TEST #1

STA 4853 **Name:**____________________________________

March 4, 2019

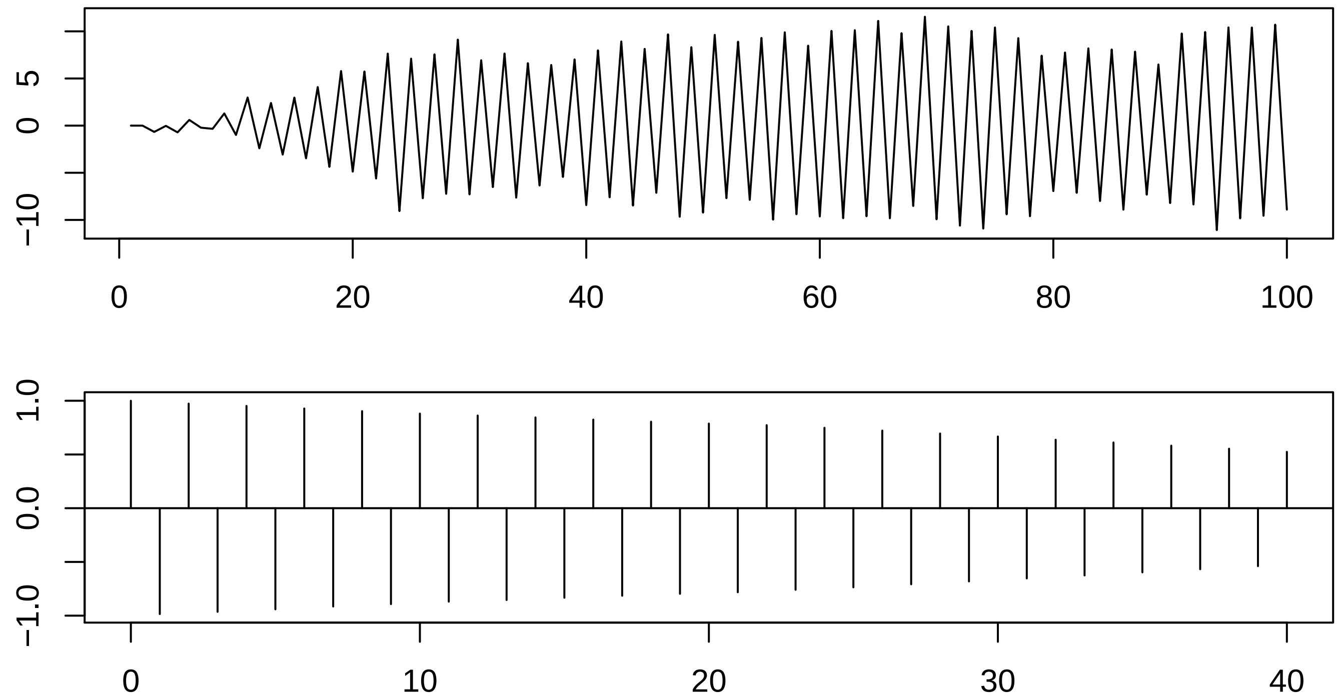
Please read the following directions.
DO NOT TURN THE PAGE UNTIL INSTRUCTED TO DO SO

Directions

- This exam is **closed book** and **closed notes**.
- There are **34** multiple choice questions.
- Circle the **single best** answer for each multiple choice question. Your choice should be made clearly.
- Always **circle the correct response**. (Sometimes the question has an empty blank or a box, but this is **NOT** where the answer goes.)
- There is no penalty for guessing.
- The exam has **13** pages.
- Each question is worth equal credit.

Problem 1. For a time series $z_1, z_2, \ldots, z_{100}$ of length 100, the two figures below give the time series plot (z_t versus t) and the estimated ACF up to lag 40. Which response best completes this sentence: This time series appears to be a realization of a _________ process.

a) stationary AR(1) **b**) stationary AR(2) **c**) random shock **d**)⋆ non-stationary
e) stationary random walk **f**) MA(1) **g**) MA(2) **h**) MA(3)



Problem 2. For a stationary MA(2) process, the value of ρ_0, the ACF at lag zero, is _________.

a) 0 **b**)⋆ 1 **c**) 0.5 **d**) 2
e) σ_a^2 **f**) $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$ **g**) $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$ **h**) $\sigma_a^2\psi_0\psi_2$

