Cancer Model Univariate Analysis Report

2024-02-08

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 sortied 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.
WorstArea	91.9%	90.4%	97.5%	90.0%	93.8%	82.7%	91.1%
MeanConcavePoints	91.2%	94.8%	91.0%	91.3%	92.9%	81.6%	90.4%
WorstConcavePoints	89.3%	95.1%	87.4%	89.9%	91.1%	78.1%	88.5%
MeanConcavity	88.4%	91.9%	89.4%	88.1%	90.6%	75.5%	87.3%
MeanArea	87.0%	88.8%	90.8%	85.7%	89.8%	72.0%	85.7%
AreaError	86.8%	84.7%	96.4%	83.6%	90.2%	71.8%	85.6%
WorstConcavity	84.7%	90.7%	84.3%	84.8%	87.4%	68.3%	83.4%
PerimeterError	81.9%	84.0%	88.0%	79.8%	85.9%	60.8%	80.1%
RadiusError	81.5%	83.7%	87.7%	79.5%	85.6%	60.0%	79.7%
MeanCompactness	79.1%	86.5%	79.0%	79.1%	82.6%	56.9%	77.2%
WorstCompactness	78.7%	83.9%	81.8%	77.7%	82.8%	54.9%	76.6%
SmoothnessError	51.1%	67.9%	42.0%	54.3%	51.9%	8.5%	45.9%
SymmetryError	50.1%	67.3%	39.8%	53.6%	50.0%	7.2%	44.7%
TextureError	49.2%	64.4%	42.6%	51.5%	51.3%	2.9%	43.6%
MeanFractalDimension	48.3%	63.1%	42.6%	50.3%	50.8%	0.6%	42.6%
FractalDimensionError	38.3%	51.4%	30.0%	41.2%	37.9%	-17.7%	30.2%
MeanSymmetry	36.6%	49.2%	36.4%	36.6%	41.9%	-26.0%	29.1%
MeanSmoothness	35.5%	48.2%	37.8%	34.7%	42.4%	-29.6%	28.2%
WorstFractalDimension	33.7%	45.5%	28.0%	35.7%	34.7%	-28.4%	24.9%
WorstSmoothness	32.3%	44.9%	34.7%	31.5%	39.2%	-35.8%	24.5%

This table shows an overview of the results for the variables in this file, representing those whose average test score are ranked between 1 and 20 of the variables passed to the Cancer Model.

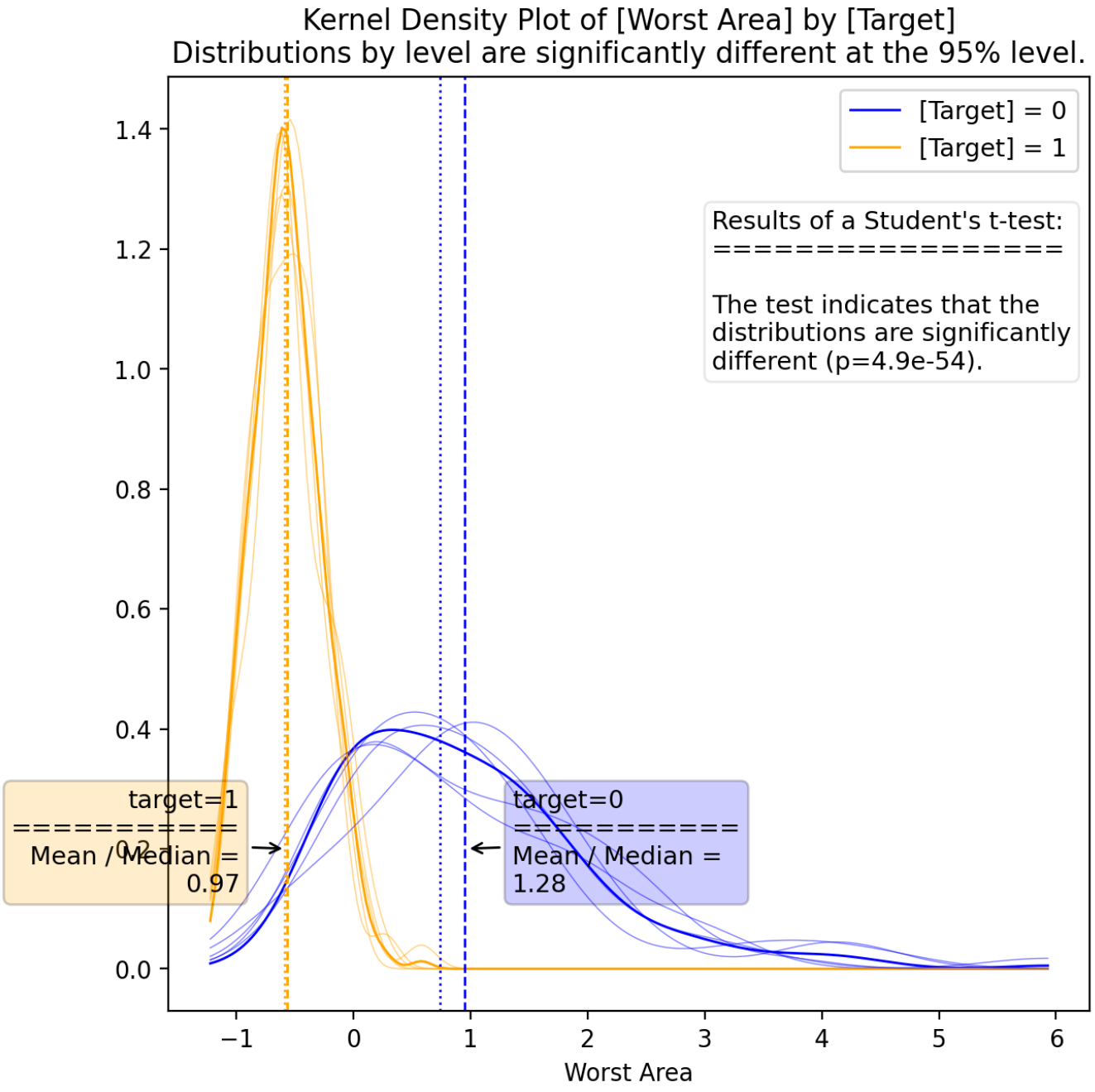
Univariate Report

Worst Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.9e+00	-5.4e+00	-5.3e+00	-4.9e+00	-5.1e+00	-3.3e-04	2.4e-01
Fitted p-Value	1.8e-26	5.0e-26	1.3e-25	5.9e-28	4.2e-26	8.2e-05	5.1e-26
Fitted Std. Err.	0.459	0.516	0.508	0.448	0.485	0.000	0.030
Conf. Int. Lower	-5.8e+00	-6.5e+00	-6.3e+00	-5.8e+00	-6.1e+00	-4.9e-04	3.0e-01
Conf. Int. Upper	-4.0e+00	-4.4e+00	-4.3e+00	-4.0e+00	-4.2e+00	-1.7e-04	1.9e-01
Train Accuracy	91.9%	92.5%	91.2%	91.3%	92.7%	91.9%	0.7%
Val Accuracy	92.7%	90.1%	93.8%	95.0%	89.3%	91.9%	2.4%
Train AUC	90.0%	90.8%	89.4%	89.2%	90.7%	90.0%	0.8%
Val AUC	90.6%	87.2%	91.4%	94.5%	88.0%	90.0%	2.9%
Train F1	93.8%	94.3%	93.2%	93.4%	94.5%	93.8%	0.6%
Test F1	94.6%	92.5%	95.7%	95.7%	91.4%	93.8%	1.9%
Train Precision	90.0%	91.4%	88.9%	90.3%	91.3%	90.4%	1.0%
Val Precision	91.9%	86.1%	93.9%	93.3%	88.1%	90.4%	3.4%
Train Recall	97.8%	97.3%	97.8%	96.7%	97.8%	97.5%	0.5%
Val Recall	97.5%	100.0%	97.5%	98.2%	94.9%	97.5%	1.8%
Train MCC	82.9%	83.9%	81.8%	80.9%	84.2%	82.7%	1.4%
Val MCC	83.9%	80.0%	85.1%	89.9%	77.8%	82.7%	4.7%
Train Log-Loss	2.92	2.70	3.16	3.15	2.63	2.91	0.25
Val Log-Loss	2.62	3.57	2.23	1.80	3.85	2.91	0.87

Univariate Report

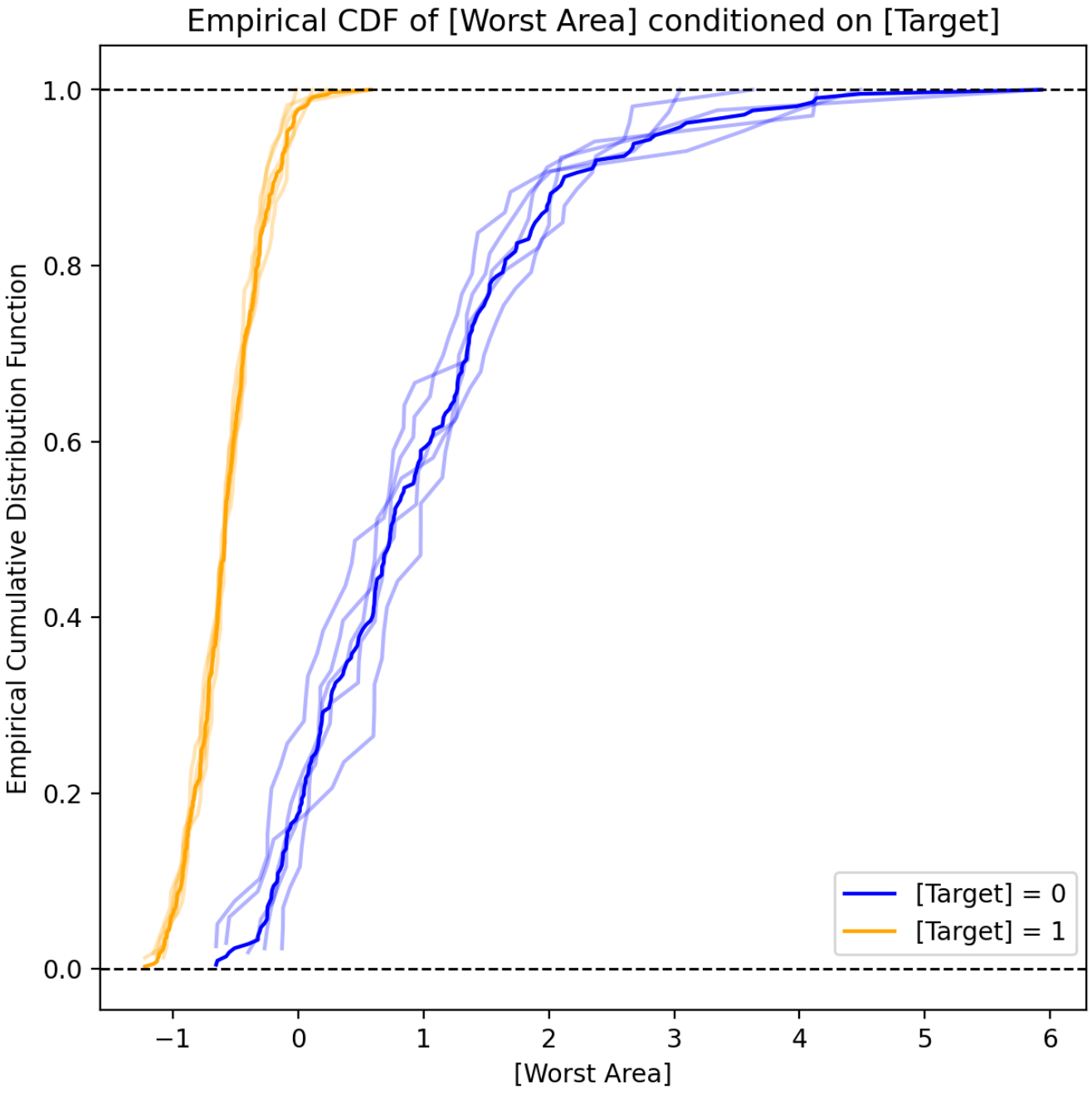
Worst Area - Kernel Density Plot



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

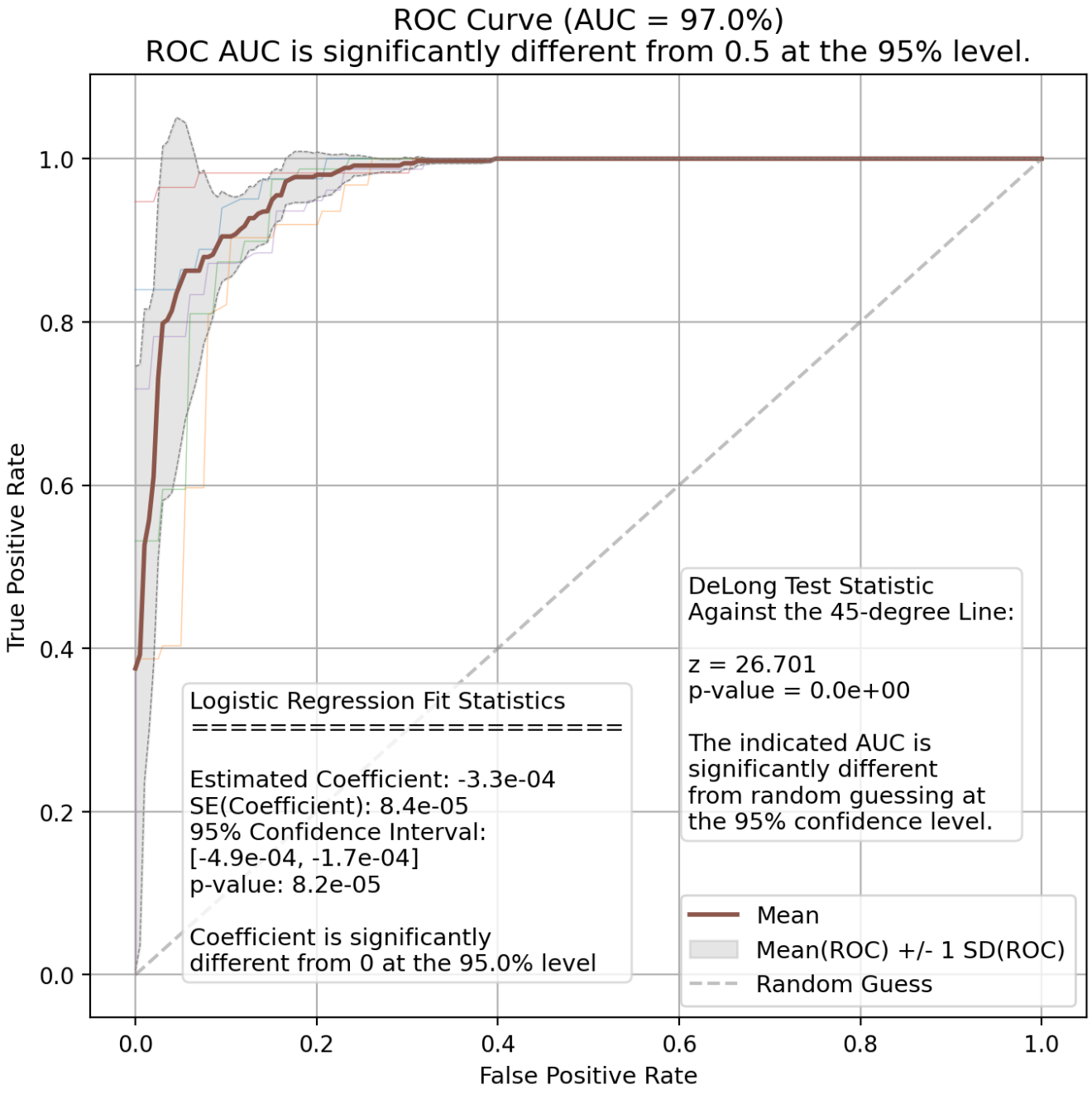
Worst Area - Empirical CDF Plot



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

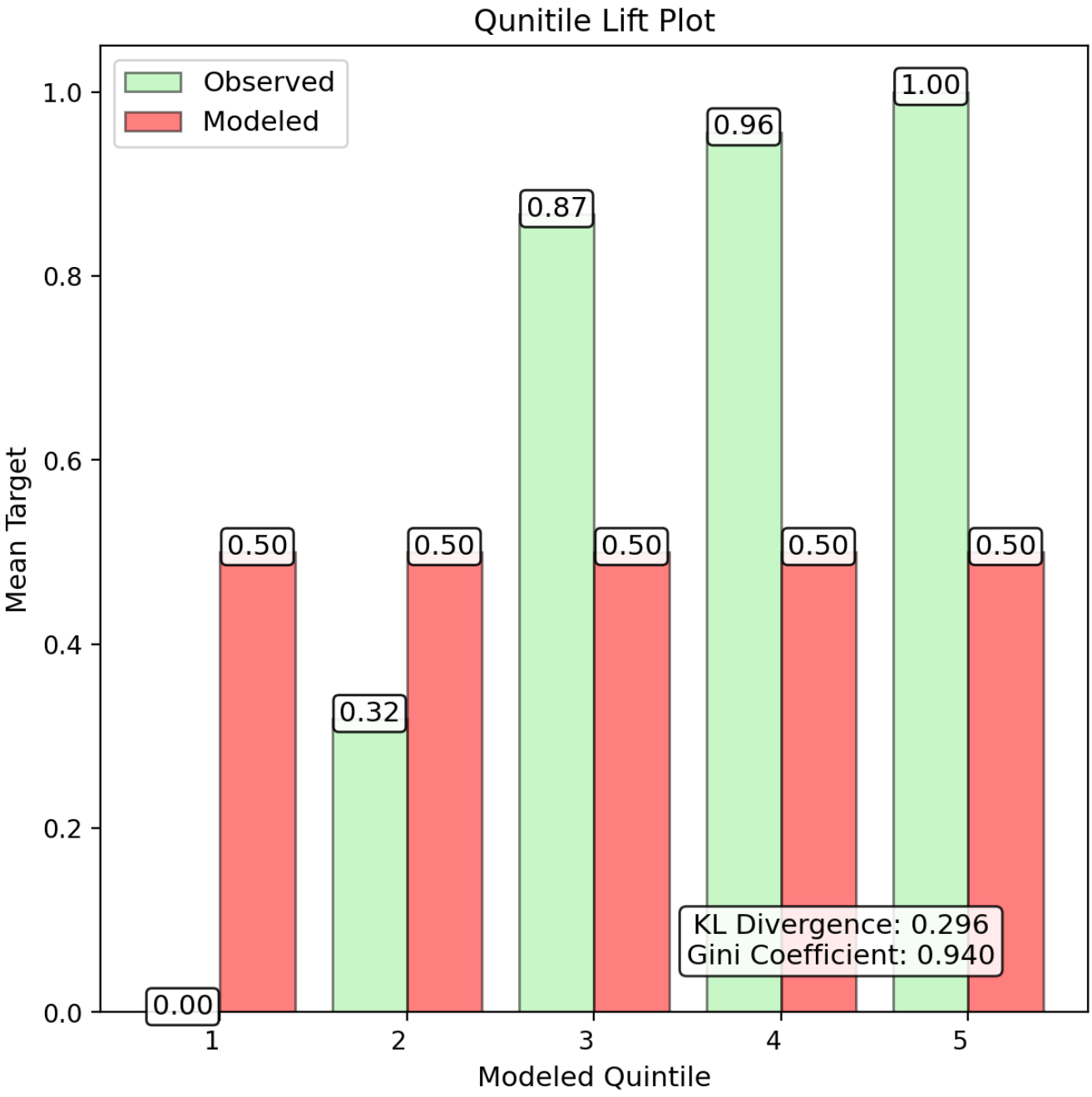


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst 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.

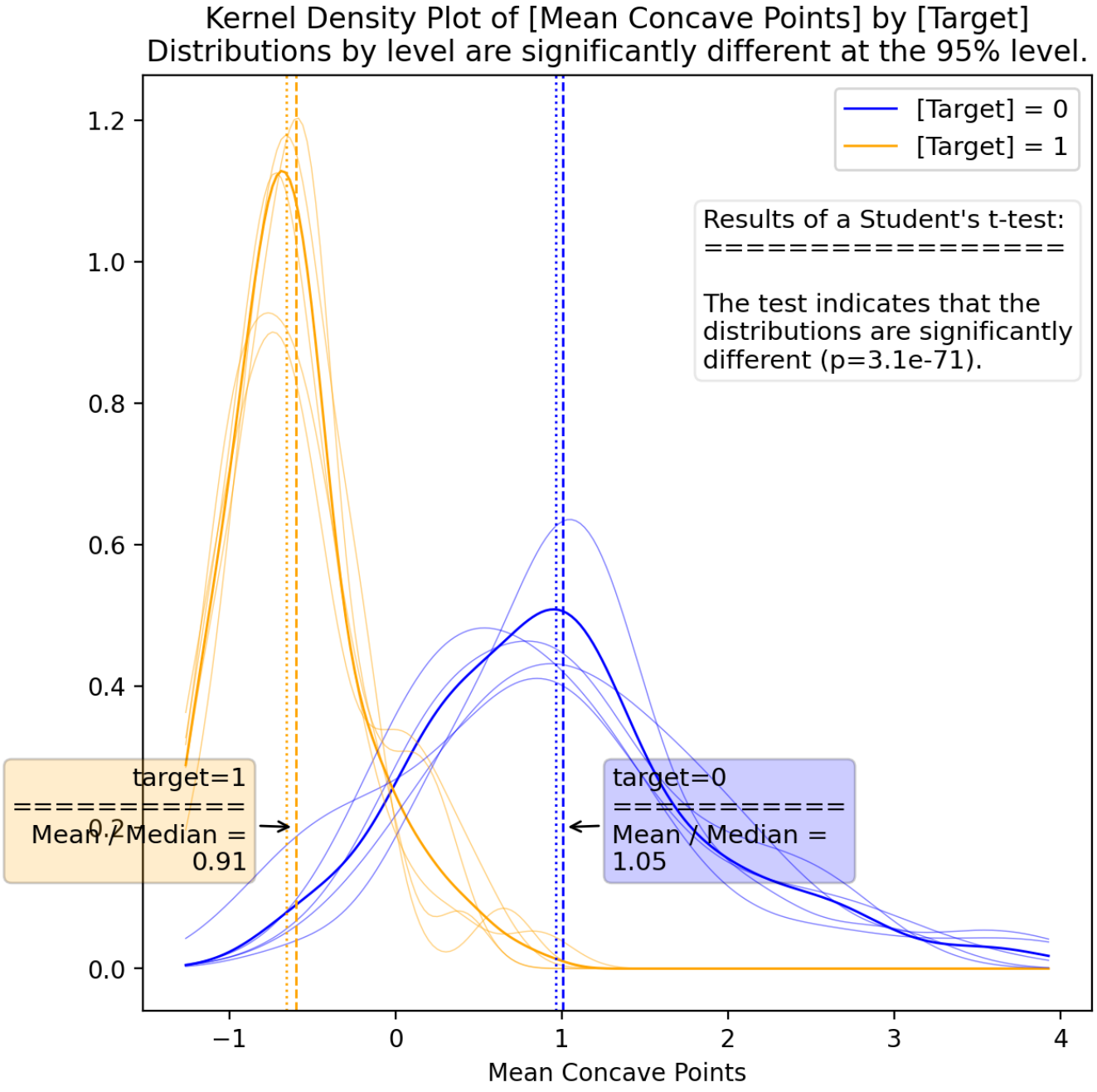
Univariate Report

Mean Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.9e+00	-3.7e+00	-3.7e+00	-3.8e+00	-4.4e+00	-9.1e+00	2.9e-01
Fitted p-Value	1.3e-29	2.2e-31	4.3e-30	2.6e-31	7.8e-27	1.9e-09	3.5e-27
Fitted Std. Err.	0.347	0.314	0.328	0.326	0.409	1.507	0.038
Conf. Int. Lower	-4.6e+00	-4.3e+00	-4.4e+00	-4.4e+00	-5.2e+00	-1.2e+01	3.6e-01
Conf. Int. Upper	-3.2e+00	-3.0e+00	-3.1e+00	-3.2e+00	-3.6e+00	-6.1e+00	2.2e-01
Train Accuracy	92.1%	90.4%	90.6%	90.2%	92.5%	91.2%	1.1%
Val Accuracy	88.7%	96.0%	94.7%	91.0%	87.8%	91.2%	3.6%
Train AUC	91.8%	90.5%	90.6%	90.1%	93.1%	91.3%	1.2%
Val AUC	90.3%	95.8%	94.5%	91.8%	86.4%	91.3%	3.7%
Train F1	93.6%	92.2%	92.1%	92.2%	93.9%	92.9%	0.9%
Test F1	90.8%	96.8%	96.2%	91.6%	90.1%	92.9%	3.1%
Train Precision	94.1%	94.3%	94.0%	94.1%	97.3%	94.8%	1.4%
Val Precision	97.2%	96.8%	97.4%	98.0%	86.9%	94.8%	4.7%
Train Recall	93.1%	90.2%	90.3%	90.3%	90.7%	91.0%	1.2%
Val Recall	85.2%	96.8%	94.9%	86.0%	93.6%	91.0%	5.3%
Train MCC	83.4%	79.8%	80.5%	79.2%	84.5%	81.6%	2.3%
Val MCC	77.5%	91.6%	87.7%	82.8%	74.5%	81.6%	7.0%
Train Log-Loss	2.83	3.47	3.40	3.54	2.72	3.17	0.38
Val Log-Loss	4.07	1.43	1.91	3.24	4.40	3.17	1.31

Univariate Report

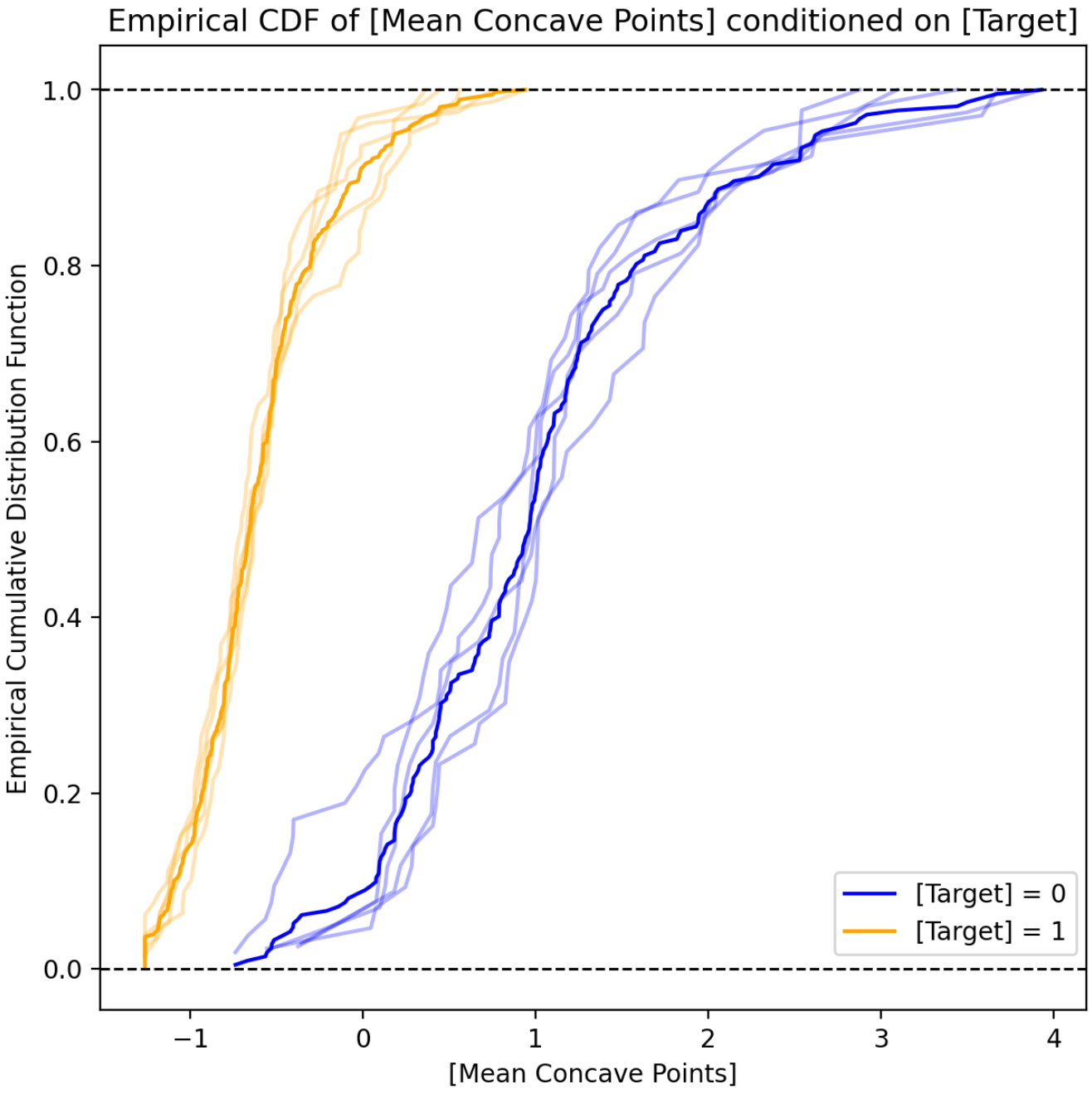
Mean Concave Points - Kernel Density Plot



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

Univariate Report

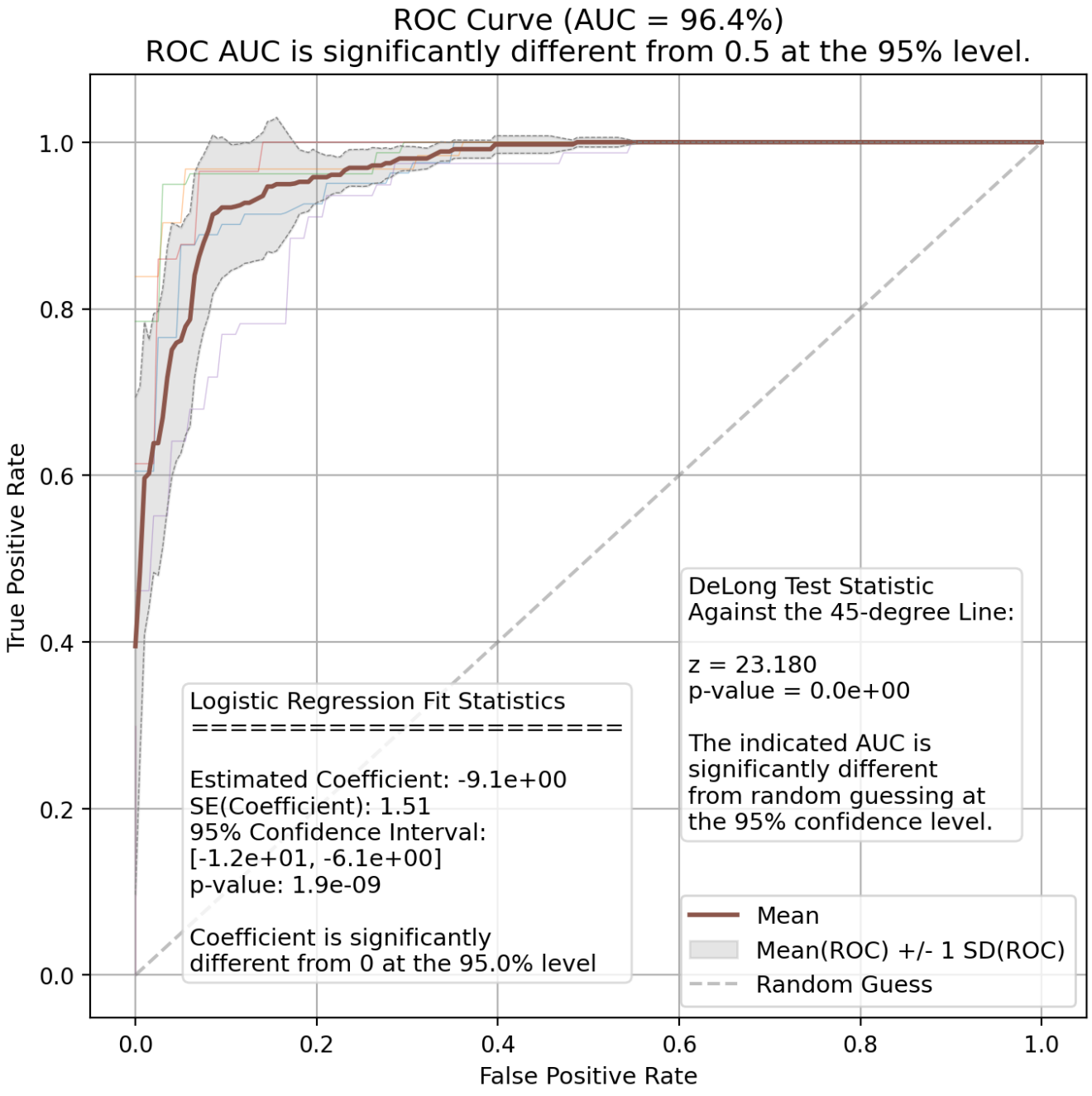
Mean Concave Points - Empirical CDF Plot



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

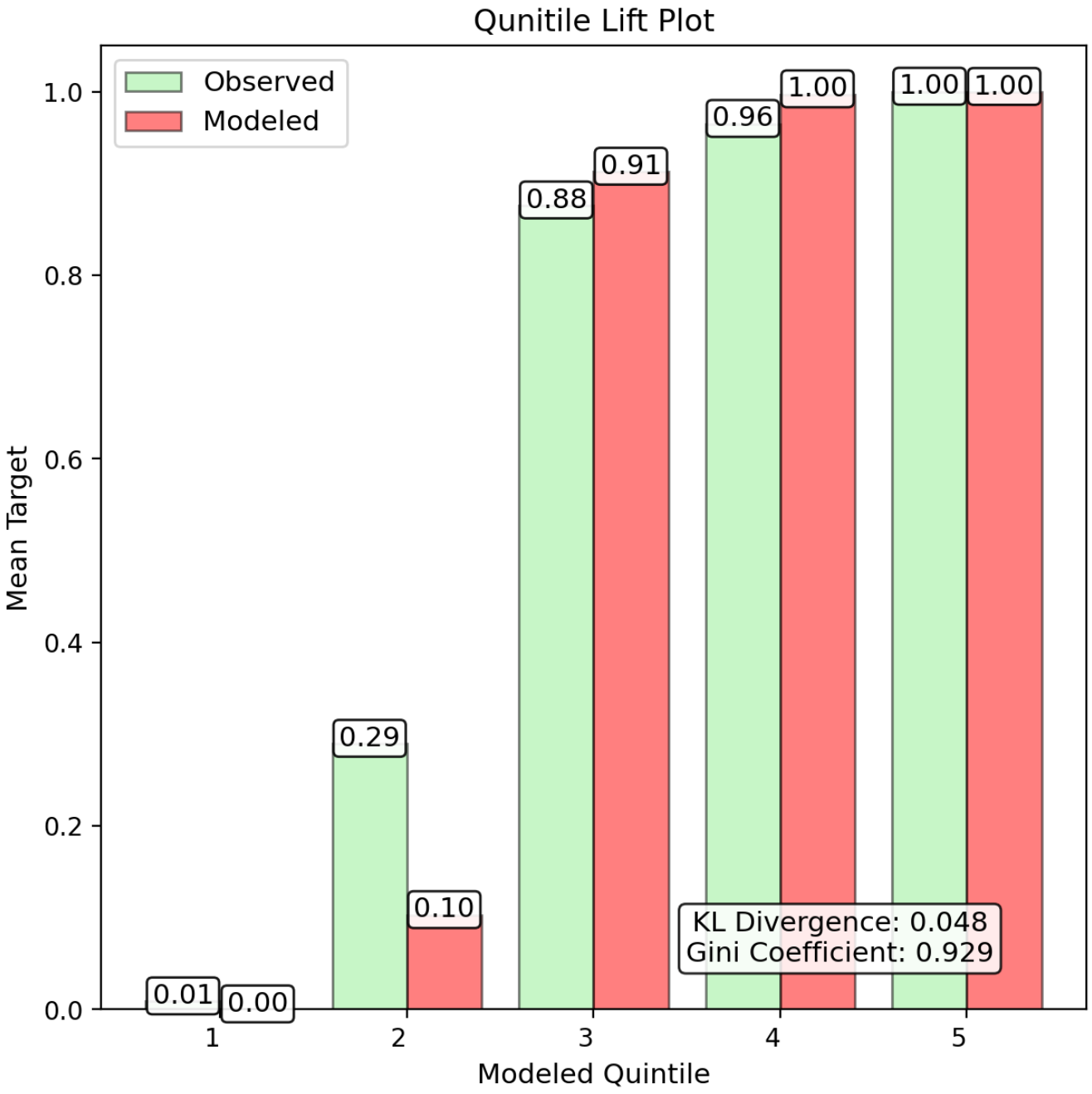


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean 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.

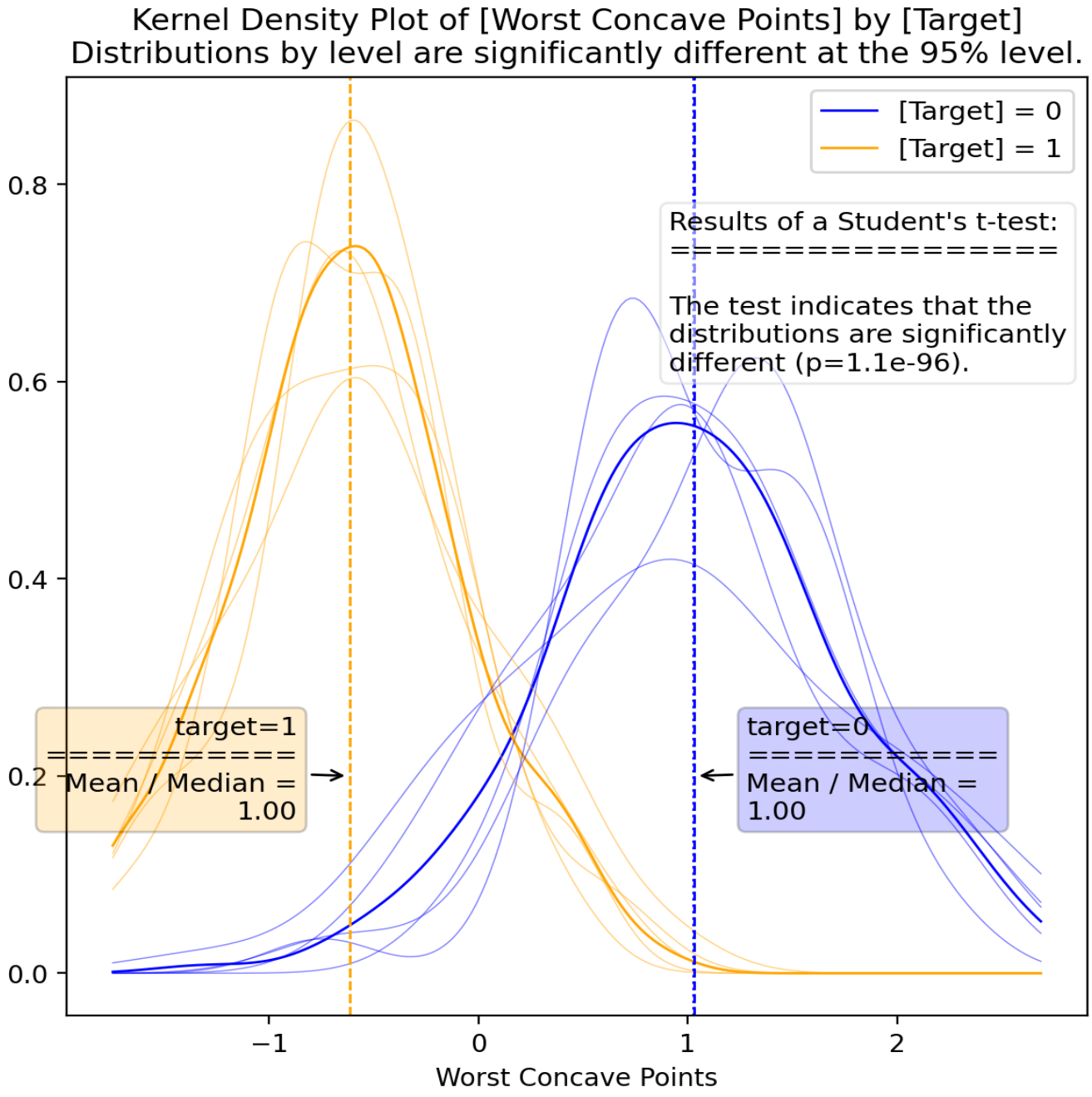
Univariate Report

Worst Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.7e+00	-3.3e+00	-3.4e+00	-3.4e+00	-3.7e+00	-2.4e+00	1.9e-01
Fitted p-Value	3.3e-27	1.6e-29	2.4e-28	5.1e-29	5.4e-27	2.0e-04	2.5e-27
Fitted Std. Err.	0.343	0.290	0.308	0.307	0.341	0.651	0.024
Conf. Int. Lower	-4.4e+00	-3.8e+00	-4.0e+00	-4.0e+00	-4.3e+00	-3.7e+00	2.3e-01
Conf. Int. Upper	-3.0e+00	-2.7e+00	-2.8e+00	-2.8e+00	-3.0e+00	-1.1e+00	1.4e-01
Train Accuracy	91.2%	88.0%	89.3%	88.1%	90.4%	89.3%	1.4%
Val Accuracy	84.7%	93.1%	92.0%	94.0%	84.7%	89.3%	4.6%
Train AUC	91.6%	88.7%	89.7%	88.9%	91.5%	89.9%	1.4%
Val AUC	86.1%	93.9%	93.5%	94.5%	84.5%	89.9%	4.8%
Train F1	92.7%	90.1%	90.9%	90.2%	92.1%	91.1%	1.2%
Test F1	87.4%	94.1%	94.0%	94.5%	87.0%	91.1%	3.9%
Train Precision	95.4%	94.4%	94.2%	94.9%	97.2%	95.1%	1.2%
Val Precision	94.3%	98.2%	98.6%	98.1%	88.2%	95.1%	4.5%
Train Recall	90.2%	86.1%	87.8%	86.0%	87.5%	87.4%	1.7%
Val Recall	81.5%	90.3%	89.9%	91.2%	85.9%	87.4%	4.1%
Train MCC	81.9%	75.6%	78.1%	75.6%	80.7%	78.1%	2.9%
Val MCC	69.3%	86.2%	82.9%	88.2%	68.5%	78.1%	9.4%
Train Log-Loss	3.16	4.31	3.87	4.30	3.46	3.86	0.51
Val Log-Loss	5.52	2.50	2.87	2.16	5.50	3.86	1.66

