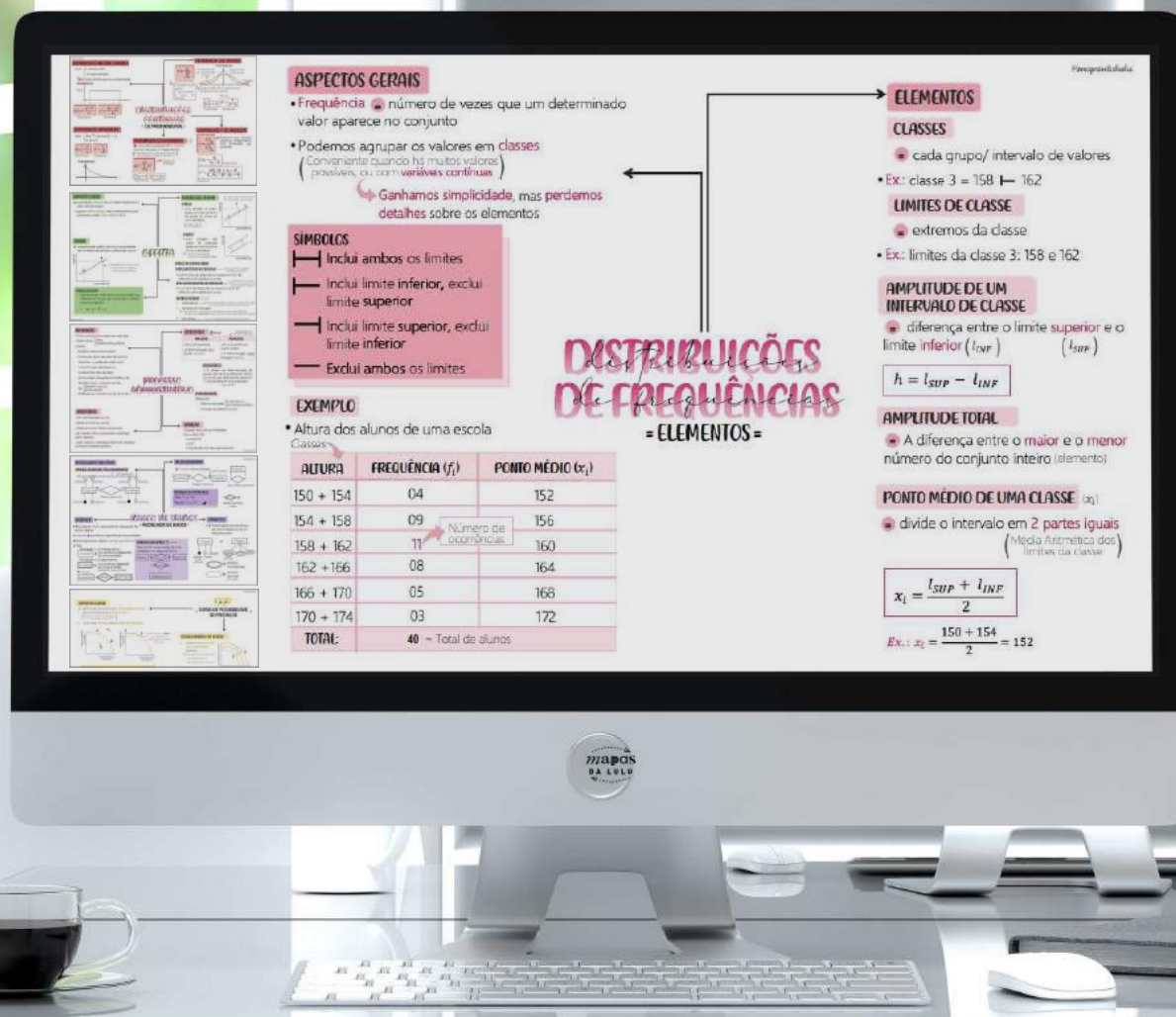




estatística

MAPAS MENTAIS PARA CONCURSOS PÚBLICOS

SEJA MUITO BEM-VINDO!

Obrigada por adquirir os Mapas da Lulu 3.0! Tenho certeza de que esse material fará toda a diferença em seus estudos e será um atalho para a sua tão sonhada aprovação!

Para quem ainda não me conhece, meu nome é Laura Amorim (@lulu.concurseira), tenho 28 anos, e, após pouco mais de um ano e meio de estudos, fui aprovada em quatro concursos públicos: Auditor Fiscal do Estado de Santa Catarina (7º lugar), Auditor Fiscal do Estado de Goiás (23º lugar), Consultor Legislativo (4º lugar) e Agente da Polícia Federal (primeira fase), tendo superado uma concorrência de mais de mil candidatos por vaga!

Aprendi que a revisão, muitas vezes ignorada, é a parte mais importante (e essencial!) do aprendizado! Após testar vários métodos, percebi que os meus mapas mentais são, com toda certeza, os melhores instrumentos de estudo e revisão. Ao longo da minha preparação, fiz e utilizei mais de 700 mapas mentais, desenvolvendo e aperfeiçoando um método próprio de sua construção até chegar aos Mapas da Lulu 3.0, aos quais você terá acesso a partir de agora:

Os Mapas da Lulu 3.0 visam, sobretudo, otimizar suas revisões e aumentar seu número de acertos de questões, te ajudando a chegar mais rápido à aprovação! Após resolver mais de 14.700 questões de concursos públicos nos últimos dois anos, percebi quais são os assuntos mais cobrados pelas bancas e suas principais pegadinhas, e todo esse conhecimento foi incorporado em meus mapas para que você, que confia no meu trabalho, possa sair na frente dos seus concorrentes!

Ah, e se você não quiser perder minhas dicas de estudos e motivação diárias, inscreva-se no meu canal do Youtube: Lulu Concurseira e no meu Instagram: @lulu.concurseira. Já somos uma comunidade de mais de 220 mil concurseiros em busca do mesmo sonho: a aprovação!



Um beijo,
Laura Amorim
@laura.amorimc

PIRATARIA É CRIME

ATENÇÃO:

Este produto é para uso pessoal. Não compartilhe o seu material.

Pessoal, os Mapas da Lulu são resultado de mais de dois anos de dedicação aos estudos. Ainda hoje, reservo boa parte do meu dia para produzir conteúdo, responder dúvidas, aconselhar e dar dicas sobre concursos públicos gratuitamente por meio dos meus perfis no Instagram (@laura.amorimc e @mapasdalulu) e no Youtube (Laura Amorim).

Nunca tive a pretensão de ganhar muito dinheiro com a venda desse material, até mesmo porque prestei concurso público para, dentre outros motivos, alcançar a estabilidade e segurança financeira que queria.

Mas preciso cobrir meus custos com site, servidores, distribuição, design e também minhas horas de trabalho empregadas, debruçada sobre a escrivaninha, dores nas costas, cansaço físico e mental.

São mais de 1.600 Mapas Mentais, com tempo médio de uma hora e meia para elaboração de cada um deles. Recebo menos de 50 centavos por hora trabalhada, para poder contribuir para sua aprovação.

Em razão disso, já agradecida pelo carinho e compreensão de todos, peço que **NÃO COMPARTILHE O MATERIAL** por nenhum meio (sites, e-mail, grupos de WhatsApp ou Facebook...). Se você vir qualquer compartilhamento suspeito, peço que denuncie essa fonte ilegal, por favor e também me envie no contato@mapasdalulu.com.br. **Pirataria é crime** e pode resultar penas de até QUATRO anos de prisão, além de multa (art. 184, CP).

O compartilhamento do material pelo aluno importará em seu bloqueio imediato.

Agradeço a todos pelo enorme carinho e respeito. Espero que aproveitem muito os Mapas da Lulu.

Um beijo,
Laura Amorim

ÍNDICE

1. ESTATÍSTICA

1.1 Distribuições de Frequências	06
1.2 Apresentação de Dados	08
1.3 Médias	10
1.4 Medidas Separatrizes	13
1.5 Moda	16
1.6 Medidas de Dispersão	18
1.7 Análise Combinatória	21
1.8 Probabilidade	22
1.9 Variáveis Aleatórias Discretas	24
1.10 Variáveis Aleatórias Contínuas	26
1.11 Distribuições Discretas de Probabilidade	27
1.12 Distribuições Contínuas de Probabilidade	29
1.13 Amostragem	31
1.14 Estimadores	32
1.15 Intervalos de Confiança	34

ÍNDICE

1. ESTATÍSTICA

1.16 Testes de Hipóteses	36
1.17 Análise de Variância	39
1.18 Regressão Linear Simples	41

ASPECTOS GERAIS

- **Frequência** = número de vezes que um determinado valor aparece no conjunto
- Podemos agrupar os valores em **classes**
(Conveniente quando há muitos valores possíveis, ou com **variáveis contínuas**)
- Ganhamos simplicidade, mas perdemos detalhes sobre os elementos

SÍMBOLOS

- ┌ Inlui ambos os limites
- └ Inlui limite inferior, exclui limite superior
- ┐ Inlui limite superior, exclui limite inferior
- ─ Exclui ambos os limites

EXEMPLO

- Altura dos alunos de uma escola
- Classes

ALTURA	FREQUÊNCIA (f_i)	PONTO MÉDIO (x_i)
150 + 154	04	152
154 + 158	09	156
158 + 162	11	160
162 + 166	08	164
166 + 170	05	168
170 + 174	03	172
TOTAL:	40 = Total de alunos	

Número de ocorrências

distribuições de frequências
= ELEMENTOS =

ELEMENTOS

CLASSES

- = cada grupo/ intervalo de valores
- Ex.: classe 3 = 158 ┌ 162

LIMITES DE CLASSE

- = extremos da classe
- Ex.: limites da classe 3: 158 e 162

AMPLITUDE DE UM INTERVALO DE CLASSE

- = diferença entre o limite superior e o limite inferior (l_{INF}) (l_{SUP})

$$h = l_{SUP} - l_{INF}$$

AMPLITUDE TOTAL

- = A diferença entre o maior e o menor número do conjunto inteiro (elemento)

PONTO MÉDIO DE UMA CLASSE (x_i)

- = divide o intervalo em 2 partes iguais
(Média Aritmética dos limites da classe)

$$x_i = \frac{l_{SUP} + l_{INF}}{2}$$

Ex.: $x_i = \frac{150 + 154}{2} = 152$

FREQUÊNCIA ABSOLUTA SIMPLES

- = Número de dados na respectiva classe (f_i)
(elementos)
- Soma das frequências simples de todas as classes = total de elementos (n)

$$\sum f_i = n$$

FREQUÊNCIA RELATIVA SIMPLES (Normalmente em porcentagem)

- = **razão** entre a frequência simples da respectiva classe e a frequência total:

$$f_{ri} = \frac{f_i}{n}$$

DENSIDADE DE FREQUÊNCIA

(= Razão entre a frequência da classe e sua amplitude:

$$d = \frac{f_i}{h}$$

EXEMPLO (Frequências simples)

- Altura dos alunos de uma escola

ALTURA	FREQUÊNCIA (f_i)	FREQUÊNCIA RELATIVA (f_{ri})
150 + 154	04	4/40 = 0.1 (10%)
154 + 158	09	9/40 = 0.225 (22,5%)
158 + 162	11	11/40 = 0.275 (27,5%)
162 + 166	08	8/40 = 0.2 (20%)
166 + 170	05	5/40 = 0.125 (12,5%)
170 + 174	03	3/40 = 0.075 (7,5%)
TOTAL:	40 (n)	1.00 (100%)

Absoluta simples

DISTRIBUIÇÕES DE FREQUÊNCIAS

= TIPOS DE FREQUÊNCIAS =

FREQUÊNCIAS ACUMULADAS

- Pode-se calcular por frequências absolutas ou relativas.

