CLIMA BRONQUIOLITE

CNPq FIOCRUZ Sala de Situação de Saúde FS·UnB SUS MINISTÉRIO DA SAÚDE GOVERNO FEDERAL BRASIL UNIÃO E RECONSTRUÇÃO

Curso

Análise de Dados de Saúde e Clima: Estatísticas para Políticas Públicas

Análise de correlação – Revisão

Módulo 2: Análise descritiva e de correlação – 06/08/2024

Renata Yokota
Felipe Freitas

Conteúdo

CLIMA
BRONQUIOLITE

Revisão

Conceitos básicos de estatística descritiva

1. Tipos de variáveis

1.1. Variáveis quantitativas (numéricas)

Importante para escolha do melhor método de como apresentar e analisar os dados

Contínua

- Medida em escala contínua
- Número **infinito** de valores
- Exemplos: temperatura (ºC), umidade relativa (%), pluviosidade (mm)

Discreta

- Número **finito** de valores
- Geralmente números inteiros
- Não podem assumir valores intermediários entre dois valores consecutivos
- Exemplos: número de moradores no domicílio, número de filhos

1.2. Variáveis qualitativas (categóricas)

Binárias ou dicotômicas

- Duas categorias
- Sim/Não

Ordinal

- Categorias naturalmente ordenadas.
- Exemplos: faixa etária; grau de escolaridade

Nominal

- Categorias não ordenadas
- Exemplo: raça/cor (branca, preta, parda, amarela, indígena); sexo

CLIMA BRONQUIOLITE

1.2. Variáveis qualitativas (categóricas)

Binárias ou dicotômicas

- Duas categorias
- Sim/Não

Ordinal

- Categorias naturalmente ordenadas.
- Exemplos: faixa etária; grau de escolaridade

Nominal

- Categorias não ordenadas
- Exemplo: raça/cor (branca, preta, parda, amarela, indígena); sexo

Hierarquia dos tipos de variáveis

Mais informação

Continua	Valores únicos
Discreta	Magnitude
Ordinal	Ordenação lógica
Nominal	Múltiplas categorias
Binária/dicotômica	Duas categorias

1.Tipos de variáveis

Desfecho

- Foco do estudo
- Variável que buscamos maior conhecimento com o estudo

Exposição

- Fatores que podem influenciar a ocorrência ou magnitude do desfecho

Y

Variável resposta
Variável dependente

X

Variável explanatória
Variável independente

Tipo de desfecho → **Analise estatística**

1.Tipos de variáveis

Qual a influência das variáveis climáticas na incidência de bronquiolite em crianças menores de 5 anos no Brasil?
Desfecho?
Exposições?

Desfecho

- Foco do estudo
- Variável que buscamos maior conhecimento com o estudo

Y

Variável resposta
Variável dependente

Exposição

- Fatores que podem influenciar a ocorrência ou magnitude do desfecho

X

Variável explanatória
Variável independente

Tipo de desfecho → **Analise estatística**

CLIMA BRONQUIOLITE

2. Medidas de tendência central

2.1. Média

$$\bar{x} = \frac{\sum x}{n}$$

- Influenciada por valores extremos (*outliers*)
- Apropriada para variáveis quantitativas

A regressão linear é baseada na média

CLIMA
BRONQUIOLITE

2. Medidas de tendência central

2.1. Média – Exemplo: Idade

Indivíduo	Idade (x)
1	15
2	20
3	20
4	30
5	25
6	25
7	35
8	40
9	30
10	20
$\sum$	

2. Medidas de tendência central

2.1. Média – Exemplo: Idade

Indivíduo	Idade (x)
1	15
2	20
3	20
4	30
5	25
6	25
7	35
8	40
9	30
10	20
∑	260

CLIMA BRONQUIOLITE

2. Medidas de tendência central

2.1. Média – Exemplo: Idade

Indivíduo	Idade (x)	Média ($\overline{x}$)	Erro ($x - \overline{x}$)
1	15	26	
2	20	26	
3	20	26	
4	30	26	
5	25	26	
6	25	26	
7	35	26	
8	40	26	
9	30	26	
10	20	26	
$\sum$	260	-	

2. Medidas de tendência central

2.1. Média – Exemplo: Idade

Indivíduo	Idade (x)	Média ($\overline{x}$)	Erro ($x - \overline{x}$)
1	15	26	-11
2	20	26	-6
3	20	26	-6
4	30	26	4
5	25	26	-1
6	25	26	-1
7	35	26	9
8	40	26	14
9	30	26	4
10	20	26	-6
$\sum$	260	-	0

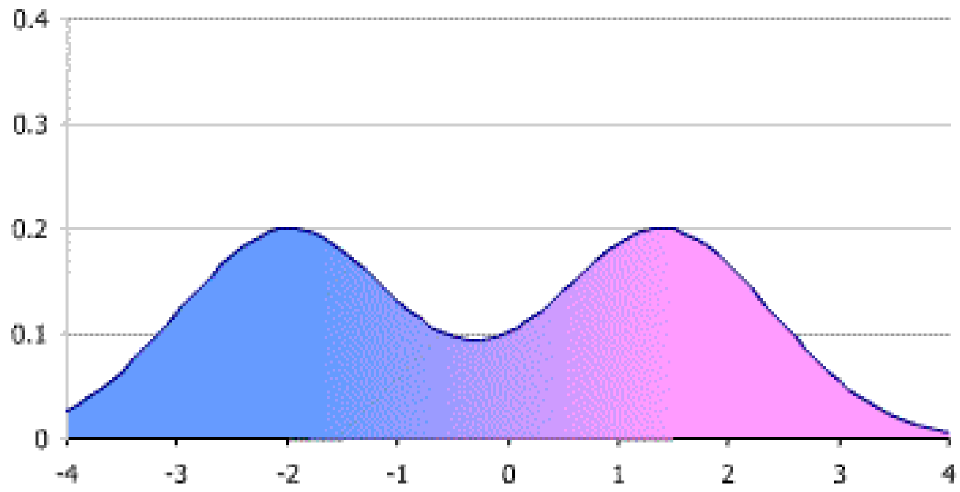
2. Medidas de tendência central

2.2. Mediana (P50)

- Divide a distribuição na metade
- Utilizada quando a distribuição possui valores extremos

2.3. Moda

- Medida mais frequente
- Distribuição bimodal



3. Medidas de variabilidade e dispersão

3.1. Intervalo

Intervalo = valor máximo – valor mínimo

- Baseado apenas em 2 observações
- Não informa como as demais observações estão dispersas entre esses 2 valores
- Influenciado por *outliers*

