TEST #1

STA 4853 **Name:**____________________________________

February 29, 2016

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 32 questions.
- 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 **15** pages.
- Each question is worth equal credit.

Problem 1. If you fit a regression model, and then plot the **residuals versus the fitted** (predicted) values, what do you expect to see **in this plot** (at least roughly) if the regression assumptions are valid?

a) the residuals will follow (roughly) a straight line with positive slope

b) the residuals will decay to zero gradually without a cutoff

c) the residuals will display an approximate cutoff to zero after lag p

d)⋆ The residuals form a band which remains centered at zero with a constant vertical width

e) All (or nearly all) of the p-values of the residuals will fall in the band, most of them being well inside the band

Problem 2. If you fit a regression model with n observations (cases) and p covariates, a value of the leverage H is considered "large" if it _________

a) is more than two times the standard error

b)⋆ exceeds $2(p+1)/n$

c) falls outside the range from -3 to 3

d) falls outside the range from -2 to 2

e) is close to 1 or larger

Problem 3. Suppose you are given a time series $z_1, z_2, \ldots, z_n$ which has been generated from an ARMA(2,1) process. If you use SAS PROC ARIMA to fit an ARMA(2,1) model to this series, SAS will compute **residuals** which ...

a) are the errors in the parameter estimates

b) give the deviations of the random shocks from their true values

c) minimize the sum of squared parameter estimates

d) are estimates of the errors in the parameter estimates

e) minimize the AIC

f)⋆ are estimates of the random shocks a_t

g) give the distance between z_t and the regression line

Problem 4. An ARMA(p,q) process is stationary if and only if _________ satisfy the conditions required for a _________ process to be stationary. Circle the response below with the choices (separated by a semi-colon) which correctly fill in the two blanks.

a)⋆ $\phi_1, \ldots, \phi_p$; AR(p)

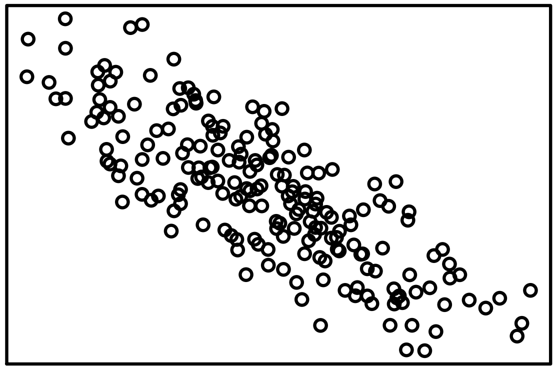
b) $\theta_1, \ldots, \theta_q$; ARMA(q,p)

c) $\phi_1, \ldots, \phi_p$; MA(q)

d) $\theta_1, \ldots, \theta_q$; MA(q)

Problem 5. For a stationary time series $z_1, \ldots, z_n$, suppose the scatterplot of z_t versus z_{t-4} looks like the plot given below. Then you know _________

a) $\rho_4 > 0$ **b**)$\star$ $\rho_4 < 0$ **c**) $\mu_z > 0$ **d**) $\mu_z < 0$ **e**) $\phi_4 < 0$ **f**) $\phi_4 > 0$
g) $\phi_{44} < 0$ **h**) $\phi_{44} > 0$ **i**) $\sigma_z^2 < 0$ **j**) $C > 0$ **k**) $C < 0$



Problem 6. A mean-centered ARMA(2,4) process may be written in backshift notation as

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

where $\phi(B)$ is a backshift polynomial given by _________

a) $\phi_1 B + \phi_2 B^2$
b) $\phi_1 B + \phi_2 B^2$
c) $1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4$
d) $\theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4$
e)$\star$ $1 - \phi_1 B - \phi_2 B^2$
f) $1 + \phi_1 B + \phi_2 B^2$

Problem 7. In the plots of the sample ACF and PACF, SAS displays a band marked at two standard errors ($\pm 2s(r_k)$ and $\pm 2s(\hat{\phi}_{kk})$). For an MA(3) process, we expect nearly all the spikes in the _________ to lie within the band.

a)$\star$ ACF at lags 4, 5, 6, 7,...
b) ACF at lags 3, 4, 5, 6, 7,...
c) PACF at lags 4, 5, 6, 7,...
d) PACF at lags 3, 4, 5, 6, 7,...

