Chapter 17

Statistics

	page
Probability distributions	408
Gaussian distribution	408
Binomial distribution	410
Poisson distribution	410
Chi-square distribution, confidence levels	413
χ^2 test of distributions	414
Propagation of errors	415
Least-squares fit to a straight line	415
Bibliography	416

© Cambridge University Press • Provided by the NASA Astrophysics Data System

Gaussian distribution

$$P_G(x, \mu, \sigma) = (\sigma\sqrt{(2\pi)})^{-1} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

where

x = value of random observation,
μ = mean value of parent distribution,
σ = standard deviation of parent distribution,
σ^2 = variance.

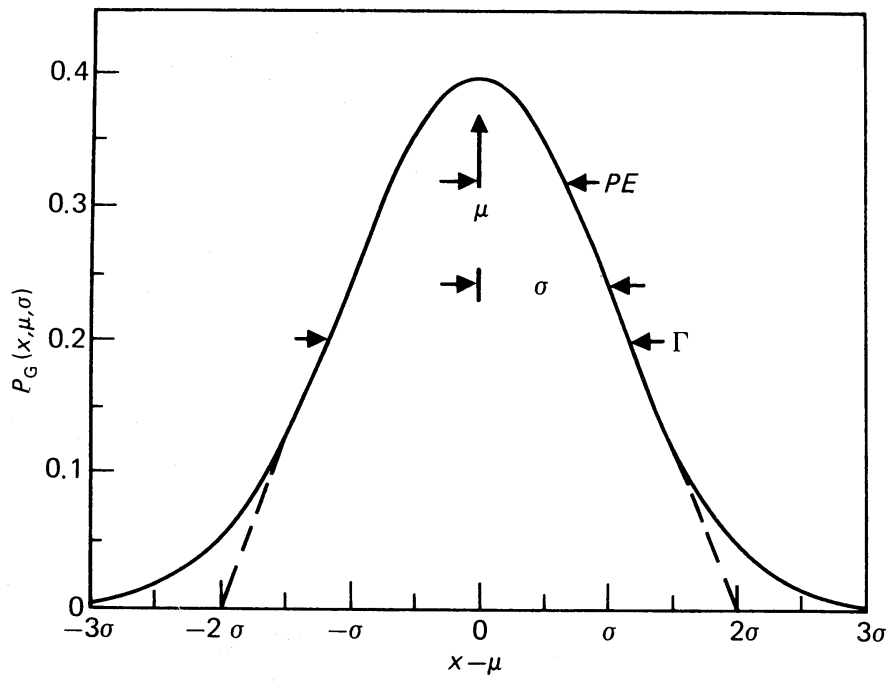
FWHM:

$$\Gamma = 2.354\sigma.$$

Probable error:

