TESTE DE HIPÓTESES

- Uma população X tem uma função de distribuição F que apresenta **aspectos desconhecidos**.
- Os aspectos desconhecidos podem dizer respeito à função de distribuição (**inferência não paramétrica**) ou simplesmente a algun(s) parâmetro(s) (**inferência paramétrica**).
- A ideia associada ao teste de hipóteses é estabelecer uma **conjectura** sobre os aspectos desconhecidos da distribuição e verificar se a informação existente suporta ou não esta conjectura.
- Primeiros conceitos:
  - **Hipótese estatística:** qualquer conjectura sobre aspectos desconhecidos de F.
  - **Hipótese paramétrica:** qualquer conjectura que diga apenas respeito ao parâmetro (escalar ou vector) da distribuição. A forma funcional de F é supostamente 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 "$\mu = 1/3$" corresponde a uma hipótese paramétrica.

- **Hipótese nula e hipótese alternativa**
    - Inferência paramétrica: $X \sim f(x|\theta)$, $\theta \in \Theta$, θ desconhecido
    - Qualquer hipótese paramétrica estabelece uma partição do espaço-parâmetro Θ em Θ_0 e Θ_1.
        - Partição: $\Theta_0 \cup \Theta_1 = \Theta$, $\quad \Theta_0 \cap \Theta_1 = \varnothing$
        - $H_0 : \theta \in \Theta_0$ é a hipótese a testar e $H_1 : \theta \in \Theta_1$ é a hipótese alternativa.
    - À hipótese H_0 dá-se o nome de **hipótese nula**, designação tradicional que geralmente corresponde ao *satus quo* ou a algo que se pretende manter; a hipótese H_1 é designada por **hipótese alternativa**.

Exemplo – Pretende aferir-se se determinada moeda é equilibrada. Como se sabe, o resultado do lançamento de uma moeda pode ser modelado por uma distribuição de Bernoulli de parâmetro θ (probabilidade de um "sucesso", por exemplo, saída da «face»). A hipótese nula é então $H_0 : \theta = 1/2$ e a alternativa é $H_1 : \theta \neq 1/2$ por forma a que a união dos valores correspondentes às duas hipóteses seja o espaço-parâmetro.

Neste caso, como $\Theta = (0,1)$, pode-se, de forma mais rigorosa, redefinir H_1 para tomar em consideração este facto. No entanto, na maioria das situações tal não é necessário uma vez que a restrição está implícita.

- **Teste de hipóteses**
    - Um teste de hipóteses é uma regra que permite especificar um subconjunto do espaço-amostra, $W \subset \mathrm{X}$, tal que:
        - se $(x_1, x_2, ..., x_n) \in W$ rejeita-se H_0 (aceita-se H_1);
        - se $(x_1, x_2, ..., x_n) \notin W$ aceita-se H_0 (rejeita-se H_1).

      O conjunto W chama-se **região crítica** ou **região de rejeição**
    - O complementar do conjunto W representa-se por $\overline{W}$ e designa-se por **região de aceitação**.
- Assim um teste estatístico introduz uma partição do espaço-amostra em duas regiões, a região de rejeição e a região de aceitação,

$$W \cup \overline{W} = \mathrm{X}, \qquad W \cap \overline{W} = \varnothing.$$

  Tenha-se presente que quer W quer $\overline{W}$ são subconjuntos de $\Re^n$.
- Em muitos casos de interesse prático consegue-se evitar esta dificuldade recorrendo a uma estatística $T = T(X_1, X_2, ..., X_n)$, designada por **estatística-teste**. Trabalha-se então com o espaço-amostra relativo à estatística *T*, ou seja, com o conjunto de todos os valores particulares $t = T(x_1, x_2, ..., x_n)$.
- Nestas circunstâncias, um teste de hipóteses estabelece uma regra que permite determinar um conjunto W_T tal que: 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 adoptada se a hipótese nula for rejeitada.
- Uma estatística-teste, $T = T(X_1, X_2, ..., X_n)$.
- Uma região crítica, W_T.

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.

Teste de hipóteses – erros de 1ª e 2ª espécie

Decisão tomada	H_0 **verdadeira**	H_0 **falsa**
Rejeitar H_0	Erro de 1ª espécie	Decisão correcta
Aceitar H_0	Decisão correcta	Erro de 2ª espécie

Teste mais potente

Lema de Neyman-Pearson

- **Hipóteses Simples *vs* 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,

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

- Para o teste associado com a região crítica *W*, a probabilidade de cometer o erro de 1ª espécie,

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

  designa-se por **dimensão do teste.**

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

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

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

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

  designa-se por **potência do teste**.

- Probabilidades dos erros e das decisões correctas.

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

Decisão tomada	H_0 **verdadeira**	H_0 **falsa**
Rejeitar H_0	Erro de 1ª espécie Probabilidade $=\alpha$	Decisão correcta Probabilidade $=\beta$
Aceitar H_0	Decisão correcta Probabilidade $=1-\alpha$	Erro de 2ª espécie Probabilidade $=1-\beta$

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

a) Recolhe-se apenas uma observação, x, e que define-se. como região de rejeição $W = \{x : x > 12.5\}$. Calculem-se as probabilidades associadas com cada tipo de erro ($\alpha \approx 0.1056$, $1-\beta \approx 0.2266$).

b) Mesma pergunta mas assumindo que a amostra tem dimensão 2 e que $W = \{(x_1, x_2) : x_1 + x_2 > 25\}$ ou, de forma equivalente, $W = \{(x_1, x_2) : \bar{x} > 12.5\}$. ($\alpha \approx 0.0384$, $1-\beta \approx 0.1446$, isto é, melhoraram-se as 2 probabilidades).

