TEST #1

STA 4853 **Name:**____________________________________

March 3, 2014

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 31 questions. Three of them are "fill in the blank." The rest are multiple choice.
- Circle a **single** answer for each multiple choice question. Your choice should be made clearly.
- On the multiple choice questions, 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 **14** pages.
- Each question is worth equal credit.

The following information applies to the next 2 problems:

SAS PROC ARIMA gives a plot of the sample ACF along with a band. For an MA(**4**) process, which choices make the following statement true:

The spike in the sample ACF at ☐ will lie **inside** the band about ☐ of the time.

Problem 1. Choices for Box 1: (Circle the correct choice)

a) lag 1 **b)** lag 2 **c)** lag 3 **d)** lag 4 **e)** lag 5

Problem 2. Choices for Box 2: (Circle the correct choice)

a) 1% **b)** 2% **c)** 5% **d)** 10% **e)** 90% **f)** 95% **g)** 98% **h)** 99%

Problem 3. If b, c, and d are constants, and X and Y are independent random variables, then $\text{Var}(bX + cY + d) =$

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

Problem 4. The theoretical **Inverse ACF (IACF)** of an **AR**(p) process ...

a) is the same as the **PACF** of an **MA**(p) process.
b) is the same as the **PACF** of an **AR**(p) process.
c) is the same as the **ACF** of an **AR**(p) process.
d) is the same as the **IACF** of an **MA**(p) process.
e) is the same as the **ACF** of an **MA**(p) process.
f) is the same as the **IACF** of an **MA**(q) process.

Problem 5. The equation

$$(1 - 0.5B + 0.4B^2 - 0.3B^3 - 0.2B^4)z_t = 2.5 + (1 - 0.4B + 0.8B^2)a_t$$

describes a _________ process. (Circle the correct response.)

a) ARMA(5,3) **b)** ARMA(3,5) **c)** ARMA(4,2) **d)** ARMA(2,4)
e) MA(6) **f)** AR(8) **g)** MA(8) **h)** AR(6) **i)** mean-centered

Problem 6. For estimating the parameters of an ARMA process, if the shocks a_t are independent and approximately normally distributed with mean zero and constant variance, the preferred method of estimation is ________. (Circle the correct response.)

a) OLS **b)** ML **c)** CLS **d)** ULS **e)** AIC **f)** SBC **g)** the default

Problem 7. A population of N individuals has heights $X_1, X_2, \ldots, X_N$. Suppose we measure the heights of a random sample of n individuals and compute the sample variance s_x^2 for this sample. The quantity s_x^2 is an estimate of the population variance σ_x^2. Which of the following is a formula for the population variance σ_x^2?

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

b) $\dfrac{1}{N}\sum_{i=1}^{N}(X_i - \mu_x)^2$ where $\mu_x = \dfrac{1}{N}\sum_{i=1}^{N} X_i$

c) $\dfrac{c(X,X)}{s_x s_x}$ where $s_x = \sqrt{s_x^2}$

d) $\dfrac{\mathrm{Cov}(X,X)}{\sigma_x \sigma_x}$ where $\sigma_x = \sqrt{\sigma_x^2}$

e) $\dfrac{\sigma_z^2}{1-\phi_1^2}$ where σ_a^2 is the random shock variance

f) $\sigma_a^2 \sum_{i=0}^{q-k} \psi_i \psi_{i+k}$ where σ_a^2 is the random shock variance

g) $\dfrac{\sum_{t=1}^{n-k}(X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$

Problem 8. Suppose you know the values of a response variable Y and p covariates $X_1, X_2, \ldots, X_p$ for each of the individuals in a random sample of size n. You wish to use SAS PROC REG to fit the regression of Y on $X_1, X_2, \ldots, X_p$. SAS PROC REG requires that the data be arranged in a SAS data set with ________.

a) $p+1$ rows and n columns

b) p rows and n columns

c) $n+1$ rows and p columns

d) $n-1$ rows and p columns

e) n rows and $p+1$ columns

f) p rows and $n-1$ columns

Problem 9. The sample autocorrelation at lag k is denoted r_k. Which of the following is the formula for $s(r_k)$, the approximate standard error of r_k?

