UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESPECIALIZAÇÃO EM ESTATÍSTICA

Bioestatística

Dimensionamento de amostras

Professora: Arminda Lúcia Siqueira

Alunos: Zilma Reis, Ricardo Veiga e Augusto Filho

Belo Horizonte - MG

Cara monitora, a numeração dos exercícios não correspondem a da lista.

Exercício 1. *Um departamento local de saúde deseja estimar a prevalência de tuberculose entre crianças com menos de cinco anos de idade em sua localidade. Quantas crianças deveriam ser incluídas na amostra para que a prevalência possa ser estimada com precisão de 5,3 ou* 1% *do verdadeiro valor e com* 95% *de confiança, se a verdadeira taxa é improvável exceder* 20%*?*

Resolução*:*

Como há o interesse em estudar a proporção de elementos na população que possuem determinada característica, devemos fazer uso da seguinte equação:

$$n = \frac{z^2_{1-\alpha/2} P\,(1-P)}{d^2}$$

Ou de forma equivalente, mas usando a precisão relativa $\varepsilon = \frac{|\hat{\theta}-\theta|}{\theta}$.

$$n = z^2_{1-\alpha/2} \frac{(1-P)}{\varepsilon^2 P}$$

No enunciado é possível obtermos os seguintes valores de interesse: $d = 5,3\%$ *e* $d = 1\%$*, com um nível de confiança de* 95% *e uma taxa de* $P = 20\%$.

Portanto, precisamos encontrar o percentil de ordem 95% *em uma tabela de distribuição normal.*

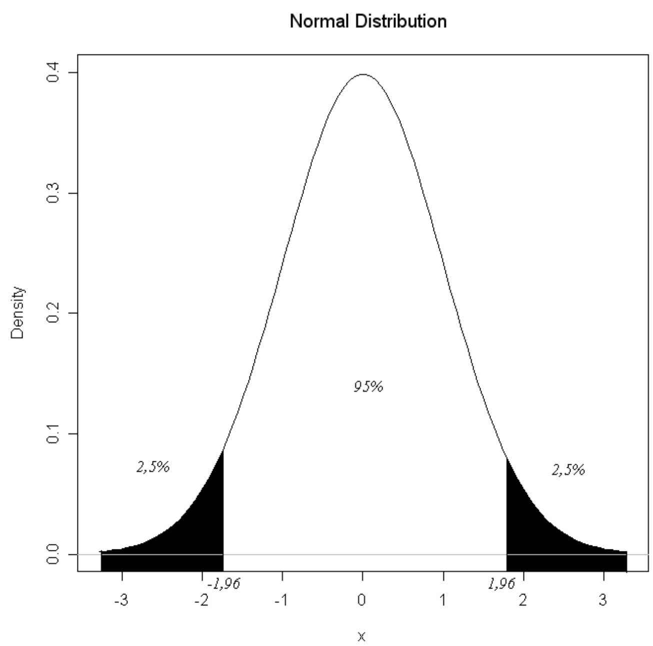

Figura 1: Percentil da distribuição normal

Logo $z_{1-\alpha/2} = 1,96$. *Para* $d = 5\%$ *temos o seguinte cálculo para o tamanho da amostra:*

$$\begin{aligned} n &= \frac{z^2_{1-\alpha/2} P\,(1-P)}{d^2} \\ n &= \frac{(1,96)^2\,0,20 \cdot (1-0,80)}{(0,05)^2} \\ n &= 245,9 \end{aligned}$$

Portanto, para $d = 5\%$ *o número de crianças incluídas na amostra será de 256.*

Para $d = 3\%$ *temos o seguinte cálculo para o tamanho da amostra:*

$$\begin{aligned} n &= \frac{z^2_{1-\alpha/2} P\,(1-P)}{d^2} \\ n &= \frac{(1,96)^2\,0,20 \cdot (1-0,80)}{(0,03)^2} \\ n &= 683 \end{aligned}$$