- Comentário: Alterando a fronteira da região crítica, isto é o valor de *k*, obtêm-se outros valores para α e β. 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 como se pode ver na figura

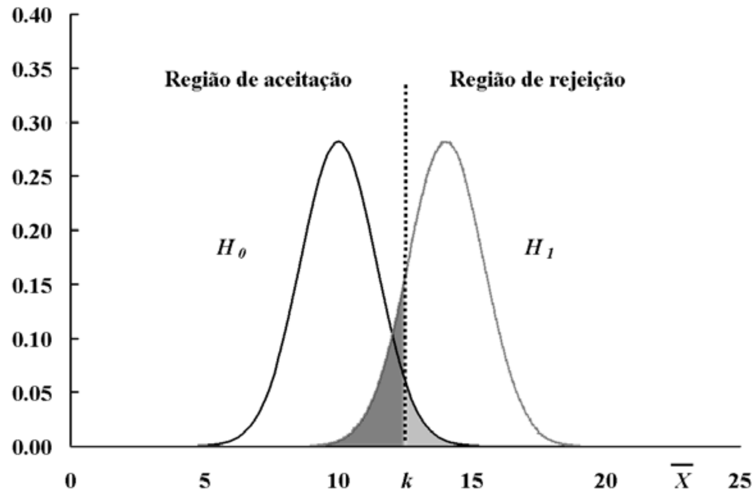

Fig. 8.3 – Probabilidade de um erro de 1ª. espécie (sombreado claro) e de um erro de 2ª. espécie (sombreado mais escuro) quando k=12.5.

- **Ideia importante**: A redução das duas probabilidades (ou de uma delas, supondo a outra fixa) só se consegue aumentando a dimensão da amostra.

- 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, α, em 0.1, 0.05 ou 0.01 ou seja apenas se rejeita H_0 se houver uma forte evidência estatística contra esta hipótese.

Definição – Teste mais potente

Fixada a probabilidade do erro de 1ª espécie (dimensão do teste), o teste mais potente é aquele em que a escolha da região crítica minimiza a probabilidade do erro de 2ª espécie. Diz-se também que esta **região crítica** é a **mais potente**.

- Como definir testes mais potentes? Recorre-se ao **lema de Neyman-Pearson**.

Teorema 8.1 (lema de Neyman-Pearson) – Suponha-se que $(X_1, X_2, \ldots, X_n)$ é uma amostra casual de população com função densidade,

$$f(x \mid \theta), \quad \theta \in \{\theta_0, \theta_1\}.$$

Seja C um número real positivo e $W \subset \Re^n$ o conjunto do espaço-amostra definido pelas condições,

$$\frac{\prod_{i=1}^{n} f(x_i \mid \theta_1)}{\prod_{i=1}^{n} f(x_i \mid \theta_0)} > C \Leftrightarrow (x_1, x_2, \ldots, x_n) \in W,$$

$$P\left[(X_1, X_2, \ldots, X_n) \in W \mid \theta = \theta_0\right] = \alpha .$$

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

***Dem.*:** Murteira (1990b).

Exemplo: Retome-se o exemplo anterior e alargue-se ligeiramente o seu enquadramento:

- Universo normal com variância conhecida σ_0^2
- $H_0 : \mu = \mu_0$ contra $H_1 : \mu = \mu_1 > \mu_0$
- amostra de dimensão n

Aplicação do Lema (as contas são apresentadas de forma detalhada no livro):

1. Construir a função de verosimilhança $\prod_{i=1}^{n} f(x_i \mid \mu) = \left(\frac{1}{\sigma_0 \sqrt{2\pi}} \right)^n \exp\left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$

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

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

3. Obter o valor fronteira utilizando a dimensão do teste

$$\alpha = P(\bar{X} > k \mid \mu = \mu_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} > \frac{k - \mu_0}{\sigma_0 / \sqrt{n}} \right) = 1 - \Phi\left(\frac{k - \mu_0}{\sigma_0 / \sqrt{n}} \right)$$

$$\text{logo } \Phi\left(\frac{k-\mu_0}{\sigma_0/\sqrt{n}}\right)=1-\alpha \quad \Rightarrow \quad k=\mu_0+z_\alpha\frac{\sigma_0}{\sqrt{n}} \text{ onde } \Phi(z_\alpha)=1-\alpha$$

- A primeira fase de aplicação do lema de Neyman-Pearson permite definir dois 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.
- Regra intuitiva: No teste de médias, variâncias ou proporções a região de rejeição está do lado da alternativa quando se utiliza a estatística "natural".
  - Exemplo: Estávamos a testar $H_0: \mu=\mu_0$ contra $H_1: \mu=\mu_1>\mu_0$
    - Estatística natural: teste sobre μ, utiliza-se $\overline{X}$
    - $\mu_1>\mu_0$ logo a região será do tipo $\bar{x}>k$
    - O valor concreto de k tem de ser determinado como anteriormente.
  - Esta regra é intuitiva (só a utilizar nas situações habituais) e só se aplica para médias variâncias e proporções e não para outros parâmetros!

Exemplo (8.4 do livro) – Seja $(0.3, 0.7, 0.9, 1.1, 1.3)$ uma amostra casual de dimensão 5 proveniente de uma população com distribuição exponencial,

$$f(x \mid \theta) = \theta \exp\{-\theta x\} \quad (x > 0, \theta > 0).$$

Pretende-se testar, $H_0 : \theta = 2$ contra $H_1 : \theta = 1$, com $\alpha = 5\%$.

Duas possibilidades à partida:

- Lema de Neyman-Pearson: $\frac{L(1)}{L(2)} > C \Leftrightarrow \sum_{i=1}^{5} x_i > k$,
- Regra intuitiva: $\mu = 1/\theta$ logo passa-se ao teste equivalente $H_0 : \mu = 1/2$ contra $H_1 : \mu = 1$

  Região de rejeição do tipo $\bar{x} > c \Leftrightarrow \sum_{i=1}^{5} x_i > 5c = k$

Em qualquer dos casos tem-se $2\theta \sum_{i=1}^{5} X_i \sim \chi^2_{(10)}$, logo

$$0.05 = P\left(\sum_{i=1}^{5} X_i > k \mid \theta = 2\right) = P\left(2\theta \sum_{i=1}^{5} X_i > 4k\right),$$

ou seja, $4k = 18.3$ e $k = 4.575$. Assim,

$$W = \left\{(x_1, x_2, ..., x_n) : \sum_{i=1}^{5} x_i > 4.575\right\}$$

Se se observar $\sum_{i=1}^{5} x_i = 4.3$, aceita-se H_0, ou seja, aceita-se $\theta = 2$.