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

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

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

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

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

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

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

Problem 10. For a time series z_t, the expression $B^3 z_t$ means

a) z_{t+3} **b)** z_{3t} **c)** $3z_t$ **d)** $-3z_t$ **e)** z_{t-3} **f)** $z_t + 3$ **g)** $z_t - 3$

Problem 11. For a stationary AR(1) process, it is always true that $\rho_3 = \ldots$

a) $\phi_1 + \rho_2$ **b)** $\phi_1 - \rho_2$ **c)** $\phi_1^3 \rho_2$ **d)** $3\rho_1$ **e)** $\rho_1/3$ **f)** $\phi_1 \rho_2$ **g)** $\rho_2/3$

Problem 12. For a time series z_t, the expression $B^j B^k z_t$ is equal to

a) $z_{t-j} z_{t-k}$

b) z_{t+j+k}

c) $B^{j+k} z_t$

d) $z_{t+j} z_{t+k}$

e) $B^j z_t B^k z_t$

f) z_{jk}

g) z_{t-jk}

Problem 13. Which one of the following types of processes is always stationary, regardless of the values of its parameters? (In the responses below, assume that p, d and q are positive integers.)

a) MA(q) **b)** ARMA(p, q) **c)** ARIMA(p, d, q) **d)** random walk **e)** AR(p)

The following information applies to the next two problems.

Suppose that $\{z_t\}$ is a stationary ARMA process and $\{a_t\}$ is the sequence of random shocks used to generate $\{z_t\}$.

Problem 14. When is $E(a_s a_t) = 0$?

a) always **b)** never **c)** if $s = t$ **d)** if $s \neq t$ **e)** only if $s > t$ **f)** only if $s < t$

Problem 15. When is $E(z_s a_t) = 0$?

a) always **b)** never **c)** if $s = t$ **d)** if $s \neq t$ **e)** if $s > t$ **f)** if $s < t$

Problem 16. Suppose you have used SAS to estimate (fit) several 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) OLS **b)** AIC **c)** Durbin-Watson **d)** Ljung-Box Q **e)** t-value **f)** Cook's D

Problem 17. The expression ∇z_t means ...

a) $z_{t-1} - z_t$ **b)** $z_{t+1} - z_t$ **c)** $z_t - z_{t-1}$ **d)** $z_t - z_{t+1}$ **e)** $(B-1)z_t$ **f)** $(1+B)z_t$

Problem 18. If z_t is an ARIMA(0,2,1) process, then $\nabla^2 z_t$ is a __________ process. (Circle the correct response)

a) MA(1) **b)** MA(3) **c)** AR(2) **d)** AR(3) **e)** random shock
f) random walk **g)** ARIMA(2,2,1) **h)** ARIMA(0,2,3) **i)** ARMA(2,1) **j)** ARMA(1,2)

Problem 19. Which one of the following types of processes has no transient initial phase; it reaches its stationary behavior immediately. (In the responses below, assume that p, d and q are positive integers.)

a) ARMA(p, q) **b)** AR(p) **c)** MA(q)

The following information applies to the next three problems.

A process is generated by

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

Problem 20. What kind of process is this?

a) ARMA(1,1) **b)** ARMA(2,1) **c)** ARMA(1,2) **d)** AR(1)
e) AR(2) **f)** AR(3) **g)** MA(1) **h)** MA(2) **i)** MA(3)

Problem 21. What are the requirements for this process to be stationary?