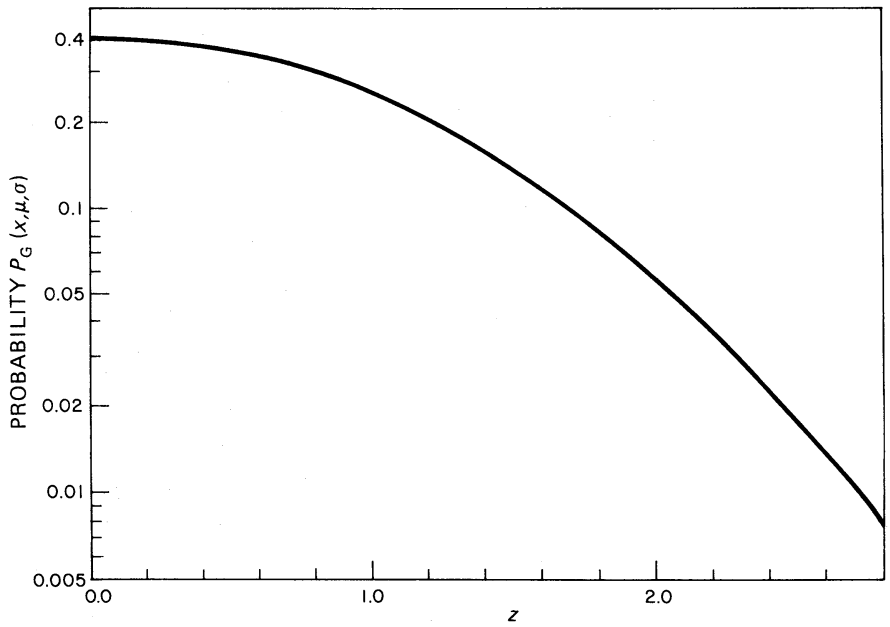
$$\text{PE} = 0.6745\sigma.$$

Gaussian probability distribution $P_G(x, \mu, \sigma)$ vs. $x - \mu$; $\Gamma = 2.354\sigma$ probable error (PE) $= 0.6745\,\sigma$. (Adapted from Bevington, P. R., ; *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Book Company, 1969.)



© Cambridge University Press • Provided by the NASA Astrophysics Data System

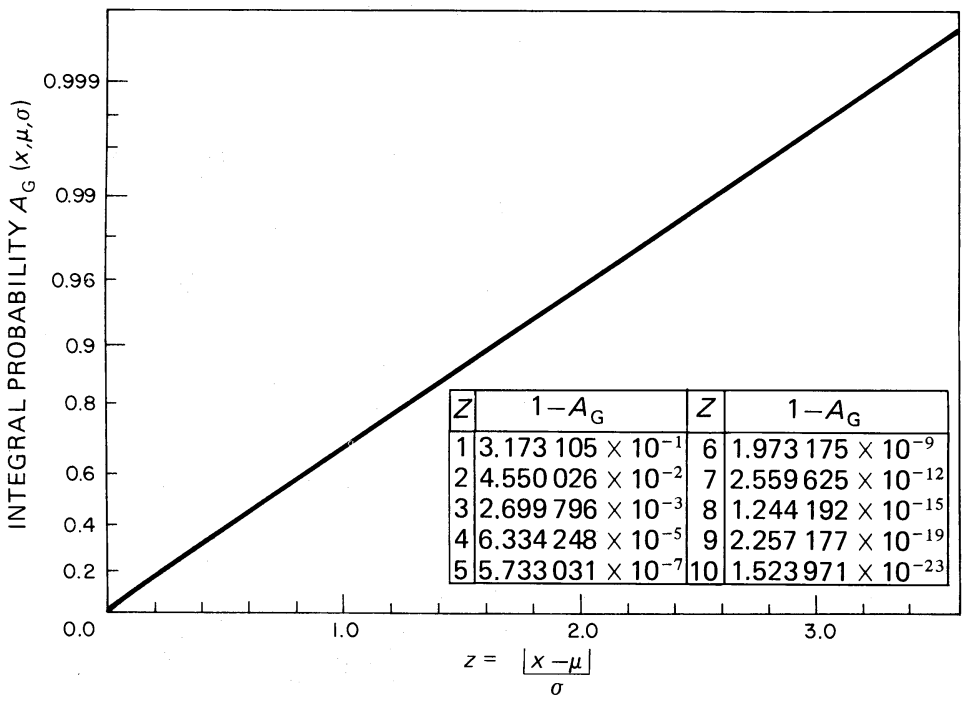
The Gaussian probability distribution $P_G(x, \mu, \sigma)$ vs. $z = |x - \mu|/\sigma$.



The integral of the Gaussian probability distribution $A_G(x, \mu, \sigma)$ vs. $z = |x - \mu|/\sigma$, where

$$A_G(x, \mu, \sigma) = \int_{\mu - z\sigma}^{\mu + z\sigma} P_G(x, \mu, \sigma)\, dx$$

(Adapted from Bevington, P. R., *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Book Company, 1969.)