FREQUÊNCIA ACUMULADA CRESCENTE

se decrescente, faz-se o mesmo procedimento, de baixo para cima.

1. Copiar a freq. absoluta da 1ª classe
2. Para o cálculo da frequência seguinte: **somar** a frequência acumulada anterior com a absoluta da classe correspondente

ALTURA	FREQUÊNCIA (f_i)	FREQ. ACUMULADA (f_{ac})
150 + 154	04	04
154 + 158	09	13
158 + 162	11	24
162 + 166	08	32
166 + 170	05	37
170 + 174	03	<u>40</u>
TOTAL:	40 (n)	40

A freq. acumulada da última classe = total de elementos (n)

- A **frequência acumulada** de uma classe, indica o número de **elementos menores** que seu limite superior

DIAGRAMA DE RAMOS E FOLHAS

- Outra representação de dados em rol

• **ROL:** 11 11 12 13 13 20 21 21 30 31 33

42 65 72 73

1	11233
2	011
3	013
4	2
6	5
7	23

1ª coluna
dezena

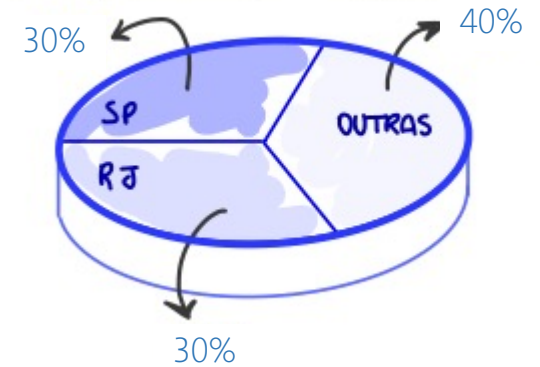
Unidades
correspondentes

APRESENTAÇÃO DE DADOS

GRÁFICO DE SETORES (PIZZA)

- Útil para apresentar para mostrar **divisão** de um todo em partes

CIDADE NATAL x FREQUÊNCIA



- Cada setor circular é **proporcional** à respectiva **frequência**

é possível encontrar o **ângulo** por regra de três:

100% → 360°

30% → x

CIDADE NATAL x FREQUÊNCIA

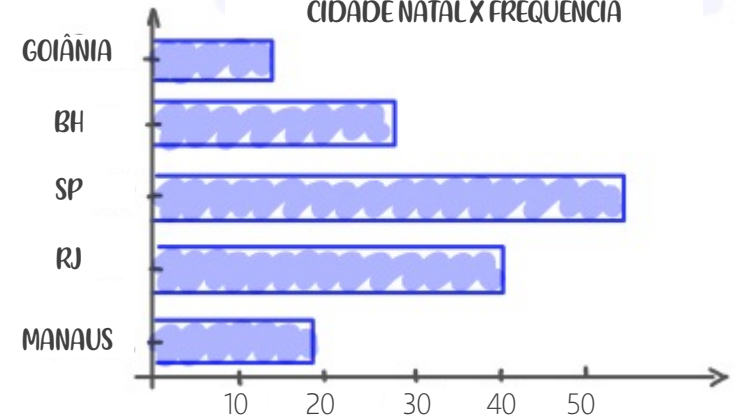


GRÁFICO DE COLUNAS OU BARRAS JUSTAPOSTAS

- Para dados agrupados por valor ou atributo
- **Ex.:** idade x frequência

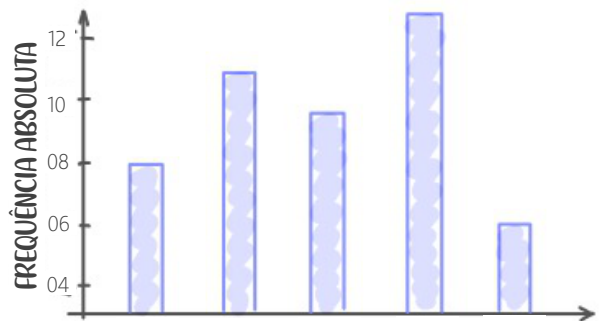


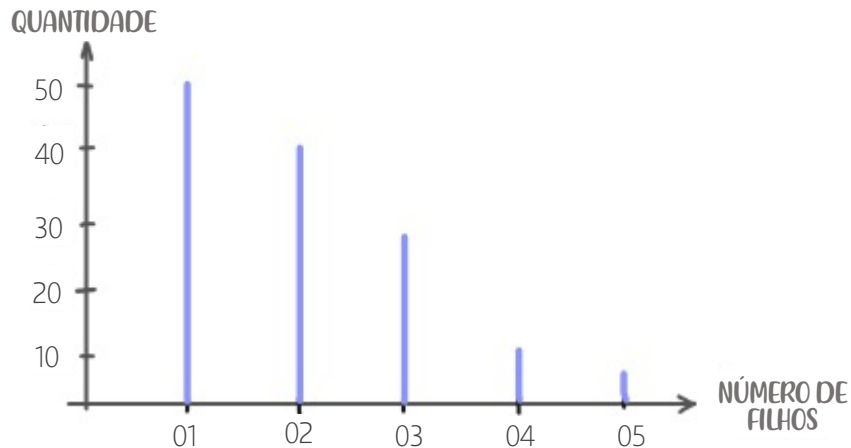
GRÁFICO DE LINHAS

- Usados na representação de séries temporais



GRÁFICO DE HASTES OU BASTÕES

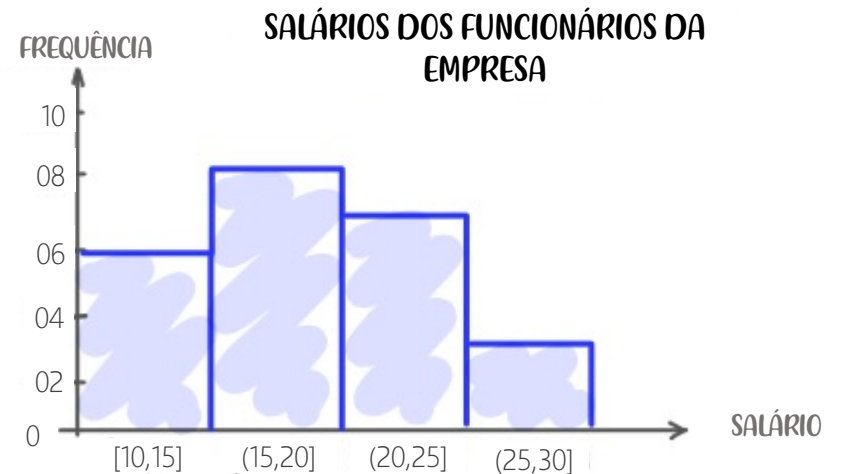
- Usados para representar dados não agrupados em classes
(Normalmente = dados discretos)



APRESENTAÇÃO DE DADOS

HISTOGRAMAS

- Usados para representar dados agrupados em classes
(Normalmente = dados contínuos)
- Relaciona classe ↔ frequências por retângulos contínuos
- Área de cada retângulo é **proporcional** à **frequência**



MÉDIA ARITMÉTICA SIMPLES

$$\bar{x} = \frac{\text{soma dos termos}}{\text{número de termos}}$$

Ex.: média aritmética simples dos números 3, 5, 9, 2, 11

$$\bar{x} = \frac{3 + 5 + 9 + 2 + 11}{5}$$

$$= \frac{30}{5} \rightarrow \bar{x} = 6$$

Se todos os números forem substituídos por $\bar{x}$, a **soma** dos termos será **preservada**.

$$\text{soma} = \bar{x} \cdot n$$

PROPRIEDADES DA MÉDIA ARITMÉTICA

- Sempre existe e é única

$$\text{menor } n^{\circ} \text{ do conjunto} \leq \bar{x} \leq \text{maior } n^{\circ} \text{ do conjunto}$$

- A **soma dos desvios** em relação à média é **nula**
- A **soma** do quadrado **dos desvios** em relação à média é **mínima**

- Somando-se uma constante c** a todos os números, a

nova média ($\bar{x}'$) será: $\bar{x}' = \bar{x} + c$ (O equivalente para subtração)

- Multiplicando-se** todos os números por **uma constante c**,

a nova média ($\bar{x}'$) será: $\bar{x}' = \bar{x} \cdot c$ (O equivalente para divisão)

Use para simplificar os termos antes de calcular as médias

MÉDIAS

MÉDIA ARITMÉTICA PONDERADA

- Como a simples, mas os elementos (x_i) podem ter **pesos diferentes** (p)
(Como em uma prova, em que as questões de uma matéria vale mais que de outra)

$$\bar{x}_p = \frac{\text{soma dos termos multiplicados pelos respectivos pesos}}{\text{soma dos pesos}}$$

- Ex.: média aritmética ponderada dos seguintes números e seus pesos:

- 3, peso 2
- 4, peso 1
- 2, peso 5

$$\bar{x}_p = \frac{3 \cdot 2 + 4 \cdot 1 + 2 \cdot 5}{2 + 1 + 5}$$

$$= \frac{6 + 4 + 10}{8}$$

$$= \frac{20}{8}$$

$$\bar{x}_p = 2,5$$

MÉDIAS



MÉDIA PARA DADOS AGRUPADOS POR VALOR

- = mesma ideia da média ponderada
- Normalmente para dados **discretos**
- **Ex.:** idade dos alunos de uma escola

valor	Número de ocorrências	
IDADE (x_i)	FREQUÊNCIA (f_i)	$x_i \cdot f_i$
25	21	525
30	47	1.410
34	54	1.836
38	41	1.558
41	37	1.517
TOTAL:	200	6.846

(= total de alunos)

Você calcula! (coluna auxiliar)

= 25.21

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i}$$

$$= \frac{6.846}{200} \rightarrow \bar{x} = 34,23 \text{ anos}$$

MÉDIA PARA DADOS AGRUPADOS POR CLASSE

- Normalmente para dados **contínuos**
- **Ex.:** altura dos alunos de uma escola

classe	Número de ocorrências		Você calcula!
ALTURA	FREQUÊNCIA (f_i)	PONTO MÉDIO (x_i)	$x_i \cdot f_i$
150-154	4	152	608
154-158	9	156	1.404
158-162	11	160	1.760
162-166	8	164	1.312
166-170	5	168	840
170-174	3	172	516
TOTAL:	40		6.440

(= total de alunos)

= 152.4

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{6.440}{40} \rightarrow \bar{x} = 161 \text{ cm}$$

CÁLCULO DO PONTO MÉDIO (x_i)

$$x_i = \frac{\text{limite inferior da classe} + \text{limite superior da classe}}{2}$$

$$\text{Ex.: } x_1 = \frac{150 + 154}{2} = 152$$

MÉDIAS



MÉDIA GEOMÉTRICA

- Raiz n-ésima do produto dos termos (n = número de termos)

$$G = \sqrt[n]{x_1 \cdot x_2 \dots x_n}$$

- Ex.: média geométrica dos termos 3, 8, 9:
3 termos (n)

$$G = \sqrt[3]{3 \cdot 8 \cdot 9} \quad G = \sqrt[3]{216}$$

$$G = 6$$

Se todos os números forem substituídos por G, o produto dos termos será preservado

produto dos termos = $\underbrace{G \cdot G \dots G}_n$

DESIGUALDADE DAS MÉDIAS



- Para uma sequência de números positivos

$$\bar{x} \geq G \geq H \quad \left(\text{Só é igual quando todos os números forem iguais} \right)$$

MÉDIA HARMÔNICA

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

- = inverso da média aritmética dos inversos:

$$H = \left(\frac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n} \right)^{-1} \quad \left(\text{Fórmula alternativa} \right)$$

Se todos os número forem substituídos por $\bar{H}$, a soma dos inversos dos termos será preservada

soma dos inversos = $\frac{n}{\bar{H}}$

- Ex.: média harmônica dos termos 3, 4, 9:
3 termos (n)

$$H = \frac{3}{\frac{1}{3} + \frac{1}{4} + \frac{1}{9}}$$

$$H = \frac{3}{\frac{12 + 9 + 4}{36}} = \frac{3}{\frac{25}{36}} = \frac{108}{25}$$

$$H = 4,32$$

MEDIDAS SEPARATRIZES

medidas separatrizes

= MEDIANA =

MEDIDAS SEPARATRIZES

- **Dividem** os dados em **partes**
- É necessário que os dados estejam dispostos em **ordem** crescente (ou decrescente)
- Dispostos em "rol"

MEDIANA (Md)

- = **número** que se encontra no **centro** de uma série de números.

MEDIANA PARA DADOS NÃO-AGRUPADOS

NÚMERO ÍMPAR DE TERMOS

3, 4, 5, 7, 8, 9, 10, 12, 13

4 elementos 4 elementos

mediana

- **Mediana** = termo de ordem $\frac{n+1}{2}$

NÚMERO PAR DE TERMOS

3, 4, 5, 7, 8, 9, 10, 12, 13, 15

4 elementos 4 elementos

mediana = $\frac{8+9}{2} = 8.5$ (ponto médio)

- **Mediana** = média aritmética entre o termo de ordem $n/2$ e $n/2 + 1$

PROPRIEDADES

CAI MUITO!

- A mediana **não** é influenciada pelo **valores extremos** do rol (depende da posição)
- **Somando-se** (subtraindo-se) uma constante **c** de todos os valores → a mediana também é somada/subtraída de **c**.

$$M_d' = M_d + c$$

$$M_d' = M_d - c$$

- **Multiplicando-se** (dividindo-se) todos os valores por uma constante **c** → a mediana também é multiplicada/dividida por **c**.

$$M_d' = M_d \cdot c$$

$$M_d' = M_d : c$$

- A **soma** dos **módulos dos desvios** da sequência de números x_i em relação a um número é **mínima** se em relação à **mediana**

MEDIANA PARA DADOS AGRUPADOS

SEM INTERVALOS DE CLASSE

- Ex.: notas de alunos em uma classe

EXEMPLO 1:

Número de alunos

NOTAS	FREQÜÊNCIA (f_i)	FREQÜÊNCIA ACUMULADA (f_{ac})
2	2	2
4	6	8
6	10	18
8	12	30
10	9	39
TOTAL:	39 (n)	39

Total = 39 (ímpar) Md = número na posição $(n+1)/2$

Md = 20º termo

Ele está na classe de nota 8!

Logo, Md = 8

EXEMPLO 2:

NOTAS	FREQÜÊNCIA (f_i)	FREQÜÊNCIA ACUMULADA (f_{ac})
2	2	2
4	6	8
6	10	18
8	12	30
10	6	36
TOTAL:	36 (n)	36

Total = 36 (par)

Md = média entre $n/2$ e $n/2+1$

Md = média entre $x_{(18)}$ e $x_{(19)}$

= 6

= 8

Logo, Md = 7

MEDIANAS PARA DADOS AGRUPADOS

EM CLASSES

ALTURA	FREQÜÊNCIA (f_i)	FREQÜÊNCIA ACUMULADA (f_{ac})
40-50	2	2
50-60	5	7 $f_{ac_{ant}}$
60-70	7	14
70-80	8	22
80-90	3	25
TOTAL:	25 (n)	25

- 1º passo.: determinar a classe mediana

Encontrar a classe onde esteja a frequência acumulada $n/2$

$$n/2 = 25/2 = 12,5$$

Está entre f_{ac} 7 e 14 → logo, classe mediana = 60 - 70

- 2º passo: aplicar a fórmula:

$$M_d = l_i + \left[\frac{n/2 - f_{ac_{anterior}}}{f_i} \right] \cdot h$$

DO EXEMPLO

$$M_d = 60 + \left[\frac{12,5 - 7}{7} \right] \cdot 10$$

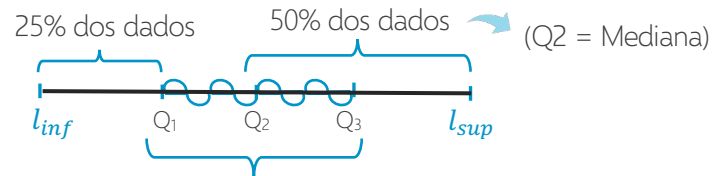
$$M_d = 67,85$$

l_i	Limite inferior
$f_{ac_{ant}}$	Frequência acumulada da classe anterior
h	Amplitude da classe → 70 - 60
f_i	Frequência simples da classe mediana

MEDIDAS SEPARATRIZES = MEDIANA =

QUARTIL

- Divide os dados em **4 partes** de mesma frequência
 São **3 quartis** com **25% dos dados** cada



= Amplitude interquartílica = $Q_3 - Q_1$

- Amplitude semi-interquartílica = $(Q_3 - Q_1)/2$



ATENÇÃO!

Q_1 = mediana entre l_{inf} e Q_2

Q_3 = mediana entre l_{sup} e Q_2

FÓRMULAS

(Procedimento análogo ao da mediana)



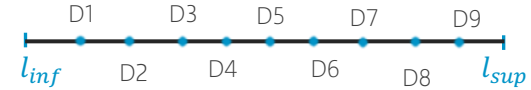
$$Q_1 = l_i + \left[\frac{1 \cdot \frac{n}{4} - fac_{ant}}{f_i} \right] \cdot h$$

$$Q_2 = l_i + \left[\frac{2 \cdot \frac{n}{4} - fac_{ant}}{f_i} \right] \cdot h$$

$$Q_3 = l_i + \left[\frac{3 \cdot \frac{n}{4} - fac_{ant}}{f_i} \right] \cdot h$$

DECIL

- Divide os dados em **10 partes** de mesma frequência
 São **9 decis** com **10% dos dados** cada



FÓRMULAS

$$D_K = l_i + \left[\frac{k \cdot \frac{n}{10} - fac_{ant}}{f_i} \right] \cdot h$$

PERCENTIL

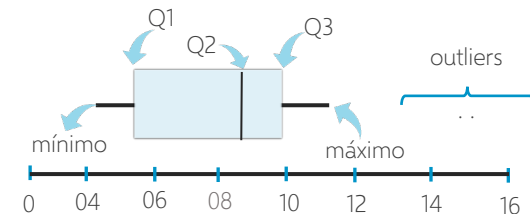
- Divide os dados em **100 partes** de mesma frequência
 (são **99 percentis** com **1% dos dados** cada)

FÓRMULAS

$$P_K = l_i + \left[\frac{k \cdot \frac{n}{100} - fac_{ant}}{f_i} \right] \cdot h$$

BOX PLOT

- Gráficos que usam os **quartis** para a **representação** de dados
- Pode ser **horizontal** ou **vertical**



medidas
separatrizes

ASPECTOS GERAIS

- = valor que aparece com **maior frequência**
- Um conjunto de valores pode ter **mais de uma moda**

MODA PARA DADOS NÃO-AGRUPADOS

- $X = \{1, 3, 9, 16, 20, 21, 21, 34\}$ = conjunto **amodal**
- $X = \{1, 3, 9, 16, 16, 16, 20, 21, 21, 34\}$ = conjunto **unimodal**
- $X = \{1, 3, 9, 16, 16, 20, 21, 21, 34\}$ = conjunto **bimodal**

MODA PARA DADOS AGRUPADOS

SEM INTERVALOS DE CLASSE

- A moda é aquele valor com f_i **maior** ↗ Frequência simples
- Ex.: notas de alunos em uma classe

NOTAS	FREQUÊNCIA (f_i)
2	2
4	4
6	10
8	12
10	9
TOTAL:	39 (n)

Moda = 8
(Nota 8!)

