**TEST #2**

**STA 4853** **Name:**______________________________________

**May 1, 2015**

**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 27 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 **10** pages.
- Each question is worth equal credit.

The next two problems involve this situation:

Suppose you have chosen to model time series data $z_1, z_2, \ldots, z_n$ using an AR(1) process $z_t = C + \phi_1 z_{t-1} + a_t$ with $C = 1$, $\phi_1 = 0.8$, and $\sigma_a = 1.5$. You decide to forecast future values $z_{n+1}$, $z_{n+2}$, $z_{n+3}$, ... of this time series.

**Problem 1.** As you forecast further and further into the future, the long run forecasts $\widehat{z}_{n+k}$ will _________

**a)** continue to gradually increase indefinitely
**b)** be white noise
**c)** be a random walk
**d)** continue to gradually decrease indefinitely
**e)** converge to a straight line with non-zero slope
**f)** converge to a repeating pattern
**g)**⋆ converge to a limiting constant

**Problem 2.** The confidence interval widths for the long run forecasts will _________

**a)** continue to gradually increase indefinitely
**b)** be white noise
**c)** be a random walk
**d)** continue to gradually decrease indefinitely
**e)** converge to a straight line with non-zero slope
**f)** converge to a repeating pattern
**g)**⋆ converge to a limiting constant

---

**Problem 3.** The ARMA(2,2) process

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

will be invertible if _________

**a)** $|\theta_1| < 1$
**b)** $|\phi_1| < 1$
**c)** the roots of the MA polynomial are strictly inside the unit circle
**d)** the roots of the AR polynomial are strictly inside the unit circle
**e)**⋆ $|\theta_2| < 1$, $\theta_2 + \theta_1 < 1$, and $\theta_2 - \theta_1 < 1$
**f)** $|\phi_2| < 1$, $\phi_2 + \phi_1 < 1$, and $\phi_2 - \phi_1 < 1$

**Problem 4.** Suppose you wish to construct a transfer function model explaining the response series $Y_t$ in terms of the input series $X_t$. You find that $X_t$ and $Y_t$ are not jointly stationary, but, after differencing, the differenced series $X_t^* = (1-B)X_t$ and $Y_t^* = (1-B)Y_t$ are jointly stationary. You then identify a transfer function and ARMA noise model for the differenced series which is

$$Y_t^* = \frac{B^b\omega(B)}{\delta(B)} X_t^* + \frac{\theta(B)}{\phi(B)} a_t \,.$$

In terms of the original series $X_t$ and $Y_t$, this model may be written as _________

**a)** $Y_t = \dfrac{B^b\omega(B)}{\delta(B)} X_t + \dfrac{(1-B)\theta(B)}{\phi(B)} a_t \,.$

**b)**⋆ $Y_t = \dfrac{B^b\omega(B)}{\delta(B)} X_t + \dfrac{\theta(B)}{(1-B)\phi(B)} a_t \,.$

**c)** $Y_t = \dfrac{B^b\omega(B)}{\delta(B)} (1-B)X_t + \dfrac{\theta(B)}{\phi(B)} a_t \,.$

**d)** $Y_t = \dfrac{B^b\omega(B)}{(1-B)\delta(B)} X_t + \dfrac{\theta(B)}{\phi(B)} a_t \,.$

**e)** $(1-B)Y_t = \dfrac{B^b\omega(B)}{\delta(B)} X_t + \dfrac{\theta(B)}{\phi(B)} a_t \,.$

**f)** $\dfrac{Y_t}{1-B} = \dfrac{B^b\omega(B)}{\delta(B)} X_t + \dfrac{\theta(B)}{\phi(B)} a_t \,.$

**Problem 5.** After fitting a time series model, the plot of the residuals versus the one-step-ahead forecasts is often used to determine if the _________

**a)** mean of the series varies with time
**b)** ACF of the series varies with time
**c)** residuals are normally distributed
**d)**⋆ variability of the residuals changes with the level of the series
**e)** one-step-ahead forecasts change with the level of the series
**f)** series needs further differencing
**g)** one-step-ahead forecasts are normally distributed

