

## TESTE DE HIPÓTESES

- A população  $X$  tem uma função de distribuição  $F$  que apresenta **aspectos desconhecidos**.
- A ideia associada ao teste de hipóteses é **estabelecer uma conjectura sobre os aspetos desconhecidos da distribuição e verificar se a informação existente suporta ou não esta conjectura**.
- Primeiros conceitos:
  - **Hipótese estatística (definição 8.1)**: qualquer conjectura sobre aspectos desconhecidos de  $F$ .
  - **Hipótese paramétrica (definição 8.2)**: qualquer conjectura que diga **apenas** respeito ao **parâmetro** (escalar ou vector) da distribuição. **A forma funcional de  $F$  é suposta conhecida**.

Neste capítulo apenas nos vamos interessar pelas **hipóteses paramétricas**

**Exemplo** – Considere-se uma população da qual se estuda um atributo representado pela v.a.  $X$ .

- A conjectura “ $X$  é uma variável aleatória com distribuição exponencial” constitui uma **hipótese estatística não paramétrica**.
- Caso se assuma que  $X$  tem distribuição exponencial, a conjectura “ $\mu = 5$ ” corresponde a uma **hipótese estatística paramétrica**.

- **Hipótese nula e hipótese alternativa**

Seja  $X \sim f(x|\theta)$ ,  $\theta \in \Theta$ ,  $\theta$  desconhecido. Qualquer hipótese paramétrica estabelece uma partição do espaço-parâmetro  $\Theta$  em  $\Theta_0$  e  $\Theta_1$  (Partição:  $\Theta_0 \cup \Theta_1 = \Theta$ ,  $\Theta_0 \cap \Theta_1 = \emptyset$ )

$H_0 : \theta \in \Theta_0$  **hipótese nula** (geralmente, corresponde ao *satus quo* – algo que se pretende manter)

$H_1 : \theta \in \Theta_1$  **hipótese alternativa**.

Uma hipótese estatística diz-se **simples** quando o respectivo subconjunto do espaço-parâmetro é **formado apenas por um elemento**; diz-se **composta** no caso contrário.

**Exemplo 8.2 (livro)** – Pretende aferir-se se determinada **moeda é equilibrada**. Codifica-se a saída de “face” pelo valor 1. O modelo teórico é assim dado por  $X \sim Ber(\theta)$ .

Espaço-parâmetro?

Hipótese nula?

Hipótese alternativa?

- **Teste de hipóteses** (definição 8.3 do livro)

- Um **teste de hipóteses é uma regra** que permite especificar um subconjunto do espaço-amostra,  $W \subset X$ , tal que:

- se  $(x_1, x_2, \dots, x_n) \in W$  rejeita-se  $H_0$  (logo utiliza-se  $H_1$ );
- se  $(x_1, x_2, \dots, x_n) \notin W$  não se rejeita  $H_0$ .

O conjunto  $W$  chama-se **região crítica** ou **região de rejeição**. O complementar do conjunto  $W$  representa-se por  $\bar{W}$  e corresponde à **região de não rejeição**.

- Assim um teste estatístico introduz uma partição do espaço-amostra em duas regiões,  $W$  e  $\bar{W}$ , tais que  $W \cup \bar{W} = X$  e  $W \cap \bar{W} = \emptyset$ .

- Em quase todos os casos de interesse prático trabalha-se com uma **estatística-teste**  $T = T(X_1, X_2, \dots, X_n)$

Nestas circunstâncias, a região de rejeição é definida em termos da estatística, isto é,

se  $t \in W_T$  rejeita-se  $H_0$ ;      se  $t \notin W_T$  não se rejeita  $H_0$ .

O conjunto  $W_T$  continua a chamar-se **região de rejeição** ou **região crítica**.

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

- A **hipótese nula**,  $H_0$ , que é defendida até a evidência mostrar o contrário.
  - A **hipótese alternativa**,  $H_1$ , que é adotada se a hipótese nula for rejeitada.
  - Uma **estatística-teste**,  $T = T(X_1, X_2, \dots, X_n)$  e **uma região crítica**,  $W_T$ .
- **Exemplo** – Retomando-se o exemplo anterior (moeda é equilibrada?) um possível teste poderia consistir em observar  $n = 100$  lançamentos da moeda e definir a estatística-teste,  $T$ , como sendo o nº de vezes em que saiu caras. A região de rejeição poderia ser dada por  $W_T = \{t: t < 40 \vee t > 60\}$ .
- Note-se que, qualquer que seja a decisão que se venha a tomar em função da amostra observada, ela poderá sempre conduzir-nos a uma conclusão errada. Com  $\theta = 0.5$  pode-se perfeitamente observar um valor de  $T \in W_T$  (a probabilidade é cerca de 3.5%) e, portanto, rejeitar  $H_0$ .
- De forma “oposta”, com  $\theta \neq 0.5$ , pode-se observar um valor de  $T \notin W_T$ . Para calcular a probabilidade de um erro desta natureza torna-se necessário postular um valor concreto para  $\theta$ , por exemplo  $\theta = 0.65$ . Neste caso a probabilidade será cerca de 17.24%

- Sendo o teste de hipóteses efectuado com base numa amostra e não no universo, a decisão a tomar pode estar errada. Devem assim considerar-se dois tipos de erros.

### Erros de 1ª e 2ª espécie

Situação	$H_0$ verdadeira	$H_0$ falsa
<b>Decisão tomada</b>		
<b>Rejeitar <math>H_0</math></b>	<b>Erro de 1ª espécie</b>	<b>Decisão correcta</b>
<b>Não rejeitar <math>H_0</math></b>	<b>Decisão correcta</b>	<b>Erro de 2ª espécie</b>

## Hipótese simples contra hipótese simples

- Considere-se o caso mais simples em que o **espaço-parâmetro é composto por apenas dois pontos**,  $\Theta = \{\theta_0, \theta_1\}$  e suponha-se que se quer testar,

$$H_0 : \theta = \theta_0 \quad \text{contra} \quad H_1 : \theta = \theta_1.$$

- Neste enquadramento, torna-se possível calcular a probabilidade associada a cada um dos tipos de erro, definida a região crítica  $W$ .
- Para o teste associado com a região crítica  $W$ , a **probabilidade de cometer o erro de 1ª espécie**,

$$\alpha = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) = P[(X_1, X_2, \dots, X_n) \in W \mid \theta = \theta_0],$$

designa-se por **dimensão do teste**.

- **A probabilidade de cometer o erro de 2ª espécie** é dada por,

$$1 - \beta = P(\text{aceitar } H_0 \mid H_0 \text{ falsa}) = P[(X_1, X_2, \dots, X_n) \notin W \mid \theta = \theta_1].$$

Esta probabilidade não tem designação especial, mas a probabilidade do acontecimento contrário, isto é, a **probabilidade de não cometer o erro de 2ª espécie**,

$$\beta = P(\text{rejeitar } H_0 \mid H_0 \text{ falsa}) = P[(X_1, X_2, \dots, X_n) \in W \mid \theta = \theta_1].$$

designa-se por **potência do teste**.

Probabilidade dos erros de 1ª e 2ª espécie e decisões correctas

<b>Decisão tomada</b>	$H_0$ verdadeira	$H_0$ falsa
Rejeitar $H_0$	<b>Erro de 1ª espécie</b> Probabilidade = $\alpha$	Decisão correcta Probabilidade = $\beta$
Não rejeitar $H_0$	Decisão correcta Probabilidade = $1 - \alpha$	<b>Erro de 2ª espécie</b> Probabilidade = $1 - \beta$

**Exemplo:** Seja  $X$  uma população  $N(\mu, \sigma^2 = 4)$ . Pretende-se testar  $H_0 : \mu = 10$  contra  $H_1 : \mu = 14$ .

Admita-se por agora que se tinha fixado  $W = \{\bar{x} : \bar{x} > k\}$  sendo  $k=12.5$  e que a amostra tinha dimensão  $n = 2$ .

