

Mining of producer recorded data; using beef calf and cow live-weight data as a case study

S. Walsh^{1,2*}, N. McHugh¹, R.D. Evans³, C. Murphy² and D.P. Berry¹

¹ Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, P61 P302, Co. Cork, Ireland; ² Department of Biological Sciences, Munster Technological University, Bishopstown Campus, T12 P928, Co. Cork, Ireland; ³ Irish Cattle Breeding Federation, Link Road, Ballincollig, P31 D452, Cork, Ireland; *shauna.walsh@teagasc.ie

Background: Animal live-weight contributes to profitability in beef herds and is a key determinant of overall efficiency of the beef sector. The objective was to develop a novel editing criteria for anomaly detection of beef cow and calf live-weight data. Live-weight data from five sources (i.e., professionally-recorded, owned-scales, borrowed-scales, scales hired from a depot, other) were available from the Irish Cattle Breeding Federation.

Methods: A number of alternative methods were used for anomaly detection including: generation of within-herd regression estimates, partial correlations between cow and calf live-weight records and mahalanobis distance. Across each method a value was calculated for each herd based on their distance from the assumed gold standard (i.e. professional recorded herds).

Results: Results suggest anomaly detection should use both within-herd regression and partial correlation estimates jointly. A contingency table was formed with herds grouped into five classes according to their herd-level regression and partial correlation estimates between cow and calf live-weight. Based on mahalanobis distance 20% of herds were considered anomalies, however the data source did not impact data quality. Herd characteristics including herd size, calving season, month of live-weight recording, proportion of dams born within the herd of weighing, and proportion of dams of dairy breed origin per herd were shown to be associated with the quality of data recorded.

Conclusion: Results from this research will contribute to the formation of a specific editing criteria for live-weight data and assist herds with poor recording in improving data quality, thereby resulting in improved data submitted for use in national genetic evaluations.

Keywords: data mining, beef, live-weight, anomaly detection.