UNIT 1

CHAPTER 2

SOME STANDARD DISTRIBUTIONS

1. **BINOMIAL DISTRIBUTION:**

   Bernoulli Experiment: Suppose we perform an experiment with two possible outcomes; either success or failure. Success happens with probability p, hence failure occurs with probability q = 1 – p.

   The Binomial experiment means Bernoulli experiment which is repeated n times. The Binomial distribution is used to obtain the probability of observing x successes in n trials, with the probability of success on a single trial denoted by p. The Binomial Distribution assumes that p is fixed for all trials. Here n and p are called the parameters of Binomial Distribution.

   The conditions needed to be satisfied for a Binomial experiment are as follows:

   (1) The number of trials is n and n is a fixed number.

   (2) The outcome of a given trial is either a ''success'' or ''failure''.

   (3) The probability of success (p) remains constant from trial to trial.

   (4) The trials are independent ; the outcome of a trial is not affected by the outcome of any other trial.

Definition:

A discrete random variable X is said to follow Binomial Distribution with parameters (n,p), if its p.m.f. is given by

$$P(x) = P(X=x) = {}^{n}C_{x}p^{x}q^{n-x}, \qquad x=0,1,2,\ldots,n; \quad 0<p<1; \quad p+q=1$$

$$= 0, \qquad \text{otherwise.}$$

It is denoted as X~B(n,p)

Note: (i) The Binomial Distribution satisfies two conditions, necessary for p.m.f.

$P(x) \geq 0$ (for x=0,1,2,….,n) are always positive.

Also $\Sigma P(x) = 1$ (for x=0,1,2,…n)

Let us consider

$\Sigma P(x) = \Sigma\, {}^nC_x p^x q^{n-x}$

$= {}^nC_0 p^0 q^{n-0} + {}^nC_1 p^1 q^{n-1} + {}^nC_2 p^2 q^{n-2} + {}^nC_3 p^3 q^{n-3} + \ldots\ldots. + {}^nC_n p^n q^{n-n}$

$= (q + p)^n = 1$

(ii) Binomial probabilities are successive terms of binomial expansion of $(q + p)^n$.

Properties :

(1) **Mean** = np
(2)**Variance** = npq

Note: Mean = np and Variance =npq

Since q < 1 => npq < np

i.e., variance < mean

(3) **Mode:** If (n+1)p = k is an integer then Binomial Distribution is bimodal. Modal values are (k-1) and k. but if (n+1)p is not an integer, then Binomial Distribution is unimodal. Modal value is integral part of (n+1)p.

(4) **Skewness and Kurtosis:**

Measure of Skewness β_1 and Measure of Kurtosis β_2 are :

$\beta_1 = \mu_3^2/\mu_2^3 = (1 - 2p)^2/npq$

⇨ $\gamma_1 = \sqrt{B_1} = (1 - 2p)/\sqrt{npq}$

$\beta_2 = \mu_4/\mu_2^2 = 3 + (1 - 6pq)/npq$

⇨ $\gamma_2 = \beta_2 - 3 = (1 - 6pq)/npq$

Binomial Distribution is symmetric if $p = 0.5$, $q = 0.5$, if $p < 0.5$, then the distribution is positively skewed and if $p > 0.5$, then the distribution is negatively skewed.

(5)If $X \sim B(n,p)$ and $Y \sim B(n,p)$ and if X and Y are independently distributed then the distribution of (X+Y) is also Binomial with parameters (n+m,p), i.e., $(X+Y) \sim B(n+m,p)$

Example 1: An unbiased coin is tossed four times. Calculate the probability of obtaining (i) no heads, (ii) at least one head and (iii) more heads than tails.

Solution: Let X = number of times head is observed when an unbiased coin is tossed four times.

X follows Binomial Distribution with n=4 and p=0.5

We have p.m.f. of Binomial Distribution as

$P(x) = P(X=x) = {}^nC_x p^x q^{n-x}$ $\quad x=0,1,2,\ldots,n;\ \ 0<p<1;\ \ p+q=1$

$=0,$ $\quad$ otherwise.

(i) Probability of obtaining no heads = P(X=0)
$P(X=0) = (0.5)^4 = 0.0625$
(ii) Probability of obtaining at least one head = $P(X \geq 1)$
$P(X \geq 1) = 1 - P(X=0) = 1 - 0.0625 = 0.9375$
(iii) Probability of obtaining more heads than tails is $P(X=4)+P(X=3)$
$P(X=4) = {}^4C_4(0.5)^4(0.5)^{4-4} = (0.5)^4 = 0.0625$
$P(X=3) = {}^4C_3(0.5)^3(0.5)^{4-3} = 4 * (0.5)^4 = 0.25$

Hence, $P(X=4)+P(X=3)= 0.0625 + 0.25 = 0.3125$

Example 2: For a Binomial variate X with parameters (n,p); p=q and P(X=2)=P(X=3), find (i) P(X=1) and (ii) P(X>1).

