

How Many Samples Are Enough?

Introduction

Recently, I had the good fortune to be selected to go on a CIMM lecture tour across Canada to present a talk on quality control to geoscientists in the mining industry. After several of these talks, a number of geoscientists asked me the question: “how many samples are enough?”

Now, I have heard this question asked many times before, and am aware of the ‘conventional wisdom’ that a dataset of approximately 30 samples is generally considered to be ‘enough’ for geochemical applications. Consequently, although my immediate answer was “30”, the widespread interest in this question amongst geoscientists motivated me

to investigate the origins of this ‘rule of thumb’.

In the geological literature, this concept apparently dates back to a publication by Bob Garrett (1979). In that paper, Bob noted that the standard error of the mean is a function of the inverse of the square root of the number of samples. Consequently, he plotted n versus $1/\sqrt{n}$ and pointed out that the resulting curve basically ‘leveled out’ at approximately $n = 30$. As a result, Bob concluded that there is no real advantage to using more than 30 samples to estimate the mean, because no substantial further decrease in the estimation error of the mean can be achieved by such effort.

Clearly Bob’s analysis is highly practical. Unfortunately, the point at which Bob’s curve ‘levels out’ is a function of both the scale one plots the curve at, and the ‘eye of the beholder’, making use of such an approach rather arbitrary. Bob readily acknowledged this, and pointed out that the number of samples required “will depend on the particular situation” (Garrett 1979), a prudent caveat. Unfortunately, this hasn’t really prevented the widely quoted and employed “30 samples” from being used by probably the majority of geoscientists as anything but a hard and fast rule (i.e., immutable and inviolable) that should apply in all situations. Until now, I can’t say that I have been anything but guilty of employing this erstwhile ‘constant’ in an array of geochemical applications where its use may have been inappropriate. After I understood the reason for the conventional wisdom of “30 samples”, I was justifiably chagrined, and decided to see if I could derive a more comprehensive way to determine “how many samples are enough” that doesn’t have the inherent subjectivity of the present ‘conventional wisdom’.

In undertaking this task, it became immediately apparent that Bob Garrett (1979) was on the right track, as the standard error of a statistic provides information about the confidence one can place in an estimate of that statistic. However, experience shows that geoscientists are many times concerned with the relative error of a statistical estimate instead of the absolute error. As a result, to determine “how many samples are enough”, one should probably also consider the relative standard error of that statistic.

Because Bob Garrett’s suggested ‘rule of thumb’ (1979) is based on the formula of the standard error of the mean, the ‘30 samples’ that Bob advocated applies only to the estimate of a mean of a set of observations. It is likely that the estimation of other statistics (e.g., the standard deviation, correlation coefficient, etc.) would require different numbers of samples in order to obtain an adequate estimate. Fortunately, the mean is a statistic that is commonly investigated by geoscientists, and the formula for the standard error of the mean ($se_{\bar{x}}$) is:

$$se_{\bar{x}} = \frac{s_x}{\sqrt{n}} \quad (1)$$

where n is the number of samples and s_x is the standard deviation. Because the relative standard error of the mean ($rse_{\bar{x}}$) is merely the standard error of the mean divided by the mean, the relative standard error statistic is:

$$rse_{\bar{x}} = \frac{se_{\bar{x}}}{\bar{x}} = \frac{\frac{s_x}{\sqrt{n}}}{\bar{x}} = \frac{s_x}{\bar{x}\sqrt{n}} = \frac{CV}{\sqrt{n}}, \quad (2)$$

continued on page 4

How Many Samples Are Enough? continued from page 2

where CV is coefficient of variation, equal to the standard deviation divided by the mean ($= s_x / \bar{x}$).

Now, the standard approach to making decisions about the quality of an estimate of a statistic involves the use of formal inferential significance tests. Unfortunately, these commonly require assumptions about the distribution of the data (e.g., that the data are normally distributed). In my experience, geochemical data only very rarely exhibit distributions similar to ideal statistical distributions (e.g., Sinclair 1976). As a result, any method to determine “how many samples are enough” probably should not assume a statistical distribution for one’s data. This is particularly true given that when one asks such a question, one typically has, or can afford, only a small number of samples to begin with, and thus one has, and will continue to have, very little evidence to deduce what distribution the data actually derive from.

