

Estatística Basica II



Universidade Estadual Paulista, Júlio de Mesquita Filho - UNESP

clayton.pereira@unesp.br



Por que estudar Estatística

Referências e Fontes das Imagens

- [Python Data Science Handbook](#)

Distribuição de Frequência

	Parâmetro	Estimador
Definição	Descreve uma característica (medida) do conjunto de dados	Estimativa de um parâmetro
Conjunto de dados	População	Amostra
Valor	Fixo (normalmente desconhecido)	Calculado e aleatório

Amostra (Exemplo): Quero descobrir a renda média das pessoas que trabalham no Brasil.

Posso perguntar para todo mundo?

Amostra (Exemplo): Quero descobrir a renda média das pessoas que trabalham no Brasil.

Posso perguntar para todo mundo?

- ☐ **População:** Pessoas brasileiras que possuem carteira de trabalho assinada.

Amostra (Exemplo): Quero descobrir a renda média das pessoas que trabalham no Brasil.

Posso perguntar para todo mundo?

- ☐ **População:** Pessoas brasileiras que possuem carteira de trabalho assinada.
- ☐ **Amostra:** Pessoas que responderam a uma pesquisa no centro da cidade.
- ☐ **Unidade amostral:** Indivíduo que respondeu a pesquisa.

Amostra (Exemplo): Quero descobrir a renda média das pessoas que trabalham no Brasil.

Posso perguntar para todo mundo?

- ☐ **População:** Pessoas brasileiras que possuem carteira de trabalho assinada.
- ☐ **Amostra:** Pessoas que responderam a uma pesquisa no centro da cidade.
- ☐ **Unidade amostral:** Indivíduo que respondeu a pesquisa.
- ☐ **Variável (característica):** informações da **unidade amostral**, como estado de residência, gênero, idade, profissão e salário.

Amostra (Exemplo):

Nome	Estado	Escolaridade	Gênero	Idade	Profissão	Salário
João	SP	Pós Graduação Incomp.	M	36	Programador	R\$ 7.000
José	RJ	Fundamental Incomp.	M	28	Pedreiro	R\$ 3.500
Maria	MS	Fundamental Completo	F	19	Trainee	R\$ 1.000
Marcia	RJ	Pós Graduação Comp.	F	42	Web Develop	R\$ 10.000
Felipe	SP	Superior Completo	M	50	Professor	R\$ 4.800
Pedro	MG	Superior Incomp.	M	46	Programador II	R\$ 4.000
Joyce	MG	Ensino Básico	F	39	Designer	R\$ 1.700

Variáveis

- Os dados da coluna **IdCustomer** são de qual tipo?

	A	B	C
1	IdCustomer	Name	Points
2	aa3eaf74-6d9c-4859-b733-5a18a3b2f71b	afonso_rf	681
3	2d3d2dce-d353-4961-ad39-46723efe2100	marlicmartins	1731
4	5f8fcbe0-6014-43f8-8b83-38cf2f4887b3	teomewhy	6349
5	ca95ef2a-5129-40f4-acbc-2ced25940032	tdlupus	421
6	65662aff-44d6-4f06-b9d9-07445c6e5943	kozat0	820
7	98b960e7-6b7b-45b7-b7af-60bca40e04b9	gu1z17	0
8	b2f9d026-0727-4125-b84b-c60af3148a15	dcha0tic	2804
9	6e1bc660-02eb-49ac-ae6-592328504897	jesak_	434
10	95ef4c1e-b21a-468e-b419-ad08333c4948	marciormr	120
11	e7c79c1a-7693-4e63-a844-c755e0bc3bd3	gabriel_aka	1063
12	c4f241f8-2140-4f72-b1a3-904fcc51e38c	otal_xandao	598
13	e407fad9-4f76-4c12-98ea-d70d1c8d0d97	juligaioso	1460
14	cf317911-6a16-403b-9ea0-ad9b90d44a1a	danszk	131
15	d92138ad-7b6f-4b5b-8e0d-5041fe5b943a	tak3_iteasy	4

