TEST #1

STA 4853 **Name:**____________________________________

March 7, 2018

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.
- 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 **10** pages.
- Each question is worth equal credit.

Problem 1. If you have time series data $z_1, z_2, \ldots, z_n$ and you use SAS PROC ARIMA to estimate the parameters in the ARMA model

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

SAS will also compute residuals. These residuals are estimates of ________.

a)$\star$ a_t **b)** $\phi_1 - \hat{\phi}_1$ **c)** $\theta_1 - \hat{\theta}_1$ **d)** $z_t - C - \phi_1 z_{t-1}$ **e)** z_t
f) the transformed series **g)** the differences $z_t - z_{t-1}$ **h)** the response variable

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

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

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

Problem 3. A stationary ARMA(p,q) process with mean zero can be expressed as ________.

a) $\dfrac{\theta(B)}{\phi(B)}z_t$ **b)** $\dfrac{\phi(B)}{\theta(B)}z_t$ **c)**$\star$ $\dfrac{\theta(B)}{\phi(B)}a_t$ **d)** $\dfrac{\phi(B)}{\theta(B)}a_t$

e) $C + \dfrac{\phi(B)}{\theta(B)}a_t$ **f)** $C + \dfrac{\phi(B)}{\theta(B)}\tilde{z}_t$ **g)** $\dfrac{\theta(B)}{\phi(B)}\tilde{z}_t$ **h)** $\dfrac{\phi(B)}{\theta(B)}\tilde{z}_t$

Problem 4. 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) σ_X^2 **b)** σ_X **c)** $\dfrac{1}{n}EX$ **d)** $\dfrac{1}{n}\sigma_X^2$ **e)** $\dfrac{1}{n}\overline{X}$ **f)** $\dfrac{1}{n}s_X^2$ **g)**$\star$ EX **h)** $\overline{X}$

Problem 5. If a time series exhibits a non-stationary ________, then differencing the series will sometimes lead to a stationary series. Which of the following words can correctly fill this blank?

1. ACF 2. mean 3. variance

Select the correct choice from the list below. The possibilities range from none of the words to all of the words. (The words are referred to by their numbers.)

a) none **b)** 1 **c)**$\star$ 2 **d)** 3 **e)** 1, 2 **f)** 1, 3 **g)** 2, 3 **h)** 1, 2, 3

Problem 6. If you take a realization from an already stationary ARMA process and difference it, the sample **IACF** of the resulting differences will usually _________

- **a)**$\star$ decay to zero very slowly
- **b)** decay to zero very rapidly
- **c)** decay to zero exponentially
- **d)** decay to zero sinusoidally
- **e)** have alternating exponential decay
- **f)** have an approximate cutoff after lag 0
- **g)** have an approximate cutoff after lag 1
- **h)** have an approximate cutoff after lag 2

Problem 7. 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) σ_a^2 **b)** σ_z^2 **c)** C **d)** ϕ_1 **e)** ϕ_2 **f)** 1 **g)** 0.5 **h)**$\star$ 0

Problem 8. For an AR(1) process with $\phi_1 = -0.5$, the lag 3 autocorrelation ρ_3 is equal to _________.

a) 0.125 **b)**$\star$ -0.125 **c)** 0.25 **d)** -0.25 **e)** 0.5 **f)** -0.5 **g)** 0.167 **h)** -0.167

Problem 9. A stationary AR(1) process

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

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

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

Problem 10. To simulate a realization from an AR(3) process, we need k starting values $z_1, z_2, \ldots, z_k$. What is the value of k?

