NITROGEN LEACHING FROM SHEEP-GRAZED HILL COUNTRY: ESTIMATING THE TRUE POPULATION MEAN FOR NON-NORMALLY DISTRIBUTED DATA

Coby Hoogendoorn, Catherine Lloyd-West and Brian Devantier

AgResearch, Grasslands Research Centre,
Tennent Drive, Private Bag 11008, Palmerston North 4442, New Zealand
coby.hoogendoorn@agresearch.co.nz

Abstract
Nitrogen (N) leaching data obtained from grazed pastures rarely conform to a normal distribution. The data generally has a large number of very low values originated from areas with little N leaching (non-urine patch areas), and a smaller number from points with large values (urine patch areas, including patch overlap). Transformation of the data to another scale is often necessary to meet assumptions for many statistical analyses. Whilst data transformation allows the appropriate testing of treatment effects on N leaching, an estimate of the amount leached per treatment accompanied by an indication of the uncertainty surrounding that number, is also of key interest.

For normally distributed data, the sample mean of the raw data (raw mean) is the appropriate estimate of the true population mean, with its standard error being a descriptor of the variability surrounding that mean. However, for non-normally distributed data the raw mean is no longer an appropriate estimate of the true population mean. In this case the back-transformed mean and 95% confidence intervals are the measure of central tendency commonly used to provide an estimate of the true population mean. There are a number of methods, additional to the simple back-transformed mean, used to provide estimates of the true population mean for non-normally distributed datasets, and each has their particular relevance depending on the nature of the data.

N leaching data collected from low (0-12°) and medium (13 – 25°) slopes from a recently completed 3-year study in hill country is used to illustrate the effect of using different methods of estimating the true population mean when summarising non-normally distributed data. The study area had a slope class mix of 16, 56 and 28% low, medium and steep slopes (>25°), respectively, and was stocked with sheep at 11 SU ha^{-1}. We present three different methods of estimating the true population means for the two treatments in our non-normally distributed dataset and discuss the appropriateness of each method.

Introduction
The general protocol for conducting scientific research is to pose a question, develop a hypothesis based on prior information, design and conduct an experiment to test the hypothesis, perform appropriate statistical analyses on the data, and then interpret and discuss the results - accepting or rejecting the null hypothesis. As scientists it is assumed that we are skilled and ethical at developing our questions and hypotheses and at designing our experiments. We also have very prescribed and accepted ways of collecting our data, and then we are generally good at getting professional advice on the appropriate statistical analysis of our data. When we communicate our results to our target audiences - through peer

reviewed publications, reports, posters and oral presentations - we usually strive to present a number that represents an estimate of the true population mean of a treatment to summarise our findings. As an audience we also often wish to have one number, or set of numbers to remember and compare with other published work. However, data from biological systems do not often conform to a normal distribution and so our data are frequently non-normally distributed. Giving a valid estimate of the true population mean and an appropriate measure of uncertainty surrounding that estimate, which neatly summarises our data, is therefore challenging.

A measure of central tendency is a single value that attempts to describe a set of data or treatment effect by identifying the central position within that set of data. The raw mean (often called the average) is the most popular and well known measure of central tendency, but there are others, such as the raw median. For a normally distributed sample, the raw mean and the median can both be legitimately used as measures of central tendency, because in a perfectly symmetrical distribution they are both equal. In this situation, the raw mean is preferred as the best measure of central tendency because it is the measure that, unlike the median, includes all the values in the dataset for its calculation, and any change in any of the values will affect the value of the mean.

As a dataset becomes skewed the raw mean loses its ability to provide the best estimate of central tendency of that data because the skewness is pulling the raw mean away from the typical value and in the direction of the skew. In these situations, the raw median is generally considered to be a better representative of the central location of the data. The simplest method of assessing normality is to look at the frequency distribution histogram of one's data. The most important things to look at are the skewness and kurtosis (i.e. peakiness) of the curve. Visual appraisals should always be supported by formal tests for normality as a best practice in statistics and most statistical packages have this function. While data transformations allow us to perform robust statistical analyses, it is important to remember that when we use a transformation, we are choosing the scale (log, square root etc.) for ease of statistical analysis and not to get the answer we desire or expect, given our prior knowledge and institutional memory.