Consequently, a method to assess “how many samples are enough” should be ‘distribution-free’ in order to have maximum utility. Fortunately, derivation of the formula for the standard error of the mean, and thus the formula for the relative standard error of the mean, involves only first-principles calculus, whereby a first order Taylor polynomial expansion of the formula for the mean is used to propagate the errors in a group of measurements into the error in the mean of that group of measurements (Stanley 1990; Appen-

dix). Thus, the absolute and relative standard errors of the mean formulae (Equations 1 and 2) are not derived using any assumption about the distribution of the data. Consequently, these formulae are ‘distribution free’, and can be used to determine “how many samples are enough” for data derived from any and all distributions (e.g., normal, Poisson, binomial, hypergeometric, exponential, multi-modal, etc.).

An assessment of the literature demonstrates that all standard error formulae are inverse functions of the number of samples used in calculating the standard error of a statistic (e.g., Arsham, un-dated). Aside from indicating that as n increases, estimates of the mean become more precise, this functional relation makes it possible to determine “how many samples are enough” to obtain a reliable estimate of the mean using the absolute or relative standard error of the mean formula.

This can be done by first determining the tolerance level one is willing to accept in one’s estimate of the mean (e.g., an absolute error of 5 ppm [1 std. dev.] on the mean estimate of a set of trace element concentrations, or say 5% relative error on that mean estimate). By assigning this tolerance level to the absolute or relative standard error of the mean ($se_{\bar{x}}$ or $rse_{\bar{x}}$) in Equations 1 or 2, and then algebraically manipulating these equations such that they are expressed in terms of the number of samples, we obtain:

continued on page 5

Note: This EXPLORE article has been extracted from the original EXPLORE Newsletter. Therefore, page numbers may not be continuous and any advertisement has been masked.

How Many Samples Are Enough? continued from page 4

$$n = \left\{ \frac{s}{se_{\bar{x}}} \right\}^2, \quad (3)$$

for the absolute standard error of the mean, and:

$$n = \left\{ \frac{CV}{rse_{\bar{x}}} \right\}^2, \quad (4)$$

for the relative standard error of the mean.

Functional analysis of Equations 3 and 4 reveals that the number of samples required to obtain an adequate estimate of the mean is dependent on both the confidence level chosen ($se_{\bar{x}}$ or $rse_{\bar{x}}$), and either the absolute (s) or relative (CV) variation in the data. This result is intuitive, as samples from a distribution that exhibits little absolute or relative variation (i.e., with a low value of s or CV) will each be relatively good estimates of the mean of that distribution, and so only a small number of such samples will be necessary to obtain a good estimate of the mean. In contrast, samples from a distribution that exhibits high absolute or relative variation (i.e., with a high value of s or CV) will mostly be poor estimates of the mean, and so a much larger number of these samples will be required to obtain a mean estimate of adequate quality.

Figures 1 and 2 graphically illustrate how many samples are necessary to estimate the mean of measurements from a suite of samples at a variety of absolute and relative precision levels ($se_{\bar{x}}$ or $rse_{\bar{x}}$). Note that in both cases, the larger the absolute or relative variation in the data (s or CV), the more samples are necessary to obtain a given level of confidence in the estimate of the mean.

Unfortunately, determining “how many samples are enough” to estimate a mean requires prior knowledge of the absolute or relative variation in the data (s or CV). If no samples have been analyzed, this information will not be available. As a result, in order to estimate the mean of a suite of samples to a certain level of precision, one has to approach this problem iteratively. If one can make a ‘provisional’ estimate of the absolute or relative variation (s or CV), either using a guess, assumption, experience, samples that have already been analyzed, or results from historical datasets, and identify an acceptable tolerance level ($se_{\bar{x}}$ or $rse_{\bar{x}}$), then one can make a provisional estimate of “how many samples are enough” using Equations 3 or 4, respectively. Collecting and analyzing this provisional number of samples will provide (additional) data to make an (improved) estimate of the