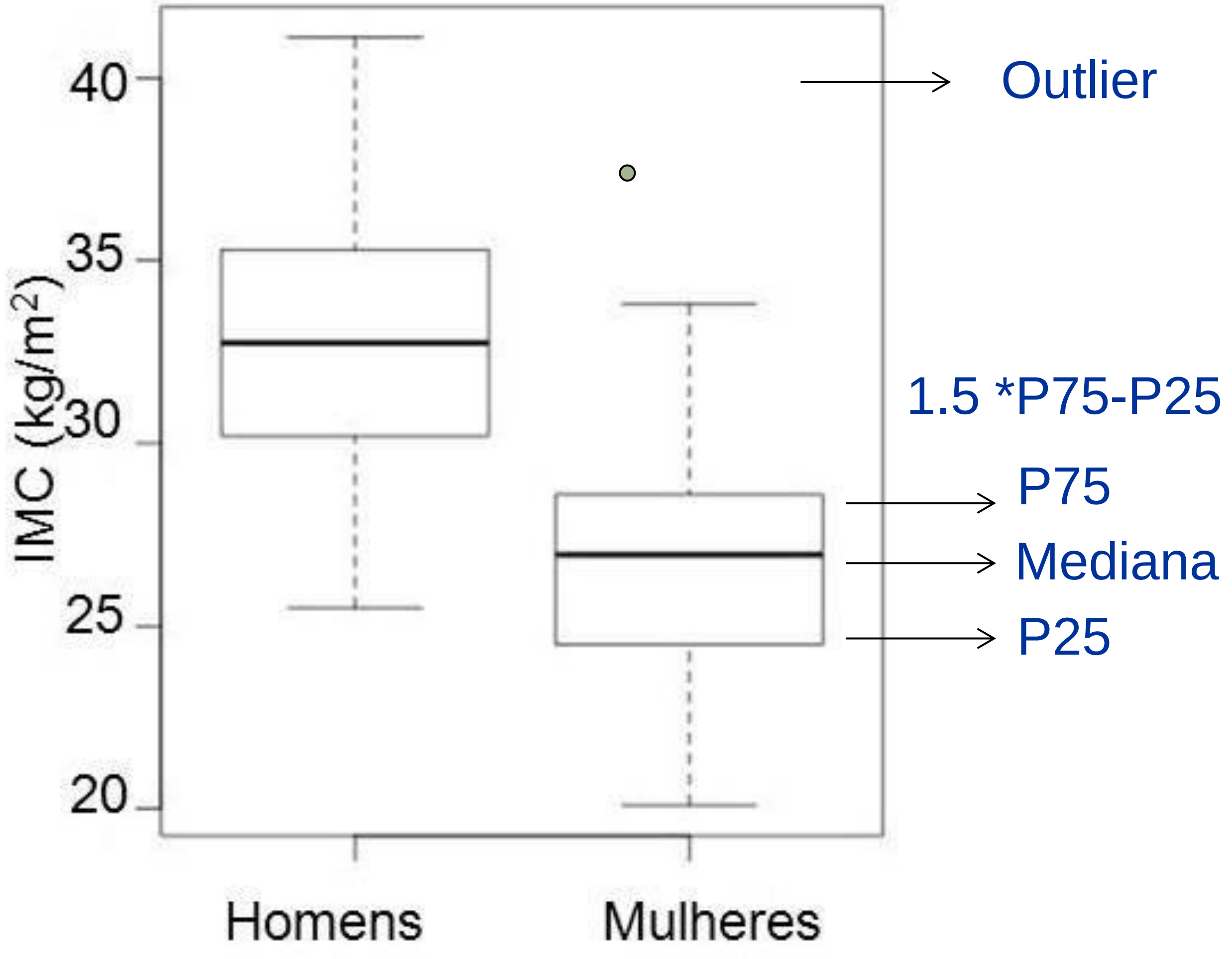
3.2. Intervalo interquartil

Intervalo interquartil = P75-P25

- Dispersão da distribuição entre P25 e P75
- Não é muito influenciado por valores extremos
- Utilizado no box plot
  - frequente
  - Distribuição bimodal

3. Medidas de variabilidade e dispersão

Box Plot



3. Medidas de variabilidade e dispersão

3.3. Variância

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

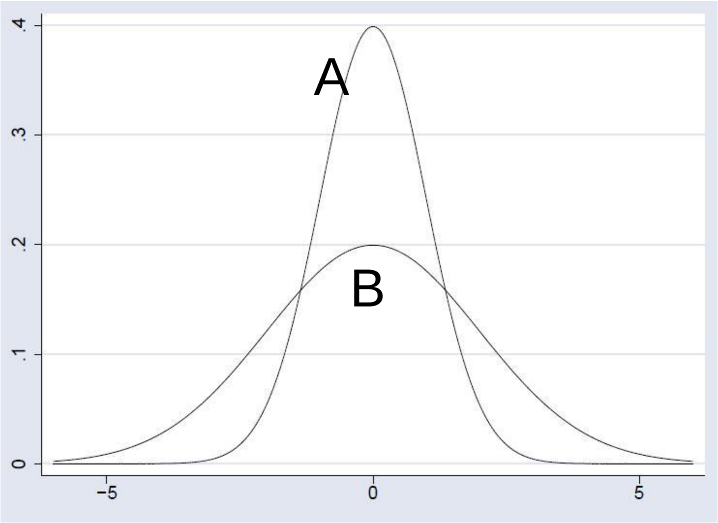
Média do quadrado dos desvios (erros)

- s^2 = variância amostral
- n = tamanho amostral
- x = variável de interesse
- $\bar{x}$ = média das observações

Por que a variância é elevada ao quadrado?

Grau de liberdade (GL) da variância $(n-1)$

- n – 1 desvios são independentes
- 1 desvio pode ser calculado por meio dos outros: A soma dos outros "n" desvios=0
- Preciso estimar a média para calcular a variância – "gasto de 1 GL"
- Em qual das duas curvas a média e mais informativa?



3. Medidas de variabilidade e dispersão

3.3. Variância – Exemplo: Idade

Indivíduo	Idade (x)	Média ($\bar{x}$)	Erro ($x - \bar{x}$)	$(x - \bar{x})^2$
1	15	26	-11	121
2	20	26	-6	36
3	20	26	-6	36
4	30	26	4	16
5	25	26	-1	1
6	25	26	-1	1
7	35	26	9	81
8	40	26	14	196
9	30	26	4	16
10	20	26	-6	36
$\sum$	260	-	0	540

3. Medidas de variabilidade e dispersão

3.3. Exemplo: Idade – cálculo da variância

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

- $\sum(x-\bar{x})^2$ **= 540**
- **n=10**
- s^2 **= 540/9 = 60**

Desvantagem: medida elevada ao quadrado

CLIMA
BRONQUIOLITE

3. Medidas de variabilidade e dispersão

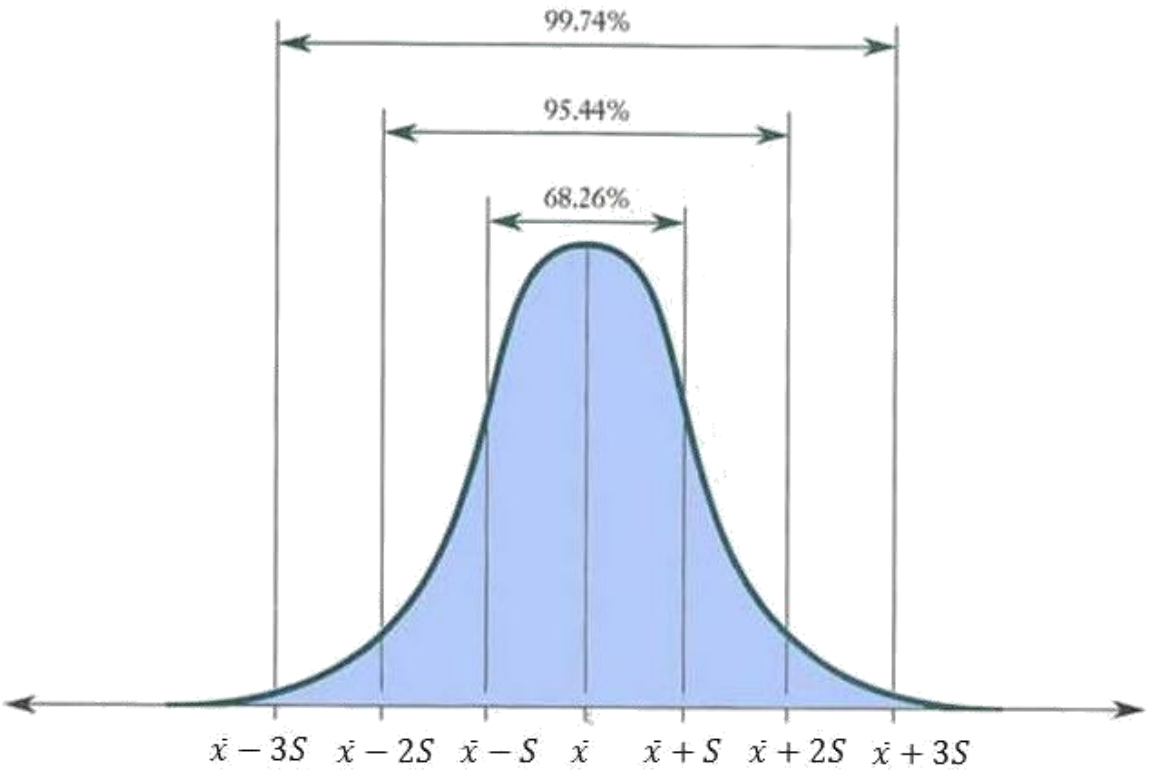
3.4. Desvio-padrão

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

- s= desvio-padrão amostral
- n = tamanho amostral
- x = variável de interesse
- $\bar{x}$ = média das observações

