Escola Superior de Agricultura “Luiz de Queiroz”
Universidade de São Paulo

Regressão Polinomial e Análise da Variância

Piracicaba
Setembro 2015

Vimos que...

Tratamentos

- **qualitativos.** Exemplos: Variedades de milho, clones de eucalípto, raça, etc.
- **quantitativos.** Exemplos: Nível de adubação, época de semeadura, quantidade de água, teor de nutriente no solo, etc.

Exemplo

Ragazzi (1979) utilizou um experimento inteiramente casualizado com quatro repetições para estudar o efeito de 7 doses de gesso: 0, 50, 100, 150, 200, 250 e 300 kg/ha sobre diversas características do feijoeiro. Para a característica peso de 1000 sementes, obteve os resultados apresentados na Tabela 1.

Tabela: Peso de 1000 sementes de feijão, em g, em função da dose de gesso, em kg/ha

Dose	Peso de 1000 sementes, em g			
0	134,8	139,7	147,6	132,3
50	161,7	157,7	150,3	144,7
100	160,7	172,7	163,4	161,3
150	169,8	168,2	160,7	161,0
200	165,7	160,0	158,2	151,0
250	171,8	157,3	150,4	160,4
300	154,5	160,4	148,8	154,0

Exemplo

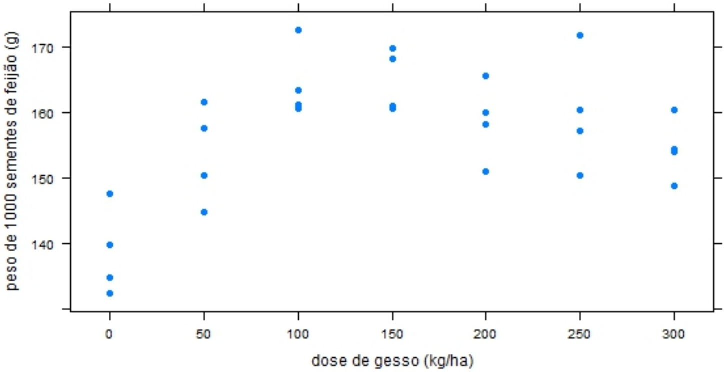


Figura: Peso de 1000 sementes de feijão, em g, em função da dose de gesso, em kg/ha

Exemplo

Tabela: Quadro da análise da variância

Fonte de Variação	gl	SQ	QM	F	Pr>Fc
Doses	6	1941,83	323,64	7,67	0,00018763
Resíduo	21	886,34	42,21		
Total	27	2828,17			

$H_0 : \mu_{D0} = \mu_{D1} = \mu_{D2} = \ldots = \mu_{D6}$
H_1 : pelo menos duas médias diferem entre si.

Exemplo

Tabela: Quadro da análise da variância

Fonte de Variação	gl	SQ	QM	F	Pr>Fc
Doses	6	1941,83	323,64	7,67	0,00018763
Resíduo	21	886,34	42,21		
Total	27	2828,17			

H_1 : pelo menos duas médias diferem entre si.

Há efeito de Dose

Relação funcional

Fatores quantitativos $\Rightarrow$ Relação funcional entre a variável resposta (y) e os níveis desses fatores (x).

Modelo

$$y = f(x) + \epsilon,$$

em que $f(x)$ é uma função desconhecida.

Objetivos:

- Obter uma função que represente $f(x)$ aproximadamente;
- Obter o nível de x que leva à máxima/mínima resposta;
- ...

Relação funcional

Função Polinomial de grau p

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p + \epsilon$$

Características:

- Fácil ajuste;
- Interpretação limitada ao intervalo de estudo;

