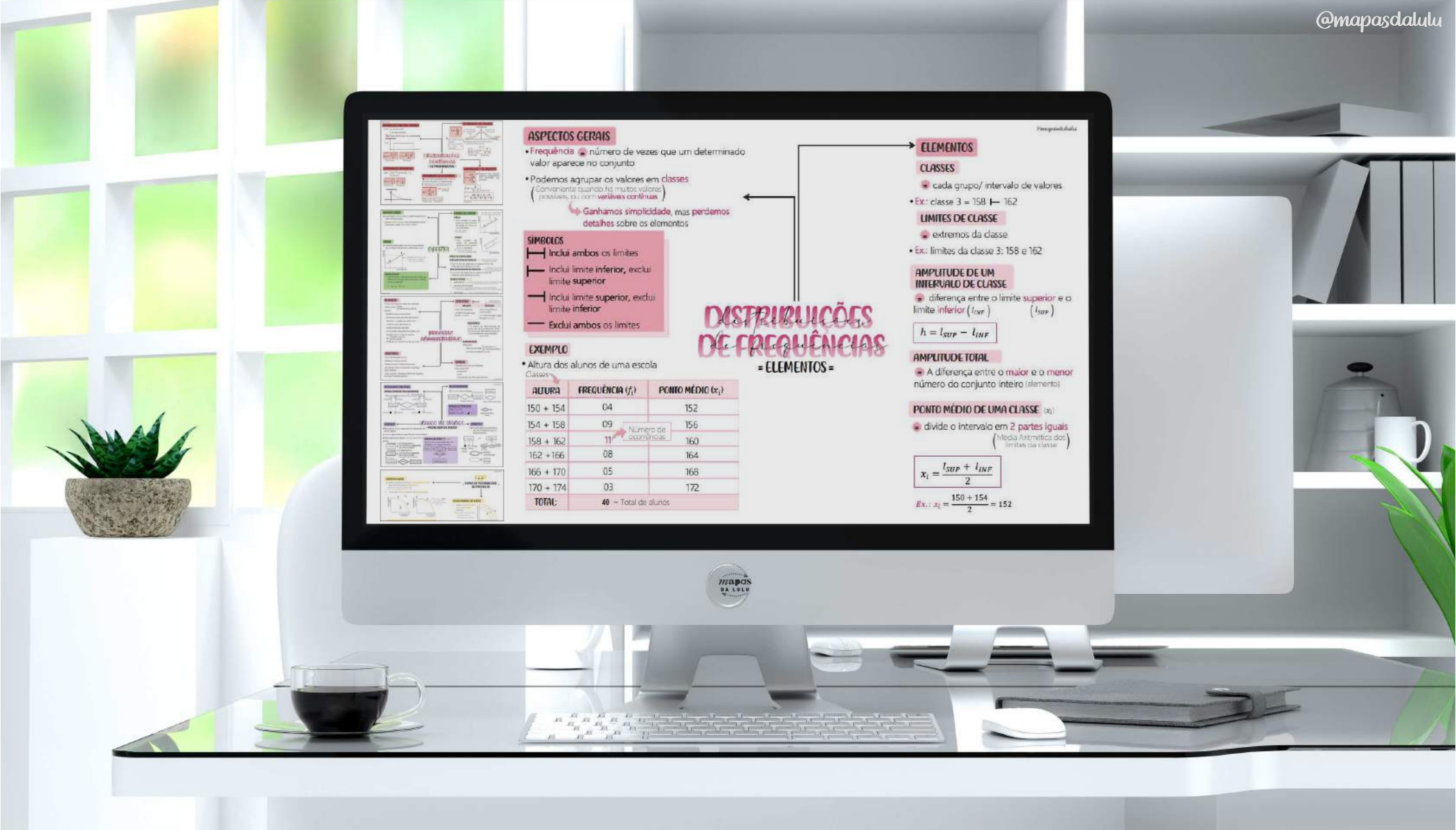


ESTATÍSTICA

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 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

ÍNDICE

1. ESTATÍSTICA

DISTRIBUIÇÕES DE FREQUÊNCIAS

distribuições de frequências

= ELEMENTOS =

@mapasdalulu

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

- $\vdash\!\dashv$ Inclui ambos os limites
- $\vdash$ Inclui limite inferior, exclui limite superior
- $\dashv$ Inclui 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

ELEMENTOS

CLASSES

= cada grupo/ intervalo de valores

- Ex.: classe 3 = 158 $\vdash$ 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 (l_{SUP}) e o limite inferior (l_{INF})

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

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)

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

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

DISTRIBUIÇÕES DE FREQUÊNCIAS
distribuições de frequências

= TIPOS DE FREQUÊNCIAS =

FREQUÊNCIA ABSOLUTA SIMPLES

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

$$\boxed{\sum f_i = n}$$

FREQUÊNCIA RELATIVA SIMPLES
(Normalmente em porcentagem)

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

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

DENSIDADE DE FREQUÊNCIA

= Razão entre a frequência da classe e sua amplitude:

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

EXEMPLO
(Frequências simples)

- Altura dos alunos de uma escola

ALTURA	FREQUÊNCIA (f_i)	FREQUÊNCIA RELATIVA (F_{ri})
150 ⊢ 154	04 (Absoluta simples)	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%)

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	40
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

