TEST #1

STA 4853 **Name:**____________________________________

March 1, 2021

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 **35** multiple choice questions.
- Select the **single best** answer for each multiple choice question.
- For each question, type the **single lower case letter** of the correct response into the "Fill in the blank" box in Canvas.
- There is no penalty for guessing.
- The exam has **19** pages.
- Each question is worth equal credit.
- The exam is closed book and closed notes. NO books, notes, computers, or internet resources are allowed. No communication with other humans about the exam is allowed during the exam.
- You must attend the Zoom meeting and remain unmuted with your video on during the entire exam. You must remain visible on Zoom during the entire exam.
- Always be on the look-out for bad page breaks. Sometimes the output needed to answer a question goes onto another page.

Problem 1. If there is a regression model with time ordered data in which the regression errors have negative serial correlation, then ________.

a) the normal probability plot of the errors will have a negative slope

b) the time series plot of the errors will have a negative slope

c) negative errors tend to be followed by positive errors

d) negative errors tend to be followed by negative errors

e) the plot of the fitted values versus the residuals will have a negative slope

f) the plot of the residuals versus the covariates will have a negative slope

g) there is a negative correlation between the errors and the fitted values

h) there is a negative correlation between the errors and the covariates

Problem 2. Suppose $a_1, a_2, a_3, \ldots$ is a random shock sequence, and ψ_0, ψ_1, ψ_2 are constants. Then the expression

$$\psi_1^2 E a_2^2 + \psi_2^2 E a_1^2 + \psi_0 \psi_1 E a_3 a_1 + \psi_1 \psi_2 E a_2 a_1$$

is equal to ________.

a) $\psi_1^2 + \psi_2^2 + \psi_0\psi_1 + \psi_1\psi_2$

b) $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$

c) $\sigma_a^2 \psi_0 \psi_2$

d) $\dfrac{\sigma_a^2}{1 - \psi_1^2}$

e) $\dfrac{\sigma_a^2}{1 - \psi_0 - \psi_1 - \psi_2}$

f) $\dfrac{\sigma_a^2}{1 - \psi_0^2 - \psi_1^2 - \psi_2^2}$

g) $\sigma_a^2(\psi_1^2 + \psi_2^2)$

h) $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$

Problem 3. Suppose $\{z_t\}$ is a stationary ARMA(p, q) process. Then the expression

$$E(z_t + z_{t-1} + z_{t-2} + a_t + a_{t-1})$$

is equal to ________.

a) $\dfrac{C}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$

b) C

c) $\sigma_a^2 \sum_{i=0}^{q} \psi_i^2$

d) $1 + \phi_1 + \phi_1^2$

e) $3\mu_z$

f) $3\mu_z + 2\sigma_a^2$

g) $\dfrac{C}{1 - \phi_1}$

h) $\dfrac{\sigma_a^2}{1 - \phi_1^2}$

Problem 4. Suppose you have time series data $z_1, z_2, \ldots, z_n$ which is a realization of a stationary process, and you use this data to compute the sample autocorrelations $r_1, r_2, \ldots, r_k$ up to some lag k. Using the values $r_1, r_2, \ldots, r_k$, it is possible to compute an estimate of ________.

a) θ_k **b)** ϕ_k **c)** ϕ_{kk} **d)** μ_z **e)** σ_z^2 **f)** σ_a^2 **g)** C **h)** ρ_{2k}

Problem 5. Suppose $\{a_t\}$ is a random shock sequence, $|\phi_1| < 1$, and $C \neq 0$. We choose a starting value z_1, and use the rule $z_t = C + \phi_1 z_{t-1} + a_t$ to generate a sequence $z_1, z_2, z_3, \ldots$. After an initial short-term phase, the sequence $z_1, z_2, z_3, \ldots$ converges to ________

a) a straight line with positive slope
b) zero
c) a straight line with negative slope
d) a process with exponential growth
e) white noise
f) within a band of width two standard errors about zero
g) within a band of constant width about zero
h) a stationary process

Problem 6. Suppose $\{z_t\}$ is a stationary ARMA(p, q) process. Then the expression

$$E(C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t)$$

is equal to ________.

a) $\phi_1^2\sigma_z^2 + \phi_2^2\sigma_z^2 + \sigma_a^2$
b) $C + \phi_1^2\sigma_z^2 + \phi_2^2\sigma_z^2 + \sigma_a^2$
c) $C + \phi_1\sigma_z^2 + \phi_2\sigma_z^2 + \sigma_a^2$
d) $\phi_1\sigma_z^2 + \phi_2\sigma_z^2$
e) $C + \phi_1\mu_z + \phi_2\mu_z$
f) $\phi_1\mu_z + \phi_2\mu_z$
g) $\phi_1^2\mu_z + \phi_2^2\mu_z$
h) $\phi_1^2\mu_z + \phi_2^2\mu_z + \sigma_a^2$

Problem 7. Let ρ denote the population correlation between X and Y. If $\rho = 0$, then we know (roughly) that there is ________ between X and Y.

a) no linear relationship
b) a strong linear relationship
c) a strong quadratic relationship
d) a strong relationship of some kind
e) no quadratic relationship
f) no relationship of any kind

Problem 8. The figure 12.06 found in the second row of numbers in the table below is the value of a Chi-Square statistic for testing the null hypothesis ________.