Variáveis

- ☐ Os dados da coluna **IdCustomer** são de qual tipo?
- **Qualitativa Nominal!!!**

	A	B	C
1	IdCustomer	Name	Points
2	aa3eaf74-6d9c-4859-b733-5a18a3b2f71b	afonso_rf	681
3	2d3d2dce-d353-4961-ad39-46723efe2100	mariicmartins	1731
4	5f8fcbe0-6014-43f8-8b83-38cf2f4887b3	teomewhy	6349
5	ca95ef2a-5129-40f4-acbc-2ced25940032	tdlupus	421
6	65662aff-44d6-4f06-b9d9-07445c6e5943	kozat0	820
7	98b960e7-6b7b-45b7-b7af-60bca40e04b9	gu1z17	0
8	b2f9d026-0727-4125-b84b-c60af3148a15	dcha0tic	2804
9	6e1bc660-02eb-49ac-ae6-592328504897	jesak_	434
10	95ef4c1e-b21a-468e-b419-ad08333c4948	marciormr	120
11	e7c79c1a-7693-4e63-a844-c755e0bc3bd3	gabriel_aka	1063
12	c4f241f8-2140-4f72-b1a3-904fcc51e38c	otal_xandao	598
13	e407fad9-4f76-4c12-98ea-d70d1c8d0d97	juligaioso	1460
14	cf317911-6a16-403b-9ea0-ad9b90d44a1a	danszk	131
15	d92138ad-7b6f-4b5b-8e0d-5041fe5b943a	tak3_iteasy	4

Variáveis

☐ E as demais colunas?

	A	B	C
1	IdCustomer	Name	Points
2	aa3eaf74-6d9c-4859-b733-5a18a3b2f71b	afonso_rf	681
3	2d3d2dce-d353-4961-ad39-46723efe2100	mariicmartins	1731
4	5f8fcbe0-6014-43f8-8b83-38cf2f4887b3	teomewhy	6349
5	ca95ef2a-5129-40f4-acbc-2ced25940032	tdlupus	421
6	65662aff-44d6-4f06-b9d9-07445c6e5943	kozat0	820
7	98b960e7-6b7b-45b7-b7af-60bca40e04b9	gu1z17	0
8	b2f9d026-0727-4125-b84b-c60af3148a15	dcha0tic	2804
9	6e1bc660-02eb-49ac-ae6-592328504897	jesak_	434
10	95ef4c1e-b21a-468e-b419-ad08333c4948	marciormr	120
11	e7c79c1a-7693-4e63-a844-c755e0bc3bd3	gabriel_aka	1063
12	c4f241f8-2140-4f72-b1a3-904fcc51e38c	otal_xandao	598
13	e407fad9-4f76-4c12-98ea-d70d1c8d0d97	juligaioso	1460
14	cf317911-6a16-403b-9ea0-ad9b90d44a1a	danszk	131
15	d92138ad-7b6f-4b5b-8e0d-5041fe5b943a	tak3_iteasy	4

Medidas Resumo:

Também conhecida como Estatística Descritiva

- Valores que resumem e descrevem um conjunto de dados de forma concisa, permitindo uma compreensão mais rápida e fácil das características centrais e distribuição dos dados.

Medidas Resumo: Tabela de Frequencia

- ☐ Utilizadas para descrever a distribuição de **variáveis qualitativas**

Medidas Resumo: Tabela de Frequencia

- ☐ Utilizadas para descrever a distribuição de **variáveis qualitativas**
- ☐ Dessa forma é possível identificar quanto cada nível (ou valor) da variável ocorre e sua representação em todo o conjunto.

Medidas Resumo: Frequencia Absoluta

- Chamamos de frequência absoluta o número de vezes que um mesmo dado apareceu dentro de um conjunto.

Medidas Resumo: Frequencia Absoluta

- ☐ Chamamos de frequência absoluta o número de vezes que um mesmo dado apareceu dentro de um conjunto.
- ☐ Independentemente do tipo de pesquisa que estamos fazendo, é inevitável calcular a frequência absoluta.