Problem 3. An AR(3) process with $\phi_1 = \phi_2 = \phi_3 = 0.4$ will _________.

a) vary about a constant mean with a constant variance
b)⋆ be non-stationary
c) have an ACF which decays rapidly to zero
d) have an ACF with a cutoff to zero after lag 3
e) have a PACF which decays to zero

Problem 4. For a stationary ARMA(1, 1) process $\{z_t\}$, the autocovariances $\gamma_k = \text{Cov}(z_t, z_{t-k})$ satisfy $\gamma_k =$ _________.

a) 0 for $k > 1$ **b)** 0 for $k > 2$ **c)** ϕ_1^k for all k
d) $\phi_1 \gamma_{k-1}$ for all k **e)**$\star$ γ_{-k} for all k **f)** σ_z^2 for $k > 1$

Problem 5. Suppose X is a random variable, and b and c are constants. Then $\text{Var}(bX + c) =$ _________.

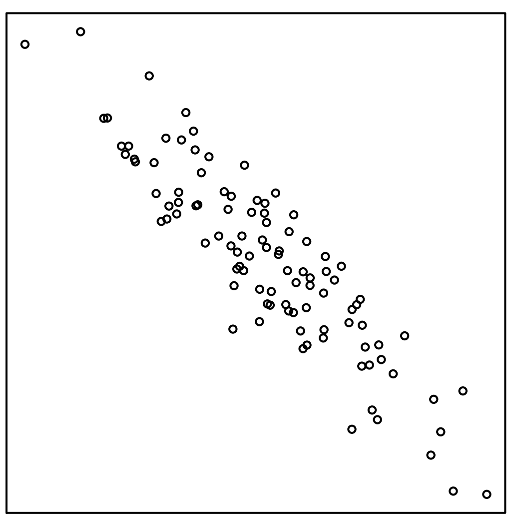
a) $b^2 + c^2$ **b)** $b + c^2$ **c)** $b + c$ **d)** $b^2\text{Var}(X) + c^2$
e) $b\text{Var}(X)$ **f)** $b\text{Var}(X) + c$ **g)**$\star$ $b^2\text{Var}(X)$ **h)** $b^2\text{Var}(X) + c$

Problem 6. To generate a realization $z_1, z_2, \ldots, z_n$ from an AR(1) process with parameters C, ϕ_1, σ_a^2, we begin by choosing a starting value z_1, and then setting z_2 equal to _________.

a)$\star$ $C + \phi_1 z_1 + a_2$ **b)** $C + \phi_1 a_2 + z_1$ **c)** $Cz_1 + \phi_1 a_2 + \sigma_a^2$ **d)** $Ca_1 + \phi_1 z_1 + a_2$
e) $Ca_2 + z_1 - \phi_1 a_1$ **f)** $C + a_2 - \phi_1 a_1$ **g)** $C + a_2 - \phi_1 a_1$ **h)** $Cz_1 - \phi_1 a_1 + a_2$

Problem 7. The scatterplot below displays a sample with a sample correlation of exactly $r =$ _________. (Make your best guess.)

a) -1 **b)**$\star$ $-.9$ **c)** $-.5$ **d)** 0 **e)** $+.5$ **f)** $+.9$ **g)** $+1$

