Introdução aos experimentos fatoriais

Importância e especificação

Prof. Walmes Zeviani

walmes@ufpr.br

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná

Atualizado em 2018-08-21

Plano de aula

Justificativa

- A maioria dos fenômenos geralmente depende de vários fatores.
- Estudar os fatores isoladamente pode dificultar compreender e otimizar processos.
- Experimentos fatoriais são importantes pois permitem estudar simultâneamente mais de um fator.

Objetivos

- Introduzir principais os conceitos sobre experimentos fatoriais.
- Fazer a expecificação do modelo e quadro de análise de variância.

Experimentos fatoriais

- Experimentos fatoriais: estudo de **mais de um fator ao mesmo tempo**.
- Níveis dos fatores combinados formam as **celas ou pontos experimentais**.
- As formas de combinar os fatores se chama arranjo.
- Cada arranjo recebe uma classficação.
- OS efeitos dos fatores podem apresentar **interação**.

Tipos de arranjo fatorial

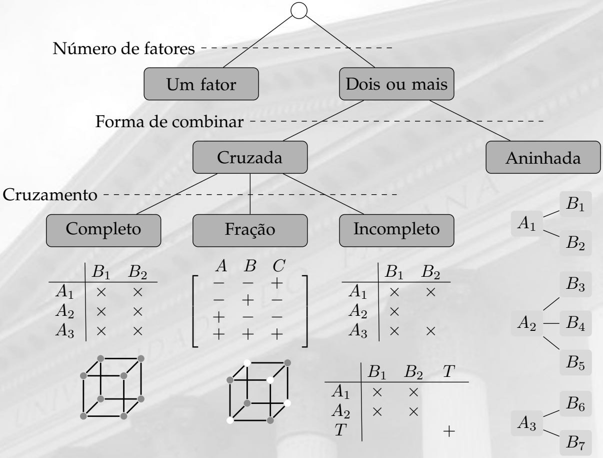


Figura 1. Tipos de arranjos fatoriais comuns em experimentos planejados.

Especificação do modelo

Considere um experimento fatorial duplo completamente cruzado com r repetições de cada ponto experimental, denotado por A $\times$ B, ou seja, o fator A com A níveis e um fator B com B níveis com a presença de todas as combinações. Os pontos experimentais são os seguintes:

$$\begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_B \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_B \\ \vdots & \vdots & \ddots & \vdots \\ a_A b_1 & a_A b_2 & \cdots & a_A b_B \end{bmatrix}. \tag{1}$$

O modelo estatístico

O modelo matemático associado ao experimento é

$$y_{abr} = \mu + \alpha_a + \beta_b + \gamma_{ab} + \epsilon_{abr}, \quad \epsilon_{abr} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2). \tag{2}$$

- y_{abr} é o valor observado na combinação r-ésima combinação ab dos fatores A e B.
- μ é uma constante que incide em todas as observações e representa a média da resposta sob a ausência de efeito dos pontos experimentais.
- α_a é o efeito do a-ésimo nível do fator A.
- β_b é o efeito do b-ésimo nível do fator B.
- γ_{ab} é o efeito da interação entre nível a e b dos fatores.
- ϵ_{ijr} é o erro experimental, assumido ser independente e ter distribuição normal de média 0 e variância comum.

A matriz do modelo

- A matriz do modelo contém colunas associadas a cada um dos fatores (cor de fundo).
- Com o contraste de zerar o primeiro nível, a primeira coluna é removida (colunas em vermelho).

$$
X = \begin{array}{c}
\begin{array}{cccccccccccc} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \beta_1 & \beta_2 & \gamma_{11} & \gamma_{21} & \gamma_{31} & \gamma_{12} & \gamma_{22} & \gamma_{32} \end{array} \\
\begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
=
\begin{array}{c}
\begin{array}{cccccc} \mu & \alpha_2 & \alpha_3 & \beta_2 & \gamma_{22} & \gamma_{32} \end{array} \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1
\end{bmatrix}
\end{array}
$$

