THE UNIVERSITY OF TEXAS AT AUSTIN
McCombs School of Business

STA 372.5 **Spring 2019**

HOMEWORK #9 – Due Wednesday, April 24

1. The data shown on the next page are annual sales data (in millions of dollars) for a company over a 20-year period. The data are annual sales so there is no seasonality.

   Letting Y_t represent sales in year t, the appropriate model for these data is the ARIMA(1, 1, 0) model:

   $$Y_t = Y_{t\text{-}1} + W_t$$

   where

   $$W_t = \mu + \beta(W_{t\text{-}1} - \mu) + \varepsilon_t \qquad\qquad \varepsilon_t \text{ iid } N(0, \sigma^2).$$

   Using all available information in the R output on the following pages, answer parts (a) and (b).

   (a) What are the forecasts for Y_{21} and Y_{22} that have been replaced with asterisks under *forecast(result_ARIMA, h=2)* in the R output?

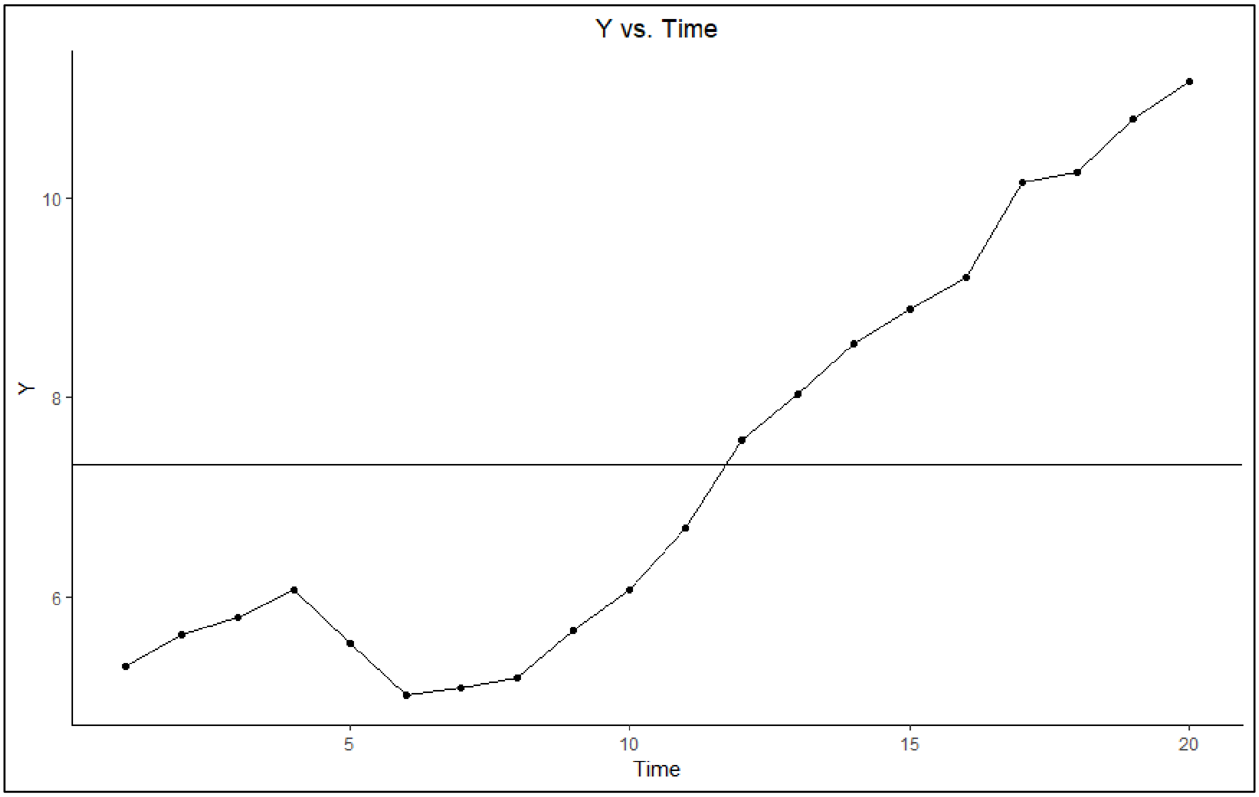
   (b) What are the 95% confidence intervals for Y_{21} and Y_{22} that have been replaced with asterisks under *forecast(result_ARIMA, h=2)* in the R output?