Z	$1-A_G$	Z	$1-A_G$
1	$3.173\,105 \times 10^{-1}$	6	$1.973\,175 \times 10^{-9}$
2	$4.550\,026 \times 10^{-2}$	7	$2.559\,625 \times 10^{-12}$
3	$2.699\,796 \times 10^{-3}$	8	$1.244\,192 \times 10^{-15}$
4	$6.334\,248 \times 10^{-5}$	9	$2.257\,177 \times 10^{-19}$
5	$5.733\,031 \times 10^{-7}$	10	$1.523\,971 \times 10^{-23}$

© Cambridge University Press • Provided by the NASA Astrophysics Data System

Binomial distribution

$$P_B(x, n, p) = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}$$

If p is the probability that an event will occur, then in a random group of n independent trials, P_B is the probability that the event will occur x times.

$$\mu = \text{mean} = np$$

$$\sigma^2 = \text{variance} = np(1-p)$$

Poisson distribution

$$P_p(x, \mu) = \frac{\mu^x}{x!}\, e^{-\mu}$$

P_p is the probability of observing x events when the average for a large number of trials is μ events.

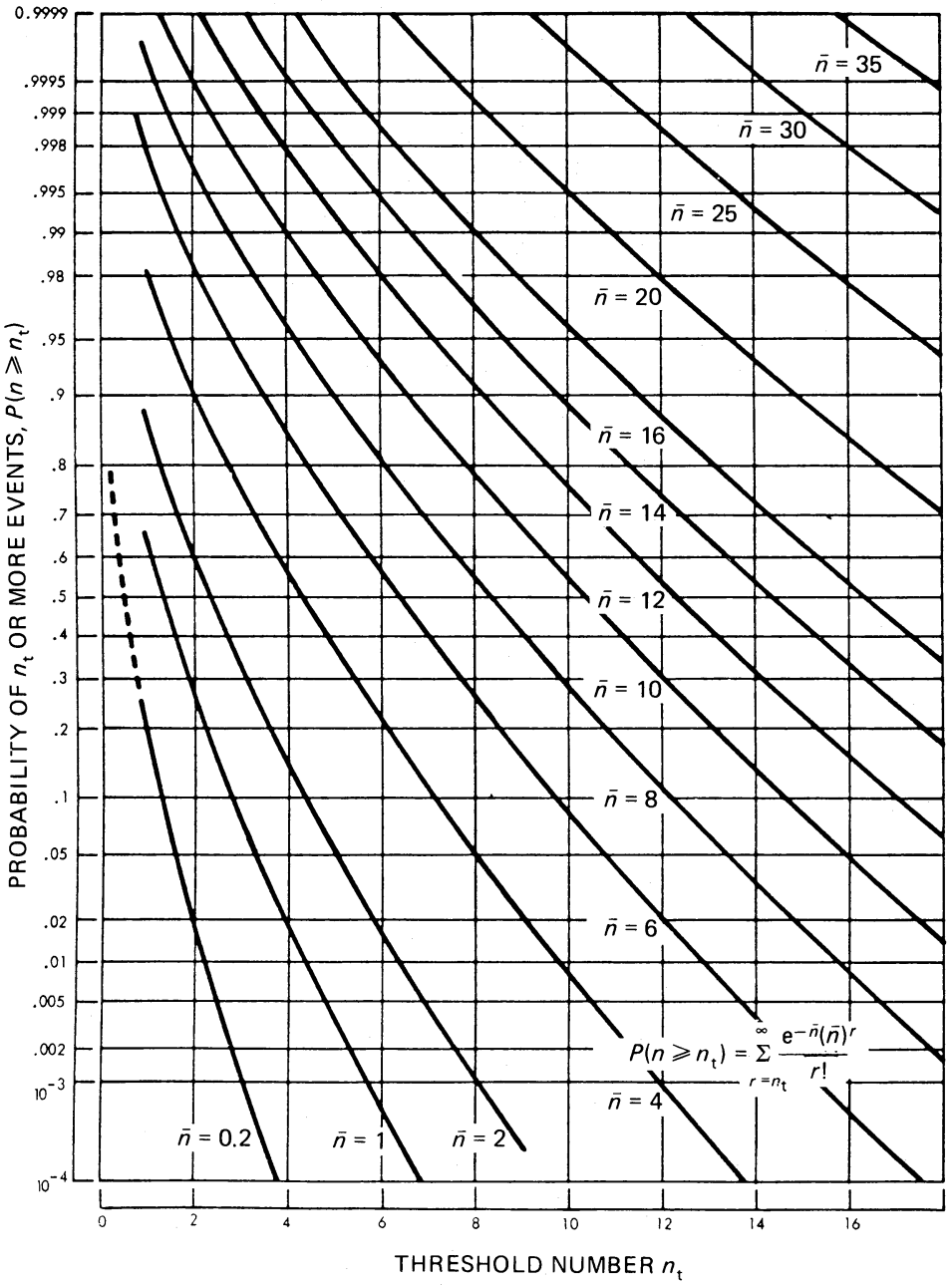
$$\sigma^2 = \text{variance} = \mu = \text{mean}$$

