



ESALQ

Escola Superior de Agricultura Luiz de Queiroz
Universidade de São Paulo

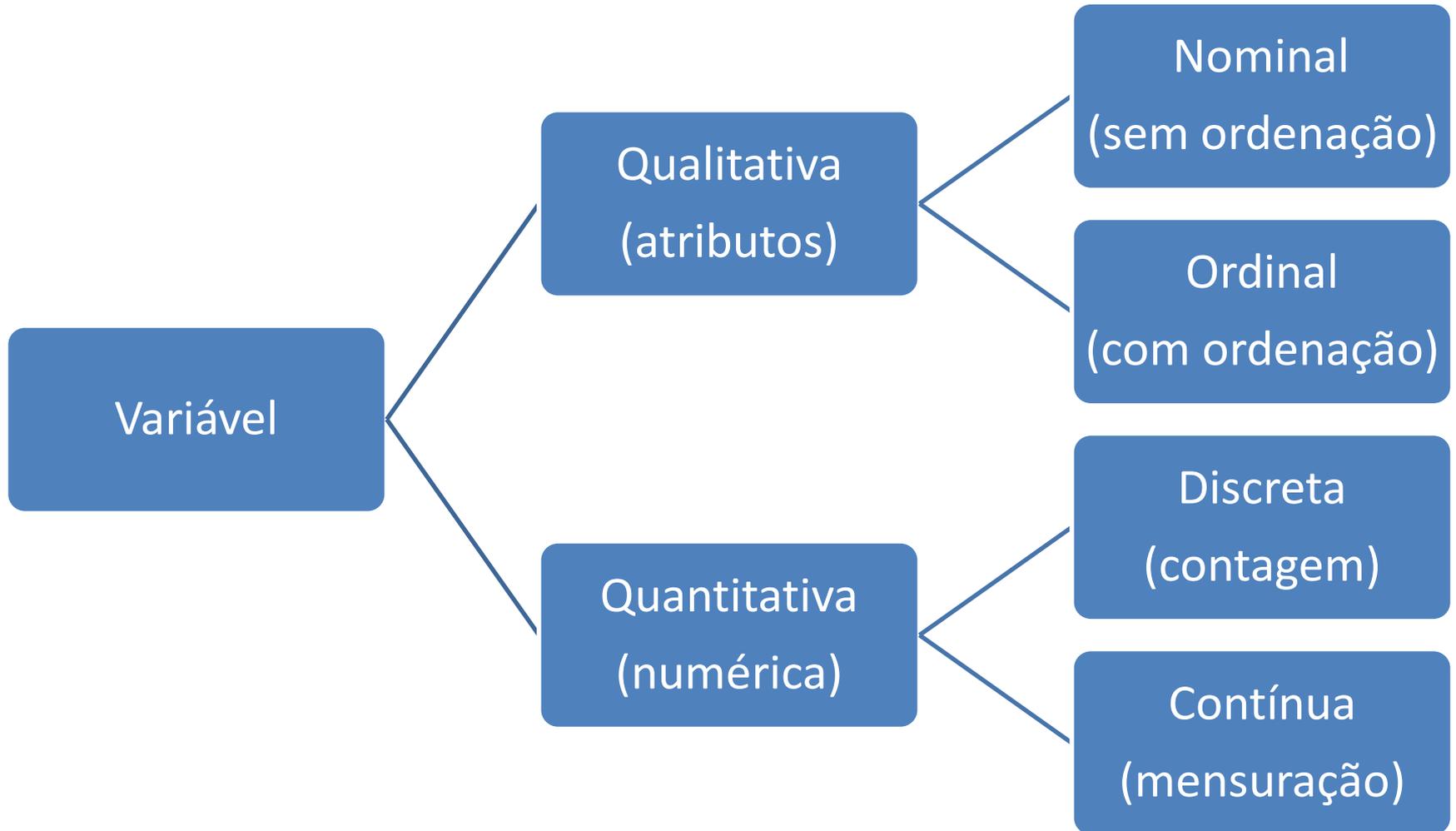


LCE2112 – Estatística Aplicada às Ciências Sociais e Ambientais

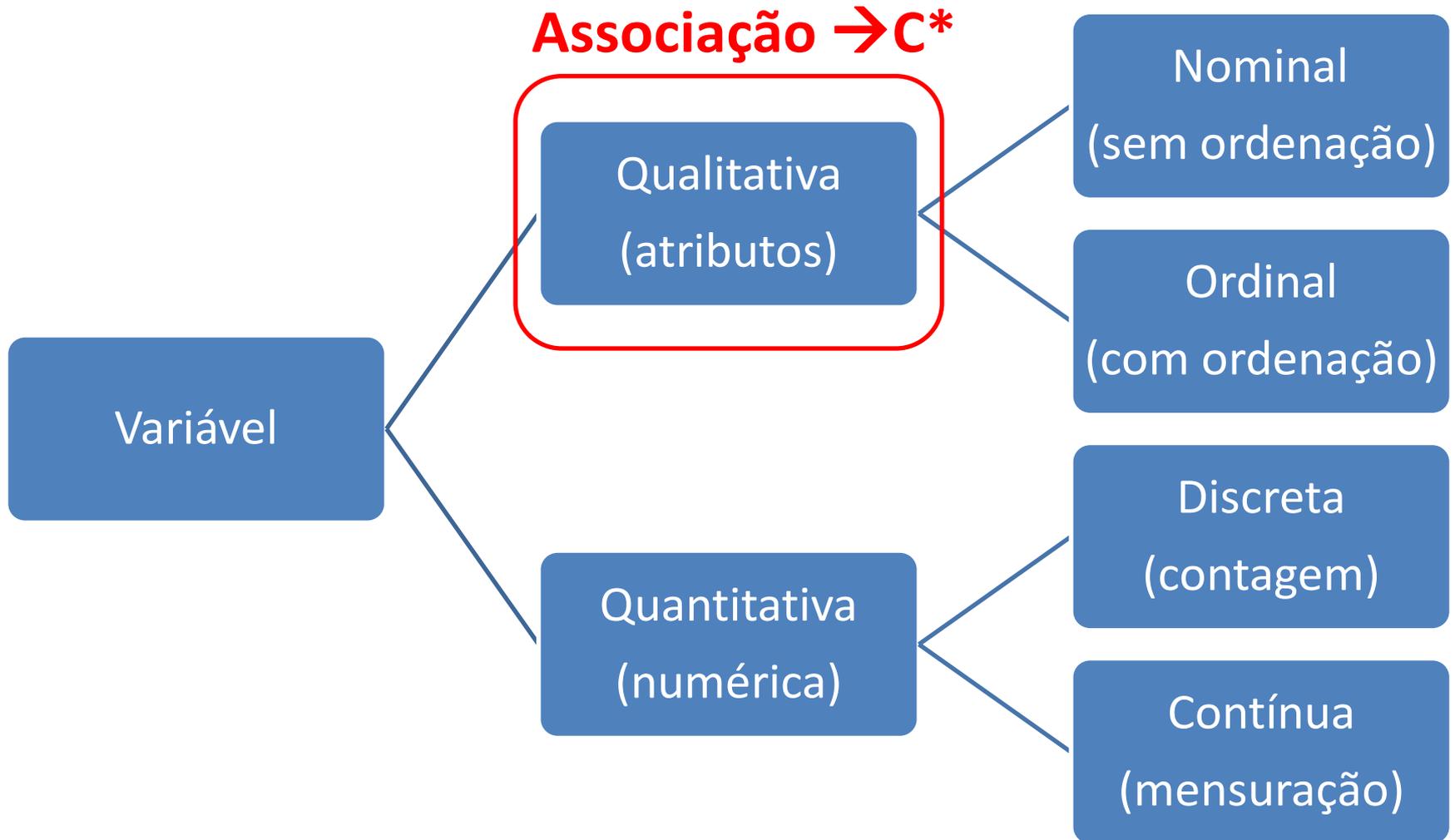
Profa. Taciana Villela Savian

tvsvian@usp.br

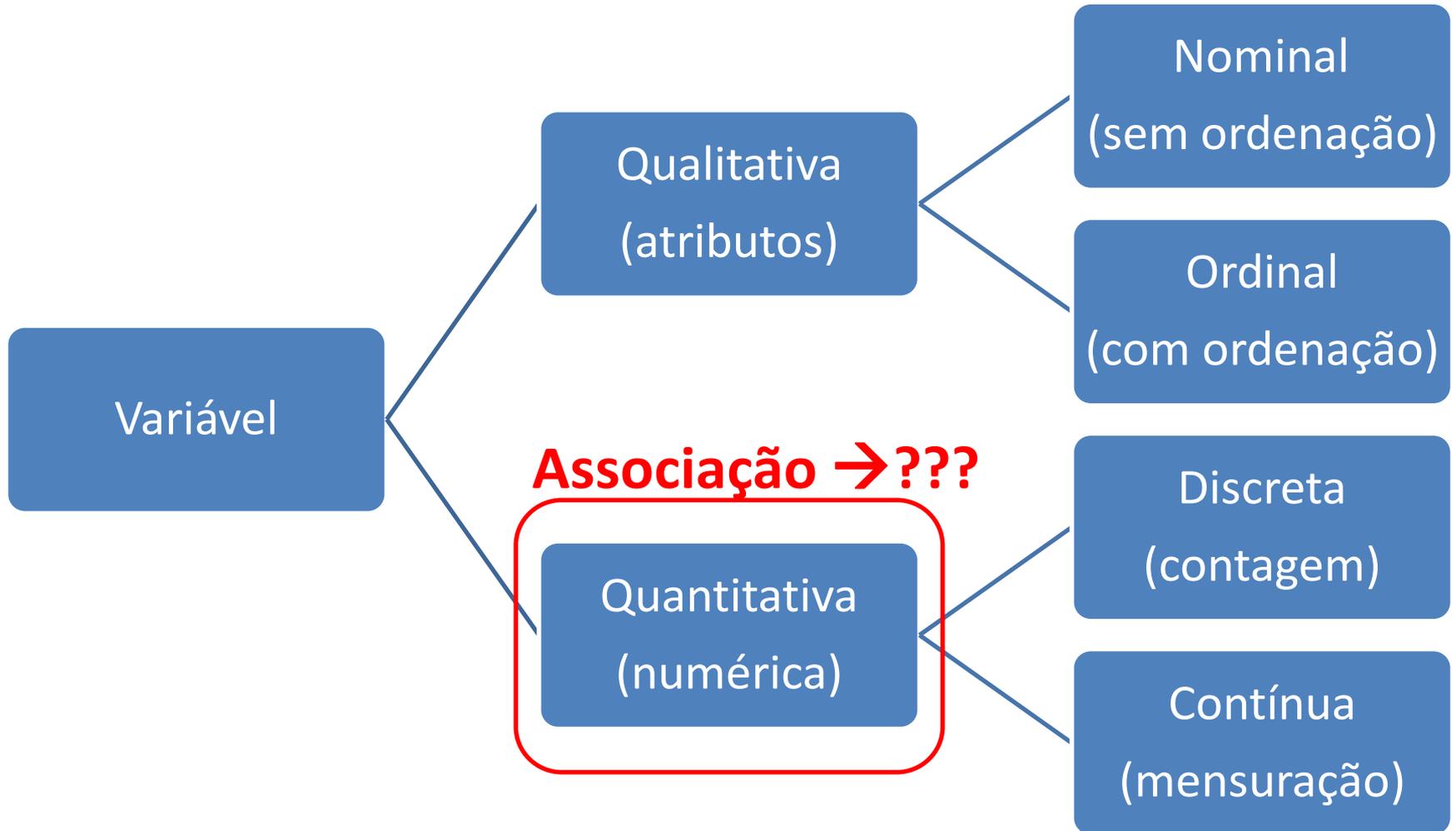
Tipos de Variáveis



Tipos de Variáveis



Tipos de Variáveis



Análise bidimensional

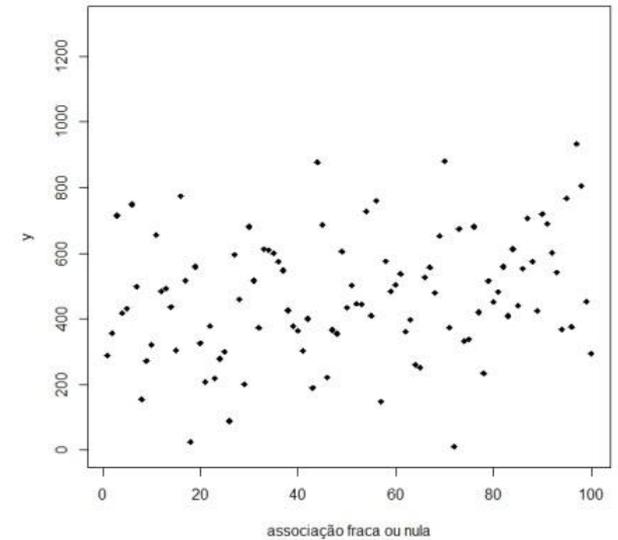
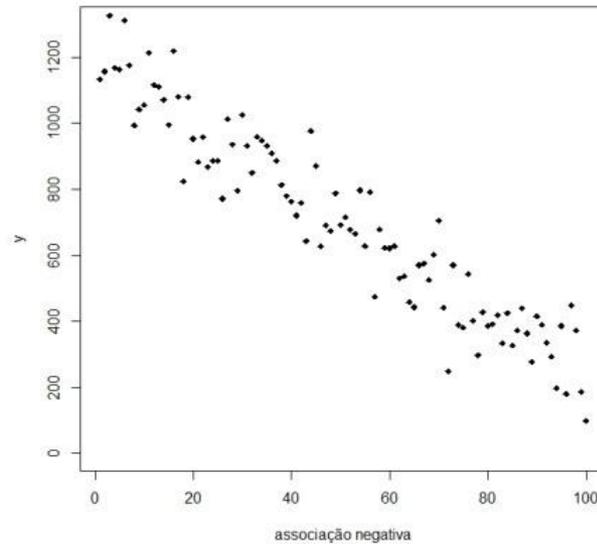
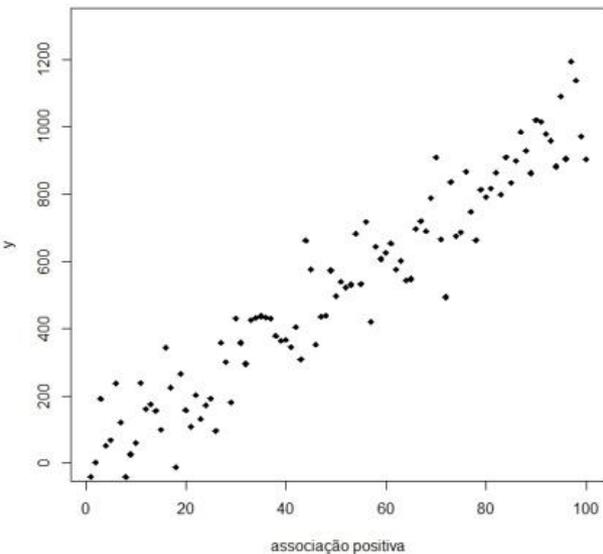
- Também pode ser de interesse do pesquisador conhecer se **duas variáveis quantitativas** têm alguma relação, isto é, se valores altos (ou baixos) de uma das variáveis implicam em valores altos (ou baixos) da outra variável.
- Por exemplo:
 - Pode-se verificar se existe associação entre a taxa de desemprego e a taxa de migração de moradores em uma cidade;
 - Entre verba investida em educação e taxa de desemprego;
 - Investimento em propagandas de um produto e retorno nas vendas, etc.

Análise bidimensional

- O primeiro estágio na determinação do grau e forma de associação entre duas variáveis quantitativas (numéricas) é fazer um gráfico de pontos dos dados observados, denominado de “diagrama de dispersão”.
- Esse diagrama nada mais é do que a representação gráfica dos pares de valores em um sistema cartesiano.

Análise bidimensional

- Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é uma das variáveis quantitativas e y é a outra variável quantitativa.