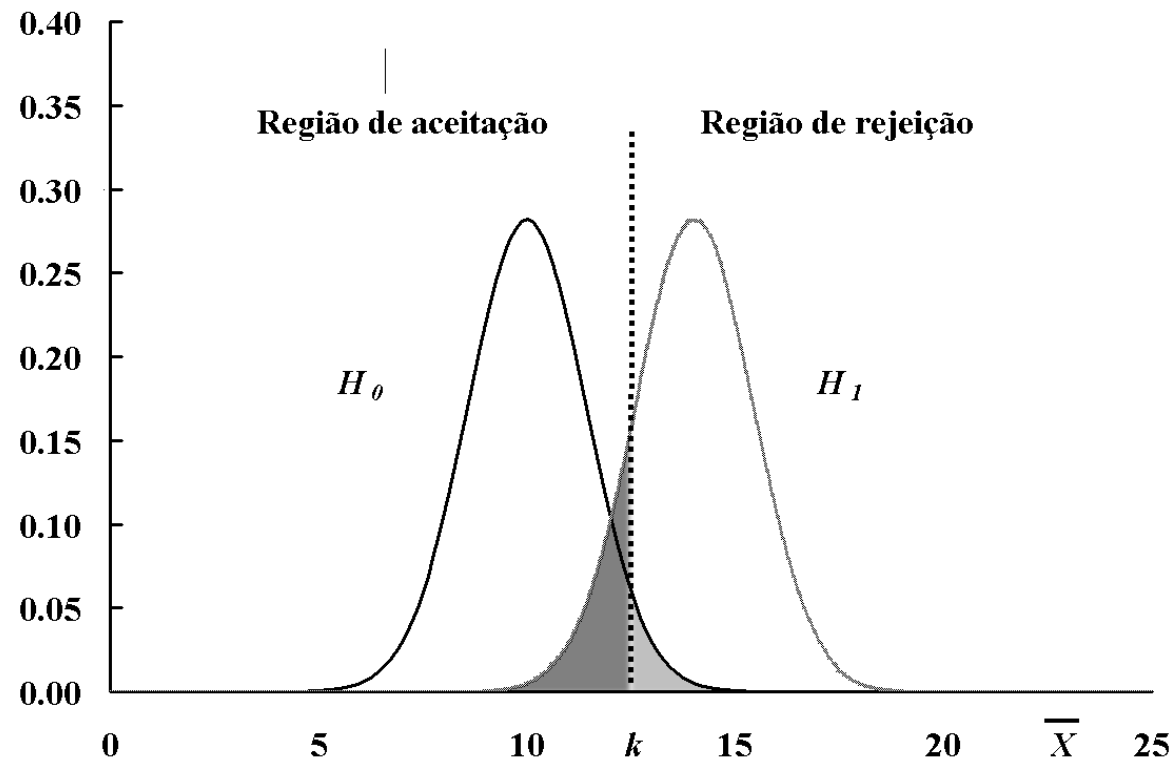
$$\begin{aligned}\alpha &= P(\text{Rej } H_0 | H_0 \text{ verd}) = P(\bar{X} > 12.5 | \mu = 10) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{12.5 - 10}{2/\sqrt{2}} \mid \mu = 10\right) = 1 - \Phi\left(\frac{2.5}{\sqrt{2}}\right) \\ &= 1 - \Phi(1.77) = 1 - 0.9616 = 0.0384\end{aligned}$$

$$\begin{aligned}1 - \beta &= P(\text{n\~ao rej } H_0 | H_0 \text{ falsa}) = P(\bar{X} \leq 12.5 | \mu = 14) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{12.5 - 14}{\frac{2}{\sqrt{2}}} \mid \mu = 14\right) \\ &= \Phi\left(-\frac{1.5}{\sqrt{2}}\right) = 1 - \Phi(1.06) = 1 - 0.8554 = 0.1446\end{aligned}$$

Alternativamente, em vez da probabilidade do erro de 2ª espécie, pode-se calcular a potência do teste  $\beta = 1 - 0.1446 = 0.8554$  ou, se não se tivesse previamente calculado  $1 - \beta$ , fazendo

$$\beta = P(\text{Rej } H_0 | H_0 \text{ falsa}) = P(\bar{X} > 12.5 | \mu = 14) = 1 - \Phi\left(\frac{12.5 - 14}{\frac{2}{\sqrt{2}}}\right) = 0.8554$$





**Fig. 8.3** – Probabilidade de um erro de 1ª espécie (sombreado claro) e de um erro de 2ª espécie (sombreado mais escuro) quando  $k=12.5$  ( $n = 2$ )

O que permite sublinhar alguns aspetos importantes:

- A redução das duas probabilidades (ou de uma delas, supondo a outra fixa) só se consegue aumentando a dimensão da amostra.
- Alterando a fronteira da região crítica, isto é o valor de  $k$ , obtêm-se outros valores para  $\alpha$  e  $\beta$ .
- Neste caso, aumentando (diminuindo) o valor de  $k$ , diminui-se (aumenta-se) a probabilidade de erro de 1ª espécie e em contrapartida aumenta-se (diminui-se) a probabilidade de erro de 2ª espécie para o mesmo valor de  $n$ .

	Dim. amostra	RC $W = \{\bar{x} : \bar{x} > k\}$	$\alpha = P(T \in W_T \mid \mu = 10)$	$1 - \beta =$ $= P(T \notin W_T \mid \mu = 14)$
a)	n=1	$W = \{x_1 : x_1 > 12.5\}$	0.1056	0.2266
b)	n=2	$W = \{(x_1, x_2) : \bar{x} > 12.5\}$	0.0384 ↓	0.1446 ↓
c)	n=2	$W = \{(x_1, x_2) : \bar{x} > 13.5\}$	0.0068 ↓	0.3632 ↑

Não se podendo minimizar simultaneamente a probabilidade dos 2 tipos de erros, torna-se necessário introduzir um critério alternativo.

A filosofia que se vai seguir consiste em

- Fixar a probabilidade de um erro de 1ª espécie num valor adequado
- Fixado  $\alpha$ , definir a região de rejeição que minimiza  $1 - \beta$  ou que maximiza a potência do teste

Esta filosofia baseia-se em 2 considerações muito importantes:

- As consequências dos erros de 1ª e 2ª espécies podem ser bem diferentes e, portanto, não existe fundamento para lhes dar peso idêntico;
- O lema de Neyman-Pearson permite definir o teste mais potente nas condições anteriores

O lema de Neyman-Pearson será objeto de tratamento mais adiante. “Saltando” as especificidades matemáticas, pode-se enunciar a seguinte regra intuitiva:

**No teste de médias, variâncias ou proporções a região de rejeição está do lado da alternativa quando se utiliza a estatística “natural”.**

Assim, o procedimento prático para realizar determinado teste será:

- Fixar  $\alpha$  num valor adequado ao problema que se quer resolver
- Escolher a estatística de teste (cuja distribuição, sob  $H_0$ , deve ser conhecida)
- Definir a região de rejeição  $W$
- Realizar o teste e concluir (só nesta última etapa se utiliza a amostra observada)

**Exemplo 1:** Retome-se o exemplo anterior  $H_0 : \mu = 10$  contra  $H_1 : \mu = 14$  numa população normal de variância conhecida e igual a 4. Vai utilizar-se uma amostra de dimensão  $n = 2$

- Fixar  $\alpha$ , por exemplo  $\alpha = 0.05$
- Teste para a média da população  $\rightarrow$  estatística “natural”:  $\bar{X}$

Sendo a variância conhecida utiliza-se, sob  $H_0$ ,  $Z = \frac{\bar{X}-10}{2/\sqrt{2}} \sim n(0, 1)$

Está escolhida a estatística de teste.

– Região de rejeição: Como a alternativa é  $H_1 : \mu = 14 > 10 = \mu_0$  vem  $W = \{(x_1, x_2) : \bar{x} > k\}$

Para determinar  $k$ , faz-se

$$\alpha = P(\text{Rej } H_0 | H_0 \text{ verd}) = P(\bar{X} > k | \mu = 10) = P\left(Z > \frac{k - 10}{2/\sqrt{2}}\right) = 1 - \Phi\left(\frac{k - 10}{\sqrt{2}}\right)$$

Logo

$$\Phi\left(\frac{k-10}{\sqrt{2}}\right) = 1 - \alpha = 0.95 \quad \text{e portanto } \frac{k-10}{\sqrt{2}} = 1.645 \text{ isto é } k = 10 + 1.645 \sqrt{2} = 12.3264$$

A região de rejeição é então  $W = \{(x_1, x_2) : \bar{x} > 12.3264\}$

– Realização do teste: Calcula-se  $\bar{x}$  com base na amostra observada e decide-se em conformidade.

