Rev. Brasil. Genet. 15, 3, 615-622 (1992)
(Brazil. J. Genetics)

APPLICATION OF A PROBABILISTIC MODEL TO THE SELECTION OF A BEEF CATTLE SUBPOPULATION

Célia M. Torres Cordeiro[1], José B. de Freitas Trovo[1], Rosângela S. Afanasieff[2], Alexander G. Razook[3] and Luiz M. Bonilha-Neto[3]

ABSTRACT

The objective of the present study was to determine the truncation point for weaning weight for selection of a specific beef cattle subpopulation consisting of 378-day old animals whose weight was greater than or equal to a given value. The probabilities of selecting animals whose weaning weights were greater than a given value for several subpopulations of interest were calculated from the bivariate normal distribution of weaning weight and weight at 378 days. Data from five years of feeding tests were used to check the normality of the distribution and to estimate the bivariate normal correlation coefficient. The results obtained using this theoretical model showed that recovery of more than 90% of the subpopulations of interest, consisting of 378-day old animals weighing more than 1.28, 1.64 and 2.05 standard deviation units (SDU) above the mean can be obtained by accepting truncation points ≥ 0.5, 0.8 and 1.1 SDU above the mean weaning weight, respectively. Estimates were checked for accuracy by simulation. For herd sizes larger than 250 animals, the simulation results showed good approximation to theoretical values at truncation points ≤ 0.8 SDU above mean weaning weight.

INTRODUCTION

The beef cattle breeding program coordinated and executed by the "Instituto de Zootecnia do Estado de São Paulo" (Institute of Animal Technology of the State of São

[1] Centro Nacional de Recursos Genéticos e Biotecnologia, CENARGEN-EMBRAPA, Caixa Postal 102372, 70770 Brasília, DF, Brasil.

[2] Departamento de Estudos e Pesquisa, EMBRAPA, Caixa Postal 04315, 70770 Brasília, DF, Brasil.

[3] Estação Experimental de Zootecnia de Sertãozinho, IZ, Caixa Postal 63, 14180 Sertãozinho, SP. Send correspondence to A.G.R.

Paulo) at the Experimental Station of Sertãozinho includes a phase of animal selection within each herd based on weaning weight (adjusted for 210 days of age on the basis of the weight gain of each animal). Selected animals are then tested for weight gain over a period of 168 days under identical feeding and management conditions (Razook *et al.*, 1984a,b). At the end of the test, animals are selected on the basis of higher final weight (adjusted for 378 days of age on the basis of their weight gain).

In view of the high cost of tests of this kind, it would be useful establish a criterion for selection at weaning. This would permit the breeder to determine truncation points, i.e., a minimum standard for weaning weight in order to insure that a given percentage of animals that will be selected as the elite of their herds at 378 days of age will be sent to the testing center.

Young and Weller (1961) established selection levels independent for each of two variables to obtain an expected value for one or both in the remaining population. Smith and Quass (1982) used a more general approach to the problem, which is the maximization of the expected value for a variable in the remaining population. In both cases, the authors considered the restriction that limits the percentage of the population to be retained, and adopted the multivariate normal probabilistic model. In the present study, we use the same model and we also made an attempt to establish truncation levels. However, the major objective is fulfilled by the use of the probability distribution of one variable in a subpopulation of interest and not of its expected values.

MATERIALS AND METHODS

Let the random variables Y_1 and Y_2 be the weaning weight and weight at the end of the test, adjusted for 210 and 378 days, respectively, from the weight gain curve for each animal. By assuming that Y_1 and Y_2 have a bivariate normal distribution with means μ_1 and μ_2, variances $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ and correlation coefficient ρ, it is possible to work with the standardized variables without a loss in generality:

$$X_1 = \frac{Y_1 - \mu_1}{\sigma_1} \quad \text{and} \quad X_2 = \frac{Y_2 - \mu_2}{\sigma_2},$$

and propose a probabilistic approach to the selection criterion regardless of the means and variances of these variables in the herd. On this basis, the probability (P) that satisfies the objective proposed is that associated with the event "weaning weight x_1 higher than x_1 in subpopulations whose final weight x_2 is higher than a given x_2", formally presented in the expression below:

$$P = \frac{P[(X_1 \geq X_1) \cap (X_2 \geq X_2)]}{P(X_2 \geq X_2)} \quad (1)$$

To obtain these probabilities, three subpopulations of interest (SPI) were taken, consisting of animals considered to be elite, i.e., having a weight at 378 days, starting from a given y_2 value, such that:

for SPI 10%, $P(Y_2 \geq y_2) = 0.10$ when $X_2 \geq 1.28155$
for SPI 5%, $P(Y_2 \geq y_2) = 0.05$ when $X_2 \geq 1.64485$
for SPI 2%, $P(Y_2 \geq y_2) = 0.02$ when $X_2 \geq 2.05375$.