Análise bidimensional

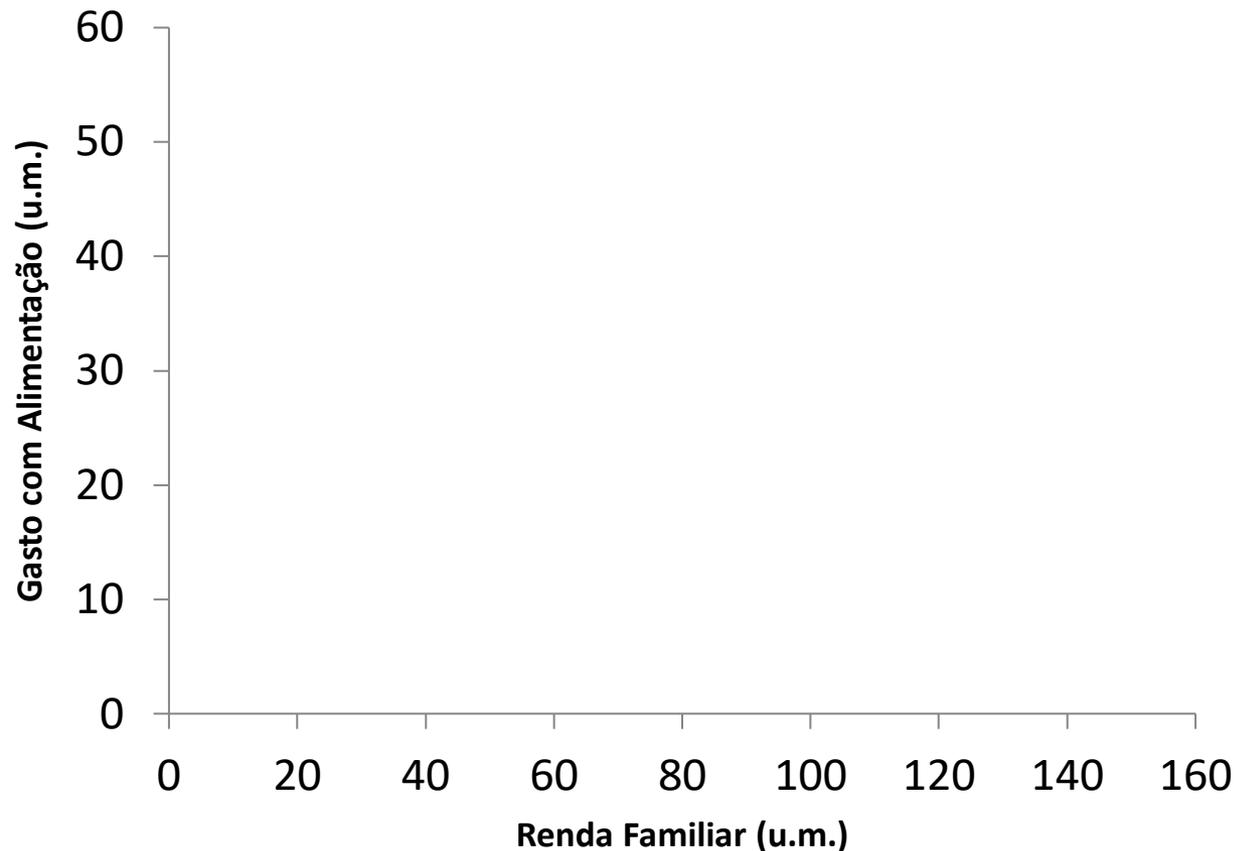
- **Exemplo:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 10 famílias.

Renda	3	5	10	20	20	40	60	70	120	150
Gasto	1,5	2	6	12	15	10	20	25	40	50

Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é a renda familiar y é o gasto com alimentação

Análise bidimensional

- Gráfico de Dispersão



Análise bidimensional

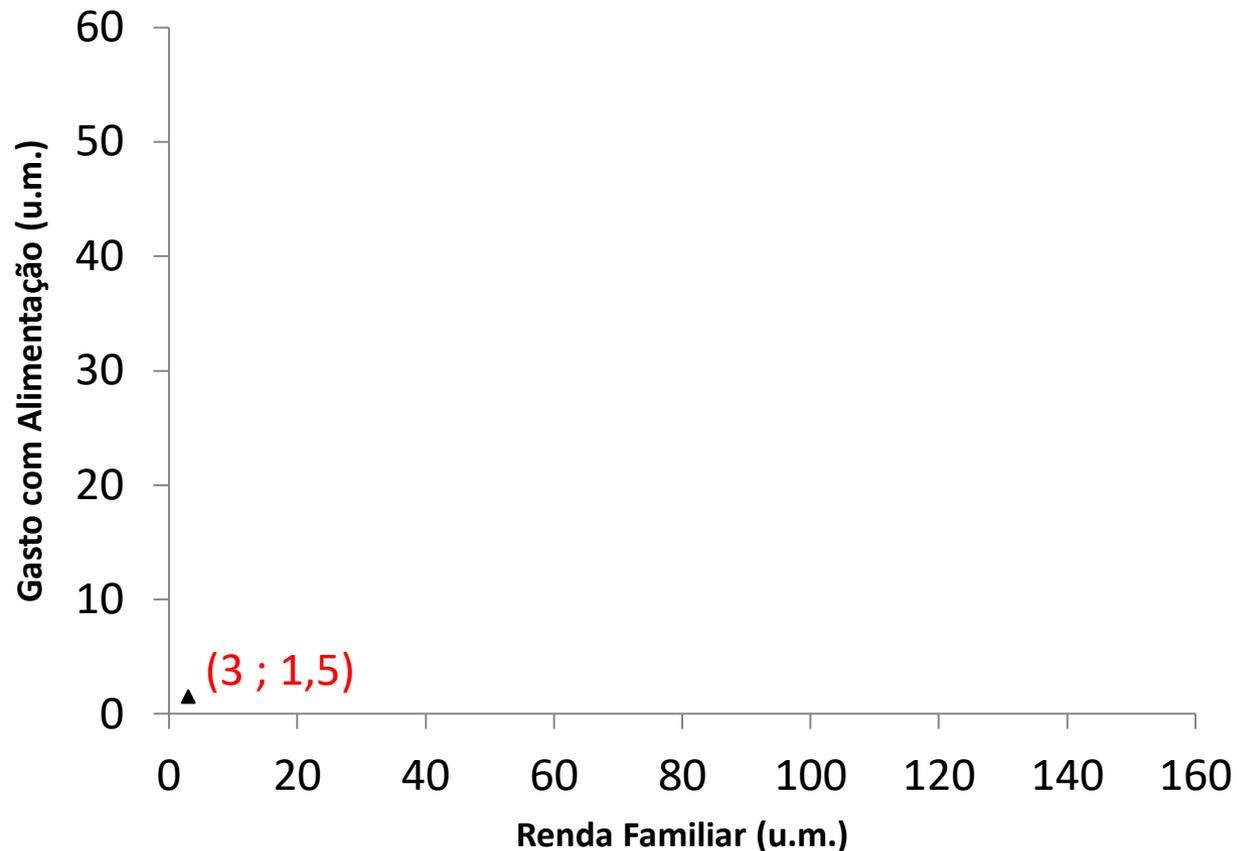
- **Exemplo:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 10 famílias.