X_μ

$X_{\mu:\alpha}$

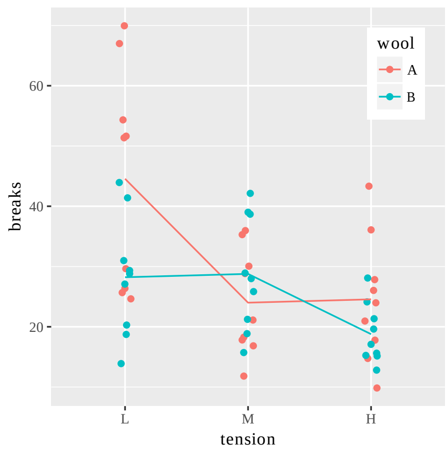
$X_{\mu:\beta}$

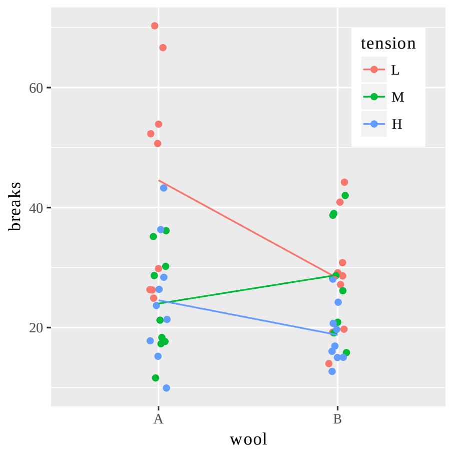
$X_{\mu:\gamma}$

Um exemplo de experimento fatorial

```r
# data(warpbreaks)
# str(warpbreaks)
library(ggplot2)

gg1 <- ggplot(data = warpbreaks,
              mapping = aes(x = tension, y = breaks, color = wool)) +
    geom_jitter(width = 0.05) +
    stat_summary(mapping = aes(group = wool), geom = "line", fun.y = mean) +
    theme(legend.position = c(0.95, 0.95),
          legend.justification = c(1, 1))

gg2 <- ggplot(data = warpbreaks,
              mapping = aes(x = wool, y = breaks, color = tension)) +
    geom_jitter(width = 0.05) +
   stat_summary(mapping = aes(group = tension), geom = "line", fun.y = mean) +
    theme(legend.position = c(0.95, 0.95),
          legend.justification = c(1, 1))

gridExtra::grid.arrange(gg1, gg2, nrow = 1)
```

Um exemplo de experimento fatorial





Questões importantes

- Quantos níveis cada fator tem?
- Quantos pontos experimentais existem?
- Quantas repetições para cada ponto experimental?
- Existe efeito de `wool`?
- Existe efeito de `tension`?
- Existe interação entre `wool` e `tension`?

Análise de variância

```
m0 <- lm(breaks ~ tension * wool, data = warpbreaks)
anova(m0)


## Analysis of Variance Table
##
## Response: breaks
##              Df Sum Sq Mean Sq F value    Pr(>F)
## tension       2 2034.3 1017.13  8.4980 0.0006926 ***
## wool          1  450.7  450.67  3.7653 0.0582130 .
## tension:wool  2 1002.8  501.39  4.1891 0.0210442 *
## Residuals    48 5745.1  119.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# summary(m0)
```

Análise de variância

```r
# Matrizes.
y <- cbind(m0$model[, "breaks"])
X <- model.matrix(m0)
a <- m0$assign

# Partições crescentes da matriz X.
X_0 <- cbind(X[, a <= 0])
X_01 <- cbind(X[, a <= 1])
X_02 <- cbind(X[, a <= 2])
X_03 <- cbind(X[, a <= 3])

# Funções.
proj <- function(X) X %*% solve(t(X) %*% X) %*% t(X)
tr <- function(X) sum(diag(X))

# Matrizes de projeção.
H_0 <- proj(X_0)
H_01 <- proj(X_01)
H_02 <- proj(X_02)
H_03 <- proj(X_03)
I <- diag(nrow(X))
```

Análise de variância

```r
P <- list("resíduo" = I - H_03,
          "A:B"     = H_03 - H_02,
          "B"       = H_02 - H_01,
          "A"       = H_01 - H_0)

# anova(m0)
t(sapply(P,
         FUN = function(p) {
             c(DF = tr(p), SQ = t(y) %*% p %*% y)
         }))


##          DF        SQ
## resíduo 48 5745.1111
## A:B      2 1002.7778
## B        1  450.6667
## A        2 2034.2593
```

Próxima aula

- Mais sobre experimentos fatoriais.
- Introdução aos fatoriais 2^k.

Avisos

- Sabatina 03 disponível no Moodle!