To calculate the probability defined in (1) we used the MDBNOR routine and the PROBNOR function available in the IMSL (1980) and SAS (1985) manuals, respectively, with X_1 ranging from -0.2 to 1.3 for each SPI.

The probabilities calculated above, since they were obtained according to a probabilistic model for infinite populations, are applicable in an exact manner for large herds, for which this model is pertinent. Thus, it is necessary to establish herd size, N, starting from which the proportion $\hat{p}$ of animals with weaning weights greater than a given value in the SPI can be considered an accurate estimate of P in (1). For this purpose, we chose a few situations characterized by distinct combinations of N, truncation point in weaning weight (three of them were selected) and SPI. A total of 1000 samples of size N of a standardized bivariate normal distribution with a given correlation coefficient were simulated for each situation. The values of $\hat{p}$ and of its mean value, $\bar{p}$, were calculated for each simulated sample and from all samples in each situation, respectively. The behavior of $\hat{p}$ was evaluated by its mean square error in relation to P, MSE ($\hat{p}$), and by its coefficient of variation, CV($\hat{p}$), estimated as follows for each situation:

$$MSE(\hat{p}) = (\bar{p} - P)^2 + var(\hat{p}),$$
$$CV(\hat{p}) = 100\,(\sqrt{MSE(\hat{p})})/P.$$

A CV(p) not exceeding 10% was adopted as the criterion, indicating that in a given situation $\hat{p}$ was an accurate estimate of P. This criterion was used to insure a probability of at least 0.75 for the occurrence of deviations of the $|\hat{p} - P|$ type, not exceeding 0.2 of P, according to the inequality of Chebyshev.

Data for weaning weight and weight gain at the end of the weight gain test for all male animals born from 1981 to 1986 as part of the Nelore cattle herd of the Experimental Station of Sertãozinho, Institute of Animal Tecnology, were used to test the hypothesis of a bivariate normal distribution and to obtain information about the nature of the correlation between the two weights. The test used was that proposed by Malkovich and Afifi (1973), which is represented by the following equation:

$$V_j = (Y_j - \overline{Y})' S^{-1} (Y_j - Y), \; j = 1, \dots, n,$$

where n is sample size, Y_j is the *j*th pair vector observed (Y1, Y2), $\overline{Y}$ is the observed mean vector, and S the matrix of sample variances and covariances.

The V_j statistic was calculated for each observation and its sample distribution function $S_n(V)$ was compared with the distribution function $\chi^2_{(2)}$, $F_n(V)$. This comparison utilizes the Kolmogorov-Smirnov statistic:

$$K = \max. \mid S_n(V) - F_n(V) \mid$$

RESULTS AND DISCUSSION

Table I shows the estimated correlation coefficients and the results of the normality test. The hypothesis of normality was rejected for 1982. However, the assumption required by the methodology proposed was considered to be realistic, taking into account that the data for this particular year may reflect disturbances caused by the removal of part of the animals belonging to this population which were used for other purposes.

Table I - Estimates of bivariate normal parameters and statistics values for the test of this distribution.

Year	ρ	Herd size	Mean weight (kg) at weaning (Y_1)	Mean weight (kg) at the end of the test (Y_2)	K (maximum absolute deviation)[a]
1981	0.80	110	175.2	308.4	0.085 NS/[a]
1982	0.60	82	170.2	391.7	0.224 *
1983	0.76	83	174.6	290.3	0.087 NS
1984	0.80	79	173.2	290.7	0.083 NS
1985	0.70	67	165.9	295.1	0.050 NS
1986	0.79	60	179.3	306.5	0.077 NS

[a]Results evaluated at the 5% (bilateral) level of significance.

On the basis of the results presented in Table I, a 0.80 correlation coefficient was established as the parameter of the bivariate normal. Thus, all the results obtained in the present sutdy are conditioned to this parametric value.

Infinite populations

Table II presents the probabilities (P) associated with the event generically defined in (1) for the subpopulations of interest, involving animals with weights at the end of the test corresponding to $X_2 \geq 1.28$ (10% SPI), $X_2 \geq 1.64$ (5% SPI) and $X_2 \geq 2.05$ (2% SPI) standard deviation units (SDU) above the mean. It can be seen that the probability of X_1 being higher than or equal to a given value increases with the conditioning of SPI to the occurrence of a more rare event, i.e., the objective of selection is to recover animals with greater weights at the end of the test. Thus, with 10% SPI, 55% of the population consists of animals with weaning weights exceeding 1.3 SDU above the mean, whereas with 5% SPI and 2% SPI, these animals make up 69% and 83% of the respective populations. The results in table II indicate that animals with weaning weights below the herd mean, $X_1 < 0$, represent a very small proportion (less than 2%) of the SPI studied, so that it would be of no advantage to submit these animals to weight gain tests. This practice would barely increase the probability of recovering the animals considered to be the herd elite in the SPI studied.

