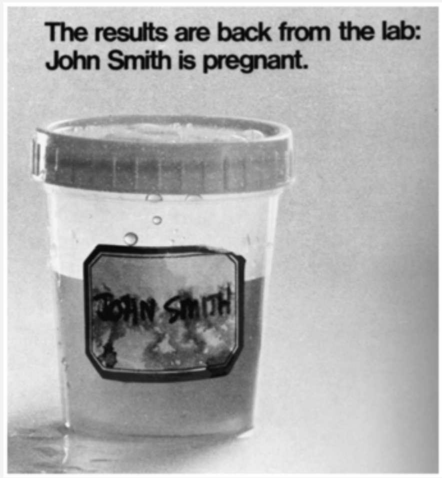
The importance of reliable data!



Exploring Chemical Analysis, 3th Ed., D.C. Harris, WH Freeman, NY, 2005

Statistics

- Data quality heart of trustworthy data
- Rule 702 of federal Rules of Evidence
  - Calls for quality assurance
  - Need statistics

What is the molecular mass of $CaCO_3$?

1. 68.09 g/mol
2. 100.1 g/mol
3. 100.08 g/mol
4. 100.086 g/mol
5. 100.0862 g/mol
6. 100.09 g/mol
7. 204.26 g/mol
8. No idea / none of the above

M_{Ca} = 40.078 g/mol
M_C = 12.01 g/mol
M_O = 15.9994 g/mol

What is 4.2601 rounded to two significant figures?

1. 4
2. 4.2
3. 4.3
4. 4.26
5. 4.260
6. No idea / none of the above

Sig figs: special case

- When rounding to the digit to the left of a *5, *50 or *5000, round to nearest even digit
- E.g. 43.550000 rounded to 3 s.f. is 43.6
  – 43.550001 rounded to 3 s.f. is 43.6
- E.g. 42.500 rounded to 2 s.f. is 42
  – 42.5501 rounded to 3 s.f. is 43

What is the number 4.25000 rounded to two significant figures?

1. 4
2. 4.2
3. 4.3
4. 4.25
5. 4.250
6. No idea / none of the above

Sig figs

0.1920
Four sig figs; the zero furthest to the right is known

1920
Three or four sig figs; do not know if last zero is known

1920.0
Five sig figs; last zero is known

How many significant figures are in the number 4.200×10^{-6}?

1. One
2. Two
3. Three
4. Four
5. Six
6. No idea / none of the above

What is the number 42.5001 rounded to two significant figures?

1. 40
2. 42
3. 43
4. 42.5
5. 42.50
6. No idea / none of the above

724500 rounded to three significant figures is…?

1. 725
2. 724500
3. 720000
4. 725000
5. 724000
6. No idea / none of the above

Error in chemical analyses

- Mean (average, arithmetic mean)
  - Serves as the central "best" value for a set of replicate measurements
  - For a set of measurements, x_i, where i = 1,2,3…

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \qquad \bar{x} = \text{mean} \qquad N = \text{number of measurements}$$

- Median

Is the middle result when the data are arranged in order of size

- Ideally, Mean = Median, but this is rarely true

Definitions

- Replicates:
  - Samples from the same system that are carried through an analysis in an identical manner
  - Forensic samples (e.g. blood) often taken in duplicates (think Olympics)
- Error:
  - Deviation of a result from the accepted true value
- Precision:
  - Deviation of a result from the average of a set of results:

$$d_i = \left| x_i - \bar{x} \right| \qquad d_i = \text{deviation of result i} \quad x_i = \text{result i} \quad \bar{x} = \text{mean}$$

- Accuracy:
  - Deviation of a result from the accepted true value

$$E_i = x_i - x_t \qquad E_i = \text{error of result i} \quad x_i = \text{result i} \quad x_t = \text{"true" value}$$

- Error can be determined in a single measurement; precision cannot

Definitions

- Absolute error: has units & sign (±)

$$E_i = x_i - x_t \qquad \begin{aligned} E_i &= \text{error of result i} \\ x_i &= \text{result i} \\ x_t &= \text{"true" value} \end{aligned}$$

- Relative error: error expressed as a percent & has a sign (±)

$$E_r = \frac{x_i - x_t}{x_t} \times 100\ \% = \text{percent error}$$

- Relative error can be expressed in ppt or ppm:

$$E_r = \frac{x_i - x_t}{x_t} \times 10^6 = \text{ppm error}$$

Forensic Chemistry 13

If an analyst determined a pill to contain 0.15 g drug X, and the accepted value was 0.25 g, what is the absolute error?