Falta de Ajuste

Mais de uma observação da variável resposta por nível do fator

$\Downarrow$

Verificação da **Falta de Ajuste**

Falta de Ajuste = Desvios de Regressão

Polinômio

Se I é o número de níveis do fator quantitativo

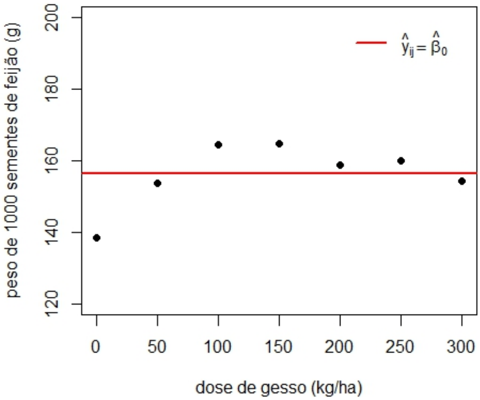
$$\Downarrow$$

Ajuste de um polinômio de no máximo grau $(I - 1)$

No exemplo

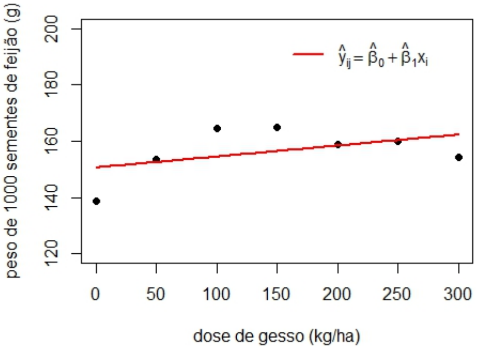
$I = 7$ **doses** de gesso, 0, 50, 100, 150, 200, 250 e 300. Logo podemos ajustar um **polinômio de grau no máximo 6**.

Polinômios: Possíveis ajustes



Não há efeito de dose!

Polinômios: Possíveis ajustes



$$\underbrace{\beta_0 + \beta_1 x}_{\text{modelo ajustado}} + \underbrace{\beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{termos que podemos adicionar no modelo}}$$

Polinômios: Possíveis ajustes

$$\underbrace{\beta_0 + \beta_1 x}_{\text{RL}} + \underbrace{\beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{Desvios de Regressão}}$$

Causas de Variação	gl	
Doses	6	
Regressão Linear		1
Desvios de Regressão		5
Resíduo	21	
Total	27	

Hipóteses:

- Regressão Linear
  $H_0 : \beta_1 = 0 | \beta_0$ está no modelo
  $H_1 : \beta_1 \neq 0 | \beta_0$ está no modelo

Polinômios: Possíveis ajustes

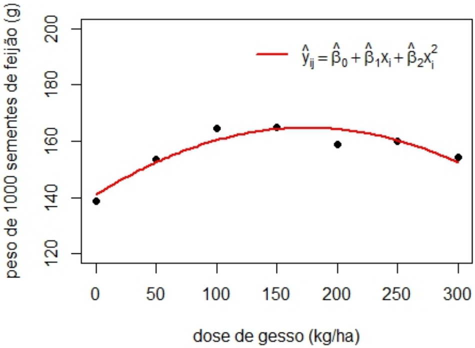
$$\underbrace{\beta_0 + \beta_1 x}_{\text{RL}} + \underbrace{\beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{Desvios de Regressão}}$$

Causas de Variação	gl	
Doses	6	
Regressão Linear		1
Desvios de Regressão		5
Resíduo	21	
Total	27	

Hipóteses:

- Desvios de Regressão
  $H_0 : \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 = 0 | \beta_0, \beta_1$ estão no modelo
  $H_1 : \beta_k \neq 0 | \beta_0, \beta_1$ estão no modelo, para algum $k = 2, \ldots, 6$

Polinômios: Possíveis ajustes



$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2}_{\text{modelo ajustado}} + \underbrace{\color{red}{\beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}}_{\color{red}{\text{termos que podemos adicionar no modelo}}}$$

Polinômios: Possíveis ajustes

$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2}_{\text{modelo quadrático}} + \underbrace{\beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{desvios de regressão}}$$

Causas de Variação	gl	
Doses	6	
Regressão Linear		1
Regressão Quadrática		1
Desvios de Regressão		4
Resíduo	21	
Total	27	

Hipóteses:

- Regressão Quadrática
  $H_0 : \beta_2 = 0 | \beta_0, \beta_1$ estão no modelo
  $H_1 : \beta_2 \neq 0 | \beta_0, \beta_1$ estão no modelo