Table II - Probability (p) of $X_1 \geq x_1$ conditioned to the occurrence of the event $X_2 \geq x_2$.

x_1 (SDU above mean weaning weight)	Expected selection %	Probability of $X_1 \geq x_1$ in the subpopulations of interest (SPI)		
		SP 10% $X_2 \geq 1.28$	SPI 5% $X_2 \geq 1.64$	SPI 2% $X_2 \geq 2.05$
1.3	9.68	0.551	0.695	0.834
1.2	11.50	0.610	0.747	0.870
1.1	13.56	0.666	0.794	0.901
1.0	15.86	0.719	0.835	0.926
0.9	18.40	0.768	0.871	0.945
0.8	21.18	0.812	0.901	0.961
0.7	24.19	0.850	0.925	0.972
0.6	27.42	0.883	0.945	0.981
0.5	30.85	0.911	0.960	0.987
0.4	34.45	0.933	0.972	0.991
0.3	38.20	0.951	0.981	0.994
0.2	42.07	0.965	0.987	0.996
0.1	46.01	0.977	0.991	0.998
0.0	50.00	0.983	0.994	0.999
- 0.1	53.98	0.988	0.996	0.999
- 0.2	57.92	0.992	0.997	0.999

Finite populations

The results of the simulation (Table III) indicate that for truncation points at weaning equal to or lower than the herd mean, the sampling proportions represent an acceptable approximation of P (Table II) for herd sizes starting from 80 animals in all SPI studied. The reuslts in Table II are also acceptable for truncation points at weaning equal to or lower than 0.8 SDU above the mean for all SPI studied only if herd size is 250 animals or more. These results indicate that when selection at weaning rates of about 20% are adopted according to the criterion of greatest weaning weight in order to select animals that at the end of the test will have a weight of more than 1.28 SDU above the mean, for example, the expected proportion of those animals that are being effectively selected is approximately 0.81. For a truncation point of 1.3 SDU above the mean, the results of simulation indicate that the probabilities, P, in Table II cannot be approximated by the proportions observed in the sample for any of the SPI studied and for any of the herd sizes considered. This means that the practice of establishing this minimum weight standard for selection at weaning is risky because the proportion of elite animals effectively selected is submitted to elevated variability around P or its expected value, as can be seen in Table III.

Table III - Statistical data obtained by the simulation of 1000 samples in each condition of herd size, subpopulation of interest (SPI) and truncation point at weaning.

Selection at weaning of animals with weights greater than[a]		SPI 10% - Animals with a final weight greater than 1.28[a]				
		Herd size at weaning				
		80	200	250	400	500
0.0	$\bar{p}$	0.9864	0.9846	0.9838	0.9848	0.9839
	MSE ($\hat{p}$)	0.0018	0.0008	0.0007	0.0004	0.0003
	CV ($\hat{p}$)	4.3	2.9	2.6	2.0	1.7
0.8	$\bar{p}$	0.8083	0.8125	0.8140	0.8155	0.8151
	MSE ($\hat{p}$)	0.0221	0.0078	0.0063	0.0039	0.0032
	CV ($\hat{p}$)	18.3	10.9	9.7	7.7	6.9