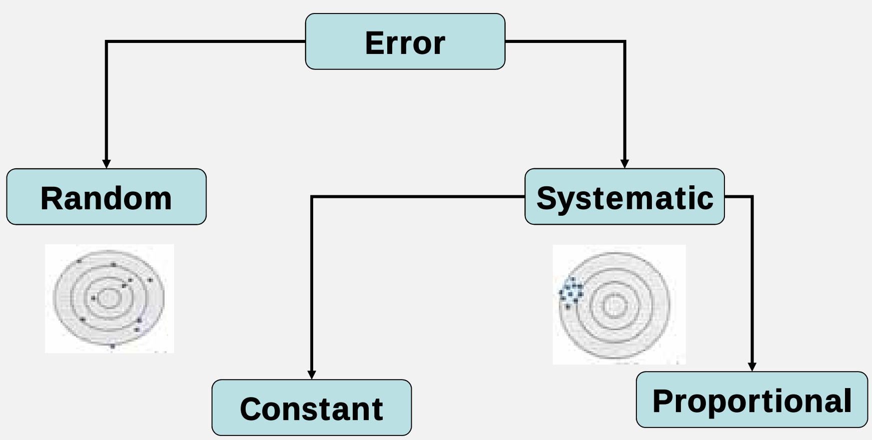
1. 0.4 g
2. -0.1 g
3. 0.1 g
4. 0.1%
5. -0.1%
6. None of the above

Forensic Chemistry 14

If an analyst determined a pill to contain 0.15 g drug X, and the accepted value was 0.25 g, what was the relative error?

1. 40 g
2. 40%
3. -40%
4. 0.4%
5. -0.4%
6. None of the above

Types of errors



E.g. balance reads 0.03 g too heavy
1.00 g measured = 0.97 g actual (+3% error)
100.00 g measured = 99.97 g actual (0.03 % error)

E.g. balance always -0.8% off
1.00 g measured = 1.008 g actual
100.00 g measured = 100.8 g actual

Systematic error

- Has definite value, assignable cause, and affects all replicates similarly

- Origins of systematic errors:
  1) Instrumental errors
     faulty calibration
     use outside appropriate range
     instabilities in components
     etc.

  2) Method error
     unrepresentative sampling
     incomplete chemistry
     impure/unstable reagents

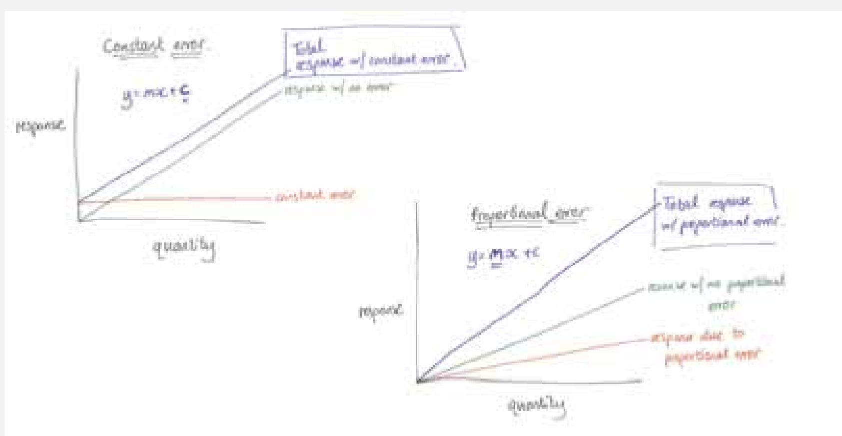
Systematic error

- Has definite value, assignable cause, and affects all replicates similarly

- Origins of systematic errors:
  3) Operator errors
     Misreading dials, scales, menisci
     Poor control of conditions (temp, pH etc.)
     Poor technique (e.g. blowout pipeting)
     etc.

