4 Testes de Hipóteses

4.1 Introdução e generalidades

Contexto:

- População X segue distribuição de probabilidades que apresenta aspectos desconhecidos: estipula-se apenas que a sua distibuição é um elemento de uma classe $\mathcal{F}$
- No âmbito da inferência paramétrica, essa classe é indexada parametricamente: $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$
- Uma hipótese estatística é uma conjectura sobre aspectos desconhecidos da distribuição de X
- No caso paramétrico, tal conjectura traduz-se numa afirmação sobre o parâmetro que indexa $\mathcal{F}$

Exemplo 4.1 *Suponha-se que $\mathcal{F} = \{f : f \text{ é simétrica}\}$. Um exemplo de uma hipótese estatística neste contexto é a conjectura "f é gaussiana". Se $\mathcal{F} = \{N(\mu, 1) : \mu \in \mathbb{R}\}$, um exemplo de uma hipótese estatística é a conjectura de que $\mu = 3$.* ■

Neste capítulo trataremos exclusivamente de hipóteses paramétricas.

- Neste contexto, qualquer hipótese estatística induz uma partição no espaço paramétrico: $\Theta = \Theta_0 \cup \Theta_1$, com $\Theta_0 \cap \Theta_1 = \varnothing$, onde Θ_0 é o conjunto dos valores de θ que verificam a conjectura.
- **Hipótese nula** — $H_0 : \theta \in \Theta_0$
- **Hipótese alternativa** — $H_1 : \theta \in \Theta_1$
- Um teste de hipóteses corresponde a uma estratégia para averiguar qual das duas hipóteses é melhor suportada pelos dados observados $x_1, \ldots, x_n$
- O tratamento dado às hipóteses em confronto é assimétrico; tradicionalmente, a hipótese nula corresponde ao *status quo* que só será abandonado se se recolher evidência substancial
- Uma hipótese estatística diz-se **simples** se especifica completamente a distribuição da população X, ou seja, quando determina um único valor para o parâmetro desconhecido
- Hipóteses que não são simples dizem-se **compostas**

Exemplo 4.2 *Se $\mathcal{F} = \{N(\mu, 1) : \ \mu \in \mathbb{R}\}$, um exemplo de uma hipótese nula simples é $H_0 : \ \mu = 3$, que conduz a $\Theta_0 = \{3\}$. Um exemplo de uma hipótese nula composta é $H_0 : \mu > 3$, que corresponde a $\Theta_0 =]3, +\infty[$.* ■

Exemplo 4.3 *Pretende-se aferir se determinada moeda é equilibrada. O resultado do lançamento de uma moeda pode ser modelado por uma distribuição de Bernoulli de parâmetro* $\theta \in (0,1)$, *onde* θ *é, e.g., a probabilidade de se obter uma "face". A hipótese nula é então* $H_0 : \theta = 1/2$, *e a alternativa é* $H_1 : \theta \neq 1/2$. *Se a hipótese a testar fosse a de que a probabilidade de obter uma face é maior do que a de obter uma coroa, então teríamos as hipóteses* $H_0 : \theta > 1/2$ *contra* $H_1 : \theta \leq 1/2$. ■

Teste de hipóteses

- Um teste de hipóteses é uma regra que permite especificar um subconjunto do espaço amostral, $W \subset \mathcal{X}$, tal que
  - se $(x_1, \ldots, x_n) \in W$, rejeita-se H_0 (aceita-se H_1)
  - se $(x_1, \ldots, x_n) \notin W$, aceita-se H_0 (rejeita-se H_1)
- Ao conjunto W dá-se o nome de **região crítica** ou **região de rejeição**; o seu complementar, $\bar{W}$, é designado por **região de aceitação**
- Um teste estatístico induz então uma partição do espaço amostral $\mathcal{X} = W \cup \bar{W}$; $W \cap \bar{W} = \varnothing$
- Na maior parte dos casos práticos, a região de rejeição é tal que

$$(x_1, \ldots, x_n) \in W \Leftrightarrow T(x_1, \ldots, x_n) \in W_T$$

  para alguma estatística $T = T(X_1, \ldots, X_n)$ e conjunto W_T

- Nestas circunstâncias, a decisão de rejeitar ou não rejeitar H_0 é baseada exclusivamente no valor observado da estatística T, a chamada **estatística de teste**. Designa-se então W_T por região de rejeição ou região crítica

Em resumo, os ingredientes de um teste de hipóteses são:

- A hipótese nula, H_0, que usualmente é a posição aceite "até prova em contrário"
- A hipótese alternativa, H_1, que é adoptada se se recolher evidência para rejeitar H_0
- Uma estatística de teste $T = T(X_1, \ldots, X_n)$ cujo valor observado determinada a decisão a tomar
- Uma região crítica W_T: se $T(x_1, \ldots, x_n) \in W_T$, rejeitar H_0

Erros de 1^a e de 2^a espécie

Decisão	H_0 verdadeira	H_0 falsa
Rejeitar H_0	erro de 1^a espécie	decisão correcta
Aceitar H_0	decisão correcta	erro de 2^a espécie

- Ao erro que consiste em rejeitar H_0 quando H_0 é efectivamente verdadeira dá-se o nome de erro de tipo I ou erro de 1^a espécie;
- Ao erro que consiste em aceitar H_0 quando H_0 é efectivamente falsa dá-se o nome de erro de tipo II ou erro de 2^a espécie;
- note-se que estes erros não podem ocorrer simultaneamente
- tradicionalmente, diz-se "não rejeitar H_0" em vez de "aceitar H_0"

4.2 Hipóteses simples: Lema de Neyman-Pearson

- Considere-se a situação em que ambas as hipóteses em confronto são simples: $\Theta = \{\theta_0, \theta_1\}$ e quer-se testar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$
- Considere-se um teste definido pela região crítica W.
- A probabilidade de cometer um erro de 1^a espécie

$$\alpha = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) = P[(X_1, \ldots, X_n) \in W \mid \theta = \theta_0]$$

  é designada por **dimensão do teste**
- O complementar da probabilidade de cometer um erro de 2^a espécie

$$\beta = P(\text{rejeitar } H_0 \mid H_0 \text{ falsa}) = P[(X_1, \ldots, X_n) \in W \mid \theta = \theta_1]$$

  designa-se por **potência do teste**

Probabilidades dos erros de 1^a e de 2^a espécie

Decisão	H_0 verdadeira	H_0 falsa
Rejeitar H_0	erro de 1^a espécie Probabilidade $= \alpha$	decisão correcta Probabilidade $= \beta$
Aceitar H_0	decisão correcta Probabilidade $= 1 - \alpha$	erro de 2^a espécie Probabilidade $= 1 - \beta$

Interpretação prática

- Quando se observa a amostra concreta $(x_1, \ldots, x_n)$,
  - ou $(x_1, \ldots, x_n) \in W$, caso em que se rejeita H_0 ao nível de $\alpha \times 100\%$
  - ou $(x_1, \ldots, x_n) \notin W$, caso em que não existe evidência contra H_0 suficiente para rejeitar H_0 ao nível de $\alpha \times 100\%$
- Qual o erro cometido, se algum?
- Se se rejeita H_0, é incorrecto afirmar que a probabilidade de se estar a cometer um erro é α! Note-se que o que teria interesse era a probabilidade $P(H_0 \text{ verdadeira} \mid (x_1, \ldots, x_n) \in W)$
- Se se repetir o processo de observar uma amostra de tamanho n um número muito grande de vezes, se H_0 for verdadeira então será rejeitada incorrectamente aproximadamente $\alpha \times 100\%$ das vezes; se H_0 for falsa, será correctamente rejeitada aproximadamente $\beta \times 100\%$ dos casos

