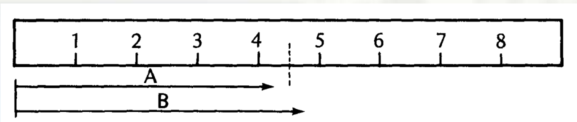
Data Analysis Summary

Errors of observation

- Error: the difference between the observed value of any physical quantity and the 'actual' value.
- Human error: due to observational reasons.
- Systematic error: due to the instruments and it determines the accuracy of the reading.
- Random error: due to environmental reasons or nature of the measurement
- Note difference between accuracy and precision.

Estimation of Errors



- The maximum possible error is 0.5 cm
- Therefore A = 4.0 ± 0.5 and B = 5.0 ± 0.5
- Absolute error is 0.5 cm, but relative (percentage) error changes.
- Relative error of A is 0.5/4.0 x 100% = 12.5%
- If C is 50, then relative error of C is 0.5/50 x 100% = 1%.

Propagation of Errors

If $x = a + b$, then $\Delta x = \Delta a + \Delta b$

If $x = a - b$, then $\Delta x = \Delta a + \Delta b$

If $x = a \times b$, then $\dfrac{\Delta x}{x} = \dfrac{\Delta a}{a} + \dfrac{\Delta b}{b}$

If $x = \dfrac{a}{b}$, then $\dfrac{\Delta x}{x} = \dfrac{\Delta a}{a} + \dfrac{\Delta b}{b}$

If $x = ka$, where k is a constant, then $\Delta x = k\Delta a$

In general, if $u = u(x, y, z, \ldots)$,

$$\text{then} \quad \delta u = \frac{\partial u}{\partial x}\delta x + \frac{\partial u}{\partial y}\delta y + \frac{\partial u}{\partial z}\delta z + \Lambda$$

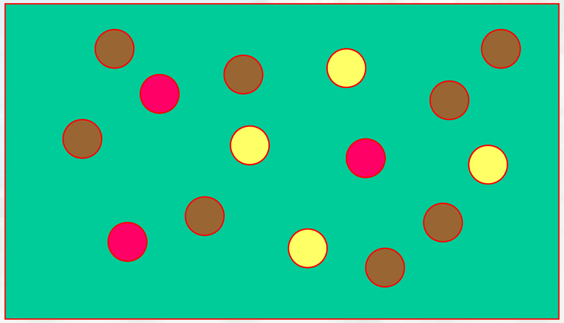
Examples

Given a = 10 ± 2, b = 7 ± 3.

- d = a + b = 17 ± (2+3).
- e = a - b = 3 ± (2+3).
- f = a x b = 70 ± Δf
  where Δf = (2/10 + 3/7) x 70 = 44.
- g = a/b = 1.43 ± Δg
  where Δg = (2/10 + 3/7) x 1.43 = 0.90.

Statistical Nature of Random Error

•No “accurate” value
•Mean value with uncertainty



e.g. radioactive decay

Terminologies

- Expectation value
- Frequency distribution function
- Mean value
  – For discrete x values, the mean
  $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$
  – For continuous distribution,
  $$\bar{x} = \frac{\int_{x_1}^{x_2} x f(x) dx}{\int_{x_1}^{x_2} f(x) dx}$$
  – The mean value is only a sample mean, it may be quite different from the true mean value.

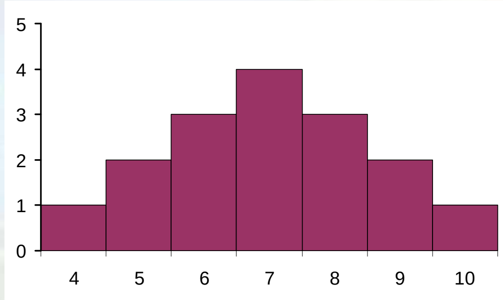
Terminologies

- Deviation = $x_i - \bar{x}$
- Sample variance s^2
  $$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \quad \text{or} \quad s^2 = \frac{\int_{x_1}^{x_2} (x-\bar{x})^2 f(x) dx}{\int_{x_1}^{x_2} f(x) dx}$$
- Standard deviation = $(\text{variance})^{1/2}$ = s
- Normalization – a distribution function is said to be normalized if the total area under the distribution curve is equal to 1. i.e.
  $$\int_{-\infty}^{\infty} f(x) dx = 1$$