Portanto, para $d = 3\%$ *o número de crianças incluídas na amostra será de 683.*

O mesmo cálculo acima para $d = 1\%$, *nos levará ao seguinte resultado:*

$$\begin{aligned} n &= \frac{z^2_{1-\alpha/2} P\,(1-P)}{d^2} \\ n &= \frac{(1,96)^2\,0,20 \cdot (1-0,80)}{(0,01)^2} \\ n &= 6146,6 \end{aligned}$$

Portanto, para $d = 1\%$ *o número de crianças que deverá ser incluída na amostra é de 6.147 crianças.*

A seguir, veremos o cálculo do tamanho da amostra escrito no R. Tomou-se o cuidado de não realizar nenhuma programação e somente utilizou-se o R como planilha de cálculos:

Para $d = 5\%$, *temos:*

```
> z = 1.96
> d = 0.05
> p = 0.2
> n = (z^2 * p * (1 - p))/d^2
```

O valor de n para d=5% é 246

Para $d = 3\%$, temos:

```
> d = 0.03
> n = (z^2 * p * (1 - p))/d^2
```

O valor de n para d=3% é 683

Para $d = 1\%$, temos:

```
> d = 0.01
> n = (z^2 * p * (1 - p))/d^2
```

O valor de n para d=1% é 6147

A seguir, veremos os valores de n reunidos em uma tabela para melhor visualização:

Tabela 1: Tamanho da amostra

d	tamanho da amostra
5%	246
3%	683
1%	6147

Exercício 2. *Um estudo tem como objetivo estimar a média da glicemia em jejum de uma certa população. a partir do estudo piloto, encontrou-se* $\overline{X} = 87$ mg/dL *e* $s = 13$. *Qual deverá ser o tamanho de amostra* (n) *necessário para estimar esta média com* 95% *de confiança e precisão de* 5% 3% *e* 1 mg/dL*?*

Resolução*:*

Para encontrarmos o tamanho da amostra n necessário para estimar a média com um nível de confiança de 95% *e uma precisão de 5,3 e 1 mg/dL, devemos utilizar a seguinte fórmula:*

$$n = z^2_{1-\alpha/2} \cdot \frac{\sigma^2}{d^2}$$

Ou de forma equivalente, mas usando a precisão relativa $\varepsilon = \frac{|\hat{\theta}-\theta|}{\theta}$.

$$n = z^2_{1-\alpha/2} \frac{\sigma^2}{\varepsilon^2 \mu^2}$$

No enunciado é possível obtermos os seguintes valores de interesse: $d = 5,3$*;* $d = 1$, *com um nível de confiança de* 95%*;* $\overline{x} = 87$ mg/dL *e* $s = 13$ mg/dL.

Portanto, precisamos encontrar o percentil de ordem 95% *em uma tabela de distribuição normal.*

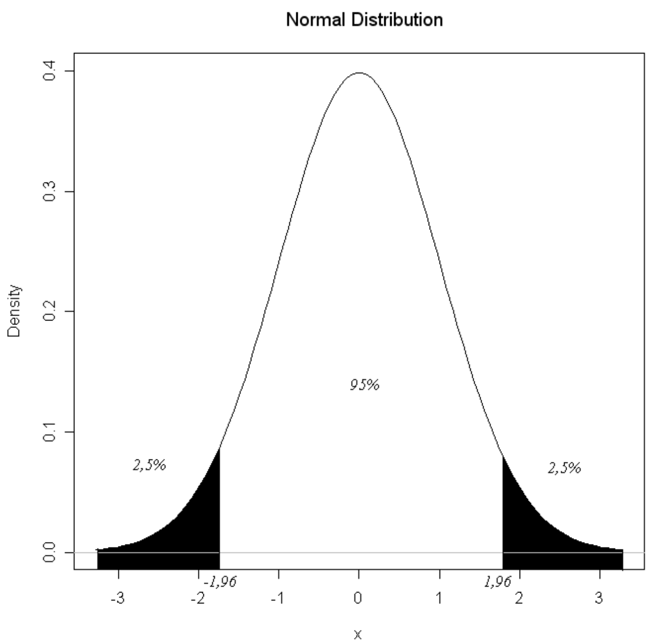

Figura 2: Percentil da distribuição normal

Logo $z_{1-\alpha/2} = 1,96$. *Quando* $d = 5$ *o cálculo para o tamanho da amostra é:*

