On Establishing Reference Values

J. H. Lumsden and K. Mullen*

ABSTRACT

In order to establish a range of reference values for any characteristic one can use Gaussian or nonparametric techniques, whichever are most appropriate. One has the choice of calculating tolerance intervals or percentile intervals. A tolerance interval is said to contain, say 95% of the population with probability, say 0.90. A percentile interval simply calculates the values between which 95% of the observations fall. If the data can be said to have a Gaussian distribution, the same precision can be obtained with smaller sample sizes than using the nonparametric techniques. In some cases, data which are not Gaussian can be transformed into a Gaussian form and hence make use of the more efficient Gaussian techniques. In both cases, the data should be checked for outliers or rogue observations and these should be eliminated if the testing procedure fails to imply that they are an integral part of the data.

RÉSUMÉ

Lorsqu'on veut tracer une courbe de valeurs de référence, quelle que soit la caractéristique dont il s'agit, on peut utiliser les techniques non paramétriques ou celles de Gauss, en choisissant celles qui semblent les mieux appropriées. On a aussi le choix de calculer les intervalles de tolérance ou ceux des centiles. On dit qu'un intervalle de tolérance contient, v.g. 95% d'une population avec probabilité, v.g. 0.90. Un intervalle de centile calcule simplement les valeurs entre lesquelles se situent 95% des observations. Si on peut affirmer que les données possèdent une distribution équivalente à celle de Gauss, il est possible d'obtenir la même précision, avec un échantillonnage moins considérable, que donneraient les techniques non paramétriques. Dans certains cas on peut transformer des données de façon à les rendre équivalentes à celles de Gauss et de ce fait recourir à ses techniques, reconnues comme plus efficaces. Dans un cas comme dans l'autre, il faut vérifier les données et en éliminer les trompeuses, si la technique qu'on emploie ne les reconnaît pas comme partie intégrante de l'ensemble des données.

INTRODUCTION

A commonly recurring problem of a veterinary diagnostic laboratory is to establish reference values for a particular characteristic. By reference intervals one means a range of probable values of that characteristic for healthy animals. Values outside of this range are suggestive of a lack of good health. Generally one assumes that if the animals are healthy, then their values of the characteristic will have a particular distribution, whereas the values of nonhealthy animals will have another different distribution. This concept is somewhat limiting because although a healthy animal is one without disease, a nonhealthy one can occur in many different ways due, for example to many possible different diseases, each one presumably leading to another distribution of values. A clinician would prefer to know the probability of a certain disease being present given a test result in the referent value. Referent values vary by test, by population, by devise and by the predictive value desired. In human medicine where only one species is involved, determination of referent values is in the developmental phase. In veterinary medicine when there are many species and breeds, universal agreement on referent values must be regarded only as a future goal. It seems reasonable and necessary therefore at this time to at least try to establish the distribution for healthy animals with each methodology.

*Department of Pathology (Lumsden) and Department of Mathematics and Statistics (Mullen), University of Guelph, Guelph, Ontario N1G 2W1.

Submitted June 6, 1977.

The approach taken in this article is to establish the distribution of values for clinically healthy animals and from this to calculate reference intervals or a normal range. This normal range is usually a pair of numbers within which, for example, 95% of the values can be expected to lie. This implies that 5% or one in 20 healthy animals will have values outside the normal range. One could calculate a pair of numbers which contain 99% of the population, leaving only 1% of healthy animals outside of the range but naturally such a range would be wider than the 95% limits, making it more likely to include the values for unhealthy animals (false negatives). For this reason and because 95% limits are widely accepted in human medicine, we shall deal with limits which contain 95% of the healthy values.

The whole concept of reference values and the normal range is controversial and summarized by Henry and Reed (5). Despite the controversy however, there is still a definite need for a range of normal or reference values. This article discusses some of the more important approaches to its establishment and the attendant problems involved.