**Alternativamente, e de forma talvez mais simples**, pode-se definir a região de rejeição em termos da estatística de teste  $Z = \frac{\bar{X}-10}{2/\sqrt{2}}$ , fazendo-se  $W = \{(x_1, x_2) : z > 1.645\}$

Observada a amostra (10.7, 13.3) obtém-se  $\bar{x} = 12$  ou  $z = \frac{12-10}{\sqrt{2}} = 1.4142$  (valores equivalentes) que conduzem à não rejeição de  $H_0$ .

**Exemplo 2:** Observou-se a amostra (15.2, 20.1, 17.4, 15.3) de uma população normal com o propósito de testar  $H_0: \mu = 20$  contra  $H_1: \mu = 15$  .

- a) Assumindo que a variância da população é 16, realize o teste para  $\alpha = 0.01$ .
- b) Qual a potência do teste definido na alínea anterior?
- c) Considerando agora que a variância da população é desconhecida, efectue o teste proposto para  $\alpha = 0.05$ .

a)  $H_0: \mu = 20$  contra  $H_1: \mu = 15$

– Teste para a média da população  $\rightarrow$  estatística “natural”:  $\bar{X}$

Sendo a variância conhecida utiliza-se, sob  $H_0$ ,  $Z = \frac{\bar{X}-20}{\sqrt{4}/\sqrt{4}} = (\bar{X} - 20) \sim n(0, 1)$

–  $H_1: \mu = 15 < 20$  logo  $W = \{(x_1, \dots, x_4): z \leq k^*\}$  ou  $W = \{(x_1, \dots, x_4): \bar{x} \leq k\}$

$k^*$ :  $\alpha = 0.01 = P(\text{Rej } H_0 | H_0 \text{ verd}) = P(Z \leq k^*) = \Phi(k^*)$  logo  $k^* = -2.326$  e portanto  $W = \{(x_1, \dots, x_4): z \leq -2.326\}$ . Como  $z = \bar{x} - 20 = -3 < -2.326$  logo **rejeita-se  $H_0$** .

**alternativamente**

$k$ :  $\alpha = 0.01 = P(\text{Rej } H_0 | H_0 \text{ verd}) = P(\bar{X} \leq k) = \Phi(k - 20)$  logo  $k - 20 = -2.326$  e portanto  $W = \{(x_1, \dots, x_4): \bar{x} \leq 17.674\}$  e, como  $\bar{x} = 17 < 17.674$ , **rejeita-se  $H_0$** .

b)

$$\beta = P(\text{Rej } H_0 | H_0 \text{ falsa}) = P(\bar{X} \leq 17.674 | \mu = 15) = \Phi\left(\frac{17.674 - 15}{\sqrt{4}/\sqrt{4}}\right) = \Phi(2.674) = 0.9962$$

c)

– Teste para a média da população → estatística “natural”:  $\bar{X}$

Sendo a variância desconhecida utiliza-se, sob  $H_0$ ,  $T = \frac{\bar{X}-20}{s'/\sqrt{4}} = \frac{\bar{X}-20}{s'/2} \sim t_{(3)}$

–  $H_1: \mu = 15 < 20$  logo  $W = \{(x_1, \dots, x_4): t \leq k^*\}$  ou  $W = \{(x_1, \dots, x_4): \bar{x} \leq k\}$

$k^*$ :  $\alpha = 0.05 = P(\text{Rej } H_0 | H_0 \text{ verd}) = P(T \leq k^*)$  logo  $k^* = -2.353$  e portanto  $W =$   
 $\{(x_1, \dots, x_4): t \leq -2.353\}$ . Como  $t = \frac{\bar{x}-20}{s'/2} = \frac{17-20}{2.3022/2} = -2.606 < -2.353$ , rejeita-se  $H_0$ .



## Teste mais potente. Lema de Neyman-Pearson

- Na impossibilidade de minimizar simultaneamente os dois tipos de erros, torna-se necessário definir uma abordagem que permita considerá-los de alguma forma.
  - **Filosofia de Neyman-Pearson:** fixar a probabilidade associada ao erro de 1ª espécie e minimizar a probabilidade do erro de 2ª espécie ou, dito de outra forma, **fixar a dimensão** do teste e **maximizar a sua potência**.
  - Esta forma de proceder atribui **maior importância ao erro de 1ª espécie**, uma vez que é fixado num valor conveniente, enquanto a potência vem a maior possível dentro dos condicionantes existentes.
  - Geralmente, fixa-se a dimensão do teste,  $\alpha$ , em 0.1, 0.05 ou 0.01. Apenas se rejeita  $H_0$  se houver uma forte evidência estatística contra esta hipótese.
- **Teste mais potente:** Fixada a probabilidade de cometer o erro de 1ª espécie (dimensão do teste) o teste mais potente é aquele em que se escolhe a região crítica que minimiza a probabilidade do erro de 2ª espécie (ou seja maximiza a potência).
- Como definir testes mais potentes?  
Recorrendo ao **lema de Neyman-Pearson**.

### Lema de Neyman-Pearson – (Livro)

$X \sim f(x|\theta)$ ,  $\theta \in \{\theta_0, \theta_1\}$  e  $(X_1, X_2, \dots, X_n)$  amostra casual dessa população,

Seja  $C > 0$ , e  $W \subset \mathfrak{R}^n$  o conjunto do espaço-amostra definido por

$$\frac{\prod_{i=1}^n f(x_i | \theta_1)}{\prod_{i=1}^n f(x_i | \theta_0)} > C \Leftrightarrow (x_1, x_2, \dots, x_n) \in W, \quad \text{em que } P[(X_1, X_2, \dots, X_n) \in W | \theta = \theta_0] = \alpha.$$

O teste associado com a região crítica  $W$  é o teste mais potente de dimensão  $\alpha$  para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ .

#### Exemplo anterior:

Seja  $X$  uma população  $N(\mu, \sigma^2)$  com  $\sigma^2 = 4$ .

Quer-se testar  $H_0 : \mu = 10$  contra  $H_1 : \mu = 14$ .

Amostra casual de dimensão  $n$ .

Deduza a região crítica mais potente de dimensão 0.05

## Fase 0

Construir a função de verosimilhança  $\prod_{i=1}^n f(x_i | \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}$

## Fase 1

Escrever a condição sobre a razão de verosimilhanças e simplificá-la

$$\frac{\prod_{i=1}^n f(x_i | \mu_1)}{\prod_{i=1}^n f(x_i | \mu_0)} = \frac{L(\mu_1 | x_1, x_2, \dots, x_n)}{L(\mu_0 | x_1, x_2, \dots, x_n)} > C \Leftrightarrow \bar{x} > k$$

## Fase 2

Obter o valor fronteira, utilizando a dimensão do teste

$$\alpha = P(\bar{X} > k | \mu = \mu_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\frac{k - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\text{logo } \Phi\left(\frac{k - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \alpha \Rightarrow k = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \text{ onde } \Phi(z_\alpha) = 1 - \alpha$$

Se  $\alpha = 0.05$  e  $n = 2$  vem  $k = 10 + 1.645 \times \frac{2}{\sqrt{2}} = 12.3264$  ficando definida a região crítica

$$W = \{(x_1, x_2) : \bar{x} > 12.3264\}$$

Uma vez obtida a região de rejeição realiza-se o teste calculando a média observada da amostra e comparando-a com o valor de referência.

Por simples curiosidade, pode-se partir de  $k = 12.3264$  e, efetuando-se as operações inversas daquelas que se efetuaram na fase 1, obter o valor inicial  $C = \exp\left(-\frac{n(\mu_1^2 - \mu_0^2) - 2nk(\mu_1 - \mu_0)}{2\sigma_0^2}\right) \approx 1.92$

Neste caso, rejeitamos  $H_0$  quando a verosimilhança de  $\mu = 14$  é 1.92 vezes superior à verosimilhança de  $\mu = 10$

## Resumindo e voltando ao caso geral

- A primeira fase de aplicação do lema de Neyman-Pearson permite definir 2 aspectos essenciais:
  - **a estatística-teste**
  - **o tipo de região de rejeição** que lhe está associado: inferior ou superior a uma constante  $k$ .
- A segunda fase serve para definir o valor da constante  $k$  em função da dimensão do teste que se encontra pré-fixada.
- **CUIDADO:** O lema só tem utilidade prática quando se conhece **a distribuição por amostragem da estatística-teste**.
- **O lema serve de base à regra intuitiva:** No teste de médias, variâncias ou proporções a região de rejeição está do lado da alternativa quando se utiliza a estatística “natural”.

