British Geological Survey

NATURAL ENVIRONMENT RESEARCH COUNCIL

Within site and between site nested analysis of variance (ANOVA) for Geochemical Surveys using MS EXCEL

Internal Report IR/02/043

BRITISH GEOLOGICAL SURVEY

INTERNAL REPORT IR/02/043

Within site and between site nested analysis of variance (ANOVA) for Geochemical Surveys using MS EXCEL

C C Johnson

Key words

Report; Geochemistry, ANOVA, MS EXCEL, quality control.

Bibliographical reference

JOHNSON C.C. 2002. Within site and between site nested analysis of variance (ANOVA) for Geochemical Surveys using MS EXCEL. *British Geological Survey Internal Report*, IR/02/043. 25 pp.

Keyworth, Nottingham British Geological Survey 2002

BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham and Edinburgh; see contact details below or shop online at www.thebgs.co.uk

The London Information Office maintains a reference collection of BGS publications including maps for consultation.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.

The British Geological Survey is a component body of the Natural Environment Research Council.

Keyworth, Nottingham NG12 5GG

☎ 0115-936 3241 Fax 0115-936 3488
e-mail: sales@bgs.ac.uk
www.bgs.ac.uk
Shop online at: www.thebgs.co.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

☎ 0131-667 1000 Fax 0131-668 2683
e-mail: scotsales@bgs.ac.uk

London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE

☎ 020-7589 4090 Fax 020-7584 8270
☎ 020-7942 5344/45 email: bgslondon@bgs.ac.uk

Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU

☎ 01392-445271 Fax 01392-445371

Geological Survey of Northern Ireland, 20 College Gardens, Belfast BT9 6BS

☎ 028-9066 6595 Fax 028-9066 2835

Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB

☎ 01491-838800 Fax 01491-692345

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1EU

☎ 01793-411500 Fax 01793-411501
www.nerc.ac.uk

Contents

FIGURES

TABLES

Summary

This report describes a method of nested analysis of variance (ANOVA) using a visual basic macro in Microsoft EXCEL 2000. Nested ANOVA analysis is used in geochemical surveys to test the suitability of the sampling and analytical methodology and quantifies the within site and between site variability. If the between site variability is greater than the within site variability then the survey methodology is satisfactory. BGS uses this quality control procedure for its national and international geochemical surveys. However, the restricted availability of specialist statistical software to do nested ANOVA has often resulted in a variety of software packages being used for different projects. This report describes a method of nested ANOVA that can be done using the widely available Microsoft EXCEL. Data input format and result reporting is similar to that used on the Geochemical Baseline Survey of the Environment (G-BASE) Project.

1 Introduction

Analysis of variance (ANOVA) is one of several procedures that is used to monitor the quality and representativity of geochemical results. Nested ANOVA analysis done on duplicate and replicate samples can be used to give a quantitative measure of the variability of results within and between sampling sites. For the methods of a geochemical survey results to be considered valid, the variation in results between different sites should be much more than the variation in results from samples collected at the same site.

This is determined by calculating the variance components expressed as a percentage of the total and a summary table of such data is shown in Table 1 .

	Variance Component (%)			
	Sites	**Duplicates**	**Replicates**	**Total**
Ag	2.7193	0.0000	97.2807	100
As	99.5744	0.2556	0.1700	100
Ba	99.5641	0.2944	0.1415	100
Be	95.6789	0.0000	4.3211	100
Bi	95.4449	0.0000	4.5551	100
Cd	13.8361	0.0000	86.1639	100
Co	68.6439	0.0000	31.3561	100
Cr	99.0051	0.2933	0.7016	100
Cu	99.0679	0.0921	0.8399	100
Fe_2O_3	98.9610	0.9041	0.1349	100
Li	98.3441	0.0000	1.6559	100
MnO	99.0357	0.8001	0.1642	100
Mo	79.3723	11.9387	8.6890	100
Nb	99.5499	0.4251	0.0250	100
Ni	93.1000	0.0000	6.9000	100
Pb	97.6080	2.1987	0.1934	100
Sb	90.2153	2.3122	7.4725	100
Sn	94.0362	0.2867	5.6771	100
Sr	99.8779	0.0774	0.0447	100
Te	0.0000	0.0000	100	100
Th	92.8791	0.0000	7.1209	100
TiO_2	98.3582	1.4452	0.1965	100
U	71.1965	0.1202	28.6833	100
V	99.0381	0.3177	0.6442	100
W	88.0914	8.8383	3.0702	100
Y	96.9617	1.8054	1.2330	100
Zn	99.4340	0.4571	0.1089	100
Zr	98.7967	1.0113	0.1920	100

