

齿长度增长相同。三种绘图方法中,我更推荐R包中的`interaction2wt()`函数,因为它能展示任意复杂度设计(双因素方差分析、三因素方差分析等)的主效应(箱线图)和交互效应。

此处没有涵盖模型假设检验和均值比较的内容,因为它们只是之前方法的一个自然扩展而已。而且,该设计是均衡的,故而不担心效应顺序的影响。

## 9.6 重复测量方差分析

所谓重复测量方差分析,即受试者被测量不止一次。本节重点关注含一个组内和一个组间因子的重复测量方差分析(这是一个常见的设计)。示例来源于生理生态学领域,研究方向是生命系统的生理和生化过程如何响应环境因素的变异(此为应对全球变暖的一个非常重要的研究领域)。基础安装包中的`co2`数据集包含了北方和南方牧草类植物 *Echinochloa crus-galli* (Potvin, Lechowicz, Tardif, 1990) 的寒冷容忍度研究结果,在某浓度二氧化碳的环境中,对寒带植物与非寒带植物的光合作用率进行了比较。研究所用植物一半来自于加拿大的魁北克省,另一半来自美国的密西西比州。

首先,我们关注寒带植物。因变量是二氧化碳吸收量(`uptake`),单位为`ml/L`,自变量是植物类型`Type`(魁北克VS密西西比州)和七种水平( $95\sim 1000\text{ }\mu\text{mol/m}^2\text{ sec}$ )的二氧化碳浓度(`conc`)。另外,`Type`是组间因子,`conc`是组内因子。分析过程见代码清单9-7。

代码清单9-7 含一个组间因子和一个组内因子的重复测量方差分析

```
> wlb1 <- subset(CO2, Treatment=='chilled')
> fit <- aov(uptake ~ conc*Type + Error(Plant/(conc)), wlb1)
> summary(fit)

Error: Plant
      Df Sum Sq Mean Sq F value    Pr(>F)
Type    1 2667.24  2667.24   60.414 0.001477 **
Residuals  4   176.60    44.15
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Plant:conc
      Df Sum Sq Mean Sq F value    Pr(>F)
conc    1  888.57   888.57  215.46 0.0001253 ***
conc:Type 1  239.24   239.24   58.01 0.0015952 **
Residuals  4    16.50     4.12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
      Df Sum Sq Mean Sq F value    Pr(>F)
Residuals 30  869.05    28.97

> par(las=2)
> par(mar=c(10,4,4,2))
> with(wlb1, interaction.plot(conc,Type,uptake,
                             type="b", col=c("red","blue"), pch=c(16,18),
                             main="Interaction Plot for Plant Type and Concentration"))
```

```
> boxplot(uptake ~ Type*conc, data=w1b1, col=(c("gold", "green")),
  main="Chilled Quebec and Mississippi Plants",
  ylab="Carbon dioxide uptake rate (umol/m^2 sec)")
```

