Determination of Confidence Intervals in Non-normal Data: Application of the Bootstrap to Cocaine Concentration in Femoral Blood

Brigitte Desharnais[a,b,*], Félix Camirand Lemyre[c], Pascal Mireault[a], Cameron D. Skinner[b]

[a]*Department of Toxicology, Laboratoire de sciences judiciaires et de médecine légale*
1701 Parthenais Street, Montréal, Québec, Canada H2K 3S7
[b]*Department of Chemistry & Biochemistry, Concordia University*
7141 Sherbrooke Street West, Montréal, Québec, Canada H4B 1R6
[c]*Department of Mathematics, Université de Sherbrooke*
2500 Boulevard de l'Université, Sherbrooke, Québec, Canada J1K 2R1

Abstract

Calculating the confidence interval is a common procedure in data analysis and is readily obtained from normally distributed populations with the familiar $\bar{x} \pm (t \times s)/\sqrt{n}$ formula. However, when working with non-normally distributed data, determining the confidence interval is not as obvious. For this type of data, there are fewer references in the literature, and they are much less accessible. We describe, in simple language, the percentile and bias-corrected and accelerated variations of the bootstrap method to calculate confidence intervals. This method can be applied to a wide variety of parameters (mean, median, slope of a calibration curve, etc.) and is appropriate for normal and non-normal data sets. As a worked example, the confidence interval around the median concentration of cocaine in femoral blood is calculated using bootstrap techniques. The median of the non-toxic concentrations was 46.7 ng/mL with a 95% confidence interval of 23.9–85.8 ng/mL in the non-normally distributed set of 45 postmortem cases. This method should be used to lead to more statistically sound and accurate confidence intervals for non-normally distributed populations, such as reference values of therapeutic and toxic drug concentration, as well as situations of truncated concentration values near the limit of quantification or cutoff of a method.

*Author to whom correspondence should be addressed.
Email: brigitte.desharnais@msp.gouv.qc.ca

1. Introduction

Determination of population statistical parameters (e.g., μ and σ) are common goals in forensic toxicology calculations. Unfortunately, most often the toxicologist must settle for estimates of these parameters, such as confidence intervals, based on sample measurements (e.g., $\mu = \bar{x} \pm (t \times s)/\sqrt{n}$). A fundamental assumption of these parametric calculations is that the underlying population is normally distributed. Parametric calculations (μ and σ based on $\bar{x}$ and s) are incorrect when the data are non-normally distributed [1]. However, there are several situations in forensic toxicology where calculations with non-normal data are required. For example, the distribution of blood concentrations of a given drug in a population (e.g., blood concentrations of methamphetamine in driving under influence cases) is often log-normal [2]. Quantification results might also yield non-normal data at the lower limit of quantification (LOQ) or cutoff. Indeed, all values below the LOQ/cutoff are reported as "not quantified"; the distribution of values is therefore truncated and non-symmetrical. In these situations, resampling techniques provide a viable alternative to the traditional parametric approaches by making no assumptions about the underlying population distribution. This powerful statistical approach is widely accepted in the statistics community and is gaining acceptance in other disciplines [3, 4]. Unfortunately, the literature, and until recently the software, describing and performing these calculations has not been readily accessible to the Chemistry and Bioanalysis communities. In order to fill-in this gap, we will focus on one resampling technique called "the bootstrap" due to its simplicity, efficacy and appropriateness to the task of calculating confidence intervals. The method is then applied in the calculation of a confidence interval for the median concentration of the approximately log-normal distribution of cocaine concentrations in postmortem femoral blood in the 2012–2013 cases in the province of Québec (Canada).

2. Materials and methods

2.1. Selection of cases

All 45 postmortem cases in Québec where non-toxic concentrations of cocaine was found in the femoral blood in years 2012–2013 were selected for this analysis. Six cases with cocaine concentrations ranging from 494 to 2 880 ng/mL were determined to be intoxications by the forensic toxicologist and were not included in the analyzed data set. Femoral blood was collected by the forensic pathologist in a BD Vacutainer storage tube containing 100 mg of sodium fluoride and 20 mg of potassium oxalate from the femoral vein and stored at 4 °C.