s = desvio-padrão da amostra
σ = desvio-padrão da população

- > variação em relação à **média**, > s
- Muito influenciado por *outliers*



3.4. Exemplo: Idade – cálculo do desvio-padrão, média e mínimo dos quadrados

Id	Idade	Média=10		Média=24		Média=26		Média=30		Média=40	
		$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP
1	15	25		81		121		225		625	
2	20	100		16		36		100		400	
3	20	100		16		36		100		400	
4	30	400		36		16		0		100	
5	25	225		1		1		25		225	
6	25	225		1		1		25		225	
7	35	625		121		81		25		25	
8	40	900		256		196		100		0	
9	30	400		36		16		0		100	
10	20	100		16		36		100		400	
$\sum$	260	3100	18,6	580	8,0	540	7,7	700	8,8	2500	16,7

$$s^2 = \frac{\sum(x-\bar{x})^2}{(n-1)} = \frac{540}{9} = 60$$

$$s = \sqrt{s^2} = \sqrt{60} = 7,7$$

CLIMA BRONQUIOLITE

3.4. Exemplo: Idade – cálculo do desvio-padrão, média e mínimo dos quadrados

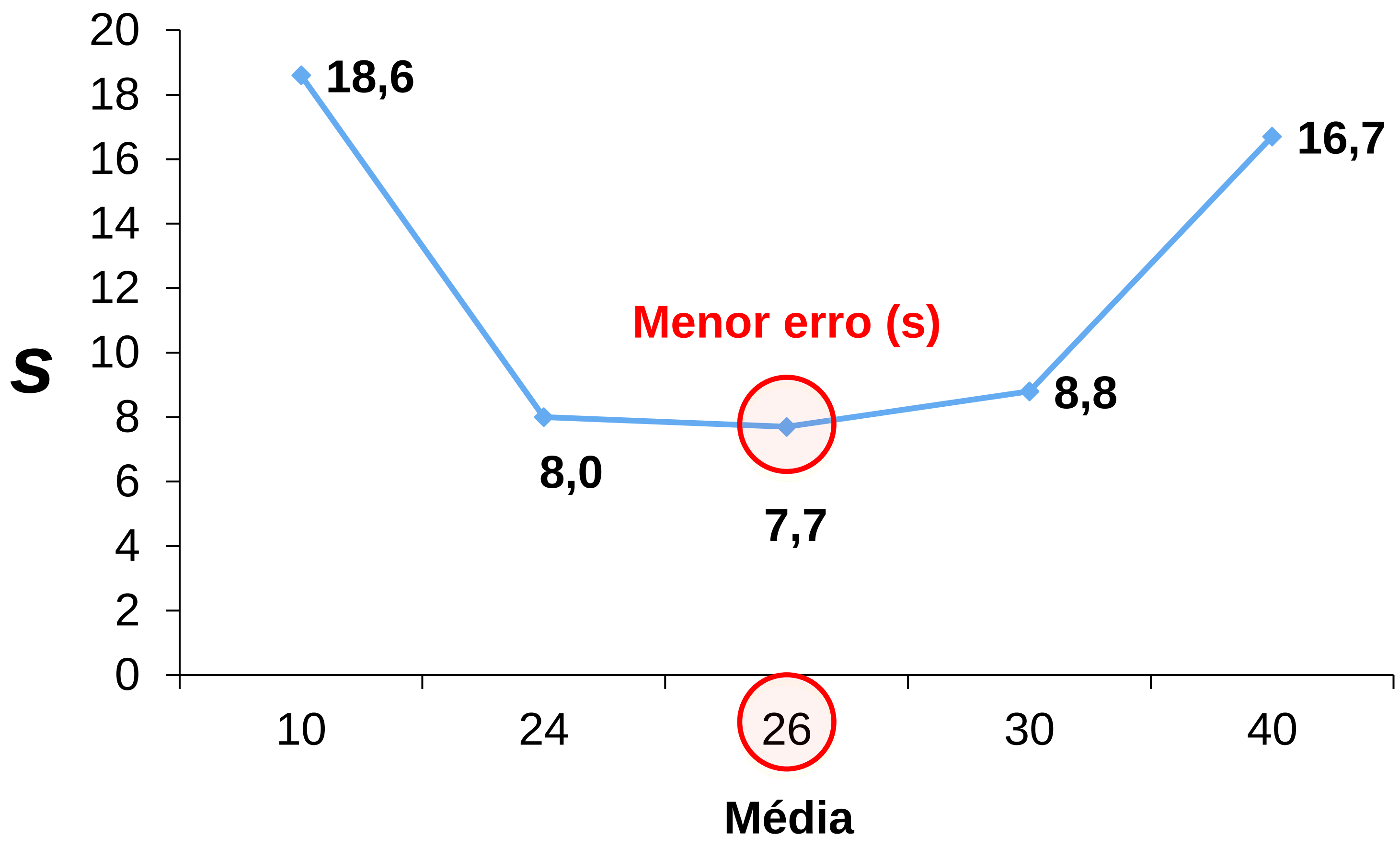
Id	Idade	Média=10		Média=24		Média=26		Média=30		Média=40	
		$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP	$(x-\overline{x})^2$	DP
1	15	25		81		121		225		625	
2	20	100		16		36		100		400	
3	20	100		16		36		100		400	
4	30	400		36		16		0		100	
5	25	225		1		1		25		225	
6	25	225		1		1		25		225	
7	35	625		121		81		25		25	
8	40	900		256		196		100		0	
9	30	400		36		16		0		100	
10	20	100		16		36		100		400	
$\sum$	260	3100	18,6	580	8,0	540	7,7	700	8,8	2500	16,7

