Probabilidades e Estatística

Testes de Hipóteses

Prof. Caldeira Duarte e Prof.ª Anabela Pereira

Actualizado pelo Prof. António Sardinha (Fevereiro de 2013)
Actualizado pela Prof.ª Dina Salvador (Fevereiro de 2017)

Departamento de Matemática

IPS Instituto Politécnico de Setúbal
Escola Superior de Tecnologia de Setúbal

6 TESTES DE HIPÓTESES

6.1 Generalidades

Todos os dias temos de tomar decisões respeitantes a determinadas populações, com base em amostras das mesmas (decisões estatísticas). Nesta tomada de decisões é útil formular hipóteses sobre as populações, hipóteses essas que podem ou não ser verdadeiras. A essas hipóteses chamamos hipóteses estatísticas, as quais geralmente se baseiam em afirmações sobre as distribuições de probabilidade das populações ou alguns dos seus parâmetros. Por vezes estas hipóteses, ao serem formuladas, têm por único objectivo serem rejeitadas.

Exemplo 6.1. Se queremos decidir se uma dada moeda está viciada, formulamos a hipótese de que a moeda seja "*honesta*", isto é, que a probabilidade de sair por exemplo cara seja $p = 0.5$. Da mesma forma, se queremos decidir se um produto é melhor do que outro, podemos formular a hipótese de que não existe diferença entre ambos os produtos. ∎

Desta forma os testes de hipóteses podem considerar-se uma segunda vertente da inferência estatística, tendo por objectivo verificar, a partir de dados observados numa ou várias amostras, a validade de certas hipóteses relativas a uma ou várias populações.

6.2 Princípios da realização dos testes de hipóteses

1. De uma forma geral emite-se uma certa hipótese a testar denominada **Hipótese Nula** e representada por H_0:

   (a) em seguida medimos o desvio observado em certas características da amostra e calculamos a probabilidade, se H_0 for verdadeira, do desvio ser "importante";

   (b) se a probabilidade anterior for "relativamente elevada" (isto é, superior a um nível de significância, α, previamente definido), consideramos plausivel H_0 e aceitamo-la, pelo menos provisoriamente; quando um teste não rejeita H_0 não se pode concluir que esta seja verdadeira, mas apenas que não está em desacordo com os factos observados, como tal utiliza-se a expressão *não rejeitar* H_0 em vez de *aceitar* H_0;

   (c) se, pelo contrário, a probabilidade for "pequena" (isto é, inferior a um nível de significância, α, previamente definido), o desvio observado mostra-se pouco compatível com H_0 e rejeitamo-la. Desta forma admitimos, implicitamente, a validade da outra hipótese, denominada por **Hipótese Alternativa** e representada por H_1.

2. O conjunto dos valores observados para os quais H_0 é admissível forma a **Região de Aceitação** (representada por RA). Os restantes valores formam a **Região de Rejeição** ou **Região Crítica** (representada por RC).

3. Consoante o número de elementos em análise num teste, #, podemos distinguir diferentes formas de especificar H_0 (que traduz a situação estacionária, sendo usual colocar nesta hipótese a igualdade) e H_1, considerando, por exemplo, θ^* estimador de θ:

(a) hipótese simples (ou composta) contra hipótese composta (em que $\#\{\theta_{H_0}\} = 1$ (ou $\#\{\theta_{H_0}\} > 1$) e $\#\{\theta_{H_1}\} > 1$). Podemos neste tipo de testes estar perante,

i. **Teste Bilateral** que apresenta duas regiões críticas como vemos na Figura 6.1.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

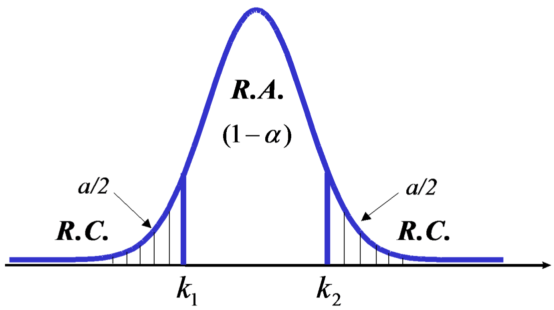

Figura 6.1: Teste Bilateral.

Como,

$$P\left(\theta^* \geq k_1\right) = P\left(\theta^* \leq k_2\right)$$

então,

$$P\left(\text{Rej. } H_0/H_0V\right) = P(\theta^* \in RC/\theta = \theta_0) = \alpha \Leftrightarrow \begin{cases} P\left(\theta^* \leq k_1/\theta = \theta_0\right) = \frac{\alpha}{2} \\ P\left(\theta^* \geq k_2/\theta = \theta_0\right) = \frac{\alpha}{2} \end{cases}.$$

ii. **Teste Unilateral Esquerdo** que apresenta a região crítica à esquerda como vemos na Figura 6.2.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

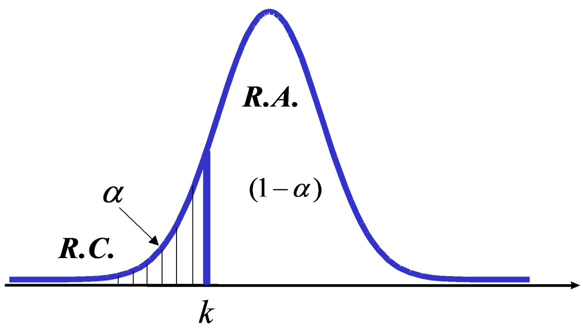

Figura 6.2: Teste Unilateral Esquerdo.

$$P\left(\text{Rej. } H_0/H_0V\right) = P(\theta^* \in RC/\theta = \theta_0) = P\left(\theta^* \leq k/\theta = \theta_0\right) = \alpha.$$

iii. **Teste Unilateral Direito** que apresenta a região crítica à direita como vemos na Figura 6.3.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

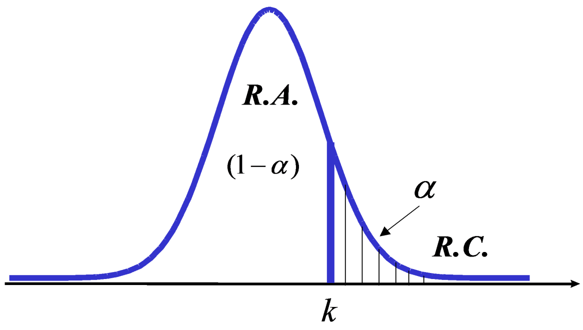

Figura 6.3: Teste Unilateral Direito.

$$P\left(\text{Rej. } H_0/H_0V\right) = P(\theta^* \in RC/\theta = \theta_0) = P\left(\theta^* \geq k/\theta = \theta_0\right) = \alpha.$$

(b) hipótese simples contra hipótese simples (em que $\#\left\{\theta_{H_0}\right\} = 1$ e $\#\left\{\theta_{H_1}\right\} = 1$).

