

Cancer Model Univariate Analysis Report

2024-02-22

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.
Log1p Mean_radius	75.0%	84.8%	68.4%	76.1%	75.7%	51.8%	72.0%
Log1p Radius_error	71.8%	86.0%	64.2%	74.0%	73.5%	46.9%	69.4%
Mean Texture	71.9%	81.0%	71.8%	72.0%	76.1%	42.8%	69.3%
Log1p Mean_compactness	70.4%	86.5%	57.1%	72.6%	68.8%	46.2%	66.9%
Mean Smoothness	67.5%	80.4%	63.4%	68.9%	70.9%	36.6%	64.6%
Log1p Mean_texture	65.3%	75.6%	56.4%	66.6%	64.6%	33.3%	60.3%
Log1p Mean_smoothness	62.2%	74.4%	51.8%	64.0%	61.1%	28.3%	57.0%
Mean Symmetry	55.3%	70.0%	49.3%	57.2%	57.9%	14.1%	50.6%
Texture Error	55.3%	70.0%	49.3%	57.2%	57.9%	14.1%	50.6%
Log1p Mean_fractal_dimension	52.9%	70.5%	46.3%	55.6%	55.9%	10.8%	48.6%
Log1p Mean_symmetry	47.5%	63.0%	28.3%	52.0%	39.1%	4.4%	39.0%
Mean Fractal_dimension	44.7%	58.0%	40.8%	46.0%	47.9%	-7.8%	38.3%
Log1p Texture_error	43.6%	53.7%	36.7%	45.2%	43.6%	-9.7%	35.5%

This table shows an overview of the results for the variables in this file, representing those whose average test score are ranked between 16 and 28 of the variables passed to the Cancer Model.

Univariate Report

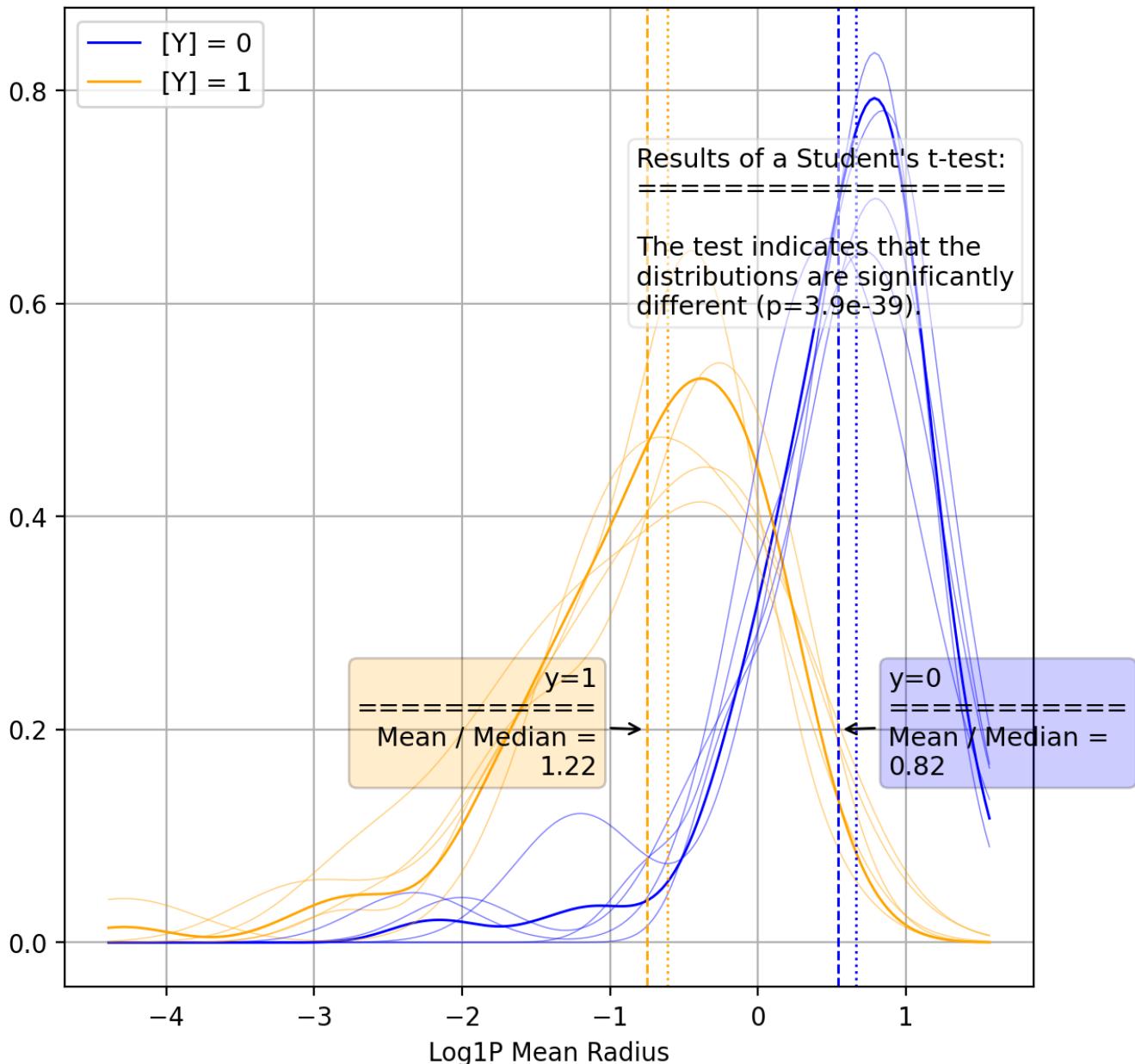
Log1P Mean Radius - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-3.21	2.3e-14	0.420	-4.03	-2.38	75.4%	76.9%	77.1%	87.1%	69.2%	52.8%
Fold-2	-2.56	1.2e-13	0.345	-3.24	-1.89	86.1%	86.8%	86.1%	94.4%	79.1%	73.5%
Fold-3	-3.07	4.4e-16	0.377	-3.80	-2.33	78.0%	80.4%	77.6%	95.0%	65.5%	61.2%
Fold-4	-3.21	1.8e-15	0.403	-4.00	-2.42	83.3%	83.8%	84.7%	89.3%	80.6%	66.9%
Fold-5	-2.92	1.1e-15	0.364	-3.63	-2.21	78.4%	81.0%	78.4%	95.2%	66.7%	61.9%
mean	-2.98	1.7e-18	0.340	-3.65	-2.32	75.0%	76.1%	75.7%	84.8%	68.4%	51.8%
std	0.27	5.0e-14	0.030	0.32	0.22	4.3%	3.7%	4.3%	3.8%	7.1%	7.6%

Univariate Report

Log1P Mean Radius - Kernel Density Plot

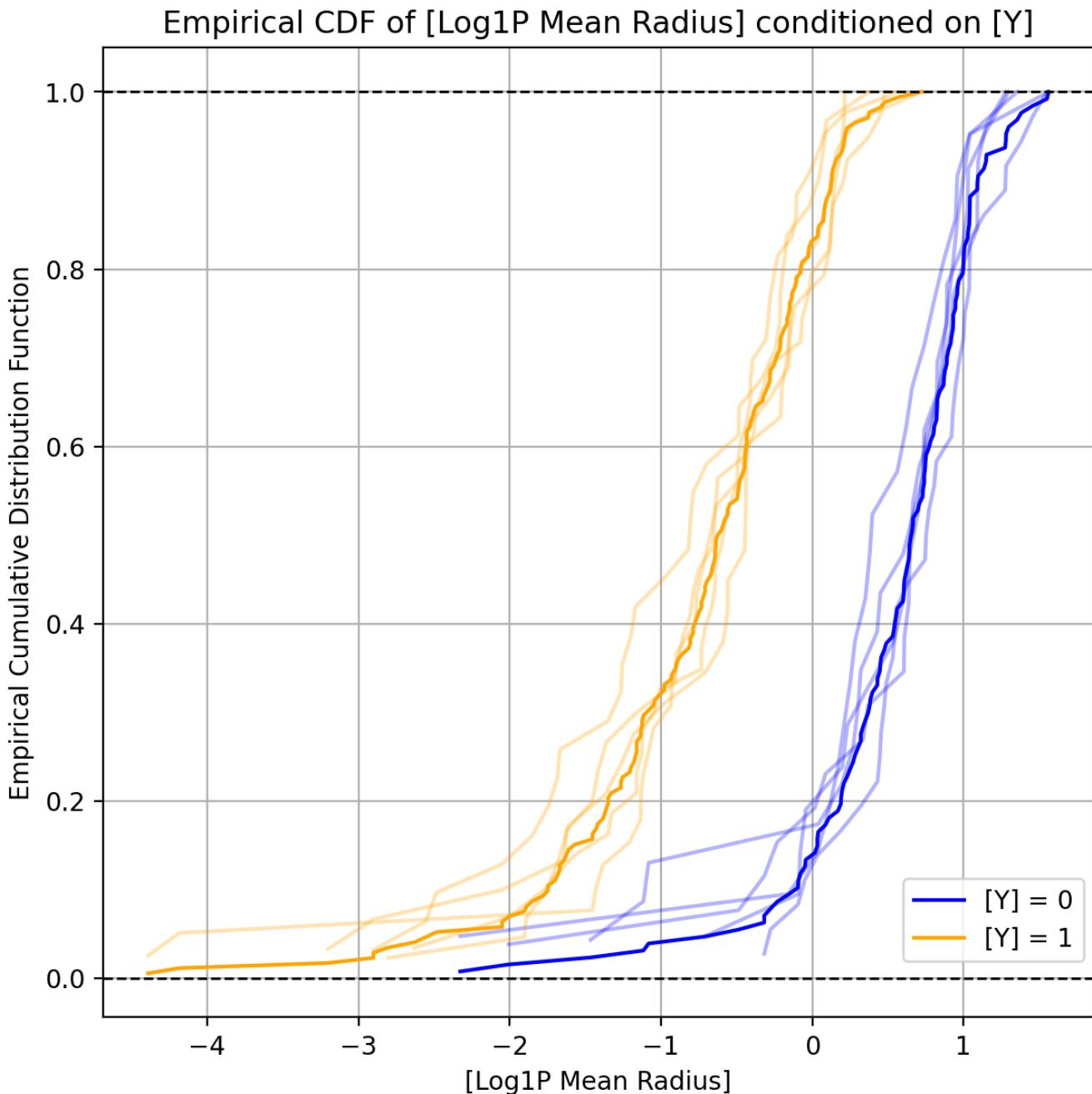
Kernel Density Plot of [Log1P Mean Radius] by [Y].
Distributions by level are significantly different at the 95% level.



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

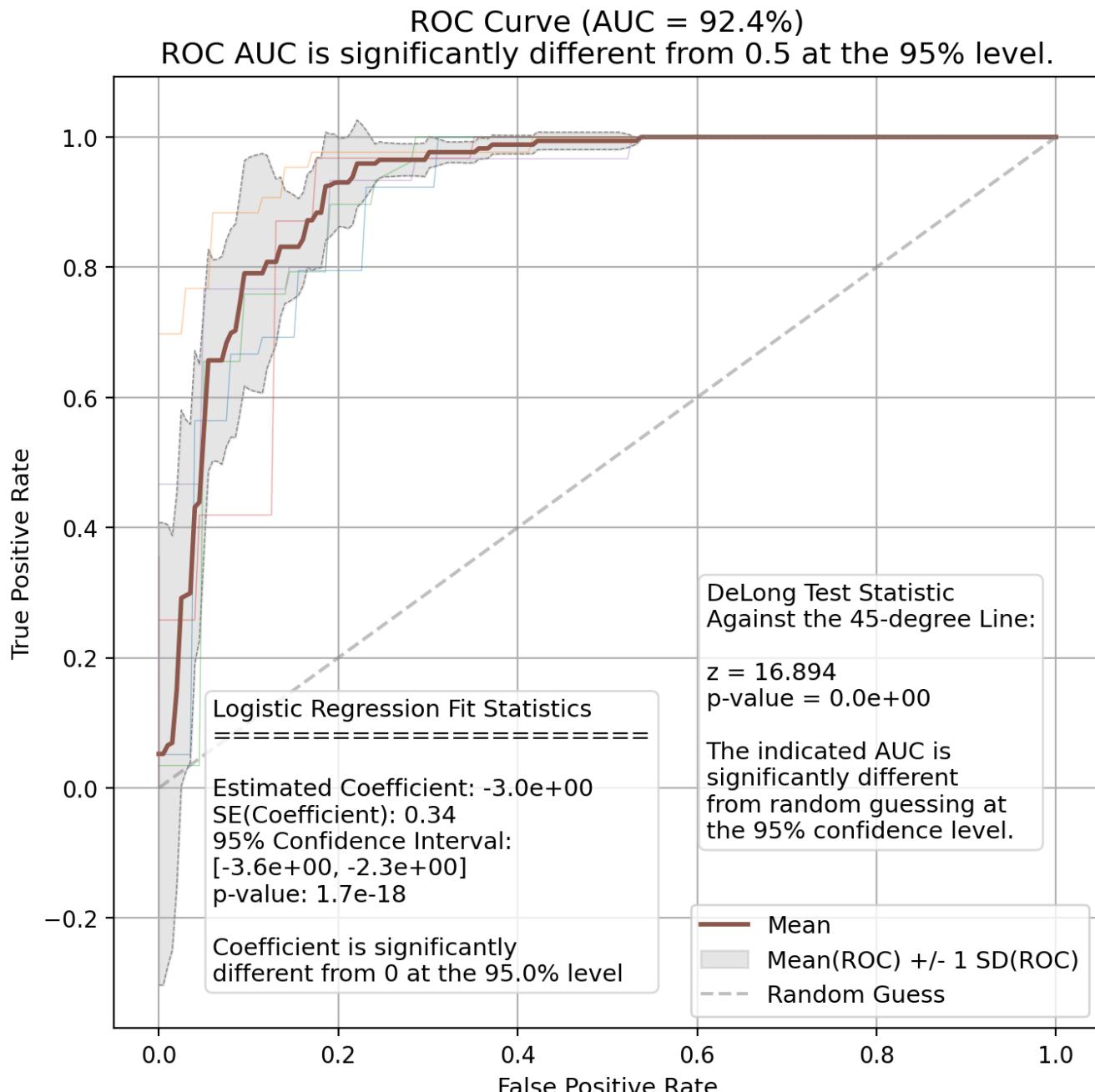
Log1P Mean Radius - Empirical CDF Plot



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

Log1P Mean Radius - ROC Curve

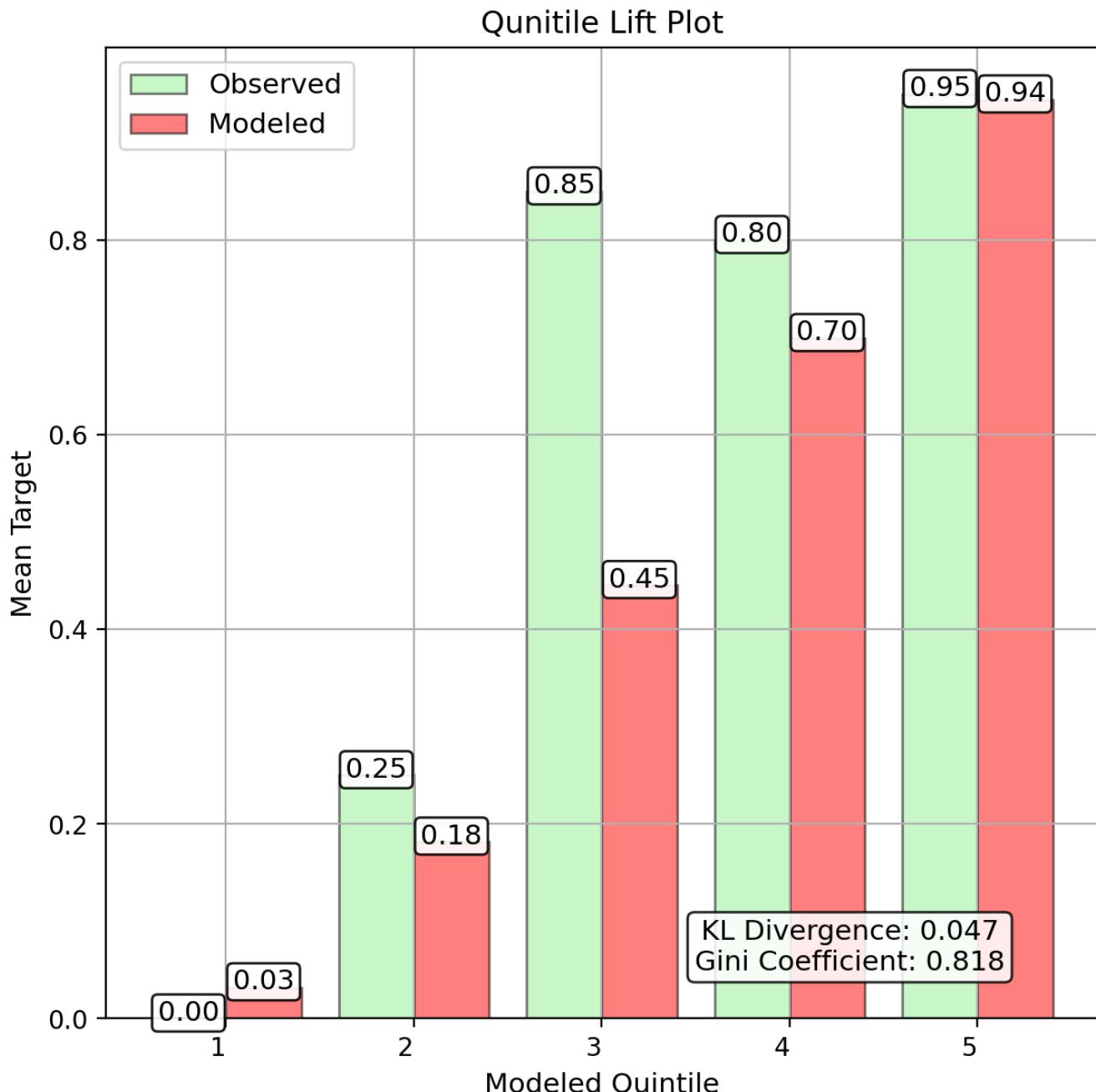


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Log1P 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

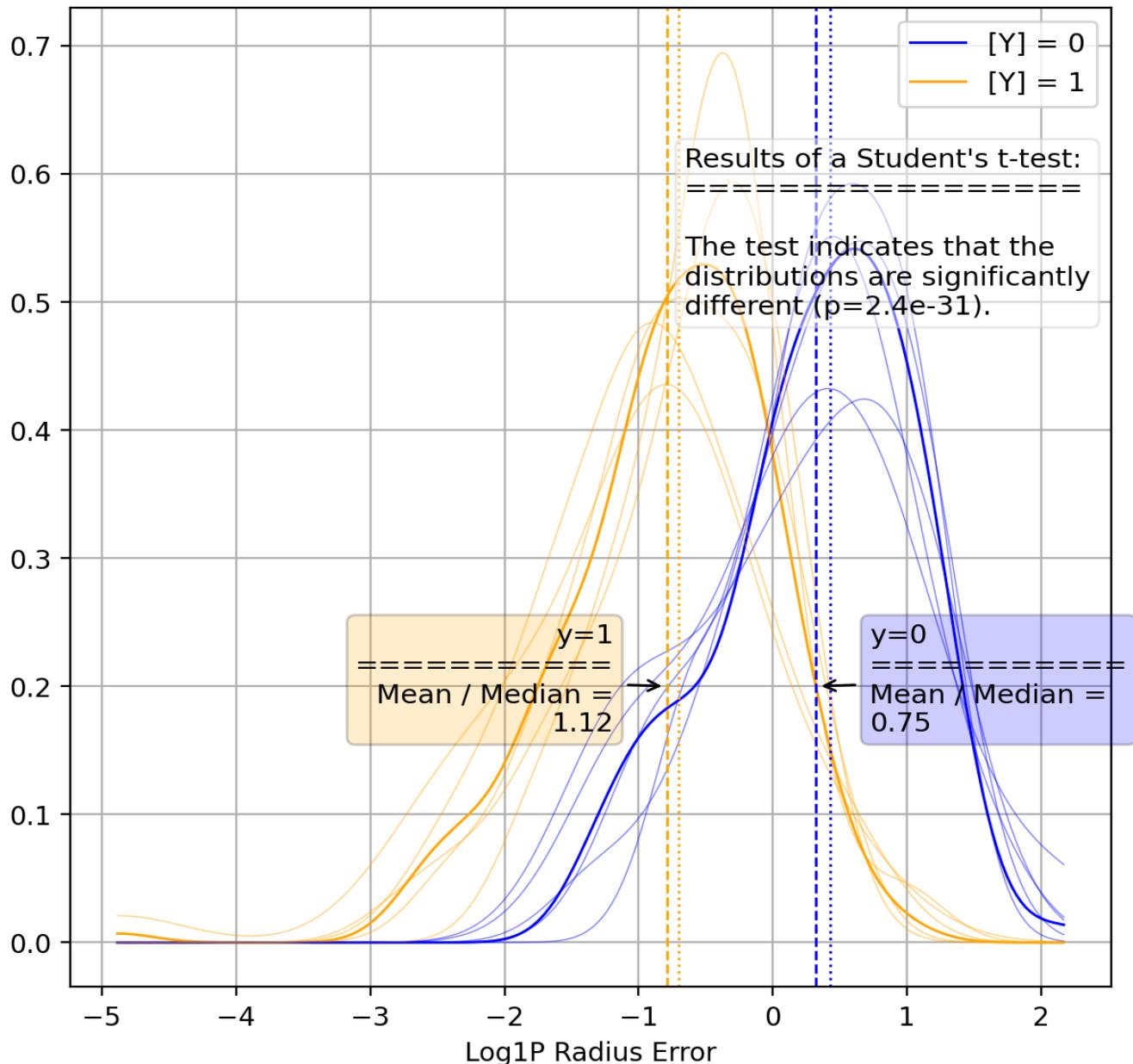
Log1P Radius Error - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-2.37	3.4e-16	0.291	-2.94	-1.80	66.2%	67.6%	70.0%	80.0%	62.2%	34.0%
Fold-2	-1.97	2.9e-15	0.250	-2.46	-1.48	66.7%	69.5%	65.1%	84.4%	52.9%	39.9%
Fold-3	-2.27	2.8e-17	0.269	-2.80	-1.75	63.6%	66.9%	64.3%	81.8%	52.9%	33.6%
Fold-4	-2.12	3.1e-16	0.259	-2.63	-1.61	83.6%	83.4%	86.5%	88.9%	84.2%	65.8%
Fold-5	-2.15	3.2e-16	0.264	-2.67	-1.64	75.8%	77.1%	80.0%	88.2%	73.2%	51.5%
mean	-2.17	6.3e-20	0.238	-2.64	-1.71	71.8%	74.0%	73.5%	86.0%	64.2%	46.9%
std	0.15	1.2e-15	0.015	0.18	0.12	8.3%	7.1%	9.7%	3.9%	13.5%	13.7%

Univariate Report

Log1P Radius Error - Kernel Density Plot

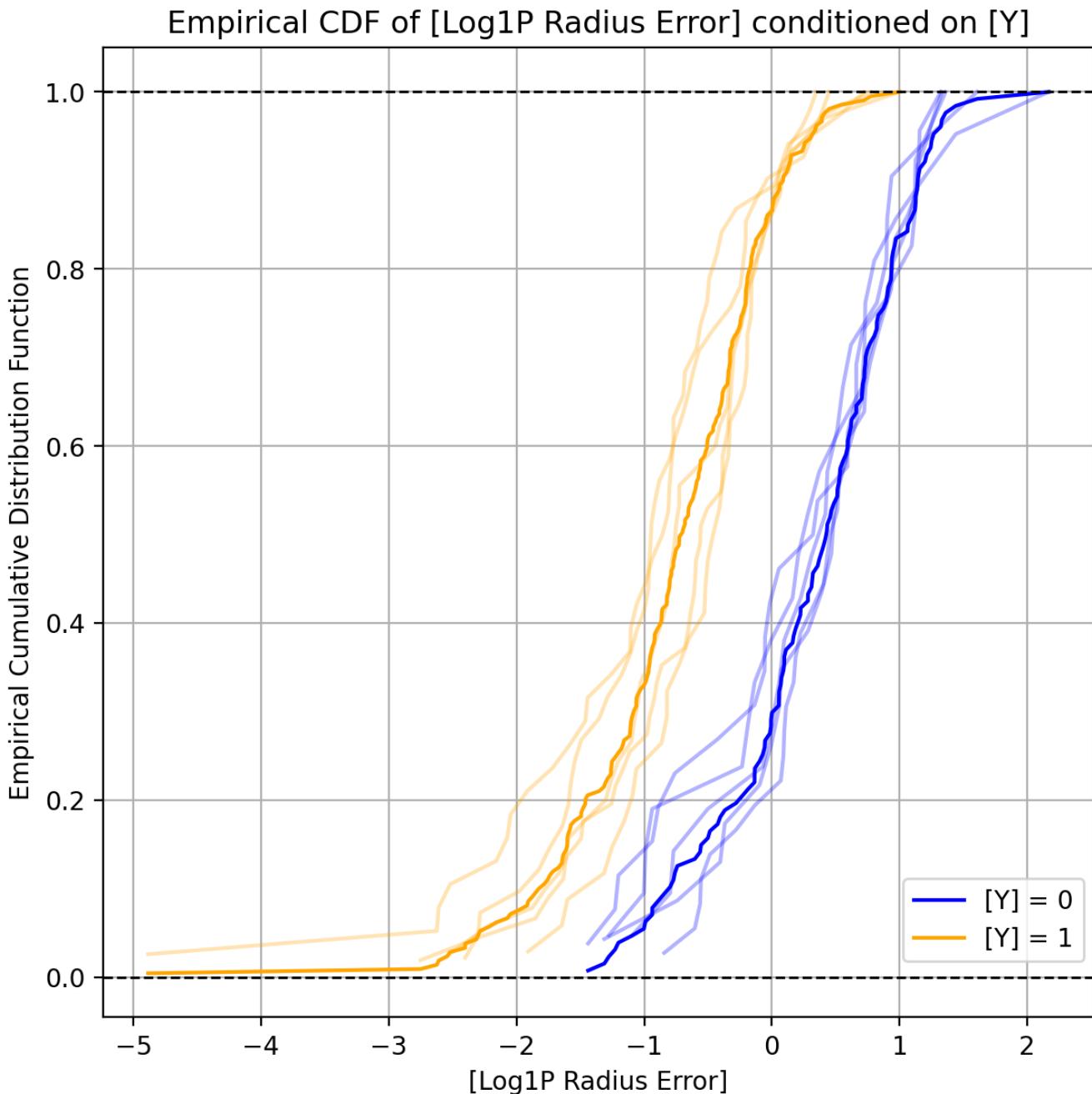
Kernel Density Plot of [Log1P Radius Error] by [Y].
Distributions by level are significantly different at the 95% level.