**Problem 6.** For a specified ARIMA$(p, d, q)$ process $z_t$, suppose you know all the values of the series $z_t$ and the random shocks $a_t$ for times $t \leq n$ (call this information set $\mathcal{I}_n$), and you wish to predict future values $z_{n+k}$. If you are using "squared error loss," then the best predictions $\widehat{z}_{n+k}$ are given by $\widehat{z}_{n+k} =$ _________

**a)**⋆ $E(z_{n+k} \,|\, \mathcal{I}_n)$   **b)** $\mathrm{Var}(z_{n+k} \,|\, \mathcal{I}_n)$   **c)** $\sigma_a\sqrt{1+\psi_1^2+\cdots+\psi_k^2}$

**d)** $z_n$ (the last known value)   **e)** $a_n$   **f)** $\widehat{z}_n$   **g)** 0

**Problem 7.** Suppose you are modeling a long time series (of length $n$) named `Z`, you have chosen the correct model, and the parameter estimates are accurate. The output file created by the `FORECAST` statement contains columns with the headings `Z, FORECAST, STD, L95, U95`, and `RESIDUAL`. Prior to time $n$ (the end of the series), you expect _________ of the values in column _________ to lie between the values in columns `L95` and `U95`. (Circle the choice below with the responses that best complete the two blanks.)

**a)** about 95%, `FORECAST`
**b)** less than 95%, `FORECAST`
**c)** nearly all, `Z`
**d)**$\star$ about 95%, `Z`
**e)** about 95%, `RESIDUAL`
**f)** nearly all, `RESIDUAL`
**g)** more than 95%, `RESIDUAL`
**h)** about 95%, `STD`
**i)** nearly all, `STD`

**Problem 8.** Suppose $z_t$ is a **quarterly** series with a **non**stationary mean. If the first differences $\nabla z_t$ appear stationary but have substantial autocorrelations at lags 1, 4, and 8, which of the following options might you wish to pursue? (More than one may be reasonable.)

1. Try a model without any differencing.
2. Try differencing at lag 1 a second time.
3. Try replacing differencing at lag 1 by differencing at lag 4.
4. Try a seasonal model which includes a seasonal term at lag 4.
5. Try an MA(1) model on $\nabla z_t$.
6. Try an AR(1) model on $\nabla z_t$.

Select the pair of options which seem most reasonable and circle your choice **below**. (Do NOT circle items on the list above!)

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

**Problem 9.** You wish to predict a quantity $X$ on the basis of information $\mathcal{I}$. If you will be forced to pay \$100 unless your prediction $\widehat{X}$ is within $\varepsilon$ of X (where $\varepsilon$ is small), then your best prediction is the _________ of the conditional distribution of $X$ given $\mathcal{I}$.

**a)**$\star$ mode **b)** variance **c)** standard deviation
**d)** mean **e)** median **f)** minimum **g)** $P$-value

**Problem 10.** In `PROC ARIMA`, an `ESTIMATE` statement which includes

```
INPUT=(3$(1,2)/(1)X)
```

is instructing SAS to use a transfer function of the form ________

**a)** $\dfrac{(1-B^3)(\omega_1 B-\omega_2 B^2)}{1-\delta_1 B}X_t$

**b)**⋆ $\dfrac{B^3(\omega_0-\omega_1 B-\omega_2 B^2)}{1-\delta_1 B}X_t$

**c)** $\dfrac{B^3(1-\omega_1 B-\omega_2 B^2)}{\delta_0-\delta_1 B}X_t$

**d)** $\dfrac{3B(1-\omega_1 B-\omega_2 B^2)}{\delta_0-\delta_1 B}X_t$

**e)** $\dfrac{(1-B^3)(\omega_1 B-\omega_2 B^2)}{1-B}X_t$

**f)** $\dfrac{3B(1-\omega_1 B-\omega_2 B^2)}{1-B}X_t$

**Problem 11.** The theoretical ACF of a purely seasonal ARIMA$(0,0,2)_{\mathbf{10}}$ process will be non-zero only at the lags ________

