NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

SORS6102

FACULTY OF APPLIED SCIENCE

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH

SORS6102 FORECASTING

MSc. Operations Research Part II

NOVEMBER 2024 EXAMINATION

Time : 3 hours

Total Marks: 100

Candidates may attempt **ALL**. For all questions where necessary clearly show all your work for full credit.

A1. MULTIPLE CHOICE [1 Mark each]

(a) Which time series model assumption are you testing when you perform a runs test?

(A) stationarity
(B) independence
(C) normality
(D) invertibility

(b) Consider an invertible MA(2) process

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

Which statement is true?

(A) Its PACF can decay exponentially or in a sinusoidal manner depending on the roots of the MA characteristic polynomial.
(B) It is always stationary.

(C) Its ACF is nonzero at lags $k = 1$ and $k = 2$ and is equal to zero when $k > 2$.
(D) All of the above.

(c) Which function is best suited to determine the orders of a mixed ARMA (p, q) process?

(A) EACF
(B) PACF
(C) ACF
(E) a bivariate cross-correlation function

(d) True or False. When compared to Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC) is more likely to favor overly-large models (i.e., models with more parameters).

(A) True
(B) False

(e) Which value of θ in the MA(1) family produces the same ACF?as an MA(1) process with $\theta = 0.5$.?

(A) -0.5
(B) 0
(C) 1
(D) 2

(f) Here is the R output from fitting an ARMA(1, 1) model to a data set:

```
> arima(cows,order=c(1,0,1),method='CSS') # conditional least squares
Coefficients:
              ar1 ma1 intercept
          0.6625 0.6111 58.7013
s.e. 0.0616 0.0670 1.6192
```

What is the estimate of the autoregressive parameter ϕ ?

(A) -0.6625
(B) -0.6111
(C) 0.6625
(D) 0.6111

(g) True or False. If $\{\nabla Y_t\}$ is a stationary process, then $\{Y_t\}$ must be stationary.

(A) True
(B) False

(h) What is a likelihood function?

(A) A function that describes how extended autocorrelations are computed
(B) A function that is maximized to find model estimates
(C) A function that can determine the order of seasonal and nonseasonal differencing

(D) A function that can be used to test for a unit root

(i) In an analysis, we have determined the following:

(1.) The sample PACF for the series $\{Y_t\}$ has a slow decay.

(2.) The mean and variance for the series $\{Y_t\}$ are constant over time.

(3.) The first difference process $\{\nabla Y_t\}$ has a sample ACF with a very large spike at lag 1 , a smaller spike at lag 2, and no spikes elsewhere.

Which model is most consistent with these observations?

(A) MA(1)

(B) ARI(2, 1)

(C) AR(2)

(D) IMA(1, 2)

(j) Suppose that $\{Y_t\}$ is a white noise process with a sample size of $n = 100$. If we performed a simulation to study the sampling variation of r_1, the lag one sample autocorrelation, about 95 percent of our estimates r_1 would fall between

(A) -0.025 and 0.025

(B) -0.05 and 0.05

(C) -0.1 and 0.1

(D) -0.2 and 0.2

(k) Consider the nonseasonal process defined by

$$(1 + 0.6B)(1 - B)Y_t = \left(1 - B + 0.25B^2\right) e_t$$

This process is identified by which ARIMA model?

(A) ARIMA(1, 1, 2)

(B) ARIMA(2, 1, 1)

(C) ARIMA(2, 2, 1)

(D) ARIMA(2, 1, 2)

(l) You have performed a Ljung-Box test to determine if an ARMA(1, 1) model is suitable for a data set. The p-value for the test is equal to 0.329 . What should you do?

(A) Reject H_0 and look for a better model.

(B) Do not reject H_0 on the basis of this test.

(m) In Chapter 3, we talked about using regression methods to detrend time series data that exhibited deterministic trends (e.g., linear, quadratic, seasonal, etc.). For all of the regression output in R to be applicable, we needed the regression model errors to have zero mean, constant variance, and be independent. What additional assumption was needed?