2.2. LC–MS-MS quantification

Cocaine was spiked in bovine blood at concentrations of 5, 10, 50, 100, 500 and 1 000 ng/mL to determine the calibration function. Cocaine-D_3 was used as an internal standard and produced a quadratic, $1/x$ weighted, calibration model. Solid phase extraction (SPE) of the analytes was performed using Oasis cartridges (HLB 3cc, product WAT094226; Waters, Mississauga, ON, Canada). A volume of 2 mL of blood was extracted and reconstituted in 100 mL of 15:85 methanol:ammonium formate (10 mM) after evaporation of the eluate. An aliquot (5 mL) was injected on an Agilent 1200 HPLC equipped with an AB Sciex 4000 QTrap mass detector. A 25-min step/ramp gradient from 10 mM ammonium formate + 0.2% formic acid to methanol was run on an Agilent Zorbax Eclipse C18 column (100 × 2.1 mm, 3.5 μm). Quantitative analysis was performed with m/z transition 305.1/183.0 (^{13}C mass was used in Q1 to reduce the signal), whereas transition 304.1/91.0 was used to confirm the identity of the analyte. The ratio of area of the analyte peak to the area of the internal standard peak was used as the response to determine the calibration function.

2.3. Statistical calculations

2.3.1. Normality testing and distribution of concentration values

MATLAB (version 7.12.0 R2011a; The Mathworks, Natick, MA, USA) was used to perform all calculations. The script used for the data analysis is available, with extensive comments, in Supplementary Data S2. The reader is encouraged to reproduce the results using it as a reference. Publication figures were produced in Excel 2010 (Microsoft, Redmond, WA, USA).

Non-normal distribution of the cocaine blood concentrations was confirmed using the Kolmogorov–Smirnov test (`kstest` function) on the standardized concentration data (`zscore` function). The `zscore` function applies the transformation $X_T = {(X - \bar{x})}/{s}$ to every data point (X), where X_T is the corresponding standardized data point, $\bar{x}$ is the original data average and s is the original data standard deviation. The standardized data has a mean of 0 and a standard deviation of 1. The `kstest` function tests the standardized data against the null hypothesis that the data follow a standard normal distribution, as illustrated in Figure S3 of Supplementary Data S1. A $P < 0.05$ indicates that the data do not follow a normal distribution (95% confidence level).

A histogram of the cocaine concentrations was obtained using 7 bins (`hist` function) and plotted in Excel. A good rule of thumb for the production of histograms is that the number of bins should be equal to the square root of the

sample size, up to a number of 25 bins [5].

2.3.2. Percentile bootstrap and number of resamples

For readers unfamiliar with the resampling process, a detailed description is given in the Section 3. The medians of 2 000 resamples were obtained using the `bootstrp(nboot, bootfun, d1)` function. The function's arguments are `nboot`, the number of bootstrap calculations (resamplings), `bootfun`, the name of the function that will operate on each of the bootstrap resamples and `d1`, the (column) vector containing the cocaine concentration data (`Coc_FB`) from which each resample is drawn. In this case, the median was calculated, therefore `@median` was used as the `bootfun` argument. The @ symbol instructs MATLAB to treat the text (`median`) as a function call. Thus, the MATLAB command was written as: `result = bootstrp(2000, @median, Coc_FB)`. Thirty (30) different sets of 2 000 median values were obtained and used to produce a distribution histogram of the medians.

To study the effect of the number of resamples on the calculated confidence intervals, the `bootci` function was used as: `bootci(nboot, {bootfun, d1}, 'type', 'name')`. This function can take a series of arguments that refine the calculation, but here only the following were needed: `nboot`, `bootfun` and `d1`, which were described previously and the `'type'`, `'name'` pair which defines the type of bootstrap variation and was set to `'per'` (percentile; i.e., `'type'`, `'per'`). The number of resamples (`nboot`) was set sequentially at 50, 100, 500, 1 000 and 2 000. Thus, for a `nboot` of 2 000 the command was written as: `CI = bootci(2000, {@median, Coc_FB}, 'type', 'per')`. Calculation of confidence intervals was repeated 30 times for each `nboot` value. The α value was kept at the default 0.05 (95% confidence level), but could be changed, if desired, by adding an argument (`'alpha'`, alpha value) to the `bootci` function argument list.

