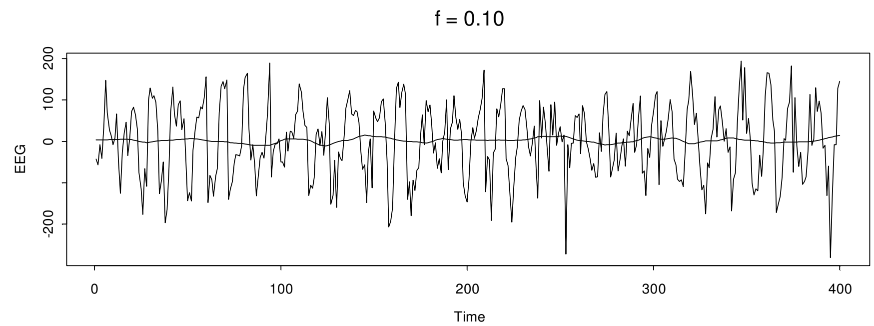
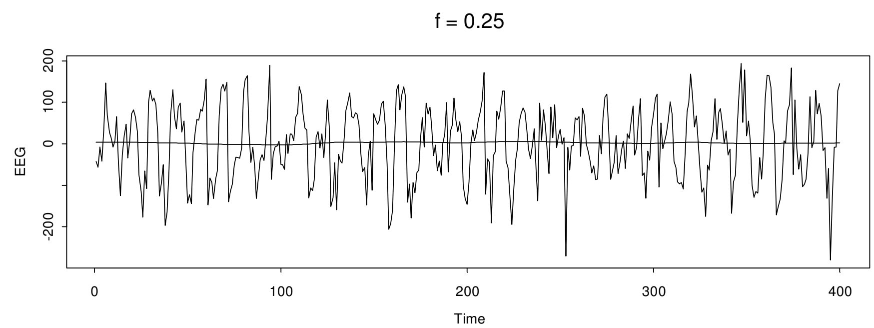
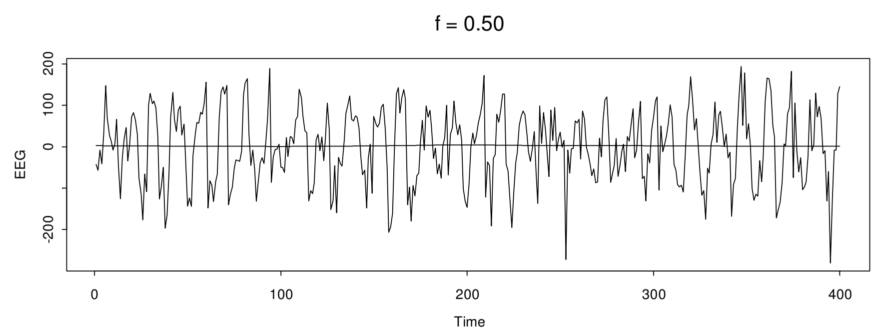
Time Series Homework #2 Solutions

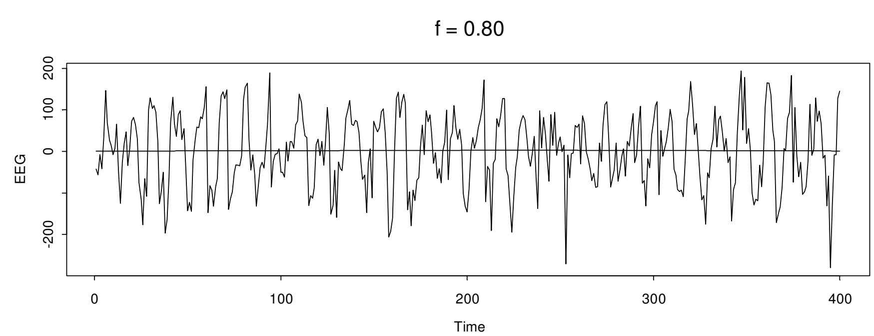
1. a. (3 pts)

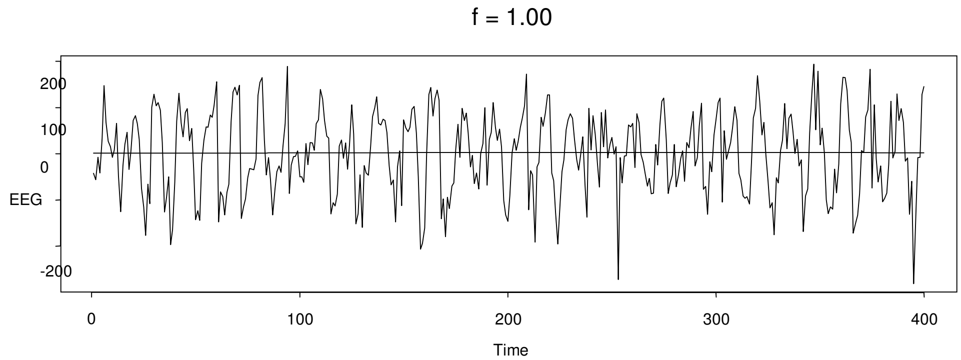
Below are the plots of the lowess estimates of the first 400 observations of the EEG data using f = 0.10, 0.25, 0.50, 0.80, and 1.00, respectively, to smooth. The smoothing parameter does not seem to have any impact on the trend, as there is not one.









