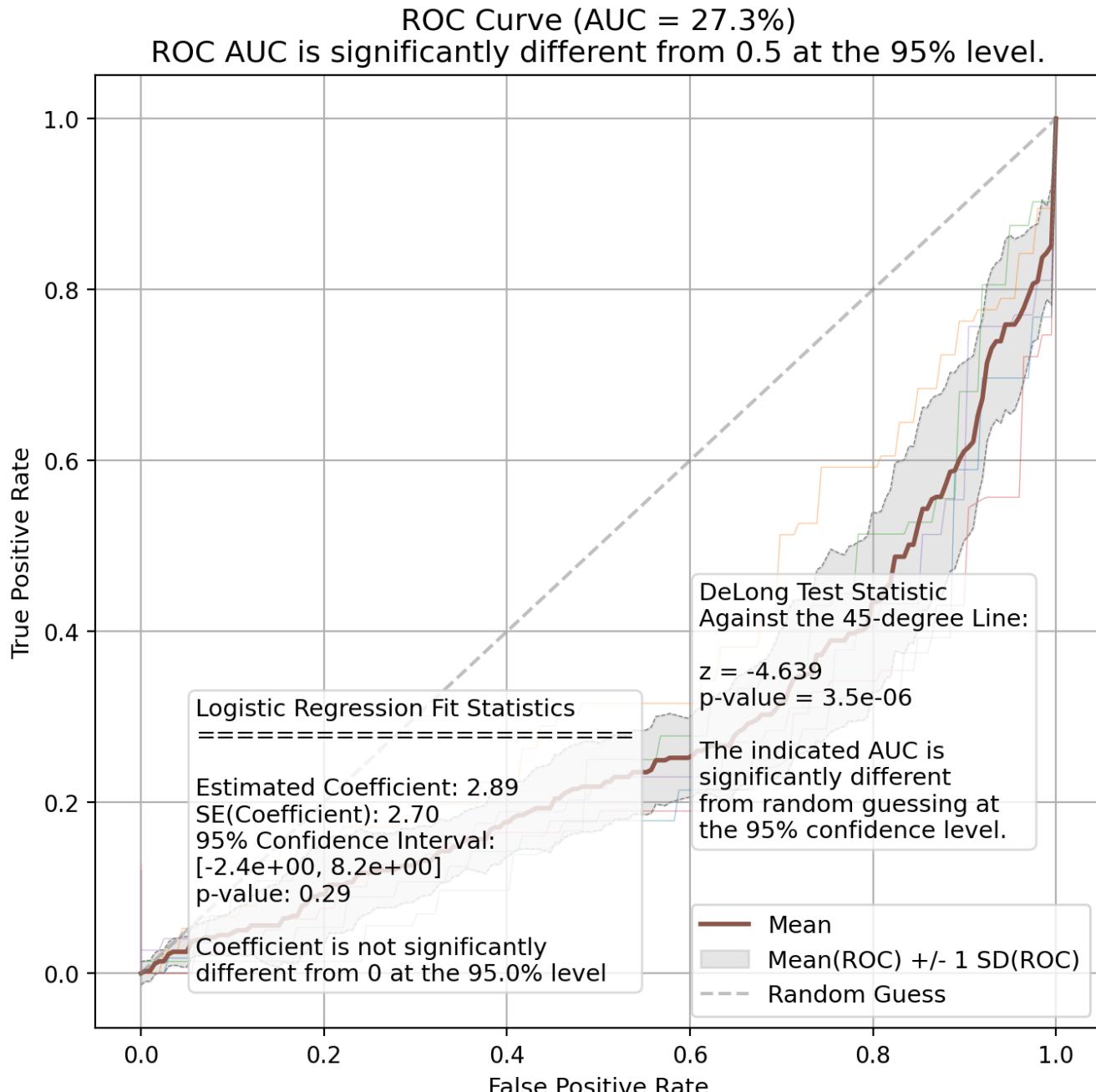
Compactness Error - Empirical CDF Plot



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



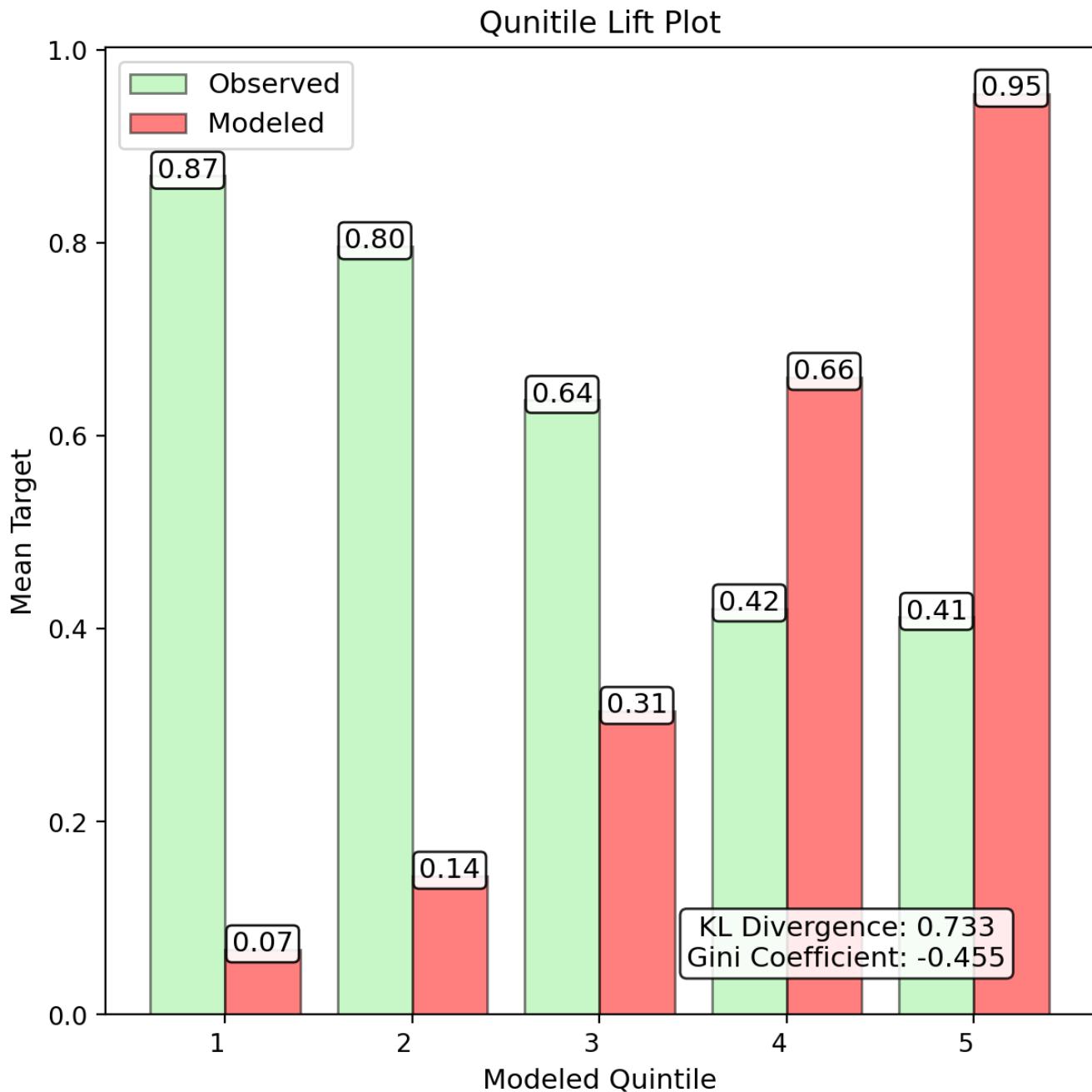
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

