Chapter 5
Determination of outliers

Determination of Outliers

- Odd or extreme data in a given set of measurement results must be rejected with statistical justification
- **PaHTa's Rule,**
- **Chauvenet ω Test**
- **Dixon's Q Test**
- **Thompson's *tau* τ Test**
- Other outlier tests include: Grubb's, Cochran's, Bartlett's, Hartley's, Levene's and Brown–Forsythe's.

Determination of Outliers

- **PaHTa's Rule**
- For a series of test values, $x_1, x_2, \ldots., x_n$
- If $v_i = x_i - \bar{x}$, with standard deviation s, and,
- if $v_i > 3*s$, then,
- x_i is a significant outlier.

Determination of Outliers

- **Chauvenet Test**
- For a series of test values, $x_1, x_2, \ldots, x_n$
- If $v_i = x_i - \bar{x}$, with standard deviation *s*, and
- when $|v_i| > \omega_i * s$, where ω_i is referred to the Chauvenet's table, based on the number of data, n, then the extreme value x_i is an outlier.

Table 2: Chauvenet's ω_i *values against number of data, n*

n	ω	n	ω	n	ω
3	1.38	**13**	2.07	**23**	2.3
4	1.53	**14**	2.10	**24**	2.3
5	1.65	**15**	2.13	**25**	2.3
6	1.73	**16**	2.15	**30**	2.4
7	1.80	**17**	2.17	**40**	2.5
8	1.86	**18**	2.20	**50**	2.6
9	1.92	**19**	2.22	**75**	2.7
10	1.96	**20**	2.24	**100**	2.8
11	2.00	**21**	2.26	**200**	3.0
12	2.03	**22**	2.28	**500**	3.2

Determination of Outliers

- **Dixon's Q Test**
- *Q = | suspected value – nearest value | / (largest value – smallest value)*
- For values $X_1, X_2, \ldots, X_{n-1}, X_n$ where X_n is suspected to be extremely high:
  - For sets of 3 through 7 values:
  - $Q = (X_n - X_{n-1}) / (X_n - X_1)$
  - For sets of 8 through 12 values:
  - $Q = (X_n - X_{n-1}) / (X_n - X_2)$
  - For sets of 13 through 40 values:
  - $Q = (X_n - X_{n-2}) / (X_n - X_3)$

Determination of outliers

- For values X_1, X_2, ….., X_{n-1}, X_n where X_1 is suspected to be extremely low:
  - For sets of 3 through 7 values:
  - $Q = (X_2 - X_1) / (X_n - X_1)$
  - For sets of 8 through 12 values:
  - $Q = (X_2 - X_1) / (X_{n-1} - X_1)$
  - For sets of 13 through 40 values:
  - $Q = (X_3 - X_1) / (X_{n-2} - X_1)$

Compare the estimated Q against the critical values at n sets of tests in the Dixon's Table (Table 3 next page).

If Q > Dixon's critical value, the result is an outlier with 95% confidence.

Table 3 : Critical Values for the Dixon Test				
Test Criteria	**N**	**95%**	**99%**	
	3	0.970	0.994	
D(3…7) = [x(2) – x(1)] / [x(n) – x(1)]	4	0.829	0.926	
Or	5	0.710	0.821	
D(3…7) = [x(n) – x(n-1)]) / [x(n) – x(1)]	6	0.628	0.740	
(Whichever is the greater)	7	0.569	0.680	
	8	0.608	0.717	
D(8…12) = [x(2) – x(1)] / [x(n-1) – x(1)]	9	0.564	0.672	
Or	10	0.530	0.635	
D(8…12) = [x(n) – x(n-1)]) / [x(n) – x(2)]	11	0.502	0.605	
(Whichever is the greater)	12	0.479	0.579	

Critical Values for Dixon's Test

$D(13...40) = [x(3) - x(1)] / [x(n-2) - x(1)]$

Or

$D(13...40) = [x(n) - x(n-2)] / [x(n) - x(3)]$

(Whichever is the greater)

13	0.611	0.697	
14	0.586	0.670	
15	0.565	0.647	
16	0.546	0.633	
17	0.529	0.610	
18	0.514	0.594	
19	0.501	0.580	
20	0.489	0.567	
21	0.478	0.555	
22	0.468	0.544	
23	0.459	0.535	
24	0.451	0.526	
25	0.443	0.517	
26	0.436	0.510	
27	0.429	0.502	
28	0.423	0.495	
29	0.417	0.489	
30	0.412	0.483	
31	0.407	0.477	
32	0.402	0.472	
33	0.397	0.467	
34	0.393	0.462	
35	0.388	0.458	
36	0.384	0.454	
37	0.381	0.450	
38	0.377	0.446	
39	0.374	0.442	
40	0.371	0.438	

Example of Dixon's test

Repeated drug assay values

	Original		Sorted		
	Trial #	Value, %	Trial #	Value, %	
	1	98.0	6	96.8	
	2	98.5	1	98.0	
	3	99.0	2	98.5	
	4	98.6	4	98.6	
	5	99.3	7	98.8	
	6	96.8	3	99.0	
	7	98.8	5	99.3	
	8	99.4	8	99.4	

- Dixon's *Q* value = (98.0–96.8)/(99.3–96.8) = 1.20/2.50 = 0.48
- Dixon's critical value for $n = 8$ is 0.608 (95% confidence)
- Conclusion: Value 96.8% is not an outlier.

Determination of outliers

- **Thompson's tau (τ) Test**

145	155	153	154	158	161	148	155
147	146	156	155	159	160	172	160
157	153	147	154	157	158	149	152

- $\bar{x} =$ 154.6 and s = 6.00
- Note the suspected outlier : 172
- Calculate the ***absolute*** delta value δ = suspected value – mean x
- Here, δ = 172 – 154.6 = 17.4
- Use Thompson's critical value table (95% confidence)

Table of Critical Values for Thompson's τ at 95% Confidence Level

Sample size	τ	Sample size	τ
3	1.150	21	1.889
4	1.393	22	1.893
5	1.572	23	1.896
6	1.656	24	1.899
7	1.711	25	1.902
8	1.749	26	1.904
9	1.777	27	1.906
10	1.798	28	1.908
11	1.815	29	1.910
12	1.829	30	1.911
13	1.840	31	1.913
14	1.849	32	1.914
15	1.858	33	1.916
16	1.865	34	1.917
17	1.871	35	1.919
18	1.876	36	1.920
19	1.881	37	1.921
20	1.885	38	1.922

Determination of outliers

- From the table, $\tau = 1.899$
- Calculate the product of τ and s :
- Here, $\tau s = 1.899 \times 6.00 = 11.4$
- Compare δ and the product τs
- If $\delta > \tau s$, then the suspected value is an outlier or else, the suspected value is not an outlier.
- In this case, as $17.4 > 11.4$, the data 172 is indeed an outlier.