2.3.3. Bootstrap variations

To study the effect of the variation of the bootstrap method used, confidence intervals were obtained with the `bootci` function `'type'` set to `'per'` (percentile) and `'bca'` (bias corrected and accelerated) variations at 2 000 resamples. Calculation of the confidence intervals was repeated 30 times for each bootstrap method.

3. Results and discussion

3.1. Normality testing and distribution of concentration values

The original set of concentrations is available in Table 1. The 45 postmortem cocaine cases follow a non-normal distribution as shown in Figure 1 and as confirmed by the Kolmogorov-Smirnov test ($P = 0.0084$). Visually, the distribution of the concentration values is far from a normal distribution and seems to be closer to a log-normal distribution, which is frequently encountered in this kind of forensic bioanalysis [2] and is representative of a non-normal data distribution. Expected histograms from a normal and a log-normal distribution are shown in Supplementary Data S1 for reference.

Table 1: Cocaine concentration (ng/mL) in femoral blood of 45 suspicious death cases

7.67, 10.3, 10.8, 13.8, 14.0, 14.0, 14.9, 15.0, 15.1, 16.8, 17.5, 17.8, 18, 19.4, 21.6, 23.1, 23.9,
27.5, 28.8, 36.0, 37.2, 40.5, 46.7, 49.3, 61.3, 79.0, 79.2, 82.0, 85.8, 88.6, 93.7, 95.1, 99.1, 118,
122, 128, 177, 243, 256, 287, 340, 349, 398, 420, 437

Figure 1: Histogram of the concentrations of cocaine in femoral blood
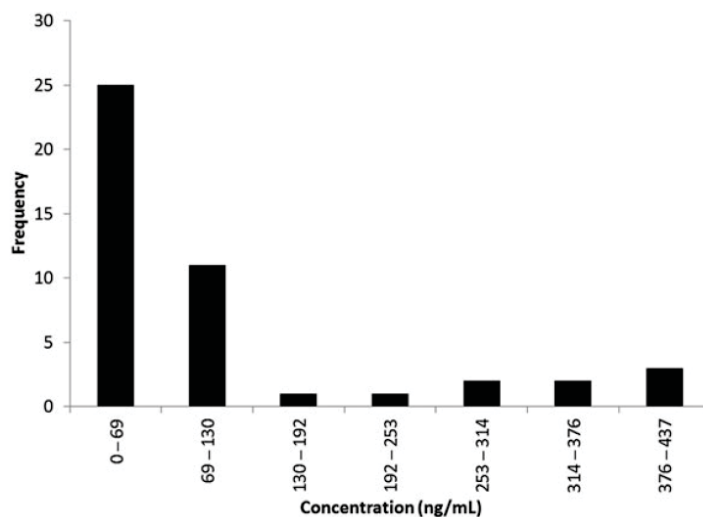
There are two widely used measures of central tendency. The mean is used with symmetrical data distributions whereas the median is more appropriate with skewed data distributions such as this data set. The median concentration of this

5

data was 46.7 ng/mL.

3.2. The bootstrap

The rigorous method of determining population parameters (i.e., μ, σ, median) is to sample the entire population (e.g., the cocaine concentrations in femoral blood) to allow direct calculation of the population parameter(s) and their distributions. Unfortunately, this technique is usually not feasible due to the size of the population. Instead, the population is sampled and the population parameters (μ, σ) are inferred from the sampled observations. From this sample, an estimator for each unknown parameter (e.g., $\bar{x}$ estimates μ, σ estimates s) is computed. This expediency usually requires making an assumption about the underlying distribution (i.e., normality). In cases where the underlying distribution is misspecified the inferred population parameters can be dramatically in error [6].