But, how do we estimate the population mean for non-normal datasets that reflects the true nature of the dataset or population, often with limited numbers (less than what we would feel is ideal for our type of measurements)? For data transformed to logarithms, for example, one can express just the log transformed mean and confidence interval but this is generally meaningless to an audience. For right-skewed data, the back-transformed means are generally smaller than the mean of the untransformed data because the transformation reduces the influence of large observations in the calculation of the mean. Although the back-transformed mean (and confidence intervals) are less subject to distortion by the unusually large values in the tail of the positively skewed distribution of the data, the back-transformed mean really gives us an estimate of the median of the dataset, which may not be considered as the most appropriate estimate of the true population mean for non-normally distributed data.

The aim of this paper is to raise awareness and promote thought and discussion amongst workers who gather, disseminate, read and interpret information gathered from biological systems. We do this by exploring different, yet statistically valid, ways to express a measure of central tendency or estimate of the true population mean for a non-normally distributed dataset. To illustrate this we use a dataset obtained from a recently completed three year study on nitrate N leaching in sheep grazed hill country in New Zealand.

The dataset

Nitrate leaching information was obtained in a 3 year study using 96 *in situ* lysimeters (15 x 30 cm) positioned in low (0-12°) and medium (13-25°) slope areas of southern North Island sheep-grazed hill country. The study area was divided into 3 hillside blocks, all southwest facing; each block was divided into two 0.4 ha paddocks, with each paddock having similar areas of low, medium and high (>25°) slope. Two grazing intensity treatments were imposed in each block; one paddock in each block receiving an intensive grazing treatment (14 stock units (SU) ha^{-1}) and the other an extensive grazing treatment (8 SU ha^{-1}). The trial design was a split plot, with grazing intensity as the main plot and slope class as the sub-plot.

Sixteen lysimeters were installed in each of the six paddocks with 8 lysimeters each located in low and medium slope areas, giving an overall trial total of 48 lysimeters in each slope class. The lysimeters were installed approximately 8 months before collection of leachate began. Leachate collection occurred after every 100 mm rain or approximately monthly, with approximately 12 collection events per year over the three years. At each collection, leachate samples were weighed and a subsample taken for nitrate N analysis. The total amount of nitrate N leached from each lysimeter at each collection period was calculated by multiplying the weight of leachate by the nitrate N concentration and then summed for each collection to give an annual amount of nitrate N leached.

Both the annual and 3-year (Figure 1) mean (Figure 1) annual nitrate N leaching data were non-normally distributed, and highly skewed to the right. Transformation of the data was necessary to meet the assumptions required for performing the parametric statistical analyses. The dataset followed a log normal distribution and, once log transformed, an analysis of variance was performed according to a split plot design.

There was a highly significant (P<0.001) effect of slope class, but not of grazing intensity (P=0.211), on the amount of nitrate N leached. Since there was no significant interaction between slope class and grazing intensity (P=0.18), the data from the two grazing intensities were combined for each slope class.

Expressing estimates of the true population mean for each of the two slope categories proved challenging. The raw means were 124 and 16 kg nitrate N ha^{-1} for low and medium slopes, respectively (Table 1), and are clearly not a fair measure to describe the central tendency or the true population mean of these two respective datasets. The raw median values (58 and 1 kg nitrate N ha^{-1}, respectively) are somewhat more representative as they are less influenced by the few higher values than the mean. In published literature, there are a number of approaches used to express the true population mean for a non-normally distributed dataset and some methods are more commonly used in certain disciplines. In the field of computer sciences, for example, a refreshing perspective is presented in a paper by John Mashey (2004) entitled “War of the Benchmark Means: Time for a Truce”, and is well worth reading.