Medidas Resumo: Frequencia Absoluta

- ☐ Chamamos de frequência absoluta o número de vezes que um mesmo dado apareceu dentro de um conjunto.
- ☐ Independentemente do tipo de pesquisa que estamos fazendo, é inevitável calcular a frequência absoluta.
- ☐ Podemos denotar por n_i a quantidade de ocorrência (frequencia) do i -ésimo nível da variável

Medidas Resumo: Frequencia Absoluta

- Quanto estamos trabalhando com estatística, precisamos decidir **qual variável** será aplicado a técnica estatística.

Medidas Resumo: Frequencia Absoluta

- ☐ Quanto estamos trabalhando com estatística, precisamos decidir **qual variável** será aplicado a técnica estatística.
- ☐ Independentemente do tipo de pesquisa que estamos fazendo, é inevitável calcular a frequência absoluta.

Medidas Resumo: Frequencia Absoluta

- Quanto estamos trabalhando com estatística, precisamos decidir **qual variável** será aplicado a técnica estatística.
- Independentemente do tipo de pesquisa que estamos fazendo, é inevitável calcular a frequência absoluta.
- Podemos denotar por n_i a quantidade de ocorrência (frequencia) do i -ésimo nível da variável

Medidas Resumo: Frequencia Relativa

- Representa a proporção ou a porcentagem de ocorrências de um determinado evento em comparação ao número total de um conjunto de dados

Medidas Resumo: Frequencia Relativa

- Representa a proporção ou a porcentagem de ocorrências de um determinado evento em comparação ao número total de um conjunto de dados
- É a **frequência absoluta** dividida pelo número total de elementos no conjunto e então multiplicada por 100 para obter a representação em percentual.

Medidas Resumo: Frequencia Relativa

- ☐ Representa a proporção ou a porcentagem de ocorrências de um determinado evento em comparação ao número total de um conjunto de dados
- ☐ É a **frequência absoluta** dividida pelo número total de elementos no conjunto e então multiplicada por 100 para obter a representação em percentual.
- ☐ Fornece informações sobre a importância relativa de cada valor em relação ao todo, destacando a proporção ou peso que cada valor representa no conjunto.

Medidas Resumo: Frequencia Relativa

□ f_i = frequência absoluta

Medidas Resumo: Frequencia Relativa

- f_i = frequência absoluta
- n = total de elementos (ou soma das frequências)

Medidas Resumo: Frequencia Relativa

- f_i = frequência absoluta
- n = total de elementos (ou soma das frequências)
- $f_r\%$ = frequência relativa em percentual

Medidas Resumo: Frequência Relativa

- f_i = frequência absoluta
- n = total de elementos (ou soma das frequências)
- $f_r\%$ = frequência relativa em percentual

$$f_r\% = \left(\frac{f_i}{n} \right) \times 100$$

Medidas Resumo: Frequencia Relativa

- O intuito desse tipo de análise, é possibilitar que através do resultado obtido, permita **inferir** sobre os dados da população.

Medidas Resumo: Frequencia Relativa

- O intuito desse tipo de análise, é possibilitar que através do resultado obtido, permita **inferir** sobre os dados da população.
- Podemos entender como uma forma de **normalizar** os dados baseado no valor da porcentagem adquirida na **amostra**.

Medidas Resumo:

Tabela de Frequencia II (continuação)

- Na aula passada, verificamos esses dois tipos de frequencia em nossos dados:

Intervalos	Freq. Absoluta	Freq. Relativa
12.50 a 13.50	2	$2/20 = 0,1 \rightarrow 10\%$
13.51 a 14.50	3	$3/20 = 0,15 \rightarrow 15\%$
14.51 a 15.50	7	$7/20 = 0,35 \rightarrow 35\%$
15.51 a 16.50	5	$5/20 = 0,25 \rightarrow 25\%$
16.51 a 17.50	3	$3/20 = 0,15 \rightarrow 15\%$

Handwritten notes: A red bracket on the left groups the intervals from 12.50 to 16.50, labeled with a red 'K'. A blue bracket at the bottom groups the intervals from 15.51 to 17.50, labeled with a blue 'a'.

