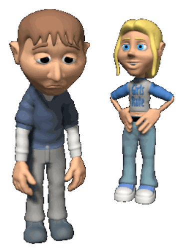
## Rejection of data.

Outliers

Rule of the huge error

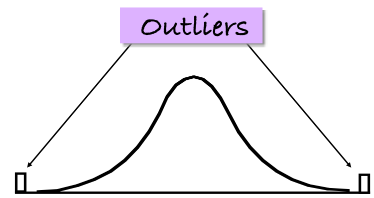
Dixon test - Q test

Grubbs test



## Rejection of data

Sometimes we know that a data point looks bad (outlier). We can't just pitch it out - there must be a basis for rejection data.




## Outliers

- Values that do not belong to a population.
- Can be based on knowing that the value is truly different or
- Demonstrated that it falls outside of a specified probability.

When rejecting data resulting from replicate measurements, you need to use an established statistical method.

## Rule of the Huge Error

Assumes that you have some idea as to what the standard deviation should be or can calculate it.

If M > 4 then you can reject the point.

This is simply a crude t test. It is only useful for discarding obviously 'bad' data. No need for any sort of 'data table.'

$$M = \frac{|suspect - mean|}{s}$$

## Dixon test

Assumes

Mean and standard deviation are unknown.

Data is normally distributed.

Steps

1. Rank the data: $x_1 < x_2 < \ldots < x_n$
2. Choose confidence level
3. Calculate ratio (based on n)
4. Look up proper value
5. If ratio > table value then reject

Also called the 'Q test.'

## Dixon test - ratios

The ratio used is based on the number of data points and if you are evaluating the highest or lowest value.

| # of points | Test | Low | High |
|---|---|---|---|
| 3 - 7 | $\tau_{10}$ | $\frac{X_2 - X_1}{X_n - X_1}$ | $\frac{X_n - X_{n-1}}{X_n - X_1}$ |
| 8 - 10 | $\tau_{11}$ | $\frac{X_2 - X_1}{X_{n-1} - X_1}$ | $\frac{X_n - X_{n-1}}{X_n - X_2}$ |
| 11 - 13 | $\tau_{21}$ | $\frac{X_3 - X_1}{X_{n-1} - X_1}$ | $\frac{X_n - X_{n-2}}{X_n - X_2}$ |
| 14 - 25 | $\tau_{22}$ | $\frac{X_3 - X_1}{X_{n-2} - X_1}$ | $\frac{X_n - X_{n-2}}{X_n - X_3}$ |

## Dixon test - partial table

| Statistic | n | Risk of false rejection. 0.5% | 1% | 5% | 10% |
|---|---|---|---|---|---|
| $\tau_{10}$ | 3 | .994 | .988 | .941 | .886 |
| | 4 | .926 | .889 | .765 | .679 |
| | 5 | .821 | .780 | .642 | .557 |
| | 6 | .740 | .698 | .560 | .482 |
| | 7 | .680 | .637 | .507 | .434 |
| $\tau_{11}$ | 8 | .725 | .683 | | |
| | 9 | .677 | .635 | | |
| | 10 | .639 | .679 | | |
| $\tau_{21}$ | 11 | .713 | .642 | | |
| | 12 | .675 | .615 | | |
| | 13 | .649 | | | |
| $\tau_{22}$ | 14 | .674 | | | |
| | 15 | .647 | | | |

rejection

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Peak area | | use τ10 calculation | |
| 2 | 14613 | | | |
| 3 | 14734 | | (Xn - X n-1) / (Xn - X1) | |
| 4 | 14805 | | 0.68106618 | |
| 5 | 14992 | | | |
| 6 | 15104 | | | |
| 7 | 15045 | | Dixon values for n = 9 | |
| 8 | 15301 | | 1% | 0.635 |
| 9 | 15307 | | 0.5% | 0.677 |
| 10 | 16789 | | | |
| 11 | | | Data point can be rejected. | |

grubbs Dixon Sheet3

Ready Sum=0

## Example

rejection

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Peak area | | use τ11 calculation | |
| 2 | 14613 | | | |
| 3 | 14745 | | (Xn - X n-1) / (Xn - X2) | |
| 4 | 14734 | | 0.72504892 | |
| 5 | 14805 | | | |
| 6 | 14992 | | | |
| 7 | 15104 | | Dixon values for n = 10 | |
| 8 | 15045 | | 1% | 0.679 |
| 9 | 15301 | | 0.5% | 0.639 |
| 10 | 15307 | | | |
| 11 | 16789 | | Point can be rejected. | |

grubbs Dixon Sheet3

Ready Sum=0

## Example

## Grubbs test

This approach requires calculation of the mean and standard deviation

- Rank points
- Pick suspect point
- Calculate mean and standard deviation using all points.
- Calculate T. T = |mean - suspect| / $s_X$
- Look up T on table.
- If T > table value then reject it.

## Grubbs test - partial table

| n | Risk of false rejection 0.1% | 1% | 5% |
|---|---|---|---|
| 3 | 1.155 | 1.155 | 1.153 |
| 4 | 1.496 | 1.492 | 1.463 |
| 5 | 1.780 | 1.749 | 1.672 |
| 6 | 2.011 | 1.944 | 1.822 |
| 7 | 2.201 | 2.097 | 1.938 |
| 8 | 2.358 | 2.221 | 2.032 |
| 9 | 2.492 | 2.323 | 2.110 |
| 10 | 2.606 | 2.410 | 2.176 |

| Fe, ppm | | | |
|---|---|---|---|
| 1.45 | Mean | 1.51 | |
| 1.47 | Std. Dev. | 0.05601587 | |
| 1.47 | | | |
| 1.48 | T | abs(mean - suspect) / Std. Dev. | |
| 1.49 | | 2.57070002 | |
| 1.50 | | | |
| 1.50 | Table values for n = 10 | | |
| 1.52 | 0.10% | 2.606 | |
| 1.53 | 1% | 2.41 | |
| 1.65 | 5% | 2.176 | |
| | We can discard the suspect value | | |
| | with only a 1% of false rejection. | | |

## Grubbs example