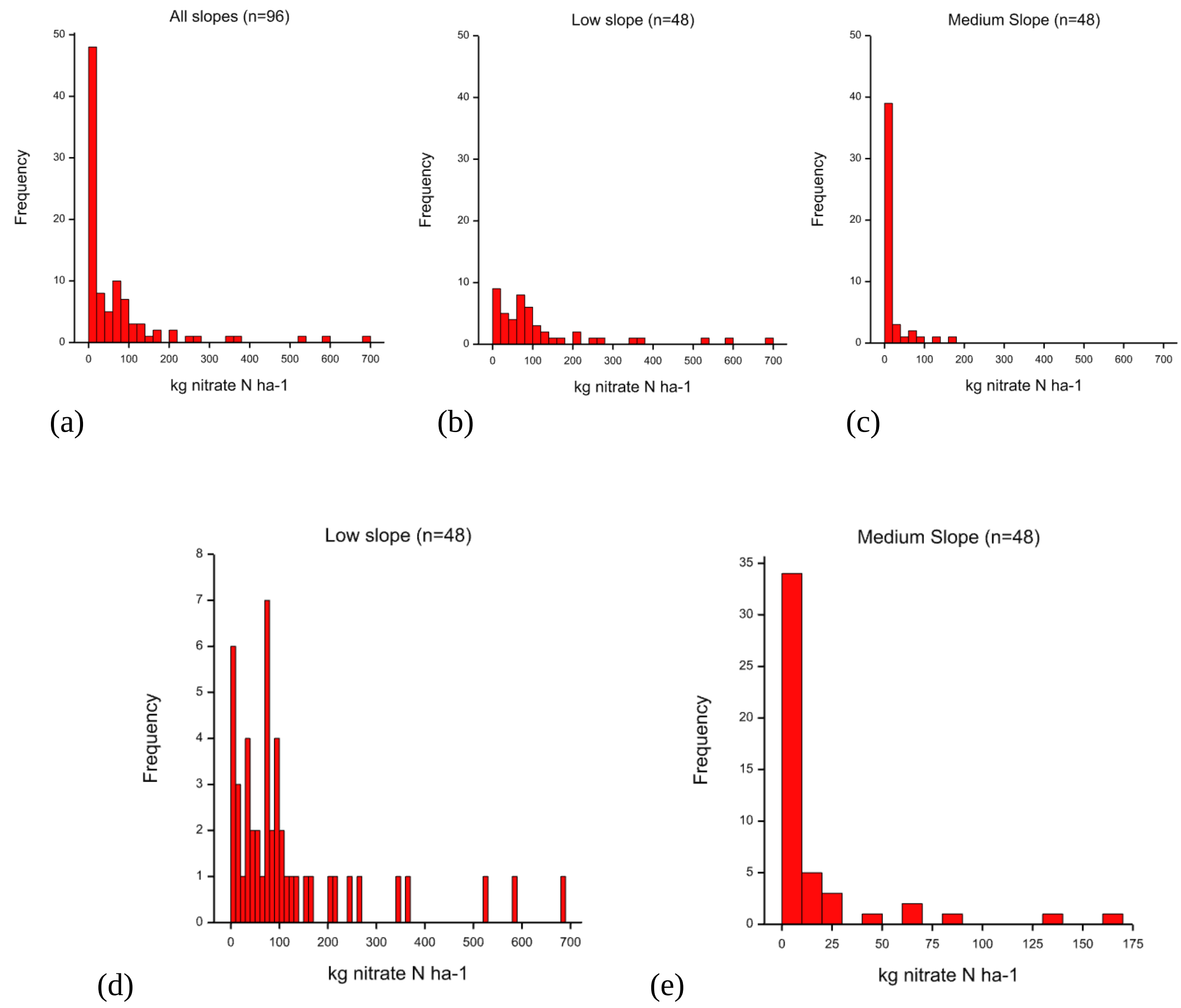

Figure 1. Frequency distributions of 3-year mean annual nitrate N leached per lysimeter for (a) both slope classes combined (n=96), (b) low slope only (n=48) and (c) medium slope only (n=48) using identical scales for all axes. Frequency distributions of (d) low and (e) medium slopes with a scale difference between the slope classes.

