Capítulo 9

Teste de Hipóteses

AMG, JFO (v10 – 2018)
adaptado de: *Estatística*,
Rui Campos Guimarães,
José A. Sarsfield Cabral

Slide 9.-1

Conteúdo

Slide 9.0

Resultados de aprendizagem

- Descrever a metodologia de um teste de hipóteses
- Verificar as condições de aplicação das fórmulas dos testes de hipótese
- Realizar teste de hipóteses envolvendo:
  - valores esperados,
  - proporções binomiais,
  - variâncias,
  - razões entre variâncias,
  - diferenças de valores esperados e
  - diferenças de proporções binomiais
- Realizar testes de hipóteses unilaterais e bilaterais
- Definir as hipóteses subjacentes a um teste de hipótese, a estatística de teste e a sua distribuição, estabelecer a regra de decisão e, finalmente, calcular a estatística de teste e tomar a decisão
- Determinar o valor de prova e a potência de teste num teste de hipóteses
- Usar ICs na realização de teste de hipóteses
- Derivar e realizar testes de hipóteses “exactos”
- Conhecer o criticismo à metodologia dos testes hipóteses

Slide 9.1

9.1 Teste de hipóteses: conceito e procedimento básico

- Permite verificar se dados amostrais (ou estimativas obtidas a partir deles) são ou não compatíveis com determinadas populações (ou valores dos correspondentes parâmetros populacionais)
- Permitem assim tomar decisões usando dados reais (amostrais)
- Respostas possíveis: afirmativa ou negativa

Atenção: em ambos os casos corre-se o risco de errar

- O teste de hipóteses permite controlar esse risco

(http://digitalfirst.bfwpub.com/stats_applet/stats_applet_15_reasoning.html)

Procedimento básico

1. Definição das hipóteses
2. Identificação da estatística a utilizar no teste e caracterização da sua distribuição
3. Estabelecimento da regra de decisão, pela especificação do nível de significância do teste
4. Cálculo da estatística de teste e tomada de decisão

Slide 9.2

9.1.1 Definição das hipóteses

Estratégia básica

Demonstração da validade de uma hipótese (designada por hipótese alternativa) partindo do princípio que uma outra hipótese (designada por hipótese nula) é verdadeira, tentando mostrar a sua inverosimilhança

(procedimento semelhante ao da demonstração por redução ao absurdo)

- *Hipótese nula (H_0)*: representa o *status quo*
- *Hipótese alternativa (H_1)*: representa a hipótese em estudo

$$H_0 \text{ falsa com elevada probabilidade } \Rightarrow H_1 \text{ muito provavelmente verdadeira}$$

$$H_0 \text{ não puder ser rejeitada } \Rightarrow H_1 \text{ provavelmente falsa}$$

Slide 9.3

Formulação de hipóteses

- A *hipótese alternativa* é construída a partir das condições concretas do problema que se pretende resolver e traduz a conjectura que se pretende verificar;

  Contém sempre uma desigualdade ($<, >, \neq$) e nunca uma igualdade;
- A *hipótese nula* é a hipótese complementar de H_1 e é considerada verdadeira ao longo do teste, até que haja prova em contrário (i.e., evidência estatística) que permita rejeitá-la;

  A rejeição de H_0 leva à aceitação da hipótese alternativa;
- A hipótese nula conterá *sempre* uma igualdade ($=, \geq, \leq$);

  Na prática utiliza-se sempre $=$ na hipótese nula (mesmo quando faz sentido considerar uma situação de $\geq$ ou $\leq$) por ser a que mais se aproxima de H_1, que é complementar a H_0;
- O teste diz-se *unilateral* se a hipótese alternativa for $<$ ou $>$ e diz-se *bilateral* se a hipótese alternativa for $\neq$

Slide 9.4

Exemplo

Numa determinada empresa industrial, uma peça é fabricada automaticamente, em grandes quantidades, por duas máquinas A e B idênticas. As duas máquinas distinguem-se apenas pelo facto de a máquina A ser mais velha e, logo, mais usada.

Recolheram-se duas amostras aleatórias, uma de cada máquina, cada uma com 100 peças. Tendo-se encontrado 9 peças defeituosas na amostra proveniente da máquina A e 2 na amostra proveniente da máquina B.

Coloca-se a questão de saber se as proporções de peças defeituosas produzidas em A e B são idênticas

Resolução – Hipóteses possíveis:

$$p_A = p_B, \quad p_A \neq p_B, \quad p_A < p_B \quad \text{e} \quad p_A > p_B$$

Dado o enunciado, queremos verificar se o envelhecimento provoca um aumento na proporção de peças defeituosas produzidas, logo:

$$H_1 : p_A > p_B$$
$$H_0 : p_A = p_B$$

Slide 9.5

9.1.2 Identificação da estatística de teste e caracterização da sua distribuição

Estatística de teste (ET)

- Estatística que é utilizada para verificar a plausibilidade da hipótese nula
- ET será sempre a estatística que corresponde, na amostra, ao parâmetro que se pretende testar
- Necessário conhecer a distribuição de ET quando se admite que H_0 é verdadeira na população

Estabelecimento da distribuição da ET baseia-se:

- Combinação linear de v.a. Normais ou T.L.C.
- Amostras aleatórias (preferencialmente simples) e de grande dimensão
- Poderá ser necessário assumir a Normalidade da população

Slide 9.6

Exemplo (cont.)

A estatística que corresponde à diferença entre proporções populacionais, $p_A - p_B$, é a diferença entre proporções amostrais, $\frac{Y_A}{N_A} - \frac{Y_B}{N_B}$.

Se admitirmos que as amostras são independentes e de grande dimensão:

$$\frac{\left(\frac{Y_A}{N_A} - \frac{Y_B}{N_B}\right) - (p_A - p_B)}{\sqrt{\frac{p_A(1-p_A)}{N_A} + \frac{p_B(1-p_B)}{N_B}}} \rightsquigarrow N(0,1)$$

Como $H_0 : p_A = p_B \Leftrightarrow p_A - p_B = 0$, ou seja $p_0 = p_A = p_B$:

$$\frac{\left(\frac{Y_A}{N_A} - \frac{Y_B}{N_B}\right) - 0}{\sqrt{\frac{p_0(1-p_0)}{N_A} + \frac{p_0(1-p_0)}{N_B}}} \rightsquigarrow N(0,1)$$

com p_0 estimado por $\hat{p}_0 = \dfrac{Y_A + Y_B}{N_A + N_B}$, já que como H_0 é verdadeira Y_A e Y_B serão provenientes de distribuições Binomiais com o mesmo parâmetro p_0, logo

$$\frac{\left(\frac{Y_A}{N_A} - \frac{Y_B}{N_B}\right) - 0}{\sqrt{\hat{p}_0(1-\hat{p}_0) \cdot \left(\frac{1}{N_A} + \frac{1}{N_B}\right)}} \rightsquigarrow N(0,1)$$

Slide 9.7

9.1.3 Estabelecimento da regra de decisão, pela especificação do nível de significância do teste

- *Decisão de rejeitar ou não a hipótese nula* – fundamenta-se no valor que ET toma, no pressuposto que H_0 é verdadeira;

  Se, nestas condições, este valor for muito improvável então será porque a hipótese nula não é verdadeira.

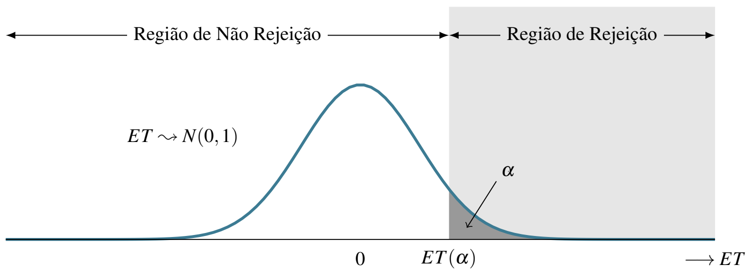
- *Como se fixa o valor a partir do qual se considera improvável que H_0 seja verdadeira* (*Regra de decisão*)?

  Define-se um *valor crítico* ($ET(\alpha)$) a partir do qual se rejeita H_0, ou seja cria-se uma região de rejeição: se o valor de ET cair nessa região H_0 é rejeitada.

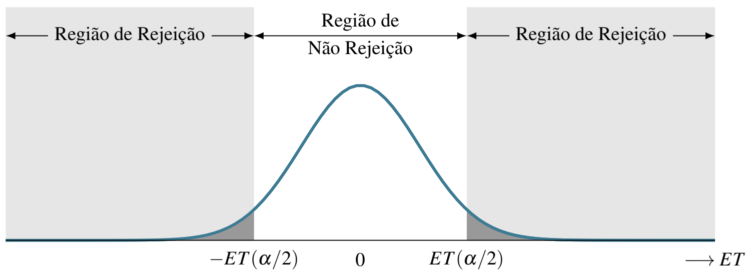
- *Nível de significância do teste* – probabilidade α de, no caso de H_0 ser verdadeira, o resultado da estatística de teste ET cair na zona de rejeição, isto é, a probabilidade de rejeitarmos H_0 quando de facto H_0 é verdadeira – *Erro do Tipo I*.

Slide 9.8

- Região de rejeição: teste unilateral à direita



- Região de rejeição: teste bilateral



Slide 9.9

9.1.4 Cálculo da estatística de teste e tomada de decisão

- Corresponde ao cálculo do valor de ET (substituição dos valores amostrais na expressão de ET) e consequente aplicação da regra de decisão (se ET pertence ou não à região de rejeição de H_0).

