

Aplicativos Computacionais para Geografia - Regressão Linear Simples

AMÂNDIO L. A. TEIXEIRA*
LÚCIA H. O. GERARDI*

INTRODUÇÃO

O presente trabalho tem por objetivo, apresentar um programa para cálculo de regressão linear simples, entre duas variáveis, para n casos.

A REGRESSÃO LINEAR SIMPLES

1) O QUE É

A Regressão Linear Simples é um método de estimativa do relacionamento entre duas variáveis, que expressa uma delas em termos de uma função linear da outra, do tipo

$$y = \pm a \pm bx.$$

Como é sabido, o coeficiente de correlação mede o grau de associação entre dois conjuntos de valores de duas variáveis. Embora significativa, a correlação não diz de que forma as duas variáveis estão relacionadas, de tal forma que não se pode prever uma variável a partir de outra, nem se pode verificar anomalias ou desvios na relação entre os dois conjuntos. Se se quiser resolver estes problemas tem-se que utilizar o recurso da análise de regressão e do estudo dos resíduos.

Análise de regressão consiste simplesmente no ajustamento de uma linha, a mais ajustada possível (best fit) a um conjunto de pontos dispostos num diagrama de dispersão. Esta linha pode ser interpretada como 1. um resumo da relação entre duas variáveis, como 2. uma inferência em relação à qual são considerados os desvios de cada ponto real plotado ou ainda, como 3. um recurso para interpolar ou prever valores desconhecidos de uma variável a partir dos valores da outra.

* Departamento de Planejamento Regional, IGCE - UNESP - Rio Claro.

2) COMO SE FAZ

Assim como a correlação, a regressão não diz quando há relações causais entre as variáveis, porém, no caso da regressão, é imprescindível que se defina a variável dependente (y) e a variável independente (x) definição esta baseada no bom senso. Por exemplo, se há relação entre quantidade de precipitação e volume das colheitas é muito mais provável que a primeira variável influencie a segunda do que vice-versa, de modo que se pode chamar a precipitação de variável independente e as colheitas de variável dependente.

O critério mais comumente utilizado para se ajustar uma reta a um conjunto de pontos e definir sua função matemática é chamado de *método dos mínimos quadrados*, uma vez que se procura ajustar uma reta que induza ao mínimo os desvios de cada ponto em relação a ela.

A equação geral da reta, ajustada segundo este método, é $\hat{y} = \pm a \pm bx$ onde y e x são as variáveis e a e b parâmetros que indicam respectivamente a origem e o gradiente ou inclinação da reta. Note que \hat{y} significa um valor estimado de y para um dado valor de x.

O cálculo de a e b é realizado pelas seguintes fórmulas:

$$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - (\sum x_i)^2 / n} \quad (\text{coeficiente de regressão})$$

onde

$$a = y - bx \quad (\text{base constante})$$

x_i = valor da variável independente para o i-ésimo caso

y_i = valor da variável dependente para o i-ésimo caso

n = número total de casos

b = coeficiente de regressão

a = base constante.

Uma vez ajustada a reta, pode-se através dela fazer-se predição de valores de y (\hat{y}) para cada novo valor de x. Da mesma maneira tendo-se os valores estimados de y (\hat{y}), pode-se avaliar os desvios destes em relação aos valores observados, expressos pelos valores residuais.

O grau de ajustamento da reta de regressão ou, a medida de quanto da variabilidade dos valores observados da variável dependente é apreendido pela regressão, é expresso pelo coeficiente de determinação (r^2), calculado pela fórmula.

$$r^2 = \frac{\sum (\hat{y}_i - y)^2}{\sum (y_i - y)^2}$$

O programa a seguir sugere uma forma de automatização de processo.