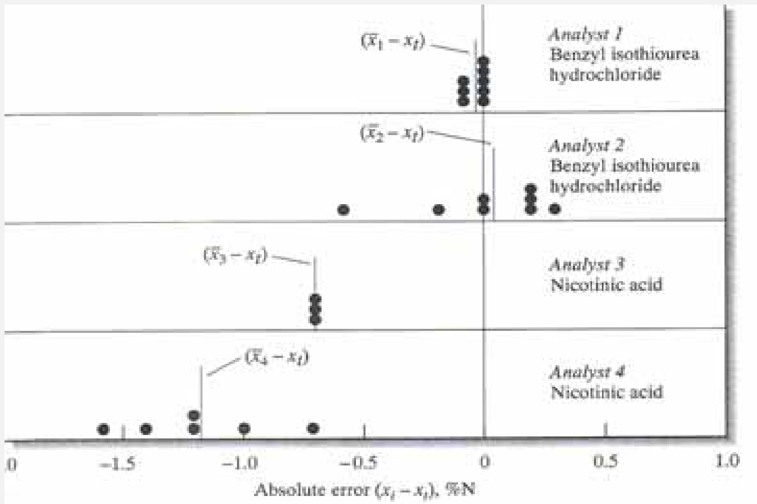
Constant v. proportional errors

Constant v. proportional errors



Accuracy and Precision

Analysis of nitrogen in two compounds by two different methods



Forensic Chemistry 21

2. Propagation of error

The method of using derivatives is used for ALL calculations regarding the propagation of error. Sometimes the solutions are simple; sometimes complicated.

E.g. 1.

$$y = a + b - c$$

$$s_y = \sqrt{\left(\frac{\partial y}{\partial a}\right)^2 s_a^2 + \left(\frac{\partial y}{\partial b}\right)^2 s_b^2 + \left(\frac{\partial y}{\partial c}\right)^2 s_c^2}$$

To calculate the error of y, find the derivatives of y w.r.t. each of the variables and plug them into the master equation.

Forensic Chemistry 22

E.g. 2.

$$y = \frac{a.b}{c}$$

$$s_y = \sqrt{\left(\frac{\partial y}{\partial a}\right)^2 s_a^2 + \left(\frac{\partial y}{\partial b}\right)^2 s_b^2 + \left(\frac{\partial y}{\partial c}\right)^2 s_c^2}$$

Forensic Chemistry

23

Significant Figures

Based on what we know about standard deviations, we can no longer determine sig. figs. the way we used to; i.e. we have to use propagation of error.

$\frac{24 \times 4.52}{100.0} = 1.0848$. Using sig. figs. we would round to 1.1

But, assuming ± 1 in the last digits (as we should),

$$e_{answer} = 1.0848\sqrt{\left(\frac{1}{24}\right)^2 + \left(\frac{0.01}{4.52}\right)^2 + \left(\frac{0.1}{100}\right)^2} = 0.04508 = 0.05$$

Therefore answer $= 1.08 \pm 0.05$ **Three sig. figs.**

$\frac{24 \times 4.02}{100.0} = 0.9648$. Using sig. figs. we would round to 0.96

As before,

$$e_{answer} = 0.9648\sqrt{\left(\frac{1}{24}\right)^2 + \left(\frac{0.01}{4.02}\right)^2 + \left(\frac{0.1}{100}\right)^2} = 0.04028$$

Therefore, answer $= 0.96 \pm 0.04$ **Two sig. figs.**

Forensic Chemistry

24

$$\frac{(12.18\pm0.08)(23.04\pm0.07)}{3.247\pm0.006} = ?$$

$$\%e_y = \sqrt{\%e_a^2 + \%e_b^2 + \%e_c^2}$$

1. 86.4266091777±0.122
2. 86.43±0.007
3. 86.4±0.6
4. 86.4±0.2
5. 86±1
6. None of the above