方差分析表表明在0.01的水平下，主效应类型和浓度以及交叉效应类型x浓度都非常显著，图9-9中通过interaction.plot()函数展示了交互效应。

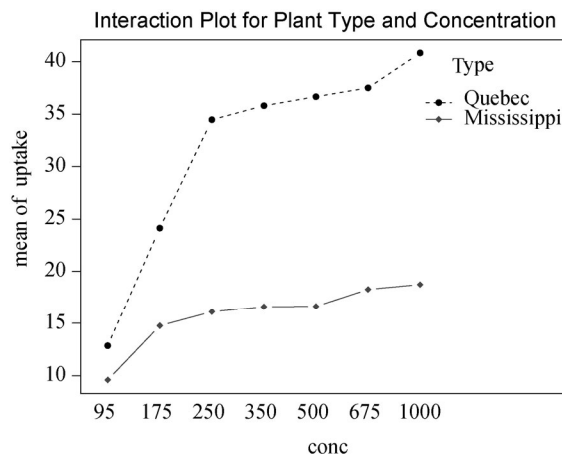


图9-9 CO<sub>2</sub>浓度和植物类型对CO<sub>2</sub>吸收的交互影响。图形由interaction.plot()函数绘制

若想展示交互效应其他不同的侧面，可以使用boxplot()函数对相同的数据画图，结果见图9-10。

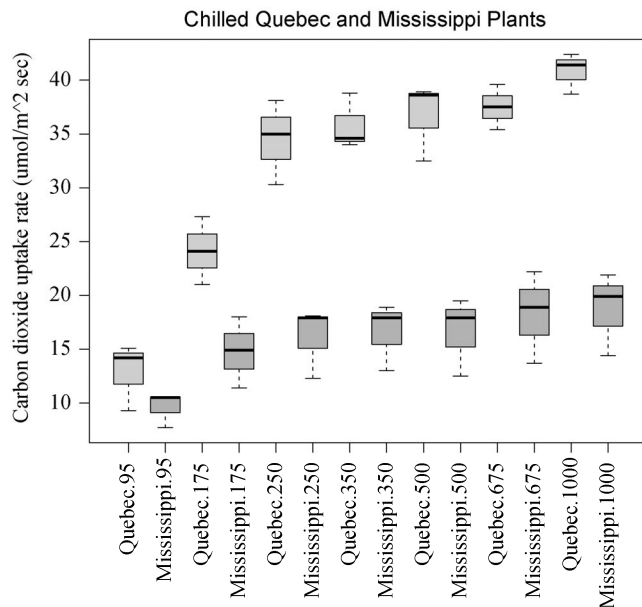


图9-10 CO<sub>2</sub>浓度和植物类型对CO<sub>2</sub>吸收的交互效应。图形由boxplot()函数绘制

从以上任意一幅图都可以看出，魁北克省的植物比密西西比州的植物二氧化碳吸收率高，而且随着CO<sub>2</sub>浓度的升高，差异越来越明显。

**注意** 通常处理的数据集是宽格式 (wide format)，即列是变量，行是观测值，而且一行一个受试对象。9.4节中的litter数据框就是一个很好的例子。不过在处理重复测量设计时，需要有长格式 (long format) 数据才能拟合模型。在长格式中，因变量的每次测量都要放到它独有的行中，CO<sub>2</sub>数据集即该种形式。幸运的是，5.6.3节的reshape包可方便地将数据转换为相应的格式。

### 混合模型设计的各种方法

在分析本节关于CO<sub>2</sub>的例子时，我们使用了传统的重复测量方差分析。该方法假设任意组内因子的协方差矩阵为球形，并且任意组内因子两水平间的方差之差都相等。但在现实中这种假设不可能满足，于是衍生了一系列备选方法：

- 使用lme4包中的lmer()函数拟合线性混合模型 (Bates, 2005)；
- 使用car包中的Anova()函数调整传统检验统计量以弥补球形假设的不满足 (例如 Geisser-Greenhouse校正)；
- 使用nlme包中的gls()函数拟合给定方差-协方差结构的广义最小二乘模型 (UCLA, 2009)；
- 用多元方差分析对重复测量数据进行建模 (Hand, 1987)。

以上方法已超出本书范畴，如果你对 these 方法感兴趣，可以参考Pinheiro & Bates (2000)、Zuur et al. (2009)。

目前为止，本章都只是对单个因变量的情况进行分析，在下一节，我们将简略介绍多个结果变量的设计。

## 9.7 多元方差分析

当因变量 (结果变量) 不止一个时，可用多元方差分析 (MANOVA) 对它们同时进行分析。以MASS包中的UScereal数据集为例[Venables, Ripley (1999)]，我们将研究美国谷物中的卡路里、脂肪和糖含量是否会因为储存架位置的不同而发生变化。其中1代表底层货架，2代表中层货架，3代表顶层货架。卡路里、脂肪和糖含量是因变量，货架是三水平 (1、2、3) 的自变量。分析过程见代码清单9-8。

### 代码清单9-8 单因素多元方差分析

```
> library(MASS)
> attach(UScereal)
> y <- cbind(calories, fat, sugars)
> aggregate(y, by=list(shelf), FUN=mean)
```