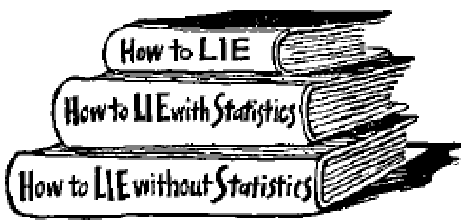
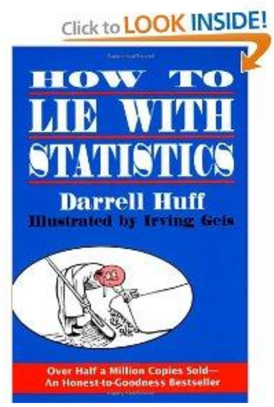
Statistics - Lying without sinning?

- "**Lies, damned lies, and statistics**"





1954

Statistics - Lying without sinning?

In North Dakota, 54 Million Beer Bottles by the side of the Road
April 01 2002

South Dakota's Pierre Capital Journal reports (Mar. 1) that "an average of **650** beer cans and bottles are tossed **per mile** of road **annually**." The statistic is attributed to Dennis W. Brezina, an activist against drunk-driving.

But how did he come up with his data? According to the Journal, Brezina traveled "highways across the nation to determine whether the problem he perceived was widespread. He made two trips to South Dakota, one in 1998 and another in 2000." He counted "cans and bottles in ditches in May of both years" and claimed to have found an average of "**one** beer can or bottle every **16 feet** when walking randomly selected stretches of ditch."

But the math appears a little blurry. The web site of the South Dakota Department of Transportation claims that the state "has **83,472** miles of highways, roads and streets." Assuming Brezina's estimate is correct, South Dakotans appear to be world-class litterbugs, tossing aside approximately **54,256,800** bottles or cans every year. According to the Census Bureau there are **754,844** people in South Dakota. So, according to Brezina, the average resident throws at least **71** beer bottles or cans on the side of the road every **year**.

For more
Check out
www.STATS.org

Statistics for Quantitative Analysis

- Statistics: Set of mathematical tools used to describe and make judgments about data
- Type of statistics we will talk about in this class has important assumption associated with it:

Experimental variation in the population from which samples are drawn has a normal (Gaussian, bell-shaped) distribution.

Normal distribution

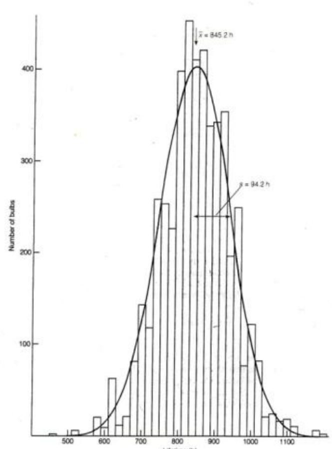


Figure 4-1 Bar graph and Gaussian curve describing the lifetime of a hypothetical set of electric light bulbs. The smooth curve has the same mean, standard deviation, and area as the bar graph. Any finite set of data, however, will differ from the bell-shaped curve.

- Infinite members of group: *population*
- Characterize population by taking *samples*
- The larger the number of samples, the closer the distribution becomes to normal
- Equation of normal distribution:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Normal distribution

- Estimate of mean value of population = μ
- Estimate of mean value of samples = $\bar{x}$

Mean = $\bar{x} \;=\; \dfrac{\sum_i x_i}{n}$

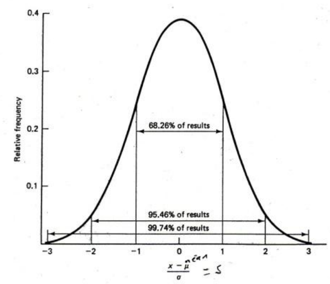


Figure 2.2 Normal distribution curve; relative frequencies of deviations from the mean for a normally distributed infinite population; deviations $(x - \mu)$ are in units of σ.

Normal distribution

- Degree of scatter (measure of central tendency) of population is quantified by calculating the *standard deviation*
- Std. dev. of population = σ
- Std. dev. of sample = s

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$$

- Characterize sample by calculating $\bar{x} \pm s$

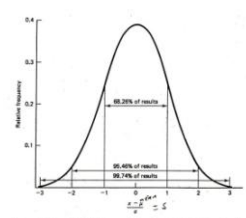
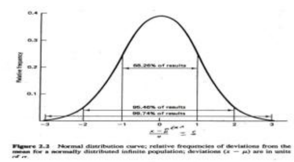


Figure 2.2 Normal distribution curve; relative frequencies of deviations from the mean for a normally distributed infinite population; deviations $(x - \mu)$ are in units of σ.