Exemplo 4.4 *Considere-se uma amostra casual de dimensão* $n = 1$ *proveniente de uma população* $X \sim N(\mu, 25)$. *Pretende-se testar*

$$H_0 : \mu = 15 \text{ versus } H_1 : \mu = 25$$

Uma região de rejeição sensata parece ser $W = \{x : x > k\}$ *para algum* k *fixo. Por exemplo, para* $k = 24$ *tem-se*

$$\alpha = P(X > 24 \mid \mu = 15) = 1 - \Phi[(24 - 15)/5] \approx 0.0359$$

$$\beta = P(X > 24 \mid \mu = 25) = 1 - \Phi[(24 - 25)/5] \approx 0.5793$$

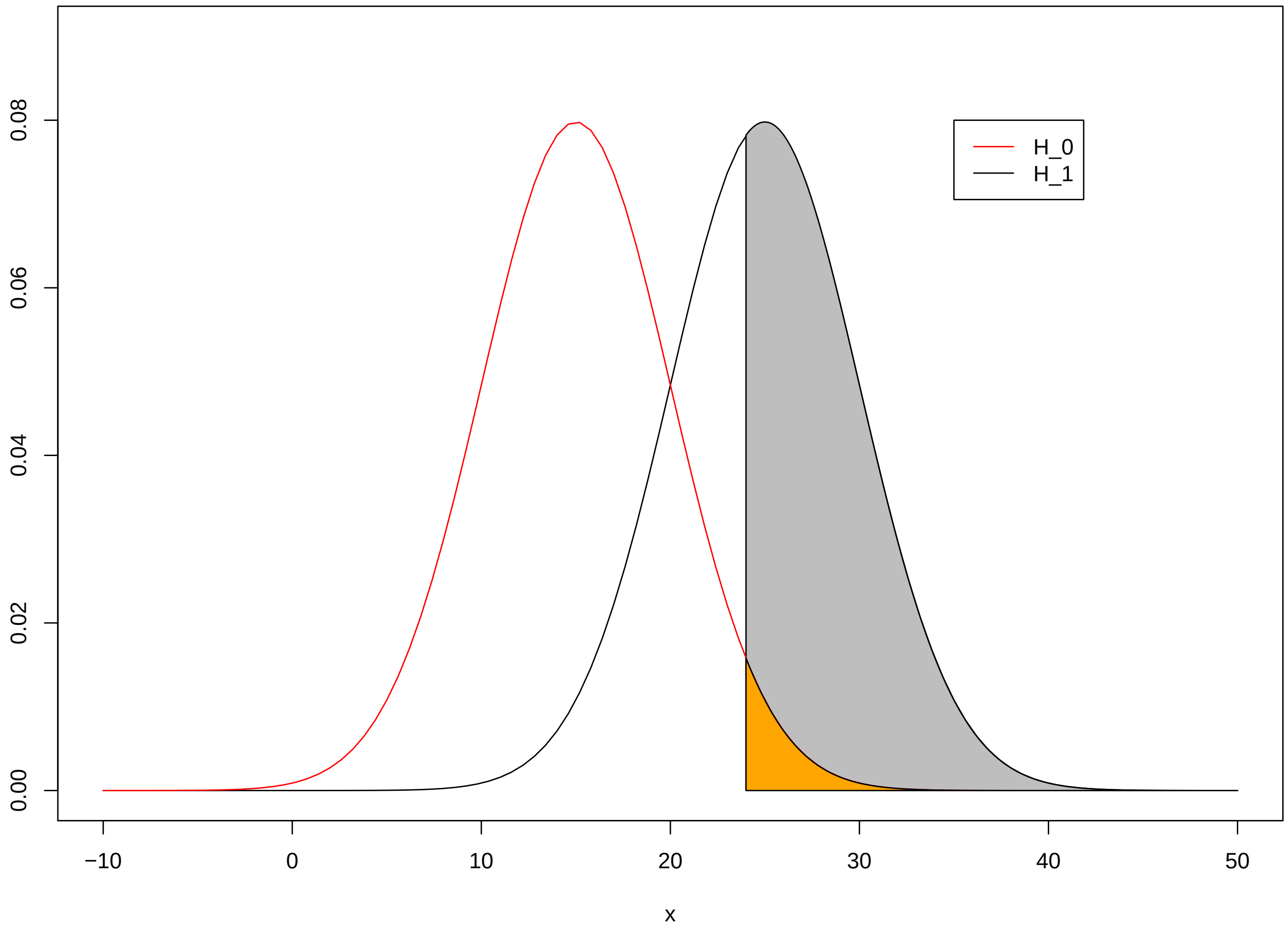


Figura 1: Dimensão do teste (laranja) e potência (cinzento)

- Note-se como uma diminuição de α implica uma diminuição de β (ou um aumento de $1-\beta$): é impossível minimizar simultaneamente a probabilidade de os dois tipos de erro para uma dimensão fixa da amostra

- Estratégia de Neyman-Pearson: fixar α (erro de 1^a espé cie ou dimensão do teste) num valor adequado (tradicionalmente, $\alpha = 0.01, 0.05$ ou 0.10) e procurar o teste com menor erro de 2^a espécie, ou maior potência: o **teste mais potente** de dimensão α

- Esta estratégia evidencia mais uma vez o carácter assimétrico do tratamento dado às duas hipóteses em confronto: precavemo-nos contra uma incorrecta rejeição de H_0

Definição 4.1 Teste mais potente: *Fixada a dimensão do teste (probabilidade de erro de 1^a espécie), o teste mais potente é aquele cuja região crítica minimiza a probabilidade de erro de 2^a espécie, ou maximiza a potência. Diz-se que a região crítica associada é a mais potente.* ■

Teorema 4.1 Lema de Neyman-Pearson: *Seja $X_1, \ldots, X_n$ uma amostra casual proveniente de uma população com função densidade $f(\cdot \mid \theta)$, $\theta \in \{\theta_0, \theta_1\}$.*

Seja $c > 0$, $\alpha \in (0, 1)$, e considere-se $W \subset \mathcal{X}$ definido pelas condições

$$\frac{\prod_{i=1}^n f(x_i \mid \theta_1)}{\prod_{i=1}^n f(x_i \mid \theta_0)} = \frac{L(\theta_1 \mid x_1, \ldots, x_n)}{L(\theta_0 \mid x_1, \ldots, x_n)} > c \Leftrightarrow (x_1, \ldots, x_n) \in W$$

$$P[(X_1, \ldots, X_n) \in W \mid \theta = \theta_0] = \alpha \ .$$

Então, o teste associado à região crítica W é o mais potente de dimensão α para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. ∎

Exemplo 4.5 *Seja* $X_1, \ldots, X_n$ *uma amostra casual proveniente de uma população* $N(\mu, 1)$. *Determinar o teste mais potente de dimensão* $\alpha \in (0, 1)$ *para testar* $H_0 : \mu = \mu_0$ *verus* $H_1 : \mu = \mu_1$, *com* $\mu_1 > \mu_0$

1. *Construir a função verosimilhança:*

$$L(\mu \mid x_1, \ldots, x_n) \propto \exp\left[\frac{-1}{2}(n\mu^2 - 2n\bar{x}\mu)\right]$$

2. *Calcular a razão de verosimilhanças:*

$$\frac{L(\mu_1 \mid x_1, \ldots, x_n)}{L(\mu_0 \mid x_1, \ldots, x_n)} = \exp\left\{\frac{-1}{2}[n(\mu_1^2 - \mu_0^2) - 2n\bar{x}(\mu_1 - \mu_0)]\right\}$$

3. *Simplificar a condição sobre a razão de verosimilhanças:*

$$\frac{L(\mu_1 \mid x_1, \ldots, x_n)}{L(\mu_0 \mid x_1, \ldots, x_n)} > c \Leftrightarrow \bar{x} > k$$

