

# Estatística

## Análise gráfica

---

Dra. Agatha Rodrigues

[agatha.srodrigues@gmail.com](mailto:agatha.srodrigues@gmail.com)

Agradecimento ao professor Dr. Alessandro Sarnaglia por disponibilizar alguns slides utilizados nessa apresentação

# Análise gráfica

---

# Análise descritiva

O que fazer com as observações que coletamos?

## Estatística descritiva

Resumir os dados;  
Entender os dados.

## Como fazer?

- Tabelas de frequências.
- Medidas descritivas ou resumo
  - Medidas de posição
  - Medidas de dispersão

Depende do tipo de variável

## Resumindo...

### Medidas de posição

Mínimo, Máximo, Moda, Média, Mediana, Percentis.

### Medidas de dispersão

Amplitude, Intervalo interquartil, Variância, Desvio Padrão, Coeficiente de Variação.

# Análise descritiva

	Grupo	Média	Mediana	Variância	Desvio padrão	Mínimo	Máximo	Amplitude interquartil
Medida do Colo	Placebo	22,63	21,80	101,81	10,09	5,83	43,40	9,75
	Progesterona	25,79	28,60	130,25	11,41	6,60	46,30	19,20
Nº de contrações	Placebo	5,23	5,00	9,36	3,06	1,00	10,00	6,00
	Progesterona	3,88	3,00	6,86	2,62	1,00	9,00	5,00
IG do parto	Placebo	36,02	36,43	5,39	2,32	29,43	38,29	2,50
	Progesterona	36,17	37,00	5,78	2,40	30,29	38,43	3,89

É possível resumir os dados observados por meio de gráficos!

**É possível resumir os dados observados por meio de gráficos!**

## Por que fazer um gráfico?

- Pode ser a maneira mais adequada para resumir e apresentar um conjunto de dados.
- Tem a vantagem de facilitar a compreensão de uma determinada situação que queira ser descrita, permitindo uma interpretação rápida e visual das suas principais características.

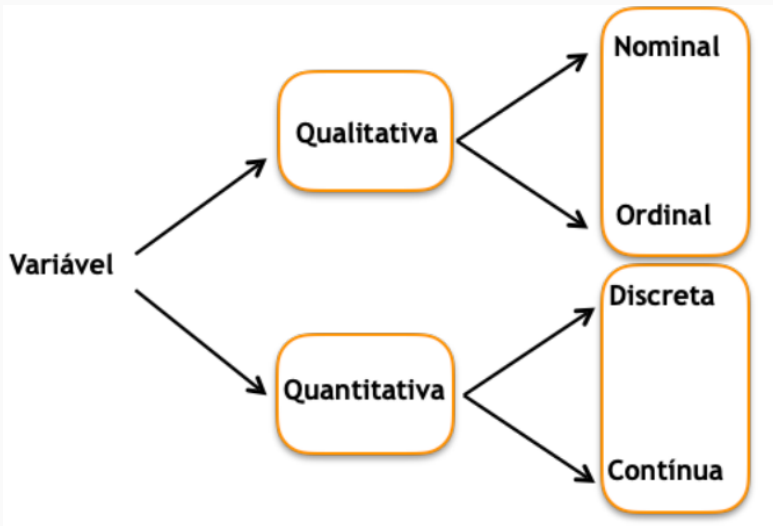
**É possível resumir os dados observados por meio de gráficos!**

## Por que fazer um gráfico?

- Pode ser a maneira mais adequada para resumir e apresentar um conjunto de dados.
- Tem a vantagem de facilitar a compreensão de uma determinada situação que queira ser descrita, permitindo uma interpretação rápida e visual das suas principais características.

## Qual o melhor gráfico?

Depende do tipo de variável de interesse e também da informação que deseja ser transmitida.





1. Gráficos para Variáveis Qualitativas
2. Gráficos para Variáveis Quantitativas
3. Gráficos para duas Variáveis
  - Qualitativa e Qualitativa
  - Qualitativa e Quantitativa
  - Quantitativa e Quantitativa

## 1. Gráficos para Variáveis Qualitativas

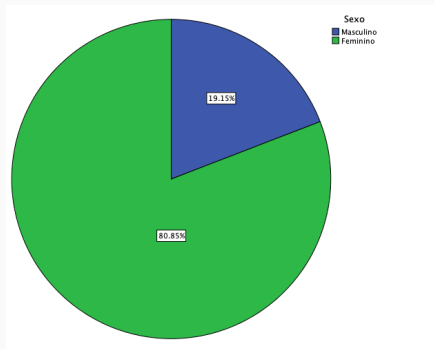
### Exemplo: Sexo dos alunos

Sexo	Freq. absoluta	Freq. relativa (%)
Masculino	9	19,1
Feminino	38	80,9
Total	47	100,0

## Variáveis qualitativas

### 1. Gráfico de setores (pizza)

Um círculo é dividido em tantos setores quantas forem as categorias da variável. A área de cada setor é proporcional à frequência relativa (ou absoluta) da categoria.

