# PROBABILITY AND SAMPLING DISTRIBUTIONS

**Seema Jaggi and P.K. Batra**
**I.A.S.R.I., Library Avenue, New Delhi - 110 012**
**pkbatra@iasri.res.in**

The concept of probability plays an important role in all problems of science and every day life that involves an element of uncertainty. **Probabilities** are defined as relative frequencies, and to be more exact as limits of relative frequencies. The relative frequency is nothing but the proportion of time an event takes place in the long run.

When an experiment is conducted, such as tossing coins, rolling a die, sampling for estimating the proportion of defective units, several outcomes or events occur with certain probabilities. These events or outcomes may be regarded as a variable which takes different values and each value is associated with a probability. The values of this variable depends on chance or probability. Such a variable is called a **random variable**. Random variables which take a finite number of values or to be more specific those which do not take all values in any particular range are called **discrete** random variables. For example, when 20 coins are tossed, the number of heads obtained is a discrete random variable and it takes values 0,1,...,20. These are finite number of values and in this range, the variable does not take values such as 2.8, 5.7 or any number other than a whole number. In contrast to discrete variable, a variable is **continuous** if it can assume all values of a continuous scale. Measurements of time, length and temperature are on a continuous scale and these may be regarded as examples of continuous variables. A basic difference between these two types of variables is that for a discrete variable, the probability of it taking any particular value is defined. For continuous variable, the probability is defined only for an interval or range. The frequency distribution of a discrete random variable is graphically represented as a histogram, and the areas of the rectangles are proportional to the class frequencies. In continuous variable, the frequency distribution is represented as a smooth curve. Frequency distributions are broadly classified under two heads:

1. Observed frequency distributions and
2. Theoretical or Expected frequency distributions

Observed frequency distributions are based on observations and experimentation. As distinguished from this type of distribution which is based on actual observation, it is possible to deduce mathematically what the frequency distributions of certain populations should be. Such distributions as are expected from on the basis of previous experience or theoretical considerations are known as **theoretical distributions** or **probability distributions**. Probability distributions consist of mutually exclusive and exhaustive compilation of all random events that can occur for a particular process and the probability of each event's occurring. It is a mathematical model that represents the distributions of the universe obtained either from a theoretical population or from the actual world, the distribution shows the results we would obtain if we took many probability samples and computed the statistics for each sample. A table listing all possible values that a random variable can take on together with the associated probabilities is called a probability distribution.

The probability distribution of X, where X is the number of spots showing when a six-sided symmetric die is rolled is given below:

| X | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| f(X) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

The probabilty distribution is the outcome of the different probabilities taken by the function of the random variable X.

Knowledge of the expected behaviour of a phenomenon or the expected frequency distribution is of great help in a large number of problems in practical life. They serve as benchmarks against which to compare observed distributions and act as substitute for actual distributions when the latter are costly to obtain or cannot be obtained at all. We now introduce a few discrete and continuous probability distributions that have proved particularly useful as models for real-life phenomena. In every case the distribution will be specified by presenting the probability function of the random variable.

## 1. Discrete Probability Distributions

### 1.1 Uniform Distribution

A uniform distribution is one for which the probability of occurrence is the same for all values of X. It is sometimes called a rectangular distribution. For example, if a fair die is thrown, the probability of obtaining any one of the six possible outcomes is 1/6. Since all outcomes are equally probable, the distribution is uniform.

**Definition:** If the random variable X assumes the values $x_1,x_2,...,x_k$ with equal probabilities, then the discrete uniform distribution is given by

$$P(X=x_i) = 1/k \quad \text{for } i = 1,2,...,k$$

**Example 1:** Suppose that a plant is selected at random from a plot of 10 plants to record the height. Each plant has the same probability 1/10 of being selected. If we assume that the plants have been numbered in some way from 1 to 10, the distribution is uniform with f(x;10) = 1/10 for x=1,...,10,.

### 1.2 Binomial Distribution

Binomial distribution is a probability distribution expressing the probability of one set of dichotomous alternatives i.e. success or failure. More precisely, the binomial distribution refers to a sequence of events which posses the following properties:

1. An experiment is performed under same conditions for a fixed number of trials say, n.
2. In each trial, there are only two possible outcomes of the experiment 'success' or 'failure'.
3. The probability of a success denoted by p remains constant from trial to trial.
4. The trials are independent i.e. the outcomes of any trial or sequence of trials do not affect the outcomes of subsequent trials.

Consider a sequence of n independent trials. If we are interested in the probability of x successes from n trials, then we get a binomial distribution where x takes the values from 0,1,…,n.

**Definition:** A random variable X is said to follow a binomial distribution with parameters n and p if its probability function is given by

$$P[X=x] = \binom{n}{x} p^x q^{n-x}, \; x = 0,1,...,n, \; 0<p<1.$$

The probability of success are the successive terms of the binomial expansion $(q+p)^n$. The probable frequencies of the various outcomes in $N$ sets of $n$ trials are $N(q+p)^n$. The frequencies obtained by this expression are known as expected or theoretical frequencies. On the other hand, the frequencies actually obtained by making experiments are called observed frequencies. Generally, there is some difference between the observed and expected frequencies but the difference becomes smaller and smaller as N increases.

**Constants of the binomial distribution**

The various constants of the binomial distribution are as follows:

Mean = $np$, Variance = $npq$. Here mean > variance

First moment $\mu_1 = 0$, Second moment $\mu_2 = npq$

Third moment $\mu_3 = npq(q-p)$, Fourth moment $\mu_4 = 3n^2p^2q^2 + npq\,(1-6pq)$

$$\beta_1 = \frac{(q-p)^2}{npq}, \; \gamma_1 = \frac{q-p}{\sqrt{npq}}; \qquad \beta_2 = 3 + \frac{1-6pq}{npq}, \; \gamma_2 = \frac{1-6pq}{npq}$$

**Properties of the binomial distribution**

1. The shape and location of the binomial distribution changes as p changes for a given n or as n changes for a given p. As p increases for a fixed n, the binomial distribution shifts to the right.
2. The mode of the binomial distribution is equal to the value of x which has the largest probability. The mean and mode are equal if np is an integer.
3. As n increase for a fixed p, the binomial distribution moves to right, flattens and spreads out. When p and q are equal, the distribution is symmetrical, for p and q may be interchanged without altering the value of any term, and consequently terms equidistant from the two ends of the series are equal. If p and q are unequal, the distribution is skewed. If p is less than 1/2, the distribution is positively skewed and when p is more than 1/2, the distribution is negatively skewed.
4. If n is large and if neither p nor q is too close to zero, the binomial distribution can be closely approximated by a normal distribution with standardized variable given by $Z = \frac{X - np}{\sqrt{npq}}$.

