Univariate Analysis Report

Cancer Model

2024-02-02

Overview

Cancer Model Univariate Analysis Report

These sorted results for the features in this report indicate the average cross-validated test scores for each feature, if it were used as the only predictor in a simple linear model. Keep in mind that these results are based on the average, without considering the standard deviation. This means that the results are not necessarily the best predictors, but they are the best on average, and provide a fine starting point for grouping those predictors that are on average better than others. This means that nothing was done to account for possible sampling variability in the 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.
worst_concavity	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
worst_compactness	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
mean_area	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
worst_area	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
worst_perimeter	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
worst_radius	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
mean_compactness	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
mean_concavity	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
concave_points_error	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
concavity_error	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
mean_concave_points	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
area_error	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
perimeter_error	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%
mean_radius	37.7%	0.00e+00	0.00e+00	50.0%	0.00e+00	0.00e+00	14.6%

This table shows an overview of the results for the variables in this file, representing those whose average test score are ranked between 17 and 31 of the variables passed to the Cancer Model.

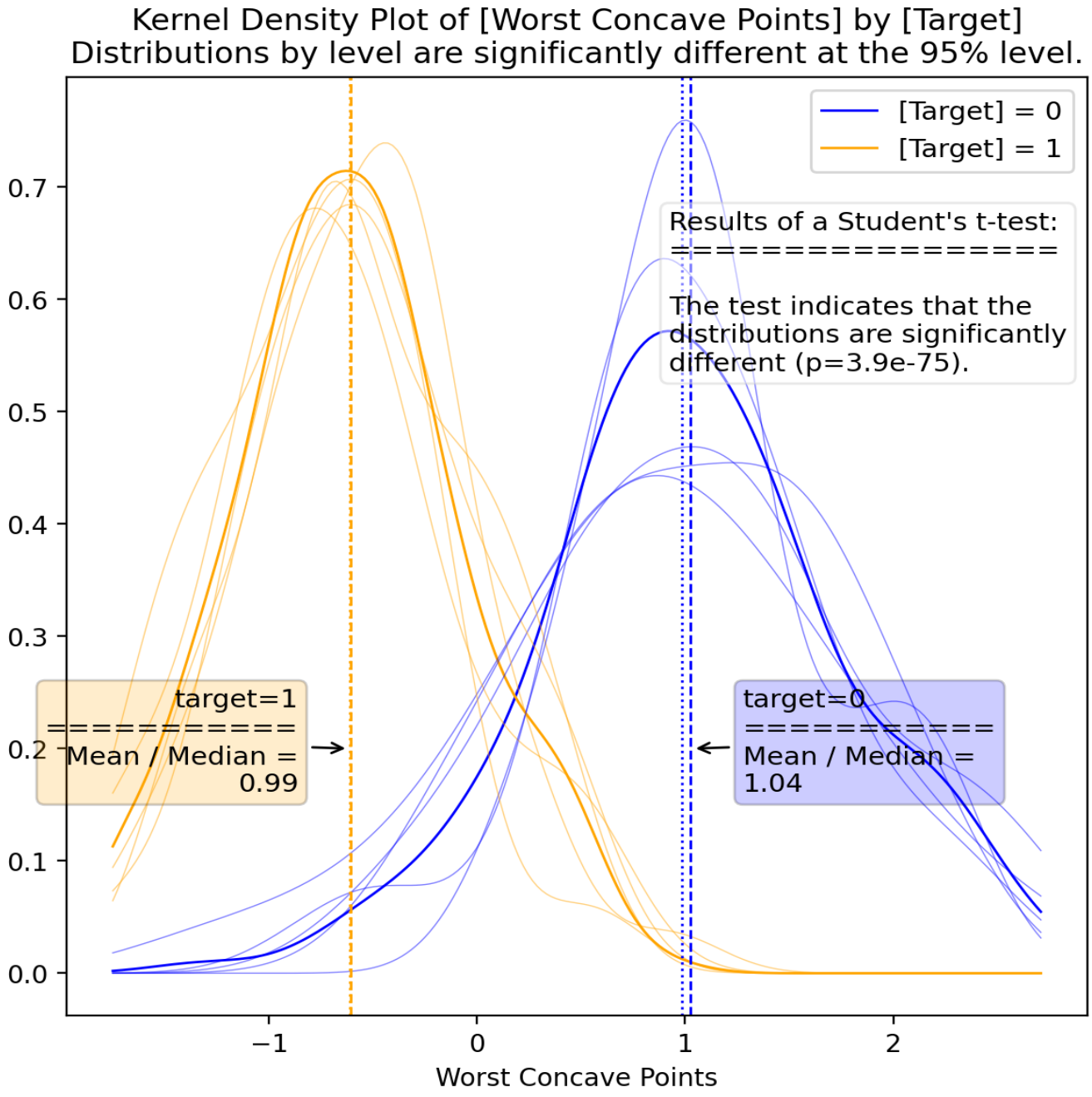
Univariate Report

Worst Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.4e+00	-3.4e+00	-3.6e+00	-3.2e+00	-3.2e+00	-3.4e+00	1.7e-01
Fitted p-Value	1.7e-23	2.0e-22	1.7e-22	5.5e-24	3.2e-24	8.5e-29	9.9e-23
Fitted Std. Err.	0.336	0.351	0.374	0.320	0.317	0.302	0.024
Conf. Int. Lower	-4.0e+00	-4.1e+00	-4.4e+00	-3.9e+00	-3.8e+00	-4.0e+00	2.2e-01
Conf. Int. Upper	-2.7e+00	-2.7e+00	-2.9e+00	-2.6e+00	-2.6e+00	-2.8e+00	1.3e-01
Train Accuracy	89.2%	89.6%	89.5%	87.3%	87.4%	88.6%	1.2%
Val Accuracy	86.3%	84.8%	84.9%	94.1%	94.0%	37.7%	4.8%
Train AUC	89.8%	90.1%	90.4%	87.9%	88.5%	89.3%	1.1%
Val AUC	87.2%	87.3%	85.0%	94.8%	93.1%	50.0%	4.2%
Train F1	91.2%	91.1%	91.2%	89.7%	89.3%	90.5%	0.9%
Test F1	87.4%	88.5%	87.7%	94.5%	95.3%	0.0%	3.9%
Train Precision	95.3%	94.5%	96.1%	94.0%	95.1%	95.0%	0.8%
Val Precision	93.8%	96.7%	90.9%	100.0%	94.4%	0.0%	3.4%
Train Recall	87.4%	87.9%	86.8%	85.7%	84.1%	86.4%	1.5%
Val Recall	81.8%	81.7%	84.7%	89.6%	96.2%	0.0%	6.2%
Train MCC	77.7%	79.0%	78.9%	73.8%	74.9%	76.8%	2.4%
Val MCC	73.4%	68.7%	68.6%	88.8%	86.9%	0.0%	9.9%
Train Log-Loss	3.90	3.75	3.78	4.58	4.55	4.12	0.42
Val Log-Loss	4.93	5.46	5.43	2.12	2.17	22.45	1.73

Univariate Report

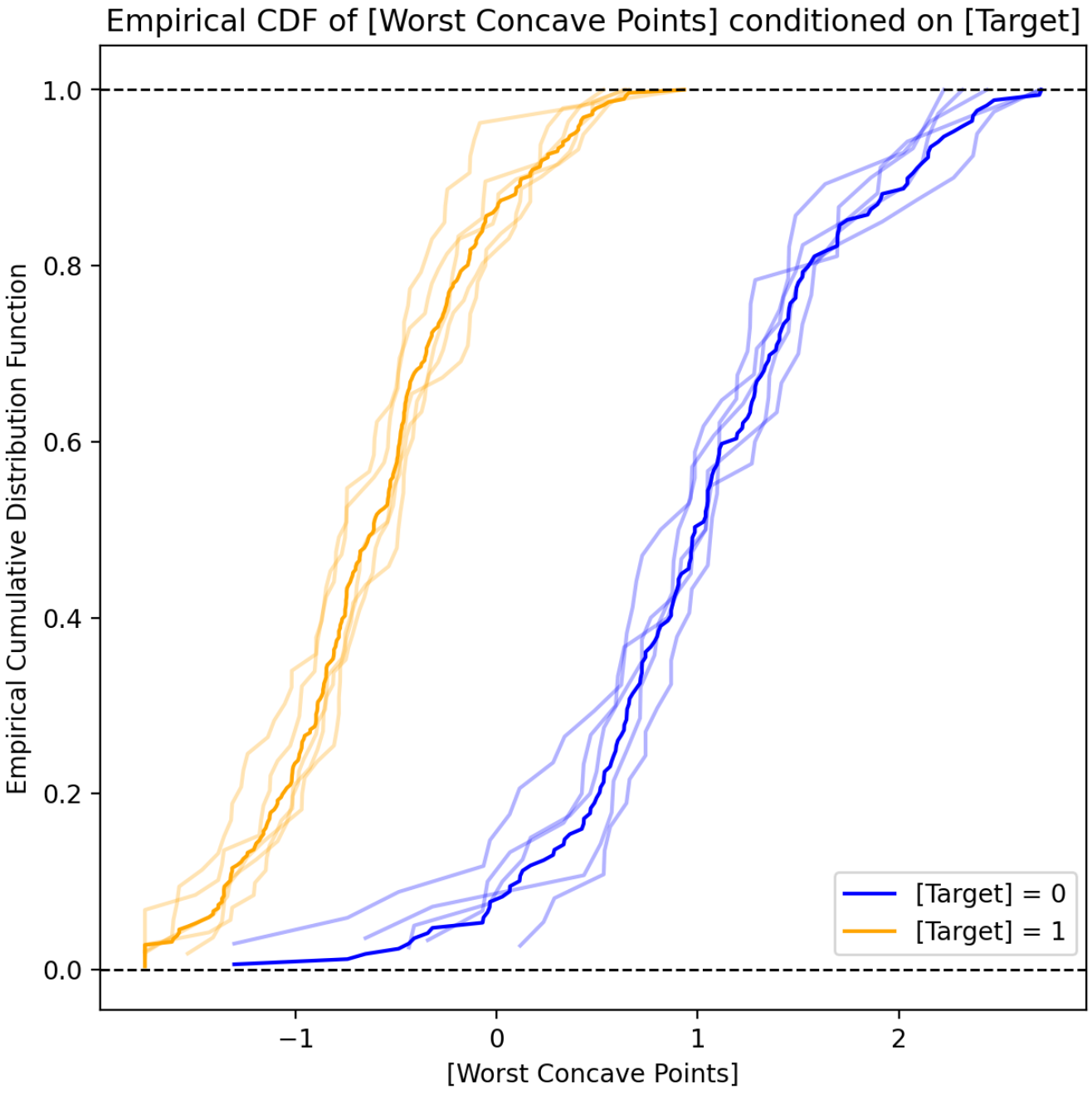
Worst Concave Points - Kernel Density Plot



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

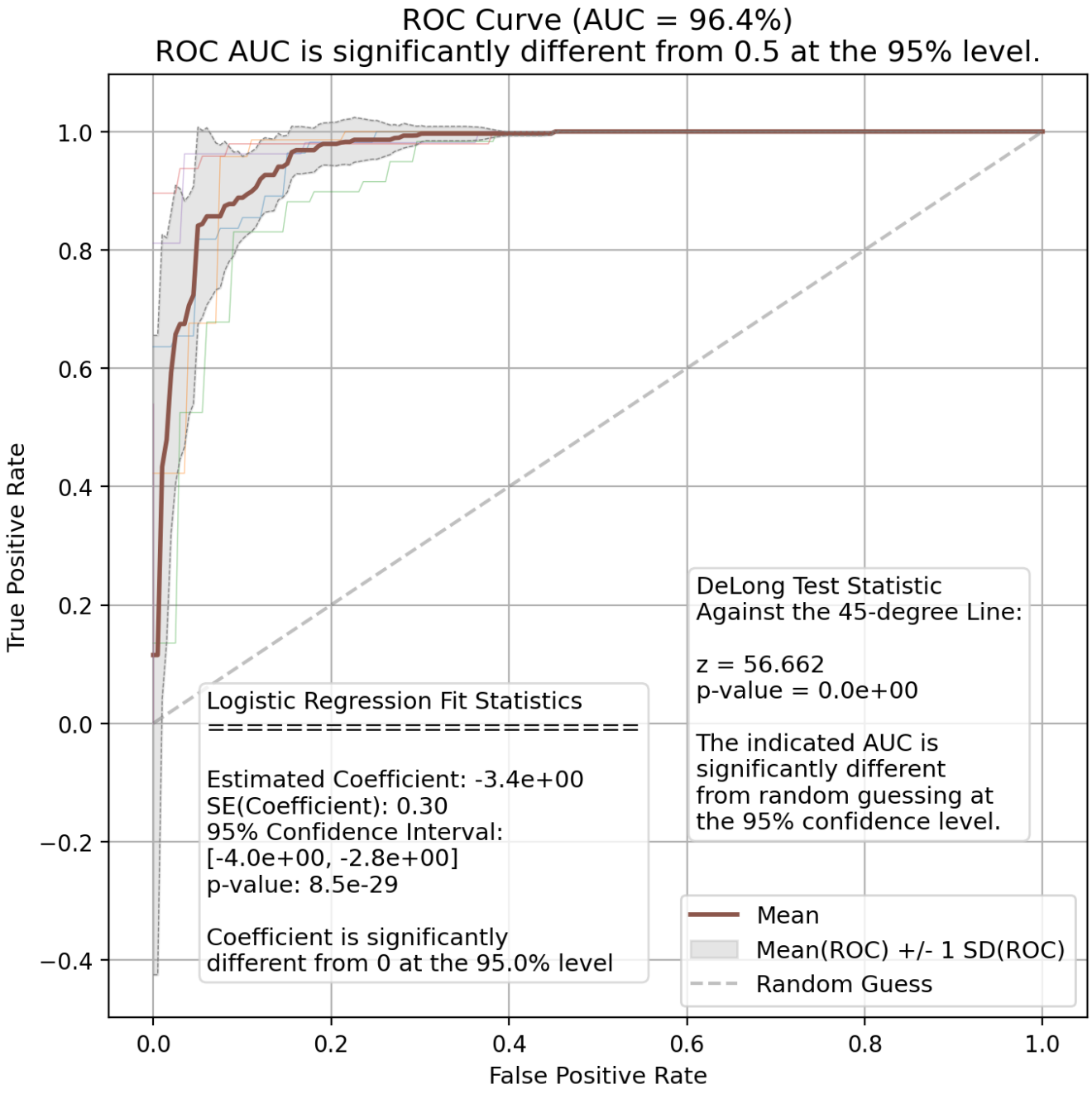
Worst Concave Points - Empirical CDF Plot



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

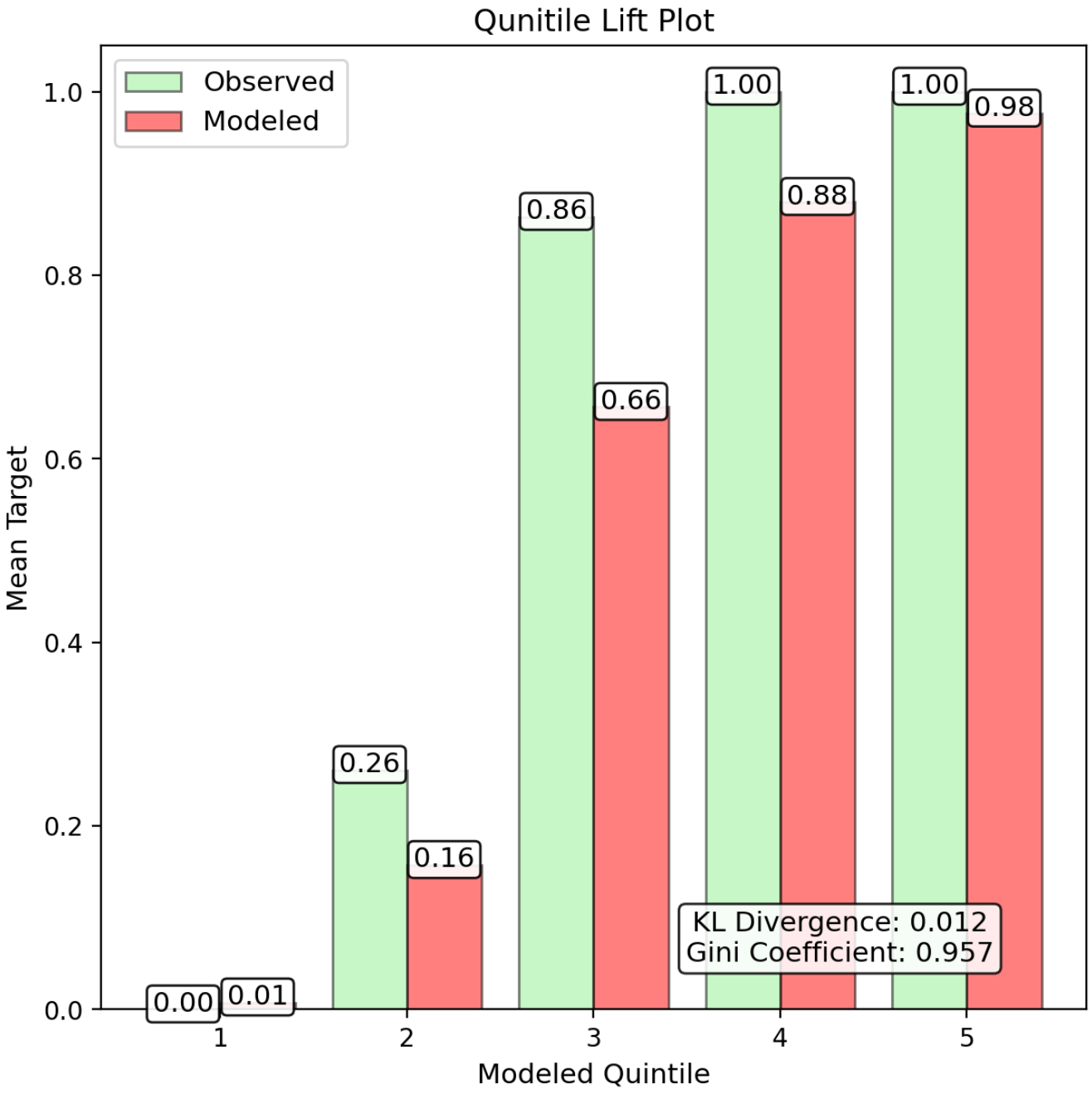


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Worst Concave Points - Quintile Lift



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

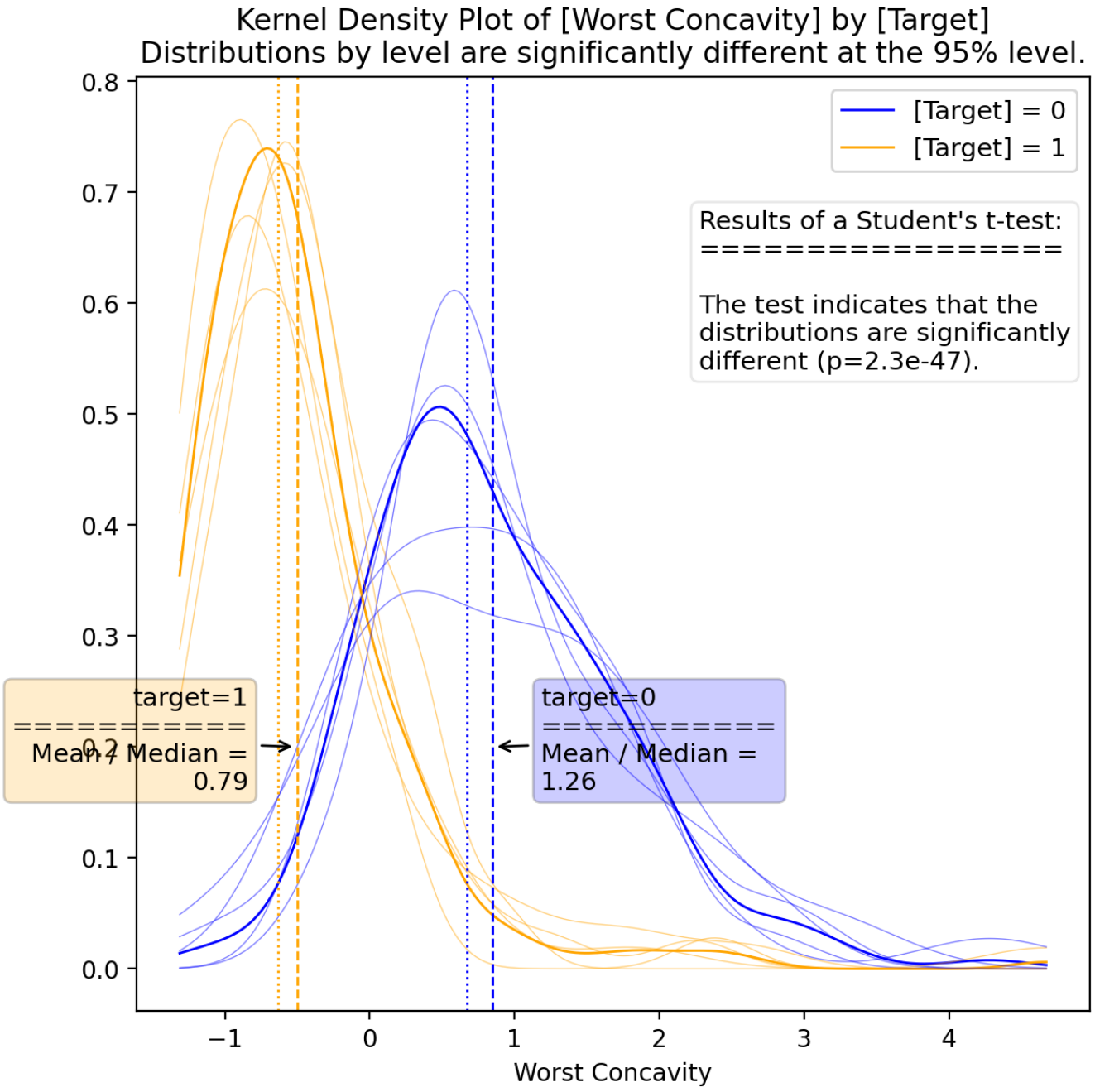
Univariate Report

Worst Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.1e+00	-2.1e+00	-2.5e+00	-2.3e+00	-2.1e+00	-2.2e+00	1.7e-01
Fitted p-Value	1.8e-22	2.2e-22	4.5e-23	4.6e-23	2.1e-22	5.9e-28	8.9e-23
Fitted Std. Err.	0.219	0.216	0.251	0.229	0.214	0.201	0.015
Conf. Int. Lower	-2.6e+00	-2.5e+00	-3.0e+00	-2.7e+00	-2.5e+00	-2.6e+00	2.0e-01
Conf. Int. Upper	-1.7e+00	-1.7e+00	-2.0e+00	-1.8e+00	-1.7e+00	-1.8e+00	1.4e-01
Train Accuracy	84.7%	85.4%	85.1%	83.8%	84.1%	84.6%	0.7%
Val Accuracy	84.2%	81.8%	82.8%	88.2%	86.7%	37.7%	2.7%
Train AUC	84.8%	85.2%	85.3%	83.7%	84.7%	84.7%	0.6%
Val AUC	84.3%	84.1%	82.7%	88.7%	84.6%	50.0%	2.2%
Train F1	87.6%	87.7%	87.7%	87.0%	86.7%	87.3%	0.5%
Test F1	86.0%	86.2%	86.0%	89.1%	89.9%	0.0%	1.9%
Train Precision	91.1%	89.4%	91.0%	90.1%	91.4%	90.6%	0.8%
Val Precision	88.5%	94.9%	89.1%	93.2%	87.5%	0.0%	3.2%
Train Recall	84.4%	86.0%	84.6%	84.0%	82.4%	84.3%	1.3%
Val Recall	83.6%	78.9%	83.1%	85.4%	92.5%	0.0%	5.0%
Train MCC	68.1%	69.8%	69.1%	65.9%	67.8%	68.1%	1.5%
Val MCC	68.1%	62.6%	64.1%	76.7%	70.9%	0.0%	5.7%
Train Log-Loss	5.51	5.26	5.38	5.84	5.72	5.55	0.24
Val Log-Loss	5.69	6.55	6.20	4.24	4.78	22.45	0.97

Univariate Report

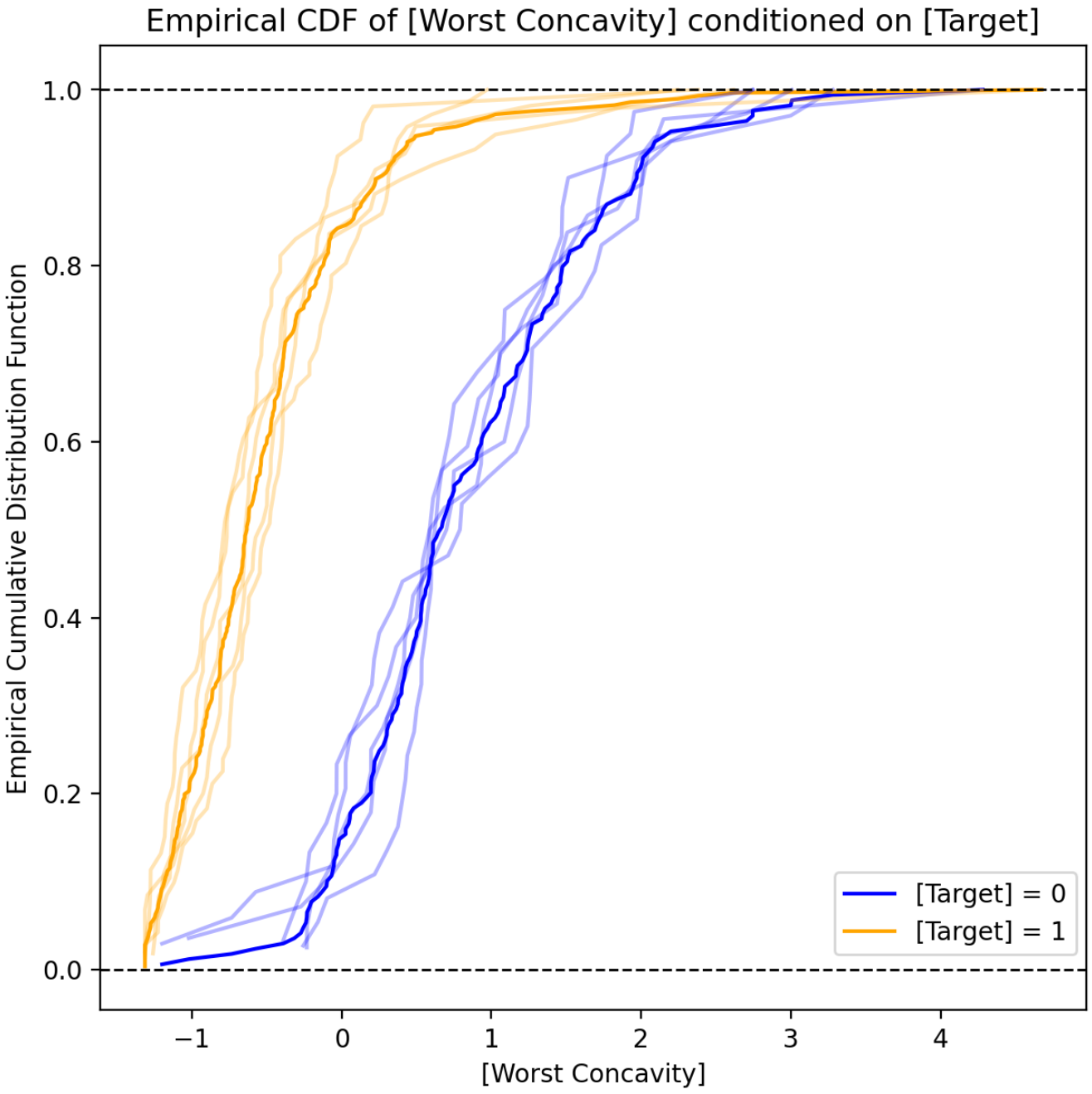
Worst Concavity - Kernel Density Plot



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

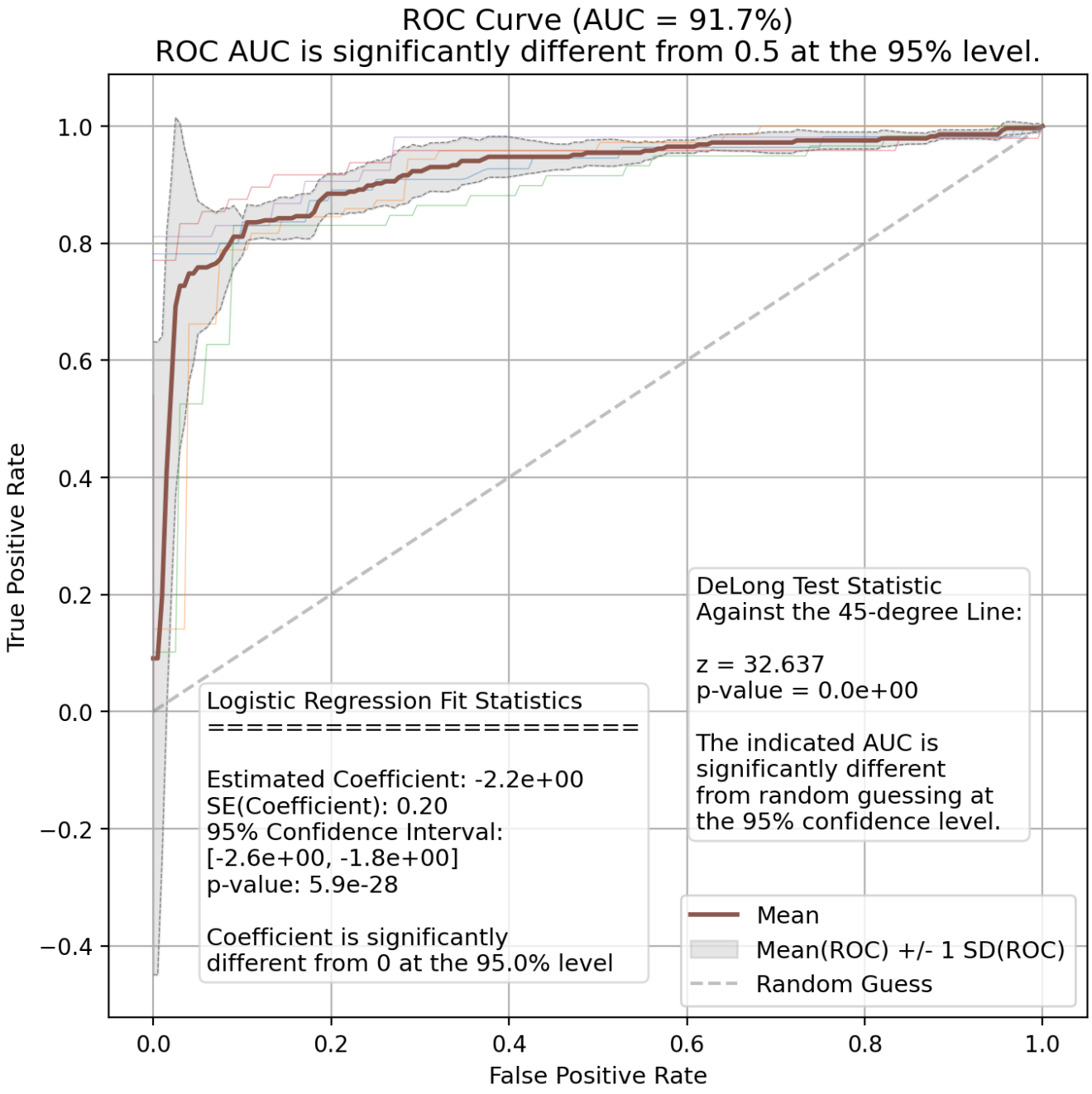
Worst Concavity - Empirical CDF Plot



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

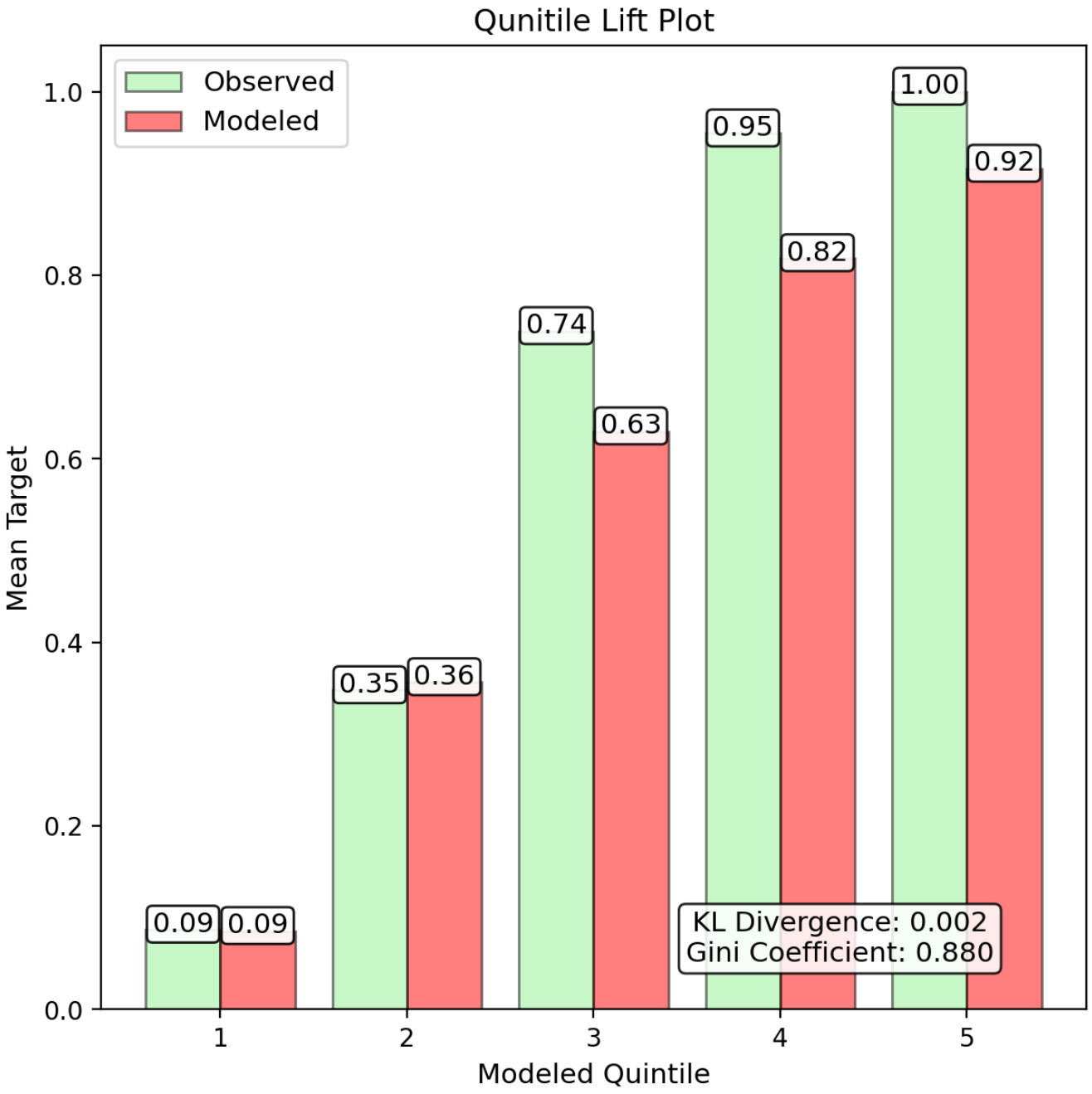


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Worst 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.