Gaussian Versus Nonparametric Ranges

Early in the use of statistical techniques to establish reference values, especially for human medicine, it was usual to assume that the sample data came from a Gaussian distribution. Later, after it was realized that many populations did not have a Gaussian shape, it was argued that it was unnecessary to make the Gaussian assumption and that nonparametric techniques were more than adequate for estimation purposes. Using the Gaussian approach one simply gave $\bar{x} \pm 2s$ as an approximate normal range and this rule was blindly applied to all data, regardless of whether they were Gaussian or not. The reaction away from Gaussian techniques was equally severe and some writers, notably Read *et al* (5), argue that the nonparametric techniques are always equally good. Although the results obtained in nonparametric cases could be applied to Gaussian cases as well, it would not be satisfactory to do so, since for the parametric cases, methods having greater efficiency can be devised by taking into account the available information regarding the functional form of the distribution (16). Further, as we shall discuss below, the sample sizes required to obtain adequate figures using nonparametric techniques are often much larger than those required assuming a Gaussian distribution. We shall argue that, when it is safe to assume a Gaussian distribution or when the data can be transformed to have a Gaussian distribution, then Gaussian techniques should be used. In other cases, the use of nonparametric techniques will be recommended.

The sequence of discussion is

(a) identifying the outliers and eliminating them.
(b) establishing whether the distribution is Gaussian or not.
(c) presenting one of the four following techniques for estimating the normal ranges:
  (i) Gaussian tolerance interval estimates
  (ii) Nonparametric tolerance interval estimates
  (iii) Gaussian percentile estimates
  (iv) Nonparametric percentile estimates.

For purposes of illustration we include three examples. One of these is data for bovine hemoglobin on 42 clinically healthy cattle in the age group two weeks to six months. The second is platelet counts on 41 cattle in the same age group. The third is serum iron measurements on 43 healthy calves in the age group one day to 14 days. The data are presented (from smallest to largest) in Tables Ia, Ib and Ic.

TABLE Ia. Hemoglobin Measurements (gm/dl) on 42 Healthy Cattle in the Age Group two weeks to six months (arranged in ascending order)

8.4	10.1	10.9	11.8	13.0
9.1	10.4	11.0	12.0	13.2
9.2	10.4	11.2	12.2	13.3
9.3	10.4	11.3	12.3	13.5
9.4	10.4	11.3	12.4	13.5
9.6	10.5	11.4	12.5	14.0
9.8	10.6	11.4	12.5	
9.9	10.8	11.5	12.8	
10.1	10.8	11.8	12.9	

Treatment of Outliers

If the number of observations is not large, then as we shall subsequently see,

TABLE Ib. Platelet Counts ($\times 10^3/\mu l$) no 41 Healthy Cattle in the Age Group two weeks to six months, (arranged in ascending order)

280	415	510	580	650	830
320	420	510	590	700	870
330	430	520	590	720	970
340	460	550	600	740	970
380	465	550	600	770	1000
380	500	560	630	800	1270
400	500	565	640	800	

TABLE Ic. Serum Iron Measurements (mg/dl) on 43 Healthy Calves in the Age Group one day to 14 days (arranged in ascending order)

27	45	63	78	102	156
28	45	64	82	106	161
35	47	65	83	121	164
37	50	67	95	127	177
38	50	69	95	135	193
45	52	75	100	136	200
45	53	77	102	142	224
					283

the nonparametric methods provide estimates which are functions of the few largest or smallest observations. Thus it is important that the experimenter satisfy himself that these observations are not contaminated by, for instance, blunders, technical or clerical errors or accidents. An outlier (or rogue observation), which is undetected and hence used in the calculation of the normal range, will in general cause that range to be wider than it should be and hence weaken the sensitivity of such a range as a predictor of unhealthiness. On the other hand one wishes to avoid the subjective elimination and discarding of data which do in fact belong to the healthy population. Since almost all criteria for outliers are based on an assumed underlying Gaussian distribution and since at this stage we have not tested to see if the data came from a Gaussian distribution, the experimenter is in a quandary. Henry and Reed (5) resolve this problem by avoiding the word outlier and ensuring simply that the data form a homogeneous group, using the ratio

$$\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$$

If this ratio is greater than 1/3, eliminate $x_{(n)}$.