Back-transformed mean - Naïve method: exp ($\bar{Y}$)

One of the most widely used ways to describe a non-normally distributed dataset is to present the exponent of the mean of the logged data, or the back-transformed mean, with its 95% confidence interval (CI) (Olivier et al, 2008). This is often termed the "naïve" method, because it assumes that the researcher knows very little about the nature of the data obtained from the system that s/he is working with. This approach is considered by some to be best utilised when exploring systems or areas that have traditionally not been well researched and so the dataset obtained has few datasets to compare with. Additionally the naïve approach is suited to small datasets where the level of replication is less than desirable. In our case, the simple back-transformed means are 31 and 1 kg nitrate N ha^{-1} for low and medium slopes, respectively (Table 1). Note that, as expected, the back-transformed means are closer to the raw median than the raw mean (they are identical for the medium slope).

Table 1. Values and 95% confidence limits (CIs) for methods to estimate the true population mean for nitrate N leached per annum in sheep grazed hill country (kg nitrate N ha^{-1} per annum). Values are rounded to the nearest kg for simplicity. BT = back-transformed.

	Estimate of the true population mean	Value	Upper CI	Lower CI
Low slope	Raw mean	124	156	92
	Raw median	58	86	29
	BT mean (naïve)	31	41	23
	BT mean (Cox method)	127	196	82
	Sichel's estimator	67	81	56
Medium Slope	Raw mean	16	26	7
	Raw median	1	1	1
	BT mean (naïve)	1	2	1
	BT mean (Cox method)	6	8	4
	Sichel's estimator	4	6	3

Back-transformed mean corrected for bias - Cox method: $(\bar{Y} + \frac{V}{2})$

Adjusting the mean of the logged values by using the sample variance is one way to account for the few higher values in the dataset. The Cox method (Zhou and Gao, 1997) uses the exponent of $(\bar{Y} + \frac{V}{2})$. In our case, we have a fairly generous dataset compared to many nitrate N leaching studies conducted on grazed pastures, although it could be argued that our level of replication was still too low (Lilburne at al. 2012). However, few data exist on nitrate N leaching in grazed hill country and we would find it difficult to find a similar dataset with which to compare our data. Nevertheless, if we felt that our dataset was sufficiently large enough to obtain a good measure of the variance of the population, the Cox method uses this variance to correct for the bias of few but large values. Using the Cox method we find that, since our variance is rather large, the back-transformed bias corrected mean is actually slightly larger than our raw mean for low slopes but not for the medium slopes (127 vs. 124 and 6 vs. 16 kg nitrate N ha^{-1} for low and medium slopes, respectively). The Cox method of bias correcting the back-transformed mean takes into account the nature of the sample data, but it assumes that the variance of the sample is a reflection of the variance of the true population and we know that for data collected from grazed grassland this may not in fact be the case (Lilburne et al 2013). In the case of our data the Cox method may not be appropriate.

Sichel's estimator: $\left(\frac{\bar{x}\,2}{\sqrt{\bar{x}^2 + s^2}} \right)$

Sichel's estimator was developed for predicting of reserves in the mining industry (Clark et al., 1987) and White et al. (1987) showed Sichel's estimator to be an appropriate measure for estimating the true population mean for non-normally distributed datasets of soil inorganic N (nitrate N plus ammonium N) content and inorganic N leaching estimates. White et al. (1987) concluded that Sichel's estimator was appropriate for smaller datasets which have large variance and a high coefficient of skewness because it adjusts the raw mean using the sample variance, but in a different way to the Cox method. In the case of our data Sichel's estimator for nitrate N leaching is 67 and 4 kg nitrate N ha^{-1} for low and medium slopes, respectively. Again values generated using Sichel's estimator are closer to the raw median than raw mean, but greater than simply taking the exponent of the mean of the logged values (back-transformed mean (BT mean – naïve) (Table 1).

