

- 1) Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 25 famílias.

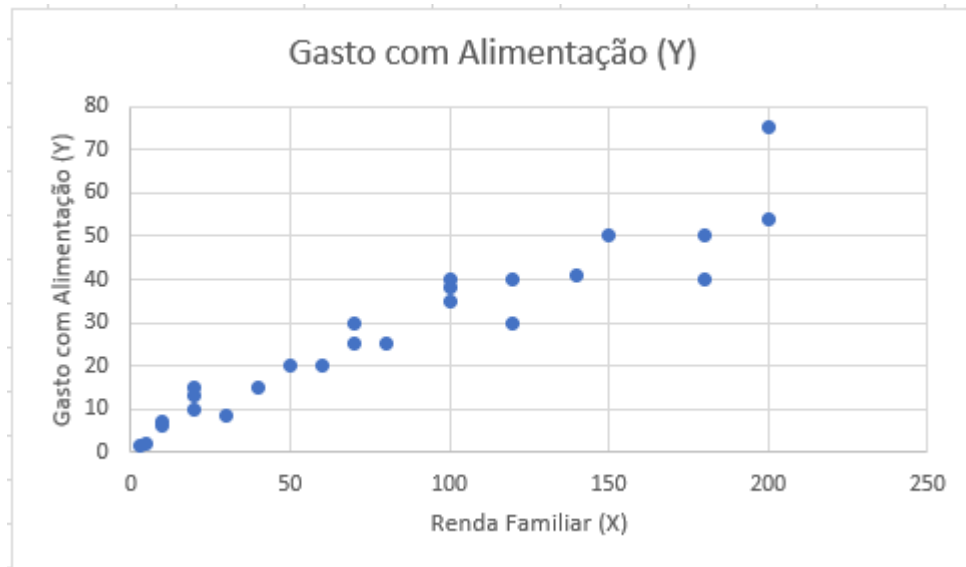
Renda Familiar (X)	Gasto com Alimentação (Y)
3	1,5
5	2,0
10	6,0
10	7,0
20	10,0
20	13,0
20	15,0
30	8,5
40	15,0
50	20,0
60	20,0
70	25,0
70	30,0
80	25,0
100	40,0
100	35,0
100	38,0
120	30,0
120	40,0
140	41,0
150	50,0
180	40,0
180	50,0
200	75,0
200	54,0

- a) (0,5 ponto) Calcular o coeficiente de correlação entre essas variáveis. O que se pode concluir a partir deste resultado?

**R: Correl = 0.9504**

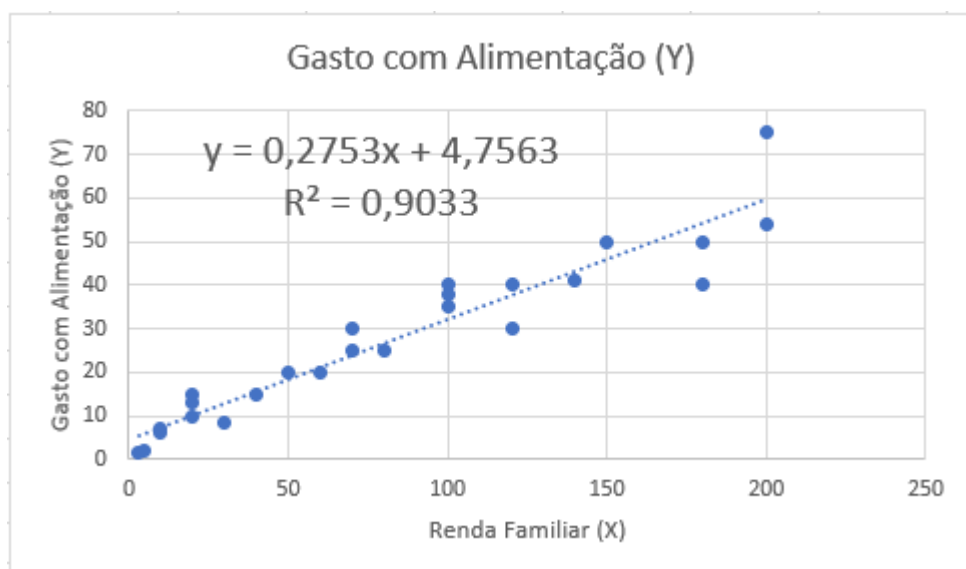
**R: A Correlação entre "gasto com alimentação" e "renda familiar" é de 95,04%, então existe uma correlação muito forte entre as variáveis "gasto de alimentação" e "renda familiar", ou seja quanto maior a quantidade de renda familiar, maior será com o gasto com alimentação**

