Módulo 7 – Sessão 3

QUANTITATIVE STATISTICAL ANALYSIS AND MODELING

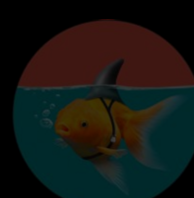
TUTORA
Carla Cardoso
Freelancer AI Manager

24 de março 2025

1

AGENDA

AMOSTRAGEM



ESTIMAÇÃO

4

AMOSTRAGEM: **MOTIVAÇÃO**

E D I T.

Na primeira sessão falamos de nem sempre é possível, ou vantajoso, estudar a População e é preferível trabalhar com uma **Amostra**.

Alguns dos cenários em que falamos foram:

- População de **elevada dimensão**
- **Custo** unitário para análise é muito elevada
- Estudo implica **destruição** das observações
- Populações **infinitas**

Mas **como** extrair uma amostra da população?

DATA SCIENCE & BUSINESS ANALYTICS Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3 5

5

AMOSTRAGEM: **PRINCIPAIS PASSOS**

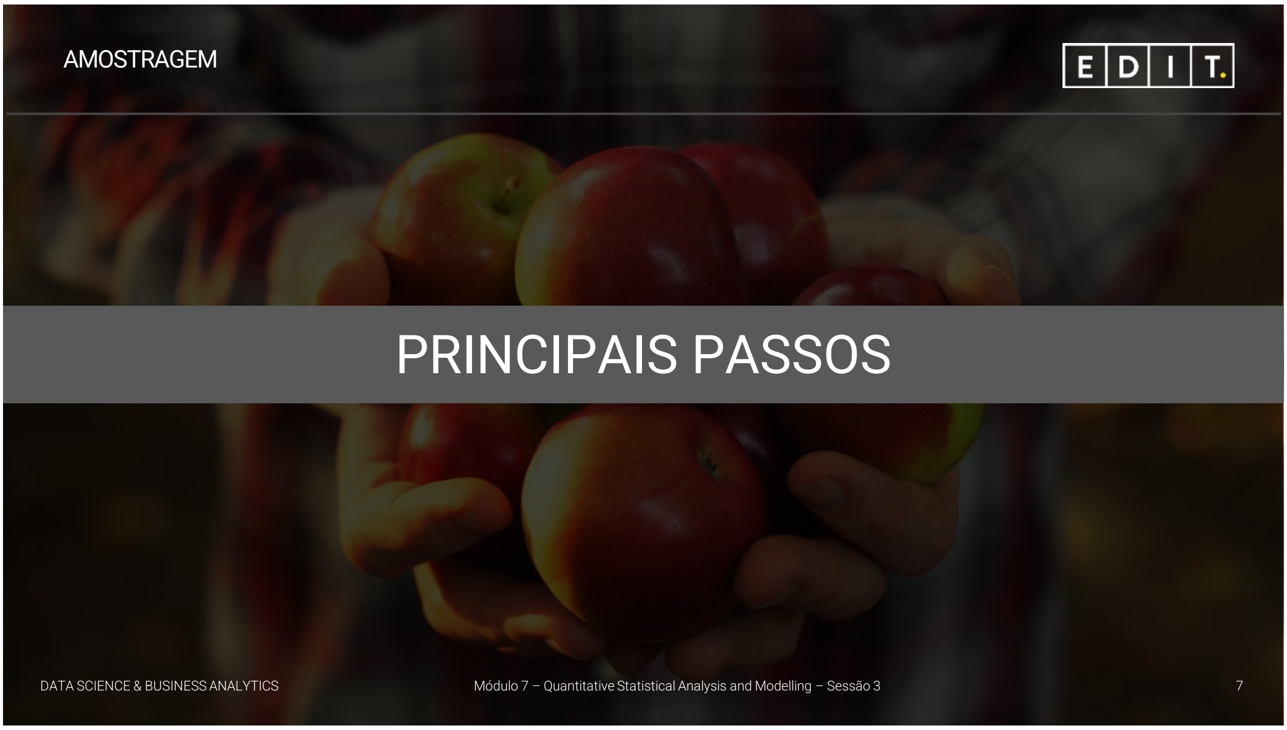
E D I T.