With non-normally distributed datasets, the 95% confidence interval is an important measure of the uncertainty surrounding measures of central tendency because it gives an indication of the likely values one can expect. Each of the measures in Table 1 has its corresponding set of CIs and it is worth noting the large difference between the upper and lower CI for some of the measures, and the overlap between the various measures and their CIs.

For each slope class there is a large difference in both the value and relativities between the different estimates. For example, the raw median amount of N leached from medium slopes is 13% of that leached from low slopes but only 2% when using the raw mean.

When these values used in the examples presented in Table 1 are translated to a per ha of grazed hill country, the amount of nitrate N leached from a nominal ha of hill country is still heavily influenced by the amount of N leaching from low slope areas, despite these areas occupying only 16% of the grazed area (Table 2). Conversely, whilst just over half the grazed area is occupied by medium slope areas, the amount of nitrate N leached per unit area in this slope class is much lower, and thus contributes less to the total amount leached per unit area of hill country. Whilst we did not measure N leaching from steep slope areas (>25°) we can assume that the amount would be even lower than from medium slope areas since (a) steep slope areas occupied less than a third of the total area, (b) less urine is likely to have been deposited in this slope class (Saggar et al. 1990) and (c) steep slope areas are likely to be even more N limited than medium slope areas and therefore any N deposited would be rapidly taken up by plants and/or incorporated into organic matter (Hoogendoorn et al. 2011).

Table 2. Amount of nitrate N leached (kg nitrate N per annum) for low and medium slope areas for a nominal 1 ha of sheep grazed hill country; where low and medium slopes occupy 16 and 56% of the land area, respectively. BT = back-transformed.

	Raw median	Raw mean	BT mean - naïve	BT mean - Cox	Sichel's estimator
Low slopes (16% of hill land area)					
value	9	20	5	20	11
95% CIs	5 - 14	15 - 25	4 - 7	13 - 31	9 - 13
Medium slopes (56% of hill land area)					
value	0.5	9	0.8	3.4	2.5
95% CIs	0.3 - 0.7	4.9 - 15	0.6 - 1.0	2.2 -5.0	1.9 - 3.4
Total leached from Low and Medium slopes (72% of land area)	10	29	6	23	14

Discussion

Choosing the most appropriate estimate of the population mean to summarise a non-normally distributed dataset is challenging. For the dataset presented here, there are few relevant studies to compare with. Lambert et al. (1985) monitored both runoff and leaching from the bottom of a catchment of similar topography and soil type to the one used in the present study and reported annual amounts of inorganic N loss of approximately 4.5 kg N ha^{-1}; comparable to the amount of nitrate N leached per ha on medium slope areas in the present study if Sichel's estimator is used to estimate the true population mean. Parfitt et al. (2009) reported a 3-year mean annual leaching loss of approximately 30 kg nitrate N ha^{-1} from low slope areas of similar hill country; comparable with the simple back-transformed mean for low slopes in our study, but lower than if Sichel's estimator is used to estimate the true population mean for low slopes.

When choosing the most appropriate estimate of central tendency for a dataset one cannot help but be influenced to some degree by one's prior knowledge, experiences and by institutional (in a science discipline sense) memory. Institutional memory may be unconsciously ingrained so that it becomes hard to challenge something found to contradict what was previously thought to have been "correct". The risk is that one becomes parochial in one's choices. It is wise to seek expert advice from statisticians who are invariably less influenced by institutional memory and can offer more objectivity in the decision making process.