APRESENTAÇÃO DE DADOS

apresentação de dados

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

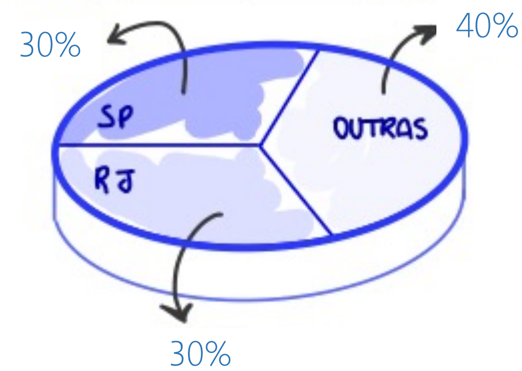
1ª coluna dezena

Unidades correspondentes

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

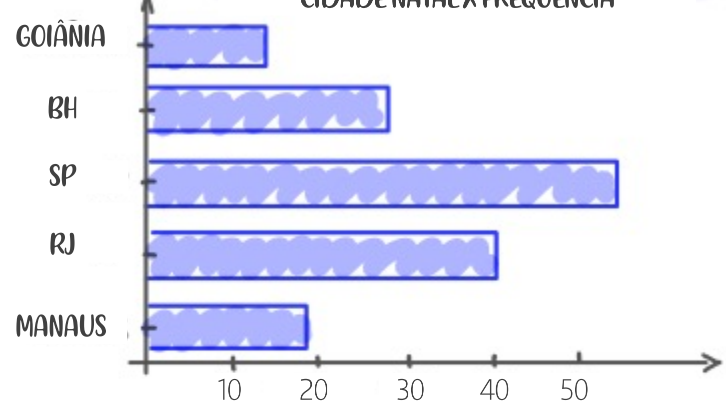
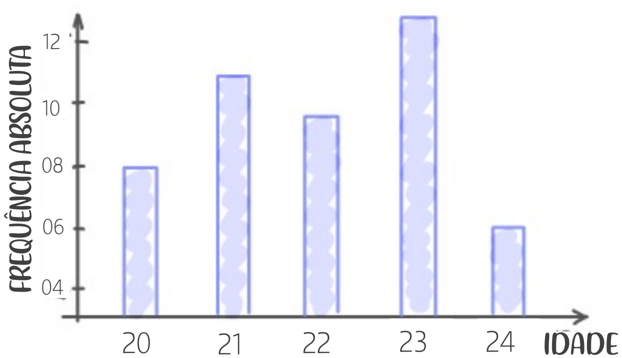


GRÁFICO DE COLUNAS OU BARRAS JUSTAPOSTAS

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



APRESENTAÇÃO DE DADOS

GRÁFICO DE LINHAS

- Usados na representação de séries temporais

ESPERANÇA DE VIDA AO NASCER

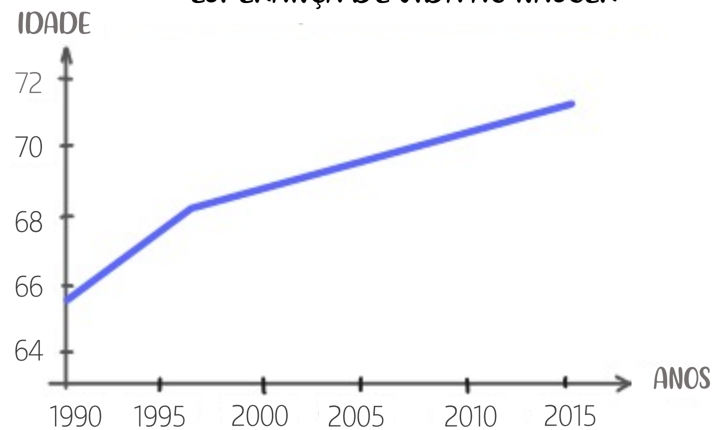
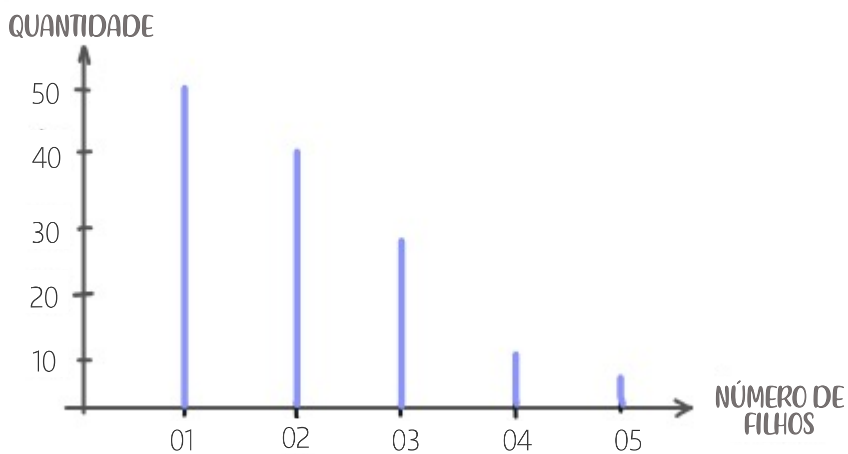


GRÁFICO DE HASTES OU BASTÕES

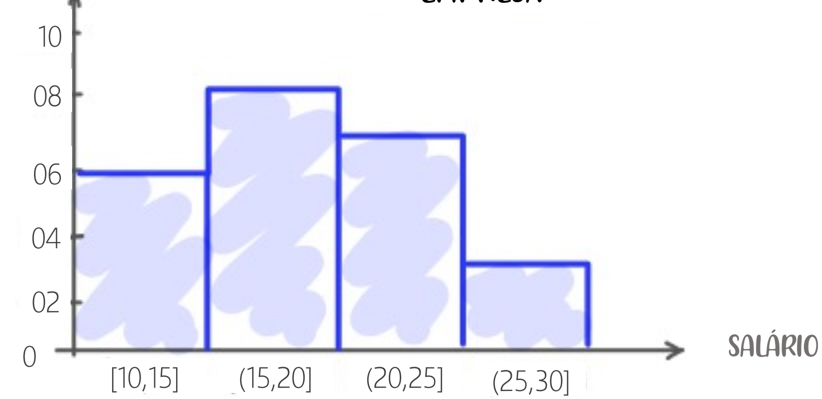
- Usados para representar dados não agrupados em classes
  (Normalmente = dados discretos)