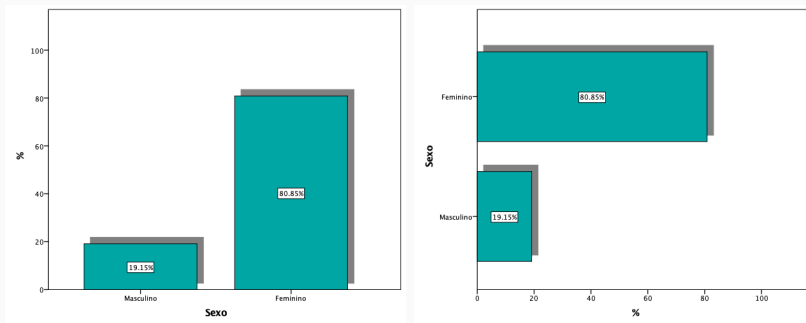


**Figura 1:** Gráfico em setores para a variável Sexo dos alunos

## Variáveis qualitativas

### 2. Gráficos de barras

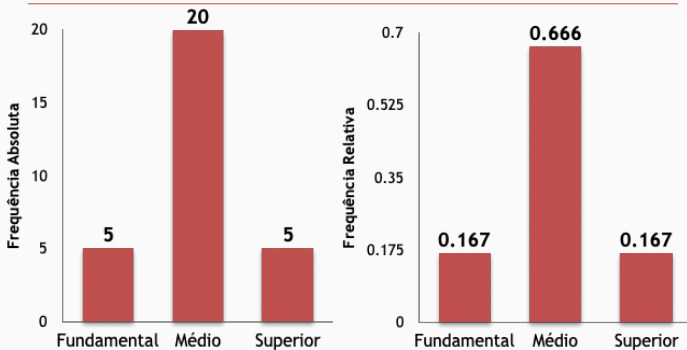
Para cada categoria da variável é desenhado um retângulo na vertical (ou horizontal) e a altura (comprimento) do retângulo corresponde à frequência absoluta ou relativa da categoria.



**Figura 2:** Gráfico em barras (vertical e horizontal) para o Sexo dos alunos

**Tabela 1.** Frequências dos 30 pacientes de **baixo risco** da Obstetrícia segundo o grau de instrução.

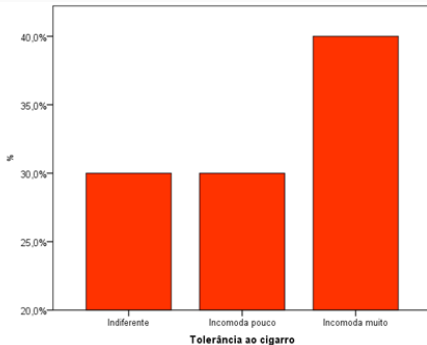
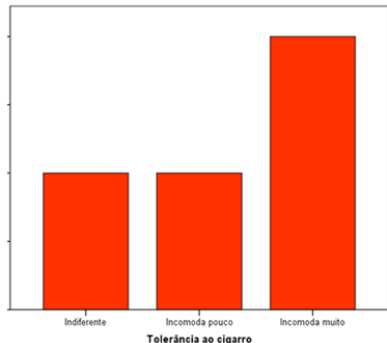
Escolaridade	Frequência absoluta	Frequência relativa	%
Fundamental	5	0,167	16,7
<b>Médio</b>	20	0,666	66,6
Superior	5	0,167	16,7
Total	30	1	100



**Figura 3:** Gráfico em barras para a variável Escolaridade para baixo risco - frequência absoluta (à esquerda) e frequência relativa (à direita).

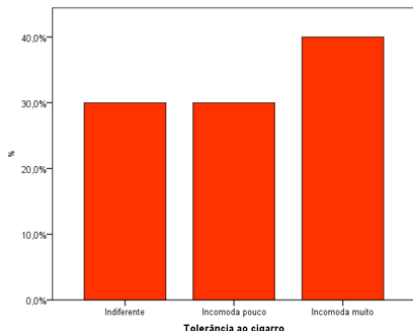
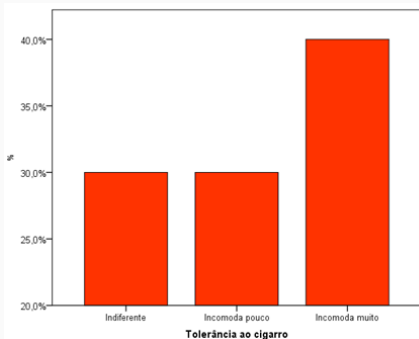
**Tabela 2.** Tabela de frequências da tolerância ao cigarro dos alunos do curso de Estatística.

Tolerância ao cigarro	n	%
Indiferente	9	30,0
Incomoda pouco	9	30,0
Incomoda muito	12	40,0
Total	30	100



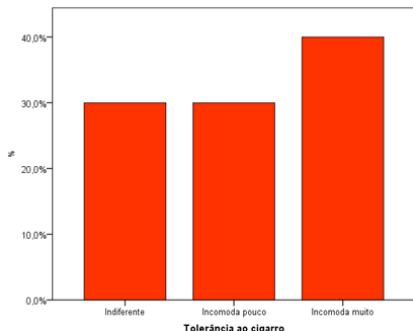
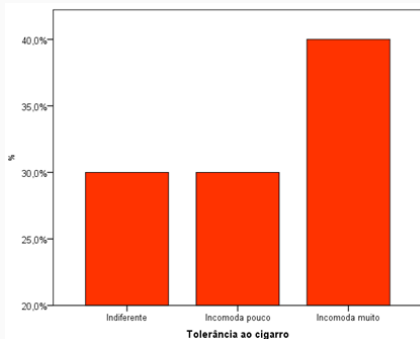
**Tabela 2.** Tabela de frequências da tolerância ao cigarro dos alunos do curso de Estatística.

Tolerância ao cigarro	n	%
Indiferente	9	30,0
Incomoda pouco	9	30,0
Incomoda muito	12	40,0
Total	30	100



**Tabela 2.** Tabela de frequências da tolerância ao cigarro dos alunos do curso de Estatística.

Tolerância ao cigarro	n	%
Indiferente	9	30,0
Incomoda pouco	9	30,0
Incomoda muito	12	40,0
Total	30	100



CAUTION WITH THE SCALE!



