

# **Univariate Analysis Report**

## **Cancer Model**

2024-02-02

# Overview

## Cancer Model Univariate Analysis Report

These sorted results for the features in this report indicate the average cross-validated test scores for each feature, if it were used as the only predictor in a simple linear model. Keep in mind that these results are based on the average, without considering the standard deviation. This means that the results are not necessarily the best predictors, but they are the best on average, and provide a fine starting point for grouping those predictors that are on average better than others. This means that nothing was done to account for possible sampling variability in the sorted results. This is a limitation of the univariate analysis, 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.
<b>texture_error</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>worst_symmetry</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>worst_smoothness</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>worst_texture</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>fractal_dimension_error</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>symmetry_error</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>mean_texture</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>smoothness_error</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>worst_fractal_dimension</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>mean_smoothness</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>mean_symmetry</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>mean_fractal_dimension</b>	62.3%	62.3%	100.0%	50.0%	76.8%	0.00e+00	58.6%
<b>radius_error</b>	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
<b>mean_perimeter</b>	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
<b>worst_concave_points</b>	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%

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 16 of the variables passed to the Cancer Model.

## Univariate Report

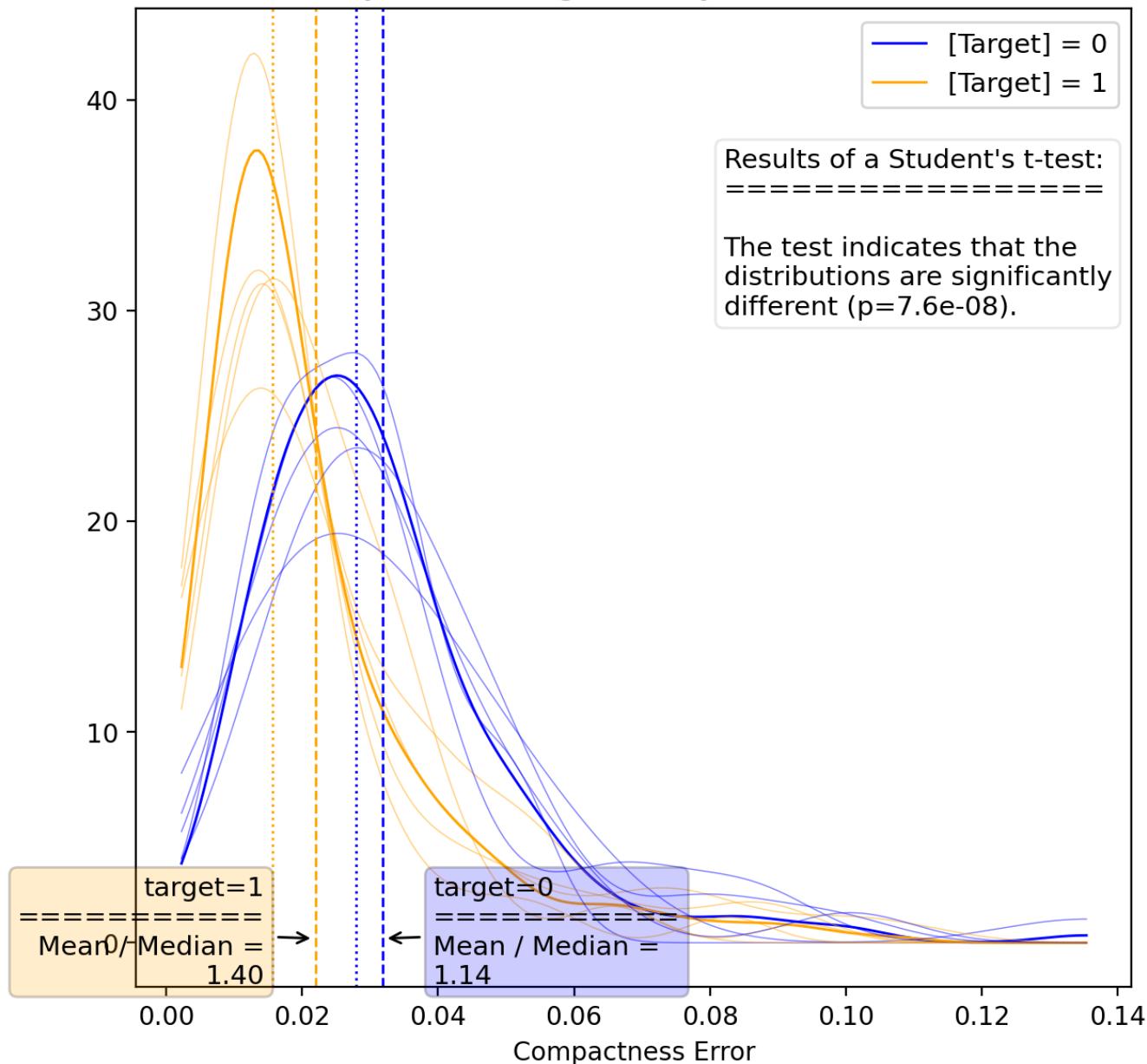
### Compactness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	3.95	1.35	3.34	6.51	4.68	3.98	1.88
<b>Fitted p-Value</b>	0.229	0.686	0.329	0.057	0.148	0.183	0.243
<b>Fitted Std. Err.</b>	3.28	3.36	3.43	3.42	3.23	2.98	0.09
<b>Conf. Int. Lower</b>	-2.5e+00	-5.2e+00	-3.4e+00	-1.9e-01	-1.7e+00	-1.9e+00	1.9e+00
<b>Conf. Int. Upper</b>	10.39	7.93	10.07	13.21	11.01	9.83	1.90
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

Compactness Error - Kernel Density Plot

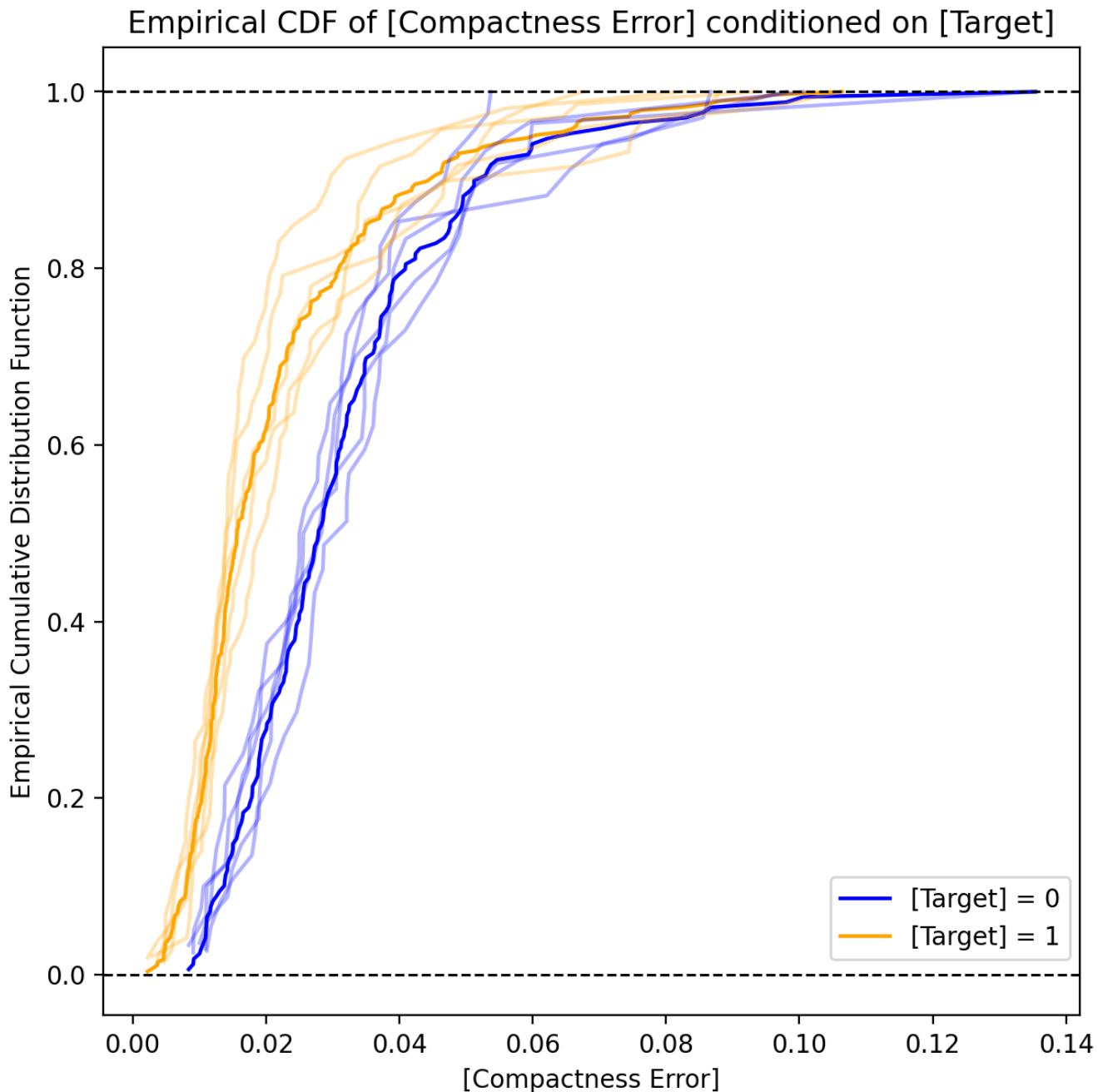
Kernel Density Plot of [Compactness Error] by [Target]  
Distributions by level are significantly different at the 95% level.



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

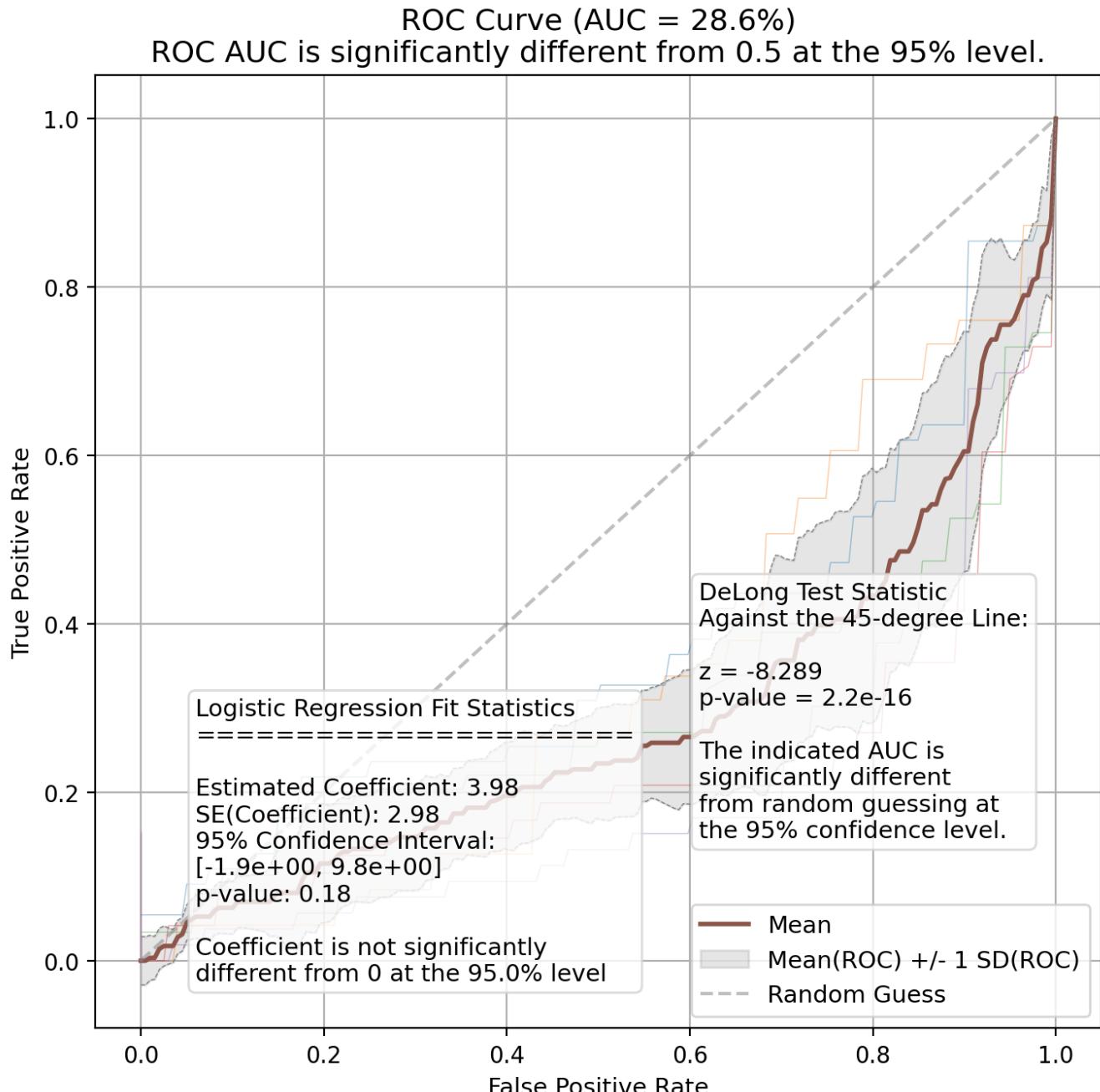
### Compactness Error - Empirical CDF Plot



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



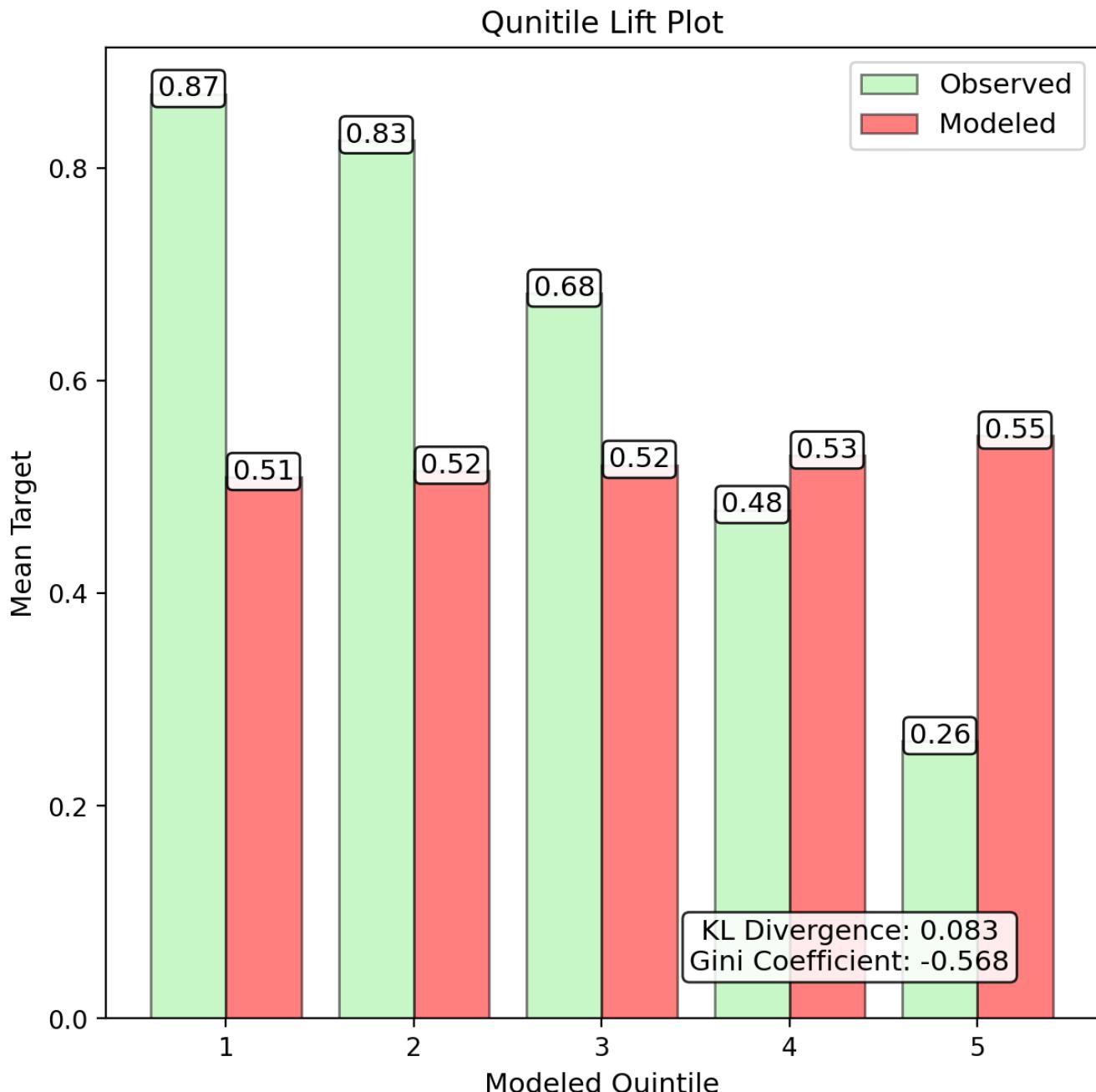
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

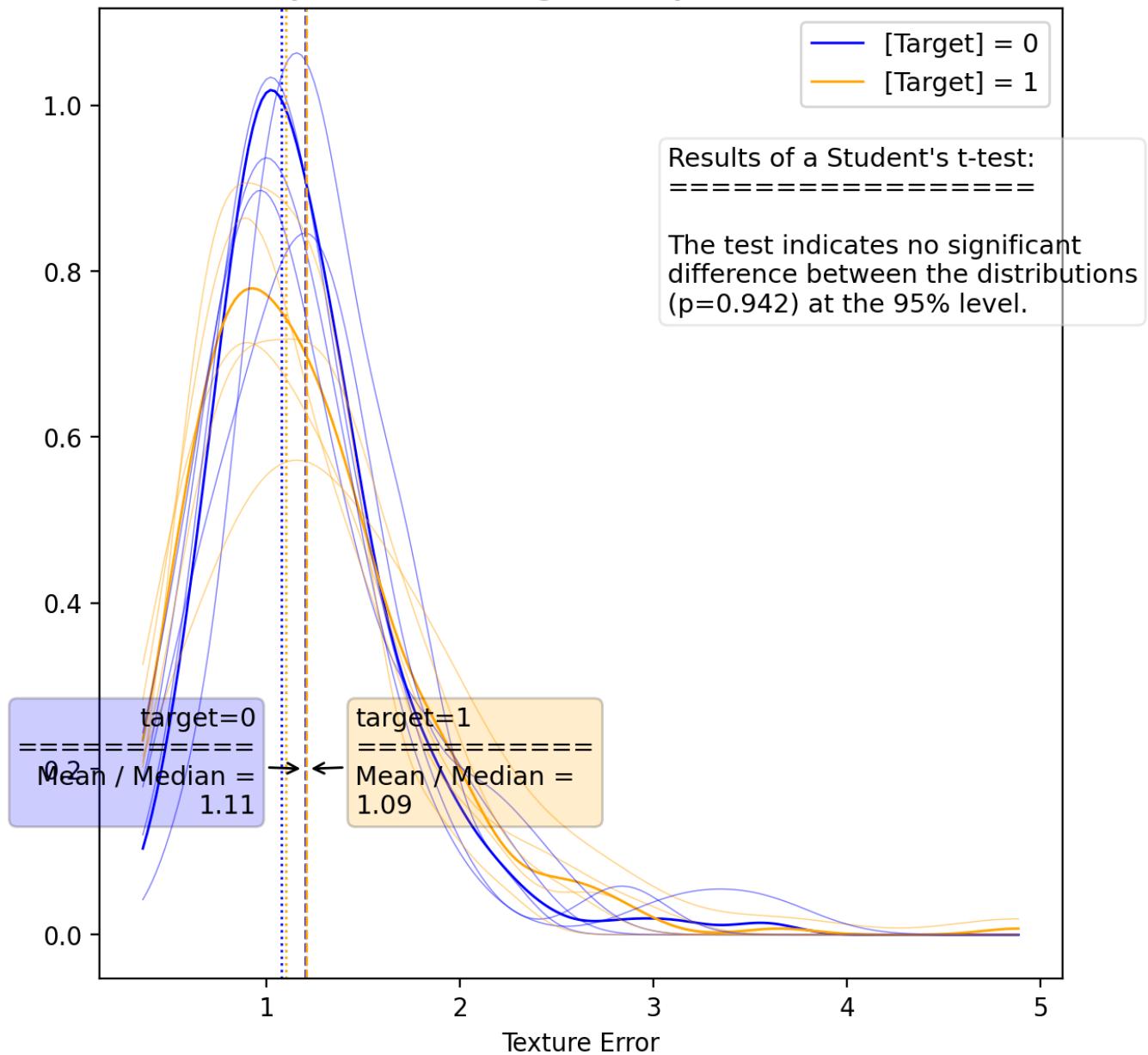
### Texture Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	0.393	0.308	0.350	0.409	0.363	0.364	0.039
<b>Fitted p-Value</b>	3.5e-06	2.0e-04	6.2e-05	1.4e-06	9.3e-06	1.3e-06	8.6e-05
<b>Fitted Std. Err.</b>	8.5e-02	8.3e-02	8.7e-02	8.5e-02	8.2e-02	7.5e-02	2.1e-03
<b>Conf. Int. Lower</b>	0.227	0.146	0.179	0.243	0.202	0.217	0.039
<b>Conf. Int. Upper</b>	0.559	0.471	0.521	0.574	0.523	0.512	0.040
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

### Texture Error - Kernel Density Plot

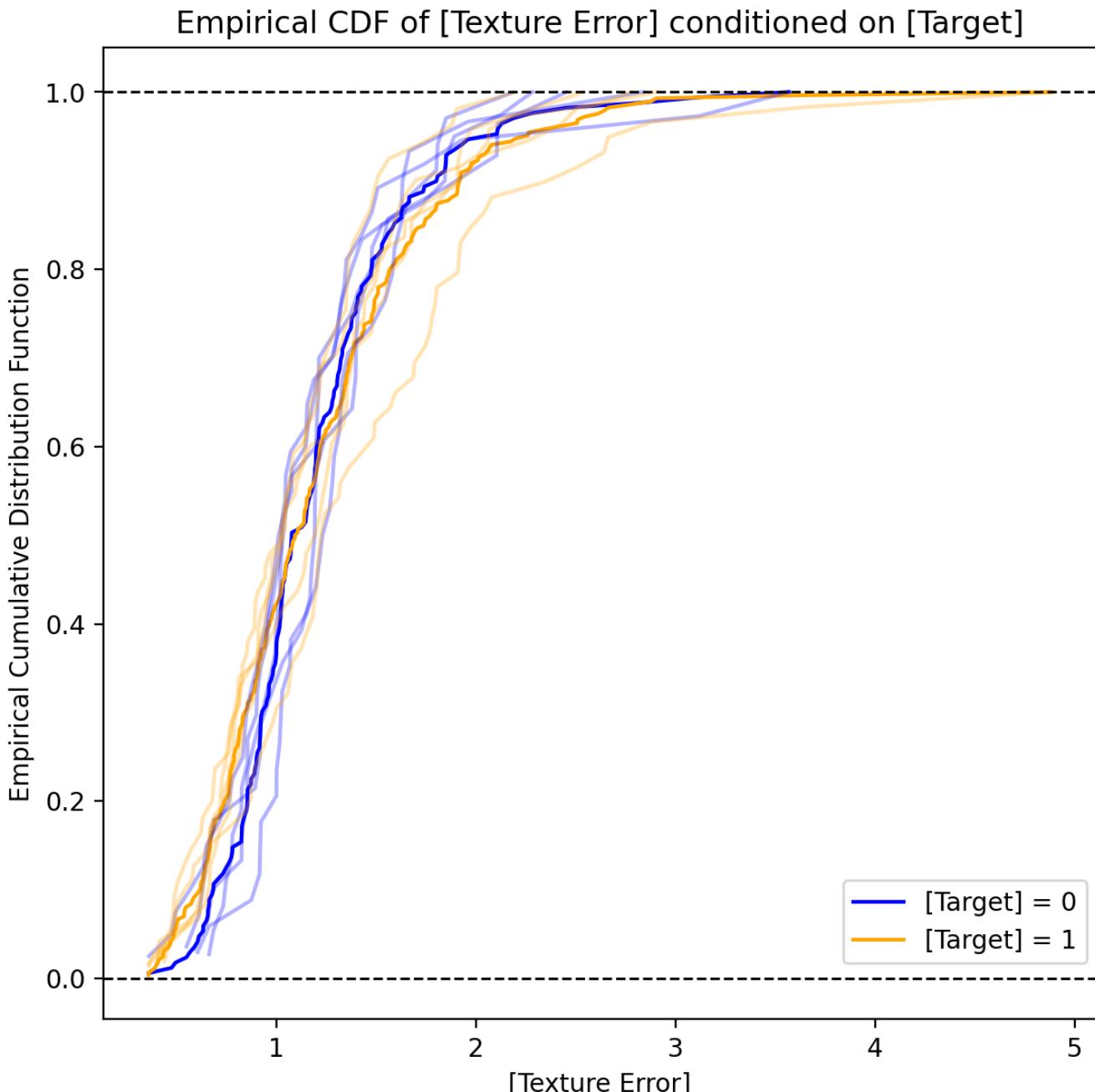
Kernel Density Plot of [Texture Error] by [Target]  
Distributions by level are not significantly different at the 95% level.



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

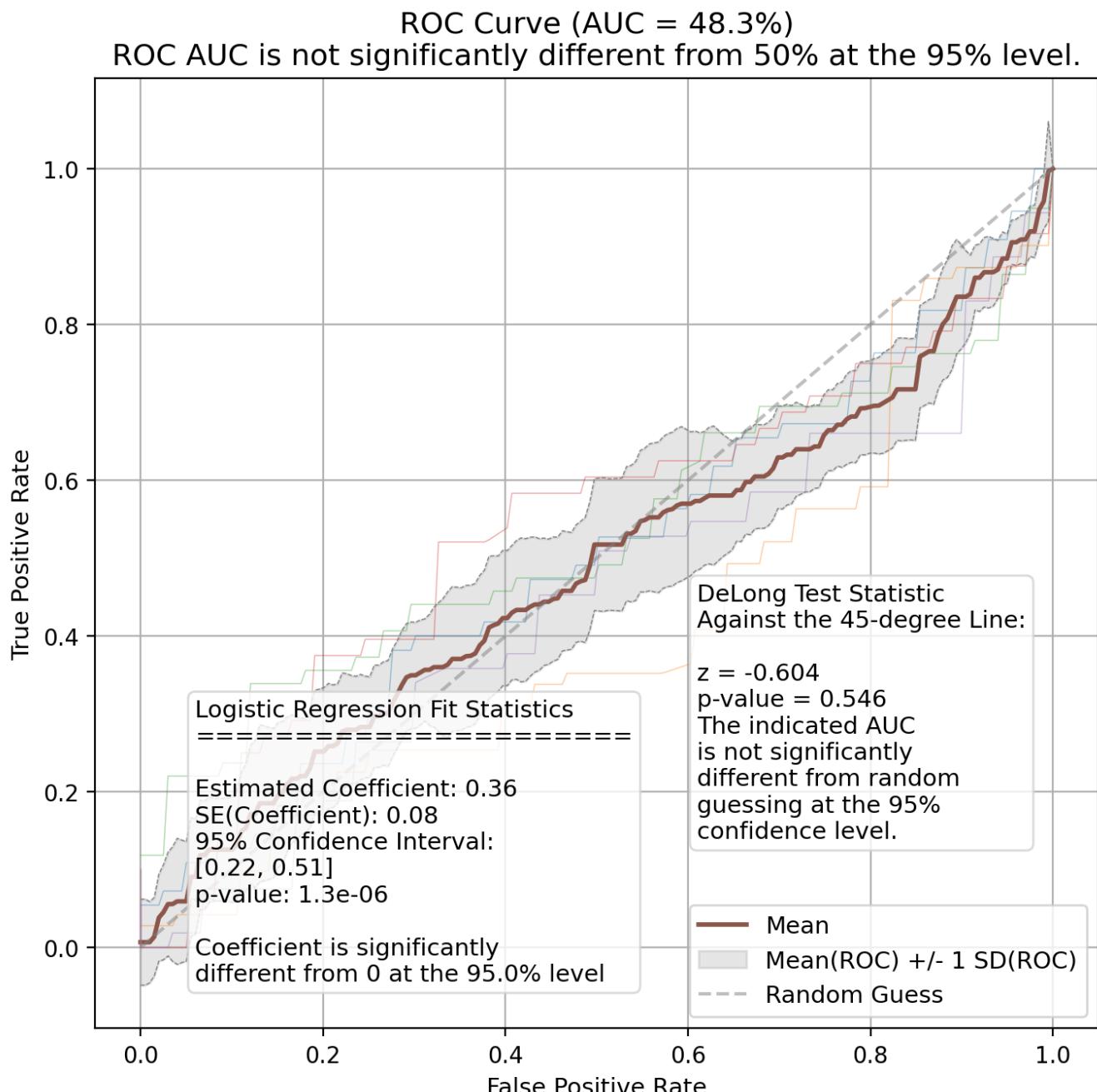
### Texture Error - Empirical CDF Plot



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



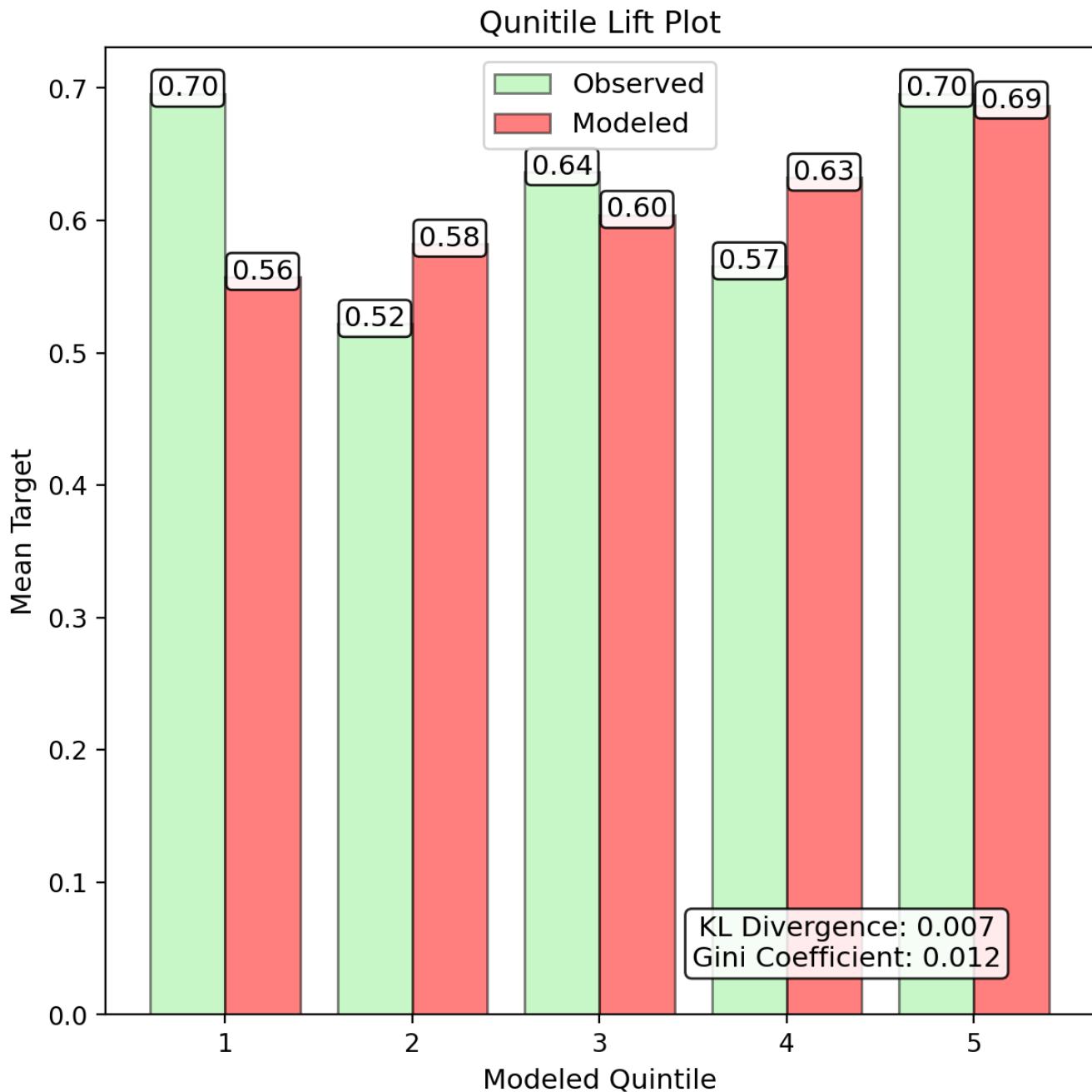
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 +/- 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