Univariate Report

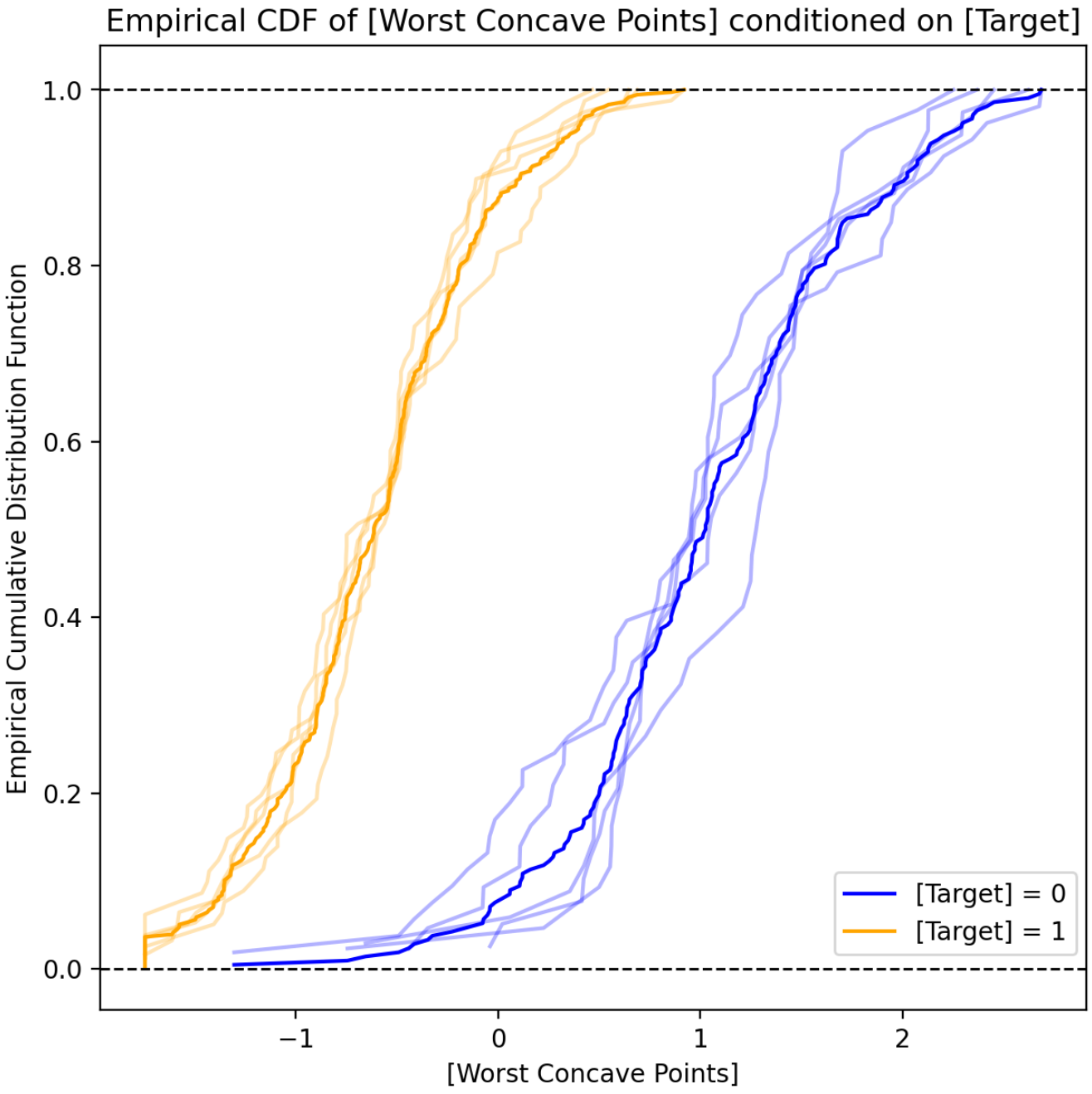
Worst Concave Points - Kernel Density Plot



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

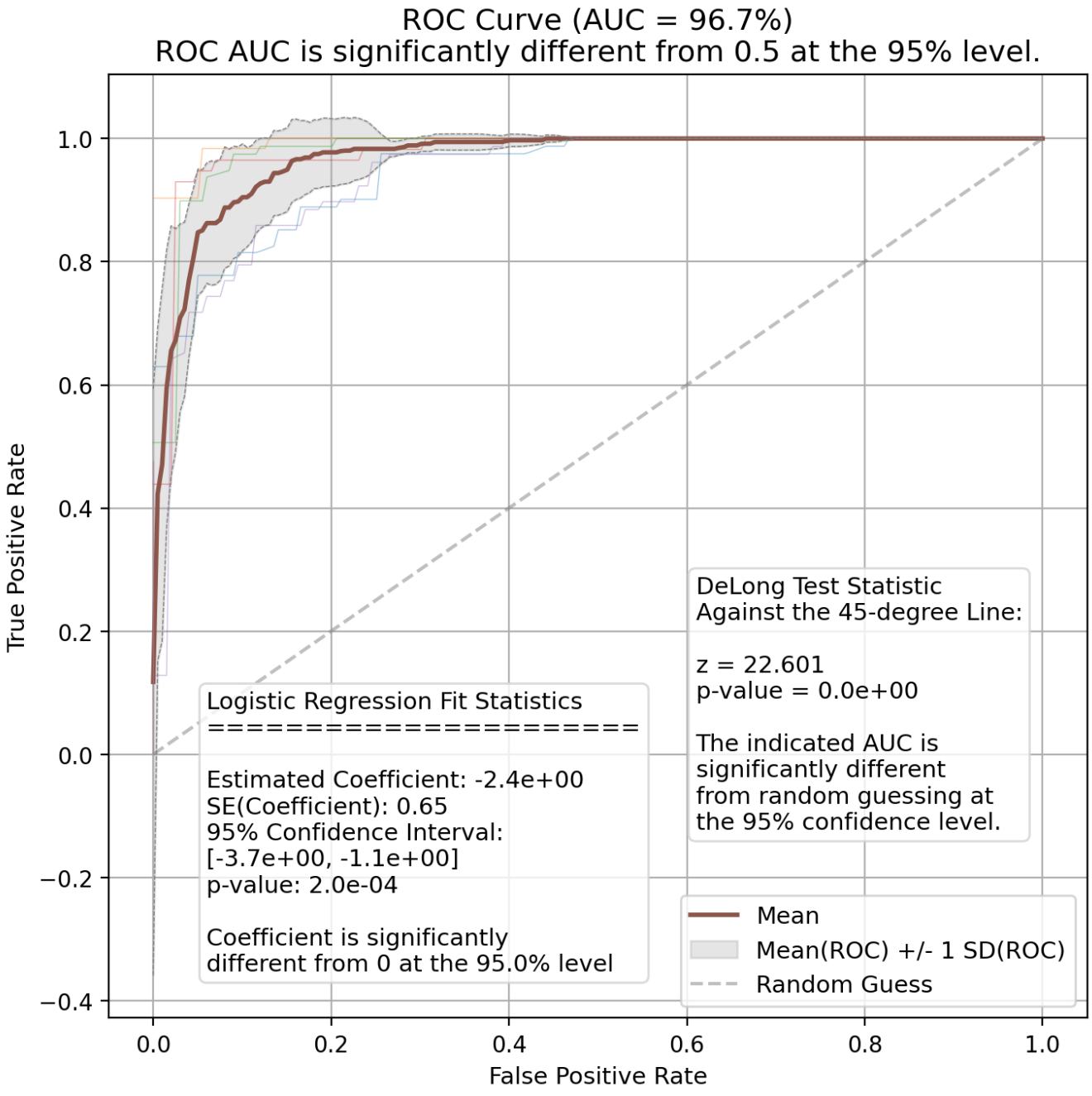
Worst Concave Points - Empirical CDF Plot



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

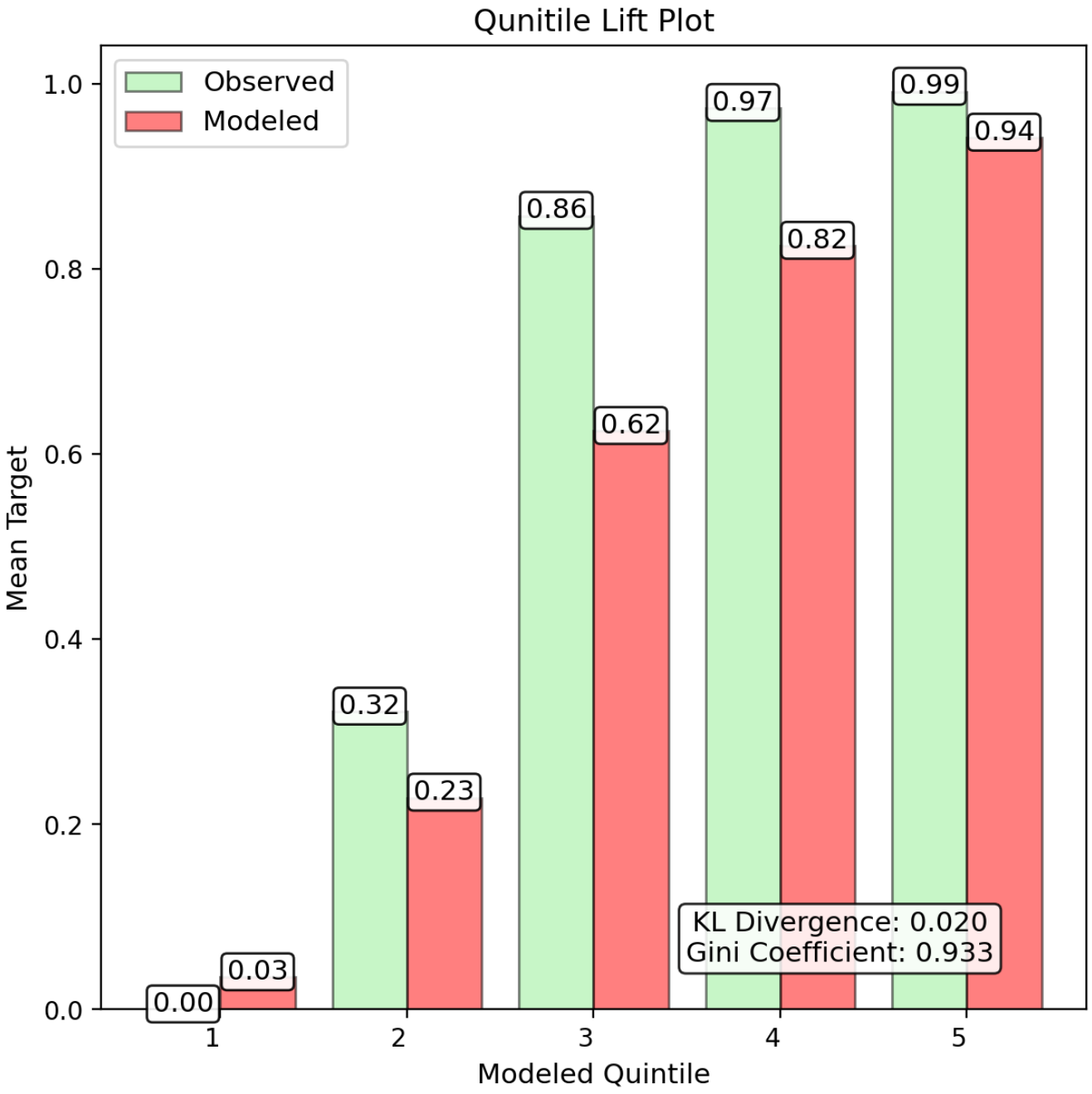


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst 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.

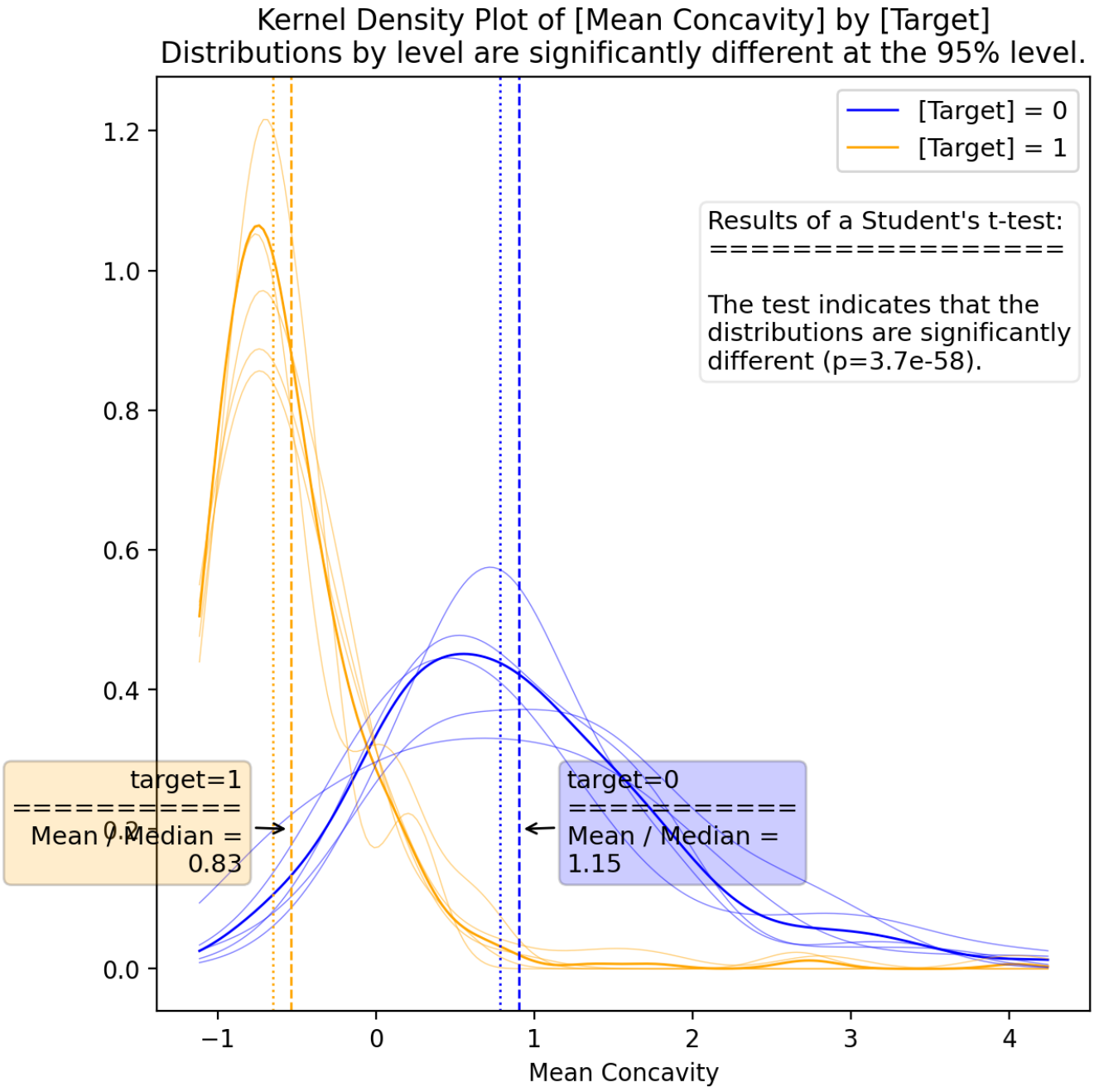
Univariate Report

Mean Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.1e+00	-2.7e+00	-2.8e+00	-2.8e+00	-3.2e+00	-4.7e+00	2.2e-01
Fitted p-Value	5.4e-30	3.8e-31	1.2e-30	7.3e-31	1.6e-29	5.6e-09	6.8e-30
Fitted Std. Err.	0.271	0.234	0.247	0.243	0.288	0.802	0.022
Conf. Int. Lower	-3.6e+00	-3.2e+00	-3.3e+00	-3.3e+00	-3.8e+00	-6.2e+00	2.7e-01
Conf. Int. Upper	-2.5e+00	-2.3e+00	-2.4e+00	-2.3e+00	-2.7e+00	-3.1e+00	1.8e-01
Train Accuracy	88.3%	88.5%	87.3%	88.3%	89.0%	88.4%	0.6%
Val Accuracy	88.7%	88.1%	92.9%	86.0%	86.3%	88.4%	2.8%
Train AUC	87.8%	88.2%	86.9%	87.7%	89.2%	88.1%	0.8%
Val AUC	89.2%	87.5%	93.3%	86.6%	84.8%	88.1%	3.2%
Train F1	90.5%	90.7%	89.5%	90.7%	91.1%	90.6%	0.6%
Test F1	91.0%	90.3%	94.8%	87.0%	88.9%	90.6%	2.9%
Train Precision	91.2%	92.3%	90.4%	91.8%	93.9%	91.9%	1.3%
Val Precision	94.7%	90.3%	97.3%	92.2%	85.7%	91.9%	4.4%
Train Recall	89.9%	89.2%	88.5%	89.7%	88.5%	89.4%	0.6%
Val Recall	87.7%	90.3%	92.4%	82.5%	92.3%	89.4%	4.2%
Train MCC	75.3%	75.6%	73.5%	74.8%	77.0%	75.5%	1.3%
Val MCC	76.3%	74.9%	84.0%	72.4%	71.3%	75.5%	5.0%
Train Log-Loss	4.21	4.16	4.58	4.23	3.95	4.18	0.23
Val Log-Loss	4.07	4.28	2.55	5.05	4.95	4.18	1.00

Univariate Report

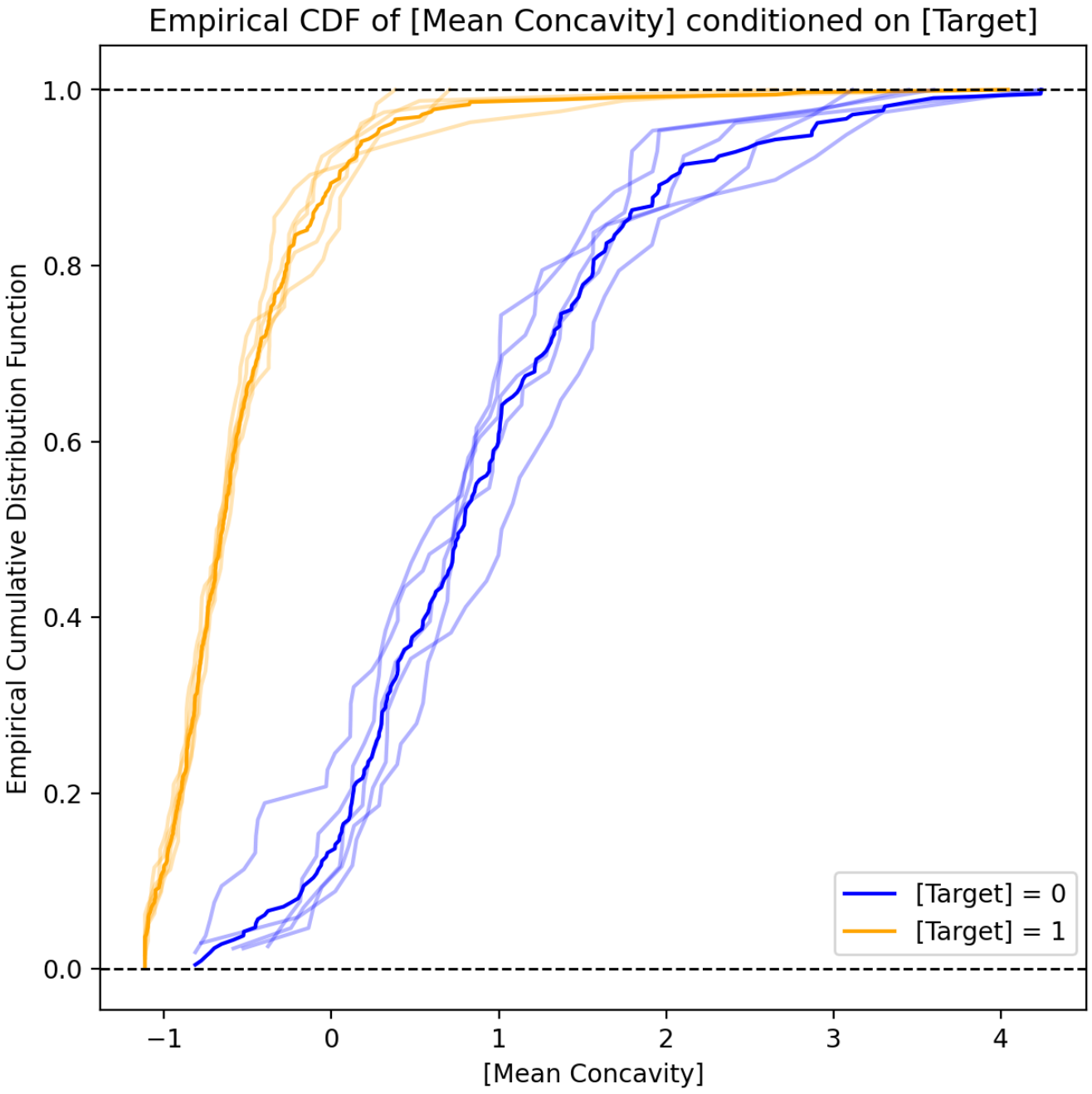
Mean Concavity - Kernel Density Plot



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

Univariate Report

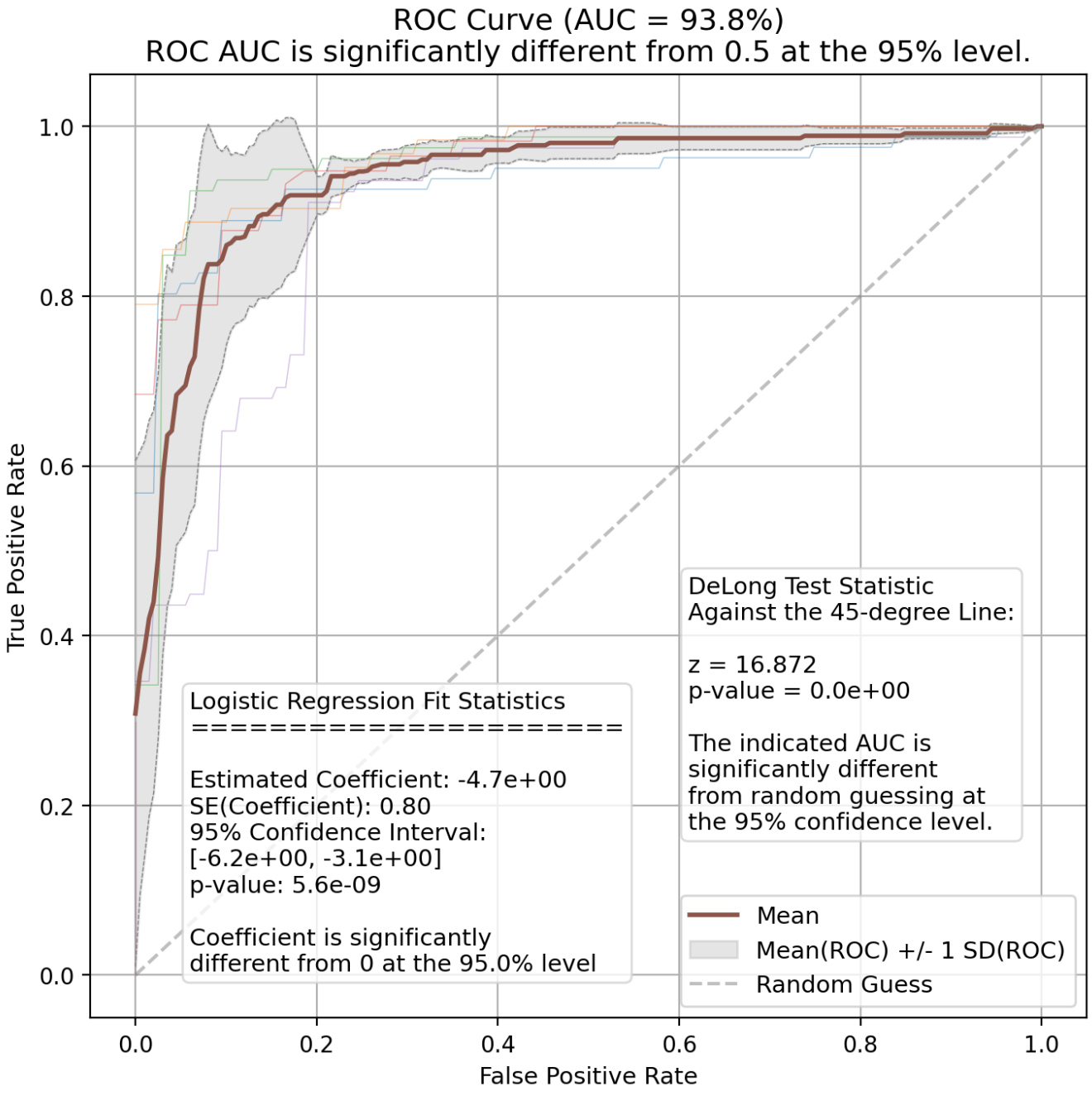
Mean Concavity - Empirical CDF Plot



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

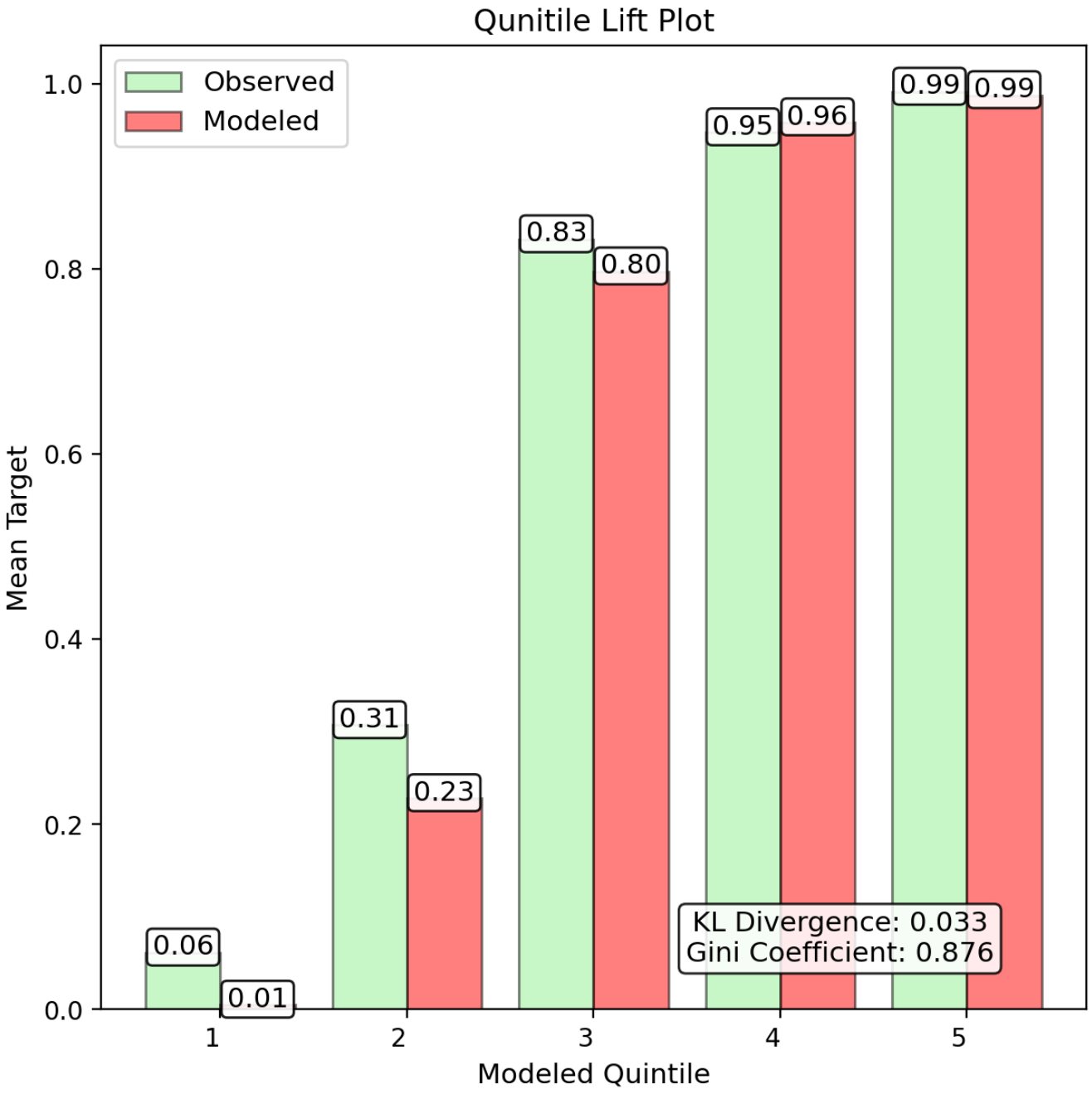


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean 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.

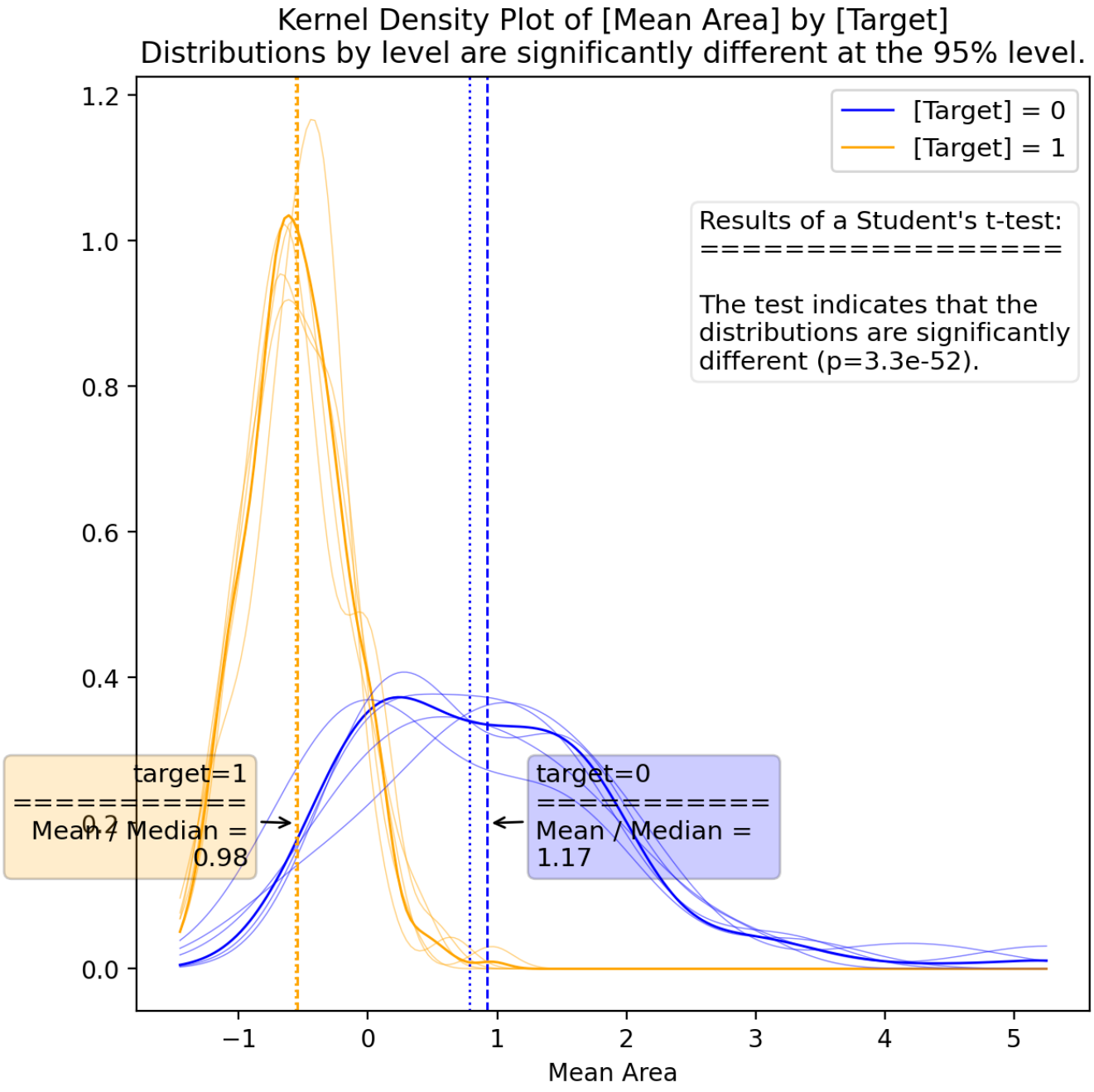
Univariate Report

Mean Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.8e+00	-4.2e+00	-4.0e+00	-3.8e+00	-3.9e+00	-2.7e-04	1.5e-01
Fitted p-Value	8.6e-27	1.9e-27	1.5e-26	3.7e-28	1.8e-26	2.0e-02	7.9e-27
Fitted Std. Err.	0.354	0.384	0.374	0.347	0.363	0.000	0.015
Conf. Int. Lower	-4.5e+00	-4.9e+00	-4.7e+00	-4.5e+00	-4.6e+00	-4.9e-04	1.8e-01
Conf. Int. Upper	-3.1e+00	-3.4e+00	-3.3e+00	-3.1e+00	-3.2e+00	-4.1e-05	1.3e-01
Train Accuracy	87.4%	88.0%	86.8%	86.1%	87.4%	87.0%	0.7%
Val Accuracy	87.1%	85.1%	90.3%	90.0%	84.0%	87.0%	2.8%
Train AUC	86.1%	86.8%	85.4%	85.2%	85.7%	85.7%	0.6%
Val AUC	85.2%	82.2%	88.8%	89.2%	84.1%	85.7%	3.0%
Train F1	90.0%	90.6%	89.5%	89.1%	90.3%	89.8%	0.6%
Test F1	90.2%	88.7%	93.0%	91.5%	86.1%	89.8%	2.7%
Train Precision	88.5%	89.7%	87.1%	89.6%	88.6%	88.8%	1.1%
Val Precision	89.2%	83.1%	93.6%	88.5%	89.0%	88.8%	3.7%
Train Recall	91.7%	91.5%	92.1%	88.7%	92.1%	90.8%	1.4%
Val Recall	91.4%	95.2%	92.4%	94.7%	83.3%	90.8%	4.8%
Train MCC	73.1%	74.2%	72.1%	70.1%	72.5%	72.0%	1.5%
Val MCC	71.3%	68.6%	77.1%	79.6%	67.4%	72.0%	5.3%
Train Log-Loss	4.54	4.31	4.74	5.00	4.53	4.69	0.26
Val Log-Loss	4.65	5.35	3.51	3.60	5.78	4.69	1.02

Univariate Report

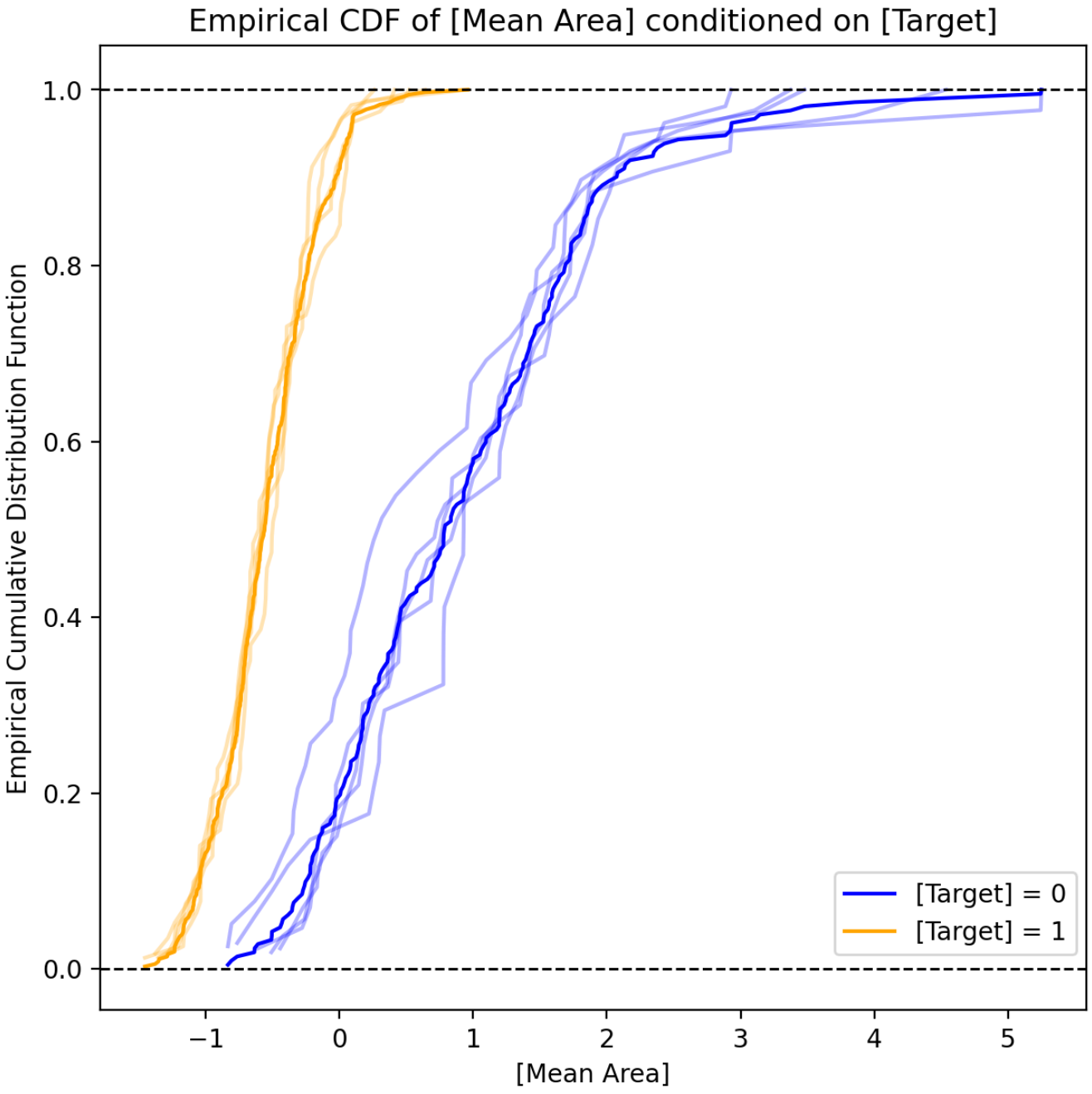
Mean Area - Kernel Density Plot



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

Univariate Report

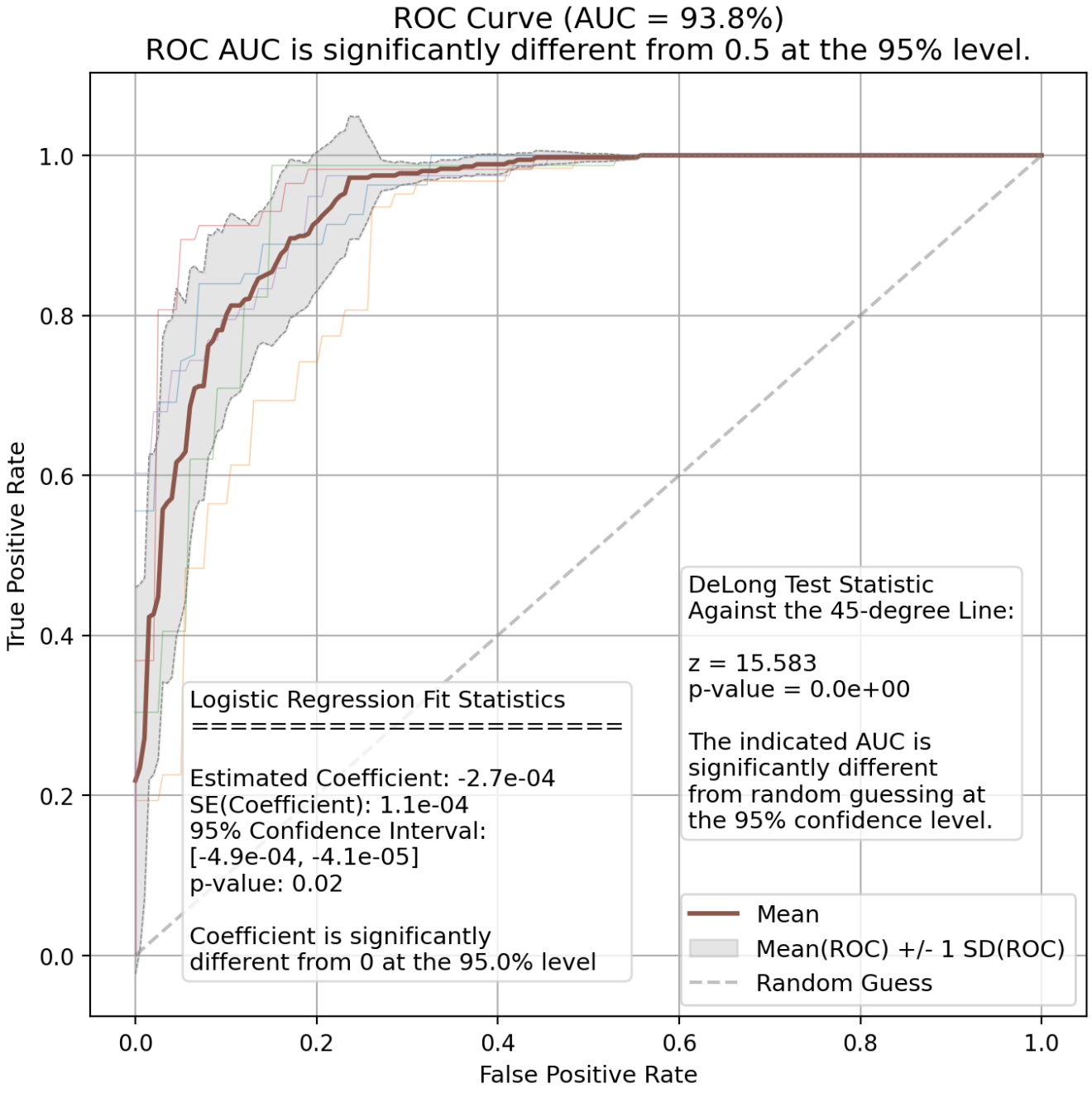
Mean Area - Empirical CDF Plot



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

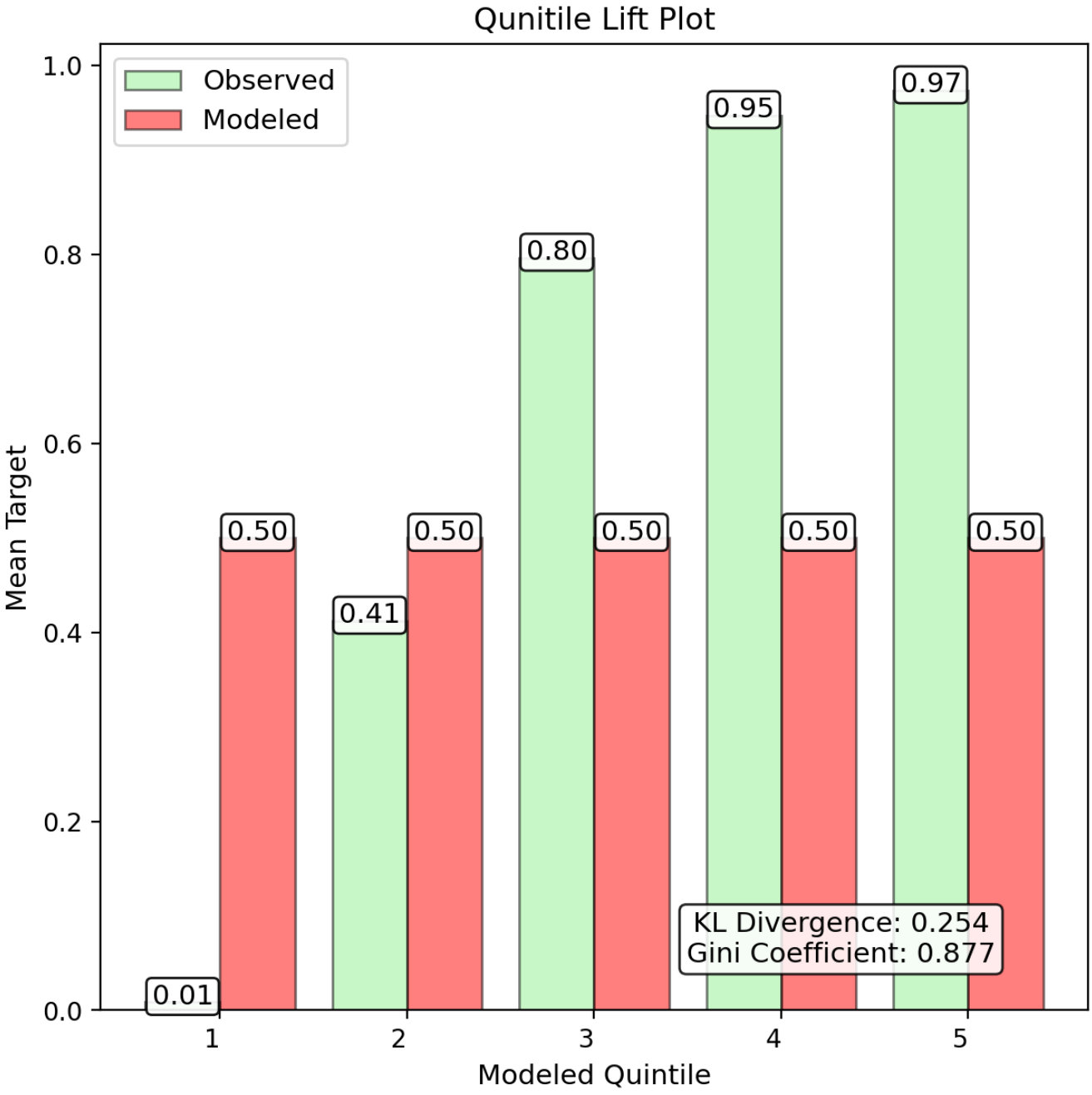


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean 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.