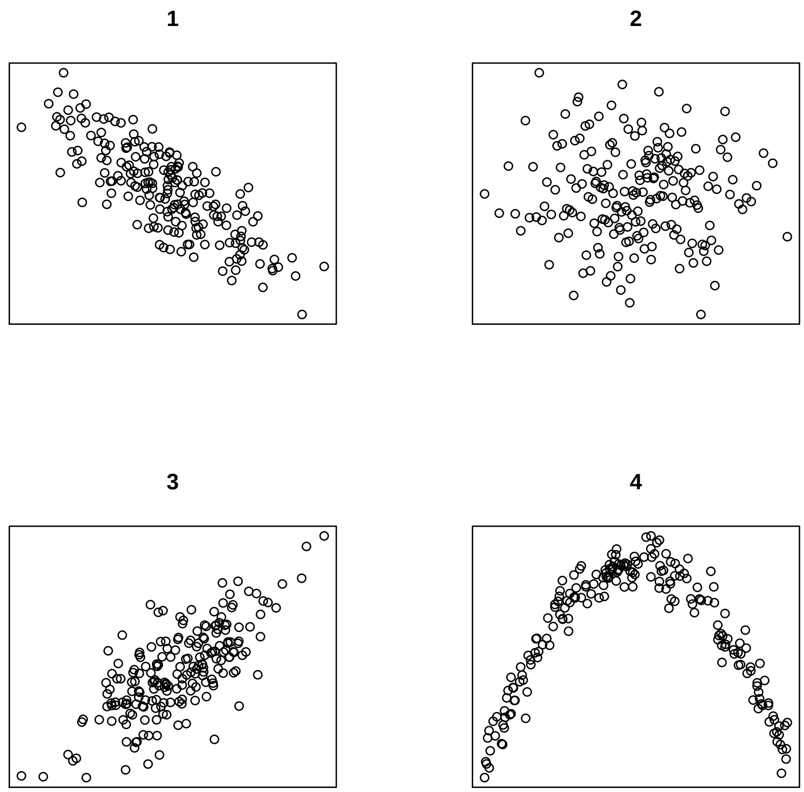
a) $|\phi_2| < 1 \,,\ \phi_2 + \phi_1 < 1 \,,\ \phi_2 - \phi_1 < 1 \,,\ |\theta_1| < 1$
b) $|\phi_1| < 1 \,,\ \phi_1 + \phi_2 < 1 \,,\ \phi_1 - \phi_2 < 1 \,,\ |\theta_1| < 1$
c) $|\phi_2| < 1 \,,\ \phi_2 + \phi_1 < 1 \,,\ \phi_2 - \phi_1 < 1$
d) $|\phi_1| < 1 \,,\ \phi_1 + \phi_2 < 1 \,,\ \phi_1 - \phi_2 < 1$
e) $|\phi_1| < 1 \,,\ |\theta_1| < 1$
f) $|\phi_2| < 1 \,,\ |\theta_1| < 1$
g) $|\phi_1| < 1$
h) $|\theta_1| < 1$

Problem 22. When the process is stationary, what is its mean μ_z?

a) $\dfrac{C}{1 - \phi_1 - \phi_2}$
b) $\dfrac{\sigma_a^2}{1 - \phi_1^2 - \phi_2^2}$
c) $\dfrac{C}{1 - \phi_1}$
d) $\dfrac{\sigma_a^2}{1 - \phi_1^2}$
e) $\dfrac{C}{1 - \theta_1}$
f) $\dfrac{\sigma_a^2}{1 - \theta_1^2}$
g) C

The following information applies to the next two problems.

The values of X and Y were observed for random samples of size $n = 200$ from four different populations (numbered 1, 2, 3, 4), and are displayed in the scatter plots given below.



Problem 23. In two of the populations, X and Y are **un**correlated. Which two are these?

a) 1, 2 **b)** 1, 3 **c)** 1, 4 **d)** 2, 3 **e)** 2, 4 **f)** 3, 4

Problem 24. In one of the populations, X and Y are independent. Which is it?

a) 1 **b)** 2 **c)** 3 **d)** 4

For a time series z_t of length $n = 10{,}000$, we used OLS to estimate the coefficients in four different regression models:

$$z_t \text{ on } z_{t-1}$$
$$z_t \text{ on } z_{t-1}, z_{t-2}$$
$$z_t \text{ on } z_{t-1}, z_{t-2}, z_{t-3}$$
$$z_t \text{ on } z_{t-1}, z_{t-2}, z_{t-3}, z_{t-4}$$

Excerpts from the regression output are given below. In this output the lagged variables are named `zlag1`, `zlag2`, `zlag3`, `zlag4`.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.00060091	0.02325	0.03	0.9794
zlag1	1	0.49999	0.00866	57.72	<.0001

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.00114	0.02202	0.05	0.9588
zlag1	1	0.66005	0.00947	69.68	<.0001
zlag2	1	-0.32022	0.00947	-33.80	<.0001