Teste de hipóteses simples contra hipóteses compostas

- Na secção anterior estudaram-se os testes de hipóteses numa situação pouco comum, em que o espaço-parâmetro é composto por apenas dois pontos, $\Theta = \{\theta_0, \theta_1\}$.
- Para cobrir situações mais realistas definem-se os conceitos de **hipótese simples** e **hipótese composta**

  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.

Exemplos:

1. Se X é uma variável aleatória com distribuição de Poisson, $X \sim \mathrm{Po}(\lambda)$, então no teste $H_0: \lambda = 143$ contra $H_1: \lambda \neq 143$, a hipótese nula é simples e a hipótese alternativa é composta; no teste $H_0: \lambda \leq 143$ contra $H_1: \lambda > 143$, ambas as hipóteses são compostas.
2. Se X é uma variável aleatória com distribuição normal $N(\mu, \sigma^2)$, então no teste $H_0: \mu = 10,\ \sigma^2 = 4$ contra $H_1: \mu \neq 10$ ou $\sigma^2 > 4$, a hipótese nula é simples e a hipótese alternativa é composta; no teste $H_0: \mu \geq 10,\ \sigma^2 = 4$ contra $H_1: \mu < 10$ ou $\sigma^2 > 4$ ambas as hipóteses são compostas.

A primeira generalização que vai fazer-se consiste em estudar o teste de hipóteses nulas simples contra hipóteses alternativas compostas mas **unilaterais**, isto é, testes do tipo,

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

Em situações deste tipo:

- Nada se altera no que se refere ao erro de 1ª espécie (dimensão do teste) já que apenas depende de H_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 $H_1: \theta > \theta_0$ (ou contra $H_1: \theta < \theta_0$), com região crítica W, é dada por, $\beta(\theta) = P(\text{rejeitar } H_0 \mid \theta) = P\left[(X_1, X_2, \ldots, X_n) \in W \mid \theta\right]$ para $\theta \in \Theta_1$, onde $\Theta_1 = \{\theta : \theta > \theta_0\}$ (ou $\Theta_1 = \{\theta : \theta < \theta_0\}$).

- 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.
- Nalgumas situações existe um teste UMP que é obtido “estendendo” o lema de Neyman-Pearson, como vimos em exemplos. Nada obriga a que existam sempre testes UMP !!!
- **A regra intuitiva (médias, variâncias e proporções) mantém-se!**

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

- Universo normal com variância conhecida σ_0^2
- $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$
- amostra de dimensão n

- Conclui-se que a região crítica definida por $\bar{x} > k$, emanada do lema de Neyman-Pearson (ou obtida aplicando a regra intuitiva), com dimensão $\alpha = P(\bar{X} > k \mid \mu = \mu_0)$, corresponde ao teste UMP.

- A função potência vem $\beta(\mu) = P(\bar{X} > k \mid \mu) = 1 - \Phi\left(\dfrac{k - \mu}{\sigma_0 / \sqrt{n}}\right), \quad \mu > \mu_0,$

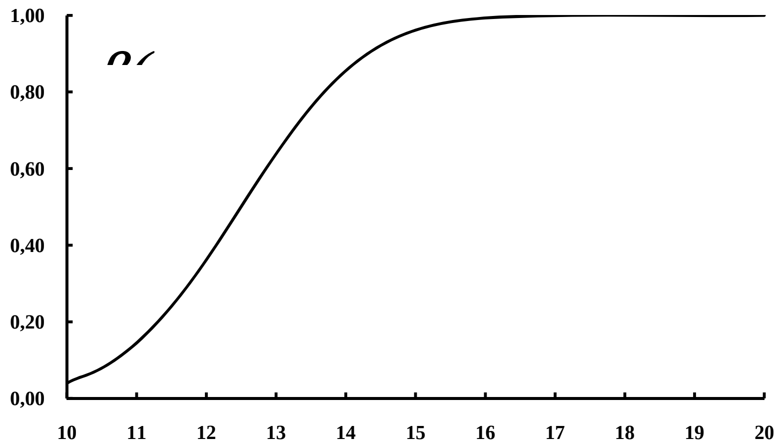

Fig. – Função potência, $H_0 : \mu = 10$ contra $H_1 : \mu > 10$. $\mu_0 = 10$, $\sigma_0 = 2$, $n = 2$, $\alpha = 0.05$ e $k = 12.5$.

O que acontece à função potência quando n **cresce, mantendo-se** α**?** Como é intuitivo a potência também cresce. No exemplo tem-se

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

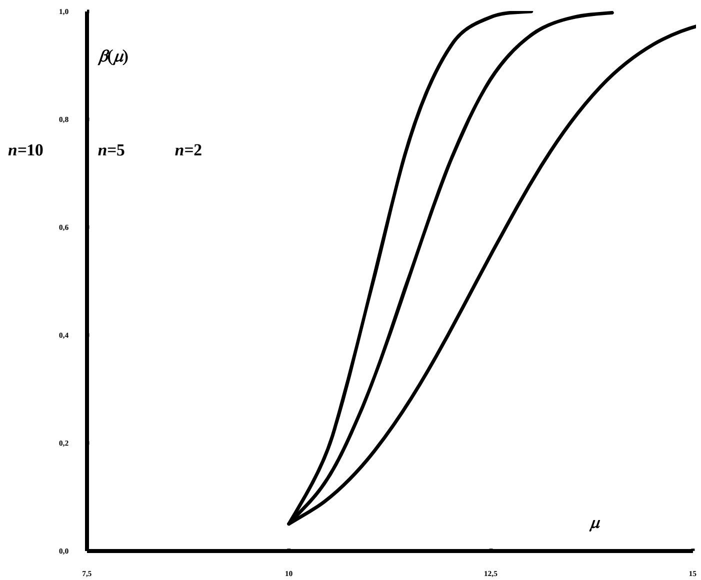

Fig. 8.7 – Função potência no teste de $H_0 : \mu = 10$ contra $H_1 : \mu > 10$

para amostras de dimensão $n = 2$, 5 e 10.