- Na prática, consiste em verificar se ET toma um valor igual ou mais extremo que o valor crítico ($ET(\alpha)$).

- Para um teste de hipótese ao parâmetro θ, as hipóteses típicas são:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1^a &: \theta < \theta_0 \quad \text{(teste unilateral à esquerda)} \\ H_1^b &: \theta \neq \theta_0 \quad \text{(teste bilateral)} \\ H_1^c &: \theta > \theta_0 \quad \text{(teste unilateral à direita)} \end{aligned}$$

- Rejeita-se a hipótese nula (H_0) quando:

$$\begin{aligned} H_1^a &: ET \leq ET(1-\alpha) \\ H_1^b &: |ET| \geq ET(\alpha/2) \\ H_1^c &: ET \geq ET(\alpha) \end{aligned}$$

Slide 9.10

Exemplo (cont.):

$$\alpha = 5\% \quad \Rightarrow \quad ET(\alpha) = 1.645$$

Regra de decisão:

- $ET > 1.645$ rejeitar H_0 e aceitar H_1
- $ET \leq 1.645$ não rejeitar H_0 (teste inconclusivo)

Com $N_A = N_B = 100$, $Y_A = 9$, $Y_B = 2$, vem:

$$\hat{p}_0 = \frac{Y_A + Y_B}{N_A + N_B} = \frac{11}{200} = 0.055$$

$$ET = \frac{\left(\frac{9}{100} - \frac{2}{100}\right) - 0}{\sqrt{0.055 \cdot (1 - 0.055) \cdot \left(\frac{1}{100} + \frac{1}{100}\right)}} = 2.171$$

Como o valor da estatística de teste ($ET = 2.171$) é mais extremo (superior) que o valor crítico ($ET(\alpha) = 1.645$), H_0 é rejeitada.

Slide 9.11

9.1.5 Comentários

Resultados possíveis de um teste de hipóteses:

	Realidade	
Decisão	H_0 verdadeira	H_0 falsa
H_0 rejeitada	*Decisão errada*	*Decisão correcta*
	(Erro do Tipo I)	(Teste útil)
(*probabilidade:*)	α	$1-\beta$
H_0 aceite	*Decisão correcta*	*Decisão errada*
	(Teste inconclusivo)	(Erro do Tipo II)
(*probabilidade:*)	$1-\alpha$	β

- A decisão errada de rejeitar H_0 quando H_0 é verdadeira designa-se por *Erro do Tipo I* e é por definição o nível de significância α (correspondendo a um falso positivo);
- A decisão errada de não rejeitar H_0 quando H_0 é falsa designa-se por *Erro do Tipo II beta* (correspondendo a um falso positivo);
- A decisão correcta de não rejeitar H_0 quando H_0 é verdadeira leva a um teste inconclusivo – recordar que o teste é realizado considerando H_0 verdadeira;
- Um teste de hipótese só é verdadeiramente útil quando se consegue rejeitar H_0 no caso de H_0 ser falsa, corresponde à *Potência de Teste*.

Slide 9.12

Ou seja:

- Erro to Tipo I: $\alpha = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira})$
- Teste inconclusivo: $1-\alpha = P(\text{Não rejeitar } H_0 | H_0 \text{ é verdadeira})$
- Erro to Tipo II: $\beta = P(\text{Não rejeitar } H_0 | H_0 \text{ é falsa})$
- Potência do teste: $1-\beta = P(\text{Rejeitar } H_0 | H_0 \text{ é falsa})$

Analogia com um julgamento:

- Num julgamento em tribunal, a hipótese nula H_0 é: o réu é inocente até prova em contrário;
- Exigem-se provas para se rejeitar a hipótese nula (i.e., condenar o réu);
- Se se exigirem "poucas provas", então a proporção de pessoas inocentes condenadas irá aumentar (erro do tipo I) assim como a proporção de pessoas culpadas condenadas (i.e., corretamente rejeitando a H_0);
- Por outro lado, se se exigir um "grande volume" de provas, então a percentagem de pessoas inocentes absolvidas irá aumentar (aceitar corretamente H_0) assim como a percentagem de pessoas culpadas absolvidas (erro do tipo II);
- Preferem condenar um inocente ou libertar um culpado?

Slide 9.13

Comentários

- Habitualmente fixa-se α num valor baixo, de forma a que se rejeitarmos H_0 (ou todo o processo está errado ou) então a probabilidade de cometermos um erro é baixa (i.e., a probabilidade de condenar um inocente);
- Não se fixa o erro de tipo II, β, pelo que quando não se consegue rejeitar H_0 diz-se que o teste é inconclusivo e não que se aceita H_0;
- Quando se consegue rejeitar a hipótese nula H_0 diz-se que o resultado é *estatisticamente significativo*, i.e., o resultado é improvável de ocorrer por razões puramente aleatórias;
- *Atenção*, significância estatística não é a mesma coisa que significância científica ou significância prática (voltaremos a esta questão mais à frente).

- É necessário verificar os pressupostos de aplicação de cada TH específico (exemplos de pressupostos: normalidade da população, tamanho da amostra, teorema do limite central, ...).

Slide 9.14

9.1.6 Relação entre Teste de Hipóteses e Intervalos de Confiança

- Uma hipótese nula
$$H_0: \ \theta = \theta_0$$
pode ser rejeitada a um nível de significância α se, e só se, o intervalo de confiança de θ a $(1-\alpha)\cdot 100\%$ não incluir o valor θ_0.

- O intervalo de confiança deverá ser compatível com a natureza de H_1, ou seja,
  - se o teste for bilateral o intervalo de confiança deverá ser bilateral e
  - se o teste for unilateral o intervalo de confiança deverá ser unilateral.

- Pode-se, portanto, realizar um teste de hipóteses recorrendo à construção de intervalos de confiança.

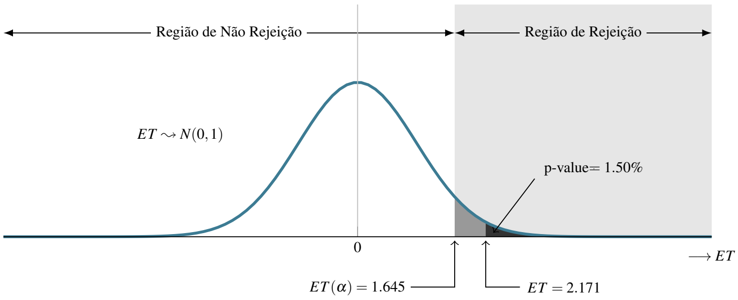
- As conclusões obtidas quando se recorre a IC para testar hipóteses são rigorosamente as mesmas se se recorresse ao procedimento "convencional".

Slide 9.15

9.2 Valor de Prova (p-value)

- É uma medida do grau com que os dados amostrais contradizem H_0;
- Corresponde à probabilidade de ET tomar um valor igual ou mais extremo que o valor observado admitindo que H_0 é verdadeira (num teste unilateral à direita vem: p-value $= P[ET \geq valor\_observado|H_0]$).

Exemplo (cont.):
Como se trata de um teste unilateral: p-value $= P(ET \geq 2.171|H_0) \approx 0.0150$



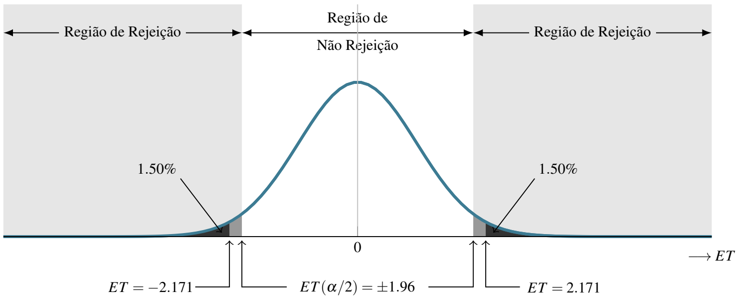
Slide 9.16

Caso tivéssemos considerado um teste bilateral, teríamos

$$H_0 : p_A = p_B \qquad H_1 : p_A \neq p_B$$

Vindo,

$$\text{p-value} = P[ET \leq -2.171|H_0 \vee ET \geq 2.171|H_0] \approx 2 \times 0.0150 = 0.03$$



Slide 9.17

9.2.1 Teste de Hipóteses – procedimento alternativo

- O valor de prova corresponde ao valor mais baixo de α que ainda permite rejeitar H_0;
- ou seja é a probabilidade, de sob H_0, se obterem dados amostrais tão ou mais extremos do que seria observado por simples acaso.

- Quando o valor de prova for muito pequeno então ou H_0 é verdadeira (e estamos perante um fenómeno raro) ou então H_0 é falsa.

- É assim possível estabelecer uma *regra de decisão alternativa*: se o valor de prova for inferior a α então rejeita-se H_0.

- Para testes bilaterais é necessário considerar a área da cauda da esquerda e a da cauda da direita (na prática, basta multiplicar a probabilidade de uma das caudas por dois).

Slide 9.18

9.2.2 Comentários

- Os dados amostrais podem contradizer H_0 em maior ou menor grau;
- O valor de prova pode ser visto como uma medida informal desse grau de contradição (ter em atenção que é uma medida não linear).

- Note-se que o nível de significância α é um valor que é fixado "arbitrariamente" por quem realiza o teste de hipótese;
- No exemplo que temos seguido H_0 é rejeitada para $\alpha = 5\%$, e se tivéssemos considerado $\alpha = 1\%$?