As principais etapas de um processo amostral incluem:

1. Definir a **variável de estudo**
2. Identificar a **população** alvo
3. Identificar a **base de amostragem**
4. Escolher uma **técnica amostral**
5. Determinar a **dimensão** da amostra
6. **Selecionar** os elementos da amostra
7. Recolher a **informação** necessária dos elementos da amostra

DATA SCIENCE & BUSINESS ANALYTICS Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3 6

6



AMOSTRAGEM

PRINCIPAIS PASSOS

7

PRINCIPAIS PASSOS: **VARIAVEL DE ESTUDO**

1. Variável de estudo

É necessário ter claro o que pretendemos com a nossa amostra.
Vamos analisar alguns exemplos:

- Resultado de eleições → p
- Processo de controlo de qualidade → p
- Variabilidade do caudal da foz de um rio → σ
- Gap salarial → $\mu_1 - \mu_2$

Estas métricas podem implicar **técnicas diferentes** para escolha das variáveis, especialmente se implicam mais do que uma população.

9

PRINCIPAIS PASSOS: **POPULAÇÃO**

2. **Identificar a população alvo**

Trata-se da população sobre a qual queremos extrair as nossas conclusões:

Nos exemplos anteriores:

- Resultado de eleições → **eleitores**
- Processo de controlo de qualidade → **produtos**
- Variabilidade do caudal da foz de um rio → **caudal da foz**
- Gap salarial → **empregados**

PRINCIPAIS PASSOS: **BASE DE AMOSTRAGEM**

3. **Identificar a base de amostragem**

Nem sempre temos acesso a toda a população para servir de base à extração da amostra, por isso começamos por definir a nossa base de sondagem.

Nos exemplos anteriores:

- Resultado de eleições → **lista telefónica?**
- Processo de controlo de qualidade → **fabrica X?**
- Variabilidade do caudal da foz de um rio → **último ano?**
- Gap salarial → **empregados > 6 meses antiguidade?**

AMOSTRAGEM

TÉCNICAS AMOSTRAIS

14

TÉCNICAS AMOSTRAIS: **CENSOS vs. SONDAGEM**

Existem **diferentes métodos** para extrair amostras, e isso deve-se às **diferentes necessidades** mas também a **diferentes restrições** que podem impedir de realizar amostragens mais complexas.

- **Censos** – Quando falamos em censos estamos a falar de procurar analisar a **totalidade da população**, ainda que tal seja praticamente impossível.
- **Sondagem** – por contraponto, a sondagem considera apenas **parte da população** (amostra) como base do estudo. É este o caso que vamos detalhar.

15

TÉCNICAS AMOSTRAIS: **MÉTODOS**

Os métodos podem ser organizados em **2 tipos**:

Métodos probabilísticos

. Amostragem aleatória simples
· Amostragem aleatória Estratificada
. Amostragem por clusters (ou cachos)
· Amostragem aleatória Sistemática

Métodos Não Probabilísticos

. Amostragem por conveniência
· Bola de Neve
· Amostragem por quotas
· Grupos objectivo

TÉCNICAS AMOSTRAIS: **MÉTODOS PROBABILISTICOS**

Métodos Probabilísticos

São amostragens no qual o processo de seleção tem por base uma **função aleatória**, atribuindo a cada elemento da população uma determinada **probabilidade de inclusão conhecida**.

- **Seleção aleatória**
- **Elimina** o **erro sistemático** e o **enviesamento** da seleção

- Implica **conhecimentos especializados** e por isso **mais custos**
- Implica conhecer **bem a população de interesse**, o que nem sempre é fácil

TÉCNICAS AMOSTRAIS: **MÉTODOS NÃO-PROBABILISTICOS**

EDIT.

São amostragens nas quais há uma **escolha determinada** dos elementos da amostra. Essa escolha pode ser feita de acordo com **critérios do analista** ou derivado a **limitações no estudo**.