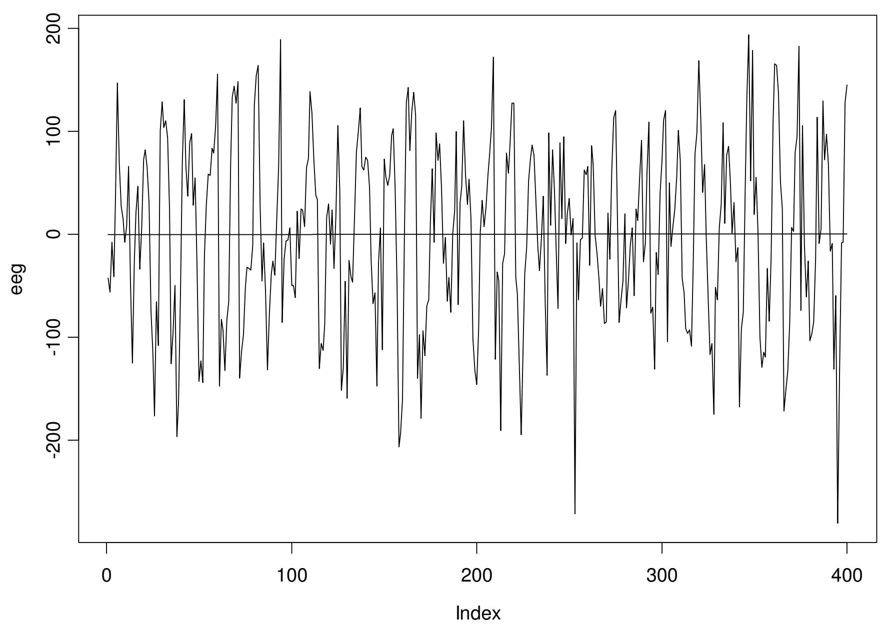


b. (3 pts)

The regression line is

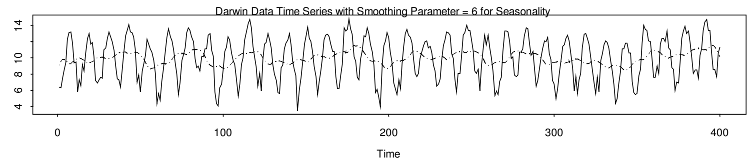
$$EEG = -0.2968 + 0.0015 * Time.$$

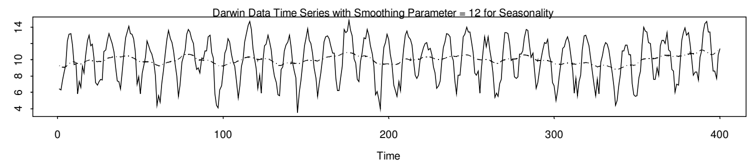
The slope is negligible and therefore, the regression line adequately represents the trend (as previously stated, there does not appear to be a trend).

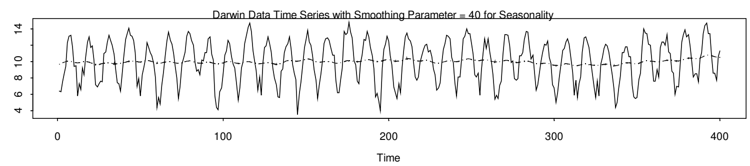


2. a. (2 pts)

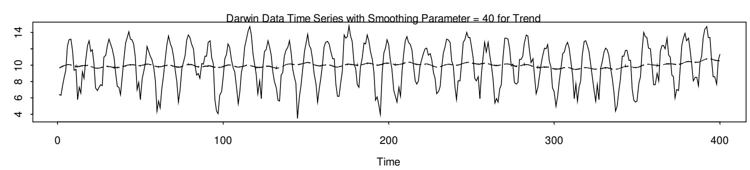
For seasonality, the smoothing parameters chosen were q = 6, 12, 40. The smoothing parameter of 12 is the best as the season is best represented by a yearly cycle. When the smoothing parameter is 6, the average is rather sporadic. When the smoothing parameter is 12, the cycle looks stable. When the smoothing parameter is 40, the data is smoothed too much.

