TEST #1

STA 4853 **Name:**____________________________________

March 6, 2023

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

Directions

- This exam is **closed book** and **closed notes**.
- 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.)
- Each question is worth equal credit.
- There is no penalty for guessing.
- There are **33** multiple choice questions.
- The exam has **15** pages.

Problem 1. Suppose there is a very large population of N individuals, each with values of a response variable Y and p covariates $X_1, X_2, \ldots, X_p$. We randomly sample n individuals from this population. They have observed values of $(Y, X_1, \ldots, X_p)$ which we write as

$$(Y_i, X_{1,i}, X_{2,i}, \ldots, X_{p,i}), \quad i = 1, \ldots, n.$$

For a linear regression model of Y on $X_1, X_2, \ldots, X_p$ written in our usual notation, which ONE of the following is NOT typically assumed?

- **a)** $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are $N(0, \sigma^2)$.
- **b)** The errors ε are independent of the covariates $X_1, X_2, \ldots, X_p$ (i.e., ε is independent of X_1, ε is independent of X_2, ε is independent of X_3), etc.
- **c)**$\star$ The covariates $X_1, X_2, \ldots, X_p$ are independent (i.e., X_1 and X_2 are independent, X_1 and X_3 are independent, X_2 and X_3 are independent, etc.).
- **d)** $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + \varepsilon_i$, $i = 1, 2, \ldots, n$.
- **e)** $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent (i.e., ε_1 and ε_2 are independent, ε_1 and ε_3 are independent, ε_2 and ε_3 are independent, etc.)

Problem 2. In a regression model for data with p covariates and a sample of size n, which of the following would be considered an example of a large or unusual value of a Studentized Residual (RStudent)? That is, a case with this value of RStudent might lead you to investigate further to determine what is going on. (Is this case an outlier? typographical error? Etc.)

a) $3/\sqrt{n}$ **b)**$\star$ -3.9 **c)** $2(n-(p+1))/n$ **d)** $3(p+1)/n$ **e)** 1.3 **f)** 0

Problem 3. For an AR(p) process, the _________ has (or have) a cutoff to zero after lag p; the last nonzero value is at lag p, and it is exactly zero for lags greater than p.

What words correctly complete the sentence above?

a) sample ACF **b)** white noise probabilities **c)** chi-square values
d)$\star$ theoretical PACF **e)** sample PACF **f)** theoretical ACF

Problem 4. Suppose $\{z_t\}$ is a stationary ARMA process and $\{a_t\}$ is the sequence of random shocks used to generate $\{z_t\}$. In the list of choices below, all of them **except one** are definitely equal to zero, but one of them might **not** be. Which one is that?

a) $E(a_5 a_7)$ **b)** $E(a_5 a_3)$ **c)** $E(a_5 z_3)$ **d)**$\star$ $E(a_5 z_7)$ **e)** $E(z_5 a_7)$

Problem 5. Suppose you have used SAS to estimate (fit) several time series models which all have acceptable residual diagnostics. Which of the following is the name of a statistic or test you can use to compare and choose among them?

a)$\star$ AIC **b)** Durbin-Watson **c)** Ljung-Box Q **d)** t-value **e)** Cook's D **f)** OLS

Problem 6. If $\rho_j = 0$ for all $j \geq k$, then an approximate standard error for r_k is given by ________.

a) $\left(1 - \sum_{j=1}^{k-1} r_j\right)^{-1/2} n^{1/2}$ b) $\left(1 - 2\sum_{j=1}^{k-1} r_j\right)^{-1/2} n^{1/2}$

c) $\left(1 - 2\sum_{j=1}^{k-1} r_j^2\right)^{1/2} n^{1/2}$ d)$\star$ $\left(1 + 2\sum_{j=1}^{k-1} r_j^2\right)^{1/2} n^{-1/2}$

e) $\left(1 + \sum_{j=1}^{k-1} r_j^2\right)^{-1/2} n^{-1/2}$ f) $\left(1 + \sum_{j=1}^{k-1} r_j\right)^{1/2} n^{-1/2}$

Problem 7. When you fit a regression model $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$ to data, the statistical software will compute residuals (and many other quantities). The residuals are estimates of the ________.

a) standard errors $\mathrm{SE}(\hat{\beta}_i)$ b) predicted values c) $\hat{\beta}_i - \hat{\beta}$

d) Root MSE $\hat{\sigma}$ e) P-values f) covariates

g) fitted values h) true responses Y i)$\star$ true errors ε

Problem 8. The Ljung-Box test statistic $Q(12)$ computed on the raw data appears in the SAS output produced by the IDENTIFY statement of PROC ARIMA in the table entitled "**Autocorrelation Check for White Noise**" in the row labeled **12**. The purpose of this statistic is to test the null hypothesis that ________.

a) the ACF is zero at lag 12, that is, $\rho_{12} = 0$

b) the PACF is zero at lag 12, that is, $\phi_{12,12} = 0$

c) all the sample autocorrelations r_k lie inside the "two standard error" band after lag 12

d) all the sample autocorrelations $r_1, r_2, \ldots, r_{12}$ lie inside the "two standard error" band

e)$\star$ the first 12 autocorrelations are zero, that is, $\rho_1 = \rho_2 = \cdots = \rho_{12} = 0$

f) there is a cutoff in the ACF after lag 12, that is, $\rho_k = 0$ for all $k > 12$

g) there is a cutoff in the PACF after lag 12, that is, $\phi_{kk} = 0$ for all $k > 12$

Problem 9. If you fit a linear regression model for a response Y on two covariates X_1, X_2, the residual for the i^{th} case in the data set can be computed using the formula ________.

