

CL:AIRE

CONTAMINATED LAND: APPLICATIONS IN REAL ENVIRONMENTS



Chartered
Institute of
Environmental
Health

Guidance on Comparing Soil Contamination Data with a Critical Concentration

May 2008

The Chartered Institute of Environmental Health

CL:AIRE

with support from The Soil and Groundwater Technology Association

SAGTA

This *Guidance on Comparing Soil Contamination Data with a Critical Concentration* is published jointly by the Chartered Institute of Environmental Health (CIEH), Chadwick Court, 15 Hatfields, London, SE1 8DJ, tel: +44 (0)20 7928 6006 and CL:AIRE, 7th Floor, 1 Great Cumberland Place, London, W1H 7AL, tel: +44 (0)20 7258 5321. Downloadable copies may be obtained from the CIEH website www.cieh.org.uk and the CL:AIRE website www.claire.co.uk

The CIEH is a registered charity no. 290350. CL:AIRE is a registered charity no. 1075611.

The contents are copyright but permission is hereby given to reproduce them in part for non-commercial purposes provided appropriate acknowledgement is made.



This Guidance is endorsed for use in Scotland by the Royal Environmental Health Institute of Scotland.

Disclaimer: This guidance is offered in good faith but neither the publishers nor the authors can accept any liability whatsoever for any loss or damage arising from its interpretation or use nor should it be taken to represent the views of government. The guidance offers ways of incorporating statistical techniques into decision-making when assessing land contamination data but the techniques presented are not the only available methods. Whilst every effort has been made to provide a comprehensive and practical document, readers are advised to review the suitability of their chosen method and ensure they are comfortable with the relevant statistical principles, or to seek further advice if necessary, before applying this guidance.

May 2008

ISBN10 1-904306-79-9
ISBN13 978-1-904306-79-5

Preface

This document provides advice on the use of statistical techniques in the assessment of soil contamination data. It forms one part of the package of improved UK guidance highlighted in Defra's discussion paper *Assessing risks from land contamination – a proportionate approach. Soil guideline values: the Way Forward* (CLAN 06/2006) on the role and use of Soil Guideline Values (SGV) for managing the risks associated with soil contamination. Its overall aim is to increase understanding amongst stakeholders of the role that statistics can play in quantifying the uncertainty attached to estimates of the mean concentration of contaminants in soil thereby creating a more informed basis for regulatory decision-making.

The guidance presents a structured process that users can follow when employing statistical techniques for data assessment purposes. The process directs users to (and reinforces the importance of) other relevant guidance on sampling and data assessment and encourages practitioners to undertake appropriate scrutiny and organisation of data in preparation for statistical testing.

It is expected that the guidance will be useful to all those with an interest in or responsibility for land contamination. Note, however, that the statistical procedures set out in the document are drafted from the perspective of a regulator operating under either the planning system or Part 2A of the Environmental Protection Act 1990.

Acknowledgements

The guidance was prepared by an expert group working under the direction of Judith Lowe, then Defra's technical policy adviser. Other members of the working group and authors of the note were:

- Dr Mary Harris (formerly of Bureau Veritas, now an independent consultant)
- Paolo Masi (formerly of ES International and now of Foster Wheeler Italiana, Environment Division)
- Mathew Worboys (Atkins Environment)
- Peter Witherington (RSK Group plc)

Lucy Thomas and Tom Knowles (RSK Group plc) and Tim Wright (Atkins Environment) assisted the group by reviewing draft text and helping in the preparation of worked examples.

The authors would like to express their thanks for comments on drafts of the note provided by the following people and organisations:

Dr John Heathcote (Dounreay Site Restoration Ltd)
On behalf of the Soil and Groundwater Technology Association

Nick Barber (Nexia Solutions Ltd)
On behalf of the Soil and Groundwater Technology Association

Rob Ivens
On behalf of Mole Valley Borough Council

Simon Firth (WorleyParsons Komex) and Mr Paul Sheehan (Parsons Brinkerhoff)
On behalf of the Environmental Industries Commission Ltd

Steve Moreby (Gloucester City Council)
On behalf of the Chartered Institute of Environmental Health

Comments were also provided by the Environment Agency.

Core funding for preparation of the guidance note was provided by Defra and industry via members of the Soil and Groundwater Technology Association. Due acknowledgement is also given for the further financial support provided by Magnox Electric Ltd on behalf of the Nuclear Decommissioning Authority.

The project was administered by CL:AIRE.

Preparation for publication was undertaken by the Chartered Institute of Environmental Health.

Contents

| | | | |
|--|-----------|-------------------|--|
| Summary | 6 | Appendices | 45 |
| 1.0 Background | 10 | Appendix A | Statistical tables 46 |
| 1.1 Purpose and scope of the guidance | 10 | Appendix B | Directions on the use of one type of outlier test 49 |
| 1.2 Target audience | 11 | Appendix C | Directions for checking the normality of data 51 |
| 2.0 Structure of the guidance | 12 | Appendix D | Worked examples 58 |
| 3.0 Introduction to basic statistical principles | 14 | | |
| 3.1 Background | 14 | | |
| 3.2 Statistical distribution of data | 15 | | |
| 3.3 Null and alternative hypotheses | 16 | | |
| 4.0 Key aspects of designing sampling strategies | 18 | | |
| 4.1 Basic principles | 18 | | |
| 4.2 The importance of the conceptual model | 20 | | |
| 4.3 Unbiased sampling data | 20 | | |
| 5.0 Reviewing data and creating appropriate datasets | 23 | | |
| 5.1 Introduction | 23 | | |
| 5.2 Reviewing data quality | 23 | | |
| 5.3 Creating (an) appropriate dataset(s) | 23 | | |
| 6.0 Procedures for the application of statistical tests | 27 | | |
| 6.1 Introduction | 27 | | |
| 6.2 Defining the null and alternative hypotheses | 27 | | |
| 6.3 Selecting an appropriate statistical test | 29 | | |
| 6.4 Applying (an) appropriate statistical test(s) | 33 | | |
| 6.5 Worked examples | 40 | | |
| 7.0 Documenting decisions | 41 | | |
| References | 42 | | |
| Key statistical symbols and terms | 43 | | |

Summary

Purpose and scope

People sometimes regard 'statistics' as a complex topic that can be difficult to apply to practical problems and may lead to confusion, rather than clarity, about the key issues. As is in any other field, statistics also has its own set of conventions and terminology which may be difficult for non-statisticians to deal with initially, although everyone accepts that statistical 'rules' have to be respected to obtain valid results.

However, statistics can offer a potentially powerful way of objectively evaluating the evidence for and against particular propositions. For example, a statistical approach has the useful attribute of enabling decision-makers to reach conclusions about the available evidence with at least some understanding of the chances that they may be wrong.

One of the aims of this guidance is to overcome some of the negative perceptions about statistics by providing sufficient statistical background to allow land contamination specialists to use some statistical techniques competently in just one important area of assessment, that is, when comparing measured concentrations of contaminants in soil against some, user defined, 'critical concentration' or indicator of risk.

The guidance approaches this in the contexts of the two main reasons for such assessments:

- to establish whether land is suitable for a new use under the land use planning system;
- to determine whether land falls within the scope of Part 2A of the Environmental Protection Act 1990.

The main body of the guidance contains background information on sampling and statistics so that any data obtained with the intention of using statistical techniques will be suitable for that purpose, and so assessors will know in broad terms why and how statistical tests are structured in the way described. The guidance also contains

step-by-step procedures on the mechanics of carrying out the relevant statistical methods. This information should help users to understand and apply the methods in a robust way.

The main purpose of this non-technical summary is to highlight the core concept behind the guidance, and to encourage users to persevere with some of the more challenging statistical content of the guidance itself.

Core concept

The main thing to bear in mind when reading and using the guidance presented here is that the assessment of land under the planning system and Part 2A starts from two different perspectives.

In planning, the key question will usually be along the lines of "can we confidently say that the level of contamination on this land is low relative to some appropriate measure of risk?"

The key question under Part 2A will usually be "can we confidently say that the level of contamination on this land is high relative to some appropriate measure of risk?"

In statistical terms, these types of question are handled through the use of formal hypotheses – the Null Hypothesis and the Alternative Hypothesis.

Statistical tests are structured so as to be able to show (with a defined level of confidence) which of the two hypotheses is most likely to be true in any particular case. Note, however, that the outcome of testing is always expressed in terms of whether the Null Hypothesis can be rejected.

By convention, the Null Hypothesis is the starting proposition against which the key question (as expressed by the Alternative Hypothesis) can be tested. The confidence level of the test relates to the degree of confidence with which the Null Hypothesis can be rejected. It follows that if the Null Hypothesis can be rejected with a high degree of confidence, the assessor can be

fairly safe in concluding that the Alternative Hypothesis (the one the assessor is really interested in) is the correct one.

Under the land use planning system, where the aim is to demonstrate 'suitability for use':

- the Null Hypothesis is that the level of contamination [in the land of interest] **is the same as, or higher than**, the critical concentration; and
- the Alternative Hypothesis is that the level of contamination **is lower than** the critical concentration.

If the statistical test shows that the Null Hypothesis should be rejected, the assessor can conclude that the Alternative Hypothesis is more likely to be true, i.e. that contaminant concentrations are low relative to the critical concentration and that, potentially, the land is suitable for use. If the test shows that the Null Hypothesis should not be rejected, the assessor should conclude that contaminant concentrations may be the same as, or higher than, the critical concentration and further measures may be needed.

Under Part 2A, precisely the opposite set of propositions is in play. Thus under Part 2A:

- the Null Hypothesis is that the level of contamination in the land of interest **is the same as or lower than** the critical concentration; and
- the Alternative Hypothesis is that the level of contamination **is higher than** the critical concentration.

In this case, if the Null Hypothesis is rejected, the assessor can conclude that contaminant concentrations are high relative to the critical concentration (and a basis for intervention under the regime potentially exists). If the Null Hypothesis is not rejected, contamination levels may be the same as, or lower than, the critical concentration and a basis for regulation is unlikely to be available.

In effect, therefore, statistical tests explore two very different relationships between site derived measurements of contaminant concentrations and the critical concentration, as can be seen by Figure X (page 8).

In Figure X, sample data are shown in the form of a frequency distribution where sample results conform to a normal distribution. The 'x' axis represents contaminant concentrations and the 'y' axis the frequency at which particular concentrations occur within the sample data.

It can be seen that in the planning scenario, the statistical test is all about showing that there is a 95% probability that the true population mean falls below the critical concentration. In practice, the test involves comparing a quantity known as the 95th Upper Confidence Limit (UCL) of the true population mean with the critical concentration and deciding whether the difference is due simply to chance or is a [statistically] significant difference.

Under Part 2A, the opposite situation applies. Here, the test is all about showing that there is a 95% probability that the true population mean falls above the critical concentration and the test involves comparing the 95th Lower Confidence Limit (LCL) of the true population mean with the critical concentration.

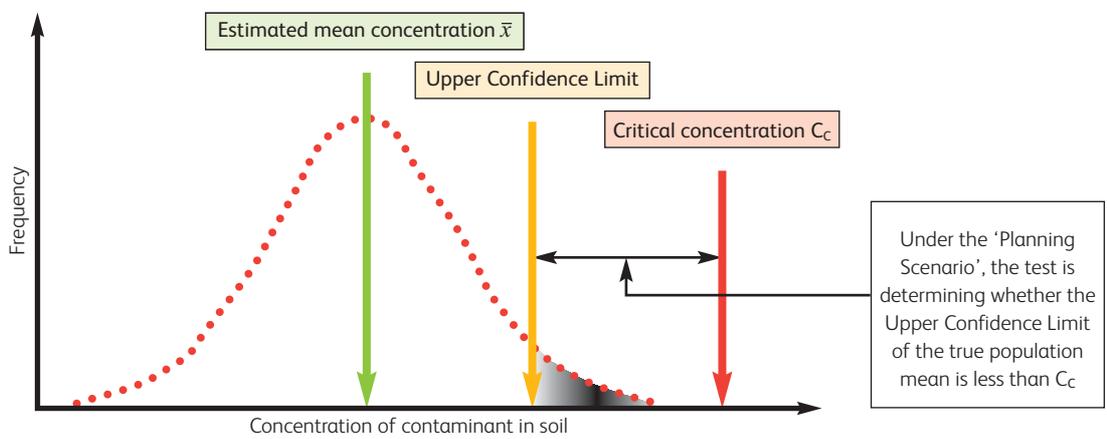
One other, potentially very useful, feature of the statistical techniques described here is that particular inferences can be drawn from the relationship between the sample mean and critical concentration which allow quick and easy screening of datasets and associated contaminants.

For example, in the planning scenario, if the sample mean is higher than the critical concentration it is clear without further calculation that the UCL of the true population mean will be higher than the critical concentration. In these circumstances it is not possible to reject the Null Hypothesis. This means the assessor can immediately identify which datasets (and associated contaminants) are unlikely to meet the planning test so that appropriate decisions and actions in relation to these data/contaminants, such as further data collection or precautionary remediation, can be taken at an early stage.

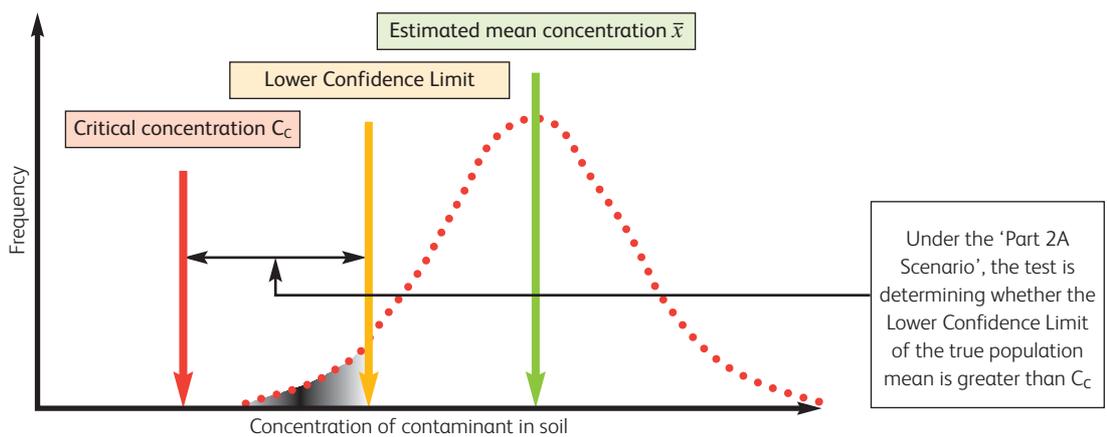
Figure X Relationship between sample data and critical concentration

Where data are normally distributed*

Planning Scenario



Part 2A Scenario



*The bell shaped curves in this figure correspond to the probability density functions of the estimated mean

Similarly, for the Part 2A scenario, if the sample mean is less than the critical concentration, the LCL must also be lower than the critical concentration and the Null Hypothesis cannot be rejected. This means that the assessor can put to one side datasets (and associated contaminants) where Part 2A is unlikely to apply, and concentrate on those more likely to fall within the remit of the regime.

Note that there will be occasions in both planning and Part 2A where no clear distinction exists between the sample data and the critical concentration, and it is not possible to reject the Null Hypothesis. In other words, it is not possible to say in one case that concentrations are clearly low or, in another, that concentrations are clearly high. In these cases, other decisions and actions have to be taken.

For a developer working in the planning system, these other decisions and actions may involve collecting further data to show clearly that contaminant concentrations are low, or undertaking remediation as a precautionary measure because it is easier, quicker and cheaper to do this than to carry out more detailed site investigation and risk assessment.

For regulators working under Part 2A, an inability to show that contaminant concentrations are high may mean collecting further sample data to demonstrate that a basis for regulation exists. Alternatively, and especially where a reasonable amount of work has been done to characterise the land, a regulator may decide that the evidence does not support intervention under Part 2A and it is better to direct available resources towards another priority site.

Other issues

There are, of course, a number of other issues which have to be considered when using statistics to help make decisions about the condition of land. These include:

- whether the data themselves are up to the task, i.e. whether there are sufficient, representative data available to apply statistical tests in a meaningful way bearing in mind that **no amount of statistical testing can compensate for insufficient, irrelevant or otherwise flawed data;**
- what statistical methods should be used, for example where the assumption of normality does not apply, and;
- the confidence level at which statistical tests should be conducted.

These, and related matters, are considered in more detail in the main text below.

There is also the wider question of the degree of risk represented by the critical concentration used in the tests. This is an important subject but not one that is covered at length in this document since other specific advice on this is available.

In preparing this guidance, and given the expected readership, the authors have attempted to keep the statistical material within reasonable bounds whilst maintaining technical accuracy. Nevertheless the authors are conscious that some users may feel overwhelmed by some of the more detailed statistical content.

In these circumstances, the authors can only encourage users to bear with the statistics because the potential prize – clear and defensible decisions based firmly on the evidence – is definitely worth the effort.

1.0 Background

This guidance is the result of one of a series of actions that were initiated by the Soil Guideline Value (SGV) Taskforce and subsequently promoted by Defra with the aim of improving UK guidance on the role and use of Soil Guideline Values in managing land contamination.¹

The note focuses specifically on the analysis of soil contamination data using statistical techniques. It was prompted by concerns expressed by participants at a SGVTF sponsored workshop held in London on 27th March 2006 about the adequacy of current guidance on the statistical testing of land contamination data and the way that it is applied in practice.

The following particular issues were highlighted:

- the fragmented nature of current guidance in this area which is considered difficult to relate to the practical procedures normally used to collect and evaluate data for risk management purposes;
- a lack of knowledge about established statistical conventions leading to the application of statistical tests to inappropriate data or, alternatively, the use of unsuitable statistical tests where the data themselves may be valid;
- the potential for misuse of the then current key guidance on this topic (CLR7, 2002)² because it does not prompt assessors to ask the correct questions about the purpose of statistical testing in particular applications, with the result that false conclusions may be drawn about the condition of land and inappropriate decisions be made on how the land should be handled.

1.1 PURPOSE AND SCOPE OF THE GUIDANCE

This note considers the role that statistical analysis can play in the assessment of land contamination data, the conditions that have to be met if a statistical approach is to be adopted and the techniques that are appropriate for use in particular cases.

The guidance is structured around a decision flow chart which describes how data collection, data review and statistical testing interact to produce defensible conclusions about the condition of land. The flow chart shows the decision logic for managing land contamination in the two key UK regulatory regimes of the land use planning system and Part 2A of the Environmental Protection Act 1990 (the 'EPA').

The guidance aims to improve the way in which statistical testing is applied to land contamination data by:

- explaining in simple terms the scientific basis for statistical testing so that appropriate data are collected in the field;
- encouraging the appropriate scrutiny and organisation of data into meaningful datasets in preparation for statistical testing;
- ensuring that the correct statistical questions are asked about the available evidence on the condition of land (that is, the data on contaminant concentrations in soil) so that appropriate inferences are drawn from the results of testing.

The guidance focuses on soil contamination data of the type that is routinely gathered to support human health risk assessment projects. Subject to the statistical validity of the sampling strategies used, the same principles could be applied to other, similar, types of data (such as leachates prepared from soil samples) and to support related parts of the risk management process (for example, during or on completion of remediation).