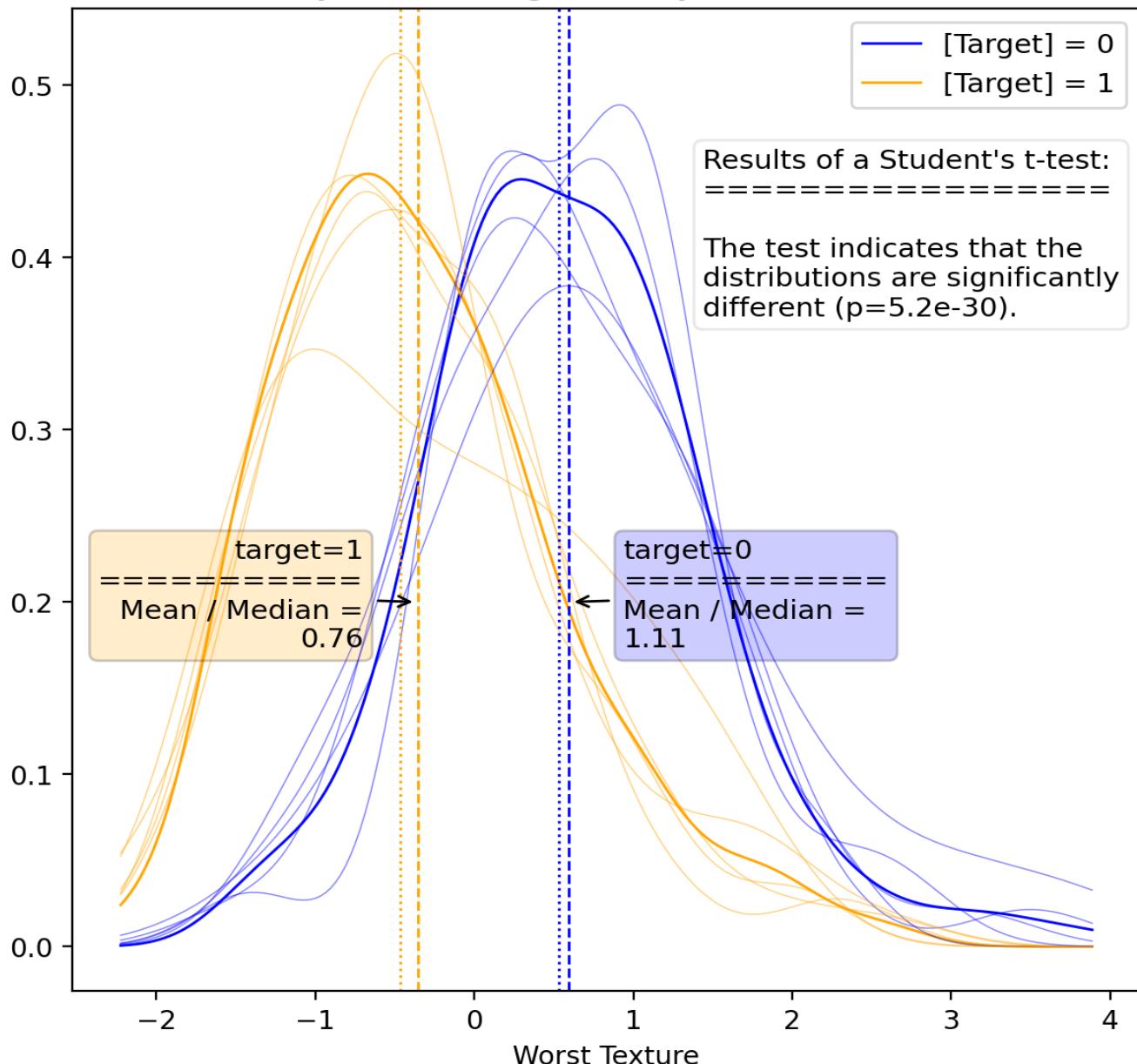
Worst Texture - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.1e+00	-1.0e+00	-1.1e+00	-1.1e+00	-1.0e+00	1.1e-02	4.7e-02
Fitted p-Value	1.1e-18	1.7e-16	1.6e-17	4.8e-17	6.7e-17	7.3e-04	6.6e-17
Fitted Std. Err.	0.129	0.124	0.126	0.130	0.124	0.003	0.003
Conf. Int. Lower	-1.4e+00	-1.3e+00	-1.3e+00	-1.3e+00	-1.3e+00	4.6e-03	5.2e-02
Conf. Int. Upper	-8.9e-01	-7.8e-01	-8.3e-01	-8.3e-01	-7.9e-01	1.7e-02	4.3e-02
Train Accuracy	71.7%	71.4%	70.9%	72.8%	71.8%	28.1%	0.7%
Val Accuracy	67.4%	74.6%	76.1%	68.7%	73.0%	28.1%	3.8%
Train AUC	72.7%	72.0%	71.4%	73.6%	72.8%	27.4%	0.8%
Val AUC	69.2%	74.9%	78.0%	69.4%	72.0%	27.4%	3.7%
Train F1	75.4%	75.3%	74.6%	76.8%	75.3%	34.6%	0.8%
Test F1	69.4%	78.3%	80.0%	71.7%	78.3%	34.6%	4.7%
Train Precision	83.5%	82.3%	81.1%	83.8%	83.0%	40.3%	1.1%
Val Precision	81.0%	83.6%	89.7%	78.8%	81.2%	40.3%	4.2%
Train Recall	68.8%	69.4%	69.1%	70.9%	68.9%	30.3%	0.8%
Val Recall	60.7%	73.7%	72.2%	65.8%	75.7%	30.3%	6.2%
Train MCC	43.9%	42.7%	41.7%	45.5%	44.1%	-43.8%	1.5%
Val MCC	37.7%	48.5%	53.1%	38.1%	43.0%	-43.8%	6.7%
Train Log-Loss	10.20	10.32	10.50	9.79	10.16	25.91	0.26
Val Log-Loss	11.75	9.16	8.60	11.28	9.72	25.91	1.36