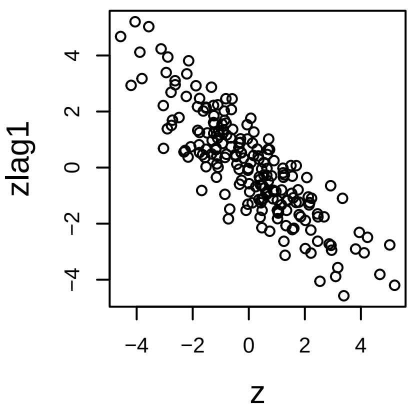
a) $(\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2}) - Y_i$ b) $(\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2}) - \hat{Y}_i$

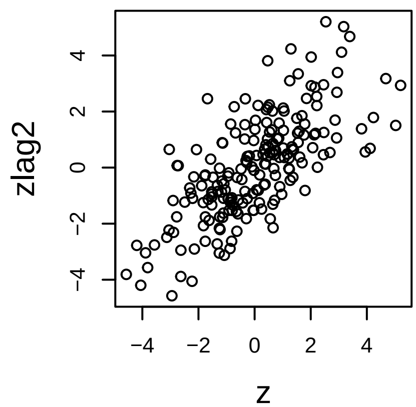
c)$\star$ $Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2})$ d) $\hat{Y}_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2})$

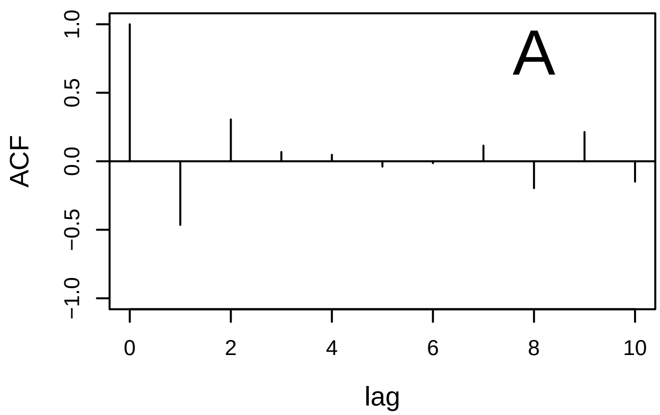
e) $Y_i + (\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2})$ f) $\hat{Y}_i + (\hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2})$

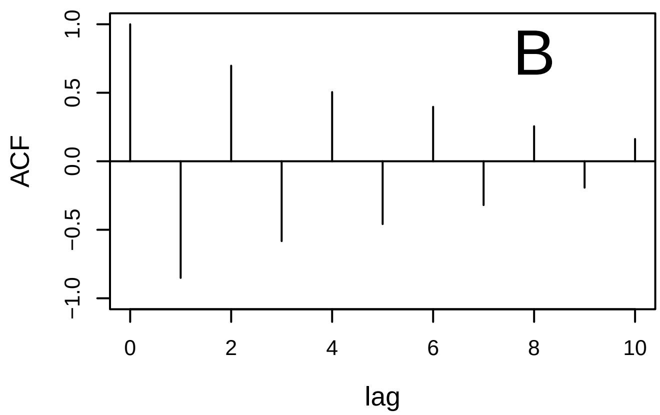
Problem 10. A time series $z_1, z_2, \ldots, z_{200}$ has been used to construct two lagged scatter plots which are the top two plots given below. The left plot (zlag1 versus z) displays the points (z_t, z_{t-1}). The right plot (zlag2 versus z) displays the points (z_t, z_{t-2}). Below these plots are four sample autocorrelation function (ACF) plots labeled A, B, C, D which are given up to lag 10. One of these sample ACF plots is for the series z_t. Which is it?

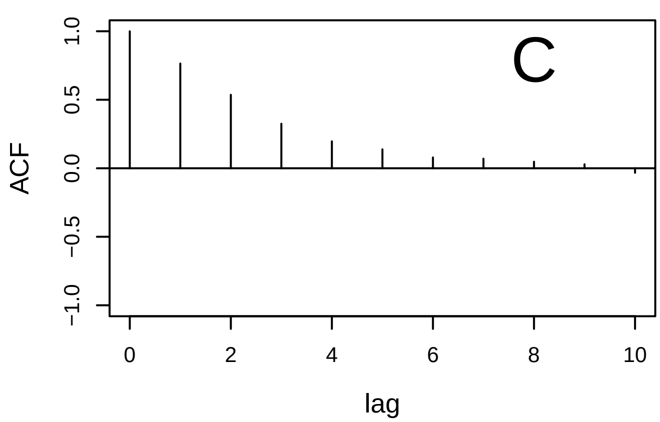
a) A **b**)$\star$ B **c**) C **d**) D

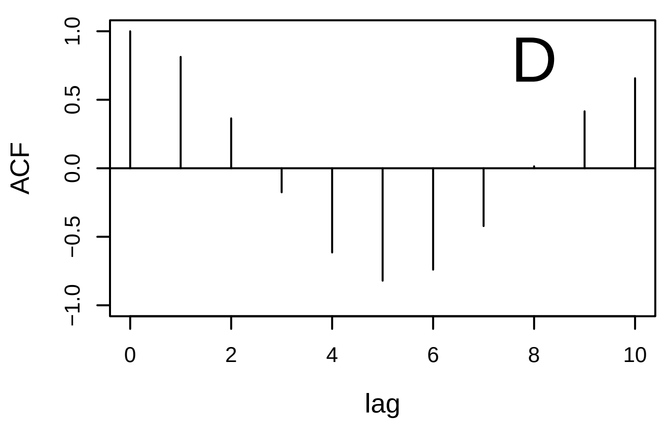












Problem 11. The sample ACF plot produced by SAS PROC ARIMA has a band marked at two standard errors (at $\pm 2s(r_k)$). If $\rho_j = 0$ for all $j \geq 5$, then the spike in the sample ACF at lag 5 will be **inside** the band with probability approximately _________.