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) 0

b) C

c) ψ_1^2

d) ψ_1^6

e) $\sigma_a^2(\psi_0\psi_1\psi_2\psi_3 + \psi_2\psi_3\psi_4\psi_5)$

f) $\sigma_a^2(\psi_0\psi_1\psi_2\psi_3\psi_4\psi_5)$

g) $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \psi_4^2 + \psi_5^2)$

h) $\sigma_a^2(\psi_0\psi_1\psi_2 + \psi_3\psi_4\psi_5)$

i) $\sigma_a^2(\psi_0\psi_2 + \psi_1\psi_3 + \psi_2\psi_4 + \psi_3\psi_5)$

j) $\sigma_a^2(\psi_0\psi_4 + \psi_1\psi_5)$

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

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

m) $\dfrac{\sigma_a^2}{1-\psi_2}$

n) $\dfrac{\sigma_a^2}{1-\psi_2^2}$

o) $\dfrac{\sigma_a^2}{1-\psi_6}$

p) $\dfrac{\sigma_a^2}{1-\psi_6^2}$

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

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

Problem 8. Which of these responses gives the value of μ_z?

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

Problem 9. Which of these responses gives the value of σ_z^2?

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

Problem 10. Which of these responses gives the value of $\gamma_2 = \mathrm{Cov}(z_t, z_{t-2})$?

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

Problem 11. Which of these responses gives the value of $\gamma_6 = \text{Cov}(z_t, z_{t-6})$?

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

Problem 12. 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 ; **not** needed

b) reject ; needed

c)$\star$ do **not** reject ; **not** needed

d) do **not** reject ; needed

Problem 13. If a stationary ARMA(p, q) process with both $p > 0$ and $q > 0$ is written in MA(∞) form

$$\tilde{z}_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \dots ,$$

then the ψ-weights ψ_k will _________ as $k \to \infty$.

a) cutoff to zero after lag p

b) cutoff to zero after lag q

c)$\star$ eventually decay to zero

d) **not** converge to zero

e) increase in magnitude

f) will have no values inside the unit circle

g) will have a constant variance

Problem 14. As you add more covariates to a regression model, the value of R-squared always _________

a)$\star$ increases

b) decreases

c) eventually decays to zero

d) becomes more accurate

e) becomes less accurate

f) becomes statistically significant

g) becomes statistically insignificant

Problem 15. The backshift expression $(4 + 2B^2 + 3B^5)Z_t$ is equal to ________

a) $4Z_t + 2Z_{t+1}^2 + 3Z_{t+2}^5$
b) $4Z_t^4 + 2Z_{t-2}^2 + 3Z_{t-5}^3$
c) $4Z_t^4 + 2Z_{t+2}^2 + 3Z_{t+5}^3$
d)$\star$ $4Z_t + 2Z_{t-2} + 3Z_{t-5}$
e) $4Z_t + 2Z_{t+2} + 3Z_{t+5}$
f) $4Z_t + 2Z_t^2 + 3Z_t^5$
g) $4Z_t + 2Z_{t-1}^2 + 3Z_{t-2}^5$

Problem 16. The backshift expression $B(C + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1})$ is equal to ________

a) $C_{t+1} + \phi_2 z_t + a_{t+1} - \theta_2 a_t$
b) $C + \phi_2 z_{t-2} + a_{t-1} - \theta_2 a_{t-2}$
c) $0 + \phi_1 z_{t-2} + a_{t-1} - \theta_1 a_{t-2}$
d) $0 + \phi_0 z_{t-2} + a_{t-1} - \theta_0 a_{t-2}$
e) $0 + \phi_2 z_{t+1} + a_{t+1} - \theta_2 a_{t+2}$
f)$\star$ $C + \phi_1 z_{t-2} + a_{t-1} - \theta_1 a_{t-2}$
g) $C_{t-1} + \phi_0 z_{t-2} + a_{t-1} - \theta_0 a_{t-2}$