## Testes envolvendo hipóteses compostas unilaterais

Depois de estudar os testes de hipóteses numa situação pouco comum, em que o espaço-parâmetro é composto por apenas dois pontos,  $\Theta = \{\theta_0, \theta_1\}$  vamos agora testar uma hipótese simples contra alternativa **unilateral**)

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

Em situações deste tipo:

- Nada se altera no que se refere à probabilidade do erro de 1ª espécie já que apenas depende de  $H_0$ .

$$\text{Recorde-se que } \alpha = P(\text{Rej } H_0 | H_0 \text{ verd}) = P(\text{Rej } H_0 | \theta = \theta_0)$$

- Deixa de haver um valor para a probabilidade de um erro de 2ª espécie e para a potência do teste, sendo necessário definir o conceito de **função potência** (definição 8.7 do livro)

A função potência do teste  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta > \theta_0$  com região crítica  $W$ , é dada por,  
 $\beta(\theta) = P(\text{rejeitar } H_0 | \theta) = P[(X_1, X_2, \dots, X_n) \in W | \theta]$  para  $\theta \in \Theta_1$ , onde  $\Theta_1 = \{\theta : \theta > \theta_0\}$ .

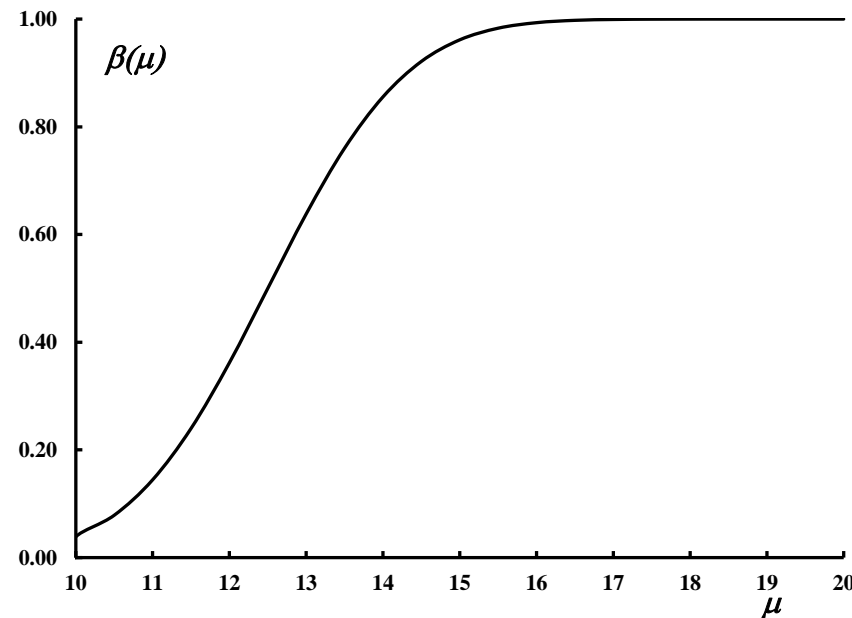
- Em vez de teste mais potente fala-se agora em teste **uniformemente mais potente (UMP)** – Uma definição formal está dada no livro (definição 8.8). Em termos intuitivos, **um teste é UMP se for o mais potente contra cada uma das hipóteses simples abrangidas por  $H_1$ .**

- Nas situações que vamos considerar existe, geralmente, um teste UMP que é obtido “generalizando” o lema de Neyman-Pearson (teorema de Karlin-Rubin) mas **nada obriga a que existam sempre testes UMP!!!**
- **A regra intuitiva (médias, variâncias e proporções) mantém-se** já que a relação entre qualquer valor da alternativa e o valor especificado pela hipótese nula é sempre a mesma: todos superiores ou todos inferiores. Daí a importância de se considerar uma alternativa **unilateral**.

**Exemplo:** Adapte-se o exemplo que se utilizou anteriormente:

- Universo normal com variância conhecida  $\sigma^2$
  - $H_0 : \mu = \mu_0$  contra  $H_1 : \mu > \mu_0$
  - amostra de dimensão  $n$
- 
- Estatística natural  $\rightarrow \bar{X}$
  - Região crítica: Como em  $H_1$  se tem  $\mu > \mu_0$ , a região crítica será dada por  $\bar{x} > k$  (generalização do lema de Neyman-Pearson ou regra intuitiva).
  - Para obter  $k$ , fixa-se  $\alpha$  e utiliza-se  $\alpha = P(\bar{X} > k \mid \mu = \mu_0)$ , obtendo-se o teste UMP.
  - A função potência vem  $\beta(\mu) = P(\bar{X} > k \mid \mu) = 1 - \Phi\left(\frac{k - \mu}{\sigma / \sqrt{n}}\right)$ ,  $\mu > \mu_0$ .

- Exemplifique-se para  $\mu_0 = 10$ ,  $\sigma = 2$ ,  $n = 2$ ,  $\alpha = 0.05$ 
  - Para  $\alpha = 0.05$ ,  $k = 10 + 1.645 \times 2 / \sqrt{2} = 12.3264$  (já tinha sido visto)
  - $\frac{k - \mu}{\sigma / \sqrt{n}} = \frac{12.3264 - \mu}{1.4142}$  logo  $\beta(\mu) = 1 - \Phi\left(\frac{12.3264 - \mu}{1.4142}\right)$
  - Função para  $\mu > 10$  (para  $\mu = 10$  tem-se  $\beta(\mu) = \alpha$ ). Quanto mais afastado de 10 estiver  $\mu$ , maior a potência.



**Fig. 8.6** – Função potência para testar  $H_0 : \mu = 10$  contra  $H_1 : \mu > 10$ .

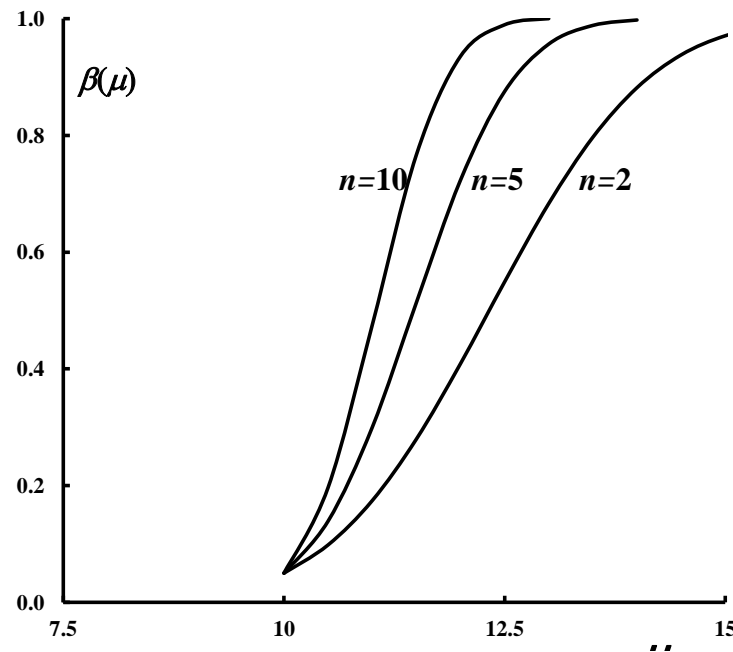


### O que acontece à função potência quando $n$ cresce, mantendo-se $\alpha$ ?

Como é intuitivo a potência também cresce já que se dispõe de mais informação.

No exemplo vem

$$\beta_n(\mu) = P[\bar{X} > k(n) | \mu] = P\left[\frac{\bar{X} - \mu}{2/\sqrt{n}} > \frac{k(n) - \mu}{2/\sqrt{n}}\right], \quad \mu > 10, \text{ com } k(n) = 10 + 1.645 \times \frac{2}{\sqrt{n}},$$



**Fig. 8.7** – Função potência no teste de  $H_0 : \mu = 10$  contra  $H_1 : \mu > 10$  para amostras de dimensão  $n = 2, 5$  e  $10$ .

