*Res. Lett. Inf. Math. Sci.* (2003) **4**, 173-180
Available online at http://iims.massey.ac.nz/research/letters/

# Calculating animal performance from limited liveweight measurements of the population

G.C. Wake
*Centre for Mathematics in Industry, IIMS, Massey University at Albany, Auckland, N.Z.*
*g.c.wake@massey.ac.nz*

A.B. Pleasants
*Mathematical Biology Unit, AgResearch, Private Bag 3123,Hamilton,N.Z.*
*tony.pleasants@agresearch.co.nz*

**Abstract**
We consider the problem of estimating the distribution of carcass weights in a flock of animals from estimates made on a truncated sample. This arises when a farmer chooses the heaviest lambs for slaughter and then measurements are made by the meat processor. This enables a farmer to answer two questions: what proportion of the animals remaining exceed a nominated carcass weight, and/or what carcass weight is exceeded by a nominated proportion of the population? Estimates of these statistics and their uncertainties are derived and are exact if the animal weights are normally distributed. These calculations can be the basis of decisions about future feeding and drafting strategies, important for farmers producing animals on contracts for future delivery. An example is given based on 1000 lambs using a cut-off weight of 15.5 kg with mean of this upper group of 16 kg. Using a realistic estimate of a standard deviation (of the weighing scales) of 0.3 kg, this gives an estimated mean of 14.6± 0.04 kg, with a standard deviation of 0.94±0.044 kg, and that 75% of the lambs in the population exceed 13.9±0.07 kg. The proportion of lambs that exceed 14.5 kg is then between 51.3% and 55.7%.

## Introduction

Integrated management of the food supply chain requires livestock farmers to meet objective specifications on the weight of animals sold for slaughter. Generally these animals need to be above a nominated carcass weight and to have a minimum fat cover to be acceptable. The farmer estimates carcass weight before slaughter by the animal live weight. It is common to record only the live weights of those animals selected for slaughter. When the animals are slaughtered, only the carcass weights of those selected animals will be available for the meat processor.

However, information on the liveweight distribution of the animals not selected for slaughter is valuable in farm decision making. For example, estimates of the numbers of lambs growing to future slaughter weights would be useful in decisions about the resources required in order to achieve this. This paper describes the methodology to do this by deriving a method for estimating the frequency distribution of the live-weights or carcass weights of the animals not selected for slaughter from the data recorded for the animals that are slaughtered.

Assuming the frequency distribution for animal live-weight is normal, we begin by deriving the estimators for the mean and standard deviation of the animal live-weights in the whole population when only the mean and proportion of the animals selected for slaughter are known. From these estimates any statistic of the animals not selected for slaughter can be calculated. We then use this to find the uncertainty in the statistics of interest for the unselected population, and to apply these estimates to practical situations. Finally we note that the assumption of normality is based on it being the most likely, but we can extend this methodology easily to any other distribution, for example a t-distribution. So it is a generic question, which can arise for any probability distribution. Future work will extend this to the case where the multivariate weight and fat distribution is truncated for both variates.

**The Model**

Assume that the liveweight frequency distribution is normal with mean μ and standard deviation σ. Suppose, as is current practice, that the farmer weighs his/her animals, but only records the mean live-weight ($\overline{w}_L$) of the animals selected for slaughter. Then $\overline{w}_L$ is the mean of a truncated normal density. The farmer also knows the actual numbers, and hence the proportion of the population (α) of the animals selected for slaughter. This knowledge is expressed mathematically as:

$$\begin{aligned} &\int_W^\infty \frac{e^{\frac{-(y-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\,dy = \alpha \\ &\int_W^\infty \frac{y e^{\frac{-(y-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma\alpha}\,dy = \overline{w}_L \end{aligned} \qquad (1)$$

where the liveweights of the animals selected for slaughter (but not recorded) is y, and W is the cut – off liveweight – the liveweight above which animals are selected for slaughter. The parameter α is included in the second equation to express the fact that we are calculating the mean of the upper portion.

This is shown in the diagram given in Figure 1, which uses the values of the weights in the section "An Application" described later.