## 2. Gráficos para Variáveis Quantitativas

### Exemplo: IMC dos alunos

Variável	n	Min	Média	Mediana	Max	Dp	CV
IMC	46	17,75	23,04	22,28	33,39	3,65	15,8%

## 2. Gráficos para Variáveis Quantitativas

### Exemplo: IMC dos alunos

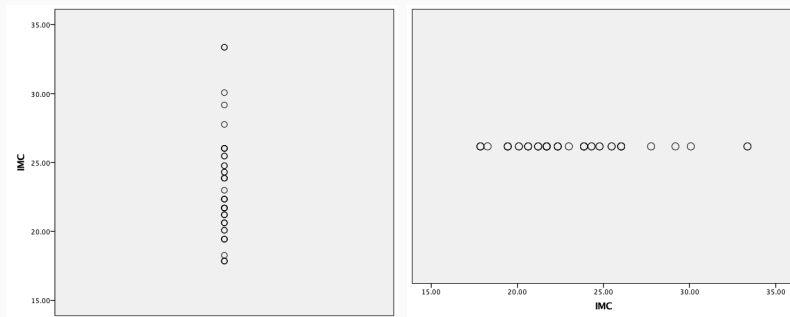
Variável	n	Min	Média	Mediana	Max	Dp	CV
IMC	46	17,75	23,04	22,28	33,39	3,65	15,8%

Faz sentido fazer um gráfico de setores para a variável IMC?

## Variáveis quantitativas

### 1. Gráfico de pontos ou *Dotplot*

O *Dotplot* representa cada observação obtida em uma escala horizontal (ou vertical), permitindo visualizar a distribuição dos dados ao longo deste eixo.

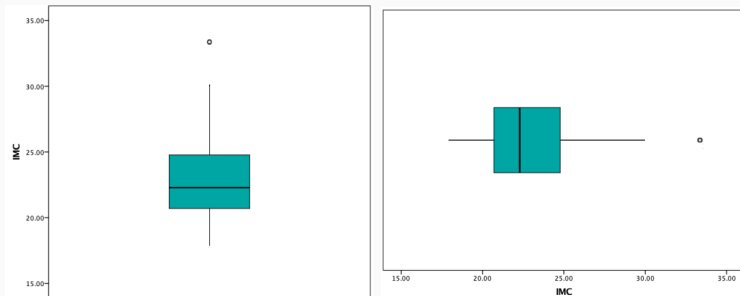


**Figura 4:** Gráfico dotplot vertical (à esquerda) e horizontal (à direita) do IMC dos alunos.

## Variáveis quantitativas

### 2. Gráfico de caixa ou *Boxplot*

O *Boxplot* representa através de um retângulo o 1º quartil (Q1), mediana e 3º quartil (Q3).



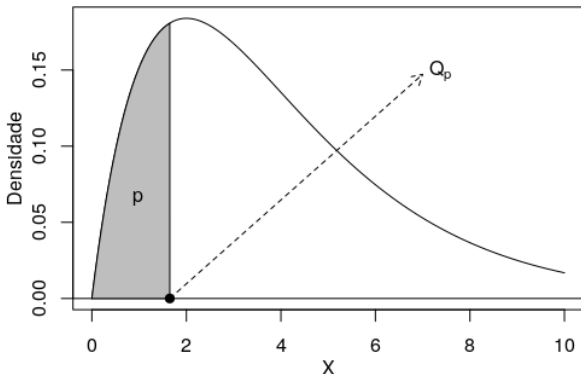
**Figura 5:** Gráfico boxplot vertical (à esquerda) e horizontal (à direita) do IMC dos alunos.

# Relembrando - Quantil

O que são?

Medidas baseadas na ordenação dos dados!

O quantil  $Q_p$  é o valor que separa os  $100p\%$  menores valores dos  $100(1 - p)\%$  maiores valores.



# Relembrando - Quantil

Quantis especiais::

**Percentil  $j$  ( $P_j$ ),  $j = 1, 2, \dots, 99$**  é o quantil  $Q_{j/100}$ :

$$P_1 = Q_{0.01}, P_2 = Q_{0.02}, \dots, P_{99} = Q_{0.99},$$

isto é, separatrizes de 1% em 1%.

**Decil  $j$  ( $D_j$ ),  $j = 1, 2, \dots, 9$**  é o quantil  $Q_{j/10}$ :

$$D_1 = Q_{0.1}, D_2 = Q_{0.2}, \dots, D_9 = Q_{0.9},$$

isto é, separatrizes de 10% em 10%.

**Quartil  $j$  ( $Q_{rj}$ ),  $j = 1, 2, 3$**  é o quantil  $Q_{j/4}$ :

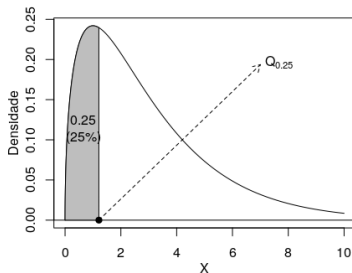
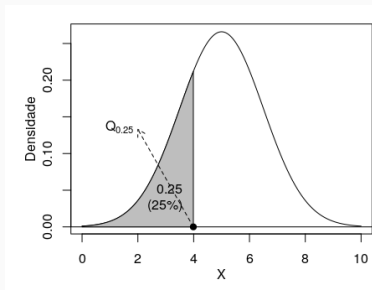
$$Q_{r1} = Q_{0.25}, Q_{r2} = Q_{0.5}, Q_{r3} = Q_{0.75},$$

isto é, separatrizes de 25% em 25%.

Note que  $Q_{0.5} = P_{50} = D_5 = Q_{r2} = \text{Mediana}$ .

