TEST #1

STA 4853 **Name:**__________________________________

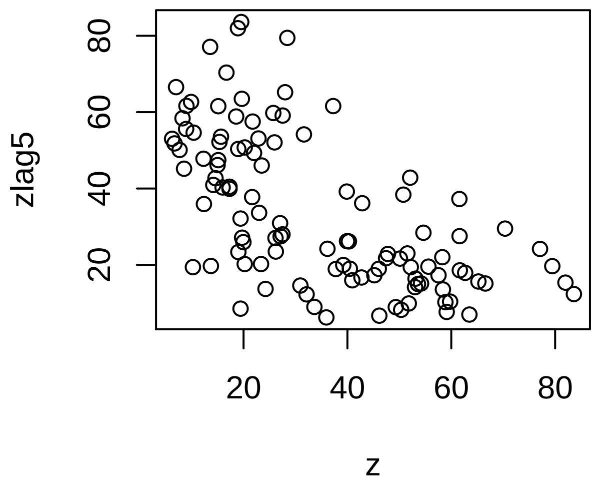
March 4, 2020

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

Suppose you are given data consisting of a time series $z_1, z_2, z_3, \ldots, z_{113}, z_{114}$ of length 114. Using all of this data, you construct the plot of z_t versus z_{t-5} (z lagged by 5) given below.



Problem 1. What is the number of points displayed in this plot?

a) 117 **b**) 116 **c**) 115 **d**) 114 **e**) 113 **f**) 112 **g**) 111
h) 110 **i**)⋆ 109 **j**) 108 **k**) 107 **l**) 106 **m**) 105 **n**) 104

Problem 2. For the time series z_t, the value of r_5, the sample autocorrelation at lag 5, will be _________.

a) between 20 and 80 **b**) between 20 and 50
c) between 50 and 80 **d**) between 0 and .5
e) between .5 and 1 **f**) between −80 and −20
g)⋆ between −1 and 0 **h**) between 0 and 1

Problem 3. The values of AIC and SBC (also called BIC) are used to compare different ARMA models for a time series. The quantities AIC and SBC depend on the likelihood L, the number of estimated parameters k, and the number of residuals n (which is often equal to the length of the time series). Both AIC and SBC contain a penalty term which penalizes model complexity. In SBC, this penalty term multiplies the number of estimated parameters by _________.

a) $\ln(L)$ **b**) -2 **c**)⋆ $\ln(n)$ **d**) 2
e) n **f**) L **g**) k **h**) $\ln(k)$

Problem 4. Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ denote the regression errors. SAS displays two different p-values for the Durbin-Watson statistic DW. One of them is $P(DW < DW_{\text{observed}})$. If this p-value is small, you reject the null hypothesis H_0 in favor of the alternative that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ _________.

a) are independent
b) are white noise
c)⋆ display positive serial correlation
d) display negative serial correlation
e) have increasing mean
f) have decreasing mean
g) have increasing variance
h) have decreasing variance

Problem 5. 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) $\tilde{z}_t = C/(1 - \phi_1 - \phi_2)$
d) $\tilde{z}_t = C/(1 - \phi_1 - \phi_2 + \theta_1)$
e) $\tilde{z}_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \tilde{a}_t - \theta_1 \tilde{a}_{t-1}$
f)⋆ $\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t - \theta_1 a_{t-1}$
g) $\tilde{z}_t = C/(1 - \phi_1 - \phi_2 - \theta_1)$
h) $\tilde{z}_t = C/(1 + \phi_1 + \phi_2)$

Suppose $a_1, a_2, a_3.a_4$ are random shocks (that is, they are independent $N(0, \sigma_a^2)$ random variables) and ψ_0, ψ_1, ψ_2 are constants.