**Importance of the binomial distribution**

The binomial probability distribution is a discrete probability distribution that is useful in describing an enormous variety of real life events. For example, an experimenter wants to know the probability of obtaining diseased trees in a random sample of 10 trees if 10 percent of the trees are diseased. The answer can be obtained from the binomial probability distribution. The binomial distribution can be used to know the distribution of the number of seeds germinated out of a lot of seeds sown.

**Example 2:** The incidence of disease in a forest is such that 20% of the trees in the forest have the chance of being infected. What is the probability that out of six trees selected, 4 or more will have the symptoms of the disease?

**Solution:** The probability of a tree having being infected is

$$p = \frac{20}{100} = \frac{1}{5}$$

and the probability of not being infected = $1 - \frac{1}{5} = \frac{4}{5}$

Hence the probability of 4 or more trees being infected out of 6 will be

$$= \binom{6}{4}\left(\frac{1}{5}\right)^4 + \binom{6}{5}\left(\frac{1}{5}\right)^5\binom{4}{5} + \binom{6}{6}\left(\frac{1}{5}\right)^6\binom{4}{5}^6 = \frac{53}{3125}$$

**Fitting a binomial distribution**

When a binomial distribution is to be fitted to the observed data, the following procedure is adopted:

1. Evaluate mean of the given distribution and then determine the values of p and q. Expand the binomial $(q+p)^n$. The number of terms in the expanded binomial is equal to one more than n.
2. Multiply each term of the expanded binomial by N (the total frequency) for obtaining the expected frequency in each category.

**Exercise:** The following data shows the number of seeds germinating out of 10 on damp filter for 80 sets of seeds. Fit a binomial distribution to this data.

| X: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| f: | 6 | 20 | 28 | 12 | 8 | 6 | 0 | 0 | 0 | 0 | 0 |

**Step 1:** Calculate $\overline{X} = \frac{\sum fX}{\sum f}$

**Step 2:** Find p and q using mean = $np$.

**Step 3:** Expand the binomial $80(q+p)^{10}$ and find expected frequencies.

The generalization of the binomial distribution is the **multinomial distribution**. Whereas in case of binomial distribution, there are only two possible outcomes on each experimental trial, in the multinomial distribution there are more than two possible outcomes on each trial. The assumptions underlying the multinomial distribution are analogous to the binomial distribution. These are:

1. An experiment is performed under the same conditions for a fixed number of trials, say, n.
2. There are k outcomes of the experiment which may be referred to $e_1$, $e_2$, $e_3$,...,$e_k$. Thus the sample space of possible outcomes on each trial shall be:
   $S = \{ e_1, e_2, e_3,...,e_k\}$
3. The respective probabilities of the various outcomes i.e., $e_1$, $e_2$, $e_3$,...,$e_k$ denoted by $p_1$, $p_2$, $p_3$,...,$p_k$ respectively remain constant from trial to trial.
   $p_1 + p_2 + p_3 + ... + p_k = 1$
4. The trials are independent.

## 1.3 Poisson Distribution

Poisson distribution is a discrete probability distribution and is very widely used in statistical work. This distribution is the limiting form of the binomial distribution as n becomes infinitely large and p approaches to zero in such a way that np=λ remains constant. A Poisson distribution may be expected in cases where the change of any individual event being a success is small. The distribution is used to describe the behaviour of rare events.

**Definition:** A random variable X is said to follow a Poisson distribution with parameter λ if the probability function is given by

$$P[X=x] = \frac{e^{-\lambda}\,\lambda^{x}}{x!},\ x = 0,1,... \qquad \text{where } e = 2.7183.$$

**Constants of the Poisson distribution**

The various constants of the Poisson distribution are

Mean = $\lambda$, Variance = $\lambda$ , [Here mean = variance]

First moment $\mu_1 = 0$, Second moment $\mu_2 = \lambda$

Third moment $\mu_3 = \lambda$, Fourth moment $\mu_4 = \lambda + 3\lambda^2$

$$\beta_1 = \frac{1}{\lambda}, \quad \gamma_1 = \frac{1}{\sqrt{\lambda}}, \quad \beta_2 = 3 + \frac{1}{\lambda}, \quad \gamma_2 = \frac{1}{\lambda}$$

**Properties of the Poisson distribution**

1. As λ increases, the distribution shifts to the right, i.e. the distribution is always a skewed distribution.
2. Mode: When $\lambda$ is not an integer then unique mode i.e. m=[$\lambda$].
   When $\lambda$ is an integer then bimodal i.e. m=$\lambda$ and m=$\lambda$-1.
3. Poisson distribution tends to normal distribution as $\lambda$ becomes large.

**Importance of the Poisson distribution**

In general, the Poisson distribution explains the behaviour of discrete variates where the probability of occurrence of the event is small and total number of possible cases is sufficiently large. For example, it is used in quality control statistics to count the number of defects of an item, or in biology to count the number of bacteria, or in physics to count the number of particles emitted from a radioactive substance, or in insurance problems to count the number of casualties etc. The Poisson distribution is also used in problems dealing with the inspection of manufactured products with the probability that any piece is defective is very small and the lots are very large. Also used to know the probability of mutations in a DNA segment.

Note also that the only variable needed to generate these distributions is λ, the average occurrence/interval. Moreover, in biology situations often occur where knowing the probability of no events P(0) in an interval is useful. When x = 0, equation simplifies to:

$$P(0) = e^{-\lambda}$$

For example, we might want to know the fraction of uninfected cells for a known average ($\lambda$) multiplicity of virus infection (MOI). Other times, we need to know the average mutation

rate/base pair, but our sequencing determines nearly all wild type sequence, P(0). In each case, if we can determine either $\lambda$ or P(0), we can solve for the other.

**The Standard Deviation (SD):** The uncertainty (expressed as ±1 SD) in the measurement of a number of random events equals the square root of the total number of events i.e.

$$SD = \sqrt{Total\ \ Events}$$

We use radioactive decay and its detection to illustrate this feature of the Poisson distribution for two reasons. Most biologists have some experience with radioactivity measurements; more important, radioactive decay is a true random process. In fact, it is the only truly random process known in nature. For this latter reason, we can make confident predictions about its behavior.

Suppose we have a radioactive sample that registers about 1000 cpm. We want to report our measurement along with an uncertainty expressed as a standard deviation (SD). We could count the sample 10 times for one minute each and then calculate the mean and SD of the 10 determinations. However, the important property of processes described by the Poisson distribution is that the SD is the square root of the total counts registered To illustrate, the table shows the results of counting our radioactive sample for different time intervals (with some artificial variability thrown in).

| **Time (min)** | **Total Counts** | **SD (counts)** | **Reported cpm** | **SD (in cpm)** | **Relative Error as %** |
|---|---|---|---|---|---|
| **0.1** | 98 | 10 | 980 | 100 | 10 |
| **1.0** | 1,020 | 32 | 1020 | 32 | 3 |
| **10** | 9,980 | 100 | 998 | 10 | 1 |
| **100** | 101,000 | 318 | 1010 | 3 | 0.3 |

Reported cpm is Total Counts/Time; SD (in cpm) is SD (counts)/Time;
and Relative Error is SD (in cpm)/Reported cpm, expressed as %.