The raw mean of a dataset is the most common method of estimating the true population mean in agricultural science and is often reported, even when the dataset from which it is derived is non-normally distributed. It is questionable whether the raw mean should be used at all in this instance because there are a number of more suitable options to choose from depending on one's prior knowledge.

Ideally if one had some understanding of the nature of the system one is sampling, one could make an *a priori* decision about the most appropriate method of estimating the true population mean to summarize a dataset. For example, if one expects non-normality, and one knows that the variance is likely to be large, then Sichel's estimator may be the most appropriate measure to use. Conversely, if one's dataset is not expected to have high variance and/or is a relatively large dataset for the type of data one is collecting, then the back-transformed mean with bias correction may be most appropriate. If very little is known about the nature of the system one is studying then the simple back-transformed mean or the raw median may be the most appropriate measure to use.

If the decision on which method of estimating the true population mean to use is made *a posteriori* then one risks choosing the most "palatable" or "acceptable" measure in terms of institutional memory when presenting one's data in reports and scientific publications. Nevertheless, it is up to the researcher to provide the rationale and justification for the method used in a manner that the target audience is able to understand. It is also the author's responsibility to explicitly state the nature of the distribution of the dataset, provide information on data transformation if any, and inform the reader/audience of the rationale for the method given. Additionally, presenting the uncertainty surrounding that value is paramount, as the audience needs to be able to put into context the information presented. Presenting a number with little or no indication of the uncertainty surrounding that number can be misleading.

Acknowledgements
This work was funded by the Foundation for Research, Science and Technology- Rural Futures programme (C10X0809) and AgResearch Core Funding. Helpful discussions were had with statisticians at AgResearch, Siva Ganesh, Neil Cox and John Koolaard. This dataset would not exist without the work of the Ballantrae Farm staff John Napier and Brad Gordon and the AgResearch Technical team of Barry Rolle, Frances Miller and Phil Theobald.

References

Clark I. 1987. Turning the tables – an interactive approach to the traditional estimation of reserves. Journal of the South African Institute of Mining and Metallurgy 87(10):293-306.

Hoogendoorn CJ, Bowatte S, Tillman RW. 2011. Simple models of carbon and nitrogen cycling in New Zealand hill country pastures: exploring impacts of intensification on soil C and N pools. New Zealand Journal of Agricultural Research 54(4): 221-249.

Lambert MG, Devantier BP, Nes P, Penny PE. 1985. Losses of nitrogen, phosphorus, and sediment in runoff from hill country under different fertiliser and grazing management regimes. New Zealand Journal of Agricultural Research 28:371-379.

Lilburne L, Carrick S, Webb T, Moir J. 2012. Computer evaluation of methods to sample nitrate leached from grazed pasture. Soil Use and management 28:19-26

Mashey J. 2004. War of the benchmark means: time for a truce. ACM Sigarch Computer Architecture News 32(4):1-13.

Olivier J, Johnson WD, Marshall GB. 2008. The logarithmic transformation and the geometric mean in reporting experimental IgE results: what are they and when and why to use them? Annals of Allergy Asthma Immunology 100(6):625-626.

Parfitt RL, Mackay AD, Ross DJ, Budding PJ. 2009. Effects of soil fertility on leaching losses of N, P and C in hill country. New Zealand Journal of Agricultural Research 52:69-80.

Saggar S, Mackay AD, Hedley MJ, Lambert, MG, Clark DA. 1990. A nutrient-transfer model to explain the fate of phosphorus and sulphur in a Grazed Hill-Country pasture. Agriculture, Ecosystems and Environment 30(3/4):295-315.

White RE, Haigh R, MacDuff JH. 1987. Frequency distributions and spatially dependent variability of ammonium and nitrate concentrations in soil under grazed and ungrazed grassland. Fertiliser Research 11:193-208.

Zhou X-H, Gao S. 1997. Confidence intervals for the log-normal mean. Statistics in Medicine 16:783-790.