4. *Determinar* k *através da dimensão do teste:*

$$P(\bar{X} > k \mid \mu = \mu_0) = \alpha \Leftrightarrow 1 - \Phi[\sqrt{n}(k - \mu_0)] = \alpha \Leftrightarrow k = \mu_0 + \Phi^{-1}(1 - \alpha)/\sqrt{n}$$

■

Observações:

- A aplicação do lema é normalmente feita em duas fases: determinar a estatística de teste e a forma da região crítica — no exemplo, a estatística de teste é $\bar{X}$ e a região de rejeição é da forma $W = \{\bar{x} > k\}$ para algum k. Na segunda fase encontra-se k usando a condição de o teste ter dimensão α

- A estatística de teste será sempre uma estatística suficiente

- Em muitos casos, o valor concreto de θ_1 não é relevante para definir W: é apenas necessário saber se $\theta_0 > \theta_1$ ou se, pelo contrário, $\theta_1 < \theta_0$. No entanto, a potência do teste MP depende do valor concreto de θ_1

Observações:

- Regra geral: se θ é uma média, variância ou proporção, a estatística de teste é um estimador de θ, e a região de rejeição "está do lado da alternativa"
- Em universos discretos, não se consegue em geral encontrar um teste com a dimensão α fixada. O lema de Neyman-Pearson aplica-se a universos discretos, e garante que se um teste de dimensão α é da forma indicada, então ele é o mais potente
- Por vezes recorre-se à aleatorização: o teste que rejeita H_0 com probabilidade 1 se

$$\frac{L(\theta_1 \mid x_1, \ldots, x_n)}{L(\theta_0 \mid x_1, \ldots, x_n)} > c$$

e com probabilidade ε se

$$\frac{L(\theta_1 \mid x_1, \ldots, x_n)}{L(\theta_0 \mid x_1, \ldots, x_n)} = c$$

é o teste mais potente com a sua dimensão, que é dada por

$$P(\text{rejeitar } H_0 \mid H_0) = P(R > c \mid H_0) + \varepsilon P(R = c \mid H_0)$$

onde $R = L(\theta_1 \mid X_1, \ldots, X_n)/L(\theta_0 \mid X_1, \ldots, X_n)$

Exemplo 4.6 *Considere-se uma população* $B(1,\theta)$ *e suponha-se que se pretende testar as hipóteses* $H_0 : \theta = \theta_0$ *contra* $H_1 : \theta = \theta_1$ *com* $\theta_0 > \theta_1$.

É fácil verificar que

$$\frac{L(\theta_1 \mid x_1, \ldots, x_n)}{L(\theta_0 \mid x_1, \ldots, x_n)} > c \Leftrightarrow \sum_{i=1}^{n} x_i \leq k \ .$$

Se $n = 10$ *e* $\theta_0 = 0.4$, *com* $T = \sum_{i=1}^{10} X_i$, *tem-se*

$$P(T \leq 1 \mid \theta = 0.4) = 0.0463$$

$$P(T \leq 2 \mid \theta = 0.4) = 0.1672$$

pelo que $W = \{\sum_{i=1}^{10} x_i \leq 1\}$ *é o teste mais potente de dimensão* 0.0463 *e* $W = \{\sum_{i=1}^{10} x_i \leq 2\}$ *é o teste mais potente de dimensão* 0.1672. *O teste de dimensão 0.05 rejeita* H_0 *sempre se* $\sum_{i=1}^{10} x_i \leq 1$ *e com probabilidade* $0.0301 = (0.05 - P(T \leq 1 \mid \theta = 0.4))/P(T = 2 \mid \theta = 0.4)$ *se* $\sum_{i=1}^{10} x_i = 2$. ■

4.3 Hipótese simples contra hipótese composta

- Estudamos agora a situação em que H_0 é simples e H_1 é composta, mas unilateral:

$$H_0 : \theta = \theta_0 \text{ contra } H_1 : \theta > \theta_0 \text{ (ou contra } H_1 : \theta < \theta_0)$$

- a estratégia continua a ser a de fixar o erro de 1^a espécie
- no entanto, neste contexto o erro de 2^a espécie é uma função de θ, para $\theta \in \Theta_1$
- Temos neste caso a função potência e o teste uniformemente mais potente:

Definição 4.2 Função potência: *A função potência do teste*

$$H_0 : \theta = \theta_0 \ \mathit{contra}\ H_1 : \theta > \theta_0 \ (\mathit{ou\ contra}\ H_1 : \theta < \theta_0)$$

com região crítica W é dada por

$$\beta(\theta) = P(\mathit{rejeitar}\ H_0 \mid H_0\ \mathit{falsa}) = P((X_1, \ldots, X_n) \in W \mid \theta), \quad \theta \in \Theta_1$$

onde $\Theta_1 =]\theta_0, +\infty[$ *(ou* $\Theta_1 =]-\infty, \theta_0[$*).* ■

Definição 4.3 Teste uniformemente mais potente (UMP): *Ao testar a hipótese nula* $H_0 : \theta = \theta_0$ *contra a alternativa* $H_1 : \theta > \theta_0$ *(ou contra* $H_1 : \theta < \theta_0$*), considerem-se dois testes de dimensão* α*:* T_1 *com função potência* $\beta(\theta)$ *e* T_2 *com função potência* $\beta^\star(\theta)$*. Dizemos que* T_1 *é uniformente mais potente do que* T_2 *se* $\beta(\theta) \geq \beta^\star(\theta)$ *para todo o* $\theta \in \Theta_1$*.*

Um teste T *de dimensão* α *diz-se uniformemente mais potente para as hipóteses em causa se for uniformemente mais potente do que qualquer outro teste de dimensão* α*.* ■

Exemplo 4.7 *Seja* $X_1, \ldots, X_n$ *uma amostra casual proveniente de uma população* $N(\mu, 1)$*. Pretende-se testar* $H_0 : \mu = \mu_0$ *contra* $H_1 : \mu > \mu_0$*.*

Vimos que o teste mais potente para $H_0 : \mu = \mu_0$ *contra* $H_1 : \mu = \mu_1$*, onde* $\mu_1 > \mu_0$ *é dado pela região crítica* $W = \{(x_1, \ldots, x_n) : \bar{x} > \mu_0 + \Phi^{-1}(1-\alpha)/\sqrt{n}\}$*. Este teste não depende do valor de* μ_1 *(desde que* $\mu_1 > \mu_0$*), e consequentemente é o teste uniformente mais potente para as hipóteses em causa.* ■

Observação: O exemplo acima sugere que existem circustâncias em que o teste UMP pode ser encontrado através do lema e Neyman-Peason, já que o teste encontrado não depende do particular valor escolhido para a alternativa. Antes de prosseguir, estendemos o contexto um pouco mais.

4.4 Hipótese nula unilateral contra alternativa unilateral

- A hipóteses em confronto são do tipo

$$H_0 : \theta \leq \theta_0 \text{ contra } H_1 : \theta > \theta_0$$

ou

$$H_0 : \theta \geq \theta_0 \text{ contra } H_1 : \theta < \theta_0$$

- A probabilidade de erro de 1^a espécie é agora também uma função de $\theta \in \Theta_0$. Define-se a dimensão do teste como o erro máximo de 1^a espécie:

$$\alpha = \sup_{\theta \in \Theta_0} P(\text{rejeitar } H_0 \mid \theta)$$

- Generalização do conceito de função potência:

$$\beta(\theta) = P[(X_1, \ldots, X_n) \in W \mid \theta], \quad \theta \in \Theta$$

- Para $\theta \in \Theta_0$, $\beta(\theta)$ é a probabilidade de erro de 1^a espécie; para $\theta \in \Theta_1$, $\beta(\theta)$ é a probabilidade de não cometer um erro de 2^a espécie.

