

Psych 6136: Loglinear models tutorial

Survival on the *Titanic*

These exercises examine the fitting of various loglinear models to data about survival on the *Titanic*, a 4-way table giving the cross-classification of 2201 passengers and crew, according to

- Gender (G): M vs. F
- Age (A): Adult vs. Child
- Class (C): 1st, 2nd, 3rd, Crew
- Survival (S): Died vs. Survived

The data is obtained using `data(Titanic)`. The R code for this exercise is contained in the file `titanic-loglin.R` on the course web page. Some of this is shown inline below. You should experiment with other analyses, commands and plots.

1. One slight complication here is that there are 8 cells with zero frequencies. Four of these (male and female children in 1st and 2nd class who died) should be considered sampling zeros, but 4 (children among the crew) should probably be considered structural zeros (cells where data could not occur). In these analyses, you can treat these all as sampling zeros by adding a small number to each cell.

```
library(MASS)      # for loglm()
library(vcd)       # for mosaic, aka plot.loglm()
data(Titanic)
Titanic <- Titanic + 0.5 # adjust for 0 cells
```

2. It is natural to consider Survival as the natural response variable, and the remaining variables as explanatory. Therefore, all models should include the high-order term among age, gender and class. Therefore, the **minimal null model** is `[AGC][S]`, which asserts that survival is jointly independent of Age, Sex and Class. Fit this model, and obtain a mosaic plot. Interpret the pattern of the residuals in this mosaic plot.

```
titanic.mod1 <- loglm(~ (Class * Age * Sex) + Survived, data=Titanic)
titanic.mod1
plot(titanic.mod1, main="Model [AGC][S]")
```

3. Fit a main effects model for survival, `[AGC][AS][GS][CS]`, that includes an association of survival with each of age, gender and class. Is this an adequate fit? What does the pattern of residuals tell you about remaining associations?

```
titanic.mod2 <- loglm(~ (Class * Age * Sex) + Survived*(Class + Age + Sex),
  data=Titanic)
titanic.mod2
plot(titanic.mod2, main="Model [AGC][AS][GS][CS]")
```

Note that, rather than specifying a new model completely, you can use `update()` to add terms to an existing model. In the formula for `update()`, "." stands for "what was there before." The model `titanic.mod2` could be obtained more simply as

```
titanic.mod2 <- update(titanic.mod1, . ~ . + Survived*(Class+Age+Sex))
```

4. What model would you use to allow an interaction of Age and Gender in their effect on Survival? Fit this model as above, and obtain the mosaic plot. [Hint: update by adding `Survived*Age*Sex`, to obtain the model `titanic.mod3`.]

5. Compare the three models using `anova()`

```
anova(titanic.mod1, titanic.mod2, titanic.mod3, test="chisq")
```