

Cancer Model Univariate Analysis Report

2024-02-15

Overview

Cancer Model Univariate Analysis Report

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

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

	Accuracy	Precision	Recall	AUC	F1	MCC	Ave.
Worst Area	93.0%	90.9%	98.6%	91.2%	94.6%	85.2%	92.2%
Pc1	92.1%	94.3%	93.0%	91.8%	93.6%	83.3%	91.3%
Worst Radius	90.4%	91.7%	93.0%	89.5%	92.3%	79.4%	89.4%
Worst Perimeter	89.5%	91.5%	91.5%	88.8%	91.5%	77.6%	88.4%
Mean Concave Points	88.6%	93.9%	87.3%	89.0%	90.5%	76.6%	87.7%
Mean Area	88.6%	88.2%	94.4%	86.7%	91.2%	75.5%	87.4%
Mean Concavity	87.7%	91.3%	88.7%	87.4%	90.0%	74.2%	86.5%
Mean Radius	86.8%	90.0%	88.7%	86.2%	89.4%	72.1%	85.5%
Mean Perimeter	86.8%	90.0%	88.7%	86.2%	89.4%	72.1%	85.5%
Area Error	86.8%	85.9%	94.4%	84.4%	89.9%	71.7%	85.5%
Worst Concave Points	84.2%	96.5%	77.5%	86.4%	85.9%	70.6%	83.5%
Worst Concavity	84.2%	89.6%	84.5%	84.1%	87.0%	67.2%	82.8%
Perimeter Error	82.5%	83.1%	90.1%	80.0%	86.5%	62.0%	80.7%
Radius Error	80.7%	82.7%	87.3%	78.5%	84.9%	58.3%	78.7%
Mean Compactness	78.9%	86.2%	78.9%	79.0%	82.4%	56.7%	77.0%
Worst Compactness	75.4%	83.1%	76.1%	75.2%	79.4%	49.4%	73.1%
Worst Smoothness	74.6%	85.0%	71.8%	75.5%	77.9%	49.4%	72.4%
Mean Texture	71.9%	81.0%	71.8%	72.0%	76.1%	42.8%	69.3%
Concavity Error	70.2%	75.3%	77.5%	67.8%	76.4%	36.0%	67.2%
Mean Smoothness	67.5%	80.4%	63.4%	68.9%	70.9%	36.6%	64.6%

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

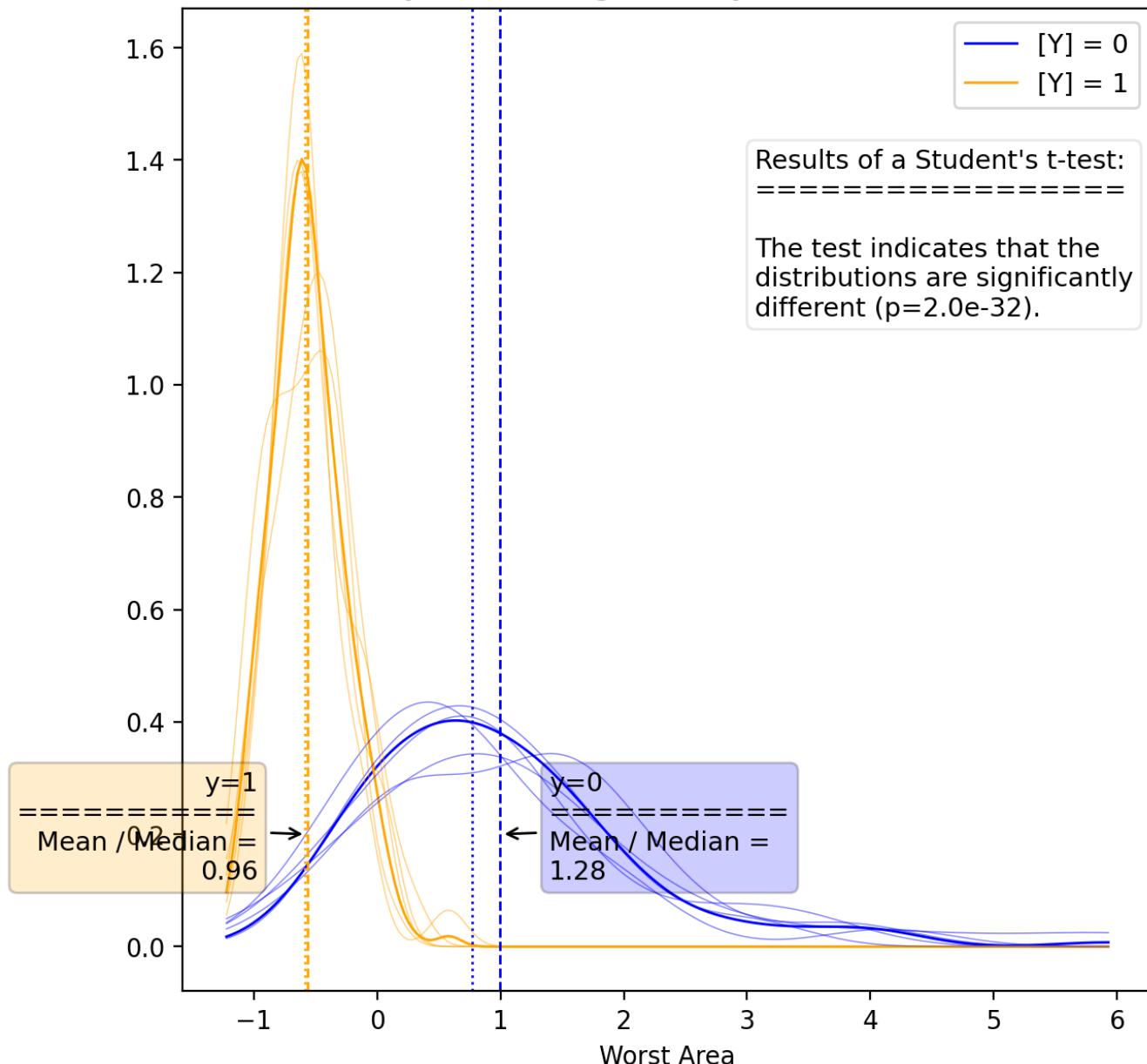
Worst Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-5.58	-5.68	-5.62	-5.82	-5.67	-5.67	0.09
Fitted p-Value	1.4e-16	1.8e-15	4.1e-15	3.3e-15	1.1e-15	3.9e-19	1.6e-15
Fitted Std. Err.	0.675	0.713	0.715	0.739	0.707	0.634	0.023
Conf. Int. Lower	-6.90	-7.07	-7.02	-7.27	-7.05	-6.91	0.13
Conf. Int. Upper	-4.25	-4.28	-4.22	-4.37	-4.28	-4.43	0.06
Train Accuracy	92.2%	90.5%	91.4%	91.6%	91.4%	91.5%	0.6%
Val Accuracy	89.7%	91.1%	90.5%	91.0%	92.1%	93.0%	0.9%
Train AUC	90.5%	88.6%	89.7%	89.7%	89.7%	89.7%	0.7%
Val AUC	88.1%	91.0%	86.5%	89.6%	88.8%	91.2%	1.7%
Train F1	94.0%	92.8%	93.3%	93.6%	93.3%	93.5%	0.4%
Test F1	91.9%	92.0%	93.3%	93.0%	94.4%	94.6%	1.0%
Train Precision	91.1%	91.0%	89.8%	90.3%	89.7%	90.4%	0.6%
Val Precision	89.5%	90.9%	89.1%	90.9%	91.3%	90.9%	1.0%
Train Recall	97.2%	94.7%	97.0%	97.1%	97.1%	96.7%	1.1%
Val Recall	94.4%	93.0%	98.0%	95.2%	97.7%	98.6%	2.1%
Train MCC	83.3%	78.7%	81.8%	82.0%	81.8%	81.7%	1.7%
Val MCC	77.9%	82.1%	78.2%	80.7%	81.5%	85.2%	1.9%
Train Log-Loss	2.80	3.44	3.10	3.03	3.11	3.07	0.23
Val Log-Loss	3.73	3.19	3.41	3.23	2.86	2.53	0.32

Univariate Report

Worst Area - Kernel Density Plot

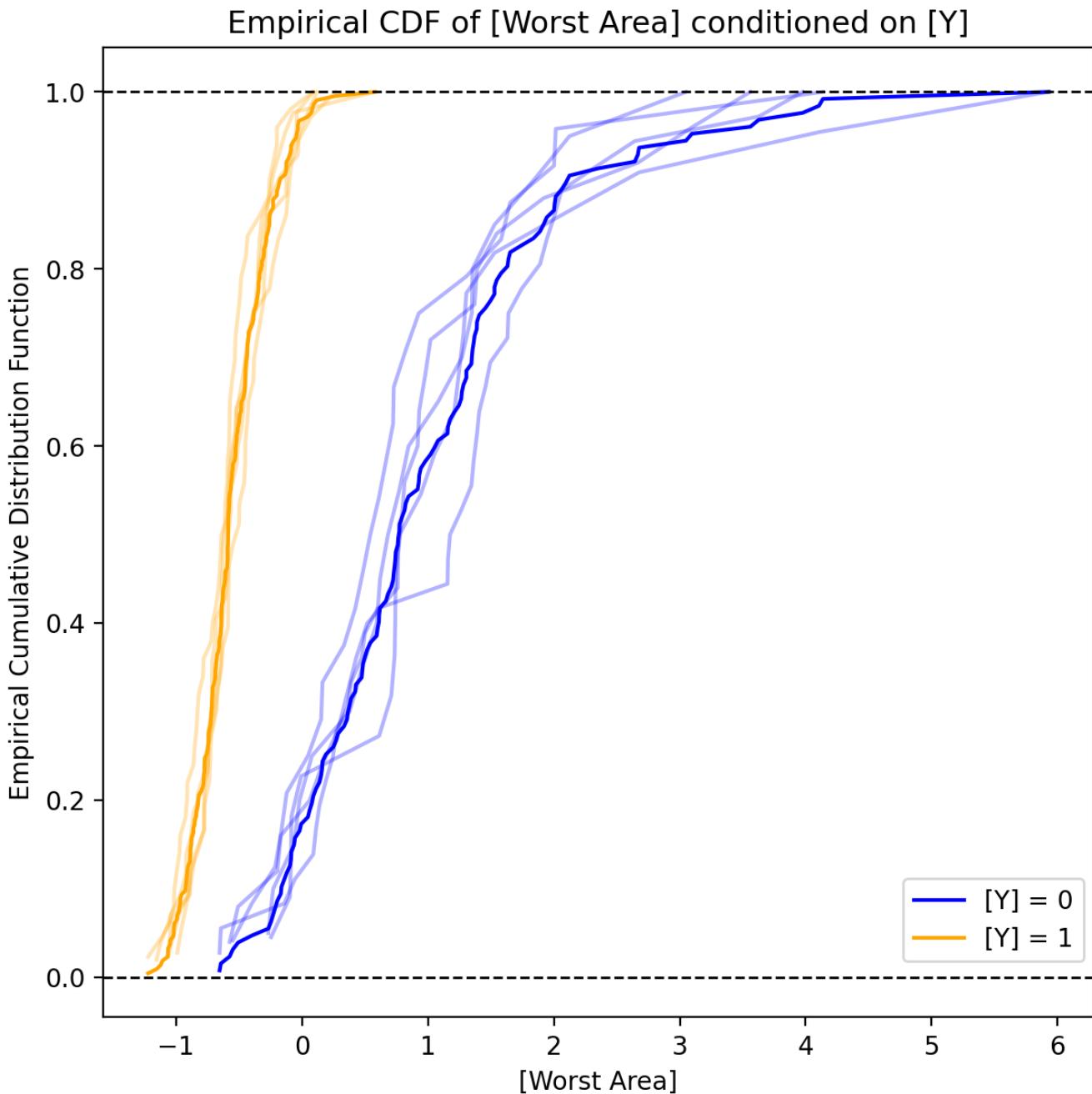
Kernel Density Plot of [Worst Area] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

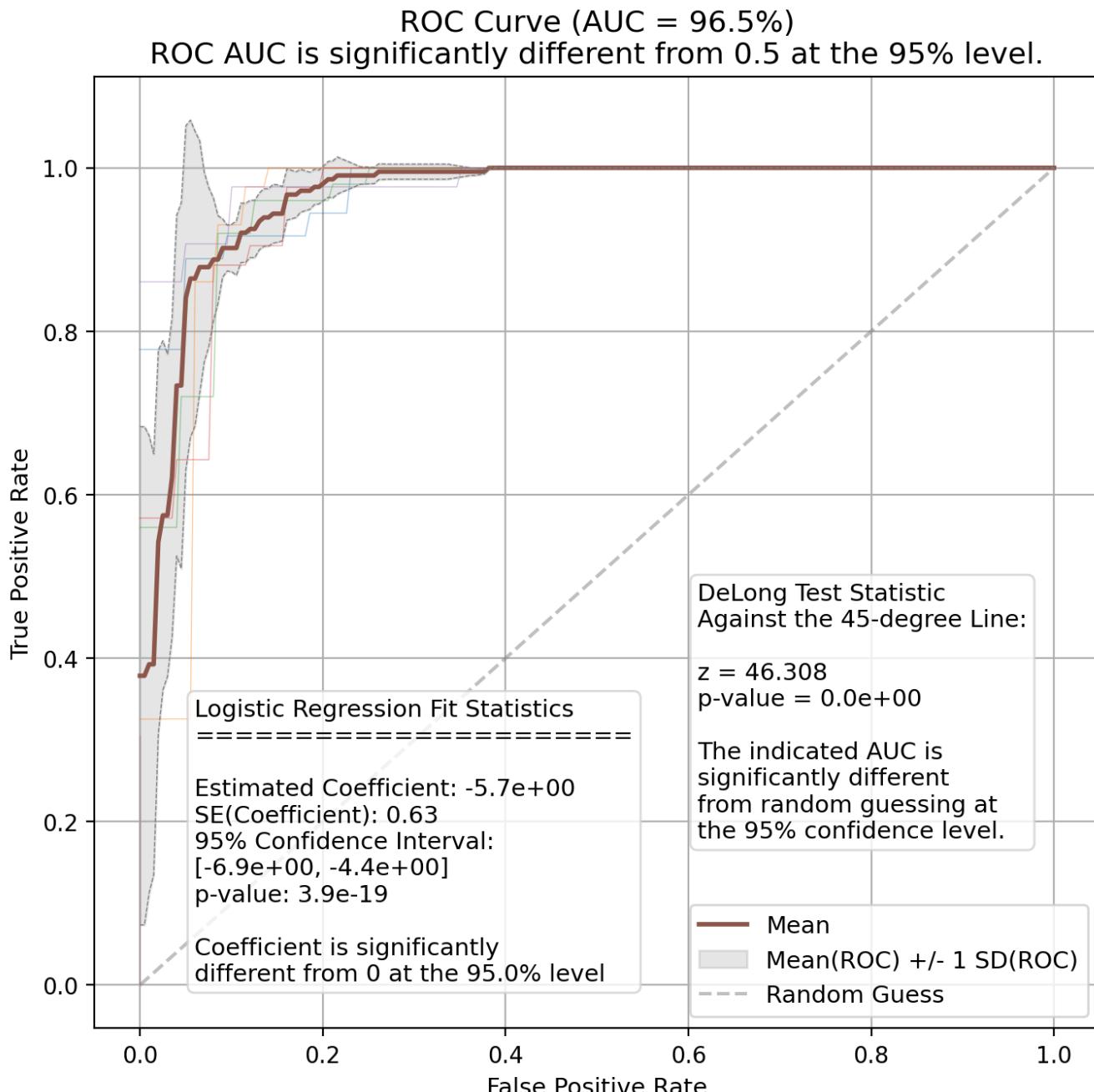
Worst Area - Empirical CDF Plot



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

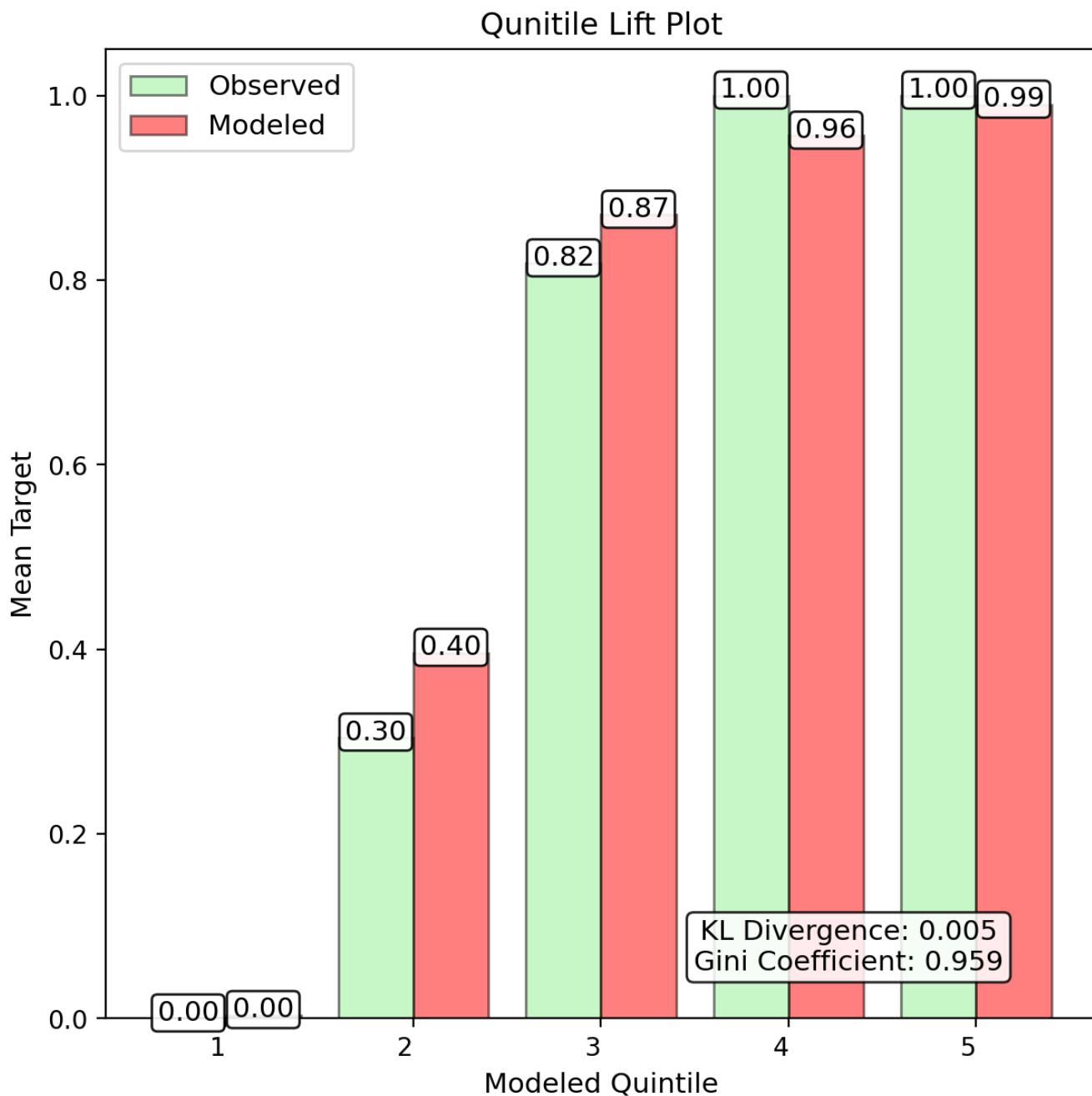


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

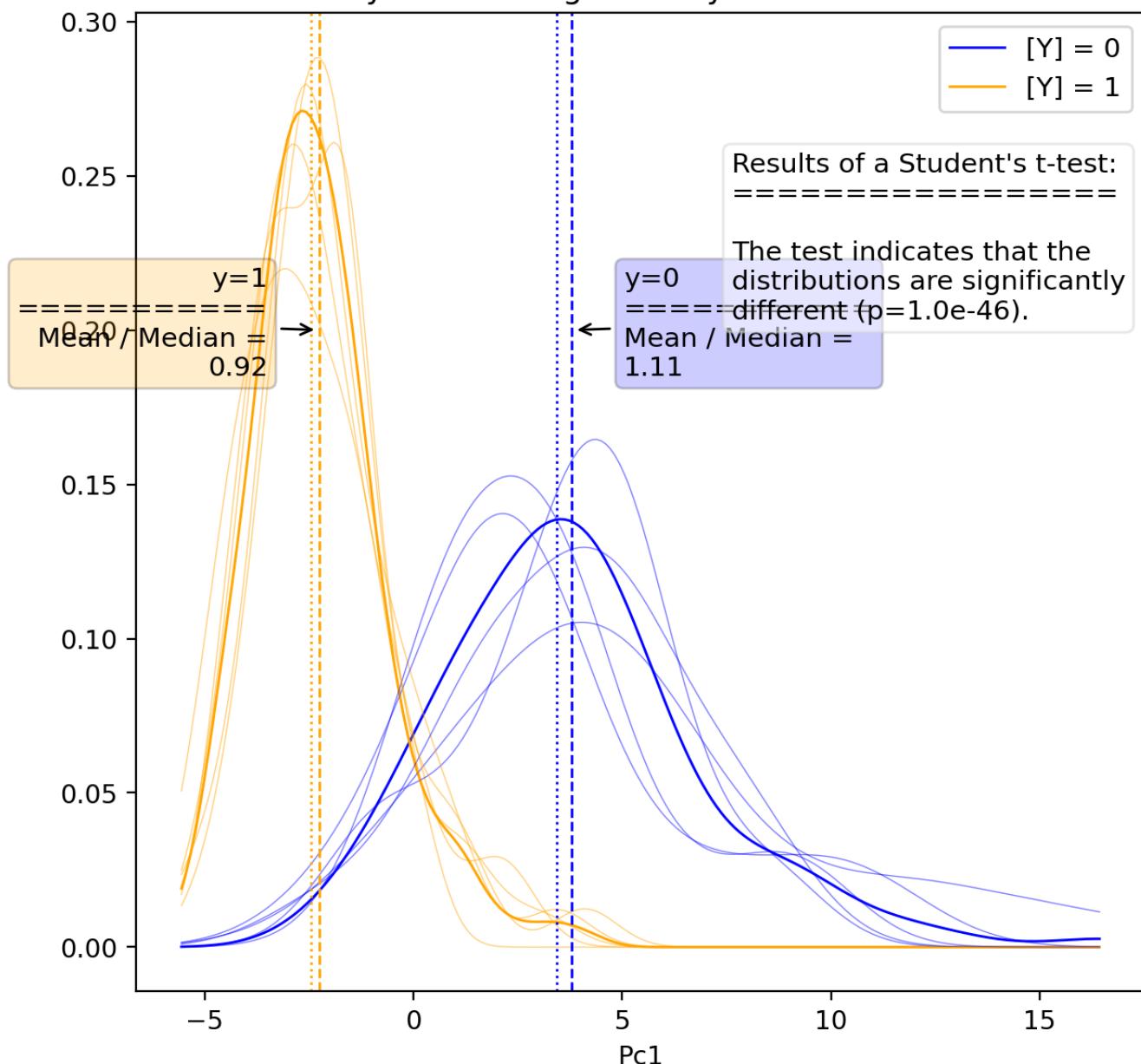
Pc1 - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-1.21	-1.27	-1.24	-1.27	-1.10	-1.21	0.07
Fitted p-Value	6.7e-18	3.8e-16	4.1e-16	3.2e-16	4.9e-18	9.2e-21	2.0e-16
Fitted Std. Err.	0.140	0.156	0.153	0.156	0.127	0.130	0.013
Conf. Int. Lower	-1.48	-1.58	-1.54	-1.58	-1.35	-1.47	0.10
Conf. Int. Upper	-0.934	-0.964	-0.943	-0.968	-0.852	-0.959	0.047
Train Accuracy	90.5%	91.6%	91.8%	92.0%	90.3%	91.5%	0.8%
Val Accuracy	91.4%	88.6%	90.5%	89.6%	96.8%	92.1%	3.2%
Train AUC	90.3%	91.8%	91.7%	91.6%	90.2%	91.3%	0.8%
Val AUC	92.2%	88.6%	89.8%	90.0%	96.3%	91.8%	3.0%
Train F1	92.3%	93.4%	93.2%	93.6%	92.0%	93.1%	0.7%
Test F1	92.8%	89.4%	92.9%	91.4%	97.7%	93.6%	3.1%
Train Precision	93.6%	95.7%	94.4%	94.1%	93.4%	94.3%	0.9%
Val Precision	97.0%	90.5%	93.9%	94.9%	97.7%	94.3%	2.9%
Train Recall	91.0%	91.2%	92.1%	93.0%	90.6%	92.1%	1.0%
Val Recall	88.9%	88.4%	92.0%	88.1%	97.7%	93.0%	4.0%
Train MCC	79.8%	82.0%	82.8%	82.9%	79.7%	82.0%	1.6%
Val MCC	82.6%	77.1%	78.7%	78.5%	92.7%	83.3%	6.3%
Train Log-Loss	3.44	3.03	2.97	2.89	3.50	3.07	0.28
Val Log-Loss	3.11	4.11	3.41	3.77	1.14	2.85	1.16

Univariate Report

Pc1 - Kernel Density Plot

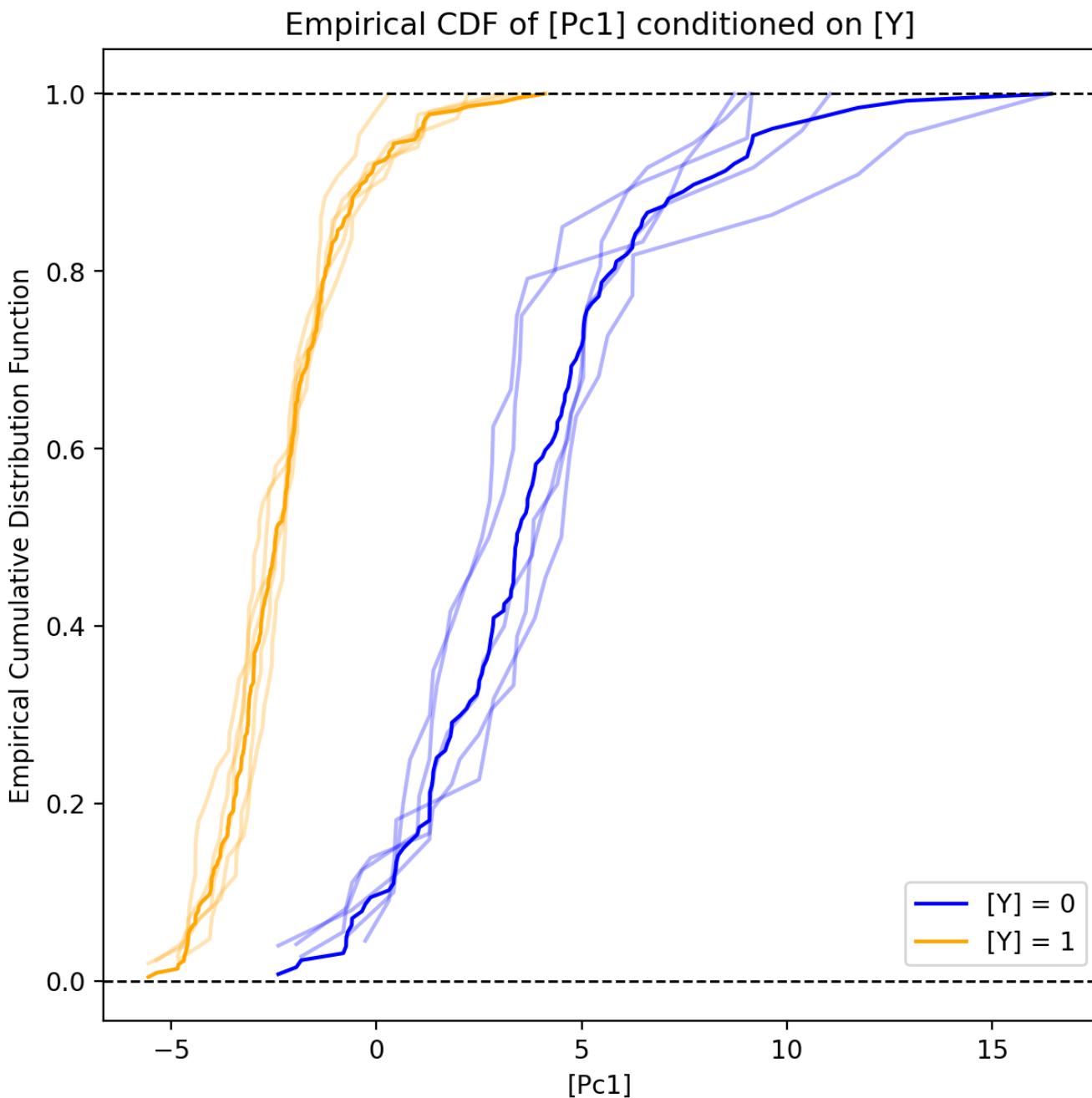
Kernel Density Plot of [Pc1] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

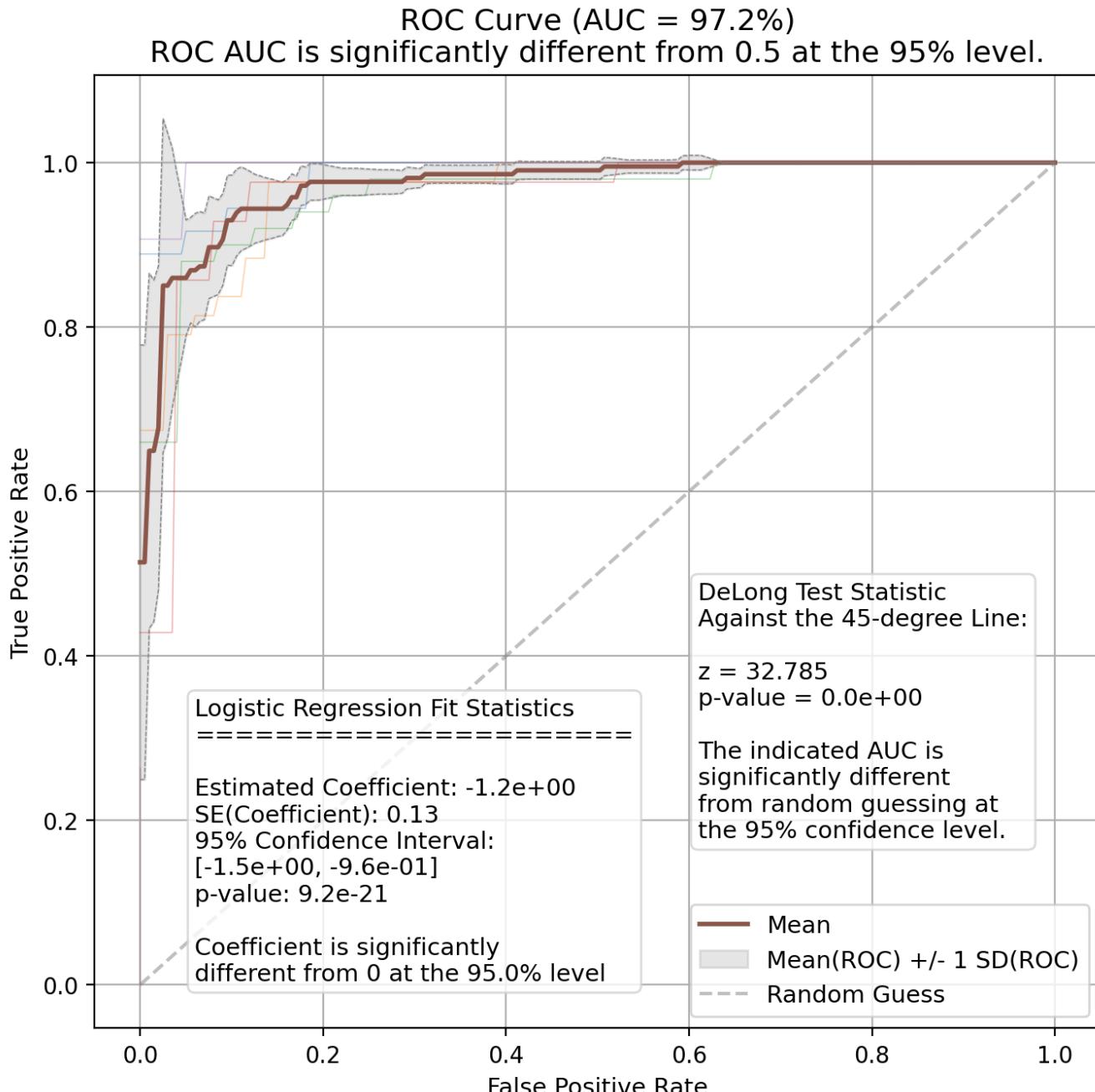
Pc1 - Empirical CDF Plot



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

Pc1 - ROC Curve

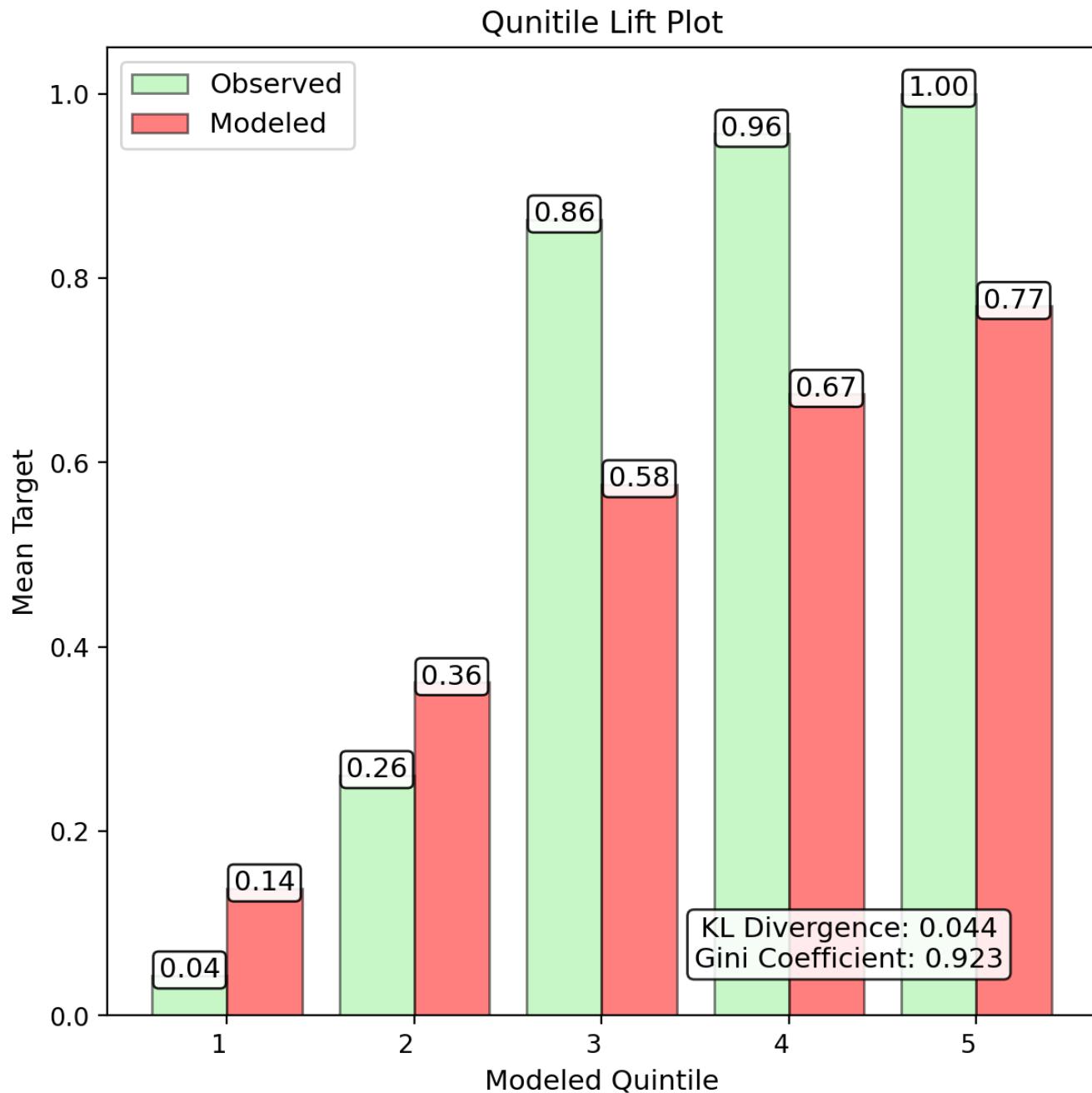


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Pc1 - 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

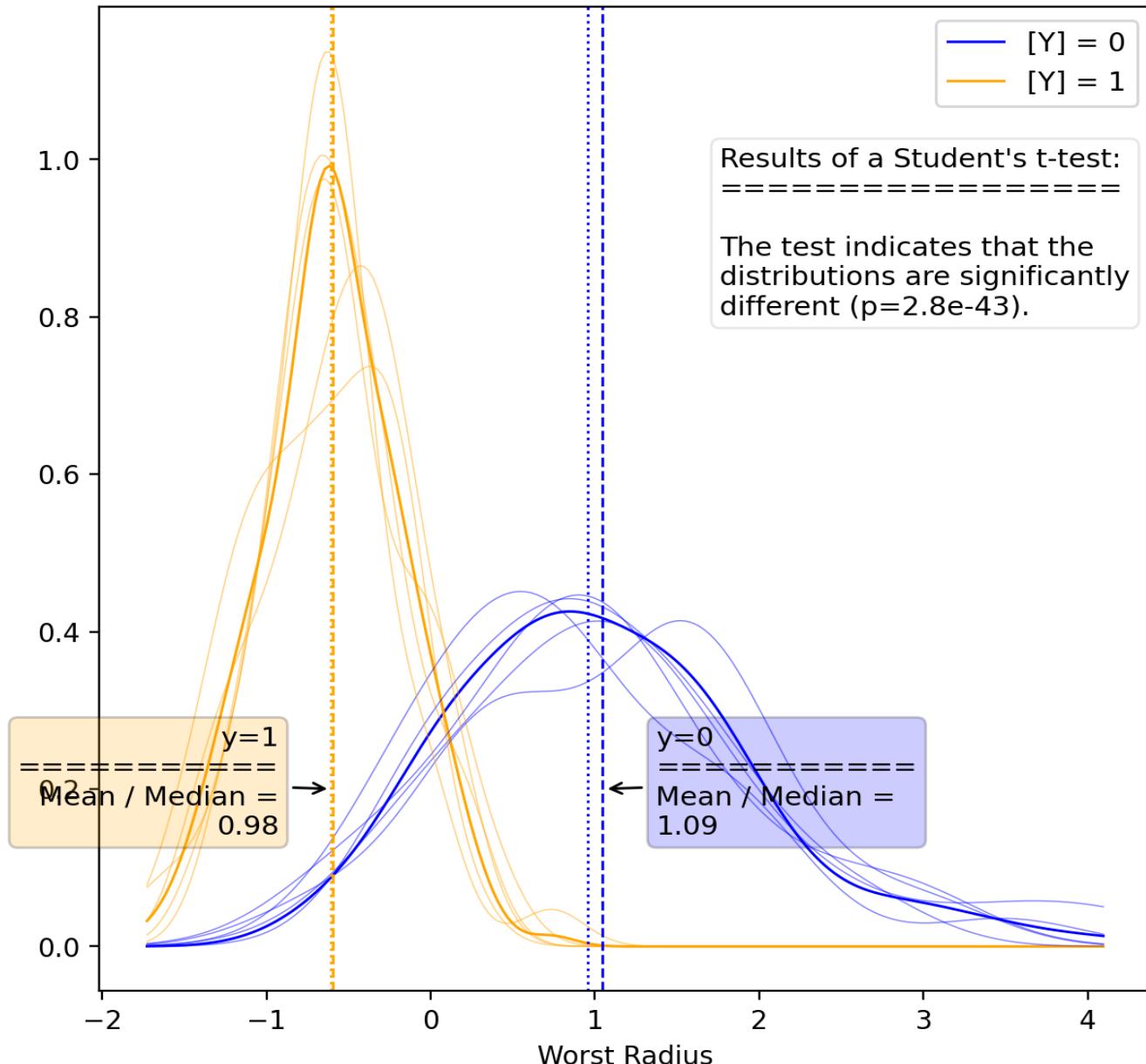
Worst Radius - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-5.24	-5.44	-5.02	-5.60	-5.26	-5.30	0.22
Fitted p-Value	3.1e-15	9.9e-14	1.4e-14	1.7e-13	2.1e-14	1.7e-17	7.1e-14
Fitted Std. Err.	0.664	0.731	0.652	0.760	0.688	0.623	0.045
Conf. Int. Lower	-6.54	-6.88	-6.30	-7.09	-6.61	-6.52	0.31
Conf. Int. Upper	-3.94	-4.01	-3.74	-4.11	-3.91	-4.08	0.14
Train Accuracy	90.5%	90.5%	90.6%	91.2%	90.3%	90.9%	0.4%
Val Accuracy	91.4%	88.6%	93.2%	88.1%	93.7%	90.4%	2.6%
Train AUC	89.9%	90.4%	89.9%	90.8%	89.5%	90.5%	0.5%
Val AUC	93.1%	88.9%	90.7%	87.2%	91.3%	89.5%	2.2%
Train F1	92.4%	92.5%	92.4%	93.0%	92.2%	92.7%	0.3%
Test F1	92.5%	89.2%	95.1%	90.5%	95.5%	92.3%	2.8%
Train Precision	92.7%	94.5%	91.6%	93.5%	91.4%	93.4%	1.3%
Val Precision	100.0%	92.5%	92.5%	90.5%	93.3%	91.7%	3.6%
Train Recall	92.1%	90.6%	93.3%	92.4%	93.0%	92.1%	1.0%
Val Recall	86.1%	86.0%	98.0%	90.5%	97.7%	93.0%	5.9%
Train MCC	79.6%	79.5%	80.2%	81.3%	79.4%	80.7%	0.8%
Val MCC	83.8%	77.4%	84.5%	74.5%	85.2%	79.4%	4.8%
Train Log-Loss	3.44	3.44	3.37	3.16	3.50	3.28	0.13
Val Log-Loss	3.11	4.11	2.44	4.30	2.29	3.48	0.93