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.00090301	0.02137	0.04	0.9663
zlag1	1	0.73775	0.00971	76.01	<.0001
zlag2	1	-0.48000	0.01121	-42.84	<.0001
zlag3	1	0.24193	0.00970	24.93	<.0001

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.00087067	0.02096	0.04	0.9669
zlag1	1	0.78472	0.00981	79.96	<.0001
zlag2	1	-0.57316	0.01196	-47.91	<.0001
zlag3	1	0.38507	0.01196	32.20	<.0001
zlag4	1	-0.19413	0.00981	-19.78	<.0001

Use the output above to determine estimates of the first three partial autocorrelations: $\hat{\phi}_{11}$, $\hat{\phi}_{22}$, $\hat{\phi}_{33}$ (i.e. the first three values in the PACF). Write your answers (to **five** decimal places) in the blanks given below.

Problem 25. $\hat{\phi}_{11} =$ ____________

Problem 26. $\hat{\phi}_{22} =$ ____________

Problem 27. $\hat{\phi}_{33} =$ ____________

Problem 28. On the **next page** of this exam is a panel of graphs containing regression diagnostics. These were obtained from a regression of a response variable Y on three covariates X_1, X_2, X_3. Based on these graphs, which one of the following statements is true?

a) There is 1 case with an unusual response value.

b) There is 1 case which has a large influence on the estimated parameters and predicted values.

c) There is 1 case with unusual covariate values

d) There are 3 cases with unusual response values.

e) There are 3 cases which have a large influence on the estimated parameters and predicted values.

f) There are 3 cases with unusual covariate values.

Problem 29. On the **page after the next** is a single page of output (produced by the IDENTIFY statement in PROC ARIMA) for a series z_t. Using this output, select a reasonable ARMA model for this series from the list below. (Circle the correct response.)

a) ARMA(1,1)	**b**) ARMA(1,0)	**c**) ARMA(2,0)
d) ARMA(0,1)	**e**) ARMA(0,2)	**f**) random shocks

The **last three** pages of the exam give output (produced by the IDENTIFY statement in PROC ARIMA) for a series z_t and its first and second differences. Using this output, select a reasonable ARIMA(p, d, q) model for this series. Specify your answer in the next two questions.