Medidas Resumo:

Frequencia Absoluta Acumulada

- A **frequência absoluta acumulada** ou, **frequência acumulada**, é a soma das frequências absolutas simples de cada variável.

Medidas Resumo:

Frequencia Absoluta Acumulada

- A **frequência absoluta acumulada** ou, **frequência acumulada**, é a soma das frequências absolutas simples de cada variável.
- Nessa frequência, os valores numéricos são somados, acumulando, de uma variável para a outra, até a última variável estudada.

Medidas Resumo:

Frequencia Absoluta Acumulada

- A **frequência absoluta acumulada** ou, **frequência acumulada**, é a soma das frequências absolutas simples de cada variável.
- Nessa frequência, os valores numéricos são somados, acumulando, de uma variável para a outra, até a última variável estudada.
- O crescimento de F_1 ajuda a perceber onde os dados se concentram (acúmulo rápido) ou se estão bem distribuídos.

Medidas Resumo:

Frequencia Absoluta Acumulada

- Podemos denotar N_i a ocorrência acumulada do i -ésimo nível da variável no conjunto de dados, dado o acúmulo das frequências dos níveis anteriores

Medidas Resumo:

Frequencia Absoluta Acumulada

- Podemos denotar N_i a ocorrência acumulada do i -ésimo nível da variável no conjunto de dados, dado o acúmulo das frequências dos níveis anteriores
- F_i = frequência absoluta acumulada até a i -ésima classe
- f_i = frequência absoluta da i -ésima classe
- $\sum_{j=1}^i f_j$ = somatório das frequências até a classe i

$$F_i = \sum_{j=1}^i f_j$$

Medidas Resumo: Frequencia Relativa Acumulada

- A **frequência relativa acumulada**, é a percentagem relativa à frequência acumulada, somam-se as frequências relativas acumuladas da classe em que estamos e e das anteriores.
- f_j = frequência absoluta da j -ésima classe
- F_i = frequência absoluta acumulada até a i -ésima classe
- n = total de observações
- F_{ri} = frequência relativa acumulada até a i -ésima classe

$$F_{ri} \% = \left(\frac{F_i}{n} \right) \times 100$$

Medidas Resumo: Interpretação dos dados

- ☐ Vamos analisar os dados apresentados na tabela abaixo, como podemos interpretar os resultados?

A	B	C	D	E	F
Classe (Nota)	f _i	F _i	f _r (%)	F _r (%)	
0-2	2	2	10	10	
2-4	4	6	20	30	
4-6	6	12	30	60	
6-8	5	17	25	85	
8-10	3	20	15	100	

Medidas Resumo: Interpretação dos dados

- Vamos pegar como exemplo a linha da classe 4–6, teremos:
 - $f_i = 6$: 6 alunos tiraram entre 4 e 6.
 - $F_i = 12$: até essa faixa, 12 alunos no total já foram contabilizados.
 - $f_r = 30\%$: 30% da turma tirou entre 4 e 6.
 - $F_r = 60\%$: 60% dos alunos tiraram até 6..

Exercício: Analizando a tabela abaixo, apresente os valores de frequência de gênero.

Nome	Estado	Escolaridade	Gênero	Idade	Profissão	Salário
João	SP	Pós Graduação Incomp.	M	36	Programador	R\$ 7.000
José	RJ	Fundamental Incomp.	M	28	Pedreiro	R\$ 3.500
Maria	MS	Fundamental Completo	F	19	Trainee	R\$ 1.000
Marcia	RJ	Pós Graduação Comp.	F	42	Web Develop	R\$ 10.000
Felipe	SP	Superior Completo	M	50	Professor	R\$ 4.800
Pedro	MG	Superior Incomp.	M	46	Programador II	R\$ 4.000
Joyce	MG	Ensino Básico	F	39	Designer	R\$ 1.700

Nos vemos na próxima aula

□ Bons estudos, abraços!