(A) invertibility

(B) normality

(C) unbiasedness

(D) strict stationarity

(n) The analysis of time series data must account for the fact that data measured close in time are very often (choose one)

(A) independent

(B) discrete

(C) identical

(D) correlated

(o) For a time series Y_t, the series of first differences is defined as $Y_t - Y_{t-1}$. What is the series of second differences?

(a) $Y_t - 2Y_{t-1}$

(b) $Y_t - Y_{t-2}$

(c) $Y_t - 2Y_{t-1} + Y_{t-2}$

(d) $Y_t - Y_{t-1} - Y_{t-2}$

[15]

A2. Consider two random variables, X and Y. Suppose $\mathrm{E}(X) = 6$, $\mathrm{var}(X) = 9$, $\mathrm{E}(Y) = 0$, $\mathrm{var}(Y) = 4$, and $\mathrm{corr}(X, Y) = 0.25$. Find the following, showing all your steps:

(a) $\mathrm{var}(X + Y)$, and [3]

(b) $\mathrm{corr}(X + Y, X - Y)$. [3]

A3. ARIMA models include a parameter, d, that controls the number of times a time series is differenced before being modeled by an ARMA process.

(a) Why is differencing a time series sometimes necessary? [3]

(b) How could you choose the amount of differencing required for a particular time series? [2]

A4. (a) Define (weak) stationarity. [3]

(b) Give an example of a non-stationary process. [2]

(c) Based on the visual representation of the time series shown in Figure 1, does the series appear stationary? Justify your answer. [3]