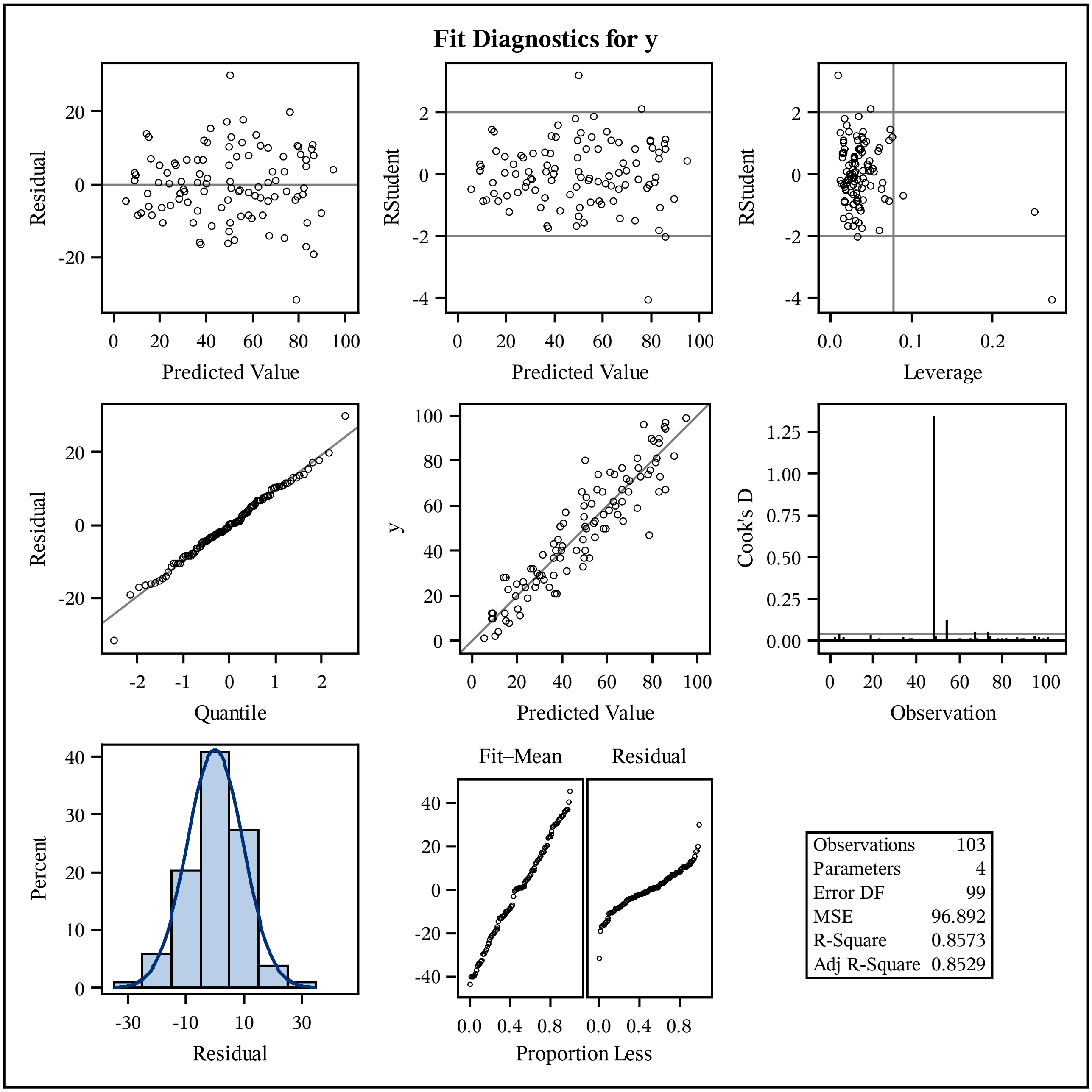
Problem 30. What value of d should be used?

a) $d = 0$	**b**) $d = 1$	**c**) $d = 2$	**d**) $d = 3$

Problem 31. What values of p and q should be used?

a) $p = 1, q = 1$	**b**) $p = 2, q = 2$	**c**) $p = 0, q = 2$	**d**) $p = 0, q = 3$
e) $p = 1, q = 0$	**f**) $p = 2, q = 0$	**g**) $p = 3, q = 0$	**h**) $p = 0, q = 1$