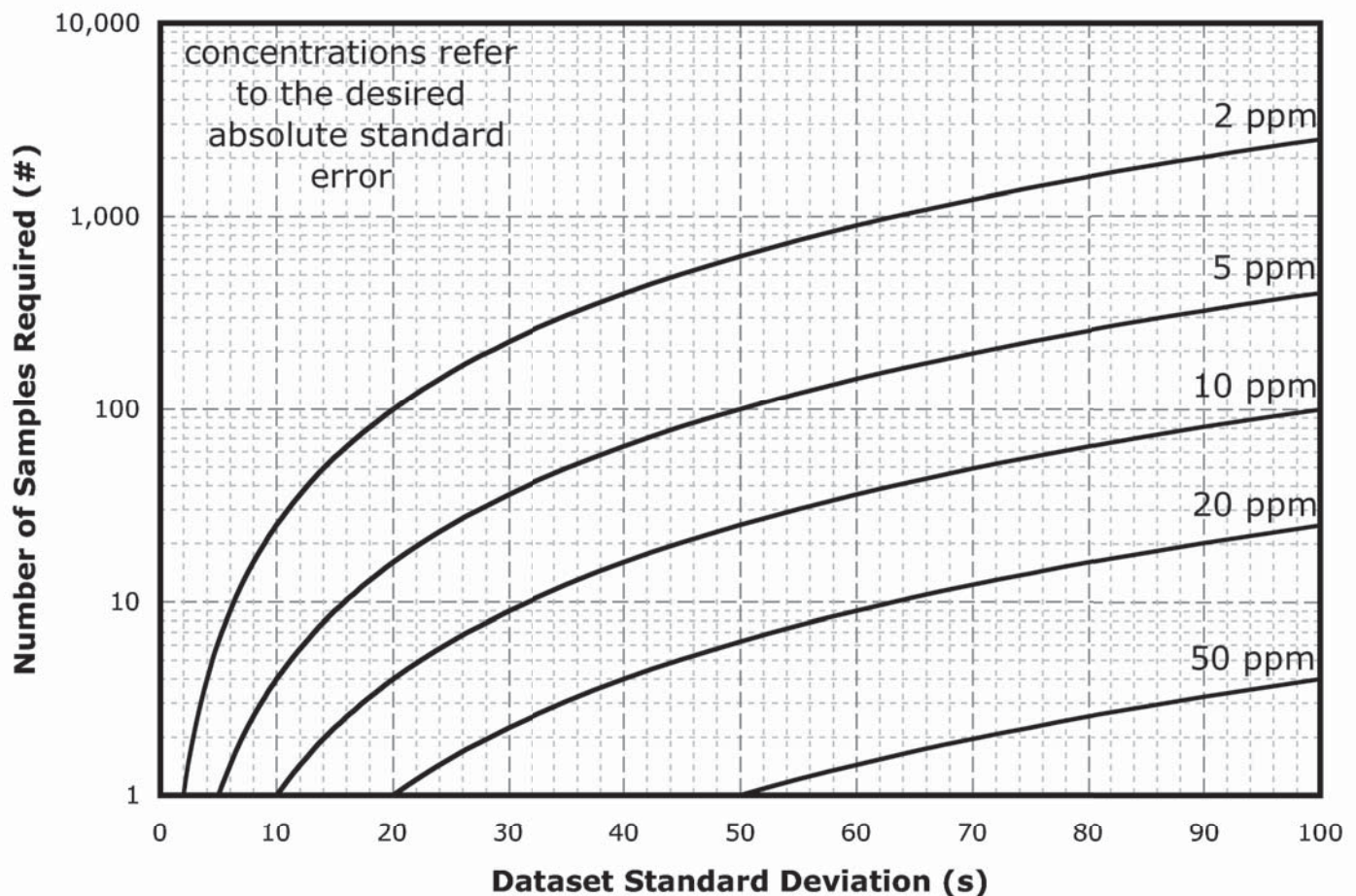


Figure 1. Number of samples (n) required to obtain an estimate of the mean with a desired level of absolute estimation error ($se_{\bar{x}}$). Note that the number of samples required is a function of the standard deviation (s), a measure of the variation in the measurements, and the number of samples (n), and thus is not, universally, 30.