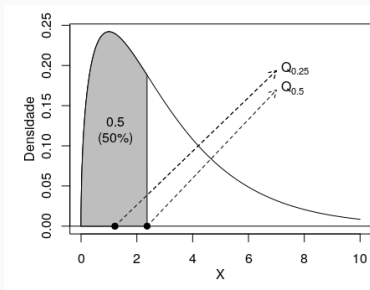
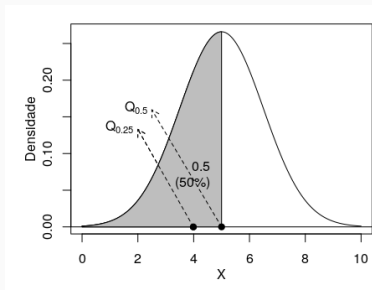
# Relembrando - Quantil

Os quartis fornecem informações fundamentais sobre a distribuição dos dados.



# Relembrando - Quantil

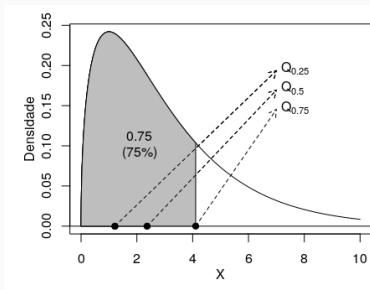
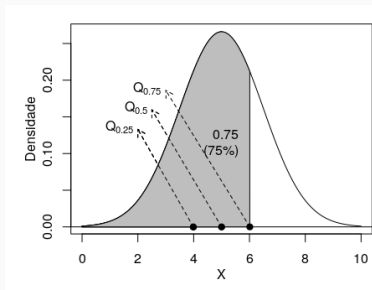
Os quartis fornecem informações fundamentais sobre a distribuição dos dados.





# Relembrando - Quantil

Os quartis fornecem informações fundamentais sobre a distribuição dos dados.



## Variáveis quantitativas

### 2. Gráfico de caixa ou *Boxplot*

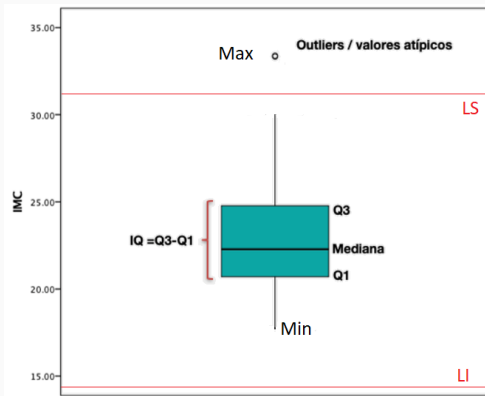
#### Utilidade

Da uma ideia de ...

1. posição: a posição central é dada pela mediana;
2. dispersão: a medida de dispersão é dada por  $IQ = Q3 - Q1$ , em que  $IQ$  é o **intervalo-interquartil**;
3. caudas: avaliadas pelos
  - Limite inferior  $LI = Q1 - 1,5 * IQ$ , e
  - Limite superior  $LS = Q3 + 1,5 * IQ$ ;
4. dados discrepantes, atípicos ou *outliers*: são considerados valores atípicos aqueles menores que  $LI$  e maiores que  $LS$ .
5. assimetria: a assimetria é examinada pelas caudas e dados discrepantes.

## Variáveis quantitativas

### 2. Gráfico de caixa ou *Boxplot*



#### LS

$$\max(IMC) = 33,39$$

$$LS = Q3 + 1,5 * IQ$$

$$Q3 = 24,9 \text{ e } IQ = 4,27$$

$$LS = 24,9 + 1,5 * 4,27 = 31,30$$

#### LI

$$\min(IMC) = 17,75$$

$$LI = Q1 - 1,5 * IQ$$

$$Q1 = 20,63 \text{ e } IQ = 4,27$$

$$LI = 20,63 - 1,5 * 4,27 = 14,22$$

**Figura:** Gráfico boxplot do IMC dos alunos.

Os cinco valores:  $x_{(1)}$  (mínimo),  $Q_1$ ,  $Q_2$ ,  $Q_3$  e  $x_{(n)}$  (máximo) são importantes para se ter uma boa ideia da assimetria da distribuição dos dados.

Para uma distribuição simétrica ou aproximadamente simétrica, deveríamos ter:

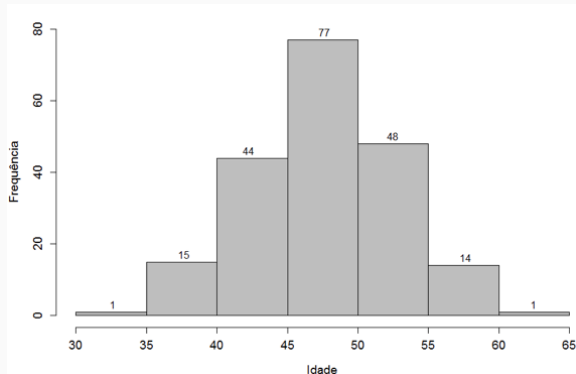
- $Q_2 - x_{(1)} \approx x_{(n)} - Q_2$ ;
- $Q_2 - Q_1 \approx Q_3 - Q_2$ ;
- $Q_1 - x_{(1)} \approx x_{(n)} - Q_3$ ;

A diferença  $Q_2 - x_{(1)}$  é chamada *dispersão inferior* e  $x_{(n)} - Q_2$  é a *dispersão superior*.

## Variáveis quantitativas

### 3. Histograma

É um gráfico de retângulos contíguos de base dada pelas faixas de valores da variável quantitativa.

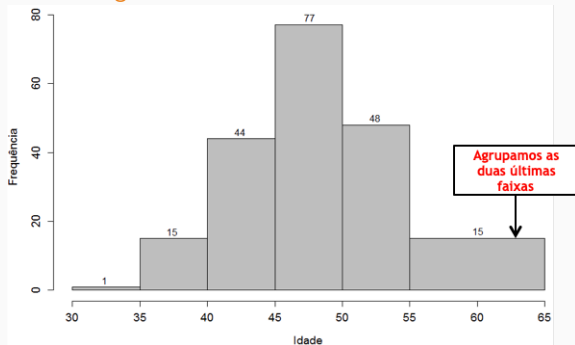


Neste histograma, no eixo y (eixo vertical) está a frequência das faixas de valores do eixo x (eixo horizontal).