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

Neste caso estamos sempre perante um teste unilateral considerado esquerdo, se $\theta_0 > \theta_1$, ou um teste unilateral considerado direito se $\theta_0 < \theta_1$.

4. Existe uma relação entre a teoria da estimação, envolvendo intervalos de confiança, e a teoria relativa aos testes de hipóteses. **Quando trabalhamos com testes de hipóteses bilaterais podemos efectivamente utilizar os intervalos de confiança para testar hipóteses (pois o intervalo de confiança coincide com a região de aceitação, e portanto rejeita-se H_0 se o valor do parâmetro especificado em H_0 não pertencer ao intervalo de confiança).** Resultado análogo para testes unilaterais, exigiriam intervalos de confiança unilaterais, os quais, embora de rara aplicação prática, são possíveis de definir, mas que não foram abordados no capítulo anterior.

5. Um teste de hipóteses nem sempre conduz a decisões correctas pois a análise de uma amostra pode, como é evidente, falsear as conclusões. Como tal podemos encontrar-nos perante quatro situações distintas apresentadas na tabela seguinte:

Decisão Tomada	Situação Real H_0 Verdadeira	H_0 Falsa
Rejeita-se H_0	Erro de 1^a espécie (α)	Decisão correcta (π)
Não se rejeita H_0	Decisão correcta	Erro de 2^a espécie (β)

(a) Num erro de 1^a espécie (cuja probabilidade se representa por α, ou **nível de significância do teste**) rejeita-se H_0, sendo esta verdadeira, logo

$$\alpha = P(\text{Rej. } H_0/H_0V).$$

(b) Num **erro de 2^a espécie** (cuja probabilidade se representa por β) não se rejeita H_0, sendo esta falsa (ou H_1 verdadeira), logo

$$\beta = P(\text{Não Rej. } H_0/H_0F) = P(\text{Não Rej. } H_0/H_1V).$$

Num teste unilateral direito, os dois erros podem geometricamente representar-se como mostra a Figura 6.4 em que a função cujo gráfico está a tracejado representa o comportamento do verdadeiro parâmetro da população.

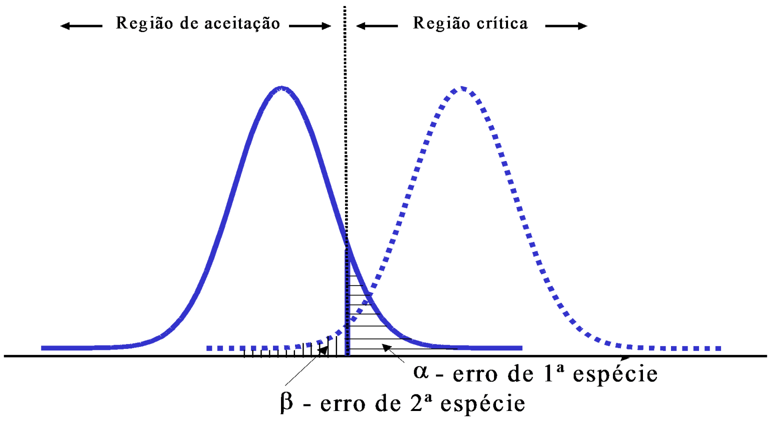

Figura 6.4: Erros de 1^a e 2^a espécie.

Num teste bilateral, os dois erros podem geometricamente representar-se como mostra a Figura 6.5, em que as três últimas funções representam comportamentos possíveis da função densidade de probabilidade da população, com os respectivos erros de 1 ^{a} e 2 ^{a} espécies.

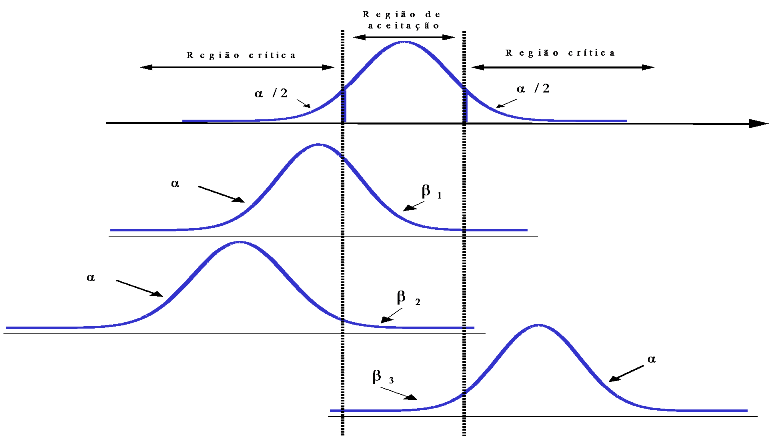

Figura 6.5: Erros num Teste Bilateral.

(c) Chama-se **função potência** de um teste e representa-se por π à probabilidade de rejeitar H_0 quando esta é falsa (decisão correcta). Então podemos dizer que dado o erro de 2 ^{a} espécie β, a função potência é o seu complementar

$$\pi = P(\text{Rej. } H_0/H_0F) = P(\text{Rej. } H_0/H_1V) = 1 - \beta.$$

Esta probabilidade é função do grau de falsidade de H_0, logo a probabilidade de rejeição é tanto mais elevada, quanto mais falsa for H_0. Conclui-se então que a relação entre a probabilidade de rejeição de H_0 e o grau de falsidade da mesma constituem a função potência do teste, isto é, quanto maiores forem os valores da função potência, menor é o erro de 2 ^{a} espécie cometido, logo, melhor a qualidade do teste (teste mais potente).

