ANÁLISE DE REGRESSÃO POR POLINÔMIOS ORTOGONAIS

Prof. Ana Lúcia S. S. Mateus

1.1 INTRODUÇÃO

USO: Quando os tratamentos ou níveis de um fator são quantitativos;

- Avaliar os níveis crescentes de nitrogênio na produção de milho
- Avaliar o efeito da adição de água na resistência de um bloco de concreto
- Avaliar o efeito de diferentes doses de um remédio na concentração de histamina no sangue
- Determinar a melhor temperatura sobre a hidrólise da lactose em leite pasteurizado

- A REGRESSÃO é uma técnica de análise que utiliza a relação entre duas ou mais variáveis quantitativas para determinar um modelo matemático de forma que o efeito de uma possa ser previsto por meio de outra variável.

- Na análise de experimentos, em geral utiliza-se modelos polinomais.

- Qualquer modelo de regressão linear e não linear pode ser utilizado.

- O modelo de regressão de grau p, é dado por:

$$Y = \beta_o + \beta_1 X + \beta_2 X^2 + \ldots + \beta_p X^p$$

X (variável independente, explicativa, regressora ou preditora)

Y (variável dependente)

- Quando os níveis do tratamento são equidistantes e com igual número de repetições pode-se utilizar POLINÔMIOS ORTOGONAIS para ajustar o modelo de regressão.

Fórmula geral da Regressão:

$$\hat{Y} = \overline{Y} + B_1 M_1 P_1 + B_2 M_2 P_2 + \ldots + B_p M_p P_p$$

$$\overline{Y} = \frac{G}{IJ}$$

$$B_1 = \frac{\hat{Y}_1}{rK_1} \qquad B_2 = \frac{\hat{Y}_2}{rK_2} \qquad \ldots \qquad B_p = \frac{\hat{Y}_p}{rK_p}$$

$$em \;\; que: \left\{ \hat{Y}_p = \sum_{i=1}^{I} C_{pi} T_i \right. \qquad r - número\; de\; repetições$$

$$M_1, M_2, \ldots, M_p \quad (tabela)$$

$$K = \sum_{i=1}^{I} C_{pi}^2$$

$$P_1 = x \qquad P_2 = x^2 - \frac{n^2 - 1}{12} \quad \ldots. \qquad n - número\; de\; tratamentos$$

$$x = \frac{X - \bar{X}}{q}$$

$$\bar{X} - \textit{média dos tratamentos}$$

$$q - \textit{é a amplitude entre os níveis}$$

- Soma de Quadrados da Regressão

$$\boxed{SQ\,\mathrm{Re}\,g_p = \frac{\hat{Y}_p^2}{r\sum c_{pi}^2} = \frac{\hat{Y}_p^2}{rK_p}}$$

Teste F na Regressão ...

$$\begin{cases} \beta_1 = 0 \\ \beta_1 \neq 0 \end{cases} \quad \textit{Efeito Linear}$$

$$\begin{cases} \beta_2 = 0 \\ \beta_2 \neq 0 \end{cases} \quad \textit{Efeito Quadrático}$$

$$\vdots$$

$$\begin{cases} \beta_p = 0 \\ \beta_p \neq 0 \end{cases} \quad \textit{Efeito de grau } p$$

PASSOS PARA REALIZAÇÃO DA REGRESSÃO NA ANOVA

1- Definir os efeitos da regressão a serem testados. A cada efeito de regressão tem-se 1 grau de liberdade;

2- Definir as Soma de Quadrados de cada um dos efeitos;

3- Testar os efeitos usando Quadrado Médio do Resíduo;

4- Por meio do teste F, definir o grau de polinômio que melhor se ajusta às médias;

5- Determinar o grau de ajuste, através do coeficiente de determinação;

6- Obter as estimativas do parâmtero do modelo de regressão escolhido.