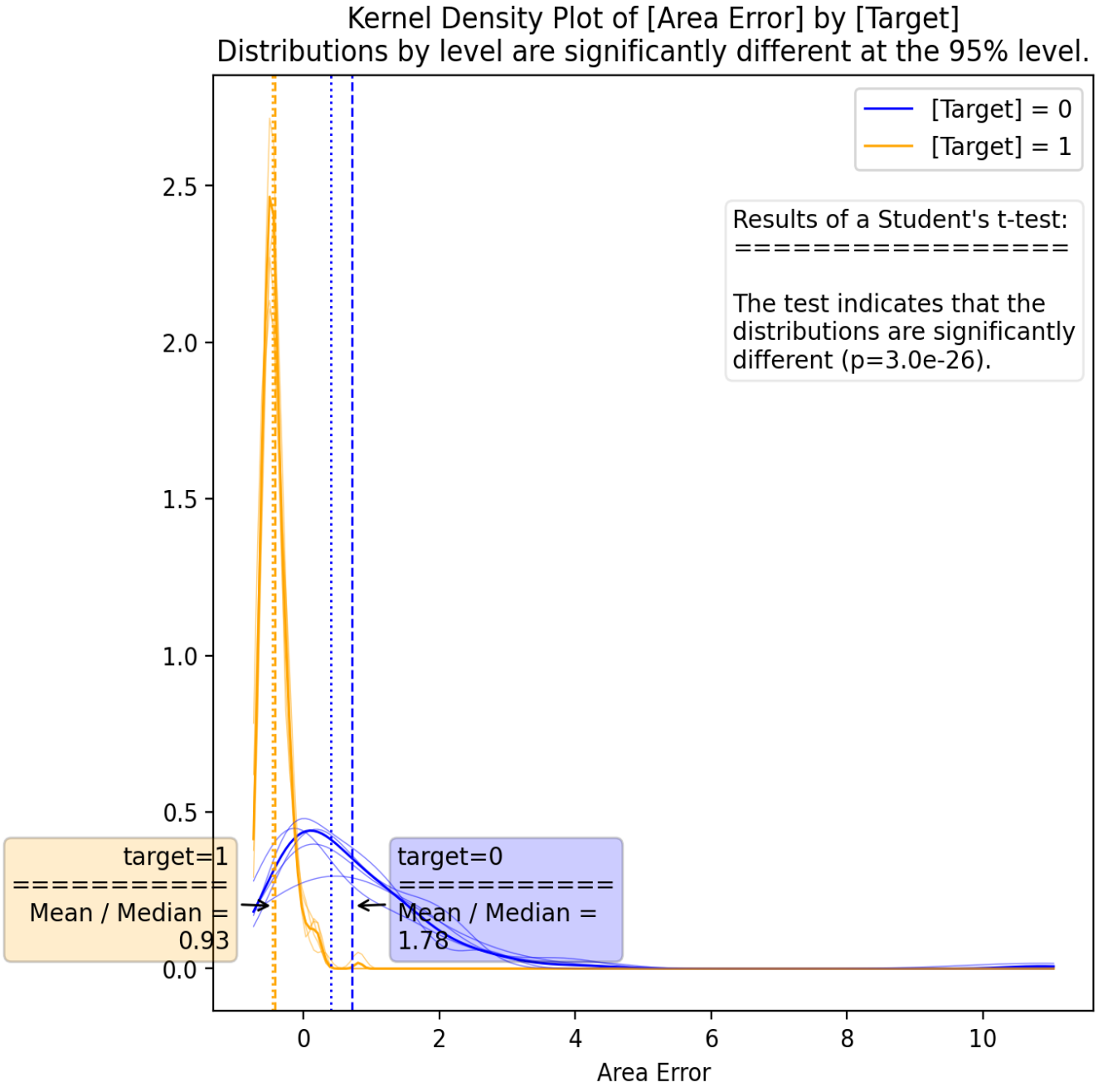
Univariate Report

Area Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.7e+00	-5.1e+00	-4.4e+00	-4.4e+00	-4.7e+00	-9.6e-03	2.6e-01
Fitted p-Value	8.0e-31	1.2e-32	2.5e-31	1.1e-32	2.2e-30	2.2e-07	9.1e-31
Fitted Std. Err.	0.411	0.426	0.380	0.373	0.411	0.002	0.022
Conf. Int. Lower	-5.5e+00	-5.9e+00	-5.2e+00	-5.2e+00	-5.5e+00	-1.3e-02	3.0e-01
Conf. Int. Upper	-3.9e+00	-4.2e+00	-3.7e+00	-3.7e+00	-3.9e+00	-5.9e-03	2.2e-01
Train Accuracy	86.3%	87.4%	84.9%	86.1%	87.2%	86.8%	1.0%
Val Accuracy	87.9%	80.2%	90.3%	91.0%	84.0%	86.8%	4.5%
Train AUC	83.1%	84.3%	81.6%	82.7%	83.7%	83.6%	1.0%
Val AUC	84.2%	75.3%	86.3%	90.1%	81.1%	83.6%	5.6%
Train F1	89.7%	90.6%	88.6%	89.8%	90.6%	90.2%	0.8%
Test F1	91.2%	85.7%	93.3%	92.4%	87.7%	90.2%	3.2%
Train Precision	83.9%	85.5%	82.0%	85.1%	85.4%	84.7%	1.5%
Val Precision	86.7%	76.9%	90.5%	88.7%	80.6%	84.7%	5.7%
Train Recall	96.4%	96.3%	96.4%	95.0%	96.4%	96.4%	0.6%
Val Recall	96.3%	96.8%	96.2%	96.5%	96.2%	96.4%	0.3%
Train MCC	71.0%	72.9%	68.5%	69.5%	72.2%	71.8%	1.8%
Val MCC	73.0%	58.8%	76.3%	81.8%	67.3%	71.8%	8.8%
Train Log-Loss	4.94	4.54	5.45	5.00	4.61	4.75	0.36
Val Log-Loss	4.36	7.14	3.51	3.24	5.78	4.75	1.64

Univariate Report

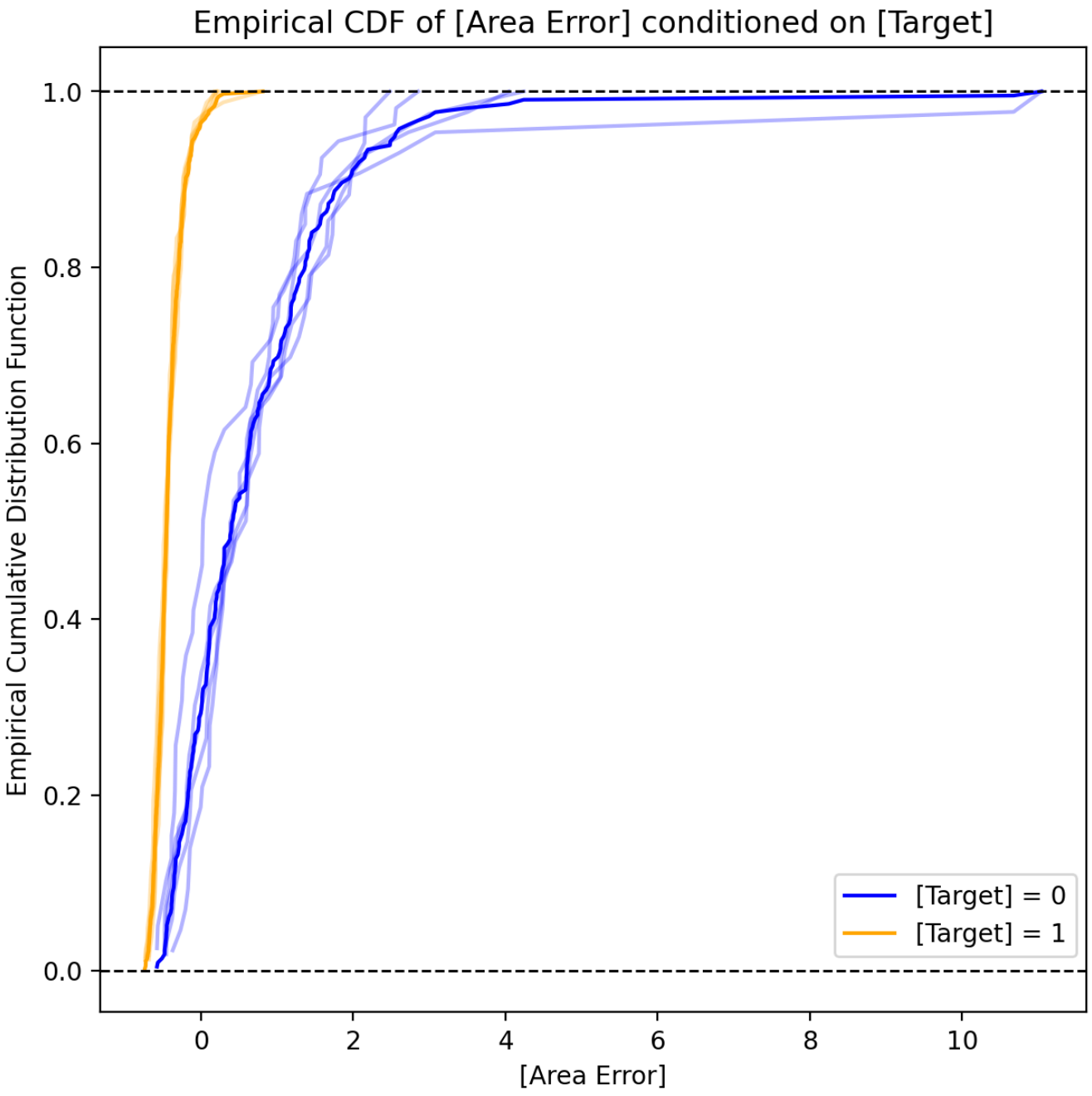
Area Error - Kernel Density Plot



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

Univariate Report

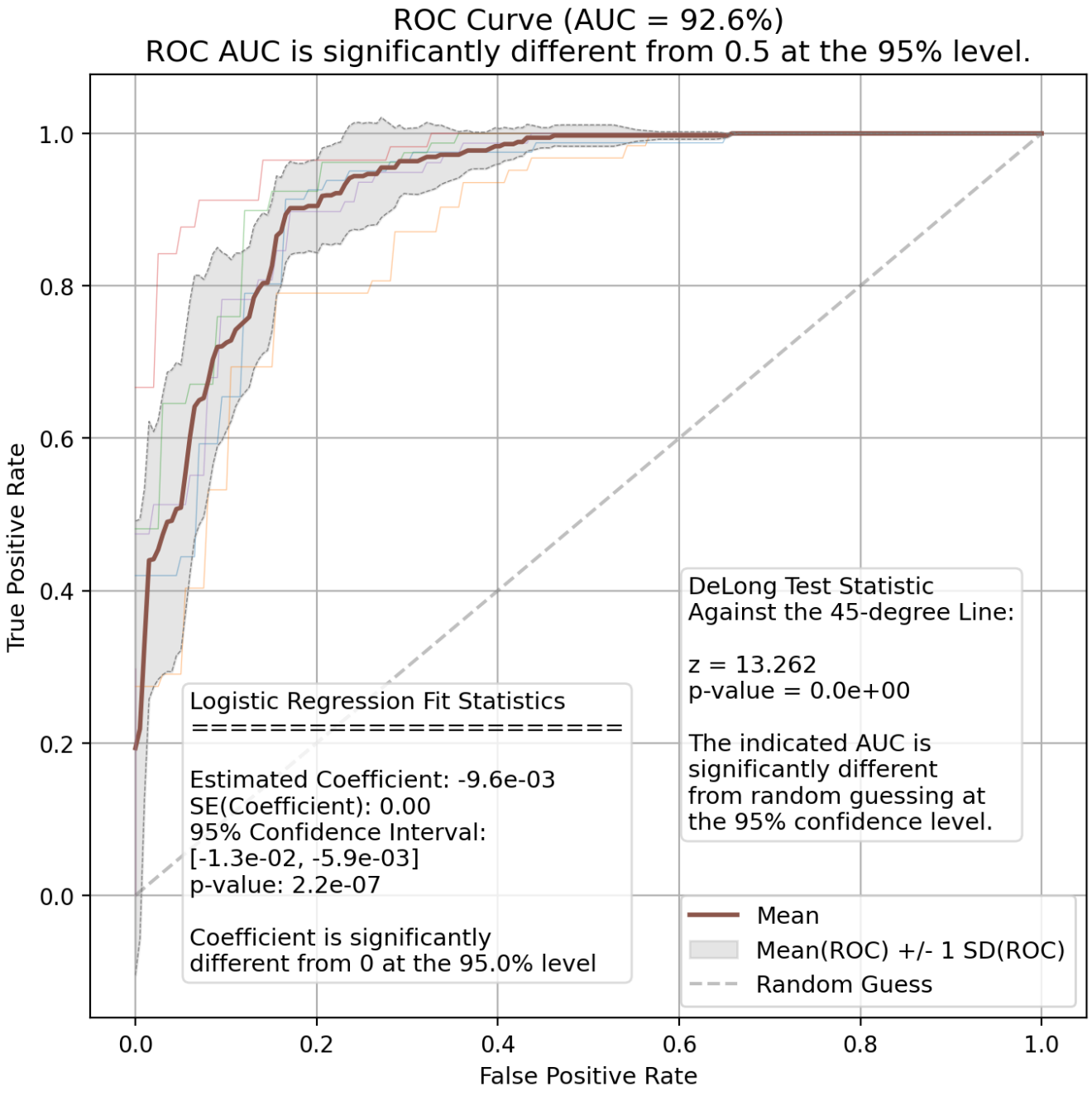
Area Error - Empirical CDF Plot



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

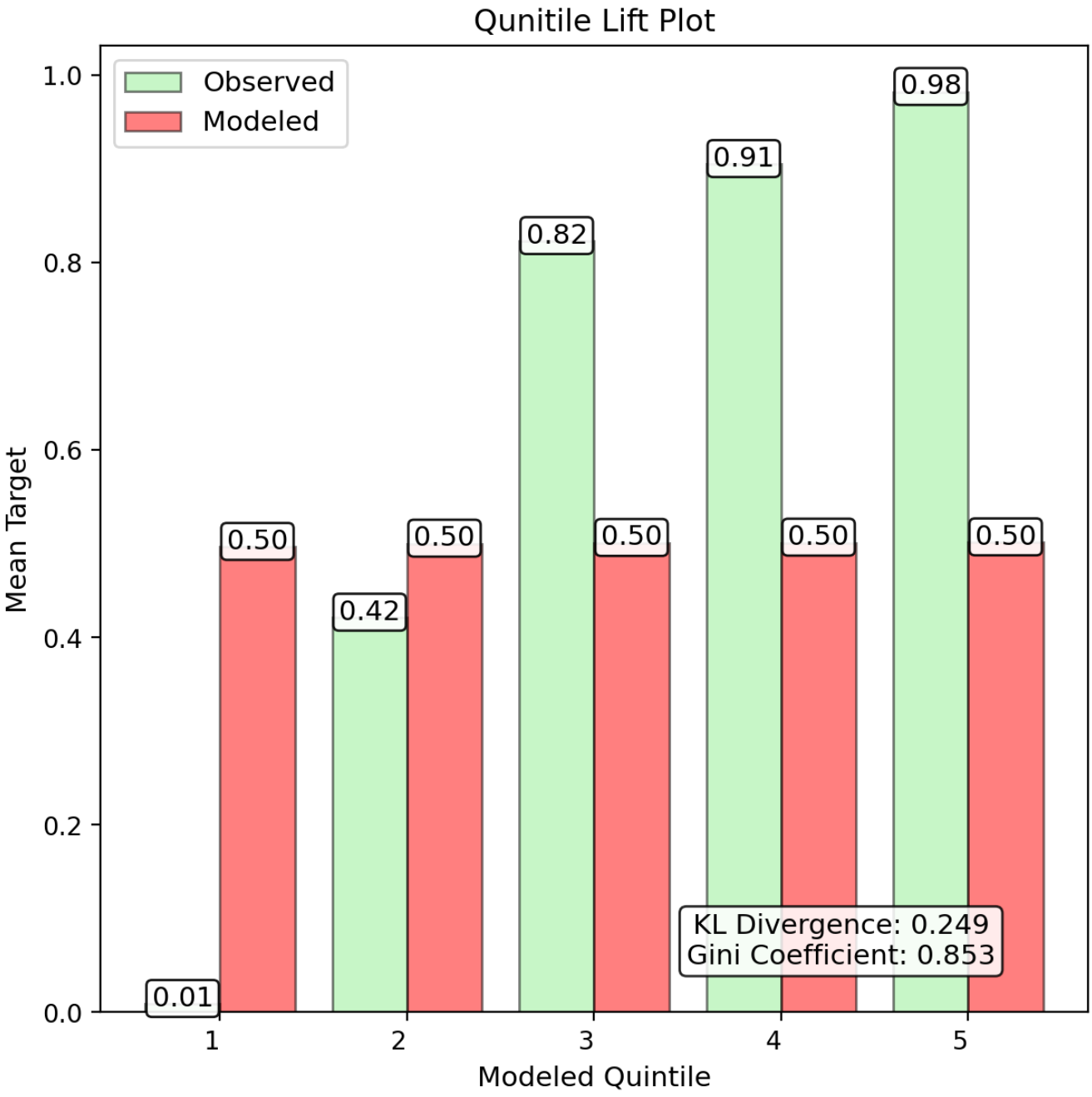


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

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.

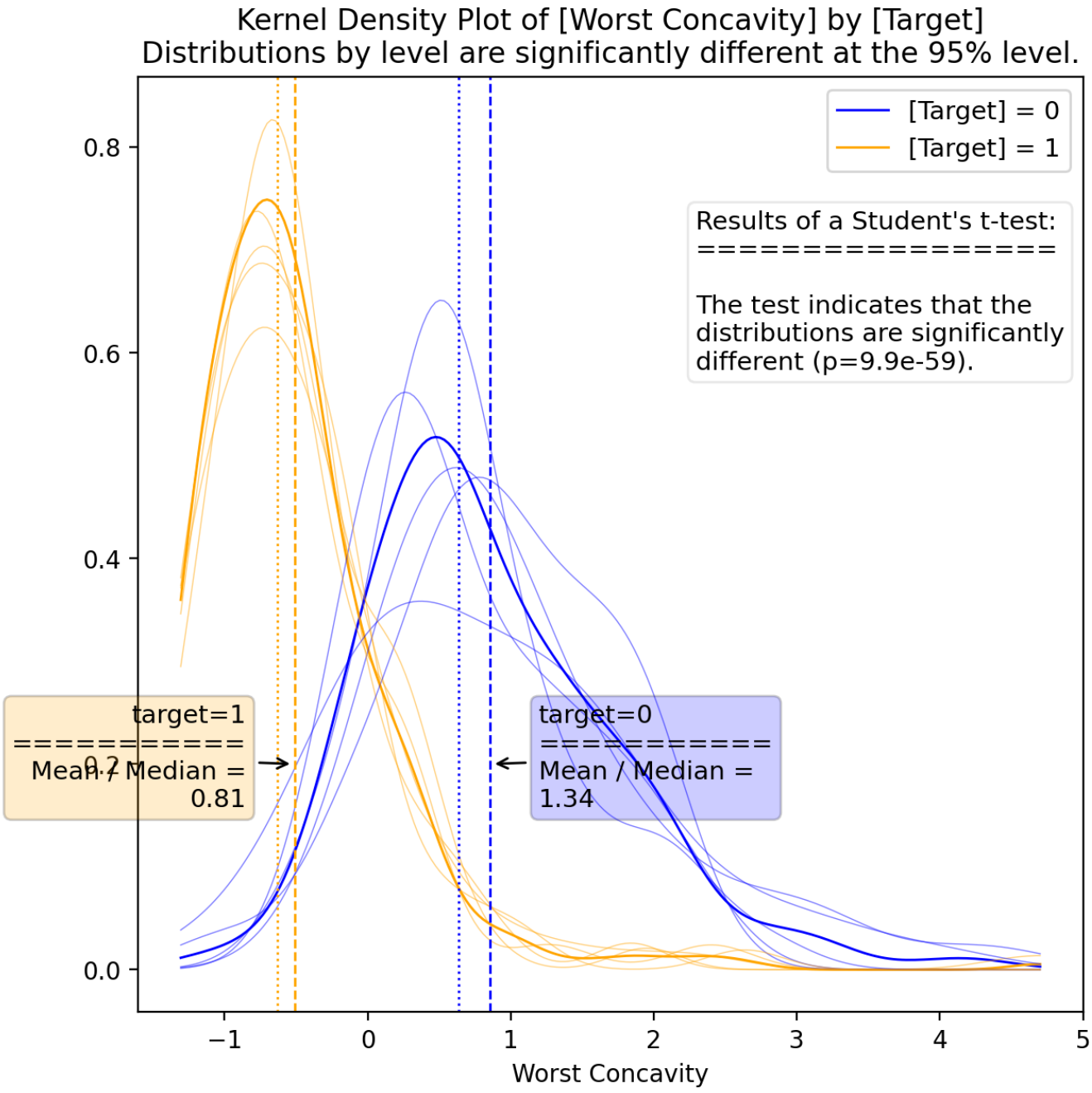
Univariate Report

Worst Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.6e+00	-2.2e+00	-2.3e+00	-2.3e+00	-2.4e+00	-1.1e+00	1.3e-01
Fitted p-Value	2.0e-28	2.0e-28	1.0e-27	1.1e-28	1.3e-27	2.0e-05	5.6e-28
Fitted Std. Err.	0.232	0.202	0.211	0.207	0.222	0.258	0.012
Conf. Int. Lower	-3.0e+00	-2.6e+00	-2.7e+00	-2.7e+00	-2.9e+00	-1.6e+00	1.6e-01
Conf. Int. Upper	-2.1e+00	-1.8e+00	-1.9e+00	-1.9e+00	-2.0e+00	-5.9e-01	1.1e-01
Train Accuracy	84.7%	83.8%	85.7%	84.9%	85.2%	84.7%	0.7%
Val Accuracy	84.7%	91.1%	81.4%	84.0%	84.7%	84.7%	3.6%
Train AUC	84.9%	83.9%	85.7%	84.8%	85.9%	84.8%	0.8%
Val AUC	84.5%	92.3%	82.5%	84.5%	83.9%	84.8%	3.9%
Train F1	87.2%	86.6%	88.0%	87.8%	87.7%	87.4%	0.6%
Test F1	87.9%	92.3%	85.7%	85.2%	87.3%	87.4%	2.8%
Train Precision	90.6%	90.1%	90.2%	90.7%	92.8%	90.7%	1.1%
Val Precision	90.8%	98.2%	92.6%	90.2%	86.2%	90.7%	4.3%
Train Recall	84.1%	83.4%	86.0%	85.0%	83.2%	84.3%	1.2%
Val Recall	85.2%	87.1%	79.7%	80.7%	88.5%	84.3%	3.9%
Train MCC	68.6%	66.4%	70.6%	68.2%	69.8%	68.3%	1.6%
Val MCC	67.3%	82.6%	60.9%	68.4%	68.2%	68.3%	8.0%
Train Log-Loss	5.51	5.85	5.14	5.46	5.35	5.51	0.26
Val Log-Loss	5.52	3.21	6.70	5.77	5.50	5.51	1.29

Univariate Report

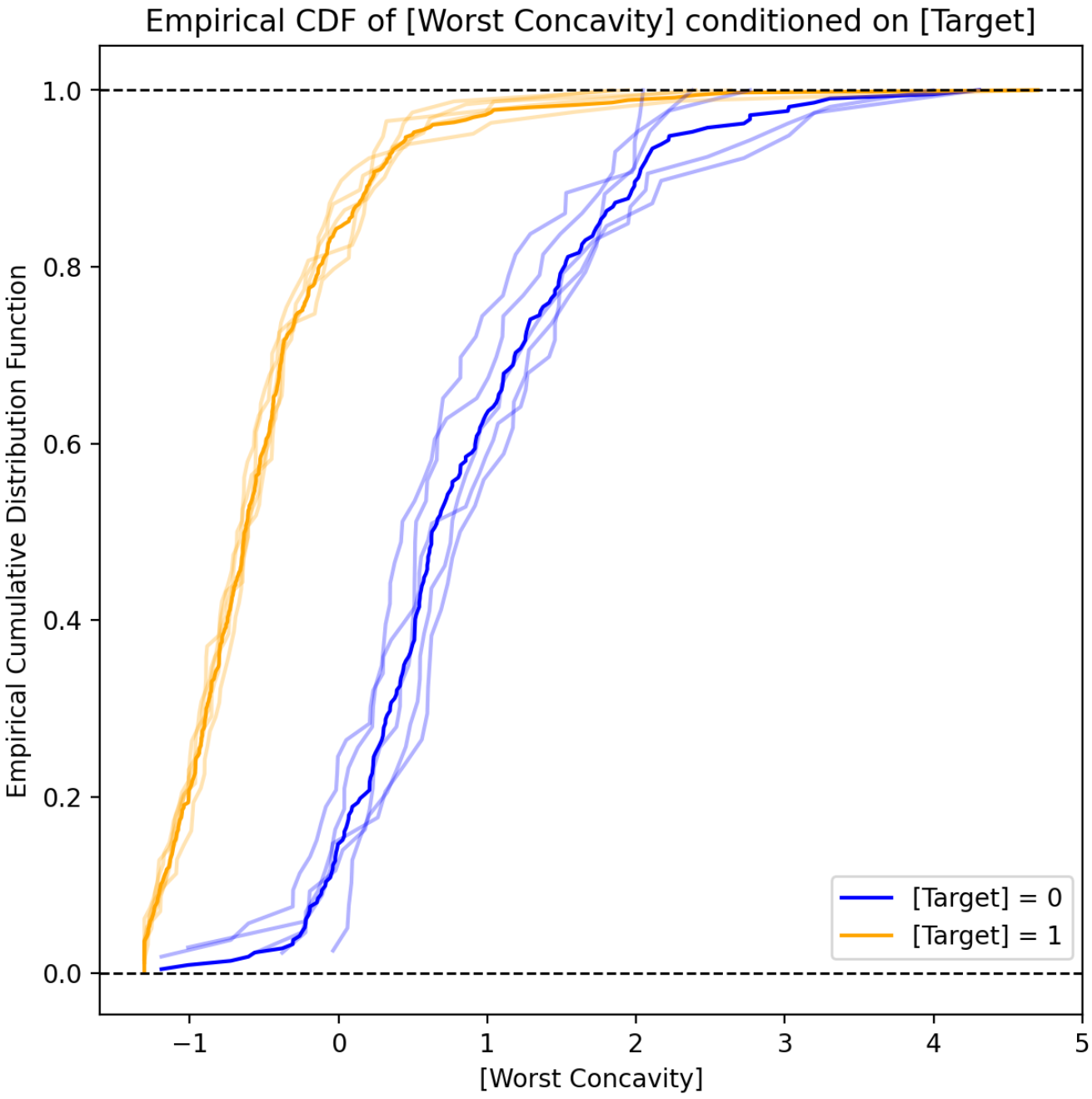
Worst Concavity - Kernel Density Plot



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

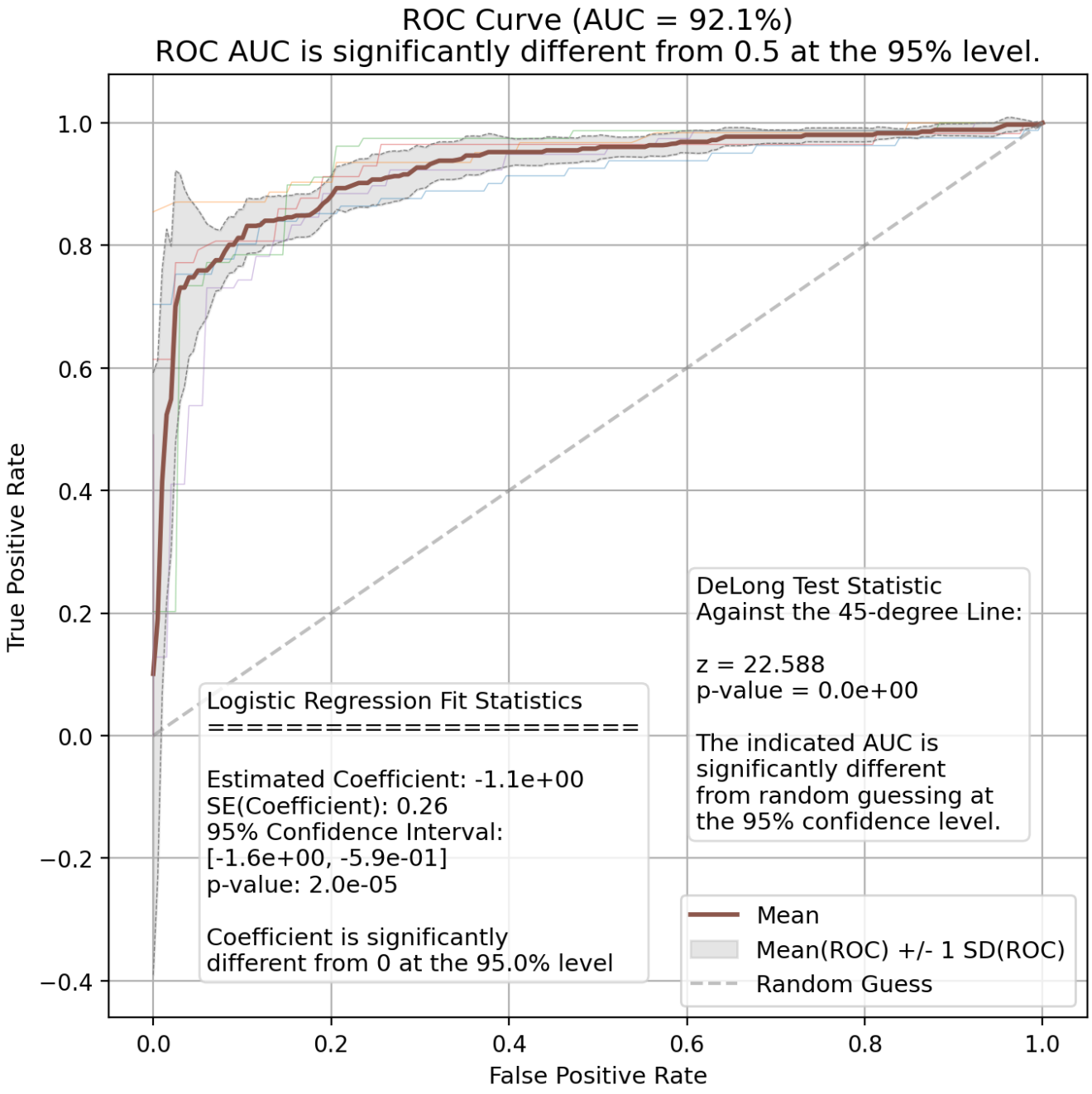
Worst Concavity - Empirical CDF Plot



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

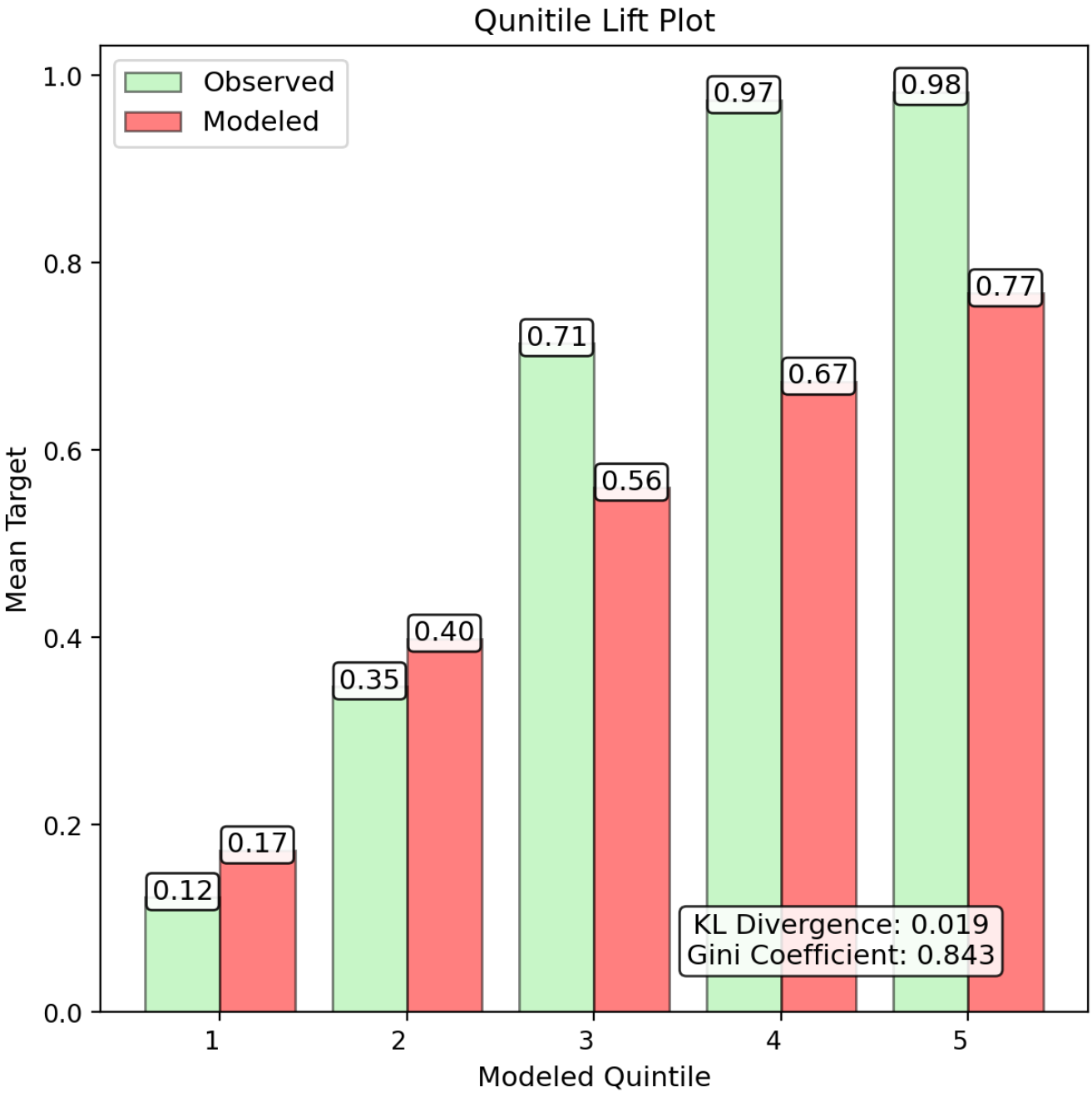


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst 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.

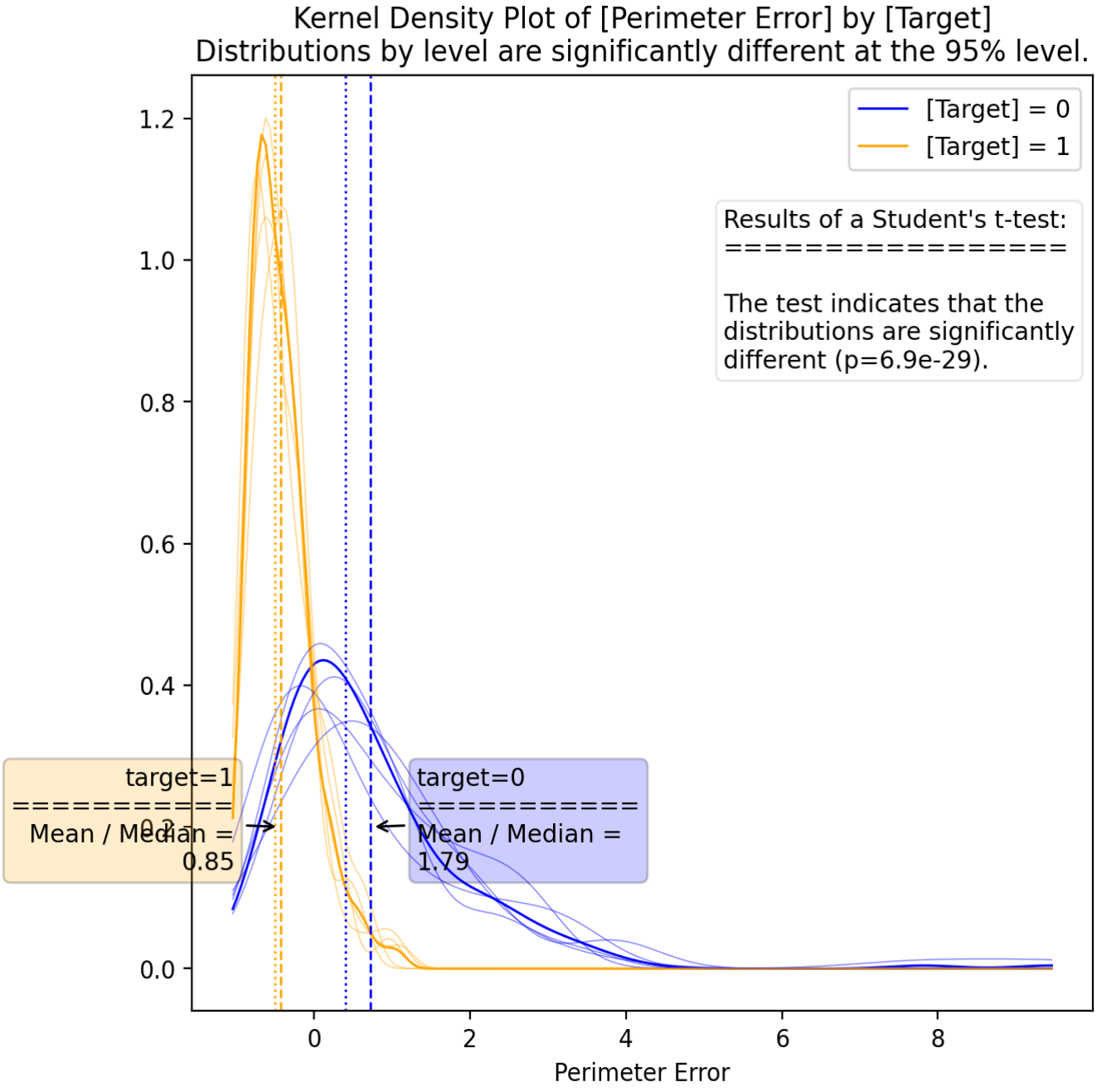
Univariate Report

Perimeter Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.0e+00	-3.2e+00	-2.9e+00	-2.8e+00	-3.0e+00	-5.8e-02	1.4e-01
Fitted p-Value	1.1e-26	1.9e-28	8.4e-27	1.3e-27	2.1e-26	1.9e-02	8.6e-27
Fitted Std. Err.	0.281	0.286	0.270	0.256	0.287	0.025	0.013
Conf. Int. Lower	-3.6e+00	-3.7e+00	-3.4e+00	-3.3e+00	-3.6e+00	-1.1e-01	1.7e-01
Conf. Int. Upper	-2.5e+00	-2.6e+00	-2.4e+00	-2.3e+00	-2.5e+00	-9.6e-03	1.2e-01
Train Accuracy	81.6%	83.1%	82.7%	80.8%	82.6%	81.9%	0.9%
Val Accuracy	83.1%	78.2%	80.5%	88.0%	80.9%	81.9%	3.7%
Train AUC	79.4%	81.4%	80.4%	78.3%	80.6%	79.8%	1.2%
Val AUC	80.5%	73.7%	78.5%	88.3%	78.2%	79.8%	5.4%
Train F1	85.6%	86.8%	86.4%	85.3%	86.6%	85.9%	0.6%
Test F1	87.3%	84.1%	85.7%	89.1%	85.2%	85.9%	2.0%
Train Precision	83.0%	85.5%	82.6%	83.4%	85.1%	84.0%	1.3%
Val Precision	85.7%	76.3%	88.0%	92.5%	79.1%	84.0%	6.6%
Train Recall	88.4%	88.1%	90.6%	87.3%	88.2%	88.0%	1.2%
Val Recall	88.9%	93.5%	83.5%	86.0%	92.3%	88.0%	4.2%
Train MCC	60.3%	63.4%	63.1%	57.7%	62.0%	60.8%	2.4%
Val MCC	62.1%	53.5%	55.4%	76.0%	60.2%	60.8%	8.9%
Train Log-Loss	6.64	6.08	6.24	6.92	6.25	6.52	0.34
Val Log-Loss	6.10	7.85	7.02	4.33	6.88	6.52	1.33