a) $H_0 : \phi_{12,12} = 0$

b) $H_0 : \phi_{7,7} = \phi_{8,8} = \cdots = \phi_{12,12} = 0$

c) $H_0 : \mu_z = 0$

d) $H_0 : \sigma_z^2 = 0$

e) $H_0 : r_{12} = 0$

f) $H_0 : \rho_1 = \rho_2 = \cdots = \rho_{12} = 0$

g) $H_0 : \rho_{12} = 0$

h) $H_0 : \rho_7 = \rho_8 = \cdots = \rho_{12} = 0$

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.70	6	0.5829	0.143	0.079	0.001	-0.096	0.088	-0.035
12	12.06	12	0.4412	-0.200	-0.127	0.004	0.050	0.083	0.031
18	13.16	18	0.7819	-0.043	-0.064	0.008	0.023	-0.044	0.027
24	20.13	24	0.6892	-0.138	0.020	0.005	-0.003	0.015	-0.183

Problem 9. The output in the table above would lead us to suspect the data being analyzed is ________.

a) an MA(1) process

b) an AR(12) process

c) an MA(12) process

d) a random walk

e) an ARMA(12,12) process

f) white noise

g) non-stationary

h) an AR(1) process

Problem 10. Let $\{a_t\}$ be a random shock sequence. Consider the two random variables

$$5 + a_1 + 2a_2 + 3a_3 \qquad \text{and} \qquad 5 + 3a_4 + 2a_5 + a_6\,.$$

These random variables ________.

a) are independent

b) are uncorrelated, but not independent

c) have a chi-square distribution

d) have an expected value of zero

e) have a variance of zero

f) have a t distribution

g) have a positive correlation

h) have a negative correlation

Problem 11. Generating a realization from the process

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

requires _________ starting values for the series z_t.

a) 0 **b)** 1 **c)** 2 **d)** 3 **e)** 4 **f)** 5 **g)** 6

Problem 12. If $\{z_t\}$ is an AR(1) process, then z_{101} and a_{99} _________.

a) have skewed distributions
b) have positive means
c) decay exponentially
d) have a cutoff to zero
e) are correlated
f) are independent

Problem 13. Suppose you observe X and Y (height and weight) for a random sample of n individuals: $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. For X and Y, the **sample covariance** is _________.

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

d) $E(Y-\mu_y)^2$ **e)** $\frac{\text{Cov}(X,Y)}{\sigma_x\sigma_y}$ **f)** $\frac{c(X,Y)}{s_x s_y}$

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

Problem 14. Suppose you are interested in the height X of individuals in some large population. The population average of the heights can be written as _________.

a) $\frac{1}{n}\sigma_X^2$ **b)** $\frac{1}{n}\overline{X}$ **c)** $\frac{1}{n}s_X^2$ **d)** σ_X^2 **e)** σ_X **f)** $\frac{1}{n}EX$ **g)** EX **h)** $\overline{X}$

Problem 15. For an AR(1) process with $\phi_1 = -0.5$, the lag 3 autocorrelation ρ_3 is equal to _________.

a) 0.167 **b)** −0.167 **c)** 0.125 **d)** −0.125 **e)** 0.25 **f)** −0.25 **g)** 0.5 **h)** −0.5

Problem 16. A stationary AR(1) process

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

has $\text{Var}(z_t) =$ __________.

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

Problem 17. 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) $\left(1 + \frac{1}{2}\sum_{j=1}^{k-1} r_j\right)^{-1/2} n^{+1/2}$ **b)** $\left(1 - \frac{1}{2}\sum_{j=1}^{k-1} r_j\right)^{-1/2} n^{+1/2}$

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

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

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

Problem 18. After fitting a regression model, if the magnitude of the t-value t_i for the estimated regression coefficient $\hat{\beta}_i$ is **small**, then we __________ the null hypothesis $H_0 : \beta_i = 0$ and conclude that the variable X_i is __________ in our model. Circle the response below with the choices (separated by a semi-colon) which correctly fill in the two blanks.

a) reject ; needed

b) reject ; **not** needed

c) do **not** reject ; **not** needed

d) do **not** reject ; needed

Problem 19. As you add more covariates to a regression model, the value of R-squared always __________

a) becomes more accurate

b) becomes less accurate

c) becomes statistically significant

d) becomes statistically insignificant

e) increases

f) decreases

g) eventually decays to zero

Problem 20. Which of the following is a correct expression for the **population correlation** between X and Y?

a) $E(X - \overline{X})(Y - \overline{Y})$

b) $\dfrac{E(X - \mu_x)(Y - \mu_y)}{\sigma_x \sigma_y}$

c) $\dfrac{c(X,Y)}{s_x s_y}$

d) $\dfrac{\text{Cov}(X,Y)^2}{\sigma_x^2 \sigma_y^2}$

e) $\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$

f) $\frac{1}{N} \sum_{i=1}^{N} (X_i - \mu_x)^2 (Y_i - \mu_y)^2$

Problem 21. An ARMA process is constructed from a sequence of random shocks a_t which are __________ random variables.

a) independent $N(0, \sigma_a^2)$ **b)** serially correlated **c)** autocorrelated **d)** skewed
e) positive **f)** negatively correlated **g)** non-stationary **h)** increasing

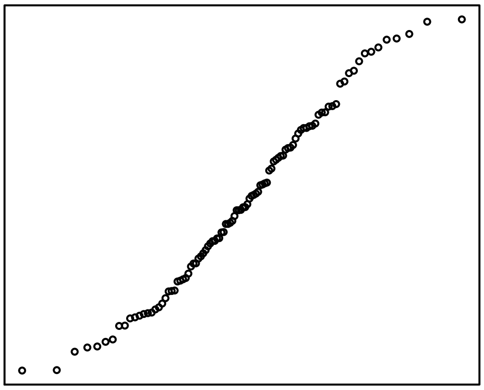
---

Suppose $\{Y_t\}$ and $\{X_t\}$ are time series, and you fit a regression model $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$.

