Analytical chemistry

Lecture №12

Random errors in chemical analysis

"Facts are stubborn, but statistics are much more pliable."

—Mark Twain

Statistical Data Treatment and Evaluation

1. Defining a numerical interval around the mean of a set of replicate results within which the population mean can be expected to lie with a certain probability. This interval is called the ***confidence interval***. The confidence interval is related to the standard deviation of the mean.

2. Determining the number of replicate measurements required to ensure that an experimental mean falls within a certain range with a given level of probability.

3. Estimating the probability that (a) an experimental mean and a true value or (b) two experimental means are different, that is, whether the difference is real or simply the result of random error. This test is particularly important for discovering systematic errors in a method and determining whether two samples come from the same source.

4. Determining at a given probability level whether the precision of two sets of measurements differs.

5. Comparing the means of more than two samples to determine whether differences in the means are real or the result of random error. This process is known as analysis of variance.

6. Deciding whether to reject or retain a result that appears to be an outlier in a set of replicate measurements.

Some practical problems with relevant statistical tests

Practical problems/ Relevant tests

1. One result in a replicate set differs rather widely from the rest. Is it a significant result? Examine for gross error. Apply ***Q-test.***

2. Two operators analysing the same sample by the same method obtain results with different spreads. Is there a significant difference in precision between the results? Examine data for unreliable results. Apply ***F-test.***

3. A new method of analysis is being tested by the analysis of a standard sample with an accurately known composition. Is the difference between the experimental value and the accepted value significant?

Examine data for unreliable results. Apply ***t-test****.*

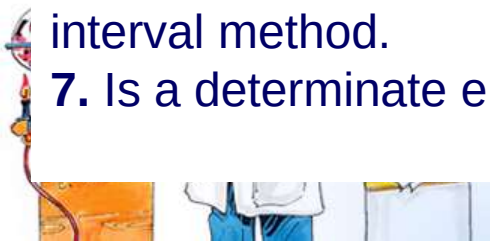
4. Two independent methods of analysis have been used to analyse a sample of unknown composition. Is the difference between the two results significant and thus indicative of an error in one method?

Examine data for unreliable results. Establish that both sets have similar precisions by *F-test. Apply t-test.*

5. With what confidence can the mean of a set of experimental results be quoted as a measure of the true mean? Calculate the confidence interval.

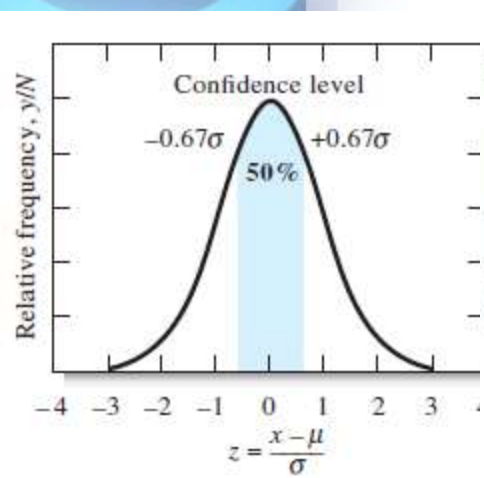
6. If the standard deviation for a method is known, how many results must be obtained to provide a reasonable estimate of the true mean? Use the confidence interval method.

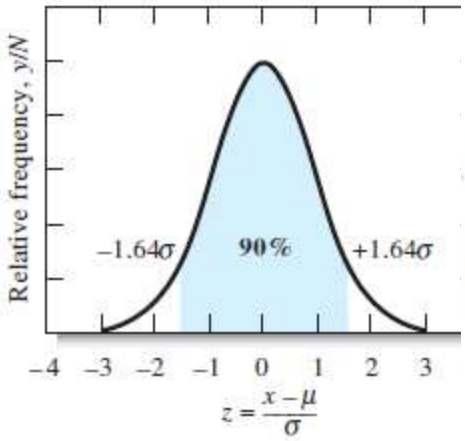
7. Is a determinate error fixed or proportional? Graphical plot of results.

