ESTIMAÇÃO POR INTERVALO (INTERVALOS DE CONFIANÇA)

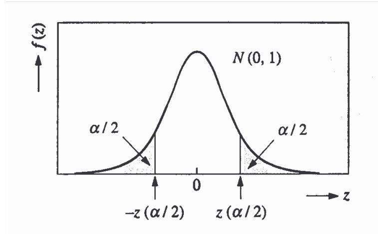
Cada um dos métodos de estimação pontual permite associar a cada parâmetro populacional um estimador. Ora a cada estimador estão associadas tantas estimativas diferentes quantas as amostras utilizadas para o seu cálculo. De um modo geral nenhuma destas estimativas irá coincidir com o valor do parâmetro da população e não é possível obter qualquer informação relativa ao seu rigor. Esta impossibilidade de associar a uma dada estimativa o respectivo grau de confiança, constitui a grande limitação dos métodos de estimação pontual. Este problema é ultrapassado recorrendo à estimação por intervalo.

Admita-se então que temos uma população $X \sim N(\mu, \sigma^2)$ e que é seleccionada uma amostra aleatória de dimensão n. Para essa amostra é calculada a respectiva média amostral cujo valor é $\bar{x}$. **O objectivo é definir um intervalo que com uma dada probabilidade 1 - α (p.ex: 95%, 99%), inclua o verdadeiro valor do parâmetro μ da população**.

Sabemos que:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

defina-se agora $z(\alpha/2)$ como o valor da v.a. Z que verifica $P[Z > z(\alpha/2)] = \alpha/2$. Então $P[Z < -z(\alpha/2)] = \alpha/2$ e portanto $P[-z(\alpha/2) < Z < z(\alpha/2)] = 1 - \alpha$.



Então:

$$P\left[-z(\alpha/2) < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z(\alpha/2)\right] = 1 - \alpha$$

que se pode escrever como:

$$P\left[\mu - z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

ou como:

$$P\left[\bar{X} - z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

De acordo com a expressão anterior o intervalo:

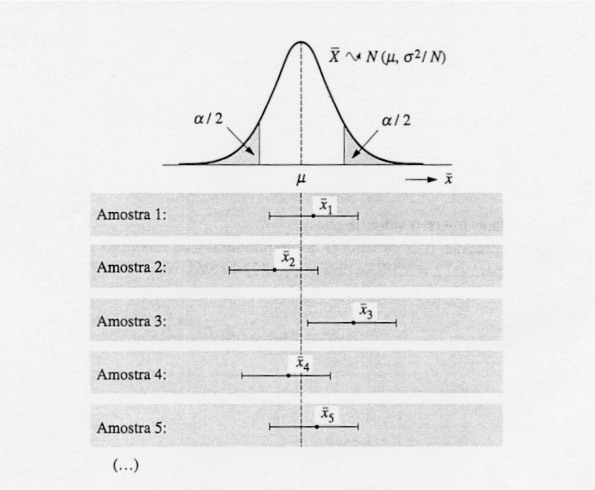
$$\left[\bar{X} - z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}} \; , \; \bar{X} + z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}\right]$$

incluirá o valor de μ com probabilidade 1 - α .

Este intervalo designa-se por **intervalo de confiança para o valor esperado a (1 - α).100%**. Os extremos deste intervalo são **os limites de confiança a (1 - α).100%**. O valor de $z(\alpha/2)\cdot\frac{\sigma}{\sqrt{n}}$, que representa a **semiamplitude do intervalo de confiança**, corresponde ao erro máximo que, com a confiança especificada, se pode cometer na estimativa de μ.

NOTA:

- O valor de α representa, em média, a proporção de vezes em que o intervalo de confiança não contém o parâmetro que se pretende estimar.



- Outro aspecto a salientar prende-se com a simetria do intervalo de confiança relativamente ao valor do estimador pontual $\overline{X}$.

Para quaisquer valores α_1 e α_2 não simétricos que satisfaçam:

$$\alpha_1 + \alpha_2 = \alpha$$

os intervalos

$$\left[\overline{X} - z(\alpha_1)\cdot\frac{\sigma}{\sqrt{n}} \;,\; \overline{X} + z(\alpha_2)\cdot\frac{\sigma}{\sqrt{n}} \right]$$

são todos eles intervalos de confiança de μ a (1-α).100% , porém com amplitudes diferentes.

