Practice Questions for Ch. 3: Solutions

STAT 485/685 E100/G100: Applied Time Series Analysis

Fall 2020, Simon Fraser University

Question 1

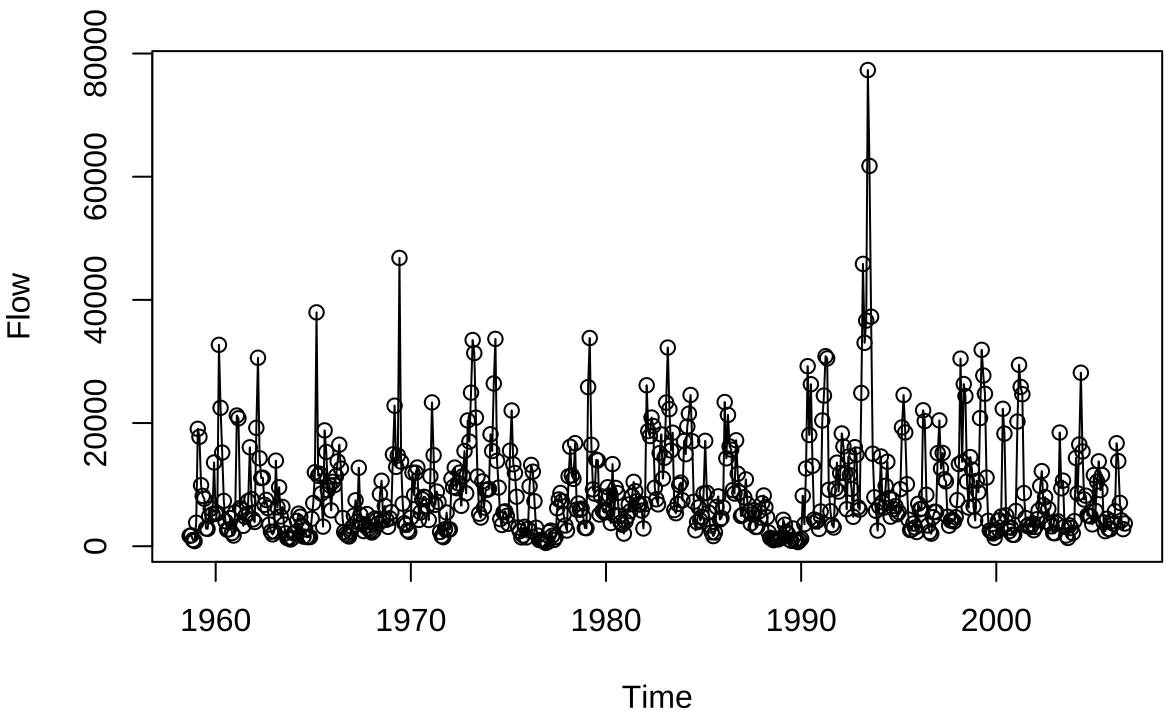
1a)

We'll choose the "flow" dataset in the TSA package. This dataset gives the flow (in cubic feet per second) of the Iowa River, measured at Wapello, Iowa for the period Sept. 1958 - Aug. 2006.

We load it in and plot it as follows:

```
library(TSA)
```

```
data(flow)
plot(flow, xlab='Time', ylab='Flow', type='o')
```



1b)

This time series appears to have a fairly constant mean. There are certainly several points that could be outliers, but its mean appears to be generally constant.

Just to make this problem a little more interesting, we will fit a linear trend to the dataset (since a constant trend is just a special case of the linear trend).

1c)

In order to see how times are coded up in this dataset, we'll do the following:

```r
time(flow)
```