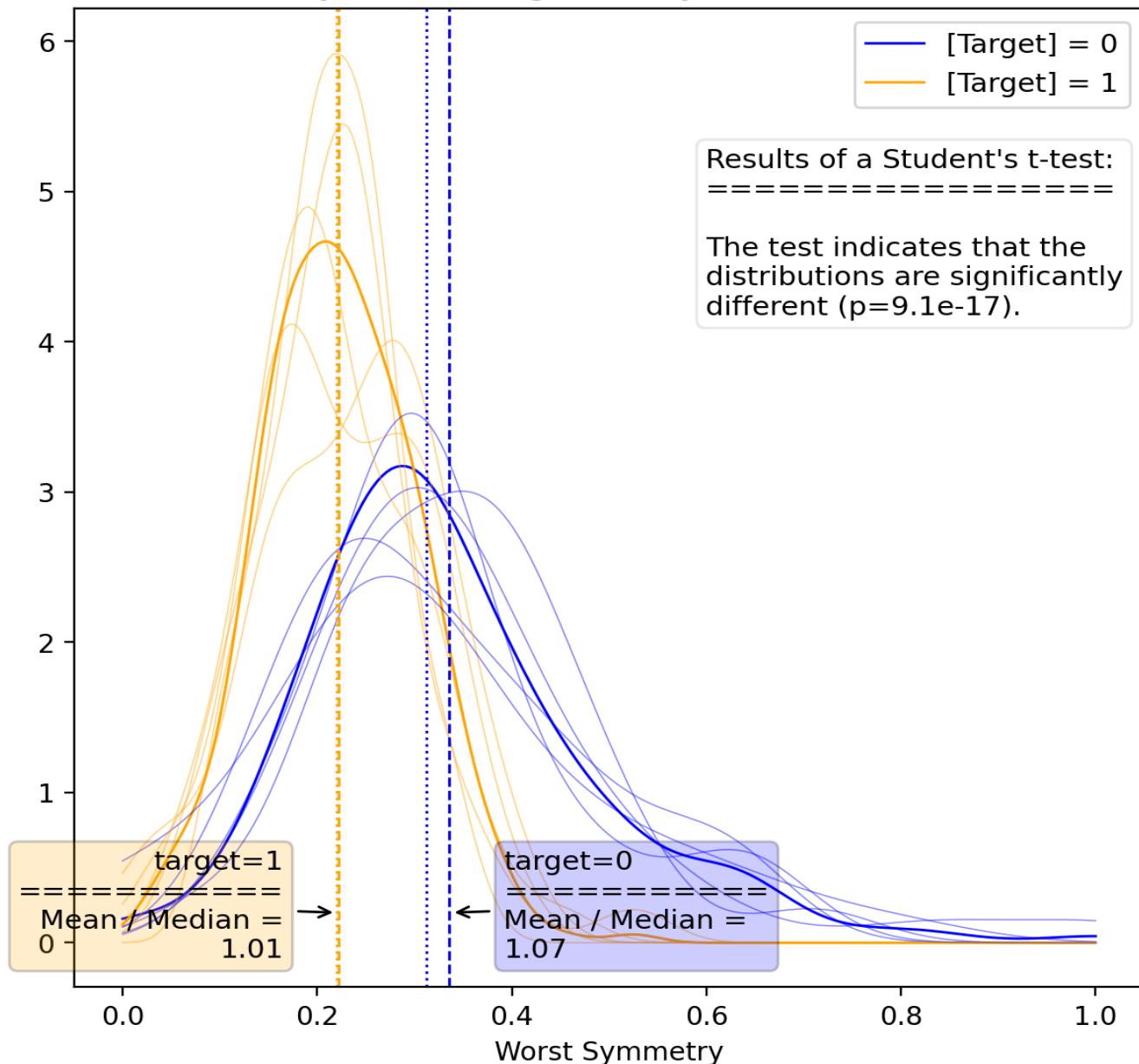
### Worst Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	0.466	0.067	0.270	0.611	0.330	0.346	0.205
<b>Fitted p-Value</b>	0.203	0.850	0.453	0.095	0.351	0.283	0.291
<b>Fitted Std. Err.</b>	0.366	0.357	0.359	0.366	0.354	0.322	0.005
<b>Conf. Int. Lower</b>	-2.5e-01	-6.3e-01	-4.3e-01	-1.1e-01	-3.6e-01	-2.9e-01	2.0e-01
<b>Conf. Int. Upper</b>	1.183	0.767	0.974	1.329	1.023	0.977	0.213
<b>Train Accuracy</b>	64.4%	60.7%	62.7%	64.6%	62.9%	63.1%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	64.5%	56.5%	63.9%	62.3%	6.1%
<b>Train AUC</b>	50.4%	50.4%	50.0%	50.4%	50.4%	50.3%	0.2%
<b>Val AUC</b>	50.0%	50.0%	51.5%	50.0%	50.0%	50.0%	0.7%
<b>Train F1</b>	78.3%	75.4%	77.1%	78.4%	77.2%	77.3%	1.2%
<b>Test F1</b>	73.3%	83.5%	78.1%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.3%	60.6%	62.7%	64.5%	62.8%	63.0%	1.6%
<b>Val Precision</b>	57.9%	71.7%	64.1%	56.5%	63.9%	62.3%	6.1%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	7.1%	6.6%	0.0%	7.0%	6.7%	6.1%	3.1%
<b>Val MCC</b>	0.0%	0.0%	13.7%	0.0%	0.0%	0.0%	6.1%
<b>Train Log-Loss</b>	12.82	14.17	13.44	12.76	13.37	13.31	0.57
<b>Val Log-Loss</b>	15.18	10.19	12.79	15.69	13.03	13.60	2.19

## Univariate Report

### Worst Symmetry - Kernel Density Plot

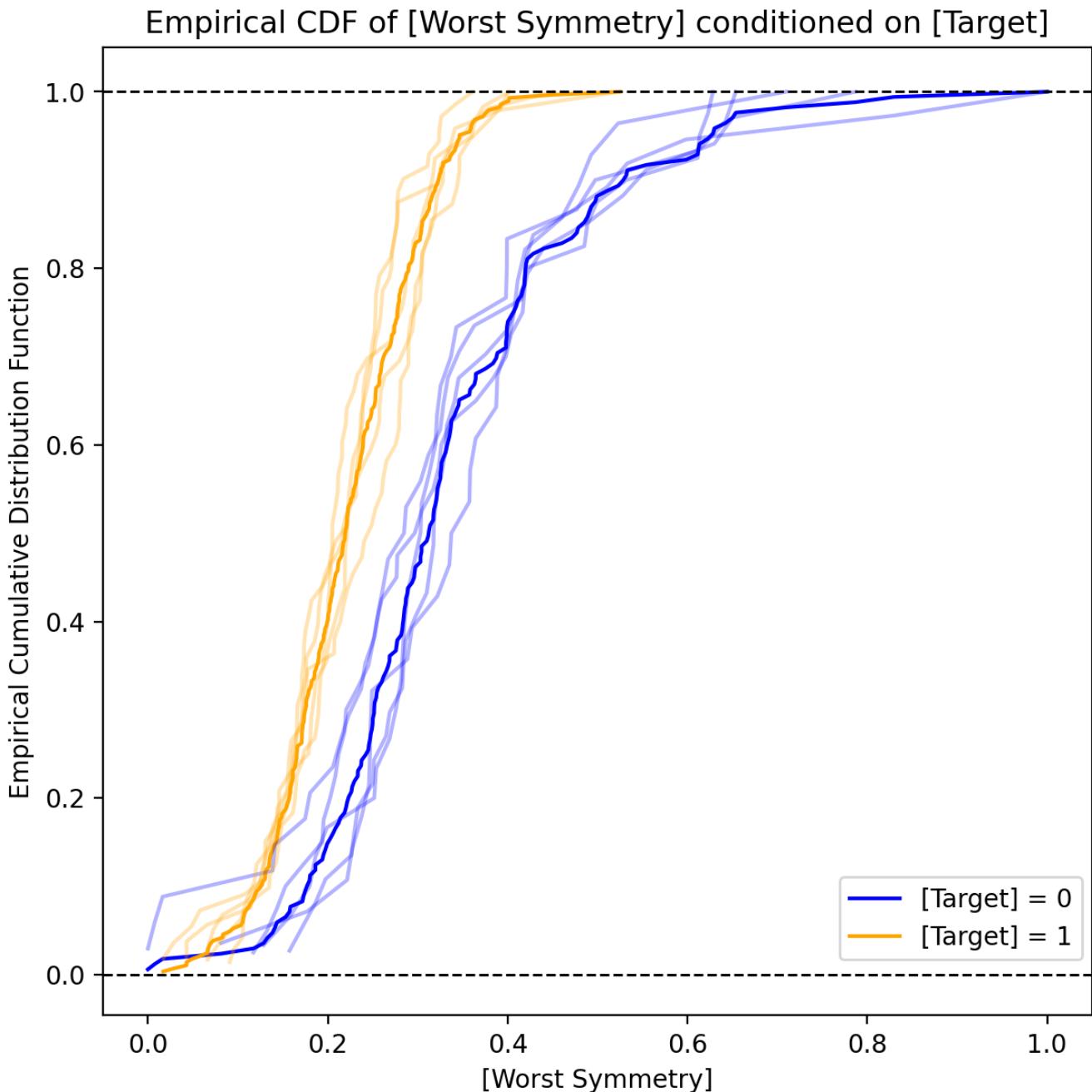
Kernel Density Plot of [Worst Symmetry] by [Target]  
Distributions by level are significantly different at the 95% level.



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



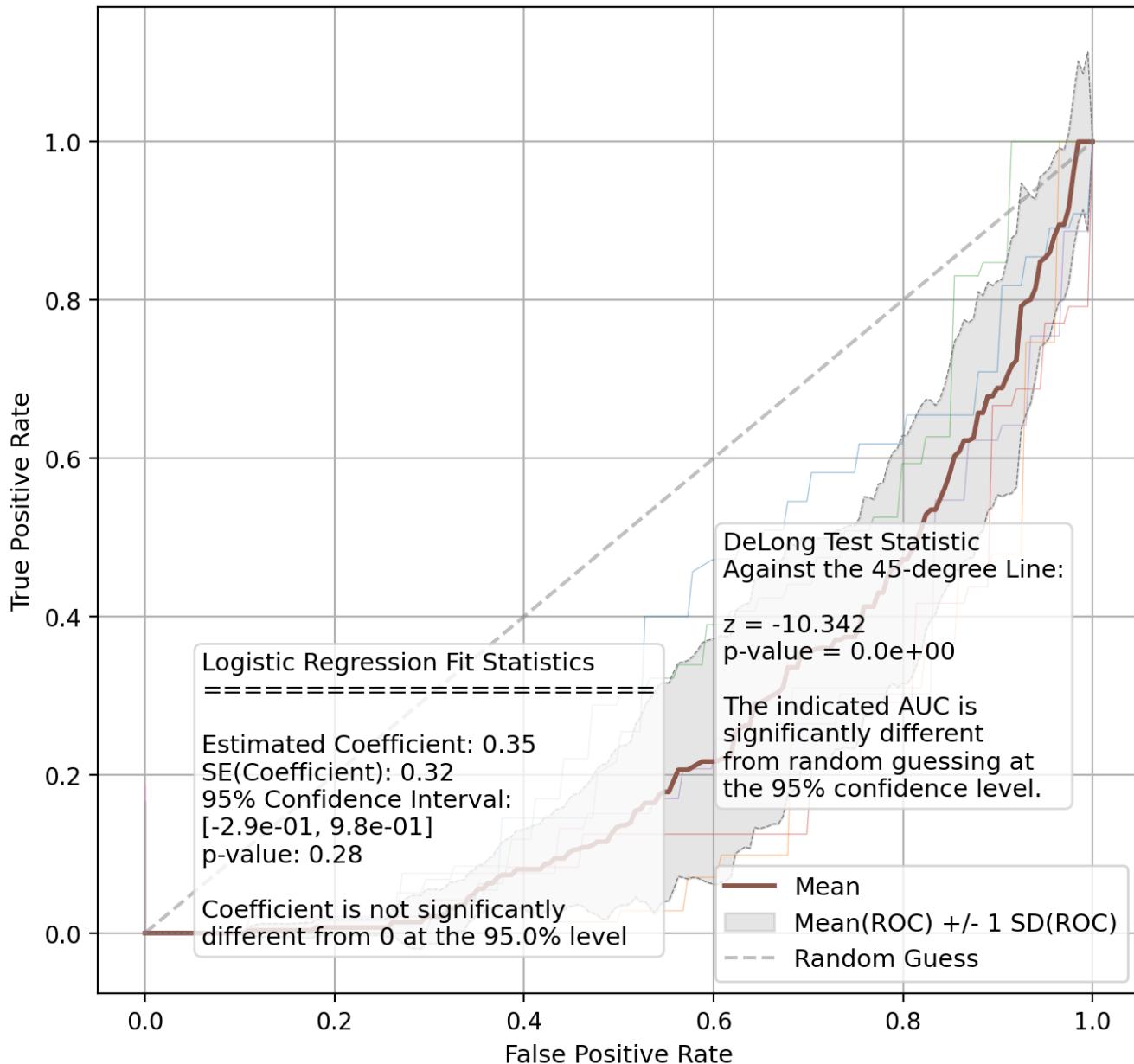
This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly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

ROC Curve (AUC = 24.6%)

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



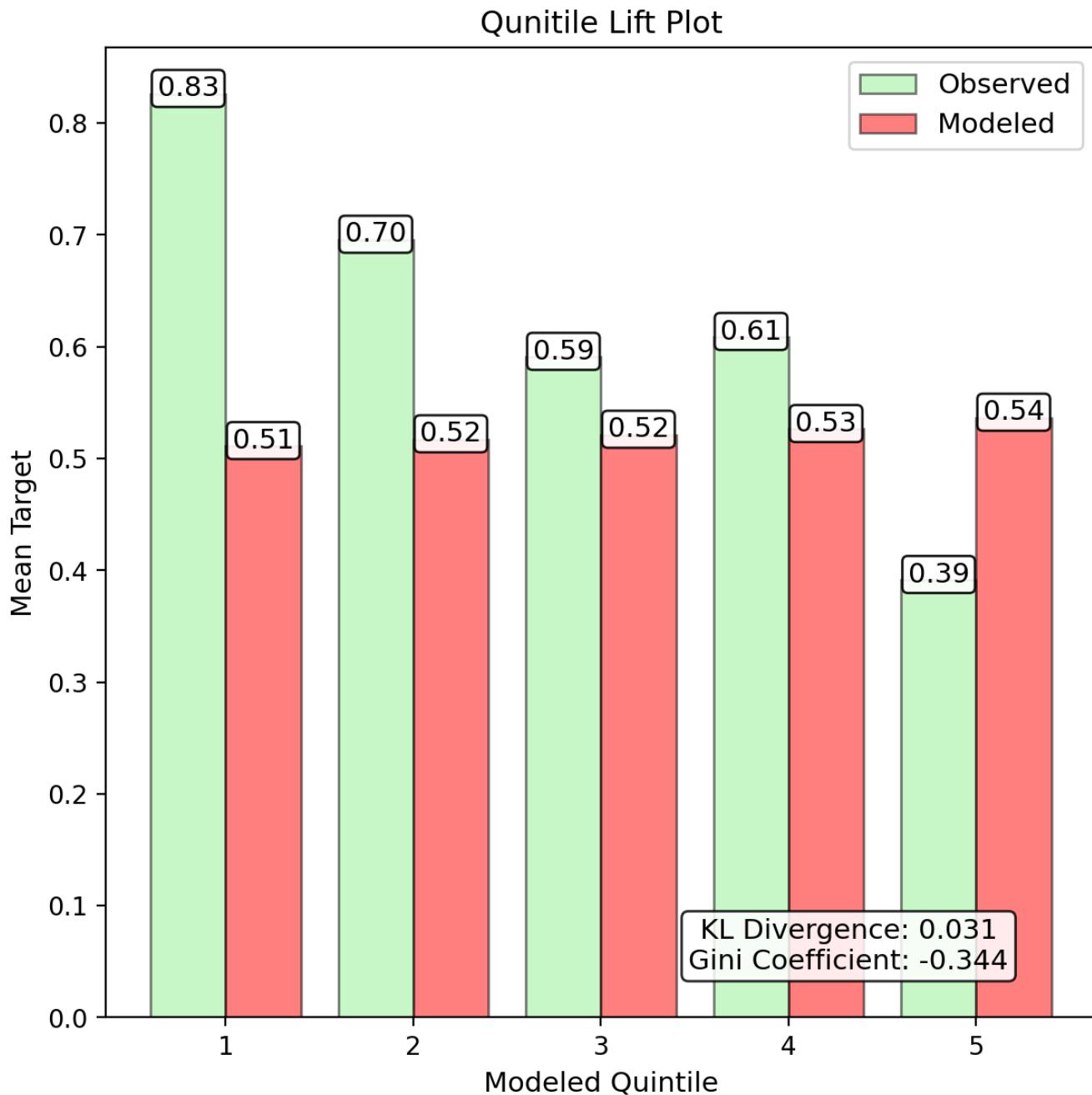
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 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

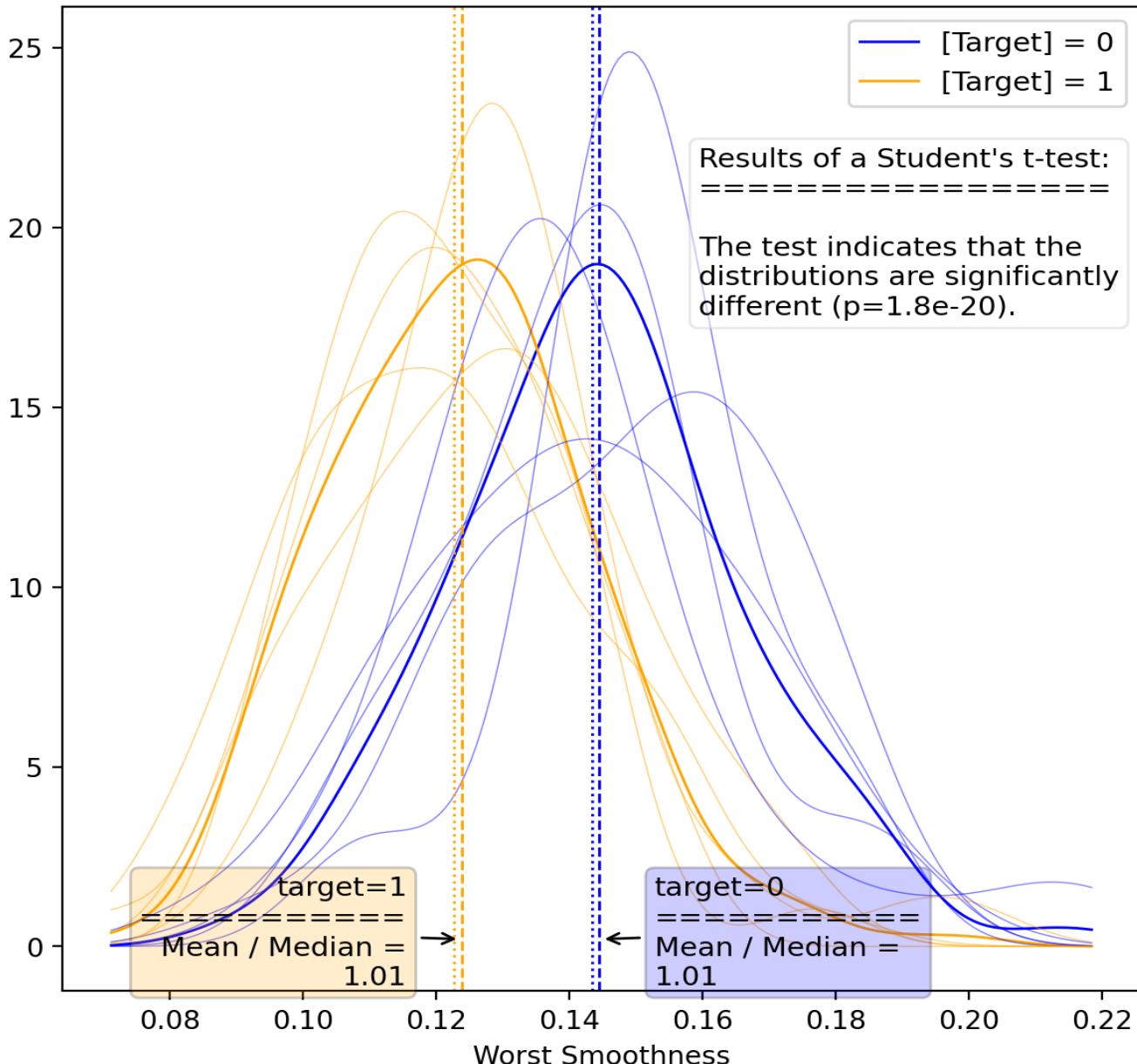
### Worst Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	3.01	1.94	2.76	3.13	2.63	2.70	0.47
<b>Fitted p-Value</b>	2.1e-04	1.6e-02	5.8e-04	8.3e-05	8.1e-04	1.6e-04	7.0e-03
<b>Fitted Std. Err.</b>	0.812	0.805	0.801	0.795	0.786	0.715	0.010
<b>Conf. Int. Lower</b>	1.42	0.36	1.19	1.57	1.09	1.29	0.47
<b>Conf. Int. Upper</b>	4.60	3.51	4.33	4.69	4.17	4.10	0.47
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

Worst Smoothness - Kernel Density Plot

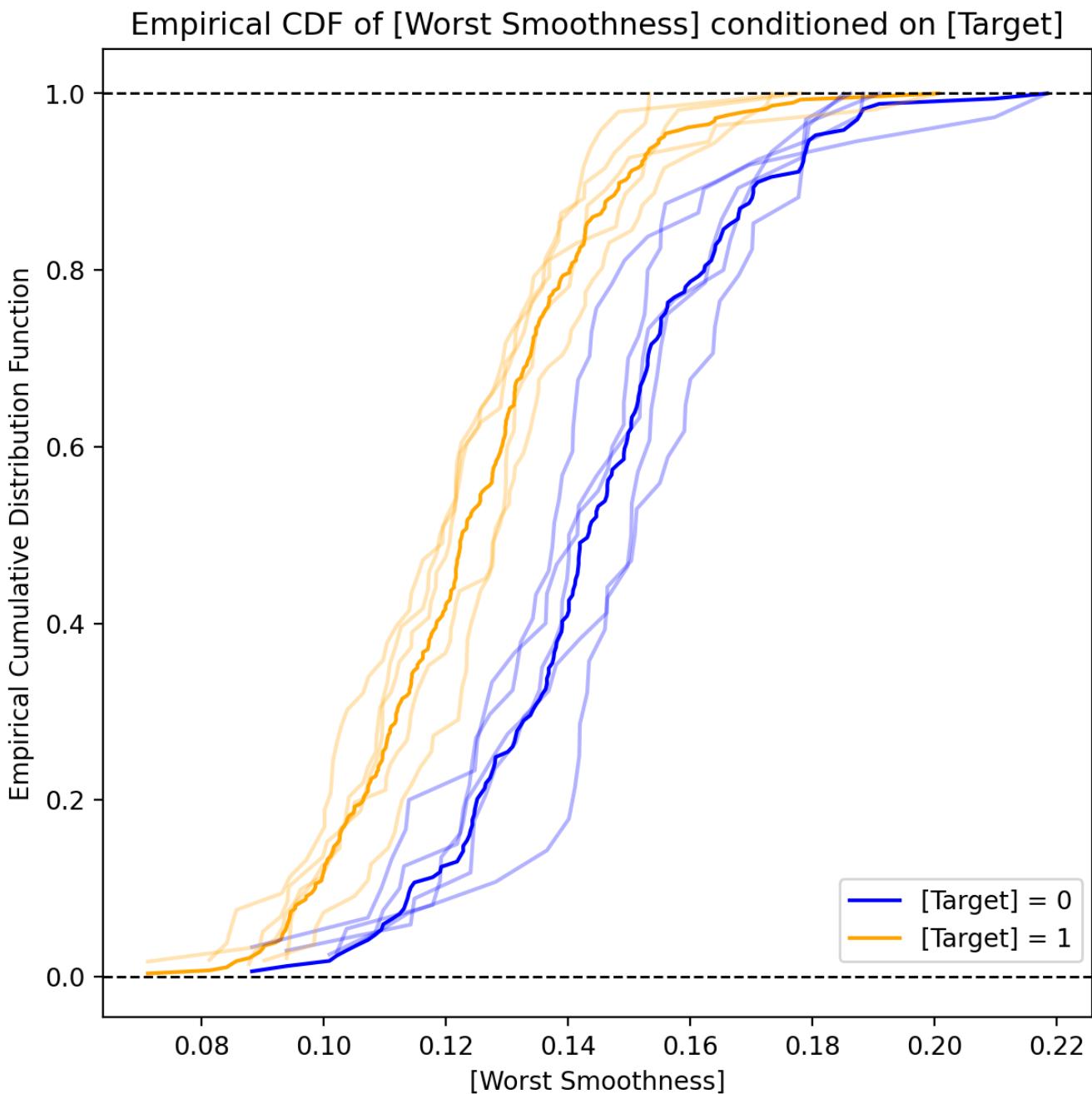
Kernel Density Plot of [Worst Smoothness] by [Target]  
Distributions by level are significantly different at the 95% level.



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

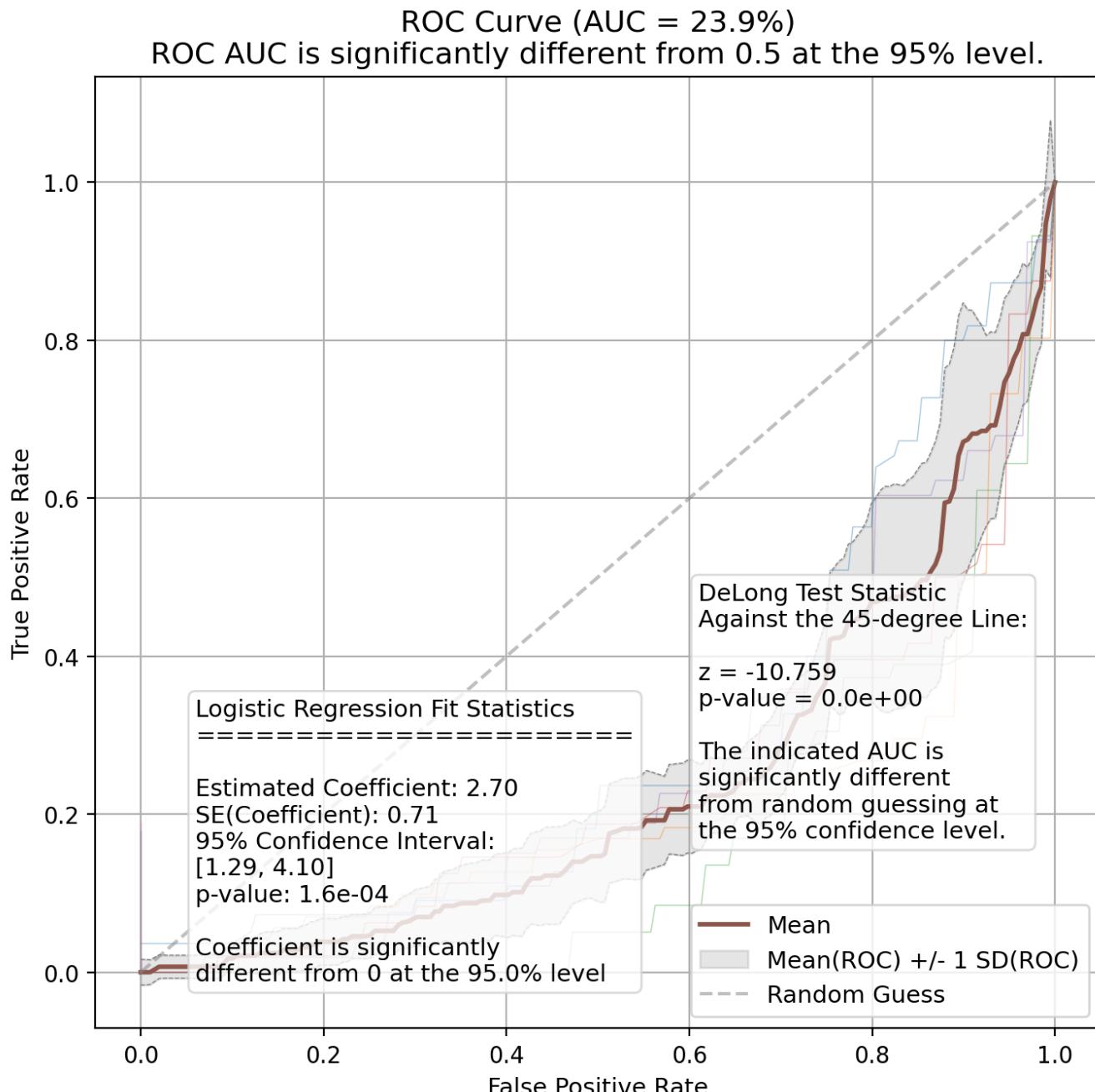
Worst Smoothness - Empirical CDF Plot



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



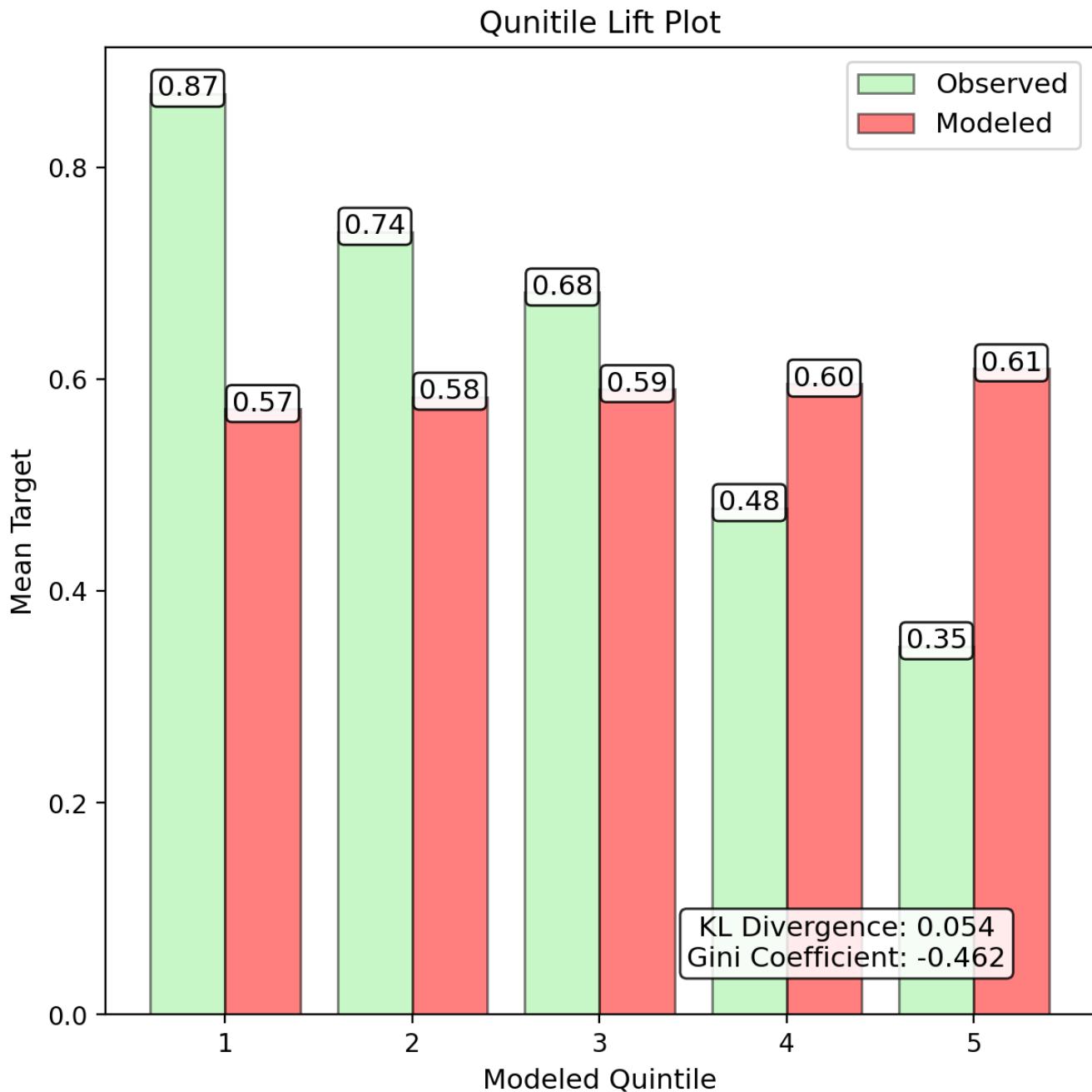
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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.