Sempre que a estatística a partir da qual se definem os intervalos de confiança, apresentar uma distribuição unimodal simétrica, o intervalo simétrico em relação à estatística ($\alpha_1 = \alpha_2 = \alpha/2$) é o de menor amplitude e portanto aquele que deve ser calculado.

As excepções a esta regra são situações em que o objectivo é definir intervalos de confiança unilaterais (ilimitados superiormente ou ilimitados inferiormente).

ESPECIFICAÇÃO DE INTERVALOS DE CONFIANÇA

A especificação de um intervalo de confiança para um parâmetro implica conhecer:

- Um estimador do parâmetro em causa
- A distribuição desse estimador
- Uma estimativa pontual do parâmetro

INTERVALOS DE CONFIANÇA PARA O VALOR ESPERADO (μ)

I) Amostra de grande dimensão. População qualquer.

De acordo com o teorema do limite central temos que, neste caso:

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Em geral o desvio padrão da população, σ, é desconhecido, sendo estimado através do desvio padrão amostral, S:

$$S = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}$$

(S: estimador desvio padrão amostral; s: estimativas)

Uma vez que se admitiu que a amostra é de elevada dimensão, o erro de estimação é desprezável e podemos admitir que:

$$S \approx \sigma \qquad \text{(constante)}$$

e portanto:

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{n}} \approx \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Então o intervalo de confiança para o valor esperado μ a (1-α).100% é dado por:

$$\left[\overline{X} - z(\alpha/2) \cdot \frac{S}{\sqrt{n}} \ , \ \overline{X} + z(\alpha/2) \cdot \frac{S}{\sqrt{n}}\right]$$

II) Amostra de pequena dimensão. População Normal.

Neste caso já não é válido considerar que:

$$S \approx \sigma \qquad \text{(constante)}$$

e portanto também já não é válido admitir que:

$$\frac{\overline{X} - \mu}{S / \sqrt{n}} \approx \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Então, para definir o intervalo de confiança é necessário determinar a distribuição da v.a. :

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}$$

Notemos que:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{S/\sigma} \sim \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}}$$

e como $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ e S/σ são v.a. independentes, resulta da definição da distribuição t de Student que:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

sendo portanto o intervalo de confiança para o valor esperado μ a $(1-\alpha).100\%$ dado por:

$$\left[\overline{X} - t_{n-1}(\alpha/2)\cdot\frac{S}{\sqrt{n}}\ ,\ \overline{X} + t_{n-1}(\alpha/2)\cdot\frac{S}{\sqrt{n}}\right]$$

INTERVALOS DE CONFIANÇA PARA A PROPORÇÃO BINOMIAL ($P = \frac{Y}{n}$)

Vimos já anteriormente que $P = \frac{Y}{n}$ era um estimador para a proporção binomial p e que, sob determinadas condições, a distribuição de $P = \frac{Y}{n}$ é dada por:

$$P = \frac{Y}{n} \sim N\left(p, \frac{p\cdot(1-p)}{n}\right)$$

e portanto os limites do intervalo de confiança para $P = \frac{Y}{n}$ são dados por:

$$\frac{Y}{n} \pm z(\alpha/2)\cdot\sqrt{\frac{p\cdot(1-p)}{n}} = \frac{Y}{n} \pm z(\alpha/2)\cdot\sigma$$

Uma vez que o valor de σ depende do parâmetro desconhecido p, poderá para amostras de elevada dimensão, ser substituído por um qualquer valor do seu estimador $P = \frac{Y}{n}$ resultando em:

$$\sigma = \sqrt{\frac{Y/n\cdot(1-Y/n)}{n}} = \sqrt{\frac{Y\cdot(n-Y)}{n^3}}$$

e portanto:

$$\frac{\frac{Y}{n}-p}{\sqrt{\frac{Y\cdot(n-Y)}{n^3}}}\sim N(0,1)$$

sendo o intervalo de confiança para a proporção binomial p a (1-α).100% dado por:

$$\left[\frac{Y}{n}-z(\alpha/2)\cdot\sqrt{\frac{Y\cdot(n-Y)}{n^3}}\ ,\ \frac{Y}{n}+z(\alpha/2)\cdot\sqrt{\frac{Y\cdot(n-Y)}{n^3}}\right]$$

