09/28/1999

Dealing with outliers

There are many statistical tests available for identifying outliers. One that is relatively easy to use is the Grubbs test.

$$\mathbf{Z} = \frac{\text{mean - questionable data point}}{\text{SD}}$$

- Only test the highest (and/or lowest) value
- Ignore the sign of the "Z" value….is always "+"
- Include the suspected outlier when calculating mean and SD
- If the calculated Z-value is greater than the criterion Z value for that number of data points, then the value is an outlier

N	Critical Z
18	2.65
19	2.68
20	2.71
21	2.73
22	2.76
23	2.78
24	2.80
25	2.82
26	2.84
27	2.86
28	2.88
29	2.89
30	2.91
35	2.98
40	3.04
50	3.13
60	3.20

Exercise

DATE	%R
11/01	81
11/04	87
11/07	90
11/10	93
11/13	80
11/16	82
11/19	91
11/22	94
11/25	83
11/28	**110**
12/01	91
12/04	92
12/07	80
12/10	88
12/13	94
12/16	92
12/19	83
12/22	**80**
12/25	91
12/28	95

1. Calculate the mean and SD
   Mean = 88.85 SD = 7.278 Limits =67 - 110.7
   N = 20

2. Test the high value (110)
   $Z = \frac{110 - 88.85}{7.278} = 2.9058$

3. Test the low value (80)
   $Z = \frac{88.85 - 80}{7.278} = 1.2159$

4. Discard outliers; since Z_{110} > criterion, 110 is an outlier
   re-calculate mean and SD

   Mean =87.737 SD= 5.4553 Limits = 71 – 104

NOTE: Step 4 may also require a re-check for additional outliers

Wis. Dept. of Natural Resources
Laboratory Certification Program
Regional Lab Auditing Staff