

# Meaning of concepts



**ASSIGN  
BUSTER**

**MEANING OF CONCEPTS RANDOM SAMPLING** Random sampling is a sampling technique where we select a group of  $s$  (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has a known, but possibly non-equal, chance of being included in the sample.

For example, within each group the choice of cases to constitute the sample should be based on chances selection. If the element of randomness is not introduced, bias is likely to enter and make the sample unrepresentative.

#### **PROBABILITY DISTRIBUTION**

A probability distribution describes the values and probabilities associated with a random event. The values must cover all of the possible outcomes of the event, while the total probabilities must sum to exactly 1, or 100%.

For example, a single coin flip can take values ' Heads or Tails' with a probability of exactly  $1/2$  for each; these two values and two probabilities make up the probability distribution of the single coin flipping event.

#### **CONTINUOUS PROBABILITY DISTRIBUTION**

A continuous distribution describes events over a continuous range, where the probability of a specific outcome is zero.

For example, a dart thrown at a dartboard has essentially zero probability of landing at a specific point, since a point is vanishingly small, but it has some probability of landing within a given area. The probability of landing within the small area of the bulls' eye would (hopefully) be greater than landing on an equivalent area elsewhere on the board. A smooth function that describes the probability of landing anywhere on the dartboard is the probability distribution of the dart throwing event. The integral of the probability density function (pdf) over the entire area of the dartboard (and, perhaps, the wall

surrounding it) must be equal to 1, since each dart must land somewhere.

### EXPECTED VALUE

Expected value is another word for ‘mean’ or ‘average’.

For example, the expected value of a six-sided die roll is 3.5. The concept is similar to the mean. The expected value of random variable  $X$  is typically written  $E(X)$  or  $\mu$  (mu).

### NORMAL DISTRIBUTION

The normal distribution, also called the Gaussian distribution. In it each member of the family may be defined by two parameters, location and scale: the mean (“average”,  $\mu$ ) and variance (standard deviation squared)  $\sigma^2$ , respectively.

For example, the sampling distribution of the sample mean is approximately normal, even if the distribution of the population from which the sample is taken is not normal.

### TYPE I ERROR

Type I error, also known as an “error of the first kind”, “ $\alpha$  error”, or a “false positive”, the error of rejecting a null hypothesis when it is actually true. It occurs when we are observing a difference when in truth there is none.

For example, a pregnancy test with a positive result (indicating that the woman taking the test is pregnant) has produced a false positive in the case where the woman is not pregnant.

### TYPE II ERROR

Type II error, also known as an “error of the second kind”, “ $\beta$  error”, or a “false negative”, the error of failing to reject a null hypothesis when the alternative hypothesis is the true state of nature. In other words, this is the error of failing to observe a difference when in truth there is one. This type of

error can only occur when the statistician fails to reject the null hypothesis. In the example of a pregnancy test, a type II error occurs if the test reports false when the woman is, in fact, pregnant.

#### STANDARD ERROR

The standard error of a method of measurement or estimation is the estimated standard deviation of the error in that method. Specifically, it estimates the standard deviation of the difference between the measured or estimated values and the true values.

#### DEGREE OF FREEDOM

The number of degrees of freedom generally refers to the number of independent observations in a sample minus the number of population parameters that must be estimated from sample data.

For example, the exact shape of a t distribution is determined by its degrees of freedom. When the t distribution is used to compute a confidence interval for a mean score, one population parameter (the mean) is estimated from sample data. Therefore, the number of degrees of freedom is equal to the sample size minus one.

#### NULL HYPOTHESIS

A null hypothesis ( $H_0$ ) is a hypothesis set up to be nullified or refuted in order to support an alternative hypothesis. When used, the null hypothesis is presumed true until statistical evidence, in the form of a hypothesis test, indicates otherwise.

For example, one may want to compare the test scores of two random samples of men and women, and ask whether or not one population has a mean score different from the other. A null hypothesis would be that the mean score of the male population was the same as the mean score of the

female population.

#### ALTERNATIVE HYPOTHESIS

The alternative hypothesis (or maintained hypothesis or research hypothesis) is the possibility that an observed effect is genuine and the null hypothesis is the rival possibility that it has resulted from random chance. For example: In the trial of Sally Clark, a solicitor accused of killing both her babies, pediatrician Sir Roy Meadow testified that the probability of two infants in the same family dying of natural causes was 1 in 73, 000, 000. If natural death is the null hypothesis and murders the alternative hypothesis, then the p-value is 1/73, 000, 000. The smallness of this value means that the null hypothesis that the deaths had had natural causes should be rejected and therefore murder concluded.

#### POWER OF THE TEST

The power of a statistical test is the probability that the test will reject a false null hypothesis (that it will not make a Type II error). As power increases, the chances of a Type II error decrease. Therefore power is equal to  $1 - \beta$ . For example, to test the null hypothesis that the mean scores of men and women on a test do not differ, samples of men and women are drawn, the test is administered to them, and the mean score of one group is compared to that of the other group using a statistical test. The power of the test is the probability that the test will find a statistically significant difference between men and women, as a function of the size of the true difference between those two populations.

#### T - TEST

A t-test is any statistical hypothesis test in which the test statistic has a Student's t distribution if the null hypothesis is true. It is applied when sample

sizes are small enough that using an assumption of normality and the associated z-test leads to incorrect inference.

#### CHI SQUARE DISTRIBUTION

This is not a test, but a distribution. The Chi-square distribution (chi-squared or  $\chi^2$  distribution) is derived from the Normal distribution. It is the distribution of a sum of squared Normal distributed variables. The best-known situations in which the chi-square distributions are used are the common chi-square tests for goodness of fit of an observed distribution to a theoretical one, and of the independence of two criteria of classification of qualitative data.

Describe the normal percentages of distribution in Normal Curve

The normal distribution and standard deviations:-

Approximately 68% of scores will fall within one standard deviation of the mean. It should be memorized is between Z-scores of -1.00 and +1.00.

Approximately 95% of scores will fall within two standard deviations of the mean. It should be memorized is between Z-scores of -2.00 and +2.00.

Approximately 99% of scores will fall within three standard deviations of the mean. It should be memorized is between Z-scores of -3.00 and +3.00.