PROGRAMA

```
LOAD REGLIN
ÇLIST

5 D$ = ""
6 PRINT D$; "BLOAD ÇMINI, A$6100"
10 REM * PROGRAMA REGLIN *
20 HOME : PRINT "ESTE PROGRAMA C
    ALCULA REGRESSAO LINEAR SIMP
    LES ENTRE DUAS VARIAVEIS"
30 PRINT "A VARIAVEL DEPENDENTE
    DEVE SER 'Y'"
40 FLASH : PRINT : PRINT "LIGUE
    A IMPRESSORA!": NORMAL
50 GET S$
60 HOME
70 INPUT "QUANTOS CASOS? "; N
80 DIM A(N,2), B(N)
90 SX = 0: SY = 0: XX = 0: YY = 0: XY
    = 0: Y2 = 0
100 REM ** ENTRADA DE DADOS **
110 FOR I = 1 TO N
120 PRINT "ENTRE 'X,Y' PARA O CA
    SO "; I; " : "; INPUT A(I,1), A
    (I,2)
125 REM ** CALCULO DOS PARAMETR
    OS DA RETA **
130 SX = SX + A(I,1)
140 SY = SY + A(I,2)
150 XX = XX + A(I,1) ^ 2
160 YY = YY + A(I,2) ^ 2
170 XY = XY + A(I,1) * A(I,2)
180 NEXT I
190 SX = SX / N
200 SY = SY / N
210 B = (XY - N * SX * SY) / (XX -
    N * SX ^ 2)
220 A = SY - B * SX
230 VY = YY / N - SY ^ 2
```

```

235 REM ** IMPRESSAO DOS DADOS
    INICIAIS E PARAMETROS DA RE
    TA **
240 PRINT "DADOS INICIAIS": PRINT
    "X", "Y"
250 PRINT : PRINT
260 FOR I = 1 TO N
270 PRINT A(I,1),A(I,2)
280 NEXT I
290 PRINT : PRINT : PRINT "A", "B
    ": PRINT A,B: PRINT : PRINT
295 REM ** CALCULO E IMPRESSAO
    DO "Y-CALC" E "RESIDUOS" **

300 YY = 0: PRINT "Y-CALC", "RESID
    UOS": PRINT
310 FOR I = 1 TO N
320 YY = A + B * A(I,1)
330 YI = A(I,2) - YY
340 PRINT YY,YI
350 Y2 = Y2 + YY ^ 2
360 NEXT I
365 REM ** CALCULO E IMPRESSAO
    DO COEFICIENTE DE AJUSTAMEN
    TO **
370 YV = Y2 / N - SY ^ 2
380 R2 = YV / VY
390 PRINT : PRINT : PRINT "COEFI
    CIENTE DE AJUSTAMENTO: ";R2
395 REM ** INTERPOLA E EXTRAPO
    LA DADOS **
400 PRINT : PRINT "DESEJA CALCUL
    AR 'Y' PARA ALGUM OUTRO VALO
    R DE 'X'? (S/N): "
410 GET C$: IF C$ = "S" THEN GOTO
    430
415 INPUT "ENTRE OS VALORES DE X
    PARA PLOTAR A RETA DE REGRE

```

```

    SSAO (XMIN,XMAX)";X1,X2
416 Y1 = A + B * X1:Y2 = A + B *
    X2
420 GOTO 600
430 INPUT "ENTRE O VALOR DE X: "
    ;X
440 Y = A + B * X: PRINT "Y= ";Y
450 PRINT "MAIS ALGUM?"
460 GOTO 410
600 PRINT : PRINT : PRINT "QUER
    GRAFICO? (S/N):"
610 GET C$
620 IF C$ = "S" THEN GOTO 640
630 GOTO 650
640 GOSUB 1000
645 GET R$: TEXT
650 END
990 REM ** PLOTA O GRAFICO DE
    DISPERSAO - RETA DE REGRESSA
    O **
1000 POKE 232,0: POKE 233,97
1005 HGR2
1010 ROT= 0: SCALE= 1
1020 HPLLOT 40,30 TO 40,150 TO 24
    0,150
1030 V = 1:C = 200: GOSUB 4000
1035 C = C / (G - P)
1037 GOSUB 5000
1038 X1 = INT ((X1 - P) * C) + 4
    0:X2 = INT ((X2 - P) * C) +
    40
1040 V = 2:C = 110: GOSUB 4000
1045 C = C / (G - P)
1050 GOSUB 5000
1054 Y1 = INT (150 - (Y1 - P) *
    C):Y2 = INT (150 - (Y2 - P)
    * C)
1055 HPLLOT X1,Y1 TO X2,Y2

```

```

1060 FOR J = 1 TO N
1070 A$ = "+"
1080 H% = A(J,1)
1090 G% = A(J,2)
1100 GOSUB 6000
1110 NEXT J
1120 A$ = "GRAFICO DE DISPERSAO -
      RETA DE REGRESSAO":@ = 32:Z
      = 20: GOSUB 1500
1125 A$ = "Y":@ = 25:Z = 30: GOSUB
      1500
1130 A$ = "X":@ = 230:Z = 155: GOSUB
      1500
1140 RETURN
1500 M = LEN (A$): FOR I = 1 TO
      M
1510 @ = @ + 6
1520 W$ = MID$ (A$,I,1)
1530 F = ASC (W$) - 31
1540 DRAW F AT @,Z
1545 NEXT I
1550 RETURN
2000 FOR I = 1 TO N
2010 B(I) = A(I,V)
2020 NEXT I
2030 RETURN
2900 REM ** ORDENA VALORES **
3000 FOR I = 1 TO N - 1
3010 FOR J = I + 1 TO N
3020 IF (B(I) ( = B(J)) THEN GOTO
      3060
3030 M = B(I)
3040 B(I) = B(J)
3050 B(J) = M
3060 NEXT J
3070 NEXT I
3080 RETURN