Polinômios: Possíveis ajustes

$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2}_{\text{modelo quadrático}} + \underbrace{\beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{desvios de regressão}}$$

Causas de Variação	gl	
Doses	6	
Regressão Linear		1
Regressão Quadrática		1
Desvios de Regressão		4
Resíduo	21	
Total	27	

Hipóteses:

- Desvios de Regressão
  $H_0 : \beta_3, \beta_4, \beta_5, \beta_6 = 0 | \beta_0, \beta_1, \beta_2$ estão no modelo
  $H_1 : \beta_k \neq 0 | \beta_0, \beta_1, \beta_2$ estão no modelo, para algum $k = 3, \ldots, 6$

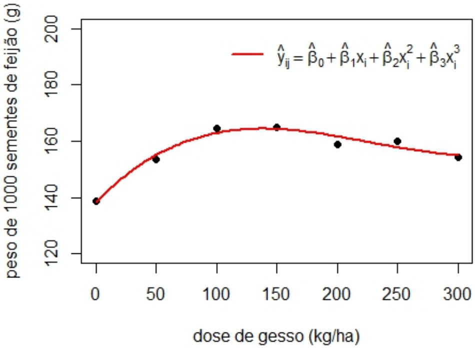
Polinômios: Possíveis ajustes

Procedimento...

- Se Desvios de Regressão for **não significativo** $\Rightarrow$ verificar a significância da Regressão Quadrática;

- Se Desvios de Regressão for **significativo** $\Rightarrow$ continuar "procurando" o modelo.

Polinômios: Possíveis ajustes



$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3}_{\text{ajustado}} + \underbrace{\beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{termos que podemos adicionar no modelo}}$$

Polinômios: Possíveis ajustes

$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3}_{\text{modelo cúbico}} + \underbrace{\beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{desvios de regressão}}$$

Causas de Variação	gl	
Doses	6	
Regressão Linear		1
Regressão Quadrática		1
Regressão Cúbica		1
Desvios de Regressão		3
Resíduo	21	
Total	27	

Hipóteses:

- Regressão Cúbica
  $H_0 : \beta_3 = 0 | \beta_0, \beta_1, \beta_2$ estão no modelo
  $H_1 : \beta_3 \neq 0 | \beta_0, \beta_1, \beta_2$ estão no modelo

Polinômios: Possíveis ajustes

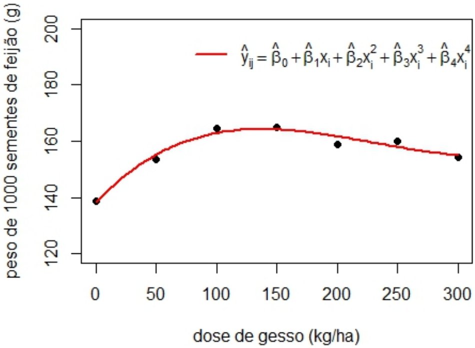
$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3}_{\text{modelo cúbico}} + \underbrace{\beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{desvios de regressão}}$$

Causas de Variação	gl	
Doses	6	
Regressão Linear		1
Regressão Quadrática		1
Regressão Cúbica		1
Desvios de Regressão		3
Resíduo	21	
Total	27	

Hipóteses:

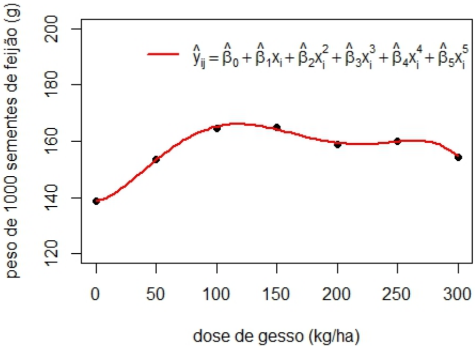
- Desvios de Regressão
  $H_0 : \beta_4, \beta_5, \beta_6 = 0 | \beta_0, \beta_1, \beta_2, \beta_3$ estão no modelo
  $H_1 : \beta_k \neq 0 | \beta_0, \beta_1, \beta_2, \beta_3$ estão no modelo, para algum $k = 4, 5, 6$