Métodos Não-Probabilísticos

- **Processo mais barato**, dado que não obriga à analise da população
- **Implementação** tende a ser **mais simples**

- Pode comprometer a **representatividade da população**
- Reduzida capacidade de **generalização** dos resultados obtidos

18

19

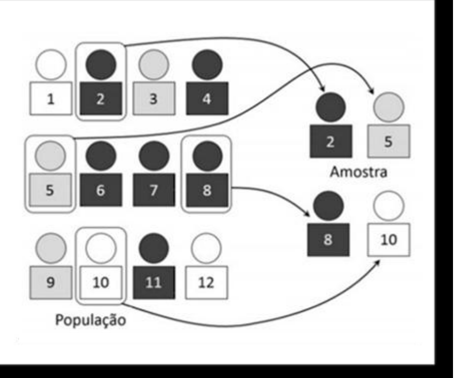
MÉTODOS PROBABILISTICOS: **AMOSTRAGEM ALEATÓRIA SIMPLES**

- **Amostragem Aleatória Simples**

A amostragem aleatória simples, é a uma das técnica de amostragem mais frequente, e atribui a **todos os elementos** que constituem a população a **mesma probabilidade** de ser selecionado para a amostra.

É o esquivamente a fazer um **sorteio** junto dos elementos da população.

Existem diferentes formas de extrair uma amostra aleatória, sendo a mais simples atribuir um **número aleatório** a cada elemento da população, ordenar a população segundo esse valor e extrair os n elementos do topo.



20

MÉTODOS PROBABILISTICOS: **AMOSTRAGEM ALEATÓRIA SISTEMÁTICA**

- **Amostragem Aleatória Sistemática**

O processo difere do aleatório simples pelo facto de **apenas o primeira elemento** é escolhido de forma verdadeiramente aleatória, sendo os restantes elementos **escolhidos segundo um critério** que é aplicado de forma sistemática a uma lista com os elementos da população (ex.: lista telefónica).

O processo consistem dois passos:

1. **Identificar primeiro elemento** de forma aleatória
2. Definir o **intervalo X entre indivíduos selecionados** e o sujeito selecionado seguinte
3. Escolher os elementos da amostra com base no **primeiro elemento** e depois indo **de X em X** indivíduos

21

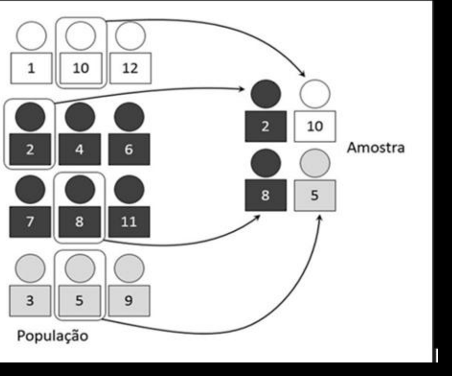
MÉTODOS PROBABILISTICOS: **AMOSTRAGEM ESTRATIFICADA**

❖ **Amostragem Estratificada**

Este tipo de amostragem procura garantir a **representatividade de determinados estratos** (subconjuntos) da população dada a sua importância para o estudo.

O processo consistem dois passos:

1. **Agrupar os indivíduos** segundo o seu **estrato**
2. Extrair uma **amostra aleatória simples** de dentro de cada um dos estratos **respeitando o peso de cada extrato**, ou seja, se o estrato A representa 20% da população, então 20% dos elementos da amostra devem ser do estrato A e esses devem ser selecionados aleatoriamente.



22

MÉTODOS PROBABILISTICOS: **AMOSTRAGEM POR CLUSTERS**

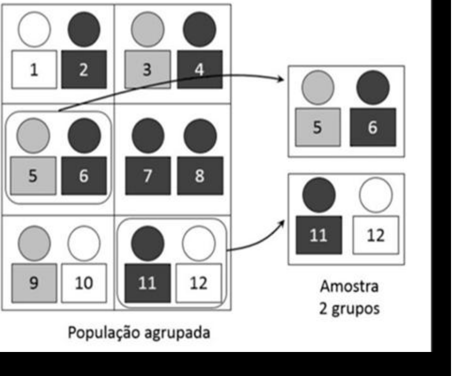
❖ **Amostragem por Clusters**

Neste tipo de amostragem tira-se partido da existência de **grupos** na população (ex.: escolas) cujas características os tornam **representativos da população**, ou seja, eles contém a diversidade/variabilidade da população (ex.: escola ABC).