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 1958
## 1959 1959.000 1959.083 1959.167 1959.250 1959.333 1959.417 1959.500 1959.583
## 1960 1960.000 1960.083 1960.167 1960.250 1960.333 1960.417 1960.500 1960.583
## 1961 1961.000 1961.083 1961.167 1961.250 1961.333 1961.417 1961.500 1961.583
## 1962 1962.000 1962.083 1962.167 1962.250 1962.333 1962.417 1962.500 1962.583
## 1963 1963.000 1963.083 1963.167 1963.250 1963.333 1963.417 1963.500 1963.583
## 1964 1964.000 1964.083 1964.167 1964.250 1964.333 1964.417 1964.500 1964.583
## 1965 1965.000 1965.083 1965.167 1965.250 1965.333 1965.417 1965.500 1965.583
## 1966 1966.000 1966.083 1966.167 1966.250 1966.333 1966.417 1966.500 1966.583
## 1967 1967.000 1967.083 1967.167 1967.250 1967.333 1967.417 1967.500 1967.583
## 1968 1968.000 1968.083 1968.167 1968.250 1968.333 1968.417 1968.500 1968.583
## 1969 1969.000 1969.083 1969.167 1969.250 1969.333 1969.417 1969.500 1969.583
## 1970 1970.000 1970.083 1970.167 1970.250 1970.333 1970.417 1970.500 1970.583
## 1971 1971.000 1971.083 1971.167 1971.250 1971.333 1971.417 1971.500 1971.583
## 1972 1972.000 1972.083 1972.167 1972.250 1972.333 1972.417 1972.500 1972.583
## 1973 1973.000 1973.083 1973.167 1973.250 1973.333 1973.417 1973.500 1973.583
## 1974 1974.000 1974.083 1974.167 1974.250 1974.333 1974.417 1974.500 1974.583
## 1975 1975.000 1975.083 1975.167 1975.250 1975.333 1975.417 1975.500 1975.583
## 1976 1976.000 1976.083 1976.167 1976.250 1976.333 1976.417 1976.500 1976.583
## 1977 1977.000 1977.083 1977.167 1977.250 1977.333 1977.417 1977.500 1977.583
## 1978 1978.000 1978.083 1978.167 1978.250 1978.333 1978.417 1978.500 1978.583
## 1979 1979.000 1979.083 1979.167 1979.250 1979.333 1979.417 1979.500 1979.583
## 1980 1980.000 1980.083 1980.167 1980.250 1980.333 1980.417 1980.500 1980.583
## 1981 1981.000 1981.083 1981.167 1981.250 1981.333 1981.417 1981.500 1981.583
## 1982 1982.000 1982.083 1982.167 1982.250 1982.333 1982.417 1982.500 1982.583
## 1983 1983.000 1983.083 1983.167 1983.250 1983.333 1983.417 1983.500 1983.583
## 1984 1984.000 1984.083 1984.167 1984.250 1984.333 1984.417 1984.500 1984.583
## 1985 1985.000 1985.083 1985.167 1985.250 1985.333 1985.417 1985.500 1985.583
## 1986 1986.000 1986.083 1986.167 1986.250 1986.333 1986.417 1986.500 1986.583
## 1987 1987.000 1987.083 1987.167 1987.250 1987.333 1987.417 1987.500 1987.583
## 1988 1988.000 1988.083 1988.167 1988.250 1988.333 1988.417 1988.500 1988.583
## 1989 1989.000 1989.083 1989.167 1989.250 1989.333 1989.417 1989.500 1989.583
## 1990 1990.000 1990.083 1990.167 1990.250 1990.333 1990.417 1990.500 1990.583
## 1991 1991.000 1991.083 1991.167 1991.250 1991.333 1991.417 1991.500 1991.583
## 1992 1992.000 1992.083 1992.167 1992.250 1992.333 1992.417 1992.500 1992.583
## 1993 1993.000 1993.083 1993.167 1993.250 1993.333 1993.417 1993.500 1993.583
## 1994 1994.000 1994.083 1994.167 1994.250 1994.333 1994.417 1994.500 1994.583
## 1995 1995.000 1995.083 1995.167 1995.250 1995.333 1995.417 1995.500 1995.583
## 1996 1996.000 1996.083 1996.167 1996.250 1996.333 1996.417 1996.500 1996.583
## 1997 1997.000 1997.083 1997.167 1997.250 1997.333 1997.417 1997.500 1997.583
## 1998 1998.000 1998.083 1998.167 1998.250 1998.333 1998.417 1998.500 1998.583
## 1999 1999.000 1999.083 1999.167 1999.250 1999.333 1999.417 1999.500 1999.583
## 2000 2000.000 2000.083 2000.167 2000.250 2000.333 2000.417 2000.500 2000.583
## 2001 2001.000 2001.083 2001.167 2001.250 2001.333 2001.417 2001.500 2001.583
```