HISTOGRAMAS

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

SALÁRIOS DOS FUNCIONÁRIOS DA EMPRESA



MÉDIAS

médias

MÉDIA ARITMÉTICA SIMPLES

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

Ex.: média aritmética simples dos números $\underbrace{3, 5, 9, 2, 11}_{\text{5 termos (n)}}$

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

$= \dfrac{30}{5} \rightarrow \boxed{\bar{x} = 6}$

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

$soma = \bar{x} \, . \, n$

PROPRIEDADES DA MÉDIA ARITMÉTICA

- Sempre existe e é única

$$\boxed{\text{menor nº do conjunto} \leq \bar{x} \leq \text{maior nº do conjunto}}$$

- 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: $\boxed{\bar{x}' = \bar{x} + c}$ (O equivalente para subtração)
- Multiplicando-se todos os números por uma constante c, a nova média ($\bar{x}'$) será: $\boxed{\bar{x}' = \bar{x} . c}$ (O equivalente para divisão)

Use para simplificar os termos antes de calcular as 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)

$$\boxed{\overline{x_p} = \dfrac{\text{soma dos termos multiplicados pelos respectivos pesos}}{\text{soma dos pesos}}}$$

- Ex.: média aritmética ponderada dos seguintes números e seus pesos:
  - 3, peso 2
  - 4, peso 1
  - 2, peso 5

$$\overline{x_p} = \dfrac{3 . 2 + 4 . 1 + 2 . 5}{2+1+5}$$

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

$$= \dfrac{20}{8}$$

$$\boxed{\overline{x_p} = 2{,}5}$$

MÉDIAS 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 / Você calcula! (coluna auxiliar)

IDADE (x_i)	FREQUÊNCIA (f_i)	$x_i . f_i$
25	21	525
30	47	1.410
34	54	1.836
38	41	1.558
41	37	1.517
TOTAL:	200	6.846

= 25.21

(= total de alunos)

$$\boxed{\bar{x} = \frac{\sum x_i . f_i}{\sum f_i}}$$

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

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)	$x_i . f_i$
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

= 152.4

(= total de alunos)

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

CÁLCULO DO PONTO MÉDIO (x_i)

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

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

MÉDIAS

médias

MÉDIA GEOMÉTRICA

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

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

- Ex.: média geométrica dos termos $\underbrace{3, 8, 9}_{3 \text{ termos (n)}}$:

$$G = \sqrt[3]{3 . 8 . 9} \qquad G = \sqrt[3]{216}$$

$$\boxed{G = 6}$$

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

$$\text{produto dos termos} = \underbrace{\mathrm{G} . \mathrm{G} \dots \mathrm{G}}_{n}$$

DESIGUALDADE DAS MÉDIAS

CAI MUITO!

- Para uma sequência de números positivos

$$\boxed{\bar{x} \geq G \geq H}$$

(Só é igual quando todos os números forem iguais)

MÉDIA HARMÔNICA

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

= inverso da média aritmética dos inversos:

$$H = \left(\frac{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}{n} \right)^{-1}$$

(Fórmula alternativa)

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

$$\text{soma dos inversos} = \frac{\mathrm{n}}{\overline{\mathrm{H}}}$$

Ex.: média harmônica dos termos $\underbrace{3, 4, 9}_{3 \text{ termos (n)}}$:

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

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

$$\boxed{H = 4{,}32}$$

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

mediana

$$\underbrace{3,\ 4,\ 5,\ 7}_{\text{4 elementos}},\ 8,\ \underbrace{9,\ 10,\ 12,\ 13}_{\text{4 elementos}}$$

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

NÚMERO PAR DE TERMOS

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

$$\underbrace{3,\ 4,\ 5,\ 7}_{\text{4 elementos}},\ 8,\ 9,\ \underbrace{10,\ 12,\ 13,\ 15}_{\text{4 elementos}}$$

- 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.

$\boxed{M_d' = M_d + c}$ $\boxed{M_d' = M_d - c}$

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

$\boxed{M_d' = M_d . c}$ $\boxed{M_d' = M_d : c}$

- 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	FREQUÊNCIA (f_i)	FREQÊNCIA 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= 20º termo

Md= número na posição (n+1)/2

Ele está na classe de nota 8!

Logo, Md = 8

EXEMPLO 2:

NOTAS	FREQUÊNCIA (f_i)	FREQÊNCIA 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

MEDIDAS SEPARATRIZES

medidas separatrizes

= MEDIANA =

MEDIANAS PARA DADOS AGRUPADOS

EM CLASSES

ALTURA	FREQUÊNCIA (fi)	FREQÊNCIA ACUMULADA (fac)
40-50	2	2
50-60	5	7 fac_{ant}
l_i 60-70	f_i 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:

$$M_d = l_i + \left[\frac{{}^{n}/_{2} - fac_{anterior}}{f_i}\right] . h$$

DO EXEMPLO

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

$$M_d = 67{,}85$$

l_i	Limite inferior
fac_{ant}	Frequência acumulada da classe anterior
h	Amplitude da classe
f_i	Frequência simples da classe mediana

70 - 60

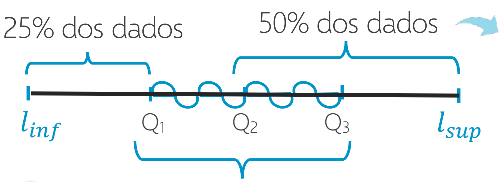
MEDIDAS SEPARATRIZES

medidas separatrizes

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!