Considere-se agora que quer  $H_0$  quer  $H_1$  são hipóteses compostas unilaterais

$H_0 : \theta \leq \theta_0$  contra  $H_1 : \theta > \theta_0$  ou  $H_0 : \theta \geq \theta_0$  contra  $H_1 : \theta < \theta_0$

- Considere-se o primeiro caso  $H_0 : \theta \leq \theta_0$  contra  $H_1 : \theta > \theta_0$  já que o segundo é em tudo semelhante
- A probabilidade do erro de 1ª espécie passa a depender do valor de  $\theta$  em  $H_0$  que se considere. Define-se a **dimensão do teste** como sendo o valor máximo desta probabilidade. Este valor corresponde a utilizar o ponto fronteira  $\theta = \theta_0$ . Em termos práticos está-se a proceder como se se tivesse  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta > \theta_0$ .

Consequência importante: **A igualdade está sempre em  $H_0$**

- **Exemplo:** Universo normal com variância conhecida  $\sigma^2$

- $H_0: \mu \leq 10$  contra  $H_1: \mu > 10$

- amostra de dimensão  $n$

$$\alpha = P(\text{Rej } H_0 | H_0 \text{ verd}) = P(\text{Rej } H_0 | \mu \leq 10)$$

Estatística natural  $\rightarrow \bar{X}$

Região crítica: Como qualquer valor em  $H_1$  é superior a qualquer valor de  $H_0$ , a região crítica será dada por  $\bar{x} > k$

Para obter  $k$ , fixa-se  $\alpha$ , a dimensão do teste, e utiliza-se  $\alpha = P(\bar{X} > k | \mu = 10)$ , obtendo-se o teste UMP.

Repare-se que a probabilidade de um erro de 1ª espécie é dada por

$$\alpha(\mu) = P(\bar{X} > k | \mu) = 1 - \Phi\left(\frac{k - \mu}{\sigma/\sqrt{n}}\right) \quad \mu \leq 10$$

Quanto maior for  $\mu$ , menor  $\frac{k - \mu}{\sigma/\sqrt{n}}$  e portanto menor  $\Phi\left(\frac{k - \mu}{\sigma/\sqrt{n}}\right)$  logo maior  $\alpha(\mu)$ . Escolher o pior caso equivale assim a escolher  $\mu = 10$ .

## Testes bilaterais

$$H_0 : \theta = \theta_0 \text{ contra } H_1 : \theta \neq \theta_0$$

- Numa situação destas é fácil entender que **não existe, na generalidade dos casos, um teste UMP**. Para tal retome-se o exemplo que se tem vindo a tratar e admita-se que se queria testar,

$$H_0 : \mu = 10 \text{ contra } H_1 : \mu \neq 10,$$

para uma população normal de variância igual a 4. Daquilo que se viu anteriormente, quando  $\mu < 10$  a região deveria ser **toda para a esquerda** e, quando  $\mu > 10$ , **toda para a direita**. Esta situação não pode ocorrer simultaneamente.

- Recorre-se a uma regra intuitiva que consiste em definir uma região de rejeição nas duas caudas da distribuição da estatística-teste, atribuindo igual probabilidade, seja  $\alpha/2$ , a cada uma das duas sub-regiões.

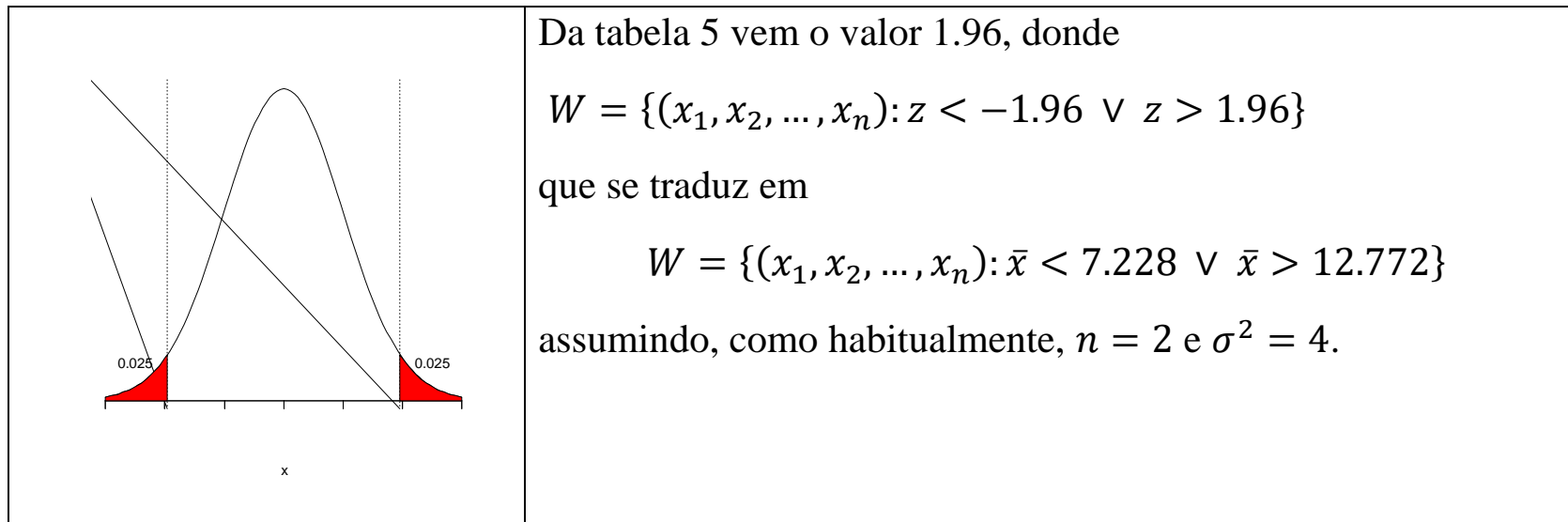
- Retomando o exemplo de população normal com variância conhecida

$$H_0 : \mu = 10 \text{ contra } H_1 : \mu \neq 10,$$

Estatística natural  $\rightarrow \bar{X}$  utilizando-se  $Z = \frac{\bar{X}-10}{\sigma/\sqrt{n}} \sim n(0, 1)$

Fixar  $\alpha$ , por exemplo  $\alpha = 0.05$

Região crítica: Sendo  $H_1$  bilateral utiliza-se a regra empírica que consiste em considerar a região de rejeição nas 2 caudas da distribuição



## Valor-p

- Num teste de hipóteses, **fixada a dimensão**  $\alpha$ , o resultado consiste em rejeitar (ou não rejeitar)  $H_0$ , não se tendo em conta se o valor da estatística-teste se situa longe ou perto do limiar de rejeição.
- O valor-p é uma forma alternativa de reportar o resultado de um teste que permite ultrapassar esta limitação.
- **valor-p** (definição 8.9 do livro) - Seja  $T(x_1, x_2, \dots, x_n) = t_{\text{obs}}$  o valor observado (com a amostra concreta) para a estatística de teste. **O valor-p é a probabilidade,  $p_{\text{obs}}$ , de observar um valor tão ou mais desfavorável para  $H_0$ , admitindo  $H_0$  verdadeira.**

**O valor-p mede a evidência que os dados fornecem a favor de  $H_0$ .** Assim, **quanto menor for o valor-p** ( $p_{\text{obs}}$ ) menor é a consistência dos dados com a hipótese e, portanto, **mais se rejeita  $H_0$** ; por exemplo, se  $p_{\text{obs}} = 0.0001$ , rejeitamos sem problemas  $H_0$  mas se  $p_{\text{obs}} = 0.23$  não rejeitaremos  $H_0$ . O problema surge com os valores “cinzentos”, por exemplo,  $p_{\text{obs}} = 0.053$ . Caso tenhamos fixado previamente a dimensão desejada para o teste comparamos o valor-p com  $\alpha$ :

Valor-p  $< \alpha \rightarrow$  Rejeita-se  $H_0$

Valor-p  $> \alpha \rightarrow$  Não se rejeita  $H_0$

### Para calcular o valor- $p$

- Obter a distribuição da estatística de teste assumindo que  $H_0$  (ou o seu valor limite se for uma hipótese composta) é verdadeira.
  - Definir acontecimento mais improvável do que o observado (mais para o lado da alternativa)
  - Calcular a sua probabilidade.
- **Exemplos:**  $X \sim N(\mu; 4)$ ,  $n = 16$   $\sigma^2 = 4$  tendo-se observado  $\bar{x} = 9$ .
    1.  $H_0 : \mu = 10$  (ou  $\mu \geq 10$ ) contra  $H_1 : \mu < 10$
    2.  $H_0 : \mu = 10$  contra  $H_1 : \mu \neq 10$

Em qualquer dos casos a estatística de teste vai ser  $\bar{X}$  que tem distribuição  $\bar{X} \sim N(\mu; 1/4)$  ou, de forma equivalente  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim n(0, 1)$ .

