Inferência Estatística

Amostragem

Prof. Antonio Estanislau Sanches
2017

Inferência estatística

é o conjunto de metodologia que apoiam na formulação de conclusões sobre as características de uma POPULAÇÃO a partir de uma parte dessa população, denominada de AMOSTRA.

População ou Universo

É a coleção de unidades individuais com uma ou mais características comuns, que se pretende estudar

<u>Exemplos</u>

- Alunos de uma escola;
- Crianças de 0 à 5 anos de um orfanato;
- Agregados familiar de uma província;
- Carteiras dentro do campus da N Lins;
- Automóveis da cidade de Manaus.

Introdução à Amostragem

Se uma população for muito grande, para estudá-la demandará muito trabalho e geralmente com resultados falhos.

Recorre–se então, a uma AMOSTRA, sendo esta uma PARTE representativa da POPULAÇÃO em dimensões reduzidas, porém, com as mesmas características.

Exemplo 1. Tendo uma escola 400 alunos, podemos colher uma amostra de 40 alunos para estudar o comportamento da variável ALTURA, apenas nesses alunos.

Exemplo 2. O censo mostra a existência de 15 mil agregados familiares na província de Manica. Objetivando analisar a variável “número médio dos agregados por família”, podemos estudar o comportamento dessa variável em 601 agregados.

Uma AMOSTRA tem que ser:

Representativa →conter em proporção tudo o que a população possui qualitativa e quantitativamente;

Imparcial →todos os elementos da população tem igual oportunidade de fazer parte da amostra;

Uma AMOSTRA é a redução de uma população em dimensões menores, porém, sem perda de suas características.

O processo de definição da amostra chama-se AMOSTRAGEM

Amostragem Probabilistica e Não Probabilistica

Métodos Probablisticos (Aleatórios)

Todos os elementos da população tem uma probabilidade conhecida, diferente de zero, de pertencer à amostra. Desta forma, a amostragem probabilística implica um sorteio com regras bem determinadas.

Métodos Não Probablisticos (Não Aleatórios)

Quando não é possível designar uma probabilidade para cada elemento da população, dizemos que a amostragem é não probabilística.

Amostragens Probabilísticas:	➢ Aleatória Simples ➢ Estratificada ➢ Por Clusters ➢ Multi-Etapas

Estatística Descritiva

Introdução a amostragem

População ou universo – Coleção de unidades individuais com uma ou mais características comuns, que se pretendem estudar.

Amostra – redução representativa da População a dimensões menores, porém sem perda da característica.

- ✓ **Representativa** – conter em proporção tudo o que a população possui qualitativa e quantitativamente;
- ✓ **Imparcial** – todos os elementos da população tem igual oportunidade de fazer parte da amostra.

Estatística Descritiva

Introdução a amostragem

Erro amostral – é a diferença entre um resultado amostral e o verdadeiro resultado populacional; tais erros resultam de flutuações amostrais aleatórias.

- ✓ **Quanto maior o tamanho da amostra, menor o erro amostral.**
- ✓ **Amostras desnecessariamente grandes acarretam desperdício de tempo e de dinheiro;**
- ✓ **Amostras excessivamente pequenas podem levar a resultados não confiáveis.**

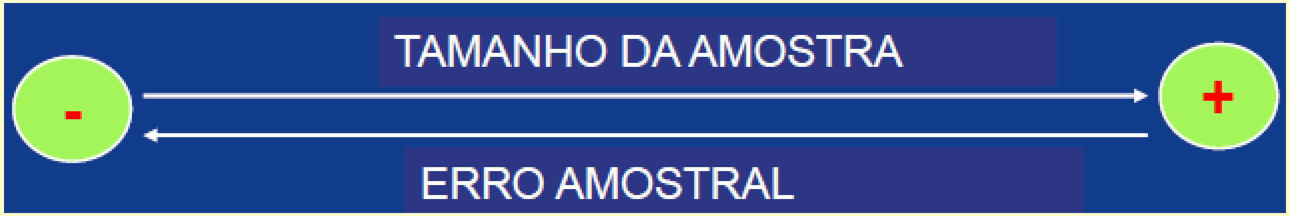
ERRO AMOSTRAL

Não há dúvida de que uma amostra não representa perfeitamente uma população. Ou seja, a utilização de uma amostra implica na aceitação de uma margem de erro que se denomina **ERRO AMOSTRAL**.