- Nestas circunstâncias, para que populações se pode garantir a existência de testes UMP, ou seja, de testes de tamanho α cuja função potência seja a maior possível para $\theta \in \Theta_1$?

Definição 4.4 Razão de verosimilhança monótona (RVM): *A família* $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$, *onde* Θ *é um intervalo real, tem RVM na estatística* $T = T(X_1, \ldots, X_n)$ *quando para quaisquer* $\theta' > \theta$, *(*$\theta', \theta \in \Theta$*), a razão de verosimilhanças*

$$\frac{L(\theta' \mid x_1, \ldots, x_n)}{L(\theta \mid x_1, \ldots, x_n)}$$

é função monótona não decrescente de T. ■

Observações:

- o importante é que a razão de verosimilhança seja uma função monótona de uma estatística: se for não crescente em T será não decrescente em $S = -T$.
- Distribuições com RVM incluem $\mathrm{Po}(\lambda)$, $\mathrm{N}(\mu, 1)$, $\mathrm{Ex}(\lambda)$...

Teorema 4.2 Teorema de Karlin-Rubin: *Se $(X_1, \ldots, X_n)$ for uma amostra casual de uma população com RVM na estatística T, então para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$, os testes com região de rejeição da forma $W = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > k\}$ são UMP da sua dimensão, que é dada por $P(rejeitar\ H_0 \mid \theta_0)$.* ■

Observação Importante: O teorema cobre também os casos em que a razão de verosimilhanças é monótona não crescente, ou em que as hipóteses são do tipo $H_0 : \theta \geq \theta_0$ contra $H_1 : \theta < \theta_0$. Basta considerar a parameterização $\psi = -\theta$ ou $\psi = 1/\theta$ e a estatística $W = -T$ ou $W = 1/T$. Resumo:

	crescente	decrescente
$H_0 : \theta \leq \theta_0$	$> t_0$	$< t_0$
$H_0 : \theta \geq \theta_0$	$< t_0$	$> t_0$

Exemplo 4.8 *Considere-se uma população* $Ex(\theta)$ *e as hipóteses* $H_0 : \theta \geq \theta_0$ *contra* $H_1 : \theta < \theta_0$*. A função de verosimilhança escreve-se*

$$L(\theta) \propto \theta^n \exp(-\theta t)$$

onde $t = \sum x_i$*. Se* $\theta' > \theta$*, então*

$$L(\theta')/L(\theta) = (\theta'/\theta) \exp[-(\theta' - \theta)t]$$

que é uma função descrescente em t*, crescente em* $s = -t$*.*

Considerando a parameterização $\lambda = 1/\theta$*, as hipóteses escrevem-se* $H_0 : \lambda \leq \lambda_0$ *contra* $H_1 : \lambda > \lambda_0$*, onde* $\lambda_0 = 1/\theta_0$*. A razão de verosimilhanças, para* $\lambda' > \lambda$ *é*

$$L(\lambda')/L(\lambda) = (\lambda/\lambda') \exp[-(1/\lambda' - 1/\lambda)t]$$

que é crescente em t*. A região de rejeição UMP é então* $W = \{(x_1, \ldots, x_n) : \sum_{i=1}^n x_i > t_0\}$*.*

Se $\theta_0 = 2$, $\alpha = 0.05$ e $n = 1$, vem $\beta(\theta) = P(T > t_0 \mid \theta) = \exp(-\theta t_0)$, que é uma função descrescente de θ. Logo,

$$\alpha = \sup_{\theta \geq 2} \beta(\theta) = \beta(2) \Rightarrow e^{-2t_0} = \alpha \Rightarrow t_0 = -[\ln \alpha]/2 \approx 1.4979$$

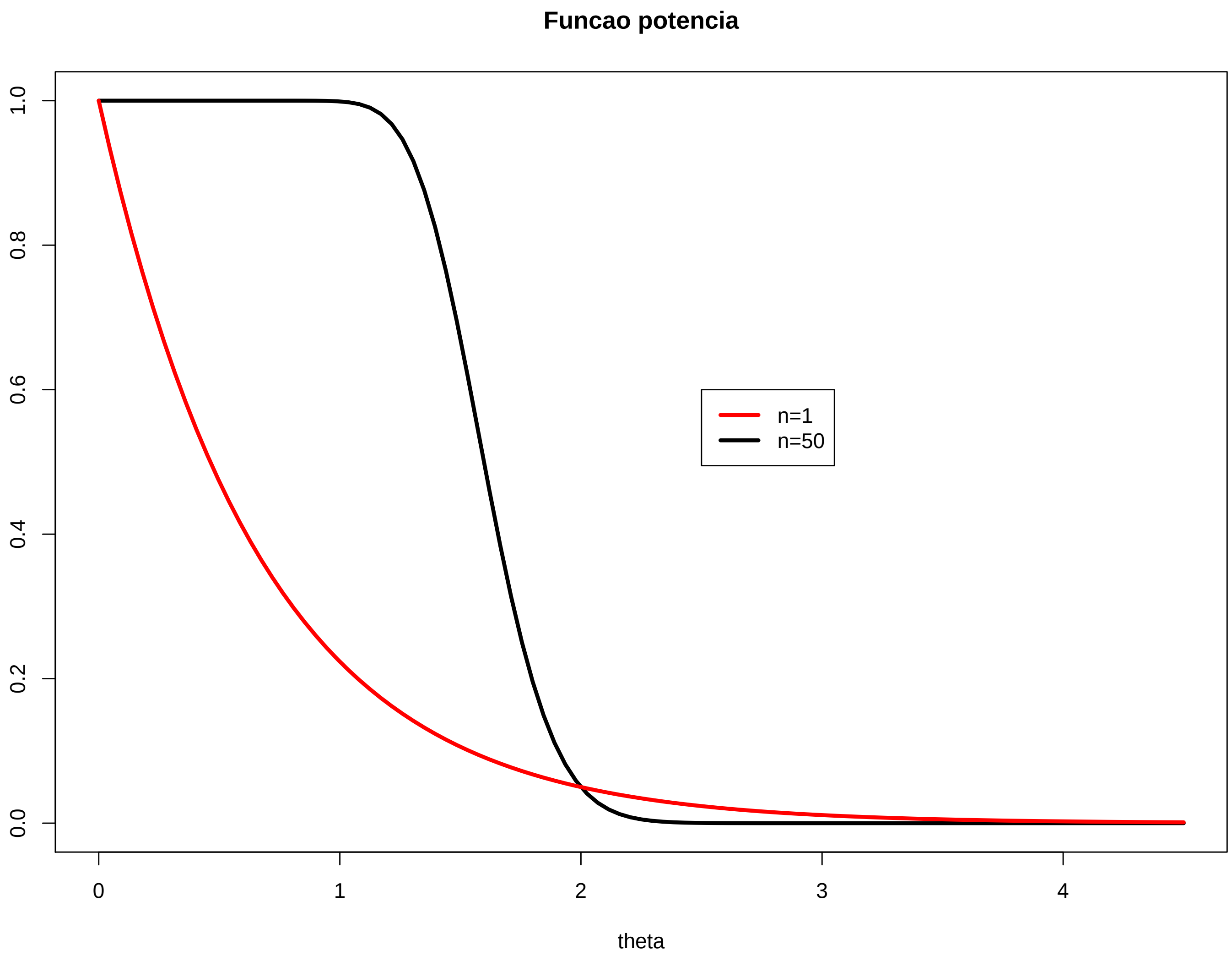
Assim, $\beta(\theta) \approx \exp(-1.4979\theta)$, $\theta > 0$.