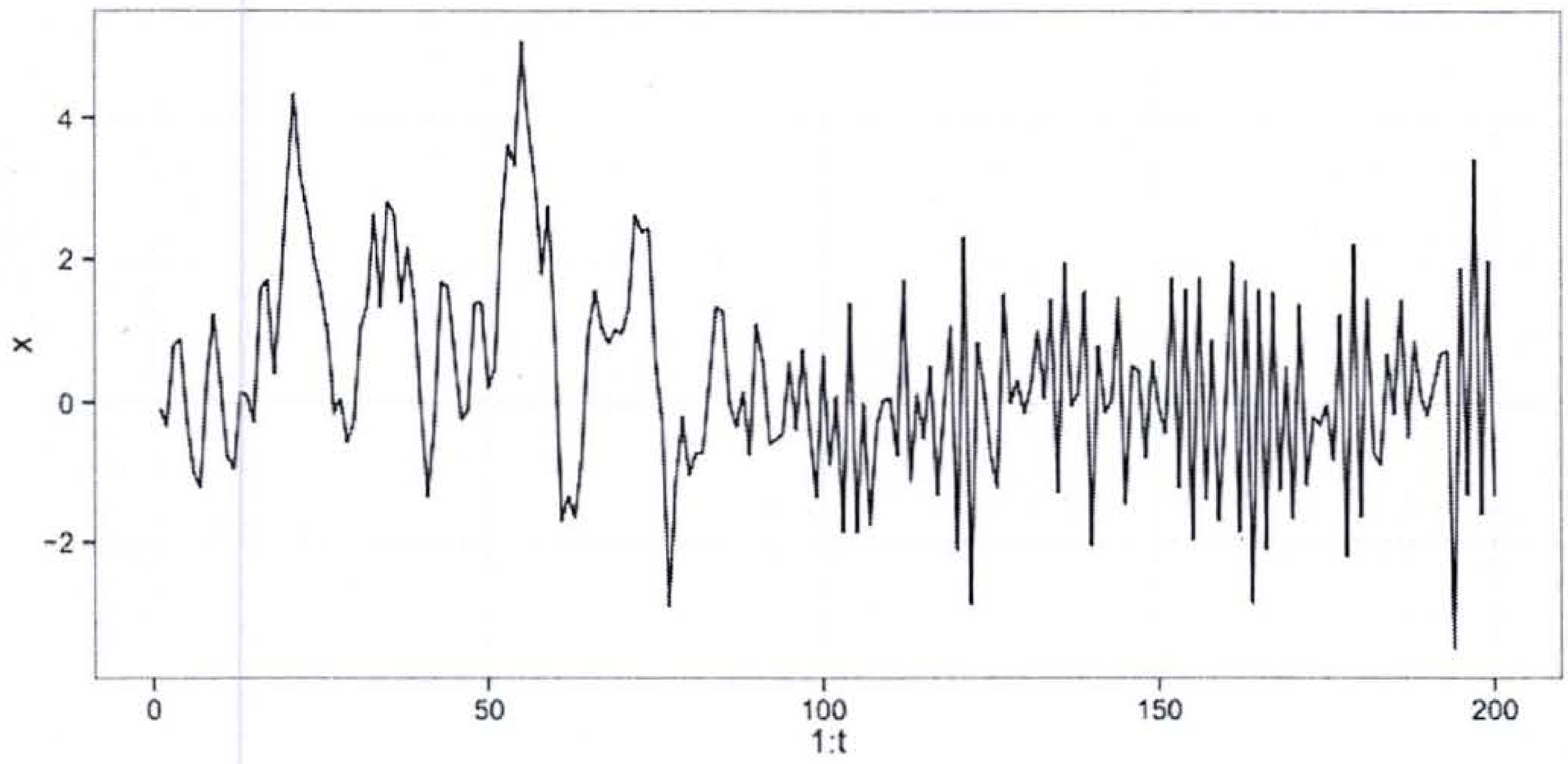

Figure 1: Time Series Plot

A5. Suppose $\{e_t\}$ is a normal white noise process with mean zero and variance $\sigma_e{}^2$. Let $\{Y_t\}$ be a process defined as: $Y_t = e_t e_{t-1}$.

(a) Find the mean function and the autocovariance function of Y_t. [6]

(b) Is the time series $\{Y_t\}$ stationary? Explain your answer. [3]

(a) A simple trend model for a time series Y_t might be specified as $Y_t = \mu_t + X_t$. Explain briefly in words what each of μ_t and X_t signify in this model. [4]

(b) A data analyst fit a linear trend model with Y_1 as the response, and the AIC of the linear model was 453.7. The analyst fit a quadratic trend model with the same Y_t as the response, and the AIC of the quadratic model was 446.2. What conclusions can you draw? [3]

(c) A data analyst fit a linear trend model with Y_1 as the response, and the AIC of the linear model was 624.3. Because a residual analysis showed possible non constant spread of the residuals, the analyst fit a linear trend model with the logarithm of Y_t as the response, and the AIC of the log-transformed linear model was 342.8. What conclusions can you draw? [3]

(d) Given the data $Y_1 = 10, Y_2 = 9$, and $Y_3 = 9.5$, we wish to fit an IMA(1,1) model without a constant term.

(i) Find the conditional least squares estimate of θ. [5]

(i) Estimate σ^2. [5]

A6. The white blood cell count was measured for a patient over a period of 36 days. A plot of the time series is in Figure 2.