If the data are believed *a priori* to be Gaussian, then below we discuss two of the more important tests for outliers and treat the data of Tables Ia, Ib and Ic with them. For purposes of notation, we denote our sample of observations as $x_1, x_2, \ldots x_n$ and the same sample but in ordered form (from smallest to largest as in Table Ia, Ib and Ic) as $x_{(1)}, x_{(2)}, \ldots x_{(n)}$. In general the smallest observation, $x_{(1)}$, or the largest observation, $x_{(n)}$, will be the suspected value. Henry and Reed (5) recommend Dixon's r statistics denoted by r_{10}, as follows:

(a) if $x_{(n)}$ is suspected, $r_{10} = \dfrac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$

(b) if $x_{(1)}$ is suspected, $r_{10} = \dfrac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}$

The table of critical values of r_{10} is given in reference (1). If r_{10} is larger than the critical value then the suspected value is eliminated.

A second test of outliers due to Grubbs (4), sometimes known as Grubbs T-statistic, which was shown by Ferguson (3) to have a greater probability of detecting true outliers is as follows:

(a) if $x_{(n)}$ is suspected, $T_n = \dfrac{x_{(n)} - \bar{x}}{s}$

(b) if $x_{(1)}$ is suspected, $T_1 = \dfrac{\bar{x} - x_{(1)}}{s}$

(where s is the sample standard deviation). One rejects the suspected value if T_n or T_1 is greater than its critical value given in reference (4).

For the data of Table Ia, if the first observation is suspected, then using Dixon's statistic, $r_{10} = (9.1\text{-}8.4)/(14.0\text{-}8.4) = 0.125$ which is smaller than the critical value and thus we do not eliminate the value 8.4. The same conclusion is reached with Grubbs T-statistic. For the data of Table Ib, if 1270 is suspected then from Dixon's statistic we get $r_{10} = (1270\text{-}1000)/(1270\text{-}280) = 0.27$ and we do not reject the value. Using Grubbs statistic however, with $\bar{x} = 602.56$, $s = 212.63$ $\left(\text{where } s = \sqrt{\dfrac{\Sigma \bar{x}_i^2 - n\bar{x}^2}{n-1}}\right)$, the value of T_n is 3.1389, which is larger than the critical value. Thus the value 1270 should be rejected as an outlier. If the largest (smallest) observation is rejected, repeat the process with the new largest (smallest) value.

In general, we recommend the application of Grubb's T-statistic at a low level of significance (α) so that good observations will rarely be rejected. For the data of Tables Ia, Ib and Ic, with the level of

significance of 1% we conclude:

Ia: no outlier (as seen later, these data are Gaussian).

Ib: 1270 is an outlier (as seen later, these data are Gaussian).

Ic: no outlier (as seen later, these data are not Gaussian).

Testing for Gaussian distribution

Perhaps the two most frequently used tests to ascertain that data came from a Gaussian distribution are

(a) The chi-squared goodness of fit test

(b) The Kolmogorov-Smirnov test.

In the chi-squared test, the data are changed into k classes, their observed and expected values for each class compared using the chi-squared distribution, whose degrees of freedom are adjusted to allow for the estimation of the mean and variance of the original population (see reference (12), chapter 9). The data in Table II have $\bar{x} = 585.875$, $s = 186.1894$, $X^2 = 2.8817$ which with 4-1-2 = 1 degrees of freedom is not significant at 5% level.

TABLE II. Frequency Distribution of the 40 Platelet Counts for Healthy Cattle in Age Group two Weeks to six Months

Class	Class limit	Class frequency	Expected Frequency
1	280 - 450	10	9.3404
2	450 - 620	16	13.4264
3	620 - 790	7	11.7146
4	790 - 960	4	4.5168
5	960 - 1130	3	0.9100
Total		40	39.9982

The second method to examine the data for Gaussian distribution is the Kolmogorov-Smirnov test. A discussion of this test is given in (6). The critical values of the test statistic D, when the mean and variance are unknown and must be estimated from the data, is given by Lilliefors (7). He gives the 10%, 5% and 1% significance points which are reproduced below (Table III).