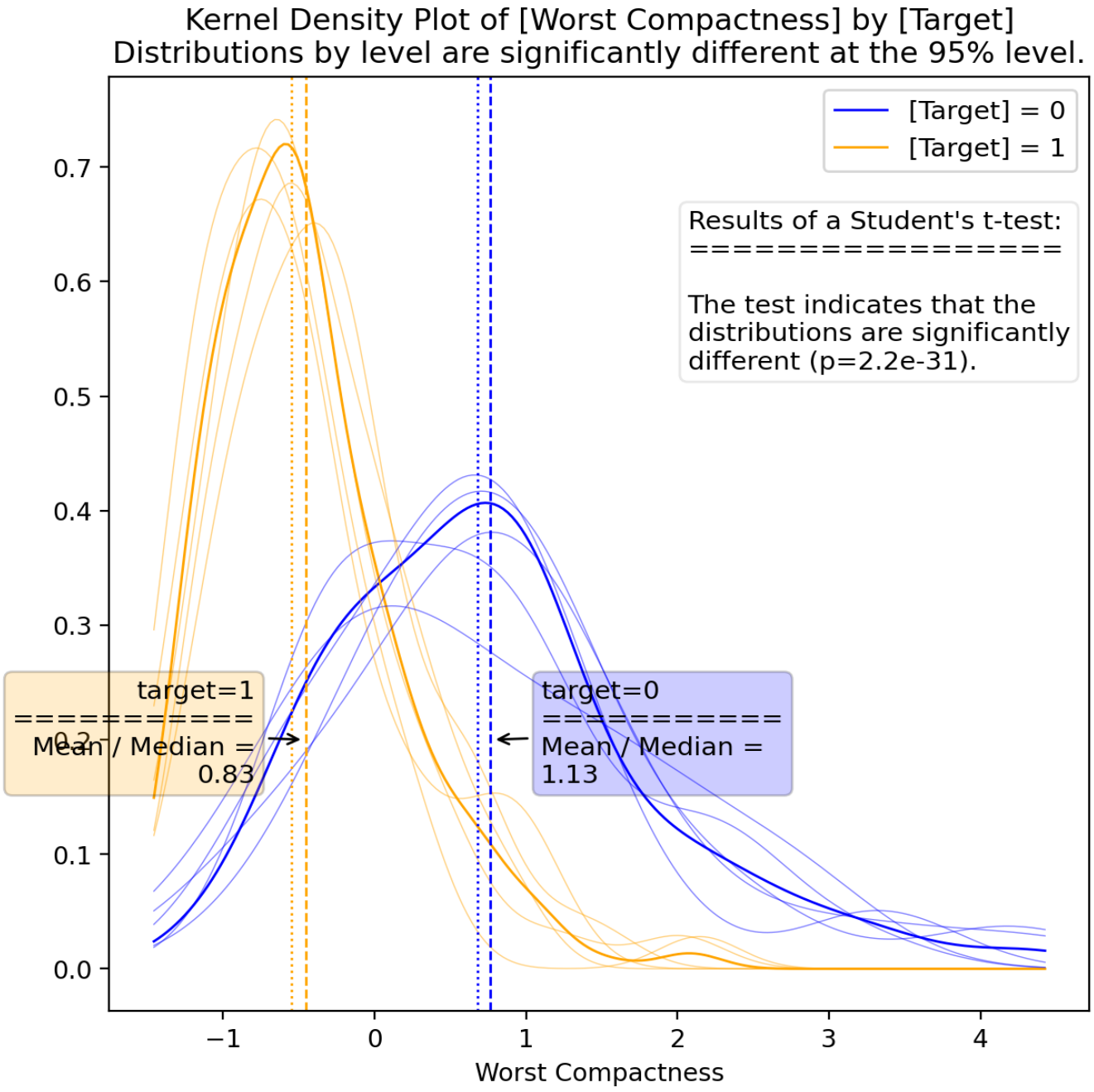
Univariate Report

Worst Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.0e+00	-1.9e+00	-2.1e+00	-1.9e+00	-1.8e+00	-1.9e+00	9.5e-02
Fitted p-Value	1.7e-21	3.3e-20	5.8e-21	1.5e-20	4.1e-20	2.0e-25	1.7e-20
Fitted Std. Err.	0.210	0.204	0.219	0.201	0.200	0.185	0.008
Conf. Int. Lower	-2.4e+00	-2.3e+00	-2.5e+00	-2.3e+00	-2.2e+00	-2.3e+00	1.1e-01
Conf. Int. Upper	-1.6e+00	-1.5e+00	-1.6e+00	-1.5e+00	-1.4e+00	-1.6e+00	8.0e-02
Train Accuracy	79.7%	78.1%	79.3%	77.3%	76.9%	78.2%	1.2%
Val Accuracy	72.6%	78.8%	74.2%	82.4%	84.3%	37.7%	5.0%
Train AUC	79.1%	77.2%	78.8%	76.1%	76.3%	77.5%	1.4%
Val AUC	71.9%	79.8%	72.2%	82.5%	82.7%	50.0%	5.4%
Train F1	83.7%	81.8%	83.0%	82.0%	81.0%	82.3%	1.1%
Test F1	76.4%	84.0%	79.7%	83.9%	87.9%	0.0%	4.4%
Train Precision	86.2%	82.2%	85.5%	83.8%	83.6%	84.2%	1.6%
Val Precision	76.4%	91.7%	79.7%	86.7%	87.0%	0.0%	6.2%
Train Recall	81.4%	81.4%	80.6%	80.3%	78.5%	80.4%	1.2%
Val Recall	76.4%	77.5%	79.7%	81.2%	88.7%	0.0%	4.9%
Train MCC	57.0%	54.3%	56.7%	51.4%	51.8%	54.2%	2.6%
Val MCC	43.9%	54.9%	44.4%	64.6%	65.8%	0.0%	10.6%
Train Log-Loss	7.31	7.90	7.47	8.18	8.33	7.84	0.44
Val Log-Loss	9.86	7.65	9.30	6.36	5.65	22.45	1.82

Univariate Report

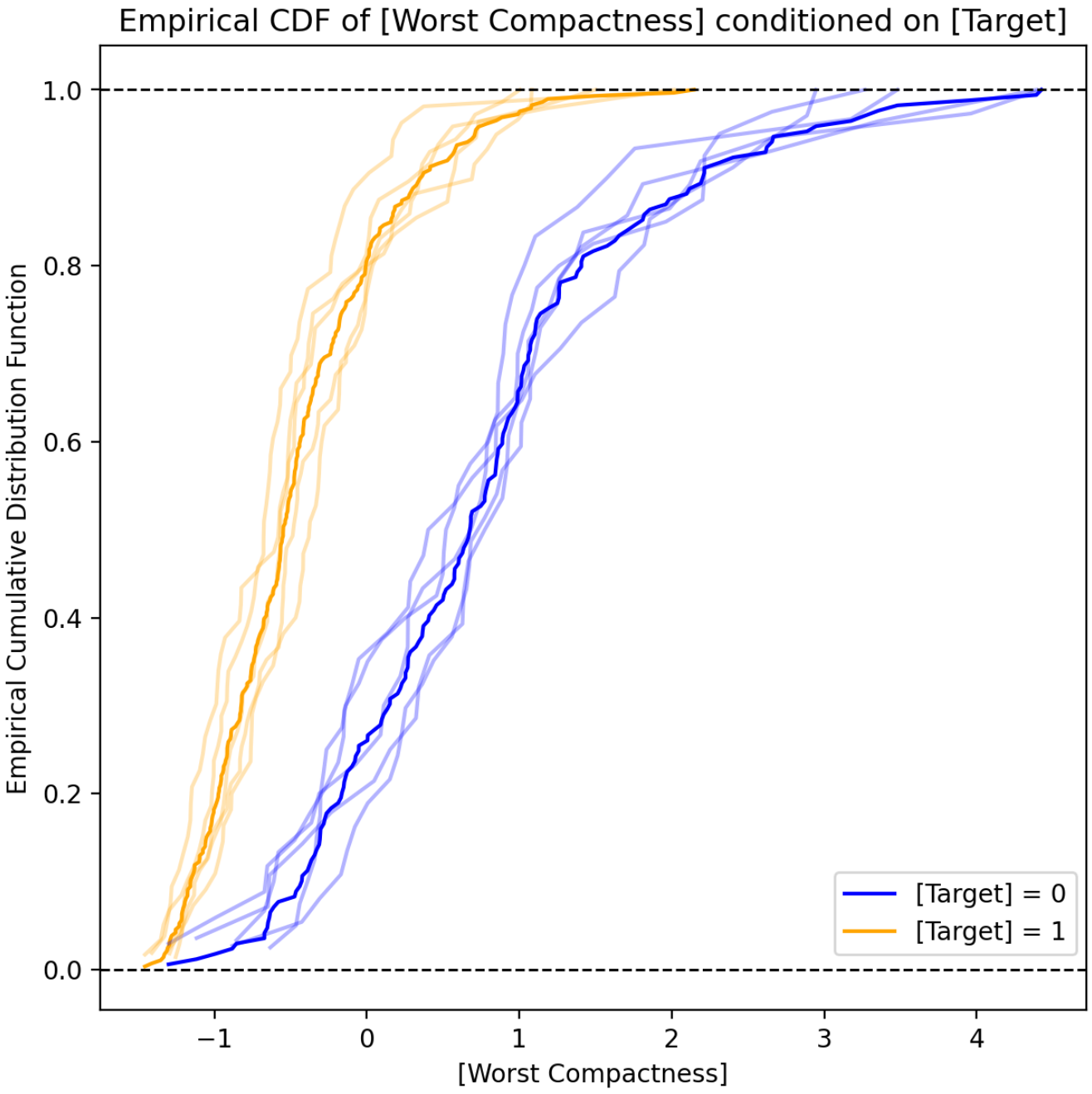
Worst Compactness - Kernel Density Plot



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

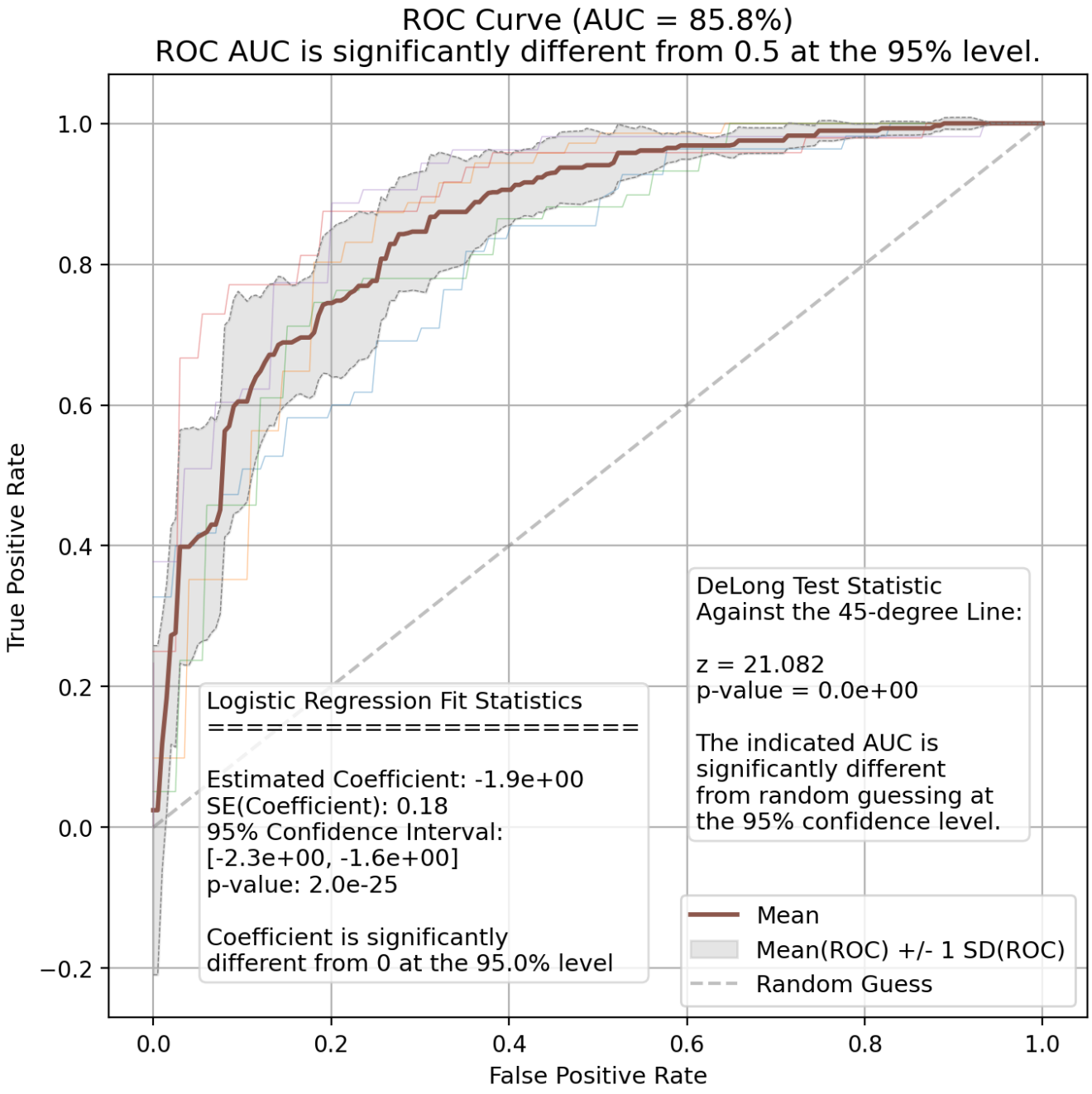
Worst Compactness - Empirical CDF Plot



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

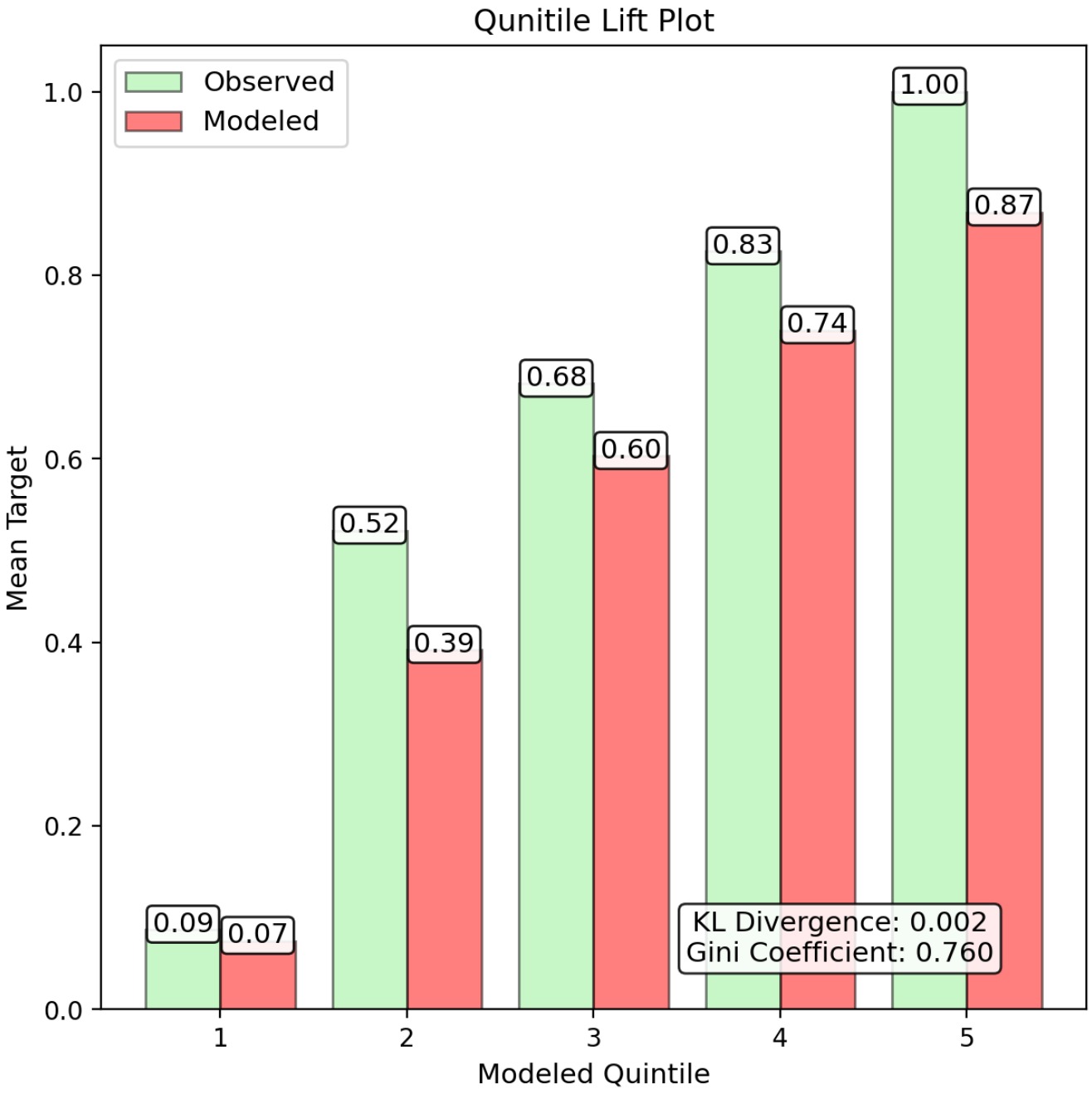


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Worst 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.

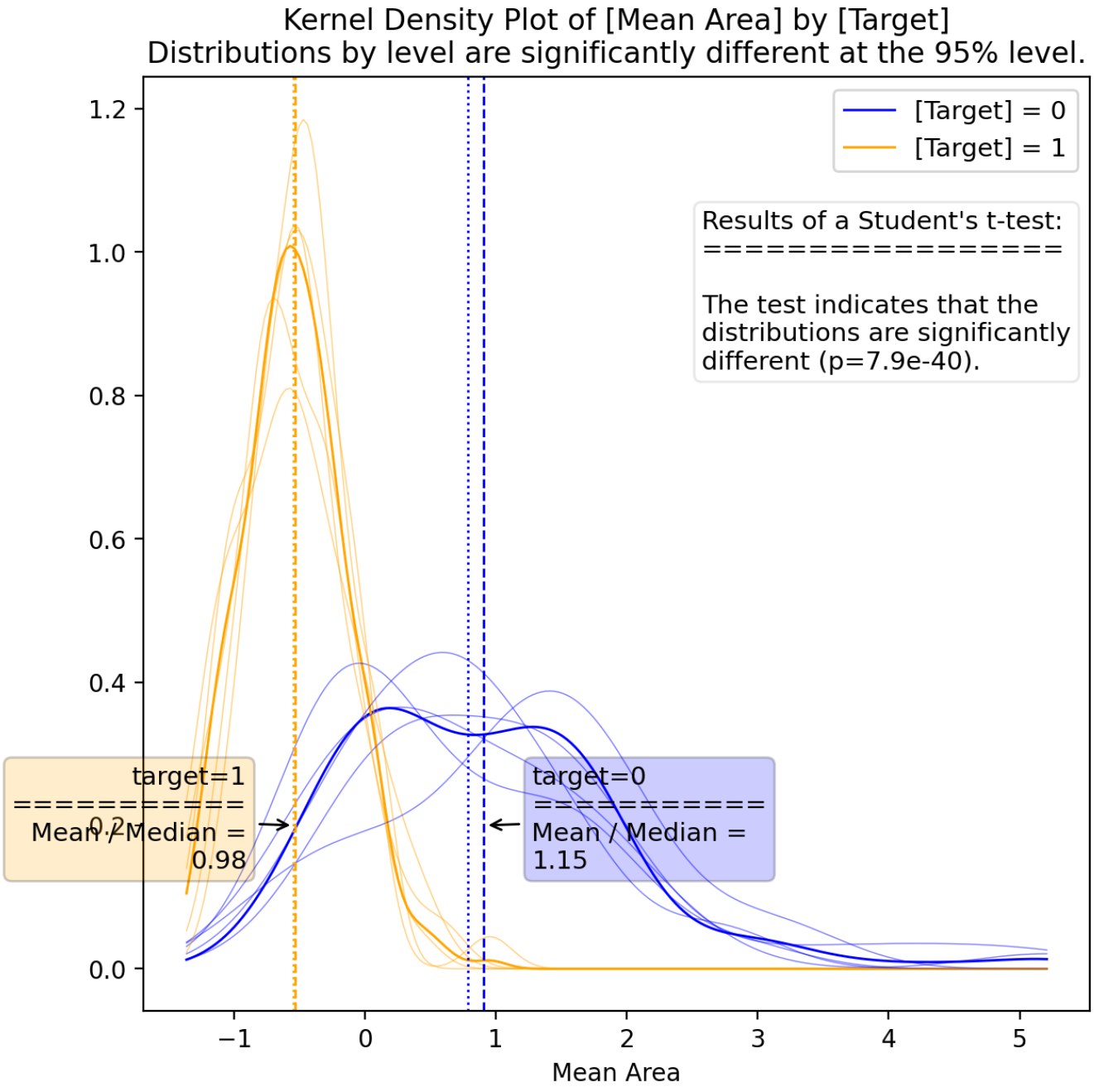
Univariate Report

Mean Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.6e+00	-3.8e+00	-3.7e+00	-3.8e+00	-3.7e+00	-3.7e+00	1.0e-01
Fitted p-Value	1.6e-22	3.6e-21	9.5e-22	1.7e-22	3.4e-22	3.8e-27	1.4e-21
Fitted Std. Err.	0.367	0.407	0.385	0.391	0.383	0.345	0.015
Conf. Int. Lower	-4.3e+00	-4.6e+00	-4.4e+00	-4.6e+00	-4.5e+00	-4.4e+00	1.3e-01
Conf. Int. Upper	-2.9e+00	-3.0e+00	-2.9e+00	-3.0e+00	-3.0e+00	-3.0e+00	7.8e-02
Train Accuracy	86.1%	87.1%	86.5%	86.5%	85.8%	86.4%	0.5%
Val Accuracy	87.4%	83.8%	86.0%	85.9%	89.2%	37.7%	2.0%
Train AUC	84.0%	86.4%	85.3%	84.8%	84.6%	85.0%	0.9%
Val AUC	87.7%	77.9%	84.0%	85.6%	87.2%	50.0%	4.0%
Train F1	89.4%	89.4%	89.3%	89.6%	88.7%	89.3%	0.3%
Test F1	88.7%	89.0%	89.3%	87.5%	91.7%	0.0%	1.6%
Train Precision	87.6%	88.9%	88.7%	88.5%	88.1%	88.4%	0.5%
Val Precision	92.2%	86.7%	87.1%	87.5%	89.3%	0.0%	2.3%
Train Recall	91.3%	89.8%	89.9%	90.8%	89.3%	90.2%	0.8%
Val Recall	85.5%	91.5%	91.5%	87.5%	94.3%	0.0%	3.6%
Train MCC	69.4%	72.9%	70.9%	70.3%	69.4%	70.6%	1.5%
Val MCC	74.7%	58.7%	69.5%	71.3%	76.2%	0.0%	6.9%
Train Log-Loss	5.01	4.66	4.88	4.87	5.14	4.91	0.18
Val Log-Loss	4.55	5.83	5.04	5.09	3.91	22.45	0.71