Renda	3	5	10	20	20	40	60	70	120	150
Gasto	1,5	2	6	12	15	10	20	25	40	50

Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é a renda familiar y é o gasto com alimentação

Análise bidimensional

- Gráfico de Dispersão



Análise bidimensional

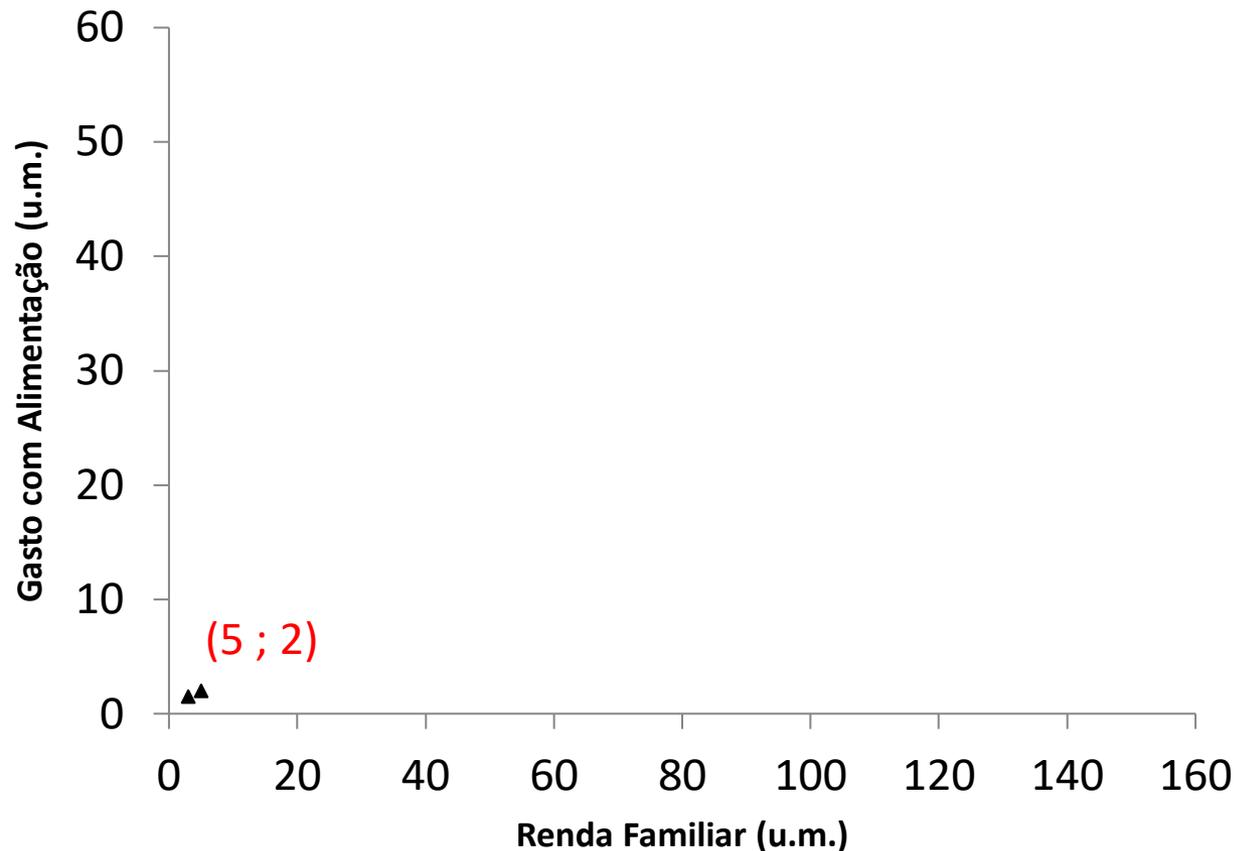
- **Exemplo:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 10 famílias.

Renda	3	5	10	20	20	40	60	70	120	150
Gasto	1,5	2	6	12	15	10	20	25	40	50

Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é a renda familiar y é o gasto com alimentação

Análise bidimensional

- Gráfico de Dispersão



Análise bidimensional

- **Exemplo:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 10 famílias.

Renda	3	5	10	20	20	40	60	70	120	150
Gasto	1,5	2	6	12	15	10	20	25	40	50

Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é a renda familiar y é o gasto com alimentação

Análise bidimensional

- Gráfico de Dispersão

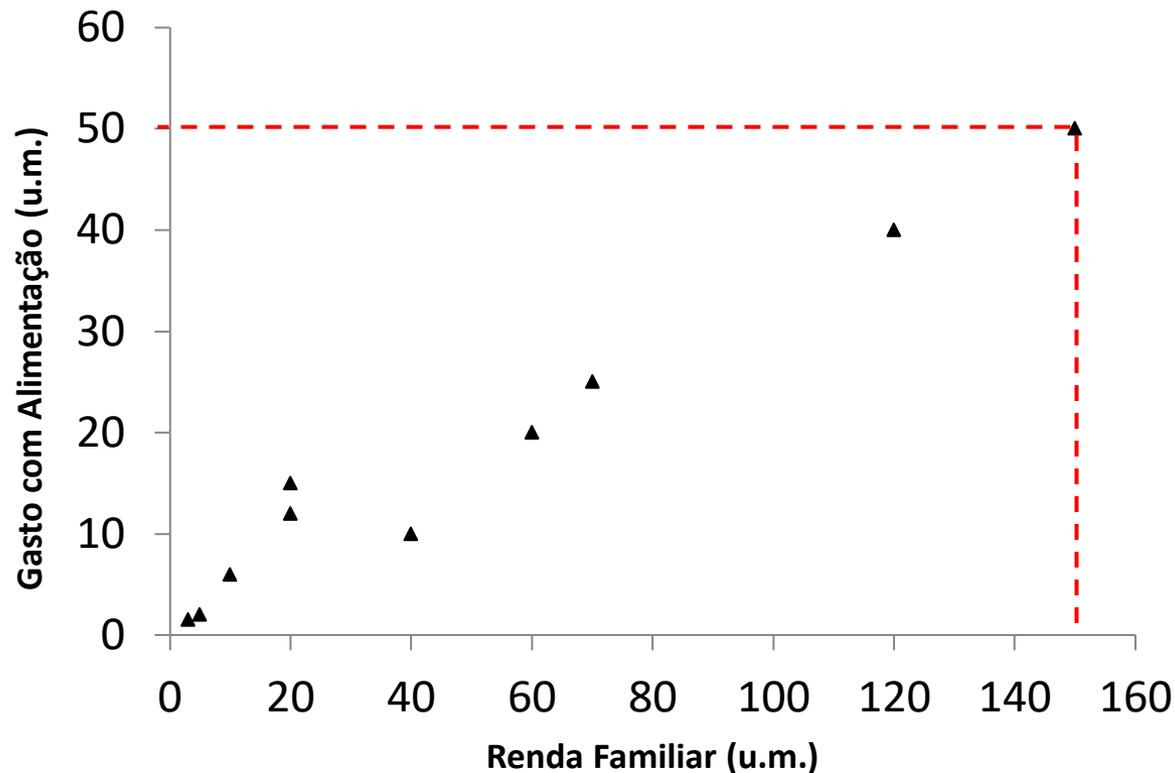
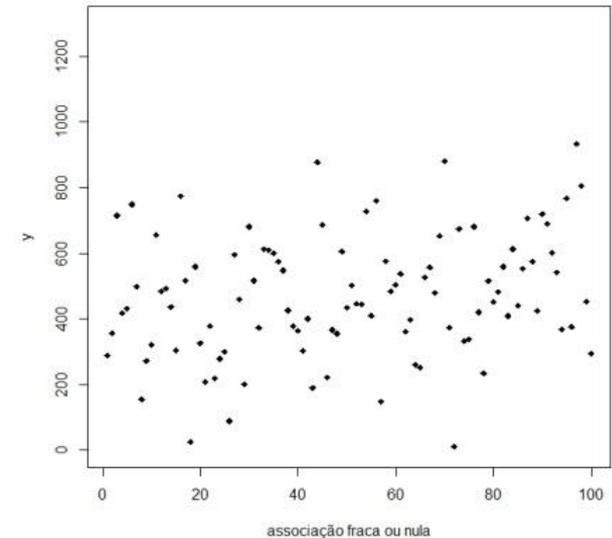
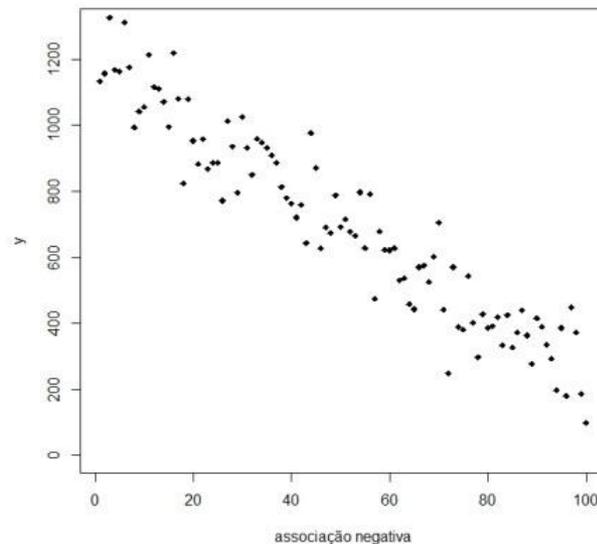
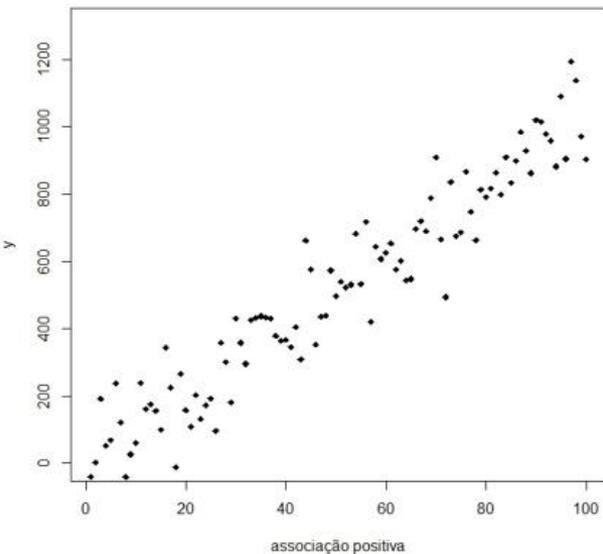