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

Problem 11. 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)** $\left(1 - 2\sum_{j=1}^{k-1} r_j^2\right)^{-1/2} n^{+1/2}$

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

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

Problem 12. Suppose $\{z_t\}$ is a stationary ARMA process generated using the sequence of random shocks $\{a_t\}$. The expression

$$E\left[\,(5\tilde{z}_{t-1} + a_t)\,\tilde{z}_{t-3}\,\right]$$

can be shown to be equal to _________.

a)$\star$ $5\gamma_2$ **b)** $5\gamma_2 + \sigma_a^2$ **c)** $5\gamma_2 + \sigma_z^2$ **d)** $25\gamma_2$ **e)** $25\gamma_2^2$
f) $5\rho_2$ **g)** $5\rho_2 + \sigma_a^2$ **h)** $5\rho_2 + \sigma_z^2$ **i)** $25\rho_2$ **j)** $25\rho_2^2$

MINIC is a _________ a time series. It finds the values of p and q of the ARMA(p, q) model which minimizes an estimate of the _________.

Problem 13. Which of the following phrases correctly fills in the **first** blank above?

a) test of the normality of the residuals obtained by fitting an ARMA(p, q) model to
b) test of the significance of an ARMA(p, q) model for
c) test of the residual ACF when fitting an ARMA(p, q) model to
d) test of the stationarity of an ARMA(p, q) model for
e) method for estimating the parameters of an ARMA(p, q) model for
f)$\star$ method to help identify the orders p and q of an appropriate ARMA model for

Problem 14. Which of the following choices correctly fills in the **second** blank above?

a) PACF **b)** Constant Estimate **c)** Chi-Square **d)** ACF
e) Likelihood **f)** Variance Estimate **g)**$\star$ SBC **h)** Standard Error Estimate

Problem 15. Suppose we know the values of a time series $\{z_t\}$ at times $1, 2, \ldots n$, that is, we know $z_1, z_2, z_3, \ldots, z_n$. The series $\{z_{t-5}\}$ (z_t lagged by 5) will have _________

a)$\star$ 5 missing values at the beginning of the series
b) 4 missing values at the beginning of the series
c) 5 missing values at the end of the series
d) 4 missing values at the end of the series
e) no missing values

Problem 16. For a stationary process $\{z_t\}$, the quantity $\text{Cov}(z_t, z_{t-k})$ is represented by the symbol _________

a) ψ_k **b)** ϕ_{kk} **c)** θ_k **d)** ϕ_k **e)**$\star$ γ_k **f)** ρ_k

Problem 17. If the theoretical **P**ACF of a stationary ARMA process has a cutoff (to zero) after lag 3, then the theoretical **I**ACF (Inverse Autocorrelation Function) will _________.

a) undergo sinusoidal decay
b) undergo alternating exponential decay
c) decay exponentially
d) decay to zero very slowly
e) decay to zero very rapidly
f)$\star$ have a cutoff after lag 3
g) decay exponentially starting at lag 3
h) undergo sinusoidal decay after lag 3

Problem 18. A realization from a process with a **non**-stationary mean will usually _________.

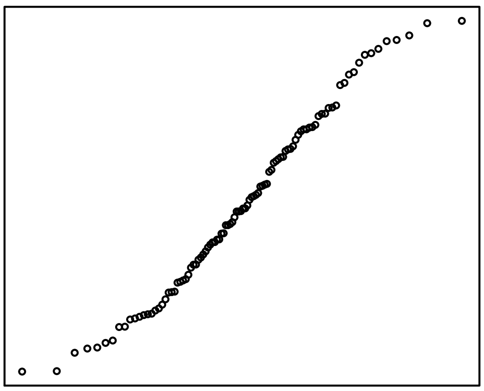
a) require a log transformation
b) require a square root transformation
c) require a square transformation
d) require a reciprocal transformation
e) require a large AR order p in its model
f) require a large MA order q in its model
g) require large values of both p and q in its model
h) have a sample PACF which decays very slowly to zero
i)$\star$ have a sample ACF which decays very slowly to zero
j) have a sample IACF which decays very slowly to zero

