



JAPANESE
INDUSTRIAL
STANDARD

Translated and Published by
Japanese Standards Association

JIS Z 8101-1 : 1999

**Statistics—
Vocabulary and symbols—
Part 1 : Probability and
general statistical terms**

ICS 01.040.03; 03.120.30

Descriptors : vocabulary, marks, symbols, probability theory, quality control

Reference number : JIS Z 8101-1 : 1999 (E)

Z 8101-1 : 1999

Foreword

This translation has been made based on the original Japanese Industrial Standard established by the Minister of International Trade and Industry through deliberations at the Japanese Industrial Standards Committee in accordance with the Industrial Standardization Law. By this establishment JIS Z 8101 : 1981 was withdrawn and replaced with this Standard. This Standard has been prepared based on the first edition issued in 1993 of ISO 3534-1.

JIS Z 8101 : 1999 consists of the following parts, under the general title *Statistics—Vocabulary and symbols*:

- Part 1 : *Probability and general statistical terms*
- Part 2 : *Statistical quality control terms*
- Part 3 : *Design of experiments*

Date of Establishment: 1999-05-20

Date of Public Notice in Official Gazette: 1999-05-20

Investigated by: Japanese Industrial Standards Committee
Divisional Council on Basic Items

JIS Z 8101-1:1999, First English edition published in 2000-07

Translated and published by: Japanese Standards Association
4-1-24, Akasaka, Minato-ku, Tokyo, 107-8440 JAPAN

In the event of any doubts arising as to the contents,
the original JIS is to be the final authority.

© JSA 2000

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

Printed in Japan

Contents

	Page
Introduction	1
Scope	1
1 Terms used in the theory of probability	1
2 General statistical terms	9
Annex A (normative) Symbols used in JIS Z 8101-1	21

Statistics—Vocabulary and symbols— Part 1 : Probability and general statistical terms

Introduction This Japanese Industrial Standard has been prepared based on ISO 3534-1, *Statistics—Vocabulary and symbols—Part 1 : Probability and general statistical terms* published in 1993 as the 1st edition. The Sections 3 and 4 in ISO 3534-1 are removed from this Standard because it has been decided that the technical contents of the Sections 3 and 4 in ISO 3534-1 : 1993 were to be transferred to Part 2, by ongoing policy of ISO/TC 69/SC 1/WG 2 with respect to the work of revision of ISO 3534-1 and 2.

Scope This Standard defines probability and general statistical terms which may be used in the drafting of Japanese Industrial Standards. In addition, it defines symbols for a limited number of these terms.

The terms are classified under the following main headings:

- Terms used in the theory of probability
- General statistical terms

1 Terms used in the theory of probability

1.1 probability

A limiting value of the relative frequency by which a specific result occurs when a long-run trial is continued under the same condition. In general, it is defined as a real number in the scale of $[0, 1]$ attached to a random event. Generally, probability of event A is denoted by $\Pr(A)$.

Informative reference : There is subjective probability which expresses a degree of belief that an event will occur.

1.2 random variable, variate

A variable that may take any of the values according to a probability rule. The probability rule is described by a probability distribution. A random variable that may take only isolated values is said to be a discrete (random) variable and a random variable which may take continuous values is said to be a continuous (random) variable. The data expressed with the discrete random variable is said to be **discrete variable** and the data expressed with the continuous random variable is said to be **continuous variable**.

1.3 probability distribution (of a random variable)

A probability that the random variable takes any given value, or a function giving the probability that a random variable belongs to a given set of values. The probability on the whole set of values of the random variable equals 1. In particular, when the probability that the random variable X takes a value not more than x is considered to be a function of x , it is said to be a **distribution function** and denoted with $F(x) = \Pr\{X \leq x\}$.

For each value x_i taken by the discrete variable, its probability is called **probability mass function** which is denoted with $p_i = \Pr\{X = x_i\}$.

In the case of continuous variable, the derivative of distribution function is called **probability density function** which is denoted with $f(x) = \frac{dF(x)}{dx}$.

$f(x) dx$ corresponds to the probability that the random variable X belongs to an infinitesimal section $[x, x + dx]$ and is said to be probability element.

