

Profile Analysis

Podemos usá-la quando realizamos p tratamentos (ex.: perguntas de um questionário) em 2 ou mais grupos (populações). Todas as respostas devem estar na mesma unidade—por exemplo, uma escala de 1 até 5. Além disso, assume-se que as respostas de um grupo são independentes das respostas dos outros grupos.

Sejam $\mu_1 = [\mu_{11}, \mu_{12}, \dots, \mu_{1p}]$ e $\mu_2 = [\mu_{21}, \mu_{22}, \dots, \mu_{2p}]$ as médias dos dois grupos p/ os p tratamentos. O gráfico destas médias de um grupo, ligadas por segmentos de reta, é chamado *profile*. A análise compara estas médias em 3 etapas:

1. Os perfis são paralelos?

$$H_0^{(1)} : \mu_{1,i} - \mu_{1,i-1} = \mu_{2,i} - \mu_{2,i-1}, \text{ para } i = 2, 3, \dots, p.$$

2. Assumindo que os perfis são paralelos, eles são também coincidentes?

$$H_0^{(2)} : \mu_{1,i} = \mu_{2,i}, \text{ para } i = 1, 2, \dots, p.$$

3. Assumindo que os perfis são coincidentes, todas as médias dos tratamentos são iguais?

$$H_0^{(3)} : \mu_{11} = \mu_{12} = \dots = \mu_{1p} = \mu_{21} = \mu_{22} = \dots = \mu_{2p}.$$

Em forma matricial, podemos escrever $H_0^{(1)} : C\mu_1 = C\mu_2$, onde C vale

$$C_{(p-1) \times p} = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

As diferenças têm médias amostrais $C\bar{x}_1$ e $C\bar{x}_2$. Assumindo variâncias iguais nas duas populações, a variância amostral das diferenças vale CS_pC' , onde

$$\begin{aligned} S_p &= \frac{n_1 - 1}{n_1 + n_2 - 2} S_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2, \\ S_1 &= \frac{\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)'}{n_1 - 1}, \\ S_2 &= \frac{\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)(x_{2j} - \bar{x}_2)'}{n_2 - 1}. \end{aligned}$$

S_1 e S_2 são as matrizes de variâncias e covariâncias das populações 1 e 2, respectivamente.

Portanto, o vetor de diferenças da população 1 tem distribuição $N_{p-1}(C\mu_1, C\Sigma C')$ na população 1, e distribuição $N_{p-1}(C\mu_2, C\Sigma C')$ na população 2.

Rejeitamos $H_0^{(1)} : C\mu_1 = C\mu_2$ (perfis paralelos) com confiança α quando

$$T^2 = (\bar{x}_1 - \bar{x}_2)' C' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) CS_p C' \right]^{-1} C(\bar{x}_1 - \bar{x}_2) > c^2,$$

onde

$$c^2 = \frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} F_{p-1, n_1 + n_2 - p}(\alpha).$$

Assumindo que os perfis são paralelos, para verificar que são coincidentes basta testar a hipótese de que as somas

das médias na população 1 é igual a soma das médias na população 2, ou seja, $1'\mu_1 = \mu_{11} + \mu_{12} + \dots + \mu_{1p} = \mu_{21} + \mu_{22} + \dots + \mu_{2p} = 1'\mu_2$. Podemos então usar um teste t univariado.

Rejeitamos $H_0^{(2)} : 1'\mu_1 = 1'\mu_2$ (perfis coincidentes) com confiança α quando

$$\begin{aligned} T^2 &= 1'(\bar{x}_1 - \bar{x}_2) \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) 1'S_p 1 \right]^{-1} 1'(\bar{x}_1 - \bar{x}_2) \\ &= \left(\frac{1'(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) 1'S_p 1}} \right)^2 \\ &> t_{n_1 + n_2 - 2}^2(\alpha/2) = F_{1, n_1 + n_2 - 2}(\alpha). \end{aligned}$$

Se as médias são coincidentes, podemos usar os pontos das duas populações para obter a média amostral

$$\bar{x} = \frac{\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_2} x_{2j}}{n_1 + n_2} = \frac{n_1}{n_1 + n_2} \bar{x}_1 + \frac{n_2}{n_1 + n_2} \bar{x}_2.$$

Neste caso, resta verificar se as médias dos tratamentos são todas iguais, ou seja, $\mu_1 = \mu_2 = \dots = \mu_p$.

