Name: Example Solutions

STAT 520 – Test 1 – Fall 2023

1) If a time series displays a mean function that is not constant, name two general approaches you could take to achieve a series that is stationary. State an advantage of each approach.

Differencing – there is no need to specify a trend model

Detrending – can use the estimated trend model to characterize the time series

2) 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

3) 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.

μ_t = overall mean function

X_t = stochastic component (random noise/error)

4) For what type of time series data (i.e., having what pattern or shape in a plot of the time series) are the seasonal means model and the harmonic regression model commonly used? Which of these two models makes a stronger assumption about the shape of the trend model?

A periodic (oscillating seasonally) pattern

Harmonic regression makes a stronger assumption about the shape.

5) A data analyst fit a linear trend model with Y_t 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?

The quadratic model is preferred since it has a lower AIC.

6) A data analyst fit a linear trend model with Y_t as the response, and the AIC of the linear model was 624.3. Because a residual analysis showed possible nonconstant 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?

None – we cannot compare the AICs of two models with different response variables.

7) A student wrote the following statement: "If a time series has a constant mean function over time, then that time series is stationary." In a sentence or two, carefully assess whether or not this statement is valid.

That is not enough to establish that the series is (even weakly) stationary. You also need the variance function to be constant, and the autocovariance function to depend only on the lag, not on t.

8) 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) $Y_t - Y_{t-1} - Y_{t-2}$

9) Let $\{Y_t\}$ be a stationary process. For such a process, how does $\text{var}(Y_t)$ compare to $\text{var}(Y_{t-1})$? Briefly explain.

$\text{var}(Y_t) = \text{var}(Y_{t-1})$ since the variance is constant if the series is stationary.

10) Consider two random variables, X and Y. Suppose $E(X) = 3$, $\text{var}(X) = 9$, $E(Y) = 0$, $\text{var}(Y) = 4$, and $\text{corr}(X, Y) = -0.2$. Find the following, showing all your steps:

(a) $\text{var}(X + Y)$

Note $\text{corr}(X,Y) = -0.2 = \dfrac{\text{cov}(X,Y)}{\sqrt{(9)(4)}}$

$\Rightarrow \text{cov}(X,Y) = (-0.2)(6) = -1.2$

$\Rightarrow \text{var}(X+Y)$

$= \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y)$

$= 9 + 4 + 2(-1.2)$

$= \boxed{10.6}$

(b) $\text{corr}(X + Y, 2X - Y)$ [Hint: You can use the fact that $\text{var}(2X - Y) = 44.8$ here.]

$\text{cov}(X+Y, 2X-Y) = 2\,\text{cov}(X,X) - \text{cov}(X,Y) + 2\,\text{cov}(X,Y) - \text{cov}(Y,Y)$

$= 2(9) - (-1.2) + 2(-1.2) - 4$

$= 18 + 1.2 - 2.4 - 4 = 12.8$

$\Rightarrow \text{corr}(X+Y, 2X-Y) = \dfrac{12.8}{\sqrt{(10.6)(44.8)}} = \boxed{0.5874}$

11) Suppose $\{e_t\}$ is a normal white noise process with mean zero and variance 1. Let $\{Y_t\}$ be a process defined as:
$Y_t = e_t + 0.5e_{t-2}$.

(a) Find the autocovariance function for this process. Write the autocovariance function as a piecewise function for various values of the lag, specifically for lags $k = 0, 1, 2$, and $k > 2$. Show your work where applicable.