Univariate Report

Worst Radius - Kernel Density Plot

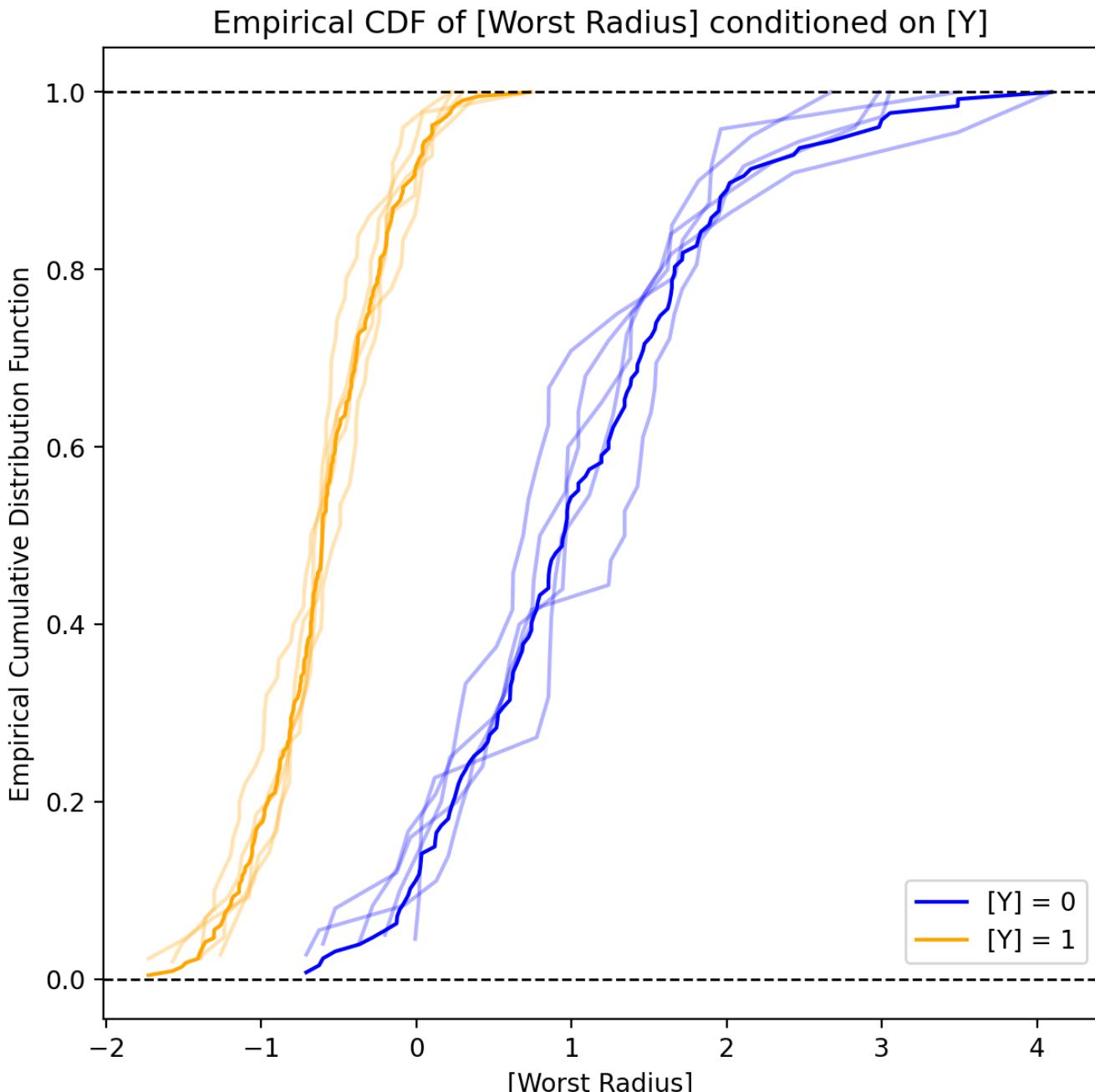
Kernel Density Plot of [Worst Radius] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

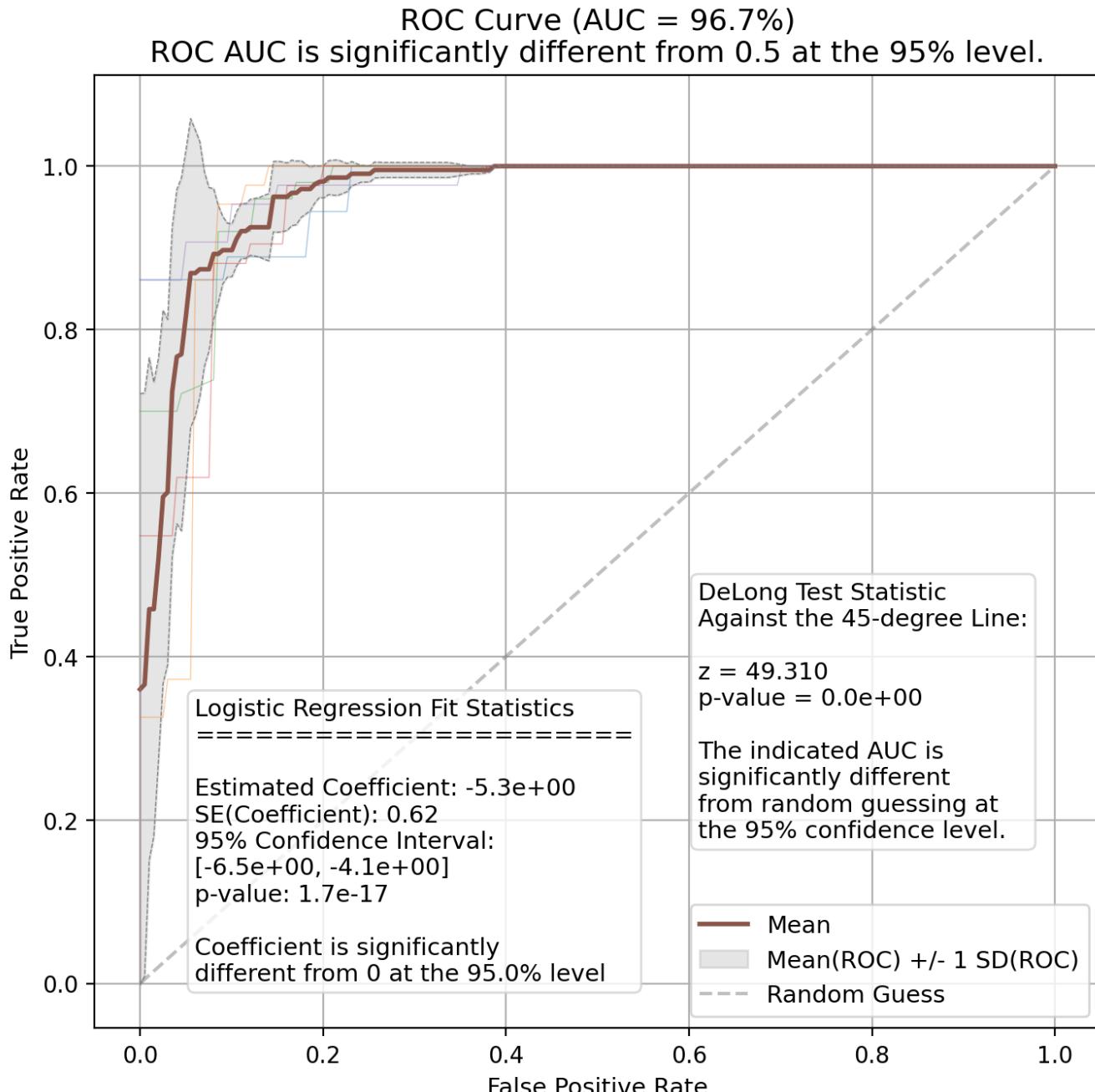
Worst Radius - Empirical CDF Plot



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

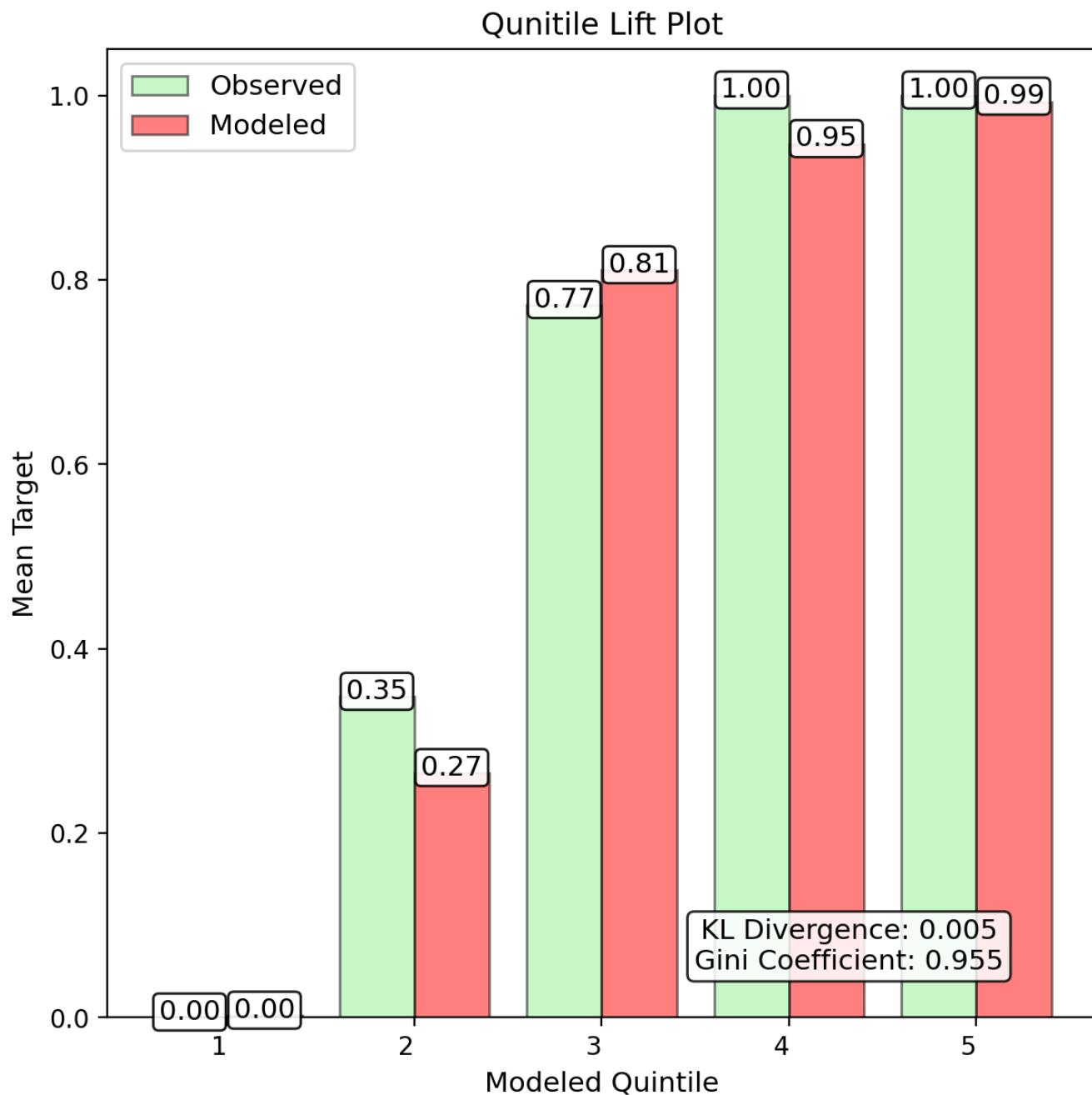


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

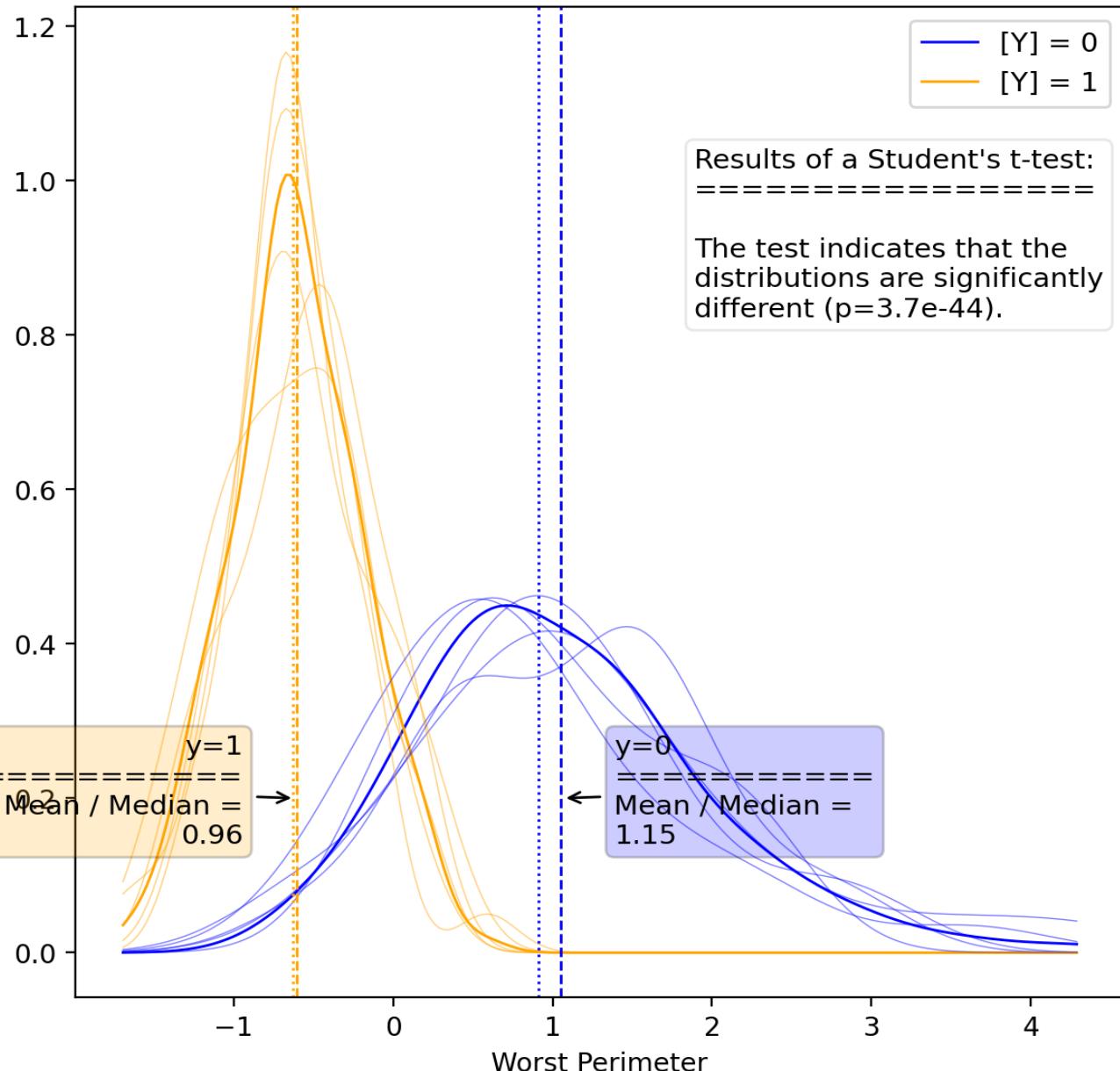
Worst Perimeter - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-5.57	-5.64	-5.40	-5.85	-5.34	-5.55	0.20
Fitted p-Value	4.8e-15	1.1e-13	3.4e-14	2.0e-13	1.0e-14	2.1e-17	8.4e-14
Fitted Std. Err.	0.711	0.759	0.712	0.796	0.691	0.654	0.043
Conf. Int. Lower	-6.96	-7.12	-6.80	-7.41	-6.70	-6.83	0.28
Conf. Int. Upper	-4.17	-4.15	-4.00	-4.29	-3.99	-4.27	0.12
Train Accuracy	90.8%	90.8%	91.8%	92.0%	91.0%	91.5%	0.5%
Val Accuracy	91.4%	88.6%	91.9%	89.6%	95.2%	89.5%	2.6%
Train AUC	90.5%	90.7%	91.5%	91.6%	90.6%	91.1%	0.5%
Val AUC	92.2%	89.1%	89.7%	89.2%	93.8%	88.8%	2.1%
Train F1	92.6%	92.9%	93.3%	93.6%	92.7%	93.2%	0.4%
Test F1	92.8%	88.9%	94.1%	91.6%	96.6%	91.5%	2.9%
Train Precision	93.7%	94.5%	93.8%	94.1%	92.9%	93.8%	0.6%
Val Precision	97.0%	94.7%	92.3%	92.7%	95.5%	91.5%	1.9%
Train Recall	91.6%	91.2%	92.7%	93.0%	92.4%	92.5%	0.8%
Val Recall	88.9%	83.7%	96.0%	90.5%	97.7%	91.5%	5.6%
Train MCC	80.5%	80.2%	82.7%	82.9%	81.0%	81.9%	1.2%
Val MCC	82.6%	77.9%	81.3%	77.9%	88.9%	77.6%	4.5%
Train Log-Loss	3.31	3.30	2.97	2.89	3.24	3.07	0.20
Val Log-Loss	3.11	4.11	2.92	3.77	1.72	3.79	0.92

Univariate Report

Worst Perimeter - Kernel Density Plot

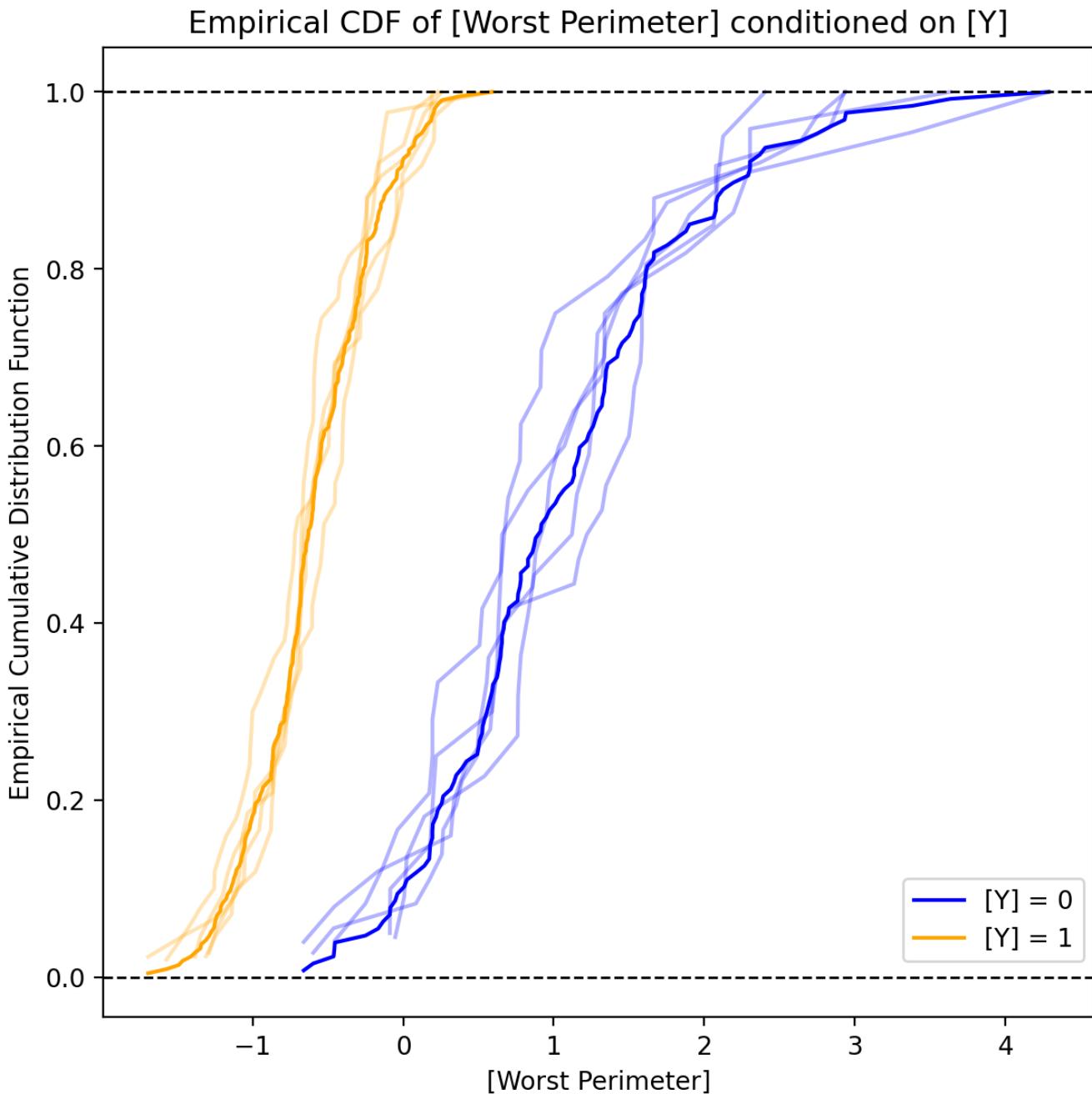
Kernel Density Plot of [Worst Perimeter] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

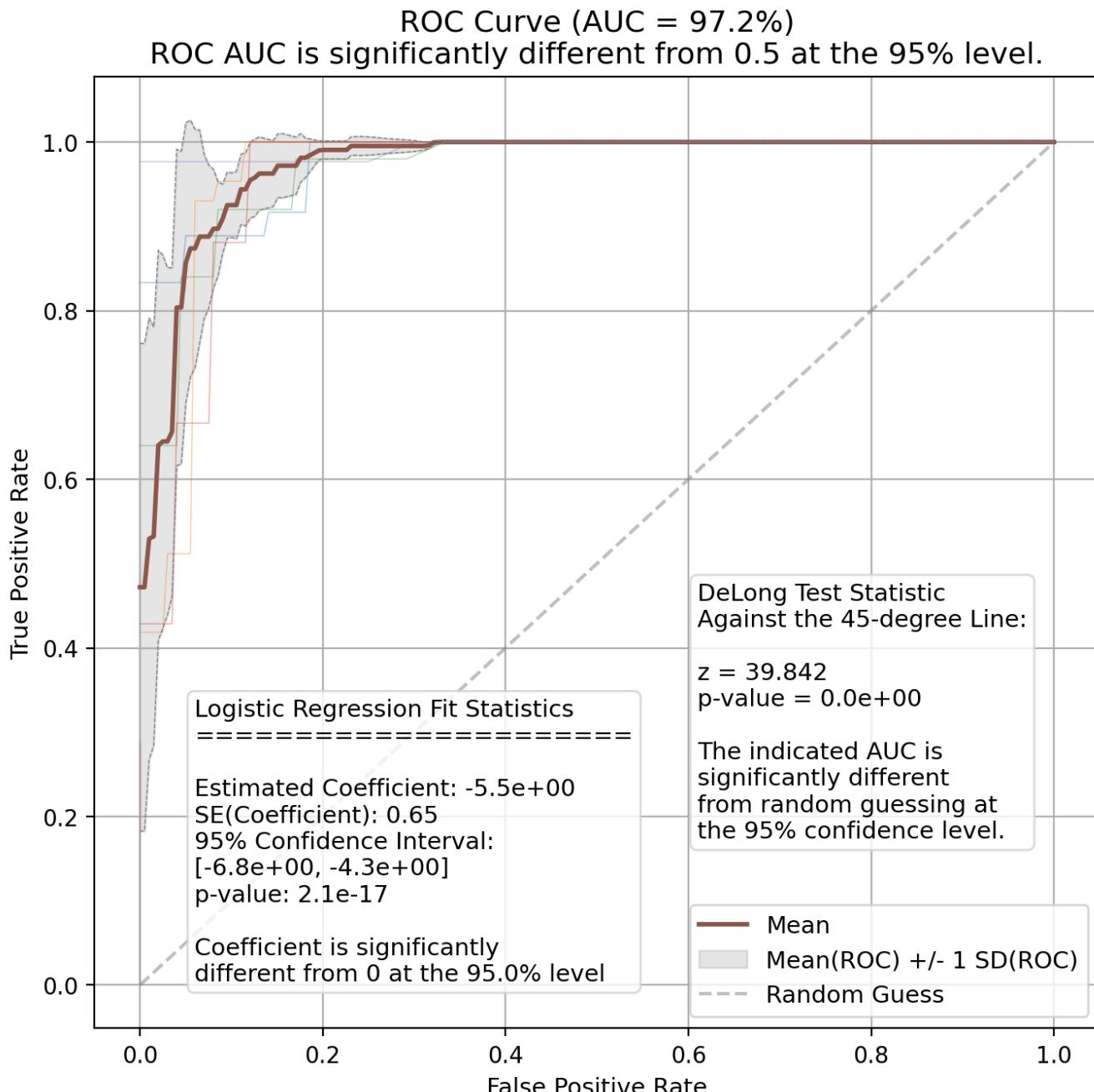
Worst Perimeter - Empirical CDF Plot



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

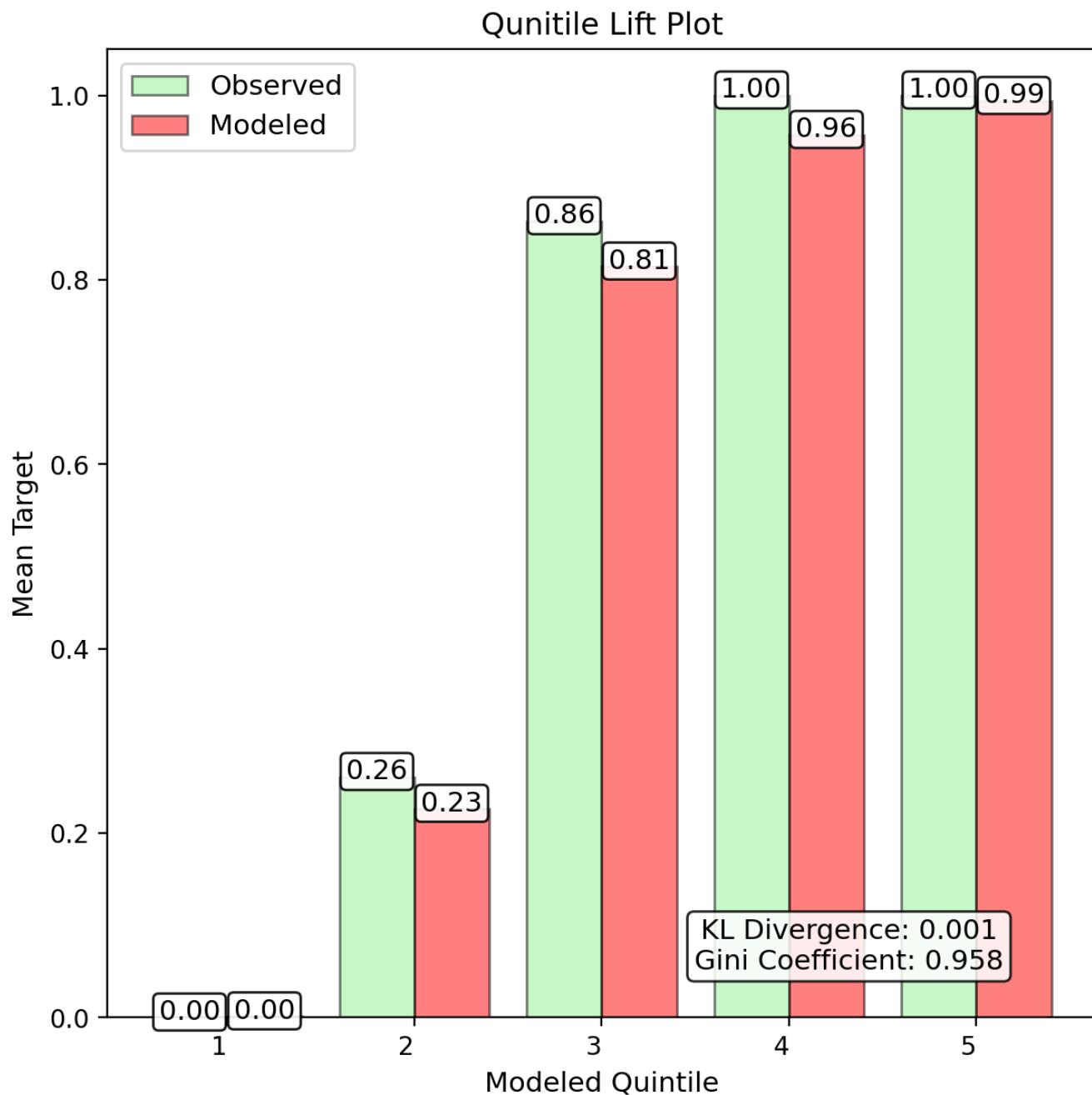


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

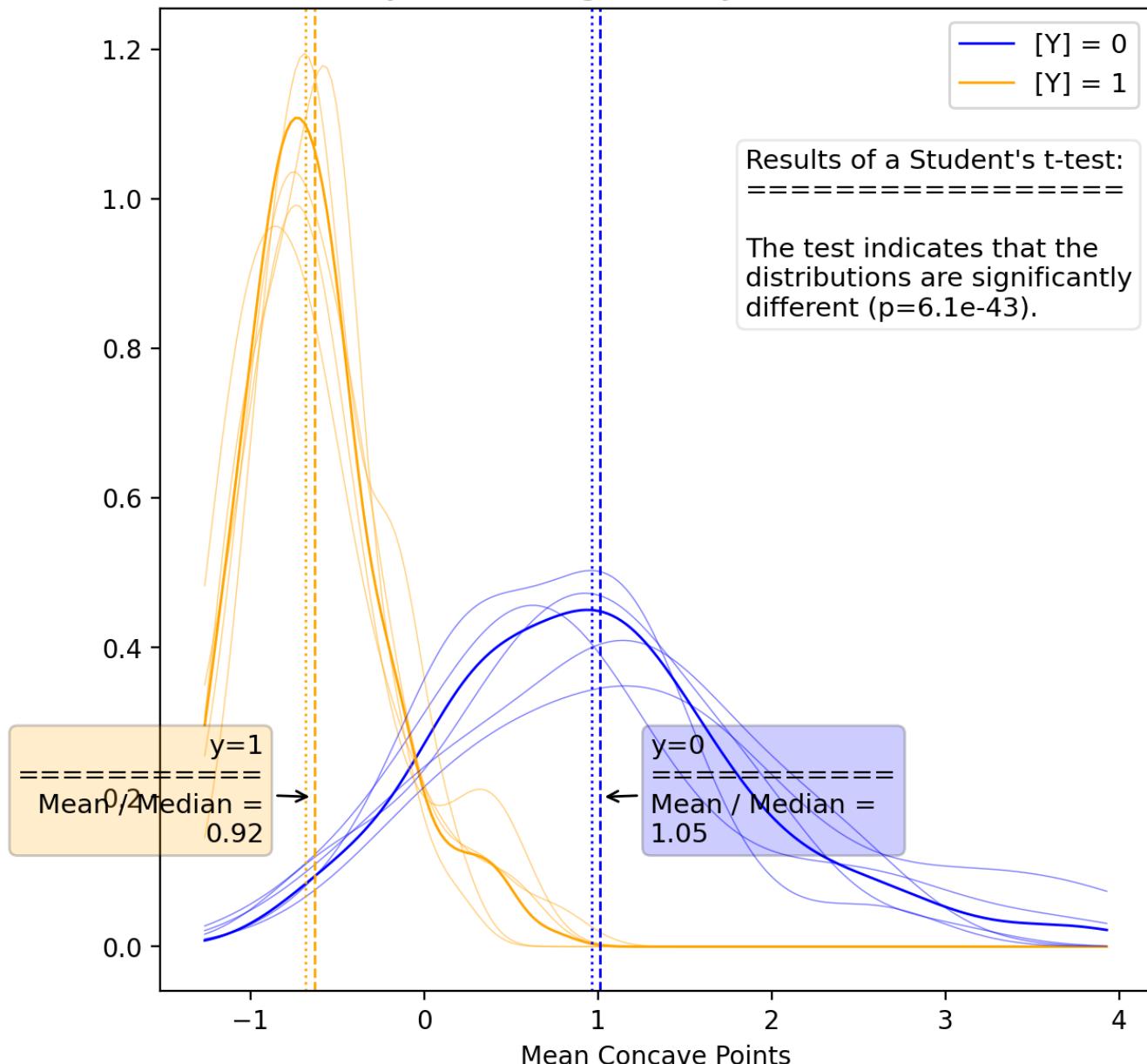
Mean Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-4.56	-4.98	-4.66	-4.42	-4.25	-4.55	0.27
Fitted p-Value	6.5e-18	2.6e-15	2.7e-16	1.1e-17	8.0e-18	6.4e-21	1.1e-15
Fitted Std. Err.	0.529	0.629	0.569	0.516	0.494	0.485	0.053
Conf. Int. Lower	-5.60	-6.21	-5.78	-5.43	-5.21	-5.50	0.38
Conf. Int. Upper	-3.52	-3.74	-3.54	-3.40	-3.28	-3.60	0.17
Train Accuracy	92.2%	92.7%	92.5%	91.6%	91.4%	91.8%	0.6%
Val Accuracy	91.4%	89.9%	90.5%	92.5%	95.2%	88.6%	2.1%
Train AUC	91.7%	92.6%	92.3%	91.5%	91.2%	91.5%	0.6%
Val AUC	93.1%	89.8%	89.8%	91.6%	93.8%	89.0%	1.9%
Train F1	93.8%	94.4%	93.9%	93.2%	92.9%	93.4%	0.6%
Test F1	92.5%	90.7%	92.9%	94.1%	96.6%	90.5%	2.2%
Train Precision	93.8%	95.8%	94.4%	94.6%	94.0%	94.3%	0.8%
Val Precision	100.0%	90.7%	93.9%	93.0%	95.5%	93.9%	3.5%
Train Recall	93.8%	93.0%	93.3%	91.9%	91.8%	92.5%	0.9%
Val Recall	86.1%	90.7%	92.0%	95.2%	97.7%	87.3%	4.4%
Train MCC	83.3%	84.3%	84.3%	82.3%	81.9%	82.6%	1.1%
Val MCC	83.8%	79.6%	78.7%	84.0%	88.9%	76.6%	4.1%
Train Log-Loss	2.80	2.61	2.70	3.03	3.11	2.96	0.21
Val Log-Loss	3.11	3.65	3.41	2.69	1.72	4.11	0.76

Univariate Report

Mean Concave Points - Kernel Density Plot

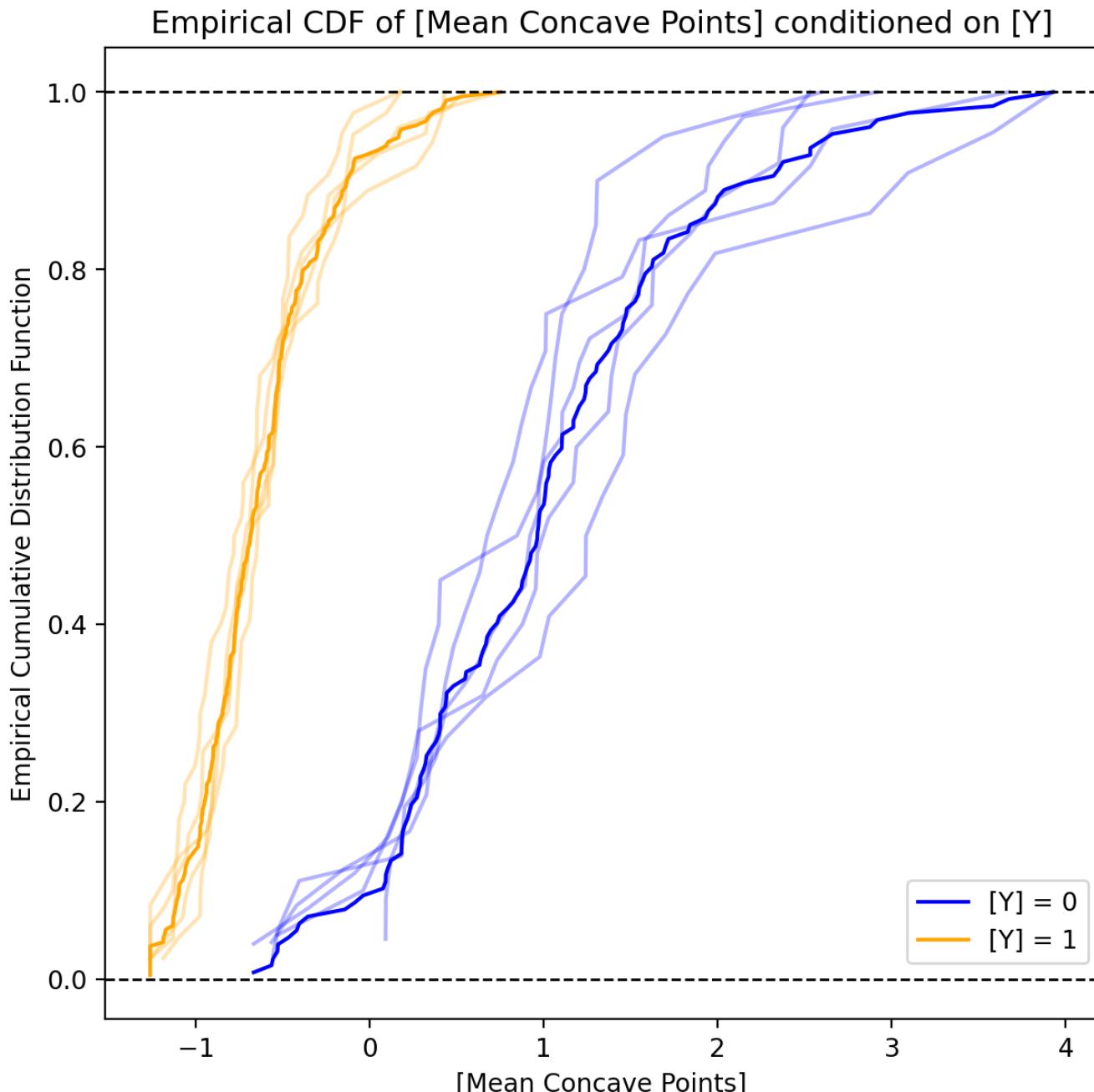
Kernel Density Plot of [Mean Concave Points] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

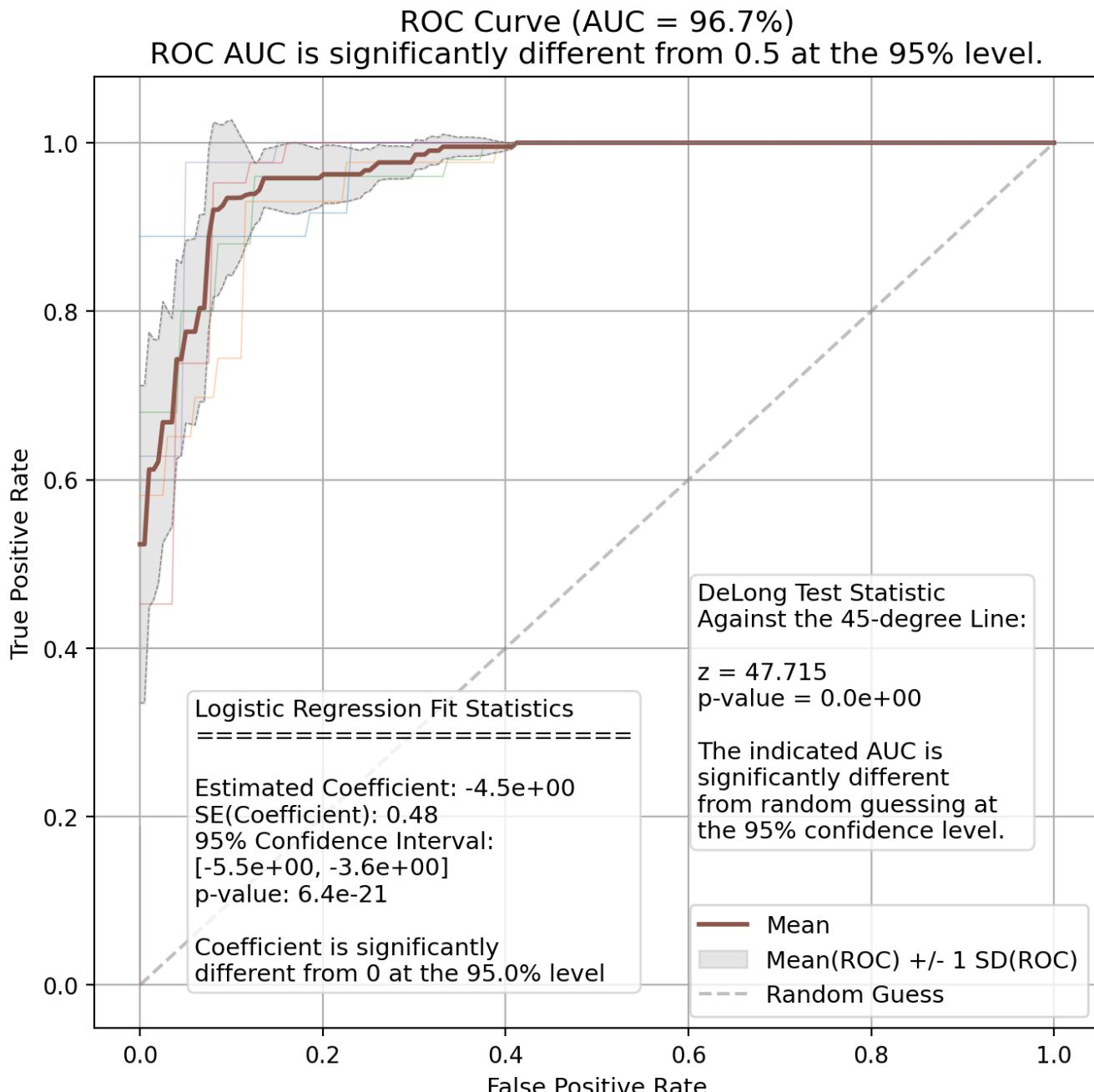
Mean Concave Points - Empirical CDF Plot



