

MODELO MATEMÁTICO PARA PREVISÃO DE VENDAS: REGRESSÃO LINEAR SIMPLES

MATHEMATICAL MODEL FOR SALES FORECASTING: SIMPLE LINEAR REGRESSION

ANASTÁCIO PASCOAL ESPANDI CANHANGA^a

Resumo

O presente artigo tem como objetivo apresentar uma das aplicabilidades da Matemática por meio do modelo de regressão linear simples. Fez-se por meio do mesmo modelo a previsão de vendas de uma loja. Utilizou-se para o problema modelo os dados coletados no registro de caixa de uma loja para, aproveitando a incerteza do seu gerente, fazer previsão das vendas, utilizou-se também a folha de cálculo Excel para elaboração dos gráficos e tabelas que se impõem, bem como os respetivos cálculos. Fez-se uma abordagem explicativa de uma função geral para um modelo econométrico com especificações nas suas variáveis. Passou-se em revista aos elementos essenciais que nos proporcionam uma melhor análise da regressão linear simples e os adaptamos às variáveis em estudo que são o custo com publicidade e vendas dos produtos.

Palavras-chave: Regressão linear simples, modelo matemático, previsão de receitas.

Abstract

The present article has as objective to present one of the interesting applicability of the quite simple Mathematics through the model of simple linear regression. It was done through a store sales model. For the elaboration of the model, data were collected in the cash register of a store. Taking advantage of the uncertainty of its manager, the model focused on forecasting sales. It was also used the Excel spreadsheet for the elaboration of the graphics and tables that are imposed, as well as the respective calculations. It was made an explanatory approach of a general function for an econometric model with specifications in their variables. We reviewed the essential elements that provide us with a better analysis of simple linear regression and adapted them

^aEscola Superior Politécnica do Bié - Universidade José Eduardo dos Santos, Cuito, Angola;
ORCID:<https://orcid.org/0000-0003-3255-8087> E-mail: pascoalkanhanga@hotmail.com

to the variables under study, which are the cost of advertising and sales of products.

Keywords: Simple linear regression, mathematical model, revenue forecast.

MSC2020: 62J05

1 Introdução

A motivação deste trabalho foi a preocupação do gerente da recém nascida empresa CHEIRUS LDA acerca do investimento que ele estava fazendo em publicidade e se este investimento iria se refletir em um aumento das vendas num futuro próximo. É daí a preocupação do gerente da empresa CHEIRUS LDA, se o investimento que estava fazendo na publicidade teria resultado ainda num futuro próximo. Este momento de incerteza do gerente, deu-nos, então, oportunidade de mostrarmos uma abordagem quantitativa que fizesse uma previsão das receitas face aos investimentos com a publicidade. Para o efeito, a regressão linear nos proporcionaria as melhores condições para estudar este problema e apresentar uma possível previsão para ajudar o gerente a tomar melhor decisão.

Fez-se a recolha de dados por meio dos registros de caixa da loja, no mês de Janeiro e princípio do mês de Fevereiro, mas também fez-se um levantamento de dados relacionados com a satisfação dos clientes que afluíram à loja no período já referenciado, mas no entanto, por causa do comportamento das variáveis custo e receita, preferiu-se focar-se nos dados da primeira coleta de dados.

Fez-se o estudo do problema recorrendo-se ao método de regressão linear, o que permitiu fazer a análise dos dados proporcionados pela loja e cujos resultados estão apresentados neste trabalho. Assim, o objetivo deste trabalho é resolver um problema de incerteza, divulgando a aplicabilidade da Matemática.