The guidance does not address in detail the overall process of assessing and managing the risks that may be associated with land contamination since this is already set out in the Model Procedures for the Management of Land Contamination, CLR11.³ Nor does the guidance cover the particular activities that have to be carried out to generate factual data and to

¹ Defra, *Assessing risks from land contamination – a proportionate approach. Soil Guideline values: the Way Forward. CLAN 06/2006*, November 2006

² Defra/Environment Agency, *Overview of the development of guideline values and related research, CLR7, 2002*

³ Defra/Environment Agency, *Model Procedures for the Management of Land Contamination, CLR11, 2004*

estimate and evaluate risks (see Box 1). Technical advice on these matters is already available and further material is in preparation.

Box 1 – Topics outside the scope of this guidance

- Development and refinement of conceptual models
- Development of appropriate soil sampling strategies
- Collection and testing of soil samples for contamination
- Use of generic and site specific assessment criteria for risk assessment

Nevertheless, all of these activities have a bearing on whether, and to what extent, a statistical approach to data assessment is valid and helpful in making decisions about individual sites. For this reason, this note highlights aspects of these activities which are key to a statistical approach and directs readers to relevant sources of advice.

The guidance assumes that users are familiar with the policy and legislative framework for managing land contamination in the UK^{4,5} and how good technical practice is meant to operate to ensure that particular legal requirements are met.

1.2 TARGET AUDIENCE

This guidance has been prepared for the use of all stakeholders (regulators, land owners and occupiers, consultants, contractors and others) who are responsible for managing land contamination issues in the UK and who have a common interest in ensuring that correct decisions are made in response to the presence of contamination. It is expected that most users will find this material relevant to their day-to-day operations and, for this reason, much of the text has been drafted from the simple perspective of ‘an assessor’.

Note, however, that the statistical procedures set out in Section 6 of the guidance adopt the particular perspective of a regulator operating under either the planning system or Part 2A of the EPA 1990. This has been done in the expectation that clarity about how regulators handle statistical information will encourage all stakeholders to collect, prepare, understand and present data in an appropriate manner. This should ensure a greater degree of confidence in both the outcome of statistical testing and the decision-making process as a whole.

⁴ Defra Circular 01/2006, *Environmental Protection Act 1990: Part 2A, Contaminated Land*, Sept 2006 (England); Scottish Executive, *Paper SE/2006/44, Environmental Protection Act 1990: Part IIA, Contaminated Land Statutory Guidance: Edition 2*, May 2006 (Scotland); WAG, *Part 2A Statutory Guidance on Contaminated Land (2006)*, Dec 2006 (Wales)

⁵ Office of the Deputy Prime Minister, *Planning Policy Statement 23: Planning and Pollution Control, Annex 2, Development on Land Affected by Contamination*, Oct 2004 (England); Scottish Executive, *Planning Advice Note (PAN) 33 – Development of Contaminated Land*, Oct 2000 and *PAN 51 – Planning and Environmental Protection Regulation*, Oct 2006 (Scotland); WAG, *Planning Policy Wales*, March 2002 (Wales)

2.0 Structure of the guidance

The guidance is arranged around a decision flow chart (Figure 1) which sets out the various steps that should be followed when intending to incorporate statistical techniques in the analysis of soil contamination data.

It is divided into two broad parts:

Sections 3, 4 and 5 – which address the basic scientific principles that underpin a statistical approach to data analysis and the sampling and data processing that have to be carried out to obtain suitable datasets.

Sections 6 and 7 – which address the selection and use of statistical tests appropriate to particular regulatory contexts and how the outcome of testing should be reported.

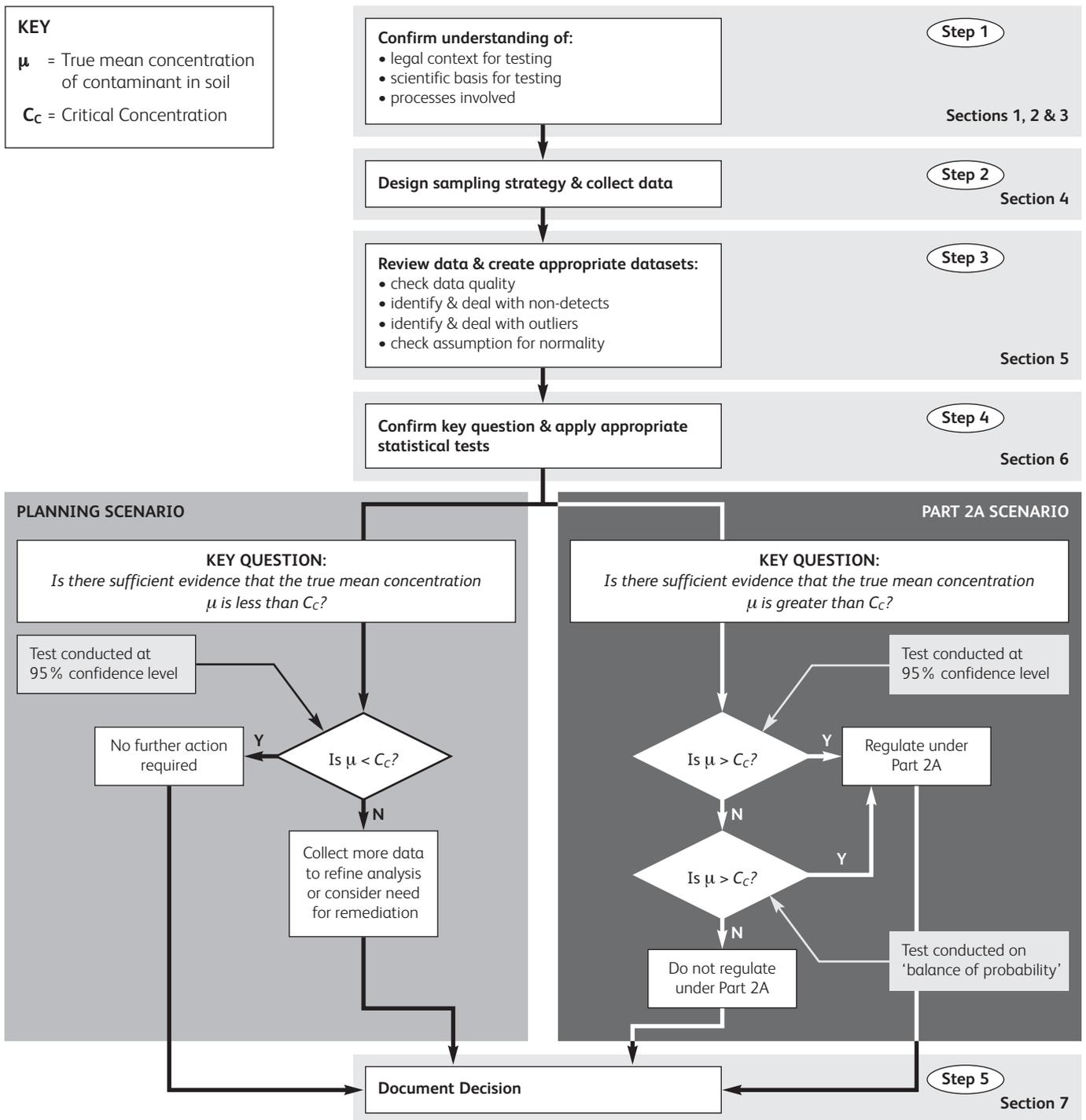
Users are strongly advised to read the background information on statistics and statistical conventions given in Section 3 before they attempt to apply the statistical tests described in Section 6. Since the use of statistics requires that unbiased sample data are available, users should also refer to Section 4 which highlights particular sampling issues that are key to the use of statistical techniques. Note, however, that it is not the intention of this guidance to deal at length with sampling design and implementation since extensive guidance on this is available elsewhere.

It is important to bear in mind that the correct application of statistical tests to soil contamination data relies on assessors being clear about the purpose of testing (and how to frame the corresponding statistical questions). They also need a good understanding of the soil populations from which sample data are collected, of how sample data should be collected to ensure they are unbiased and representative of the body of soil being assessed, and of what inferences can be drawn from the data in both spatial and statistical terms.

Note also that although the statistical approach described in this document aims to provide an objective means of comparing contaminant concentration data with some critical concentration or measure of risk, the choice of an appropriate critical concentration in any particular case is outside the scope of this guidance.

For ease of reference, a list of key statistical symbols and terms is provided at the end.

Figure 1 Decision Flow Chart for Comparing Site Data against a Critical Concentration



3.0 Introduction to basic statistical principles

3.1 BACKGROUND

To say anything meaningful at all about the level of risk that a particular contaminant in soil may pose to human health or the environment, information must be available on the concentration of that contaminant in a defined body of soil.

Normally, assessors are most interested in the mean (or average) concentration of contaminants as this is a reasonable guide to the amount of exposure that an individual will experience over a long period of time, or to the overall impact of a contaminant on other receptors such as groundwater.⁶

Clearly it is not feasible to examine every part of the soil to determine the true mean concentration of the contaminant and therefore assessors have to rely on sampling data to provide an estimate of the mean. In addition, and since it is unlikely that the contaminant is uniformly present at the same concentration throughout the soil mass, it is usually necessary to collect a number of samples from the soil of interest to determine the range or spread of concentrations that may exist. Knowing the variability in contaminant concentration is crucial since it allows quantification of the amount of uncertainty around the estimate of the mean that is attributable to variations in soil concentration.⁷

In simple terms, a statistical approach to sampling and analysis formalises the way in which data are collected, uncertainty about the true mean concentration of contaminants in the soil of interest is quantified, and decisions are made about the level of risk that may exist based on the strength of the evidence provided by the sample data.

For example, if the data provide evidence which is strongly in favour of one set of conclusions (and one course of action) over

another, then decision-makers can proceed with confidence. Alternatively, if the evidence for a particular conclusion and action is less strong, decision-makers can proceed with that action at least knowing the chances of being wrong. Alternatively, they can take steps to strengthen the evidence, for example by collecting further or different types of data.

A statistical approach to sampling and data analysis therefore offers a potentially powerful way of supporting decisions about the condition of land and how it should be regarded in both technical and legal terms. Employing statistics requires, however, that certain conventions are observed with regard to data collection and processing and how questions about the evidence base should be framed.

To successfully apply a statistical approach, the key points to remember are:

- (a) the conceptual model⁸ plays a key role in defining what is a meaningful dataset for statistical analysis;
- (b) strictly, the application of statistical tests is valid only in relation to unbiased sample data which are representative of the soil of interest and have been collected using a sampling density appropriate for the area and depth of soil of interest, likely degree of heterogeneity and the nature of the risk assessment;
- (c) most commonly used statistical tests assume that data are normally distributed although it is possible to use non-parametric methods if the normality of the data cannot be demonstrated;
- (d) most statistical tests are framed in terms of formal hypotheses (so-called Null and Alternative Hypotheses) and since these are determined by the question being posed of the data, and affect the interpretation of the outcome and the method of calculation, it is important to use the correct form for the application in hand.

⁶ Environment Agency, *Secondary model procedure for the development of appropriate soil sampling strategies for land contamination*, R&D Technical Report P5-066/TR, 2000

⁷ Note that actual variations in soil concentration are not the only source of uncertainty in data: sampling and measurement in both the field and the laboratory can make a contribution – see Ramsey, M.H., *Improving the reliability of contaminated land assessment using statistical methods: Part 1 – Basic principles and concepts*, CL:AIRE Technical Bulletin TB7, 2004

⁸ Conceptual Model – a representation of the characteristics of a site in graphical or written form that shows the possible relationships between contaminants, pathways and receptors (CLR11, 2004)

Points (a) and (b) relate to the design of appropriate soil sampling strategies and are discussed in Section 4 of this note. Points (c) and (d) relate to statistical theory and convention and are considered in more detail below.

Note that the statistical information presented below is a highly summarised account of some of the key issues. Gilbert (1987) provides a good basic reference text on the use of statistical methods in managing environmental data.⁹

3.2 STATISTICAL DISTRIBUTION OF DATA

In statistical theory, the spread of data values together with the frequency at which individual values (for a particular attribute) occur, provides an estimate of the amount of variation that is present within a population of possible values.

If there is little variation, most of the values cluster around the mean value and large departures from the mean are rare. If there is much variation in the population, a much wider and flatter distribution of values is observed in which considerable departures from the mean value are quite common.

The distribution of values for a particular population is described by two key parameters (see Box 2); the true population mean ' μ ' and true population standard deviation ' σ '.

Box 2 – Key terms

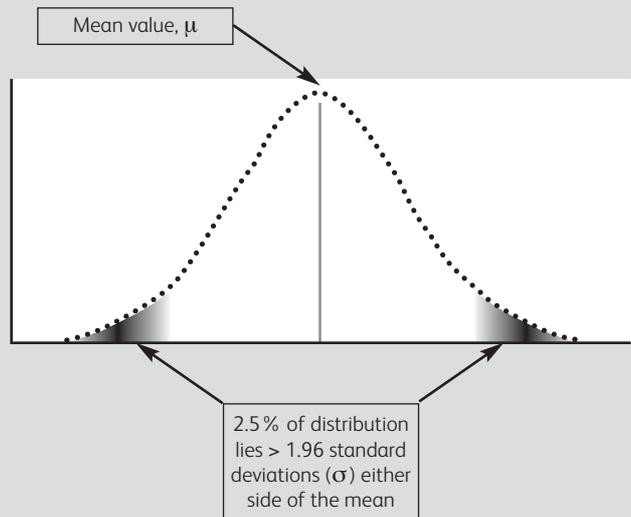
- μ true mean of the values in the population
- σ true standard deviation of the values in the population
- \bar{x} mean of values observed in samples collected from the population (note: not the average value obtained from duplicate analyses of the same sample)
- s unbiased standard deviation of values observed in samples collected from the population

The true mean and standard deviation of the population of values are not (usually) known, therefore, it is necessary to use the sample mean ' \bar{x} ' and sample unbiased standard deviation ' s ' as estimates of ' μ ' and ' σ '. It is worth noting that if it were possible to collect large numbers of samples from the population, the sample mean and standard deviation would provide increasingly more precise estimates of the corresponding population parameters.

A normally distributed dataset follows a classic bell shaped curve where the data are symmetrically distributed and the curve follows a particular mathematical form (see Box 3 where the 'x' axis represents contaminant concentrations and the 'y' axis the frequency with which particular concentration values occur).

In a normally distributed dataset, about 95% of values fall within approximately two standard deviations from the mean value and 5% of values fall outside the central area, within the two 'tails' of the curve (2.5% in each tail).

Box 3 – Key features of a normally distributed dataset



⁹ Gilbert, R O (1987) *Statistical methods for environmental pollution monitoring*. John Wiley and Sons

Note, however, that in land contamination applications, assessors are usually interested only in either the upper or lower tails of the distribution (see the discussion in Section 6).

The logic of many common statistical tests (known as parametric tests), including the 'Mean Value Test' set out in CLR7 (2002) and some described later in this guidance, relies on the assumption that the population from which samples are drawn share these particular characteristics of the normal distribution. The assumption of normality cannot, however, be taken for granted and the robustness of the statistical test may suffer where significant departures from normality occur.

Section 5 of this guidance contains advice on how to examine the statistical distribution of datasets and Section 6 describes the use of non-parametric tests where the normality of the data cannot be assumed.

3.3 NULL AND ALTERNATIVE HYPOTHESES

In a conventional approach, statistical tests are used to decide whether the available evidence supports a particular hypothesis – the Null Hypothesis – or an alternative hypothesis.

The Null Hypothesis (H_0) is put forward as a starting position for testing either because it is believed to be true (but has yet to be proved) or because it can be used to create the basis of an argument or proposition.

The point of the test is to establish whether, on the basis of the available data, the strength of evidence favours the Null Hypothesis or is so removed from what would be expected if the Null Hypothesis were true, that the Alternative Hypothesis (H_1) is more likely to be true.

The outcome of statistical testing is always expressed in terms of H_0 . Therefore the outcome is either:

- H_0 is not rejected; or
- H_0 is rejected in favour of H_1 .

Box 4 – Type I and Type II Errors

| | DECISION | |
|----------------|--|--|
| | H_0 is rejected | H_0 is not rejected |
| H_0 is true | Type I Error Probability (Type I Error) = α | Correct decision Probability = $1-\alpha$ |
| H_0 is false | Correct decision Probability = $1-\beta$ | Type II Error Probability of Type II Error = β |

If H_0 is **not rejected**, this does not necessarily mean that the Null Hypothesis is true – only that there is insufficient evidence to justify rejecting it in favour of H_1 .

If H_0 is **rejected**, it means that there is good evidence to suggest that H_1 is more likely to be true.

One of the underlying design objectives of a statistical test is to ensure a defined low probability (α or significance level) of wrongly rejecting the Null Hypothesis when, in fact, it is true.

If H_0 is wrongly rejected (that is, H_0 is rejected when it is in fact true) a Type I Error occurs (see Box 4). The opposite situation applies if the Null Hypothesis is not rejected when in fact it is false. This type of error is referred to as a Type II Error and the probability of it occurring is denoted β .

Statistical power ($1-\beta$) is the probability of correctly rejecting the Null Hypothesis when it is false and is a function of the sample size, the variation in measured values, the difference between the mean and the critical concentration, and the significance level, α . **It is worth noting that smaller sample sizes tend to result in lower power** which means that it is more difficult for the assessor correctly to reject the Null Hypothesis and hence obtain a clear-cut outcome from statistical testing (see also the discussion on the inherent power of different tests in Section 6.3.1).

Typically, the significance level of a test (α) is set in advance at 0.05 (5%), however, the actual value of α obtained in a statistical test will vary, depending on the data being tested. Since this parameter can be calculated for a particular dataset, it is possible for assessors to determine the chances of making a Type I Error. This is a useful way of looking at the strength of evidence (for rejecting or not rejecting the Null Hypothesis) and of supporting any conclusions drawn from the test.

Statistical convention dictates how significance tests are structured and should be interpreted. Setting the Null Hypothesis (or creating the basis of the argument in any particular case) is nevertheless within the control of the assessor (see the example in Box 5).¹⁰

Asking a different key question changes the Null and Alternative Hypotheses and the interpretation of the outcome of the test, as well as the mechanics of the statistical calculation. It follows, therefore, that there has to be clarity about the question the test is designed to answer.

This has important implications in land contamination cases as discussed later in this guidance, since the question being asked will be different depending on the regulatory context.

Box 5 – Example Null and Alternative Hypotheses

The key question for a site going through the planning system is:

Is there sufficient evidence that the true mean concentration of a contaminant in soil (μ) is less than some critical concentration (C_c)?

In this case:

- the Null Hypothesis is that μ is equal to or greater than C_c
- the Alternative Hypothesis (*the question the test is designed to answer*) is that μ is less than C_c
- let the chosen significance level = 0.05 (confidence level = 0.95)

The relevant statistical test involves comparing a conservative estimate of the true population mean (i.e. the 95% Upper Confidence Limit, or 95% UCL) with the critical concentration, C_c (see Section 6).

Since the 95% UCL is at most times greater than the true population mean, it follows that if the 95% UCL is less than C_c , the assessor will know (with a defined high level of confidence) that the true population mean (μ , the value which is not known) is also likely to be less than C_c .

¹⁰ The authors recognise that the formulation of the Null Hypotheses for both the planning and Part 2A scenarios adopted in this document departs from normal statistical practice. For the purposes of this guidance, however, this is considered justified as it reinforces for non-statistician users the fact that the key question changes, depending on the scenario, and that the statistical tests target different parts of the data distribution in each case

4.0 Key aspects of designing sampling strategies

4.1 BASIC PRINCIPLES

The starting point for a statistical approach to data analysis comes well before the point at which an assessor begins to review the factual data obtained from intrusive site investigation.