Q1 = mediana entre l_{inf} e Q_2

Q3 = mediana entre l_{sup} e Q_2

FÓRMULAS

(Procedimento análogo ao da mediana)

DECORE!

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

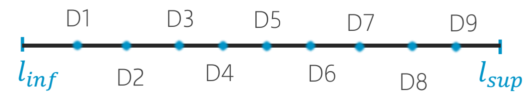
$$Q_2 = l_i + \left[\frac{2.{}^{n}/_{4} - fac_{ant}}{f_i}\right].h$$

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

DECIL

- Divide os dados em 10 partes de mesma frequência

São 9 decis com 10% dos dados cada



FÓRMULAS

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

PERCENTIL

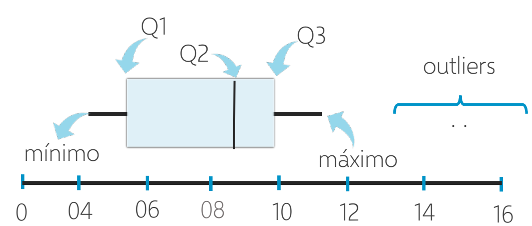
- Divide os dados em 100 partes de mesma frequência (são 99 percentis com 1% dos dados cada)

FÓRMULAS

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

BOX PLOT

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



MODA

moda

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

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

$$\boxed{M_o{}' = M_o + c} \qquad \boxed{M_o{}' = M_o - c}$$

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

$$\boxed{M_o{}' = M_o . c} \qquad \boxed{M_o{}' = M_o : c}$$

MODA

MODA DE PEARSON

DECORE!

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

média

moda

mediana

- 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 f_{ant}
l_i 70-80	f_M 8 classe modal
80-90	3 f_{post}
TOTAL:	25 (n)

moda bruta = 75

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

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

CAI MUITO!

MODA PARA DADOS AGRUPADOS EM CLASSES

MODA DE CZUBER

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

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

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

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

Alguns livros usam a fórmula assim

MODA DE KING

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

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

ATENÇÃO!

MEDIDAS DE DISPERSÃO

medidas de dispersão

ASPECTOS GERAIS

= Analisa o afastamento dos dados

- Exemplos:
  - Amplitudes (Total, interquartílica ....)
  - Desvios (Quartílico, médio, padrão...)
  - Variância
  - Coeficiente de Variação

AMPLITUDE TOTAL

$$\boxed{A = x_{máx} - x_{min}}$$

= 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 ⟶ a amplitude não é alterada.
- Multiplicando-se (dividindo-se) todos os valores por uma constante c ⟶ a amplitude também é multiplicada (ou dividida) por c.

DESVIOS

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

$$\boxed{d_i = x_i - m}$$

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

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:

Os pesos são as frequências (f_i)

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

VARIÂNCIA (Populacional)

= média aritmética dos quadrados dos desvios

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

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

$\sigma^2 =$ 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

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

σ : Desvio padrão populacional

S : Desvio padrão amostral

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

MEDIDAS DE DISPERSÃO

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 c^2

VARIÂNCIA E DESVIO PADRÃO

PARA DADOS AGRUPADOS

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

Deve-se multiplicar cada resultado pela respectiva frequência

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

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
- $C_v^{\ 2}$ = variância relativa

ANÁLISE COMBINATÓRIA

análise combinatória

@mapasdalulu

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:

$= p_1 \cdot p_2 \cdot \ldots \; p_n$ (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?

$= n \,.\, (n-1) \,.\, (n-2) \,.\, \ldots \,.\, 1 = n!$

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:

$$\boxed{= \frac{n!}{r!\,.\,s!}}$$

(Como uma "correção" pela existência das repetições)

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)

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

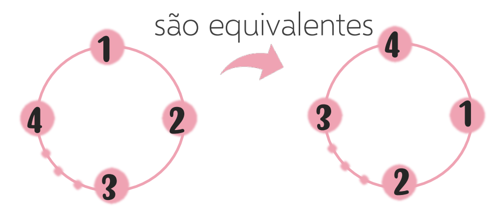
COMBINAÇÃO COM REPETIÇÃO

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

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

PERMUTAÇÃO CIRCULAR

ATENÇÃO!



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

PROBABILIDADE

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

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

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

PROPRIEDADES

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

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

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

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

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

COMBINAÇÃO DE EVENTOS

UNIÃO

$A \cup B$ → Ocorre se ocorrer A ou B ou ambos

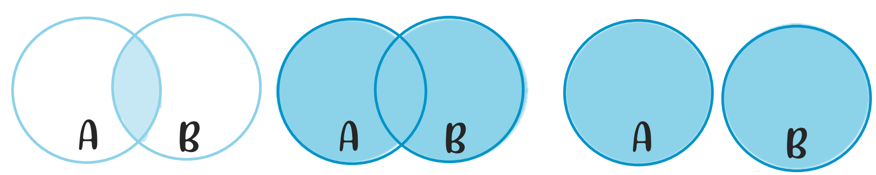
INTERSECÇÃO

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

COMPLEMENTAR

$\bar{A}$ → Ocorre se não ocorrer A

REPRESENTAÇÃO GRÁFICA



$A \cap B$ $A \cup B$

mutuamente excludentes

CASOS ESPECIAIS

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

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

@mapasdalulu

PROBABILIDADE
probabilidade
= CONDICIONAL =

ASPECTOS GERAIS

= Probabilidade de um evento B ocorrer dado que o evento A ocorreu

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

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

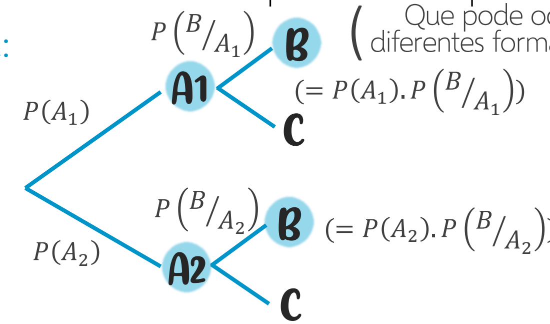
$n(A \cap B)$: nº de elementos da interseção entre A e B

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.