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

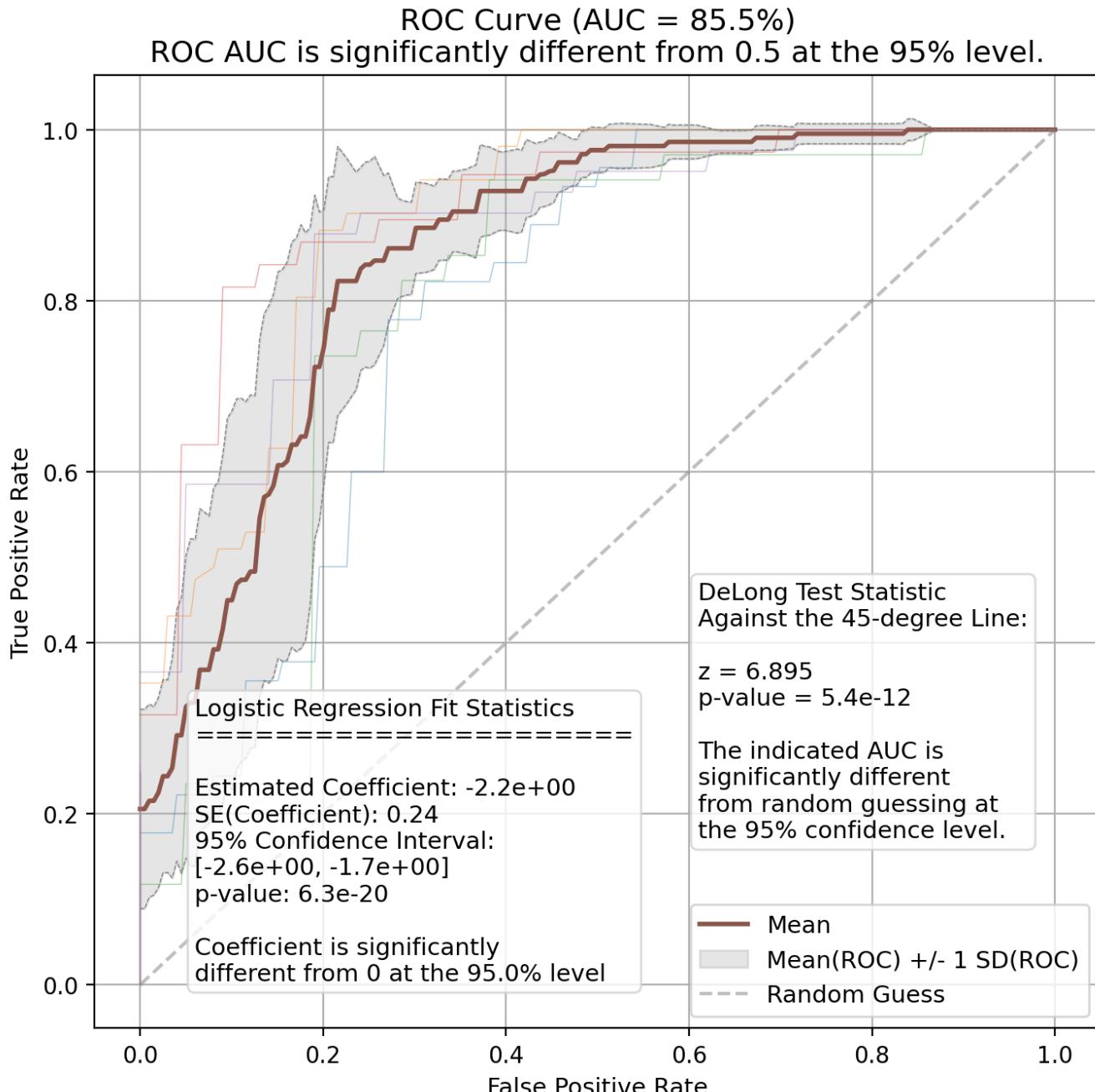
Log1P Radius Error - Empirical CDF Plot



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

Log1P Radius Error - ROC Curve

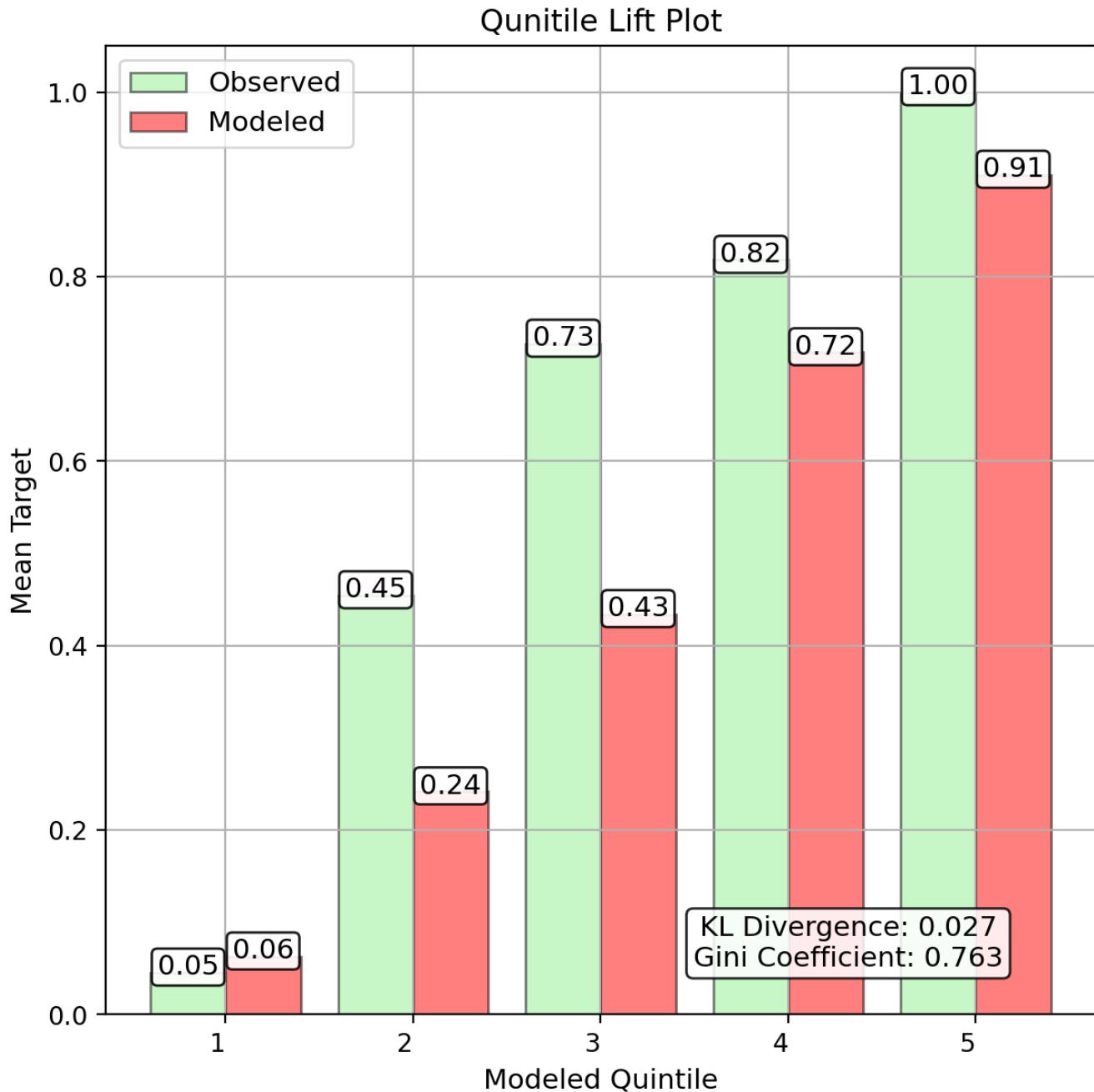


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Log1P 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

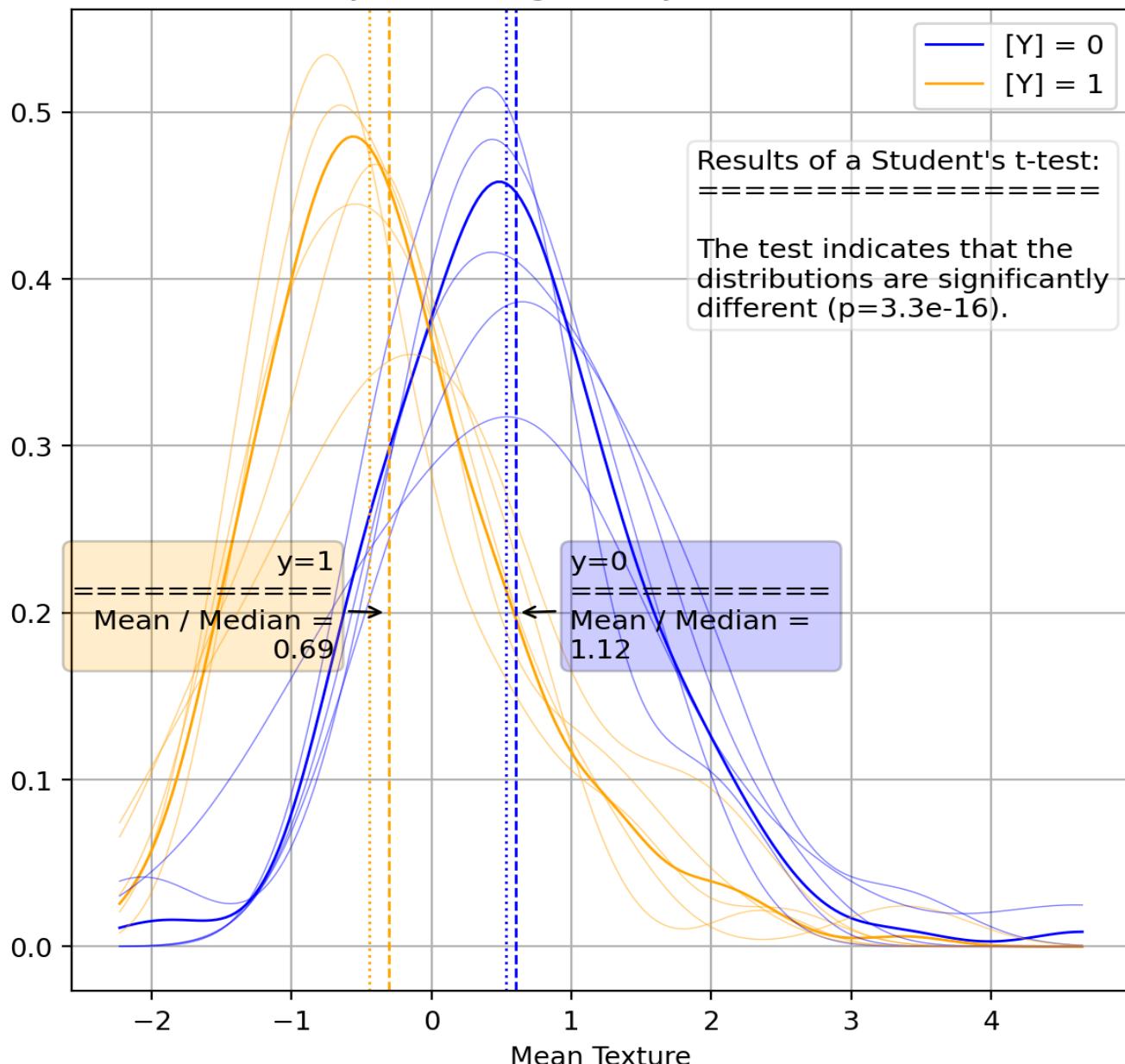
Mean Texture - Results

	Fitted Coef.	Fitted p-Value	Fitted	Conf.	Conf.	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
			Std. Err.	Int. Lower	Int. Upper						
Fold-1	-0.909	1.1e-08	0.159	-1.22	-0.597	78.9%	79.3%	82.4%	87.5%	77.8%	56.9%
Fold-2	-0.918	5.7e-09	0.158	-1.23	-0.609	71.1%	70.4%	75.5%	76.9%	74.1%	40.4%
Fold-3	-0.988	3.1e-10	0.157	-1.30	-0.680	71.4%	71.4%	75.8%	80.6%	71.4%	41.7%
Fold-4	-1.224	5.4e-12	0.178	-1.57	-0.876	61.3%	63.0%	64.7%	75.9%	56.4%	25.1%
Fold-5	-1.029	1.9e-10	0.162	-1.35	-0.713	75.8%	75.9%	80.5%	86.1%	75.6%	49.7%
mean	-1.009	3.5e-12	0.145	-1.29	-0.725	71.9%	72.0%	76.1%	81.0%	71.8%	42.8%
std	0.128	5.0e-09	0.009	0.14	0.112	6.7%	6.2%	6.9%	5.3%	8.5%	11.9%

Univariate Report

Mean Texture - Kernel Density Plot

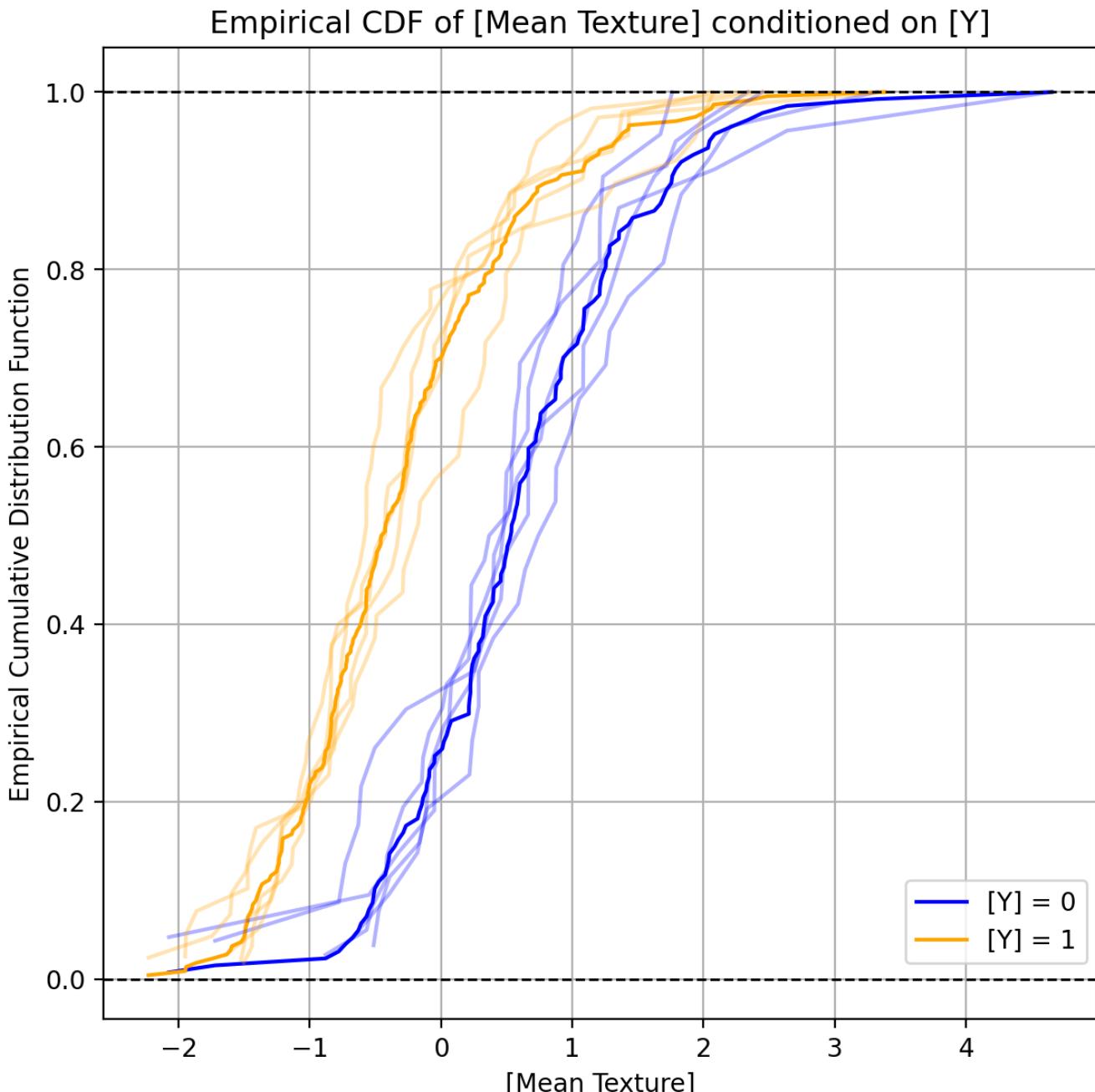
Kernel Density Plot of [Mean Texture] by [Y].
Distributions by level are significantly different at the 95% level.



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

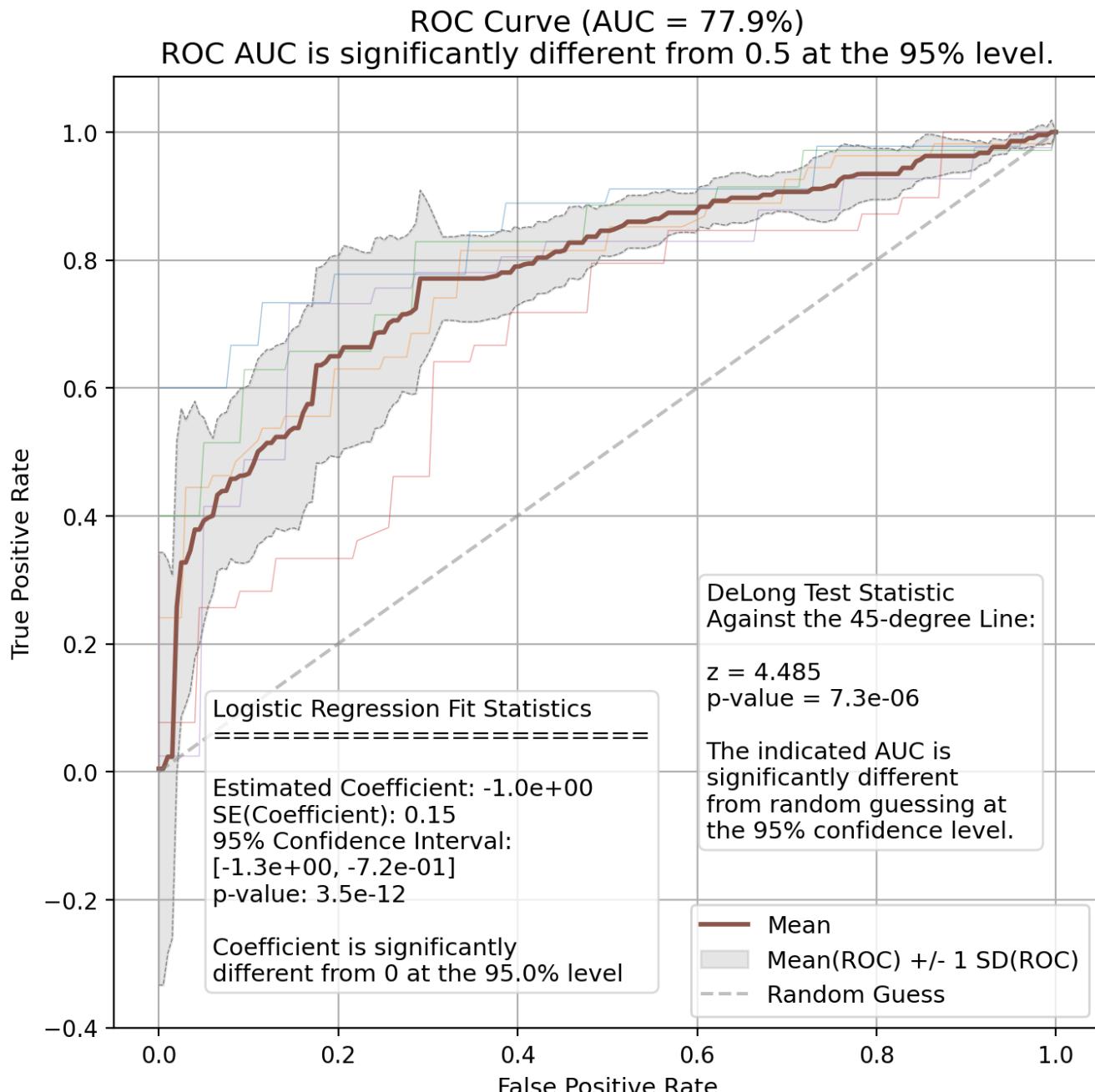
Mean Texture - Empirical CDF Plot



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

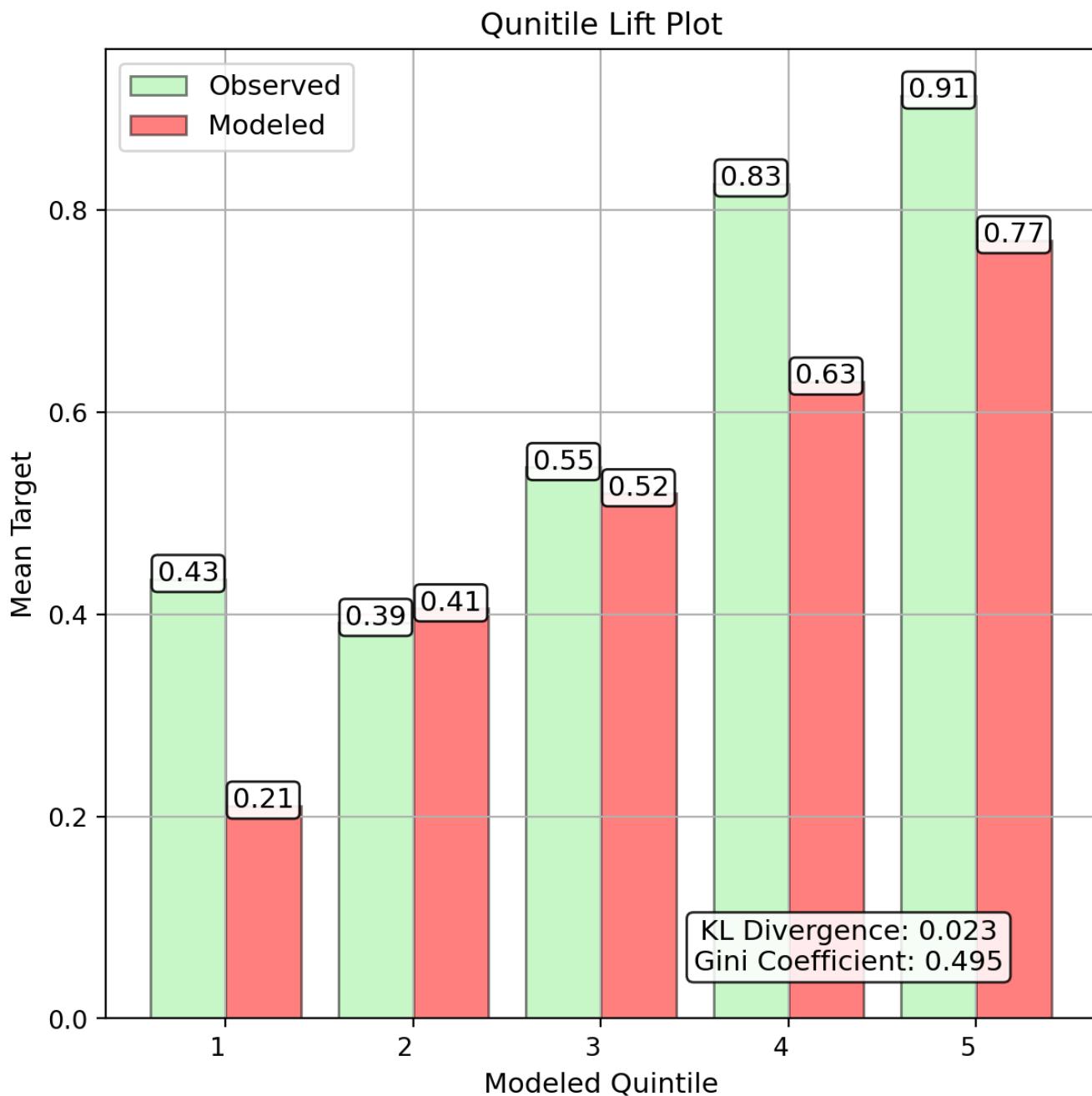


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

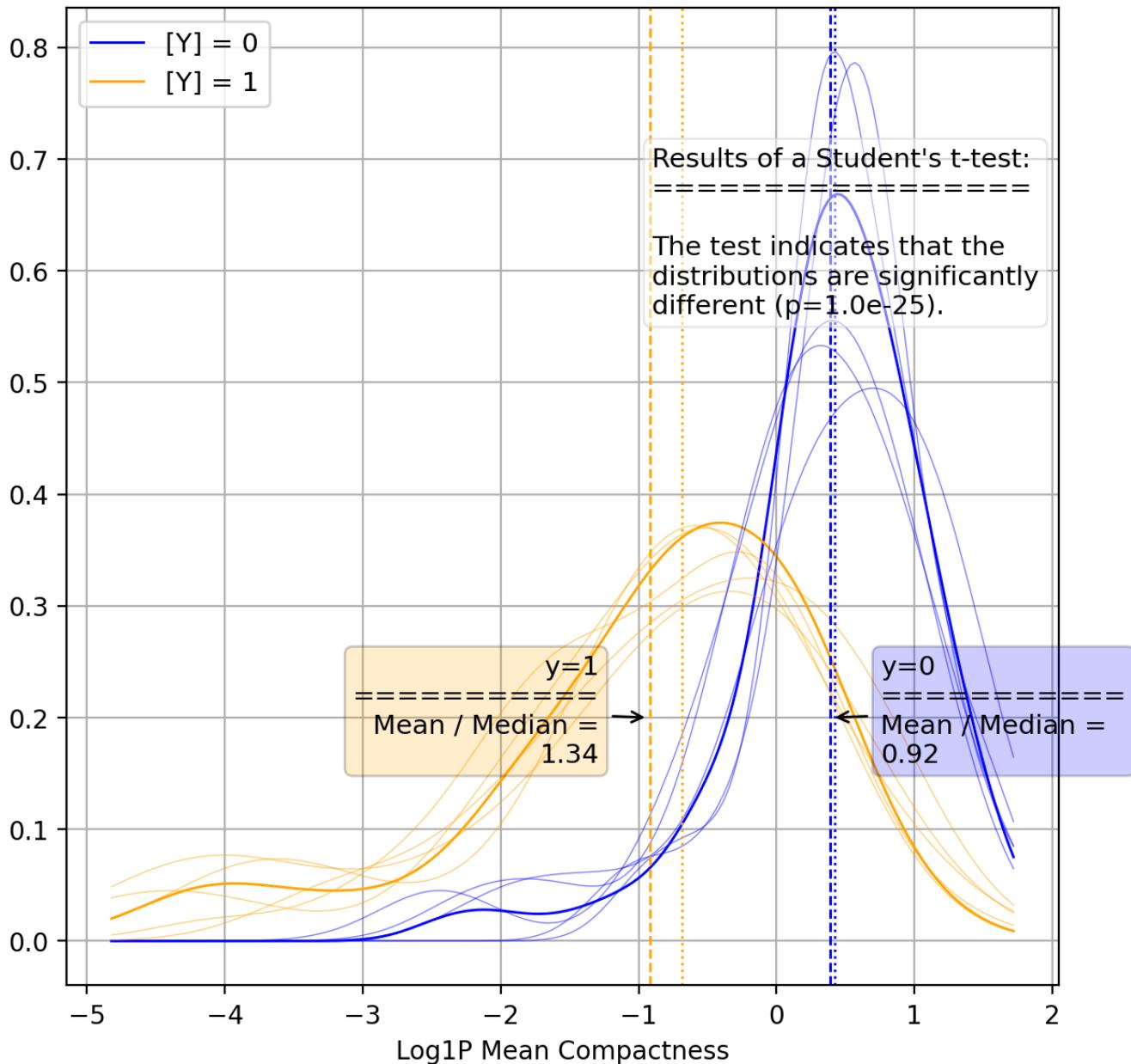
Log1P Mean Compactness - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-1.91	6.6e-13	0.266	-2.44	-1.39	70.5%	71.8%	71.0%	81.5%	62.9%	43.4%
Fold-2	-1.59	6.7e-11	0.243	-2.06	-1.11	75.0%	76.5%	73.0%	90.0%	61.4%	54.5%
Fold-3	-2.01	1.2e-13	0.271	-2.54	-1.48	65.3%	68.4%	63.8%	83.3%	51.7%	37.4%
Fold-4	-1.82	6.6e-13	0.253	-2.31	-1.32	72.7%	74.7%	72.7%	87.0%	62.5%	49.5%
Fold-5	-1.60	4.6e-12	0.231	-2.05	-1.14	76.8%	79.5%	78.7%	92.3%	68.6%	57.3%
mean	-1.78	2.9e-15	0.225	-2.22	-1.34	70.4%	72.6%	68.8%	86.5%	57.1%	46.2%
std	0.19	2.9e-11	0.016	0.22	0.16	4.5%	4.3%	5.3%	4.5%	6.1%	8.1%

