

TAMANHO AMOSTRAL ENVOLVENDO MÚLTIPLOS CARACTERES: UMA APLICAÇÃO À DIGESTIBILIDADE EM BOVINOS¹

ALFREDO RIBEIRO DE FREITAS², AIRTON MANZANO² e SÉRGIO NOVITA ESTEVES³

RESUMO - O objetivo deste trabalho foi estimar o tamanho amostral mínimo (n) para comparar tratamentos em experimentos de consumo e de digestibilidade com bovinos, envolvendo múltiplos caracteres. Foram utilizados dados de digestibilidade de 72 novilhas com média de 18 meses de idade e 250 kg de peso. O experimento foi realizado na Embrapa-Centro de Pesquisa de Pecuária do Sudeste, São Carlos, SP, de 1988 a 1989, em delineamento inteiramente casualizado, com nove tratamentos organizados em esquema fatorial 3 x 3 (três grupos genéticos: Canchim, ½ Canchim + ½ Nelore e Nelore, e três níveis de proteína bruta: 6, 10 e 13%, com oito repetições cada, sendo a unidade experimental a novilha). Foram analisados consumo de ração (g/kg^{0.75}) por quilograma de peso metabólico, energia digestível, nitrogênio retido (NR), NR (mg/kg^{0.75}) e digestibilidades da matéria seca, proteína bruta, fibra em detergente neutro e fibra em detergente ácido. O valor mínimo de n, que permite detectar diferenças significativas (Δ) entre vetores de médias de tratamentos, foi obtido por meio de um programa SAS (Statistical Analysis System), considerando modelo de distribuição normal t-variada, média zero e matriz de covariância Σ , estatística T² de Hotelling, distribuição F com parâmetro de não centralidade (δ^2_{Δ}), erros do tipo I (α), poder do teste (1- β) e Δ . O valor de n variou de 6 a 47, sendo mais influenciado por alteração nos valores de Δ , do que nos valores de α e poder do teste.

Termos para indexação: distribuição F, diferença mínima significativa, Hotelling.

SAMPLE SIZE INVOLVING MULTIPLE TRAITS: AN APPLICATION TO DIGESTIBILITY IN CATTLE

ABSTRACT - The objective of this work was to estimate minimum sample size (n) for comparison of treatments in experiments of consumption and digestibility of cattle in a multitrait analysis. Data from digestibility of 72 heifers aging approximately 18 months and with 250 kg of live weight were used. A completely randomized experiment was conducted at Embrapa-Centro de Pesquisa de Pecuária do Sudeste, São Carlos, SP, Brazil, from 1988 to 1989, considering the combination of nine treatments organized in a factorial 3 x 3 (three genetic groups: Canchim, Nelore and ½ Canchim + ½ Nelore, and three dietary levels of crude protein: 6, 10 and 13%, and eight heifers each). The end points analyzed were feed consumption (g/kg^{0.75}), digestible energy, retained nitrogen (RN), RN (mg/kg^{0.75}), and digestibilities of dry matter, crude protein, neutral detergent fiber and acid detergent fiber. The minimum sample size required to detect significant differences (Δ) between mean vectors was calculated by a SAS program considering a t-variate normal distribution with mean vector 0 and covariance matrix Σ , Hotelling's T² test, F-distribution with noncentrality parameter (δ^2_{Δ}) and combinations of type I error (α) and power of the test (1- β) and Δ . The calculated n ranged from 6 to 47, and was affected more by variation in Δ than variation in α or power of the test.

Index terms: F-distribution, minimum significant difference, Hotelling's.

INTRODUÇÃO

Em ensaios de digestibilidade em bovinos, fatores como o uso de animais fistulados, instalações, disponibilidade de animais, entre outros,

¹ Aceito para publicação em 30 de outubro de 1998.

² Eng. Agr., Ph.D., Embrapa-Centro de Pesquisa de Pecuária do Sudeste (CPPSE), Caixa Postal 239, CEP 13560-970 São Carlos, SP. E-mail: ribeiro@cppsse.embrapa.br

³ Méd. Vet., Ph.D., Embrapa-CPPSE.

limitam ou dificultam a condução de pesquisas na área. Deste modo, deve-se determinar o número de animais a ser utilizado nos experimentos, de modo a fazer inferências precisas.