Comparing every other line shows that a 100-fold increase in counting time increases SD, but only by 10-fold. At the same time, the relative error decreases by 10-fold. The general point here is that the experimenter can report the 1000 cpm value to any degree of precision desired simply by choosing the appropriate time interval for measurement. There is no advantage whatever in using multiple counting periods. Thus, counting error is distinguished from experimental error in that the latter can only be estimated with multiple measurements.

**Fitting a Poisson distribution**

The process of fitting a Poisson distribution involves obtaining the value of λ, i.e., the average occurrence, and to calculate the frequency of 0 success. The other frequencies can be very easily calculated as follows:

$$N(P_0) = Ne^{-\lambda};\quad N(P_1) = N(P_0) \times \frac{\lambda}{1};\quad N(P_2) = N(P_1) \times \frac{\lambda}{2};\quad N(P_3) = N(P_2) \times \frac{\lambda}{3}, \text{ etc.}$$

**Exercise:** The following mutated DNA segments were observed in 325 individuals:

| Mutated DNA segments | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Number of individuals | 211 | 90 | 19 | 5 | 0 |

Fit a Poisson distribution to the data.

**Step 1:** Calculate the mean
**Step 2:** Find the different terms $N(P_0)$, $N(P_1)$,... i.e. the expected frequencies.

### 1.4 Negative Binomial Distribution

The negative binomial distribution is very much similar to the binomial probability model. It is applicable when the following conditions hold good:

1. An experiment is performed under the same conditions till a fixed number of successes, say c, are achieved.
2. In each trial, there are only two possible outcomes of the experiment 'success' or 'failure'
3. The probability of a success denoted by p remains constant from trial to trial.
4. The trials are independent i.e. the outcome of any trial or sequence of trials do not affect the outcomes of subsequent trials.

The only difference between the binomial model and the negative binomial model is about the first condition.

Consider a sequence of Bernoulli trials with p as the probability of success. In the sequence, success and failure will occur randomly and in each trial the probability of success will be p. Let us investigate how much time will be taken to reach the $r^{th}$ success. Here r is fixed, let the number of failures preceding the $r^{th}$ success be x (=0,1,...). The total number of trials to be performed to reach the $r^{th}$ success will be x+r. Then the probability that $r^{th}$ success occurs at $(x+r)^{th}$ trial is

$$P(X=x) = \binom{x+r-1}{r-1} p^r q^x \; ; \quad x=0,1,2,....$$

**Example 3:** Suppose that 30% of the items taken from the end of a production line are defective. If the items taken from the line are checked until 6 defective items are found, what is the probability that 12 items are examined?

**Solution:** Suppose the occurrence of a defective item is a success. Then the probability that there will be 6 failures preceding the 6th success will be given by:

$$\binom{6+6-1}{6-1} (.30)^6 (.70)^6 = 0.0396.$$

**Note:** If we take r=1, i.e. the first success, then $P[X=x] = pq^x$, x=0,1,2,... which is the probability distribution of X, the number of failures preceding the first success. This distribution is called as **Geometric distribution**.

### 1.5 Hypergeometric Distribution:

The hypergeometric distribution occupies a place of great significance in statistical theory. It applies to sampling without replacement from a finite population whose elements can be classified into two categories - one which possess a certain characteristic and another which does not possess that characteristic. The categories could be male, female, employed unemployed etc. When n random selections are made without replacement from the population, each subsequent draw is dependent and the probability of success changes on each draw. The following conditions characterise the hypergeometric distribution:

1. The result of each draw can be classified into one of the two categories.
2. The probability of a success changes on each draw.
3. Successive draws are dependent.
4. The drawing is repeated a fixed number of times.

The hypergeometric distribution which gives the probability of r successes in a random sample of n elements drawn without replacement is;

$$P(r) = \frac{\binom{N-X}{n-r}\binom{X}{r}}{\binom{N}{n}} \quad \text{for } r=0,1,2...,[n,X]$$

The symbol [n, X] means the smaller of n or X. This distribution may be used to estimate the number of wild animals in forests or to estimate the number of fish in a lake. The hypergeometric distribution bears a very interesting relationship to the binomial distribution. When N increases without limit, the hypergeometric distribution approaches the binomial distribution. Hence, the binomial probabilities may be used as approximation to hypergeometric probabilities when n/N is small.

## 2. Continuous Probability Distributions

### 2.1 Normal Distribution

The normal distribution is "probably" the most important distribution in Statistics. It is a probability distribution of a continuous random variable and is often used to model the distribution of discrete random variable as well as the distribution of other continuous random variables. The basic form of normal distribution is that of a bell, it has single mode and is symmetric about its central values. The flexibility of using normal distribution is due to the fact that the curve may be centered over any number on the real line and it may be made flat or peaked to correspond to the amount of dispersion in the values of random variable. The versatility in using the normal distribution as probability distribution model is depicted in Fig. i.

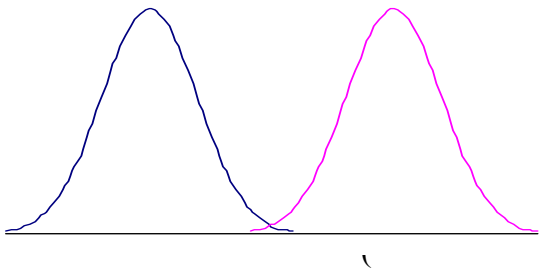


Fig.-i

Many quantitative characteristics have distribution similar in form to the normal distribution's bell shape. For example height and weight of people, the IQ of people, height of trees, length of leaves etc. are typically the type of measurements that produces a random variable that can be successfully approximated by normal random variable. The values of random variables are produced by a measuring process and measurements tend to cluster symmetrically about a central value.

**Definition**: A random variate X, with mean μ and variance $\sigma^{2,}$ is said to have normal distribution, if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad ; \qquad -\infty < x < \infty;\ -\infty < \mu < \infty;\ \sigma > 0$$

Here $\pi$ is a mathematical constant equal to 3.14156.

**Standard Normal Distribution**: If X is a normal random variable with mean μ and standard deviation σ, then $\frac{X-\mu}{\sigma}$ is a standard normal variate with zero mean and standard deviation 1. The probability density function of standard normal variable Z is

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

**Area under the normal Curve**

For normal variable X,

P( a < X < b) = Area under y = f (x) from X = a to X = b as shown in Fig. (ii).

