

*Preliminaries*

```
> library(DAAG)
```

*Exercise 1*

Repeat the principal components calculation omitting the points that appear as outliers in Figure 13.1, and redo the regression calculation. What differences are apparent, in loadings for the first two principal components and/or in the regression results?

The following repeats the calculations that are described in the text.

```
> not.na <- complete.cases(socsupport[, 9:19])
> not.na[36] <- FALSE
> ss.pr <- princomp(as.matrix(socsupport[not.na, 9:19]), cor = TRUE)
> ss.lm <- lm(BDI[not.na] ~ ss.pr$scores[, 1:6], data = socsupport)
> attach(socsupport)
> plot(BDI[not.na] ~ ss.pr$scores[, 1], col = as.numeric(gender[not.na]),
+      pch = as.numeric(gender[not.na]), xlab = "1st principal component",
+      ylab = "BDI")
> topleft <- par()$usr[c(1, 4)]
> legend(topleft[1], topleft[2], col = 1:2, pch = 1:2, legend = levels(gender))
> detach(socsupport)
```

Examination of Figure 13.1 makes it clear that we need to omit points for which BDI is greater than 35. We determine the relevant row numbers:

```
> (1:95)[socsupport$BDI > 35]
```

```
[1] 36 68 95
```

Row 36 had already been omitted. We need, additionally, to omit rows 68 and 95. The following repeats the calculations given above, but now with observations 36, 68 and 95 omitted:

```
> not3.na <- complete.cases(socsupport[, 9:19])
> not3.na[c(36, 68, 95)] <- FALSE
> ss3.pr <- princomp(as.matrix(socsupport[not3.na, 9:19]), cor = TRUE)
> ss3.lm <- lm(BDI[not3.na] ~ ss3.pr$scores[, 1:6], data = socsupport)
> attach(socsupport)
> plot(BDI[not3.na] ~ ss3.pr$scores[, 1], col = as.numeric(gender[not3.na]),
+      pch = as.numeric(gender[not3.na]), xlab = "1st principal component",
+      ylab = "BDI")
> topleft <- par()$usr[c(1, 4)]
> legend(topleft[1], topleft[2], col = 1:2, pch = 1:2, legend = levels(gender))
> detach(socsupport)
```

The following (shown in the left panel below) compares the loadings, with (*x*-axis) and without (*y*-axis) rows 68 and 95.