As already noted, any intention to use a statistical approach to data analysis must be established before samples are collected since significance tests rely on unbiased sample data being available (see Section 4.3). The context within which the investigation and data analysis are to be carried out (i.e. as part of redevelopment or for regulatory purposes under Part 2A) should also be apparent at this stage. Once these issues are decided, the collection of sample data should follow established good practice for intrusive site investigation.

There is already substantial UK and international technical literature on how to design and implement intrusive site

investigation projects in land contamination applications **and therefore sampling design and implementation are not covered in detail in this guidance.**

A selection of key UK guidance documents on these matters is listed in Box 6.

CLR11 sets out the overall role of site characterisation data in supporting risk assessment and management in land contamination applications with the expectation that any data collected will be:

- **relevant** – to the process of assessing and managing risk;
- **sufficient** – for the purpose of characterising contaminants, pathways and receptors so that risks can be adequately assessed and managed;
- **reliable** – to the extent that measurements or observations accurately reflect true or likely site conditions rather than bias or defects in the sampling or analysis methodology;

Box 6 – Key guidance on Site Characterisation

| | |
|--|--|
| CLR 11, 2004 ³ | Sets out the overall framework for the technical activities involved in assessing and managing the risks associated with land contamination and signposts a range of other guidance on specific topics |
| Guide to good practice for the development of conceptual models, 2001 ¹¹ | Discusses the information to be considered and issues to be addressed in the development of conceptual models for land contamination applications |
| Code of practice for the investigation of potentially contaminated sites, 2001 ¹² | Contains technical advice on the design and implementation of site characterisation (including intrusive site investigation) activities for contaminated land focusing on the selection and use of different field sampling and monitoring techniques; collection, handling and transport of samples; and reporting of field observations and related data |
| Technical aspects of site investigation, 2000 ¹³ | Provides guidance for project managers, rather than specialists, on the technical aspects of site investigation |
| Secondary model procedure for the development of appropriate soil sampling strategies, 2000 ⁶ | Addresses key design issues for soil sampling including zoning, sampling patterns and number of samples |

¹¹ Environment Agency, *Guide to good practice for the development of conceptual models and the selection and application of mathematical models of contaminant transport processes in the subsurface*, 2001

¹² British Standards Institution, *Investigation of potentially contaminated sites*, Code of Practice, BS:10175, 2001

¹³ Environment Agency, *Technical aspects of site investigation in relation to land contamination, Volume 1*, R&D Technical Report P5-065/TR, 2000 & Volume 2 Text Supplements

- *transparent* – in the sense that the origins and provenance of the data are clear and unambiguous.

Within this overall framework, several key elements of sampling design are relevant to a statistical approach to data analysis as shown in Box 7.

Only when assessors are satisfied that the sampling design has delivered data which are suitable for statistical analysis should the procedures described in Section 6 of this document be applied.

Box 7 – Key elements of sampling design relevant to the statistical analysis of data

Important for creating meaningful datasets & for understanding variations in contaminant concentrations

Creates basic building blocks for individual datasets

Note the condition for unbiased sample data if statistical testing is to be carried out

Important for understanding variations in concentration with depth & for defining relevant datasets

Factors which affect the technical validity of individual sample results & hence their inclusion or exclusion from particular datasets

Will determine Method Detection Limit (MDL) & may affect number of non-detects in dataset with implications for outcome of statistical testing

A4 KEY DESIGN PARAMETERS FOR SOIL SAMPLING

The main technical decisions that assessors have to make when developing soil sampling strategies are:

- Should the same sampling strategy be applied to the whole of the site, or should different strategies be developed for specific areas or zones?
- Should sampling be carried out on a staged basis or as a single exercise?
- What substances should be looked for and what other parameters should be measured?
- Where and how many sampling locations should be planned for?
- At what depth should samples be collected or tests conducted?
- What type and how much sample should be collected and how should it be handled?
- What frequency of sampling or testing should be carried out?
- What field techniques should be used to collect samples or conduct tests?
- What laboratory techniques should be used to test samples?

Text taken from: *Secondary model procedure for the development of appropriate soil sampling strategies for land contamination*, R&D Technical Report, P5 – 066 TR, Environment Agency, 2000

4.2 THE IMPORTANCE OF THE CONCEPTUAL MODEL

It is now generally accepted that the conceptual model of pollutant linkages, developed initially using documentary information about a site and its setting is key to the design of appropriate intrusive site investigations and to the generation of good quality data for risk assessment.

The conceptual model can help to inform decisions on the zoning of sites – either to break down a large site into more manageable units or because there are good reasons for varying the sampling design within different zones.

Site zoning also offers a means of understanding and, in some cases, minimising the uncertainty associated with variations in the actual concentration of contaminants in soil.

For example, large variations in soil contamination data may be expected due to the way in which land has been used or because of the behaviour of contaminants in the environment (see Box 8).

Box 8 – Variations in Contaminant Concentration

Possible reasons for variations in observed concentrations of contaminants are:

Previous uses of land – particular parts of a former industrial process may be more likely than others to have resulted in ground contamination leading to marked variations in contaminant concentrations across a site.

Physico-chemical properties of the contaminant – for example, ‘made ground’ may be highly contaminated but underlying natural ground may be largely unaffected unless contaminants are soluble or otherwise liable to move from made ground into natural ground.

Laboratory or sampling errors – anomalously high (or low) concentration values relative to the bulk of results in a dataset (and contrary to the pattern suggested by the conceptual model or visual observation) may occur due to laboratory error or because of flaws in the sampling technique, such as cross-contamination or the loss of reactive or volatile substances.

In these cases, investigators may commence site investigation work with the specific intention of sampling from one or more different soil populations and knowing precisely how datasets are to be handled for statistical analysis purposes.

In other cases, large variations (including the presence of extreme values or ‘outliers’) may be identified only when the data are available for review and sorting into meaningful datasets. In these cases, decisions have to be made as to whether the results reflect real variations in concentration within the soil of interest, the presence of a site-related feature which could not have been anticipated by the conceptual model (such as a small, previously unidentified area of fill) or have arisen for other reasons, such as laboratory or sampling errors.

Whatever the circumstances, it is important to use the conceptual model as a tool to understand and explain both the spatial distribution of soil contamination data and its statistical distribution (for example where results suggest that more than one population of soil is present).

4.3 UNBIASED SAMPLING DATA

In land contamination applications, sampling data are usually collected in one of two main ways:

- through judgmental (targeted or biased) sampling;
- through systematic (non-targeted or unbiased) sampling.

In a judgmental sampling approach, samples are collected to provide information about soil conditions in relation to a particular feature which the conceptual model has suggested is present on the site. For example, assessors may be keen to know whether and to what extent contaminants have escaped from a sub-surface tank, or to determine the location of a particular feature such as the edge of a landfill. In these situations, sampling locations are deliberately targeted in the area of interest – around the tank or along the assumed boundary line.

In systematic sampling, each part of the site or body of soil has an equal chance of being sampled and there is no bias in the selection of sampling locations. One advantage of this type of approach is that it allows assessors to obtain the best overall coverage of the area of interest given the available (that is, budgeted) number of sampling locations.

In the UK, systematic sampling is often achieved by positioning a [theoretical] regular sampling grid of known spacing over the area to be sampled with sampling points located at grid nodes.¹⁴ There are some, fairly uncommon, circumstances where such an approach may not result in truly representative site data, for example where soil contaminants are themselves aligned in a systematic pattern that could be missed by the sampling grid. An example might be a tank storage farm where both tanks and any associated distribution network might be arranged in a regular grid formation. Depending on grid spacing and alignment, in these circumstances it is possible for a regular grid sampling pattern to miss potential contamination sources giving an unreliable picture of the condition of the site. Used sensibly, however, regular grid sampling has much to commend it in terms of overall simplicity and ease of use in the field.

A possible alternative approach is so-called random sampling, which involves sampling at an appropriate number of randomly-generated locations. A random sampling strategy has the advantage of being truly unbiased thereby reducing the risk of missing systematic patterns of contamination which might be not known at the design stage. Given sufficient samples, it may therefore supply a more appropriate estimate of the site-wide mean contaminant concentration.¹⁵

When developing sampling strategies and determining the number and location of sampling points to calculate mean

concentrations, it is also important to be clear about the scale of sampling and how this relates to the type of risk assessment being carried out. Key issues include the area of land ('averaging area') which the mean value is meant to represent and how likely it is that some defined area of elevated contamination (a 'hotspot') may be missed by the proposed sampling pattern. Reference 6 provides guidance on how averaging areas relevant to particular types of risk assessment might be defined.

The depth of sampling is also an important design issue to consider when collecting data for risk assessment purposes and is a key parameter in defining the mass of soil of interest (see Reference 6). For example, in human health applications, assessors are usually interested in near surface soil conditions. If there are concerns about the risks posed by leachable contaminants for the water environment, soil sampling may need to extend to greater depths.

The main issue to note as far as this guidance is concerned, however, is that **strictly, only systematic or unbiased sampling data should be subject to statistical testing** since it is only from this type of data that inferences can be drawn about conditions within the sampled body of soil as a whole.

This should not be taken to mean that judgmental sampling has no value or that data obtained from judgmental sampling cannot be used to develop a broad understanding of the nature or spatial distribution of the contamination on a site. Indeed, judgmental sampling is often used to good effect to obtain early confirmation of the type and concentration of contaminants in suspect locations. Nor is it to say that small deviations from an otherwise genuinely unbiased sampling plan, for example to avoid physical obstructions or site services, need necessarily preclude the use of statistical techniques.

¹⁴ Note that 'the area to be sampled' could be a specific, and in that sense a 'targeted' sub-area or part of the site which the assessor has decided requires particular attention

¹⁵ United States Environmental Protection Agency, Office of Environmental Information, Washington DC 20460, *Guidance on choosing a sampling design for environmental data collection*, EPA QA/G-5S, 2002



However, **any inferences about conditions in unsampled locations drawn from the statistical analysis of judgmental sample data are of doubtful validity and should be avoided.**

For this reason, it is recommended that data obtained using a combination of judgmental and systematic sampling approaches are collated and considered separately, and that the formal use of statistical techniques is confined to unbiased sample data only.

Several types of sampling strategy have been illustrated in the literature but a complete description is beyond the scope of this guidance. Further details on approaches to sampling can be found in References 6 and 13. The United States Environmental Protection Agency (USEPA) has also produced guidance on this issue (Reference 15) which contains a useful summary of the relevant issues and possible approaches.

5.0 Reviewing data and creating appropriate datasets

5.1 INTRODUCTION

The key objective of this stage of data analysis is to obtain good quality and meaningful datasets that justify the application of statistical tests.

Assuming that assessors have followed good technical practice when designing the soil sampling strategy, they will already have a view about the number and type of datasets that will be subject to statistical testing. Before applying statistical tests, however, assessors should review the data to ensure that they are complete and free from obvious error, and to check that they do not depart significantly from the distributional assumptions underpinning the relevant tests.

Typically, soil contamination data will be held in electronic spreadsheet or database form; this may have been supplied directly by the laboratory or compiled from laboratory certificates or on-site records and measurements. Collating the data in a spreadsheet or a database has a number of advantages. For example, it means that it is relatively easy and straightforward to:

- manually or automatically QA/QC and scrutinise the data – for example to identify anomalous or invalid data, missing units etc.;
- sort the data – for example into datasets for individual contaminants, zones, type of sample, sample depth etc., and ultimately,
- calculate the basic statistics needed for statistical testing, such as sample mean and sample unbiased standard deviation.

It also means that data can be presented in summary form (although it is important to remember that this does not obviate the need also to provide data in its original form, such as Certificates of Analysis and site log book records). Electronic storage means that it is relatively easy to produce spatial representations of the data, such as ‘blobby plots’ (or proportional symbol maps) showing contaminant concentrations across the site by depth and in relation to site boundaries and other features. Statistical presentations can include frequency histograms and probability plots.

Graphical representations can be extremely useful in helping assessors to better visualise the pattern of contamination on the site – for example whether more than one ‘population’ of soil is present, whether extreme values (outliers) exist and are spatially related, and whether assumptions about the statistical distribution of data are justified. This ensures that the most robust and meaningful datasets for statistical testing are created.

5.2 REVIEWING DATA QUALITY

An important first step in the assessment of any land contamination data is an evaluation of data quality. As a minimum, this should include:

- checks on the ‘completeness’ of the data – for example, are results available for all sample locations, contaminants, sample types, sample depths etc.?
- checks on the accuracy of the data – are results correctly identified, for example, by sample location, depth, type, sampling date?
- identification of obviously anomalous results such as elevated values that are unexpected given the conceptual model and (say) a field description in a borehole log – this may indicate a labelling or laboratory error;
- identification of invalid data – for example where the field or laboratory record indicates that sample integrity may have been compromised.

Once these basic checks have been carried out and the reasons for any discrepancies have been identified and documented, the assessor is in a position to move on to the next step in the statistical assessment of the data.

5.3 CREATING (AN) APPROPRIATE DATASET(S)

The next task is to create appropriate datasets for statistical testing. There are three key elements:

- dealing with non-detects;
- dealing with outliers;
- understanding the statistical distribution of data.

This may be an iterative process since examination of the data with respect to these three elements may lead to changes in the composition of datasets to better reflect true or likely conditions on the site.

5.3.1 Dealing with non-detects

Data generated from chemical analysis in the laboratory will often contain 'non-detects', that is, sample results which are reported by the laboratory as 'less than' a specified minimum value – usually the limit of detection, or Method Detection Limit, MDL – for the particular type of sample and laboratory analysis method used. In these cases, the actual concentration of the contaminant in the sample is unknown although it will lie somewhere between zero and the detection limit.

Assuming that the MDL of the relevant laboratory test is set at a sensible level (i.e. lower, and ideally at least 10 x lower, than the critical concentration which the assessor intends to use in the data analysis), the presence of non-detects in the dataset nonetheless tells the assessor something useful about the condition of the land – that is, that [parts] of it appear to be uncontaminated.

The presence of non-detects, however, also creates practical difficulties for statistical analysis since 'less than' values cannot be used to compute the key statistics used in the tests. A decision therefore has to be made about the value to be assumed for samples presenting as non-detects. For example, it could be assumed that non-detect samples contain the contaminant at the same concentration as the MDL or at some lesser concentration such as 50 % of the detection limit.

The presence of non-detects (and the choice of any values as substitutes for them) can have implications for the estimation of key parameters such as sample mean and

sample unbiased standard deviation, upon which many statistical tests rely. Where a dataset contains relatively few non-detects, the effect on key parameters, and hence on the outcome of the statistical tests themselves, is likely to be small, however, where a substantial number of the values in the dataset present as non-detects, particularly where the MDL is close in value to the critical concentration used in the test, there could be a much more significant effect on the outcome of testing.

The following may be helpful in deciding how to proceed where non-detects appear in the data:¹⁶

- where the proportion of non-detects within the dataset is less than 10 to 15 %, non-detect values should simply be replaced by a 'small' number (e.g. MDL or 0.5MDL);
- where the proportion of results within the dataset is greater than 10 to 15 %, especially where the MDL is close to the critical concentration, non-detects can be substituted-for as before but note that the values selected may have a large effect on the outcome of testing in which case a sensitivity check could be carried out to establish the effect of substituting different values on the outcome of statistical testing. Reference 17 describes possible statistical methods for adjusting the dataset for non-detects¹⁷ and Reference 18 contains a good discussion on censored datasets (i.e. datasets containing non-detects).¹⁸

Remember also that the presence of large numbers of non-detects within a dataset might indicate that zoning or data sorting decisions may need to be reviewed and revised. For example, it might indicate that the dataset includes a large number of samples from natural (uncontaminated) ground and a relatively small number of more contaminated 'made ground' samples (which may be more appropriately considered as a separate dataset).

¹⁶ Note also that sample size affects the procedures to be used: clearly a dataset in which 1 sample in 4 is below the MDL tells the assessor much less about the true or likely condition of a soil population than one in which 25 out of 100 samples present as non-detects. For a fuller discussion, see Reference 17

¹⁷ United States Environmental Protection Agency, Office of Environmental Information, Washington DC, *Data Quality Assessment: Statistical Methods for Practitioners*, EPA QA/G-9S, EPA/3240/B-06/003, February 2006

¹⁸ Daniels, W M & Higgins, N A (2002) *Environmental distributions and the practical utilisation of censored environmental data*, NRPB report NRPB-R02008

5.3.2 Dealing with outliers

Site investigation data can comprise contaminant concentrations spanning several orders of magnitude. This, amongst other things, reflects heterogeneity in soil conditions, the variability of contaminant distributions at large and small scales, differing sources of contamination, and uncertainty associated with laboratory analysis.

Extreme values can also result from inaccuracy in sampling, chain of custody and laboratory analysis processes, measurement system problems, transcription or data entry errors and the use of incorrect units in reporting and recording analytical results.

The failure to remove true outliers from a dataset or, conversely, the removal of values which are not in fact outliers, obviously has consequences for the outcome of statistical testing. Therefore if outliers are identified, assessors have to decide whether they represent genuine soil concentrations or are the result of an error. This means that possible reasons for the presence of outliers should be explored. For example, assessors should:

- re-examine field records (e.g. trial pit and borehole logs) to establish whether observations made at the time the samples were collected can explain the results obtained;
- re-check laboratory certificates to determine whether there has been a data entry error;
- review sampling and sample handling protocols to see if there has been an obvious breach in procedures.

Subject to time and resource constraints, and assuming sufficient material is still available and is in a suitable condition, assessors may also consider asking for samples to be retested to confirm the original results. Note, however, that the failure to confirm an original test result may not necessarily mean that the original result was wrong, especially where contaminant concentrations are highly variable.

In general, however, outliers should be excluded from a dataset ONLY where they:

- **are obviously and demonstrably the result of an error that can be identified and explained** – in which case the correct value should be identified and the dataset amended, where possible, or the erroneous value excluded with justification, or
- **clearly indicate that more than one soil population exists within the dataset and this can be justified by (or informs the further development of) the conceptual model** – in which case the different population expressed by the outlier(s) should be explored in more detail, either by reviewing and refining zoning decisions and treating outlier values as a separate population or even individually or, if necessary, by undertaking further site sampling to verify conditions in the vicinity of outlier values.

In all other cases, outlying data should be assumed to be genuine and reflective of the full range of soil concentrations to which receptors may be exposed.

Various ‘outlier’ tests are available that assessors can use to identify anomalous data in a dataset, each with their own advantages and disadvantages (see Reference 17 for details). Appendix B contains guidance on carrying out one such test – Grubb’s Test.¹⁹

Outliers can also be identified by cross-comparison of the concentrations of several similar determinants in the same area. This procedure, which identifies so-called multivariate outliers, is beyond the scope of this document.

5.3.3 Assumptions about the distribution of data

Statistical tests are often based on assumptions about the distribution of the data being tested. Before applying a statistical test it is therefore important to know what these assumptions are and if they are reasonable for the dataset under scrutiny.

¹⁹ Grubb, F (1969) *Procedures for detecting outlying observations in samples*, Technometrics, 11, 1-21

The one-sample t-test described in Section 6.3 of this guidance relies on the assumption that the data being assessed are approximately normally distributed.

The alternative (non-parametric) one-sided Chebychev Theorem, also described in Section 6.3 of this guidance, is recommended where the normality of the data cannot be demonstrated since this test makes no assumption about the shape of the distribution.