  Nesta situação H_0 não seria rejeitada, pois o valor de prova (1.5%) seria superior a α (1%).

- A indicação do valor de prova permite a realização do teste de hipóteses com qualquer nível de significância (α) usando o procedimento alternativo.

- Na realização de um teste de hipóteses deve-se apresentar sempre o valor de prova.

Slide 9.19

9.2.3 Criticismo

- Medidas absolutas sobre a raridade de um evento nunca são boas opções sobre a evidência a favor ou contra uma determinada hipótese.

- Os valores de prova têm sido utilizados de modo "abusivo" como forma de mostrar a validade, a importância ou a relevância prática de determinado estudo ou investigação.

- É muito difícil com um valor de prova ou um resultado de um teste de hipóteses distinguir entre relevância prática e significância estatística.

- Na realização de um teste de hipóteses é conveniente apresentar-se sempre o correspondente Intervalo de Confiança (permite ter uma melhor ideia sobre a relevância prática).

Slide 9.20

9.3 Potência de Teste

- A *potência* de um teste é a probabilidade de rejeitar correctamente H_0, i.e., quando H_0 é efectivamente falsa.

- A potência, como o próprio nome sugere, é uma coisa boa, logo pretende-se mais potência.

- Como vimos atrás, um erro do tipo II (uma coisa má, como o próprio nome sugere) é não rejeitar H_0 quando H_0 é falsa;

  A probabilidade de se cometer um erro do tipo II designa-se por β.

- Assim, *potência* $= 1 - \beta$

- Para se avaliar um erro do tipo II é forçoso transformar, com base em informação adicional, a hipótese alternativa H_1 numa igualdade;

  Para um teste unilateral à direita ao parâmetro θ, H_1 viria

  $$H_1 : \theta = \theta_0$$

Slide 9.21

Exemplo (cont.)

Exemplo (cont.)

- No exemplo que estamos a tratar temos:

  $$H_1 : p_A > p_B$$

  a região de rejeição é definida por $ET > 1.645$ e a de não rejeição por $ET < 1.645$.

- Admitindo que H_0 é falsa e, em particular, é falsa porque $p_A = p_B + \sigma$, vem

  $$H_1 \;:\; p_A - p_B = \sigma$$

  Nota: σ representa a diferença entre p_A e p_B (ou seja, é um valor numérico); neste caso considerou-se a diferença como sendo igual ao desvio padrão (σ) da variável "diferença entre proporções amostrais" (facilita os cálculos).

- Denotando-se por X a v.a. "diferença entre proporções amostrais", quando H_0 é verdadeira significa que $\mu_X = 0$ e $\sigma_X = \sigma$, vindo a distribuição de ET,

  $$ET_{H_0} = \frac{X - \mu_X}{\sigma_X} = \frac{X - 0}{\sigma} \rightsquigarrow N(0,1)$$

Slide 9.22

Exemplo (cont.)

- Quando H_1 é verdadeira significa que $\mu_X = \sigma$ e $\sigma_X = \sigma$, vindo a distribuição de ET

  $$ET_{H_1} = \frac{X - \mu_X}{\sigma_X} = ET_{H_0} + \frac{\sigma}{\sigma} = ET_{H_0} + 1 \rightsquigarrow N(1,1)$$

  Nota: optou-se por representar ET_{H_1} em função de ET_{H_0}, para facilitar a comparação entre as distribuições da ET nas duas situações;

- O valor crítico X_c que corresponde a $ET_{H_0}(5\%)$:

  $$ET(5\%) = 1.645 = Z = \frac{X - \mu_X|H_0}{\sigma_X} = \frac{X_c - 0}{\sigma} \quad \longrightarrow \quad X_c = 1.645 \cdot \sigma$$

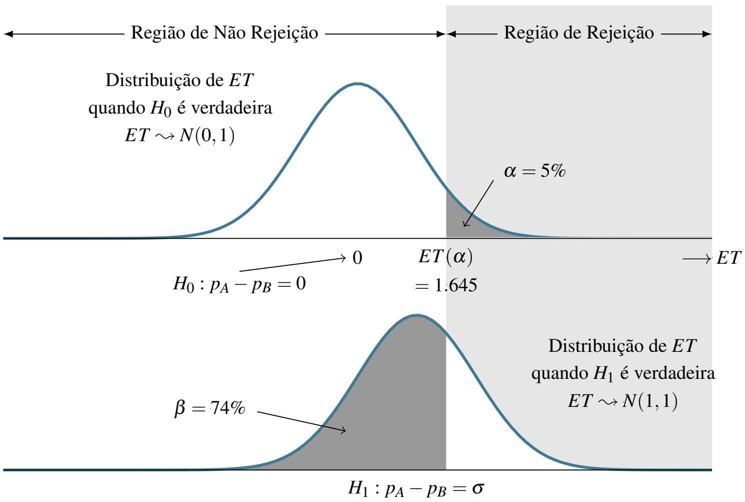
- Nestas condições, a potência de teste, $1-\beta$, vem:

  $$1 - \beta = P\left[Z = \frac{X - \mu_X|H_1}{\sigma_X} > \frac{X_c - \sigma}{\sigma}\right] = P\left[Z > \frac{1.645 \cdot \sigma - \sigma}{\sigma}\right] = 0.26$$

- A probabilidade de se incorrer num erro do tipo II, β, corresponde à probabilidade de não haver rejeição de H_0, isto é: $\beta = 1 - 0.26 = 0.74$.

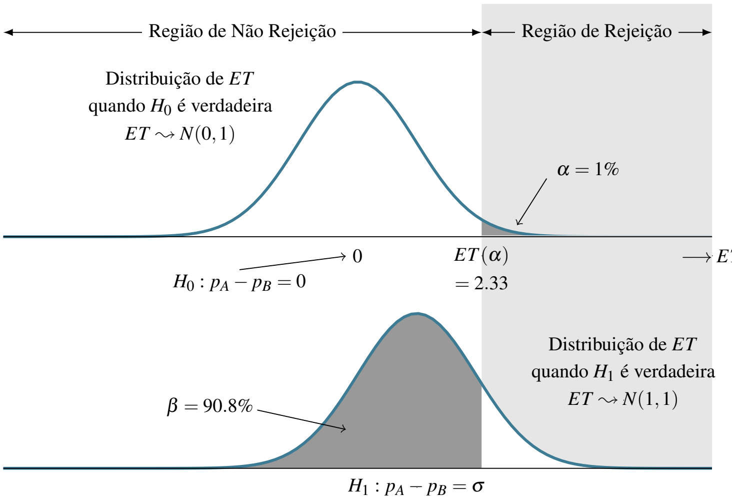
Slide 9.23

- A figura seguinte ilustra o cálculo da potência para o exemplo em estudo,



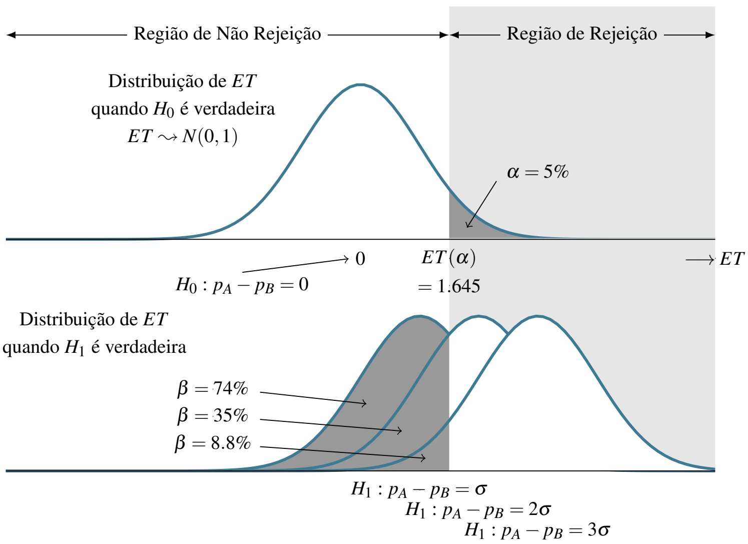
Slide 9.24

- Redução do nível de significância de $\alpha = 5\%$ para $\alpha = 1\%$



Slide 9.25

- A potência de teste, $1-\beta$, depende não só de α mas também do *afastamento* entre a hipótese alternativa H_1 e a hipótese nula H_0

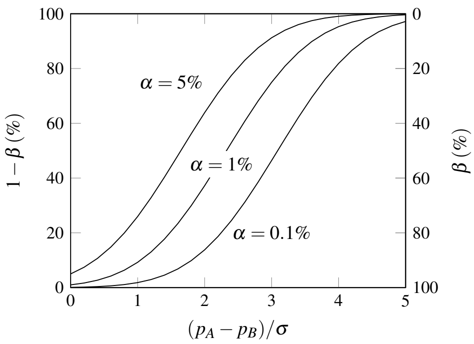


Slide 9.26

Curva da Potência de Teste

- A curva da potência de teste pode ser obtida fazendo variar H_1 de forma sistemática e mantendo α constante

- Curva da potência de teste, em função de $(p_A - p_B)/\sigma$, do exemplo que tem vindo a ser usado, para vários valores de α:



Slide 9.27

- Seria desejável que α e β fossem ambos pequenos, mas quando α diminui β aumenta, diminuindo consequentemente a potência de teste (ver figura anterior);

- A única possibilidade de diminuir simultaneamente α e β é aumentando a dimensão da amostra, este efeito é consequência da diminuição da variância da distribuição de ET;

- *Atenção*: amostras exageradamente grandes podem, no limite, conduzir à rejeição, por excesso de potência, de H_0 aproximadamente verdadeiras (ver exemplo seguinte);

- Notar que a potência aumenta quer com o aumento de α quer com o aumento do tamanho da amostra.