The test statistic is

$$D = \max|S_n(x) - F(x)|$$

where $F(x)$ is the theoretical distribution function of x and $S_n(x)$ is the sample distribution function. ($S_n(x)$ is calculated as 1/42, 2/42, 3/42, etc. We illustrate the calculation of $F(x)$ for $x = 8.4$. The value of $\bar{x}$ and s for the data of Table Ia is 11.25 and 1.40 respectively. Thus the standardized normal value of $x = 8.4$ is $z = (8.4 - 11.25)/1.40 = -2.04$. The probability of a standardized normal value less than -2.04, for normal tables in reference (12) is 0.0207. Other values are found similarly). In Table IV we show the calculations of D for the data of Table Ia. Since in Table IV, the largest difference (marked with an asterisk) is less than the 5% critical value of .13671 ($.886/\sqrt{42}$), we conclude that the data do not depart significantly from the Gaussian shape. For the data of Table Ib, the value of D was calculated and

TABLE III.

Sample Size (n)	Level of significance (α) $\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
5	0.315	0.337	0.405
10	0.239	0.258	0.294
15	0.201	0.220	0.257
20	0.174	0.190	0.231
25	0.165	0.180	0.203
30	0.144	0.161	0.187
over 30	$0.805/\sqrt{n}$	$0.886/\sqrt{n}$	$1.031/\sqrt{n}$

found to be less than the 5% critical value of 0.14008 ($0.886/\sqrt{40}$). Thus we conclude that the original population is Gaussian. But for the data of Table Ic the largest value of D is greater than the 5% critical value of 0.13511 ($0.886/\sqrt{43}$) and hence we conclude that data of Table Ic is non-Gaussian. The observations declared as outliers are not considered while testing for normality.

Since the discriminating ability of the Kolmogorov-Smirnov test is generally greater than that of the chi-squared test (8) and also since the test statistic does not depend on the data's being grouped, we recommend the Kolmogorov-Smirnov test as the appropriate test of Gaussian distribution.

Gaussian Tolerance Intervals

If the data prove to come from a Gaussian distribution then we may calculate the tolerance interval (an interval which has probability 0.90 of containing 95% of the population) as follows. If L_1 and L_2 are the lower and upper limits of the

interval then

$$L_1 = \bar{x} - ks, L_2 = \bar{x} + ks$$

where values of k, which have been composed from the paper by Weissberg-Beatty (17) are given below for a few values of n.

Sample size n	K
10	3.0183
20	2.5642
30	2.4130
40	2.3336
50	2.2834
60	2.2485
70	2.2222
80	2.2018
90	2.1852
100	2.1716

For sample sizes other than shown here, one can obtain the exact values of k directly from reference (17) but good practical results (differing at worst in the second decimal place) can be obtained by linearly interpolating in the accompanying table. Thus for n = 38, the interpolated figure is 2.3495, whereas the exact figure is 2.3465. For the data of Table Ia, Ib and Ic we have

	n	$\bar{x}$	s	k	L_1	L_2
Table Ia	42	11.25	1.396	2.322	8.01	14.49
Table Ib	40	585.88	186.189	2.334	151.31	1020.45
Table Ic	43	96.26	58.657	2.316	− 39.59	232.11

Note that the data of Table Ic are not Gaussian and hence should not use the above technique. We have included it here to illustrate that the improper use of this technique leads to absurd results (a negative lower limit).