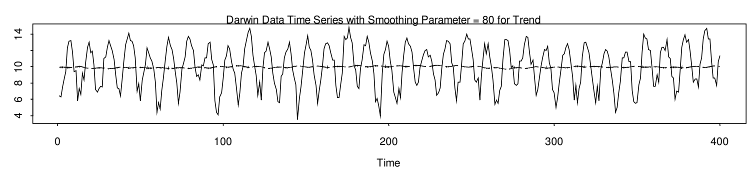


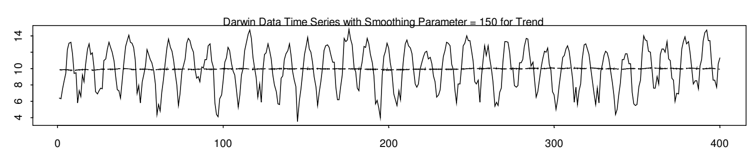




For trend, the smoothing parameters chosen were q = 40, 80, 150. The smoothing parameter of 150 is the best as a straight line best represents the trend. When the smoothing parameter is 40, the average is tries to follow the data too much. When the smoothing parameter is 80, the average still follows the data, but not as much. When the smoothing parameter is 150, the average is practically a straight line, representing the trend.



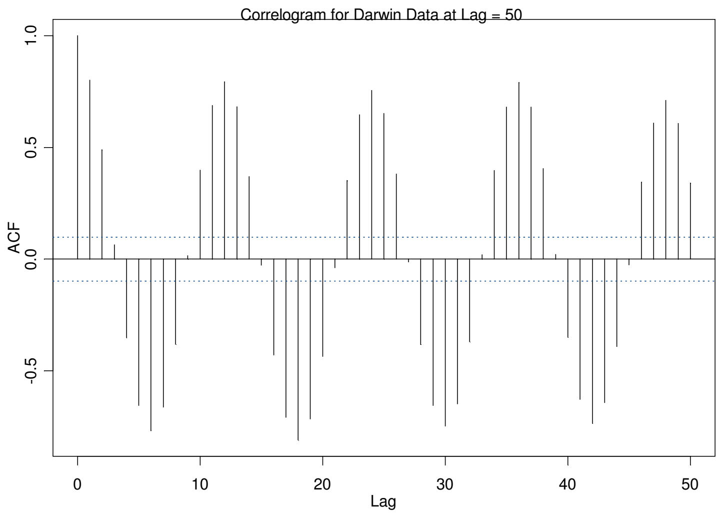




b. (3 pts)

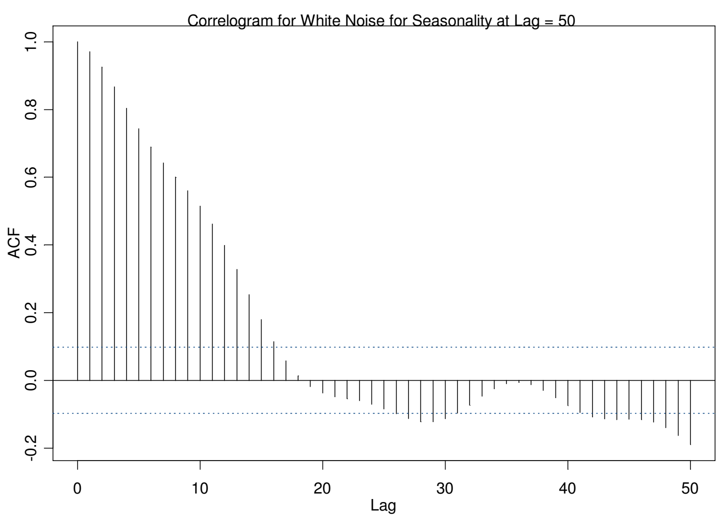
The correlogram for the Darwin data is shown below. This plot clearly shows the seasonality of the data. The period is 12, as shown in this plot.

Series : darwin

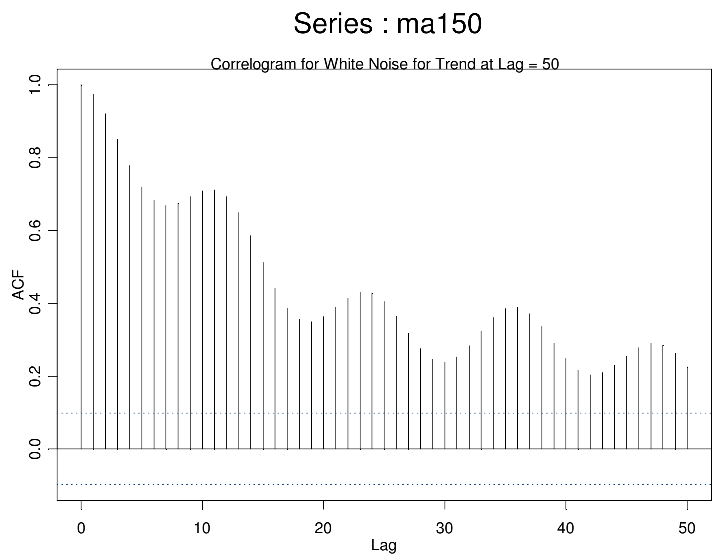


When the moving average for seasonality is plotted, we see the correlogram shown below. It shows that the seasonality has been removed (for the most part) from the data. The plot does not show a significant periodicity like the above graph shows.