Num teste bilateral o gráfico da função potência tem a forma de um V como se visualiza na Figura 6.6. Um V estreito (com um declive acentuado) indica que o valor do parâmetro definido na hipótese nula e os diversos valores da hipótese alternativa estão bem discriminados; se pelo contrário, o V for largo, indica uma fraca discriminação nos valores formulados nas hipóteses.

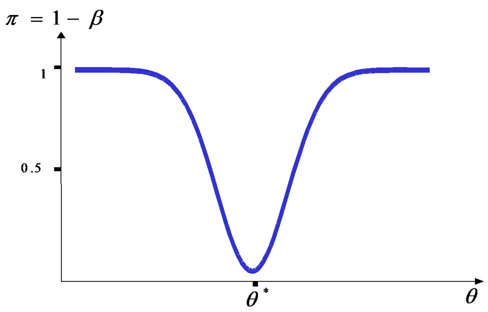

Figura 6.6: Função potência de um Teste Bilateral.

Num teste unilateral direito o gráfico da função potência tem a forma de um S como se observa na Figura 6.7. Mais uma vez, o declive acentuado indica que o valor do parâmetro definido na hipótese nula e diversos valores da hipótese alternativa estão bem discriminados; se pelo contrário o declive for pouco acentuado indica uma fraca discriminação nos valores formulados nas hipóteses.

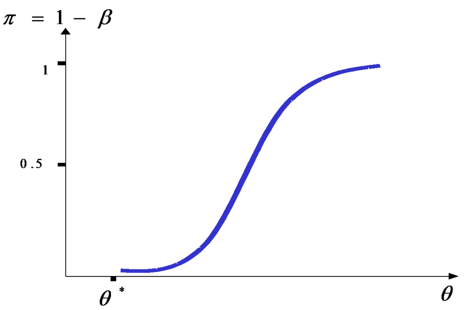

Figura 6.7: Função potência num Teste Unilateral Direito.

(d) É através das probabilidades α e β que se procura o melhor teste de hipóteses, sendo o teste ideal o que minimiza simultâneamente ambos os valores. No entanto, e como α e β variam em sentidos contrários, tal não é possível. O que na maior parte dos casos se faz (com base no Teorema de *Neyman-Pearson* [1]) consiste em fixar α (para amostras de dimensão n) para tentar minimizar β. Note-se ainda que é possível fixar α e β *a priori*, ficando n livre; no entanto este método que se baseia em valores pequenos de α e β conduz a valores de n grandes, o que nem sempre é conveniente.

6. Os erros anteriores não podem ser completamente evitados, no entanto, pode-se manter pequena a probabilidade de os cometer. Na prática fixa-se um limite superior de risco de erro de 1 ^{a} espécie (α), que na maior parte dos casos se situa entre 1% e 5% ($\alpha = 0.01$ até $\alpha = 0.05$). Este limite, ou nível de significância do teste, é que permite definir a condição de rejeição de H_0.

6.3 Testes de Hipóteses Paramétricos

Nos testes de hipóteses paramétricos ou realizados a parâmetros de uma população, e ao contrário dos intervalos de confiança, em vez de procurarmos uma estimativa ou um intervalo para o parâmetro, vamos admitir um valor hipotético para o mesmo e depois utilizar a informação da amostra para rejeitar ou não esse valor. Nos casos que em seguida apresentamos, vamos debruçar-nos apenas sobre populações com distribuições normais (ou aproximadamente normais).

Comecemos por enunciar, de uma forma geral, **duas metodologias equivalentes** a utilizar num teste de hipóteses paramétrico.

Passemos a enunciar, de uma forma geral, uma **primeira metodologia**:

1. formulação das hipóteses (identificando o parâmetro a testar, θ , e o respectivo estimador, θ^*);
2. fixação do erro de 1 ^{a} espécie ou nível de significância do teste
$$\alpha = P(\text{Rej. } H_0/H_0V);$$
3. escolha da estatística (também denominada por estatística de teste ou variável fulcral) e respectiva distribuição amostral adequadas;
4. cálculo de RC a partir do nível de significância do teste, α;
5. com base na amostra calcula-se o estimador θ^* do parâmetro θ, e aplica-se a regra de decisão:
$$\begin{cases} \text{se } \theta^* \in RC \Rightarrow \text{rejeitar } H_0 \\ \text{se } \theta^* \in RA \Rightarrow \text{não rejeitar } H_0. \end{cases}$$

[1] Página 307 e seguintes de Bento Murteira, *Probabilidades e Estatística*, Volume II, McGraw-Hill,1990.

De seguida enuncia-se, de uma forma geral, uma **segunda metodologia:**

1. formulação das hipóteses (identificando o parâmetro a testar, θ);
2. fixação do erro de 1 ^{a} espécie ou nível de significância do teste
$$\alpha = P(\text{Rej. } H_0/H_0V);$$
3. escolha da estatística (também denominada por estatística de teste ou variável fulcral) e respectiva distribuição amostral adequadas;
4. cálculo de RC a partir do nível de significância do teste, α, e com base na distribuição amostral;
5. com base na amostra calcula-se o valor observado da estatística de teste, U^*, e aplica-se a regra de decisão:
$$\begin{cases} \text{se } U^* \in RC \Rightarrow \text{rejeitar } H_0 \\ \text{se } U^* \in RA \Rightarrow \text{não rejeitar } H_0. \end{cases}$$

6.3.1 Testes de Hipóteses para a Média

Neste caso a estatística a utilizar é

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Como vimos no capítulo anterior, consoante os restantes parâmetros sejam ou não conhecidos e a dimensão da amostra seja grande ou pequena, vamos utilizar diferentes estatísticas de teste e respectivas distribuições amostrais. Para os testes de hipóteses este procedimento repete-se, logo vamos utilizar:

1. se σ é conhecido,
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1);$$
2. se σ é desconhecido e $n \geq 30$,
$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim \mathcal{N}(0, 1);$$
3. se σ é desconhecido e $n < 30$,
$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}.$$

Exemplo 6.2. De um universo normal, de média e variância desconhecidas, foi retirada uma amostra aleatória de 9 observações, cujos resultados foram:

$$\sum_{i=1}^{9} x_i = 36 \quad \text{e} \quad \sum_{i=1}^{9} x_i^2 = 162.$$

Proceda ao seguinte ensaio de hipóteses:

$$\begin{cases} H_0 : \mu = 5 \\ H_1 : \mu = 6 \end{cases}$$

para um nível de significância de 5%.