NonGaussian Tolerance Intervals

If the data prove to be nonGaussian, then the tolerance intervals are based on the ordered data, that is, on the data ranked from smallest to largest Wilks (18). Now however, the number of observations is much more important. If we call l_1 and l_n the smallest and the largest observation respectively, then we can say that the probability is 0.9 that 95% of the population lies between l_1 and l_n if n, the number of observations, is greater than 80. If the value of n is less than 80, then either the probability is less than 0.9 or the percentage of the population included is less than 95%. Because of the importance of the lowest and highest values, duplicate analysis for these values are recommended.

If one thinks that the smallest and largest observations are not very reliable, then he might prefer to use the second smallest and second largest (call these s_1 and s_n respectively). Then in order to be able to say that the probability is 0.9 that 95% of the population lies between s_1 and s_n, he needs a sample of at least 140 observations. For the data of Table Ia and Ib which are Gaussian, it is preferable to calculate the tolerance intervals using Gaussian techniques. For the data of Table Ic, we must use nonGaussian techniques. In that case, from the work and table of Somerville (15), the probability is only 0.636 that 95% of the population lies between l_1 (=27) and l_n (=283). This is a universal problem of nonparametric intervals, namely, that they require larger sample sizes than if Gaussian for similar coverage, or put another way, they give less coverage for equal sample sizes.

Gaussian Percentile Estimates

The p^{th} percentile P is the point on the distribution below which p percent of the observations lie. Our interest centres on the 2.5^{th} percentile and the 97.5^{th} percentile (having 95% of the distribution between them). If the distribution is Gaussian with known mean μ and known variance σ^2, then, calling L and U the 2.5^{th} and 97.5^{th} percentiles respectively,

$$L = \mu - 1.96\,\sigma$$

$$U = \mu + 1.96\sigma$$

Since μ and σ are unknown, they may be replaced by $\bar{x}$ and s. Because s is a biased

estimator of σ (see the appendix), then σ must be replaced by cs, where c is a number which depends on the sample size. If $\hat{L}$ and $\hat{U}$ are the estimates of L and U then

$$\left.\begin{aligned}\hat{L} &= \bar{x} - ks \\ \hat{U} &= \bar{x} + ks\end{aligned}\right\} \text{where } k = 1.96c .$$

Sahney (14) has shown how to calculate confidence interval estimates for U and L. 95% confidence intervals for U are given by:

$$U \pm ts\sqrt{\frac{1}{n} + k^2T}$$

where t is the upper 97.5 percentage point of the t-distribution with n-1 degrees of freedom, s is the sample estimate of the standard deviation and k and T are as given in the appendix.

For the data of Tables Ia and Ib we have the following results:

Table	L	U	UCL for L	LCL for L	UCL for U	LCL for U
Ia.....	8.500	14.000	9.251	7.744	14.751	13.249
Ib.....	218.598	953.153	321.209	115.987	1055.763	850.541

UCL = upper 95% confidence limit. LCL = 95% confidence limit. We see immediately that the estimates of the 2.5th and 97.5th percentiles are not the same as the 95% tolerance intervals and there is no reason why they should be. They are essentially two different approaches to the same problem. The tolerance interval establishes an interval which contains a prescribed portion of the population *with a definite probability,* 0,90, whereas the percentile estimates form an interval which contains the same portion of the population but with no probability attached. That is why when one calculates the percentile estimates, he needs also to calculate a confidence interval about them. In general the tolerance intervals will be further apart than the percentile estimates, but that is because they give more information, namely, the associated probability.

Although we are more interested in comparing Gaussian versus nonGaussian techniques in this paper than in recommending tolerance intervals or percentile estimates, we may note that tolerance intervals are probably preferable for the reason given in the previous paragraph, but that not much error will be incurred if percentile estimates are used.

NonParametric Percentile Estimates

If the data do not have a Gaussian distribution, then one estimates the population values of L and U using the sample 2.5th and 97.5th values.