There are two main ways of testing the assumption about the normality of the data distribution:

- using graphical presentations such as frequency histograms and probability plots;
- using statistical tests, such as the Shapiro-Wilk normality test.

Appendix C contains guidance on probability plots and the Shapiro-Wilk normality test.

Note that, in general, probability plots are more accurate in assessing the normality of data distributions than are frequency histograms and are therefore preferred. Frequency histograms can, however, be useful in providing an early indication of the shape of a dataset and of identifying those distributions, such as highly skewed or multi-modal types, that clearly depart from normal. Frequency histograms have an additional advantage in that they can be readily produced using most commercially available computer software.

Assessors should also note the following:

- techniques such as normality tests should be used with care because they can provide misleading results. For example, with large datasets, minor deviations from normality may be flagged as statistically significant even though small deviations from a normal distribution will not affect the reliability of the one sample t-test.

Conversely, datasets with a small sample size more easily pass normality tests. Failing, however, to detect non-normality in a small dataset is unlikely to compromise the validity of the one sample t-test.

- although this guidance provides an alternative to the one sample t-test for datasets that clearly depart from a normal distribution, the method based on the Chebychev Theorem is not as powerful a test as the one sample t-test. This may affect the assessor's ability to obtain a clear cut outcome from statistical testing (see the discussion in Section 6.3.1).

A more detailed discussion on this can be found in Reference 17.

It is important to note that this guidance does not recommend log transformation as a means of normalising soil contamination data when calculating confidence limits. International literature²⁰ and empirical studies²¹ discuss alternative and more effective methods for skewed datasets, such as those based on the Chebychev Theorem.

²⁰ Singh A.K., Singh A. and Engelhardt M. (1997), *The lognormal distribution in environmental applications*, US EPA Technology Support Center Issue EPA/600/R-97/006

²¹ Masi, P & P. Morgan, *Statistical assessment of contaminated land: some implications of the 'mean value test'*, CL:AIRE Technical Bulletin B12, 2006

6.0 Procedures for the application of statistical tests

6.1 INTRODUCTION

This part of the guidance sets out procedures for comparing the concentration of a contaminant in soil, as determined by sampling and laboratory analysis, with a critical concentration (C_c).²²

The procedures describe statistical tests to assess with a stated degree of confidence the difference between the true population mean (represented by the sample mean, \bar{x}) and C_c .

Two different assessment methods are presented:

- the one sample t-test (parametric test) which can be used where it can be shown that data are normally distributed; and
- a method based on the one-sided Chebychev Theorem (non-parametric) which can be used where data are not normally distributed.

Further details on the two tests are given in Section 6.3.

Application of the tests to soil contamination data is discussed in relation to regulation under planning legislation (the Planning Scenario) and regulation under Part 2A of the EPA 1990 (the Part 2A Scenario).

6.2 DEFINING THE NULL AND ALTERNATIVE HYPOTHESES

Section 3.3 of this note explains why it is important to frame the Null and Alternative Hypotheses in a way that is statistically robust and allows the assessor to draw the correct conclusions from the data. Since different key questions apply in each of the two regulatory scenarios considered in this note, this section begins with a discussion of the Null and Alternative Hypotheses that apply in each case.

6.2.1 Planning scenario

In this case, formulation of the Null and Alternative Hypotheses is taken from the perspective of a **regulator operating under planning guidance and reviewing the data analysis carried out by a developer**.

For this scenario, it is important to recognise that it is the developer's responsibility to address the potential or actual presence of contamination on land by carrying out an appropriate risk assessment and, if unacceptable risks exist, by implementing a suitable programme of remediation. It is also the developer's task to carry out the relevant data collection, data analysis and statistical testing.

When deciding the confidence level at which statistical tests are carried out, developers should bear in mind that the higher the confidence level, the less chance there is of wrongly concluding that contaminant concentrations are low relative to the critical concentration (i.e. less chance of wrongly concluding that the land is suitable for development when it may not be). This might be a particular consideration if there are plans for independent sampling of the land.

It is therefore recommended that unless there are very good reasons for selecting an alternative confidence level, statistical tests carried out under the planning scenario should be conducted at the 95% confidence level.

Under the planning scenario, the regulator is likely to be presented with reports containing data (including statistical analysis) on two main occasions: when the developer presents the risk assessment report for the land and, if remediation has been carried out, when a verification report is submitted setting out the condition of land following remediation.

²² Note that the selection of an appropriate critical concentration (C_c) is a matter for the assessor to decide on a site-specific basis and by reference to the relevant published literature. For the purposes of this guidance, C_c is assumed to be a measure of 'unacceptable' risk

In considering such reports, the regulator will wish to see evidence which demonstrates (with a high degree of confidence) that the land is 'suitable for use'; that is, that the concentration of contaminants is low relative to an appropriate critical concentration.

For the planning scenario, therefore, the regulator should check whether the developer has considered the following key question:

Is there sufficient evidence that the true mean concentration of the contaminant (μ) is less than the critical concentration (C_c)?

and, referring back to Section 3.3 of this note, whether the appropriate Null and Alternative Hypotheses have been adopted:

$$H_0 : \mu \geq C_c$$

$$H_1 : \mu < C_c$$

If statistical testing suggested that H_0 **should not be rejected**, then the developer should have concluded that the true mean concentration of the contaminant may be equal to, or greater than, C_c . In these circumstances, the developer should have considered:

- collecting further data about the condition of the land – since this may alter the outcome of the test, for example by providing a more accurate estimate of the population mean and standard deviation of the contaminant in soil; or
- undertaking remediation as a precaution or (if remediation has already been carried out) carrying out further or a different type of remediation.

The regulator should then expect to see a revised data analysis using additional information on the condition of the land.

If the outcome of statistical testing suggested that the Null Hypothesis should

be rejected in favour of the Alternative Hypothesis, H_1 , then the developer should have concluded that there is **good evidence that the true mean concentration of the contaminant is less than the critical concentration and no further action need be taken**.

Assuming that the test was conducted at a confidence level of 95%, the chance of this being the wrong decision would be less than 5%.

In other words, based on the available information **and assuming that sampling and testing have been carried out according to published good technical practice, that the data are representative of the land under scrutiny at an appropriate scale and that an appropriate critical concentration has been selected**, in this latter case the developer (and the regulator) should have a high level of confidence that the land is 'suitable for use'.

6.2.2 Part 2A scenario

In this case, formulation of the Null and Alternative Hypotheses is taken from the point of view of a **regulator working under the Part 2A legislation**.

Bearing in mind this guidance focuses primarily on soil contamination and its implications for human health, the regulator's responsibility in this case is to decide whether the land is in such a condition that it represents "a significant possibility of significant harm" to human health.

Ideally, the regulator should make this decision at the 95% confidence level on the basis that if the Null Hypothesis is rejected with a high degree of confidence, the regulator has an immediate and robust basis for determination. Under Part 2A, however, the decision can be made on the basis only of the 'balance of probabilities'.

In statistical terms, this means that the regulator should start by conducting the test at the 95% confidence level (significance level of 5%) but, if the Null Hypothesis cannot be rejected at this level, may also determine land at a lesser but still

defensible confidence level of 51 % or more (significance level of 49 % or less). For the Part 2A scenario, the key task for the regulator is to show that the true mean concentration of the contaminant in soil **is greater than** the critical concentration. Therefore the key question is:

Is there sufficient evidence that the true mean concentration of the contaminant (μ) is greater than the critical concentration (C_c)?

Referring back to Section 3.3 of this guidance, the appropriate Null and Alternative Hypotheses for this proposition are:

$$H_0 : \mu \leq C_c$$

$$H_1 : \mu > C_c$$

At the 95 % confidence level, if the outcome of the statistical test suggests that H_0 **should be rejected**, the regulator should conclude that there is good evidence that the true mean concentration of the contaminant is greater than C_c . In other words, and subject to meeting all the legal requirements for determination of land set out in the primary legislation and statutory guidance to Part 2A, having good quality data and an appropriate critical concentration, the regulator could conclude that the land meets the description of contaminated land. Further, in drawing this conclusion, the regulator could proceed to determination knowing that the chance of being wrong is less than 5 %.

If the outcome of the test at the 95 % level is that H_0 **should not be rejected**, the regulator should conclude that the true mean concentration of the contaminant may be less than, or equal to, C_c – in other words, he or she cannot conclude that the land would meet the legal definition of contaminated land.

The regulator may then explore the actual strength of evidence against the Null Hypothesis being true (see Step 13 of Box 10) with two possible outcomes:

- if the evidence suggests that the probability against the Null Hypothesis being true is small (less than 51 %), the regulator **should not reject** the Null Hypothesis and should not proceed to determination because even on a balance of probabilities, it is unlikely that the land would meet the legal definition of contaminated land.
- if the evidence suggests that the probability against the Null Hypothesis being true is less than 95 % but still greater than 51 %, the regulator **can still reject** the Null Hypothesis and proceed to determination on a 'balance of probabilities' basis although clearly in this case, the determination will be based on a (statistically) lower degree of confidence than 95 %.

In all cases, the significance tests should be applied only if the regulator is satisfied that all sampling and testing has been carried out according to good technical practice and that the data are representative of the land under scrutiny at an appropriate scale.

6.3 SELECTING AN APPROPRIATE STATISTICAL TEST

6.3.1 Test methods

As previously discussed, testing for the Null and Alternative Hypotheses described above requires a decision to be made about the distribution of the data since different statistical tests make different assumptions about the shape of the distribution.

The one sample t-test assumes that the data are identically and independently distributed²³ and come from an appropriate normally distributed population. The test is not sensitive to a moderate departure from normality but is sensitive to the presence of outliers. Note that this test is the same as the 'Mean Value Test' described in CLR7

²³ In probability theory, a sequence or other collection of random variables is independently and identically distributed if each has the same probability distribution as the others and all are mutually independent

2002 **but only when applied in the context of the planning scenario and with a normally distributed dataset.**

The one-sided Chebychev Theorem (or Chebychev Inequality) is not based on any distributional assumption. Methods which use the Chebychev Theorem are not strictly true statistical tests and are based on the assumption that 's', the estimate of the true population standard deviation ' σ ' is "close enough" to the true value. However, when the assumption of an approximate normal distribution is not valid, an assessment based on this theorem does give a reliable upper bound of the 95 % Upper Confidence Limit (UCL) and lower bound of the 5 % Lower Confidence Limit (LCL) respectively and is therefore preferred to the one sample t-test.

These tests have been selected for the following reasons:

- over the last twenty years, research into the subject has led to the identification of about twenty different formulae for dealing with the problem of evaluating Upper and Lower Confidence Limits for the population mean under different working scenarios;
- most of these formulae are appropriate only for particular cases;
- some formulae, in spite of the conceptual complication, tend to reproduce the results of the one sample t-test (e.g. bootstrap and jackknife re-sampling);
- some others tend consistently to over-estimate the 95 % UCL, especially with small datasets, and are highly sensitive to the presence of non-detects (e.g. H-Statistics).²¹

For the purposes of this guidance, therefore, it was considered appropriate to refine all the available formulae down to the two that are applicable to a wide variety of cases whilst avoiding conceptual complications as far as possible.

When considering which of the two tests to use, however, assessors should bear in mind that, in general, the one sample t-test is more powerful than the method based on the Chebychev Theorem. In the latter case,

the method calculates a more cautious confidence interval than results from the one sample t-test because there is less certainty about the shape of the distribution. This means that, other things being equal, it is more difficult to reject the Null Hypothesis using the Chebychev Theorem than using the one sample t-test. It is hence more difficult for the assessor to show a clear outcome from statistical testing.

In other words, using the Chebychev method it is more difficult to show that contaminant concentrations are clearly lower than the critical concentration for a case being considered under planning, and to show that concentrations are clearly higher than the critical concentration where testing is carried out in the context of Part 2A.

Given that the one sample t-test is also not sensitive to moderate departures from normality, it is recommended that assessors use the t-test **unless** there is good evidence that the dataset departs significantly from normality. Note that reviewing and adjusting datasets by re-zoning the site (**provided this is consistent with the conceptual model**) and/or collecting additional data may help in generating datasets that better satisfy the assumption of normality.

6.3.2 How the tests work

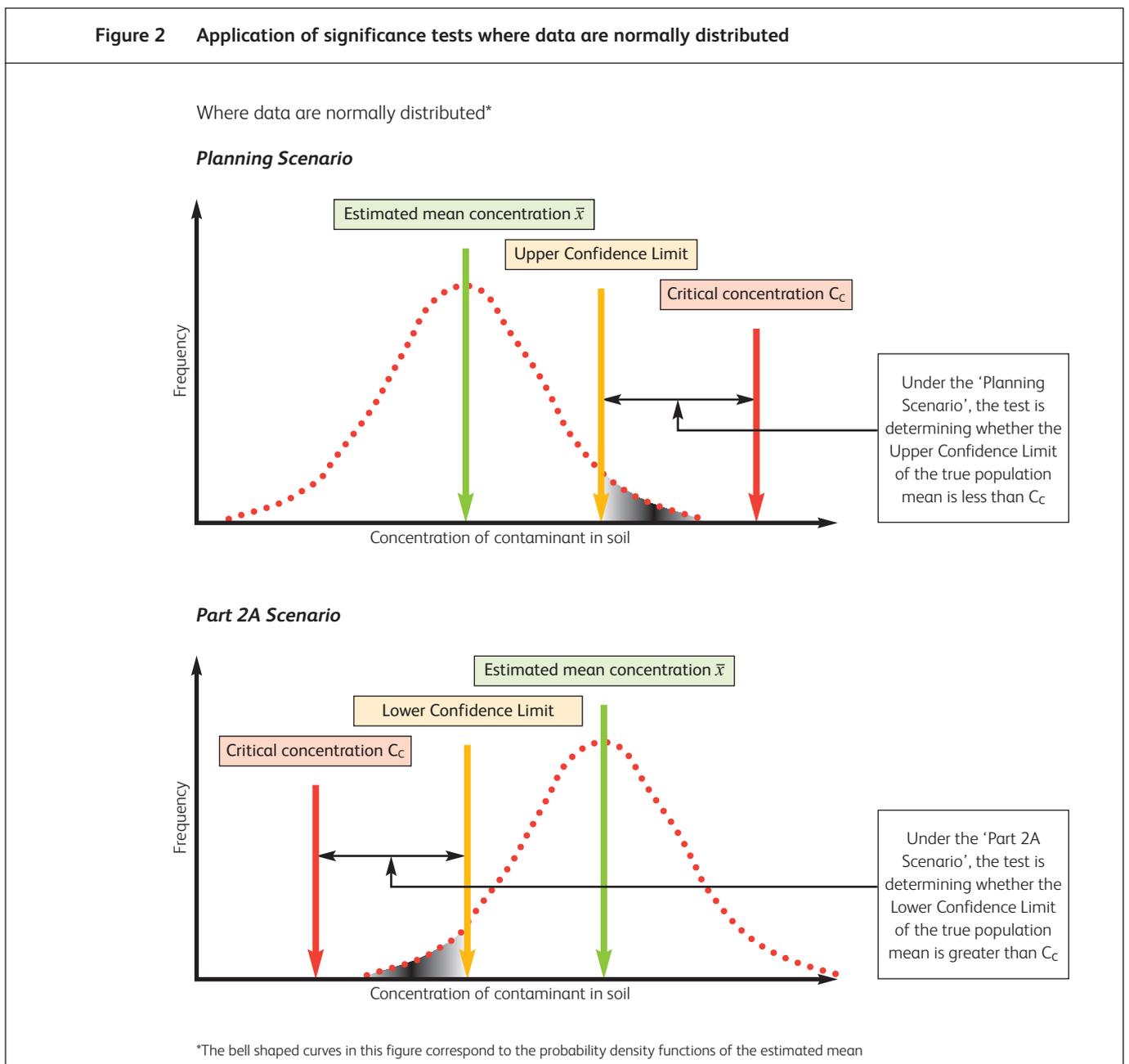
It is important to note that the assessment methods presented in this guidance perform different functions depending on the scenario being considered. Understanding these different functions and presenting the distinctions graphically helps to explain what the tests are designed to do.

For example, in the planning scenario, where data are normally distributed the statistical test in effect makes a comparison between a value larger than the sample mean (in this case the UCL) and the critical concentration in order to draw conclusions about the condition of the land under scrutiny. In the Part 2A scenario, the comparison is between the LCL and the critical concentration (see Figure 2). Confidence limits are calculated using the "t" statistic and take into account the spread (or standard deviation) of values in the dataset being considered.

Note that in Figure 2, sample data are shown in the form of a normal distribution, where the 'x' axis represents contaminant concentration and the 'y' axis the frequency at which particular concentrations occur within the sample data.

Where the normality of the data cannot be demonstrated, conventional assumptions about the shape of the distribution do not apply. In this case, although the proposed assessment method operates according to the same principles as before, the confidence

Figure 2 Application of significance tests where data are normally distributed



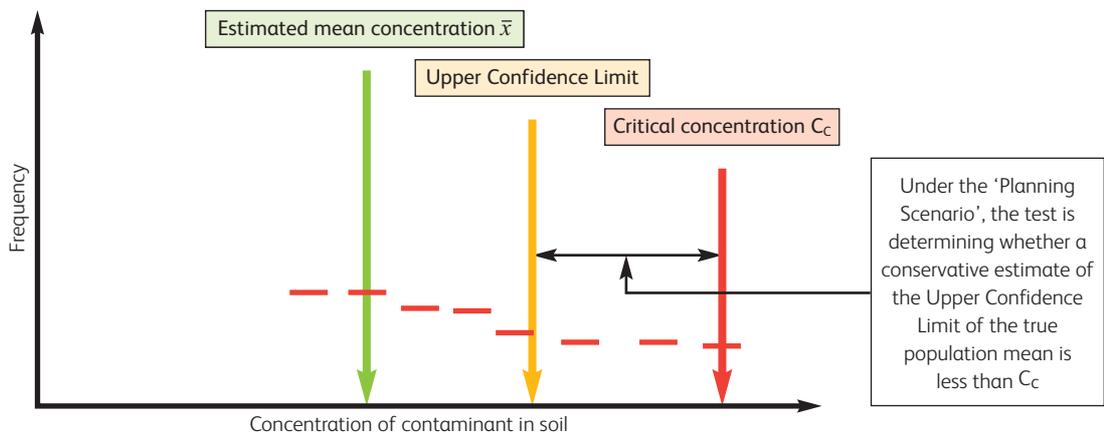
limits are estimated using a different set of tables and critical values derived from the Chebychev Theorem (see Figure 3). In order to distinguish this test from the one sample t-test, the letter “k” is used rather than the letter “t” (see Box 9 and 10).

In Figure 3, the x and y axes respectively represent contaminant concentrations and frequency of occurrence as before, however, in this figure the dashed line is intended to show that there is some uncertainty about the exact shape of the distribution.