Problem 17. Which of the following is a correct expression for the **population correlation** between X and Y?

a) $\dfrac{\mathrm{Cov}(X,Y)^2}{\sigma_x^2 \sigma_y^2}$
b) $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$
c) $\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_x)^2(Y_i - \mu_y)^2$
d) $E(X - \overline{X})(Y - \overline{Y})$
e)$\star$ $\dfrac{E(X - \mu_x)(Y - \mu_y)}{\sigma_x \sigma_y}$
f) $\dfrac{c(X,Y)}{s_x s_y}$

Problem 18. If a series $z_1, z_2, \ldots, z_n$ is non-stationary in some way, examination of the sample ACF of the **entire** series may be helpful in detecting that the series ________

a) does **not** have a constant variance
b)$\star$ does **not** have a constant mean
c) does **not** have a constant ACF
d) is weakly stationary but **not** strictly stationary
e) has a large mean
f) has a large variance

Problem 19. An ARMA process is constructed from a sequence of random shocks a_t which are _________ random variables.

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

Problem 20. In SAS PROC ARIMA, the letters CLS refer to a _________.

a) test for serial correlation
b) test for normality of the residuals
c) method of forecasting
d) test for stationarity based on least squares
e)$\star$ method of parameter estimation
f) test for correlation based on least squares

Problem 21. 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".

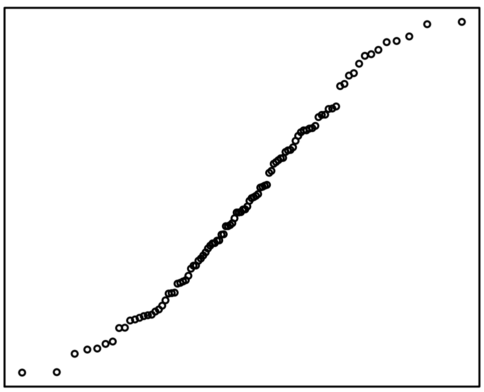
a)$\star$ $AIC_1 < AIC_2$ and $SBC_1 < SBC_2$
b) $AIC_1 < AIC_2$ and $SBC_1 > SBC_2$
c) $AIC_1 > AIC_2$ and $SBC_1 < SBC_2$
d) $AIC_1 > AIC_2$ and $SBC_1 > SBC_2$
e) $AIC_1 \approx AIC_2$ and $SBC_1 \approx SBC_2$
f) $AIC_1 \approx SBC_1$ and $AIC_2 \approx SBC_2$
g) $AIC_1 > SBC_1$ and $AIC_2 > SBC_2$

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

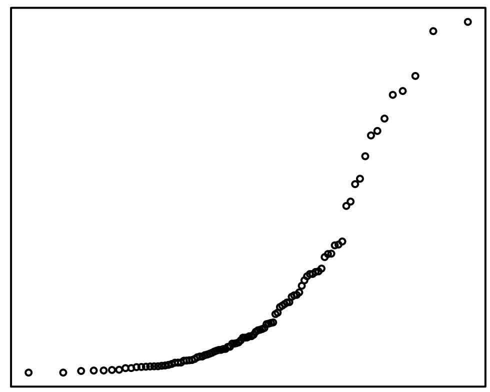
Problem 22. If the regression residuals exhibit strong serial correlation, then the time series plot of the residuals (the residuals plotted in time order) might resemble one of the following plots. Which one?

