# PSTAT 174/274 Fall 2024 – Lab Assignment 4

Due by midnight (11:59pm) on November 27(Wednesday), 2024.

**1.Modelling Global Temperature** We consider the *gtemp_ocean* data from the *astsa* package. Our aim is to fit an appropriate ARIMA$(p, d, q)$ model and to evaluate our model's performance using residual diagnostics before constructing a 10 year ahead forecast.

1. Using the following code to load the data and produce a time series plot. Is the data stationary? Comment on any possible linear or seasonal trends and how best to remove them.

```
library(astsa)
plot.ts(gtemp_ocean)
```

2. Difference the data lag 1 and produce a new time series plot. Does the data now appear stationary? You may want to use the function `diff(,)`.

3. Produce both an ACF and a PACF of the differenced data and comment on your observations. What potential ARIMA$(p, d, q)$ models do they suggest? (Hint: Recall that the $d$ parameter just indicates how many times we needed to differencing to obtain stationarity.)

4. This time use the `sarima()` function from the *astsa* package to fit your selected ARIMA model.

5. Using our final model produce a 20 year ahead forecast for global temperature using the `sarima.for()` function from *astsa* package. Comment on your forecast.

**2. Modelling AirPassengers** In this question we shall be analyzing the *AirPassengers* data set from the *astsa* package. Our aim is to determine an appropriate model from the SARIMA family.

1. Begin by producing a time series plot of the data using the `plot.ts()` function. Note your observations about any trends, seasonality and stationarity (is the variance constant?).

2. We investigate possible transformations to obtain a stationary time series on which we can consider potential SARIMA models. Use the following codes to compute 3 transformed time series:
   (1) log.data taking the natural log of the data;
   (2) dlog.data taking the difference (lag 1) of the log data; and
   (3) ddlog.data taking the difference (lag 12) of the differenced log data.

```
# computing transformed series
log.data <- log(AirPassengers)
dlog.data <- diff(log.data, 1)
ddlog.data <- diff(dlog.data, 12)
```

Create a matrix of all four series called **plot.data** using the `cbind()` function and use `plot.ts()` to produce our combined plot.

3. We continue our data exploration by producing both an ACF and PACF plot for our twice differenced log data using the `acf2()` function. Note your observations about the significant spikes in autocorrelation and partial autocorrelation.

4. Let us first consider a non-seasonal ARMA(1,1) model which we fit using the `sarima()` function (this time fitting the model to the twice differenced log data).

5. Use your ACF and PACF plot above to suggest possible parameters for the parameters SARIMA$(p, d, q)(P, D, Q)s$. Consider two potential models and fit each in turn using the `sarima()` function.

6. Produce a 24 month ahead forecast for our data using the `sarima.for()` function. Does our forecast appear reasonable?