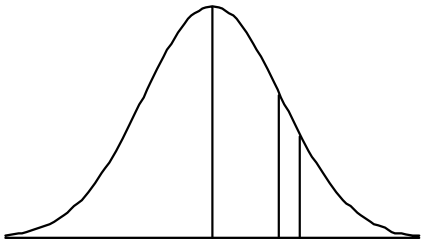
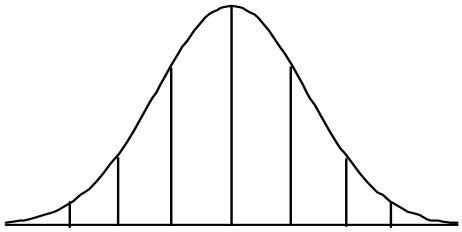


Fig.-ii: Area representing P[a <X <b] for a normal random variable

The probability that X is between a and b (b > a) can be determined by computing the probability that Z is between (a - μ) / σ and (b - μ) / σ. It is possible to determine the area in Fig- ii by using tables (for Areas under normal curve) rather then by performing any mathematical computations. Probability associated with a normal random variable X can be determined from Table-1 given at the end.



μ -3σ μ -2σ μ -σ μ μ +σ μ +2σ μ +3σ

Fig. iii

As indicated in Fig.(iii) for any normal distribution, 68.26% of the Z values lie within one standard deviation of mean, 95.44% of values lie within 2 standard deviations of mean and 99.73% of values lie within three standard deviations of mean. By using the fact that the normal distribution is symmetric about its mean (zero in this case) and the total area under curve is 1 (half to the left of zero and half to right), probabilities of standard normal variable

of the form $P = \int_a^{\infty} zdz$ = P (Z>a) are provided in Table 1 in the end. Using table 1 probabilities that Z lies in any interval on real line may be determined.

**Properties of normal distribution**

1. The normal curve is symmetrical about the mean x=μ
2. The height of normal curve is at its maximum at the mean. Hence the mean and mode of normal distribution coincides. Also the number of observations below the mean in a normal distribution is equal to the number of observations above the mean. Hence mean and median of normal distribution coincides. Thus for normal distribution mean = median = mode.
3. The normal curve is unimodal at x =μ
4. The point of inflexion occurs at μ ± σ
5. The first and third quartiles are equidistant from the median.
6. The area under normal curve is distributed as follows
   (a) μ - σ and μ + σ covers 68.26% of area
   (b) μ - 2σ and μ + 2σ covers 95.44% of area
   (c) μ - 3σ and μ + 3σ covers 99.73% of area

**Importance of normal distribution**

1. Of all the theoretical distributions, the normal distribution is the most important and is widely used in statistical theory and work. The most important use of normal distribution is in connection with generalization from a limited number of individuals observed on individuals that have not been observed. It is because of this reason that the normal distribution is the core heart of sampling theory. The distribution of statistical measures such as mean or standard deviation tends to be normal when the sample size is large. Therefore, inferences are made about the nature of population from sample studies.
2. The normal distribution may be used to approximate many kinds of natural phenomenon such as length of leaves, length of bones in mammals, height of adult males, intelligence quotient or tree diameters. For example, in a large group of adult males belonging to the same race and living under same conditions, the distribution of heights closely resembles the normal distribution.
3. For certain variables the nature of the distribution is not known. For the study of such variables, it is easy to scale the variables in such a way as to produce a normal distribution. It is indispensable in mental test study. It is reasonable to assume that a selected group of children of a given age would show a normal distribution of intelligence test scores.

**Exercises**

1. The average rainfall in a certain town is 50 cm with a standard deviation of 15 cm. Find the probability that in a year the rainfall in that town will be between 75 and 85 cm.
2. The average fuelwood value in Kcal/kg of subabul plant is found to be 4,800 with standard deviation of 240. Find the probability that the subabul plant selected at random has fuelwood value greater than 5,280 Kcal/Kg.

## 3. Sampling Distributions

The word population or universe in Statistics is used to refer to any collection of individuals or of their attributes or of results of operations which can be numerically specified. Thus, we may speak of the populations of weights, heights of trees, prices of wheat, etc. A population with finite number of individuals or members is called a finite population. For instance, the population of ages of twenty boys in a class is an example of finite population. A population with infinite number of members is known as infinite population. The population of pressures at various points in the atmosphere is an example of infinite population. A part or small section selected from the population is called a sample and the process of such selection is called sampling. Sampling is resorted to when either it is impossible to enumerate the whole population or when it is too costly to enumerate in terms of time and money or when the uncertainty inherent in sampling is more than compensated by the possibilities of errors in complete enumeration. To serve a useful purpose sampling should be unbiased and representative.

The aim of the theory of sampling is to get as much information as possible, ideally the whole of the information about the population from which the sample has been drawn. In particular, given the form of the parent population we would like to estimate the parameters of the population or specify the limits within which the population parameters are expected to lie with a specified degree of confidence. It is, however, to be clearly understood that the logic of the theory of sampling is the logic of induction, that is we pass from particular (i.e., sample) to general (i.e., population) and hence all results will have to be expressed in terms of probability. The fundamental assumption underlying most of the theory of sampling is random sampling which consists in selecting the individuals from the population in such a way that each individual of the population has the same chance of being selected.

**Population and Sample Statistics**

**Definition:** In a finite population of N values $X_1, X_2, ..., X_N$; of a population of characteristic X, the population mean ($\mu$) is defined as

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$$

and population standard deviation($\sigma$) is defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2}$$

**Definition:** If a sample of n values $x_1, x_2, ..., x_n$ is taken from a population set of values, the sample mean ($\bar{x}$) is defined as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and, the sample standard deviation (s) is defined as

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

**Sampling distribution of sample mean**

When different random samples are drawn and sample mean or standard deviation is computed, in general the computed statistics will not be same for all samples. Consider artificial example, where the population has four units 1,2,3,4 possessing the values 2,3,4,6 for the study variable. Then we will have 6 possible samples, if units are drawn without replacement. The possible samples with sample mean are given below:

**Different possible samples of size 2 without replacement**

| S.No. | Possible samples | Sample mean |
|---|---|---|
| 1 | (1, 2) | 2.5 |
| 2 | (1, 3) | 3.0 |
| 3 | (1, 4) | 4.0 |
| 4 | (2, 3) | 3.5 |
| 5 | (2, 4) | 4.5 |
| 6 | (3, 4) | 5.0 |

Though sample means are not the same from sample to sample, the average of sample means is 3.75 which is the same as population mean. The variance of sample means is 0.73.