Univariate Report

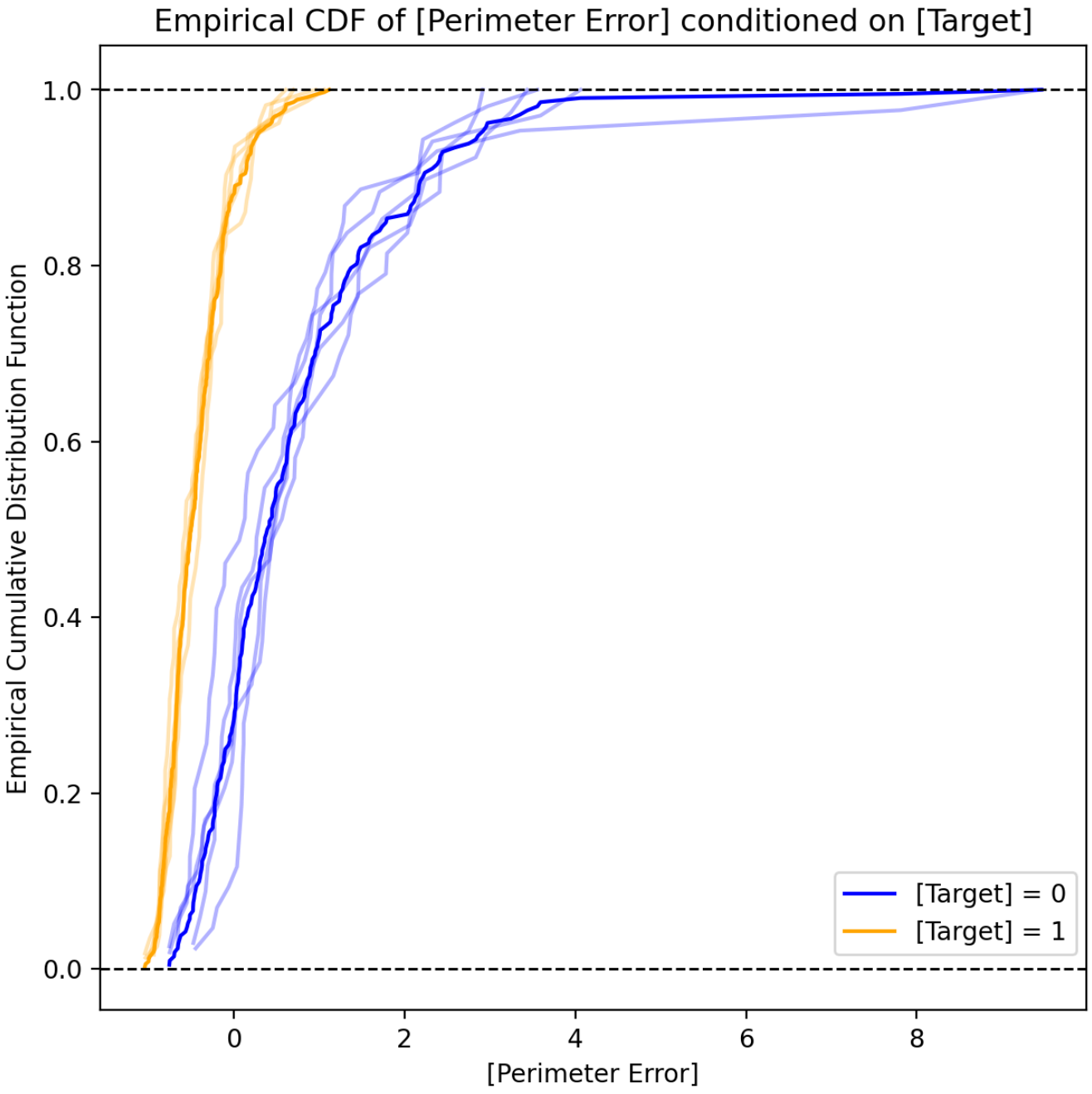
Perimeter Error - Kernel Density Plot



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

Univariate Report

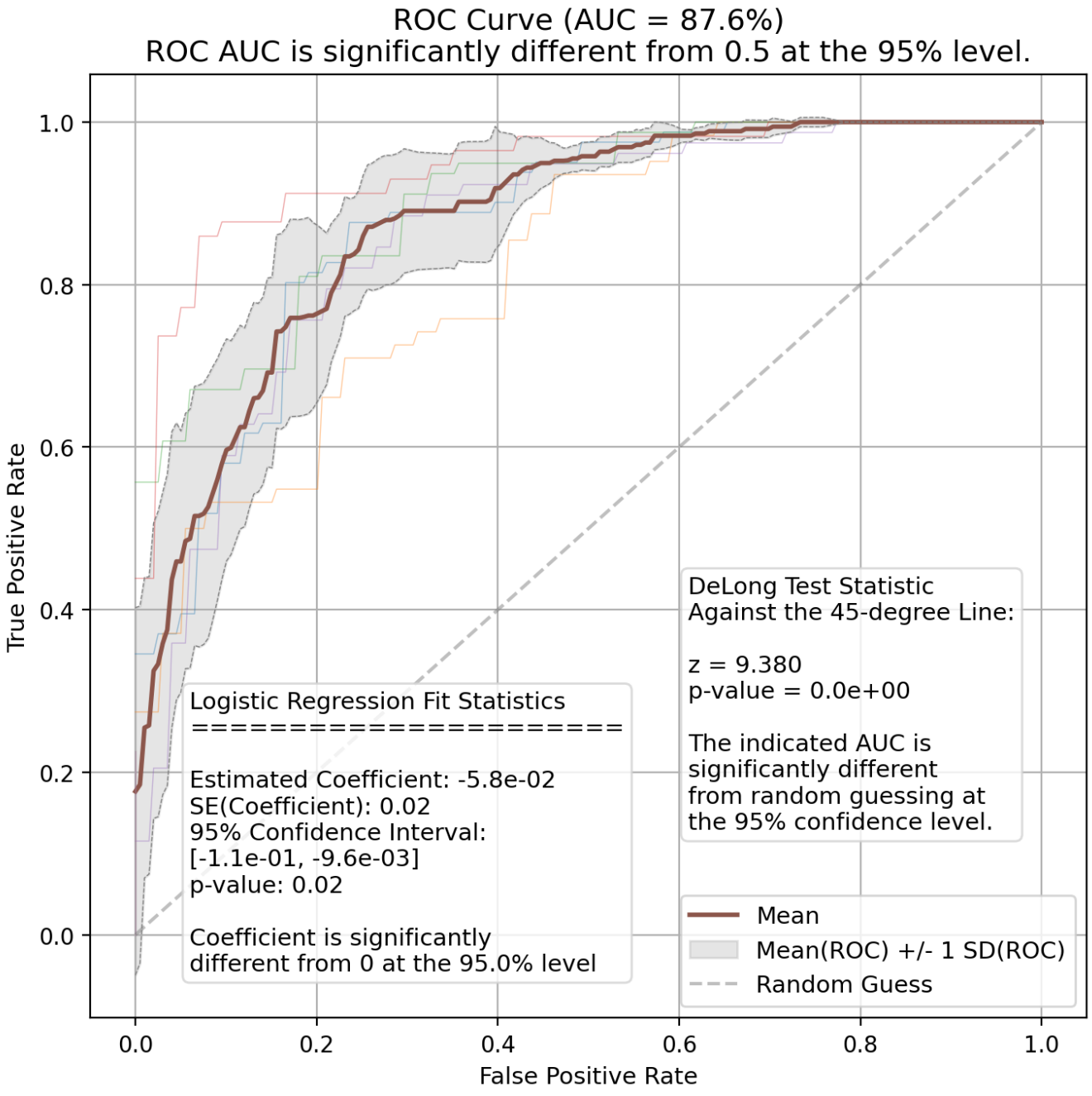
Perimeter Error - Empirical CDF Plot



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

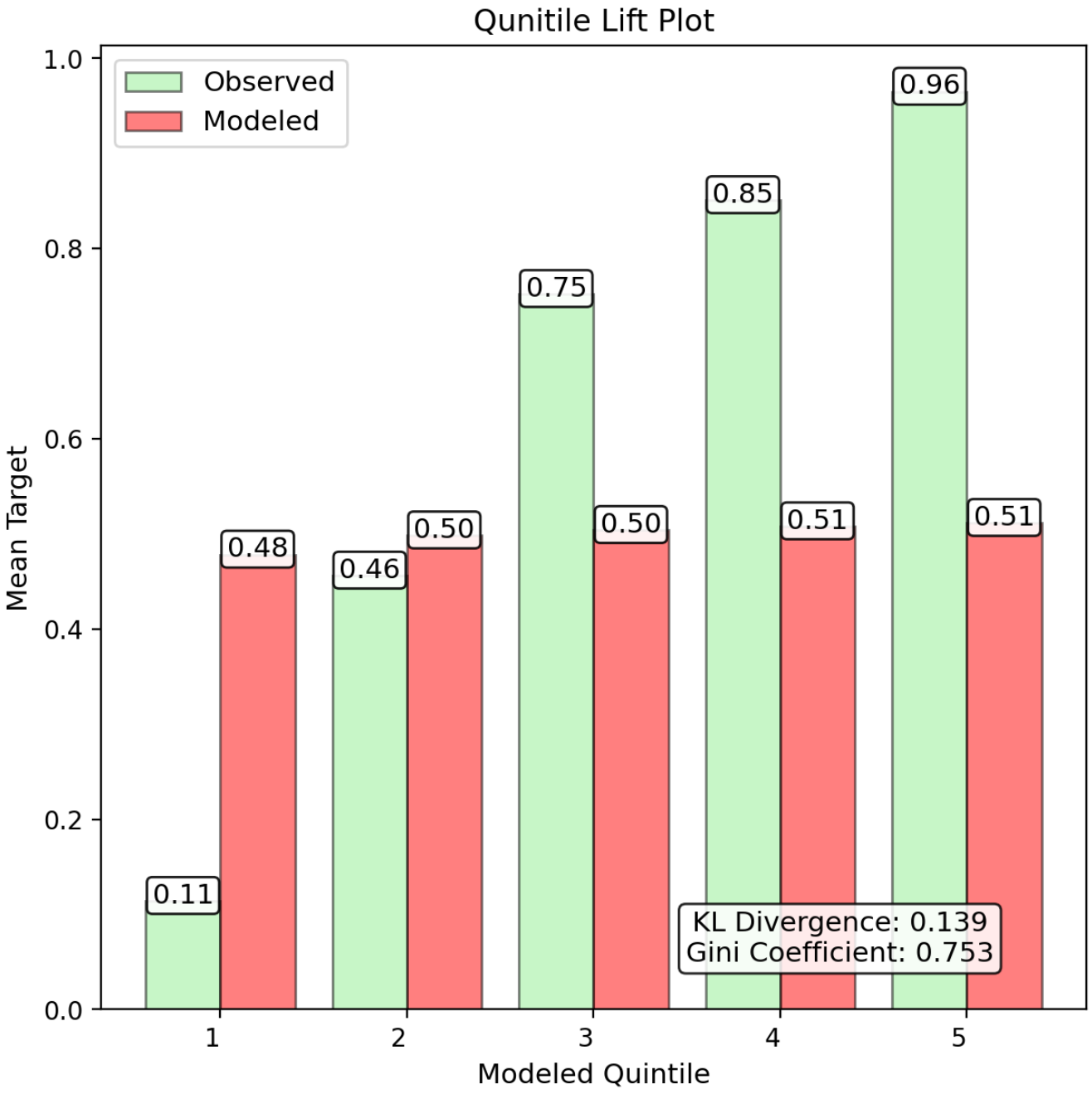


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

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.

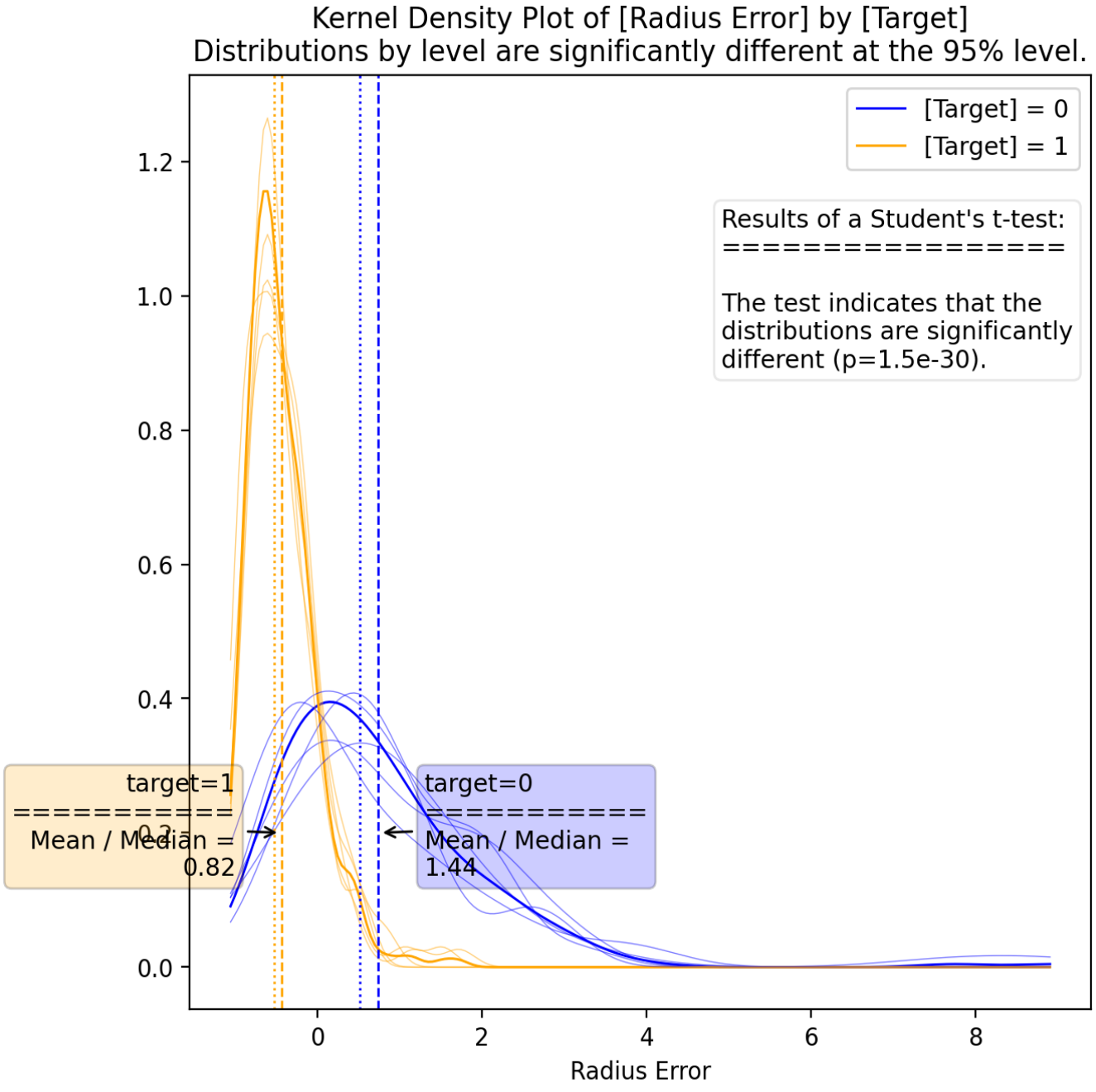
Univariate Report

Radius Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.8e+00	-2.9e+00	-2.6e+00	-2.6e+00	-2.8e+00	-4.0e-01	1.4e-01
Fitted p-Value	6.5e-27	4.0e-28	2.6e-26	3.1e-27	6.0e-26	2.1e-02	2.5e-26
Fitted Std. Err.	0.261	0.262	0.247	0.237	0.267	0.175	0.012
Conf. Int. Lower	-3.3e+00	-3.4e+00	-3.1e+00	-3.0e+00	-3.3e+00	-7.5e-01	1.6e-01
Conf. Int. Upper	-2.3e+00	-2.4e+00	-2.1e+00	-2.1e+00	-2.3e+00	-6.0e-02	1.1e-01
Train Accuracy	81.1%	83.8%	81.4%	80.6%	82.2%	81.5%	1.2%
Val Accuracy	82.3%	76.2%	84.1%	85.0%	80.9%	81.5%	3.4%
Train AUC	78.9%	82.0%	79.3%	78.5%	80.1%	79.5%	1.4%
Val AUC	80.4%	71.6%	81.9%	85.1%	78.5%	79.5%	5.0%
Train F1	85.3%	87.3%	85.3%	85.0%	86.3%	85.6%	1.0%
Test F1	86.4%	82.6%	88.5%	86.5%	85.0%	85.6%	2.2%
Train Precision	82.7%	85.9%	82.1%	84.0%	84.8%	83.7%	1.6%
Val Precision	86.4%	75.0%	89.6%	88.9%	79.8%	83.7%	6.3%
Train Recall	88.0%	88.8%	88.8%	86.0%	87.8%	87.7%	1.2%
Val Recall	86.4%	91.9%	87.3%	84.2%	91.0%	87.7%	3.2%
Train MCC	59.3%	64.8%	60.3%	57.6%	61.0%	60.0%	2.7%
Val MCC	60.8%	48.8%	62.8%	69.8%	60.0%	60.0%	7.6%
Train Log-Loss	6.80	5.85	6.72	6.99	6.42	6.65	0.44
Val Log-Loss	6.39	8.56	5.74	5.41	6.88	6.65	1.24

Univariate Report

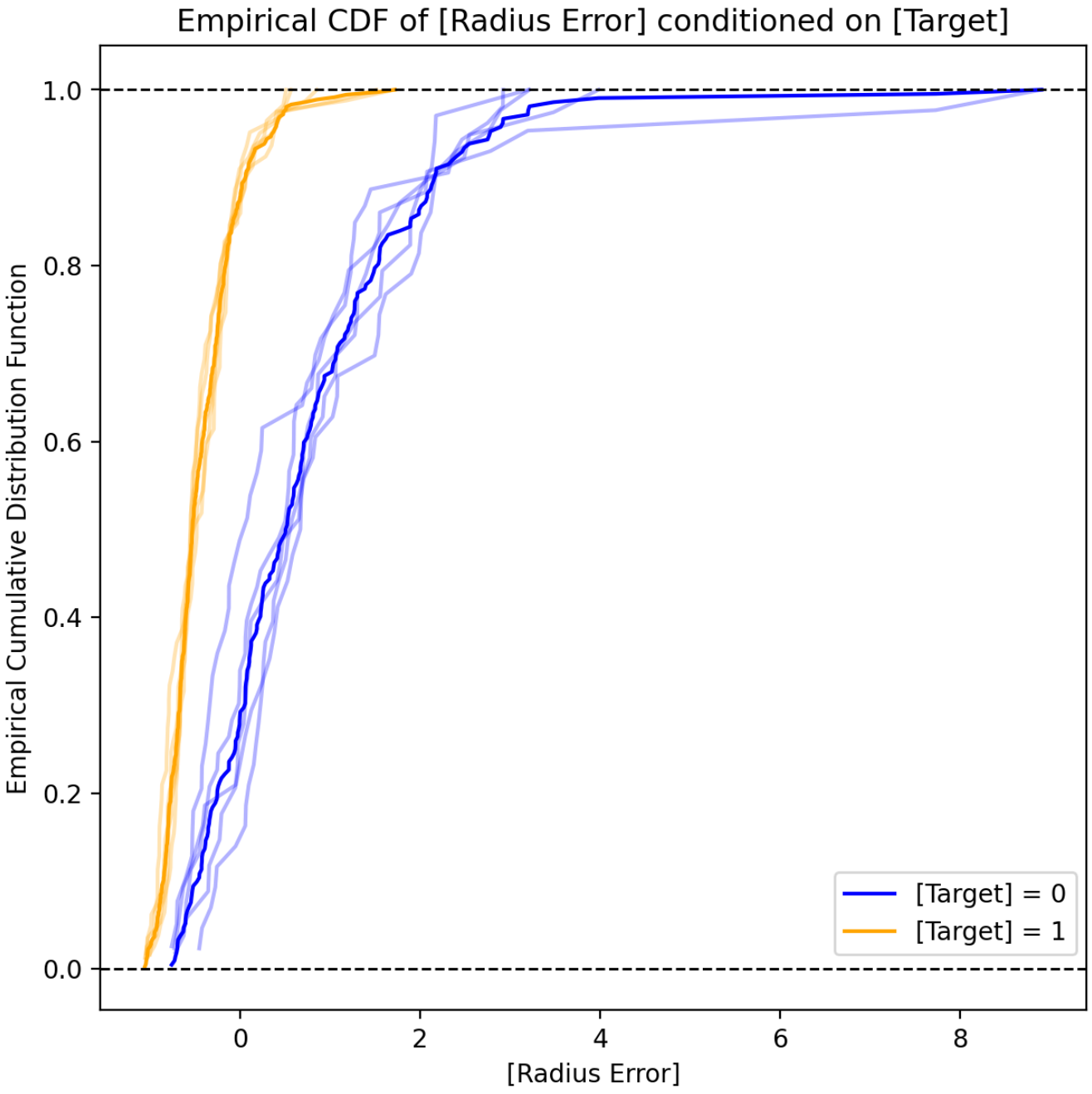
Radius Error - Kernel Density Plot



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

Univariate Report

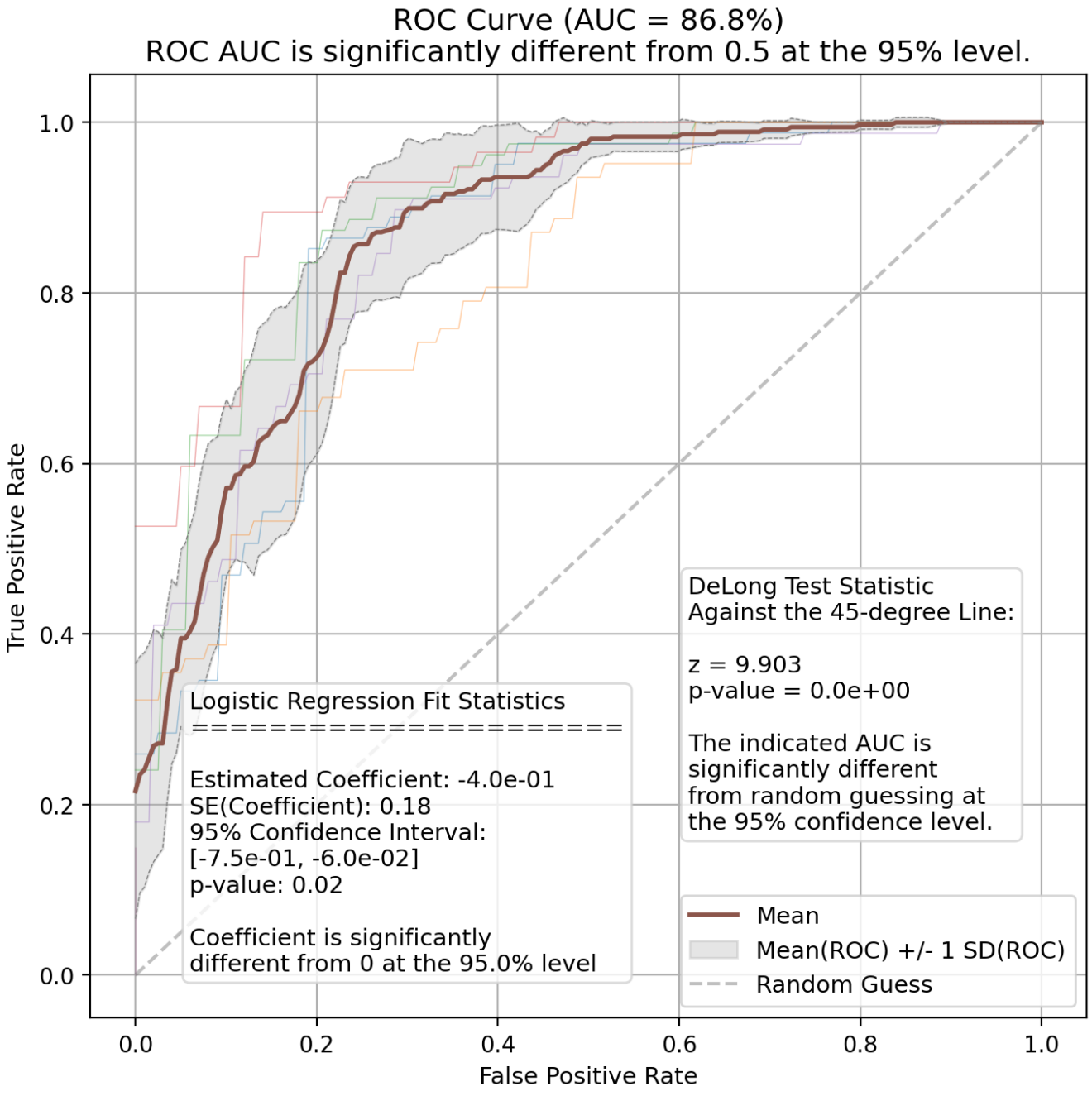
Radius Error - Empirical CDF Plot



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

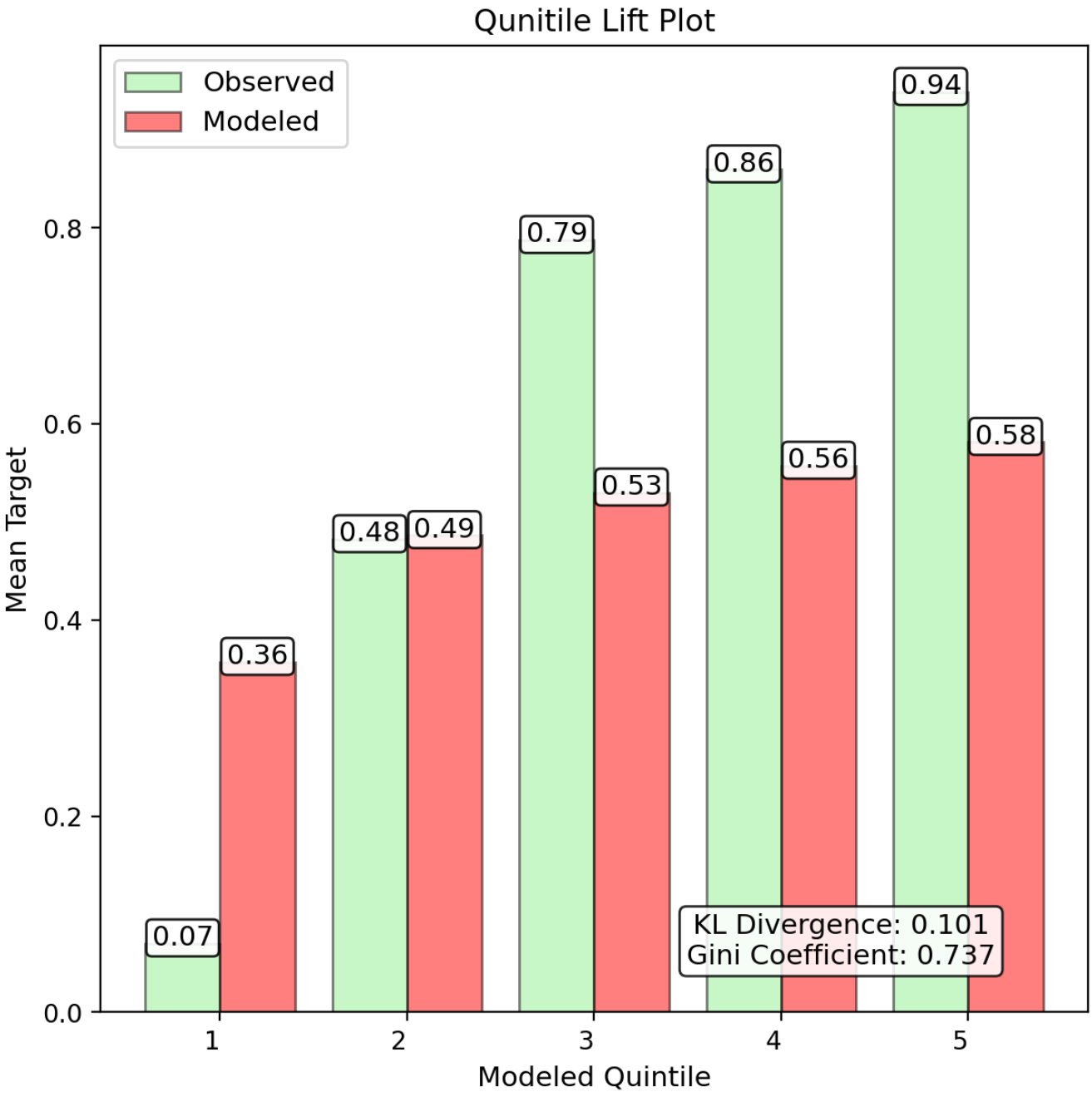


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

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.

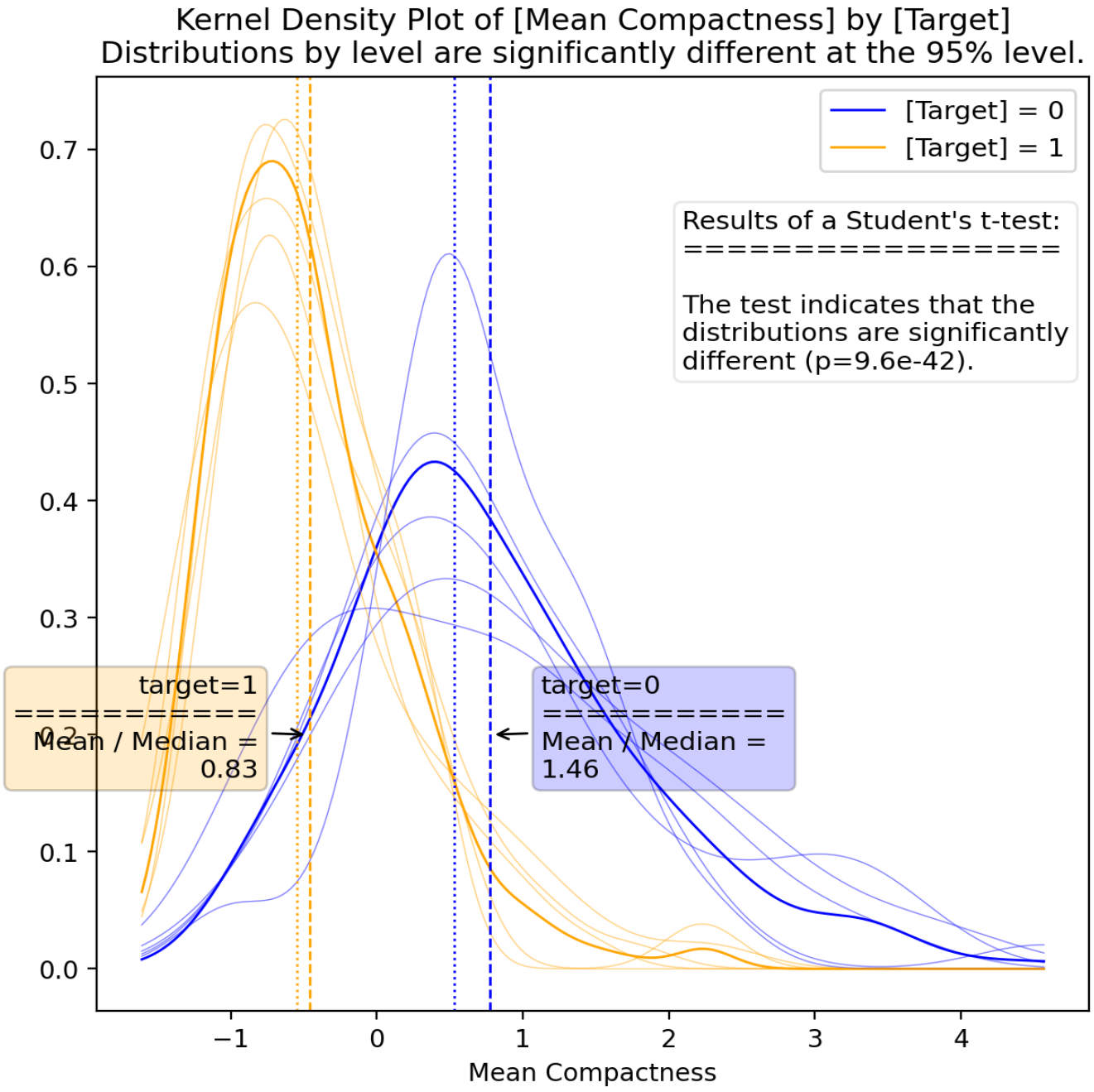
Univariate Report

Mean Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.9e+00	-1.8e+00	-1.9e+00	-2.0e+00	-2.0e+00	-5.5e-01	9.2e-02
Fitted p-Value	5.0e-25	2.2e-25	1.1e-25	1.2e-25	1.2e-24	4.4e-01	4.8e-25
Fitted Std. Err.	0.188	0.170	0.180	0.187	0.197	0.718	0.010
Conf. Int. Lower	-2.3e+00	-2.1e+00	-2.2e+00	-2.3e+00	-2.4e+00	-2.0e+00	1.1e-01
Conf. Int. Upper	-1.6e+00	-1.4e+00	-1.5e+00	-1.6e+00	-1.6e+00	8.5e-01	7.3e-02
Train Accuracy	79.6%	78.6%	78.5%	78.3%	80.4%	79.1%	0.9%
Val Accuracy	77.4%	81.2%	81.4%	83.0%	74.8%	79.1%	3.4%
Train AUC	79.2%	78.7%	78.5%	78.0%	81.1%	79.1%	1.2%
Val AUC	78.9%	80.9%	81.7%	83.9%	73.1%	79.1%	4.1%
Train F1	83.1%	82.2%	81.6%	82.3%	83.6%	82.6%	0.8%
Test F1	81.1%	84.3%	85.9%	83.8%	79.5%	82.6%	2.6%
Train Precision	85.4%	86.5%	85.2%	85.9%	89.4%	86.5%	1.7%
Val Precision	89.6%	86.4%	91.4%	91.7%	77.1%	86.5%	6.0%
Train Recall	80.8%	78.3%	78.4%	79.0%	78.5%	79.0%	1.0%
Val Recall	74.1%	82.3%	81.0%	77.2%	82.1%	79.0%	3.6%
Train MCC	57.5%	56.1%	56.1%	54.6%	60.2%	56.9%	2.1%
Val MCC	55.2%	61.0%	59.9%	67.3%	47.1%	56.9%	7.5%
Train Log-Loss	7.37	7.70	7.75	7.84	7.08	7.54	0.32
Val Log-Loss	8.14	6.78	6.70	6.13	9.08	7.54	1.21

Univariate Report

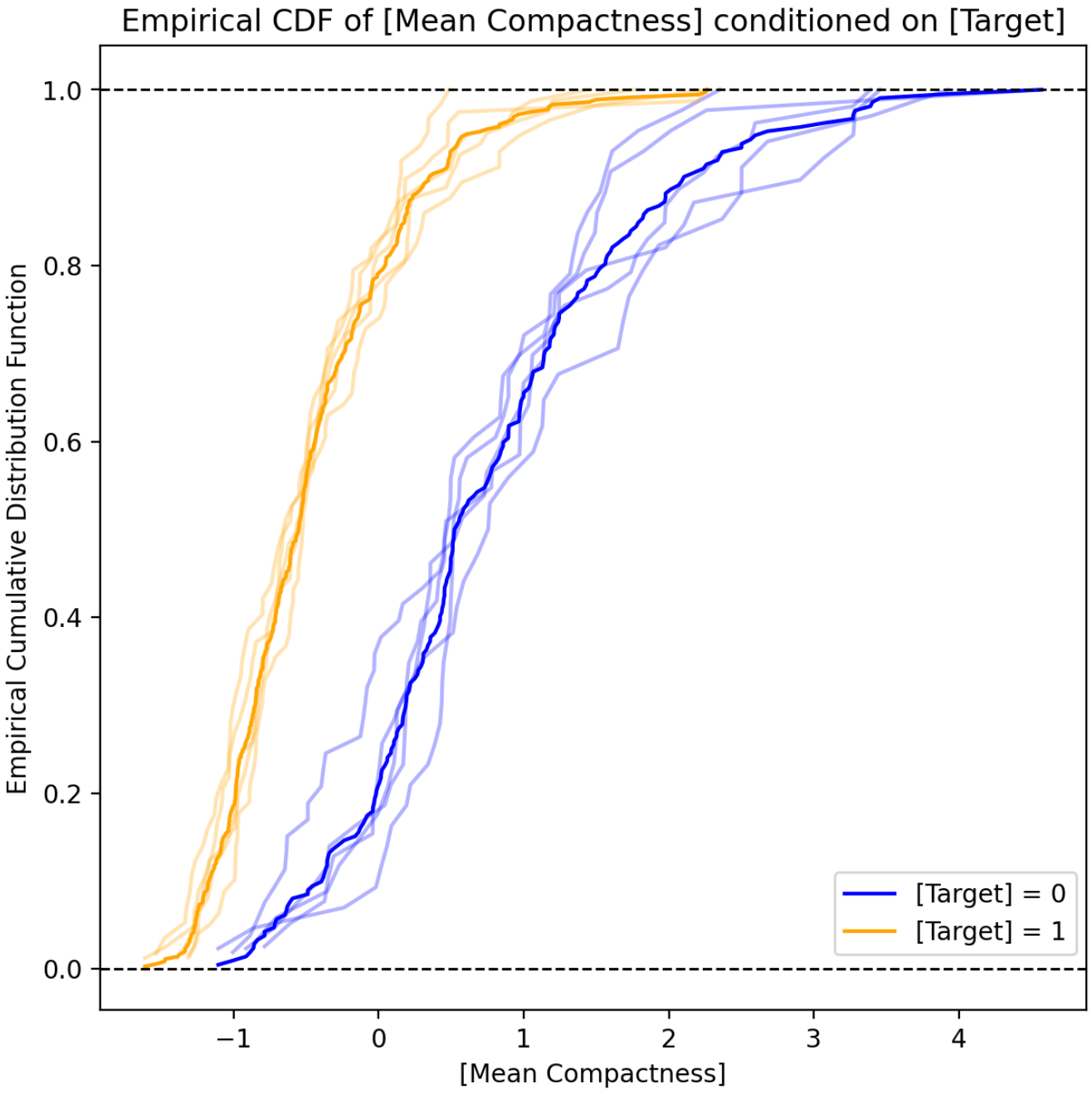
Mean Compactness - Kernel Density Plot



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

Univariate Report

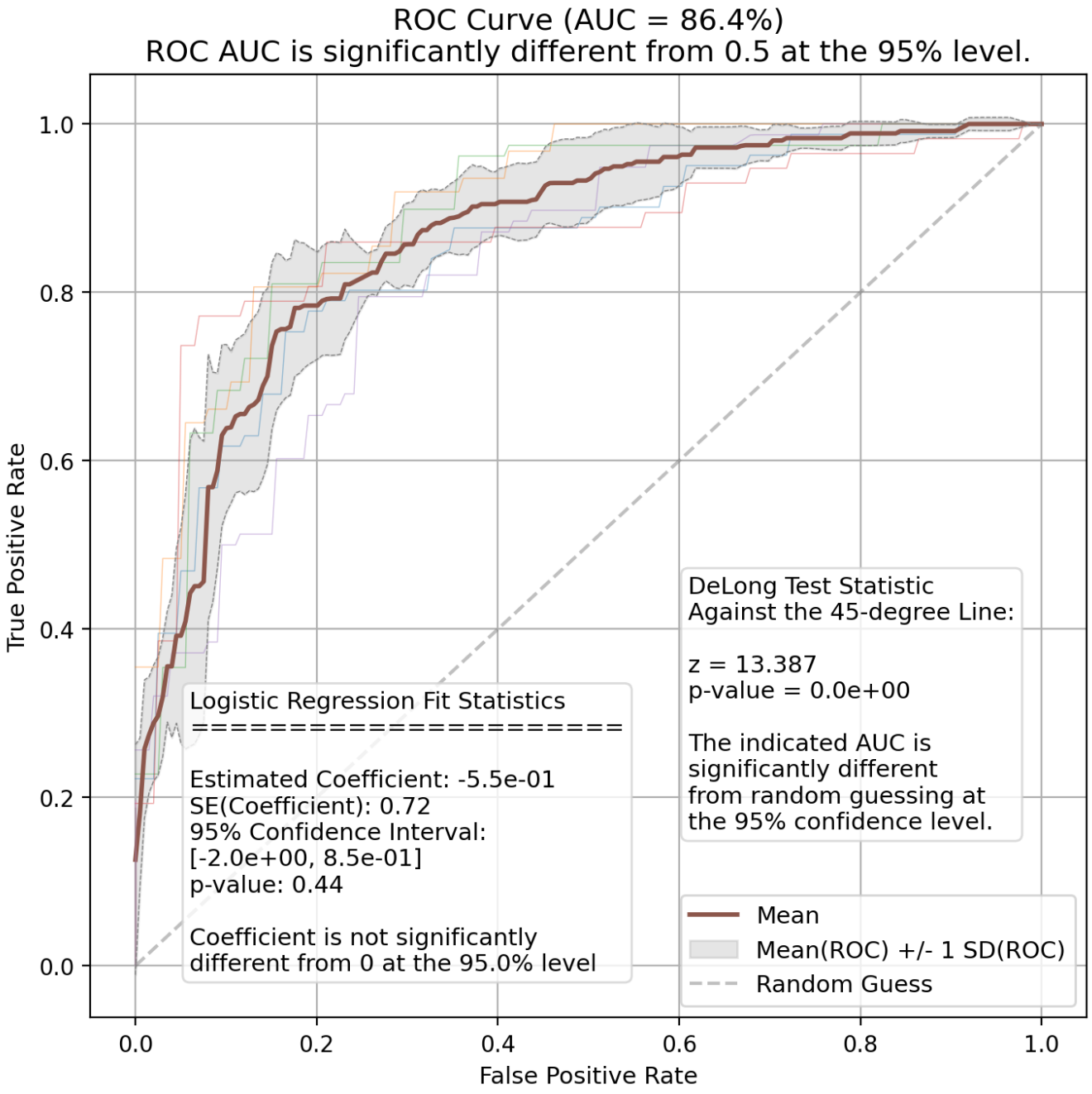
Mean Compactness - Empirical CDF Plot



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

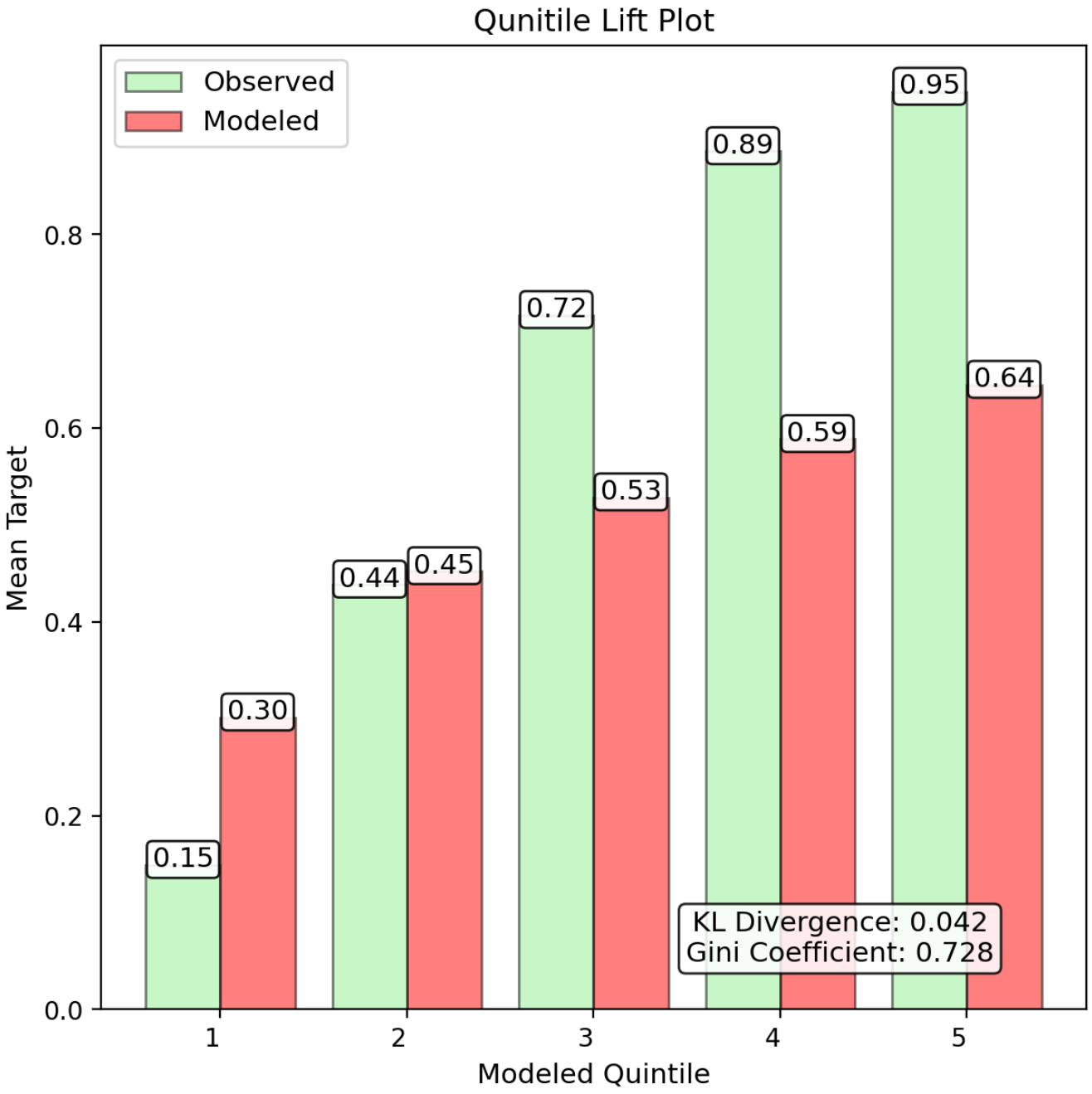


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean 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.

