# MSCA31010: Linear & Non-Linear Models Winter Quarter 2023 Assignment 3

Question 1 (20 points)

1. (10 points) For each categorical predictor,
   * Generate a vertical bar chart that shows the odds of Churn for each category.
   * Display the categories in the order of descending odds of Churn.
   * Add a reference line to indicate the overall odds of Churn.
   * Comment on whether it may affect the target variable.

Chart, bar chart

Description automatically generatedChart, bar chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

1. (10 points). For each interval predictor,
   * Generate a horizontal boxplot grouped by the target categories.
   * Add a reference line to indicate the overall mean of the interval predictor.
   * Comment on whether it may affect the target variable.

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Question 2 (30 points)

1. Please provide a summary report of the Backward Selection. The report should include (1) the step number, (2) the predictor removed, (3) the number of non-aliased parameters in the current model, (4) the log-likelihood value of the current model, (5) the Deviance Chi-squares statistic between the current and the previous models, (6) the corresponding Deviance Degree of Freedom, and (7) the corresponding Chi-square significance.

Table

Description automatically generated

1. (10 points). Please show a table of the complete set of parameters of your final model (including the aliased parameters). Besides the parameter estimates, please also include the standard errors, and the 95% asymptotic confidence intervals. Conventionally, aliased parameters have missing or zero standard errors and confidence intervals.

Table

Description automatically generated

1. What is the predicted probability of Churn for a customer with the following profile? Contract One year is Month-to-month, Dependents is No, Gender is Male, InternetService is Fiber optic, MultipleLines is No phone service, PaperlessBilling is Yes, Partner is No, PhoneService is No, SeniorCitizen is Yes, MonthlyCharges is 70, Tenure is 29, and TotalCharges is 1400.

Predicted probability of Churn: **0.6073319944090094**

Question 3 (30 points)

1. (10 points). What is the McFadden’s R-squared, the Cox-Snell’s R-squared, the Nagelkerke’s Rsquared, and the Tjur’s Coefficient of Discrimination?

McFadden's R-squared : **0.27057873690192324**

Cox-Snell's R-squared : **0.26899990403441476**

Nagelkerke's R-squared : **0.39218554903069747**

Tjur's coefficient of discrimination : **0.2907638199477376**

1. (10 points). What is the Area Under Curve value?

Area Under Curve value : **0.8411964188949088**

1. (10 points). What is the Root Average Squared Error value?

Root Average Squared value : **0.3708047029352164**

Question 4 (20 points)

1. (10 points). Please generate the Kolmogorov-Smirnov Chart. What is the Kolmogorov-Smirnov statistic and the corresponding probability threshold for Churn? What is the misclassification rate if we use this probability threshold?

Chart, line chart

Description automatically generated

Kolmogorov-Smirnov statistic : **0.5244664390313967**

Corresponding probability threshold for Churn : **0.2579355595500369**

Misclassification rate : **0.26493174061433444**

1. (10 points). Please generate the properly labelled Precision-Recall chart with a No-Skill line. According to the F1 Score, what is the probability threshold for Churn? What is the misclassification rate if we use this probability threshold?

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

F-1 Score : **0.6274157303370785**

Corresponding probability threshold for Churn : **0.32902080792001165**

Misclassification rate : **0.23577929465301478**