

## Stat 5421, Fall 2006: Data for exam 3

Marketing researchers have long believed that companies that have a large share of a market for a particular type of product can maintain market leadership over many years, even decades. These data are from a study in the late 1990s that started with a book written in 1923 that listed the market leaders in 100 different categories based on a survey of about 1000 people. The market share of these companies in 1997 was then determined for 95 of these categories (five categories were no long in existence). The results are shown below, and in the file `mshare.txt` on the data page of the class webpage:

```
> file <- "http://www.stat.umn.edu/~sandy/courses/5421/data/mshare.txt"
> mshare <- read.table(url(file), header = TRUE)
> xtabs(y ~ with(mshare, type:P1923) + P1997, data = mshare)
```

```

                P1997
with(mshare, type:P1923) C1 C2 C3 C4 C5 C6 Fail
Durable:N1              7  2  6  3  1  8  19
Durable:N2              3  2  0  0  3  7  13
Durable:N3              1  1  0  1  1  1  12
Durable:N4              1  0  1  1  1  3   1
Durable:N5              0  0  2  0  1  0   1
Nondurable:N1          15  5  3  5  6  9   8
Nondurable:N2           4  4  2  3  3 11  14
Nondurable:N3           1  2  1  0  3  5  13
Nondurable:N4           0  1  0  0  1  8   7
Nondurable:N5           0  0  1  0  0  5   1
```

The variables in the data file are `y`, the cell counts, `type`, the category of goods, either durable, like cars or appliances, or nondurable, like food or clothing; `P1923`, the market position in 1923, and `P1997`, the position in 1997. The category labels for `P1923` are `N1`, `N2`, `N3`, `N4` and `N5`, corresponding to ranked number 1, number 2, . . . , number 5. The category labels for `P1997` are `C1` for ranked first, `C2` for ranked second, `C3` for ranked third, `C4` for ranked 4 or 5, `C5` for ranks 6 to 10, `C6` for ranks higher than 10 but still in business, and `Fail` if out of business.

1. Does it appear that market leadership has endured over this period? That is, are 1923 market position and 1997 market position independent or dependent? Does any loglinear model other than the saturated model provide a good fit to these data?
2. The category “Fail” is different from the others because it represents companies that were no longer competing in 1997. What happens if you ignore all companies in this category and redo question 1?
3. Finally, consider collapsing all of the 1997 market position categories together, so that the “new” 1997 variable has only two categories, “Failed” and “Not Failed.” Is there a relationship between 1923 position and this variable? Does it depend on type of good? Hint: In R, you can create a factor with the values “Failed” and “NotFailed” as follows:

```
> mshare$F1997 <- factor(ifelse(mshare$P1997 == "Fail",
+                               "Failed", "NotFailed"))
```