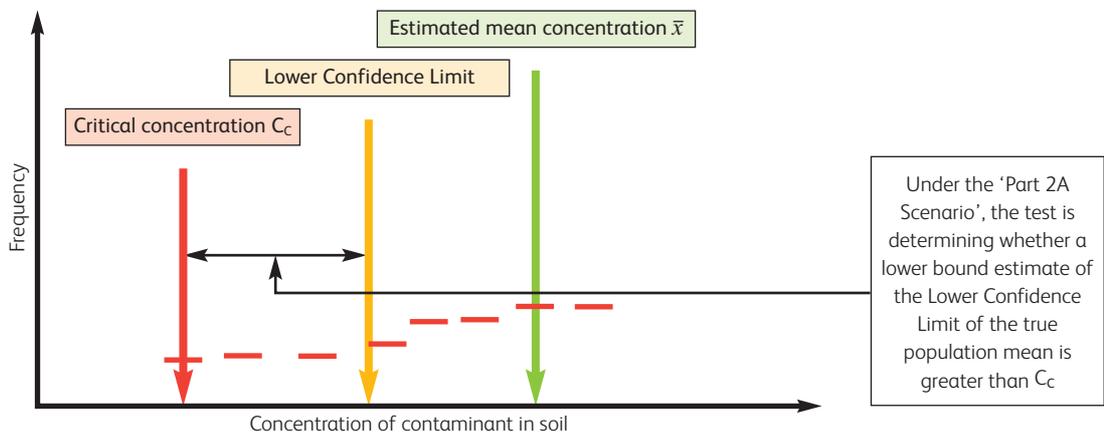
Figure 3 Application of significance tests where data are not normally distributed

Where data are not normally distributed

Planning Scenario



Part 2A Scenario



One other, and potentially very useful, feature of the statistical tests described here is that particular inferences can be drawn from the relationship between the sample mean and critical concentration which allow quick and easy screening of datasets and associated contaminants.

For example, in the planning scenario, if the sample mean is higher than the critical concentration it is clear without further calculation that the UCL of the true population mean will be higher than the critical concentration. In these circumstances it is not possible to reject the Null Hypothesis (see Step 4 in Box 9). This means the assessor can immediately identify which datasets (and associated contaminants) are unlikely to meet the planning test so that appropriate decisions and actions in relation to these data/contaminants can be taken at an early stage. This may involve further data collection and assessment, or remediation.

Similarly, for the Part 2A scenario, if the sample mean is less than the critical concentration, the LCL must also be lower than the critical concentration and the Null Hypothesis cannot be rejected (see Step 4 in Box 10). This means that the assessor can put to one side datasets (and associated contaminants) where Part 2A is unlikely to apply, and concentrate on those more likely to fall within the remit of the regime.

6.4 APPLYING (AN) APPROPRIATE STATISTICAL TEST(S)

This part of the guidance sets out step-by-step procedures for applying significance tests to datasets under the planning and Part 2A scenarios.

The procedures include steps for checking data quality, identifying and dealing with non-detects and outliers, testing for the normality of the data and applying the relevant test as discussed in Section 5.0.

Reference is also made to the following Appendices:

Appendix A – Statistical tables

Appendix B – Directions on the use of one type of outlier test

Appendix C – Directions for checking the normality of data

When using these procedures, readers may find it useful to refer to the list of statistical notation given at the end of this note. When applying the tests, it is also important to note the following:

- superficially, the tests for the different scenarios may appear very similar, however there are many subtle differences between the two and it is important to choose the correct test for the particular application and to progress carefully and systematically through each procedure. Key steps and calculations are highlighted for ease of reference.
- “Less than” and “more than” signs and positive and negative outcomes to calculations should be respected and taken at face value as they are important to both the design and interpretation of the tests; similarly, users should observe the relevant statistical conventions and terminology when reporting the results of testing. For example, the outcome should always be described in terms of the evidence for accepting or rejecting the Null Hypothesis rather than accepting/rejecting the Alternative Hypothesis.
- it is possible to develop simple spreadsheet tools to support statistical testing and these will really be essential where large amounts of data and several contaminants require evaluation. As is usually the case, the quality of such spreadsheet tools should be checked to ensure the correct results will be obtained. In addition to illustrating how to apply the tests, the worked examples given in Appendix D of this note may be helpful for quality control purposes.

6.4.1 Planning scenario

Box 9 sets out the procedure for applying a statistical test to a dataset assuming the planning scenario.

| Box 9 – Applying a statistical test to a dataset under the Planning scenario | |
|--|--|
| Step 0 | Review the dataset and confirm that it satisfies the data quality criteria discussed in Section 4.0 and 5.0 of this guidance. |
| Step 1 | <p>Confirm the key question and associated Null Hypothesis (H_0) and Alternative Hypothesis (H_1) for the Planning scenario:</p> <p>Key question Is there sufficient evidence that the true mean concentration of the contaminant (μ) is less than the critical concentration (C_C)?</p> <p>H_0 $\mu \geq C_C$ i.e., the true mean concentration is equal to, or greater than, the critical concentration</p> <p>H_1 $\mu < C_C$ i.e., the true mean concentration is less than the critical concentration</p> |
| Step 2 | <p>Verify assumptions and robustness with regard to non-detects, outliers and normality as follows:</p> <p>i If the dataset contains:</p> <ul style="list-style-type: none"> • non-detects – proceed with the simple substitution method and/or review zoning decisions as discussed in Section 5.3.1 noting that the choice of substitute value may have a large effect on the outcome of the test where a moderate to large (>10 – 15%) proportion of the data are substituted; • no non-detects – proceed to Step 2ii. <p>ii Investigate statistical outliers and anomalous concentrations as discussed in Section 5.3.2.</p> <p>iii Check the normality of the data distribution following the procedures in Appendix C.</p> |
| Step 3 | Calculate \bar{x} (sample mean) and s (sample standard deviation) – see “Key statistical symbols and terms” for the relevant formulae. |
| Step 4 | <p>If $\bar{x} > C_C$, conclude that H_0 cannot be rejected. Go to Step 14.</p> <p>If $\bar{x} < C_C$, go to Step 5.</p> |
| Step 5 | <p>If the dataset distribution does NOT deviate significantly from normality, follow directions for the one-sample t-test (Step 6).</p> <p>If the dataset distribution deviates significantly from normality, follow directions for the one-sided Chebychev Theorem (Step 10).</p> |

Note the direction of the sign & the short cut to Step 14 if \bar{x} is more than C_C . Since the sample mean is already greater than C_C , we know without further calculations that the UCL will be greater than C_C (see also Figure 2 and 3 for a graphical representation).

Box 9 – Applying a statistical test to a dataset under the Planning scenario *continued*

| | |
|---------|--|
| Step 6 | <p>Apply the one sample t-test²⁴:</p> <p>Calculate the one-sample t statistic, $t_0 = \frac{\bar{x} - C_C}{\frac{s}{\sqrt{n}}}$</p> |
| Step 7 | Use Table A.1 in Appendix A to find $t_{(n-1, 0.95)}$ where n = number of samples in the dataset and $n - 1$ = degrees of freedom. |
| Step 8 | If $t_0 < -t_{(n-1, 0.95)}$, reject H_0 . |
| Step 9 | <p>Estimate p_1 (the level of evidence against H_0)²⁵ as follows:</p> <ul style="list-style-type: none"> i In Table A.1 find t_p, the value of t that is closest to $-t_0$ corresponding to $n - 1$ degrees of freedom. ii Find p_1, the value of $1 - \alpha$ that corresponds to t_p. iii Note that H_0 must not be rejected unless $p_1 \geq 0.95$. Go to Step 14. |
| Step 10 | <p>Apply the one-sided Chebychev Theorem²⁶:</p> <p>Calculate the quantity $k_0 = \frac{\bar{x} - C_C}{\frac{s}{\sqrt{n}}}$</p> |
| Step 11 | Use Table A.2 in Appendix A to find $k_{0.05}$, the value of k corresponding to $\alpha = 0.05$. |
| Step 12 | Let $k_{crit} = -k_{0.05}$. If $k_0 < k_{crit}$, reject H_0 . |

Note “less than” sign & negative value for $t_{(n-1, 0.95)}$ so from Table A.1, if

$$t_{(n-1, 0.95)} = 1.833$$

$$-t_{(n-1, 0.95)} = -1.833$$

Note the negative sign: so from Table A.2, if

$$k_{0.05} = 4.36$$

$$-k_{0.05} = k_{crit} = -4.36$$

Note negative sign for t_0 so from Table A.1, if

$$t_0 = 1.833$$

$$-t_0 = -1.833$$

²⁴ Here the assessor can also calculate the 95% UCL of the sample mean as:

$$UCL_{0.95} = \bar{x} + \left(t_{(n-1, 0.95)} \times \frac{s}{\sqrt{n}} \right)$$

²⁵ Where $p_1 = 1 - p$, and p is the so-called *p-value*

²⁶ The assessor here can also calculate the 95% UCL of the sample mean as:

$$UCL_{0.95} = \bar{x} + \left(k_{(0.05)} \times \frac{s}{\sqrt{n}} \right)$$

Box 9 – Applying a statistical test to a dataset under the Planning scenario *continued*

| | |
|---------|---|
| Step 13 | Estimate p_1 (the level of evidence against H_0) ²⁵ using Table A.2 as follows: i In Table A.2 find k_1 , the value of k closest to $-k_0$. ii Read the value of α_1 , the value of α that corresponds to k_1 . iii Calculate: $p_1 = 1 - \alpha_1$ p_1 represents a conservative (under-) estimate of the evidence against H_0 . iv Note that H_0 must not be rejected unless $p_1 > 0.95$. Go to Step 14. |
| Step 14 | If H_0 is NOT rejected (that is, the evidence suggests that μ is equal to, or greater than, C_C) the developer may have options on how to proceed (see Section 6) including collecting further data and re-running the significance test at the same confidence level, or undertaking remediation on a precautionary basis. |
| Step 15 | Document the process followed and the outcome of the test – see Section 7. |

Note the negative sign so:

If $k_0 = -2.14$

$-k_0 = 2.14$

From Table A.2, $k_1 = 2.13$

$\alpha = 0.18$ (18%) &

$p_1 = 82\%$

6.4.2 Part 2A scenario

Box 10 sets out the procedure for applying a statistical test to a dataset assuming the Part 2A scenario.

| Box 10 – Applying a statistical test to a dataset under the Part 2A scenario | | | | | | | | | | |
|--|---|---|---|--|-------|----------------|---|-------|-------------|--|
| Step 0 | Review the dataset and confirm that it satisfies the data quality criteria discussed in Section 4.0 and Section 5.0 of this guidance. | | | | | | | | | |
| Step 1 | <p>Confirm the key question and associated Null Hypothesis (H_0) and Alternative Hypothesis (H_1) for the Part 2A scenario:</p> <table border="0"> <tr> <td>Key question</td> <td colspan="2">Is there sufficient evidence that the true mean concentration of the contaminant (μ) is greater than the critical concentration (C_C)?</td> </tr> <tr> <td>H_0</td> <td>$\mu \leq C_C$</td> <td>i.e., the true mean concentration is equal to, or less than, the critical concentration</td> </tr> <tr> <td>H_1</td> <td>$\mu > C_C$</td> <td>i.e., the true mean concentration is greater than the critical concentration</td> </tr> </table> | Key question | Is there sufficient evidence that the true mean concentration of the contaminant (μ) is greater than the critical concentration (C_C)? | | H_0 | $\mu \leq C_C$ | i.e., the true mean concentration is equal to, or less than, the critical concentration | H_1 | $\mu > C_C$ | i.e., the true mean concentration is greater than the critical concentration |
| Key question | Is there sufficient evidence that the true mean concentration of the contaminant (μ) is greater than the critical concentration (C_C)? | | | | | | | | | |
| H_0 | $\mu \leq C_C$ | i.e., the true mean concentration is equal to, or less than, the critical concentration | | | | | | | | |
| H_1 | $\mu > C_C$ | i.e., the true mean concentration is greater than the critical concentration | | | | | | | | |
| Step 2 | <p>Verify assumptions and robustness with regard to non-detects, outliers and normality as follows:</p> <ol style="list-style-type: none"> i If the dataset contains: <ul style="list-style-type: none"> • non-detects – proceed with the simple substitution method and/or review zoning decisions as discussed in Section 5.3.1 noting that the choice of substitute value may have a large effect on the outcome of the test where a moderate to large (>10 - 15%) proportion of the data are substituted; • no non-detects – proceed to Step 2ii. ii Investigate statistical outliers and anomalous concentrations as discussed in Section 5.3.2. iii Check the normality of the data distribution following the procedures in Appendix C. | | | | | | | | | |
| Step 3 | Calculate \bar{x} (sample mean) and s (sample standard deviation) – see “Key statistical symbols and terms” for the relevant formulae. | | | | | | | | | |
| Step 4 | <p>If $\bar{x} < C_C$, conclude that H_0 cannot be rejected. Go to Step 15.</p> <p>If $\bar{x} > C_C$, go to Step 5.</p> | | | | | | | | | |

Note the direction of the sign & the short cut to Step 15 if \bar{x} is less than C_C . Since the sample mean is less than C_C , we know already without further calculations that the LCL will be less than C_C (see also Figure 2 and 3 for a graphical representation).

Box 10 – Applying a statistical test to a dataset under the Part 2A scenario *continued*

| | |
|---------|---|
| Step 5 | <p>If the dataset distribution does NOT deviate significantly from normality, follow directions for the one-sample t-test (Step 6).</p> <p>If the dataset distribution deviates significantly from normality, follow directions for the one-sided Chebychev Theorem (Step 10).</p> |
| Step 6 | <p>Apply the one sample t-test²⁷:</p> <p>Calculate the one-sample t statistic, $t_0 = \frac{\bar{x} - C_C}{\frac{s}{\sqrt{n}}}$</p> |
| Step 7 | Use Table A.1 in Appendix A to find $t_{(n-1, 0.95)}$ where n = number of samples in the dataset and $n - 1$ = degrees of freedom. |
| Step 8 | If $t_0 > t_{(n-1, 0.95)}$, reject H_0 . |
| Step 9 | <p>Estimate p_1 (the level of evidence against H_0)²⁵ as follows:</p> <ul style="list-style-type: none"> i In Table A.1 find t_p, the value of t that is closest to t_0 corresponding to $n - 1$ degrees of freedom. ii Find p_1, the value of $1 - \alpha$ that corresponds to t_p. iii Note that H_0 must not be rejected unless $p_1 \geq 0.95$. Go to Step 14. |
| Step 10 | <p>Apply the one-sided Chebychev Theorem²⁸:</p> <p>Calculate the quantity $k_0 = \frac{\bar{x} - C_C}{\frac{s}{\sqrt{n}}}$</p> |
| Step 11 | Use Table A.2 in Appendix A to find $k_{0.05}$, the value of k corresponding to $\alpha = 0.05$. |
| Step 12 | Let $k_{crit} = k_{0.05}$. If $k_0 > k_{crit}$, reject H_0 . |

Note there is no negative sign here. In the case of $n-1=15$ degrees of freedom and $t_0 = 2.129$, from Table A.1, the value for t_p is $t_p = 2.131$

²⁷ The assessor here can also calculate the LCL as:

$$LCL_{0.95} = \bar{x} - \left(t_{(n-1, 0.95)} \times \frac{s}{\sqrt{n}} \right)$$

²⁸ The assessor here can also calculate the LCL as:

$$LCL_{0.95} = \bar{x} - \left(k_{(0.05)} \times \frac{s}{\sqrt{n}} \right)$$

Box 10 – Applying a statistical test to a dataset under the Part 2A scenario *continued*

| | |
|---------|---|
| Step 13 | <p>Estimate p_1 and P_1 (the lower and upper bounds of the evidence against H_0) as follows:</p> <ul style="list-style-type: none"> i To calculate p_1 (lower bound of evidence against H_0) use Table A.2 to find k_1, the value of k closest to k_0. ii Read the value of α_1, the value of α that corresponds to k_1. iii Calculate: $p_1 = 1 - \alpha_1$ p_1 represents the lower bound of the evidence against H_0. iv To calculate P_1 (the upper bound of evidence against H_0) enter Table A.1 at $n - 1$ degrees of freedom. v In this row, find t_p, the value of t closest to t_0. vi Find P_1, the value of $1 - \alpha$ corresponding to t_p. vii P_1 represents the upper bound of the evidence against H_0. viii The level of evidence against H_0, therefore, is a probability value within the range $[p_1 : P_1]$. |
| Step 14 | <p>If H_0 is NOT rejected at a 95 % level of confidence, the regulator should reapply the test on a ‘balance of probabilities’ basis (see Section 6.2). To do this:</p> <ul style="list-style-type: none"> i If the one-sample t-test was adopted (Step 6), estimate the evidence against H_0 as in Step 9 using Table A.1. Provided $p_1 > 0.51$, H_0 can be rejected. ii If the Chebychev Theorem was adopted, estimate the evidence against H_0 as in Step 13 using Tables A.1 and A.2 and make a judgment about whether H_0 can be rejected, bearing in mind that the balance of probabilities implies a probability of more than 0.50. |
| Step 15 | <p>If the outcome of the test suggests that H_0 should not be rejected, the regulator may decide to collect additional representative samples of the land being assessed and re-run the significance test. <i>[Note, however, that the aim of Part 2A is to ensure that resources are directed at the most pressing problems first. If a reasonable amount of sampling has failed to indicate that land could be contaminated land (even on a ‘balance of probabilities basis’), regulators may consider that assessing other priority land in their areas represents a better use of resources and is more in keeping with Part 2A requirements than carrying out exhaustive sampling of a particular area of land].</i></p> |
| Step 16 | <p>Document the process followed and the outcome of the test – see Section 7.</p> |

Note there is no negative sign here.
In the case of $k_0 = 2.40$, the value for k_1 from Table A.2, should be $k_1 = 2.38$, corresponding to $\alpha_1 = 0.15$



6.5 WORKED EXAMPLES

Appendix D contains two worked examples which illustrate how the procedures can be applied and the results interpreted using real data for each of the two scenarios. For ease of reference, the results of calculations are shown for each step in the procedure.

7.0 Documenting decisions

It is likely that the statistical tests described in this guidance will be conducted as part of wider risk assessment or remediation projects. Much of the key contextual information that should be recorded during the formal documentation of statistical testing is also relevant to risk assessment and remediation activities and therefore may have already been addressed as part of the risk assessment or remediation report.

Irrespective of the form of reporting, however, the following matters should be clearly addressed in any written account of the purpose and outcome of statistical testing:

- the regulatory context within which the test is conducted;
- the rationale for collecting and testing samples including confirmation of the scale of sampling and that unbiased sample data have been used;
- details of sampling methods used;
- the process followed to create relevant datasets and the outcome of data quality checks including the identification of any data that have been discarded (with full justification);
- the methods used to handle non-detects and outliers including identification of any substitute values, any discarded data (with justification) and records of any revised datasets to which statistical tests have been applied;
- the outcome of testing for normality (including details/justification of the confidence level at which any relevant tests were conducted);
- the Null and Alternative Hypotheses that have been used to frame the tests;
- the methods used to calculate key statistics;
- the type of statistical test used with justification;
- the outcome of the test;
- the interpretation of the outcome of the test;
- recommended next step(s).

