# The Dixon Q-Test for Detecting Outliers

## Introduction to Hypothesis Testing

# Outliers

**Outliers**

- In Science, it is quite often the case that an outlier measurement is the result of faulty or unclean equipment, or a data entry error etc.
- **Important:** Care must be take to assess that the measurement is an outlier, rather than an unusual result that is in fact genuine.
- It is good practice not to remove outliers from an overall analysis (*not permanently anyway*).
- However you may omit suspected outliers and run the analysis a second time, then present all of the obtained results, with and without the outliers.

# Outliers

There may be an outlier, or multiple outliers, present in the data.

There are several formal hypothesis tests to determine presence of an outlier. The main ones that we will use are

- Grubbs' Test
- Dixon Test

## Dixon Q Test

- The Dixon's Q test, or simply the Q test, is used for identification and rejection of outliers.
- **(Important)** - This test is based upon the assumption of normality.
- This test should be used sparingly and never more than once in a data set.

# Hypothesis Testing : Steps when using Critical Value approach

1. Formally State the Null and Alternative Hypothesis
   - **ALWAYS** include a short written description of both hypotheses.
   - State hypotheses in mathematical notation where possible.
2. Calculate Test Statistic (TS).
3. Determine the Critical Value (CV).
4. Compare Test Statistic and Critical Value to make your conclusions

# Hypothesis Testing : Steps when using Critical Value approach

- Consider TS as a measure of the strength of evidence, and CV as a threshold that we have to cross for our evidence to be considered sufficiently strong.
- We usually use absolute value of Test Statistic $|TS|$.
- Key comparison

$$\text{is } |TS| > CV?$$

# Hypothesis Testing : Steps when using Critical Value approach

Key comparison

$$\text{is } |TS| > CV?$$

Yes: We **reject the Null** Hypothesis.
We have sufficient evidence against Null Hypothesis.

No: We **fail to reject** Null hypothesis.
We do not have sufficient evident against Null Hypothesis.

N.B. Note the terminology that we are using. Also note exactly what our conclusion is.

# Outliers

- Lets try out this Four Step Process for the Dixon Q-Test.

**Dixon Q Test: Example**

Consider the data set:

$$0.189,\ 0.167,\ 0.187,\ 0.183,\ 0.186,$$
$$0.182,\ 0.181,\ 0.184,\ 0.181,\ 0.177$$

**Testing Outliers with The Dixon Test**

**Step 1:** Hypotheses for the Dixon Test.

$H_0$ No Outlier Present in Data

$H_1$ There is an Outlier Present in Data

# Outliers

**Step 2:** Dixon Q Test Statistic

To apply a Q test for suspicious data, arrange the data in order of increasing values and calculate Q as defined:

$$\boxed{Q = \frac{\text{gap}}{\text{range}}}$$

Where gap is the absolute difference between the outlier in question and the closest number to it.

## Dixon Q Test: Example

Consider the data set:

0.189, 0.167, 0.187, 0.183, 0.186,
0.182, 0.181, 0.184, 0.181, 0.177

Now rearrange in increasing order:

0.167, 0.177, 0.181, 0.181, 0.182,
0.183, 0.184, 0.186, 0.187, 0.189

**Dixon Q Test: Example**

We hypothesize 0.167 is an outlier.
Calculate The Test Statistic $Q_{Test}$:

$$Q_{TS} = \frac{\text{gap}}{\text{range}}$$

$$Q_{TS} = \frac{0.177 - 0.167}{0.189 - 0.167} = 0.455.$$

# Outliers

**Step 3:** Dixon Q Critical Value

| 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 |

*Here: N is the sample size (it is usually denoted as n).*

# Outliers

**Step 4** : Dixon Q Test Decison Rule

- Choose the Critical Value based on sample size and significance level $\alpha$.
- In this table we work on the basis of confidence level.
- Let's use 95% as our confidence level.
  (i.e. 5% significance, i.e. $\alpha = 0.05$)

## Dixon Q Test: Example

- If $Q_{TS} > Q_{CV}$ , where $Q_{CV}$ is a critical value corresponding to the sample size and confidence level, then reject the null hypothesis.
- If $Q_{TS} \leq Q_{CV}$ , we fail to reject. null hypothesis. i.e. Not enough evidence.
- At 95% confidence, $Q_{TS} \leq Q_{CV}$ i.e $0.455 \leq 0.466$
- Therefore we dont have enough evidence to classify the lowest value 0.167 as an outlier.