Erro Amostral é a diferença entre um resultado amostral e o verdadeiro resultado populacional; tais erros resultam de flutuações amostrais aleatórias.

Não podemos evitar a ocorrência do ERRO AMOSTRAL, porém podemos limitar seu valor através da escolha de uma amostra de tamanho adequado.

Obviamente, o **ERRO AMOSTRAL** e o **TAMANHO da AMOSTRA** seguem sentidos contrários. Quanto maior o tamanho da amostra, menor o erro cometido e vice-versa.



ERRO AMOSTRAL

Considere uma **POPULAÇÃO** de média **μ** e uma **AMOSTRA** dessa população, cuja média seja representada por **Ẍ**;

O **ERRO AMOSTRAL**, será medido pela diferença entre:

$$\mathbf{E = \ddot{X} - \mu \quad sendo \ E = Z^{*}\sigma/\sqrt{n}}$$

*e **Z** obtido da tabela de distribuição normal ou pela função*

INV.NORMP(α/2)**, quando α representa o **grau de confiança

EXEMPLO: O CREA/AM deseja pesquisar o salário médio dos engenheiros de Manaus. Já o CONFEA informa que o **desvio padrão** do salário da categoria em âmbito nacional é de R$ 6.250,00 ; entrevistados 600 engenheiros, obteve-se uma média salarial de R$ 9.600,00 Calcular o erro amostral, para um grau de confiança de 95% sendo a população infinita.

$\sigma = 6250$; $\ddot{X} = 9600$; $n = 600$; $\alpha = 0,05$ e $\alpha/2 = 0,025$ calcular **E = ?**

ERRO AMOSTRAL - Solução

$\sigma = 6250$; $\ddot{X} = 9600$; $n = 600$; $\alpha = 0{,}05$ e $\alpha/2 = 0{,}025$ calcular **E = ?**

Z = INV.NORMP.N(1 – 0,025) ou Z = 1,96

$E = Z^*\sigma/\sqrt{n}$ ⇔ $E = (1{,}96 * 6250) / \sqrt{600}$ ⇔ E = 500

INTERVALO DE CONFIANÇA:

$P[(\ddot{X}-E)<\mu<(\ddot{X}+E)]$ ou $P[(9.600 – 500) < \mu < (9.600 + 500)]$ ou

IC: $P[9.100 < \mu < 10.100]$; c/ 95% de certeza

INTERVALO DE CONFIANÇA: *Para a média populacional* μ *estimada pela média amostral* $\ddot{X}$ *é um intervalo de valores, acima e abaixo dessa média, dentro do qual se acredita que possa estar a verdadeira média populacional* μ*, com um grau de confiança de tantos pontos percentuais, no caso do exemplo: 95%, determinado por* **Z**.

A determinação do tamanho de uma amostra é problema de grande importância:

- Amostras muito grandes acarretam desperdício de tempo e de dinheiro;

- Amostras muito pequenas podem levar a resultados não confiáveis.

Fórmula para cálculo do tamanho da amostra

- N = Tamanho da população
- E_0 = erro amostral tolerável
- n_0 = primeira aproximação do tamanho da amostra

$$\boxed{n_0 = \frac{1}{E_0^2}}$$

- n = tamanho da amostra

$$\boxed{n = \frac{N.n_0}{N + n_0}}$$

Estatística Descritiva

Introdução a amostragem

Tamanho da amostra para populações muito grandes.