Making normality assumptions is only required because it is too costly to collect large sets of data from the population. In the bootstrap resampling statistical approach, the simple assumption is made that the collected sample data, having being drawn from the population, are the best available representative of the population. Since it is the best available representation of the population, any new set of data drawn from the sample (i.e., resampled data), and any parameter (B) calculated from the *resampled* data, is also characteristic of the population. If the sample set is resampled with replacement *nboot* times then a set of parameters $(B_1$ to $B_{nboot})$ can be "collected" very inexpensively. In the bootstrap method, each of the sets is collected using unbiased resampling where, conceptually, a value is drawn out of the sample data set at random, noted and replaced in the sample data set until a set equal in size to the original has been collected. Clearly, the make-up of the resampled sets is governed by random chance and consequently the calculated parameter set will take on a range of values with a sampling distribution similar to that of the population provided that *nboot* is large enough to allow the distribution to converge, that is, to reach a stable profile [7]. Notably, the bootstrap approach reproduces the distribution of the original population rather than making assumptions about it.

To utilize this set of parameters for confidence interval determination, the B_1 to B_{nboot} parameter values are ordered in ascending order and the bottom and top $(\alpha/2) \times nboot$ B values are determined [8]. For example, if 2 000 resamples are taken and the resulting 2 000 median values are ordered and labeled B_1 to B_{2000}, then the lower limit of the 95% confidence interval $(\alpha = 0.05)$ of the median sits at B_{51} (the first $(^{0.05}/2) \times 2000$ values are excluded) and the corresponding upper limit is at B_{1949}. The confidence interval thus includes 95% of the calculated

medians. A confidence interval for the median is interpreted in the same way as a confidence interval for the mean, i.e., 95% of the time the population median value lies within the confidence interval boundaries.

Bootstrap resampling was applied to the femoral blood cocaine concentration data. Thirty (30) sets of 2 000 resamples (*nboot*) were obtained, and the median was calculated for each of the resamples. A distribution of the 2 000 medians was obtained for each of the 30 sets; the overall distribution is summarized in Figure 2. The error bars (% coefficient of variation) illustrate the variability of the distributions of the 30 sets of medians. Notably, the relative standard deviations in the first and the last three bins corresponding to 16–28, 94–105, 105–116 and 116–128 ng/mL are large and went up to a maximum of 84%, whereas in the six central bins the relative standard deviation is always below 6%. This is a result of the random sampling process, which produces quite variable individual median values at the extremities of the median distribution. However, the values found at the confidence interval boundaries, $(\alpha/2) \times nboot$, are constant provided that *nboot* is large enough to eliminate these unstable values.

Figure 2: Distribution of medians for 2 000 resamples. Error bars represent the standard deviation of frequency over 30 bootstrap samples of 2 000 resamples each
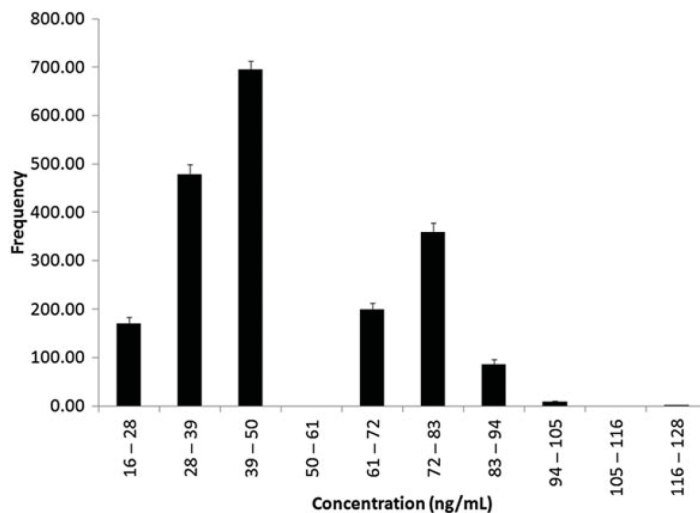
One point of debate is how many resamples are required to obtain reliable confidence intervals. A potential of n^n different resamples exist in a sample set of n values so the resampling process nearly always represents a very small fraction of this number. Just as in sampling from any population, there are no hard and fast rules. It takes relatively few measurements to establish the value of the