2 Modelo matemático - modelo estatístico

Um modelo deve ser entendido como sendo uma representação simplificada de uma determinada realidade, por meio de símbolos. No entanto, um modelo matemático é uma representação abstrata, por símbolos, de informações concretas ou relacionadas ao realizável por certos conjuntos de relações. Suponhamos a seguinte relação:

$$R = f(C) \tag{1}$$

Sendo que R estabelece a receita ou vendas efetuadas em função do custo C do investimento realizado, por exemplo, com a publicidade. No entanto, o modelo (1) continua sendo um modelo matemático teórico e precisa-se aplicar a este um pendur econométrico. Para isto podemos adotar uma função que relaciona linearmente as variáveis custo e receita:

$$R = \beta_0 + \beta_1 C + u \quad (2)$$

A reta:

$$\hat{R} = \hat{\beta}_0 + \hat{\beta}_1 C \quad (3)$$

é chamada de reta de regressão estimada e será o modelo linear que irá dar o valor aproximado ou esperado da receita ou venda em função de um determinado custo. A constante β_0 é o intercepto da reta de regressão e a constante β_1 é a inclinação da reta. Observamos que se $\beta_1 = 0$, o valor esperado da receita, \hat{R} , não dependerá do valor do custo C ; pois será uma função constante. Como desejamos que as variáveis \hat{R} e C estejam linearmente relacionadas, para o modelo, devemos ter $\beta_1 \neq 0$.

A constante u nos dá informação da variabilidade ou erro aleatório cometido ao se fazer à aproximação dos dados reais sobre a receita, R , pelos valores encontrados na reta de regressão estimada, \hat{R} . Portanto, u é o erro que se comete na previsão, deste modo, está associado à variável resposta.

A equação (2), representa o modelo linear simples.^[1]

2.1 Coeficiente de correlação linear

Considere uma amostra de dados sobre os custos e receitas (vendas) da loja, (C_i, R_i) , $i = 1, \dots, n$, de tamanho n ; e, sejam \bar{R} a média dos dados sobre as receitas e \bar{C} a média sobre os custos coletados.

Designamos de coeficiente de correlação linear a expressão:

$$r = \frac{\sum_{i=1}^n R_i C_i - n \bar{R} \bar{C}}{\sqrt{(\sum_{i=1}^n C_i^2 - n \bar{C}^2)(\sum_{i=1}^n R_i^2 - n \bar{R}^2)}} \quad (4)$$

Este coeficiente, coeficiente de correlação de Pearson^[4], é uma medida do grau de dependência linear as duas variáveis C e R , sendo que $-1 \leq r \leq 1$.

Quando o coeficiente de correlação for se $r = 1$ ou $r = -1$, então a correlação linear entre as variáveis R e C é perfeita, sendo que para $r > 0$, o coeficiente angular

da reta é positivo e para $r < 0$, o coeficiente angular da reta é negativo. Se o coeficiente de correlação tiver valor próximo de zero, isto significa que o modelo linear não é adequado para descrever a relação entre as variáveis dependente e independente do problema. A reta, neste caso, não é uma boa aproximação para os dados coletados..[3]

2.2 Reta regressão estimada

A reta de regressão estimada dada em (3) é:

$$\hat{R} = \hat{\beta}_0 + \hat{\beta}_1 C$$

Partindo do coeficiente de correlação linear, temos:

$$S_{RC} = \sum_{i=1}^n R_i C_i - n \bar{R} \bar{C}$$

$$S_{CC} = \sum_{i=1}^n C_i^2 - n \bar{C}^2$$

$$S_{RR} = \sum_{i=1}^n R_i^2 - n \bar{R}^2$$

Obtem-se que:

$$\hat{\beta}_1 = \frac{S_{RC}}{S_{CC}} \quad e \quad \hat{\beta}_0 = \bar{R} - \hat{\beta}_1 \bar{C} \quad (5)$$

3 Abordagem do problema

Para pôr em prática o que teóricamente analisou-se na primeira parte, eis a seguir um problema de uma nova pequena empresa de venda de cosméticos na cidade do Cuito, CHEIRUS LDA, na província do Bié, Angola. A referida empresa estreou no mercado no final de Novembro. No entanto, em condições de incerteza, o gerente precisou saber se o investimento que estava a fazer na publicidade não valeria a pena ou se era um elemento que se devia levar a sério para aumentar as vendas, atrair para os seus produtos os clientes para depois fidelizá-los. Para ajudar o gerente, foi elaborado um estudo quantitativo para tomada de decisão. Fez-se uma coleta de dados através dos registros do caixa do mês de Janeiro e Fevereiro das vendas feitas naqueles meses, bem como os registros dos investimentos com publicidade feitos pela

empresa. Cujos resultados coletados resultaram na seguinte tabulação:

Tabela 1: Dados coletados - Custos e Receitas (vendas)

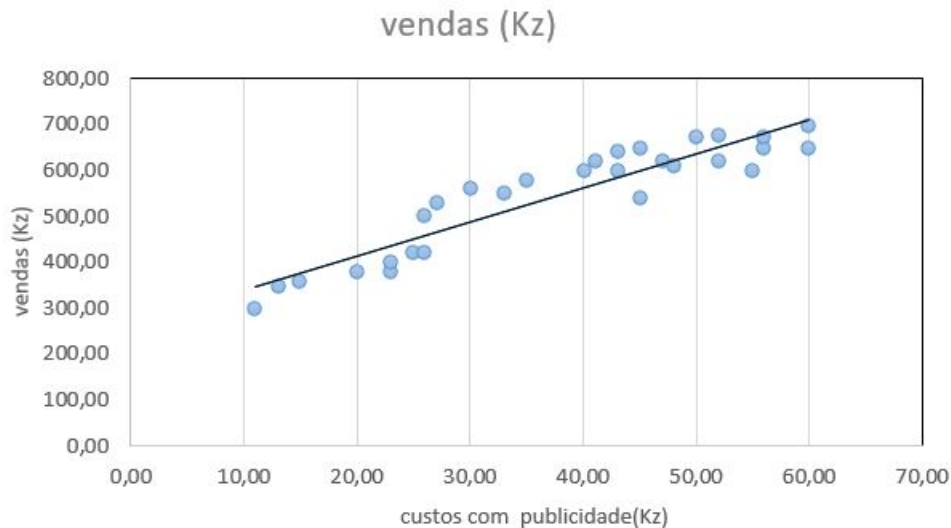
Observações ao dia	custo em mil kwanzas(Kz)	venda em mil Kz
1	11,00	300,00
2	13,00	350,00
3	15,00	360,00
4	20,00	380,00
5	23,00	380,00
6	23,00	400,00
7	25,00	420,00
8	26,00	500,00
9	26,00	420,00
10	27,00	530,00
11	30,00	560,00
12	33,00	550,00
13	35,00	580,00
14	40,00	600,00
15	41,00	620,00
16	43,00	600,00
17	43,00	640,00
18	45,00	540,00
19	45,00	650,00
20	47,00	620,00
21	48,00	610,00
22	50,00	675,00
23	52,00	678,00
24	52,00	620,00
25	55,00	600,00
26	56,00	650,00
27	56,00	675,00
28	60,00	700,00
29	60,00	650,00

Fonte: Elaborado pelo autor com os dados recolhidos.

3.1 Correlação custo e venda

Então, fez-se a representação gráfica das variáveis custo (C) e venda (R) a fim de averiguar qual a correlação entre estas variáveis, ou seja, a análise do coeficiente de definição e análise da variabilidade das variáveis:

Figura 1: Linha de tendência

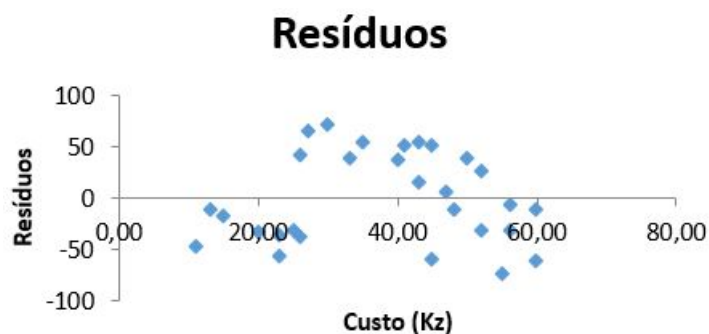


Fonte: Elaborado pelo autor a partir do programa Excel.