$$s^2 = \frac{\sum(x-\bar{x})^2}{(n-1)} = \frac{540}{9} = 60$$

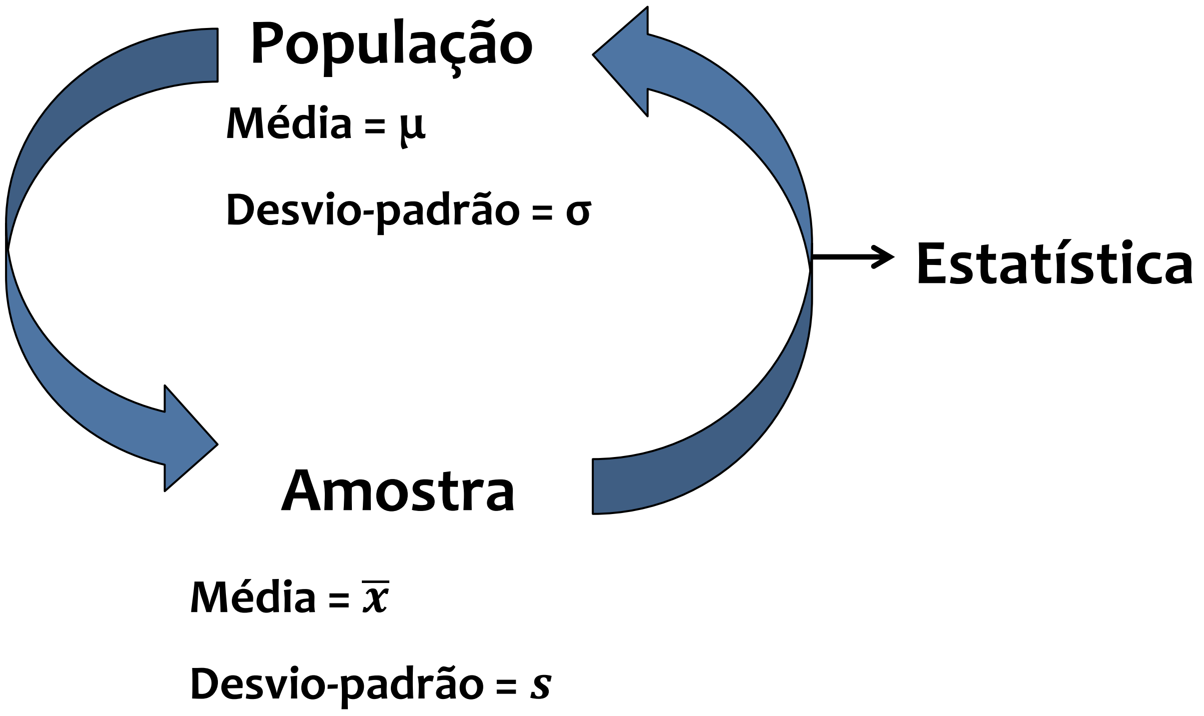
$$s = \sqrt{s^2} = \sqrt{60} = 7,7$$

CLIMA
BRONQUIOLITE

3.4. Exemplo: Idade – Média e mínimo dos quadrados



4. Variação amostral



- Diferença entre μ e $\bar{x}$: variação amostral

4. Variação amostral

4.1. Erro-padrão

- Amostras independentes de mesmo tamanho, da mesma população – dificilmente apresentarão a mesma média
- **Distribuição amostral** – distribuição das médias dessas amostras da mesma população
  - A média da distribuição das amostras = média da população
  - **Erro-padrão da média amostral:**

$$Ep = \frac{s}{\sqrt{n}}$$

Mede a precisão da média amostral como estimativa da média da população

CLIMA BRONQUIOLITE

4. Variação amostral

4.1. Exemplo

População (N=4)

x = idade

- x_1=10
- x_2=20
- x_3=30
- x_4=40

$$\bar{x} = 25$$

Amostras aleatórias de 2 elementos com reposição

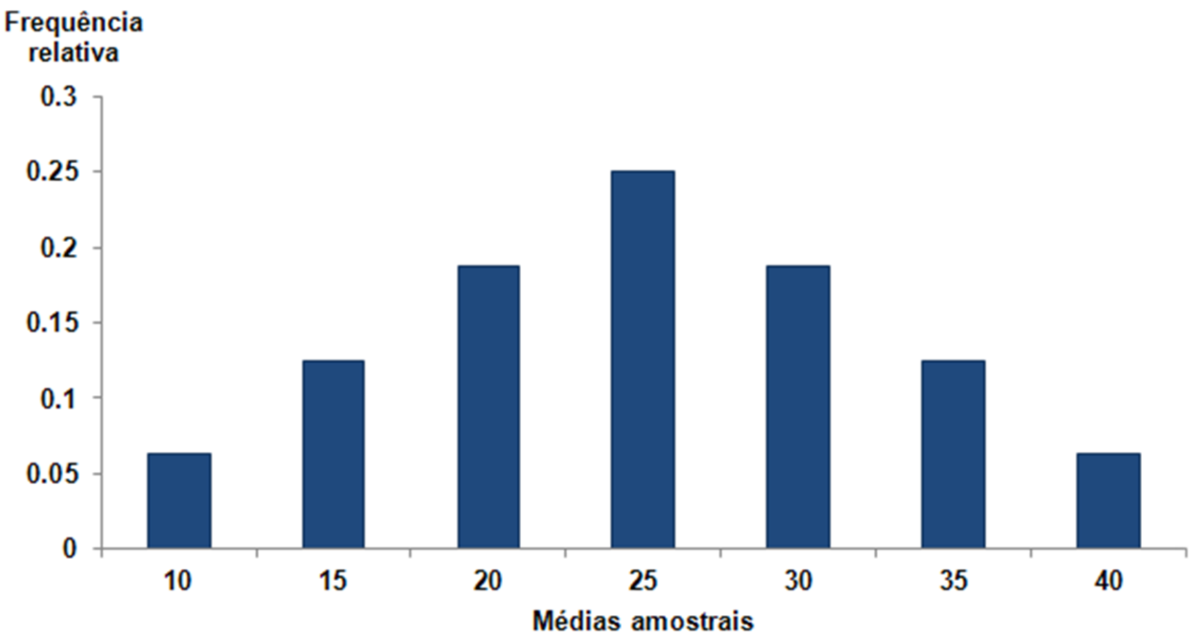
$2^n = 2^4 = 16$ amostras

4. Variação amostral

Médias amostrais possíveis

$\overline{x}$	Frequência	Frequência relativa
10	1	0,0625
15	2	0,1250
20	3	0,1875
25	4	0,2500
30	3	0,1875
35	2	0,1250
40	1	0,0625

4. Variação amostral



Teorema do limite central

A **distribuição amostral** das médias é **normal**, mesmo quando as observações individuais não possuem distribuição normal, desde que o tamanho amostral não seja muito pequeno (n>30?)