continued on page 6

How Many Samples Are Enough? continued from page 5

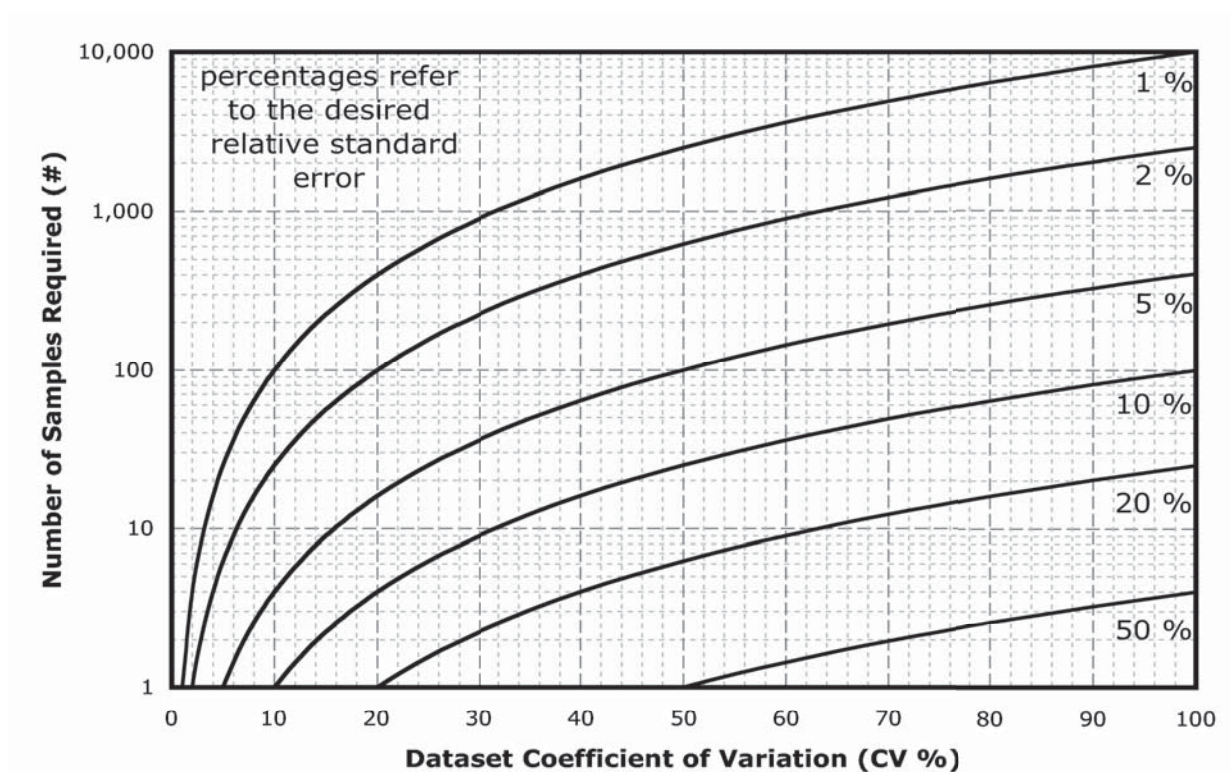


Figure 2. Number of samples (n) required to obtain an estimate of the mean with a desired level of relative estimation error ($rse_{\bar{x}}$). Note that the number of samples required is a function of the coefficient of variation (CV), a measure of the relative variation in the measurements, and the number of samples (n), and thus is also not, universally, 30.

continued on page 7

How Many Samples Are Enough? continued from page 6

absolute or relative variation (s or CV), and these can then be used to make a better estimate of “how many samples are enough”, again using Equations 3 or 4. Obviously, a poor initial estimate of the amount of variation that exists will lead to a poor estimate of the number of samples required, but because this procedure can be undertaken iteratively, better estimates of the absolute or relative variation can be obtained as more and more samples are collected, such that an adequate estimate of the mean, at whatever confidence level is required, can eventually be obtained.

The following example illustrates how the above approach can be used to determine “how many samples are enough.”