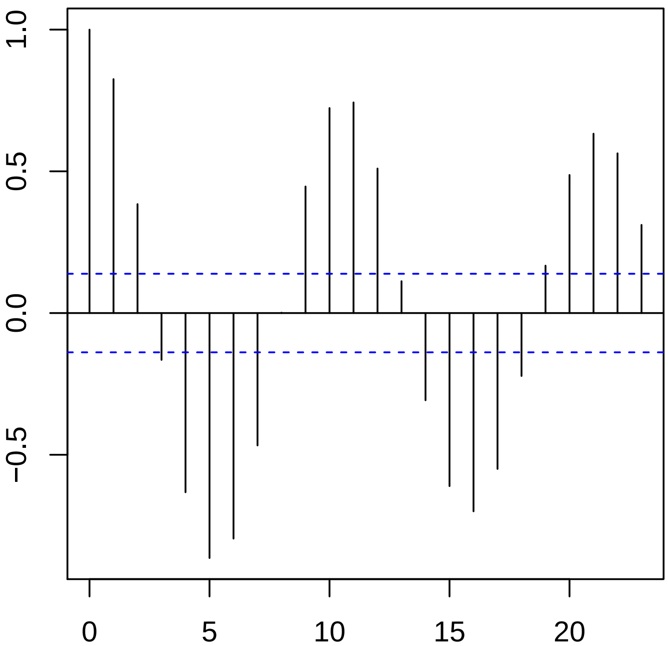


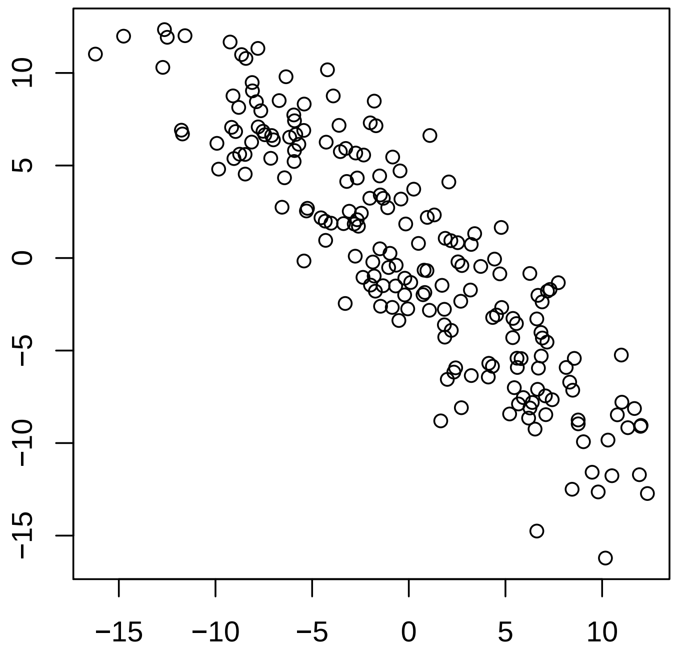
Problem 8. Let B denote the backshift operator. For a time series $\{z_t\}$, we know that $Bz_{t-3} =$ _________.

a) z_{t-1} **b)** z_{t-2} **c)** z_{t-3} **d)**$\star$ z_{t-4} **e)** z_{t-5} **f)** z_{t-6}

Problem 9. The left figure below is the sample ACF of a time series $\{z_t\}$ of length 200; it displays the ACF up to lag 23. The right figure below is a lagged scatterplot made from $\{z_t\}$ using a lag of k. That is, it is a plot showing the points (z_t, z_{t-k}). What value of k was used in making this plot?

a) 1 **b**) 3 **c**)$\star$ 5 **d**) 8 **e**) 10 **f**) 13





Problem 10. Suppose X is a random variable, and b and c are constants. Then $E(bX + c) =$ _________.

a) $b(EX)$ **b**) $b^2(EX) + c$ **c**)$\star$ $b(EX) + c$ **d**) $b^2(EX)$
e) $b^2(EX) + c^2$ **f**) $b^2 + c^2$ **g**) $b + c^2$ **h**) $b + c$

Problem 11. Suppose $\{Y_t\}$ and $\{X_t\}$ are time series, and you fit a regression model $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$. For β_1, the SAS output will display values for all of the following: an estimate, a standard error, a t-value, and a p-value. If the residuals exhibit strong serial correlation, one of these values is probably still reasonable, but the others could be way off. Which of the values is probably reasonable?

a)$\star$ estimate **b**) standard error **c**) t-value **d**) p-value

Problem 12. A process which is weakly stationary, but **not** strictly stationary ________.

a) has a constant mean and variance, but changes its ACF over time

b) has a constant ACF, but the mean changes over time.

c)⋆ has a constant mean, variance, and ACF, but changes its behavior in some other way over time

d) has a constant ACF, but the variance changes over time

e) has the same joint distribution of (z_t, z_{t+1}) for all t, but the distribution of z_t changes over time

f) has the same joint distribution of (z_t, z_{t+1}, z_{t+k}) for all t, but the joint distribution of (z_t, z_{t+1}) changes over time

g) has the same distribution of z_t for all t, but the joint distribution of (z_t, z_{t+1}) changes over time

Problem 13. Suppose we have three time series $\{Y_t\}$, $\{X_{1,t}\}$, $\{X_{2,t}\}$ of length 240, and we fit the model

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t$$

using standard regression (ordinary least squares). The figure below gives a time series plot of the residuals from this regression. Based on the given information, which one of the following statements is correct?

a) The Durbin-Watson test will **NOT** reject the null hypothesis in this case.