Solution: Given X is a binomial variate with parameters (n,p);

p=q , since p+q=1, therefore p=q=0.5

Also, $P(X=2) = P(X=3)$

$^nC_2(0.5)^2(0.5)^{n-2} = {}^nC_3(0.5)^3(0.5)^{n-3}$

- ⇨ $n = 5$

(i) $P(X=1) = {}^5C_1(0.5)^1(0.5)^{5-1}$

$=5 * (0.5)^5$

$=0.15625$

(ii) $P(X>1) = 1 – P(X \leq 1) = 1 – [P(X=0) + P(X=1)]$

$= 1 – 0.1875$

$= 0.8125$

Example 3: For a Binomial variate mean is 3 and variance is 2, find (i) P(X = 1) & P(X > 2).

Solution: Let X~B(n,p).

Given mean = np = 3 ………..(i)

And variance = npq = 2 ……….(ii)

- ⇨ Variance/mean = 2/3
- ⇨ npq/np = 2/3
- ⇨ q = 2/3

Since 1 – q = p

- ⇨ 1 – 2/3 = p

⇨ $1/3 = p$

Putting $p=1/3$ and $q=2/3$ in (ii) we get

$npq = 2$

⇨ $n*1/3*2/3 = 2$

⇨ $n = 9$

(i) $P(X = 1) = {}^9C_1(1/3)^1(2/3)^{9-1}$
$= 0.1170$

(ii) $P(X > 2) = 1 – P(X \leq 2) = 1 – [P(X=0) + P(X=1) + P(X =2)]$
$= 1 – 0.3771$
$= 0.6229$

Practice Sums:

1. Five unbiased coins are tossed 640 times. Find in how many tosses you expect (i) all tails; (ii) three heads and (iii) exactly four tails.
2. If discrete random variable $X \sim B(6,p)$. Find p if $9P(X=4)=P(X=2)$.
3. For a binomial variate mean is 3 and $15P(X=0)=2P(X=1)$, find $P(X=5)$

Real Life Applications of Binomial Distributions:

1. Binomial theorem is used in computing areas such as in distribution of IP addresses. With binomial distribution , the automatic allocation of IP addresses is possible.
2. Another field that uses Binomial Distribution as the important tools is the nation's economic prediction. Economists use binomial distribution theorem to count probabilities to predict the way the economy will behave in the next few years. To be able to come up with realistic predictions, binomial theorem is used in this field.
3. Binomial Distribution has also been a great use in the architecture industry in design of infrastructure. It allows engineers to calculate the magnitudes of the projects and thu

delivering accurate estimates of not only the costs but also time required to construct them.

4. The Binomial Distribution is used when a researcher is interested in the occurrence of an event and not in its magnitude. For instance, in a clinical trial, a patient may survive or die. The researcher studies the number of survivors, but not how long the patient survives after treatment.
5. Other situations in which Binomial Distributions arise are quality control, public opinion surveys, medical research and insurance problems.

NORMAL DISTRIBUTION:

The Normal Distribution is the most used statistical distribution. The principal reason is normality arises naturally in many physical, biological and social measurement situations. A good number of random variables occurring in practice actually are normally distributed or very close to it. For example: height and intelligence are approximately normally distributed; measurement errors also often have a normal distribution.

Definition: A continuous random variable X is said to follow normal distribution with parameters μ and σ^2, if its probability density function is given by

$f(x) = 1(/\sigma\sqrt{2\pi})e^{1/2(x-\mu/\sigma)^2}$; $-\infty< x <\infty, -\infty< \mu <\infty, \sigma > 0.$