É fundamental determinar o número de indivíduos necessários (n) para testar os efeitos entre tratamentos, quando um conjunto de t caracteres é analisado no animal. Para a determinação de n , admitiu-se que a amostra era proveniente de uma população com distribuição normal multivariada, sendo adotado o modelo $Y = X\beta + \varepsilon_i$ de dimensões $n \times p$, $n \times k$, $k \times p$ e $n \times p$, respectivamente, sendo p o número de variáveis dependentes, k o número de parâmetros e n o número de observações. Definem-se ainda $y_{ij} = (y_{ij}^{(1)}, y_{ij}^{(2)}, \dots, y_{ij}^{(p)})$ como sendo o vetor de resposta do j -ésimo indivíduo do i -ésimo tratamento e avaliado em p variáveis e, μ_i ($i = 1, \dots, t$), o vetor de média $p \times 1$ associado ao tratamento i .

O objetivo deste trabalho foi estimar o número de indivíduos necessários para detectar significância entre vetores de médias de tratamentos em experimentos de consumo e de digestibilidade em bovinos.

MATERIAL E MÉTODOS

Obtenção dos dados

Foram utilizados dados de digestibilidade de 72 novilhas com média de idade de 18 meses e peso de 250 kg. O experimento, em que a unidade experimental foi representada pela novilha, foi realizado na Embrapa-Centro de Pesquisa de Pecuária do Sudeste (CPPSE), São Carlos, SP, de 1988 a 1989, e constou de três períodos. Em cada período, com 24 animais, foram avaliados três grupos genéticos (Canchim, 1/2 Canchim + 1/2 Nelore e Nelore) e uma dieta. O primeiro período (6% de PB) iniciou em fevereiro de 1988, o segundo (10% de PB), em agosto de 1988, e o terceiro (13% de PB), em maio de 1989. A relação concentrado: volumoso das rações de cada período foi de 25:75, 40:60 e 60:40, respectivamente. Cada período constou de 121 dias, sendo 21 para adaptação dos animais à dieta e ao confinamento, 90 para avaliação do consumo voluntário, e dez para ganho de peso e conversão alimentar, sendo três para adaptação e sete para colheita de fezes e urina, para determinação dos coeficientes de digestibilidade e do balanço de N. Os caracteres analisados foram o consumo de ração por

quilograma de peso metabólico ($\text{kg}^{0,75}$) (CPM); digestibilidade da energia bruta (%) (DEB); digestibilidade da fibra em detergente ácido (%) (DFDA); digestibilidade da fibra em detergente neutro (%) (DFDN); digestibilidade da matéria seca (%) (DMS); digestibilidade da proteína bruta (%) (DPB); N retido (g/dia) (NR) e N retido por quilograma de peso metabólico ($\text{mg}/\text{kg}^{0,75}$) (NPM).

Método de análise

Os dados de digestibilidade, considerando confundimento entre efeitos de períodos e dietas, foram analisados por meio do procedimento GLM do SAS (SAS, 1993), conforme modelo abaixo:

$$y_{ijk} = \mu + D_i + G(D)_{j(i)} + \beta(X_{ijk} - \bar{X}) + \varepsilon_{jk(i)}, \text{ em que}$$

y_{ijk} = valor observado no k -ésimo animal do j -ésimo grupo genético da i -ésima dieta;

μ = efeito médio;

D_i = efeito principal da dieta;

$G(D)_{j(i)}$ = efeito do j -ésimo grupo genético dentro da i -ésima dieta;

$\beta(X_{ijk} - \bar{X})$ = coeficiente de regressão linear (β) entre y_{ijk} e X_{ijk} (peso do k -ésimo animal da j -ésima dieta do i -ésimo grupo genético); \bar{X} é a média dos valores de X ;

$\varepsilon_{jk(i)}$ = efeito aleatório, suposto ter distribuição normal, média zero e variância σ^2 , $\varepsilon_i \sim N(0, \sigma^2)$.