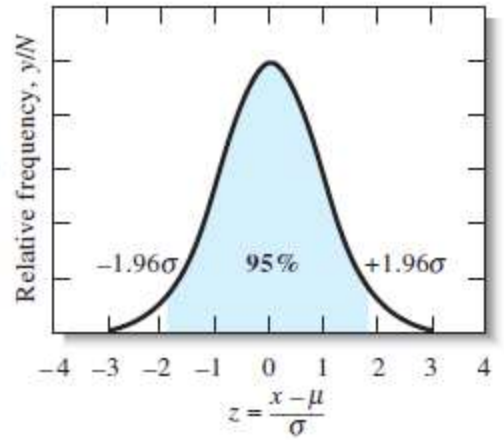


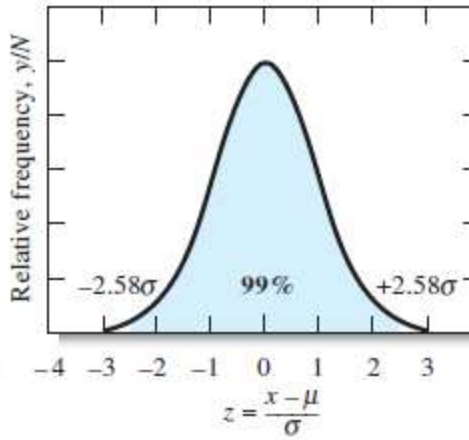
Confidence Intervals

In most quantitative chemical analyses, the true value of the mean cannot be determined because a huge number of measurements (approaching infinity) would be required. With statistics, however, we can establish an interval surrounding the experimentally determined mean *x within which the population mean m is expected* to lie with a certain degree of probability. This interval is known as the ***confidence interval*****.** Sometimes the limits of the interval are called ***confidence limits*****.** For example, we might say that it is 99% probable that the true population mean for a set of potassium measurements lies in the interval 7.25 ± 0.15%. Thus, the probability that the mean lies in the interval from 7.10 to 7.40% is 99%. The size of the confidence interval, which is computed from the sample standard deviation, depends on how well the sample standard deviation *S estimates the population* standard deviation. If *S is a good estimate of s, the confidence interval can be significantly* narrower than if the estimate of S is based on only a few measurement values.









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

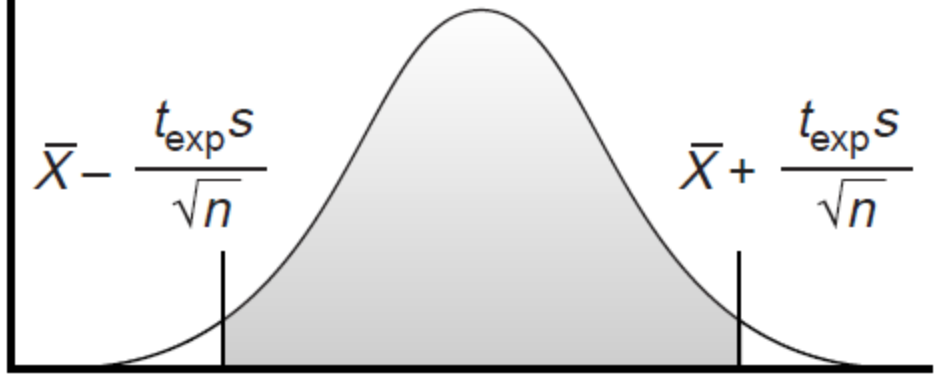
$$\text{Confidence limit} = \bar{x} \pm \frac{ts}{\sqrt{N}}$$

One approach for validating a new analytical method is to analyze a standard sample containing a known amount of analyte, m. The method's accuracy is judged by determining the average amount of analyte in several samples, *X, and using* a significance test to compare it with m.

