

# MAT02014 - Planejamento de Experimentos II

## Experimentos com múltiplos fatores

Rodrigo Citton P. dos Reis  
[rodrigocpdosreis@gmail.com](mailto:rodrigocpdosreis@gmail.com)

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática e Estatística  
Departamento de Estatística

Porto Alegre, 2018

# Delineamentos fatoriais com múltiplos fatores

# Introdução

- Vimos que delineamentos fatoriais de dois fatores são mais eficientes do que estudar cada fator separadamente em projetos de um fator.
- Da mesma forma, quando muitos fatores estão em estudo, é mais eficiente estudá-los juntos em um planejamento fatorial **multifatorial** do que estudá-los separadamente em grupos de dois usando delineamentos fatoriais de dois fatores.
- Quando múltiplos fatores são estudados simultaneamente, o poder para detectar efeitos principais é maior sobre o que seria em experimentos fatoriais separados de dois fatores.
- Além disso, a possibilidade de detectar interações entre qualquer um dos fatores é necessária.

qualquer um dos fatores é possível.

## Introdução

- Se os fatores fossem estudados separadamente (em delineamentos fatoriais de dois fatores), interações de dois fatores só poderiam ser detectadas entre fatores estudados juntos no mesmo desenho.
- Em um fatorial multifatorial, não só é possível detectar interações de dois fatores entre qualquer par de fatores, mas também é possível detectar **interações de ordem superior** entre grupos de fatores.
- Uma interação de três fatores entre os fatores  $A$ ,  $B$  e  $C$ , por exemplo, significa que o efeito do fator  $A$  difere dependendo da combinação dos níveis dos fatores  $B$  e  $C$ .

# Introdução

- As combinações de **tratamento** em um fatorial multifatorial consistem em todas as combinações possíveis dos níveis de todos os fatores.
- **Pergunta 1:** Faça a combinação dos níveis de 3 fatores, no caso em que cada um destes apresenta dois níveis.
- **Pergunta 2:** Como você faria isto no seu software?
- O modelo para análise é uma extensão do modelo proposto para delineamentos fatoriais de dois fatores, e a análise pode ser feita usando a função `lm` do R similar aos exemplos mostrados anteriormente.

# Exemplo

- Considere um exemplo de um planejamento fatorial multifatorial em pesquisa de *marketing*.
- Uma empresa cujas vendas são feitas *on-line* por meio de uma página da *Web* gostaria de aumentar a proporção de visitantes de seu site que se inscrevem em seus serviços configurando de forma ideal sua página da *web*.

# Exemplo

- Para comprar na empresa, os clientes devem se inscrever e preencher um formulário fornecendo seu endereço de *e-mail* junto com outros campos necessários.
- Uma vez que um cliente se inscreve, a empresa tem informações de contato para seu banco de dados e pode enviar por *e-mail* com propagandas, ofertas especiais e assim por diante.
- A empresa gostaria de experimentar testando diferentes configurações de sua página da *Web* para ver se eles podem aumentar o número de visitantes do site que realmente se inscreveram.



# Exemplo

- As unidades experimentais neste estudo serão os indivíduos que visitam o site da empresa.
- A resposta é **binária**: o cliente ou **se inscreve** ou **não se inscreve**.
- Os fatores em estudo são características que alteram a aparência da página da *Web*.
  - O **fator  $A$**  representa as alternativas de segundo plano (plano de fundo) para a página com três opções.
  - O **fator  $B$**  é o tamanho da fonte na faixa principal, com três níveis.
  - O **fator  $C$**  é a cor do texto, com duas alternativas.
  - O **fator  $D$**  é uma escolha entre um botão de inscrição ou um link.

# Exemplo

- Com base nesses fatores, existem  $3 \times 3 \times 2 \times 2 = 36$  possíveis configurações da página da Web ao considerar todas as combinações possíveis dos níveis de cada fator.
- Um experimento fatorial de **quatro fatores** consiste em atribuir aleatoriamente visitantes ao site a uma das possíveis configurações e registrar sua resposta binária.
- Existem variáveis ocultas que podem afetar a chance de um visitante do site se inscrever.
  - Por exemplo, a ordem de posição em que o link (para o site da empresa) aparece em uma pesquisa na Web para os produtos que eles vendem, promoções oferecidas pelos concorrentes e atratividade dos sites dos concorrentes.
- A designação aleatória de cada visitante sequencial do site

para uma das configurações alternativas em estudo deve minimizar a chance de viés de alterações nas variáveis ocultas ao longo do tempo.