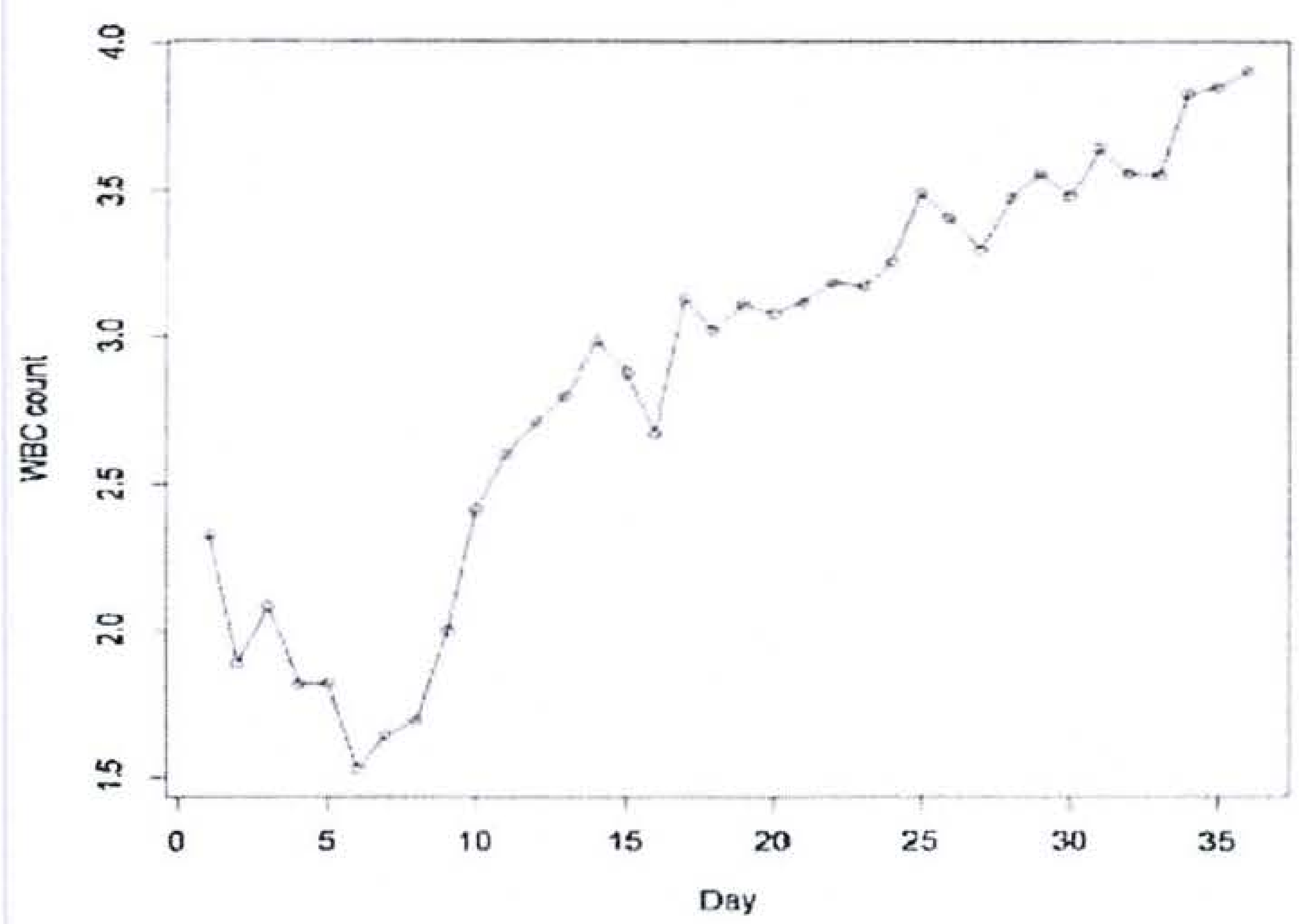

Figure 2: Plot of time series

(a) An analyst decided to fit a linear trend model to this time series. Briefly discuss why you do or do not agree with this choice, based on an initial look at the data. [4]

(b) Summary output from the 'lm' function in R is given below. Write the equation of the fitted linear time trend model and comment on the output. [4]

```
lm(formula = blood.ts ~ time(blood.ts))

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.741011   0.084288   20.66   <2e-16 ***
time(blood.ts) 0.062127   0.003973   15.64   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2476 on 34 degrees of freedom
Multiple R-squared:  0.8779,    Adjusted R-squared:  0.8744
```

(c) Various plots (Figure 3-Figure 5) of the standardized residuals for this linear trend model fit are given below. For each one, write a brief comment explaining what can be concluded about the stochastic component of the model, based on the respective plot. [9]