ATENÇÃO!

A moda não é a f_i ,
mas o valor em si

moda

PROPRIEDADES DA MODA

- A moda **não** é influenciada pelos **valores extremos** do rol (Depende apenas do número de vezes que cada valor aparece!)
- **Somando-se** (subtraindo-se) uma constante c de todos os valores → a moda também é somada ou subtraída de c

$$M_o' = M_o + c$$

$$M_o' = M_o - c$$

- **Multiplicando-se** (Dividindo-se) todos os valores por uma constante c → a moda também é multiplicada/dividida por c

$$M_o' = M_o \cdot c$$

$$M_o' = M_o : c$$

moda

MODA DE PEARSON



$$Mo = 3.Md - 2\bar{X}$$

moda

mediana

média

- Utilize apenas quando a questão pedir **expressamente**

MODA PARA DADOS AGRUPADOS EM CLASSES

MODA BRUTA

Classe com maior frequência

- É o **ponto médio** da **classe modal**

- Ex.:

ALTURA	FREQUÊNCIA (f _i)
40-50	2
50-60	5
60-70	7 <i>f_{ant}</i>
<i>l_i</i> 70-80	<i>f_M</i> 8 classe modal
80-90	3 <i>f_{post}</i>
TOTAL:	25 (n)

moda bruta = 75

se a classe modal for:
primeira: *f_{ant}* = 0
última: *f_{post}* = 0

<i>l_i</i>	Limite inferior
<i>h</i>	Amplitude de classe
<i>f_{ant}</i>	Frequência da classe anterior
<i>f_{post}</i>	Frequência da classe posterior
<i>f_M</i>	Frequência simples da classe modal



MODA PARA DADOS AGRUPADOS EM CLASSES

MODA DE CZUBER

$$M_o = l_i + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] \cdot h$$

$$\Delta_1 = f_M - f_{ant}$$

$$\Delta_2 = f_M - f_{post}$$

$$M_o = l_i + \left[\frac{f_M - f_{ant}}{(f_M - f_{ant}) + (f_M - f_{post})} \right] \cdot h$$

Alguns livros usam a fórmula assim

MODA DE KING

$$M_o = l_i + \left[\frac{f_{post}}{f_{ant} + f_{post}} \right] \cdot h$$

Essas fórmulas de Czuber e King só podem ser aplicadas se as **amplitudes** das classes (h) forem todas **iguais**



MEDIDAS de dispersão

ASPECTOS GERAIS

= Analisa o **afastamento** dos dados

• Exemplos:

- Amplitudes (Total, interquartílica)
- Desvios (Quartilício, médio, padrão...)
- Variância
- Coeficiente de Variação

AMPLITUDE TOTAL

$$A = x_{\max} - x_{\min}$$

= Diferença entre o **maior** e o **menor** valor de um conjunto de dados

PROPRIEDADES

- **Somando-se** (subtraindo-se) uma constante **c** de todos os valores $\rightarrow$ a amplitude **não** é alterada.
- **Multiplicando-se** (dividindo-se) todos os valores por uma constante **c** $\rightarrow$ a amplitude também é multiplicada (ou dividida) por **c**.

DESVIOS

= **Diferença** entre um número do conjunto (x_i) em relação a um número (m).

$$d_i = x_i - m$$

PROPRIEDADES

- Soma **algébrica** dos desvios **relação à média** ($\bar{x}$) é zero

$$\sum d_i = \sum (x_i - \bar{x}) = 0$$

- A soma dos **quadrados** dos desvios ($\sum d_i^2$) em relação a um número **m** é mínima quando **m = $\bar{x}$**
- A soma dos módulos dos desvios ($\sum |d_i|$) em relação a um número **m** é mínima quando **m = mediana**.

MEDIDAS de dispersão

DESVIO MÉDIO

= média dos módulos dos desvios dos termos (x_i) em relação à média ($\bar{x}$)

$$D_m = \frac{\sum(|x_i - \bar{x}|)}{n}$$

- Em **dados agrupados**, deve-se calcular a média **ponderada** dos desvios:

$$D_m = \frac{\sum(|x_i - \bar{x}| \cdot f_i)}{n}$$

Os pesos são as frequências (f_i)

VARIÂNCIA (Populacional)

= **média** aritmética dos **quadrados** dos **desvios**

$$\sigma^2 = \frac{\sum(|x_i - \bar{x}|)^2}{n}$$

- Ou $\sigma^2 = \overline{x^2} - (\bar{x})^2$

σ^2 = média dos quadrados – quadrado da média

DESVIO PADRÃO (Populacional)

= **raiz** quadrada da **variância**:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

- A variância e o desvio padrão serão **nulos** quando **todos os termos forem iguais**.

SIMBOLOGIA:

σ^2 : Variância populacional

S^2 : Variância amostral

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1} \rightarrow S^2 = [\overline{x^2} - (\bar{x})^2] \cdot \frac{n}{n - 1}$$

σ : Desvio padrão populacional

S : Desvio padrão amostral

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

MEDIDAS de dispersão

VARIÂNCIA E DESVIO PADRÃO

PROPRIEDADES

- **Somando-se** (subtraindo-se) uma constante **c** de todos os valores: a variância e o desvio padrão **não** se alteram
- **Multiplicando-se** (dividindo-se) todos os valores por uma constante **c**:
 - O **desvio padrão** também é multiplicado (dividido) por **c**
 - A **variância** é multiplicada (dividida) por **c²**

VARIÂNCIA E DESVIO PADRÃO

PARA DADOS AGRUPADOS

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{n}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{n - 1}$$

→ Deve-se multiplicar cada resultado pela respectiva frequência

COEFICIENTE DE VARIAÇÃO

$$C_v = \frac{\sigma}{\bar{x}}$$

= Razão entre o desvio padrão e a média (É adimensional)

- Dá uma **noção relativa** da dispersão dos dados: permite "ver" se o desvio padrão é relevante quando comparado à média dos dados
- **C_v² = variância relativa**

PRINCÍPIO FUNDAMENTAL DA CONTAGEM

- = princípio multiplicativo
- Em experimentos que ocorrem em várias etapas sucessivas e independentes
 - p_1 = nº de possibilidades na 1ª etapa
 - p_n = nº de possibilidades na n-ésima etapa
- **Número total** de formas de o acontecimento ocorrer:
 - = $p_1 \cdot p_2 \cdot \dots \cdot p_n$ (multiplicação)

PRINCÍPIO ADITIVO

- Em experimentos que podem ser realizados de p modos ou q modos
- **Número total** de formas de o acontecimento ocorrer = $p+q$ (soma)

PERMUTAÇÃO SIMPLES

- De quantas maneiras é possível **ordenar** n objetos distintos?
- = $n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1 = n!$

PERMUTAÇÃO COM REPETIÇÃO

- De quantas maneiras é possível **ordenar** n objetos, sendo alguns deles **repetidos**?
- **Exemplo:** Um termo r_x repetidos e um s_x repetidos:

$$= \frac{n!}{r! \cdot s!} \quad \left(\text{Como uma "correção" pela existência das repetições} \right)$$

ANÁLISE COMBINATÓRIA

COMBINAÇÃO SIMPLES

- De quantas maneiras podemos formar **subconjuntos** de p elementos a partir de um conjunto de n elementos? (A ordem dos elementos não importa)

$$C_{n_1p} = C_n^p = \binom{n}{p} = \frac{n!}{p! (n-p)!}$$

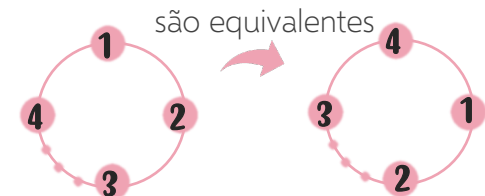
COMBINAÇÃO COM REPETIÇÃO

- De quantas maneiras podemos escolher p elementos a partir de um conjunto com n variedades? (Elementos de uma mesma variedade são considerados repetidos)

$$CR_n^p = \frac{(n+p-1)!}{p! (n-1)!}$$

PERMUTAÇÃO CIRCULAR

⚠ ATENÇÃO!



- Número total de permutações circulares de n objetos distintos = $(n-1)!$

PROBABILIDADE

CONCEITOS

ESPAÇO AMOSTRAL (U)

= Conjunto de **todos** os resultados possíveis

EVENTO

= Todo **subconjunto** do espaço amostral
 evento impossível = $\emptyset$ (Conjunto Vazio)

PROBABILIDADE

- Considera-se que cada elemento de U tem a **mesma chance** de ocorrer

$$P(A) = \frac{n(A)}{n(U)}$$

$n(A)$: nº de elementos do evento A
 $n(U)$: nº de elementos do espaço amostral

COMBINAÇÃO DE EVENTOS

UNIÃO

$A \cup B \rightarrow$ Ocorre se ocorrer A **ou** B **ou** ambos

INTERSECÇÃO

$(A \cap B) \rightarrow$ Ocorre se ocorrer A **e** B (ou seja, ambos)

COMPLEMENTAR

$\bar{A} \rightarrow$ Ocorre se **não** ocorrer A

PROPRIEDADES

- Se evento = espaço amostral $\rightarrow$ evento é **certo**
- Se evento = $\emptyset \rightarrow$ evento é **impossível**

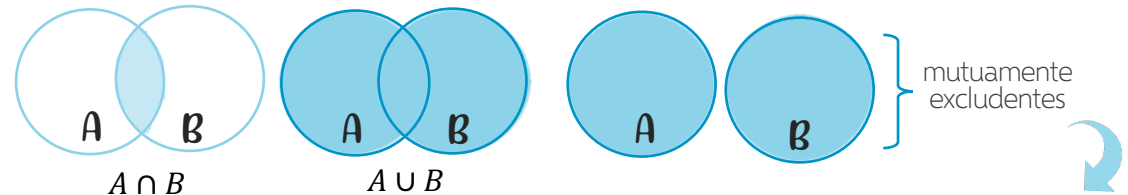
$$0 \leq P(A) \leq 1 \quad (P(U) = 1)$$

$$P(A) + P(\bar{A}) = 1$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$(se A \cap B = \emptyset, P(A \cup B) = P(A) + P(B))$$