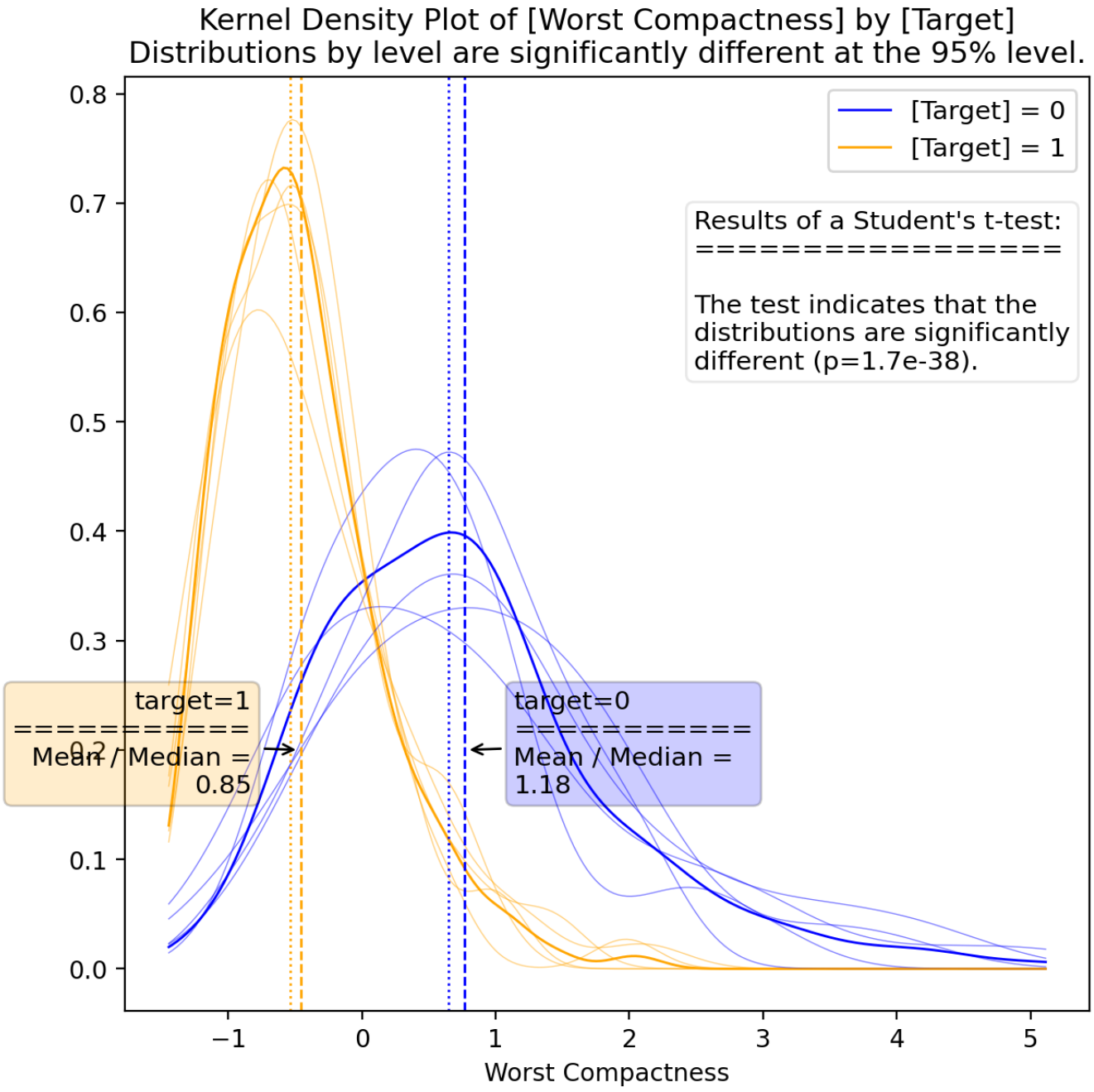
Univariate Report

Worst Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.1e+00	-1.9e+00	-2.0e+00	-2.1e+00	-2.1e+00	-5.5e-01	6.1e-02
Fitted p-Value	2.6e-25	5.4e-26	2.4e-25	2.7e-26	4.2e-25	5.1e-02	1.6e-25
Fitted Std. Err.	0.198	0.184	0.193	0.198	0.199	0.283	0.006
Conf. Int. Lower	-2.5e+00	-2.3e+00	-2.4e+00	-2.5e+00	-2.4e+00	-1.1e+00	7.3e-02
Conf. Int. Upper	-1.7e+00	-1.6e+00	-1.6e+00	-1.7e+00	-1.7e+00	2.3e-03	5.0e-02
Train Accuracy	79.3%	76.5%	78.7%	78.7%	80.4%	78.7%	1.4%
Val Accuracy	79.8%	83.2%	81.4%	79.0%	74.8%	78.7%	3.1%
Train AUC	78.1%	75.7%	77.7%	77.4%	79.9%	77.7%	1.5%
Val AUC	78.0%	83.0%	80.0%	78.7%	73.1%	77.7%	3.6%
Train F1	83.3%	80.8%	82.5%	83.1%	84.1%	82.8%	1.2%
Test F1	84.5%	86.0%	86.3%	81.4%	79.5%	82.8%	3.0%
Train Precision	83.3%	83.2%	82.7%	84.2%	86.7%	83.9%	1.6%
Val Precision	85.0%	88.1%	89.2%	82.1%	77.1%	83.9%	4.9%
Train Recall	83.3%	78.6%	82.4%	82.0%	81.7%	81.8%	1.8%
Val Recall	84.0%	83.9%	83.5%	80.7%	82.1%	81.8%	1.4%
Train MCC	56.1%	50.6%	55.4%	54.3%	58.6%	54.9%	2.9%
Val MCC	55.7%	65.1%	57.9%	57.3%	47.1%	54.9%	6.4%
Train Log-Loss	7.45	8.47	7.67	7.69	7.08	7.66	0.51
Val Log-Loss	7.27	6.07	6.70	7.57	9.08	7.66	1.13

Univariate Report

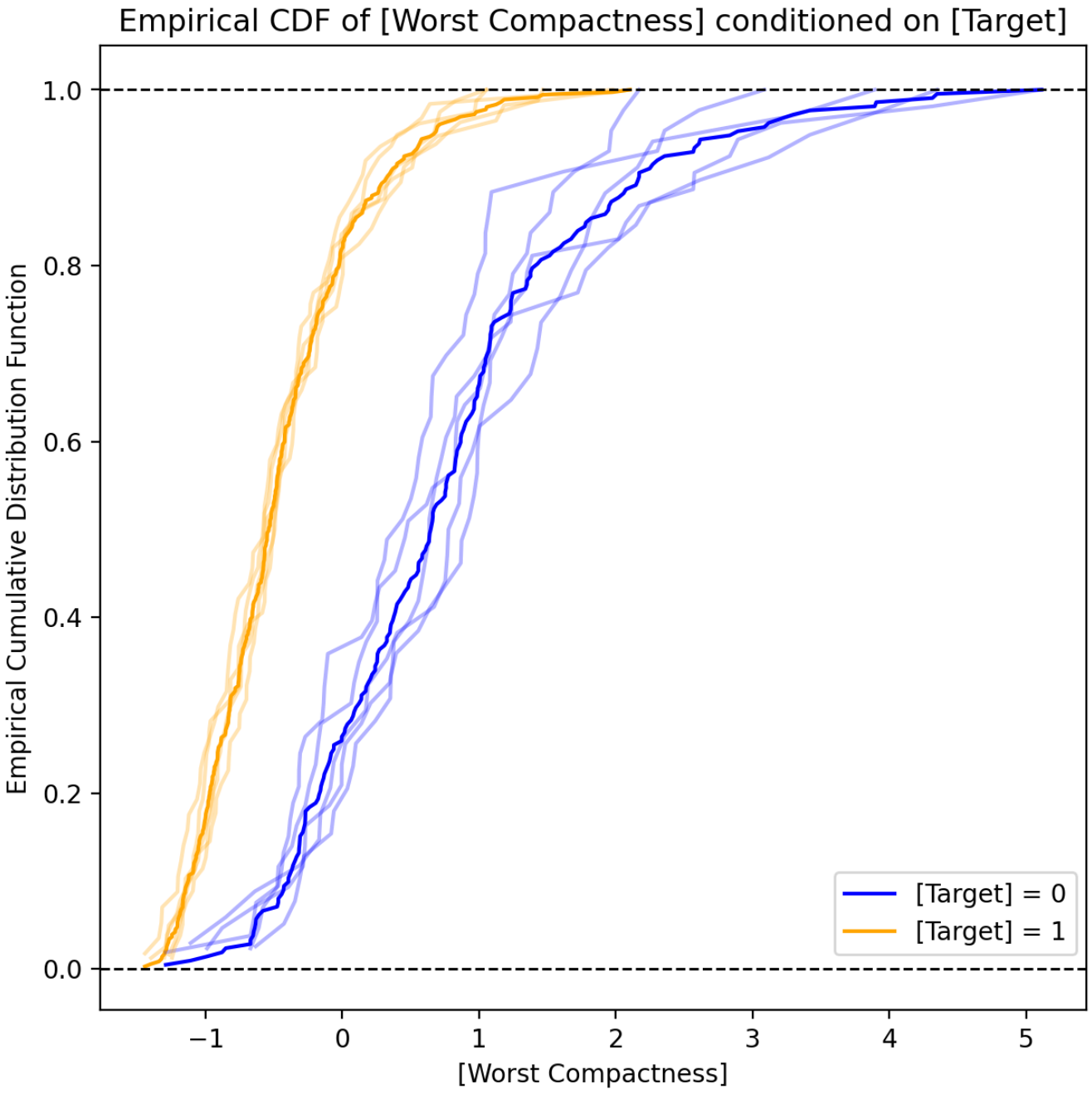
Worst Compactness - Kernel Density Plot



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

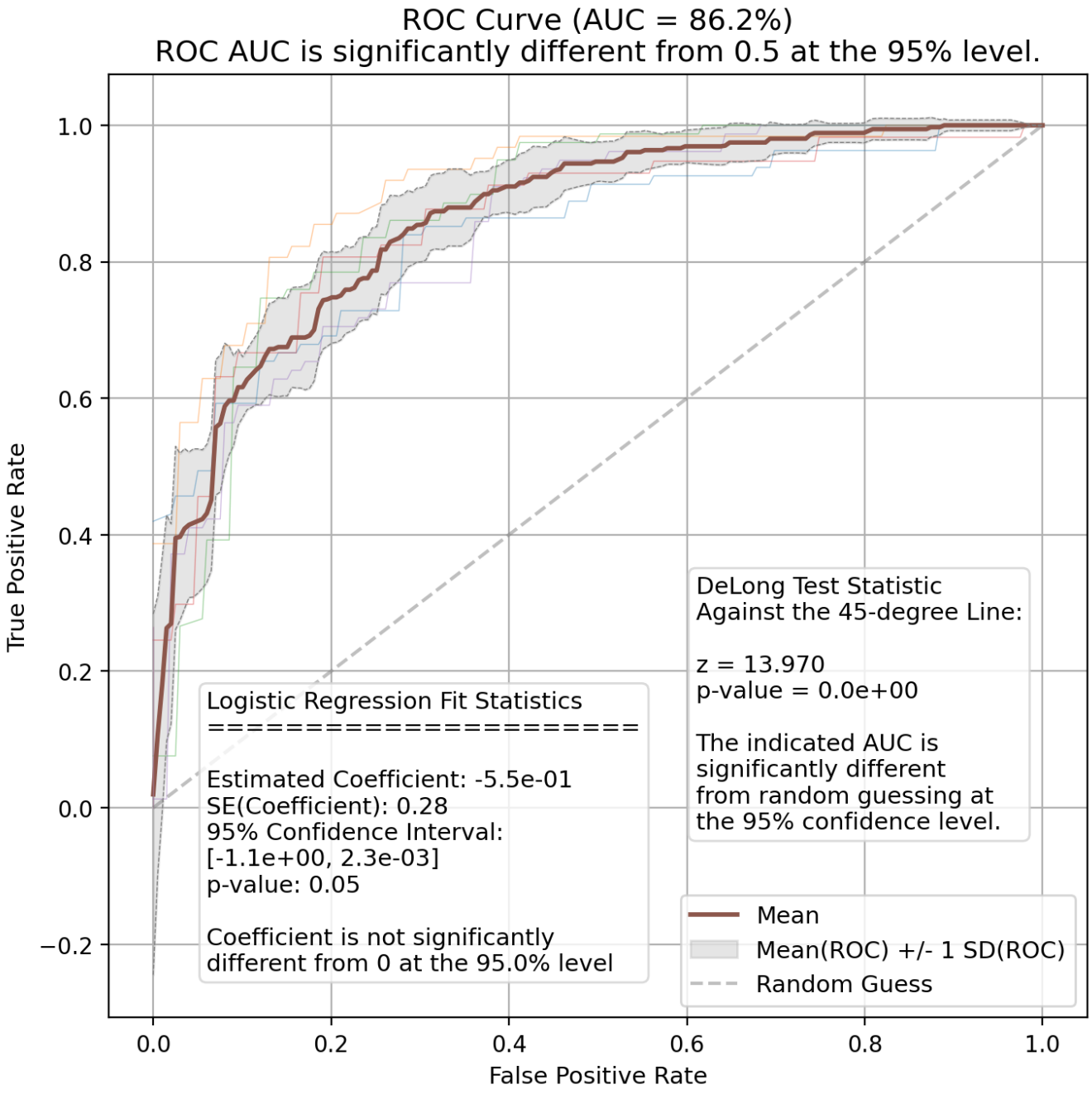
Worst Compactness - Empirical CDF Plot



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

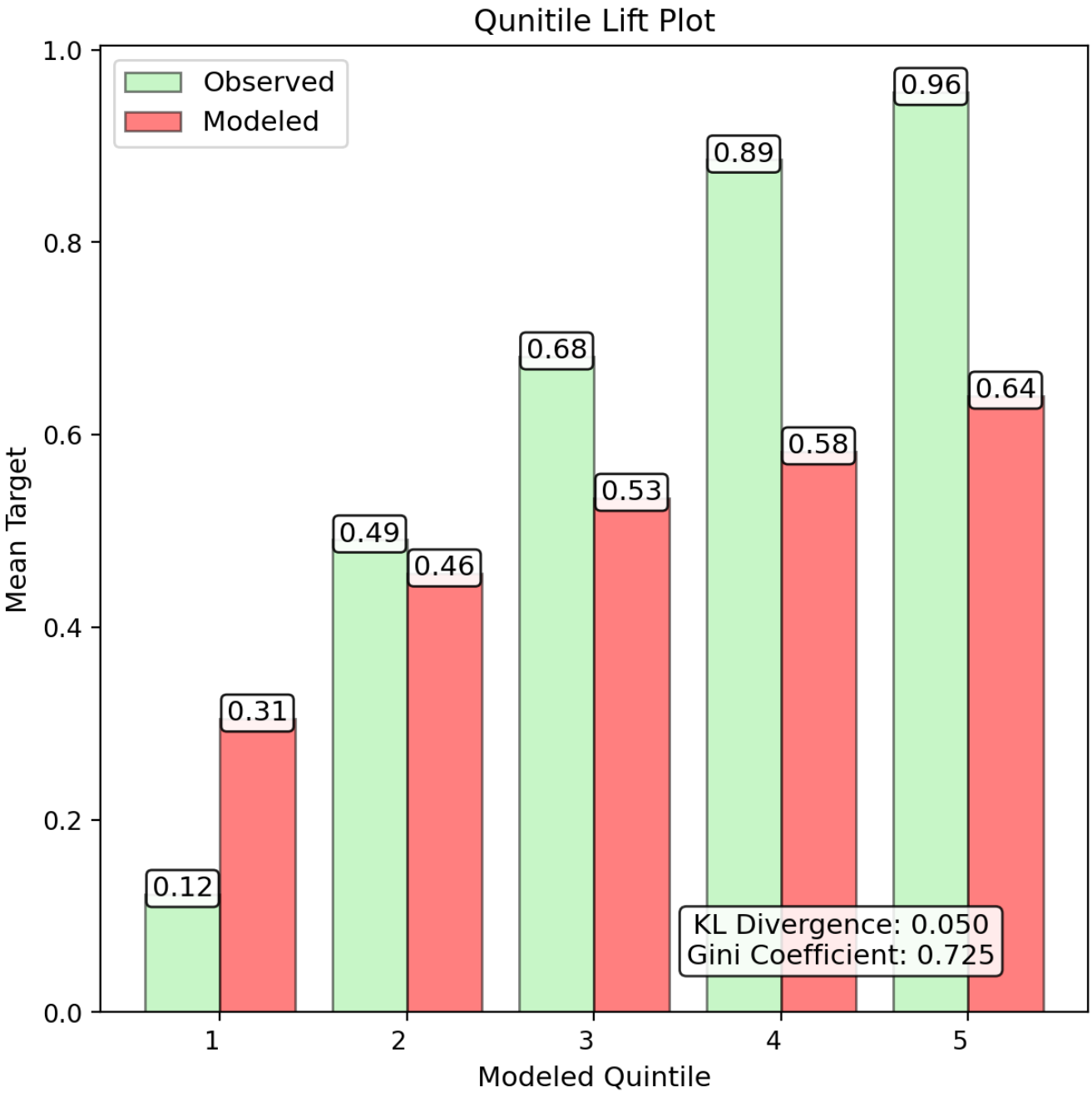


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst 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.

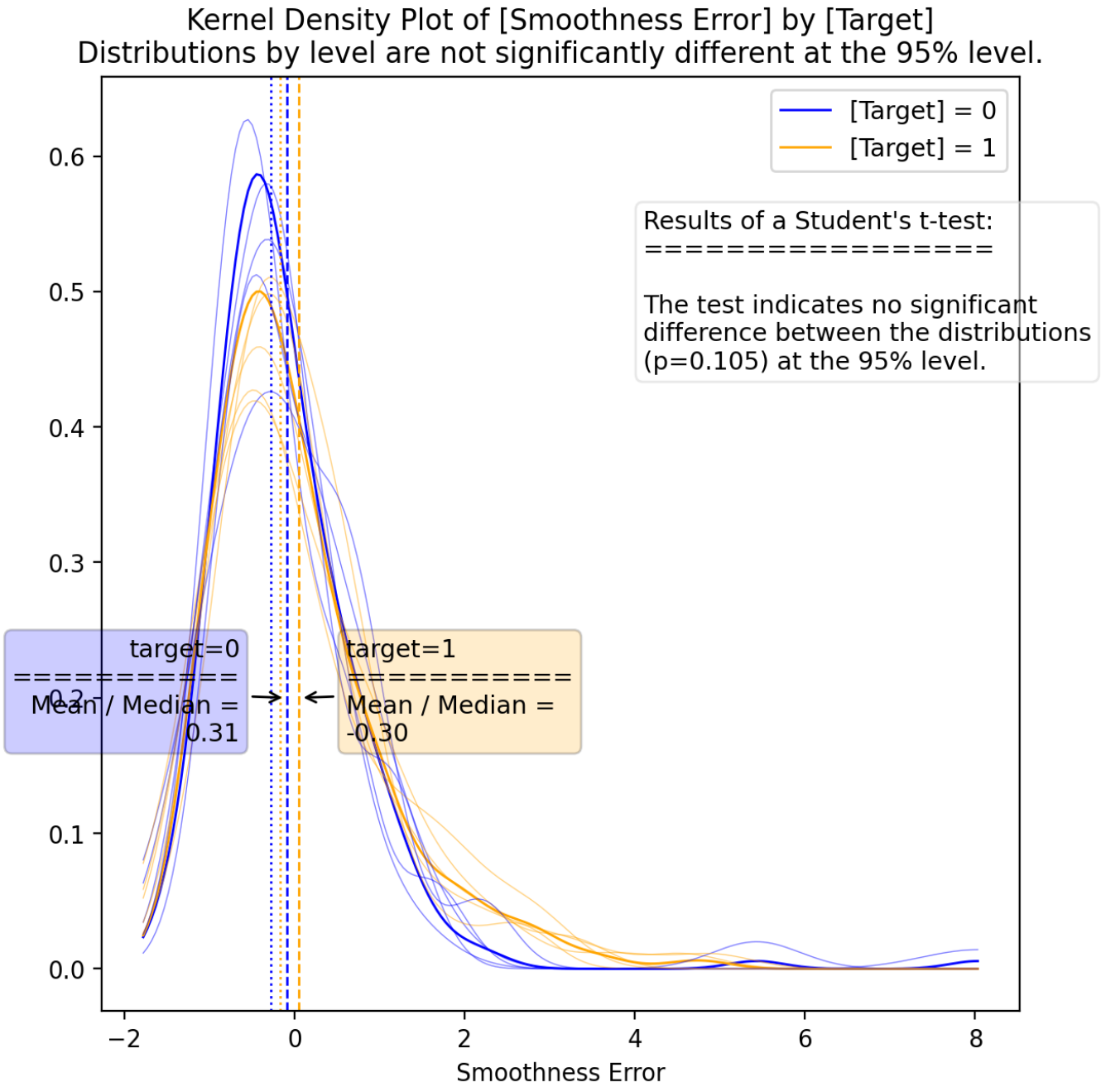
Univariate Report

Smoothness Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	0.180	0.124	0.091	0.154	0.104	69.880	0.036
Fitted p-Value	0.082	0.182	0.341	0.106	0.270	0.000	0.109
Fitted Std. Err.	1.0e-01	9.3e-02	9.6e-02	9.5e-02	9.4e-02	1.2e+01	4.2e-03
Conf. Int. Lower	-2.3e-02	-5.8e-02	-9.6e-02	-3.3e-02	-8.0e-02	4.7e+01	3.1e-02
Conf. Int. Upper	0.383	0.307	0.279	0.341	0.288	92.918	0.043
Train Accuracy	52.4%	49.8%	50.9%	52.2%	50.0%	51.1%	1.2%
Val Accuracy	46.8%	57.4%	52.2%	46.0%	53.4%	51.1%	4.8%
Train AUC	55.2%	53.0%	53.4%	55.8%	53.3%	54.3%	1.3%
Val AUC	50.5%	60.1%	58.3%	47.2%	56.1%	54.3%	5.4%
Train F1	53.1%	50.5%	50.9%	53.5%	51.2%	51.9%	1.4%
Test F1	48.4%	58.3%	55.7%	44.9%	52.0%	51.9%	5.4%
Train Precision	68.2%	66.7%	65.2%	70.9%	67.6%	67.9%	2.1%
Val Precision	66.0%	73.2%	79.1%	53.7%	67.3%	67.9%	9.5%
Train Recall	43.5%	40.7%	41.7%	43.0%	41.2%	42.0%	1.2%
Val Recall	38.3%	48.4%	43.0%	38.6%	42.3%	42.0%	4.1%
Train MCC	10.3%	5.9%	6.9%	11.5%	6.5%	8.5%	2.5%
Val MCC	1.0%	20.0%	15.7%	-5.6%	12.3%	8.5%	10.6%
Train Log-Loss	17.17	18.10	17.71	17.21	18.02	17.61	0.44
Val Log-Loss	19.18	15.35	17.22	19.46	16.78	17.61	1.72

Univariate Report

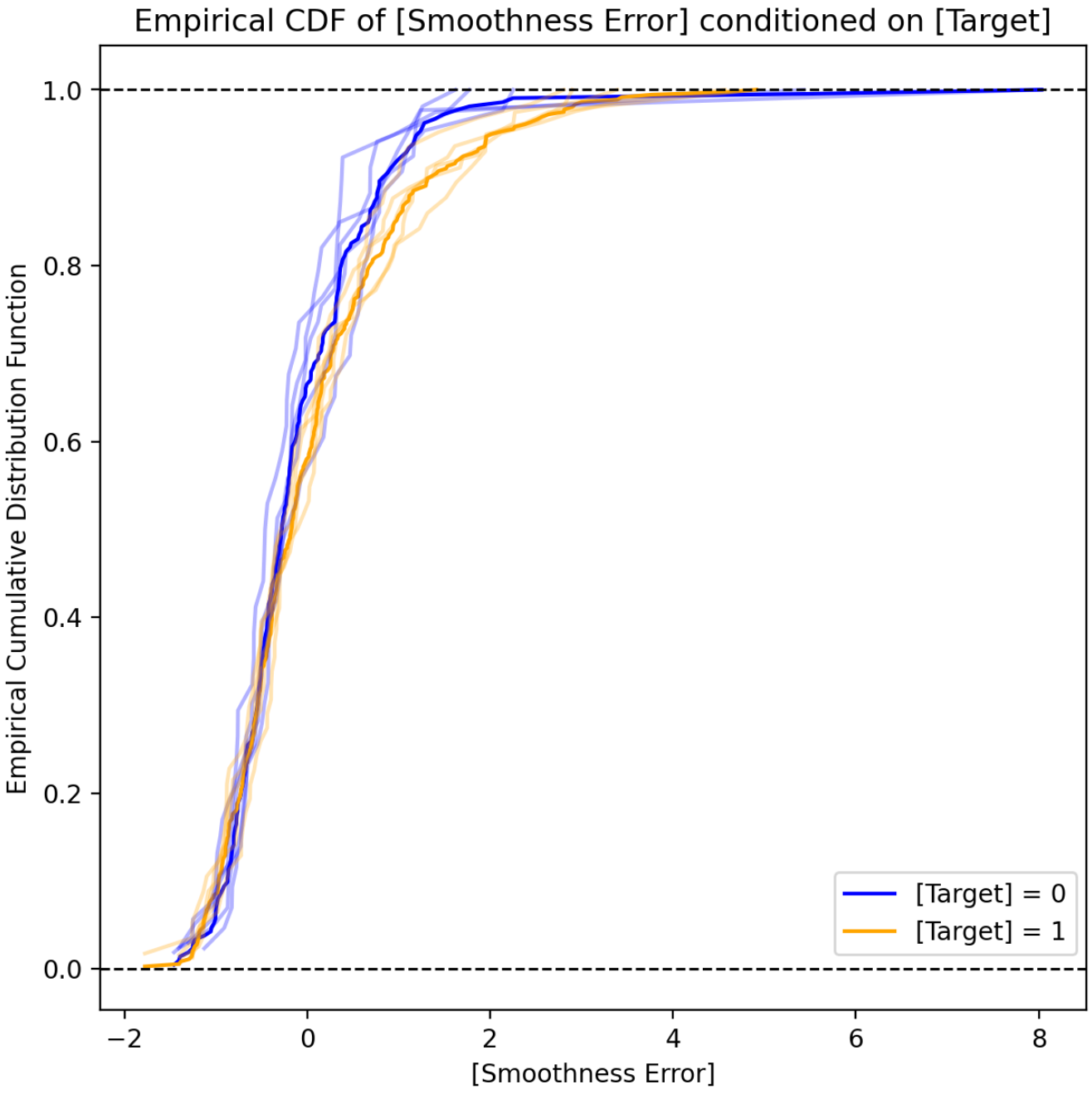
Smoothness Error - Kernel Density Plot



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

Univariate Report

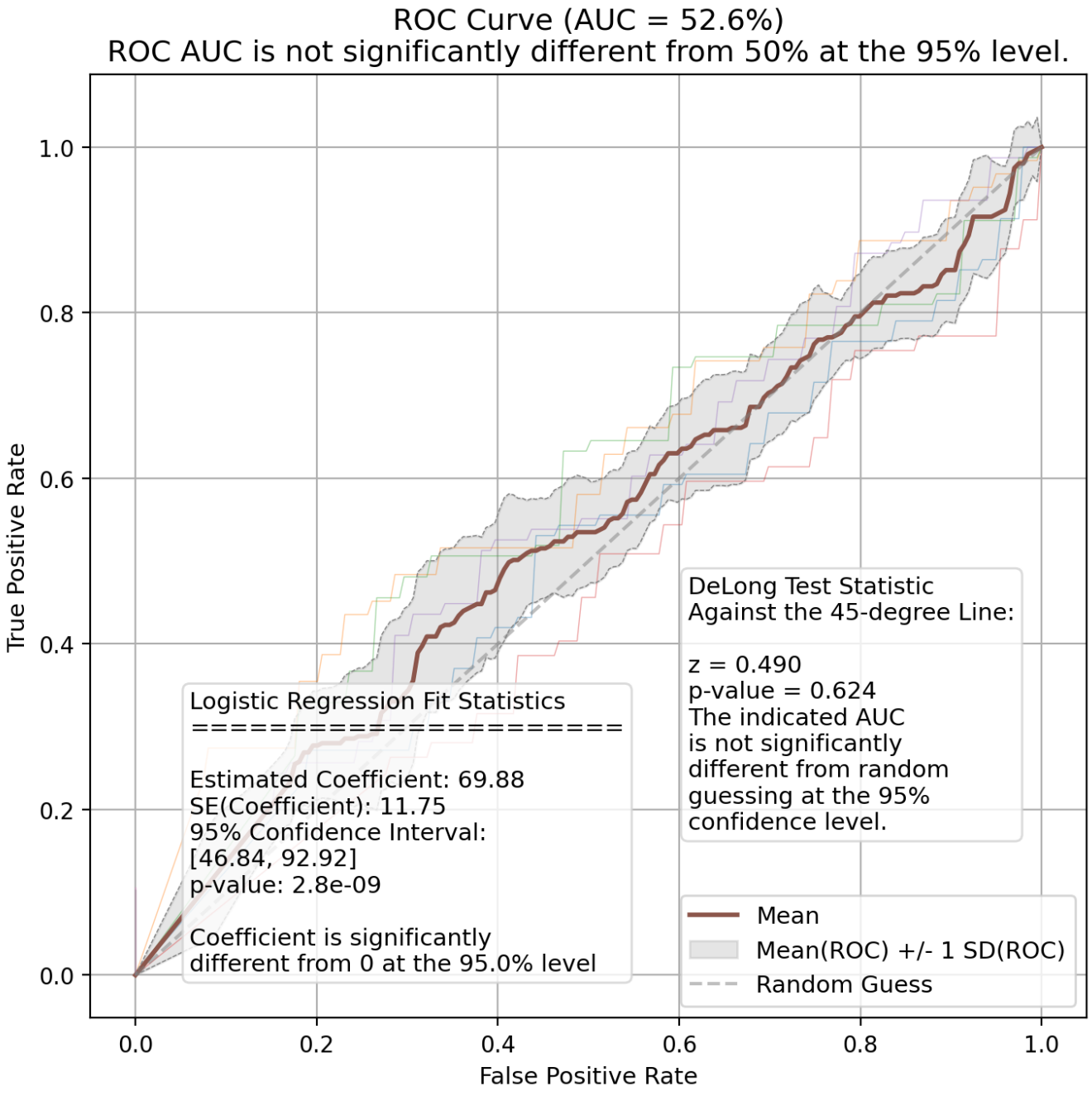
Smoothness Error - Empirical CDF Plot



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

Univariate Report

Smoothness Error - ROC Curve

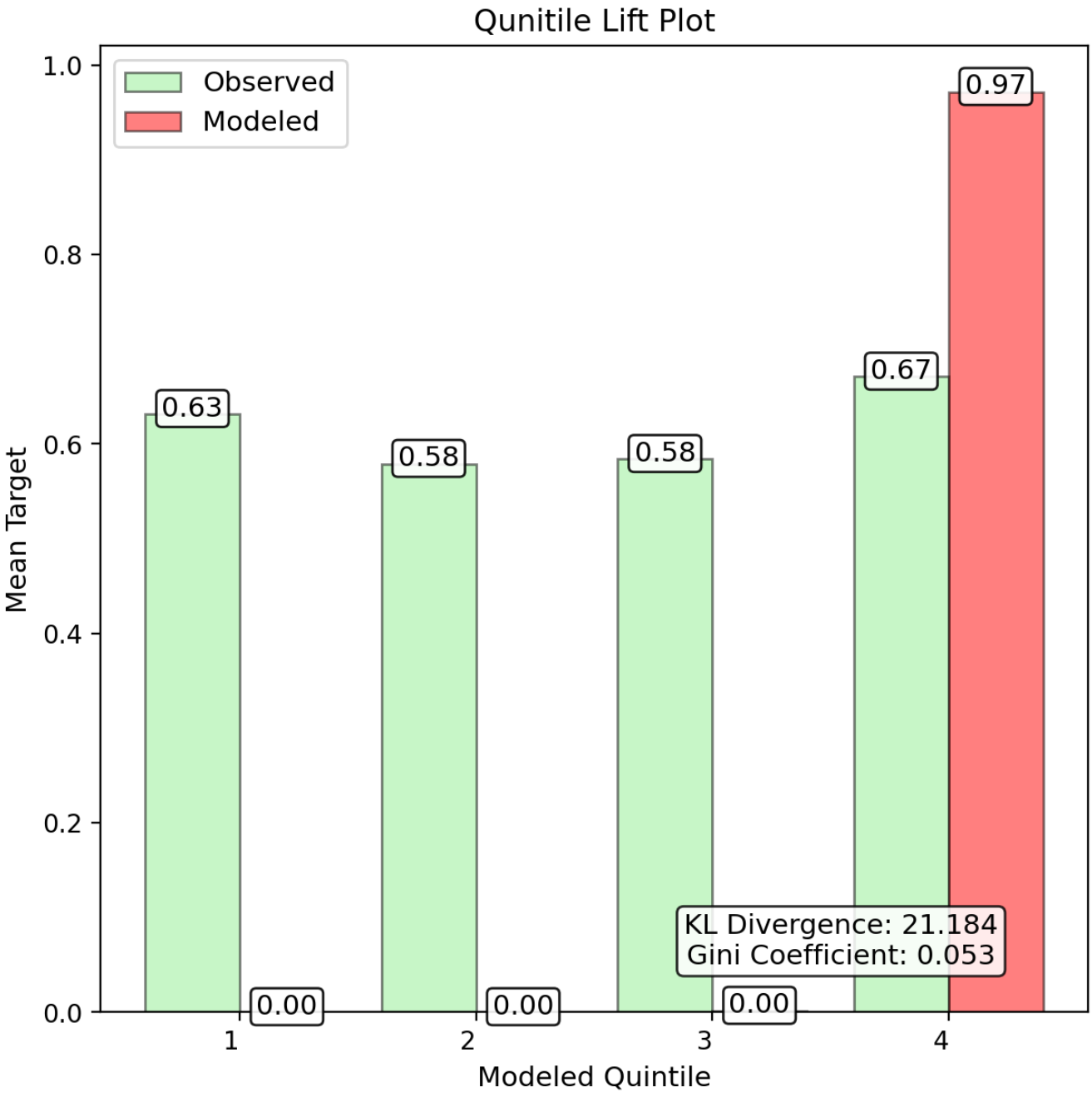


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Smoothness Error - Quintile Lift



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

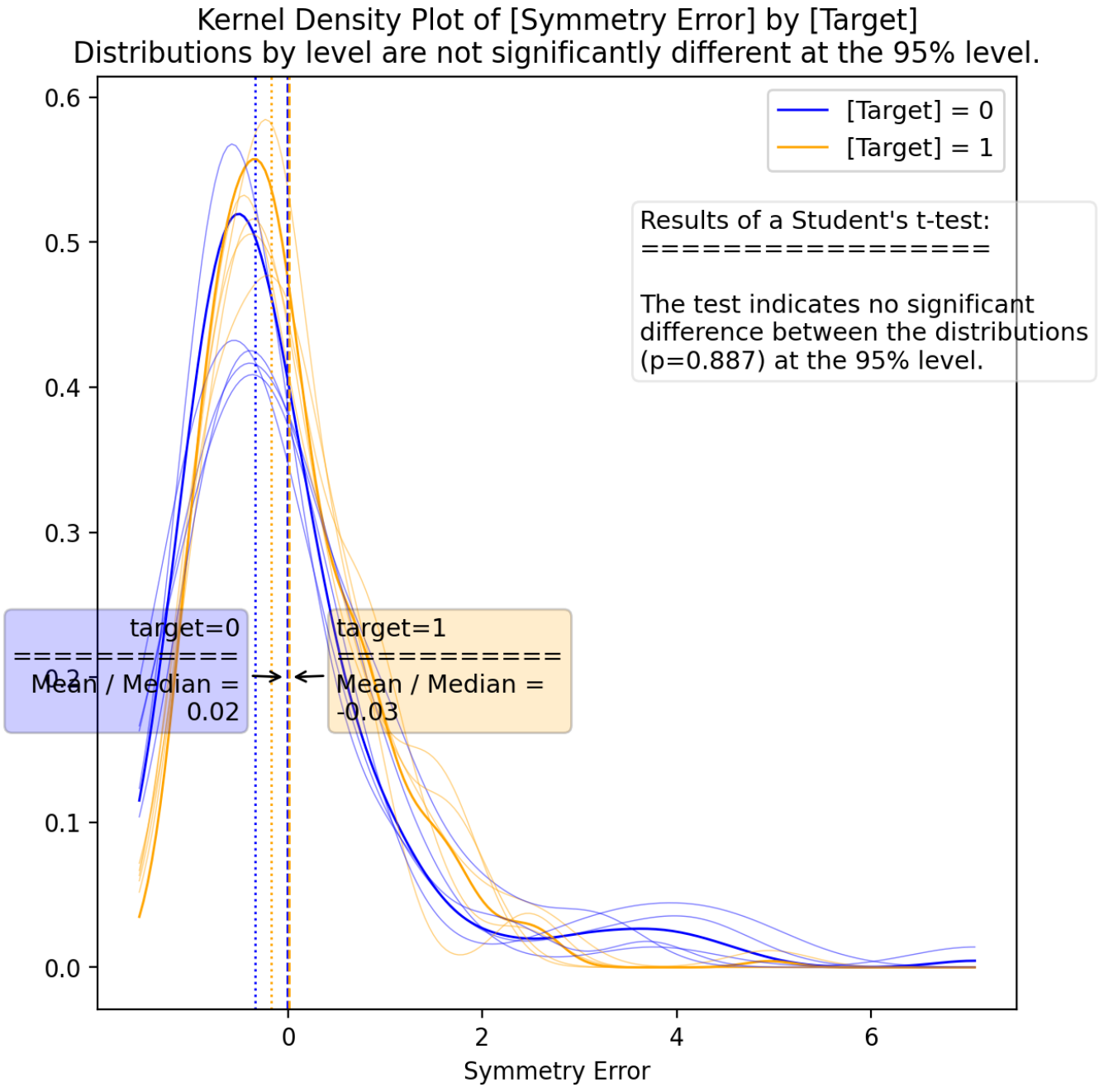
Univariate Report

Symmetry Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.6e-02	3.3e-02	3.4e-02	3.4e-02	-2.6e-02	2.2e+01	3.0e-02
Fitted p-Value	0.876	0.724	0.723	0.711	0.776	0.000	0.069
Fitted Std. Err.	1.0e-01	9.2e-02	9.5e-02	9.1e-02	9.2e-02	4.0e+00	3.5e-03
Conf. Int. Lower	-2.1e-01	-1.5e-01	-1.5e-01	-1.4e-01	-2.1e-01	1.4e+01	3.4e-02
Conf. Int. Upper	0.180	0.213	0.219	0.212	0.155	29.806	0.028
Train Accuracy	49.2%	50.2%	50.2%	49.5%	50.9%	50.1%	0.7%
Val Accuracy	53.2%	49.5%	50.4%	52.0%	46.6%	50.1%	2.6%
Train AUC	46.1%	53.7%	53.2%	53.5%	47.1%	53.6%	3.8%
Val AUC	49.5%	54.1%	56.2%	52.8%	43.3%	53.6%	5.0%
Train F1	59.1%	50.5%	49.2%	49.7%	61.3%	50.0%	5.7%
Test F1	63.3%	45.2%	54.1%	52.9%	57.3%	50.0%	6.6%
Train Precision	59.1%	67.6%	65.1%	68.4%	61.6%	67.3%	4.0%
Val Precision	64.9%	67.7%	76.7%	60.0%	54.7%	67.3%	8.3%
Train Recall	59.1%	40.3%	39.6%	39.0%	60.9%	39.8%	11.2%
Val Recall	61.7%	33.9%	41.8%	47.4%	60.3%	39.8%	12.0%
Train MCC	-7.8%	7.4%	6.5%	7.0%	-5.7%	7.2%	7.6%
Val MCC	-1.0%	8.7%	11.7%	5.5%	-13.8%	7.2%	10.1%
Train Log-Loss	18.31	17.94	17.94	18.21	17.69	17.99	0.24
Val Log-Loss	16.86	18.20	17.86	17.30	19.26	17.99	0.92