Fórmula

$$n = \frac{N \cdot n_0}{N + n_0} \quad \Rightarrow \quad n_0 = \frac{1}{E_0^2}$$

- ✓ n – tamanho da amostra;
- ✓ N – tamanho da população;
- ✓ n_o – primeira aproximação para o tamanho da amostra;
- ✓ E_o – erro amostral tolerável.

Exemplo cálculo do tamanho da amostra

N = 200 famílias

E_0 = erro amostral tolerável = 4% (E_0 = 0,04)

$$n_0 = \frac{1}{E_0^2}$$

$n_0 = 1/(0{,}04)^2$ = 625 famílias

n (tamanho da amostra corrigido) =

n = 200x625/200+625 = 125000/825 = 152 famílias

$$n = \frac{N.n_0}{N + n_0}$$

E se a população fosse de 200.000 famílias?

n = (200.000)x625/(200.000 +625) = 623 famílias

Observe=se que se N é muito grande, não é necessário considerar o tamanho exato N da população. Nesse caso, o cálculo da primeira aproximação já é suficiente para o cálculo.

$$\boxed{n = n_0 = \frac{1}{E_0^2}}$$

Estatística Descritiva

Introdução a amostragem

Tamanho da amostra para populações muito grandes.

Qual o tamanho da amostra representativa para uma população de 200 indivíduos considerando uma margem de erro de 4% ?

$$n_0 = \frac{1}{0{,}04^2} = 625 \quad \Rightarrow \quad n = \frac{200 \cdot 625}{200 + 625} = 151{,}5 \cong 152$$

76%

e se a população fosse de 200.000 indivíduos?

$$n = \frac{200000 \cdot 625}{200000 + 625} = 623$$

0,3%

Tamanho da amostra ...

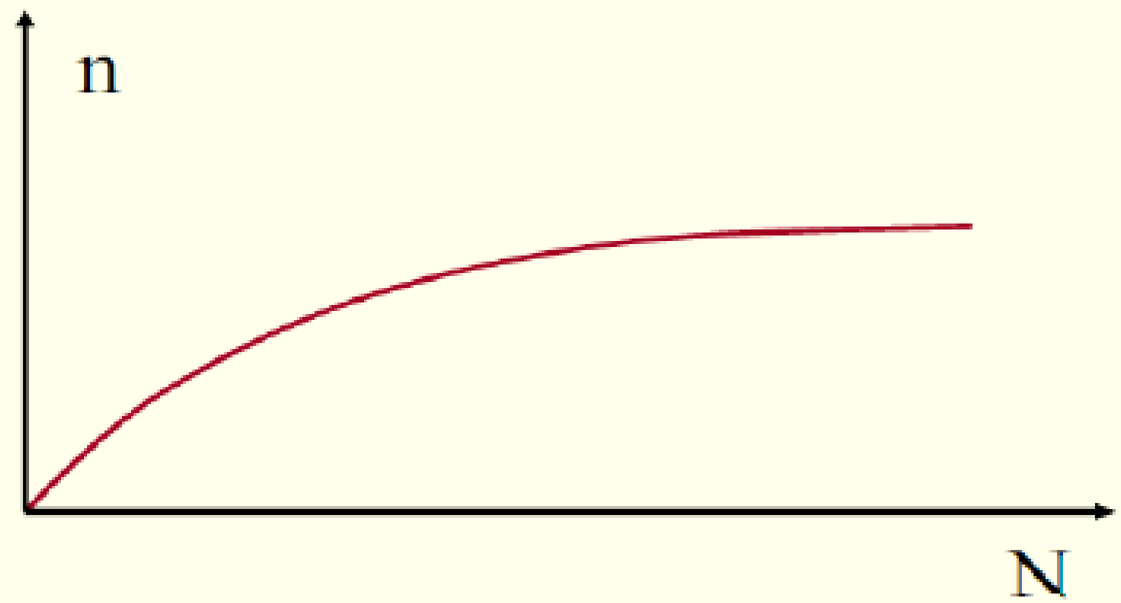
Observe que: N = 200 famílias, E_0 = 4%

n = 152 famílias ➔ 76% da população

Observe que: N = 200.000 famílias, E_0 = 4%

n = 623 famílias ➔ 0,3% da população

Logo, é errôneo pensar que o tamanho da amostra deve ser tomado como um percentual do tamanho da população para ser representativa