Se $n = 50$, $T = \sum X_i \sim \mathrm{G}(n, \theta)$, logo

$$\beta(\theta) = 1 - F_{\chi^2(2n)}(2t_0\theta)$$

que é uma função descrecente de θ. Assim, $\alpha = 1 - F_{\chi^2(2n)}(4t_0)$ e portanto $t_0 = F^{-1}_{\chi^2(100)}(0.95)/4 \approx 31.0855$. Assim, $\beta(\theta) \approx 1 - F_{\chi^2(100)}(62.171\theta)$.

```
curve(1-pchisq(62.171*x,100),from=0,to=4.5,ylab="",xlab="theta",
      main="Funcao potencia")
curve(exp(-1.4979*x),add=T,col=2)
legend(2.5,0.6,c("n=1","n=50"),col=c(2,1),lty=c(1,1))
```


Funcao potencia
n=1
n=50
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
theta

4.5 Hipótese simples contra composta bilateral

- Em geral, $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$
- Neste caso, regra geral não existem testes UMP: por exemplo, considere-se a situação em que $X \sim \mathrm{N}(\mu, 1)$ e se pretende testar $H_0 : \mu = 0$ contra $H_1 : \mu \neq 0$. Se se considerar $\mu > 0$, a região crítica UMP é $\bar{x} > k$, enquanto que se considerar a alternativa $\mu < 0$, a região crítica UMP é da forma $\bar{x} < k'$. Prova-se então que não pode existir teste UMP para as hipóteses em causa.
- Solução: limitar a classe de testes onde se procura o óptimo, ou seja, procurar o teste UMP entre a classe de testes de tamanho α centrados. Esse teste, quando existe, é denominado de teste UMPU — uniformly most powerful unbiased.
- Na prática: duas regiões de rejeição nas abas da distribuição da estatística de teste, distribuindo metade do erro de 1^a espécie $(\alpha/2)$ por cada uma delas
- No exemplo acima, $\alpha/2 = P(\bar{X} > k \mid \mu = 0)$ e $\alpha/2 = P(\bar{X} < k' \mid \mu = 0)$

4.6 Ensaios de significância

- Os ensaios de significância distinguem-se dos testes de hipóteses pela ausência da hipótese alternativa, pelo menos formalmente — só se fazem cáculos probabilísticos sob H_0
- Baseados numa estatística de teste T apropriada à hipótese nula em causa
  - a distribuição por amostragem de T é conhecida sob H_0, pelo menos aproximadamente
  - determinado tipo de valores observados de T lançam dúvidas sobre a validade de H_0. Por exemplo, ao testar $H_0 : \mu > 0$ onde μ é o valor médio de uma população normal, $T = \bar{X}$ e valores pequenos de T põem em causa a validade de H_0
  - Observado $t_{obs} = T(x_1, \ldots, x_n)$, calcular a probabilidade de observar, sob H_0, valores de T tão ou mais incompatíveis com H_0 do que t_{obs}, no exemplo acima, $p = P(T < t_{obs})$ — valor-p (p-value)
  - Se p pequeno, ou H_0 é verdadeira e obteve-se por acaso um acontecimento pouco provável sob H_0, ou os dados observados não foram efectivamente gerados por H_0
  - acredita-se na segunda explicação e rejeita-se H_0 se p pequeno — dados são significativamente inconsistentes com H_0, há evidência siginificativa contra H_0
- Ausência formal de hipótese alternativa

Valor-p

- Na perspectiva de Neyman-Peason, α é fixo e rejeita-se H_0 ao nível α se $(x_1, \ldots, x_n) \in W$. Independentemente de quão longe os dados se encontram do limiar da região de rejeição, a decisão é a mesma
- o valor-p corresponde a uma tentativa de medir a evidência pró-H_0 utilizando uma quantidade que depende dos dados observados, e que será tanto maior quanto mais longe do limiar da região de rejeição estiverem os dados observados
- Dados que conduzam a $p = 0.001$ contêm maior evidência contra H_0 do que dados que conduzam a $p = 0.1$
- Importante: o valor-p não corresponde à probabilidade de H_0 ser verdadeira. O uso do valor-p como medida de evidência contra H_0 é um assunto controverso ao nível dos fundamentos da Estatística

Procedimento:

- Fixar o nível de significância α (usualmente em $1\%, 5\%$ ou 10%)
- Calcular o valor observado da estatística de teste, t_{obs}
- Calcular o valor-p: probabilidade de observar valores de T tão ou mais inconsistentes com H_0 que t_{obs} (ou o seu supremo, se H_0 composta)
- Se $p \leq \alpha$, p é pequeno, ensaio significativo, rejeitar H_0 ao nível α
- se $p > \alpha$, p é grande, ensaio não significativo, não rejeitar H_0 ao nível α
- podemos simplesmente reportar p como uma medida da evidência pró-H_0 veiculada pelos dados

Exemplo 4.9 *Considere-se uma população* $N(\mu, 1)$ *e as hipóteses nulas*

1. $H_0 : \mu = 10$
2. $H_0 : \mu \geq 10$

Em ambos os casos, a estatística de teste é $\bar{X} \mid \mu \sim N(\mu, 1/n)$. *Se* $\bar{x} = 10.5$ *e* $n = 16$,

1. *valores tão ou mais desfavoráveis com* H_0 *correspondem a valores de* $\bar{X}$ *tão ou mais distantes de 10 que o valor observado, ou seja,*

$$p = P(|\bar{X} - 10| > |10 - 10.5| \mid \mu = 10) = 2 \times [1 - \Phi(0.5 \times \sqrt{16})] = 0.0455$$

2. *valores tão ou mais desfavoráveis a* H_0 *correspondem a valores de* $\bar{X}$ *menores ou iguais que o valor observado, ou seja,*

$$p = \sup_{\mu \geq 10} P(\bar{X} \leq 10.5 \mid \mu) = \sup_{\mu \geq 10} \Phi[(10.5 - \mu)\sqrt{16}] = \Phi[(10.5 - 10)\sqrt{16}] = 0.9772$$

4.7 Aplicações

- Universo normal: testes para a média de uma população normal (variância conhecida e desconhecida); testes para a variância de uma população normal
- Dois universos nomais: testes para a diferença das médias e o quociente das variâncias
- Testes para universos não normais: resultados assintóticos

Amostras emparelhadas: Uma amostra emparelhada de populações normais é $(X_i, Y_i), i = 1, \ldots, n$ tais que

1. (X_i, Y_i) independente de (X_j, Y_j) se $i \neq j$
2. $X_i \sim N(\mu_X, \sigma_X^2)$; $Y_i \sim N(\mu_Y, \sigma_Y^2)$
3. X_i e Y_j não são em geral independentes: correpondem e.g. a características diferentes medidas no mesmo indivíduo

Pretende-se testar $H_0 : \mu_X = \mu_Y$, ou seja, $H_0 : \mu_X - \mu_Y = 0$. Como

$$Z_i = X_i - Y_i \overset{iid}{\sim} N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}), \quad i = 1, \ldots, n$$

deduz-se sem dificuldade que

$$T = \frac{\bar{Z} - (\mu_X - \mu_Y)}{S'_Z/\sqrt{n}} \sim t(n-1)$$

sendo $\bar{Z} = \bar{X} - \bar{Y}$ e $S_Z'^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2/(n-1)$.

- Reduz-se o problema a um já tratado.
- Note-se que não é necessário fazer suposições acerca de σ_X^2 e σ_Y^2, nem conhecer σ_{XY}.