Univariate Report

Worst Texture - Kernel Density Plot

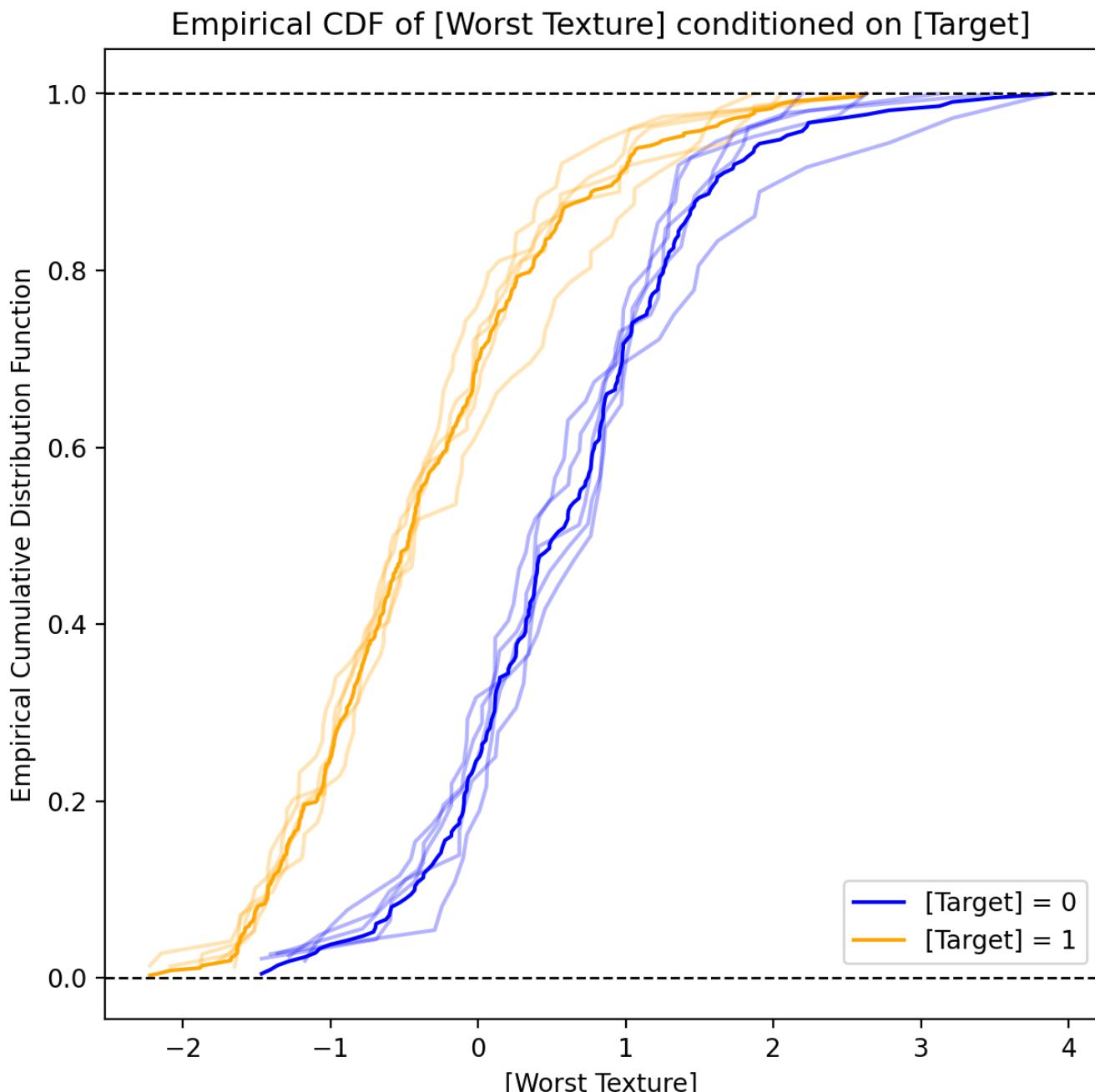
Kernel Density Plot of [Worst Texture] by [Target]
Distributions by level are significantly different at the 95% level.