Teste de hipóteses compostas contra hipóteses compostas

- Teste $H_0:\theta\in\Theta_0$ contra $H_1:\theta\in\Theta_1$, $\Theta=\{\Theta_0,\Theta_1\}$
- Tal como no caso anterior, há situações em que existe um **teste UMP** que é obtido "estendendo a filosofia" do lema de Neyman-Pearson, como vimos em exemplos. Mas só para testes

$$H_0:\theta\geq\theta_0 \text{ contra } H_1:\theta<\theta_0 \quad \text{ou} \quad H_0:\theta\leq\theta_0 \text{ contra } H_1:\theta>\theta_0.$$

- Prob. do erro de 1ª espécie

$$\alpha(\theta)=P(\underline{X}\in W\,|\,\theta\in\Theta_0)$$

- Dimensão do teste

$$\alpha=\sup_{\theta}\{P(\underline{X}\in W\,|\,\theta\in\Theta_0)\}$$

- Prob. do erro de 2ª espécie e potência do teste, podendo-se estender a definição

$$\beta(\theta)=P(\text{rejeitar } H_0\,|\,\theta)=P[(X_1,X_2,\ldots,X_n)\in W\,|\,\theta]=\begin{cases}\alpha(\theta), & \theta\in\Theta_0\\ \beta(\theta), & \theta\in\Theta_1\end{cases}.$$

Def.: Família de distribuições com **RVM** (Razão de Verosimilhanças Monótona)

A família $\{f(x\,|\,\theta),\theta\in\Theta\}$ tem RVM na estatística T, quando para $\theta_2>\theta_1$, $L(\theta_2)/L(\theta_1)$ é função não decrescente (monótona) de $t=T(\underline{x})$.

Ex.: A família Bernoulli tem RVM na estatística $T = \sum_{i=1}^{n} X_i$. A família Normal, com variância conhecida também.

Teorema de Karlin-Rubin: Considere uma amostra casual $\underline{X} = (X_1,..., X_2)$ de dimensão n de uma população com distribuição com RVM na estatística *T*, então para ensaiar

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

os testes da forma $T(\underline{x}) > t_0$, onde t_0 é determinado em função da dimensão desejada, são UMP.

Nota: a adaptação ao ensaio $H_0 : \theta \ge \theta_0$ contra $H_1 : \theta < \theta_0$ é imediata.

Ex: $X \sim Exponencial(\theta)$

Teste de hipóteses simples contra hipóteses compostas, "bilaterais"

Testes bilaterais

$H_0 : \theta = \theta_0$ 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.

- Quando $\mu > 10$ a região crítica do teste UMP é definida pela condição $\bar{x} > k$.
- Quando $\mu < 10$ a região crítica do teste UMP é definida pela condição $\bar{x} < k'$.
- Recorre-se a uma regra intuitiva que consiste em definir uma região de rejeição nas duas abas da distribuição da estatística–teste, atribuindo igual probabilidade, seja $\alpha / 2$, a cada uma das duas sub-regiões.
- Retomando o exemplo vem, para um teste de dimensão 5% e uma amostra com 25 observações, a região de rejeição definida por, $W = \{(x_1,..., x_{25}) : \bar{x} < 9.216 \vee \bar{x} > 10.784\}$.

Ensaios de Significância

- Os ensaios de significância distinguem-se dos ensaios de hipóteses, na medida em que no seu procedimento estatístico não há qualquer intervenção de distribuições probabilísticas que não decorram da hipótese nula.
- Esta distinção é uma das raízes do confronto entre Fischer (t. significância) e Neymann-Pearson (t. hipóteses)
- Para Fisher a perspectiva evidencial é a que interessa ao cientista e não a "análise custo-benefício".
- O aspecto fundamental é encontrar uma **estatística-teste**, ou medida de afastamento, T:
    - Quando a hipótese nula é verdadeira, a sua distribuição é conhecida, ou aproximada.
    - Quanto mais afastado da hipótese nula for o valor de "T observado" $H_0 : \mu = 10$, mais se põe em dúvida H_0.
    - T observado: $t_{obs} = T(x_1, ..., x_n)$
    - $p_{obs} = P(T \geq t_{obs} \mid H_0)$ ou $p_{obs} = P(T \geq t_{obs} \mid H_0)$ ou $p_{obs} = P(|T| \geq t_{obs} \mid H_0)$ é o nível de significância ou "valor –*p*" (*p*-value).
    - Qto menor for o valor o valor-*p* menor é a evidência dos dados em suportar a hipótese nula.
- Nos ensaios de significância não há concretização de hipótese alternativa

Valor-*p*

- Num teste de hipóteses, fixada a dimensão α, 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, forma esta que permite ultrapassar esta limitação.

Definição – valor-*p*

Dada a amostra concreta $(x_1, x_2, ..., x_n)$, o valor observado para a estatística-teste escreve-se,

$$T(x_1, x_2, ..., x_n) = t_{\text{obs}}.$$

O valor-*p* ou nível de significância associado ao valor observado da estatística-teste é a probabilidade, p_{obs}, de obter este valor ou outro mais desfavorável para a hipótese nula, admitindo que esta hipótese é verdadeira.

- O valor-*p* mede a evidência que os dados fornecem a favor de H_0. Assim, quanto menor for p_{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 $p_{\text{obs}} = 0.23$ aceitaremos H_0. O problema surge com os valores "cinzentos", por exemplo se $p_{\text{obs}} = 0.068$.

- 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 (depende de H_0) do que aquele que se observou
    - Calcular a sua probabilidade
- **Exemplos:** $X \sim N(\mu;4)$**,** $n=16$ tendo-se observado $\bar{x}=11$.
    1. $H_0:\mu\geq 10$ contra $H_1:\mu<10$
    2. $H_0:\mu=10$ contra $H_1:\mu\neq 10$

Em ambos os casos a estatística de teste vai ser $\overline{X}$ que tem distribuição $\overline{X} \sim N(\mu;1/4)$**.**

Na situação 1 os casos tão ou mais desfavoráveis correspondem a observar uma media amostral inferior ou igual ao valor observado (valor mais pequenos de μ tendem a originar amostrar com médias amostrais inferiores), logo $p_{\text{obs}}=P(\overline{X}\leq 11\,|\,\mu=10)$.