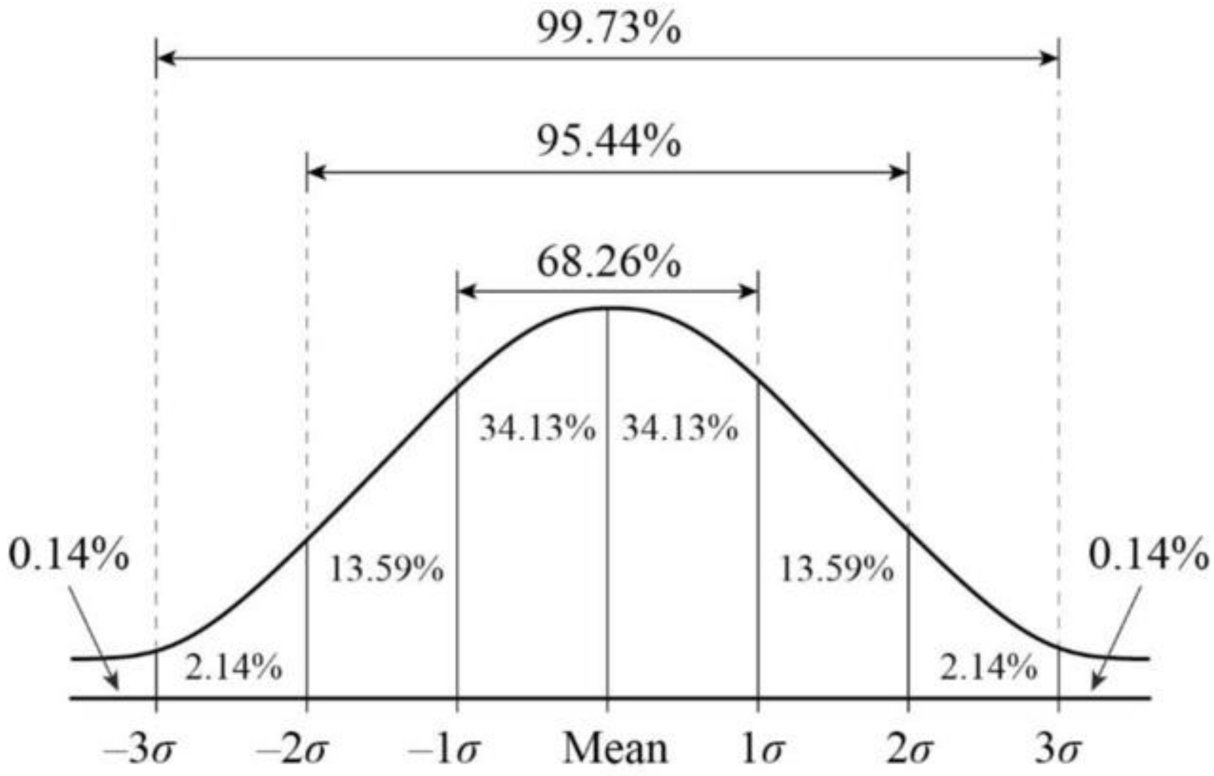
And it is denoted as $X \sim N(\mu , \sigma^2)$

Note: If $\mu = 0$ and $\sigma^2 = 1$ then the normal variable is known as standard normal variate and generally denoted by Z i.e., if Z is a standard normal variate then $Z\sim N(0,1)$ and its p.d.f. is given by

$Ff(z) = (1/\sqrt{2\pi})e^{-z^2/2}$; $-\infty< x <\infty$

Probability curve: The Probability Density curve of N(μ , σ^2) is a bell-shaped, symmetric about μ and Mesokurtic .

A graphical representation of a normal curve is as given below:



Properties of the Normal Distribution:

1. If X ~ N(μ, σ^2), then its p.d.f. is given by
   f(x) =1(/σ√2π)$e^{1/2(x-\mu/\sigma)^2}$;   -∞< x <∞, -∞< μ <∞, σ > 0
2. Lim (x-> -∞) f(x) = Lim (x-> ∞) f(x) = 0, the value of f(x) approaches zero as x approaches negative infinity or positive infinity.
3. Total area under the normal curve is 1.
4. f(μ - x) = f(μ + x); the p.d.f. is symmetric about μ. The normal curve y = f(x) is a bell symmetric curve, symmetric about x = μ. As x moves away from μ, the curve comes closer and closer to X axis and extends till infinity on both the sides without touching the X axis.

5. Distribution is symmetric distribution ,i.e., mean = median = mode = μ.
6. Measure of skewness = $\beta_1 = 0$ and measure of Kurtosis = $\beta_2 = 3$.
7. Area under normal curve between:

Ordinate	μ ± σ	μ ± 2σ	μ ± 3σ	μ ±0.6745σ	μ ± 1.645σ	μ ± 1.96σ	μ ± 2.326	μ ± 2.575σ
Area(%)	68.27	95.45	99.73	50	90	95	98	99

8. The cdf for the normal distribution is given by
   $F(x) = P(X \leq x) = \int 1(/\sigma\sqrt{2\pi})e^{1/2(x-\mu/\sigma)^2}\,dt$, -∞ to +∞.
9. If $X_1, X_2, \ldots, X_n$ are normal variates such that $X_i \sim N(\mu_i, \sigma_i^2)$ for I = 1,2,….n. then $\Sigma l_i X_i \sim N(\Sigma l_i \mu_i, \Sigma l_i^2 \sigma_i^2)$ where l_i are constants ,i.e., linear combination of normal variates is also a normal variate.
10. If $X \sim N(\mu, \sigma^2)$, then $Z = [(X - \mu)/\sigma] \sim N(0,1)$, i.e., Z follows standard normal distribution with mean 0 and variance 1.
    The p.d.f. of Z is $\Phi(z) = (1/\sqrt{2\pi})e^{-z^2/2}$, $-\infty < z < +\infty$.