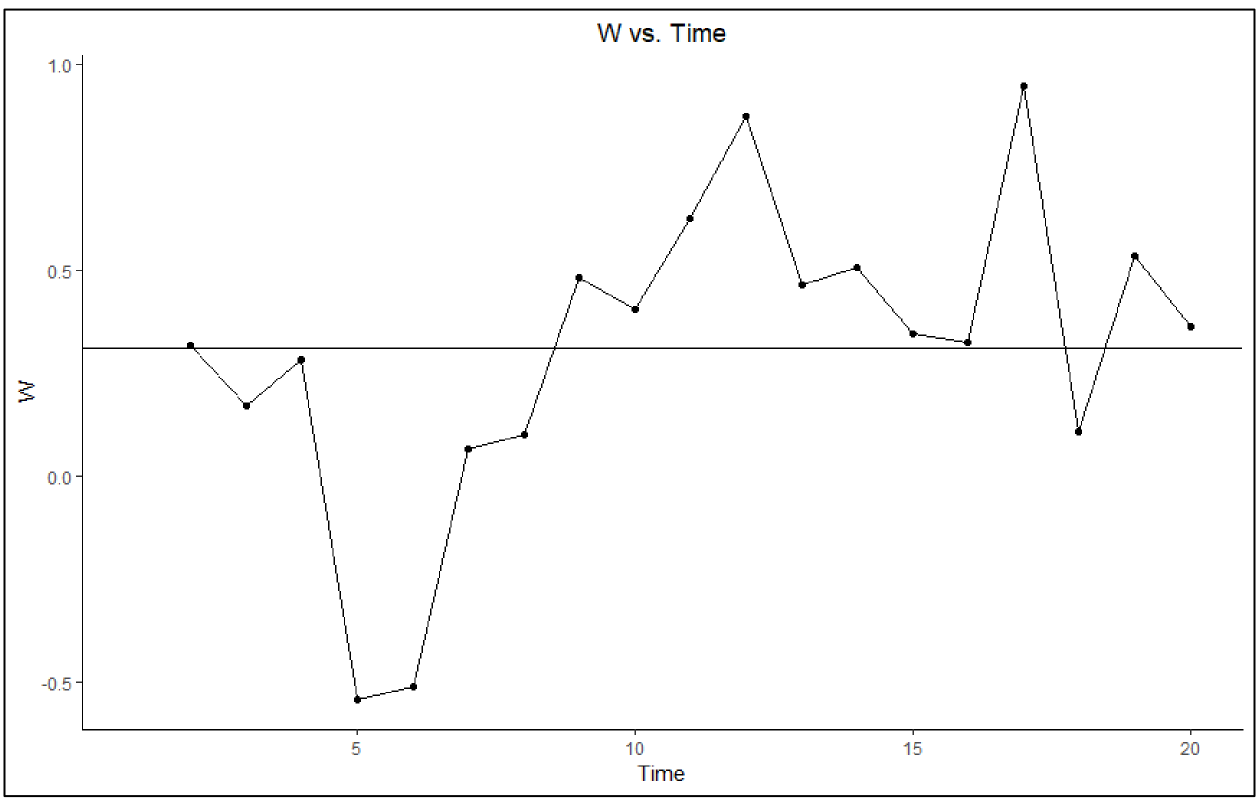
```
head(data_table)
tail(data_table)
```

```
  time        Y          W
1    1 5.304990         NA
2    2 5.621624  0.3166335
3    3 5.790759  0.1691358
4    4 6.072214  0.2814548
5    5 5.529907 -0.5423071
6    6 5.019658 -0.5102490

   time         Y         W
15   15  8.883935 0.3454396
16   16  9.206596 0.3226613
17   17 10.151273 0.9446764
18   18 10.259602 0.1083293
19   19 10.793419 0.5338172
20   20 11.156535 0.3631154
```

Plots of *Y* vs. *Time* and *W* vs. *Time*





```
------------------------------------------------------------------------------------------------------------------------
#
#   Forecast using Arima command
#
Y_time_series <- ts(Y)
result_ARIMA <- Arima(Y_time_series, order=c(1,1,0), include.constant=TRUE)
result_ARIMA
------------------------------------------------------------------------------------------------------------------------
```

```
Series: Y_time_series
ARIMA(1,1,0) with drift

Coefficients:
         ar1   drift
      0.4265  0.3103
s.e.  0.1981  0.1260

sigma^2 estimated as 0.1195:  log likelihood=-5.82
AIC=17.64   AICc=19.24   BIC=20.48
```