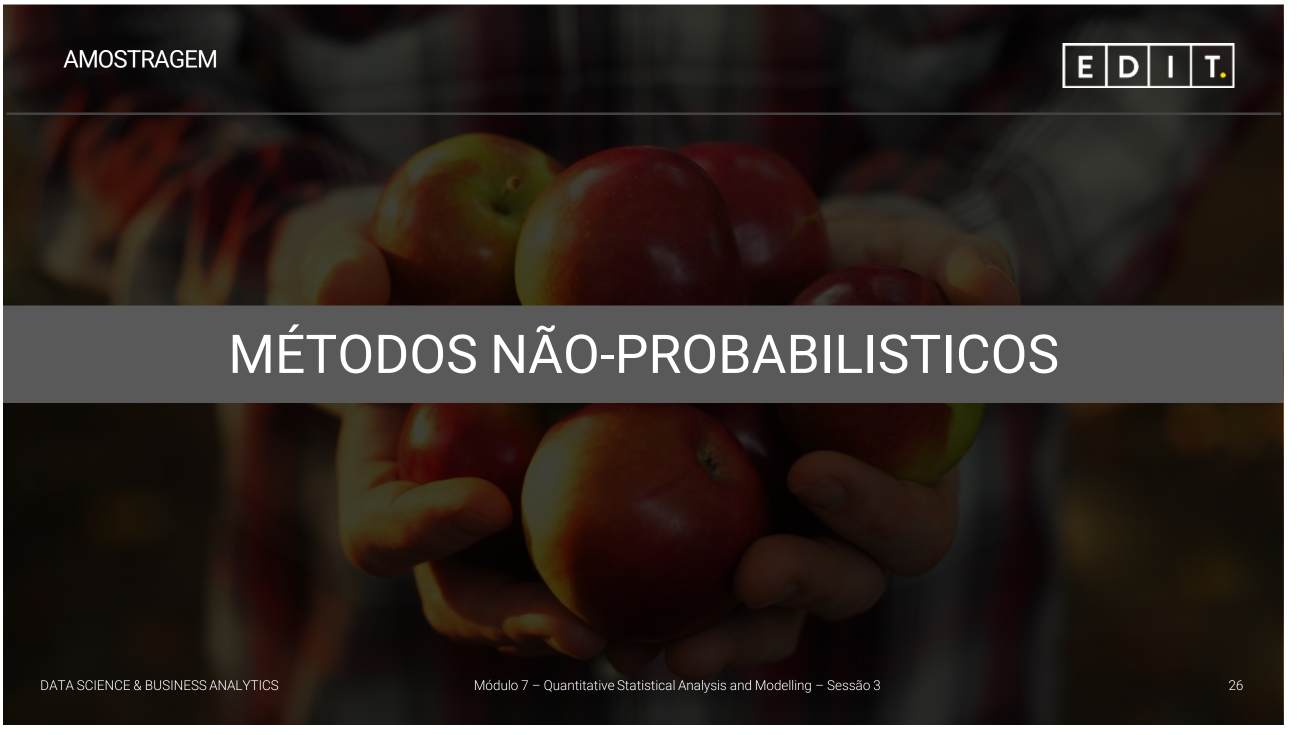
O processo consistem dois passos:

1. **Identificar os grupos** representativos
2. Considerar **todos** os indivíduos dos grupos identificados para a amostra

Em comparação com a mostragem estratificada, este processo permite **poupança de custos e tempo**, aumentando a sua eficiência, ainda que à custa de uma possível **perda de precisão**.



25



AMOSTRAGEM

E D I T.

MÉTODOS NÃO-PROBABILISTICOS

DATA SCIENCE & BUSINESS ANALYTICS — Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3 — 26

26

MÉTODOS NÃO-PROBABILISTICOS

E D I T.