A estatística para o estudo do parâmetro μ é $\bar{X}$. Como desconhecemos a variância da população e $n < 30$, utilizamos a estatística de teste e a distribuição amostral:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{9}}} \sim t_{(8)}.$$

Impõe-se então calcular $\bar{x}$ e s:

$$\bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = 4$$

$$s^2 = \frac{1}{8}\left(\sum_{i=1}^{9} x_i^2 - 9\bar{x}^2\right) = \frac{1}{8}\left(162 - 9\,(4)^2\right) = 2.25 \Rightarrow s = 1.5.$$

Partindo de

$$T = \frac{\bar{X} - \mu}{\frac{1.5}{\sqrt{9}}} \sim t_{(8)}$$

e de $\alpha = 0.05$, vamos calcular RC de um teste unilateral direito (dado que H_1 está sempre associada a RC) como podemos ver na Figura 6.8.

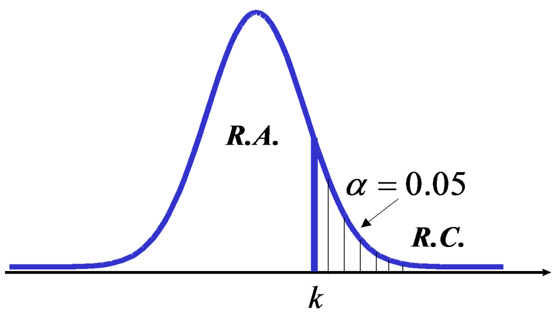

Figura 6.8: Região crítica associada ao teste.

Para tal podemos seguir **duas metodologias equivalentes** mas que correspondem a escalas diferentes no cálculo da região crítica.

Na primeira metodologia, k (que separa RA de RC) é obtido a partir da estatística $\bar{X}$, assim como a tomada de decisão consiste em verificar se a estimativa da média amostral se situa em RC ou RA.

Na segunda metodologia, k é obtido a partir da estatística de teste T, assim como a tomada de decisão consiste em verificar se a estimativa da estatística de teste, T^*, se situa em RC ou RA.

Ambos os procedimentos são equivalentes, variando apenas a escala utilizada para o cálculo de RC, assim como para a tomada de decisão.

Vamos começar por resolver este exemplo através da primeira metodologia:

$$P(\text{Rej. } H_0/H_0V) = \alpha \Leftrightarrow P\left(\bar{X} \in RC/H_0V\right) = P\left(\bar{X} \geq k/\mu = 5\right) = 0.05 \Leftrightarrow$$
$$\Leftrightarrow P\left(T \geq \frac{k-\mu}{\frac{1.5}{\sqrt{9}}}/\mu = 5\right) = 0.05 \Leftrightarrow P\left(T < \frac{k-5}{\frac{1.5}{3}}\right) = 0.95 \Leftrightarrow$$
$$\Leftrightarrow \frac{k-5}{\frac{1.5}{3}} = 1.86 \Leftrightarrow k = 5.93.$$

Então $RC = [5.93, +\infty[$, como podemos visualizar na Figura 6.9.

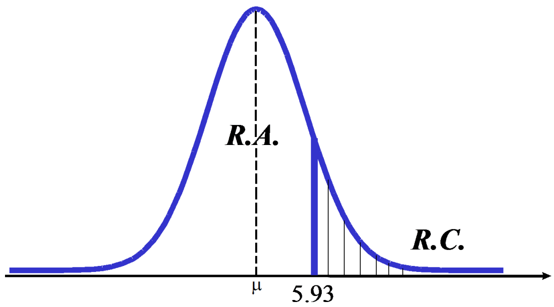

Figura 6.9: Região crítica calculada em função de $\bar{X}$.

Como $\bar{x} = 4 < 5.93$ se encontra na região de aceitação (RA), não se rejeita H_0.

Na segunda metodologia, k é obtido a partir da estatística de teste T, sendo RC calculado a partir do quantil que lhe corresponde, isto é, a RC começa a partir do quantil referente a $(1-\alpha) = (1-0.05) = 0.95$. Como

$$t_{(8);0.95} = 1.86, \text{ então, } RC = [1.86, +\infty[\,.$$

Podemos desta forma visualizar na Figura 6.10 a mudança de escala da região crítica utilizando a 2 ^{a} metodologia.

Como a estimativa de T é dada por

$$T^* = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{9}}} = \frac{4-5}{\frac{1.5}{3}} = -2.0 \in RA,$$

não se rejeita H_0. ■

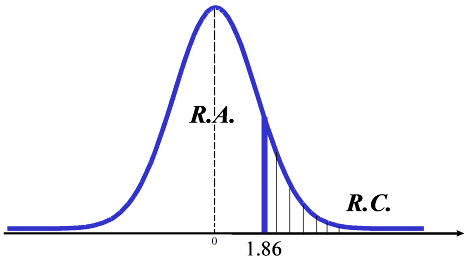

Figura 6.10: Região Crítica calculada em função de T.

Exemplo 6.3. Para $X \sim \mathcal{N}(\mu, 100)$, $n = 25$, $\bar{x} = 980$ e $\alpha = 0.05$, vamos calcular RC, erros de 2^a espécie e a função potência para

$$\begin{cases} H_0 : \mu = 1000 \\ H_1 : \mu < 1000 \end{cases}$$

A estatística para o estudo do parâmetro μ é $\bar{X}$. Como conhecemos a variância da distribuição, utilizamos a estatística de teste e a distribuição amostral

$$Z = \frac{\bar{X} - \mu}{\frac{100}{\sqrt{25}}} \sim \mathcal{N}(0, 1).$$

Temos um teste unilateral esquerdo tal como podemos ver na Figura 6.11.