The probability of observing at least s events is:

$$P_p(x \geqslant s, \mu) = \sum_{x=s}^{\infty} \frac{\mu^x}{x!}\, e^{-\mu}$$

© Cambridge University Press • Provided by the NASA Astrophysics Data System

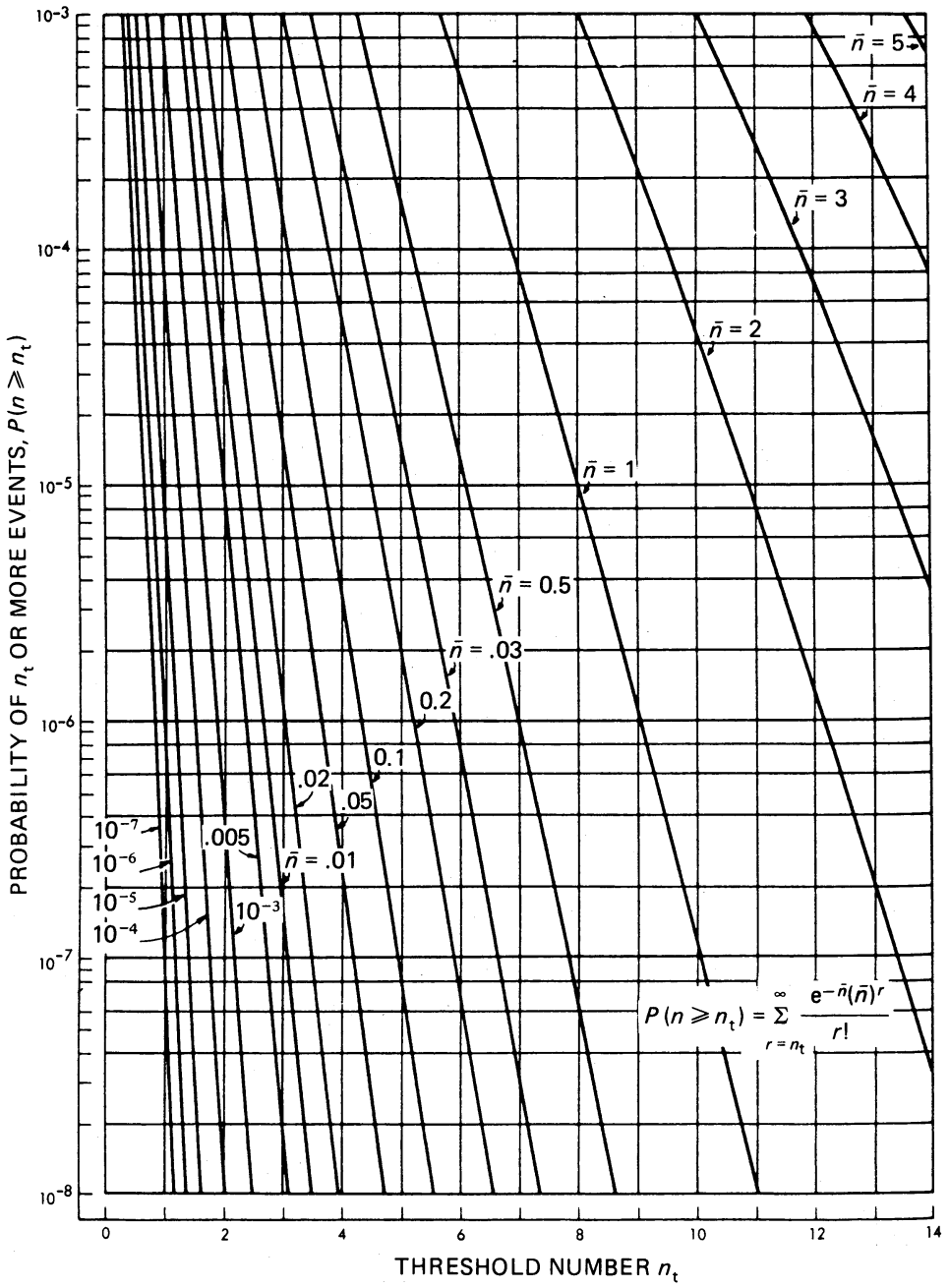
Probability of n_t or more random events with Poisson distribution when the expected or mean number of events is $\bar{n}$ as a function of the threshold number n_t. (From RCA Electro-Optics Handbook, 1974.)