Rejeitamos $H_0^{(3)} : C\mu = 0$ (médias dos tratamentos iguais) com confiança α se

$$(n_1 + n_2) \bar{x}' C' [CSC']^{-1} C \bar{x} > c^2,$$

onde S é a matriz de covariâncias amostrais calculada a partir de todas as $n_1 + n_2$ observações, e

$$c^2 = \frac{(n_1 + n_2 - 1)(p - 1)}{n_1 + n_2 - p + 1} F_{p-1, n_1 + n_2 - p + 1}(\alpha).$$

Exemplo:

- x_1 = resposta em escala de 8 pontos p/ Questão 1
- x_2 = resposta em escala de 8 pontos p/ Questão 2
- x_3 = resposta em escala de 5 pontos p/ Questão 3
- x_4 = resposta em escala de 5 pontos p/ Questão 4

Escala de 8 pontos: (1) Extremamente negativa, (2) Muito negativa, (3) Moderadamente negativa, (4) Levemente negativa, (5) Levemente positiva, (6) Moderadamente positiva, (7) Muito positiva, e (8) Extremamente positiva.

Escala de 5 pontos: (1) Nada, (2) Só um pouco, (3) Algum, (4) Bastante, e (5) Muitíssimo.

População 1: $n_1 = 30$ homens casados.

População 2: $n_2 = 30$ mulheres casadas.

exemplo.png

```
> plot(x1,type="o",col="red")
> lines(x2,type="o",col="blue")
>
> require(MASS) % Para usar a função inverso 'ginv'
> p=4
> n1=30
> n2=30
> alpha=.05
> x1 = c(6.833, 7.033, 3.967, 4.700)
> x2 = c(6.633, 7.000, 4.000, 4.533)
> Sp = matrix(c(.606, .262, .066, .161, .262, .637,
               .173, .143, .066, .173, .810, .029,
               .161, .143, .029, .306), nrow=4)
> C = matrix(c(-1, 0, 0, 1, -1, 0, 0, 1, -1, 0, 0, 1),
             nrow=3)
> C
      [,1] [,2] [,3] [,4]
[1,]  -1    1    0    0
[2,]   0   -1    1    0
[3,]   0    0   -1    1
> T2 = t(x1-x2) %*% t(C) %*%
      ginv( (1/n1+1/n2) * C %*% Sp %*% t(C) ) %*%
      C %*% (x1-x2)
> T2
      [,1]
[1,] 1.00536
> c2 = (n1+n2-2)*(p-1)/(n1+n2-p)*
      qf(alpha,p-1,n1+n2-p,lower.tail=F)
> c2
[1] 8.605018
> T2 > c2
      [,1]
[1,] FALSE

> T2 = (sum(x1-x2)/sqrt((1/n1+1/n2)*sum(Sp)))^2
> T2
[1] 0.5016973
> qf(alpha,1,n1+n2-2,lower.tail=F)
[1] 4.006873
```

```
> T2 > qf(alpha,1,n1+n2-2,lower.tail=F)
      [,1]
[1,] FALSE
```

Não faz sentido testar se os tratamentos são iguais, pois foram medidos em escalar distintas.

Farei o teste apenas p/ ilustrar a ferramenta R.

Vou usar Sp no lugar de S.

```
> x = n1/(n1+n2)*x1 + n2/(n1+n2)*x2
```

```
> T2 = (n1+n2)*t(x) %*% t(C) %*%
      ginv(C %*% Sp %*% t(C)) %*% C %*% x
> T2
      [,1]
[1,] 927.8113
```

```
> c2 = (n1+n2-1)*(p-1)/(n1+n2-p+1)*
      qt(alpha,p-1,n1+n2-p+1,lower.tail=F)
> c2
[1] 722.0931
> T2 > c2
      [,1]
[1,] TRUE
```

Quando a amostra é pequena é necessário verificar se a população é normal antes de realizar os testes de hipótese.