Ex.: (Que pode ocorrer de diferentes formas: A1, A2 ...)



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

TEOREMA DE BAYES

DECORE!

- Para descobrir a probabilidade de ocorrer A1, dado que B ocorreu

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

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

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

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

VARIÁVEIS ALEATÓRIAS

variáveis aleatórias

CONCEITO

= variável associada a uma distribuição de probabilidade

podem assumir diferentes valores

MODA

= valor com maior probabilidade

ESPERANÇA MATEMÁTICA

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

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

PROPRIEDADES

- $E(k.x) = k.E(x)$
- $E(x+k) = E(x) + k$
- $E(x+y) = E(x) + E(y)$
- $E(k) = k$

PEGADINHA!

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

Se X e Y forem variáveis independentes:

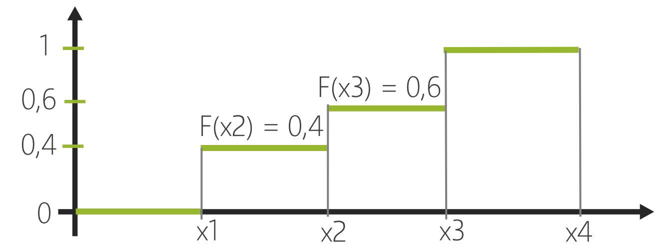
- $E(x.y) = E(x).E(y)$

FUNÇÃO DE DISTRIBUIÇÃO

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

(= Probabilidade de a variável aleatória assumir valores ≤ 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

$$\boxed{\sigma^2 = E(X^2) - E(X)^2}$$

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

Lembrando que E(x)=μ

VARIÁVEIS ALEATÓRIAS

variáveis aleatórias

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

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

- 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 $\boxed{Cov(x,y) = E(x.y) - E(x).E(y)}$

"esperança do produto – produto das esperanças"

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

Assume valores em **[-1,1]**

- ρ(x,y) =1 correlação linear perfeita positiva
- ρ(x,y)= -1 correlação linear perfeita negativa

- Se x e y são variáveis **independentes:**
  - ρ(x,y)=0
  - Cov(x,y)=0

Mas cuidado: se ρ(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.Cov(x,y)
- Cov(x+y,z) = Cov(x,z) + Cov(y,z)

VARIÂNCIA DA SOMA E DA DIFERENÇA

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

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

COEFICIENTE DE VARIAÇÃO

CAI MUITO!

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

VARIÂNCIA RELATIVA

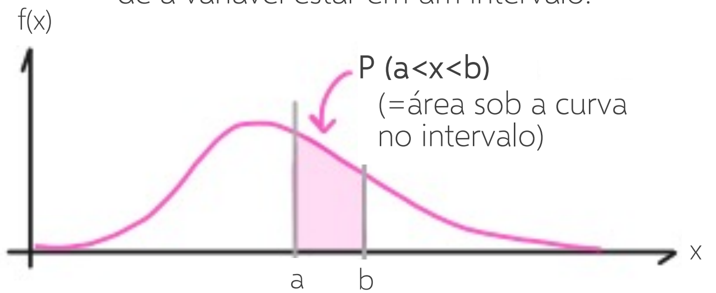
$$V_R = {C_V}^2 = \frac{\sigma^2}{\mu^2}$$

VARIÁVEIS CONTÍNUAS

CONCEITOS

= variável que pode assumir qualquer valor dentro de um intervalo

- Deve existir uma função f tal que
  - f(x) ≥ 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)

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

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

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

MODA

- = valor x que maximiza f(x)

MEDIANA

- = número m tal que:

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

MÉDIA

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

VARIÂNCIA

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

ou

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

@mapasdalulu

DISTRIBUIÇÕES DISCRETAS

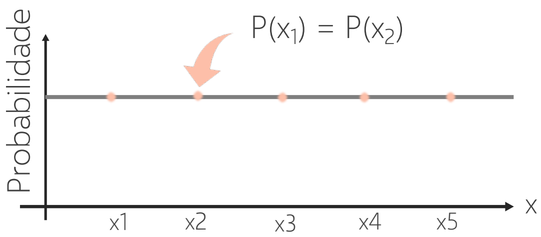
distribuições discretas

= DE PROBABILIDADE =

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

$$\begin{cases} P(\text{sucesso})=p \\ P(\text{fracasso})=q=(1-p) \end{cases}$$

$\sigma^2 = p\,.\,q$

$\boxed{E(x) = p}$ (Esperança)

$\boxed{\sigma^2 = p\,(1-p)}$ (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

$$P(x=k) = \underbrace{q\,.\,q\,.\,...\,q}_{(k-1)\text{ fracassos}}\,.\,p$$

1 sucesso

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

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

$\boxed{\sigma^2 = \frac{q}{p^2}}$ (Variância)

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)

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

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

$\sigma^2 = n\,.\,p\,.\,q$

$\boxed{E(x) = n\,.\,p}$ (Esperança)

$\boxed{\sigma^2 = n\,.\,p\,(1-p)}$ (Variância)

@mapasdalulu

DISTRIBUIÇÕES DISCRETAS
= DE PROBABILIDADE =

DISTRIBUIÇÃO HIPERGEOMÉTRICA

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

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

$$E(x) = n\,.\,p$$

(Esperança)

$$p = \frac{S}{N} \; e \; q = 1-p$$

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

(Variância)

DISTRIBUIÇÃO DE POISSON

DECORE!