Univariate Report

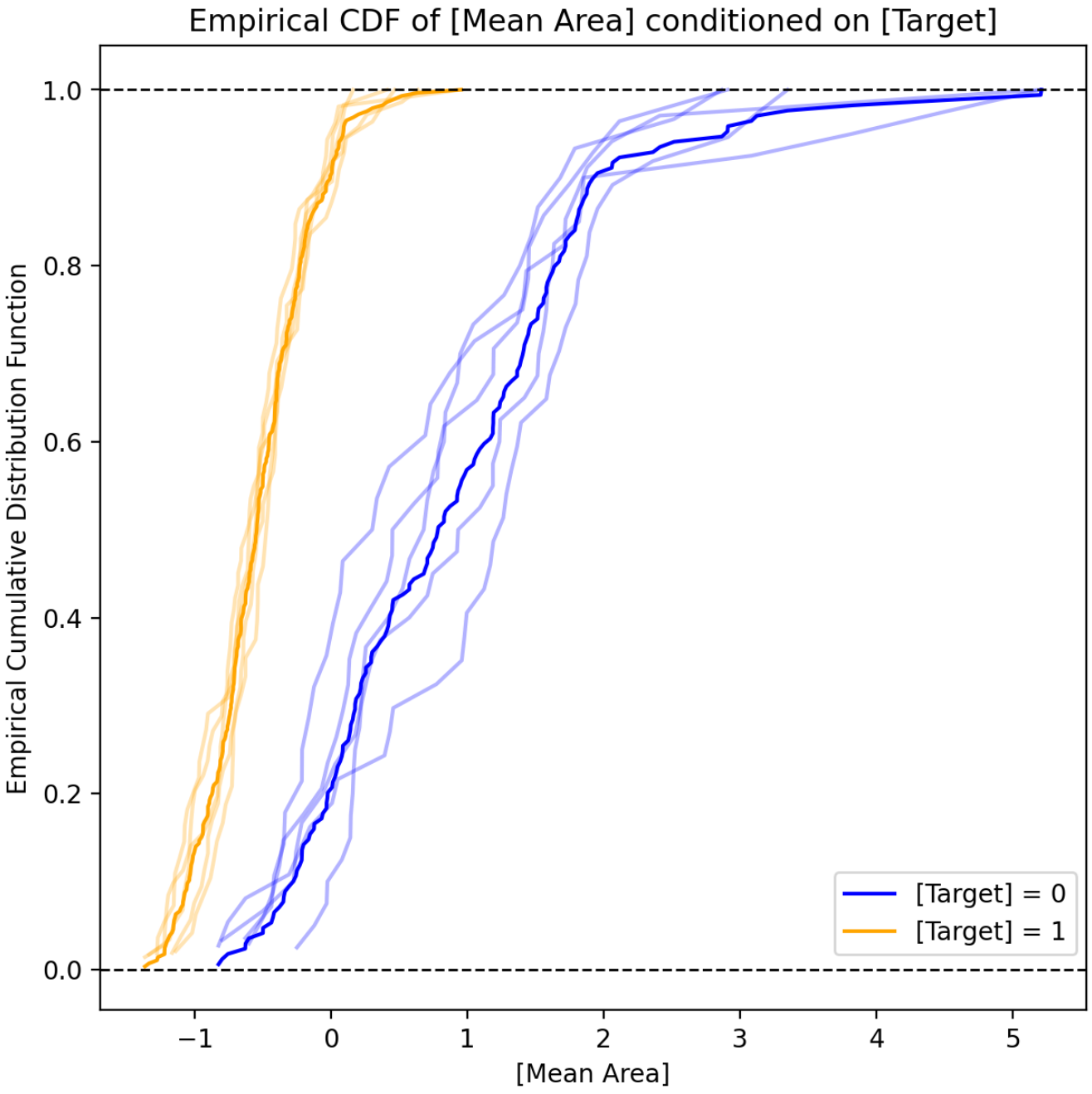
Mean Area - Kernel Density Plot



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

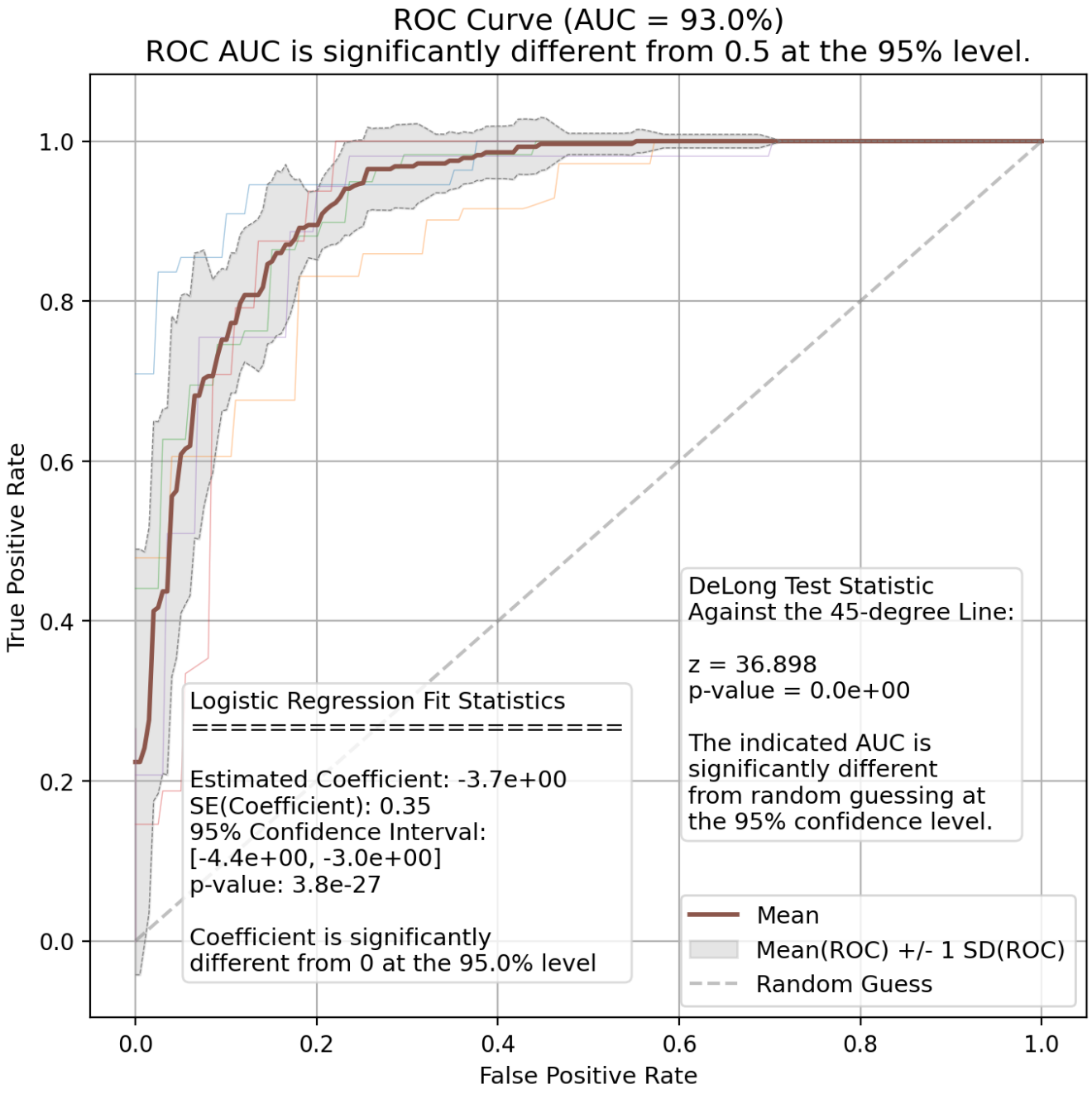
Mean Area - Empirical CDF Plot



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

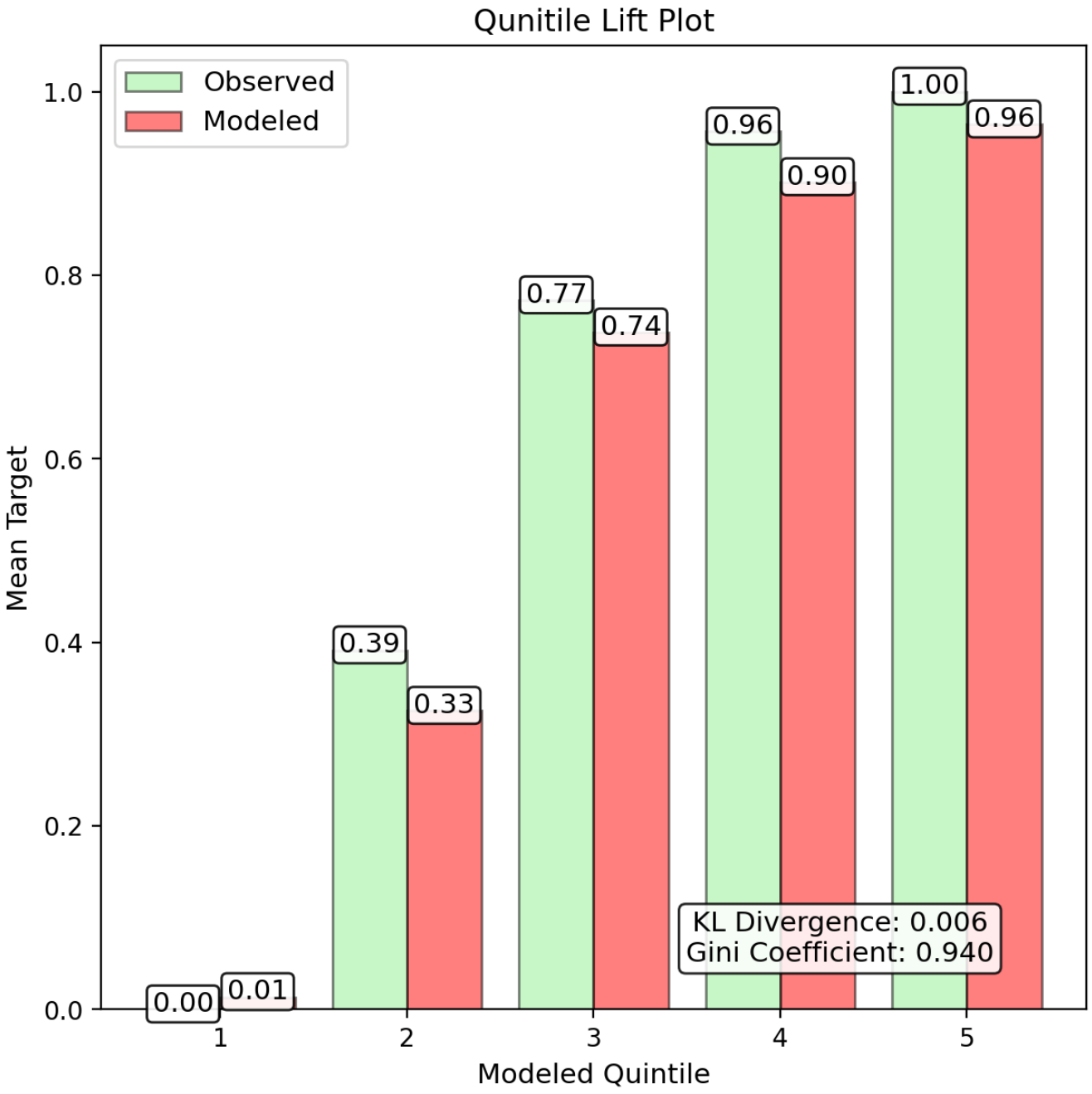


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Mean Area - Quintile Lift



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

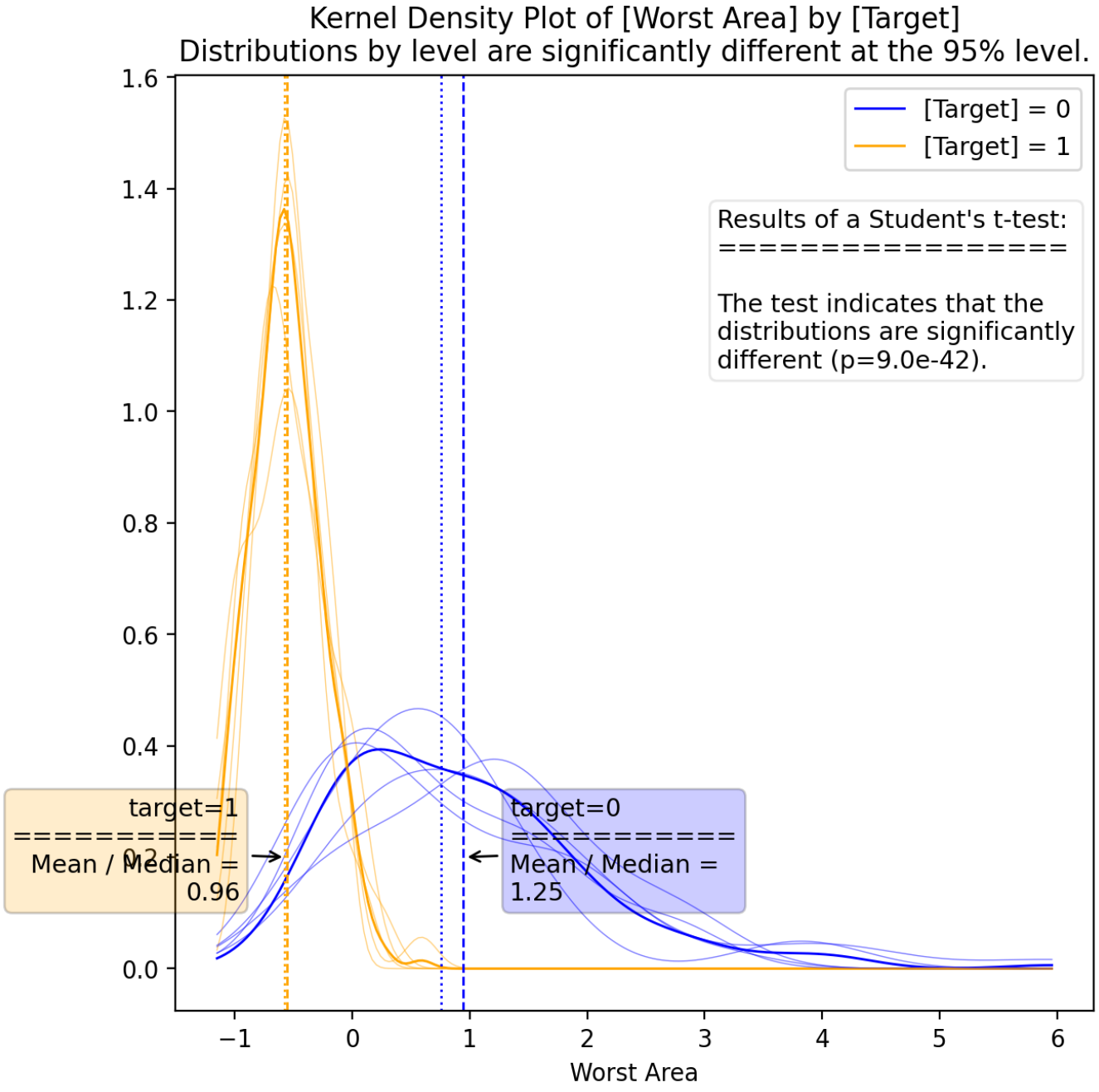
Univariate Report

Worst Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.7e+00	-5.2e+00	-4.9e+00	-5.1e+00	-5.0e+00	-4.9e+00	1.8e-01
Fitted p-Value	9.6e-23	2.0e-20	1.5e-21	8.1e-22	2.2e-21	1.3e-26	8.5e-21
Fitted Std. Err.	0.477	0.556	0.511	0.528	0.527	0.463	0.029
Conf. Int. Lower	-5.6e+00	-6.2e+00	-5.9e+00	-6.1e+00	-6.0e+00	-5.9e+00	2.4e-01
Conf. Int. Upper	-3.7e+00	-4.1e+00	-3.9e+00	-4.0e+00	-4.0e+00	-4.0e+00	1.3e-01
Train Accuracy	90.6%	92.1%	91.4%	90.8%	91.1%	91.2%	0.6%
Val Accuracy	93.7%	87.9%	90.3%	92.9%	91.6%	37.7%	2.3%
Train AUC	87.8%	90.8%	89.6%	88.6%	89.3%	89.3%	1.1%
Val AUC	93.5%	81.8%	88.0%	91.9%	89.1%	50.0%	4.5%
Train F1	93.0%	93.7%	93.4%	93.1%	93.2%	93.3%	0.3%
Test F1	94.5%	91.9%	92.7%	94.1%	93.7%	0.0%	1.1%
Train Precision	88.9%	90.5%	90.2%	90.2%	90.0%	89.9%	0.6%
Val Precision	94.5%	88.3%	89.1%	88.9%	89.7%	0.0%	2.5%
Train Recall	97.4%	97.2%	96.9%	96.2%	96.6%	96.9%	0.5%
Val Recall	94.5%	95.8%	96.6%	100.0%	98.1%	0.0%	2.1%
Train MCC	79.4%	83.6%	81.6%	79.8%	81.0%	81.1%	1.7%
Val MCC	87.0%	68.9%	79.0%	86.3%	81.8%	0.0%	7.3%
Train Log-Loss	3.40	2.83	3.09	3.31	3.20	3.17	0.22
Val Log-Loss	2.28	4.37	3.49	2.54	3.04	22.45	0.83

Univariate Report

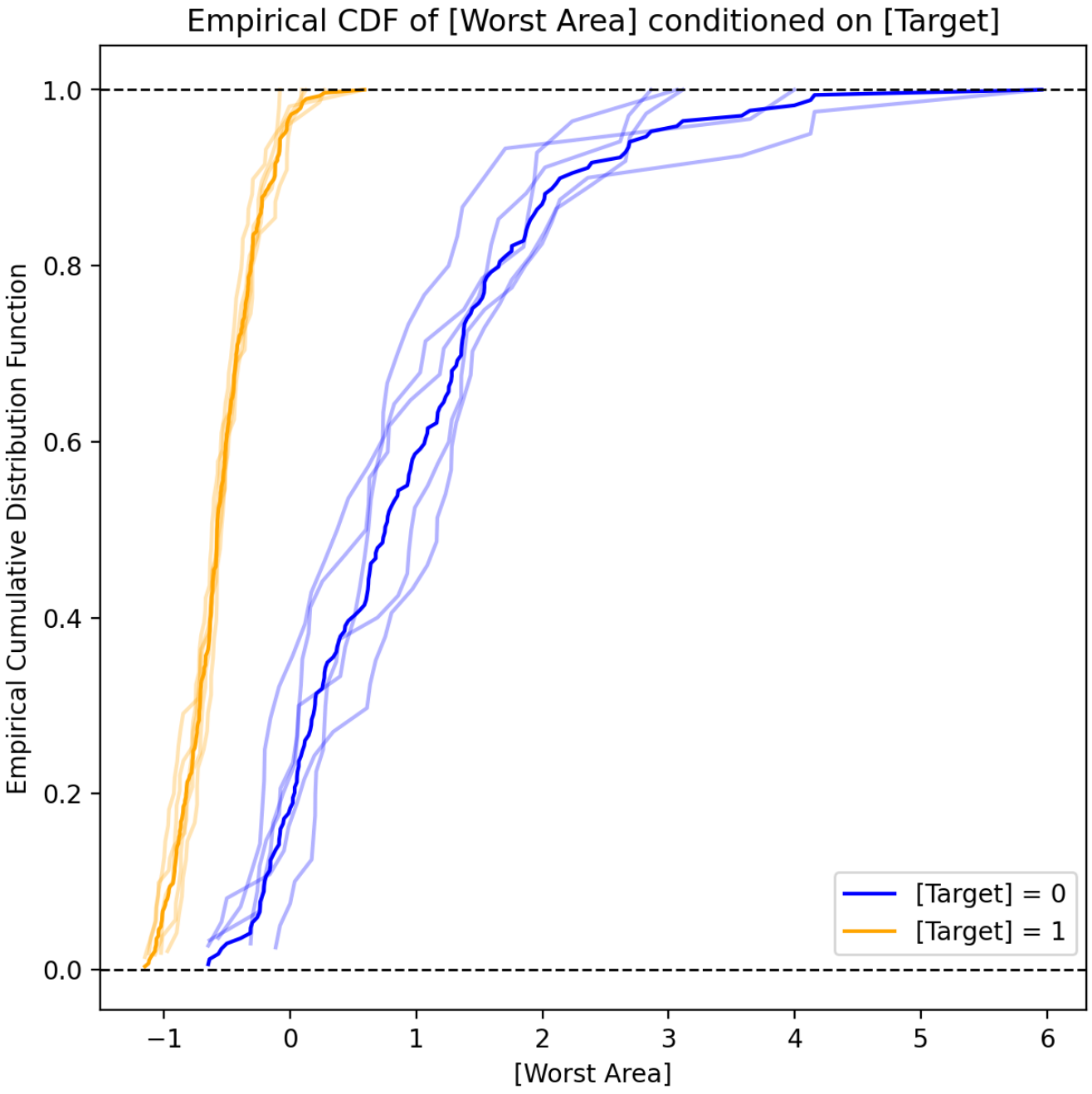
Worst Area - Kernel Density Plot



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

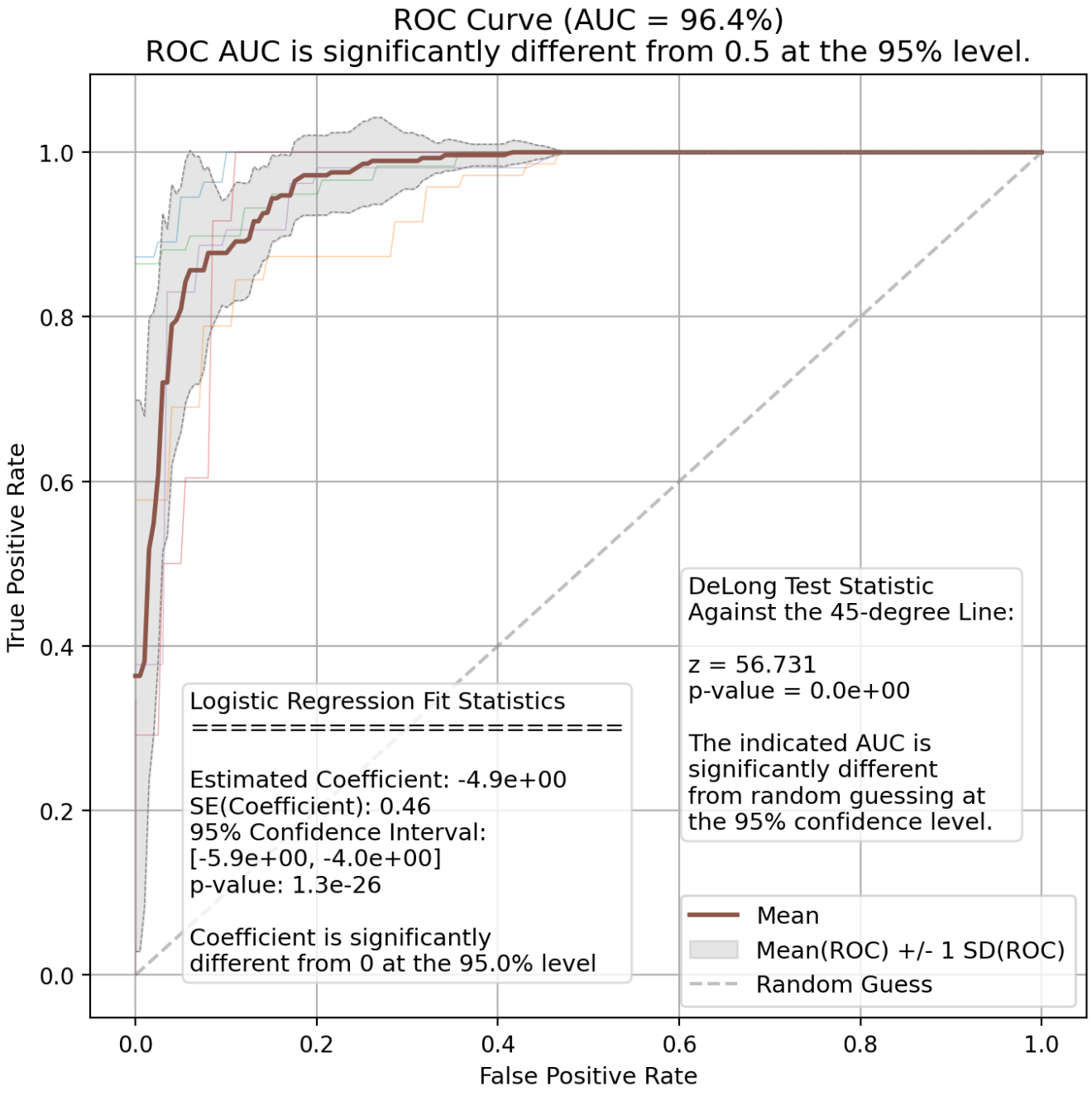
Worst Area - Empirical CDF Plot



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

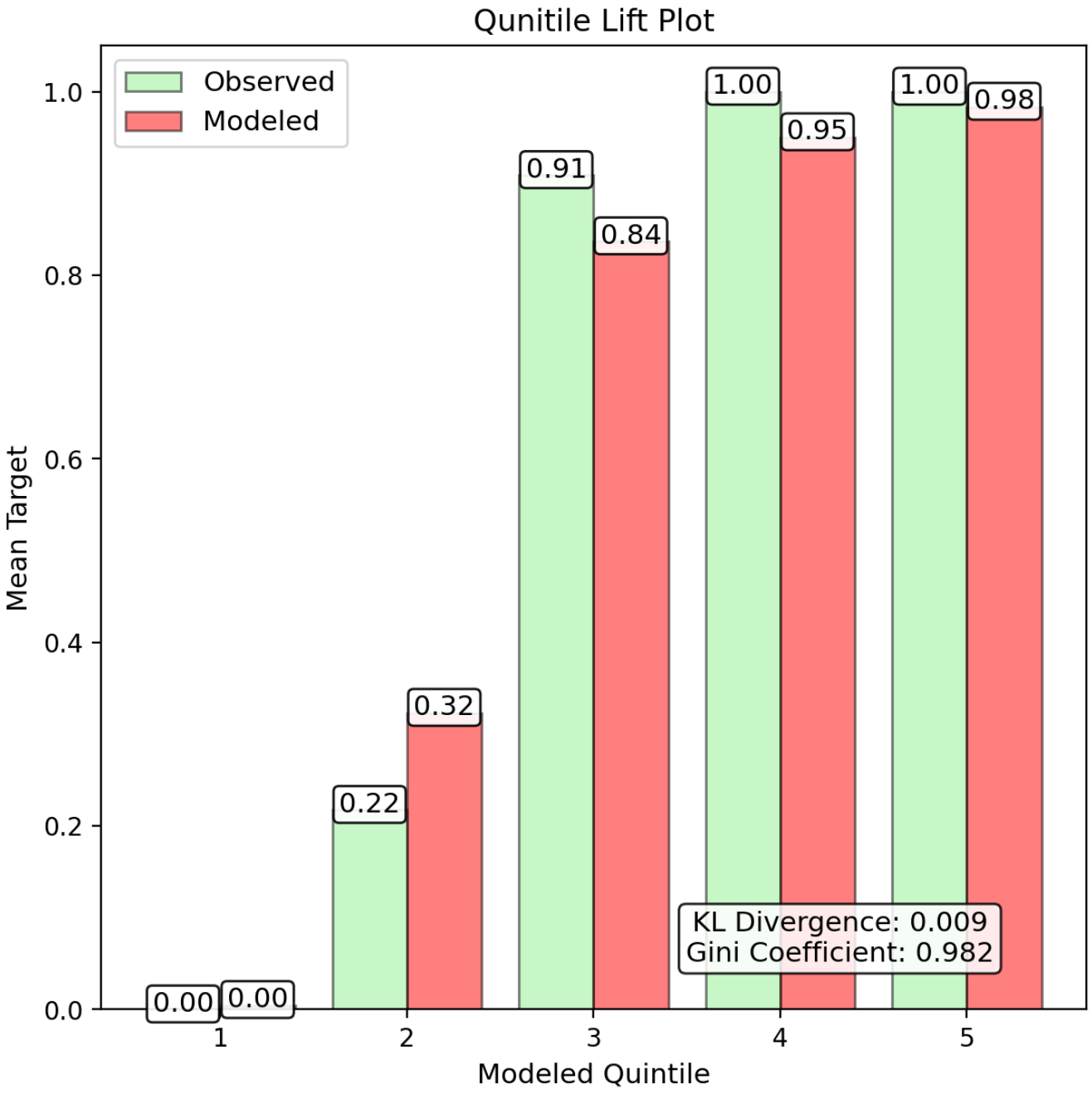


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Worst Area - Quintile Lift



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

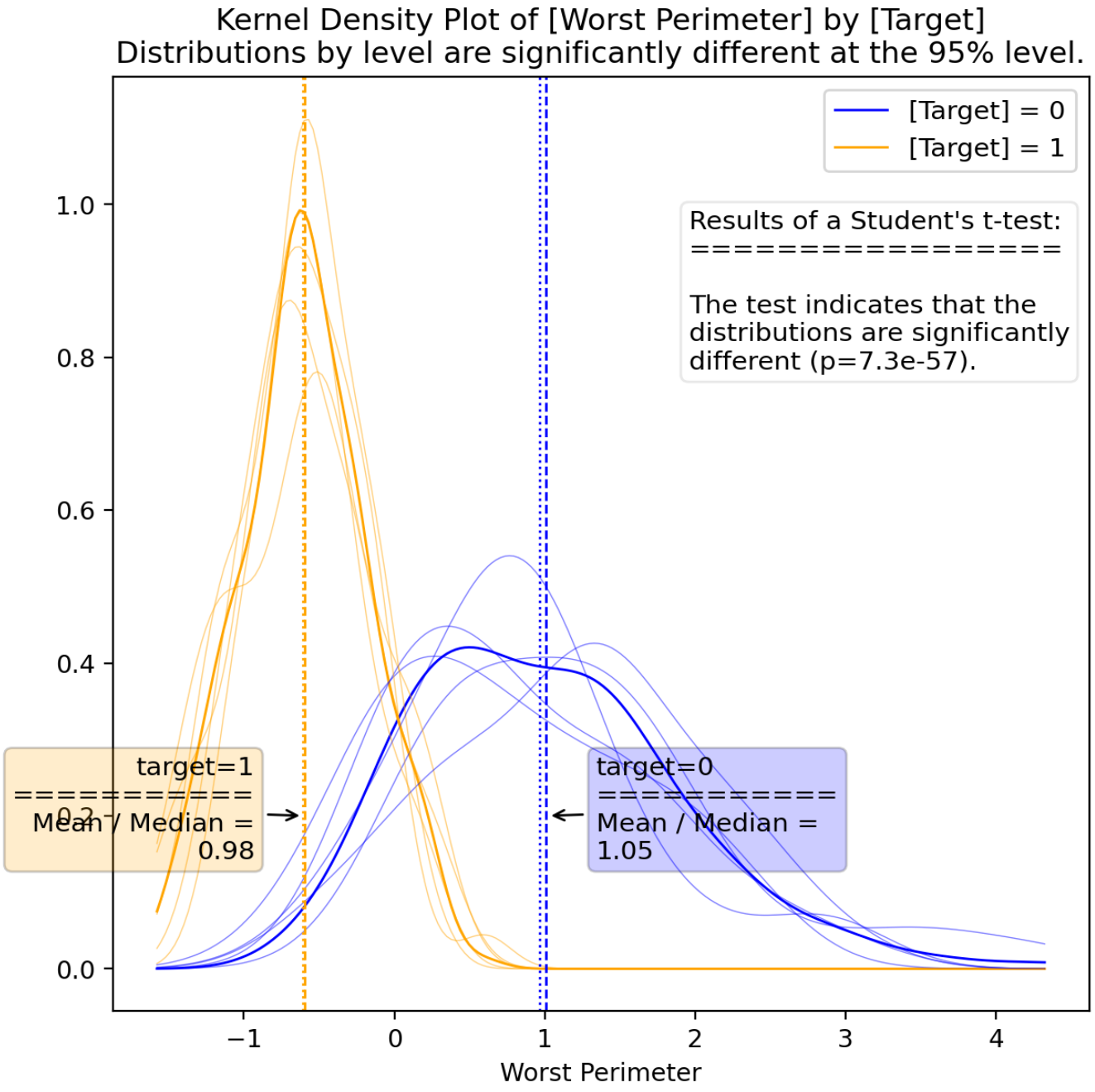
Univariate Report

Worst Perimeter - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.7e+00	-5.0e+00	-4.6e+00	-4.8e+00	-4.7e+00	-4.8e+00	1.5e-01
Fitted p-Value	1.4e-20	4.2e-19	3.8e-20	1.6e-20	3.8e-20	9.8e-25	1.8e-19
Fitted Std. Err.	0.500	0.561	0.505	0.512	0.516	0.463	0.024
Conf. Int. Lower	-5.6e+00	-6.1e+00	-5.6e+00	-5.8e+00	-5.8e+00	-5.7e+00	1.9e-01
Conf. Int. Upper	-3.7e+00	-3.9e+00	-3.7e+00	-3.7e+00	-3.7e+00	-3.8e+00	1.0e-01
Train Accuracy	90.0%	91.6%	90.6%	89.7%	89.8%	90.3%	0.8%
Val Accuracy	91.6%	85.9%	89.2%	92.9%	92.8%	37.7%	3.0%
Train AUC	88.8%	91.4%	90.4%	89.1%	89.5%	89.9%	1.1%
Val AUC	92.7%	82.6%	87.8%	92.8%	91.4%	50.0%	4.4%
Train F1	92.3%	93.0%	92.4%	91.9%	91.7%	92.3%	0.5%
Test F1	92.2%	90.1%	91.7%	93.8%	94.4%	0.0%	1.7%
Train Precision	91.5%	93.8%	93.7%	92.7%	93.0%	92.9%	0.9%
Val Precision	100.0%	90.1%	90.2%	93.8%	92.7%	0.0%	4.0%
Train Recall	93.1%	92.1%	91.2%	91.2%	90.6%	91.6%	1.0%
Val Recall	85.5%	90.1%	93.2%	93.8%	96.2%	0.0%	4.1%
Train MCC	78.1%	82.5%	80.1%	77.8%	78.4%	79.4%	2.0%
Val MCC	84.4%	65.1%	76.6%	85.6%	84.2%	0.0%	8.6%
Train Log-Loss	3.60	3.04	3.39	3.70	3.68	3.49	0.28
Val Log-Loss	3.04	5.10	3.88	2.54	2.61	22.45	1.07