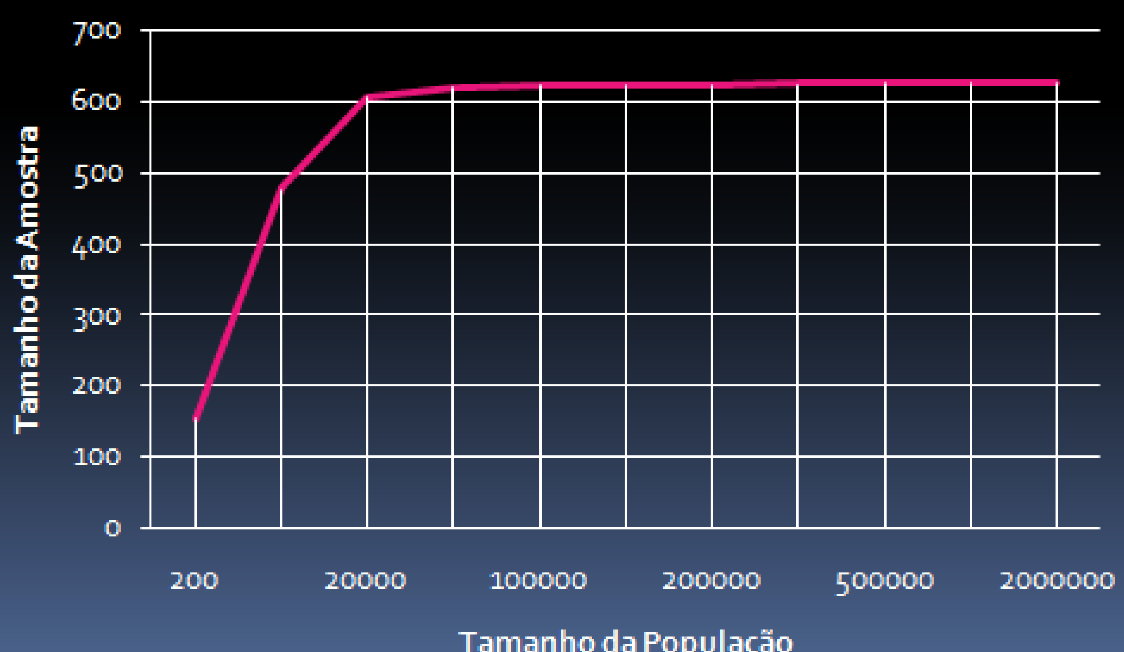
Estatística Descritiva

Introdução a amostragem

Tamanho da amostra para populações muito grandes.

Para E_o = 4%

N = 200	n = 152	76%
N = 200.000	n = 623	0,3%



Estatística Descritiva

Introdução a amostragem

Tamanho da amostra para populações muito grandes.

Para $E_0 = 4\%$

N = 200	n = 152	76%
N = 200.000	n = 623	0,3%

N => grande

TAMANHO DA AMOSTRA

$$n_0 = \frac{1}{E_0^2}$$

Exercício Tamanho da amostra ...

Numa pesquisa para uma eleição presidencial, qual deve ser o tamanho de uma amostra aleatória simples, se se deseja garantir um erro amostral não superior a 2% ?

Numa empresa com 1000 funcionários, deseja-se estimar a percentagem dos favoráveis a certo treinamento. Qual deve ser o tamanho da amostra aleatória simples que garanta um erro amostral não superior a 5%?

$N =$

$E_0 =$

$n_0 =$

$n =$

Exercício Tamanho da amostra ...