= distribuição binomial, em que p é muito pequeno e n muito grande $\left(\begin{smallmatrix} n\to\infty \\ p\to 0\end{smallmatrix}\right)$

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

(Esperança)

$$\sigma^2 = n\,.\,p\,.\,q \quad \left(\begin{smallmatrix} p\to 0, \\ q\to 1\end{smallmatrix}\right)$$

$$\sigma^2 = \lambda$$

(Variância)

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

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

e = número de Euler = 2,718...

@mapasdalulu

DISTRIBUIÇÕES CONTÍNUAS

distribuições contínuas

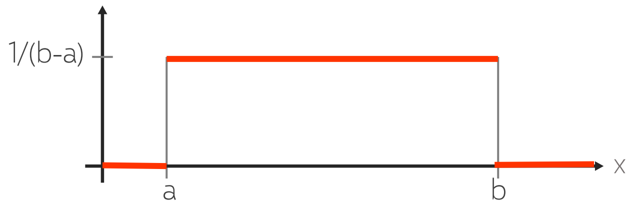
= DE PROBABILIDADE =

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

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

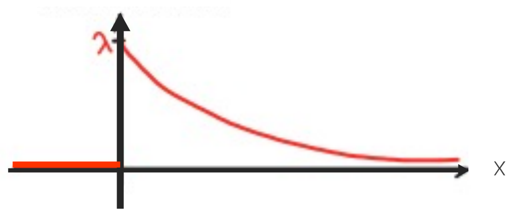
$$E(x) = \frac{1}{\lambda}$$

(Esperança)

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

(Variância)

PROBABILIDADE

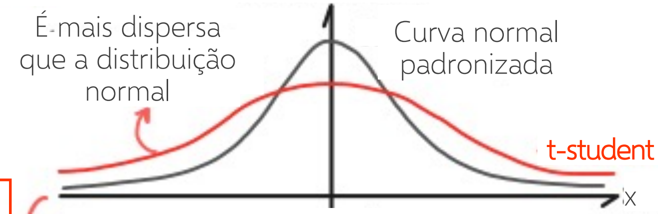


DISTRIBUIÇÃO t DE STUDENT

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

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

PROBABILIDADE



Quanto maior for k, mais a curva se aproxima da normal

Se k>1:

$$E(x) = 0$$

(Esperança)

Se k>2:

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

(Variância)

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)

$$\sigma^2 = 2.k$$

(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)

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

(Variância)

DISTRIBUIÇÕES CONTÍNUAS

= DISTRIBUIÇÃO NORMAL =

ASPECTOS GERAIS

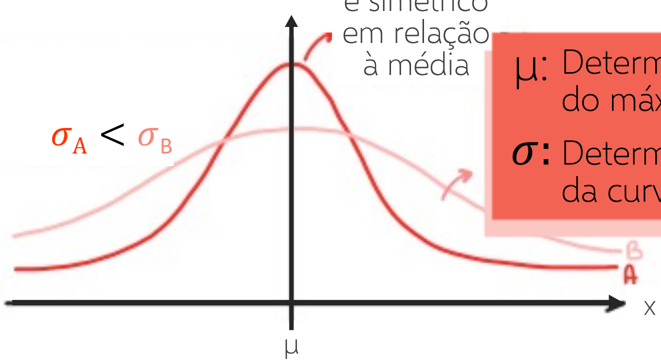
- só depende da **média** (μ) e **desvio padrão** (σ)

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

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

MÉDIA = MEDIANA = MODA DECORE!

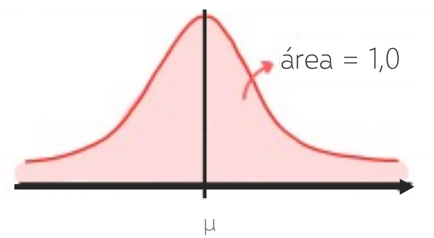
REPRESENTAÇÃO GRÁFICA

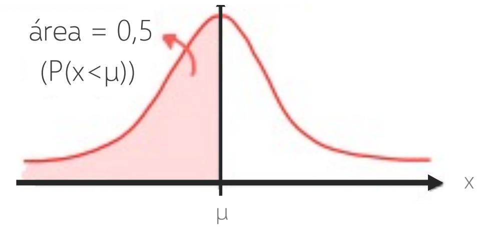


μ: Determina a posição do máximo da curva

σ: Determina o formato da curva

PROPRIEDADES DA CURVA NORMAL





DISTRIBUIÇÃO NORMAL PADRÃO N(0,1)

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

Número a transformar

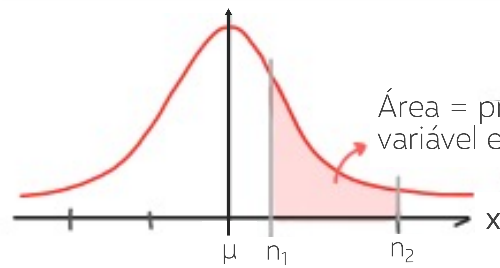
$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

CAI MUITO!

= Número correspondente na distribuição padrão

- É a distribuição usada na tabela para consultas.

CÁLCULO DA PROBABILIDADE DE INTERVALOS



Área = probabilidade de a variável estar entre **n1** e **n2**

PASSO A PASSO

- Transforme os valores nos correspondentes aos da normal padrão
- Consultar: **P1** (z<n1) e **P2**(z<n2)

$$\boxed{P(n_1 < x < n_2) = P_2\,(z < n_2) - P_1\,(z < n_1)}$$

AMOSTRAGEM

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 todos os seus elementos

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.