Univariate Report

Log1P Mean Compactness - Kernel Density Plot

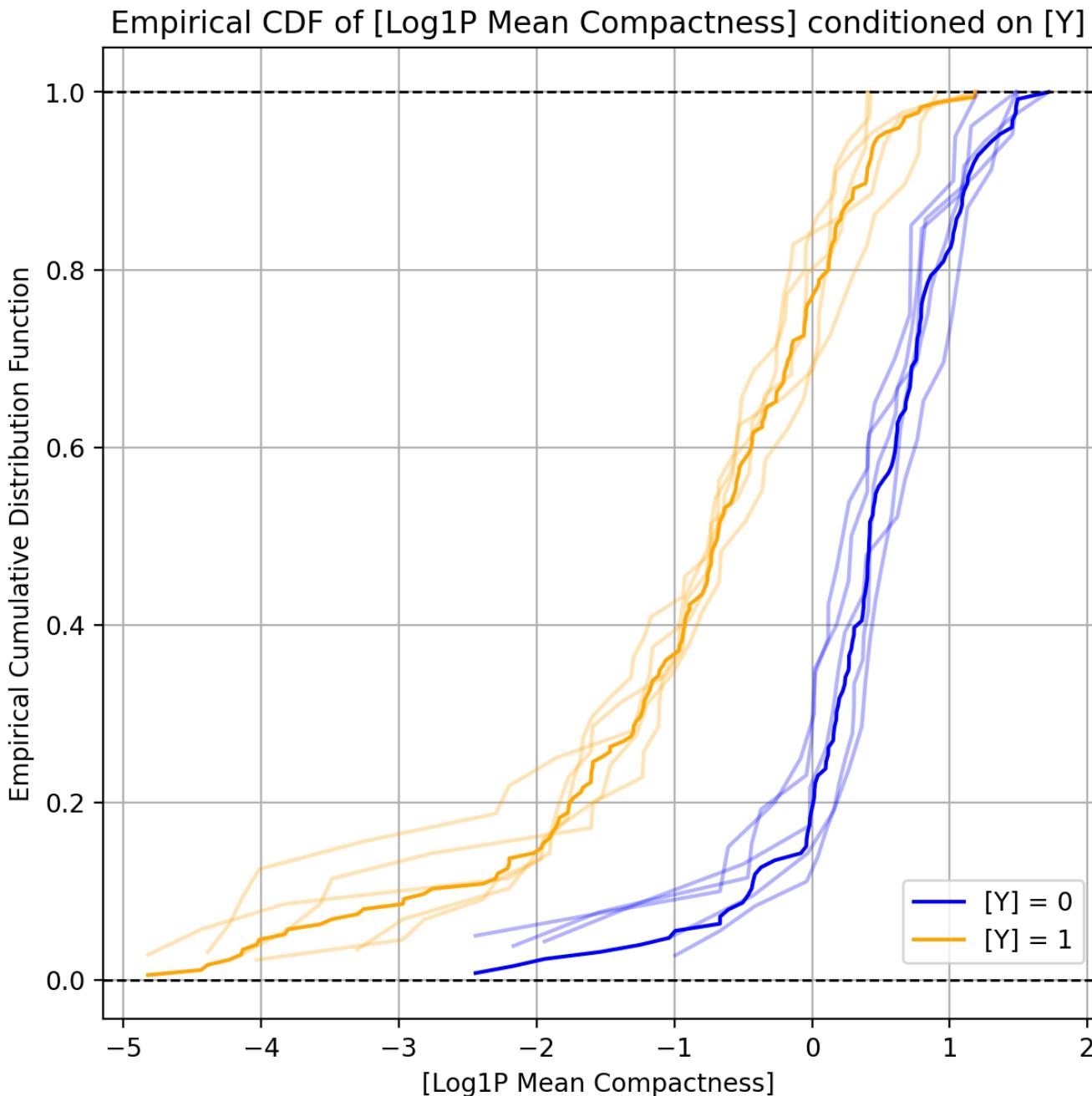
Kernel Density Plot of [Log1P Mean Compactness] by [Y].
Distributions by level are significantly different at the 95% level.



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

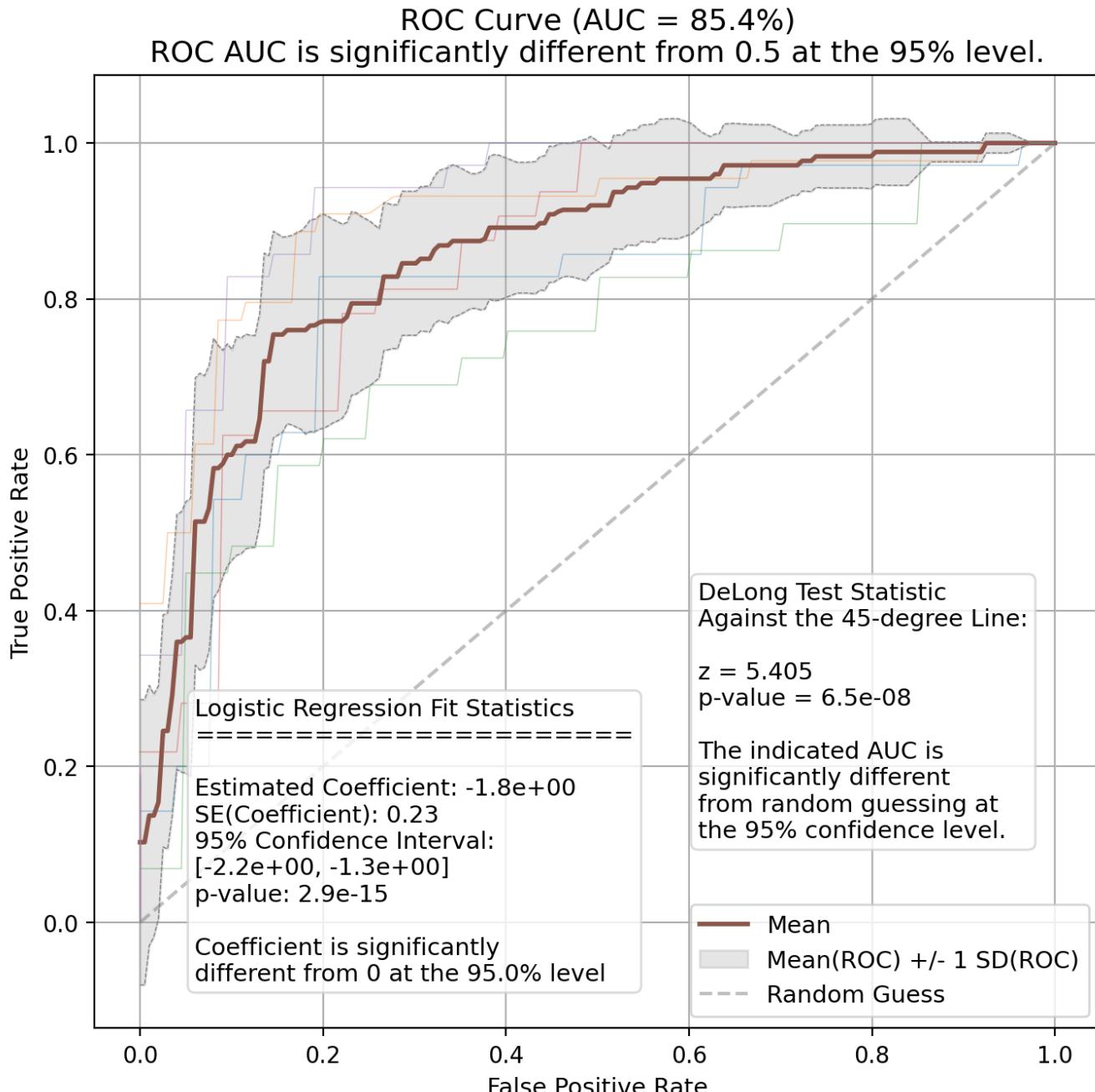
Log1P Mean Compactness - Empirical CDF Plot



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

Log1P Mean Compactness - ROC Curve

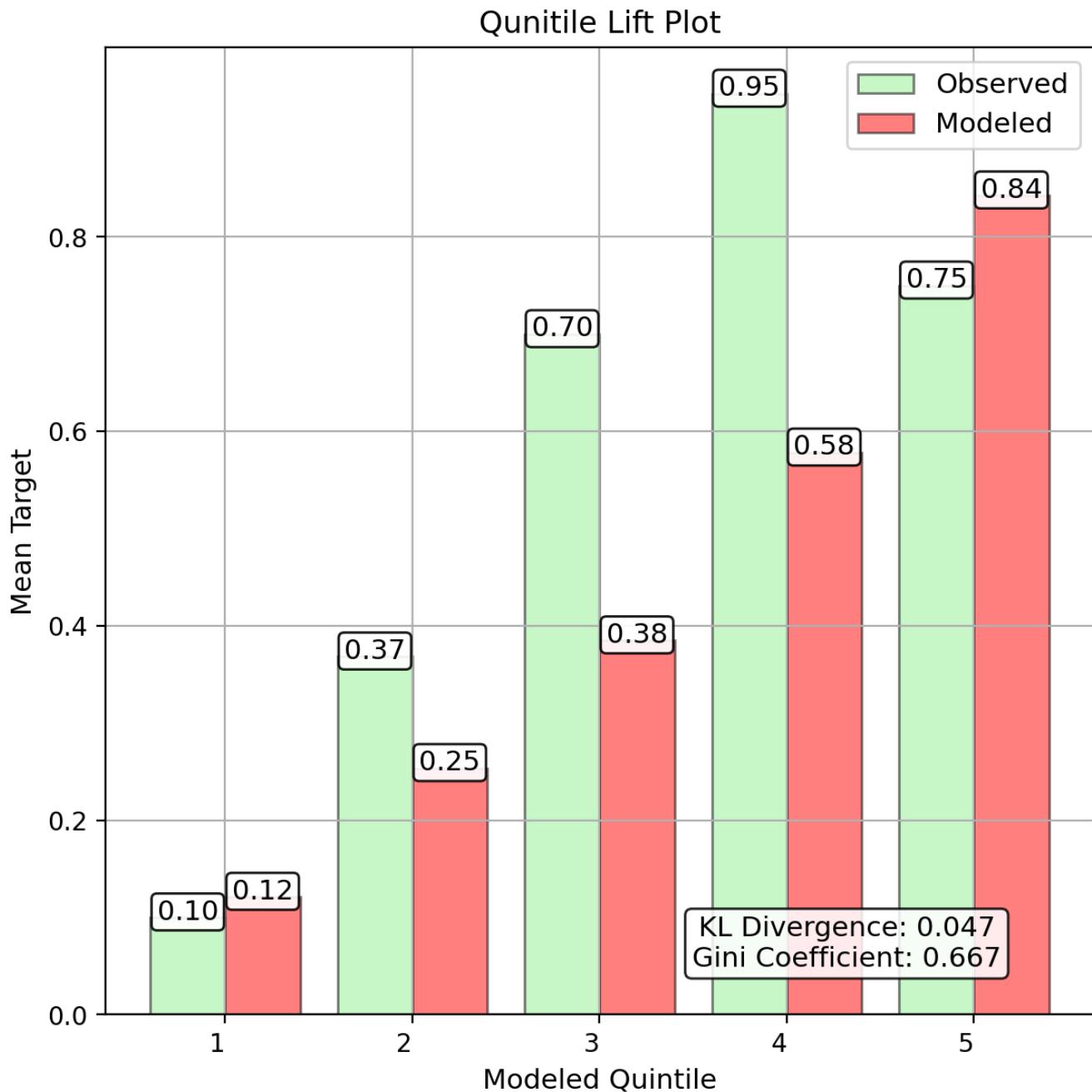


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Log1P Mean Compactness - Quintile Lift



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

Univariate Report

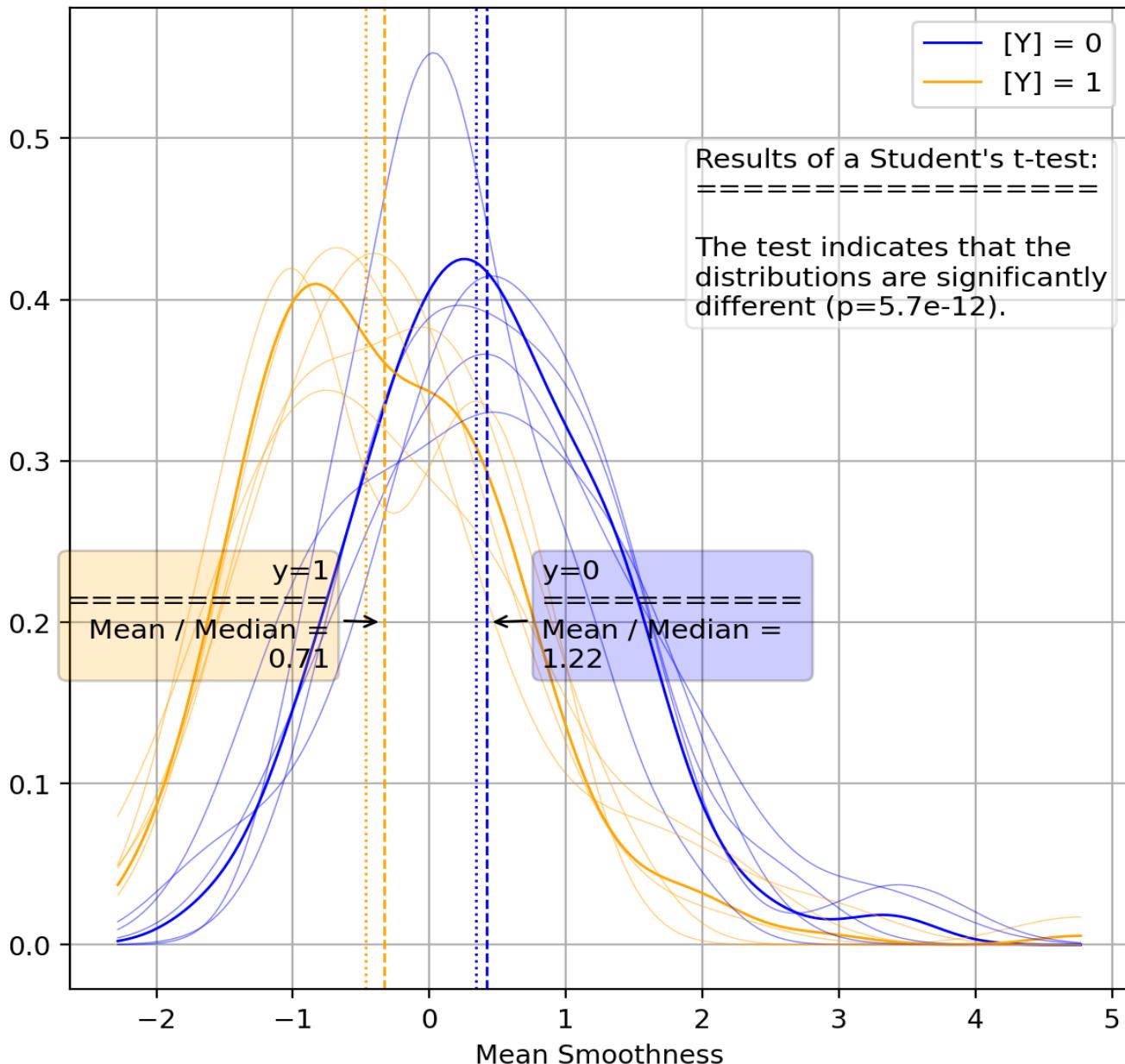
Mean Smoothness - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-0.900	2.0e-09	0.150	-1.19	-0.606	66.2%	65.2%	72.1%	75.6%	68.9%	29.7%
Fold-2	-0.723	1.6e-06	0.151	-1.02	-0.427	67.8%	69.0%	70.1%	79.1%	63.0%	37.2%
Fold-3	-0.905	2.4e-09	0.152	-1.20	-0.608	66.1%	65.2%	71.6%	75.0%	68.6%	29.8%
Fold-4	-0.862	1.8e-08	0.153	-1.16	-0.562	66.1%	67.7%	69.6%	80.0%	61.5%	34.3%
Fold-5	-0.784	4.9e-08	0.144	-1.07	-0.502	59.7%	62.5%	63.8%	78.6%	53.7%	23.9%
mean	-0.836	4.5e-10	0.134	-1.10	-0.573	67.5%	68.9%	70.9%	80.4%	63.4%	36.6%
std	0.079	7.3e-07	0.004	0.08	0.077	3.2%	2.5%	3.3%	2.2%	6.2%	5.1%

Univariate Report

Mean Smoothness - Kernel Density Plot

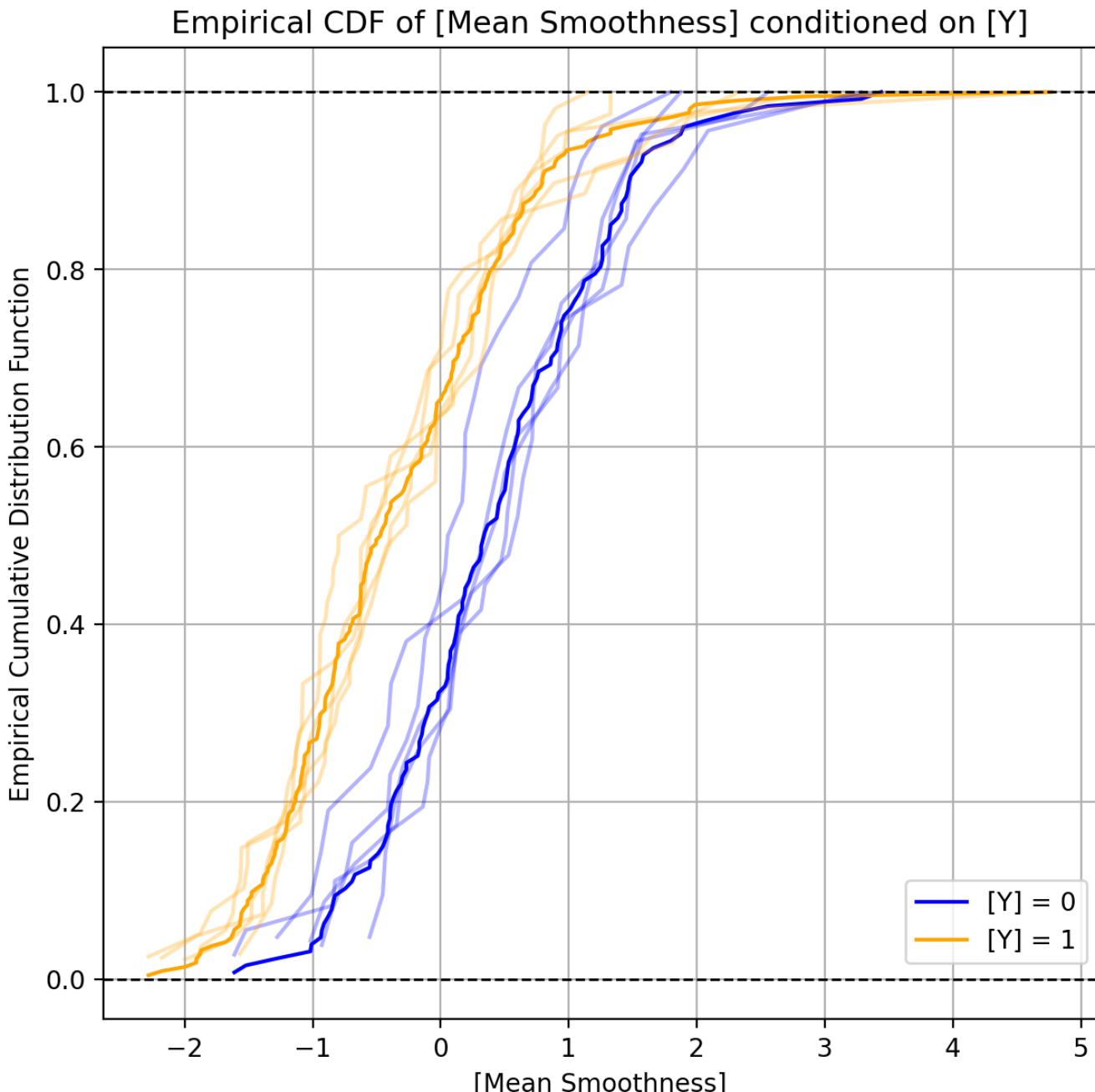
Kernel Density Plot of [Mean Smoothness] by [Y].
Distributions by level are significantly different at the 95% level.