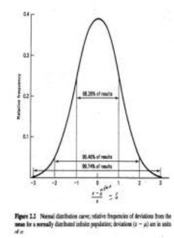
Standard deviation and the normal distribution

- Standard deviation defines the shape of the normal distribution (particularly width)



- Larger std. dev. more scatter about the mean, worse precision.

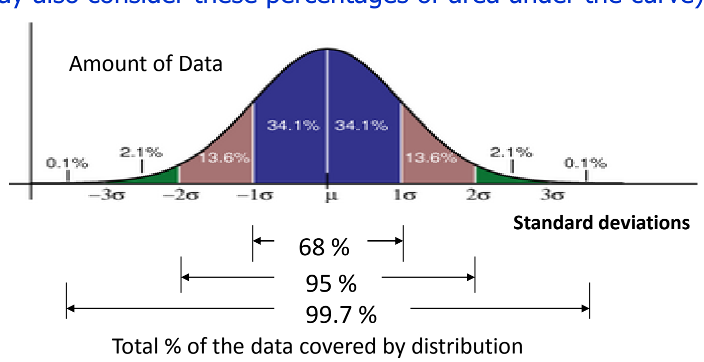
- Smaller std. dev. means less scatter about the mean, better precision.



Standard deviation and the normal distribution

- There is a well-defined relationship between the std. dev. of a population and the normal distribution of the population.

- (May also consider these percentages of area under the curve)



Example of mean and standard deviation calculation

Consider Cu data: 5.23, 5.79, 6.21, 5.88, 6.02 nM

$\bar{x}$ = 5.826 nM → 5.8_2 nM

s = 0.368 nM → 0.3_6 nM

Answer: $5.8_2 \pm 0.3_6$ nM **or 5.8 ± 0.4 nM**

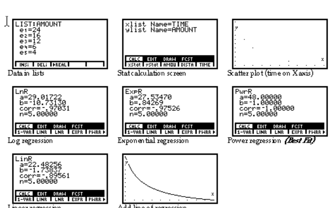
Learn how to use the statistical functions on your calculator. Do this example by longhand calculation once, and also by calculator to verify that you'll get exactly the same answer. Then use your calculator for all future calculations.

Learn to use your calculator's statistical functions to calculate

mean and standard deviation. You'll save yourself a lot of work.

http://www.willamette.edu/~mjaneba/help/TI-85-stats.htm

http://www2.ohlone.edu/people2/joconnell/ti/



Relative standard deviation (rsd) or coefficient of variation (CV)

rsd or CV = $\left(\frac{s}{\bar{x}}\right)100$

From previous example,

rsd = $(0.3_6$ nM/5.8_2 nM) 100 = **$6._1$% or 6%**

Standard error

- Tells us that standard deviation of set of samples should decrease if we take more measurements
- Standard error = $s_{\bar{x}} = \frac{s}{\sqrt{n}}$
- Take twice as many measurements, s decreases by $\sqrt{2} \approx 1.4$
- Take 4x as many measurements, s decreases by $\sqrt{4} = 2$
- There are several quantitative ways to determine the sample size required to achieve a desired precision for various statistical applications. Can consult statistics textbooks for further information; e.g. J.H. Zar, *Biostatistical Analysis*

Variance

Used in many other statistical calculations and tests

Variance = s^2

From previous example, s = 0.3_6

$s^2 = (0.3_6)^2 = 0.129$ (not rounded because it is usually used in further calculations)

Average deviation

- Another way to express degree of scatter or uncertainty in data. Not as statistically meaningful as standard deviation, but useful for small samples.

$$\bar{d} = \frac{\sum_i (|x_i - \bar{x}|)}{n}$$

Using previous data:

$$\bar{d} = \frac{|5.23 - 5.8_2| + |5.79 - 5.8_2| + |6.21 - 5.8_2| + |5.88 - 5.8_2| + |6.02 - 5.8_2|}{5}$$

$$\bar{d} = 0.25 \rightarrow 0.2_5 \; or \; 0.2 \; nM$$

$$Answer: \quad 5.8_2 \pm 0.2_5 \; nM \; or \; 5.8 \pm 0.2 \; nM$$

Relative average deviation (RAD)

$$RAD = \left(\frac{\bar{d}}{\bar{x}}\right)100 \quad (as\ percentage)$$

$$RAD = \left(\frac{\bar{d}}{\bar{x}}\right)1000 \quad (as\ parts\ per\ thousand,\ ppt)$$