Series : ma12



When the moving average for trend is plotted, we see the correlogram shown below. It shows that when the trend is removed (or attempted to be removed) using a moving average, the autocorrelations are all significant and do not decrease as quickly as the above two plots. There appears to still be a seasonality of approximately 12.

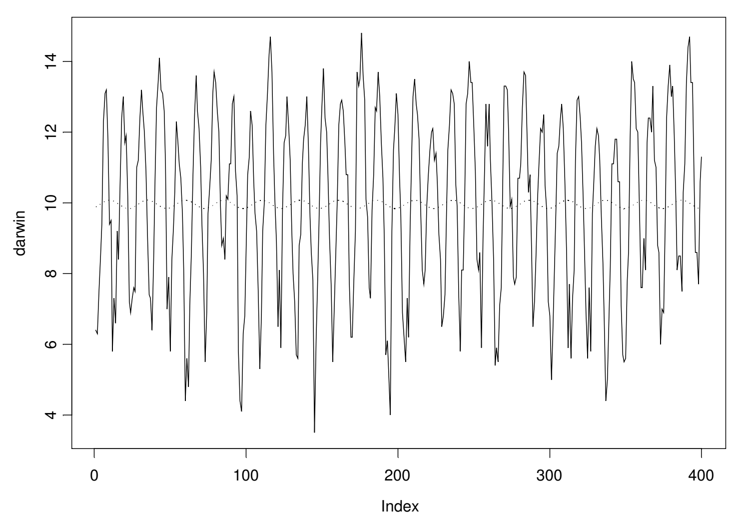


c. (3 pts)

The first plot has the regression equation

$$Pressure = 9.956459 + 0.07241695 * \sin(0.25 * Time) - 0.09366165 * \cos(0.25 * Time)$$

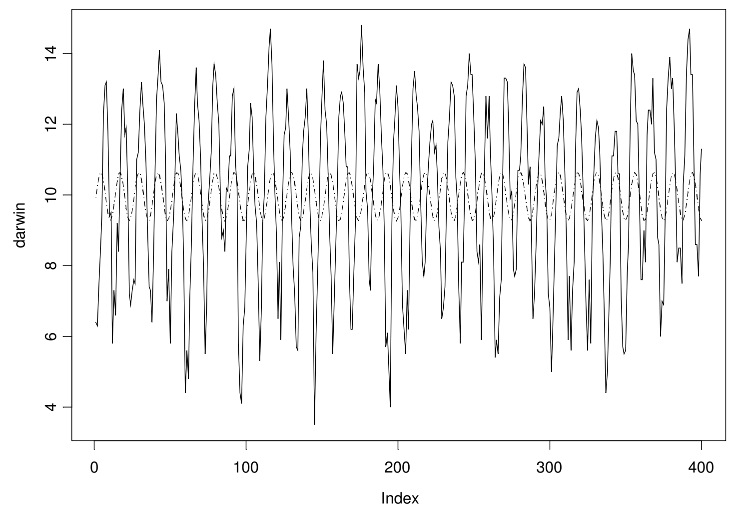
This plot does not represent the seasonality well as it is still only mimicking the data.



The second plot has the regression equation

$$\text{Pressure} = 9.954402 + 0.5837419 * \sin(0.5 * \text{Time}) - 0.3599607 * \cos(0.5 * \text{Time})$$

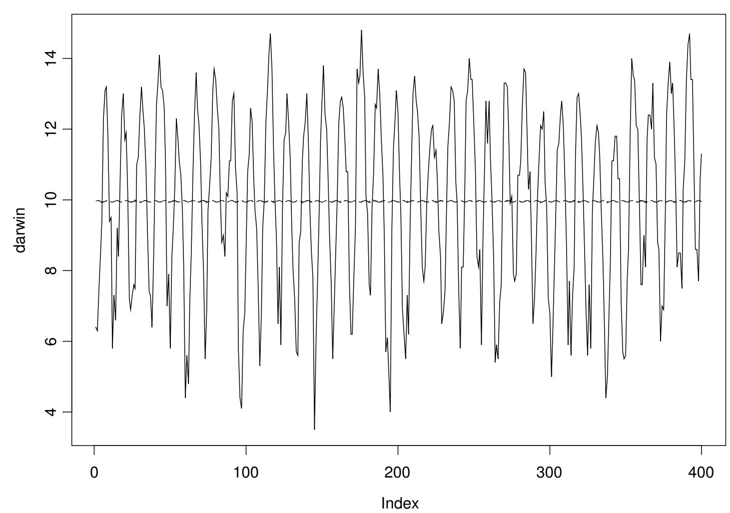
This plot does represent the seasonality well as it has the same number of periods as the number or years and it shows a nice cyclic pattern.