REPRESENTAÇÃO GRÁFICA



CASOS ESPECIAIS

$A \cup B =$ Espaço amostral $\rightarrow$ eventos **exaustivos**

$A \cap B = \emptyset \rightarrow$ eventos mutuamente excludentes/exclusivos

ASPECTOS GERAIS

= Probabilidade de um evento B ocorrer **dados** que o evento A ocorreu

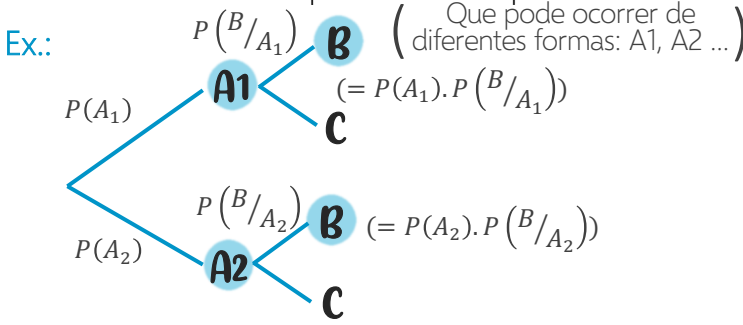
$$P(B/A) = \frac{n(A \cap B)}{n(A)}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$n(A \cap B)$: nº de elementos da interseção entre A e B
 $n(A)$: nº de elementos do evento A

TEOREMA DA PROBABILIDADE TOTAL

• Para descobrir a **probabilidade total** de o evento B ocorrer no caso em que B ocorre após o evento A.



$$P(B) = P(A_1).P(B/A_1) + P(A_2).P(B/A_2)$$

probabilidade
= CONDICIONAL =

TEOREMA DA MULTIPLICAÇÃO

$$P(A \cap B) = P(A).P(B/A)$$

• Se os eventos forem **independentes**:

$$P(A \cap B) = P(A).P(B)$$

CAI MUITO!

• Eventos **independentes**: = a ocorrência do evento A **não** influi na ocorrência de B

TEOREMA DE BAYES DECORE!

• Para descobrir a **probabilidade** de ocorrer A1, dados que B ocorreu

$$P(A_1/B) = \frac{P(A_1 \cap B)}{P(B)}$$

(Se decorar o teorema, vai agilizar as resoluções!)

$$P(A_1/B) = \frac{P(A_1).P(B/A_1)}{P(A_1).P(B/A_1) + P(A_2).P(B/A_2)}$$

VARIÁVEIS ALEATÓRIAS

CONCEITO

- = variável associada a uma **distribuição de probabilidade**
 - podem assumir diferentes valores

MODA

- = valor com **maior** probabilidade

ESPERANÇA MATEMÁTICA

$$\mu = E(x) = \sum_{i=1}^n X_i \cdot P(x_i)$$

1. Multiplique cada valor por sua probabilidade
2. Some tudo

PROPRIEDADES

- $E(k \cdot x) = k \cdot E(x)$
- $E(x + k) = E(x) + k$
- $E(x + y) = E(x) + E(y)$
- $E(k) = k$



Mas cuidado: se $E(x \cdot y) = E(x) \cdot E(y)$,
não necessariamente x e y são
variáveis independentes

Se X e Y forem **variáveis independentes**:

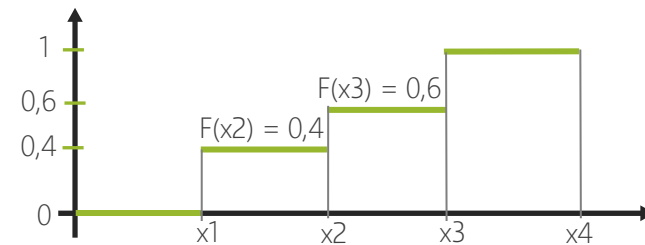
- $E(x \cdot y) = E(x) \cdot E(y)$

FUNÇÃO DE DISTRIBUIÇÃO

$$F(x) = P(X \leq x)$$

(= Probabilidade de a variável aleatória
assumir valores $\leq$ valor em questão)

- Seu gráfico é uma **função escada**:



MEDIANA

- = valor de x em que a função de distribuição **ultrapassa 50%** (0,5) pela primeira vez

VARIÂNCIA E DESVIO PADRÃO

$$\sigma^2 = E(X - E(X))^2$$

Medida do grau de dispersão da
distribuição em torno da média

ou

$$\sigma^2 = E(X^2) - E(X)^2$$

$$\sigma = \sqrt{\sigma^2}$$

Lembrando que
 $E(x) = \mu$

COVARIÂNCIA (COV [X,Y]) E CORRELAÇÃO ($\rho[X,Y]$)

$$COV(x, y) = E(x - \mu_x) \cdot (y - \mu_y)$$

Se y **aumenta** quando x **aumenta**, $Cov(x,y) > 0$

Se y **diminui** quando x **aumenta**, $Cov(x,y) < 0$

• Pode assumir **qualquer** valor real

ou
$$Cov(x, y) = E(x \cdot y) - E(x) \cdot E(y)$$

"esperança do produto – produto das esperanças"

$$\rho(x, y) = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$$

Assume valores em $[-1, 1]$

- $\rho(x,y) = 1$ correlação linear perfeita positiva
- $\rho(x,y) = -1$ correlação linear perfeita negativa

• Se x e y são variáveis **independentes**:

- $\rho(x,y) = 0$
- $Cov(x,y) = 0$

Mas cuidado: se $\rho(x,y) = 0$ ou $Cov(x,y) = 0$, não necessariamente x e y são variáveis independentes

PROPRIEDADES

- $Cov(x, x) = Var(x)$
- $Cov(k, x) = 0$
- $Cov(k, x, y) = k \cdot Cov(x, y)$
- $Cov(x+y, z) = Cov(x, z) + Cov(y, z)$

VARIÁVEIS ALEATÓRIAS

VARIÂNCIA DA SOMA E DA DIFERENÇA

$$V(x + y) = V(x) + V(y) + 2 \cdot Cov(x, y)$$

$$V(x - y) = V(x) + V(y) - 2 \cdot Cov(x, y)$$

$$V(ax + by) = a^2 \cdot V(x) + b^2 \cdot V(y) + 2 \cdot a \cdot b \cdot Cov(x, y)$$

$$V(ax - by) = a^2 \cdot V(x) + b^2 \cdot V(y) - 2 \cdot a \cdot b \cdot Cov(x, y)$$

COEFICIENTE DE VARIAÇÃO

 CAI MUITO!

$$C_v = \frac{\sigma}{\mu} = \frac{\text{desvio padrão}}{\text{média}}$$

VARIÂNCIA RELATIVA

$$V_R = C_v^2 = \frac{\sigma^2}{\mu^2}$$

CONCEITOS

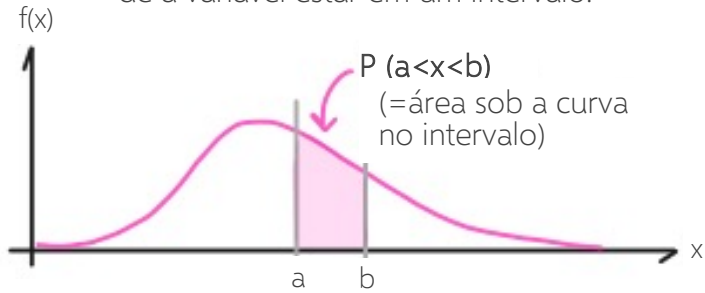
= variável que pode assumir **qualquer valor** dentro de um intervalo

• Deve existir uma **função f** tal que

- $f(x) \geq 0$, para todo x real
- $\int_{-\infty}^{\infty} f(x)dx = 1$ → = função densidade de probabilidade

• A probabilidade de a variável assumir um **valor específico** é nula!

→ Nos interessamos nas probabilidades de a variável estar em um intervalo:



$$P(a < x < b) = \int_a^b f(x)dx$$

FUNÇÃO DE DISTRIBUIÇÃO (F(x))

(ou função de distribuição acumulada)

$$F(x) = P(X \leq x)$$

$$F(k) = \int_{-\infty}^k f(x)dx$$

$$\frac{dF(x)}{dx} = f(x)$$

VARIÁVEIS CONTÍNUAS

MODA

- = valor **x** que maximiza $f(x)$

MEDIANA

- = número **m** tal que:

$$\int_{-\infty}^m f(x)dx = 50\%$$

MÉDIA

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

VARIÂNCIA

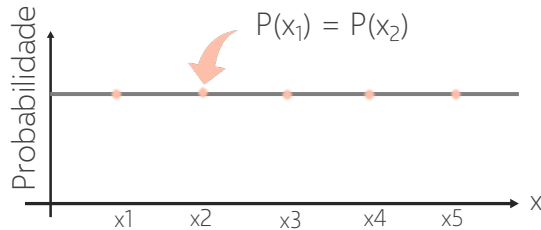
$$VAR(x) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx$$

ou $VAR(x) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \mu^2$

DISTRIBUIÇÃO UNIFORME DISCRETA

= Todos os elementos têm a **mesma probabilidade** de ocorrer

$$E(x) = \frac{\sum x_i}{n}$$



DISTRIBUIÇÃO DE BERNOULLI

= Há **dois resultados** possíveis (Sucesso e fracasso) em um **único** experimento

$$\begin{cases} P(\text{sucesso}) = p \\ P(\text{fracasso}) = q = (1-p) \end{cases} \quad \sigma^2 = p \cdot q$$

$$E(x) = p$$

(Esperança)

$$\sigma^2 = p(1-p)$$

(Variância)

DISTRIBUIÇÃO GEOMÉTRICA

= há **dois resultados** possíveis (Sucesso e fracasso) e deseja-se saber a probabilidade de sucesso exatamente no **k-ésimo ensaio** (k-1) fracassos → 1 sucesso

$$P(x = k) = \underbrace{q \cdot q \cdot \dots \cdot q}_{(k-1) \text{ fracassos}} \cdot p \quad \text{1 sucesso}$$

$$P(x = k) = q^{(k-1)} \cdot p$$

$$E(x) = \frac{1}{p}$$

(Esperança)

$$\sigma^2 = \frac{q}{p^2}$$

(Variância)

DISTRIBUIÇÕES DISCRETAS = DE PROBABILIDADE =

DISTRIBUIÇÃO BINOMIAL

= Há **dois resultados** possíveis (Sucesso e fracasso), e o experimento é realizado **várias vezes** (n vezes). O resultado de um ensaio não interfere no outro (São independentes)

$$P(x = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$E(x) = n \cdot p$$

(Esperança)

$$\sigma^2 = n \cdot p \cdot q$$

(Variância)

DISTRIBUIÇÕES discretas = DE PROBABILIDADE =

DISTRIBUIÇÃO HIPERGEOMÉTRICA

= probabilidade de, ao retirarmos, **sem reposição**, n elementos de um conjunto de N elementos, saíam k sucessos, de S sucessos presentes no conjunto.
(e $n-s$ fracassos)

$$P(x = k) = \frac{\binom{S}{k} \cdot \binom{N-S}{n-k}}{\binom{N}{n}}$$

$$E(x) = n \cdot p$$

(Esperança) $p = \frac{S}{N}$ e $q = 1 - p$

$$\sigma^2 = n \cdot p \cdot q \cdot \frac{N-n}{N-1}$$

(Variância)

DISTRIBUIÇÃO DE POISSON



= distribuição **binomial**, em que p é muito **pequeno** e n muito **grande**

$$\left(\begin{matrix} n \rightarrow \infty \\ p \rightarrow 0 \end{matrix} \right)$$

$$\sigma^2 = n \cdot p \cdot q \rightarrow \left(\begin{matrix} p \rightarrow 0 \\ q \rightarrow 1 \end{matrix} \right)$$

$$E(x) = \lambda = n \cdot p$$

(Esperança)

$$E(x) = \sigma^2 = \lambda$$

$$\sigma^2 = \lambda$$

(Variância)

$$P(x = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

e = número de Euler = 2,718...