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

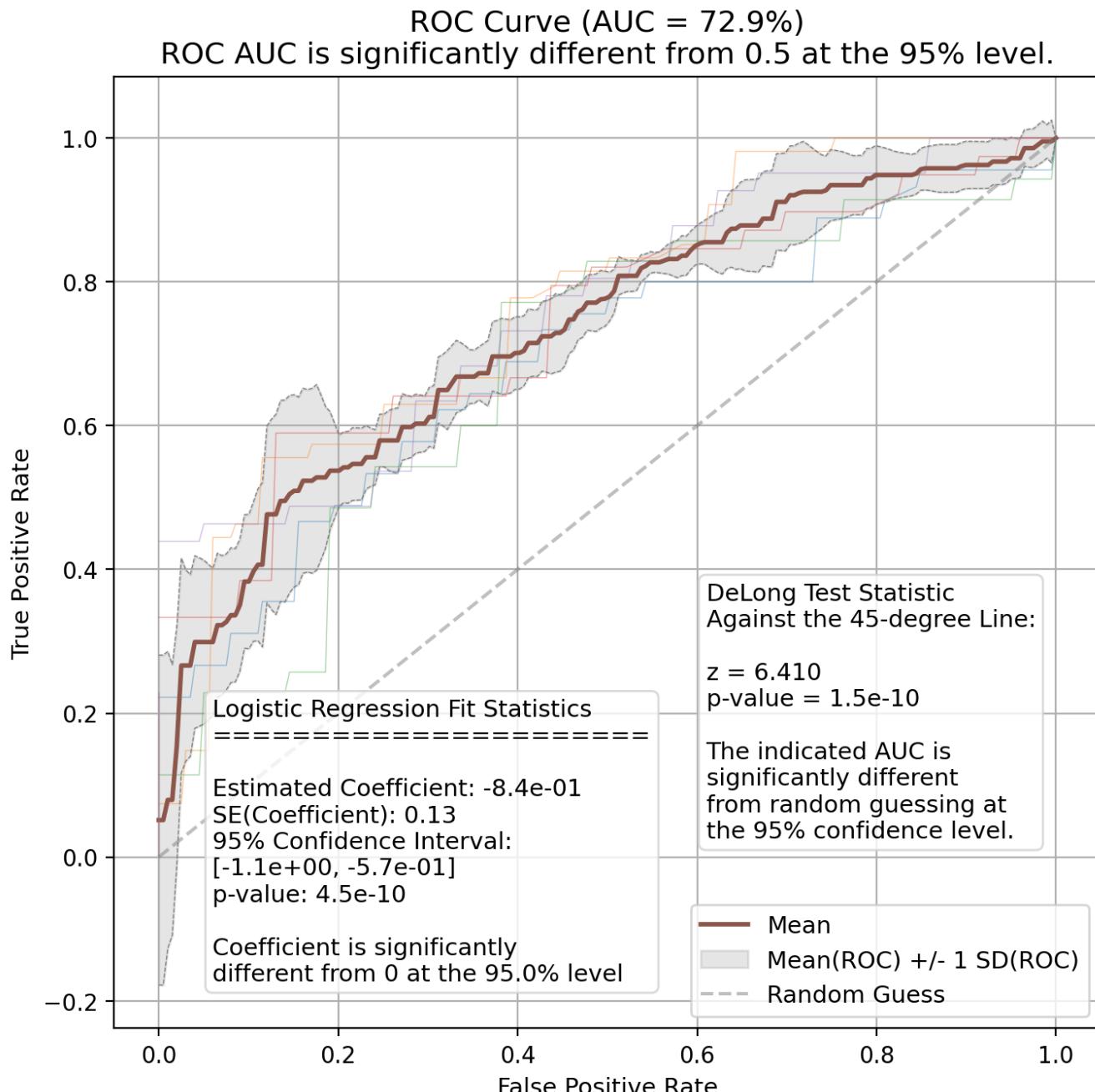
Mean Smoothness - Empirical CDF Plot



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

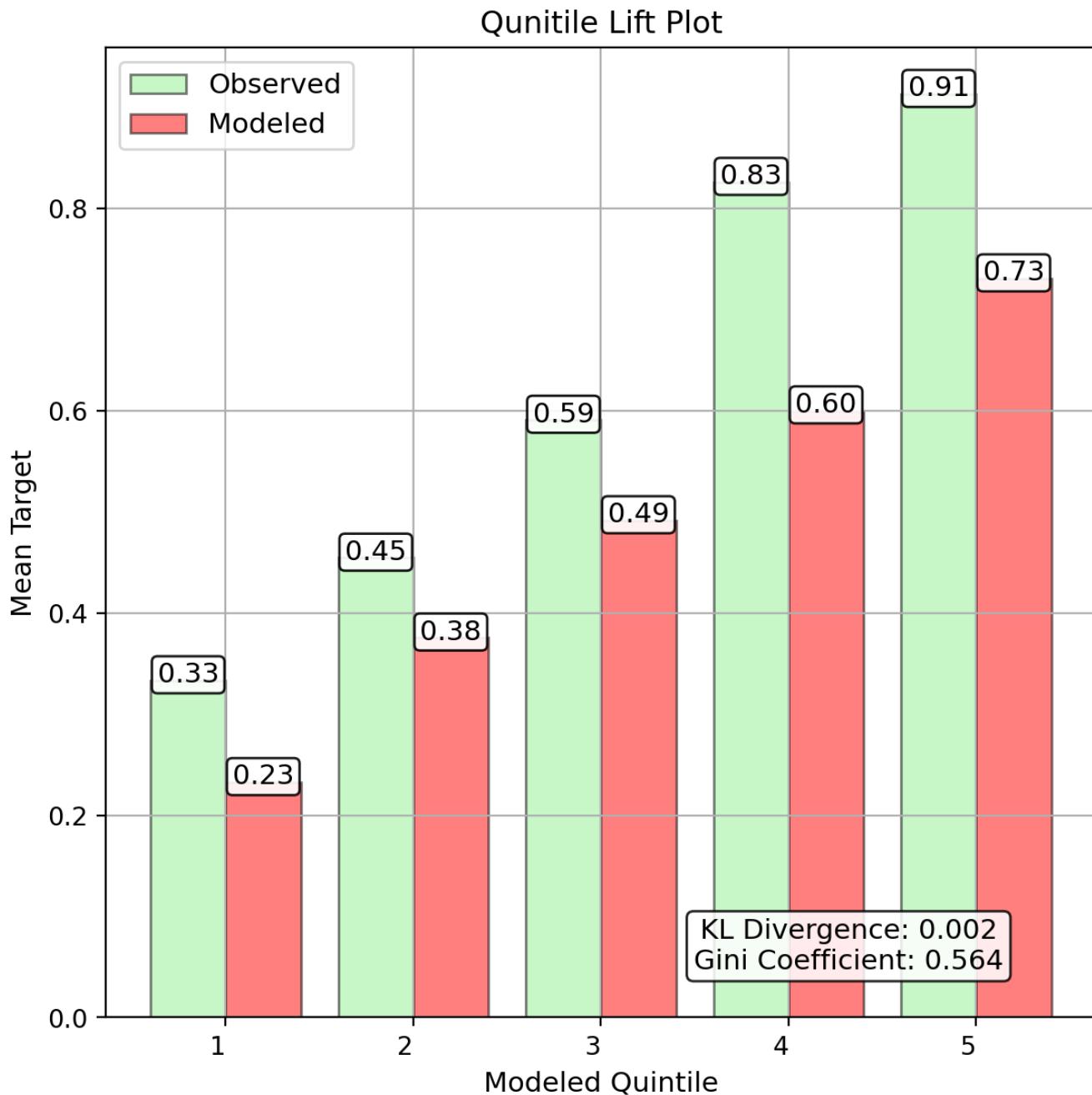


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Univariate Report

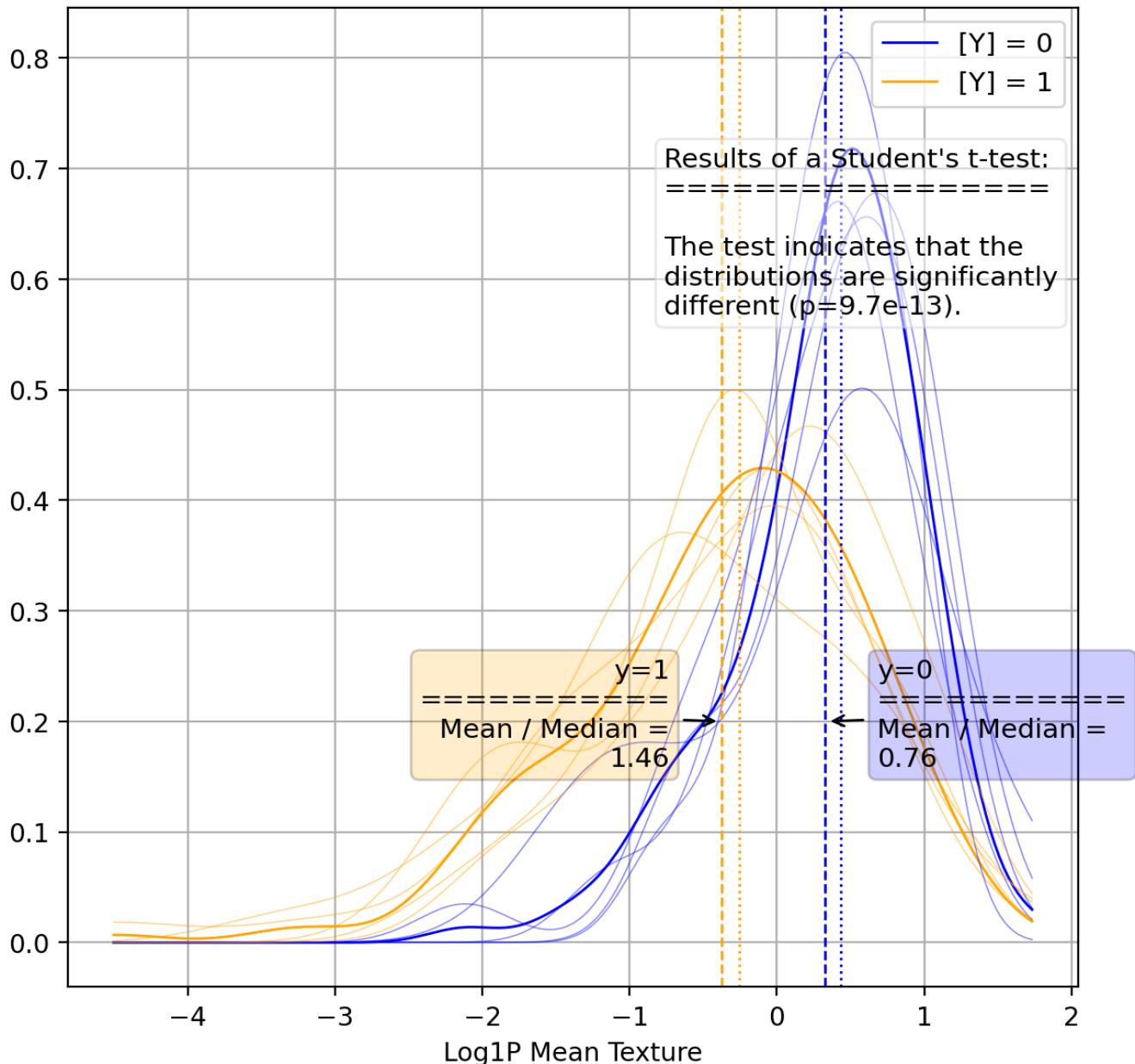
Log1P Mean Texture - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-0.98	5.0e-07	0.195	-1.37	-0.599	74.6%	75.2%	75.4%	82.1%	69.7%	50.2%
Fold-2	-1.08	1.5e-07	0.206	-1.48	-0.677	65.0%	65.7%	65.0%	72.2%	59.1%	31.3%
Fold-3	-1.07	3.4e-08	0.193	-1.44	-0.688	64.6%	65.9%	63.8%	75.0%	55.6%	31.9%
Fold-4	-1.33	9.4e-10	0.218	-1.76	-0.906	54.7%	57.3%	52.0%	68.4%	41.9%	15.1%
Fold-5	-1.06	2.9e-08	0.190	-1.43	-0.682	75.0%	77.8%	76.4%	91.3%	65.6%	54.5%
mean	-1.10	7.6e-10	0.179	-1.45	-0.749	65.3%	66.6%	64.6%	75.6%	56.4%	33.3%
std	0.13	2.1e-07	0.011	0.15	0.115	8.4%	8.2%	10.0%	9.1%	10.7%	15.9%

Univariate Report

Log1P Mean Texture - Kernel Density Plot

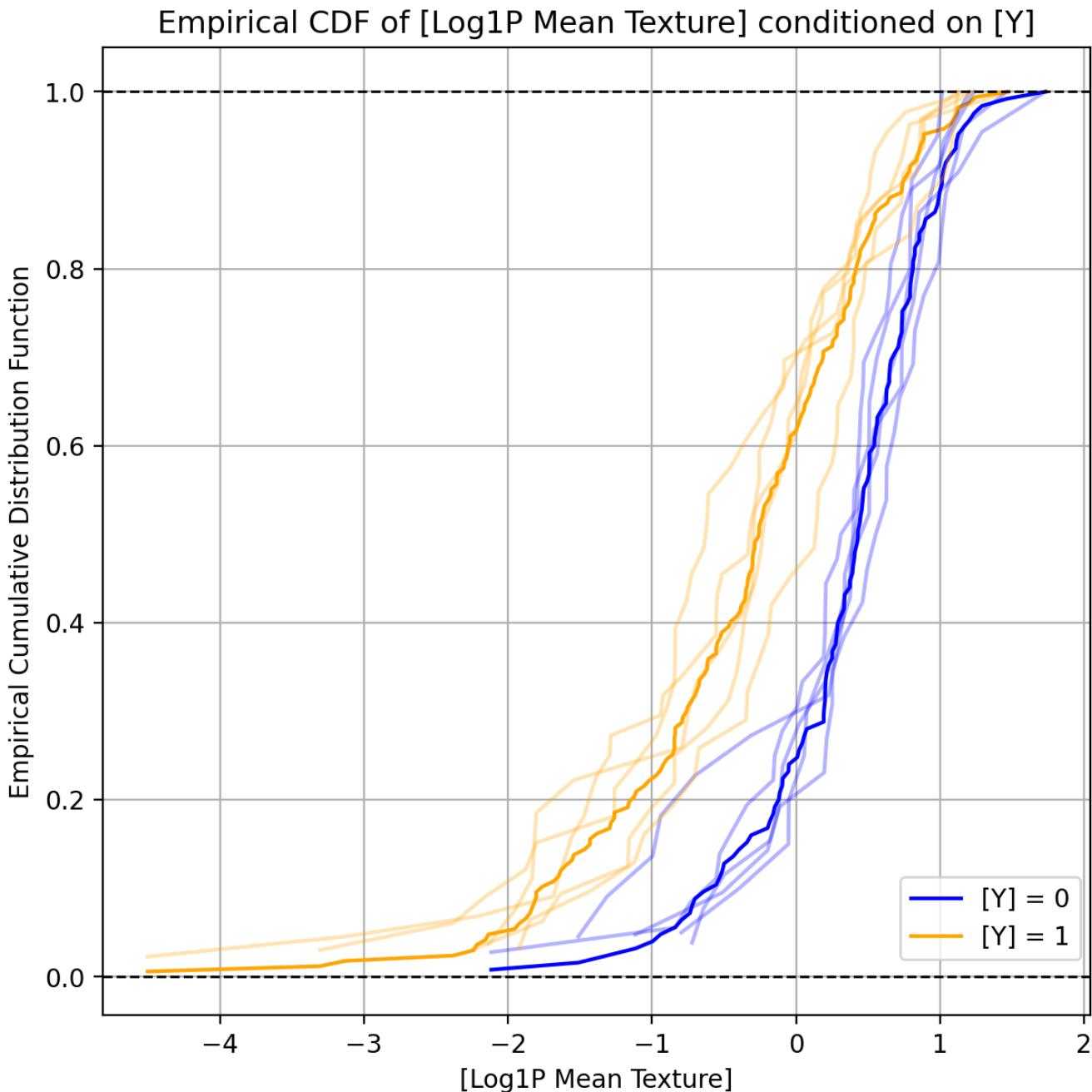
Kernel Density Plot of [Log1P Mean Texture] by [Y].
Distributions by level are significantly different at the 95% level.



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

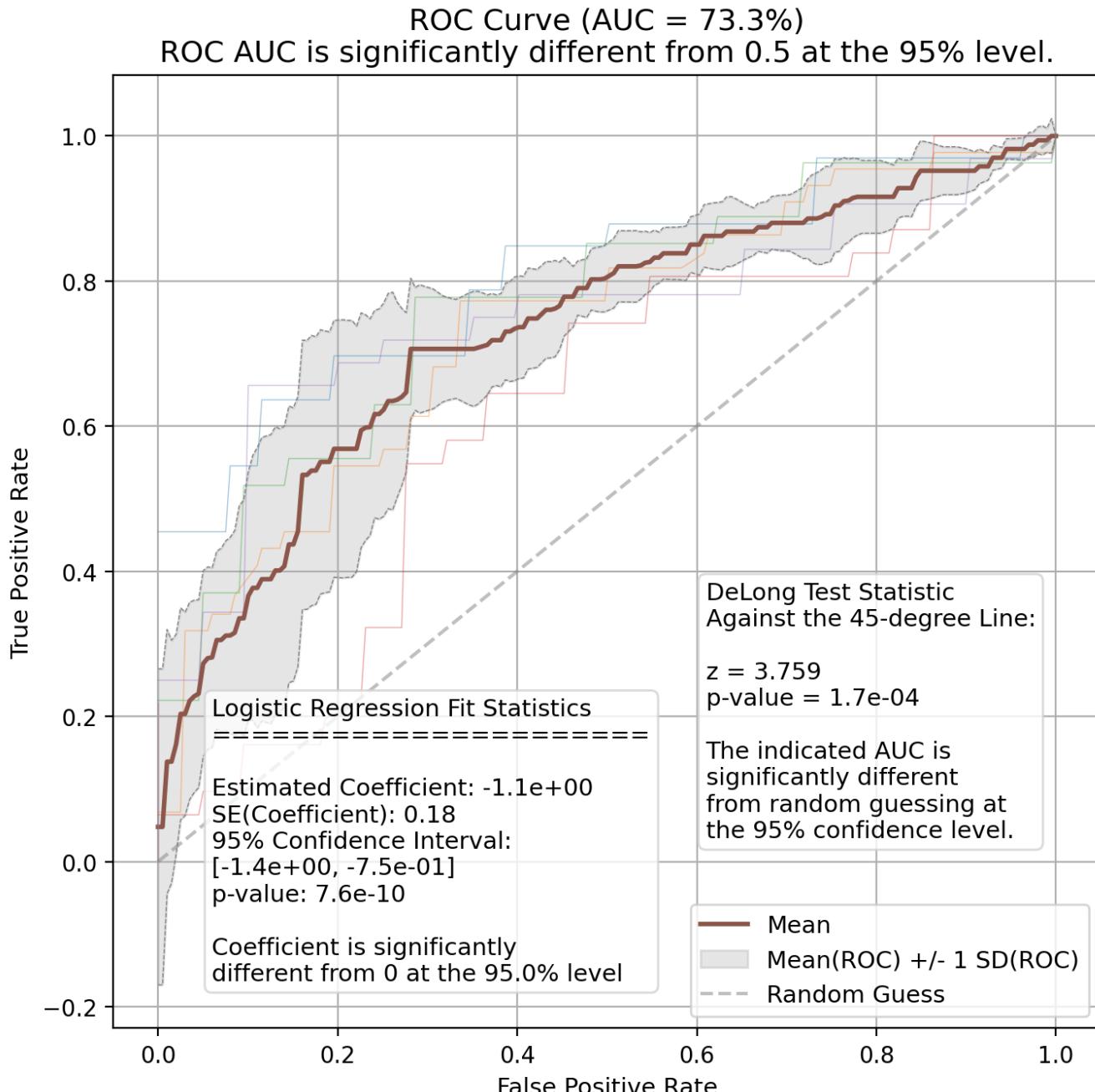
Log1P Mean Texture - Empirical CDF Plot



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

Log1P Mean Texture - ROC Curve

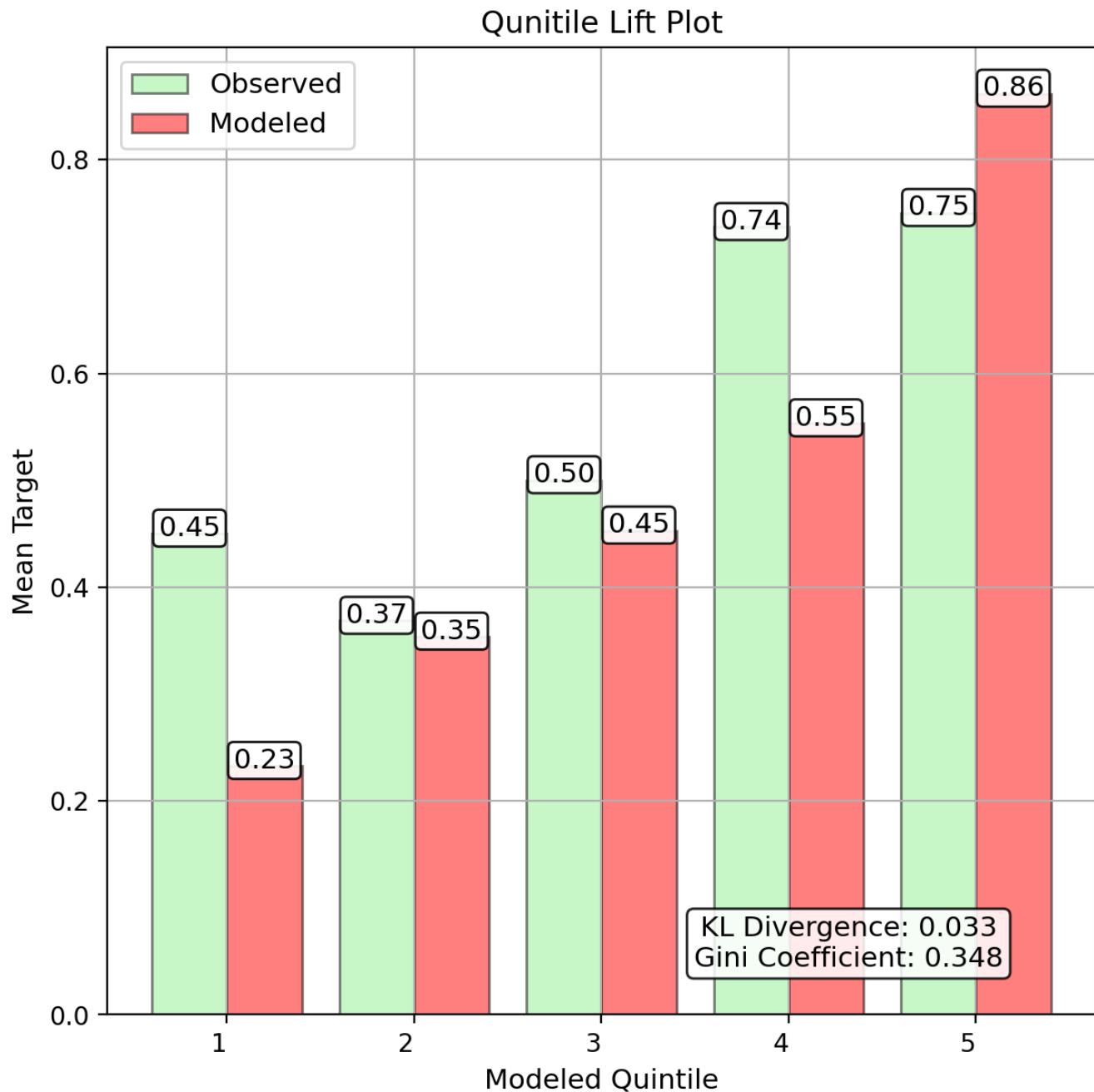


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Log1P 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

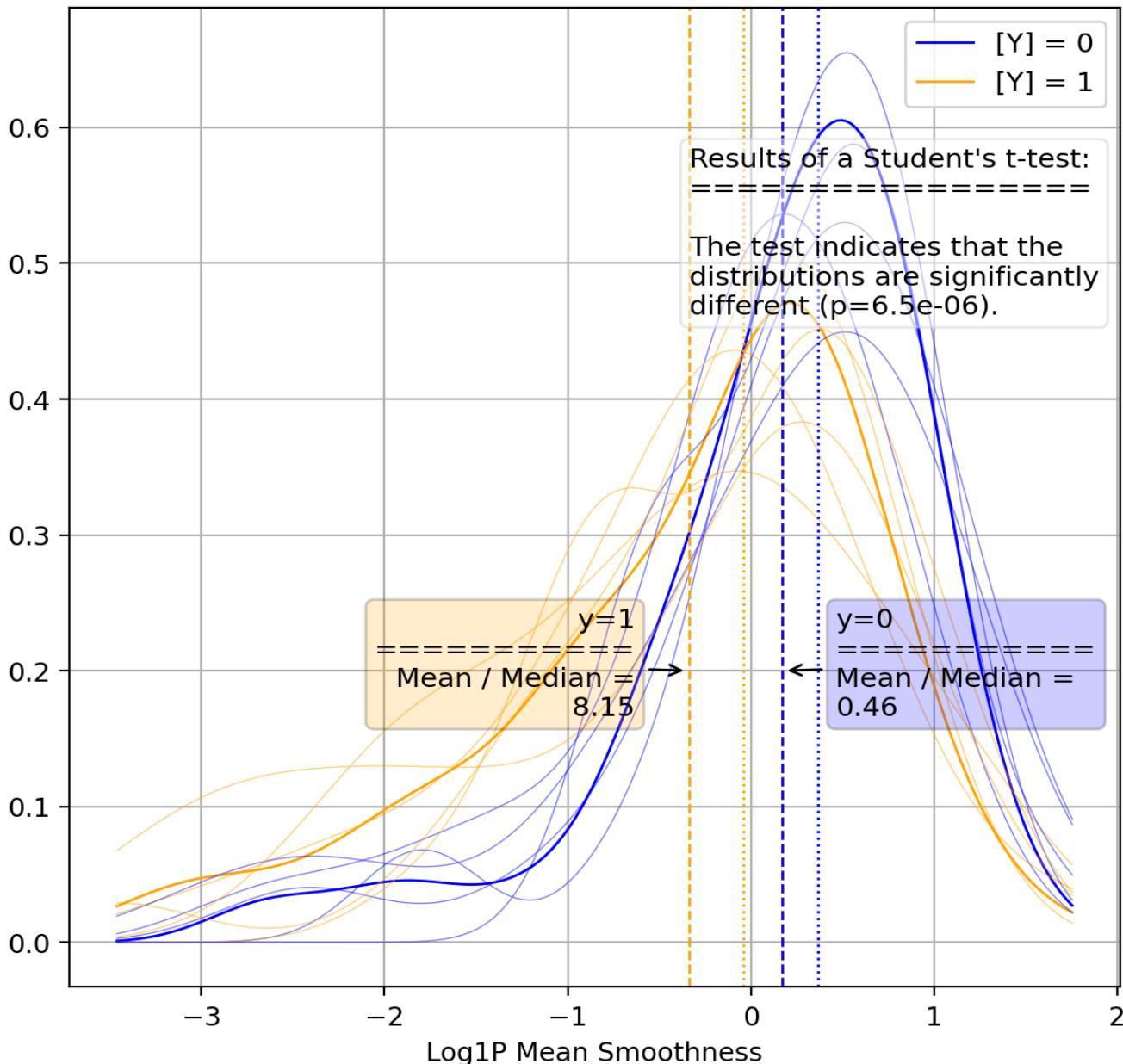
Log1P Mean Smoothness - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-0.698	4.5e-05	0.171	-1.033	-0.363	61.3%	61.3%	64.7%	68.8%	61.1%	22.4%
Fold-2	-0.514	2.3e-03	0.169	-0.845	-0.183	59.7%	60.9%	50.8%	71.4%	39.5%	24.0%
Fold-3	-0.664	6.5e-05	0.166	-0.990	-0.338	63.0%	63.8%	65.3%	72.7%	59.3%	27.3%
Fold-4	-0.653	4.5e-05	0.160	-0.966	-0.339	58.3%	59.8%	52.4%	68.8%	42.3%	20.7%
Fold-5	-0.590	1.3e-04	0.154	-0.893	-0.288	51.0%	54.0%	46.8%	64.7%	36.7%	8.5%
mean	-0.625	1.9e-05	0.146	-0.911	-0.338	62.2%	64.0%	61.1%	74.4%	51.8%	28.3%
std	0.073	1.0e-03	0.007	0.076	0.072	4.6%	3.6%	8.5%	3.1%	11.5%	7.2%

Univariate Report

Log1P Mean Smoothness - Kernel Density Plot

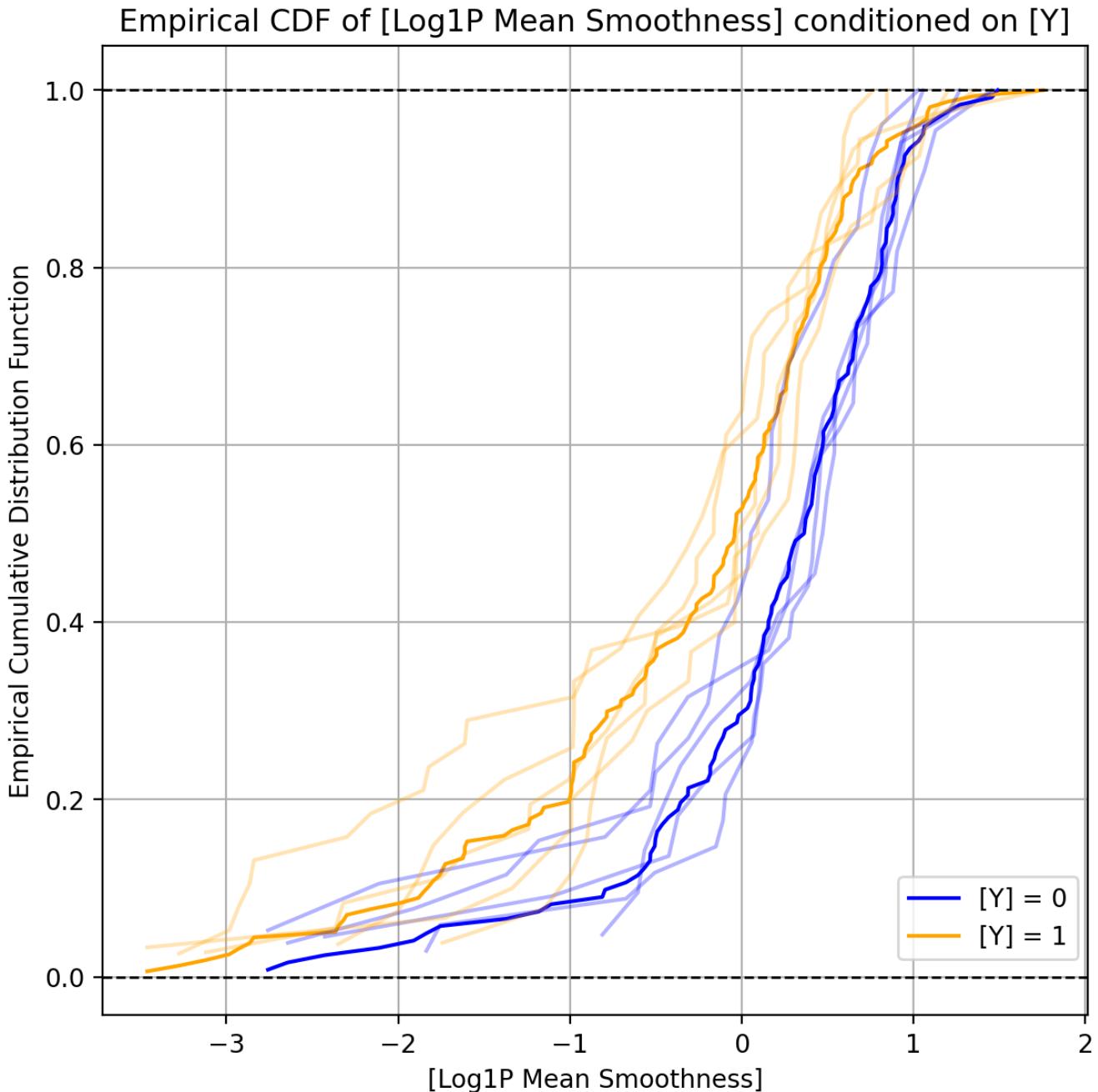
Kernel Density Plot of [Log1P Mean Smoothness] by [Y].
Distributions by level are significantly different at the 95% level.