Case History

The North American Soil Geochemical Landscapes Project (NASGLP) is a tri-national program begun in 2006 that seeks to provide a consistent set of C-horizon soil geochemical data across North America (Natural Resources Canada 2008). Collection of soil samples at a density of one sample per $\sim 770 \text{ km}^2$ (Goodwin et al. 2009) on a random, stratified grid was conducted in Nova Scotia, Canada as part of a pilot study in 2007 and 2008. Analytical results ($< 63 \mu\text{m}$ size fraction, nitric-perchloric-hydrochloric-hydrofluoric acid digestion, ICP-MS finish) confirm that As concentrations in Nova Scotia are generally elevated relative to the Canadian Soil Quality Guideline (Maximum Allowable Concentration; MAC) of 12 ppm (Canadian Council of Ministers of the Environment 2009; *pers. comm.*, Terry Goodwin, Nova Scotia Department of Natural Resources), as concentrations range from 2.2 to 345.7 ppm (cf., the average As concentration in soils in Southern California is 1.5 ppm; $n = 1086$; Chernoff et al., un-dated).

The MAC's comprising the Canadian Soil Quality Guidelines have been developed to determine whether a soil concentration is at or below a level where no appreciable risk to human and environmental health is expected. If the maximum soil concentration at a site exceeds the MAC, a potential problem exists, requiring further study and possible remediation (Willis 2006).

Results from the NASGLP indicate that south of the Cobequid-Chedabucto Fault Zone (CCFZ) within the Meguma Terrane, 75 % of the samples have As concentrations that exceed the MAC (*pers. comm.* Goodwin 2010), a probable consequence of the abundance of arsenopyrite (up to 15 %) in Meguma Supergroup meta-sedimentary rocks comprising the basement to that terrane. In contrast, north of the CCFZ, only 25 % of the samples have As concentrations that exceed the MAC (*pers. comm.* Goodwin 2010), a proportion that is still high, probably because erosion products from the Meguma Supergroup were deposited in Carboniferous basins north of the CCFZ. These observations indicate that many uncontaminated sites in Nova Scotia would be considered ‘contaminated’ using the national MAC for As because of the naturally high background As concentrations that exist in the province.

As a result, it is probably appropriate to identify an alternative ‘made in Nova Scotia’ As concentration (*pers.*

comm. Terry Goodwin, NS DNR) for local use, in lieu of the Canadian MAC of 12 ppm. How this As concentration will be identified is not entirely clear, because MAC's are not merely a series of threshold separating background from anomalous concentrations, as they have been developed using “conservative and protective assumptions, and (risk assessment) principles/methods to ensure protection of both sensitive human and ecological receptors under typical conditions” (Willis 2006). As a result, the natural frequency distribution of As concentrations in soil is only one factor contributing to the assignment of an MAC; other factors include analytical geochemistry methods and detection limits; industrial production and method-of-use factors; physical and chemical properties; mineral and water speciation; exposure mechanisms; environmental fate and behavior in soils, waters, air, microbes, plants and animals; bio-availability and bio-accumulation; metabolic behavior; acute and chronic toxicity levels (e.g., LD_{50} s) in a variety of organisms (e.g., microbes, invertebrates, plants, livestock, wildlife, experimental animals, and humans); treatment and remediation ease, cost, and success probability; and benchmarks established by other jurisdictions (e.g., the US EPA). Unfortunately, the large number of factors that influence the assignment of an MAC make most MAC's appear to have been chosen subjectively because the ‘influence weights’ each factor is assigned are not known, and many of these factors are not quantitative.

Nevertheless, in order to establish an appropriate local MAC, the frequency distribution of As concentrations in soils needs to be understood so that the range of background concentrations can be identified using probability plot analysis (Sinclair 1976). Unfortunately, the only Nova Scotia-wide soil survey that has been undertaken is that of the NASGLP, and it provides only 72 samples (including 3 field duplicates; Goodwin et al. 2009). This is generally inadequate to characterize the distribution of As in Nova Scotia (with a landmass of $53,338 \text{ km}^2$), and so very many more samples will need to be collected and analyzed.

As a result, in 2010, the Nova Scotia Department of Natural Resources (NS DNR) intends to collect additional soil samples using the same sampling and analysis protocols as the NASGLP, in order to obtain a better understanding of the As concentrations in the province. This will create a

continued on page 8

How Many Samples Are Enough? continued from page 7

larger database with a higher sample density than the original NASGLP survey.