Samples & populations

μ = mean value of the population (system)

$\bar{x}$ = estimate of μ, because we only take small samples from the system

As $n_{\text{sample}} \rightarrow N_{\text{population}}$, $\bar{x} \rightarrow \mu$

Samples & populations

Precision of a population of data is described by population standard deviation, σ :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \mu\right)}{N}}$$

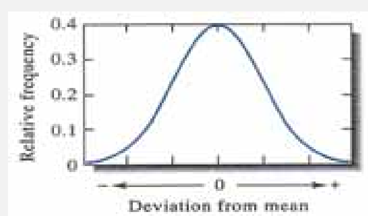
Precision of a sample of data is described by sample standard deviation, *s*:

$$s = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{N}\left(d_i\right)^2}{n-1}}$$

As $\mathrm{n}_{\mathrm{sample}} \rightarrow \mathrm{N}_{\mathrm{population}}$, $s \rightarrow \sigma$ just as $\bar{x} \rightarrow \mu$

Properties of Gaussian curves

- Plot of relative frequency vs. deviation from mean
- Mean occurs at the central point
  - i.e. no deviation from mean
  - occurs at highest relative frequency
- Are symmetrical
- Exponential decrease in frequency as d increases



Properties of Gaussian curves

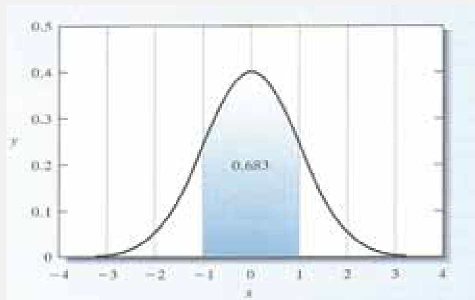
- Area under the Gaussian curve

$$area = \int_{-x}^{x} \frac{e^{-(x-\mu)^2 / 2\sigma^2}}{\sigma\sqrt{2\pi}} dx$$

If we want the area between $+\sigma$ and $-\sigma$

$$area = \int_{-\sigma}^{\sigma} \frac{e^{-(x-\mu)^2 / 2\sigma^2}}{\sigma\sqrt{2\pi}} dx$$

$$area = 0.683$$



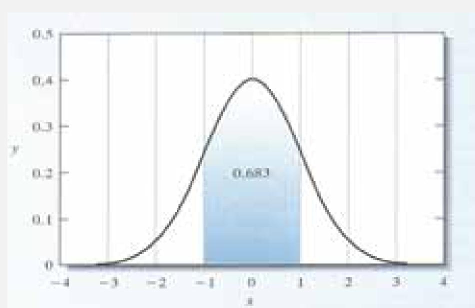
Forensic Chemistry 29

Properties of Gaussian curves: probability

- Area under the Gaussian curve
- There is a 68.3% probability that a measurement will be within ±1σ of the mean.

$$area = \int_{-\sigma}^{\sigma} \frac{e^{-(x-\mu)^2 / 2\sigma^2}}{\sigma\sqrt{2\pi}} dx$$

$$area = 0.683$$



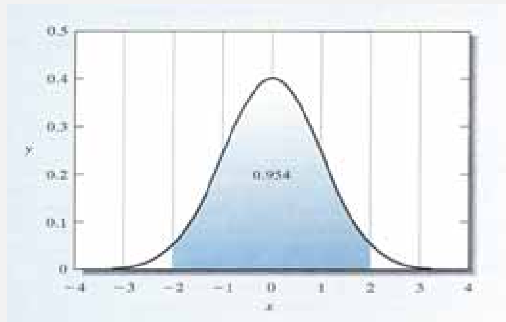
Forensic Chemistry

30

Properties of Gaussian curves: probability

- There is a 95.4% probability that a measurement will be within $\pm 2\sigma$ of the mean.

$$area = \int_{-2\sigma}^{2\sigma} \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}dx$$
$$area = 0.954$$



Forensic Chemistry

31

Properties of Gaussian curves: probability

- What is the probability that a measurement will be greater than the mean?

$$area = \int_{0}^{+\infty} \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}dx$$

Forensic Chemistry

32

Confidence intervals

An experimental determination of Pb in a bullet using ICP-OES was repeated10 times. The mean and standard deviation were 0.022 %w/w and 0.003 %w/w, respectively. What is the 99% confidence interval of the mean?