$k=0$: $cov(Y_t, Y_t) = cov(e_t + 0.5e_{t-2}, e_t + 0.5e_{t-2})$

$= cov(e_t, e_t) + 0.5\,cov(e_t, e_{t-2}) + 0.5\,cov(e_{t-2}, e_t) + 0.25\,cov(e_{t-2}, e_{t-2})$

$= 1 + 0 + 0 + 0.25(1) = 1.25$

$k=1$: $cov(Y_t, Y_{t-1}) = cov(e_t + 0.5e_{t-2}, e_{t-1} + 0.5e_{t-3})$

$= cov(e_t, e_{t-1}) + 0.5\,cov(e_t, e_{t-3}) + 0.5\,cov(e_{t-2}, e_{t-1}) + 0.25\,cov(e_{t-2}, e_{t-3})$

$= 0 + 0 + 0 + 0 = 0$

$k=2$: $cov(Y_t, Y_{t-2}) = cov(e_t + 0.5e_{t-2}, e_{t-2} + 0.5e_{t-4})$

$= cov(e_t, e_{t-2}) + 0.5\,cov(e_t, e_{t-4}) + 0.5\,cov(e_{t-2}, e_{t-2}) + 0.25\,cov(e_{t-2}, e_{t-4})$

$= 0 + 0 + 0.5(1) + 0.25 = 0.5$

For $k > 2$, $cov(Y_t, Y_{t-k})$ will be zero since it is easy to see there will be no overlapping subscripts.

So

$$\gamma_k = \begin{cases} 1.25 & \text{if } k=0 \\ 0 & \text{if } k=1 \\ 0.5 & \text{if } k=2 \\ 0 & \text{if } k>2 \end{cases}$$

(b) Find the autocorrelation function for this process (show work when applicable). Write the autocorrelation function as a piecewise function for same values of the lag as you considered in part (a).

$$\rho_k = corr(Y_t, Y_{t-k}) = \frac{cov(Y_t, Y_{t-k})}{\sqrt{var(Y_t)\, var(Y_{t-k})}}$$

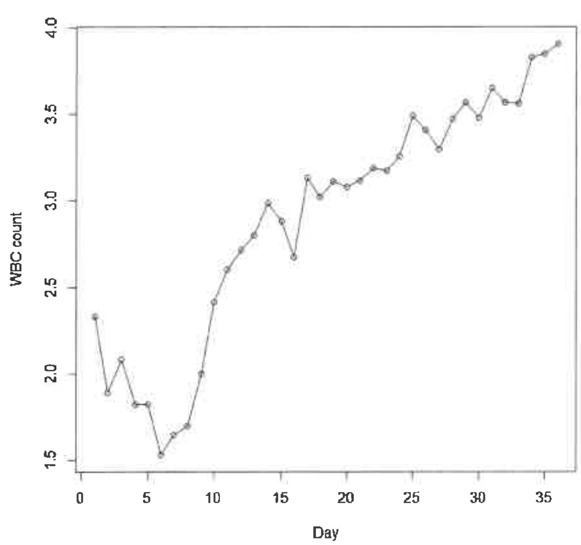
$\underline{k=0}$: Clearly $corr(Y_t, Y_t) = 1$. $\underline{k=1}$. From (a), clearly $corr(Y_t, Y_{t-1}) = 0$. $\underline{k=2}$: $corr(Y_t, Y_{t-2}) = \frac{0.5}{\sqrt{(1.25)(1.25)}}$

$= \frac{0.5}{1.25} = 0.4$

And for $k>2$, $corr(Y_t, Y_{t-k}) = 0$.

$$\Rightarrow \rho_k = \begin{cases} 1 & \text{if } k=0 \\ 0 & \text{if } k=1 \\ 0.4 & \text{if } k=2 \\ 0 & \text{if } k>2 \end{cases}$$

12) The white blood cell count was measured for a patient over a period of 36 days. A plot of the time series is given below.



(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.