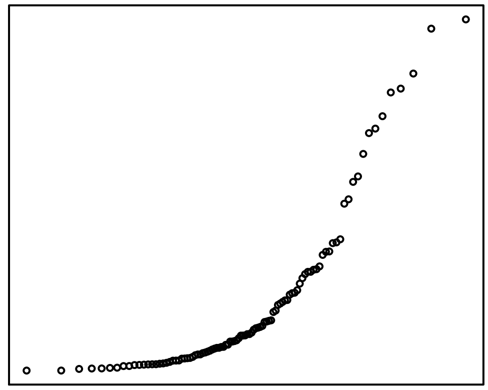
Problem 22. If the regression residuals exhibit strong serial correlation, then the time series plot of the residuals (the residuals plotted in time order) might resemble one of the following plots. Which one?

a) A **b)** B **c)** C **d)** D

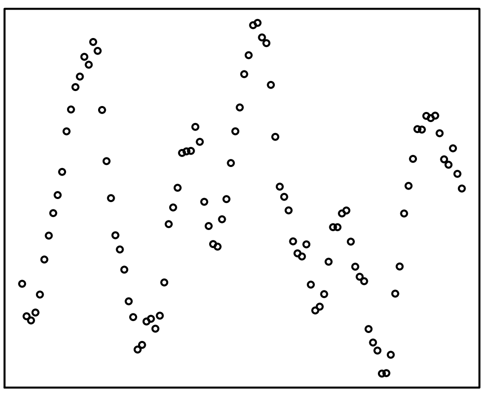
A



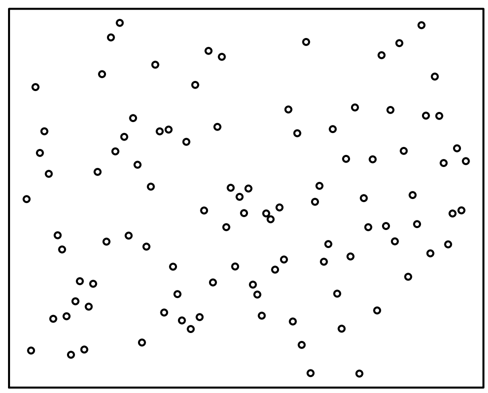
B



C



D



Problem 23. Continuing in the same situation as the previous question: 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) estimate **b)** p-value **c)** t-value **d)** standard error

Problem 24. Which one of the following is a true relationship between the autocorrelations (ρ_k) and autocovariances (γ_k)?

