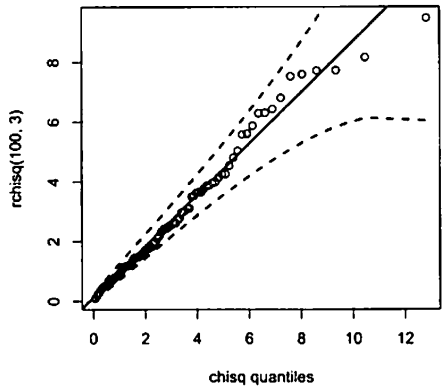

Figure 3.5 Quantile-comparison plot of a sample of size $n = 100$ from the $\chi^2(3)$ distribution against the distribution from which the sample was drawn.

the distribution from which it was drawn, producing Figure 3.5:

```
> set.seed(124) # for reproducibility
> qqPlot(rchisq(100, 3), distribution="chisq", df=3)
```

The points should, and do, closely match the straight line on the graph, with the fit a bit worse for the larger values in the sample. The confidence envelope suggests that these deviations for large values are to be expected, as they reflect the long right tail of the $\chi^2(3)$ density function.

3.1.4 BOXPLOTS

The final univariate display that we describe is the *boxplot*. Although boxplots are most commonly used to compare distributions among groups (as in Section 3.2.2), they can also be drawn to summarize a single sample, providing a quick check of symmetry and the presence of outliers. Figure 3.6 shows a boxplot for `income`, produced by the `Boxplot` function in the **car** package:[6]

```
> Boxplot( ~ income, data=Prestige)

[1] "general.managers"          "lawyers"
[3] "physicians"                "veterinarians"
[5] "osteopaths.chiropractors"
```

The variable to be plotted is given in a *one-sided formula*: a tilde (`~`) followed by the name of the variable. This variable is contained in the data frame `Prestige`, and the `data` argument is used to tell the function where to find the data. Most graphical functions that use a formula accept a `data` argument.

[6]The standard R `boxplot` function can also be used to draw boxplots, but `Boxplot` is more convenient, automatically identifying outliers, for example; indeed, `Boxplot` is simply a front-end to `boxplot`.

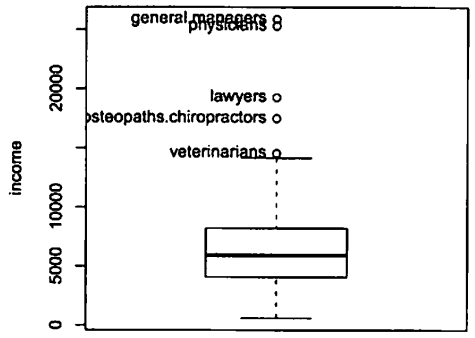

Figure 3.6 Boxplot of `income`. Several outlying observations were labeled automatically.

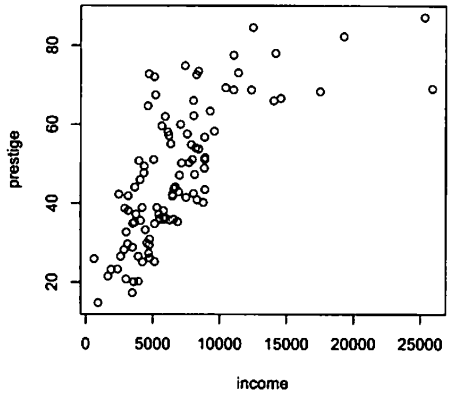

Figure 3.7 Simple scatterplot of `prestige` versus `income` for the Canadian occupational-prestige data.

3.2 Examining Relationships

3.2.1 SCATTERPLOTS

A *scatterplot* is the familiar graph of points with one quantitative variable on the horizontal or x-axis and a second quantitative variable on the vertical or y-axis. Understanding, and using, scatterplots is at the heart of regression analysis. There is typically an asymmetric role of the two axes, with the y-axis reserved for a response variable and the x-axis for a predictor.

The generic `plot` function is the primary tool in R for drawing graphs in two dimensions. What this function produces depends on the values of its first one or two arguments.[7] If the first two arguments to `plot` are numeric vectors, then we get a scatterplot, as in Figure 3.7:

```
> with(Prestige, plot(income, prestige))
```

[7]The behavior of generic functions such as `plot` is discussed in Sections 1.4 and 8.7, and more information about the `plot` function is provided in Section 3.2.3 and in Chapter 7 on R graphics.