Using previous data,

RAD = $(0.\,2_5/5.8_2)$ 100 = $4._2$ **or 4%**

RAD = $(0.\,2_5/5.8_2)$ 1000 = 42 ppt
→ **$4._2 \times 10^1$ or 4×10^1 ppt ($^0/_{00}$)**

Some useful statistical tests

- To characterize or make judgments about data
- Tests that use the *Student's t distribution*
  - Confidence intervals
  - Comparing a measured result with a "known" value
  - Comparing replicate measurements (comparison of means of two sets of data)

Table 4-2

Values of Student's t

Degrees of freedom	Confidence level (%)						
	50	90	95	98	99	99.5	99.9
1	1.000	6.314	12.706	31.821	63.657	127.32	636.619
2	0.816	2.920	4.303	6.965	9.925	14.089	31.598
3	0.765	2.353	3.182	4.541	5.841	7.453	12.924
4	0.741	2.132	2.776	3.747	4.604	5.598	8.610
5	0.727	2.015	2.571	3.365	4.032	4.773	6.869
6	0.718	1.943	2.447	3.143	3.707	4.317	5.959
7	0.711	1.895	2.365	2.998	3.500	4.029	5.408
8	0.706	1.860	2.306	2.896	3.355	3.832	5.041
9	0.703	1.833	2.262	2.821	3.250	3.690	4.781
10	0.700	1.812	2.228	2.764	3.169	3.581	4.587
15	0.691	1.753	2.131	2.602	2.947	3.252	4.073
20	0.687	1.725	2.086	2.528	2.845	3.153	3.850
25	0.684	1.708	2.068	2.485	2.787	3.078	3.725
30	0.683	1.697	2.042	2.457	2.750	3.030	3.646
40	0.681	1.684	2.021	2.423	2.704	2.971	3.551
60	0.679	1.671	2.000	2.390	2.660	2.915	3.460
120	0.677	1.658	1.980	2.358	2.617	2.860	3.373
∞	0.674	1.645	1.960	2.326	2.576	2.807	3.291

Note: In calculating confidence intervals, σ may be substituted for s in Equation 4-6 if you have a great deal of experience with a particular method and have therefore determined its "true" population standard deviation. If σ is used instead of s, the value of t to use in Equation 4-6 comes from the bottom row of Table 4-2.

From D.C. Harris (2003) *Quantitative Chemical Analysis*, 6th Ed.

Confidence intervals

- Quantifies how far the true mean (μ) lies from the measured mean, $\bar{x}$. Uses the mean and standard deviation of the sample.

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

where t is from the t-table and n = number of measurements.

Degrees of freedom (df) = n - 1 for the CI.

Example of calculating a confidence interval

Consider measurement of dissolved Ti in a standard seawater (NASS-3):

