SIRE EVALUATION WITH SPARSE AND MESSY DATA

K. A. Rathie

Department of Primary Industries, Newstead, Queensland, 4006

"Messy" is used in its statistical sense as applying to cross-classified data with unequal replication and some missing cells. Some small (60 to 250 df) sets of Queensland dairy data, comprising milk and fat production records from daughters of bulls in progeny test teams for 1977 and 1978, have been analysed by two methods. Firstly, sires have been evaluated by an unmodified (simple) contemporary comparison (CC). Secondly, analyses have been made using Harvey's (1977) LSML76 least-squares and maximum likelihood computer program package. The analyses to be described were least-squares using a fixed model, as interest in sire evaluation is in specific bulls rather than bulls in general.

All data were within-years, within-breeds, and comprised first lactations only - the standard Queensland conditions for bull proving records. Both years were drought-affected, causing losses of daughter records on the predominantly non-irrigated farms. Jersey losses for 1978 were so severe that no proven sire could be declared.

The dependent variables were milk yield and fat yield, which generally gave similar analysis of variance results. Main effects and covariates, due to the small numbers of observations per set and per cell, could not all be simultaneously analysed in the one set of data. Main effects studied were season of calving (months July to September inclusive, October-November, December to February inclusive, and March to June inclusive), districts, and sires. Covariates were age at calving (in months) and herd average milk yield of cows aged over 4 years; the latter was taken as a crude index of level of feeding and management.

District effects were always highly significant, and sires and seasons sometimes significant (at 5% level), with age at calving as the covariate (or with no covariate). The herd average milk yield, as a covariate, itself explained much of the total variation, but was useless in "sharpening-up" sire evaluations and other main effects - all became clearly non-significant. Other finer-tuned management covariates may prove more useful.

Season of calving effects were greatest in Friesians; the other two breeds (AIS and Jerseys) studied had most observations concentrated in season 1 (July to September), particularly the Jerseys. The Friesian records were evenly distributed over all seasons.

Age at calving within seasons, or within sires, was significant on several occasions.

Interactions, apparently real, were not uncommon - e.g., sires X seasons, sires X districts - but could have been artifacts of the small sample sizes. The obvious next step of looking at the set of all Queensland lactations, instead of just those for daughters of the current progeny test sires, could not be followed due to restrictions on data handling while the lactation records remain on the present computer. This situation will alter in July 1980. The bull proving records do differ from other records in that some data (sire, age at calving) are more reliably known for the former.

Ignoring interactions, where present at a significance level of 15% or higher, sometimes altered sire rankings considerably. Thus it is necessary to know if the interactions are real. Least-squares sire estimates for no-interaction models gave rankings virtually identical to those for the CC method.

REFERENCE

HARVEY, W.R. (1977) User's guide for LSML76. Mixed model least-squares and maximum likelihood computer program - Ohio State University: November 1977.

* * * *