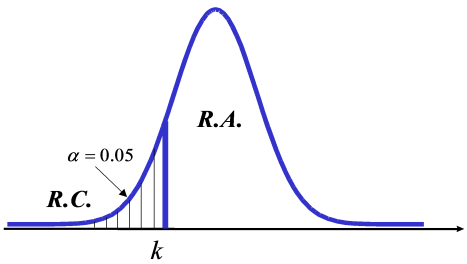

Figura 6.11: Região crítica associada ao teste.

Utilizando a primeira metodologia temos:

$$P(\text{Rej. } H_0/H_0V) = \alpha \Leftrightarrow P\left(\bar{X} \in RC/H_0V\right) = P\left(\bar{X} \leq k/\mu = 1000\right) = 0.05 \Leftrightarrow$$
$$\Leftrightarrow P\left(Z \leq \frac{\bar{X} - \mu}{\frac{100}{5}}/\mu = 1000\right) = 0.05 \Leftrightarrow P\left(Z \leq \frac{k - 1000}{20}\right) = 0.05 \Leftrightarrow$$
$$\Leftrightarrow \frac{k - 1000}{20} = -1.645 \Leftrightarrow k = 967.1$$

Então $RC =]-\infty, 967.1]$ como se visualiza na Figura 6.12.

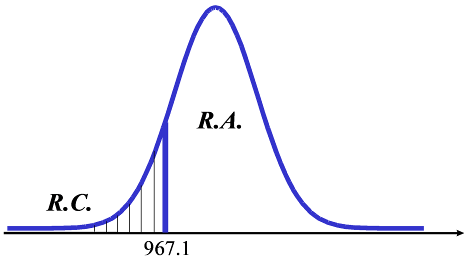

Figura 6.12: Região crítica calculada em função de $\bar{X}$.

Como $\bar{x} = 980 \in RA$, não se rejeita H_0.

Utilizando a segunda metodologia,

$$z_{0.05} = -z_{0.95} = -1.645, \text{ isto é, } RC = \left]-\infty, -1.645\right].$$

Como a estimativa de Z é dada por

$$Z^* = \frac{\bar{x} - \mu}{\frac{100}{\sqrt{25}}} = \frac{980 - 1000}{\frac{100}{\sqrt{25}}} = -1 \in RA,$$

não se rejeita H_0.

Embora a segunda metodologia seja mais rápida, a primeira é mais directa quando pretendemos calcular o erro de 2 ^{a} espécie, como em seguida veremos:

$$\begin{aligned} \beta &= P(\text{Não Rej. } H_0/H_0F) = P(\bar{X} \in RA/H_1V) = \\ &= P(\bar{X} > k/\mu < 1000) = P(\bar{X} > 967.1/\mu < 1000) = \\ &= 1 - P\left(Z \leq \frac{967.1 - \mu}{20}/\mu < 1000\right). \end{aligned}$$

Atribuindo alguns valores a μ, por exemplo 999, 990, 970, 950, 930 e 910, calculamos o respectivo erro de 2 ^{a} espécie e correspondente função potência cujos valores se encontram na tabela seguinte:

μ	$\beta(\mu)$	$\pi(\mu)$
999	0.9446	0.0554
990	0.8729	0.1271
970	0.5557	0.4443
950	0.1977	0.8023
930	0.0318	0.9682
910	0.0022	0.9978

Podemos ainda expressar graficamente estas duas funções através da Figura 6.13.

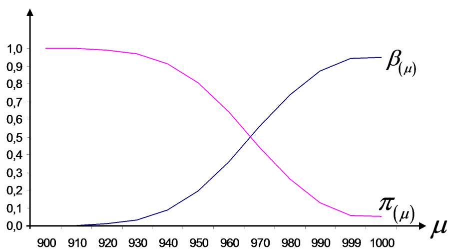

Figura 6.13: Erro de 2^a espécie e correspondente função potência do teste.

■

6.3.2 Testes de Hipóteses para a Diferença de Duas Médias

Nestes casos há que diferenciar mais uma vez as estatísticas de teste e respectivas distribuições amostrais a utilizar:

1. se os desvios padrão são conhecidos,

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1);$$

2. se os desvios padrão são desconhecidos, $n_1 \geq 30$ e $n_2 \geq 30$,

$$T = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{N}(0,1);$$

3. se os desvios padrão são desconhecidos (e iguais), $n_1 < 30$ ou $n_2 < 30$,

$$T = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}} \sim t_{(n_1+n_2-2)}.$$

Exemplo 6.4. A altura média de 50 atletas de um dado clube que tiveram bons resultados em competições desportivas, é de 68.2 polegadas, com desvio padrão de 2.5 polegadas, enquanto que um grupo de 50 atletas do mesmo clube com resultados inferiores nessas competições tem altura média de 67.5 polegadas com desvio padrão de 2.8 polegadas. Vamos testar a hipótese de que os atletas que obtiveram bons resultados nas competições são, em média, mais altos do que os restantes (com $\alpha = 0.05$).

Devemos então proceder ao teste de hipóteses:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$$

o que significa confrontar a inexistência de diferença entre as médias das alturas dos dois grupos de atletas, contra a altura média do 1 ^{o} grupo de atletas ser superior à do 2 ^{o} grupo,

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}.$$

Como os desvios padrão são desconhecidos, $n_1 \geq 30$ e $n_2 \geq 30$ utilizamos a estatística de teste e a distribuição amostral

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}}} \sim \mathcal{N}(0,1).$$

Vamos resolver o exemplo recorrendo à segunda metodologia.