INTERVALOS DE CONFIANÇA PARA A VARIÂNCIA DE UMA POPULAÇÃO NORMAL (σ^2)

Vimos já que se de uma população Normal, $N(\mu,\sigma^2)$, forem seleccionadas amostras aleatórias de dimensão n com variância amostral S^2, então a v.a. :

$$(n-1)\cdot\frac{S^2}{\sigma^2}\sim\chi^2_{n-1}$$

Consideremos agora dois valores $\left(\chi^2_{n-1}\right)_A$ e $\left(\chi^2_{n-1}\right)_B$ tais que:

$$P\left[\left(\chi^2_{n-1}\right)_A<\chi^2_{n-1}<\left(\chi^2_{n-1}\right)_B\right]=1-\alpha$$

Substituindo na equação anterior $\left(\chi^2_{n-1}\right)$ por $(n-1)\cdot\frac{S^2}{\sigma^2}$ obtém-se:

$$P\left[\left(\chi^2_{n-1}\right)_A<(n-1)\cdot\frac{S^2}{\sigma^2}<\left(\chi^2_{n-1}\right)_B\right]=1-\alpha$$

ou:

$$P\left[\frac{1}{\left(\chi^2_{n-1}\right)_A}>\frac{\sigma^2}{(n-1)\cdot S^2}>\frac{1}{\left(\chi^2_{n-1}\right)_B}\right]=1-\alpha$$

a que podemos ainda dar outro aspecto:

$$P\left[\frac{(n-1)\cdot S^2}{\left(\chi^2_{n-1}\right)_A}>\sigma^2>\frac{(n-1)\cdot S^2}{\left(\chi^2_{n-1}\right)_B}\right]=1-\alpha$$

ou finalmente:

$$P\left[\frac{(n-1)\cdot S^2}{\left(\chi^2_{n-1}\right)_B}<\sigma^2<\frac{(n-1)\cdot S^2}{\left(\chi^2_{n-1}\right)_A}\right]=1-\alpha$$

O intervalo de confiança para a variância σ^2 a (1-α).100% é dado por:

$$\left[\frac{(n-1)\cdot S^2}{\left(\chi^2_{n-1}\right)_B} \, , \, \frac{(n-1)\cdot S^2}{\left(\chi^2_{n-1}\right)_A} \right]$$

Neste caso a distribuição não é simétrica existindo portanto a dificuldade de definir os valores $\left(\chi^2_{n-1}\right)_A$ e $\left(\chi^2_{n-1}\right)_B$ que conduzem ao intervalo de confiança de menor amplitude. Por razões de simplicidade é habitual escolher:

$$\left(\chi^2_{n-1}\right)_B = \chi^2_{n-1}(\alpha/2)$$

$$\left(\chi^2_{n-1}\right)_A = \chi^2_{n-1}(1-\alpha/2)$$

e assim a expressão final para o intervalo de confiança é:

$$\left[\frac{(n-1)\cdot S^2}{\chi^2_{n-1}(\alpha/2)} \, , \, \frac{(n-1)\cdot S^2}{\chi^2_{n-1}(1-\alpha/2)} \right]$$

INTERVALOS DE CONFIANÇA PARA A RAZÃO ENTRE VARIÂNCIAS DE POPULAÇÕES NORMAIS

Admita-se que σ^2_A e σ^2_B correspondem às variâncias de duas populações Normais A e B. Considere-se também que, com base em amostras independentes de dimensão n_A e n_B respectivamente, se obtêm os estimadores para aquelas variâncias, isto é S^2_A e S^2_B. Então:

$$(n_A - 1)\cdot \frac{S^2_A}{\sigma^2_A} \sim \chi^2_{n_A - 1}$$

e

$$(n_B - 1)\cdot \frac{S^2_B}{\sigma^2_B} \sim \chi^2_{n_B - 1}$$

resultando que:

$$\frac{S^2_A/\sigma^2_A}{S^2_B/\sigma^2_B} \sim \frac{\chi^2_{n_A-1}/(n_A-1)}{\chi^2_{n_B-1}/(n_B-1)}$$

Atendendo à definição da distribuição F temos então que:

$$\frac{S^2_A/\sigma^2_A}{S^2_B/\sigma^2_B} \sim F_{n_A-1,n_B-1}$$

uma vez que se admite que as variáveis S^2_A e S^2_B são independentes (pois são obtidas a partir de amostras independentes).