- **Amostragem por Conveniência:** Esta técnica é muito comum e consiste em selecionar uma amostra da população que seja acessível. Ou seja, os indivíduos que entram nessa pesquisa são selecionados porque eles estão prontamente disponíveis, não porque eles foram selecionados por meio de um critério estatístico.

- **Bola de Neve:** A amostra por bola de neve é uma técnica de amostragem não probabilística onde os indivíduos selecionados para serem estudados convidam novos participantes da sua rede de amigos e conhecidos.

- **Amostragem por quotas**: Podemos encontrar a amostragem por quotas como a versão não probabilística da amostra estratificada, onde cada estrato é representado por indivíduos selecionados por métodos não-probabilísticos, como bola de neve.

DATA SCIENCE & BUSINESS ANALYTICS — Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3 — 27

27

28

EXERCICIO 1

EDIT.

Considerem novamente o ficheiro **laptop_pricing** definam uma variável aleatória **X** com distribuição Binomal em função da variável **price**.

Segue um exemplo:

X = "Proporção de portáteis no mercado abaixo dos 700€"

Extraiam 3 amostras com n = 50, com base nos seguintes critérios:

1. Amostra **aleatória simples**
2. Amostra **Estratificada por RAM_GB**
3. Amostragem por **Clusters** considerando os clusters**: Lenovo e Toshiba**

Calculem, para a **população** (ou seja, todo o ficheiro) o **valor esperado** e a **variância de X**. Depois, para **cada** uma das vossas **amostras**, calculem a média e o desvio padrão.

Que **conclusões** podes tirar sobre o método amostral que melhor permite calcular o **valor real de X**?

DATA SCIENCE & BUSINESS ANALYTICS — Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3 — 29

29

EXERCICIO 1

E D I T.

Equipa 1
Rui
Tamara

Equipa 2
Gonçalo
Alexandre

Equipa 3
José F.
Carolina M.

Equipa 4
Nuno
Yhoanna

Equipa 5
Susana
Stéfane

Equipa 6
Filipa
Andreia

Equipa 7
Joana
José P.M.

Equipa 8
Ana
João
Carolina L.

DATA SCIENCE & BUSINESS ANALYTICS — Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3 — 30

30

31

DIMENSÃO DA AMOSTRA: **FATORES DETERMINANTES**

Fatores determinantes na dimensão da amostra

- **Precisão** e **Confiança** (requeridas para os resultados)
- **Variabilidade** da população no que respeita à característica em estudo
- **Distribuição** amostral do estimador utilizado na estimação do parâmetro
- **Orçamento** disponível para o estudo
- **Tempo** previsto dedicar ao processo de recolha de dados

DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

Muitas vezes estes 2 termos aparecem juntos, precisão e confiança significam **coisas muito diferentes**.

- **Precisão** – Erro que estou disposto a assumir na minha conclusão.

DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

Muitas vezes estes 2 termos aparecem juntos, precisão e confiança significam **coisas muito diferentes**.

❖ **Precisão** – Erro que estou disposto a assumir na minha conclusão.

Por exemplo: Na estimativa da proporção de eleitores que vão votar no candidato X, qual o erro que estou disposto a admitir?

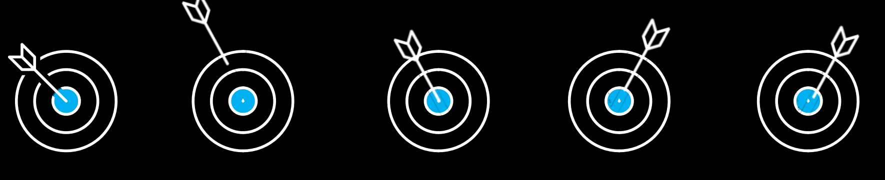
Imaginem **3 estudos** com os seguintes resultados com respeito ao intervalo de confiança para **percentagem de eleitores** a votar no candidato. **Qual dos estudos vocês acham ter mais qualidade?**

- Estudo 1: intervalo de confiança [90% : 98%]
- Estudo 2: intervalo de confiança[45% : 50%]
- Estudo 3: intervalo de confiança[12% : 13%] → erro = 1%

DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

Mas falta-nos informação para comparar a qualidade dos estudos. Falta-nos a **confiança**.