Table 1: Summary of nested ANOVA results for a regional geochemical data set from Morocco

The amount of importance we should attach to the ANOVA method depends on how representative the replicate and duplicate sampling is, and to what extent the geochemical data fits in with the ideal statistical models for data distribution. For the UK Regional Geochemical Mapping there are ANOVA samples for one out of every 93 sites sampled and the analysis is done without considering any of the site characteristics (such as catchment lithology). Furthermore, the geochemical data rarely follows the ideal normal population distribution so poly-modal non-normal distributions, with outlying values, will not give strictly valid ANOVA results. Inspection of the ANOVA data beforehand (e.g. plotting cumulative probability plots) will help to establish the applicability of the ANOVA method and may be used to explain high percentage of variance between samples collected at the same site. Converting the results to $\log_{10}$ will reduce the effect of outlying results.

It should also be noted that although duplicate samples are collected at the same site, this is strictly speaking not true. Two samples cannot be collected from the same point so there will be some variance component because of this. The significance of this will vary with different sample media.

Any discussion of the ANOVA results should take into account these limitations discussed above. Nevertheless, the results can be used to indicate the elements for which geochemical variability between sites is most significant. A good discussion of ANOVA analysis for evaluation of precision requirements for geochemical analysis is given by Ramsey and others (1992). This work recommends that analytical variance should be <4% and duplicate variance <16%. The term nested (sometimes referred to as hierarchical) is used when fewer than all levels of one factor occur within each level of the other factor.

In order to look at "within" site and "between" site variability it is necessary for the data set to contain duplicate and replicate samples. A duplicate sample is a sample that is collected from the same sampling site as an original sample. A replicate is an original sample subdivided in the laboratory. A duplicate sample will therefore indicate sampling variability within a single site whereas a replicate will indicate the variability of the laboratory analysis. For the G-BASE project the replicates are sub-samples of the duplicate samples (see Figure 1) so the nested ANOVA is made up of the results from four samples originating from a single site.

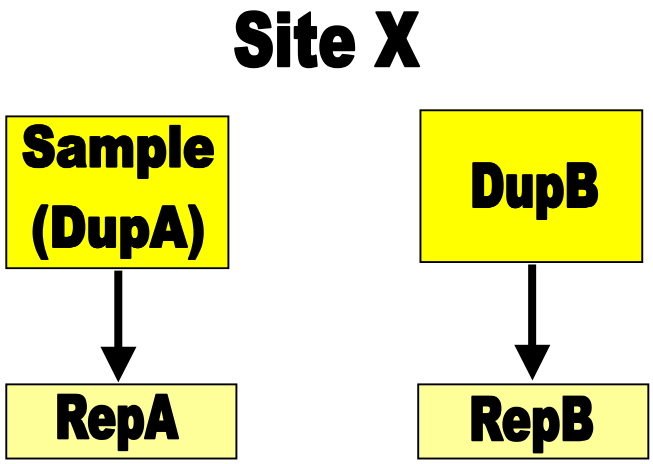

Figure 1: Explanation of the G-BASE method of duplicate and replicate numbering

This report does not cover the theory behind the random nested ANOVA method (see Sinclair, 1983). It is based on an unpublished note by Johnson (1995).

Nested ANOVA is not a simple statistical procedure and methods vary (particularly in data input format) between statistical packages. The G-BASE project has used SAS and MINITAB software but these relatively expensive packages have not generally been available for use in overseas geochemical surveys because of licencing restrictions. It is much more satisfactory to carry out the nested ANOVA in a software package that is universally available. Microsoft Office is a software package widely used and Microsoft EXCEL can be used to do much of the statistical analysis necessary for regional geochemical surveys (e.g. summary statistics and histograms, Johnson, 2000).

The method described here is therefore widely applicable, particularly for international geochemical surveys where specialist statistical software in the local language is not available. It is a simple procedure to edit the macro to report the summary table with column headings in, say for example, French.

1.1 NESTED ANOVA ANALYSIS BY MICROSOFT EXCEL 2000

Microsoft Excel has "Data Analysis" option from the Tools menu (if installed) that brings up the dialogue shown in Figure 2.

