GEOSCIENCES DEPARTMENT

STANDARD OPERATING PROCEDURE

TITLE: Outlier Determination

WRITTEN BY: Bruce Wegter

APPROVAL

Supervisor Greg Rahn **Date:** 02/10/2010

Dept. Representative Eugene Domack **Date:** 2/10/2010

Effective Date:

1. PURPOSE

This document describes the statistical procedures that will be used to determine when an experimentally derived value is considered to be an outlier and not representative of the sample result population.

2. SCOPE

This SOP sets the guidelines for determining, documenting, and dealing with a value that is considered to be an outlier.

3. RESPONSIBILITY

It is the responsibility of all persons assigned to a task which can and/or will involve standard and sample data to understand and use these procedures.

4. DEFINITIONS

4.1. Outlier - an **outlier** is defined as an observation that is generated from a different model or a different distribution than was the main "body" of data. Although this definition implies that an outlier may be found anywhere within the range of observations, it is natural to suspect and examine as possible outliers only the extreme sample values in a population. The rejection of suspect observations must be based exclusively on an objective criterion and not on subjective or intuitive grounds. This can be achieved by using statistically sound tests for the detection of outliers.

4.2. Dixon's Q-test for the Detection of Outliers - This is the simpler test of this type (this is usually the only one described in textbooks of Analytical Chemistry in the chapters of data treatment). This test allows examination of one (and only one) observation from a small set of replicate observations (typically 3 to 10) to determine if the value can be statistically rejected or not. In case of the detection and rejection of an outlier, a Q-test **cannot** be reapplied on the set of the remaining experimentally determined observations. This test is typically applied to “standards” which are analyzed in replicate.

5. PROCEDURES

5.1. Prior to implementation of any statistical outlier testing (e.g., Dixon Q-Test) on a suspected data point, an"assignable cause" for the outlier value should be sought. If no assignable cause is indicated for the suspected aberrant value, then the Dixon Q-Test should be implemented.

5.2. Application of the Dixon Q-Test:

5.2.1. Arrange the number of values "N" in ascending order.

$$x_1 < x_2 < \ldots < x_N$$

5.2.2. Calculate the experimental Q-value (Q_{exp}). This is a ratio of the difference of the suspect value from the next nearest value divided by the range of the values. For testing x_1 or x_N (as the only possible outliers) we use the following Q_{exp} values:

$$Q_{exp} = (x_2 - x_1) / (x_N - x_1) \quad \text{or} \quad Q_{exp} = (x_N - x_{N-1}) / (x_N - x_1)$$

5.2.3. The obtained Q_{exp} value is compared to a critical Q-value (Q_{crit}) found in the following table. This critical value should correspond to the **confidence level** (CL) we have decided to apply in the test (at minimum: CL=95%).

Table of critical values of Q_{crit}

N	Q_{crit} (CL:90%)	Q_{crit} (CL:95%)	Q_{crit} (CL:99%)
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

[from: *D.B. Rorabacher, Anal. Chem. 63 (1991) 139*]

5.2.4. If $Q_{exp} > Q_{crit}$, then the suspect value can be characterized as an outlier and it can be rejected, if not, the suspect value must be retained and used in all subsequent calculations.

5.2.5. All calculations used for this determination shall be included with the data. This could be an additional EXCEL worksheet, hard copy attached to the standards info, or other convenient method.

7. REFERENCE DOCUMENTS

7.1. *D.B. Rorabacher, Anal. Chem. 63 (1991) 139*

8. REVISIONS AND REASONS

8.1. Original