DISTRIBUIÇÃO UNIFORME CONTÍNUA

$$f(x) = \begin{cases} k, & \text{se } a \leq x \leq b \\ 0, & \text{caso contrário} \end{cases}$$

Função distribuição de probabilidade

PROBABILIDADE



$$E(x) = \frac{a+b}{2}$$

(Esperança)

$$\sigma^2 = \frac{(b-a)^2}{12}$$

(Variância)

DISTRIBUIÇÃO EXPONENCIAL

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & \text{se } x \geq 0 \ (\lambda > 0) \\ 0, & \text{se } x < 0 \end{cases}$$

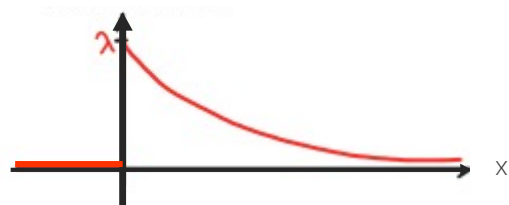
$$E(x) = \frac{1}{\lambda}$$

(Esperança)

$$\sigma^2 = \frac{1}{\lambda^2}$$

(Variância)

PROBABILIDADE



distribuições contínuas = DE PROBABILIDADE =

DISTRIBUIÇÃO QUI-QUADRADO (χ^2)

= soma dos quadrados de **k variáveis** normais reduzidas e independentes
(Importante para o estudo de intervalos de confiança e testes de hipóteses)

$$\chi_k^2 = \sum_{i=1}^k Z_i^2$$

k graus de liberdade

$$E(x) = k$$

(Esperança)

$$\sigma^2 = 2 \cdot k$$

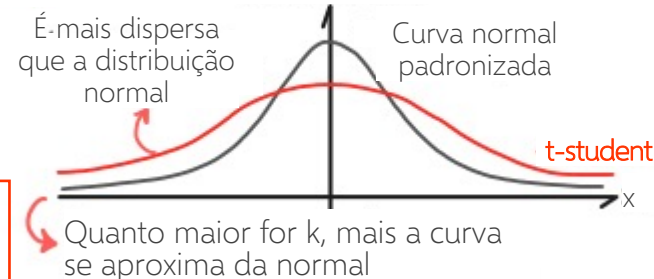
(Variância)

DISTRIBUIÇÃO t DE STUDENT

$$t = \frac{Z}{\sqrt{\frac{\chi^2}{K}}}$$

Z: Variável normal padrão
 χ^2 : Qui-quadrado
k: graus de liberdade

PROBABILIDADE



Se $k > 1$:

$$E(x) = 0$$

(Esperança)

Se $k > 2$:

$$\sigma^2 = \frac{k}{k-2}$$

(Variância)

DISTRIBUIÇÃO F DE SNEDECOR

$$F = \frac{\chi_1^2 / k_1}{\chi_2^2 / k_2}$$

Relaciona duas variáveis aleatórias independentes com distribuição qui-quadrado

$$E(x) = \frac{k_2}{k_2 - 2}$$

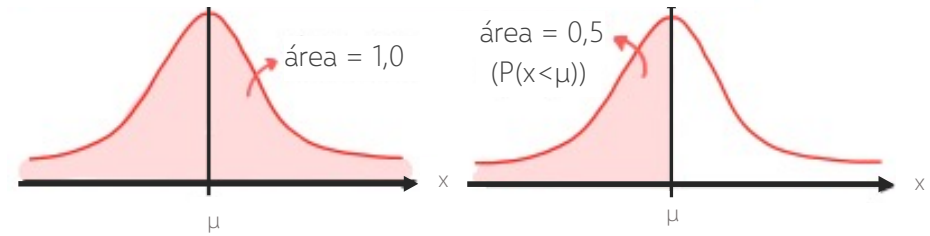
(Esperança)

$$\sigma^2 = \frac{2k_2^2 \cdot (k_1 + k_2 - 2)}{k_1 \cdot (k_2 - 2)^2 \cdot (k_2 - 4)}$$

(Variância)

DISTRIBUIÇÕES CONTÍNUAS = DISTRIBUIÇÃO NORMAL =

PROPRIEDADES DA CURVA NORMAL



DISTRIBUIÇÃO NORMAL PADRÃO $N(0,1)$

- **Transformar** uma normal qualquer em uma normal padrão:

$$Z = \frac{X - \mu}{\sigma}$$

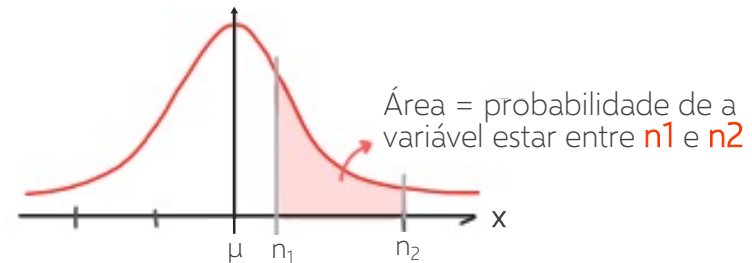
Número a transformar

CAI MUITO!

= Número correspondente na distribuição padrão

- É a distribuição usada na tabela para consultas.

CÁLCULO DA PROBABILIDADE DE INTERVALOS



PASSO A PASSO

- Transforme os valores nos correspondentes aos da normal padrão
- Consultar: $P_1(z < n_1)$ e $P_2(z < n_2)$

$$P(n_1 < x < n_2) = P_2(z < n_2) - P_1(z < n_1)$$

ASPECTOS GERAIS

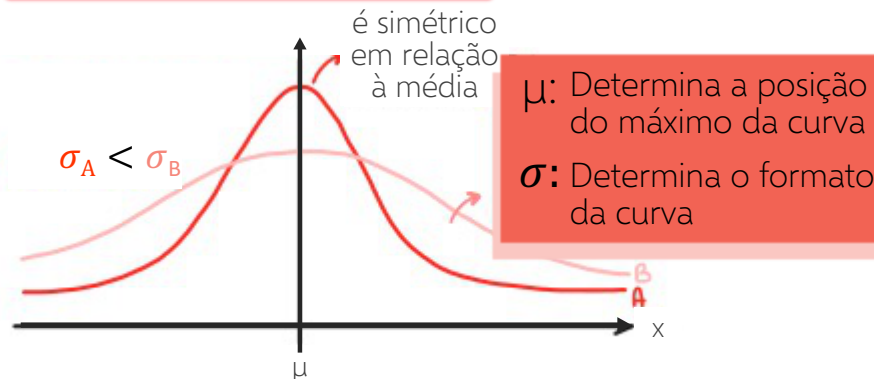
- = só depende da **média** (μ) e **desvio padrão** (σ)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2 \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

- = $N(\mu, \sigma^2)$ (Muitas bancas só colocam isso, você tem que saber o que é cada termo!) **ATENÇÃO!**

MÉDIA = MEDIANA = MODA **DECORE!**

REPRESENTAÇÃO GRÁFICA



ASPECTOS GERAIS

- = seleção de uma **amostra** a partir de uma população

AMOSTRAGEM ALEATÓRIA SIMPLES

- = todos os elementos têm a **mesma** probabilidade de serem selecionados.
- Casos:
 - População **infinita** ou **com reposição**
 - ↪ os valores observados são independentes
 - População **finita** ou **sem reposição**:
 - ↪ os valores observados **não** são independentes

AMOSTRAGEM POR ESTRATIFICAÇÃO

- = divide-se a população em **estratos** e realiza-se uma amostragem aleatória simples em cada um
- Os estratos são **homogêneos** (Têm baixa variabilidade), mas entre os estratos há uma grande **heterogeneidade**.

AMOSTRAGEM POR CONGLOMERADOS

- = divide-se a população em subconjuntos, (Com uma **baixa variabilidade** entre eles, mas **alta variabilidade** dentro de cada subconjunto)
- Sorteia-se um dos subconjuntos e analisa-se

AMOSTRAGEM SISTEMÁTICA

- = ordenam-se os dados → selecionam-se elementos **igualmente espaçados**
- **Ex.:** de 7 em 7, de 10 em 10 ...