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

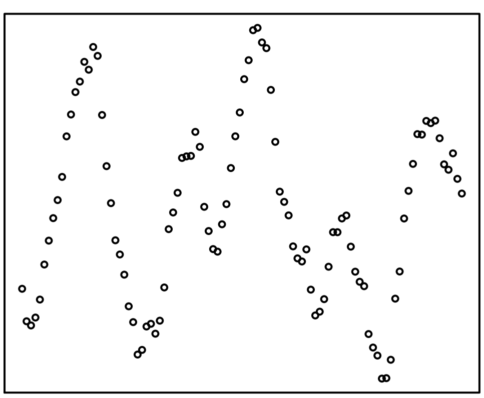
A



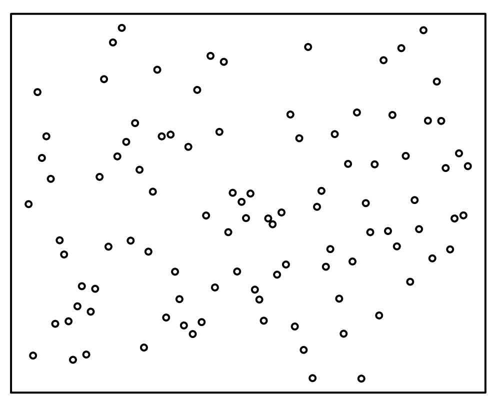
B



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

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

a) $\gamma_k = \dfrac{\rho_k}{\rho_0}$ **b)**$\star$ $\rho_k = \dfrac{\gamma_k}{\gamma_0}$ **c)** $\rho_k = \gamma_1^k$

d) $\gamma_k = \rho_1^k$ **e)** $\rho_k = \gamma_1 \rho_{k-1}$ **f)** $\gamma_k = \rho_1 \gamma_{k-1}$

Problem 25. The autocorrelations for a stationary AR(1) process obey the recursion _________

a) $\rho_k = \phi_1 \rho_{k+1} + a_k$
b) $\rho_k = C + \phi_1 \rho_{k-1} + a_k$
c)$\star$ $\rho_k = \phi_1 \rho_{k-1}$
d) $\tilde{\rho}_k = \phi_1 \tilde{\rho}_{k-1} + a_k$
e) $\rho_k = \phi_1(\rho_1 + \cdots + \rho_{k-1})$
f) $\rho_k = \dfrac{C}{1 - \rho_1 - \cdots - \rho_{k-1}}$
g) $\rho_k = \dfrac{\sigma_a^2}{1 - \rho_1^2 - \cdots - \rho_{k-1}^2}$

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

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

Problem 27. Assuming all the regression assumptions are valid, compute an approximate 95% confidence interval for β_1.

a) (8.04, 11.96) **b)** $(-68.4, 88.4)$ **c)** (39.02, 40.98) **d)** (6.08, 13.92)

e) (38.04, 41.96) **f)**$\star$ (20.4, 59.6) **g)** (39.951, 40.049) **h)** (9.804, 10.196)

In the remaining problems, you are asked to use the SAS output in the following five pages to classify five time series (z1, z2, z3, z4, z5) into the following categories:

- RS (random shocks)
- AR(1)
- AR(2)
- AR(3)
- MA(1)
- MA(2)
- MA(3)
- NCV (Series with non-constant variance)
- NCM (Series with non-constant mean)

No category is used more than once. All except the last two categories are stationary series.

Problem 28. What is z1?

a) RS **b)** AR(1) **c)** AR(2) **d)** AR(3) **e)** MA(1)
f)⋆ MA(2) **g)** MA(3) **h)** NCV **i)** NCM

Problem 29. What is z2?

a) RS **b)** AR(1) **c)** AR(2) **d)** AR(3) **e)** MA(1)
f) MA(2) **g)** MA(3) **h)**⋆ NCV **i)** NCM

Problem 30. What is z3?

a) RS **b)** AR(1) **c)** AR(2) **d)**⋆ AR(3) **e)** MA(1)
f) MA(2) **g)** MA(3) **h)** NCV **i)** NCM

Problem 31. What is z4?

a) RS **b)** AR(1) **c)** AR(2) **d)** AR(3) **e)** MA(1)
f) MA(2) **g)** MA(3) **h)** NCV **i)**⋆ NCM

Problem 32. What is z5?

a)⋆ RS **b)** AR(1) **c)** AR(2) **d)** AR(3) **e)** MA(1)
f) MA(2) **g)** MA(3) **h)** NCV **i)** NCM