Univariate Report

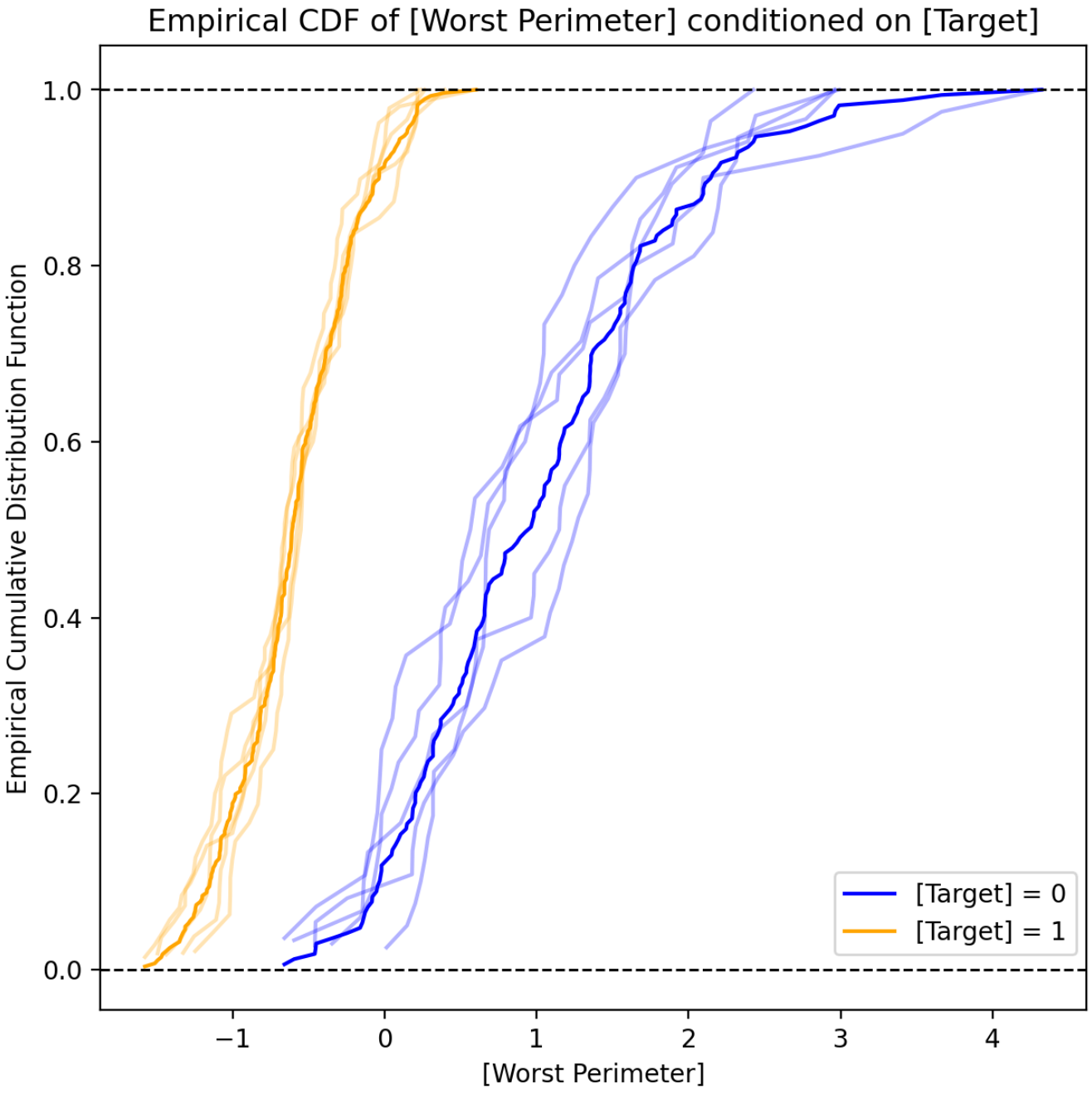
Worst Perimeter - Kernel Density Plot



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

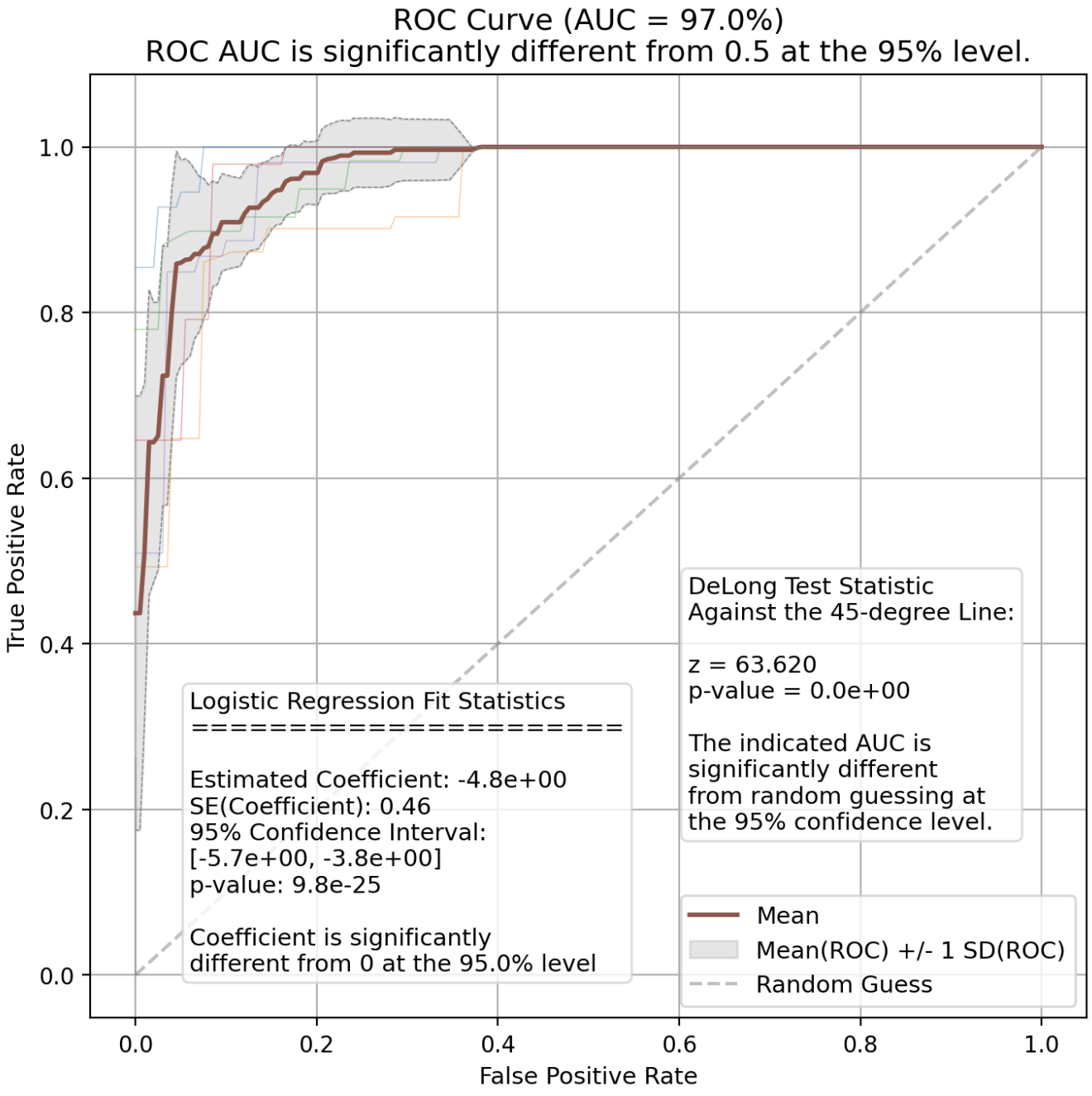
Worst Perimeter - Empirical CDF Plot



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

Univariate Report

Worst Perimeter - ROC Curve

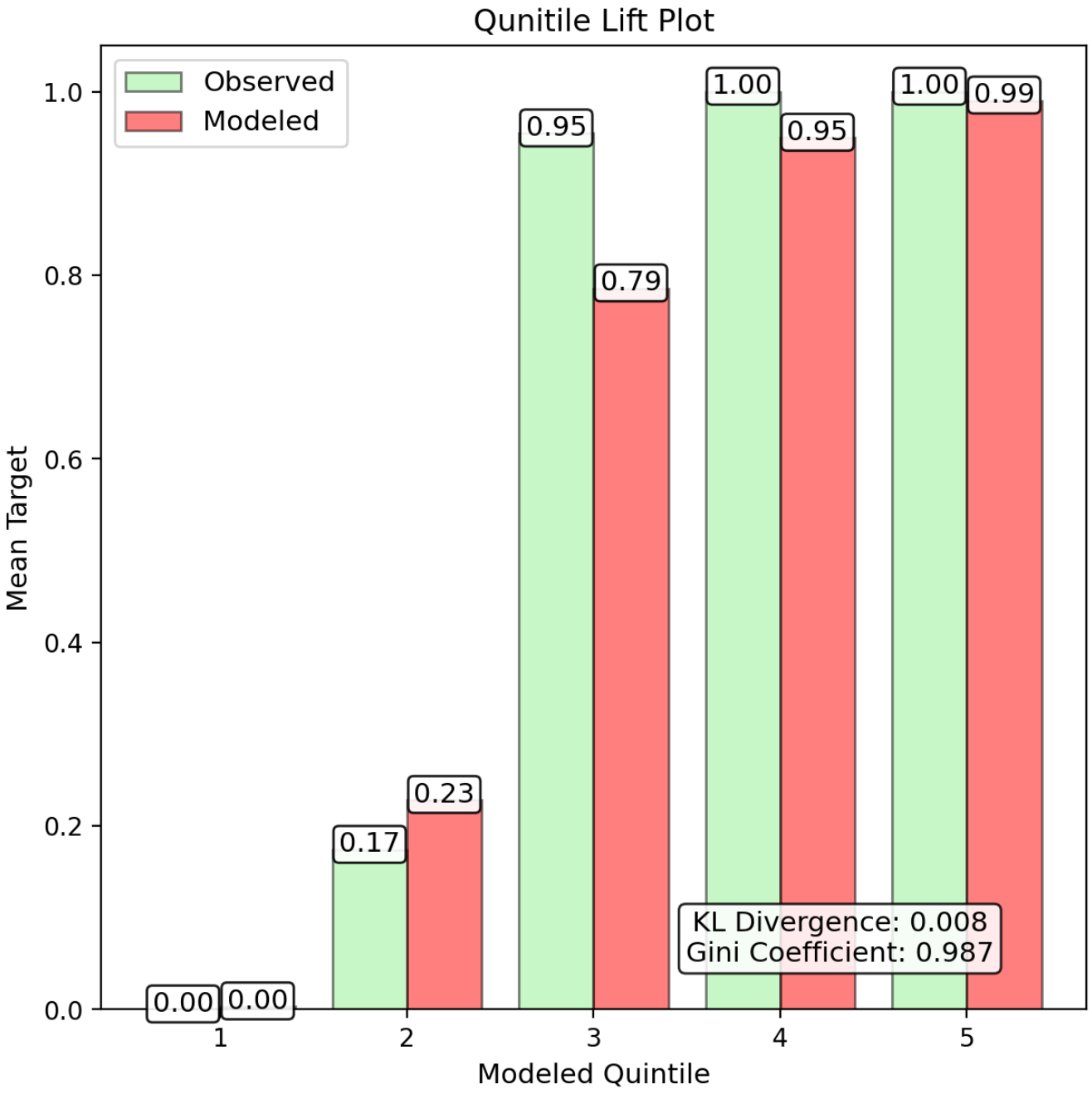


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data. Significance of the ROC curve is determined based on the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Perimeter - Quintile Lift



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

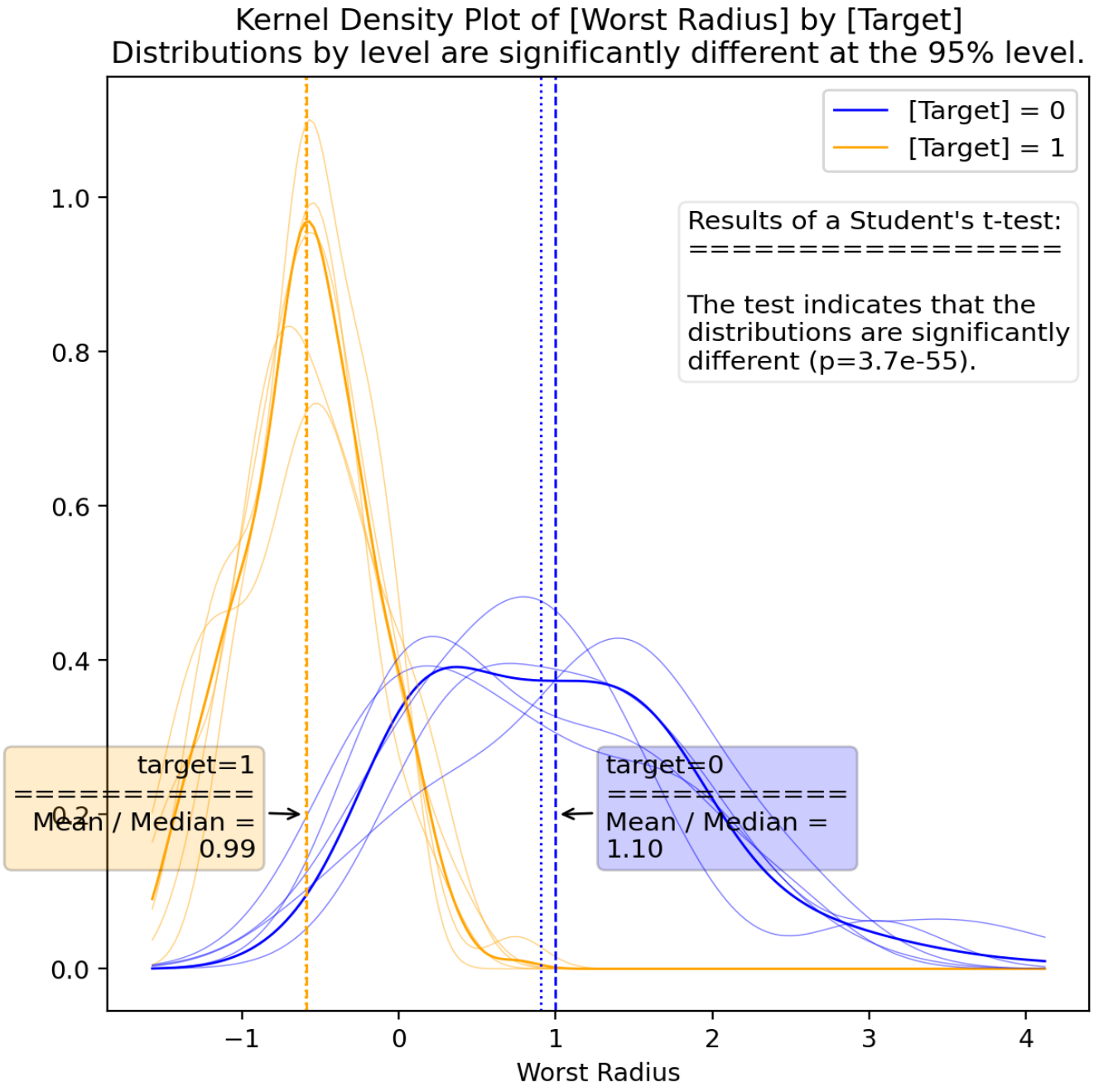
Univariate Report

Worst Radius - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.5e+00	-4.8e+00	-4.5e+00	-4.6e+00	-4.7e+00	-4.6e+00	1.5e-01
Fitted p-Value	2.2e-20	5.7e-19	6.4e-20	2.3e-20	7.3e-20	1.7e-24	2.4e-19
Fitted Std. Err.	0.484	0.542	0.490	0.498	0.513	0.451	0.023
Conf. Int. Lower	-5.4e+00	-5.9e+00	-5.4e+00	-5.6e+00	-5.7e+00	-5.5e+00	1.9e-01
Conf. Int. Upper	-3.5e+00	-3.8e+00	-3.5e+00	-3.6e+00	-3.7e+00	-3.7e+00	1.0e-01
Train Accuracy	88.6%	90.4%	88.4%	88.4%	89.2%	89.0%	0.9%
Val Accuracy	90.5%	83.8%	91.4%	91.8%	88.0%	37.7%	3.3%
Train AUC	87.4%	90.4%	88.2%	87.9%	88.9%	88.6%	1.2%
Val AUC	91.8%	80.1%	90.1%	91.5%	86.9%	50.0%	4.9%
Train F1	91.2%	92.0%	90.6%	90.8%	91.3%	91.2%	0.5%
Test F1	91.1%	88.7%	93.3%	92.8%	90.6%	0.0%	1.8%
Train Precision	90.6%	93.3%	92.2%	92.2%	92.5%	92.1%	1.0%
Val Precision	100.0%	88.7%	91.8%	91.8%	90.6%	0.0%	4.3%
Train Recall	91.8%	90.7%	89.0%	89.5%	90.1%	90.2%	1.1%
Val Recall	83.6%	88.7%	94.9%	93.8%	90.6%	0.0%	4.5%
Train MCC	75.1%	80.2%	75.6%	75.0%	77.3%	76.7%	2.2%
Val MCC	82.6%	60.2%	81.3%	83.2%	73.9%	0.0%	9.7%
Train Log-Loss	4.10	3.44	4.18	4.19	3.88	3.96	0.32
Val Log-Loss	3.41	5.83	3.10	2.97	4.34	22.45	1.19

Univariate Report

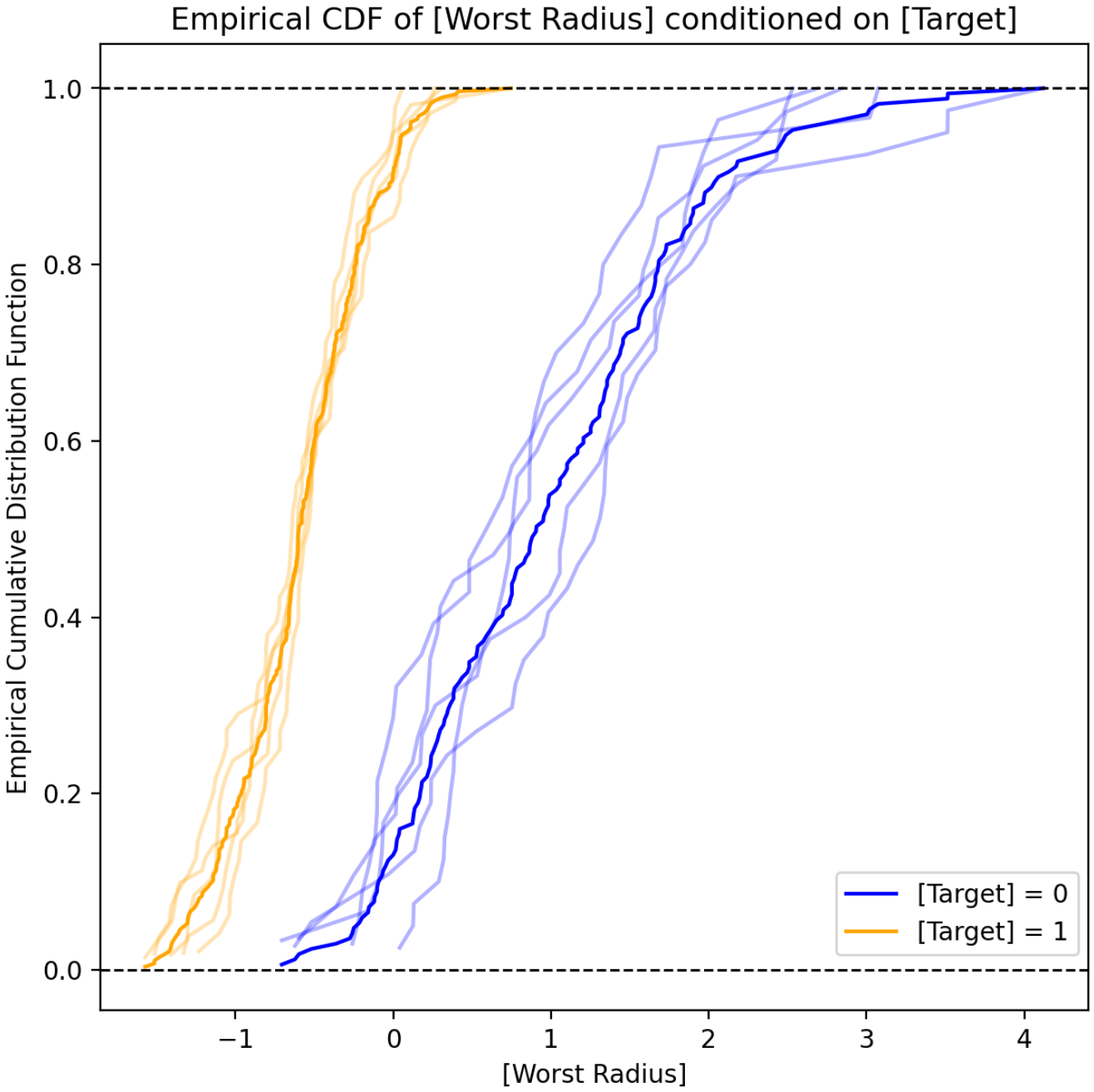
Worst Radius - Kernel Density Plot



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

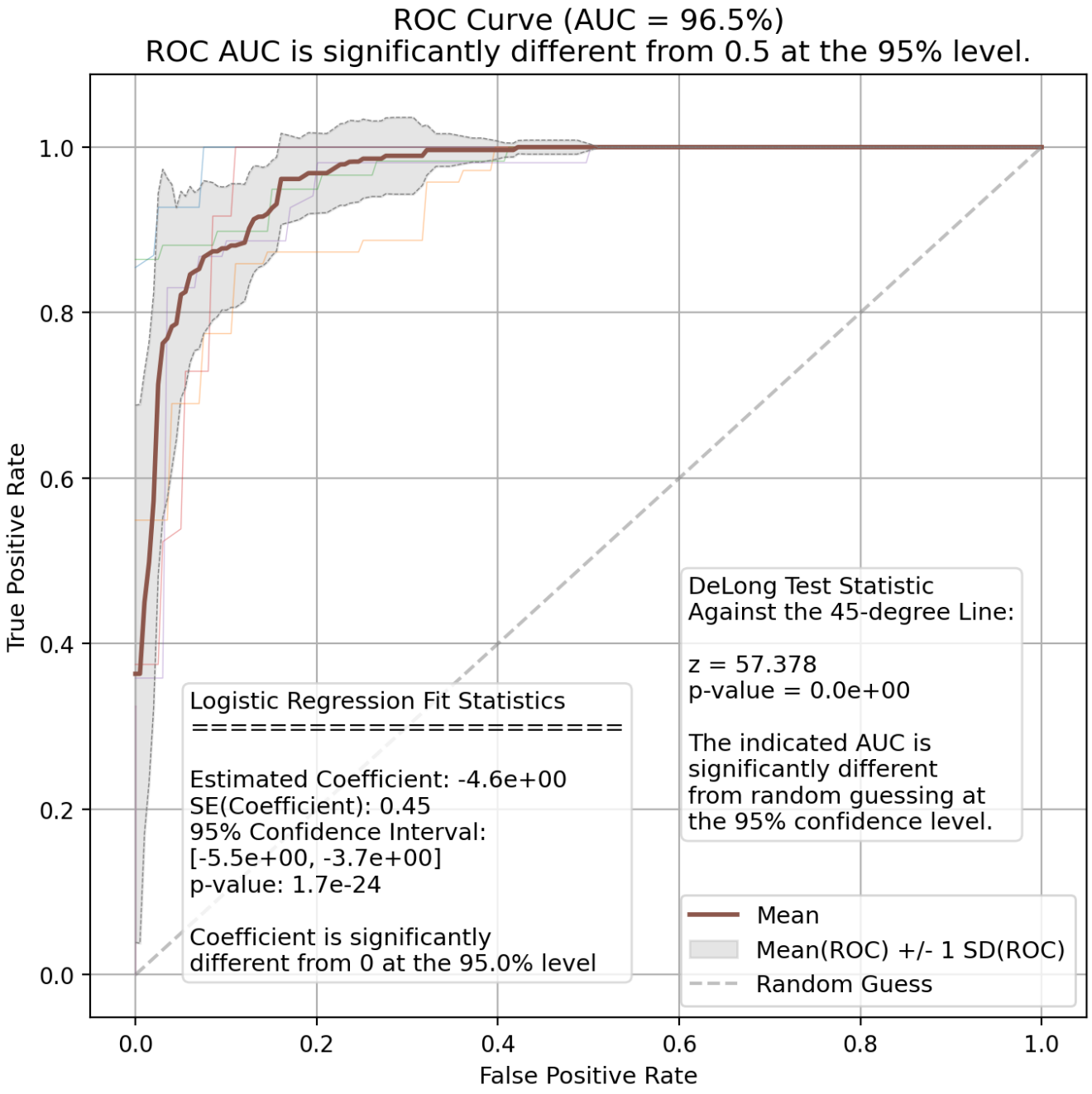
Worst Radius - Empirical CDF Plot



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 Radius - ROC Curve

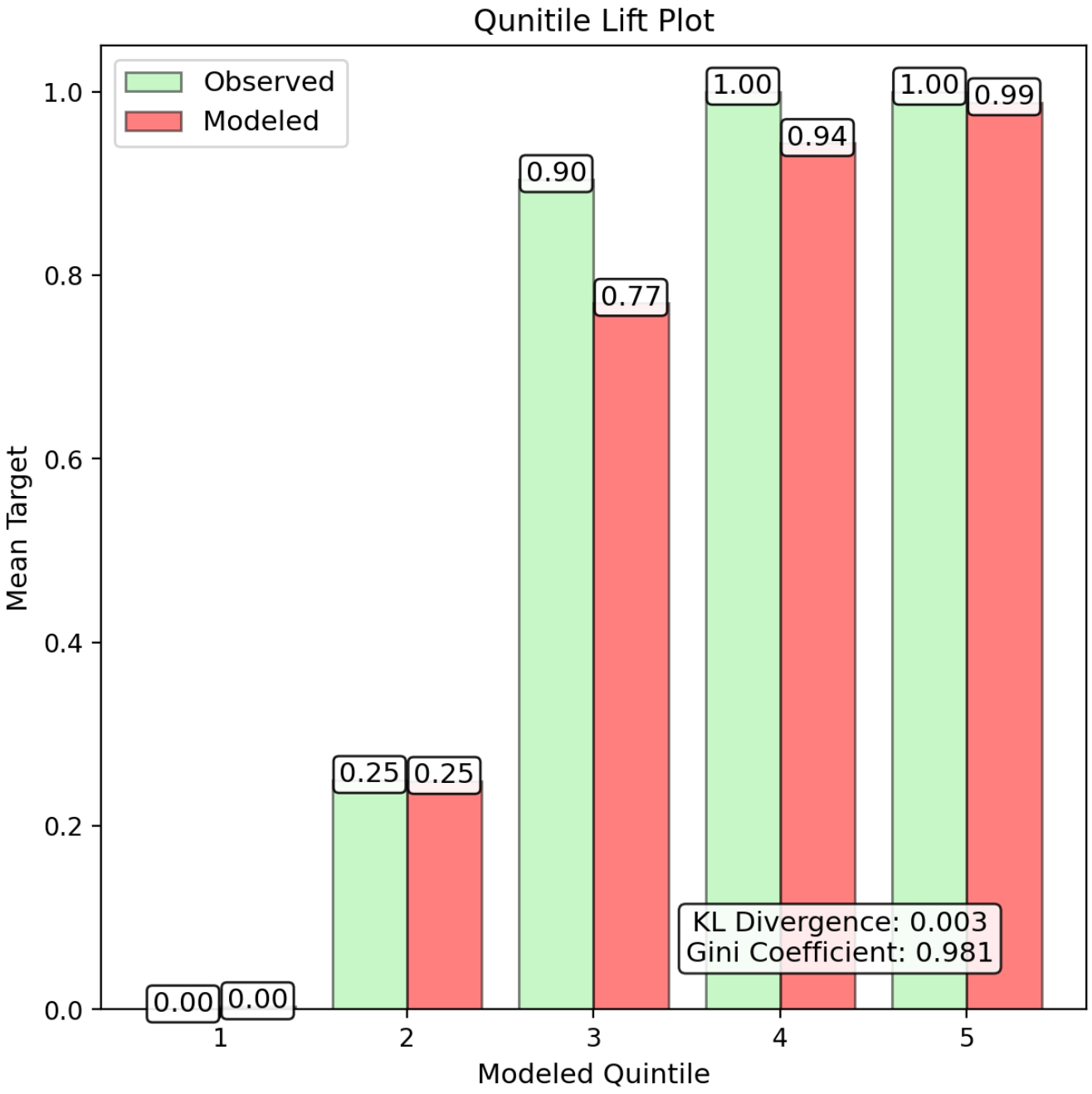


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data. Significance of the ROC curve is determined based on the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Worst Radius - 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.

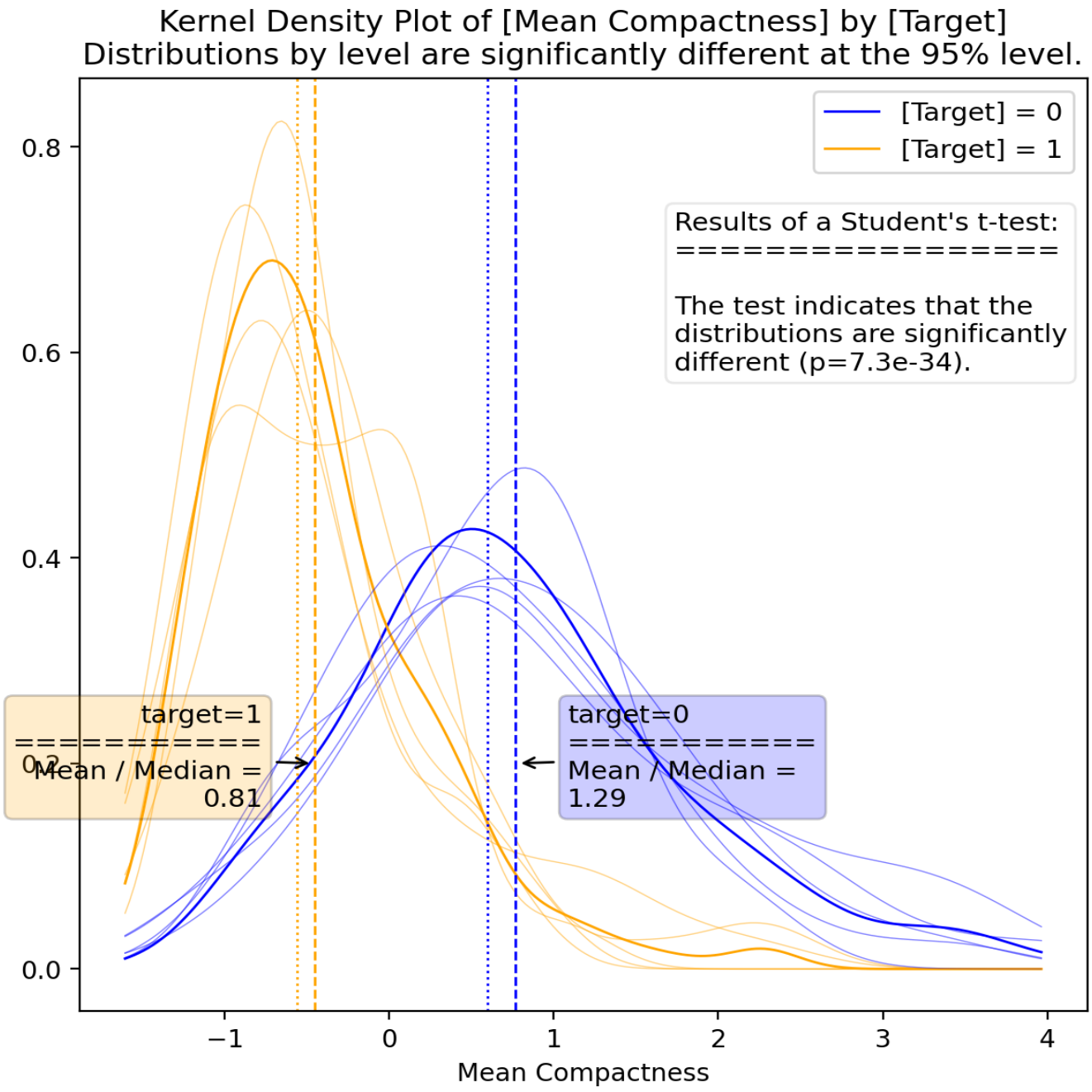
Univariate Report

Mean Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.9e+00	-1.7e+00	-2.0e+00	-1.7e+00	-1.7e+00	-1.8e+00	1.6e-01
Fitted p-Value	1.2e-21	4.5e-20	9.0e-21	7.3e-20	2.6e-19	4.9e-25	1.1e-19
Fitted Std. Err.	0.203	0.188	0.214	0.183	0.186	0.173	0.013
Conf. Int. Lower	-2.3e+00	-2.1e+00	-2.4e+00	-2.0e+00	-2.0e+00	-2.1e+00	1.8e-01
Conf. Int. Upper	-1.5e+00	-1.4e+00	-1.6e+00	-1.3e+00	-1.3e+00	-1.5e+00	1.3e-01
Train Accuracy	80.6%	80.6%	79.6%	78.4%	78.8%	79.6%	1.0%
Val Accuracy	75.8%	75.8%	79.6%	84.7%	83.1%	37.7%	4.1%
Train AUC	81.1%	80.5%	79.6%	78.3%	79.3%	79.7%	1.1%
Val AUC	75.3%	77.7%	80.2%	85.2%	81.7%	50.0%	3.8%
Train F1	83.9%	83.5%	82.9%	82.4%	82.0%	82.9%	0.8%
Test F1	78.9%	81.2%	82.9%	85.7%	86.8%	0.0%	3.2%
Train Precision	89.3%	86.1%	87.0%	86.6%	87.4%	87.3%	1.2%
Val Precision	79.6%	91.2%	88.5%	90.7%	86.8%	0.0%	4.7%
Train Recall	79.2%	80.9%	79.3%	78.6%	77.3%	79.0%	1.3%
Val Recall	78.2%	73.2%	78.0%	81.2%	86.8%	0.0%	5.0%
Train MCC	60.2%	60.3%	57.9%	55.0%	57.0%	58.1%	2.2%
Val MCC	50.5%	50.5%	58.5%	69.9%	63.5%	0.0%	8.4%
Train Log-Loss	7.01	6.99	7.37	7.79	7.65	7.37	0.37
Val Log-Loss	8.73	8.74	7.36	5.51	6.08	22.45	1.48