Numa pesquisa para uma eleição presidencial, qual deve ser o tamanho de uma amostra aleatória simples, se se deseja garantir um erro amostral não superior a 2% ?

$$n_0 = \frac{1}{E_0^2}$$

$n = n_0 = 1/(0,02)^2 = 1/0,0004 = 2500$ eleitores

Numa empresa com 1000 funcionários, deseja-se estimar a percentagem dos favoráveis a certo treinamento. Qual deve ser o tamanho da amostra aleatória simples que garanta um erro amostral não superior a 5%?

$N = 1000$ empregados

E_0 = erro amostral tolerável = 5% ($E_0 = 0,05$)

$$n_0 = \frac{1}{E_0^2}$$

$n_0 = 1/(0,05)^2 = 400$ empregados

$n = 1000 \times 400/(1000+400) = 286$ empregados

$$n = \frac{N.n_0}{N + n_0}$$

Fórmula para cálculo do tamanho da amostra para uma estimativa confiável da MÉDIA POPULACIONAL (μ):

$$n = \frac{(Z_{\alpha/2} \times S)^2}{E^2}$$

Onde:

n → Número de indivíduos na amostra;
Z α/2 → Valor crítico que corresponde ao grau de confiança desejado;
s → Desvio-padrão populacional da variável estudada;
E → Margem de erro ou ERRO MÁXIMO DE ESTIMATIVA. Identifica a diferença máxima entre a MÉDIA AMOSTRAL ($\ddot{X}$) e a verdadeira MÉDIA POPULACIONAL (μ) ⇔ [$\ddot{X}$ - μ];
α → Nível de significância.

Os valores de confiança mais utilizados e os valores de Z correspondentes.

Valores críticos associados ao grau de confiança na amostra

Grau de Confiança	α	Valor Crítico Zα/2
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

INV.NORMP.N(0,95) =	**1,645**	ou (1-α) = 0,90 ≡ 90%
INV.NORMP.N(0,975) =	**1,960**	ou (1-α) = 0,95 ≡ 95%
INV.NORMP.N(0,98) =	**2,054**	ou (1-α) = 0,96 ≡ 96%
INV.NORMP.N(0,995) =	**2,576**	ou (1-α) = 0,99 ≡ 99%

Exemplo

Um economista deseja estimar a renda média para o primeiro ano de trabalho de um bacharel em direito. Quantos valores de renda devem ser tomados, se o economista deseja ter 95% de confiança em que a média amostral esteja a menos de $500,00 da verdadeira média populacional? Suponha que saibamos, por um estudo prévio, que para tais rendas, s = $6.250,00.

Resolução

$$n = \frac{(Z_{\alpha/2} \times S)^2}{E^2}$$

Exemplo

Um economista deseja estimar a renda média para o primeiro ano de trabalho de um bacharel em direito. Quantos valores de renda devem ser tomados, se o economista deseja ter 95% de confiança em que a média amostral esteja a menos de $500,00 da verdadeira média populacional? Suponha que saibamos, por um estudo prévio, que para tais rendas, s = $6.250,00.

Resolução

Queremos determinar o tamanho **n** da amostra, dado que $\alpha = 0{,}05$ (95% de confiança) $Z_{\alpha/2}=1{,}96$.

Desejamos que a média amostral seja a menos de $ 500 da média populacional, de forma que E = 500

Supondo S = 6.250 e aplicamos a equação, obtendo:

$$n = \frac{(Z_{\alpha/2} \times S)^2}{E^2}$$

$$n=((1{,}96*6250)\ /500)^2 = 601$$

Com tal amostra teremos 95% de confiança em que a média amostral $\ddot{X}$ defira em menos de R$500,00 da verdadeira média populacional μ.