Problem 6. What is the value of $E[(\psi_0 a_3 + \psi_1 a_2 + \psi_2 a_1)^2]$?

a)⋆ $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$
b) $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$
c) $\sigma_a^2\psi_0\psi_2$
d) 0
e) $\sigma_a^2/(1 - \psi_0 - \psi_1 - \psi_2)$
f) $\sigma_a^2/(1 - \psi_1^2)$
g) $\sigma_a^2/(1 - \psi_0^2 - \psi_1^2 - \psi_2^2)$
h) $\sigma_a^2/(1 - \psi_0\psi_1 - \psi_1\psi_2)$

Problem 7. What is the value of $E[(\psi_0 a_4 + \psi_1 a_3 + \psi_2 a_2)(\psi_0 a_3 + \psi_1 a_2 + \psi_2 a_1)]$?

a) $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$
b)⋆ $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$
c) $\sigma_a^2\psi_0\psi_2$
d) 0
e) $\sigma_a^2/(1 - \psi_0 - \psi_1 - \psi_2)$
f) $\sigma_a^2/(1 - \psi_1^2)$
g) $\sigma_a^2/(1 - \psi_0^2 - \psi_1^2 - \psi_2^2)$
h) $\sigma_a^2/(1 - \psi_0\psi_1 - \psi_1\psi_2)$

Problem 8. Let z_t be a stationary process. For any integer k the autocovariance is defined by $\gamma_k = \text{Cov}(z_t, z_{t-k}) = \text{Cov}(z_t, z_{t+k})$. In terms of the autocovariances, the autocorrelation ρ_k can be written as _________.

a) γ_1^k

b)$\star$ $\dfrac{\gamma_k}{\gamma_0}$

c) $\dfrac{\gamma_0}{1-\gamma_k}$

d) $\dfrac{\gamma_0}{1-\gamma_k^2}$

e) $\dfrac{\gamma_0}{1-\gamma_1-\gamma_2-\cdots-\gamma_k}$

f) $\dfrac{\gamma_0^2}{1-\gamma_1^2-\gamma_2^2-\cdots-\gamma_k^2}$

g) γ_0^k

h) $\phi_1\gamma_{k-1}$

Problem 9. The ACF and PACF plots produced by SAS PROC ARIMA both have bands about zero marked at two standard errors. For a stationary AR(p) process, after lag p we expect _________.

a) all (or nearly all) of the values in the **ACF** to lie within the band, most of them being well inside the band

b) the values in the **PACF** to decay gradually to zero

c)$\star$ all (or nearly all) of the values in the **PACF** to lie within the band, most of them being well inside the band

d) the values in the **ACF** to have an approximate cutoff to zero.

e) the theoretical **ACF** to be exactly equal to zero

f) the theoretical **PACF** to decay gradually to zero

Problem 10. For a stationary AR(3) process _________.

a) $\rho_3 \neq 0$ and $\rho_k = 0$ for $k > 3$

b) the band about the ACF has constant width

c)$\star$ $\phi_{33} \neq 0$ and $\phi_{kk} = 0$ for $k > 3$

d) the band about the PACF has increasing width

e) the autocovariances satisfy $\gamma_k = 0$ for $k > 3$

Problem 11. The table following this problem was produced by running the MINIC option of SAS PROC ARIMA on some time series data. According to this table, one plausibly good tentative model choice for this time series is _________.

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

f) MA(1) **g)** MA(2) **h)** MA(3) **i)** MA(4) **j)** MA(5)

k) ARMA(0,0) **l)** ARMA(1,1) **m)** ARMA(1,2) **n)**$\star$ ARMA(2,1) **o)** ARMA(2,2)

Minimum Information Criterion						
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	9.934709	9.740732	9.549696	9.445997	9.35974	9.306752
AR 1	9.224105	9.204932	9.201066	9.199017	9.199204	9.199919
AR 2	9.199627	9.197606	9.198216	9.198714	9.199619	9.20047
AR 3	9.198472	9.198318	9.198721	9.199591	9.200496	9.201311
AR 4	9.197776	9.198624	9.199534	9.200344	9.20109	9.201906
AR 5	9.198579	9.199474	9.200385	9.200989	9.201896	9.202815

Problem 12. The log of the Lynx time series exhibits oscillations of a somewhat periodic nature. Which of the following time series models (with an appropriate choice of parameter values) is also capable of exhibiting somewhat periodic oscillatory behavior?

a)$\star$ AR(2) **b)** MA(2) **c)** AR(1) **d)** MA(1) **e)** random shocks

Problem 13. When $\phi_{kk} = 0$, an approximate standard error for the estimated value $\hat{\phi}_{kk}$ is given by $s(\hat{\phi}_{kk}) =$ __________.