Ex.1: For a standard normal variable Z, find the area using the table:

(i) To the left of Z = 1.3, (ii) to the left of Z = -1.2, (iii) to the right of Z = 1.6 , (iv) between Z = 0 and Z = -2.

Solution:

(i) To the left of Z = 1.3
$P(Z < 1.3) = \Phi(1.3) = 0.9032$

(ii) To the left of Z = -1.2
$P(Z < -1.2) = \Phi(-1.2) = 1 - \Phi(1.2) = 1 - 0.88493 = 0.11507$

(iii) To the right of Z = 1.6
$P(Z > 1.6) = 1 - P(Z < 1.6) = 1 - \Phi(1.6) = 1 - 0.9452 = 0.0548$

(iv) Between Z = 0 and Z = -2 = area between Z = 0 and Z = 2 (using the property of symmetry)
$P(-2 < Z < 0) = P(0 < Z < 2) = 0.47725$

Ex.2: The time to pass through a queue to begin self-service at a cafeteria has been found to be N(15,9). Find the probability that an arriving customer waits between 14 and 17 minutes.

Solution:

The probability that an arriving customer waits between 14 and 17 minutes is determined as follows:

$$P(14 \leq X \leq 17) = F(17) - F(14) = \Phi(17 - 15/3) - \Phi(14 - 15/3)$$

$$= \Phi(0.667) - \Phi(-0.333)$$

$$= 0.7476 - [1- \Phi(0.333)]$$

$$= 0.7476 - [1- 0.6304]$$

$$= 0.3780$$

Application of Normal Distribution:

One of the most important applications of the normal distribution is the analysis of errors of measurement made in astronomical observations, error that occur because of imperfect instruments and imperfect observers. Gauss in 1809 showed that errors were fit well by this distribution.

The normal distribution is the foundation for statistical inference. It is one of the most commonly used distributions. This distribution is important due to its wide use. Many things which are measured on continuous scale closely follow normal distributions.

For example:

(i) Heights or weights of people
(ii) Size of items produced by machines
(iii) Errors in measurements
(iv) IQ measurements
(v) Percentage of marks.

Normal distribution is useful in Statistical Quality Control, Statistical Inference, Reliability theory, Operations Research, Educational and Psychological Statistics, Theory of Sampling, Design of Experiments and so on. Normal Distribution is viewed as limiting distribution of several distributions like Binomial, Poisson etc. In nature and technology we very often deal with distributions which are very close to normal distribution. Even though the parental distribution is non-normal, using Central Limit Theorem the distribution of sample mean tends to normal distribution.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

CHI-SQUARE DISTRIBUTION:

The Chi-Square Distribution is the distribution of the sum of squared independently distributed standard normal variates. The degrees of freedom of the distribution is equal to the number of standard normal variates being summed. Therefore, Chi-Square with one degree of freedom , written as $\chi^2(1)$, is simply the distribution of a single normal variate squared.

If Z_1, Z_2,…….,Z_n are n independent identically distributed (i.i.d.) standard normal variates, then ΣZ_i^2 follows chi square distribution with n degrees of freedom (d.f.) ($\chi^2(n)$).

Since the square of a standard normal variate is a Chi-square variate with 1 degree of freedom, i.e., if X is normally distributed with mean μ and standard deviation σ, then $[(X - \mu)/ \sigma]^2$ is a Chi-Square variate χ^2

With 1 d.f. The distribution of Chi-square depends on the degrees of freedom. There is a different distribution for each number of degrees of freedom.

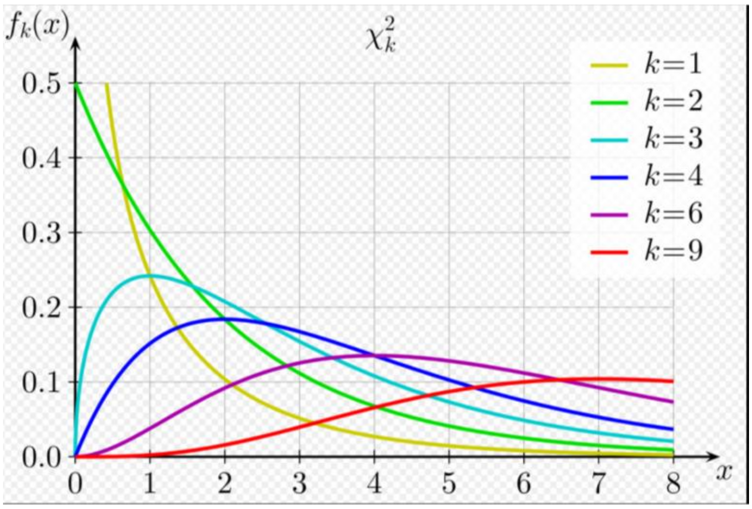
Definition:

Let $X_1, X_2, \ldots\ldots, X_n$ be independent identically distributed $N(\mu, \sigma^2)$; then $\Sigma[(X - \mu)/\sigma]^2$ follows $\chi^2(n)$. The parameter of χ^2 distribution is its degrees of freedom.

