Exam 2-Spring 2023 (65 points)

1 Exercise 1 (10 points)

. (a) Explain how k-fold cross-validation is implemented. What are the advantages and dis-advantages of k-fold cross-validation relative to LOOCV? Explain.

(b) True or false, we can use the R^2 criterion when selecting the best model among all possible subsets using a specified set of predictors. Explain.

2 Exercise 2 (7.5 points)

Identify the following models as ARMA(p, q) models. Identify p and q.

(a) $x_t + 0.25x_{t-2} = w_t + 2w_{t-1}$

(b) $x_t + 0.6x_{t-1} - 0.1x_{t-2} - 0.2x_{t-3} = w_t$

(c) $x_t = \frac{1}{4}(w_{t-3} + w_t + w_{t-1} + w_{t-2})$

3 Exercise 3 (11.5 points)

Evaluate the mean and covariance function for each of the following processes. In each case, determine whether or not the process is stationary. Here, w_t is i.i.d. N(0,1).

(a) $x_t = 2 + tw_2$

(b) $y_t = \nabla x_t = x_t - x_{t-1}$ is first order differencing of the series in part(a).

4 Exercise 4 (11.5 points)

Suppose $Cov(x_t, x_{t+h}) = \gamma(h)$ is free of t but that $E(x_t) = 4t$.

(a) Is x_t stationary? Explain.

(b) Let $y_t = 7 - 4t + x_t$. Is y_t stationary? Show your work.

5 Exercise 5 (8.5 points)

Consider the following AR(2) model:

$x_t = -0.0625x_{t-2} + w_t$, where w_t is a white noise.

(a) Rewrite the model using back-shift operator.

(b) Find the polynomial that corresponds to the operator in (a) and find its roots.

6 Exercise 6 (16 points)

From a data set of a hospital 100 adults patients, we attempt to develop equations for estimation of SYSTOLIC BLOOD PRESSURE (response variable). The predictor variables considered are:

x_1 : WEIGHT

x_2 : HEIGHT

x_3 : SERUM.CHOL

x_4 : IQ

x_5 : AGE

x_6 : SMOKER(Yes = 1, No = 0)

x_7 : SODIUM .

Use Table 1 and Fig 1 and Figure 2 to answer the following questions:

(a) Report the best 3 variables and 4 variables models respectively according to the best subset selection. (Use table 1)

(b) Report the best subset model according to C_p, BIC, and adjusted R^2 respectively.

(c) Figure 2 shows the plots for tests MSE associated with the best model of each size for k-fold cross validation and for validation set approach. Which size model would you select according to k-fold cross validation, and according to validation set respectively? Explain.

(d) In Figure 2, one would like to choose between the four variables and the five variables model. Would you use validation set approach or k-fold cross validation? What size model would you choose? Justify your answer.

7 Exercise 7 :Bonus (3 points)

Figure 3 represents the international airline passengers series (monthly total, Jannuary 1949 to December 1960). The series clearly exhibits a **strong seasonal effect**($period = 12$: monthly), an **increasing trend**, and an **increasing variability**. Explain in details how you can make this series a stationary looking realization.