# *Das model*



## *Das model*

- A probabilidade de um visitante do site se inscrever pode ser expressa pelo modelo:

$$p_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \delta_l + \alpha\delta_{il} + \beta\delta_{jl} + \alpha\beta\delta_{ijl} + \gamma\delta_{kl} + \alpha\gamma\delta_{ikl} + \beta\gamma\delta_{jkl} + \alpha\beta\gamma\delta_{ijkl},$$

- $\alpha_i$  representa o efeito do plano de fundo;
- $\beta_j$  representa o efeito do tamanho da fonte na faixa principal;
- $\gamma_k$  representa o efeito da cor do texto;
- $\delta_l$  representa o efeito do botão vs o link para inscrição.

# Exemplo

- O experimento foi conduzido através da construção de 36 sites, consistindo de todas as combinações possíveis dos quatro fatores descritos.
- Cada cliente em potencial que visitou o site da empresa durante o período de teste foi redirecionado **aleatoriamente** para uma das 36 configurações.
- O número de visitantes  $n_{ijkl}$  da versão  $ijkl$  e o número que se inscreveu  $x_{ijkl}$  foram registrados.
- O  $x_{ijkl}$  é então distribuído conforme uma distribuição binomial:

$$B(x_{ijkl}, n_{ijkl}, p_{ijkl}) = \binom{n_{ijkl}}{x_{ijkl}} p_{ijkl}^{x_{ijkl}} (1 - p_{ijkl})^{(n_{ijkl} - x_{ijkl})}$$

# Exemplo (R)

```
library(daewr)  
data(web)  
head(web)
```

```
##      A B C D visitors  signup  
## 1 1 1 1 1      1016      22  
## 2 1 1 1 2      1145      16  
## 3 1 1 2 1      1145      17  
## 4 1 1 2 2      1082      19  
## 5 1 2 1 1      1121      28  
## 6 1 2 1 2      1103      28
```

# Exemplo

- Poderíamos tentar transformar os dados (as proporções amostrais).
- Uma alternativa mais elegante, e com resultados teóricos interessantes é o uso do **modelo logístico**.
- O modelo logístico é um modelo que faz parte da classe dos **modelos lineares generalizados (MLG ou GLM)**.

(Pausa)

- Para avaliarmos as hipóteses usuais, utilizaremos a **análise do desvio** (*Analys of Deviance*; ANODEV).
  - É uma generalização da análise de variância para os MLG.



# Exemplo (R)

- **Sua vez:** ao nível de 5% de significância, o que podemos falar dos efeitos principais? E dos efeitos de interação?