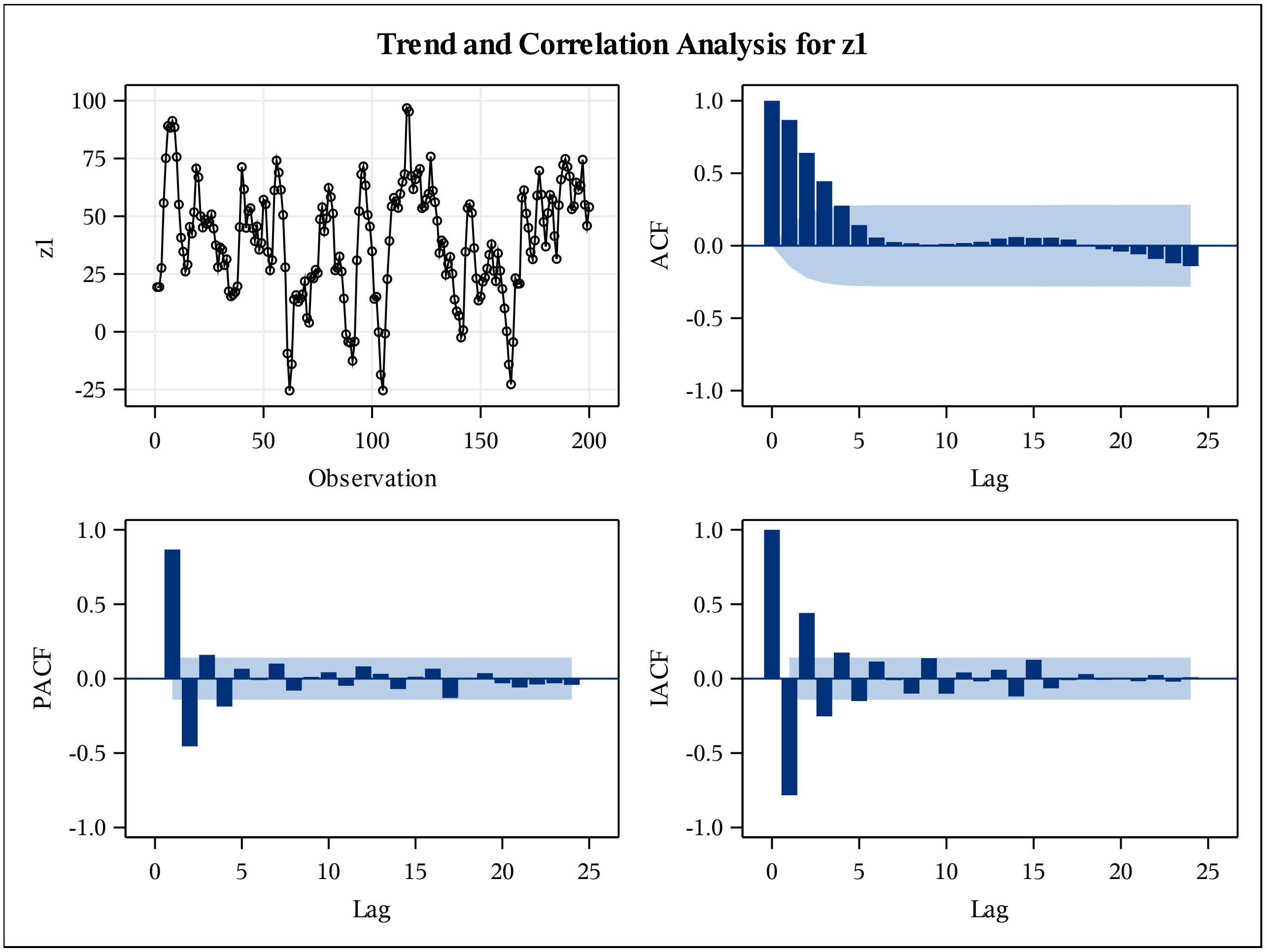
a) $\gamma_k = \rho_1^k$ **b)** $\rho_k = \gamma_1 \rho_{k-1}$ **c)** $\gamma_k = \rho_1 \gamma_{k-1}$

d) $\gamma_k = \dfrac{\rho_k}{\rho_0}$ **e)** $\rho_k = \dfrac{\gamma_k}{\gamma_0}$ **f)** $\rho_k = \gamma_1^k$

Problem 25. Running SAS PROC ARIMA on a time series (named z1) produced the output given below. What is a reasonable model for this time series?

a) random shocks **b**) MA(1) **c**) MA(2)

d) AR(1) **e**) AR(2) **f**) ARMA(1,1)



The next four questions use the SAS output described below which is given on the following four pages.

A simulated data set has 103 observations and four variables, a response variable Y and three covariates X1, X2, X3. The output on the following pages contains: (1) a matrix of scatterplots for the covariates X1, X2, X3; (2) some printed output from SAS PROC REG obtained by regressing the response variable Y on all three covariates X1, X2, X3; (3) the usual panel of plots produced by SAS PROC REG displaying some case diagnostics and other items; and (4) a printout of the first 24 observations giving the values of Y, X1, X2, X3, the predicted (fitted) values, residuals, Cook's D, Leverage, and RStudent.

Three unusual observations (i.e., rows) have been planted in this data set.

Problem 26. Two of the unusual observations are easily visible in the pairwise scatterplots of X1, X2, X3, and these observations have been circled in the plots. What are these observations? (Select the correct pair of observation numbers.)

a) 7, 15 **b**) 4, 20 **c**) 8, 17 **d**) 1, 20 **e**) 8, 16
f) 7, 13 **g**) 4, 16 **h**) 3, 8 **i**) 13, 15 **j**) 1, 17

Problem 27. One of the unusual points has a much greater effect on the regression model (such as on the the estimated parameters and predicted values) than the other two. Which observation is this?