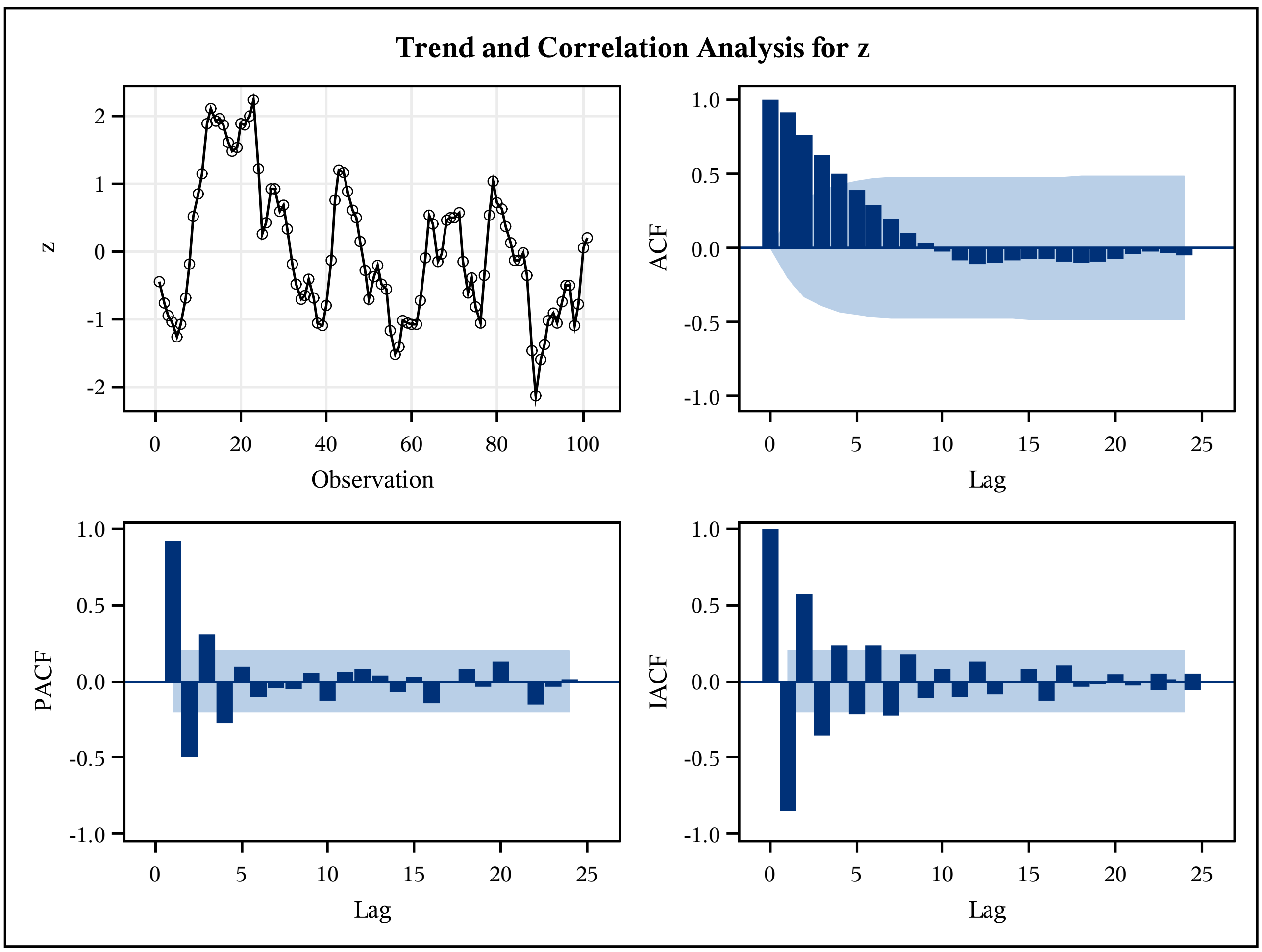
The REG Procedure
Model: MODEL1
Dependent Variable: y



The ARIMA Procedure

Name of Variable = z	
Mean of Working Series	-9.9E-6
Standard Deviation	0.995022
Number of Observations	101