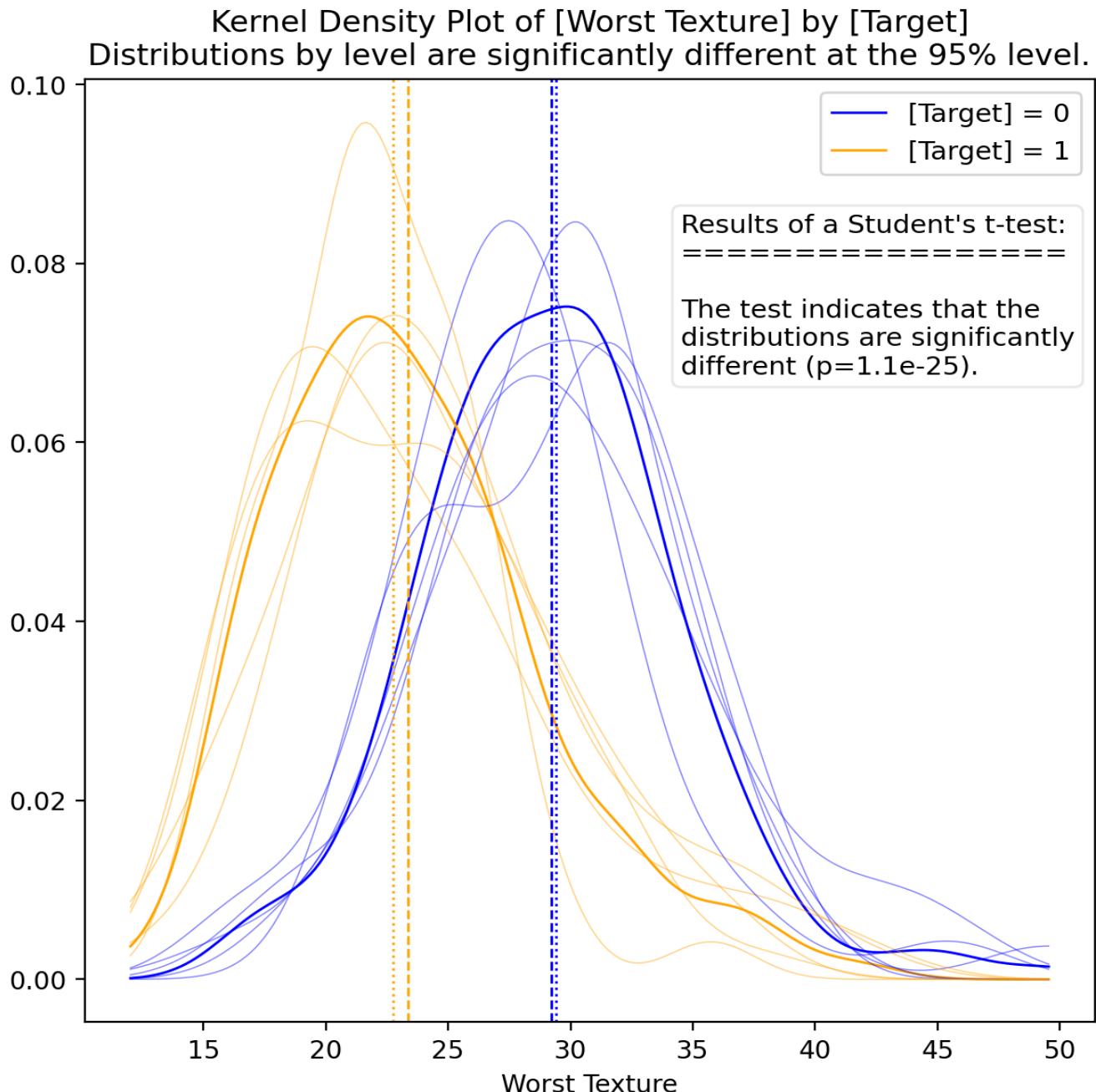
## Univariate Report

### Worst Texture - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	1.3e-02	7.4e-03	1.0e-02	1.3e-02	1.1e-02	1.1e-02	2.3e-03
<b>Fitted p-Value</b>	1.2e-03	6.7e-02	1.2e-02	1.7e-03	4.7e-03	2.4e-03	2.8e-02
<b>Fitted Std. Err.</b>	4.1e-03	4.0e-03	4.1e-03	4.0e-03	4.0e-03	3.6e-03	5.2e-05
<b>Conf. Int. Lower</b>	5.3e-03	-5.1e-04	2.3e-03	4.8e-03	3.5e-03	3.9e-03	2.3e-03
<b>Conf. Int. Upper</b>	2.1e-02	1.5e-02	1.8e-02	2.1e-02	1.9e-02	1.8e-02	2.4e-03
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

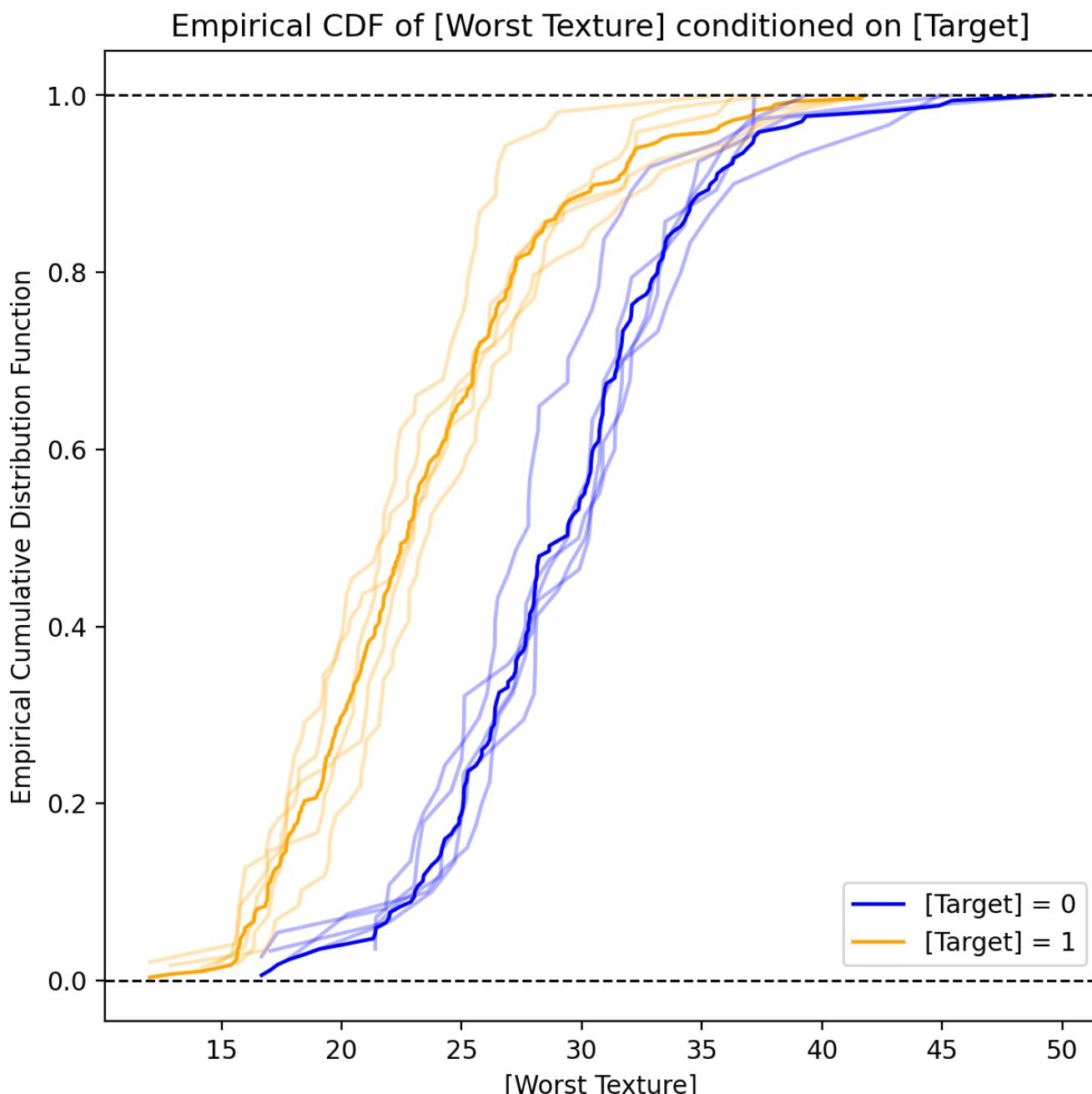
### Worst Texture - Kernel Density Plot



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

## Univariate Report

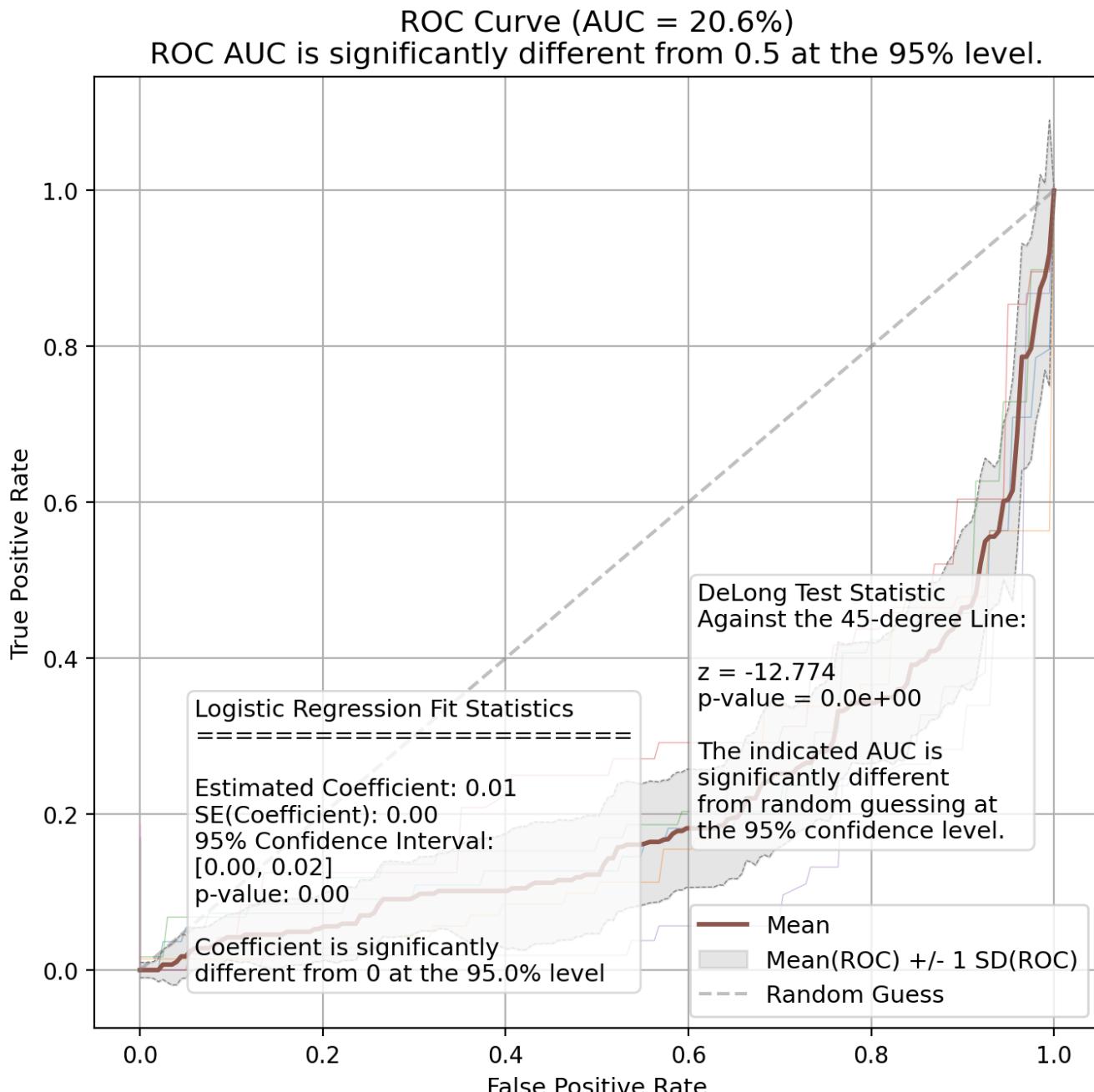
Worst Texture - Empirical CDF Plot



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



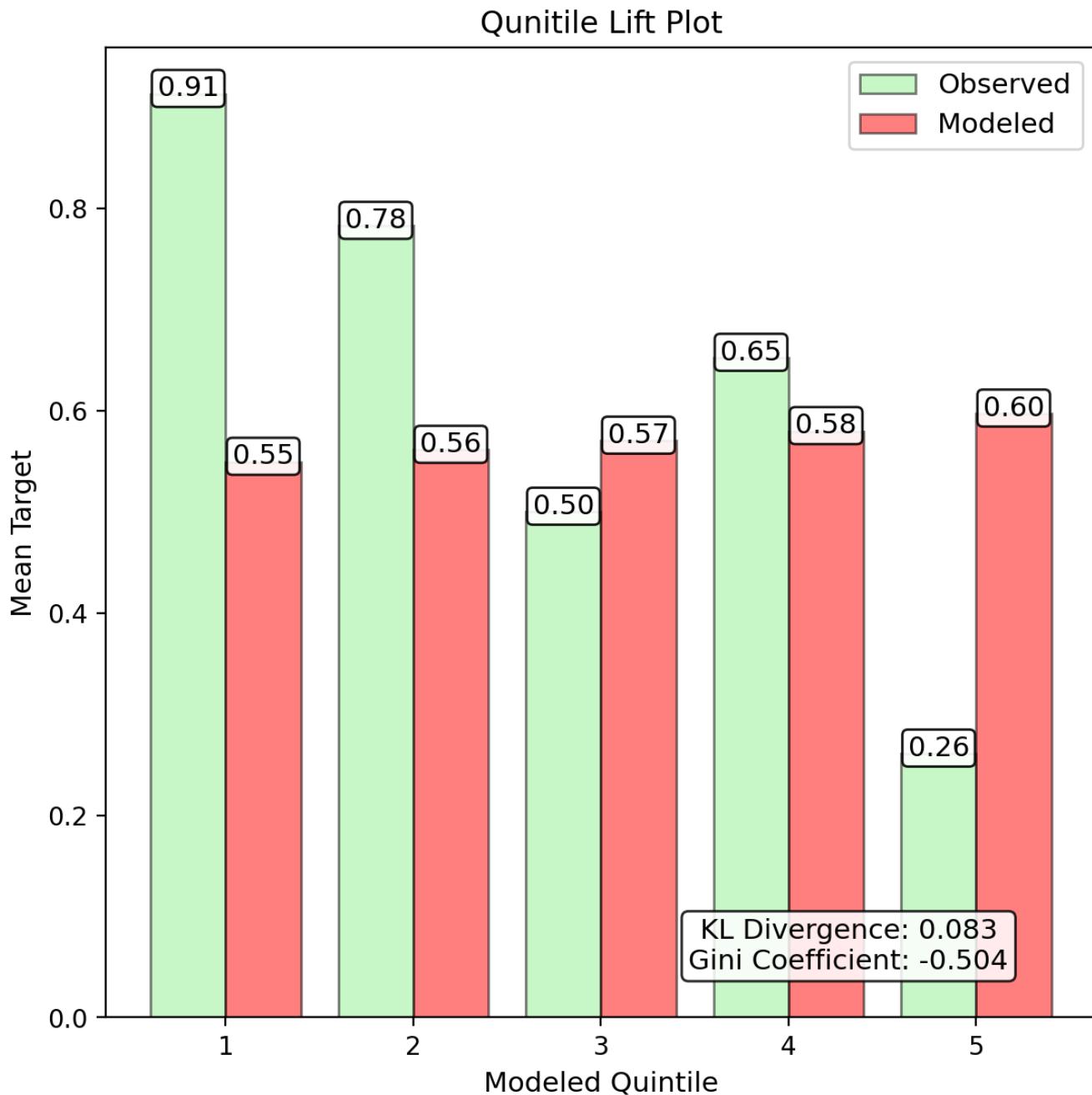
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

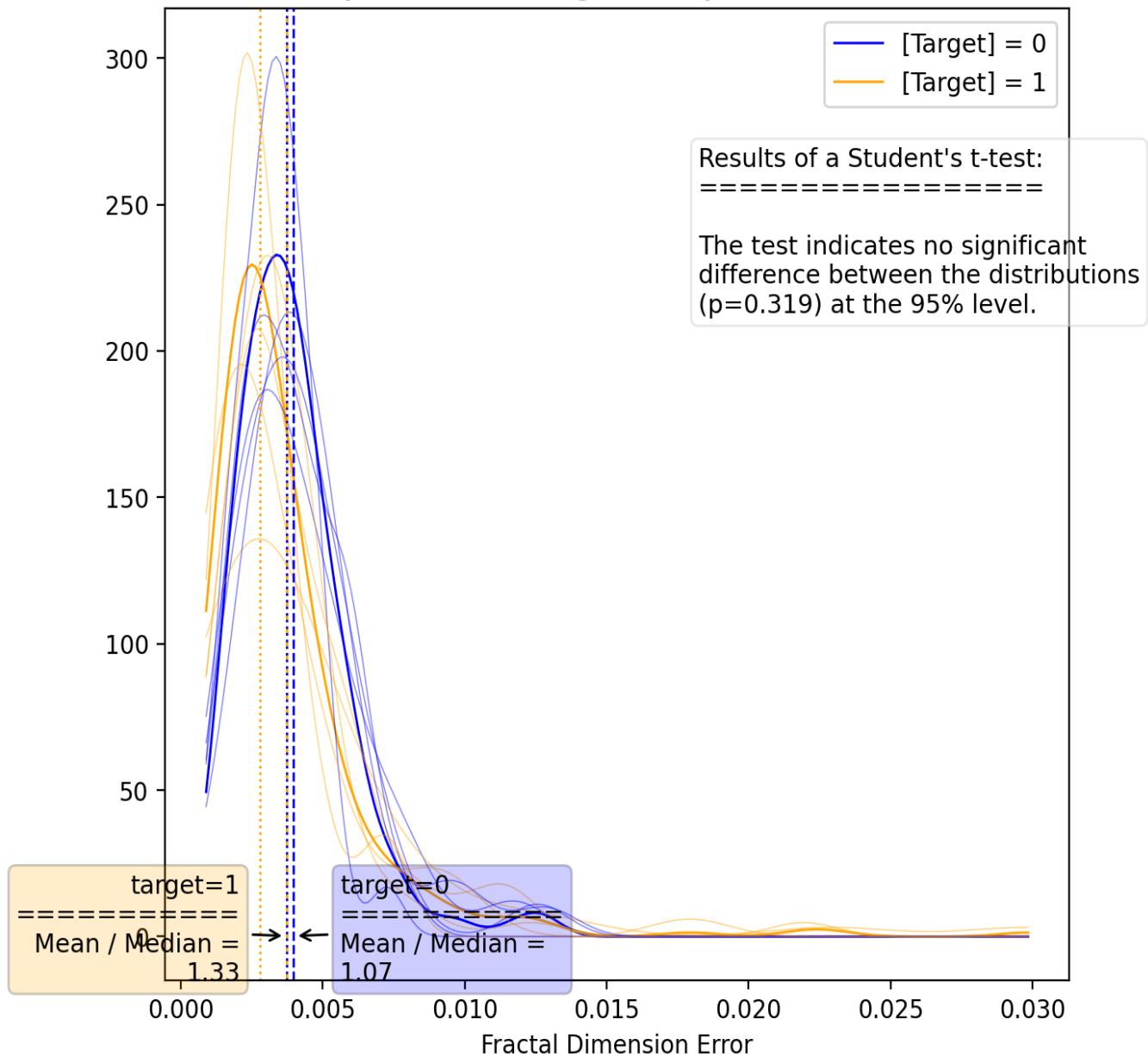
## Fractal Dimension Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	78.02	64.18	81.44	106.11	83.74	82.45	15.13
<b>Fitted p-Value</b>	1.3e-03	7.6e-03	1.5e-03	3.7e-05	4.0e-04	1.8e-04	3.1e-03
<b>Fitted Std. Err.</b>	24.23	24.04	25.63	25.73	23.64	22.01	0.96
<b>Conf. Int. Lower</b>	30.53	17.06	31.20	55.69	37.41	39.32	14.04
<b>Conf. Int. Upper</b>	125.50	111.30	131.67	156.53	130.07	125.59	16.37
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

### Fractal Dimension Error - Kernel Density Plot

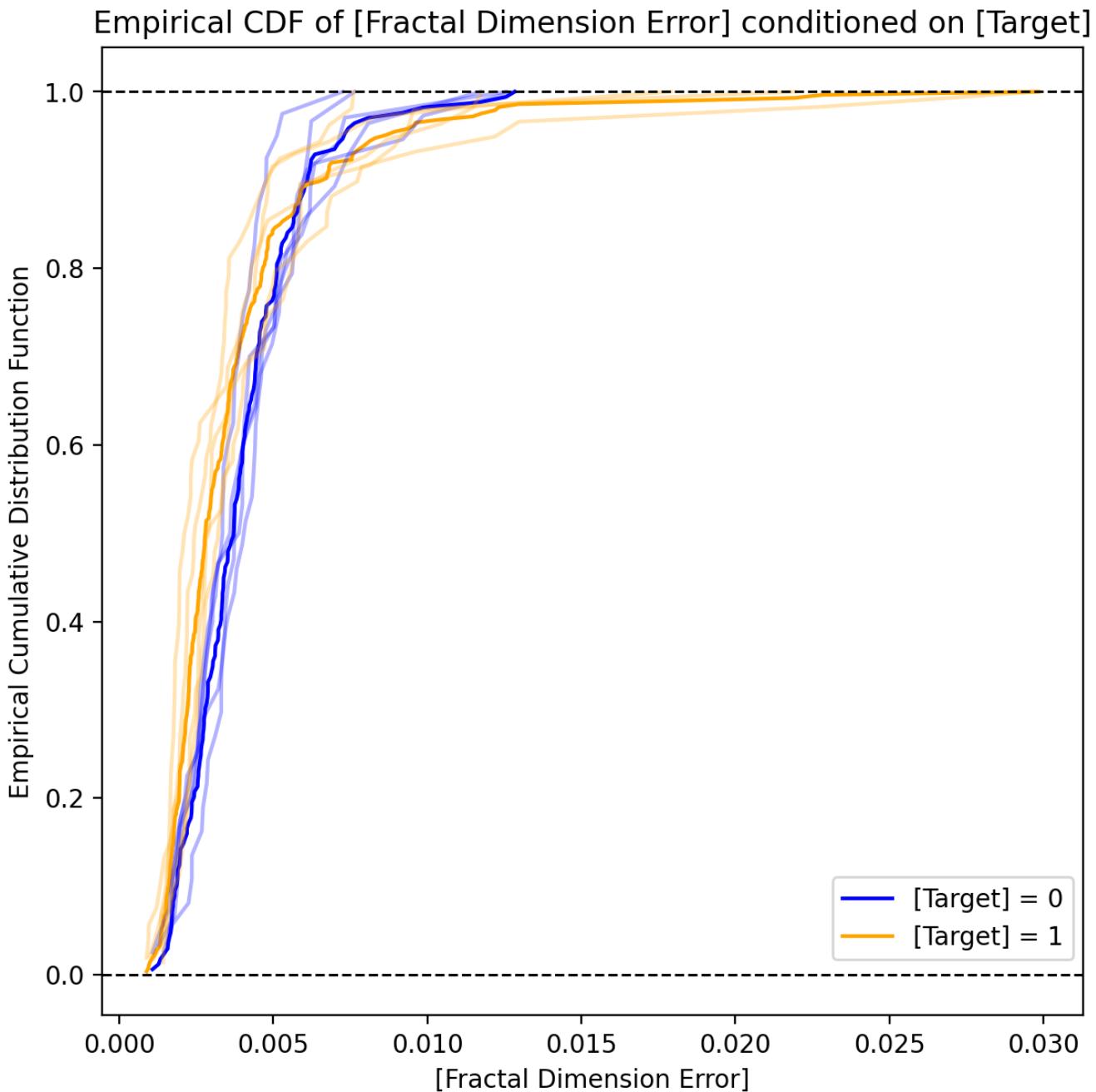
Kernel Density Plot of [Fractal Dimension Error] by [Target]  
Distributions by level are not significantly different at the 95% level.