The probability density function of $X = \chi^2(n)$ [χ^2 with n d.f.] is given by

$$f(x) = [1/2^{n/2}\Gamma(n/2)]e^{-x/2}X^{(n/2)-1}, \qquad x \geq 0$$

$$= 0, \qquad \text{otherwise.}$$



p.d.f of Chi-square distribution with different d.f (k)

Properties of Chi-square distribution:

1. The Mean of χ^2 distribution is equal to the number of degrees of freedom (n).
2. The variance of χ^2 distribution is equal to 2n.
3. The mode of χ^2 distribution is equal to (n – 2). So mode exits if n>2.
4. Since chi-square values are always positive, the chi-square curve lies in the first quadrant and is positively skewed. As n increases, the curve tends to the symmetric curve.
5. Since chi-square value increases with the increase in the degrees of freedom, there is a new Chi-square distribution with every increase in the number of degrees of freedom. As the degrees of

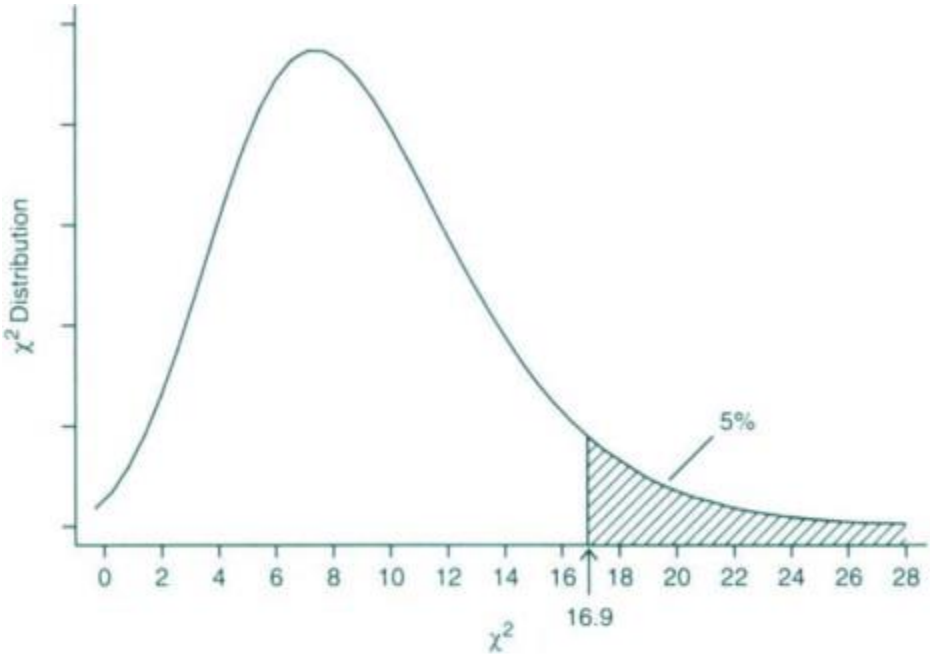
freedom increases, Chi-square distribution tends to normal distribution.

6. When two Chi-square χ_1^2 and χ_2^2 are independent , each having χ^2 distribution with n and m degrees of freedom respectively , then their sum will follow χ^2 distribution with (n+m) degrees of freedom.

Cumulative Probability and the Chi-Square Distribution:

The chi-square distribution is constructed so that the total area under the curve is equal to 1. The area under the curve between 0 and a particular chi-square value is a cumulative probability associated with that chi-square value. The expression for c.d.f. is :

$F(x) = [1/2^{n/2}\Gamma(n/2)] \int e^{-x/2}X^{(n/2)-1}\ dx$ (0 to x)



The shaded region in the above graph is the area which represents a cumulative probability associated with a chi-square statistic. It is the critical region.

Ex .1: Let X ~ $\chi^2(10)$. Find (i) P(X < 7.27) (ii) P(X > 11.78) (iii) Mode of the distribution (iv) median.