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) $\dfrac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$ **d)** $\dfrac{\sum_{t=1}^{n} e_t^2}{\sum_{t=2}^{n}(e_t - e_{t-1})^2}$

e)$\star$ $n^{-1/2}$ **f)** $n^{1/2}$ **g)** 1.96 **h)** 2×1.96

Problem 14. When performing a regression analysis of a response variable Y on covariates $X_1, X_2, \ldots, X_p$, cases with large values of the Leverage (also known as H) are ones with __________.

a)$\star$ unusual values of the covariates
b) large residuals
c) large serial correlation
d) large autocorrelation
e) spikes outside the shaded bands
f) insignificant P-values
g) unusual values of the response variable

h) a large influence on the estimated parameters and predicted values

Problem 15. If you use SAS PROC REG to perform a regression analysis of a response variable Y on covariates $X_1, X_2, \ldots, X_p$, the output will include a table of parameter estimates which gives an estimated regression parameter $\hat{\beta}_i$, a standard error $\text{SE}(\hat{\beta}_i)$, a t-Value, and a P-value for each covariate X_i in the model. If the P-value listed for X_i is small, then we will _________.

a) drop X_i from the model

b) decide that case i is an outlier

c)$\star$ reject the null hypothesis $H_0 : \beta_i = 0$

d) accept the null hypothesis $H_0 : \beta_i = 0$

e) decide that the residuals are NOT white noise

f) decide that the residuals ARE white noise

Problem 16. Let $r_1, r_2, r_3, \ldots$ be the sample autocorrelations of a stationary time series of length n. The statistic

$$Q(m) = n(n+2) \sum_{k=1}^{m} \frac{r_k^2}{n-k}$$

is used to test _________.

a) $H_0 : \phi_{mm} = 0$

b) $H_0 : \theta_m = 0$

c) $H_0 : \phi_m = 0$

d) the normality of the residuals

e) whether the variance is constant

f) whether the ACF is constant

g)$\star$ $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0$

h) $H_0 : \rho_m = 0$

Problem 17. Under the appropriate null hypothesis, the statistic $Q(m)$ defined in the previous problem will have approximately a _________.

a) $N(0,1)$ distribution

b) $N(\mu_z, \sigma_z^2)$ distribution

c) $N(0, \sigma_a^2)$ distribution

d) t-distribution with $n - m - 1$ degrees of freedom

e)$\star$ χ_m^2 distribution

f) $F_{m,n}$ distribution

Problem 18. The mean of a stationary ARMA(p, q) process is equal to _________.

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

b) C

c) $1 + 2\sum_{j=1}^{p-1} r_j^2$

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

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

f) $\sum_{i=1}^{\infty} \psi_i a_{t-i}$

Problem 19. Suppose a random sample of n individuals has values of X and Y given by $(X_1, Y_1), (X_2, Y_2,), \ldots, (X_n, Y_n)$. If the sample correlation r 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) lie exactly on a line with slope equal to one
b)$\star$ lie exactly on a line with positive slope
c) are very close to a line with slope exactly equal to one
d) are very close to a line with positive slope
e) are very close to the regression line $Y = \beta_0 + \beta_1 X + \varepsilon$
f) lie in a band of constant width centered about zero.
g) lie in a band of constant width centered about one.

Problem 20. Let $\{a_t\}$ be a sequence of random shocks. An AR(2) process is defined by _________.

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

Problem 21. By a process of repeated substitution, a stationary AR(1) process with $C = 0$ can be written in the form _________.

a) $z_t = a_t - \sum_{i=1}^{\infty} \theta_i z_{t-i}$
b) $z_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2}$
c) $z_t = a_t + \sum_{i=1}^{\infty} \phi_1^i z_{t-i}$
d)$\star$ $z_t = a_t + \sum_{i=1}^{\infty} \phi_1^i a_{t-i}$
e) $z_t = a_t + \phi_1 z_{t-1} + \phi_1^2 z_{t-2}$
f) $z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
g) $z_t = a_t + \phi_1 z_{t-1} + \phi_1^2 a_{t-2}$
h) $z_t = a_t + \theta_1 z_{t-1} + \theta_1^2 a_{t-2}$