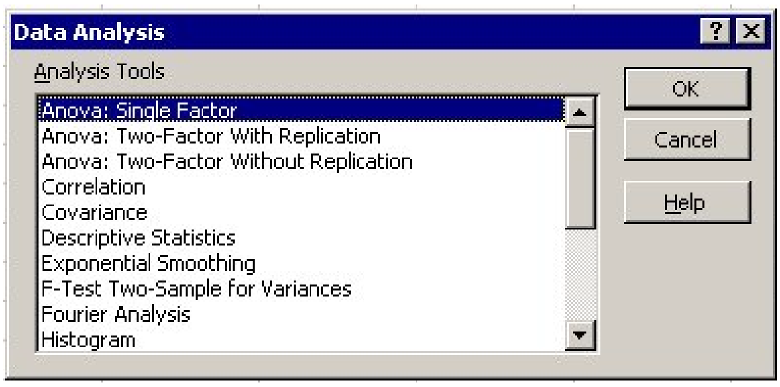

Figure 2: Microsoft Excel 2000 Data Analysis dialogue box

Nested ANOVA analysis is **not** an option[1] and the data analysis **cannot** be used be performed using this data analysis tool. For nested ANOVA an EXCEL visual basic macro has been written that will reformat and log the raw data, calculate the variance components for each element and create a summary table of the results. The equations and formulae for calculating the variance are based on those presented by Sinclair (1983). A listing of the macro is given in Appendix 1 and is available in a blank workbook file called *nested_Anova_blank.xls*[2].

[1] An explanation of the the ANOVA options that are available can be found at http://www.mastep.sjsu.edu/learn/anova.htm

[2] For those who have access to the internal BGS server this is available on the corporate shared drive in the G-BASE project directory at : ……gbase/Sample Analysis/

2 Procedure for Nested Analysis of Variance using MS EXCEL 2000

The procedure has two stages. Firstly, the data (in a standard format) is copied into an EXCEL workbook containing the "nested-ANOVA" macro. The data input format (see below) is similar to that used for the G-BASE project. Secondly, the macro is run to produce a summary percentage variance table. The macro is currently set to process a data set with a maximum of 4000 control sites and 50 elements. These are merely a check to stop uncontrolled looping and could be increased if necessary.

2.1 DATA FORMAT

Appendix 2 gives an example of the format for the input data. The correct data input format is essential and the following should be used as a checklist prior to actually doing the ANOVA analysis:

1) The first cell (A1) should contain a title for the data
2) The second row should contain the column headers and these headers will be used in the final summary table. Column one is the site/sample number, column two the control sample description (e.g. DUPA), and subsequent columns the data to be interpretted
3) The data records should commence on the third row. **The data should not contain zeros or minus data**.
4) Data for each set of control samples (i.e. two duplicates and two replicates) **must be** listed in the order DUPA, REPA, DUPB, and REPB (see Appendix 2)
5) The worksheet should be named "Data" and it should be the **only** worksheet in the workbook.

2.2 DOING THE ANOVA ANALYSIS

When the data is correctly set out as described above the actual data processing can commence. This is simply done by running the nested-Anova macro in EXCEL (shortcut key Ctrl+Shift+R).

The macro then does the following:

1) The data layout is checked and the user informed as to how many sites and elements will be processed. If the data layout is incorrect the user will be informed or it will be obvious from the "number of sites and elements to be processed" message.
2) A new worksheet is created and named after the element being processed
3) The four control samples (DUPA, REPA, DUPB and REPB) are reformatted from four rows in a single column to one row of four columns
4) The data is log transformed ($\log_{10}$)
5) The degrees of freedom and various summations as described by Sinclair (1983) are performed and listed on the data sheet. The layout of a worksheet for a single element ANOVA analysis described in Appendix 3

6) The variance component (expressed as a percentage) is taken from each element sheet and placed in a neatly formatted table. This table is placed in a worksheet called "Summary" which is displayed when the macro ends successfully.

Urban Soil Control data

Element	Between Site %	Between Sample %	Within Sample %	Total %
MnO	96.03	3.92	0.05	100
FE2O3	96.62	3.36	0.01	100
V	97.85	2.09	0.06	100
Cr	93.46	5.55	0.99	100
Co	94.00	5.62	0.38	100
Ba	97.39	2.56	0.05	100
Ni	95.96	3.83	0.21	100
Cu	98.87	1.08	0.06	100
Zn	92.64	7.34	0.02	100
As	97.87	1.82	0.31	100
Mo	93.59	3.23	3.17	100
Pb	96.51	3.43	0.06	100

Table 2: Summary table of results from nested ANOVA analysis of control data set

2.3 TESTING THE PROCEDURE

This procedure has been tested using a control data set analysed in SAS and described by Lister (2002). The table of variance components is shown in Table 2 and is consistent with the results obtained by Lister (2002).

