Comparison of means with student's "t":

Comparing a measured result with a "known" value:

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

Problem:

You purchased a standard reference material coal sample certified by the National Institute of Standard and Technology to contain 3.19 wt% sulfur. you are testing a new analytical method to see whether it can reproduce the known value. The measured values are 3.29, 3.22, 3.30 and 3.23 wt% sulfur, giving a mean of $\bar{x} = 3.26_0$ and a standard deviation of $s = 0.04_1$. Does your answer agree with the known answer?

Soln

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

where $\bar{x}$ – mean.

n – number of observation.

s – Standard deviation.

$\bar{x} = 3.26_0$

$s = 0.04_1$

$n = 4$

$$t_{calculated} = \frac{|3.26_0 - 3.19|}{0.04_1}\sqrt{4}$$

$$= 3.4_1$$

From the Students t_{table}, look in the 95% Confidence Column across from $\overset{4-1}{(n-1)} = 3$ degrees of freedom to find $t_{table} = 3.182$.

Here

$$t_{calculated}(3.4_1) > t_{table}(3.182)$$

You conclude that your result is different from the known value.

Comparing Replicate Measurements:

$$\boxed{t_{calculated} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{pooled}} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}$$

$$S_{pooled} = \sqrt{\frac{\sum_{Set\,1}(x_i - \bar{x}_1)^2 + \sum_{Set\,2}(x_j - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Table 4-3 from the Book

Mass of gas isolated by Lord Rayleigh: 4

From air (g)	From chemical decomposition (g)
2.310 17	2.301 43
2.309 86	2.298 90
2.310 10	2.298 16
2.310 01	2.301 82
2.310 24	2.298 69
2.310 10	2.299 40
2.310 28	2.298 49
~~2.310 28~~	2.298 89
	~~2.298 89~~
$\bar{x} = 2.310\ 11$	$\bar{x} = 2.299\ 47$

Standard deviation

$s = 0.00014_3$

$s = 0.001\ 38$

Soln

Gas from air

$\bar{x}_1 = 2.31011$

$S_1 = 0.00014_3$

$n_1 = 7$

Gas from chemical source.

$\bar{x}_2 = 2.29947$

$S_2 = 0.00138$

$n_2 = 8$

$$\therefore S_{Pooled} = \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(0.00014_3)^2(7-1) + (0.00138)^2(8-1)}{7+8-2}}$$

$$= 0.00102$$

$$\therefore t_{calculated} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{Pooled}} \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$t_{Calculated} = \frac{|2.31011 - 2.29947|}{0.00102}\sqrt{\frac{7 \times 8}{7+8}}$$

$$= 20.2$$

For Degrees of freedom

$7 + 8 - 2 = 13$

$\therefore$ in the t_{table} 13 d.f lies between

2.228 and 2.131 for 95% confidence.

Here

$t_{Calculated} > t_{table}$

$\therefore$ the difference is significant.

Comparing Individual differences:
(Paired t test)

$$t_{calculated} = \frac{\bar{d}}{S_d}\sqrt{n}$$

Where

$$S_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

Where

$\bar{d}$ is the average difference between methods A and B.

n is the number of pairs.

Example :

4-14. The Ti content (wt%) of five different ore samples (each with a different Ti content) was measured by each of two methods.

Sample	Method 1	Method 2
A	0.0134	0.0135
B	0.0144	0.0156
C	0.0126	0.0137
D	0.0125	0.0137
E	0.0137	0.0136

Soln

Sample	Method 1	Method 2	d_i	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
A	0.0134	0.0135	−0.0001	+0.0006	3.6×10^{-7}
B	0.0144	0.0156	−0.0012	−0.0005	2.5×10^{-7}
C	0.0126	0.0137	−0.0011	−0.0004	1.6×10^{-7}
D	0.0125	0.0137	−0.0012	−0.0005	2.5×10^{-7}
E	0.0137	0.0136	+0.0001	+0.0008	6.4×10^{-7}
			$\bar{d} = \bar{0}.00070$		$\Sigma = 16.6 \times 10^{-7}$

Standard deviation

$$Sd = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n-1}}$$

$$= \sqrt{\frac{16.6 \times 10^{-7}}{5-1}} = \sqrt{\frac{16.6 \times 10^{-7}}{4}}$$

$$= 6.4 \times 10^{-4}$$

$$t_{calculated} = \frac{\bar{d}}{S_d}\sqrt{n}$$

$$= \frac{0.00070}{0.00064}\sqrt{5}$$

$$= 2.4.$$

From the Students t for 4 degrees of freedom 2.776.