Confidence limits

An experimental determination of Pb in a bullet using ICP-OES was repeated 10 times. The mean and standard deviation were 0.022 %w/w and 0.003 %w/w, respectively. What are the 99% confidence limits?

Confidence limits

In 5 replicate analyses, a baggie was found to contain an average of 12.5 % w/w crack cocaine with a s.d. of 0.4 % w/w.
What is the 95 % confidence interval for the true cocaine content?

1. 12.5 ± 0.1%
2. 12.5 ± 0.2%
3. 12.5 ± 0.3%
4. 12.5 ± 0.4%
5. 12.5 ± 0.5%
6. 12.5 ± 0.6%
7. None of the above

Comparing an experimental mean with the true value

A procedure for the analysis of Pb in kerosene was tested on a sample that is known to contain 0.123 ppm Pb (x_t). The results were 0.112, 0.118, 0.115 and 0.119 ppm Pb.

To what degree can we be confident that our method does not contain bias?

Comparing an experimental mean with the true value

A new technique is being developed for the analysis of lead isotope ratios in gun shot residues. To test the accuracy of the method, an analyst analyzed a NIST sample that is known to have a $^{208}Pb/^{206}Pb$ ratio of 2.1764. The results of the analysis were 2.1776, 2.1779, 2.1773, 2.1791, 2.1774.

Do you have any evidence (at 99% CL) to think that the method might be inaccurate?

Comparing two experimental means

- This can be used to test
  - 1) whether two samples are significantly different
  - 2) whether two methods are significantly different
- We still use the T-test (and the t-tables)
- Different samples and different methods might have different standard deviations
  - This has to be accounted for by pooling the deviations of the means

Comparing two experimental means

$$\bar{x}_a - \bar{x}_b = \pm t s_{pooled} \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

$$\text{or, } t = \frac{\bar{x}_a - \bar{x}_b}{s_{pooled}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}}$$

Comparing two experimental means

$$\bar{x}_a - \bar{x}_b = \pm t s_{pooled} \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

$$\text{or, } t = \frac{\bar{x}_a - \bar{x}_b}{s_{pooled}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}}$$

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1) + \ldots}{n_1 + n_2 \ldots - \text{number data sets}}}$$

If $t > t_{crit}$ then reject the hypothesis that the means are the same.

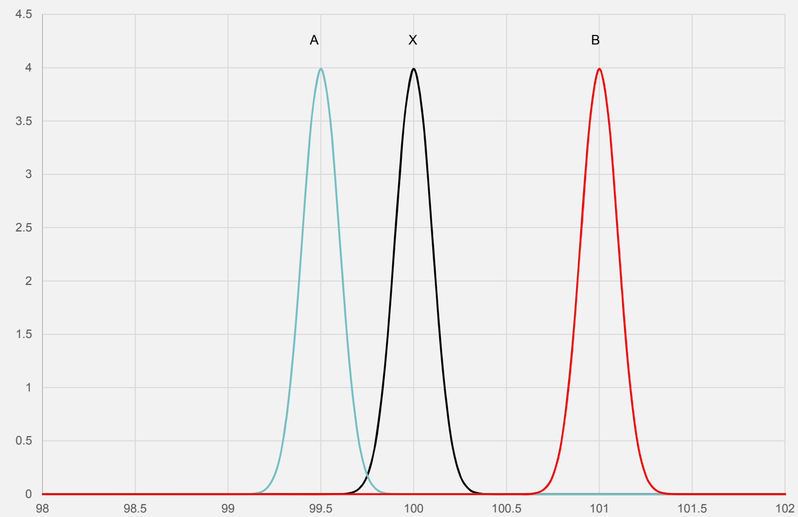
Comparing two experimental means

Glass in a broken window from a crime scene is compared with glass shards found on a suspect. Five analyses of the glass from the crime scene gave chromium content of 0.013±0.003% w/w and Zinc content of 2.74±0.02 ppm. Five analyses of the glass from the suspect gave values of 0.009±0.002% and 2.88±0.04 ppm, respectively.