$$\begin{aligned} n &= z_{1-\alpha/2}^2 \cdot \frac{\sigma^2}{d^2} \\ n &= (1,96)^2 \frac{(13)^2}{(5)^2} \\ n &= 25,97 \end{aligned}$$

Portanto, para $d = 5$ *o tamanho da amostra necessário para estimar a média é 26.*

Para $d = 3$ *o cálculo para o tamanho da amostra é:*

$$\begin{aligned} n &= z_{1-\alpha/2}^2 \cdot \frac{\sigma^2}{d^2} \\ n &= (1,96)^2 \frac{(13)^2}{(3)^2} \\ n &= 72,14 \end{aligned}$$

Portanto, para $d = 3$ *o tamanho da amostra necessário para estimar a média é 73.*

O mesmo cálculo acima para $d = 1$, *nos levará ao seguinte resultado:*

$$\begin{aligned} n &= z_{1-\alpha/2}^2 \cdot \frac{\sigma^2}{d^2} \\ n &= (1,96)^2 \frac{(13)^2}{(1)^2} \\ n &= 649,2 \end{aligned}$$

Desta forma, para $d = 1$ *o tamanho da amostra necessário para estimar a média é de 650.*

A seguir, veremos os valores de n reunidos em uma tabela:

Tabela 2: Tamanho da amostra

d	tamanho da amostra
5	26
3	73
1	650

Exercício 3. *Uma pesquisa mostrou que o peso médio de homens acima dos 55 anos de idade com recente diagnóstico de doença do coração era de 90 kg. entretanto, suspeita-se que agora o peso médio de tais homens é um poco menor. Qual tamanho de amostra seria necessário para testar, ao nível de* 5% *de significância e poder de 80,90 ou* 99%, *se este peso médio permanece o mesmo ou se ele teria diminuído de 90 para 85 kg com um desvio padrão estimado de 20 kg?*

Resolução:

Temos: $\mu_1 = 90$; $\mu_2 = 85$; $\alpha = 5\%$ *com* $1 - \beta = 80\%;\ 90\%$ *e* 99%.

Logo, o percentil de ordem 95% *é encontrado a seguir:*

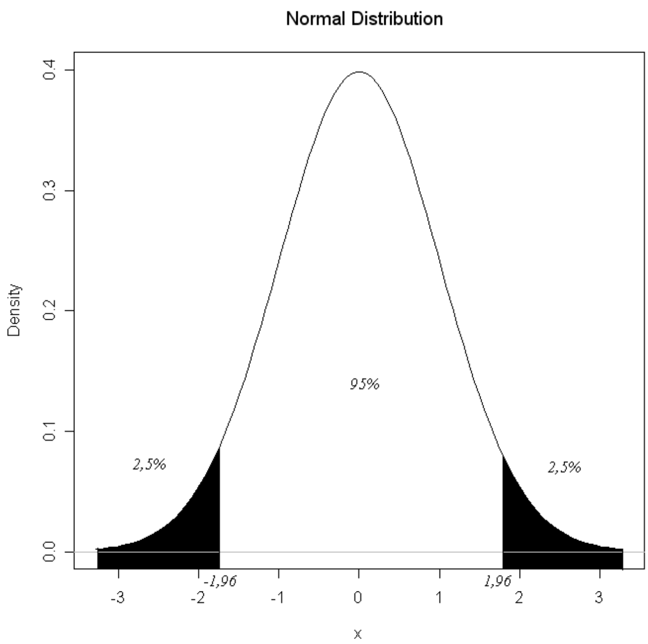

Figura 3: Percentil da distribuição normal

Para um poder de 80%, *temos o seguinte valor tabelado:*

$$z_{1-\beta} = 0,84$$

Para um poder de 90%, *temos o seguinte valor tabelado:*

$$z_{1-\beta} = 1,28$$

Para um poder de 99%, *temos o seguinte valor tabelado:*

$$z_{1-\beta} = 2,33$$

Logo para $z_{1-\beta} = 0,84$, temos o seguinte valor para o tamanho da amostra:

$$n = \frac{\sigma^2 \left[Z_{1-\alpha/2} + Z_{1-\beta}\right]^2}{(\mu - \mu_0)^2}$$

Temos:

$$\begin{aligned} n &= \frac{(20)^2 \left[1,96 + 0,84\right]^2}{(90 - 85)^2} \\ n &= 125,4 \end{aligned}$$

Logo, o tamanho da amostra para um poder de 80% *é 126.*

Para um poder de 90%*, temos:*

$$\begin{aligned} n &= \frac{(20)^2 \left[1,96 + 1,28\right]^2}{(90 - 85)^2} \\ n &= 168 \end{aligned}$$

Desta forma, o tamanho da amostra para um poder de 90% *é de 168.*

Para um poder de 99%*, temos:*

$$\begin{aligned} n &= \frac{(20)^2 \left[1,96 + 2,33\right]^2}{(90 - 85)^2} \\ n &= 294,5 \end{aligned}$$

Desta forma, o tamanho da amostra para um poder de 99% *é de 295.*

A seguir, temos os valores resumidos em uma tabela:

Tabela 3: Tamanho da amostra

n	$z_{1-\alpha/2}$	$z_{1-\beta}$	μ_0	μ_1	σ
126	1,96	0,84	90	85	20
168	1,96	1,28	90	85	20
295	1,96	2,33	90	85	20

Computacionalmente, temos:

```
> mu1 = 90
> mu2 = 85
> sigma = 20
> beta1 = 0.84
> beta2 = 1.28
> beta3 = 2.33
```

```
> alpha = 1.96
> n1 = (sigma^2 * (alpha + beta1)^2)/(mu1 - mu2)^2
> n1

[1] 126

> n2 = (sigma^2 * (alpha + beta2)^2)/(mu1 - mu2)^2
> n2

[1] 168

> n3 = (sigma^2 * (alpha + beta3)^2)/(mu1 - mu2)^2
> n3

[1] 295
```

Exercício 4. *Suponha que se queira determinar a prevalência de obesidade em duas populações diferentes definidas pela condição sócio-econômica. De um estudo piloto obtiveram-se as seguintes estimativas.* $\hat{P}_1 = 0,40$ *e* $\hat{P}_2 = 0,28$. *Considerando-se os dois grupos de mesmo tamanho, nível de significância de* 5% *e precisão de* 10%, 5% *e* 1%, *determinar o tamanho dos grupos.*

Resolução*:*

Temos P_1 *e* P_2 *as proporções de unidades na população 1 e 2, respectivamente, que possuem a característica de interesse e* $\hat{P}_1$ *e* $\hat{P}_2$ *os seus respectivos estimadores.*

Desta forma, temos:

$$n = \frac{z^2_{1-\alpha/2}\left[\hat{P}_1\left(1-\hat{P}_1\right)+\hat{P}_2\left(1-\hat{P}_2\right)\right]}{d^2} \tag{1}$$

Logo, encontraremos o tamanho da amostra para $\hat{P}_1 = 0,40$*;* $\hat{P}_2 = 0,28$*;* $\alpha = 0,05$*; e* $d = 10\%, 5\%$ *e* 1%.

- Primeiramente encontraremos o valor de $z^2_{1-\alpha/2}$.