Remarks : When the probability distribution of random variable X is given, it is said that “ X equals probability distribution” or “variable X which equals probability distribution”.

1.4 bivariate distribution function

A function giving, for every pair of probability variables (X, Y), the probability that X be less than or equal to x and Y be less than or equal to y . It is denoted with $F(x, y) = \Pr\{X \leq x, Y \leq y\}$. For the multivariate random variables ($X, Y, \dots$), **multivariate distribution function** $F(x, y, \dots) = \Pr\{X \leq x, Y \leq y, \dots\}$ is defined as well. The probability distribution of a set of two or more variables is called **joint probability distribution**.

1.5 marginal probability distribution

A joint probability distribution of a subset of k_1 random variables from a probability distribution of k random variables. For example, in the case of three random variables (X, Y, Z), there are three bivariate marginal probability distributions of (X, Y), (X, Z) and (Y, Z) and three univariate marginal probability distributions corresponding to X, Y and Z .

1.6 conditional probability distribution

A probability distribution of a subset of k_1 random variables from a probability distribution of k random variables when the other $(k - k_1)$ variables have fixed values. For example, in the case of probability distribution of two random variables (X, Y), there are conditional probability distribution of X which is fixed as $Y = y$ and conditional probability distribution Y which is fixed as $X = x$.

1.7 independence

Two random variables X and Y are independent if, and only if, their distribution functions are related by $F(x, y) = F(x, \infty) \cdot F(\infty, y) = G(x) \cdot H(y)$.

$G(x) = F(x, \infty)$ and $H(y) = F(\infty, y)$ are the marginal probability distribution functions of X and Y respectively.

Remarks 1 For discrete independent random variables, their joint probabilities are related by the product of the marginal probability distributions $\Pr\{X = x_i, Y = y_j\} = \Pr\{X = x_i\} \cdot \Pr\{Y = y_j\}$.

For continuous independent random variables, their joint density functions $f(x, y)$ are related by $f(x, y) = g(x) \cdot h(y)$, namely, the product of the marginal density functions $g(x)$ and $h(y)$.

2 Two events are independent if the probability that both occur is equal to the product of the probabilities of the two events.

1.8 parameter

- (1) When considering a group of population distribution $f(x; \theta_1, \theta_2, \dots, \theta_p)$, the constants $\theta_1, \theta_2, \dots, \theta_p$, that, if they are designated, a distribution is determined. For example, the normal distribution is determined by the two parameters of arithmetic mean μ and standard deviation σ , and the Poisson distribution is determined by one parameter of arithmetic mean μ .
- (2) In a wider sense, the numerical value determined by the probability distribution. In this sense, the moments of the probability distribution such as arithmetic means, dispersions, skewness and kurtosis are all said to be parameters. The parameter is a term to be distinguished from the statistic of the same denomination defined as sample.

1.9 correlation

The relationship between distribution rules of two random variables. In many cases, only the degree of linear relationship is indicated.

1.10 quantile, fractile

The p -quantile is the value of the random variable for which the distribution function equals p or jumps from a value less than p to a value greater than p . When the probability p is expressed with 100 %, it is called **100 p percentile**.

- Remarks 1 If the distribution function takes a specific value p in an interval of the random variables, then any value in this interval may be considered as the p -quantile, provided $0 \leq p \leq 1$.
- 2 The value of the random variable corresponding to $p = \frac{1}{2}$ is called **median**.
 - 3 The value of the random variable corresponding to $p = \frac{1}{4}$ and $p = \frac{3}{4}$ is called **quartile**.

1.11 mode

The value(s) of a random variable at the maximum value of the random variable in the case of discrete distribution and at the maximum value of the density function in the case of continuous distribution. In polymodal case, the value of a random variable which gives respective local maximum value(s).

1.12 expectation

For a discrete distribution, the value of $E(X) = \sum x_i p_i$ when the probability that the value x_i occurs is $p_i = \Pr\{X = x_i\}$. For the continuous distribution which has the probability density function $f(x)$, the value of $E(X) = \int x f(x) dx$. If carrying out many measurements to obtain the average of the measurements, it becomes the value near the expectation.

The expectation $E[g(X)]$ of the function $g(X)$ is given by $E[g(X)] = \sum g(x_i) p_i$ or $E[g(X)] = \int g(x) f(x) dx$.