Exemplo: Efeito de doses de gesso na cultura do feijoeiro (*Phaseolus vulgaris* L.) , Ragazzi (1979) utilizou um experimento em DIC com 4 repetições, para estudar os efeitos de 7 doses de gesso (Tratamentos): 0, 50 ,100, 150 , 200, 250 e 300 Kg/ha sobre diversas características do feijoeiro. Para a característica: peso de 1000 sementes, tem-se os resultados:

1.2 OBTENÇÃO DA ANOVA

Tabela 1: Peso de 1000 sementes, em gramas

Tratamentos (Kg)	REPETIÇÕES 1	2	3	4	Totais
1 – 0	134,8	139,7	147,6	132,3	554,4
2 – 50	161,7	157,7	150,3	144,7	614,4
3 – 100	160,7	172,7	163,4	161,3	658,1
4 – 150	169,8	168,2	160,7	161,0	659,7
5 – 200	165,7	160,0	158,2	151,0	634,9
6 – 250	171,8	157,3	150,4	160,4	639,9
7 – 300	154,5	160,4	148,8	154,0	617,7
					4.379,1

Tabela 2: ANOVA preliminar

FV	GL	SQ	QM	Fc	Ft
Tratamentos	6	1.941,83	323,64	7,67**	3,81
Resíduo	21	886,34	42,21	-	-
Total	27	2.828,17	-	-	-

	Fontes de Variação		GL
	Regressão Linear	(ou de 1^{o} grau)	1
	Regressão Quadrática	(ou de 2^{o} grau)	1
	Regressão Cúbica	(ou de 3^{o} grau)	1
Não tem interesse prático	Regressão de 4^{o} Grau		1
	Regressão de 5^{o} Grau		1
	Regressão de 6^{o} Grau		1
	Tratamentos		(6)

Fontes de Variação		GL
Regressão Linear	(ou de 1^{o} grau)	1
Regressão Quadrática	(ou de 2^{o} grau)	1
Regressão Cúbica	(ou de 3^{o} grau)	1
Desvio de Regressão		(3)
Tratamentos		(6)

Tabela 3: Coeficientes (C_i) e totais (T_i) para o experimento com 7 tratamentos

Totais de Tratamentos (Ti)	COEFICIENTES PARA n=7 NÍVEIS		
	1º GRAU	2º GRAU	3º GRAU
	C_{1i}	C_{2i}	C_{3i}
T_1=554,4	-3	+5	-1
T_2=614,4	-2	0	+1
T_3=658,1	-1	-3	+1
T_4=659,7	0	-4	0
T_5=634,9	+1	-3	-1
T_6=639,9	+2	0	-1
T_7=617,7	+3	+5	+1
K	28	84	6
M	1	1	1/6

SQ Regressão Linear

$$Y_{RL} = \sum_{i=1}^{I} C_{1i} T_i \qquad SQY = \frac{\hat{Y}^2}{r\sum c_i^2} \qquad SQRL = \frac{(\hat{Y}_{RL})^2}{rK_1}$$

$$Y_{RL} = -3T_1 - 2T_2 - 1T_3 + 0T_4 + 1T_5 + 2T_6 + 3T_7 \qquad \Rightarrow \hat{Y}_{RL} = 217,7$$

$$r = 4$$

$$K_1 = 28$$

$$SQRL = \frac{(217,7)^2}{4 * 28} = 423,15$$

$$Y_{RQ} = \sum_{i=1}^{I} C_{2i} T_i \qquad SQRQ = \frac{(\hat{Y}_{RQ})^2}{rK_2}$$

SQ Regressão Quadrática

SQ Regressão Cúbica

$$Y_{RC} = \sum_{i=1}^{I} C_{3i} T_i \qquad SQRC = \frac{(\hat{Y}_{RC})^2}{rK_3}$$

SQ Regressão Linear

$$Y_{RL} = \sum_{i=1}^{I} C_{1i} T_i \qquad SQY = \frac{\hat{Y}^2}{r \sum c_i^2} \qquad SQRL = \frac{(\hat{Y}_{RL})^2}{rK_1}$$

$$Y_{RL} = -3T_1 - 2T_2 - 1T_3 + 0T_4 + 1T_5 + 2T_6 + 3T_7 \qquad \Rightarrow \hat{Y}_{RL} = 217,7$$

$$r = 4$$

$$K_1 = 28$$

$$SQRL = \frac{(217,7)^2}{4 * 28} = 423,15$$

$$Y_{RQ} = \sum_{i=1}^{I} C_{2i} T_i \qquad SQRQ = \frac{(\hat{Y}_{RQ})^2}{rK_2}$$

SQ Regressão Quadrática

SQ Regressão Cúbica

$$Y_{RC} = \sum_{i=1}^{I} C_{3i} T_i \qquad SQRC = \frac{(\hat{Y}_{RC})^2}{rK_3}$$

