3.7 Exercises

1. For the following series, find an appropriate Box-Cox transformation in order to stabilise the variance.
   - `usnetelec`
   - `usgdp`
   - `mcopper`
   - `enplanements`
2. Why is a Box-Cox transformation unhelpful for the `cangas` data?
3. What Box-Cox transformation would you select for your retail data (from Exercise 3 in Section 2.10)?
4. For each of the following series, make a graph of the data. If transforming seems appropriate, do so and describe the effect. `dole`, `usdeaths`, `bricksq`.
5. Calculate the residuals from a seasonal naïve forecast applied to the quarterly Australian beer production data from 1992. The following code will help.

   ```
   beer <- window(ausbeer, start=1992)
   fc <- snaive(beer)
   autoplot(fc)
   res <- residuals(fc)
   autoplot(res)
   ```

   Test if the residuals are white noise and normally distributed.

   ```
   checkresiduals(fc)
   ```

   What do you conclude?
6. Repeat the exercise for the `WWWusage` and `bricksq` data. Use whichever of `naive()` or `snaive()` is more appropriate in each case.
7. Are the following statements true or false? Explain your answer.

   a. Good forecast methods should have normally distributed residuals.

b. A model with small residuals will give good forecasts.
c. The best measure of forecast accuracy is MAPE.
d. If your model doesn't forecast well, you should make it more complicated.
e. Always choose the model with the best forecast accuracy as measured on the test set.

8. For your retail time series (from Exercise 3 in Section 2.10):

a. Split the data into two parts using

```
myts.train <- window(myts, end=c(2010,12))
myts.test <- window(myts, start=2011)
```

b. Check that your data have been split appropriately by producing the following plot.

```
autoplot(myts) +
  autolayer(myts.train, series="Training") +
  autolayer(myts.test, series="Test")
```

c. Calculate forecasts using `snaive` applied to `myts.train`.

```
fc <- snaive(myts.train)
```

d. Compare the accuracy of your forecasts against the actual values stored in `myts.test`.

```
accuracy(fc,myts.test)
```

e. Check the residuals.

```
checkresiduals(fc)
```

Do the residuals appear to be uncorrelated and normally distributed?

f. How sensitive are the accuracy measures to the training/test split?

9. `visnights` contains quarterly visitor nights (in millions) from 1998 to 2016 for twenty regions of Australia.

a. Use `window()` to create three training sets for `visnights[,"QLDMetro"]`, omitting the last 1, 2 and 3 years; call these train1, train2, and train3, respectively. For example `train1 <- window(visnights[, "QLDMetro"], end =`

`c(2015, 4))`.

b. Compute one year of forecasts for each training set using the `snaive()` method. Call these `fc1`, `fc2` and `fc3`, respectively.

c. Use `accuracy()` to compare the MAPE over the three test sets. Comment on these.

10. Use the Dow Jones index (data set `dowjones`) to do the following:

a. Produce a time plot of the series.
b. Produce forecasts using the drift method and plot them.
c. Show that the forecasts are identical to extending the line drawn between the first and last observations.
d. Try using some of the other benchmark functions to forecast the same data set. Which do you think is best? Why?

11. Consider the daily closing IBM stock prices (data set `ibmclose`).

a. Produce some plots of the data in order to become familiar with it.
b. Split the data into a training set of 300 observations and a test set of 69 observations.
c. Try using various benchmark methods to forecast the training set and compare the results on the test set. Which method did best?
d. Check the residuals of your preferred method. Do they resemble white noise?

12. Consider the sales of new one-family houses in the USA, Jan 1973 – Nov 1995 (data set `hsales`).

a. Produce some plots of the data in order to become familiar with it.
b. Split the `hsales` data set into a training set and a test set, where the test set is the last two years of data.
c. Try using various benchmark methods to forecast the training set and compare the results on the test set. Which method did best?
d. Check the residuals of your preferred method. Do they resemble white noise?