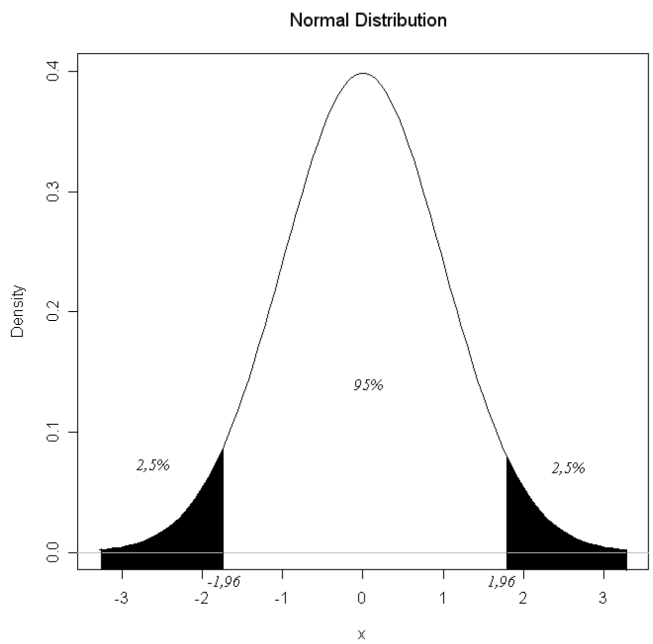

Figura 4: Valor tabelado

Portanto para o percentil $z^2_{1-\alpha/2}$, temos 1,96. Utilizando a fórmula (1), temos:

- Calcularemos n, para $d = 10\%$:

$$\begin{aligned} n &= \frac{(1,96)^2\,[0,40\,(1-0,40)+0,28\,(1-0,28)]}{(0,10)^2} \\ n &= \frac{1,696}{0,010} \\ n &= 169,6 \end{aligned}$$

  Desta forma, o tamanho do primeiro grupo com um desvio absoluto de 10% é igual a 170.

- Para $d = 5\%$, o tamanho da amostra será de:

$$\begin{aligned} n &= \frac{(1,96)^2\,[0,40\,(1-0,40)+0,28\,(1-0,28)]}{(0,05)^2} \\ n &= \frac{1,696}{0,0025} \\ n &= 678,6 \end{aligned}$$

  Desta forma, o tamanho do segundo grupo com um desvio absoluto de 5% é igual a 679.

- Para $d = 1\%$, o tamanho da amostra será de:

$$\begin{aligned} n &= \frac{(1,96)^2\,[0,40\,(1-0,40)+0,28\,(1-0,28)]}{(0,010)^2} \\ n &= \frac{1,696}{0,0001} \\ n &= 16964,5 \end{aligned}$$

  Desta forma, o tamanho do terceiro grupo com um desvio absoluto de 1% é igual a 16965.

  A seguir, veremos o mesmo procedimento através de forma computacional. Utilizou-se o R como planilha de cálculo. Usou-se o Sweave para transpor os resultados do R para LaTeX.

Aqui, a única preocupação foi a utilização do software como planilha de cálculo e não houve a preocupação de implementar um programa. Desta forma, temos:

```
> z = 1.96
> p1chapel = 0.4
> p2chapel = 0.28
> d = 0.10
> n = z^2 * (p1chapel * (1 - p1chapel) + p2chapel * (1 - p2chapel))/(d^2)
> n

[1] 169.6451 # o valor de n para d = 10%

> d = 0.05
> n = z^2 * (p1chapel * (1 - p1chapel) + p2chapel * (1 - p2chapel))/(d^2)
> n

[1] 678.5802 # o valor de n para d = 5%

> d = 0.01
> n = z^2 * (p1chapel * (1 - p1chapel) + p2chapel * (1 - p2chapel))/(d^2)
> n

[1] 16964.51 # o valor de n para d = 1%
```

Lembrando que no cálculo do tamanho da amostra, deve-se trabalhar com o próximo número inteiro.

Os valores obtidos para os diferentes d com $\hat{P}_1 = 0,40$; $\hat{P}_2 = 0,28$ e $\alpha = 0,05$ são mostrados a seguir:

Tabela 4: Tamanho da amostra

d	n
1%	16.965
5%	679
10%	170