- b) (0,5 ponto) Construa o diagrama de dispersão da variável gasto com alimentação (Y) em função da renda familiar (X).



- c) (0,5 ponto) Obtenha um modelo linear do gasto com alimentação em função da renda familiar. O que se pode concluir sobre o valor do coeficiente angular da reta encontrada?

**R: O "gasto de alimentação (Y)" aumenta, em média, 0,2753 vezes a cada "renda familiar (X)".**



- d) (0,5 ponto) Obtenha e interprete o coeficiente de determinação do modelo obtido.

**R: 90.3% do aumento do gasto de alimentação é explicado devido ao aumento do valor da renda familiar.**

- 2) Uma empresa que vende por correio componentes de computadores pessoais, software e hardware possui um depósito geral para a distribuição dos produtos. Atualmente, a administração se encontra examinando o processo de distribuição deste depósito e está interessado em estudar os fatores que afetam os custos de distribuição do depósito. Atualmente um pequeno cargo de manipulação se adiciona ao pedido, independentemente da quantidade pela que se fizeram. Foram coletados dados correspondentes de 24 meses e respeito aos custos de distribuição de depósito, as vendas e número de pedidos. A continuação apresenta os resultados:

Mês	Y	X <sub>1</sub>	X <sub>2</sub>	Mês	Y	X <sub>1</sub>	X <sub>2</sub>
1	52,95	386	4.015	13	62,98	372	3.977
2	71,66	446	3.806	14	72,30	328	4.428
3	85,56	512	5.309	15	58,99	408	3.964
4	63,69	401	4.262	16	79,38	491	4.582
5	72,81	457	4.296	17	94,44	527	5.582
6	68,44	458	4.097	18	59,74	444	3.450
7	52,46	301	3.213	19	90,50	623	5.079
8	70,77	484	4.809	20	93,24	596	5.735
9	82,03	517	5.237	21	69,33	463	4.269
10	74,39	503	4.732	22	53,71	389	3.708
11	70,84	535	4.413	23	98,18	547	5.387
12	54,08	553	2.921	24	66,80	415	4.161

Onde:

Y: Custo de distribuição (em milhares de dólares);

X<sub>1</sub>: Vendas (em milhares de dólares);

X<sub>2</sub>: Número de pedidos.

- a) (0,5 ponto) Ajuste os dados a um modelo de regressão com duas variáveis regressoras e interprete as estimativas dos parâmetros do modelo.

**R: O Custo de distribuição (Y) aumenta, 0,0418 vezes em cada "Vendas (X<sub>1</sub>) e em 0,0141 vezes em cada Número de pedidos(X<sub>2</sub>) e -9,81 (em milhares de dólares) de Custo de distribuição quando não há vendas e pedidos ” .**

- b) (0,5 ponto) Estime o custo de distribuição do depósito mensal da empresa quando as vendas são \$ 400.000 dólares e o número de pedidos é de 4.500.

**R: O custo de distribuição segundo o modelo de regressão é de 70,43 em milhares de dólares.**

- c) (1 ponto) Calcule e interprete o coeficiente de determinação e o coeficiente de determinação ajustado. **Como quanto maior o número de variáveis regressoras, maior vai ser o valor do coeficiente de determinação, então o valor do coeficiente de determinação deverá a ser ajustado. O valor do coeficiente de determinação é de 87,4%, e o coeficiente de determinação ajustado é de 86,19% de qualidade de ajuste.**

- d) (1 ponto) Verifique estatisticamente se existe uma relação significativa entre o custo de distribuição e as variáveis regressoras ou explicativas (vendas e número de pedidos), ao nível de significância de 5%.