$\therefore t_{calculated}(2.4) < t_{table}(2.776)$

The $t_{calculated}$ value is less than t_{table} value. The difference is not significant.

Comparison of standard deviations with the F test:

$$F_{calculated} = \frac{S_1^2}{S_2^2}$$

The F-test tells the whether two standard deviations are "significantly" different from each other. F is the quotient of the squares of the standard deviations.

We always put the larger standard deviation in the numerator so that $F \geq 1$.

If $F_{calculated} > F_{table}$, then the difference is significant.

Example:

In $\boxed{\text{Table A-3 in the Book}}$, the larger standard deviation is $S_1 = 0.00158$

$n_1 = 8$ measurement

$S_2 = 0.000143$

$n_2 = 7$ measurement.

Solution:

To answer, find F with the following equation.

$$F_{calculated} = \frac{S_1^2}{S_2^2}$$

$$= \frac{(0.00158)^2}{(0.000143)^2} = 93.1$$

We look for F_{table}:

In the <u>column</u> with 7 degrees of freedom for S_1 (because degree of freedom $= n-1$).

and the row with 6 degrees of freedom for S_2 (because d.f. $n_1 - 1$).

$F_{\text{Calculated}}$ (93.1) > F_{Table} (4.21).

The standard deviations are different from each other above the 95% confidence level.

$\boxed{A-19.}$ Hydrocarbons in the cab of an automobile were measured during trips on the New Jersey Turnpike and trips through the Lincoln Tunnel connecting New York and New Jersey. The total concentrations (± standard deviations) of m- and p-xylene were.

Turnpike : 31.4 ± 30.0 $\mu g/m^3$ (32 measurements)

Tunnel : 52.9 ± 29.8 $\mu g/m^3$ (32 measurements)

Do these results differ at the 95% confidence level? At the 99% confidence level?

Soln:

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

$$S_{pooled} = \sqrt{\frac{S_1^2 (n_1 - 1) + S_2^2 (n_2 - 1)}{n_1 + n_2 - 2}}$$

$\bar{x}_1 = 31.4$; $S_1 = 30.0$; $n = 32$

$\bar{x}_2 = 52.9$; $S_2 = 29.8$; $n = 32$

$$\therefore S_{pooled} = \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(30.0)^2(32-1) + (29.8)^2(32-1)}{32+32-2}}$$

$$= \sqrt{\frac{(30.0)^2(31) + (29.8)^2(31)}{62}}$$

$S_{pooled.} = 29.9$

$$t_{calculated} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{pooled}} \sqrt{\frac{n_1 . n_2}{n_1 + n_2}}$$

$$t_{calculated} = \frac{|52.9 - 31.4|}{29.9} \sqrt{\frac{32 \times 32}{32 + 32}}$$

$$= 2.88.$$

Degrees of freedom $32 + 32 - 2 = \underline{\underline{62}}.$

ie. D.F $= 62$

In the t table this 62 is close to 60 degrees of freedom. ~~which is~~

@ 95% confidence level.

$t_{calculated}\ (2.88) > t_{table}\ (2.000)$

@ 99% confidence level

$t_{calculated}\ (2.88) > t_{table}\ (2.660)$

∴ The difference is significant at the 95 and 99% levels

Q test for Bad Data:

Sometimes one datum is inconsistent with the remaining data. You can use the Q test to help decide whether to retain or discard a questionable datum.

$$Q_{calculated} = \frac{Gap}{range}.$$

Gap - is the difference between the questionable point and the nearest value.

range = Largest value - Lowest value.

$\boxed{A-22}$. Using the Q test, decide whether the value 216 should be rejected from the set of results 192, 216, 202, 195, and 204.

<u>Soln</u>

Arrange the data according to increasing order of the number.

$$\underbrace{192,\ 195,\ 202,\ \overbrace{204,\ 216}^{\text{Gap}}}_{\text{Range}}. \qquad \boxed{n = 5}$$

$$Q_{\text{calculated}} = \frac{\text{Gap}}{\text{Range}}$$

$$= \frac{(216 - 204)}{(216 - 192)} = \frac{12}{24} = 0.50$$

From the Q table. for $n = 5$:

$$Q_{\text{table}} = 0.64$$

$\therefore Q_{\text{calculated}}\ (0.50) < Q_{\text{table}}\ (0.64)$, $\therefore$ <u>Retain 216</u>