

"There are three kinds of lies: lies, damned lies, and statistics."

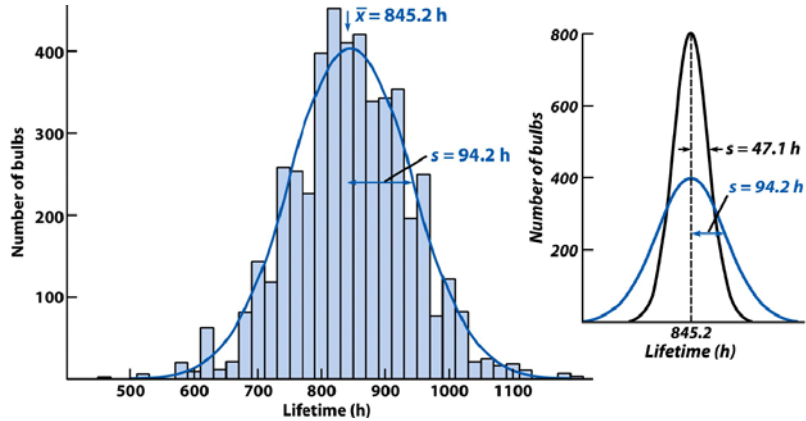
Benjamin Disraeli, 1895 (British statesman)

## H 4-1 Gaussian Distribution

If a measurement is repeated many times a statistical treatment of the data can provide an indication of the reliability of the results. If the errors associated with the measurement are completely random then the Central Limit Theorem assures that the data will follow a mathematical form called a **Gaussian distribution** (bell-shaped curve). In the limit of an infinite number of measurements the histogram below becomes the population distribution denoted by the solid line.

The histogram on the right was drawn to have the same mean, standard deviation, and area as the smooth curve. This is not true for a finite number of measurements. Only in the infinite limit will they be the same.

The population distribution (infinite  $n$ ) is characterized by a population mean (average - center of the symmetric distribution),  $\mu$ , and a population standard deviation (measure of the width of the distribution),  $\sigma$  (Harris uses symbol  $s$  for a finite distribution). Another useful measure is the square of the standard deviation,  $\sigma^2$ , known as the variance.



$$\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

A Gaussian curve,  $y_{\text{Gaussian}}$  is totally defined by these variables:  $y_{\text{Gaussian}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

### Obtaining the Probability

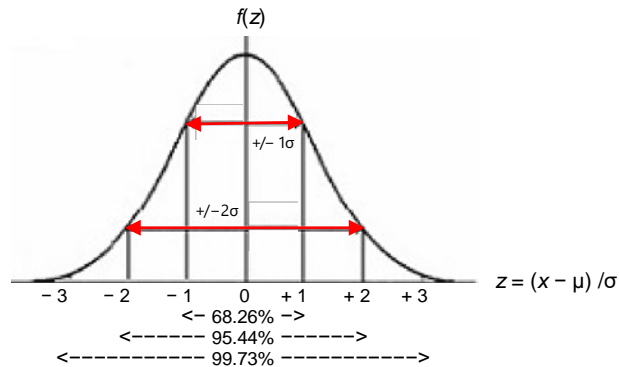
All measurements contain experimental error. However, by a judicious use of statistics one can associate a probability with the result. To convert the Gaussian distribution into a probability density a new variable  $z = (x - \mu) / \sigma$  is introduced. Then the Gaussian curve becomes the **normal Gaussian error curve**.

$$y = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$f(0) = \frac{1}{\sqrt{2\pi}} = 0.3989$$

$$f(1) = \frac{1}{\sqrt{2\pi e}} = 0.2419$$

( $x = \mu \pm \sigma$  are inflection points)



Integration of the population distribution between two limits as from  $x = \mu - z\sigma$  to  $x = \mu + z\sigma$  gives the probability of obtaining a value of  $x$  between these limits. The integral limits that give 95% of the population distribution area (a probability of 0.95) are **95% confidence limits**. The symmetric Gaussian distribution requires that these limits be equidistant from  $\mu$ . The integral of a Gaussian between finite limits is not analytic and the area under the curve is given in tables.

How can one characterize the reliability of data which is not infinite? First one needs to obtain the mean and standard deviation of the data. The finite distribution is characterized by a **sample mean** (sample average),  $\langle x \rangle$ , and a **sample standard deviation**,  $s$ . Another useful measure is the square of the standard deviation  $s^2$ , the sample **variance**. The quantity  $n - 1$  below is called the degrees of freedom.

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^n x_i \quad s = \sqrt{\sum_i (x_i - \langle x \rangle)^2 / (n - 1)}$$

## H 4-2 Comparison of Standard Deviations with F Test

To determine whether two standard deviations for two different sets of measurements are or are not statistically the same obtain the means and standard deviations for each set

$n_1$  measurements,  $\langle x_1 \rangle$  mean,  $s_1$  standard deviation

$n_2$  measurements,  $\langle x_2 \rangle$  mean,  $s_2$  standard deviation

and determine

$$F_{\text{calculated}} = s_1^2 / s_2^2 \quad \text{where } F \geq 1$$

If  $F_{\text{calculated}} > F_{\text{table}}$  (table given below or found in a spreadsheet) the difference is significant and the two measurements are statistically different.