```

```

4000 GOSUB 2000: GOSUB 3000:P =
      B(1):G = B(N)
4010 RETURN
4900 REM ** TRANSFORMA PARA A
      ESCALA DO GRAFICO **
5000 FOR I = 1 TO N
5005 IF V = 2 THEN GOTO 5011
5010 A(I,V) = INT ((A(I,V) - P) *
      C) + 40: GOTO 5020
5011 A(I,V) = INT (150 - (A(I,V)
      - P) * C)
5020 NEXT I
5030 RETURN
6000 W$ = LEFT$ (A$,1)
6020 F = ASC (W$) - 31
6030 DRAW F AT H% - 1,G% - 4
6040 RETURN

```

3) EXEMPLO DE APLICAÇÃO

A figura 1 mostra a folha de cálculos para a regressão linear simples. No exemplo deseja-se verificar a relação existente entre a precipitação (variável dependente) e a umidade relativa do ar (variável independente) na cidade de Fortaleza (Tabela 1).

A listagem a seguir demonstra os resultados obtidos, tendo sido utilizado o programa proposto.

RUN

ESTE PROGRAMA CALCULA REGRESSAO LINEAR SIMPLIS ENTRE DUAS VARIAVEIS
A VARIÁVEL DEPENDENTE DEVE SER 'Y'

LIGUE'A' IMPRESSORAa

QUANTOS CASOS? 12

ENTRE 'X,Y' PARA O CASO 1 :?85,225.2

ENTRE 'X,Y' PARA O CASO 2 :?87,154.6

ENTRE 'X,Y' PARA O CASO 3 :?87,251,1

TABELA 1 — Precipitação e umidade relativa na cidade de Fortaleza

Meses	p (mm)	u (%)
Janeiro	225,2	87
Fevereiro	154,6	87
Março	251,1	87
Abril	339,8	89
Mai	194,9	90
Junho	499,3	91
Julho	218,5	83
Agosto	36,0	78
Setembro	6,7	77
Outubro	38,8	75
Novembro	3,1	78
Dezembro	2,7	77

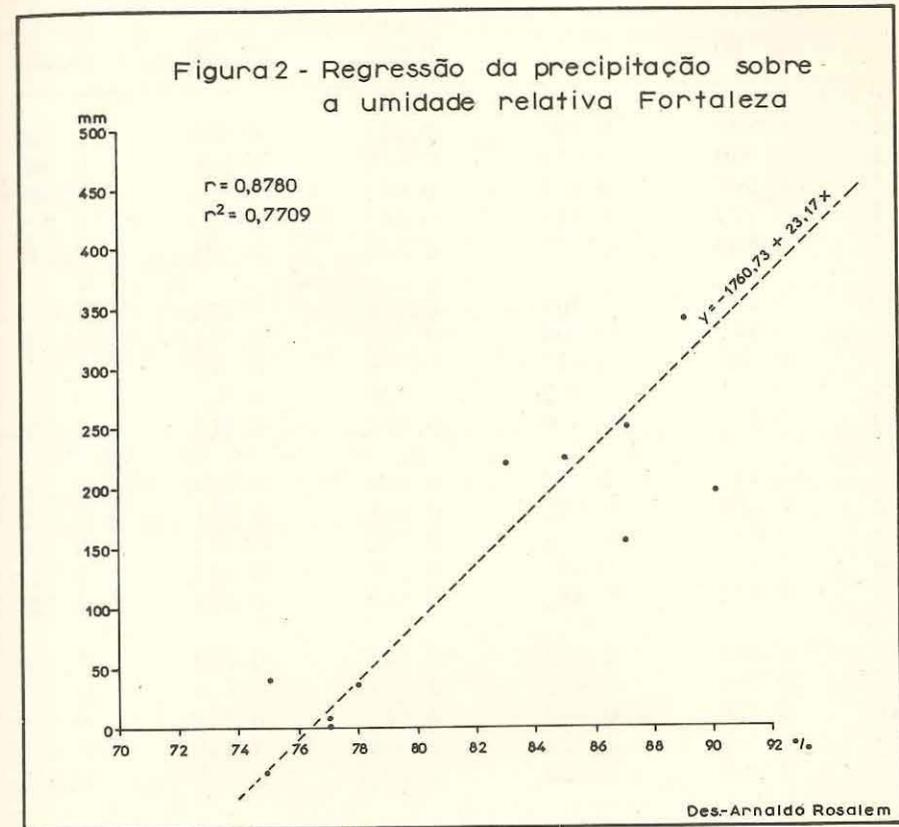
Y-CALC	RESIDUOS
209.409471	15.7905288
256.558485	-101.958485
256.558485	-5.45848489
303.707499	36.0925014
327.282005	-132.382006
350.856512	148.443488
162.260457	56.2395426
44.3879232	-8.38792317
20.8134163	-14.1134163
-26.3355974	65.1355974
44.3879232	-41.2879232
20.8134163	-18.1134163