The third plot has the regression equation

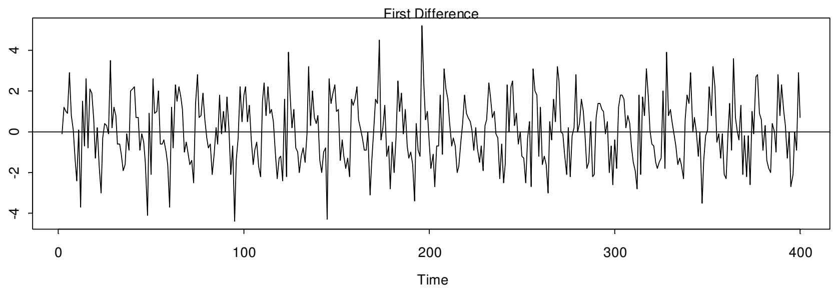
$$\text{Pressure} = 9.956895 + 0.01739335 * \sin(\text{Time}) - 0.01634599 * \cos(\text{Time})$$

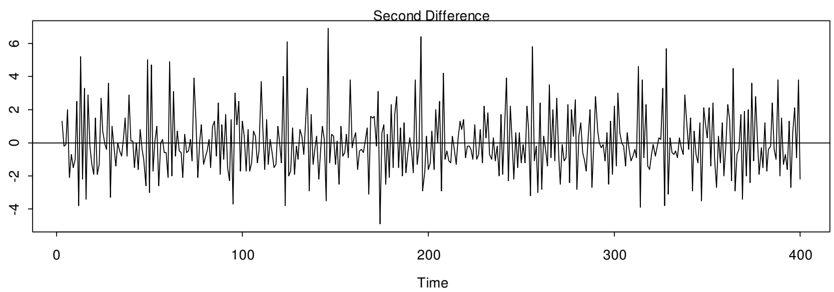
This plot does not represent the seasonality well as smoothes the seasonality too much.



d. (3 pts)

The first difference reduces the noise more than the second difference. The second difference looks like it increases the white noise in the data.





3. (4 pts)

$$x_t = ax_{t-1} + w_t$$
$$iid\ w_t \sim N(0, \sigma^2)$$
$$a \in (-1, 1)$$

$$X_t = aX_{t-1} + w_t$$
$$\Rightarrow X_1 = aX_0 + w_1$$
$$\Rightarrow X_2 = aX_1 + w_2 = a(aX_0 + w_1) + w_2 = a^2 X_0 + aw_1 + w_2$$
$$\Rightarrow X_t = a^k X_{t-k} + \sum_{j=0}^{k-1} a^j w_{t-j}$$
$$As\ k \to \infty,\ X_t = \sum_{j=0}^{\infty} a^j w_{t-j}$$

$$\begin{aligned} E(X_t) &= E\left(\sum_{j=0}^{\infty} a^j w_{t-j}\right) \\ &= \sum_{j=0}^{\infty} a^j E(w_{t-j}) \\ &= 0 \end{aligned}$$

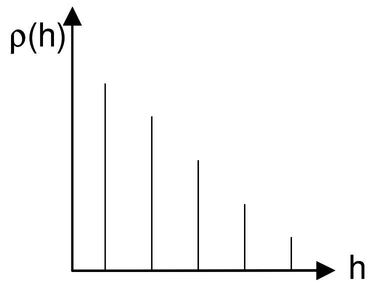
$$\begin{aligned} E(X_t X_{t+h}) &= E\left(\sum_{j=0}^{\infty} a^j w_{t-j} \sum_{k=0}^{\infty} a^k w_{t+h-k}\right) \\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} a^j a^k E(w_{t-j} w_{t+h-k}) \end{aligned}$$