7

parameter but additional measurements are required to establish the distribution. When calculating 90–95% confidence intervals, it is generally agreed that the number of resamples should be between 1 000 and 2 000 [9]. By using many resamples it is possible to consistently determine the confidence interval boundaries without undue influence from outlier values. The weight of the extreme values is reduced by the sampling process, which is a big advantage of the bootstrap technique. This is illustrated by calculating 30 replicates of the confidence interval of the median for resample values of 50, 100, 500, 1 000 and 2 000 of the femoral blood cocaine concentration data. The results (Table 2) illustrate that with fewer resamples a larger range of values for confidence interval boundaries are obtained. In this case, the lower and upper boundary values obtained with 50 resamples show a range of 7.8 and 15.2 ng/mL, respectively. This is reduced to 0.8 and 2.8 ng/mL when 500 resamples or more are used. Values obtained for the boundaries therefore stabilize for $nboot > 500$ to form a confidence interval of 23.9–85.8 ng/mL around the data set's median, 46.7 ng/mL. The critical point of stabilization will vary from one data set to another, but as suggested by most practitioners [9], using 1 000 or 2 000 resamples to calculate the distribution of the median insures that the boundaries are not greatly influenced by an aberration in the sampling process.

Table 2: 95% confidence interval boundaries for different number of resamples

Number of resamples (n_{boot})	Median of the lower boundary (range)	Median of the upper boundary (range)
50	23.7 (7.8)	86.5 (15.2)
100	23.9 (4.4)	85.8 (11.7)
500	23.9 (0.8)	85.8 (2.8)
1,000	23.9 (0.8)	85.8 (2.8)
2,000	23.9 (0.8)	85.8 (2.8)

The median and range of the boundaries were obtained from 30 bootstrap replicates.

3.3. Bootstrap variations

The percentile bootstrap method described above is conceptually the simplest but is appropriate only when the distribution of the estimator of a parameter (median, mean, etc.) is normally distributed for the large sample size [8]. It is important to note that the distribution of a calculated estimator for a given parameter can be normal even though the underlying data set is not normal and vice versa. The bootstrap variation (percentile or other) is chosen according to the distribution of the calculated parameter. Typically, the samples means will be normally distributed as it follows the central limit theorem [1]. However, the

median distributions tend to be less symmetrical, as Figure 2 shows. If a notably skewed distribution of parameter is obtained, calculation of the confidence interval boundaries, namely the number of parameter values to remove at the tails of the distribution, must be modified. This version of the bootstrap technique is called the bias corrected and accelerated (BCA) method. It requires the calculation of a bias coefficient (z_0), which reflects the skew of the distribution, and an acceleration coefficient (α), which reflects the way variance changes across the distribution [10]. Calculation of these coefficients is too involved to be carried out manually but statistical programs such as MATLAB or R have confidence interval functions that utilize the BCA variation of the bootstrap method. Both methods were tested to estimate the confidence interval of cocaine in femoral blood. In this case, no difference was observed between the results obtained from the two variations. Both gave a 95% confidence interval of 23.9–85.8 ng/mL with ranges of values obtained for the lower and upper boundaries of 0.8 and 2.8 ng/mL, respectively. The more skewed the distribution of the parameters is, the larger the difference between the two methods is expected. As a precautionary measure, the use of the BCA bootstrap is encouraged, but the percentile method can often be employed, as evidenced here.

3.4. Bootstrap execution and applications

While the bootstrap approach is conceptually simple, execution is numerically intensive and beyond simple manual calculation. Bootstrap resampling and calculations can be performed using different software applications with varying degrees of ease. Probably the most readily accessible to forensic toxicologists is Excel, but its use is limited due to the lack of built-in bootstrap functions. Instead a variety of Excel functions can be used to resample the original data as extensively described, with examples, by Rochowicz [11]. A simpler option is to use Minitab (State College, PA, USA). Minitab is a very user-friendly statistics application with a suite of operations including some resampling functions, which can be used to perform percentile bootstrap as described by Moshonov [12]. Unfortunately, BCA bootstrap cannot be performed simply with either Excel or Minitab. Undoubtedly, the most efficient option is to use a powerful application like MATLAB or its open-source counterpart R, as these applications have bootstrap functions built-in and are full-fledged programing and data processing environments with a wide community of users and experts. It is therefore feasible to perform sophisticated calculations with a single command line, as shown in Section 2. In addition to the BCA and percentile methods used here, other related methods are available and easily explored and utilized.