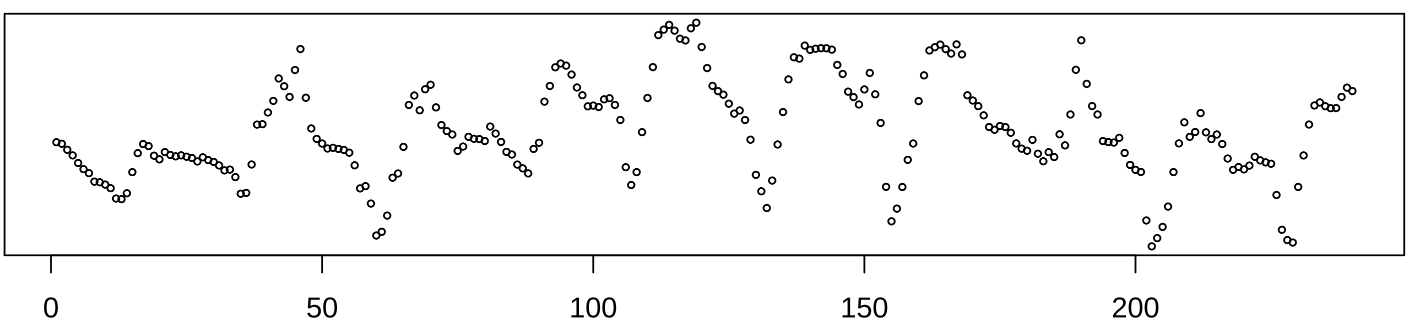
b) The residuals appear to be a random shock sequence.

c) Several observations have unusual response values.

d) Several observations have unusual covariate values.

e) The covariate $X_{2,t}$ is highly significant.

f)⋆ The residuals exhibit considerable serial correlation.



Problem 14. If we generate an AR(1) process $\{z_t\}$ with $|\phi_1| < 1$ starting from an initial value z_1, the process $\{z_t\}$ will always _________.

- **a)** reach its stationary behavior immediately
- **b)** reach its stationary behavior after a delay of one time period
- **c)** have all or nearly all of its values within a band of two standard errors about zero
- **d)** decay exponentially to zero
- **e)** have a cutoff to zero after lag one
- **f)**$\star$ converge to a stationary process after an initial transient phrase
- **g)** behave like random shocks after a delay of one time period

Problem 15. Which of the following is a correct expression for $\mathrm{Cov}(X, Y)$, the covariance between the random variables X and Y?

a) $\displaystyle \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2(Y_i - \overline{Y})^2$ **b)** $\overline{X}\,\overline{Y}$ **c)** $E(X - \mu_X)^2(Y - \mu_Y)^2$

d) $s_x^2 s_y^2$ **e)**$\star$ $E(X - \mu_X)(Y - \mu_Y)$ **f)** $E(X - \mu_X)\,E(Y - \mu_Y)$

g) $s_x s_y$ **h)** $\overline{X}^2\,\overline{Y}^2$ **i)** $\displaystyle \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 Y_i^2$

Problem 16. An ARMA(2,1) process is defined by _________

- **a)** $z_t = C + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
- **b)** $z_t = C + \phi_1 z_1 + \phi_2 z_2 + a_t - \theta_1 a_1$
- **c)** $z_t = C + \phi_1 z_1 + a_t - \theta_1 a_1 - \theta_2 a_2$
- **d)**$\star$ $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1}$
- **e)** $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} - \theta_1 a_{t-1}$
- **f)** $z_t = C + \phi_1 z_{t-1} - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
- **g)** $z_t = C + \phi_1 z_1 + \phi_2 z_2 - \theta_1 a_1$
- **h)** $z_t = C + \phi_1 z_1 - \theta_1 a_1 - \theta_2 a_2$

Problem 17. Attic temperature and wind speed have been observed at hourly intervals for 240 consecutive hours. The graph given below is a scatter plot of the attic temperature (Y) versus the wind speed (X). What would you expect **solely on the basis of this scatter plot?**

a)$\star$ In a regression of Y on X, the coefficient of X will be small and positive.

b) In a regression of Y on X, the coefficient of X will be large and negative.

c) The PACF of Y will cutoff to zero.

d) The PACF of X will cutoff to zero.

e) The ACF of Y will decay to zero slowly.