Appendix 1 : Listing of Nested_Anova macro

```
Sub Nested_Anova()
'
' Macro to do nested ANOVA analysis
' Macro written 18/02/2002 by Chris Johnson
'
' Keyboard Shortcut: Ctrl+Shift+R
'
'declare  variables
'

Dim Msg, Style, title, Help, Ctxt, Response, MyString
Dim myRange As Range
Dim c As Object
Dim d As Object
Dim e As Object
Dim Results(1 To 50, 1 To 4) As Single
Dim  ncols,  nrows,  replicates,  duplicates,  nrecords,  sites,  n,  i,  k,  j  As
Integer
Dim element, listofnames(50) As String
Dim tit, site As String

'check name of active sheet which must be "Data" and check that there is only
one sheet in the workbook
If ActiveSheet.name <> "Data" Or Sheets.Count > 1 Then
MsgBox "Program terminated as your currently active sheet is not called Data
or you have more than one worksheet in your workbook"
Stop
End If

'read in title
tit = ActiveSheet.Cells(1, 1).Value

'determine size of dataset
    'move to first row of data
    Range("A3:A3").Select
    Set c = ActiveCell
    ncols = 0
    'count the number of colums. Exit if number of columns gets to 50
    Do Until IsEmpty(c.Value)
        ncols = ncols + 1
        Set c = c.Offset(0, 1)
            If ncols = 50 Then
            MsgBox "Program terminated as number of columns greater than 50"
            Stop
            End If
    Loop
'number of elements to be processed is no cols -2
ncols = ncols - 2

'move to first row of data
    Range("A3:A3").Select
    Set c = ActiveCell
    nrows = 0
    'count the number of rows. Exit if number of rows gets to 4000
    Do Until IsEmpty(c.Value)
        nrows = nrows + 1
        Set c = c.Offset(1, 0)
            If nrows = 4000 Then
```

```
            MsgBox "Program terminated as number of rows of data greater than 4000"
            Stop
            End If
    Loop


Msg = "You have chosen to process " & ncols & " elements and " & nrows / 4 & " sites. Do you want to continue ?"     ' Define message.
Style = vbYesNo + vbCritical + vbDefaultButton1    ' Define buttons.
title = "Parameters selected"    ' Define title.
Help = "DEMO.HLP"    ' Define Help file.
Ctxt = 1000    ' Define topic
        ' context.
        ' Display message.
Response = MsgBox(Msg, Style, title, Help, Ctxt)
If Response = vbYes Then    ' User chose Yes.
    MyString = "Yes"    ' Perform some action.
Else    ' User chose No.
    MyString = "No" ' Perform some action.
    MsgBox "Processing terminated"
    End
End If



'start processing each element column
For n = 1 To ncols
Worksheets("Data").Activate
'activate first column in first row of data on data sheet
    Range("A2:A2").Select

    'move to column label
    Set d = ActiveCell.Offset(0, 1 + n)
    element = d.Value
    Set d = d.Offset(1, 0)
    'add a worksheet and name it after the element being processed

    Sheets.Add
    ActiveSheet.name = element

    'place lables
        ActiveSheet.Cells(1, 1).Value = tit
        ActiveSheet.Cells(2, 1).Value = element
        ActiveSheet.Cells(3, 1).Value = "Site"
        ActiveSheet.Cells(3, 2).Value = "DupA"
        ActiveSheet.Cells(3, 3).Value = "RepA"
        ActiveSheet.Cells(3, 4).Value = "DupB"
        ActiveSheet.Cells(3, 5).Value = "RepB"
        ActiveSheet.Cells(3, 6).Value = "Log10(DupA)"
        ActiveSheet.Cells(3, 7).Value = "Log10(RepA)"
        ActiveSheet.Cells(3, 8).Value = "Log10(DupB)"
        ActiveSheet.Cells(3, 9).Value = "Log10(RepB)"
        ActiveSheet.Cells(3, 10).Value = "A=DupA+RepA"
        ActiveSheet.Cells(3, 11).Value = "B=DupB+RepB"
        ActiveSheet.Cells(3, 12).Value = "C=A+B"
        ActiveSheet.Cells(3, 13).Value = "Results"

     'activate cell for first information
    Range("A4:A4").Select
    Set c = ActiveCell

    For j = 1 To nrows / 4
    Set d = d.Offset(0, -(n + 1))
    c.Value = d.Value
```

```