Data: 1.34, 1.15, 1.28, 1.18, 1.33, 1.65, 1.48 nM

DF = $n - 1 = 7 - 1 = 6$

$\bar{x}$ = 1.3_4 nM or 1.3 nM

s = 0.1_7 or 0.2 nM

95% confidence interval

$t_{(df=6,95\%)} = 2.447$

CI_{95} = 1.3 ± 0.16 or 1.3 ± 0.2 nM

50% confidence interval

$t_{(df=6,50\%)} = 0.718$

CI_{50} = 1.3 ± 0.05 nM

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

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

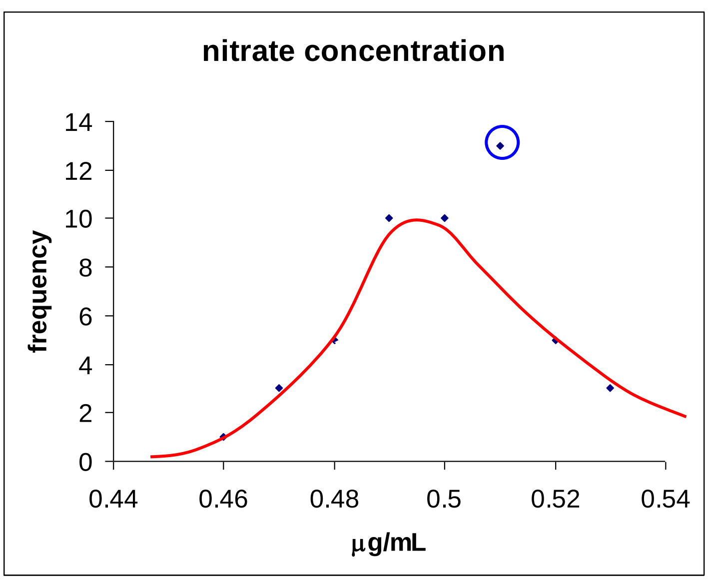
Nitrate Concentrations (μg/mL)

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	
0.51	0.51	0.51	0.5	0.51	0.49	0.52	0.53	0.5	0.47	
0.51	0.52	0.53	0.48	0.49	0.5	0.52	0.49	0.49	0.5	
0.49	0.48	0.46	0.49	0.49	0.48	0.49	0.49	0.51	0.47	
0.51	0.51	0.51	0.48	0.5	0.47	0.5	0.51	0.49	0.48	
0.51	0.5	0.5	0.53	0.52	0.52	0.5	0.5	0.51	0.51	
0.506	**0.504**	**0.502**	**0.496**	**0.502**	**0.492**	**0.506**	**0.504**	**0.5**	**0.486**	**mean**

average	**0.4998**
stdev	**0.01647**

mg/mL	frequency
0.53	**3**
0.52	**5**
0.51	**13**
0.5	**10**
0.49	**10**
0.48	**5**
0.47	**3**
0.46	**1**

Let’s Graph the Data!



Confidence Interval Exercise

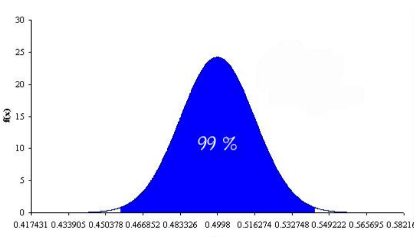
$$\bar{x} = \mu \pm t \cdot s_m = \mu \pm t \cdot \frac{s}{\sqrt{n}}$$

Calculate the 95, 98 and 99 % confidence intervals

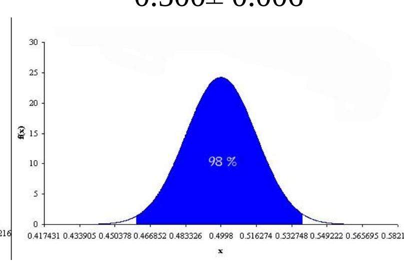
For the nitrate concentration data

95 %	0.500 ± 0.005
98 %	0.500 ± 0.006
99 %	0.500 ± 0.006
50 %	0.500 ± 0.002

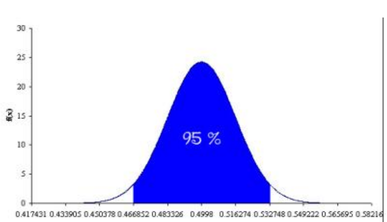
0.500 ± 0.006



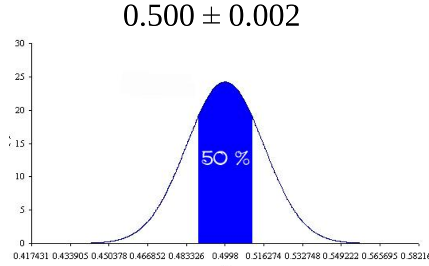
0.500 ± 0.006



0.500 ± 0.005



0.500 ± 0.002



Testing a Hypothesis (Significance Tests)

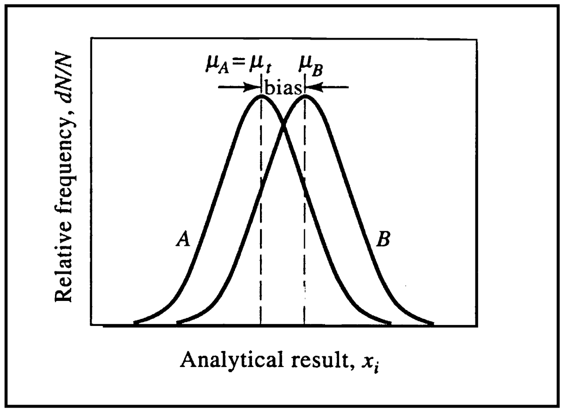
Carry out measurements on an accurately known standard.