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

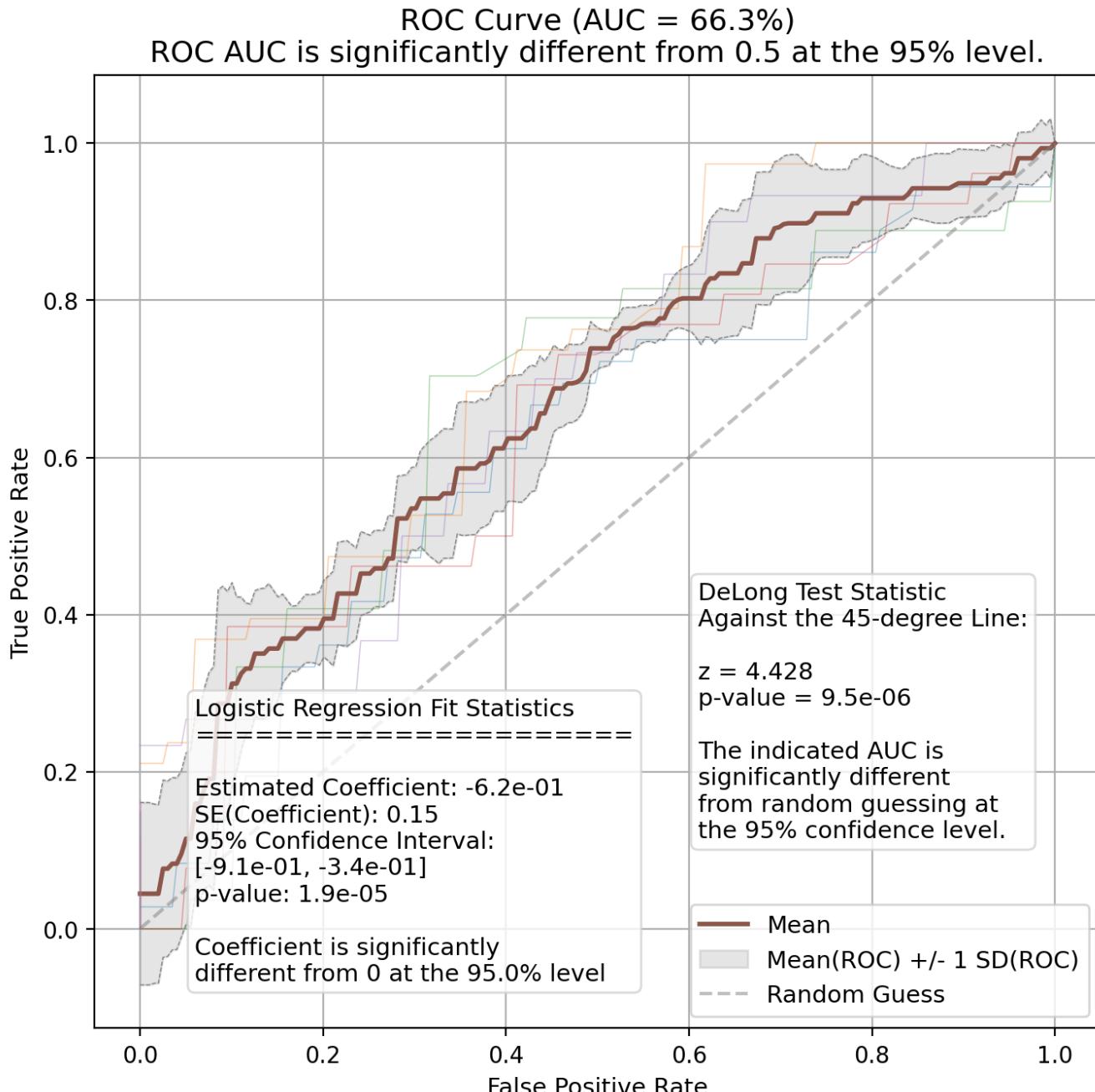
Log1P Mean Smoothness - Empirical CDF Plot



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

Log1P Mean Smoothness - ROC Curve

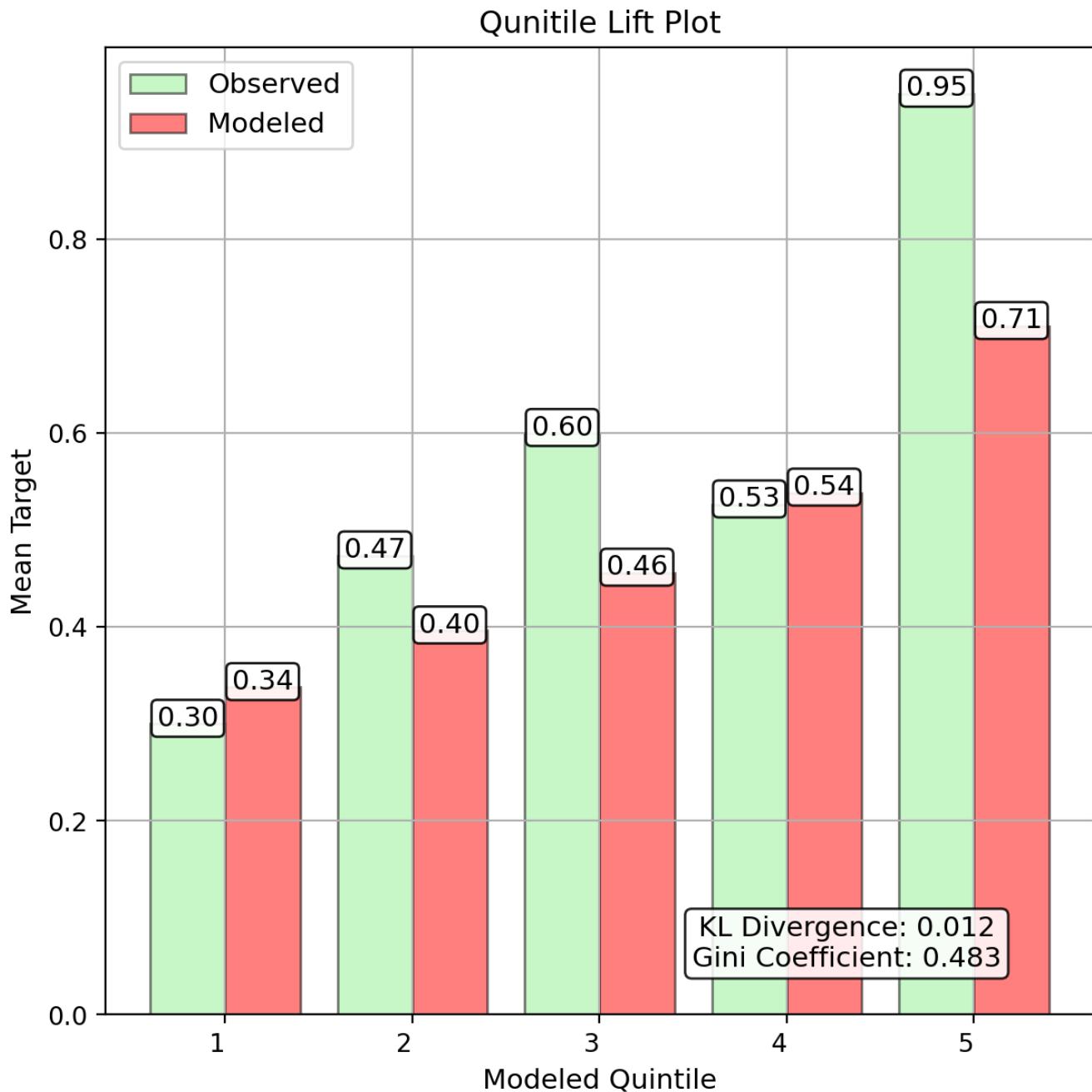


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

Log1P Mean Smoothness - Quintile Lift



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

Univariate Report

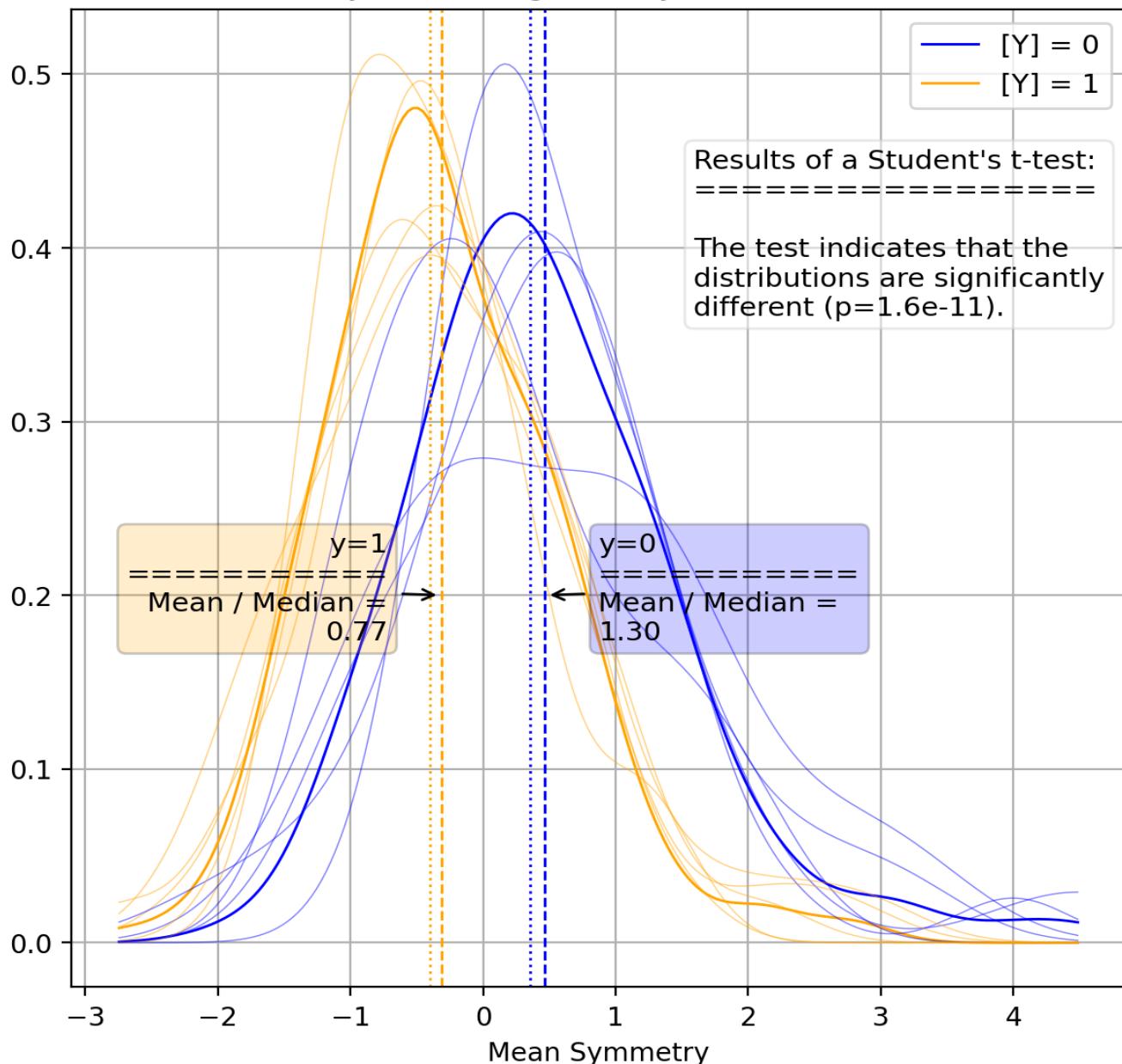
Mean Symmetry - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf.		Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
				Int. Lower	Int. Upper						
Fold-1	-0.987	2.9e-09	0.166	-1.31	-0.661	54.9%	52.3%	63.6%	65.1%	62.2%	4.5%
Fold-2	-0.766	6.4e-07	0.154	-1.07	-0.465	70.0%	69.9%	73.8%	77.6%	70.4%	39.2%
Fold-3	-0.830	3.7e-08	0.151	-1.13	-0.534	73.2%	73.8%	76.9%	83.3%	71.4%	46.2%
Fold-4	-1.003	2.1e-09	0.167	-1.33	-0.675	61.3%	62.1%	65.7%	74.2%	59.0%	23.4%
Fold-5	-0.914	1.5e-08	0.161	-1.23	-0.598	67.7%	69.8%	72.2%	83.9%	63.4%	37.5%
mean	-0.896	3.3e-10	0.143	-1.18	-0.617	55.3%	57.2%	57.9%	70.0%	49.3%	14.1%
std	0.101	2.8e-07	0.007	0.12	0.088	7.3%	8.6%	5.6%	7.7%	5.4%	16.6%

Univariate Report

Mean Symmetry - Kernel Density Plot

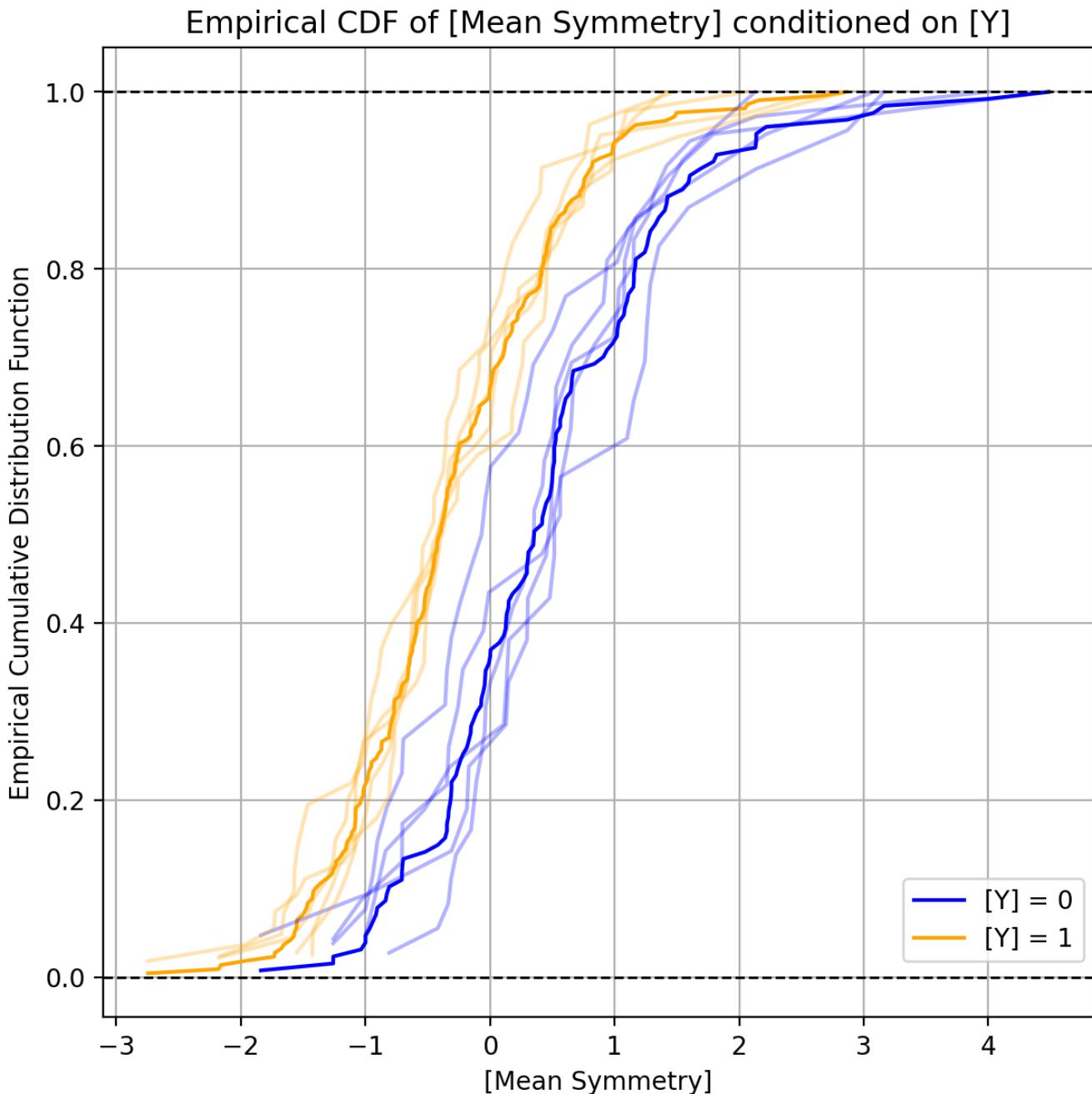
Kernel Density Plot of [Mean Symmetry] by [Y].
Distributions by level are significantly different at the 95% level.



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

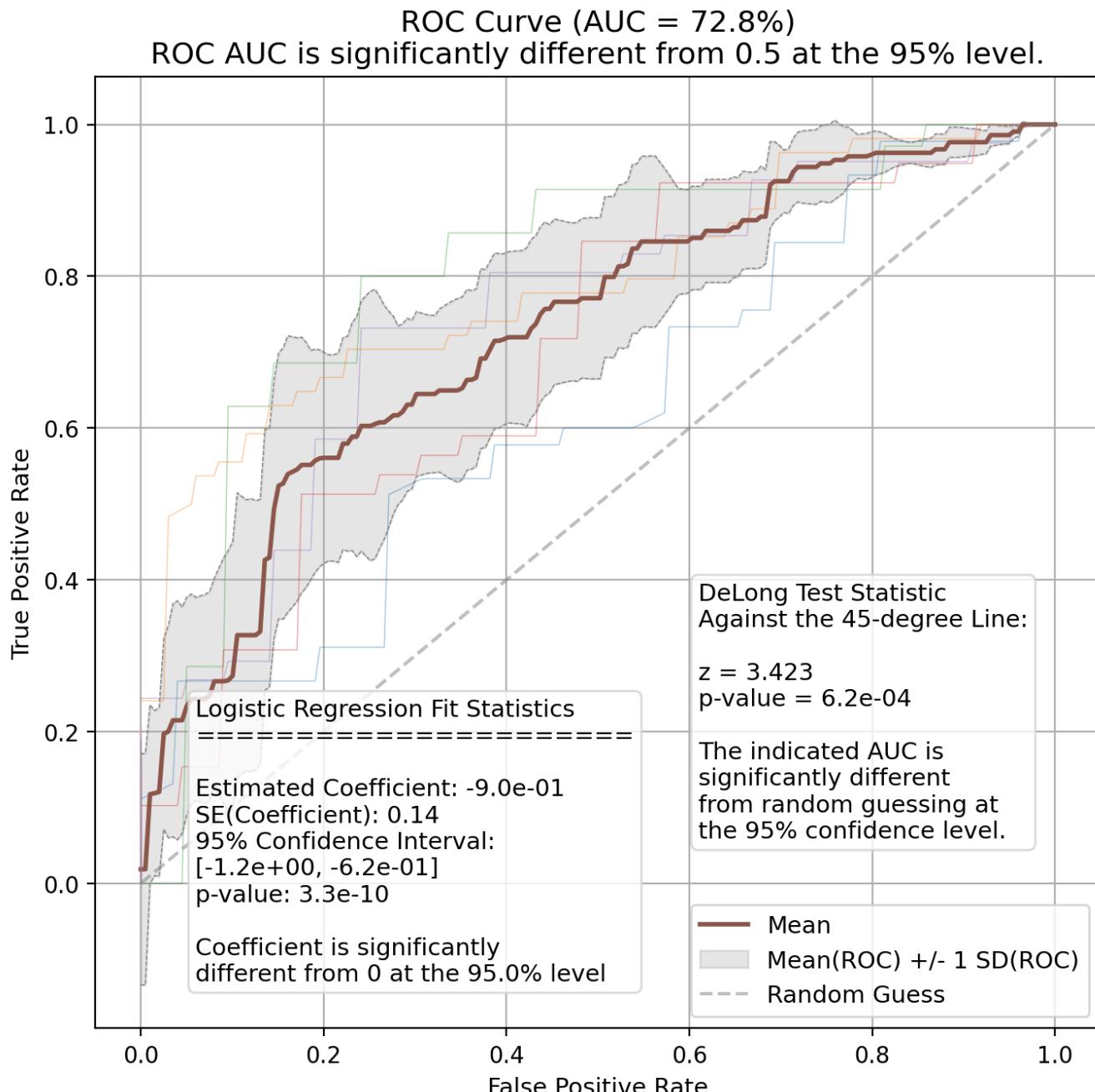
Mean Symmetry - Empirical CDF Plot



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

Univariate Report

Mean Symmetry - ROC Curve

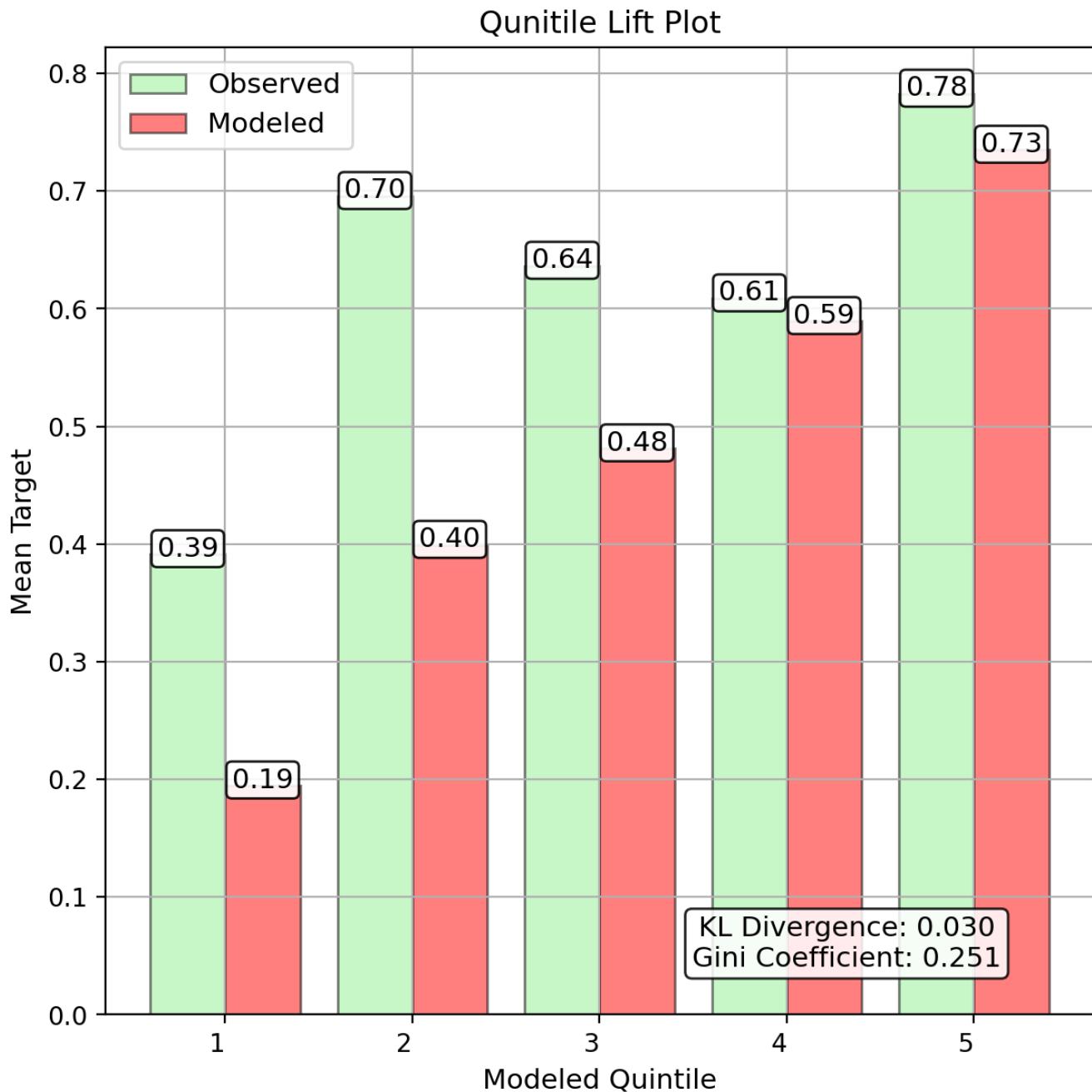


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Mean Symmetry - Quintile Lift