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

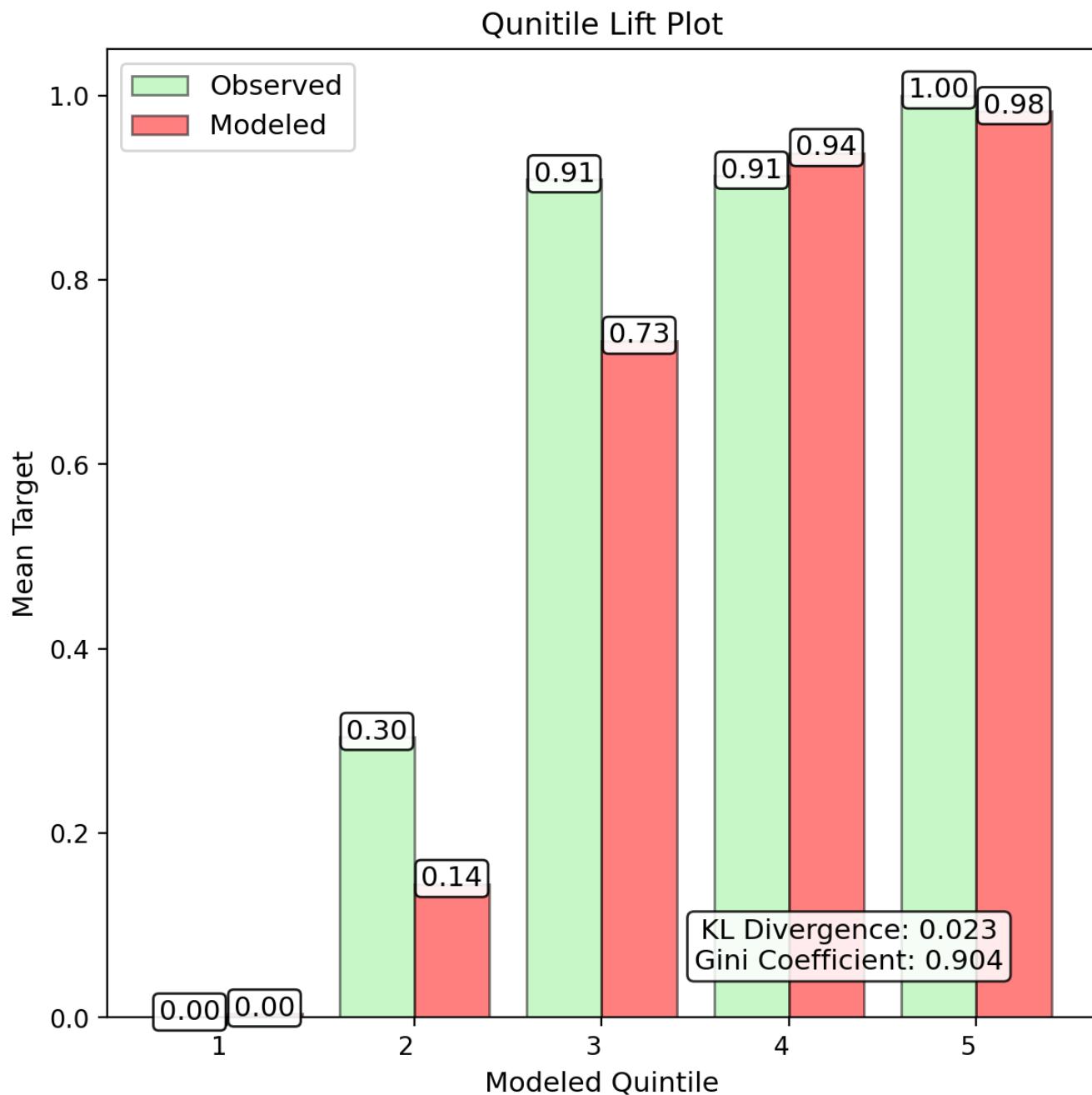


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

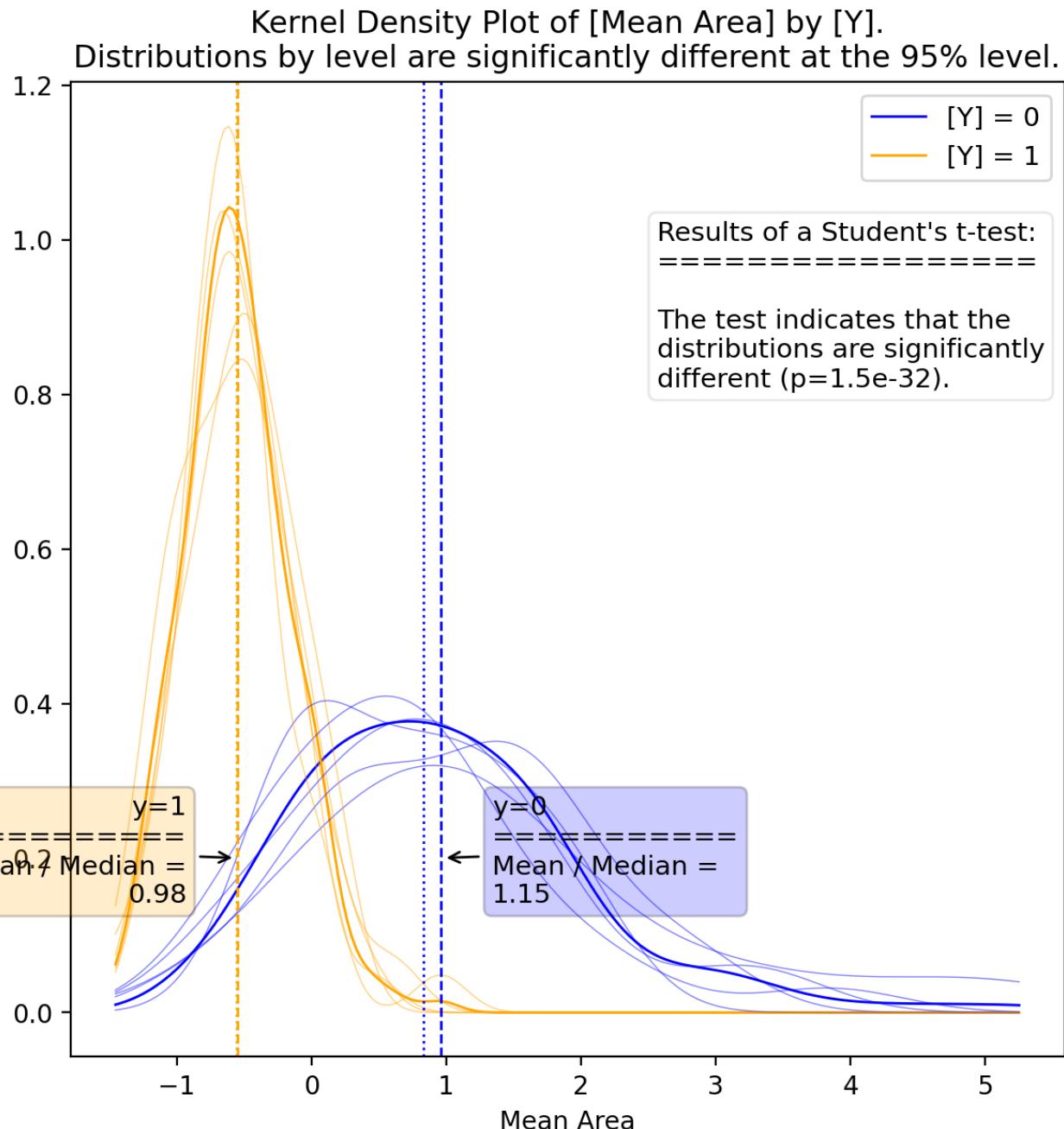
Univariate Report

Mean Area - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-4.15	-4.28	-4.25	-4.30	-4.20	-4.23	0.06
Fitted p-Value	1.5e-17	1.1e-16	3.2e-16	1.6e-16	7.3e-17	1.4e-20	1.2e-16
Fitted Std. Err.	0.486	0.517	0.521	0.521	0.504	0.455	0.015
Conf. Int. Lower	-5.10	-5.30	-5.27	-5.32	-5.19	-5.13	0.09
Conf. Int. Upper	-3.19	-3.27	-3.23	-3.28	-3.22	-3.34	0.04
Train Accuracy	86.6%	86.3%	88.4%	88.3%	89.2%	88.3%	1.3%
Val Accuracy	89.7%	87.3%	90.5%	88.1%	88.9%	88.6%	1.3%
Train AUC	85.6%	85.4%	86.9%	86.9%	87.9%	87.0%	1.0%
Val AUC	89.9%	87.5%	86.5%	87.2%	83.8%	86.7%	2.2%
Train F1	89.3%	89.3%	90.8%	90.9%	91.4%	90.8%	1.0%
Test F1	91.4%	88.1%	93.3%	90.5%	92.3%	91.2%	2.0%
Train Precision	89.3%	90.4%	88.4%	89.3%	89.4%	89.5%	0.7%
Val Precision	94.1%	90.2%	89.1%	90.5%	87.5%	88.2%	2.4%
Train Recall	89.3%	88.3%	93.3%	92.4%	93.6%	92.1%	2.4%
Val Recall	88.9%	86.0%	98.0%	90.5%	97.7%	94.4%	5.4%
Train MCC	71.2%	70.0%	75.3%	74.8%	77.0%	74.7%	2.9%
Val MCC	78.6%	74.7%	78.2%	74.5%	74.0%	75.5%	2.2%
Train Log-Loss	4.84	4.95	4.18	4.21	3.89	4.23	0.46
Val Log-Loss	3.73	4.56	3.41	4.30	4.00	4.11	0.46

Univariate Report

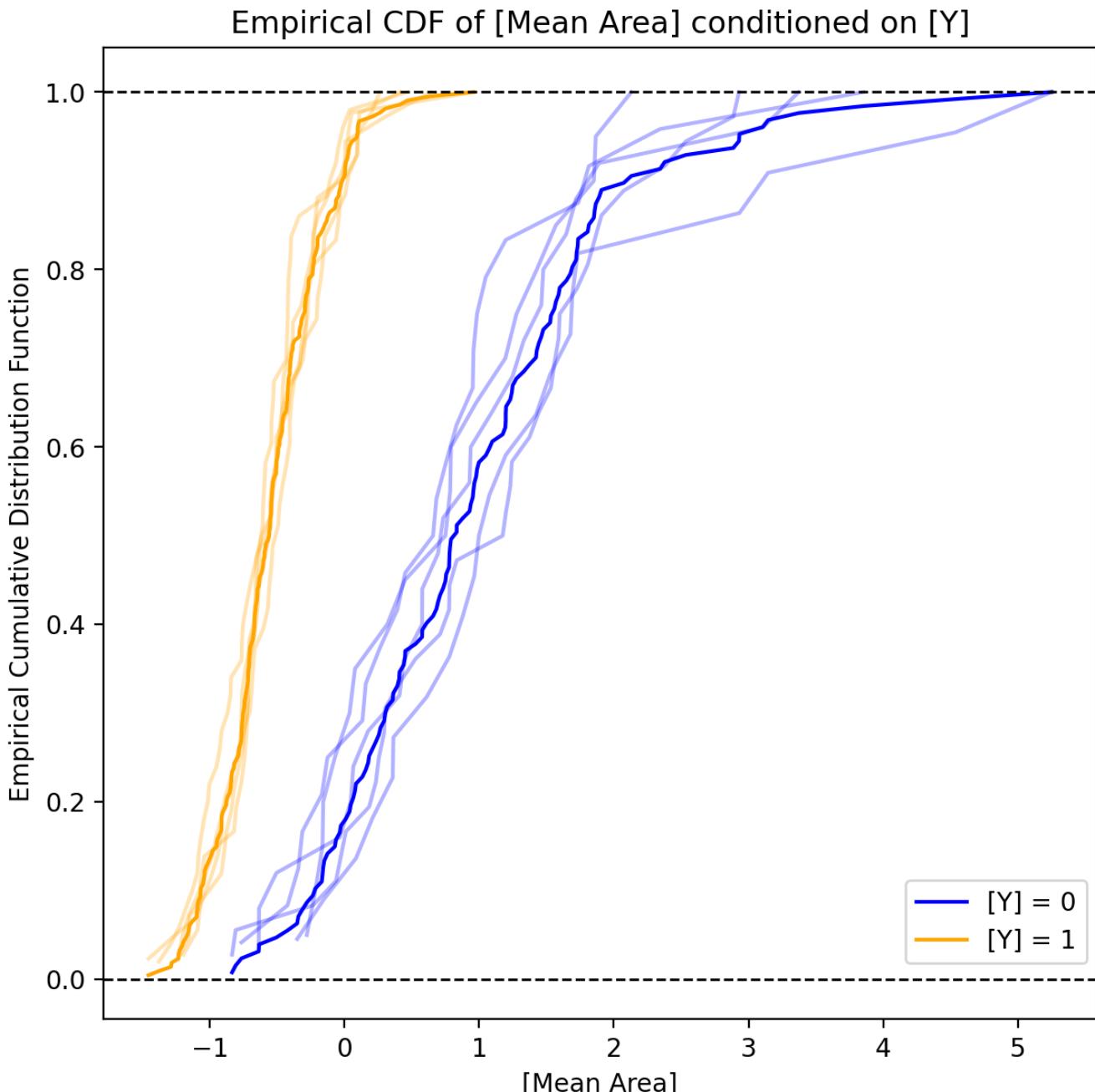
Mean Area - Kernel Density Plot



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

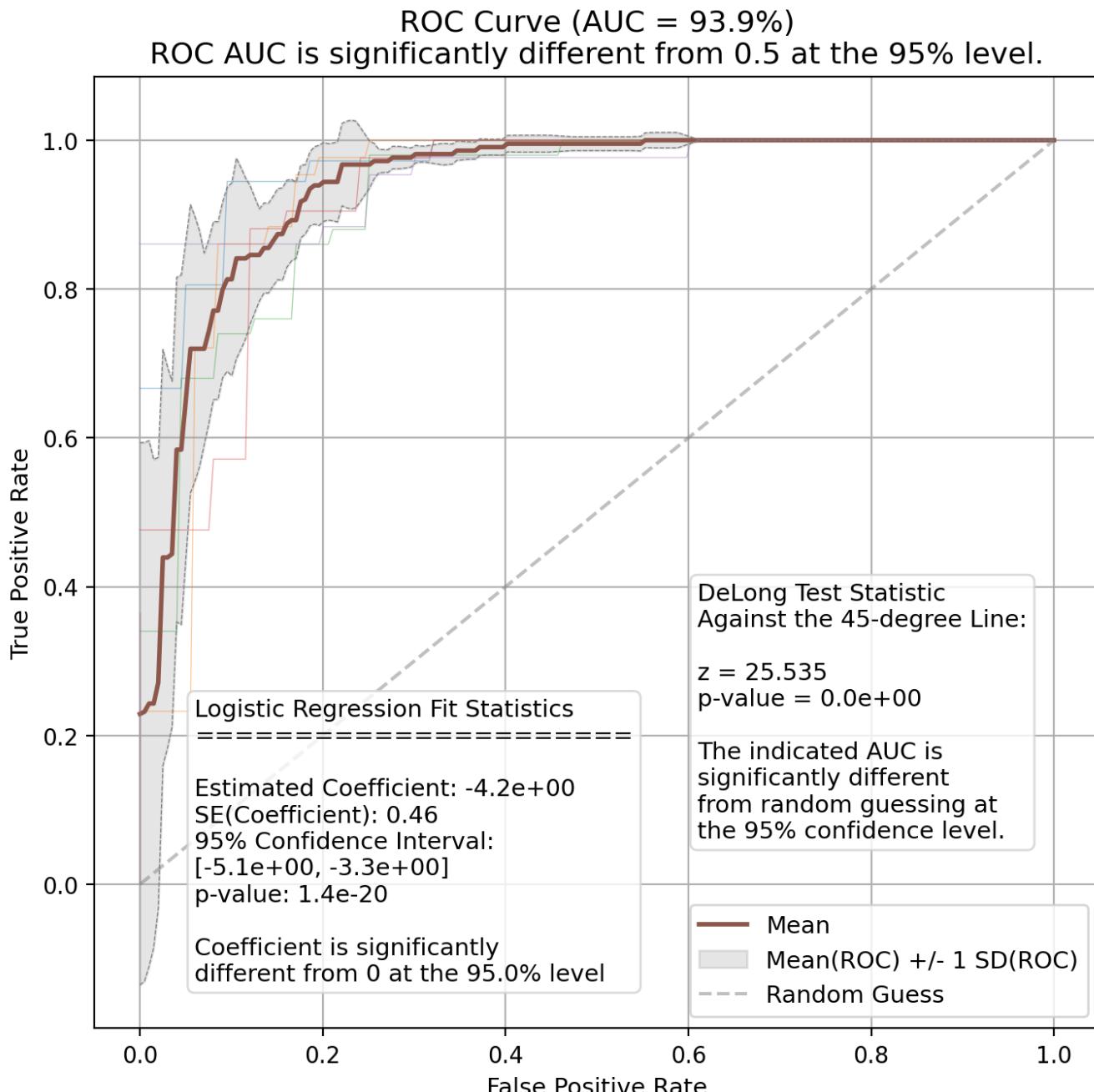
Mean Area - Empirical CDF Plot



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

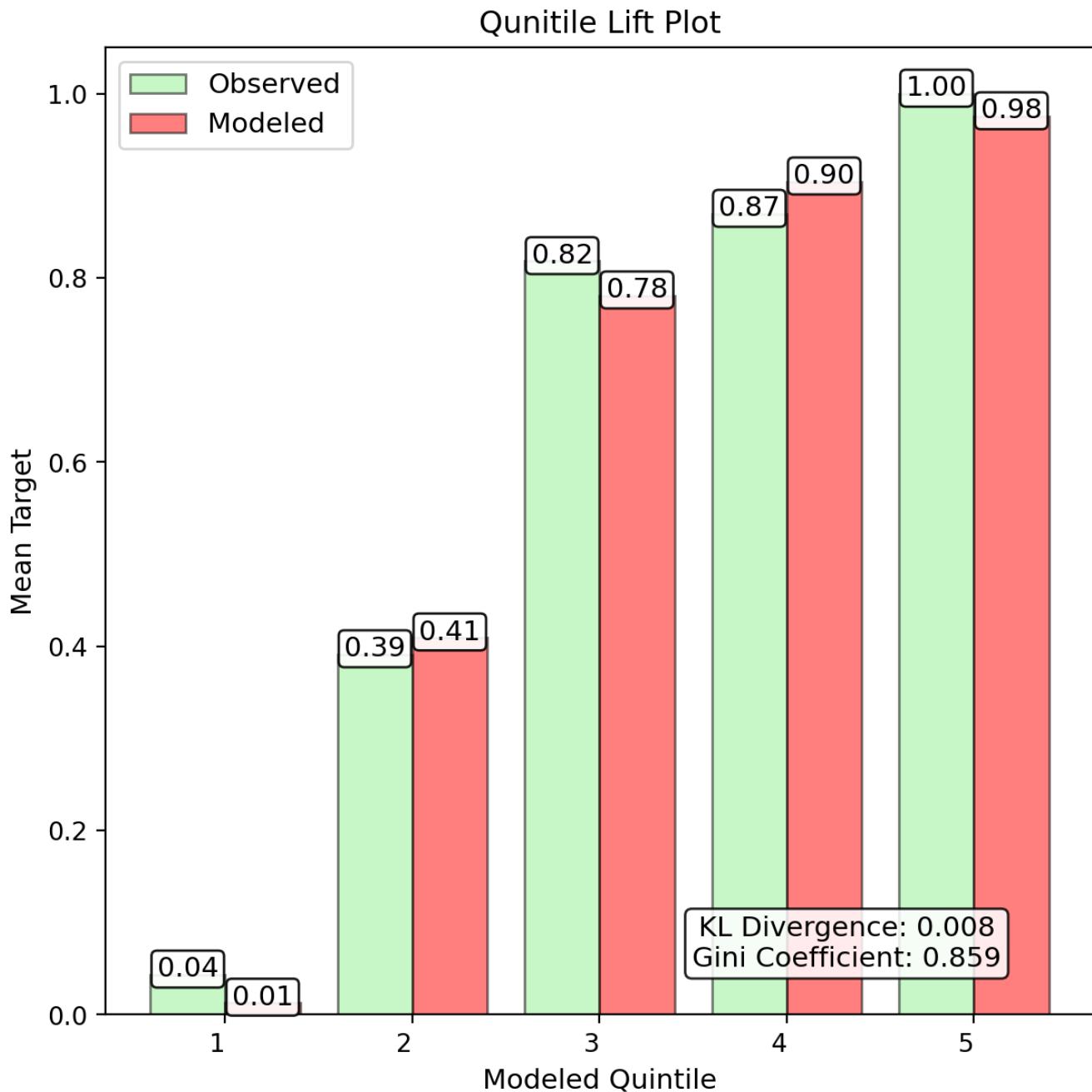


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

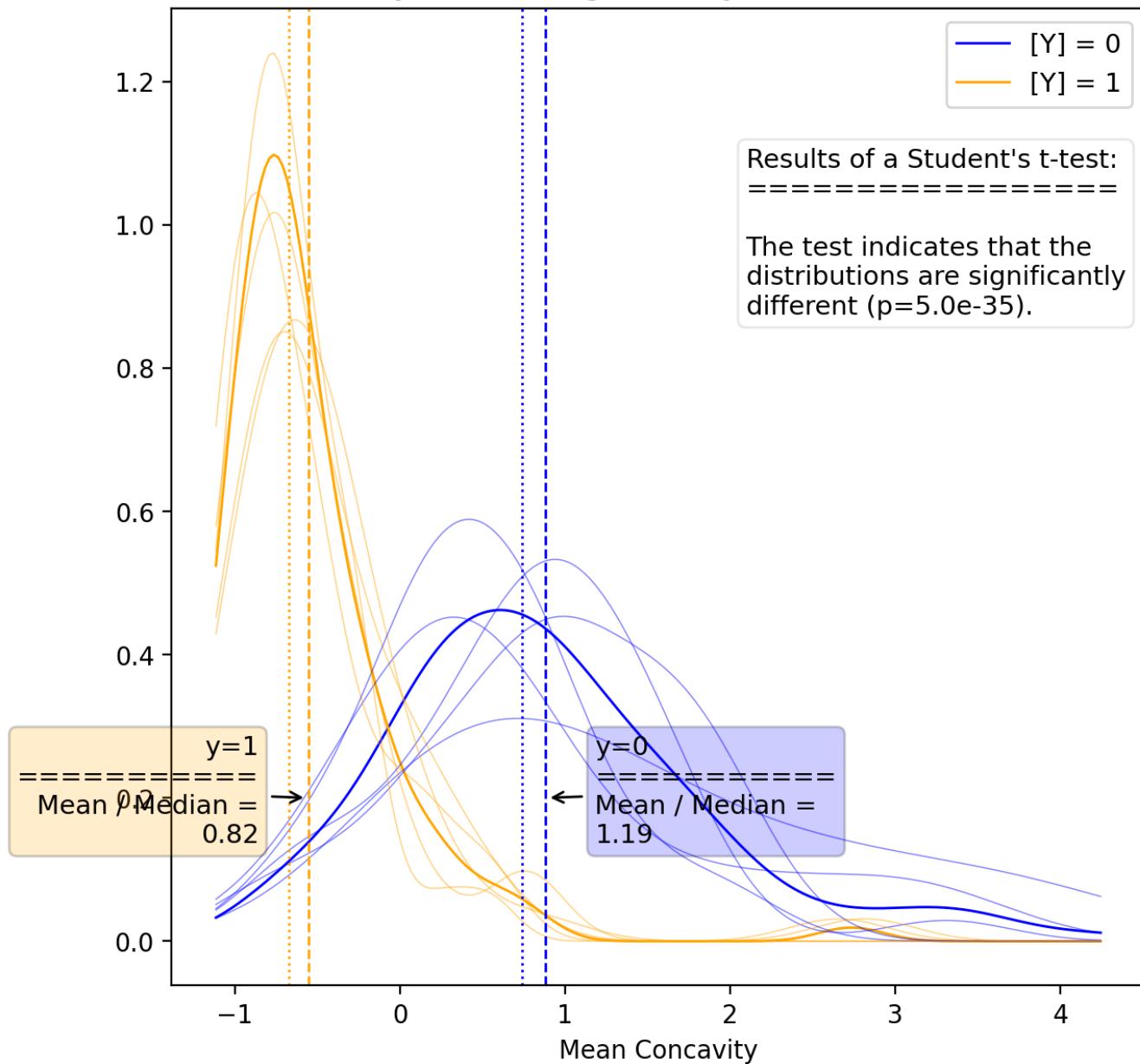
Mean Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-3.01	-3.45	-3.01	-3.39	-2.95	-3.15	0.24
Fitted p-Value	1.8e-20	1.3e-18	3.0e-19	1.4e-19	1.4e-19	4.9e-24	5.4e-19
Fitted Std. Err.	0.325	0.392	0.335	0.375	0.326	0.311	0.031
Conf. Int. Lower	-3.65	-4.22	-3.67	-4.13	-3.59	-3.76	0.30
Conf. Int. Upper	-2.38	-2.68	-2.35	-2.66	-2.31	-2.54	0.18
Train Accuracy	87.6%	89.3%	89.1%	89.1%	88.1%	88.6%	0.7%
Val Accuracy	93.1%	86.1%	86.5%	88.1%	90.5%	87.7%	2.9%
Train AUC	87.6%	89.2%	88.8%	88.9%	88.1%	88.3%	0.7%
Val AUC	93.6%	86.3%	85.7%	88.0%	87.7%	87.4%	3.1%
Train F1	89.9%	91.6%	91.1%	91.1%	90.1%	90.7%	0.7%
Test F1	94.3%	86.7%	89.8%	90.2%	93.2%	90.0%	3.0%
Train Precision	92.3%	93.9%	91.9%	92.8%	92.1%	92.3%	0.8%
Val Precision	97.1%	90.0%	91.7%	92.5%	91.1%	91.3%	2.7%
Train Recall	87.6%	89.5%	90.2%	89.5%	88.3%	89.3%	1.0%
Val Recall	91.7%	83.7%	88.0%	88.1%	95.3%	88.7%	4.4%
Train MCC	74.1%	77.1%	77.2%	76.9%	75.3%	75.9%	1.4%
Val MCC	85.8%	72.3%	69.9%	75.0%	77.6%	74.2%	6.1%
Train Log-Loss	4.46	3.85	3.91	3.95	4.28	4.12	0.26
Val Log-Loss	2.49	5.02	4.87	4.30	3.43	4.43	1.06

Univariate Report

Mean Concavity - Kernel Density Plot

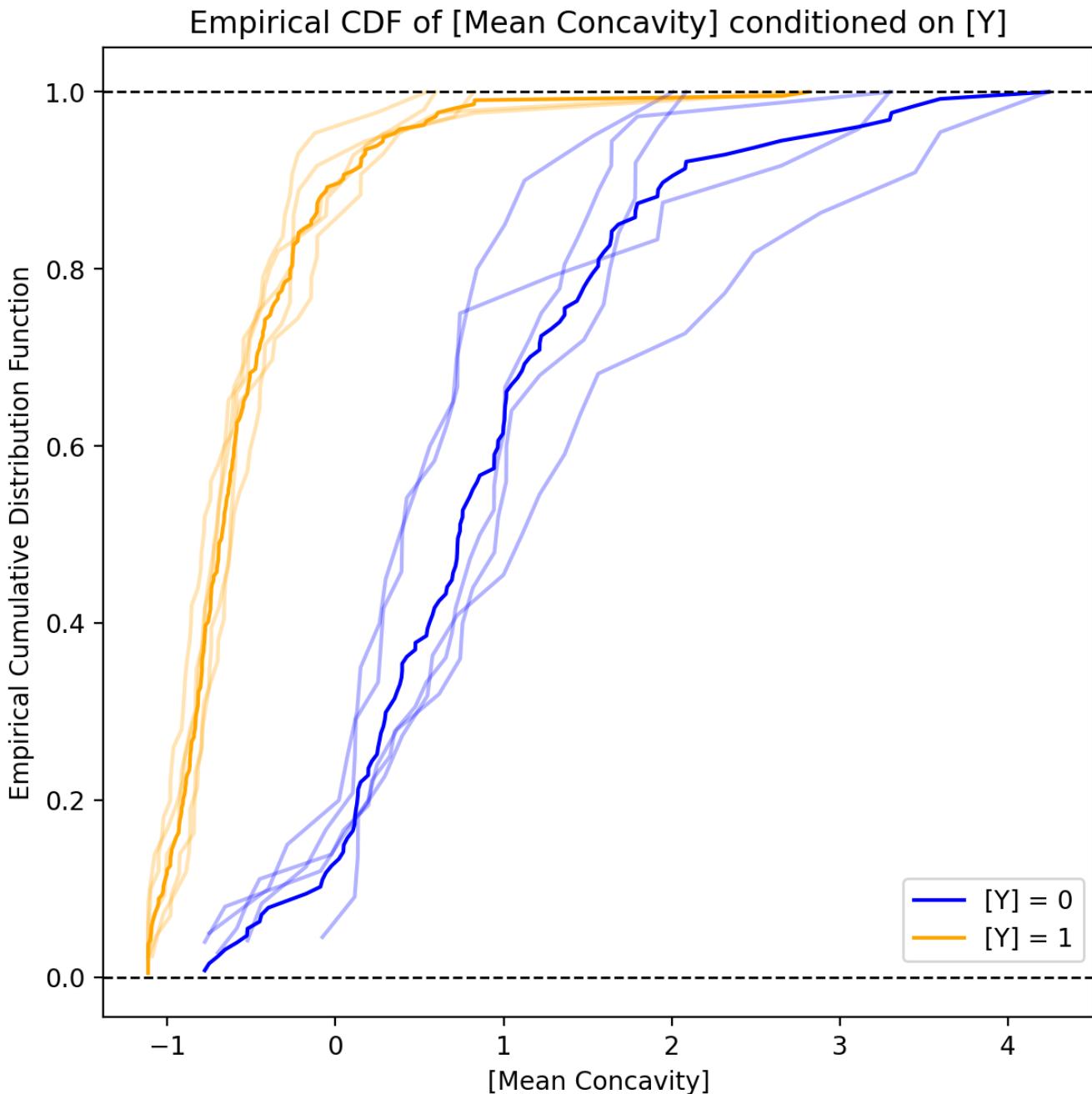
Kernel Density Plot of [Mean Concavity] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

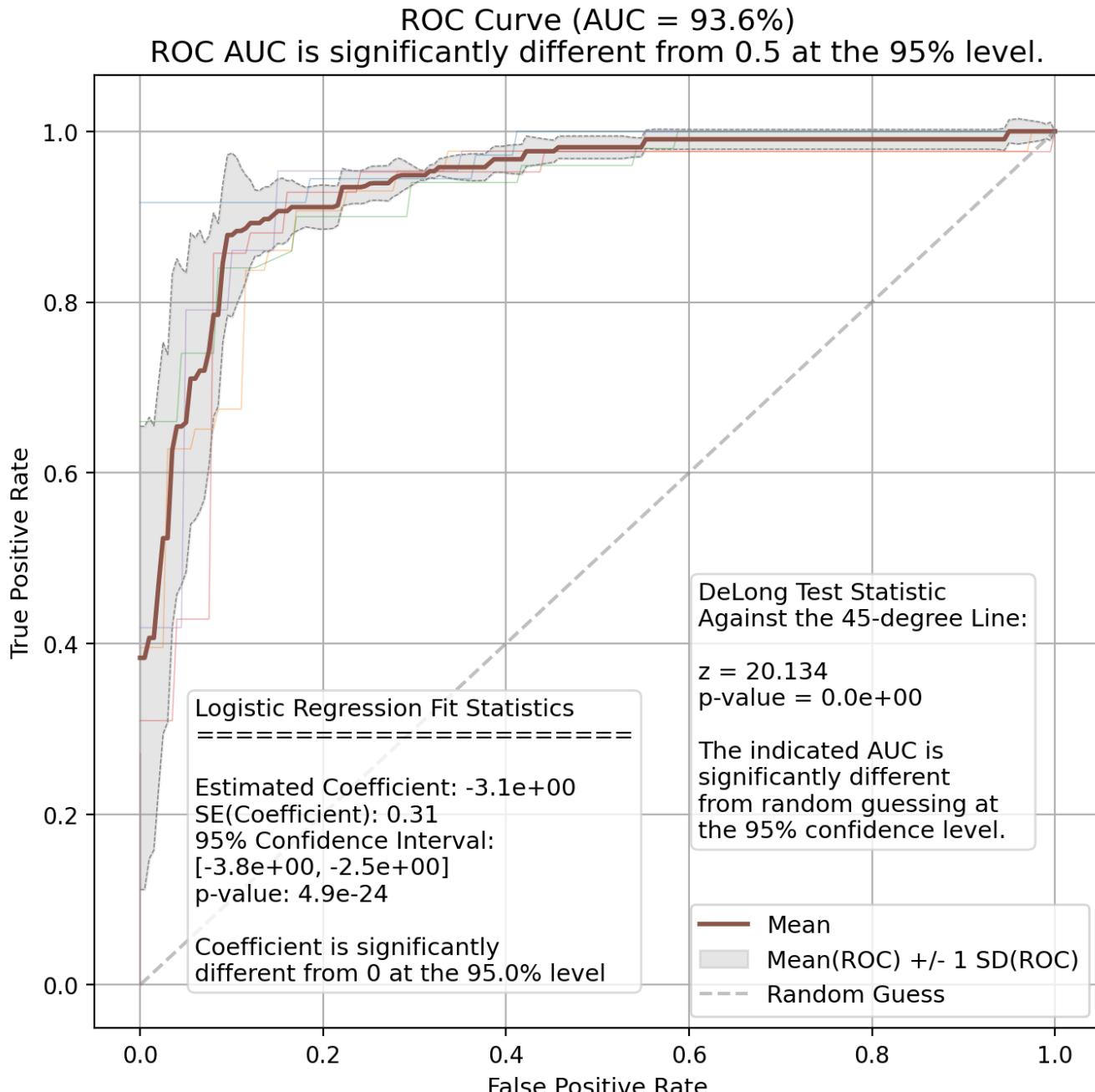
Mean Concavity - Empirical CDF Plot



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

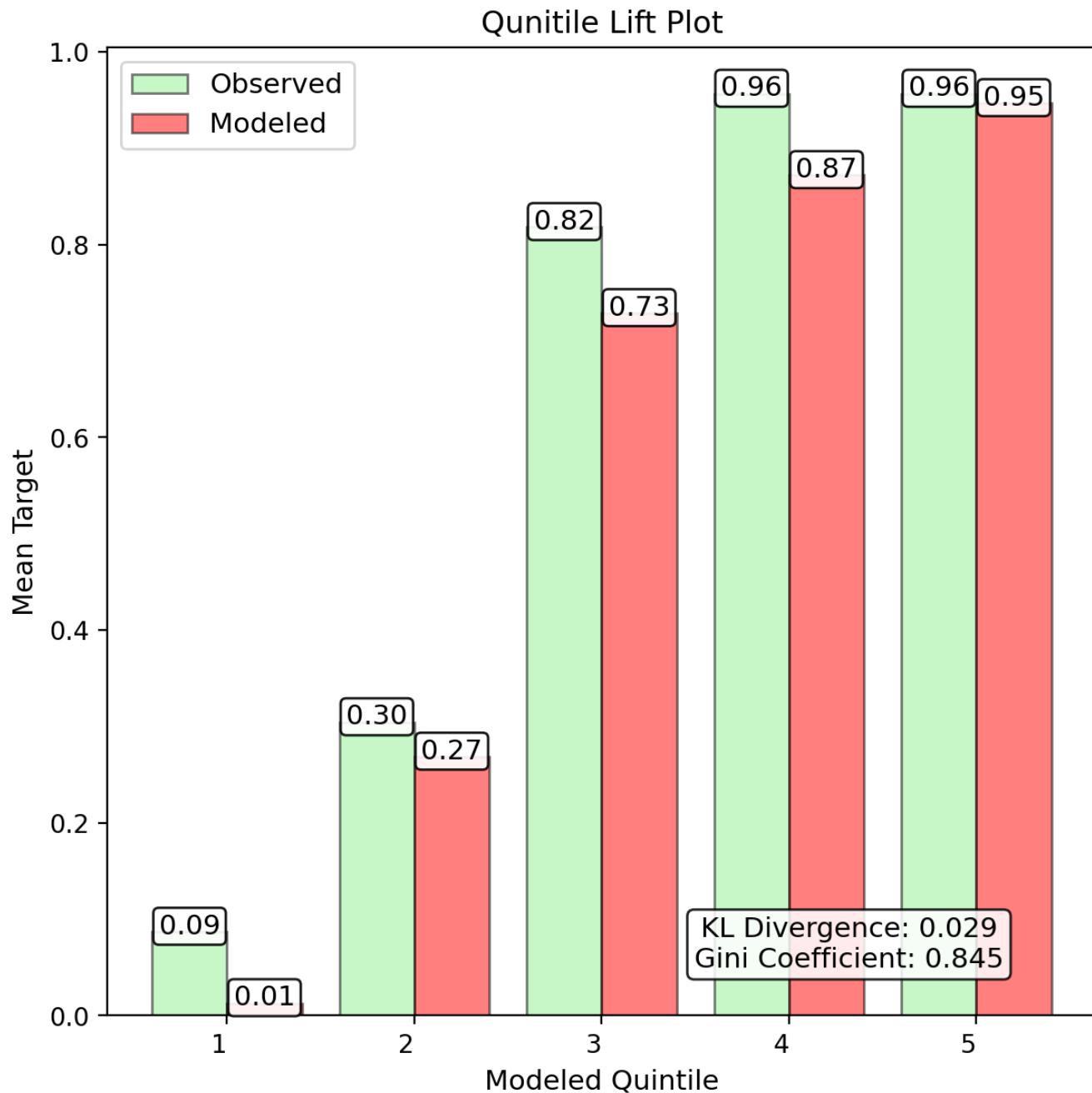


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

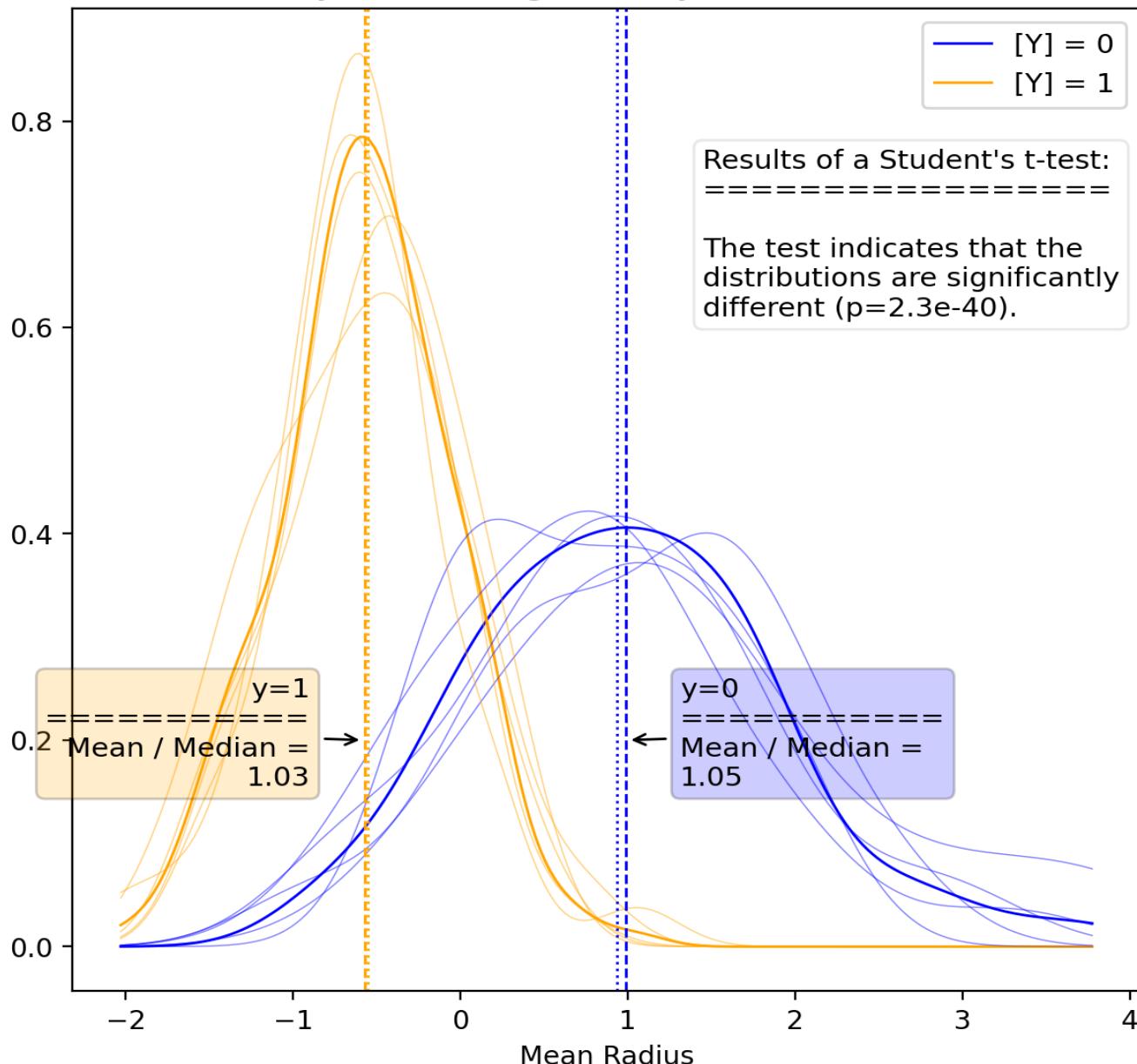
Mean Radius - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-3.65	-3.80	-3.66	-3.77	-3.58	-3.69	0.09
Fitted p-Value	5.9e-17	5.5e-16	4.6e-16	6.2e-16	1.7e-16	5.6e-20	2.4e-16
Fitted Std. Err.	0.437	0.469	0.451	0.466	0.434	0.403	0.016
Conf. Int. Lower	-4.51	-4.72	-4.55	-4.68	-4.43	-4.48	0.12
Conf. Int. Upper	-2.80	-2.88	-2.78	-2.85	-2.73	-2.90	0.06
Train Accuracy	85.5%	85.9%	86.1%	86.1%	86.7%	86.5%	0.4%
Val Accuracy	84.5%	86.1%	86.5%	88.1%	84.1%	86.8%	1.6%
Train AUC	85.6%	86.1%	86.0%	86.2%	86.4%	86.5%	0.3%
Val AUC	85.7%	86.5%	83.5%	88.0%	81.7%	86.2%	2.5%
Train F1	88.1%	88.8%	88.5%	88.6%	89.0%	88.9%	0.3%
Test F1	86.6%	86.4%	90.2%	90.2%	88.4%	89.4%	1.9%
Train Precision	91.0%	92.4%	90.4%	91.4%	90.4%	91.6%	0.8%
Val Precision	93.5%	92.1%	88.5%	92.5%	88.4%	90.0%	2.4%
Train Recall	85.4%	85.4%	86.6%	86.0%	87.7%	86.4%	1.0%
Val Recall	80.6%	81.4%	92.0%	88.1%	88.4%	88.7%	4.9%
Train MCC	69.8%	70.3%	71.2%	71.1%	72.2%	71.9%	0.9%
Val MCC	69.5%	72.8%	68.6%	75.0%	63.4%	72.1%	4.5%
Train Log-Loss	5.22	5.09	4.99	5.00	4.80	4.86	0.16
Val Log-Loss	5.59	5.02	4.87	4.30	5.72	4.74	0.57