    Set d = d.Offset(0, 1 + n)
    Set c = c.Offset(0, 1)
        For k = 1 To 4
                With c
                c.Value = d.Value
                Set c = c.Offset(0, 4)
                c.Value = Log10(d.Value)
                Set c = c.Offset(0, -3)
                Set d = d.Offset(1, 0)
                End With

        Next k
        Set c = c.Offset(1, -5)
       Next j

'do calculations
' variable names take on cell reference from Johnson (1995)

nrecords = nrows
duplicates = 2
replicates = 2
sites = nrecords / (duplicates + replicates)

Worksheets(element).Activate
Range("J4:J4").Select
Set e = ActiveCell
                        For j = 0 To (nrows / 4) - 1
                              e.Offset(j, 0).Value = e.Offset(j, -4).Value
+ e.Offset(j, -3).Value
                              e.Offset(j, 1).Value = e.Offset(j, -2).Value
+ e.Offset(j, -1).Value
                              e.Offset(j, 2).Value = e.Offset(j, 0).Value +
e.Offset(j, 1).Value
                        Next j

 Range("M4:M4").Select
 Set e = ActiveCell
 Set myRange = Range(Cells(4, 12), Cells((4 + (sites) - 1), 12))
 e.Value = Application.WorksheetFunction.Sum(myRange)

 Range("M5:M5").Select
 Set e = ActiveCell
 Set myRange = Range(Cells(4, 6), Cells((4 + (sites) - 1), 9))
 e.Value = Application.WorksheetFunction.SumSq(myRange)
 k2 = e.Value
 Range("M6:M6").Select
 Set e = ActiveCell
 Set myRange = Range(Cells(4, 10), Cells((4 + (sites) - 1), 11))
 e.Value = (Application.WorksheetFunction.SumSq(myRange)) / 2
 k3 = e.Value
 Range("M7:M7").Select
 Set e = ActiveCell
 Set myRange = Range(Cells(4, 12), Cells((4 + (sites) - 1), 12))
 e.Value = (Application.WorksheetFunction.SumSq(myRange)) / (2 * 2)
 k4 = e.Value
Range("M8:M8").Select
 Set e = ActiveCell
 Set myRange = Range(Cells(4, 13), Cells(4, 13))
 e.Value = (Application.WorksheetFunction.SumSq(myRange)) / (duplicates *
replicates * (sites))
    k5 = e.Value

'Degrees of Freedom
    'TOTAL
    q12 = (sites * duplicates * replicates) - 1
```

```
    'SITE
    q14 = sites - 1
    'SAMPLE
    q16 = sites * (duplicates - 1)
    'ERROR
    q18 = sites * duplicates * (replicates - 1)

'Sum of squares
    'SITE
    r14 = k4 - k5
    'SAMPLE
    r16 = k3 - r14 - k5
    'ERROR
    r18 = k2 - k5 - r14 - r16
    'TOTAL
    r12 = r14 + r16 + r18

'Mean of Squares
    'SITE
    s14 = r14 / q14
    'SAMPLE
    s16 = r16 / q16
    'ERROR
    s18 = r18 / q18
    'TOTAL
    s12 = r12 / q12

'F(calc)
    'SITE
    t14 = s14 / s16
    'SAMPLE
    t16 = s16 / s18

'Variance component
    'SITE
    u14 = (s14 - s16) / (duplicates * replicates)
    'SAMPLE
    u16 = (s16 - s18) / replicates
    'ERROR
    u18 = s18 / 1
    'TOTAL
    u12 = u14 + u16 + u18

'variance of total
    'SITE
    v14 = (u14 / u12) * 100
    'SAMPLE
    v16 = (u16 / u12) * 100
    'ERROR
    v18 = (u18 / u12) * 100
    'TOTAL
    v12 = (u12 / u12) * 100


'place all the results in the last column of the current worksheet
Worksheets(element).Activate
Range("n4:n4").Select
Set e = ActiveCell
e.Value = q12
Set e = e.Offset(1, 0)
e.Value = q14
Set e = e.Offset(1, 0)
e.Value = q16
Set e = e.Offset(1, 0)
e.Value = q18
```

```
Set e = e.Offset(1, 0)
e.Value = r12
Set e = e.Offset(1, 0)
e.Value = r14
Set e = e.Offset(1, 0)
e.Value = r16
Set e = e.Offset(1, 0)
e.Value = r18
Set e = e.Offset(1, 0)
e.Value = s12
Set e = e.Offset(1, 0)
e.Value = s14
Set e = e.Offset(1, 0)
e.Value = s16
Set e = e.Offset(1, 0)
e.Value = s18
Set e = e.Offset(1, 0)
e.Value = t14
Set e = e.Offset(1, 0)
e.Value = t16
Set e = e.Offset(1, 0)
e.Value = u12
Set e = e.Offset(1, 0)
e.Value = u14
Set e = e.Offset(1, 0)
e.Value = u16
Set e = e.Offset(1, 0)
e.Value = u18
Set e = e.Offset(1, 0)
e.Value = v12
Set e = e.Offset(1, 0)
e.Value = v14
Set e = e.Offset(1, 0)
e.Value = v16