Slide 9.28

Exemplo (cont.)
Admita-se que a cada uma das máquinas A e B anteriormente referidas se adaptou um dispositivo capaz de detectar e separar peças defeituosas. Ao fim de uma longa série de 100 000 unidades produzidas em cada uma das máquinas, verificou-se que se produziram 8 000 defeituosas na máquina A e 7 800 na máquina B. Com base nestes números pretende-se efectuar o seguinte teste ao nível de significância de 5%:

$$H_0 : p_A = p_B$$
$$H_1 : p_A > p_B$$

- A ET é dada por $ET = \dfrac{\left(\frac{Y_A}{N_A} - \frac{Y_B}{N_B}\right) - 0}{\sqrt{\hat{p}_0(1-\hat{p}_0)\cdot\left(\frac{1}{N_A}+\frac{1}{N_B}\right)}}$
- H_0 será rejeitada se $ET > ET(\alpha = 5\%) = 1.645$
- Com $\hat{p}_0 = \dfrac{Y_A+Y_B}{N_A+N_B} = \dfrac{8000+7800}{100000+100000} = 0.079$
- Resulta que $ET = \dfrac{0.080-0.078}{\sqrt{0.079\cdot(1-0.078)\cdot\left(\frac{2}{100000}\right)}} = 1.658$
- Como $ET > ET(5\%)$, H_0 é rejeitada apesar de a diferença entre as proporções de defeituosas ser de apenas 0.2% ...

Slide 9.29

9.3.1 Exemplo – Parking Tickets

- As a part of her evaluation of municipal employees, the city manager audits the parking tickets issued by city parking officers to determine the number of tickets that were contested by the car owner and found to be improperly issued. In past years, the number of improperly issued tickets per officer had a normal distribution with mean $\mu = 380$ and $\sigma = 35.2$.
- Because there has recently been a change in the city's parking regulations, the city manager suspects that the mean number of improperly issued tickets has increased.
- An audit of 50 randomly selected officers is conducted to test whether there has been an increase in improper tickets. Use the sample data given here and $\alpha = 1\%$ to test the research hypothesis that the mean number of improperly issued tickets is greater than 380. The audit generates the following data: $n = 50$ and $y = 390$.

Nota: exemplo retirado de "An Introduction to Statistical Methods and Data Analysis", 6ed, by R. Lyman Ott and Michael Longnecker, Brooks/Cole, Cengage Learning

Slide 9.30

Resolução:

1) Definição das hipóteses:

   Queremos verificar se o valor esperado de multas incorrectas é superior a 390, logo:

$$H_0 : \mu = 390$$
$$H_1 : \mu > 390$$

2) Identificação de ET e definição da respectiva distribuição:

   Como a amostra é de grande dimensão ($n \geq 50$), podemos aproximar σ por s e o T.L.C. garante que ET segue uma distribuição Normal,

$$ET = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{(n)}} \approx \frac{\bar{y}-\mu_0}{s/\sqrt{(n)}} \rightsquigarrow N(0,1)$$

3) Regra de decisão:

   O valor crítico que define a região de rejeição é $ET(1\%) = 2.33$, rejeitando-se H_0 se $ET > 2.33$;

4) Cálculo de ET e tomada de decisão:

$$ET = \frac{390-380}{35.2/\sqrt{(50)}} = \frac{10}{35.2/7.07} = 2.01$$

   Como ET não é superior ao valor crítico ($ET = 2.01 < ET(1\%) = 2.33$) H_0 não é rejeitada, concluindo-se ser um teste inconclusivo.

Slide 9.31

- Porque o valor de ET não ultrapassa o valor crítico podemos ser tentados a aceitar H_0 como verdadeira (i.e., concluir que $\mu = 380$);

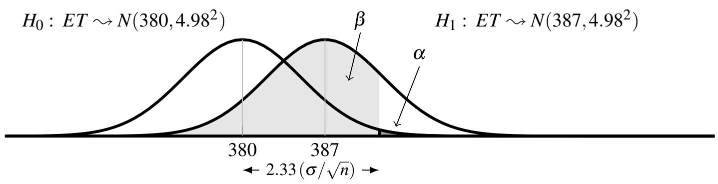
- O principal problema desta conclusão é o desconhecimento do valor do erro do tipo II (β), i.e., a probabilidade de aceitar H_0 incorrectamente;

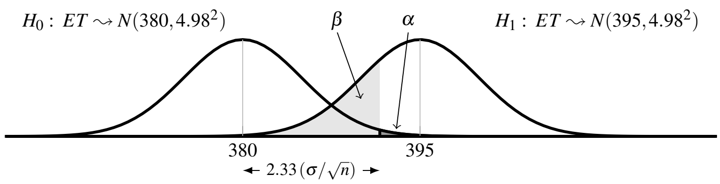
- Devemos sim dizer que se trata de um teste inconclusivo, ou seja que não temos suficiente evidência estatística para rejeitar H_0;

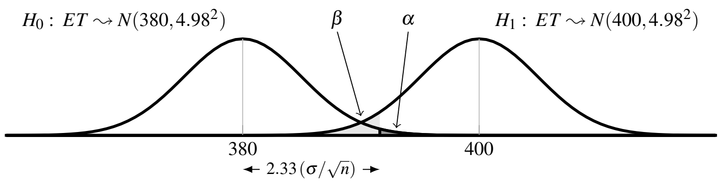
- A conclusão de aceitação ou não de H_0 deverá ser baseada no valor β, se este valor for baixo para alternativas razoáveis de valores de μ então H_0 poderá ser aceite;

- De seguida apresentam-se 3 gráficos que ilustram o cálculo de β para 3 valores alternativos de μ (387, 395 e 400), tendo-se optado por desenhar as distribuições originais e não as distribuições padronizadas.

Slide 9.32





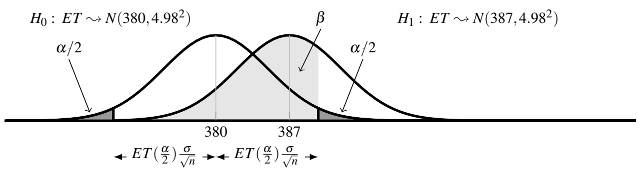


Slide 9.33

- Expressão geral para o cálculo do erro do tipo II (β) em função de μ na hipótese alternativa (μ_1) para testes unilaterais ao valor esperado ($\beta(\mu_1)$):

$$\beta(\mu_1) = P\left[ET < ET(\alpha) - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}\right]$$

- Para testes bilaterais é necessário considerar o valor crítico nas duas caudas. Exemplo para $H_1 : \mu = 387$ e $\alpha = 5\%$,



- Uma aproximação passa por considerar apenas o teste unilateral para $\alpha/2$ (no gráfico acima consiste em ignorar o valor crítico da cauda esquerda da distribuição de H_1),

$$\beta(\mu_1) \approx P\left[ET < ET(\alpha/2) - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}\right]$$

- A potência de um teste unilateral é sempre inferior à potência do teste bilateral associado.

Slide 9.34

9.3.2 Comentários

Comentários

- Apresentou-se o método genérico para calcular o erro do tipo II (β), e consequentemente para a potência de teste.

- Este método genérico foi ilustrado com exemplos envolvendo o teste ao valor esperado, para situações em que ET siga uma distribuição Normal, e o teste à diferença de proporções amostrais, a partir dos quais se pode adaptar as fórmulas apresentadas para testes semelhantes.

- De entre os testes apresentados na secção seguinte, há alguns para os quais o cálculo de β (ou da potência de teste) não é imediato:
  - teste à variância e teste à razão de variâncias,
  - teste ao valor esperado quando ET segue uma distrib. t de Student,

  nestes casos a expressão de cálculo de β não é trivial obrigando a cálculos iterativos e/ou recurso a técnicas de simulação.

- Uma das principais aplicações da potência de teste consiste na: determinação do tamanho da amostra a recolher de forma a que se possa detectar diferenças iguais ou superiores a um determinado valor com uma determinada potência de teste (pergunta habitual em exames ...).

Slide 9.35

9.4 Testes mais comuns

Quadro Resumo

Dispersão (Variância)	Uma amostra	População normal Amostra de qualquer dimensão	Teste do χ^2
	Duas amostras independentes	Populações normais Amostras de quaisquer dimensões	Teste F
Localização (Valor Esperado)	Uma amostra	População qualquer Amostra de grande dimensão	Teste Z
		População normal Amostra de pequena dimensão	Teste t
	Duas amostras independentes	Populações quaisquer Amostras de grandes dimensões	Teste Z
		Populações normais Amostras de pequenas dimensões	Teste t
	Duas amostras emparelhadas	Populações quaisquer Amostras de grandes dimensões	Teste Z
		Populações normais Amostras de pequenas dimensões	Teste t
Localização (Proporção Binomial)	Uma amostra	População dicotómica Amostra de grande dimensão	Teste Z
	Duas amostras independentes	Populações dicotómicas Amostras de grandes dimensões	Teste Z

Slide 9.36

9.4.1 Teste à variância de uma população normal (Teste do χ^2)

Um teste à variância pode incluir as seguintes hipóteses:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2,\ \sigma^2 > \sigma_0^2,\ \sigma^2 < \sigma_0^2$$

A estatística de teste é:

$$ET = (N-1)\frac{S^2}{\sigma_0^2}$$

onde N é a dimensão da amostra aleatória com variância amostral S^2.