Frequently, the analyst wishes to decide whether there is a statistical difference between the results obtained using two different procedures, that is, whether they both indeed measure the same thing. The ***t-test*** *is very useful for such comparisons.*

The *t statistic is often called* **Student's t.** Student was the name used by W. S. Gossett when he wrote the classic paper on *t that appeared in 1908.*

Comparing $\bar{x}$ ***to*** μ



Relationship between confidence intervals and results of a significance test. The shaded area under the normal distribution curves shows the apparent confidence intervals for the sample based on t_{exp}.

There are several ways and several different types of situations in which a *t-test* can be used. Consider the following cases:

1. You have taken a certified single sample for which the analytical result is exactly known or is known with a degree of certainty much higher than you expect from your test method. You analyze the same sample by the test method a number of times with an objective to determine if there is no difference between the certified value and the mean value obtained by your method at a specified degree of certainty.

2. The situation is the same as above except that the uncertainty of the certified value or the standard deviation of measurements by a reference method is not negligible. You compare repeated measurements of the same single sample made by the reference method with repeated measurements made by the test method. This is often referred to as *t-test by comparison of the means. The number of measurements in the two* measurement sets need not be the same.

3. Often the intent is to compare a newly developed method with another, and a certified reference standard is not available. Further, even if a reference standard is available, it can only check a method at only one concentration and it will be desirable to check the applicability of the method spanning the entire range of concentrations in which the method is to be used. You take a *number of different samples spanning* the concentration range of interest. All samples are divided in two parts; one set is analyzed by the benchmark method and the other by the test method. The pairs of analytical results thus generated are subjected to the *paired t-test to determine if the* two methods produce results that are statistically different at a specified confidence level.

4. You want to compare two sample populations that are unrelated to each other. Note that unlike the preceding example, in this case, two different sample populations are tested and the numbers in each population do not have to be equal. This type of *t-test can be subdivided in two groups: (a) when the variance or standard deviations of* the two sample sets being compared are statistically the same*, the two sample sets are said to be homoscedastic, (b) when the* variance of the two sample sets are statistically different, the sample populations are *heteroscedastic.*

THE STUDENT *t-TEST* - ARE THERE DIFFERENCES IN THE METHODS?

The *t-test is used to determine if* two sets of measurements are statistically different. If $t_{calc} > t_{table}$, *then the two data sets* are significantly different at the chosen confidence level.

t-Test When a Reference Value Is Known

"Known" value would typically be a certified value from a standard reference material (SRM)

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{N}}$$

$$t_{calc} = (\bar{x} - \mu)\frac{\sqrt{N}}{s}$$

Values of *t* for ν Degrees of Freedom for Various Confidence Levels[a]

	Confidence Level		
ν	90%	95%	99%
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.500
8	1.860	2.306	3.355
9	1.833	2.262	3.250

You are developing a procedure for determining traces of copper in biological materials using a wet digestion followed by measurement by atomic absorption spectrophotometry. In order to test the validity of the method, you obtain a certified reference material and analyze this material. Five replicas are sampled and analyzed, and the mean of the results is found to be 10.8 ppm with a standard deviation of ±0.7 ppm. The listed reference value is 11.7 ppm. Does your method give a statistically correct value relative to the certified value at the 95% confidence level?

$$t_{calc} = (\bar{x} - \mu)\frac{\sqrt{N}}{s}$$

$$= (10.8 - 11.7)\frac{\sqrt{5}}{0.7}$$

$$= 2.9$$

There are five measurements, so there are four degrees of freedom *(N − 1), we see that the tabulated value of t at the 95% confidence level is 2.776.*

Because $t_{calc} > t_{table}$, *one is 95% sure that your procedure is producing a value* that is statistically different from the true value. Your procedure will be considered unacceptable until you can find the source of the discrepancy and fix it.