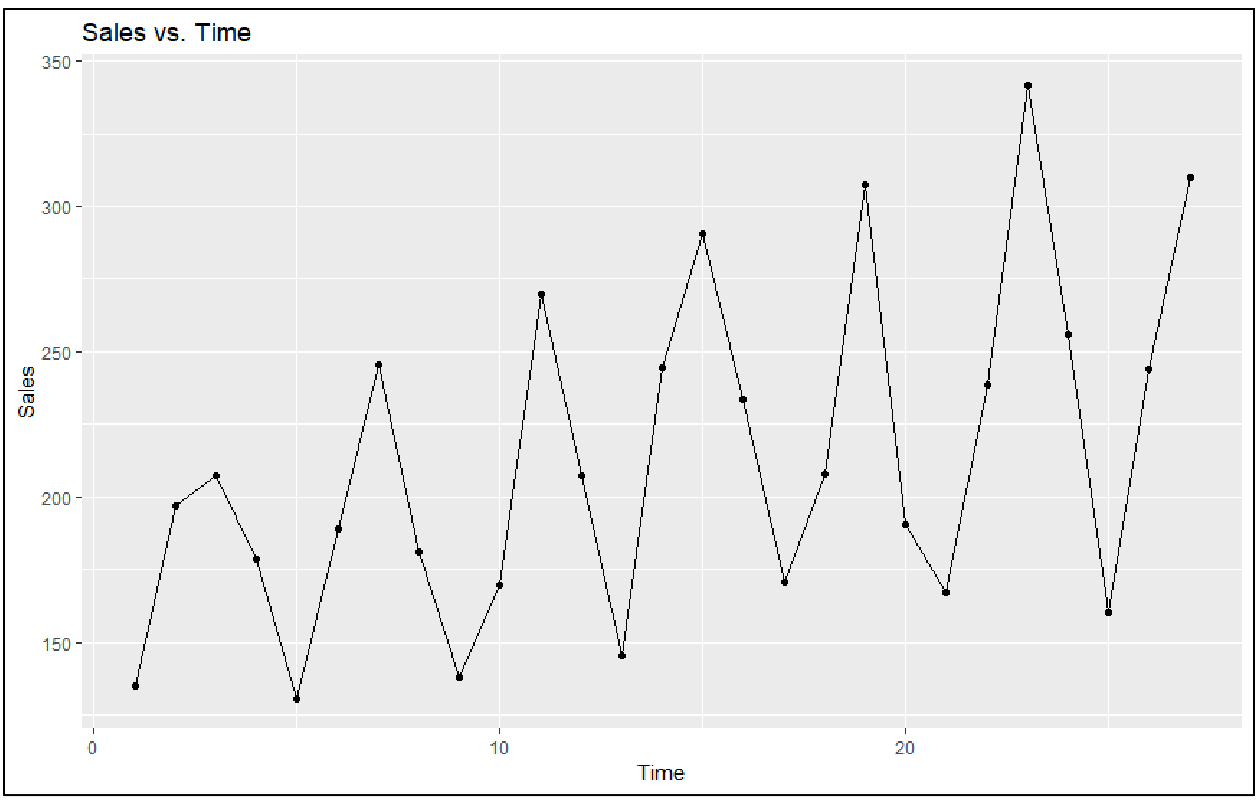
```
------------------------------------------------------------------------------------------------------------------------
forecast(result_ARIMA, h=2)
------------------------------------------------------------------------------------------------------------------------
```

```
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
21       ******** ******** ******** ******** ********
22       ******** ******** ******** ******** ********
```

2. Briggs and Stratton is the world's largest manufacturer of small, air-cooled engines for outdoor power equipment; the company also manufactures locks for use in automobiles and trucks. In the last fiscal year, approximately 93% of company sales (and about 96% of profits) were generated by engines and parts, and about 7% of sales (and about 4% of profits) came from the company's lock operations. Also, in the same year about 86% of the company's engines were used in walk-type power mowers, riding-type lawn mowers, garden tractors and tillers, snow-throwers, etc. Export sales accounted for about 20% of total sales. The company manufacturers and services automotive locks and related products for a variety of North American car and truck manufacturers.

   Your assignment is to develop forecasts of the total sales of Briggs and Stratton Corporation for four quarters beyond the last sales observation available.

   Sales data, in millions of dollars, are available for the past 27 quarters and are plotted below. A strong seasonal pattern is present in sales as well as an upward trend. The data are available in the file STA372_Homework9_Question2.dat on the *Data sets* page of the Canvas class website. The file contains *Time* in the first column, *Quarter* in the second column and *Sales* in the third column.



(a) Why is it important to analyze log(*Sales*) rather than *Sales*?

(b) Use the *stl* command to seasonally adjust log(*Sales*) and store the results in *log_A*.

For parts (c) – (h), use the first 23 observations for estimation and the last four observations for out-of-sample validation.

(c) Use the *ets* command to select the best exponential smoothing model for *log_A*. Be sure to allow for additive and multiplicative errors, and additive damped and non-damped trends. It is not necessary to allow for a multiplicative trend. Store the results in the object *result_ets*. Are the errors from the best model independent? Use the *forecast* command to forecast *log_A* in quarters 24-27.

(d) Use the commands

```
accuracy_result_ets <- result_ets %>% forecast(h=4) %>% accuracy(log_A_time_series)
accuracy_result_ets[,c("RMSE", "MAE")]
```

to compute the out-of-sample RMSE and MAE for the four forecasts of *log_A*.

(e) Use the *Arima* command to estimate an ARIMA(1, 1, 0) model for *log_A*. Do not include a drift term (which is the default, or equivalently, set *include.constant=FALSE*). Are the residuals from this model independent? Use the *forecast* command to forecast *log_A* in quarters 24-27.

(f) Use the *accuracy* command to compute the out-of-sample RMSE and MAE for the four forecasts of *log_A*.

(g) Which model (exponential smoothing or ARIMA(1, 1, 0)) should be used to forecast *log_A* in quarters 28-31? Why?

(h) Using the appropriate model and all 27 observations, compute a forecast and 95% confidence interval for *Sales* in quarter 28.

In practice, you would want to compute forecasts and confidence intervals for *Sales* in quarters 29, 30 and 31 as well. However, these are tedious calculations to do by hand. We will discuss in class in the near future a command to automate the calculation of the forecasts and confidence intervals for *Sales*.