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

Univariate Report

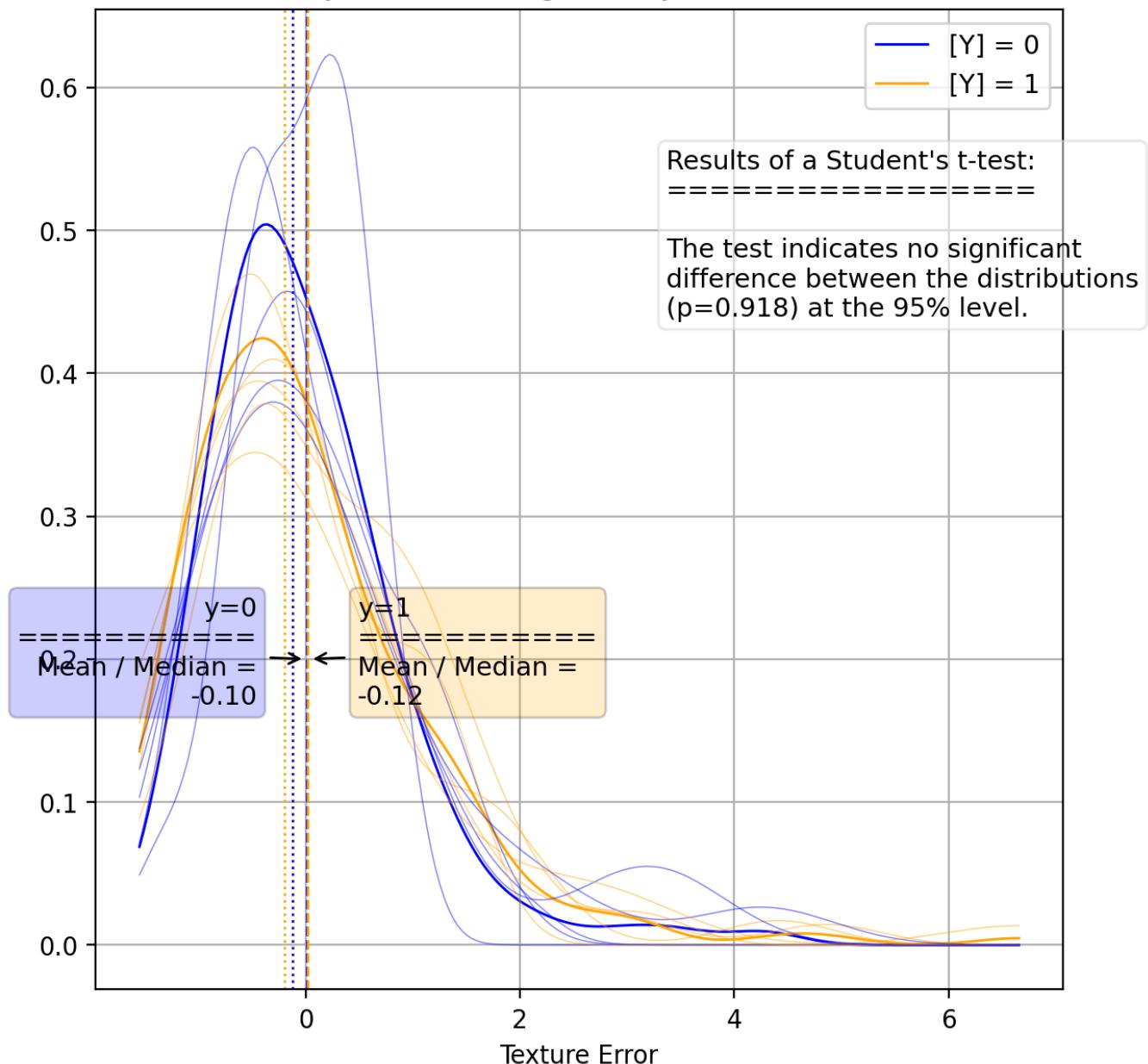
Texture Error - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	2.2e-02	0.844	0.114	-0.200	0.245	40.8%	44.4%	40.0%	56.0%	31.1%	-11.3%
Fold-2	-8.1e-03	0.945	0.117	-0.238	0.222	51.1%	48.6%	60.0%	58.9%	61.1%	-2.8%
Fold-3	4.1e-02	0.700	0.107	-0.168	0.251	39.3%	41.0%	41.4%	52.2%	34.3%	-17.8%
Fold-4	-5.8e-03	0.960	0.114	-0.229	0.218	46.8%	45.2%	54.8%	58.8%	51.3%	-9.3%
Fold-5	4.2e-02	0.736	0.124	-0.201	0.285	45.2%	49.2%	46.9%	65.2%	36.6%	-1.5%
mean	1.9e-02	0.857	0.103	-0.183	0.220	55.3%	57.2%	57.9%	70.0%	49.3%	14.1%
std	2.4e-02	0.118	0.006	0.027	0.027	4.7%	3.4%	8.6%	4.8%	12.8%	6.6%

Univariate Report

Texture Error - Kernel Density Plot

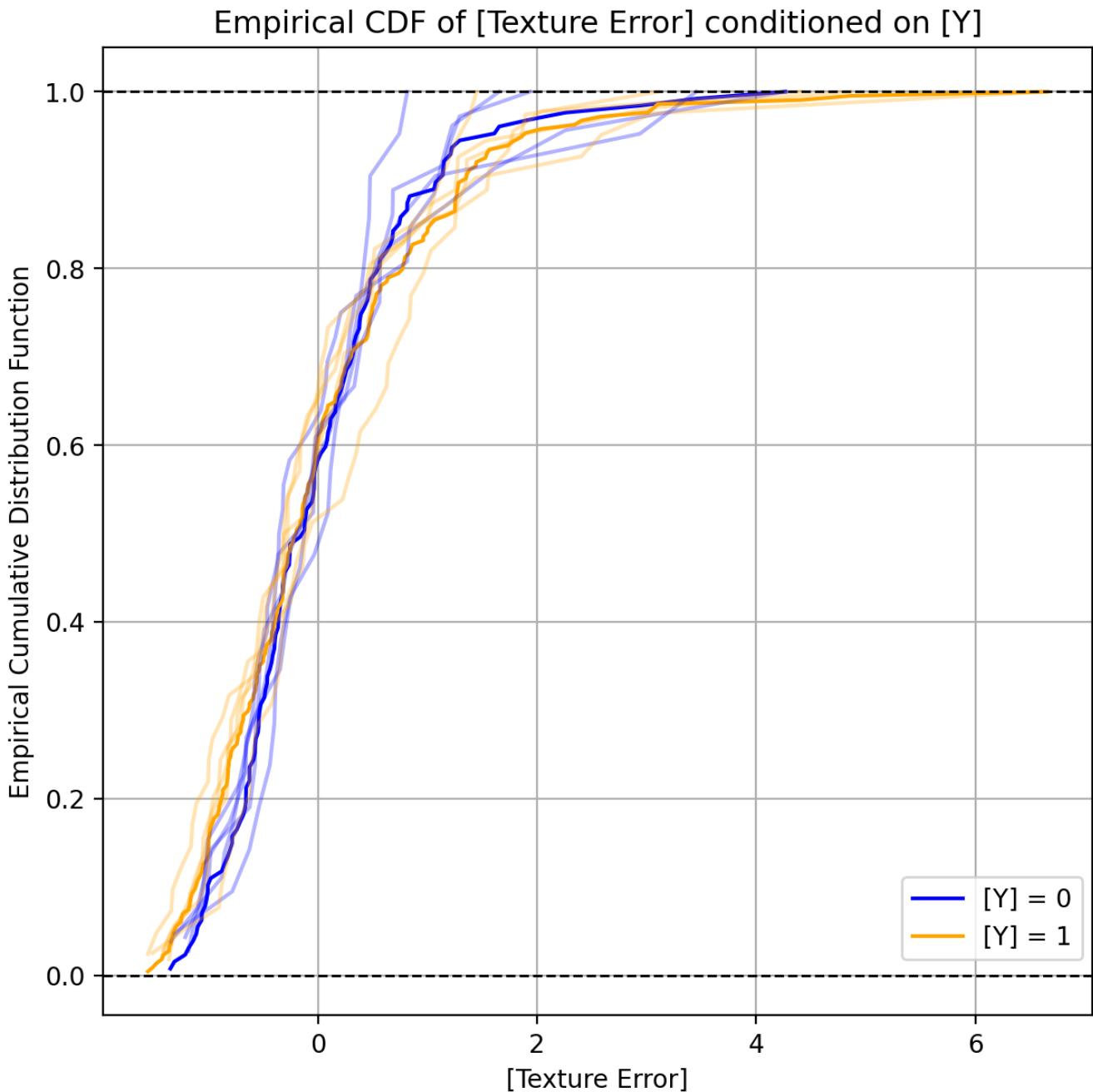
Kernel Density Plot of [Texture Error] by [Y].
Distributions by level are not significantly different at the 95% level.



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

Univariate Report

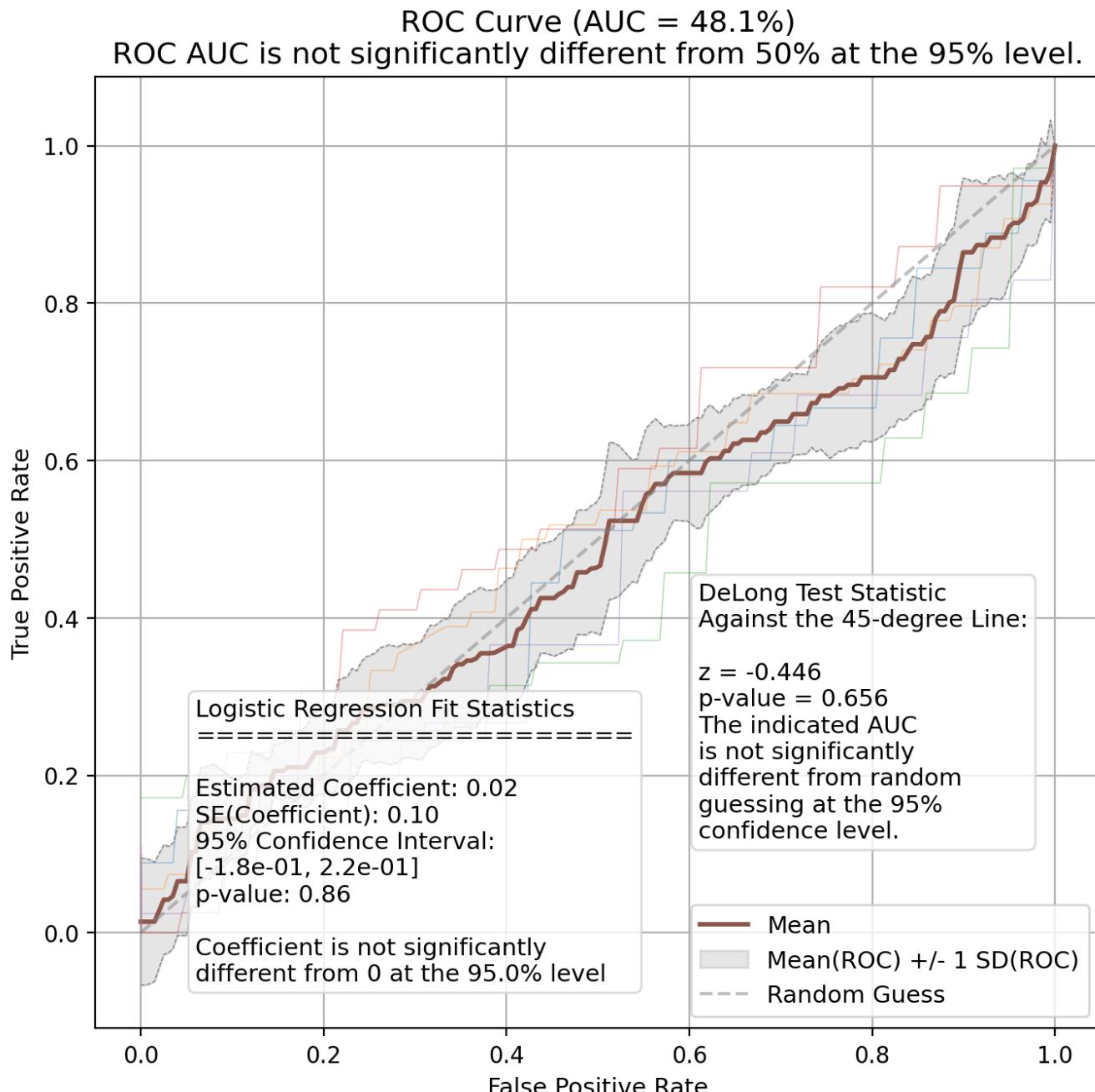
Texture Error - Empirical CDF Plot



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

Univariate Report

Texture Error - ROC Curve

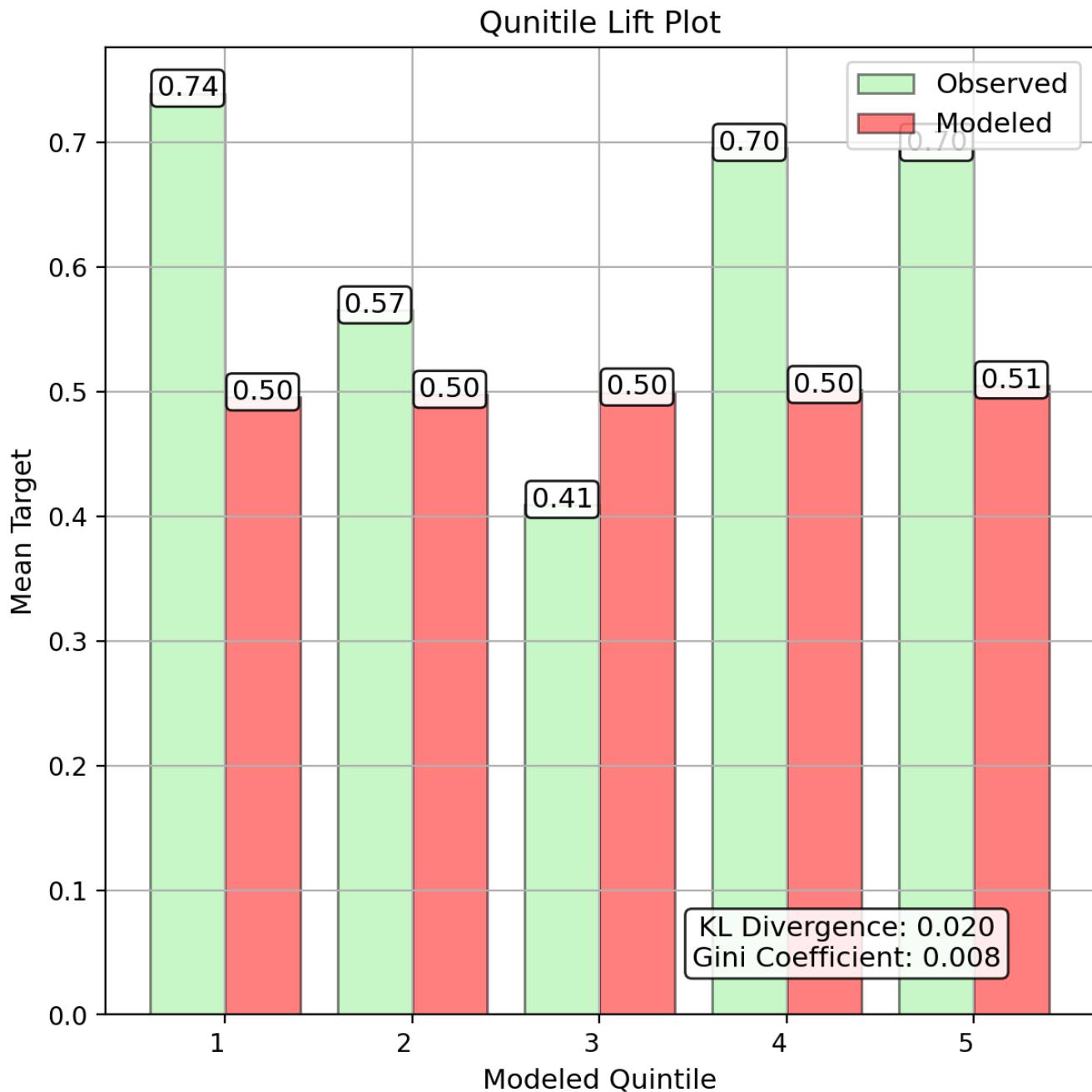


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Texture Error - Quintile Lift



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

Univariate Report

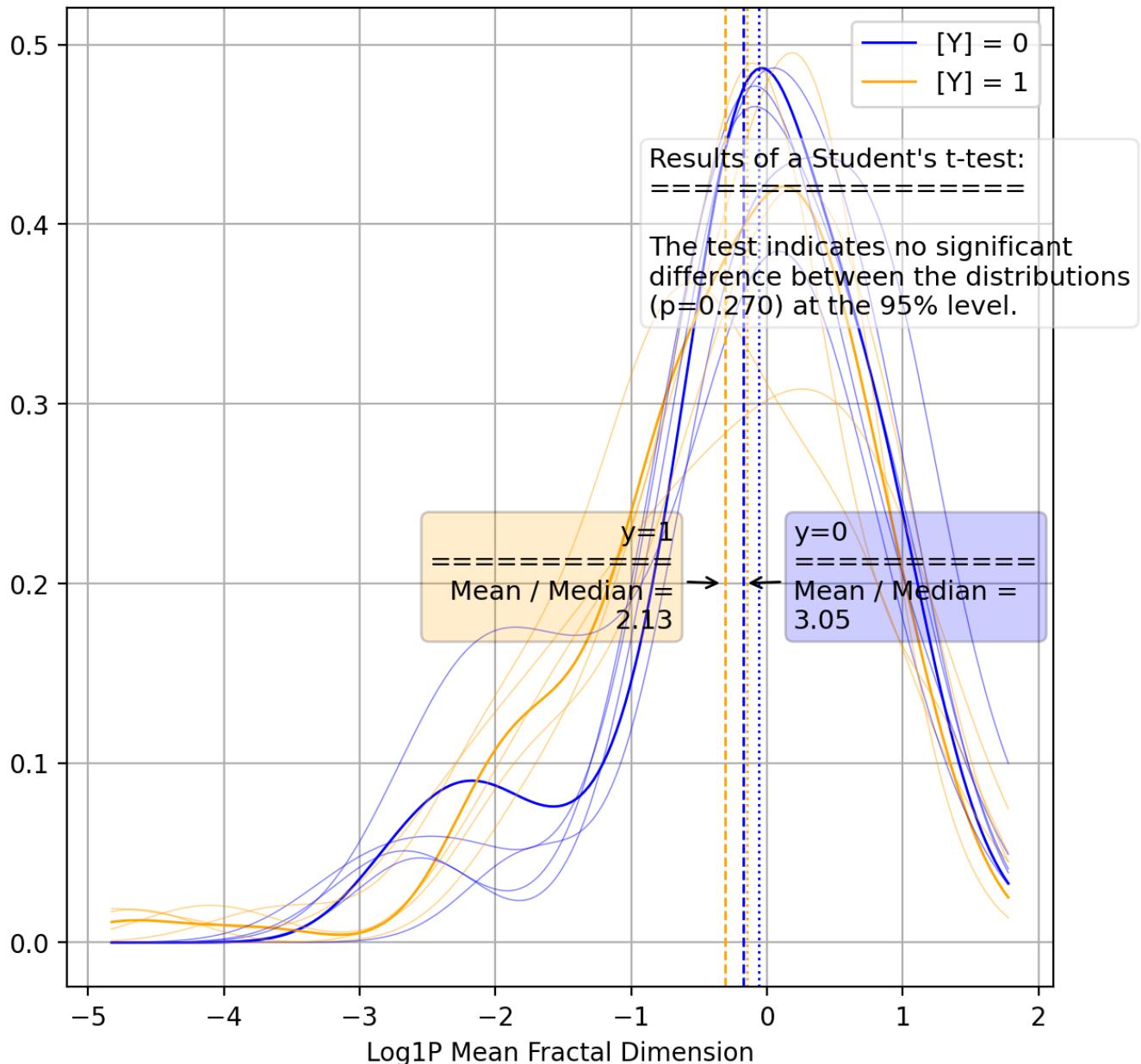
Log1P Mean Fractal Dimension - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-0.205	1.1e-01	0.129	-0.457	4.8e-02	61.7%	60.7%	69.3%	76.5%	63.4%	20.0%
Fold-2	-0.440	2.7e-03	0.147	-0.727	-1.5e-01	45.8%	46.9%	47.1%	55.6%	40.8%	-6.2%
Fold-3	-0.244	5.8e-02	0.129	-0.496	8.5e-03	50.0%	52.7%	55.6%	71.4%	45.5%	5.1%
Fold-4	-0.209	8.5e-02	0.122	-0.447	2.9e-02	53.7%	61.1%	52.8%	82.4%	38.9%	22.6%
Fold-5	-0.271	3.2e-02	0.127	-0.520	-2.3e-02	46.4%	51.8%	48.3%	70.0%	36.8%	3.4%
mean	-0.267	2.1e-02	0.116	-0.494	-4.0e-02	52.9%	55.6%	55.9%	70.5%	46.3%	10.8%
std	0.097	4.3e-02	0.009	0.114	8.0e-02	6.5%	6.1%	8.9%	10.0%	10.7%	12.1%

Univariate Report

Log1P Mean Fractal Dimension - Kernel Density Plot

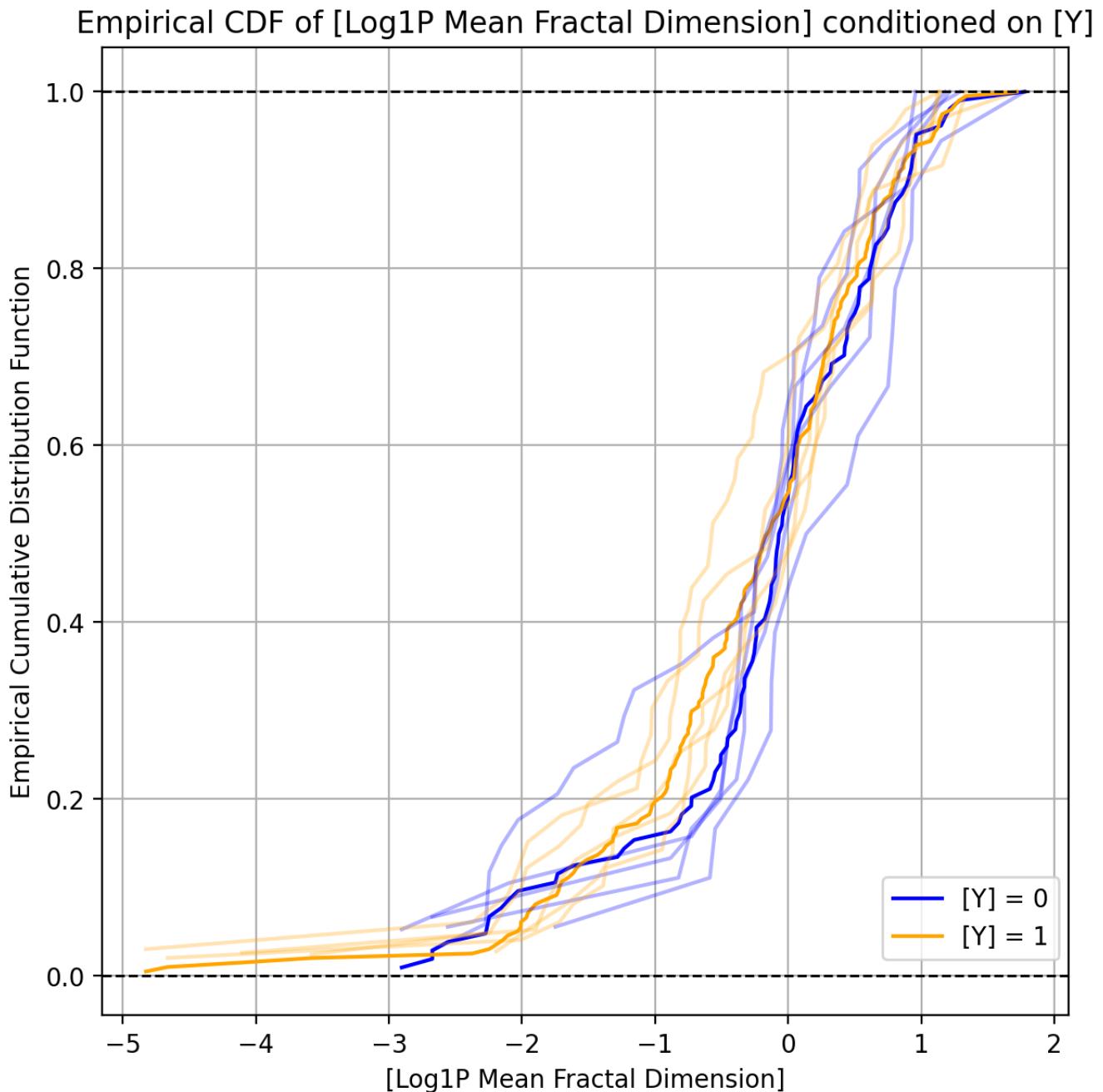
Kernel Density Plot of [Log1P Mean Fractal Dimension] by [Y].
Distributions by level are not significantly different at the 95% level.