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

- Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é uma das variáveis quantitativas e y é a outra variável quantitativa;



Análise bidimensional

- Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é uma das variáveis quantitativas e y é a outra variável quantitativa;

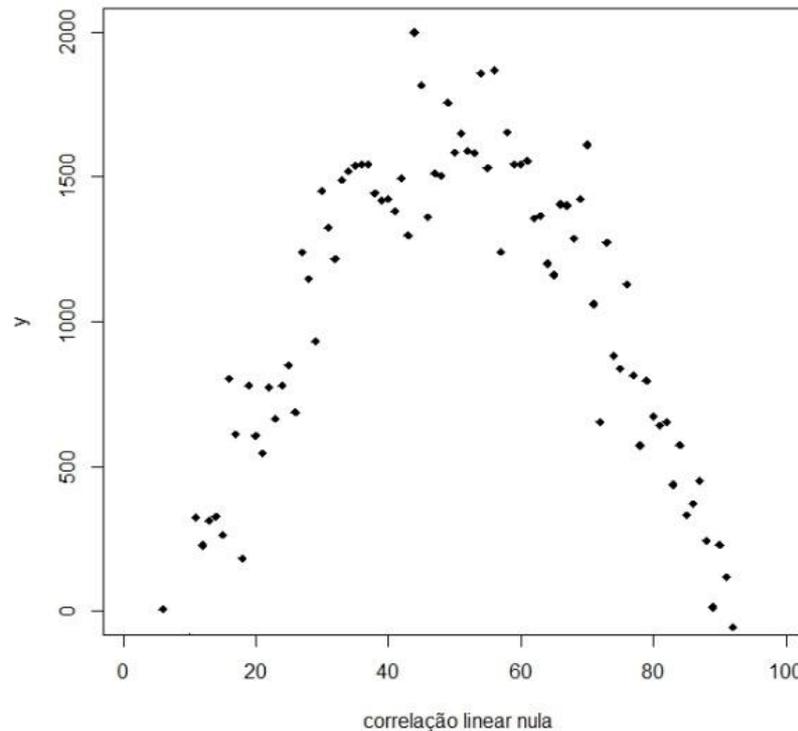


Figura. Esquema de forte associação entre duas variáveis aleatórias quantitativas, porém com coeficiente de correlação linear nulo. Fonte: Autoria própria.

Análise bidimensional

- **Exemplo:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 10 famílias.

Renda	3	5	10	20	20	40	60	70	120	150
Gasto	1,5	2	6	12	15	10	20	25	40	50

Quantificar a associação entre essas duas variáveis?

Coeficiente de Correlação de Pearson (r)

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$r = \frac{SPXY}{\sqrt{SQX \times SQY}}$$

Em que: SPXY – Soma de Produtos de X e Y;

SQX – Soma de Quadrados de X;

SQY – Soma de Quadrados de Y;

Como calcular cada um dos termos?

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$SP_{XY} = \left(\sum_{i=1}^n X_i Y_i \right) - \frac{(\sum_{i=1}^n X_i) \times (\sum_{i=1}^n Y_i)}{n}$$

X – Renda	3	5	10	20	20	40	60	70	120	150
Y – Gasto	1,5	2	6	12	15	10	20	25	40	50
XY	4,5	10	60	240	300	400	1200	1750	4800	7500

$$\sum_{i=1}^n X_i = 3 + 5 + \dots + 150 = 498$$

$$\sum_{i=1}^n Y_i = 1,5 + 2 + \dots + 50 = 181,5$$

$$\sum_{i=1}^n X_i Y_i = 4,5 + 10 + \dots + 7500 = 16.264,5$$

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$SPXY = 16.264,5 - \frac{498 \times 181,5}{10} = 7.245,8$$

X – Renda	3	5	10	20	20	40	60	70	120	150
Y – Gasto	1,5	2	6	12	15	10	20	25	40	50
XY	4,5	10	60	240	300	400	1200	1750	4800	7500

$$\sum_{i=1}^n X_i = 3 + 5 + \dots + 150 = 498$$

$$\sum_{i=1}^n Y_i = 1,5 + 2 + \dots + 50 = 181,5$$

$$\sum_{i=1}^n X_i Y_i = 4,5 + 10 + \dots + 7500 = 16.264,5$$

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$SQX = \left(\sum_{i=1}^n X_i^2 \right) - \frac{(\sum_{i=1}^n X_i)^2}{n}$$

X – Renda	3	5	10	20	20	40	60	70	120	150
Y – Gasto	1,5	2	6	12	15	10	20	25	40	50
XY	4,5	10	60	240	300	400	1200	1750	4800	7500

$$\sum_{i=1}^n X_i = 3 + 5 + \dots + 150 = 498$$

$$\sum_{i=1}^n X_i^2 = 3^2 + 5^2 + \dots + 150^2 = 47.934$$

$$SQX = 47.934 - \frac{(498)^2}{10} = 23.133,6$$

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$SQY = \left(\sum_{i=1}^n Y_i^2 \right) - \frac{(\sum_{i=1}^n Y_i)^2}{n}$$

X – Renda	3	5	10	20	20	40	60	70	120	150
Y – Gasto	1,5	2	6	12	15	10	20	25	40	50
XY	4,5	10	60	240	300	400	1200	1750	4800	7500

$$\sum_{i=1}^n Y_i = 1,5 + 2 + \dots + 50 = 181,5$$

$$\sum_{i=1}^n Y_i^2 = 1,5^2 + 2^2 + \dots + 50^2 = 5.636,2$$

$$SQY = 5.636,2 - \frac{(181,5)^2}{10} = 2.342,0$$

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$r = \frac{SPXY}{\sqrt{SQX \times SQY}}$$

Em que: SPXY – Soma de Produtos de X e Y;

SQX – Soma de Quadrados de X;

SQY – Soma de Quadrados de Y;

$$r = \frac{SPXY}{\sqrt{SQX \times SQY}} = \frac{7.245,8}{\sqrt{23.133,6 \times 2.342,0}} \\ = 0,98$$

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$r = \frac{SPXY}{\sqrt{SQX \times SQY}}$$

Propriedades:

- r é sempre um número entre -1 e 1;
- r = 0 **não indica independência entre as variáveis**, indica apenas que não existe uma relação LINEAR entre as variáveis;
- |r| próximo a 1, indica alta associação entre as variáveis (positiva/negativa)
- |r| próximo a 0, indica não **relação linear** entre as variáveis;
- |r| próximo a 0,5, indica associação moderada.

Análise bidimensional

- **Coeficiente de Correlação de Pearson (r)**

$$r = \frac{SPXY}{\sqrt{SQX \times SQY}}$$

$$r = 0,98$$

Segundo o resultado da correlação obtida ($r=0,98$), pode-se notar que há uma forte correlação linear entre as variáveis renda familiar e gasto com alimentação. Nota-se que à medida que a renda familiar aumenta o gasto com alimentação (em unidades monetárias) também aumenta, o que é coerente com o gráfico de dispersão apresentado anteriormente.