@mapasdalulu

ESTIMADORES

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): $\boxed{\varepsilon = \hat{\theta} - \theta}$

SIMBOLOGIA

IMPORTANTE!

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

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

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

n = número de elementos da amostra

$$E(\bar{X}) = \mu \qquad S_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

FATOR DE CORREÇÃO PARA POPULAÇÕES FINITAS

- As variáveis $x_1, \ldots, x_n$ 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($\mathrm{E} = \widehat{(\theta)} - \theta$)
  - 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)

@mapasdalulu

ESTIMADORES

DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

CAI MUITO!

- Proporção p de sucessos da população

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

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

$$E(X) = p \qquad VAR(X) = p(1-p)$$

$(\sigma^2 = p\,.\,q)$

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

X = distribuição binomial

$$E(X) = n\,.\,p \qquad VAR(X) = n\,.\,p\,.\,q$$

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

$$\begin{cases} E(\hat{p}) = p \\ VAR(\hat{p}) = \dfrac{p\,.\,q}{n} \end{cases}$$

É um estimador:

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

FATOR DE CORREÇÃO PARA POPULAÇÕES FINITAS

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

DECORE!

N= número de elementos da população

DISTRIBUIÇÃO AMOSTRAL DA VARIÂNCIA

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

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1} \qquad VAR(S^2) = \frac{2\sigma^4}{n-1}$$

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

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

= $(\bar{X} - Z_o \,.\, \sigma_X)$ $\bar{X}$ $(\bar{X} + Z_o \,.\, \sigma_X)$

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

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

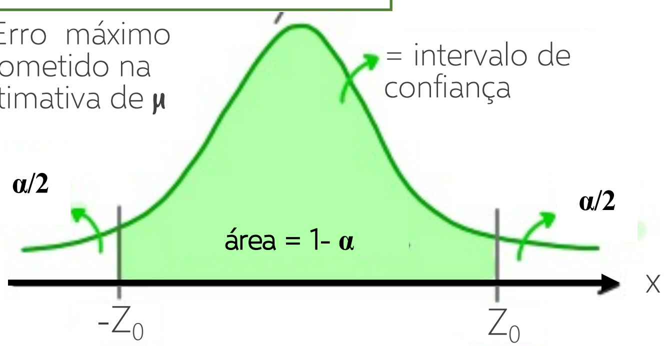
$\sigma_{\bar{X}}$: desvio padrão da média amostral

- Amplitude do intervalo → $A = 2.Z_o \,.\, \sigma_{\bar{X}}$

$$\boxed{\text{erro máximo} = \frac{A}{2} = Z_o \,.\, \sigma_{\bar{X}}}$$

= Erro máximo cometido na estimativa de μ

= intervalo de confiança



PARA μ QUANDO σ^2 É DESCONHECIDO

INTERVALO DE CONFIANÇA

= $(\bar{X} - t_o \,.\, \sigma_X)$ $\bar{X}$ $(\bar{X} + t_o \,.\, \sigma_X)$

- 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

$$S_{\bar{X}}^2 = \frac{\sigma^2}{n} \quad \text{(População infinita ou amostragem com reposição)}$$

$$S_{\bar{X}}^2 = \frac{\sigma^2}{n} \,.\, \frac{N-n}{N-1} \quad \text{(População finita ou amostragem sem reposição)}$$

- Amplitude do intervalo → $A = 2.t_o \,.\, \sigma_{\bar{X}}$

$$\boxed{\text{erro máximo} = \frac{A}{2} = t_o \,.\, \sigma_{\bar{X}}}$$

INTERVALOS DE CONFIANÇA

PARA UMA PROPORÇÃO

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

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

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

- Amplitude do intervalo → $A = 2.Z_o \,.\, S_{\hat{p}}$

$$\boxed{\text{erro máximo} = \frac{A}{2} = Z_o \,.\, S_{\hat{p}}}$$

- Usamos a distribuição normal

PARA A VARIÂNCIA

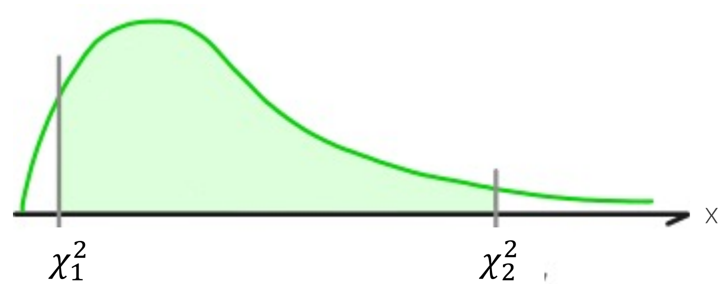
IMPORTANTE!

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

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

n-1 graus de liberdade

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



TESTES DE HIPÓTESES

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

TIPO DE TESTES

DECORE!

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ → 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

IMPORTANTE!

ERRO TIPO I = rejeitar H_0, quando for verdadeira

- P(erro tipo I)= α (Nível de significância do teste)
- Nível de confiança: 1- α

ERRO TIPO II = aceitar H_0, quando for falsa

- P(erro tipo II) = β
- Poder do teste: 1- β (Probabilidade de rejeitar H_0, quando for falsa)

PEGADINHA!