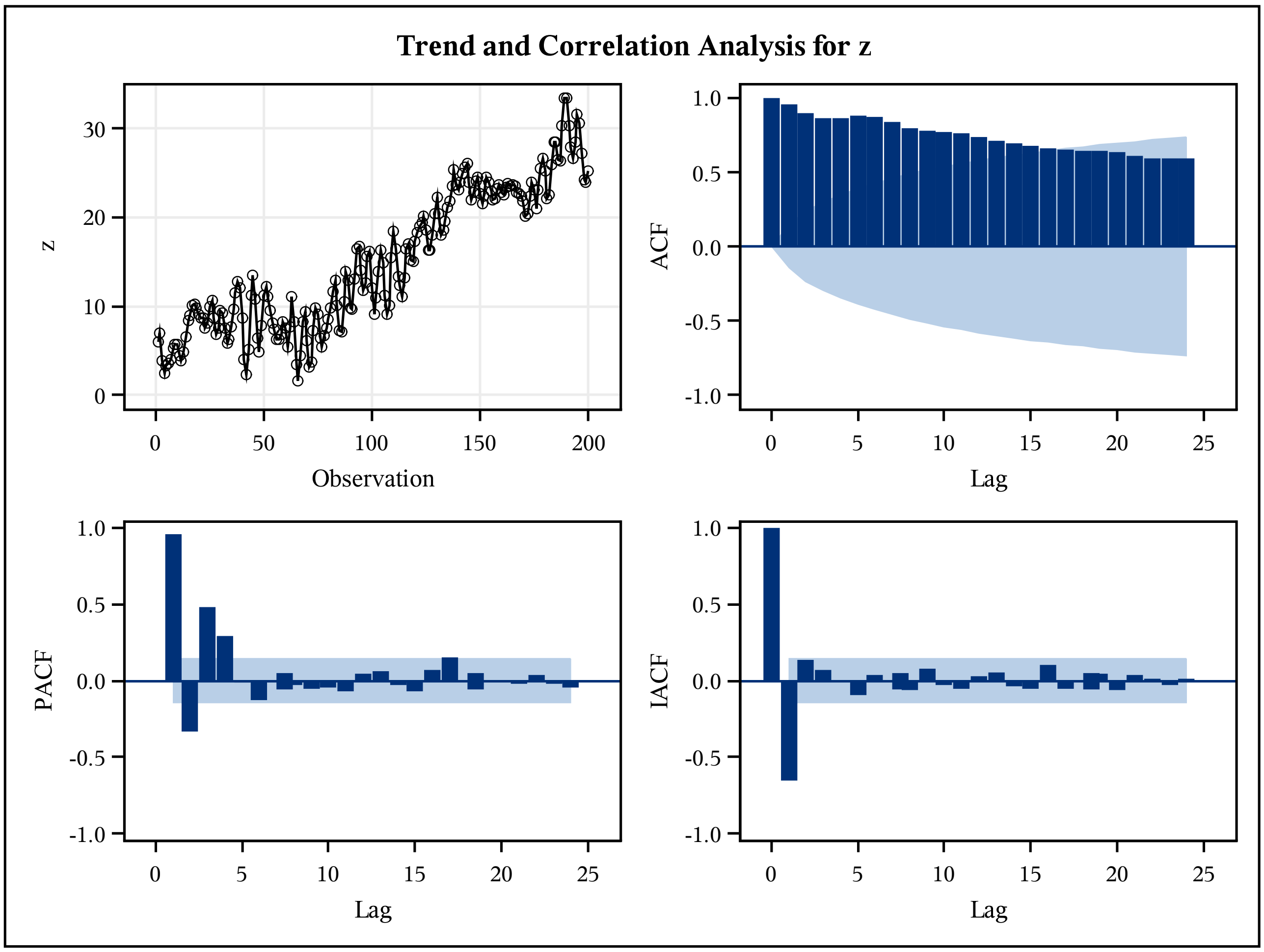
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	242.44	6	<.0001	0.917	0.760	0.624	0.504	0.391	0.289
12	250.07	12	<.0001	0.194	0.104	0.031	-0.028	-0.080	-0.106
18	256.08	18	<.0001	-0.100	-0.086	-0.078	-0.079	-0.094	-0.104
24	258.68	24	<.0001	-0.096	-0.072	-0.040	-0.021	-0.031	-0.054



The ARIMA Procedure

Name of Variable = z	
Mean of Working Series	15.01998
Standard Deviation	7.921258
Number of Observations	200