O Coeficiente de Correlação Linear encontrado é 0,8638 o que significa que há uma forte relação entre as variáveis custo de publicidade e venda (receitas).

A tendência para se adotar o modelo linear pode também ser observado nas Figuras 2, 3 e 4 mostradas abaixo. Olhando para o gráfico da Figura 2, comparando com a Figura 4, temos indícios de que os erros são normalmente distribuídos, já que os pontos seguem o comportamento da reta. Na Figura 2, podemos ver que os resíduos estão aleatoriamente em torno da linha central, o que diz que os resíduos são independentes. Não tem um padrão aleatório, ou seja, a variável explicativa C não afeta sistematicamente o resultado. Por outro lado, os valores apresentados pela linha ajustada não são iguais aos valores observados, porém são aproximados, visto que os dados observados seguem a mesma tendência da linha ajustada, de acordo com o gráfico da Figura 3.

Figura 2: Representação de Resíduos



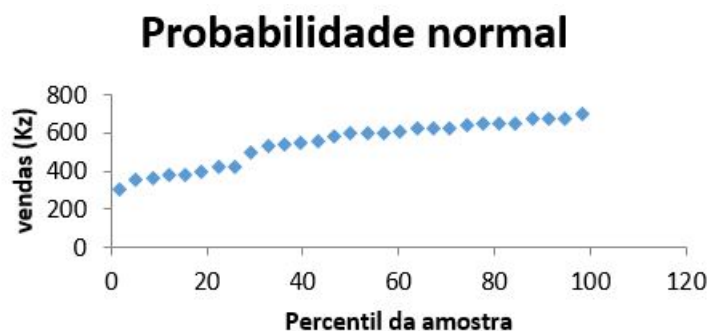
Fonte: Elaborado pelo autor a partir do programa Excel.

Figura 3: Representação do Ajuste de Linha



Fonte: Elaborado pelo autor a partir do programa Excel.

Figura 4: Representação da Probabilidade de distribuição Normal



Fonte: Elaborado pelo autor a partir do programa Excel.

Elaborando os cálculos usando as fórmulas dadas em (5) ou através do programa

Excel, tem-se a equação da reta ajustada:

$$\hat{R} = 265,1563 + 7,42588 * C \quad (6)$$

Ou seja, sem investimento algum na publicidade, a receita que se espera ter por dia é 265,16 mil Kwanzas e portanto, por cada investimento que se faz na publicidade, as vendas aumentam em média a 7,4259 mil Kwanzas.

Figura 5: Resumo dos resultados

<i>Estatística de regressão</i>	
R múltiplo	0,929390869
R-Quadrado	0,863767388
R-quadrado ajustado	0,858721736
Erro padrão	44,38507931
Observações	29

Fonte: Elaborado pelo autor a partir do programa Excel.

3.2 Análise de significância usando ANOVA

Os resultados dos cálculos estatísticos pertinentes ao problema podem ser exibidos em uma tabela denominada tabela de Análise de Variância ou Tabela ANOVA. Portanto, a ANOVA testa a significância da regressão.

Figura 6: Análise da variância - ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>	
Regressão	1	337251,1858	337251,1858	171,19043	3,34581E-13	
Resíduo	27	53190,95217	1970,035265			
Total	28	390442,1379				

	Coeficientes	Erro padrão	Stat t	Valor-p	95%inferior	95%superior
Interseção	265,1563	23,05179	11,5026311	6,49629E-12	217,8579162	312,4546684
Vendas(em mil Kwanzas)	7,42588	0,567555	13,0839761	3,34581E-13	6,261352348	8,590406786

Fonte: Elaborado pelo autor a partir do programa Excel.

Como podemos observar, o F calculado 3,346, que indica a significância da regressão, é maior que o F tabelado 0,05 o que implica que o teste é significativo ao

nível de significância, α , considerado. Caso, em teste de hipótese, em que deve-se rejeitar a hipótese nula a favor da hipótese alternativa.[5, p. 6-7]

Como consequência da rejeição da hipótese nula, a reta estimada (6) é o modelo adequado para descrever o problema proposto.