© Cambridge University Press • Provided by the NASA Astrophysics Data System

Poisson distribution (*cont.*)

Probability of n_t or more random events with Poisson distribution when the expected or mean number of events is $\bar{n}$ as a function of the threshold number n_t. Curves for $10^{-7} < \bar{n} < 10^{-4}$ are approximate. (From *RCA Electro-Optics Handbook*, 1974.)



© Cambridge University Press • Provided by the NASA Astrophysics Data System

Chi-square distribution, confidence levels

The chi-square distribution for n_D degrees of freedom is given by:

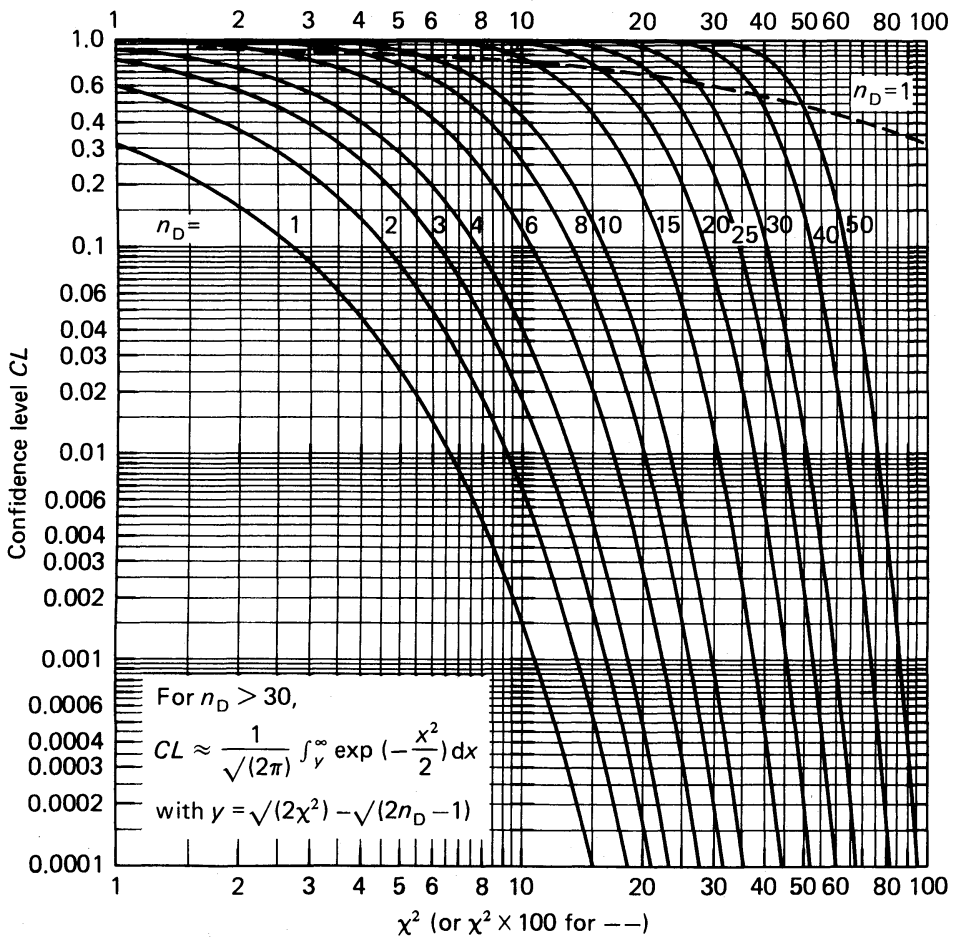
$$P_{n_D}(\chi^2)d\chi^2 = \frac{1}{2^h\Gamma(h)}(\chi^2)^{h-1}\,e^{-\chi^2/2}d\chi^2$$