4. Variação amostral

4.1. Desvio-padrão x Erro-padrão

Desvio-padrão

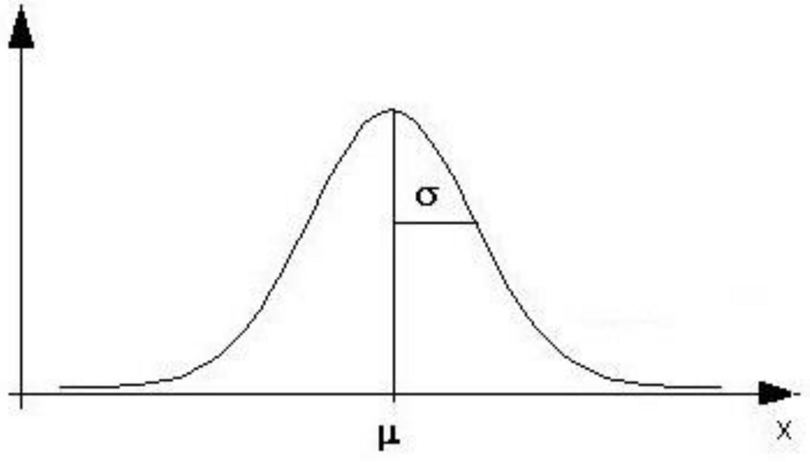
- Variabilidade da população

Erro-padrão

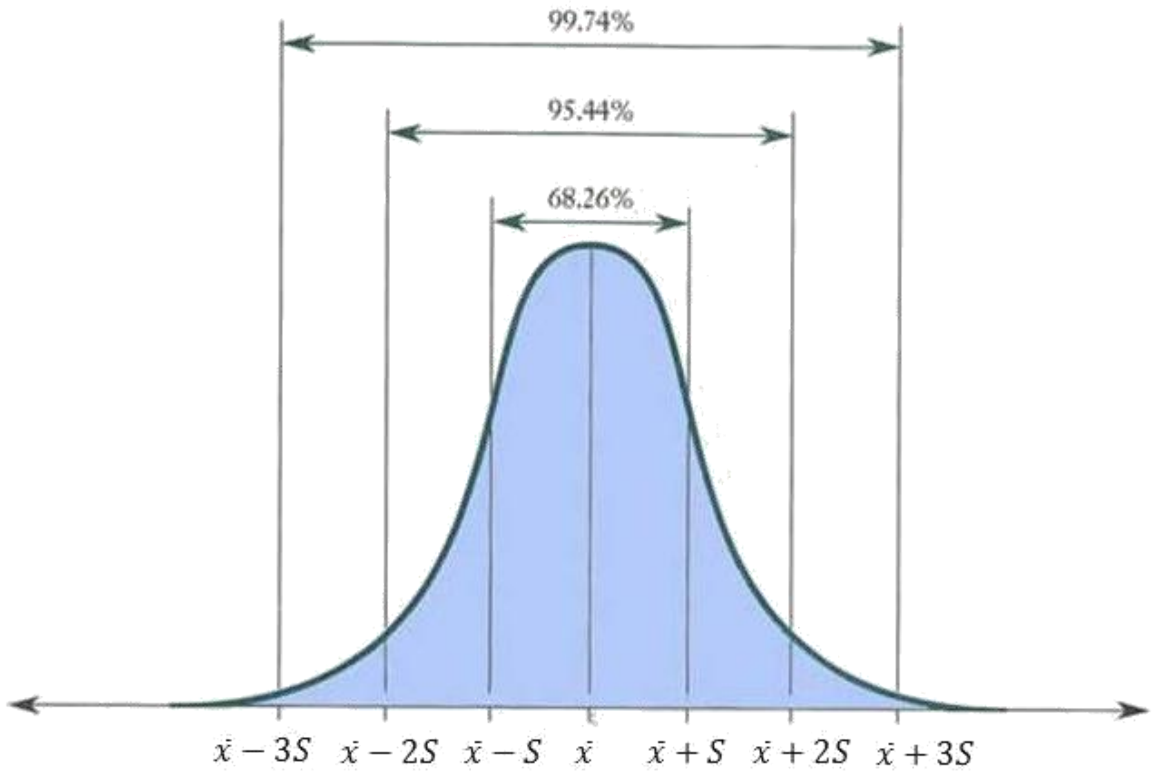
- Variabilidade da média amostral
- Indica o quanto a média amostral está próxima da média populacional

5. Distribuição Normal

5.1. Propriedades



- Formato de sino, com caudas assintóticas ao eixo x $(-\infty; +\infty)$
- Curva simétrica em relação à média
- Média = Moda = Mediana
- Area sob a curva = 1



95% da população apresenta valores de x entre $-1{,}96s$ e $+1{,}96s$

5. Distribuição Normal

Exemplo: Temperatura média diária de Brasília no mês de julho: T~N(25,3²)

$\bar{x} = 25°\text{C}$ $\qquad\qquad s = 3$

Conclusões

- 68% dos dias do mês de julho ("população de estudo") apresentam temperatura entre (25-3) 22ºC e (25+3) 28ºC
- 95% dos dias do mês de julho ("população de estudo) apresentam temperatura entre (25-6) 19ºC e (25+6) 31ºC
- Praticamente todos os dias do mês de julho apresentam temperatura entre (25-9) 14ºC e (25+9) 34ºC
- Probabilidade que uma pessoa da população tenha glicemia entre 25 e 28 é 34%

5. Distribuição Normal

5.2. Curva Normal Padronizada

(Normal reduzida)

- Se uma variável possui distribuição normal, a mudança da unidade de medida de uma variável não altera sua distribuição
- Mudança da média – a curva é movida ao longo do eixo x
- Mudança do desvio-padrão – alteração da altura e largura da curva
- Curva normal com

  média = 0 e desvio-padrão = 1 → N(0,1)

5.3. Padronização

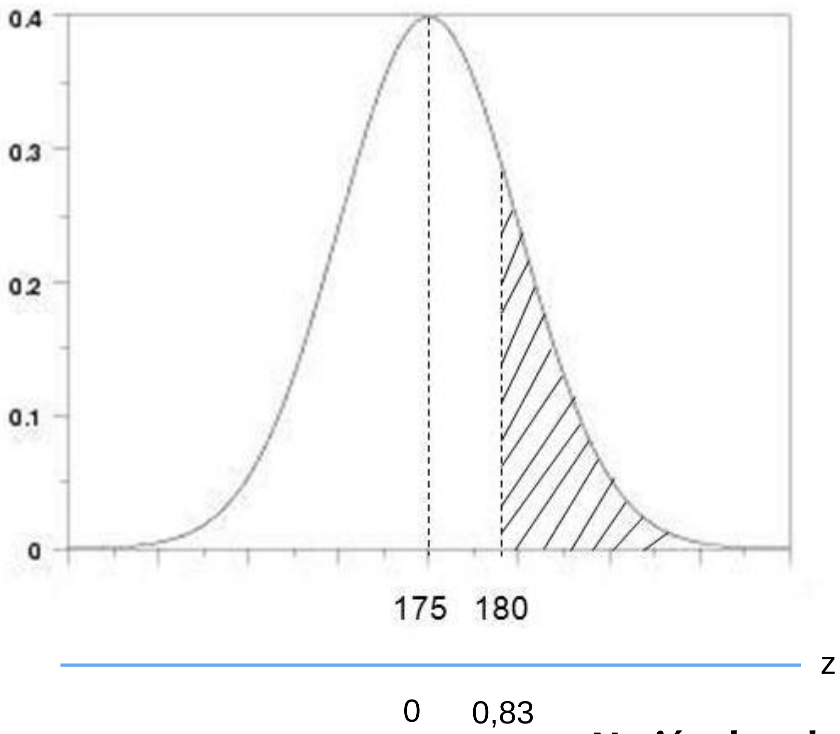
- Subtração da média e divisão pelo desvio-padrão

$$z(escore\ z) = \frac{x - \bar{x}}{s}$$

- z: diferença, em unidades de desvio-padrão, entre um valor de x e a média.

Exemplo: Seleção de jovens com no mínimo 180cm de altura

$\bar{x} = 175cm$ $\quad s = 6$ $\quad N = 140$



Variável padronizada

$$x = 175, z(escore\ Z) = \frac{x - \bar{x}}{s} = \frac{(175 - 175)}{6} = 0$$

$$x = 180, z(escore\ Z) = \frac{x - \bar{x}}{s} = \frac{(180 - 175)}{6} = 0{,}83$$

- Área entre z=0 e z=0,83 (Tabela) = 0,2967
- Área além de 0,83 (0,5 – 0,2967) = 0,2033

Conclusão

- 20,33% (28 indivíduos) da população tem altura maior que 180cm

6. Intervalo de Confiança (IC)

6.1. $IC_{95\%}$ da média

$$IC95\% = \bar{x} \pm \underbrace{1{,}96}_{z_{\alpha/2} = 0{,}025}\left(\frac{s}{\sqrt{n}}\right)$$

Interpretação

Se fossem selecionadas várias amostras independentes e aleatórias da mesma população e, para cada população, o IC95% fosse calculado, 95% dos IC conteriam a verdadeira média populacional e 5% dos IC não conteriam

6.2. $IC_{95\%}$ da diferença de duas médias

$$IC95\% = (\overline{x_1} - \overline{x_2}) \pm 1{,}96\left(\frac{s_1}{\sqrt{n_1}} + \frac{s_2}{\sqrt{n_2}}\right)$$

- Erro-padrão da diferença entre médias – combinação dos erros de x_1 e x_2

7. Teste de hipótese

7.1 Definição das hipóteses

- Sempre comparam 2 ou mais parâmetros

- Tipos de hipóteses estatísticas:

  - **Hipótese nula (H_0):** ausência de diferença entre os parâmetros (exemplo: média)

  - **Hipótese alternativa (H_1):** hipótese contrária à hipótese nula. Hipótese que o pesquisador quer confirmar

CLIMA BRONQUIOLITE

7. Teste de hipótese

7.1 Definição das hipóteses

- Sempre comparam 2 ou mais parâmetros
- Tipos de hipóteses estatísticas:
  - **Hipótese nula (H_0):** ausência de diferença entre os parâmetros (exemplo: média)
  - **Hipótese alternativa (H_1):** hipótese contrária à hipótese nula. Hipótese que o pesquisador quer confirmar