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

Univariate Report

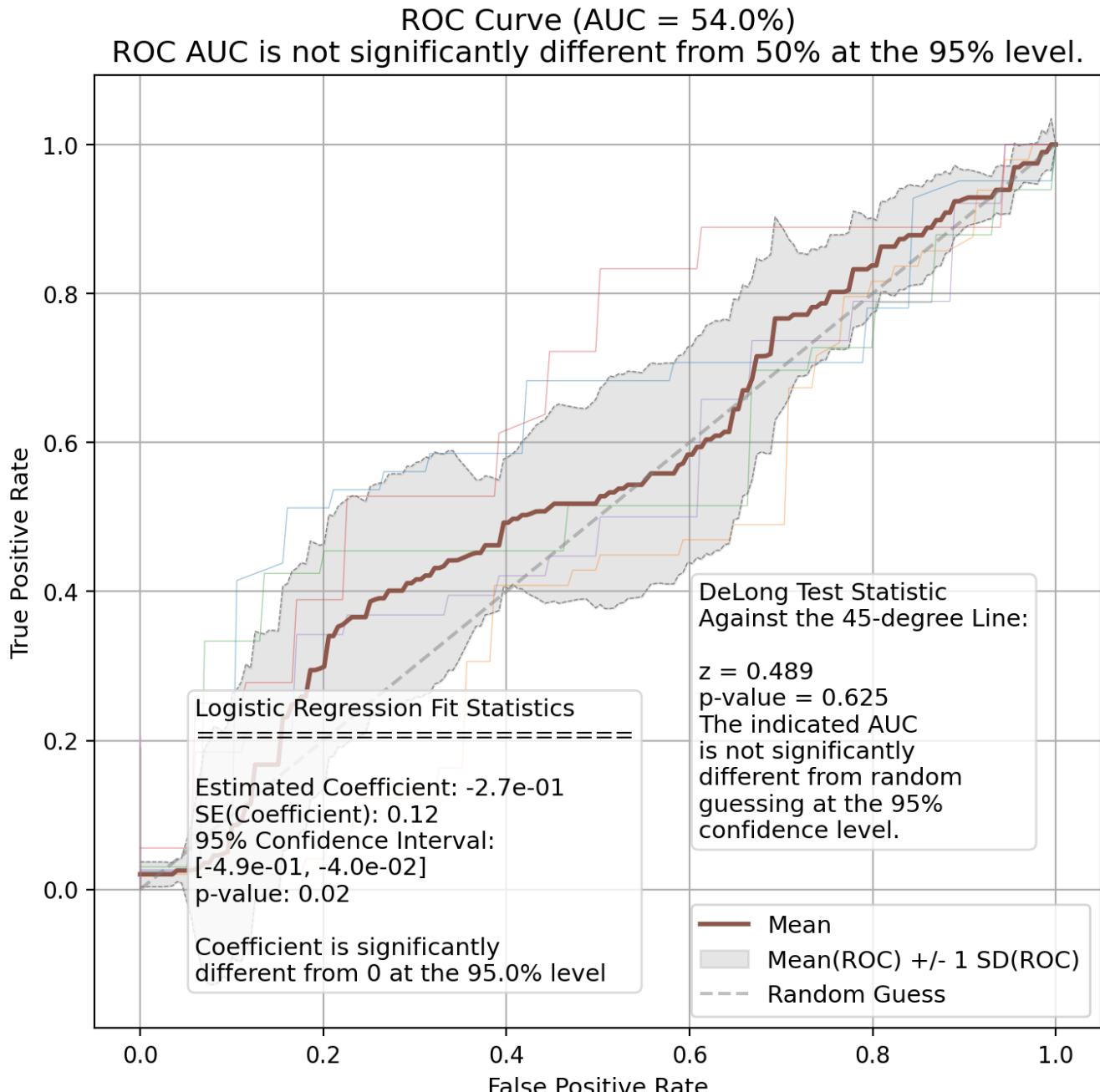
Log1P Mean Fractal Dimension - Empirical CDF Plot



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

Univariate Report

Log1P Mean Fractal Dimension - ROC Curve

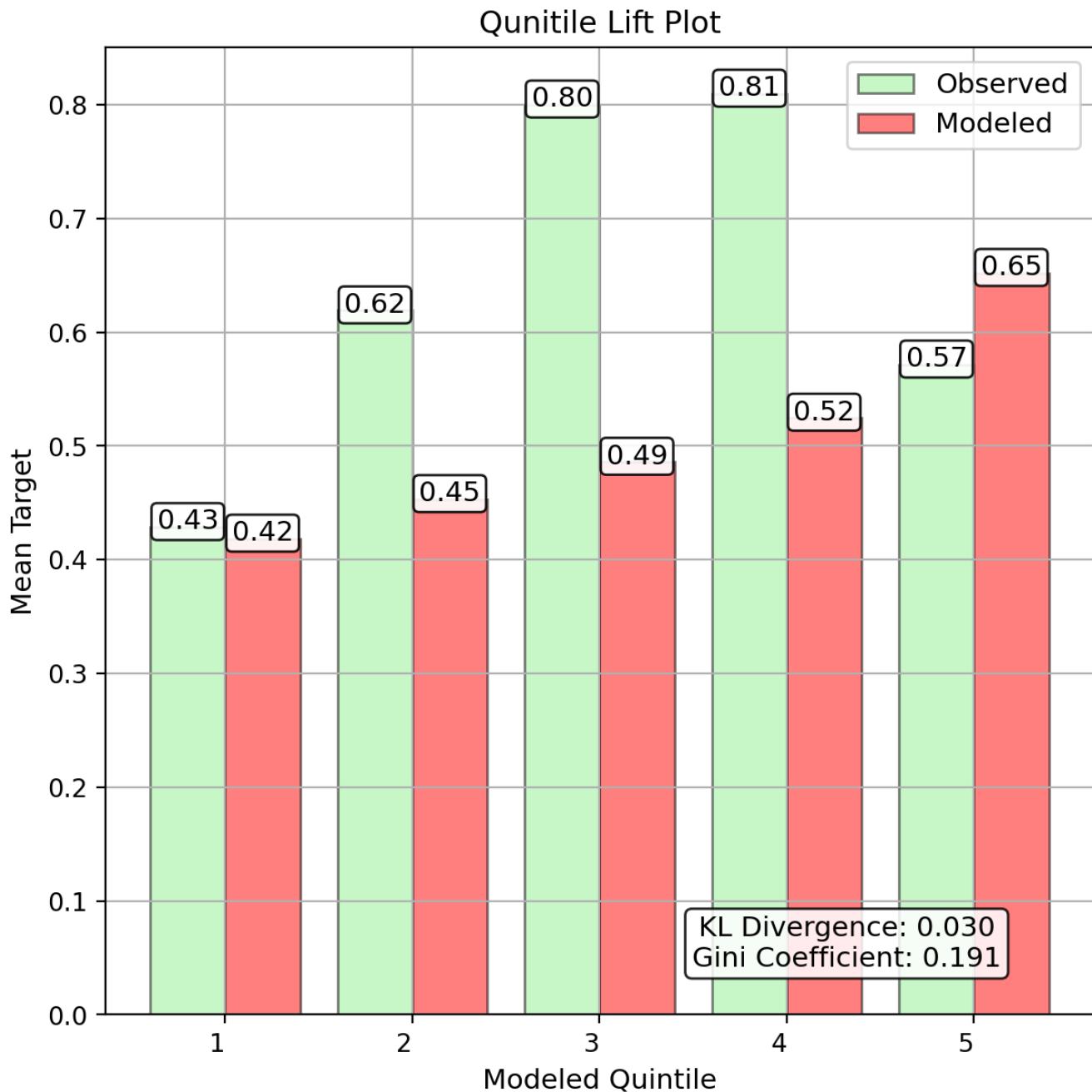


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Log1P Mean Fractal Dimension - Quintile Lift



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

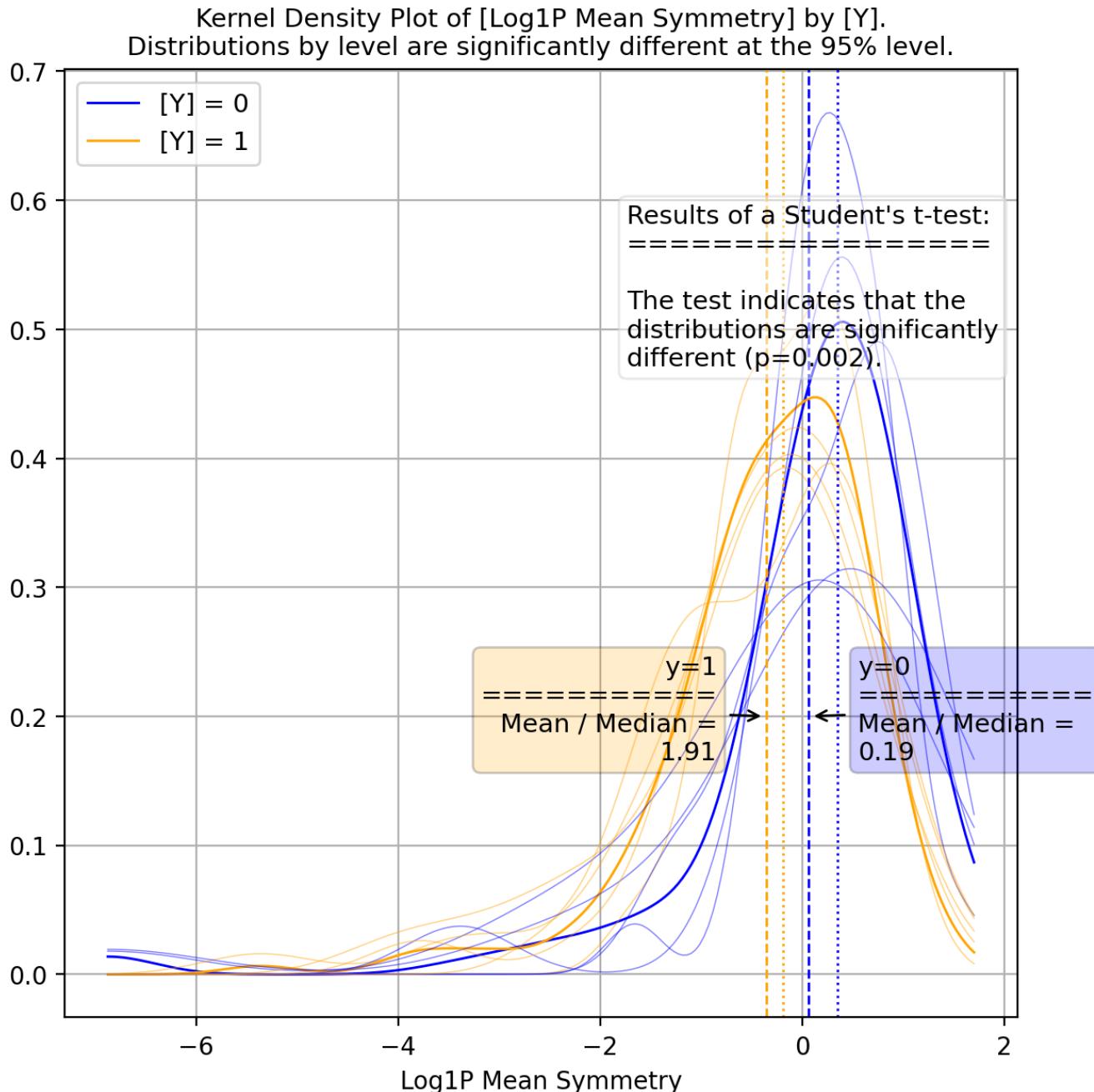
Univariate Report

Log1P Mean Symmetry - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	-0.708	3.2e-05	0.170	-1.042	-3.7e-01	48.3%	49.2%	48.3%	56.0%	42.4%	-1.6%
Fold-2	-0.259	5.2e-02	0.134	-0.521	2.6e-03	69.1%	70.8%	66.7%	83.3%	55.6%	42.9%
Fold-3	-0.355	9.8e-03	0.138	-0.625	-8.5e-02	67.4%	68.8%	66.7%	78.9%	57.7%	37.9%
Fold-4	-0.364	7.4e-03	0.136	-0.631	-9.8e-02	57.4%	60.0%	58.2%	72.7%	48.5%	19.8%
Fold-5	-0.553	4.5e-04	0.158	-0.863	-2.4e-01	61.5%	63.9%	61.5%	76.2%	51.6%	27.8%
mean	-0.435	9.2e-04	0.131	-0.692	-1.8e-01	47.5%	52.0%	39.1%	63.0%	28.3%	4.4%
std	0.180	2.2e-02	0.016	0.212	1.5e-01	8.4%	8.6%	7.6%	10.5%	6.0%	17.5%

Univariate Report

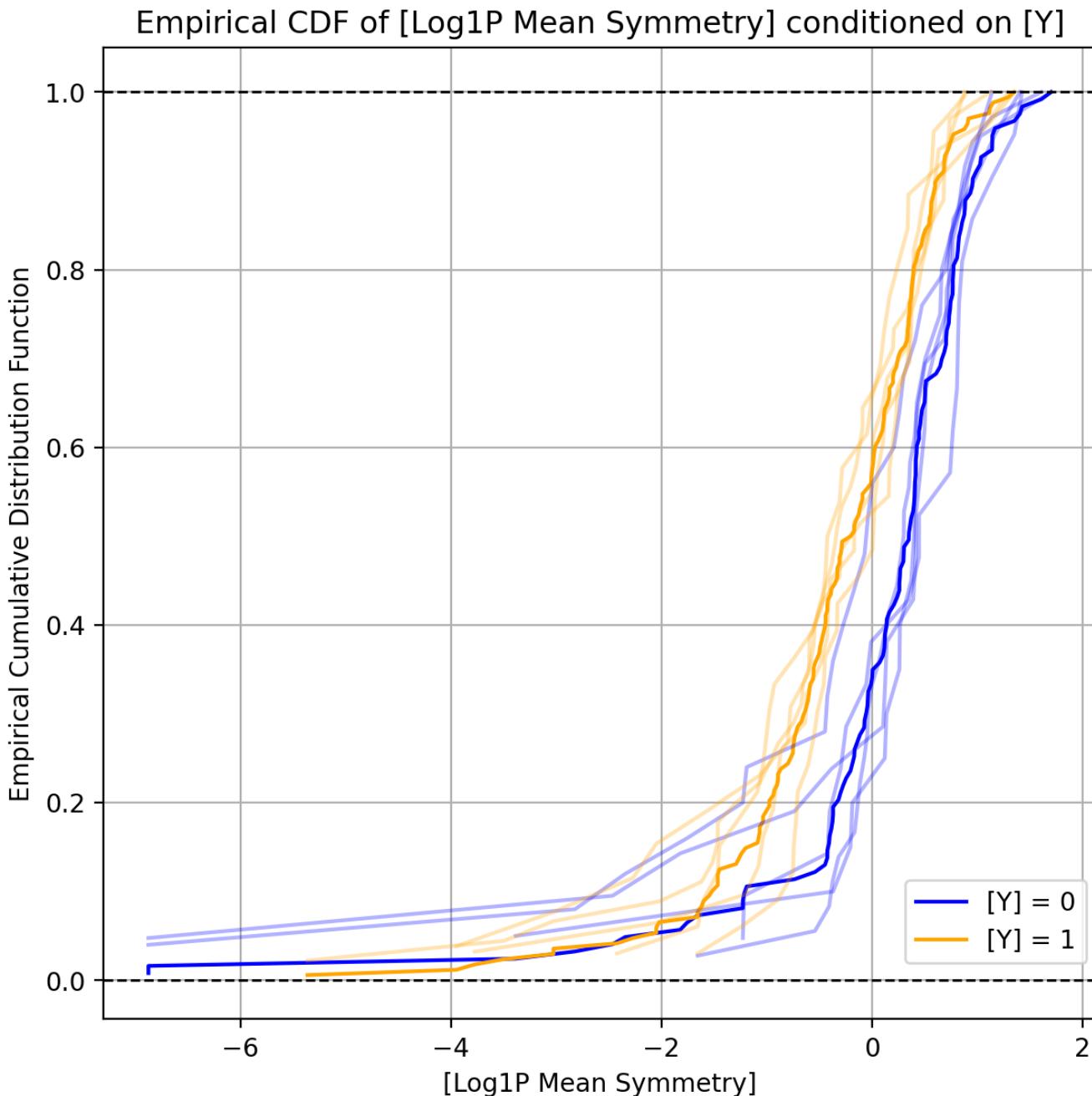
Log1P Mean Symmetry - Kernel Density Plot



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

Univariate Report

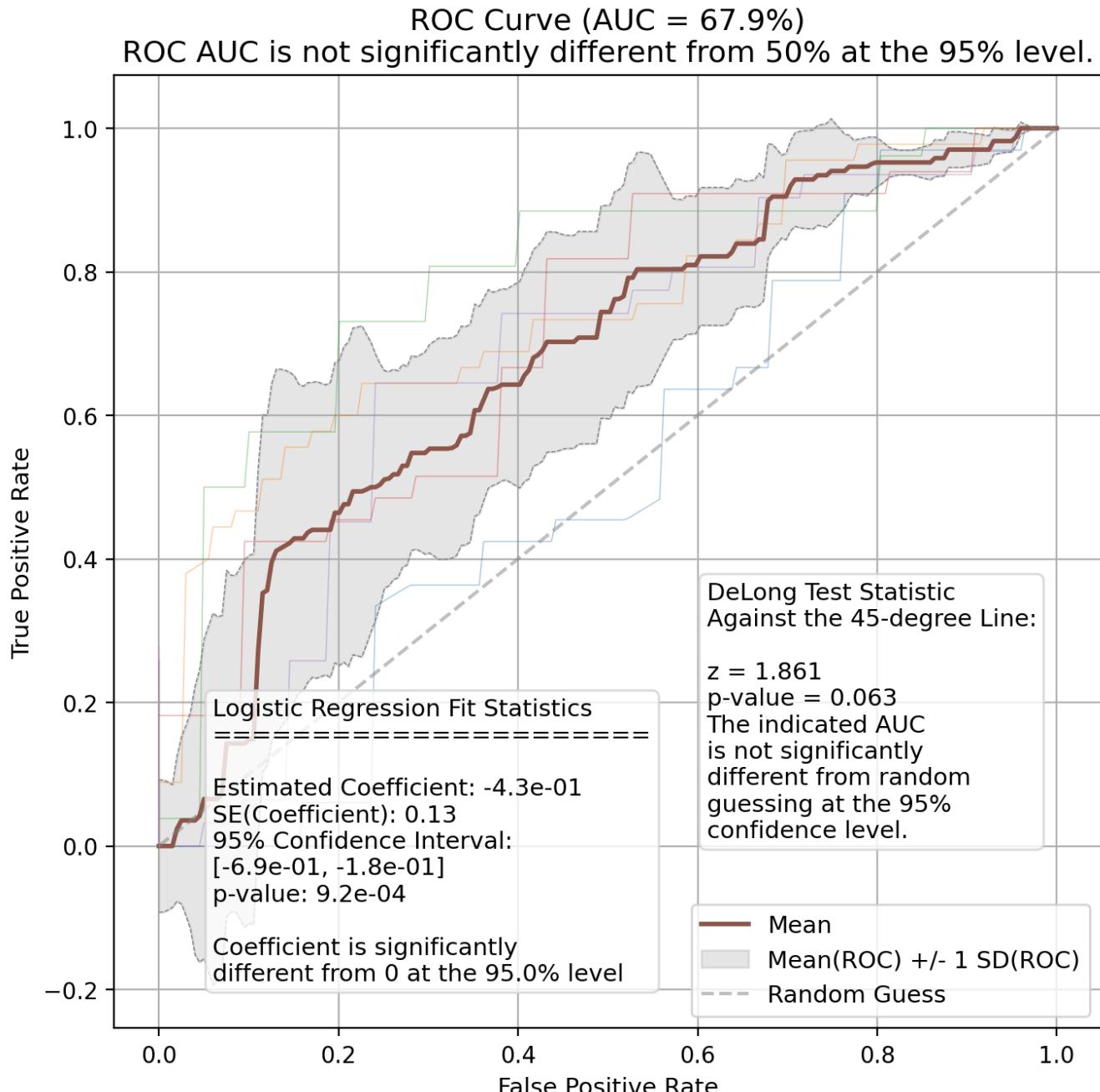
Log1P Mean Symmetry - Empirical CDF Plot



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

Univariate Report

Log1P Mean Symmetry - ROC Curve

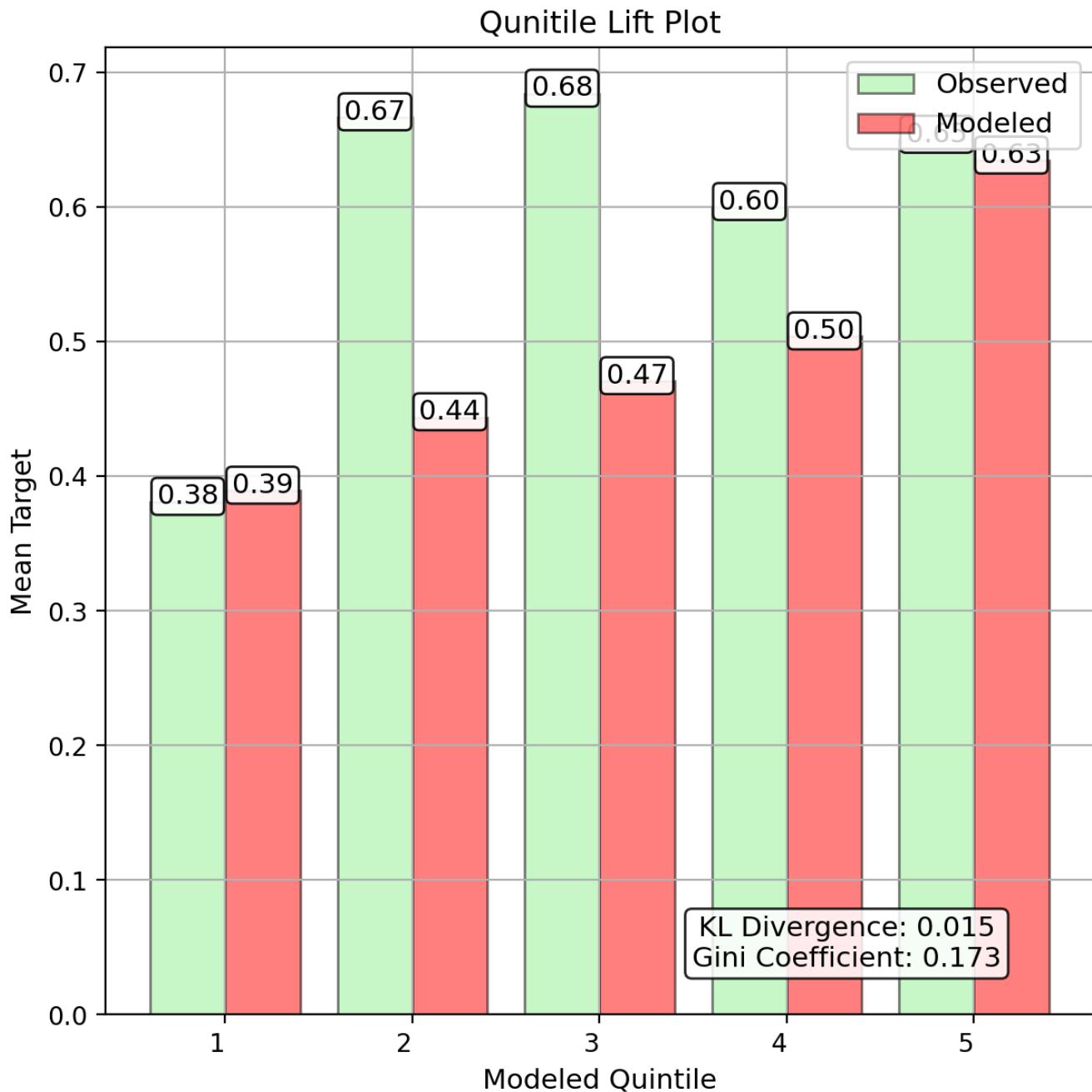


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Log1P Mean Symmetry - Quintile Lift



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

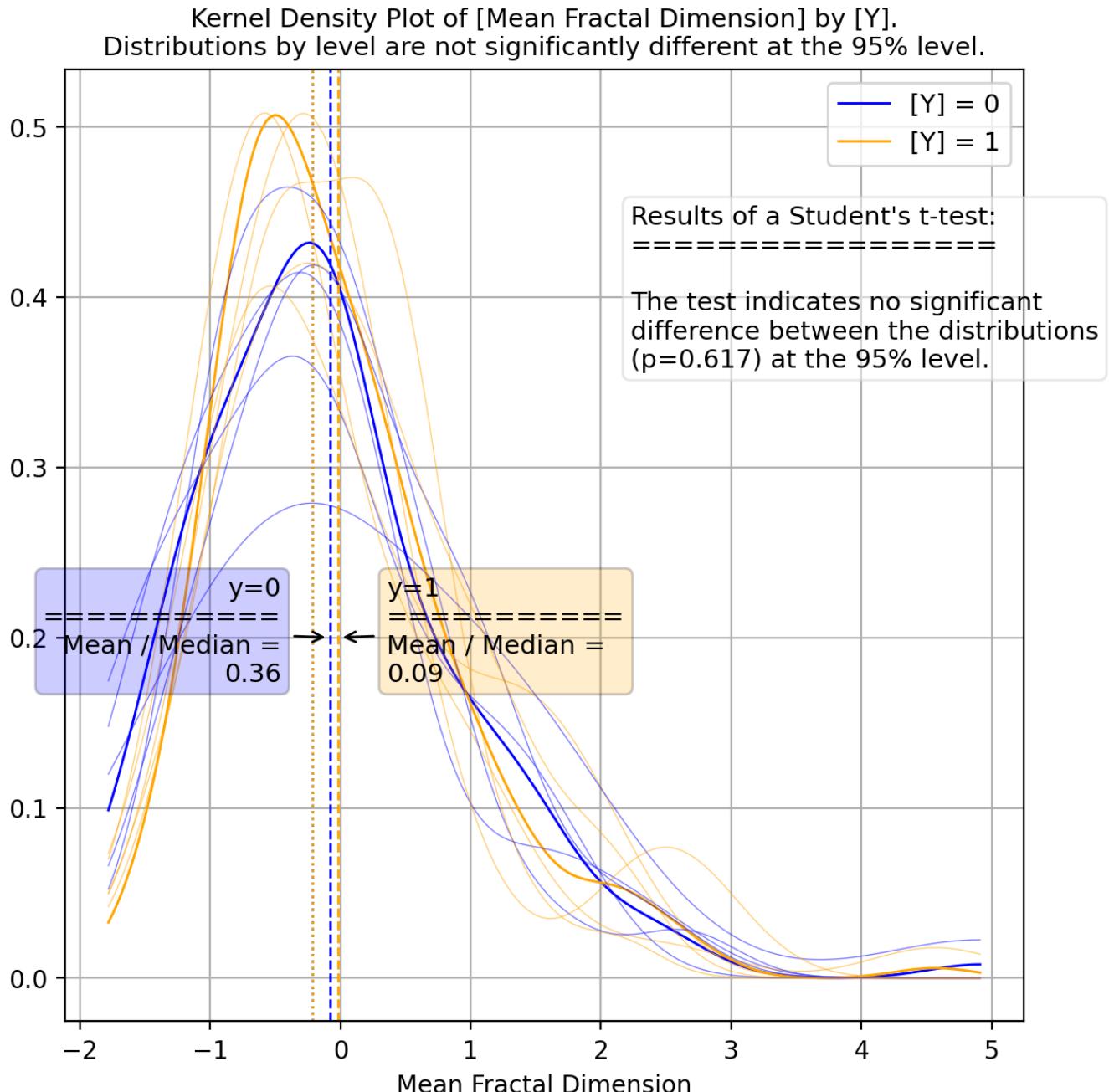
Univariate Report

Mean Fractal Dimension - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int. Lower	Conf. Int. Upper	Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
Fold-1	3.4e-02	0.793	0.129	-0.219	0.287	42.3%	47.1%	38.8%	59.1%	28.9%	-6.0%
Fold-2	3.9e-02	0.752	0.124	-0.203	0.281	54.4%	56.0%	55.9%	66.7%	48.1%	11.9%
Fold-3	-1.5e-02	0.902	0.124	-0.257	0.227	46.4%	43.8%	55.9%	57.6%	54.3%	-12.2%
Fold-4	9.9e-02	0.454	0.132	-0.160	0.358	46.8%	47.9%	50.7%	60.7%	43.6%	-4.1%
Fold-5	1.9e-02	0.873	0.122	-0.220	0.259	53.2%	53.0%	60.3%	68.8%	53.7%	5.7%
mean	3.4e-02	0.765	0.113	-0.187	0.254	44.7%	46.0%	47.9%	58.0%	40.8%	-7.8%
std	4.1e-02	0.179	0.004	0.035	0.048	5.1%	4.9%	8.3%	4.9%	10.4%	9.6%