Considerem-se agora dois valores desta distribuição $F_{n_A-1,n_B-1}(\alpha/2)$ e $F_{n_A-1,n_B-1}(1-\alpha/2)$ tais que:

$$P\left[F_{n_A-1,n_B-1}(1-\alpha/2) < F_{n_A-1,n_B-1} < F_{n_A-1,n_B-1}(\alpha/2)\right] = 1-\alpha$$

e portanto:

$$P\left[F_{n_A-1,n_B-1}(1-\alpha/2) < \frac{S_A^2/\sigma_A^2}{S_B^2/\sigma_B^2} < F_{n_A-1,n_B-1}(\alpha/2)\right] = 1-\alpha$$

ou ainda:

$$P\left[\frac{1}{F_{n_A-1,n_B-1}(1-\alpha/2)} > \frac{\sigma_A^2/\sigma_B^2}{S_A^2/S_B^2} > \frac{1}{F_{n_A-1,n_B-1}(\alpha/2)}\right] = 1-\alpha$$

ou de outro modo:

$$P\left[\frac{1}{F_{n_A-1,n_B-1}(1-\alpha/2)} \cdot \frac{S_A^2}{S_B^2} > \frac{\sigma_A^2}{\sigma_B^2} > \frac{1}{F_{n_A-1,n_B-1}(\alpha/2)} \cdot \frac{S_A^2}{S_B^2}\right] = 1-\alpha$$

e finalmente:

$$P\left[\frac{1}{F_{n_A-1,n_B-1}(\alpha/2)} \cdot \frac{S_A^2}{S_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{n_A-1,n_B-1}(1-\alpha/2)} \cdot \frac{S_A^2}{S_B^2}\right] = 1-\alpha$$

O intervalo de confiança a (1-α).100% para a razão entre as variâncias das duas populações normais σ_A^2/σ_B^2 é então:

$$\left[\frac{1}{F_{n_A-1,n_B-1}(\alpha/2)} \cdot \frac{S_A^2}{S_B^2} \; , \; \frac{1}{F_{n_A-1,n_B-1}(1-\alpha/2)} \cdot \frac{S_A^2}{S_B^2}\right]$$

INTERVALOS DE CONFIANÇA PARA A DIFERENÇA ENTRE OS VALORES ESPERADOS DE DUAS POPULAÇÕES $(\mu_A - \mu_B)$

I) Amostras independentes de grandes dimensões, populações quaisquer

Sejam μ_A e μ_B os valores esperados das populações A e B e σ_A^2 e σ_B^2 as suas variâncias. Considere que a partir destas populações se obtêm amostras independentes de dimensão N_A e N_B com base nas quais se determinam os estimadores dos valores esperados, $\overline{X}_A$ e $\overline{X}_B$, e das variâncias, S_A^2 e S_B^2.

Uma vez que estamos a tratar com amostras de elevada dimensão, podemos considerar que:

$$S_A^2 \approx \sigma_A^2 \quad \text{e} \quad S_B^2 \approx \sigma_B^2$$

por outro lado, o teorema do limite central permite-nos afirmar que, quaisquer que sejam as formas das distribuições de A e B teremos:

$$\overline{X}_A \sim N\left(\mu_A, \frac{\sigma_A^2}{n_A}\right) \approx N\left(\mu_A, \frac{S_A^2}{n_A}\right)$$

e

$$\overline{X}_B \sim N\left(\mu_B, \frac{\sigma_B^2}{n_B}\right) \approx N\left(\mu_B, \frac{S_B^2}{n_B}\right)$$

