SAMPLE SIZE FOR BODY WEIGHT IN BEEF CATTLE IN REPEATED MEASUREMENTS EXPERIMENTS

C.C.P. Paz[1,2], A.R. Freitas[3], I.U. Packer[2], D.D. Tambasco[3], L.C.A. Regitano[3] and M.M. Alencar[3]

[1]Instituto de Zootecnia, Centro de Genética e Reprodução, CP 60, 13460-000, Nova Odessa-SP-Brasil
[2] Esalq-USP, Depto Produção Animal Av. Pádua Dias,11 CP 9, 13418-900 Piracicaba-SP-Brasil
[3] Embrapa Pecuária Sudeste, CP 339, 13560-970 São Carlos-SP-Brasil.

INTRODUCTION

Animal production has experimented considerable genetic improvement in some performance traits. Improvement of the animal breeding programs to obtain further may require molecular marker-assisted selection, which requires a identification of candidate genes or anonymous genetic markers associated with the traits of interest. The use of candidate genes has been proposed to direct the search for QTL (Quantitative Trait Loci). The polymorphism for growth hormone (GH) gene has been associated with growth traits (Rocha *et al*., 1992; Unanian *et al*., 2000) and carcass composition and meat quality (Taylor *et al*., 1998). Polymorphisms of kappa-casein (K-Cas) and beta-lactoglobulin (B-Lac) have been associated with growth traits (Moody *et al*., 1996). The influence of candidate genes on performance traits of cattle has been analyzed considering each trait individually, which is a difficult way to detect significant effects. An alternative method is the use of repeated measurements (RM) that has increased substantially in beef cattle breeding studies in recent years. Jensen (1982) and Vonesh (1983) showed greater efficiency in the use of RM designs. Results given by Vonesh and Schork (1986) showed greater efficiency of RM analyses regarding to sample size. An important contribution of sample size in RM was given by Freitas *et al.* (1999) for scrotal circumference of Nellore cattle. Proper planning reduces the risks of conducting a study that will not produce useful results and determines the most sensitive design for the available resources. In studies of candidate genes or molecular markers associated with QTL, the cost of laboratory analyses is very important and need to be considered in the planning of the experiment. The purpose of this study was to estimate sample size required for an experiment which associates candidate genes (GH, K-Cas and B-Lac) with body weight in crossbred beef cattle, assuming the body weights of the animals as repeated measurements.

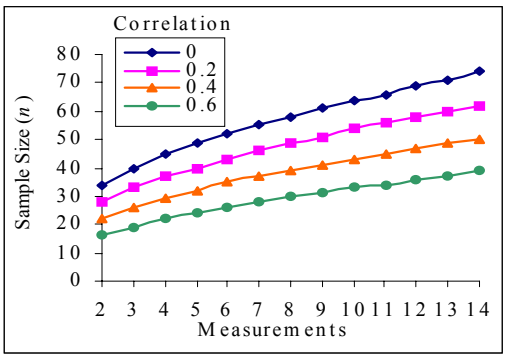
MATERIAL AND METHODS

This study used data collected on 213 animals (75 ½Canchim+½Nellore, 74 ½Aberdeen Angus+½Nellore and 64 ½Simental+½Nellore) born in 1998 and 1999 in Southeast Brazil. The data were 14 measurements of weight considered as RM, collected at birth, weaning (7 months of age) and monthly from 8 to 19 months of age. The model used to determine sample size in repeated measurements analysis was : $Y_{ijk} = \mu + \varepsilon_{ijk}$ ($i=1, ..., n$), $\varepsilon_{ijk} \sim$ IID N_p ($\mathbf{0}$, Σ), where $Y'_i = (Y_{i1}, ..., Y_{ip})$ is the response vector of the i^{th} subject across p repeated measurements, $\mu' = (\mu_1, ... \mu_p)$ is the mean response vector, ε_i is the experimental error, and N_p ($\mathbf{0}$, Σ) is a p-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix Σ. Considering the standard model for repeated measure (Little *et al*., 1998), the μ effect in this