**a)** 1, 2, 10, 20

**b)** 10, 20, 30, 40, ...

**c)** 1, 2, 10, 20, 30, 40, ...

**d)** 1, 2, 10, 11, 12, 13, ...

**e)**⋆ 10, 20

**f)** 20

**g)** 10, 11, 12, 13, ...

**Problem 12.** For an AR(1) process $z_t = C + \phi_1 z_{t-1} + a_t$, if you are given $\mathcal{I}_n$ (information up to time $n$), the forecast $\widehat{z}_{n+2}$ is ________

**a)**⋆ $C + \phi_1 C + \phi_1^2 z_n$

**b)** $C - \phi_1 a_n - \phi_1 a_{n-1}$

**c)** $C - \phi_1 a_n$

**d)** $C + z_n$

**e)** $C$

**f)** $z_n$

**g)** $\phi_1 z_n$

**Problem 13.** Suppose the series $\{X_t\}$ and $\{Y_t\}$ are jointly stationary and we wish to construct a transfer function model $Y_t = C + v(B)X_t + N_t$. In the pre-whitening process, the goal is to apply a filter $f(B)$ to both $X_t$ and $Y_t$ to produce new series $X'_t = f(B)X_t$ and $Y'_t = f(B)Y_t$ such that _________.

- **a)** $X'_t$ and $Y'_t$ are uncorrelated
- **b)** $X'_t$ and $Y'_t$ are jointly stationary
- **c)** $X'_t$ and $Y'_t$ are independent
- **d)** their CCF is close to zero
- **e)** both $X'_t$ and $Y'_t$ are white noise
- **f)**$\star$ $X'_t$ is white noise
- **g)** $Y'_t$ is white noise

**Problem 14.** The AR-polynomial of an ARIMA$(2,0,1)(1,0,2)_{12}$ process is _________

- **a)**$\star$ $(1-\phi_1 B-\phi_2 B^2)(1-\Phi_1 B^{12})$
- **b)** $1-\phi_1 B-\phi_2 B^2-\Phi_1 B^{12}$
- **c)** $(1-\phi_1 B)(1-\Phi_1 B^{12}-\Phi_2 B^{24})$
- **d)** $1-\phi_1 B-\Phi_1 B^{12}-\Phi_2 B^{24}$
- **e)** $(1-\phi_1 B-\phi_2 B^2)(1-\Phi_1 B^{12}-\Phi_2 B^{24})$
- **f)** $1-\phi_1 B-\phi_2 B^2-\Phi_1 B^{12}-\Phi_2 B^{24}$
- **g)** $(1-B)^2(1-B^{12})(1-\phi_1 B-\phi_2 B^2)$

**Problem 15.** $\nabla_s z_t =$ _________

**a)** $(1-B)^s z_t$ **b)** $(B-1)^s z_t$ **c)** $(1-B)^t z_s$ **d)** $z_{t+s}-z_{t+s-1}$

**e)** $z_{t+s}-z_t$ **f)**$\star$ $z_t-z_{t-s}$ **g)** $z_{t-s}-z_{t-s-1}$

**Problem 16.** The long range forecasts from a seasonal ARIMA model with seasonal lag $S$, orders of differencing $d=0$ and $D=1$, and a non-zero constant $C \neq 0$ will converge to _________

- **a)**$\star$ a repeating pattern (with period $S$) added to a straight line with non-zero slope
- **b)** a repeating pattern with period $S$
- **c)** a straight line with non-zero slope
- **d)** the estimated mean of the process
- **e)** the estimated variance of the process
- **f)** a value depending mainly on the last values of the series
- **g)** a straight line with slope depending mainly on the last values of the series

**Problem 17.** Suppose you are given a time series $X_1, X_2, \ldots, X_n$. You decide to log transform the data and work with the transformed series $Y_t = \log X_t$. You find a good ARIMA model for $Y_t$ and use this to forecast the value $Y_{n+5}$. SAS gives you a 95% confidence interval $(L, U)$ for $Y_{n+5}$. From this you can obtain a 95% confidence interval for $X_{n+5}$. What is this confidence interval?