While we have used bootstrap calculations to calculate the median and its

confidence interval on non-normally distributed data, this method can be used equally well on normally distributed data to produce results that are consistent with parametric statistics. This approach can also be extended to calculate other parameters, and their confidence intervals, such as the mean, standard deviation, variance, regression parameters with/without homoscedastic data, process capability, etc. [9]. As expected, to calculate these parameters, the sample data are resampled multiple times and the desired parameter (and its distribution) is calculated. Naturally, the bootstrap approach can also be extended to hypothesis testing and yields tests that are akin to Student's t-test, ANOVA, etc. that are especially useful when samples have unequal variances and/or non-normal distributions [13].

4. Conclusion

While this introduction has not been extensive, we hope that it is clear that bootstrap methods can be a valuable tool for data analysis in forensic toxicology, especially for non-normally distributed data where parametric methods fail. For example, the bootstrap could be applied to obtain more accurate description of drug use profiles, reference values for therapeutic and toxic concentrations and measured concentrations near the LOQ or cutoff of an analytical method. We would urge readers to investigate these methods so that they become more widely known, and used, in the community.

5. Supplementary data

Supplementary data are available at Journal of Analytical Toxicology online.

6. Acknowledgment

References

[1] D. Massart, B. Vandeginste, L. Buydens, S. De Jong, P. Lewi, J. Smeyers-Verbeke, The normal distribution, in: Handbook of Chemometrics and Qualimetrics: Part A, volume 20A of *Data Handling in Science and Technology*, Elsevier, Amsterdam, Netherlands, 1997, pp. 47–72.

[2] C. D. Klaassen, Casarett & Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill Education / Medical, New York, United States, 7 edition, 2007.

[3] H. Schäfer, An application of the bootstrap in clinical chemistry, in: P.-H. Jöckel, P. Rothe, P. Sendler (Eds.), Bootstrapping and Related Techniques, Springer, Berlin, Germany, 1992, pp. 213–217.

[4] F. Z. Sebestyén, Application of the Bootstrap Method for chemical data from the Boda Claystone Formation, Acta Geologica Hungarica 45 (2002) 207–217.

[5] D. Massart, B. Vandeginste, L. Buydens, S. De Jong, P. Lewi, J. Smeyers-Verbeke, Statistical description of the quality of processes and measurements, in: Handbook of Chemometrics and Qualimetrics: Part A, volume 20A of *Data Handling in Science and Technology*, Elsevier, Amsterdam, Netherlands, 1997, pp. 21–46.

[6] S. Litière, A. Alonso, G. Molenberghs, The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models, Statistics in Medicine 27 (2008) 3125–3144.

[7] D. Massart, B. Vandeginste, L. Buydens, S. De Jong, P. Lewi, J. Smeyers-Verbeke, Robust statistics, in: Handbook of Chemometrics and Qualimetrics: Part A, volume 20A of *Data Handling in Science and Technology*, Elsevier, Amsterdam, Netherlands, 1997, pp. 339–378.

[8] L. Simar, An Invitation to the Bootstrap: Panacea for Statistical Inference?, Institut de Statistique, Université Catholique de Louvain, Louvain, Belgique, 2008.

[9] J. Carpenter, J. Bithell, Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, Statistics in Medicine 19 (2000) 1141–1164.

[10] T. J. Diciccio, J. P. Romano, A review of bootstrap confidence intervals, Journal of the Royal Statistical Society: Series B (Methodological) 50 (1988) 338–354.

[11] J. Rochowicz, Bootstrapping analysis, inferential statistics and EXCEL, Electronic Journal of Spreadsheets in Education 4 (2010) 1–23.

[12] H. Moshonov, Bootstrap Confidence Intervals, University of Toronto (ND).

[13] N. I. Fisher, P. Hall, On bootstrap hypothesis testing, Australian Journal of Statistics 32 (1990) 177–190.