a) **.**20 **b**) **.**10 **c**) **.**05 **d**) **.**025 **e**) **.**01

f) **.**99 **g**) **.**975 **h**)$\star$ **.**95 **i**) **.**90 **j**) **.**80

Problem 12. For a random sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, the formula for the **sample covariance** $c(X, Y)$ is ________.

a) $\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2$

b) $\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2(Y_i - \overline{Y})^2$

c)$\star$ $\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$

d) $\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2(Y_i - \overline{Y})^2}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$

e) $\dfrac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\cdot\overline{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right)}}$

f) $\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$

Problem 13. Suppose you have a time series $z_1, z_2, \ldots, z_n$ which is a very long (i.e., n is very large) realization of a stationary ARMA process. You fit a regression of z_t on $z_{t-1}, z_{t-2}, z_{t-3}, z_{t-4}$ by creating a SAS data set containing `z` (the time series $\{z_t\}$) and the lagged variables `zlag1, zlag2, zlag3, zlag4` and then using SAS PROC REG to fit the regression model

$$\texttt{z} = \beta_0 + \beta_1\,\texttt{zlag1} + \beta_2\,\texttt{zlag2} + \beta_3\,\texttt{zlag3} + \beta_4\,\texttt{zlag4} + \varepsilon$$

and obtain the estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$. After this, you use SAS PROC ARIMA to find the sample PACF $\hat{\phi}_{11}, \hat{\phi}_{22}, \ldots$ of the series $\{z_t\}$. When you compare $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ and $\hat{\phi}_{11}, \hat{\phi}_{22}, \hat{\phi}_{33}, \hat{\phi}_{44}$, what should you expect to see? (Choose the **single** best response. Note that $x \approx y$ means "x and y are nearly equal".)

a) $\hat{\beta}_1 \approx \hat{\phi}_{11}$ **b)** $\hat{\beta}_2 \approx \hat{\phi}_{22}$ **c)** $\hat{\beta}_3 \approx \hat{\phi}_{33}$ **d)**$\star$ $\hat{\beta}_4 \approx \hat{\phi}_{44}$

e) All of (a) through (d) **f)** None of (a) through (d)

Problem 14. In SAS PROC ARIMA, the code `ESTIMATE P=3;` will estimate the parameters of which of the following models?

a) $z_t = C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$

b) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$

c)$\star$ $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + a_t$

d) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1}$

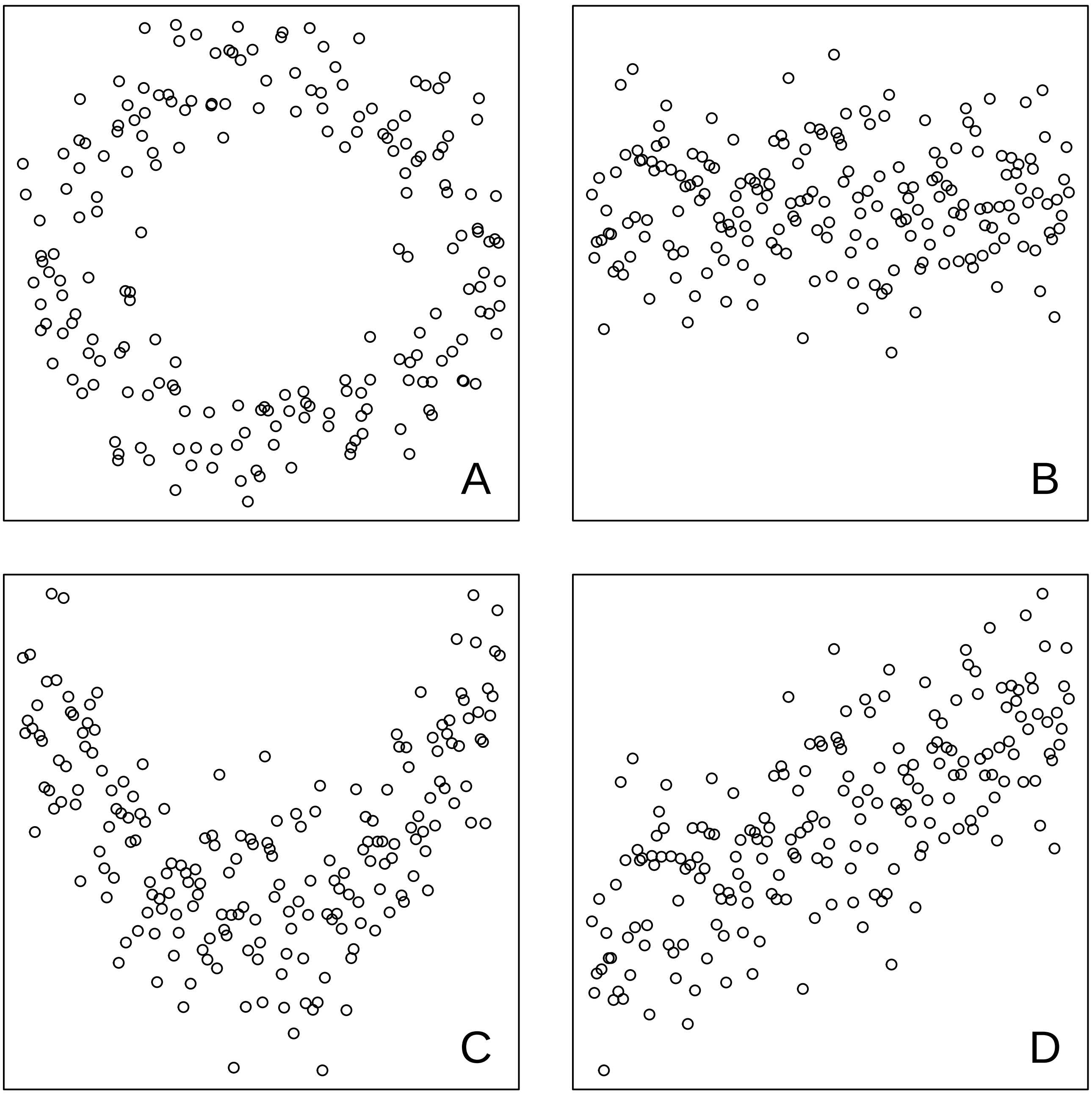
e) $z_t = C + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

f) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

Problem 15. There are four different populations named A, B, C, D. A random sample of 200 people was drawn from each population and the values of X and Y (e.g., height and weight) were recorded for each person. Scatterplots of (X, Y) are given below for each of the four samples, labeled by the population from which they were drawn. In **at least one** of these populations X and Y are **independent**. Select the single response which correctly specifies the population or populations in which X and Y are independent.

a) A **b**)⋆ B **c**) C **d**) D **e**) A, B

f) A, C **g**) A, D **h**) B, C **i**) B, D **j**) C, D



Problem 16. For a regression model which includes a covariate X_i, we use the notation β_i and $\widehat{\beta}_i$ for the corresponding parameter and its estimate, and t_i for the t-value.

A Table of Parameter Estimates produced by SAS PROC REG is given below. If the regression assumptions are valid in this situation, what is the correct interpretation of the P-value 0.0193 for the covariate X_3 in the last row of this table?

a) When $\beta_3 = 0.13317$, the probability that $|t_3|$ is at least 2.38 is 0.0193.

b) When $\beta_3 = 0.13317$, the probability that $|\widehat{\beta}_3|$ is at least 0.13317 is 0.0193.

c) When $\beta_3 = 0$, the probability that $|t_3|$ is less than 2.38 is 0.0193.

d) When $\beta_3 = 0$, the probability that $|\widehat{\beta}_3|$ is less than 0.13317 is 0.0193.

e) When $\beta_3 = 0.13317$, the probability that $|t_3|$ is less than 2.38 is 0.0193.

f) When $\beta_3 = 0.13317$, the probability that $|\widehat{\beta}_3|$ is less than 0.13317 is 0.0193.

g)$\star$ When $\beta_3 = 0$, the probability that $|t_3|$ is at least 2.38 is 0.0193.

h) When $\beta_3 = 0$, the probability that $|\widehat{\beta}_3|$ is at least 0.13317 is 0.0193.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-62.01022	21.11760	-2.94	0.0042
x1	1	-1.56231	0.05560	-28.10	<.0001
x2	1	1.53837	0.05795	26.55	<.0001
x3	1	0.13317	0.05595	2.38	0.0193

Problem 17. Suppose that z_t is an AR(1) process generated by

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

By making appropriate substitutions one may show that z_t also satisfies __________.

a)$\star$ $z_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 z_{t-3}$

b) $z_t = a_t - \phi_1 a_{t-1} - \phi_1^2 a_{t-2} - \phi_1^3 z_{t-3}$

c) $z_t = a_t - \theta_1 a_{t-1} - \theta_1^2 a_{t-2} - \theta_1^3 z_{t-3}$

d) $z_t = a_t + \theta_1 a_{t-1} + \theta_1^2 a_{t-2} + \theta_1^3 z_{t-3}$

e) $z_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3}$

f) $z_t = a_t - \phi_1 a_{t-1} - \phi_1^2 a_{t-2} - \phi_1^3 a_{t-3}$

g) $z_t = a_t - \theta_1 a_{t-1} - \theta_1^2 a_{t-2} - \theta_1^3 a_{t-3}$

h) $z_t = a_t + \theta_1 a_{t-1} + \theta_1^2 a_{t-2} + \theta_1^3 a_{t-3}$

Problem 18. Suppose we use SAS PROC ARIMA to fit an ARMA model to a time series $\{z_t\}$. For this model, the residual diagnostics all look good and the table of parameter estimates is given below. Based on this information, what would we usually do in this situation and why?

a)$\star$ We would drop the term labeled AR1,**3** because its P-value is large.
b) We would retain the term labeled AR1,**3** because the residual diagnostics are good.
c) We would drop the term labeled AR1,**3** because dropping it would increase the AIC.
d) We would retain the term labeled AR1,**3** because dropping it would decrease the AIC.
e) We would drop the term labeled AR1,**2** because its P-value is small.
f) We would drop the term labeled AR1,**2** because the residual diagnostics are good.
g) We would retain the term labeled AR1,**2** because dropping it would decrease the AIC.
h) We will retain the term labeled AR1,**2** because dropping it would improve the residual diagnostics.

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	14.51694	0.37734	38.47	<.0001	0
AR1,1	0.65760	0.10232	6.43	<.0001	1
AR1,2	-0.41047	0.11703	-3.51	0.0005	2
AR1,3	-0.05807	0.10408	-0.56	0.5769	3

Problem 19. If X, Y, Z are random variables and b, c, d are constants, then

$$\mathrm{Var}(bX + cY + Z + d) = \underline{\hspace{2cm}}.$$

a) $b^2\mathrm{Var}(X) + c^2\mathrm{Var}(Y) + d^2$
b) $b\mathrm{Var}(X) + c\mathrm{Var}(Y) + \mathrm{Var}(Z)$
c) $b\mathrm{Var}(X) + c\mathrm{Var}(Y) + \mathrm{Var}(Z) + d$
d)$\star$ $b^2\mathrm{Var}(X) + c^2\mathrm{Var}(Y) + \mathrm{Var}(Z)$
e) $b\mathrm{Var}(X) + c\mathrm{Var}(Y)$
f) $b^2\mathrm{Var}(X) + c^2\mathrm{Var}(Y) + \mathrm{Var}(Z) + d^2$

Problem 20. An ARMA(2,1) process is defined by _________

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

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 21. Which of these responses gives the value of μ_z?