Univariate Report

Mean Radius - Kernel Density Plot

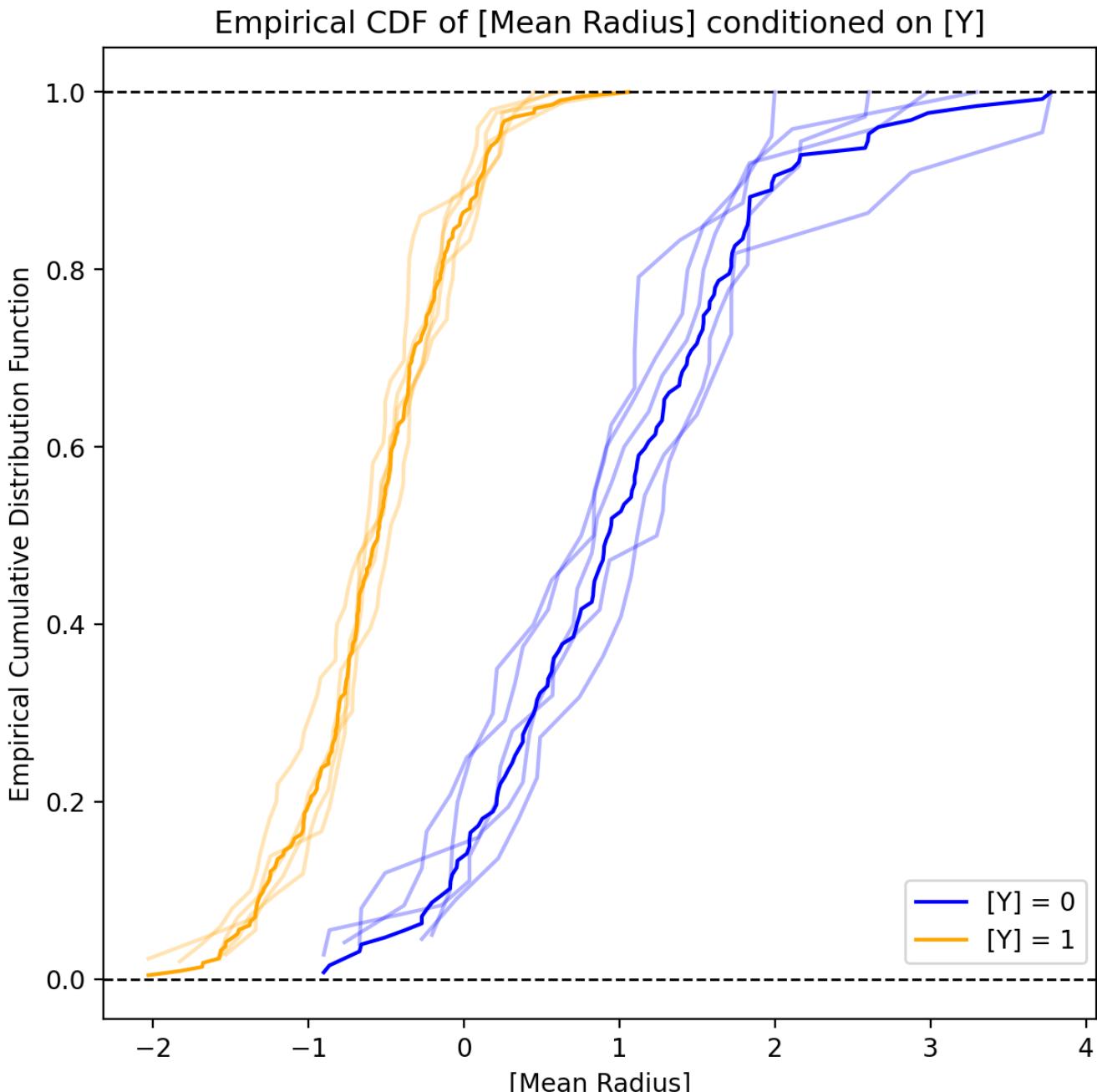
Kernel Density Plot of [Mean Radius] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

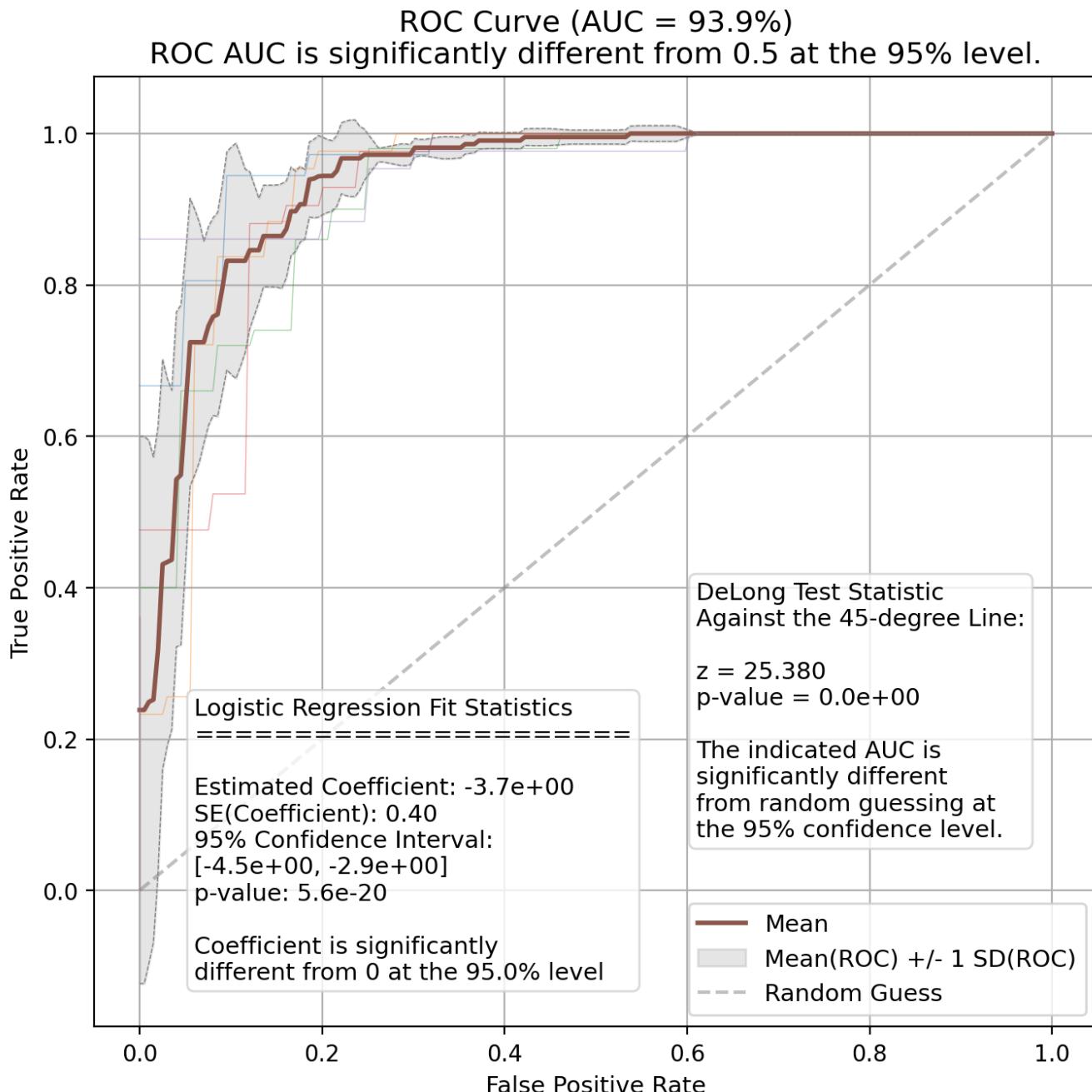
Mean Radius - Empirical CDF Plot



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

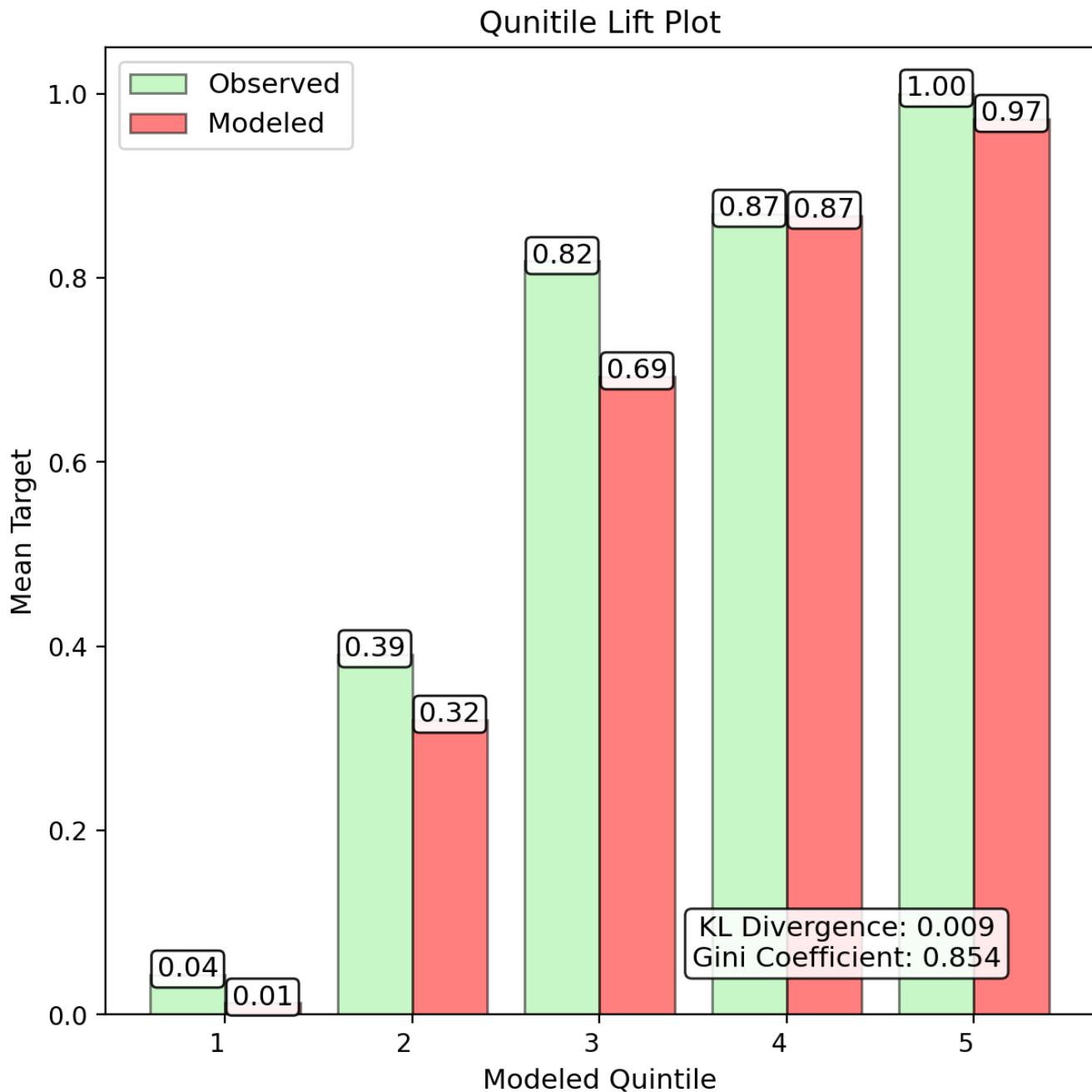


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

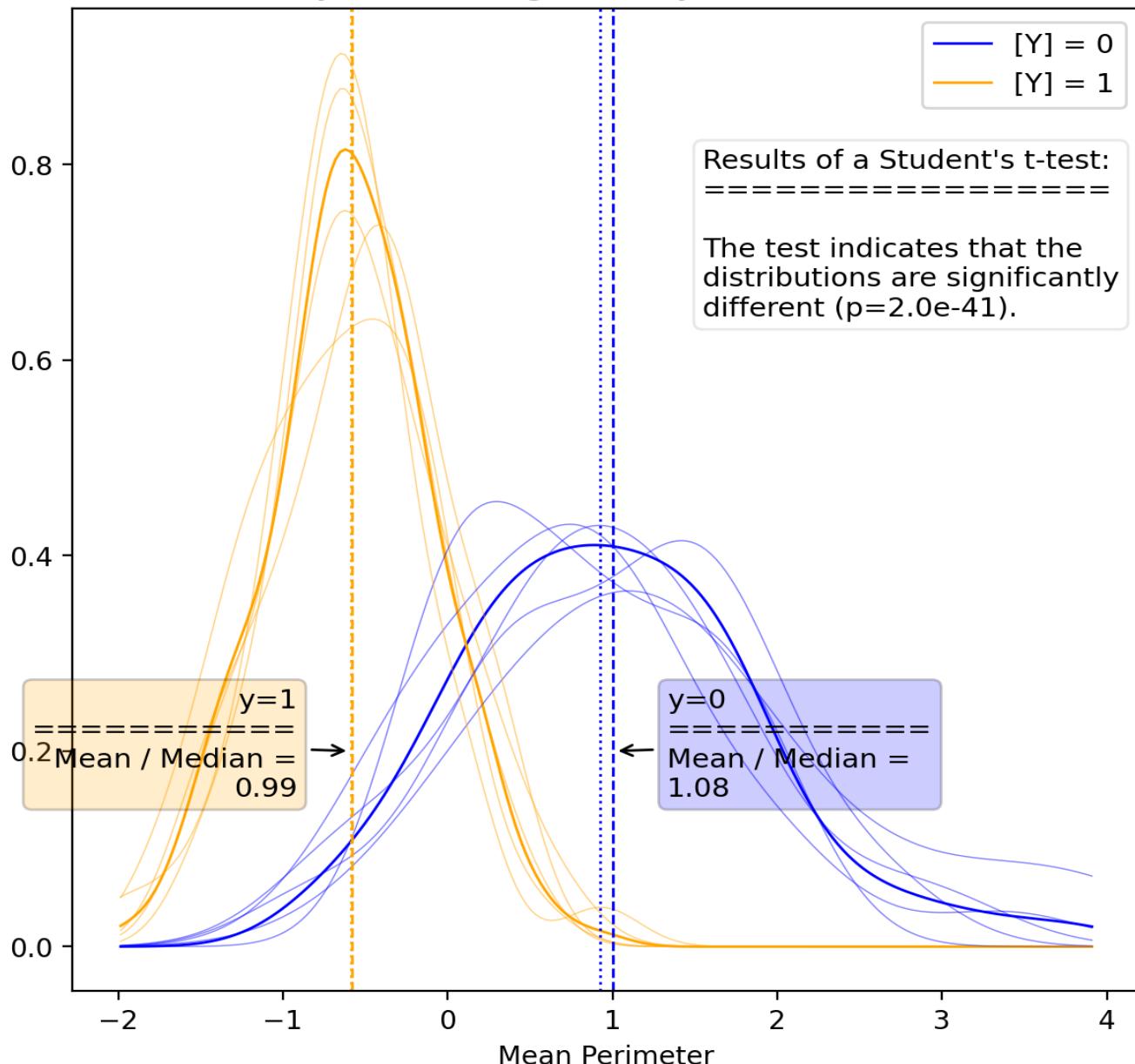
Mean Perimeter - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-4.00	-4.17	-3.94	-4.07	-3.83	-4.00	0.13
Fitted p-Value	7.5e-17	9.4e-16	4.0e-16	8.2e-16	1.4e-16	6.6e-20	3.9e-16
Fitted Std. Err.	0.480	0.518	0.484	0.506	0.464	0.438	0.022
Conf. Int. Lower	-4.95	-5.18	-4.89	-5.06	-4.74	-4.86	0.17
Conf. Int. Upper	-3.06	-3.15	-2.99	-3.08	-2.92	-3.14	0.09
Train Accuracy	88.0%	88.5%	87.6%	87.2%	87.1%	87.4%	0.6%
Val Accuracy	86.2%	87.3%	86.5%	88.1%	88.9%	86.8%	1.1%
Train AUC	87.9%	88.9%	87.8%	87.2%	87.0%	87.4%	0.7%
Val AUC	88.0%	87.7%	83.5%	88.0%	86.5%	86.2%	1.9%
Train F1	90.2%	90.9%	89.7%	89.6%	89.2%	89.7%	0.7%
Test F1	87.9%	87.8%	90.2%	90.2%	92.0%	89.4%	1.8%
Train Precision	92.4%	94.3%	92.3%	92.0%	91.4%	92.1%	1.1%
Val Precision	96.7%	92.3%	88.5%	92.5%	90.9%	90.0%	3.0%
Train Recall	88.2%	87.7%	87.2%	87.2%	87.1%	87.4%	0.5%
Val Recall	80.6%	83.7%	92.0%	88.1%	93.0%	88.7%	5.3%
Train MCC	74.8%	75.9%	74.5%	73.3%	73.2%	73.7%	1.1%
Val MCC	73.8%	75.1%	68.6%	75.0%	74.1%	72.1%	2.7%
Train Log-Loss	4.33	4.13	4.45	4.60	4.67	4.55	0.22
Val Log-Loss	4.97	4.56	4.87	4.30	4.00	4.74	0.40

Univariate Report

Mean Perimeter - Kernel Density Plot

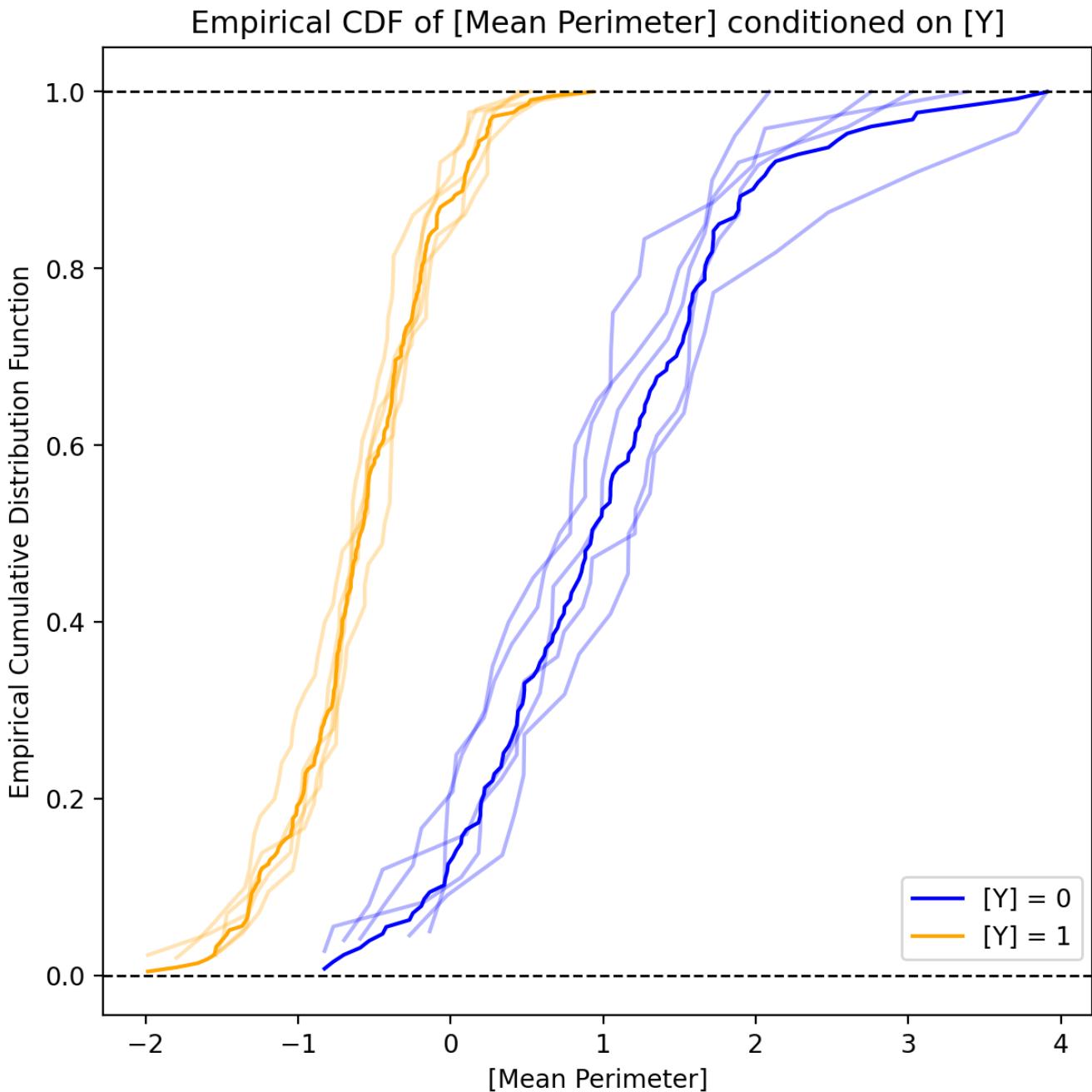
Kernel Density Plot of [Mean Perimeter] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

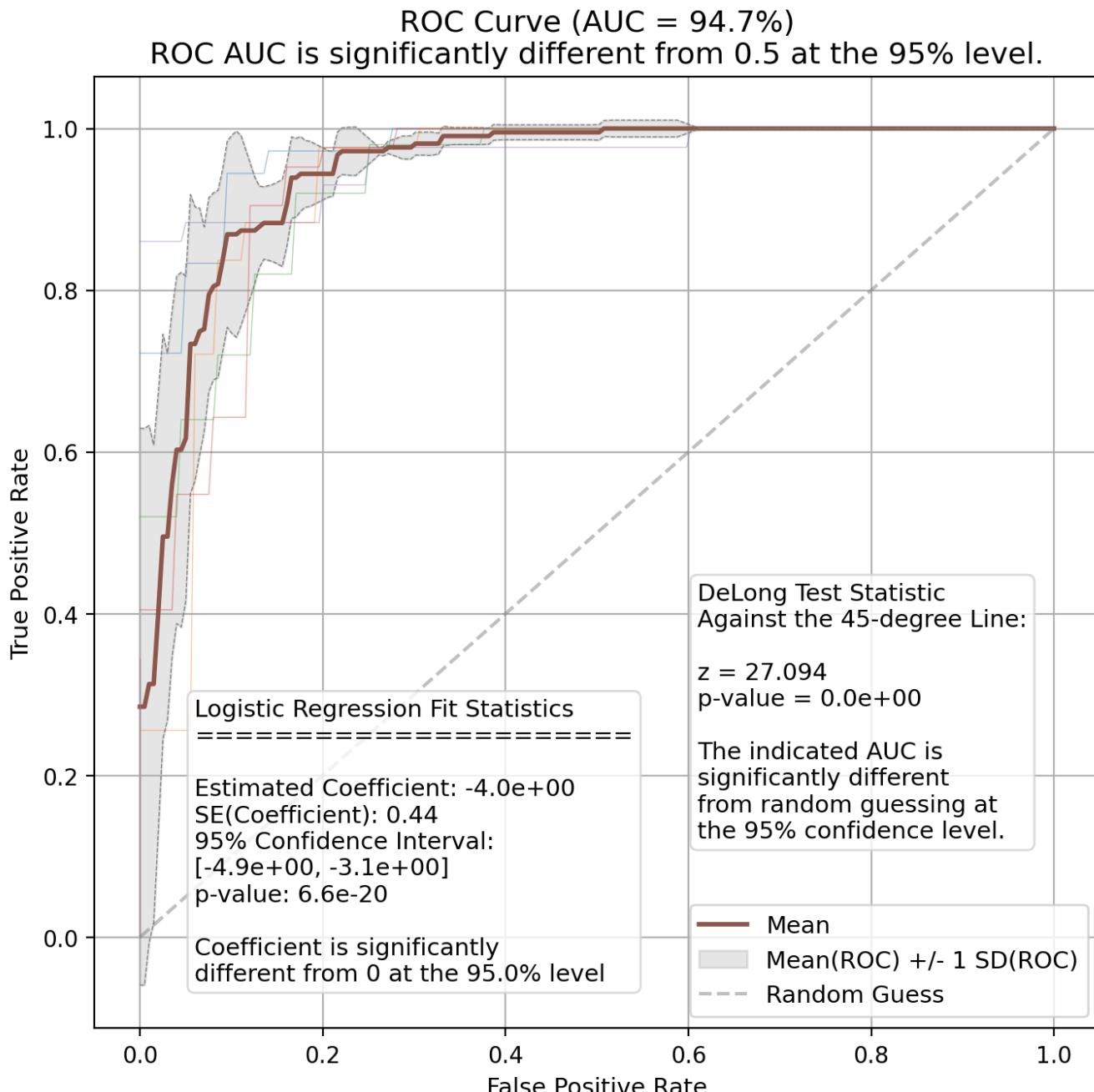
Mean Perimeter - Empirical CDF Plot



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

Univariate Report

Mean Perimeter - ROC Curve

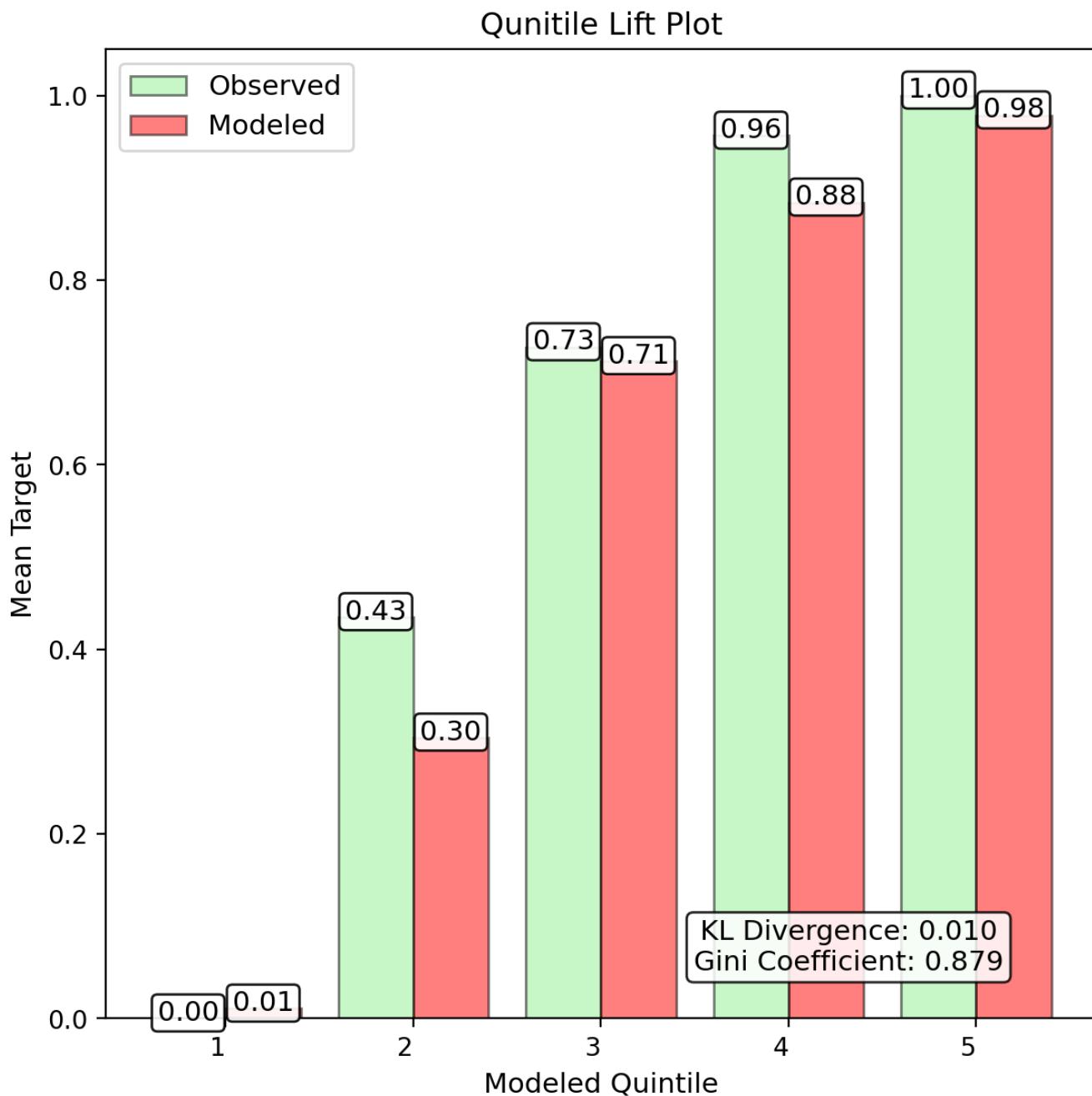


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Significance of the ROC curve is determined based on 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 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

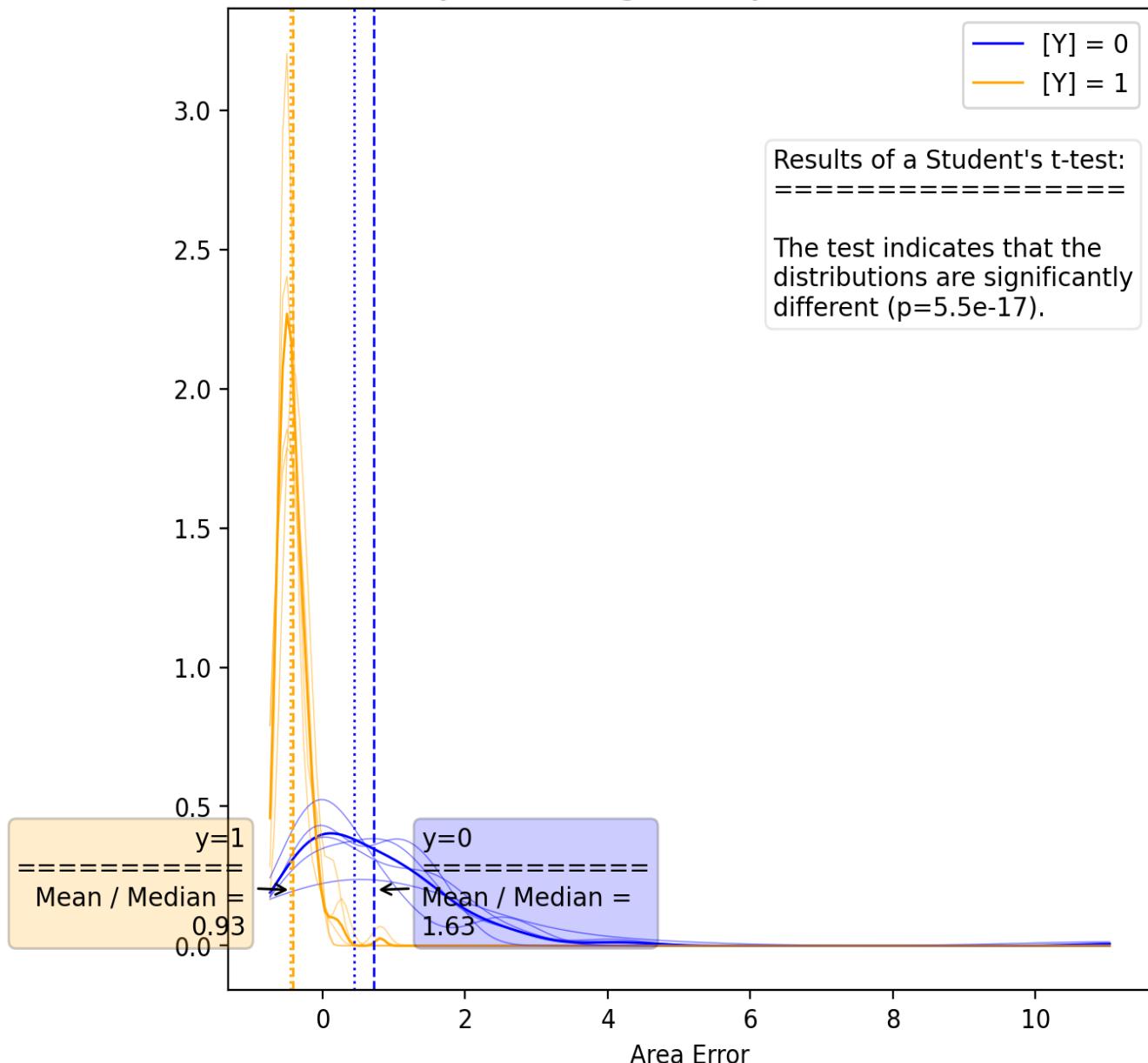
Area Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-4.66	-4.95	-4.92	-4.75	-4.75	-4.80	0.12
Fitted p-Value	1.7e-20	1.2e-18	6.1e-19	1.9e-19	8.7e-20	6.3e-24	5.1e-19
Fitted Std. Err.	0.502	0.561	0.553	0.527	0.522	0.476	0.024
Conf. Int. Lower	-5.64	-6.05	-6.00	-5.78	-5.78	-5.73	0.17
Conf. Int. Upper	-3.68	-3.85	-3.83	-3.72	-3.73	-3.87	0.08
Train Accuracy	87.3%	86.6%	86.9%	85.8%	86.0%	86.8%	0.6%
Val Accuracy	86.2%	88.6%	82.4%	89.6%	85.7%	86.8%	2.8%
Train AUC	84.0%	82.6%	83.9%	82.3%	82.7%	83.4%	0.8%
Val AUC	84.5%	87.5%	75.1%	86.0%	80.2%	84.4%	5.0%
Train F1	90.5%	90.4%	90.1%	89.4%	89.5%	90.2%	0.5%
Test F1	89.2%	90.5%	88.1%	92.3%	90.1%	89.9%	1.6%
Train Precision	85.1%	85.4%	84.1%	83.8%	83.0%	84.5%	1.0%
Val Precision	86.8%	82.7%	81.4%	85.7%	85.4%	85.9%	2.3%
Train Recall	96.6%	95.9%	97.0%	95.9%	97.1%	96.7%	0.6%
Val Recall	91.7%	100.0%	96.0%	100.0%	95.3%	94.4%	3.5%
Train MCC	72.7%	70.1%	72.6%	69.4%	70.7%	71.8%	1.5%
Val MCC	70.4%	78.8%	58.4%	78.6%	66.0%	71.7%	8.7%
Train Log-Loss	4.59	4.81	4.72	5.13	5.06	4.76	0.23
Val Log-Loss	4.97	4.11	6.33	3.77	5.15	4.74	1.00

Univariate Report

Area Error - Kernel Density Plot

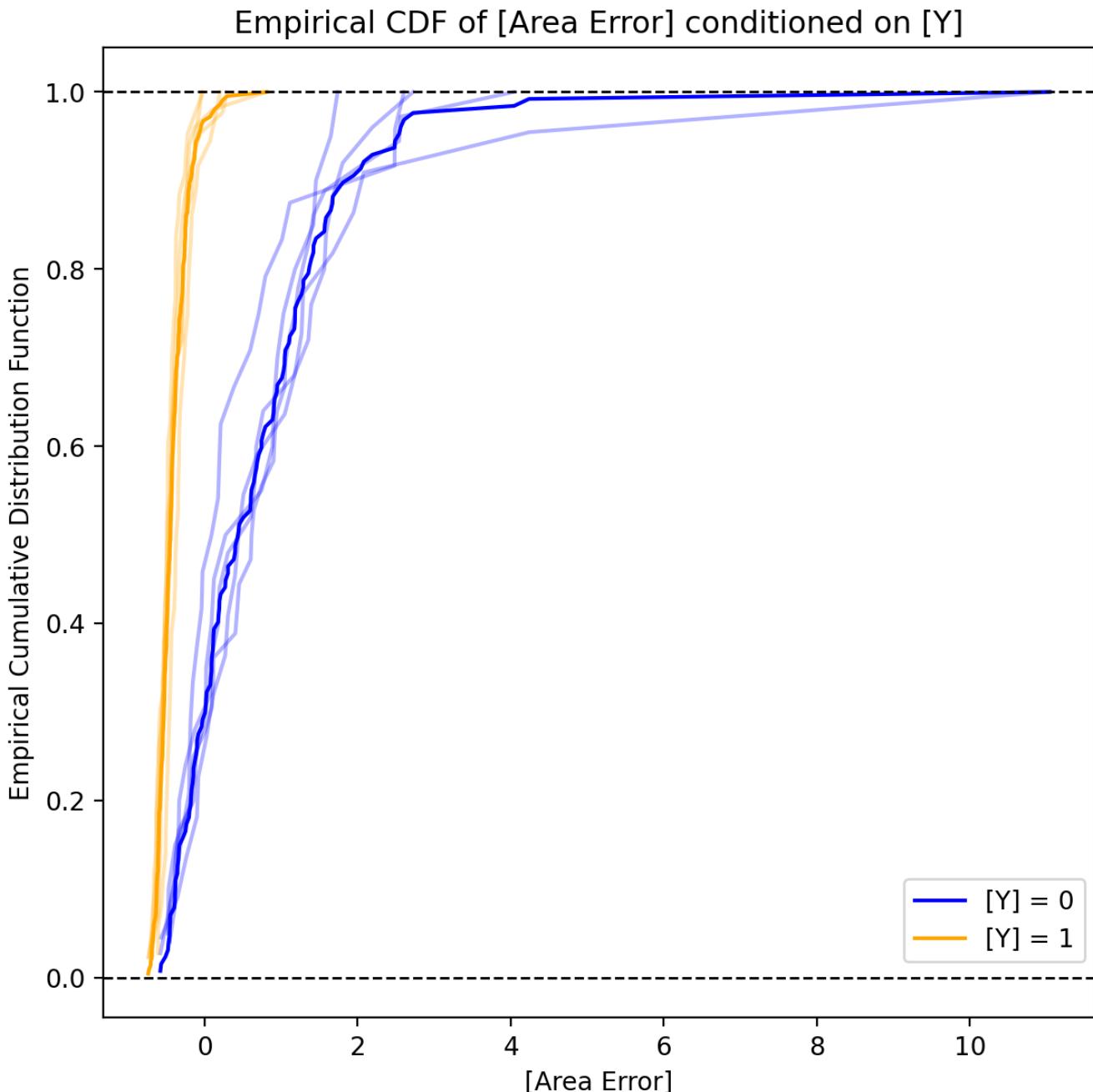
Kernel Density Plot of [Area Error] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