Before determining the amount of Na_2CO_3 in an unknown sample, a student decides to check her procedure by analyzing a sample known to contain 98.76% w/w Na_2CO_3. Five replicate determinations of the %w/w Na_2CO_3 in the standard were made with the following results

98.71% 98.59% 98.62% 98.44% 98.58%

Is the mean for these five trials significantly different from the accepted value at the 95% confidence level (0.05)?

The mean and standard deviation for the five trials are

$$\overline{X} = 98.59 \qquad s = 0.0973$$

The test statistic is

$$t_{exp} = \frac{|\mu - \overline{X}| \times \sqrt{n}}{s} = \frac{|98.76 - 98.59| \times \sqrt{5}}{0.0973} = 3.91$$

The critical value for *t(0.05,4), i*s found in table, s 2.78. Since t_{exp} is greater than t(0.05,4). At the 95% confidence level the difference between X and m is significant and cannot be explained by indeterminate sources of error. There is evidence, therefore, that the results are affected by a determinate source of error.

Interpreting the confidence interval

- For a 95% CI, there is a 95% probability that the true mean (μ) lies between the range 1.3 ± 0.2 nM, or between 1.1 and 1.5 nM
- For a 50% CI, there is a 50% probability that the true mean lies between the range 1.3 ± 0.05 nM, or between 1.25 and 1.35 nM
- Note that CI will decrease as *n* is increased
- Useful for characterizing data that are regularly obtained; e.g., quality assurance, quality control

Comparison of the Means of Two Methods.

$$\sqrt{\frac{N_1 N_2}{N_1 + N_2}} \Big/ s_p$$

$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{s_p} \sqrt{\frac{N_1 N_2}{N_1 + N_2}}$$

$$s_p = \sqrt{\frac{\sum (x_{a1} - \bar{x}_1)^2 + \sum (x_{a2} - \bar{x}_2)^2}{N_1 + N_2 - 2}}$$

To apply the comparison of means *t-test between two methods, it is necessary* that both methods have statistically the same standard deviation. This must first be verified by using the ***F-test.***

Comparing S^2 to σ^2

When a particular type of sample is analyzed on a regular basis, it may be possible to determine the expected, or true variance, S^2, for the analysis. This often is the case in clinical labs where hundreds of blood samples are analyzed each day. Replicate analyses of any single sample, however, results in a sample variance, ***S^2***. *A statistical* comparison of S^2 to σ^2 *provides useful information about whether the analysis* is in a state of "statistical control."

F-test

Statistical test for comparing two variances to see if their difference is too large to be explained by indeterminate error.

A manufacturer's process for analyzing aspirin tablets has a known variance of 25. A sample of ten aspirin tablets is selected and analyzed for the amount of aspirin, yielding the following results

254 249 252 252 249 249 250 247 251 252

Determine whether there is any evidence that the measurement process is not under statistical control at a = 0.05.

The variance for the sample of ten tablets is 4.3. A two-tailed significance test is used since the measurement process is considered out of statistical control if the sample's variance is either too good or too poor.

$$F_{exp} = \frac{\sigma^2}{s^2} = \frac{25}{4.3} = 5.8$$

The critical value for *F(0.05,∞,9) from the table is 3.33. Since F is greater* than *F(0.05,∞,9), we accept* hypothesis that the analysis is not under statistical control. One explanation for the unreasonably small variance could be that the aspirin tablets were not selected randomly.

Tests of Significance—Is There a Difference?

The ***F-test*** *is used to determine if* two variances are statistically different.

$$F = \frac{s_1^2}{s_2^2}$$

S^2_1 is greater than or equal to S^2_2

$$F_{exp} = \frac{s^2}{\sigma^2} \quad \text{or} \quad F_{exp} = \frac{\sigma^2}{s^2}$$

$(s^2 > \sigma^2)$ $(\sigma^2 > s^2)$

Values of *F* at the 95% Confidence Level

	$\nu_1 = 2$	3	4	5	6	7	8
$\nu_2 = 2$	19.0	19.2	19.2	19.3	19.3	19.4	19.4
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23
10	4.10	3.71	3.48	3.33	3.22	3.14	3.07
15	3.68	3.29	3.06	2.90	2.79	2.71	2.64
20	3.49	3.10	2.87	2.71	2.60	2.51	2.45
30	3.32	2.92	2.69	2.53	2.42	2.33	2.27

If the calculated *F value exceeds a tabulated F* value at the selected confidence level, then there is a significant difference between the variances of the two methods.