Quando H_0 é verdadeira (e a população é Normal) a ET segue:

$$ET \rightsquigarrow \chi^2_{N-1}$$

(*Condições*: poderá ser necessário admitir a normalidade da população)

Slide 9.37

9.4.2 Teste à razão de variâncias entre duas populações normais (Teste Z)

- Um teste à razão entre as variâncias de duas populações normais pode incluir as seguintes hipóteses:

$$H_0 : \frac{\sigma_A^2}{\sigma_B^2} = r_0$$
$$H_1 : \frac{\sigma_A^2}{\sigma_B^2} \neq r_0,\ \frac{\sigma_A^2}{\sigma_B^2} < r_0,\ \frac{\sigma_A^2}{\sigma_B^2} > r_0$$

A estatística de teste é:

$$ET = r_0 \cdot \frac{S_A^2}{S_B^2}$$

- No caso particular de $r_0 = 1$, temos:

$$H_0 : \sigma_A^2 = \sigma_B^2$$
$$H_1 : \sigma_A^2 \neq \sigma_B^2,\ \sigma_A^2 < \sigma_B^2,\ \sigma_A^2 > \sigma_B^2$$

e

$$ET = \frac{S_A^2}{S_B^2}$$

- Quando H_0 é verdadeira (e as populações são Normais) a ET segue:

$$ET \rightsquigarrow F_{N_A-1,N_B-1}$$

(*Condições:* poderá ser necessário admitir a normalidade das populações)

Slide 9.38

9.4.3 Teste ao valor esperado de uma população – amostra de grande dimensão, população qualquer (Teste Z)

Um teste ao valor esperado de uma população quando a amostra é de grande dimensão pode incluir as seguintes hipóteses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0,\ \mu > \mu_0,\ \mu < \mu_0$$

A estatística de teste é:

$$ET = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{N}}}$$

onde N é a dimensão da amostra aleatória com média amostral $\bar{X}^2$ variância amostral S^2.

Quando H_0 é verdadeira (e a amostra de grande dimensão) a ET segue:

$$ET \rightsquigarrow N(0,1)$$

(*Condições*: necessário verificar se a amostra é de grande dimensão)

Slide 9.39

9.4.4 Teste ao valor esperado de uma população – amostra de pequena dimensão, população normal (Teste t)

Quando a amostra é de pequena dimensão o teste ao valor esperado pode incluir as seguintes hipóteses (idênticas às do teste anterior):

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0,\ \mu > \mu_0,\ \mu < \mu_0 \end{aligned}$$

A estatística de teste é também idêntica à utilizada no teste anterior

$$ET = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{N}}}$$

Quando H_0 é verdadeira (e a amostra de grande dimensão) a ET segue uma distribuição t de $Student$

$$ET \rightsquigarrow t_{N-1}$$

(*Condições*: poderá ser necessário admitir a normalidade da população)

Slide 9.40

9.4.5 Teste à diferença entre valores esperados de duas populações (amostras independentes) – amostras de grandes dimensões, populações quaisquer (Teste Z)

- Se os valores esperados das populações forem μ_A e μ_B, as hipóteses do teste são:

$$\begin{aligned} H_0 &: \mu_A - \mu_B = \delta_0 \\ H_1 &: \mu_A - \mu_B \neq \delta_0,\ \mu_A - \mu_B > \delta_0,\ \mu_A - \mu_B < \delta_0 \end{aligned}$$

- Se *não se puder admitir* que as variâncias das duas populações são iguais (pode ser verificado recorrendo ao teste da razão de variâncias), a estatística de teste é,

$$ET = \frac{(\overline{X}_A - \overline{X}_B) - \delta_0}{\sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}}}$$

Quando H_0 é verdadeira (e as amostras são independentes e de grandes dimensões) a estatística de teste segue uma distribuição Normal $ET \rightsquigarrow N(0,1)$.

Condições:

necessário verificar se as amostras são de grandes dimensões e se são independentes;

caso *não se possa* admitir igualdade entre as variâncias das duas populações.

Slide 9.41

- Se *se puder admitir* que as variâncias das duas populações são iguais (pode ser verificado recorrendo ao teste da razão de variâncias), a S^2 pode ser estimado por:

$$S^2 = \frac{(N_A - 1) \cdot S_A^2 + (N_B - 1) \cdot S_B^2}{N_A + N_B - 2}$$

Nestas condições, a estatística de teste é:

$$ET = \frac{(\overline{X}_A - \overline{X}_B) - \delta_0}{S \cdot \sqrt{(1/N_A) + (1/N_B)}}$$

Quando H_0 é verdadeira (e as amostras são independentes e de grandes dimensões) a estatística de teste segue uma distribuição Normal,

$$ET \rightsquigarrow N(0,1)$$

Condições:

necessário verificar se as amostras são de grandes dimensões e se são independentes;

caso *se possa* admitir a igualdade entre as variâncias das duas populações.

Slide 9.42

9.4.6 Teste à diferença entre valores esperados de duas populações (amostras independentes) – amostras de pequenas dimensões, populações Normais (Teste t)

- As hipóteses são, neste caso, as mesmas:

$$H_0 : \mu_A - \mu_B = \delta_0$$
$$H_1 : \mu_A - \mu_B \neq \delta_0,\ \mu_A - \mu_B > \delta_0,\ \mu_A - \mu_B < \delta_0$$

- Se *não se puder admitir* que as variâncias das duas populações são iguais, i.e., são diferentes (pode ser verificado recorrendo ao teste da razão de variâncias), a estatística de teste é:

$$ET = \frac{(\overline{X}_A - \overline{X}_B) - \delta_0}{\sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}}}$$

Quando H_0 é verdadeira (e as amostras são independentes e as populações Normais) $ET \rightsquigarrow t_{GL}$, com

$$GL = \frac{\left(\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}\right)^2}{\frac{\left(\frac{S_A^2}{N_A}\right)^2}{N_A - 1} + \frac{\left(\frac{S_B^2}{N_B}\right)^2}{N_B - 1}}$$

Condições:

necessário verificar se as amostras são independentes;

poderá ser necessário admitir a normalidade das populações.

Slide 9.43

- Se *se puder admitir* que as variâncias das duas populações são iguais (pode ser verificado recorrendo ao teste da razão de variâncias), a estatística de teste será:

$$ET = \frac{(\overline{X}_A - \overline{X}_B) - \delta_0}{S \cdot \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}}$$

Quando H_0 é verdadeira (e as amostras são independentes e as populações Normais) $ET \rightsquigarrow t_{GL}$, com

$$GL = N_A + N_B - 2$$

Condições:

necessário verificar se as amostras são independentes;

poderá ser necessário admitir a normalidade das populações.

Slide 9.44

9.4.7 Teste à diferença entre valores esperados de duas populações (amostras emparelhadas) – amostras de grandes dimensões, populações quaisquer (teste Z) e amostras de pequenas dimensões e populações Normais (teste t)

- *Amostra emparelhada* – amostra bivariada constituída por pares ordenados cujos termos medem a mesma grandeza.
- Exemplo:

  Idades do marido e da mulher de seis casais seleccionados ao acaso: (31, 27), (82, 76), (21, 22), (27, 27), (53, 50), (67, 62).

  As idades dos dois elementos do casal não parecem ser independentes.

- Nestes casos interessará analisar o comportamento da amostra univariada constituída pela diferença entre pares de observações;
- Como as amostras não são independentes a distribuição de $X_A - X_B$ é difícil de caracterizar $\longrightarrow$ considerar $\Delta = X_A - X_B$.

#	Amostra emparelhada (X_A, X_B)	Diferença $\Delta = X_A - X_B$
1	(x_{A1}, x_{B1})	$\Delta_1 = x_{A1} - x_{B1}$
2	(x_{A2}, x_{B2})	$\Delta_2 = x_{A2} - x_{B2}$
...	...	...
N	(x_{AN}, x_{BN})	$\Delta_N = x_{AN} - x_{BN}$
	$\bar{x}_A, \bar{x}_B$	$\bar{\Delta} = \sum_{i=1}^{N}(\Delta_i/N)$

- Retomando o exemplo das idades dos casais, pretende-se verificar se o valor esperado das idades dos maridos (A) é, ou não, superior ao valor esperado das idades das mulheres (B).
- Sendo $\Delta = X_A - X_B$, as hipóteses do teste são as seguintes:

$$H_0 : \mu_\Delta = 0$$
$$H_1 : \mu_\Delta > 0$$

$\Delta = \{31-27, 82-76, 21-22, 47-47, 53-50, 67-62\}$

$$\overline{\Delta} = 2.83 \qquad S_\Delta^2 = 2.787^2$$

$$ET = \frac{\overline{\Delta} - \Delta_0}{\frac{S_\Delta}{\sqrt{N}}} = \frac{2.83 - 0}{\frac{2.787}{\sqrt{6}}} = 2.49 > t_{6-1}(0.05) = 2.015$$

- Conclui-se então que, em média, a diferença entre a idade dos maridos e das esposas é superior a zero e é estatisticamente significativa, com valor de prova igual 2.76%.

