

Solutions to Exam—style question (4)

- (a) The time series plot also shows seasonality, with visitor numbers peaking in August and dipping in July.
- (b) Although we have *monthly* data here, there are only four seasons per year (July, August, September and October) since the water park is not open in any other month. Thus, we use the formula for *quarterly* moving averages, and so

$$y_3^* = \frac{262 + 2(901 + 841 + 389) + 402}{8} = 615.75$$

and

$$y_4^* = \frac{901 + 2(841 + 389 + 402) + 1035}{8} = 650.$$

- (c) Now there are 39 observations in the original dataset; however, $n = 35$ since we are modelling the *moving averages*, and we will not have any moving averages for the first and last *two* time-points. This will give

$$S_{TY^*} = \sum ty^* - n\bar{t}\bar{y}^* = 966734 - 35 \times \frac{700}{35} \times \frac{42121.6}{35} = 124302$$

and

$$S_{TT} = \sum t^2 - n\bar{t}^2 = 17570 - 35 \times \left(\frac{700}{35}\right)^2 = 3570.$$

So we get

$$\hat{\beta}_1 = \frac{S_{TY^*}}{S_{TT}} = \frac{124302}{3570} = 34.81849$$

and

$$\hat{\beta}_0 = \bar{y}^* = \hat{\beta}_1 \bar{t} = \frac{42121.6}{35} - 34.81849 \times \frac{700}{35} = 507.1045.$$

This gives the following trend line for the series of moving averages:

$$Y^* = 507.1045 + 34.81849T + \epsilon.$$

- (d) The *seasonal indices*, also known as *seasonal effects*, are the average deviations from the trend line for each season. We use the seasonal effects to adjust a forecast given by the trend line alone to account for the average effect of the season for which we are making a forecast. For example, if our variable is usually below average in a particular season,

the seasonal effect for that season is likely to be negative and so we adjust the value given by the trend line downwards; conversely, if our variable is usually *above* average in a particular season, the seasonal effect will be positive and so we adjust the value given by the trend line upwards.

- (e) The partial autocorrelation function indicates that there is significant lag 1 autocorrelation, but nothing more long-term than this (r_1 , which is approximately 0.6, is the only partial autocorrelation above the significance bands). This would suggest an autoregressive model of order 1 might be appropriate (i.e. an $AR(1)$ model), and this is what has been fitted in **Minitab** in the output to the right-hand-side.

item [(f)]

- (i) The full estimated model is

$$\begin{aligned} Y_t &= \mu + \alpha_1(Y_{t-1} - \mu) + \epsilon_t \\ &= 0.1228 + 0.5820(Y_{t-1} - 0.1228) + \epsilon_t. \end{aligned}$$

- (ii) August 2010 will be time-point 42 (think about it – look at the original time series plot!), and so the trend would be

$$Y^* = 507.1045 + 34.81849 \times 42 = 1969.481.$$

The seasonal effect for August is +317.734, giving a *full forecast* of

$$0.13407 + 1969.481 + 317.734 = 2287.349,$$

i.e. about 2287 visitors.

- (iii) The 95% CI for this forecast is given by:

$$(-5.38638 + 1969.481 + 317.349, 5.65454 + 1969.481 + 317.349)$$

i.e.

$$(2281, 2292) \text{ visitors.}$$