Because initial results indicate that soils developed over different lithologies in Nova Scotia exhibit different As concentrations, the NS DNR has divided the province into eight general but distinct lithozones (soils developed over different bedrock lithologies). As part of their effort to determine the frequency distribution of As concentrations in these lithozones, they intend to estimate a number of statistics that contribute to the understanding of the distribution of As concentrations in soils developed over each. These statistics will include the mean, standard deviation, and quantiles, among others, and it is hoped that at least a crude approximation of the frequency distribution of As concentrations in each lithozone can also be obtained from these data.

Two of the lithozones defined by the NS DNR are underlain by: (i) ‘granite’, including but not limited to the peraluminous South Mountain batholith, and (ii) ‘lower Paleozoic meta-greywacke and slate’ (the Meguma Supergroup). NASGLP individual soil sample As concentration results from these lithozones are presented in Table 1, along with their means, standard deviations and coefficients of variation (Goodwin *et al.* 2009). Although roughly similar numbers of NASGLP samples were collected and analyzed (17 and 15, respectively), the As concentration statistics for soils from these lithozones are significantly different.

Overall, As concentrations are high, with average concentrations of 13 and 47 ppm for granite and lower Paleozoic meta-greywacke and slate, respectively (Table 1). Furthermore, soils developed over lower Paleozoic meta-greywacke and slate exhibit more variable As concentrations than granite (standard deviations of 85 and 7 ppm, and coefficients of variation of 182 and 55 %, respectively; Table 1). Note that although these Meguma Supergroup rocks are known to contain As-bearing saddle reef gold deposits (Sangster & Smith 2007), all of the randomly located NASGLP samples ended up being collected away from known mineralization, so the higher As concentrations may be associated with yet-to-be discovered mineralization, or are merely samples from the high end of an elevated background As frequency distribution. In either case, these elevated As concentrations justify health and environmental concern. Additionally, the standard deviations and coefficients of variation of As concentrations over different lithozones indicate that a larger

Table 1. Arsenic concentrations (ppm) and statistics from soil samples collected as part of the NASGLP in Nova Scotia (Goodwin *et al.* 2009).

Samples & statistics	Over granite lithozone	Over lower Paleozoic meta-greywacke & slate lithozone
1	13.8	11.4
2	4.8	81.1
3	12.3	37.3
4	18.5	15.8
5	12.1	29.9
6	33.1	16.1
7	2.2	13.4
8	15.3	5.9
9	15.3	345.7
10	17.4	26.2
11	15.7	10.2
12	13.7	41.9
13	4.7	22.2
14	15.9	25.2
15	3.0	16.9
16	18.3	
17	10.8	
Mean	13.35	46.61
Standard Deviation	7.36	84.77
CV (%)	55.12	181.85

number of samples will be necessary to determine a reliable average As concentration in soils over lower Paleozoic meta-greywacke and slate, than in soils over granite.

Example Calculation

Fortunately, because at least some NASGLP As concentrations have already been measured from these two lithozones, we can use these data to provide ‘provisional’ estimates of the relative variations (CVs) of As concentrations in soils over these two lithozones (Table 1). These can then be used to obtain estimates of the number of samples (n) required to calculate mean As concentrations for each lithozone to a certain precision level. After collecting and analyzing n additional samples, better estimates of the relative variations (CVs) can then be made using this larger dataset, and these can be used to further refine/confirm the number of samples necessary to reliably estimate the mean As concentrations in the various lithozones. Obviously, this process can proceed iteratively until sufficiently reliable mean concentrations have been obtained.

How Many Samples Are Enough? continued from page 8

Table 2 illustrates the number of samples required to estimate mean As concentrations in each lithozone at various tolerances, based on the results from the initial NASGLP samples. For example, for the soils overlying the granite lithozone, the calculated sample size necessary to achieve a 10% tolerance level ($rse_{\bar{x}}$) is, coincidentally, $(55.12/10)^2 = 30.38$ samples. In contrast, the number of samples necessary to estimate the mean As concentration in soils over the lower Paleozoic meta-greywacke and slate lithozone is $(181.85/10)^2 = 330.69$. Obviously, the larger coefficient of variation for As concentrations in this lithozone necessitates the collection and analysis of far more samples.

Table 2. Calculated number of samples required to obtain estimates of the mean As concentration at various tolerance levels for the granite and lower Paleozoic meta-greywacke and slate lithozones of Nova Scotia.