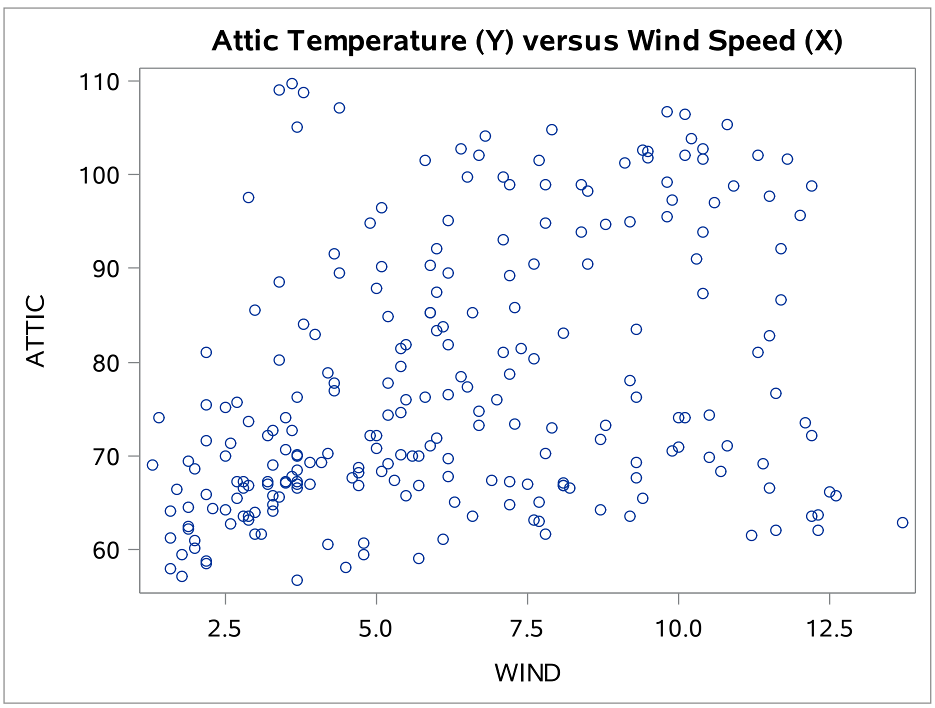
f) The ACF of Y will decay to zero rapidly.

g) The time series plot of X will exhibit a strong daily (24 hour) pattern.

h) The ACF of X will decay to zero slowly.

i) The ACF of X will decay to zero rapidly.

1



Problem 18. For the METHOD option in the ESTIMATE statement of PROC ARIMA, what is the preferred method of estimation when the random shocks are independent and approximately normally distributed with constant variance?

a) CLS **b)** ULS **c)**$\star$ ML **d)** AIC **e)** SBC **f)** OLS **g)** MINIC

Problem 19. Suppose you have a time series $z_1, z_2, \ldots, z_n$ which is very long (n is extremely large) and has been generated using

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \,.$$

If you fit the following regression model to this time series data

$$z_t = \beta_0 + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \beta_3 z_{t-3} + \beta_4 z_{t-4} + \varepsilon \,,$$

then $\hat{\beta}_4$ (the estimate of β_4) will be very close to _________.

a) 1 **b**) 0.5 **c**)$\star$ 0 **d**) σ_a^2 **e**) σ_z^2 **f**) C **g**) ϕ_1 **h**) ϕ_2

Problem 20. Suppose you fit a regression model and then construct a plot of the residuals versus fitted values (also known as the predicted values). In this plot the residuals are plotted on the y-axis and fitted values on the x-axis. If the regression assumptions are valid, we expect roughly that _________

a) the plot will follow a straight line
b) the fitted values will be normally distributed
c) the residuals will decay to zero
d)$\star$ the residuals will form a band centered at zero with constant vertical width
e) 95% of the fitted values will be between -2 and 2
f) positive residuals will tend to be followed by positive residuals
g) positive residuals will tend to be followed by negative residuals