To estimate the pth percentile one uses the (n+1)p/100 order statistic which is easily obtained from the given data. For p = 2.5 and n = 43 (as in Table Ic) (n+1)p/100 = 1.1. In Table Ic, the first order statistic is 27 and the second is 28. Linearly interpolating gives an estimate of the 1.1th to be 27.1. Similarly since the largest observation is 283 and the second largest is 224 and (n+1) (97.5)/100 = 41.9, the estimate of the 97.5 percentile is 277.1. For the data of Tables Ia, Ib and Ic the following results are obtained:

Table Ia: (8.453, 13.963)
Table Ib: (281.000, 999.250)
Table Ic: (27.1, 277.1)

Confidence intervals for the population percentiles have been calculated and given (5). In general for n less than 120, it is not possible to obtain two-sided confidence intervals. Thus for the data of Tables Ia, Ib and Ic no confidence intervals can be calculated. Henry and Reed consider a larger sample size example and use their table to calculate a 90% two-sided confidence interval for the nonparametric estimates of the 2.5 and 97.5 percentiles.

We see again, in the comparison of Gaussian versus nonparametric percentile estimation that the Gaussian methods give usable results for smaller sample sizes than the nonparametric methods. This alone is a major factor in favor of using the Gaussian methods if they are applicable.

Use of Transformations

If the evidence indicates that the data do not come from a Gaussian population,

two courses of action are open to us. One of these, the use of nonparametric techniques has already been discussed. A second approach is to transform the data in such a way that the resulting transformed data are

TABLE IV. Calculation of D, the Kolmogorov-Smirnov Statistics for the Data of Table Ia

	x	$S_n(x)$	F(x)	$\lvert\Delta\rvert = \lvert S_n(x) - F(x)\rvert$
1.	8.4	0.0238	0.0207	0.0031
2.	9.1	0.0476	0.0618	0.0142
3.	9.2	0.0714	0.0721	0.0007
4.	9.3	0.0952	0.0823	0.0129
5.	9.4	0.1190	0.0934	0.0256
6.	9.6	0.1429	0.1190	0.0239
7.	9.8	0.1667	0.1492	0.0175
8.	9.9	0.1905	0.1685	0.2200
9.	10.1	0.2143	0.2061	0.0082
10.	10.1	0.2381	0.2061	0.0320
11.	10.2	0.2619	0.2266	0.0353
12.	10.4	0.3857	0.2709	0.0148
13.	10.4	0.3095	0.2709	0.0386
14.	10.4	0.3333	0.2709	0.0624
15.	10.5	0.3571	0.2946	0.0625*
16.	10.6	0.3809	0.3228	0.0581
17.	10.8	0.4048	0.3745	0.0303
18.	10.8	0.4286	0.3745	0.0541
19.	10.9	0.4524	0.4013	0.0511
20.	11.0	0.4762	0.4286	0.0476
21.	11.2	0.5000	0.4840	0.0160
22.	11.3	0.5238	0.5159	0.0079
23.	11.3	0.5476	0.5159	0.0317
24.	11.4	0.5714	0.5438	0.0276
25.	11.4	0.5952	0.5438	0.0514
26.	11.5	0.6190	0.5714	0.0476
27.	11.8	0.6429	0.6517	0.0088
28.	11.8	0.6667	0.6517	0.0150
29.	12.0	0.6905	0.7054	0.0149
30.	12.2	0.7143	0.7517	0.0374
31.	12.3	0.7381	0.7734	0.0353
32.	12.4	0.7619	0.7939	0.0320
33.	12.5	0.7857	0.8133	0.0276
34.	12.8	0.8095	0.8133	0.0038
35.	12.8	0.8333	0.8665	0.0332
36.	12.9	0.8571	0.8810	0.0239
37.	13.0	0.8810	0.8943	0.0133
38.	13.2	0.9048	0.9177	0.0129
39.	13.3	0.9286	0.9278	0.0008
40.	13.5	0.9524	0.9463	0.0061
41.	13.5	0.9762	0 9463	0.0299
42.	14.0	1.0000	0.9750	0.0250

Critical value from Table III = 0.13671

Gaussian distributed (or approximately so). This being so, the tolerance limits or percentiles can be calculated for the transformed Gaussian data and by using the inverse of the transformation, the tolerance limits or percentiles can be found for the original data. This will be illustrated subsequently.