Experimental value is different from the true value.

Is the difference due to a systematic error (bias) in the method - or simply to random error?

Assume that there is *no* bias (**NULL HYPOTHESIS**), and calculate the probability that the experimental error is due to random errors.

Figure shows (A) the curve for the true value ($\mu_A = \mu_t$) and (B) the experimental curve (μ_B)



Comparing a measured result with a "known" value

- "Known" value would typically be a certified value from a standard reference material (SRM)
- Another application of the *t* statistic

$$t_{calc} = \frac{|known\ value - \bar{x}|}{s}\sqrt{n}$$

Will compare t_{calc} to tabulated value of *t* at appropriate df and CL.

df = n -1 for this test

Comparing a measured result with a "known" value--example

Dissolved Fe analysis verified using NASS-3 seawater SRM
Certified value = 5.85 nM
Experimental results: $5.7_6 \pm 0.1_7$ nM (n = 10)

$$t_{calc} = \frac{|known\,value - \bar{x}|}{s}\sqrt{n} \quad = \quad \frac{|5.85 - 5.7_6|}{0.1_7}\sqrt{10} \quad = \quad 1.674$$

(Keep 3 decimal places for comparison to table.)

Compare to t_{table}; df = 10 - 1 = 9, 95% CL

$t_{table(df=9,95\%\ CL)}$ = 2.262

If $|t_{calc}| < t_{table}$, results are not significantly different at the 95% CL.

If $|t_{calc}| \geq t_{table}$, results are significantly different at the 95% CL.

For this example, $t_{calc} < t_{test}$, so experimental results are not significantly different at the 95% CL. THE NULL HYPOTHESIS IS MAINTAINED and no BIAS at the 95 % confidence level.

Comparing replicate measurements or comparing means of two sets of data

- Another application of the t statistic
- Example: Given the same sample analyzed by two different methods, do the two methods give the "same" result?

$$t_{calc} = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{pooled}}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

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

Will compare t_{calc} to tabulated value of t at appropriate df and CL.

df = $n_1 + n_2 - 2$ for this test

Comparing replicate measurements or comparing means of two sets of data—example

Ewww!

Determination of nickel in sewage sludge using two different methods

Method 1: Atomic absorption spectroscopy

Data: 3.91, 4.02, 3.86, 3.99 mg/g

$\bar{x}_1$ = 3.94_5 mg/g

s_1 = 0.07_3 mg/g

n_1 = 4

Method 2: Spectrophotometry

Data: 3.52, 3.77, 3.49, 3.59 mg/g

$\bar{x}_2$ = 3.5_9 mg/g

s_2 = 0.1_2 mg/g

n_2 = 4

Comparing replicate measurements or comparing means of two sets of data—example

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}} = \sqrt{\frac{(0.07_3)^2(4-1)+(0.1_2)^2(4-1)}{4+4-2}} = 0.0993$$

$$t_{calc} = \frac{|\bar{x}_1-\bar{x}_2|}{s_{pooled}}\sqrt{\frac{n_1 n_2}{n_1+n_2}} = \frac{|3.94_5-3.5_9|}{0.0993}\sqrt{\frac{(4)(4)}{4+4}} = 5.056$$

Note: Keep 3 decimal places to compare to t_{table}.

Compare to t_{table} at df = 4 + 4 – 2 = 6 and 95% CL.
$t_{table(df=6,95\%\ CL)}$ = 2.447

If $|t_{calc}| < t_{table}$, results <u>are not</u> significantly different at the 95%. CL.

If $|t_{calc}| \geq t_{table}$, results <u>are</u> significantly different at the 95% CL.

Since $|t_{calc}|$ (5.056) $\geq t_{table}$ (2.447), results from the two methods <u>are</u> significantly different at the 95% CL.

Comparing replicate measurements or comparing means of two sets of data

Wait a minute! There is an important assumption associated with this *t*-test:

It is assumed that the standard deviations (i.e., the precision) of the two sets of data being compared <u>are not significantly different</u>.

• How do you test to see if the two std. devs. are different?