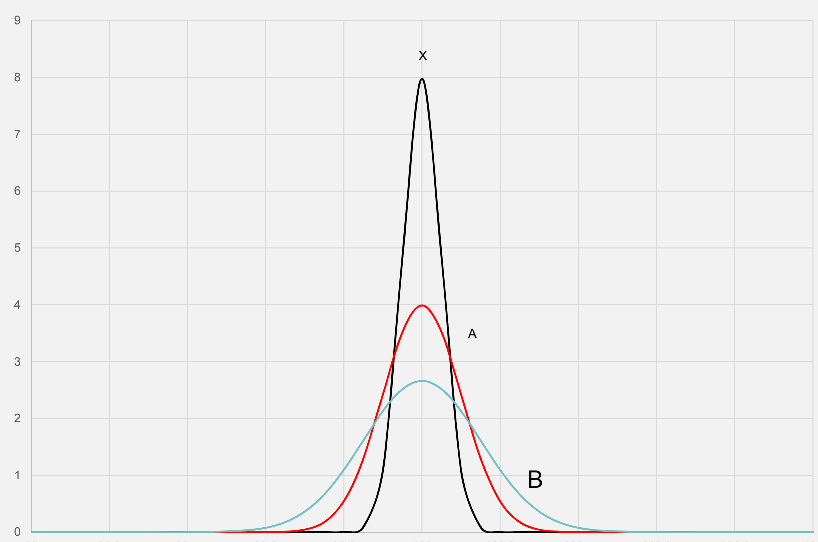
1) Does either impurity suggest a significant difference between the pieces of glass at the 95% CL?

2) Does this prove innocents or guilt?

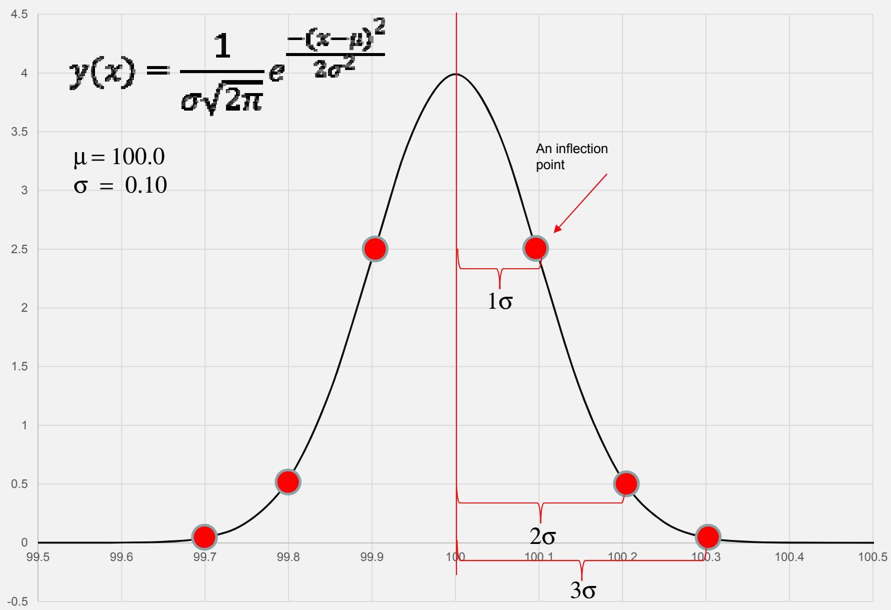
Description	Equation	Variable
Arithmetic Mean	$\bar{x} = \sum_i \frac{x_i}{n}$	$x_i = the\ ith\ value$ $n = total\ number\ of\ values$
Degrees of Freedom	$v = (n-1)$	
Variance	$V(x) = \frac{\sum_i (x_i - \bar{x})^2}{v}$	
Standard Deviation	$s = \sqrt{V(x)} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{v}}$	
Normal Distribution	$y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$\mu = population\ mean$ $\sigma = population\ standard\ deviation$
Confidence Limits	$\mu = x \pm t(\frac{s}{\sqrt{n}})$	$t = confidence\ interval\ value$



