Adjusting Dairy Yield Records From 3 Times to 2 Time Milking Using a Random Regression Model

A.S. Leaflet R1895

Neirlson Santos, Graduate Student in Animal Breeding, and A. E. Freeman, Distinguished Professor of Agriculture

Introduction

Most cows are milked twice a day (2X) or three times a day (3X). If all cows within a herd are milked the same number of times and decisions are strictly within herd then milking frequency does not need to be considered for culling decisions. However, for estimating transmitting abilities to use for making genetic decisions records need to be standardized. All yield records, milk, fat, and protein, are adjusted to a 305-day lactation length, to a 2 times (2X) a day milking, and to a mature equivalent basis for estimating transmitting abilities.

Adjusting records from 3X to 2X was done by factors “agreed upon” in 1956. No research had been done to improve 3X to 2X conversions until we published the factors that are used currently. They were developed from data from the Midstates Record Processing Center. The work in progress will use data selected from all over the nation. The percentages of cows milked 3X in the United States are CA-20, MI-15, NY-25, NC-15, and UT-25. This justifies the need to obtain adjustment factors to correct milking 3X to 2X basis.

A preliminary study on a subset of the national data showed that a test-day model, more specifically a random regression model (RRM), would fit best to our data and answer our research goals. Test-day models use DHIA data and analyze the data separately for each test day.

There are many advantages to using a test-day model. It provides a more precise fit to the data than ordinary models, which use records on a lactation basis as has been used in dairy breeding research. Accuracy of evaluation of animals could be increased using a test-day model because each piece of information (test day record) can be used instead of just one (305-day lactation record). Because test-day models account for correlated records from test-day to test-day within cows they use more information. As new information is available, it can be used immediately. Therefore, this could decrease the generation interval. Also, because the assumption to estimate transmitting abilities is that all cows get equal treatment, this is more likely to be true on a test-day basis.

Among some types of test-day models, a random regression model has the advantage of accounting for the variation of the lactation curve that exists among the animals and therefore allows us to consider the difference in persistence of production among animals. To evaluate this difference, regression coefficients on days in milk are estimated for each cow. These regression coefficients must be considered as random variables, because each cow has a different slope to her lactation curve.

Materials and Methods

A model is a mathematical equation that tries to describe the factors that influence milk, fat, and protein production. Even though the model will never be able to explain all the variation in production, we can often explain 60 to 80 % of the variation.

We will use the following random regression model to analyze the data:

$$y = htd + pr + hys + mf + pr \times mf + dp \times mf + sdim \times mf + dim \times mf + ls \times mf + b_1(sdim) + b_2(ldim) + b_3(dp) + a_i + y_1(sdim) + y_2(ldim) + pe + Q_1(sdim) + Q_2(ldim) + e$$

Where:

y = test day yield for milk, fat and protein;

htd = fixed effect of herd-test-date. This effect represents the test date effect in which the test day was taken. This effect was coded within herd;

hys = herd-year-season. Four seasons per year were used the first one starting with cows calving in January;

pr = parity number. This effect was coded as 1 for lactation number 1, and 2 for lactation number 2 up to lactation number 5;

dp = days pregnant. This effect represents how many days the cow was pregnant up to the time the test day was taken. Open cows increase production early in lactation and have lower production late in lactation when they conceive early;

mf = milking frequency. This effect represents how many times the cow was milked in the day that the test data was taken; 3X milking generally results in more production;

dim = days in milking. This represents how many days the cow has been milked up to day that the test data was taken;

sdim = square root of days in milk;

ldim = logarithm of days in milk. These sdim and ldim were chosen for this model because they provide a good fit of the lactation curve;

a = this represents the cow’s additive genetic effect on yield;

$y_1 y_2$ = random regression coefficients related with the animal effect. These coefficients help to estimate the lactation curve that is specific for each animal;

pe = permanent environmental effect. This effect takes into account the different measures that happen for a cow within lactation, and also across lactation;

Q_1Q_2 = random regression coefficients related with the permanent environmental effect. Again, these coefficients help to estimate the lactation curve that is specific for each animal; and

e = a residual effect in test-day production that is not accounted for by other effects in the model.

The effects in the model that have an "*x*" are interaction terms. For example, the term "*pr x mf*" is the interaction of parity (*pr*) with milking frequency (*mf*). This is interpreted that the effect of milking frequency on yield is not the same for first calf heifer and older cows. Stated differently, when cows are changed from 2X to 3X this effect on yield is not the same for heifers and cows, or vice versa. The other interaction terms with milking frequency have similar interpretation.

This is a very complicated model. However, if we are going to estimate accurately the effects of 3X relative to 2X milking we need to account for as much of the variation in test-day yield as possible. There will be biases in estimated breeding values, if there are biases in the factors that adjust records to a 2X basis. Two types of adjustment factors have been used: additive and multiplicative factors. The first type is used when the variance does not increase as the mean increases and the second type, when the variance increases as the mean increases.

The adjustment factors can be obtained by considering the average level of the factor with the highest average as the standard and make the adjustment factor value for this level equal 0, in case of additive factor, and 1 in case of multiplicative factor. The adjustment factor values, for the other levels, would be 0 plus the difference between the mean of the standard level and the mean for each other level, in case of the additive adjustment, and the ratio between the mean of the standard level and the mean for each other level, in case of multiplicative adjustment.

Adjustment factors are only obtained for factors that explain a reasonable portion of the total variation. Some articles have considered reasonable the amount of 5% of the total variation. For all factors and interactions of the full model that showed significant *F*-test and explained more than 5% of the total variation, the residual variance within each level will be computed.

Data

We intend to use test day data provided by USDA for some U.S. states. The states are California, Utah, Washington, Wisconsin, Pennsylvania, New York, Iowa, Florida, North Carolina, Texas and Arizona.

Results

The results for this project will be available next year.