Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES (SIMS)
MASTER OF SCIENCE IN STATISTICAL SCIENCES
~~END OF SEMESTER EXAMINATION~~ SPECIAL EXAMINATION
STA 8203: TIME SERIES AND FORECASTING

DATE: 22nd April 2021 Time: 3 Hours

Instructions
1. This examination consists of **FOUR** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question One

1.1. Define the *autocorrelation function* (ACF) for a stationary time series. (1 marks:)

1.2. Derive the ACF for the MA (2) series generated by the scheme (5 marks)

$$Y_t = w_t + \theta_1 w_{t-1} + \theta_1 w_{t-2}$$

where w_t is a white noise series.

1.3. Derive the ACF for the AR (1) series generated by the scheme. (4 marks)

$$Y_t = \emptyset Y_{t-1} + w_t$$

where $|\phi| < 1$ and w_t is a white noise series.

1.4. For the ARMA (1,2) model $Y_t = 0.8Y_{t-1} + w_t + 0.7w_{t-1} + 0.6w_{t-2}$ (6 marks)

show that

(i) $\rho(k) = 0.8\rho(k-1)$, for $k \geq 3$ and
(ii) $\rho(2) = (0.8\rho(1) + 0.6\sigma_w^2)/\gamma(0)$

Where Y_t is the original series data and w_t is a series of white noise.

1.5. Describe, in general terms, the behavior you would expect to see for the *estimated* ACF computed from the set of observations $Y_1 \ldots, Y_T$ generated by the random walk model. (4 marks)

$$Y_t = Y_{t-1} + w_t$$

Question Two

The traffic department would like to analyse the number of accidental deaths occurring monthly for a period of 5 years.

2.1. In order to investigate possible non-stationarity of the series, the following three plots are produced from the series. What kind of effects do each of the plots show and how would these effects influence the analysis of the series?

(3 marks)

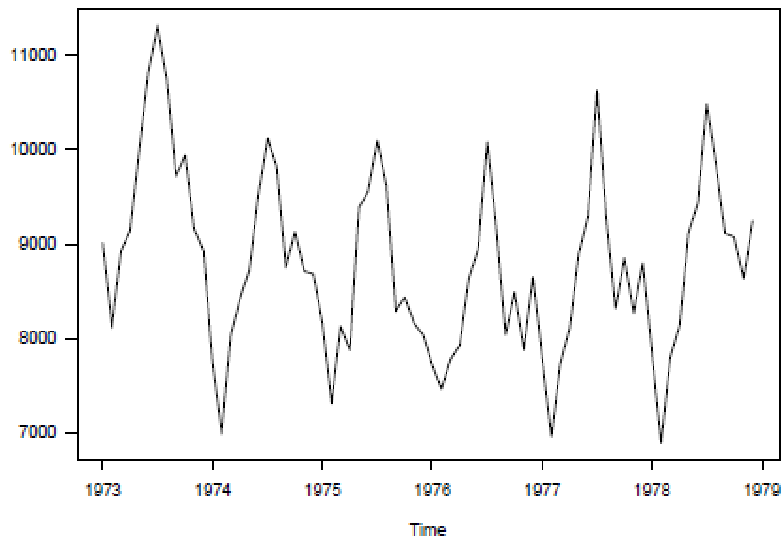


Figure 1: A plot of the original accident time series Y_t.

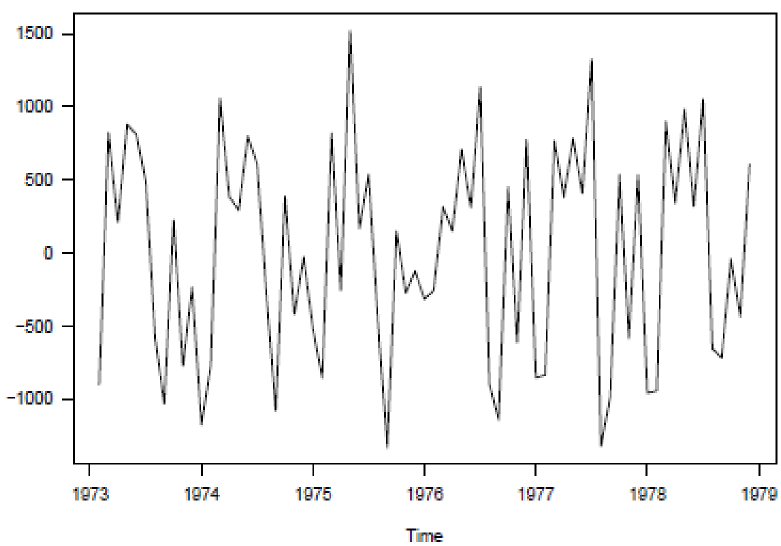


Figure 2: A plot of the differenced accident time series $Y_t - Y_{t-1}$.