Statistical Models

- Binomial distribution
- Poisson distribution
- Gaussian or Normal distribution
  - standard deviation $\sigma = (\text{mean})^{1/2}$
  - distribution is symmetric about mean

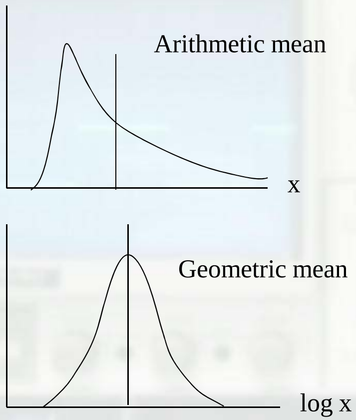
Example - mean



x-value
4
5
5
6
6
6
7
7
7
7
8
8
8
9
9
10

Mean value
= (4x1 + 5x2 + 6x3 +7x4 + 8x3 + 9x2 + 10x1) /16 = 7

Arithmetic & Geometric Mean



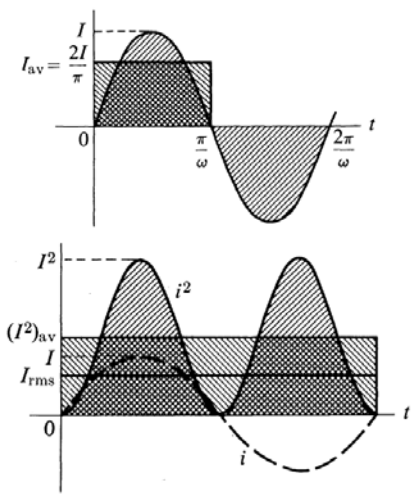
e.g. the size distribution of atmospheric aerosols is Guassian when the size is plotted in a log scale.

Mean and Root-mean-squared Value

- R.M.S. equals the square root of the mean of the squared value.
- E.g. for five measurement results: 9.81, 9.80, 10.0, 9.60, 9.71
  - Mean = (9.81 + 9.80 + 10.0 + 9.60 + 9.71)/5 = 9.784
  - R.M.S. =

$$\sqrt{\frac{9.81^2 + 9.80^2 + 10.0^2 + 9.60^2 + 9.71^2}{5}} = 9.785$$

R.M.S. of AC



$$I_{av} = \frac{\omega}{\pi}\int_0^{\pi/\omega} I \sin \omega t \, dt = \frac{2I}{\pi}$$

$$i^2 = I^2 \sin^2 \omega t = \frac{1}{2}I^2 - \frac{1}{2}I^2 \cos 2\omega t$$

$$(I^2)_{av} = \frac{I^2}{2}$$

$$I_{rms} = \sqrt{(I^2)_{av}} = \frac{I}{\sqrt{2}}$$

Example - standard deviation

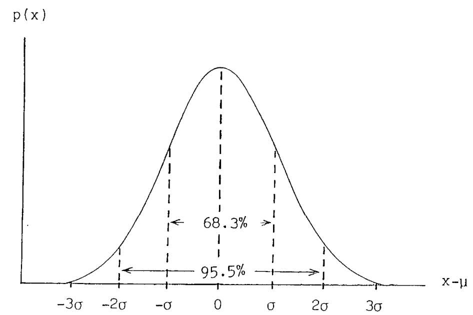
What is the standard deviation of the previous distribution?

1. $$\sigma = \sqrt{\frac{\sum_1^n (x-\bar{x})^2}{n-1}} = 1.6$$

2. $$\sigma = \sqrt{\bar{x}} = \sqrt{7} = 2.6$$

Confidence Levels

The confidence of observing a value within the range μ-r to μ+r is equal to the probability of its occurrence given by $\int_{\mu-r}^{\mu+r} P(x)dx$.



Range r	0.675 σ	σ	2 σ	3 σ
Confidence	50 %	68.3 %	95.4 %	99.7 %

Example - relative error

X = 100

σ = 10

σ/x = 0.1 = 10 %

X = 100 ± 10

at 68.3 % C.L.