- Quando as amostras são emparelhadas as hipóteses são:

$$H_0 : \mu_\Delta = \Delta_0$$
$$H_1 : \mu_\Delta \neq \Delta_0,\ \mu_\Delta > \Delta_0,\ \mu_\Delta < \Delta_0$$

- A estatística de teste é dada por $\quad ET = \dfrac{\bar{\Delta} - \Delta_0}{S_\Delta / \sqrt{N}}$
- Quando H_0 é verdadeira

  Para amostras de grande dimensão, populações quaisquer (teste Z),

$$ET \rightsquigarrow N(0, 1)$$

  Para amostras de pequena dimensão, populações Normais (teste t),

$$ET \rightsquigarrow t_{N-1}$$

Nota
Há situações (por exemplo, quando se quer comparar o efeito de determinado tratamento comparando-se os resultados com e sem tratamento) em que faz sentido efectuar testes com base em diferenças relativas $\left(\frac{X_C - X_S}{X_S}\right)$ em vez de diferenças absolutas $(X_A - X_B)$.

9.4.8 Teste à proporção Binomial – amostra de grande dimensão (Teste Z)

- Numa população constituída por elementos de dois tipos o valor p, que corresponde à proporção de elementos de um dos dois tipos, designa-se por proporção binomial.
- Se uma amostra de dimensão N inclui Y elementos de um dos tipos a proporção amostral é $\hat{P} = \frac{Y}{N}$ que corresponde a um valor particular do estimador $\hat{p} = \frac{y}{N}$.
- Se a amostra é de grande dimensão: $\dfrac{Y}{N} \rightsquigarrow N\left(\mu = p, \sigma^2 = \dfrac{p(1-p)}{N}\right)$.
- As hipóteses a considerar num teste relativo a proporção binomial são:

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0,\ p > p_0,\ p < p_0$$

- A estatística de teste é: $\qquad ET = \dfrac{\frac{Y}{N} - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}$
- Quando H_0 é verdadeira $ET \rightsquigarrow N(0,1)$

Nota: repare-se que, no denominador, figura o desvio padrão da proporção binomial, quando H_0 é verdadeira (ou seja $p = p_0$).

9.4.9 Teste à diferença entre duas proporções Binomiais – amostras de grandes dimensões (Teste Z)

- As hipóteses a considerar num teste relativo a diferenças entre duas proporções binomiais são:

$$H_0 : p_A - p_B = p_0$$
$$H_1 : p_A - p_B \neq p_0,\ p_A - p_B > p_0,\ p_A - p_B < p_0$$

A estatística de teste é: $$ET = \frac{\left(\frac{Y_A}{N_A} - \frac{Y_B}{N_B}\right) - p_0}{\sqrt{\frac{Y_A \cdot (N_A - Y_A)}{N_A^3} + \frac{Y_B \cdot (N_B - Y_B)}{N_B^3}}}$$

Quando H_0 é verdadeira $ET \rightsquigarrow N(0,1)$.

- No caso particular de $p_0 = 0$ (i.e., $H_0 : p_A = p_B$), e estatística de teste é,

$$ET = \frac{\left(\frac{Y_A}{N_A} - \frac{Y_B}{N_B}\right)}{\sqrt{\hat{p}_0 \cdot (1 - \hat{p}_0) \cdot \left(\frac{1}{N_A} + \frac{1}{N_B}\right)}} \qquad \left(\text{com } \hat{p}_0 = \frac{Y_A + Y_B}{N_A + N_B}\right)$$

Quando H_0 é verdadeira $ET \rightsquigarrow N(0,1)$.

Slide 9.49

9.5 Testes "Exactos"

O que fazer quando não se verificam os pressupostos de um teste?

- Os testes apresentados na secção anterior, denominados testes mais comuns, baseiam-se na aproximação da distribuição de ET à distribuição Normal, sendo por isso também conhecidos como "testes aproximados";
- A grande vantagem dos "testes aproximados" é a sua simplicidade e consequentes baixos requisitos a nível de cálculo;
- A grande desvantagem é a necessidade de verificação de pressupostos e condições de aproximação;
- O que fazer quando estes não se verificarem (amostras de pequenas dimensões e/ou então populações não Normais)?

⇒ Recorrer a outro tipo de testes que também seguem a metodologia dos testes de hipóteses (Definição das hipóteses; Distribuição de ET; Regra de decisão; Tomada de decisão), tais como:

  - Testes "Exactos";
  - Testes Não-paramétricos;
  - IC por Bootstrapping e Testes de Permutação.

  Os testes "exactos" serão abordados de seguida, enquanto as restantes opções não fazem parte do programa de MEst.

Slide 9.50

Testes "Exactos"

- Seguem a mesma metodologia dos testes "aproximados";
- Utiliza-se a distribuição "exacta" de ET (cálculos podem requerer a utilização de computadores e/ou calculadoras);
- Iremos estudar apenas testes "exactos" triviais: envolvendo a proporção Binomial (p) e o parâmetro λ de distribuições de Poisson;
- Testes envolvendo parâmetros de populações Hipergeométricas seguem uma lógica semelhante;
- Teste de Hipóteses "exactos" que comparem parâmetros de duas populações não são triviais já que se desconhece a distribuição de ET.

- *Atenção* – as designações "aproximado" e "exacto" não têm necessariamente uma conotação negativa e positiva, devendo-se sim a:
  - "Teste Aproximado" – quando é obtido a partir da aproximação da distribuição amostral à distribuição Normal;
  - "Teste Exacto" – quando é obtido directamente a partir da distribuição amostral;
  - → "Testes Exactos" dão origem a menores níveis de significância como forma a garantir os níveis de significância nominais, na prática significa inferência mais conservadora.

Slide 9.51

9.5.1 Exemplo – Proporção de componentes defeituosos

- Num processo que, no passado, era adoptado no fabrico de determinado componente de uma máquina, a proporção de componentes defeituosos era de 5%. Um novo processo – mais barato e mais rápido que o anterior – acabou de ser desenvolvido e admite-se que a proporção de componentes defeituosos possa aumentar. Nos primeiros 14 componentes fabricados, 2 revelaram-se defeituosos.
- Que conclusões pode retirar destes resultados, ao nível de significância de 10%?

Resolução:

- Pode haver a tentação de considerar um teste à diferença de proporções antes (p_A) e depois (p_D) da alteração do processo; tal não é correcto uma vez que o valor de p_A é conhecido ($p_A = 0.5\%$);
- Trata-se assim de um teste à proporção amostral, onde:

  Y Número de componentes defeituosos numa amostra de dimensão 14 retirada depois da alteração do processo;

  p_D Proporção de componentes defeituosos depois da alteração do processo.

Slide 9.52

- Definição das hipóteses,

  $H_0 : p_D = p_A \qquad H_1 : p_D > p_A$

- Admitindo independência entre os 14 elementos da amostra $Y \rightsquigarrow B(N = 14, p_D)$;
- Dada a dimensão da amostra ($N = 14$) a aproximação da Binomial pela Normal não é válida, pelo que se irá realizar um teste "exacto" usando a própria distribuição Binomial.

- Estatística de teste: $ET = Y$

  Dado que se considera H_0 verdadeira temos $ET \rightsquigarrow B(14, 0.05)$

- Regra de decisão: $\alpha = 5\%$, H_0 será rejeitada se:
  - o valor de prova for inferior a $\alpha = 5\%$ (mais fácil);
  - o valor de ET for inferior ao valor crítico $ET(5\%) = 3$.

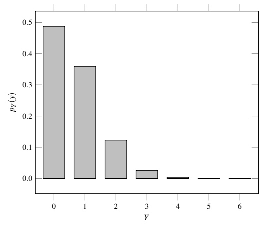
Slide 9.53

- Cálculo do valor crítico:

  $p_Y(0) = 0.4877,\ p_Y(1) = 0,3593,\ p_Y(2) = 0.1229,\ p_Y(3) = 0.02588,\ \ldots$
- Procura-se o maior valor de y que deixa à direita pelo menos $\alpha = 5\%$, logo

  $p(ET \geq 2) = 1 - 0.4877 - 0.3593 = 0.1583$

  $p(ET \geq 3) = 0.1583 - 0.1229 = 0.0301 \quad \Rightarrow \quad 2 < ET(5\%) < 3$



- Logo como $ET = 2 < ET(5\%) \Rightarrow H_0$ não é rejeitada (teste inconclusivo).

- Alternativamente e de forma mais fácil,

  p-value $= p(ET \geq 2) = 0.1583 > \alpha = 5\%$, logo H_0 não é rejeitada.

Slide 9.54

9.5.2 Exemplo – Computer chip manufacturer

- A computer chip manufacturer claims that no more than 2 percent of the chips it sends out are defective. An electronics company, impressed with this claim, has purchased a large quantity of such chips. To determine if the manufacturer's claim can be taken literally, the company has decided to test a sample of 300 chips.
- If 10 of these 300 chips are found to be defective, should the manufacturer's claim be rejected?

(Problema retirado do livro "Introduction to Probability and Statistics for Engineers and Scientists", 4ed, de Sheldon Ross)

Esboço de resolução:

- Definição das hipóteses
  $H_0 : p = 0.02 \qquad H_1 : p > 0.02$

Slide 9.55

- Estatística de teste:

  Teste "exacto":
  $ET = Y \rightsquigarrow B(300, 0.02)$ ou
  Aproximação à Normal:
  $ET = Y/N \rightsquigarrow N(Np = 6, Np(1-p) = 5.88)$.