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

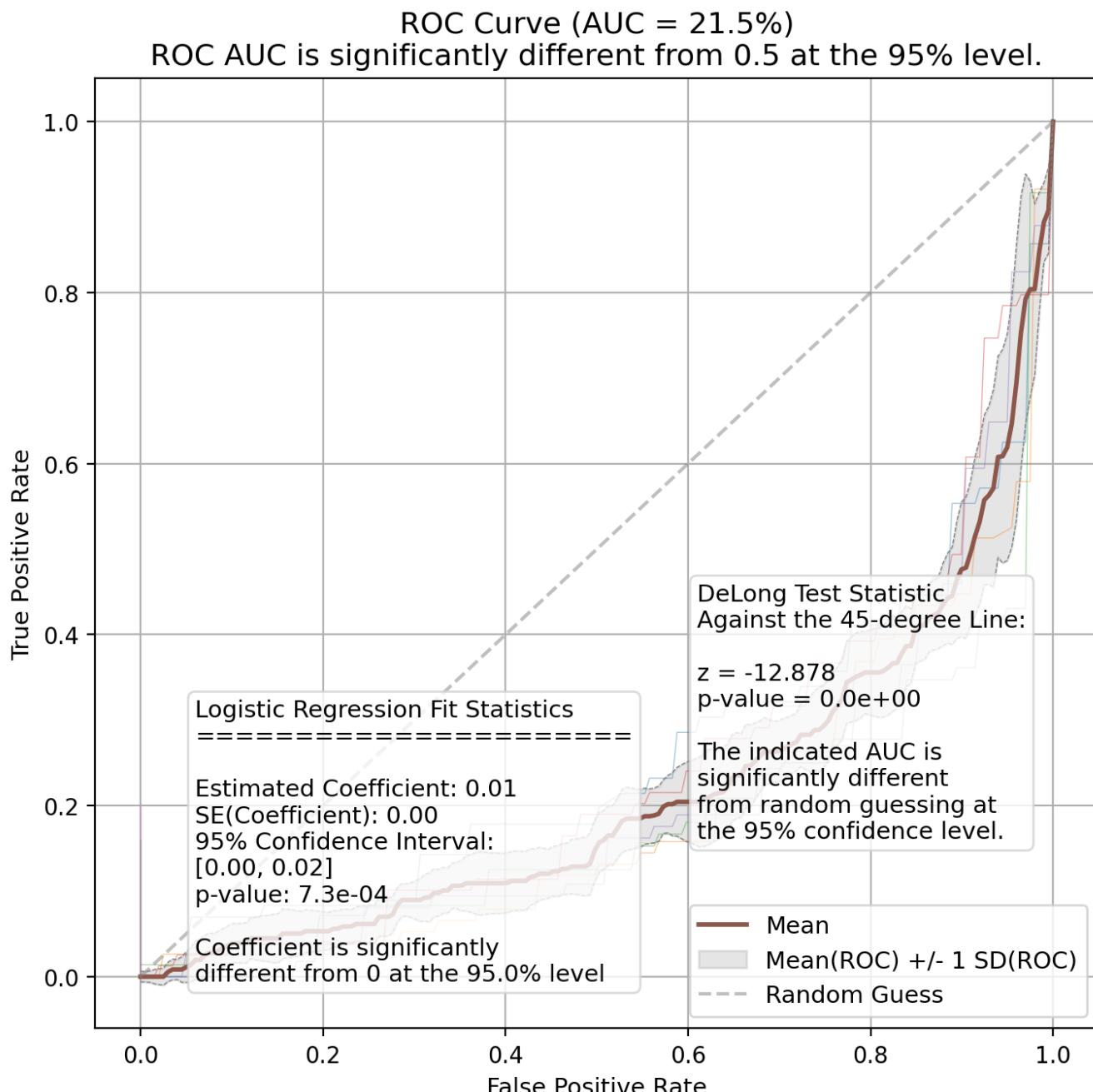
Worst Texture - Empirical CDF Plot



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



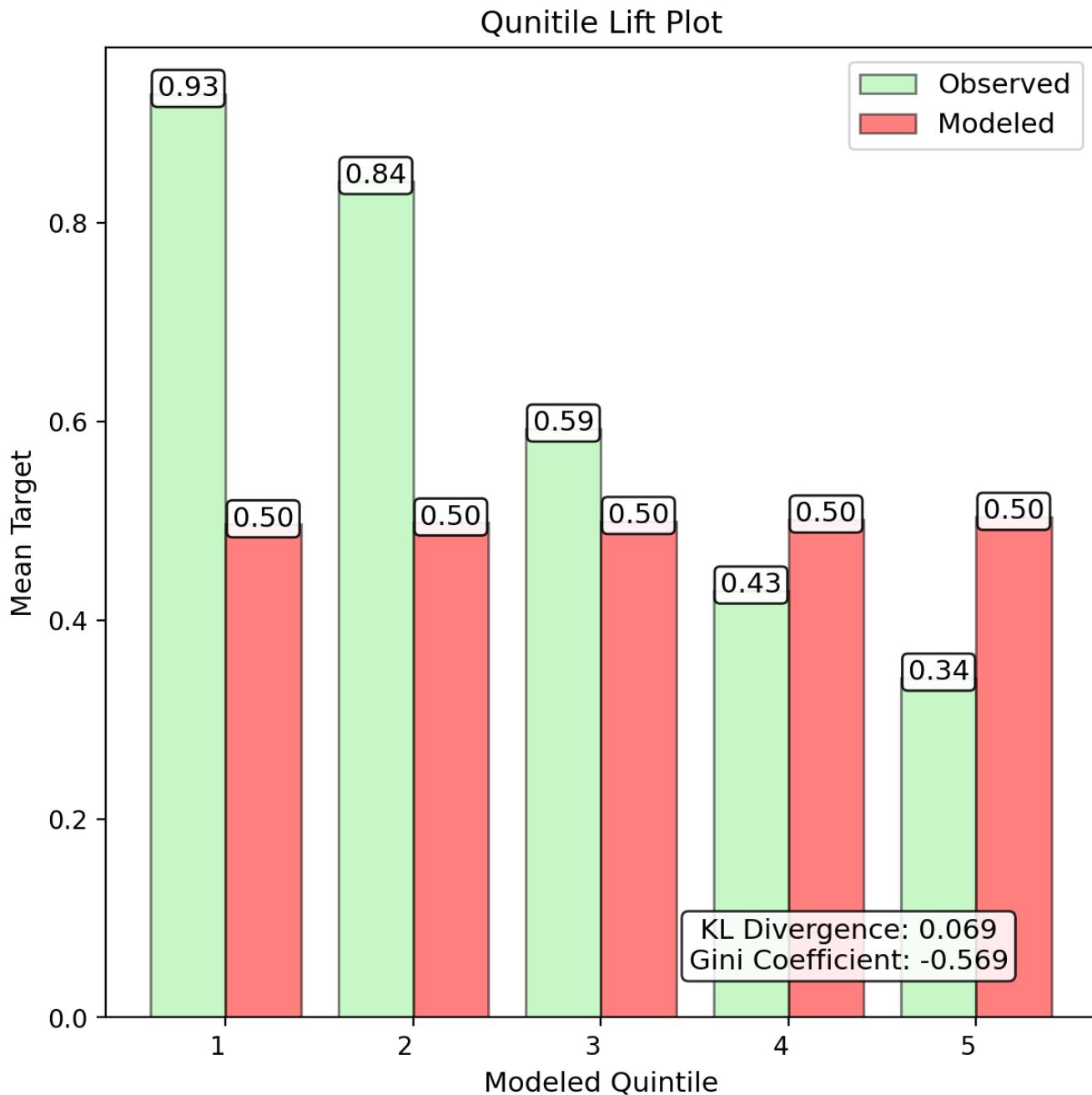
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

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

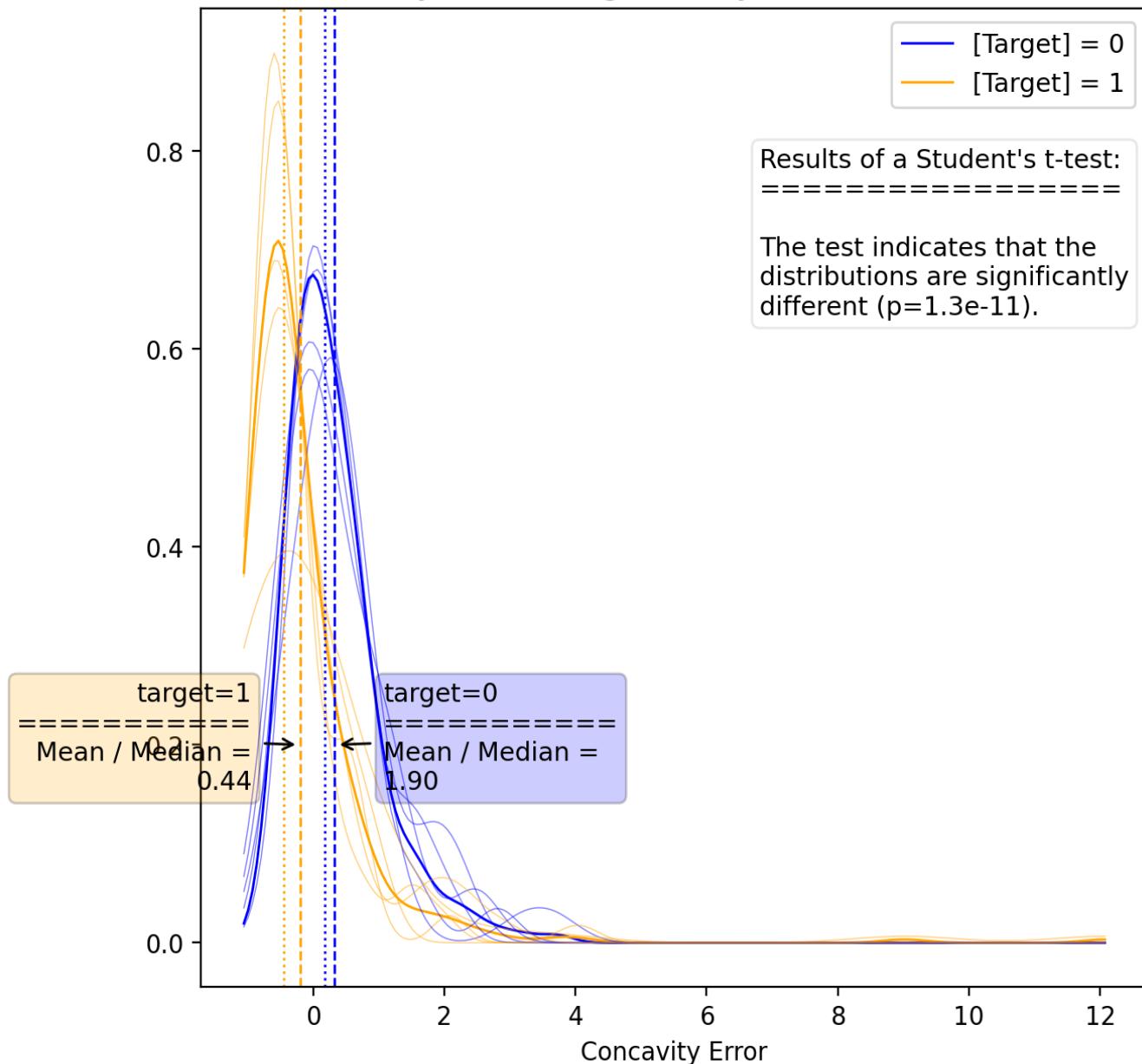
Concavity Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-6.5e-01	-1.3e+00	-6.7e-01	-5.7e-01	-8.1e-01	7.4e-01	3.1e-01
Fitted p-Value	1.8e-06	1.7e-14	1.9e-06	3.4e-05	1.1e-07	7.0e-01	1.5e-05
Fitted Std. Err.	0.136	0.176	0.141	0.138	0.153	1.915	0.017
Conf. Int. Lower	-9.1e-01	-1.7e+00	-9.5e-01	-8.4e-01	-1.1e+00	-3.0e+00	3.5e-01
Conf. Int. Upper	-3.8e-01	-1.0e+00	-4.0e-01	-3.0e-01	-5.1e-01	4.5e+00	2.8e-01
Train Accuracy	70.2%	73.6%	70.7%	69.6%	71.8%	29.0%	1.6%
Val Accuracy	75.0%	63.1%	71.6%	75.6%	69.6%	29.0%	5.0%
Train AUC	68.0%	72.6%	68.4%	67.6%	70.1%	31.0%	2.1%
Val AUC	73.0%	61.4%	71.2%	73.2%	67.1%	31.0%	5.0%
Train F1	76.4%	78.5%	76.7%	75.9%	77.3%	29.1%	1.0%
Test F1	80.0%	69.8%	77.0%	80.7%	76.2%	29.1%	4.3%
Train Precision	76.4%	80.5%	75.5%	76.6%	77.6%	39.0%	1.9%
Val Precision	78.0%	71.2%	82.5%	77.0%	76.7%	39.0%	4.0%
Train Recall	76.4%	76.5%	77.9%	75.2%	77.0%	23.2%	1.0%
Val Recall	82.1%	68.4%	72.2%	84.8%	75.7%	23.2%	6.8%
Train MCC	36.1%	44.5%	37.2%	35.0%	40.1%	-38.0%	3.8%
Val MCC	46.8%	22.5%	40.7%	48.0%	34.0%	-38.0%	10.5%
Train Log-Loss	10.73	9.51	10.58	10.94	10.16	25.59	0.56
Val Log-Loss	9.01	13.29	10.25	8.80	10.97	25.59	1.82

