DIRECTIONS:

- This homework is based on the first week of Time Series lectures. If you wish, you may view a **recorded lecture** to review Topic 11 Part 1 Examples 2 and 3.

  (Access Select Video Recordings from the main Stats website. Headphones are recommended.)

- It may be convenient to print these directions and write answers in the available space.
- Data files for all exercises are found on the MINITAB Data Sets link from the course website.

A. Chicago Cubs Home Attendance

The Chicago Cubs are a professional baseball team, one of the oldest active teams in American sports. The data set `Cubs` shows average home-game attendance for the years 1916–2007.

The response variable is

$$y_t = \text{ Average Cubs home-game attendance in year } t$$

In this problem we choose between three competing forecast models:

- MA(2) or moving average with time span $k = 2$ years
- MA(4) or moving average with time span $k = 4$ years
- Simple Regression Trend model

based on *historical* performance, then we measure which model actually performs best for *future* predictions, all from the vantage point of the year 1960.

A1. DIRECTIONS:

1. Open the `Cubs` data. Plot attendance figures over the years 1916–2007:
   Stat > Time Series > Time Series Plot > (Simple is OK) > (Select Attendance) > OK

2. Notice that the "Index" variable on the horizontal axis is not very informative! So modify the plot by adding a "time stamp" to more clearly see attendance patterns over different years:
   Stat > Time Series > Time Series Plot > (Simple is OK) > (Select Attendance) > Time/Scale > Stamp > (Select Year as the time stamp) > OK > OK

A2. For this next part, imagine that you are the business manager for the Cubs in the year 1960. Your goal is to predict Cubs attendance for the *next* ten years (1960–1969) based on the *past* ten years (1950–1959.)

DIRECTIONS:

- Focus on the relevant data by plotting Cubs attendance over the 1950's and 1960's only. (Be sure to use the appropriate time stamp.)

QUESTIONS:

(a) Find the *historical* accuracy MSE = Mean Squared prediction Error for the MA(2) model over the 1950's, step-by-step:

1. Find predictions $\widehat{y}_t$.
2. Find prediction errors $e_t = y_t - \widehat{y}_t$.
3. Find squared prediction errors e_t^2.
4. Find MSE by averaging the squared errors.

The following worksheet provides a *template* for your calculations. (It may also help to refer to Example 1 in the Topic 11 Notes.)

Year t	**Attendance** y_t	**MA(2)** $\widehat{y}_t$	**Prediction Error** $e_t = y_t - \widehat{y}_t$	**Squared Prediction Error**
1950	14,948	*	*	*
1951	11,616	*	*	*
1952	13,309	13,282	+27	729.00
1953	9,918	12,462.5	−2544.5	6,474,480.25
1954	9,717	11,613.5	−1896.5	3,596,712.25
1955	11,374	_____	_____	_____
1956	9,001	_____	_____	_____
1957	8,598	10,187.5	−1589.5	2,526,510.25
1958	12,726	8799.5	+3,926.5	15,417,402.25
1959	11,146	_____	_____	_____
				_____

(continued)

(b) Find the *historical* accuracy MSE for the MA(4) model over the 1950's, step-by-step.

The following worksheet provides a *template* for your calculations.

Year t	**Attendance** y_t	**MA(4)** $\widehat{y}_t$	**Prediction Error** $e_t = y_t - \widehat{y}_t$	**Squared Prediction Error**
1950	14,948	*	*	*
1951	11,616	*	*	*
1952	13,309	*	*	*
1953	9,918	*	*	*
1954	9,717	12,447.75	−2730.75	7,456,995.56
1955	11,374	_____	_____	_____
1956	9,001	_____	_____	_____
1957	8,598	10,002.5	−1404.5	1,972,620.25
1958	12,726	9672.5	+3053.5	9,323,862.25
1959	11,146	_____	_____	_____
				_____

(c) Confirm your hand calculations for the MA(2) and MA(4) models by using the MINITAB Moving Average Time Series procedure. Does MINITAB show the same answers for MSE?

(d) Based on recent history (1950's), which MA model should the manager choose in 1960 to forecast attendance in the 1960's, MA(2) or MA(4)?

(e) Add a third model to the mix: Forecast trend by simple linear regression:

Stat > Regression > Regression
> (Response = Attend-50's, Predictor = Year-50s) > OK

What is the model's historical performance MSE? MSE = _____

(Hint: "SS Residual Error" in the regression output is the "sum of squared prediction errors" so just divide this number by 10 to get MSE.)

Note: "MS Residual Error" in the regression output is not the correct answer since this statistic divides total squared error by $(n-2) = 10-2 = 8$ "degrees of freedom" instead of 10.