Problem 21. For an AR(1) process $\{z_t\}$ with $C = 0$, one can show by repeated substitution that _________

a) $z_t = a_t + \phi_1 z_{t-1} + \phi_1^2 z_{t-2} + \phi_1^3 z_{t-3} + \phi_1^4 z_{t-4}$
b) $z_t = a_t - \phi_1 z_{t-1} - \phi_1^2 z_{t-2} - \phi_1^3 z_{t-3} - \phi_1^4 z_{t-4}$
c) $z_t = a_t - \phi_1 a_{t-1} - \phi_1^2 a_{t-2} - \phi_1^3 a_{t-3} - \phi_1^4 z_{t-4}$
d)$\star$ $z_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \phi_1^4 z_{t-4}$
e) $z_t = a_t + \phi_1 z_{t-1} - \phi_1 a_{t-1} + \phi_1^2 z_{t-2} - \phi_1^2 a_{t-2}$
f) $z_t = a_t - \phi_1 z_{t-1} + \phi_1 a_{t-1} - \phi_1^2 z_{t-2} + \phi_1^2 a_{t-2}$

Problem 22. For a given time series $z_1, z_2, \ldots, z_n$, the sample PACF value $\hat{\phi}_{33}$ may be approximated by ________

a) the ACF value r_3 when n is large

b) the ψ-weight ψ_3 when n is large

c) the AR coefficient ϕ_3 when n is large

d) the MA coefficient θ_3 when n is large

e)$\star$ the coefficient of z_{t-3} in a regression of z_t on $z_{t-1}, z_{t-2}, z_{t-3}$

f) the coefficient of a_{t-3} in a regression of z_t on $a_{t-1}, a_{t-2}, a_{t-3}$

g) the coefficient of a_{t-3} in a regression of a_t on $a_{t-1}, a_{t-2}, a_{t-3}$

Problem 23. Consider a regression model for a response Y and two covariates X_1 and X_2. For a random sample of n individuals, the model has the form:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i \,, \quad i = 1, \ldots, n \,.$$

Here is a list of statements:

1. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent.
2. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent of the responses $Y_1, \ldots, Y_n$.
3. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent of the covariates $(X_{1,i}, X_{2,i})$, $i = 1, \ldots, n$.
4. The errors are independent of the residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$.
5. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are $N(0, \sigma^2)$.

Which two of these statements are **not** assumptions of the standard regression model? (Circle the single correct choice below.)

a) 1, 2	**b)** 1, 3	**c)** 1, 4	**d)** 1, 5	**e)** 2, 3
f) 2, 4	**g)** 2, 5	**h)** 3, 4	**i)** 3, 5	**j)** 4, 5

The questions below involve the following situation:

Suppose we write an MA(5) process in the form

$$z_t = C + \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \psi_4 a_{t-4} + \psi_5 a_{t-5}$$

where $\psi_0 = 1$ and $\psi_i = -\theta_i$ for $i = 1, 2, 3, 4, 5$.

Consider the following list of responses:

a) $\dfrac{\sigma_a^2}{1-\psi_1}$ **b)** $\dfrac{\sigma_a^2}{1-\psi_1^2}$

c) $\dfrac{\sigma_a^2}{1-\psi_2}$ **d)** $\dfrac{\sigma_a^2}{1-\psi_2^2}$

e) $\dfrac{\sigma_a^2}{1-\psi_6}$ **f)** $\dfrac{\sigma_a^2}{1-\psi_6^2}$

g) $\dfrac{\sigma_a^2}{1-\psi_1-\psi_2-\psi_3-\psi_4-\psi_5}$ **h)** $\dfrac{\sigma_a^2}{1-\psi_1^2-\psi_2^2-\psi_3^2-\psi_4^2-\psi_5^2}$

i) 0 **j)** C

k) ψ_1^2 **l)** ψ_1^6

m) $\sigma_a^2(\psi_0\psi_1\psi_2\psi_3 + \psi_2\psi_3\psi_4\psi_5)$ **n)** $\sigma_a^2(\psi_0\psi_1\psi_2\psi_3\psi_4\psi_5)$

o) $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \psi_4^2 + \psi_5^2)$ **p)** $\sigma_a^2(\psi_0\psi_1\psi_2 + \psi_3\psi_4\psi_5)$

q) $\sigma_a^2(\psi_0\psi_2 + \psi_1\psi_3 + \psi_2\psi_4 + \psi_3\psi_5)$ **r)** $\sigma_a^2(\psi_0\psi_4 + \psi_1\psi_5)$