Before proceeding with the technique of transforming data, let us try to allay some of the misgivings that seem to accompany the transforming of data. Firstly, the use of transformations in the biological, chemical, medical and veterinary sciences is not new. Thus pH values use logarithms, while dilution levels in microbiological titrations are reciprocals.

In some cases the transformation needed to make the data Gaussian is known a priori. For instance data whose distribution is skewed to the left can be made Gaussian by taking the logarithm of each observation. In most cases, however the "proper" transformation is selected based on experience or trial and error.

Three common transformations are: (a) logarithmic (b) square root and (c) reciprocal. For data expressed as a percentage the arc-sine transformation may be appropriate.

We illustrate the use of transformation for the data of Table Ic, with the logarithmic transformation. First we show the transformed data (using the natural logarithm transformation), in ordered form.

The data of Table IC, transformed using the natural logarithm transformation

3.30	4.16	4.80
3.33	4.17	4.84
3.56	4.20	4.91
3.61	4.23	4.91
3.64	4.32	4.96
3.81	4.34	5.05
3.81	4.36	5.08
3.81	4.41	5.10
3.81	4.42	5.18
3.85	4.55	5.26
3.91	4.55	5.30
3.91	4.61	5.41
3.95	4.62	5.65
3.97	4.62	
4.14	4.66	

For these data $\bar{x} = 4.3972$, $s = 0.5904$, $T_n = 2.12$. In terms of the transformed data $ln\ 283 = 5.65$ is not considered to be an outlier. The Kolmogorov-Smirnov test leads us to conclude that the transformed data are Gaussian.

To calculate the Gaussian tolerance interval, $k = 2.319$.

$$x \pm ks = (3.028, 5.766)$$

The inverse transformation (e^x) gives as a 95% tolerance interval the values (20.66, 319.26).

The 2.5^{th} and 97.5^{th} percentile estimates are found by looking up k in the appendix ($k = 1.9718$ for $n = 43$) then

$$\hat{L} = 4.3972 - (1.9718)\,(.5904) = 3.23$$

$$\hat{U} = 4.3972 + (1.9718)\,(.5904) = 5.56$$

which when inverse-transformed give as percentile estimates (25.36, 260.17). For the original data the values of $\hat{U}$, $\hat{L}$ and their 95% confidence intervals are given below.

$\hat{L}$	$\hat{U}$	UCL for L	LCL for L	UCL for U	LCL for U
25.36	260.17	28.32	15.12	436.46	232.97

SUMMARY

In order to use data to establish reference values, care must be taken to ensure that the techniques being used are valid and appropriate. Thus one first tries to identify outliers which are then eliminated. Grubbs T-statistic is the suggested method here, although it does assume that the data are Gaussian. If this assumption is impossible to make then $x_{(n)}$ can be tested by comparing the ratio $(x_{(n)}-x_{(n-1)})/(x_{(n)}-x_{(1)})$ to 1/3.

Once the data have been purified of outliers, they can be tested for Gaussianness using the Kolmogorov-Smirnov test and compared to the significance values in Table III.

If the data are Gaussian, one can calculate tolerance intervals, or percentile estimates (using the values of k given in the appendix).

If the data are nonGaussian, one can calculate the tolerance intervals (using Somerville's tables ([15]), but note should be made that for small samples, the desired probability of 0.9 may be impossible to obtain, and for moderate samples, the values of $x_{(1)}$ and $x_{(n)}$ will be the indicated ones. If one uses nonparametric percentile estimates, then reference (5) will be needed to calculate confidence intervals for them and these cannot be calculated for sample sizes of less than 120.

An alternative method is to use a transformation to a Gaussian form and then use Gaussian tolerance intervals or percentile estimates.