Tolerance level (%)	# of samples over the granite lithozone	# of samples the lower Paleozoic meta-greywacke & slate lithozone
1	3038	33069
2	760	8267
5	122	1323
10	30	331
20	8	83
30	3	37
40	2	21
50	1	13
100	0	3

The results in Table 2 illustrate that significant differences exist between these two lithozones. The approximately 3 times larger CV, 181.85, observed in soil samples from the lower Paleozoic meta-greywacke and slate lithozone makes it necessary to collect and analyze approximately 10 times more samples to obtain the same relative standard error of the mean in each lithozone. Obviously, the relatively high variability in As concentrations in the lower Paleozoic meta-greywacke and slate lithozone makes estimating the mean more difficult. Nevertheless, using Equation 4 and the strategy described above to determine “how many samples are enough” will allow the NS DNR to estimate, in advance, how many samples should be collected to obtain adequate, or at least the best available, estimates of the average As soil concentration in all eight lithozones within Nova Scotia. Based on the results in Table 2, it is likely that the additional samples that they intend to collect will not be sufficient to obtain mean concentration estimates with highly acceptable errors for all eight lithozones. Nevertheless, this procedure will allow them to obtain the best estimates possible given their survey constraints.

Discussion

Given that elevated As concentrations represent a potential hazard in Nova Scotia, the concentrations of other elements measured in the NASGLP with crustally anomalous background geochemical concentrations should also be assessed to obtain other ‘made in Nova Scotia’ MAC’s, where appropriate (e.g., U, whose concentrations are known to also be elevated in Nova Scotia because of the abundance of per-aluminous granite, and the presence of Carboniferous basins that host sediments shed by these eroding intrusions). Such an analysis will provide additional estimates of the numbers of samples necessary to estimate the mean concentrations of other elements for each lithozone, and these numbers of samples are likely to be different than those calculated for As, above. As a result, before collecting additional samples in the follow-up survey, the NS DNR will have to balance their tolerance for mean concentration estimation error for each element in each lithozone with a number of other factors (e.g., bio-accumulation and bio-availability; toxicity; treatment and remediation ease, cost, and success probability; etc.) to determine how many samples to collect from each lithozone, so that they maximize the quality of the resulting data.

Now that we understand all of the factors that control “how many samples are enough”, let us consider the estimation error we would obtain in terms of relative standard error on the mean if we were to use the ‘conventional wisdom’ of 30 samples (Garrett 1979) to estimate the mean As concentration in the granite and lower Paleozoic meta-greywacke and slate lithozones, above. Using Equation 2 and the observed coefficients of variation in each lithozone, we can see that in soils over the lower Paleozoic meta-greywacke and slate lithozone, there will be about 33% error on the mean As concentration estimate, and in soils over the granite lithozone, there will be only about 10% error on the mean As concentration estimate. These precision levels are significantly different, and may or may not be adequate for the purposes described above, likely because “how many samples are enough” is a function of more than just the number of samples (Garrett 1979), as the variability of the data (in absolute or relative terms) and the desired absolute or relative standard error on the mean also figure prominently.

Conclusions

Formulae have been derived that allow determination of how many samples are necessary to obtain an ‘adequate’ estimate of the mean of sample measurements. These derivations do not assume that any specific data distribution exists, making this approach perfectly general and applicable to geochemical (and other) data collected by the mineral exploration/mining (and other) industries and disciplines.

The formulae derived to determine “how many samples are enough” do not support what has become the ‘conventional wisdom’ that 30 samples are adequate to obtain a good estimate of the mean, as other factors such as confidence level and variation in the underlying dataset also exert control. Rather, these results indicate that a quantitative strategy

continued on page 10

How Many Samples Are Enough? continued from page 9

can be used to more comprehensively address the unspecified “particular situations” that Bob Garrett alluded to in his 1979 paper when he first investigated the question: “how many samples are enough?”

Acknowledgements

Many thanks are extended to Terry Goodwin, geochemist for the Province of Nova Scotia Department of Natural Resources, for providing the data used in the example to illustrate how to determine “how many samples are enough”. This contribution also benefited from a helpful review by Dr. Robert Garrett of the Geological Survey of Canada.