- Regra de decisão: $\alpha = 5\%$.

- Cálculo do valor de prova do teste "exacto":
  p-value $= p(ET \geq 10) = 1 - \sum_{i=0}^{9} C_i^{300} 0.02^i (1-0.03)^{300-i} = 8.18\%$.

- Cálculo do valor de prova do teste "aproximado":
  p-value $= p\left(\frac{Y-6}{\sqrt{5.88}} \geq \frac{9.5-6}{\sqrt{5.88}}\right) \approx p(Z \geq 1.443) = 7.45\%$.

- Como p-value $> \alpha = 5\%$ então H_0 não é rejeitada (teste inconclusivo).

Slide 9.56

9.5.3 Exemplo – Management claim

- Management's claim that the mean number of defective computer chips produced daily is not greater than 25 is in dispute.
- Test this hypothesis, at the 5 percent level of significance, if a sample of 5 days revealed 28, 34, 32, 38, and 22 defective chips.

(Problema retirado do livro "Introduction to Probability and Statistics for Engineers and Scientists", 4ed, de Sheldon Ross)

Esboço de resolução:

- Because each individual computer chip has a very small chance of being defective, it is probably reasonable to suppose that the daily number of defective chips is approximately a Poisson random variable, with mean, say, λ.

Slide 9.57

- To see whether or not the manufacturer's claim is credible, we shall test the hypothesis,
  $H_0 : \lambda = 25 \qquad H_1 : \lambda > 25$

- Now, under H_0, the total number of defective chips produced over a 5-day period is Poisson distributed (since the sum of independent Poisson random variables is Poisson) with a mean no greater than 125 (5×25). Since the number of observed defective chips over a 5 day period is equal to 154 ($28+34+32+38+22$), it follows that the p-value of the data is given by,

  p-value $= p(ET \geq 154) = 1 - \sum_{i=0}^{153} \frac{e^{-\lambda} \cdot \lambda^i}{i!} = 0.66\%$

- Since p-value $< \alpha = 5\%$, the manufacturer's claim is rejected at the 5 percent (as it would be even at the 1 percent) level of significance

Slide 9.58

9.5.4 Caso de Estudo – “Can Dolphins Communicate?”

- A primeira experiência do caso de estudo “Can Dolphins Communicate?” pode ser analisada recorrendo a um Teste “Exacto” com as seguintes hipóteses:

  H_0: não há comunicação entre os golfinhos ($p = 0.5$, ou seja o Buzz responde à sorte)

  H_1: há comunicação entre os golfinhos ($p > 0.5$, ou seja o Buzz não responde à sorte)

- Já a segunda experiência do caso de estudo “Can Dolphins Communicate?” pode ser analisada recorrendo a um Teste “Exacto” ou a um Teste “Aproximado”.

Slide 9.59

9.6 Comentários

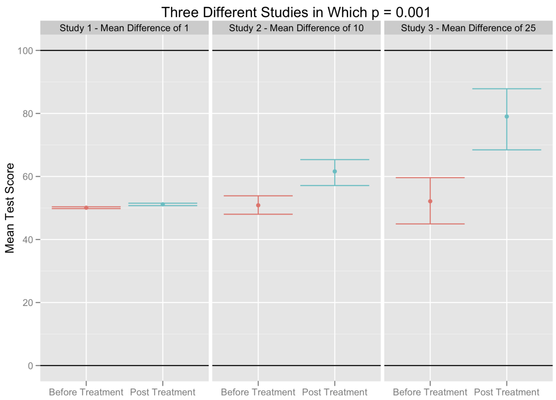
- No capítulo seguinte serão apresentadas outras técnicas de inferência (testes não-paramétricos, testes de permutação e bootstrapping) que também seguem a metodologia de teste de hipóteses aqui apresentada.

- De uma forma geral, a realização de inferência baseada na metodologia de teste de hipóteses (TH) é equivalente à realizada com base em intervalos de confiança (IC), já que a fórmula do IC resulta da inversão da expressão de ET para o teste equivalente;
- De entre os THs e ICs estudados, a única excepção onde tal equivalência não se verifica é nos THs e ICs para a proporção binomial para amostras de pequena dimensão, já que a inversão da expressão de ET não é directa;
- Uma boa aproximação para o IC para a proporção binomial para amostras de pequena dimensão quando $\alpha = 5\%$ consiste em somar 2 sucessos e 2 insucessos ao estimador de p ($\hat{p} = (Y+2)/(N+4)$) e usar a fórmula do IC apresentada atrás (dando origem ao IC de Agresti/Caffo).

Slide 9.60

9.6.1 Criticismo e Recomendações

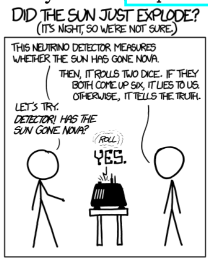
- Desde a sua introdução, nos anos vinte do século XX, a Metodologia de Teste de Hipóteses tem sido objecto de grande discussão e polémica;
- De seguida, apresentam-se as principais críticas à Metodologia de Teste de Hipóteses, assim como as principais recomendações para as minimizar.



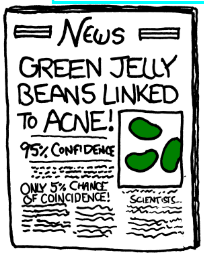
Slide 9.61

Null Hypothesis (https://xkcd.com/892/)

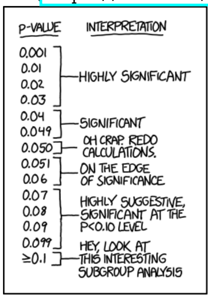
Frequentists vs. Bayesians (https://xkcd.com/1132/)



Significant (https://xkcd.com/882/)



P-Values (https://xkcd.com/1478/)



Slide 9.62

Five criticisms of significance testing – John D. Cook (link)

1. Andrew Gelman: In reality, null hypotheses are nearly always false. Is drug A identically effective as drug B? Certainly not. You know before doing an experiment that there must be some difference that would show up given enough data.
2. Jim Berger: A small p-value means the data were unlikely under the null hypothesis. Maybe the data were just as unlikely under the alternative hypothesis. Comparisons of hypotheses should be conditional on the data.
3. Stephen Ziliak and Deirdra McCloskey: Statistical significance is not the same as scientific significance. The most important question for science is the size of an effect, not whether the effect exists.
4. William Gosset: Statistical error is only one component of real error, maybe a small component. When you actually conduct multiple experiments rather than speculate about hypothetical experiments, the variability of your data goes up.
5. John Ioannidis: Small p-values do not mean small probability of being wrong. In one review, 74% of studies with p-value 0.05 were found to be wrong.

Slide 9.63

9.6.2 Comentários finais

Apesar do criticismo e limitações mencionadas, a metodologia do Teste de Hipóteses apresenta aspectos meritórios e positivos. Destacam-se as seguintes recomendações na sua utilização:

– Uso correcto da metodologia do Teste de Hipóteses;
– Apresentar sempre o valor de prova (*p-value*) de um Teste de Hipóteses;
– Apresentar também o Intervalo de Confiança associado, porque permite uma melhor avaliação da significância prática de um resultado estatisticamente significativo;
– Avaliar a potência de teste, calculando-a para diferentes valores realistas da hipótese alternativa.

Como principais alternativas à metodologia de Teste de Hipóteses, destacam-se (tópicos não abordados):

– Recurso a simulação, tirando partido do poder computacional actualmente disponível;
– Inferência bayesiana.

Slide 9.64

9.7 Exercícios

1. Um determinado processo de fabrico foi modificado por forma a melhorar a qualidade do produto final. Após alguns testes com o novo método de fabrico, o responsável pelo produto desconfia que o novo processo é significativamente mais demorado. De seguida apresentam-se os resultados de uma amostragem com os dois métodos de fabrico.

   Método antigo: $\bar{X}_A = 3.5 \quad S_A^2 = 0.961^2 \quad N_A = 20$

   Método novo: $\bar{X}_N = 4.25 \quad S_N^2 = 0.979^2 \quad N_N = 17$

   Verifique, recorrendo à técnica de intervalos de confiança, se há ou não um aumento significativo do tempo de produção de uma unidade.

   No caso de a resposta ser afirmativa estime o valor esperado e o desvio padrão desse aumento.

   Slide 9.65

   Resolução:

- Comecemos por analisar se as variâncias das duas populações podem ser consideradas iguais, sendo necessário admitir que as duas populações seguem distribuições normais,

  $H_0 : \sigma_N^2 = \sigma_A^2$

  $H_1 : \sigma_N^2 > \sigma_A^2$

  Nota: considerou-se $H_1 : \sigma_N^2 > \sigma_A^2$ por os dados amostrais assim o sugerirem ($S_N^2 > S_A^2$), uma alternativa seria considerar $H_1 : \sigma_N^2 \neq \sigma_A^2$.

  O IC vem:

  $$\left[\frac{1}{F_{16,19}(5\%)} \cdot \frac{0.979^2}{0.961^2}, +\infty\right[= [0.468, +\infty[$$

  O teste é inconclusivo (i.e., não se conseguiu rejeitar H_0), não havendo assim evidência de $\sigma_N^2 > \sigma_A^2$.