Totais Tratamentos	1º Grau	2º Grau	3º Grau	$\hat{Y}_{RL} = \sum_{i=1}^{I} C_{1i} T_i$	$\hat{Y}_{RQ} = \sum_{i=1}^{I} C_{2i} T_i$	$\hat{Y}_{RC} = \sum_{i=1}^{I} C_{3i} T_i$
554.4	-3	5	-1	-1663.2	2772	-554.4
614.4	-2	0	1	-1228.8	0	614.4
658.1	-1	-3	1	-658.1	-1974.3	658.1
659.7	0	-4	0	0	-2638.8	0
634.9	1	-3	-1	634.9	-1904.7	-634.9
639.9	2	0	-1	1279.8	0	-639.9
617.7	3	5	1	1853.1	3088.5	617.7
Totais				217.7	-657.3	61
K	28	84	6			
M	1	1	0.1667			

$$SQRQ = \frac{(\hat{Y}_{RQ})^2}{rK_2} = \frac{(-657.3)^2}{4*84} = 1285.84$$

$$SQRC = \frac{(\hat{Y}_{RC})^2}{rK_3} = \frac{(61)^2}{4*6} = 155.04$$

Tabela 4: Análise de variância do experimento, para o estudo da regressão

FV	GL	SQ	QM	Fc	Ft
Regressão Linear	1	423,15	423,15	$10{,}02^{**}$	8,02
Regressão Quadrática	1	1.285,84	1.285,84	$30{,}46^{**}$	8,02
Regressão Cúbica	1	155,04	155,04	$3{,}67^{ns}$	8,02
Desvios de regressão	3	77,80	25,93	$0{,}61^{ns}$	4,87
(Tratamentos)	(6)	(1.941,83)	-	-	
Resíduo	21	886,34	42,21	-	
Total	27	2.828,17	-	-	

Quando o teste F para Desvios de regressão for significativo, isso indica que existe alguma regressão significativa, de grau maior que o terceiro, e, se tivermos interesse em estudá-la, devemos desdobrar os Desvios de regressão.

$$F_{(1,21,1\%)} = 8{,}02 \qquad F_{(3,21,1\%)} = 4{,}87$$

FÓRMULA DA REGRESSÃO CONFORME EXEMPLO

$$\boxed{\hat{Y} = \bar{Y} + B_1 M_1 P_1 + B_2 M_2 P_2}$$

$$\bar{Y} = \frac{G}{IJ} = \frac{4379,1}{7*4} = 156,3964$$

$$B_1 = \frac{\hat{Y}_{RL}}{rK_1} = \frac{217,7}{4*28} = 1,9438$$

$$B_2 = \frac{\hat{Y}_{RQ}}{rK_2} = \frac{-657,3}{4*84} = -1,9563$$

$$M_1 = 1\ (tabela)$$

$$M_2 = 1\ (tabela)$$

$$P_1 = x \qquad\qquad P_2 = x^2 - \frac{n^2-1}{12} = x^2 - \frac{7^2-1}{12} = x^2 - 4$$

$$\hat{Y} = 156,3964 + 1,9438x - 1,9563(x^2 - 4) \rightarrow \boxed{\hat{Y} = 164,2216 + 1,9438x - 1,9563x^2} \quad \textbf{(1)}$$

$$x = \frac{X - \bar{X}}{q}$$

$$\bar{X} = \frac{0 + 50 + 100 + 150 + 200 + 250 + 300}{7} = 150$$

q é a diferença entre dois valores sucessivos de X = 50

Substituindo x em (1)

$$\boxed{\hat{Y} = 140,7835 + 0,2737X - 0,000783X^2 \quad (0 \leq X \leq 300)}$$

PREDIÇÃO .....