Como estamos perante um teste unilateral direito, RC começa a partir do quantil associado a $(1-\alpha) = (1-0.05) = 0.95$, logo,

$$RC = [z_{0.95}, +\infty[\, = [1.645, +\infty[\,.$$

Sendo $(\bar{x}_1 - \bar{x}_2) = (68.2 - 67.5) = 0.7$, a estimativa de Z é dada por

$$Z^* = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}}} = \frac{0.7 - 0}{\sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}}} = 1.319 \in RA.$$

Conclui-se que não se rejeita H_0, isto é, não se prova que a altura média dos dois grupos de atletas difira para o valor de α considerado. ■

A tomada de decisão face ao resultado de um teste de hipóteses não dá garantia de que estejamos a agir de forma correcta pois basta alterar o tipo de teste (por exemplo de unilateral para bilateral), o nível de significância ou a dimensão da amostra, para que o resultado do teste possa ser completamente diferente. Esta situação está ilustrada no exemplo seguinte.

Exemplo 6.5. Se, para o exemplo anterior, quiséssemos que a diferença observada entre as alturas médias, de 0.7 polegadas, fosse significativa, qual deveria ser a dimensão das amostras (mantendo a igualdade entre as mesmas)?

Neste caso, pretendemos que $Z^* \in RC$, isto é,

$$\begin{aligned}
Z^* \geq 1.645 &\Leftrightarrow \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{2.5^2}{n} + \frac{2.8^2}{n}}} \geq 1.645 \Leftrightarrow \frac{0.7 - 0}{\sqrt{\frac{2.5^2}{n} + \frac{2.8^2}{n}}} \geq 1.645 \Leftrightarrow \\
&\Leftrightarrow 1.645\sqrt{\frac{2.5^2}{n} + \frac{2.8^2}{n}} \leq 0.7 \Leftrightarrow 1.645\frac{\sqrt{14.09}}{\sqrt{n}} \leq 0.7 \Leftrightarrow \\
&\Leftrightarrow \sqrt{n} \geq \frac{1.645 \times \sqrt{14.09}}{0.7} \Rightarrow n \geq 78.
\end{aligned}$$

■

Exemplo 6.6. Os quocientes de inteligência (QI) de 16 estudantes de um dado bairro de uma cidade apresentaram uma média de 107 com um desvio padrão de 10; entretanto, noutro bairro da mesma cidade, analisaram-se 14 estudantes cujos QI tinham uma média de 112 e um desvio padrão de 8. Há diferenças significativas entre os QI dos dois grupos (considere $\alpha = 0.01$ e as populações normais com $\sigma_1 = \sigma_2$)?

Para resolver esta questão vamos elaborar o seguinte teste de hipóteses bilateral:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

O que significa confrontar a inexistência de diferença significativa entre as médias dos QI contra a existência dessa mesma diferença,

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}.$$

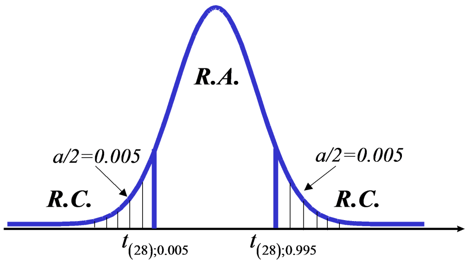

Figura 6.14: Teste bilateral.

Como as populações são Normais, $\sigma_1 = \sigma_2$, $n_1 < 30$ e $n_2 < 30$, vamos utilizar,

$$T = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{16} + \frac{1}{14}\right) \frac{15\times10^2+13\times8^2}{28}}} \sim t_{(28)}.$$

Estamos perante um teste bilateral traduzido na Figura 6.14.

A região crítica calcula-se fazendo,

$$RC = \left]-\infty, t_{(28);0.005}\right] \cup \left[t_{(28);0.995}, +\infty\right[=]-\infty, -2.763] \cup [2.763, +\infty[\,.$$

Como $(\bar{x}_1 - \bar{x}_2) = (107 - 112) = -5$, a estimativa de T é dada por

$$T^* = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{16} + \frac{1}{14}\right) \frac{15\times10^2+13\times8^2}{28}}} = \frac{-5-0}{\sqrt{\frac{8745}{784}}} = -1.4971 \in RA$$

não se rejeita H_0, isto é, conclui-se que não existem diferenças significativas entre as médias dos QI dos dois grupos. ■

6.3.3 Teste de Hipóteses para uma Proporção

Neste caso, e considerando p^* a proporção observada na amostra, estimativa da proporção desconhecida (p) da população, vamos utilizar a variável aleatória

$$Z = \frac{p^* - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0,1)$$

para grandes amostras (na prática $n \geq 30$).

Chama-se a atenção para o facto da estatística utilizada nos Intervalos de Confiança ser uma aproximação desta estatística de teste (no respeitante à variância de p^*).

6.3.4 Teste de Hipóteses para a Diferença de Duas Proporções

Neste caso, e considerando p_1^* e p_2^* as proporções observadas nas amostras, estimativas das proporções desconhecidas (p_1 e p_2) das populações, vamos utilizar

$$Z = \frac{(p_1^* - p_2^*) - (p_1 - p_2)}{\sqrt{\frac{p_1^*\left(1-p_1^*\right)}{n_1} + \frac{p_2^*\left(1-p_2^*\right)}{n_2}}} \sim \mathcal{N}(0,1)$$

para grandes amostras (na prática $n_1 \geq 30$ e $n_2 \geq 30$).

Exemplo 6.7. Sabendo que existem dois grupos de individuos X e Y (cada um com 100 indivíduos) portadores de uma doença, aplica-se um antibiótico apenas ao 1 ^{o} grupo. De resto, ambos os grupos são tratados em condições idênticas. Constata-se que, nos grupos X e Y 73 e 65 dos indivíduos, respectivamente, se curaram da doença. Teste a hipótese de que o antibiótico não é eficiente para o nível de significância de 0.01.