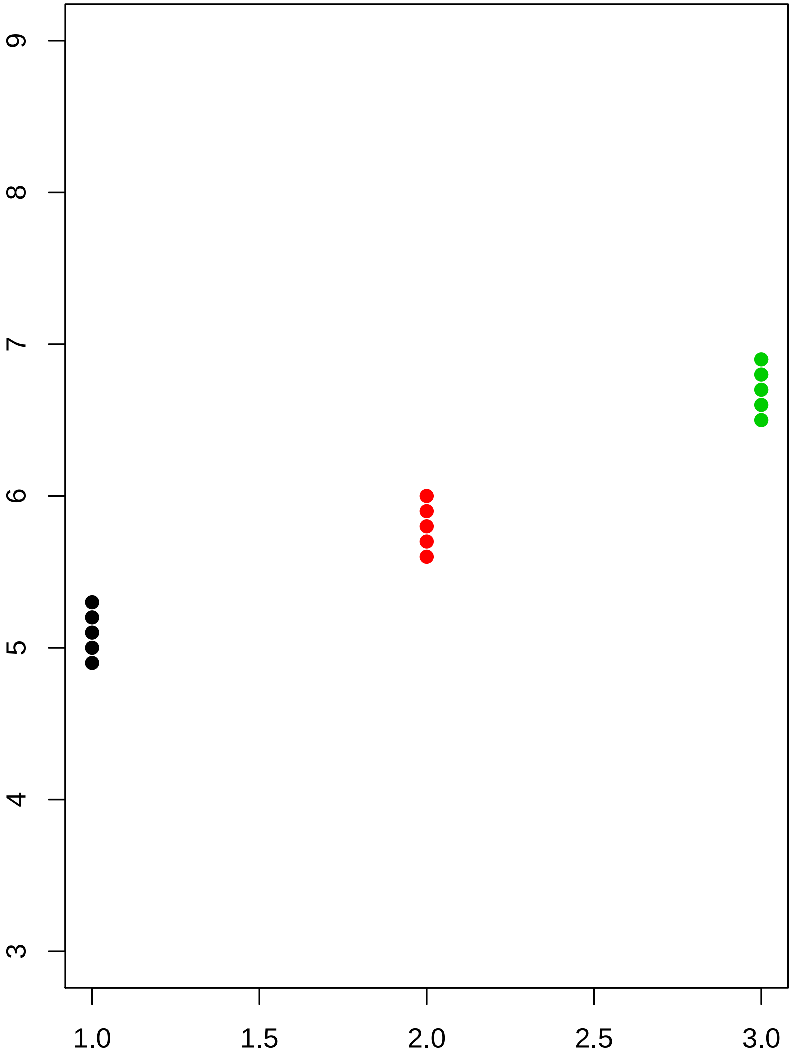
4.8 ANOVA — Análise de variância

Introdução

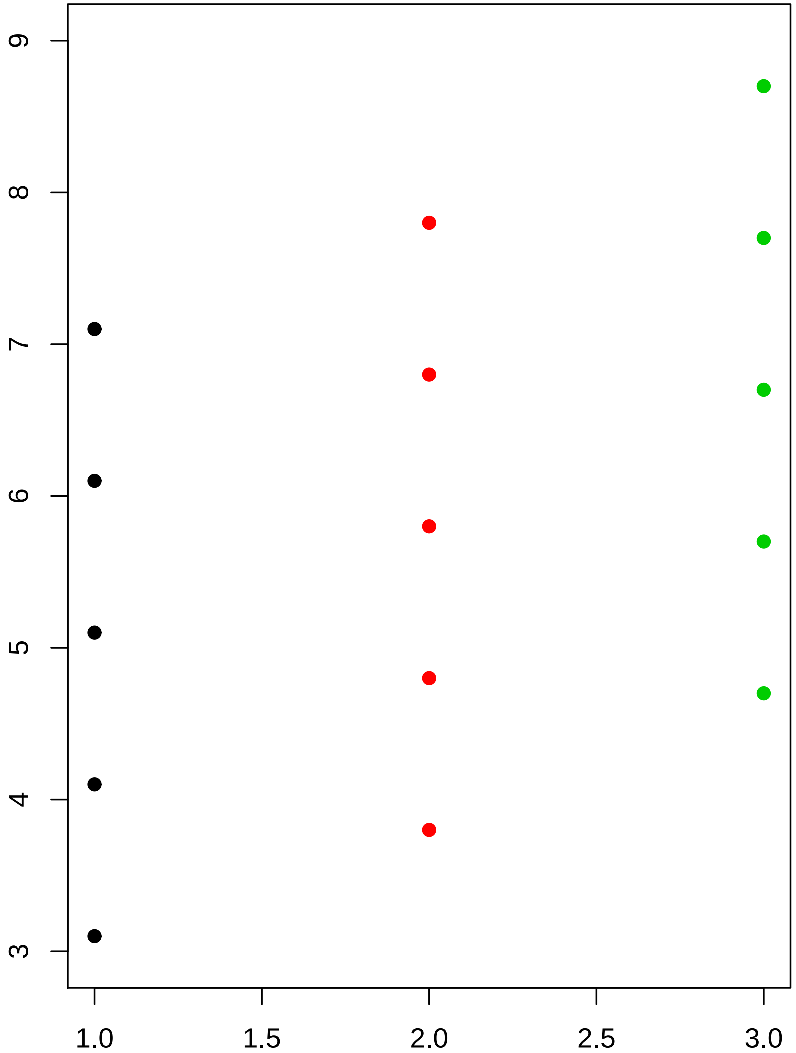
- Objectivo: testar a igualdade das médias de várias populações normais (de variância comum), atribuindo a uma ou mais causas a eventual diferença entre as médias
- Exemplo: Pretende-se averiguar se o rendimento médio é idêntico no Norte, Centro e Sul de determinado país
- Hipótese de interesse: $H_0 : \mu_1 = \mu_2 = \ldots = \mu_m$
- As m populações podem ser definidas de várias formas, que influenciam a forma de testar a hipótese

Exemplo 4.10 *Suponha-se que se pretende testar a igualdade do rendimento médio no Norte, Centro e Sul de um país: populações A, B e C. Recolhe-se uma amostra casual simples de dimensão 5 em cada uma das populações e observa-se que* $\bar{x}_A = 5.7$, $\bar{x}_B = 5.8$, $\bar{x}_C = 6.7$.

Será a variabilidade patente nas médias das amostras compatível com a hipótese nula? Depende!



(a)



(b)

- em cada caso, a variabilidade patente nas três amostras é similar
- no caso (a), essa variabilidade é pequena comparada com a variabilidade patente nas médias das amostras, facto que nos leva intuitivamente a pôr em causa a hipótese nula
- a comparação entre a variabilidade dentro de cada amostra e a variabilidade entre as amostras é portanto um factor chave na avaliação da plausibilidade de H_0
- Análise de variância: ANOVA — técnica com aplicações mais vastas

4.9 Análise com um factor: classificação simples

- *one-way* ANOVA
- A definição das populações é feita tendo em conta apenas um critério, designado por *factor*. Cada uma das populações corresponde a um *nível* do factor, que apresenta m níveis
- Exemplos de factores: sexo, nível de instrução, faixa etária
- A rejeição de H_0 permite concluir que as m populações não apresentam comportamento idêntico. No entanto,
- Só é legítimo considerar este factor como a causa deste comportamento distinto se se puder garantir a homogeneidade das populações face a todos os outros factores que poderiam ser relevantes para a explicação do fenómeno
- Variáveis confundidoras: imagine-se que se conclui que homens e mulheres apresentam concentrações médias de hemoglobina significativamente diferentes. Isso pode ser devido à maior incidência de fumadores em uma das populações. Possível solucão: incluir o factor "fumador" na análise...
- Note-se que o contrário de todos iguais corresponde a pelo menos dois diferentes, e não a todos diferentes...