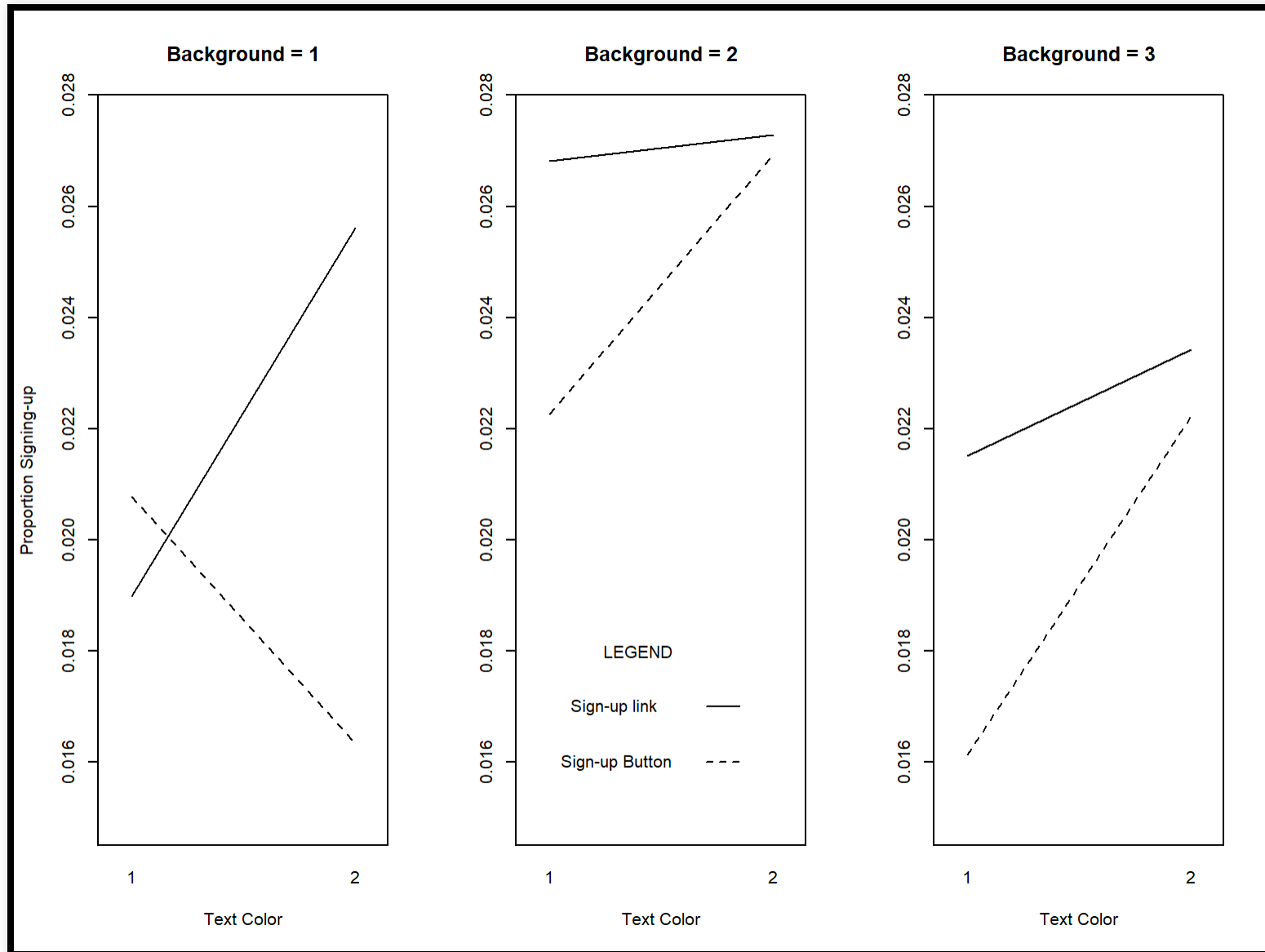
```
modb <- glm( cbind(signup, visitors-signup) ~ A * B * C * D,  
             data = web, family = binomial )  
anova(update(modb, .~ A+B + A:B + C + A:C + B:C + A:B:C + D + A:D + B:D  
            test = "Chisq")
```

```
## Analysis of Deviance Table  
##  
## Model: binomial, link: logit  
##  
## Response: cbind(signup, visitors - signup)  
##  
## Terms added sequentially (first to last)  
##  
##  
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)  
## NULL                35      56.641  
## A                 2    10.2540    33     46.387 0.005934 **  
## B                 2     6.9301    31     39.457 0.031272 *  
## C                 1     2.9632    30     36.493 0.085182 .  
## D                 1     4.6165    29     31.877 0.031666 *  
## A:B              4     6.3093    25     25.568 0.177213  
## A:C              2     0.6821    23     24.886 0.711026  
## B:C              2     4.0303    21     20.855 0.133300
```

# Exemplo (R)

```
prop <- web$signup / web$visitors
webp <- data.frame(web, prop)
par(mfrow = c(1,3))
webp1 <- subset(webp, A == 1)
interaction.plot(webp1$C, webp1$D, webp1$prop, type = "l",
                 legend=FALSE, ylim = c(.015,.0275), main = "Background = ",
                 xlab = "Text Color", ylab = "Proportion Signing-up")
webp2 <- subset(webp, A == 2 )
interaction.plot( webp2$C, webp2$D, webp2$prop, type = "l", legend = FALSE,
                 xlab = "Text Color", ylab = " ")
lines( c(1.7,1.85), c(.016,.016), lty = 2)
lines( c(1.7,1.85), c(.017,.017) ,lty = 1)
text(1.3, .017, "Sign-up link ")
text(1.3, .016, "Sign-up Button" )
text(1.4, .018, "LEGEND" )
webp3 <- subset(webp, A == 3)
interaction.plot(webp3$C, webp3$D, webp3$prop, type = "l",
                 legend=FALSE, ylim = c(.015,.0275), main="Background = ")
```

# Exemplo (R)



# Conclusões

- Ao usar o tipo de segundo plano 2 ou tipo 3, pode-se observar que a alteração da cor do texto de preto = 1 para branco = 2 causa um aumento na proporção de visitantes do site que se inscrevem.
- O aumento (representado pela inclinação das linhas) é maior quando um botão de inscrição é usado, em vez de um link de inscrição, porque a taxa de inscrição geral é maior quando um link é usado, independentemente da cor do texto.
  - Não há muito espaço para melhorias.
- No entanto, quando o tipo de plano de fundo 1 é usado, o efeito da cor do texto é completamente diferente.
  - Nesse caso, alterar a fonte de preto = 1 para branco = 2 causa uma diminuição na proporção quando um botão

de inscrição é usado, e há um grande aumento na proporção de inscritos ao mudar de fonte preta para branca ao usar um link de inscrição.

- Este é o oposto dos efeitos vistos para os tipos de fundo 2 e 3.

# Para casa (para aula)

- Construa a tabela de análise de variância para o caso multifatorial com resposta normal.
- Discuta outras aplicações que você considera que o delineamento fatorial com múltiplos fatores é apropriado.