**Theorem**: If a random variable X is normally distributed with mean $\mu$ and standard deviation $\sigma$, and a simple random sample of size n has been drawn, then the sample average is normally distributed (for all sample sizes n) with a mean $\mu$ and standard deviation $\sigma/\sqrt{n}$.

**Central limit theorem**: Let $x_1, x_2, ..., x_n$ be a simple random sample of size n drawn from an infinite population with a finite mean $\mu$ and standard deviation $\sigma$. Then random variable has a limiting distribution that is normal with a mean $\mu$ and standard deviation $\sigma/\sqrt{n}$.

### 3.1 Chi-Square Distribution

**Definition:** A random variable X is said to have $\chi^2$ distribution with v degrees of freedom if its probability density function (p.d.f.) is

$$f(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)} e^{-x/2} x^{(n/2)-1},\ 0 \le x < \infty$$

If samples of size n are drawn repeatedly from a normal population with variance $\sigma^2$, and the sample variance $s^2$ is computed for each sample, we obtain the value of a statistic $\chi^2$. The distribution of the random variable $\chi^2$, called chi-square, defined by

$$\chi^2 = (n - 1)\, s^2 / \sigma^2$$

is referred to as $\chi^2$ distribution with n-1 degrees of freedom(v). The mean of $\chi^2$ distribution equals the number of degree of freedom. The variance is twice its mean. Mode is v-2.

Let $\alpha$ be positive probability and let X have a chi-square distribution with v degrees of freedom, Then $\chi^2_\alpha(v)$ is a number such that

$$P[X \ge \chi^2_\alpha(v)] = \alpha$$

Thus $\chi^2_{\alpha}(v)$ is $100(1-\alpha)$.. percentile or upper $100\alpha$. percent point of chi-square distribution with v degrees of freedom. Then $100\alpha$. percentile point is the number $\chi^2_{(1-\alpha)}(v)$ such that $P[X \le \chi^2_{1-\alpha}(v)] = 1-\alpha$, i.e. the probability to right of $\chi^2_{1-\alpha}(v)$ is $1-\alpha$.

**Properties of $\chi^2$ variate**

1. Sum of independent $\chi^2$-variates is a $\chi^2$-variate.
2. $\chi^2$ distribution tends to normal distribution as v is large.

Table 2 gives values of $\chi^2_{\alpha}(v)$ for various values of $\alpha$ and v.

**Example 4:** Let X have a chi-square distribution with seven degrees of freedom using Table-2, $\chi^2_{0.05}(7) = 14.07$ and $\chi^2_{0.95}(7) = 2.167$.

**Example 5:** Let X have a chi-square distribution with five degrees of freedom using Table-2, $P(1.145 \le X \le 12.83) = F(12.83)-F(1.145) = 0.975-0.050 = 0.925$ and $P(X \le 15.09) = 1-F(15.09) = 1-.099 = 0.01$

### 3.2 t - Distribution

If Z is a random variable N(0,1), U is a $\chi^2$ (v) variate and if Z and U are independent, then

$$T = Z / \sqrt{U}$$

has a t - distribution with v degrees of freedom, its probability density function is

$$f(x) = \frac{1}{\sqrt{v}\, B(\frac{1}{2}, \frac{v}{2})} \frac{1}{[1 + \frac{x^2}{v}]^{(v+1)/2}}, \quad 0 \le x < \infty$$

Distribution of T is completely determined by number ν. Table 3 gives the values of t-corresponding with the various values of the probability of a random variable falling outside the limits $\pm t$. Graph of probability density function of T is symmetrical with respect to vertical axis t = 0. For t-distribution,

$$\text{Mean} = 0 ; \qquad \text{Variance} = \frac{v}{v-2}, \ \nu \ge 2$$

For $\nu = 1$, the mean and variance do not exist. Let $t_{\alpha}(\nu)$ denote the constant for which

$$P[T \ge t_{\alpha}(\nu)] = \alpha$$

When T has 't' distribution with ν degrees of freedom. $t_{\alpha}(\nu)$ is the upper $100\alpha$ percent point of t distribution with ν degrees of freedom.

**Example 6:** Let T have a t-distribution with 7 degree of freedom, then from Table-3 we have

$P(T \le 1.415) = 0.90$

$P(T \ge 1.415) = 1 - P(T \le 1.415) = 0.10$

$P(-1.895 \le T \le 1.415) = 0.90 - 0.05 = 0.85$

**Example 7:** Let T have a t distribution with a variance of 5 / 4 ($\nu = 10$) then

$P(-1.812 \le T \le 1.812) = 0.90$

$t_{0.05}(10) = 1.812$ ; $t_{0.01}(10) = 2.764$ ; $t_{0.99}(10) = -2.764$,

### 3.3 F Distribution

One of the most important distributions in applied statistics is the F distribution. F distribution is defined to be the ratio of two independent chi-square variates, each divided by their respective degrees of freedom.

$$F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2},$$

where $\chi_1^2$ is a value of a chi-square distribution with $\nu_1$ degrees of freedom and $\chi_2^2$ is a value of a chi-square distribution with $\nu_2$ degrees of freedom. The mathematical form of the p.d.f. of F distribution is.

$$f(x) = \frac{(\upsilon_1/\upsilon_2)^{\upsilon_1/2}}{B(\frac{\upsilon_1}{2}, \frac{\upsilon_2}{2})} \frac{x^{(\upsilon_1/2)-1}}{[1+\frac{\upsilon_1}{\upsilon_2}x]^{(\upsilon_1+\upsilon_2)/2}}, \quad -\infty \le x < \infty \qquad 0 < x < \infty$$

To obtain an f value, select a random sample of size $n_1$ from a normal population having variance $\sigma_1^2$ and compute $s_1^2/\sigma_1^2$. An independent random sample of size $n_2$ is then selected from a second normal population having variance $\sigma_2^2$ and compute $s_2^2/\sigma_2^2$. The ratio of two quantities $s_1^2/\sigma_1^2$ and $s_2^2/\sigma_2^2$ produces an f value. The distribution of all possible f values is called F distribution .The number of degrees of freedom associated with the sample variance in numerator is stated first, followed by the number of degrees of freedom associated with the sample variance in the denominator. Thus the curve of F distribution depends not only on the two parameters $\nu_1$ and $\nu_2$ but also on the order in which we state them. Once these two values are given we can identify the curve.

Let $f_\alpha$ be the f value above which we find an area equal to $\alpha$. Table-4 gives values of $f_\alpha$ only for $\alpha = 0.05$ and $\alpha = 0.01$ for various combinations of the degrees of freedom $\nu_1$ and $\nu_2$. Hence the f value with 6 and 10 degrees of freedom, leaving an area of 0.05 to the right, is $f_{0.05} = 3.22$.