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

Problem 22. 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)**⋆ **p)** **q)** **r)**

Problem 23. 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)**⋆ **r)**

Problem 24. 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 25. Suppose a population of N individuals has values of X and Y given by (X_1, Y_1), $(X_2, Y_2,)$, ..., (X_N, Y_N). If the population correlation ρ between X and Y is exactly equal to 1, then the N points $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ ________.

- **a)** are very close to a line with slope exactly equal to one
- **b)** are very close to a line with positive slope
- **c)** are very close to the regression line $Y = \beta_0 + \beta_1 X + \varepsilon$
- **d)** lie exactly on a line with slope equal to one
- **e)**⋆ lie exactly on a line with positive slope
- **f)** lie in a band of constant width centered about zero.
- **g)** lie in a band of constant width centered about one.

Problem 26. Suppose that

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1}$$

is a stationary process with mean μ_z. Then the mean centered process $\tilde{z}_t = z_t - \mu_z$ satisfies ________.

- **a)** $\tilde{z}_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \tilde{a}_t - \theta_1 \tilde{a}_{t-1}$
- **b)** $\tilde{z}_t = C + \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \tilde{a}_t - \theta_1 \tilde{a}_{t-1}$
- **c)**⋆ $\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t - \theta_1 a_{t-1}$
- **d)** $\tilde{z}_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \tilde{a}_t - \theta_1 \tilde{a}_{t-1}$
- **e)** $\tilde{z}_t = C/(1 - \phi_1 - \phi_2)$
- **f)** $\tilde{z}_t = C/(1 - \phi_1 - \phi_2 + \theta_1)$
- **g)** $\tilde{z}_t = C/(1 - \phi_1 - \phi_2 - \theta_1)$
- **h)** $\tilde{z}_t = C/(1 + \phi_1 + \phi_2)$

Problem 27. For a random shock process, the theoretical ACF ________

- **a)** decays rapidly to zero
- **b)** will have nearly all values within the two standard error band
- **c)** will have an exact cutoff after lag 1
- **d)** is non-stationary
- **e)** is independent of future values
- **f)**⋆ is exactly zero for all nonzero lags
- **g)** will have an approximate cutoff after lag 1

Problem 28. A stationary AR(1) process

$$z_t = C + \phi_1 z_{t-1} + a_t$$

has $\mathrm{Var}(z_t) =$ ________.

a) $\dfrac{\sigma_a^2}{1 - \phi_1}$ **b)** $\dfrac{\sigma_a}{1 - \phi_1}$ **c)** $\dfrac{C}{1 - \phi_1}$ **d)** $\dfrac{C^2}{1 - \phi_1^2}$ **e)** $\dfrac{C}{1 - \phi_1^2}$ **f)**⋆ $\dfrac{\sigma_a^2}{1 - \phi_1^2}$

Problem 29. Suppose you are analyzing a time series $z_1, z_2, \ldots, z_n$ using SAS PROC ARIMA. In the output produced by the IDENTIFY statement, the values of the Ljung-Box test statistics $Q(6)$, $Q(12)$, $Q(18)$, $Q(24)$ are given along with their corresponding P-values. If $\{z_t\}$ is actually just a random shock sequence, then we expect that _________

a) most of the ACF values will lie outside the band

b) most of the PACF values will lie outside the band

c) $H_0 : \rho_1 = \rho_2 = \cdots = \rho_6 = 0$ will be rejected

d) $H_0 : Q(6) = Q(12) = Q(18) = Q(24) = 0$ will be rejected

e)$\star$ all four P-values will be $> .05$

f) all four P-values will be $\leq .05$

g) $Q(6)$, $Q(12)$, $Q(18)$, $Q(24)$ will all be large

Problem 30. In PROC ARIMA the statement

ESTIMATE P=(2,4) ;

will fit which of the following models?

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) $z_t = C + \phi_2 z_{t-2} + a_t - \theta_4 a_{t-4}$

d) $z_t = C + \phi_4 z_{t-4} + a_t - \theta_2 a_{t-2}$

e) $z_t = C + a_t - \theta_2 a_{t-2} - \theta_4 a_{t-4}$

f)$\star$ $z_t = C + \phi_2 z_{t-2} + \phi_4 z_{t-4} + a_t$

Problem 31. In regression, large values of H (the leverage) identify _________

a) influential covariates in the model

b) covariates which can be dropped

c) covariates with serial correlation

d) covariates which should be retained

e) serial correlation in the residuals

f) cases with unusual response values

g) influential cases in the data

h)$\star$ cases with unusual covariate values

Problem 32. This and the following pages give some SAS output for a time series $\{z_t\}$ of length $n = 400$. This page gives the time series plot, the following two pages give the usual IDENTIFY statement output for the first and second half of the series (i.e., the first 200 and last 200 observations, respectively), and the last page gives histograms for the first and second half of the series. This series belongs to one of the five categories given below. Which is it?