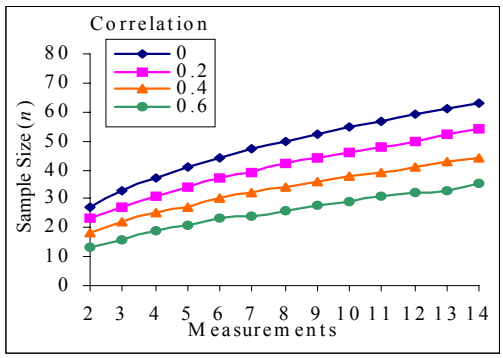
case also included fixed effects such as contemporaneous and genotype group (τ_i), random effect of animal within τ (d_{ij}), k times (t_k) and interaction of τ_i with t_k $(\tau t)_{ik}$ effects. The test to reject or accept the null hypothesis of equal measurements effects $H_0 = \mu_1 = ... = \mu_p$, is based on the statistic $T^2 = n\overline{Y}'C(C'\Sigma C)^{-1}C'\overline{Y}$, where $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$, $S = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \overline{Y})(Y_i - \overline{Y})'$ is the sample covariance matrix, positive defined, and C' is any (p-1) x p orthogonal contrast matrix. The T^2 statistic is distributed according to the Hotelling T^2 with (p-1) and (n–1) degrees of freedom (df) and noncentrality parameter $\delta^2 = n\mu'C(C'\Sigma C)^{-1}C'\mu$. Under true H_0, $F=(n-p+1)[(n-1)(p-1)]^{-1}T^2$ was obtained, which has distribution F with (p-1) and (n-p+1) df and noncentrality parameter δ^2; for a particular α, then it rejects H_0 if $F > F(p-1, n-p+1; \delta^2)$. The sample size ($n$) is determined by assuming several values of μ and Σ, in which H_0 is rejected. It was specified for any pair of RM a minimum difference (Δ), subject to the restriction $|\mu_j - \mu_k| = \Delta$ for any $j \neq k$ (Scheffé, 1959), whose significance should be detected, considering a level of probability α and power of test (1-β). The minimum value of δ^2 subject to the restriction $|\mu_j - \mu_k| = \Delta$, defined by δ^2_{Δ} is equal to $n\Delta^2/\max_{j<k}\{\sigma^2_j + \sigma^2_k - 2\sigma_{jk}\}$ where σ^2_j and σ^2_k (j<k) are the variances and σ_{jk} is the covariance associated to traits j and k, respectively. Considering Σ, any variance-covariance matrix, defined positive satisfying $\rho_{jk} \geq 0$, for all j<k, it can be demonstrated that $n\Delta^2/[(2\sigma^2_{max}(1-\rho_{min})] \leq \delta^2_{\Delta}$ is appropriated for estimating sample size n (Vonesh and Schork, 1986). In this expression, ρ_{min} is the lower correlation coefficient between repeated measures, $\sigma^2_{max} = \text{maximum}(\sigma^2_j)$ and Δ is measured in units of σ_{max}. Using this expression, the n estimated for $p \geq 2$ RM, in functions of p-1 and n-p+1, α and power of test (1-β), were obtained by SAS program that considered an integral and a noncentral F-distribution (Hardison *et al.*, 1983; Vonesh and Schork, 1986). The main point of this study was based on the fact that repeated measures (p=14) of body weight from birth to approximately 19 months of age were adequately described by a sigmoid growth curve, so, as n increases, the confidence interval becomes lower and the curve approximates the population growth curve better.