COEFICIENTE DE AJUSTAMENTO: .770946389

DESEJA CALCULAR 'Y' PARA ALGUM OUTRO VALOR DE 'X'? (S/N):
 ENTRE OS VALORES DE X PARA FLITAR A RETA DE REGRESSAO (XMIN,XMAX)75,95

QUER GRAFICO? (S/N):

O programa além de calcular os parâmetros da reta de regressão, apresenta também, o resultado sob forma gráfica.



Face aos resultados obtidos pode-se concluir que:

1 — As duas variáveis se relacionam na razão de 0.8780, coeficiente este significativo para $\alpha = 0,001$, segundo pode ser observado na tabela 2

TABELA 2 — Coeficiente de correlação r de Karl Pearson
(Valores significativos)

gl	Nível de significância				
	0.100	0.050	0.020	0.010	0.001
1	0.988	0.997	0.999	0.999	0.999
2	0.900	0.950	0.980	0.990	0.999
3	0.805	0.878	0.934	0.959	0.991
3	0.729	0.811	0.882	0.917	0.974
5	0.669	0.755	0.833	0.875	0.951
6	0.622	0.707	0.789	0.834	0.925
7	0.582	0.666	0.750	0.798	0.898
8	0.549	0.632	0.716	0.765	0.872
9	0.521	0.602	0.685	0.735	0.847
10	0.497	0.576	0.658	0.708	0.823
11	0.476	0.553	0.634	0.684	0.801
12	0.458	0.532	0.612	0.661	0.780
13	0.441	0.514	0.592	0.641	0.760
14	0.426	0.497	0.574	0.623	0.742
15	0.412	0.482	0.558	0.606	0.725
16	0.400	0.468	0.543	0.590	0.708
17	0.389	0.456	0.529	0.574	0.693
18	0.378	0.444	0.516	0.561	0.679
19	0.369	0.433	0.503	0.549	0.665
20	0.360	0.423	0.492	0.537	0.652
25	0.323	0.381	0.445	0.487	0.597
30	0.296	0.349	0.409	0.449	0.554
35	0.275	0.325	0.381	0.418	0.519
40	0.257	0.304	0.358	0.393	0.490
45	0.243	0.288	0.338	0.372	0.465
50	0.231	0.273	0.322	0.354	0.443
60	0.211	0.250	0.295	0.325	0.408
70	0.195	0.232	0.274	0.302	0.380
80	0.183	0.217	0.257	0.283	0.357
90	0.173	0.205	0.242	0.267	0.338
100	0.164	0.195	0.230	0.254	0.321

FONTE: Gerardi e Silva (1981), pág. 160.

2 — Analisando-se o gráfico, ou os resultados listados, pode-se ter uma idéia dos desvios dos valores de precipitação para cada mês que poderão ser objeto de análises qualitativas, de acordo com o interesse do pesquisador.

3 — É possível interpolar ou extrapolar novos valores de y para qualquer valor de x em que se tenha interesse. Este recurso permite completar lacunas em séries de dados ou projetar uma seqüência.

BIBLIOGRAFIA

- Gerardi, L. H. O. e Silva, B. C. N. (1981) — *Quantificação em Geografia*, São Paulo, DIFEL, 161 pp.
 Smith, D. M. (1977) — *Patterns in Human Geography*, Harmondsworth, Penguin Books Ltd., 373 pp.
 Taylor, P. J. (1977) — *Quantitative Methods in Geography an introduction to spatial analysis*, Atlanta, Houghton Mifflin Co, 386 pp.
 Yeates, M. (1974) — *An introduction to quantitative analysis in Human Geography*, New York, Mc Graw-Hill, 300 pp.

ABSTRACT: — APPLICATION PROGRAMS FOR GEOGRAPHY: SIMPLE LINEAR REGRESSION: — 8 bits microcomputer program in BASIC language is presented for the solution of simple linear regression, with two variables for 'N' cases

An application exercise is also given.