No mesmo problema e utilizando uma margem de erro maior, no caso $ 1.000,00 determine o tamanho da amostra:

Resolução:

$Z\ \alpha/2 = 1,96$

$S = 6.250$

$E = 1.000$

N = ?

$$n = \frac{(Z_{\alpha/2} \times S)^2}{E^2}$$

No mesmo problema e utilizando uma margem de erro maior, no caso $ 1.000,00 determine o tamanho da amostra:

Resolução:

Queremos determinar o tamanho **n** da amostra, dado que α = 0,05 (95% de confiança) *Z* α/2=1,96.
Desejamos que a média amostral seja a menos de $ 500 da média populacional, de forma que E = 1.000
Supondo S = 6.250 e aplicamos a equação, obtendo:

$$n = \frac{(Z_{\alpha/2} \times S)^2}{E^2}$$

N = ((1,96*6250)/1000)^2 => N = 150

Quando o desvio-padrão “S” for desconhecido, devemos utilizar um valor preliminar obtido por :

1. Utilizar a aproximação S = amplitude/4 ;

2. Realizar um estudo piloto, iniciando o processo de amostragem. Com base numa coleção de pelo menos 31 valores amostrais selecionados aleatoriamente, calcular o desvio-padrão amostral S e utilizá-lo num processo de refinado com a obtenção de mais dados amostrais.

3. Substituir o desvio padrão populacional σ (desconhecido) por um múltiplo ou fração desse mesmo desvio padrão. Por exemplo: substituindo o E (erro amostral) da fórmula, por 2σ, nesse caso, o fator σ é automaticamente eliminado.

$$n = \left(z.\frac{\sigma}{e} \right)^2$$

Quando o desvio-padrão "S" for desconhecido, devemos utilizar um valor preliminar obtido por :

Uma universidade está pesquisando o tempo médio diário que alunos das faculdades do país passam assistindo televisão. A pesquisa determina um grau de confiança de 96 % e deseja-se a estimativa com um erro de 0,25 horas, ou quinze minutos. Um estudo piloto, realizado especificamente para esta pesquisa mostrou um desvio padrão de 1,87 horas. Neste caso o desvio padrão populacional σ foi estimado pelo desvio padrão amostral s = 1,87 horas em um trabalho anterior.

Porém, se não existisse esse trabalho prévio, o que normalmente ocorre, poderia se supor o erro amostral como ***20%*** *de* **σ**, *ou seja,* **E = 0,20σ**. *Nesse caso, quantas entrevistas seriam necessárias para essa pesquisa? – (***Caso 3***)*

Para um **α = 96%** teremos um **Z = 2,054**

$$n = \left(z . \frac{\sigma}{E} \right)^2$$

$$n = \left(2{,}054 * \frac{\sigma}{0{,}20\sigma} \right)^2 \quad ou \quad n = 105$$

Quando o desvio-padrão “S” for desconhecido, devemos utilizar um valor preliminar obtido por :

A cadeia de lanchonetes Marreco Donaldo, MD, esta interessada em conhecer o gasto médio por pessoa, dos clientes de uma cadeia de lanchonetes concorrente. Os executivos da MD admitem que a despesa de um cliente possa variar entre R$ 3,00 e R$ 15,00. Eles desejam um grau de confiança de 98 % e admitem uma margem de erro de R$ 0,25. Pede-se calcular o tamanho da amostra para esta pesquisa? – (**Caso 1**)

$$\sigma = \frac{amplitude}{4} \qquad s = \frac{15-3}{4} = 3,0$$

Para um **α = 98%** teremos um **Z = 2,33**

$$n = \left(z.\frac{\sigma}{E}\right)^2 \qquad n = \left(2,33.\frac{3}{0,25}\right)^2 = 781,76$$

ou $n = 782$

Quando o desvio-padrão “S” for desconhecido, devemos utilizar um valor preliminar obtido por :

No problema do cálculo do erro amostral da pesquisa dos salários dos engenheiros na cidade de Manaus, como estimar a quantidade de entrevistas, sabendo que o erro amostral E=500 e grau de confiança 95%*?* – (**Caso 2**)

Como não conhecemos o desvio padrão da população, realizamos uma pesquisa amostral entre 30 engenheiros, escolhidos randomicamente. A pesquisa apresentou o resultado da tabela abaixo:

26.300	20.000	22.800	19.700	19.300
18.200	14.600	21.000	25.000	22.400
4.200	19.800	19.800	23.500	5.300
9.800	8.500	15.400	16.000	12.500
18.200	7.600	16.400	12.500	24.200
13.400	6.800	26.000	21.300	5.900

$$S = 6.592,62$$

Um **α = 95%** gera **Z = 1,96**

$$n = \left(z.\frac{\sigma}{E} \right)^2 \qquad n = \left(1,96.\frac{6.592,62}{500} \right)^2$$

$$ou \quad n = 688$$

Cálculo do tamanho da amostra quando a população for FINITA:

Nesse caso, o ERRO AMOSTRAL (E) deverá sofrer uma correção, ou seja, o ERRO deverá ser multiplicado pelo Coeficiente de População Finita = CPF.

$$E = z.\frac{\sigma}{\sqrt{n}} \qquad CPF = \sqrt{\frac{N-n}{N-1}} \qquad E = z.\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

Onde:

N = tamanho da população

n = tamanho da amostra

Existe, uma condição a ser satisfeita: o numero de amostras colhidas, **n**, deverá ser igual a pelo menos **5 %** do tamanho da população **N**.

O município de Arapiraca deseja fazer uma pesquisa do peso de papéis descartados mensalmente pelas residências da cidade, para planejamento da coleta de lixo. O peso médio do papel descartado em um mês, por uma amostra de 62 residências, foi de 9,4281 kg. e o desvio padrão dessa amostra foi de 4,1681 kg. Como a cidade possui 2.637 residências, deseja-se conhecer o intervalo de confiança para esta média, com grau de confiança de 95 %.

$n = 62 \quad ; \quad N = 2.637$

$\ddot{X} = 9{,}4281 \quad ; \quad S = 4{,}1681$

Um **α = 95%** gera **Z = 1,96**

Cálculo do ERRO AMOSTRAL quando a população for FINITA:

n = 62 ; N = 2.637

Ẍ = 9,4281 ; S = 4,1681

Um **α = 95%** gera **Z = 1,96**

$$E = z.\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$E = 1,96.\frac{4,1681}{\sqrt{62}}.\sqrt{\frac{2.637-62}{2.637-1}} = 1,025 \text{ kg}$$

INTERVALO DE CONFIANÇA:

P[(Ẍ-E)<μ<(Ẍ+E)] ou P[(9,4281 – 1,025) < μ < (9,4281 + 1,025)] ou

Intervalo de Confiança: P[8,4031 < μ < 10,4531]

Cálculo do tamanho da amostra quando a população for FINITA:

$$E = z.\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

gera:

$$n = \frac{z^2.\sigma^2.N}{E^2.(N-1)+z^2.\sigma^2}$$

Cálculo do tamanho da amostra quando a população for FINITA:

No caso anterior o município de Arapiraca deseja fazer uma pesquisa do peso de papéis descartados mensalmente pelas residências da cidade, para planejamento da coleta de lixo. O peso médio do papel descartado, por uma amostra teste em 31 residências, foi de 9,4281 kg e o desvio padrão dessa amostra foi de 4,1681 kg. Como a cidade possui 2.637 residências, deseja-se calcular o tamanho da amostra a ser coletada, com grau de confiança de 95 %. $E = 1{,}025$; $N = 2.637$

$\ddot{X} = 9{,}4281$; $S = 4{,}1681$

Um **α = 95%** gera **Z = 1,96**

$$n = \frac{z^2.\sigma^2.N}{E^2.(N-1)+z^2.\sigma^2}$$

$$n = \frac{1{,}96^2.4{,}1681^2.2637}{1{,}025^2.(2637-1)+1{,}96^2.4{,}1681^2} = 62$$

Fórmula para cálculo do tamanho da amostra com base na estimativa da proporção populacional (*p*):

$$n = \frac{Z_{\alpha/2}^{2} \cdot p \cdot q}{E^{2}}$$

Onde:

n	=	Número de indivíduos na amostra
$Z_{\alpha/2}$	=	Valor crítico que corresponde ao grau de confiança desejado.
p	=	Proporção populacional de indivíduos que pertence a categoria que estamos interessados em estudar.
q	=	Proporção populacional de indivíduos que NÃO pertence à categoria que estamos interessados em estudar (q = 1 – p).
E	=	Margem de erro ou ERRO MÁXIMO DE ESTIMATIVA. Identifica a diferença máxima entre a PROPORÇÃO AMOSTRAL e a verdadeira PROPORÇÃO POPULACIONAL (p).

