Scientific Papers, Animal Science, Series D, vol. LV
CD-ROM ISSN 2285-5769, ISSN-L 2285-5750

ROBUST REGRESSION MODELS FOR PREDICTING THE LEAN MEAT PROPORTION OF LAMBS CARCASSES

Cristina XAVIER, Vasco CADAVEZ

Mountain Research Center (CIMO), ESA - Instituto Politécnico de Bragança, Campus de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal. Phone: + 351 273 303304, Fax: + 351 273, E-mail: vcadavez@ipb.pt

Corresponding author: vcadavez@ipb.pt

Abstract

The aim of this study was to develop and evaluate robust regression models for predicting the carcass composition of lambs. One hundred and twenty lambs (34 females and 86 males) were slaughtered and their carcasses were cooled for 24 hours. The subcutaneous fat thickness (C12) was measured between the 12th and 13th rib, and the left side of carcasses was dissected and the proportions of lean meat (LMP) was calculated. A multiple regression model was fitted using robust regression (RR) methods, and the results were compared to ordinary least squares (OLS) estimates. For RR methods, the Bisquare and Welsch weighting functions were used, and model fitting quality was evaluated by the following statistics: the root mean square error (RMSE), the median absolute deviation (MAD), the mean absolute error (MAE), and the coefficient of determination (R^2). The parameters obtained by RR presented lower standard error for C12 measurement (decreases by 12% when compared with OLS estimates). The RR methods or weighted least squares methods represents a good alternative to OLS approach for modelling the LMP of lambs carcass. In this study, the Bisquare weighting method presented the best results, however other weighting functions are available and should be tested and compared in the near future.

Key words*: carcass, quality, ordinary least squares, robust regression*

INTRODUCTION

An accurate objective system for the classification of carcasses at the slaughter-line is of great importance for meat industry, and research results attained by [4,1,2] indicate that the lean meat proportion (LMP) of lamb carcasses can be predicted by simple models using the hot carcass weight (HCW) and one fat depth measurement as explanatory variables.

Multiple regression models has been the most important statistical tool applied to develop models to predict carcass composition. The ordinary least square (OLS) method has been generally used considering this purpose, however data sets, usually, contain outliers which presents higher residuals. Thus, OLS estimates can be biased when the distribution of the residuals is not normal, specially when the residuals distribution is heavy-tailed [3]. When this occurs the OLS estimates are unreliable. The removal from the data set of influential observations can be a solution to this problem, however is observations are real this procedure leads to looses of information.

Robust regression (RR) methods are resistant to unusual data, and the most common general method of RR is M-estimation developed by [5]. When data are linear and residuals are normally distributed, the OLS and RR estimates are similar. The results will be quite different when the residuals are not normally distributed or when data contains a considerable number of outliers [8]. Thus, RR can be used, as a complement to OLS estimates, to handle unusual data records that are not following the general data trend, since RR is resistant to outliers and often performs better in the presence of heteroscedastic residuals.

In the case of models for predicting carcass composition, RR methods can be used to estimate equations parameters, whilst the unusual values are taken into account.Thus, the aim of this study was to compare OLS and RR methods for modelling the LMP of lamb carcasses.

MATERIAL AND METHOD

Data. One hundred and thirty five (45 females and 90 males) of Churra Galega Bragançana breed were randomly selected from the experimental flock of the Escola Superior Agrária de Bragança. The lambs were slaughtered, and their carcasses were weighed approximately 30 min after slaughter in order to obtain the hot carcass weight (HCW). After chilling at 4ºC for 24 hours, the carcasses were halved through the centre of the vertebral column, and the kidney knob and channel fat were removed and weighed. During quartering of the carcasses, the subcutaneous fat thickness (C12, mm) between the 12th and 13th rib was measured with a calliper. The left side of each carcass was dissected into muscle, subcutaneous fat, inter-muscular fat, and bone plus remainder (major blood vessels, ligaments, tendons, and thick connective tissue sheets associated with muscles). The carcasses' lean meat proportion (LMP) was calculated as a proportion of the total tissues in the carcasses.

Statistical analysis

In linear regression, the ordinary least squares (OLS) estimator is sensible to extreme observations, to overcome this problem [5] developed the a general method of robust regression (RR), the M-estimation. This class of estimators represents a generalization of maximum-likelihood estimation. Thus, in this study we used the following multiple regression model used was:

$$LMP = \alpha_0 + \alpha_1 \times LMP + \alpha_2 \times C12 + \varepsilon_i$$

where α_0, α_1, and α_2 are regression coefficients and ε_i is the error term.

All statistical analyses were undertaken using the software "R" [6], and the add-on package "robustbase" [7] was used to fit model using RR estimators.

