EAR 419/619 – Aqueous Geochemistry

Lab #3

Data Analysis + Statistics

Water samples were collected from the Meadowbrook Creek in order to evaluate chemistry changes along the stream in an urban setting. Various instruments were used to determine solute concentrations in the water samples (refer to “Extended Reading” for instrument details). Here, students will evaluate calibration standards to ensure data quality and estimate analytical error (the uncertainty in a data value attributed to the instrument used for analysis). Reported values that fall outside the calibration curve are considered *censored data* and need to be designated as *below detection limit* (< DL or BDL) or *above detection limit* (> DL), with the detection limit defined here as the lowest and highest calibration standard concentrations.

**Objectives**

* Construct calibration curves from calibration standards
* Perform statistical analyses to determine analytical uncertainty for each solute
* Plot water chemistry data in order to compare the chemical composition of water samples collected from different sites; graphs will be shared with the class during a later class period

**Data**

* Data will be obtained from Excel spreadsheets uploaded to Blackboard:
  + Concentrations of major cations (K, Na, Mg, Ca) and minor cations (total Fe, As, Ba, Cr, Sr, Zn, Mn, Al) in water samples
  + Field measurements (EC, pH, temperature, distance upstream)

**Definitions**

* calibration curve – method for determining the concentration of a substance in an unknown sample by determining instrument response to standards with known concentrations
* error – the deviation of a measurement from its “true” value
* analytical error – the uncertainty in a given value that is associated with the ability of the instrument to measure that value; a measure of the accuracy of instrument predictions, calculated as the standard error of the linear regression (see below)
* standard deviation – the amount of variation in a population; STDEV function in excel
* blank – a solution that contains very low concentration of solute (e.g., ultrapure water); used to assess background contamination and instrument noise
* method blank – a solution with the same *matrix* as your samples; used to assess if a solvent or extra solute (e.g., HCl, HNO3) influences instrument response
* analyte – a chemical species that is being identified and measured
* matrix – the components of a sample other than the analyte of interest; the “background”

**Report**

1. (25 pts) Demonstration of calibration curve
   1. In the ‘Calibration Curve’ tab, for each of the three elements: Na, Ca, and As, construct a calibration curve from their calibration standards. Include a linear regression, linear regression equation, and R2 value.
   2. Report linear regression statistics in a table that identifies each parameter (See page 3 of Lab 3 instruction). Calculate the analytical standard error for Na, Ca, and As concentrations.
   3. For the samples, use the raw absorbance data to calculate Na, Ca, and As concentrations in mg/L, then convert to mmol/L. For samples below the detection limit, write “<DL”.
2. (10 pts) ICP-OES data

The processed data are provided for you as mg/L. In the ICP-OES data tabs, fill out the tables that list the concentration of each cation or anion in mmol/L. Then answer the following questions in one to two paragraphs each.

* 1. Which cation(s) were the most abundant? Which were the least abundant? Did concentrations of the most abundant cation differ amongst water samples? If so, how?
  2. Did concentrations of the metals (K, Na, Mg, Ca, As, Sr, Cr, Zn, Ba, Mn, Al, Fe) differ amongst water samples, and if so, how?

1. (5 pts) Compile the final data in the summary tab. Report here the concentrations of the listed species in solution.
2. (20 pts) Provide at least two graphs that plot water chemistry data. The purpose of the graphs is to visually communicate how water chemistry varies between different water sources. You may choose any of the water chemistry parameters to plot, but focus on parameters that will address your question and hypothesis. Add error bars to the data points in your plots, using the analytical error of the solute as the error value. Refer to this link (<https://support.microsoft.com/en-us/office/add-change-or-remove-error-bars-in-a-chart-e6d12c87-8533-4cd6-a3f5-864049a145f0>) for how to add error bars in Excel.
3. Upload results to Blackboard in one Excel spreadsheet (data) AND one PDF file (graphs and answers). For all files, **include your last name in the file name**.

**Linear Regression**

The LINEST function in Excel can provide statistics on a line drawn to best fit a series of data points. The equation for a line is y = mx + b, where the dependent y-value is a function of the independent x-value, m is the slope of the line, and b is the intercept. This value will be used to determine analytical error for the calibration curves but can also be used to determine whether linear relationships between dependent and independent variables are significant for any set of data. See ‘Calibration Curve’ tab to see how these calculations are performed.

**Standard error of the x-value (sx), used here for analytical error, is calculated as sx = sy/m**

LINEST in Excel:

1) One the same tab containing known x and y values, enter “=linest(known\_y’s, known\_x’s, true, true)” into an empty cell, replacing “known\_y’s” and “known\_x’s” with the array of cell numbers that hold the values

2) hit “Enter” to fill the block of cells with the linear regression statistics in the order shown below

|  |  |  |
| --- | --- | --- |
|  | A | B |
| 1 | mn | b |
| 2 | sen | seb |
| 3 | R2 | sey |
| 4 | F | df |
| 5 | ssreg | ssresid |

mn: slope of the line

b: value of the y-intercept

sen: standard error value for the coefficient *m*

seb: standard error value for the y-intercept *b*

R2: coefficient of determination (should be > 0.95 for a calibration curve; ideally R2 = 0.99, or even 0.9999 for a very precise instrument)

sey: standard error for the y-value

F: F-statistic; used to determine if the trend is significant by comparing to F-critical (a value obtained from a table)

df: degrees of freedom; used to find F-critical from a table of F-values

ssreg: the regression sum of squares; a measure of the difference between actual y-values and y-values reported by the line

ssresid: the residual sum of squares

From office.microsoft.com: In regression analysis, Microsoft Excel calculates for each point the squared difference between the y-value estimated for that point and its actual y-value. The sum of these squared differences is called the residual sum of squares, ssresid. Microsoft Excel then calculates the total sum of squares, sstotal. When const = TRUE, or omitted, the total sum of squares is the sum of the squared differences between the actual y-values and the average of the y-values. Then regression sum of squares, ssreg, can be found from: ssreg = sstotal - ssresid. The smaller the residual sum of squares is, compared with the total sum of squares, the larger the value of the coefficient of determination, R2, which is an indicator of how well the equation resulting from the regression analysis explains the relationship among the variables. R2 equals ssreg/sstotal.