Exemplo

Um assistente social deseja saber o tamanho da amostra (**n**) necessário para determinar a proporção da população atendida por uma Unidade de Saúde, que pertence ao município de Cariacica. Não foi feito um levantamento prévio da proporção amostral e, portanto, seu valor é desconhecido. Ela quer ter 90% de confiança que o **erro máximo de estimativa (E)** seja de ±5% (ou 0,05). Quantas pessoas necessitam ser entrevistadas?.

Resolução

$$n = \frac{Z^2_{\alpha/2} \cdot p \cdot q}{E^2}$$

OBS: **Quando os valores de "*p*" e "*q*" forem desconhecidos, considera-se *p* = *q* = 0,5 ou seja, *p*q* = 0,25**

Exemplo

Um assistente social deseja saber o tamanho da amostra (**n**) necessário para determinar a proporção da população atendida por uma Unidade de Saúde, que pertence ao município de Cariacica. Não foi feito um levantamento prévio da proporção amostral e, portanto, seu valor é desconhecido. Ela quer ter 90% de confiança em sua pesquisa e um **erro máximo de estimativa (E)** de ±5% (ou 0,05). Quantas pessoas necessitam ser entrevistadas?.

Resolução

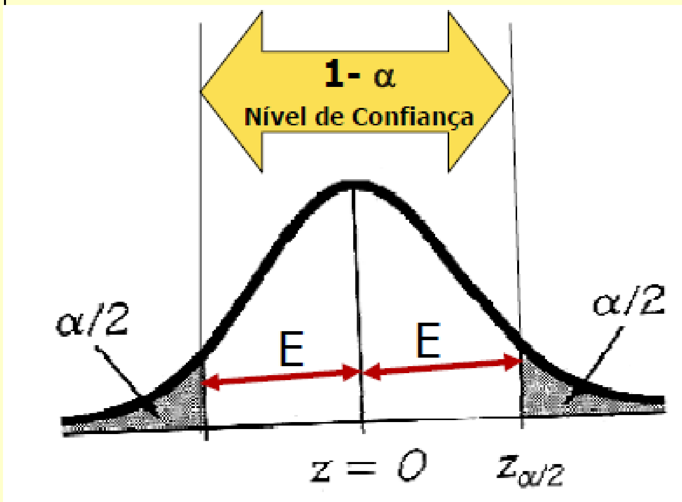
Considerando que o valor da proporção amostral de atendimentos para pessoas de Cariacica não é conhecida. Utilizamos a equação indicada para determinar o tamanho da amostra. Sabemos que, para 90% de confiança teremos o valor crítico ($Z\alpha/2$) = 1,645, conforme ***INV.NORMP.N(0,95) = 1,645***.

$$n = \frac{Z_{\alpha/2}^2 \cdot p \cdot q}{E^2}$$

$$n = \frac{\left[Z_{\alpha/2}\right]^2 \cdot 0{,}25}{E^2} = \frac{1{,}645^2 \cdot 0{,}25}{0{,}05^2} = 270{,}6 = 271$$

Mas se o Nível de Confiança for de 90% porque utilizar 0,95 como argumento da função INV.NORMP ?

Resposta: fórmula, $n = \frac{(Z_{\alpha/2} \times S)^2}{E^2}$ podemos obter $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$



Haverá uma probabilidade de **1-α** da média amostral conter um erro não superior a E. Por outro lado, haverá uma probabilidade **α = (α/2) + (α/2)** da média amostral conter um erro superior a E.

INV.NORMP.N(0,95) = 1,645.

F I M