Let $t - j = t + h - k \rightarrow k = h + j$

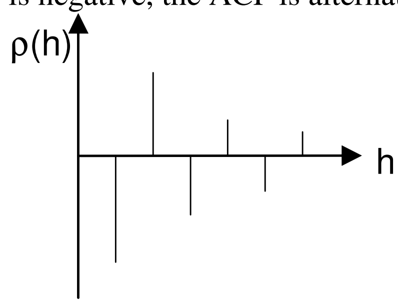
$$\begin{aligned} E(X_t X_{t+h}) &= \begin{cases} \sigma_w^2 \sum_{j=0}^{\infty} a^j a^{j+h} & for\ k = j + h \\ 0 & otherwise \end{cases} \\ &= \begin{cases} \sigma_w^2 a^h \sum_{j=0}^{\infty} a^{2j} & for\ k = j + h \\ 0 & otherwise \end{cases} \\ &= \begin{cases} \dfrac{\sigma_w^2 a^h}{1 - a^2} & for\ k = j + h \\ 0 & otherwise \end{cases} \end{aligned}$$

$$\rho(h) = \frac{\sigma_w^2 a^h / 1-a^2}{\sigma_w^2 / 1-a^2} = a^h \ \ for\ a \in (-1,1)$$

When *a* is positive, the ACF is monotonically decreasing and non-negative.



When *a* is negative, the ACF is alternately positive and negative and tends toward 0.



4. a. (1 pt)

Model: $M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - t) + \beta_3 (T_t - t)^2 + \beta_4 P_t + \beta_5 P_{t-4} + w_t$

Model	R^2	AIC
1. $M_t = \beta_0 + \beta_1 t + w_t$	0.21	5.38
2. $M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - t) + w_t$	0.38	5.14
3. $M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - t) + \beta_3 (T_t - t)^2 + w_t$	0.45	5.03
4. $M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - t) + \beta_3 (T_t - t)^2 + \beta_4 P_t + w_t$	0.60	4.73
5. $M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - t) + \beta_3 (T_t - t)^2 + \beta_4 P_t + \beta_5 P_{t-4} + w_t$	0.61	4.70

F-Statistic, (4) vs (5) ≈20.8 > $F_{1,200}(0.001)$ =11.15.

Adding P_{t-4} to the model increased the R^2 (which is to be suspected when adding parameters to a model), but also decreased the AIC. The F-statistic shows that the model including P_{t-4} makes a better model.

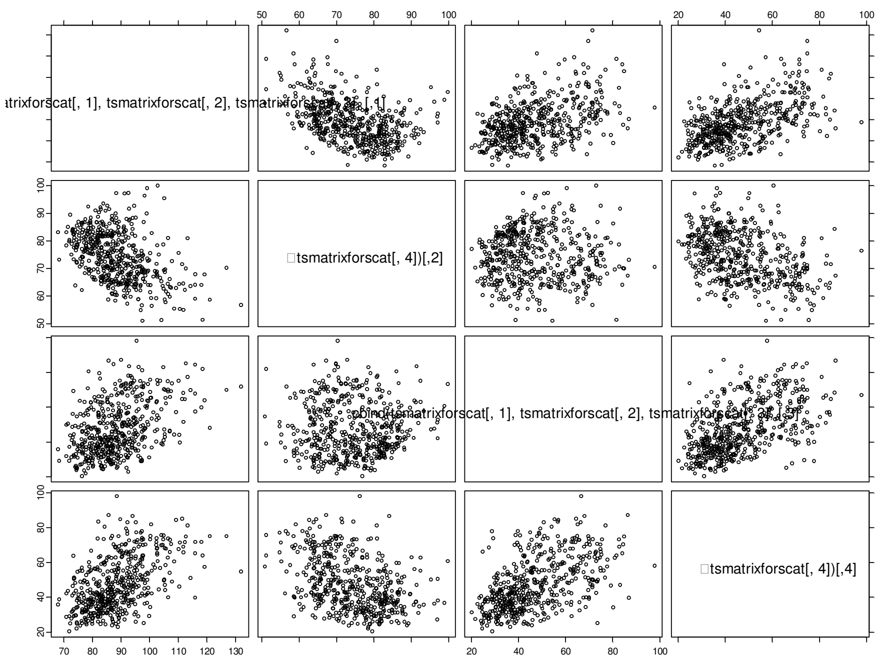
b. (3 pts)

Correlations between the series are shown below.

	M_t	T_t	P_t	P_{t-4}
M_t	1.00000	-0.43696	0.44229	0.52100
T_t		1.00000	-0.01482	-0.39908
P_t			1.00000	0.53405
P_{t-4}				1.00000

The correlation between mortality and the particulate count four weeks prior is slightly stronger than that of mortality and the particulate count. This further strengthens the fact that adding the particulate count four weeks prior was a good idea.

Below is the scatterplot matrix for all pairwise variables.