Univariate Report

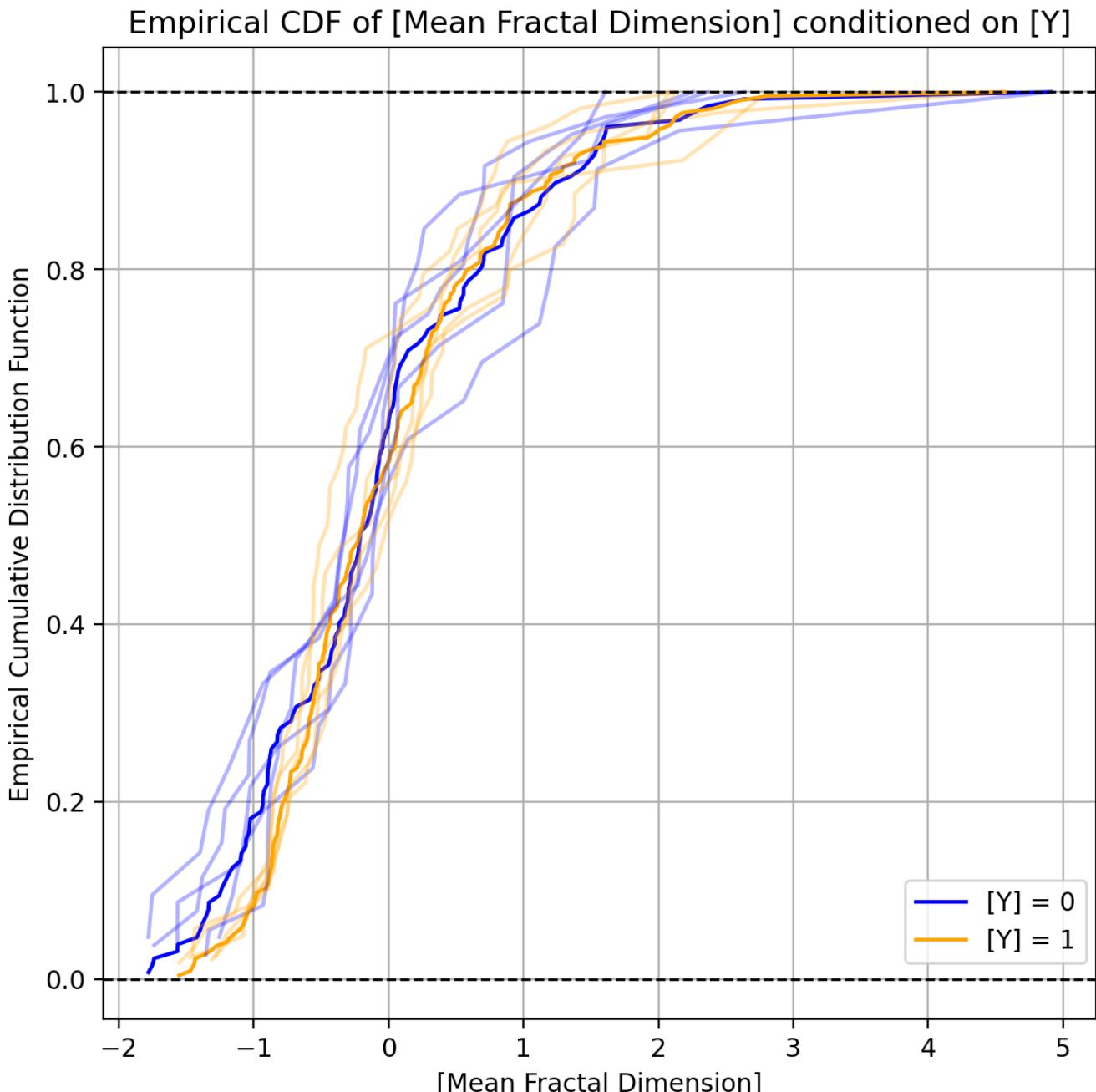
Mean Fractal Dimension - Kernel Density Plot



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

Univariate Report

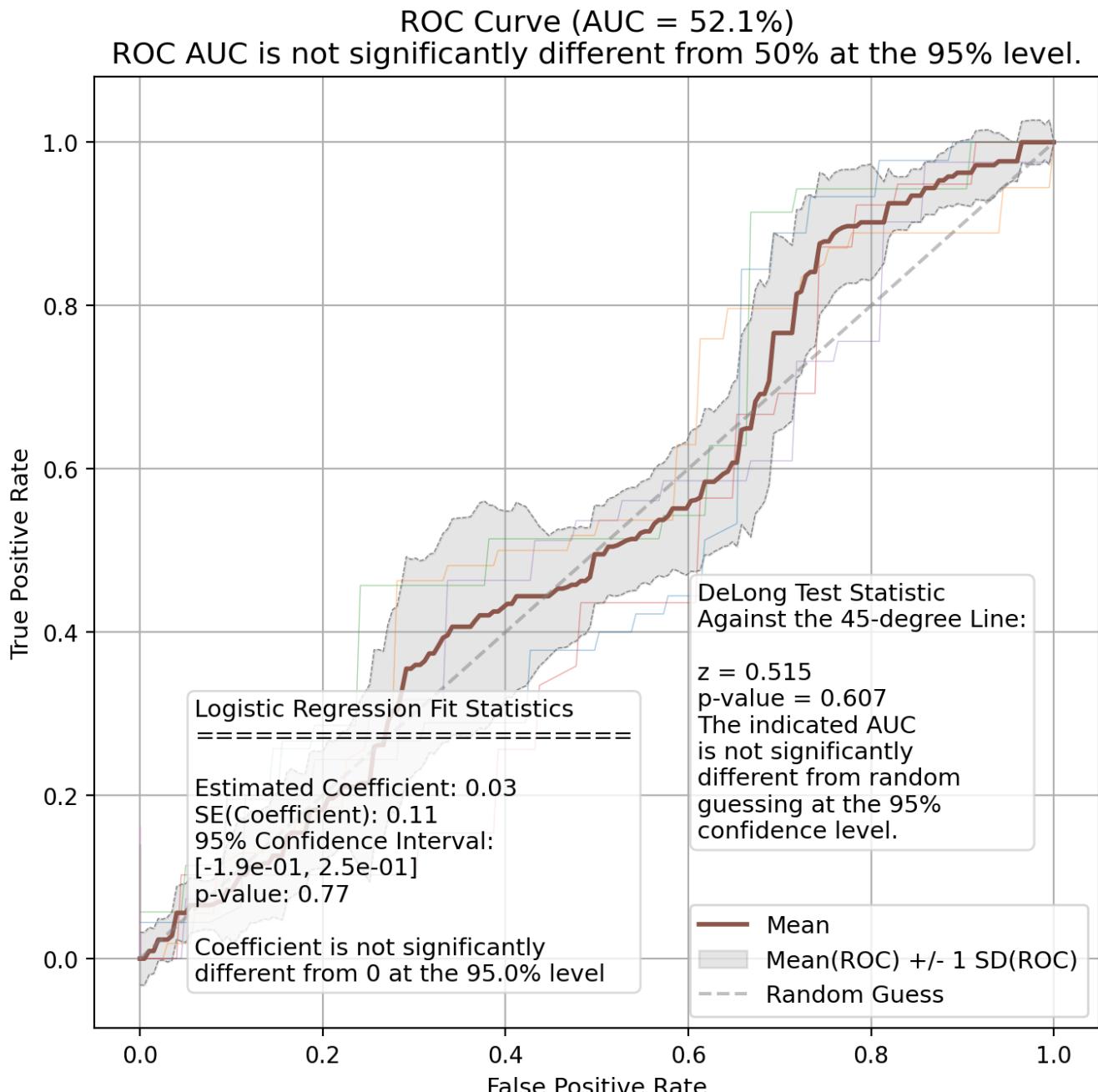
Mean Fractal Dimension - Empirical CDF Plot



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

Univariate Report

Mean Fractal Dimension - ROC Curve

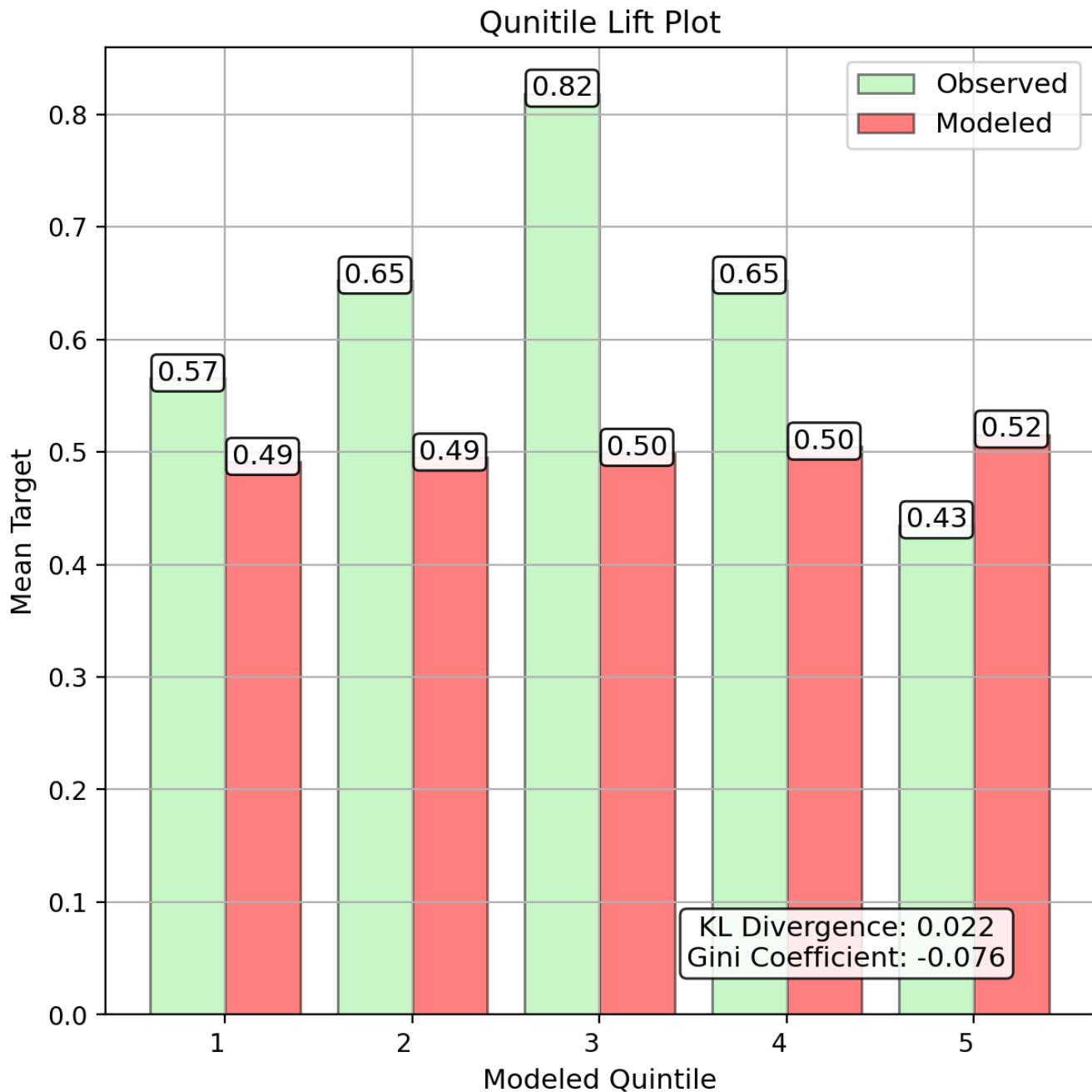


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Mean Fractal Dimension - Quintile Lift



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

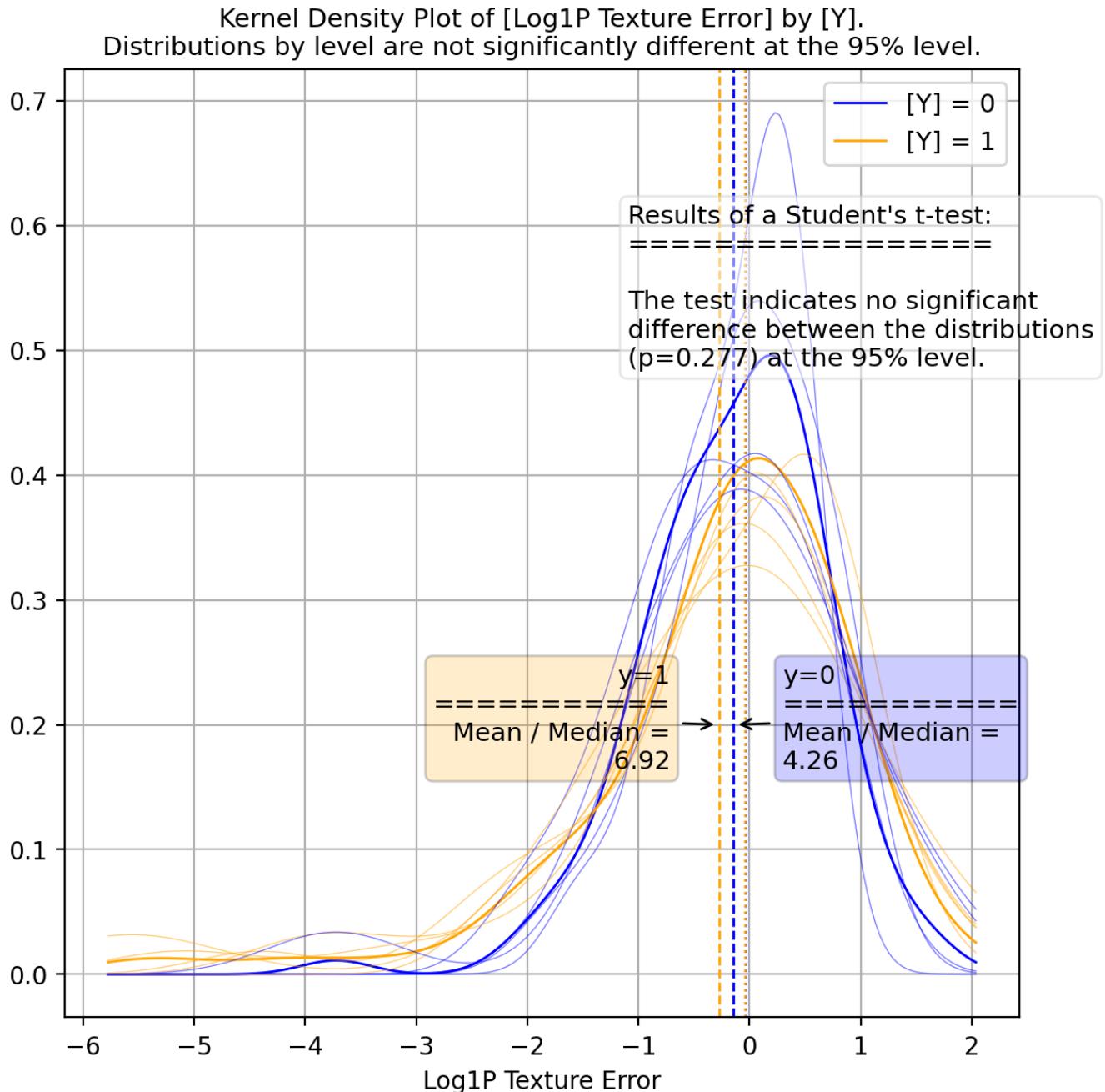
Univariate Report

Log1P Texture Error - Results

	Fitted Coef.	Fitted p-Value	Fitted Std. Err.	Conf. Int.		Test Accuracy	Test AUC	Test F1	Test Precision	Test Recall	Test MCC
				Lower	Upper						
Fold-1	-0.162	1.7e-01	0.119	-0.394	7.1e-02	46.7%	51.2%	44.8%	65.0%	34.2%	2.4%
Fold-2	-0.235	7.1e-02	0.130	-0.490	2.0e-02	43.0%	43.4%	45.8%	51.4%	41.3%	-13.1%
Fold-3	-0.137	2.6e-01	0.121	-0.374	1.0e-01	52.9%	54.2%	55.6%	65.2%	48.4%	8.2%
Fold-4	-0.202	8.5e-02	0.117	-0.433	2.8e-02	47.4%	50.3%	50.0%	65.2%	40.5%	0.5%
Fold-5	-0.212	8.8e-02	0.124	-0.456	3.1e-02	49.0%	50.8%	51.9%	63.6%	43.8%	1.6%
mean	-0.189	8.3e-02	0.109	-0.402	2.5e-02	43.6%	45.2%	43.6%	53.7%	36.7%	-9.7%
std	0.040	7.9e-02	0.005	0.046	3.4e-02	3.6%	4.0%	4.4%	6.0%	5.2%	7.9%

Univariate Report

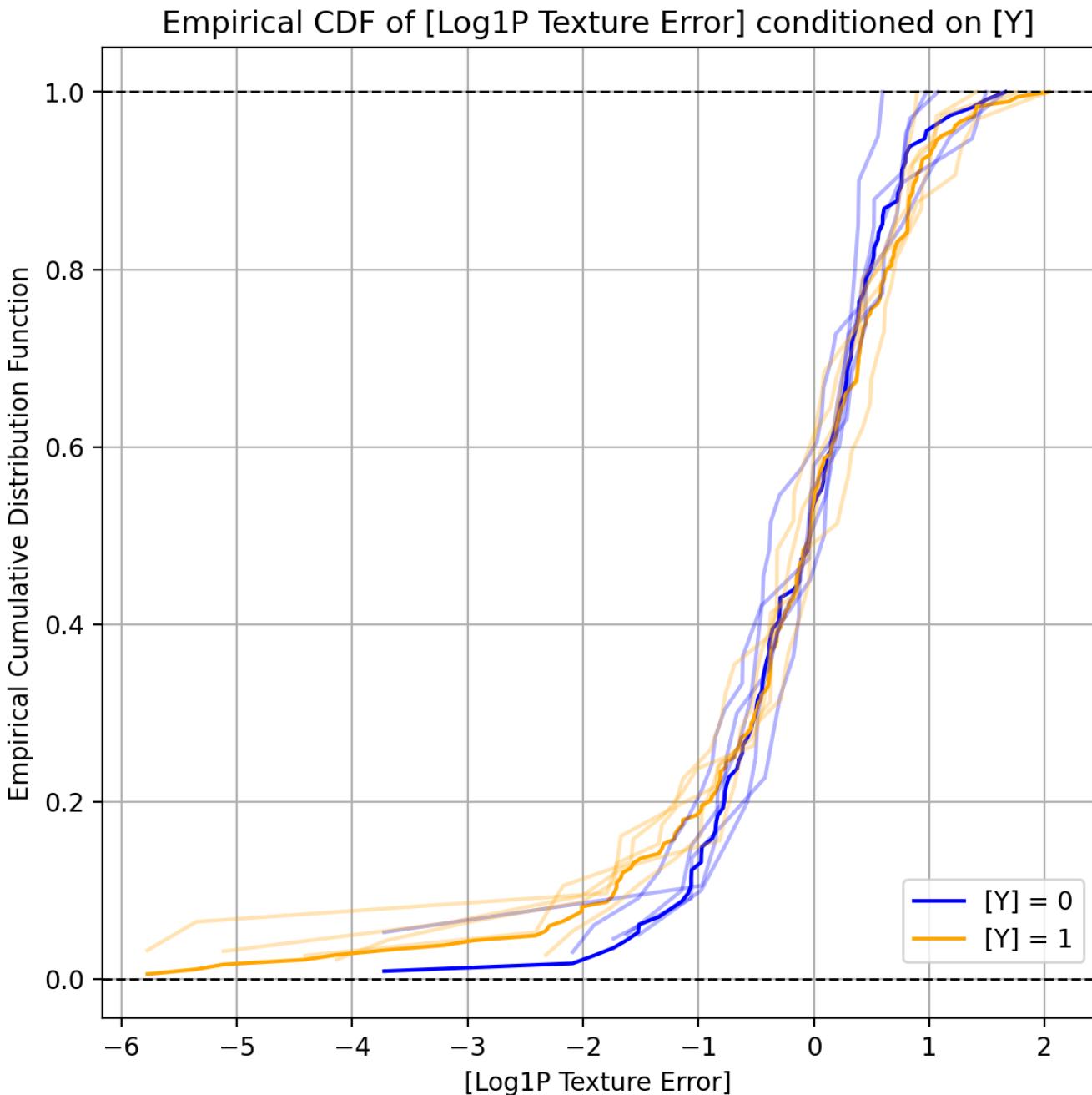
Log1P Texture Error - Kernel Density Plot



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

Univariate Report

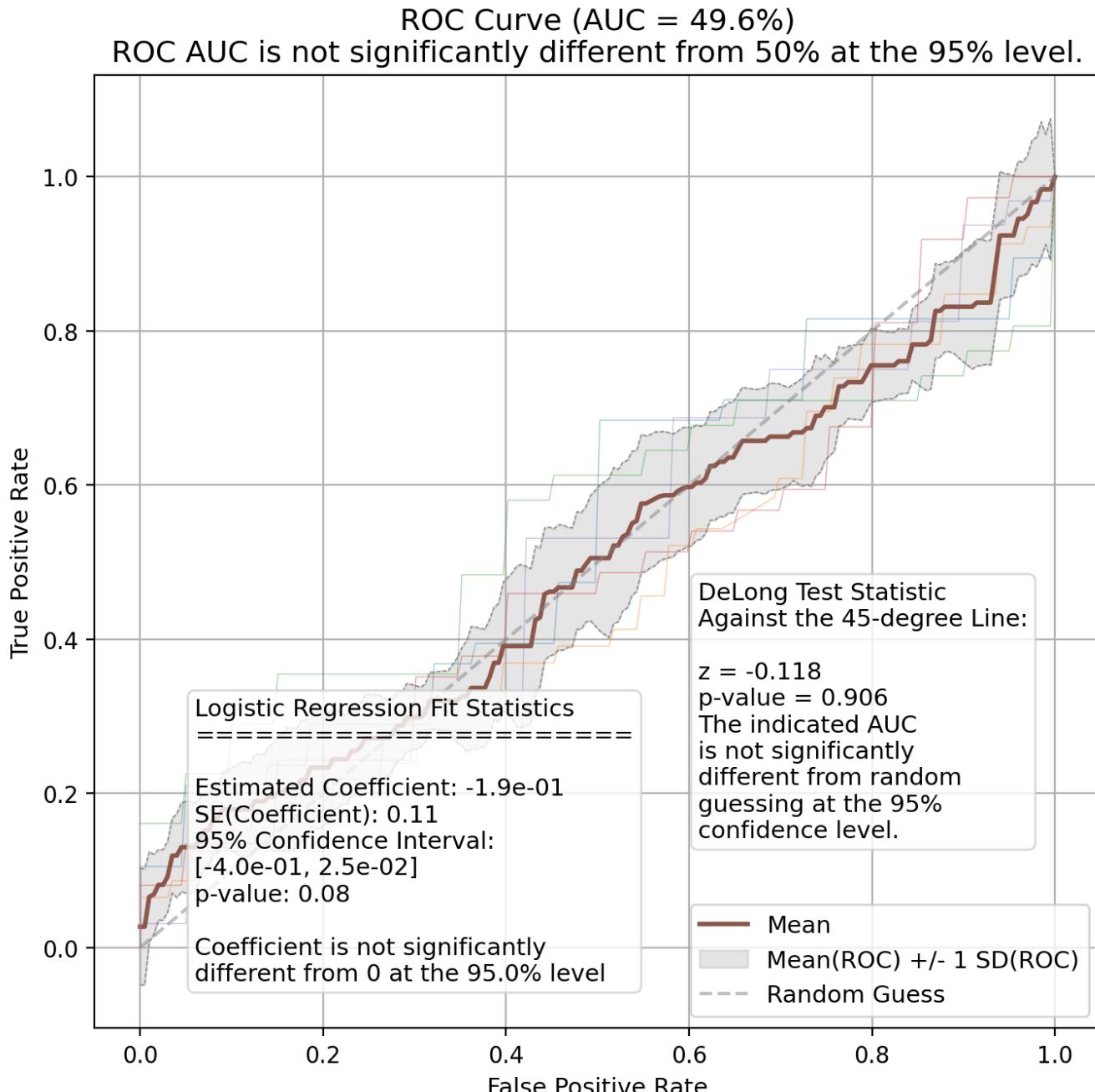
Log1P Texture Error - Empirical CDF Plot



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

Univariate Report

Log1P Texture Error - ROC Curve

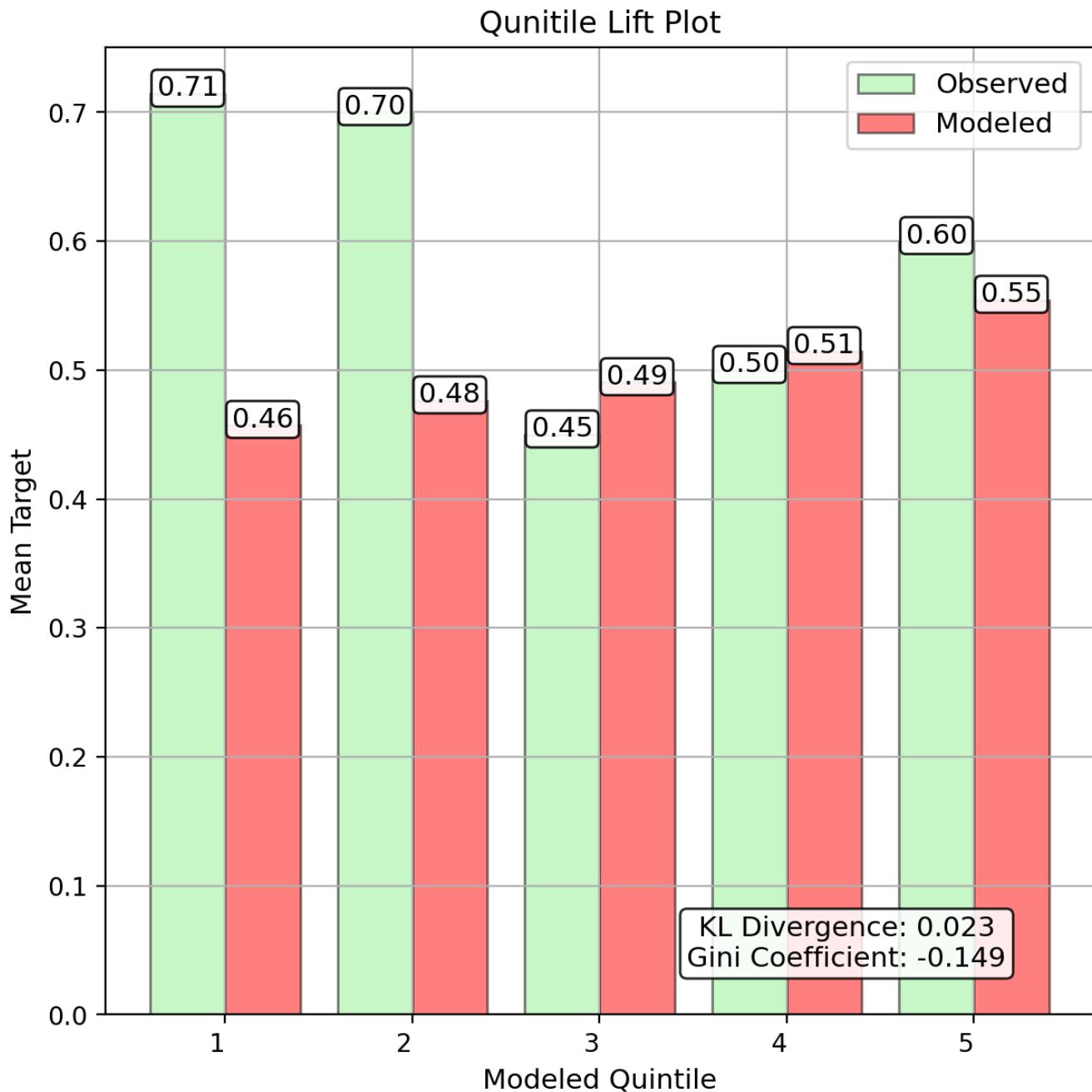


This plot shows the receiver operating characteristic (ROC) curve for the target variable in total and for each fold. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. This is based on a simple Logistic Regression model with no regularization, no intercept,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

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

Univariate Report

Log1P Texture Error - Quintile Lift



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