**Figura 6:** Histograma da variável Idade (n=200).

## Variáveis quantitativas

### 3. Histograma

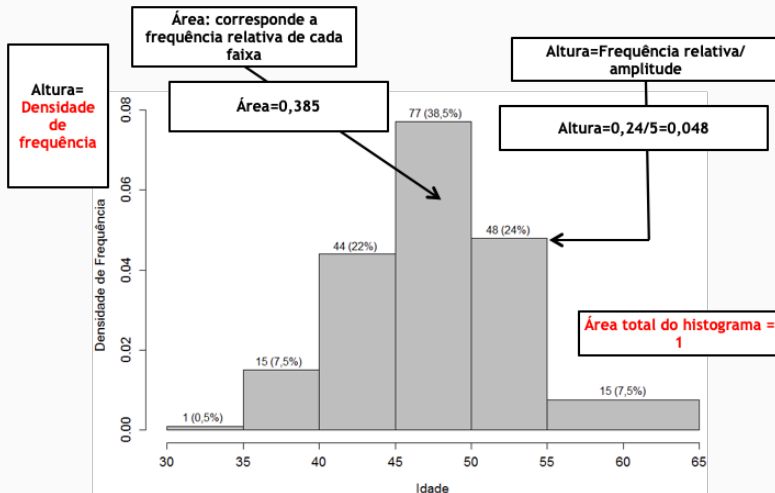


**Figura 7:** Histograma da variável Idade agrupando as duas últimas faixas.

**Solução:** Usar a *densidade de frequência*, dada por: frequência relativa/amplitude.

- Em alguns casos, pode ser necessário utilizar classes de **amplitudes diferentes**.
- Note que a faixa [35,40) tem a mesma frequência de [55,65]. Mas é injusto ter a mesma altura no histograma pois as amplitudes são diferentes.

## 3. Histograma

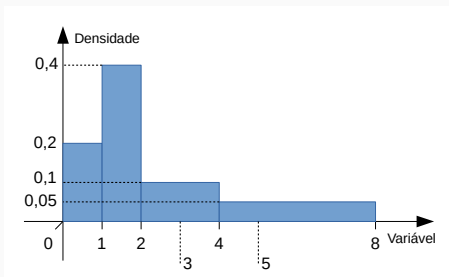


**Figura 8:** Histograma da variável Idade (n=200) com densidade de frequência no eixo y.

## Variáveis quantitativas

### 3. Histograma

A probabilidade de estar em determinada região é aproximada pela área do histograma sobre essa região.

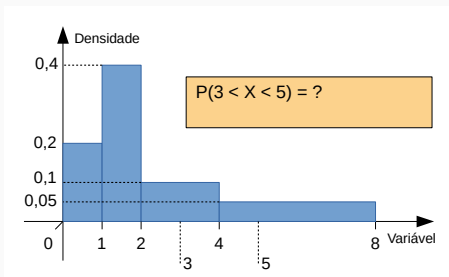




## Variáveis quantitativas

### 3. Histograma

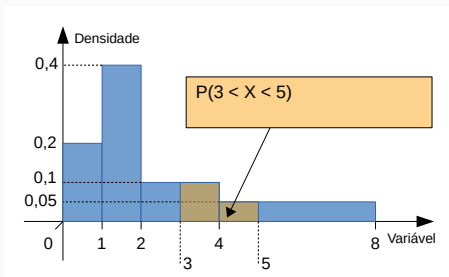
A probabilidade de estar em determinada região é aproximada pela área do histograma sobre essa região.



## Variáveis quantitativas

### 3. Histograma

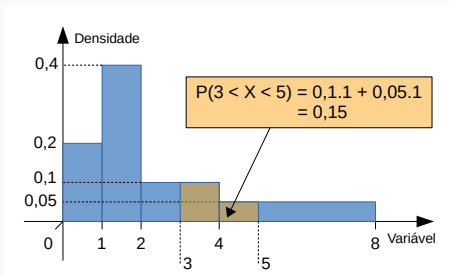
A probabilidade de estar em determinada região é aproximada pela área do histograma sobre essa região.



## Variáveis quantitativas

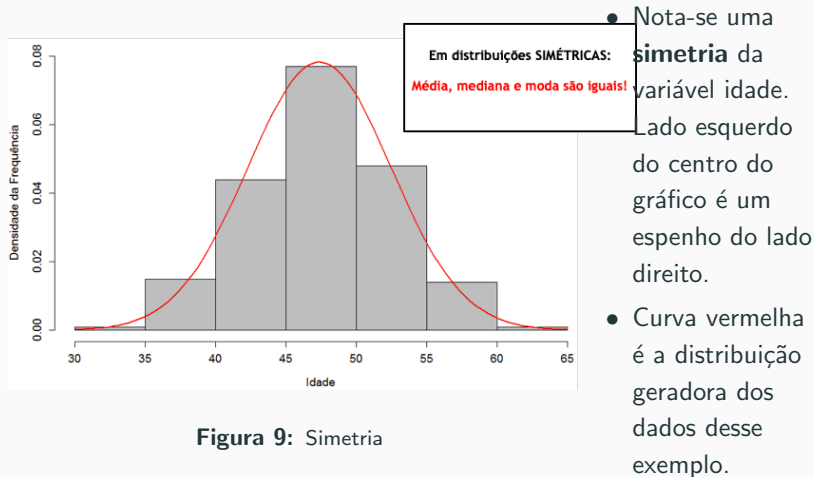
### 3. Histograma

A probabilidade de estar em determinada região é aproximada pela área do histograma sobre essa região.