Para propósitos de estimação do tamanho amostral, interessa obter a matriz de covariância amostral (S) e sua inversa (S^{-1}), sem a influência de tratamentos específicos; assim, por meio do procedimento GLM eliminaram-se dos dados observados os efeitos devidos a D_i e $G(D)_{j(i)}$ e $\beta(X_{ijk} - \bar{X})$.

Para testar a hipótese da igualdade de dois vetores μ_1 e μ_2 , $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$, tem-se:

$$F = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)' S^{-1} (\bar{y}_1 - \bar{y}_2)$$

A estatística F possui graus de liberdade p e $(n_1 + n_2 - p - 1)$ e parâmetro de não-centralidade $\delta^2 = n_1 n_2 (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) / (n_1 + n_2)$. Para um dado α , rejeita-se H_0 se $F \geq F(p, n_1 + n_2 - p - 1; \delta^2)$.

Uma vez que não se conhecem todos os valores de μ_1, \dots, μ_t e Σ para os quais se deseja aceitar ou rejeitar H_0 na comparação de dois vetores i e j , especificou-se um vetor de diferença mínima significativa (Δ), em valor absoluto, dada por $|\mu_i - \mu_j| \geq \Delta$, para algum $j \neq k$ (Scheffé, 1959), que se deseja detectar a um nível α de probabilidade e poder do teste $(1 - \beta)$. O valor mínimo de δ^2 sujeito à restrição $|\mu_j - \mu_k| = \Delta$, definido por δ_{Δ}^2 , é igual a $n\Delta^2 / \max_{j < k} \{\sigma_j^2 + \sigma_k^2 - 2\sigma_{jk}\}$, em que σ_j^2 e σ_k^2 ($j < k$) são as

variâncias e σ_{jk} , a covariância, associadas às características j e k , respectivamente. Vonesh & Schork (1986) demonstram que essa expressão é apropriada para estimar o tamanho amostral n .

Estimativas de n , para $t \geq 2$ tratamentos, em função de v_1 , v_2 , α , poder do teste $(1 - \beta)$ e de δ^2_{Δ} , foram obtidas iterativamente por meio de um programa SAS que utiliza as funções FPROB e FINV (Hardison et al., 1983; Vonesh & Schork, 1986). O valor de n foi obtido em função de α (0,01 e 0,05), poder do teste (0,95 e 0,90) e δ^2_{Δ} . Para o cálculo da Δ , considerou-se os valores $0,5\sigma$, $1,0\sigma$, $1,5\sigma$, $2,0\sigma$ e $2,5\sigma$, sendo σ vetor de desvio obtido da matriz de covariância (Tabela 1).

RESULTADOS E DISCUSSÃO

Da matriz de variância-covariância dos dados (Tabela 1), calculou-se $\max_{j < k} (s^2_j + s^2_k - 2s_{jk}) = (217,4826 + 50,8182 - 2 \times 15,8558) = 236,6392$. Utilizando-se este valor e ainda $D/s_{MAX} = 1,0$ para $s_{MAX} = \sqrt{217,4826} = 14,7473$, obteve-se o valor de δ^2_{Δ} (função de n , valor a ser estimado), como sendo $\delta^2_{\Delta} = 0,9204n$.

As Tabelas 2 a 4 apresentam as estimativas de n para α variando de 0,01 a 0,05, poder do teste $(1 - \beta)$

de 0,80 a 0,95 e Δ de $1,0\sigma$ a $2,5\sigma$. Uma vez que estas estimativas foram obtidas considerando a variabilidade de cada caráter, sem a influência de tratamentos específicos, e ainda os dois maiores valores da matriz de variância-covariância, inferências sobre os valores de n podem ser feitas para o conjunto de caracteres estudado (Vonesh & Schork, 1986).

Como exemplo, seja o planejamento de um experimento de digestibilidade com bovinos em que se deseja comparar 10 tratamentos, considerando ainda a avaliação dos caracteres citados neste trabalho, que são normalmente utilizados neste tipo de pesquisa. Se pretender detectar diferença significativa entre dois vetores de médias de tratamentos a 5% de probabilidade ($\alpha = 0,05$), considerando que a diferença entre os dois vetores seja no máximo igual a um desvio-padrão ($\Delta = 1,0\sigma$) e poder do teste $F(1-\beta)$ igual a 0,80, há necessidade de 28 indivíduos (Tabela 2). Para esta mesma situação, ao adotar maior rigor na comparação de vetores de médias de tratamentos, ou seja, diferença significativa a 1% de probabilidade ($\alpha = 0,01$) e poder do teste igual a 0,95, seriam necessários 47 indivíduos, conforme observa-se na Tabela 4.

