The aims of this Computing Practical are to

- examine the serial dependence structure of a time series using the correlogram,
- show that trend removal is important in establishing the dependence structure in a time series, but that it can introduce additional autocorrelation,
- examine a time series for cyclic patterns using the periodogram,
- interpret the periodogram, and
- examine a test for white noise based on the periodogram.

We will analyse two datasets in this practical: the `UKLungDeaths` data we considered in Computing Practical 2, and data on levels of lutenizing hormone (LH) in blood samples. To obtain the data for this session enter R and type

```
> library(ts)
> data(UKLungDeaths)
> data(lh)
```

Recall that the `UKLungDeaths` dataset contains three time series on monthly total, male and female deaths from common lung diseases in the UK for the years 1974 to 1979.

It is a good idea to set up the on-line help at the beginning of your R session using `help.start()`. Remember to keep using the `help` facility throughout this exercise.

1. The functions `ts.union` and `ts.intersect` bind together multiple time series. The time axes are aligned and only observations at times that appear in all the series are retained with `ts.intersect`.

   For the male and female deaths series, type

   ```
   > deaths <- ts.union(mdeaths, fdeaths)
   > deaths  # look at what has been created
   ```

   The function `aggregate` can be used to change the frequency of the time base. For example, to obtain quarterly sums or annual means of deaths, try

   ```
   > aggregate(deaths, 4, sum)
   > aggregate(deaths, 1, mean)
   ```

2. We will now examine the data more closely. Plot both the male and female time series and their associated correlograms. You can do this using

   ```
   > plot.ts(deaths)
   > acf(deaths)
   ```

   The last command is the same as

   ```
   > acf(ts.union(mdeaths, fdeaths))
   ```

Note that approximate 95% confidence limits are shown for the autocorrelation plots: these are for an independent series for which $\rho_k = 0$.

Describe the obvious features of the correlograms. (Ignore the two cross-series for now.)

3. For each of the male and female series, find a simple moving average of order three and calculate the residual series (i.e. the original minus the smoothed series); you saw how to do this in Computing Practical 2. Plot the two residual series and their associated correlograms.

   Based on these plots, what do you think would be a reasonable description (or model) for these data?

   Note that you can either analyse `mdeaths` and `fdeaths` separately, or use the bivariate series `deaths` you created above. For example, you could use

```
> deaths.ma3 <- filter(deaths, rep(1,3)/3)
                # fit the 3-point moving average
> res.ma3 <- deaths - deaths.ma3
> plot.ts(res.ma3)
```

   The residual series will have `NA` for the first and last values so to avoid problems, redefine the series using

```
> res.ma3  <-  window(res.ma3,start=c(1974,2),end=c(1979,11))
> res.ma3  <-  as.ts(res.ma3)
```

   Now you can type

```
> acf(res.ma3)
```

   If you analyse each series separately, it may be helpful to put your plots together on one page; to do this, type

```
> par(mfrow=c(2,2))
```

   If you use the bivariate series structure `deaths`, `par(mfrow=c(2,1))` may be useful. (If later you want a single plot to a page, type `par(mfrow=c(1,1))`.)

4. * The workhorse function for spectral analysis in R is `spectrum`, which with its default options computes the periodogram ordinates and plots the periodogram on the log scale. (The function `spectrum` calls `spec.pgram` to do most of the work.)

   Note that by default, `spectrum` removes a linear trend (using least squares) from the series before estimating the periodogram.

   For example, we can type

```
> spectrum(deaths)
```

to plot the 'raw periodogram' for both the male and female series. Remember we discussed in lectures why R plots the periodogram on the log scale (in fact, it plots $10 \log_{10} I(\omega)$ versus ω): this is because the raw periodogram as an estimate of the spectral density of the series has approximately constant variation on the log scale.

You can suppress the raw periodogram plot but create the periodogram ordinates by

```
> deaths.spec <- spectrum(deaths, plot=FALSE)
> deaths.spec$spec
    # these are the periodogram ordinates for both series
> deaths.spec$freq  # these are the frequencies omega
```

To obtain the periodogram as in Diggle (1990), type

```
> plot.ts(deaths.spec$spec, xlab="Fourier frequency",
        main="Periodogram")
```

The first series gives the periodogram ordinates for UK males; the second is for the females.

Plot the periodograms for the male and females series and describe the features in the two plots. In particular, which frequencies are dominant? What are the implications, if any, for modelling the annual cycle?

5. The male data can be analysed separately using `deaths.spec$spec[,1]`, and similarly for the female series. For example, to plot the periodogram ordinates against the frequency in cycles, type

```
> plot(deaths.spec$freq, deaths.spec$spec[,1],  type="l",
   xlab="frequency in cycles", ylab="P. ordinates for mdeaths")
```

Repeat this command using `plot.ts` instead of `plot` and observe what happens.

Now plot the periodogram separately for the female series.

6. Re-create the periodograms but now use the option `detrend=FALSE`. Comment on the plots.

7. Repeat the periodogram calculations for each of the residual series and comment on what you see.

8. * The periodogram ordinates can be used to examine the question of whether a time series consists of white noise. We do this by plotting the *cumulative periodogram* using the function

```
> cpgram(<name of series>)
```

Plot cumulative periodograms for the male and female series and interpret your results.

Repeat for both the residual series. Again, what do you conclude?

The dataset `data(lh)` is a time series of levels of lutenizing hormone (LH) in blood samples taken at 10-minute intervals over an eight-hour period. LH plays an important role in the reproductive system and the series is from the early follicular phase of the subject's menstrual cycle. One of the questions posed by these data is whether there are cyclic patterns due to pulsatile release of LH.

9. Plot the LH time series, sample autocorrelation function, periodogram and cumulative periodogram. Is there any evidence of serial dependence and/or cycling in this series? If so, describe these features.

   Can you suggest an appropriate model for these data?

   The R `help(lh)` file is given on the next page.

Questions 4 and 8 above are to be handed in as part of Assignment 2.

Assignment 2 is due by noon on Friday 17th September (Week 8).

Patty Solomon
August 2004

lh {ts} R Documentation

Luteinizing Hormone in Blood Samples

Description

A regular time series giving the luteinizing hormone in blood samples at 10 mins intervals from a human female, 48 samples.

Usage

```
data(lh)
```

Source

P.J. Diggle (1990) *Time Series: A Biostatistical Introduction.* Oxford, table A.1, series 3

[Package Contents]