Forensic Chemistry



Forensic Chemistry



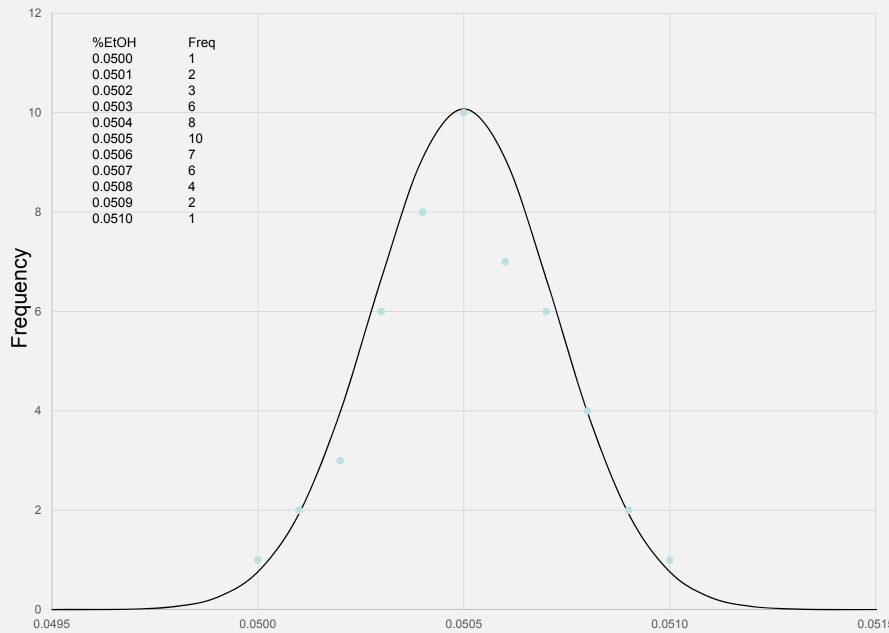
Forensic Chemistry

t values for confidence intervals

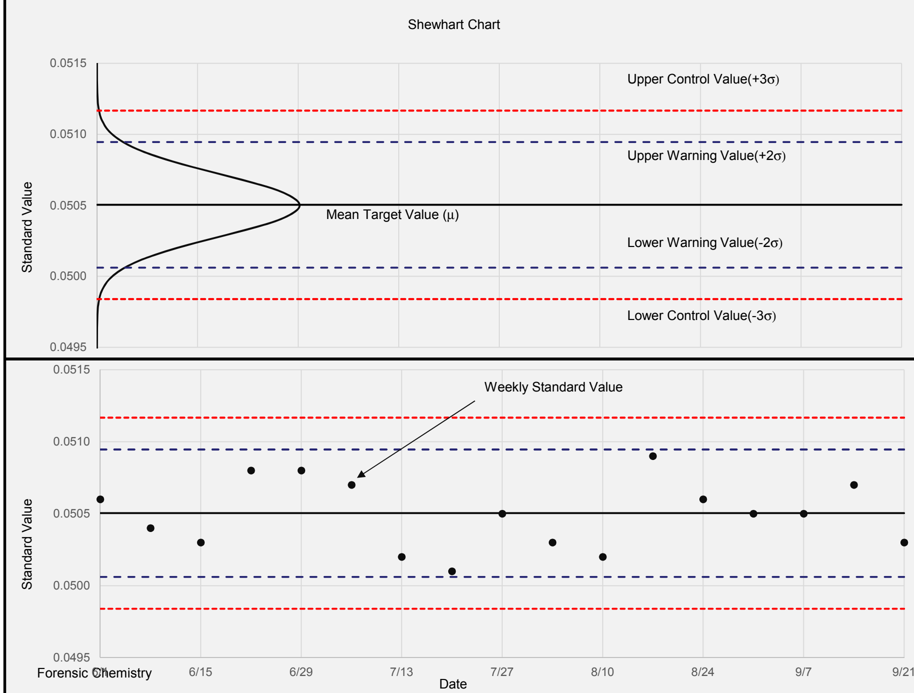
V	95%	99%
1	12.71	63.66
2	4.3	9.92
3	3.18	5.84
4	2.78	4.6
5	2.57	4.03
10	2.23	3.17
20	2.09	2.85
30	2.04	2.75
50	2.01	2.68
100	1.98	2.63