If *the variances variances* being compared are significantly different. If *Fcalc < Ftable, they are* statistically the same.

Note that the F-test can be used to simply test whether or not two sets of data have statistically similar precisions or not.

Can use to answer a question such as: *Do method one and method two provide similar precisions for the analysis of the same analyte?*

You are developing a new colorimetric procedure for determining the glucose content of blood serum. You have chosen the standard Folin-Wu procedure with which to compare your results. From the following two sets of replicate analyses on the same sample, determine whether the variance of your method differs significantly from that of the standard method.

Your Method (mg/dL)	*Folin-Wu Method (mg/dL)*
127	130
125	128
123	131
130	129
131	127
126	125
129	
mean ($\bar{x}_1$) 127	mean ($\bar{x}_2$) 128

$$s_1^2 = \frac{\sum (x_{i1} - \bar{x}_1)^2}{N_1 - 1} = \frac{50}{7 - 1} = 8.3$$

$$s_2^2 = \frac{\sum (x_{i2} - \bar{x}_2)^2}{N_2 - 1} = \frac{24}{6 - 1} = 4.8$$

$$F = \frac{8.3}{4.8} = 1.7_3$$

In short, statistically, your method does as well as the established procedure.

A new gravimetric method is developed for iron(III) in which the iron is precipitated in crystalline form with an organoboron "cage" compound. The accuracy of the method is checked by determining the iron in an ore sample and comparing with the results using the standard precipitation with ammonia and weighing the Fe_2O_3 formed after ignition of the $Fe(OH)_3$ precipitated. The results, reported as % Fe for each analysis, were as follows

x_{1i}	$x_{1i} - \bar{x}_1$	$(x_{1i} - \bar{x}_1)^2$	x_{2i}	$x_{2i} - \bar{x}_2$	$(x_{2i} - \bar{x}_2)^2$
20.10	0.45	0.202	18.89	0.35	0.122
20.50	0.85	0.722	19.20	0.04	0.002
18.65	1.00	1.000	19.00	0.24	0.058
19.25	0.40	0.160	19.70	0.46	0.212
19.40	0.25	0.062	19.40	0.16	0.026
19.99	0.34	0.116		$\sum (x_{2i} - \bar{x}_2)^2 = 0.420$	
	$\sum (x_{1i} - \bar{x}_1)^2 = 2.262$				

Test Method	*Reference Method*
20.10%	18.89%
20.50	19.20
18.65	19.00
19.25	19.70
19.40	19.40
19.99	
$\bar{x}_1 = 19.65\%$	$\bar{x}_2 = 19.24\%$

Is there a significant difference between the two methods?

$$F = \frac{s_1^2}{s_2^2} = \frac{2.262/5}{0.420/4} = 4.31$$

This is less than the tabulated value (6.26), so the two methods have comparable standard deviations and the *t*-test can be applied

$$s_p = \sqrt{\frac{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2}{N_1 + N_2 - 2}}$$

$$= \sqrt{\frac{2.262 + 0.420}{6 + 5 - 2}} = 0.546$$

$$\pm t = \frac{19.65 - 19.24}{0.546} \sqrt{\frac{(6)(5)}{6 + 5}} = 1.2_3$$

The tabulated *t* for nine degrees of freedom ($N_1 + N_2 - 2$) at the 95% confidence level is 2.262, so there is no statistical difference in the results by the two methods.