where h (for 'half') $= n_D/2$.

The confidence level CL associated with a given value of n_D and an observed χ^2 is the probability of χ^2 exceeding the observed value:

$$CL = \int_{\chi^2}^{\infty} d\chi^2 P_{n_D}(\chi^2)$$

χ^2 confidence level vs. χ^2 for n_D degrees of freedom. (From *Review of Particle Properties*, Lawrence Berkeley Laboratory, University of California, Berkeley, CA, 1982.)



© Cambridge University Press • Provided by the NASA Astrophysics Data System

χ^2 test of distributions

$$\chi^2 = \sum_{j=1}^{n} \frac{[f(x_j) - NP(x_j)]^2}{NP(x_j)},$$

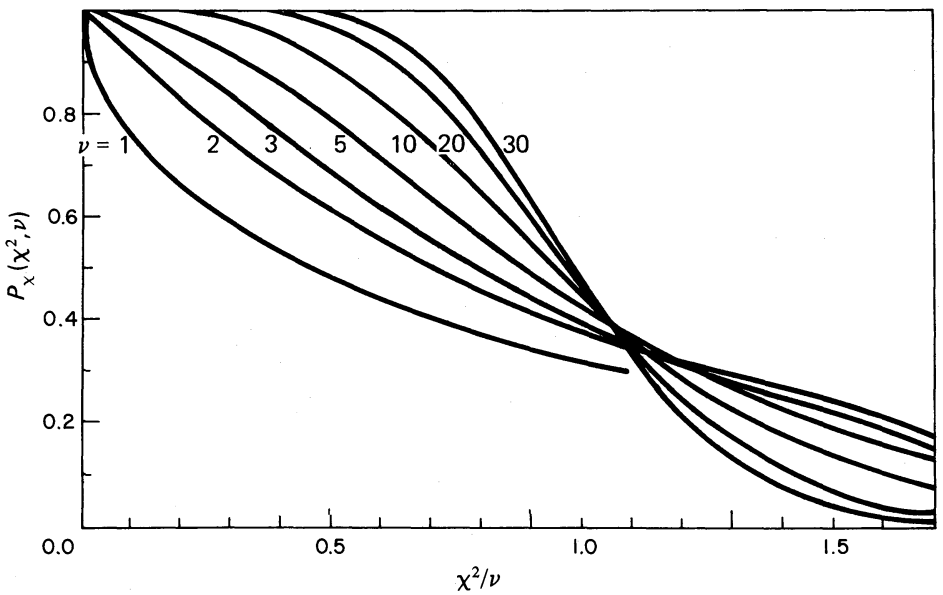
where

$f(x_j)$ = observed frequency distribution of possible observation x_j,
n = number of different values of x_j observed,
N = total number of measurements,
$P(x_j)$ = theoretical probability distribution.
$\chi^2_\nu = \chi^2/\nu$ (for χ^2 tests, χ^2_ν should be $\cong 1$),
ν = degrees of freedom = n − number of parameters calculated from the data to describe the distribution.

The probability $P_\chi(\chi^2, \nu)$ of exceeding χ^2 vs. the reduced chi-square $\chi^2_\nu = \chi^2/\nu$ and the number of degrees of freedom ν.

$$P_\chi(\chi^2, \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_{\chi^2}^{\infty} (x^2)^{1/2(\nu-2)} e^{-x^2/2} dx^2$$

(Adapted from Bevington, P. R., *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Book Company, 1969.)



© Cambridge University Press • Provided by the NASA Astrophysics Data System

Propagation of errors

Sample mean:

$$\bar{x} = \frac{1}{N}\sum x_i \simeq \mu, \quad \text{parent mean.}$$

Sample variance:

$$s^2 = \frac{1}{N-1}\sum (x_i - \bar{x})^2 \simeq \sigma^2, \quad \text{parent variance.}$$

For $x = f(u, v, \ldots)$

$\bar{x} = f(\bar{u}, \bar{v}, \ldots)$, the most probable value for x, and