(continued)

(f) Does the Simple Regression Trend model outperform the MA(2) and MA(4) models historically? (This is sometimes referred to as "in-sample" performance.) Explain why or why not.

Note: Historical performance does not guarantee future success. Which model makes the best *future forecasts* over the 1960's? (This is sometimes referred to as "out-of-sample" performance.)

(g) To save time we'll use MINITAB instead of hand calculations to measure MA(2) accuracy over the 1960's. Since 1958 and 1959 attendance figures are available for the 1960 forecast:

$$\widehat{y}_{1960} = \frac{y_{1959} + y_{1958}}{2}$$

we apply the Moving Average procedure to the combined 1950's-1960's data, then extract the errors for 1960–1969:

- Stat > Time Series > Moving Average > (Choose variable Attend-50s-60s) > (MA length = 2) > Time > (Select Year-50s-60s as time stamp) > Storage > (Check Residuals box) > OK > OK

- Notice that a new MINITAB column named "RESI1" has appeared. Cut-and-paste the last 10 years of residuals (for years 1960-1969) from this column into a new column. Label the new column "MA2 Errors".

- Label another (empty) column "MA2 Squared Errors". Use the MINITAB Calculator to square the errors:

  Calc > Calculator > (Store result in MA2 Squared Errors)

  > (Select MA2 Errors, calculate the expression MA2 Errors**2) > OK

- Finally, calculate MSE by selecting the Sum function from the scroll-down box in the MINITAB Calculator and evaluating the expression

  SUM(MA2 Squared Errors)/10

  (Store the answer into any empty column.)

  MSE = _____

(h) Find MSE for future forecasts (1960–1969) for the MA(4) model similarly.

MSE = _____

(continued)

(i) Find MSE for future forecasts (1960-1969) for the Simple Regression Trend model:

- Stat > Regression > Regression
  > (Response = Attend-50s, Predictor = Year-50s) > Options
  > (Enter Year-60s into "Prediction Intervals" box, check Storage Fits)
  > OK > OK

- Notice that a new column "PFIT1" shows the Regression forecasts for 1960–1969.

  Label a new column "Reg Errors" and fill the column by subtracting "PFIT1" from "Attend-60s" in the MINITAB Calculator.)

  Continue on to calculate MSE, as you did in (g) and (h).

  MSE = _____

(j) Which model provides the most accurate future forecasts over 1960–1969: MA(2), MA(4), or Simple Regression Trend? Explain.

Is this the same model that performed best historically over 1950–1959?

(k) Compare the MINITAB plots for MA(2) and MA(4). Can you explain why MA(2) performs better than MA(4) over 1960–1969, even though MA(4) performed better over 1950–1959?

B. Chicago Cubs, Revisited

Continue from Problem A. Compare the same three models:

- MA(2) or moving average with time span $k = 2$ years
- MA(4) or moving average with time span $k = 4$ years
- Simple Regression Trend model

from the vantage point of the year 2000. Your goal is to predict Cubs attendance for all *future* years (2000–2007) based on the *past* ten years (1990–1999.)

DIRECTIONS:

- Make a time-series plot over the relevant data (1990–2007) with appropriate time stamp. Can you guess which model will do best over 1990–1999? Over 2000–2007?
- Cut-and-paste columns in the `Cubs` data set to create new columns, as needed.

QUESTIONS:

(a) Find MSE for MA(2) over historical data (1990–1999).

(b) Find MSE for MA(4) over historical data.

(c) Find MSE for Simple Regression Trend over historical data.

(d) Which model should you choose in the year 2000 to make future forecasts, based on historical performance?

(e) Find MSE for MA(2) for future forecasts (2000–2007).

(f) Find MSE for MA(4) for future forecasts.

(g) Find MSE for Simple Regression Trend for future forecasts.

(h) Which model provides the most accurate future forecasts?

(i) Is the best model historically the same as the best model for future forecasts?

(j) Multiple Choice. (Choose the best answer: (A) or (B).)

SRT performs poorly in the 2000's since SRT forecasts are consistently

(A) too low
(B) too high

compared to actual Cubs attendance in each of the years 2000–2007.

(Hint: You may wish to refer to your MINITAB worksheet or make a plot.)

C. Exponential Smoothing for Gas Sales (first 10 weeks of data for Topic 11 Example 1)

(a) Fill in the blanks below for exponential smoothing with ($w = 0.10$). The first three calculations are shown below to help you get started.

(Round all calculations to 2 decimal places.)