3.3 Resultados das previsões

Na tabela 4 apresentamos os resultados das previsões, obtidos por meio do programa Excel.

Figura 7: Previsões

<i>Observação</i>	<i>Previstas Vendas (mil Kz)</i>	<i>Resíduos</i>	<i>Resíduos padrão</i>
1	346,8409675	-46,84096752	-1,074696886
2	361,6927267	-11,69272666	-0,268272361
3	376,5444858	-16,54448579	-0,379588815
4	413,6738836	-33,67388363	-0,772597574
5	435,9515223	-55,95152233	-1,283725123
6	435,9515223	-35,95152233	-0,824854633
7	450,8032815	-30,80328146	-0,706735843
8	458,229161	41,77083897	0,958370267
9	458,229161	-38,22916103	-0,877111693
10	465,6550406	64,3449594	1,476300152
11	487,9326793	72,0673207	1,653478338
12	510,210318	39,789682	0,912915544
13	525,0620771	54,93792287	1,260469579
14	562,191475	37,80852503	0,86746082
15	569,6173545	50,38264547	1,15595546
16	584,4691137	15,53088633	0,356333271
17	584,4691137	55,53088633	1,274074251
18	599,3208728	-59,3208728	-1,361029898
19	599,3208728	50,6791272	1,162757796
20	614,1726319	5,827368065	0,133700362
21	621,5985115	-11,5985115	-0,266110733
22	636,4502706	38,54972936	0,88446666
23	651,3020298	26,69797023	0,612545534
24	651,3020298	-31,30202977	-0,718178887
25	673,5796685	-73,57966847	-1,688176926
26	681,005548	-31,00554804	-0,711376551
27	681,005548	-6,005548038	-0,137788439
28	710,7090663	-10,70906631	-0,245703725
29	710,7090663	-60,70906631	-1,39287995

Fonte: Elaborado pelo autor a partir do programa Excel.

Figura 8: Probabilidade de vendas

<i>Percentil</i>	<i>Vendas (mil Kz)</i>
1,724137931	300
5,172413793	350
8,620689655	360
12,06896552	380
15,51724138	380
18,96551724	400
22,4137931	420
25,86206897	420
29,31034483	500
32,75862069	530
36,20689655	540
39,65517241	550
43,10344828	560
46,55172414	580
50	600
53,44827586	600
56,89655172	600
60,34482759	610
63,79310345	620
67,24137931	620
70,68965517	620
74,13793103	640
77,5862069	650
81,03448276	650
84,48275862	650
87,93103448	675
91,37931034	675
94,82758621	678
98,27586207	700

Fonte: Elaborado pelo autor a partir do programa Excel.

4 Conclusão

A regressão linear simples é um modelo matemático que pode ser usado em muitos casos como ferramenta para fazer previsões quantitativas. A utilização de softwares ajuda de forma mais rápida a buscar possíveis soluções de problemas modelados e tomar decisões de forma rápida e assertiva. Depois de analisados os dados com os elementos essenciais da regressão linear simples, podemos concluir que com as

mesmas condições do comportamento de mercado, foi possível fazer previsão de alguns números com vista a dar resposta ao problema com um nível de confiança de 95%.

Referências

- [1] HOFFMANN, R.. Análise de regressão: uma introdução à econometria [recurso eletrônico] / Rodolfo Hoffmann. - 5. ed. Piracicaba: O Autor, 2016. 393 p.
- [2] EDWARDS, A. L.; An Introduction to Linear Regression and Correlation. San Francisco, CA: W. H. Freeman, 1976.
- [3] HILL, C., GRIFFITHS, W.; JUDGE, G.. Econometria, Tradução de Alfredo Alves de Farias, São Paulo: Saraiva, 2000.
- [4] MUKAKA, M.M. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. Malawi Medical Journal; 24(3): 69-71 September 2012.
- [5] VEAZIE, P. J.. Understanding Statistical Testing. SAGE Open January-March 2015: 1 –9