References

- British Standards Institution, Investigation of potentially contaminated sites, Code of Practice, BS:10175, 2001
- Daniels, W M & Higgins, N A (2002) Environmental distributions and the practical utilisation of censored environmental data, NRPB report NRPB-R02008
- Defra Circular 01/2006, Environmental Protection Act 1990: Part 2A, Contaminated Land, September 2006
- Defra, Assessing risks from land contamination – a proportionate approach. Soil Guideline values: the Way Forward. CLAN 06/2006, November 2006
- Defra/Environment Agency, Model procedures for the management of land contamination, CLR11, 2004
- Defra/Environment Agency, Overview of the development of guideline values and related research, CLR7, 2002
- Environment Agency, Guide to good practice for the development of conceptual models and the selection and application of mathematical models of contaminant transport processes in the subsurface, 2001
- Environment Agency, Secondary model procedure for the development of appropriate soil sampling strategies for land contamination, R&D Technical Report P5-066/TR, 2000
- Environment Agency, Technical aspects of site investigation in relation to land contamination, Volume 1, R&D Technical Report P5-065/TR, 2000 & Volume 2 Text Supplements
- Gilbert, R O (1987) Statistical methods for environmental pollution monitoring. John Wiley and Sons
- Grubb, F (1969) Procedures for detecting outlying observations in samples. *Technometrics*, 11, 1-21
- Masi, P and P. Morgan, Statistical assessment of contaminated land: some implications of the 'mean value test', CL:AIRE Technical Bulletin TB12, October 2006
- Office of the Deputy Prime Minister, Planning Policy Statement 23: Planning and Pollution Control, Annex 2, Development on Land Affected by Contamination, 2004
- Ramsey, M. H., Improving the reliability of contaminated land assessment using statistical methods: Part 1 – Basic Principles and Concepts, CL:AIRE Technical Bulletin TB7, March 2004
- Scottish Executive, Paper SE/2006/44, Environmental Protection Act 1990: Part IIA, Contaminated Land Statutory Guidance: Edition 2, 2006
- Scottish Executive, Planning Advice Note 33 – Development of Contaminated Land, 2000
- Scottish Executive, Planning Advice Note 51 – Planning and Environmental Protection Regulation, 2006
- Shapiro, S.S. and Francia, R.S. (1972), An approximate analysis of variance test for normality. *Journal of American Statistical Association*, 67(337), 215-216
- Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611
- Singh A.K., Singh A. and Engelhardt M. (1997), The lognormal distribution in environmental applications, US EPA Technology Support Center Issue EPA/600/R-97/006
- United States Environmental Protection Agency, Office of Environmental Information, Washington DC, Data quality assessment: statistical methods for practitioners, EPA QA/G-9S, EPA/3240/B-06/003, February 2006
- United States Environmental Protection Agency, Office of Environmental Information, Washington DC 20460, Guidance on choosing a sampling design for environmental data collection, EPA QA/G-5S, 2002
- Welsh Assembly Government, Part 2A Statutory Guidance on Contaminated Land (2006), 2006
- Welsh Assembly Government, Planning Policy Wales, 2002

Key statistical symbols and terms

| | | | |
|-----------|--|---------------------|--|
| μ | True (unknown) mean concentration of the contaminant in soil | H_0 | Null Hypothesis – a theory put forward for testing because it is believed to be true (but has yet to be proved) or because it creates a basis for an argument or proposition |
| σ | True (unknown) standard deviation of the concentration of the contaminant in soil | H_1 | Alternative Hypothesis – the converse of the Null Hypothesis and the question that the test is designed to answer |
| x_i | Measured concentration of the contaminant in the i^{th} sample | α | Probability of wrongly rejecting the Null Hypothesis when it is true |
| \bar{x} | Sample mean (average concentration of contaminant concentrations for the samples in the dataset under scrutiny) | β | Probability of not rejecting the Null Hypothesis when it is false |
| | $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ | P | The probability of obtaining the computed test statistic (e.g. t_0) or one even less likely when the Null Hypothesis is true |
| s | Sample unbiased standard deviation (the standard deviation of the contaminant concentrations in the dataset under scrutiny). | p_1 | The level of evidence against the Null Hypothesis – the stronger the evidence against the Null Hypothesis the greater the value of p_1 |
| | $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ | t_0 | Test statistic for the one-sample t-test |
| C_C | Critical concentration | k_0 | Test statistic for the test method based on the one-sided Chebychev inequality |
| n | Sample size (the number of data in the dataset under scrutiny) | $t_{(n-1), \alpha}$ | α^{th} critical value of the Student's t-distribution with $n-1$ degrees of freedom |
| UCL | Upper Confidence Limit (of the true population mean) | k_α | α^{th} critical value based on the one-sided Chebychev Theorem |
| LCL | Lower Confidence Limit (of the true population mean) | | |



Appendices

Appendix A Statistical tables

Table A.1 $t_{n-1, 1-\alpha}$ for the one-sample t-test

Degrees of freedom = number of data in dataset minus 1

| Deg of freedom | $1-\alpha$ | | | | | | | | | | | | | | |
|----------------|------------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--|--|--|--|
| | 0.51 | 0.6 | 0.7 | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | | | | |
| 1 | 0.031 | 0.325 | 0.727 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | | | | |
| 2 | 0.028 | 0.289 | 0.617 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | | | | |
| 3 | 0.027 | 0.277 | 0.584 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | | | | |
| 4 | 0.027 | 0.271 | 0.569 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | | | | |
| 5 | 0.026 | 0.267 | 0.559 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | | | | |
| 6 | 0.026 | 0.265 | 0.553 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | | | | |
| 7 | 0.026 | 0.263 | 0.549 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | | | | |
| 8 | 0.026 | 0.262 | 0.546 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | | | | |
| 9 | 0.026 | 0.261 | 0.543 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | | | | |
| 10 | 0.026 | 0.260 | 0.542 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | | | | |
| 11 | 0.026 | 0.260 | 0.540 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | | | | |
| 12 | 0.026 | 0.259 | 0.539 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | | | | |
| 13 | 0.026 | 0.259 | 0.538 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | | | | |
| 14 | 0.026 | 0.258 | 0.537 | 0.692 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | | | | |
| 15 | 0.025 | 0.258 | 0.536 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | | | | |
| 16 | 0.025 | 0.258 | 0.535 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | | | | |
| 17 | 0.025 | 0.257 | 0.534 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | | | | |
| 18 | 0.025 | 0.257 | 0.534 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | | | | |
| 19 | 0.025 | 0.257 | 0.533 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | | | | |
| 20 | 0.025 | 0.257 | 0.533 | 0.687 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | | | | |
| 21 | 0.025 | 0.257 | 0.532 | 0.686 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | | | | |
| 22 | 0.025 | 0.256 | 0.532 | 0.686 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | | | | |
| 23 | 0.025 | 0.256 | 0.532 | 0.685 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | | | | |
| 24 | 0.025 | 0.256 | 0.531 | 0.685 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | | | | |
| 25 | 0.025 | 0.256 | 0.531 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | | | | |

Table A.1 $t_{n-1,1-\alpha}$ for the one-sample t-test continued

| Deg of freedom | 1- α | | | | | | | | | | | | |
|----------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--|--|
| | 0.51 | 0.6 | 0.7 | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | | |
| 26 | 0.025 | 0.256 | 0.531 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | | |
| 27 | 0.025 | 0.256 | 0.531 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | | |
| 28 | 0.025 | 0.256 | 0.530 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | | |
| 29 | 0.025 | 0.256 | 0.530 | 0.683 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | | |
| 30 | 0.025 | 0.256 | 0.530 | 0.683 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | | |
| 40 | 0.025 | 0.255 | 0.529 | 0.681 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | | |
| 50 | 0.025 | 0.255 | 0.528 | 0.679 | 0.849 | 1.047 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 | | |
| 60 | 0.025 | 0.254 | 0.527 | 0.679 | 0.848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | | |
| 70 | 0.025 | 0.254 | 0.527 | 0.678 | 0.847 | 1.044 | 1.294 | 1.667 | 1.994 | 2.381 | 2.648 | | |
| 80 | 0.025 | 0.254 | 0.526 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.374 | 2.639 | | |
| 90 | 0.025 | 0.254 | 0.526 | 0.677 | 0.846 | 1.042 | 1.291 | 1.662 | 1.987 | 2.368 | 2.632 | | |
| 100 | 0.025 | 0.254 | 0.526 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.364 | 2.626 | | |
| 120 | 0.025 | 0.254 | 0.526 | 0.677 | 0.845 | 1.041 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | | |
| 140 | 0.025 | 0.254 | 0.526 | 0.676 | 0.844 | 1.040 | 1.288 | 1.656 | 1.977 | 2.353 | 2.611 | | |
| 160 | 0.025 | 0.254 | 0.525 | 0.676 | 0.844 | 1.040 | 1.287 | 1.654 | 1.975 | 2.350 | 2.607 | | |
| 200 | 0.025 | 0.254 | 0.525 | 0.676 | 0.843 | 1.039 | 1.286 | 1.653 | 1.972 | 2.345 | 2.601 | | |
| 10000000 | 0.025 | 0.253 | 0.524 | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | | |

Table A.2 Alpha levels for the test based on the Chebychev Theorem

From Example 1,
 $k_0 = -5.699$
 therefore $-k_0 = 5.699$
 therefore $k_1 = 5.69$
 therefore $\alpha = 0.03$ (3%)

| alpha | k | alpha | k |
|-------|------|-------|------|
| 0.01 | 9.95 | 0.26 | 1.69 |
| 0.02 | 7.00 | 0.27 | 1.64 |
| 0.03 | 5.69 | 0.28 | 1.6 |
| 0.04 | 4.90 | 0.29 | 1.56 |
| 0.05 | 4.36 | 0.30 | 1.53 |
| 0.06 | 3.96 | 0.31 | 1.49 |
| 0.07 | 3.64 | 0.32 | 1.46 |
| 0.08 | 3.39 | 0.33 | 1.42 |
| 0.09 | 3.18 | 0.34 | 1.39 |
| 0.10 | 3.00 | 0.35 | 1.36 |
| 0.11 | 2.84 | 0.36 | 1.33 |
| 0.12 | 2.71 | 0.37 | 1.30 |
| 0.13 | 2.59 | 0.38 | 1.28 |
| 0.14 | 2.48 | 0.39 | 1.25 |
| 0.15 | 2.38 | 0.40 | 1.22 |
| 0.16 | 2.29 | 0.41 | 1.20 |
| 0.17 | 2.21 | 0.42 | 1.18 |
| 0.18 | 2.13 | 0.43 | 1.15 |
| 0.19 | 2.06 | 0.44 | 1.13 |
| 0.20 | 2.00 | 0.45 | 1.11 |
| 0.21 | 1.94 | 0.46 | 1.08 |
| 0.22 | 1.88 | 0.47 | 1.06 |
| 0.23 | 1.83 | 0.48 | 1.04 |
| 0.24 | 1.78 | 0.49 | 1.02 |
| | 0.25 | 1.73 | |

Appendix B Directions on the use of one type of outlier test

Statistical Notation for the Outlier Test (Grubbs Test)^{29,30}

The meaning of the symbols used in the outlier test is given below:

- μ true (unknown) mean concentration of contaminant in soil.
- x_i the soil concentration value in the i^{th} sample.
- \bar{x} sample mean (the average of the concentrations in the dataset under scrutiny):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- s sample unbiased standard deviation (the standard deviation of the concentrations in the dataset under scrutiny):

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- n sample size, the number of data in the dataset under scrutiny.

Procedure for the Outlier Test

1. The Null Hypothesis:
There are no outliers in the dataset
2. The Alternative Hypothesis:
The maximum value in the dataset is an outlier
3. Verify assumptions:
 - i) Normality
This outlier test assumes that the other data values, except for the suspect observation, are normally distributed. Use the procedure described in Appendix C to assess the normality of the rest of the dataset, once the suspected outlier(s) has been eliminated.
4. If the dataset proves to be not normally distributed, consider transforming the data by taking the natural logarithm and check the normality of the transformed dataset.
5. Set $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ to be the n observations (or the natural logarithms of the n -observations) ordered from the smallest to the largest.
6. Calculate \bar{x} and s , the sample arithmetic average and unbiased standard deviation of x_1, x_2, \dots, x_n respectively.
7. Take the maximum value, x_n and standardise it by subtracting the mean \bar{x} and dividing by the standard deviation s :

$$T_n = \frac{x_n - \bar{x}}{s}$$
8. Choose a level of significance $\alpha = 0.01$, $\alpha = 0.05$ or $\alpha = 0.1$ to carry out the outlier test (note that the higher α , the more likely it is that high values will be identified as outliers in the dataset under scrutiny) and use Table B.1 to find T_{crit} , the critical value corresponding to the sample size n and the chosen level of significance α .
9. If $T_n > T_{crit}$ it is justifiable to consider x_n as an outlier.
10. Go back to Step 3 and repeat the procedure for the second highest value in the dataset and continue until T_n is found to be less than or equal to T_{crit} .
11. **Identifying a value as an outlier does not necessarily mean that the value should be removed from the dataset. Treat the identified outliers as described in Section 5 of the guidance above.**

Note: Sometimes the presence of more than one anomalous high value in a dataset can prevent the outlier test identifying the maximum value as an actual outlier. When more than one anomalous high value is considered likely to affect a dataset, good practice is to test first whether the lower of these high values is an actual outlier.

²⁹ This is the same test as the Maximum Value Test given in CLR7 2002

³⁰ Note that this is a test for the upper outlier only

Table B.1 Critical values for the outlier test

| Sample size | Level of Significance α | | | Sample size | Level of Significance α | | |
|-------------|--------------------------------|-------|-------|-------------|--------------------------------|-------|-------|
| | 0.01 | 0.05 | 0.1 | | 0.01 | 0.05 | 0.1 |
| 3 | 1.155 | 1.153 | 1.148 | 27 | 3.049 | 2.698 | 2.52 |
| 4 | 1.492 | 1.462 | 1.425 | 28 | 3.068 | 2.714 | 2.536 |
| 5 | 1.749 | 1.671 | 1.602 | 29 | 3.086 | 2.73 | 2.551 |
| 6 | 1.944 | 1.822 | 1.729 | 30 | 3.103 | 2.745 | 2.565 |
| 7 | 2.097 | 1.938 | 1.828 | 31 | 3.119 | 2.76 | 2.579 |
| 8 | 2.221 | 2.032 | 1.909 | 32 | 3.135 | 2.773 | 2.592 |
| 9 | 2.323 | 2.11 | 1.977 | 33 | 3.15 | 2.787 | 2.605 |
| 10 | 2.41 | 2.176 | 2.036 | 34 | 3.164 | 2.799 | 2.618 |
| 11 | 2.484 | 2.234 | 2.088 | 35 | 3.178 | 2.812 | 2.63 |
| 12 | 2.549 | 2.285 | 2.134 | 36 | 3.191 | 2.824 | 2.641 |
| 13 | 2.607 | 2.331 | 2.176 | 37 | 3.204 | 2.835 | 2.652 |
| 14 | 2.658 | 2.372 | 2.213 | 38 | 3.216 | 2.846 | 2.663 |
| 15 | 2.705 | 2.409 | 2.248 | 39 | 3.228 | 2.857 | 2.674 |
| 16 | 2.747 | 2.443 | 2.279 | 40 | 3.239 | 2.868 | 2.684 |
| 17 | 2.785 | 2.475 | 2.309 | 41 | 3.251 | 2.878 | 2.694 |
| 18 | 2.821 | 2.504 | 2.336 | 42 | 3.261 | 2.887 | 2.704 |
| 19 | 2.853 | 2.531 | 2.361 | 43 | 3.272 | 2.897 | 2.713 |
| 20 | 2.884 | 2.557 | 2.385 | 44 | 3.282 | 2.906 | 2.722 |
| 21 | 2.912 | 2.58 | 2.408 | 45 | 3.292 | 2.915 | 2.731 |
| 22 | 2.939 | 2.603 | 2.429 | 46 | 3.301 | 2.924 | 2.739 |
| 23 | 2.963 | 2.624 | 2.449 | 47 | 3.31 | 2.933 | 2.748 |
| 24 | 2.987 | 2.644 | 2.468 | 48 | 3.319 | 2.941 | 2.756 |
| 25 | 3.009 | 2.663 | 2.486 | 49 | 3.328 | 2.949 | 2.764 |
| 26 | 3.029 | 2.681 | 2.503 | 50 | 3.337 | 2.957 | 2.772 |

From Example 1:
For n = 33 and $\alpha = 0.05$, $T_{crit} = 2.787$

Appendix C Directions for checking the normality of data

Checking the Normality of Data

It is recommended that both visual and numerical assessments are undertaken to check the assumption regarding the distribution of data. Although a statistical test for normality can identify when deviation from normality starts to be significant, only a visual assessment of the data can support the interpretation of which part of a dataset differs most from the normal distribution.

An intuitive and powerful means for visually assessing normality of a dataset is the so-called probability plot, and an appropriate statistical test for datasets with a sample size of up to 50 data is the Shapiro-Wilk test.

The Shapiro-Wilk test is sensitive to the departure from normality in the tails of a distribution and therefore provides a reliable check on whether a test based on normality (like the one-sample t-test) can be applied on the dataset under scrutiny. Other statistical tests are available for testing normality of a dataset, and these should be considered for datasets larger than 50 data (Shapiro and Francia, 1972).³¹

When testing for the normality of an approximately normal large dataset (for example, a dataset that appears to be normal on a visual assessment) it is sometimes possible for the test to fail (i.e. normality is rejected). If a test against the mean (such as those presented in the main text) is to be applied to a dataset, failure to prove normality suggests that the Chebychev theorem should be used rather than the one-sample t-test.

This is likely not to be an issue of practical concern at most sites where, after appropriate site zoning, datasets do not tend to be so large, but it should be carefully considered if the dataset fails the Shapiro-Wilk test while appearing normal on visual assessment.

Statistical Notation for the Normality Test

The meaning of the symbols used in the normality tests is given below:

μ true (unknown) mean concentration of the contaminant in soil

x_i measured concentration of the contaminant in the i^{th} sample

\bar{x} sample mean (average of the concentrations in the dataset under scrutiny):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

s sample unbiased standard deviation (the standard deviation of the concentrations in the dataset under scrutiny):

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

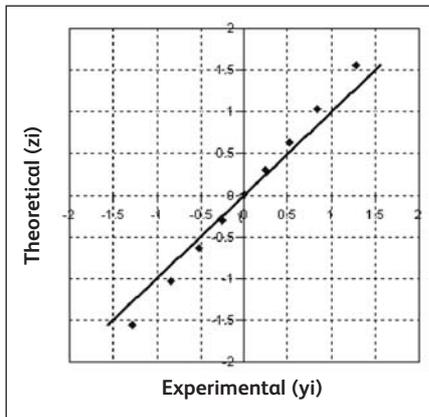
n sample size, the number of data in the dataset under scrutiny

³¹ Shapiro, S.S. and Francia, R.S. (1972) *An approximate analysis of variance test for normality*. Journal of American Statistical Association, 67(337), 215-216

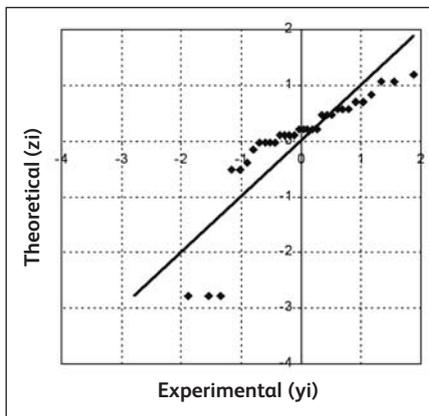
Procedure for a Probability Plot

A probability plot is a graph on which the percentiles of a dataset are plotted versus those of a theoretical (e.g. normal) distribution (see the example plots below).

Data approximately normally distributed



Data not normally distributed



The rationale behind building such a graph is that if the empirical and the theoretical distributions are the same, the points in the probability plot are aligned along a 45 degree line.

Whenever a probability plot is used to test normality, the dataset under scrutiny can be considered approximately normal where there is no significant disagreement between the points of the probability plot and the 45 degree line. Particular attention should be paid to the high values (upper tail) in a probability plot: a disagreement between these values and the 45 degrees line is often an indication that the dataset under scrutiny cannot be considered approximately normal.