Problem 24. Which of these responses gives the value of μ_z?

a)	**b)**	**c)**	**d)**	**e)**	**f)**	**g)**	**h)**	**i)**
j)$\star$	**k)**	**l)**	**m)**	**n)**	**o)**	**p)**	**q)**	**r)**

Problem 25. Which of these responses gives the value of σ_z^2?

a)	**b)**	**c)**	**d)**	**e)**	**f)**	**g)**	**h)**	**i)**
j)	**k)**	**l)**	**m)**	**n)**	**o)**$\star$	**p)**	**q)**	**r)**

Problem 26. Which of these responses gives the value of $\gamma_2 = \text{Cov}(z_t, z_{t-2})$?

a)	**b)**	**c)**	**d)**	**e)**	**f)**	**g)**	**h)**	**i)**
j)	**k)**	**l)**	**m)**	**n)**	**o)**	**p)**	**q)**$\star$	**r)**

Problem 27. Which of these responses gives the value of $\gamma_6 = \text{Cov}(z_t, z_{t-6})$?

a)	**b**)	**c**)	**d**)	**e**)	**f**)	**g**)	**h**)	**i**)⋆
j)	**k**)	**l**)	**m**)	**n**)	**o**)	**p**)	**q**)	**r**)

Problem 28. Large values of the Studentized Residual (RStudent in SAS Proc Reg) indicate _________

- **a**) cases with unusual covariate values
- **b**) cases with a large influence
- **c**) cases with a large leverage
- **d**)⋆ cases with unusual response values
- **e**) the presence of serial correlation
- **f**) the presence of non-normality
- **g**) the presence of non-linearity
- **h**) the presence of non-constant variance

Problem 29. Suppose you use PROC ARIMA to fit a time series model to the series z_t and your code includes the statement

ESTIMATE P=(2,4);

What model are you fitting?

- **a**) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4}$
- **b**) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \phi_4 z_{t-4} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
- **c**)⋆ $z_t = C + \phi_2 z_{t-2} + \phi_4 z_{t-4} + a_t$
- **d**) $z_t = C + a_t - \theta_2 a_{t-2} - \theta_4 a_{t-4}$
- **e**) $z_t = C + \phi_2 z_{t-2} + a_t - \theta_4 a_{t-4}$
- **f**) $z_t = C + \phi_4 z_{t-4} + a_t - \theta_2 a_{t-2}$

Problem 30. Suppose we use SAS to fit two different ARMA models (call them #1 and #2) to a stationary time series $z_1, z_2, \ldots, z_{999}, z_{1000}$. Both models fit the data about equally well (that is, their likelihood values are very close), but Model #2 is much more complicated than Model #1 (that is, #2 has more parameters than #1). If you compare the AIC and SBC values for these models, which **one** of the following statements would you expect to be true?

Note: In the responses, the subscript denotes the model number; AIC is *Akaike's Information Criterion*; SBC is *Schwarz's Bayesian Criterion* (also called BIC); and $\approx$ means "approximately equal".

a) $AIC_1 < AIC_2$ and $SBC_1 > SBC_2$
b) $AIC_1 > AIC_2$ and $SBC_1 < SBC_2$
c)$\star$ $AIC_1 < AIC_2$ and $SBC_1 < SBC_2$
d) $AIC_1 > SBC_1$ and $AIC_2 > SBC_2$
e) $AIC_1 > AIC_2$ and $SBC_1 > SBC_2$
f) $AIC_1 \approx AIC_2$ and $SBC_1 \approx SBC_2$
g) $AIC_1 \approx SBC_1$ and $AIC_2 \approx SBC_2$

The last page of the exam contains four time series plots labeled A, B, C, D. Three of these plots show non-stationary series, and one is stationary. One of the non-stationary series has non-constant mean, one has non-constant variance, and one has non-constant ACF.

Problem 31. Which one of these series is stationary?

a) **b**) **c**) **d**)$\star$

Problem 32. Which one of these series has a non-constant mean?