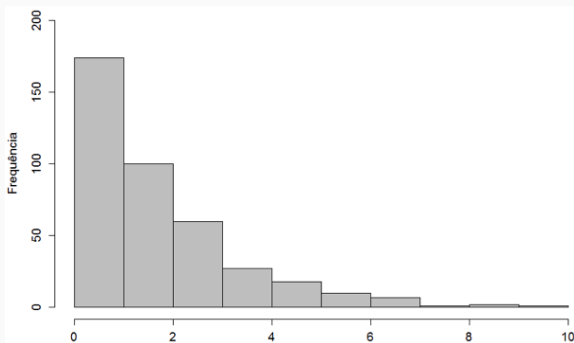
## Variáveis quantitativas

### 3. Histograma



## Variáveis quantitativas

### 3. Histograma

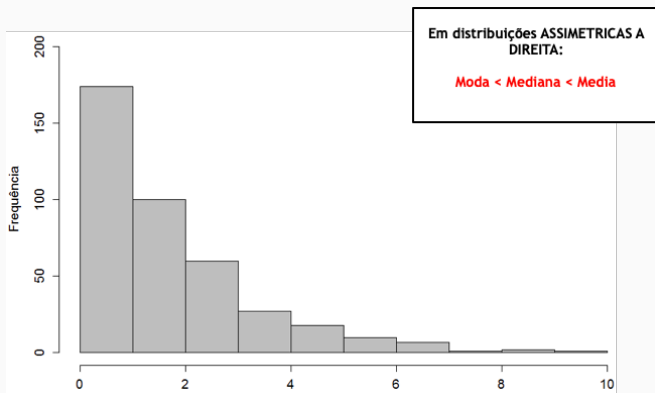


**Figura 10:** Assimetria à direita

Observe que os valores da variável mais frequentes são os mais baixos, com baixa ocorrência dos maiores valores.

## Variáveis quantitativas

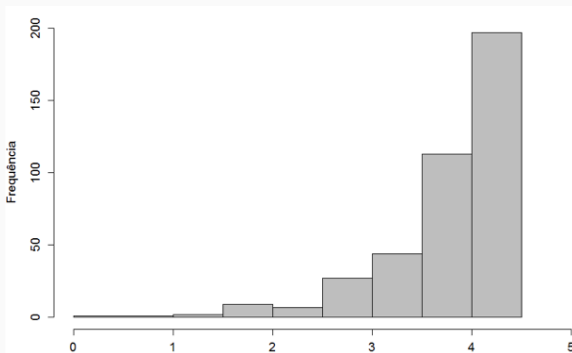
### 3. Histograma



**Figura 11:** Assimetria à direita

## Variáveis quantitativas

### 3. Histograma



**Figura 12:** Assimetria á esquerda

Note que os valores da variável mais frequentes são os mais altos, com baixa ocorrência dos menores valores.

## Variáveis quantitativas

### 3. Histograma

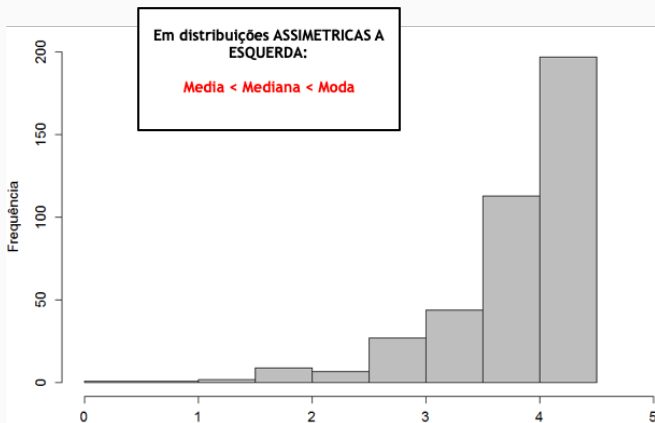
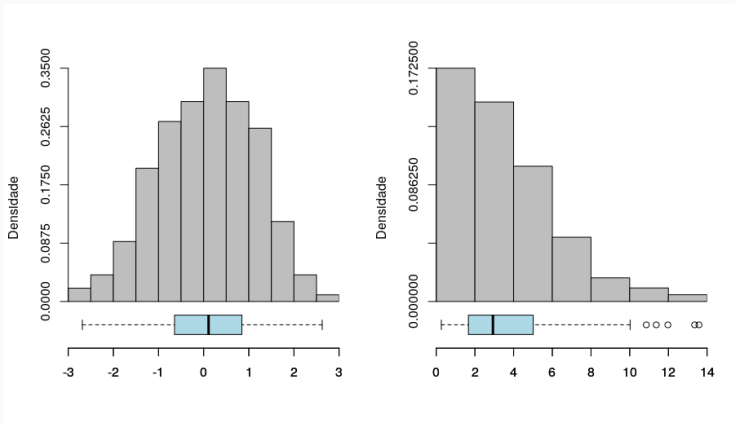


Figura 13: Assimetria á esquerda



## Relação entre boxplot e histograma



Perceba que no cenário de distribuição assimétrica à direita, a mediana está mais próxima do Q1 que do Q3 e o Q1 está mais próximo do mínimo que Q3 do máximo.

## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 3: Tabela de contingência de sexo e consumo de café (% pelo total).