7.2. Teste de hipóteses

- Z-escore

$$z = \frac{(\overline{x_1} - \overline{x_0})}{Ep} = \frac{(\overline{x_1} - \overline{x_0})}{\left(\frac{s_1}{\sqrt{n_1}} + \frac{s_2}{\sqrt{n_0}}\right)}$$

7.3. Definição do nível de significância

- $\alpha = 0{,}05$

7.4. Determinação do valor crítico do teste

- $z_{0,05} = 1{,}96$

7.5. Decisão

- Se $|z| < z_\alpha$, não rejeita H_0
- Se $|z| \geq z_\alpha$, rejeita-se H_0

7.5. Conclusão

7. Teste de hipótese

Exemplo: Diferença do IMC médio entre homens e mulheres

- IMC aferido em 100 adultos com idade de 30-40 anos

Grupo	N	Média de IMC	S	Ep da média do IMC
Homens	n_1=45	$\overline{x_1}$ = 31,5	s_1=3,6	$Ep_1 = {3,6}/{\sqrt{45}} = 0,53$
Mulheres	n_2=55	$\overline{x_1}$ = 28,7	s_2=5,7	$Ep_2 = {5,7}/{\sqrt{55}} = 0,76$

1. Definição das hipóteses

- **H_0:** O IMC médio dos homens **é igual** ao IMC médio das mulheres; H_0: $\overline{x_1} = \overline{x_2}$ ou H_0: $\overline{x_1} - \overline{x_2} = 0$
- **H_1:** O IMC médio dos homens **é diferente** ao IMC médio das mulheres; H_1: $\overline{x_1} \neq \overline{x_2}$ ou H_1: $\overline{x_1} - \overline{x_2} \neq 0$

7. Teste de hipótese

Exemplo: Diferença do IMC médio entre homens e mulheres

- IMC aferido em 100 adultos com idade de 30-40 anos

Grupo	N	Média de IMC	s	Ep da média do IMC
Homens	n_1=45	$\overline{x_1} = 31{,}5$	s_1=3,6	$Ep_1 = {}^{3{,}6}/_{\sqrt{45}} = 0{,}53$
Mulheres	n_2=55	$\overline{x_1} = 28{,}7$	s_2=5,7	$Ep_2 = {}^{5{,}7}/_{\sqrt{55}} = 0{,}76$

1. Definição das hipóteses

- **H_0:** O IMC médio dos homens **é igual** ao IMC médio das mulheres; H_0: $\overline{x_1} = \overline{x_2}$ ou H_0: $\overline{x_1} - \overline{x_2} = 0$
- **H_1:** O IMC médio dos homens **é diferente** ao IMC médio das mulheres; H_1: $\overline{x_1} \neq \overline{x_2}$ ou H_1: $\overline{x_1} - \overline{x_2} \neq 0$

2. Teste de hipóteses

$$z = \frac{(\overline{x_1} - \overline{x_0})}{Ep} = \frac{(\overline{x_1} - \overline{x_0})}{(\frac{s_1}{\sqrt{n_1}} + \frac{s_2}{\sqrt{n_0}})} = \frac{31{,}5 - 28{,}7}{(\frac{3{,}6}{\sqrt{45}} + \frac{5{,}7}{\sqrt{55}})} = \frac{2{,}8}{0{,}53 + 0{,}76} = \frac{2{,}8}{1{,}29} = 2{,}17$$

3. Definição do nível de significância

- α=0,05

4. Determinação do valor crítico do teste

- $z_{0,05}$=1,96

7. Teste de hipótese

Exemplo: Diferença do IMC médio entre homens e mulheres

5. Decisão

- Se $|z| < z_\alpha$, não rejeita H_0
- Se $|z| \geq z_\alpha$, rejeita-se H_0
- **Como $|2{,}17| > 1{,}96$, rejeita-se H_0**

6. Conclusão

O **IMC médio de homens difere significativamente do IMC médio das mulheres**, portanto a média de IMC dos dois grupos não são iguais. O IMC médio dos homens é maior do que o IMC médio de mulheres, para um nível de significância de 0,05.

7. Teste de hipótese

7.5. Teste de hipóteses para amostras pequenas (n<30)

- Baseado na distribuição t
- Robusto à violação do pressuposto de normalidade para amostras pequenas
- Também utilizado quando as duas populações comparadas possuem o mesmo desvio-padrão

Desvio-padrão "comum"

$$s = \sqrt{[\frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_2^2}{(n_1 + n_0 - 2)}]}$$

7. Teste de hipótese

7.6. Erro-padrão

$$\mathrm{Ep} = \mathrm{s}\sqrt{\left(\frac{1}{\mathrm{n}_1} + \frac{1}{\mathrm{n}_2}\right)}$$

7.7. Teste t

$$t_{\alpha,(n_1+n_2-2)} = \frac{(\overline{x_1} - \overline{x}_0)}{Ep} = \frac{(\overline{x_1} - \overline{x}_0)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

7.8. $IC_{95\%}$ – baseado na distribuição t

$$IC = (\overline{x_1} - x_2) \pm (t_{\alpha,(n_1+n_2-2)} \times Ep)$$

7. Teste de hipótese

7.9. Comparação de duas variâncias – estatística F

- Baseado na distribuição F
- Grau de liberdade (numerador e denominador): n-1

$$F_{\alpha,glN,glD} = \frac{s_{maior}^2}{s_{menor}^2}$$

8. Valor de p

8.1. Definição

Considerando que a hipótese nula é verdadeira, o valor de p é a **probabilidade** de se obter uma diferença da média de dois grupos tão grande ou maior que a diferença observada

8.2. Interpretação

- ↑ **valor de p (≥0,05):** os dados não fornecem evidência contrária à hipótese nula: existe probabilidade de que a diferença observada foi devido à variação amostral
- ↓ **valor de p (<0,05):** uma diferença tão grande quanto a observada é pouco provável de ocorrer se a hipótese nula for verdadeira; logo, existe forte evidência contrária à hipótese nula

8. Valor de p x IC95%

8.3. $IC_{95\%}$

- Intervalo de valores no qual apresentamos confiança razoável de que o parâmetro da população se encontra
- Se o $IC_{95\%}$ não contem a nulidade – o valor de p **provavelmente** é <0,05
- Depende do tamanho amostral – quanto ↓ amostra, ↑ o $IC_{95\%}$

8.4. Valor de p

- Força de evidência contrária à hipótese nula que o parâmetro na população é =0

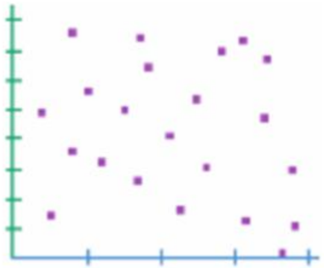
Correlação

1. Coeficiente de correlação (r)

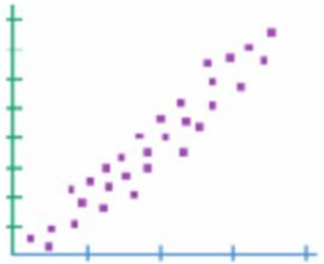
Mede a força da associação **linear** entre duas variáveis contínuas X e Y

- Não possui unidade de medida
- Fornece a direção (positiva ou negativa) da relação linear
- $-1 \leq r \leq +1$
- Antes de avaliar o valor do coeficiente de regressão (r), é importante observar o **diagrama de dispersão**

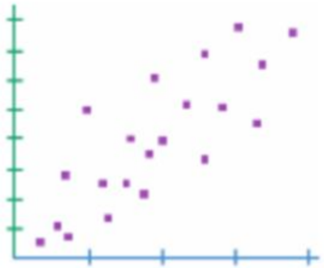
1. Coeficiente de correlação (r)