X = 10000

σ = 100

σ/x = 0.01 = 1 %

X = 10000 ± 100

at 68.3 % C.L.

Useful statistical equations

$$\sigma_x = \sqrt{\bar{x}} \qquad \frac{\sigma_x}{\bar{x}} = \frac{1}{\sqrt{\bar{x}}}$$

$$\sigma_R = \frac{\sigma_x}{t} = \sqrt{\frac{R}{t}} \qquad \frac{\sigma_R}{R} = \frac{1}{\sqrt{Rt}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\bar{x}}{n}} \qquad \frac{\sigma_{\bar{x}}}{\bar{x}} = \frac{1}{\sqrt{n\bar{x}}}$$

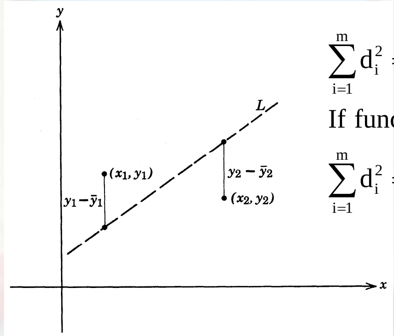
If $u = u(x, y, z, \ldots)$ and if the errors are individually small and symmetric about zero, then

$$\sigma_u^2 = \left(\frac{\partial u}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial u}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial u}{\partial z}\right)^2 \sigma_z^2 + \Lambda$$

17

Least Squares Approximations

- A curve fitting technique
- Minimize the sum of the squares of the deviations



$$\sum_{i=1}^{m} d_i^2 = \sum_{i=1}^{m} (y_i - \bar{y}_i)^2$$

If function is linear, then $\bar{y} = a_1 x + a_0$

$$\sum_{i=1}^{m} d_i^2 = \sum_{i=1}^{m} (y_i - a_1 x - a_0)^2$$

18

Least Square Fit

To minimize S with respect to a & b

$$\frac{\partial S}{\partial a_0} = \sum_{i=1}^{m} 2(y_i - a_1 x_i - a_0)(-1) = 0$$

$$\frac{\partial S}{\partial a_1} = \sum_{i=1}^{m} 2(y_i - a_1 x_i - a_0)(-x_i) = 0$$

Rearranging terms we get

$$m a_0 + \left(\sum x_i\right) a_1 = \sum y_i \quad \text{and} \quad \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

Solving the two linear equations, we have

$$a_0 = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{m \sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a_1 = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - \left(\sum x_i\right)^2}$$

19

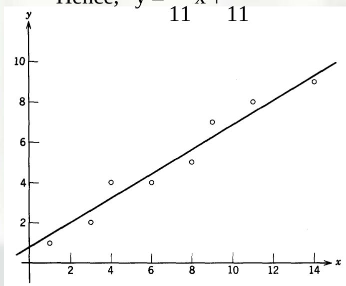
Example

	x	y	x^2	xy
	1	1	1	1
	3	2	9	6
	4	3	16	16
	6	3	36	24
	8	4	64	40
	9	7	81	63
	11	8	121	88
	14	9	196	126
sums:	56	40	524	364

$$a_0 = \frac{40 \times 524 - 56 \times 364}{8 \times 524 - (56)^2} = \frac{6}{11}$$

$$a_1 = \frac{8 \times 364 - 56 \times 40}{8 \times 524 - (56)^2} = \frac{7}{11}$$

Hence, $y = \frac{7}{11} x + \frac{6}{11}$



Other curves

Polynomial: $y = a_n x^n + a_{n-1} x^{n-1} + \Lambda + a_1 x + a_0$

Hyperbola: $y = \frac{1}{a_0 + a_1 x}$

Let $z = \frac{1}{y}$, then $z = a_0 + a_1 x$

Exponential: $y = a(b^x) \Rightarrow \log y = \log a + x \log b$

Geometric: $y = ax^b \Rightarrow \log y = \log a + b \log x$

Trigonometric: $y = a_0 + a_1 \cos \omega x$

or more generally $y = a_0 + \sum_{k=1}^{n} a_k \cos(k\omega x) + \sum_{k=1}^{n} b_k \sin(k\omega x)$