To draw a probability plot:

1. Order the data x_i ($i = 1, 2, \dots, n$) from the smallest to the largest ($x_1 < x_2 < \dots < x_n$)
2. Calculate the values y_i by subtracting the mean \bar{x} of the dataset from each x_i and divide the result by the unbiased standard deviation s :

$$y_i = \frac{x_i - \bar{x}}{s}; \quad i = 1, 2, \dots, n$$

3. Estimate the cumulative frequency of all the y_i by calculating the quantity:

$$q_i = \frac{i}{n + 1}$$

From Example 1, $n = 33$,
 $q_1 = 1 \div 34$; $q_2 = 2 \div 34$ etc.

4. Find the quantile of the standard normal distribution as:

$$z_i = N^{-1}\left(\frac{i}{n + 1}\right)$$

N^{-1} can be calculated using the NORMSINV function in Excel. From Example 1, for $i = 5$, we can calculate z_5 by writing the formula = NORMSINV(5/34), which gives $z_5 = -1.049$

Where N^{-1} is the inverse of the standard normal cumulative distribution.

5. Plot y_i (the experimental quantiles of x_i) against the z_i (theoretical quantiles of a normal distribution). If x_i ($i = 1, 2, \dots, n$) are drawn from a normal distribution, the plot should approximate to a straight 45 degree line.

Procedure for the Shapiro-Wilk Normality Test

The Shapiro-Wilk test is a method for testing the distributional assumption of normality by means of an analysis of variance type procedure (Shapiro and Wilk, 1965).³² This test is applicable to a dataset of up to 50 data. For larger datasets an alternative normality test should be considered.

Given a set of data: x_i ($i = 1, 2, \dots, n$), the Shapiro-Wilk test is applied as follows:

1. Order the data so that $x_1 < x_2 < \dots < x_n$;
2. Calculate the average \bar{x} of the data;
3. Calculate the quantity S^2 (Shapiro notation) as:

$$S^2 = \sum_{i=1}^n (x_i - \bar{x})^2;$$

Calculate for each value of x from 1 to n

4. Compute:

$$b = \sum_{i=1}^k a_{n-i+1} (x_{n-i+1} - x_i)$$

From Example 2, $n = 10$, $k = 5$ and each value of $(x_{n-i+1} - x_i)$ is: $(x_{10} - x_1)$, $(x_9 - x_2)$, $(x_8 - x_3)$, $(x_7 - x_4)$, $(x_6 - x_5)$,

where:

k is the greatest integer (whole number) less than or equal to $n/2$;
 a_{n-i+1} are given in Table C.1 below.

5. Compute the test statistic:

$$W = \frac{b^2}{S^2}$$

6. The data should not be considered as normal if the Shapiro-Wilk statistic W is less than the critical values reported in Table C.2 below.

³² Shapiro, S.S. and Wilk, M.B. (1965) *An analysis of variance test for normality (complete samples)*. *Biometrika*, 52, 591-611

Table C.1 α_{n-i+1} Coefficients for the Shapiro-Wilk Test

| i/n | α_2 | α_3 | α_4 | α_5 | α_6 | α_7 | α_8 | α_9 | α_{10} | α_{11} | α_{12} | α_{13} | α_{14} | α_{15} |
|-------|------------|------------|------------|------------|------------|------------|------------|------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1 | 0.707 | 0.707 | 0.687 | 0.665 | 0.643 | 0.623 | 0.605 | 0.589 | 0.574 | 0.56 | 0.548 | 0.536 | 0.525 | 0.515 |
| 2 | | 0 | 0.168 | 0.241 | 0.281 | 0.303 | 0.316 | 0.324 | 0.329 | 0.332 | 0.333 | 0.333 | 0.332 | 0.331 |
| 3 | | | | 0 | 0.088 | 0.14 | 0.174 | 0.198 | 0.214 | 0.226 | 0.235 | 0.241 | 0.246 | 0.25 |
| 4 | | | | | | 0 | 0.056 | 0.095 | 0.122 | 0.143 | 0.159 | 0.171 | 0.18 | 0.188 |
| 5 | | | | | | | | 0 | 0.04 | 0.07 | 0.092 | 0.11 | 0.124 | 0.135 |
| 6 | | | | | | | | | | 0 | 0.03 | 0.054 | 0.073 | 0.088 |
| 7 | | | | | | | | | | | | 0 | 0.024 | 0.043 |
| 8 | | | | | | | | | | | | | | 0 |

| i/n | α_{16} | α_{17} | α_{18} | α_{19} | α_{20} | α_{21} | α_{22} | α_{23} | α_{24} | α_{25} | α_{26} | α_{27} | α_{28} | α_{29} |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1 | 0.506 | 0.497 | 0.489 | 0.481 | 0.473 | 0.464 | 0.459 | 0.454 | 0.449 | 0.445 | 0.441 | 0.437 | 0.433 | 0.429 |
| 2 | 0.329 | 0.327 | 0.325 | 0.323 | 0.321 | 0.319 | 0.316 | 0.313 | 0.31 | 0.307 | 0.304 | 0.302 | 0.299 | 0.297 |
| 3 | 0.252 | 0.254 | 0.255 | 0.256 | 0.257 | 0.258 | 0.257 | 0.256 | 0.255 | 0.254 | 0.253 | 0.252 | 0.251 | 0.25 |
| 4 | 0.194 | 0.199 | 0.203 | 0.206 | 0.209 | 0.212 | 0.213 | 0.214 | 0.215 | 0.215 | 0.215 | 0.215 | 0.215 | 0.215 |
| 5 | 0.145 | 0.152 | 0.159 | 0.164 | 0.169 | 0.174 | 0.176 | 0.179 | 0.181 | 0.182 | 0.184 | 0.185 | 0.186 | 0.186 |
| 6 | 0.101 | 0.111 | 0.12 | 0.127 | 0.133 | 0.14 | 0.144 | 0.148 | 0.151 | 0.154 | 0.156 | 0.158 | 0.16 | 0.162 |
| 7 | 0.059 | 0.073 | 0.084 | 0.093 | 0.101 | 0.109 | 0.115 | 0.12 | 0.125 | 0.128 | 0.132 | 0.135 | 0.137 | 0.14 |
| 8 | 0.02 | 0.036 | 0.05 | 0.061 | 0.071 | 0.08 | 0.088 | 0.094 | 0.1 | 0.105 | 0.109 | 0.113 | 0.116 | 0.119 |
| 9 | | 0 | 0.016 | 0.03 | 0.042 | 0.053 | 0.062 | 0.07 | 0.076 | 0.082 | 0.088 | 0.092 | 0.097 | 0.1 |
| 10 | | | | 0 | 0.014 | 0.026 | 0.037 | 0.046 | 0.054 | 0.061 | 0.067 | 0.073 | 0.078 | 0.082 |
| 11 | | | | | | 0 | 0.012 | 0.023 | 0.032 | 0.04 | 0.048 | 0.054 | 0.06 | 0.065 |
| 12 | | | | | | | | 0 | 0.011 | 0.02 | 0.028 | 0.036 | 0.042 | 0.048 |
| 13 | | | | | | | | | | 0 | 0.009 | 0.018 | 0.025 | 0.032 |
| 14 | | | | | | | | | | | | 0 | 0.008 | 0.016 |
| 15 | | | | | | | | | | | | | | 0 |

Use these values as the first, second and third α values where $n = 6$

Use these values if $n = 16$

From USEPA (1992) after Shapiro and Wilk (1965)

Table C.1 $\alpha_{n,i+1}$ Coefficients for the Shapiro-Wilk Test continued

| i/n | α_{30} | α_{31} | α_{32} | α_{33} | α_{34} | α_{35} | α_{36} | α_{37} | α_{38} | α_{39} | α_{40} | α_{41} | α_{42} | α_{43} |
|-----|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1 | 0.425 | 0.422 | 0.419 | 0.416 | 0.413 | 0.41 | 0.407 | 0.404 | 0.402 | 0.399 | 0.396 | 0.394 | 0.392 | 0.389 |
| 2 | 0.294 | 0.292 | 0.29 | 0.288 | 0.285 | 0.283 | 0.281 | 0.279 | 0.277 | 0.276 | 0.274 | 0.272 | 0.27 | 0.268 |
| 3 | 0.249 | 0.248 | 0.246 | 0.245 | 0.244 | 0.243 | 0.242 | 0.24 | 0.239 | 0.238 | 0.237 | 0.236 | 0.235 | 0.233 |
| 4 | 0.215 | 0.215 | 0.214 | 0.214 | 0.213 | 0.213 | 0.212 | 0.212 | 0.211 | 0.21 | 0.21 | 0.209 | 0.209 | 0.208 |
| 5 | 0.187 | 0.187 | 0.188 | 0.188 | 0.188 | 0.188 | 0.188 | 0.188 | 0.188 | 0.188 | 0.188 | 0.188 | 0.187 | 0.187 |
| 6 | 0.163 | 0.164 | 0.165 | 0.166 | 0.167 | 0.167 | 0.168 | 0.168 | 0.169 | 0.169 | 0.169 | 0.169 | 0.169 | 0.17 |
| 7 | 0.142 | 0.143 | 0.145 | 0.146 | 0.148 | 0.149 | 0.15 | 0.15 | 0.151 | 0.152 | 0.153 | 0.153 | 0.154 | 0.154 |
| 8 | 0.122 | 0.124 | 0.127 | 0.128 | 0.13 | 0.132 | 0.133 | 0.134 | 0.136 | 0.137 | 0.138 | 0.138 | 0.139 | 0.14 |
| 9 | 0.104 | 0.107 | 0.109 | 0.112 | 0.114 | 0.116 | 0.118 | 0.12 | 0.121 | 0.123 | 0.124 | 0.125 | 0.126 | 0.127 |
| 10 | 0.086 | 0.09 | 0.093 | 0.096 | 0.099 | 0.101 | 0.104 | 0.106 | 0.108 | 0.109 | 0.111 | 0.112 | 0.114 | 0.115 |
| 11 | 0.07 | 0.074 | 0.078 | 0.081 | 0.084 | 0.087 | 0.09 | 0.092 | 0.095 | 0.097 | 0.099 | 0.1 | 0.102 | 0.104 |
| 12 | 0.054 | 0.059 | 0.063 | 0.067 | 0.071 | 0.074 | 0.077 | 0.08 | 0.082 | 0.085 | 0.087 | 0.089 | 0.091 | 0.093 |
| 13 | 0.038 | 0.044 | 0.049 | 0.053 | 0.057 | 0.061 | 0.065 | 0.068 | 0.071 | 0.073 | 0.076 | 0.078 | 0.08 | 0.082 |
| 14 | 0.023 | 0.029 | 0.034 | 0.04 | 0.044 | 0.048 | 0.052 | 0.056 | 0.059 | 0.062 | 0.065 | 0.068 | 0.07 | 0.072 |
| 15 | 0.008 | 0.014 | 0.021 | 0.026 | 0.031 | 0.036 | 0.04 | 0.044 | 0.048 | 0.052 | 0.055 | 0.058 | 0.06 | 0.063 |
| 16 | | 0 | 0.007 | 0.013 | 0.019 | 0.024 | 0.029 | 0.033 | 0.037 | 0.041 | 0.044 | 0.048 | 0.051 | 0.053 |
| 17 | | | | 0 | 0.006 | 0.012 | 0.017 | 0.022 | 0.026 | 0.031 | 0.034 | 0.038 | 0.041 | 0.044 |
| 18 | | | | | | 0 | 0.006 | 0.011 | 0.016 | 0.02 | 0.024 | 0.028 | 0.032 | 0.035 |
| 19 | | | | | | | | 0 | 0.005 | 0.01 | 0.015 | 0.019 | 0.023 | 0.026 |
| 20 | | | | | | | | | | 0 | 0.005 | 0.009 | 0.014 | 0.018 |
| 21 | | | | | | | | | | | | 0 | 0.005 | 0.009 |
| 22 | | | | | | | | | | | | | | 0 |
| 23 | | | | | | | | | | | | | | |

From USEPA (1992) after Shapiro and Wilk (1965)

Table C.1 α_{n-i+1} Coefficients for the Shapiro-Wilk Test *continued*

| i/n | α_{44} | α_{45} | α_{46} | α_{47} | α_{48} | α_{49} | α_{50} |
|------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| 1 | 0.387 | 0.385 | 0.383 | 0.381 | 0.379 | 0.377 | 0.375 |
| 2 | 0.267 | 0.265 | 0.264 | 0.262 | 0.26 | 0.259 | 0.257 |
| 3 | 0.232 | 0.231 | 0.23 | 0.229 | 0.228 | 0.227 | 0.226 |
| 4 | 0.207 | 0.207 | 0.206 | 0.205 | 0.205 | 0.204 | 0.203 |
| 5 | 0.187 | 0.187 | 0.186 | 0.186 | 0.186 | 0.185 | 0.185 |
| 6 | 0.17 | 0.17 | 0.17 | 0.17 | 0.169 | 0.169 | 0.169 |
| 7 | 0.154 | 0.155 | 0.155 | 0.155 | 0.155 | 0.155 | 0.155 |
| 8 | 0.141 | 0.141 | 0.142 | 0.142 | 0.142 | 0.143 | 0.143 |
| 9 | 0.128 | 0.129 | 0.129 | 0.13 | 0.131 | 0.131 | 0.132 |
| 10 | 0.116 | 0.117 | 0.118 | 0.119 | 0.12 | 0.121 | 0.121 |
| 11 | 0.105 | 0.106 | 0.107 | 0.109 | 0.11 | 0.111 | 0.111 |
| 12 | 0.094 | 0.096 | 0.097 | 0.099 | 0.1 | 0.101 | 0.102 |
| 13 | 0.084 | 0.086 | 0.088 | 0.089 | 0.091 | 0.092 | 0.093 |
| 14 | 0.075 | 0.078 | 0.079 | 0.08 | 0.082 | 0.083 | 0.085 |
| 15 | 0.065 | 0.067 | 0.069 | 0.071 | 0.073 | 0.075 | 0.076 |
| 16 | 0.056 | 0.058 | 0.061 | 0.063 | 0.065 | 0.067 | 0.069 |
| 17 | 0.047 | 0.05 | 0.052 | 0.055 | 0.057 | 0.059 | 0.061 |
| 18 | 0.038 | 0.041 | 0.044 | 0.047 | 0.049 | 0.051 | 0.053 |
| 19 | 0.03 | 0.033 | 0.036 | 0.039 | 0.041 | 0.044 | 0.046 |
| 20 | 0.021 | 0.025 | 0.028 | 0.031 | 0.034 | 0.036 | 0.039 |
| 21 | 0.013 | 0.016 | 0.02 | 0.023 | 0.026 | 0.029 | 0.031 |
| 22 | 0.004 | 0.008 | 0.012 | 0.015 | 0.019 | 0.022 | 0.024 |
| 23 | | 0 | 0.004 | 0.008 | 0.011 | 0.014 | 0.017 |
| 24 | | | | 0 | 0.004 | 0.007 | 0.01 |
| 25 | | | | | | 0 | 0.004 |

From USEPA (1992) after Shapiro and Wilk (1965)

Table C.2 Percentage Points of the W Test for n = 3 to 50

| n | 0.01 | 0.05 | n | 0.01 | 0.05 |
|----|-------|-------|----|-------|-------|
| 3 | 0.753 | 0.767 | 27 | 0.894 | 0.923 |
| 4 | 0.687 | 0.748 | 28 | 0.896 | 0.924 |
| 5 | 0.686 | 0.762 | 29 | 0.898 | 0.926 |
| 6 | 0.713 | 0.788 | 30 | 0.9 | 0.927 |
| 7 | 0.73 | 0.803 | 31 | 0.902 | 0.929 |
| 8 | 0.749 | 0.818 | 32 | 0.904 | 0.93 |
| 9 | 0.764 | 0.829 | 33 | 0.906 | 0.931 |
| 10 | 0.781 | 0.842 | 34 | 0.908 | 0.933 |
| 11 | 0.792 | 0.85 | 35 | 0.91 | 0.934 |
| 12 | 0.805 | 0.859 | 36 | 0.912 | 0.935 |
| 13 | 0.814 | 0.866 | 37 | 0.914 | 0.936 |
| 14 | 0.825 | 0.874 | 38 | 0.916 | 0.938 |
| 15 | 0.835 | 0.881 | 39 | 0.917 | 0.939 |
| 16 | 0.844 | 0.887 | 40 | 0.919 | 0.94 |
| 17 | 0.851 | 0.892 | 41 | 0.92 | 0.941 |
| 18 | 0.858 | 0.897 | 42 | 0.922 | 0.942 |
| 19 | 0.863 | 0.901 | 43 | 0.923 | 0.943 |
| 20 | 0.868 | 0.905 | 44 | 0.924 | 0.944 |
| 21 | 0.873 | 0.908 | 45 | 0.926 | 0.945 |
| 22 | 0.878 | 0.911 | 46 | 0.927 | 0.945 |
| 23 | 0.881 | 0.914 | 47 | 0.928 | 0.946 |
| 24 | 0.884 | 0.916 | 48 | 0.929 | 0.947 |
| 25 | 0.888 | 0.918 | 49 | 0.929 | 0.947 |
| 26 | 0.891 | 0.92 | 50 | 0.93 | 0.947 |

Significance level of 0.01

Significance level of 0.05

From USEPA (1992) after Shapiro and Wilk (1965)

Appendix D Worked examples

D

Example 1 Assuming the Planning Scenario

The site has been identified for redevelopment and therefore reference has been made to the procedure given in Box 9 of the main text. The primary contaminant of concern is Substance X. A site investigation was undertaken and a total of 33 samples was collected from a defined area and depth of soil and analysed for Substance X.

The objective of the statistical analysis is to determine whether there is sufficient evidence to demonstrate that the concentration of Substance X in the soil being assessed is less than a critical concentration of 50 mg/kg and therefore can be considered 'suitable for use'. Measured sample concentrations are presented in Table 1.