Univariate Report

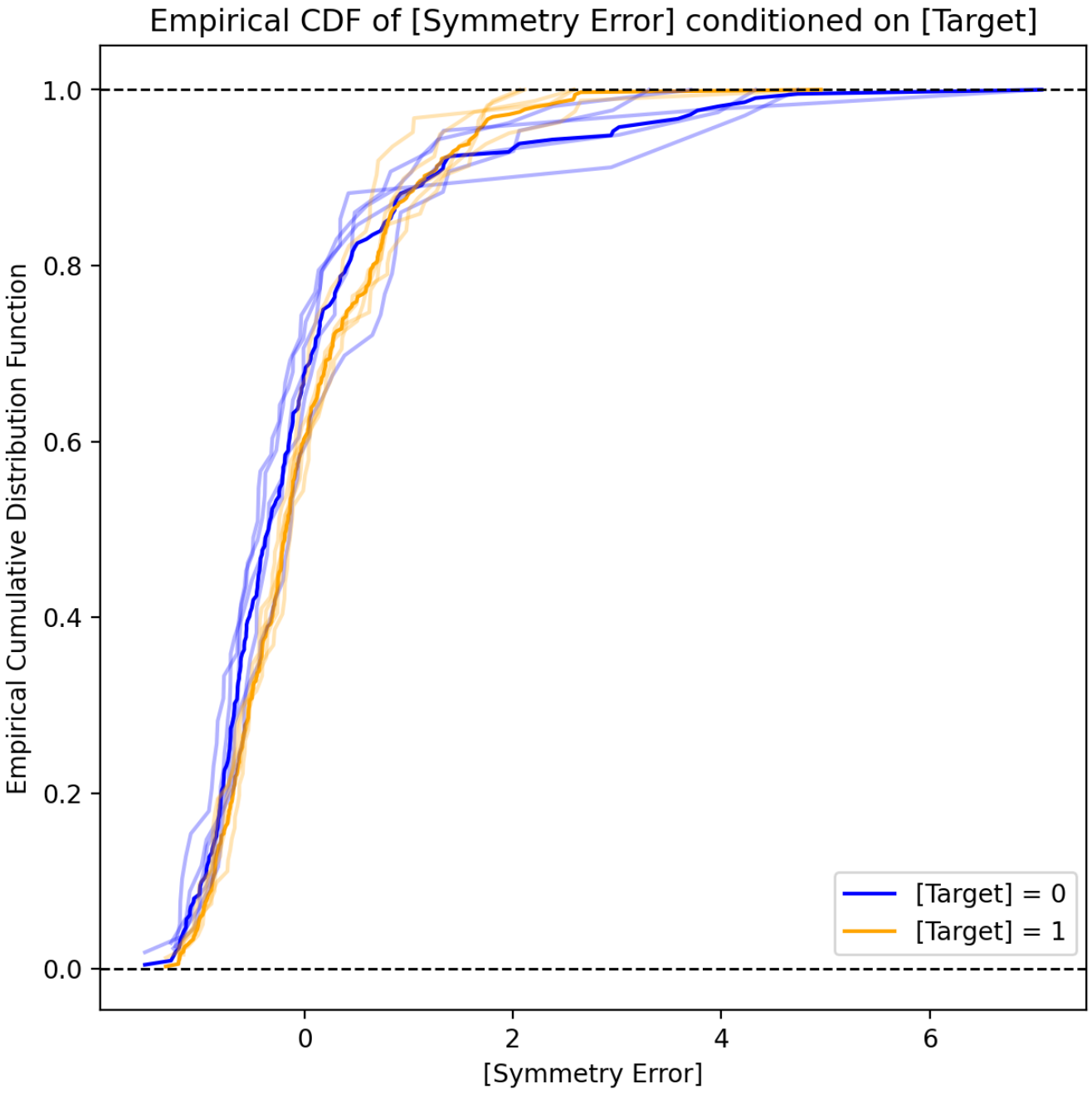
Symmetry Error - Kernel Density Plot



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

Univariate Report

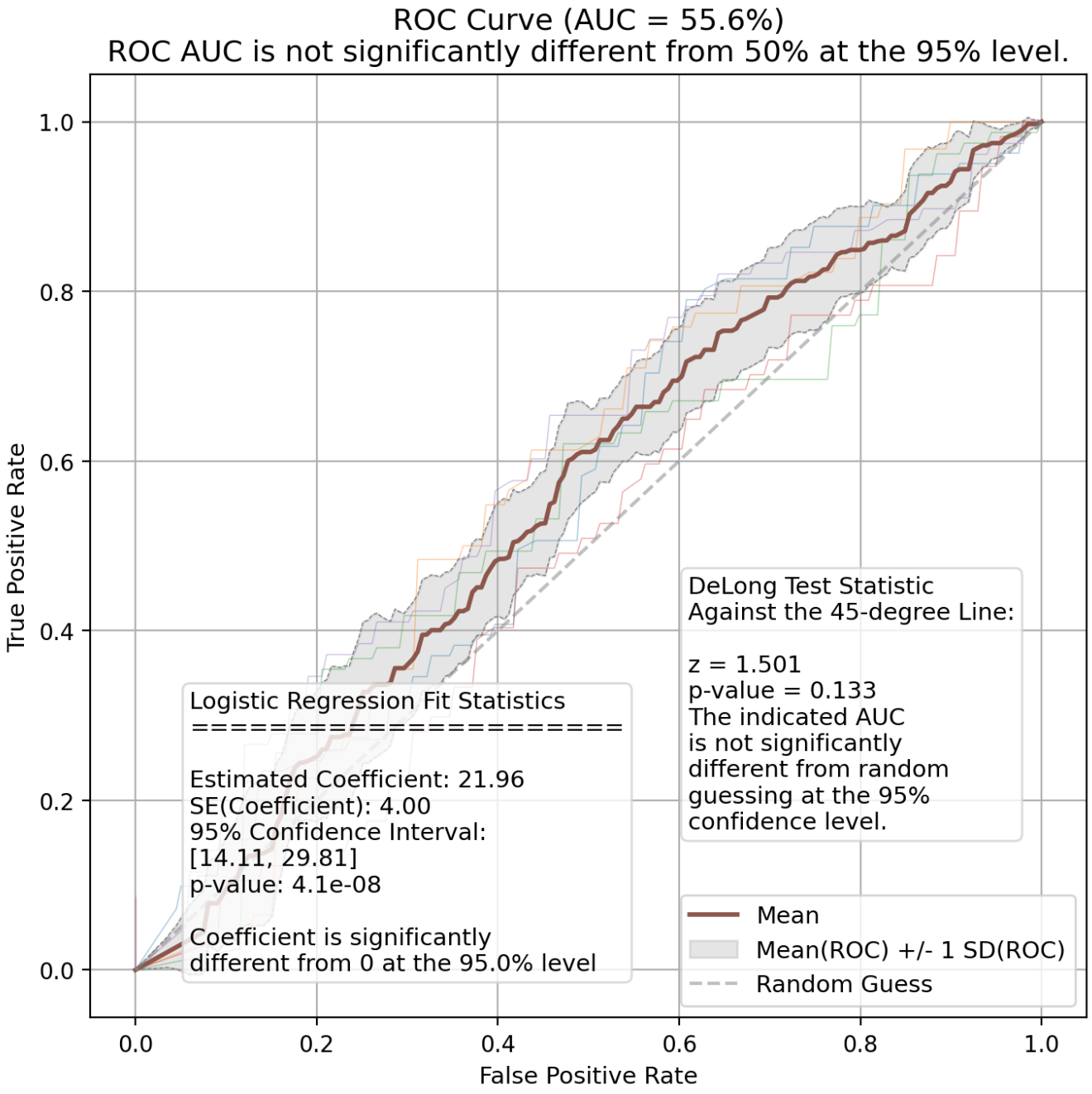
Symmetry Error - Empirical CDF Plot



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

Univariate Report

Symmetry Error - ROC Curve

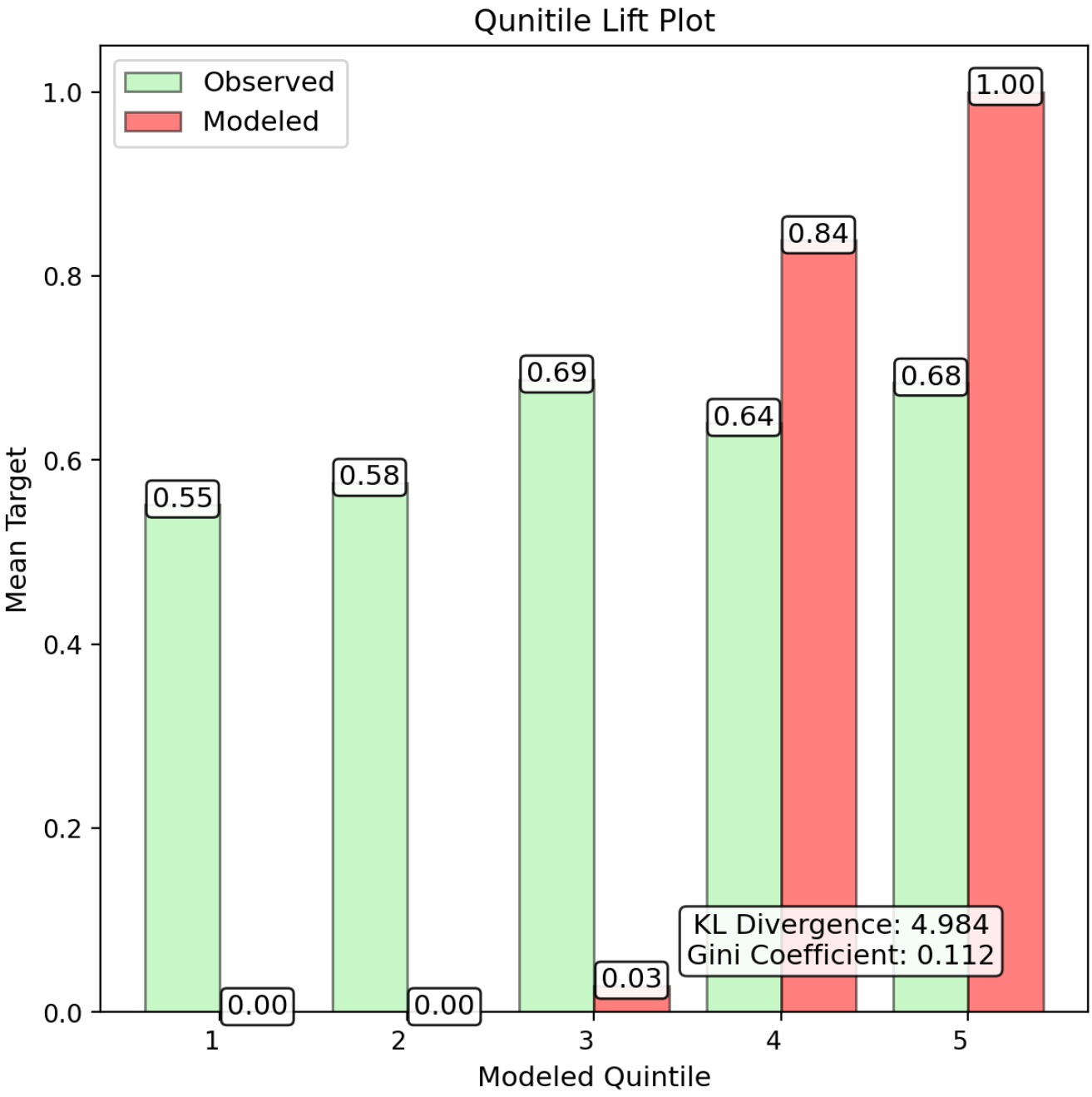


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Symmetry Error - Quintile Lift



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

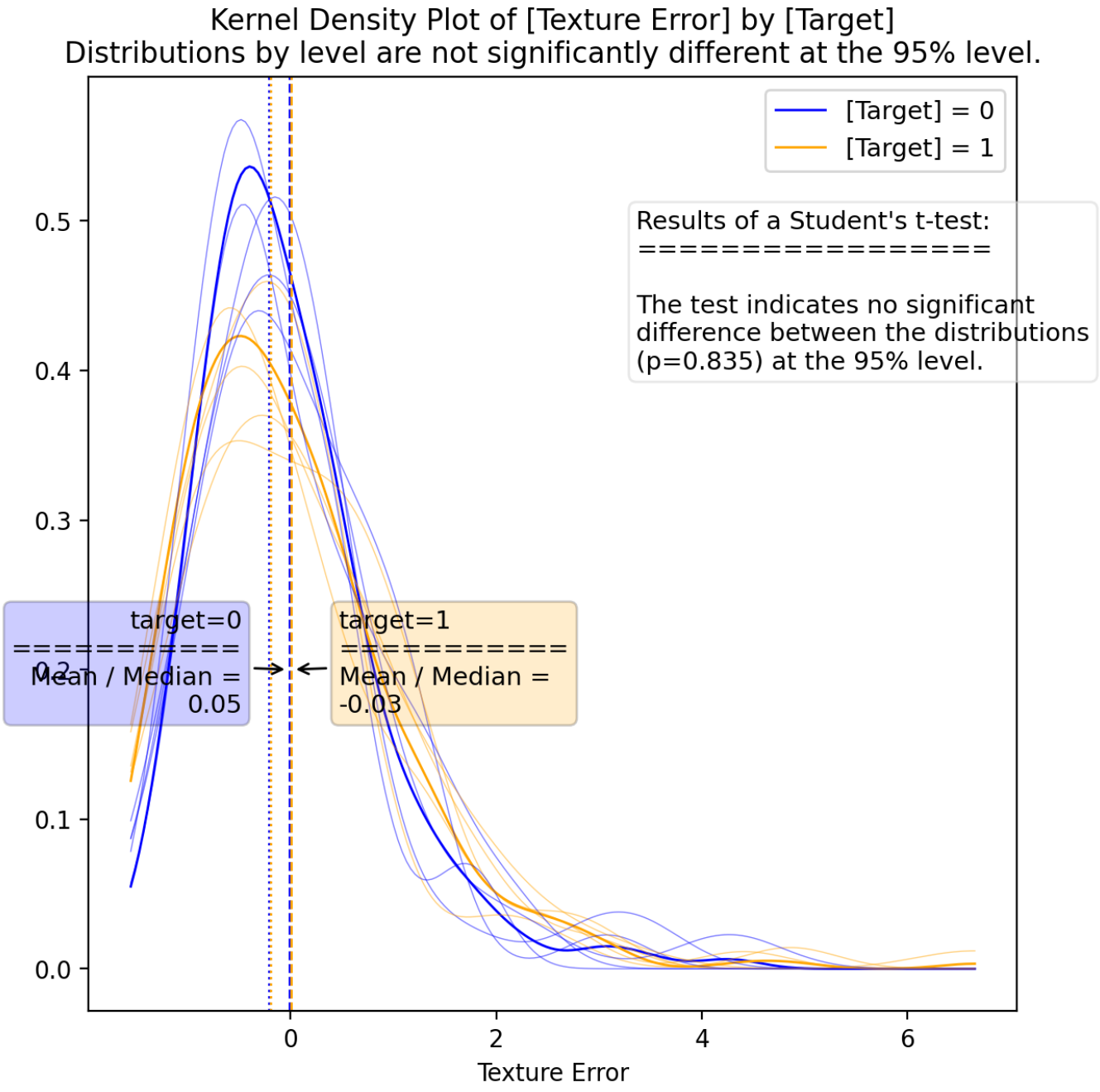
Univariate Report

Texture Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	1.9e-03	5.3e-02	3.3e-02	-2.1e-02	9.6e-03	3.6e-01	2.8e-02
Fitted p-Value	0.983	0.571	0.722	0.827	0.919	0.000	0.163
Fitted Std. Err.	9.4e-02	9.3e-02	9.4e-02	9.5e-02	9.4e-02	6.6e-02	9.4e-04
Conf. Int. Lower	-1.8e-01	-1.3e-01	-1.5e-01	-2.1e-01	-1.7e-01	2.3e-01	3.0e-02
Conf. Int. Upper	0.185	0.234	0.217	0.166	0.194	0.489	0.027
Train Accuracy	49.2%	48.9%	49.8%	51.4%	48.2%	49.2%	1.2%
Val Accuracy	50.0%	46.5%	46.0%	47.0%	48.1%	49.2%	1.6%
Train AUC	51.5%	51.1%	51.5%	49.3%	50.5%	51.5%	0.9%
Val AUC	50.3%	49.8%	48.8%	46.4%	51.0%	51.5%	1.8%
Train F1	50.7%	51.3%	51.4%	59.9%	50.8%	51.3%	4.0%
Test F1	56.3%	44.9%	52.0%	52.3%	45.2%	51.3%	5.0%
Train Precision	63.7%	64.3%	62.7%	63.4%	64.3%	64.4%	0.7%
Val Precision	65.6%	61.1%	68.8%	53.7%	60.9%	64.4%	5.7%
Train Recall	42.0%	42.7%	43.5%	56.7%	41.9%	42.6%	6.3%
Val Recall	49.4%	35.5%	41.8%	50.9%	35.9%	42.6%	7.3%
Train MCC	2.9%	2.2%	3.0%	-1.3%	1.0%	2.9%	1.8%
Val MCC	0.5%	-0.4%	-2.2%	-7.2%	2.0%	2.9%	3.6%
Train Log-Loss	18.31	18.41	18.10	17.52	18.68	18.31	0.43
Val Log-Loss	18.02	19.27	19.46	19.10	18.71	18.31	0.57

Univariate Report

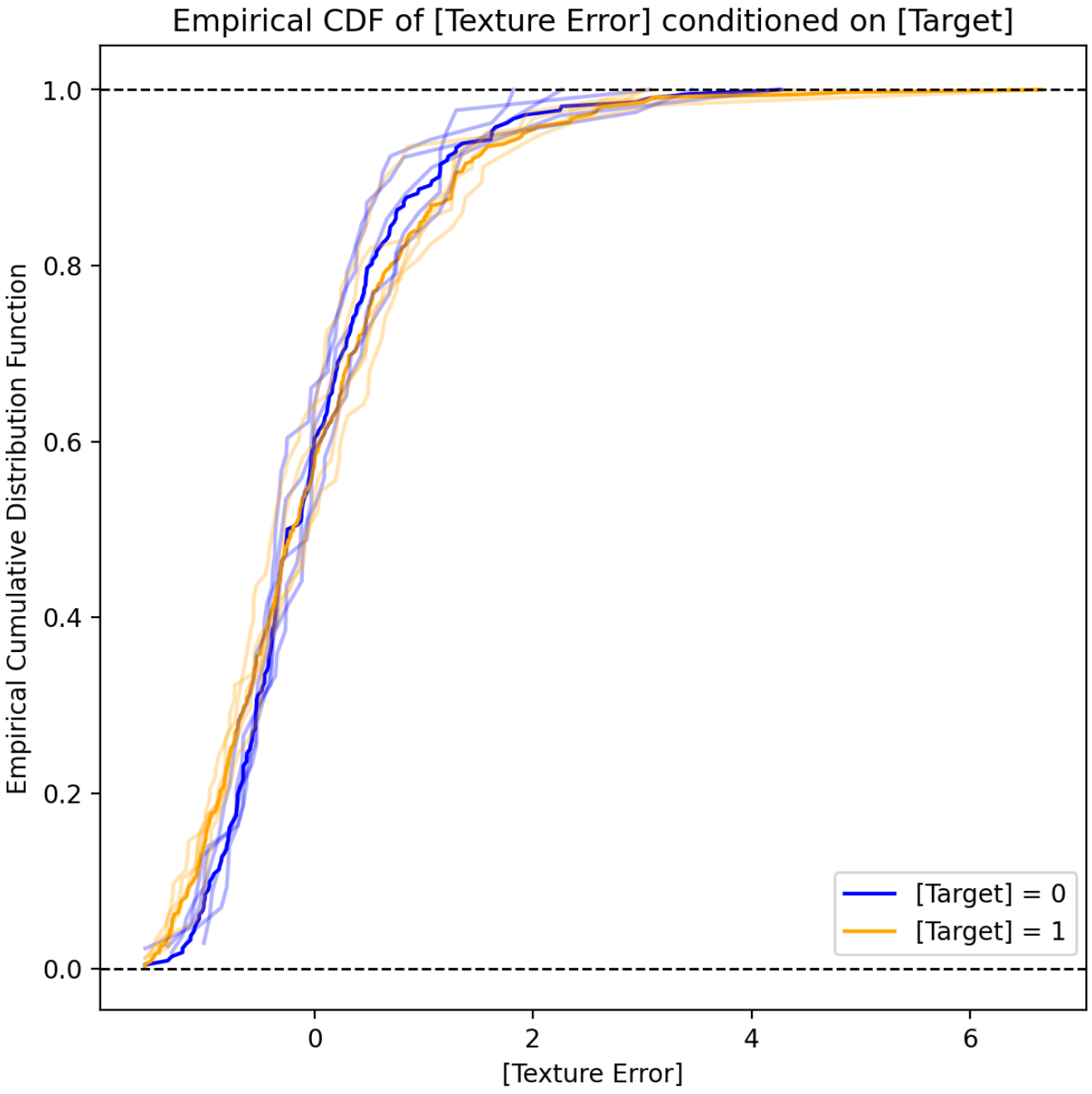
Texture Error - Kernel Density Plot



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

Univariate Report

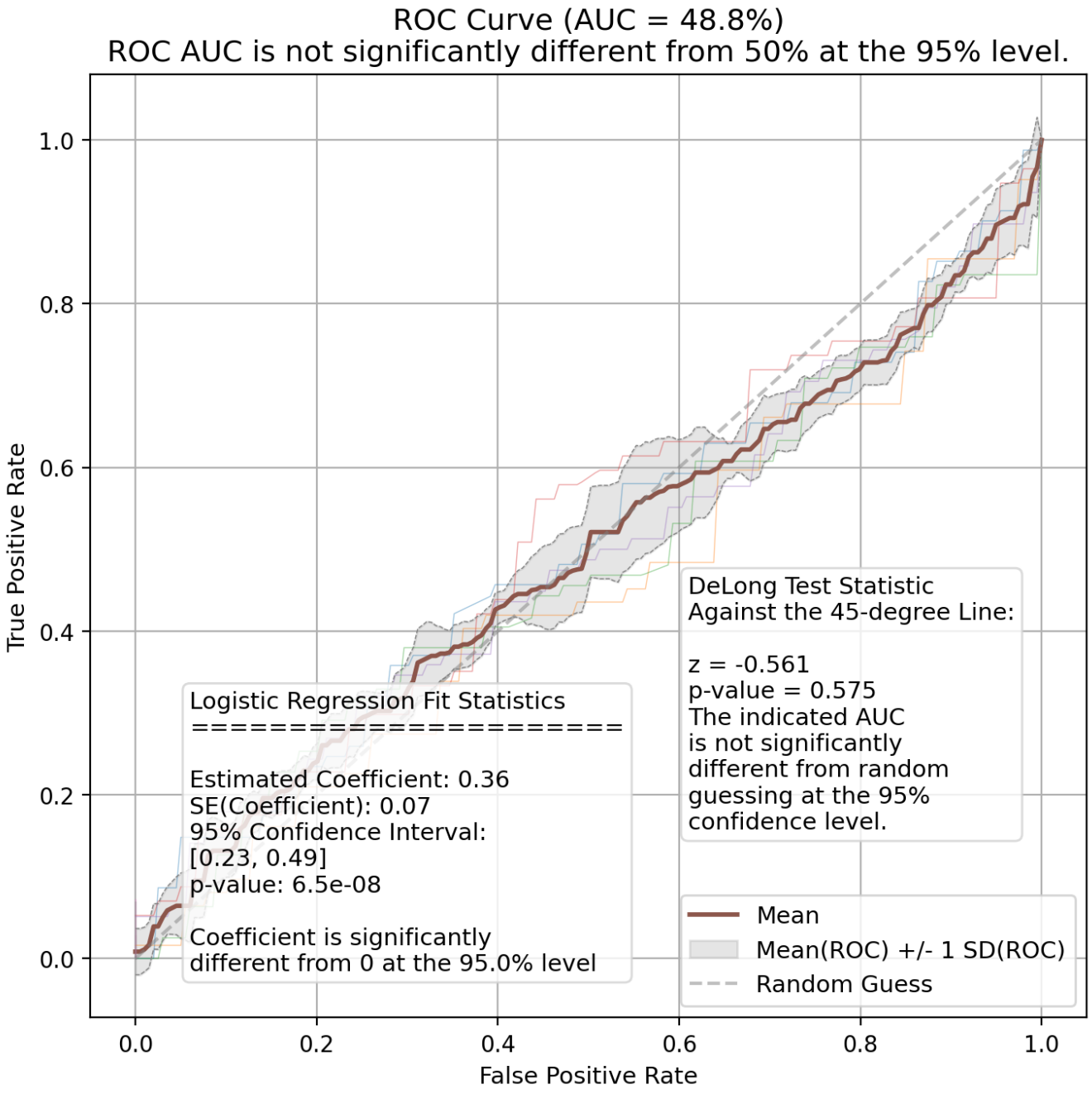
Texture Error - Empirical CDF Plot



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

Univariate Report

Texture Error - ROC Curve

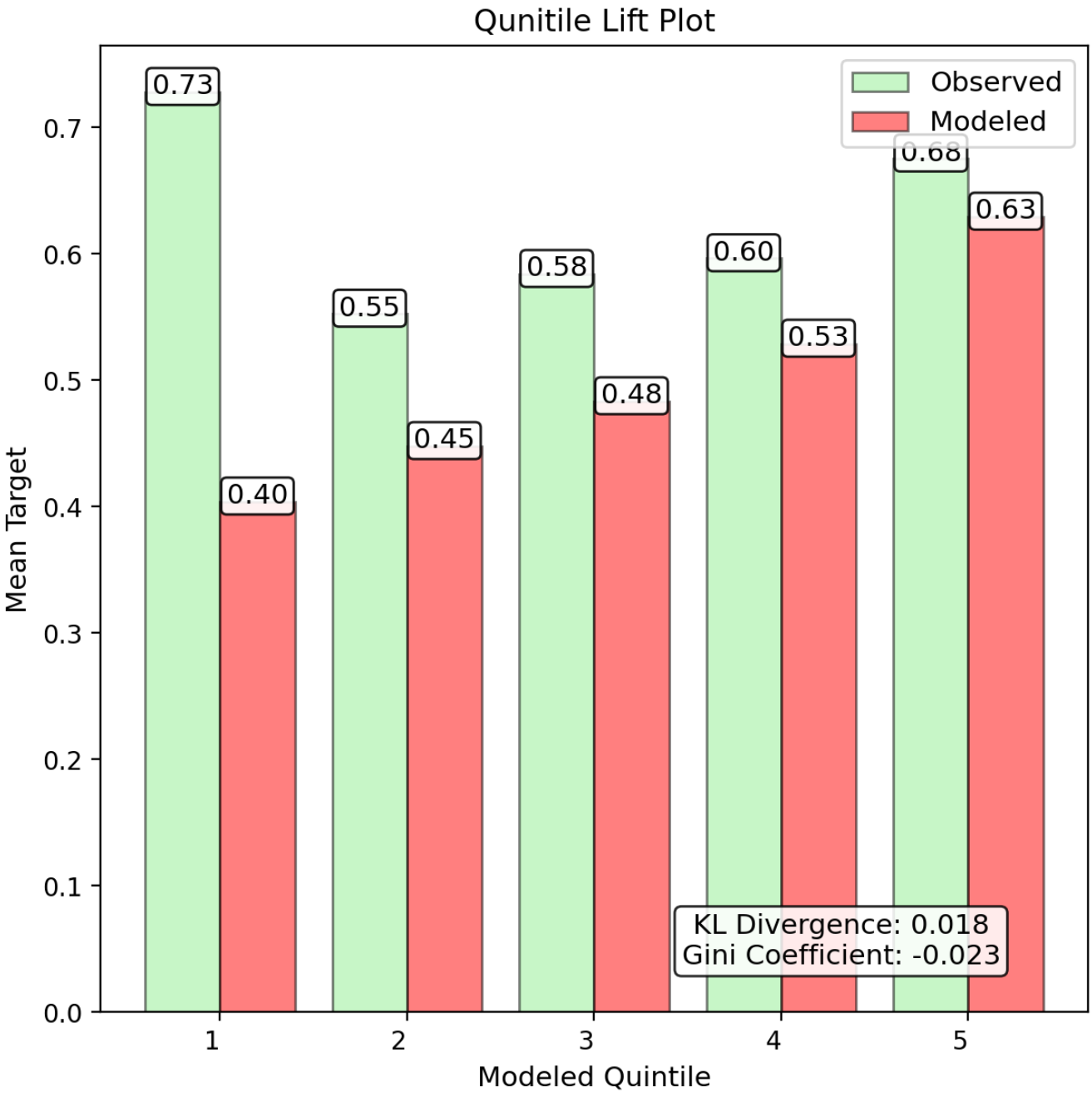


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Texture Error - Quintile Lift



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

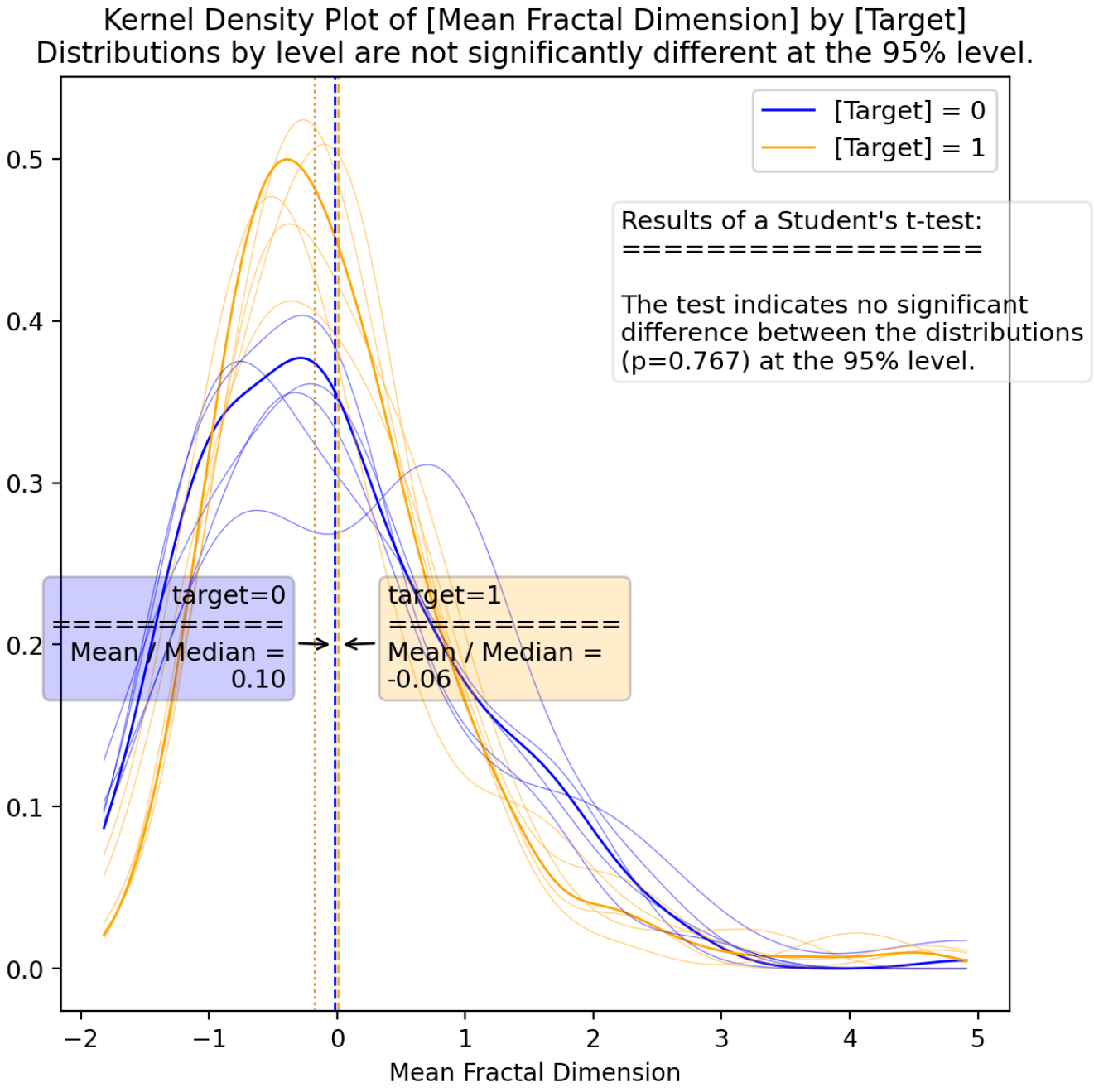
Univariate Report

Mean Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.5e-02	6.4e-02	4.4e-02	3.1e-02	-3.9e-03	8.1e+00	3.3e-02
Fitted p-Value	0.871	0.480	0.649	0.741	0.967	0.000	0.190
Fitted Std. Err.	9.4e-02	9.1e-02	9.8e-02	9.3e-02	9.4e-02	1.4e+00	2.4e-03
Conf. Int. Lower	-2.0e-01	-1.1e-01	-1.5e-01	-1.5e-01	-1.9e-01	5.4e+00	3.4e-02
Conf. Int. Upper	0.169	0.242	0.235	0.214	0.180	10.768	0.033
Train Accuracy	52.4%	48.9%	48.0%	48.8%	52.1%	48.3%	2.0%
Val Accuracy	50.0%	45.5%	51.3%	45.0%	51.1%	48.3%	3.1%
Train AUC	50.4%	51.2%	49.6%	51.0%	49.8%	50.3%	0.7%
Val AUC	48.6%	45.7%	54.3%	47.2%	49.0%	50.3%	3.3%
Train F1	60.4%	51.1%	49.9%	52.0%	60.7%	50.8%	5.3%
Test F1	58.1%	50.5%	57.4%	39.6%	59.5%	50.8%	8.3%
Train Precision	62.3%	64.4%	60.5%	65.0%	63.5%	63.1%	1.8%
Val Precision	64.2%	57.1%	74.0%	52.9%	58.8%	63.1%	8.1%
Train Recall	58.7%	42.4%	42.4%	43.3%	58.1%	42.6%	8.6%
Val Recall	53.1%	45.2%	46.8%	31.6%	60.3%	42.6%	10.6%
Train MCC	0.7%	2.4%	-0.8%	1.9%	-0.4%	0.6%	1.4%
Val MCC	-2.6%	-8.5%	7.9%	-5.9%	-2.0%	0.6%	6.2%
Train Log-Loss	17.17	18.41	18.73	18.44	17.28	18.62	0.73
Val Log-Loss	18.02	19.63	17.54	19.82	17.61	18.62	1.11

Univariate Report

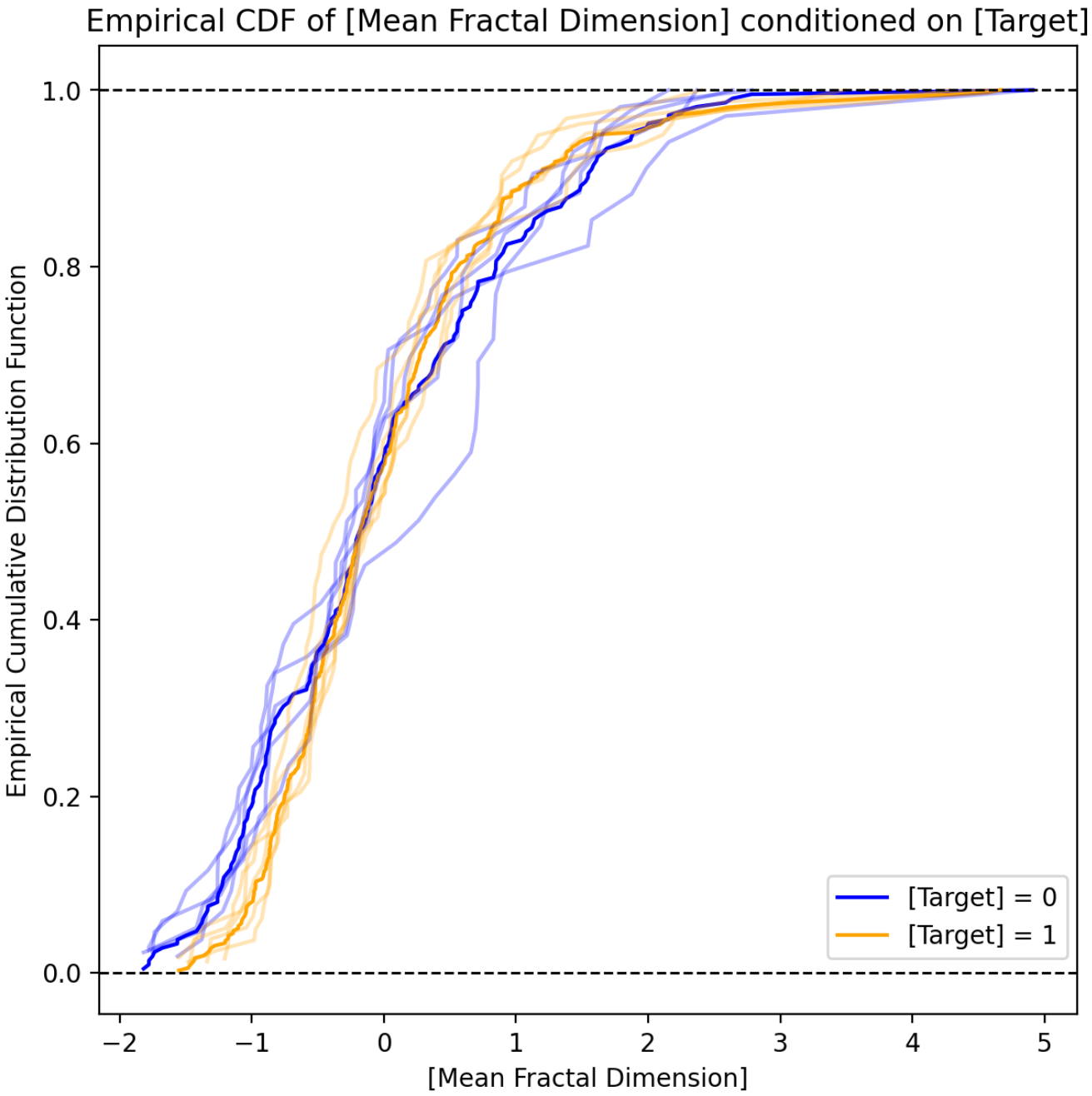
Mean Fractal Dimension - Kernel Density Plot



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

Univariate Report

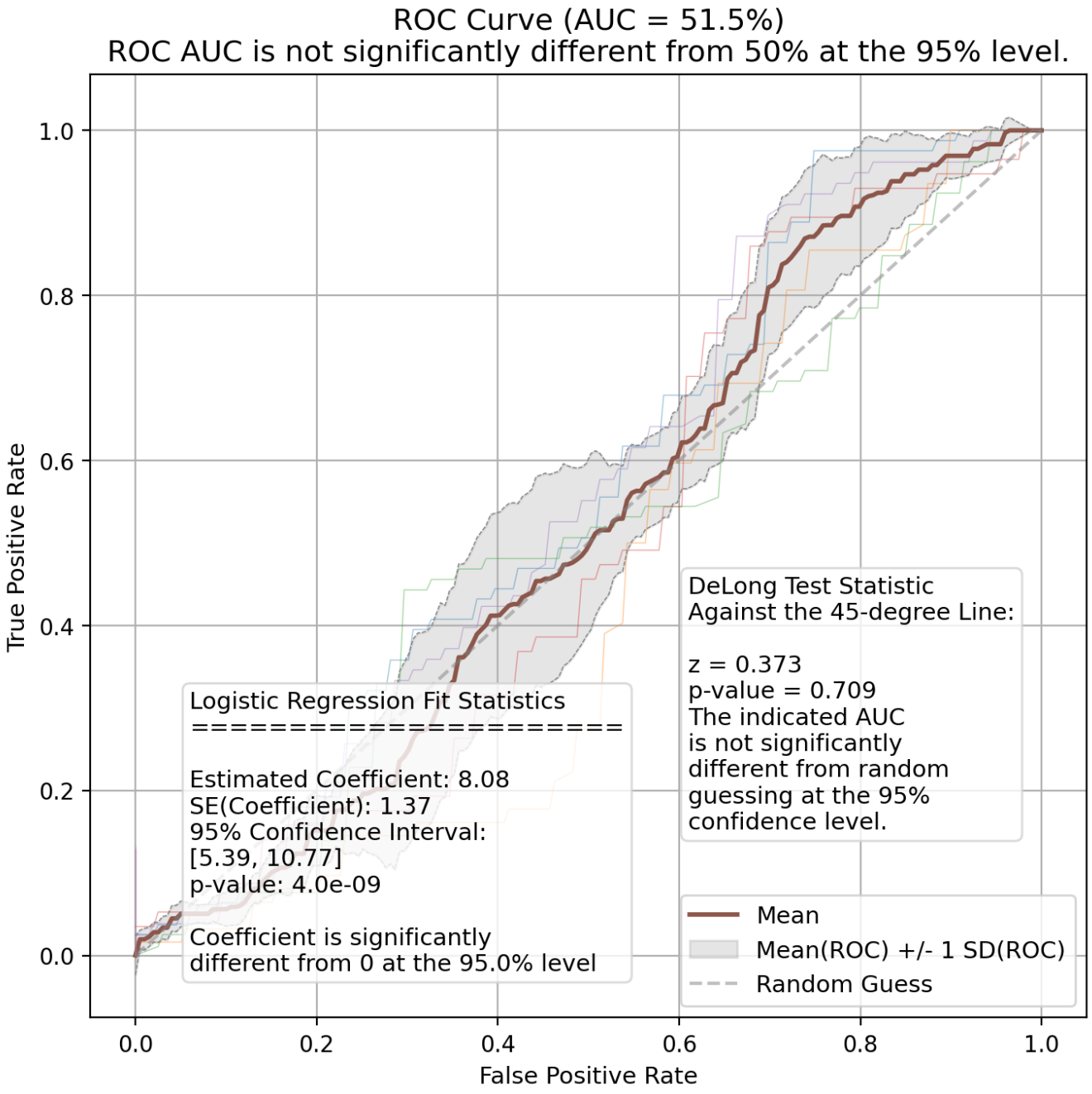
Mean Fractal Dimension - Empirical CDF Plot



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

Univariate Report

Mean Fractal Dimension - ROC Curve

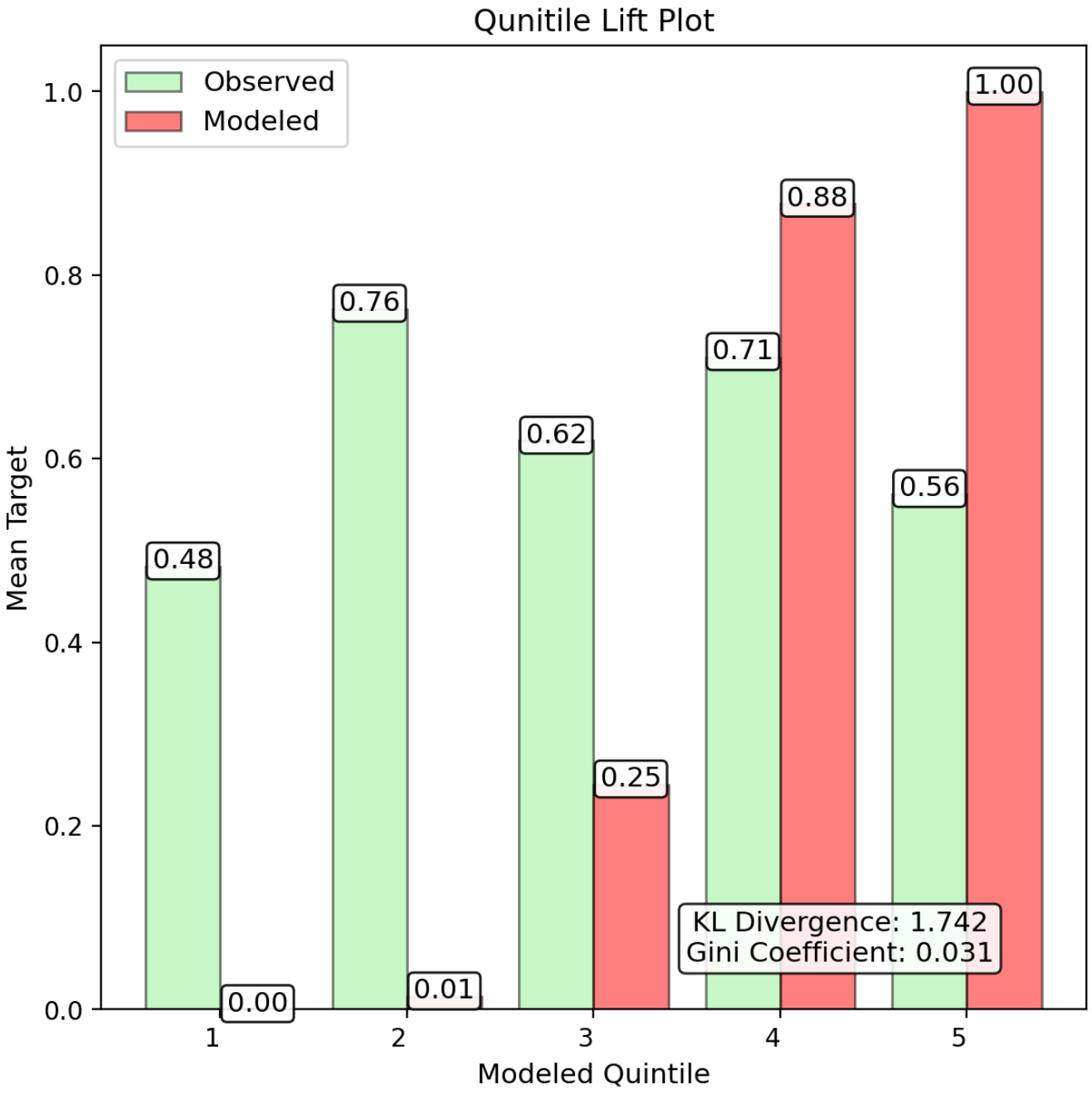


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Mean Fractal Dimension - Quintile Lift