The %w/w Na_2CO_3 in soda ash can be determined by an acid–base titration. The results obtained by two analysts are shown here. Determine whether the difference in their mean values is significant at α = 0.05.

Analyst A	Analyst B
86.82	81.01
87.04	86.15
86.93	81.73
87.01	83.19
86.20	80.27
87.00	83.94

We begin by summarizing the mean and standard deviation for the data reported by each analyst. These values are

$$\overline{X}_A = 86.83\%$$
$$s_A = 0.32$$
$$\overline{X}_B = 82.71\%$$
$$s_B = 2.16$$

The test statistic is

$$F_{exp} = \frac{s_B^2}{s_A^2} = \frac{(2.16)^2}{(0.32)^2} = 45.6$$

Since F_{exp} *is larger than the critical value of 7.15 for F(0.05, 5, 5),* the variances are significantly different is accepted. As a result, a pooled standard deviation cannot be calculated.

the test statistic, t_{exp}, is calculated

$$t_{exp} = \frac{\left|\overline{X}_A - \overline{X}_B\right|}{\sqrt{(s_A^2/n_A) + (s_B^2/n_B)}} = \frac{\left|86.83 - 82.71\right|}{\sqrt{[(0.32)^2/6] + [(2.16)^2/6]}} = 4.62$$

The critical value for *t(0.05, 5) is 2.57. Since the calculated value of* t_{exp} *is* greater than *t(0.05, 5)* the mean values for %w/w Na_2CO_3 reported by the two analysts are significantly different at the chosen significance level.

Rejection of a Result: The Q Test

Outliers

On occasion, a data set appears to be skewed by the presence of one or more data points that are not consistent with the remaining data points. Such values are called outliers. The most commonly used significance test for identifying outliers is Dixon's *Q-test.*

Dixon's *Q-test*

Statistical test for deciding if an outlier can be removed from a set of data.

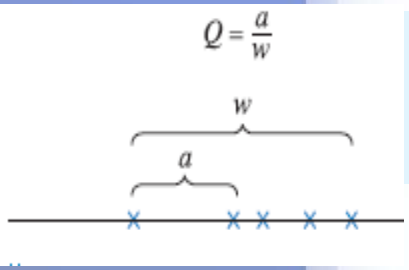
The *Q-test compares the difference between the suspected outlier and its nearest* numerical neighbor to the range of the entire data set. Data are ranked from smallest to largest so that the suspected outlier is either the first or the last data point. The value of Q_{exp} *is compared with a critical value, Q(a, n), at* a significance level of α.

Outliers

Rejection of a Result: The Q Test

In practice, it is a good idea to make more measurements to be sure, especially if the decision is a close one.

The *Q test is used to determine if* an "outlier" is due to a determinate error. If it is not, then it falls within the expected random error and should be retained.



$$Q = \frac{|x_q - x_n|}{w}$$

In this test, the absolute value of the difference between the questionable result x_q *and its nearest neighbor* x_n *is divided by the spread w of the entire set to give the quantity Q*

This ratio is then compared with critical values Q_{crit} found in Table. If Q is greater than Q_{crit}, the questionable result can be rejected with the indicated degree of confidence.

Critical Values for the Rejection Quotient, Q^*

Number of Observations	Q_{crit} (Reject if $Q > Q_{crit}$) 90% Confidence	95% Confidence	99% Confidence
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

The following set of chloride determinations on separate aliquots of a pooled serum were reported: 103, 106, 107, and 114 meq/L. One value appears suspect. Determine if it can be ascribed to accidental error, at the 95% confidence level.

The suspect result is 114 meq/L. It differs from its nearest neighbor, 107 meq/L, by 7 meq/L. The range is 114 to 103, or 11 meq/L. *Q is therefore 7/11 = 0.64. The* tabulated value for four observations is 0.829. Since the calculated *Q is less than the* tabulated *Q, the suspected number may be ascribed to random error and should not be* rejected.

Statistics: Set of mathematical tools used to describe and make judgments about data