

Introduction to S-PLUS

1 Preliminaries

You are expected to read the notes on S-PLUS (or Venables & Ripley, 2002, of which they are a modified extract) and to make use of the help facilities of S-PLUS to find out what commands to use and what commands you are given mean.

There are two colour schemes for graphics windows: `graphsheat()` gives the standard one (with a white background) and `trellis.device()` the other (with a grey background). It is your choice, but some colours only show up well in the Trellis scheme. At least in theory, if no window is open one of the appropriate type for the next plot is opened.

You can view a graph full-screen at any time by pressing F2; left-click on the plot to restore the normal view.

2 Acclimatization

Work through the introductory session in section 1.3 of your notes. You don't need to type all the commands yourself if you open the script

```
L:\splus6libs\MASS-keep\scripts\ch01.ssc
```

(navigate to it in 'My Computer' and double-click on it once you have an S-PLUS session running). You can read about using script windows in Appendix A of your notes.

3 Two-sample problems

We can try out the example in problem 4 of the week 0 problem sheet. To remind you, this was

The data in the table below are from two very similar experiments to measure acceleration due to gravity. The data given are $10^3 \times (\text{measurement in cm/sec}^2 - 980)$.

experiment 1: 95 90 76 76 87 79 77 71

experiment 2: 82 79 81 79 77 79 79 78 79 82 76 73 64

- Say how you would check that each sample comes from a normal distribution.
- Make statistical comparisons of the averages and of the variances of the two samples and interpret the results.
- What recommendation would you make concerning additional measurements if you wanted to get a 95% confidence interval of length 2 for the acceleration?

- (d) How would you compare the samples if you knew they did not come from normal distributions?

First we need to enter the data. We can do this in many ways:

1. Enter a vector using the `c()` (concatenate) function:

```
ex1 <- c(95,90,76,76,87,79,77,71)
```

2. Use `scan()`:

```
ex1 <- scan()  
95 90 76 76 87 79 77 71
```

noting that there is a final blank line.

3. Enter the data using your favourite text editor (or even Word), and then use `scan`.

```
ex1 <- scan("ex1.dat")  
ex2 <- scan("ex2.dat")
```

4. Use a data window for spreadsheet-like data entry. We leave you to explore this in Chapter 2 of the on-line Users Guide (look under the help menu for **On-line Manuals**).

You can also import data from other programs, even from spreadsheet files. See the **Import Data** item on the **File** menu.

To answer (a) we could use a normal-probability plot, also known as a QQ-plot, and also perform a Kolmogorov–Smirnov test.

```
trellis.device()  
par(pty="s") # set up a square plot  
qqnorm(ex1); qqline(ex1)  
par(pty="m") # reset  
  
# or, for a more elegant plot  
qqmath(~ ex1, distribution=qnorm, aspect="xy",  
prepanel = prepanel.qqmathline,  
panel = function(x, y, ...) {  
  panel.qqmathline(y, distribution=qnorm, ...)  
  panel.qqmath(x, y, ..., col=1)  
},  
xlab= "Quantiles of Standard Normal"  
)  
  
ks.gof(ex1)  
cdf.compare(ex1, distribution="normal", mean=mean(ex1), sd=sqrt(var(ex1)))
```

Try `ex2` too, of course.

For part (b) we need to compare the variances first (as how we compare the means will depend on the answer).

```
var.test(ex1, ex2)  
t.test(ex1, ex2, var.equal=F)
```

Part (c) needs some calculation, but is as easy to do by hand.

For part (d) the suggestion is to use the Mann-Whitney test, sometimes known as the two-sample Wilcoxon test.

```
wilcox.test(ex1, ex2)
```

Latent heat

Rice (1995, p.390) gives the following data on the latent heat of the fusion of ice (*cal/gm*):

```
Method A: 79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97
           80.05 80.03 80.02 80.00 80.02
Method B: 80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97
```

- (a) Assuming normality, test the hypothesis of equal means, both with and without making the assumption of equal variances. Compare the result with a Wilcoxon/Mann-Whitney nonparametric two-sample test.
- (b) Fit a one-way analysis of variance and compare it with your t -test. To do this set up a data frame

```
## assume vectors a and b contain the data for methods A and B ....
ice <- data.frame(lheat=c(a, b), method=rep(c("a", "b"), c(13,8)))
ice # print it to take a look at a data frame
ice.aov <- aov(lheat ~ method, data=ice)
summary(ice.aov)
```

Data frames are the most widely-used data structure in **S-PLUS**; they are made up of columns of equal lengths. Often (as here) some of the columns are numerical and some are categorical.