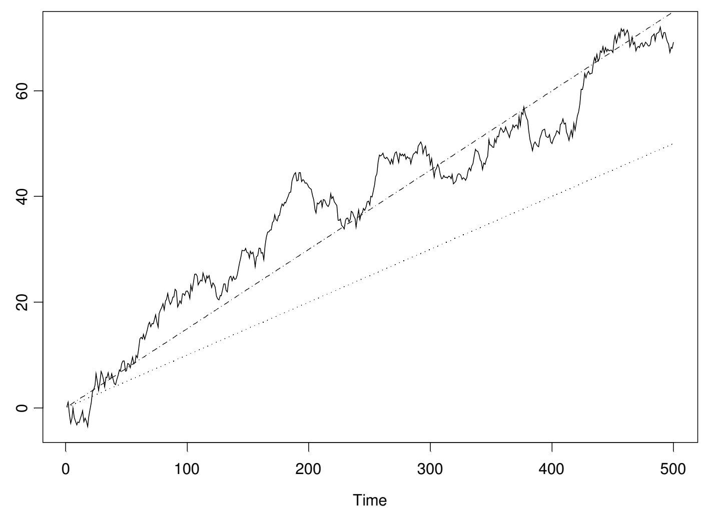
5. (3 pts)

Random Walk Model: n = 500, δ = 0.1, and σ_w = 1.

Regression Line: $\hat{x}_t = 0.1499t$

Mean Function: μ_t = 0.1t

In this case, the regression line fits the random walk data better than the mean function. The regression line will always try to fit the data best regardless of the drift. The mean function does not care what the data looks like and therefore, does not fit the data best in this case. In some cases, it may be hard to tell which line best fits the data, as they could be very similar.



6. a. (2 pts)

Model: $x_t = \beta_0 + \beta_1 t + w_t$ where w_t's iid zero means with variances = σ_w^2 and β_0 and β_1 are fixed constants.

Stationarity: Mean function must not depend on time (t) and covariance function must only depend on lag (h).

$$\begin{aligned} E(x_t) &= E(\beta_0 + \beta_1 t + w_t) \\ &= \beta_0 + \beta_1 t + E(w_t) \\ &= \beta_0 + \beta_1 t \end{aligned}$$

The process is not stationary as the mean function depends on time (t).

b. (2 pts)

Model: $\nabla x_t = x_t - x_{t-1}$

$$\begin{aligned} E(\nabla x_t) &= E(x_t - x_{t-1}) \\ &= E(x_t) - E(x_{t-1}) \\ &= E(\beta_0 + \beta_1 t + w_t) - E(\beta_0 + \beta_1(t-1) + w_{t-1}) \\ &= \beta_0 + \beta_1 t + E(w_t) - \beta_0 - \beta_1(t-1) - E(w_{t-1}) \\ &= \beta_1 \end{aligned}$$

$$\begin{aligned} \gamma(h) &= E(\nabla x_t - \beta_1)(\nabla x_{t+h} - \beta_1) \\ &= E(x_t - x_{t-1} - \beta_1)(x_{t+h} - x_{t+h-1} - \beta_1) \end{aligned}$$

Using the fact that

$$x_t = \beta_0 + \beta_1 t + w_t \Rightarrow w_t = x_t - \beta_0 - \beta_1 t$$

and

$$x_{t-1} = \beta_0 + \beta_1(t-1) + w_{t-1} \Rightarrow w_{t-1} = x_{t-1} - \beta_0 - \beta_1(t-1)$$

we get

$$w_t - w_{t-1} = x_t - x_{t-1} - \beta_1 .$$

$$\begin{aligned} \Rightarrow \gamma(h) &= E(w_t - w_{t-1})(w_{t+h} - w_{t+h-1}) \\ &= E(w_t w_{t+h} - w_t w_{t+h-1} - w_{t-1} w_{t+h} + w_{t-1} w_{t+h-1}) \\ &= \begin{cases} 2\sigma_w^2 & for\ h = 0 \\ -\sigma_w^2 & for\ |h| = 1 \\ 0 & otherwise \end{cases} \end{aligned}$$

$\Rightarrow$ The process is stationary.

c. (2 pts)

Model: $\nabla x_t = x_t - x_{t-1}$ where w_t's are replaced with y_t's iid with mean μ_y and autocovariance= $\gamma_y(h)$.

$$\begin{aligned}
E(\nabla x_t) &= E(x_t - x_{t-1}) \\
&= E(x_t) - E(x_{t-1}) \\
&= E(\beta_0 + \beta_1 t + y_t) - E(\beta_0 + \beta_1(t-1) + y_{t-1}) \\
&= \beta_0 + \beta_1 t + E(y_t) - \beta_0 - \beta_1(t-1) - E(y_{t-1}) \\
&= \beta_1 + \mu_y - \mu_y \\
&= \beta_1
\end{aligned}$$