The F-distribution is applied primarily in the analysis of variance, where we wish to test the equality of several means simultaneously. F-distribution is also used to make influences concerning the variance of two normal populations.

**Table 1:** The Normal Probability Integral or Area under the Normal Curve

$$P = \int_{a}^{\infty} zdz = \text{P (Z>a)}$$

| Z | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0. | 50000 | 49601 | 49202 | 48803 | 48405 | 48006 | 47608 | 47210 | 46812 | 46414 |
| 0.1 | | 46017 | 45620 | 45224 | 44828 | 44433 | 44038 | 43644 | 43251 | 42858 | 42465 |
| 0.2 | | 42074 | 41683 | 41294 | 40905 | 40517 | 40129 | 39743 | 39358 | 38974 | 38591 |
| 0.3 | | 38209 | 37828 | 37448 | 37070 | 36693 | 36317 | 35942 | 35569 | 35197 | 34827 |
| 0.4 | | 34458 | 34090 | 33724 | 33360 | 32997 | 32636 | 32276 | 31918 | 31561 | 31207 |
| 0.5 | | 30854 | 30503 | 30153 | 29806 | 29460 | 29116 | 28774 | 28434 | 28096 | 27760 |
| 0.6 | | 27425 | 27093 | 26763 | 26435 | 26109 | 25785 | 25463 | 25143 | 24825 | 24510 |
| 0.7 | | 24196 | 23885 | 23576 | 23270 | 22965 | 22663 | 22363 | 22065 | 21770 | 21476 |
| 0.8 | | 21186 | 20897 | 20611 | 20327 | 20045 | 19766 | 19489 | 19215 | 18943 | 18673 |
| 0.9 | | 18406 | 18141 | 17879 | 17619 | 17361 | 17106 | 16853 | 16602 | 16354 | 16109 |
| 1.0 | | 15866 | 15625 | 15386 | 15151 | 14917 | 14686 | 14457 | 14231 | 14007 | 13786 |
| 1.1 | | 13567 | 13350 | 13136 | 12924 | 12714 | 12507 | 12302 | 12100 | 11900 | 11702 |
| 1.2 | | 11507 | 11314 | 11123 | 10935 | 10749 | 10565 | 10383 | 10204 | 10027 | 098525 |
| 1.3 | 0.0 | 96800 | 95098 | 93418 | 91759 | 90123 | 88508 | 86915 | 85343 | 83793 | 82264 |
| 1.4 | | 80757 | 79270 | 77804 | 76359 | 74934 | 73529 | 72145 | 70781 | 69437 | 68112 |
| 1.5 | | 66807 | 65522 | 64255 | 63008 | 61780 | 60571 | 59380 | 58208 | 57053 | 55917 |
| 1.6 | | 54799 | 53699 | 52616 | 51551 | 50503 | 49471 | 48457 | 47460 | 46479 | 45514 |
| 1.7 | | 44565 | 43633 | 42716 | 41815 | 40930 | 40059 | 39204 | 38364 | 37538 | 36727 |
| 1.8 | | 35930 | 35148 | 34380 | 33625 | 32884 | 32157 | 31443 | 30742 | 30054 | 29379 |
| 1.9 | | 28717 | 28067 | 27429 | 26803 | 26190 | 25588 | 24998 | 24419 | 23852 | 23295 |
| 2.0 | | 22750 | 22216 | 21692 | 21178 | 20675 | 20182 | 19699 | 19226 | 18763 | 18309 |
| 2.1 | | 17864 | 17429 | 17003 | 16586 | 16177 | 15778 | 15386 | 15003 | 14629 | 14262 |
| 2.2 | | 13903 | 13553 | 13209 | 12874 | 12545 | 12224 | 11911 | 11604 | 11304 | 11011 |
| 2.3 | | 10724 | 10444 | 10170 | 099031 | 096419 | 093867 | 091375 | 088940 | 086563 | 084242 |
| 2.4 | 0.00 | 81975 | 79763 | 77603 | 75494 | 73436 | 71428 | 69469 | 67557 | 65691 | 63872 |
| 2.5 | | 62097 | 60366 | 58677 | 57031 | 55426 | 53861 | 52336 | 50849 | 49400 | 47988 |
| 2.6 | | 46612 | 45271 | 43965 | 42692 | 41453 | 40246 | 39070 | 37926 | 36811 | 35726 |
| 2.7 | | 34670 | 33642 | 32641 | 31667 | 30720 | 29798 | 28901 | 28028 | 27179 | 26354 |
| 2.8 | | 25551 | 24771 | 24012 | 23274 | 22557 | 21860 | 21182 | 20524 | 19884 | 19262 |
| 2.9 | | 18658 | 18071 | 17502 | 16948 | 16411 | 15889 | 15382 | 14890 | 14412 | 13949 |
| 3.0 | | 13499 | 13062 | 12639 | 12228 | 11829 | 11442 | 11067 | 10703 | 10350 | 10008 |
| 3.1 | 0.000 | 96760 | 93544 | 90426 | 87403 | 84474 | 81635 | 78885 | 76219 | 73638 | 71136 |
| 3.2 | | 68714 | 66367 | 64095 | 61895 | 59765 | 57703 | 55706 | 53774 | 51904 | 50094 |
| 3.3 | | 48342 | 46648 | 45009 | 43423 | 41889 | 40406 | 38971 | 37584 | 36243 | 34946 |
| 3.4 | | 33693 | 32481 | 31311 | 30179 | 29086 | 28029 | 27009 | 26023 | 25071 | 24151 |
| 3.5 | | 23263 | 22405 | 21577 | 20778 | 20006 | 19262 | 18543 | 17849 | 17180 | 16534 |
| 3.6 | | 12911 | 15310 | 14730 | 14171 | 13632 | 13112 | 12611 | 12128 | 11662 | 11213 |
| 3.7 | | 10780 | 10363 | 099611 | 095740 | 092010 | 088417 | 084957 | 081624 | 078414 | 075324 |
| 3.8 | 0.0000 | 72348 | 69483 | 66726 | 64072 | 61517 | 59059 | 56694 | 54418 | 52228 | 50122 |
| 3.9 | | 48096 | 46148 | 44274 | 42473 | 40741 | 39076 | 37475 | 35936 | 34458 | 33037 |
| 4.0 | | 31671 | 30359 | 29099 | 27888 | 26726 | 25609 | 24536 | 23507 | 22518 | 21569 |
| 4.1 | | 20658 | 19783 | 18944 | 18138 | 17365 | 16624 | 15912 | 15230 | 14575 | 13948 |
| 4.2 | | 13346 | 12769 | 12215 | 11685 | 11176 | 10689 | 10221 | 097736 | 093447 | 089337 |
| 4.3 | 0.00000 | 85399 | 81627 | 78015 | 74555 | 71241 | 68069 | 65031 | 62123 | 59340 | 56675 |
| 4.4 | | 54125 | 51685 | 49350 | 47117 | 44979 | 42935 | 40980 | 39110 | 37322 | 35612 |
| 4.5 | | 33977 | 32414 | 30920 | 29492 | 28127 | 26823 | 25577 | 24386 | 23249 | 22162 |
| 4.6 | | 21125 | 20133 | 19187 | 18283 | 2E+05 | 16597 | 15810 | 15060 | 14344 | 13660 |
| 4.7 | | 13008 | 12386 | 11792 | 11226 | 10686 | 10171 | 096796 | 092113 | 087648 | 083391 |
| 4.8 | 0.000000 | 79333 | 75465 | 71779 | 68267 | 64920 | 61731 | 58693 | 55799 | 53043 | 50418 |
| 4.9 | | 47918 | 45538 | 43272 | 41115 | 39061 | 37107 | 35247 | 33476 | 31792 | 30190 |