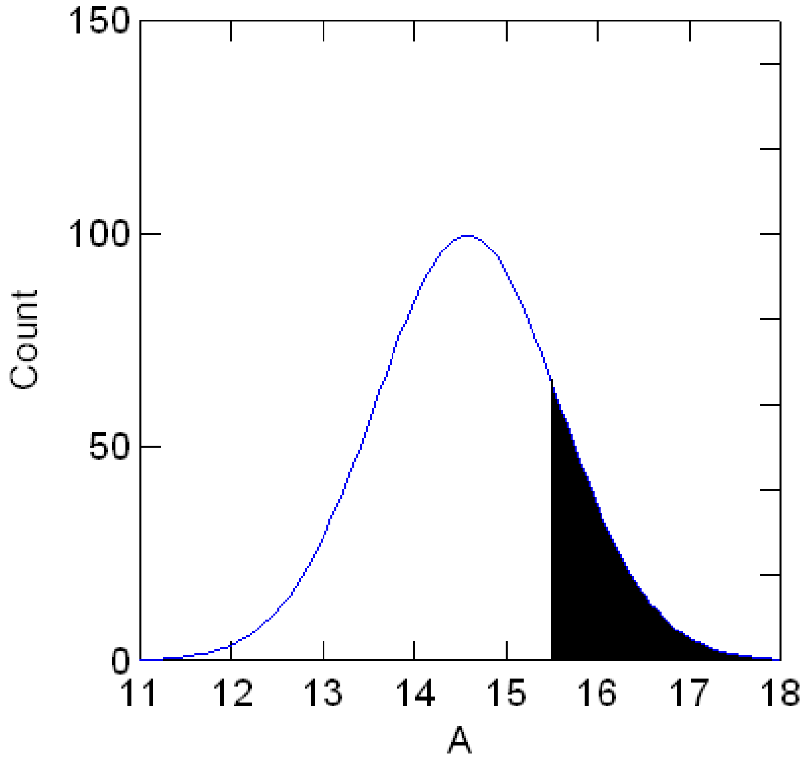


**Figure 1:** Graph of the truncated normal distribution. The fraction α of the distribution to the right is removed and has mean weight $w_L$. Here A = live weight.

The pair of equations, remarkably, has the following solution for μ and σ as shown in appendix 1:

$$\begin{aligned} \hat{\mu} &= WU - \overline{w}_L Q \\ \hat{\sigma} &= (\overline{w}_L - W)R \end{aligned} \qquad (2)$$

This provides explicit solutions for the population mean and standard deviation in terms of the mean of the selected group of animals and the proportion of animals selected. Assume $\overline{w}_L$ is normal with variance $\frac{\tilde{\sigma}^2}{n}$ where $\tilde{\sigma}^2$ is the variance of the weighing scales and n is the number of animals in the selected group. Thus, in (2) $\hat{\mu}$ will have variance, $\frac{\tilde{\sigma}^2 Q^2}{n}$ and $\hat{\sigma}$ will have variance

$\frac{\tilde{\sigma}^2 R^2}{n}$. Noting that from the equations (3) in the appendix $Q = \sqrt{2}\,erfc^{-1}(2\alpha)R$, then the covariance between $\hat{\mu}$ and $\hat{\sigma}$ will be $\frac{2\tilde{\sigma}^2 erfc^{-2}(2\alpha)R^2}{n}$ where we have for simplicity written *erfc* $^{-2}$(z) for ( *erfc* $^{-1}$(z))$^2$. Similarly with the function *erf*.

Here *erfc*$^{-1}$(2α) denotes the inverse of the complementary error function of argument 2α. To calculate these functions note that:

$$erfc(2\alpha) = 1 - erf(2\alpha),$$

where

$$erf(2\alpha) = \frac{2}{\sqrt{\pi}} \int_0^{2\alpha} e^{-t^2}\,dt,$$

$$erfc^{-1}(2\alpha) = erf^{-1}(1 - 2\alpha)$$

The inverse of the error function (*erf*(x)) is calculated using Matlab as the function *inverf.* Of course the function *erf* is closely related to the cumulative normal *F* by the relationship