**R:** Como o  $F(\text{calculado}) = 72,81421671$  é maior do que  $F(\text{tabelado}) = 3,466800112$ , pertencendo então  $F(\text{calculado})$  na região crítica, concluindo então a existência de relação significativa entre o custo de distribuição e as variáveis preditivas. Rejeitando então a hipótese nula.

- e) (1 ponto) Achar e interprete um intervalo de 95% de confiança para os parâmetros do modelo de regressão.

Se inferior de 95% , o Custo de distribuição (Y) aumenta, 0,0081 vezes em cada Vendas (X1) e em 0,0104 vezes em cada Número de pedidos(X2) e -24,34 (em milhares de dólares) de Custo de distribuição quando não há vendas e pedidos. Se Superior 95% o Custo de distribuição (Y) aumenta, 0,075472112 vezes em cada "Vendas (X1) e em 0,017805988 vezes em cada Número de pedidos(X2) e 4,729832032 (em milhares de dólares) de Custo de distribuição quando não há vendas e pedidos.

	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Interseção	-24,34118988	4,729832032
X1	0,008101639	0,075472112
X2	0,01042893	0,017805988

- f) (1 ponto) Determine se cada uma das variáveis explicativa contribuiu significativamente no modelo de regressão. Use  $\alpha=0,05$ . Baseados nestes resultados, indique o modelo de regressão que deveria ser utilizado no problema.

**R:** Segundo com a análise separada das variáveis regressoras à variável dependente, fazendo a análise do valor-p de X1 é de 0,017471759, e o valor-p de X2 é de 8,93245E-08, como as duas variáveis é menor do que o nível de significância de 0,05. Em outras palavras, uma preditora que tenha um valor-p baixo ao nível de significância provavelmente será uma adição significativa ao seu modelo, porque as alterações no valor da preditora estão relacionadas a alterações na variável resposta. Por outro lado, um valor-p maior ao nível de significância (insignificante) sugere que as mudanças na preditora não estão associadas a mudanças na resposta, como é o caso da constante que é de 0,175255865.

- g) (1 ponto) Supondo que as duas variáveis explicativas contribuem no modelo, qual delas é mais importante? Justifique porque.

**R:** Segundo com a análise separada das variáveis regressoras à variável dependente, foi calculado o coeficiente de determinação de cada variável regressora, sendo que na variável X1: Vendas o valor é de 0,4937 e o valor na variável X2: Número de pedidos o valor é de 0,8340, sendo a variável regressora X2: Número de pedidos com o maior coeficiente de determinação, sendo assim determinando como a mais importante.

- h) (1 ponto) Determine e interprete um intervalo de 99% para o custo médio de distribuição mensal quando as vendas são de \$410.000 dólares e os números de pedidos são de 4.250.

R: O custo estimado de distribuição segundo o modelo de regressão com o intervalo de confiança inferior de 99% é de 7,389959214 em milhares de dólares. E o custo estimado de distribuição segundo o modelo de regressão com o intervalo de confiança superior de 99% é de 127,2623202 em milhares de dólares.

	<i>Inferior 99,0%</i>	<i>Superior 99,0%</i>
Interseção	-29,59553461	9,984176766
X1	-0,004075011	0,087648763
X2	0,009095588	0,01913933

- i) (1 ponto) Faça uma análise residual e verifique se o modelo é adequado.

R: Existe Normalização (Probabilidade Esperada) mas no gráfico (Valores Ajustados VS Resíduos), a variância dos resíduos não é constante (heterocedasticidade), havendo uma dispersão, logo então podemos verificar os erros quanto pior é o modelo quanto mais disperso é o gráfico, quanto mais próximo os dados estão mais uniforme e assertiva é a predição, maior grau de dispersão demonstra uma maior diferença entre os valores reais na diferença entre a predição, ou seja, muitos erros, sendo assim o modelo não seja adequado.