❖ **Confiança** – Dá-nos a percentagem de vezes em que, retirada uma amostra da mesma dimensão, o parâmetro em estudo, se encontra no intervalo definido.



Confiança de 80%

DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

EDIT.

Mas falta-nos informação para comparar a qualidade dos estudos. Falta-nos a **confiança**.

- **Confiança** – Dá-nos a percentagem de vezes em que, retirada uma amostra da mesma dimensão, o parâmetro em estudo, se encontra no intervalo definido.

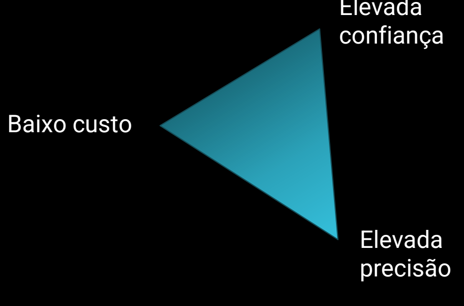
Por exemplo: Revendo os **3 estudos** com respeito não só ao **erro**, mas também à **confiança**, será que podemos chegar a conclusões diferentes? **Agora, qual dos estudos vocês acham ter mais qualidade?**

- Estudo 1: intervalo de confiança [90% : 98%] com 95% confiança
- Estudo 2: intervalo de confiança[45% : 50%] com 90% confiança
- Estudo 3: intervalo de confiança[12% : 13%] com 70% confiança

DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

EDIT.

Nós vamos querer então estudos com:



DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

DIMENSÃO DA AMOSTRA: **PRECISÃO E CONFIANÇA**

Nós vamos querer então estudos com:

Elevada confiança **95%**

Baixo custo

Elevada precisão **5%***

* Para o caso de estudo de proporções

DIMENSÃO DA AMOSTRA: **PROPORÇÕES**

Existem **formulas para o calculo** da dimensão da amostra, sendo que todas têm como base informação sobre:

- ❖ Precisão
- ❖ Confiança
- ❖ Variabilidade da população

Para o caso de uma **proporções**, estamos a falar de que **distribuição teórica**?

DIMENSÃO DA AMOSTRA: **PROPORÇÕES**

Existem **formulas para o calculo** da dimensão da amostra, sendo que todas têm como base informação sobre:

- ❖ Precisão
- ❖ Confiança
- ❖ Variabilidade da população

Para o caso de uma **proporções**, como estamos a falar de uma v.a. com distribuição Binomial (n , p), temos que a dimensão para a amostra pretendida é:

$$n = \frac{z^2 \hat{p}(1-\hat{p})}{e^2}$$

Onde ***z*** é o valor da distribuição Normal (0,1) correspondente à confiança pretendida, **e** é o erro assumido e ***p*** é a proporção estimada*.

* Recorremos a estudos prévios, a uma amostra inicial ou consideramos p = 50% que é o pior cenário

DIMENSÃO DA AMOSTRA: **PROPORÇÕES**

Para populações finitas, é necessário fazer um ajuste à formula, para ter em conta a dimensão da população:

$$n = \frac{\frac{z^2\hat{p}(1-\hat{p})}{e^2}}{1+\frac{z^2\hat{p}(1-\hat{p})}{N\,e^2}}$$

Onde **N** é a dimensão da população, ***z*** é o valor da distribuição Normal (0,1) correspondente à confiança pretendida, ***e*** é o erro assumido e ***p*** é a proporção estimada*.

Esta formula é mais complexa pois inclui um "**ajuste**" derivado do facto da população ter uma dimensão mais reduzida, mas para populações infinitas ou com **N > 10,000 as 2 formulas dão resultados equivalentes**.

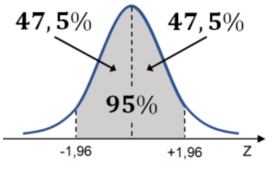
* Recorremos a estudos prévios, a uma amostra inicial ou consideramos p = 50% que é o pior cenário

DIMENSÃO DA AMOSTRA: **TABELA DISTRIBUIÇÃO NORMAL**

Como passamos de uma **nível de confiança** para o valor de ***z***?