$$X = 125 Kg / ha$$

$$\hat{Y} = peso\, de\, 1000\, sementes\, (g)$$

$$X = dose\, de\, gesso (Kg / ha)$$

$$\hat{Y}_{(125)} = 140,7835 + 0,2737(125) - 0,000783(125)^2 = 162,72\, g$$

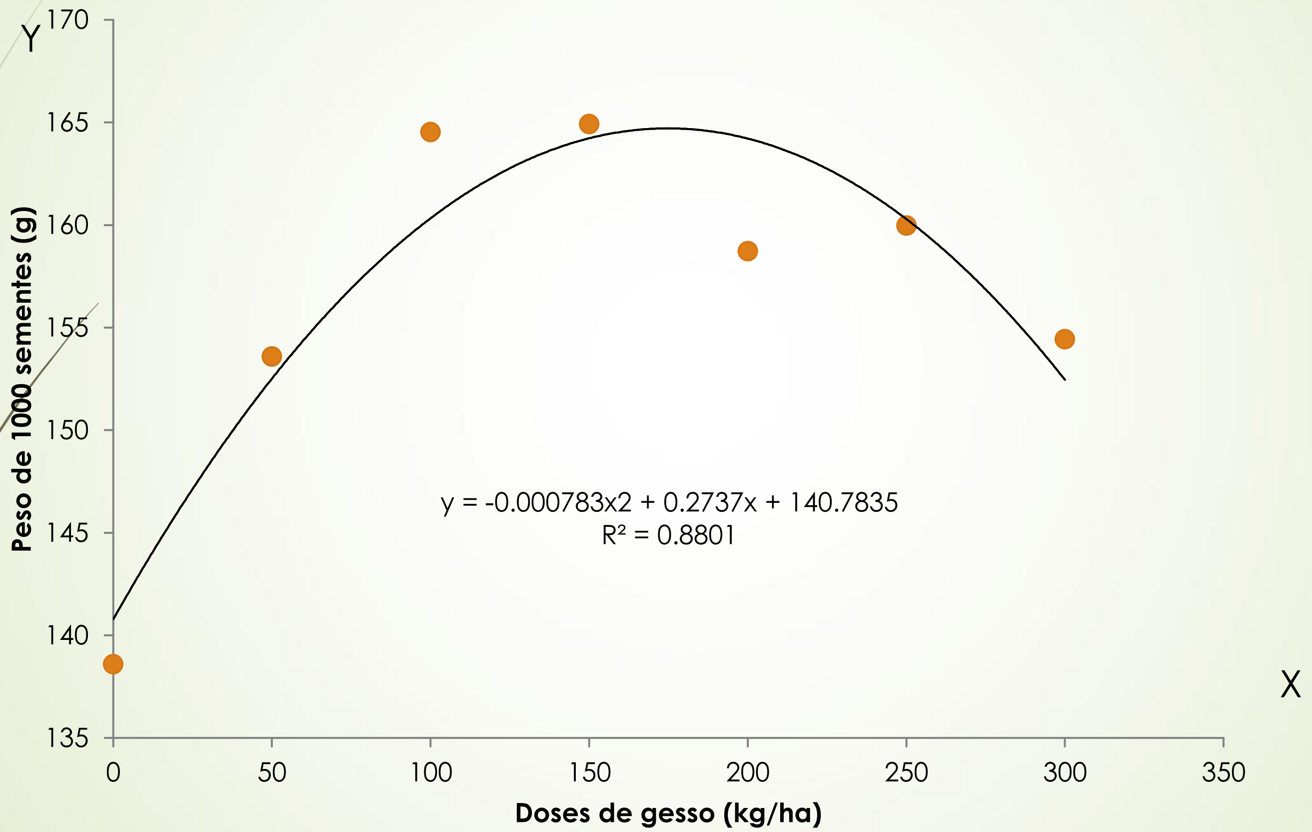
$$R^2 = \frac{S.Q.RL + S.Q.RQ}{SQTratamentos} = \frac{SQ\,\mathrm{Re}\,gressão}{SQTratamentos} = \frac{423,15 + 1285,84}{1941,83} = 0,8801 = 88,01\%$$

TRATAMENTOS (X)	Y_{obs}	$\hat{Y}$
0	138,60	140,78
50	153.60	152,51
100	164,53	160,32
150	164,93	164,22
200	158,73	164,20
250	159,98	160,27
300	154,43	152,42
Soma	1.094,80	1.094,72

Para obter o R^2, devemos somar todas as S.Q. das regressões de grau baixo até aquela que determinou o grau da equação.

R^2, representa, em proporção , o quanto da variação na resposta é explicada pela regressão em questão.

Regressão do peso de 1000 sementes em função das doses de gesso



Podemos verificar se a função tem ponto mínimo ou máximo

$$\frac{\partial \hat{Y}}{\partial X} = 0,2737 - 0,001566X$$

$$\frac{\partial^2 \hat{Y}}{\partial X^2} = -0,001566 \quad (a \ função \ tem \ máximo)$$

O ponto de máximo é aquele que anula a derivada primeira ...

$$0,2737 - 0,001566X = 0$$

$$x = 175Kg / ha$$

Então o máximo da função é dado por:

$$\hat{Y}_{(175)} = 140,7835 + 0,2737(175) - 0,000783(175)^2 = 164,70g$$