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

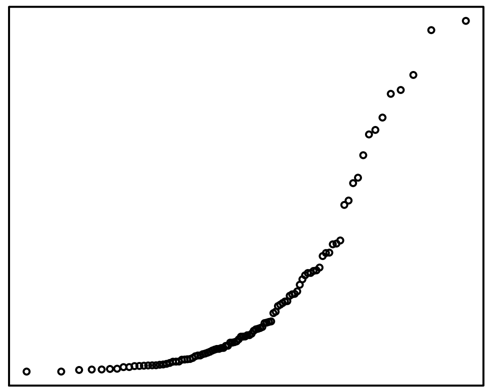
Problem 19. If the regression residuals exhibit strong serial correlation, then the time series plot of the residuals (the residuals plotted in time order) might resemble one of the following plots. Which one?

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

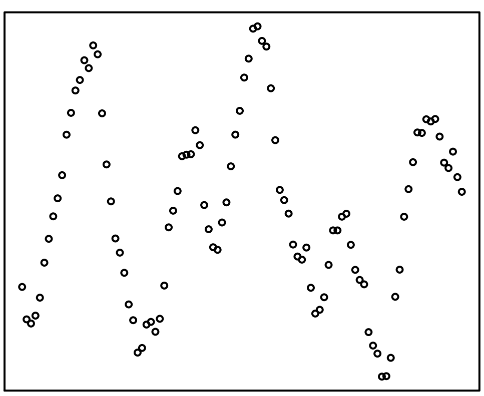
A



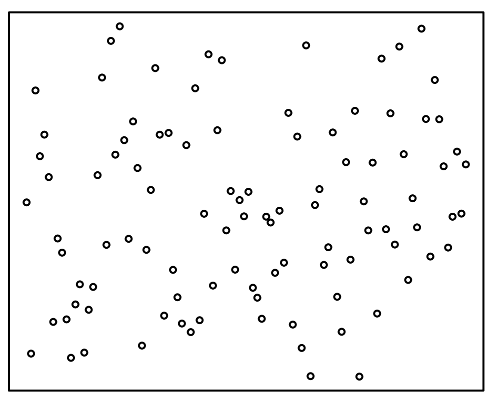
B



C



D



Problem 20. 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)$\star$ estimate **b**) standard error **c**) t-value **d**) p-value

Problem 21. 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)$\star$ $\rho_k = \dfrac{\gamma_k}{\gamma_0}$ **e)** $\rho_k = \gamma_1^k$ **f)** $\gamma_k = \dfrac{\rho_k}{\rho_0}$

Problem 22. Transforming a series z_t (that is, modeling $y_t = f(z_t)$ instead of z_t for some appropriately chosen function f such as log or square root) is useful when _________

a) the level of the series changes systematically with time.
b) the level of the series z_t changes periodically with time.
c) the level of the series z_t changes repeatedly over time.
d) the variability of the series z_t changes systematically with time.
e)$\star$ the variability of the series z_t changes systematically with the level of the series.
f) the variability of the series z_t changes periodically with time.

Problem 23. 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) $z_t = C + \phi_2 z_{t-2} + a_t - \theta_4 a_{t-4}$
b) $z_t = C + \phi_4 z_{t-4} + a_t - \theta_2 a_{t-2}$
c)$\star$ $z_t = C + \phi_2 z_{t-2} + \phi_4 z_{t-4} + a_t$
d) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4}$
e) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \phi_4 z_{t-4} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
f) $z_t = C + a_t - \theta_2 a_{t-2} - \theta_4 a_{t-4}$

Problem 24. Let L be the Likelihood value, k be the number of estimated parameters, and n be the number of residuals. Which of the following is the correct formula for SBC (Schwarz's Bayesian Criterion) also known as BIC (Bayesian Information Criterion)?