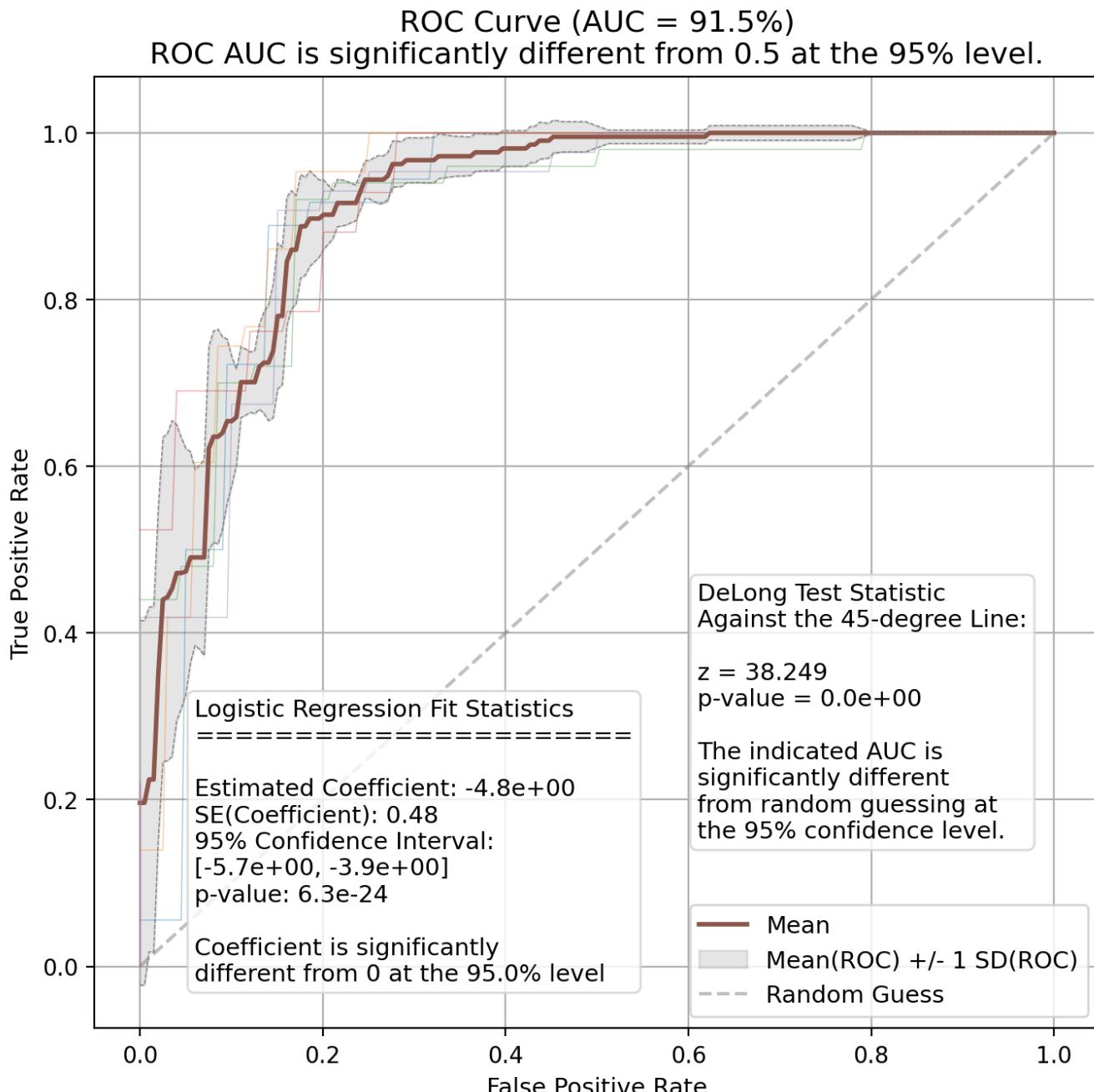
Area Error - Empirical CDF Plot



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

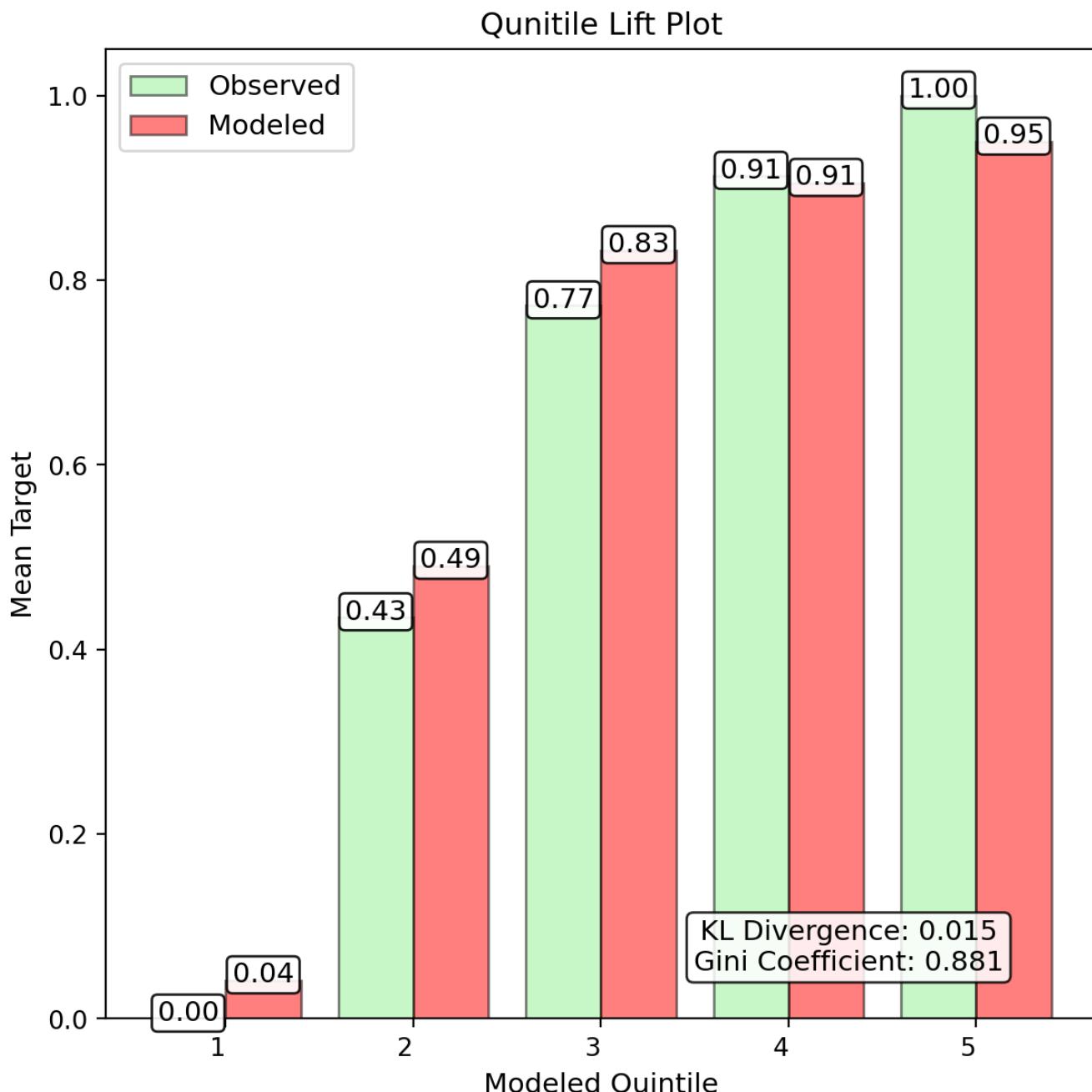


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

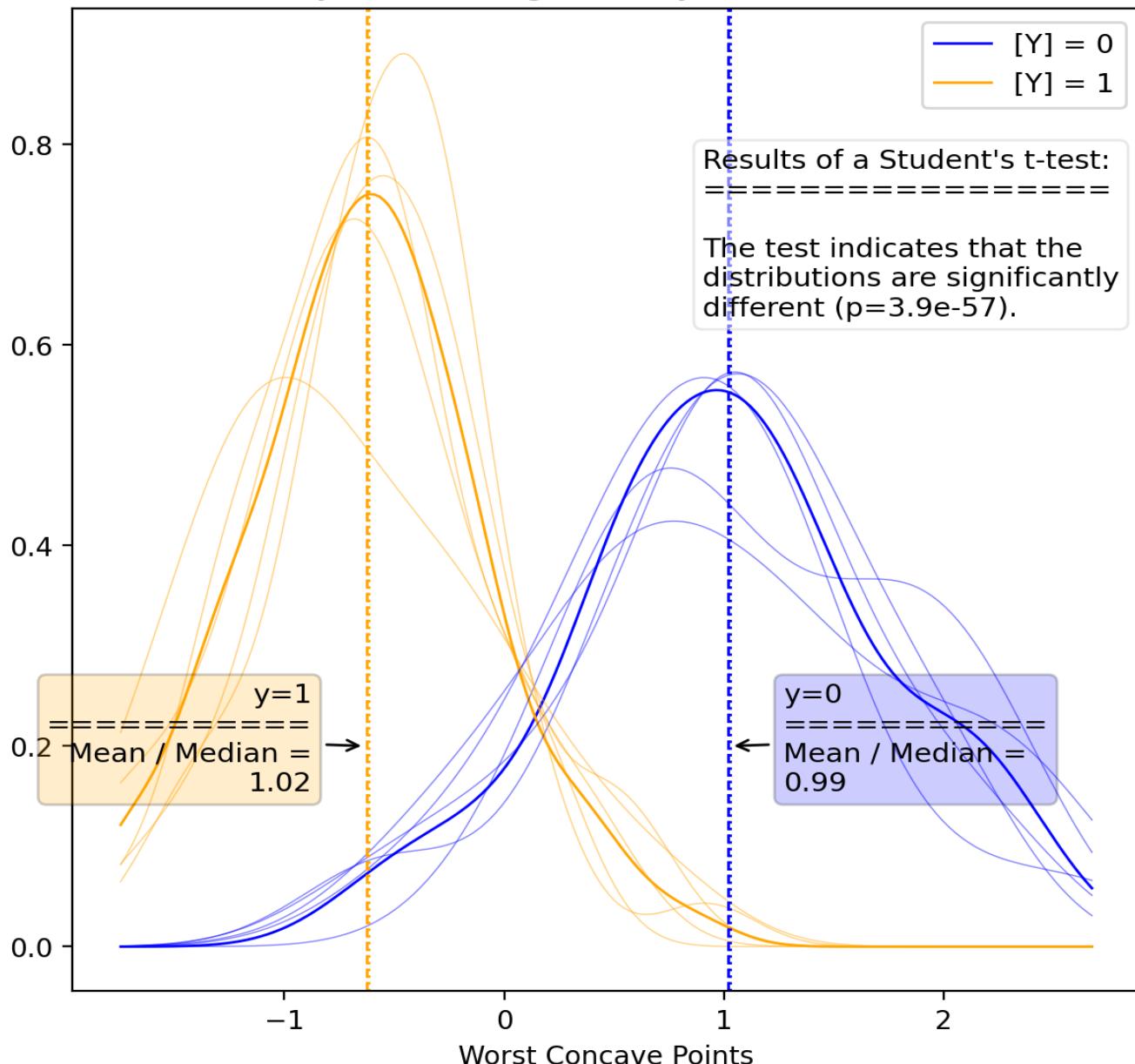
Worst Concave Points - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-3.75	-4.11	-4.22	-3.87	-3.70	-3.92	0.23
Fitted p-Value	6.4e-18	6.9e-16	2.1e-16	5.8e-17	1.9e-17	8.5e-21	2.9e-16
Fitted Std. Err.	0.435	0.510	0.513	0.463	0.435	0.419	0.039
Conf. Int. Lower	-4.61	-5.11	-5.23	-4.78	-4.55	-4.74	0.30
Conf. Int. Upper	-2.90	-3.12	-3.21	-2.97	-2.85	-3.09	0.15
Train Accuracy	89.0%	89.7%	91.0%	89.4%	89.9%	89.4%	0.7%
Val Accuracy	93.1%	87.3%	86.5%	91.0%	90.5%	84.2%	2.7%
Train AUC	89.3%	90.8%	91.1%	89.8%	90.1%	89.7%	0.7%
Val AUC	93.6%	87.5%	86.8%	91.2%	90.3%	86.4%	2.8%
Train F1	91.0%	91.7%	92.5%	91.3%	91.6%	91.3%	0.6%
Test F1	94.3%	88.1%	89.6%	92.7%	92.9%	85.9%	2.6%
Train Precision	94.0%	96.8%	94.3%	94.4%	93.9%	94.1%	1.2%
Val Precision	97.1%	90.2%	93.5%	95.0%	95.1%	96.5%	2.5%
Train Recall	88.2%	87.1%	90.9%	88.4%	89.5%	88.8%	1.4%
Val Recall	91.7%	86.0%	86.0%	90.5%	90.7%	77.5%	2.7%
Train MCC	77.3%	79.0%	81.3%	78.1%	79.2%	78.1%	1.5%
Val MCC	85.8%	74.7%	70.9%	81.3%	78.8%	70.6%	5.8%
Train Log-Loss	3.95	3.71	3.24	3.81	3.63	3.81	0.27
Val Log-Loss	2.49	4.56	4.87	3.23	3.43	5.69	0.99

Univariate Report

Worst Concave Points - Kernel Density Plot

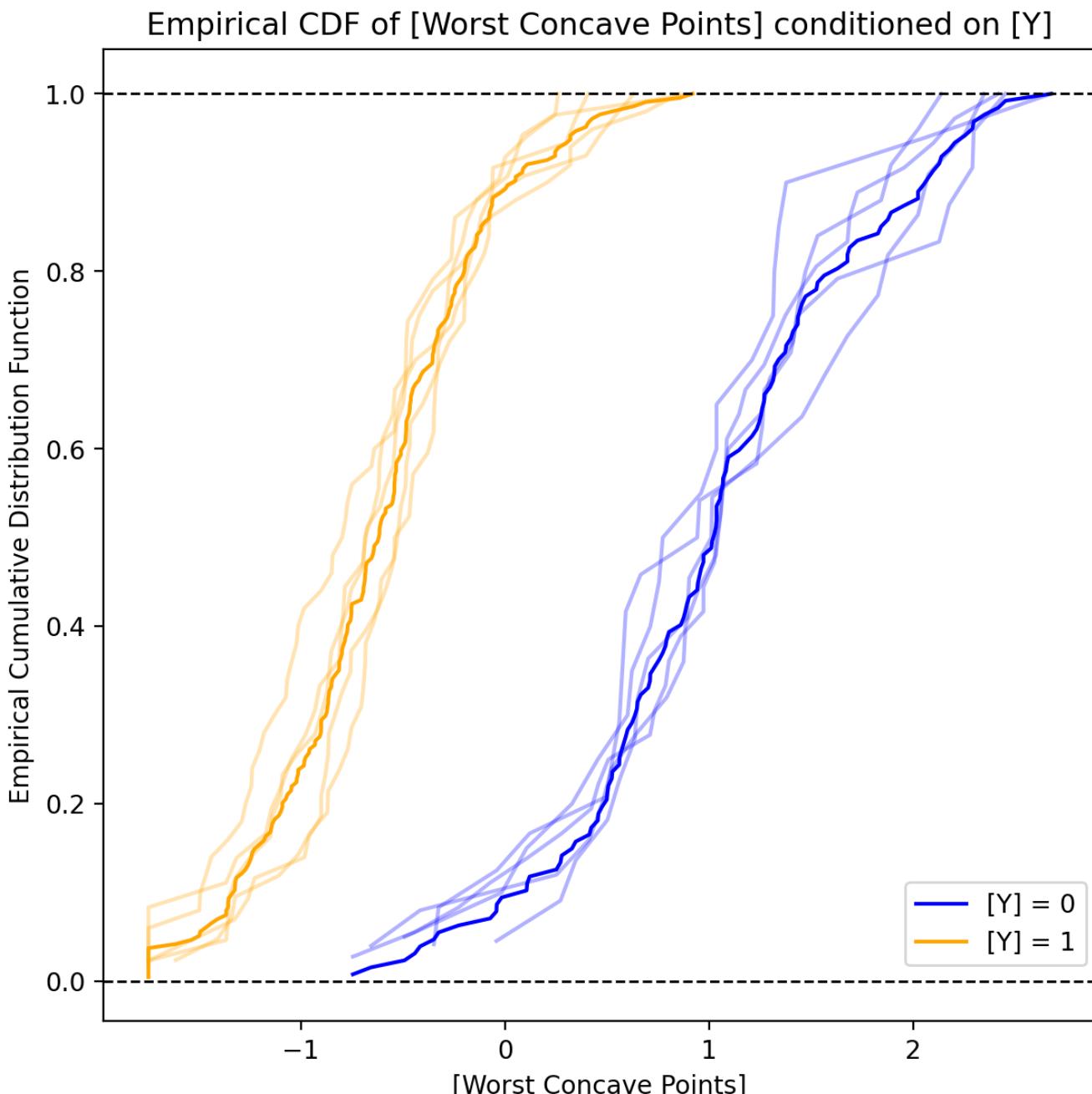
Kernel Density Plot of [Worst Concave Points] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

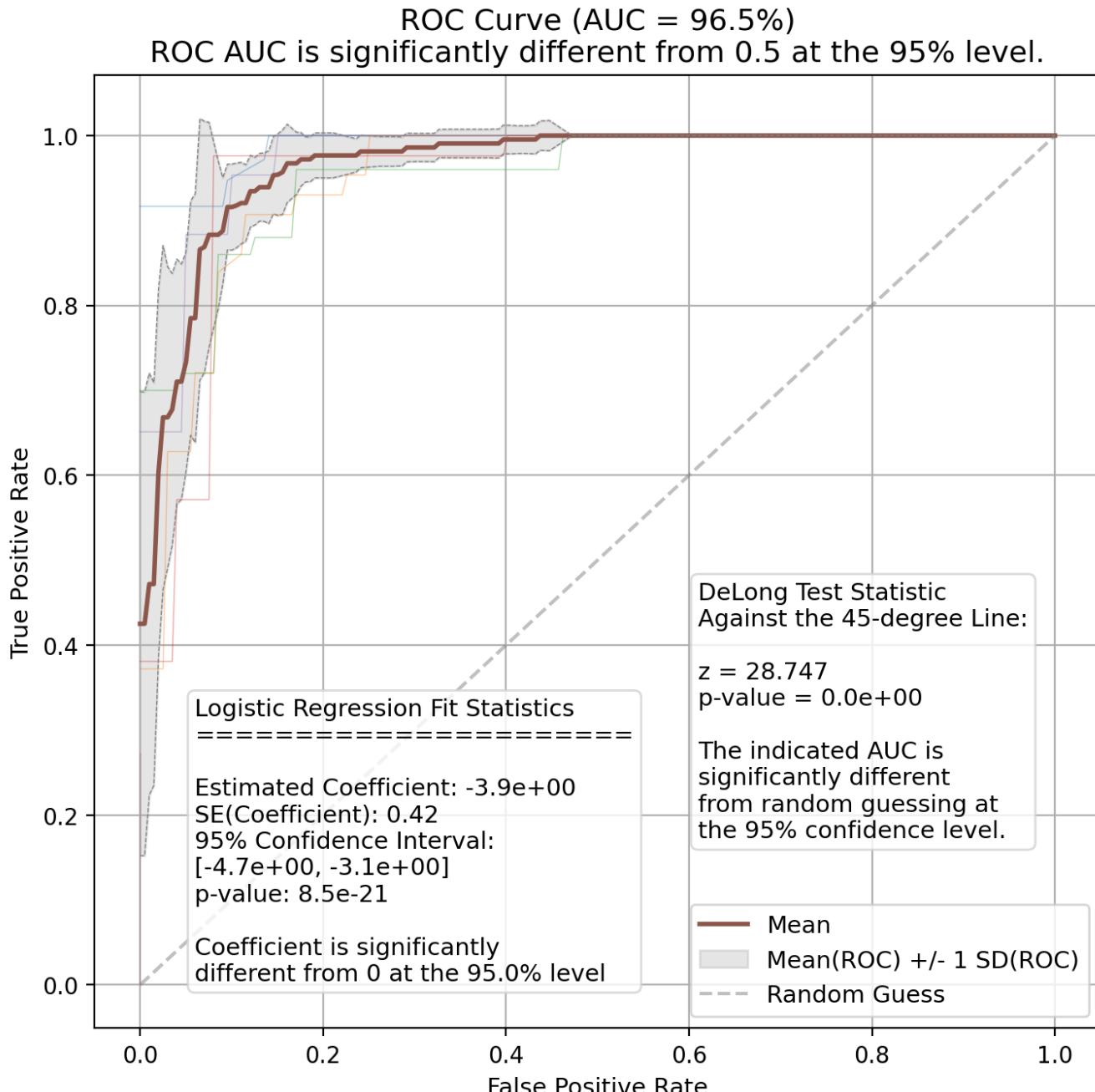
Worst Concave Points - Empirical CDF Plot



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

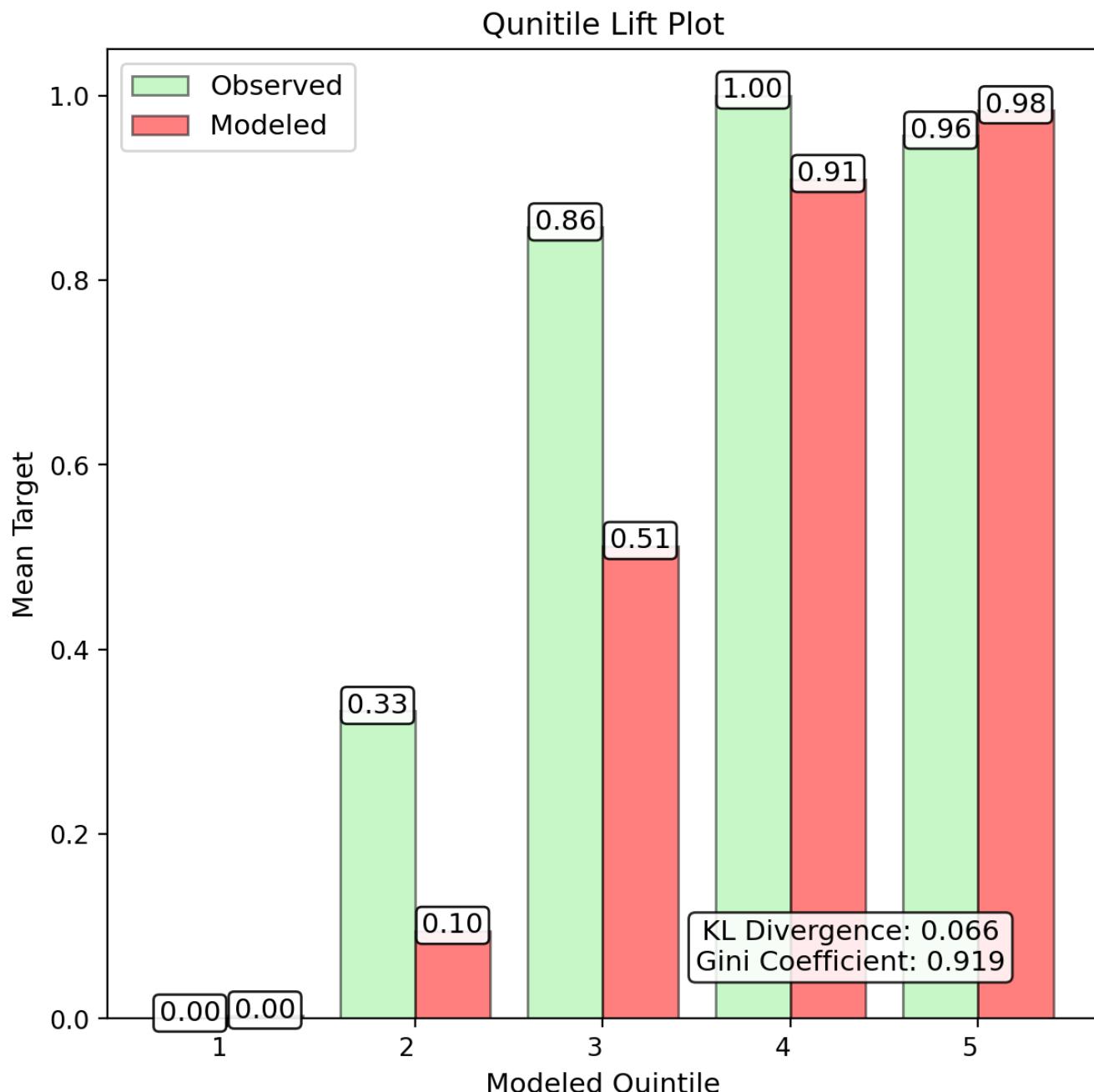


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

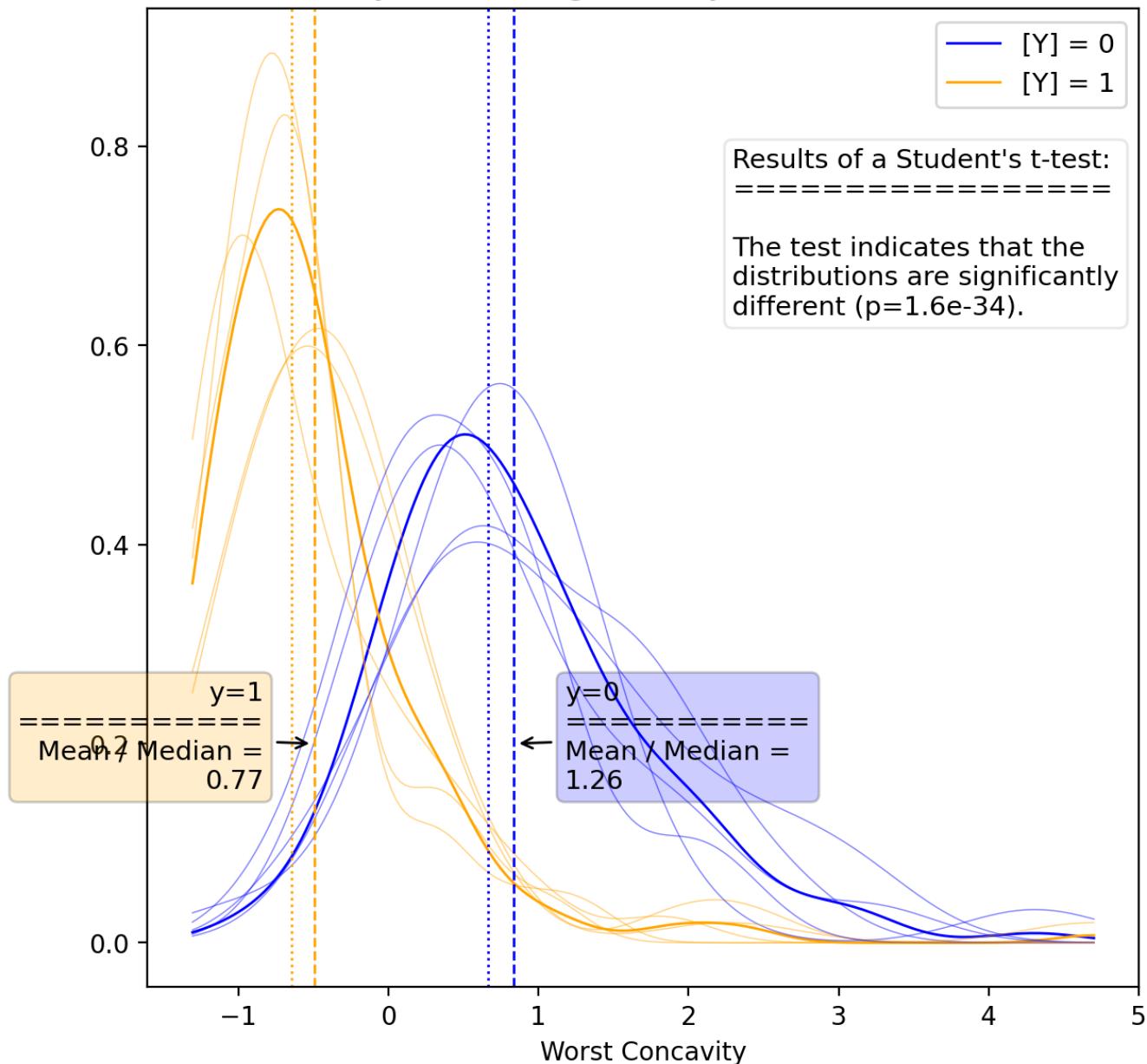
Worst Concavity - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-2.08	-2.36	-2.22	-2.48	-2.13	-2.24	0.16
Fitted p-Value	1.6e-17	8.6e-18	4.6e-17	2.4e-18	2.5e-17	1.1e-21	1.7e-17
Fitted Std. Err.	0.244	0.274	0.265	0.283	0.251	0.235	0.016
Conf. Int. Lower	-2.56	-2.90	-2.74	-3.03	-2.62	-2.70	0.20
Conf. Int. Upper	-1.60	-1.82	-1.70	-1.92	-1.63	-1.78	0.13
Train Accuracy	84.1%	84.7%	85.0%	84.7%	84.2%	84.8%	0.4%
Val Accuracy	87.9%	82.3%	82.4%	83.6%	84.1%	84.2%	2.3%
Train AUC	84.4%	85.7%	85.3%	85.0%	84.5%	85.1%	0.5%
Val AUC	88.5%	82.8%	81.6%	85.3%	81.7%	84.1%	2.9%
Train F1	86.8%	87.6%	87.3%	87.3%	86.6%	87.3%	0.4%
Test F1	89.9%	82.5%	86.6%	85.7%	88.4%	87.0%	2.8%
Train Precision	90.8%	93.4%	90.8%	91.1%	90.4%	91.3%	1.2%
Val Precision	93.9%	89.2%	89.4%	94.3%	88.4%	89.6%	2.8%
Train Recall	83.1%	82.5%	84.1%	83.7%	83.0%	83.6%	0.7%
Val Recall	86.1%	76.7%	84.0%	78.6%	88.4%	84.5%	4.9%
Train MCC	67.3%	68.9%	69.4%	68.5%	67.7%	68.7%	0.8%
Val MCC	75.5%	65.5%	61.4%	68.3%	63.4%	67.2%	5.5%
Train Log-Loss	5.73	5.50	5.40	5.52	5.70	5.50	0.14
Val Log-Loss	4.35	6.39	6.33	5.92	5.72	5.69	0.83

Univariate Report

Worst Concavity - Kernel Density Plot

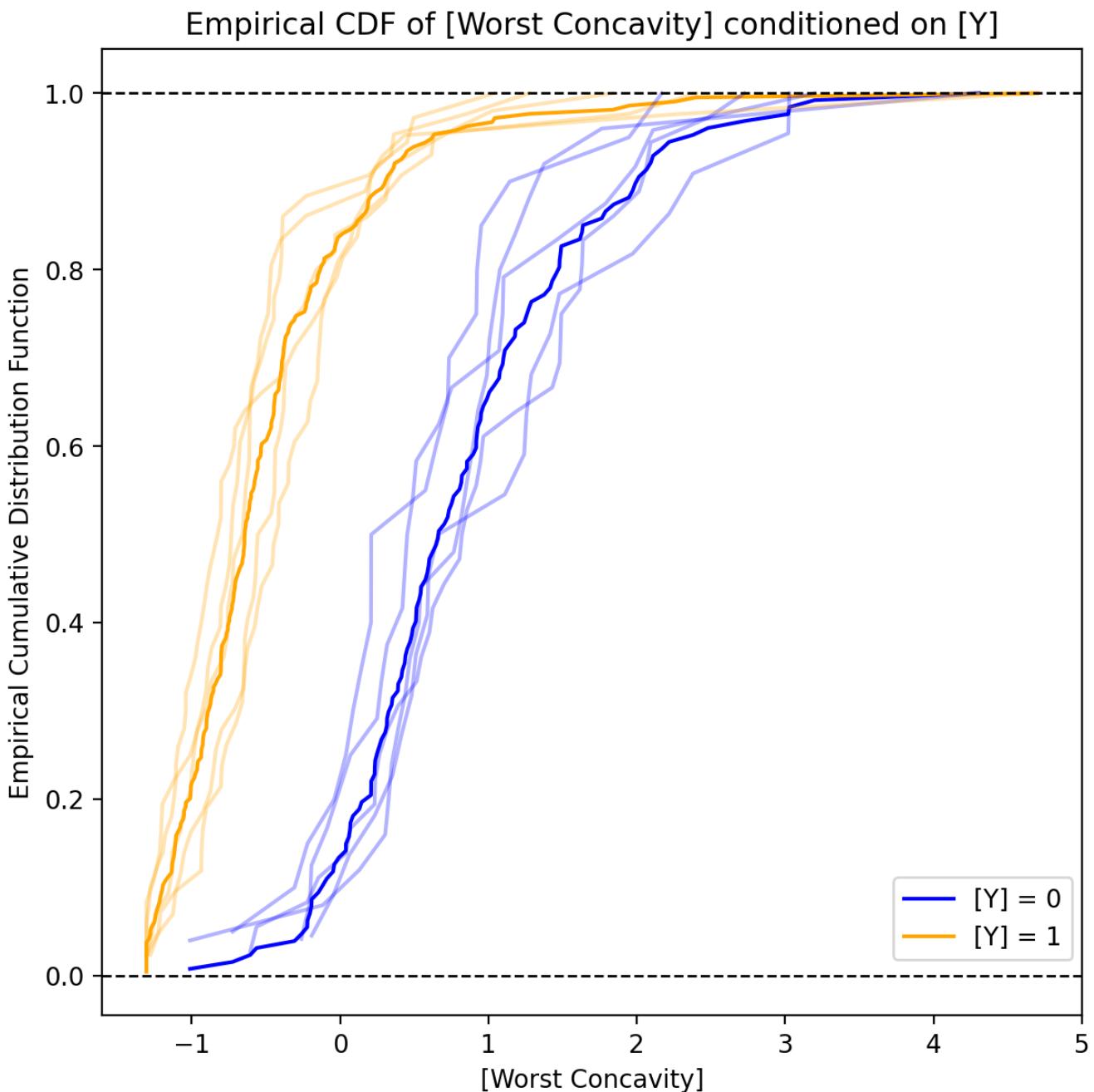
Kernel Density Plot of [Worst Concavity] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

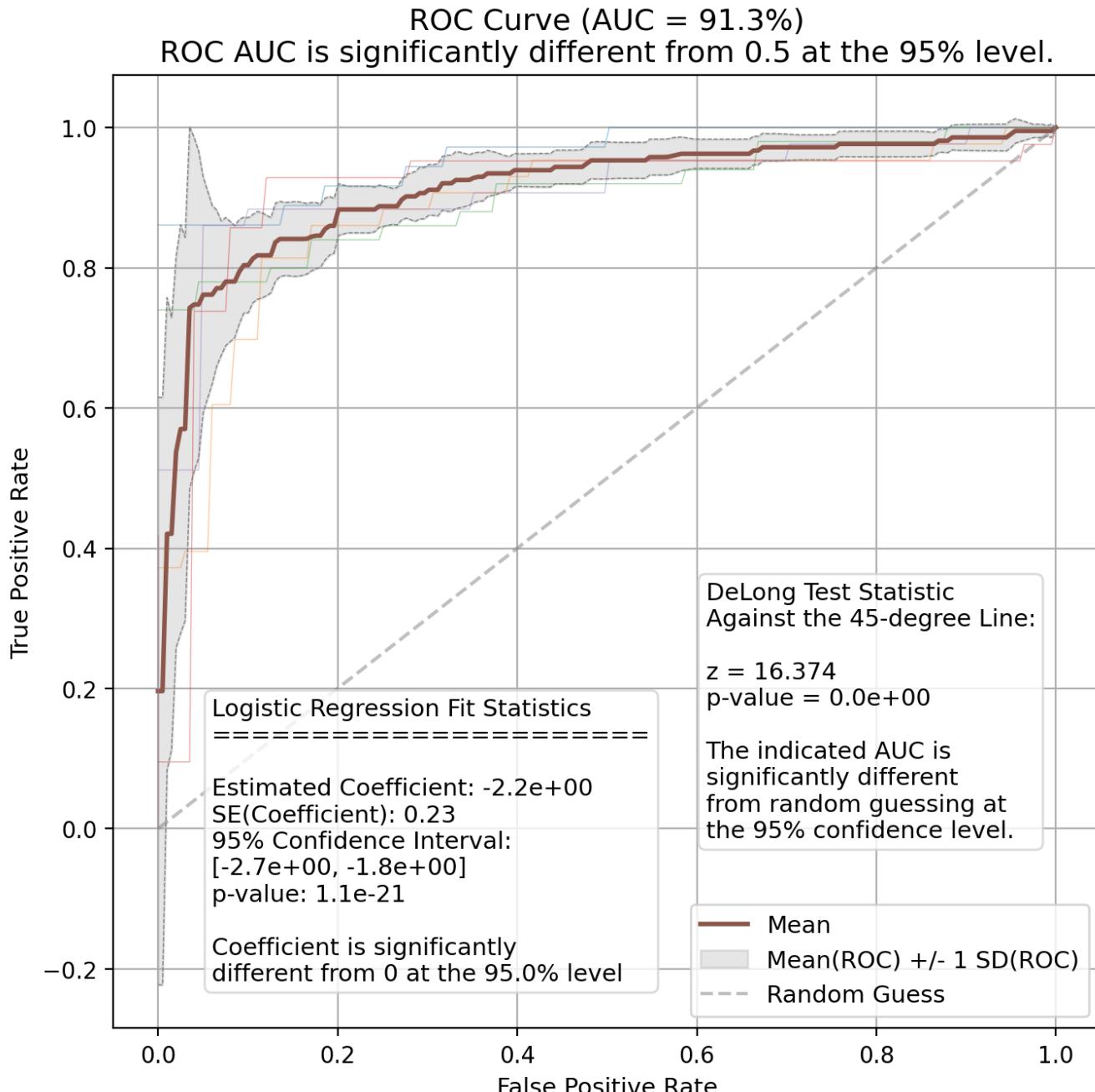
Worst Concavity - Empirical CDF Plot



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

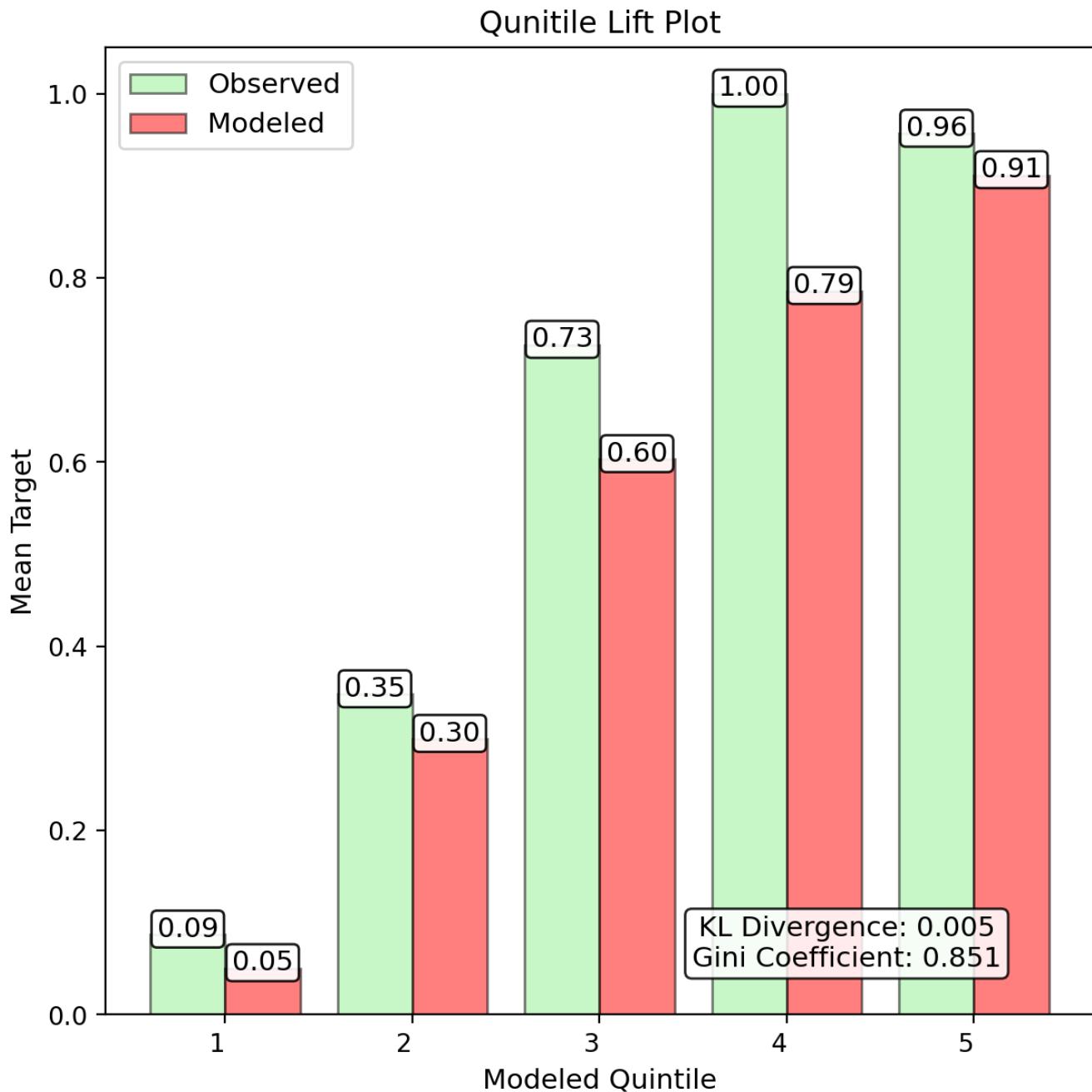


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

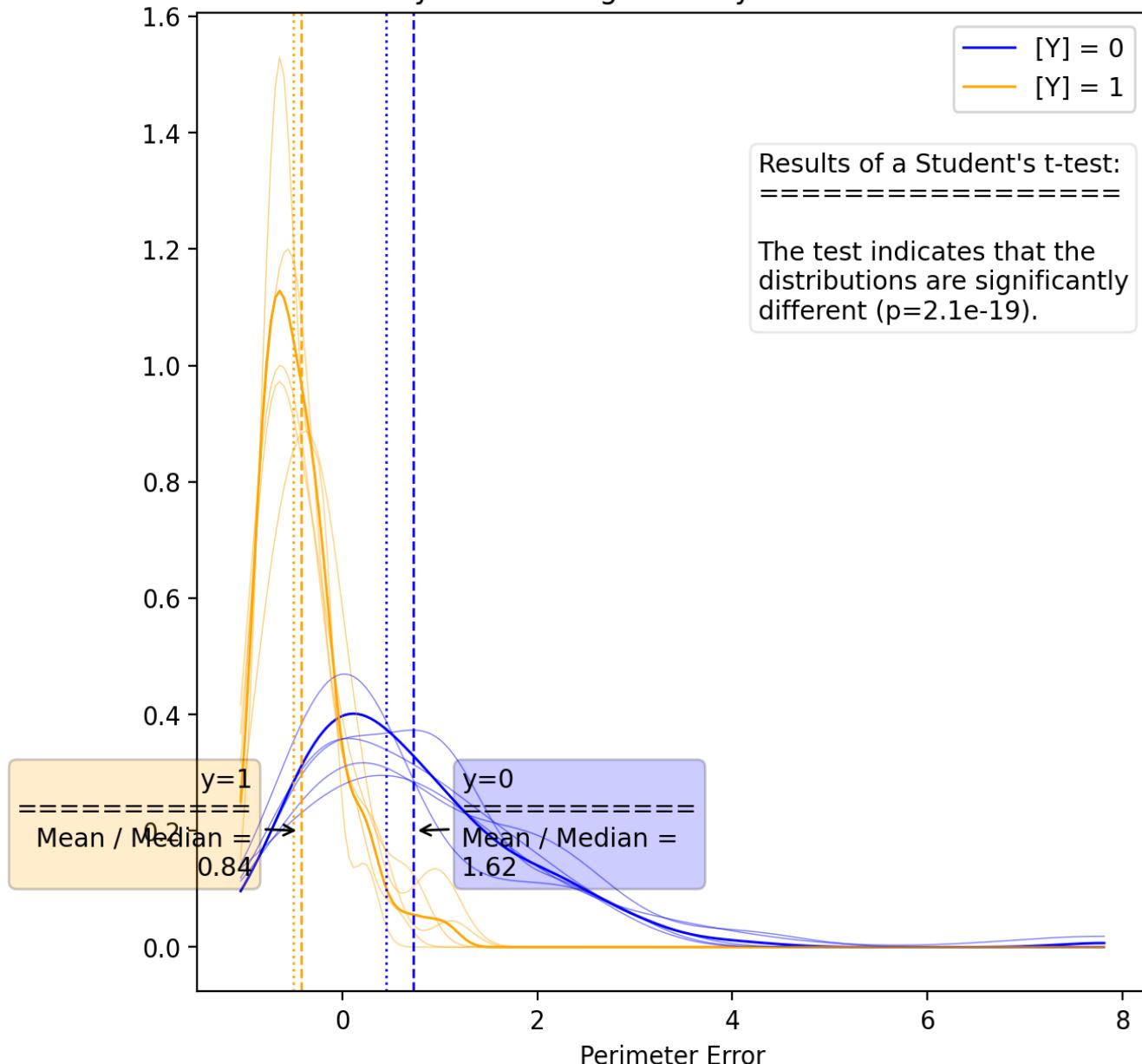
Perimeter Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-3.01	-2.72	-3.01	-2.74	-2.91	-2.87	0.14
Fitted p-Value	2.4e-18	1.0e-15	5.5e-17	1.0e-16	3.8e-17	7.4e-21	4.3e-16
Fitted Std. Err.	0.345	0.339	0.359	0.330	0.345	0.307	0.011
Conf. Int. Lower	-3.69	-3.38	-3.71	-3.38	-3.58	-3.48	0.16
Conf. Int. Upper	-2.33	-2.05	-2.30	-2.09	-2.23	-2.27	0.13
Train Accuracy	82.0%	80.2%	81.6%	81.0%	81.3%	80.9%	0.7%
Val Accuracy	81.0%	84.8%	78.4%	82.1%	79.4%	82.5%	2.5%
Train AUC	79.8%	78.1%	79.5%	79.1%	79.2%	78.9%	0.6%
Val AUC	81.2%	83.8%	75.3%	80.0%	75.5%	80.0%	3.7%
Train F1	86.0%	84.8%	85.6%	85.1%	85.3%	85.1%	0.5%
Test F1	84.1%	87.2%	84.0%	86.0%	85.1%	86.5%	1.4%
Train Precision	84.0%	84.8%	82.5%	83.7%	82.5%	83.4%	1.0%
Val Precision	87.9%	80.4%	84.0%	84.1%	84.1%	83.1%	2.6%
Train Recall	88.2%	84.8%	89.0%	86.6%	88.3%	86.9%	1.7%
Val Recall	80.6%	95.3%	84.0%	88.1%	86.0%	90.1%	5.5%
Train MCC	60.8%	56.2%	60.7%	59.0%	59.9%	58.7%	1.9%
Val MCC	61.1%	70.4%	50.7%	61.2%	51.8%	62.0%	8.1%
Train Log-Loss	6.50	7.15	6.61	6.84	6.74	6.87	0.25
Val Log-Loss	6.84	5.47	7.79	6.46	7.44	6.32	0.90

Univariate Report

Perimeter Error - Kernel Density Plot

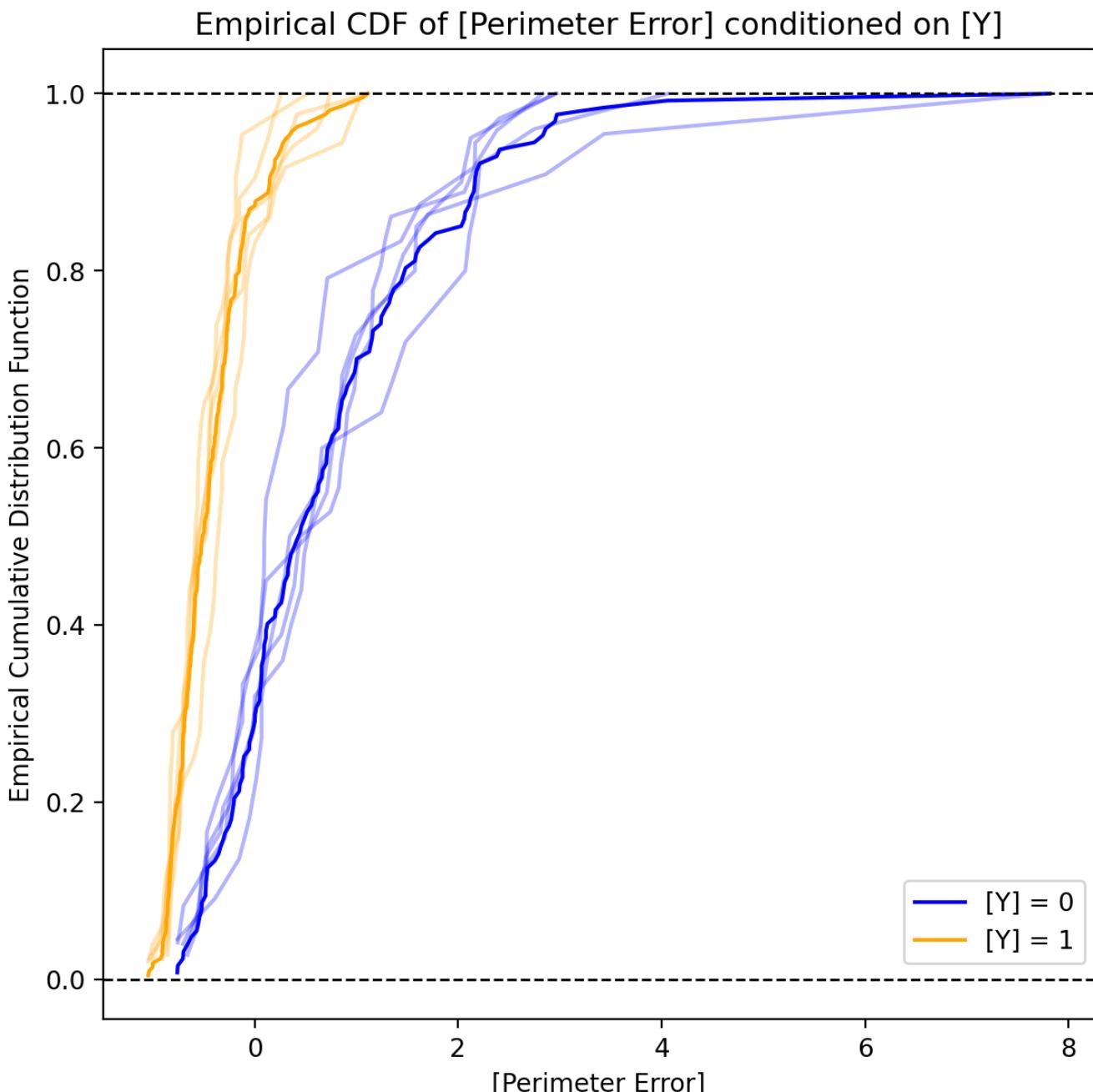
Kernel Density Plot of [Perimeter Error] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