Set e = e.Offset(1, 0)
e.Value = v18
Set e = e.Offset(1, 0)

    ' Fill array with values.

            listofnames(n) = element
            Results(n, 1) = v14
            Results(n, 2) = v16
            Results(n, 3) = v18
            Results(n, 4) = v12


    Next n

'add a worksheet and name it summary

    Sheets.Add
    ActiveSheet.name = "Summary"

'place standard text into summary table
Range("a1:a1").Select
Set e = ActiveCell
e.Value = tit
Set e = e.Offset(2, 0)
e.Value = "Element"
Set e = e.Offset(0, 1)
e.Value = "Between Site %"
Set e = e.Offset(0, 1)
e.Value = "Between Sample %"
Set e = e.Offset(0, 1)
```

```
e.Value = "Within Sample %"
Set e = e.Offset(0, 1)
e.Value = "Total %"

'fill table with results
Range("a4:a4").Select
Set e = ActiveCell
For i = 1 To ncols
e.Value = listofnames(i)

   For n = 1 To 4
   Set e = e.Offset(0, 1)
   e.Value = Results(i, n)
   Next n
Set e = e.Offset(1, -4)
Next i

'tidy up summary table
    'format numbers to two decimal places
    Columns("B:D").Select
    Selection.NumberFormat = "0.00"
    Rows("3:3").Select
    With Selection.Font
        .name = "Arial"
        .FontStyle = "Bold"
        .Size = 10
        .Strikethrough = False
        .Superscript = False
        .Subscript = False
        .OutlineFont = False
        .Shadow = False
        .Underline = xlUnderlineStyleNone
        .ColorIndex = xlAutomatic
    End With


    'set all text in colum A to be in bold
    Columns("A:A").Select
    With Selection.Font
        .name = "Arial"
        .FontStyle = "Bold"
        .Size = 10
        .Strikethrough = False
        .Superscript = False
        .Subscript = False
        .OutlineFont = False
        .Shadow = False
        .Underline = xlUnderlineStyleNone
        .ColorIndex = xlAutomatic
    End With
    'set column widths
  Columns("B:E").Select
    Selection.ColumnWidth = 12
    With Selection
        .HorizontalAlignment = xlCenter
        .VerticalAlignment = xlBottom
        .WrapText = False
        .Orientation = 0
        .ShrinkToFit = False
        .MergeCells = False
    End With
    'put a border around the table
    Range(Cells(3, 1), Cells(3 + ncols, 5)).Select
    Selection.Borders(xlDiagonalDown).LineStyle = xlNone
    Selection.Borders(xlDiagonalUp).LineStyle = xlNone
```

```
    With Selection.Borders(xlEdgeLeft)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    With Selection.Borders(xlEdgeTop)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    With Selection.Borders(xlEdgeBottom)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    With Selection.Borders(xlEdgeRight)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    Selection.Borders(xlInsideVertical).LineStyle = xlNone
    Selection.Borders(xlInsideHorizontal).LineStyle = xlNone

    'put a border around the column headings
    Range(Cells(3, 1), Cells(3, 5)).Select
    Selection.Borders(xlDiagonalDown).LineStyle = xlNone
    Selection.Borders(xlDiagonalUp).LineStyle = xlNone
    With Selection.Borders(xlEdgeLeft)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    With Selection.Borders(xlEdgeTop)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    With Selection.Borders(xlEdgeBottom)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With
    With Selection.Borders(xlEdgeRight)
        .LineStyle = xlContinuous
        .Weight = xlThin
        .ColorIndex = xlAutomatic
    End With

    'set column headings to wrap round
    Rows("3:3").Select
    With Selection
        .VerticalAlignment = xlBottom
        .WrapText = True
        .Orientation = 0
        .AddIndent = False
        .ShrinkToFit = False
        .MergeCells = False
    End With
  End Sub
          Static Function Log10(X)
          Log10 = Log(X) / Log(10#)
          End Function
```