1.  $H_0 : \mu = 10$  contra  $H_1 : \mu < 10$

Nesta situação os casos **tão ou mais desfavoráveis** correspondem a **observar uma média amostral inferior ou igual ao valor observado** (valor mais pequenos de  $\mu$  tendem a originar amostras com médias amostrais inferiores)  $\rightarrow p_{obs} = P(\bar{X} \leq 9 | \mu = 10) = \Phi\left(\frac{9-10}{2/\sqrt{16}}\right) = \Phi(-2) = 0.0228$ .

2.  $H_0 : \mu = 10$  contra  $H_1 : \mu \neq 10$

Nesta situação os casos **tão ou mais desfavoráveis** correspondem a **observar uma média amostral mais “afastada” de 10 do que o valor observado, 9, quer seja para valores superiores quer para valores inferiores**. Sendo o afastamento observado  $|9 - 10| = 1$  vem

$$\begin{aligned} p_{obs} &= P(|\bar{X} - \mu| \geq 1 | \mu = 10) = P((\bar{X} - \mu) \leq -1 | \mu = 10) + P((\bar{X} - \mu) \geq 1 | \mu = 10) \\ &= 2 \times P((\bar{X} - \mu) \geq -1 | \mu = 10) \quad \text{simetria} \\ &= 2 \times \left( 1 - P\left( \frac{\bar{X} - 10}{\frac{\sigma}{\sqrt{n}}} \leq \frac{1}{\sqrt{16}} \right) \right) = 2 \times (1 - \Phi(2.0)) = 2 \times (1 - 0.9772) \\ &\approx 2 \times 0.0228 = 0.0456. \end{aligned}$$



## Populações normais – testes de médias e variâncias - $X \sim N(\mu, \sigma^2)$

- Testes de médias **com variância,  $\sigma^2$ , conhecida**

Estatística de teste  $\bar{X} \sim N(\mu_0, \sigma^2 / n)$  ou  $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

Aplicar a regra intuitiva (no quadro 8.3 do livro apresentam-se todos os casos possíveis).

**Exemplo** (inspirado no exemplo 8.8 do livro) –  $X$  quantidade de azeite numa garrafa (dl)

$X \sim N(\mu, \sigma^2)$ ,  $\sigma^2 = 0.5$ . Quer testar-se  $H_0 : \mu = 10$  contra  $H_1 : \mu \neq 10$  com base numa amostra casual de dimensão  $n = 20$ . Suponha que se observou  $\bar{x} = 9.65$  e  $s'^2 = 0.64$ .

- a) Determine para  $\alpha = 0.05$  a região crítica a adotar e efetue o teste.
- b) Qual é o valor- $p$ ?

Solução: Formalizar o problema  $\rightarrow$  Teste bilateral para a média de uma população normal de variância conhecida (logo a informação  $s'^2 = 0.64$  não será relevante para o teste)

a) Muito semelhante ao exemplo que se acabou de ver

$$H_0 : \mu = 10 \text{ contra } H_1 : \mu \neq 10$$

Estatística de teste  $Z = \frac{\bar{x}-10}{\sigma/\sqrt{n}} \sim n(0, 1)$  já que a variância da população é conhecida

$$\alpha = 0.05 \quad \text{dado no enunciado}$$

Região crítica: Sendo  $H_1$  bilateral utiliza-se a regra empírica que consiste em considerar a região de rejeição nas 2 caudas da distribuição

$$W = \{(x_1, x_2, \dots, x_n) : z < -1.96 \vee z > 1.96\}$$

ou

$$W = \{(x_1, x_2, \dots, x_n) : \bar{x} < 9.69 \vee \bar{x} > 10.31\}$$

Utilizando  $z_{obs} = \frac{9.65-10}{\sqrt{(0.5/20)}} = -2.21$  no primeiro caso ou simplesmente  $\bar{x} = 9.65$  no segundo conclui-se que se **rejeita  $H_0$** .

$$\begin{aligned} \text{b) } p_{obs} &= P(|\bar{X} - \mu| \geq |\bar{x} - \mu| | \mu = 10) = P(|\bar{X} - 10| \geq 0.35 | \mu = 10) \\ &= 2 \times P((\bar{X} - 10) \geq 0.35 | \mu = 10) = 2 \times \left( 1 - \Phi \left( \frac{0.35}{\sqrt{\frac{0.5}{20}}} \right) \right) = 2 \times (1 - \Phi(2.21)) \\ &= 0.0272 \end{aligned}$$

Sendo o valor- $p$  pequeno (inferior à dimensão escolhida para o teste na alínea anterior) rejeita-se  $H_0$

**Alternativamente**, calcula-se  $z = \frac{9.65 - 10}{\sqrt{\frac{0.5}{20}}} = -2.21$

$$\text{e } p_{obs} = P(|Z| \geq |-2.21|) = P(|Z| \geq 2.21) = 2 \times (1 - \Phi(2.21)) = 0.0272$$

- Testes de médias com **variância desconhecida**

Raciocínio semelhante mas substituindo a estatística de teste anterior por  $T = \frac{\bar{X} - \mu_0}{S' / \sqrt{n}} \sim t(n-1)$

(Ver quadro 8.5 do livro que resume os casos possíveis)

**Exemplo 8.9** – Retome-se o exemplo anterior, supondo agora que a variância da população é desconhecida.

$$H_0 : \mu = 10 \text{ contra } H_1 : \mu \neq 10$$

Sendo a variância desconhecida, utiliza-se  $T = \frac{\bar{X} - \mu_0}{S' / \sqrt{n}} \sim t(n-1)$ , isto é,  $T = \frac{\bar{X} - 10}{S' / \sqrt{20}} \sim t(19)$

**Metodologia Neyman-Pearson:**

- Utilizando a estatística  $T$ :  $t_{\alpha/2} = 2.093$  (ver tabela da *t-student* com 19 graus de liberdade)

$$\text{Logo } W = \{(x_1, x_2, \dots, x_n) : t < -2.093 \vee t > 2.093\}$$

$$\text{Tendo-se observado } t_{obs} = \frac{9.65 - 10}{\sqrt{0.64} / \sqrt{20}} = -1.957 \text{ não se rejeita } H_0.$$

- Escrevendo a região em termos de  $\bar{x}$  e  $s'$ : Parte-se da região anterior e escrevem-se as 2 desigualdades em termos de  $\bar{x}$  e  $s'$  (menos habitual)

$$W = \left\{ (x_1, x_2, \dots, x_n): \bar{x} < 10 - 2.093 \times \frac{s'}{\sqrt{20}} \vee \bar{x} > 10 - 2.093 \times \frac{s'}{\sqrt{20}} \right\}$$

Considerando os valores observados,  $\bar{x} = 9.65$  e  $s'^2 = 0.64$ , verifica-se que a amostra não pertence à região de rejeição logo **não se rejeita**  $H_0$ .

A primeira solução é claramente mais simples do que a segunda.

**Utilizando o valor-p:** Parte-se de  $t_{obs} = \frac{9.65-10}{\sqrt{0.64}/\sqrt{20}} = -1.957$ . O “desvio” mede-se em relação ao valor 0 que corresponde à situação “ideal”, onde a média da amostra é igual à média da população.

$$p_{obs} = P(|T| \geq |-1.957|) = 2 \times P(T \geq 1.957) = 0.0652 \text{ (computador)}$$

Recorrendo às tabelas apenas se poderia concluir que o valor se situa entre 0.05 e 0.1 a não ser que se fizesse uma interpolação linear.