TABELA 1. Estimativas de variâncias (diagonal principal) e covariâncias (abaixo) do consumo por quilograma de peso metabólico (CPM), digestibilidades da energia bruta (DEB), da fibra em detergente ácido (DFDA), da fibra em detergente neutro (DFDN), da matéria seca (DMS), da proteína bruta (DPB), nitrogênio retido (NR) e do NR por kg de peso metabólico (NPM) obtidas de bovinos Canchim, 1/2 Canchim + 1/2 Nelore e Nelore, de 1988 a 1989, São Carlos, SP.

Variáveis	CPM	DMS	DPB	DEB	DFDN	DFDA	NR	NPM
CPM	0,4228							
DMS	0,0182	15,8641						
DPB	0,2113	15,6179	20,4821					
DEB	0,0476	15,7595	15,4882	15,8585				
DFDN	-0,0245	18,1995	17,0313	18,1333	26,8752			
DFDA	-0,0810	24,8268	23,4649	24,6590	31,3216	50,8682		
NR	-0,5108	11,4530	18,0325	12,0697	19,9332	15,8558	217,4826	
NPM	0,1013	1,7213	2,2315	1,7575	3,2193	2,7784	16,0364	3,6975

Para número de tratamentos superior a 10, uma aproximação razoável do número de indivíduos a ser utilizado, independentemente de qual seja a combinação de α , β e Δ , seria acrescentar 1,0 (um) ao

TABELA 2. Tamanho amostral mínimo requerido para detectar diferença mínima significativa (Δ) entre vetores de tratamentos, em função de Δ ($1,0\sigma$ a $2,5\sigma$), erro do tipo I ($\alpha = 0,05$ e $0,04$) e poder do teste ($1 - \beta = 0,80$ a $0,95$).

Número de tratamentos	Diferença mínima significativa (Δ)							
	1,0 σ	1,5 σ	2,0 σ	2,5 σ	1,0 σ	1,5 σ	2,0 σ	2,5 σ
	$\alpha = 0,05; 1 - \beta = 0,80$				$\alpha = 0,05; 1 - \beta = 0,85$			
2	14	9	7	6	16	9	7	6
3	17	10	8	7	18	11	8	7
4	19	12	9	8	20	12	10	8
5	20	13	10	9	22	14	11	10
6	22	14	12	10	24	15	12	11
7	24	16	13	11	26	16	13	12
8	25	17	14	12	27	18	14	13
9	27	18	15	14	29	19	15	14
10	28	19	16	15	30	20	17	15
	$\alpha = 0,05; 1 - \beta = 0,90$				$\alpha = 0,05; 1 - \beta = 0,95$			
2	17	10	7	6	21	11	8	7
3	20	12	9	8	23	13	10	8
4	22	13	10	9	26	15	11	9
5	24	15	11	10	28	16	12	11
6	26	16	13	11	30	18	14	12
7	28	17	14	12	32	19	15	13
8	30	19	15	13	34	20	16	14
9	31	20	16	14	35	22	17	15
10	33	21	17	15	37	23	18	16
	$\alpha = 0,04; 1 - \beta = 0,80$				$\alpha = 0,04; 1 - \beta = 0,85$			
2	15	9	7	6	17	10	7	6
3	17	11	8	7	19	11	9	7
4	20	12	10	8	21	13	10	9
5	21	14	11	10	23	14	11	10
6	23	15	12	11	25	16	12	11
7	25	16	13	12	27	17	14	12
8	27	17	14	13	28	18	15	13
9	28	19	15	14	30	19	16	14
10	30	20	16	15	32	21	17	15
	$\alpha = 0,04; 1 - \beta = 0,90$				$\alpha = 0,04; 1 - \beta = 0,95$			
2	18	11	8	7	22	12	9	7
3	21	12	9	8	24	14	10	8
4	23	14	11	9	27	15	12	10
5	26	15	12	10	29	17	13	11
6	27	17	13	11	31	18	14	12
7	29	18	14	13	33	20	15	13
8	31	19	15	14	35	21	16	14
9	33	21	17	15	37	22	18	15
10	34	22	18	16	39	24	19	17

valor de n das Tabelas de 2 a 4, para cada diferença mínima significativa entre vetores de tratamento adicional. Por exemplo, para detectar médias de tratamentos para $\Delta = 1,0\sigma$, $\alpha = 0,02$ e

