# MATH 531T: Time Series Analysis and Forecasting (Summer 2024)
# Exam #3

## Due date: Thursday, June 20 by 5:30pm

Please upload your answers in the form of **one pdf file** on Canvas. In all questions which involve R, make sure relevant R code, output, and graphs are included in the answers to each individual part of the questions. Do not put your R code in the appendix or at the end of the file. Please include the code source file, e.g., .R or .RMD as well.

---

## 1 Concepts and Theoretical Questions

1. Consider the model $X_t = \beta_0 + \beta_1 t + \beta_2 \sin(\frac{\pi}{2}t) + Z_t$ where $\{Z_t\} \sim WN(0, \sigma^2)$. Define the seasonal differencing $\nabla_d X_t = X_t - X_{t-d}$. Recall from Exam 2 you showed that $\nabla_4 X_t$ does not have a time-dependent mean. It turns out $\nabla_4 \nabla X_t$, in which $\nabla X_t = X_t - X_{t-1}$ is also stationary. The regular differencing $\nabla X_t$ eliminates the linear trend and the seasonal differencing on $\nabla X_t$, i.e. $\nabla_4 \nabla X_t$ eliminates the remaining periodic/seasonal trend. Which of the two differencing procedures ($\nabla_4 \nabla X_t$ or $\nabla_4 X_t$) do you suggest, and why? You should provide a proof for your argument.

## 2 Simulations and Data Analysis

2. Generate $n = 100$ observations from each of the following processes: ARMA(1,1), ARMA(1,0), and ARMA(0,1) with $\phi = 0.6$ and $\theta = 0.9$ using arima.sim(). Plot the sample series, ACFs, and PACFs of each and compare to the theoretical (you may use the ARMAacf() function to obtain the theoretical values (see Examples for acf and pacf for details). Compare your results with the guidelines of Table 3.1 in your textbook. Do your results match with the Table 3.1 guidelines?

3. Let $\{x_t\}$ represent the cardiovascular mortality series (cmort) discussed in Example 2.2 of your textbook. Fit an AR(2) to $\{x_t\}$ using linear regression as in Example 3.18 (ar.ols) AND in arima() using order = c(2,0,0). Both models will produce an estimate for the "intercept" parameter, but will differ in value. Research these functions to explain why these intercept values differ and provide a discussion on the topic. Provide plots of your fitted models over the data, no discussion needed on model fit.

4. The dataset *Demand.txt* includes 24 years (Jan. 1992 - Dec. 2015) of total monthly number of customers having business with a financial firm. Interest lies in performing a complete exploratory and model fitting analysis. For the model fitting analysis, consider models within the SARIMA class of models (note this includes the AR, MA, ARMA, ARIMA, or SARIMA models). You will use the first 22 years of data as the training set (for (a) - (d)), and the last two years (24 months) as the validation set (for assessing forecasting performance in (e)). You may use any code you write yourself, code you have previously obtained from other sources, code provided in class, or code provided in the following YouTube Videos on ARIMA Forecasting in R:

   (a) Perform an exploratory analysis of the time series. Checking for non-constant variance, and apply variance stabilization transformations if necessary. Discuss your thoughts and findings.

   (b) Comment on any general and seasonal trends you see in the data. Apply decomposition techniques discussed in class for extracting the general and seasonal trends: e.g. moving average smoother, exponential smoothing, and/or differencing. Test the resulting residuals (or random component) for stationarity.

(c) Using the ACF and PACF of the de-trended/de-seasoned series and any differencing values used in the previous question to discuss potential values for $p, d, q, P, D, Q$, and $s$. Note I am not looking for one particular answer here - as per our class discussions it can be difficult to really assess all of the parameter values with uniqueness based on the ACF and PACF plots.

(d) Fit several models (atleast 3) based on your previous explorations. You may fit the de-trended/de-seasoned series or the original data (with any transformation you may have applied in (a)) just be aware of any trend or season you may have already extracted. You may use the *auto.arima* function if you would like. What are the values of $p, d, q, P, D, Q$, and $s$ in your top 3 proposed models? Provide a residual analysis and comparative model criteria (e.g. AIC, AICc, etc.) for these three models. Which is your top recommended model out of the three? Why?

(e) Finally forecast your top three models over the next two years, compare your forecast for each model with the validation set. Which model performed better from a forecasting perspective? It is the same model you recommended in (d)? Overall, which model would you recommend to the firm use future analyses? Explain.