Appendix 2 : Example of data format

This is an example of the data format that is placed in the "Data" worksheet:

Urban Soil Control data

Site	Control	MnO	FE_2O_3	V	Cr	Co	Ba	Ni	Cu	Zn	As	Mo	Pb
602276	DUP A	0.059	4.62	76	69	16	425	22	16	226	11	1.7	68
602286	SUB A	0.06	4.59	75	74	16	432	22	15	225	12	2.1	68
602281	DUP B	0.057	4.4	73	69	14	442	19	16	211	10	1.9	51
602278	SUB B	0.057	4.41	75	71	14	436	20	17	211	10	1.8	52
602331	DUP A	0.065	4.16	130	105	26	603	43	28	52	4	0.1	14
602377	SUB A	0.066	4.16	127	105	26	606	44	29	51	6	0.6	15
602337	DUP B	0.06	4.26	131	104	25	593	41	28	65	6	0.7	20
602358	SUB B	0.061	4.26	131	107	25	592	42	28	64	6	0.7	20
602404	DUP A	0.067	4.26	142	111	24	651	41	31	91	9	0.5	36
602462	SUB A	0.068	4.23	140	111	24	640	39	33	90	9	0.7	36
602496	DUP B	0.081	4.88	129	101	28	631	40	36	201	8	1.1	56
602453	SUB B	0.079	4.83	126	99	29	623	41	37	202	8	1	57
602571	DUP A	0.181	4.68	86	79	17	514	24	27	246	11	1.1	104
602560	SUB A	0.182	4.7	83	80	16	515	24	27	245	12	1.5	103
602566	DUP B	0.136	4.9	84	83	18	511	25	26	205	10	1.4	86
602534	SUB B	0.133	4.94	83	81	17	511	25	27	206	11	1.4	85
602676	DUP A	0.022	2.57	211	122	16	362	43	87	120	5	1.2	69
602686	SUB A	0.021	2.56	209	123	15	364	44	87	120	5	1.2	70
602681	DUP B	0.029	2.53	217	120	20	401	42	85	178	6	0.7	110
602678	SUB B	0.029	2.53	218	118	21	405	43	85	178	5	1	110
602731	DUP A	0.026	2.39	66	76	13	103	12	7	65	7	1.8	50
602777	SUB A	0.024	2.38	66	79	13	105	12	7	65	7	1.8	50
602737	DUP B	0.042	3.39	79	85	24	115	18	8	104	11	2.6	70
602758	SUB B	0.043	3.39	82	90	25	126	18	7	104	11	2.2	70