Uma vez que se admitiu que as amostras são independentes, a diferença $\overline{X}_A - \overline{X}_B$ é a também uma v.a. com distribuição Normal e portanto:

$$\overline{X}_A - \overline{X}_B \sim N\left(\mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right) \approx N\left(\mu_A - \mu_B, \frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)$$

isto é:

$$Z = \frac{\left(\overline{X}_A - \overline{X}_B\right) - \left(\mu_A - \mu_B\right)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim N(0,1)$$

Então o intervalo de confiança a (1-α).100% para a diferença dos valores esperados $\mu_A - \mu_B$ é dado por:

$$\left[\left(\overline{X}_A - \overline{X}_B\right) - z(\alpha/2) \cdot \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}\ ,\ \left(\overline{X}_A - \overline{X}_B\right) + z(\alpha/2) \cdot \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}\right]$$

Se se admitir que as variâncias das duas populações são iguais:

$$\sigma_A^2 = \sigma_B^2 = \sigma^2$$

então:

$$\overline{X}_A - \overline{X}_B \sim N\left(\mu_A - \mu_B, \left(\frac{\sigma^2}{n_A} + \frac{\sigma^2}{n_B}\right)\right) = N\left(\mu_A - \mu_B, \sigma^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)\right)$$

neste caso é possível refinar a expressão obtida para o intervalo de confiança, estimando a variância comum σ^2, das duas populações A e B, a partir de:

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

e substituindo nessa expressão S_A^2 e S_B^2 por S^2. Então se as variâncias das populações forem iguais a expressão para o intervalo de confiança é:

$$\left[\left(\overline{X}_A - \overline{X}_B\right) - z(\alpha/2) \cdot S \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\ ,\left(\overline{X}_A - \overline{X}_B\right) + z(\alpha/2) \cdot S \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\right]$$

I) Amostras independentes de pequenas dimensões, populações quaisquer

Uma vez que agora **já não é válido** considerar:

$$S_A^2 \approx \sigma_A^2 \quad \text{e} \quad S_B^2 \approx \sigma_B^2$$

também deixa de ser válido admitir que tem distribuição N(0,1) a v.a.:

$$\frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}}$$

Seguindo um procedimento análogo ao já utilizado no caso de se trabalhar apenas com uma amostra, temos que:

$$\frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}} = \frac{\dfrac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}}{\dfrac{\sqrt{S_A^2/n_A + S_B^2/n_B}}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}}$$

$$= \frac{N(0,1)}{\sqrt{\chi_{gl}^2/gl}} \sim t_{gl}$$

isto é, aquela variável segue uma distribuição t de Student com gl graus de liberdade.

Para definir o valor de gl temos duas situações possíveis, que correspondem a podermos ou não admitir como válido que as variâncias das duas populações são iguais:

$$gl = n_A + n_B - 2 \qquad \text{se} \quad \sigma_A^2 = \sigma_B^2 = \sigma^2$$

$$gl = \frac{\left(\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}\right)^2}{\dfrac{(S_A^2/n_A)^2}{n_A - 1} + \dfrac{(S_B^2/n_B)^2}{n_B - 1}} \qquad \text{se} \quad \sigma_A^2 \neq \sigma_B^2$$

No primeiro caso o número de graus de liberdade corresponde ao número de graus de liberdade com que a variância comum das duas populações é estimada.

No segundo caso se o valor de gl não der um inteiro, deve-se utilizar o inteiro imediatamente inferior já que conduz à definição de um intervalo com uma confiança maior do que a especificada inicialmente.

Se as variâncias das populações forem iguais podemos também aqui estimar a variância comum pela fórmula usada anteriormente, isto é:

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

Então o intervalo de confiança a (1-α).100% para a diferença dos valores esperados das duas populações, $\mu_A - \mu_B$, é dado por:

$\underline{\sigma_A^2 = \sigma_B^2 = \sigma^2}$

$$(\mu_A - \mu_B) \in \left[\left(\overline{X}_A - \overline{X}_B\right) \pm t\left(\alpha/2\right) \cdot S \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right]$$

$\underline{\sigma_A^2 \neq \sigma_B^2}$

$$(\mu_A - \mu_B) \in \left[\left(\overline{X}_A - \overline{X}_B\right) \pm t\left(\alpha/2\right) \cdot \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \right]$$