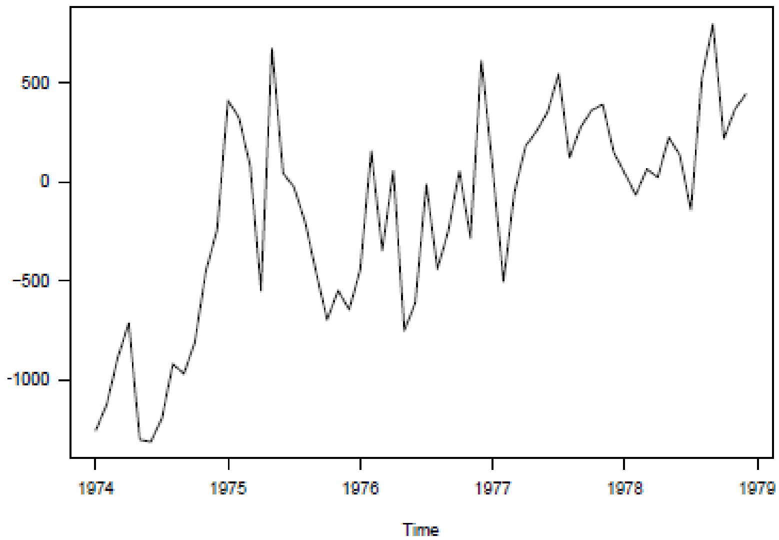


Figure 3: A plot of the seasonally differenced accident time series $Y_t - Y_{t-12}$.

2.2. The time series is read into an R data set called deaths and the following statements are issued to R.

```
> acf(diff(diff(deaths, 12), 1), 24)
> pacf(diff(diff(deaths, 12), 1), 24)
```

Describe in detail exactly what these statements do. (4 marks)

2.3. The following graphs are produced as a result of running the R statements above. Explain what kind of model structure you think that the graphs indicate is appropriate for modelling this series.

(4 marks)

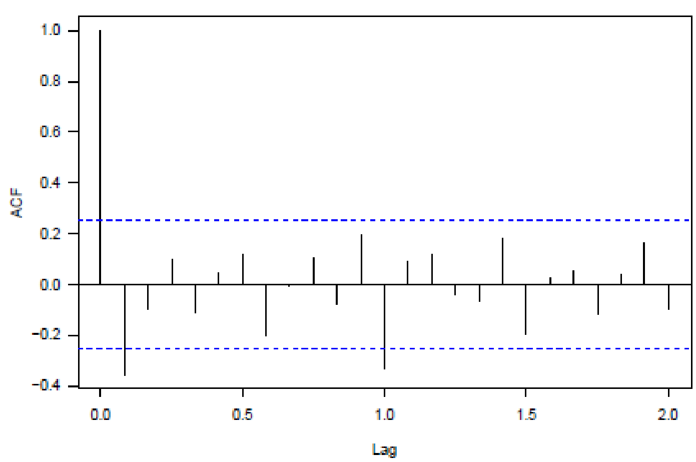


Figure 4: The first plot produced by the R statements.