- Remarks 1 The expectation of the conditional probability distribution is said to be **conditional expectation**. The conditional expectation of X given with $Y = y$ with regard to the joint distribution of X and Y becomes a function of y .

- 2 The expectation of the random variable X is also said to be population mean of X .

1.13 variance

The expectation of the square of the variable obtained by subtracting, from the random variable X , its population mean. $\sigma^2 = V(X) = E[X - E(X)]^2$.

Remarks 1 It is also said to be the variance of probability distribution that equals the random variable.

- 2 The variance of the random variable X is also said to be **population variance of X** .

1.14 standard deviation

The positive square root of the variance. $\sigma = \sqrt{V(X)}$.

Remarks : The standard deviation of the random variable X is also said to be **population standard deviation of X** .

1.15 coefficient of variation

The ratio of the standard deviation to the arithmetic mean. The coefficient of variation which relatively shows dispersion is used only in the case of a non-negative random variable. $\frac{\sqrt{V(X)}}{E(X)} = \frac{\sigma}{\mu}$.

1.16 standardized random variable

A random variable the expectation of which equals zero and the standard deviation of which equals 1.

Remarks 1 If the random variable X has an expectation equal to μ and a standard deviation equal to σ , the corresponding standardized random variable is $(X - \mu)/\sigma$.

- 2 The conversion of the random variable to zero and that of the standard deviation to 1 is said to be **standardization**.

1.17 moment of order q about the origin

The expectation $E(X^q)$ of the q th power of the random variable X .

Remarks : The arithmetic mean is the moment of order 1 about the origin.

1.18 central moment of order q

The expectation $E(X - \mu)^q$ of the q th power of the difference between X and μ when the expectation of the random variable X equals μ . It is also said to be moment of order q about the average value.

Remarks : The variance is the central moment of order 2.

1.19 skewness

A state where the probability density function or probability mass function $f(x)$ has a graph the left and the right of which are not symmetric. The degree of the skewness is the ratio of the moment of third order about the arithmetic mean μ_3 to the third power of the standard deviation σ . That is $\frac{\mu_3}{\sigma^3}$.

1.20 kurtosis

The ratio of the moment of fourth order about the arithmetic mean μ_4 to the fourth power of the standard deviation σ . That is $\frac{\mu_4}{\sigma^4}$.

1.21 covariance

For two dimensional random variables (X, Y) , the expectation of the product of the deviations from each arithmetic mean. $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$.

1.22 correlation coefficient

For two dimensional random variables (X, Y) , the ratio of the covariance to the product of their standard deviations. $\rho_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$.

Remarks 1 The suffix X and Y may be omitted if they are obvious.

2 The correlation coefficient will always lie between -1 and $+1$ inclusive.

3 If two random variables are independent, the correlation coefficient is zero. Even if the correlation coefficient is zero, two random variables are not necessarily independent.

1.23 regression curve

For two dimensional random variables (X, Y) , the function which gives the conditional expectation of Y when $X = x$ is given. This is called the regression curve of Y on X .

Remarks 1 When the regression curve of Y on X is a straight line, the regressive relation is said to be linear and called the **regression line**. In addition, this case is called the **simple linear regression**.

2 For three dimensional variables (X, Y, Z) , the conditional expectation of Z when $X = x$ and $Y = y$ are given is called the **regression surface** of Z on X and Y . When the regression surface is a plane, the regression is said to be linear and called the **regression plane surface**. This concept may be extended to four or more random variables.

1.24 uniform distribution, rectangular distribution

(1) For the continuous distribution, the distribution the probability density function of which is constant within a finite interval $[a, b]$ and zero outside this interval.

(2) For the discrete distribution, the distribution which has equal probability $\Pr(X = x_i) = \frac{1}{n}$, $i = 1, 2, \dots, n$ at n point.

1.25 normal distribution, Laplace-Gauss distribution

The distribution of a continuous random variable, the probability density function of which is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right], \quad -\infty < x < \infty.$$

Remarks : The normal distribution is determined by the arithmetic mean μ and the variance σ^2 . It is frequently denoted with the symbol $N(\mu, \sigma^2)$.