Na situação 2 os casos tão ou mais desfavoráveis correspondem a observar uma media amostral mais "afastada" de 10 do que o valor observado, 11, quer seja para valores superiores quer inferiores logo $p_{\text{obs}}=P(\left|\overline{X}-10\right|\geq 1.1\,|\,\mu=10)$.

Testes para uma e duas amostras

Exemplo: ***X*** quantidade de azeite numa garrafa (dl)

$X \sim N(\mu,\sigma^2)$, $\sigma^2 = 0.5$. Quer testar-se $H_0 : \mu = 10$, com base numa amostra casual de dimensão $n = 20$. Suponha que se observou $\bar{x} = 10.5$.

a) Determine para $\alpha = 0.05$ a região crítica a adoptar.

b) Qual é o valor-*p*?

Formalizar o problema $\rightarrow$ Teste bilateral para a média de uma população normal

a) $\alpha = 0.05$ logo $z_{\alpha/2} = 1.96$ e tem-se 2 vias **alternativas** para resolver o problema:

i. Recorrendo a Z. Região de rejeição $W_Z = \left\{ z : z < -1.96 \vee z > 1.96 \right\}$ (ver quadro) e como $z_{obs} = \dfrac{10.5-10}{\sqrt{0.5/20}} \approx 3.16$ rejeita-se H_0

ii. Recorrendo a $\bar{X}$. Região de rejeição $W_{\bar{X}} = \left\{ \bar{x} : \bar{x} < 9.69 \vee \bar{x} > 10.31 \right\}$ (ver quadro) já que $10 - 1.96 \times \sqrt{0.5/20} \approx 9.69$ e $10 + 1.96 \times \sqrt{0.5/20} \approx 10.31$. Como $\bar{x} = 10.5$ rejeita-se H_0.

b) Uma vez mais têm-se 2 alternativas

i. Recorrendo a Z. Calcular $z_{obs} = 3.16$.

$$\text{valor - } p \approx 2 \times \Pr(Z > |3.16|) = 2 \times \Pr(Z > 3.16) \approx 2(1 - 0.9992) = 0.0016$$

ii. Recorrer a $\overline{X}$. $\text{valor - }p = \Pr(\overline{X} > 10.5 \mid \mu = 10) + \Pr(\overline{X} < 9.5 \mid \mu = 10)$ logo

$$\text{valor - }p = 1 - \Phi\left(\frac{10.5 - 10}{\sqrt{0.5/20}}\right) + \Phi\left(\frac{9.5 - 10}{\sqrt{0.5/20}}\right) \approx 2 - 2\,\Phi(3.16) \approx 0.0016 \text{ (tabelas)}$$

Exemplo **–** 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 maioria absoluta das preferências naquele círculo?

Exemplo – Retome-se o exemplo anterior 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 mudança de atitude nas intenções do eleitorado?

Exemplo – Para confrontar dois tipos de máquinas de ceifar - segadeiras - um trigal foi dividido em secções longitudinais e cada duas secções adjacentes tratadas por cada uma das 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;

Máquina *B*: 5.6, 7.4, 7.3, 6.4, 7.5, 6.1, 6.6, 6.0, 5.5, 5.5.

Ao agricultor que experimenta as segadeiras interessa averiguar se a produtividade média das duas máquinas se pode considerar igual ou se existe diferença significativa que o leve a preferir uma das máquinas. Admitindo que as produtividades de ambas as máquinas possuem distribuição normal, o agricultor está inclinado a proceder, para $\alpha = 0.05$, ao teste de,

$$H_0 : \mu_A = \mu_B .$$

Admita-se que se conhecem as variâncias, seja $\sigma_A^2 = 1.5$ e $\sigma_B^2 = 1.0$.

Solução:

$z_{\text{obs}} = 1.66$ logo rejeita-se H_0. Pelo valor-*p* seríamos mais ponderados já que,

$$p_{\text{obs}} = P(\bar{X}_A - \bar{X}_B \geq 7.22 - 6.39 \mid \mu_A = \mu_B) = 1 - \Phi\left(\frac{0.83}{0.5}\right) = 1 - \Phi(1.66) \approx 0.0485 .$$

que se situa muito próximo de 5%.

Se não se conhecerem as variâncias, a estatística-teste será outra ("*t*-student"). Ver distribuições por amostragem.

Testes para 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;

Amostras emparelhadas:

Em termos intuitivos, uma amostra emparelhada consiste em submeter a mesma 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.

Por exemplo, submeter o mesmo grupo de alunos a dois tipos de testes.

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

Testes

Testar a hipótese

$$H_0 : \mu_X = \mu_Y, \text{ equivalente a } H_0 : \mu_X - \mu_Y = 0$$

- Sabe-se que $Z_i = X_i - Y_i \sim N(\mu_X - \mu_Y ; \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}),\ i = 1,2,\ldots,n$
- Como a variância é desconhecida, utiliza-se $T = \dfrac{\bar{Z} - (\mu_X - \mu_Y)}{S'_z / \sqrt{n}} \sim t(n-1)$
- Sendo $\bar{Z} = \bar{X} - \bar{Y}$ e $S'^2_z = \dfrac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2$
- O valor observado, sob H_0, da estatística-teste é dado por $T_{\text{obs}} = \dfrac{\bar{Z}}{s'_z / \sqrt{n}}$ e a região de rejeição é posta de acordo com a alternativa.
- 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

$$\{(6.3;4.2),(3.9;2.9),(5.8;5.6),(6.9;7.2),(6.4;5.2)\}$$

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

- 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 ≈ 0.0567

Testes em universos Normais bidimensionais

Seja $[(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)]$ amostra casual de pop. normal bidimensional com parâmetros

$$\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho.$$

Sejam as estatísticas

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,,\ \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i\,,\ S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\,,\ S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2\,,\ R = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{nS_x S_y},$$

Fisher e Romanovsky estudaram as distribuição por amostragem:

$(\overline{X},\overline{Y})$ e (S_X, S_Y, R) são independentes, possuindo $(\overline{X},\overline{Y})$ distribuição bi-normal, com parâmetros

$$E(\overline{X}) = \mu_X\,,\ E(\overline{Y}) = \mu_Y\,,\ \mathrm{Var}(\overline{X}) = \frac{\sigma_X^2}{n},\ \mathrm{Var}(\overline{Y}) = \frac{\sigma_Y^2}{n},\ \frac{\mathrm{Cov}(\overline{X},\overline{Y})}{\sqrt{\mathrm{Var}(\overline{X})\,\mathrm{Var}(\overline{Y})}} = \rho.$$

ρ continua a ser o coeficiente de correlação entre $\overline{X}$ e $\overline{Y}$.