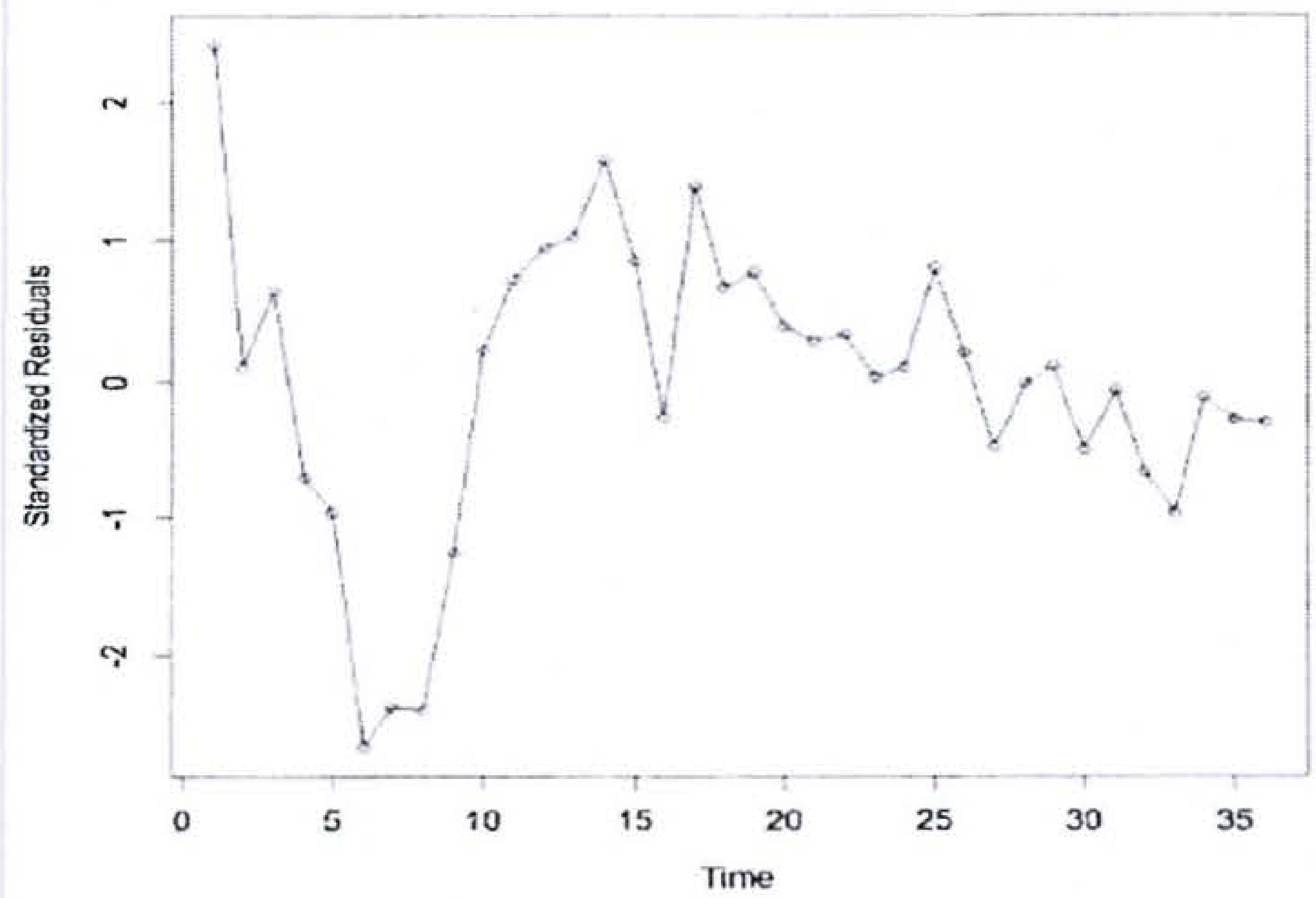

Figure 3: Redisuals vs. time plot

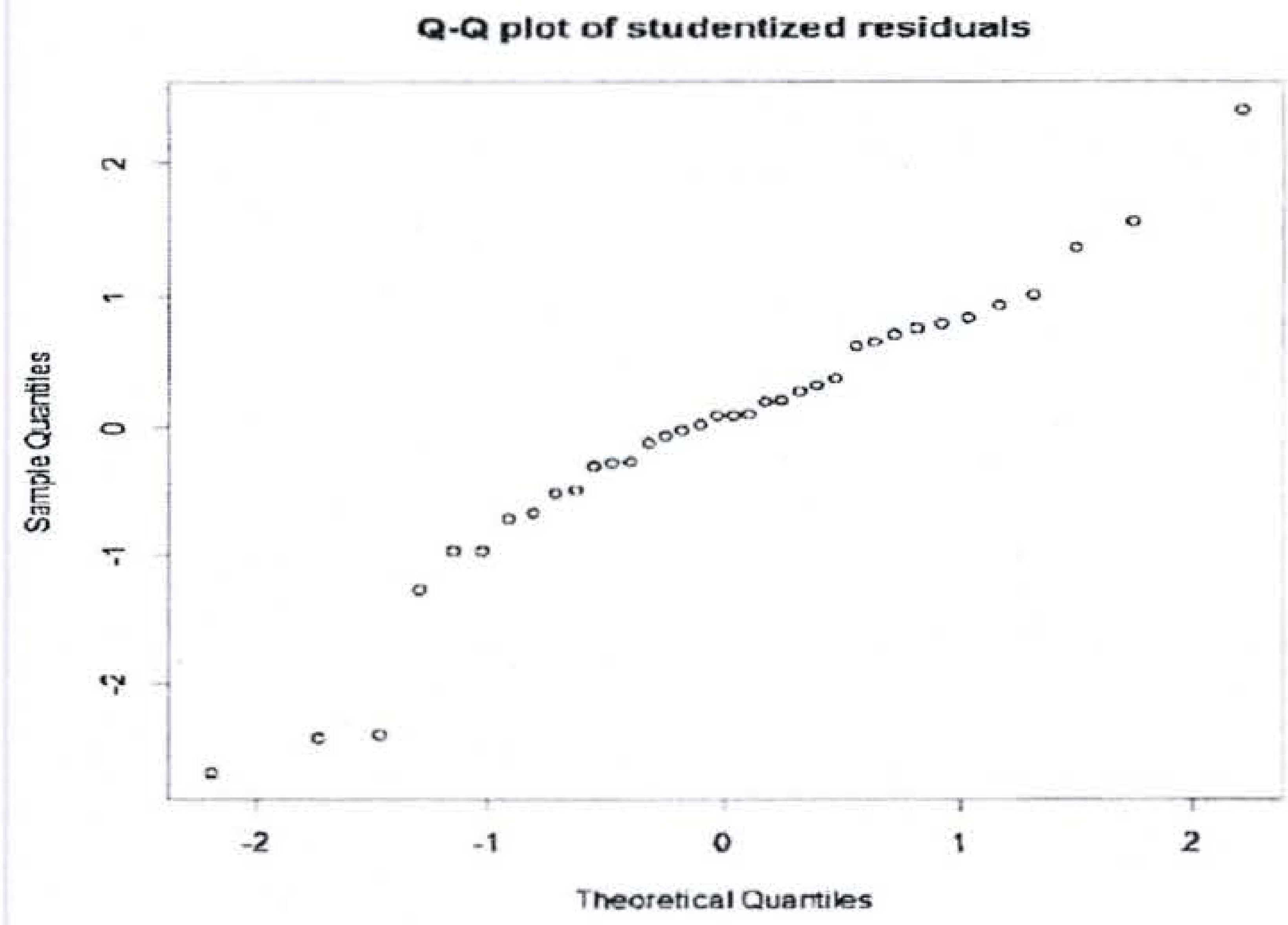

Figure 4: Normal Q-Q Plot

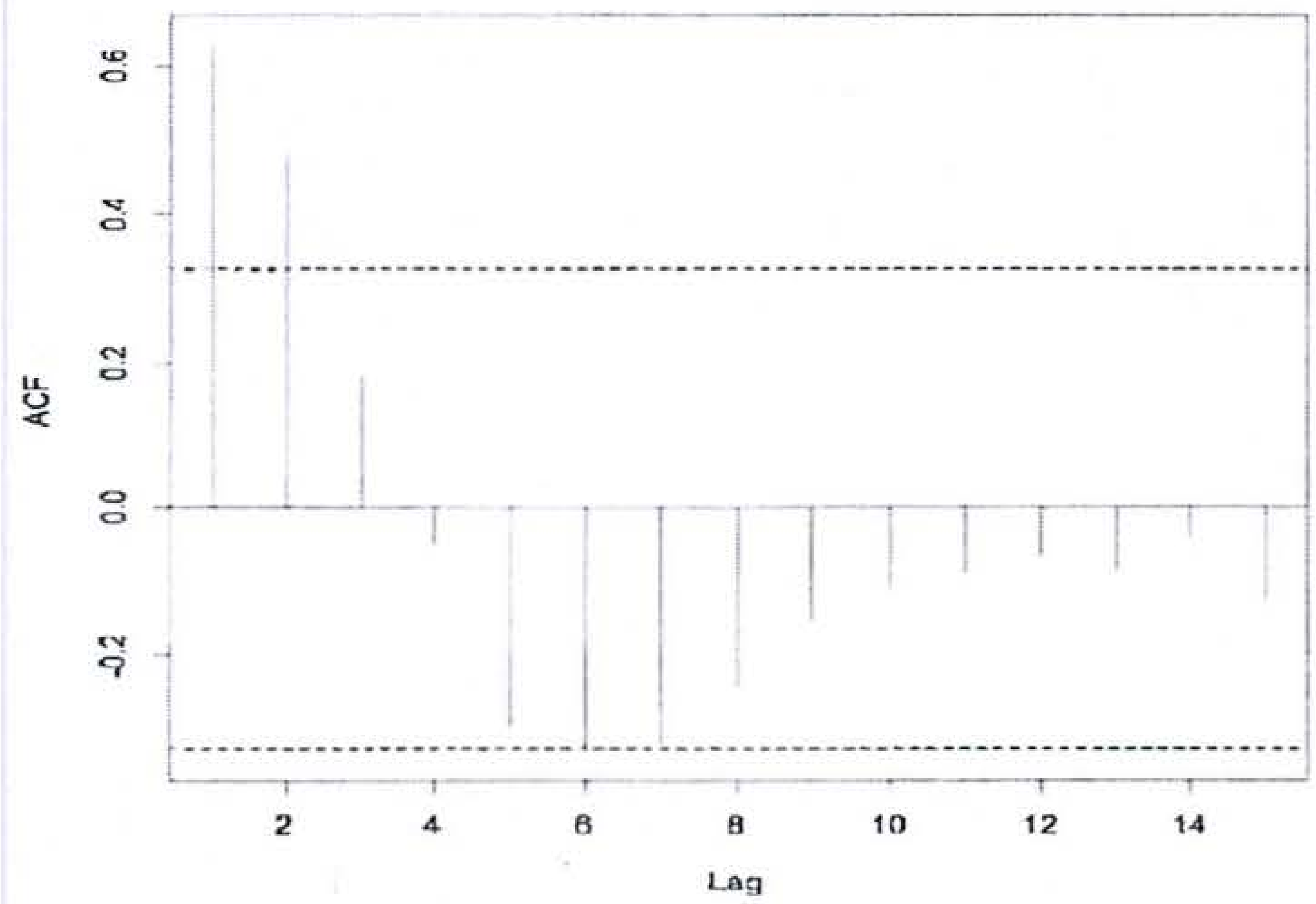

Figure 5: ACF plot of residuals

(d) Do you believe the observed number of runs for the series of studentized residuals would be less than, greater than, or approximately equal to the expected number of runs under the assumption of independence? Briefly explain your answer. [3]