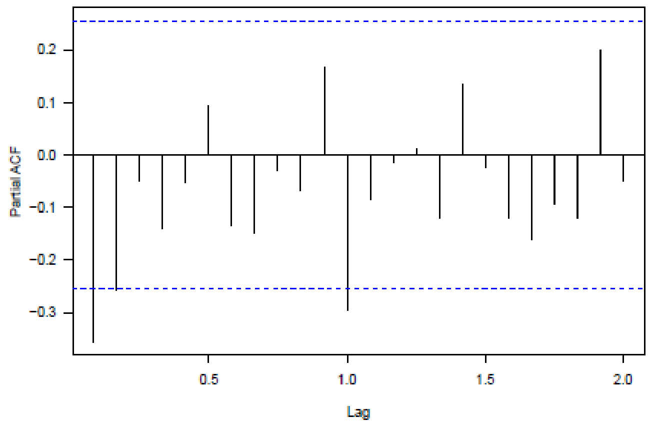
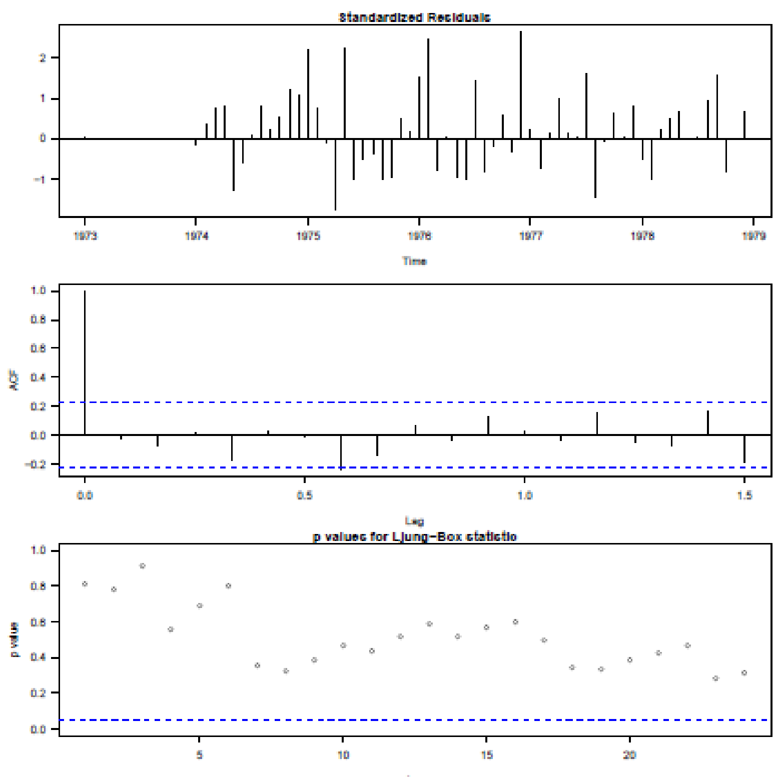


Figure 5: The second plot produced by the R statements.

2.4. Using operator notation, write down the complete model which should be fitted to the accident data series. (4 marks)

2.5. A time series model is fit to the deaths series and the following results and plots obtained. Does the model fit well? Give reasons. (5 marks)

	PARAM1	PARAM2
Estimate	-0.4264	-0.5584
SE	0.1226	0.1787
95% CI of estimate	-0.6667 to -0.1861	-0.9087 to -0.2082
sigma^2 = 99,480; Log likelihood = -425.53 and AIC = 857.06		



Question Three

3.1 You are provided with the equation below that describes the milk production (measured in thousands of litres per week).

$$f(t) = 4.3 \times cos[0.82t - 7.3] + 7.3$$

where t is time in years since January 1995.

Calculate the amplitude, the vertical offset, the phase shift, the angular frequency, and the period, and interpret the results. ____(10 marks)

3.2 The following logistic regression with sine and cosine functions was used to test for seasonality in the disease prevalence.

$$f(p) = -6.718 + 0.009822 \times sin\left(\frac{2\pi t}{12}\right) + 0.06929 \times cos\left(\frac{2\pi t}{12}\right)$$

Determine the extreme values (t_{max} *and* t_{min}), an amplitude (α), the shift parameter (θ) and interpret giving the lowest and highest prevalence ___________________________(10 marks).

Question Four

4.1. The oscillation data defined as the difference in barometric pressure was modelled by considering two possible models, AR (2) and MA (1) on differenced data. The parameter estimates computed for the two models are presented below. Use the results to answer the questions that follow.

AR (2) Model			MA (1) Model		
	Estimate	95% CI		Estimate	95% CI
Intercept	-0.0050		Intercept	-0.0051	
AR1	-0.4064	(-0.4884, -0.3243)	MA1	-0.3921	(-0.4638, -0.3205)
AR2	-0.1649	(-0.2469, -0.0829)			

(a) Write each model in the form of ARIMA_______________________________ (2 marks)

(b) Write each model for the differenced data taking Y_t, to represent the differenced observations and w_t the white noise for the differenced data. _________________(2 marks)

(c) It is often more convenient to express the models in terms of the original data, rather than the differenced data. Express each model in terms of the original data, X_t, rather than the differenced data, Y_t. ____(12 marks)

(d) The Box-Ljung test was used to determine the model that best fit the data. The p-value for AR(2) was 0.080 and for MA (1) was 0.026. State the null and alternative hypothesis under the Box-Ljung test and the conclusion on the best fit. ______________________(4 marks)