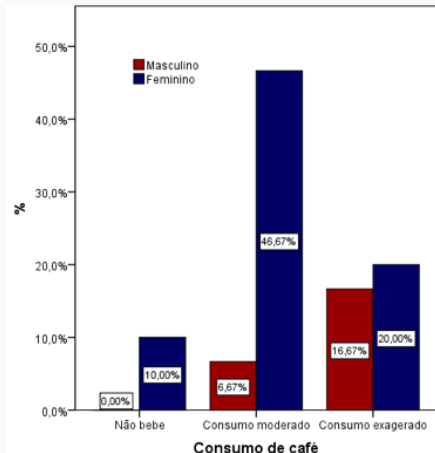
Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 6.66%	5 16.67%	7
Feminino	3 10.00%	14 46.67%	6 20.00%	23
Total	3	16	11	30 <b>100.00%</b>

## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 3: Tabela de contingência de sexo e consumo de café (% pelo total).

Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 6.66%	5 16.67%	7
Feminino	3 10.00%	14 46.67%	6 20.00%	23
Total	3	16	11	30 100.00%



## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 4: Tabela de contingência de sexo e consumo de café (% por consumo).

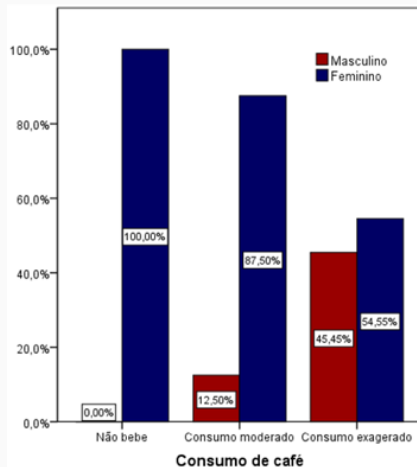
Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 12.50%	5 45.45%	7
Feminino	3 100.00%	14 87.50%	6 54.55%	23
Total	3 100.00%	16 100.00%	11 100.00%	30

## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 4: Tabela de contingência de sexo e consumo de café (% por consumo).

Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 12.50%	5 45.45%	7
Feminino	3 100.00%	14 87.50%	6 54.55%	23
Total	3 100.00%	16 100.00%	11 100.00%	30



## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 5: Tabela de contingência de sexo e consumo de café (% por sexo).

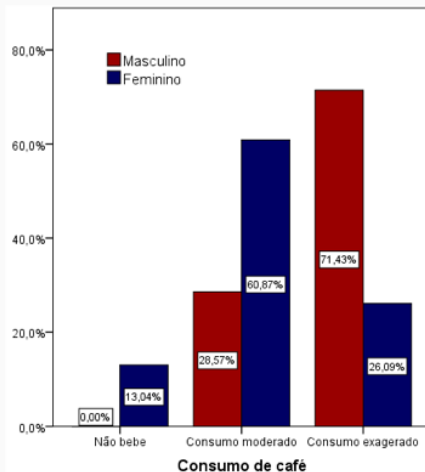
Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 28.57%	5 71.43%	7 <b>100.00%</b>
Feminino	3 13.04%	14 60.87%	6 26.09%	23 <b>100.00%</b>
Total	3	16	11	30

## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 5: Tabela de contingência de sexo e consumo de café (% por sexo).

Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 28.57%	5 71.43%	7 <b>100.00%</b>
Feminino	3 13.04%	14 60.87%	6 26.09%	23 <b>100.00%</b>
Total	3	16	11	30

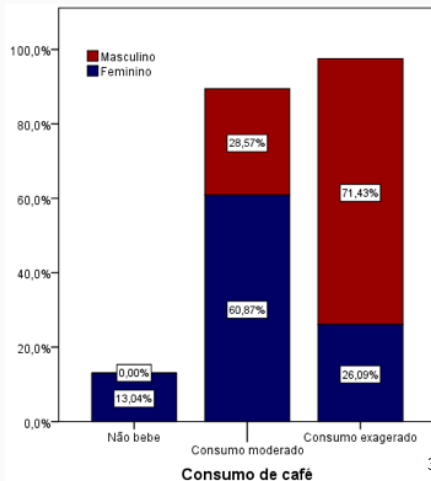


## 3. Gráficos para duas variáveis: Qualitativa e Qualitativa

### Sexo e Consumo de café

Tabela 5: Tabela de contingência de sexo e consumo de café (% por sexo).

Sexo do aluno	Consumo de café			Total
	Não bebe	Consumo moderado	Consumo exagerado	
Masculino	0 0.00%	2 28.57%	5 71.43%	7 <b>100.00%</b>
Feminino	3 13.04%	14 60.87%	6 26.09%	23 <b>100.00%</b>
Total	3	16	11	30





## 4. Gráficos para duas variáveis: Qualitativa e Quantitativa

### Exemplo: IMC dos alunos

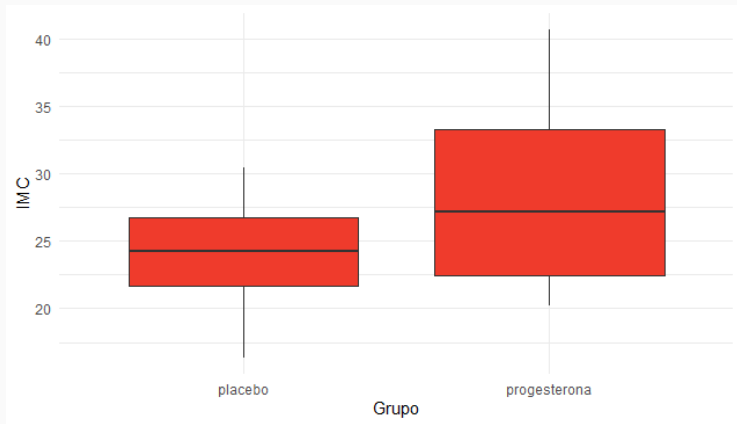
Variável	n	Min	Média	Mediana	Max	Dp	CV
IMC	46	17,75	23,04	22,28	33,39	3,65	15,8%

