3 Graphics in S-PLUS

3.1 Introduction

So far the course has concentrated on ANOVA models (for designed experiments), and you have used the `aov` function within S-PLUS to fit the models. For example, a one-way ANOVA using

```r
growth.aov <- aov(length ~ sugar, data=growth)
```

You have also seen how to fit a two-way ANOVA using, for example

```r
battery.aov <- aov(battery.life ~ material.fac + temp.fac + material.fac:temp.fac, data=battery)
```

(equivalently, `aov(battery.life ~ material.fac*temp.fac, data=battery)`).

The resulting ANOVA table is then produced using `summary(growth.aov)`, etc. Little has been covered in the way of graphical analysis, but you have seen how to plot histograms and boxplots, to examine data (and you may have explored how to undertake a normal probability plot to assess normality, using `qqnorm()`).

There are a number of graphical means which can be used to further examine your data, or to assess the assumptions of the ANOVA model, which we can consider here. We will also explore some of what S-PLUS has to offer in terms of the presentation of graphics.

We will begin by considering the use of default graphs which are available as generic functions within S-PLUS, and will then consider customisable graphics separately.

3.2 Graphics associated with ANOVA

In S-PLUS the generic plotting function `plot` can be applied to saved objects to produce default graphs associated with that object. The following are relevant examples connected with ANOVA.

3.2.1 Confidence Intervals from `multicomp`

```r
plot(multicomp(growth.aov, error.type="fwe", method="tukey", alpha=0.05)).
```

results in the plot shown overleaf, which gives an immediate display of the confidence intervals from a set of (pairwise) multiple comparisons.
Equivalently, we can add `plot=T` in a call to the `multicomp` function

```
multicomp(growth.aov,plot=T),
```

which plots the confidence intervals, shown above, in addition to the usual output.

Note that `multicomp()` is itself an object, so can be assigned, and it’s components can be individually saved and accessed. For example,

```
multicomp(growth.aov)$table
```

gives a matrix, the rows of which contain the estimated value of the pairwise contrast, it’s standard error and the lower and upper confidence limits. (Other values which can be extracted from the `multicomp` object, following the $, can be found in the S-PLUS Language Reference).

### 3.2.2 Residual Plots from an ANOVA model

`plot(growth.aov)`, for example, achieves a series of (six) diagnostic residual plots which are useful for assessing the fit/appropriateness of a fitted ANOVA (aov) model, or linear (lm) model - see Lecture 8.

These will be plotted on six different pages in a graph window. (Recall that a call to a graph function will open a new graph window, called a graphsheet, and that further calls result in the current graph being overwritten).

The graphs are:

1. Residuals against Fitted Values;
2. Square Root of Absolute Residuals against Fitted
Values; (3) Response vs Fitted Values; (4) Normal QQplot of Residuals; (5) Residual-Fit (r-f) spread plot (if the model fits well then the residuals have a small spread relative to the fit); and, (6) Cook’s Distances (which measure the influence of each observation on the overall fit).

The graphs can be seen on one page by using the commands

```r
par(mfrow=c(2,3))  # this splits the graphsheet into 2x3 panes
plot(growth.aov,smooths=T,id.n=F)
```

and are shown below. (Note that `par(mfrow=c(1,1))` should be used to reset the graphics parameters).

![Graphs showing various statistical plots for model fit assessment](image)

The required graphs can be selected using `which=c`, where `c` is the number of the appropriate graph. (For example, `which=c(1,4)`, selects (1) and (4) above). The option `smooths=T` asks that smoothed curves be added to the scatterplots, and `id.n=F` asks that no points be automatically identified as unusual. (ask=T, opens an interactive dialogue in the `commands` window from which the graphs can be selected one-by-one).
Of course, such graphs can be constructed individually (‘by hand’), using appropriate commands. For example,

```r
plot(fitted.values(growth.aov), residuals(growth.aov))
qqnorm(residuals(growth.aov))
qqline(residuals(growth.aov))
```

It is also appropriate to look at a plot of the residuals against the explanatory variables. (Note that a factor will have to be recoded as a numeric variable in order to produce a scatterplot (rather than the default boxplot). i.e.

```r
plot(as.numeric(sugar), residuals(growth.aov))
abline(h=0)
```

The `abline` function is useful to add a line to a current plot.

```r
abline(a,b) # adds a line with intercept a, and slope b
abline(h=0) # adds a horizontal line through 0 (on the y axis)
abline(v=0) # adds a vertical line through 0 (on the x axis)
```

### 3.2.3 Interaction Plots for two-way ANOVA

The interaction between factors in a two-way analysis of variance can be examined using an interaction plot. For example (for the battery operating conditions example)

```r
interaction.plot(material.fac, temp.fac, battery.life)
```

Recall that the title can be added using `title("Interaction Plot")`.