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

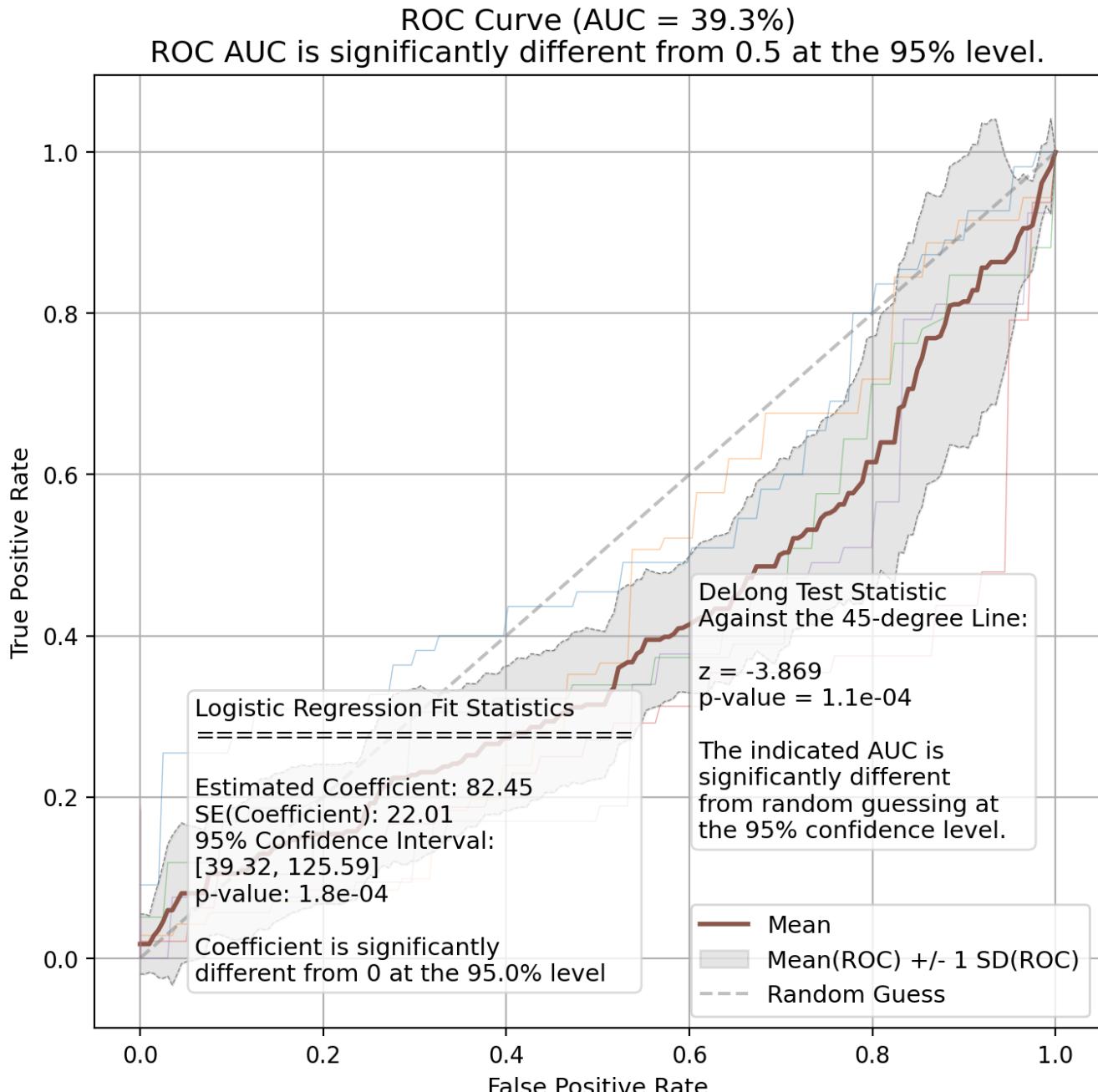
### Fractal Dimension Error - Empirical CDF Plot



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



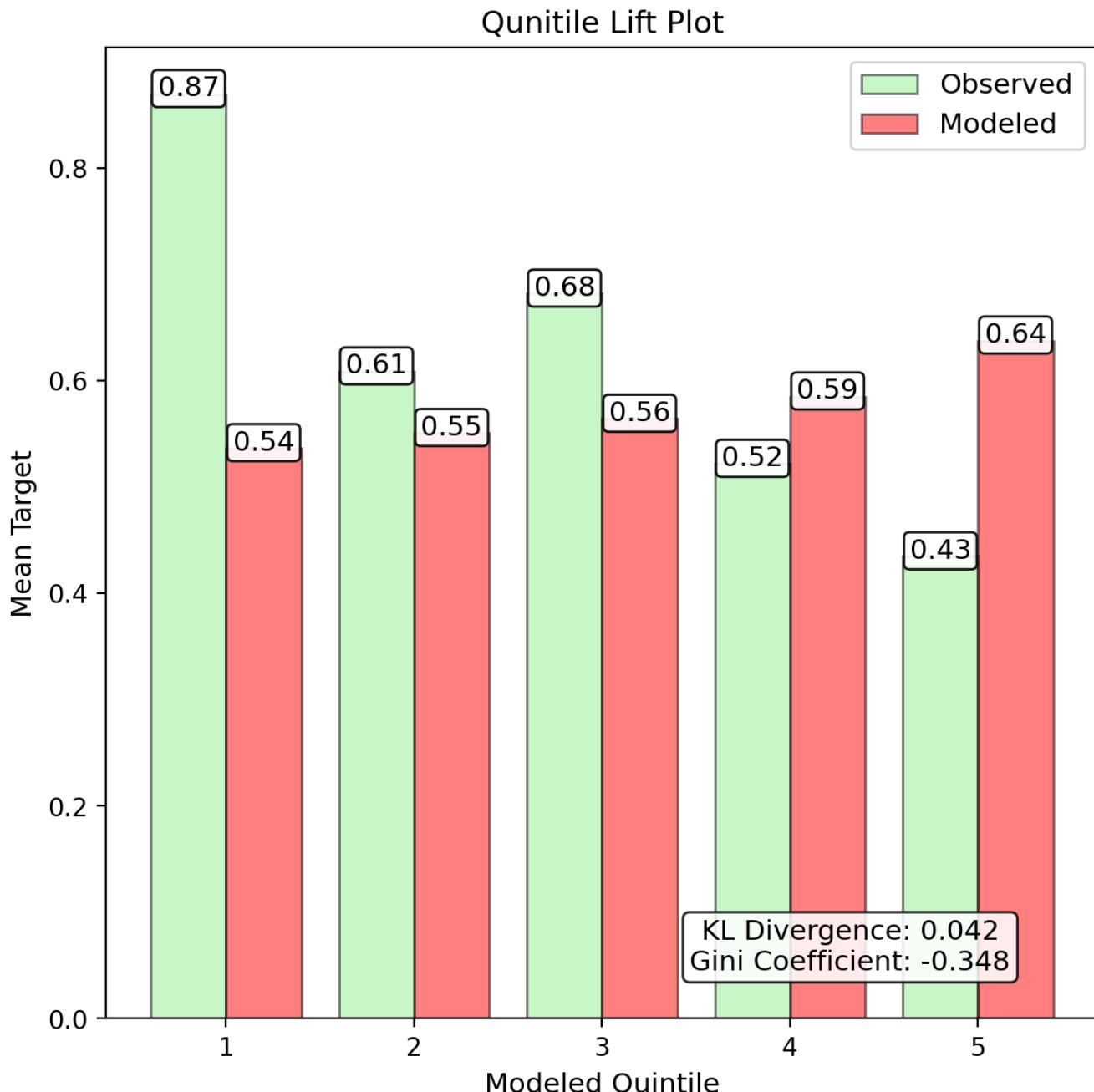
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

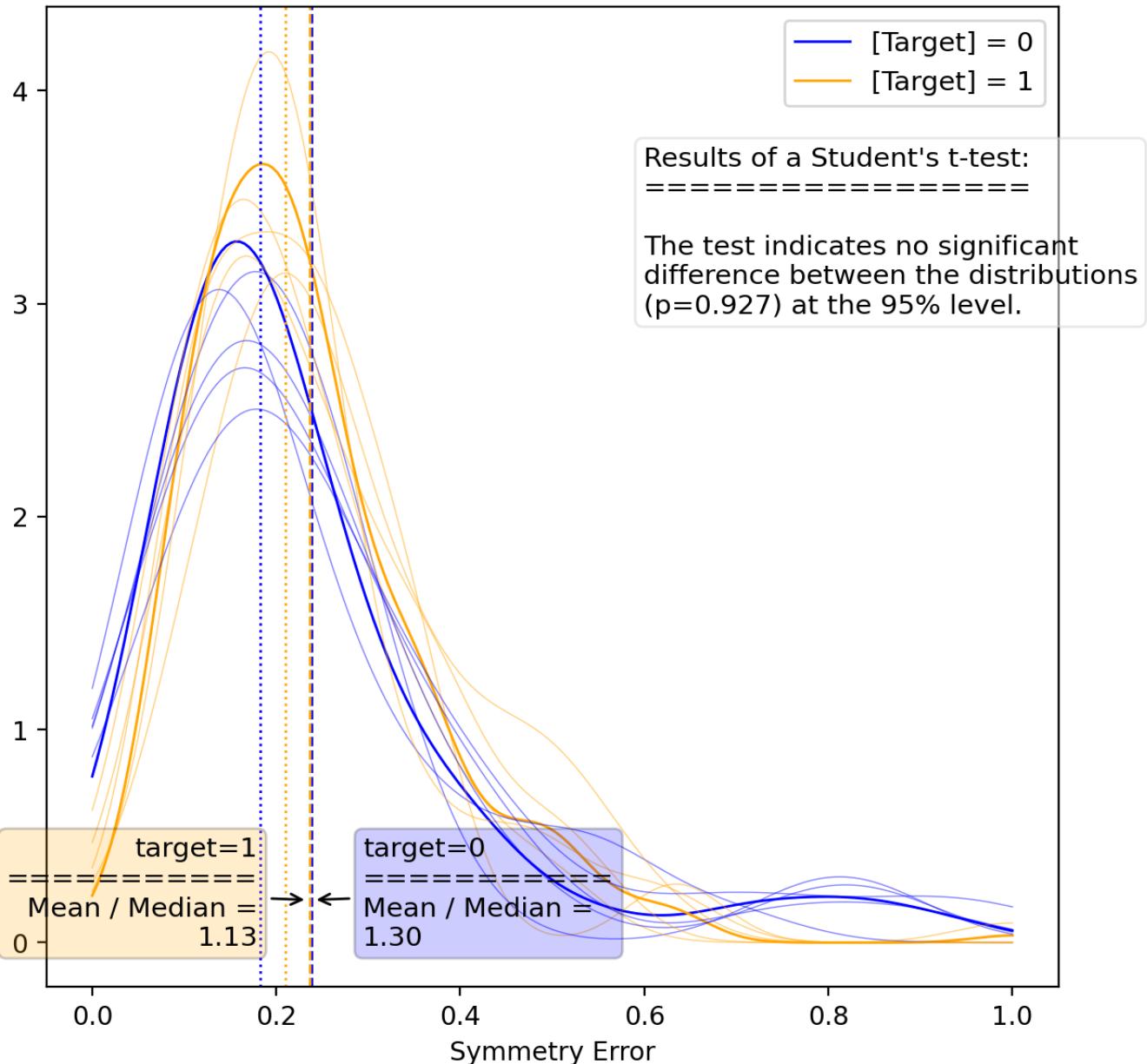
### Symmetry Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	1.65	1.26	1.43	1.97	1.53	1.57	0.27
<b>Fitted p-Value</b>	3.6e-05	1.2e-03	4.1e-04	2.2e-06	1.1e-04	1.3e-05	5.0e-04
<b>Fitted Std. Err.</b>	0.400	0.390	0.405	0.417	0.394	0.358	0.011
<b>Conf. Int. Lower</b>	0.869	0.498	0.639	1.156	0.755	0.863	0.250
<b>Conf. Int. Upper</b>	2.44	2.03	2.23	2.79	2.30	2.27	0.28
<b>Train Accuracy</b>	64.4%	60.7%	62.7%	64.6%	62.9%	63.1%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	64.5%	56.5%	63.9%	62.3%	6.1%
<b>Train AUC</b>	50.4%	50.4%	50.0%	50.4%	50.4%	50.3%	0.2%
<b>Val AUC</b>	50.0%	50.0%	51.5%	50.0%	50.0%	50.0%	0.7%
<b>Train F1</b>	78.3%	75.4%	77.1%	78.4%	77.2%	77.3%	1.2%
<b>Test F1</b>	73.3%	83.5%	78.1%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.3%	60.6%	62.7%	64.5%	62.8%	63.0%	1.6%
<b>Val Precision</b>	57.9%	71.7%	64.1%	56.5%	63.9%	62.3%	6.1%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	7.1%	6.6%	0.0%	7.0%	6.7%	6.1%	3.1%
<b>Val MCC</b>	0.0%	0.0%	13.7%	0.0%	0.0%	0.0%	6.1%
<b>Train Log-Loss</b>	12.82	14.17	13.44	12.76	13.37	13.31	0.57
<b>Val Log-Loss</b>	15.18	10.19	12.79	15.69	13.03	13.60	2.19

## Univariate Report

### Symmetry Error - Kernel Density Plot

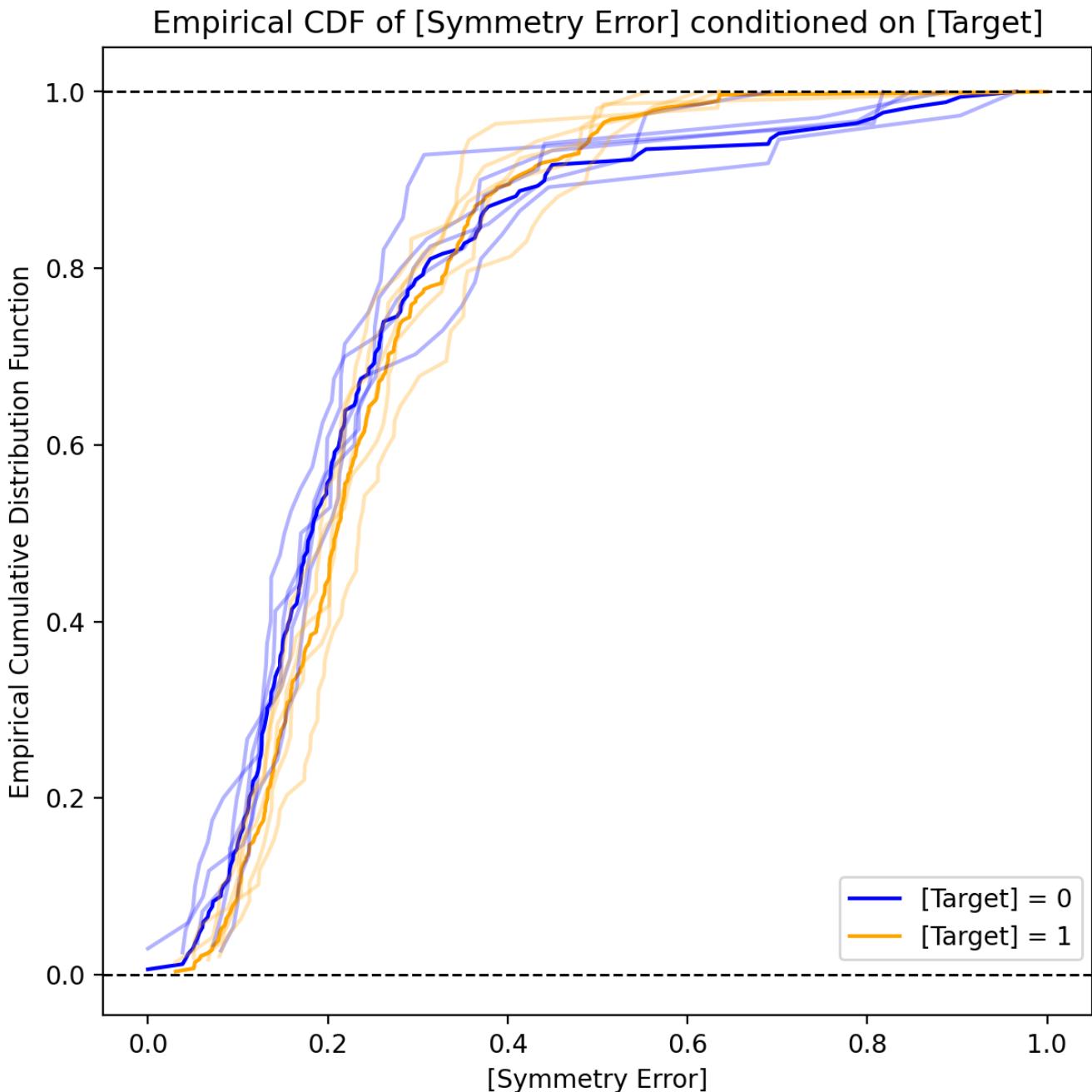
Kernel Density Plot of [Symmetry Error] by [Target]  
Distributions by level are not significantly different at the 95% level.



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

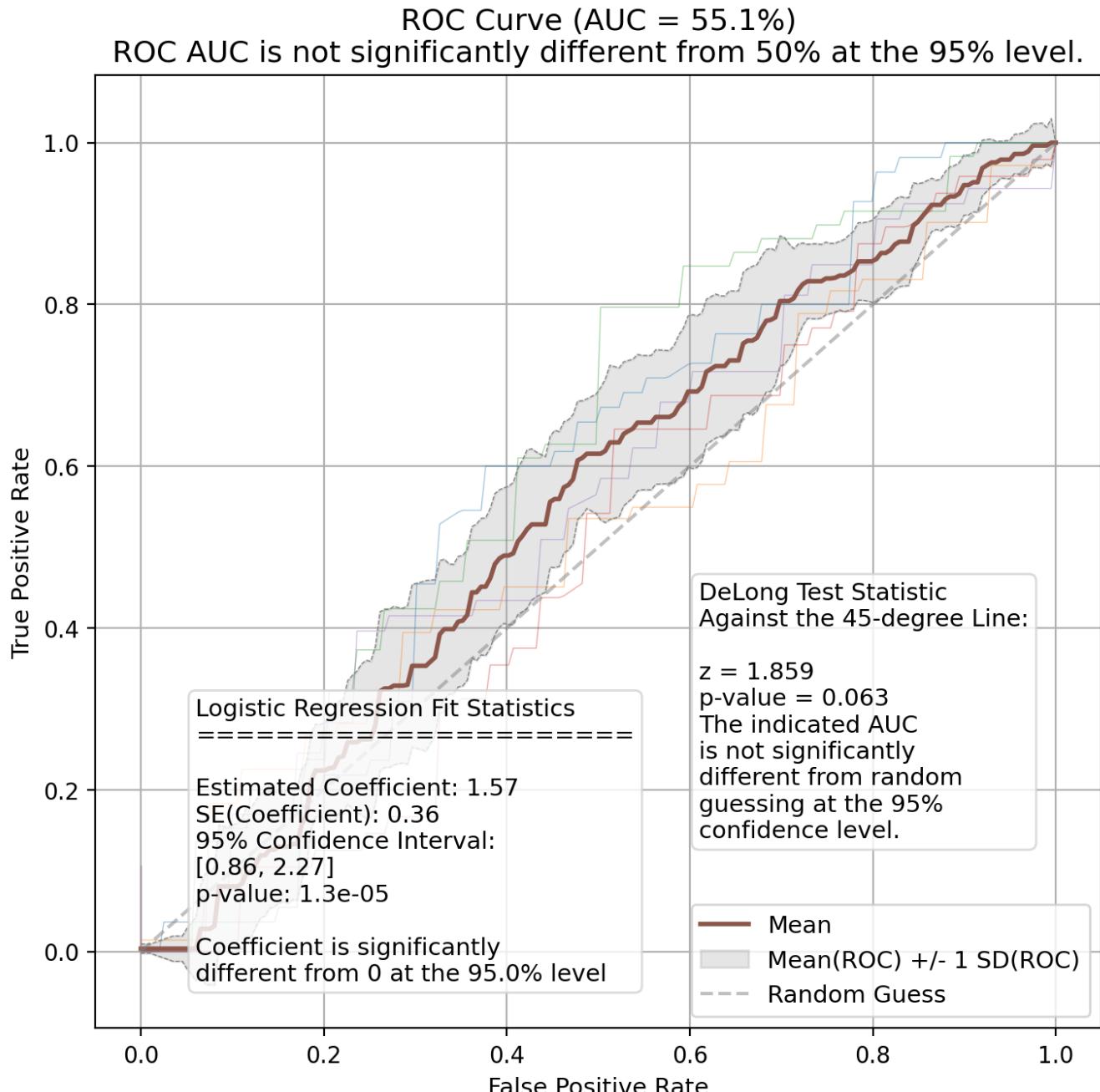
### Symmetry Error - Empirical CDF Plot



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



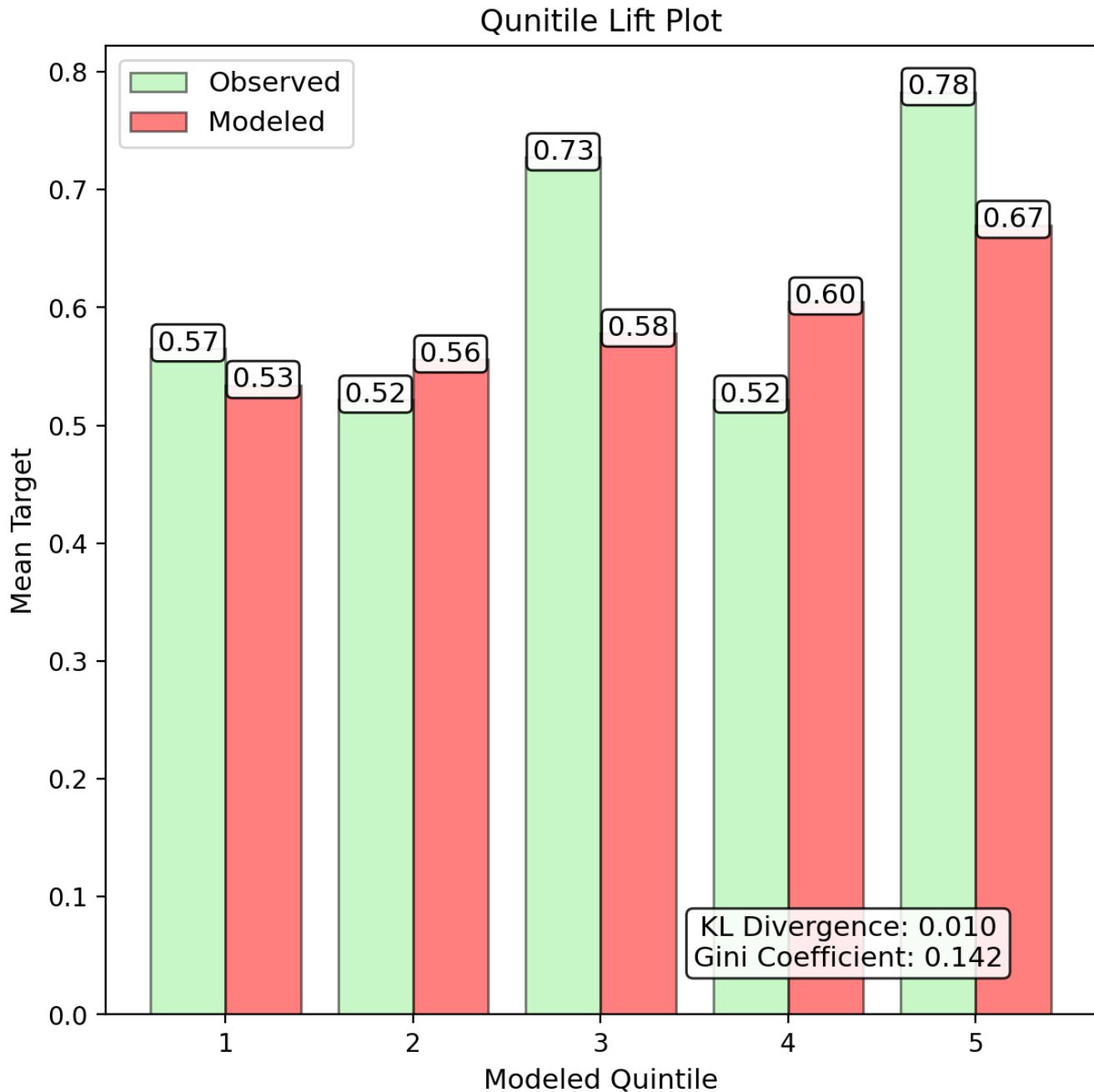
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

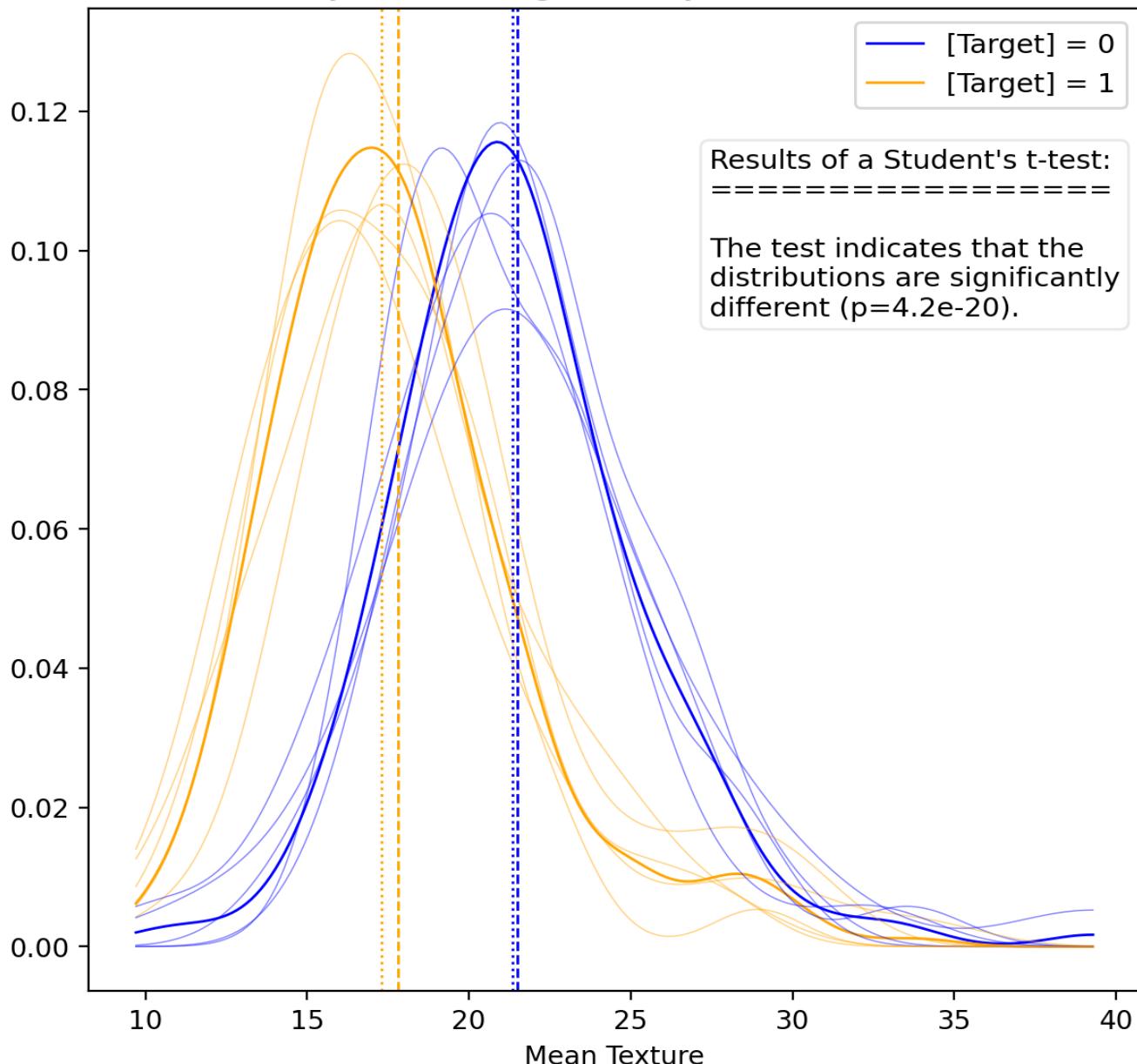
## Mean Texture - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	2.0e-02	1.1e-02	1.6e-02	1.9e-02	1.7e-02	1.7e-02	3.2e-03
<b>Fitted p-Value</b>	3.5e-04	3.3e-02	4.2e-03	4.3e-04	1.7e-03	6.6e-04	1.4e-02
<b>Fitted Std. Err.</b>	5.5e-03	5.4e-03	5.5e-03	5.4e-03	5.4e-03	4.8e-03	6.7e-05
<b>Conf. Int. Lower</b>	8.9e-03	9.1e-04	5.0e-03	8.4e-03	6.3e-03	7.0e-03	3.2e-03
<b>Conf. Int. Upper</b>	3.0e-02	2.2e-02	2.6e-02	3.0e-02	2.7e-02	2.6e-02	3.3e-03
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

### Mean Texture - Kernel Density Plot

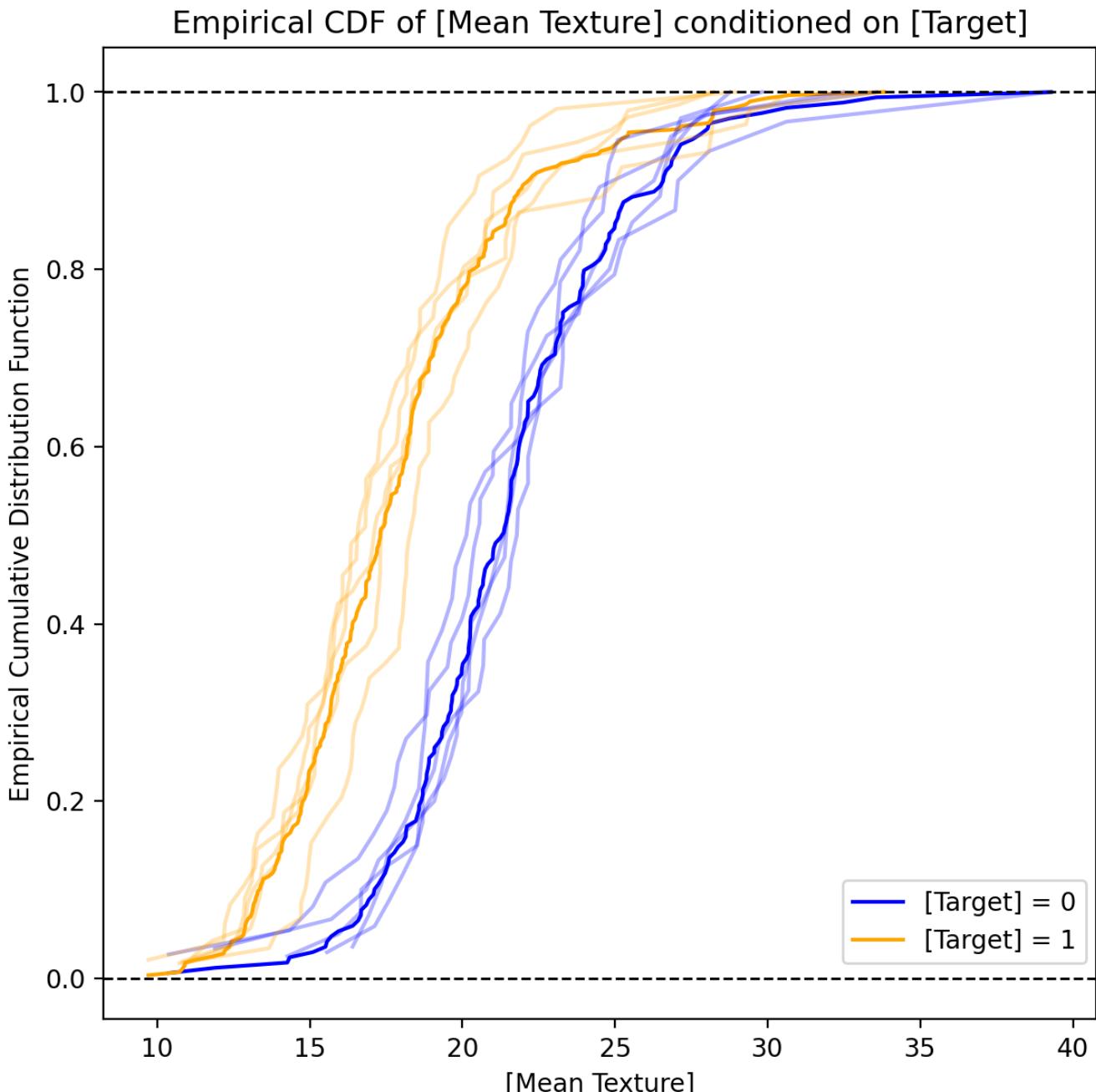
Kernel Density Plot of [Mean Texture] by [Target]  
Distributions by level are significantly different at the 95% level.