Estratificando a variável IMC por sexo

Variável	n	Min	Média	Mediana	Max	Dp	CV
Masculino	9	22,99	25,38	25,64	30,07	2,13	8,39%
Feminino	37	17,75	22,47	21,64	33,39	3,73	16,6%

## 4. Gráficos para duas variáveis: Qualitativa e Quantitativa

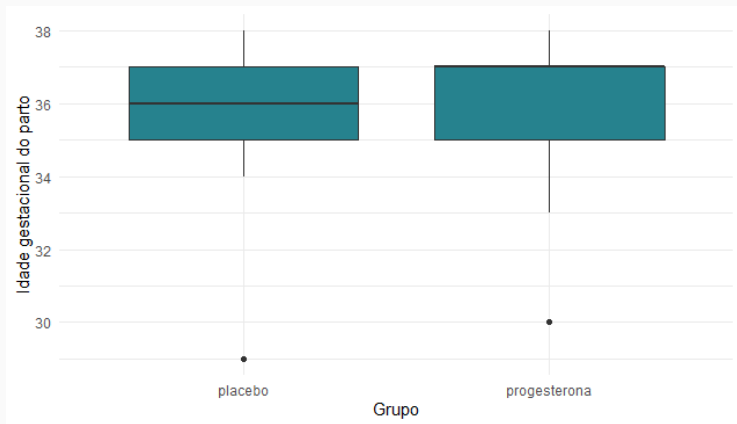
IMC por grupo (progesterona e placebo)



**Figura 14:** Gráfico boxplot do IMC por grupo.

## 4. Gráficos para duas variáveis: Qualitativa e Quantitativa

Idade gestacional do parto (IGP) por grupo (progesterona e placebo)



**Figura 15:** Gráfico boxplot do IGP por grupo.

## 5. Gráficos para duas variáveis: Quantitativa e Quantitativa

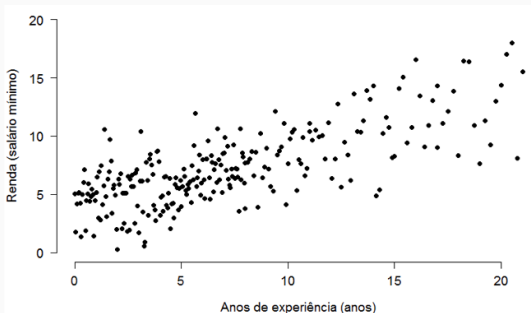
Renda (salário mínimo) e Anos de experiência em uma empresa (anos)

	A	B	C
1	ID	Anos	Renda
2	1	2,5	5,8
3	2	4,1	7,7
4	3	10,5	10,5
5	4	8,6	5,8
6	5	1,7	2,4
7	6	6,4	4,2
8	7	7,9	5,8
9	8	14,2	12,4
10	9	18,9	14,8
11	10	22,2	14

## 5. Gráficos para duas variáveis: Quantitativa e Quantitativa

### Gráfico de dispersão

É um gráfico onde pontos são usados para representar simultaneamente os valores de duas variáveis quantitativas medidas em cada elemento do conjunto de dados.



**Figura 16:** Gráfico de dispersão entre renda e anos de experiência.

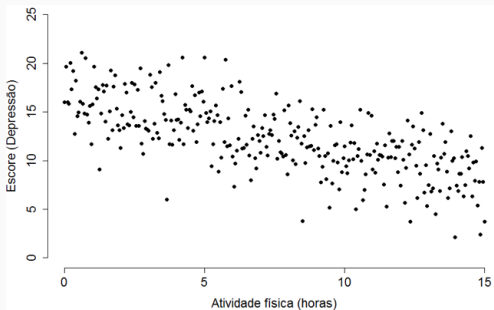
## 5. Gráficos para duas variáveis: Quantitativa e Quantitativa

### Gráfico de dispersão

Escore de Depressão e Atividade Física em horas (semana)

	A	B	C
1	ID	Atividade	Depressão
2	1	10,7	5,4
3	2	8,4	6,4
4	3	4,1	15,6
5	4	1,5	17,4
6	5	0	12,1
7	6	14,5	14,7
8	7	7,5	10,3
9	8	5,5	11,3
10	9	7,9	16,4
11	10	2,3	18,7

## 5. Gráficos para duas variáveis: Quantitativa e Quantitativa



**Figura 17:** Gráfico de dispersão entre escore de depressão e horas de atividade física.

O que acontece com o Escore de depressão?

## 5. Gráficos para duas variáveis: Quantitativa e Quantitativa

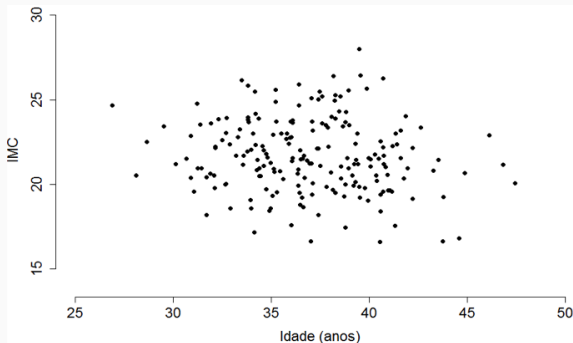
Gráfico de dispersão

IMC e Idade (anos)

	A	B	C
1	<b>ID</b>	<b>IMC</b>	<b>Idade</b>
2	1	18,4	25,4
3	2	21,8	39,4
4	3	29,1	42,6
5	4	27,8	37,8
6	5	24,9	41,2
7	6	26,7	28,4
8	7	21,5	24,2
9	8	22,9	35,7
10	9	27,8	29,9
11	10	23,4	45,8



## 5. Gráficos para duas variáveis: Quantitativa e Quantitativa



**Figura 18:** Gráfico de dispersão entre IMC e Idade.

O que se pode observar à respeito da Idade e do IMC?