

一种办法是绘制马赛克图 (mosaic plot)。在马赛克图中, 嵌套矩形面积正比于单元格频率, 其中该频率即多维列联表中的频率。颜色和/或阴影可表示拟合模型的残差值。更多图形细节可参考Meyer、Zeileis和Hornick的“The Strucplot Framework: Visualizing Multi-way Contingency Tables with vcd” (2006) 一文, 或者Michael Friendly的统计图形主页 (<http://datavis.ca>)。Steve Simon曾编辑过一个非常优秀的绘制马赛克图的概念教程, 可在<http://www.childrensmrcy.org/stats/definitions/mosaic.htm>获取。

vcd包中的mosaic()函数可以绘制马赛克图。(R基础安装中的mosaicplot()也可绘制马赛克图, 但我还是推荐vcd包, 因为它具有更多扩展功能。)以基础安装中的Titanic数据集为例, 它包含存活或者死亡的乘客数、乘客的船舱等级 (一等、二等、三等和船员)、性别 (男性、女性), 以及年龄层 (儿童、成人)。这是一个被充分研究过的数据集。利用如下代码, 你可以看到分类细节:

```
> ftable(Titanic)
      Survived No Yes
Class Sex   Age
1st  Male   Child    0  5
      Adult  118 57
      Female Child    0  1
      Adult   4 140
2nd  Male   Child    0 11
      Adult  154 14
      Female Child    0 13
      Adult   13 80
3rd  Male   Child    35 13
      Adult  387 75
      Female Child    17 14
      Adult   89 76
Crew Male   Child    0  0
      Adult  670 192
      Female Child    0  0
      Adult    3 20
```

mosaic()函数可按如下方式调用:

```
mosaic(table)
```

其中table是数组形式的列联表。另外, 也可用:

```
mosaic(formula, data=)
```

其中formula是标准的R表达式, data设定一个数据框或者表格。添加选项shade = TRUE将根据拟合模型的皮尔逊残差值对图形上色, 添加选项legend = TRUE将展示残差的图例。

例如, 使用:

```
library(vcd)
mosaic(Titanic, shade=TRUE, legend=TRUE)
```

和:

```
library(vcd)
mosaic(~Class+Sex+Age+Survived, data=Titanic, shade=TRUE, legend=TRUE)
```

都将生成图11-23。但表达式版本的代码可使你对图形中变量的选择和摆放拥有更多的控制权。

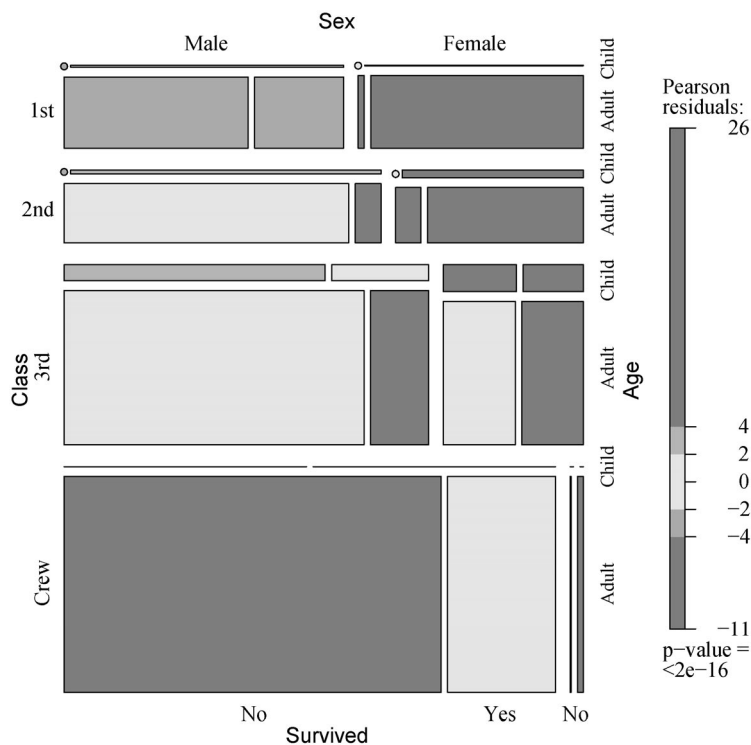


图11-23 按船舱等级、乘客性别和年龄层绘制的泰坦尼克号幸存者的马赛克图
(另见彩插图11-23)

马赛克图隐含着大量的数据信息。例如：(1)从船员到头等舱，存活率陡然提高；(2)大部分孩子都处在三等舱和二等舱中；(3)在头等舱中的大部分女性都存活了下来，而三等舱中仅有一半女性存活；(4)船员中女性很少，导致该组的Survived标签重叠（图底部的No和Yes）。继续观察，你将发现更多有趣的信息。关注矩形的相对宽度和高度，你还能发现那晚其他什么秘密吗？

扩展的马赛克图添加了颜色和阴影来表示拟合模型的残差值。在本例中，蓝色阴影表明，在假定生存率与船舱等级、性别和年龄层无关的条件下，该类别下的生存率通常超过预期值。红色阴影则含义相反。一定要运行该例子的代码，这样你可以真实感受下着色图形的效果。图形表明，在模型的独立条件下，头等舱女性存活数和男性船员死亡数超过模型预期值。如果存活数与船舱等级、性别和年龄层独立，三等舱男性的存活数比模型预期值低。尝试运行`example(mosaic)`，可以了解更多马赛克图的细节。

11.5 小结

本章中，我们学习了许多展示两个或更多变量间关系的图形方法，包括二维和三维散点图、散点图矩阵、气泡图、折线图、相关系数图 and 马赛克图。其中一些方法是标准的图形方法，而有