APPENDIX

The 2.5th and 97.5th percentiles L and U of a Gaussian distribution are

$$L = \mu - 1.96\sigma$$

$$U = \mu + 1.96\sigma$$

If μ and σ are estimated by x and s, the sample mean and standard deviation in a sample of size n, then, writing E for expected value we know that $E\bar{x} = \mu$, $E(s) \neq \sigma$. In fact, we can calculate a number c, depending on n, the sample size, so that $E(cs) = \sigma$. Thus if L and U are unbiased estimates of L and U respectively, then

$$\hat{L} = \bar{x} - ks$$
$$\hat{U} = \bar{x} + ks \qquad \text{where } k = 1.96c$$

L and U are approximately Gaussian distributed for large sample sizes with expectations L and U and variance $\frac{\sigma^2}{n} + k^2\sigma^2T$ where T depends on the sample size n. [See reference [14]. If t is the value from the student's t-distribution exceeded with probability 0.025, then a 95% confidence interval for L is

$$\hat{L} \pm ts\sqrt{\frac{1}{n} + k^2T}$$

Values of c, k, T and t are given below

n	c	k	T	t
30	1.00866	1.97697	0.01709	2.045
35	1.00738	1.97446	0.01460	2.032
40	1.00643	1.97260	0.01274	2.023
45	1.00570	1.97117	0.01130	2.015

REFERENCES

1. **DIXON, W. J.** Processing data outliers. Biometrika 22: 74-79. 1953.
2. **FAULKENBERRY, G. D. and J. DALY.** Sample size for tolerance limits on a normal distribution. Technometrics 12: 813-821. 1970.
3. **FERGUSON, T. S.** On the Rejection of Outliers. Proceedings of the Fourth Berkely Symposium. pp. 253-287. 1961.
4. **GRUBBS, F. E.** Procedures for detecting outlying observations in samples. Technometrics 11: 1-21. 1969.
5. **HENRY, R. J. and A. H. REED.** Normal values and the use of laboratory results for the detection of disease. In Clinical Chemistry: Principles and Technics. R. J. Henry, D. C. Cannon, J. W. Winkelman, Editors. Edition 2: pp. 343-371. New York: Harper & Row. 1974.
6. **HOLLANDER, M. and D. A. WOLFE.** Nonparametric Statistical Methods. New York: Wiley Publications. 1973.
7. **LILLIEFORS, H. W.** On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J. Am. Statist. Ass. 62: 399 et seq. 1967.

8. **MASSEY, F. J.** The Kolmogorov-Smirnov test for goodness of fit. J. Am. Statist. Ass. **46**: 69-78. 1951.
9. **MITRA, S. K.** Tables for tolerance intervals for a normal population based on the sample mean and mean range. J. Am. Statist. Ass. **52**: 88-94. 1957.
10. **MURPHY, R. B.** Non-parametric tolerance limits. Ann. Math. Statist. **19**: 88-94. 1948.
11. **OWEN, D. B.** Factors for two-sided tolerance intervals. Technometrics **6**: 379 et seq. 1964.
12. **REMINGTON, R. D. and M. A. SCHORK.** Statistics with Applications to the Biological and Health Sciences. New Jersey: Prentice-Hall. 1970.
13. **RESNIKOFF, G. J.** Tables to facilitate the computation of percentage points of the non-normal t-distribution. Ann. Math. Statist. **33**: 580-583. 1962.
14. **SAHNEY, A.** Tolerance Intervals versus Percentile Estimates in Setting Normal Values. M.Sc. Project, University of Guelph. 1977.
15. **SOMERVILLE, P. N.** Tables for obtaining non-parametric tolerance limits. Ann. Math. Statist. **29**: 559-601. 1958.
16. **WALD, A. and J. WOLFOWITZ.** Tolerance limits for a normal distribution. Ann. Math. Statist. **17**: 208-215. 1946.
17. **WEISSBERG, A. and G. H. BEATTY.** Tables of tolerance limit factors for the normal distribution. Technometrics **2**: 483-500. 1960.
18. **WILKS, S. S.** Determination of sample sizes for setting tolerance limits. Ann. Math. Statist. **12**: 91-96. 1941.