Table 1 Measured concentrations of Substance X in soil samples collected from across the site (mg/kg)

| | | |
|------|------|-------|
| <10 | 39.1 | 43.4 |
| <10 | 39.1 | 44.9 |
| <10 | 39.1 | 44.9 |
| 31.9 | 39.1 | 44.9 |
| 31.9 | 40.5 | 46.3 |
| 33.3 | 40.5 | 46.3 |
| 36.2 | 40.5 | 47.8 |
| 37.6 | 40.5 | 50.7 |
| 37.6 | 40.5 | 50.7 |
| 37.6 | 43.4 | 52.1 |
| 37.6 | 43.4 | 286.9 |

Data analysis according to the procedure in Box 9

| | | |
|--------|---|---|
| Step 0 | | The data were reviewed and found to meet the data quality criteria. |
| Step 1 | | <p>The planning scenario is appropriate, therefore the key question is:</p> <p>Is there sufficient evidence that the true mean concentration of Substance X (μ) is less than the critical concentration (C_c)?</p> <p>The Null Hypothesis is that the true mean is equal to, or greater than, the critical concentration.</p> <p>The Alternative Hypothesis is that the true mean concentration is less than the critical concentration.</p> |
| Step 2 | i | Three non-detects were identified. As this is about 10% of the population, they were replaced with half the level of detection, giving a value of 5 mg/kg. |

Example 1
continued

| | |
|---------------|--|
| <p>Step 2</p> | <p>ii An outlier test was performed as detailed in Appendix B using the natural logs of the data (since the dataset without the suspected outlier shows evidence of non-normality).</p> $T_n = \frac{x_n - \bar{x}}{s} = 2.85 \text{ (including the three non-detects)}$ $T_{crit} = 2.787 \text{ (for } n=33 \text{ and } \alpha = 0.05)$ <p>The maximum measured concentration (286.9 mg/kg) was therefore considered an outlier at a confidence level of 95%. The outlier test was repeated on the second highest measured value (52.1mg/kg):</p> $T_n = 0.68$ <p>The second highest measured concentration is not considered to be an outlier.</p> <p>The highest value was assessed and found to be the result of a laboratory error and, since it was impossible to find out what the correct value was, it was excluded from further assessment.³³</p> |
| <p>Step 2</p> | <p>iii A probability plot for the data was produced using the procedures outlined in Appendix C showing that the data are not aligned along the 45 degrees line.</p> <div data-bbox="603 1256 1326 1944" style="border: 1px solid black; padding: 10px; margin: 10px 0;"> </div> |

³³ Note that in this example it was possible to remove the suspected outlier because it was found to be a laboratory error. When deciding whether suspected outliers should be excluded or retained in a dataset, refer to Section 5.3.2 and the procedure given in Appendix B, particularly steps 3, 4 and 11

Example 1
continued

Step 2
continued

iii

The Shapiro-Wilk normality test was undertaken (Appendix C) on the data excluding the outlier.

| Sample | Rank _(i) | Ordered Conc. $X_{(i)}$ | $S^2_{(i)}$ | $X_{(n-i+1)}$ | $X_{(n-i+1)} - X_{(i)}$ | $\alpha_{(n-i+1)}$ | $b_{(i)}$ |
|--------|---------------------|-------------------------|-------------|---------------|-------------------------|--------------------|-----------|
| S1 | 1 | 5 | 1089.83 | 52.1 | 47.1 | 0.419 | 19.735 |
| S2 | 2 | 5 | 1089.83 | 50.7 | 45.7 | 0.290 | 13.253 |
| S3 | 3 | 5 | 1089.83 | 50.7 | 45.7 | 0.246 | 11.242 |
| S4 | 4 | 31.9 | 37.36 | 47.8 | 15.9 | 0.214 | 3.403 |
| S5 | 5 | 31.9 | 37.36 | 46.3 | 14.4 | 0.188 | 2.707 |
| S6 | 6 | 33.3 | 22.21 | 46.3 | 13 | 0.165 | 2.145 |
| S7 | 7 | 36.2 | 3.29 | 44.9 | 8.7 | 0.145 | 1.262 |
| S8 | 8 | 37.6 | 0.17 | 44.9 | 7.3 | 0.127 | 0.927 |
| S9 | 9 | 37.6 | 0.17 | 44.9 | 7.3 | 0.109 | 0.796 |
| S10 | 10 | 37.6 | 0.17 | 43.4 | 5.8 | 0.093 | 0.539 |
| S11 | 11 | 37.6 | 0.17 | 43.4 | 5.8 | 0.078 | 0.452 |
| S12 | 12 | 39.1 | 1.18 | 43.4 | 4.3 | 0.063 | 0.271 |
| S13 | 13 | 39.1 | 1.18 | 40.5 | 1.4 | 0.049 | 0.069 |
| S14 | 14 | 39.1 | 1.18 | 40.5 | 1.4 | 0.034 | 0.048 |
| S15 | 15 | 39.1 | 1.18 | 40.5 | 1.4 | 0.021 | 0.029 |
| S16 | 16 | 40.5 | 6.19 | 40.5 | 0 | 0.007 | 0.000 |
| S17 | 17 | 40.5 | 6.19 | 40.5 | 0 | | |
| S18 | 18 | 40.5 | 6.19 | 39.1 | -1.4 | | |
| S19 | 19 | 40.5 | 6.19 | 39.1 | -1.4 | | |
| S20 | 20 | 40.5 | 6.19 | 39.1 | -1.4 | | |
| S21 | 21 | 43.4 | 29.03 | 39.1 | -4.3 | | |
| S22 | 22 | 43.4 | 29.03 | 37.6 | -5.8 | | |
| S23 | 23 | 43.4 | 29.03 | 37.6 | -5.8 | | |
| S24 | 24 | 44.9 | 47.44 | 37.6 | -7.3 | | |
| S25 | 25 | 44.9 | 47.44 | 37.6 | -7.3 | | |
| S26 | 26 | 44.9 | 47.44 | 36.2 | -8.7 | | |
| S27 | 27 | 46.3 | 68.68 | 33.3 | -13 | | |
| S28 | 28 | 46.3 | 68.68 | 31.9 | -14.4 | | |
| S29 | 29 | 47.8 | 95.80 | 31.9 | -15.9 | | |
| S30 | 30 | 50.7 | 160.97 | 5 | -45.7 | | |
| S31 | 31 | 50.7 | 160.97 | 5 | -45.7 | | |
| S32 | 32 | 52.1 | 198.46 | 5 | -47.1 | | |

Example 1
continued

| | | | | | | |
|---------------------|---|---|--------------------|------|------|----------------|
| Step 2 continued | iii | $S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = 4389.00$ | | | | |
| | | $b = \sum_{i=1}^k a_{n-i+1} (x_{n-i+1} - x_i) = 56.878$ | | | | |
| | | $W = \frac{b^2}{S^2} = 0.737$ | | | | |
| | | <p>For $n = 32$</p> <table border="1"> <tr> <td>Significance Level</td> <td>0.01</td> <td>0.05</td> </tr> <tr> <td>Critical Level</td> <td>0.904</td> <td>0.930</td> </tr> </table> <p>W is less than the critical value at a significance level of 0.05, therefore the data set is not normally distributed.</p> | Significance Level | 0.01 | 0.05 | Critical Level |
| Significance Level | 0.01 | 0.05 | | | | |
| Critical Level | 0.904 | 0.930 | | | | |
| Step 3 | $\bar{x} = 38.01 \text{ mg/kg}$ $s = 11.90 \text{ mg/kg}$ | | | | | |
| Step 4 | $C_C = 50 \text{ mg/kg}$ $\bar{x} < C_C$ | | | | | |
| Step 5 | The dataset distribution deviates significantly from normality, therefore the one-sided Chebychev Theorem (Step 10) applies. | | | | | |
| Step 10 | $n = 32$ $k_0 = \frac{\bar{x} - C_C}{\frac{s}{\sqrt{n}}} = -5.699$ <p>and</p> $UCL_{0.95} = \bar{x} + \left(k_{(0.05)} \times \frac{s}{\sqrt{n}} \right) = 47.18$ | | | | | |
| Step 11 | From Table A.2, $k_{0.05} = 4.36$ | | | | | |
| Step 12 | $k_{crit} = k_{0.05} \therefore k_{crit} = 4.36$ $k_0 < k_{crit}$ <p>Therefore, H_0 can be rejected and it can be concluded that the true mean is <i>not</i> greater than the critical concentration at a confidence level of 95%. (as can be confirmed by comparing the 95% UCL with C_C).</p> | | | | | |

Example 1
continued

| | | |
|---------|---|---|
| Step 13 | i | From Table A.2, $k_1 = 5.69$ |
| | ii | Therefore $\alpha_1 = 0.03$ |
| | iii | $p_1 = 0.97$ Therefore a [conservative] estimate of the evidence <i>against</i> H_0 being true is 97%. |
| | iv | $p_1 > 95\%$ Therefore H_0 can be rejected with a high degree of confidence. |
| Step 15 | H_0 is rejected and the following outcome is recorded: The evidence suggests that μ is not greater than C_C ; that is, the area of land being assessed can be considered 'suitable for use'. | |

Example 2 Assuming the Part 2A Scenario

A site has been identified for assessment under Part 2A and therefore reference was made to Box 10 of the main text. The primary contaminant of concern is Substance Y. A site investigation was undertaken and a total of 10 samples was collected and analysed for Substance Y. Samples were also analysed for soil organic matter (SOM), and were found to have an average SOM content of 5%. The critical concentration of Substance Y for the measured SOM was determined as 41 mg/kg.

The objective of the statistical assessment is to determine whether there is sufficient evidence that the true mean concentration of Substance Y in a defined area and depth of soil is greater than the critical concentration of 41 mg/kg.

Measured sample concentrations are presented in Table 1.

Table 1 Measured concentrations of Substance Y in soil samples collected from across the site (mg/kg)

| | | |
|------|------|-------|
| 5.8 | 39.7 | 62.0 |
| 17.4 | 46.4 | 73.6 |
| 25.8 | 53.6 | 101.3 |
| 33.0 | | |

Data analysis according to the procedure in Box 10

| | | |
|--------|---|---|
| Step 0 | | The data was reviewed and found to meet the data quality criteria. |
| Step 1 | | <p>The Part 2A scenario is appropriate, therefore the key question is:</p> <p>Is there sufficient evidence that the true mean concentration of Substance Y (μ) is greater than the critical concentration (C_c)?</p> <p>The Null Hypothesis is that the true mean concentration is equal to, or less than, the critical concentration.</p> <p>The Alternative Hypothesis is that the true mean concentration is greater than the critical concentration.</p> |
| Step 2 | i | No non-detects were identified. |

Example 2
continued

| Step 2 | <p>ii</p> <p>An outlier test was performed as detailed in Appendix B.</p> <p>$T_n = 1.96$</p> <p>$T_{crit} = 2.176$ (for $n = 10$ and $\alpha = 0.05$)</p> <p>The maximum measured concentration was not considered an outlier at the 95% confidence level.</p> <p>A probability plot for the data was produced using the procedures outlined in Appendix C, showing that the dataset without the maximum value (101.3 mg/kg) is approximately normal indicating that the results of the outlier test are valid (see Appendix B, Step 3i).</p> <div data-bbox="603 1003 1326 1688"><table border="1"><caption>Approximate data points from the probability plot</caption><thead><tr><th>Experimental (yi)</th><th>Theoretical (zi)</th></tr></thead><tbody><tr><td>-1.5</td><td>-1.5</td></tr><tr><td>-1.0</td><td>-1.0</td></tr><tr><td>-0.5</td><td>-0.5</td></tr><tr><td>0.0</td><td>0.0</td></tr><tr><td>0.5</td><td>0.5</td></tr><tr><td>1.0</td><td>1.0</td></tr><tr><td>1.5</td><td>1.5</td></tr></tbody></table></div> | Experimental (yi) | Theoretical (zi) | -1.5 | -1.5 | -1.0 | -1.0 | -0.5 | -0.5 | 0.0 | 0.0 | 0.5 | 0.5 | 1.0 | 1.0 | 1.5 | 1.5 |
|-------------------|--|-------------------|------------------|------|------|------|------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|
| Experimental (yi) | Theoretical (zi) | | | | | | | | | | | | | | | | |
| -1.5 | -1.5 | | | | | | | | | | | | | | | | |
| -1.0 | -1.0 | | | | | | | | | | | | | | | | |
| -0.5 | -0.5 | | | | | | | | | | | | | | | | |
| 0.0 | 0.0 | | | | | | | | | | | | | | | | |
| 0.5 | 0.5 | | | | | | | | | | | | | | | | |
| 1.0 | 1.0 | | | | | | | | | | | | | | | | |
| 1.5 | 1.5 | | | | | | | | | | | | | | | | |

Example 2
continued

| <p>Step 2 <i>continued</i></p> | <p>iii The Shapiro-Wilk normality test was undertaken (Appendix C).</p> <table border="1" data-bbox="603 667 1453 1115"> <thead> <tr> <th>Sample</th> <th>Rank_(i)</th> <th>Ordered Conc. $X_{(i)}$</th> <th>$S^2_{(i)}$</th> <th>$X_{(n-i+1)}$</th> <th>$X_{(n-i+1)} - X_{(i)}$</th> <th>$a_{(n-i+1)}$</th> <th>$b_{(i)}$</th> </tr> </thead> <tbody> <tr><td>S1</td><td>1</td><td>5.8</td><td>1604.8036</td><td>101.3</td><td>95.500</td><td>0.574</td><td>54.817</td></tr> <tr><td>S2</td><td>2</td><td>17.4</td><td>809.9716</td><td>73.6</td><td>56.200</td><td>0.329</td><td>18.4898</td></tr> <tr><td>S3</td><td>3</td><td>25.8</td><td>402.4036</td><td>62</td><td>36.200</td><td>0.214</td><td>7.7468</td></tr> <tr><td>S4</td><td>4</td><td>33</td><td>165.3796</td><td>53.6</td><td>20.600</td><td>0.122</td><td>2.5132</td></tr> <tr><td>S5</td><td>5</td><td>39.7</td><td>37.9456</td><td>46.4</td><td>6.700</td><td>0.040</td><td>0.268</td></tr> <tr><td>S6</td><td>6</td><td>46.4</td><td>0.2916</td><td>39.7</td><td></td><td></td><td></td></tr> <tr><td>S7</td><td>7</td><td>53.6</td><td>59.9076</td><td>33</td><td></td><td></td><td></td></tr> <tr><td>S8</td><td>8</td><td>62</td><td>260.4996</td><td>25.8</td><td></td><td></td><td></td></tr> <tr><td>S9</td><td>9</td><td>73.6</td><td>769.5076</td><td>17.4</td><td></td><td></td><td></td></tr> <tr><td>S10</td><td>10</td><td>101.3</td><td>3073.5936</td><td>5.8</td><td></td><td></td><td></td></tr> </tbody> </table> <p>$S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = 7184.3$</p> <p>$b = \sum_{i=1}^k a_{n-i+1} (x_{n-i+1} - x_i) = 83.8$</p> <p>$W = \frac{b^2}{S^2} = 0.978$</p> <p>For $n = 10$</p> <table border="1" data-bbox="603 1458 1425 1536"> <tr> <td>Significance Level</td> <td>0.01</td> <td>0.05</td> </tr> <tr> <td>Critical Level</td> <td>0.781</td> <td>0.842</td> </tr> </table> <p>W is greater than the critical value showing that the dataset does not deviate significantly from normality.</p> | Sample | Rank _(i) | Ordered Conc. $X_{(i)}$ | $S^2_{(i)}$ | $X_{(n-i+1)}$ | $X_{(n-i+1)} - X_{(i)}$ | $a_{(n-i+1)}$ | $b_{(i)}$ | S1 | 1 | 5.8 | 1604.8036 | 101.3 | 95.500 | 0.574 | 54.817 | S2 | 2 | 17.4 | 809.9716 | 73.6 | 56.200 | 0.329 | 18.4898 | S3 | 3 | 25.8 | 402.4036 | 62 | 36.200 | 0.214 | 7.7468 | S4 | 4 | 33 | 165.3796 | 53.6 | 20.600 | 0.122 | 2.5132 | S5 | 5 | 39.7 | 37.9456 | 46.4 | 6.700 | 0.040 | 0.268 | S6 | 6 | 46.4 | 0.2916 | 39.7 | | | | S7 | 7 | 53.6 | 59.9076 | 33 | | | | S8 | 8 | 62 | 260.4996 | 25.8 | | | | S9 | 9 | 73.6 | 769.5076 | 17.4 | | | | S10 | 10 | 101.3 | 3073.5936 | 5.8 | | | | Significance Level | 0.01 | 0.05 | Critical Level | 0.781 | 0.842 |
|------------------------------------|---|-------------------------|---------------------|-------------------------|-------------------------|---------------|-------------------------|---------------|-----------|----|---|-----|-----------|-------|--------|-------|--------|----|---|------|----------|------|--------|-------|---------|----|---|------|----------|----|--------|-------|--------|----|---|----|----------|------|--------|-------|--------|----|---|------|---------|------|-------|-------|-------|----|---|------|--------|------|--|--|--|----|---|------|---------|----|--|--|--|----|---|----|----------|------|--|--|--|----|---|------|----------|------|--|--|--|-----|----|-------|-----------|-----|--|--|--|--------------------|------|------|----------------|-------|-------|
| Sample | Rank _(i) | Ordered Conc. $X_{(i)}$ | $S^2_{(i)}$ | $X_{(n-i+1)}$ | $X_{(n-i+1)} - X_{(i)}$ | $a_{(n-i+1)}$ | $b_{(i)}$ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S1 | 1 | 5.8 | 1604.8036 | 101.3 | 95.500 | 0.574 | 54.817 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S2 | 2 | 17.4 | 809.9716 | 73.6 | 56.200 | 0.329 | 18.4898 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S3 | 3 | 25.8 | 402.4036 | 62 | 36.200 | 0.214 | 7.7468 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S4 | 4 | 33 | 165.3796 | 53.6 | 20.600 | 0.122 | 2.5132 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S5 | 5 | 39.7 | 37.9456 | 46.4 | 6.700 | 0.040 | 0.268 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S6 | 6 | 46.4 | 0.2916 | 39.7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S7 | 7 | 53.6 | 59.9076 | 33 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S8 | 8 | 62 | 260.4996 | 25.8 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S9 | 9 | 73.6 | 769.5076 | 17.4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S10 | 10 | 101.3 | 3073.5936 | 5.8 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Significance Level | 0.01 | 0.05 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Critical Level | 0.781 | 0.842 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Step 3</p> | <p>$\bar{x} = 45.86$ mg/kg $s = 28.25$ mg/kg</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Step 4</p> | <p>$C_C = 41$ mg/kg $\bar{x} > C_C$</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Step 5</p> | <p>The dataset distribution does not deviate significantly from normality, therefore the one sample t-test (Step 6) will be applied.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Example 2
continued

| | | |
|---------|-----|--|
| Step 6 | | $n = 10$ $t_0 = \frac{\bar{x} - C_C}{\frac{s}{\sqrt{n}}} = 0.544$ <p>and</p> $LCL_{0.95} = \bar{x} - \left(t_{(n-1, 0.95)} \times \frac{s}{\sqrt{n}} \right) = 29.48$ |
| Step 7 | | $t_{(n-1, 0.95)} = 1.833$ |
| Step 8 | | $t_0 < t_{(n-1, 0.95)}$ <p>While the sample mean is greater than C_C, the possibility that the true mean is less than C_C cannot be discounted and the H_0 is not rejected at the 95 % confidence level.</p> |
| Step 9 | i | From Table A.1, $t_p = 0.543$ for 9 degrees of freedom. |
| | ii | <p>Therefore $p_1 = 0.70$</p> <p>Therefore, the level of evidence <i>against</i> H_0 being true is 70%.</p> |
| | iii | Note that the analysis progresses to Step 14. |
| Step 14 | i | <p>Since $p_1 > 0.51$</p> <p>H_0 can be rejected on a 'balance of probabilities' basis.</p> |
| Step 16 | | The evidence suggests that, on a 'balance of probabilities' basis, μ is greater than C_C ; that is, the area of land being assessed could be determined as contaminated land under Part 2A. |

CL:AIRE

CONTAMINATED LAND: APPLICATIONS IN REAL ENVIRONMENTS

CL:AIRE
7th Floor 1 Great Cumberland Place London W1H 7AL
Telephone 020 7258 5321 **Fax** 020 7258 5322
Email enquiries@claire.co.uk **Web** www.claire.co.uk
Registered Charity No. 1075611



Chartered
Institute of
Environmental
Health

Chartered Institute of Environmental Health
Chadwick Court 15 Hatfields London SE1 8DJ
Telephone 020 7928 6006 **Fax** 020 7827 5831
Email info@cieh.org **Web** www.cieh.org
Registered Charity No. 290350