

Repeatability of sheep body condition scoring

G. Refshauge^A, P. Quin^B

^A Centre for Red Meat and Sheep Development, New South Wales Department of Primary Industries, PO Box 129 Cowra NSW 2794

^B Centre for Autonomous Systems, Faculty of Engineering and IT, University of Technology Sydney 15 Broadway, Ultimo NSW 2007

^AEmail: gordon.refshauge@dpi.nsw.gov.au

Australian sheep producers are trained to tailor nutrition for groups of sheep using estimates for current body condition and future feed requirements, as well as including information about pasture availability and pasture growth outlook. The current best practice for assessing body condition score (BCS) is manual palpation of the lumbar region to feel the amount of fat and muscle tissue and the prominence of the backbones (Jeffries 1961). The value of BCS assessment is to increase whole-farm efficiency by allowing a producer to optimise the allocation of scarce feed resources and provide better control of reproduction outcomes. However, the assessment is necessarily subjective, resulting in wide disagreement between assessors and the potential for inefficiency. The objective of this study was to determine the accuracy of trained and novice assessors. This paper reports on the repeatability outcomes of three field trials, the error and a description for the modality of the histogram for each assessor.

Three field trials were undertaken at the Research Station, Cowra, NSW, to assess adult Merino ewes twice on the same day. Without conversation, using electronic ear tag recording equipment, all ewes were assessed in a sheep race. In Field trial 1 ($n = 152$), after the fifth assessor completed the race-load, all ewes were scored in the reverse direction, but without being released or mixed with other sheep. In Field trial 2 ($n = 129$), all ewes were released after each race was completed, mixed together and re-assessed. In Field trial 3 ($n = 88$), after their first assessment the ewes were mixed with another 40 ewes and all 128 were then assessed. At the commencement of Field trial 3, the trainer calibrated the producers using 6 ewes that were selected for their BCS variation. To assess user repeatability, the coefficient of determination (R^2) was calculated for the two assessments each assessor made for each sheep. The squared difference between each assessment was also calculated, the mean of all squared differences was converted to the original scale by square root to provide the error of assessment as the root mean square error (RMSE). It is assumed that an assessor with a high R^2 and low RMSE is an accurate assessor, so long as the histogram of their assessments are normal or near normal.

Assessor	Field trial	Within-assessor repeatability (R^2)	RMSE	Mean BCS	Skill	LTEM participation	Histogram
A	1	0.51	0.60	2.53	Low	Producer	Bimodal, right skew
B	1	0.75	0.33	2.87	Moderate	Producer	Right skewed
C	1	0.85	0.25	3.13	High	Researcher	Normal
D	1	0.81	0.41	3.75	Low	Consultant	Heavily skewed left
E	1	0.64	0.16	2.89	Moderate	Producer	No range detected
F	2	0.63	0.28	2.80	High	Trainer	Normal
C	2	0.74	0.31	3.37	High	Researcher	Slight left skew
H	3	0.50	0.27	3.49	Low	Producer	Normal
I	3	0.49	0.35	3.69	Low	Producer	Heavily skewed left
J	3	0.56	0.24	3.73	Low	Producer	Heavily skewed left
K	3	0.52	0.29	3.61	Low	Producer	Skewed left
L	3	0.46	0.34	3.26	Moderate	Producer	Bimodal, left skew
M	3	0.45	0.34	3.78	Low	Producer	Heavily skewed left
N	3	0.49	0.43	3.40	Low	Producer	Bimodal, left skew
O	3	0.48	0.38	3.64	High	Trainer	Slight left skew

Table. Repeatability (R^2), root mean square error (RMSE) and mean body condition score (BCS) estimates for assessors, including skill level, degree of participation in LifeTime Ewe Management (LTEM) and a description for histogram normality

This paper finds wide variation in the ability of trained and novice assessors to consistently recognise condition score, as could be expected (Kenyon *et al.* 2014). Within-assessor repeatability was moderate to high, varying from 0.45-0.85 and RMSE ranged from 0.16-0.60. Mean BCS within each field trial showed a wide range, 0.5 to 1.2 BCS. Examination of the histogram in conjunction with the R^2 and RMSE shows that assessor D had high R^2 and low RMSE but failed to detect the range in condition scores that other assessors observed. The research station flock was well managed, resulting in few very lean ewes being present for assessment. We suggest trained sheep producers be assessed for repeatability, error and distribution to provide them with the necessary feedback from which to improve.

References

Jeffries, BC. (1961) *Tasmanian Journal of Agriculture*. **32**, 19-21.

Kenyon PR, Maloney SK and Blache D. 2014. *New Zealand Journal of Agricultural Research*. **57**, 38-64.