• How do you compare two sets of data whose std. devs. are significantly different?

t-tests and the Law

Clearly, the meanings of 1.083 ± 0.007 and 1.0 ± 0.4 are very different. As a person who will either derive or use analytical results, you should be aware of this warning published in a report entitled "Principles of Environmental Analysis":

Analytical chemists must always emphasize to the public that the single most important characteristic of any result obtained from one or more analytical measurements is an adequate statement of its uncertainty interval. Lawyers usually attempt to dispense with uncertainty and try to obtain unequivocal statements: therefore, an uncertainty interval must be defined in cases involving litigation and or enforcement proceedings. Otherwise, a value of 1.001 without a specified uncertainty, for example may be views as legally exceeding a permissible level of 1.

L. K. Keith, W. Crummett, J. Deegan Jr., R. A. Libby, J. K. Taylor, and G. Wentler, *Analytical Chemis*try, **55**, 2210 (1983).

F-test to compare standard deviations

- Used to determine if std. devs. are significantly different before application of *t*-test to compare replicate measurements or compare means of two sets of data
- Also used as a simple general test to compare the precision (as measured by the std. devs.) of two sets of data
- Uses F distribution

F-test to compare standard deviations

Will compute F_{calc} and compare to $F_{table.}$

$$F_{calc} = \frac{s_1^2}{s_2^2} \quad where \quad s_1 > s_2$$

DF = n_1 - 1 and n_2 - 1 for this test.

Choose confidence level (95% is a typical CL).

Table 4-5 Critical values of $F = s_1^2/s_2^2$ at 95% confidence level

Degrees of freedom for s_2	Degrees of freedom for s_1													
	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
2	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5
3	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.81	8.79	8.74	8.70	8.66	8.62	8.53
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.63
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.36
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.67
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.58	3.51	3.44	3.38	3.23
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	2.93
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.71
10	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.70	2.54
11	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.57	2.40
12	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.30
13	3.81	3.41	3.18	3.02	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.38	2.21
14	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.31	2.13
15	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.07
16	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.19	2.01
17	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.15	1.96
18	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.11	1.92
19	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.07	1.88
20	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.84
30	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.62
∞	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.00

From D.C. Harris (2003) *Quantitative Chemical Analysis*, 6th Ed.

F-test to compare standard deviations

From previous example:
Let $s_1 = 0.1_2$ and $s_2 = 0.07_3$

$$F_{calc} = \frac{s_1^2}{s_2^2} = \frac{(0.1_2)^2}{(0.07_3)^2} = 2.70$$

Note: Keep 2 or 3 decimal places to compare with F_{table}.

Compare F_{calc} to F_{table} at df = (n_1 -1, n_2 -1) = 3,3 and 95% CL.

If $F_{calc} < F_{table}$, std. devs. <u>are not</u> significantly different at 95% CL.

If $F_{calc} \geq F_{table}$, std. devs. <u>are</u> significantly different at 95% CL.

$F_{table(df=3,3;95\%\ CL)} = 9.28$

Since F_{calc} (2.70) < F_{table} (9.28), std. devs. of the two sets of data <u>are not</u> significantly different at the 95% CL. (Precisions are similar.)

Comparing replicate measurements or comparing means of two sets of data-revisited

The use of the *t*-test for comparing means was justified for the previous example because we showed that standard deviations of the two sets of data were not significantly different.

If the F-test shows that std. devs. of two sets of data <u>are</u> significantly different and you need to compare the means, use a different version of the *t*-test →

Comparing replicate measurements or comparing means from two sets of data when std. devs. are significantly different

$$t_{calc} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

$$DF = \left\{ \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1} \right)} \right\} - 2$$

Flowchart for comparing means of two sets of data or replicate measurements

Use F-test to see if std. devs. of the 2 sets of data are significantly different or not

- Std. devs. are significantly different
  - Use the 2^{nd} version of the *t*-test ()
- Std. devs. are not significantly different
  - Use the 1^{st} version of the *t*-test (see previous, fully worked-out example)

One last comment on the F-test

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?*

Statistics in the News

Outliers Disrupt the Mean
January 01 1999

In 1984, according to Larry Gonick and Woollcott Smith, the University of Virginia announced that the mean starting salary of its graduates from the Department of Rhetoric and Communications was a very hefty $55,000 per year. But before you abandon your computer science training for speech classes, you should know that the graduating class contained a significant "outlier," or extreme data point not typical of the rest of the data set - Ralph Sampson, future NBA All-Star, who majored in speech. It would have been better to learn the median salary, the data point in the middle of the set.