Solution: Here, d.f. = n = 10

From the table of chi-square, we get

(i) $P(X < 7.27) = 1 - P(X > 7.27) = 1 - 0.7 = 0.3$

(ii) $P(X > 11.78) = 0.3$

(iii)) Mode of the distribution = $n - 2 = 10 - 2 = 8$

(iv) Median : If M = median , then $P(X > M) = 0.5$

From table, $M = 9.34$

Ex.2: Let $X_1, X_2, \ldots\ldots, X_{20}$ be independent identically distributed N(0,1). Find (i) $P(Y < 25.04)$ (ii) $P(Y > 35.02)$ (iii) $P(10.85 < Y < 28.41)$ where $Y = \Sigma X_i^2$.

Solution : Since $X_1, X_2, \ldots\ldots, X_{20}$ be independent identically distributed N(0,1), the distribution of $Y = \Sigma X_i^2$ is a chi-square with 20 d.f. Using the table we can evaluate the required probabilities.

(i) $P(Y < 25.04) = 1 - P(Y > 25.04) = 1 - 0.20 = 0.80$

(ii) $P(Y > 35.02) = 0.02$

(iii) $P(10.85 < Y < 28.41) = P(Y < 28.41) - P(Y < 10.85)$

$= [1 - P(Y > 28.41)] - [1 - P(Y > 10.85)]$

$= (1 - 0.10) - (1 - 0.95) = 0.85$

Application of chi-square Distribution:

(a) To test if the population has a specified value of the variance σ^2.

(b) Chi-square test of goodness of fit.

(c) Chi-square test for independence of attributes.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

STUDENT’t DISTRIBUTION:It is an important distibution of testing of hypothesis regarding population mean, correlation coefficient, regression coefficients and finding confidence interval for population mean. It was developed by William Sealy Gosset under the pseudonym Student. Whereas a normal distribution describes a full population, t-distribution describes samples from a full population; accordingly, the t-distribution for each sample size is different, so the sample size is the parameter of the t-distribution.

Definition:

Suppose Z is standard normal variate , i.e., Z ~ N(0,1). And V is independently distributed chi-square variate with n d.f. then

$$t = Z/\sqrt{V/n}$$

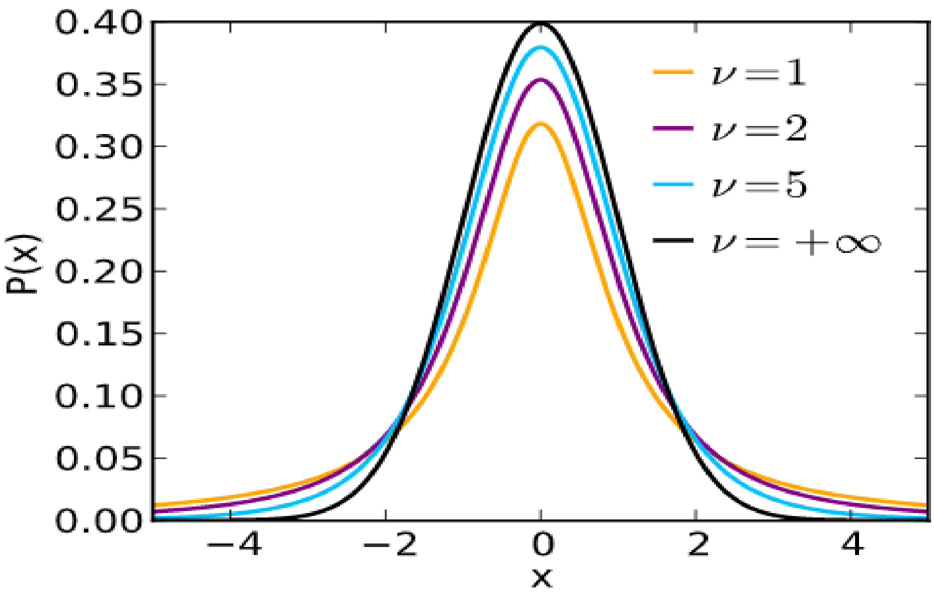
is said to follow t-distribution with n degreees of freedom. N is the parameter of t distribution . The p.d.f. of t is given by

$$f(t) = 1/\sqrt{n}\beta(n/2, ½)\ (1 + t^2/n)^{n+1/2} \qquad -\infty < t < \infty$$

where $\beta(n/2, ½) = [\Gamma(n/2)\Gamma(1/2)]/\Gamma(n+1)/2$

$\Gamma n = (n-1)\Gamma(n-1)$

$= (n-1)!$ if n is an integer.



p.d.f. of t-distribution with different d.f.

Properties of t-distribution with n d.f.:

(1) Since f(-t) = f(t). Hence it is symmetric about zero.

(2) Mean = Median = Mode = 0.

(3) Variance = n/(n-2). Note that variance exits only if n > 2.