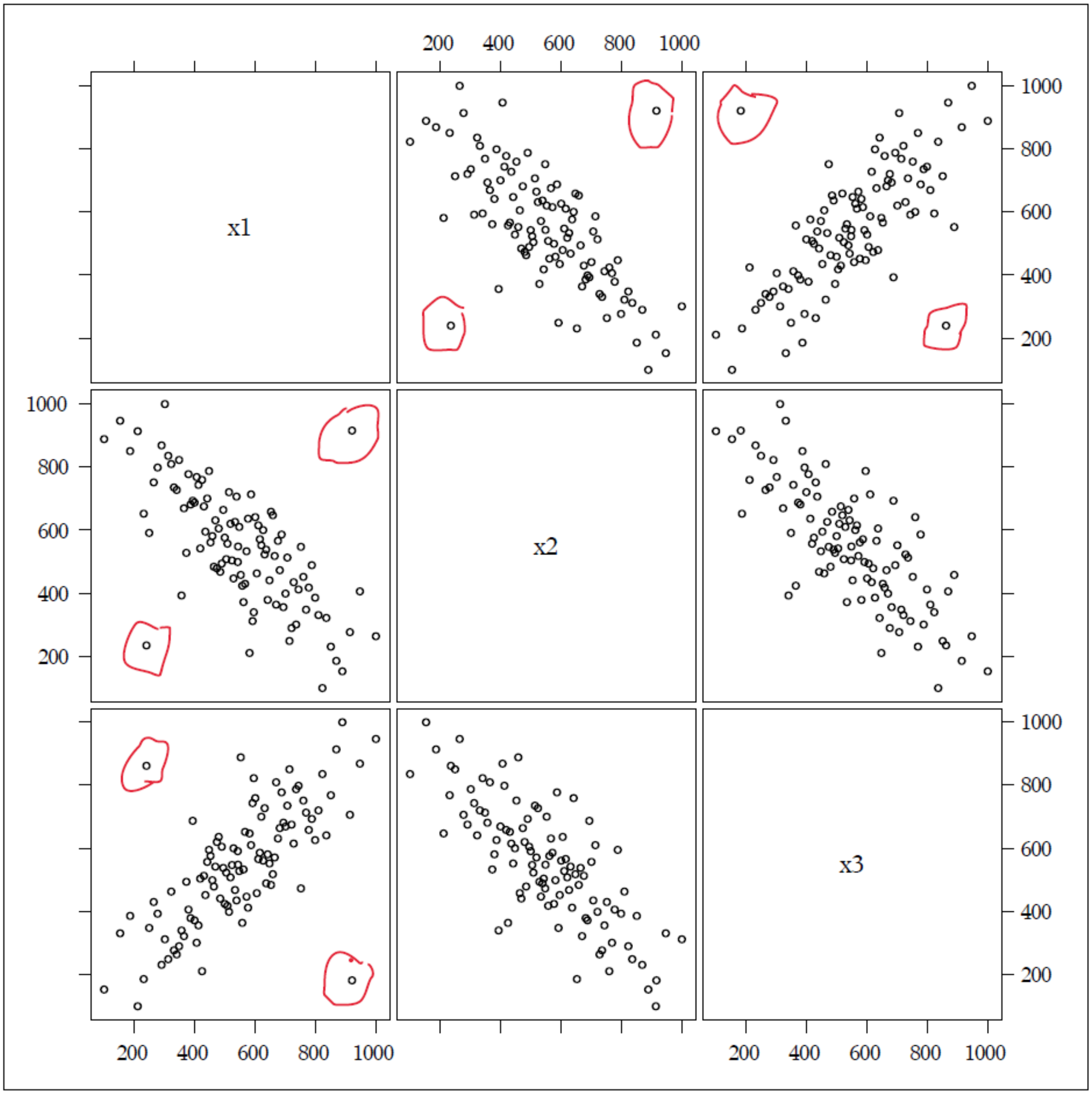
a) 1 **b**) 3 **c**) 4 **d**) 7 **e**) 8
f) 13 **g**) 15 **h**) 16 **i**) 17 **j**) 20

Problem 28. Two of the three points have unusual response values. Which observations are they?

a) 1, 20 **b**) 13, 15 **c**) 7, 16 **d**) 7, 13 **e**) 15, 16
f) 1, 7 **g**) 16, 20 **h**) 3, 8 **i**) 8 , 17 **j**) 3, 17

Problem 29. Various numbers copied from the regression output are given in the choices below. Which of these is an estimate of the variance of the regression errors ε_i?

a) 3029912 **b**) 525828 **c**) 3555740 **d**) 1009971 **e**) 5311.39538
f) 548.10680 **g**) 13.29655 **h**) 0.8521 **i**) 0.8476 **j**) 69.50810