Sobre (S_X, S_Y, R), a segunda conclusão a que chegaram refere-se à função densidade marginal do coeficiente de correlação *R*, cuja expressão é complexa. Felizmente Fisher conseguiu encontrar uma função de *R* cuja distribuição tende para a normal com muita rapidez.

A **Transformação *Z* de Fisher**, relativamente a *R* e a ρ, respectivamente,

$$Z = \frac{1}{2}\ln\left(\frac{1+R}{1-R}\right),\ \zeta = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right),$$

Mesmo para valores moderados de *n*, pode empregar-se a distribuição assintótica

$$Z \overset{a}{\sim} N\left(\zeta, \frac{1}{n-3}\right) \Leftrightarrow \sqrt{n-3}(Z-\zeta) \overset{a}{\sim} N(0,1).$$

Quando $\rho = 0$ a distribuição (exacta) por amostragem de *R*: $\sqrt{n-2}\,\dfrac{R}{\sqrt{1-R^2}} \sim t(n-2)$.

Esta distribuição exacta tem, importantes aplicações.

Testar $H_0 : \rho = 0$ contra: a) $H_1 : \rho > 0$; ou b) $H_1 : \rho < 0$; ou $H_1 : \rho \neq 0$.

A estatística-teste adequada é:

$$T = \sqrt{n-2}\,\frac{R}{\sqrt{1-R^2}} \sim t(n-2).$$

Exemplo – Recolhida uma amostra de dimensão 20 de uma população normal bidimensional, o coeficiente de correlação amostral observado é $r = 0.22$. Testar $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$ (teste bilateral), com dimensão de 0.05, a r.c. é dada por $W_T = \{t : t < -2.086 \vee t > 2.086\}$. O valor observado

$$T_{obs} = \sqrt{20-2} \times \frac{0.22}{\sqrt{1-0.22^2}} = 0.9568,$$ não se rejeitando a ausência de correlação entre as variáveis.

Testar $H_0 : \rho = \rho_0 \neq 0$,

recorre-se às transformações Z de Fisher, $Z = \frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)$, $\zeta = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$

e utiliza-se $$T = \sqrt{n-3}(Z-\zeta) \overset{a}{\sim} N(0,1).$$

Exemplo 8.17 – Recolhida uma amostra de dimensão 200 de uma população normal bidimensional, o coeficiente de correlação amostral observado é $r = 0.22$. Para testar $H_0 : \rho = 0.15$ contra $H_1 : \rho \neq 0.15$ (teste bilateral), com dimensão de 0.05, a região de rejeição é dada por $W_T = \{t : t < -1.96 \vee t > 1.96\}$. Dado H_0, o valor observado da estatística teste é

$$T_{obs} = \sqrt{200-3} \times \frac{1}{2}\left[\ln\left(\frac{1+0.22}{1-0.22}\right) - \ln\left(\frac{1+0.15}{1-0.15}\right)\right] = 1.018$$

não se rejeitando o valor 0.15 para o coeficiente de correlação.

Finalmente, também se pode utilizar a transformação *Z* de Fisher par construir intervalos de confiança para o coeficiente de correlação.

Resumo de Testes mais habituais:

Populações normais – testes de médias e variâncias

- Testes de médias com variância conhecida $X \sim N(\mu, \sigma^2)$, σ^2 conhecido.

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

Teste de $H_0 : \mu = \mu_0$ (com σ^2 conhecido) contra H_1

H_1	**Região crítica**	**valor-*p***
$\mu > \mu_0$	$z > z_\alpha$ ou $\bar{x} > \mu_0 + z_\alpha \dfrac{\sigma}{\sqrt{n}}$	$P(Z \geq z_{\text{obs}} \mid H_0)$
$\mu < \mu_0$	$z < -z_\alpha$ ou $\bar{x} < \mu_0 - z_\alpha \dfrac{\sigma}{\sqrt{n}}$	$P(Z \leq z_{\text{obs}} \mid H_0)$
$\mu \neq \mu_0$	$\lvert z \rvert > z_{\alpha/2}$ ou $\bar{x} > \mu_0 + z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \vee \bar{x} < \mu_0 - z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$	$2P\left(Z \geq \lvert z_{\text{obs}} \rvert \mid H_0\right)$
Onde: $z_{\text{obs}} = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$; $z_\alpha : P(Z > z_\alpha) = \alpha$		

- Testes de médias com variância desconhecida $X \sim N(\mu, \sigma^2)$, σ^2 desconhecido.

Raciocínio semelhante mas substituindo a estatística de teste anterior por

$$T = \frac{\overline{X} - \mu_0}{S'/\sqrt{n}} \sim t(n-1)$$

Teste de $H_0 : \mu = \mu_0$ (com σ^2 desconhecido) contra H_1

H_1	**Região crítica**	**valor-*p***
$\mu > \mu_0$	$t > t_\alpha$ ou $\overline{x} > \mu_0 + t_\alpha \frac{s'}{\sqrt{n}}$	$P(T \geq t_{\text{obs}} \mid H_0)$
$\mu < \mu_0$	$t < -t_\alpha$ ou $\overline{x} < \mu_0 - t_\alpha \frac{s'}{\sqrt{n}}$	$P(T \leq t_{\text{obs}} \mid H_0)$
$\mu \neq \mu_0$	$\lvert t \rvert > t_{\alpha/2}$ ou $\overline{x} < \mu_0 - t_{\alpha/2} \frac{s'}{\sqrt{n}} \vee \overline{x} > \mu_0 + t_{\alpha/2} \frac{s'}{\sqrt{n}}$	$2P\left(T \geq \lvert t_{\text{obs}} \rvert \middle\vert H_0\right)$
Onde: $t_{\text{obs}} = \frac{\overline{x} - \mu_0}{s'/\sqrt{n}}$; $t_\alpha : P(T > t_\alpha) = \alpha$		

- Testes sobre a variância.