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

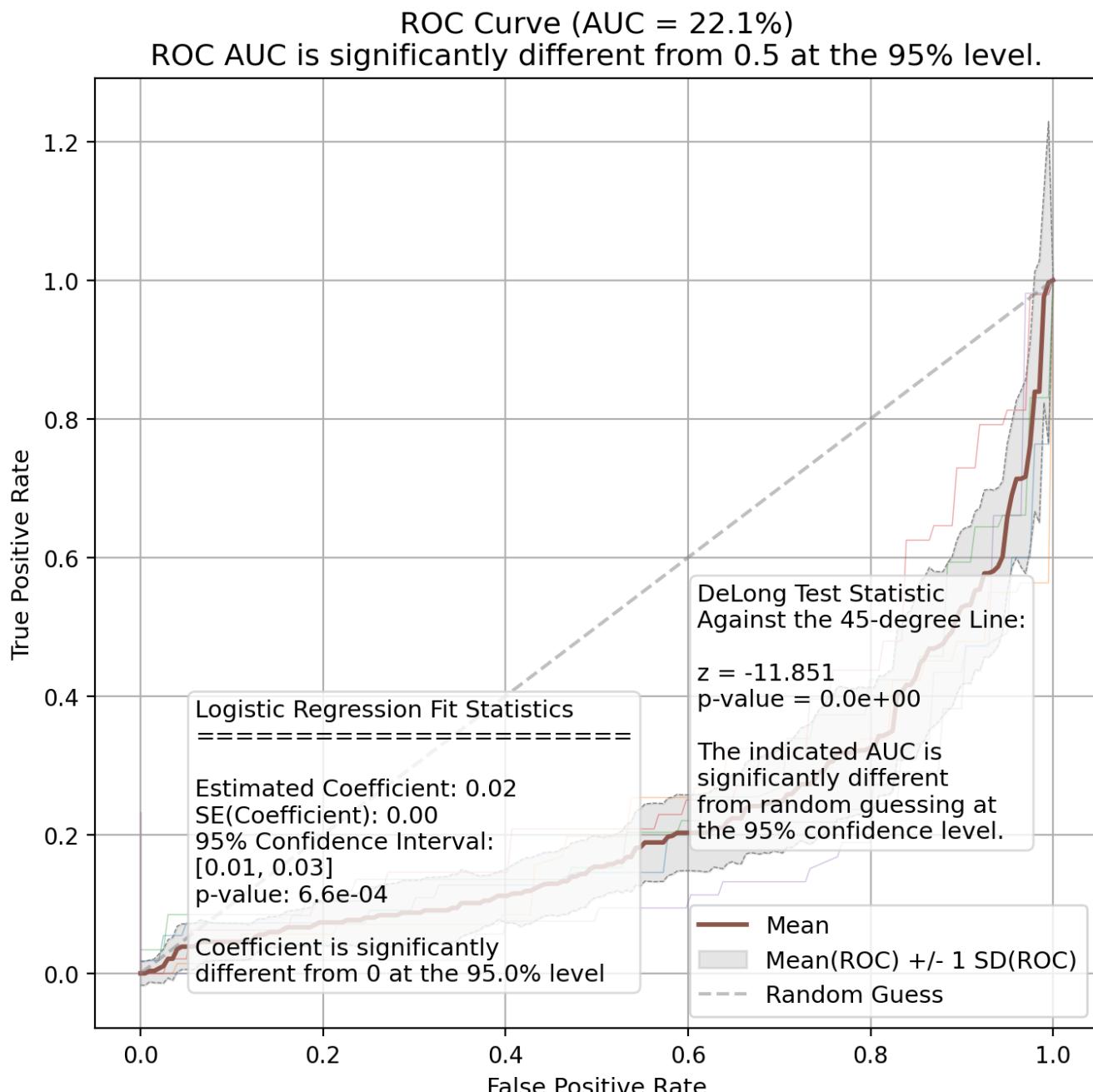
### Mean Texture - Empirical CDF Plot



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



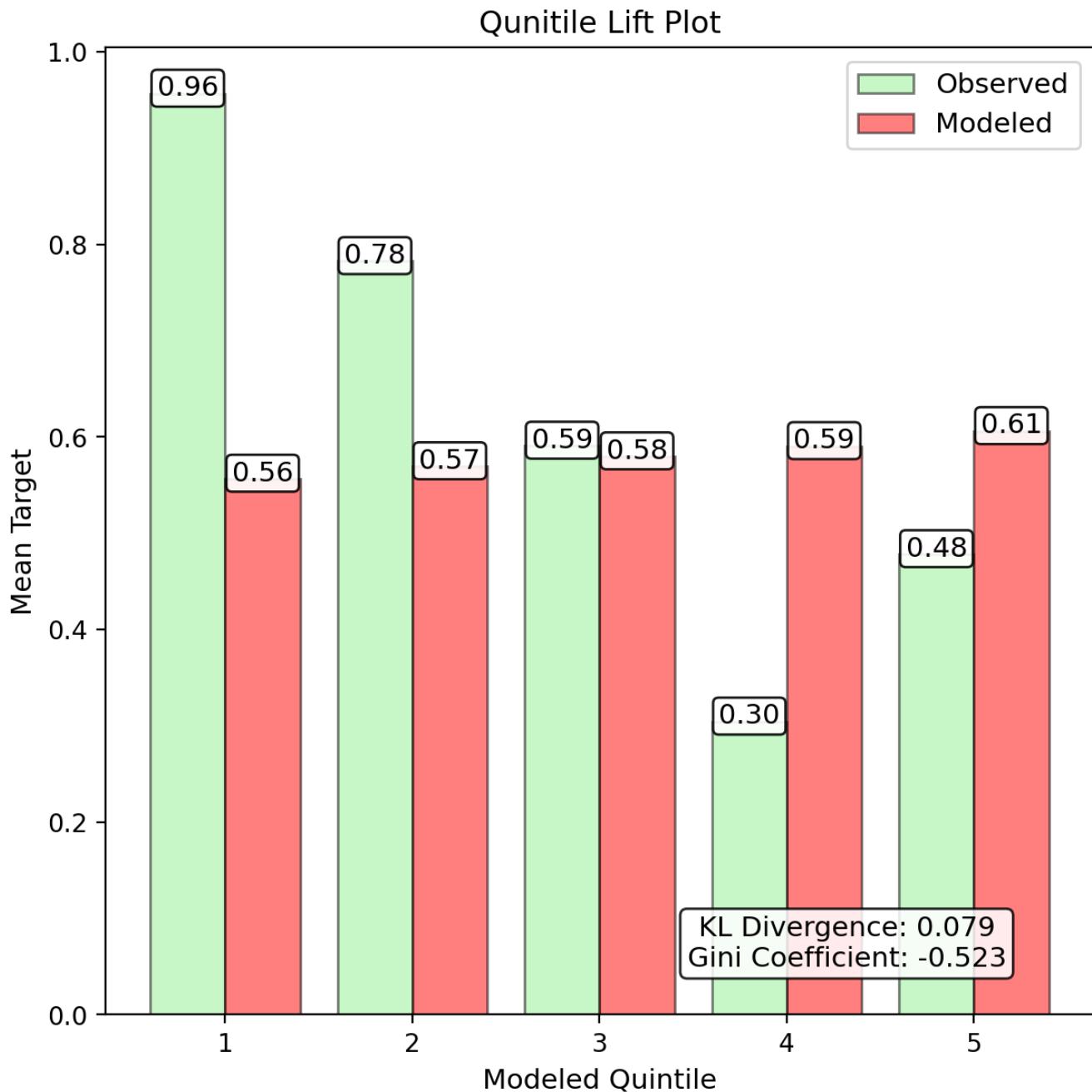
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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.

# Univariate Report

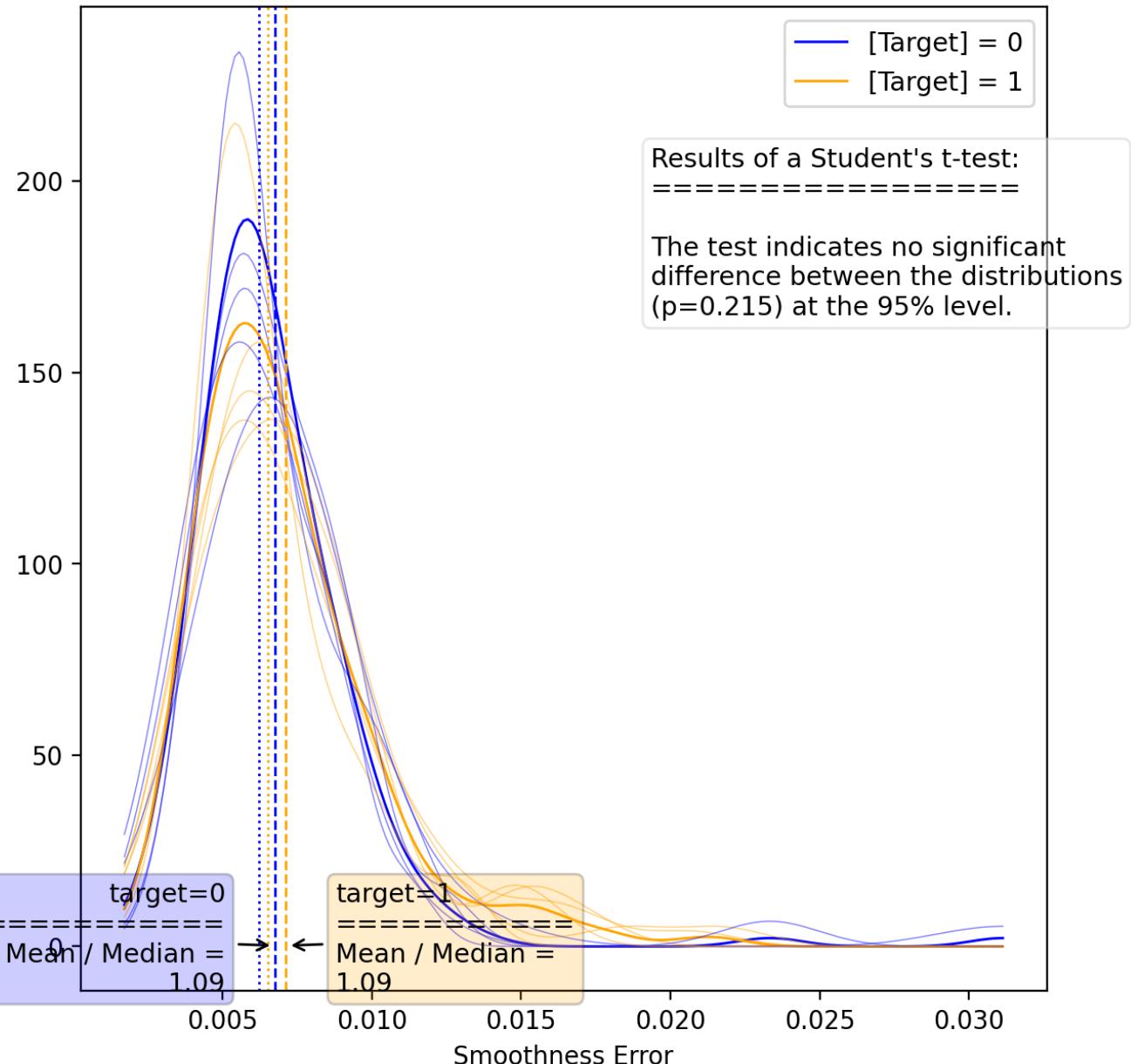
## Smoothness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	73.81	55.93	74.81	78.07	66.51	69.84	8.84
<b>Fitted p-Value</b>	8.8e-07	1.5e-04	7.6e-07	1.1e-07	4.2e-06	1.3e-07	6.7e-05
<b>Fitted Std. Err.</b>	15.01	14.76	15.13	14.69	14.45	13.23	0.27
<b>Conf. Int. Lower</b>	44.38	27.01	45.16	49.28	38.19	43.91	8.67
<b>Conf. Int. Upper</b>	103.24	84.85	104.47	106.85	94.82	95.78	9.04
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

### Smoothness Error - Kernel Density Plot

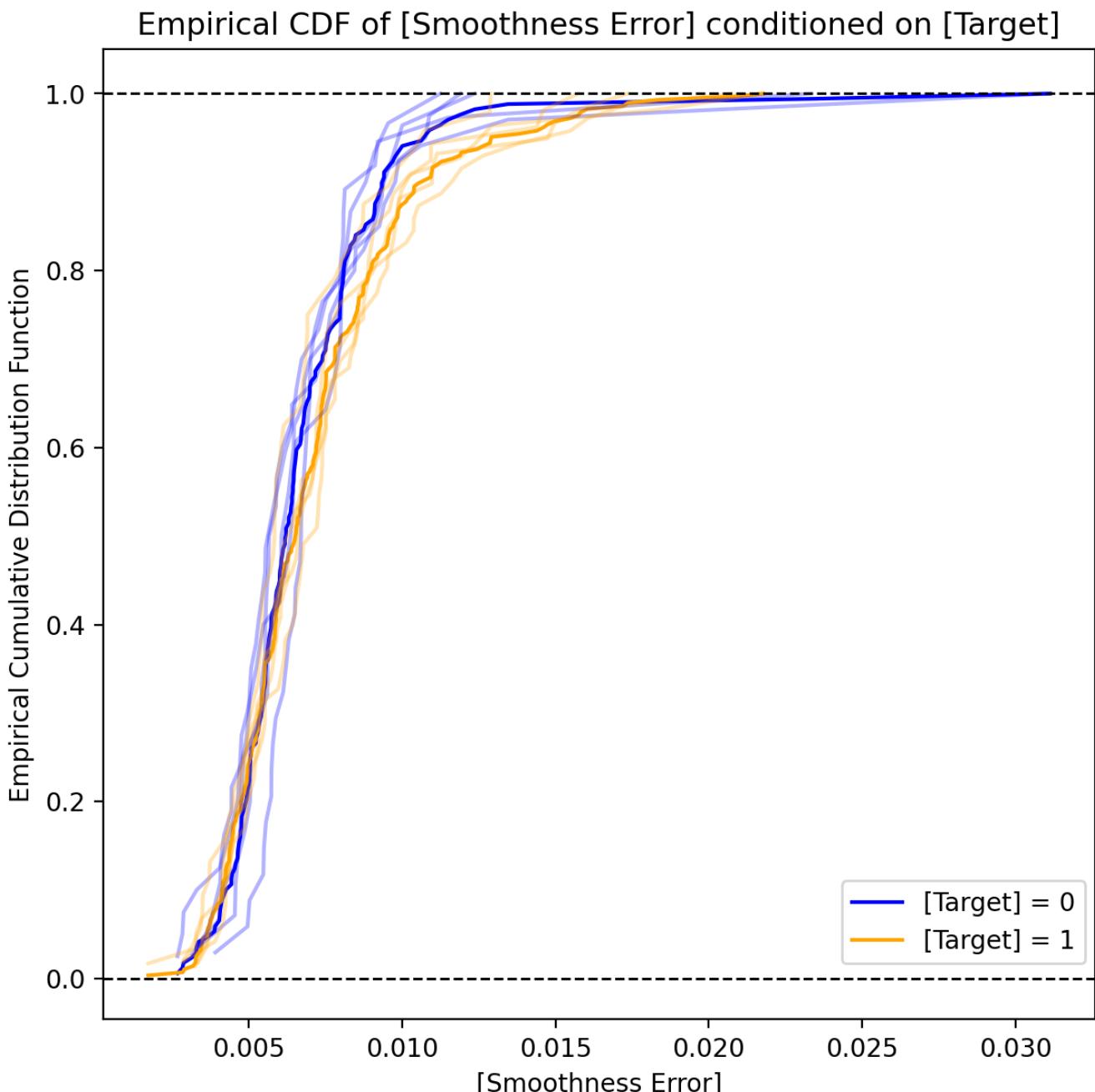
Kernel Density Plot of [Smoothness Error] by [Target]  
Distributions by level are not significantly different at the 95% level.



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

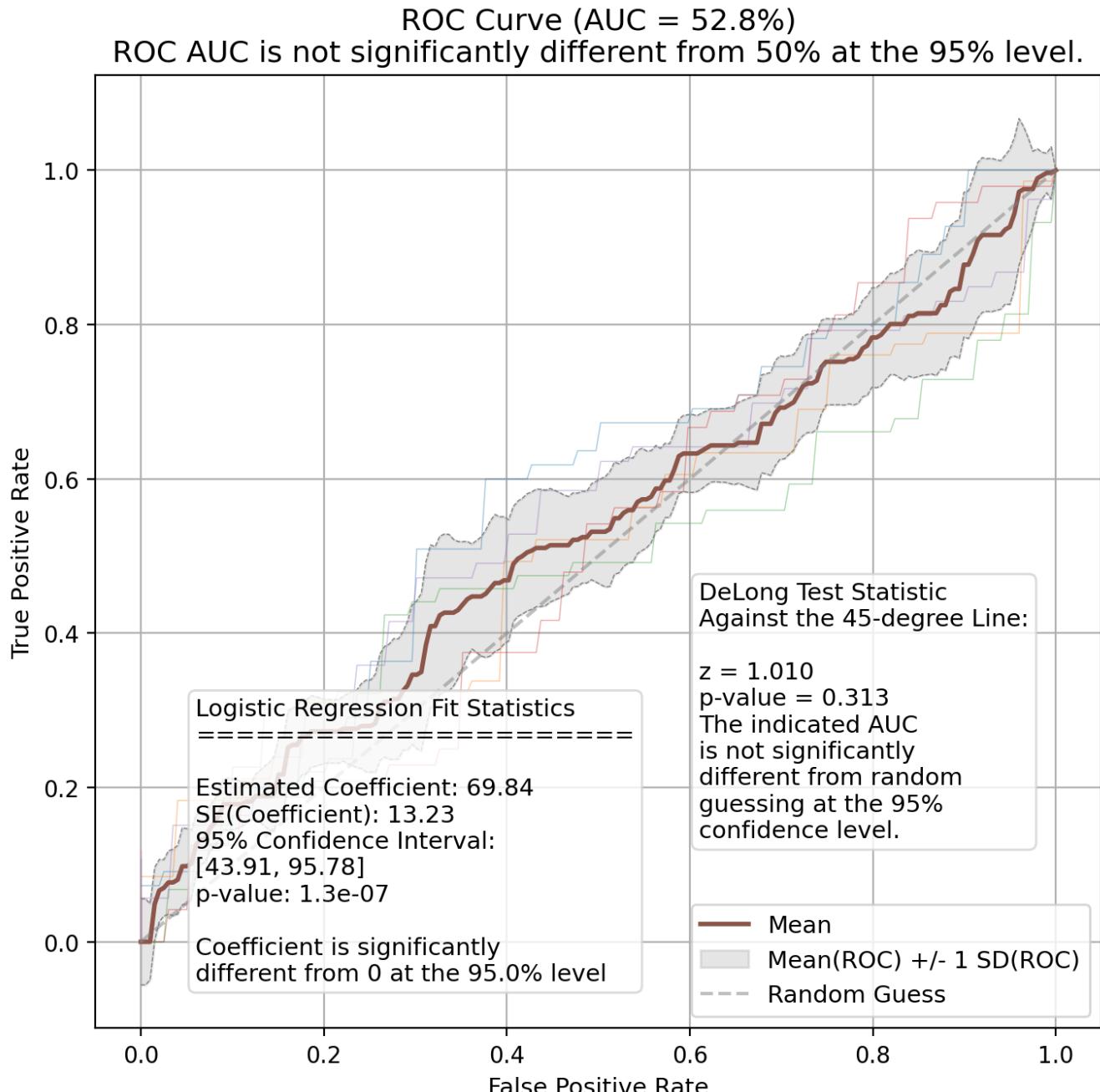
### Smoothness Error - Empirical CDF Plot



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



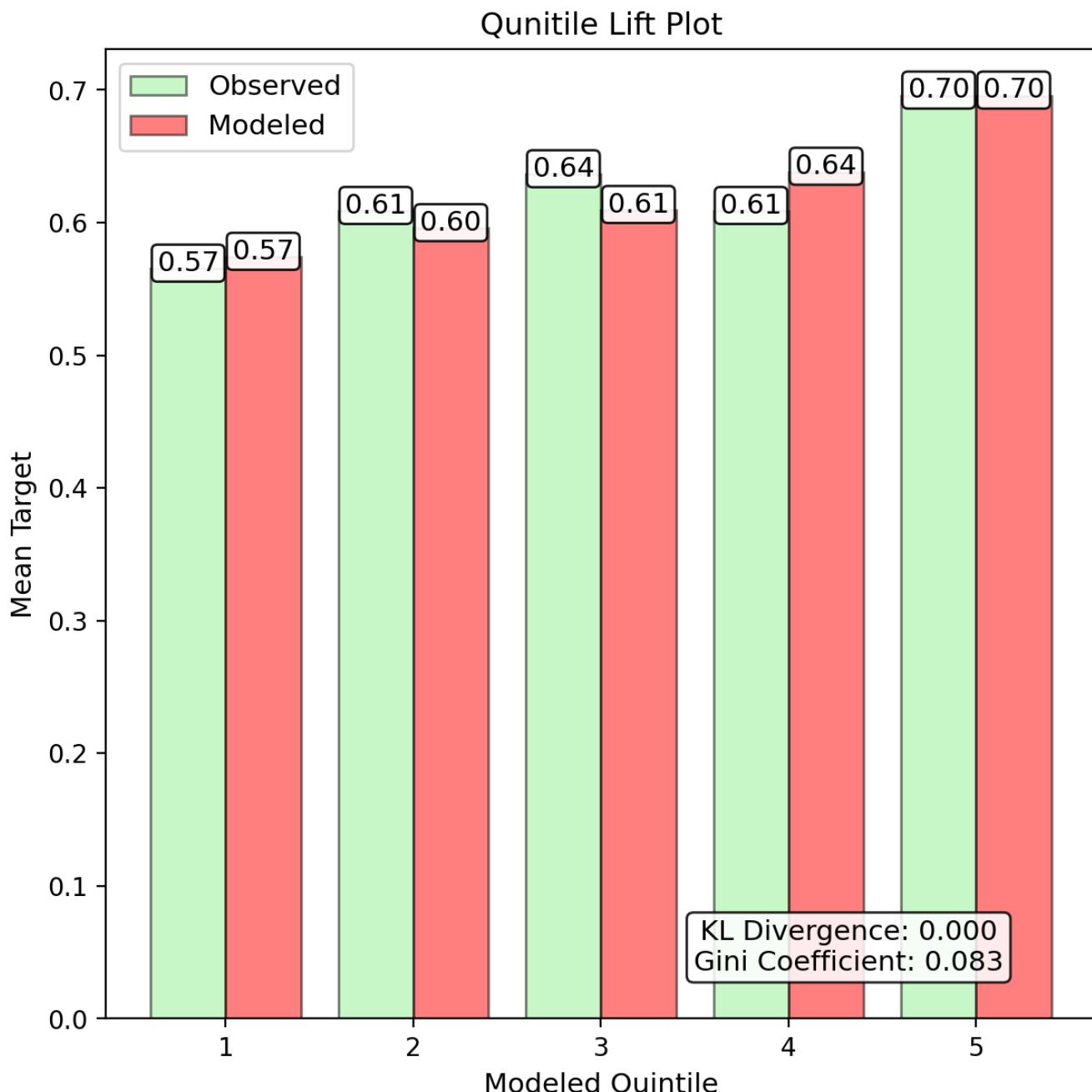
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 +/- 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

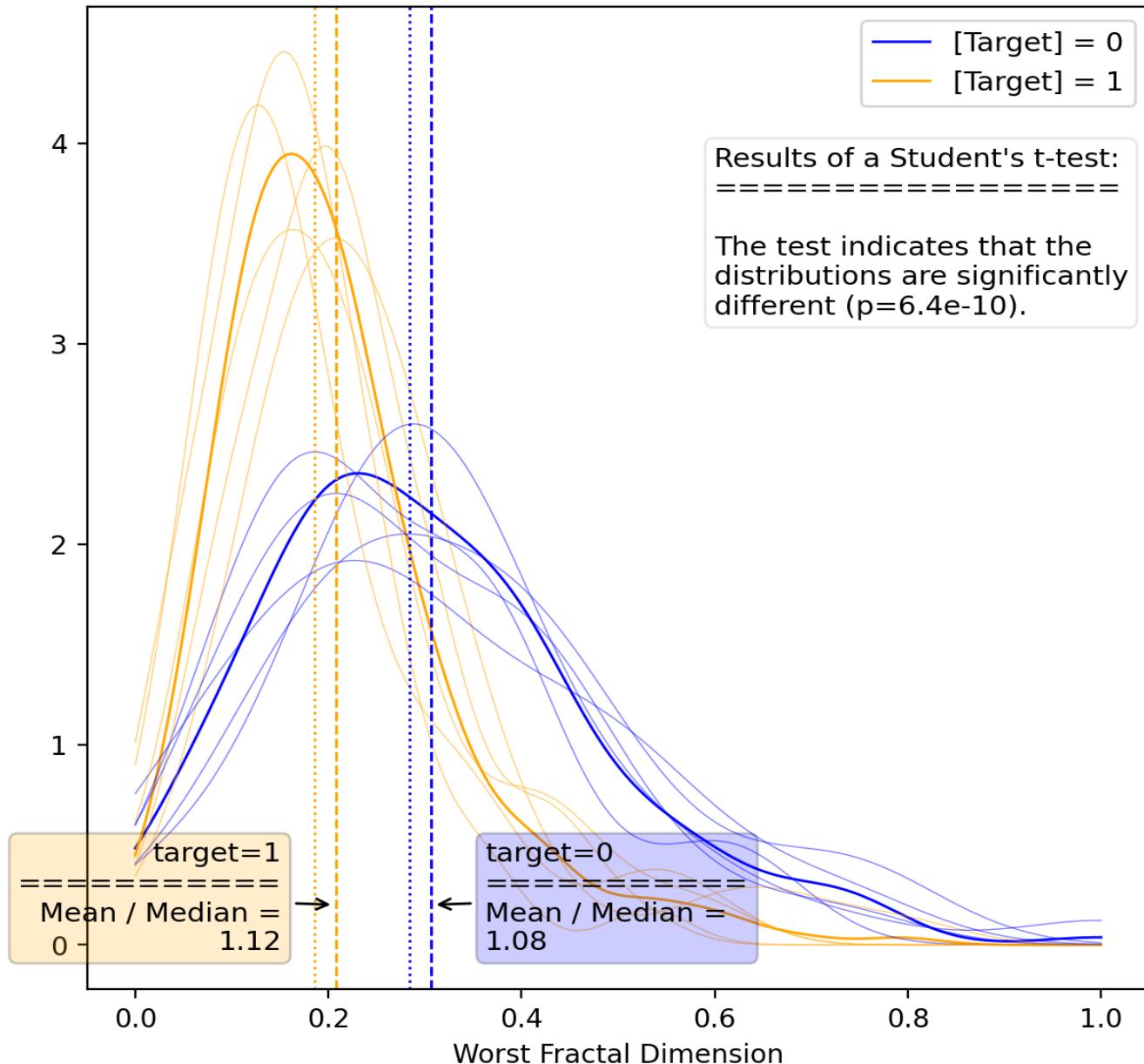
## Worst Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	0.327	0.144	0.372	0.723	0.464	0.408	0.212
<b>Fitted p-Value</b>	0.380	0.697	0.313	0.051	0.189	0.213	0.242
<b>Fitted Std. Err.</b>	0.372	0.370	0.369	0.370	0.354	0.328	0.007
<b>Conf. Int. Lower</b>	-4.0e-01	-5.8e-01	-3.5e-01	-3.1e-03	-2.3e-01	-2.3e-01	2.1e-01
<b>Conf. Int. Upper</b>	1.06	0.87	1.09	1.45	1.16	1.05	0.21
<b>Train Accuracy</b>	64.4%	60.7%	62.7%	64.6%	62.9%	63.1%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	64.5%	56.5%	63.9%	62.3%	6.1%
<b>Train AUC</b>	50.4%	50.4%	50.0%	50.4%	50.4%	50.3%	0.2%
<b>Val AUC</b>	50.0%	50.0%	51.5%	50.0%	50.0%	50.0%	0.7%
<b>Train F1</b>	78.3%	75.4%	77.1%	78.4%	77.2%	77.3%	1.2%
<b>Test F1</b>	73.3%	83.5%	78.1%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.3%	60.6%	62.7%	64.5%	62.8%	63.0%	1.6%
<b>Val Precision</b>	57.9%	71.7%	64.1%	56.5%	63.9%	62.3%	6.1%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	7.1%	6.6%	0.0%	7.0%	6.7%	6.1%	3.1%
<b>Val MCC</b>	0.0%	0.0%	13.7%	0.0%	0.0%	0.0%	6.1%
<b>Train Log-Loss</b>	12.82	14.17	13.44	12.76	13.37	13.31	0.57
<b>Val Log-Loss</b>	15.18	10.19	12.79	15.69	13.03	13.60	2.19

## Univariate Report

### Worst Fractal Dimension - Kernel Density Plot

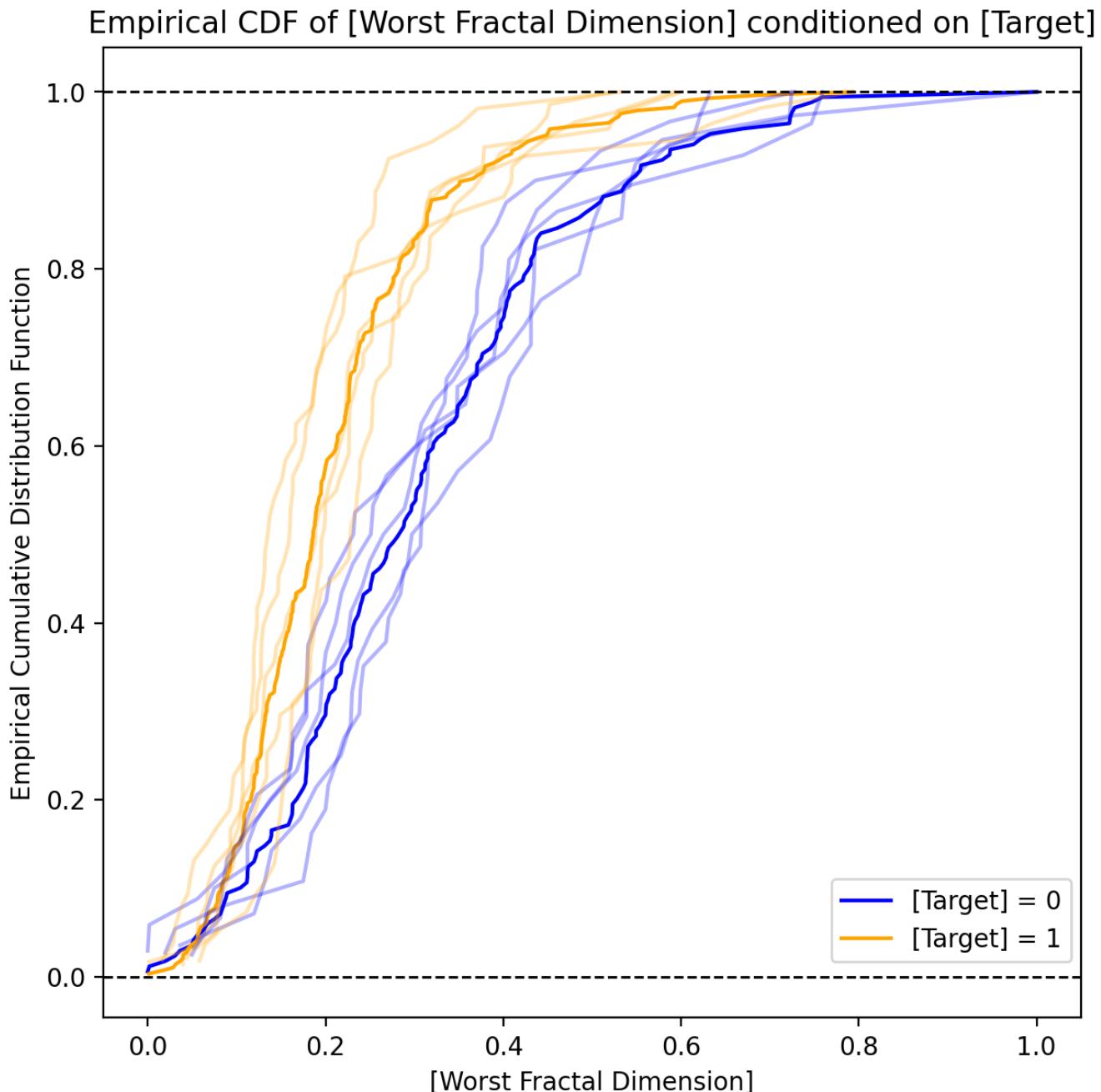
Kernel Density Plot of [Worst Fractal Dimension] by [Target]  
Distributions by level are significantly different at the 95% level.