RESULTS AND DISCUSSION

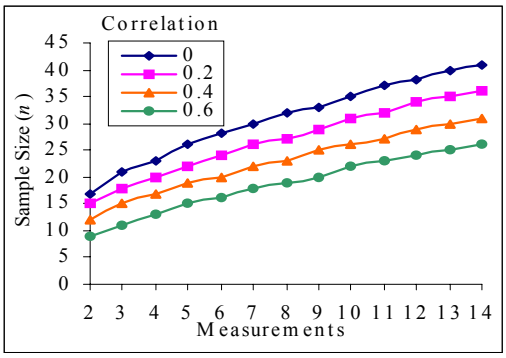
The maximum standard deviation (σ_{max}) obtained from sample variance-covariance matrix, positive defined, was 85.4235 kg and ρ_{min}=0.05. So, the n estimates showed in Figure 1, were obtained by evaluating the integral of a central and noncentral F-distribution in function of expression $0.0000685n\Delta^2/(1-\rho_{min})$; α=0.01, power of test (1-β)=0.80 and 0.90; minimum correlation (ρ_{min})=0, 0.2, 0.4 and 0.6, and detectable difference (Δ)=1.0σ, 1.5σ and 2.0σ. For including ρ_{min}, this expression takes on account the fact that the correlation between repeated measures decreases as the repeated measures become far apart; for considering σ^2_{max}, it takes on account a common fact in growth studies, that is, the variance is linearly proportional to the increment in the response function. These properties, and the fact that the sample variance-covariance matrix is positive defined, assure reliability of the sample size estimate (Brownie *et al.*, 1990; Cullis and McGilchrist, 1990). The minimum number of individuals necessary for detecting significant difference between repeated measure, increases in the following order: (Δ=2.0σ; Power=0.80); (Δ=2.0σ; Power=0.90); (Δ=1.5σ; Power=0.80); (Δ=1.5σ; Power=0.90); (Δ=1.0σ; Power=0.80) and (Δ=1.0σ; Power=0.90).



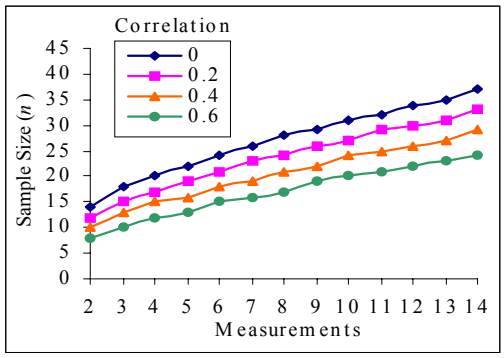
Δ= 1.0σ and Power=0.90



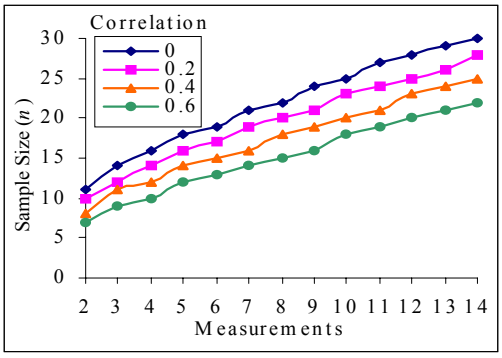
Δ= 1.0σ and Power=0.80



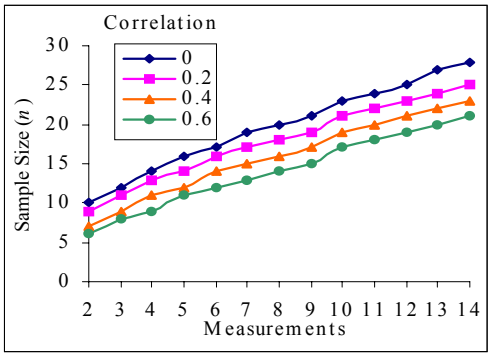
Δ= 1.5σ and Power=0.90



Δ= 1.5σ and Power=0.80



Δ= 2.0σ and Power=0.90



Δ= 2.0σ and Power=0.80

Figure 1. Estimates of sample size (*n*) in repeated measurements of body weight (BW) in beef cattle for α=0.01; power of test (1-β) =0.80 and 0.90; minimum correlation (ρ_{min})=0, 0.2, 0.4 and 0.6 and detectable difference (Δ) = 1.0σ, 1.5σ and 2.0σ