$$\sigma_x^2 \simeq \sigma_u^2\left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2\left(\frac{\partial x}{\partial v}\right)^2 + 2\sigma_{uv}^2\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + \cdots,$$

where

$$\sigma_u^2 = \lim_{N\to\infty}\frac{1}{N}\sum (u_i - \bar{u})^2, \quad \sigma_v^2 = \lim_{N\to\infty}\frac{1}{N}\sum (v_i - \bar{v})^2$$

and

$$\sigma_{uv}^2 = \lim_{N\to\infty}\frac{1}{N}\sum [(u_i - \bar{u})(v_i - \bar{v})], \quad \text{the covariance.}$$

For u and v uncorrelated, $\sigma_{uv}^2 = 0$.

Least-squares fit to a straight line

Linear function:

$$y(x) = a + bx.$$

Chi-square:

$$\chi^2 = \sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx_i)^2\right],$$

σ_i = standard deviation of the observation y_i.

Least-squares fitting procedure:

minimize χ^2 with respect to each of the coefficients a and b simultaneously.

© Cambridge University Press • Provided by the NASA Astrophysics Data System

Coefficients of least-squares fit:

$$a = \frac{1}{\Delta} \cdot \left(\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} \right)$$

$$b = \frac{1}{\Delta} \cdot \left(\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} \right)$$

$$\Delta = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2$$

Estimated variance:

$$\sigma^2 \simeq s^2 = \frac{1}{N-2} \sum (y_i - a - bx_i)^2$$

s^2 = sample variance

Uncertainties in coefficients:

$$\sigma_a^2 \simeq \frac{1}{\Delta} \sum \frac{x_i^2}{\sigma_i^2} \qquad \sigma_b^2 \simeq \frac{1}{\Delta} \sum \frac{1}{\sigma_i^2}$$

Bibliography

Statistical Methods in Experimental Physics, Eadie, W. T. *et al.*, North-Holland Publishing Co., 1971.

Data Reduction and Error Analysis for the Physical Sciences, Bevington, P. R., McGraw-Hill Book Company, 1969.

Data Analysis for Scientists and Engineers, Meyer, S. L., Peer Management Consultants, Ltd., Evanston, IL, 1986.

© Cambridge University Press • Provided by the NASA Astrophysics Data System