**Table 2:** Values of $\chi^2$ with probability P of being exceeded in random sampling at $\alpha$% level of significance; n = Number of degrees of freedom

| n | 0.99 | 0.95 | 0.90 | 0.70 | 0.50 | 0.30 | 0.20 | 0.10 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | $.0^3157$ | 0.00393 | 0.0158 | 0.148 | 0.455 | 1.074 | 1.642 | 2.706 | 3.841 | 5.412 | 6.635 |
| **2** | 0.0201 | 0.103 | 0.211 | 0.713 | 1.386 | 2.408 | 3.219 | 4.605 | 5.991 | 7.824 | 9.210 |
| **3** | 0.115 | 0.352 | 0.584 | 1.424 | 2.366 | 3.665 | 4.642 | 6.251 | 7.815 | 9.837 | 11.345 |
| **4** | 0.297 | 0.711 | 1.064 | 2.195 | 3.357 | 4.878 | 5.989 | 7.779 | 9.488 | 11.668 | 13.277 |
| **5** | 0.554 | 1.145 | 1.610 | 3.000 | 4.351 | 6.064 | 7.289 | 9.236 | 11.070 | 13.388 | 15.086 |
| **6** | 0.872 | 1.635 | 2.204 | 3.828 | 5.348 | 7.231 | 8.558 | 10.645 | 12.592 | 15.033 | 16.812 |
| **7** | 1.239 | 2.167 | 2.833 | 4.671 | 6.346 | 8.383 | 9.803 | 12.017 | 14.067 | 16.622 | 18.475 |
| **8** | 1.646 | 2.733 | 3.490 | 5.527 | 7.344 | 9.524 | 11.030 | 13.362 | 15.507 | 18.168 | 20.090 |
| **9** | 2.088 | 3.325 | 4.168 | 6.393 | 8.343 | 10.656 | 12.242 | 14.684 | 16.919 | 19.679 | 21.666 |
| **10** | 2.558 | 3.940 | 4.865 | 7.267 | 9.342 | 11.781 | 13.442 | 15.987 | 18.307 | 21.161 | 23.209 |
| **11** | 3.053 | 4.575 | 5.578 | 8.148 | 10.341 | 12.899 | 14.631 | 17.275 | 19.675 | 22.618 | 24.725 |
| **12** | 3.571 | 5.226 | 6.304 | 9.034 | 11.340 | 14.011 | 15.812 | 18.549 | 21.026 | 24.054 | 26.217 |
| **13** | 4.107 | 5.892 | 7.042 | 9.926 | 12.340 | 15.119 | 16.985 | 19.812 | 22.362 | 25.472 | 27.688 |
| **14** | 4.660 | 6.571 | 7.790 | 10.821 | 13.339 | 16.222 | 18.151 | 21.064 | 23.685 | 26.873 | 29.141 |
| **15** | 5.229 | 7.261 | 8.547 | 11.721 | 14.339 | 17.322 | 19.311 | 22.307 | 24.996 | 28.259 | 30.578 |
| **16** | 5.812 | 7.962 | 9.312 | 12.624 | 15.338 | 18.418 | 20.465 | 23.542 | 26.296 | 29.633 | 32.000 |
| **17** | 6.408 | 8.672 | 10.085 | 13.531 | 16.338 | 19.511 | 21.615 | 24.769 | 27.587 | 30.995 | 33.409 |
| **18** | 7.015 | 9.390 | 10.865 | 14.440 | 17.338 | 20.601 | 22.760 | 25.989 | 28.869 | 32.346 | 34.805 |
| **19** | 7.633 | 10.117 | 11.651 | 15.352 | 18.338 | 21.689 | 23.900 | 27.204 | 30.144 | 33.687 | 36.191 |
| **20** | 8.260 | 10.851 | 12.443 | 16.266 | 19.337 | 22.775 | 25.038 | 28.412 | 30.410 | 35.020 | 37.566 |
| **21** | 8.897 | 11.591 | 13.240 | 17.182 | 20.337 | 23.858 | 26.171 | 29.615 | 32.671 | 36.343 | 38.932 |
| **22** | 9.542 | 12.338 | 14.041 | 18.101 | 21.337 | 24.939 | 27.301 | 30.813 | 33.924 | 37.659 | 40.289 |
| **23** | 10.196 | 13.091 | 14.848 | 19.021 | 22.337 | 26.018 | 28.429 | 32.007 | 35.172 | 38.968 | 41.638 |
| **24** | 10.856 | 13.848 | 15.659 | 19.943 | 23.337 | 27.096 | 29.553 | 33.196 | 36.415 | 40.270 | 42.980 |
| **25** | 11.524 | 14.611 | 16.473 | 20.867 | 24.337 | 28.172 | 30.675 | 34.382 | 37.652 | 41.566 | 44.314 |
| **26** | 12.198 | 15.379 | 17.292 | 21.792 | 25.336 | 29.246 | 31.795 | 35.563 | 38.885 | 42.856 | 45.642 |
| **27** | 12.879 | 16.151 | 18.114 | 22.719 | 26.336 | 30.319 | 32.912 | 36.741 | 40.113 | 44.140 | 46.963 |
| **28** | 13.565 | 16.928 | 18.939 | 23.647 | 27.336 | 31.391 | 34.027 | 37.916 | 41.337 | 45.419 | 48.278 |
| **29** | 14.256 | 17.708 | 19.768 | 24.577 | 28.336 | 32.461 | 35.139 | 39.087 | 42.557 | 46.693 | 49.588 |
| **30** | 14.953 | 18.493 | 20.599 | 25.508 | 29.336 | 33.530 | 36.250 | 40.256 | 43.773 | 47.962 | 50.892 |
| **32** | 16.362 | 20.072 | 22.271 | 27.373 | 31.336 | 35.665 | 38.466 | 42.585 | 46.194 | 50.487 | 53.486 |
| **34** | 17.789 | 21.664 | 23.952 | 29.242 | 33.336 | 37.795 | 40.676 | 44.903 | 48.602 | 52.995 | 56.061 |
| **36** | 19.233 | 23.269 | 25.643 | 31.115 | 35.336 | 39.922 | 42.879 | 47.212 | 50.999 | 55.489 | 58.619 |
| **38** | 20.691 | 24.884 | 27.343 | 32.992 | 37.335 | 42.045 | 45.076 | 49.513 | 53.384 | 57.969 | 61.162 |
| **40** | 22.164 | 26.509 | 29.051 | 34.872 | 39.335 | 44.165 | 47.269 | 51.805 | 55.759 | 60.436 | 63.691 |
| **42** | 23.650 | 28.144 | 30.765 | 36.755 | 41.335 | 46.282 | 49.456 | 54.090 | 58.124 | 62.892 | 66.206 |
| **44** | 25.148 | 29.787 | 32.487 | 38.641 | 43.335 | 48.396 | 51.639 | 56.369 | 60.481 | 65.337 | 68.710 |
| **46** | 26.657 | 31.439 | 34.215 | 40.529 | 45.335 | 50.507 | 53.818 | 58.641 | 62.830 | 67.771 | 71.201 |
| **48** | 28.177 | 33.098 | 35.949 | 42.420 | 47.335 | 52.616 | 55.993 | 60.907 | 65.171 | 70.197 | 73.683 |
| **50** | 29.707 | 34.764 | 37.689 | 44.313 | 49.335 | 54.723 | 58.164 | 63.167 | 67.505 | 72.613 | 76.154 |
| **52** | 31.246 | 36.437 | 39.433 | 46.209 | 51.335 | 56.827 | 60.332 | 65.422 | 69.832 | 75.021 | 78.616 |
| **54** | 32.793 | 38.116 | 41.183 | 48.106 | 53.335 | 58.930 | 62.496 | 67.673 | 72.153 | 77.422 | 81.069 |
| **56** | 34.350 | 39.801 | 42.937 | 50.005 | 55.335 | 61.031 | 64.658 | 69.919 | 74.468 | 79.815 | 83.513 |
| **58** | 35.913 | 41.492 | 44.696 | 51.906 | 57.335 | 63.129 | 66.816 | 72.160 | 76.778 | 82.201 | 85.950 |
| **60** | 37.485 | 43.188 | 46.459 | 53.809 | 59.335 | 65.227 | 68.972 | 74.397 | 79.082 | 84.580 | 88.379 |
| **62** | 39.063 | 44.889 | 48.226 | 55.714 | 61.335 | 67.322 | 71.125 | 76.630 | 81.381 | 86.953 | 90.802 |
| **64** | 40.649 | 46.595 | 49.996 | 57.620 | 63.335 | 69.416 | 73.276 | 78.860 | 83.675 | 89.320 | 93.217 |
| **66** | 42.240 | 48.305 | 51.770 | 59.527 | 65.335 | 71.508 | 75.424 | 81.085 | 85.965 | 91.681 | 95.626 |
| **68** | 43.838 | 50.020 | 53.548 | 61.436 | 67.335 | 73.600 | 77.571 | 83.308 | 88.250 | 94.037 | 98.028 |
| **70** | 45.442 | 51.739 | 55.329 | 63.346 | 69.334 | 75.689 | 79.715 | 85.527 | 90.531 | 96.388 | 100.43 |