Continued

Table III - continued

Selection at weaning of animals with weights greater than[a]		SPI 10% - Animals with a final weight greater than 1.28[a]				
		Herd size at weaning				
		80	200	250	400	500
1.3	$\bar{p}$	0.5515	0.5466	0.5523	0.5553	0.5538
	MSE ($\hat{p}$)	0.0340	0.0127	0.0106	0.0066	0.0048
	CV ($\hat{p}$)	33.4	20.4	18.6	14.7	12.6
		SPI 5% - Animals with a final weight greater than 1.64[a]				
0.0	$\bar{p}$	**0.9959**	**0.9950**	**0.9956**	**0.9958**	**0.9958**
	MSE ($\hat{p}$)	**0.0013**	**0.0005**	**0.0004**	**0.0002**	**0.0002**
	CV ($\hat{p}$)	**3.6**	**2.3**	**1.9**	**1.4**	**1.4**
0.8	$\bar{p}$	0.8991	0.9022	**0.9012**	**0.9023**	**0.9047**
	MSE ($\hat{p}$)	0.0316	0.0096	**0.0080**	**0.0048**	**0.0037**
	CV ($\hat{p}$)	19.7	10.8	**9.8**	**7.7**	**6.7**
1.3	$\bar{p}$	0.6914	0.6967	0.6923	0.6951	0.6971
	MSE ($\hat{p}$)	0.0743	0.0233	0.0191	0.0102	0.0089
	CV ($\hat{p}$)	39.2	21.9	19.9	14.5	13.6
		SPI 2% - Animals with a final weight greater than 2.05[a]				
0.0	$\bar{p}$	**0.0\9987**	**0.9986**	**0.9992**	**0.9990**	**0.9993**
	MSE ($\hat{p}$)	**0.0013**	**0.0004**	**0.0001**	**0.0001**	**0.0001**
	CV ($\hat{p}$)	**3.6**	**1.9**	**1.2**	**1.1**	**1.0**
0.8	$\bar{p}$	0.9592	0.9592	**0.9650**	**0.9598**	**0.9614**
	MSE ($\hat{p}$)	0.0286	0.0136	**0.0082**	**0.0063**	**0.0046**
	CV ($\hat{p}$)	17.6	12.1	**9.2**	**8.3**	**7.0**
1.3	$\bar{p}$	0.8219	0.8347	0.8220	0.8345	0.8321
	MSE ($\hat{p}$)	0.1019	0.0475	0.0394	0.0191	0.0179
	CV ($\hat{p}$)	38.2	26.1	23.8	16.5	16.0

[a] Weights reported as SDU above the mean.

[b] The results in bold type accept the approximation of P by p.

ACKNOWLEDGMENT

Publication supported by FAPESP.

RESUMO

Este trabalho propõe a determinação de pontos de truncamento no peso à desmama para a seleção de animais de uma subpopulação de interesse caracterizada como aquela constituída de animais com peso igual ou superior a um dado valor aos 378 dias. Assumindo-se a distribuição Normal Bivariada para o peso à desmama e o peso aos 378 dias, obtém-se a probabilidade de se selecionar indivíduos com peso à desmama superior a um dado valor (ponto de truncamento) em diferentes subpopulações de interesse. Dados de 5 anos de prova de ganho de peso (Estação Experimental de Sertãozinho) foram utilizados para testar a validade da suposição de normalidade e para a obtenção de estimativas de coeficiente de correlação da Normal Bivariada. Os resultados obtidos com este modelo teórico indicam que a recuperação de mais de 90% das subpopulações de interesse constituídas pelos animais com peso aos 378 dias superior a 1,28, 1,64 e 2,05 u.d.p. acima da média pode ser obtida, adotando-se pontos de truncamento igual ou inferior a 0,5, 0,8 e 1,1 u.d.p. acima da média do peso à desmama, respectivamente. Verificou-se, através de simulação, o tamanho de rebanho a partir do qual a proporção de indivíduos com peso à desmama superior a dados valores (pontos de truncamento) nas subpopulações de interesse pode ser considerada uma estimativa acurada das probabilidades obtidas pelo modelo Normal Bivariado. Os resultados da simulação indicam boa concordância com os teóricos quando o tamanho do rebanho é superior a 250 animais e em pontos de truncamento à desmama igual ou inferior a 0,8 u.d.p. acima da média.

REFERENCES

IMSL, Inc. (1980). *The International Mathematical & Statistical Libraries.* Vol. 2, Edition 8, Houston, Texas: IMSL, Inc.

Malkovich, J.F. and Afifi, A.A. (1973). On tests for Multivariate Normality. *JASA 68*: 176-179.

Prova Ganho de Peso (1981/1986). Resultado Geral, Sertãozinho, SP, Instituto de Zootecnia.

Razook, A.G., Trovo, J.B.F., Pacola, L., Nascimento, J., Packer, A.U., Reichert, R.H. and Procknor, M. (1984). Novas provas de ganho de peso de Sertãozinho. a. Influências do meio sobre características observadas em bovinos Nelore e Guzerá. *B. Industr. Anim.* (São Paulo) *41*: 9-23.

Razook, A.G., Trovo, J.B.F., Pacola, L., Nascimento, J., Packer, A.U., Reichert, R.H. and Procknor, M. (1984). Novas provas de ganho de peso de Sertãozinho. b. Influências do meio e herença na seleção de um rebanho de bovinos Nelore. *B. Industr. Anim.* (São Paulo) *41*: 25-33.

SAS Institute Inc. (1985). *SAS User's Guide: Statistics*, Version 5 Edition. Cary, NC: SAS Institute Inc.

Smith, S.P. and Quass, R.L. (1982). Optimal truncation points for independent culling level selection involving two traits. *Biometrics 38*: 975-980.

Young, S.S.Y. and Weiler, J. (1961). Selection for two correlated traits by independent culling levels. *J. Genet. 57*: 329-338.

(Received July 20, 1990)