a) Does not have a constant mean

b) Does not have a constant variance

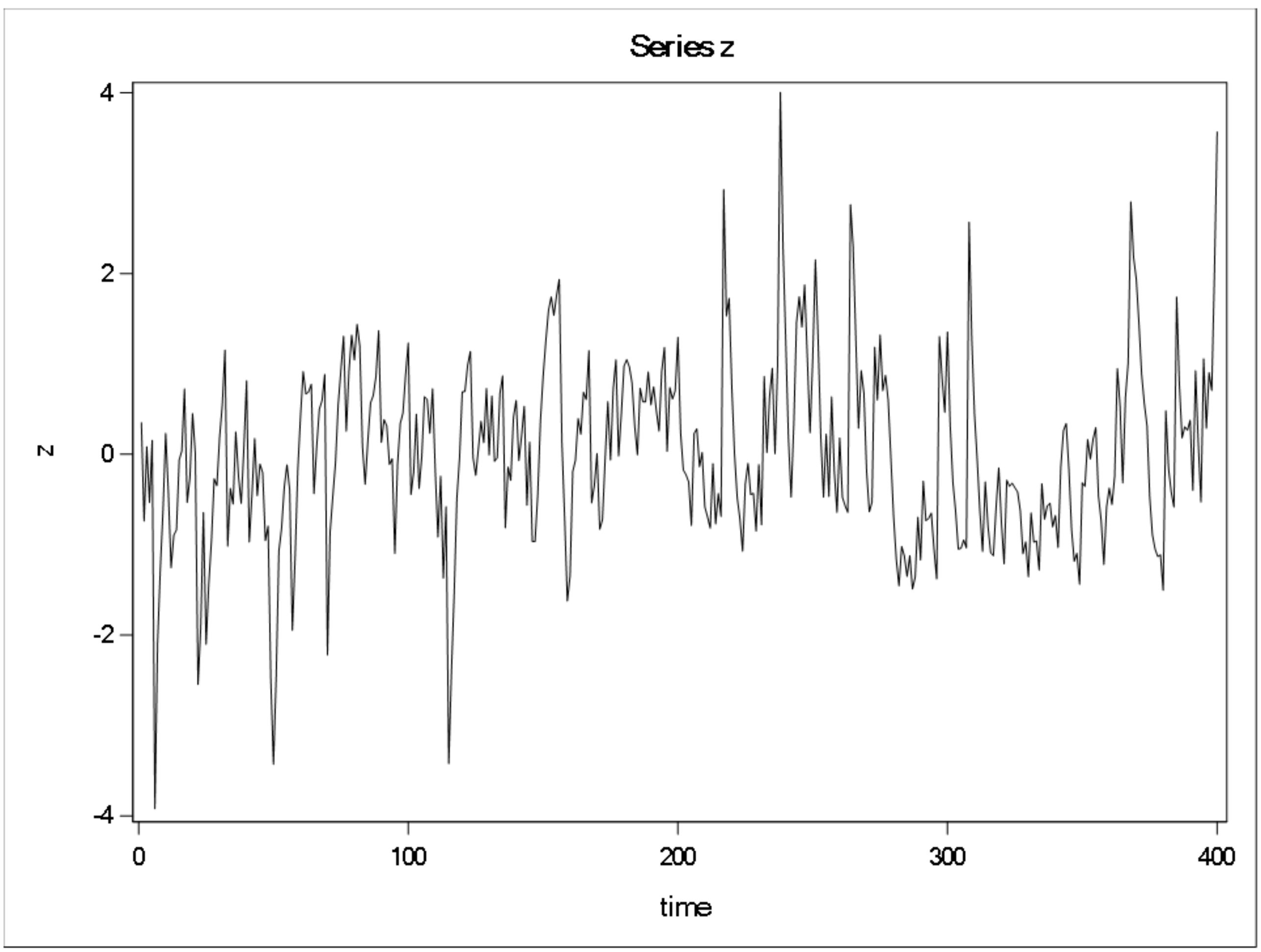
c) Does not have a constant ACF

d) Strictly stationary

e)⋆ Is weakly stationary, but not strictly stationary

Problem 33. If you were to look **only** at the following page of output which has the IDENTIFY statement output for the first half of the data, what ARMA process would you say generated this half of the series?

a)⋆ AR(1)	**b**) AR(2)	**c**) ARMA(1,1)	**d**) ARMA(3,1)	**e**) ARMA(4,1)
f) MA(1)	**g**) MA(2)	**h**) random shocks	**i**) ARMA(1,3)	**j**) ARMA(1,4)