The plot shows the mean battery life given by each of the material types at each of the three levels of temperature, and the levels correspond to the mean values for the `material.fac:temp.fac` interaction given by

```r
model.tables(battery.aov, type="means")
```

which are shown below.

```
  material.fac:temp.fac
Dim 1 : material.fac
Dim 2 : temp.fac
     1    2    3
 1  134.75  57.25  57.50
 2  155.75 119.75  49.50
 3  144.00 145.75  85.50
```
3.3 Customised Graphics in S-PLUS

S-PLUS is a very powerful tool for creating high quality graphics, suitable for publication. By appropriate use of the programming commands it is possible to customize the results of the graph calls.

3.3.1 graphsheet()

The default graph method in S-PLUS uses the graphsheet function to open the graphics device window and plot the requested graphic. Using the command

graphsheet()

before a graphics call, will result in a new graphics window being opened, so that any current graphics displayed will not be overwritten. (Selecting a number of graph calls from a script window together and running them simultaneously will result in the graphs being shown on separate pages within the current graphsheet.

The default parameters of the graphsheet, such as size, number of ‘panes’, margin widths,
etc, can be set using the \texttt{par()} function. \texttt{e.g. par(mfrow=c(2,2))} results in graphs being plotted sequentially row-by-row in a two-by-two array. (See the online Language Reference).

Note that parameter settings will be reset (to the appropriate defaults) at each session startup, but can be reset ‘by hand during a session. One method is to save the current settings so they can be reset when reequired, using, say,

\begin{verbatim}
parold <- par()
par <- parold
\end{verbatim}

\subsection*{3.3.2 Saving Graphs}

With the current graphsheet that you wish to save open and \textit{active} in S-PLUS, then selecting \texttt{File>Save As} will result in the opening of a \texttt{SAVE GRAPH SHEET AS} dialogue which allows you to save your chosen graph as an S-PLUS graph sheet, .\texttt{sgr}, which can be recalled to S-PLUS at any future time. However, the use of \texttt{File>EXPORT GRAPH} allows you to save the current graphsheet in a number of different formats which can be used in other software, \texttt{e.g. jpeg, bitmap, encapsulated postscript}.

\subsection*{3.3.3 Customising Plots}

The generic plotting function \texttt{plot} in S-PLUS is easy to use to obtain default graphics, but it also allows you to have a great deal of control over the way your graphs look by using additional commands. For example, symbols used for points can be changed, lines and text added, etc.

To give you a small flavour of what S-PLUS is capable of, you might like to consider (in advance) the pollution dataset from Examples 8 which is available from the course webpage as \texttt{pollution.sdd}. This dataset contains observations on a response variable \texttt{SO2}, the annual mean concentration of sulphur dioxide for 41 American cities, and a number of explanatory variables, \texttt{temp, manufact, popul, wind, precip}, and \texttt{pdays}. The variable \texttt{city} contains the names of the American cities from which the data were collected.

Compare the following graphics calls (shown overleaf).

\begin{verbatim}
plot(temp,SO2) # default plot
plot(temp,SO2,pch=3,xlim=c(0,80),ylim=c(0,120),main="SO2 versus temp")
\end{verbatim}

A plot with text labels indicating the cities can be produced using, for example

\begin{verbatim}
plot(temp,SO2,pch=3,main="SO2 versus temp")
text(temp+0.5,SO2-0.5,as.character(city))
\end{verbatim}
You might like to experiment with some of the other explanatory variables. If required, a very good introduction to the graphical capabilities of S-PLUS can be obtained from Chapter 3 of

Note that S-PLUS contains a suite of advance *trellis* graphics, which can be used for example to obtain a scatterplot matrix of plots concerning the response and explanatory variables.

```r
splom(~pollution[,1:7])
title("Scatterplot Matrix")
```
Appendix: A Composite Residual Plot Function

The following user defined function may be used to plot a collection of useful residual plots from any saved aov model object.

```r
resplot <- function(aov.object){
  parold <- par()
  graphsheet()
  par(mfrow=c(2,2))
  plot(aov.object,which=c(1,4,3,6),smooths=T,id.n=F)
  mtext(outer=T,"Composite Residual Plot",side=3,line=-1.5,cex=1.5)
  par <- parold
}
```

For example, using the growth data, we obtain:

```r
resplot(growth.aov)
```

Composite Residual Plot

![Composite Residual Plot](image)