Modelo estatístico: modelo de efeitos fixos

- Recolhem-se m amostras casuais, independentes, cada uma de tamanho n_i:

$$(X_{i1}, X_{i2}, \ldots, X_{in_i}), \quad i = 1, \ldots, m$$

- $X_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, m,\ j = 1, \ldots, n_i$ — note-se que a variância σ^2 é desconhecida e comum às m populações

- Modela-se o efeito do factor da forma

$$\mu_i = \mu + \alpha_i,\ i = 1, \ldots, m$$

  efeito linear e aditivo; μ é a média geral, α_i é o efeito do nível i do factor

- Formulação alternativa:

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

  onde $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

- para haver uma correspondência biunívoca entre μ_i e μ, α_i há que introduzir restrições nos α_i, as chamadas restrições de identificabilidade. Uma possibilidade, a que adoptamos aqui, é impor que $\sum_{i=1}^{m} n_i\alpha_i = 0$

O modelo de efeitos fixos assenta em cinco pressupostos cuja validade deve ser cuidadosamente ponderada:

1. $E[\varepsilon_{ij}] = 0$ $(i = 1, \ldots, m,\ j = 1, \ldots, n_i)$
2. ε_{ij} são independentes, i.e., a amostragem é casual e as amostras são independentes entre si
3. $\mathrm{var}(\varepsilon_{ij}) = \sigma^2$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, ou seja, verifica-se a homocedasticidade das populações (homogeneidade das variâncias)

   A violação deste pressuposto (heterocedasticidade) pode ter sérias consequências no que diz respeito à validade das conclusões
4. ε_{ij} têm distibuição Normal. A violação deste pressuposto pode não ter consequências sérias se a dimensão das amostras for relativamente grande (teorema do limite central)
5. o efeito do nível i do factor é representado pelo parâmetro α_i num modelo linear e aditivo

- Hipótese de interesse:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$$

$$\Leftrightarrow$$

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$$

$$\Leftrightarrow$$

$$H_0 : \sum_{i=1}^{m} \alpha_i^2 = 0$$

onde $\sum_{i=1}^{m} n_i \alpha_i = 0$

- Ideia para a construção da estatística de teste: construção de dois estimadores independentes de σ^2:
  - o primeiro é centrado independentemente da validade de H_0
  - o segundo é centrado sob H_0 mas sobrestima σ^2 se H_0 for falsa

  a estatística de teste é dada pelo quociente entre o segundo e o primeiro estimador de σ^2: valores grandes da estatística de teste são incompatíveis com H_0

Identidade fundamental da ANOVA

$$
\begin{aligned}
\text{SQT} &= \text{SQD+SQE} \\
\text{Soma de quadrados total} &= \text{SQ dentro das amostras + SQ entre as amostras}
\end{aligned}
$$

$$
\begin{aligned}
\text{SS(TO)} &= \text{SS(E)+SS(T)} \\
\text{total sum of squares} &= \text{error SS (among treatments)+ (between) treatments SS}
\end{aligned}
$$

$$
\sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{\circ\circ})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\circ})^2 + \sum_{i=1}^{m} n_i(\bar{X}_{i\circ}-\bar{X}_{\circ\circ})^2
$$

onde, com $n=\sum_{i=1}^{m} n_i$,

$$
\bar{X}_{i\circ} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i} \quad \text{(média amostral da população } i\text{)}
$$

$$
\bar{X}_{\circ\circ} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} X_{ij}}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} n_i \bar{X}_{i\circ}}{n} \quad \text{(média global)}
$$

$$\begin{aligned} \text{SQT} &= \text{SQD+SQE} \\ \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{\circ\circ})^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\circ})^2+\sum_{i=1}^{m}n_i(\bar{X}_{i\circ}-\bar{X}_{\circ\circ})^2 \end{aligned}$$

sendo SQD independente de SQE: recorde-se que numa população normal a média e a variância amostrais são independentes, logo $\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\circ})^2$ é independente de $\bar{X}_{i\circ}$ para todo o i, e portanto de $\bar{X}_{\circ\circ}$.

Teorema 4.3 *Se H_0 for verdadeira, tem-se que*

$$\frac{SQD}{\sigma^2} \sim \chi^2(n-m)$$
$$\frac{SQT}{\sigma^2} \sim \chi^2(n-1)\ .$$

seguindo-se que

$$\frac{SQE}{\sigma^2} \sim \chi^2(m-1)$$

■

Conclusão: Se H_0 for verdadeira, segue-se que

$$MQE = \frac{SQE}{m-1}$$
$$MQD = \frac{SQD}{n-m}$$

são estimadores centrados de σ^2. Além disso,

$$F = \frac{MQE}{MQD} \mid H_0 \sim F(m-1, n-m)$$

Quando H_0 não é verdadeira, continua a ter-se que

$$\frac{SQD}{\sigma^2} \sim \chi^2(n-m)$$

logo, $MQD = SQD/(n-m)$ continua a ser um estimador centrado de σ^2.

No entanto, $MQE = SQE/(n-1)$ já não o é: note-se que

$$\sum_{i=1}^{m} n_i(\bar{X}_{i\circ} - \bar{X}_{\circ\circ})^2 = \sum_{i=1}^{m} n_i \bar{X}_{i\circ}^2 - n\bar{X}_{\circ\circ}^2$$

e portanto

$$E[SQE] = \sum_{i=1}^{m} n_i[\mathrm{Var}(\bar{X}_{i\circ}) + (E[\bar{X}_{i\circ}])^2] - n[\mathrm{Var}(\bar{X}_{\circ\circ}) + (E[\bar{X}_{\circ\circ}])^2]$$

Como $\bar{X}_{i\circ} \sim \mathrm{N}(\mu + \alpha_i, \sigma^2/n_i)$, independentes, e $\bar{X}_{\circ\circ} = \sum_{i=1}^{m} n_i \bar{X}_{i\circ}/n$, segue-se que

$$\mathrm{Var}(\bar{X}_{\circ\circ}) = \frac{\sigma^2}{n} \qquad E[\bar{X}_{\circ\circ}] = \mu$$

como consequência de $\sum_{i=1}^{m} n_i\alpha_i = 0$. Alguma álgebra mostra que

$$E[SQE] = (m-1)\sigma^2 + \sum_{i=1}^{m} n_i\alpha_i^2 > (m-1)\sigma^2$$

Conclusão: Quando H_0 é falsa,

$$MQE = \frac{SQE}{m-1}$$

sobrestima σ^2.

A estatística de teste para a hipótese de interesse é então

$$F = \frac{SQE/(m-1)}{SQD/(n-m)} = \frac{MQE}{MQD} \mid H_0 \sim F(m-1, n-m)$$

e a correspondente região de rejeição é da forma

$$W = \{f > c_\alpha\}$$

onde, para um teste de dimensão α se tem $c_\alpha = F^{-1}_{F(m-1,n-m)}(1-\alpha)$.

Apresentação habitual dos resultados:

Tabela da ANOVA —classificação simples

Origem da variação	Soma de quadrados	Graus de liberdade	Médias quadráticas
Entre amostras	SQE	$m-1$	$MQE = SQE/(m-1)$
Dentro das amostras	SQD	$n-m$	$MQD = SQD/(n-m)$
Total	SQT	$n-1$	$F = MQE/MQD$

Intervalos de confiança Quando se rejeita H_0, conclui-se que as médias não são todas iguais. É geralmente de interesse procurar onde se situam as possíveis diferenças. Uma solução é construir intervalos de confiança para a média de cada uma das populações. Pode para esse efeito usar-se uma estimativa de σ^2 que usa todo o conjunto de dados:

$$\bar{x}_{i\circ} \pm t_{\alpha/2}\sqrt{\frac{MQD}{n_i}}$$

onde $t_{\alpha/2} = F^{-1}_{t(n-m)}(1-\alpha/2)$

Exemplo 4.11 *Uma fábrica de papel produz, entre outros produtos, sacos para hipermercados. Uma das experiências que o departamento técnico resolveu fazer foi ver o efeito que o factor concentração de madeira de carvalho na polpa tinha sobre a resistência do papel (medida em libras por polegada quadrada). Os níveis relevantes do factor concentração são* $m = 4$ *(5, 10, 15 e 20%) e para cada nível foram feitas* $n_i = 6$ *(*$i = 1, 2, 3, 4$*) observações conforme se indica no quadro abaixo:*