TABELA 3. Tamanho amostral mínimo requerido para detectar diferença mínima significativa (Δ) entre vetores de tratamentos, em função de Δ ($1,0\sigma$ a $2,5\sigma$), erro do tipo I ($\alpha = 0,03$ e $0,02$) e poder do teste ($1 - \beta = 0,80$ a $0,95$).

Número de tratamentos	Diferença mínima significativa (Δ)							
	$1,0\sigma$	$1,5\sigma$	$2,0\sigma$	$2,5\sigma$	$1,0\sigma$	$1,5\sigma$	$2,0\sigma$	$2,5\sigma$
	$\alpha = 0,03; 1 - \beta = 0,80$				$\alpha = 0,03; 1 - \beta = 0,85$			
2	16	10	7	6	18	10	8	7
3	19	11	9	8	20	12	9	8
4	21	13	10	9	23	14	11	9
5	23	14	11	10	25	15	12	10
6	25	16	13	11	27	16	13	11
7	26	17	14	12	28	18	14	13
8	28	18	15	13	30	19	15	14
9	30	19	16	14	32	20	16	15
10	31	21	17	15	33	22	18	16
	$\alpha = 0,03; 1 - \beta = 0,90$				$\alpha = 0,03; 1 - \beta = 0,95$			
2	20	11	8	7	23	13	9	7
3	22	13	10	8	26	15	11	9
4	25	15	11	10	28	16	12	10
5	27	16	12	11	31	18	13	11
6	29	18	14	12	33	19	15	13
7	31	19	15	13	35	21	16	14
8	33	20	16	14	37	22	17	15
9	34	22	17	15	39	23	18	16
10	36	23	18	16	40	25	19	17
	$\alpha = 0,02; 1 - \beta = 0,80$				$\alpha = 0,02; 1 - \beta = 0,8$			
2	18	10	8	7	19	11	8	7
3	20	12	9	8	22	13	10	8
4	23	14	11	9	24	15	11	10
5	25	15	12	10	27	16	13	11
6	27	17	13	12	28	18	14	12
7	28	18	14	13	30	19	15	13
8	30	19	16	14	32	20	16	14
9	32	21	17	15	34	22	17	15
10	33	22	18	16	35	23	18	16
	$\alpha = 0,02; 1 - \beta = 0,90$				$\alpha = 0,02; 1 - \beta = 0,95$			
2	21	12	9	7	25	14	10	8
3	24	14	10	9	28	16	11	9
4	27	16	12	10	30	17	13	11
5	29	17	13	11	33	19	14	12
6	31	19	14	12	35	21	16	13
7	33	20	16	14	37	22	17	14
8	35	21	17	15	49	23	18	15
9	37	23	18	16	41	25	19	17
10	38	24	19	17	43	26	20	18

TABELA 4. Tamanho amostral mínimo requerido para detectar diferença mínima significativa (Δ) entre vetores de tratamentos, em função de Δ ($1,0\sigma$ a $2,5\sigma$), erro do tipo I ($\alpha = 0,01$) e poder do teste ($1 - \beta = 0,80$ a $0,95$).