Problem 22. The process $z_t = 10 + a_t + .8a_{t-1} - .5a_{t-2}$ has coefficient values _________.

a) $\theta_1 = -.8, \theta_2 = -.5$
b)$\star$ $\theta_1 = -.8, \theta_2 = .5$
c) $\theta_1 = .8, \theta_2 = -.5$
d) $\theta_1 = .8, \theta_2 = .5$
e) $\phi_1 = -.8, \phi_2 = -.5$
f) $\phi_1 = -.8, \phi_2 = .5$
g) $\phi_1 = .8, \phi_2 = -.5$
h) $\phi_1 = .8, \phi_2 = .5$

Problem 23. Suppose you use OLS (ordinary least squares) to fit a regression model. If the residuals exhibit strong serial correlation, then

a) your data contains unusual observations which should be deleted

b) the Leverage and Cook's D values could be far off and should not be relied upon

c) the Error Sum of Squares and R-squared values could be far off and should not be relied upon

d)$\star$ the standard errors, t-values, and p-values for the parameter estimates could be far off and should not be relied upon.

e) the regression model is valid so long as the residuals are normally distributed

f) the regression model is valid so long as the residuals are independent of the covariates

g) the regression model is valid so long as the residuals are normally distributed and independent of the covariates

Problem 24. In SAS PROC ARIMA, the statement

```
ESTIMATE P=(3);
```

results in SAS fitting which of the following models?

a)$\star$ $z_t = C + \phi_3 z_{t-3} + a_t$

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

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

d) $z_t = C + a_t - \theta_3 a_{t-3}$

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

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

Problem 25. 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) $\theta_1, \ldots, \theta_q$; ARMA(q,p)

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

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

d)$\star$ $\phi_1, \ldots, \phi_p$; AR(p)

Problem 26. For a random shock process, the theoretical ACF ________

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

Problem 27. For a stationary AR(1) process

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

it is true that $\mathrm{Var}(z_t) =$ ________

a) $\mathrm{Var}(C) + \phi_1 \mathrm{Var}(z_{t-1}) + \mathrm{Var}(a_t)$
b) $C + \phi_1 \mathrm{Var}(z_{t-1}) + \mathrm{Var}(a_t)$
c) $C^2 + \phi_1^2 \mathrm{Var}(z_{t-1}) + a_t$
d) $\phi_1 \mathrm{Var}(z_{t-1}) + a_t$
e)⋆ $\phi_1^2 \mathrm{Var}(z_{t-1}) + \mathrm{Var}(a_t)$
f) $\phi_1 \mathrm{Var}(z_{t-1}) + \mathrm{Var}(a_t)$
g) $\mathrm{Var}(C) + \phi_1 \mathrm{Var}(z_{t-1}) + a_t$
h) $C + \phi_1 \mathrm{Var}(z_{t-1}) + a_t$

Problem 28. 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 form a band which remains centered at zero with a constant vertical width
b) all (or nearly all) of the p-values of the residuals will fall in the band, most of them being well inside the band
c) the residuals will follow (roughly) a straight line with positive slope
d) the residuals will decay to zero gradually without a cutoff
e) the residuals will display an approximate cutoff to zero after lag p

---

A data set contains 103 observations and four variables, a response variable Y and three covariates X_1, X_2, X_3. A regression of Y on X_1, X_2, X_3 has been performed using SAS PROC REG. Some output is attached to the end of the exam which includes some plots of regression case diagnostics, and a listing of some of the data and case diagnostics. (Many observations have been omitted to save space.)

There are three unusual observations in the data.

Problem 29. Two of the three unusual observations have unusual **covariate** values. What are these observations? (Fill in the two blanks below with the correct observation numbers.)

Answer: 24 and 48 *(Each correct response is worth 1/2 point.)*

Problem 30. Two of the observations have unusual **response** values. What are these observations. (Fill in the two blanks below with the correct observation numbers.)

Answer: <u>24</u> and <u>83</u> *(Each correct response is worth 1/2 point.)*

Problem 31. One of the observations 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 it? (Fill in the single blank below with the correct observation number.)

Answer: <u>24</u> *(A correct response is worth 1 point.)*

The last 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 32. 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 33. 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 34. 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 35. 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.