- **Testes sobre a variância**

Estatística de teste  $Q = \frac{(n-1)S'^2}{\sigma_0^2} \sim \chi^2(n-1)$  ou  $\frac{nS^2}{\sigma_0^2} \sim \chi^2(n-1)$

(Ver quadro 8.7 do livro)

**Exemplo 8.10** – Retome-se o exemplo anterior e suponha-se agora que se pretendia testar  $H_0 : \sigma^2 = 0.5$  contra  $H_1 : \sigma^2 > 0.5$  para uma dimensão de 2.5%. Que concluir?

**Recorrendo ao valor- $p$ :** Regra intuitiva → região de rejeição para o lado direito

Obter o valor observado:  $q_{obs} = \frac{19 \times 0.64}{0.5} = 24.23$

Calcular o valor- $p$ :  $p_{obs} = P(Q \geq 24.23) = 0.184$  (computador)  
entre 0.25 e 0.10 (tabelas)

Conclusão: Não se rejeita  $H_0$  já que o valor- $p$  é superior a 0.025.

Em rigor, o recurso ao valor- $p$  neste caso dispensaria ter fixado previamente a dimensão do teste já que se trata de um valor “grande”.

**Pela metodologia de Neyman-Pearson**, para  $\alpha = 0.025$ , chegar-se-ia a  $W = \{(x_1, x_2, \dots, x_n) : q > 32.852\}$ .

Calculava-se então  $q_{obs} = 24.23$  e como  $q_{obs}$  não pertence a  $W$ , não se rejeita  $H_0$

## Populações normais – testes à igualdade de duas populações

- 2 populações normais independentes:  $X \sim N(\mu_X, \sigma_X^2)$  e  $Y \sim N(\mu_Y, \sigma_Y^2)$ , sendo  $X$  e  $Y$  independentes.
- Não estando em causa a normalidade das populações, testar a igualdade das distribuições equivale a testar a igualdade das médias e/ou a igualdade das variâncias.
- Amostra casual com  $m$  observações da população  $X$  e outra com dimensão  $n$  da população  $Y$ . Sendo as variáveis independentes, também o serão as estatísticas construídas com base numa e noutra amostra.

### 1º problema: teste da igualdade das duas médias.

$$H_0 : \mu_X = \mu_Y \text{ ou } H_0 : \mu_X - \mu_Y = 0$$

como habitualmente, as alternativas podem ser de tipo unilateral ou bilateral.

Numa abordagem intuitiva, parece razoável basear a decisão a tomar na estatística  $\bar{X} - \bar{Y}$ .

Tendo presente os resultados obtidos no capítulo 6 sobre distribuições por amostragem para a diferença de médias  $\bar{X} - \bar{Y}$ , vão considerar-se 3 situações no que se refere às variâncias:

- $\sigma_X^2$  e  $\sigma_Y^2$  são **conhecidas**;
- $\sigma_X^2$  e  $\sigma_Y^2$  são **desconhecidas e iguais**;
- $\sigma_X^2$  e  $\sigma_Y^2$  são **desconhecidas e diferentes**;

Na situação 1 a estatística-teste vem dada por:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim N(0,1)$$

Na situação 2 a estatística-teste vem dada por:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(m-1)S_X'^2 + (n-1)S_Y'^2}{m+n-2}} \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$$

Na situação 3 a estatística-teste vem dada por:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X'^2}{m} + \frac{S_Y'^2}{n}}} \sim t(r) \text{ sendo } r \text{ a parte inteira de } r^* = \frac{\left(\frac{S_X'^2}{m} + \frac{S_Y'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_X'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_Y'^2}{n}\right)^2}$$



**Exemplos 8.11, 8.12 e 8.13** – Para confrontar dois tipos de máquinas ... As produtividades alcançadas foram as seguintes:

Máquina A: 8.0, 8.4, 8.0, 6.4, 8.6, 7.7, 7.7, 5.6, 5.6, 6.2;  $m=10$

Máquina B: 5.6, 7.4, 7.3, 6.4, 7.5, 6.1, 6.6, 6.0, 5.5, 5.5;  $n=10$

Dados úteis que se calculam a partir das amostras:  $\bar{x} = 7.22$ ,  $\bar{y} = 6.39$ ,  $s'_X{}^2 = 1.326$ ,  $s'_Y{}^2 = 0.6188$

Admitindo que as produtividades de ambas as máquinas possuem distribuição normal, o agricultor está inclinado a proceder, para  $\alpha = 0.05$ , ao teste de,

$$H_0 : \mu_A = \mu_B \text{ contra } H_1 : \mu_A > \mu_B,$$

para confirmar que a primeira máquina tem melhor rendimento.

Situação 1. Admitindo por agora que se conhecem as variâncias, seja  $\sigma_A^2 = 1.5$  e  $\sigma_B^2 = 1.0$  (tem de ser dado)

$$z_{\text{obs}} = 1.66 \text{ ou } p_{\text{obs}} \approx 0.0485. \text{ **Rejeita-se** mas, o valor-}p \text{ traduz uma certa “indecisão”}$$

Situação 2. Admitindo que as variâncias são desconhecidas mas iguais (a  $t$  tem 18 graus de liberdade)

$$t_{\text{obs}} = 1.882 \text{ ou } p_{\text{obs}} \approx 0.038$$

Situação 3. Admitindo que as variâncias são desconhecidas e diferentes

$$t_{\text{obs}} = 1.882 \text{ ou } p_{\text{obs}} \approx 0.0397 \quad \text{sendo } r^* = 15.97$$

## 2º problema: Teste para a igualdade das variâncias

$$H_0 : \sigma_X^2 = \sigma_Y^2 \text{ ou de forma equivalente } H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = 1,$$

Estatística de teste  $F = \frac{S_X'^2}{S_Y'^2} \sim F(m-1, n-1)$  quando  $\sigma_X^2 = \sigma_Y^2$

Ver quadro 8.11 livro

**Exemplo 8.14** - Retome-se ainda o exemplo 8.11, para testar,

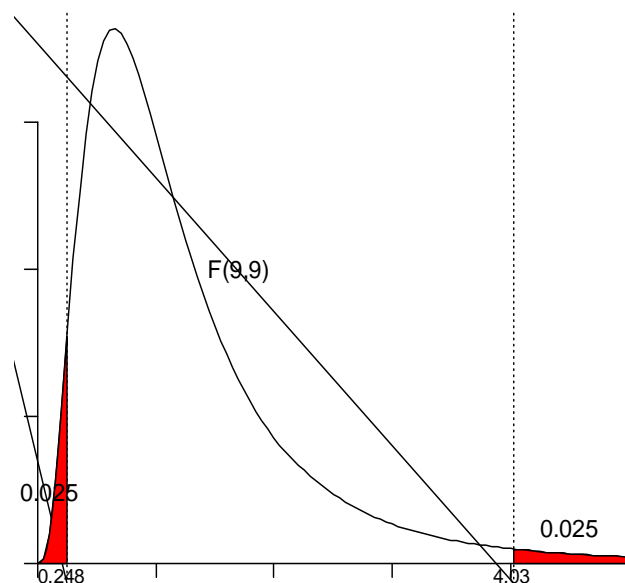
$$H_0 : \sigma_A^2 = \sigma_B^2 \text{ contra } H_0 : \sigma_A^2 \neq \sigma_B^2,$$

supondo que a dimensão do teste é  $\alpha = 5\%$ .

Para estabelecer a região de rejeição e dado que as tabelas da  $F$  são “limitadas” utiliza-se a propriedade  $F \sim F(m, n) \Rightarrow \frac{1}{F} \sim F(n, m)$ .

Neste caso o resultado pode parecer “mascarado” já que  $m = n = 9$ .

Estatística de teste  $F = \frac{S_X^2}{S_Y^2} \sim F(9,9)$



$f_2: P(F > f_2) = 0.025$   
da tabela vem  $f_2 = 4.03$

$f_1: P(F < f_1) = 0.025$   
Que pode ser obtido na máquina (computador)  
mas não na tabela.

Como  $P(F < f_1) = 0.025 \Leftrightarrow P\left(\frac{1}{F} > \frac{1}{f_1}\right) = 0.025$

e  $\frac{1}{F} \sim F(9,9)$

obtém-se  $\frac{1}{f_1} = 4.03$  e portanto  $f_1 = \frac{1}{4.03} = 0.248$

Assim  $W = \{f: f < 0.248 \vee f > 4.03\}$

Como  $f_{obs} = 2.143$ , não se rejeita  $H_0$ .