INTERVALOS DE CONFIANÇA PARA A DIFERENÇA ENTRE PROPORÇÕES BINOMIAIS $p_A - p_B$ (AMOSTRAS INDEPENDENTES DE GRANDES DIMENSÕES)

Sejam duas populações A e B constituídas por elementos de dois tipos. Seja p_A a proporção de elementos de um dos dois tipos na população A e p_B o valor correspondente para a população B. Seleccionadas independentemente duas

amostras, seja $\frac{Y_A}{n_A}$ um estimador de p_A baseado numa amostra de dimensão n_A e $\frac{Y_B}{n_B}$ o estimador de p_B baseado numa amostra de dimensão n_B.

Estando satisfeitas as condições para aproximarmos as distribuições de $\frac{Y_A}{n_A}$ e $\frac{Y_B}{n_B}$ por distribuições Normais (populações infinitas ou amostragem com reposição verificando-se ainda que n ≥ 20 e n.p > 7 ; no caso de amostragem sem reposição é também necessário garantir que a dimensão da população é grande face à dimensão da amostra) e uma vez que as amostras são independentes temos que:

$$\frac{Y_A}{n_A} - \frac{Y_B}{n_B} \sim N\left(\mu_A - \mu_B, \frac{p_A \cdot (1 - p_A)}{n_A} + \frac{p_B \cdot (1 - p_B)}{n_B} \right)$$

Então seguindo um procedimento idêntico ao utilizado anteriormente temos que o intervalo de confiança a (1-α).100% para a diferença entre as proporções binomiais, $p_A - p_B$, é dado por:

$$(p_A - p_B) \in$$

$$\in \left(\frac{Y_A}{n_A} - \frac{Y_B}{n_B} \right) \pm z(\alpha/2) \cdot \sqrt{\frac{Y_A \cdot (n_A - Y_A)}{n_A^3} + \frac{Y_B \cdot (n_B - Y_B)}{n_B^3}}$$

DIMENSIONAMENTO DE AMOSTRAS

Até agora admitimos que a dimensão das amostras utilizadas para o cálculo das estimativas pontuais estava já especificada previamente.
Contudo o problema de dimensionamento das amostras é muito importante já que:

- Se a amostra for excessivamente grande face aos objectivos que se pretendem atingir, estaremos a desperdiçar recursos na recolha e tratamento da informação.

- Se a dimensão da amostra não for suficiente para a partir dela se extraírem conclusões válidas, estaremos a cometer um erro.

A dimensão das amostras a considerar aumentará à medida que aumentem os seguintes "parâmetros" (isoladamente ou em simultâneo):

i) a precisão do intervalo de confiança (que varia na razão inversa da respectiva amplitude).

ii) o grau de confiança do intervalo, isto é, a probabilidade de este vir a incluir o verdadeiro valor do parâmetro populacional.

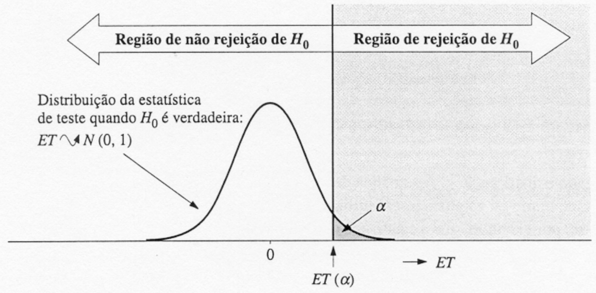


Figura 11.1. Região de rejeição de H_0 (teste unilateral à direita)

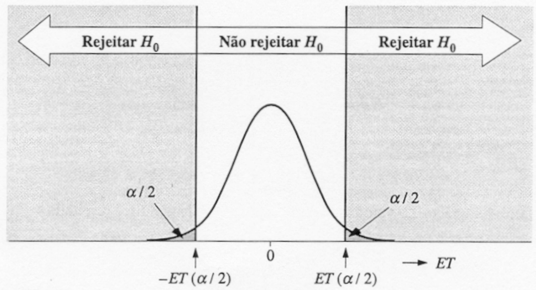


Figura 11.2. Região de rejeição de H_0 para um teste bilateral