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

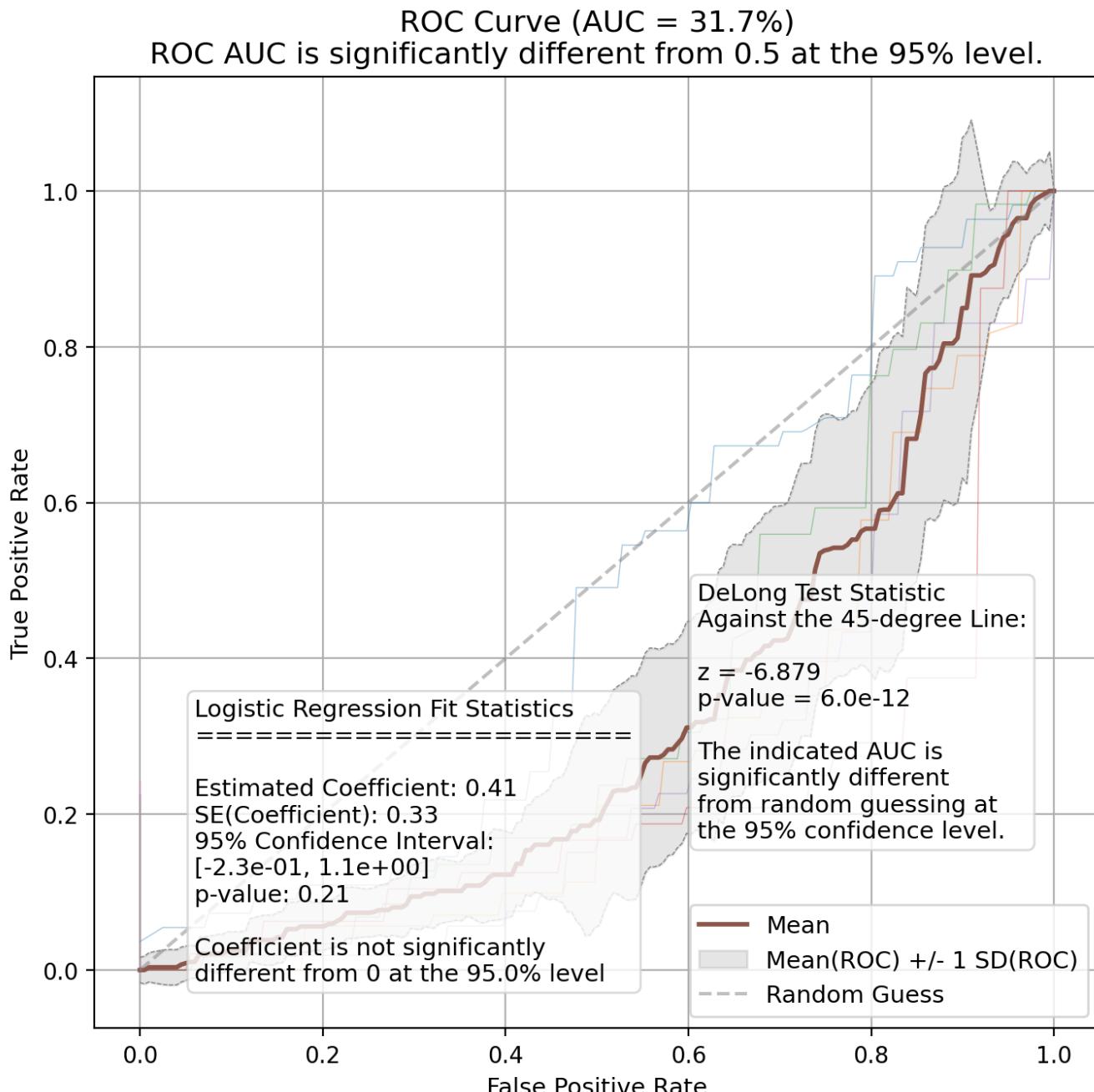
### Worst Fractal Dimension - Empirical CDF Plot



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



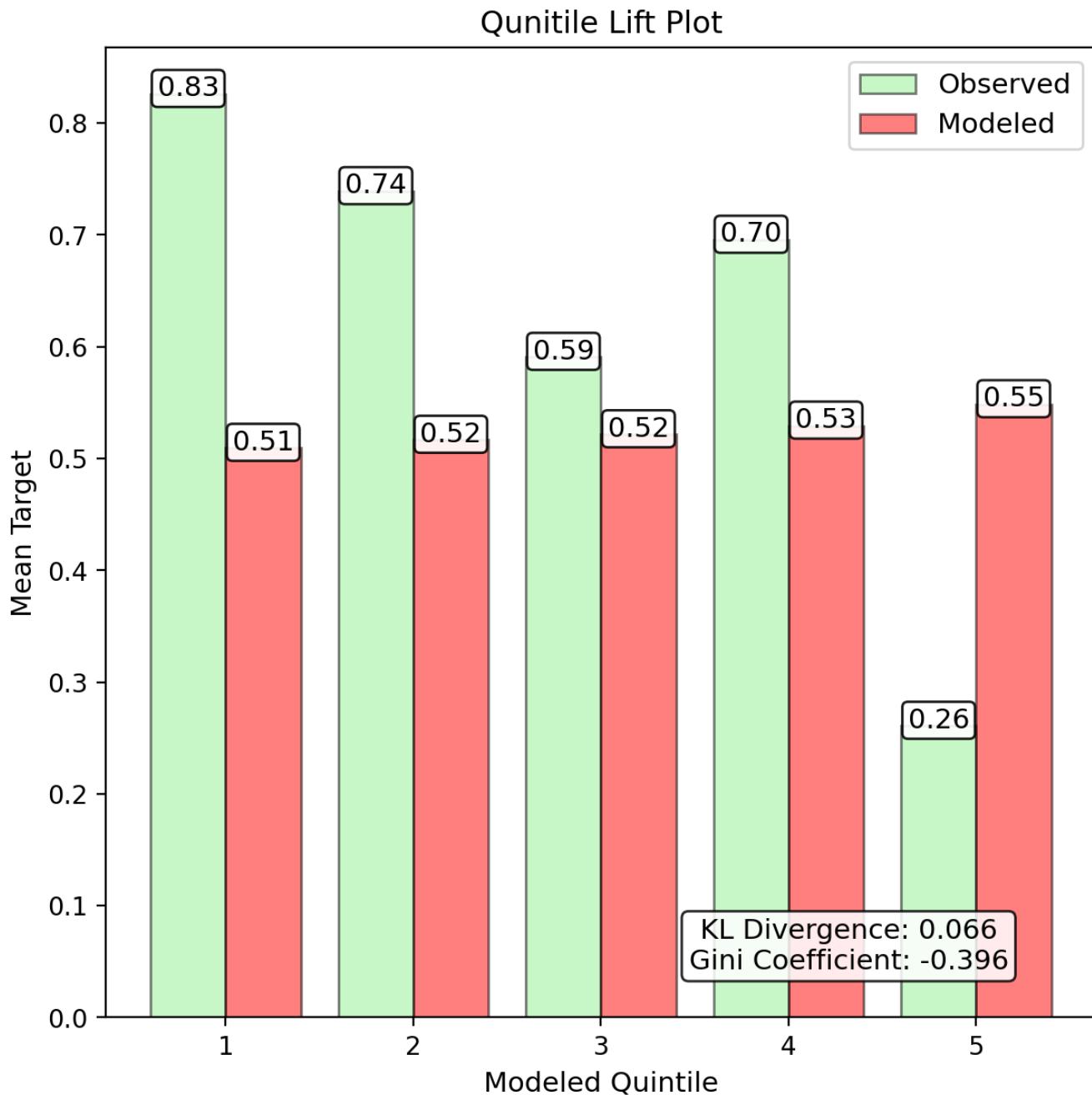
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

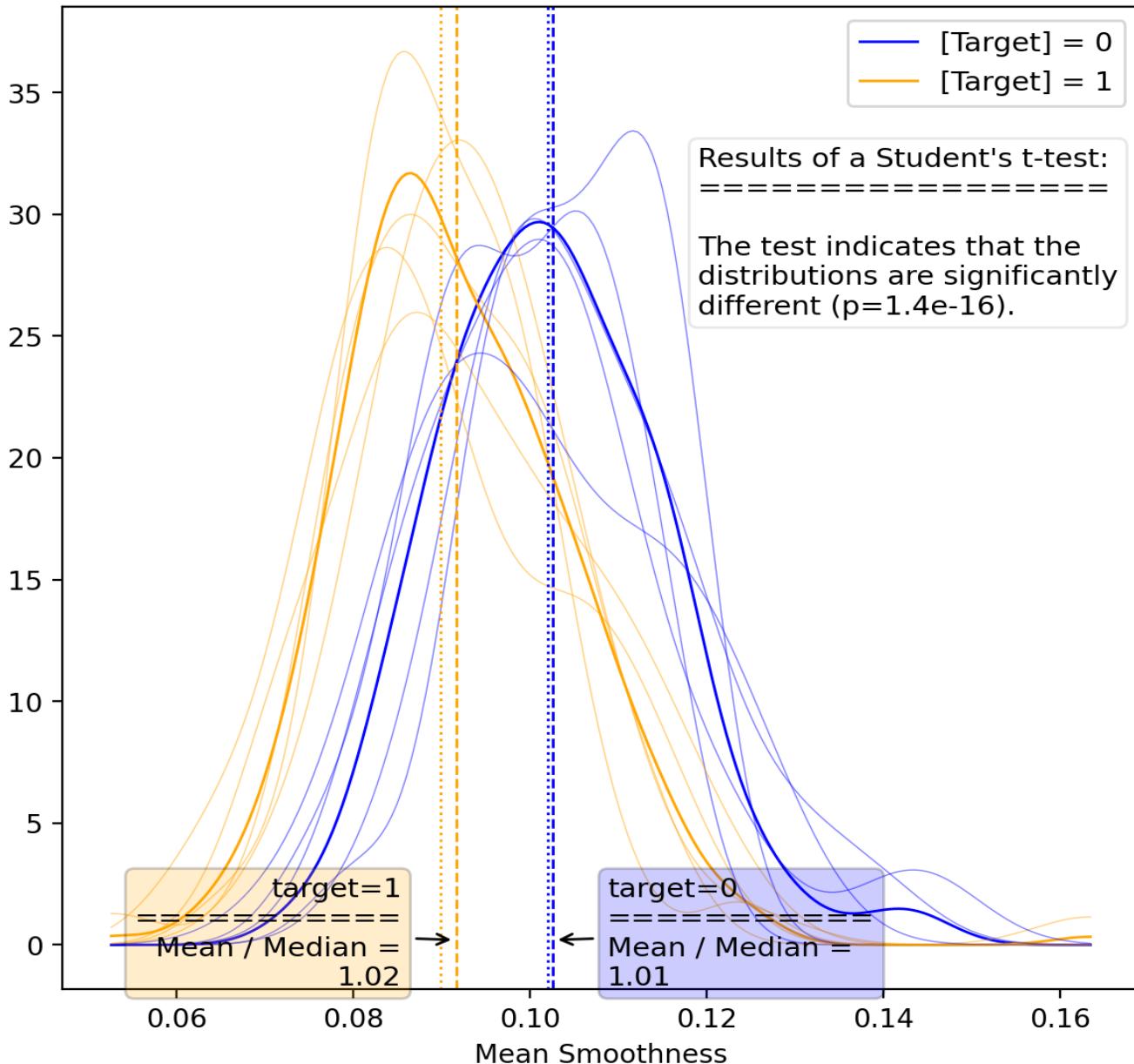
### Mean Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	4.55	3.12	4.22	4.82	4.07	4.16	0.65
<b>Fitted p-Value</b>	4.9e-05	5.0e-03	1.5e-04	1.3e-05	1.8e-04	2.7e-05	2.2e-03
<b>Fitted Std. Err.</b>	1.12	1.11	1.11	1.11	1.09	0.99	0.01
<b>Conf. Int. Lower</b>	2.35	0.94	2.03	2.66	1.94	2.22	0.65
<b>Conf. Int. Upper</b>	6.74	5.30	6.40	6.99	6.21	6.10	0.65
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

Mean Smoothness - Kernel Density Plot

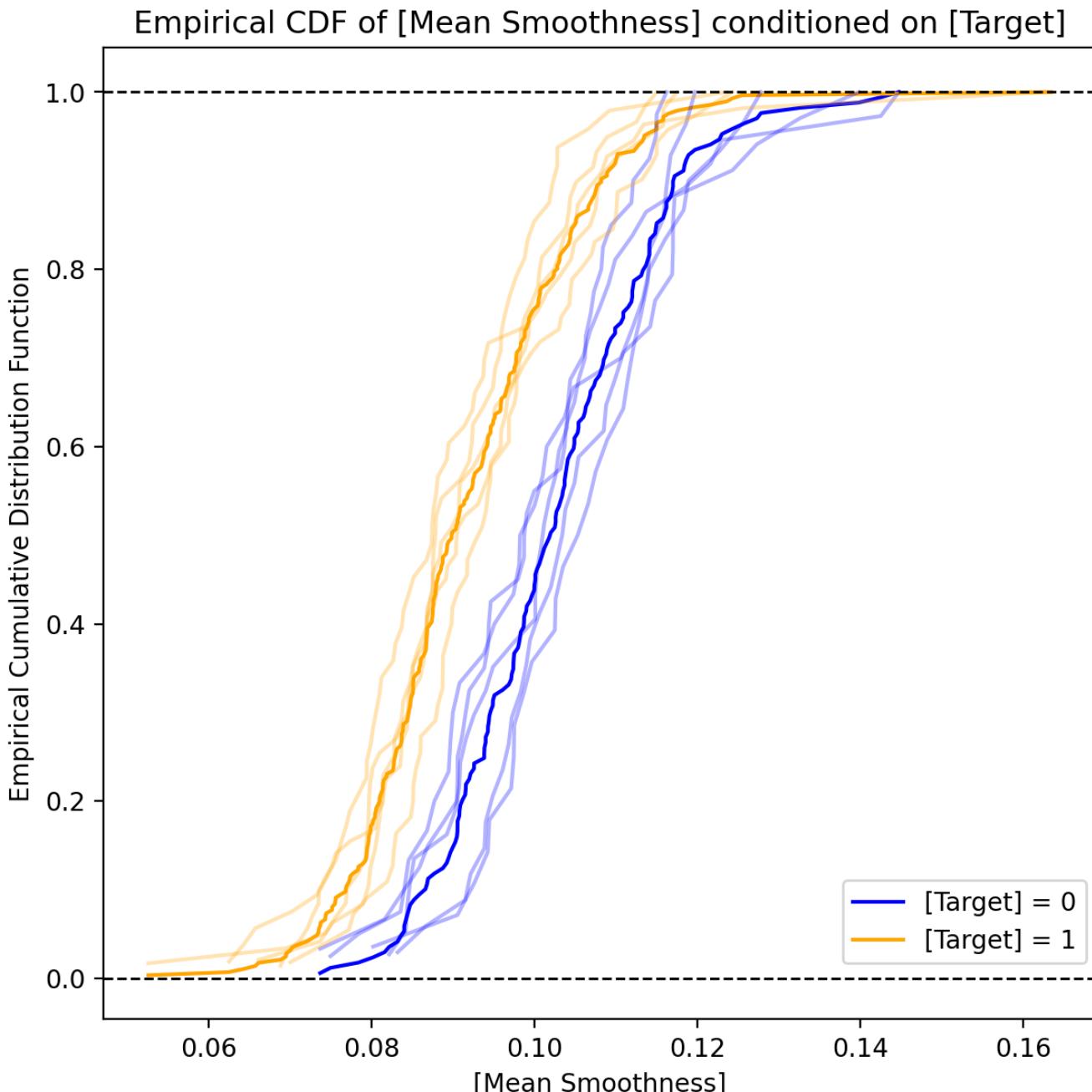
Kernel Density Plot of [Mean Smoothness] by [Target]  
Distributions by level are significantly different at the 95% level.



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



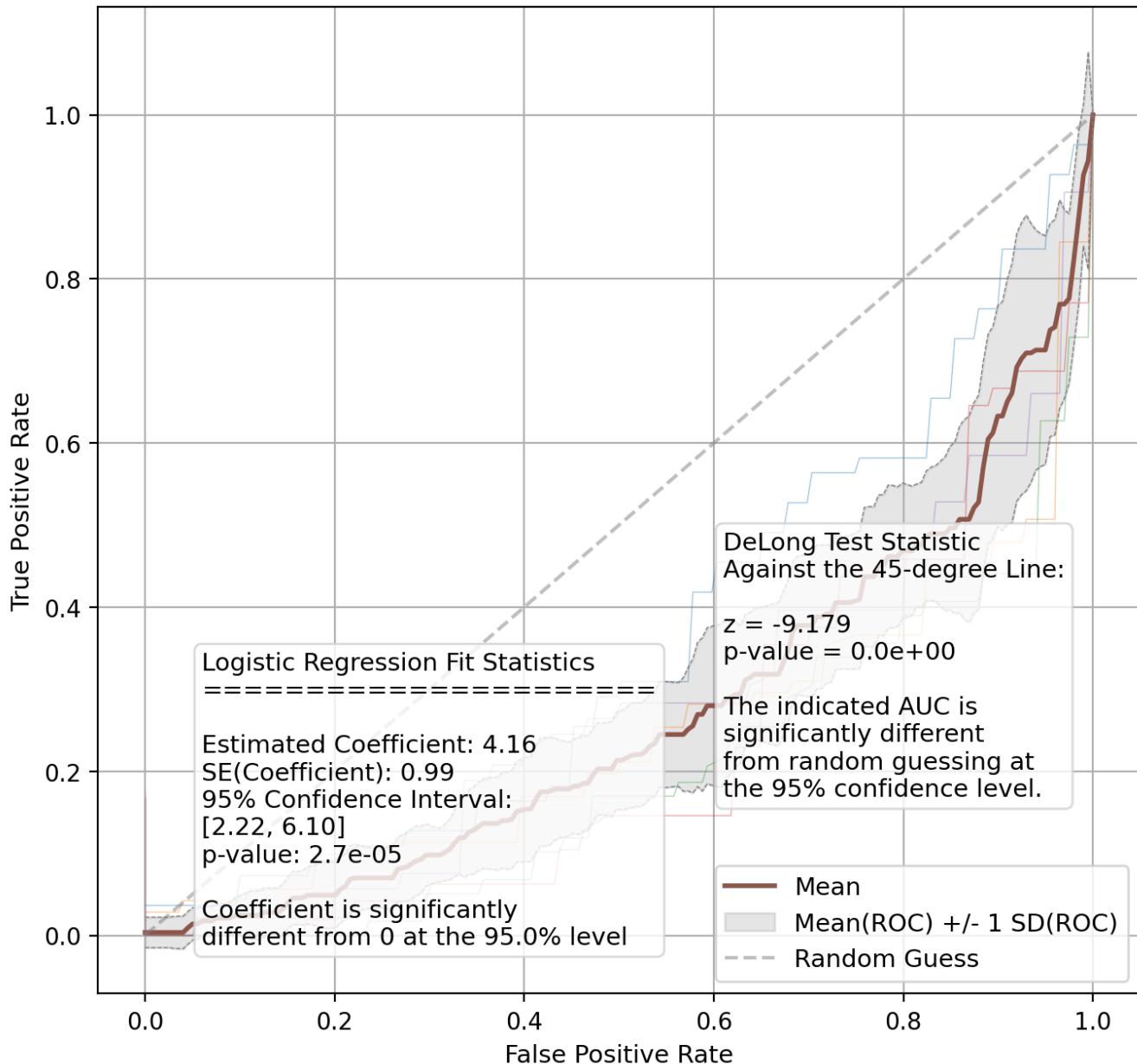
This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly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

ROC Curve (AUC = 26.8%)

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



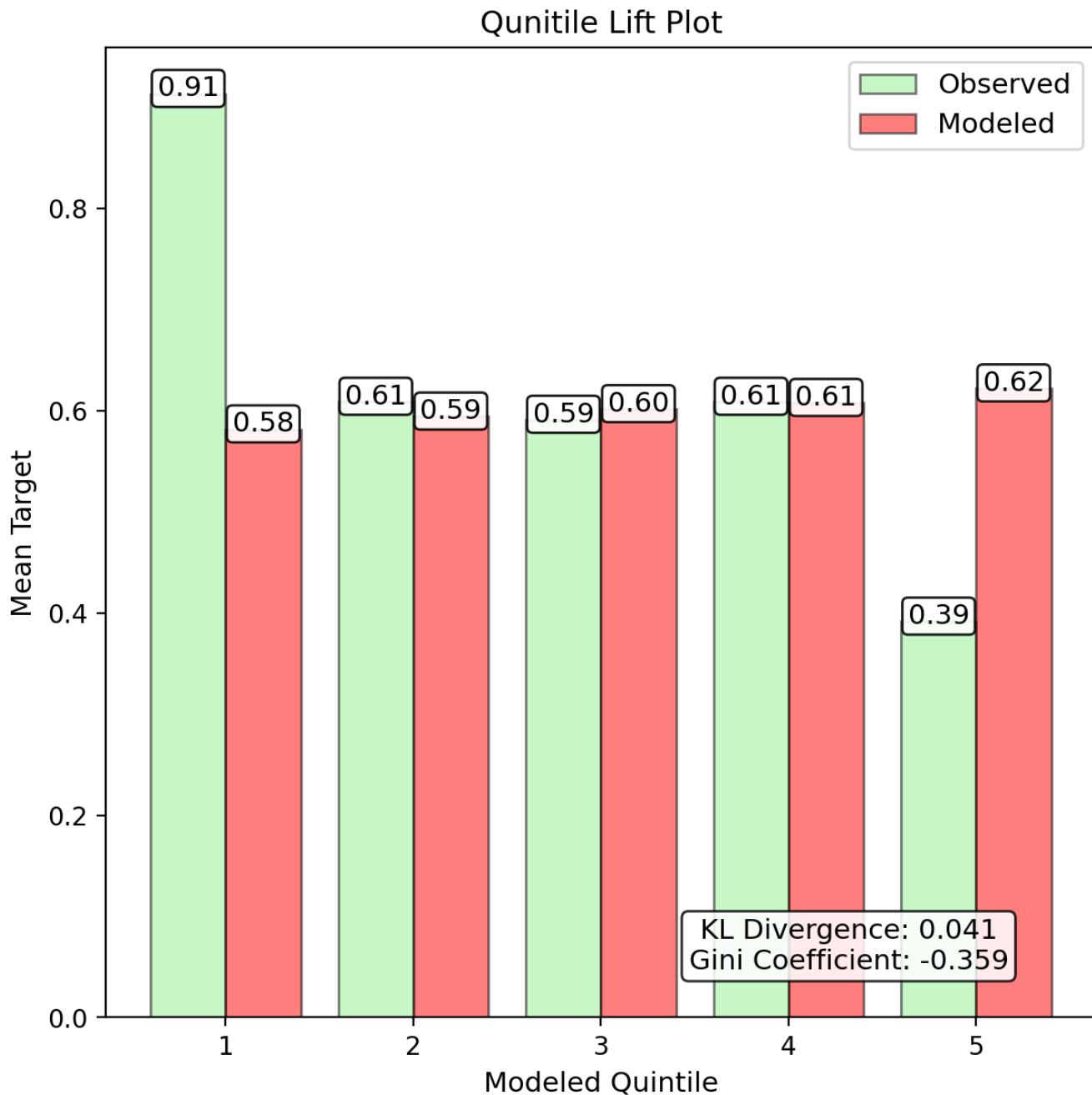
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

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

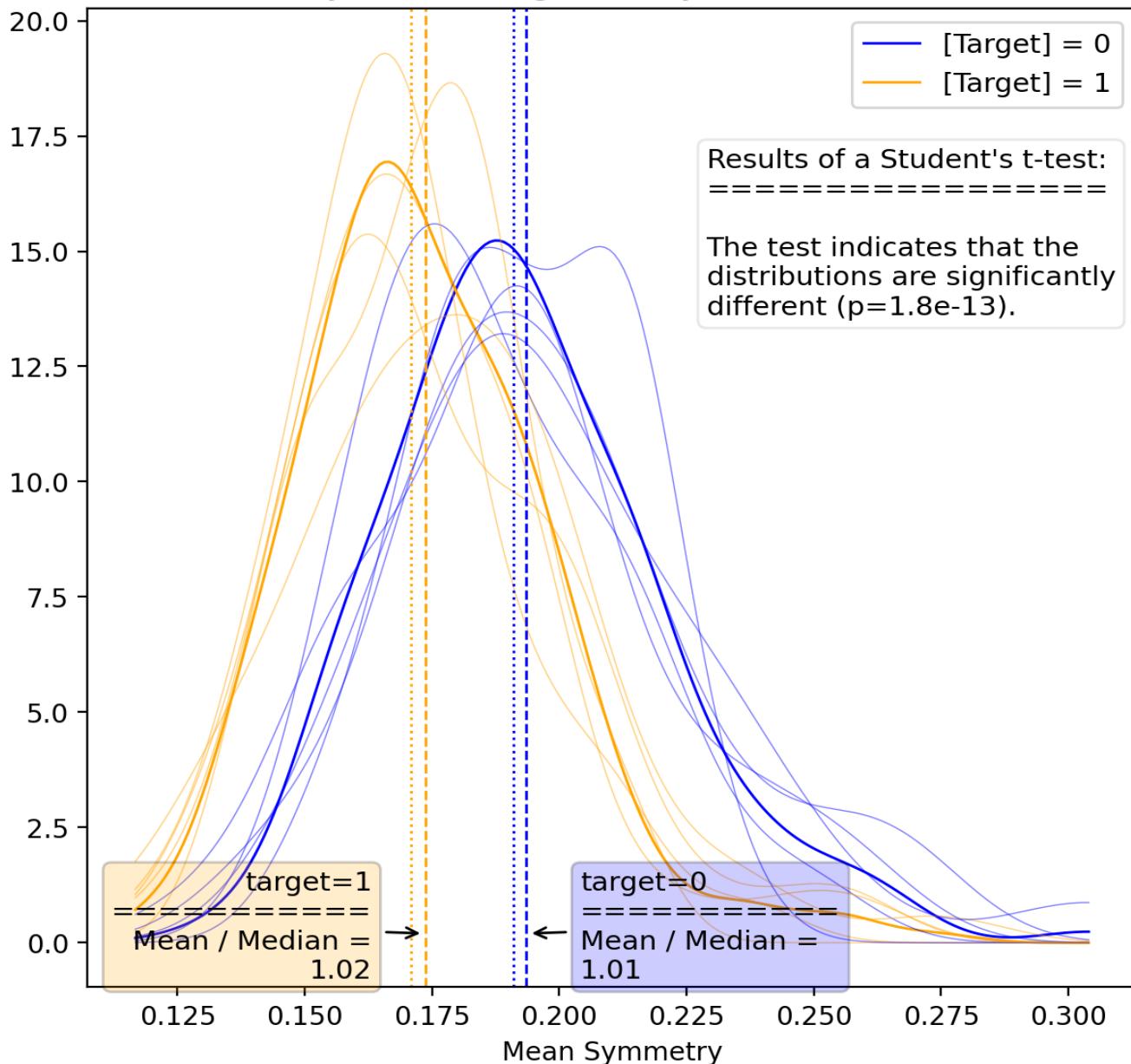
### Mean Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	2.43	1.70	2.15	2.61	2.16	2.21	0.35
<b>Fitted p-Value</b>	3.9e-05	3.7e-03	2.4e-04	8.2e-06	1.8e-04	2.4e-05	1.6e-03
<b>Fitted Std. Err.</b>	0.592	0.586	0.586	0.586	0.578	0.523	0.005
<b>Conf. Int. Lower</b>	1.27	0.55	1.00	1.47	1.03	1.19	0.34
<b>Conf. Int. Upper</b>	3.59	2.85	3.30	3.76	3.30	3.24	0.35
<b>Train Accuracy</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Val AUC</b>	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	0.0%
<b>Train F1</b>	78.2%	75.3%	77.1%	78.3%	77.0%	77.2%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.2%	77.9%	76.8%	4.5%
<b>Train Precision</b>	64.2%	60.4%	62.7%	64.3%	62.6%	62.9%	1.6%
<b>Val Precision</b>	57.9%	71.7%	63.4%	56.5%	63.9%	62.3%	6.0%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<b>Train Log-Loss</b>	12.92	14.28	13.44	12.86	13.47	13.39	0.57
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.69	13.03	13.60	2.17

## Univariate Report

### Mean Symmetry - Kernel Density Plot

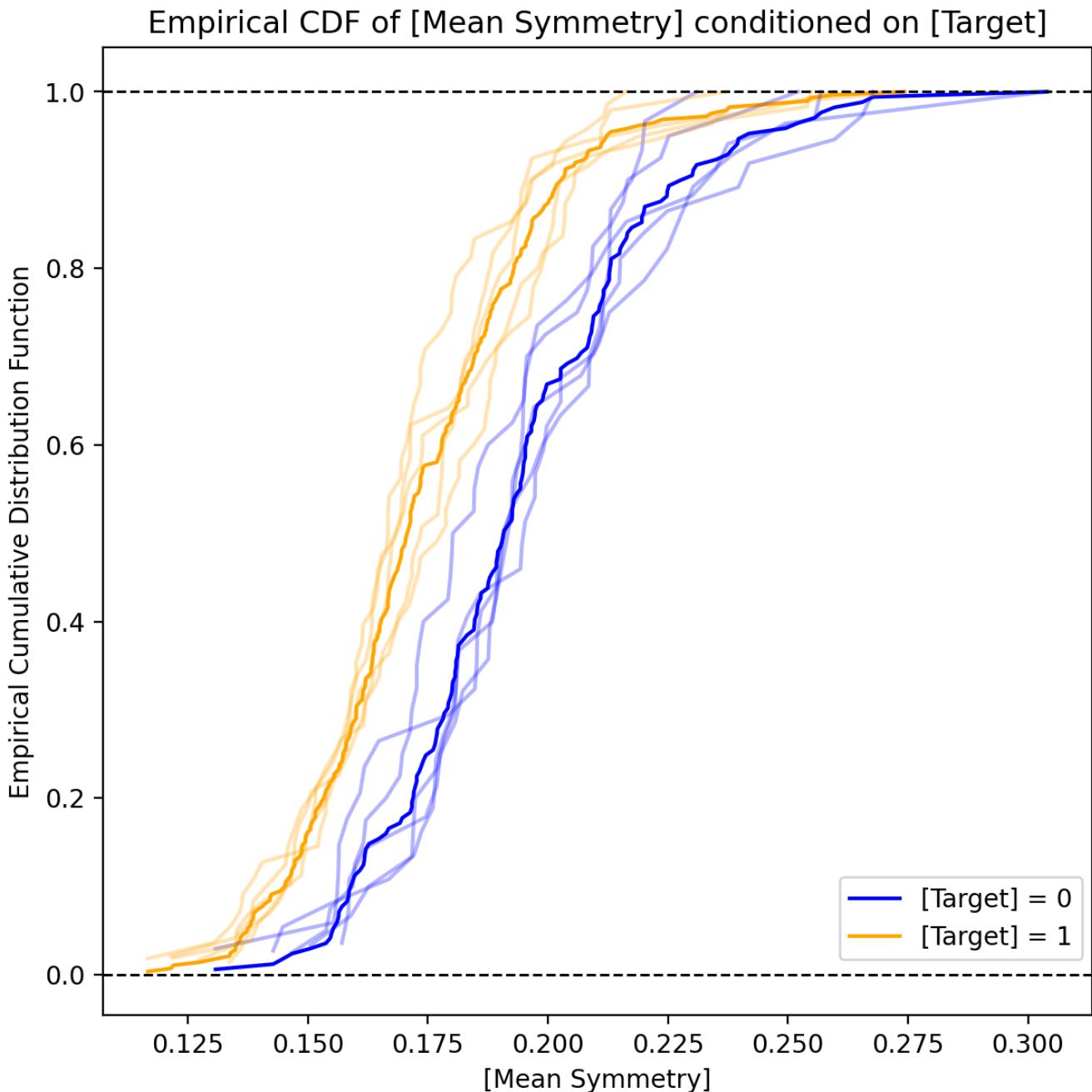
Kernel Density Plot of [Mean Symmetry] by [Target]  
Distributions by level are significantly different at the 95% level.