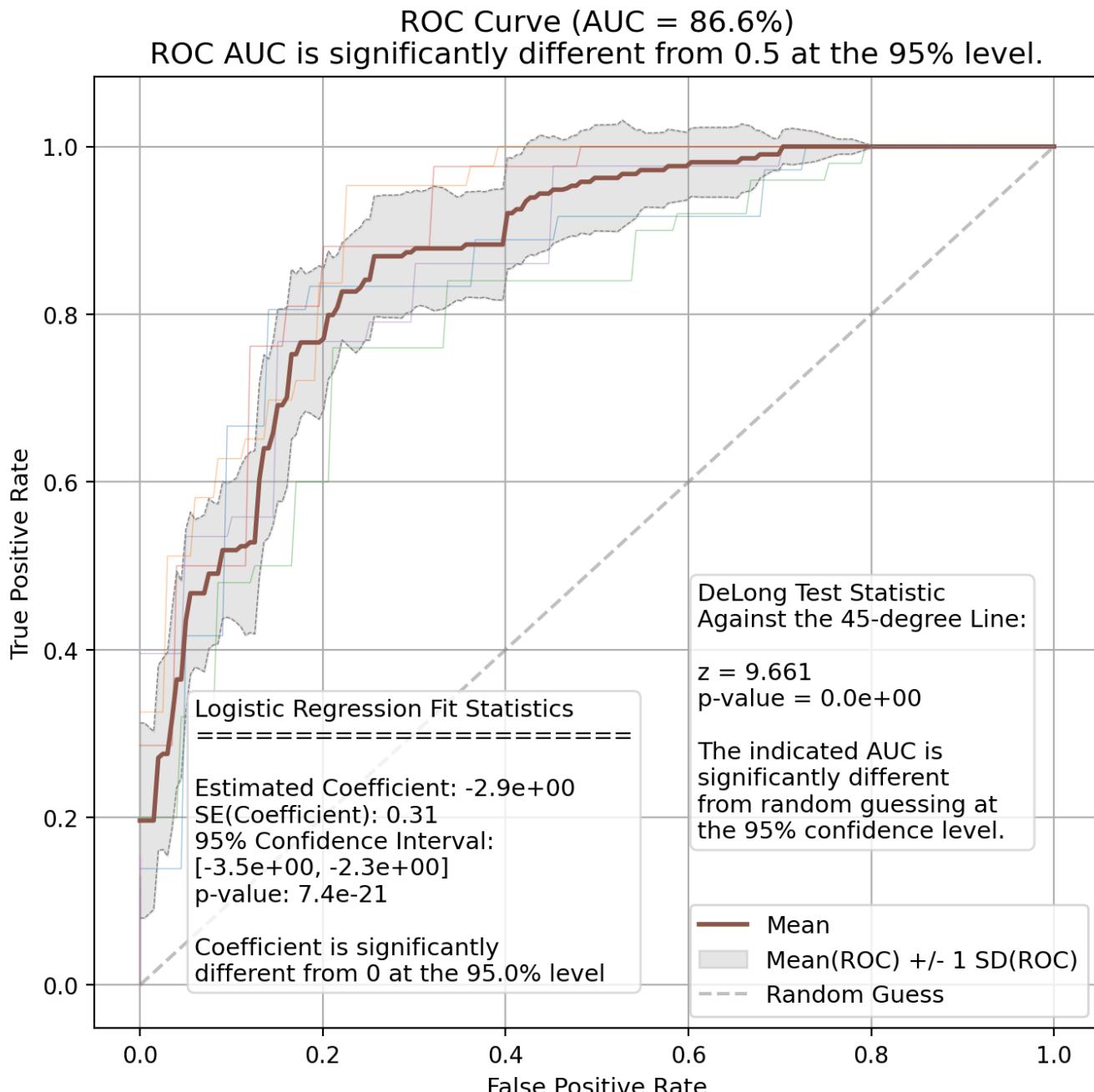
Perimeter Error - Empirical CDF Plot



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

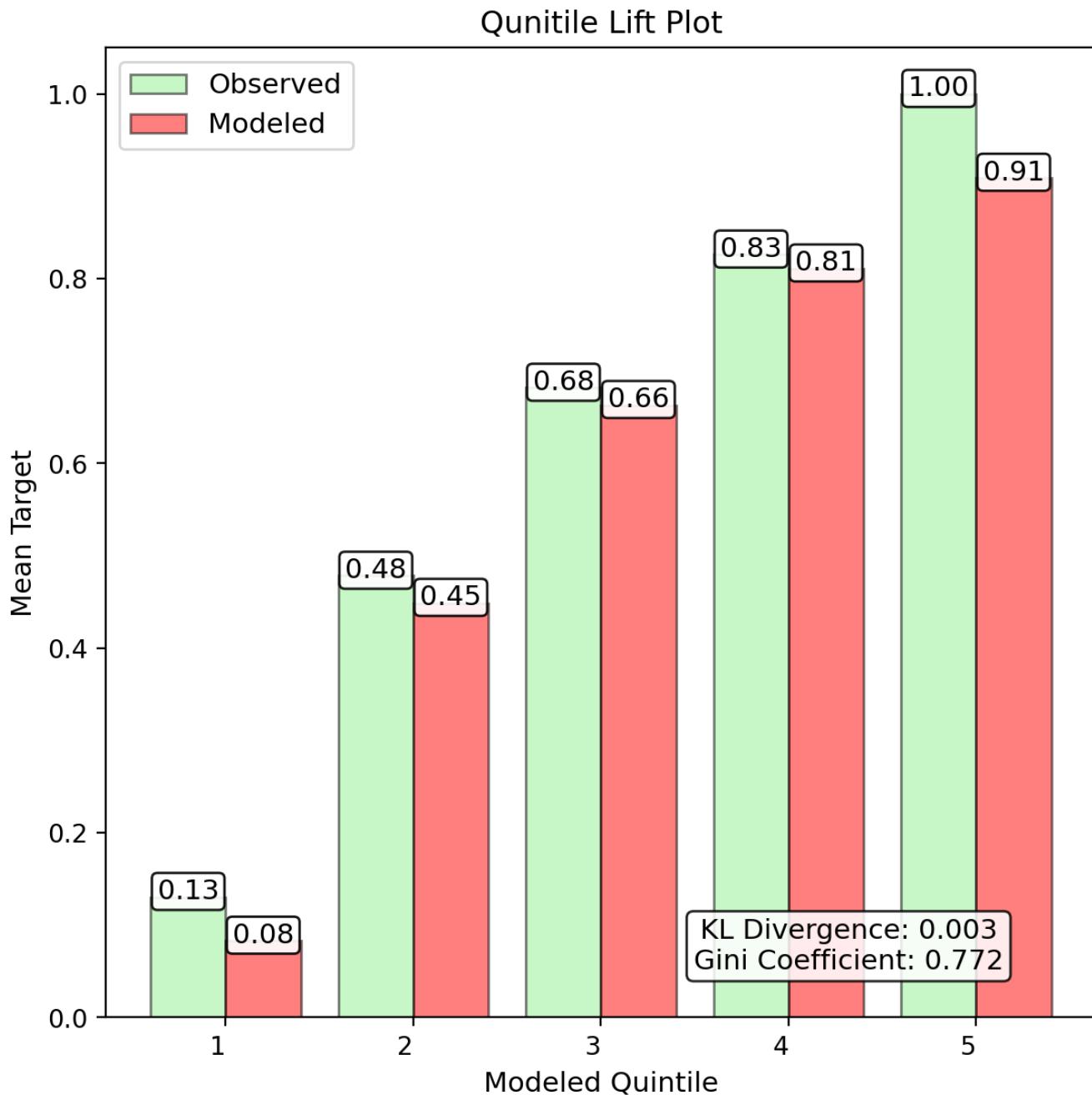


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

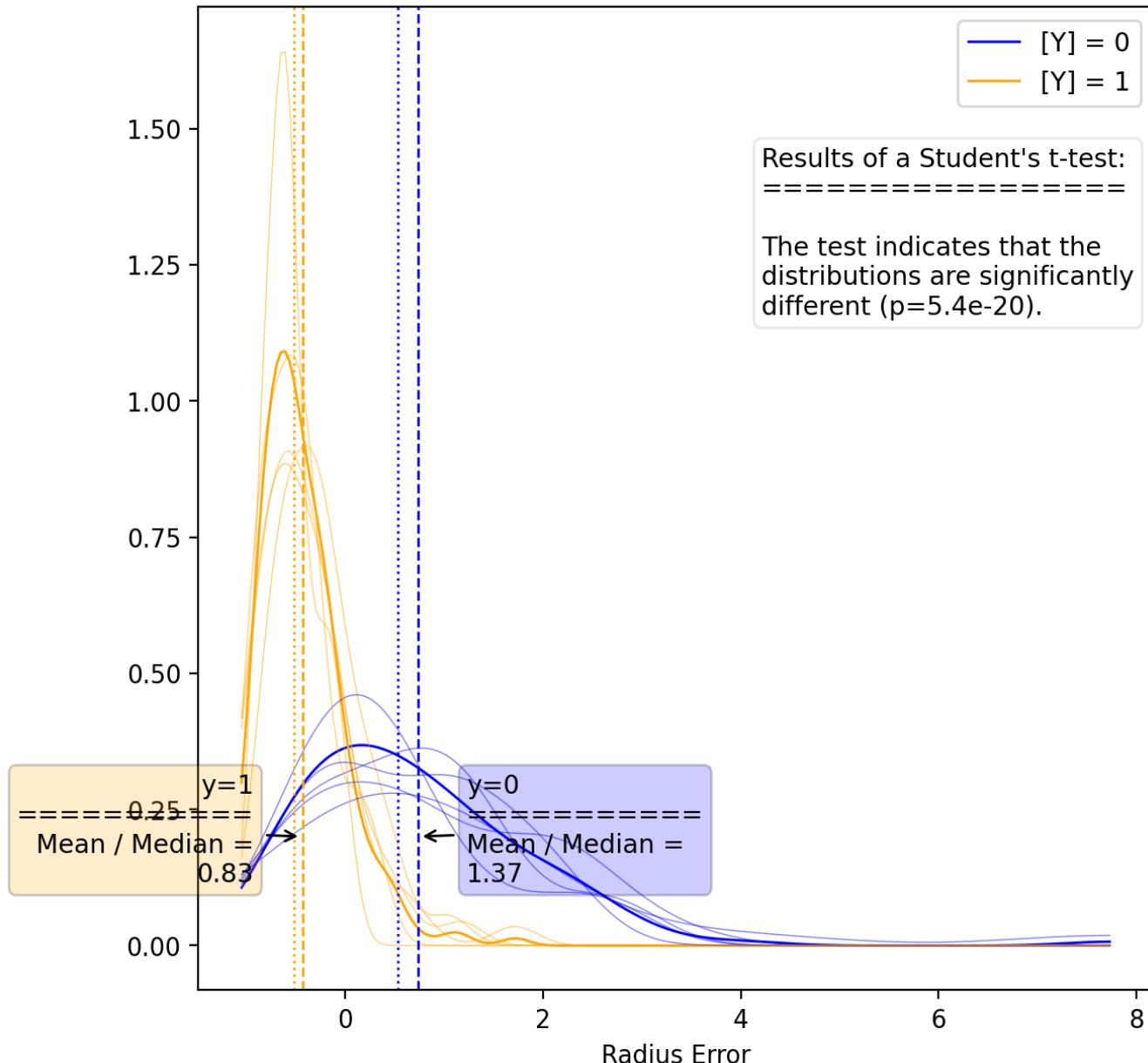
Radius Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-2.79	-2.58	-2.85	-2.66	-2.79	-2.73	0.11
Fitted p-Value	3.9e-18	2.3e-15	6.8e-17	1.3e-16	3.1e-17	1.1e-20	9.9e-16
Fitted Std. Err.	0.321	0.326	0.342	0.321	0.331	0.293	0.009
Conf. Int. Lower	-3.42	-3.22	-3.52	-3.29	-3.44	-3.31	0.12
Conf. Int. Upper	-2.16	-1.94	-2.18	-2.03	-2.14	-2.16	0.10
Train Accuracy	81.6%	79.0%	81.6%	81.4%	82.0%	80.9%	1.2%
Val Accuracy	79.3%	86.1%	81.1%	79.1%	79.4%	80.7%	3.0%
Train AUC	79.3%	77.0%	79.6%	79.2%	80.0%	78.7%	1.2%
Val AUC	80.7%	84.9%	77.3%	76.9%	75.5%	78.5%	3.8%
Train F1	85.8%	83.9%	85.5%	85.6%	85.9%	85.2%	0.8%
Test F1	81.8%	88.4%	86.3%	83.7%	85.1%	84.9%	2.5%
Train Precision	83.5%	84.1%	82.9%	83.4%	83.1%	83.1%	0.5%
Val Precision	90.0%	80.8%	84.6%	81.8%	84.1%	82.7%	3.6%
Train Recall	88.2%	83.6%	88.4%	87.8%	88.9%	87.4%	2.1%
Val Recall	75.0%	97.7%	88.0%	85.7%	86.0%	87.3%	8.1%
Train MCC	60.0%	53.8%	60.7%	59.6%	61.5%	58.6%	3.1%
Val MCC	59.6%	73.4%	56.0%	54.7%	51.8%	58.3%	8.5%
Train Log-Loss	6.62	7.57	6.61	6.71	6.48	6.87	0.44
Val Log-Loss	7.46	5.02	6.82	7.53	7.44	6.96	1.06

Univariate Report

Radius Error - Kernel Density Plot

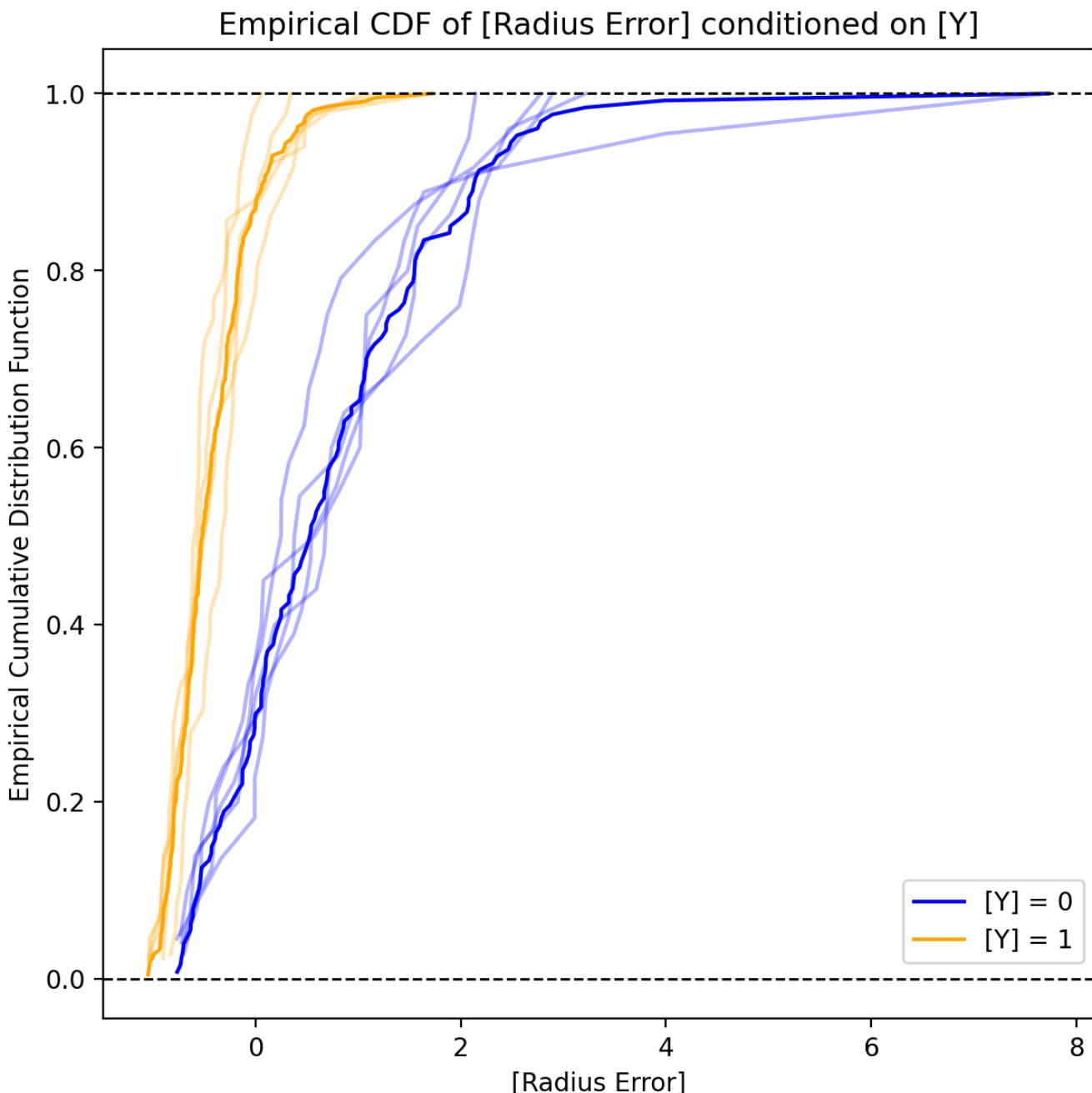
Kernel Density Plot of [Radius Error] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

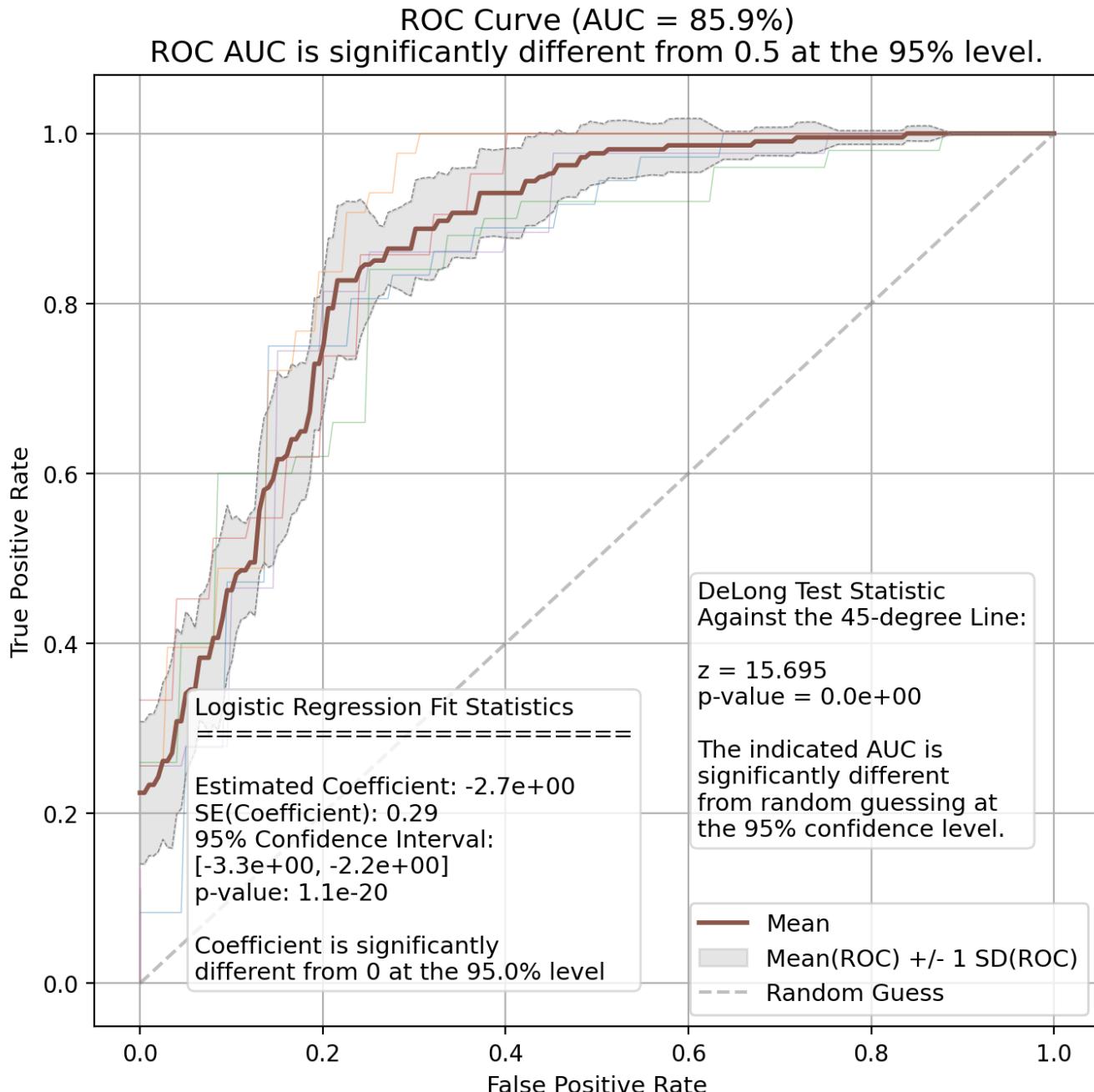
Radius Error - Empirical CDF Plot



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

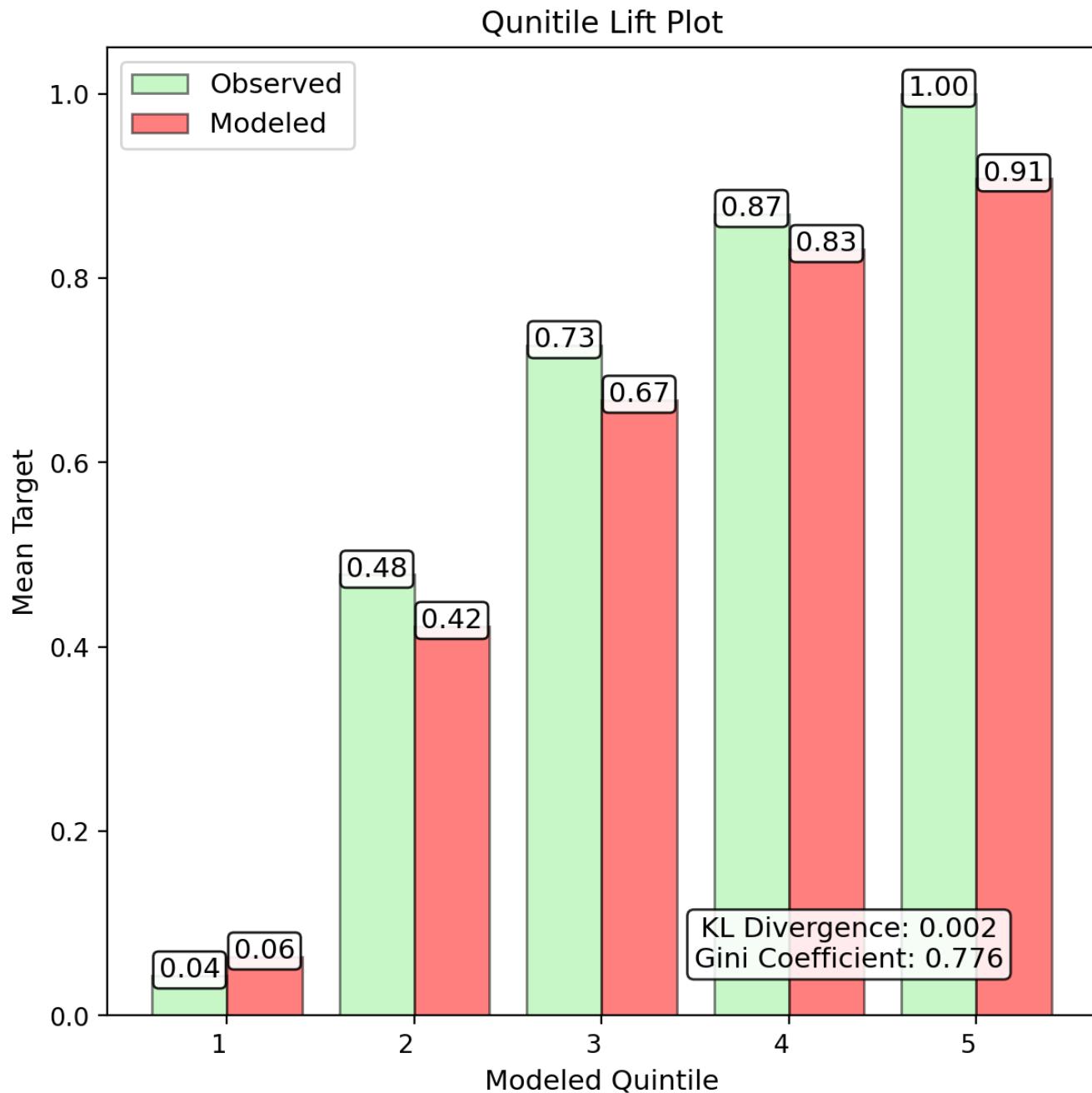


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

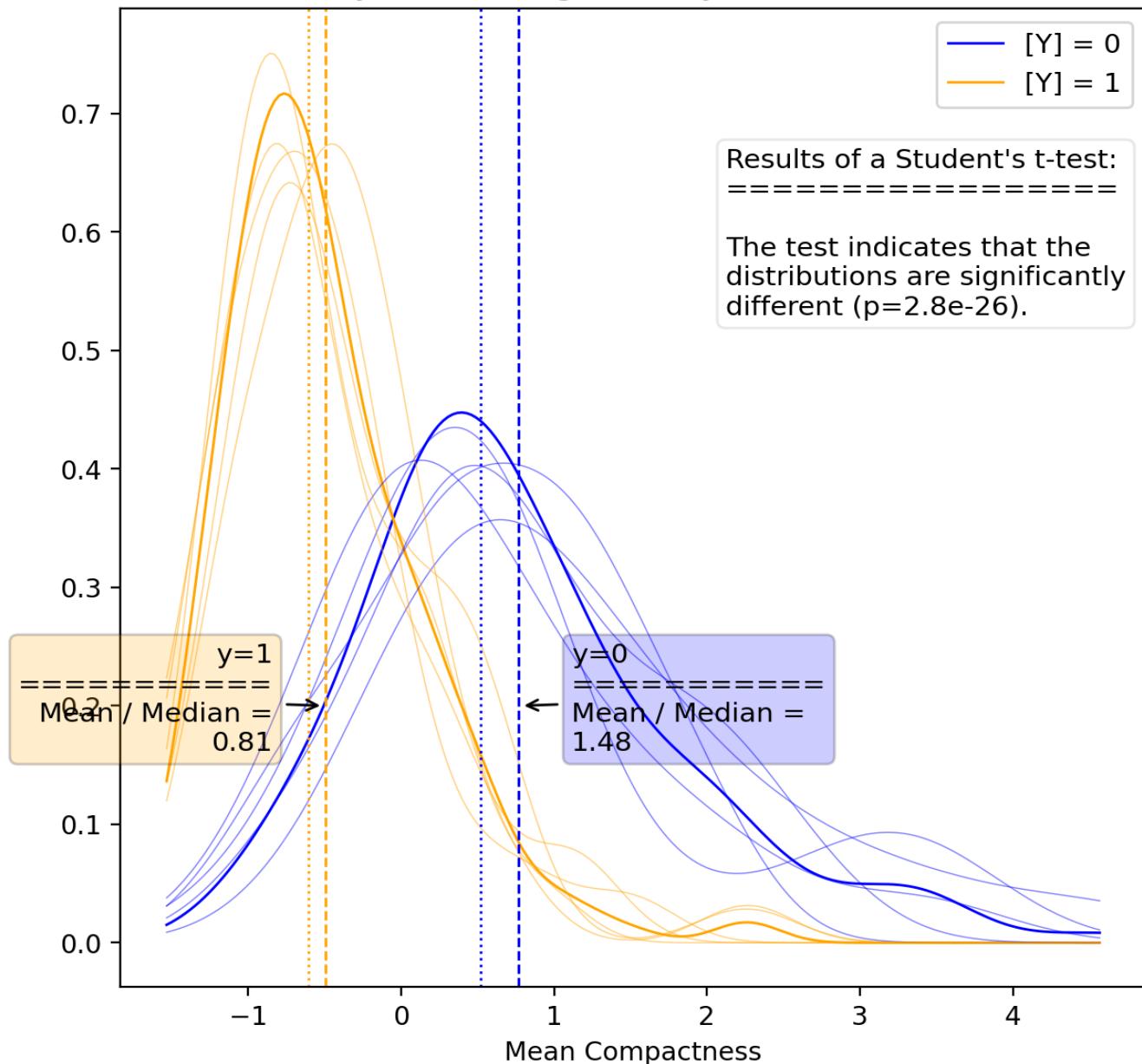
Mean Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-2.03	-2.38	-2.13	-2.15	-2.09	-2.15	0.13
Fitted p-Value	8.1e-17	2.0e-16	2.8e-16	3.6e-17	8.5e-17	1.6e-20	9.9e-17
Fitted Std. Err.	0.244	0.289	0.260	0.255	0.251	0.231	0.017
Conf. Int. Lower	-2.51	-2.94	-2.64	-2.65	-2.58	-2.60	0.16
Conf. Int. Upper	-1.56	-1.81	-1.62	-1.65	-1.60	-1.69	0.10
Train Accuracy	79.2%	82.8%	80.1%	81.0%	80.6%	81.2%	1.4%
Val Accuracy	89.7%	75.9%	82.4%	83.6%	77.8%	78.9%	5.4%
Train AUC	79.9%	83.8%	80.0%	81.5%	80.4%	81.5%	1.6%
Val AUC	90.8%	76.3%	82.7%	84.5%	75.7%	79.0%	6.2%
Train F1	82.3%	86.0%	83.3%	84.0%	83.7%	84.3%	1.4%
Test F1	91.2%	76.5%	86.3%	86.1%	83.3%	82.4%	5.4%
Train Precision	88.4%	92.0%	86.3%	89.0%	86.3%	88.7%	2.3%
Val Precision	96.9%	81.6%	91.1%	91.9%	85.4%	86.2%	6.0%
Train Recall	77.0%	80.7%	80.5%	79.7%	81.3%	80.4%	1.7%
Val Recall	86.1%	72.1%	82.0%	81.0%	81.4%	78.9%	5.1%
Train MCC	58.1%	65.0%	59.1%	61.4%	59.8%	61.6%	2.7%
Val MCC	79.6%	52.5%	62.6%	67.1%	50.2%	56.7%	11.9%
Train Log-Loss	7.51	6.19	7.15	6.84	7.00	6.76	0.49
Val Log-Loss	3.73	8.67	6.33	5.92	8.01	7.59	1.94

Univariate Report

Mean Compactness - Kernel Density Plot

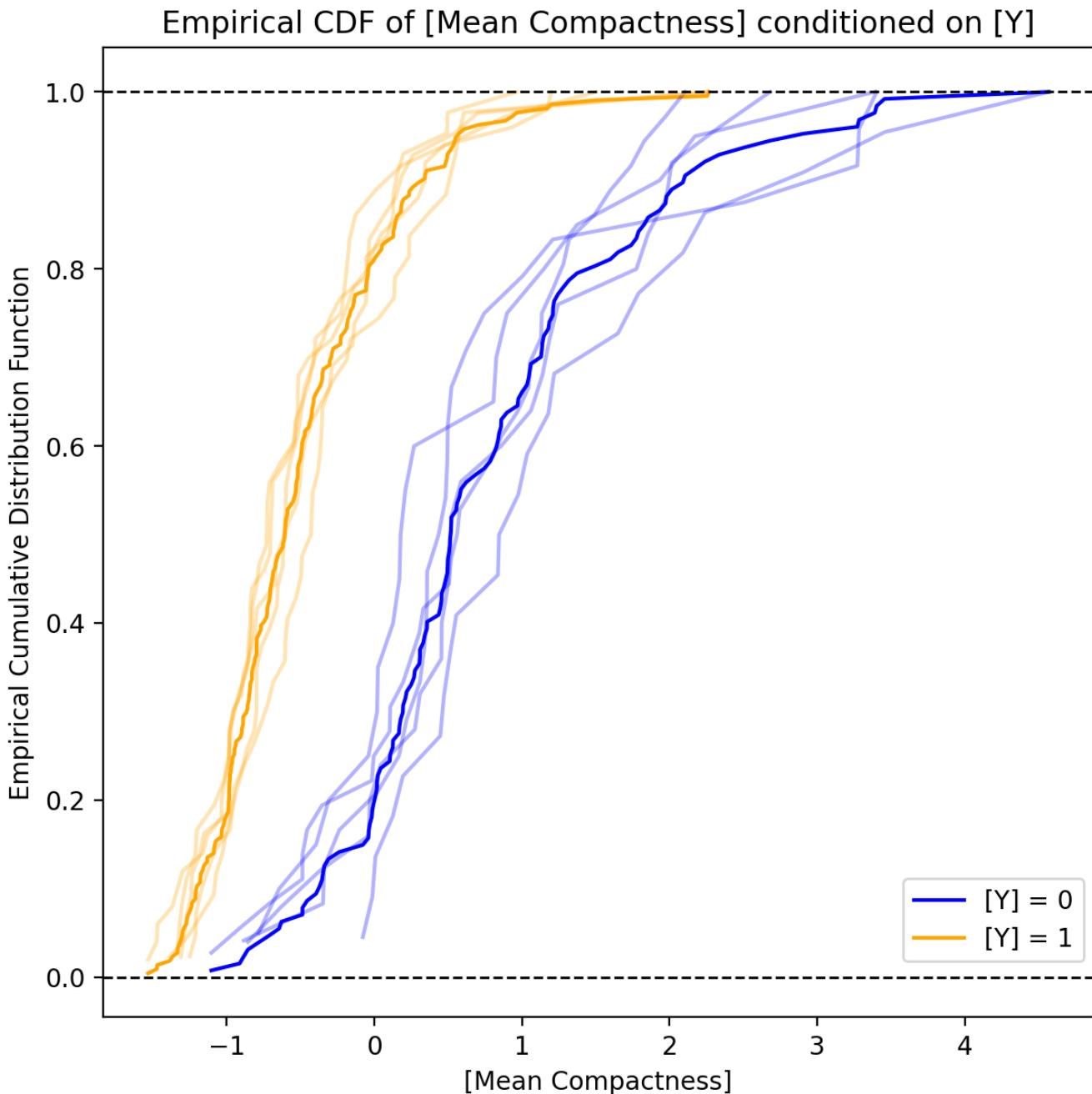
Kernel Density Plot of [Mean Compactness] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

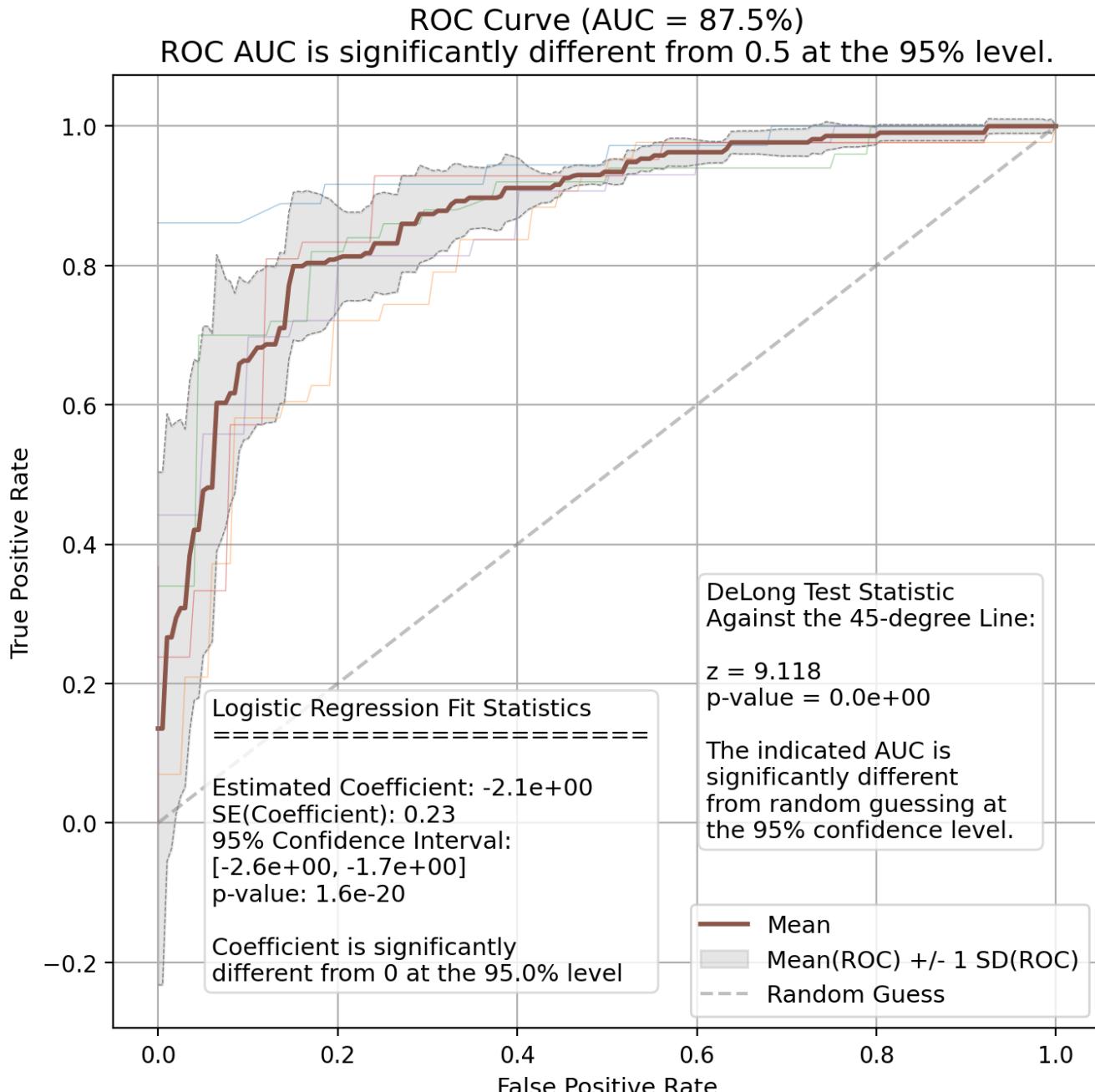
Mean Compactness - Empirical CDF Plot



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

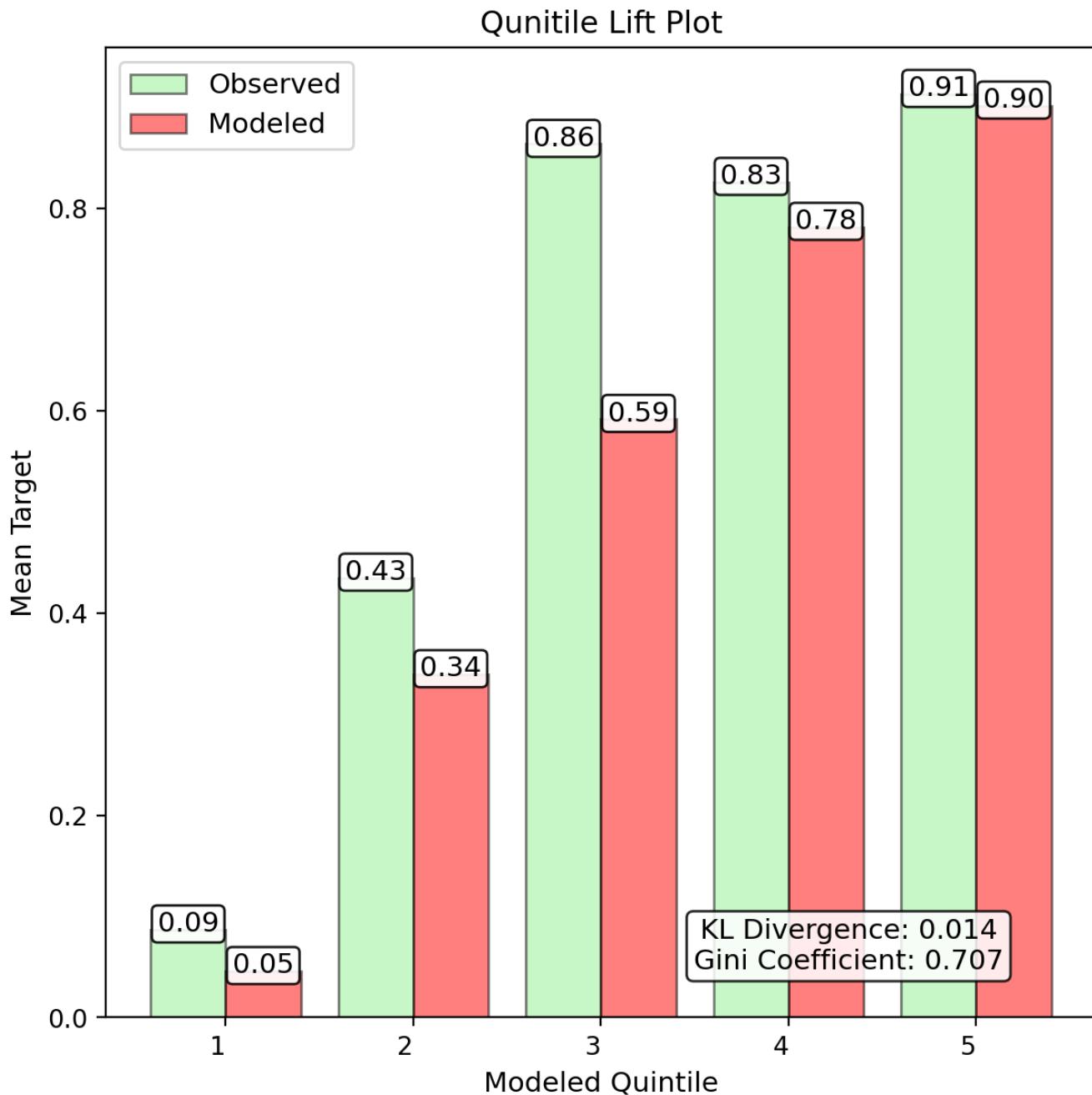


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

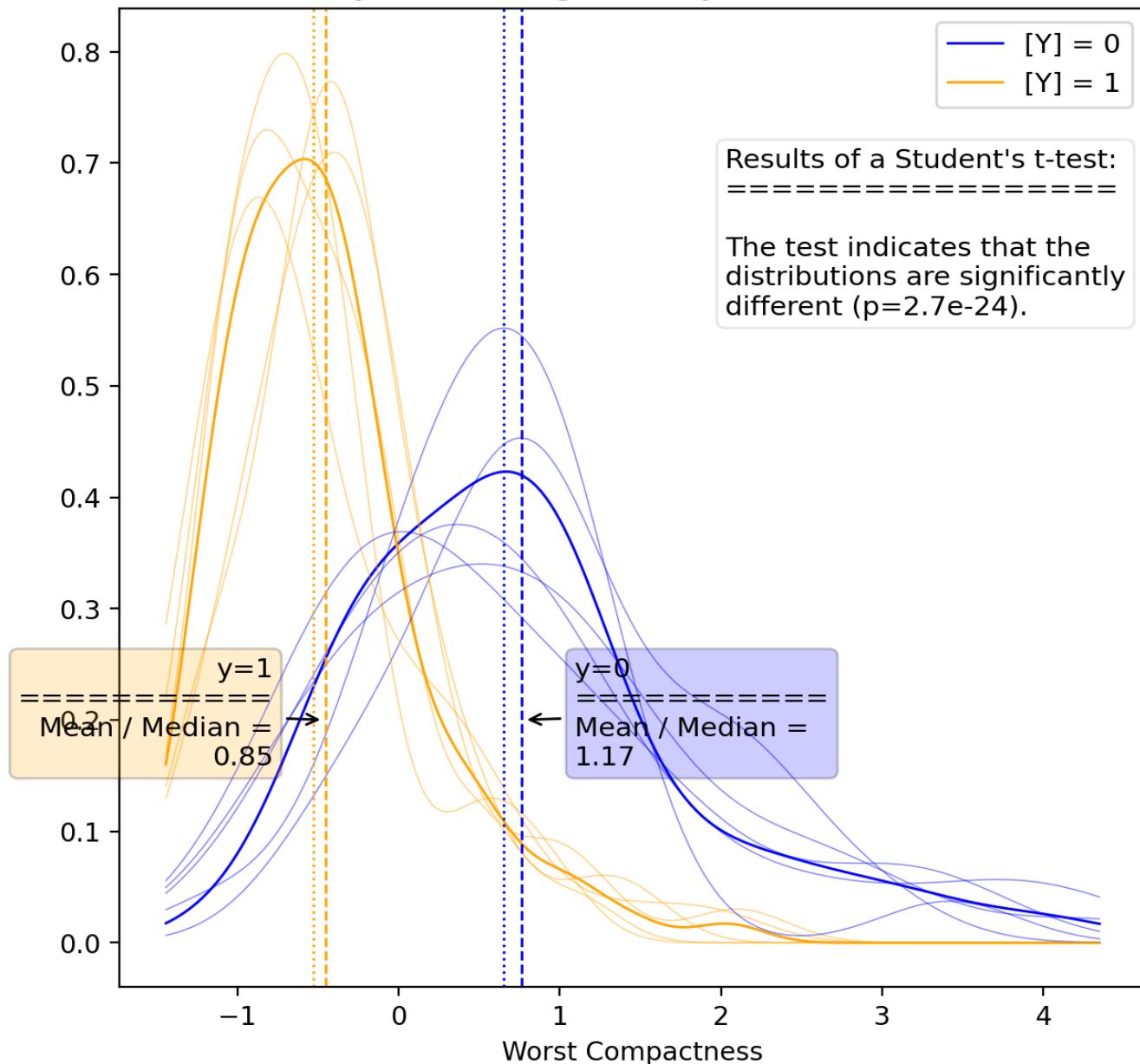
Worst Compactness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-1.92	-2.27	-2.18	-2.05	-2.02	-2.08	0.14
Fitted p-Value	1.1e-15	8.2e-17	4.8e-16	1.5e-16	2.0e-16	5.2e-20	4.3e-16
Fitted Std. Err.	0.240	0.273	0.269	0.248	0.246	0.227	0.015
Conf. Int. Lower	-2.39	-2.81	-2.71	-2.53	-2.50	-2.53	0.17
Conf. Int. Upper	-1.45	-1.74	-1.66	-1.56	-1.54	-1.64	0.11
Train Accuracy	78.4%	79.0%	80.9%	79.9%	80.2%	80.1%	1.0%
Val Accuracy	87.9%	75.9%	77.0%	80.6%	77.8%	75.4%	4.8%
Train AUC	77.0%	78.0%	79.8%	78.4%	79.5%	78.8%	1.1%
Val AUC	87.6%	75.6%	75.4%	80.5%	71.7%	75.2%	6.1%
Train F1	82.8%	83.5%	84.5%	84.1%	83.7%	84.0%	0.6%
Test F1	90.1%	78.2%	82.5%	84.0%	84.4%	79.4%	4.3%
Train Precision	83.1%	85.8%	84.2%	83.8%	84.9%	84.4%	1.1%
Val Precision	91.4%	77.3%	85.1%	87.2%	80.9%	83.1%	5.5%
Train Recall	82.6%	81.3%	84.8%	84.3%	82.5%	83.6%	1.4%
Val Recall	88.9%	79.1%	80.0%	81.0%	88.4%	76.1%	4.8%
Train MCC	53.9%	54.9%	59.6%	57.0%	58.6%	57.5%	2.4%
Val MCC	74.6%	51.4%	49.4%	59.8%	46.4%	49.4%	11.4%
Train Log-Loss	7.77	7.57	6.88	7.24	7.13	7.19	0.35
Val Log-Loss	4.35	8.67	8.28	6.99	8.01	8.85	1.74