Univariate Report

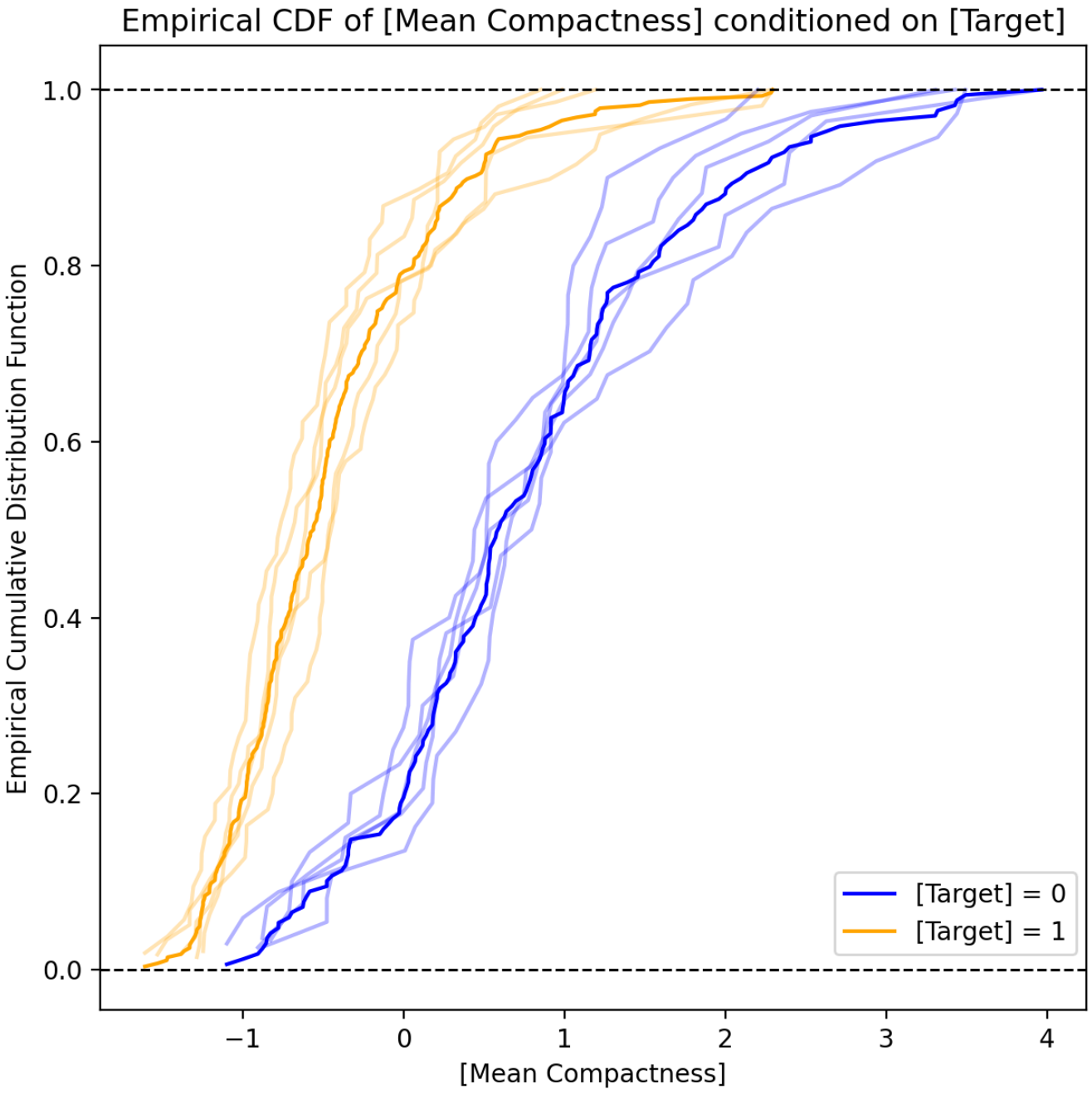
Mean Compactness - Kernel Density Plot



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

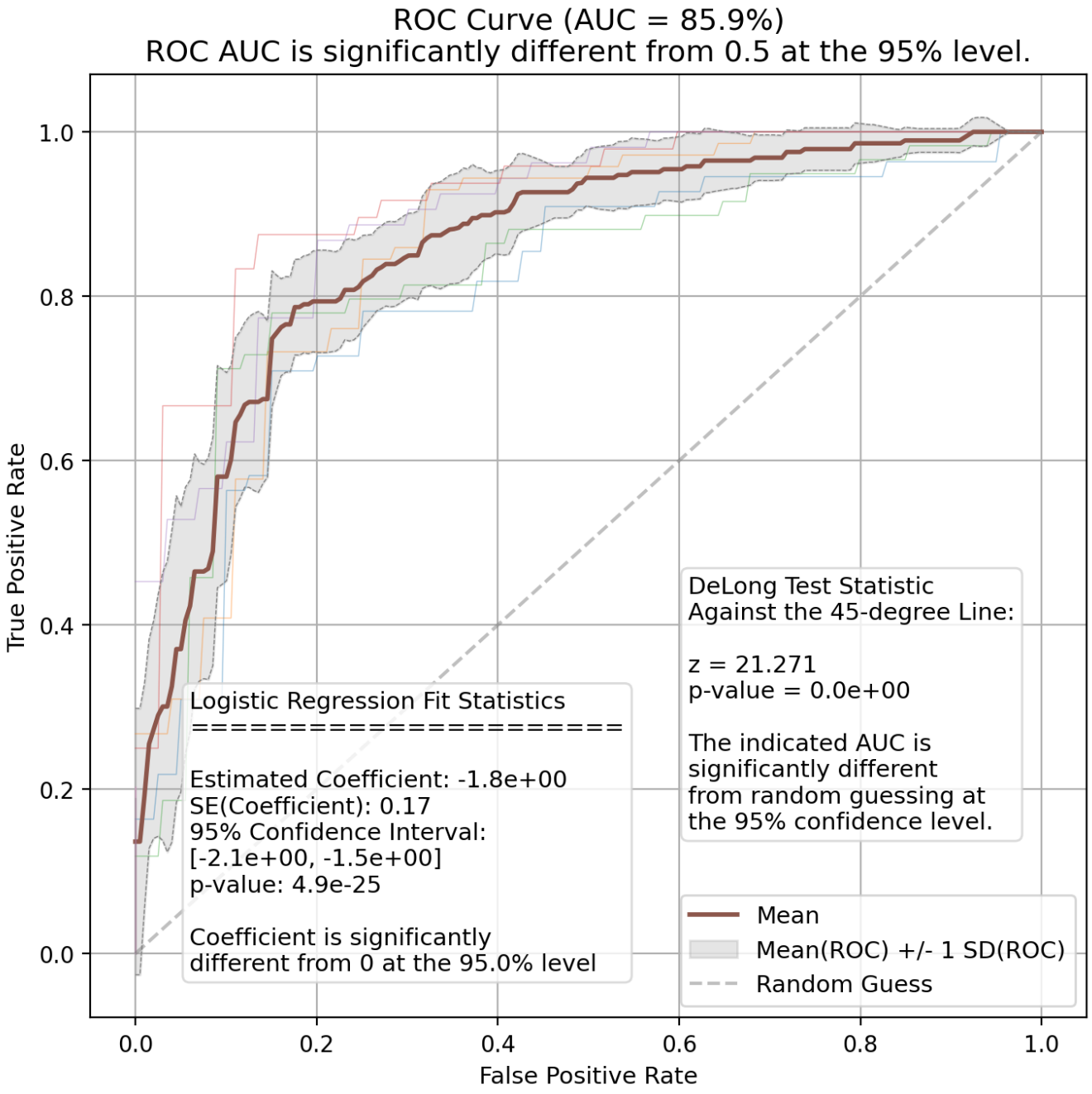
Mean Compactness - Empirical CDF Plot



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

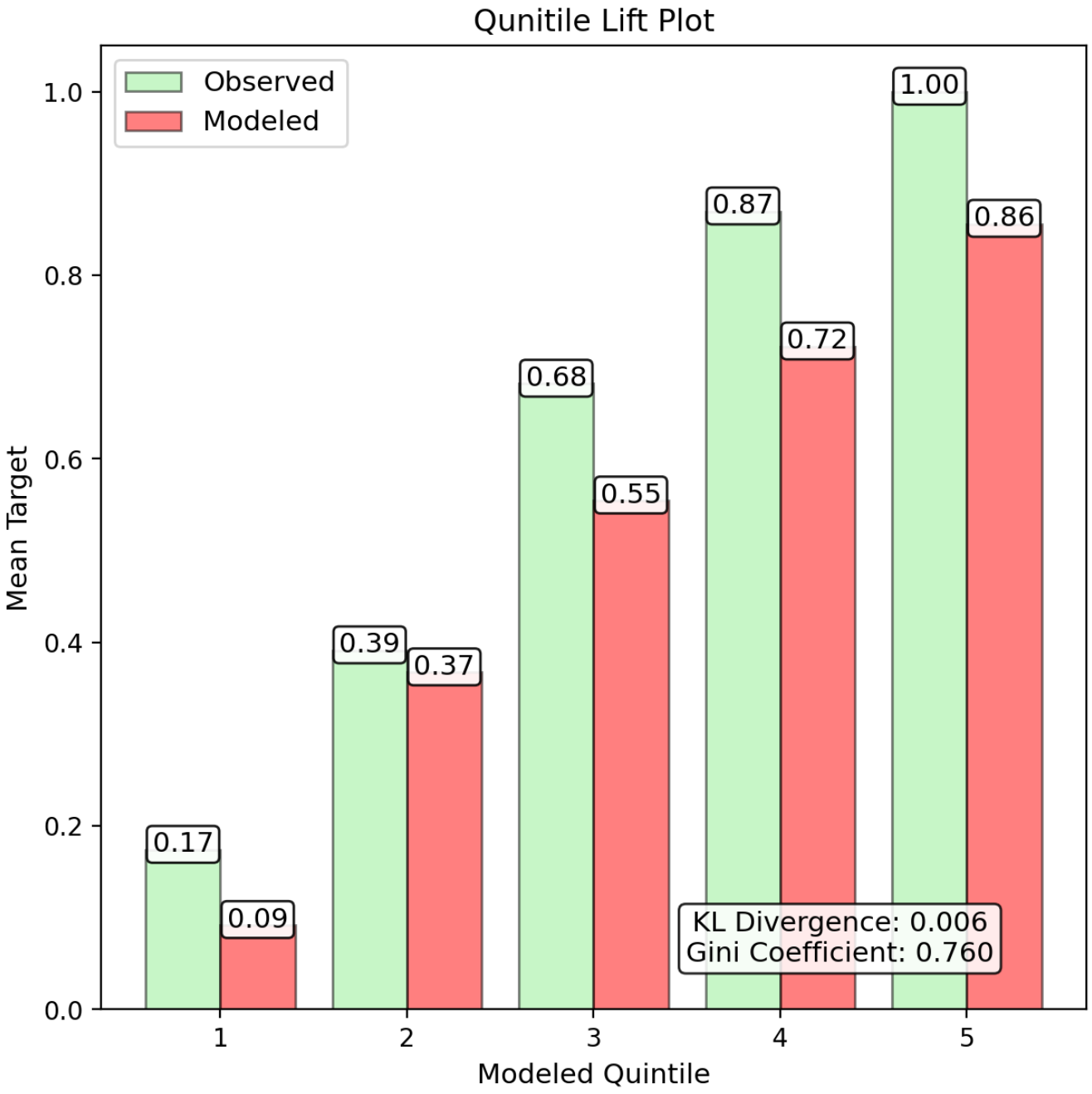


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Mean Compactness - Quintile Lift



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

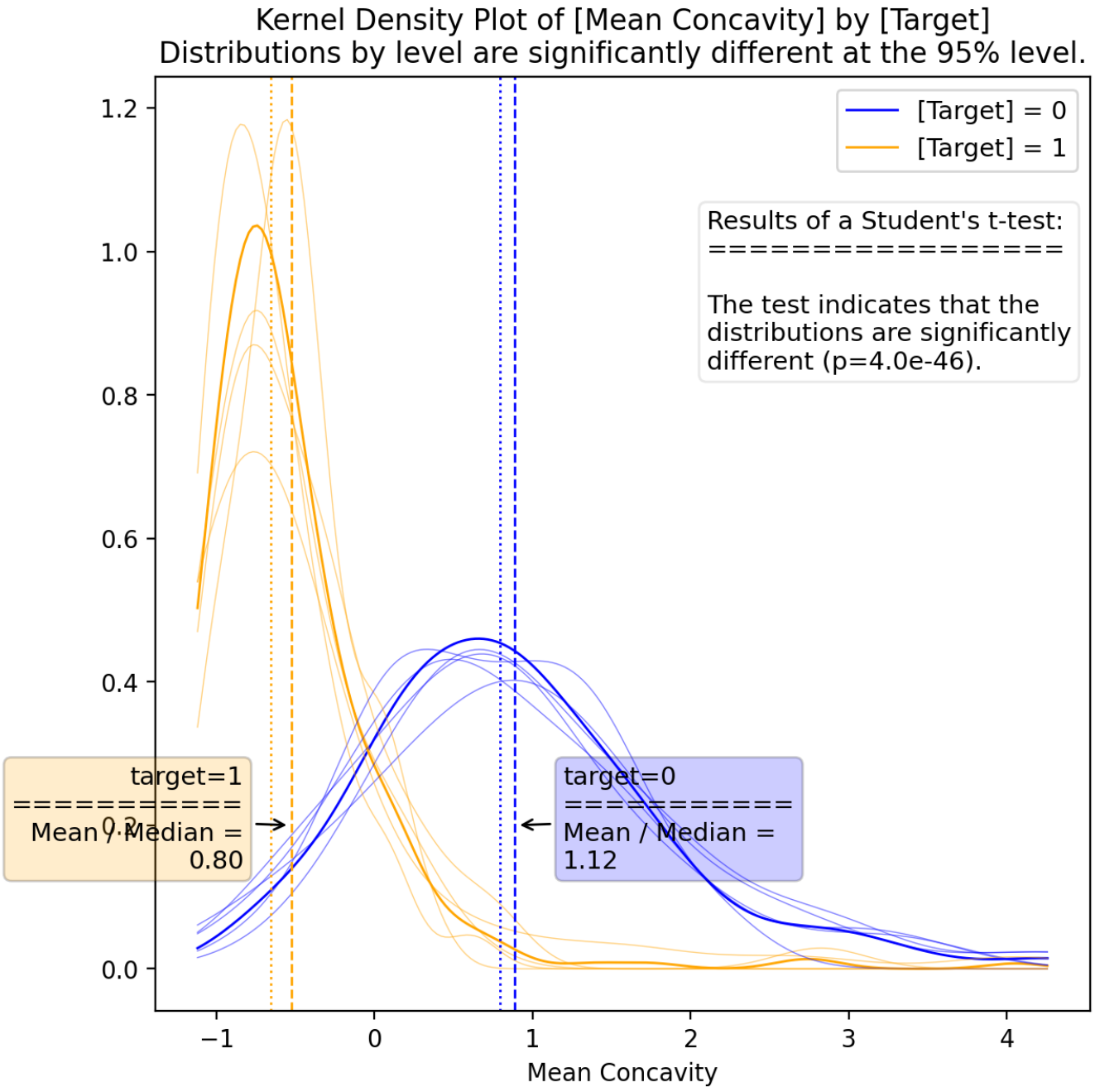
Univariate Report

Mean Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-2.7e+00	-2.4e+00	-3.4e+00	-2.7e+00	-2.5e+00	-2.7e+00	4.0e-01
Fitted p-Value	1.1e-24	2.9e-24	1.4e-24	3.4e-25	1.6e-24	1.3e-30	9.5e-25
Fitted Std. Err.	0.260	0.239	0.335	0.256	0.244	0.234	0.039
Conf. Int. Lower	-3.2e+00	-2.9e+00	-4.1e+00	-3.2e+00	-3.0e+00	-3.2e+00	4.8e-01
Conf. Int. Upper	-2.2e+00	-2.0e+00	-2.8e+00	-2.2e+00	-2.0e+00	-2.2e+00	3.2e-01
Train Accuracy	88.6%	87.1%	87.8%	87.0%	86.8%	87.5%	0.7%
Val Accuracy	83.2%	88.9%	86.0%	89.4%	90.4%	37.7%	2.9%
Train AUC	88.4%	87.0%	87.6%	86.5%	86.7%	87.3%	0.8%
Val AUC	83.1%	87.9%	85.9%	89.7%	89.6%	50.0%	2.8%
Train F1	90.9%	89.1%	90.1%	89.7%	89.2%	89.8%	0.7%
Test F1	85.2%	92.1%	88.7%	90.3%	92.5%	0.0%	3.0%
Train Precision	92.8%	90.8%	91.8%	91.3%	91.4%	91.6%	0.7%
Val Precision	86.8%	94.1%	91.1%	93.3%	92.5%	0.0%	2.9%
Train Recall	89.2%	87.4%	88.5%	88.2%	87.1%	88.1%	0.8%
Val Recall	83.6%	90.1%	86.4%	87.5%	92.5%	0.0%	3.4%
Train MCC	75.7%	73.3%	74.4%	72.2%	72.4%	73.6%	1.5%
Val MCC	65.7%	73.7%	70.6%	78.9%	79.1%	0.0%	5.7%
Train Log-Loss	4.10	4.66	4.38	4.68	4.75	4.52	0.27
Val Log-Loss	6.07	4.00	5.04	3.82	3.47	22.45	1.06

Univariate Report

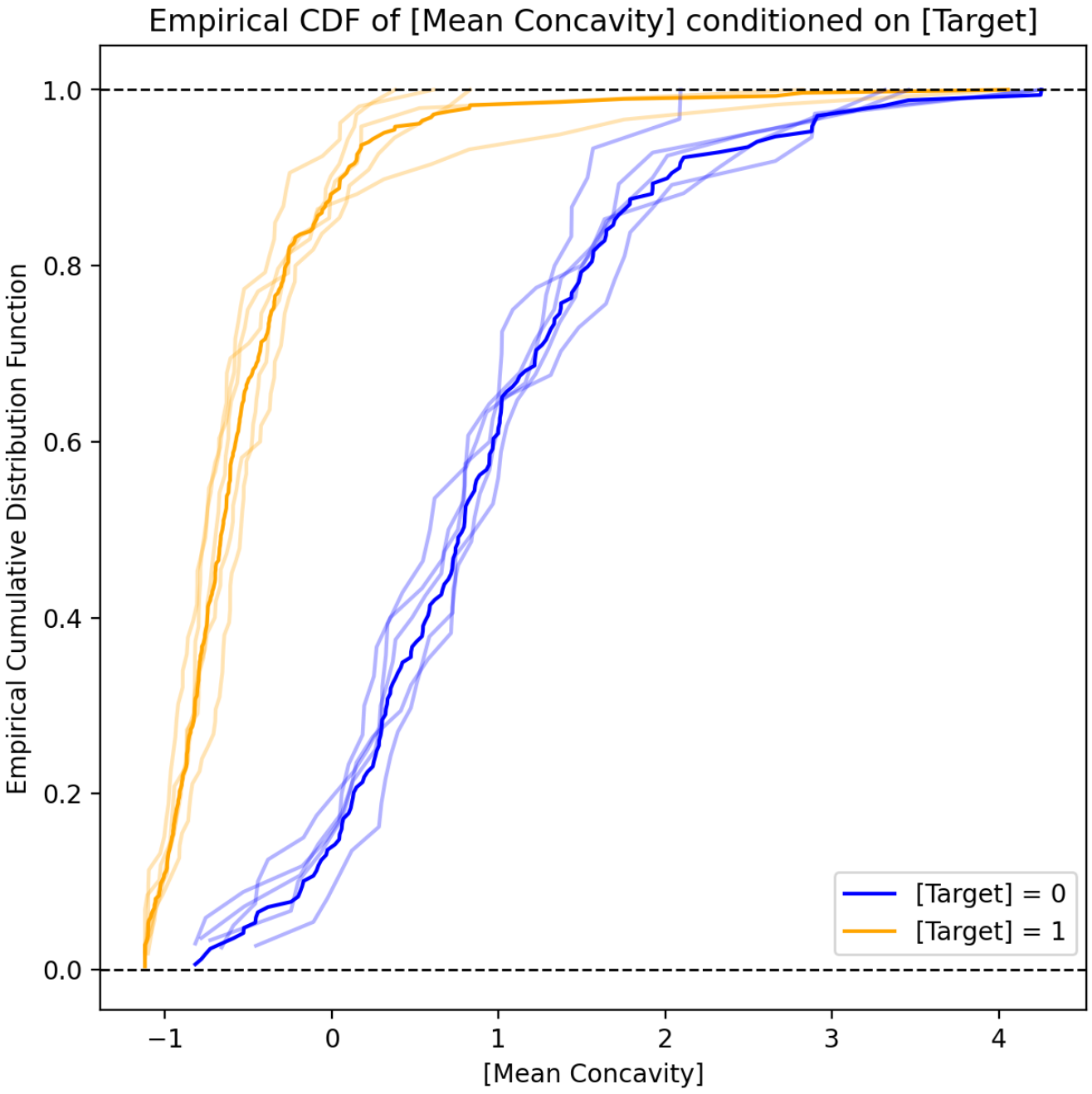
Mean Concavity - Kernel Density Plot



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

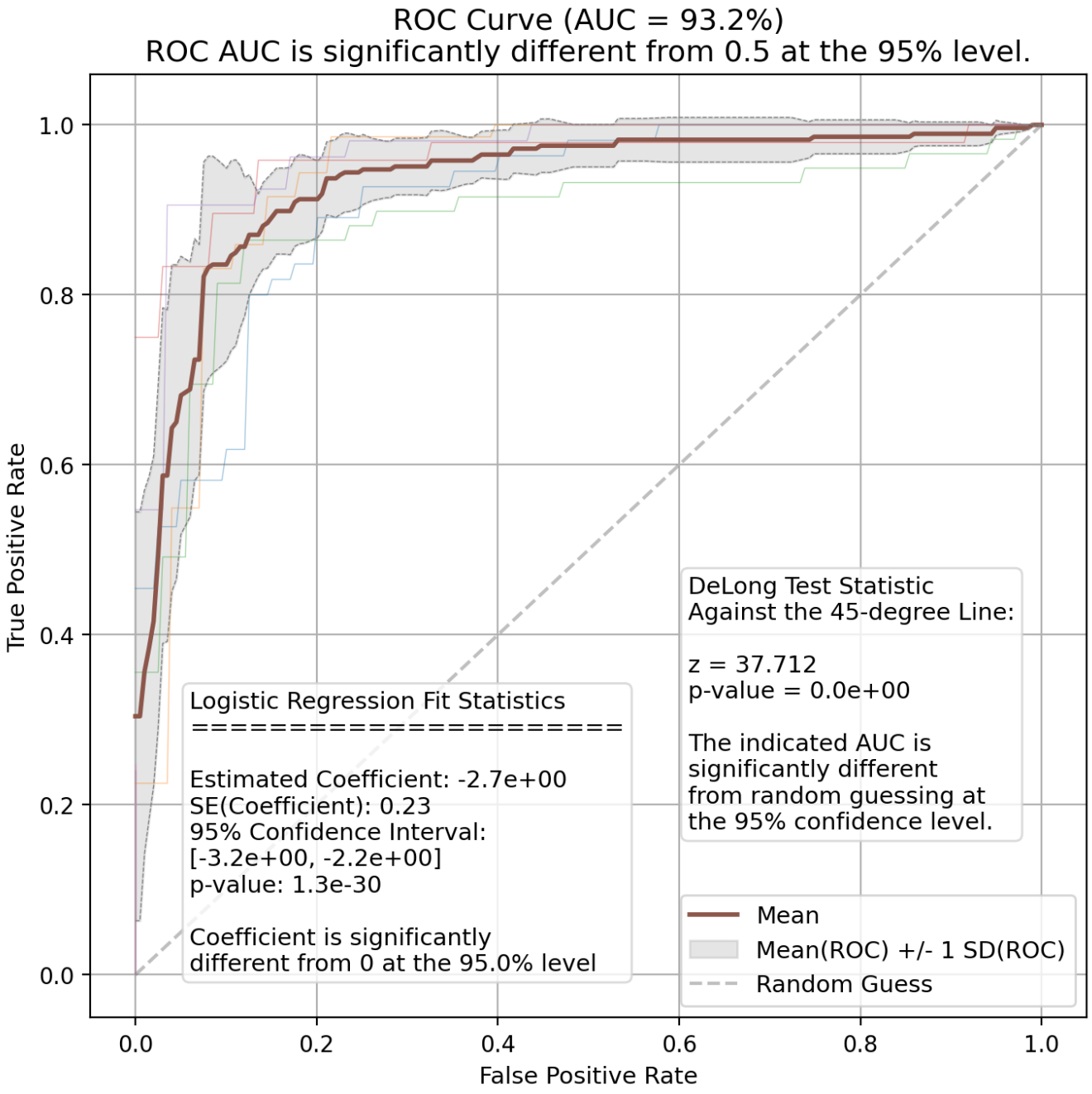
Mean Concavity - Empirical CDF Plot



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

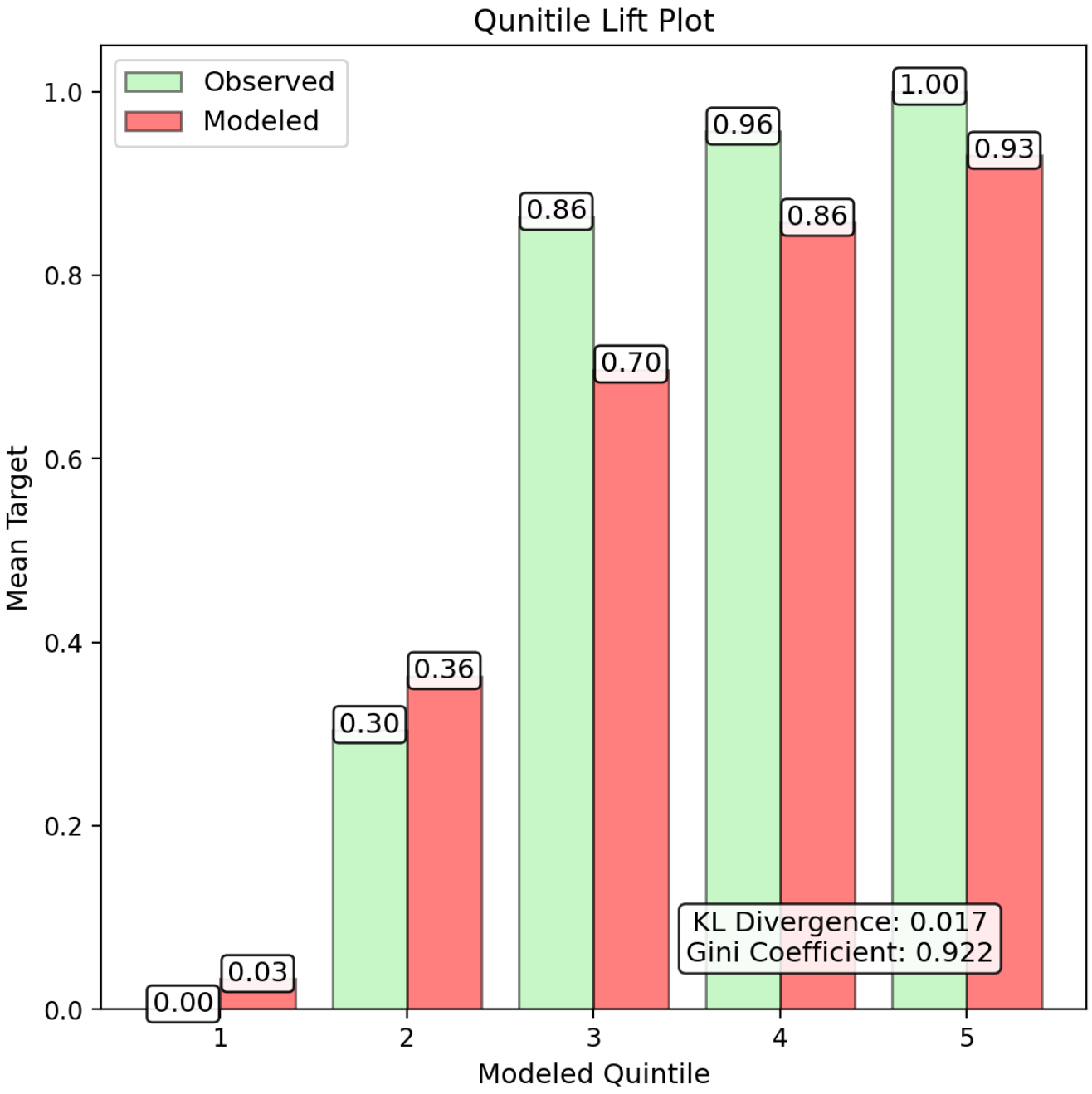


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Mean Concavity - Quintile Lift



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

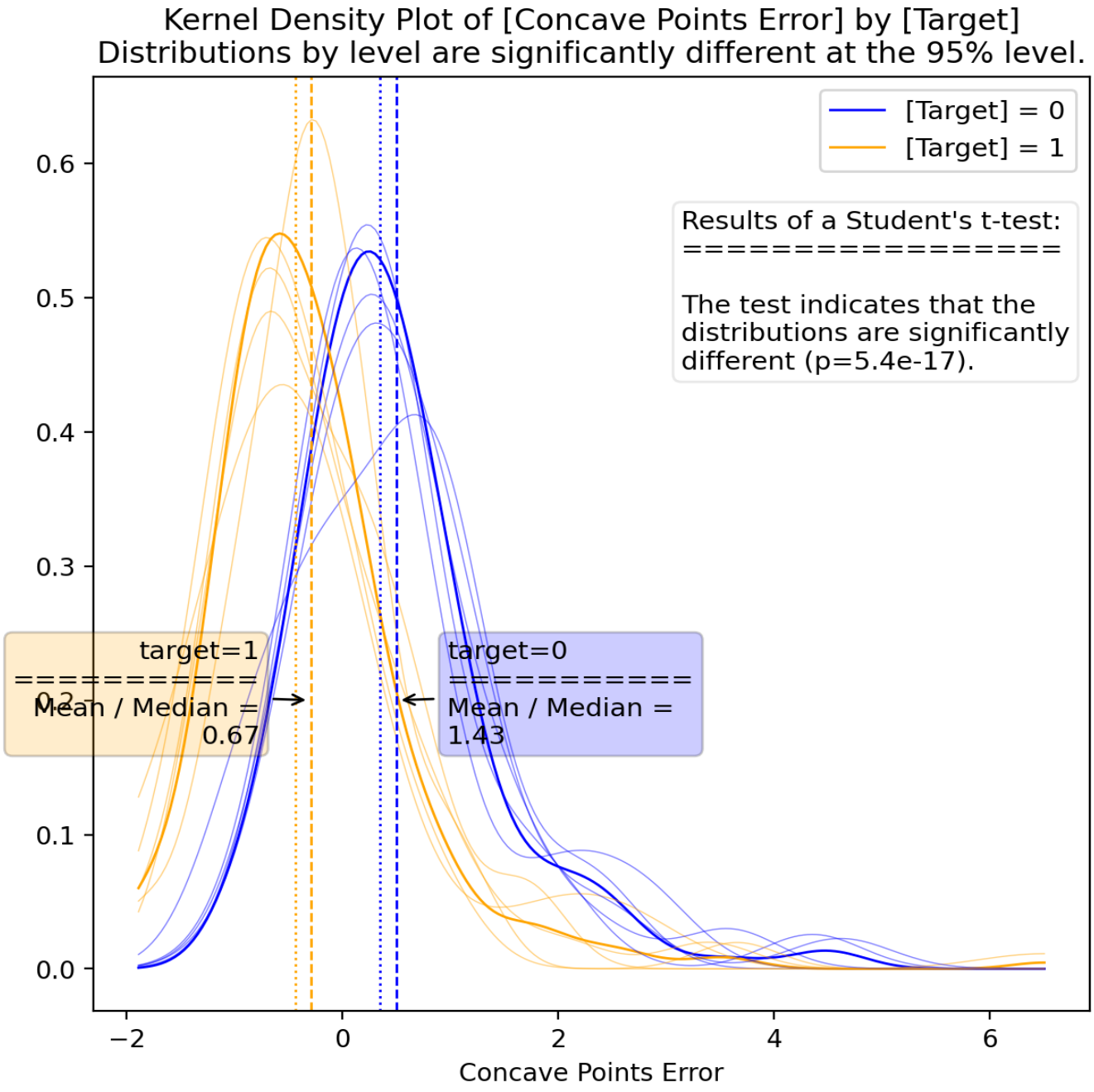
Univariate Report

Concave Points Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-1.0e+00	-9.1e-01	-1.2e+00	-9.5e-01	-8.8e-01	-9.8e-01	1.2e-01
Fitted p-Value	1.8e-10	4.0e-10	1.6e-12	2.2e-10	1.3e-09	7.3e-13	5.2e-10
Fitted Std. Err.	0.157	0.146	0.167	0.150	0.145	0.136	0.009
Conf. Int. Lower	-1.3e+00	-1.2e+00	-1.5e+00	-1.2e+00	-1.2e+00	-1.2e+00	1.4e-01
Conf. Int. Upper	-6.9e-01	-6.3e-01	-8.5e-01	-6.6e-01	-6.0e-01	-7.1e-01	1.0e-01
Train Accuracy	74.7%	73.6%	71.0%	71.9%	71.5%	72.5%	1.6%
Val Accuracy	64.2%	68.7%	78.5%	75.3%	77.1%	37.7%	6.1%
Train AUC	74.3%	73.3%	70.9%	71.2%	71.4%	72.2%	1.5%
Val AUC	64.3%	68.4%	77.4%	75.6%	75.6%	50.0%	5.6%
Train F1	79.4%	77.4%	75.5%	77.1%	75.9%	77.1%	1.5%
Test F1	67.3%	76.0%	82.8%	76.9%	81.9%	0.0%	6.2%
Train Precision	83.3%	80.1%	80.2%	81.0%	80.7%	81.1%	1.3%
Val Precision	71.4%	84.5%	84.2%	81.4%	82.7%	0.0%	5.4%
Train Recall	75.8%	74.9%	71.4%	73.5%	71.7%	73.4%	1.9%
Val Recall	63.6%	69.0%	81.4%	72.9%	81.1%	0.0%	7.7%
Train MCC	47.3%	45.9%	40.6%	41.3%	41.8%	43.4%	3.0%
Val MCC	28.3%	33.7%	54.3%	50.9%	50.8%	0.0%	11.7%
Train Log-Loss	9.11	9.52	10.45	10.13	10.27	9.90	0.56
Val Log-Loss	12.90	11.29	7.75	8.90	8.25	22.45	2.19