$$\mu = x \pm t\left(\frac{s}{\sqrt{n}}\right)$$

Forensic Chemistry



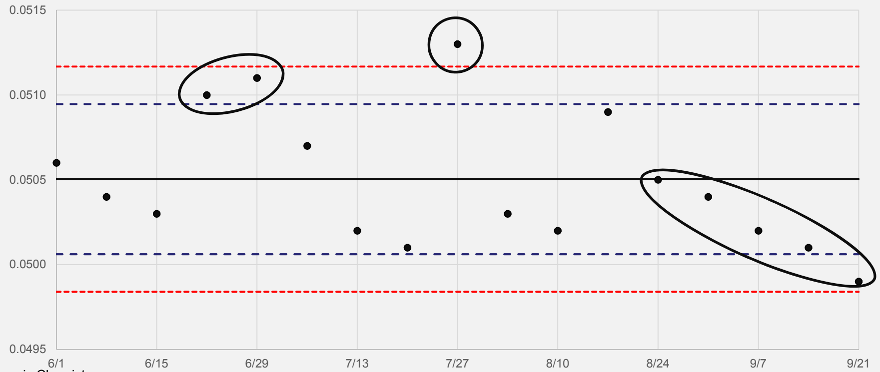
Forensic Chemistry



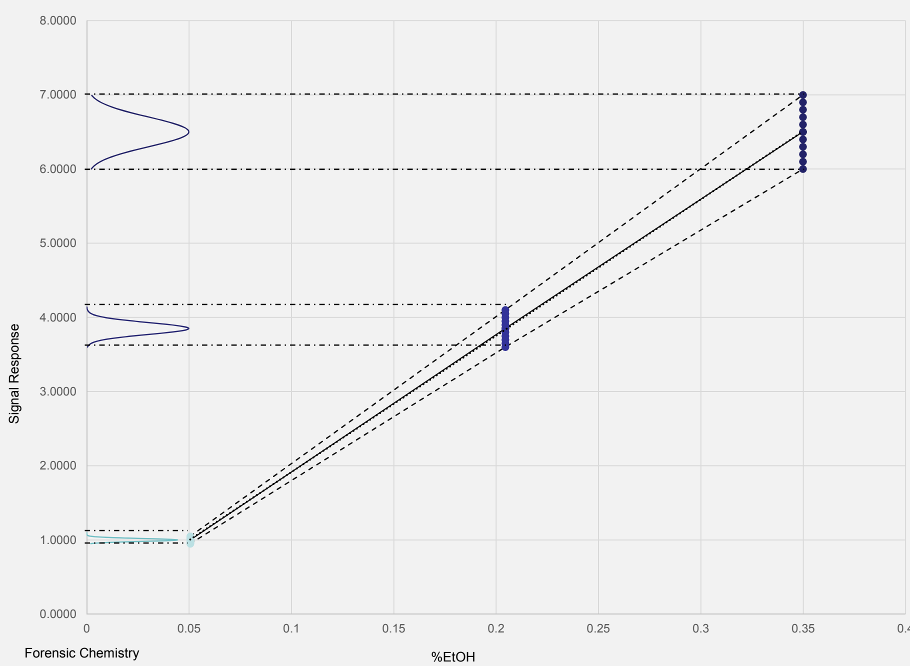
Forensic Chemistry

Run rules *are rules that are used to indicate out-of-statistical control or determinate trend situations.*

- a point lying beyond the upper and lower control limits
- 2 consecutive points lying beyond the upper or lower warning values (0.06% chance)
- 7 or more consecutive points lying on one side of the mean (0.8% chance of occurring)
- 6 consecutive points going in the same direction



Forensic Chemistry



Forensic Chemistry