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



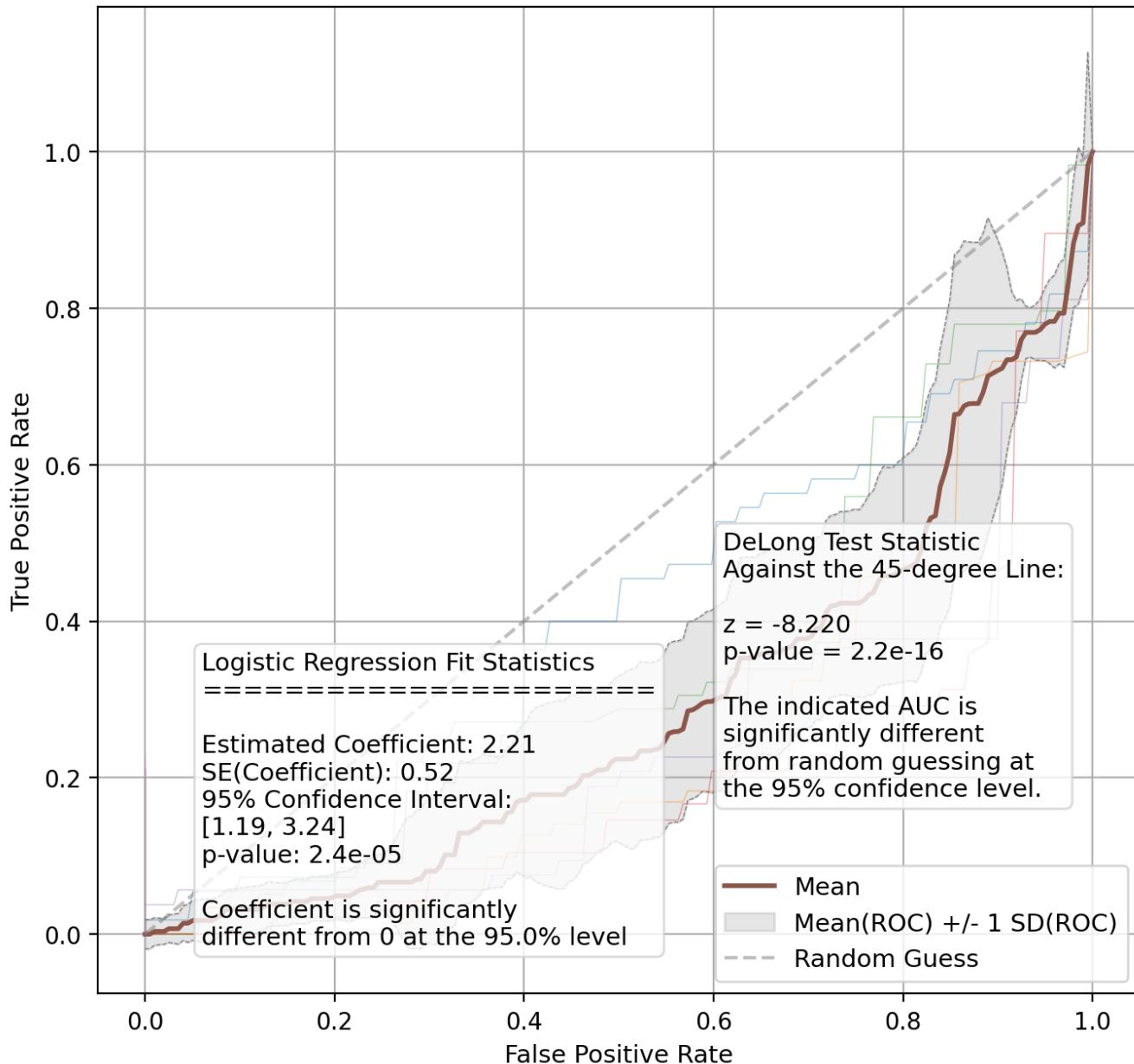
This plot shows the empirical cumulative distribution function for each level of the target variable, both in total and for each fold. The x-axis represents the feature variable, and the y-axis represents the cumulative distribution of the target variable. The cross-validation folds are included in slightly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

ROC Curve (AUC = 28.8%)

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



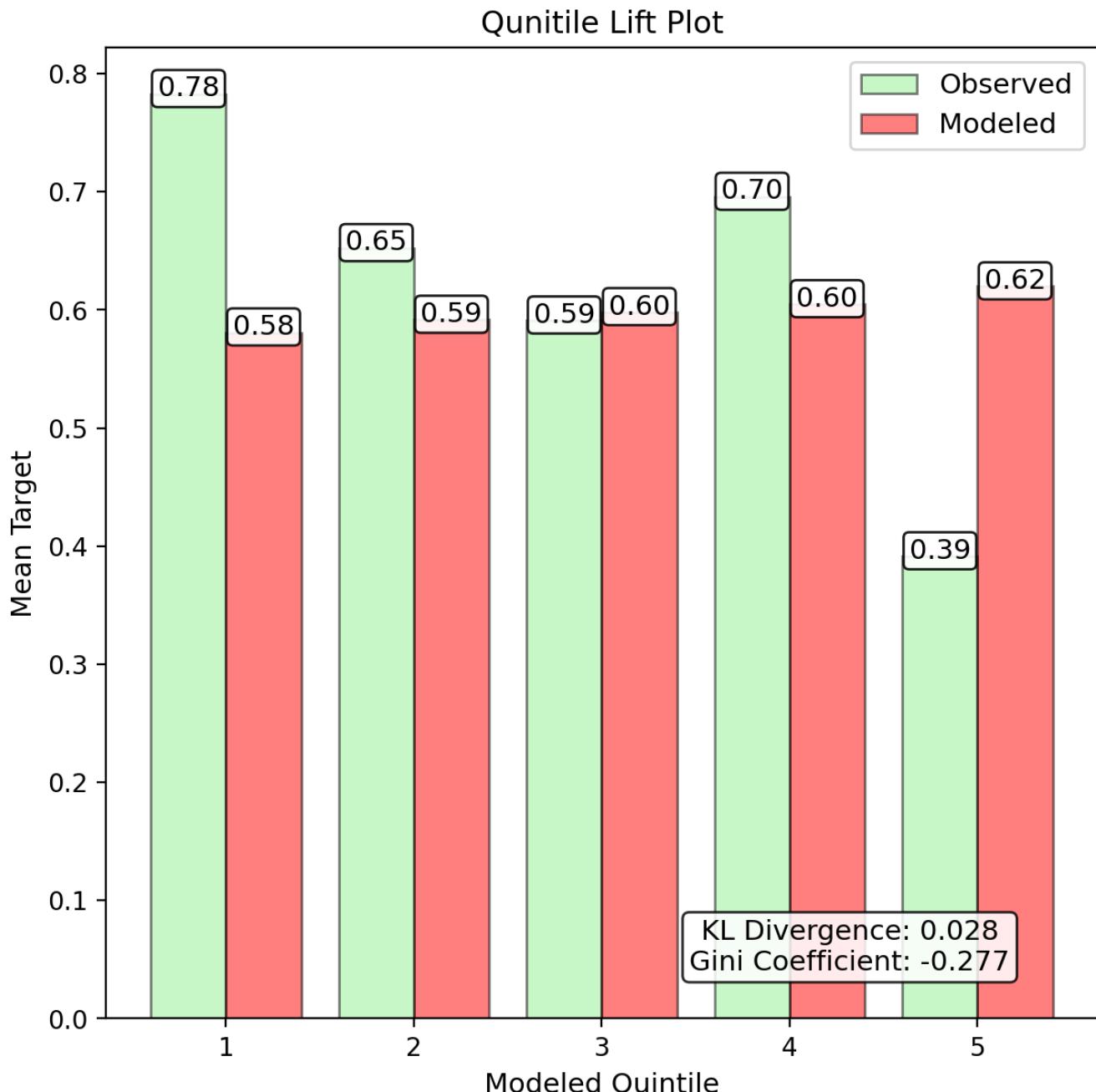
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

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

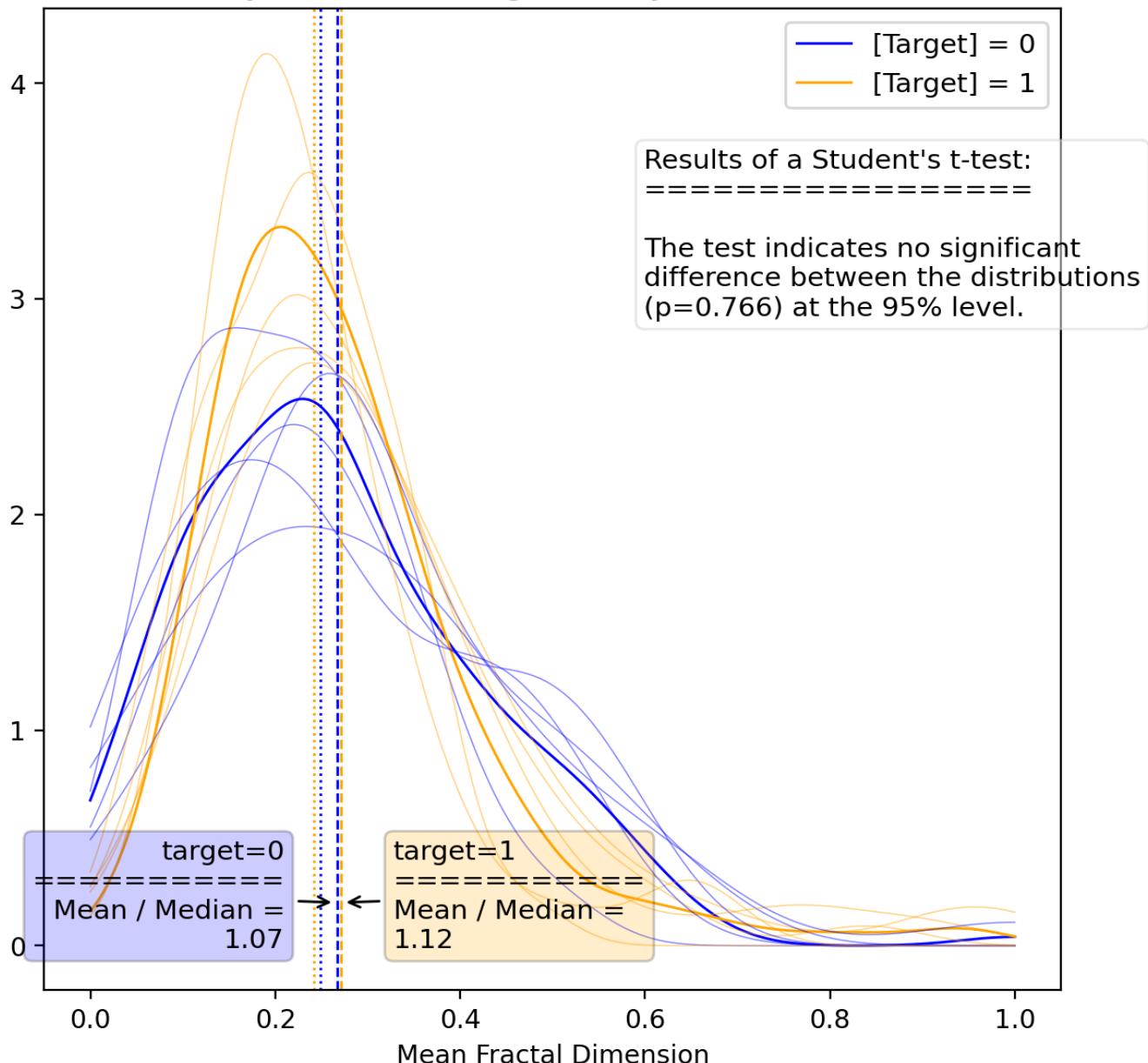
### Mean Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	1.43	1.28	1.54	1.88	1.52	1.53	0.22
<b>Fitted p-Value</b>	7.8e-05	4.3e-04	3.0e-05	3.4e-07	1.3e-05	2.3e-06	1.8e-04
<b>Fitted Std. Err.</b>	0.362	0.363	0.369	0.369	0.349	0.324	0.008
<b>Conf. Int. Lower</b>	0.721	0.566	0.816	1.159	0.835	0.896	0.217
<b>Conf. Int. Upper</b>	2.14	1.99	2.26	2.60	2.20	2.17	0.23
<b>Train Accuracy</b>	64.4%	60.7%	63.0%	64.3%	62.9%	63.1%	1.5%
<b>Val Accuracy</b>	57.9%	71.7%	63.4%	57.6%	63.9%	62.3%	5.7%
<b>Train AUC</b>	50.4%	50.4%	50.4%	50.0%	50.4%	50.3%	0.2%
<b>Val AUC</b>	50.0%	50.0%	50.0%	51.4%	50.0%	50.0%	0.6%
<b>Train F1</b>	78.3%	75.4%	77.2%	78.3%	77.2%	77.3%	1.2%
<b>Test F1</b>	73.3%	83.5%	77.6%	72.7%	77.9%	76.8%	4.3%
<b>Train Precision</b>	64.3%	60.6%	62.9%	64.3%	62.8%	63.0%	1.5%
<b>Val Precision</b>	57.9%	71.7%	63.4%	57.1%	63.9%	62.3%	5.9%
<b>Train Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Val Recall</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	0.0%
<b>Train MCC</b>	7.1%	6.6%	6.8%	0.0%	6.7%	6.1%	3.0%
<b>Val MCC</b>	0.0%	0.0%	0.0%	12.4%	0.0%	0.0%	5.6%
<b>Train Log-Loss</b>	12.82	14.17	13.34	12.86	13.37	13.31	0.55
<b>Val Log-Loss</b>	15.18	10.19	13.18	15.27	13.03	13.60	2.07

## Univariate Report

### Mean Fractal Dimension - Kernel Density Plot

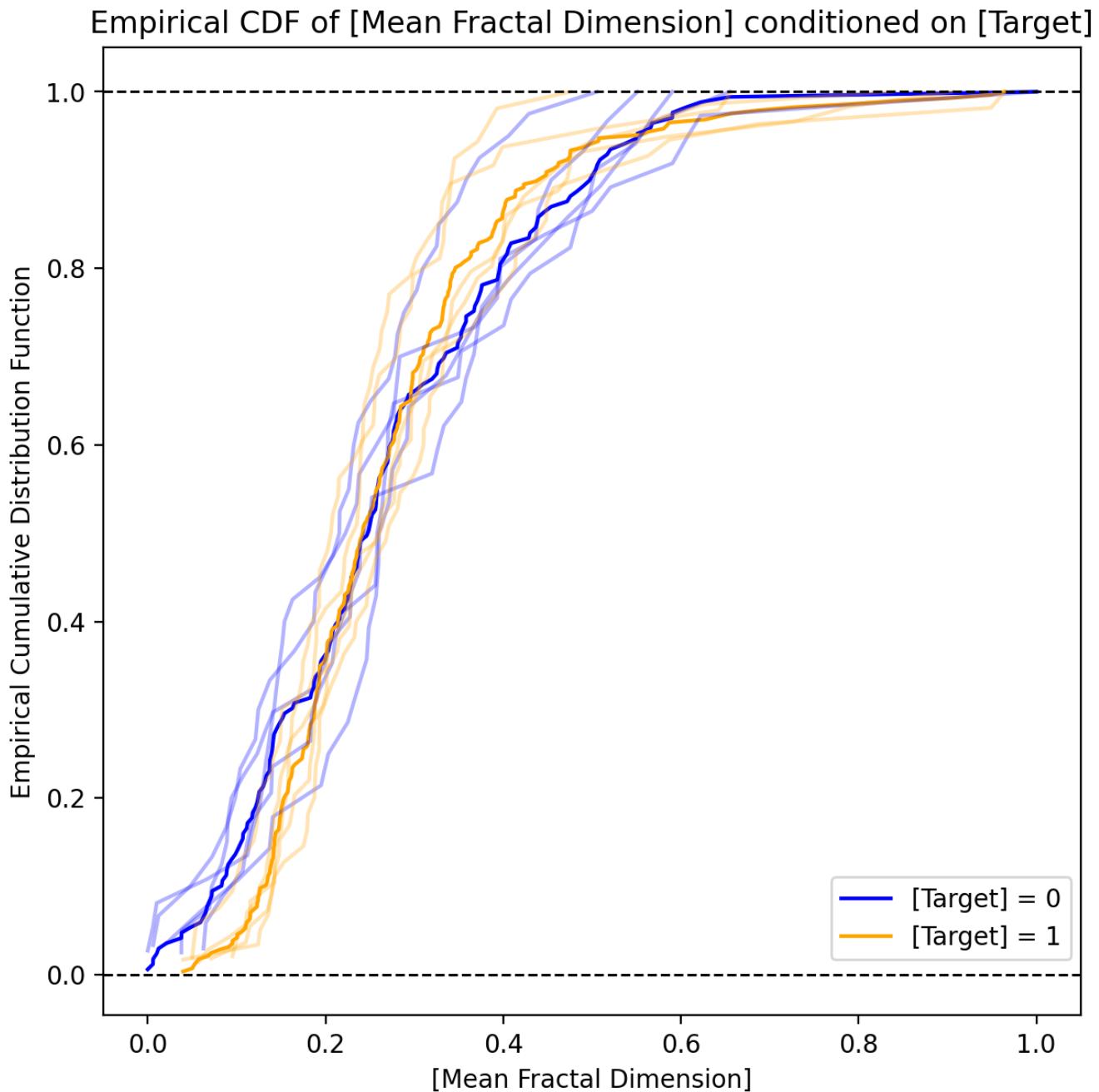
Kernel Density Plot of [Mean Fractal Dimension] by [Target]  
Distributions by level are not significantly different at the 95% level.



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

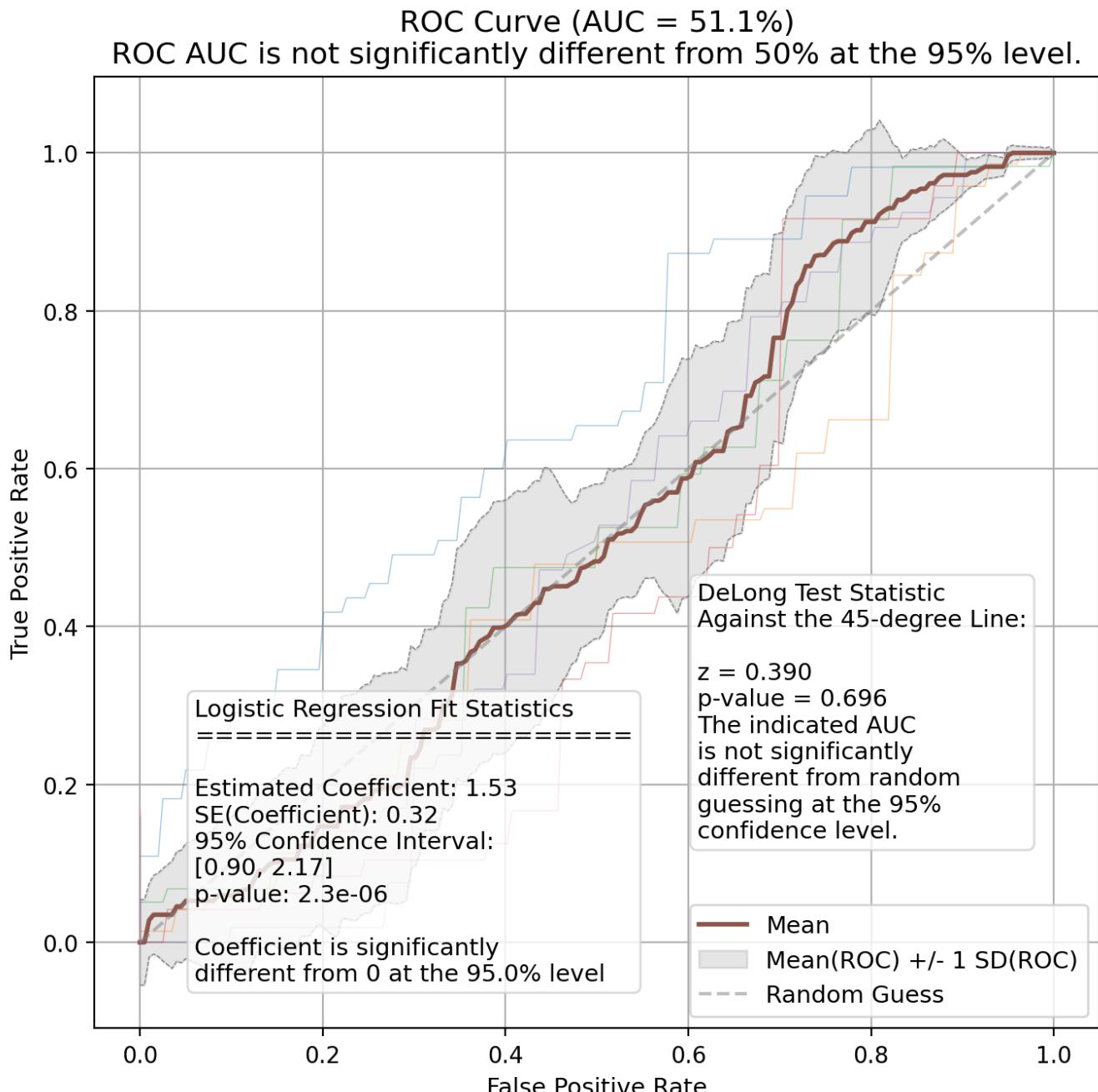
### Mean Fractal Dimension - Empirical CDF Plot



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



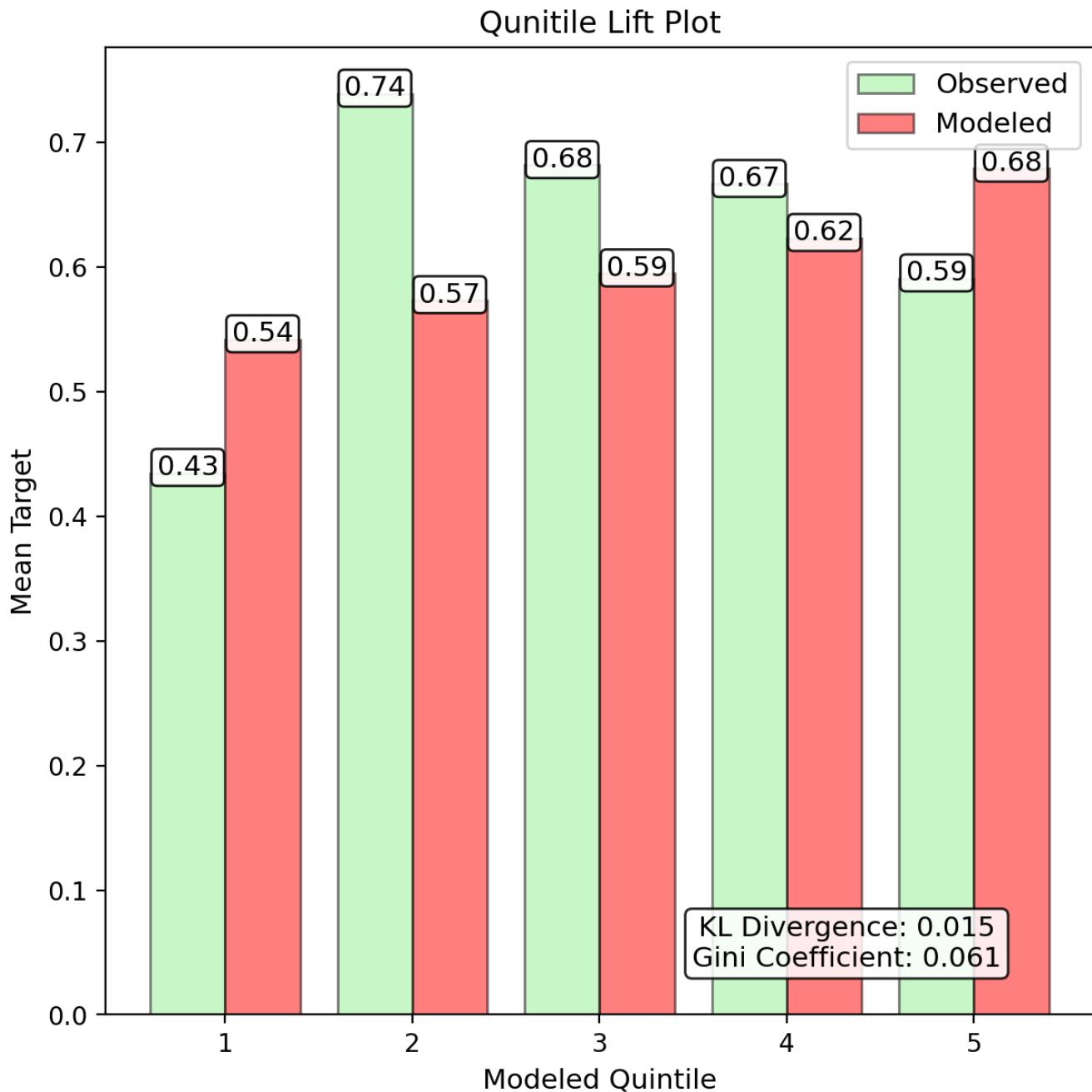
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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  $\pm$  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

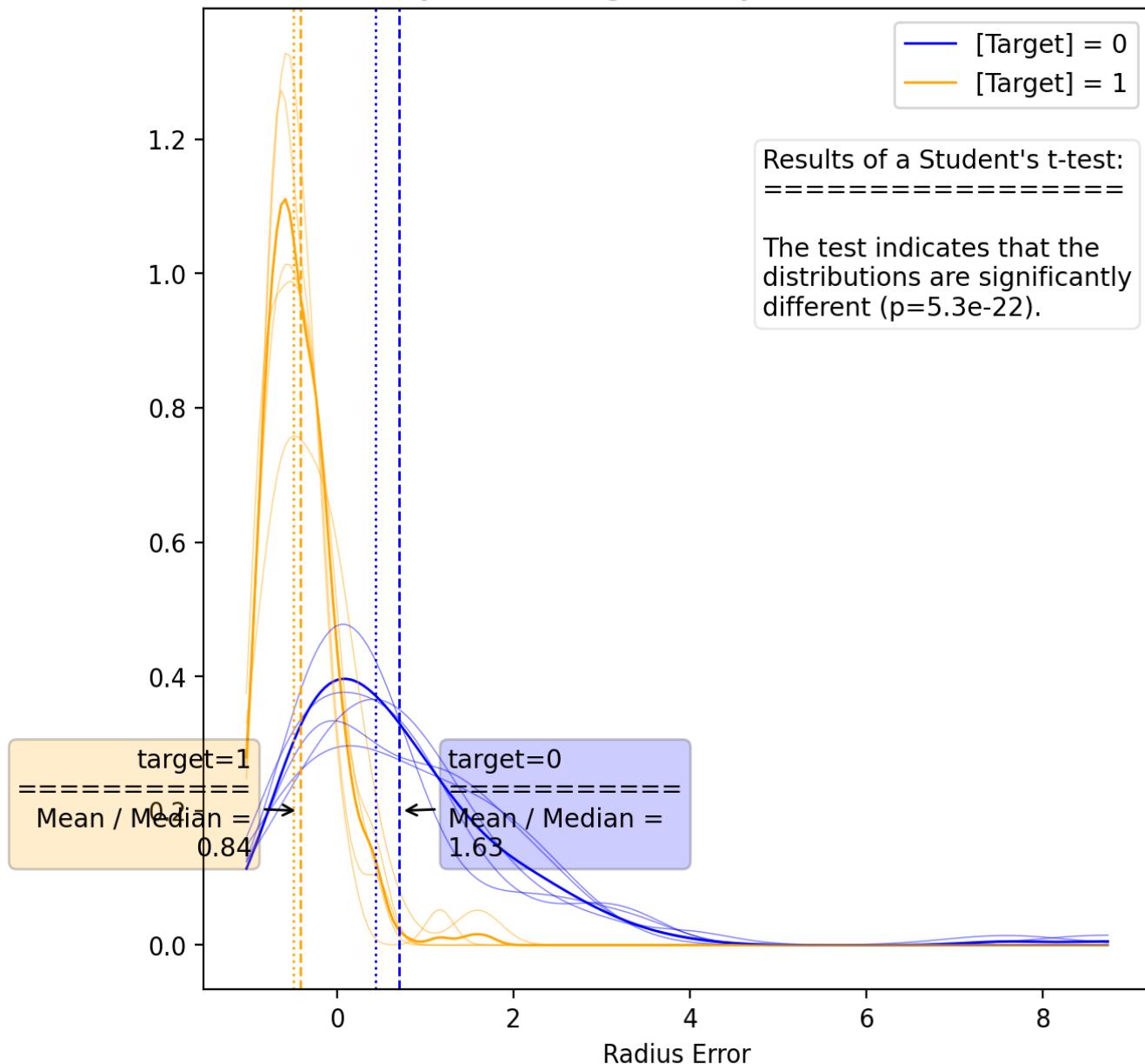
## Radius Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	-2.7e+00	-2.6e+00	-2.9e+00	-2.7e+00	-2.6e+00	-2.7e+00	1.1e-01
<b>Fitted p-Value</b>	1.4e-20	7.6e-20	7.0e-22	4.4e-21	1.9e-21	8.4e-26	3.2e-20
<b>Fitted Std. Err.</b>	0.291	0.285	0.301	0.286	0.277	0.258	0.009
<b>Conf. Int. Lower</b>	-3.3e+00	-3.2e+00	-3.5e+00	-3.3e+00	-3.2e+00	-3.2e+00	1.3e-01
<b>Conf. Int. Upper</b>	-2.1e+00	-2.0e+00	-2.3e+00	-2.1e+00	-2.1e+00	-2.2e+00	9.8e-02
<b>Train Accuracy</b>	79.7%	80.1%	81.5%	80.0%	80.9%	80.4%	0.7%
<b>Val Accuracy</b>	83.2%	81.8%	76.3%	82.4%	78.3%	37.7%	2.9%
<b>Train AUC</b>	77.2%	78.4%	78.6%	77.7%	78.8%	78.1%	0.7%
<b>Val AUC</b>	81.7%	76.5%	76.4%	80.7%	75.1%	50.0%	2.9%
<b>Train F1</b>	84.5%	84.0%	85.9%	84.6%	85.1%	84.8%	0.7%
<b>Test F1</b>	86.2%	87.5%	80.4%	85.7%	83.6%	0.0%	2.8%
<b>Train Precision</b>	82.9%	81.6%	82.3%	83.6%	83.2%	82.7%	0.8%
<b>Val Precision</b>	82.0%	86.3%	84.9%	78.9%	80.7%	0.0%	3.0%
<b>Train Recall</b>	86.1%	86.5%	89.9%	85.7%	87.1%	87.1%	1.6%
<b>Val Recall</b>	90.9%	88.7%	76.3%	93.8%	86.8%	0.0%	6.7%
<b>Train MCC</b>	55.3%	57.8%	59.6%	56.0%	58.7%	57.5%	1.8%
<b>Val MCC</b>	65.3%	54.3%	51.3%	64.7%	51.9%	0.0%	6.9%
<b>Train Log-Loss</b>	7.31	7.19	6.67	7.21	6.88	7.05	0.27
<b>Val Log-Loss</b>	6.07	6.55	8.53	6.36	7.82	22.45	1.05