Univariate Report

Worst Compactness - Kernel Density Plot

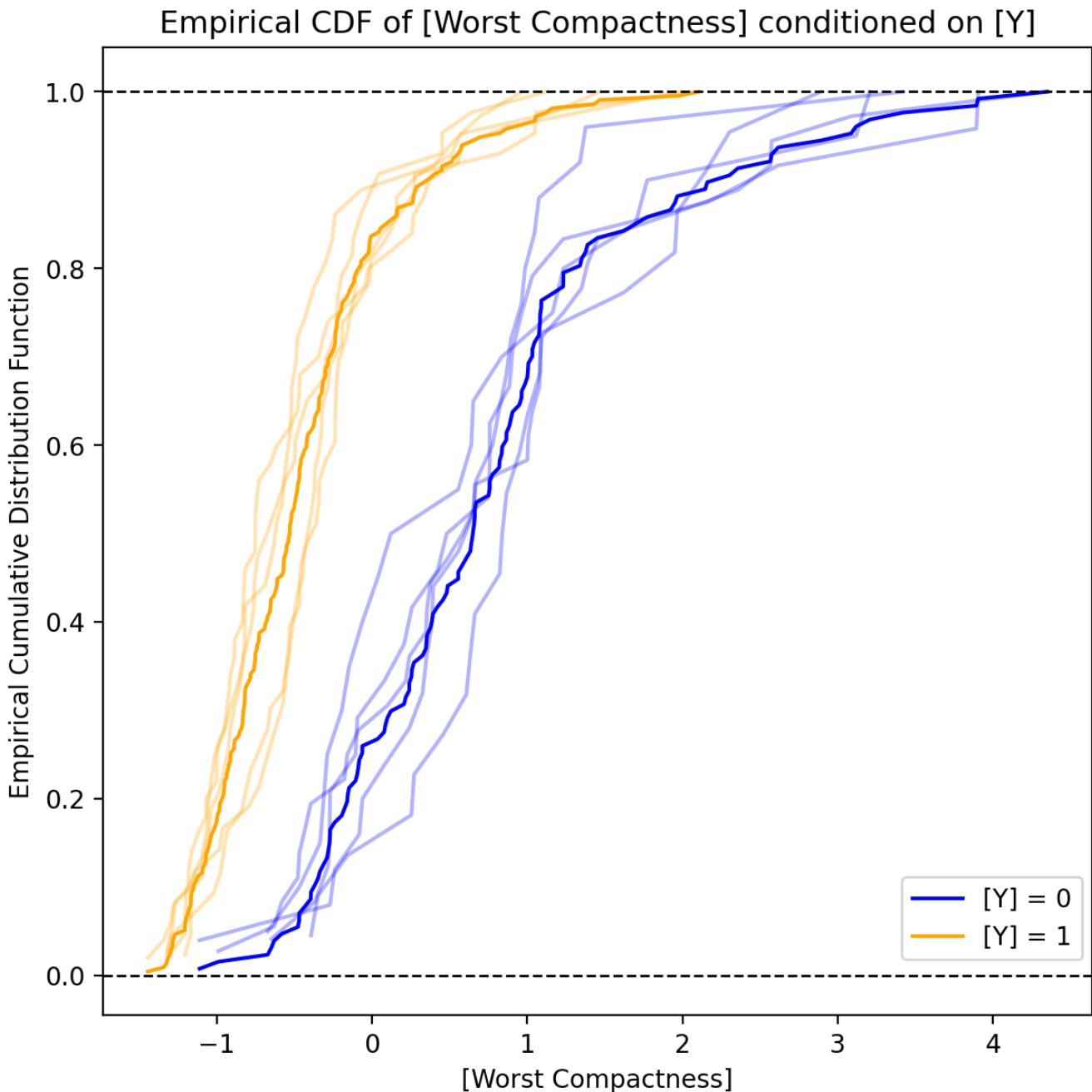
Kernel Density Plot of [Worst Compactness] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

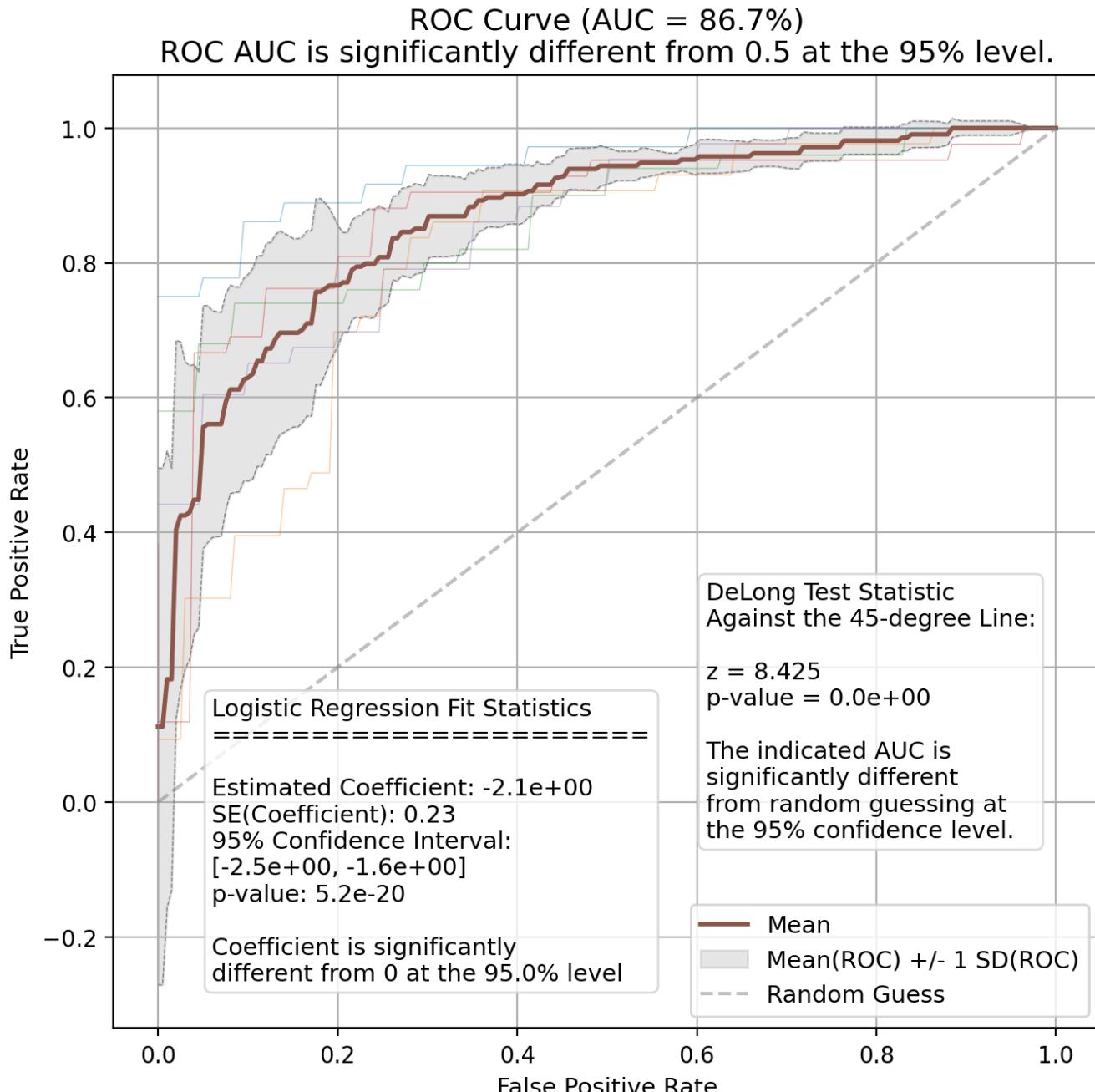
Worst Compactness - Empirical CDF Plot



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

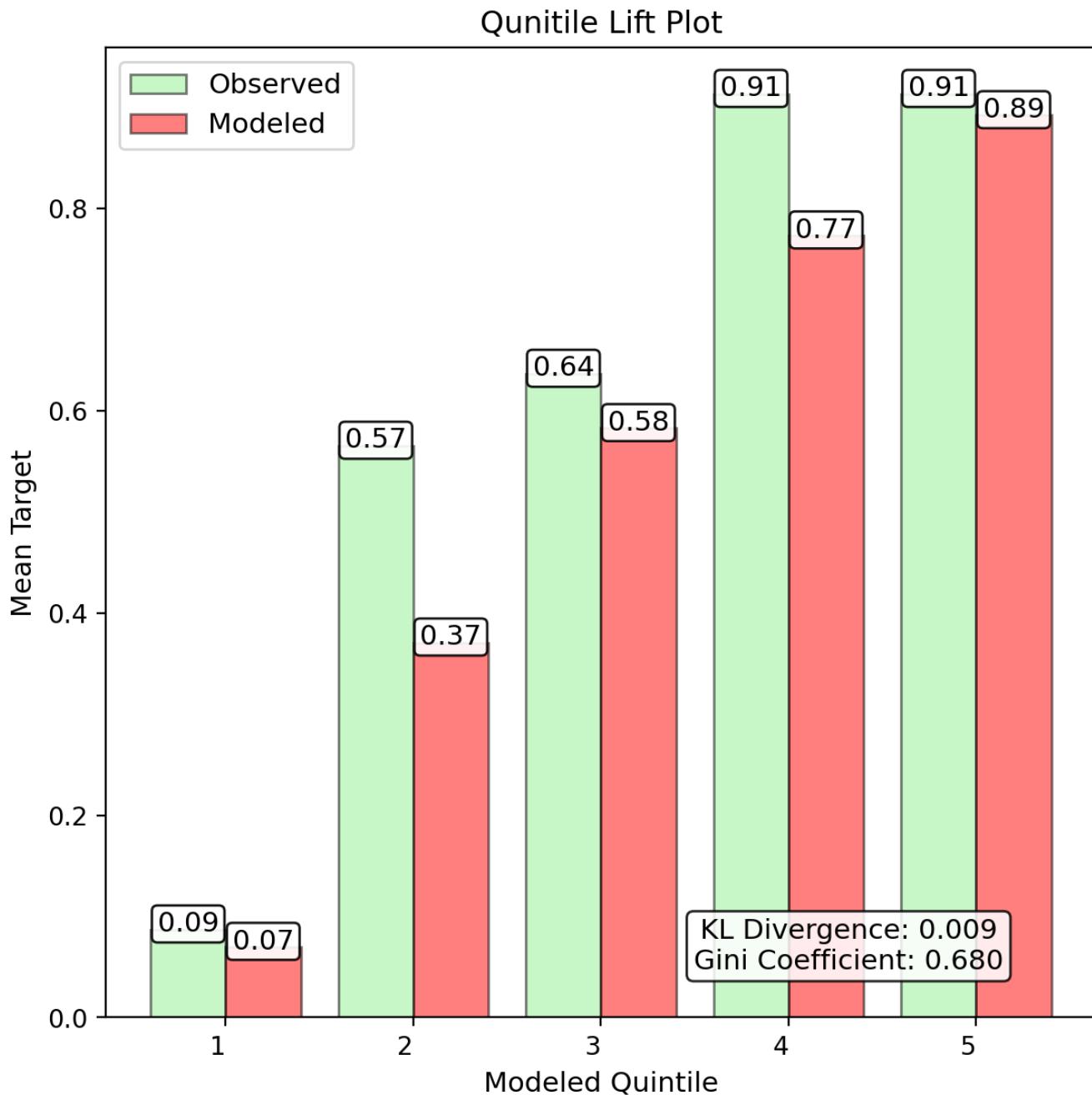


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

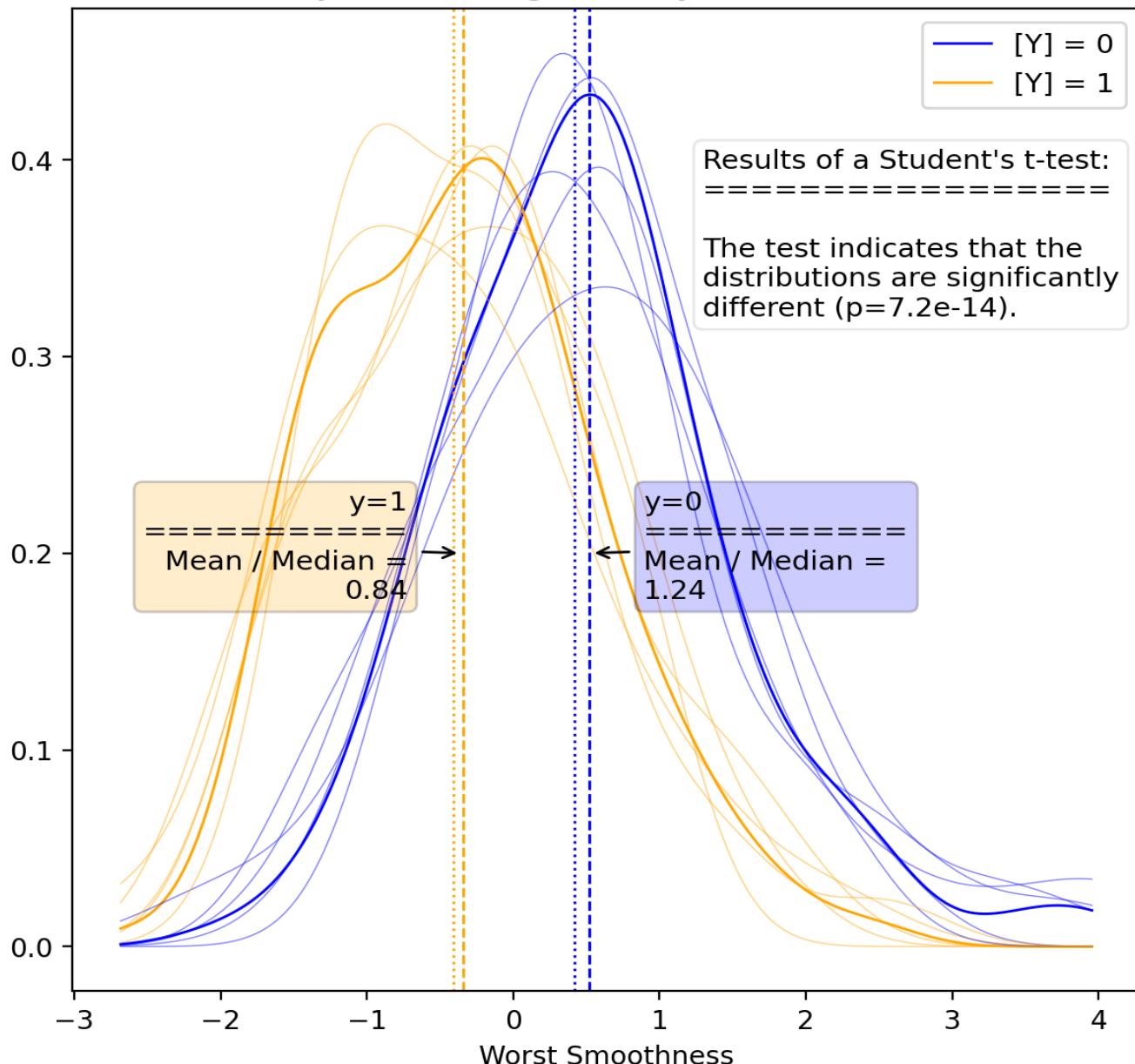
Worst Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.856	-0.956	-0.922	-0.948	-0.955	-0.926	0.042
Fitted p-Value	3.8e-09	1.5e-09	6.8e-09	5.7e-10	4.2e-10	1.5e-11	2.7e-09
Fitted Std. Err.	0.145	0.158	0.159	0.153	0.153	0.137	0.005
Conf. Int. Lower	-1.14	-1.27	-1.23	-1.25	-1.25	-1.20	0.05
Conf. Int. Upper	-0.571	-0.646	-0.610	-0.648	-0.655	-0.657	0.035
Train Accuracy	65.4%	67.6%	65.5%	66.4%	66.9%	66.3%	0.9%
Val Accuracy	70.7%	62.0%	71.6%	65.7%	63.5%	74.6%	4.3%
Train AUC	66.0%	68.7%	66.2%	67.1%	67.7%	66.9%	1.1%
Val AUC	71.1%	62.6%	71.4%	66.1%	63.9%	75.5%	4.1%
Train F1	69.8%	72.3%	69.3%	70.7%	70.5%	70.6%	1.1%
Test F1	74.6%	61.5%	77.4%	70.1%	70.1%	77.9%	6.0%
Train Precision	77.4%	81.6%	76.5%	78.2%	78.0%	78.0%	2.0%
Val Precision	80.6%	68.6%	83.7%	77.1%	79.4%	85.0%	5.7%
Train Recall	63.5%	64.9%	63.4%	64.5%	64.3%	64.5%	0.7%
Val Recall	69.4%	55.8%	72.0%	64.3%	62.8%	71.8%	6.3%
Train MCC	31.0%	35.7%	31.5%	33.0%	34.4%	32.7%	2.0%
Val MCC	41.0%	25.3%	40.6%	31.3%	26.0%	49.4%	7.7%
Train Log-Loss	12.48	11.69	12.42	12.10	11.93	12.16	0.33
Val Log-Loss	10.56	13.69	10.23	12.37	13.16	9.17	1.54

Univariate Report

Worst Smoothness - Kernel Density Plot

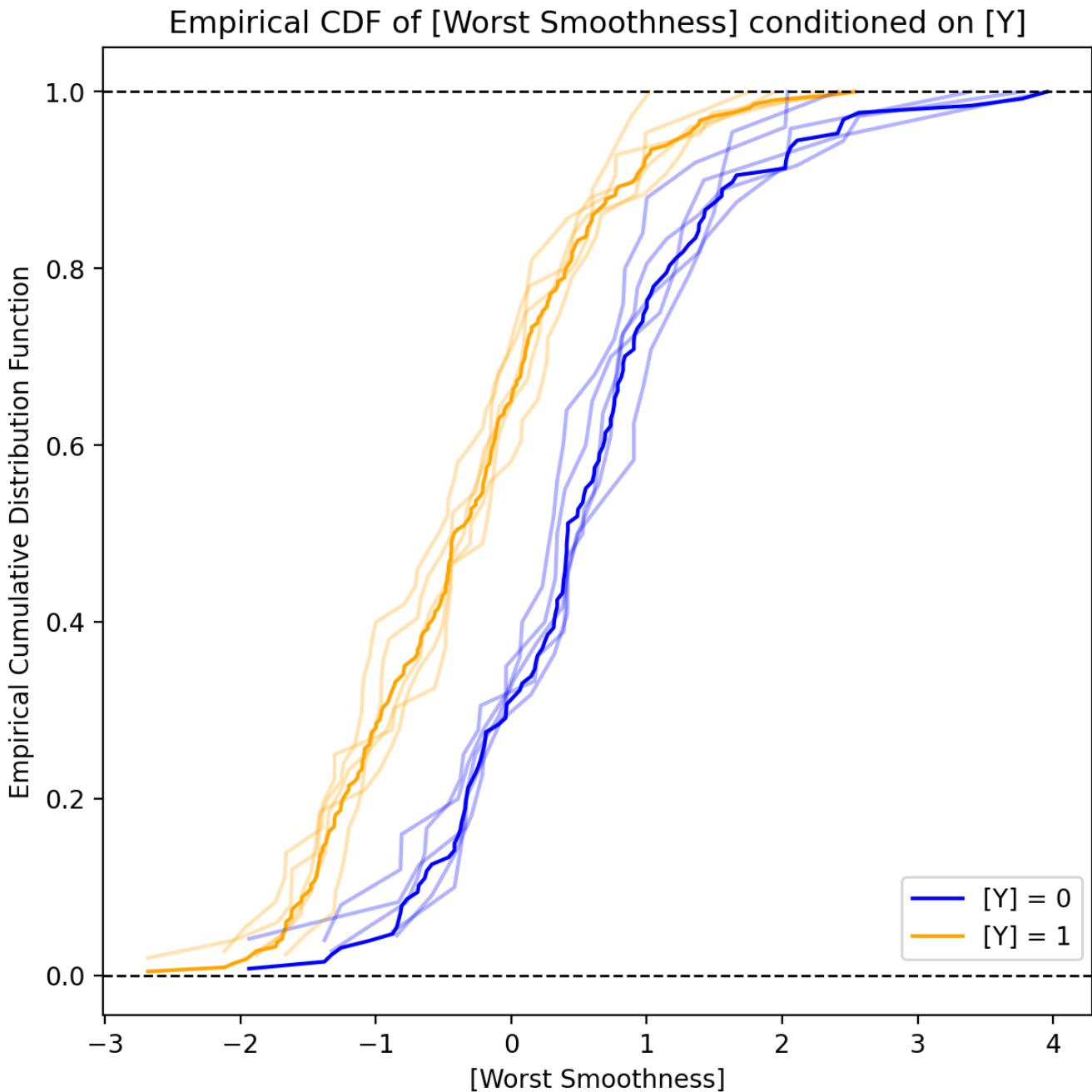
Kernel Density Plot of [Worst Smoothness] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

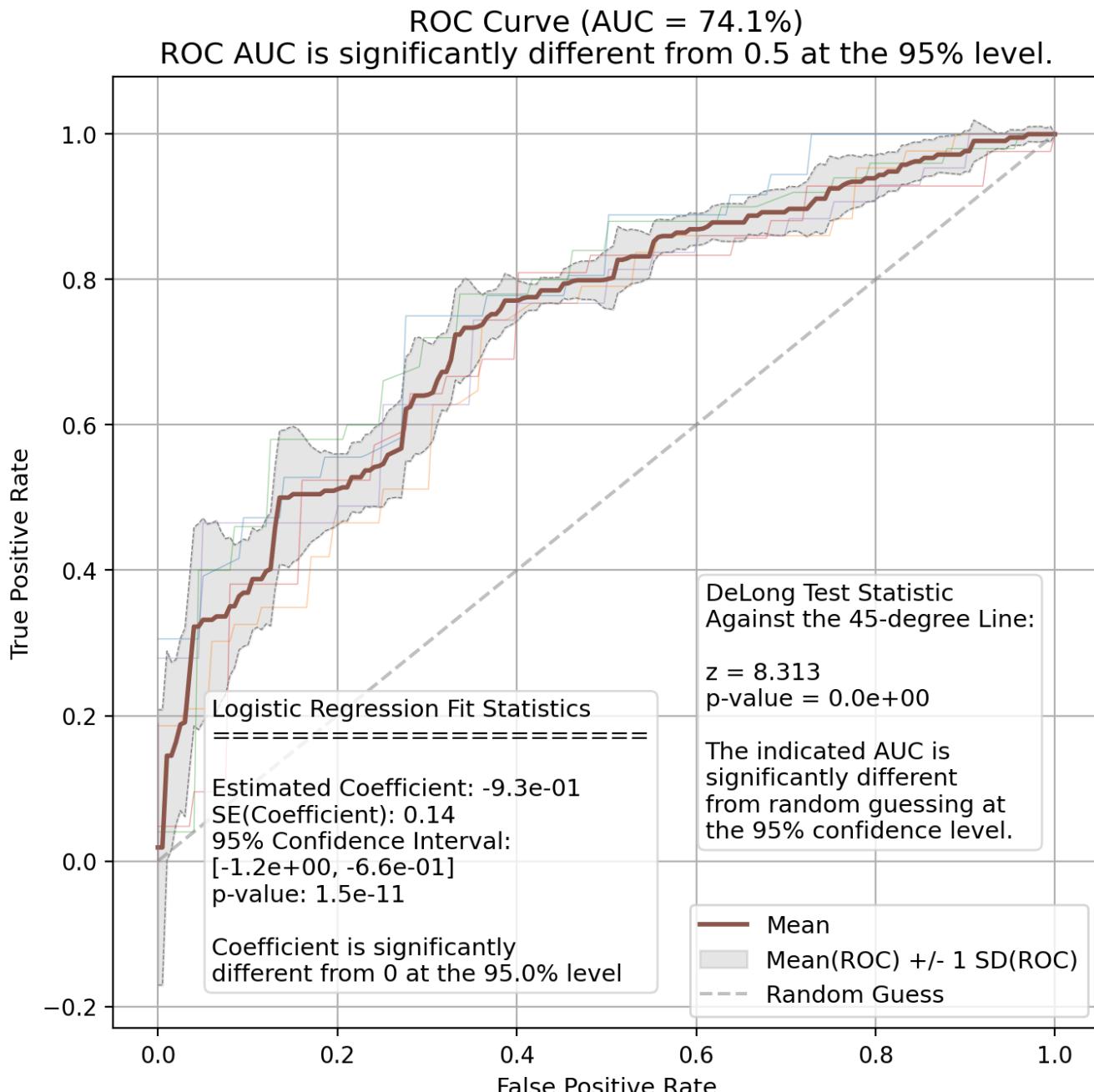
Worst Smoothness - Empirical CDF Plot



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

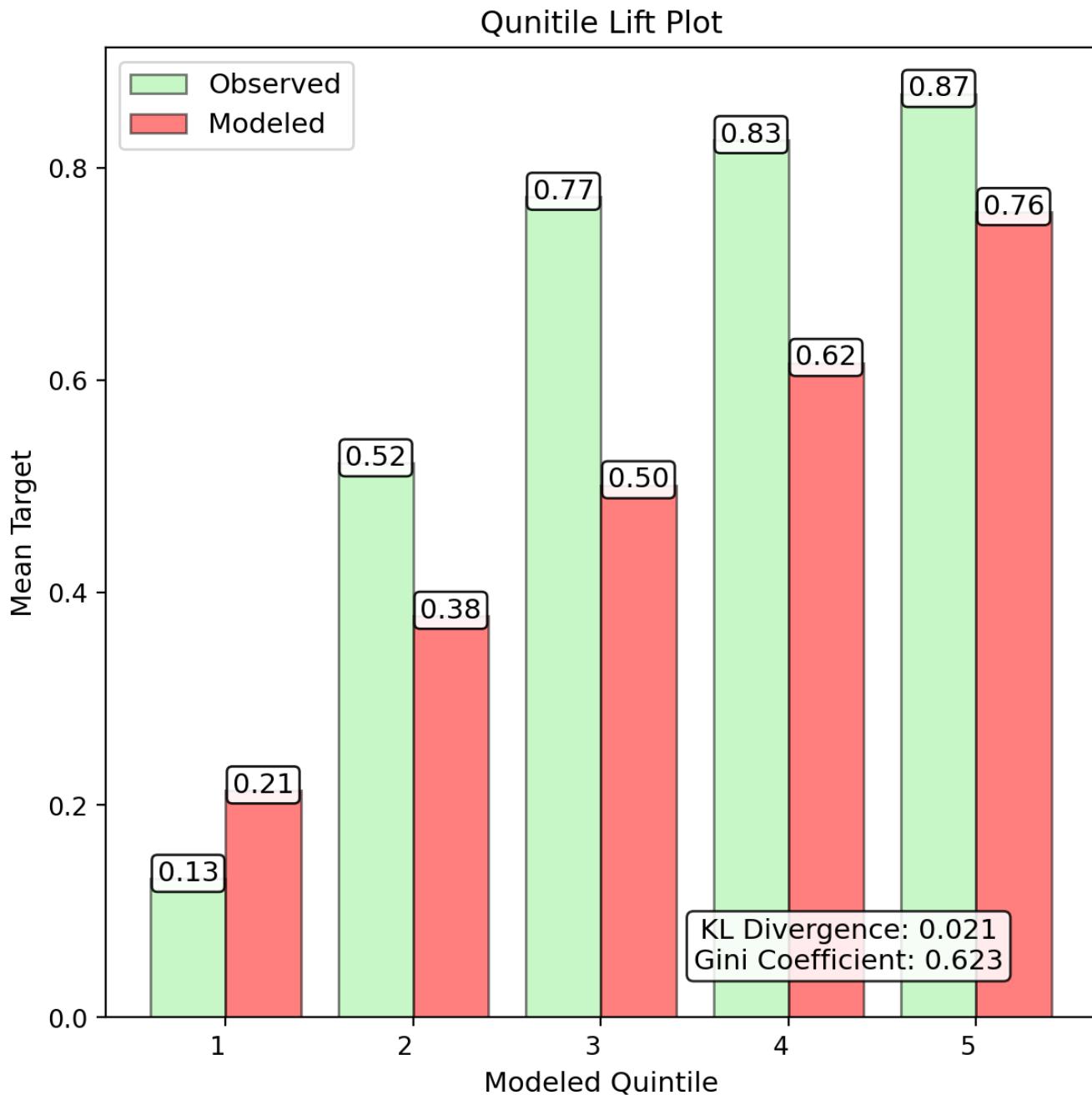


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Univariate Report

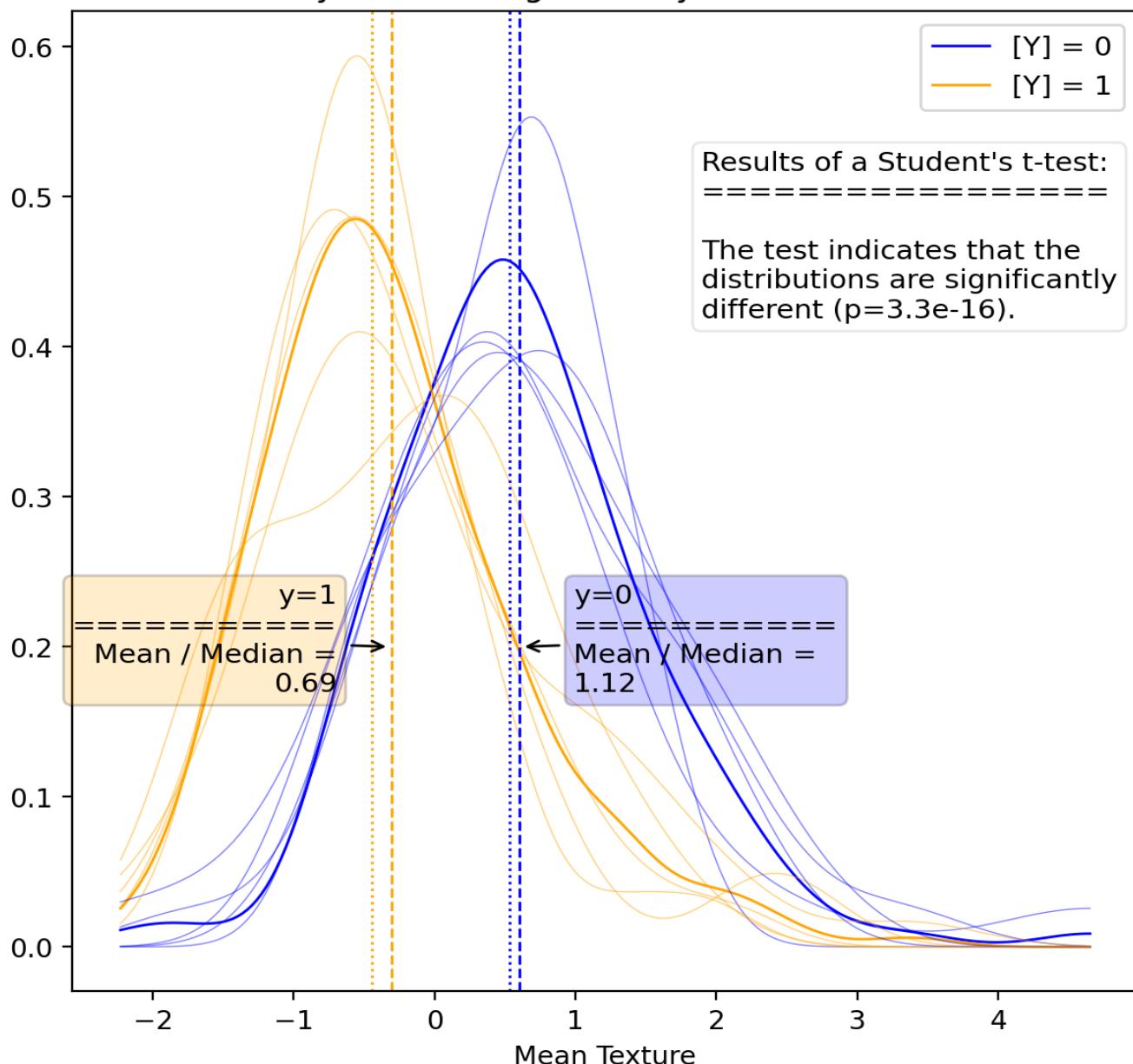
Mean Texture - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.971	-0.871	-1.271	-0.931	-1.043	-1.009	0.155
Fitted p-Value	4.1e-10	2.4e-08	3.6e-12	3.1e-09	2.1e-10	3.5e-12	1.0e-08
Fitted Std. Err.	0.155	0.156	0.183	0.157	0.164	0.145	0.012
Conf. Int. Lower	-1.28	-1.18	-1.63	-1.24	-1.37	-1.29	0.18
Conf. Int. Upper	-0.666	-0.565	-0.913	-0.623	-0.721	-0.725	0.133
Train Accuracy	70.7%	70.2%	73.4%	71.5%	73.4%	71.8%	1.5%
Val Accuracy	79.3%	77.2%	64.9%	73.1%	65.1%	71.9%	6.7%
Train AUC	71.0%	71.0%	73.7%	71.7%	73.3%	72.1%	1.3%
Val AUC	78.9%	76.8%	65.3%	73.7%	67.7%	72.0%	5.8%
Train F1	74.9%	75.0%	77.0%	75.8%	77.3%	76.0%	1.1%
Test F1	82.9%	79.5%	71.1%	76.9%	70.3%	76.1%	5.4%
Train Precision	81.0%	83.0%	82.1%	81.3%	81.3%	81.7%	0.8%
Val Precision	85.3%	77.8%	80.0%	83.3%	83.9%	81.0%	3.1%
Train Recall	69.7%	68.4%	72.6%	70.9%	73.7%	71.0%	2.1%
Val Recall	80.6%	81.4%	64.0%	71.4%	60.5%	71.8%	9.5%
Train MCC	40.8%	40.2%	46.2%	42.2%	45.6%	43.0%	2.8%
Val MCC	57.0%	53.9%	28.8%	46.0%	33.0%	42.8%	12.5%
Train Log-Loss	10.57	10.73	9.58	10.26	9.59	10.15	0.54
Val Log-Loss	7.46	8.21	12.66	9.68	12.59	10.12	2.42

Univariate Report

Mean Texture - Kernel Density Plot

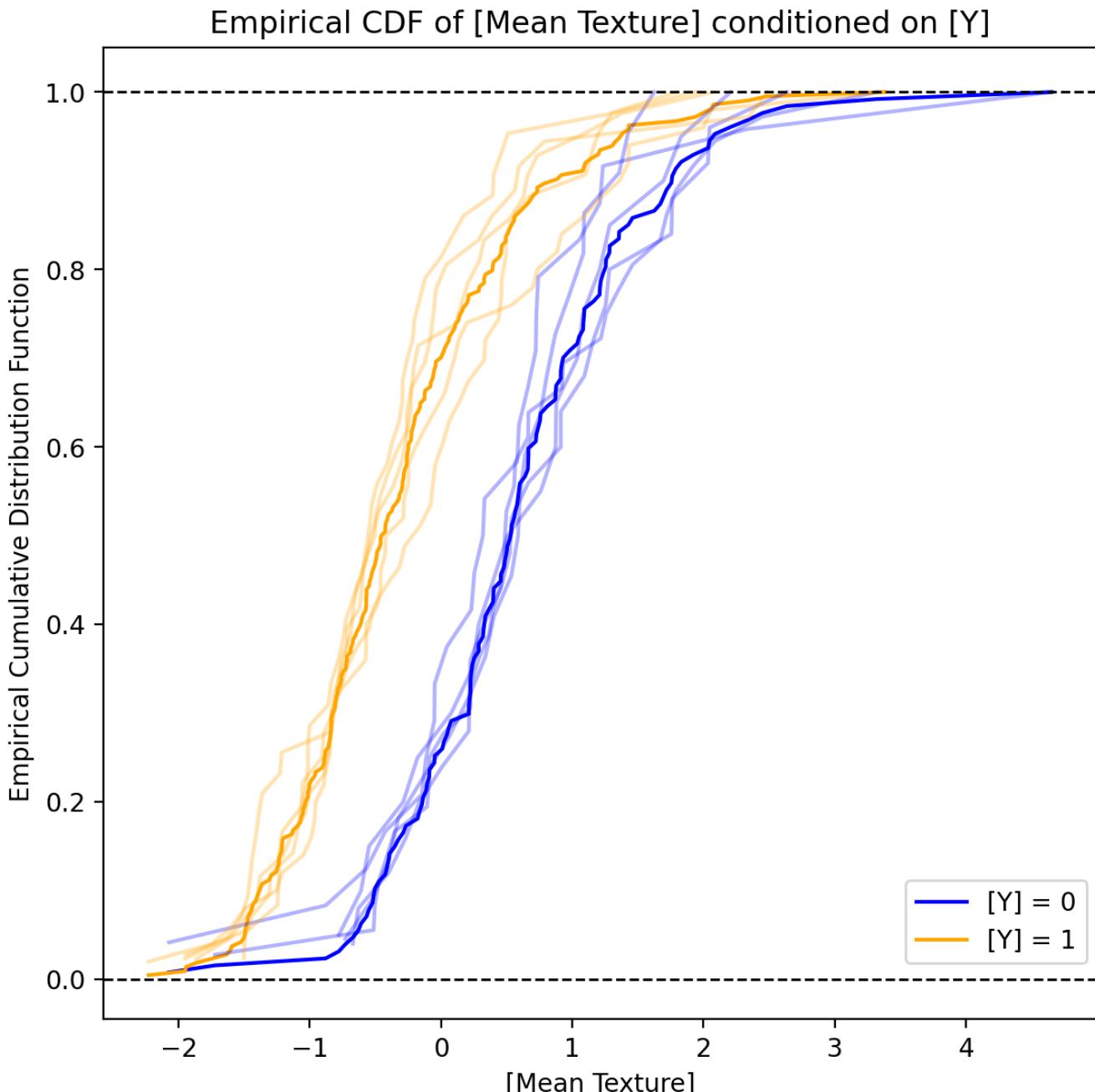
Kernel Density Plot of [Mean Texture] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