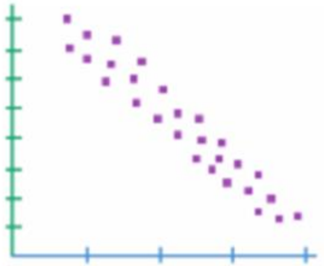
Sem correlação



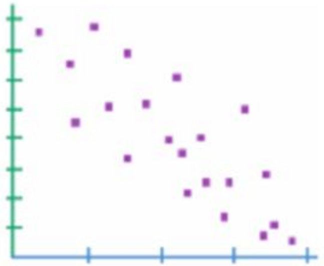
Correlação positiva forte



Correlação positiva fraca



Correlação negativa forte

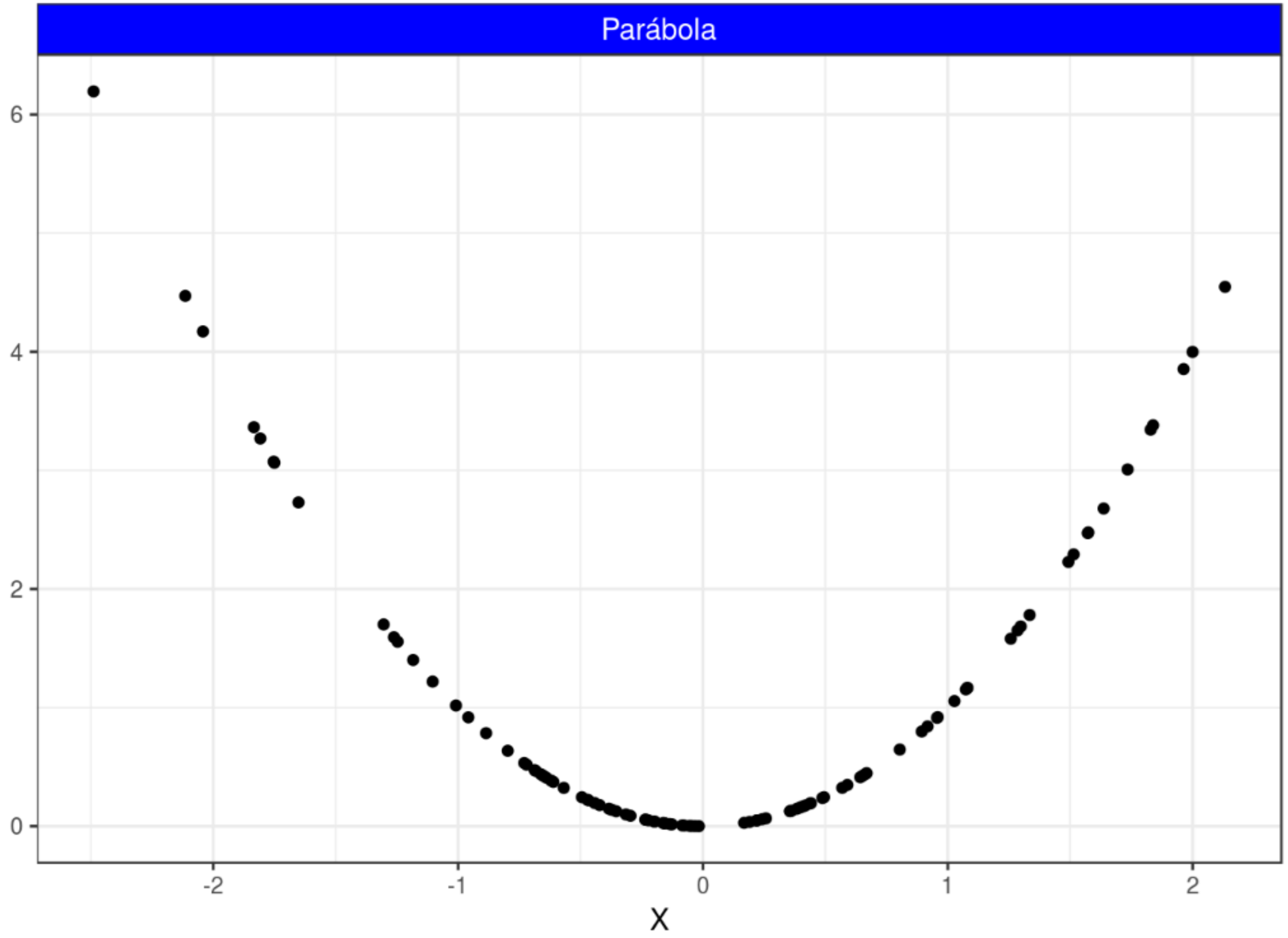


Correlação negativa fraca

- Relação linear perfeita: se conheço x, estimo y e vice-versa
- Relação Positiva: X↑ quando Y↑ ou X↓ quando Y↓
- Relação negativa: X↑ quando Y ↓ ou X ↑ quando Y↓

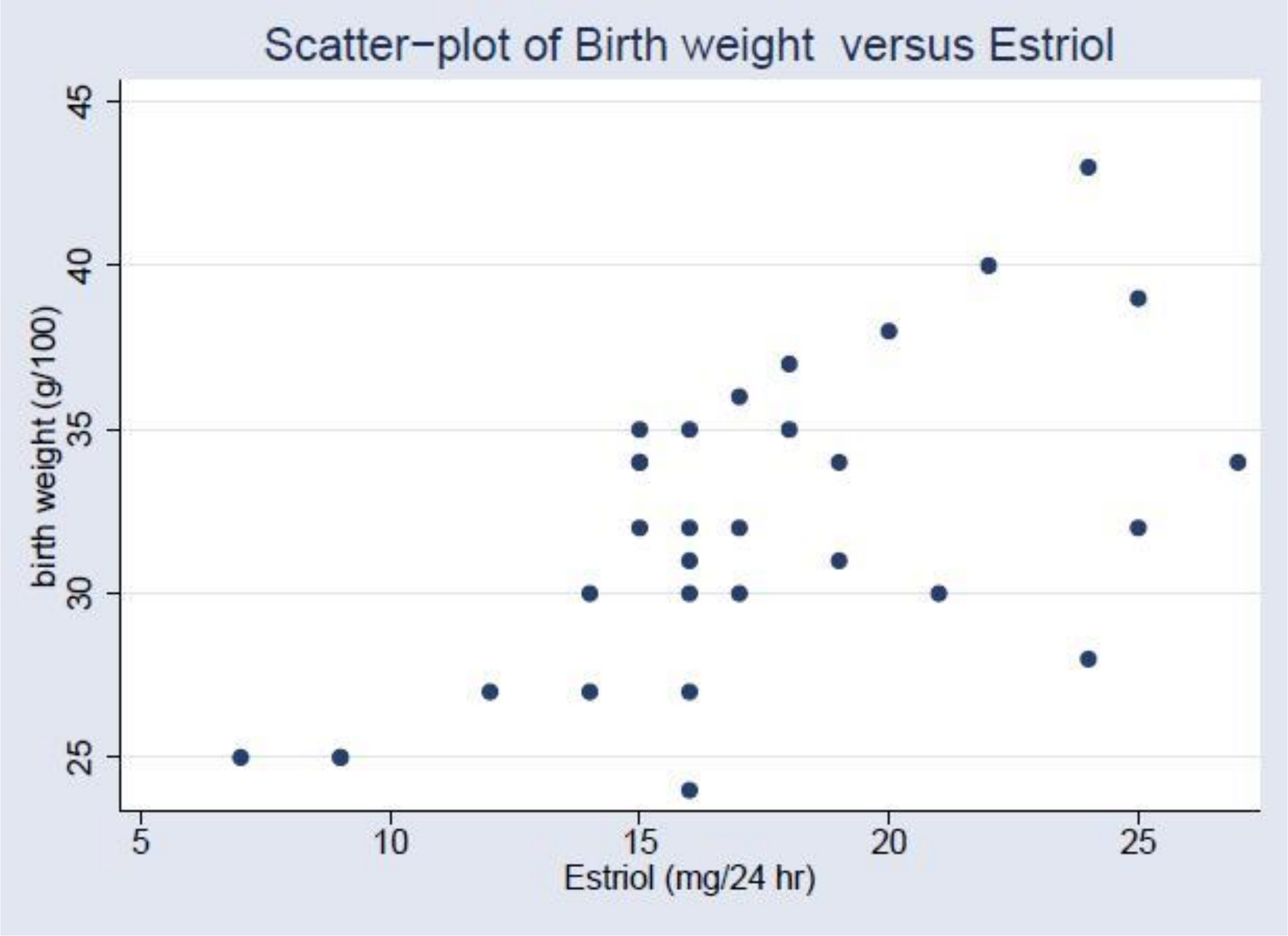
1. Coeficiente de correlação (r)