A7. (a) Suppose $\{e_t\}$ is a normal white noise process with mean zero and variance $\sigma_e{}^2$. Let $\{Y_t\}$ be a process with a linear trend over time, defined as:

$$Y_t = \beta_0 + \beta_1 t + e_t$$

(i) Explain why the time series $\{Y_t\}$ is not stationary. [2]

(ii) Consider the differenced time series $\{\nabla Y_t\}$, where $\nabla Y_t = Y_t - Y_{t-1}$. Show that the differenced time series $\{\nabla Y_t\}$ is stationary. [Hint: Note that β_0, β_1, and t are constants, not random variables, here.] [3]

(b) Consider the MA(2) process, where all the $\{e_t\}$ values are independent white noise with variance σ^2 :

$$Y_t = e_t - 0.5e_{t-1} - 0.3e_{t-2}$$

(i) Find cov $(Y_t, Y_t) =$ var (Y_t). [3]

(ii) Find cov (Y_t, Y_{t-1}) and, hence, find the lag-1 autocorrelation corr (Y_t, Y_{t-1}). [4]

(iii) Find cov (Y_t, Y_{1-2}) and, hence, find the lag-2 autocorrelation corr (Y_t, Y_{t-2}). [4]

(iv) Argue that cov $(Y_t, Y_{t-k}) = 0$ for all $k \geq 3$. [4]

END OF QUESTION PAPER