The model fitting quality was evaluate using the following statistics: the root mean square error (RMSE), the median absolute deviation (MAD), the mean absolute error (MAE), and the coefficient of determination (R^2). The RR methods assign a weight to each observation of the dataset, and in this study we compared two weighting functions Bisquare and Welsch, with the OLS estimates (Table 1). The RR methods were fitted using the default tuning constant (Table 1), which leads to coefficient estimates that are approximately 95% as statistically efficient as the OLS estimates, if the response variable has a normal distribution without outliers. In the weight functions, the value of r is $r = resid/\left(tune \times s \times \sqrt{(1-h)}\right)$, where $resid$ is the vector of residuals from the previous iteration, h is the vector of leverage values from a least-squares fit, and s is an estimate of the standard deviation of the residuals given by $s = MAD/0.6745$ .The MAD is the median absolute deviation of the residuals from their median.

Table 1. Weighting function equations for RR methods

Weight function	Equation	Default tuning constant
OLS	$w=1$	none
Bisquare	$w = (((r)) < 1) \times \left(1 - r^2\right)^2$	4.685
Welsh	$w = \exp\left(-\left(r^2\right)\right)$	2.985

RESULTS AND DISCUSSIONS

The estimated parameters and summary statistics for the three estimation methods "OLS", "Bisquare", and "Welsch" are presented in Table 2. The OLS estimates ignores the presence of outliers in the data set, these estimates differ from RR estimates. The efficiency of the RR estimates compared to the OLS estimates is expected to increase in the presence of outliers [3], which is expectable for carcass composition traits. The parameters obtained by RR are characterized by lower standard errors, the reduction of the standard errors was especially visible for C12 measurement (decreases by 12%). Thus, modelling the LMP of lambs carcasses ignoring the presence of outliers results in inefficient estimates.

Table 2. Estimation results of the three methods

Parameters	OLS	Bisquare	Welsh
Intercept	63.0	62.9	62.9
SE	0.788	0.806	0.804
Pr(>|t|)	<2e-16	<2e-16	<2e-16
HCW	0.201	0.198	0.198
SE	0.075	0.079	0.079
Pr(>|t|)	0.008	0.014	0.013
C12	-2.61	-2.58	-2.58
SE	0.210	0.185	0.185
Pr(>|t|)	<2e-16	<2e-16	<2e-16

Table 3 shows model fitting quality statistics. The RR methods presented better overall fit, indicated by the highest (0.591 for OLS, 0.651 for Bisquare, and 0.611 for Welsch).

The standard errors of the regression coefficients and the standard errors of the estimates (SEE) were larger in the OLS estimates, specially for the C12 measurement. The RMSE, the R2, and MAD statistics show that Bisquare is the best weighting function for LMP of lamb carcasses (RMSE = 2.55, R^2 = 0.651, MAD = 2.43). The RR methods presented better fitting quality statistics because they have the ability to minimize the effect of influential observations on the estimation of model parameters.

Table 3. Model fitting quality statistics for the three methods

Weighting function	MAE	RMSE	MAD	R^2
OLS	2.15	2.80	2.46	0.591
Bisquare	2.14	2.55	2.43	0.651
Welsh	2.15	2.55	2.49	0.611

CONCLUSIONS

The RR methods or weighted least squares methods represents a good alternative to OLS approach for modelling the LMP of lambs carcass. These methods can be used to minimize the effects of influence of observations (outliers) in data sets. However, if the data set do not have outliers, the results of OLS and RR will be similar. Thus, we can recommend the use of RR methods to modelling the lambs carcass quality traits. The Bisquare presented the best results in this study, however other weighting functions are available and should be tested in the near future.

REFERENCES

[1] Cadavez, V. A. P., 2009. *Prediction of lean meat proportion of lamb carcasses.* Archiva Zootechnica, 12(4), p.46–58.

[2] Cadavez, V. A. P., Henningsen, A., 2010. *The use of seemingly unrelated regression to predict the carcass composition of lambs.* Meat Science, 92, p.548:553.

[3] John, Fox, Sanford, Weisberg, 2010. *Robust Regression in R*. An Appendix to An R Companion to Applied Regression. Second edition edition, 15 December 2010. An Appendix to An R Companion to Applied Regression.

[4] Hopkins, D. L., 2008. *An industry applicable model for predicting lean meat yield in lamb carcasses*. Australian Journal of Experimental Agriculture, 48, p.757–761.

[5] Huber, P. J., 1964. *Robust estimation of a location parameter.* Annals of Mathematical Statistics, 35, p.73–101.

[6] R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[7] Rousseeuw, Peter, Christophe, Croux, Valentin, Todorov, Andreas, Ruckstuhl, Matias, Salibian-Barrera, Tobias, Verbeke, Manuel, Koller, Martin, Maechler, 2012. *Robustbase: Basic Robust Statistics*. R package version 0.9-3.

[8] Zia Ul-Saufie, Ahmad, Ahmad, Shukri Yahaya, Nor Azam Ramli, Hazrul Abdul Hamid, 2009. *Robust regression models for predicting pm10 concentration in an industrial area*. International Journal of Engineering and Technology, 2(3), p.364–370.