Coeficiente de correlação= 0, sem relação linear

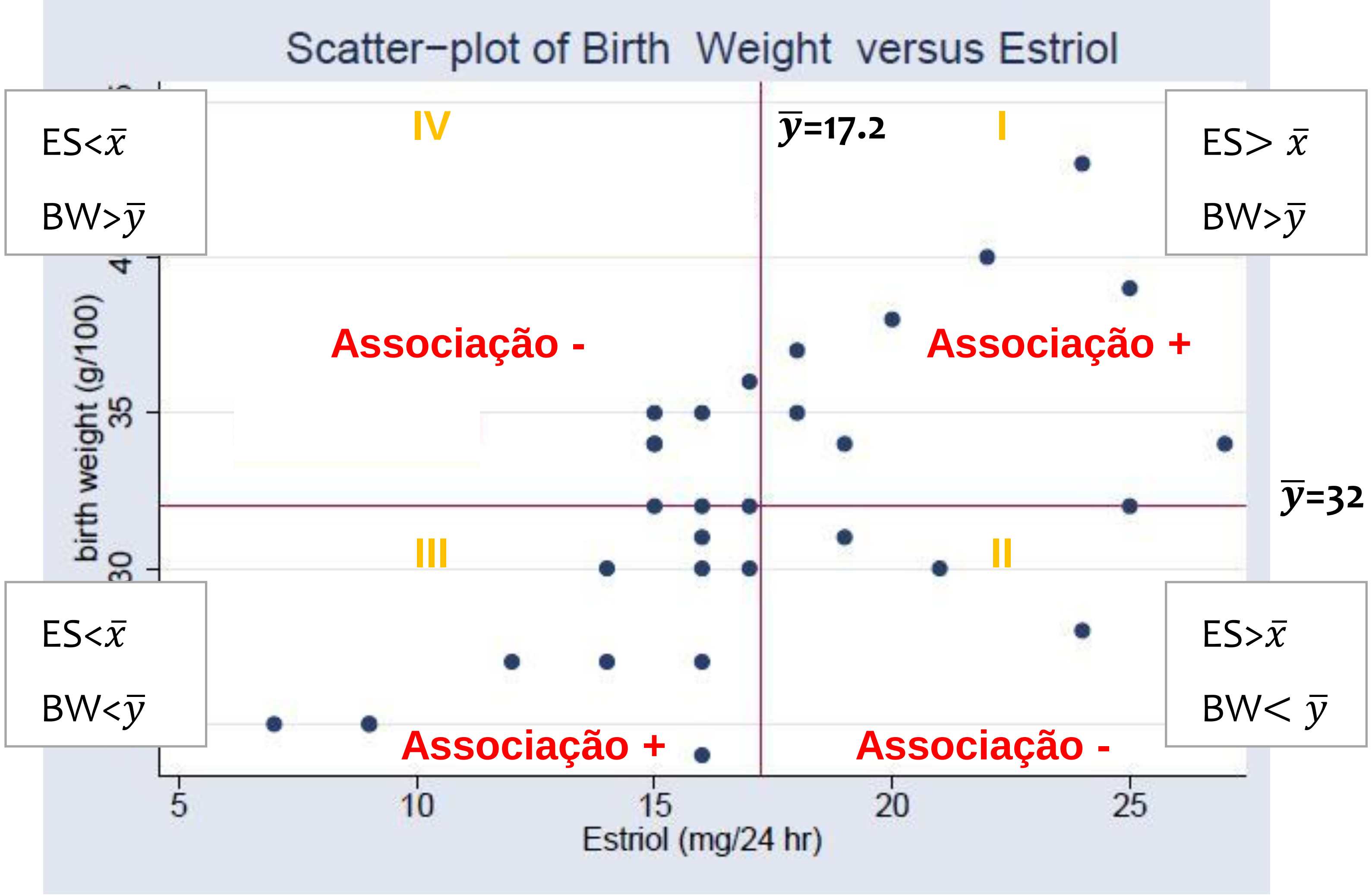


Atenção: r = 0 não significa que x e y não estão relacionados, mas que a linha reta não é útil para descrever a relação de x e y

1. Coeficiente de correlação (r)



1. Coeficiente de correlação (r)



1. Coeficiente de correlação (r)

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{[\sum(x-\overline{x})^2 \sum(y-\overline{y})^2]}}$$

r = coeficiente de correlação

x =exposição

$\bar{x}$ = média da variável de exposição

y = desfecho

$\bar{y}$ = média da variável de desfecho

1. Coeficiente de correlação (r)

1.1 Valores de referência

Valor de *r*	Classificação
0 – 0,25	Correlação baixa
0,26 – 0,50	Correlação moderada
0,51 – 0,75	Correlação Boa
> 0,75	Correlação excelente

1. Coeficiente de correlação (r)

1.2. limitações

Se a relação entre as variáveis x e y não forem linear, não fornece uma medida válida de relação

- Sensível a *outliers*
- Necessidade de distribuição simétrica entre as duas variáveis
- **Forte correlação não necessariamente significa uma relação causal**

1. Coeficiente de correlação (r)

1.3. Coeficiente de correlação de Pearson

- x e y possuem distribuição normal

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{S_x}\right)\left(\frac{y_i - \overline{y}}{S_y}\right)$$

r = coeficiente de correlação

n= tamanho amostral

x_i= valor de x para cada observação i

$\bar{x}$= média das observações i de x

y_i= valor de y para cada i observação de y

$\bar{y}$ = média das i observações de y

S_x = desvio-padrão de x

S_y = desvio-padrão de y

1. Coeficiente de correlação (r)

1.4. Coeficiente de correlação de Spearman

- x e y **não** possuem distribuição normal
- Técnica não-paramétrica
- Ordenação de x e y por valores (rank)

$$r_s = \frac{\sum_{i=1}^{n}(x_{ri} - \bar{x}_r)(y_{ri} - \bar{y}_r)}{\sqrt{[\sum_{i=1}^{n}(x_{ri} - \bar{x}_r)^2][\sum_{i=1}^{n}(y_{ri} - \bar{y}_r)^2]}}$$

r_s = coeficiente de correlação de Spearman

n= tamanho amostral

x_{ri}= valor de rank de x para cada observação i

$\overline{x_r}$ = média das i observações de x

y_{ri}= valor de rank de y para cada observação i de y

$\overline{y_r}$ = média das observações i de y

S_x = desvio-padrão de x

S_y = desvio-padrão de y

Referências

1. Kirkwood BR, Sterne JAC. Essential Medical Statistics, 2nd Edition. Blackwell Science.
2. Callegari-Jacques SM. Bioestatística – Princípios e Aplicações. Artmed, 2003.
3. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Data Regression methods in biostatistics : linear, logistic, survival, and repeated measures models. Springer, 2005.
4. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied Linear Statistical Models. McGraw-Hill, 2005.
5. Rosner B. Fundamentals in Biostatistics. Brooks/Cole, 2011.

Obrigado!

felipetmf@gmail.com

CNPq

FIOCRUZ

Sala de Situação de Saúde FS-UnB

SUS

MINISTÉRIO DA SAÚDE

GOVERNO FEDERAL

BRASIL

UNIÃO E RECONSTRUÇÃO

CLIMA

BRONQUIOLITE