Univariate Report

Concavity Error - Kernel Density Plot

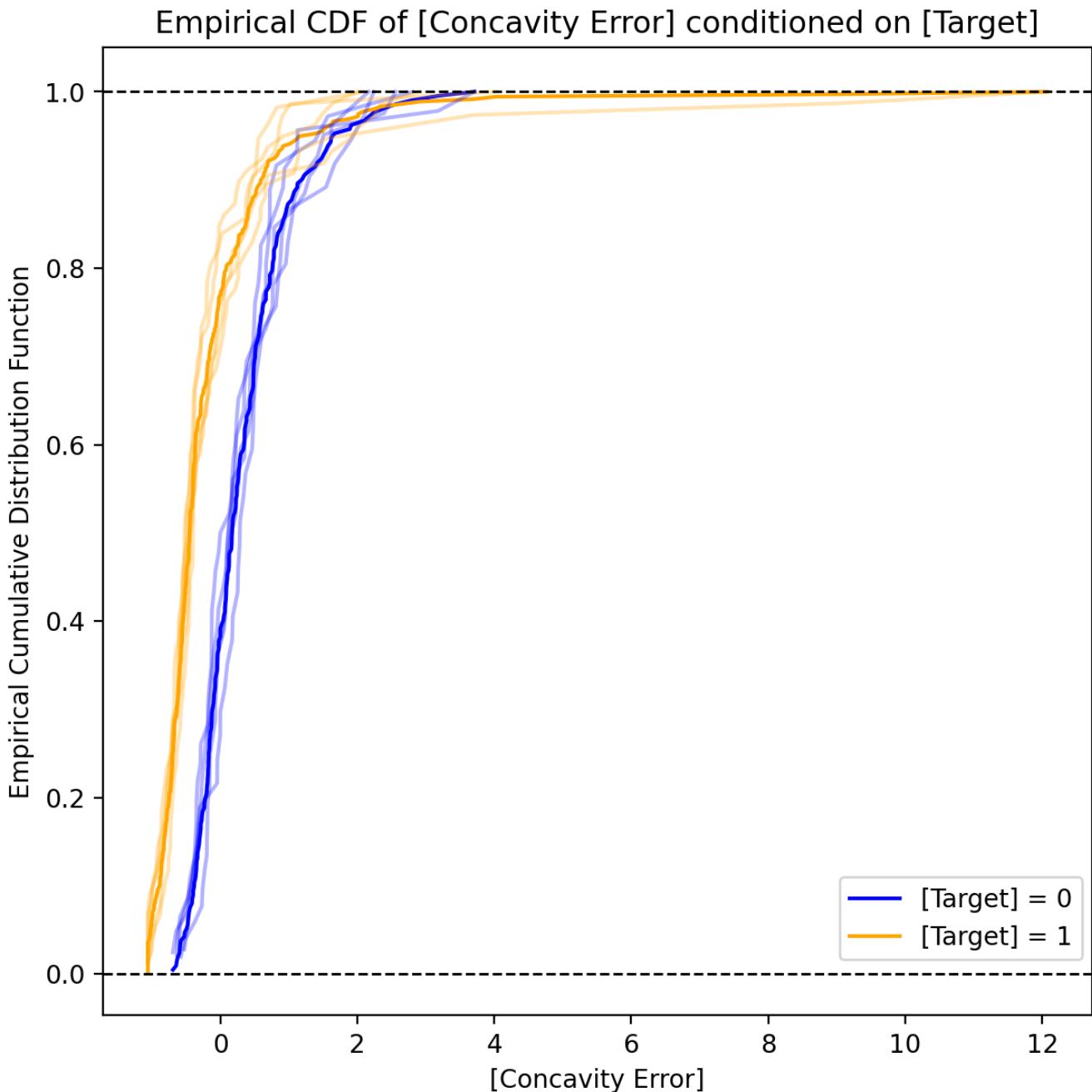
Kernel Density Plot of [Concavity Error] by [Target]
Distributions by level are significantly different at the 95% level.