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

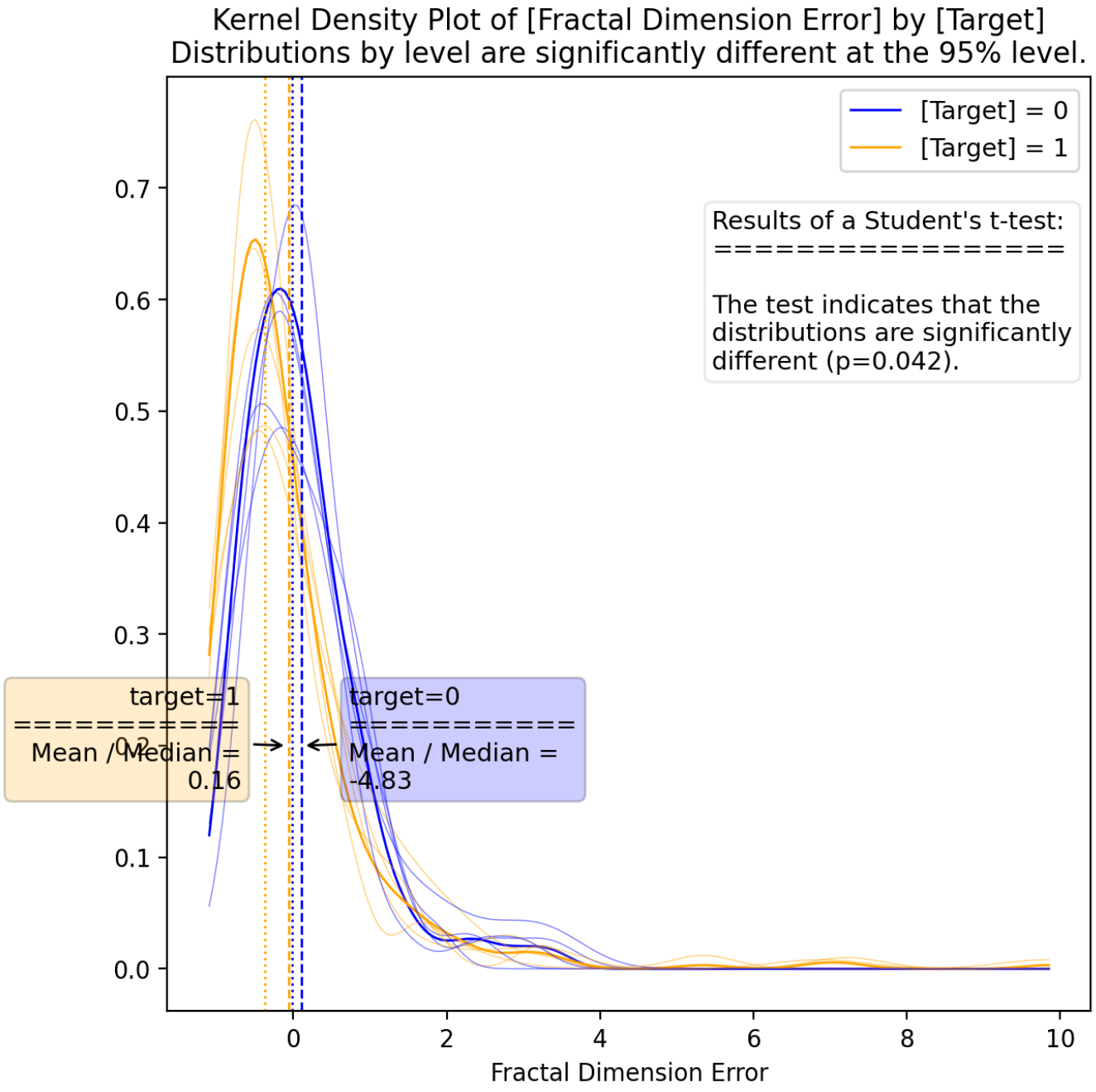
Univariate Report

Fractal Dimension Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.1e-01	-1.1e-01	-1.3e-01	-1.1e-01	-2.5e-01	7.5e+01	6.5e-02
Fitted p-Value	0.051	0.239	0.185	0.230	0.028	0.000	0.100
Fitted Std. Err.	0.110	0.092	0.100	0.092	0.113	19.625	0.010
Conf. Int. Lower	-4.3e-01	-2.9e-01	-3.3e-01	-2.9e-01	-4.7e-01	3.6e+01	8.4e-02
Conf. Int. Upper	1.3e-03	7.2e-02	6.3e-02	7.0e-02	-2.7e-02	1.1e+02	4.6e-02
Train Accuracy	64.0%	60.9%	62.5%	60.6%	61.6%	38.3%	1.4%
Val Accuracy	57.3%	66.3%	58.4%	67.0%	61.1%	38.3%	4.5%
Train AUC	61.7%	57.7%	59.9%	57.2%	58.8%	41.2%	1.8%
Val AUC	55.3%	63.1%	54.3%	65.6%	58.2%	41.2%	4.9%
Train F1	71.1%	69.2%	69.9%	69.2%	69.7%	37.9%	0.8%
Test F1	65.4%	73.8%	68.5%	72.3%	69.1%	37.9%	3.3%
Train Precision	70.9%	68.7%	68.4%	69.1%	70.2%	51.4%	1.0%
Val Precision	69.4%	70.6%	72.9%	69.4%	65.5%	51.4%	2.7%
Train Recall	71.4%	69.8%	71.6%	69.3%	69.2%	30.0%	1.1%
Val Recall	61.7%	77.4%	64.6%	75.4%	73.1%	30.0%	6.9%
Train MCC	23.5%	15.6%	20.2%	14.3%	17.5%	-17.7%	3.7%
Val MCC	10.2%	27.1%	8.2%	31.9%	17.1%	-17.7%	10.4%
Train Log-Loss	12.96	14.09	13.52	14.22	13.82	22.23	0.50
Val Log-Loss	15.41	12.13	14.99	11.89	14.03	22.23	1.61

Univariate Report

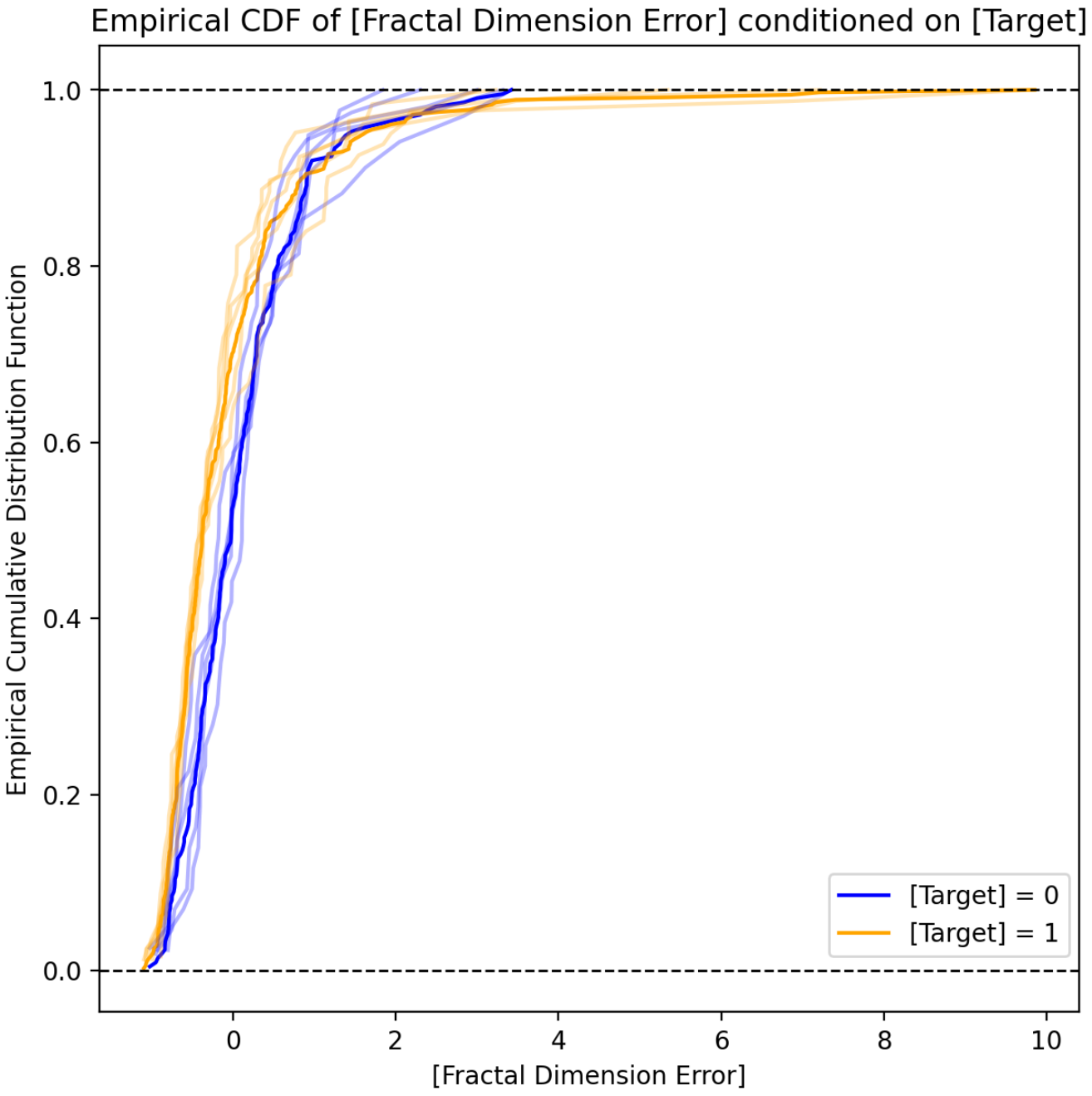
Fractal Dimension Error - Kernel Density Plot



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

Univariate Report

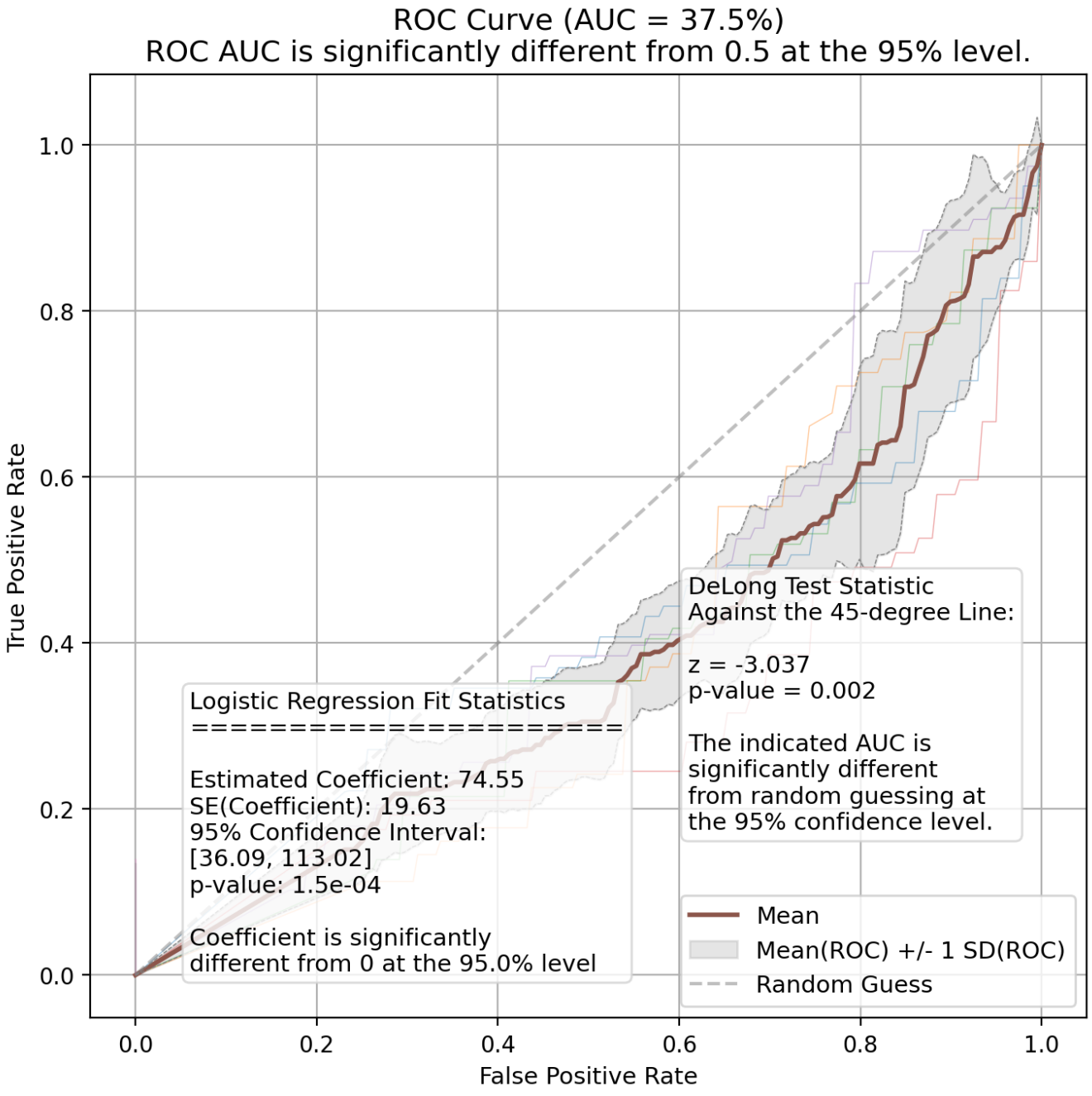
Fractal Dimension Error - Empirical CDF Plot



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

Univariate Report

Fractal Dimension Error - ROC Curve

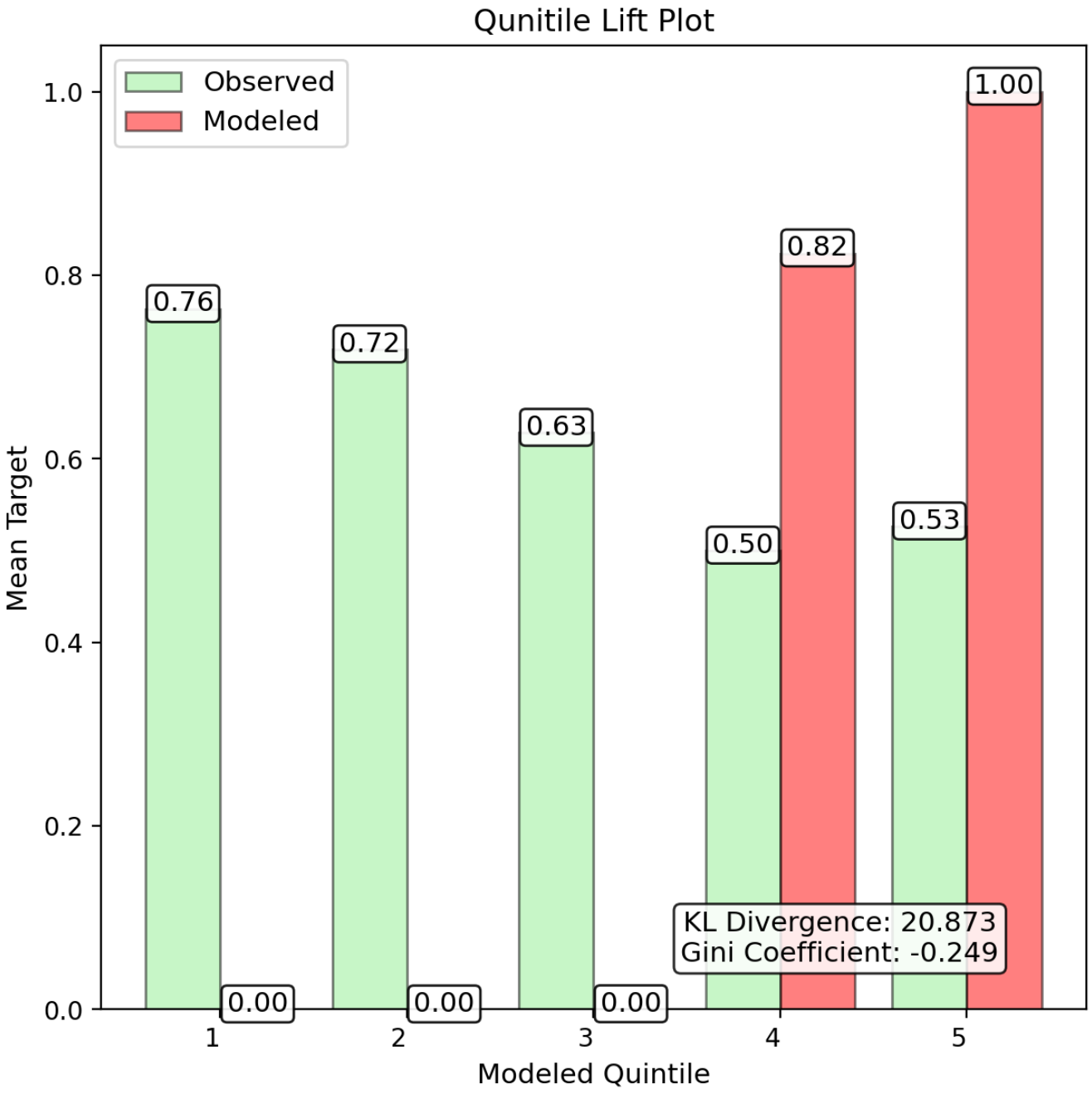


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Fractal Dimension Error - Quintile Lift



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

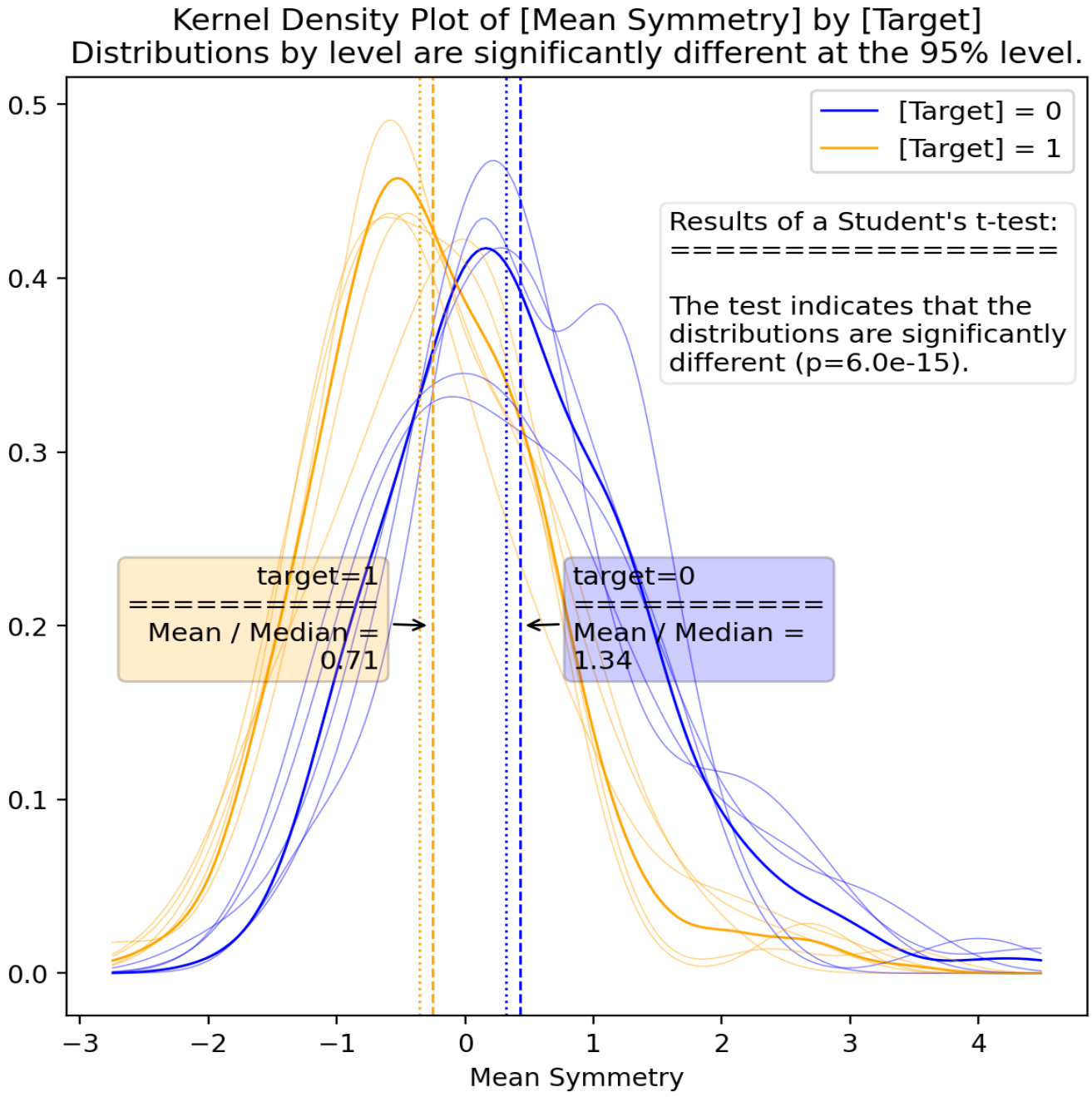
Univariate Report

Mean Symmetry - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-7.9e-01	-7.4e-01	-6.7e-01	-6.9e-01	-7.1e-01	2.2e+00	4.6e-02
Fitted p-Value	2.4e-11	7.3e-11	1.2e-09	6.1e-10	9.0e-10	2.2e-06	5.2e-10
Fitted Std. Err.	0.118	0.114	0.111	0.112	0.116	0.468	0.003
Conf. Int. Lower	-1.0e+00	-9.6e-01	-8.9e-01	-9.1e-01	-9.4e-01	1.3e+00	5.1e-02
Conf. Int. Upper	-5.6e-01	-5.2e-01	-4.6e-01	-4.7e-01	-4.8e-01	3.1e+00	4.1e-02
Train Accuracy	64.5%	64.1%	62.7%	62.0%	63.2%	36.6%	1.0%
Val Accuracy	58.1%	58.4%	67.3%	69.0%	62.6%	36.6%	5.0%
Train AUC	64.3%	64.2%	62.4%	62.1%	63.4%	36.6%	1.0%
Val AUC	59.2%	58.0%	67.4%	69.4%	61.3%	36.6%	5.0%
Train F1	69.5%	69.1%	67.7%	67.6%	68.5%	41.9%	0.8%
Test F1	63.4%	63.8%	74.1%	71.0%	68.4%	41.9%	4.6%
Train Precision	74.4%	75.5%	71.8%	74.4%	75.4%	49.2%	1.5%
Val Precision	73.8%	68.5%	82.8%	76.0%	68.8%	49.2%	5.9%
Train Recall	65.2%	63.7%	64.0%	62.0%	62.7%	36.4%	1.2%
Val Recall	55.6%	59.7%	67.1%	66.7%	67.9%	36.4%	5.5%
Train MCC	27.8%	27.5%	24.2%	23.2%	25.9%	-26.0%	2.0%
Val MCC	17.5%	15.7%	32.1%	38.4%	22.6%	-26.0%	9.7%
Train Log-Loss	12.80	12.94	13.44	13.68	13.25	22.87	0.36
Val Log-Loss	15.12	14.99	11.80	11.17	13.48	22.87	1.80

Univariate Report

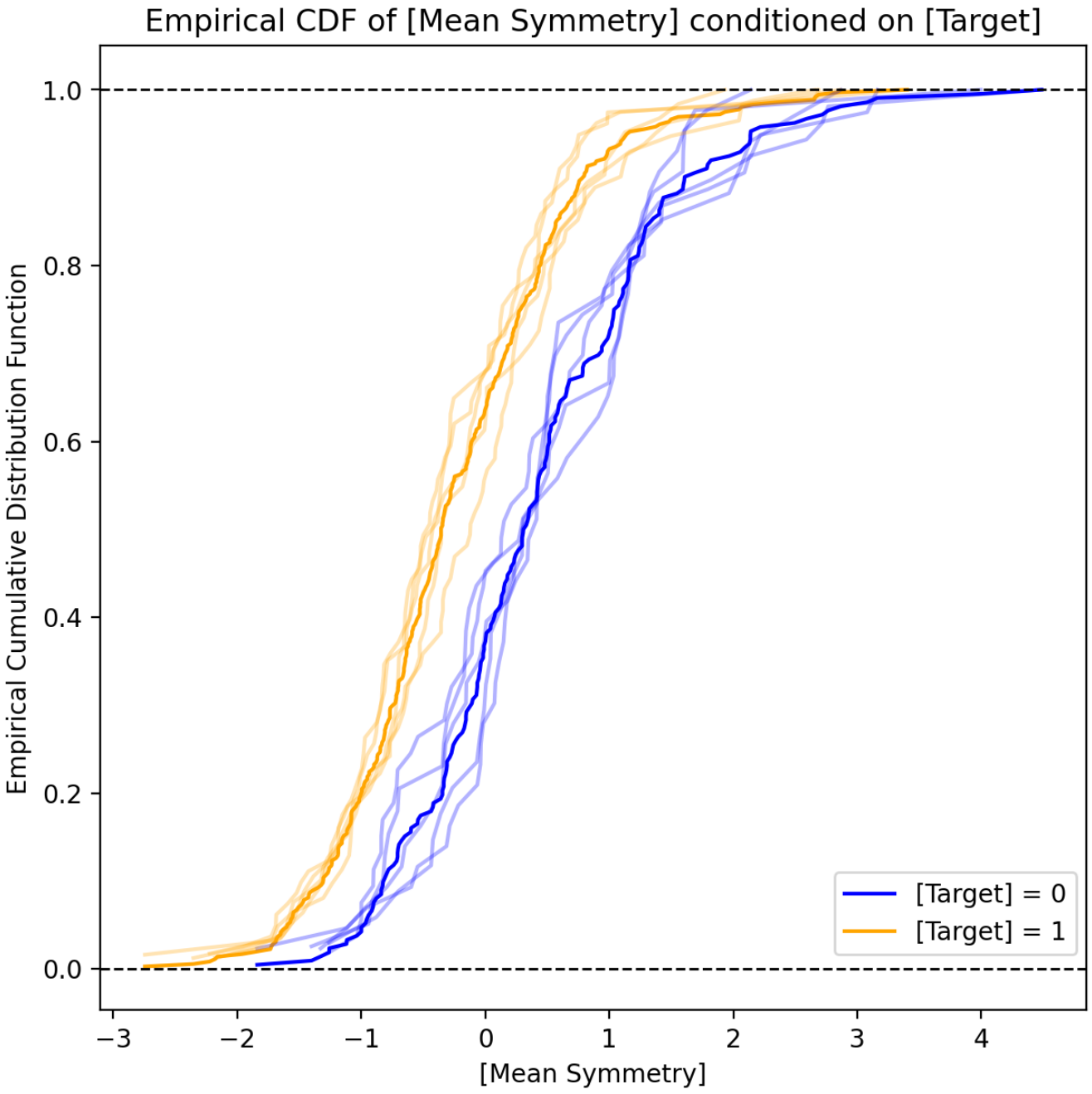
Mean Symmetry - Kernel Density Plot



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

Univariate Report

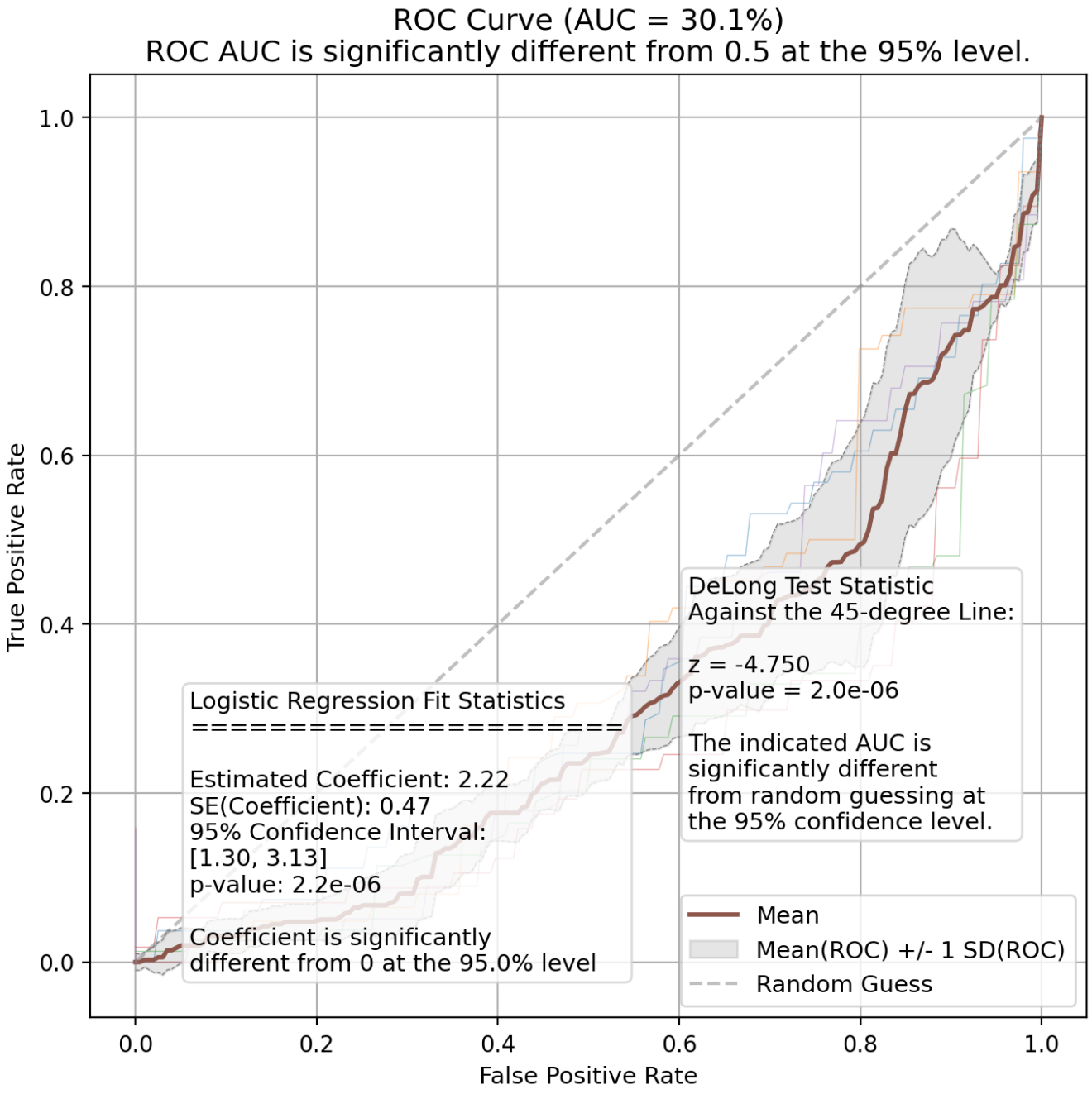
Mean Symmetry - Empirical CDF Plot



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

Univariate Report

Mean Symmetry - ROC Curve

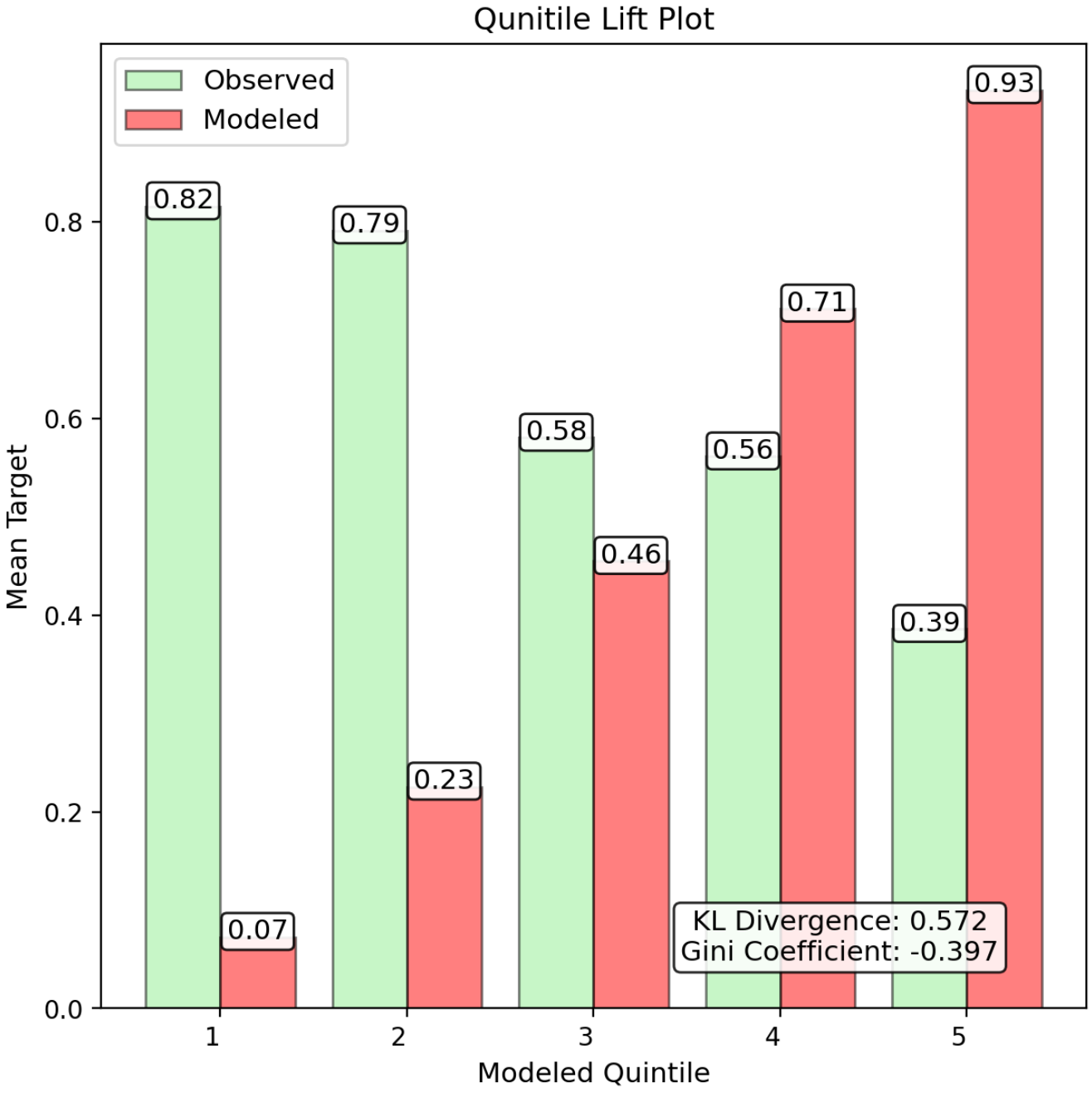


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Mean Symmetry - Quintile Lift



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

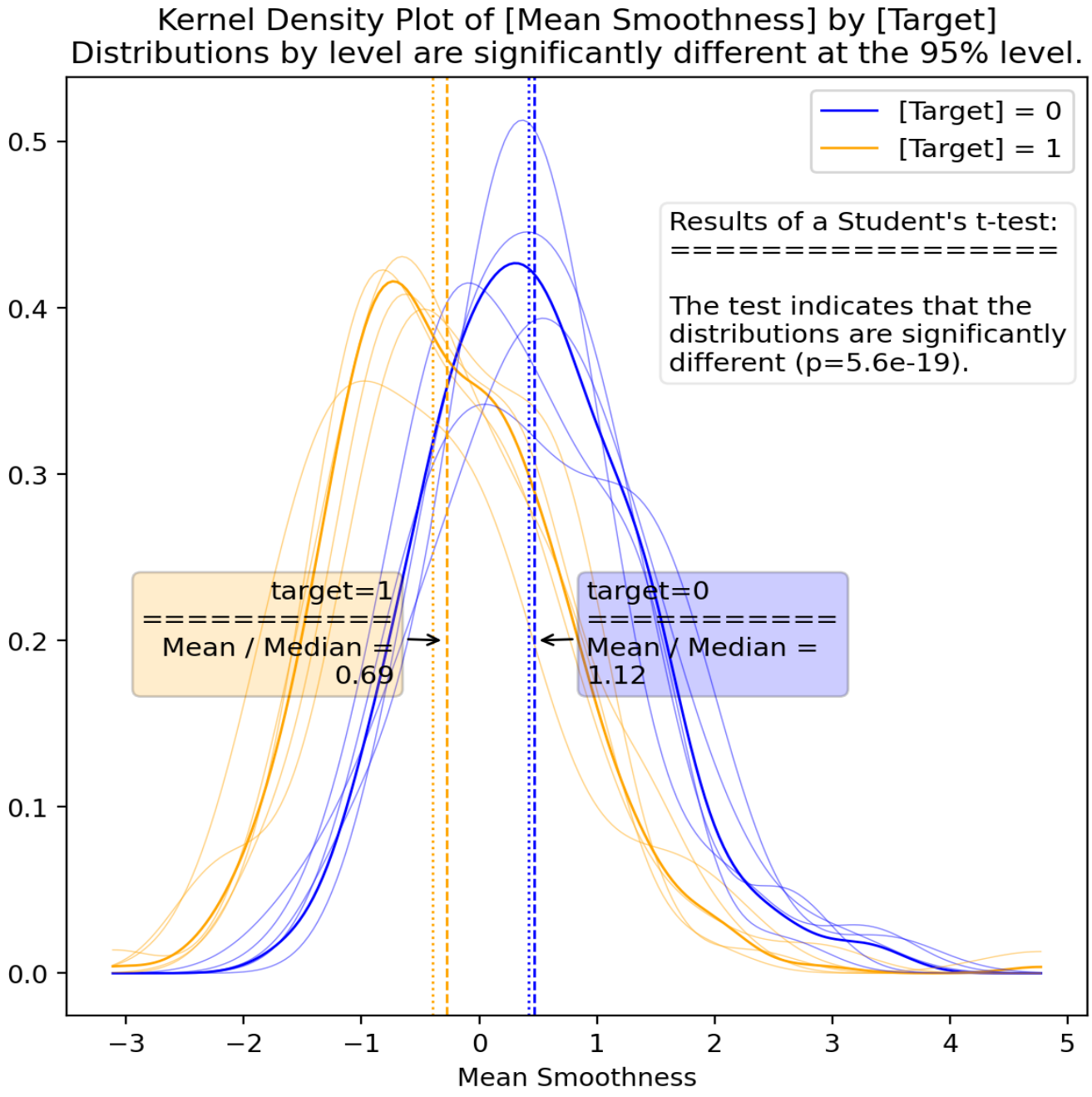
Univariate Report

Mean Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-8.0e-01	-7.7e-01	-8.2e-01	-7.7e-01	-7.6e-01	4.1e+00	2.5e-02
Fitted p-Value	5.9e-12	1.1e-11	8.8e-13	1.1e-11	3.1e-11	2.7e-06	1.1e-11
Fitted Std. Err.	0.117	0.113	0.115	0.114	0.114	0.880	0.001
Conf. Int. Lower	-1.0e+00	-9.9e-01	-1.0e+00	-1.0e+00	-9.8e-01	2.4e+00	2.7e-02
Conf. Int. Upper	-5.7e-01	-5.5e-01	-6.0e-01	-5.5e-01	-5.4e-01	5.9e+00	2.4e-02
Train Accuracy	64.7%	64.3%	64.9%	63.3%	66.2%	35.5%	1.0%
Val Accuracy	63.7%	65.3%	62.8%	70.0%	62.6%	35.5%	3.1%
Train AUC	65.2%	65.1%	65.3%	64.2%	67.5%	34.7%	1.2%
Val AUC	65.7%	66.1%	66.7%	70.3%	61.6%	34.7%	3.1%
Train F1	68.9%	68.7%	68.9%	68.0%	70.3%	42.4%	0.8%
Test F1	68.1%	69.0%	68.2%	72.2%	68.0%	42.4%	1.8%
Train Precision	76.0%	76.9%	75.0%	76.9%	79.9%	48.2%	1.8%
Val Precision	80.0%	76.5%	84.9%	76.5%	69.3%	48.2%	5.7%
Train Recall	63.0%	62.0%	63.7%	61.0%	62.7%	37.8%	1.0%
Val Recall	59.3%	62.9%	57.0%	68.4%	66.7%	37.8%	4.8%
Train MCC	29.6%	29.2%	29.8%	27.3%	33.7%	-29.6%	2.3%
Val MCC	29.9%	31.3%	30.7%	40.1%	23.1%	-29.6%	6.1%
Train Log-Loss	12.72	12.86	12.65	13.22	12.18	23.25	0.38
Val Log-Loss	13.08	12.49	13.40	10.81	13.48	23.25	1.10