x1
x2
x3
200
400
600
800
1000

The REG Procedure
Model: MODEL1
Dependent Variable: y

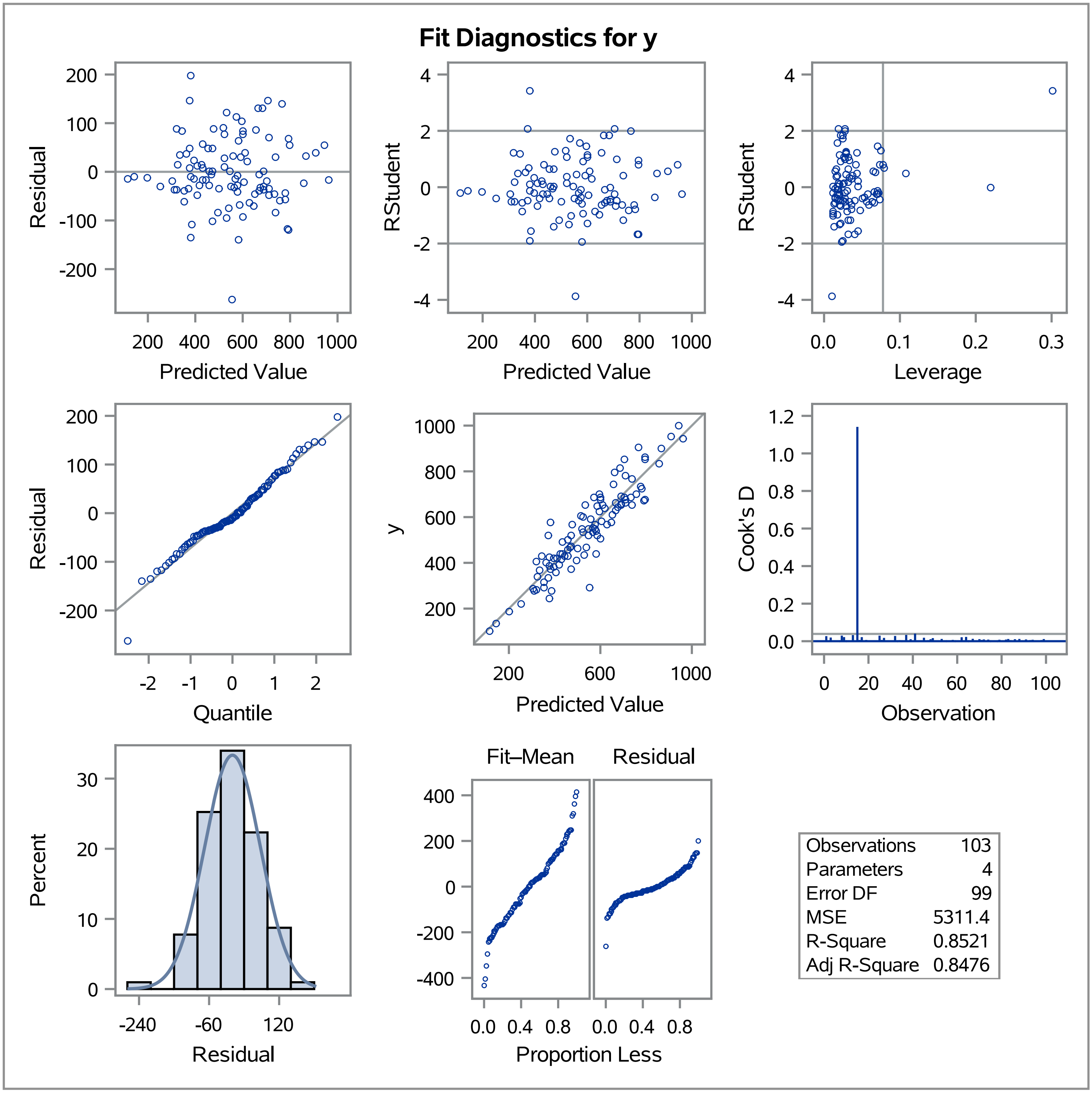
Number of Observations Read	103
Number of Observations Used	103

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	3029912	1009971	190.15	<.0001
Error	99	525828	5311.39538		
Corrected Total	102	3555740			

Root MSE	72.87932	**R-Square**	0.8521
Dependent Mean	548.10680	**Adj R-Sq**	0.8476
Coeff Var	13.29655		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	1461.70194	69.50810	21.03	<.0001
x1	1	-0.67714	0.05581	-12.13	<.0001
x2	1	-0.38145	0.06384	-5.98	<.0001
x3	1	-0.60364	0.06508	-9.27	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: y



Obs	y	x1	x2	x3	predict	resid	cookd	leverage	rstudent
1	906	364	669	322	765.665	140.335	0.02690	0.02745	1.98121
2	336	713	249	850	370.831	-34.831	0.00251	0.04042	-0.48601
3	851	399	687	372	704.917	146.083	0.01952	0.01872	2.05620
4	222	868	186	913	251.877	-29.877	0.00239	0.05113	-0.41910
5	639	561	372	533	618.191	20.809	0.00079	0.03613	0.28948
6	420	620	552	700	408.772	11.228	0.00016	0.02590	0.15532
7	690	240	235	861	689.818	0.182	0.00000	0.21916	0.00281
8	520	735	301	787	374.128	145.872	0.02967	0.02799	2.06329
9	672	331	735	277	789.999	-117.999	0.02185	0.03129	-1.65954
10	386	768	348	713	378.524	7.476	0.00007	0.02439	0.10334
11	275	913	277	706	311.647	-36.647	0.00395	0.05568	-0.51553
12	439	586	713	610	424.709	14.291	0.00047	0.04420	0.19959
13	291	543	548	548	554.190	-263.190	0.03240	0.00975	-3.87780
14	608	508	557	418	652.930	-44.930	0.00206	0.02079	-0.62107
15	578	920	915	182	379.849	198.151	1.14132	0.30140	3.42467
16	898	356	393	340	865.496	32.504	0.00674	0.10786	0.47033
17	441	576	636	412	580.372	-139.372	0.02178	0.02275	-1.96215
18	687	277	798	393	732.511	-45.511	0.00376	0.03584	-0.63405
19	489	595	340	822	432.924	56.076	0.00682	0.04228	0.78471
20	133	946	406	868	142.306	-9.306	0.00034	0.07111	-0.13182
21	565	458	581	499	628.738	-63.738	0.00246	0.01255	-0.87909
22	428	693	356	681	445.574	-17.574	0.00031	0.02043	-0.24248
23	420	798	386	626	396.231	23.769	0.00095	0.03336	0.33023
24	678	386	681	379	711.783	-33.783	0.00105	0.01884	-0.46613

The next page of the exam contains four time series plots. One of these is a realization of a stationary process, and the other three are realizations of non-stationary processes. For each series, select the correct description. Each description is used exactly once.

Problem 30. Describe series #1.

a) Stationary
b) Does **not** have a constant mean
c) Does **not** have a constant variance
d) Does **not** have a constant ACF.

Problem 31. Describe series #2.

a) Stationary
b) Does **not** have a constant mean
c) Does **not** have a constant variance
d) Does **not** have a constant ACF.