(4)The variance is always greater than 1, it is close to 1 when the degrees of freedom n is large.
(5)t-distribution has a greater dispersion than the standard normal distribution.
(6)As n -> ∞ , the distribution of t tends to standard normal distribution.
(7)It is leptokurtic curve and tends to mesokurtic as n tends to ∞.

Cumulative distribution function:

The t-distribution ranges from - ∞ to ∞ and the total area under the curve is equal to 1. The area under the curve ranges between - ∞ and a particular chi-square value is a cumulative probability associated with that t value. The expression for c.d.f. is given below:

$$F(t) = \int 1/\sqrt{n}\beta(n/2, ½) (1 + t^2/n)^{n+1/2} dt \quad (-\infty \text{ to } x)$$

We use the table values to find the cumulative probability associated with a particular t statistic.

df/α	0.9	0.5	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	1	2	3.078	6.314	12.706	31.821	64	637
2	0.142	0.816	1.386	1.886	2.92	4.303	6.965	10	31.598
3	0.137	0.765	1.25	1.638	2.353	3.182	4.541	5.841	12.929
4	0.134	0.741	1.19	1.533	2.132	2.776	3.747	4.604	8.61
5	0.132	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.718	1.134	1.44	1.943	2.447	3.143	3.707	5.959
7	0.13	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.13	0.706	1.108	1.397	1.86	2.306	2.896	3.355	5.041
9	0.129	0.703	1.1	1.383	1.833	2.263	2.821	3.25	4.781

10	0.129	0.7	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.697	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.695	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.694	1.079	1.35	1.771	2.16	2.65	3.012	4.221
14	0.128	0.692	1.076	1.345	1.761	2.145	2.624	2.977	4.14
15	0.128	0.691	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.69	1.071	1.337	1.746	2.12	2.583	2.921	4.015
17	0.128	0.689	1.069	1.333	1.74	2.11	2.567	2.898	3.965
18	0.127	0.688	1.067	1.33	1.734	2.101	2.552	2.878	3.922
19	0.127	688	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.85
21	0.127	0.686	1.063	1.323	1.721	2.08	2.518	2.831	3.819
22	0.127	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.685	1.06	1.319	1.714	2.069	2.5	2.807	3.767
24	0.127	0.685	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.684	1.058	1.316	1.708	2.06	2.485	2.787	3.725
26	0.127	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.137	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.69
28	0.127	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.649
30	0.127	0.683	1.055	1.31	1.697	2.042	2.457	2.75	3.656
40	0.126	0.681	1.05	1.303	1.684	2.021	2.423	2.704	3.551
80	0.126	0.679	1.046	1.296	1.671	2	2.39	2.66	3.46
120	0.126	0.677	1.041	1.289	1.658	1.98	2.358	2.617	3.373
Infini	0.126	0.674	1.036	1.282	1.645	1.96	2.326	2.576	3.291

The table of t-distribution.

Ex.1: If t_n is a random variable which follows t-distribution with n d.f. then compute (i) $P(t_{26} < 1.706)$ (ii) $P(t_{10} > 3.169)$ (iii) find k such that $P(t_{17} > k) = 0.2$ (iv) find c such that $P(t_{15} < c) = 0.1$

(i) $P(t_{26} < 1.706) = 1 - P(t_{26} > 1.706) = 1 - 0.1 = 0.9$
(ii) $P(t_{10} > 3.169) = 0.01$
(iii) From table we have $P(t_{17} > 1.333) = 0.2$. So $k = 1.333$
(iv) Since given probability is less than 0.5 the ordinate c is negative. We have to find k such that tab values are $P(t_{15}>k)=0.1$, then $c = -k$ (since the curve is symmetric , so the tail area are equal).
From table, $k=1.753$. hence $c = -1.753$.

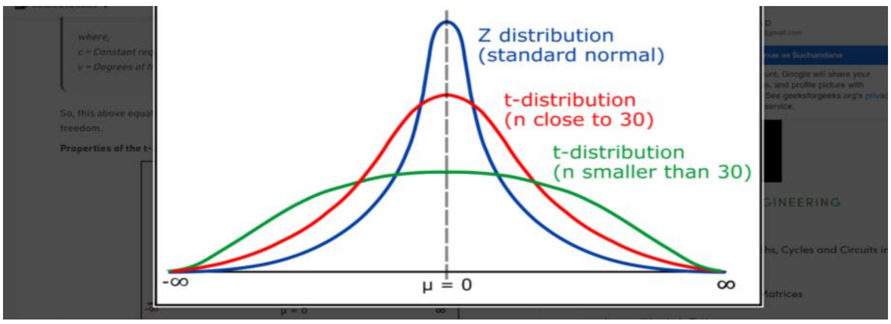
Ex.2: If t_n is a random variable which follows t-distribution with n d.f. then compute (i) $P(|t_{10}| > 1.812)$ (ii) $P(|t_8| < 2.306)$

Solution:

Here we will refer two sided probabilities from the table.