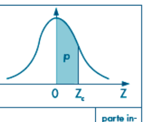
Visto que uma v.a. X ~ B (***n***, ***p***) não é mais que uma soma de v.a. independentes e $X_1, X_2, \ldots X_n$ ~ Bernouli(***p***)*, então, pelo **Teorema do Limite Central**, podemos aproximar a distribuição da variável X a uma distribuição normal, e passamos a usar os pontos de referência dessa distribuição.

Podemos então concluir que, para uma confiança de 95% o valor de ***z*** é **1,96**:



* Caso particular da Binomial com n = 1

Tabela III — Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3

DIMENSÃO DA AMOSTRA: **EXEMPLO PARA UMA PROPORÇÃO**

E D I T.

Exemplo I: Imaginem que queremos estimar a proporção de peças defeituosas num processo produtivo. O nosso estudo tem de ter as seguintes restrições:

- Precisão = 5%
- Confiança = 95%
- Proporção na da população = 10%

Aplicando a formula para calculo do n, temos:

$$n = \frac{1{,}96^2 0.1\ (1-0.1)}{0.05^2} \sim 137$$

* Recorremos a estudos prévios, a uma amostra inicial ou consideramos p = 50% que é o pior cenário

DIMENSÃO DA AMOSTRA: **EXEMPLO PARA UMA PROPORÇÃO**

E D I T.

Exemplo II: Imaginem que queremos estimar a proporção de peças defeituosas num processo produtivo. O nosso estudo tem de ter as seguintes restrições:

- Precisão = 5%
- Confiança = 95%
- Proporção na da população = 50%

Aplicando a formula para calculo do n, temos:

$$n = = \frac{1.96^2 0.5\ (1-0.5)}{0.05^2} \sim 384$$

* Recorremos a estudos prévios, a uma amostra inicial ou consideramos p = 50% que é o pior cenário

DIMENSÃO DA AMOSTRA: **EXEMPLO PARA UMA PROPORÇÃO**

Exemplo III: Imaginem que queremos estimar a proporção de peças defeituosas num processo produtivo. O nosso estudo tem de ter as seguintes restrições:

- Precisão = 5%
- Confiança = 99%
- Proporção na da população = 10%

Aplicando a formula para calculo do n, temos:

$$n = \frac{2.575^2 0.1\ (1-0.1)}{0.05^2} \sim 239$$

* Recorremos a estudos prévios, a uma amostra inicial ou consideramos p = 50% que é o pior cenário

DIMENSÃO DA AMOSTRA: **EXEMPLO PARA UMA PROPORÇÃO**

Exemplo IV: Imaginem que queremos estimar a proporção de peças defeituosas num processo produtivo. O nosso estudo tem de ter as seguintes restrições:

- Precisão = 1%
- Confiança = 95%
- Proporção na da população = 10%

Aplicando a formula para calculo do n, temos:

$$n = \frac{1.96^2 0.1\ (1-0.1)}{0.01^2} \sim 3.457$$

* Recorremos a estudos prévios, a uma amostra inicial ou consideramos p = 50% que é o pior cenário

55

EXERCICIO 2

EDIT.

Vamo-nos manter no ficheiro **laptop_pricing**, ele representa **a população** de portáteis existentes no mercado. Considerem a variável aleatória **X** com distribuição Binomal definida da seguinte forma:

X = "O portátil é da marca **Dell**"

Calculem a dimensão da amostra para os seguintes parâmetros:

1. Precisão = 1% + Confiança = 90%
2. Precisão = 5% + Confiança = 95%
3. Precisão = 10% + Confiança = 99%

Extraiam **3 amostras aleatórias**, uma para cada dimensão calculada acima, e calculem a em cada uma delas a **proporção amostral ($\hat{p}$)**.

Qual dos valores está mais próximo do **valor "real"** de p?

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 3

56

56

EXERCICIO 2

E D I T.

Equipa 1
Rui
Tamara

Equipa 2
Gonçalo
Alexandre

Equipa 3
José F.
Carolina M.