  Slide 9.66

- Analisemos agora se há ou não um aumento significativo do tempo de produção de uma unidade,

  $H_0 : \mu_N = \mu_A$

  $H_1 : \mu_N > \mu_A$

  Em face do resultado do teste anterior iremos assumir que as variâncias das duas populações são iguais, sendo a variância comum (S^2) dada por

  $$S^2 = \frac{(N_N - 1) \cdot S_N^2 + (N_A - 1) \cdot S_A^2}{N_N + N_A - 2}$$

  Nesta situação o IC para a diferença de valores esperados é dado pela expressão:

  $$(\bar{X}_N - \bar{X}_A) \pm t_{N_N+N_A-2}(5\%) \cdot S \cdot \sqrt{\frac{1}{N_N} + \frac{1}{N_A}}$$

  Substituindo,

  $$\left[(4.25 - 3.5) - t_{35}(5\%) \cdot S \cdot \sqrt{\frac{1}{17} + \frac{1}{20}}, +\infty\right[= [0.54, +\infty[$$

  Logo conseguimos rejeitar H_0, havendo assim evidência estatística de μ_N ser superior a μ_A.

  Slide 9.67

- Finalmente calculam-se as estimativas do valor esperado e do desvio padrão do aumento,

  $$\mu_{(X_N - X_A)} = E(X_N - X_A) = E(X_N) - E(X_A) \approx \bar{X}_N - \bar{X}_A = 0.75$$

  $$\sigma^2_{(X_N - X_A)} = Var(X_N - X_A) = Var(X_N) + Var(X_A) \approx S_N^2 + S_A^2 = 1.88$$

  $$\sigma_{(X_N - X_A)} \approx \sqrt{1.88} = 1.37$$

  Slide 9.68

2. Como resultado de um fim de semana prolongado de pescaria, o sr. Pescado regressou a casa com 10 cavalas e 14 sargos. O peso dos peixes pescados (em gramas) é apresentado na tabela seguinte:

Cavalas	300	285	280	355	319	275	400	310	190	320				
Sargos	370	250	471	430	250	410	300	450	260	320	210	400	425	300

   Cavalas: $\bar{X}_C = 303.4 \quad S_C^2 = 54.9^2 \quad N_C = 10$

   Sargos: $\bar{X}_S = 346.14 \quad S_S^2 = 86.2^2 \quad N_S = 14$

i) Verifique se o peso médio dos sargos é significativamente superior ao peso médio das cavalas.

ii) Se, na realidade, a diferença do peso médio dos sargos e das cavalas for de 40g, qual o nível de significância a utilizar no teste anterior por forma a que a probabilidade de cometer um erro do tipo II seja de 20%?

Slide 9.69

Resolução:

i) Comecemos por analisar se as variâncias das duas populações podem ser consideradas iguais, sendo mais uma vez necessário admitir que as populações seguem distribuições Normais,

$$H_0 : \frac{\sigma_S^2}{\sigma_C^2} = 1$$

$$H_1 : \frac{\sigma_S^2}{\sigma_C^2} > 1$$

Nota: considerou-se de novo $H_1 : \frac{\sigma_S^2}{\sigma_C^2} > 1$ por os dados amostrais assim o sugerirem ($S_S^2 > S_C^2$), uma alternativa seria considerar $H_1 : \frac{\sigma_S^2}{\sigma_C^2} \neq 1$.

Vindo,

$$ET = \frac{86.2^2}{54.9^2} = 2.46 < F_{13,9}(0.05) = 3.05$$

O teste é inconclusivo (i.e., não se conseguiu rejeitar H_0), não havendo assim evidência de $\frac{\sigma_S^2}{\sigma_C^2} > 1$.

Slide 9.70

Analisemos agora se o peso dos sargos é significativamente superior ao das cavalas,

$$H_0 : \mu_S = \mu_C$$

$$H_1 : \mu_S > \mu_C$$

Em face do resultado do teste anterior iremos assumir que as variâncias das duas populações são iguais, sendo a variância comum (S^2) dada por

$$S^2 = \frac{(N_S - 1) \cdot S_S^2 + (N_C - 1) \cdot S_C^2}{N_S + N_C - 2}$$

Nestas condições a Estatística de Teste (ET) virá,

$$ET = \frac{(\bar{X}_S - \bar{X}_C) - 0}{t_{(N_S+N_C-2)}(\alpha) \cdot \sqrt{\frac{1}{N_S} + \frac{1}{N_C}}} \rightsquigarrow t_{(N_S+N_C-2)}$$

Obtemos

$$ET = \frac{(346.14 - 303.4) - 0}{74.99 \cdot \sqrt{\frac{1}{14} + \frac{1}{10}}} = 1.38 < t_{14+10-2}(0.05) = 1.71$$

Logo não conseguimos rejeitar H_0, pelo que o teste é inconclusivo. O valor de prova é: *p-value* $= P[ET \geq 1.38|H_0] = 9.1\%$.

Slide 9.71

ii) Considerando que a diferença real do peso médio dos sargos e das cavalas é de 40g, significa que a hipótese alternativa passa a ser,

$$H_1 : \mu_S - \mu_X = \delta_0 = 40$$

A estatística de teste (ET) continua a ser,

$$ET = \frac{(\bar{X}_S - \bar{X}_C) - \delta_0}{S \cdot \sqrt{\frac{1}{N_S} + \frac{1}{N_C}}} \rightsquigarrow t_{22}$$

Nesta situação, o valor crítico (X_{VC}) não é obtido a partir de α, mas sim do valor de β,

$$t_{22}(1 - \beta = 1 - 0.2 = 0.8) = -0.858 \Rightarrow \frac{X_{VC} - 40}{74.99 \cdot \left(\frac{1}{14} + \frac{1}{10}\right)} = -0.858 \Leftrightarrow X_{VC} = 13.36$$

Finalmente,

$$P((\bar{X}_S - \bar{X}_C) > 13.36 | H_0) = P\left(t_{22}(\alpha) > \frac{13.36 - 0}{74.99 \cdot \sqrt{\frac{1}{14} + \frac{1}{10}}} = 0.43\right)$$

Vindo o nível de significância, $P(t_{22}(\alpha)) = 0.52 \Rightarrow \alpha = 33.36\%$

Notar que foi necessário "despadronizar" ET, quando H_1 é verdade, para se obter o valor crítico (X_{VC}) e, de seguida, voltar a padronizar ET, agora quando H_0 é verdade, para se obter o nível de significância (α).

Slide 9.72

3. O centro aquático de Londres, projetado pela arquiteta iraquiana Zaha Hadid, foi especialmente construído para os Jogos Olímpicos de 2012. Durante a construção do centro aquático existiriam diversas preocupações de modo a garantir níveis aceitáveis de qualidade do ar. Durante a realização dos jogos, a organização tem de manter as condições ambientais do centro bem como garantir as condições ideias da piscina olímpica para a prática das modalidades aquáticas. A Federação Internacional de Natação (FINA) adverte que a temperatura ideal para a piscina olímpica é 27^oC e que temperaturas acima dos 27ºC são particularmente prejudiciais à prática das modalidades aquáticas. A organização, que segue os critérios da FINA, estipulou que quando são registados desvios significativos, acima dos 27^oC, (a um nível de significância de 5%) o equipamento de aquecimento da piscina deve ser desligado.

Slide 9.73

A organização decidiu efetuar testes para verificar as condições da piscina, e recolheu 10 observações relativas à temperatura da piscina olímpica num determinado dia. Os dados obtidos estão apresentados no quadro seguinte. Admita que a temperatura da piscina segue uma distribuição Normal.

Observação	1	2	3	4	5	6	7	8	9	10
Temperatura (^oC)	27	30	26	25	28	28	26	25	29	29

i) Com base na amostra recolhida, diga se há necessidade de desligar o equipamento de aquecimento da piscina. Recorra a um teste de hipótese para justificar a sua resposta.

ii) Determine o valor médio da temperatura da piscina a partir do qual se deveria tomar uma decisão diferente da tomada na alínea anterior.

iii) Calcule a probabilidade de a organização conseguir detetar um desvio de temperatura da piscina de $+0.5^oC$.

Slide 9.74

4. No ano anterior, o melhor atirador da equipa dos "Andrades" conquistava pontos em cada tiro de acordo com a seguinte função de probabilidade (admita que o atirador mantém a sua performance estável ao longo do tempo):

Pontos:	2	4	6	8	10
Probabilidade:	5%	35%	35%	15%	10%

i) Num treino recente no qual efetuou 55 tiros o tal atirador obtive o fantástico resultado de 335 pontos, isto é uma média de 6.1 pontos por tiro. Com base neste resultado, teste ao nível de significância de 5% a hipótese de o atirador ter melhorado o valor esperado da pontuação por cada tiro relativamente ao que obtinha no ano anterior.

Slide 9.75

ii) Imagine que, de facto, o atirador melhorou mesmo o valor esperado da pontuação por tiro para 6.1 pontos (embora mantendo a mesma variabilidade). Nesta situação qual seria a probabilidade de, no teste anterior, não se concluir que houve uma melhoria no valor esperado da pontuação por tiro relativamente ao ano anterior.

iii) Uma das técnicas de treino utilizada pelo treinador dos "Andrades" é a seguinte: "coloca um alvo de 1cm de diâmetro a 20 metros de distância dos atiradores e cada um faz uma série de 100 tiros". No primeiro dia do recente período de treino o tal atirador acertou 35 vezes no alvo na sua série de 100 tiros. Determine, com 95% de confiança, o valor mínimo da proporção de tiros no alvo que, nessa altura, o nosso atirador obteria.

Slide 9.76