Estatística de teste $\quad Q=\dfrac{(n-1)S'^2}{\sigma_0^2}\sim\chi^2(n-1)$.

Teste de $H_0:\sigma^2=\sigma_0^2$ contra H_1

H_1	**Região crítica**	**Valor-*p***
$\sigma^2>\sigma_0^2$	$q>q_\alpha$ ou $s'^2>\dfrac{q_\alpha\,\sigma_0^2}{n-1}$	$P(Q\geq q_{\text{obs}}\mid H_0)$
$\sigma^2<\sigma_0^2$	$q<q_{1-\alpha}$ ou $s'^2<\dfrac{q_{1-\alpha}\,\sigma_0^2}{n-1}$	$P(Q\leq q_{\text{obs}}\mid H_0)$
$\sigma^2\neq\sigma_0^2$	$q<q_{1-\alpha}\vee q>q_\alpha$ ou $s'^2<\dfrac{q_{1-\alpha/2}\,\sigma_0^2}{n-1}\vee s'^2>\dfrac{q_{\alpha/2}\,\sigma_0^2}{n-1}$	2 vezes o menor dos valores acima
Onde: $q_{\text{obs}}=\dfrac{(n-1)\,s'^2}{\sigma_0^2}$; $\;q_\alpha:P(Q>q_\alpha)=\alpha$		

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 de tipo unilateral ou bilateral. Estatística $\overline{X} - \overline{Y}$.

Tendo presente os resultados obtidos sobre distribuições por amostragem para a diferença de médias $\overline{X} - \overline{Y}$, devem-se considerar três situações no que se refere às variâncias:

a) σ_X^2 e σ_Y^2 são **conhecidas**;

b) σ_X^2 e σ_Y^2 são **desconhecidas e iguais**;

c) σ_X^2 e σ_Y^2 são **desconhecidas e diferentes**.

Na situação **(a)** a estatística-teste vem dada por,

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

Teste de $H_0 : \mu_X = \mu_Y$ (com σ_X^2 e σ_Y^2 conhecidos) contra H_1

H_1	**Região crítica**	**valor-*p***
$\mu_X > \mu_Y$	$z > z_\alpha$ ou $\bar{x} - \bar{y} > z_\alpha \sqrt{\dfrac{\sigma_X^2}{m} + \dfrac{\sigma_Y^2}{n}}$	$P(Z \geq z_{\text{obs}} \mid H_0)$
$\mu_X < \mu_Y$	$z < -z_\alpha$ ou $\bar{x} - \bar{y} < -z_\alpha \sqrt{\dfrac{\sigma_X^2}{m} + \dfrac{\sigma_Y^2}{n}}$	$P(Z \leq z_{\text{obs}} \mid H_0)$
$\mu_X \neq \mu_Y$	$\lvert z \rvert > z_{\alpha/2}$ ou $\lvert \bar{x} - \bar{y} \rvert > z_{\alpha/2} \sqrt{\dfrac{\sigma_X^2}{m} + \dfrac{\sigma_Y^2}{n}}$	$2P\left(Z \geq \lvert z_{\text{obs}} \rvert \mid H_0\right)$
Onde: $z_{\text{obs}} = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_X^2}{m} + \dfrac{\sigma_Y^2}{n}}}$; $z_\alpha : P(Z > z_\alpha) = \alpha$		

Na situação **(b)** a estatística-teste vem dada por,

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

Teste de $H_0 : \mu_X = \mu_Y$ (com σ_X^2 e σ_Y^2 desconhecidos e iguais) contra H_1

H_1	**Região crítica**	**valor-*p***
$\mu_X > \mu_Y$	$t > t_\alpha$ ou $\bar{x} - \bar{y} > t_\alpha\, s^*$	$P(T \geq t_{\text{obs}} \mid H_0)$
$\mu_X < \mu_Y$	$t < -t_\alpha$ ou $\bar{x} - \bar{y} < -t_\alpha\, s^*$	$P(T \leq t_{\text{obs}} \mid H_0)$
$\mu_X \neq \mu_Y$	$\lvert t \rvert > t_{\alpha/2}$ ou $\lvert \bar{x} - \bar{y} \rvert > t_{\alpha/2}\, s^*$	$2P(T \geq \lvert t_{\text{obs}} \rvert \mid H_0)$

Onde: $t_{\text{obs}} = \dfrac{\bar{x} - \bar{y}}{s^*}; t_\alpha : P(T > t_\alpha) = \alpha$

$$s^* = \sqrt{\frac{(m-1)s_X'^2 + (n-1)s_Y'^2}{m+n-2}}\sqrt{\frac{1}{m}+\frac{1}{n}}$$

A situação **(c)** é mais complicada e utiliza-se a aproximação de Welch. A estatística-teste vem dada por,

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X'^2}{m} + \dfrac{S_Y'^2}{n}}} \sim t(r),$$

sendo r a parte inteira de,

$$r^* = \frac{\left(\dfrac{s_X'^2}{m} + \dfrac{s_Y'^2}{n}\right)^2}{\dfrac{1}{m-1}\left(\dfrac{s_X'^2}{m}\right)^2 + \dfrac{1}{n-1}\left(\dfrac{s_Y'^2}{n}\right)^2}.$$

Teste de $H_0: \mu_X = \mu_Y$ (com σ_X^2 e σ_Y^2 desconhecidos e diferentes) contra H_1