Appendix 3 : ANOVA results sheet for each element

Urban Soil Control data

Pb

Site	DupA	RepA	DupB	RepB	Log10(DupA)	Log10(RepA)	Log10(DupB)	Log10(RepB)	A=DupA + RepA	B=DupB + RepB	C=A+B	Results		
	68	68	51	52	1.832508913	1.832508913	1.707570176	1.716003344	3.6650178	3.4235735	7.0885913	387.80245	199	***Degrees of freedom -TOTAL***
602331	14	15	20	20	1.146128036	1.176091259	1.301029996	1.301029996	2.3222193	2.60206	4.9242793	788.47745	49	***Degrees of freedom -SITE***
602404	36	36	56	57	1.556302501	1.556302501	1.748188027	1.755874856	3.112605	3.5040629	6.6166679	788.46646	50	***Degrees of freedom -SAMPLE***
602571	104	103	86	85	2.017033339	2.012837225	1.934498451	1.929418926	4.0298706	3.8639174	7.8937879	787.82271	100	***Degrees of freedom -ERROR***
602676	69	70	110	110	1.838849091	1.84509804	2.041392685	2.041392685	3.6839471	4.0827854	7.7667325	751.95368	36.523769	***Sum of Squares - TOTAL***
602731	50	50	70	70	1.698970004	1.698970004	1.84509804	1.84509804	3.39794	3.6901961	7.0881361		35.869022	***Sum of Squares - SITE***
388131	28	30	27	27	1.447158031	1.477121255	1.431363764	1.431363764	2.9242793	2.8627275	5.7870068		0.6437525	***Sum of Squares - SAMPLE***
388204	63	64	92	90	1.799340549	1.806179974	1.963787827	1.954242509	3.6055205	3.9180303	7.5235509		0.0109947	***Sum of Squares - ERROR***
388371	113	115	136	138	2.053078443	2.06069784	2.133538908	2.139879086	4.1137763	4.273418	8.3871943		0.1835365	***Mean of Squares - TOTAL***
388476	155	155	197	200	2.190331698	2.190331698	2.294466226	2.301029996	4.3806634	4.5954962	8.9761596		0.7320209	***Mean of Squares - SITE***
388531	494	481	334	336	2.693726949	2.682145076	2.523746467	2.526339277	5.375872	5.0500857	10.425958		0.012875	***Mean of Squares - SAMPLE***
388604	79	77	74	75	1.897627091	1.886490725	1.86923172	1.875061263	3.7841178	3.744293	7.5284108		0.0001099	***Mean of Squares - ERROR***
388704	301	297	209	210	2.478566496	2.472756449	2.320146286	2.322219295	4.9513229	4.6423656	9.5936885		56.855771	***F(Calc) - SITE***
388876	150	150	118	117	2.176091259	2.176091259	2.071882007	2.068185862	4.3521825	4.1400679	8.4922504		117.10235	***F(Calc) - SAMPLE***
388976	1900	1900	1300	1300	3.278753601	3.278753601	3.113943352	3.113943352	6.5575072	6.2278867	12.785394		0.186279	***Variance Component - TOTAL***
389004	227	226	353	354	2.356025857	2.354108439	2.547774705	2.549003262	4.7101343	5.096778	9.8069123		0.1797865	***Variance Component - SITE***
389171	351	349	404	410	2.545307116	2.542825427	2.606381365	2.612783857	5.0881325	5.2191652	10.307298		0.0063826	***Variance Component - SAMPLE***
389276	34	35	38	38	1.531478917	1.544068044	1.579783597	1.579783597	3.075547	3.1595672	6.2351142		0.0001099	***Variance Component - ERROR***
389331	107	107	172	167	2.029383778	2.029383778	2.235528447	2.222716471	4.0587676	4.4582449	8.5170125		100	***% variance - Total***
389404	36	36	36	34	1.556302501	1.556302501	1.556302501	1.531478917	3.112605	3.0877814	6.2003864		96.514637	***% variance - Site***
600076	30	32	30	30	1.477121255	1.505149978	1.477121255	1.477121255	2.9822712	2.9542425	5.9365137		3.4263406	***% variance - Sample***
600131	27	27	28	27	1.431363764	1.431363764	1.447158031	1.431363764	2.8627275	2.8785218	5.7412493		0.0590228	***% variance - Error***
600371	44	44	41	41	1.643452676	1.643452676	1.612783857	1.612783857	3.2869054	3.2255677	6.5124731			
600476	37	38	33	34	1.568201724	1.579783597	1.51851394	1.531478917	3.1479853	3.0499929	6.1979782			
600531	154	156	155	156	2.187520721	2.1931246	2.1903317	2.1931246	4.380645	4.383456	8.764102			

References

JOHNSON, C. C. 1995. ANOVA analysis using EXCEL. *Unpublished memo from C C Johnson to J W Baldock dated 13 June 1995.*

JOHNSON, C. C. 2000. Making histograms using MS EXCEL. WORKSHOP DE CARTOGRAPHIE GEOCHEMIQUE REGIONALE. Session 5 (Geochemical data Interpretation) Rabat, Maroc juin 2000.

LISTER, T.R. 2002. Analysis of variance (ANOVA) of G-BASE sub-surface soil data from 11 urban centres in England and Wales. *British Geological Survey Internal Report*, IR/02/009.

RAMSEY, M.H., THOMPSON, M. AND HALE, M. 1992. Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. *Journal of Geochemical Exploration*, 44, 23-36.

SINCLAIR, A.J. 1983. Univariate Analysis. Chapter 3 p 69-74 in *Handbook of Exploration Geochemistry. Volume 2 - Statistics and data Analysis in Geochemical Prospecting* . HOWARTH R J.(editor). (Amsterdam: Elsevier Scientific Publishing.) ISBN 0-444-42038x