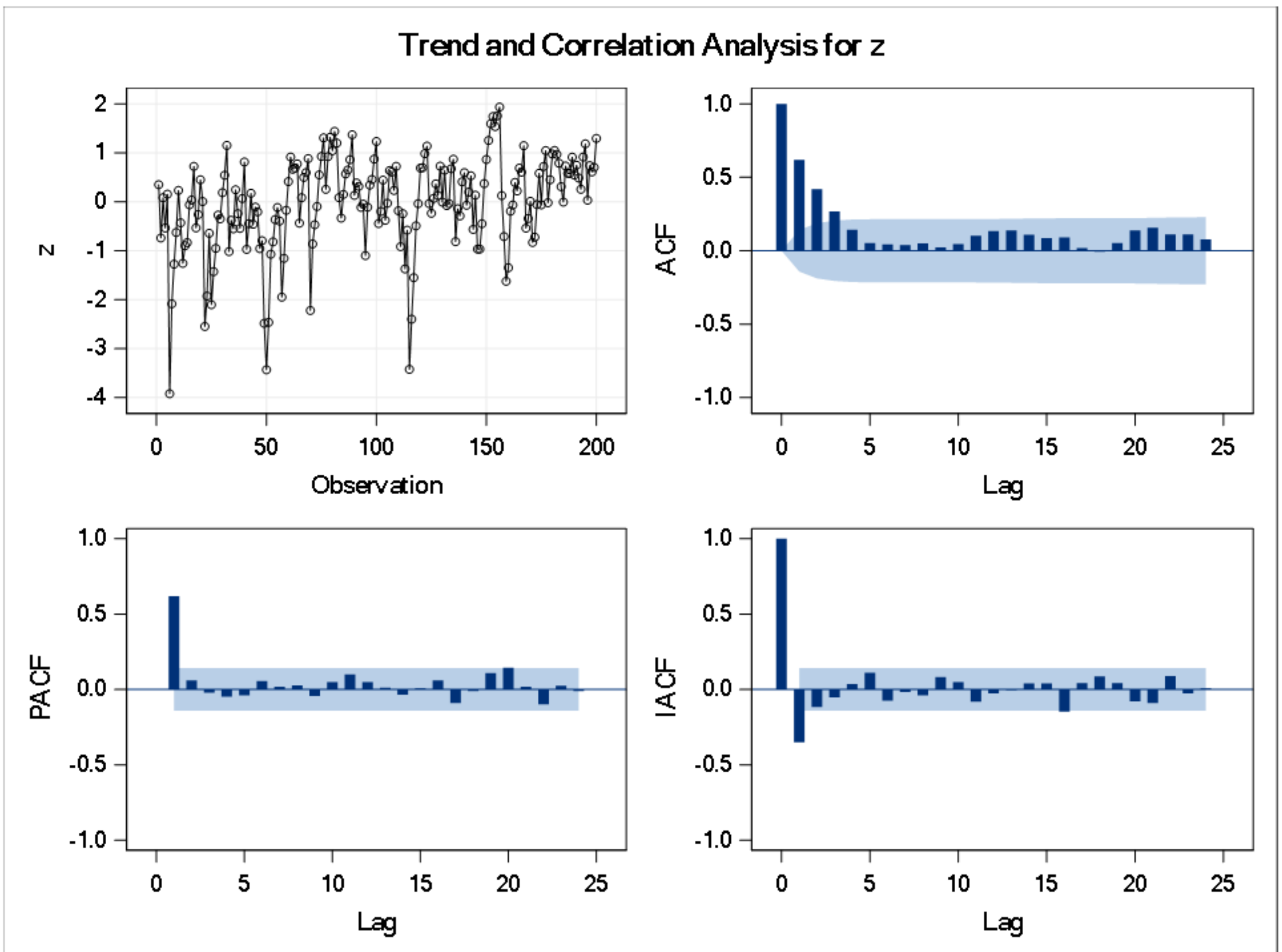


Series z -- first half

The ARIMA Procedure

Name of Variable = z	
Mean of Working Series	-0.01909
Standard Deviation	0.966402
Number of Observations	200

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	133.78	6	<.0001	0.619	0.420	0.268	0.143	0.053	0.043
12	141.21	12	<.0001	0.039	0.050	0.024	0.046	0.102	0.133
18	151.53	18	<.0001	0.138	0.110	0.087	0.092	0.018	-0.007
24	169.00	24	<.0001	0.052	0.138	0.157	0.112	0.111	0.076