Análise bidimensional

- Gráfico de Dispersão

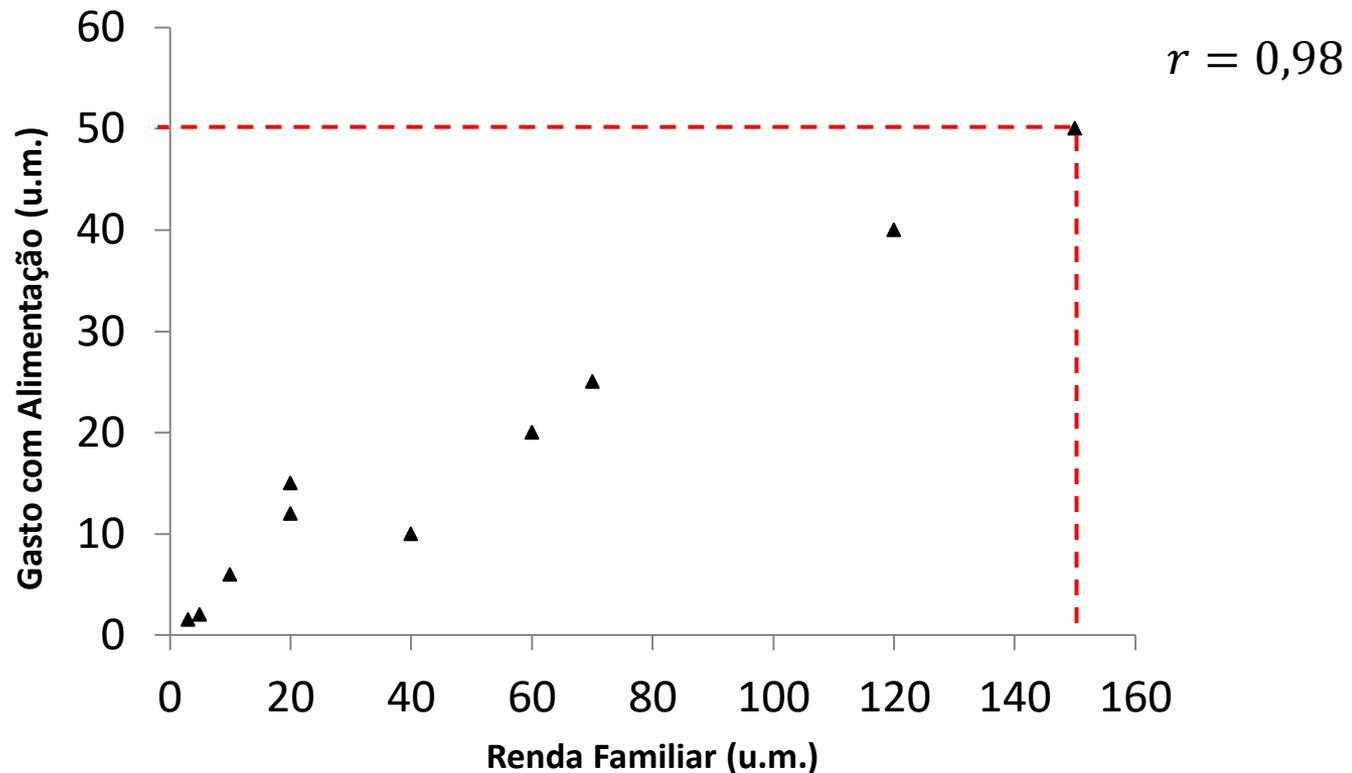


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

Os quatro conjuntos de dados a seguir foram preparados pelo estatístico F. J. Anscombe e são usados com frequência em aulas sobre correlação.

Para cada conjunto:

- Fazer o diagrama de dispersão;
- Obter o coeficiente de correlação de Pearson;
- Interpretar o resultado.

Conjunto 1		Conjunto 2		Conjunto 3		Conjunto 4	
X	Y	X	Y	X	Y	X	Y
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,10	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,10	4	5,39	19	12,50
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89

Análise bidimensional

Os quatro conjuntos de dados a seguir foram preparados pelo estatístico F. J. Anscombe e são usados com frequência em aulas sobre correlação.

Para cada conjunto:

- Fazer o diagrama de dispersão;
- Obter o coeficiente de correlação de Pearson;
- Interpretar o resultado.

Conjunto 1		Conjunto 2		Conjunto 3		Conjunto 4	
X	Y	X	Y	X	Y	X	Y
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,10	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,10	4	5,39	19	12,50
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89
r=0,82		r=0,82		r=0,82		r=0,82	

Análise bidimensional

- **Regressão Linear Simples** → Quando o objetivo é verificar a existência de uma relação funcional entre duas variáveis: uma denominada variável dependente (ou variável resposta), e a outra denominada variável independente (ou variável explicativa);
- Em outras palavras, consiste na obtenção de uma equação que tenta explicar o que ocorre na variável resposta (Y) quando variamos os valores da variável explicativa (X)

Análise bidimensional

- Assim como na análise de correlação, o diagrama de dispersão é de extrema importância na análise de regressão uma vez que o comportamento de Y em função da variável X pode se apresentar de diversas maneiras: linear, quadrático, cúbico, exponencial, logarítmico, entre outros;
- Esse comportamento dará uma primeira ideia se o modelo de regressão linear simples é condizente com o comportamento dos dados.

Análise bidimensional

- **Exemplo:** Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 10 famílias.

Renda (X)	3	5	10	20	20	40	60	70	120	150
Gasto (Y)	1,5	2	6	12	15	10	20	25	40	50

Gráfico de Dispersão = Gráfico de pontos onde plotamos os pares de dados (x,y) em que x é a renda familiar y é o gasto com alimentação