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

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

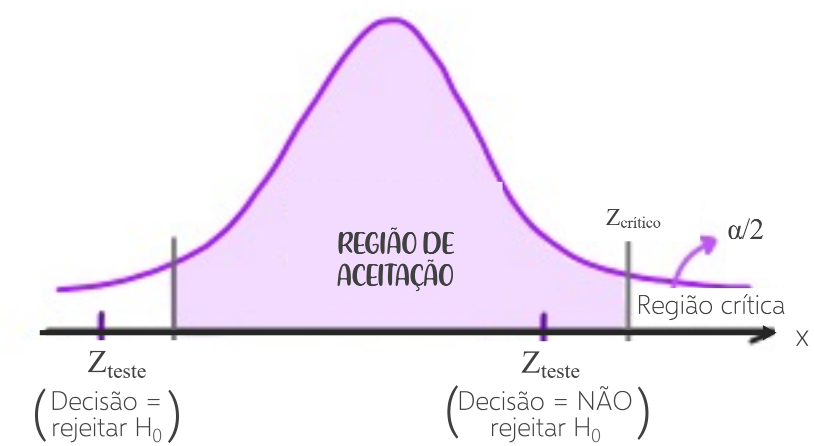
Média observada

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

Média populacional

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

- Verificamos onde Z_{teste} cai:



POPULAÇÃO NORMAL COM σ^2 DESCONHECIDO

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

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

n-1 graus de liberdade

$$S_{\bar{X}}^2 = \frac{\sigma^2}{n}$$ (População infinita ou amostragem com reposição)

$$S_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1}$$ (População finita ou amostragem sem reposição)

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:

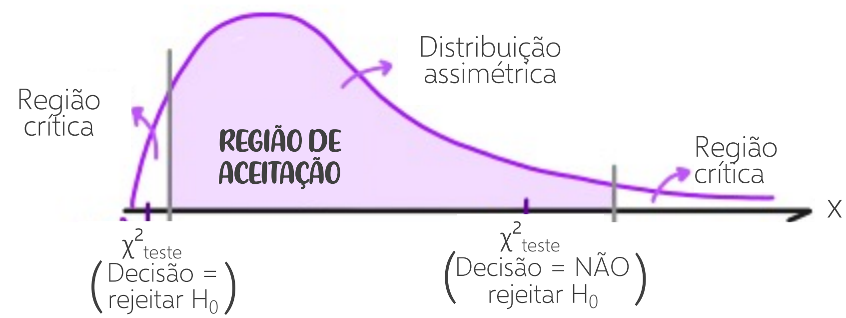
$$Z_{teste} = \frac{\hat{p} - p}{\sqrt{\frac{p \, . \, q}{n}}}$$

TESTE PARA VARIÂNCIA

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

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

n-1 graus de liberdade



ANÁLISE DE VARIÂNCIA

análise de variância

ASPECTOS GERAIS

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

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

(não especifica qual)

- Selecionam-se k amostras independentes de cada uma das populações

n_1: elementos da primeira população, ... ,
n_k: elementos da k-ésima população

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

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

$$SQ_{total} = SQden_{tro} + SQent_{re}$$

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

- As variáveis $\frac{SQdentro}{\sigma^2}$, $\frac{SQentre}{\sigma^2}$ e $\frac{SQtotal}{\sigma^2}$ têm distribuição qui-quadrado.

ANÁLISE DE VARIÂNCIA

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

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

$\chi^2_{k_1}$ e $\chi^2_{k_2}$ são variáveis aleatórias com distribuição de qui-quadrado, com k1 e k2 graus de liberdade

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

ou

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

utilizada para testar

- $H_0: \mu_0 = \mu_1 = \cdots = \mu_k$
- H_1: pelo menos uma delas é diferente das demais

HIPÓTESE	H_0
$F_{teste} > F_{crítico}$	Rejeitamos H_0
$F_{teste} < F_{crítico}$	Aceitamos H_0

REGRESSÃO LINEAR

regressão linear

ASPECTOS GERAIS

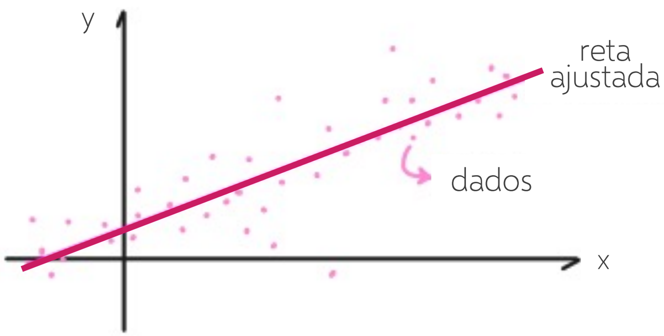
= calcular a expressão matemática que relaciona duas variáveis x e y:

$$Y = p + m \, . \, x$$

p= coeficiente linear
m= coeficiente angular

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

Exemplo:



COEFICIENTE DE CORRELAÇÃO LINEAR DE *PEARSON*

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

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:

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

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

$\hat{Y} = a + bXi$ $\quad$ $e = Y_i - \hat{Y}_i$ → 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 $\bar{Y}$ = a+b $\bar{X}$ para determinar *a*

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

RETA QUE PASSA PELA ORIGEM

- $\boxed{\alpha = 0}$ $\quad$ $Y_i = \beta X_i + \varepsilon_i$ → $\beta = \dfrac{\sum X \, . \, Y}{\sum X^2}$

REGRESSÃO LINEAR

regressão linear

ANÁLISE DA VARIÂNCIA DA REGRESSÃO

$$SQT = SQM + SQR$$

- SQT: Soma dos quadrados total
- SQM: Soma dos quadrados do modelo
- SQR: Soma dos quadrados dos resíduos

FÓRMULAS IMPORTANTES

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

Soma dos quadrados do modelo de regressão

ou

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

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

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

TABELA DA ANOVA

DECORE!

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

IMPORTANTE!

- Coeficiente de correlação: $R = \sqrt{\frac{SQM}{SQT}}$
- Coeficiente de determinação: $R^2 = 1 - \frac{SQR}{SQT}$

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

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