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

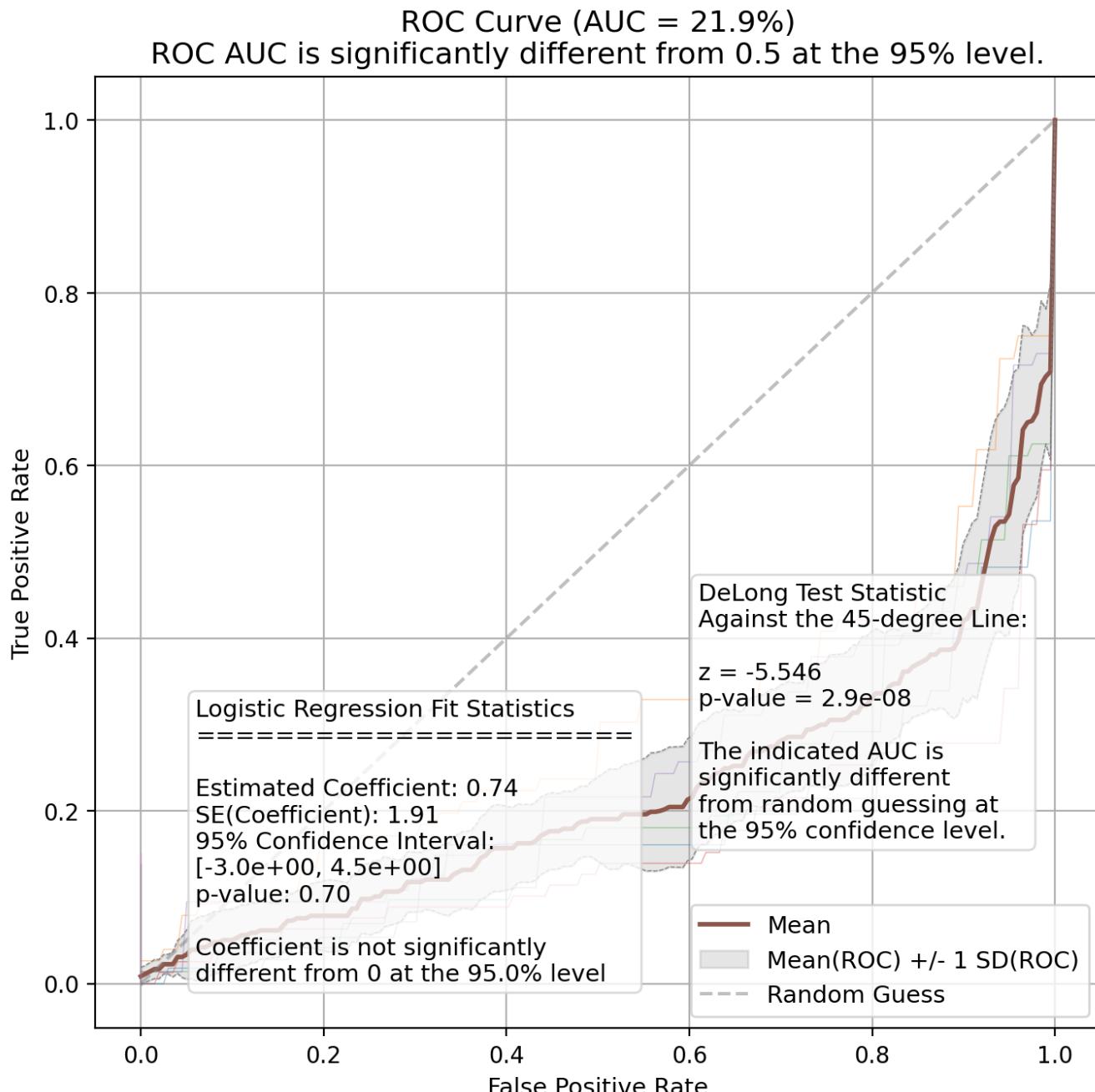
Concavity Error - Empirical CDF Plot



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

Concavity Error - ROC Curve



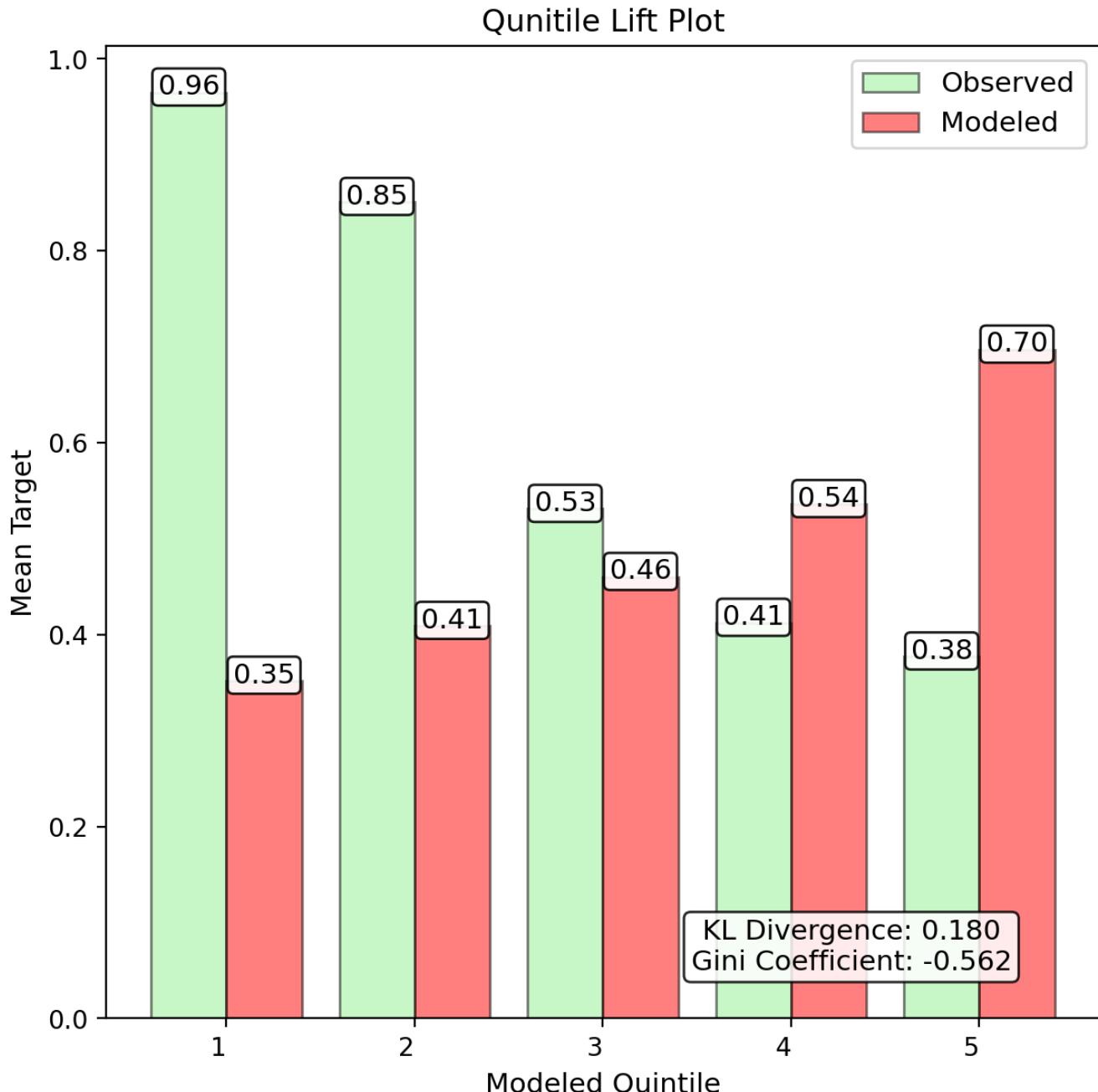
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Concavity 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

