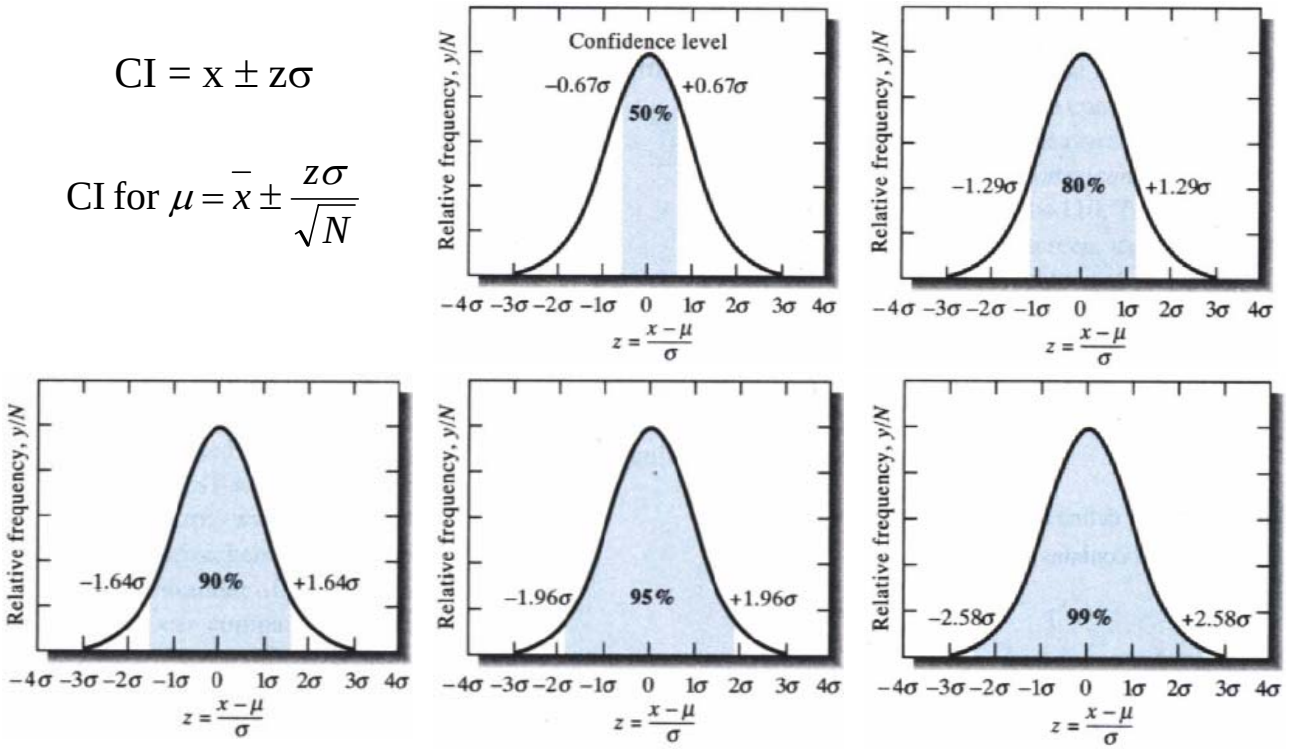
7 Statistical Data Treatment and Evaluation

7A Confidence Intervals

7A-1 Finding the confidence interval when σ is known or s is a good estimate of σ

$$CI = x \pm z\sigma$$

$$\text{CI for } \mu = \bar{x} \pm \frac{z\sigma}{\sqrt{N}}$$



Confidence Levels for Various Values of z

CL, %	50	68	80	90	95	95.4	99	99.7	99.9
z	0.67	1.00	1.28	1.64	1.96	2.00	2.58	3.00	3.29

Size of CI as a Function of the Number of Measurements Averaged (NMV)

NMV	1	2	3	4	5	6	10
Relative size of CI	1.00	0.71	0.58	0.50	0.45	0.41	0.32

Ex. 7-1 Calculate the 80% and 95% confidence limits for (a) the first entry (1108 mg/L glucose) in Ex. 6-2 and (b) the mean value (1100.3 mg/L) for month in 1 in the example. Assume that in each part s = 19 is a good estimate of σ.

(a) 80% CI = 1108 ± 1.28 × 19 = 1108 ± 24.3 mg/L, 1083.7 ~ 1132.3 mg/L

95% CI = 1108 ± 1.96 × 19 = 1108 ± 37.2 mg/L, 1070.8 ~ 1145.2 mg/L

(b) 80% $CL = 1100.3 \pm \frac{1.28 \times 19}{\sqrt{7}} = 1100.3 \pm 9.2$ mg/L, 1091.1 ~ 1109.5 mg/L

95% $CL = 1100.3 \pm \frac{1.96 \times 19}{\sqrt{7}} = 1100.3 \pm 14.1$ mg/L, 1086.2 ~ 1114.4 mg/L

Ex. 7-2 How many replicate measurements in month 1 in Ex. 6-2 are needed to decrease the 95 % confidence interval to 1100.3 ± 10.0 mg/L glucose?

