Hyndman and Athanasopoulos – Chapter 5 Answers

**Question 1**

1.a – The fitted model estimates that if daily temperature increases by one degree, we expect an average predicted increase in electricity consumption of 6.7 GW.

1.b – Yes. The results for the Breusch-Godfrey test suggest that the residuals are not autocorrelated. The correlogram also shows no significant correlations at different lags. Observation 5 could be a potential outlier, since it is a very large negative residual, more than 2 standard deviations away from the mean.

1.c – If the temperature is 15º, the model forecasts a demand of 140.6 GW. If the temperature is 35º, the model forecasts a demand of 275.7 GW

1.d – The 80% confidence interval for a temperature of 15ºC is [108.7, 172.5];

The 95% confidence interval for a temperature of 15ºC is [90.2,190.9];

The 80% confidence interval for a temperature of 15ºC is [245.2,306.2];

The 95% confidence interval for a temperature of 15ºC is [227.6,323.9];

1.e – The linear model is inadequate, since the relationship is nonlinear. It overpredicts demand on mild temperature days and underestimates the demand on very warm days.

**Question 2**

2.a – The winning time has been decreasing consistently, between 1896 and 2016. However, it appears that the relationship is not linear.

2.b – For a new time interval, we expect an average decrease of 0.25 seconds in the winning time.

2.c – The residuals seem to be heteroskedasticity and the residual for the first observation appears to be an outlier.

2.d – The model predicts a winning time of 42 seconds, with a 95% confidence interval of [39.6,44.5]. We are assuming that the model is a good approximation of reality, and we make some assumptions about the residuals:

- They have mean 0

- They are not autocorrelated

- They are unrelated to the predictor variables.

**Question 3**

3.a – It returns fractional values if Easter spans quarters 1 and 2. Or it is a dummy variable for quarters 1 or 2, depending on the year.

**Question 5**

5.a – It appears that the data peaks in December for every year. There is also an uptick in March. The data also seems to possess a linear upward trend. In 1990, the rise in December is smaller than it was in 1989, which is an unexpected pattern in the series.

5.b – Taking the natural logarithms is a wise modeling choice because the variation of the series is not constant. Taking the logs should make the seasonal variation about the same across the whole series. This should make the forecasting model simpler.

5.c – Fitted

5.d – Yes, the plot of the residuals against time suggests that the residuals present autocorrelation. The plot of residuals against fitted values seems to show no relationship between a linear combination of the predictors and the residuals. Also, the distribution of the residuals shows no sign of heteroskedasticity. The latter plot also suggests that a linear functional form is appropriate for the task at hand.

5.e – It appears that the variance of the residuals is not uniform across the months.

5.f – There is a positive linear trend; we expect that the surfing festival will have a positive impact on sales; dummies for every month are positive and statistically significant with the exception of the dummy for March. This could be a result of the inclusion of the surfing festival variable.

5.g – The results present evidence against the null hypothesis that there is no autocorrelation in the residuals. Therefore, this gives a numerical support to the conclusions drawn from question d.

**Question 6**

6.a – The more terms added to the fourier series, the closer it matches the original series. One possible drawback is that adding more terms might cause the model to overfit.

6.b – Using AICc, the best model is the one with 12 terms. The same model was found by cross-validation.

6.e – The actual data lies mostly within the prediction interval, but the forecasts for the end of the year actually overestimate the gasoline supply.