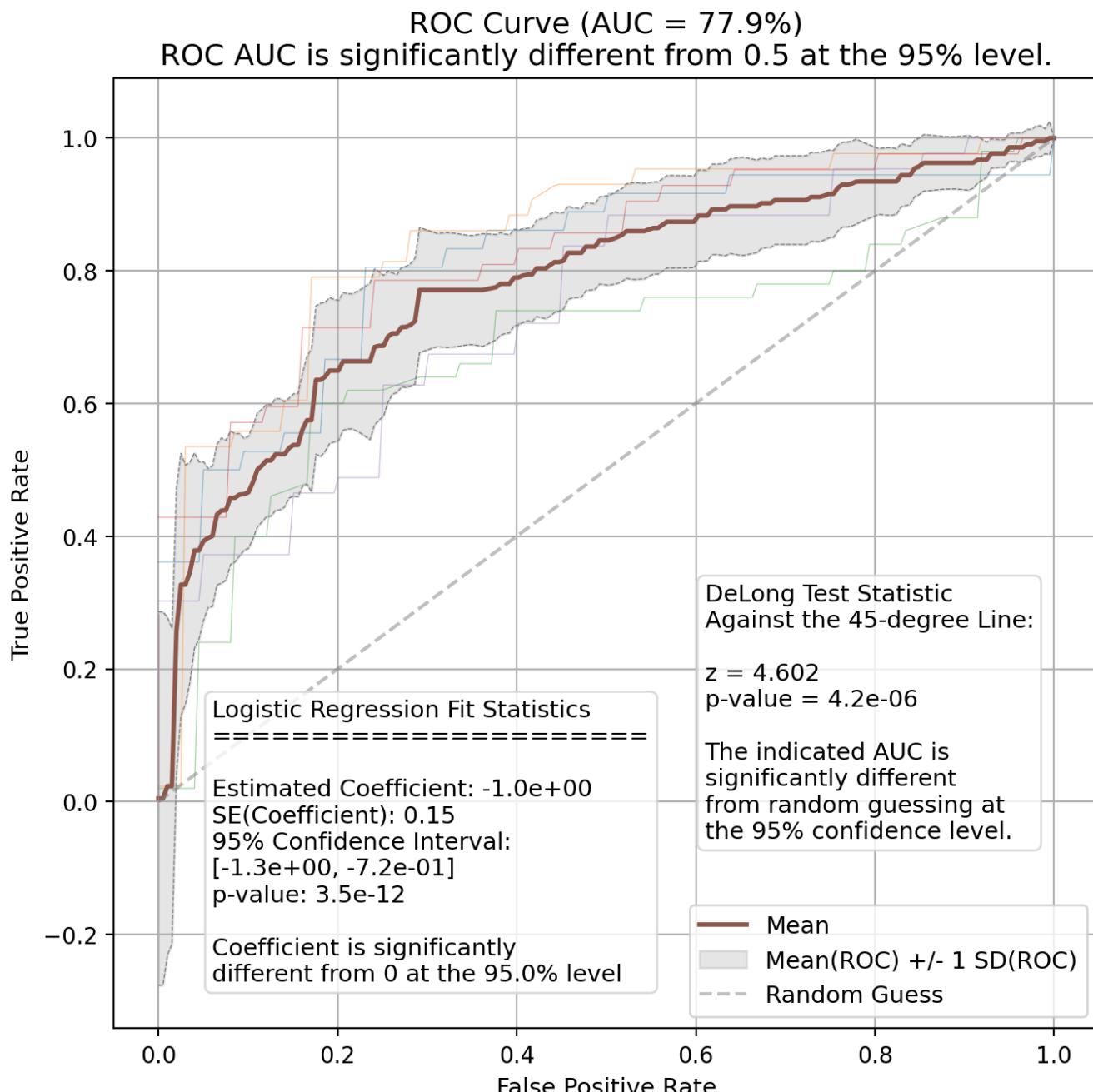
Mean Texture - Empirical CDF Plot



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

Univariate Report

Mean Texture - ROC Curve

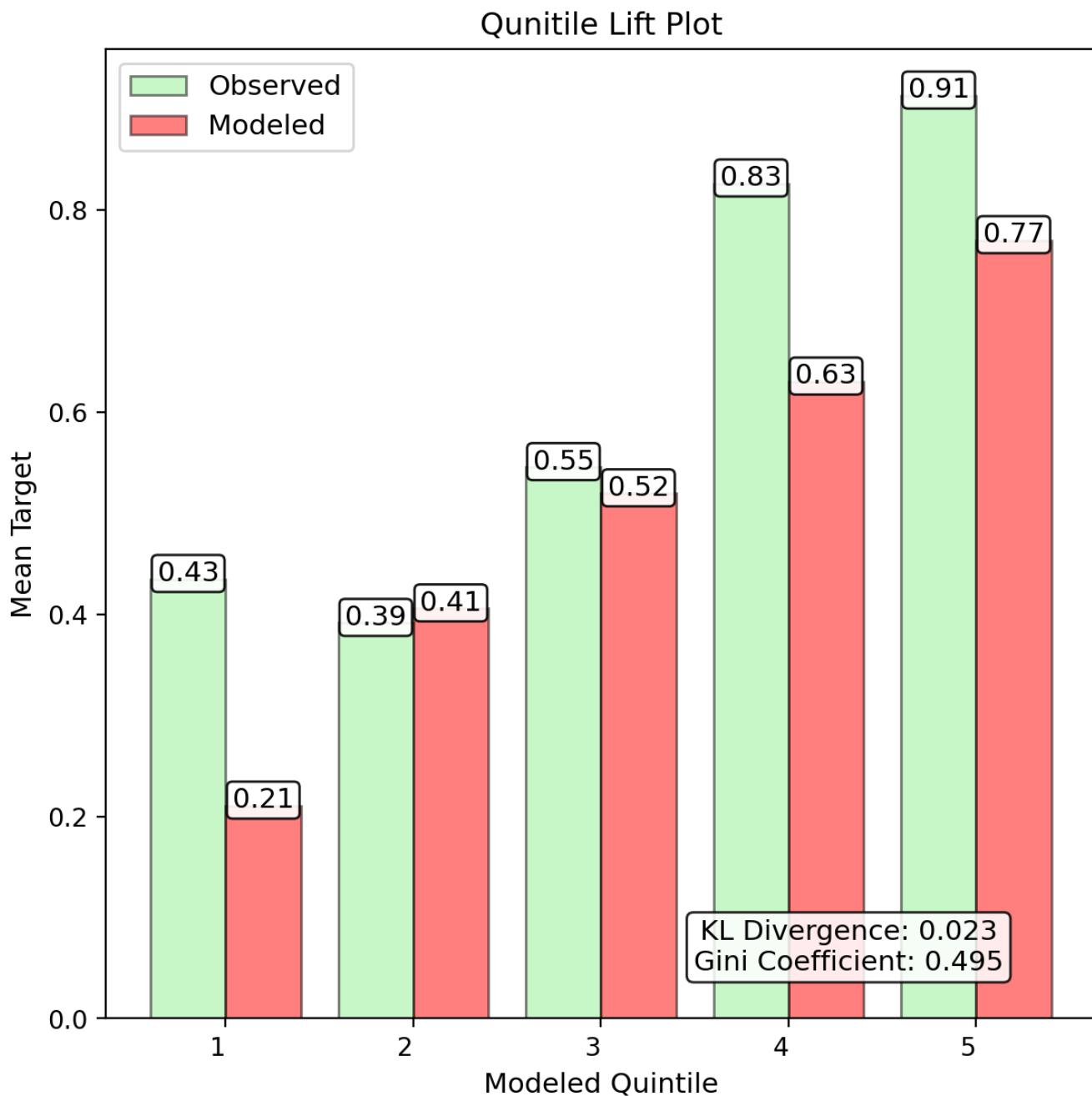


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Significance of the ROC curve is determined based on 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 Texture - Quintile Lift



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

Univariate Report

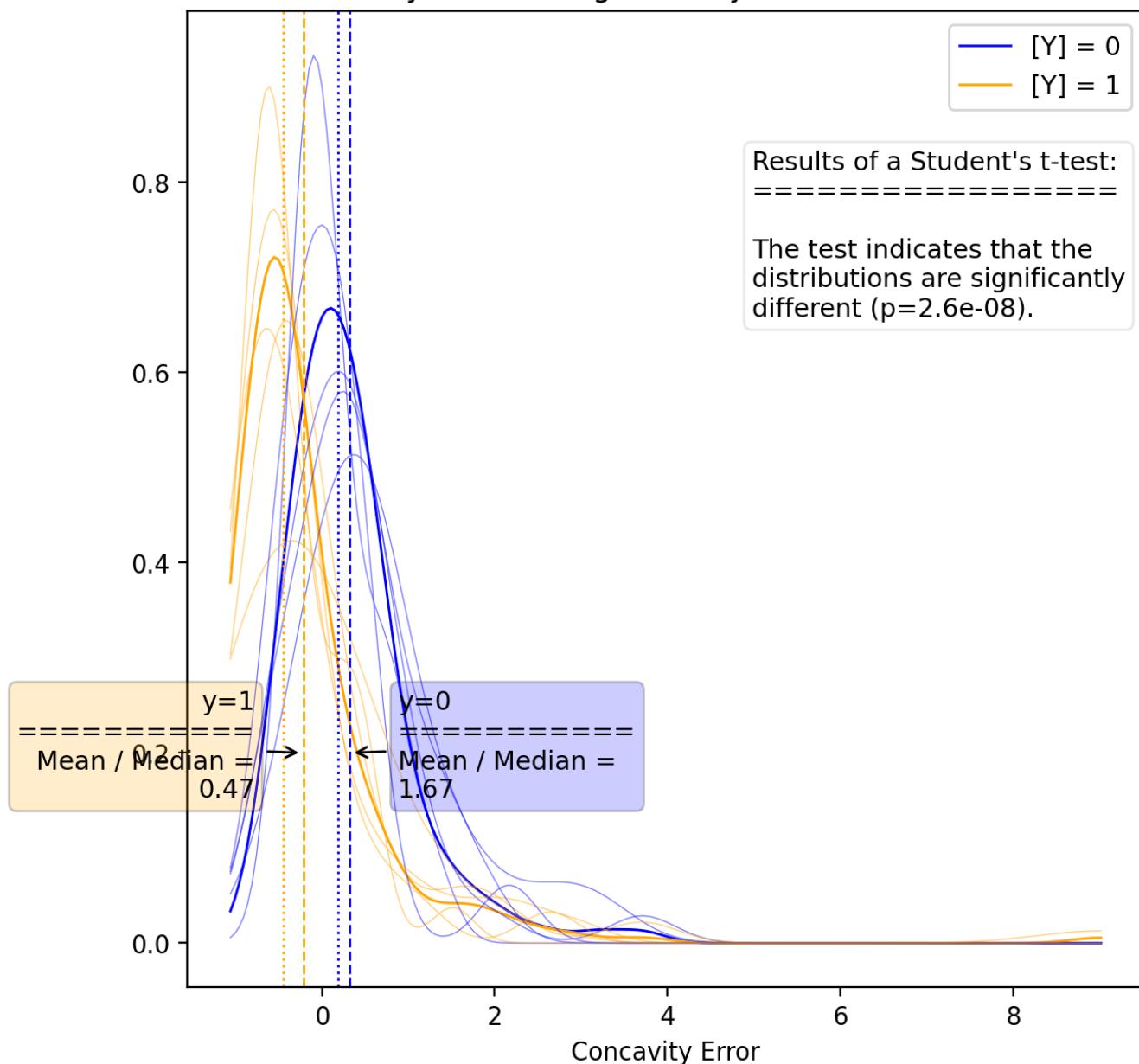
Concavity Error - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.675	-0.922	-0.877	-1.161	-0.734	-0.863	0.190
Fitted p-Value	2.3e-04	4.3e-06	1.2e-05	5.9e-08	4.3e-05	7.1e-07	9.8e-05
Fitted Std. Err.	0.183	0.201	0.201	0.214	0.179	0.174	0.014
Conf. Int. Lower	-1.03	-1.32	-1.27	-1.58	-1.09	-1.20	0.22
Conf. Int. Upper	-0.316	-0.529	-0.484	-0.742	-0.382	-0.522	0.163
Train Accuracy	70.0%	72.5%	71.9%	72.3%	73.0%	72.1%	1.2%
Val Accuracy	81.0%	72.2%	70.3%	70.1%	66.7%	70.2%	5.4%
Train AUC	68.1%	71.2%	70.5%	70.9%	71.2%	70.3%	1.3%
Val AUC	79.4%	71.9%	63.9%	71.3%	60.9%	67.8%	7.3%
Train F1	75.9%	78.2%	77.1%	77.5%	78.3%	77.8%	1.0%
Test F1	84.9%	74.4%	78.8%	73.7%	75.9%	76.4%	4.6%
Train Precision	76.6%	81.1%	77.3%	78.9%	77.6%	77.9%	1.8%
Val Precision	83.8%	74.4%	75.9%	82.4%	75.0%	75.3%	4.4%
Train Recall	75.3%	75.4%	76.8%	76.2%	78.9%	77.6%	1.5%
Val Recall	86.1%	74.4%	82.0%	66.7%	76.7%	77.5%	7.4%
Train MCC	36.0%	41.4%	40.8%	41.4%	42.7%	40.5%	2.6%
Val MCC	59.4%	43.9%	29.3%	41.3%	22.1%	36.0%	14.4%
Train Log-Loss	10.83	9.91	10.12	10.00	9.72	10.04	0.42
Val Log-Loss	6.84	10.04	10.72	10.76	12.01	10.75	1.95

Univariate Report

Concavity Error - Kernel Density Plot

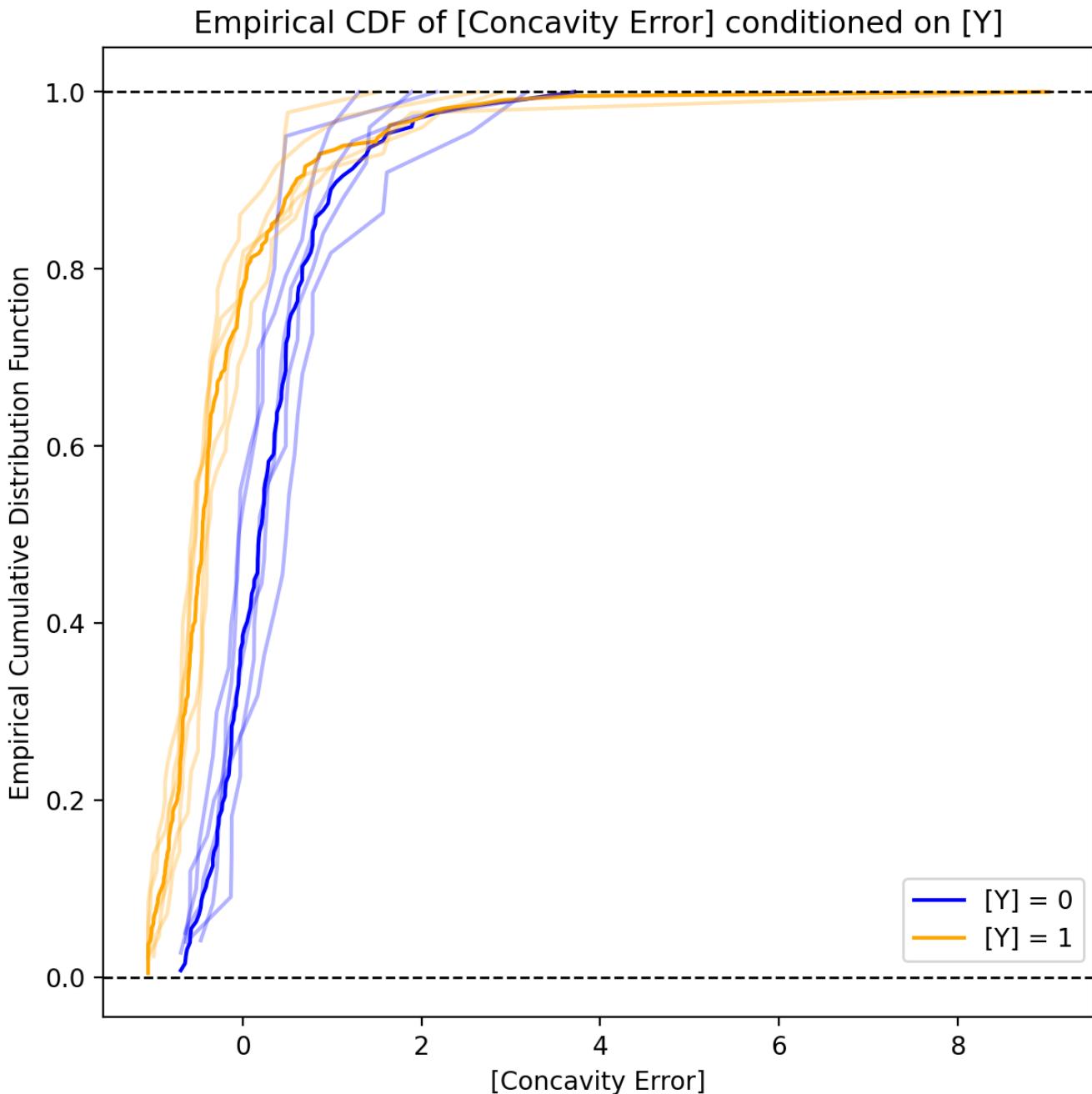
Kernel Density Plot of [Concavity Error] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

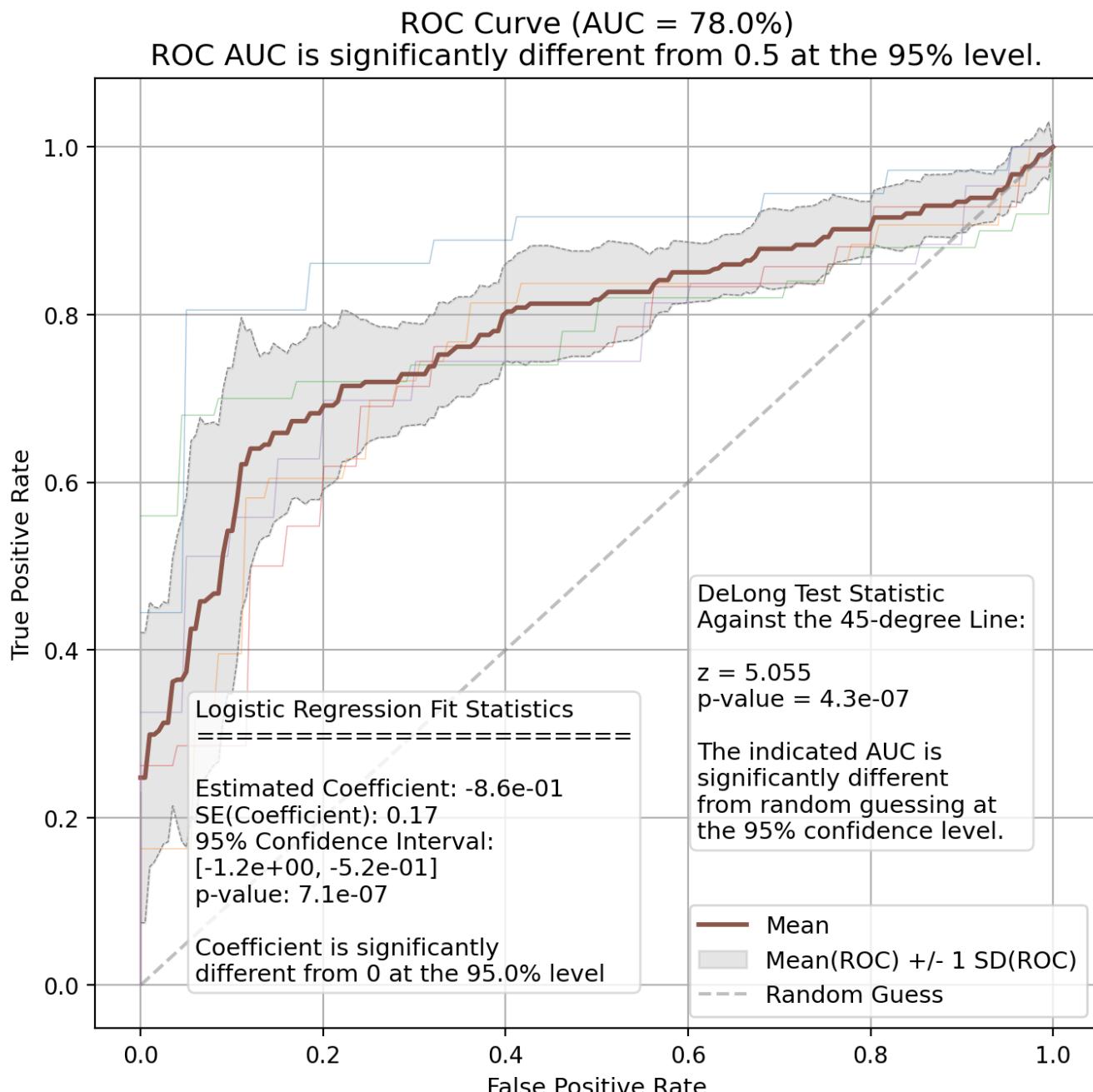
Concavity Error - Empirical CDF Plot



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

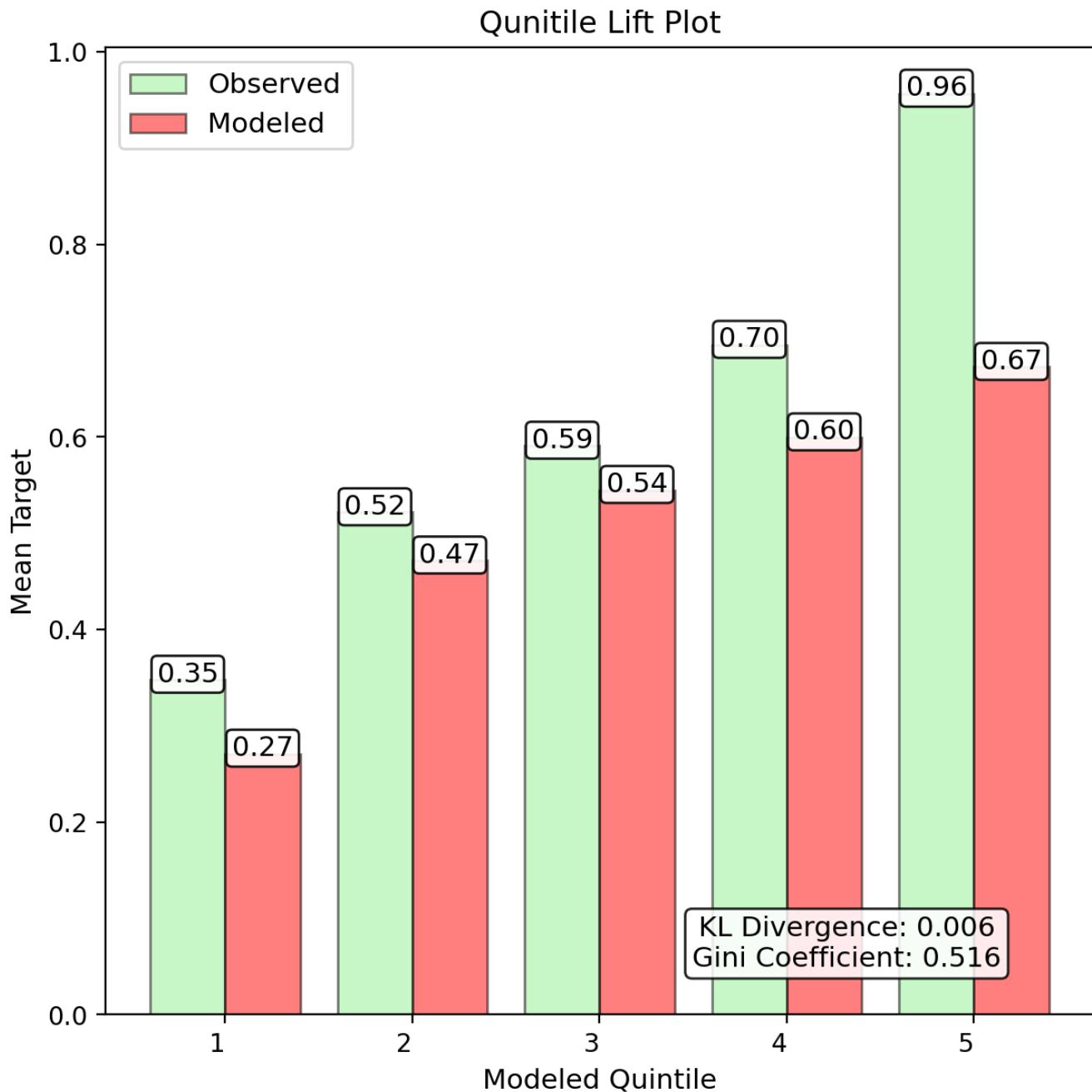


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

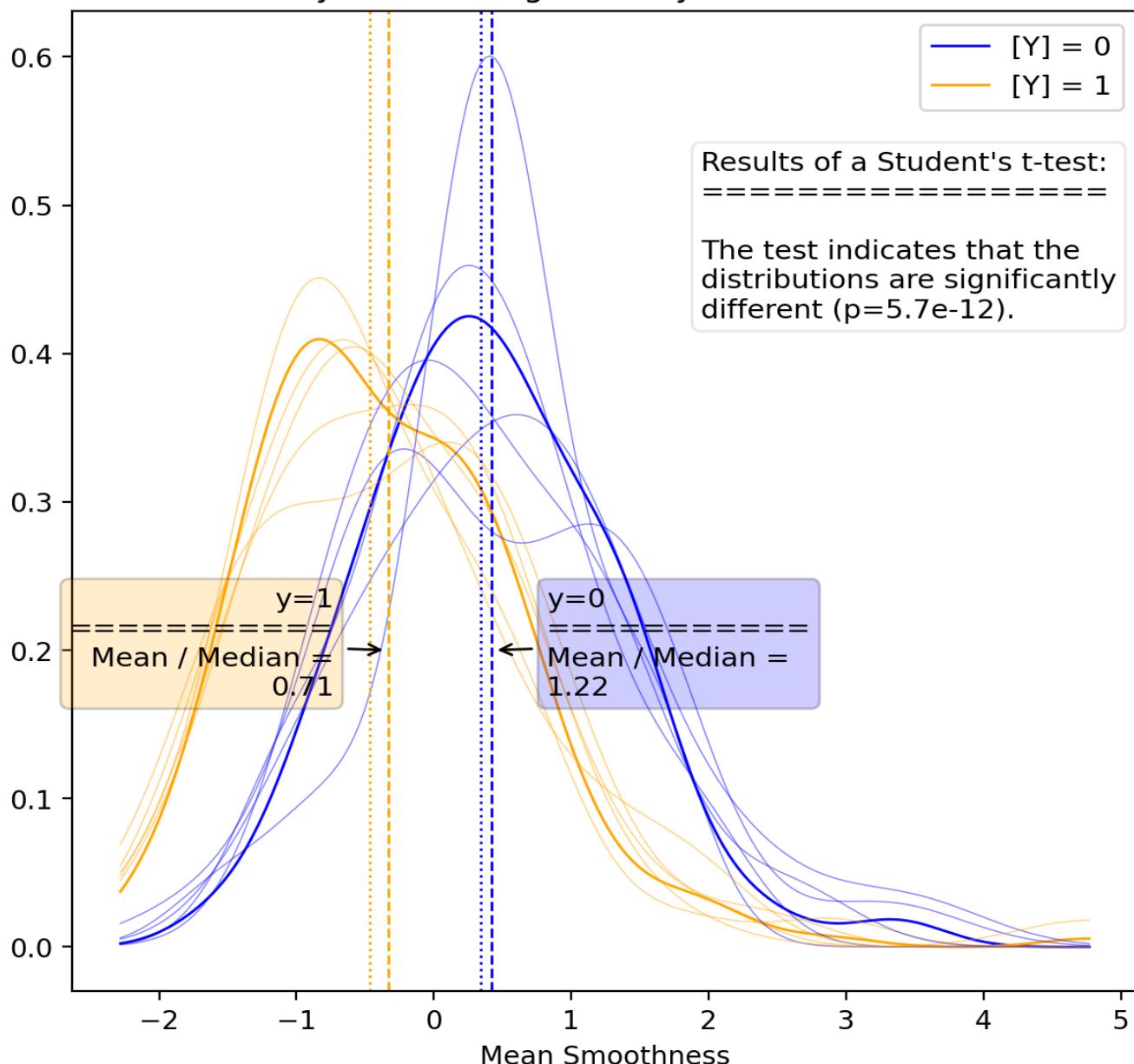
Mean Smoothness - Results

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	mean	std
Fitted Coef.	-0.839	-0.825	-0.800	-0.871	-0.842	-0.836	0.026
Fitted p-Value	1.8e-08	3.9e-08	1.3e-07	5.3e-09	1.8e-08	4.5e-10	4.9e-08
Fitted Std. Err.	0.149	0.150	0.151	0.149	0.149	0.134	0.001
Conf. Int. Lower	-1.13	-1.12	-1.10	-1.16	-1.13	-1.10	0.02
Conf. Int. Upper	-0.547	-0.531	-0.503	-0.578	-0.549	-0.573	0.028
Train Accuracy	64.7%	64.9%	66.3%	64.6%	65.8%	64.8%	0.8%
Val Accuracy	65.5%	64.6%	66.2%	65.7%	66.7%	67.5%	0.8%
Train AUC	65.3%	66.2%	67.3%	64.8%	66.6%	65.7%	1.0%
Val AUC	67.8%	64.7%	63.1%	69.4%	66.2%	68.9%	2.5%
Train F1	69.1%	69.7%	69.6%	69.4%	69.5%	68.9%	0.2%
Test F1	67.7%	65.9%	74.2%	66.7%	73.4%	70.9%	3.9%
Train Precision	76.7%	79.7%	78.0%	75.9%	77.1%	77.3%	1.5%
Val Precision	80.8%	69.2%	76.6%	85.2%	80.6%	80.4%	6.0%
Train Recall	62.9%	62.0%	62.8%	64.0%	63.2%	62.1%	0.7%
Val Recall	58.3%	62.8%	72.0%	54.8%	67.4%	63.4%	6.9%
Train MCC	29.5%	30.8%	33.7%	28.7%	32.4%	30.4%	2.1%
Val MCC	34.7%	29.3%	25.4%	38.2%	30.5%	36.6%	4.9%
Train Log-Loss	12.74	12.66	12.15	12.76	12.32	12.68	0.27
Val Log-Loss	12.43	12.77	12.18	12.37	12.01	11.70	0.29

Univariate Report

Mean Smoothness - Kernel Density Plot

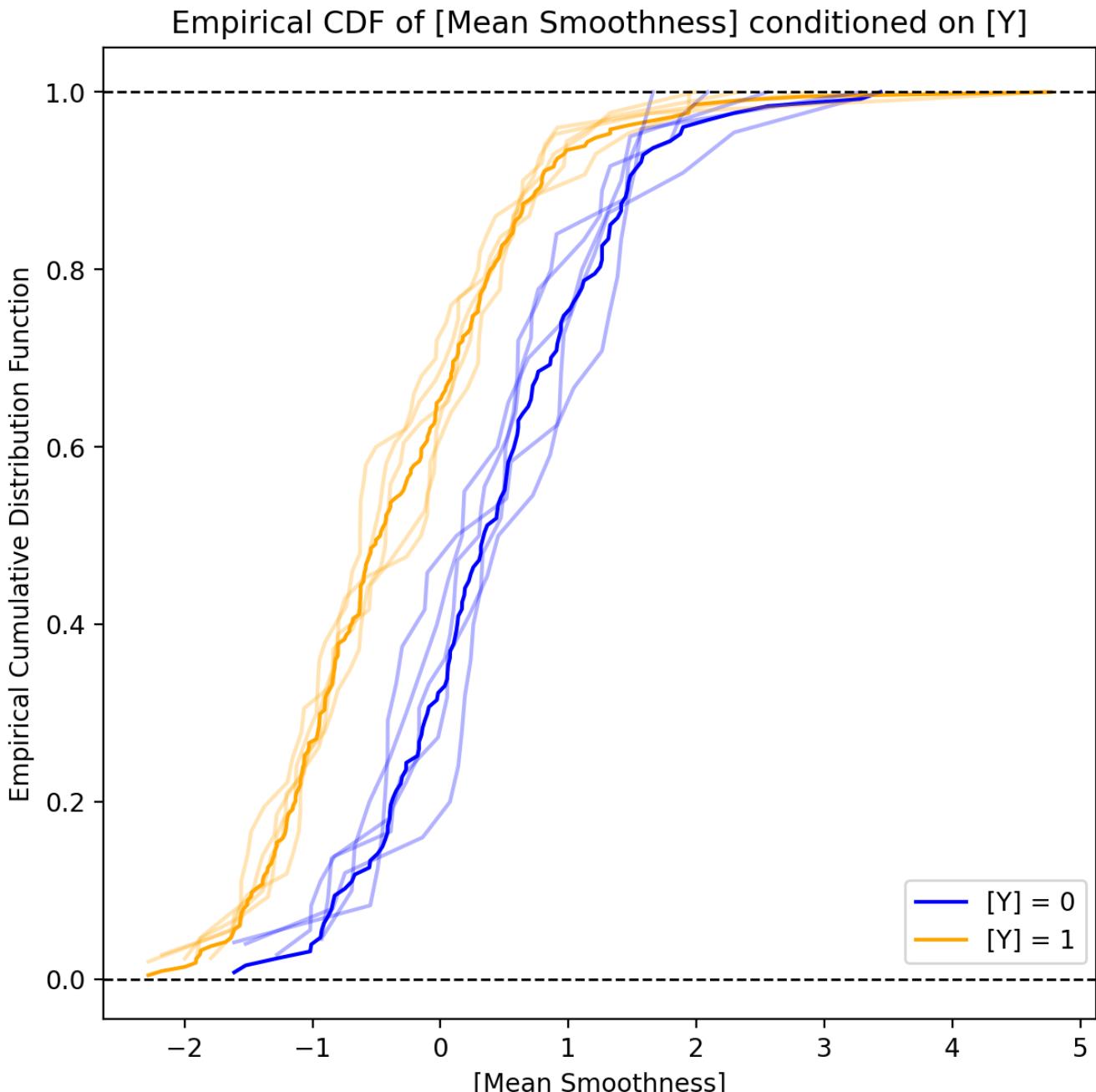
Kernel Density Plot of [Mean Smoothness] by [Y].
Distributions by level are significantly different at the 95% level.



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

Univariate Report

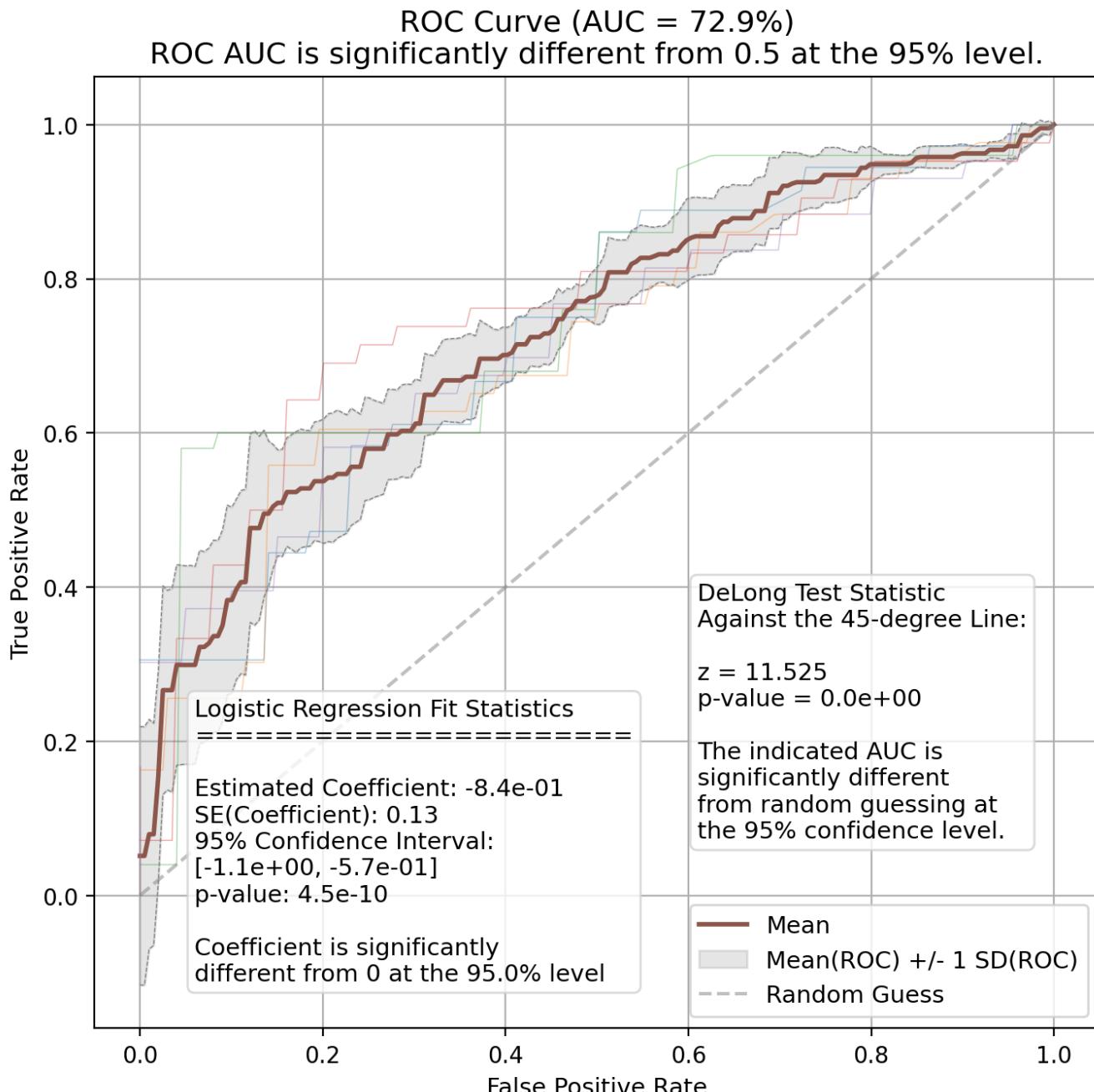
Mean Smoothness - Empirical CDF Plot



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

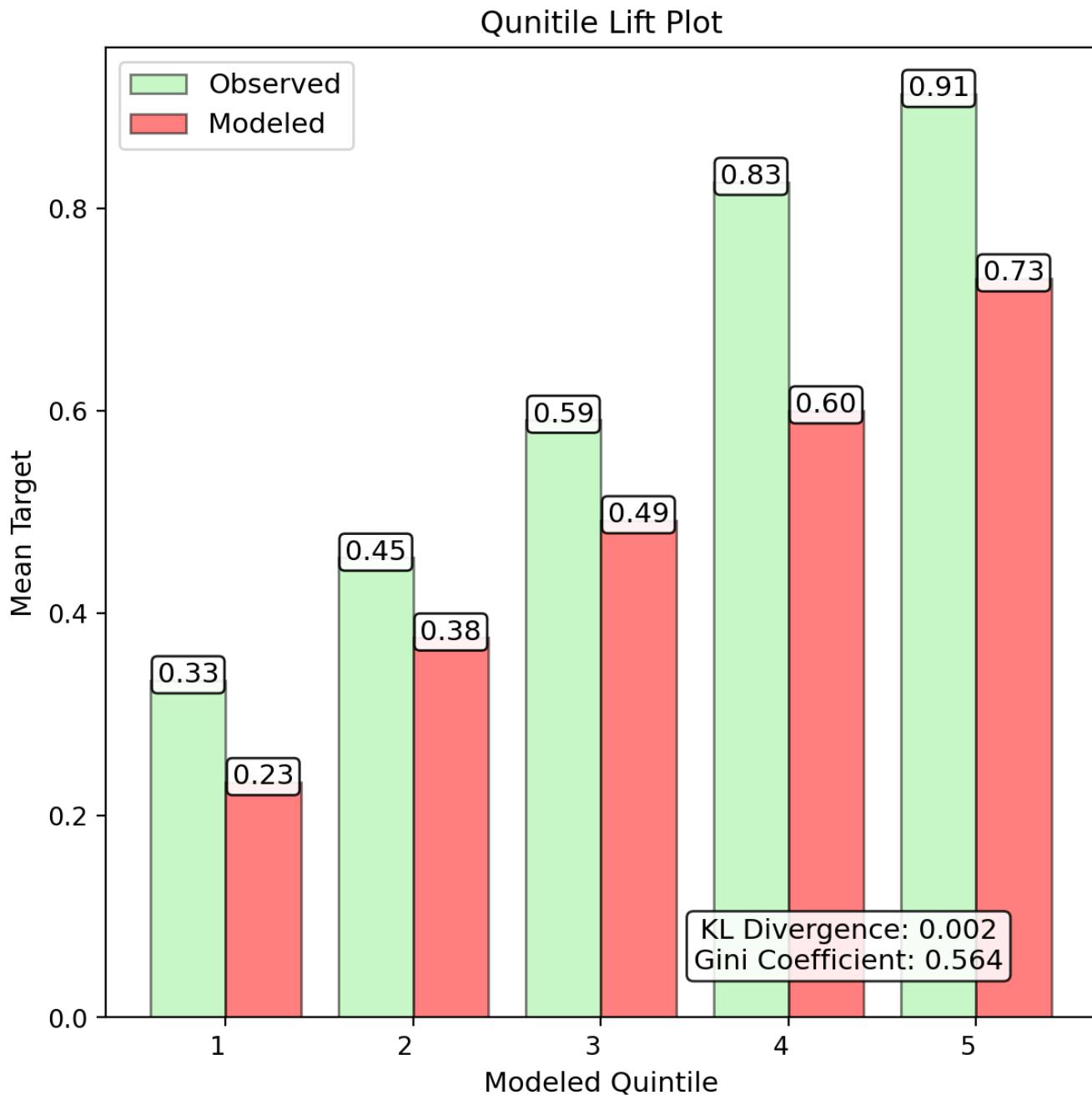


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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