H_1	**Região crítica**	**valor-*p***
$\mu_X > \mu_Y$	$t > t_\alpha$ ou $\bar{x} - \bar{y} > t_\alpha \sqrt{\frac{s_X'^2}{m} + \frac{s_Y'^2}{n}}$	$P(T \geq t_{\text{obs}} \mid H_0)$
$\mu_X < \mu_Y$	$t < -t_\alpha$ ou $\bar{x} - \bar{y} < -t_\alpha \sqrt{\frac{s_X'^2}{m} + \frac{s_Y'^2}{n}}$	$P(T \leq t_{\text{obs}} \mid H_0)$
$\mu_X \neq \mu_Y$	$\lvert t \rvert > t_{\alpha/2}$ ou $\lvert \bar{x} - \bar{y} \rvert > t_\alpha \sqrt{\frac{s_X'^2}{m} + \frac{s_Y'^2}{n}}$	$2P(T \geq \lvert t_{\text{obs}} \rvert \mid H_0)$
Onde: $t_{\text{obs}} = \dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X'^2}{m} + \frac{s_Y'^2}{n}}}$; $t_\alpha : P(T > t_\alpha) = \alpha$		

Teste para a igualdade das variâncias $H_0 : \sigma_X^2 = \sigma_Y^2$ ou de forma equivalente $H_0 : \dfrac{\sigma_X^2}{\sigma_Y^2} = 1$,

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

Teste de $H_0 : \sigma_X^2 = \sigma_Y^2$ contra H_1

H_1	**Região crítica**	**Valor-*p***
$\sigma_X^2 > \sigma_Y^2$	$F > F_\alpha$ ou $\dfrac{s_X'^2}{s_Y'^2} > F_\alpha$	$P(F \geq F_{\text{obs}} \mid H_0)$
$\sigma_X^2 < \sigma_Y^2$	$F < \dfrac{1}{F_\alpha^*}$ ou $\dfrac{s_X'^2}{s_Y'^2} < \dfrac{1}{F_\alpha^*}$	$P(F \leq F_{\text{obs}} \mid H_0)$
$\sigma_X^2 \neq \sigma_Y^2$	$F < \dfrac{1}{F_{\alpha/2}^*} \vee F > F_{\alpha/2}$ ou $\dfrac{s_X'^2}{s_Y'^2} < \dfrac{1}{F_{\alpha/2}^*} \vee \dfrac{s_X'^2}{s_Y'^2} > F_{\alpha/2}$	2 vezes o menor dos valores acima
Onde: $F_{\text{obs}} = \dfrac{s_X'^2}{s_Y'^2}$; $F_\alpha : P(F > F_\alpha) = \alpha$; $F_\alpha^* : P\left(\dfrac{1}{F} > F_\alpha^*\right) = \alpha$		

Populações não normais – grandes amostras

- Recurso ao teorema do limite central, nomeadamente $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \overset{a}{\sim} N(0,1)$ para efectuar testes sobre a média do universo quando a variância é conhecida.
- No caso em que a variância da população é desconhecida mostra-se que $Z = \dfrac{\overline{X} - \mu}{\hat{\sigma} / \sqrt{n}} \overset{a}{\sim} N(0,1)$, onde $\hat{\sigma}^2$ é um estimador consistente para σ^2, por exemplo a variância (ou a variância corrigida) da amostra.
- Para a diferença de médias utiliza-se $Z = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\hat{\sigma}_1^2}{m} + \dfrac{\hat{\sigma}_2^2}{n}}} \overset{a}{\sim} N(0,1)$, onde $\hat{\sigma}_1^2$ e $\hat{\sigma}_2^2$ são estimadores consistentes para σ_1^2 e σ_2^2, respectivamente.

- **Caso particular 1 – Universo de Bernoulli**

Teste de $H_0 : \theta = \theta_0$ (numa população de Bernoulli) contra H_1

H_1	**Região crítica**	**valor-*p***
$\theta > \theta_0$	$z > z_\alpha$ ou $\bar{x} > \theta_0 + z_\alpha \sqrt{\dfrac{\theta_0(1-\theta_0)}{n}}$	$P(Z \geq z_{\text{obs}} \mid H_0)$
$\theta < \theta_0$	$z < -z_\alpha$ ou $\bar{x} > \theta_0 - z_\alpha \sqrt{\dfrac{\theta_0(1-\theta_0)}{n}}$	$P(Z \leq z_{\text{obs}} \mid H_0)$
$\theta \neq \theta_0$	$\lvert z \rvert > z_{\alpha/2}$ ou $\bar{x} < \theta_0 - z_{\alpha/2} \sqrt{\dfrac{\theta_0(1-\theta_0)}{n}} \vee \bar{x} > \theta_0 + z_{\alpha/2} \sqrt{\dfrac{\theta_0(1-\theta_0)}{n}}$	$2P\left(Z \geq \lvert z_{\text{obs}} \rvert \middle\vert H_0\right)$
Onde: $z_{\text{obs}} = \dfrac{\bar{x} - \theta_0}{\sqrt{\dfrac{\theta_0(1-\theta_0)}{n}}}$; $z_\alpha : P(Z > z_\alpha) = \alpha$		

- **Caso particular 2 – Comparação de 2 universos de Bernoulli**

Teste de $H_0 : \theta_X = \theta_Y$ (em populações de Bernoulli) contra H_1

H_1	**Região crítica**	**valor-*p***
$\theta_X > \theta_Y$	$z > z_\alpha$ ou $\bar{x} - \bar{y} > z_\alpha \sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{m} + \frac{1}{n}\right)}$	$P(Z \geq z_{\text{obs}} \mid H_0)$
$\theta_X < \theta_Y$	$z < -z_\alpha$ ou $\bar{x} - \bar{y} < -z_\alpha \sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{m} + \frac{1}{n}\right)}$	$P(Z \leq z_{\text{obs}} \mid H_0)$
$\theta_X \neq \theta_Y$	$\lvert z \rvert > z_{\alpha/2}$ ou $\lvert \bar{x} - \bar{y} \rvert > z_{\alpha/2} \sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{m} + \frac{1}{n}\right)}$	$2P\left(Z \geq \lvert z_{\text{obs}} \rvert \mid H_0\right)$
Onde: $z_{\text{obs}} = \dfrac{\bar{x} - \bar{y}}{\sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{m} + \frac{1}{n}\right)}}$; $\hat{\theta} = \dfrac{m\,\bar{x} + n\,\bar{y}}{m+n}$; $z_\alpha : P(Z > z_\alpha) = \alpha$		