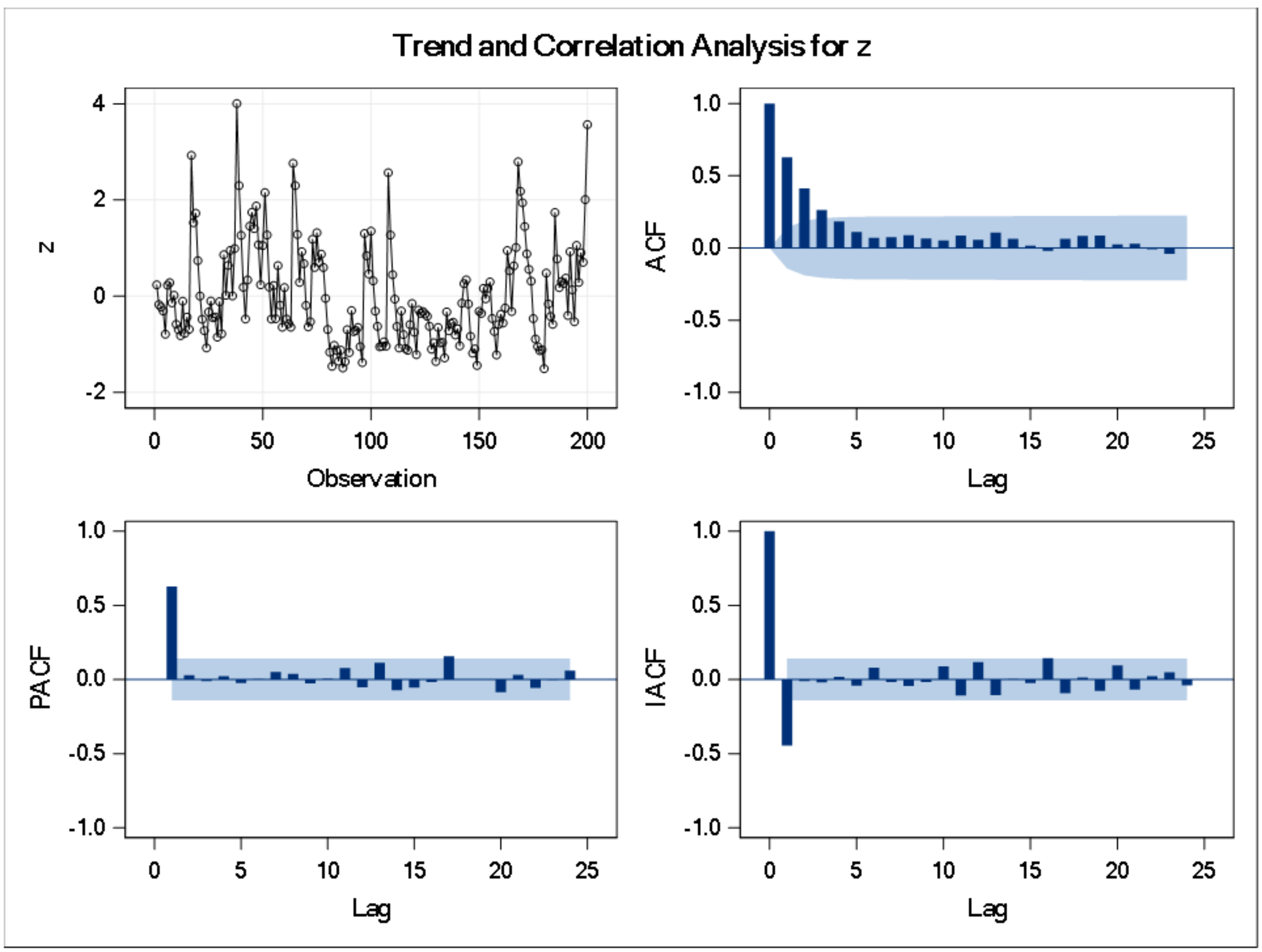


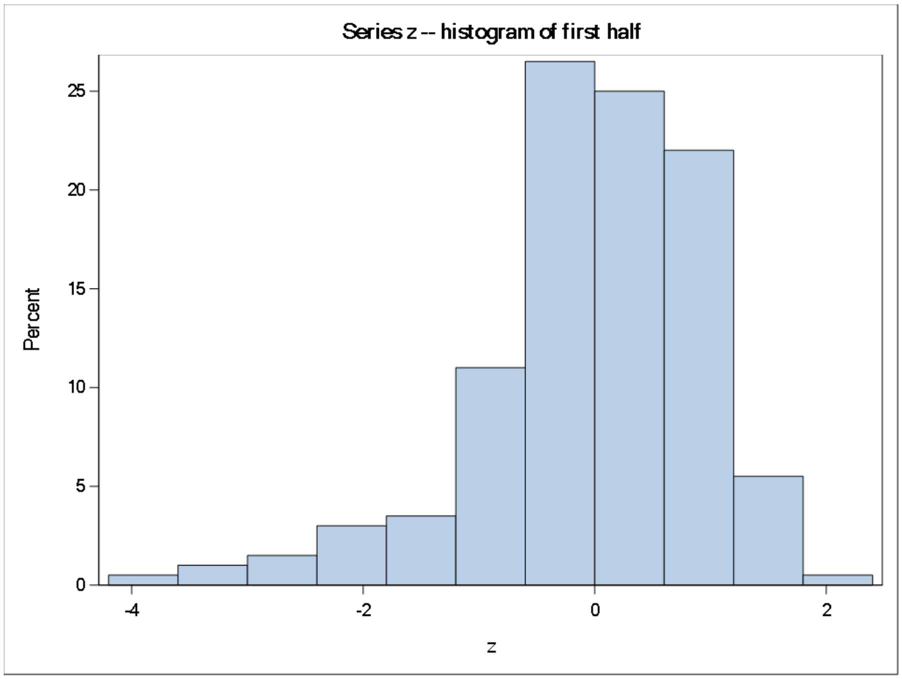
Series z -- second half

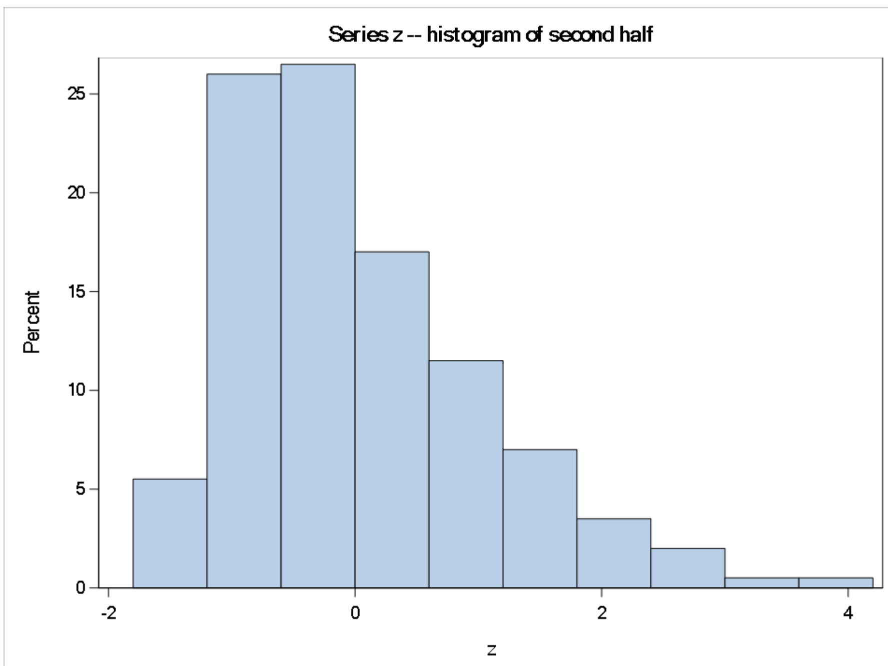
The ARIMA Procedure

Name of Variable = z	
Mean of Working Series	0.019091
Standard Deviation	1.029727
Number of Observations	200

Autocorrelation Check for White Noise									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	139.64	6	<.0001	0.628	0.413	0.264	0.183	0.111	0.070
12	146.22	12	<.0001	0.074	0.088	0.066	0.052	0.086	0.057
18	152.14	18	<.0001	0.106	0.064	0.016	-0.019	0.063	0.085
24	154.55	24	<.0001	0.086	0.025	0.030	-0.008	-0.041	-0.003




Series z -- histogram of first half
Percent
25
20
15
10
5
0
-4
-2
0
2
z


Series z -- histogram of second half
Percent
25
20
15
10
5
0
-2
0
2
4
z