

7 Análise de covariância (ANCOVA)

7.1 Introdução

Em alguns experimentos, pode ser muito difícil e até impossível obter unidades experimentais semelhantes. Por exemplo, pode-se não ter cobaias de mesmo peso, motores com o mesmo tempo de funcionamento, corpos de prova com mesmo tamanho etc. Mas, em todos estes casos, quando a condição inicial da unidade experimental for conhecida e puder ser medida, e ainda, que seja conhecido que esta condição inicial (uma variável) tenha influência sobre a variável resposta, pode-se utilizar esta informação para corrigir a variável resposta. Tal, procedimento pode ser feito utilizando-se da técnica de análise de covariância.

A variável medida na condição inicial da unidade experimental é chamada de *covariável* (ou ainda, variável auxiliar, variável concomitante). Em um mesmo experimento, pode haver mais de uma covariável.

A covariável complementa o controle local e na grande maioria das situações simplesmente o substitui.

Obviamente, a covariável necessita estar correlacionada com a variável resposta para que se possa fazer uso de tal análise.

Quando a Análise de Variância é realizada com uma ou mais covariáveis, usa-se chamar a análise de ANCOVA². A ANCOVA permite que se faça um “ajuste” do efeito de uma variável resposta que sofreu influência de uma variável ou uma causa de variação não controlada.

A ANCOVA, permite, portanto, um controle do erro experimental, aumentando a precisão do experimento. é possível também fazer o ajuste das médias dos tratamentos em função da(s) covariável(eis) e, em alguns casos, estimar observações perdidas durante o experimento.

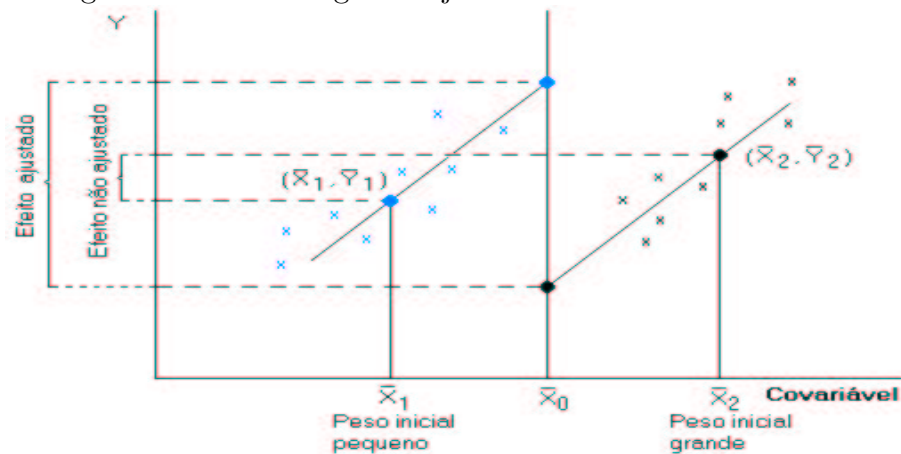
Para que uma covariável possa ser assim considerada, deve-se garantir que ela não seja afetada pelo tratamento. Por exemplo: ao utiliza-se como covariável o número de animais sobreviventes em uma gaiola, deve-se garantir que a causa da morte ou da perda dos animais durante o experimento não seja causada pelo efeito do tratamento. Neste caso, o uso da covariável é incorreto, pois será eliminado da análise uma possível fonte de variação conhecida: o próprio efeito do tratamento.

²Analysis of covariance

Um situação bastante comum é quando existem animais de pesos diferentes e a variável resposta de interesse é o peso final dos animais. Neste caso, antes do início do experimento, o peso inicial dos animais é obtido e utilizado como covariável no experimento.

Graficamente, a forma de correção da variável resposta através da ANCOVA pode ser vista na figura 14.

Figura 14: Metodologia de ajuste da análise de covariância.



7.2 Modelo estatístico

Considere um experimento com um fator e uma covariável. O modelo estatístico pode ser escrito da seguinte maneira:

$$y_{ij} = \mu + \alpha_i + \beta(X_{ij} - \bar{X}) + \epsilon_{ij} \quad (23)$$

onde

μ = constante;

α_i =efeito do i-ésimo tratamento;

X_{ij} = valor observado da covariável;

\bar{X} = média da covariável;

β = coeficiente de regressão linear entre a covariável (X) e a variável resposta (Y), com $\beta \neq 0$. Neste caso, a relação deve ser linear.

Neste modelo, assume-se que a variável resposta e a covariável estão relacionadas linearmente.

Análise de Variância

Tabela 59: Análise de covariância.