Univariate Report

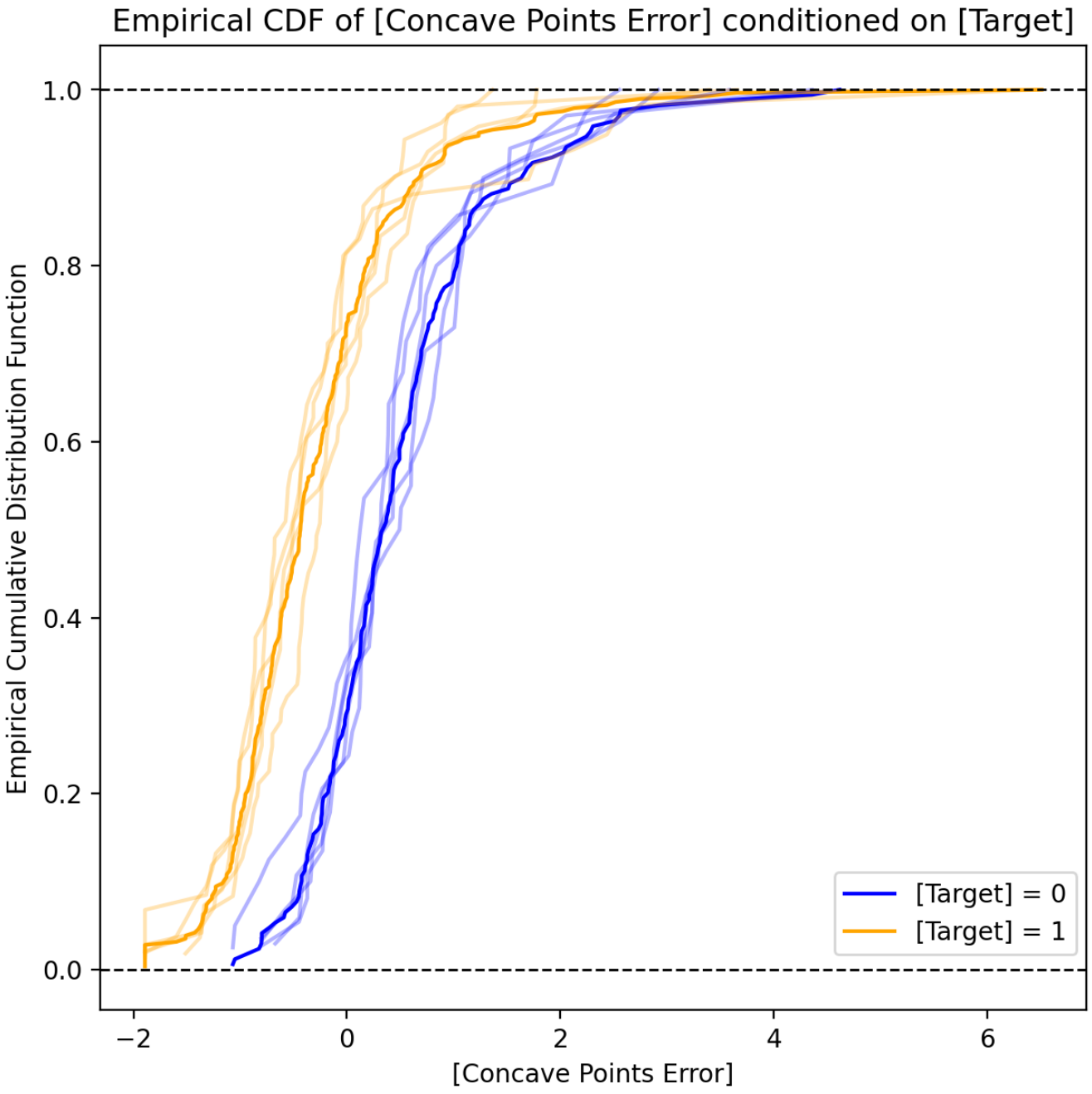
Concave Points Error - Kernel Density Plot



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

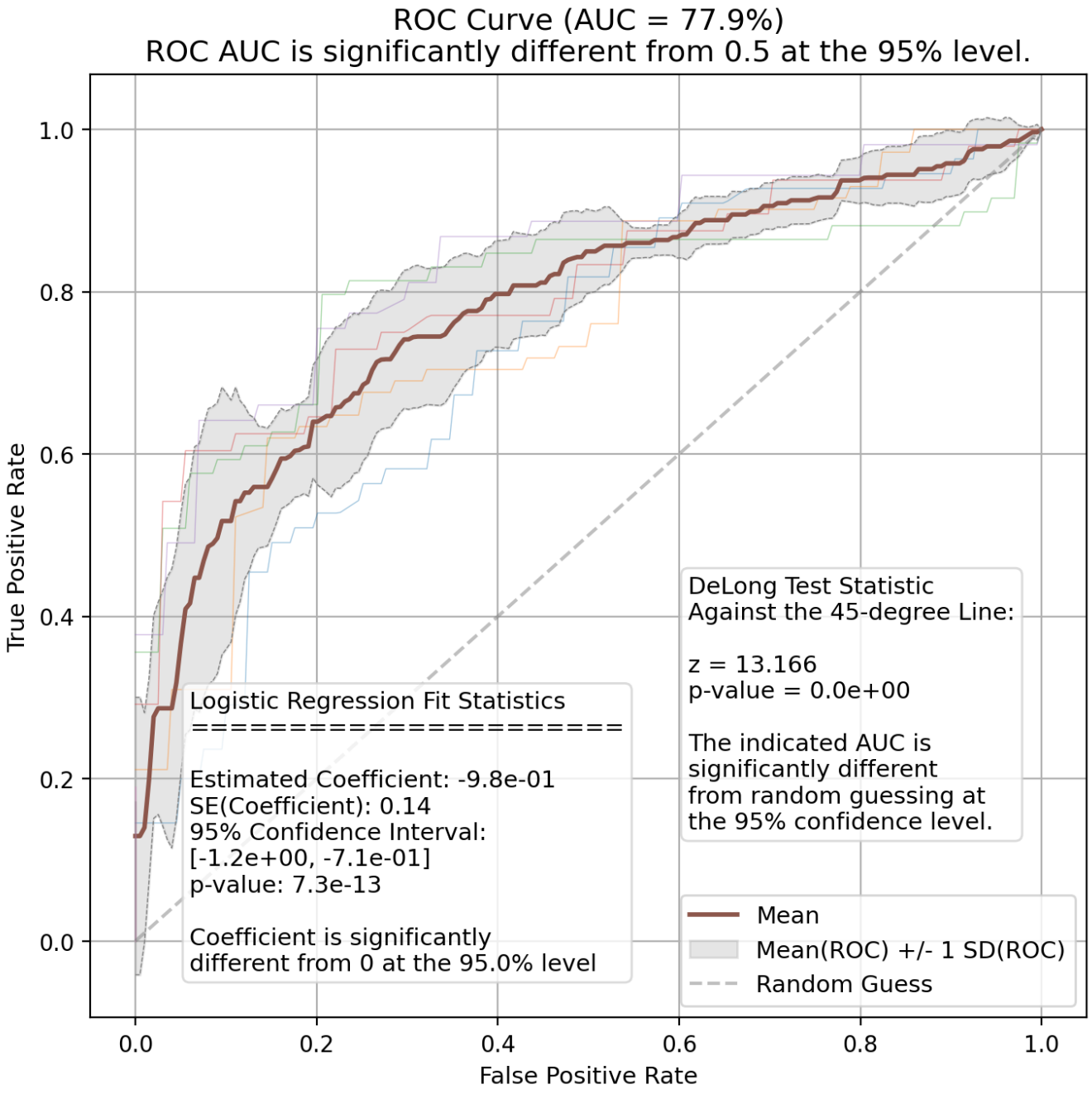
Concave Points Error - Empirical CDF Plot



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

Univariate Report

Concave Points Error - ROC Curve

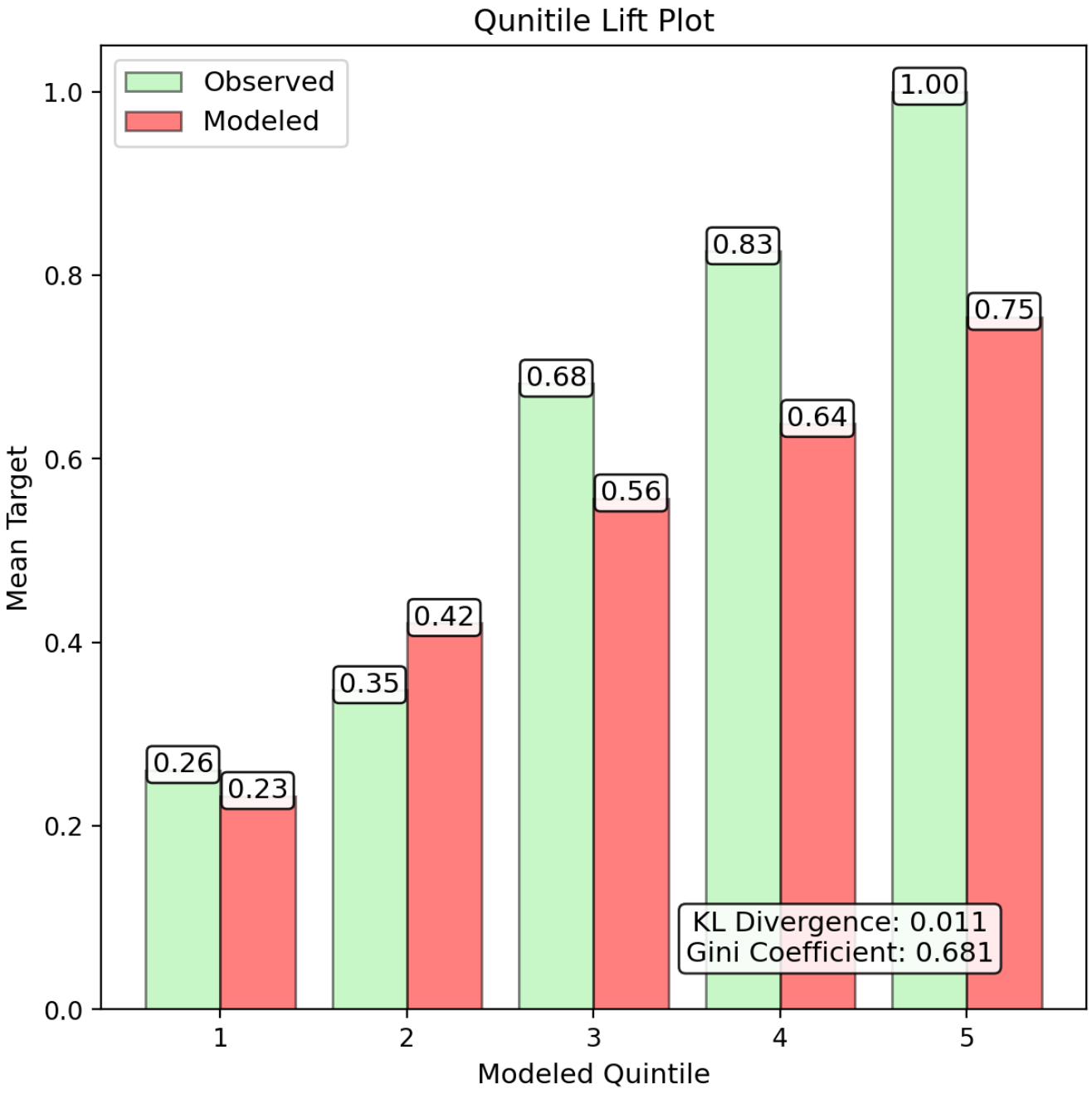


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data. Significance of the ROC curve is determined based on the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Concave Points Error - Quintile Lift



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

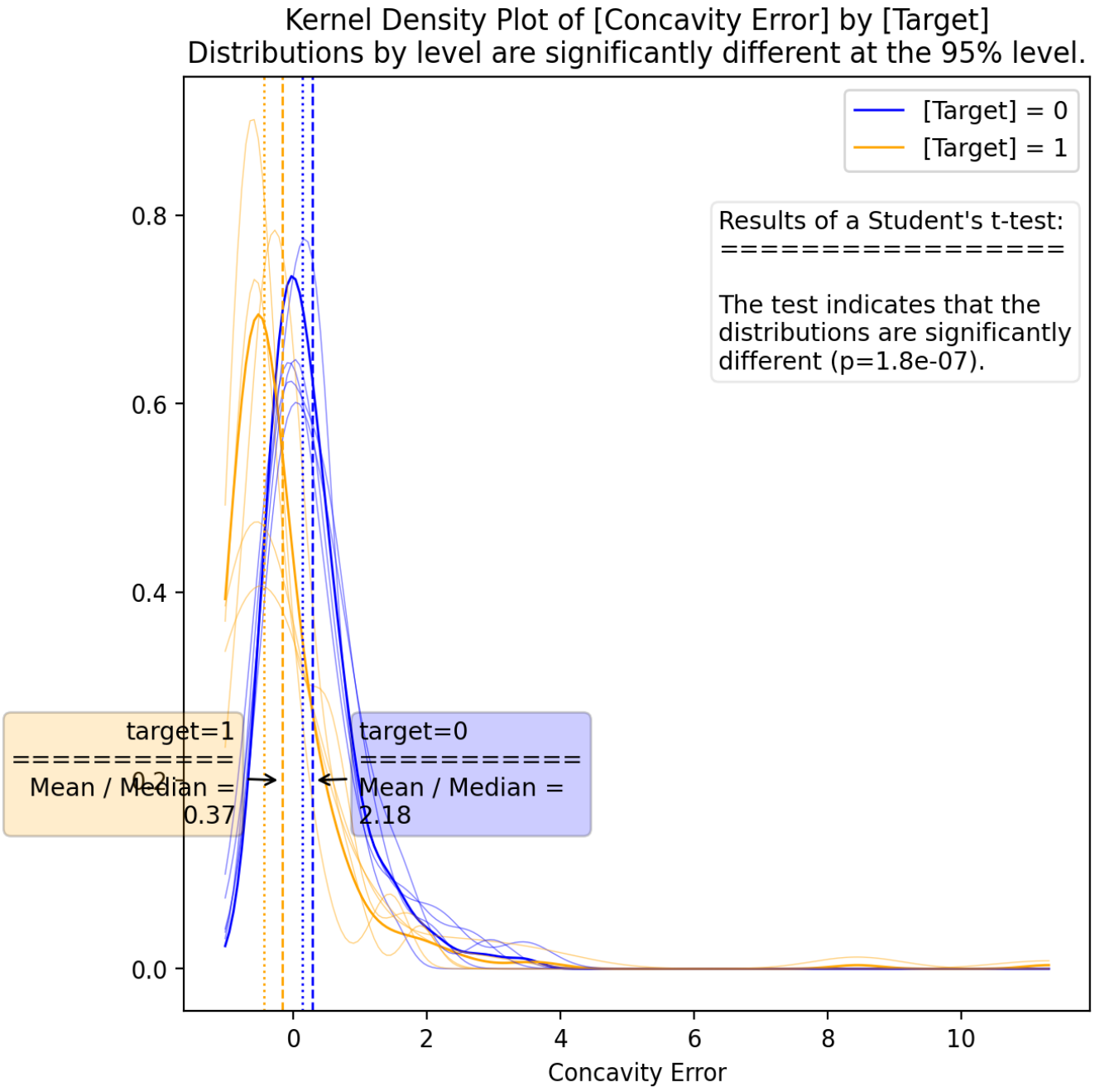
Univariate Report

Concavity Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-5.8e-01	-5.7e-01	-1.0e+00	-6.8e-01	-4.4e-01	-6.4e-01	2.1e-01
Fitted p-Value	1.7e-04	1.9e-04	5.9e-08	3.4e-05	2.3e-03	9.4e-06	9.8e-04
Fitted Std. Err.	0.154	0.154	0.186	0.165	0.144	0.144	0.016
Conf. Int. Lower	-8.8e-01	-8.8e-01	-1.4e+00	-1.0e+00	-7.2e-01	-9.2e-01	2.5e-01
Conf. Int. Upper	-2.8e-01	-2.7e-01	-6.4e-01	-3.6e-01	-1.6e-01	-3.5e-01	1.8e-01
Train Accuracy	71.7%	71.6%	69.6%	69.2%	68.5%	70.1%	1.4%
Val Accuracy	64.2%	64.6%	72.0%	74.1%	77.1%	37.7%	5.8%
Train AUC	69.2%	69.9%	67.8%	66.8%	66.9%	68.1%	1.4%
Val AUC	63.6%	61.3%	69.2%	73.4%	73.4%	50.0%	5.6%
Train F1	77.9%	76.9%	75.6%	75.8%	74.5%	76.1%	1.3%
Test F1	68.5%	73.7%	78.3%	77.6%	82.9%	0.0%	5.4%
Train Precision	77.9%	75.7%	76.2%	76.5%	75.7%	76.4%	0.9%
Val Precision	69.8%	79.0%	77.0%	76.0%	79.3%	0.0%	3.8%
Train Recall	77.9%	78.1%	74.9%	75.2%	73.4%	75.9%	2.1%
Val Recall	67.3%	69.0%	79.7%	79.2%	86.8%	0.0%	8.1%
Train MCC	38.4%	40.2%	35.4%	33.3%	33.5%	36.1%	3.0%
Val MCC	27.1%	21.0%	39.0%	47.1%	49.0%	0.0%	12.3%
Train Log-Loss	10.21	10.23	10.95	11.11	11.34	10.77	0.52
Val Log-Loss	12.90	12.74	10.08	9.33	8.25	22.45	2.08

Univariate Report

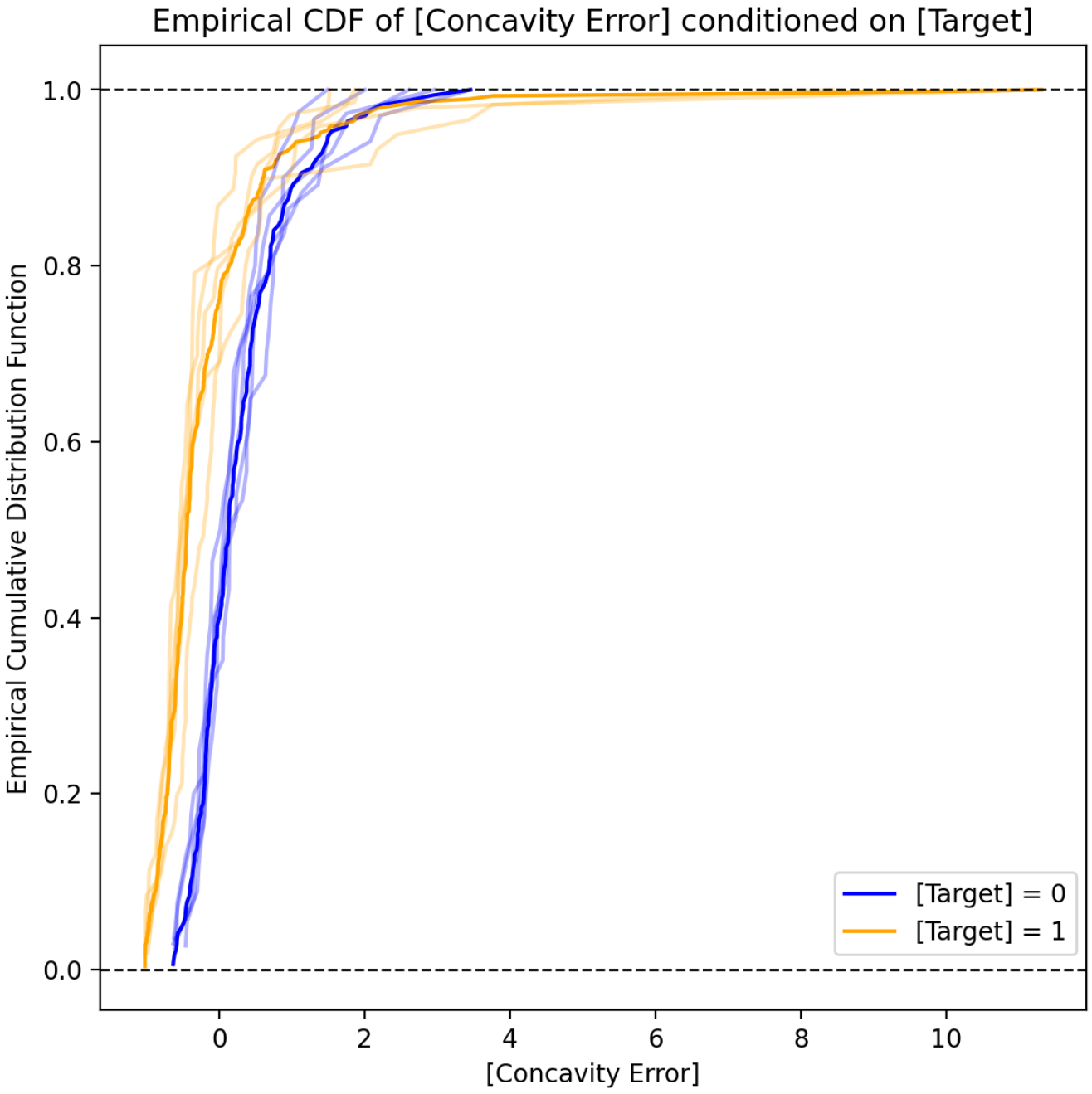
Concavity Error - Kernel Density Plot



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

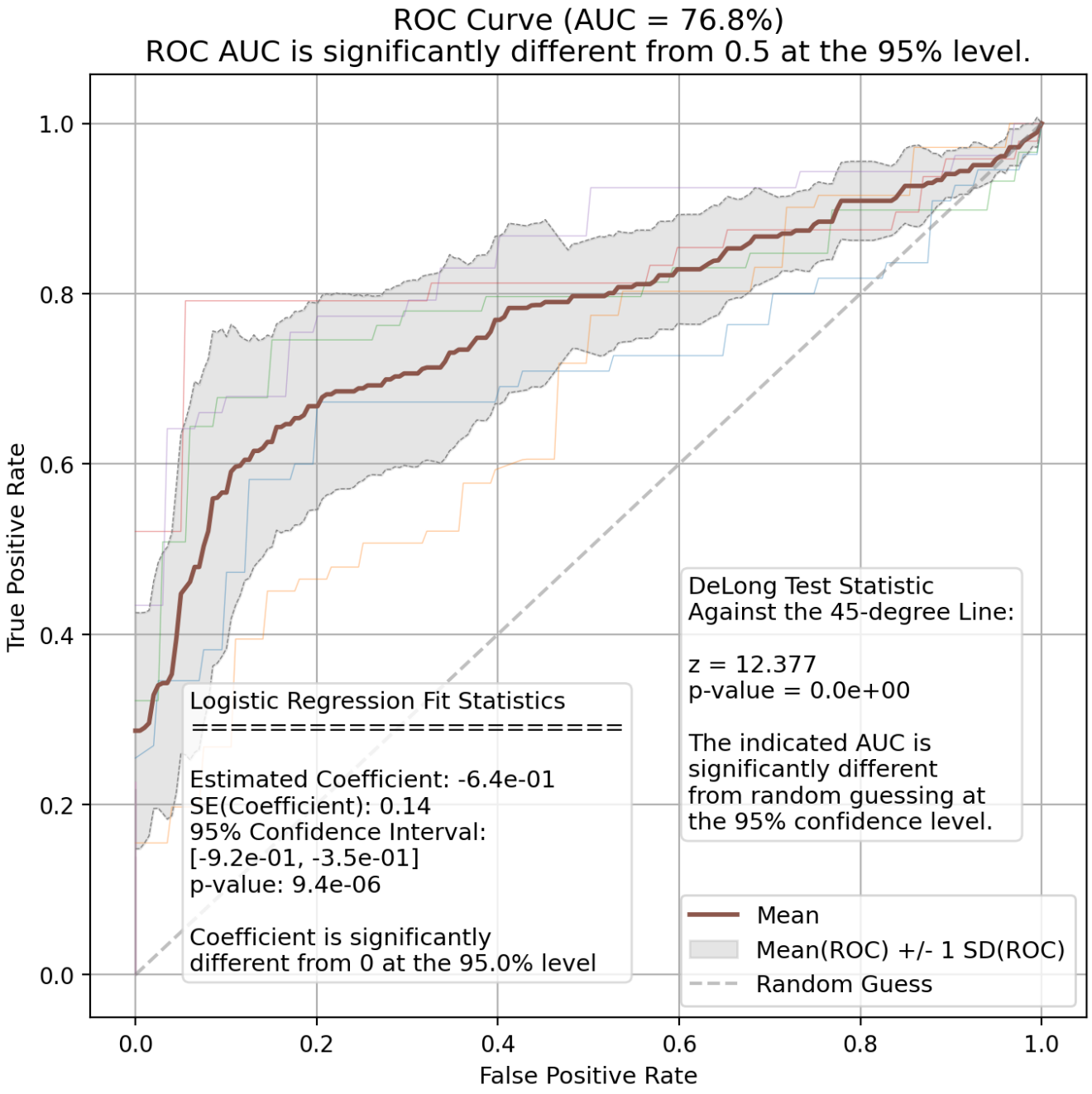
Concavity Error - Empirical CDF Plot



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

Univariate Report

Concavity Error - ROC Curve

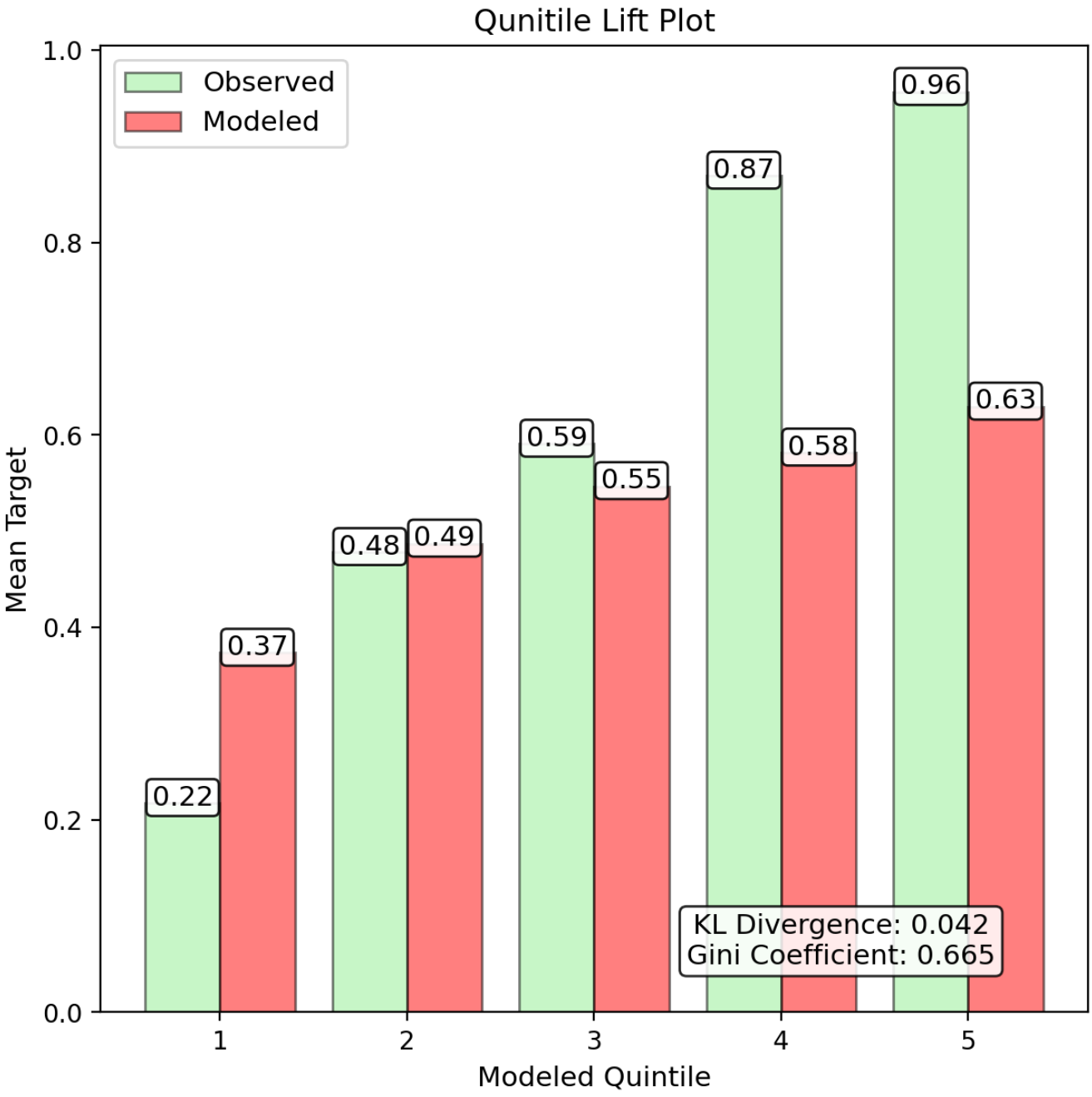


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data. Significance of the ROC curve is determined based on the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Concavity Error - Quintile Lift



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

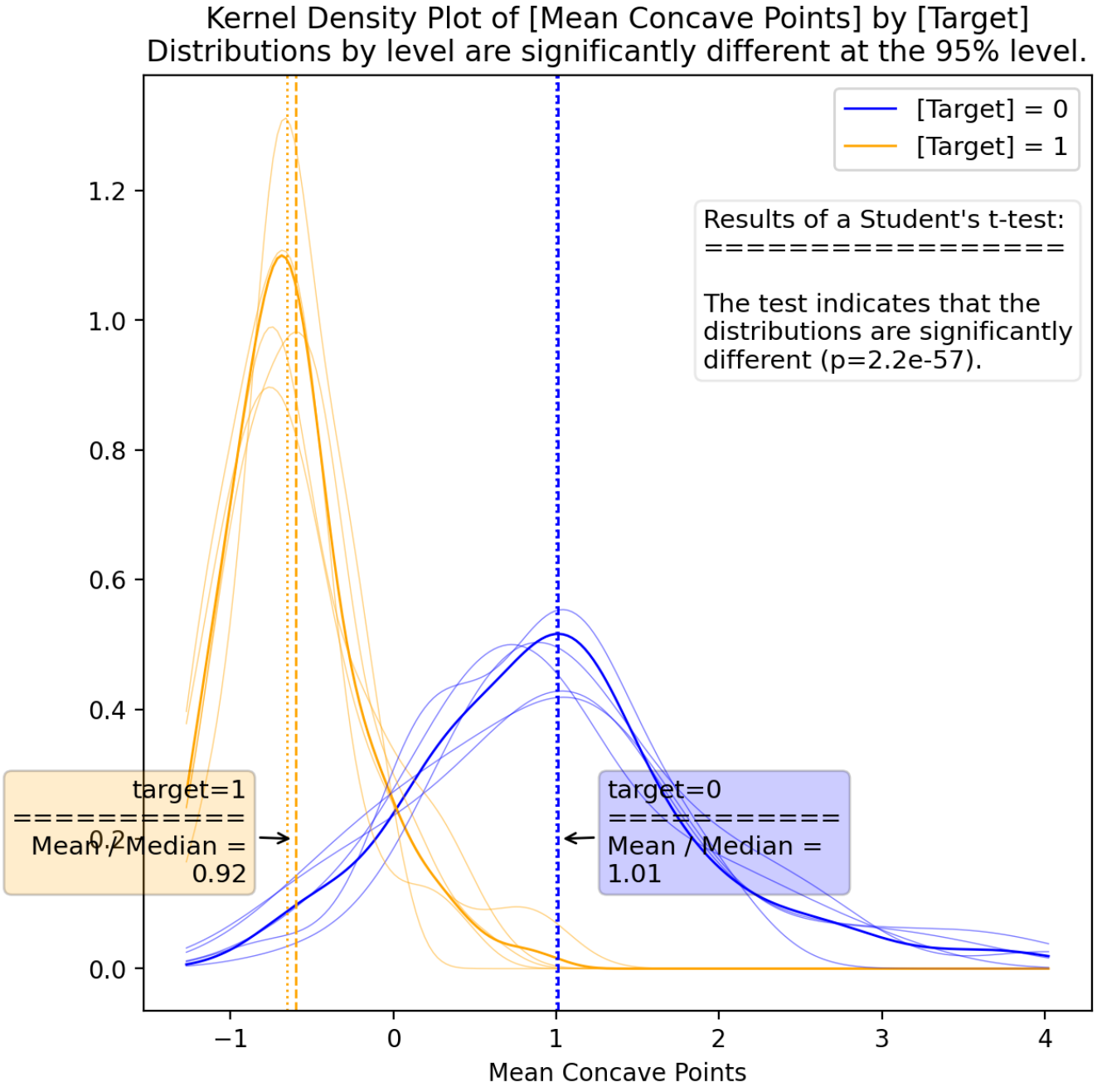
Univariate Report

Mean Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.0e+00	-3.7e+00	-4.2e+00	-3.5e+00	-3.7e+00	-3.8e+00	2.9e-01
Fitted p-Value	9.0e-24	4.2e-24	2.1e-23	2.1e-25	4.1e-25	3.9e-30	8.8e-24
Fitted Std. Err.	0.403	0.361	0.423	0.340	0.355	0.334	0.035
Conf. Int. Lower	-4.8e+00	-4.4e+00	-5.0e+00	-4.2e+00	-4.4e+00	-4.5e+00	3.6e-01
Conf. Int. Upper	-3.3e+00	-2.9e+00	-3.4e+00	-2.9e+00	-3.0e+00	-3.2e+00	2.2e-01
Train Accuracy	92.2%	90.7%	92.5%	89.5%	90.1%	91.0%	1.3%
Val Accuracy	86.3%	91.9%	84.9%	97.6%	95.2%	37.7%	5.5%
Train AUC	92.2%	90.6%	92.6%	89.3%	89.9%	90.9%	1.4%
Val AUC	86.1%	92.2%	84.4%	97.6%	95.5%	50.0%	5.8%
Train F1	93.8%	92.2%	94.0%	91.6%	91.9%	92.7%	1.1%
Test F1	88.1%	94.2%	87.9%	97.9%	96.2%	0.0%	4.6%
Train Precision	95.5%	93.3%	95.5%	93.4%	93.4%	94.2%	1.2%
Val Precision	88.9%	97.0%	89.5%	97.9%	98.0%	0.0%	4.7%
Train Recall	92.2%	91.2%	92.5%	89.9%	90.6%	91.3%	1.1%
Val Recall	87.3%	91.5%	86.4%	97.9%	94.3%	0.0%	4.8%
Train MCC	83.4%	80.8%	84.3%	77.5%	79.0%	81.0%	2.9%
Val MCC	72.0%	81.3%	68.0%	95.2%	89.8%	0.0%	11.5%
Train Log-Loss	2.80	3.34	2.69	3.80	3.58	3.25	0.48
Val Log-Loss	4.93	2.91	5.43	0.85	1.74	22.45	1.98

