# STA4003 Project Report

Yuzhou Peng

121090446

The project is divided into 3 major components: data preparation, model fitting and result output.

1. Data preparation

(1) Combine the data from 2014 to 2017 in \*\*Train\_all\*\*

(2) Extract data of 1st, 2nd, 3rd .......11th race of the racing days and form data \*\*train\_all\_i\*\* and \*\*test\_i\*\* where i = 1,2,3....11

1. Model fittting

The model includes 2 models:

(1) a linear model with \*\*Csum\*\* , \*\*WIN\_POOL.x\*\* being the response variable (y) and explanatory variable (x).

(2) a vector autoregression (\*\*VAR\*\*) model

For data of i-th race, we fit the model as follow.

Step 1: Fit a \*\*TSLM\*\* model to the first 37.5% days (rounded by \*\*floor()\*\*) of the i-th races and predict the first 30 days data in the test set. If the test set have fewer days then 30, predict all data using this model.

Step 2: Fit a vector autoregression model to the first j (j > 30) days in test set and predict data on day j+1. And thus iteratively generate all prediction of data after 30th day.

(3) The result is stored in \*\*result\_i\*\* for i-th race. \*\*result\_all\*\* is the combined set of all \*\*result\_i\*\*, containing \*\*true\_value\*\*, \*\*forecast\_value\*\*, \*\*upper\_quantile\*\* (0.95 quantile) and \*\*APE\*\* (absolute percentatage error).

1. Result output

(1) MAPE is calculated by the mean value of \*\*result\_all$APE\*\*.

(2) 0.95-quantile score is stored in \*\*QS\*\*.

Note: The calculation of QS is based on the function \*\*quantile\_score\*\* (consisting with our textbook) with input containing 0.95 quantile (\*\*upper\_quantile\*\*) and true value of Csum (\*\*true\_value\*\*)