```
## 2002 2002.000 2002.083 2002.167 2002.250 2002.333 2002.417 2002.500 2002.583
## 2003 2003.000 2003.083 2003.167 2003.250 2003.333 2003.417 2003.500 2003.583
## 2004 2004.000 2004.083 2004.167 2004.250 2004.333 2004.417 2004.500 2004.583
## 2005 2005.000 2005.083 2005.167 2005.250 2005.333 2005.417 2005.500 2005.583
## 2006 2006.000 2006.083 2006.167 2006.250 2006.333 2006.417 2006.500 2006.583
##           Sep      Oct      Nov      Dec
## 1958 1958.667 1958.750 1958.833 1958.917
## 1959 1959.667 1959.750 1959.833 1959.917
## 1960 1960.667 1960.750 1960.833 1960.917
## 1961 1961.667 1961.750 1961.833 1961.917
## 1962 1962.667 1962.750 1962.833 1962.917
## 1963 1963.667 1963.750 1963.833 1963.917
## 1964 1964.667 1964.750 1964.833 1964.917
## 1965 1965.667 1965.750 1965.833 1965.917
## 1966 1966.667 1966.750 1966.833 1966.917
## 1967 1967.667 1967.750 1967.833 1967.917
## 1968 1968.667 1968.750 1968.833 1968.917
## 1969 1969.667 1969.750 1969.833 1969.917
## 1970 1970.667 1970.750 1970.833 1970.917
## 1971 1971.667 1971.750 1971.833 1971.917
## 1972 1972.667 1972.750 1972.833 1972.917
## 1973 1973.667 1973.750 1973.833 1973.917
## 1974 1974.667 1974.750 1974.833 1974.917
## 1975 1975.667 1975.750 1975.833 1975.917
## 1976 1976.667 1976.750 1976.833 1976.917
## 1977 1977.667 1977.750 1977.833 1977.917
## 1978 1978.667 1978.750 1978.833 1978.917
## 1979 1979.667 1979.750 1979.833 1979.917
## 1980 1980.667 1980.750 1980.833 1980.917
## 1981 1981.667 1981.750 1981.833 1981.917
## 1982 1982.667 1982.750 1982.833 1982.917
## 1983 1983.667 1983.750 1983.833 1983.917
## 1984 1984.667 1984.750 1984.833 1984.917
## 1985 1985.667 1985.750 1985.833 1985.917
## 1986 1986.667 1986.750 1986.833 1986.917
## 1987 1987.667 1987.750 1987.833 1987.917
## 1988 1988.667 1988.750 1988.833 1988.917
## 1989 1989.667 1989.750 1989.833 1989.917
## 1990 1990.667 1990.750 1990.833 1990.917
## 1991 1991.667 1991.750 1991.833 1991.917
## 1992 1992.667 1992.750 1992.833 1992.917
## 1993 1993.667 1993.750 1993.833 1993.917
## 1994 1994.667 1994.750 1994.833 1994.917
## 1995 1995.667 1995.750 1995.833 1995.917
## 1996 1996.667 1996.750 1996.833 1996.917
## 1997 1997.667 1997.750 1997.833 1997.917
## 1998 1998.667 1998.750 1998.833 1998.917
## 1999 1999.667 1999.750 1999.833 1999.917
## 2000 2000.667 2000.750 2000.833 2000.917
## 2001 2001.667 2001.750 2001.833 2001.917
## 2002 2002.667 2002.750 2002.833 2002.917
## 2003 2003.667 2003.750 2003.833 2003.917
## 2004 2004.667 2004.750 2004.833 2004.917
## 2005 2005.667 2005.750 2005.833 2005.917
```