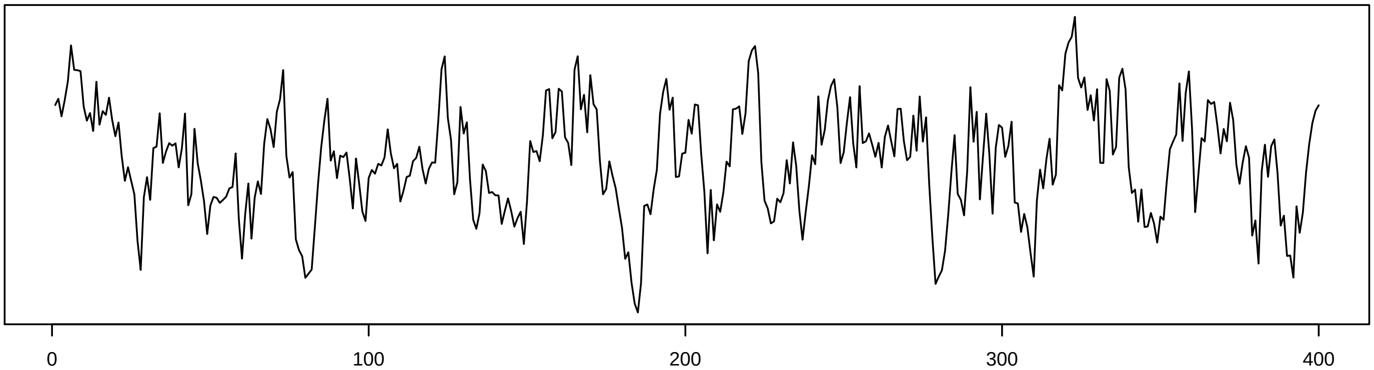
Problem 32. Describe series #3.

a) Stationary
b) Does **not** have a constant mean
c) Does **not** have a constant variance
d) Does **not** have a constant ACF.

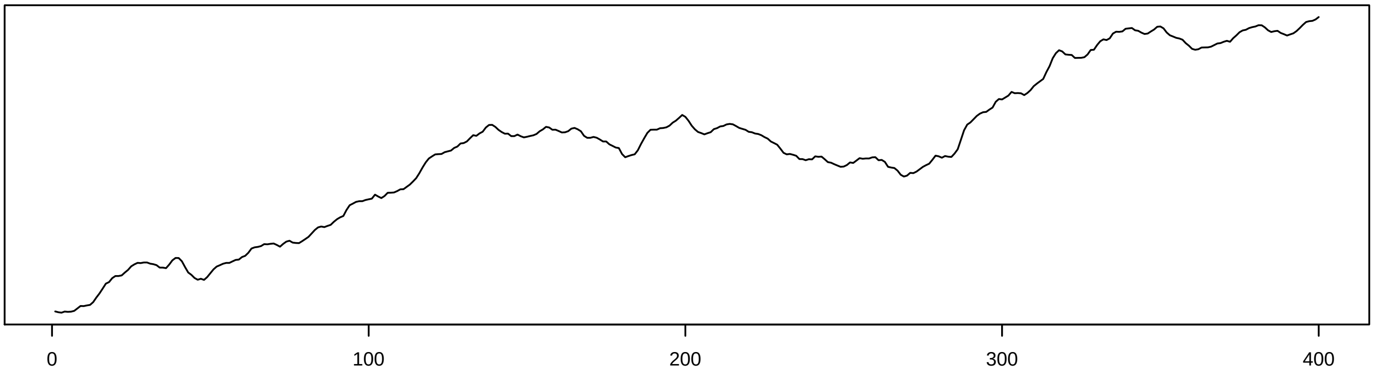
Problem 33. Describe series #4.

a) Stationary
b) Does **not** have a constant mean
c) Does **not** have a constant variance
d) Does **not** have a constant ACF.