**Table 3: P**robability of of a random variable falling outside the limits $\pm t$
n = number of degrees of freedom

| n | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 636.619 |
| 2 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 31.598 |
| 3 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 12.924 |
| 4 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.474 | 4.604 | 8.610 |
| 5 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 6.869 |
| 6 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.959 |
| 7 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 5.408 |
| 8 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 5.041 |
| 9 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.781 |
| 10 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.587 |
| 11 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.437 |
| 12 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 4.318 |
| 13 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 4.221 |
| 14 | 0.692 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 4.140 |
| 15 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 4.073 |
| 16 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 4.015 |
| 17 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.965 |
| 18 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.922 |
| 19 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.883 |
| 20 | 0.687 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.850 |
| 21 | 0.686 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.819 |
| 22 | 0.686 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.792 |
| 23 | 0.685 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.767 |
| 24 | 0.685 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.745 |
| 25 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.725 |
| 26 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.707 |
| 27 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.690 |
| 28 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.674 |
| 29 | 0.683 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.659 |
| 30 | 0.683 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.646 |
| 40 | 0.681 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.551 |
| 60 | 0.679 | 0.848 | 1.046 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.460 |
| 120 | 0.677 | 0.845 | 1.041 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.373 |
| $\infty$ | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

**Table 4**
F-table (5%)

| $n_2$ / $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 12 | 24 | ∝ |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.37 | 19.41 | 19.45 | 19.50 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.84 | 8.74 | 8.64 | 8.53 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.04 | 5.91 | 5.77 | 5.63 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.82 | 4.68 | 4.53 | 4.36 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.15 | 4.00 | 3.84 | 3.67 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.73 | 3.57 | 3.41 | 3.23 |
| 8 | 5.32.. | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.44 | 3.28 | 8.12 | 2.93 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.23 | 3.07 | 2.90 | 2.71 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.07 | 2.91 | 2.74 | 2.54 |
| 12 | 4.75 | 3.88 | 3.49 | 3.26 | 3.11 | 3.00 | 2.85 | 2.69 | 2.50 | 2.30 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.70 | 2.53 | 2.35 | 2.13 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.59 | 2.42 | 2.24 | 2.01 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.51 | 2.34 | 2.15 | 1.92 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.45 | 2.28. | 2.08 | 1.84 |
| 25 | 4.24 | 3.38 | 2.99 | 2.76 | 2.60 | 2.49 | 2.34 | 2.16 | 1.96 | 1.71 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.27 | 2.09 | 1.89 | 1.62 |
| 40 | 4..08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.18 | 2.00 | 1.79 | 1.51 |
| 60 | 4..00 | 3.15 | 2.76 | 2.52 | 2.37 | 2.25 | 2.10 | 1.92. | 1.70 | 1.39 |
| 80 | 3.96 | 3.11 | 2.72 | 2.49 | 2.33 | 2.21 | 2.06 | 1.88 | 1.65 | 1.32 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.02 | 1.83 | 1.61 | 1.25 |
| ∝ | 3.84 | 2.99 | 2.60 | 2.37 | 2.21 | 2.10 | 1.94 | 1.75 | 1.52 | 1.00 |