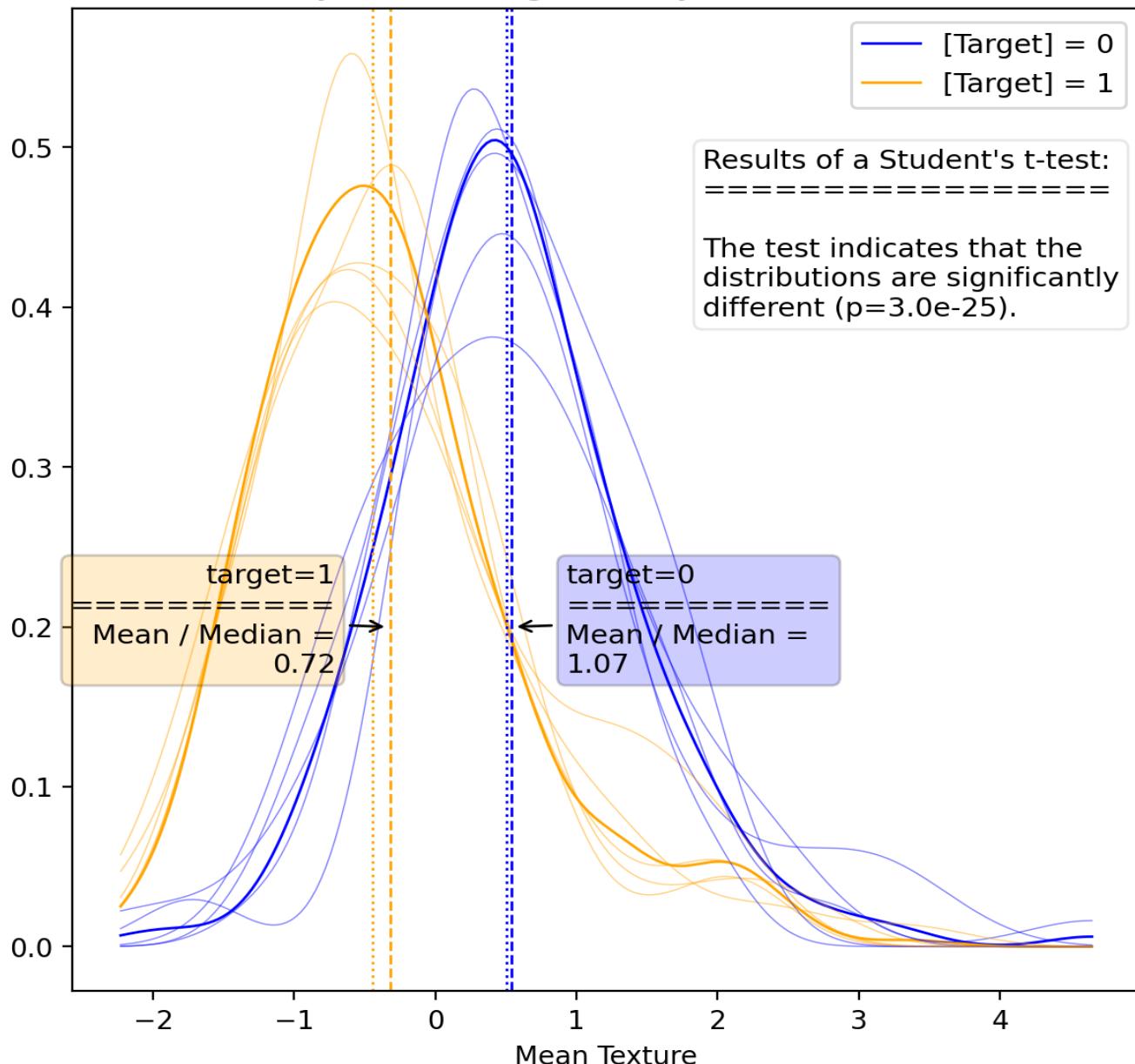
Mean Texture - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.0e+00	-8.7e-01	-9.3e-01	-9.8e-01	-9.6e-01	1.6e-02	6.3e-02
Fitted p-Value	9.8e-17	2.0e-13	1.8e-14	1.8e-14	3.5e-15	1.7e-04	8.4e-14
Fitted Std. Err.	0.126	0.119	0.121	0.128	0.121	0.004	0.004
Conf. Int. Lower	-1.3e+00	-1.1e+00	-1.2e+00	-1.2e+00	-1.2e+00	7.7e-03	6.9e-02
Conf. Int. Upper	-8.0e-01	-6.4e-01	-6.9e-01	-7.3e-01	-7.2e-01	2.5e-02	5.7e-02
Train Accuracy	73.4%	72.0%	73.0%	73.5%	73.1%	27.2%	0.6%
Val Accuracy	69.6%	77.9%	71.6%	70.2%	73.9%	27.2%	3.4%
Train AUC	73.8%	72.3%	73.4%	74.4%	73.5%	26.7%	0.7%
Val AUC	71.0%	77.9%	73.2%	70.1%	73.7%	26.7%	3.1%
Train F1	77.4%	76.2%	76.8%	77.3%	77.0%	33.2%	0.5%
Test F1	72.0%	81.4%	76.0%	74.2%	78.6%	33.2%	3.7%
Train Precision	83.5%	82.0%	82.3%	84.6%	82.6%	39.2%	1.1%
Val Precision	81.8%	85.5%	86.0%	77.8%	83.3%	39.2%	3.3%
Train Recall	72.1%	71.2%	71.9%	71.2%	72.1%	28.9%	0.5%
Val Recall	64.3%	77.6%	68.1%	70.9%	74.3%	28.9%	5.2%
Train MCC	46.2%	43.3%	45.6%	47.0%	45.7%	-45.2%	1.4%
Val MCC	41.1%	54.6%	44.0%	39.5%	46.0%	-45.2%	5.9%
Train Log-Loss	9.60	10.08	9.72	9.55	9.69	26.23	0.21
Val Log-Loss	10.97	7.98	10.25	10.73	9.40	26.23	1.21

Univariate Report

Mean Texture - Kernel Density Plot

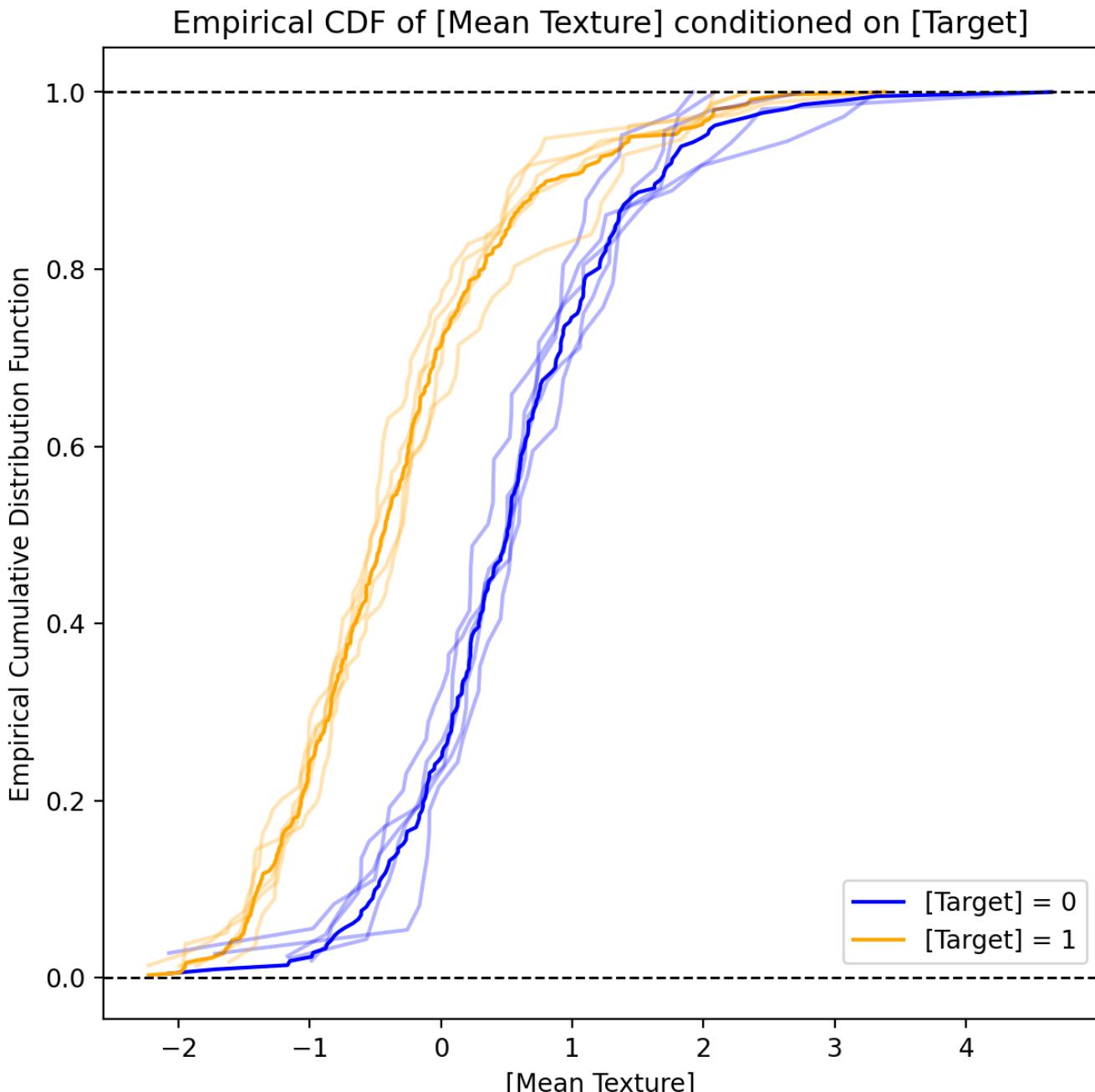
Kernel Density Plot of [Mean Texture] by [Target]
Distributions by level are significantly different at the 95% level.