AMOSTRAGEM POR CONVENIÊNCIA

- = **não** há muito critério na seleção dos elementos
- Amostragem **não probabilística**
- **Ex.:** entrevistar as pessoas que passarem na rua

AMOSTRAGEM POR JULGAMENTO

- = escolha dos elementos conforme o julgamento do responsável
- Amostragem **não probabilística**

AMOSTRAGEM POR COTAS

- = selecionam-se os elementos de forma **proporcional** às características semelhantes às da população
- Amostragem **não probabilística**
- **Ex.:** se a população é de 45% homens e 55% mulheres, refletem-se essas proporções na amostra.

amostragem

ESTIMADORES



ASPECTOS GERAIS

- **Parâmetro (θ)**: medida que descreve alguma característica numérica de uma população.
 é sempre uma constante (Não varia!)
- **Estimador ($\hat{\theta}$)**: característica numérica determinada na amostra, é uma função matemática de seus elementos.
 O valor do estimador varia com cada amostra: é uma variável aleatória com distribuição igual à da população
- **Erro amostral (ε)**: $\varepsilon = \hat{\theta} - \theta$

SIMBOLOGIA



GRANDEZA	PARÂMETRO	ESTIMADOR
genérico	θ	$\hat{\theta}$
média	μ	$\bar{x}$
variância	σ^2	S^2
desvio padrão	σ	S
coeficiente de correlação	ρ	r

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

- População **infinita** ou amostragem **com reposição**
 (Os valores da amostra são considerados valores de variáveis aleatórias independentes)

média amostral

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

elementos da amostra
 (São iid: independentes e identicamente distribuídos)

n = número de elementos da amostra

$$E(\bar{X}) = \mu$$

$$S_X^2 = \frac{\sigma^2}{n}$$

FATOR DE CORREÇÃO PARA POPULAÇÕES FINITAS

- As variáveis $x_1, \dots, x_n$ serão **dependentes**

$$S_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1}$$

N = Número de elementos da população

MÉDIA AMOSTRAL E DISTRIBUIÇÃO NORMAL

- Se a **população** tiver uma distribuição **normal**, a distribuição amostral de $\bar{x}$ também será **normal** (para qualquer tamanho de amostra)

CARACTERÍSTICAS DA MÉDIA AMOSTRAL

- $\bar{x}$ é um estimador:
 - Não viesado ($E = (\bar{\theta}) - \theta$)
 - De variância mínima
 - De mínimos quadrados (Minimiza a soma dos quadrados dos desvios)
 - De máxima verossimilhança
 - Consistente (Sua variância tende a zero quando n tende ao infinito)

ESTIMADORES



DISTRIBUIÇÃO AMOSTRAL DA VARIÂNCIA

- Estimador não **viesado** da variância: $(E(S^2) = \sigma^2)$

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

$$VAR(S^2) = \frac{2\sigma^4}{n - 1}$$

→ $\left(\frac{n-1}{\sigma^2}\right) \cdot S^2 = \chi^2_{n-1}$ (= distribuição qui-quadrado com n-1 graus de liberdade)

- Se X tem distribuição **normal**, o estimador de **máxima verossimilhança**:

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

→ Mas é um estimador **tendencioso**

DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

- Proporção **p** de **sucessos** da população

📋 CAI MUITO!

$$\begin{cases} 1, \text{ sucesso} \rightarrow P(x=1)=p \\ 0, \text{ fracasso} \rightarrow P(x=0)=q \end{cases}$$

→ $p+q=1$ (A população segue uma distribuição de bernoulli)

$$E(X) = p$$

$$VAR(X) = p(1 - p)$$

$$(\sigma^2 = p \cdot q)$$

X= variável aleatória que representa o número de **sucessos** em **n ensaios** independentes:

→ X = distribuição binomial

$$E(X) = n \cdot p$$

$$VAR(X) = n \cdot p \cdot q$$

$$\hat{p} = \frac{X}{n} = \text{Proporção de sucessos em uma amostra de n elementos}$$

→ $E(\hat{p}) = p$ → É um estimador:

- Não viesado ($E(\hat{\theta}) = \theta$)
- De mínimos quadrados
- De máxima verossimilhança

$VAR(\hat{p}) = \frac{p \cdot q}{n}$

FATOR DE CORREÇÃO PARA POPULAÇÕES FINITAS

$$VAR(\hat{p}) = \frac{p \cdot q}{n} \cdot \frac{N-n}{N-1}$$

N= número de elementos da população

💡 DECORE!

Intervalos de Confiança

μ = Média populacional

σ^2 = Variância populacional

são constantes!
(não são variáveis aleatórias)

PARA μ QUANDO O σ^2 É CONHECIDO

INTERVALO DE CONFIANÇA

$$= \left[(\bar{X} - Z_o \cdot \sigma_{\bar{X}}), \bar{X}, (\bar{X} + Z_o \cdot \sigma_{\bar{X}}) \right]$$

$\bar{X}$: valor específico obtido para a média amostral

Z_o : valor da distribuição normal padrão associada à confiança pedida no intervalo

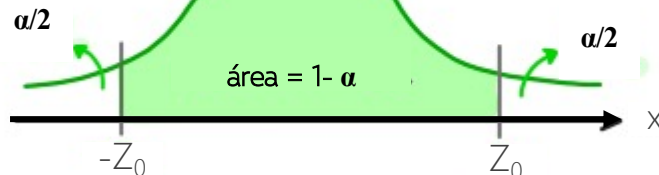
$\sigma_{\bar{X}}$: desvio padrão da média amostral

• Amplitude do intervalo $\rightarrow A = 2 \cdot Z_o \cdot \sigma_{\bar{X}}$

$$\text{erro máximo} = \frac{A}{2} = Z_o \cdot \sigma_{\bar{X}}$$

= Erro máximo cometido na estimativa de μ

= intervalo de confiança



PARA μ QUANDO σ^2 É DESCONHECIDO

INTERVALO DE CONFIANÇA

$$= \left[(\bar{X} - t_o \cdot \sigma_{\bar{X}}), \bar{X}, (\bar{X} + t_o \cdot \sigma_{\bar{X}}) \right]$$

• Usamos a distribuição **t de student**:

No lugar de Z, usamos t e consultamos a tabela t de student

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}}$$

n-1 graus de liberdade

$$\begin{cases} S_{\bar{X}}^2 = \frac{\sigma^2}{n} & \text{(População infinita ou amostragem com reposição)} \\ S_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} & \text{(População finita ou amostragem sem reposição)} \end{cases}$$

• Amplitude do intervalo $\rightarrow A = 2 \cdot t_o \cdot \sigma_{\bar{X}}$

$$\text{erro máximo} = \frac{A}{2} = t_o \cdot \sigma_{\bar{X}}$$

Intervalos de Confiança

PARA UMA PROPORÇÃO

$$= \hat{p} - Z_o \cdot \sigma_{\hat{p}} \leq p \leq \hat{p} + Z_o \cdot \sigma_{\hat{p}}$$

$$\hat{p} - Z_o \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq p \leq \hat{p} + Z_o \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

→ Sendo p desconhecida, também não sabemos $\sigma_{\hat{p}}$, usamos: $S_{\hat{p}} = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$

• Amplitude do intervalo → $A = 2 \cdot Z_o \cdot S_{\hat{p}}$

$$\text{erro máximo} = \frac{A}{2} = Z_o \cdot S_{\hat{p}}$$

• Usamos a distribuição normal

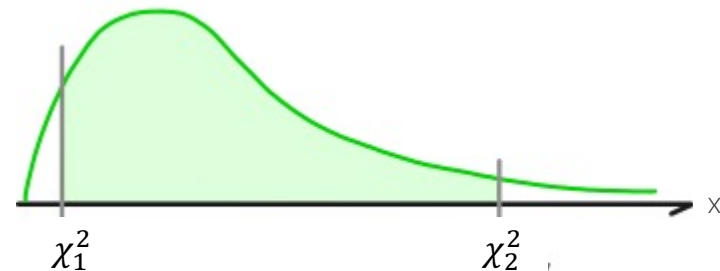
PARA A VARIÂNCIA

📢 **IMPORTANTE!**

• Usamos a distribuição **qui-quadrado** (é uma distribuição assimétrica)

$$\chi^2_{n-1} = \left(\frac{n-1}{\sigma^2} \right) \cdot S^2 \rightarrow n-1 \text{ graus de liberdade}$$

$$= \left(\frac{n-1}{\chi^2_2} \right) \cdot S^2 \leq \sigma^2 \leq \left(\frac{n-1}{\chi^2_1} \right) \cdot S^2$$



ASPECTOS GERAIS

= método para realização de **inferências**

→ Aceitaremos ou rejeitaremos H_0 com um determinado grau de **risco**

HIPÓTESES:

H_0 : Hipótese nula

H_1 : Hipótese alternativa

Funcionamento:

1. começamos com um valor suposto hipotético para o parâmetro populacional
2. coleta-se uma amostra aleatória
3. A partir de resultado obtido, decide-se se aceitamos ou não H_0

RESULTADO	H_0
Região de não rejeição (RNR)	Aceitamos H_0
Região crítica (RC)	Rejeitamos H_0

→ Resultado da amostra foi discrepante

TESTES DE HIPÓTESES

TIPO DE TESTES



$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ → teste bilateral (ou bicaudal)

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ → teste unilateral à direita (ou monocaudal)

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ → teste unilateral à esquerda (ou monocaudal)

TIPOS DE ERROS



ERRO TIPO I = **rejeitar** H_0 , quando for **verdadeira**

- $P(\text{erro tipo I}) = \alpha$ (Nível de significância do teste)
- Nível de confiança: $1 - \alpha$

ERRO TIPO II = **aceitar** H_0 , quando for **falsa**

- $P(\text{erro tipo II}) = \beta$
- Poder do teste: $1 - \beta$ (Probabilidade de rejeitar H_0 , quando for falsa)



Não há relação entre α e β .