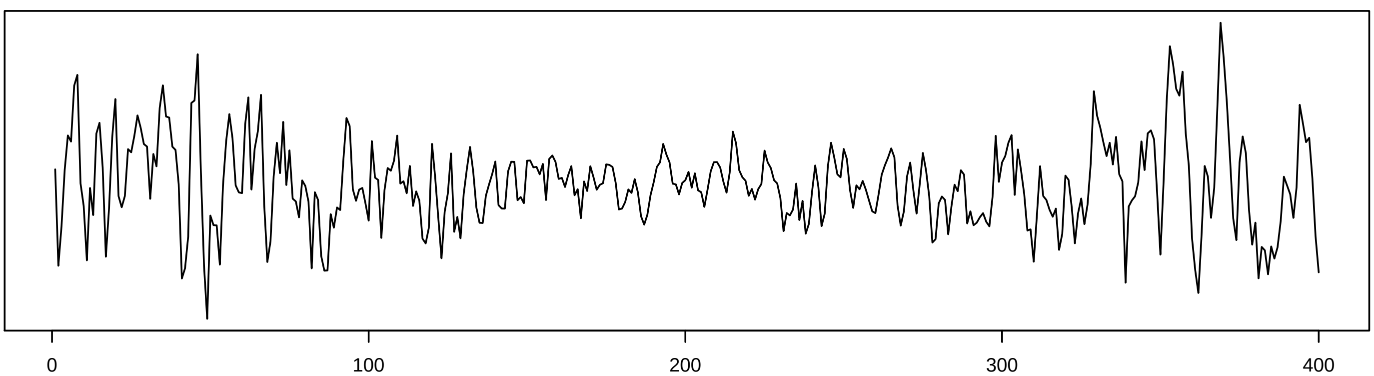
#1


0
100
200
300
400

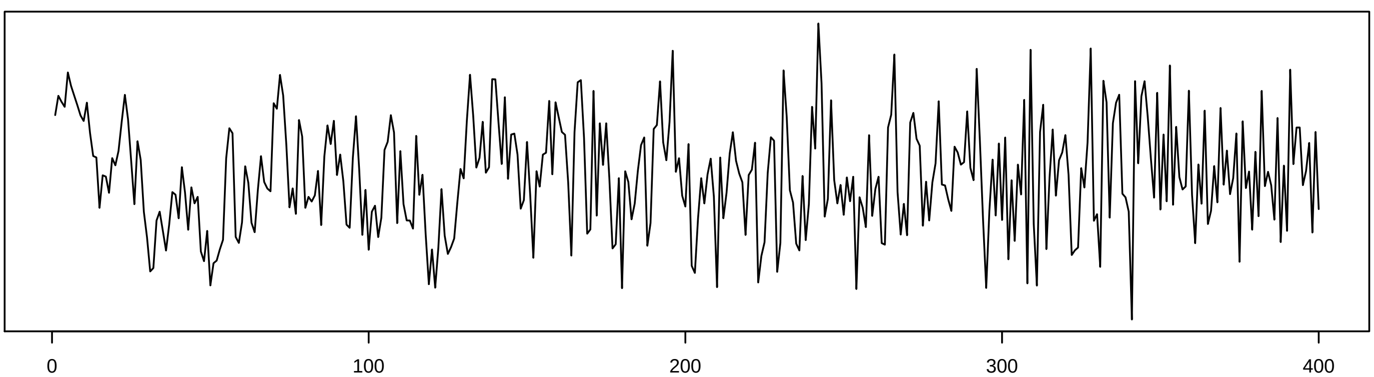
#2


0
100
200
300
400

#3


0
100
200
300
400

#4


0
100
200
300
400

A long time series `z` of length 10,000 was input to SAS, and used to create the lagged series `zlag1`, `zlag2`, `zlag3`, and `zlag4`. The first 10 observations of the series `z`, `zlag1`, `zlag2`, `zlag3`, and `zlag4` are shown in the table at the top of the following page. Using PROC REG, the following regressions were run:

- Regress `z` on `zlag1`
- Regress `z` on `zlag1`, `zlag2`
- Regress `z` on `zlag1`, `zlag2`, `zlag3`
- Regress `z` on `zlag1`, `zlag2`, `zlag3`, `zlag4`

The tables of parameter estimates from these four regressions are given on the following two pages.

Using this output one can approximate the sample PACF of the time series `z`.

Problem 34. The value of $\hat{\phi}_{22}$ is approximately equal to ________.

a) 0.91907	**b**) 1.54759	**c**) -0.68385	**d**) 1.60412	**e**) -0.81169
f) 0.08255	**g**) 1.62103	**h**) -0.97911	**i**) 0.41382	**j**) -0.20661

Problem 35. The value of $\hat{\phi}_{44}$ is approximately equal to ________.

a) 0.91907	**b**) 1.54759	**c**) -0.68385	**d**) 1.60412	**e**) -0.81169
f) 0.08255	**g**) 1.62103	**h**) -0.97911	**i**) 0.41382	**j**) -0.20661

Obs	z	zlag1	zlag2	zlag3	zlag4
1	-1.4867	.	.	.	.
2	-0.4903	-1.4867	.	.	.
3	-1.3835	-0.4903	-1.4867	.	.
4	-2.9196	-1.3835	-0.4903	-1.4867	.
5	-2.9987	-2.9196	-1.3835	-0.4903	-1.4867
6	-2.0683	-2.9987	-2.9196	-1.3835	-0.4903
7	-2.5333	-2.0683	-2.9987	-2.9196	-1.3835
8	-1.5472	-2.5333	-2.0683	-2.9987	-2.9196
9	0.3548	-1.5472	-2.5333	-2.0683	-2.9987
10	2.5219	0.3548	-1.5472	-2.5333	-2.0683

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.00051683	0.01401	-0.04	0.9706
zlag1	1	0.91907	0.00394	233.18	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.00121	0.01022	-0.12	0.9061
zlag1	1	1.54759	0.00730	212.06	<.0001
zlag2	1	-0.68385	0.00730	-93.71	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.00093326	0.01019	-0.09	0.9270
zlag1	1	1.60412	0.00997	160.93	<.0001
zlag2	1	-0.81169	0.01705	-47.59	<.0001
zlag3	1	0.08255	0.00997	8.28	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.00107	0.00997	-0.11	0.9149
zlag1	1	1.62103	0.00979	165.61	<.0001
zlag2	1	-0.97911	0.01848	-52.97	<.0001
zlag3	1	0.41382	0.01848	22.39	<.0001
zlag4	1	-0.20661	0.00979	-21.11	<.0001