a) $-2\ln(L) - k\ln(n)$ **b)** $2\ln(L) - k\ln(n)$ **c)** $-\ln(L) - k\ln(n)$

d) $-\ln(L) - 2k$ **e)**$\star$ $-2\ln(L) + k\ln(n)$ **f)** $\ln(L) - 2k$

g) $\ln(L) - k\ln(n)$ **h)** $-2\ln(L) - 2k$ **i)** $2\ln(L) - 2k$

Problem 25. When you fit a regression model with sample size n and p covariates, the value of R^2 (R-squared) tends to ________ the performance of the regression, especially when ________. (Choose the pair which best completes this sentence.)

a)⋆ overstate / p is large and n is small
b) overstate / p is small and n is large
c) overstate / parameter estimates and standard errors are large
d) overstate / parameter estimates and standard errors are small
e) understate / p is large and n is small
f) understate / p is small and n is large
g) understate / parameter estimates and standard errors are large
h) understate / parameter estimates and standard errors are small

Problem 26. 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) $z_{t+j} z_{t+k}$
d) $B^j z_t B^k z_t$
e)⋆ $B^{j+k} z_t$
f) z_{jk}
g) z_{t-jk}

Problem 27. 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) random walk b) ARMA(p, q) c) ARIMA(p, d, q) d)⋆ MA(q) 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 28. When is $E(a_s a_t) = 0$?

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

Problem 29. When is $E(z_s a_t) = 0$?

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

Problem 30. 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**) Durbin-Watson **c**) Ljung-Box Q **d**) t-value **e**) Cook's D **f**)$\star$ AIC

Problem 31. A list of stationary processes is given below. Which of these processes will be the most difficult to distinguish from a non-stationary process based on the sample ACF?

a) random shocks
b) AR(1) with $\phi_1 = 0.5$
c)$\star$ AR(1) with $\phi_1 = 0.9$
d) MA(1) with $\theta_1 = -0.5$
e) MA(1) with $\theta_1 = 0.9$
f) MA(1) with $\theta_1 = 0.5$
g) AR(1) with $\phi_1 = -0.5$

Problem 32. Suppose $a_1, a_2, a_3, \ldots$ is a random shock sequence. What is the value of

$$E[(\psi_0 a_3 + \psi_1 a_2 + \psi_2 a_1)^2] \ ?$$

a) $\psi_0^2 a_3^2 + \psi_1^2 a_2^2 + \psi_2^2 a_1^2$
b) $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$
c)$\star$ $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$
d) $\sigma_z(\psi_0 + \psi_1 + \psi_2)$
e) $\sigma_z(\psi_0\psi_1 + \psi_1\psi_2)$
f) $\sigma_a(\psi_0 + \psi_1 + \psi_2)$
g) $\sigma_a(\psi_0\psi_1 + \psi_1\psi_2)$

Suppose that, based on a random sample of **400** observations, you fit a regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ for a response variable Y on two covariates X_1 and X_2. Use the information in the table below to answer the next two questions.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t-value	Pr $> \|t\|$
Intercept	1	4.0	0.5		$< .0001$
X1	1	40.0	10.0		$< .0001$
X2	1	24.0	16.0		.1336

Problem 33. The t-value entries in the table have been left blank. What is the t-value for the X1 row of the table?

a) .25 **b)** .0625 **c)** 2.0 **d)** 0.5 **e)** 1.5 **f)** 16.0 **g)**$\star$ 4.0

Problem 34. Assuming all the regression assumptions are valid, compute an approximate 95% confidence interval for β_1.

a) (39.02, 40.98) **b)** (6.08, 13.92) **c)** (8.04, 11.96) **d)** $(-68.4, 88.4)$
e) (38.04, 41.96) **f)**$\star$ (20.4, 59.6) **g)** (39.951, 40.049) **h)** (9.804, 10.196)

Problem 35. In regression, large values of H (the leverage) identify _________

a)$\star$ cases with unusual covariate values
b) influential covariates in the model
c) covariates which can be dropped
d) covariates with serial correlation
e) covariates which should be retained
f) serial correlation in the residuals
g) cases with unusual response values
h) influential cases in the data