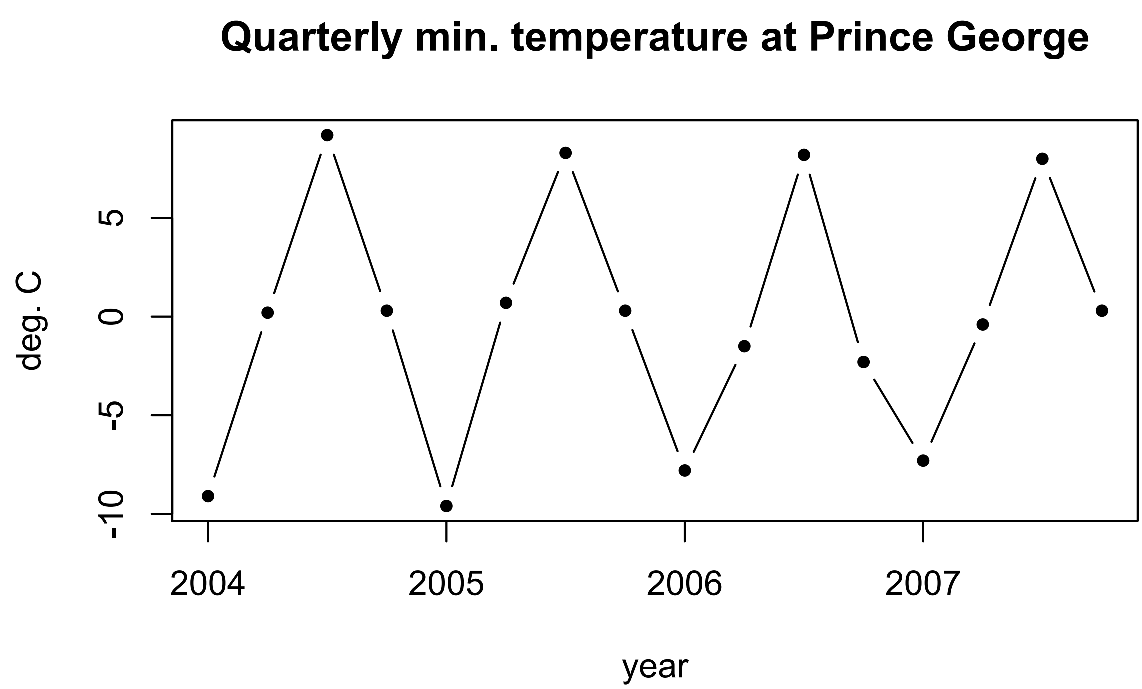
Activity solution: The Sample Autocorrelation

Recall the data on the daily minimum temperatures at Prince George, BC. Consider the quarterly means for the period 2004–2007. The figures are given below, along with the plot.

		Quarter			
		1	2	3	4
year	2004	-9.1	0.2	9.2	0.3
	2005	-9.6	0.7	8.3	0.3
	2006	-7.8	-1.5	8.2	-2.3
	2007	-7.3	-0.4	8.0	0.3
		-33.8	-1	33.7	-1.4



The sample mean of the series is $\overline{x} = -0.2$ and the standard deviation is $s = 6.2$ (in deg. C, both to 1 d.p. which is sufficient accuracy for here).

Recall:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

1. Without performing the calculation, provide a guess of the sample autocorrelation function (the sample acf, r_h) at lag 4. That is, have a guess at r_4.

   Looking at the scatter plot one would expect quite a high, positive correlation between values four time units apart. So a guess of $r_4 \approx 0.7$ would be reasonable. To verify, plot x_{t+4} against x_t for the data.

2. Using the formula, find r_4 from the data.
   Recall:
   $$r_h := \frac{\sum_{t=1}^{n-h}(x_t - \overline{x})(x_{t+h} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2} \qquad (h = 1, 2, \ldots, n-1)$$

   From the data, we should find
   $$\begin{aligned} r_4 &= \frac{\sum_{t=1}^{12}(x_t - \bar{x})(x_{t+4} - \bar{x})}{\sum_{t=1}^{16}(x_t - \bar{x})^2} \\ &= \frac{\sum_{t=1}^{12}(x_t + 0.2)(x_{t+4} + 0.2)}{(16-1) \times 6.2^2} \\ &= 0.73. \end{aligned}$$

3. Without performing the calculation, provide a guess of the sample autocorrelation function (the sample acf, r_h) at lag 2. That is, have a guess at r_2.

   Values two time points apart tend to lie on opposite sides of the mean, so we would expect the sign to be opposite to that of r_4 but the magnitude to be similar. A guess might be $r_2 \approx -0.7$.

4. Using the formula, find r_2 from the data.

   From the data, we should find
   $$\begin{aligned} r_2 &= \frac{\sum_{t=1}^{14}(x_t - \bar{x})(x_{t+2} - \bar{x})}{\sum_{t=1}^{16}(x_t - \bar{x})^2} \\ &= \frac{\sum_{t=1}^{14}(x_t + 0.2)(x_{t+2} + 0.2)}{(16-1) \times 6.2^2} \\ &= -0.85. \end{aligned}$$

5. Without performing the calculation, provide a guess of the sample autocorrelation function (the sample acf, r_h) at lag 1. That is, have a guess at r_1.

Values one time point apart tend to be on opposite sides of the mean, but about equally often x_{t+1} exceeds x_t as vice versa. So we might guess $r_1 \approx 0$.

6. Using the formula, find r_1 from the data.
   From the data, we should find

$$\begin{aligned} r_1 &= \frac{\sum_{t=1}^{15} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{16} (x_t - \bar{x})^2} \\ &= \frac{\sum_{t=1}^{15} (x_t + 0.2)(x_{t+1} + 0.2)}{(16-1) \times 6.2^2} \\ &= 0.00. \end{aligned}$$

7. A plot of r_h against the lag h is known as the *correlogram.* Sketch what you think the correlogram would look like here for lags between $h = 0$ and $h = 8$.