**a)** $(e^{\mu}, e^{\mu+\sigma^2/2})$ **b)** $(L - 1.96\sigma, U + 1.96\sigma)$ **c)** $(\log L, \log U)$

**d)** $(\sqrt{L}, \sqrt{U})$ **e)**$\star$ $(e^L, e^U)$ **f)** $(L^2, U^2)$

**Problem 18.** The next table gives the estimation output from fitting a certain model using PROC ARIMA. By *very carefully* reading this table, one can determine the model that was fit. What is the SAS code for specifying this model? Circle the correct response below.

**a)** `P=(1,2,10) Q=(3)`
**b)**$\star$ `P=(1,2)(10) Q=(3)`
**c)** `P=(1,2)(10) Q=(1)`
**d)** `P=(1,2,10) Q=(1)`
**e)** `P=(1,2,10) Q=3`
**f)** `P=3 Q=1`
**g)** `P=(3) Q=(1)`
**h)** `P=10 Q=3`

**The SAS System**

**The ARIMA Procedure**

| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | 6.72100 | 0.23875 | 28.15 | <.0001 | 0 |
| MA1,1 | -0.32195 | 0.10667 | -3.02 | 0.0025 | 3 |
| AR1,1 | 1.30023 | 0.07129 | 18.24 | <.0001 | 1 |
| AR1,2 | -0.67083 | 0.07549 | -8.89 | <.0001 | 2 |
| AR2,1 | 0.36009 | 0.09234 | 3.90 | <.0001 | 10 |

**Problem 19.** A transfer function has $v$-weights given in the table below:

| $v_0$ | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ | ... |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 2 | 1 | 0 | 0 | 0 | ... |

This can be expressed as a rational distributed lag model with particular values of $b$, $h$, and $r$ (i.e., the delay, numerator order, and denominator order) as described in lecture. What are these values?

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

**Problem 20.** The table immediately following this problem gives the ACF of a stationary process with seasonality $s = 8$. The ACF is exactly zero for all lags from 11 onward. What type of ARIMA process is this?

**a)** $(2,0,0)(1,0,0)_8$
**b)** $(1,0,0)(2,0,0)_8$
**c)** $(0,0,2)(0,0,5)_8$
**d)** $(0,0,2)(5,0,0)_8$
**e)** $(2,0,0)(0,0,5)_8$
**f)**$\star$ $(0,0,2)(0,0,1)_8$
**g)** $(0,0,2)(1,0,0)_8$
**h)** $(2,0,0)(5,0,0)_8$
**i)** $(2,0,0)(0,0,1)_8$

| LAG | ACF |
|---|---|
| 1 | -0.261194 |
| 2 | -0.223881 |
| 3 | 0.0 |
| 4 | 0.0 |
| 5 | 0.0 |
| 6 | -0.098771 |
| 7 | -0.115233 |
| 8 | 0.441176 |
| 9 | -0.115233 |
| 10 | -0.098771 |
| 11 | 0.0 |
| 12 | 0.0 |

The next two problems concern this situation: Suppose $\{z_t\}$ is an ARIMA process generated by the random shocks $\{a_t\}$. We observe **all** the values $z_t$ and $a_t$ up to time $n$. Call this set of information $\mathcal{I}_n$:

$$\mathcal{I}_n = \{z_n, z_{n-1}, z_{n-2}, \ldots, a_n, a_{n-1}, a_{n-2}, \ldots\}$$

We define $\widehat{z}_t = E(z_t \,|\, \mathcal{I}_n)$ and $\widehat{a}_t = E(a_t \,|\, \mathcal{I}_n)$.

**Problem 21.** Which of the following statements about $\widehat{z}_t$ is always true?

**a)** $\widehat{z}_t = \mu_z$ for $t > n$
**b)** $\widehat{z}_t = 0$ for $t \leq n$
**c)** $\widehat{z}_t = 0$ for $t > n$
**d)** $\widehat{z}_t = z_t$ for $t > n$
**e)** $\widehat{z}_t = \mu_z$ for $t \leq n$
**f)**$\star$ $\widehat{z}_t = z_t$ for $t \leq n$