$$F(z) = \tfrac{1}{2}\,(1 + erf(\sqrt{2}\,z)).$$

The utility of this information lies in the calculation of the quantiles of the animals in the population from which weights have not been taken. An important practical question is to calculate the lower bound on the weight of a given proportion of animals in the group. Let the given proportion be *x*, and the unknown estimated lower bound be *K* then from appendix 1:

$$K = \hat{\mu} + \sqrt{2}\hat{\sigma}\, erf^{-1}(1 - 2x)$$

From the above information *K* will a normal random variable with mean given above and variance given by:

$$\sigma_K^2 = \frac{\tilde{\sigma}^2 \left[ Q^2 + 2erf^{-2}(1-2x)R^2 + 4erfc^{-2}(2x)R^2 \right]}{n}$$

$$= \frac{\tilde{\sigma}^2 \left[ Q^2 + 6erf^{-2}(1-2x)R^2 \right]}{n}$$

An alternative question relates to estimating the proportion of animals in the group above (or below) a nominated weight given the estimates $\hat{\mu}$ and $\hat{\sigma}$. This will be given as the cumulative normal density with the above mean and standard deviation. However, the probability density of this estimate will not be normal due to the non-linearity of the relationship between the proportion above a nominated weight and the estimated parameters $\hat{\mu}$ and $\hat{\sigma}$. In this case the approximate variance can be found as in Mood *et al.* (1974) as the long expression below:

$$\sigma_\alpha^2 \cong \frac{\tilde{\sigma}^2}{n} Q^2 \left[ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(W-\mu)^2}{2\sigma^2}} \right]^2 + \frac{\tilde{\sigma}^2}{n} R^2 \left[ \left( \frac{1}{\sigma^2} - 1 \right) \int_W^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds \right]^2$$
$$+ \frac{2\tilde{\sigma}^2 \, erfc^{-2}(2\alpha)}{n} R^2 \left( \frac{1}{\sigma^2} - 1 \right) \left[ \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(W-\mu)^2}{2\sigma^2}} \right) \left( \int_W^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds \right) \right]$$

Notice that it is mostly in terms of error functions and/or the cumulative normal.

**An Application**

Suppose a farmer selects the heaviest 165 lambs from a mob of 1000 using a cut – off weight of 15.5kg. When slaughtered, the average carcass weight of the selected lambs is 16 kg. Then $\alpha$ = 0.165, W = 15.5 kg and $\overline{w}_L$ = 16 kg. The standard deviation on the weighing scales is 0.3kg. From this information we find from the above $\hat{\mu}$ = 14.6 ± 0.04 kg, and $\hat{\sigma}$ = 0.94 ± 0.044 kg. The estimated carcass weight that 75% of the lambs in the mob exceed, K = 13.9 ± 0.07 kg. The proportion of the mob with carcass weights above 14.5 kg is 0.53 ± 0.027.

**Discussion**

The standard deviations associated with the above estimates would suggest that the estimates are sufficiently accurate to be of value for management decision-making. For example, in the above case 750 lambs minus 165 slaughtered = 585 lambs are estimated to have a carcass weight heavier than 13.9kg. Alternatively, with a standard error of 0.07 there is a 95% confidence that at least 585 of the remaining lambs will exceed 13.9 – 1.96 × 0.07 = 13.7 kg.

That is, there are 585 lambs that have to gain at least 1.8 kg of carcass weight to exceed the cut–off weight with a confidence of 95%. Similarly, an estimated 365 lambs (530 – 165) need to gain at least a kilogram of carcass weight to exceed the cut–off weight.

Although it is complex, the mathematics is suitable for coding into a computer program that can present the results to a user in a form useful to easily make decisions about the feeding and management of animals for slaughter.

The estimates made using the methods in this paper can also be the basis for probabilistic constrained optimisation. For example, suppose a farmer had a known amount of feed for lambs, and he/she wished to run the lambs in two groups. If the grouping is to be decided on a weight basis then what is the optimal weight to make the split into the two groups in order to achieve the management goals with the best probability. The above provides a sound way to answer this question.