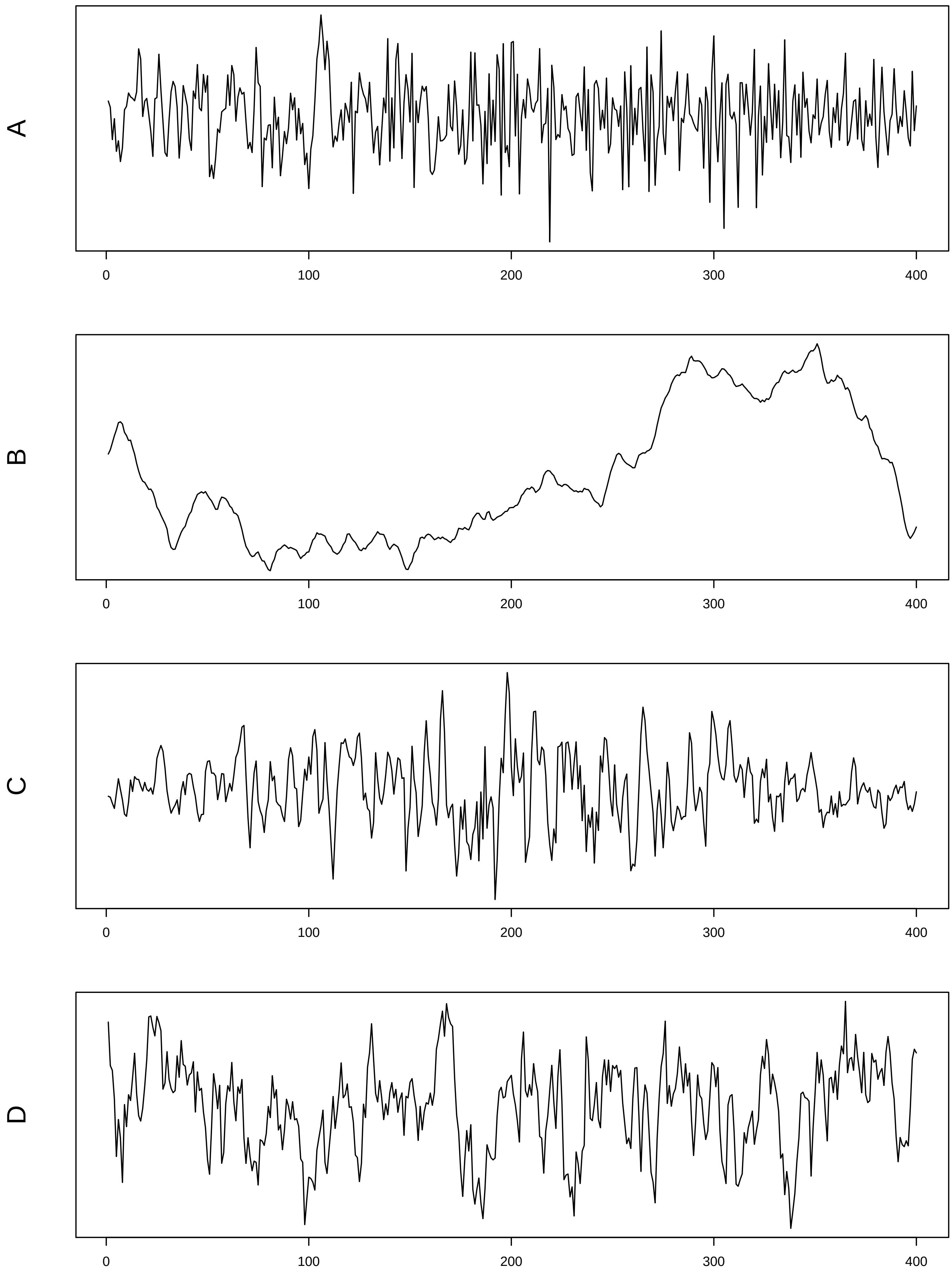
a) **b**)$\star$ **c**) **d**)

Problem 33. Which one of these series has a non-constant variance?

a) **b**) **c**)$\star$ **d**)

Problem 34. Which one of these series has a non-constant ACF?

a)$\star$ **b**) **c**) **d**)


A
0
100
200
300
400
B
0
100
200
300
400
C
0
100
200
300
400
D
0
100
200
300
400