Equipa 4
Nuno
Yhoanna

Equipa 5
Susana
Stéfane

Equipa 6
Filipa
Andreia

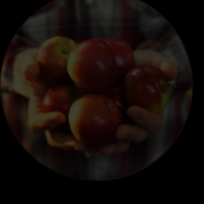
Equipa 7
Joana
José P.M.

Equipa 8
Ana
João
Carolina L.

57

AGENDA

E D I T.



AMOSTRAGEM

ESTIMAÇÃO

58

ESTIMAÇÃO: **INTRODUÇÃO**

E D I T.

É no processo de Inferência Estatística que, **a partir de dados de uma amostra**, procuramos **estimamos** características sobre a população.

Essas características podem ser diversas, sendo que as mais comuns de estimar são:

- Valor Esperado μ
- Proporção ***p***

Utilizamos os "correspondentes" indicadores na amostra como estimadores para os indicadores populacionais, ou seja:

- $\hat{\mu} = \bar{x}$
- $\hat{p} = \frac{x}{n}$

Onde é $\bar{x}$ representa a média amostral e $\frac{x}{n}$ frequência relativa amostral.

ESTIMAÇÃO: **CARACTERISTICAS DE UM BOM ESTIMADOR**

E D I T.

Ainda que não tenhamos acesso aos valor populacionais, o que procuramos é definir **estimadores** com **qualidade**.

Mas como medir a qualidade de um estimador?

Existem 2 características criticas para um bom estimado:

- ❖ E(estimador) = parâmetro **Centrado**
- ❖ $\lim_{n \to +\infty}$ var(estimador) = 0 **A variância tende para 0**

ESTIMAÇÃO: **INTERVALOS CONFIANÇA PROPORÇÕES**

Para além de estimarmos um parâmetro, também procuramos obter um intervalo de confiança para o valor obtido.

No caso de uma **proporção,** um intervalo de confiança com confiança 100(1-α)% é dado por:

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, ,\hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

ESTIMAÇÃO: **INTERVALOS CONFIANÇA PROPORÇÕES**

Exemplo:

Caso a nossa amostra seja de **50 observações**, a proporção $\hat{p}$ **= 0,3** e confiança é 95% (α **= 5%**), o intervalo de confiança para p, a proporção populacional é dada por:

$$\left[0.3 - z_{1-0,05/2}\sqrt{\frac{0,3(1-0,3)}{100}}\, , 0.3 + z_{1-0,05/2}\sqrt{\frac{0,3(1-0,3)}{100}}\right] =$$

$$= \left[0.3 - z_{0.475}\sqrt{\frac{0,21}{50}}\, , 0.3 + z_{0.475}\sqrt{\frac{0,21}{50}}\right] = = [0.3 - 1.96 * 0.06\, , 0.3 + 1.96 * 0.06] = \mathbf{[0.17\ ,\ 0.43\ ]}$$

primeira decimal de Z_s	0	1	2	3	4	5	6	7	8	9	primeira decimal de Z_s
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9

ESTIMAÇÃO: **INTERVALOS CONFIANÇA PROPORÇÕES**

Exemplo:

Caso a nossa amostra seja de **50 observações**, a proporção $\hat{p}$ **= 0,3** e confiança é 95% (α **= 5%**), o intervalo de confiança para p, a proporção populacional é dada por: **[0.17 , 0.43]**

Interpretação:

Num conjunto de **100 amostras** de **dimensão 50**, estimamos que em **95** delas o valor de p ϵ **[0.17 , 0.43].**

ESTIMAÇÃO: **INTERVALOS CONFIANÇA PROPORÇÕES**

Exemplo:

Caso a nossa amostra seja de **50 observações**, a proporção $\hat{p}$ **= 0,3** e confiança é 95% (α **= 5%**), o intervalo de confiança para p, a proporção populacional é dada por: **[0.17 , 0.43]**

Interpretação:

Num conjunto de **100 amostras** de **dimensão 50**, estimamos que o valor real de p pertence a **95** dos 100 intervalos de confiança estimados.


EXPLORATORY DATA ANALYSIS
EDIT.
Questionário
DATA SCIENCE & BUSINESS ANALYTICS
Módulo 3 – Exploratory Data Analysis – Sessão 1
65

65