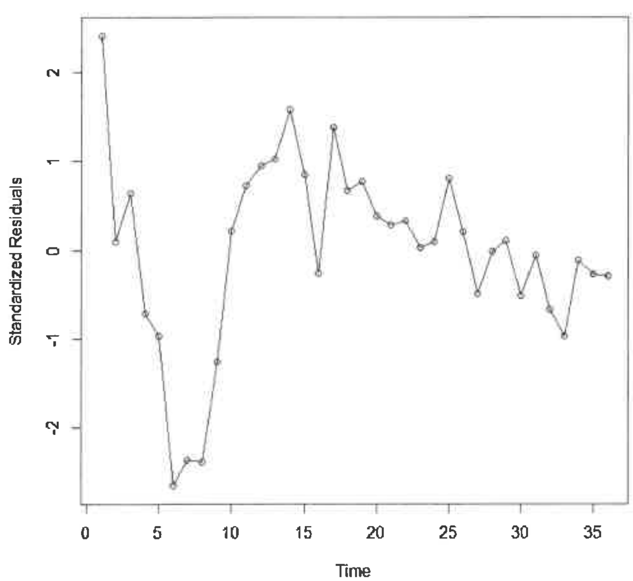
Answers may vary, but justify based on the pattern shown in the plot.

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

$$\hat{\mu}_t = 1.741 + 0.0621\, t$$

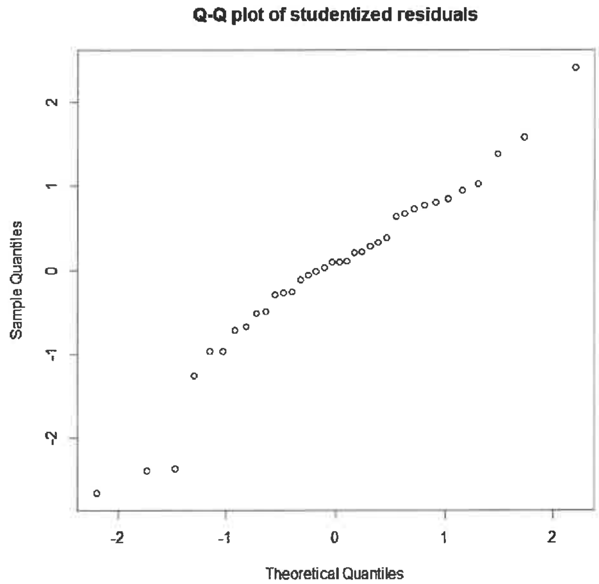
```
lm(formula = blood.ts ~ time(blood.ts))

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 (three in all) 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.

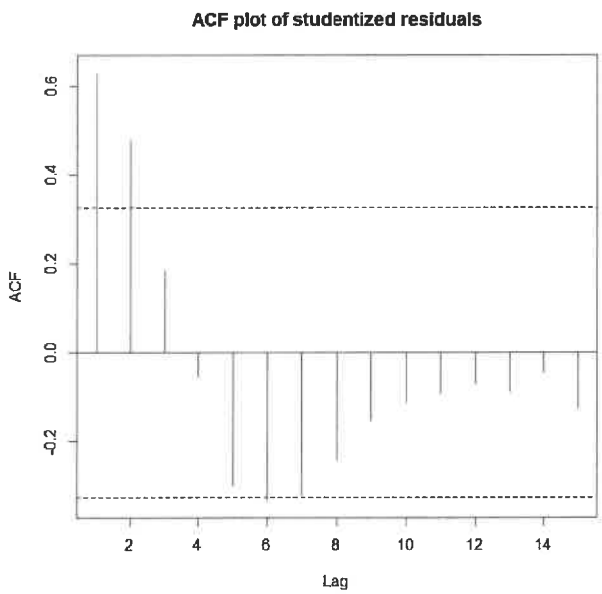


the variance function does not appear constant over time (it's greater at earlier times).

The pattern over time indicates possibly a wrong trend model was used.



The normality of the stochastic component seems reasonable
(it's OK to mention a couple of outliers)



The errors do not appear independent based on the pattern of the ACF plot.
There is positive autocorrelation at the early lags.

(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.

The observed number of runs is likely less than the expected because of the positive autocorrelation at the small lags.

Extra credit: What famous statistician died this August at the age of nearly 103?

C.R. Rao