	GL	SQ e Prod. Cruzados			Ajuste pela Regressão		
CV	GL	xx	xy	yy	y	GL	QM
Trat	a-1	T_{xx}	T_{xy}	T_{yy}			
Erro	a(n-1)	E_{xx}	E_{xy}	E_{yy}	$SQE = E_{yy} - \frac{(E_{xy})^2}{E_{xx}}$	a(n-1)-1	$\frac{SQE}{a(n-1)-1}$
Total	an -1	S_{xx}	S_{xy}	S_{yy}	$SQE' = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$	an-2	
Trat. aj.					$SQE - SQE'$	a-1	$\frac{SQE' - SQE}{a-1}$

Onde

$$S_{yy} = \sum_{ij}^{an} y_{ij} - \frac{(y_{..})^2}{an}$$

$$S_{xx} = \sum_{ij}^{an} x_{ij} - \frac{(x_{..})^2}{an}$$

$$S_{xy} = \sum_{ij}^{an} x_{ij}y_{ij} - \frac{(y_{..})(x_{..})}{an}$$

$$T_{xx} = \sum_i^a \frac{x_{i.}}{n} - \frac{(x_{..})^2}{an}$$

$$T_{xy} = \sum_i^a \frac{(x_{i.})(y_{i.})}{n} - \frac{(x_{..})(y_{..})}{an}$$

$$T_{yy} = \sum_i^a \frac{y_{i.}}{n} - \frac{(y_{..})^2}{an}$$

$$E_{yy} = S_{yy} - T_{yy}$$

$$E_{xx} = S_{xx} - T_{xx}$$

$$E_{xy} = S_{xy} - T_{xy}$$

e

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} \quad (24)$$

O valor de F para tratamentos é obtido por:

$$F = \frac{\frac{SQE' - SQE}{A-1}}{\frac{SQE}{a(n-1)-1}} \quad (25)$$

E a hipótese $H_0 : \beta = 0$ pode ser testada por

$$F = \frac{\frac{(E_{xy})^2}{E_{xx}}}{\frac{E_{yy} - \frac{(E_{xy})^2}{E_{xx}}}{a(n-1)-1}} \sim F_{\alpha}(1, a(n-1) - 1) \quad (26)$$

O ajuste da variável observada Y , pode ser entendido como

$$y_{ij} - \beta(X_{ij} - \bar{X}) = \mu + \alpha_i + \epsilon_{ij} \quad (27)$$

O ajuste de médias de tratamentos, pode ser feito da seguinte maneira

$$\bar{y}'_i = \bar{y}_i - \hat{\beta}(\bar{x}_i - \bar{x}_{..}) \quad (28)$$

7.3 Exemplo:

Considere o seguinte conjunto de dados onde foi medido a resistência de fios de algodão. A resposta avaliada foi o comprimento (cm) que o fio atingiu antes de se romper. Como cada fio possui um diâmetro diferente, e isso afeta a resistência, utilizou-se essa informação como covariável. Os dados estão na tabela 60. Três tipos de máquinas foram comparadas neste experimento.

O primeiro passo da análise é verificar se existe relação linear entre a variável e a covariável. Esta relação deve ser pelo menos aproximada.

O passo seguinte é calcular as somas de quadrados e produtos cruzados.

$$S_{yy} = \sum_{i=1}^3 \sum_{j=1}^5 y_{ij}^2 - \frac{(y_{..})^2}{an} = (36^2 + \dots + 32^2) - \frac{(603)^2}{3 \times 5} = 346,40$$

Tabela 60: Comprimento (Y) e diâmetro (X) de fios de algodão.

Observação	Máquina 1		Máquina 2		Máquina 3	
	Y	X	Y	X	Y	X
1	36	20	40	22	35	21
2	41	25	48	28	37	23
3	39	24	39	22	42	26
4	42	25	45	30	34	21
5	49	32	44	28	32	15
Total	207	126	216	130	180	106

$$S_{xx} = \sum_{i=1}^3 \sum_{j=1}^5 x_{ij}^2 - \frac{(x_{..})^2}{an} = (20^2 + \dots + 15^2) - \frac{(362)^2}{3 \times 5} = 261,73$$

$$S_{xy} = \sum_{i=1}^3 \sum_{j=1}^5 (x_{ij})(y_{ij}) - \frac{(x_{..})(y_{..})}{an} = (20 \times 36 + \dots + 15 \times 32) - \frac{(362)(603)}{3 \times 5} = 282,60$$

$$T_{yy} = \sum_{i=1}^3 \frac{(y_{i.})^2}{n} - \frac{(y_{..})^2}{an} = \frac{(207^2 + 216^2 + 180^2)}{5} - \frac{(603)^2}{15} = 140,4$$

$$T_{xx} = \sum_{i=1}^3 \frac{(x_{i.})^2}{n} - \frac{(x_{..})^2}{an} = \frac{(126^2 + 130^2 + 106^2)}{5} - \frac{(362)^2}{15} = 66,13$$

$$T_{xy} = \sum_{i=1}^3 \frac{(x_{i.})(y_{i.})}{n} - \frac{(x_{..})(y_{..})}{an} = \frac{(126 \times 207 + 130 \times 216 + 106 \times 180)}{5} - \frac{(362)(603)}{15} = 96$$

$$E_{yy} = 346,4 - 140,4 = 206,0$$

$$E_{xx} = 261,73 - 66,13 = 195,6$$

$$E_{xy} = 282,6 - 96 = 186,6$$

$$SQE' = 346,4 - \frac{(282,6)^2}{261,73} = 41,27$$

$$SQE = 206 - \frac{(186,6)^2}{195,6} = 27,99$$

$SQE' - SQE = 41,27 - 27,99 = 13,28$ que é a soma de quadrados de tratamentos ajustada para a covariável.

O Coeficiente de regressão $\hat{\beta}$ pode ser obtido por

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} = \frac{186,6}{195,6} = 0,954$$

Pode-se testar a hipótese $H_0 : \beta = 0$ por

$$F_{\beta} = \frac{(186,6)^2 / 195,6}{2,54} = 70,08$$