**Problem 22.** Which of the following statements about $\widehat{a}_t$ is always true?

**a)** $\widehat{a}_t = 0$ for $t \leq n$ and $\widehat{a}_t = \sigma_a^2$ for $t > n$
**b)** $\widehat{a}_t = a_t$ for $t \leq n$ and $\widehat{a}_t = z_t$ for $t > n$
**c)** $\widehat{a}_t = \sigma_a^2$ for $t = n$ and $\widehat{a}_t = 0$ for $t \neq n$
**d)**$\star$ $\widehat{a}_t = a_t$ for $t \leq n$ and $\widehat{a}_t = 0$ for $t > n$
**e)** $\widehat{a}_t = 0$ for $t \leq n$ and $\widehat{a}_t = a_t$ for $t > n$
**f)** $\widehat{a}_t = z_t$ for $t \leq n$ and $\widehat{a}_t = a_t$ for $t > n$
**g)** $\widehat{a}_t = \sigma_a^2$ for $t \leq n$ and $\widehat{a}_t = 0$ for $t > n$

**Problem 23.** If $\{X_t\}$ and $\{Y_t\}$ are jointly stationary, then the cross-correlation function is defined by $\rho_{xy}(s) = \ldots$

**a)** $\text{Corr}(X_{t-s}, Y_{t-s})$
**b)** $\text{Corr}(X_{t+s}, Y_{t+s})$
**c)** $\text{Corr}(X_{t+s}, Y_t)$
**d)** $\text{Corr}(X_{t+s}, Y_{t-s})$
**e)**$\star$ $\text{Corr}(X_t, Y_{t+s})$
**f)** $\text{Corr}(X_{t-s}, Y_{t+s})$

The following two questions involve this situation:

Based on time series data $Y_1, Y_2, \ldots, Y_{100}$, SAS has fit an ARIMA model and produced forecasts for future values of $Y_t$. A fragment of SAS output is given below.

| Obs | Forecast | Std Error | 95% Confidence Limits | |
|---|---|---|---|---|
| 101 | 6.5068 | 1.2111 | 4.1331 | 8.8804 |
| 102 | 6.8077 | 1.2429 | 4.3716 | 9.2437 |

We are interested in forecasts for time $t = 102$. Assuming the model is correct and the parameter estimates are accurate, based on this output we can say that $Y_{102}$ is normally distributed with a mean of $\mu =$ ☐ and a standard deviation of $\sigma =$ ☐.

**Problem 24.** Circle the correct value of $\mu$.

**a)**$\star$ 6.8077 **b)** 1.2429 **c)** 4.3716 **d)** 9.2437

**e)** 6.5068 **f)** 1.2111 **g)** 4.1331 **h)** 8.8804

**Problem 25.** Circle the correct value of $\sigma$.

**a)** 6.8077 **b)**$\star$ 1.2429 **c)** 4.3716 **d)** 9.2437

**e)** 6.5068 **f)** 1.2111 **g)** 4.1331 **h)** 8.8804

**Problem 26.** For an ARIMA$(p, d, q)$ process $\{z_t\}$, the variance of the one-step-ahead prediction errors ...

**a)** is equal to $\sigma_z^2$. **b)** depends on $\phi_1, \ldots, \phi_p$, and $\sigma_a^2$. **c)**$\star$ is equal to $\sigma_a^2$.

**d)** depends on $d$. **e)** depends on $\theta_1, \ldots, \theta_q$. **f)** depends on $d$, $\phi_1, \ldots, \phi_p$, and $\sigma_z^2$.

**Problem 27.** The minimum information criterion (MINIC) method can tentatively identify the order of a stationary and invertible ARMA process. For a range of values for $p$ and $q$, a table supplies an estimate of the ________ divided by $n$ (the series length) that would be obtained if an ARMA$(p, q)$ model were fit to the data. The models with the ________ tabled values are worth considering as possible models.

The sentences above contain two blanks. Select the response below which gives two words which correctly fill these blanks.

**a)** $P$-value, smallest **b)** $P$-value, largest **c)** AIC, smallest **d)** AIC, largest

**e)**$\star$ BIC, smallest **f)** BIC, largest **g)** Chi-Square, smallest **h)** Chi-Square, largest