Polinômios: Possíveis ajustes



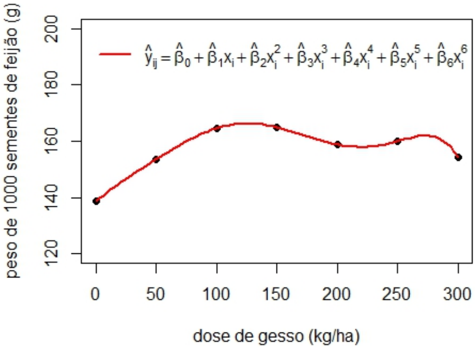
$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4}_{\text{ajustado}} + \underbrace{\color{red}{\beta_5 x^5 + \beta_6 x^6}}_{\color{red}{\text{podemos adicionar}}}$$

Polinômios: Possíveis ajustes



$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5}_{\text{ajustado}} + \underbrace{\beta_6 x^6}_{\text{podemos adicionar}}$$

Polinômios: Possíveis ajustes



$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6}_{\text{ajustado}}$$

Falta de Ajuste: generalizando

Mais de uma observação da variável resposta por nível do fator

$$\Downarrow$$

Verificação da **Falta de Ajuste**

$$\underbrace{\beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p}_{\text{modelo ajustado}} + \underbrace{\beta_{p+1} x^{p+1} + \beta_{p+2} x^{p+2} + \ldots + \beta_{I-1} x^{I-1}}_{\text{termos que podemos adicionar no modelo}}$$

Falta de Ajuste: generalizando

Hipóteses:

$$\begin{cases} H_0 : \text{Não há falta de ajuste no modelo} \\ H_1 : \text{Há falta de ajuste no modelo} \end{cases}$$

$$\begin{cases} H_0 : \beta_1 = 0|\beta_0 \text{ está no modelo} \\ H_1 : \beta_1 \neq 0|\beta_0 \text{ está no modelo} \end{cases}$$

Fontes de Variação	gl
Tratamentos	I-1
Regressão linear $(\beta_1\|\beta_0)$	1
Falta de Ajuste $(\beta_2, \ldots, \beta_{I-1}\|\beta_0, \beta_1)$	I-2
Resíduo	I(J-1)
Total	IJ-1

Falta de Ajuste: generalizando

Hipóteses:

$$\begin{cases} H_0 : \text{Não há falta de ajuste no modelo} \\ H_1 : \text{Há falta de ajuste no modelo} \end{cases}$$

$$\begin{cases} H_0 : \beta_2 = 0 | \beta_0, \beta_1 \text{ estão no modelo} \\ H_1 : \beta_2 \neq 0 | \beta_0, \beta_1 \text{ estão no modelo} \end{cases}$$

Fontes de Variação	gl
Tratamentos	I-1
Regressão linear $(\beta_1 \vert \beta_0)$	1
Regressão quadrática $(\beta_2 \vert \beta_0, \beta_1)$	1
Falta de Ajuste $(\beta_3, \ldots, \beta_{I-1} \vert \beta_0, \beta_1, \beta_2)$	I-3
Resíduo	I(J-1)
Total	IJ-1

Falta de Ajuste: generalizando

Hipóteses:

$$\begin{cases} H_0: \text{Não há falta de ajuste no modelo} \\ H_1: \text{Há falta de ajuste no modelo} \end{cases}$$

$$\begin{cases} H_0: \beta_3 = 0 | \beta_0, \beta_1, \beta_2 \text{ estão no modelo} \\ H_1: \beta_3 \neq 0 | \beta_0, \beta_1, \beta_2 \text{ estão no modelo} \end{cases}$$

Fontes de Variação	gl
Tratamentos	I-1
Regressão linear $(\beta_1 \vert \beta_0)$	1
Regressão quadrática $(\beta_2 \vert \beta_0, \beta_1)$	1
Regressão cúbica $(\beta_3 \vert \beta_0, \beta_1, \beta_2)$	1
Falta de Ajuste $(\beta_4, \ldots, \beta_{I-1} \vert \beta_0, \beta_1, \beta_2, \beta_3)$	I-4
Resíduo	I(J-1)
Total	IJ-1

Falta de Ajuste: generalizando

Observação

Aumentamos progressivamente o grau do polinômio ajustado (p) até que a **falta de ajuste** do modelo seja **não significativa** e que a conclusão do teste da hipótese:

$$\begin{cases} H_0 : \beta_p = 0 | \beta_0, \beta_1, \ldots, \beta_{p-1} \text{ estão no modelo} \\ H_1 : \beta_p \neq 0 | \beta_0, \beta_1, \ldots, \beta_{p-1} \text{ estão no modelo} \end{cases}$$

seja pela rejeição de H_0.