Número de tratamentos	Diferença mínima significativa (Δ)							
	1,0 σ	1,5 σ	2,0 σ	2,5 σ	1,0 σ	1,5 σ	2,0 σ	2,5 σ
	$\alpha = 0,01; 1 - \beta = 0,80$				$\alpha = 0,01; 1 - \beta = 0,85$			
2	20	12	9	8	22	13	9	8
3	23	14	11	9	25	15	11	9
4	26	15	12	10	27	16	12	11
5	28	17	13	11	30	18	14	12
6	30	18	14	13	32	19	15	13
7	32	20	16	14	34	21	16	14
8	33	21	17	15	36	22	17	15
9	35	22	18	16	37	23	19	16
10	37	24	19	17	39	25	20	17
	$\alpha = 0,01; 1 - \beta = 0,90$				$\alpha = 0,01; 1 - \beta = 0,95$			
2	24	14	10	8	28	15	11	9
3	27	16	12	10	31	17	13	10
4	30	17	13	11	34	19	14	12
5	32	19	14	12	36	21	16	13
6	34	21	16	13	39	23	17	14
7	37	22	17	15	41	24	18	15
8	38	23	18	16	43	25	19	17
9	40	25	19	17	45	27	21	18
10	42	26	21	18	47	28	22	19

poder do teste igual a 0,95 (Tabela 3), considerando-se 12 tratamentos, deve-se utilizar 45 indivíduos.

A Tabela 5 apresenta as estimativas dos vetores da média dos caracteres e da diferença mínima significativa de $1,0\sigma$ a $2,5\sigma$, obtidas dos dados observados. Com base nesta situação, quando se deseja detectar diferença significativa entre duas médias de um experimento com oito tratamentos, mantendo-se o poder do teste a 0,90 e níveis de probabilidade (α) variando de 0,05 a 0,01, verifica-se a necessidade de 13 a 16 indivíduos, caso a diferença entre os dois vetores seja bastante evidente ($\Delta = 2,5\sigma$, por exemplo), ou 30 a 38 indivíduos, quando a diferença for pequena ($\Delta = 1,0\sigma$). Assim, verifica-se que o número de indivíduos a ser considerado é bastante influenciado pela magnitude da Δ entre os vetores a serem comparados. Vale salientar que as estimativas obtidas nesse trabalho são válidas para experimentos de digestibilidade envolvendo os oito caracteres

TABELA 5. Médias e valores de desvio-padrão (σ) associados ao consumo por quilograma de peso metabólico (CPM), digestibilidades da matéria seca (DMS), da proteína bruta (DPB), da energia bruta (DEB), da fibra em detergente neutro (DFDN), da fibra em detergente ácido (DFDA), do nitrogênio retido (NR) e do NR por kg de peso metabólico (NPM), obtidas de bovinos Canchim, 1/2 Canchim + 1/2 Nelore e Nelore, de 1988 a 1989, São Carlos, SP.

Característica	Média	1,0 σ	1,5 σ	2,0 σ	2,5 σ
CPM	9,08	0,65	0,98	1,30	1,62
DMS	76,98	3,98	5,97	7,96	9,95
DPB	76,74	4,52	6,78	9,04	11,30
DEB	73,45	3,98	5,97	7,96	9,95
DFDN	59,05	5,18	7,77	10,36	12,95
DFDA	40,08	7,13	10,69	14,26	17,82
NR	27,97	14,75	22,12	29,50	36,87
NPM	6,42	1,92	2,88	3,84	4,80

aqui estudados, admitindo-se ainda a pressuposição de homogeneidade da matriz de variância-covariância entre os tratamentos.

CONCLUSÕES

1. O número de indivíduos necessários para detectar significância entre vetores de médias de tratamentos de dados de digestibilidade com bovinos é influenciado pelo grau adotado da diferença mínima significativa do erro do tipo I e do poder do teste.

2. O número de indivíduos é mais influenciado por alteração no grau da diferença mínima significativa do que no grau do erro tipo I e do poder do teste.

REFERÊNCIAS

- HARDISON, C.D.; QUADE, D.; LANGSTON, R.E. Nine functions for probability distributions. In: SAS INSTITUTE. **SUGI supplemental library user's guide**. Cary, NC, 1983. p.229-236.
- SAS INSTITUTE. **SAS/STAT user's guide**: statistics, versão 6. 4.ed. Cary, NC, 1993. v.2, 1686p.
- SCHEFFÉ, A. **The analysis of variance**. Toronto: John Wiley, 1959. 477p.
- VONESH, E.F.; SCHORK, M.A. Sample size analysis of repeated measurement. **Biometrics**, Baltimore, v.42, n.3, p.601-610, Sept. 1986.