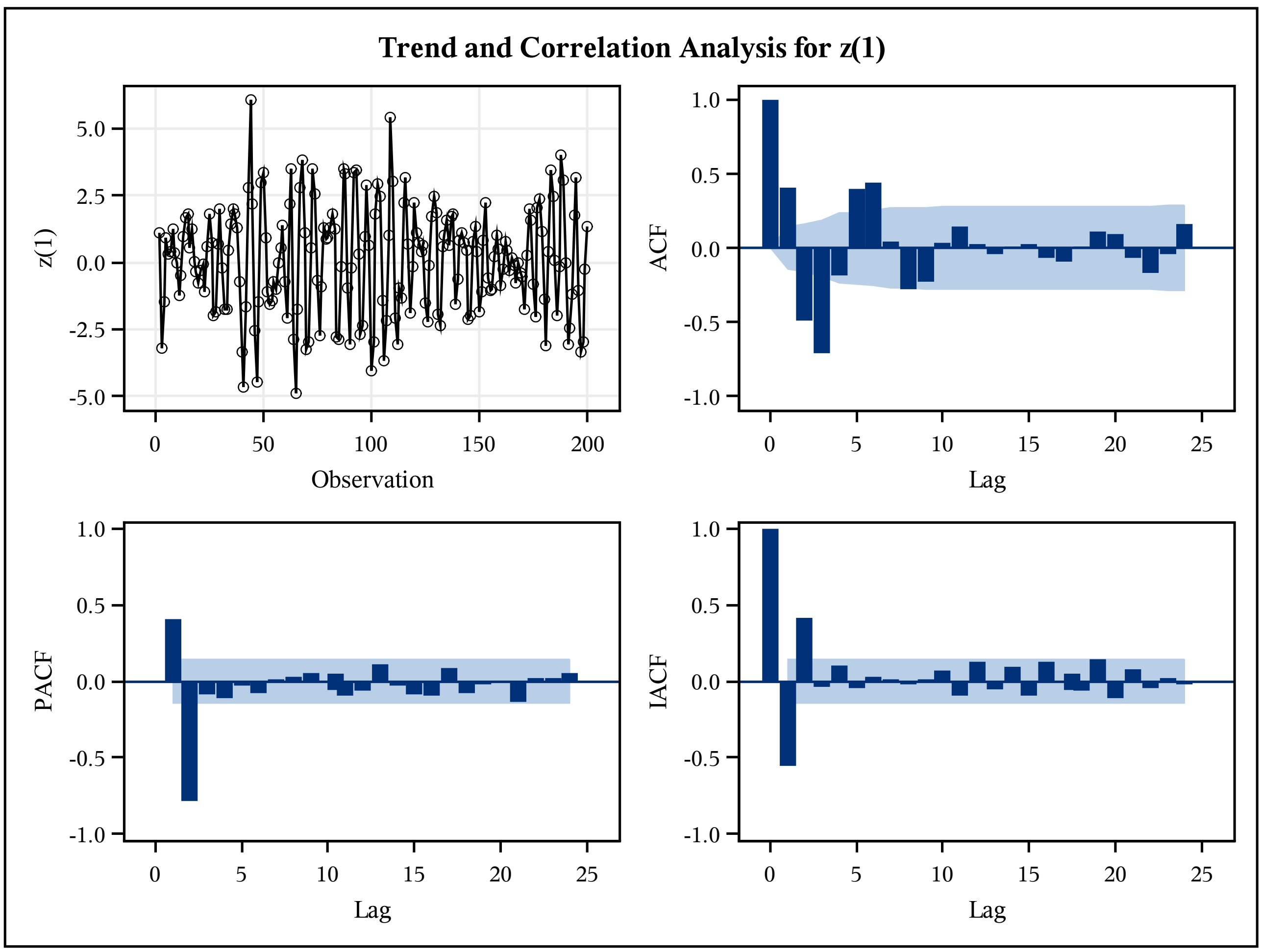
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	979.62	6	<.0001	0.961	0.897	0.862	0.868	0.883	0.873
12	1756.06	12	<.0001	0.837	0.800	0.780	0.773	0.759	0.735
18	2356.08	18	<.0001	0.713	0.698	0.682	0.664	0.651	0.646
24	2862.99	24	<.0001	0.642	0.632	0.613	0.597	0.591	0.590



The ARIMA Procedure

Name of Variable = z	
Period(s) of Differencing	1
Mean of Working Series	0.097254
Standard Deviation	1.997401
Number of Observations	199
Observation(s) eliminated by differencing	1

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	263.96	6	<.0001	0.410	-0.487	-0.706	-0.188	0.396	0.443
12	296.21	12	<.0001	0.045	-0.280	-0.224	0.038	0.143	0.029
18	299.79	18	<.0001	-0.039	0.013	0.026	-0.071	-0.095	0.006
24	318.37	24	<.0001	0.109	0.095	-0.069	-0.166	-0.045	0.164



The ARIMA Procedure

Name of Variable = z	
Period(s) of Differencing	1,1
Mean of Working Series	0.001253
Standard Deviation	2.172606
Number of Observations	198
Observation(s) eliminated by differencing	2

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	232.86	6	<.0001	0.267	-0.573	-0.629	-0.058	0.451	0.378
12	272.04	12	<.0001	-0.056	-0.322	-0.173	0.128	0.181	-0.041
18	279.49	18	<.0001	-0.097	0.037	0.090	-0.064	-0.107	0.003
24	302.93	24	<.0001	0.103	0.124	-0.059	-0.189	-0.075	0.182