**Hypothesis testing** based upon assuming that the null hypothesis is true at a certain level of probability, generally chosen to be 5%. The hypothesis is accepted if the probability for it being true is  $> 5\%$  and rejected if the probability for it being true is  $< 5\%$ .

Null hypothesis for the F test: two sets of measurements taken from populations with the same population standard deviation; all differences arise from only random variations in measurement

accept:  $F_{\text{calculated}} < F_{\text{table}} \Rightarrow$  standard deviations are not statistically different

reject:  $F_{\text{calculated}} > F_{\text{table}} \Rightarrow$  standard deviations are statistically different

## H 4-3 Confidence Intervals (inferences based on small samples)

Table 4-3 Critical values of  $F = s^2/s^2$  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	$\infty$
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
$\infty$	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

For an estimate of the uncertainty of a limited number of measurements one can use a confidence interval

$$\text{confidence interval} = \langle x \rangle \pm ts / \sqrt{n}$$

which implies that the true population mean will be found within a range of  $ts / \sqrt{n}$  of the sample mean with a confidence level (level of certainty) specified by the particular  $t$  chosen if one were to repeat the  $n$  measurements many times. Then a 95% confidence interval would include the true population mean in 95% of these sets of  $n$  measurements. Note that this implies that the uncertainty in the sample mean is reduced by more measurements by a factor of  $1/\sqrt{n}$ .

Null hypothesis for the  $t$  test: two sets of measurements taken from populations with the same mean; all differences arise from only random variations in measurement

accept:  $t_{\text{calculated}} < t_{\text{table}} \Rightarrow$  means are not statistically different

reject:  $t_{\text{calculated}} > t_{\text{table}} \Rightarrow$  means are statistically different

**Table 4-4 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	63.656	127.321	636.578
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	2.060	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
$\infty$	0.674	1.645	1.960	2.326	2.576	2.807	3.291

#### H 4-4 Comparison of Means with Student's $t$ (always use 95% confidence intervals)

**Case 1: Comparison to a Known or Standard Value**

**Case 2: Comparison of Replicate Measurements (apply  $F$  test first)**

For this case, the  $F$  test must first be applied to the two standard deviations. Whether the standard deviations are or are not statistically equivalent determines which of the set of formulas on the next page one uses. For this comparison when the standard deviations are statistically equivalent one first obtains a pooled standard deviation then uses it to calculate a value of  $t$  which is compared with  $t$  in Student's  $t$ . If the calculated  $t$  is greater than the 95% confidence level value of  $t$  the two replicates are considered different. When the standard deviations are not statistically equivalent one cannot determine the degrees of freedom from the number of measurements and it must be calculated. For two sets of data with  $n_1$  and  $n_2$  measurements, means  $\langle x_1 \rangle$  and  $\langle x_2 \rangle$ , and standard deviations  $s_1$  and  $s_2$  the two sets of formulas are

$F_{\text{calculated}} < F_{\text{table}}$	$F_{\text{calculated}} > F_{\text{table}}$
$t_{\text{calculated}} = \frac{ \langle x_1 \rangle - \langle x_2 \rangle }{s_{\text{pooled}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$	$t_{\text{calculated}} = \frac{ \langle x_1 \rangle - \langle x_2 \rangle }{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$
$s_{\text{pooled}} = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}}$	$\frac{\text{degrees of freedom}}{\text{freedom}} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$

## Case 2: Comparison of Individual Differences with Paired t Test

### 4-6 Grubbs Test for Outliers ( use 96% confidence)

To determine whether a particular data point can be excluded based upon its questionable veracity, form the Grubbs statistic,  $G$

$$G_{\text{calculated}} = \frac{|x_{\text{questionable}} - \langle x \rangle|}{s}$$

If  $G_{\text{calculated}} > G_{\text{table}}$  then the point can be excluded with the chosen confidence level (here 95%). The mean and standard deviation will need to be recalculated. Hint: generally do not exclude a data point unless you are certain that an error occurred in its measurement. **Never** exclude more than one point. Always use a  $G$  of at least a 95% confidence level.

**TABLE 4-6** Critical values of  $G$  for rejection of outlier

Number of observations	$G$ (95% confidence)
4	1.463
5	1.672
6	1.822
7	1.938
8	2.032
9	2.110
10	2.176
11	2.234
12	2.285
15	2.409
20	2.557

**NOTE:** For the  $F$ ,  $t$ , and  $G$  statistics if the calculated value is less than the table value the null hypothesis is true, you do nothing!