	Observações					
Concentração	*1*	*2*	*3*	*4*	*5*	*6*
5	*7*	*8*	*15*	*11*	*9*	*10*
10	*12*	*17*	*13*	*18*	*19*	*15*
15	*14*	*18*	*19*	*17*	*16*	*18*
20	*19*	*25*	*22*	*23*	*18*	*20*

Apresentação habitual dos resultados:

Tabela da ANOVA

Origem da variação	Soma de quadrados	Graus de liberdade	Médias quadráticas
Entre amostras	$SQE = 382.7917$	3	$MQE = 127.5972$
Dentro das amostras	$SQD = 130.1667$	20	$MQD = 6.5083$
Total	$SQT = 512.9584$	23	$F = 19.6052$

Tomando $\alpha = 0.05$ para dimensão do teste, obtém-se $c_{0.05} = 3.10$ para uma distribuição com 3 e 20 graus de liberdade ou, em alternativa, um valor-p de 0.000004. Rejeitada a igualdade das médias podem construir-se os intervalos de confiança:

$$\mu_1 : [7.83; 12.17] \qquad \mu_2 : [13.5; 17.84] \qquad \mu_3 : [14.83; 19.17] \qquad \mu_4 : [19; 23.34]$$

Nota: para efectuar os cálculos à mão, são úteis as fórmulas

$$SQE = \sum_{i=1}^{m} n_i \bar{X}_{i\circ}^2 - n\bar{X}_{\circ\circ}^2$$

$$SQD = \sum_{i=1}^{m} \sum_{j=1}^{n_i} X_{ij}^2 - \sum_{i=1}^{m} n_i \bar{X}_{i\circ}^2$$

4.10 Análise com dois factores: classificação dupla cruzada

- *Two-way* ANOVA
- A média de cada observação depende de dois factores
- Há que potencialmente considerar a forma como estes dois factores interagem
- Vamos restringir-nos à situação mais simples para ilustrar ideias:
  - O primeiro factor, designado *factor linha*, apresenta m níveis
  - O segundo factor, designado *factor coluna*, apresenta n níveis
  - A amostra observada pode ser classificada em $m \times n$ células, definidas pelos $m \times n$ cruzamentos possíveis dos níveis dos factores
  - Considera-se apenas o caso em que existe uma única observação por cada célula

- Neste quadro, a amostra pode ser representada por uma matriz $m \times n$ de elemento genérico X_{ij} que corresponde a uma observação da população que resulta do cruzamento dos níveis i do factor linha e j do factor coluna

- Modelo estatístico:
  - $X_{ij} \overset{ind}{\sim} N(\mu_{ij}, \sigma^2)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$
  - $\mu_{ij} = \mu + \alpha_i + \beta_j$
  - restrições de identificabilidade: $\sum_{i=1}^m \alpha_i = \sum_{j=1}^n \beta_j = 0$
  - sinteticamente, $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, com $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

- Hipóteses de interesse:

$$H_0^\ell : \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$$
$$H_0^c : \beta_1 = \beta_2 = \cdots = \beta_n = 0$$

que correspondem, respectivamente, a ausência de efeito linha e de efeito coluna.

$$\begin{aligned} \text{SQT} &= \text{SQL+SQC+SQR} \\ \text{Soma de quadrados total} &= \text{SQ entre as linhas + SQ entre as colunas + SQ residual} \end{aligned}$$

onde

$$\mathrm{SQT} = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{\circ\circ})^2$$

$$\mathrm{SQL} = \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{X}_{i\circ} - \bar{X}_{\circ\circ})^2$$

$$\mathrm{SQC} = \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{X}_{\circ j} - \bar{X}_{\circ\circ})^2$$

$$\mathrm{SQR} = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_{i\circ} - \bar{X}_{\circ j} - \bar{X}_{\circ\circ})^2$$

Seguindo-se um raciocínio idêntico ao desenvolvido no caso da classificação simples, segue-se que

1. Se H_0^ℓ é verdadeira, tem-se que $\sum_{i=1}^m \alpha_i^2 = 0$ e que

$$E[SQL] = (m-1)\sigma^2 \; ; \quad \frac{SQL}{\sigma^2} \sim \chi^2(m-1)$$

   Se H_0^ℓ não é verdadeira,

$$E[SQL] = (m-1)\sigma^2 + n\sum_{i=1}^m \alpha_i^2$$

2. Se H_0^c é verdadeira, tem-se que $\sum_{j=1}^n \beta_j^2 = 0$ e que

$$E[SQC] = (n-1)\sigma^2 \; ; \quad \frac{SQC}{\sigma^2} \sim \chi^2(n-1)$$

   Se H_0^c não é verdadeira,

$$E[SQC] = (n-1)\sigma^2 + m\sum_{j=1}^n \beta_j^2$$

3. Em qualquer caso,

$$E[SQR] = (n-1)(m-1)\sigma^2 \; ; \quad \frac{SQR}{\sigma^2} \sim \chi^2[(n-1)(m-1)]$$

Testes de hipóteses:

1. $H_0^\ell : \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$: usa-se a estatística de teste

$$F^\ell = \frac{SQL/(m-1)}{SQR/[(m-1)(n-1)]} \mid H_0^\ell \sim F[m-1,(m-1)(n-1)]$$

devendo rejeitar-se H_0^ℓ para valores grandes da estatística de teste. Nesse caso, conclui-se que não existe evidência para suportar a ausência de efeito linha

2. $H_0^c : \beta_1 = \beta_2 = \cdots = \beta_n = 0$: usa-se a estatística de teste

$$F^c = \frac{SQC/(n-1)}{SQR/[(m-1)(n-1)]} \mid H_0^c \sim F[n-1,(m-1)(n-1)]$$

devendo rejeitar-se H_0^c para valores grandes da estatística de teste. Nesse caso, conclui-se que não existe evidência para suportar a ausência de efeito coluna

Apresentação habitual dos resultados:

Tabela da ANOVA —classificação dupla cruzada, uma observação por célula

Origem da variação	Soma de quadrados	Graus de liberdade	Médias quadráticas
Linhas	SQL	$m-1$	$MQL = SQL/(m-1)$
Colunas	SQC	$n-1$	$MQC = SQC/(n-1)$
Residual	SQR	$(m-1)(n-1)$	$MQR = SQL/[(m-1)(n-1)]$
Total	SQT	$mn-1$	$F^{\ell} = MQL/MQR$
			$F^{c} = MQC/MQR$

Exemplo 4.12 *Foram utilizados cinco tipos de material (efeito-linha) para investigar o tempo de corte de quatro máquinas de diferentes marcas (efeito-coluna). Os dados com os tempos registados foram:*

Tipo de material	Máquina A	Máquina B	Máquina C	Máquina D
1	12	20	13	11
2	2	14	7	5
3	8	17	13	10
4	1	12	8	3
5	7	17	14	6

Tabela da ANOVA

Origem da variação	Soma de quadrados	Graus de liberdade	Médias quadráticas
Linhas	$SQL = 184$	$5 - 1$	$MQL = 46$
Colunas	$SQC = 310$	$4 - 1$	$MQC = 103.33$
Residual	$SQR = 24$	12	$MQR = 2$
Total	$SQT = 518$	$5 \times 4 - 1$	$F^\ell = 23$
			$F^c = 51.67$

valores-p:

$$P(F^\ell > 23 \mid H_0^\ell) = 1.5 \times 10^{-5}$$
$$P(F^c > 51.67 \mid H_0^c) = 3.9 \times 10^{-7}$$

É de rejeitar a hipótese de que os tempos médios de corte são os mesmos para os diferentes tipos de máquinas e de materiais.