Vamos considerar p_1 e p_2 as proporções das populações curadas, aplicando-se o antibiótico e não se aplicando o mesmo respectivamente. Devemos então decidir entre as hipóteses:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 > p_2 \end{cases}$$

o que significa confrontar a inexistência de diferenças entre proporções (antibiótico ineficiente) contra proporção de indivíduos curados no primeiro grupo ser superior à do segundo (antibiótico eficiente),

$$\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 > 0 \end{cases}$$

Utilizamos para o efeito,

$$Z = \frac{(p_1^* - p_2^*) - (p_1 - p_2)}{\sqrt{\frac{0.73\times0.27}{100} + \frac{0.65\times0.35}{100}}} \dot{\sim} \mathcal{N}(0,1).$$

Uma vez que a região crítica se situa à direita e $\alpha = 0.01$, então

$$RC = [z_{0.99}, +\infty[= [2.326, +\infty[.$$

Como $(p_1^* - p_2^*) = (0.73 - 0.65) = 0.08$, a estimativa de Z é dada por

$$Z = \frac{(p_1^* - p_2^*) - (p_1 - p_2)}{\sqrt{\frac{0.73\times0.27}{100} + \frac{0.65\times0.35}{100}}} = \frac{0.08 - 0}{\sqrt{\frac{0.73\times0.27}{100} + \frac{0.65\times0.35}{100}}} = 1.228 \in RA$$

não se rejeita H_0, não se podendo concluir que o antibiótico seja eficiente. ■

6.3.5 Testes de Hipóteses para a Variância

Neste caso, para populações Normais, vamos utilizar:

1. se μ é conhecido,
$$X^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_{(n)};$$

2. se μ é desconhecido,
$$X^2 = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Exemplo 6.8. O peso dos pacotes cheios por uma máquina de empacotamento tem uma distribuição normal com desvio padrão 0.25 kg. Extraíndo uma amostra de 20 pacotes registou-se um desvio padrão de 0.32 kg. Este aumento de variabilidade é significativo ao nível de significância de 5%?

Vamos considerar o teste de hipóteses:

$$\begin{cases} H_0 : \sigma = 0.25 \\ H_1 : \sigma > 0.25 \end{cases}$$

o que significa confrontar a inexistência do aumento de variabilidade contra o aumento da mesma.

Como a população é Normal e μ é desconhecido, utilizamos

$$X^2 = 19\frac{S^2}{\sigma^2} \sim \chi^2_{(19)}.$$

Uma vez que a região critíca se situa à direita e $\alpha = 0.05$, então,

$$RC = \left[\chi^2_{(19);0.95}, +\infty\right[= [30.1435, +\infty[\,.$$

Como $s^2 = 0.32^2 = 0.1024$, a estimativa de X^2 é dada por

$$X^2 = 19\frac{s^2}{\sigma^2} = 19\frac{0.1024}{0.25^2} = 31.13 \in RC,$$

rejeitando-se H_0, isto é, conclui-se que há um aumento significativo na variabilidade do peso dos pacotes. ■

6.3.6 Teste de Hipóteses para a Razão de Duas Variâncias

Neste teste de hipóteses, para duas populações Normais, vamos usar:

$$F = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n_1-1, n_2-1)}.$$

Exemplo 6.9. Um professor de estatística tem duas turmas, cujas notas têm um comportamento normal. A turma A tem 16 alunos e a turma B tem 21 alunos. Num exame, embora não tenha havido diferença significativa entre as notas médias, a turma A registou um desvio padrão de 9 e a turma B de 12Ṗodemos concluir que a variabilidade da turma B é superior à da turma A ($\alpha = 0.01$)?

Vamos considerar o teste de hipóteses, utilizando os indíces 1 e 2 para as turmas A e B respectivamente:

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 < \sigma_2 \end{cases}$$

O que significa confrontar a inexistência de diferença de variabilidade entre as notas das duas turmas, contra a variabilidade das notas da turma B ser superior à da turma A,

$$\begin{cases} H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1 \end{cases}$$

A estatística de teste e a distribuição amostral a utilizar é,

$$F = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(15,20)}.$$

Dado que a região critíca se situa à esquerda e $\alpha = 0.01$, temos

$$f_{(15,20);0.01} = \frac{1}{f_{(20,15);0.99}} = \frac{1}{3.09} = 0.32362.$$

Logo,

$$RC = \left]0, f_{(15,20);0.01}\right] =]0, 0.32362] .$$

Como $s_1 = 0.09$ e $s_2 = 0.12$, então

$$s_1^2 = 0.0081, \quad s_2^2 = 0.0144 \quad \text{e} \quad \frac{s_1^2}{s_2^2} = 0.5625$$

sendo a estimativa de F dada por

$$F^* = \frac{s_1^2}{s_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} = 0.5625 \times 1 = 0.5625 \in RA.$$

Conclui-se pela não rejeição de H_0, isto é, segundo os dados do problema não existe diferença de variabilidade significativa entre as notas das duas turmas. ■

6.3.7 Valor-p

No decorrer deste capítulo abordámos e exemplificámos duas metodologias que nos permitem tomar uma decisão em testes de hipóteses. Vamos agora apresentar uma **terceira metodologia**, que aqui salientamos, não só devido à sua vasta aplicação, como à flexibilidade decorrente da sua interpretação. Esta metodologia recorre à análise do valor-p (p_v) ou *p-value*, que representa a probabilidade associada ao valor da estatística de teste, considerando H_0 verdadeira e atendendo à definição de H_1.

O valor-p pode interpretar-se como o menor valor de α para o qual os dados observados indicam que H_0 deve ser rejeitada; como tal, a regra de decisão consiste em:

$$\begin{cases} \text{se } p_v \leq \alpha \Rightarrow & \text{rejeitar } H_0; \\ \text{se } p_v > \alpha \Rightarrow & \text{não rejeitar } H_0. \end{cases}$$

Numa distribuição amostral simétrica, se estivermos perante um teste unilateral esquerdo e se $p_v \leq \alpha$ ($p_v \in RC$) rejeitamos H_0 (ver a Figura 6.15).

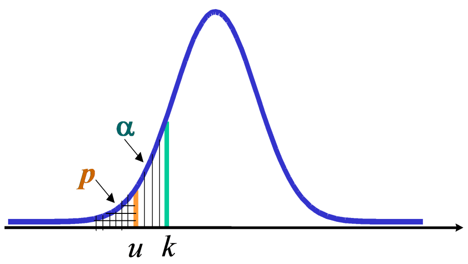

Figura 6.15: Teste Unilateral Esquerdo.