Ajuste Polinomial

Polinômios ortogonais

$$\Downarrow$$

Contrastes ortogonais

Coeficientes $\Rightarrow$ Somas de Quadrados

Coeficientes

I = 7 níveis					
	1^o grau	2^o grau	3^o grau	4^o grau	5^o grau
	-3	+5	-1	+3	-1
	-2	0	+1	-7	+4
	-1	-3	+1	+1	-5
	0	-4	0	+6	0
	+1	-3	-1	+1	+5
	+2	0	-1	-7	-4
	+3	+5	+1	+3	+1
K	28	84	6	154	84
M	1	1	1/6	7/12	7/20

Tabela do Coeficientes

Exemplo

Dose	Total	Média
0	554,4	138,600
50	614,4	153,600
100	658,1	164,525
150	659,7	164,925
200	634,9	158,725
250	639,9	159,975
300	617,7	154,425
	4379,1	

Exemplo

No nosso exemplo temos:

- Regressão Linear

Fontes de Variação	gl	SQ	QM	F	p.valor
Doses	6	1941,83	323,64	7,67	0,00018763
Regressão linear	1	423,15	423,15	10,03	0,00465
Desvios de Regressão	5	1518,68	303,74	7,20	**0,00046**
Resíduo	21	886,34	42.21		
Total	27	2828,17			

Exemplo

No nosso exemplo temos:

- Regressão Quadrática

Fontes de Variação	gl		SQ	QM	F	p.valor
Doses	6		1941,83	323,64	7,67	0,00018763
Regressão linear		1	423,15	423,15	10,03	0,00465
Regressão quadrática		1	1285,84	1285,84	30,47	2×10^{-5}
Desvios de Regressão		4	232,83	58,21	1,38	**0,27505**
Resíduo	21		886,34	42.21		
Total	27		2828,17			

Estimação dos Parâmetros

Obter β_0, β_1, ..., β_p, tais que

$$SQ = \sum_i \left[y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p)\right]^2$$

seja mínima.

Ou seja,

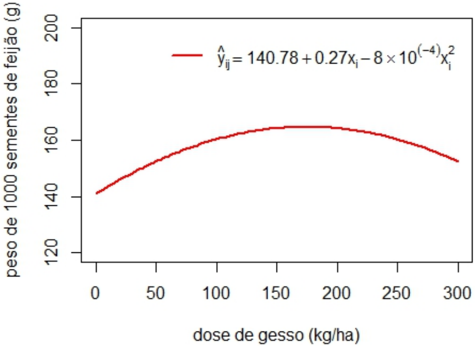
$$\left.\begin{array}{l} \frac{\partial SQ}{\partial \beta_0} = 0 \\ \frac{\partial SQ}{\partial \beta_1} = 0 \\ \ldots \\ \frac{\partial SQ}{\partial \beta_p} = 0 \end{array}\right\} \Rightarrow \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$$

Solução de Mínimos Quadrados

Exemplo

Modelo ajustado

$$\hat{\mathbf{y}} = \mathbf{140,7839286 + 0,2736250x - 0,0007825x^2}$$



Coeficiente de Determinação

Definição

$$R^2 = \frac{\text{SQ Modelo}}{\text{SQ Tratamentos}} = 1 - \frac{\text{SQ Falta de Ajuste}}{\text{SQTratamentos}}$$

$$0 \leq R^2 \leq 1$$

- Proporção da variabilidade devida a tratamentos que é explicada pelo modelo de regressão;
- Quão maior o grau do polinômio, maior será o coeficiente de determinação.