## Amostras emparelhadas

Quando se comparam as médias de duas populações é interessante considerar duas situações:

- Amostras (e populações) independentes (o que temos vindo a fazer);
- Amostras emparelhadas para procurar combater as flutuações aleatórias;

O que são amostras emparelhadas?

Em termos intuitivos, uma amostra emparelhada consiste em submeter cada elemento da amostra às duas situações em análise. Por exemplo, para comparar dois medicamentos contra a insónia, medindo o ganho em horas de sono, faz mais sentido operar com dados emparelhados, submetendo o mesmo paciente aos dois tratamentos. Se assim não for, há um vasto conjunto de características pessoais que afectam os resultados e mascaram o efeito das drogas.

Formalmente, a amostra  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ , composta por pares de observações independentes de populações normais, respetivamente,  $X \sim N(\mu_X, \sigma_X^2)$ ,  $Y \sim N(\mu_Y, \sigma_Y^2)$  é uma amostra emparelhada de populações normais.

Note-se que embora os **pares de observações sejam independentes** – cada  $(X_i, Y_i)$  é independente de  $(X_j, Y_j)$ , com  $i \neq j$  – nada se afirmou sobre a independência entre  $X$  e  $Y$  no mesmo par, **pois não há independência**.

Assim,  $Z_i = X_i - Y_i \sim N(\mu_X - \mu_Y; \sigma_{Z_i}^2)$ ,  $i = 1, 2, \dots, n$ , com

$$\sigma_{Z_i}^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}, \quad i = 1, 2, \dots, n$$

As hipóteses em teste:  $H_0: \mu_X = \mu_Y$  ou  $H_0: \mu_X - \mu_Y = 0$

\* Como a variância é desconhecida, utiliza-se

$$T = \frac{\bar{Z} - (\mu_X - \mu_Y)}{S'_z / \sqrt{n}} \sim t(n-1), \quad \text{sendo } \bar{Z} = \bar{X} - \bar{Y} \quad \text{e} \quad S'^2_z = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$$

\* O valor observado, sob  $H_0$ , da estatística-teste é dado por  $T_{\text{obs}} = \frac{\bar{z}}{s'_z / \sqrt{n}}$  e a região de rejeição é posta de acordo com  $H_1$ .

\* Nota final: O emparelhamento permite testar  $H_0$  sem necessidade de fazer qualquer suposição sobre as variâncias.

**Exemplo** – Considere-se o exemplo dos ganhos em horas de sono com os 2 medicamentos contra a insónia. Assuma-se a normalidade dos universos e suponha-se que se observou uma amostra de 5 doentes

(6.3; 4.2), (3.9; 2.9), (5.8; 5.6), (6.9; 7.2), (6.4; 5.2)

sendo o 1º elemento de cada par referente ao ganho com o medicamento A e o segundo com o medicamento B. Pretende-se testar se o medicamento A é mais eficiente, isto é

$$H_0 : \mu_A = \mu_B \text{ contra } H_1 : \mu_A > \mu_B,$$

para uma dimensão de 5%.

Solução 6.3-4.2 ... 6.4-5.2

- Constrói-se  $z_i \rightarrow 2.1 ; 1.0 ; 0.2 ; -0.3 ; 1.2$
- Calcula-se  $\bar{z} \approx 0.84$  e  $s'_z \approx 0.929$  logo  $T_{obs} \approx 2.022$

valor-p  $\approx 0.0567$

## Populações não normais – grandes amostras

Grandes amostras → Recorre-se ao teorema do limite central

### Testes sobre a média de uma população:

- utiliza-se  $Z = \frac{\bar{X} - \mu^a}{\sigma / \sqrt{n}} \sim N(0,1)$ , quando a variância é conhecida.
- utiliza-se  $Z = \frac{\bar{X} - \mu^a}{\hat{\sigma} / \sqrt{n}} \sim N(0,1)$ , quando a variância da população é desconhecida onde  $\hat{\sigma}^2$  é um estimador consistente para  $\sigma^2$ , por exemplo a variância (ou a variância corrigida) da amostra.

### Testes sobre a diferença de médias populações independentes

- utiliza-se  $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)^a}{\sqrt{\frac{\hat{\sigma}_1^2}{m} + \frac{\hat{\sigma}_2^2}{n}}} \sim N(0,1)$ , onde  $\hat{\sigma}_1^2$  e  $\hat{\sigma}_2^2$  são estimadores consistentes para  $\sigma_1^2$  e  $\sigma_2^2$ , quando estas são desconhecidas.

### Caso particular 1 – universo de Bernoulli (Quadro 8.13 do livro)

Estatística de teste  $\bar{X} \stackrel{a}{\sim} N(\theta_0, \theta_0(1-\theta_0)/n)$  ou  $Z = \frac{\bar{X} - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}} \stackrel{a}{\sim} N(0,1)$

**Exemplo 8.19** – Numa sondagem à opinião pública, em dado círculo eleitoral, foram inquiridas 1000 pessoas e houve 43% que se disseram favoráveis a determinado partido político. Será de rejeitar a hipótese deste partido ter pelo menos 50% das preferências naquele círculo?

**Formalização do teste:**  $X \sim Ber(\theta)$ , ( $X = 1$  corresponde a votar no partido)

$$H_0: \theta \geq 0.5 \quad \text{contra} \quad H_1: \theta < 0.5$$

$$\text{Estatística de teste (sob } H_0\text{): } Z = \frac{\bar{X} - 0.5}{\sqrt{\frac{0.5 \times (1-0.5)}{1000}}} = \frac{\bar{X} - 0.5}{0.0158} \sim n(0; 1)$$

$$\text{Valor observado da est.- teste: } z_{obs} = \frac{0.43 - 0.5}{0.0158} = -4.427$$

$$\text{Valor-}p: p_{obs} = P(Z < -4.427) \approx 0 \quad (\text{em rigor } 4.78 \times 10^{-6})$$

Rejeita-se claramente  $H_0$ , isto é, rejeita-se que o partido venha a ter pelo menos 50% dos votos (se nada se alterar, claro)



**Caso particular 2 – Comparação de 2 universos de Bernoulli** - Teste de  $H_0 : \theta_X = \theta_Y$  (quadro 8.14)

Estatística de teste:  $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \stackrel{a}{\sim} N(0,1)$  com  $\hat{\theta} = \frac{n_X \bar{x} + n_Y \bar{y}}{n_X + n_Y}$

**Exemplo 8.20** – Retome-se o exemplo e considere-se que, tempos depois, se efectuava uma segunda sondagem, no mesmo círculo, inquirindo 1200 pessoas. A proporção de intenções de voto passou para 45%. Em que medida os dados registados evidenciam uma progressão nas intenções do eleitorado?

**Formalização do teste:** Diferença de proporções em universos de Bernoulli – amostras independentes

Momento inicial:  $X \sim Ber(\theta_X)$   $n = 1000$ ,  $\bar{x} = 0.43$

Segundo momento:  $Y \sim Ber(\theta_Y)$   $n = 1200$ ,  $\bar{x} = 0.45$

Assume-se que as amostras são independentes.

Progressão nas intenções de voto significa  $\theta_X < \theta_Y$  logo vai-se testar

$H_0: \theta_X \geq \theta_Y$  contra  $H_1: \theta_X < \theta_Y$

Estatística de teste (sob  $H_0$ ):  $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{(\hat{\theta}(1-\hat{\theta}))\left(\frac{1}{1000} + \frac{1}{1200}\right)}} \sim n(0; 1)$

### Realização do teste:

$$\text{Valor observado da est.- teste: } \hat{\theta} = \frac{1000 \times 0.43 + 1200 \times 0.45}{1000 + 1200} = 0.44$$

$$Z_{obs} = \frac{0.43 - 0.45}{\sqrt{\left(0.44 \times 0.56 \times \left(\frac{1}{1000} + \frac{1}{1200}\right)\right)}} = -0.941$$

$$\text{Valor-}p: p_{obs} = P(Z \leq -0.941) = 1 - 0.8264 = 0.1736$$

Não se rejeita  $H_0$ , isto é, com base na informação disponível não se pode garantir que haja progressão nas intenções de voto.

A diferença pode perfeitamente dever-se ao acaso.