Análise bidimensional

- Gráfico de Dispersão

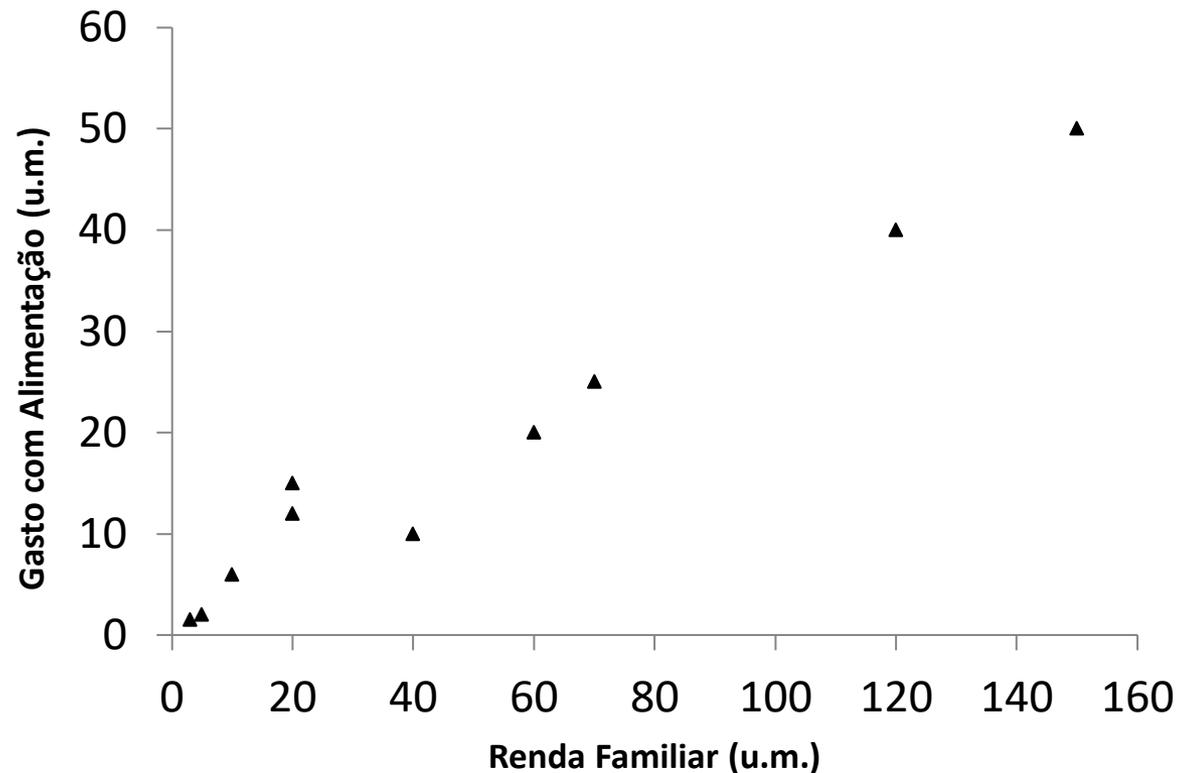


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

- Gráfico de Dispersão

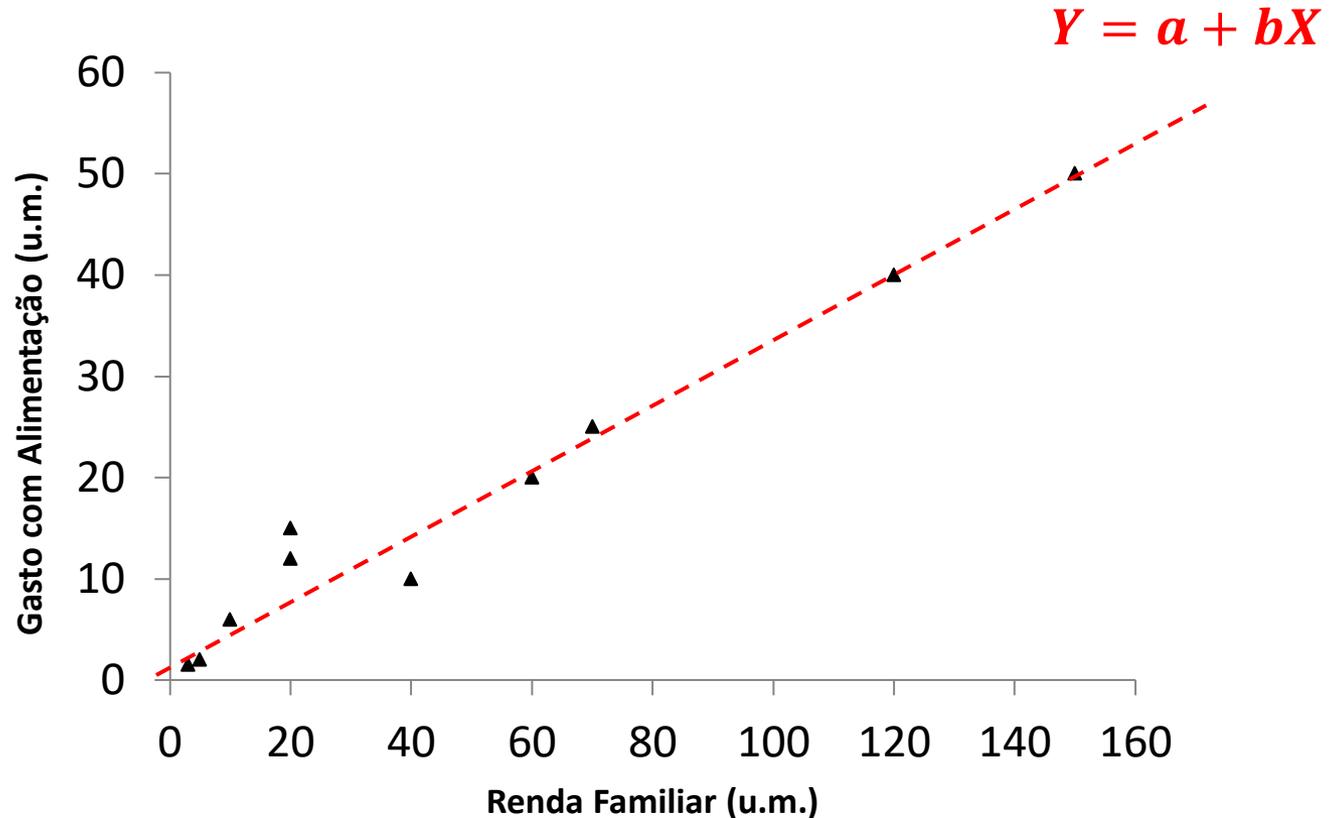


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

- O que significam os parâmetros do modelo?

Modelos Regressão Linear Simples (RLS)

$$Y = a + bX$$

a é o parâmetro do modelo RLS que representa o intercepto da reta com o eixo Y;

b denominado coeficiente angular ou coeficiente de regressão, e representa a variação de Y em função da variação de uma unidade na variável X;

Análise bidimensional

- Gráfico de Dispersão

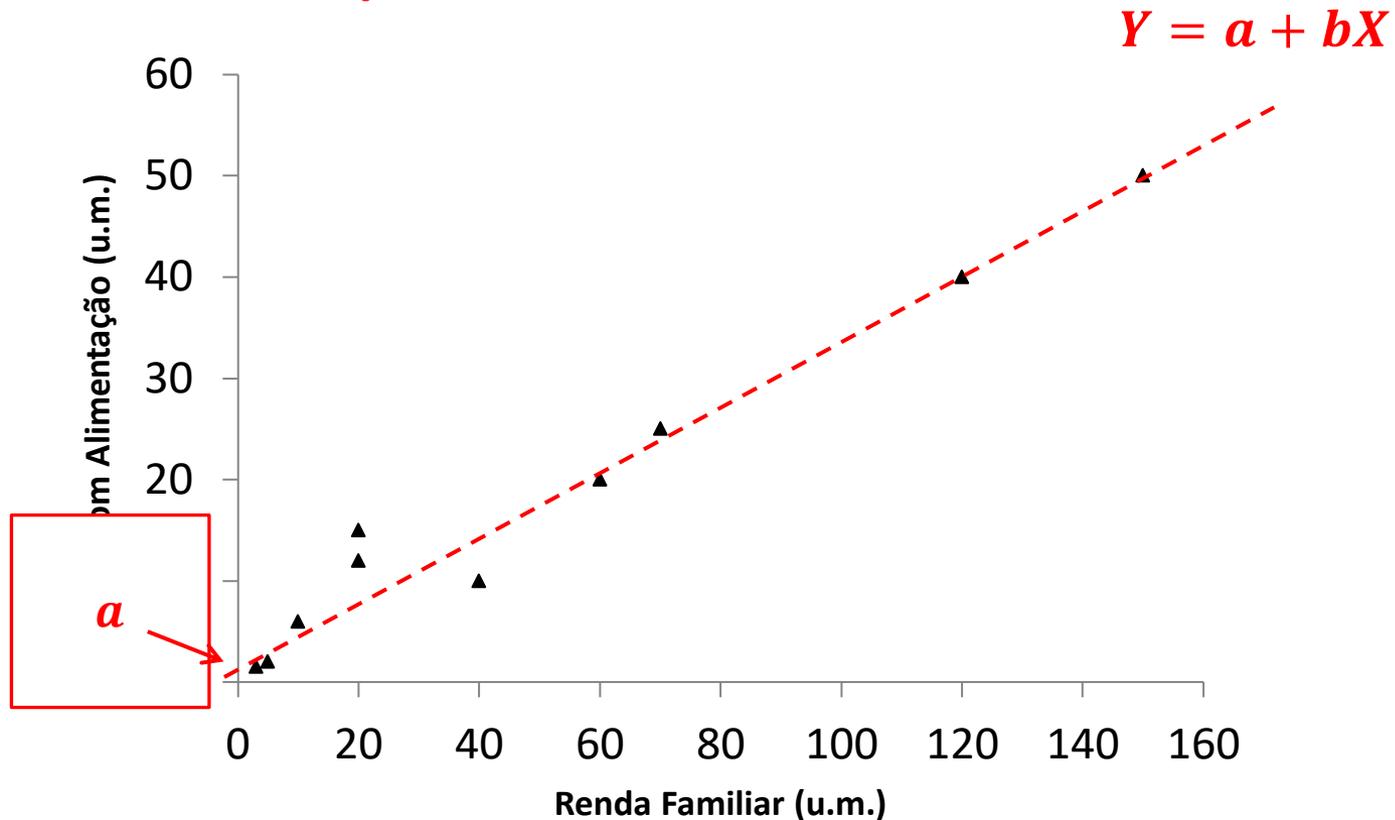


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

- Como estimar (encontrar os valores) os parâmetros ***a*** e ***b***?