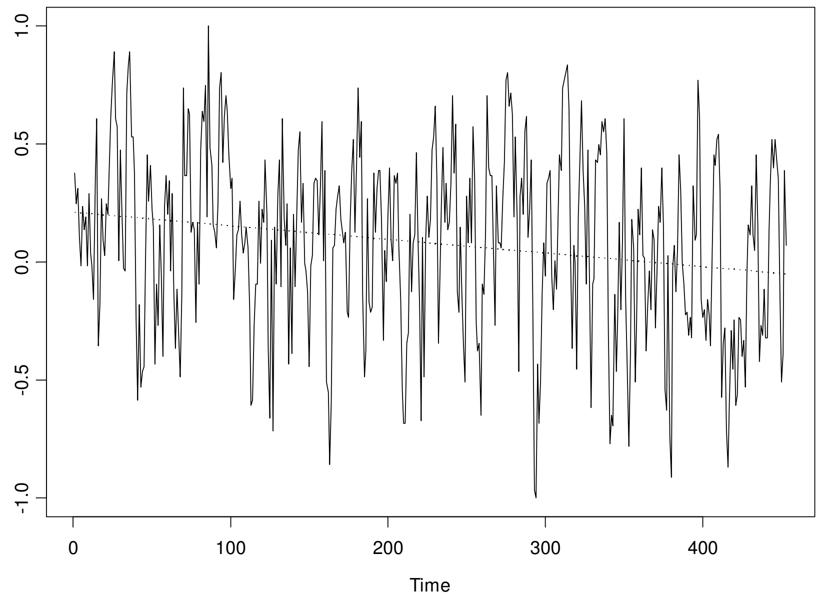
$$\begin{aligned}
\gamma(h) &= E(\nabla x_t - \beta_1)(\nabla x_{t+h} - \beta_1) \\
&= E(x_t - x_{t-1} - \beta_1)(x_{t+h} - x_{t+h-1} - \beta_1) \\
&= E(y_t - y_{t-1})(y_{t+h} - y_{t+h-1}) \\
&= E(y_t y_{t+h} - y_t y_{t+h-1} - y_{t-1} y_{t+h} + y_{t-1} y_{t+h-1}) \\
&= 2\gamma_y(h) - \gamma_y(h-1) - \gamma_y(h+1)
\end{aligned}$$

$\Rightarrow$ The process is stationary.

7. a. (1 pt)

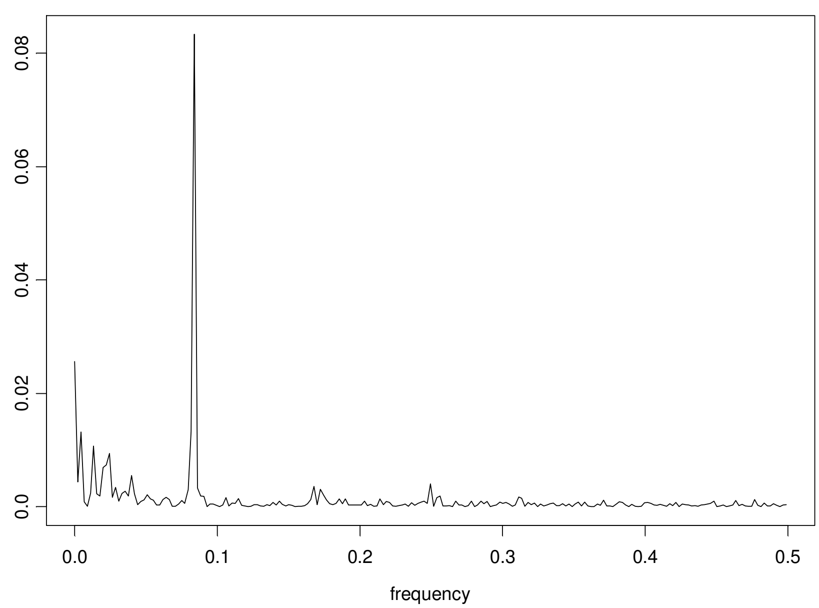
By plotting a regression line over the data and looking at the value of the slope, it is clear that there is a trend in the sea surface temperature. The regression line is

$$\text{Temperature} = 0.2109 - .0006 * \text{Time}.$$



b. (3 pts)

The frequencies of the two main peaks are around 0.083 and near 0. The minor peak does not indicate that there will be an El Nino cycle as $1/0 \cong \infty$. Thus, the major peak is the only one that indicates a cycle (about 12 months).



8. (2 pts)

A 3rd-degree polynomial fits this data very well. The regression model is

Temperature = 203.9177 – 0.2626 * Time + 0.0001 * Time^2 + 0.0000 * Time ^3

The last coefficient only looks like it is 0 because the entire value for the coefficients was not given.