$\widehat{y}_1 = y_1 = 18$ (the first forecast is just taken to be the first value)

$$\begin{aligned}\widehat{y}_2 &= (0.10)y_1 + (0.90)\widehat{y}_1 \\ &= (0.10)(18) + (0.90)(18) = 18\end{aligned}$$

$$\begin{aligned}\widehat{y}_3 &= (0.10)y_2 + (0.90)\widehat{y}_2 \\ &= (0.10)(21) + (0.90)(18) \\ &= 2.1 + 16.2 = 18.3\end{aligned}$$

Week t	**Gas Sales** y_t	**($w = 0.10$)** $\widehat{y}_t$	**Prediction Error** $e_t = y_t - \widehat{y}_t$	**Squared Prediction Error**
1	18	18	0	0
2	21	18	+3	9.00
3	19	18.3	+0.7	0.49
4	22	_____	_____	_____
5	17	_____	_____	_____
6	17	_____	_____	_____
7	20	18.40	+1.60	2.55
8	19	18.56	+0.44	0.19
9	11	18.61	−7.61	57.87
10	25	_____	_____	_____
				_____

(b) Calculate MSE for ($w = 0.10$) from the table. _____

(continued)

(c) Fill in the blanks below for exponential smoothing with ($w = 0.90$).

(Round all calculations to 2 decimal places.)

Week t	**Gas Sales** y_t	**($w = 0.90$)** $\widehat{y}_t$	**Prediction Error** $e_t = y_t - \widehat{y}_t$	**Squared Prediction Error**
1	18	18	0	0
2	21	18	+3	9.00
3	19	20.7	−1.70	2.89
4	22	_____	_____	_____
5	17	_____	_____	_____
6	17	17.47	−0.47	0.22
7	20	17.05	+2.95	8.72
8	19	19.70	−0.70	0.50
9	11	19.07	−8.07	65.13
10	25	_____	_____	_____
				_____

(d) Calculate MSE for ($w = 0.90$) from the table. _____

(e) Use MINITAB Single Exponential Smoothing to verify your MSE calculations for ($w = 0.10$) and ($w = 0.90$). Find exact MINITAB values for MSD (same as MSE.)

Use the steps below and notice that you must use an option to force MINITAB to set the first forecast equal to the first value, as we did in the table:

Stat > Time Series > Single Exponential Smoothing > (Choose variable Gas Sales) > (Use Weight 0.10) > Options > (Set K = 1) > OK > OK

(f) Use MINITAB to find MSE for ($w = 0.50$) _____

(g) Which ES model performs best over Weeks 1–10, ($w = 0.10$), ($w = 0.90$) or ($w = 0.50$)?

D. Chicago Cubs, Once More

(a) Apply MA(5), MA(10), ES ($w = 0.10$), ES ($w = 0.90$), ES ($w = 0.50$) to the *entire* time series for Chicago Cubs average home-game attendance (1916–2007.) Which model performs best over this entire time period?

(b) Cubs average home-game attendance in 2008 was 40,743. Which of the five models provides the most accurate forecast for 2008 from the vantage point of the year 2007?

(Hint: One way to do this is by forecasting one period ahead with MINITAB Exponential Smoothing.)

E. Chinese Car Ownership

China's rapid economic growth can be measured in many dimensions. The data set `China Cars` shows the number of cars owned by individuals in China (in tens of thousands) from 1990 to 2006.

DIRECTIONS:

1. Turn to Example 3 on page 29 in the Topic 11 Notes. Open the file `Blackberry Sales` and make a time-series plot to reproduce the plot on page 29. Use Week as the time stamp.

2. Also make a Fitted-Line Plot, as on page 29.

3. Apply MINITAB Trend Analysis to Blackberry Sales. Are you able to reproduce the plots on pages 30–31?

4. Produce the residuals plot at the top of page 32. Apply a MINITAB Runs Test to the residuals, as well as Partial Autocorrelation and ARIMA. Did you get the output in the Notebook? Review the remainder of the Example 3 Notes to understand the entire analysis.

5. Now apply Steps 1–4 above to the `China Cars` data, with the goal of forecasting the number of cars in 2007, 2010, and 2013.

QUESTIONS:

(a) Select the trend pattern which fits the data best. Write down the prediction equation of the best-fitting trend. What is the trend's MSE?

(b) According to the trend equation, by what percentage does Chinese car ownership grow each year?

(continued)

(c) Make a time plot for trend residuals. Does the plot appear to show autocorrelation? What is the P-value for the Runs Test applied to the trend residuals? What is the conclusion from the test, using 5% significance?

(d) Fit an AR(p) model to trend residuals. What is the value of p? Write down the prediction equation for the residuals.

(e) Find the (improved) simple forecast F_{2007} for Chinese car ownership in 2007.

(f) Find a 95% forecast for Chinese car ownership in 2010.

(g) Find a 95% forecast for Chinese car ownership in 2013. Then interpret the answer.

(end of assignment)