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

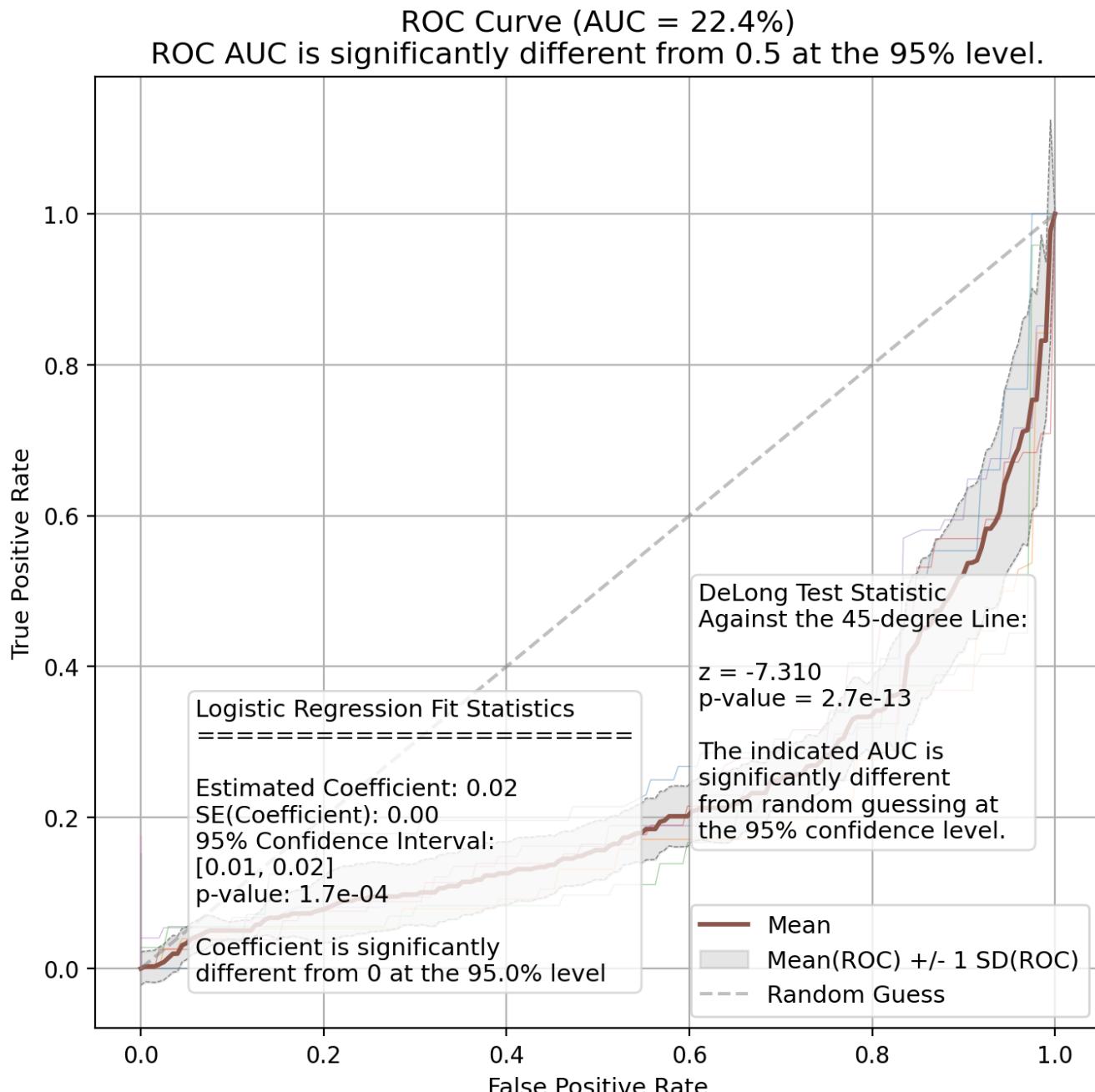
Mean Texture - Empirical CDF Plot



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

Mean Texture - ROC Curve



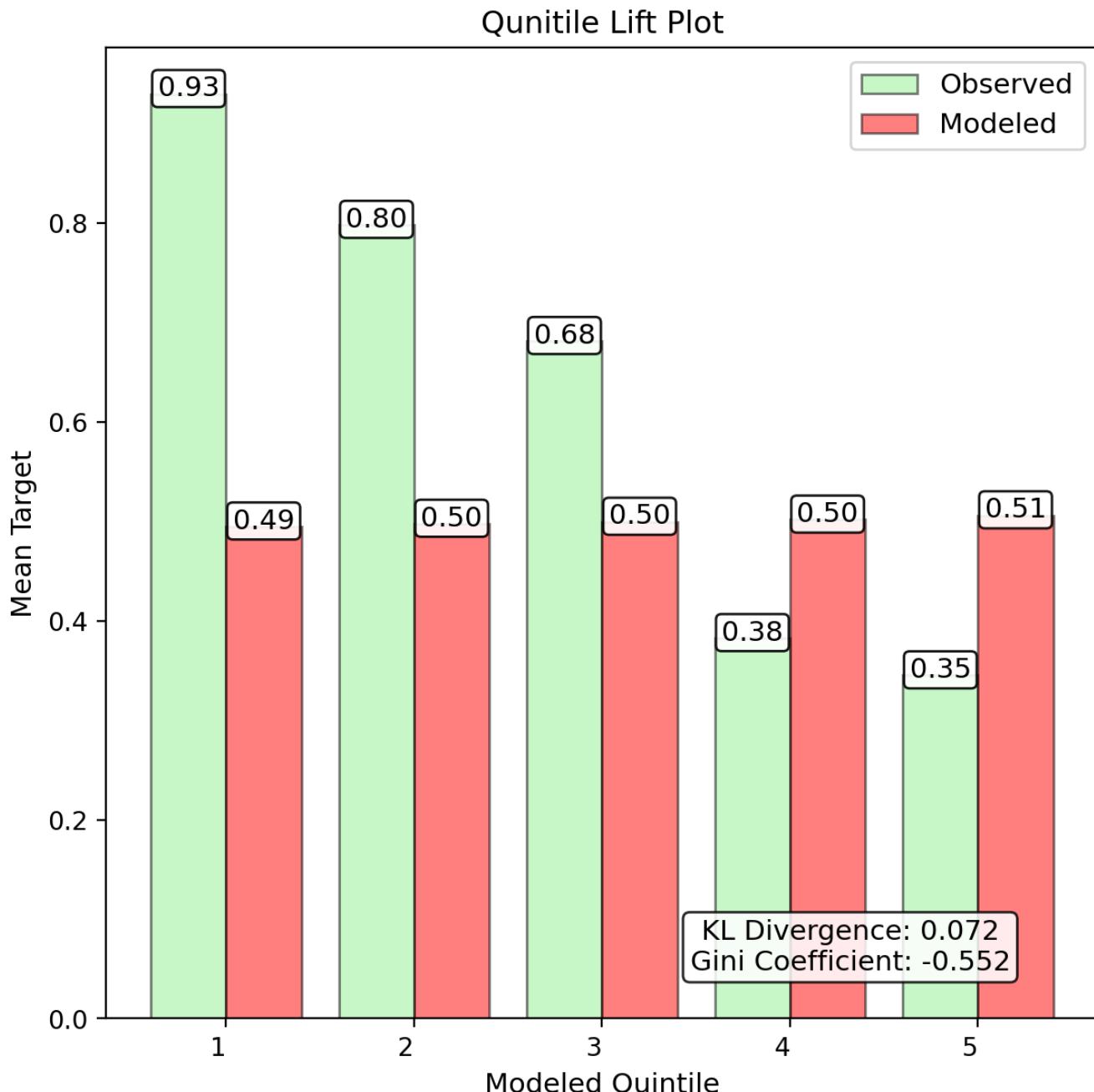
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean 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.