NÍVEL DE SIGNIFICÂNCIA (α)

- Área da região crítica = nível de significância
Taxa tolerável de erro →
- Ex.: nível de significância = 10%



TESTES DE HIPÓTESES PARA A MÉDIA

POPULAÇÃO NORMAL COM σ^2 CONHECIDO

- Passo a passo:
 - Determinar os **valores críticos** de **z**, de modo que a área da região crítica = nível de significância (α)
 - Calcular a estatística de teste padronizada

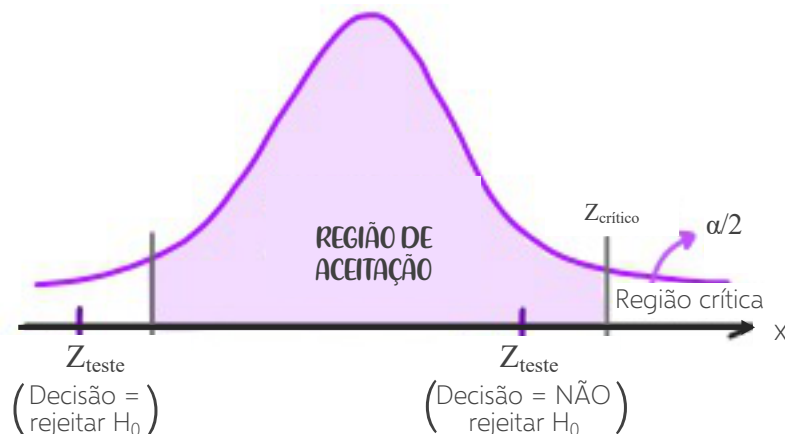
Média observada

$$Z_{teste} = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

Média populacional

$$Z_{teste} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- Verificamos onde Z_{teste} cai:



TESTES DE HIPÓTESES

POPULAÇÃO NORMAL COM σ^2 DESCONHECIDO

- Procedimento análogo ao anterior, mas usamos a distribuição **t de student**:

$$t_{teste} = \frac{\bar{X} - \mu}{S_{\bar{X}}}$$

$$\begin{cases} S_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(\begin{array}{l} \text{População infinita ou} \\ \text{amostragem com reposição} \end{array} \right) & n-1 \text{ graus de liberdade} \\ S_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \left(\begin{array}{l} \text{População finita ou} \\ \text{amostragem sem reposição} \end{array} \right) \end{cases}$$

P-VALOR

- = área delimitada pela estatística de teste  **DECORE!**
- Probabilidade de, sendo H_0 verdadeira, a variável reduzida ser maior/igual que a estatística de teste

HIPÓTESE	H_0
p-valor > α	Aceitamos H_0
p-valor < α	Rejeitamos H_0 (Estatística de teste caiu na região crítica)

TESTES DE HIPÓTESES

TESTES DE HIPÓTESES PARA PROPORÇÕES

- Uso da **distribuição binomial**
(Para n suficientemente grande, x será aproximadamente normal, assim como p)
- Procedimento análogo aos anteriores com a seguinte **estatística de teste**:

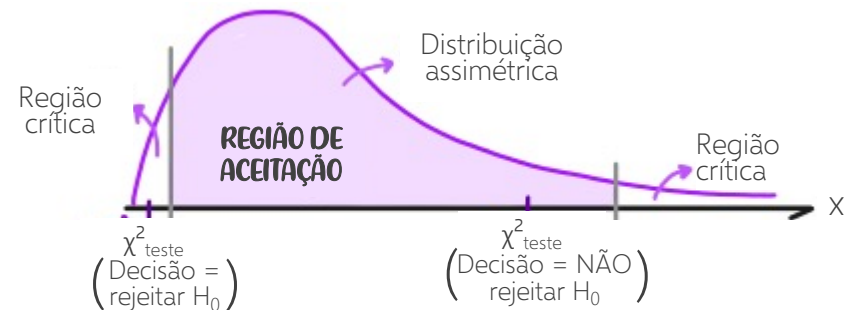
$$Z_{teste} = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

TESTE PARA VARIÂNCIA

- Usamos a distribuição **qui-quadrado**
- Procedimento análogo aos anteriores com a seguinte **estatística de teste**:

$$\chi^2_{teste} = \left(\frac{n-1}{\sigma^2} \right) \cdot S^2$$

↗ n-1 graus de liberdade



ANÁLISE DE VARIÂNCIA

ASPECTOS GERAIS

- Visa testar a hipótese de que as **médias** de k populações distintas são **iguais**:

$$\begin{cases} H_0: \mu_0 = \mu_1 = \dots = \mu_k \\ H_1: \text{pelo menos uma delas é diferente das demais} \end{cases}$$

(não especifica qual)

- Seleccionam-se **k amostras** independentes de cada uma das **populações**

$$\begin{cases} n_1: \text{elementos da primeira população, ... ,} \\ n_k: \text{elementos da k-ésima população} \end{cases}$$

$$N = \sum_{i=1}^k n_i$$

→ Total de elementos das amostras

SUPOSIÇÕES NECESSÁRIAS

- As populações devem ter distribuição **normal** + as **variâncias** das populações devem ter a **mesma variância** (homocedascia)

VALORES OBSERVADOS

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

→ Erro aleatório (tem distribuição normal)

i: população
j: elemento

SUPOSIÇÕES PARA O ERRO ALEATÓRIO

- Média dos erros aleatórios em cada grupo é 0
- E_{ij} são independentes
- A variância de E_{ij} é a mesma em todos os grupos (homocedascia)

SOMA DOS QUADRADOS

$$SQ_{total} = SQ_{dentro} + SQ_{entre}$$

SQ_{dentro} : soma dos quadrados dos erros/resíduos
 SQ_{entre} : soma dos quadrados dos tratamentos
 SQ_{total} : soma dos quadrados total (Entre os grupos)

- As variáveis $\frac{SQ_{dentro}}{\sigma^2}$, $\frac{SQ_{entre}}{\sigma^2}$ e $\frac{SQ_{total}}{\sigma^2}$ têm distribuição **qui-quadrado**.

ANÁLISE DE VARIÂNCIA

GRAUS DE LIBERDADE

$$gl_{total} = gl_{dentro} + gl_{entre}$$

$$gl_{total} = N - 1$$

$$gl_{entre} = k - 1$$

$$gl_{dentro} = N - k$$

QUADRADO MÉDIO

= quociente entre a soma dos quadrados e o respectivo número de graus de liberdade:

$$QM_{dentro} = \frac{SQ_{dentro}}{N - k}$$

$$QM_{entre} = \frac{SQ_{entre}}{k - 1}$$

$$QM_{total} = \frac{SQ_{total}}{N - 1}$$

TESTE F

F DE SNEDECOR

$$F_{k_1, k_2} = \frac{\chi_{k_1}^2 / k_1}{\chi_{k_2}^2 / k_2}$$

$\chi_{k_1}^2$ e $\chi_{k_2}^2$ são variáveis aleatórias com distribuição de qui-quadrado, com k_1 e k_2 graus de liberdade

$$F = \frac{SQ_{entre} / k - 1}{SQ_{dentro} / N - k}$$

ou

$$F_{teste} = \frac{QM_{entre}}{QM_{dentro}}$$

utilizada para testar

$$\begin{cases} H_0: \mu_0 = \mu_1 = \dots = \mu_k \\ H_1: \text{pelo menos uma delas} \\ \text{é diferente das demais} \end{cases}$$

HIPÓTESE	H ₀
F _{teste} > F _{crítico}	Rejeitamos H ₀
F _{teste} < F _{crítico}	Aceitamos H ₀

REGRESSÃO LINEAR

ASPECTOS GERAIS

= calcular a **expressão matemática** que relaciona duas variáveis x e y:

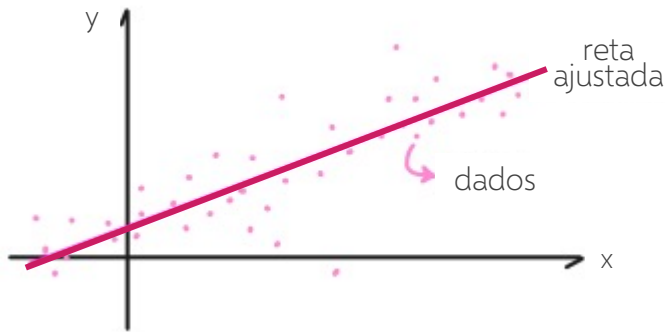
$$Y = p + m \cdot x$$

p = coeficiente linear

m = coeficiente angular

$$m = \frac{\Delta y}{\Delta x}$$

Exemplo:



COEFICIENTE DE CORRELAÇÃO LINEAR DE PEARSON

$$r = \frac{\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum[(X_i - \bar{X})^2 \cdot (Y_i - \bar{Y})^2]}}$$

Mede a força da relação linear entre 2 variáveis x e y

$-1 \leq r \leq 1$:

- Mais próximo de -1 ou 1 → Correlação mais forte
- Mais próximo de zero → Correlação mais fraca (não há relação linear entre as variáveis)

Em questões, é útil usar:

$$\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})] = \sum(X_i \cdot Y_i) - n \cdot \bar{X} \cdot \bar{Y}$$

MODELO ESTATÍSTICO

$$Y_i = \alpha + \beta X_i + \mu_i$$

x = variável independente

y = variável dependente

n = número de pares de valores observados

μ_i = erro ou desvio

$\alpha + \beta X_i$ = componente de y que varia linearmente com x

RETA DE REGRESSÃO ESTIMADA

$$\hat{Y} = a + bX_i$$

$$e = Y_i - \hat{Y}_i \rightarrow \text{desvio}$$

MÉTODO DOS MÍNIMOS QUADRADOS

= determina as estimativas de a e b dos parâmetros **minimizando a soma dos quadrados** dos desvios

Após calcular b, basta substituir na equação $\hat{Y} = a + b\bar{X}$ para determinar a

$$b = \frac{\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sum(X_i - \bar{X})^2}$$

RETA QUE PASSA PELA ORIGEM

• $\alpha = 0$

$$Y_i = \beta X_i + \varepsilon_i$$

$$\beta = \frac{\sum X \cdot Y}{\sum X^2}$$

ANÁLISE DA VARIÂNCIA DA REGRESSÃO

regressão linear

$$SQT = SQM + SQR$$

Soma dos quadrados total

Soma dos quadrados do modelo

Soma dos quadrados dos resíduos

FÓRMULAS IMPORTANTES

$$SQM = b \cdot \sum [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]$$

Soma dos quadrados do modelo de regressão

ou

$$SQM = b^2 \cdot \sum (X_i - \bar{X})^2$$

$$SQR = \sum (Y_i - \bar{Y})^2$$

É a parte do desvio total não explicada pelo modelo de regressão

TABELA DA ANOVA



FONTE DE VARIAÇÃO	GRAUS DE LIBERDADE (gl)	SOMA DE QUADRADOS	QUADRADOS MÉDIOS	F
MODELO	1	SQM	$QMM = \frac{SQM}{1}$	$F_{teste} = \frac{SMM}{QMR}$
RESÍDUOS	n-2	SQR	$QMR = \frac{SQR}{n-2}$	
TOTAL	n-1	SQT	$QMT = \frac{SQT}{n-1}$	

n = tamanho da amostra

(QMR = Estimativa de variância σ^2 residual)

COEFICIENTES



- Coeficiente de **correlação**:

$$R = \sqrt{\frac{SQM}{SQT}}$$

- Coeficiente de **determinação**:

Exprime a proporção da variação total de y que é explicada pela reta de regressão

$$R^2 = 1 - \frac{SQR}{SQT}$$

$$(0 \leq R^2 \leq 1)$$