## Univariate Report

### Radius Error - Kernel Density Plot

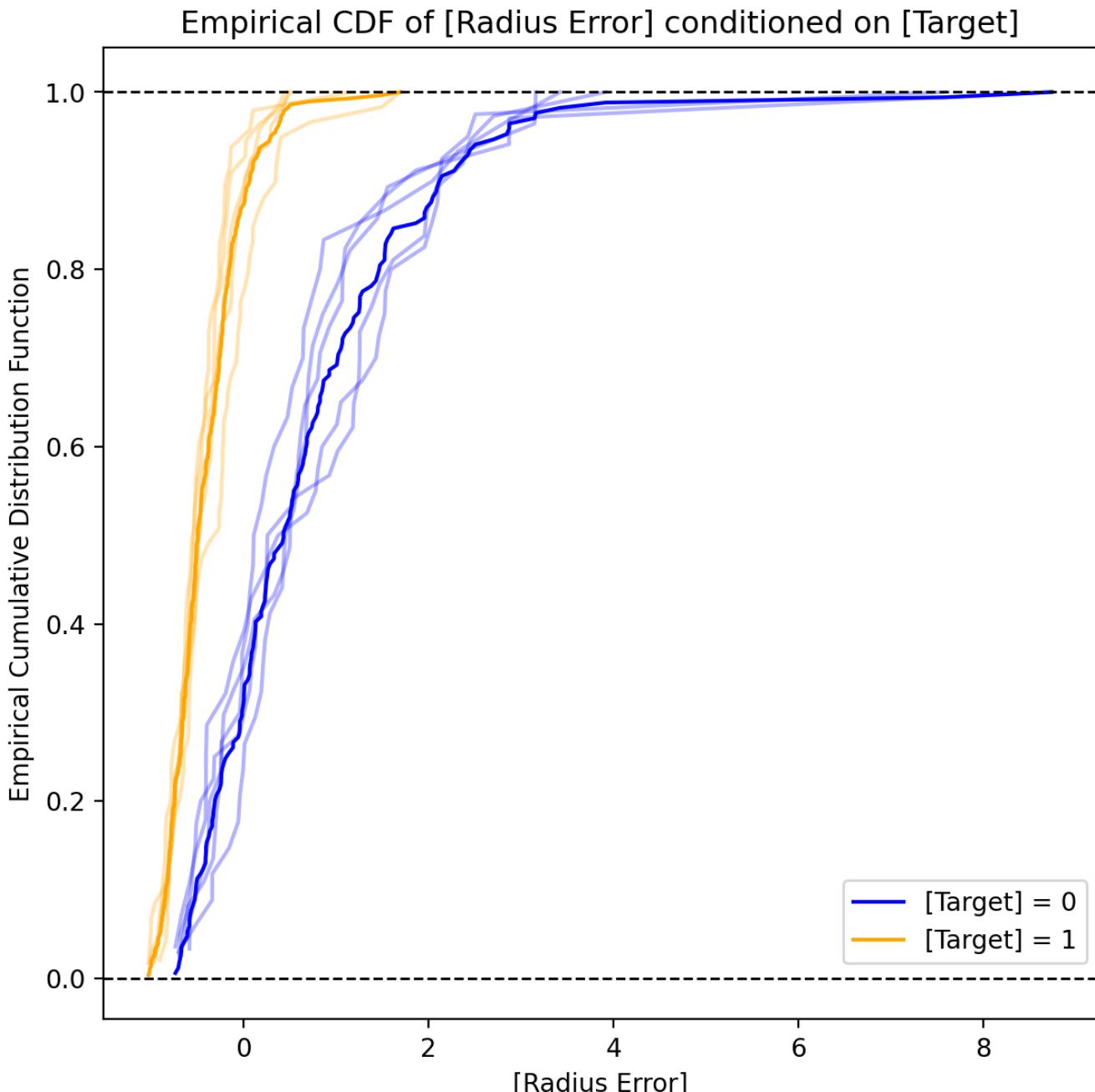
Kernel Density Plot of [Radius Error] by [Target]  
Distributions by level are significantly different at the 95% level.



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

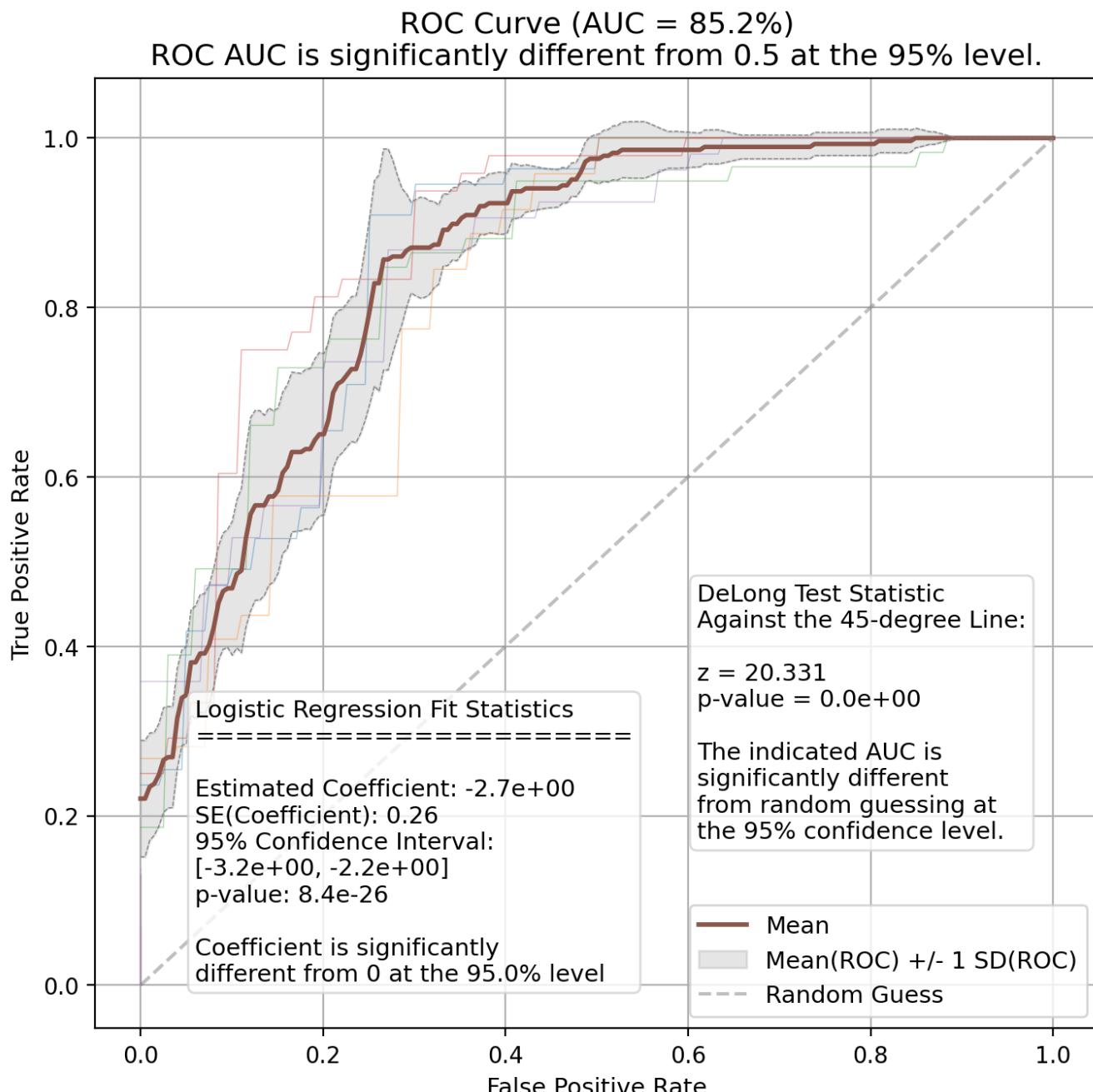
### Radius Error - Empirical CDF Plot



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



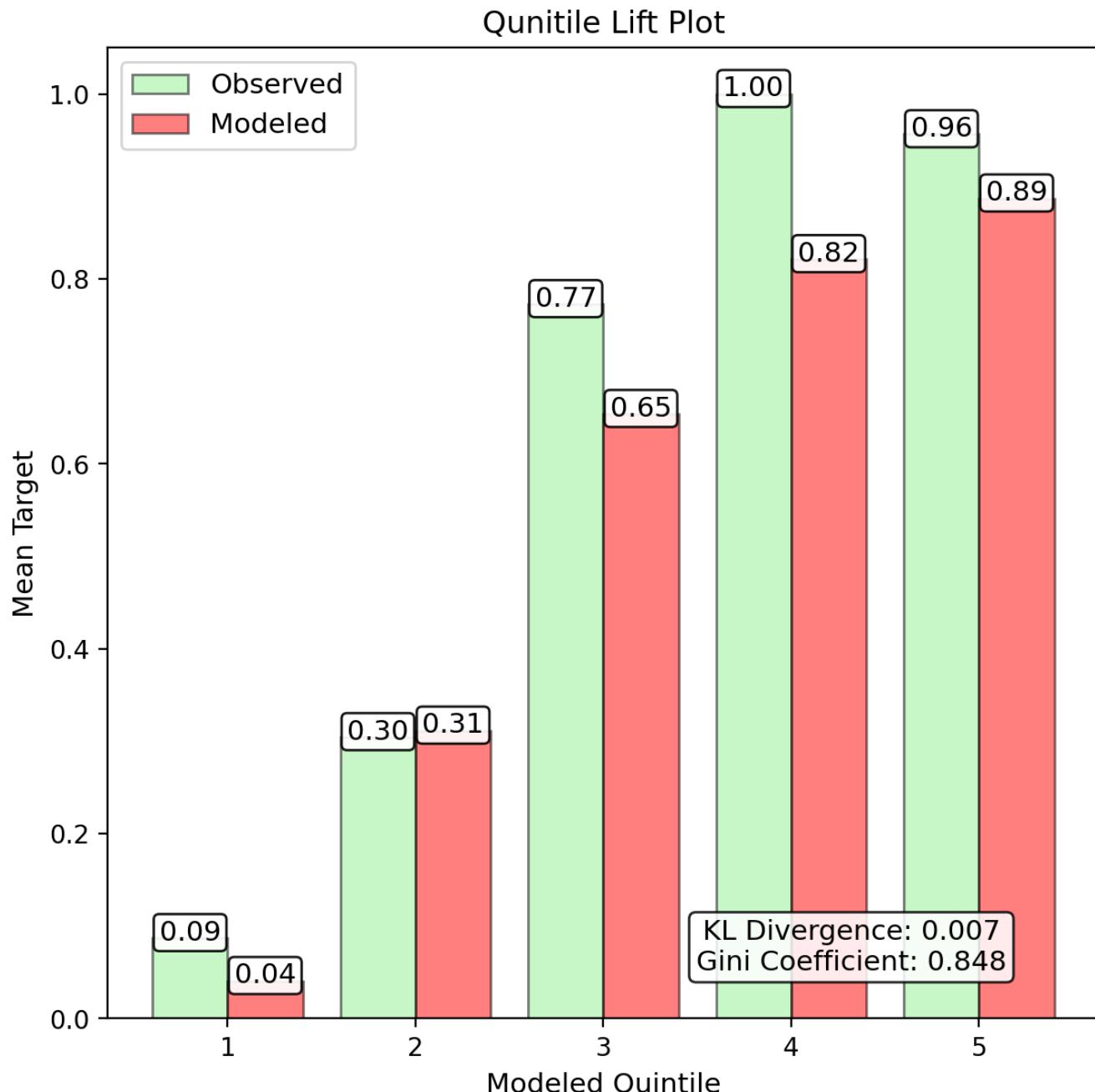
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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

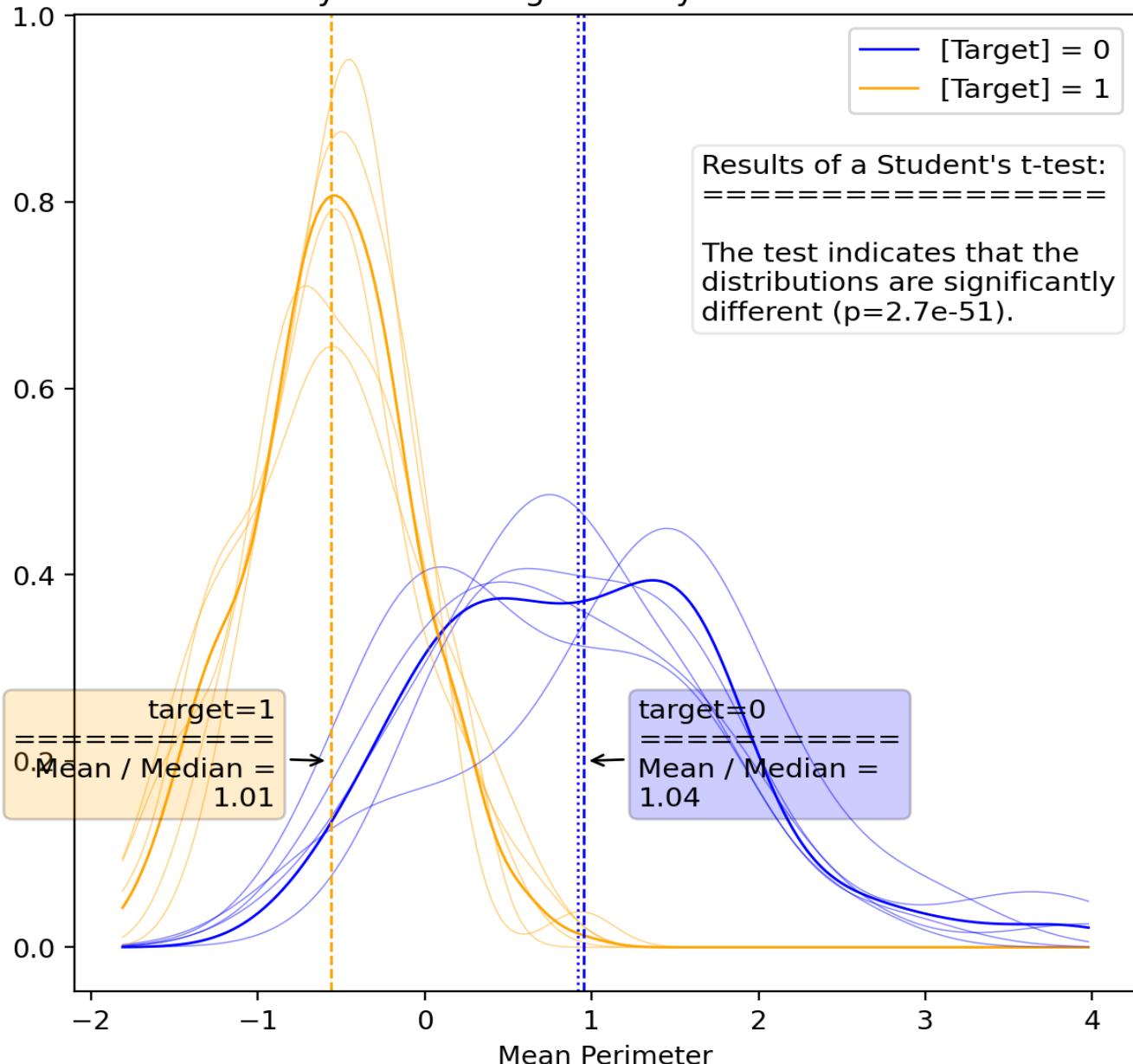
## Mean Perimeter - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
<b>Fitted Coef.</b>	-3.5e+00	-3.6e+00	-3.5e+00	-3.6e+00	-3.5e+00	-3.5e+00	7.9e-02
<b>Fitted p-Value</b>	1.6e-21	7.0e-21	3.2e-21	5.7e-22	1.1e-21	2.0e-26	2.6e-21
<b>Fitted Std. Err.</b>	0.363	0.388	0.369	0.370	0.361	0.331	0.011
<b>Conf. Int. Lower</b>	-4.2e+00	-4.4e+00	-4.2e+00	-4.3e+00	-4.2e+00	-4.2e+00	9.9e-02
<b>Conf. Int. Upper</b>	-2.7e+00	-2.9e+00	-2.8e+00	-2.8e+00	-2.7e+00	-2.9e+00	6.0e-02
<b>Train Accuracy</b>	86.1%	87.4%	87.0%	87.0%	86.6%	86.8%	0.5%
<b>Val Accuracy</b>	89.5%	84.8%	86.0%	85.9%	88.0%	37.7%	1.9%
<b>Train AUC</b>	85.2%	87.3%	86.9%	86.9%	86.5%	86.6%	0.8%
<b>Val AUC</b>	90.6%	82.9%	85.2%	85.6%	86.9%	50.0%	2.8%
<b>Train F1</b>	89.1%	89.3%	89.4%	89.7%	89.0%	89.3%	0.3%
<b>Test F1</b>	90.2%	89.2%	88.9%	87.5%	90.6%	0.0%	1.2%
<b>Train Precision</b>	89.9%	91.3%	91.7%	92.0%	91.4%	91.2%	0.8%
<b>Val Precision</b>	97.9%	91.2%	89.7%	87.5%	90.6%	0.0%	3.9%
<b>Train Recall</b>	88.3%	87.4%	87.2%	87.4%	86.7%	87.4%	0.6%
<b>Val Recall</b>	83.6%	87.3%	88.1%	87.5%	90.6%	0.0%	2.5%
<b>Train MCC</b>	70.0%	74.0%	72.8%	72.5%	71.9%	72.3%	1.4%
<b>Val MCC</b>	80.1%	64.0%	70.1%	71.3%	73.9%	0.0%	5.9%
<b>Train Log-Loss</b>	5.01	4.56	4.68	4.68	4.84	4.75	0.18
<b>Val Log-Loss</b>	3.79	5.46	5.04	5.09	4.34	22.45	0.67

## Univariate Report

### Mean Perimeter - Kernel Density Plot

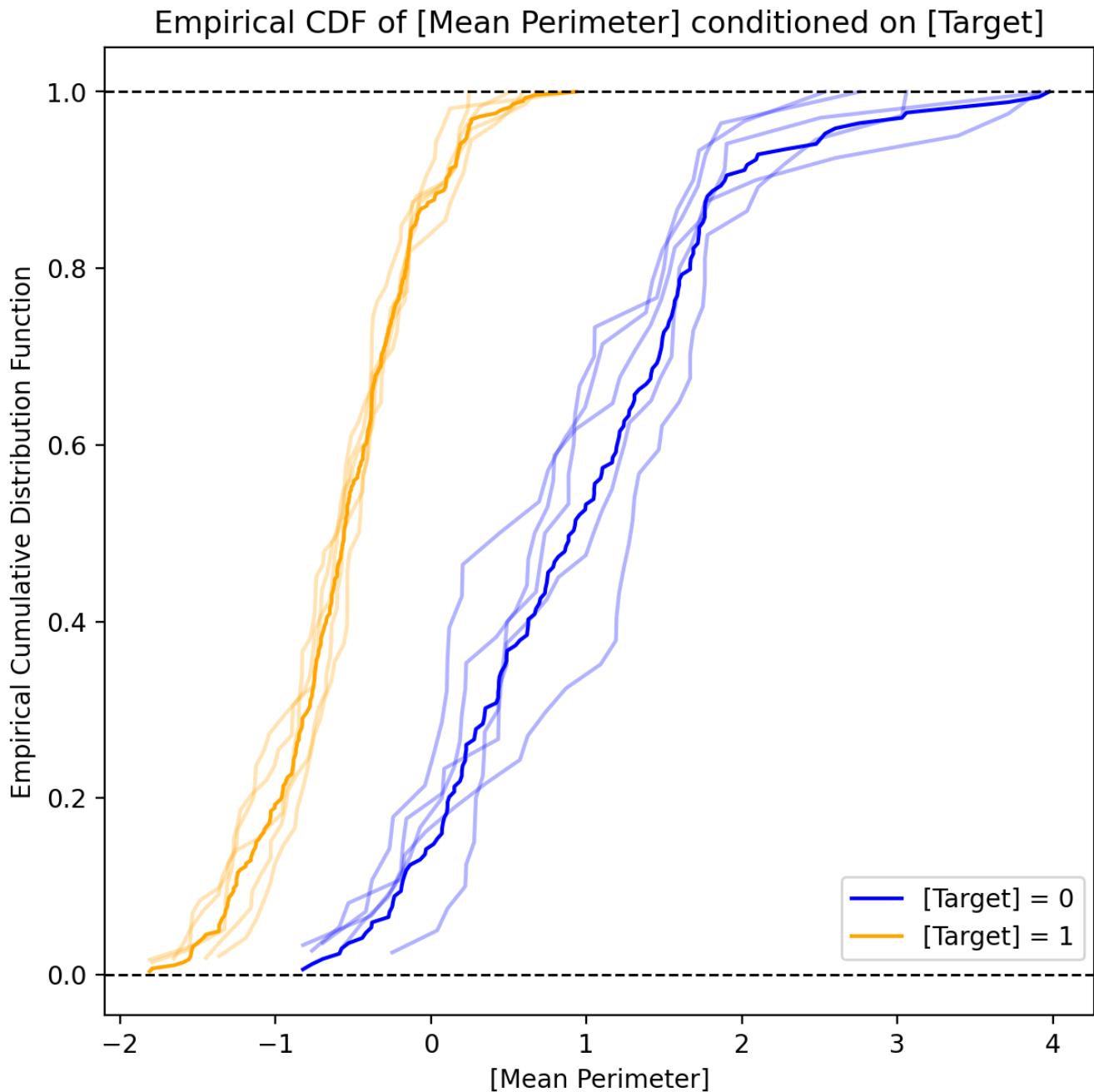
Kernel Density Plot of [Mean Perimeter] by [Target]  
Distributions by level are significantly different at the 95% level.



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

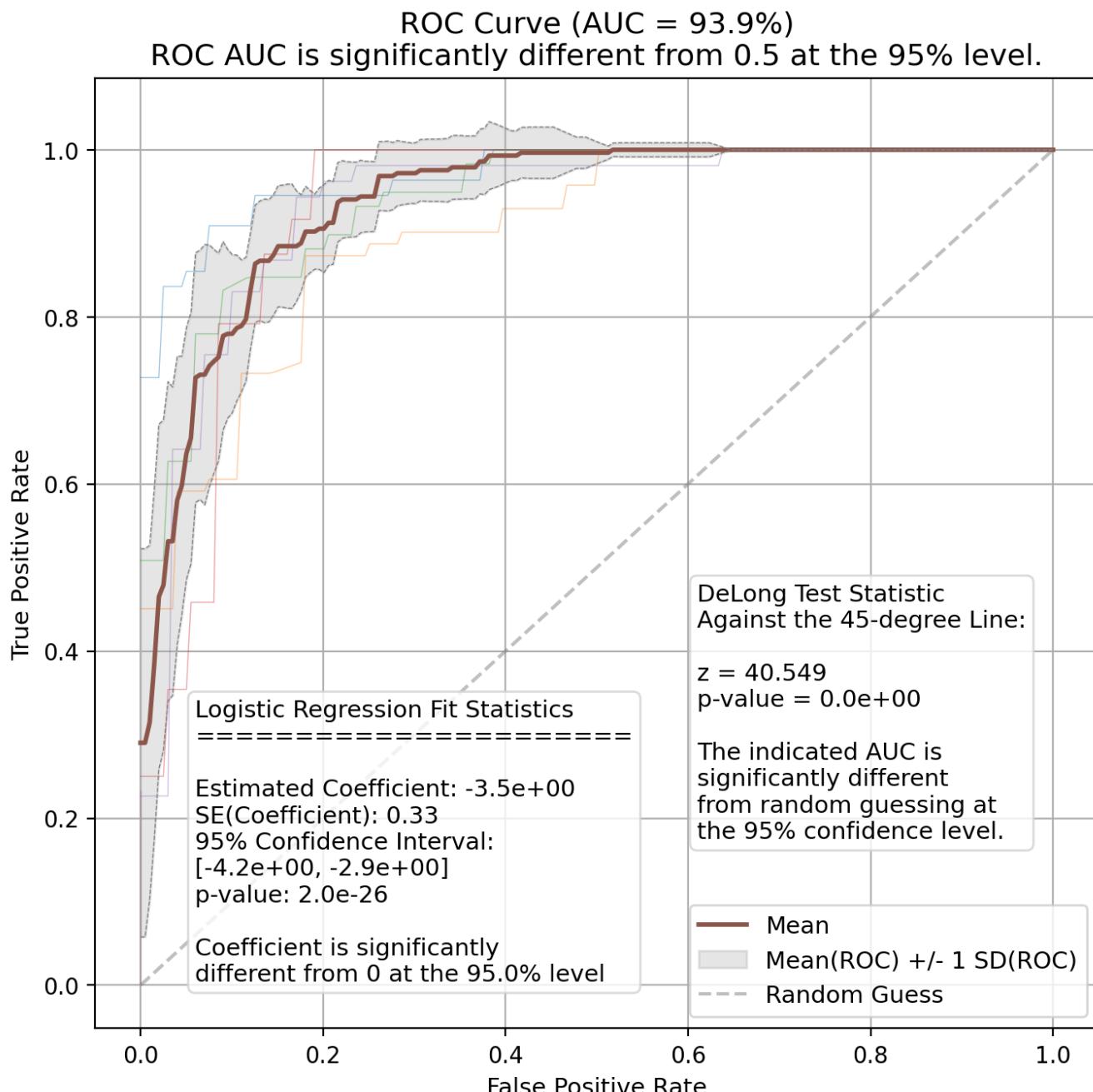
Mean Perimeter - Empirical CDF Plot



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 Perimeter - ROC Curve



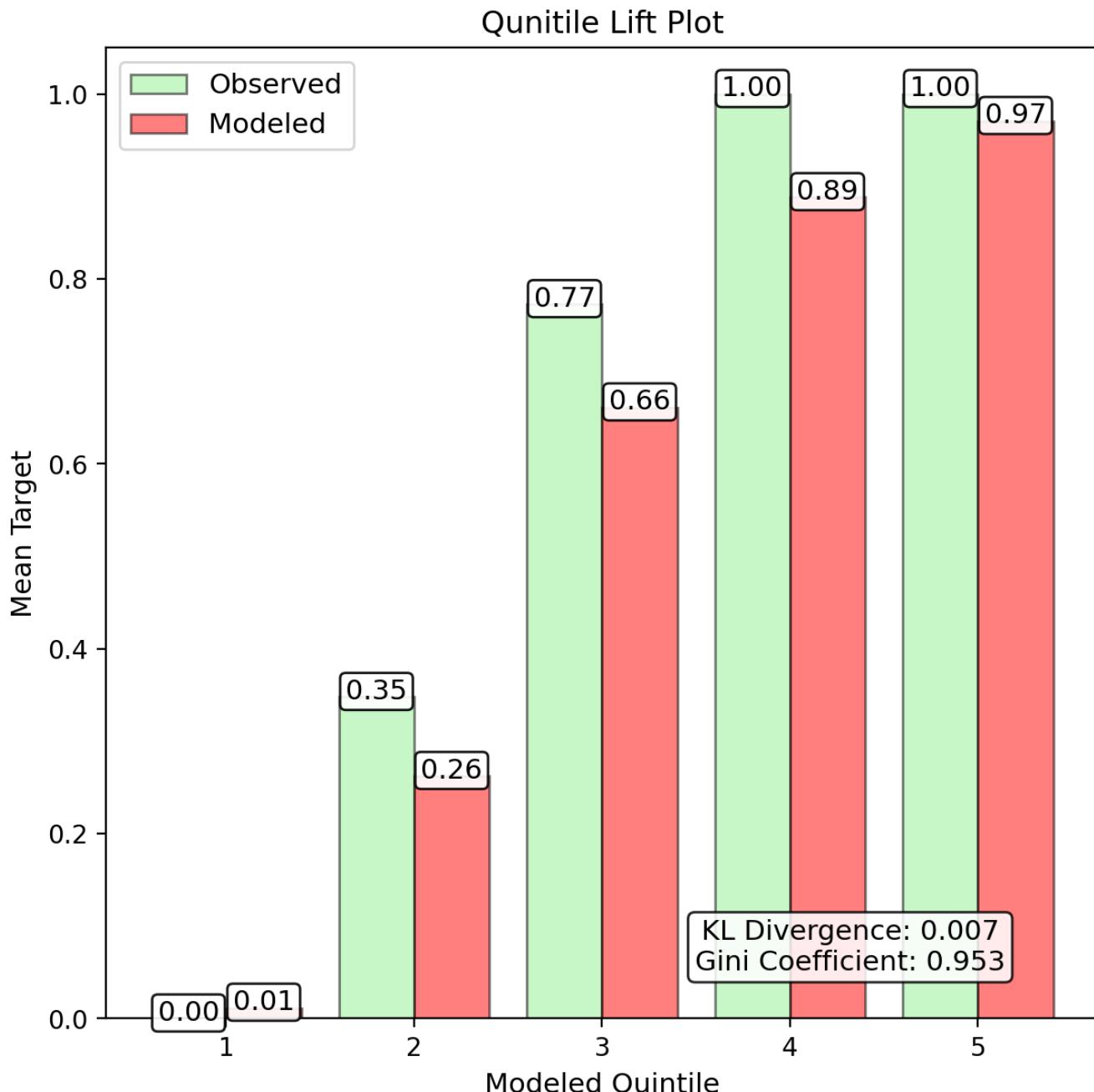
This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 Perimeter - 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.