As an example, to detect a significant difference between any two of 14 measurements, considering a minimum difference of 1.0σ, power of test=0.90, and minimum correlation equal

to 0.4, it is necessary a minimum of 50 individuals. At the same conditions, a minimum of 31 individuals are needed when Δ changes from 1.0σ to 1.5σ. Vonesh and Schork (1986) studied the sample size from three to six (p=3(1)6) measurements, seven values of Δ (1.0σ to 3.0σ), power of test of 0.80 and 0.90, and minimum correlation from 0.1 to 0.9. They observed greater reduction in the estimates of n when Δ changed from 1.0σ to 1.25σ. The determination of sample sizes as implemented in this study, plays an important role in the planning of the number of individuals requested in an experiment. Suppose a similar study is planned to evaluate the influence of candidate genes on body weight in cattle, from birth to two years of age. In this case, it is reasonable to estimate the sample size n by $0.0000685n\Delta^2/(1\text{-}0.05)$, where 0.05 is the estimated minimum correlation. Adequate planning reduces the risks of conducting a study that will not produce useful results, and provide a desired power for an effect of scientific interest.

CONCLUSION

The number of animals necessary to detect significance between repeated monthly evaluations of body weight of cattle from birth to approximately 19 months of age is influenced by a minimum difference significative (Δ), correlation among the repeated measures, type I error (α) and power of test ($1\text{-}\beta$). For a particular Δ value, it is necessary a bigger sample size (n) to prove significant difference between repeated measures response, when α moves from 0.05 to 0.01 and the power goes from 0.80 to 0.90. Independently of the power of the test, ρ_{min} and Δ, significant difference between mean of any two measurements at α=0.01 requests a sample size about 30% greater as for with α=0.05.

ACKNOWLEDGEMENTS

This study was financially supported by FAPESP and CNPq**.**

REFERENCES

Brownie, C., Boods, D.D. and Oliver, J.H. (1990) *Biometrics* **46** : 259-266.

Cullis, B.R. and McGilchrist, C.A. (1990) *Biometrics* **46** : 131-142.

Freitas, A.R., Silva, A.E.D.F. and Unanian, M. (1999) *Rev. Bras. Zootec.***28**: 279-287.

Hardison, C.D., Quade, D. and Langston, R.E. (1983) *In:* SAS Institute Inc. SUGI *Supplemental Library User's Guide*. Cary: North Caroline, SAS Institute Inc: 229-236.

Jensen, D.R. (1982) *Biometrics* **38** : 813-825.

Little, R.C., Henry, P.R. and Ammerman,C.B. (1998) *J. Anim. Sci*., **76** : 1216-1231.

Moody, D.E., Pomp, D., Newman, S. and MacNeil, M.D. (1996) *J. Anim. Sci.* **74** : 1784-1793.

Rocha, J.L., Baker, J.F., Womack, J.E. Sanders, J.O. and Taylor, J.F. (1992) *J. Anim. Sci.* **70** : 3360-3370.

Scheffé, A. (1959) The Analysis of Variance. ed. John Wiley, 477p.

Taylor, J.F., Coutinho, L.L., Herring, K.L., *et al*. (1998) *Anim. Genet.* **29** : 194-201.

Unanian, M.M., Barreto, C.C., Freitas, A.R., Cordeiro, C.M.T. and Josahkian, L.A. (2000) *Rev. Bras. Zootec.* **29** : 1380-1386.

Vonesh, E.F. (1983) *Communication in Statistics - Theory and Methods* **12** : 289-301.

Vonesh, E.F.and Schork, M.A. (1986) *Biometrics* **42** : 601-610.