The accuracy of the statistics derived here are dependent on the frequency distribution of animal weights being normal. This is not a particularly restrictive assumption since it has been shown by Wake *et al.,* (1999) that a group of animals which have had a proportion removed (e.g. selected for slaughter) so that the frequency distribution is truncated will return to a normal density rapidly (generally less than 20 days for normal lamb growth rates). Thus unless animal growth was particularly abnormal the results reported here should be stable and useful estimates of the frequency distribution of animals remaining on the farm.

Although the application discussed here is that of the management of animals for slaughter, there are a range of circumstances in which only a sample from the upper tail of the distribution is made, yet inferences are required on the whole population. For example, in measuring microbial contamination, the resolution of the test limits observations to microbe numbers above the threshold measure. Similar questions arise in the manufacturing sectors. It is emphasised that the solution given here in equations (2) and (3) is general and applies in all these and other applications.

**References**

Mood, A.M.; Boes, D.C.; Graybill, F.A. 1974: Introduction to the theory of statistics. McGraw Hill, USA.

Wake, G.C.; Soboleva, T.K.; Pleasants, A.B. 1999: The evolution of a truncated Gaussian density through time – modelling animal liveweights after selection. *In*: Wilson, R.J.;. Osaki, S.; Faddy, M.J. *Ed* Proceedings of the first Western Pacific and Third Australia-Japan Workshop on. Stochastic Models in Engineering, Technology and Management. University of Queensland. pp 547-555.

## Appendix 1. Derivation of the mean and standard deviation from the truncation normal disttribution

We first define 3 expressions which will be needed in the derivation. Let

$$U = \frac{1}{1 - 2\alpha\sqrt{\pi}e^{\left(erfc^{-1}(2\alpha)\right)^2} erfc^{-1}(2\alpha)}$$

$$Q = \frac{2\alpha\sqrt{\pi}\, erfc^{-1}(2\alpha)}{e^{-\left(erfc^{-1}(2\alpha)\right)^2} - 2\alpha\sqrt{\pi}\, erfc^{-1}(2\alpha)} \qquad (3)$$

$$R = \frac{\alpha\sqrt{2\pi}}{e^{-\left(erfc^{-1}(2\alpha)\right)^2} - 2\alpha\sqrt{\pi}\, erfc^{-1}(2\alpha)}$$

where $erfc^{-1}(2\alpha)$ denotes the inverse of the complementary error function of argument $2\alpha$. To calculate this function we note that:

$$erfc(2\alpha) = 1 - erf(2\alpha),$$

$$\text{where} \quad erf(2\alpha) = \frac{2}{\sqrt{\pi}} \int_0^{2\alpha} e^{-t^2}\, dt,$$

$$erfc^{-1}(2\alpha) = erf^{-1}(1 - 2\alpha)$$

The inverse of the error function ($erf(z)$) is calculated using Matlab as the function inverf and is of course closely related to the cumulative normal $F$ by the relationship

$$F(z) = \tfrac{1}{2}\left(1 + erf(\sqrt{2}\, z)\right).$$

The first equation in (1) can be written:

$$erfc\left(\frac{W-\mu}{\sqrt{2}\sigma}\right)=2\alpha,$$

or $$erf\left(\frac{W-\mu}{\sqrt{2}\sigma}\right)=1-2\alpha, \qquad \text{(a)}, \text{ and so}$$

$$W=\sqrt{2}\sigma\ erf^{-1}(1-2\alpha).$$

The second equation in (1) can be integrated by parts to give:

$$e^{-\frac{(W-\mu)^2}{2\sigma^2}}=\frac{\sqrt{2\pi}\alpha}{\sigma}\left(\overline{w}_L-\mu\right) \qquad \text{(b)}$$

The two equations (a) and (b) regarded as determining μ and σ in terms of W and $\overline{w}_L$ remarkably reduce to two linear equations for the two unknowns:

From (a) and (b):

$$\overline{w}_L=\mu+\frac{\sigma e^{-\left(erf^{-1}(1-2\alpha)\right)^2}}{\alpha\sqrt{2\pi}} \qquad \text{(c)}$$

Then (a) and (c) are easily solved to give:

$$\mu=WU(\alpha)-\overline{w}_L Q(\alpha)$$
$$\sigma=(\overline{w}_L-W)R(\alpha)$$

where U(α), Q(α) and R(α) are as given by equation (2).