Univariate Report

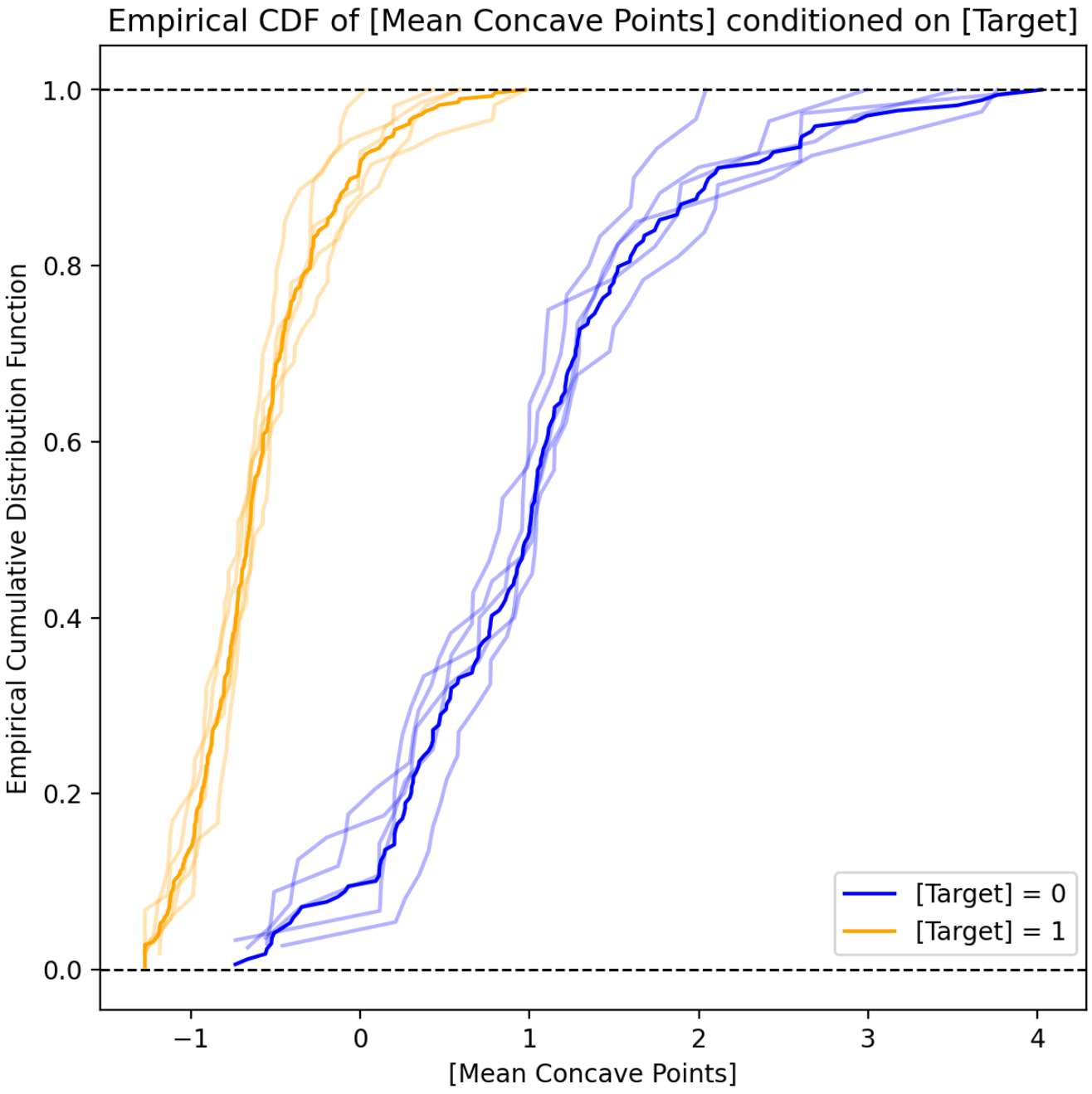
Mean Concave Points - Kernel Density Plot



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

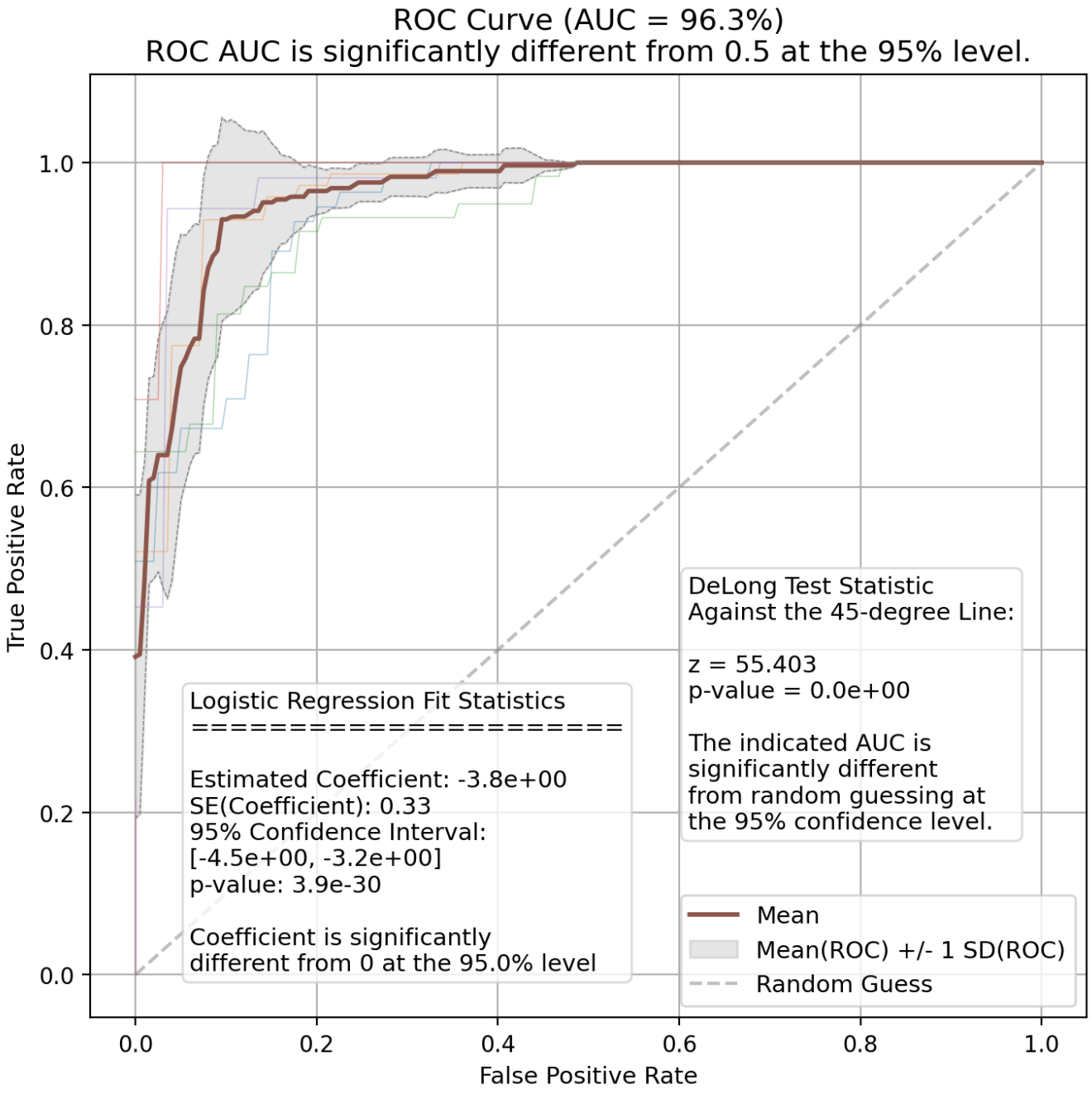
Mean Concave Points - Empirical CDF Plot



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

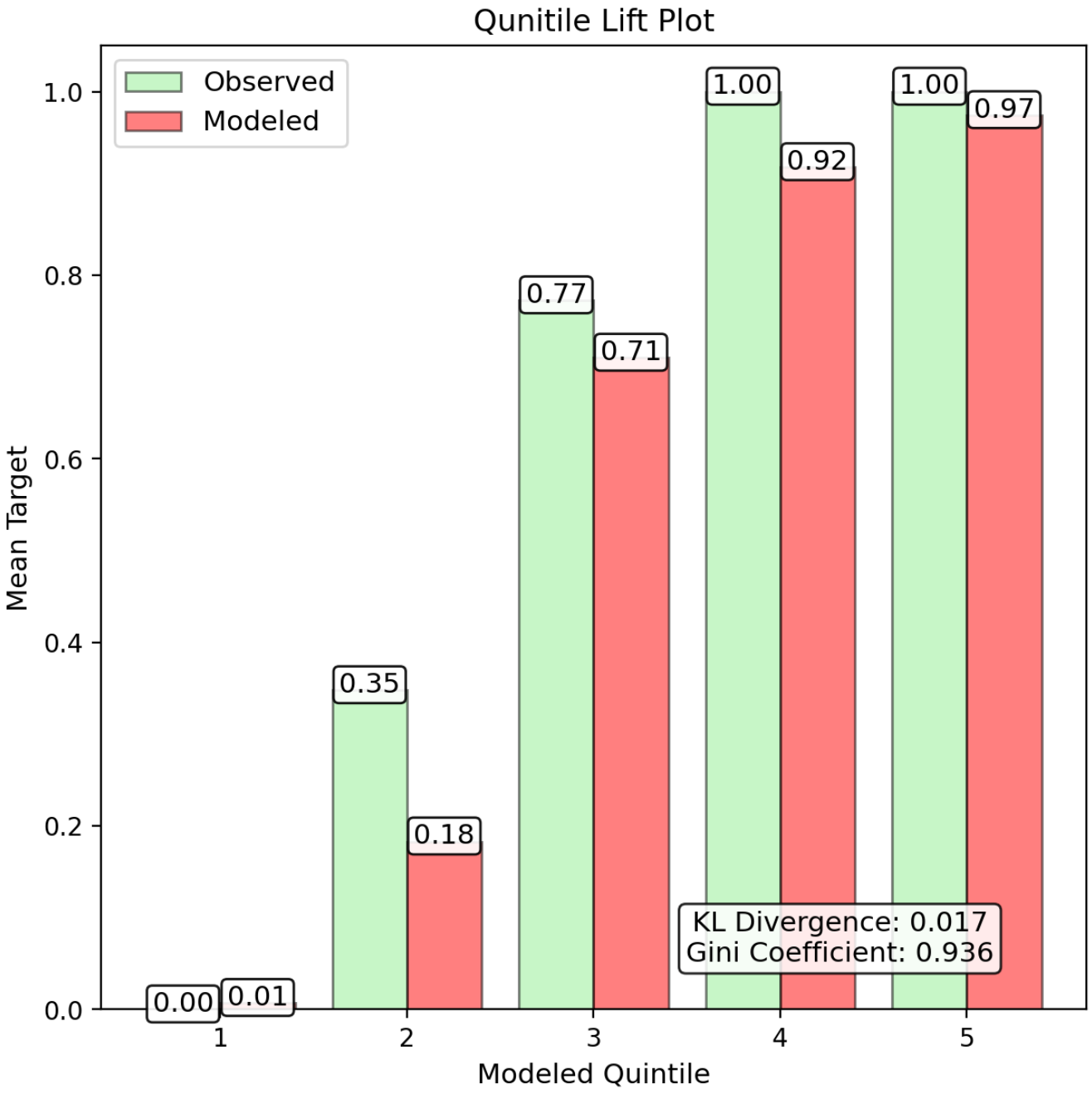


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Mean Concave Points - Quintile Lift



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

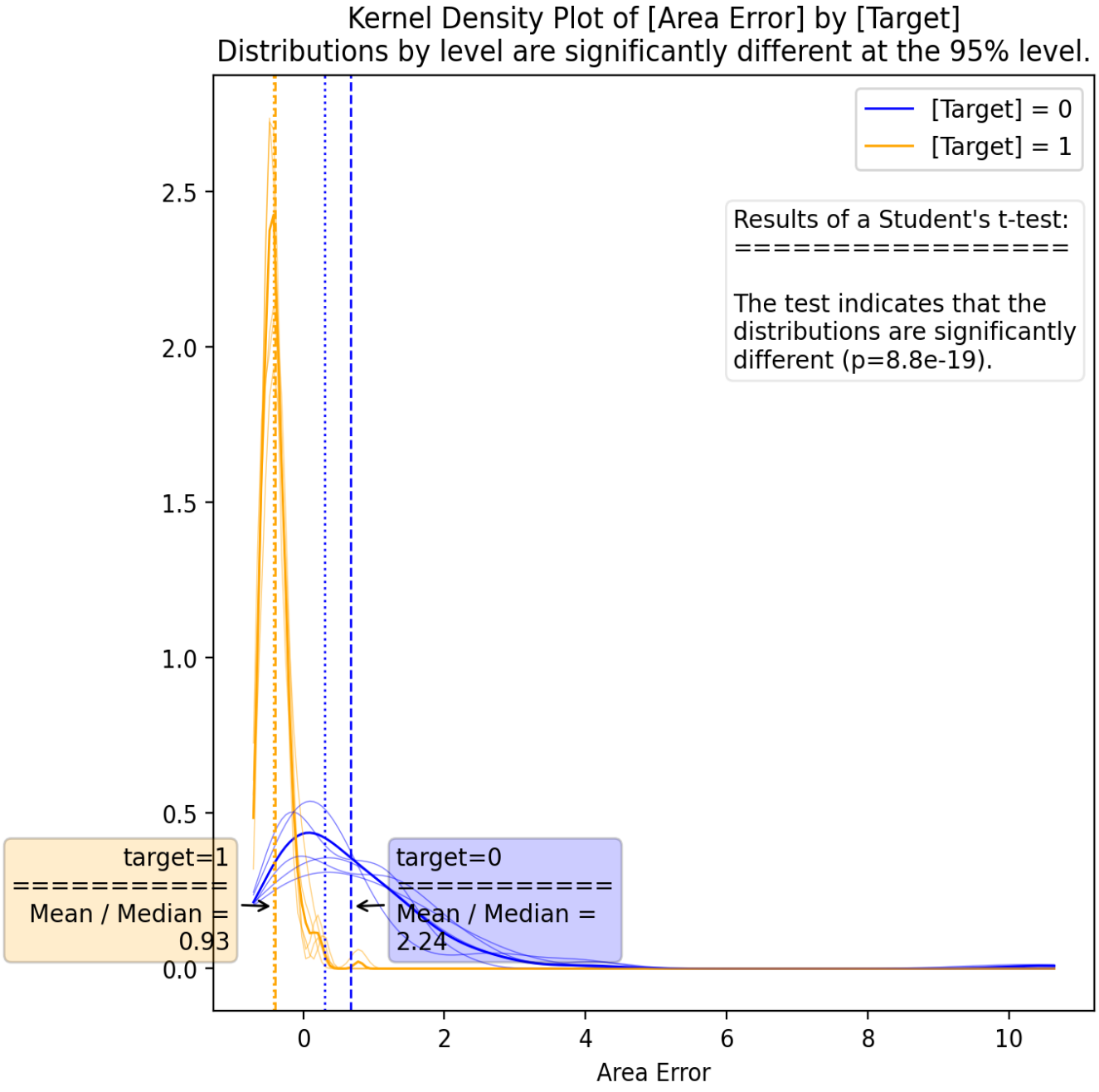
Univariate Report

Area Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-4.7e+00	-4.5e+00	-4.5e+00	-4.7e+00	-4.5e+00	-4.6e+00	1.2e-01
Fitted p-Value	3.7e-25	3.8e-24	3.5e-25	5.9e-26	1.8e-25	4.5e-31	1.6e-24
Fitted Std. Err.	0.458	0.448	0.437	0.451	0.433	0.398	0.010
Conf. Int. Lower	-5.6e+00	-5.4e+00	-5.4e+00	-5.6e+00	-5.4e+00	-5.4e+00	1.4e-01
Conf. Int. Upper	-3.8e+00	-3.7e+00	-3.7e+00	-3.9e+00	-3.7e+00	-3.8e+00	1.0e-01
Train Accuracy	85.6%	85.7%	85.4%	86.2%	85.8%	85.7%	0.3%
Val Accuracy	86.3%	85.9%	87.1%	83.5%	85.5%	37.7%	1.3%
Train AUC	81.4%	82.9%	81.3%	82.2%	82.1%	82.0%	0.7%
Val AUC	84.1%	77.2%	84.8%	81.4%	81.4%	50.0%	3.0%
Train F1	89.5%	89.0%	89.3%	90.0%	89.5%	89.5%	0.3%
Test F1	89.3%	90.8%	90.2%	87.0%	89.5%	0.0%	1.4%
Train Precision	83.8%	82.8%	82.5%	84.5%	83.3%	83.4%	0.8%
Val Precision	81.8%	85.2%	87.3%	78.3%	83.6%	0.0%	3.4%
Train Recall	96.1%	96.3%	97.4%	96.2%	96.6%	96.5%	0.5%
Val Recall	98.2%	97.2%	93.2%	97.9%	96.2%	0.0%	2.0%
Train MCC	68.3%	70.4%	69.0%	69.7%	69.6%	69.4%	0.8%
Val MCC	73.1%	63.4%	71.8%	68.3%	68.5%	0.0%	3.8%
Train Log-Loss	5.21	5.16	5.28	4.97	5.14	5.15	0.11
Val Log-Loss	4.93	5.10	4.65	5.94	5.21	22.45	0.48

Univariate Report

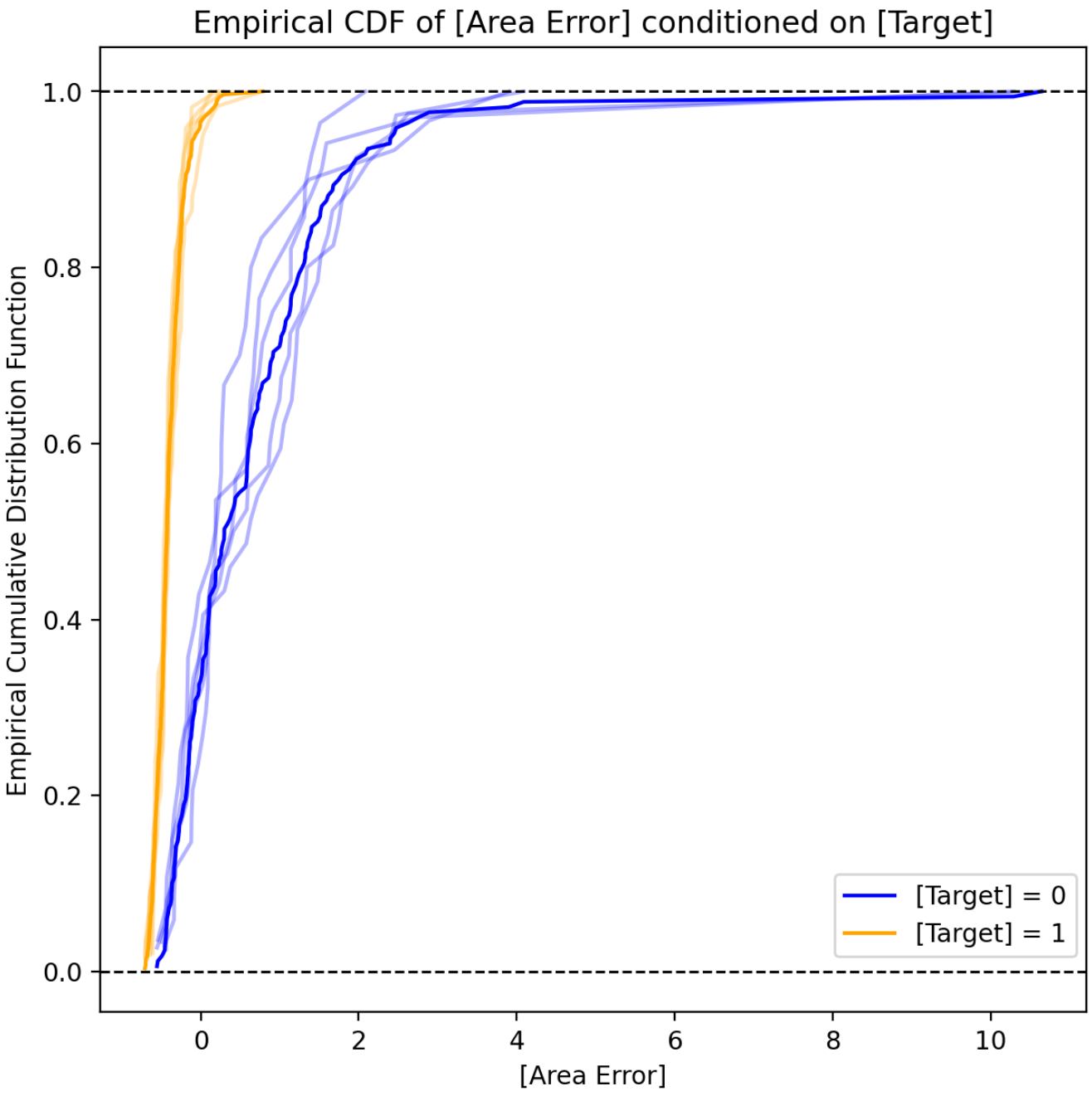
Area Error - Kernel Density Plot



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

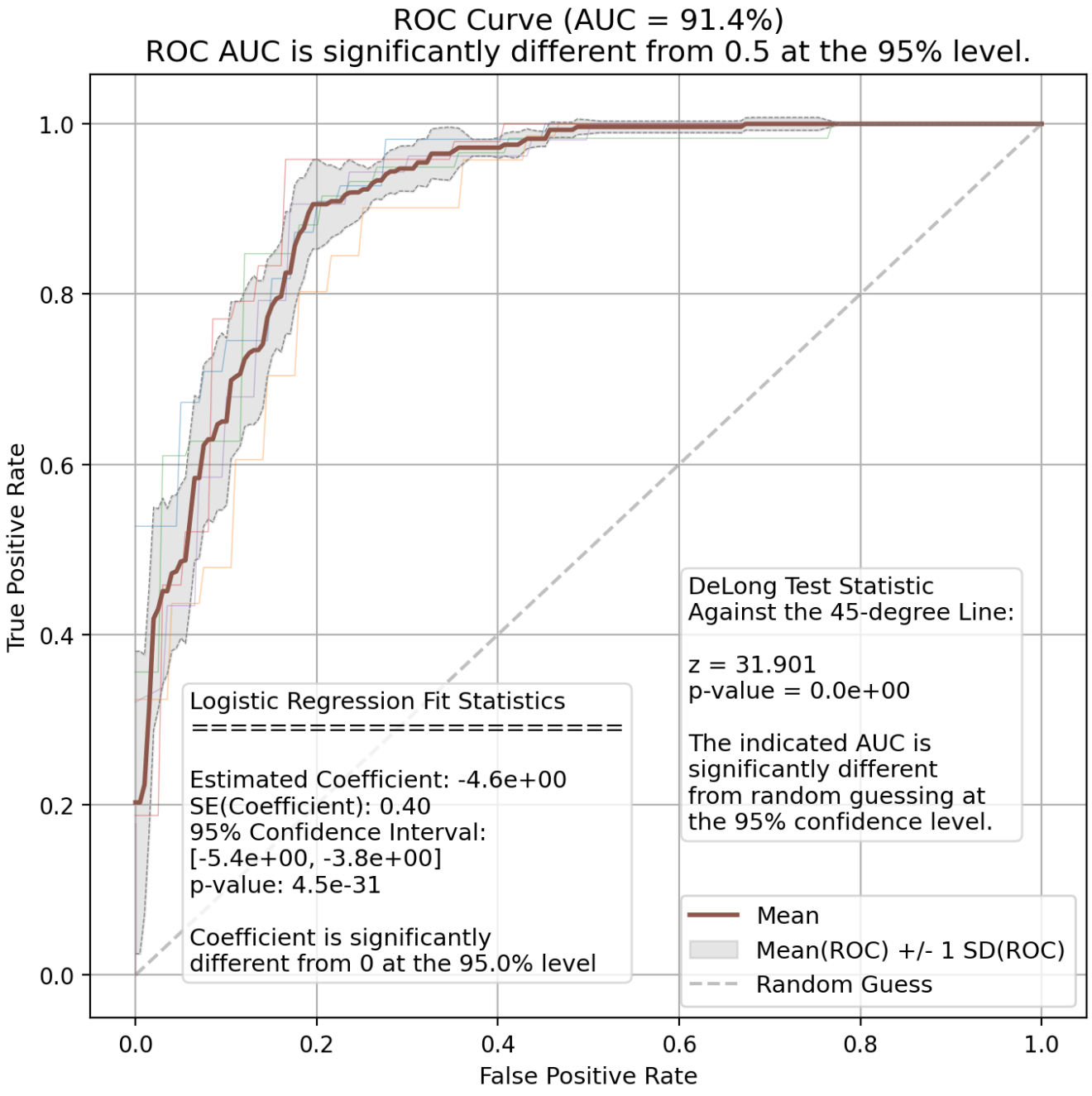
Area Error - Empirical CDF Plot



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

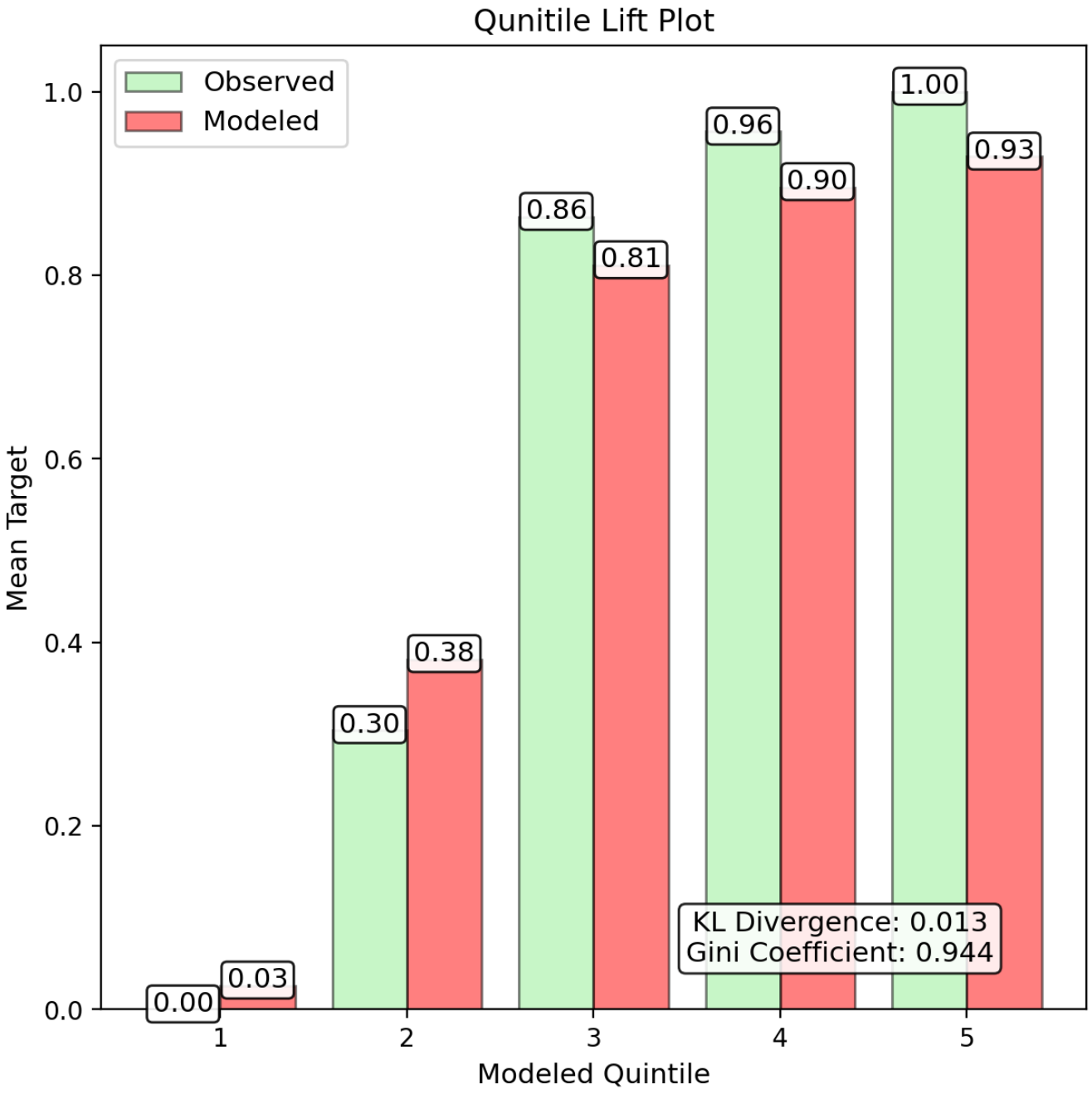


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Area Error - Quintile Lift



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

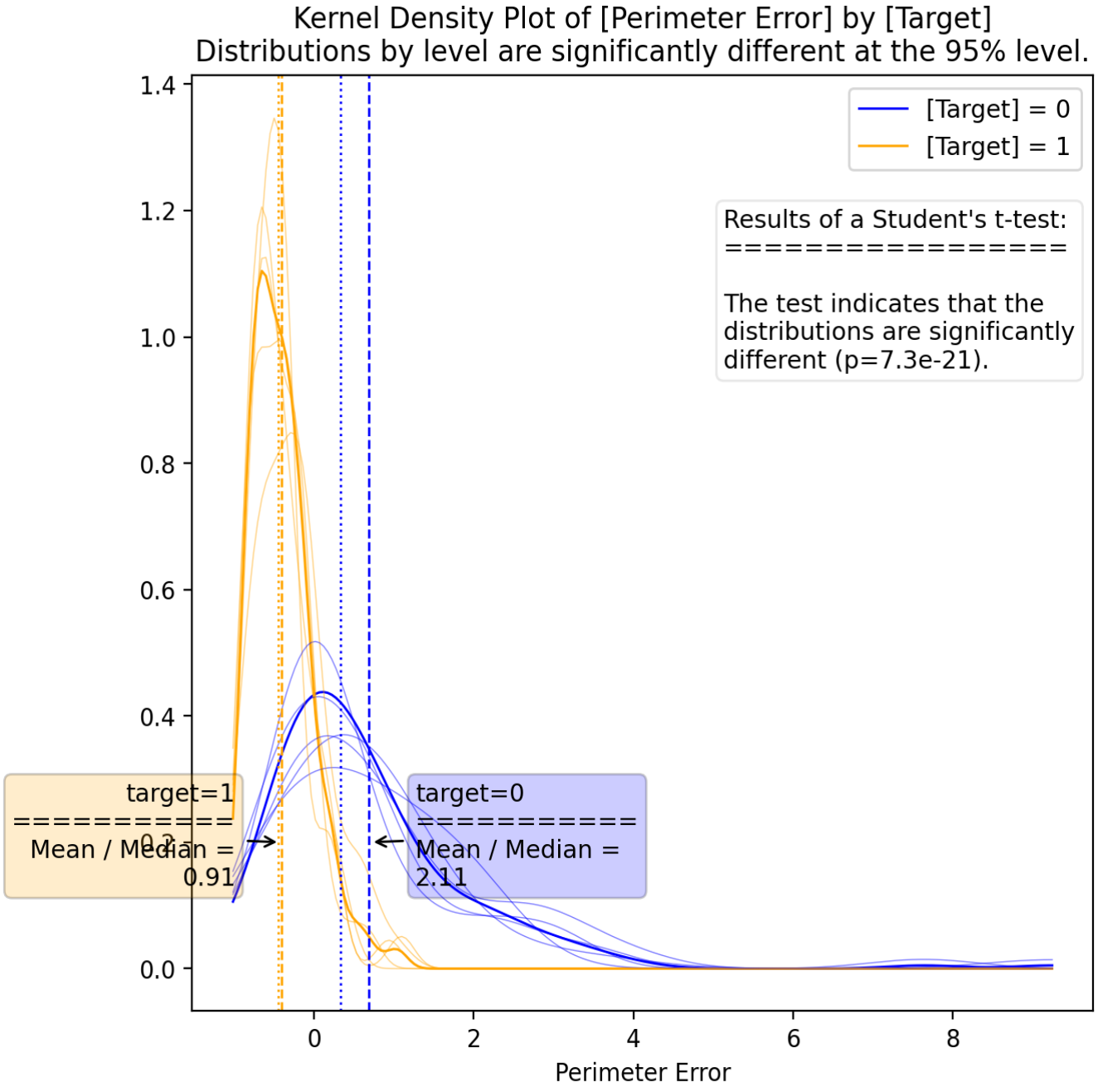
Univariate Report

Perimeter Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.1e+00	-2.9e+00	-3.0e+00	-3.0e+00	-2.9e+00	-3.0e+00	9.5e-02
Fitted p-Value	3.7e-21	2.2e-20	1.1e-21	1.6e-21	7.8e-22	3.1e-26	9.3e-21
Fitted Std. Err.	0.328	0.310	0.316	0.310	0.300	0.280	0.010
Conf. Int. Lower	-3.7e+00	-3.5e+00	-3.6e+00	-3.6e+00	-3.5e+00	-3.5e+00	1.1e-01
Conf. Int. Upper	-2.5e+00	-2.3e+00	-2.4e+00	-2.3e+00	-2.3e+00	-2.4e+00	7.8e-02
Train Accuracy	81.4%	80.9%	82.3%	81.6%	81.5%	81.5%	0.5%
Val Accuracy	82.1%	83.8%	78.5%	81.2%	81.9%	37.7%	1.9%
Train AUC	78.8%	79.4%	79.7%	79.3%	79.5%	79.4%	0.3%
Val AUC	81.1%	77.9%	78.1%	79.9%	78.6%	50.0%	1.4%
Train F1	85.8%	84.5%	86.4%	86.0%	85.5%	85.7%	0.7%
Test F1	85.0%	89.0%	82.5%	84.3%	86.5%	0.0%	2.5%
Train Precision	83.9%	82.7%	83.3%	84.6%	83.9%	83.7%	0.7%
Val Precision	82.8%	86.7%	85.5%	79.6%	82.8%	0.0%	2.7%
Train Recall	87.9%	86.5%	89.9%	87.4%	87.1%	87.8%	1.3%
Val Recall	87.3%	91.5%	79.7%	89.6%	90.6%	0.0%	4.8%
Train MCC	58.9%	59.7%	61.5%	59.5%	59.9%	59.9%	1.0%
Val MCC	63.1%	58.7%	55.0%	61.7%	59.9%	0.0%	3.1%
Train Log-Loss	6.71	6.88	6.37	6.62	6.69	6.65	0.19
Val Log-Loss	6.45	5.83	7.75	6.78	6.51	22.45	0.70

