Assignment Instructions

**To complete the assignment, you need to complete the following steps:**

**Step 1: Data preparation**

1. Load the data
2. Define variables for the probabilities of a home win and away win associated with bookmaker odds
3. Define a dummy variable = 1 if the home team loses, and zero otherwise
4. Define a variable equal to H if the home team wins and A if the visiting team wins
5. Define a variable equal to H if the home team win probability is greater than 0.5 according to the bookmaker odds and A otherwise
6. Define a variable equal to the logarithm of the ratio of the home team salaries to the visiting team salaries.

**Step 2: Estimate a logit model of home time wins depending on the log salary ratio, using the data for calendar year 2018 as the “training data”.**

1. Define a subset for the calendar year 2018 data
2. Import the logistic regression package (copy the code for this from Week 1)
3. Run the logistic regression of hwin on the log salary ratio (copy the code for this from Week 1 while changing the variable names to the ones required here)

**Step 3: Define the predicted probabilities and the predicted results, using the entire data set**

1. The predicted probability of home win can be defined using the formula 1/(1+1/(exp(b0 + b1 (logsalaryratio)))) where b0 is the constant (the intercept) in the logistic regression and b1 is the coefficient for logsalaryratio.
2. Based on the predicted probability, define the predicted result H as the outcome where the predicted home win probability is greater than 0.5, and A otherwise.

**Step 4: For games played in 2019, compare the bookmaker probabilities and model probabilities in terms of the mean number of successfully predicted outcomes and the Brier scores.**

1. Define the subset of games played in calendar year 2019
2. Define a dummy variable equal to 1 when the bookmaker result prediction is correct, and zero otherwise. Define the equivalent variable for the logit model prediction.
3. Calculate the means for each of these variables
4. Define the Brier score for the bookmaker probabilities and the Brier score for the logit model probabilities
5. Calculate the mean of each Brier score

**Beware**: even though your code might get you to the correct answer at a given point, it is sometimes possible that the way you write it might interfere with completing a further step. So even if you get the answer right, you should look at the code we supply to check if you are going the same way. In practice, there are often many ways to get to answer in Python, and we do not insist that you follow our approach exactly – but simply warn you to be aware that differences could turn out to be problematic later.