$$\text{CI} = 10.0 = \frac{z\sigma}{\sqrt{N}} = \frac{1.96 \times 19}{\sqrt{N}}, \quad \sqrt{N} = \frac{1.96 \times 19}{10.0} = 3.724$$

$N = (3.724)^2 = 13.9$ 14 measurements are needed

7A-2 Finding the confidence interval when σ is unknown

for a single measurement, *x*,

$$t = \frac{x - \mu}{s}$$

for the mean of N measurements,

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

$$\text{CI for } \mu = \bar{x} \pm \frac{t\,s}{\sqrt{N}}$$

Table 7-3 Values of t for various levels of probability

Degrees of Freedom	80%	90%	95%	99%	99.9%
1	**3.08**	**6.31**	**12.7**	**63.7**	**637**
2	**1.89**	**2.92**	**4.30**	**9.92**	**31.6**
3	**1.64**	**2.35**	**3.18**	**5.84**	**12.9**
4	**1.53**	**2.13**	**2.78**	**4.60**	**8.61**
5	**1.48**	**2.02**	**2.57**	**4.03**	**6.87**
6	**1.44**	**1.94**	**2.45**	**3.71**	**5.96**
7	**1.42**	**1.90**	**2.36**	**3.50**	**5.41**
8	**1.40**	**1.86**	**2.31**	**3.36**	**5.04**
9	**1.38**	**1.83**	**2.26**	**3.25**	**4.78**
10	**1.37**	**1.81**	**2.23**	**3.17**	**4.59**
15	**1.34**	**1.75**	**2.13**	**2.95**	**4.07**
20	**1.32**	**1.73**	**2.09**	**2.84**	**3.85**
40	**1.30**	**1.68**	**2.02**	**2.70**	**3.55**
60	**1.30**	**1.67**	**20..**	**2.62**	**3.46**
∞	**1.28**	**1.64**	**1.96**	**2.58**	**3.29**

Ex. 7-3 A chemist obtained the following data for the alcohol content of a sample of blood: % C_2H_5OH: 0.084, 0.089 and 0.079. Calculate the 95% confidence interval for the mean assuming (a) the three results obtained are the only indication of the precision of the method and (b) from previous experience on hundreds of samples, we know that the standard deviation of the method s = 0.005% C_2H_5OH and is a good estimate of σ.

(a) $\sum x_i = 0.084 + 0.089 + 0.079 = 0.252, \quad \bar{x} = 0.252 / 3 = 0.084$

$$\sum x_i^2 = 0.007056 + 0.007921 + 0.006241 = 0.021218$$

$$s = \sqrt{\frac{0.021218 - (0.252)^2 / 3}{3 - 1}} = 0.0050\%\ \text{C}_2\text{H}_5\text{OH}$$

$$95\%\,\text{CI} = \bar{x} \pm \frac{t\,s}{\sqrt{N}} = 0.084 \pm \frac{4.30 \times 0.0050}{\sqrt{3}} = 0.084 \pm 0.012\%\ \text{C}_2\text{H}_5\text{OH}$$

(b) s = 0.005% is a good estimate of σ

$$95\%\,\text{CI} = \bar{x} \pm \frac{z\sigma}{\sqrt{N}} = 0.084 \pm \frac{1.96 \times 0.0050}{\sqrt{3}} = 0.084 \pm 0.006\%\,\text{C}_2\text{H}_5\text{OH}$$

7C Detection of Gross Errors

The Q TEST

$$Q_{\text{exp}} = \frac{|x_q - x_n|}{w} = \frac{|x_q - x_n|}{|x_{high} - x_{low}|}$$

if $Q_{exp} > Q_{crit}$, reject x_q

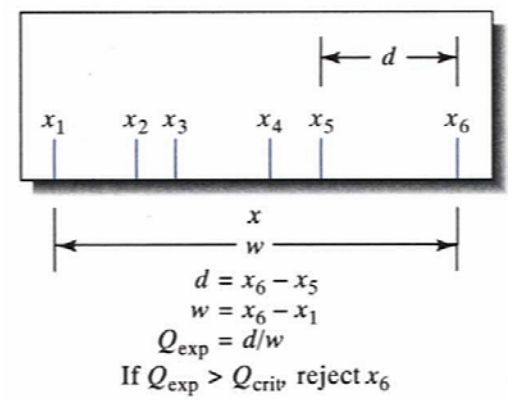


Fig 7-3 The Q test for outliers.

Table 7-4 Critical Values for the Rejection Quotient Q

No. of Observation	Q_{crit} (Reject if $Q_{exp} > Q_{crit}$)		
	90% CL	95% CL	99% CL
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

Ex. 7-6 The analysis of a calcite sample yielded CaO percentages of 55.95, 56.00, 56.04, 56.08 and 56.23. The last value appears anomalous; should it be retained or rejected?

$$Q_{\text{exp}} = \frac{56.23 - 56.08}{56.23 - 55.95} = \frac{0.15}{0.28} = 0.54$$

at the 90 % confidence level, $Q_{crit} = 0.64$,

$0.54 < 0.64$, retain the outlier.