Univariate Report

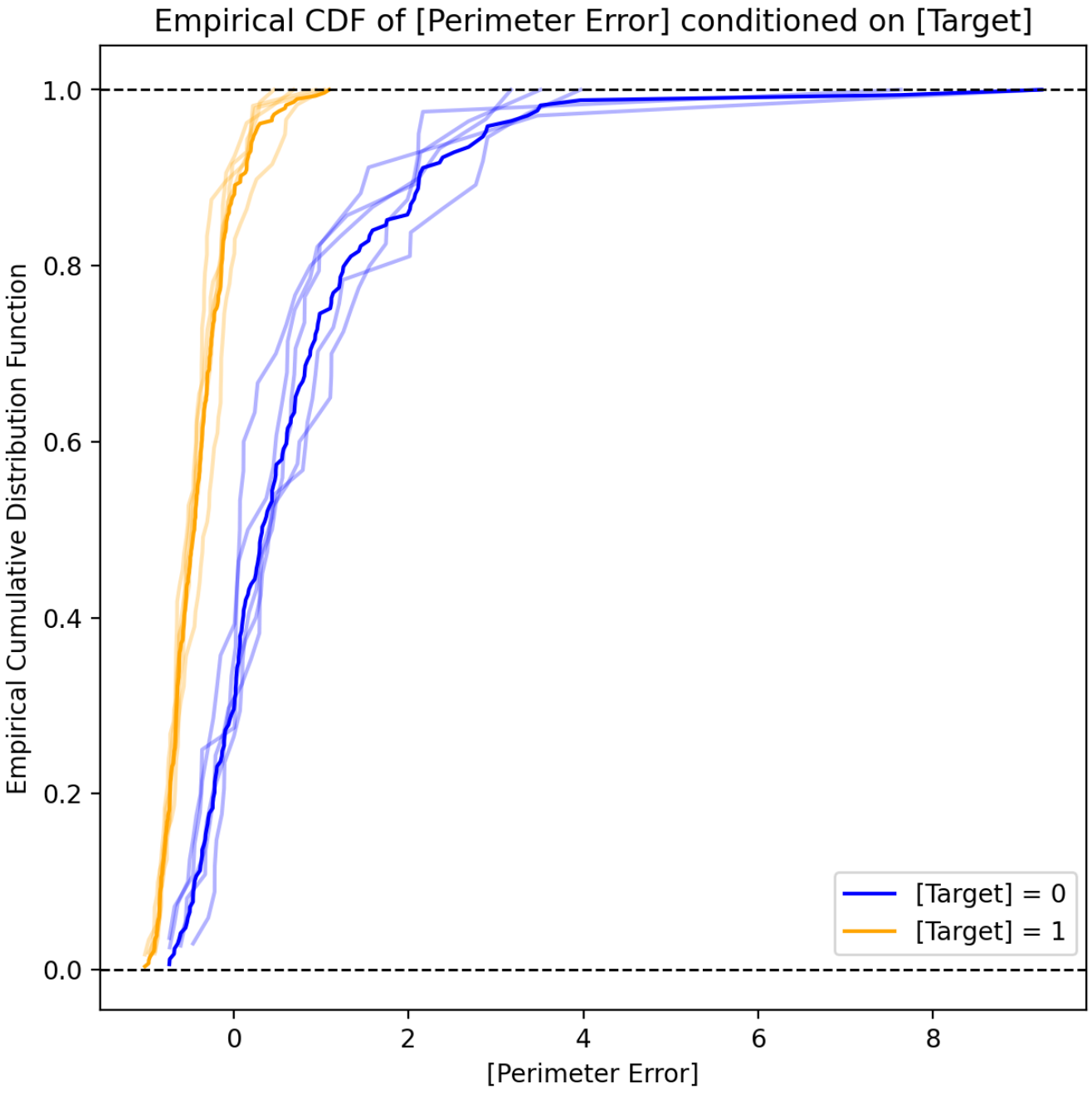
Perimeter Error - Kernel Density Plot



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

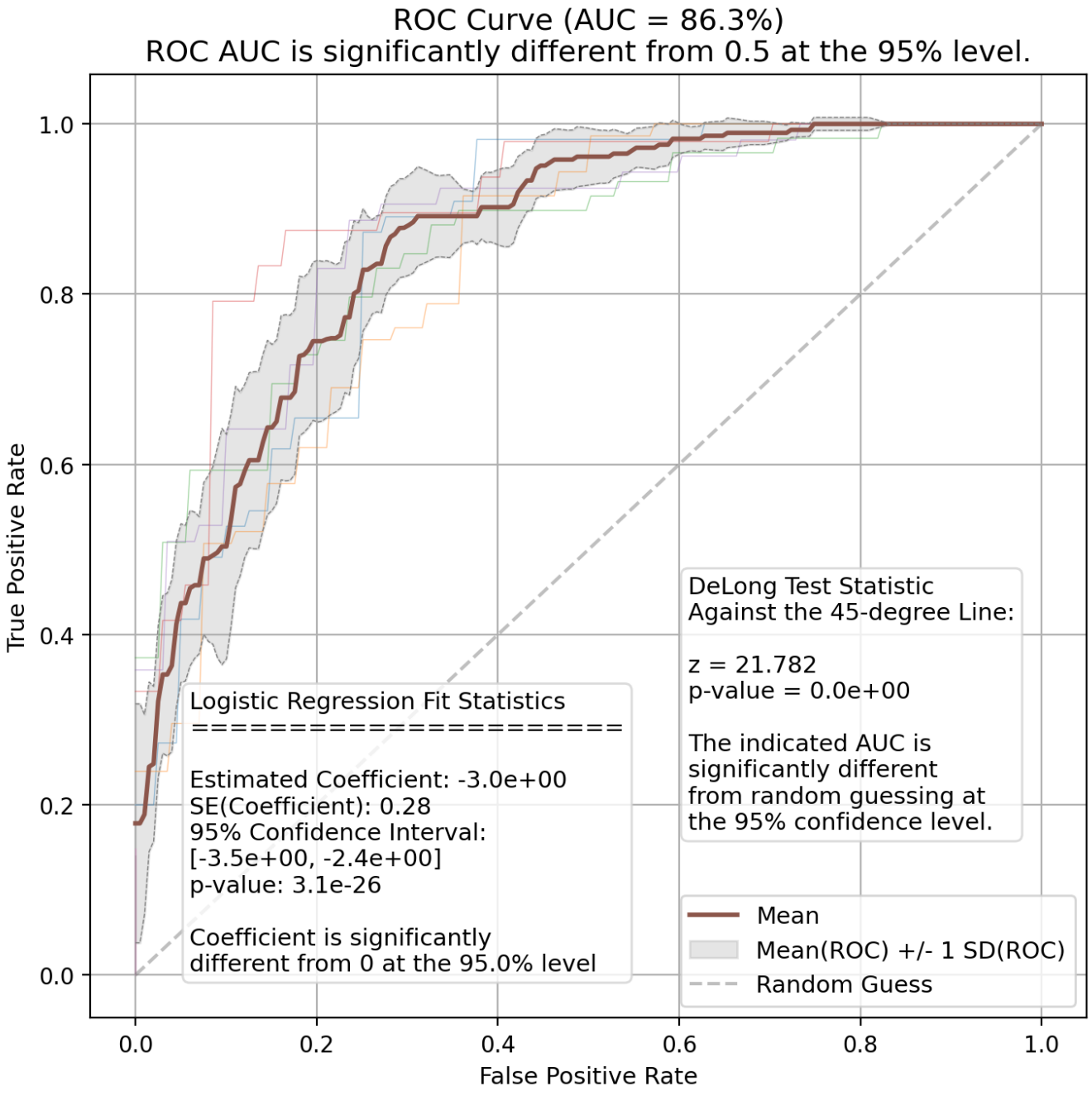
Perimeter Error - Empirical CDF Plot



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

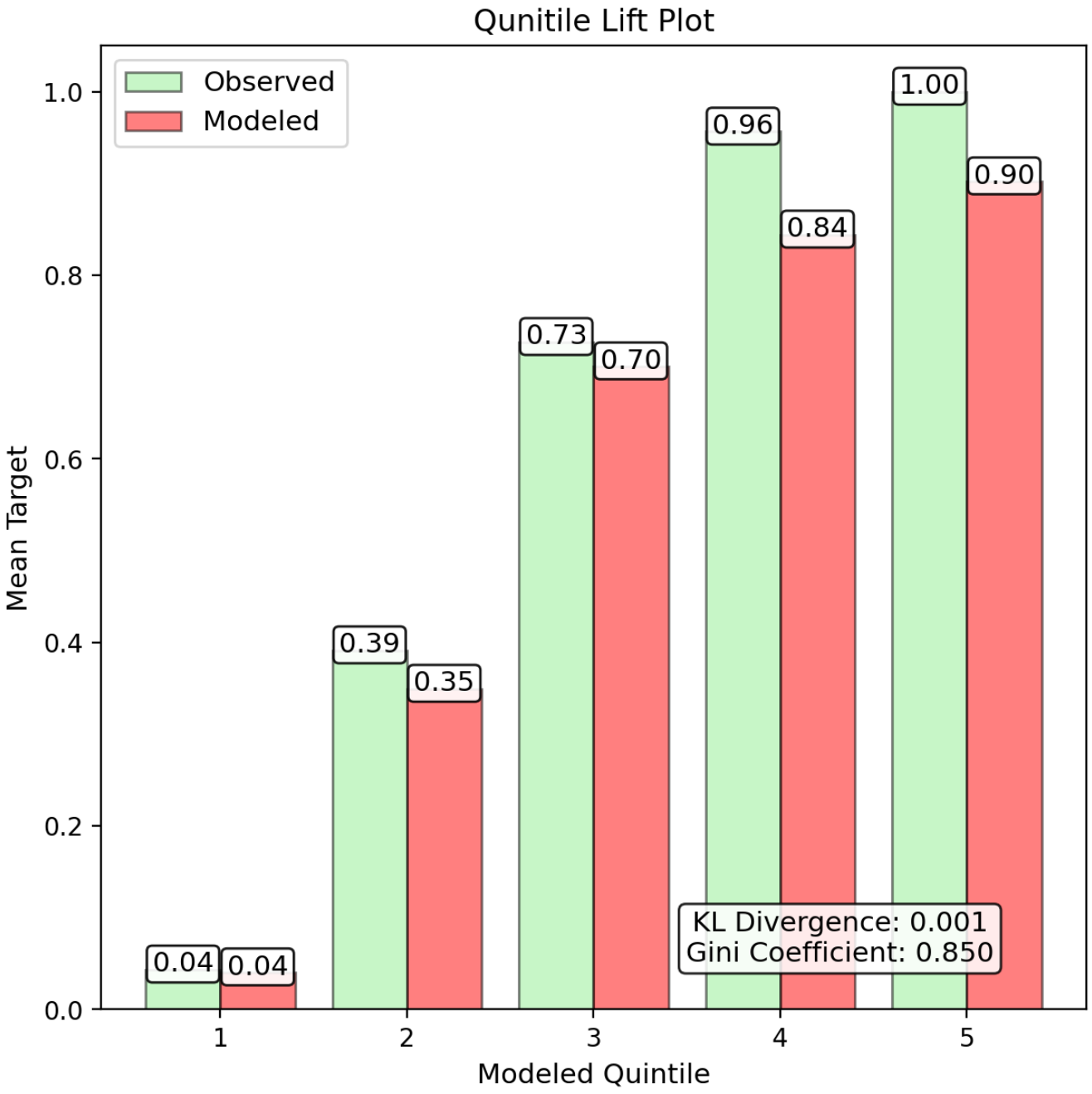


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

distribution to get the p-value.

Univariate Report

Perimeter Error - Quintile Lift



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

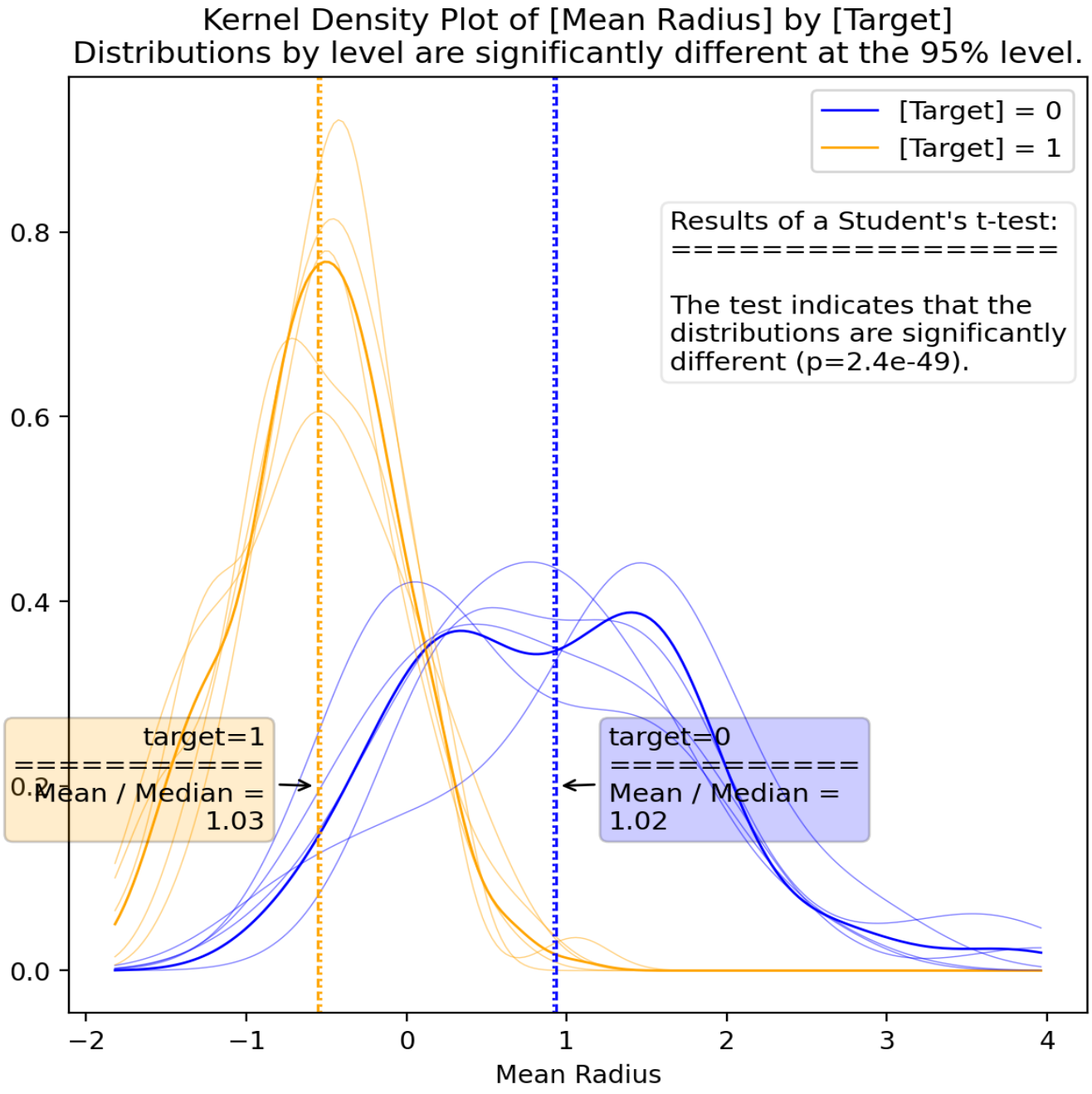
Univariate Report

Mean Radius - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Agg. Mean	Agg. SD
Fitted Coef.	-3.2e+00	-3.4e+00	-3.2e+00	-3.3e+00	-3.2e+00	-3.3e+00	7.8e-02
Fitted p-Value	3.3e-21	1.8e-20	7.1e-21	1.5e-21	2.4e-21	5.8e-26	6.6e-21
Fitted Std. Err.	0.338	0.364	0.343	0.349	0.342	0.310	0.010
Conf. Int. Lower	-3.9e+00	-4.1e+00	-3.9e+00	-4.0e+00	-3.9e+00	-3.9e+00	9.7e-02
Conf. Int. Upper	-2.5e+00	-2.7e+00	-2.5e+00	-2.6e+00	-2.6e+00	-2.7e+00	6.0e-02
Train Accuracy	85.0%	86.8%	85.9%	85.9%	86.0%	85.9%	0.6%
Val Accuracy	89.5%	82.8%	86.0%	85.9%	85.5%	37.7%	2.4%
Train AUC	84.0%	86.9%	85.8%	85.7%	85.8%	85.7%	1.0%
Val AUC	90.6%	79.4%	85.2%	85.6%	85.1%	50.0%	4.0%
Train F1	88.2%	88.8%	88.5%	88.8%	88.6%	88.6%	0.2%
Test F1	90.2%	87.9%	88.9%	87.5%	88.5%	0.0%	1.0%
Train Precision	89.0%	91.2%	90.7%	91.2%	90.6%	90.5%	0.9%
Val Precision	97.9%	88.6%	89.7%	87.5%	90.2%	0.0%	4.1%
Train Recall	87.4%	86.5%	86.3%	86.6%	86.7%	86.7%	0.4%
Val Recall	83.6%	87.3%	88.1%	87.5%	86.8%	0.0%	1.8%
Train MCC	67.6%	72.9%	70.5%	70.2%	70.7%	70.4%	1.9%
Val MCC	80.1%	58.1%	70.1%	71.3%	69.2%	0.0%	7.8%
Train Log-Loss	5.41	4.76	5.08	5.07	5.04	5.07	0.23
Val Log-Loss	3.79	6.19	5.04	5.09	5.21	22.45	0.85

Univariate Report

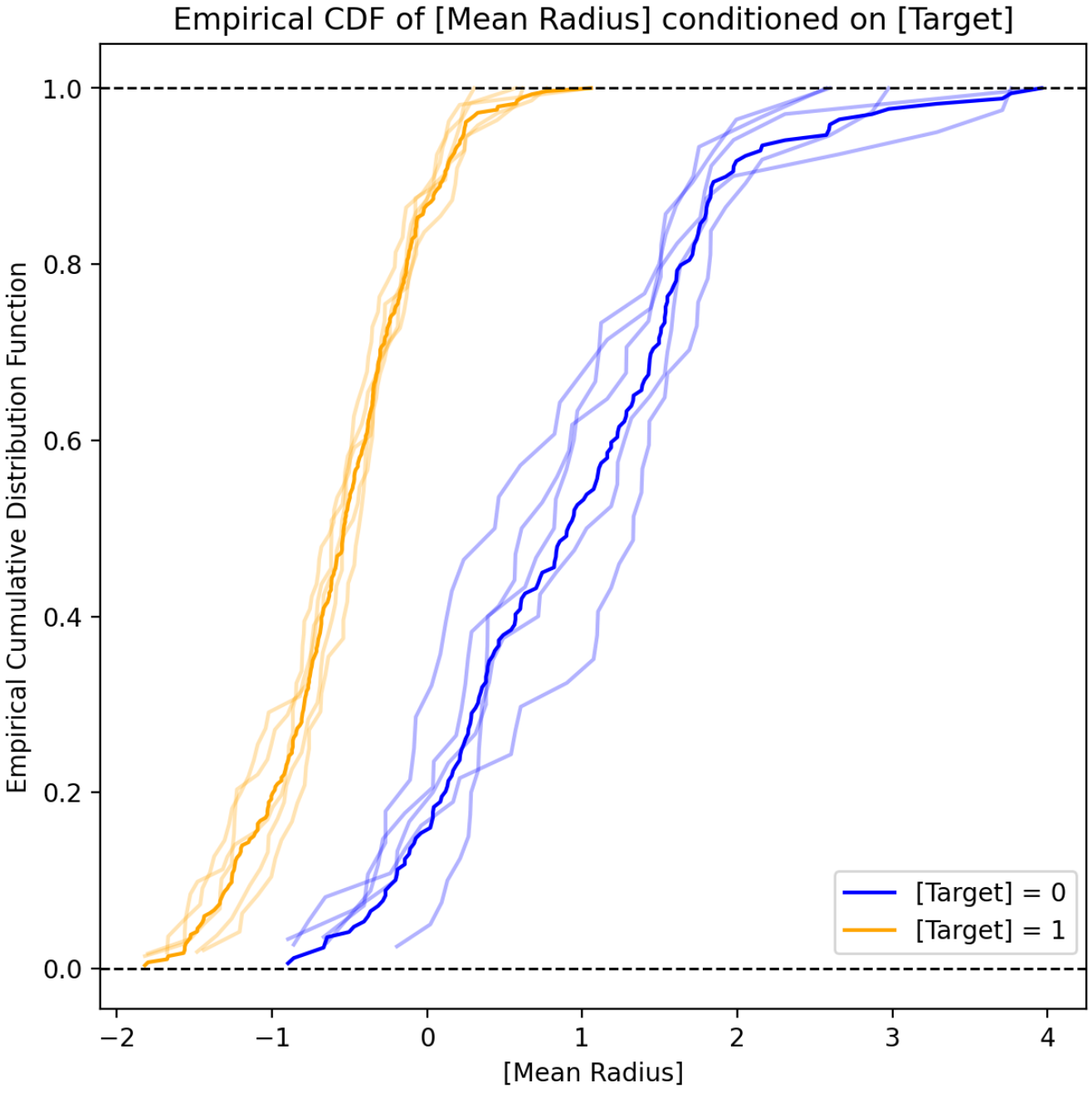
Mean Radius - Kernel Density Plot



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

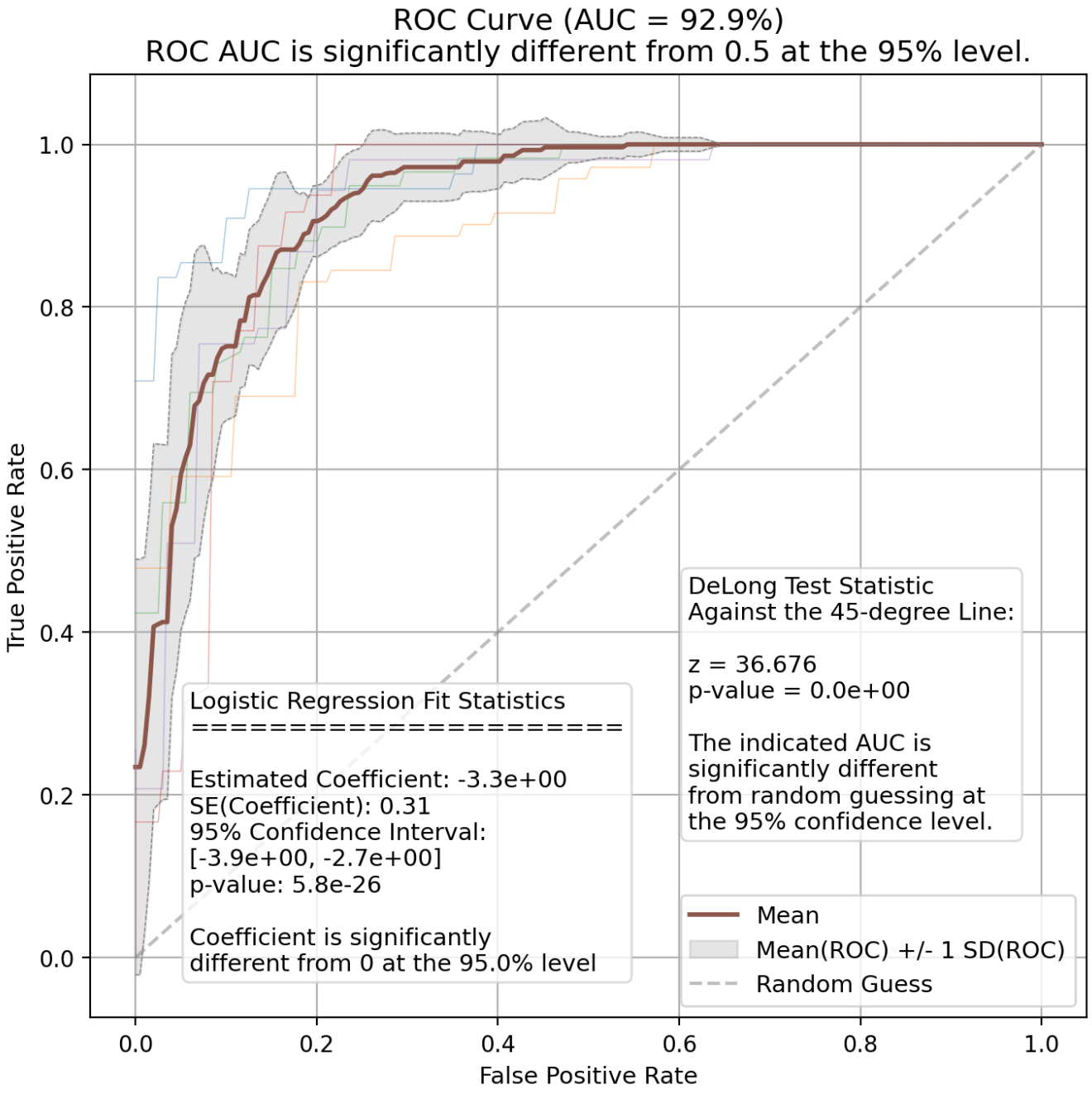
Mean Radius - Empirical CDF Plot



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 Radius - ROC Curve

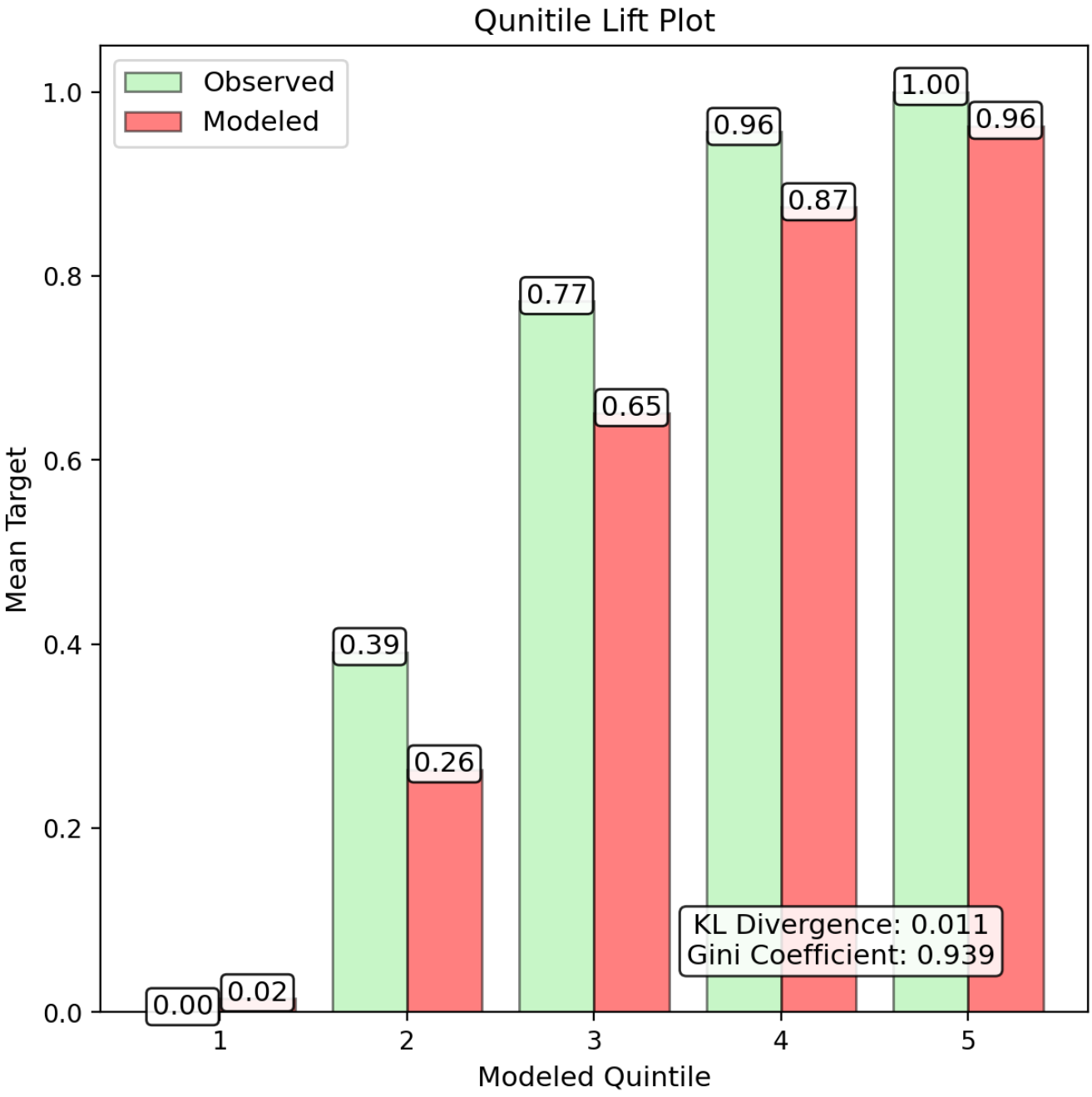


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept, and no other features. Annotations are on the plot to help understand the results of the model, including the coefficient, standard error, and p-value for the feature variable. The cross-validation folds are used to create the grey region around the mean ROC curve to help understand the variability of the data. Significance of the ROC curve is determined based on the method from DeLong et al. (1988). In brief, the AUC is assumed to be normally distributed, and the z-score is calculated based on the AUC and the standard error. This z-score is compared to a +/- two standard deviations from a standard normal

distribution to get the p-value.

Univariate Report

Mean Radius - 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.