Desta forma é possível fazer uma análise da tomada de decisão para múltiplos valores do nível de significância, permitindo uma visão mais rica do teste realizado.

À semelhança do que foi feito nas anteriores metodologias, enuncia-se, de uma forma geral, a **terceira metodologia:**

1. formulação das hipóteses (identificando o parâmetro a testar, θ);
2. fixação do erro de 1 ^{a} espécie ou nível de significância do teste

$$\alpha = P(\text{Rej. } H_0/H_0V);$$

3. escolha da estatística (também denominada por estatística de teste ou variável fulcral) e respectiva distribuição amostral adequadas;
4. com base na amostra calcula-se o valor observado da estatística de teste,U^*;
5. cálculo de p_v, de acordo com a simetria ou não da distribuição amostral e com o tipo de teste de hipóteses, e com base no valor de p_v, aplica-se a regra de decisão:

$$\begin{cases} \text{se } p_v \leq \alpha \Rightarrow & \text{rejeitar } H_0; \\ \text{se } p_v > \alpha \Rightarrow & \text{não rejeitar } H_0. \end{cases}$$

Para o cálculo de p_v é necessário ter em conta, como já referimos, H_1. Numa **distribuição amostral simétrica** (Normal reduzida ou t de *Student*), supondo que temos uma estimativa U^*, de uma distribuição amostral U, podemos estar perante três situações distintas:

1. Se o teste é unilateral esquerdo,

$$p_v = P\left(U \leq U^*\right);$$

2. Se o teste é unilateral direito,

$$p_v = P\left(U \geq U^*\right);$$

3. Se o teste é bilateral,

$$\begin{aligned} &\text{se } \; U^* > 0 : p_v = P\left(U \leq -U^*\right) + P\left(U \geq U^*\right); \\ &\text{se } \; U^* < 0 : p_v = P\left(U \leq U^*\right) + P\left(U \geq -U^*\right). \end{aligned}$$

Exemplo 6.10. Retomando o exemplo 6.4 em que $Z^* = 1.319$ e o teste é unilateral direito, se pretendessemos tomar uma decisão com base em p_v faríamos o seguinte:

$$p_v = P\left(Z \geq Z^*\right) = P\left(Z \geq 1.319\right) = P\left(Z \geq 1.32\right) = 1 - \Phi\left(1.32\right) = 1 - 0.9066 = 0.0934.$$

Como o $p_v = 0.0934 > \alpha = 0.05$ se situa em RA (ver a Figura 6.16), não se rejeita H_0. Como o $p_v = 0.0934 > \alpha = 0.05$ não se rejeita H_0. Podemos ainda acrescentar que H_0 só se rejeitaria para valores de $\alpha \geq 0.0934$. ∎

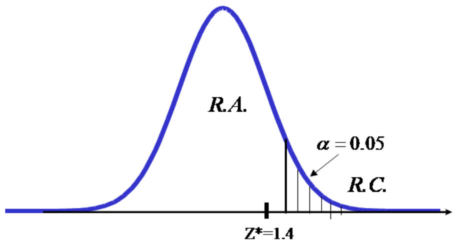

Figura 6.16: Teste Unilateral Direito.

Numa **distribuição amostral assimétrica** (Qui-Quadrado ou F de Snedcor), supondo que temos uma estimativa V^*, de uma distribuição amostral V, para o cálculo de p_v podemos estar novamente perante três situações distintas:

1. Se o teste é unilateral esquerdo,
$$p_v = P\left(V \leq V^*\right);$$
2. Se o teste é unilateral direito,
$$p_v = P\left(V \geq V^*\right);$$
3. Se o teste é bilateral,
$$p_v \text{ é igual ao dobro do menor dos dois valores anteriores.}$$

Exemplo 6.11. Retomando o exemplo 6.9 em que $F^* = 0.5625$ e o teste é unilateral esquerdo, se pretendessemos tomar uma decisão com base em p_v faríamos o seguinte:

$$p_v = P\left(F_{(15,20)} \leq F^*\right) = P\left(F_{(15,20)} \leq 0.5625\right) = 1 - P\left(F_{(20,15)} \leq \frac{1}{0.5625}\right) =$$
$$= 1 - P\left(F_{(20,15)} \leq 1.8\right) \approx 1 - 0.9 = 0.1.$$

Como o $p_v \approx 0.1 > \alpha = 0.01$ se situa em RA (ver a Figura 6.17), não se rejeita H_0.

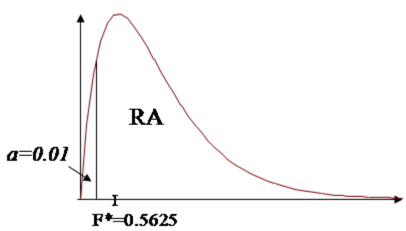

Figura 6.17: Teste Unilateral Esquerdo.

Podemos ainda acrescentar que H_0 só se rejeitaria para valores de $\alpha \geq 0.1$. ∎

Referências

[1] FISZ, M., *Probability Theory and Mathematical Statistics*, Jonh Wiley & Sons, Inc., New York, 1963.

[2] GUIMARÃES, R.C. e CABRAL, J.A.S., *Estatística*, McGraw-Hill de Portugal, Lisboa, 1997.

[3] MURTEIRA, B.J.F., *Probabilidades e Estatística*, McGraw-Hill de Portugal, Lisboa, 1979.

[4] SPIEGEL, M.R., *Probabilidade e Estatística*, Colecção Schaum, McGraw-Hill do Brasil, São Paulo, 1978

[5] OLIVEIRA, J.T., *Probabilidades e Estatística*, vol. I, Escolar Editora, Lisboa,1967.

[6] MELLO, F., *Introdução aos Métodos Estatísticos*, vol. I e II, Cadernos do Instituto de Orientação Profissional, Lisboa, 1973.

[7] MONTGOMERY, D.; RUNGER, G., *Applied Statistics and Probability for Engineers*, Jonh Wiley & Sons, Inc., New York, 2003.