

交互性地对图中的点进行鉴别。从图形上看，观测点“Wheaties Honey Gold”和“Wheaties”异常，数据集似乎违反了多元正态性。可以删除这两个点再重新分析。

方差-协方差矩阵同质性即指各组的协方差矩阵相同，通常可用Box's M检验来评估该假设。由于R中没有Box's M函数，可以通过网络搜索找到合适的代码。另外，该检验对正态性假设很敏感，会导致在大部分案例中直接拒绝同质性假设。也就是说，对于这个重要的假设的检验，我们目前还没有一个好方法[但是可以参考Anderson（2006）和Silva et al.（2008）提供的一些有趣的备选方法，虽然在R中还没有实现]。

最后，还可以使用mvoutlier包中的ap.plot()函数来检验多元离群点。代码如下：

```
library(mvoutlier)
outliers <- ap.plot(y)
outliers
```

自己尝试一下，看看会得到什么结果吧！

## 9.7.2 稳健多元方差分析

如果多元正态性或者方差-协方差均值假设都不满足，又或者你担心多元离群点，那么可以考虑用稳健或非参数版本的MANOVA检验。稳健单因素MANOVA可通过rrcov包中的Wilks.test()函数实现。vegan包中的adonis()函数则提供了非参数MANOVA的等同形式。代码清单9-10是Wilks.test()的应用。

代码清单9-10 稳健单因素MANOVA

```
library(rrcov)
> Wilks.test(y,shelf,method="mcd")

Robust One-way MANOVA (Bartlett Chi2)

data: x
Wilks' Lambda = 0.511, Chi2-Value = 23.71, DF = 4.85, p-value =
0.0002143
sample estimates:
  calories    fat sugars
1      120 0.701    5.66
2      128 1.185   12.54
3      161 1.652   10.35
```

从结果来看，稳健检验对离群点和违反MANOVA假设的情况不敏感，而且再一次验证了存储在货架顶部、中部和底部的谷物营养成分含量不同。

## 9.8 用回归来做 ANOVA

在9.2节中，我们提到ANOVA和回归都是广义线性模型的特例。因此，本章所有的设计都可以用lm()函数来分析。但是，为了更好地理解输出结果，需要弄明白在拟合模型时，R是如何处理类别型变量的。

以9.3节的单因素ANOVA问题为例，即比较五种降低胆固醇药物治疗法（trt）的影响。

```
> library(multcomp)
> levels(cholesterol$trt)

[1] "1time" "2times" "4times" "drugD" "drugE"

首先，用aov()函数拟合模型：

> fit.aov <- aov(response ~ trt, data=cholesterol)
> summary(fit.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt	4	1351.37	337.84	32.433	9.819e-13 ***
Residuals	45	468.75	10.42		

现在，用lm()函数拟合同样的模型。结果见代码清单9-11。

#### 代码清单9-11 解决9.3节ANOVA问题的回归方法

```
> fit.lm <- lm(response ~ trt, data=cholesterol)
> summary(fit.lm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.782	1.021	5.665	9.78e-07 ***
trt2times	3.443	1.443	2.385	0.0213 *
trt4times	6.593	1.443	4.568	3.82e-05 ***
trtdrugD	9.579	1.443	6.637	3.53e-08 ***
trtdrugE	15.166	1.443	10.507	1.08e-13 ***

Residual standard error: 3.227 on 45 degrees of freedom  
Multiple R-squared: 0.7425, Adjusted R-squared: 0.7196  
F-statistic: 32.43 on 4 and 45 DF, p-value: 9.819e-13

我们能发现什么？因为线性模型要求预测变量是数值型，当lm()函数碰到因子时，它会用一系列与因子水平相对应的数值型对照变量来代替因子。如果因子有k个水平，将会创建k-1个对照变量。**R提供了五种创建对照变量的内置方法（见表9-6），你也可以自己重新创建（此处不做介绍）。默认情况下，对照处理用于无序因子，正交多项式用于有序因子。**

表9-6 内置对照

对照变量创建方法	描 述
contr.helmert	第二个水平对照第一个水平，第三个水平对照前两个的均值，第四个水平对照前三个的均值，以此类推
contr.poly	基于正交多项式的对照，用于趋势分析（线性、二次、三次等）和等距水平的有序因子
contr.sum	对照变量之和限制为0。也称作偏差找对，对各水平的均值与所有水平的均值进行比较
contr.treatment	各水平对照基线水平（默认第一个水平）。也称作虚拟编码
contr.SAS	类似于contr.treatment，只是基线水平变成了最后一个水平。生成的系数类似于大部分SAS过程中使用的对照变量

以对照（treatment contrast）为例，因子的第一个水平变成了参考组，随后的变量都以它为

标准。可以通过 `contrasts()` 函数查看它的编码过程。

```
> contrasts(cholesterol$trt)
      2times 4times drugD drugE
1time      0      0      0      0
2times      1      0      0      0
4times      0      1      0      0
drugD       0      0      1      0
drugE       0      0      0      1
```

若患者处于drugD条件下，变量drugD等于1，其他变量2times、4times和drugE都等于0。无需列出第一组的变量值，因为其他四个变量都为0，这已经说明患者处于1time条件。

在代码清单9-11中，变量trt2times表示水平1time和2times的一个对照。类似地，trt4times是1time和4times的一个对照，其余以此类推。从输出的概率值来看，各药物条件与第一组（1time）显著不同。

通过设定contrasts选项，可以修改lm()中默认的对照方法。例如，若使用Helmert对照：

```
fit.lm <- lm(response ~ trt, data=cholesterol, contrasts="contr.helmert")
```

还能通过 `options()` 函数修改R会话中的默认对照方法，例如，

```
options(contrasts = c("contr.SAS", "contr.helmert"))
```

将设定无序因子默认对比方法为contr.SAS，有序因子的为contr.helmert。虽然我们一直都是在线性模型范围中讨论对照方法的使用，但是在R中，你完全可以将其应用到其他模型中，包括第13章将会介绍的广义线性模型。

## 9.9 小结

本章中，我们回顾了基本实验和准实验设计的分析方法，包括ANOVA/ANCOVA/MANOVA。然后通过组内和组间设计的示例介绍了基本方法的使用，如单因素ANOVA、单因素ANCOVA、双因素ANOVA、重复测量ANOVA和单因素MANOVA。

除了这些基本分析，我们还回顾了模型的假设检验，以及应用多重比较过程来做综合检验的方法。最后，对各种结果可视化方法也进行了探索。如果你对用R分析DOE(Design Of Experiment, 实验设计)感兴趣，请参阅“CRAN Task View: Design of Experiments (DoE) & Analysis of Experimental Data”<sup>①</sup>中Groemping(2009)提供的方法。

第8章和第9章已经涵盖了各领域研究者常用的统计方法。在下一章中，我们将介绍功效分析。功效分析可以帮助我们在给定置信度的情况下，判断达到要求效果所需的样本大小，这一点对于研究设计非常重要。

<sup>①</sup> CRAN上实验设计的Task View页面地址为<http://cran.r-project.org/web/views/ExperimentalDesign.html>。——译者注