$$Y = a + bX$$

$$\hat{a} = \frac{\sum_{i=1}^n Y_i}{n} - \hat{b} \times \frac{\sum_{i=1}^n X_i}{n}$$

$$\hat{b} = \frac{SPXY}{SQX} \longrightarrow \begin{aligned} SPXY &= \sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n} \\ SQX &= \sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \end{aligned}$$

Análise bidimensional

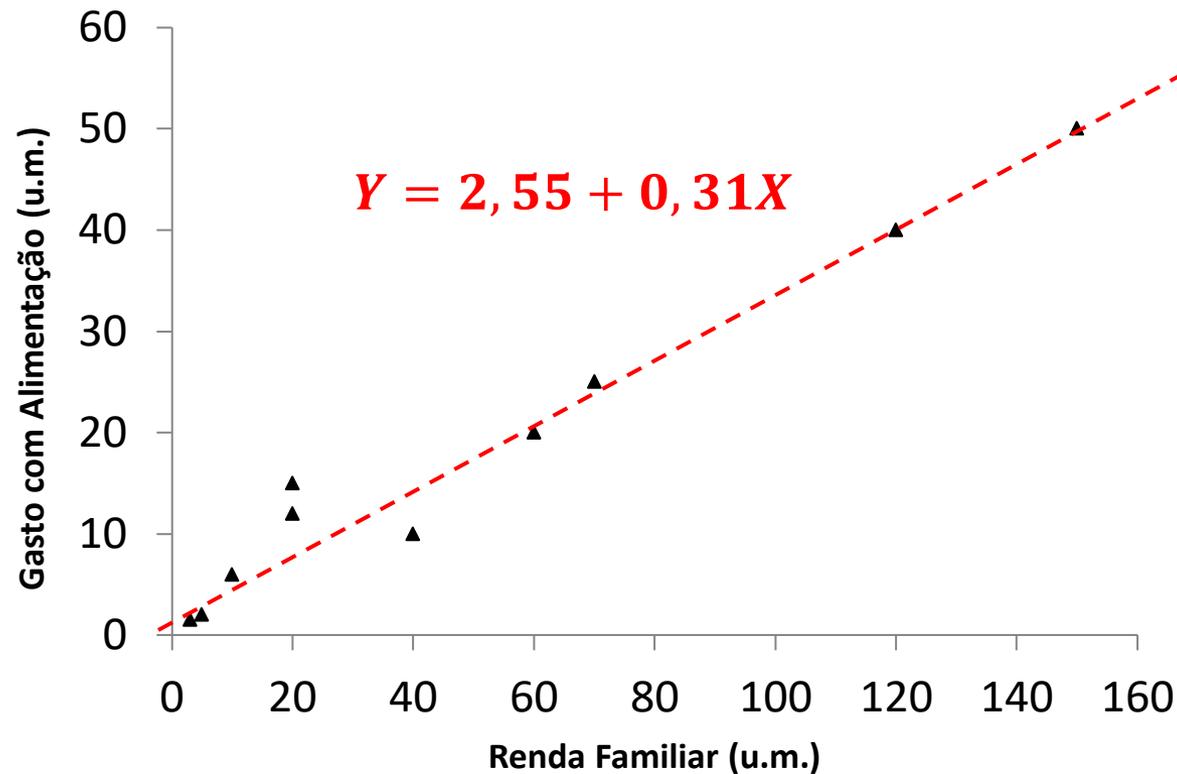


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

Conclusão?

- Em relação as estimativas dos parâmetros, pode-se concluir que “ $a=2,55$ ” representa o intercepto do modelo com o eixo Y , ou seja, a estimativa do gasto com alimentação quando a renda familiar for zero ($x_i=0$). Veja que nem sempre a estimativa do parâmetro “ a ” tem uma interpretação prática (embora tenha interpretação estatística);
- Já a estimativa do parâmetro “ b ”, que representa o coeficiente angular da reta ajustada, expressa a variação em Y em função da variação de uma unidade na variável X, ou seja, para o aumento de uma unidade monetária na renda familiar, o gasto com alimentação aumenta em $b=0,31$ unidades monetárias.

Análise bidimensional

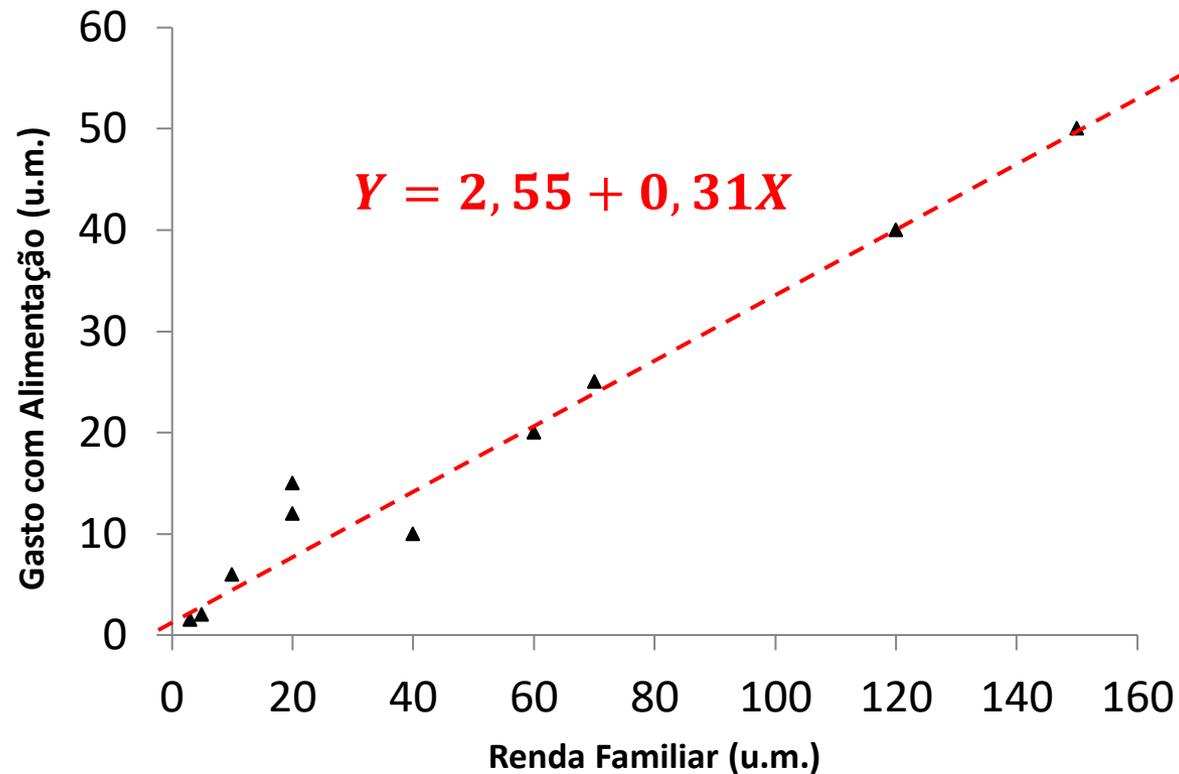


Figura 1. Gráfico de dispersão do gasto com alimentação e renda familiar (em unidades monetárias) de uma amostra de 10 famílias.

Análise bidimensional

Qual seria o gasto com alimentação (em unidades monetárias) se a renda familiar fosse de 80 u. m.?

Análise bidimensional

Qual seria o gasto com alimentação (em unidades monetárias) se a renda familiar fosse de 80 u. m.?

$$Y = 2,55 + 0,31X$$

$$Y = 2,55 + 0,31(80)$$

$$Y = 2,55 + 24,8$$

$$Y = 27,35$$

Exercícios

- Fazer a lista que está no e-disciplina Aula _03;
 - Realizar à mão;
 - Reproduzir no excel;