Evaluating questionable data points using the Q-test

- Need a way to test questionable data points (outliers) in an unbiased way.
- Q-test is a common method to do this.
- Requires 4 or more data points to apply.

Calculate Q_{calc} and compare to Q_{table}

Q_{calc} = gap/range

Gap = (difference between questionable data pt. and its nearest neighbor)

Range = (largest data point – smallest data point)

Evaluating questionable data points using the Q-test--example

Consider set of data; Cu values in sewage sample:
9.52, 10.7, 13.1, 9.71, 10.3, 9.99 mg/L

Arrange data in increasing or decreasing order:
9.52, 9.71, 9.99, 10.3, 10.7, 13.1

The questionable data point (outlier) is 13.1

Calculate $Q_{calc} = \frac{gap}{range} = \frac{(13.1-10.7)}{(13.1-9.52)} = 0.670$

Compare Q_{calc} to Q_{table} for *n* observations and desired CL (90% or 95% is typical). It is desirable to keep 2-3 decimal places in Q_{calc} so judgment from table can be made.

$Q_{table\ (n=6, 90\%\ CL)} = 0.56$

TABLE 2.3

Rejection Quotient, Q, at Different Confidence Limits[a]

No. of Observations	Confidence level Q_{90}	Q_{95}	Q_{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
15	0.338	0.384	0.475
20	0.300	0.342	0.425
25	0.277	0.317	0.393
30	0.260	0.298	0.372

[a]Adapted from D. B. Rorabacher, *Anal. Chem.*, 63 (1991) 139.

From G.D. Christian (1994) *Analytical Chemistry*, 5th Ed.

Evaluating questionable data points using the Q-test--example

If $Q_{calc} < Q_{table}$, do not reject questionable data point at stated CL.

If $Q_{calc} \geq Q_{table}$, reject questionable data point at stated CL.

From previous example,
Q_{calc} (0.670) > Q_{table} (0.56), so reject data point at 90% CL.

Subsequent calculations (e.g., mean and standard deviation) should then exclude the rejected point.

Mean and std. dev. of remaining data: $10.0_4 \pm 0.4_7$ mg/L

Q or G outlier test?

$$G_{calc} = \frac{|questionable\_value - \bar{x}|}{s}$$

reject if $G_{calc} > G_{table}$

G (95 % confidence)	Number of Observations
1.463	4
1.672	5
1.822	6
1.938	7
2.032	8
2.11	9
2.176	10
2.234	11
2.285	12
2.409	15
2.557	20

$$Q_{calc} = \frac{gap}{range}$$

reject if $Q_{calc} > Q_{table}$

Q (90 % confidence)	Number of Observations
0.76	4
0.64	5
0.56	6
0.51	7
0.47	8
0.44	9
0.41	10

No. of observations	90%	95%	99% confidencelevel
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

Rejection of outlier recommended if $Q_{calc} > Q_{table}$ for the desired confidence level.

Note:1. The higher the confidence level, the less likely is rejection to be recommended.

2. Rejection of outliers can have a marked effect on mean and standard deviation, esp. when there are only a few data points. *Always try to obtain more data*.

Q Test for Rejection of Outliers

The following values were obtained for the concentration of nitrite ions in a sample of river water: 0.403, 0.410, 0.401, 0.380 mg/l. Should the last reading be rejected?

$$Q_{calc} = |0.380 - 0.401| / (0.410 - 0.380) = 0.7$$

But Q_{table} = 0.829 (at 95% level) for 4 values

Therefore, $Q_{calc} < Q_{table}$, and we cannot reject the suspect value.

Suppose 3 further measurements taken, giving total values of:

0.403, 0.410, 0.401, 0.380, 0.400, 0.413, 0.411 mg/l. Should 0.380 still be retained?

$$Q_{calc} = |0.380 - 0.400| / (0.413 - 0.380) = 0.606$$

But Q_{table} = 0.568 (at 95% level) for 7 values

Therefore, $Q_{calc} > Q_{table}$, and rejection of 0.380 is recommended.

But note that 5 times in 100 it will be wrong to reject this suspect value! Also note that if 0.380 is retained, s = 0.011 mg/l, but if it is rejected, s = 0.0056 mg/l, i.e. precision appears to be twice as good, just by rejecting one value.