(i) $P(|t_{10}| > 1.812) = 0.1/2 = 0.05$
(ii) $P(|t_8| < 2.306) = 1 - P(|t_8| > 2.306) = 1 - 0.05/2 = 1 - .025 = 0.975$

Comparison between Normal curve and corresponding t-curve:



The difference is that the t distribution is a leptokurtic but lies below the peak of the normal distribution. As the number of d.f. in t-distribution increases, the peak goes very near to the peak of the normal distribution.

Application of t-distribution:

1. t-test for significance has a number of single mean, population variance being unknown.
2. t-test for significance of the difference between two sample means, the population variances being equal but unknown.

F-DISTRIBUTION:

Like t-distribution, Snedecor's F-distribution is useful in statistical inference. F-statistic is the ratio of two independent chi-square variates by their respective degrees of freedom. This statistic follows G.W.Snedecor's F-distribution with (n.m) d.f.

Definition:

If X is a χ^2 variate with n degrees of freedom and Y is another χ^2 variate with m degrees of freedom, then F- statistic is defined as

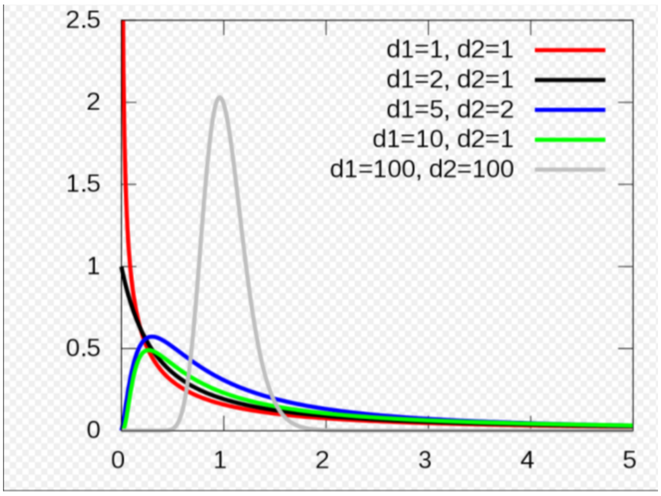
$$F = (X/n)/(Y/m)$$

And it ia said to follow F distribution with (n,m) d.f. and (n,m) are the parameters of the distribution . the p.d.f. of F distribution is given by

$$f(F) = [(n/m)^{n/2}\ (F)^{n/2-1}]/\beta(n/2\ ,\ m/2)[1 + (n/m)F]^{(n+m)/2}\ ,\quad F > 0$$

$$= 0\ ,\qquad \text{otherwise.}$$

Symbolically , we write F ~ F(n,m)



The above graph shows the p.d.f. of F for different values of (n,m).

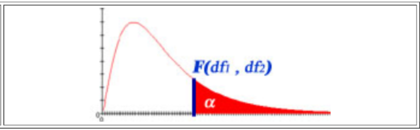
Properties of F(n,m):

1. Mean is m/m-2.
2. Mean is always less than one.
3. Mode is [m(n-2)]/[(m+2)n] which is less than one.
4. It is positively skewed as mean – mode > 0.
5. Variance is $[2m^2(n+m-2)]/n(m-2)^2(m-4)$. Variance exists only if m > 4.
6. If F ~ F(n,m) then 1/F ~ F(m,n) , i.e., reciprocal of F is also F, only the degrees of freedom are interchanged.
7. If F ~ F(n,n), then the median is one.

Note:

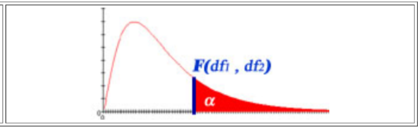
1. If t ~ t_n then t^2 follows F(1,n)
2. Since reciprocal of F is also F, the degrees of freedom are only interchanged.
3. P(F(n,m)> c) = α then P(F(m,n) > 1/c) = 1 – α
4. Table provides values of a such that P(F > a) = α for α = 0.05 or 0.01.

F Table for α = 0.05	



/	df_1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df_2=1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

F Table for α = 0.01



/	df_1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df_2=1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847	6106.321	6157.285	6208.730	6234.631	6260.649	6286.782	6313.030	6339.391	6365.864
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000

Ex.2: Find (i) $P[F(24,20) > 2.03]$,   (ii) $P[F(20, 24) < 0.49]$

Solution:

From the table we can find these values:

(i) $P[F(24,20) > 2.03] = 0.05$

(ii) $P[F(20, 24) < 0.49] = P[F(24, 20) > 1/0.49] = P[F(24, 20) > 2.04] = 0.05$