```
## 2006
```

It appears that each time unit corresponds to 1 year. So, separate months are coded up as fractions added to each year. For example: January 1959 is 1959.000 (= 1959 + 0/12), February 1959 is 1959.083 (= 1959 + 1/12), . . . , December 1959 is 1959.917 (= 1959 + 11/12).

1d)

We fit the linear trend model to this data:

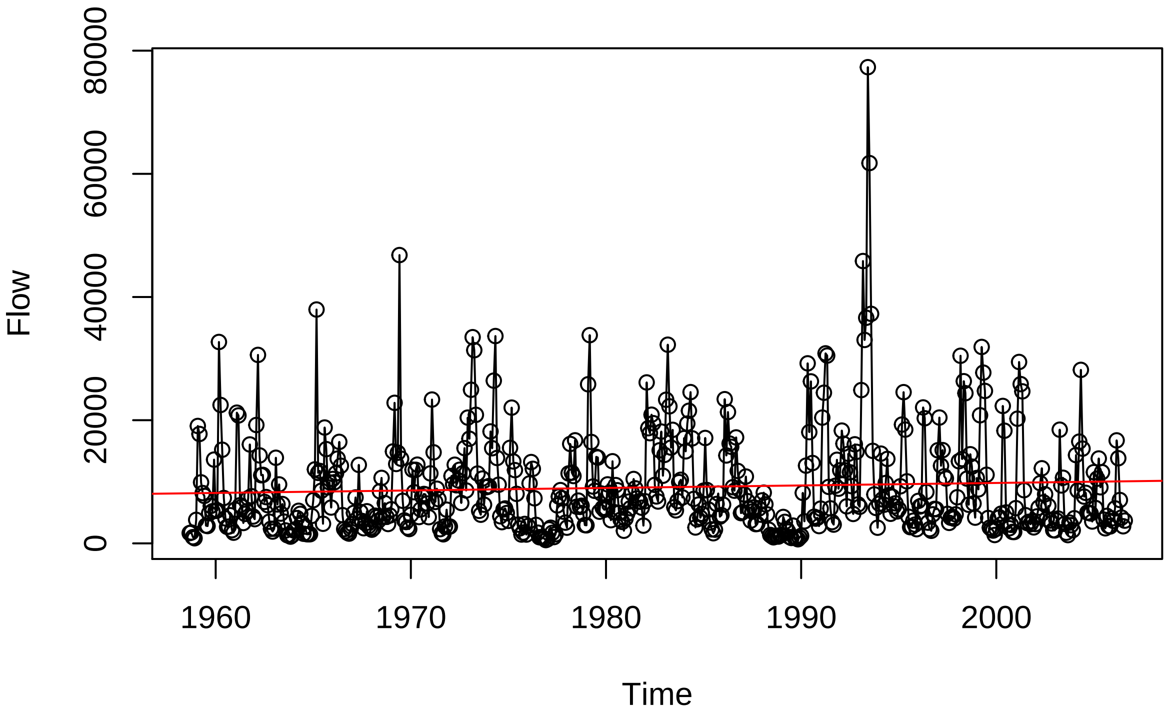
```r
lin.model <- lm(flow ~ time(flow))
summary(lin.model)
```

```
##
## Call:
## lm(formula = flow ~ time(flow))
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8730  -5683  -2859   3053  67781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -71308.76   50393.17  -1.415    0.158
## time(flow)      40.56      25.42   1.596    0.111
##
## Residual standard error: 8452 on 574 degrees of freedom
## Multiple R-squared:  0.004416,   Adjusted R-squared:  0.002682
## F-statistic: 2.546 on 1 and 574 DF,  p-value: 0.1111
```

The estimate of the intercept is $\hat{\beta}_0 = -71,308.76$, and the estimate of the slope is $\hat{\beta}_1 = 40.56$.

We then add it to the plot:

```r
plot(flow, xlab='Time', ylab='Flow', type='o')
abline(lin.model, col='red')
```



We notice that the slope is very small when compared to the actual y-values, so the line is nearly constant.

1e)

The equation for the estimated mean river flow at any time t is:

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 t = -71,308.76 + 40.56t.$$

1f)

From 1c), we saw that this dataset spans from Sept. 1958 to Aug. 2006. We decide to pick the time Oct. 2001. Since October is the 10th month, this corresponds to 2001 + 9/12 = 2001.75. We can confirm this using the table in 1c).

1g)

The estimate of the mean river flow in Oct. 2001 is:

$$\hat{\mu}_{2001.75} = -71,308.76 + 40.56(2001.75) = 9,882.2.$$

If we wanted a more precise calculation (using all the significant digits), we could use R to do this:

```
mu_hat <- coef(lin.model)[1] + coef(lin.model)[2]*2001.75
print(as.numeric(mu_hat))
```

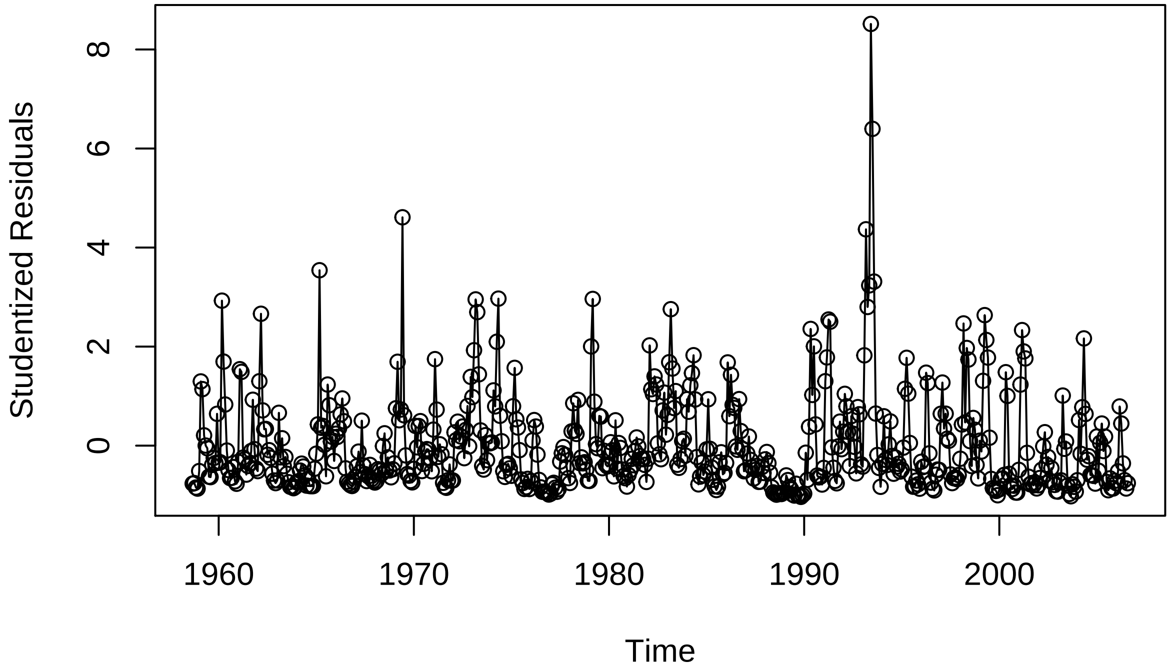
```
## [1] 9877.311
```

We see that the values are slightly different, due to rounding error in using the first method.

1h)

We plot the residuals vs. time for this model:

```
plot(y=rstudent(lin.model), x=as.vector(time(flow)), type='o',
         ylab='Studentized Residuals', xlab='Time')
```



The first thing that we notice is that the plot of residuals vs. time looks very similar to the original dataset. This is to be expected, because the slope of the line was very small (compared to the y-values), and so the line is nearly horizontal. Since the residuals are:

$$\hat{X}_t = Y_t - \hat{\mu}_t,$$

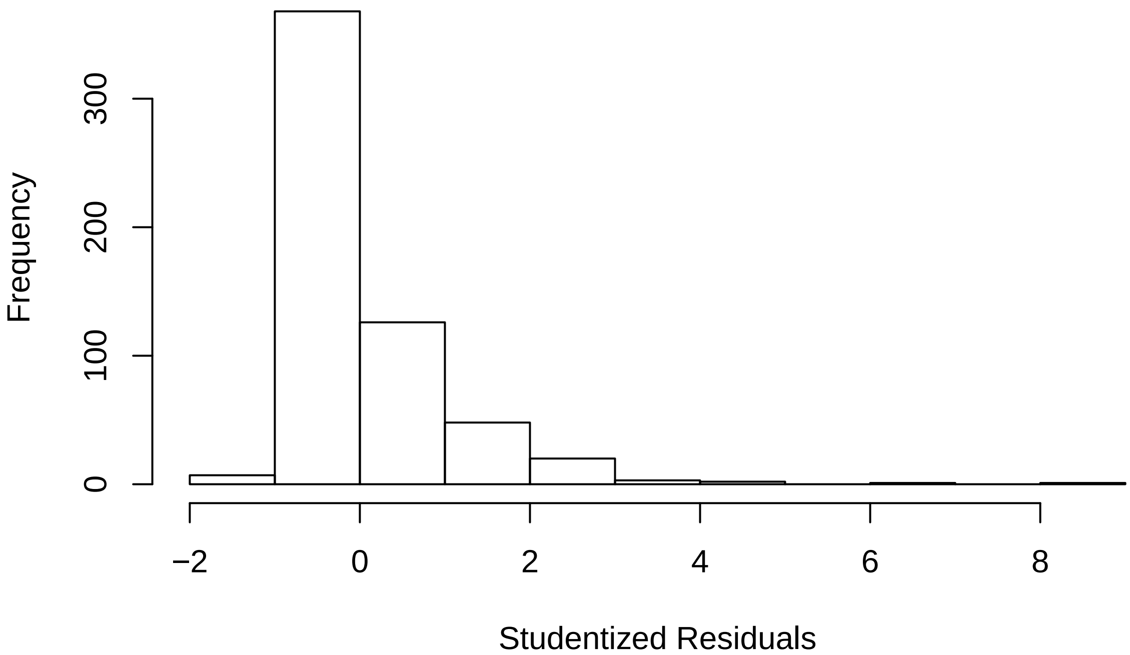
and $\hat{\mu}_t$ is nearly a constant, this means that the residuals $\hat{X}_t$ look very similar to the data Y_t.

Looking at the patterns in the plot, we see that there is a fairly random scatter about zero. Points don't "hang together" too much, and the variances look fairly constant for different values of time.

However, we see what appears to be an outlier around 1993.

Also, while the residuals are randomly scattered about zero, they don't appear to be randomly scattered symmetrically – they are much more likely to fall high above zero than they are to fall far below it. This suggests to us that the residuals are right-skewed, and therefore not normally distributed. We can test this suspicion using the histogram of residuals:

```r
hist(rstudent(lin.model), xlab='Studentized Residuals', main='')
```



Question 2

2a)

We begin by fitting the "cosine curve plus linear time trend model", using the code given to us in the question:

```r
data(co2)
har. <- harmonic(co2, 1)
my.model <- lm(co2 ~ har. + time(co2))
summary(my.model)
```

```
##
## Call:
## lm(formula = co2 ~ har. + time(co2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1804 -1.7916 -0.1045  1.8986  5.1809
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -3.332e+03  1.280e+02  -26.03   <2e-16 ***
## har.cos(2*pi*t)    3.851e+00  2.868e-01   13.43   <2e-16 ***
## har.sin(2*pi*t)    5.596e+00  2.874e-01   19.47   <2e-16 ***
## time(co2)          1.851e+00  6.401e-02   28.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.329 on 128 degrees of freedom
## Multiple R-squared:  0.9109, Adjusted R-squared:  0.9088
```

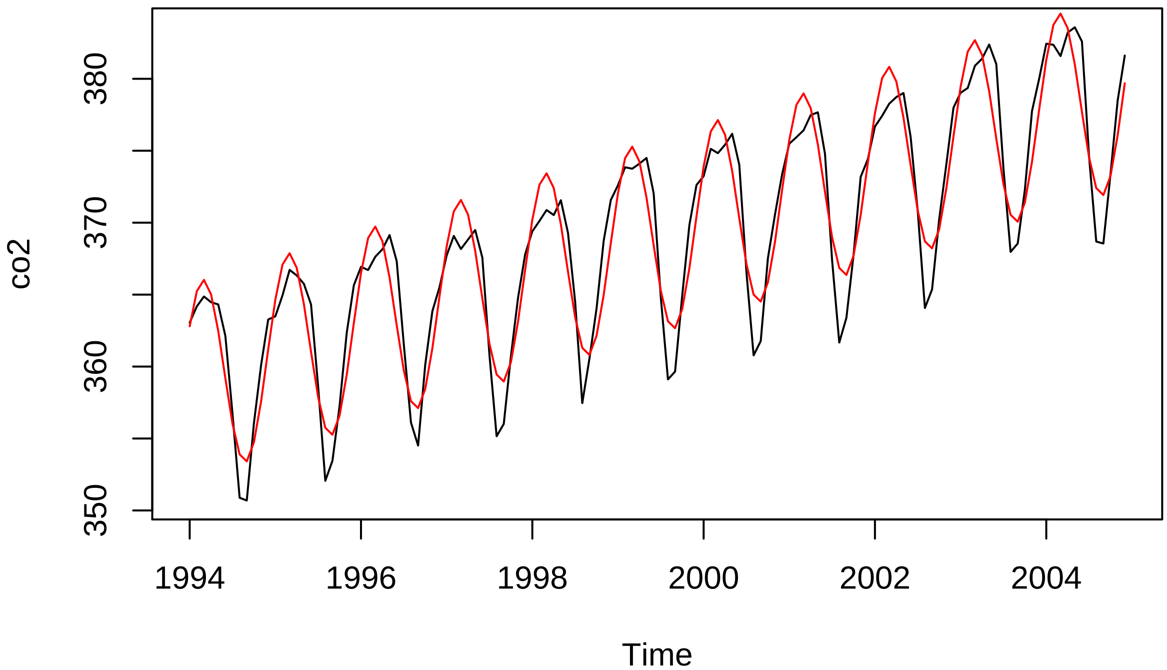
```
## F-statistic: 436.3 on 3 and 128 DF,  p-value: < 2.2e-16
```

The estimates are: $\hat{\beta}_0 = -3,332$, $\hat{\beta}_1 = 3.851$, $\hat{\beta}_2 = 5.596$ and $\hat{\beta}_3 = 1.851$.

2b)

Next, we plot the data along with the fitted trend:

```
plot(co2)
lines(x=as.vector(time(co2)), y=as.vector(fitted(my.model)), col='red')
```

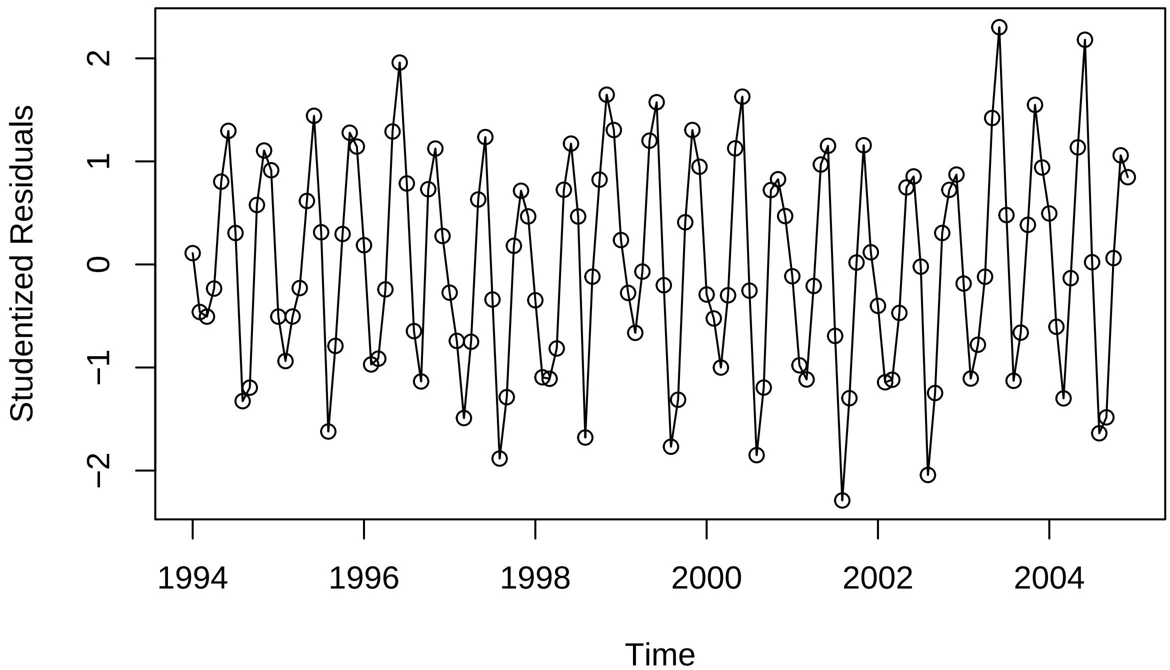


The fitted cosine curve + linear trend fits the data fairly well, although it overestimates most of the extreme values (both minima and maxima). But it captures both the cyclical behaviour and the linear trend fairly well.

2c)

We plot the residuals vs. time for this model:

```
plot(y=rstudent(my.model), x=as.vector(time(co2)), type='o',
         ylab='Studentized Residuals', xlab='Time')
```



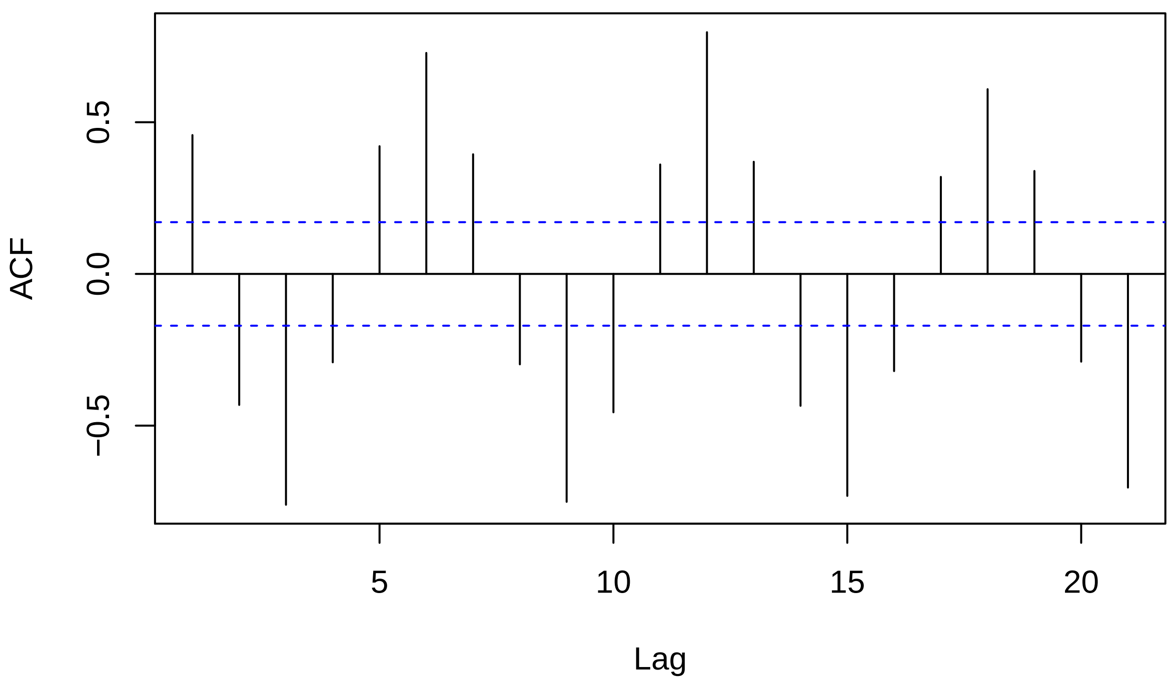
This appears to be a somewhat random scatter about zero. However, if we look closely, there does appear to be some sort of "smoothness" in the movement from one residual to the next; in other words, they do appear to "hang together" a little bit. This suggests that there may be some correlations in the residuals.

The variability appears to be constant across time, however.

2d)

Next we create the sample ACF plot for the residuals of this model:

```
acf(rstudent(my.model), main='')
```



This plot shows us the sample ACF of the residuals (which is an estimate of the autocorrelation function ρ_k of the residuals, and therefore an estimate of the autocorrelation function of $\{X_t\}$), as a function of k.

In this plot, we can clearly see that there are some correlations in the residuals, as supported by our observations in 2c). Many of the estimates of the ρ_k's fall outside of the blue dashed lines. For instance, the estimate of $\rho_1 = Corr(\hat{X}_t, \hat{X}_{t-1})$ is approximately 0.5, meaning that residuals that are 1 point apart are moderately positively correlated. We also see many other significant correlations – both positive and negative.

This is telling us that the process $\{X_t\}$ is almost certainly not a white noise process, since the residuals are not independent observations (and so the X_t's are probably also not independent observations).