References

- ARSHAM, H. (un-dated). Statistical Thinking for Managerial Decisions. University of Baltimore website, <http://home.ubalt.edu/ntsbarsh/business-stat/opre504.htm>, downloaded July 1, 2010.
- CANADIAN COUNCIL OF MINISTERS OF THE ENVIRONMENT. 2009. Canadian Environmental Quality Guidelines: Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health, Summary Tables, Update 7.0, Canadian Council of Ministers of the Environment website, http://www.ccme.ca/assets/pdf/rev_soil_summary_tbl_7.0_e.pdf, September 2007, downloaded July 1, 2010.
- CHERNOFF, G., BOSAN, W. & OUDIZ, D. (un-dated). Determination of a Southern California Regional Background Arsenic Concentration in Soil. California Department of Toxic Substances Control website, <http://www.dtsc.ca.gov/upload/Background-Arsenic.pdf>, downloaded July 1, 2010.
- GARRETT, R.G. 1979. Sampling considerations for regional geochemical surveys. In: Current Research, Part A. Geological Survey of Canada Paper, 79-1A, 197-205.
- GOODWIN, T.A., MCISAAC, E.M. & FRISKE, P.W.B. 2009. The North American Soil Geochemical Landscapes Project: Report on the 2008 Sampling Program and Preliminary 2007 Results. In: Mineral Resources Branch, Report of Activities 2008. Nova Scotia Department of Natural Resources, ME 2009-1, 45-51.
- NATURAL RESOURCES CANADA. 2008. Earth Sciences Sector, Environment and Health Program, North American Soil Geochemical Landscapes Project. Natural Resources Canada website, http://ess.nrcan.gc.ca/eh-esh/trinat/index_e.php, January 3, 2008, downloaded July 1, 2010.
- SANGSTER, A.L. & SMITH, P.K. 2007. Metallogenic summary of the Meguma gold deposits, Nova Scotia. In: GOODFELLOW, W.D. (ed) Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods. Geological Association of Canada, Mineral Deposits Division, Special Publication, 5, 723-732.
- SINCLAIR, A.J. 1976. Applications of Probability Graphs in Mineral Exploration. Association of Exploration Geochemists, Special Volume 4.
- STANLEY, C.R. 1990. Error propagation and regression on Pearce element ratio diagrams. In: J.K. RUSSELL & C.R. STANLEY (eds) Theory and Application of Pearce Element Ratios to Geochemical Data Analysis, Geological Association of Canada Short Course Notes, 8, 179-216.
- WILLIS, R. 2006. Scientific Basis of Environmental Quality Guidelines and Their Application at Contaminated Sites. Nova Scotia government website, <http://www.gov.ns.ca/nse/contaminatedsites/forum/docs/RobWillis.pdf>, April 5, 2006, downloaded July 1, 2010.

Clifford R. Stanley

*Dept. of Earth & Environmental Science,
Acadia University
Wolfville, Nova Scotia,
CANADA, B4P 2R6
(902) 585-1344 (office), (902) 670-0817 (mobile)
Email: cliff.stanley@acadiau.ca*

Appendix

Suppose that a statistic (f) is a function of n measurements of x (x_i), all of which have a common error (se_x). Provided that the x measurements are independent and identically distributed (iid), because they come from samples

continued on page 11

How Many Samples Are Enough? continued from page 10

of a single geological entity for which we are calculating a statistic, propagation of the errors in the x_i into f can be achieved using the following generative formula:

$$se_f^2 \cong \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 se_{x_i}^2, \quad (5)$$

(Stanley 1990).

In order to derive the standard error of the mean (in this case, our f), we propagate the error in x_i through the formula for the mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p x_i. \quad (6)$$

This requires the partial derivatives of the mean with respect to x_i :

$$\frac{\partial \bar{x}}{\partial x_i} = \frac{1}{n}. \quad (7)$$

Substituting these partial derivatives into Equation 5, using the variance of x () as the common variance in all x_i 's, and setting all covariances equal to zero, we obtain the well known formula for the standard error of the mean (Equation 1):

$$se_{\bar{x}} = \sqrt{\sum_{i=1}^n \left(\frac{1}{n} \right)^2 se_x^2} = \sqrt{n \left(\frac{1}{n^2} \right) s_x^2} = \frac{s_x}{\sqrt{n}}. \quad (8)$$

