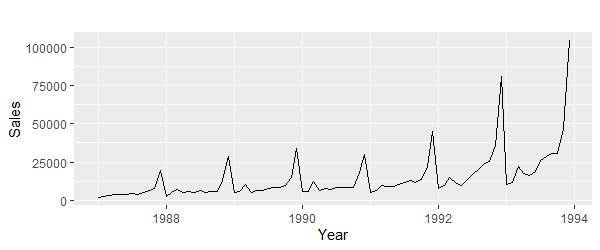
**Chapter 5.5 Exercise**

The data set fancy concerns the monthly sales figures of a shop which opened in January 1987 and sells gifts, souvenirs, and novelties. The shop is situated on the wharf at a beach resort town in Queensland, Australia. The sales volume varies with the seasonal population of tourists. There is a large influx of visitors to the town at Christmas and for the local surfing festival, held every March since 1988. Over time, the shop has expanded its premises, range of products, and staff.

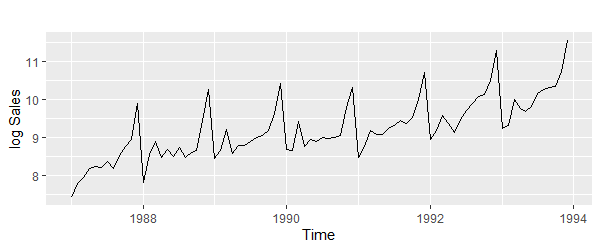
1. Produce a time plot of the data and describe the patterns in the graph. Identify any unusual or unexpected fluctuations in the time series.

This data has an upward season trend because the pattern repeats. There are two spikes in the data: one during Christmas and one during March due to the surfing festival.

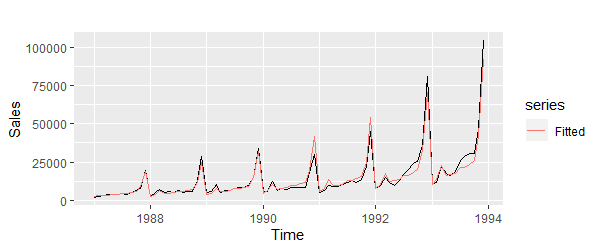
1. Explain why it is necessary to take logarithms of these data before fitting a model.

It is important to take logarithms of data before fitting for several reasons:

1. They constrain the forecasts to stay positive
2. The changes in log value are relative changes on the original scale
3. Logarithms make the data pattern more linear and more constant compared to the original data

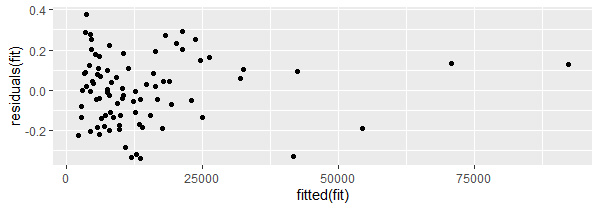


1. Use R to fit a regression model to the logarithms of these sales data with a linear trend, seasonal dummies and a “surfing festival” dummy variable.



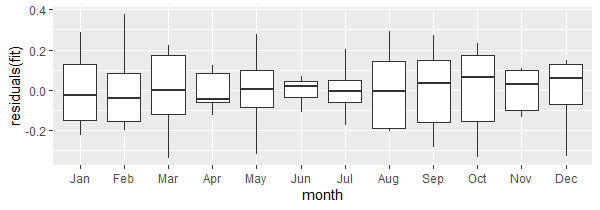
1. Plot the residuals against time and against the fitted values. Do these plots reveal any problems with the model?





The residuals appear to be random, are not serially correlated and are nonlinear

1. Do boxplots of the residuals for each month. Does this reveal any problems with the model?

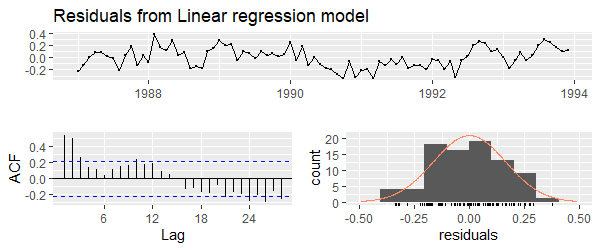


The boxplot does not reveal any problems with the model.

1. What do the values of the coefficients tell you about each variable?

* Trend – log sales increase every year by 0.022 on average
* SeasonXX – show an increase in sales compared to January of the same year
  + Season2 – 0.25
  + Season3 – 0.27
  + Season4 – 0.38
  + Season5 – 0.41
  + Season6 – 0.45
  + Season7 – 0.61
  + Season8 – 0.59
  + Season9 – 0.67
  + Season10 – 0.75
  + Season11 – 1.21
  + Season12 – 1.96
* festivalTRUE – shows that sales increase an average of 0.5 compared to non-festival months

1. What does the Breusch-Godfrey test tell you about your model?



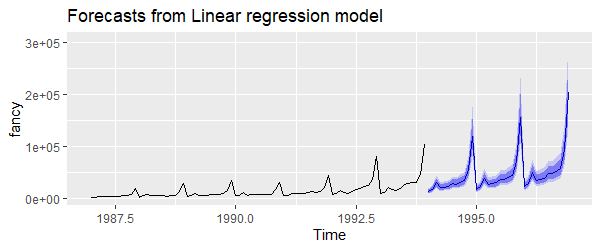
Breusch-Godfrey test for serial correlation of order up to 17

data: Residuals from Linear regression model

LM test = 37.954, df = 17, p-value = 0.002494

According to the p-value obtained from the Breusch-Godfrey test these residual finding are significant

1. Regardless of your answers to the above questions, use your regression model to predict the monthly sales for 1994, 1995, and 1996. Produce prediction intervals for each of your forecasts.



1. Transform your predictions and intervals to obtain predictions and intervals for the raw data.

Since we passed lambda=0 inside the function tslm the data was fit directly to the raw time series data

1. How could you improve these predictions by modifying the model?

We did have to transform the model with a log function so that it would be more linear, so taking that into account would improve the model