Univariate Report

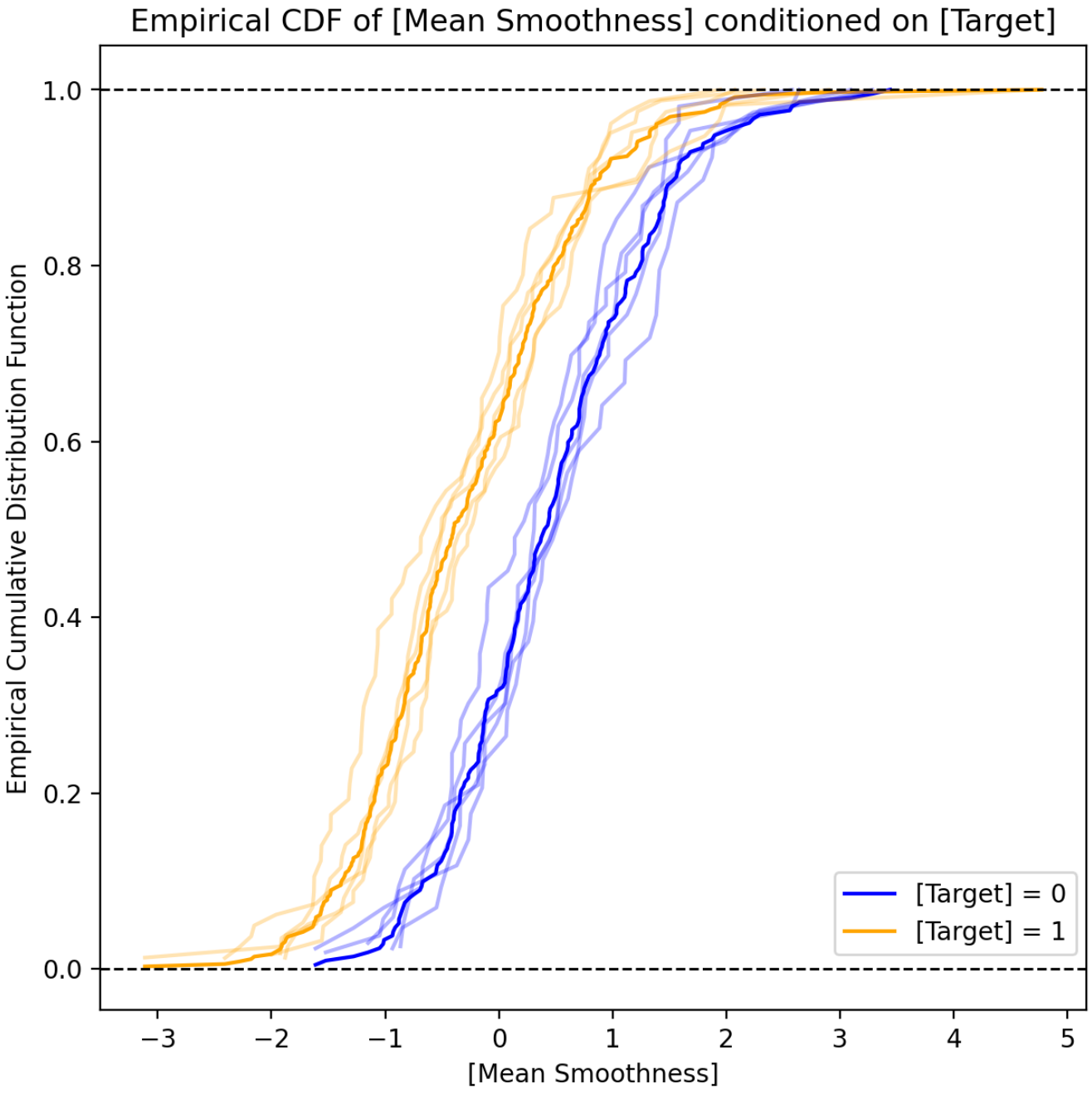
Mean Smoothness - Kernel Density Plot



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

Univariate Report

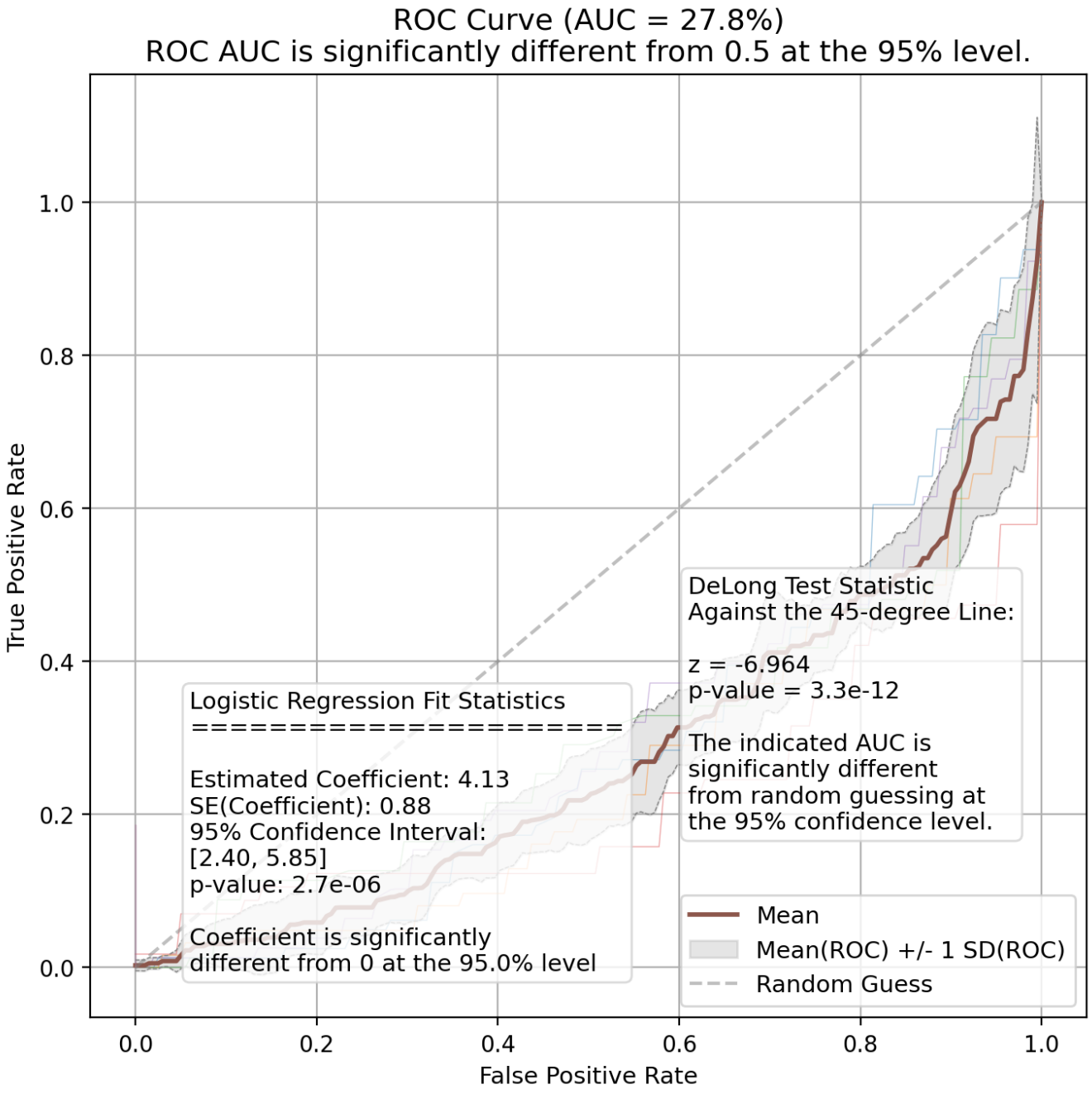
Mean Smoothness - Empirical CDF Plot



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

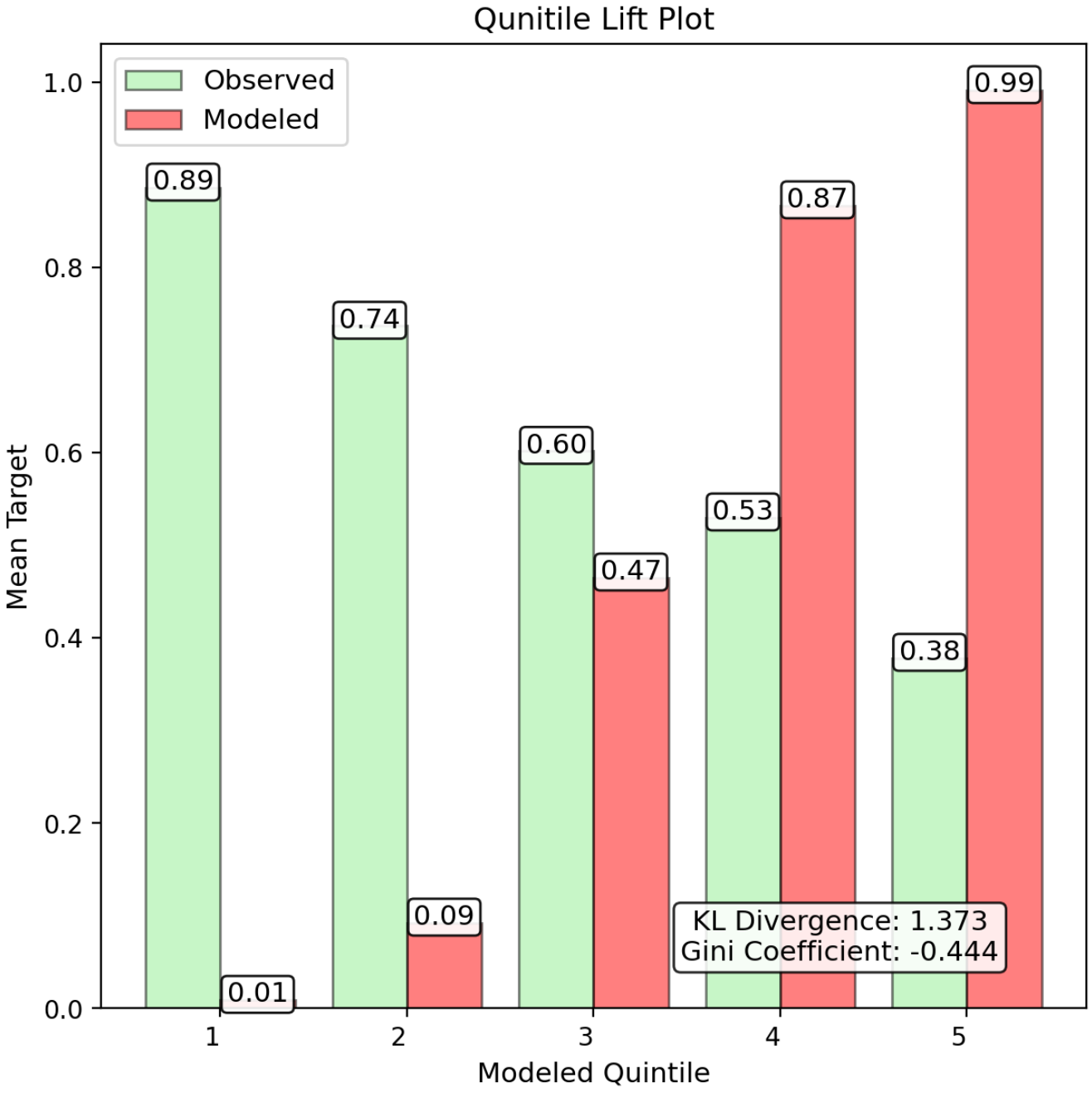


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Mean 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.

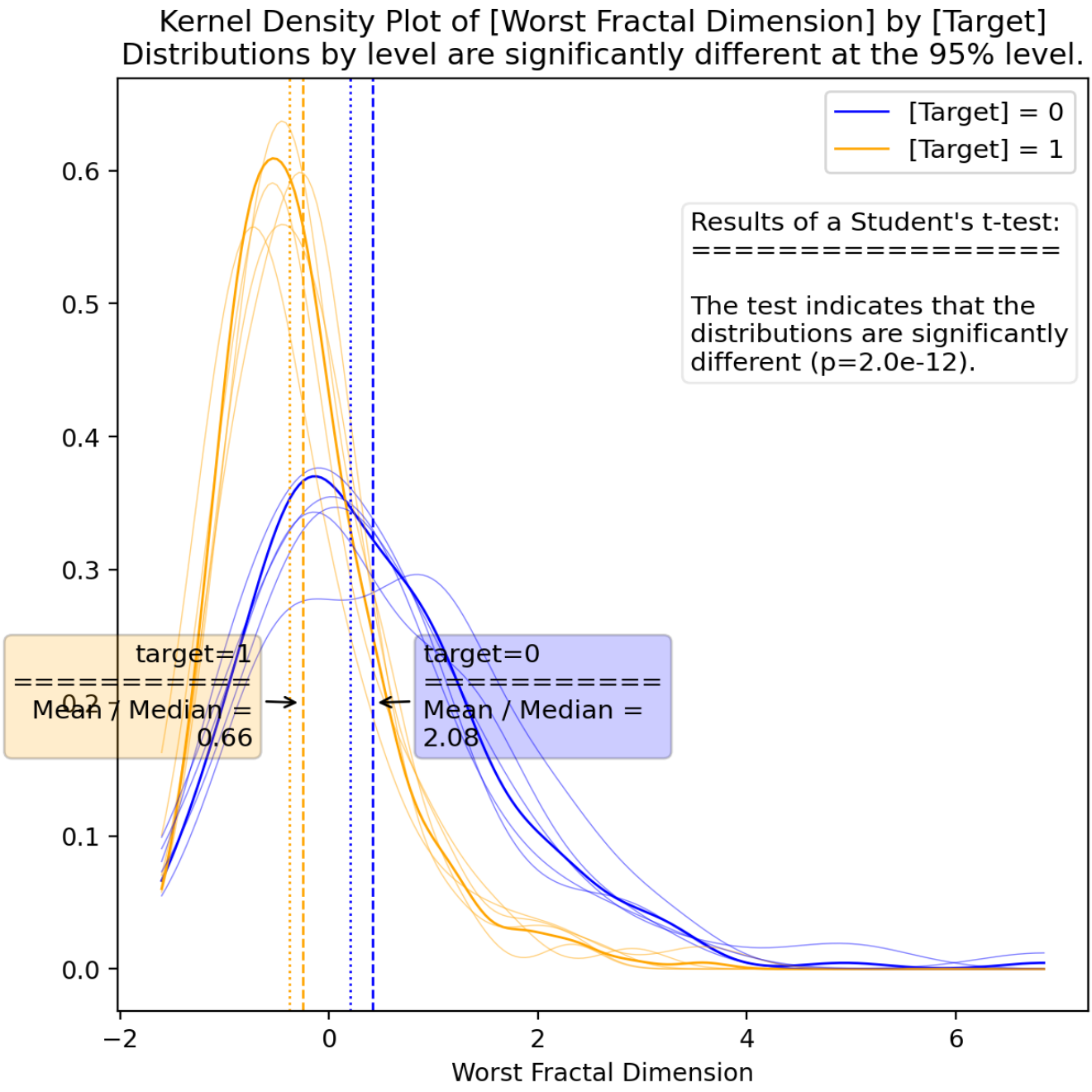
Univariate Report

Worst Fractal Dimension - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-8.0e-01	-7.0e-01	-7.4e-01	-7.8e-01	-7.5e-01	4.2e+00	3.6e-02
Fitted p-Value	1.0e-10	2.4e-09	9.1e-10	3.1e-10	5.3e-10	2.1e-05	9.3e-10
Fitted Std. Err.	0.123	0.118	0.120	0.124	0.121	0.996	0.002
Conf. Int. Lower	-1.0e+00	-9.4e-01	-9.7e-01	-1.0e+00	-9.9e-01	2.3e+00	4.1e-02
Conf. Int. Upper	-5.6e-01	-4.7e-01	-5.0e-01	-5.4e-01	-5.2e-01	6.2e+00	3.2e-02
Train Accuracy	67.4%	65.8%	65.1%	66.5%	66.4%	33.7%	0.9%
Val Accuracy	62.1%	68.3%	70.8%	65.0%	65.6%	33.7%	3.3%
Train AUC	65.7%	63.9%	63.4%	64.5%	64.3%	35.7%	0.9%
Val AUC	59.0%	67.5%	68.2%	63.6%	64.2%	35.7%	3.7%
Train F1	73.5%	72.4%	71.4%	73.3%	73.2%	34.7%	0.9%
Test F1	70.4%	73.3%	78.1%	70.6%	71.3%	34.7%	3.2%
Train Precision	74.2%	73.7%	71.5%	74.9%	74.4%	45.5%	1.3%
Val Precision	71.8%	75.9%	81.9%	67.7%	70.9%	45.5%	5.5%
Train Recall	72.8%	71.2%	71.2%	71.7%	72.0%	28.0%	0.7%
Val Recall	69.1%	71.0%	74.7%	73.7%	71.8%	28.0%	2.2%
Train MCC	31.2%	27.5%	26.8%	28.6%	28.3%	-28.4%	1.7%
Val MCC	17.7%	34.5%	34.8%	27.7%	28.5%	-28.4%	6.9%
Train Log-Loss	11.74	12.32	12.57	12.07	12.10	23.88	0.31
Val Log-Loss	13.66	11.42	10.53	12.62	12.38	23.88	1.20

Univariate Report

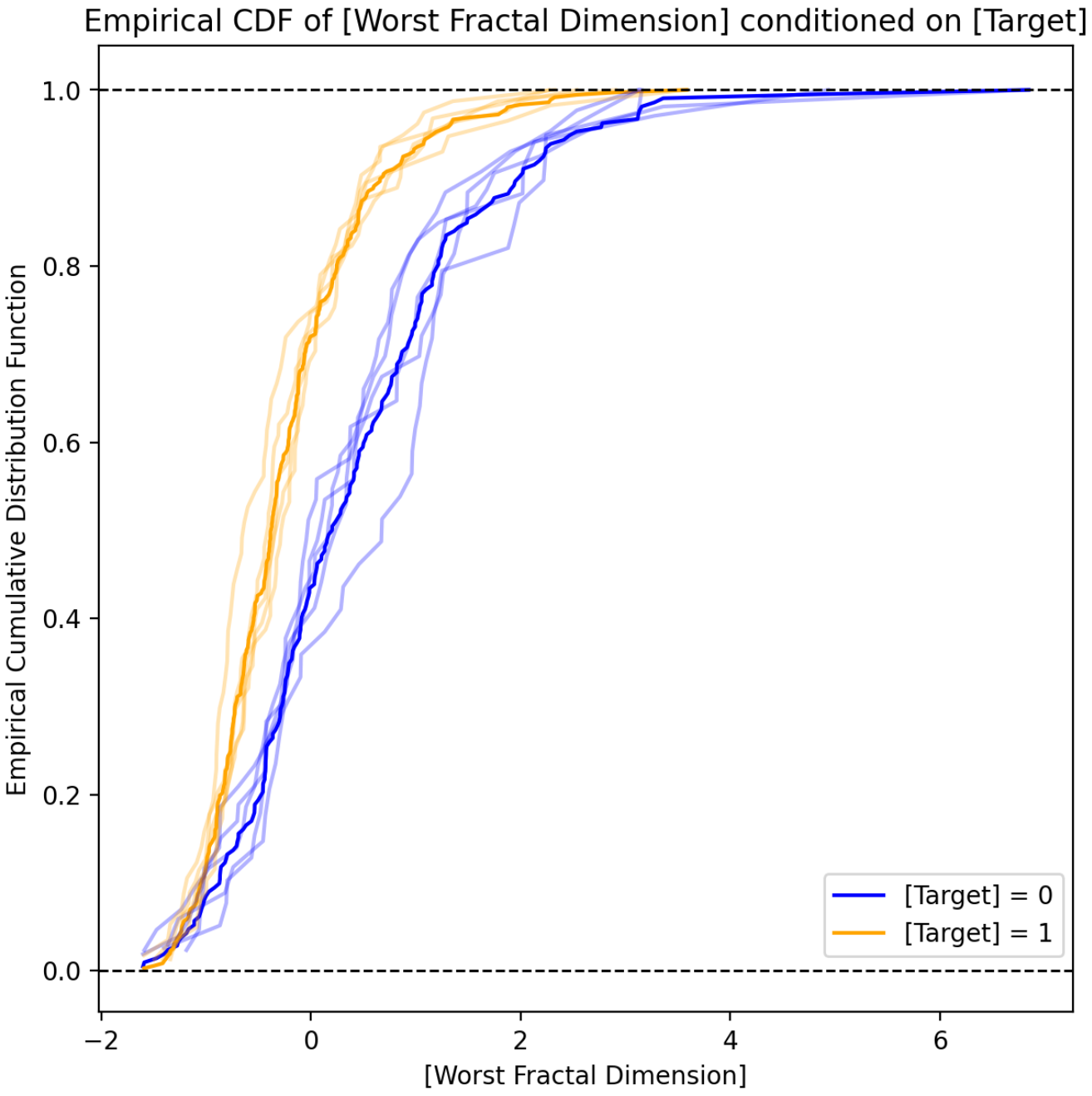
Worst Fractal Dimension - Kernel Density Plot



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

Univariate Report

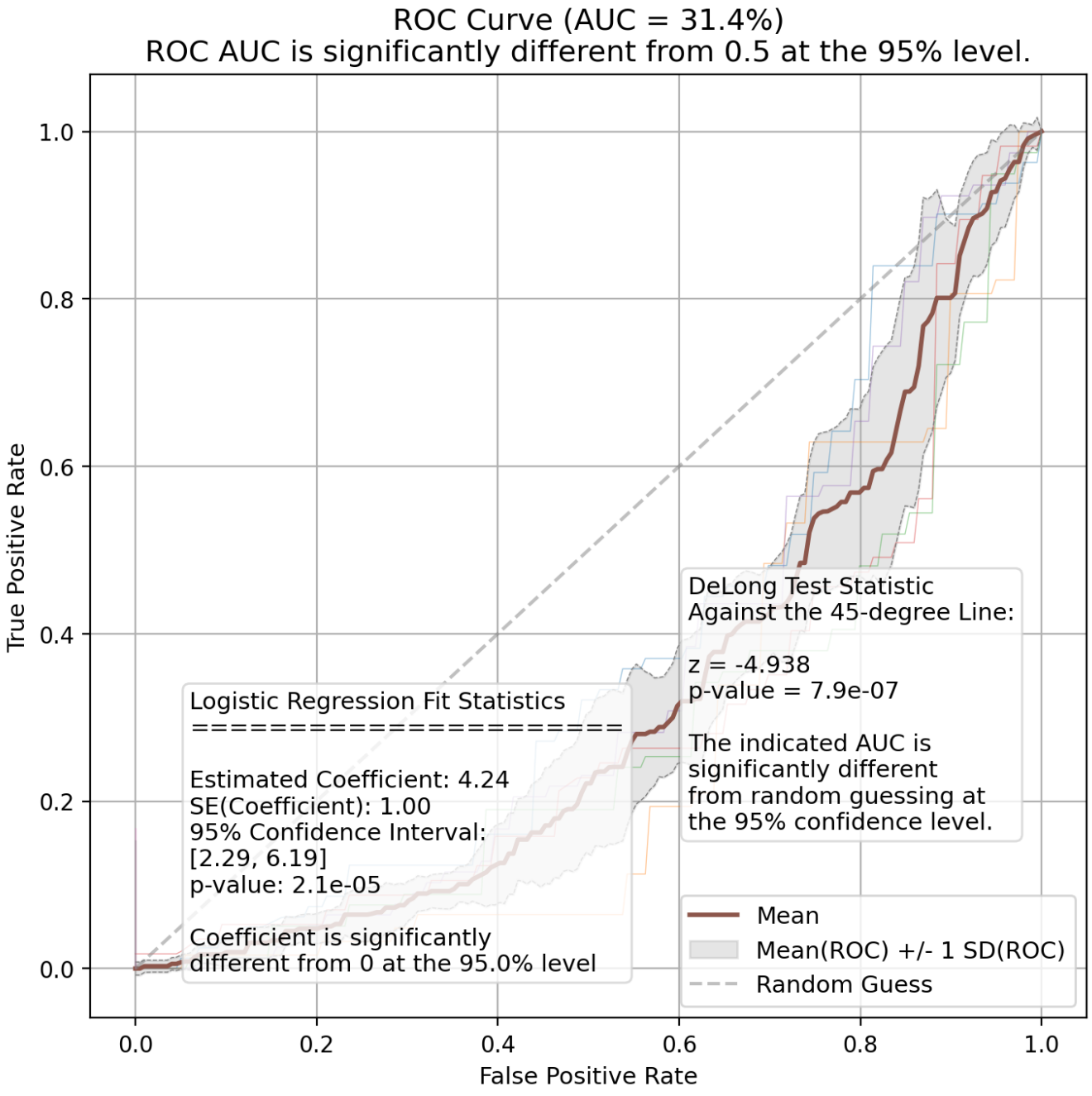
Worst Fractal Dimension - Empirical CDF Plot



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

Univariate Report

Worst Fractal Dimension - ROC Curve

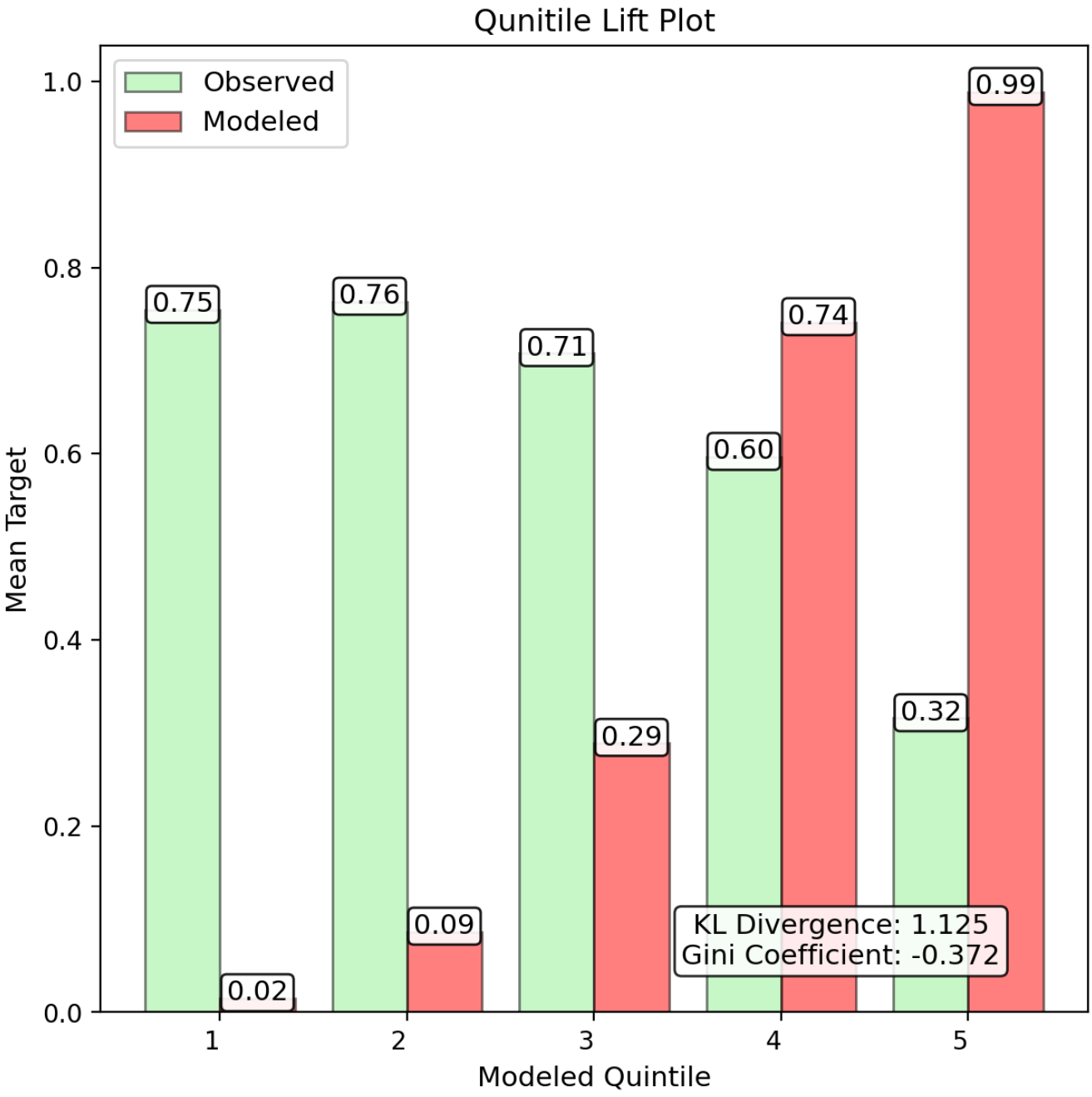


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Worst Fractal Dimension - Quintile Lift



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

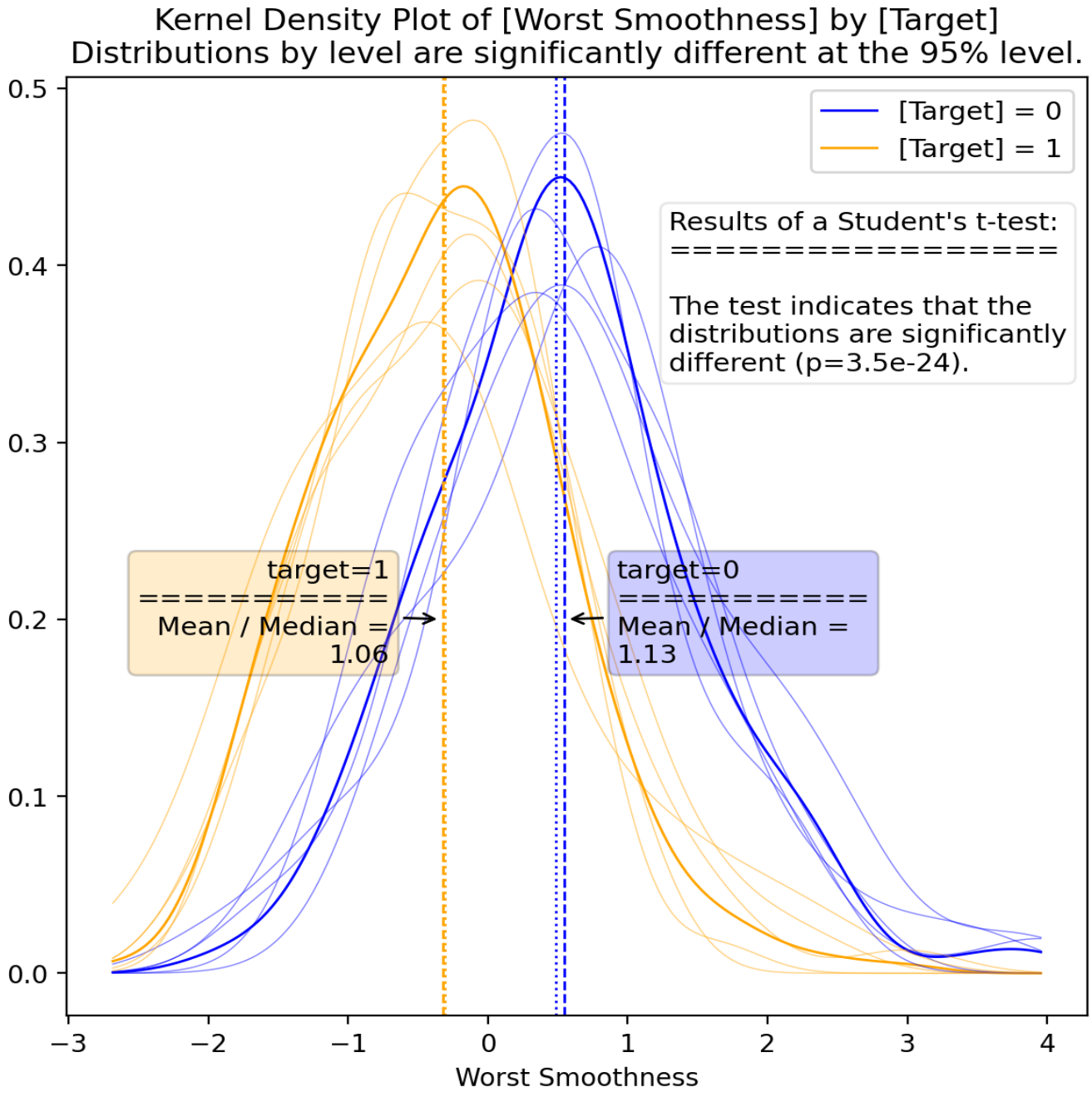
Univariate Report

Worst Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-9.1e-01	-1.0e+00	-1.0e+00	-1.0e+00	-9.1e-01	2.7e+00	5.9e-02
Fitted p-Value	4.4e-14	8.9e-16	5.0e-16	8.5e-16	4.7e-14	2.4e-05	2.4e-14
Fitted Std. Err.	0.121	0.124	0.128	0.126	0.121	0.636	0.003
Conf. Int. Lower	-1.1e+00	-1.2e+00	-1.3e+00	-1.3e+00	-1.1e+00	1.4e+00	6.5e-02
Conf. Int. Upper	-6.7e-01	-7.6e-01	-7.9e-01	-7.7e-01	-6.8e-01	3.9e+00	5.3e-02
Train Accuracy	67.4%	68.2%	68.4%	67.2%	68.9%	32.3%	0.7%
Val Accuracy	70.2%	65.3%	65.5%	72.0%	64.9%	32.3%	3.3%
Train AUC	67.8%	69.0%	69.0%	68.4%	69.9%	31.5%	0.8%
Val AUC	72.3%	68.0%	66.9%	71.4%	64.8%	31.5%	3.1%
Train F1	71.6%	72.3%	71.9%	71.4%	73.1%	39.2%	0.7%
Test F1	74.1%	66.7%	71.9%	75.4%	68.9%	39.2%	3.6%
Train Precision	77.9%	80.2%	78.6%	80.7%	81.5%	44.9%	1.5%
Val Precision	85.5%	81.4%	83.3%	75.4%	72.9%	44.9%	5.4%
Train Recall	66.3%	65.8%	66.2%	64.0%	66.3%	34.7%	1.0%
Val Recall	65.4%	56.5%	63.3%	75.4%	65.4%	34.7%	6.8%
Train MCC	34.5%	36.7%	37.2%	35.3%	38.4%	-35.8%	1.5%
Val MCC	42.4%	35.4%	31.1%	42.9%	29.1%	-35.8%	6.3%
Train Log-Loss	11.74	11.48	11.38	11.84	11.19	24.39	0.26
Val Log-Loss	10.75	12.49	12.44	10.09	12.66	24.39	1.18

Univariate Report

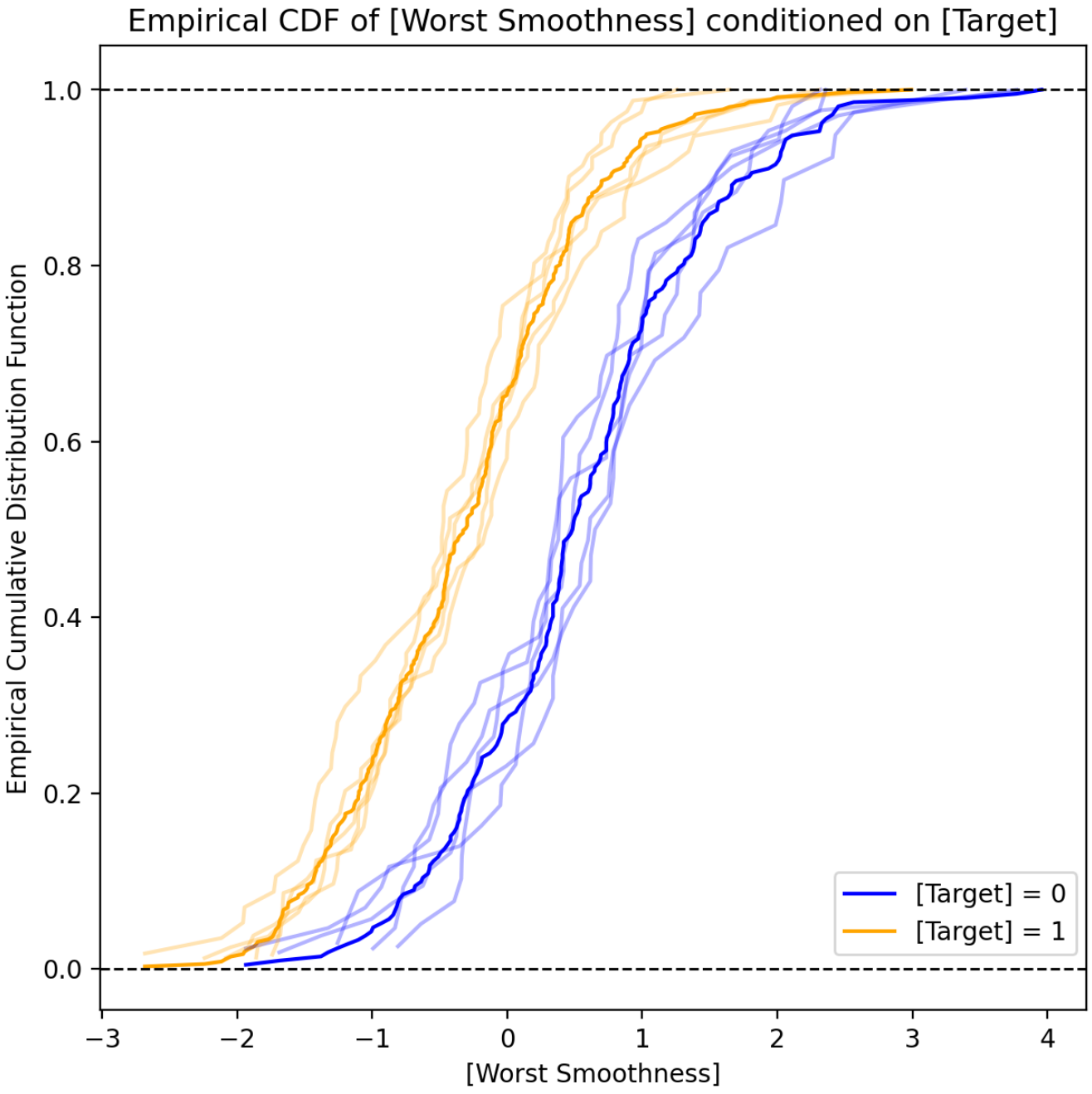
Worst Smoothness - Kernel Density Plot



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

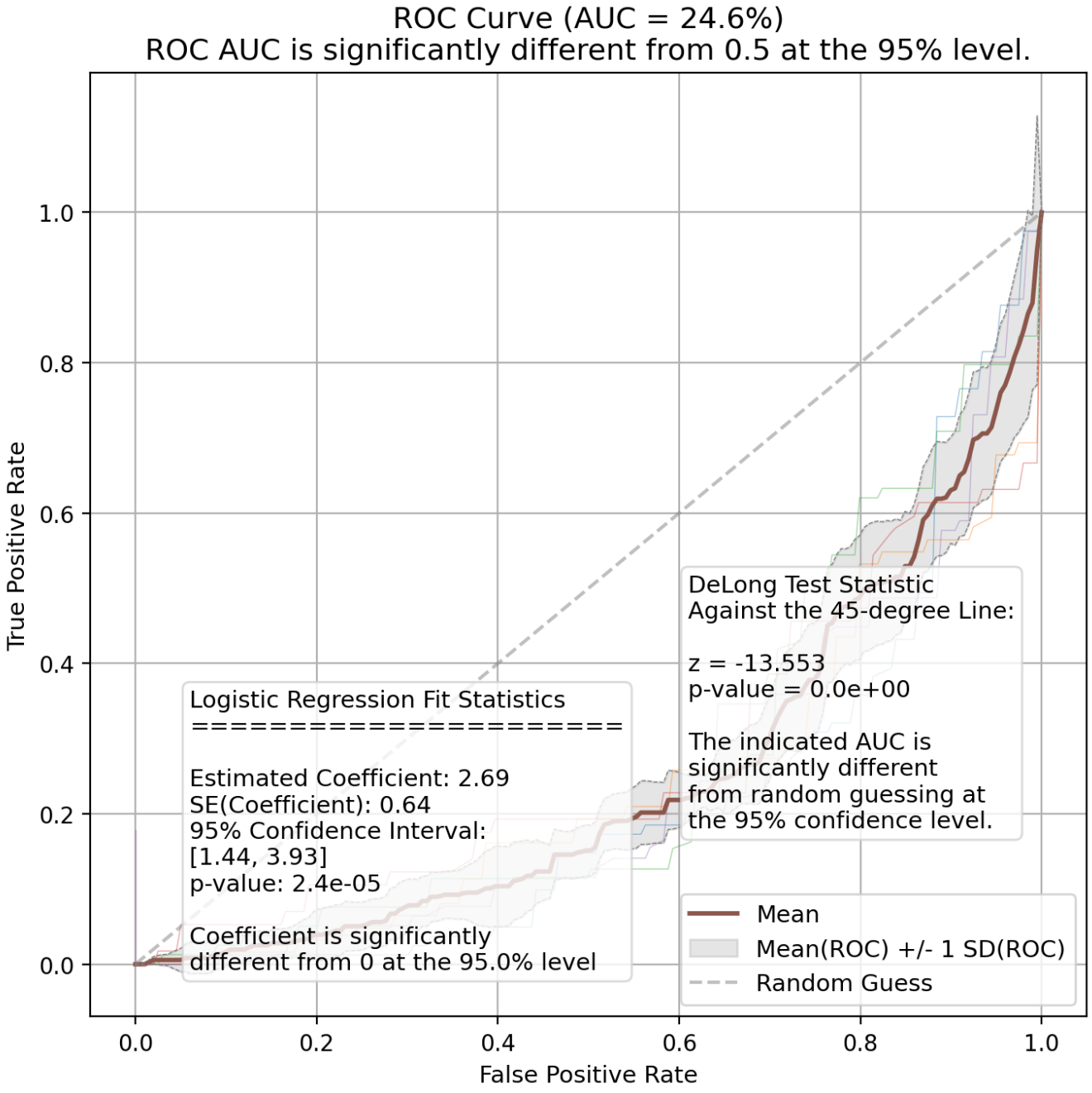
Worst Smoothness - Empirical CDF Plot



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

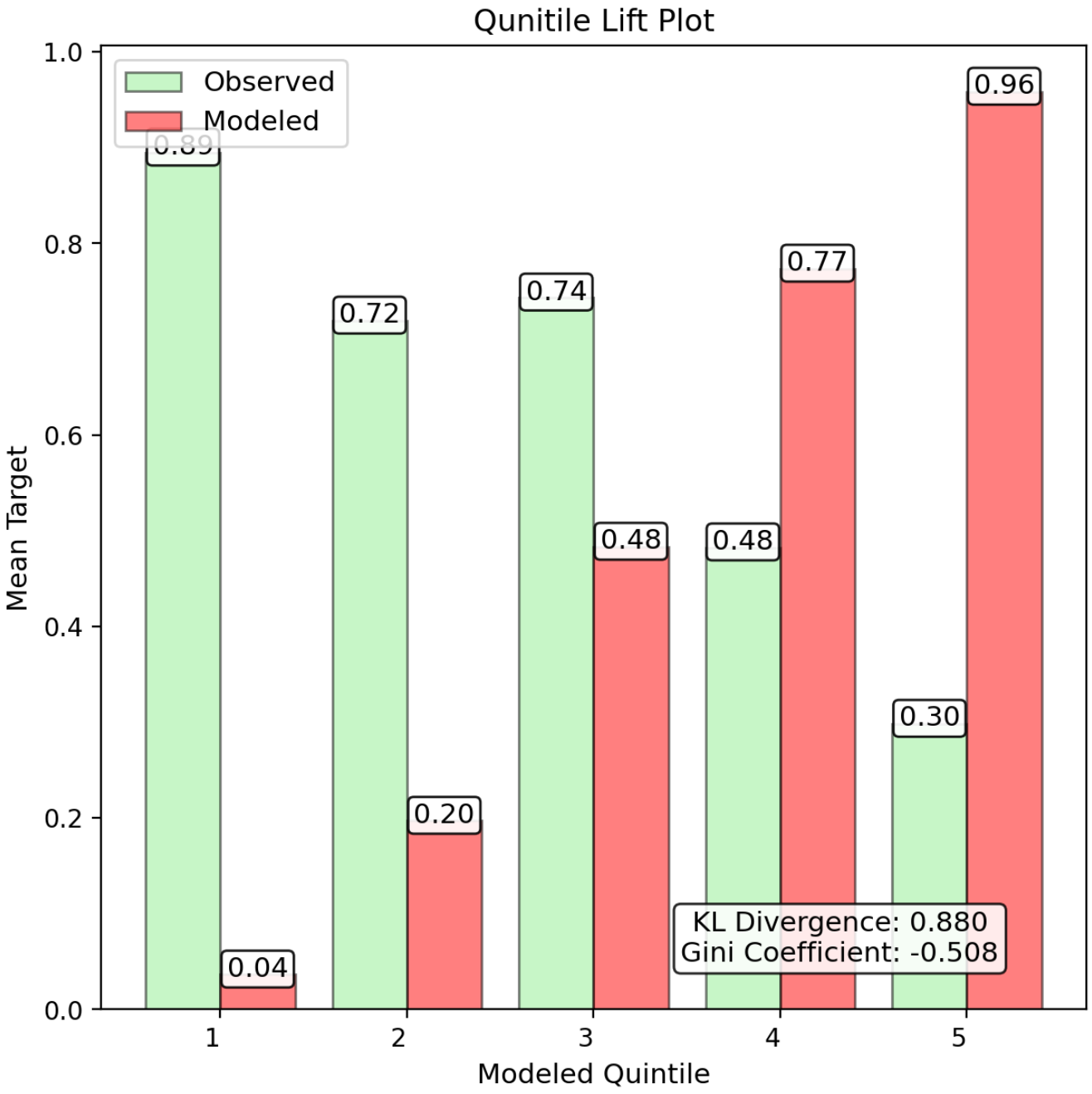


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, 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 a modified version the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and I calculate the empirical standard error from the cross-validated AUC values. I then calculate a z-score for the AUC, and use the z-score to calculate a p-value. The p-value is then used to determine the significance of the AUC. This is a simple test, and should be used with caution.

Univariate Report

Worst 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.