BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

innovate achieve lead

BITS Pilani

Session No 6
Probability Distributions

(24th/25th June 2023)

Session : 6 Agenda

Distributions:-

- Bernoulli
- Binomial
- Poisson
- Normal distributions.
- Introduction to t-distribution
- F-Distribution and
- Chi Square distributions

Bernoulli Distribution

Definition

A random variable 'X' is said to have Bernouli distribution if its probability mass function is given by

$$p(x) = \begin{cases} p^x \, q^{1-x}, \; x = 0,\, 1 \\ 0, \; \text{elsewhere} \end{cases}$$

Mean & Variance

Binomial Distribution

Definition

A random variable 'X' is said to have Binomial distribution if its probability mass function is given by

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0, 1, 2, 3, \ldots, n \\ 0, & \text{elsewhere} \end{cases}$$

Binomial Distribution

- Binomial distribution is a discrete probability distribution.
- Binomial distribution will be applied under the following experimental conditions

1) The number of trails (n) is finite

2) The trails are independent of each other

3) The probability of success p is constant for each trail.

4) Each trail results in two mutually exclusive events known as success and failure.

Mean & variance

$$P(x) = {}^nC_x\, p^x q^{n-x}, \quad x = 0, 1, 2, \dots n$$

$$\text{mean} = \mu = E(x) = \sum x\, P(x)$$

$$= \sum x\, {}^nC_x\, p^x q^{n-x} = \sum x \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

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

$$= np \sum_{x=1}^{n} \underbrace{\frac{(n-1)!}{(x-1)!\,[(n-1)-(x-1)]!}}_{(n-1)C_{x-1}} p^{x-1} q^{(n-1)-(x-1)}$$

$$= np\,(q+p)^{n-1}$$

$$(q+p)^n = {}^nC_0\, q^n \cdot p^0 + {}^nC_1\, q^{n-1} \cdot p + \dots + {}^nC_n\, q^0 p^n$$

$$= np$$

innovate achieve lead

Variance $\sigma^2 = E(x^2) - [E(x)]^2$

$E(x^2) = \sum x^2 p(x) = \sum x^2\, {}^nC_x\, p^x q^{n-x}$

$= \sum [x(x-1) + x]\, {}^nC_x\, p^x q^{n-x}$ $\quad \sum x p(x) \rightarrow$ mean $= np$

$= \sum x(x-1) \cdot {}^nC_x\, p^x q^{n-x} + \sum x\, {}^nC_x\, p^x q^{n-x}$

$= \sum x(x-1) \dfrac{n!}{x!\,(n-x)!} p^x q^{n-x} + np$

$= \sum \dfrac{n(n-1)(n-2)!}{(x-2)!\,[(n-2)-(x-2)]!} p^2 p^{(x-2)} \cdot q^{(n-2)-(x-2)} + np$

$= n(n-1)p^2 \sum {}^{(n-2)}C_{x-2}\, p^{x-2} q^{n-x} + np$

$= n(n-1)p^2 + np$

Variance $= \sigma^2 = n(n-1)p^2 + np - (np)^2$

$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2}$

$= np(1-p) = npq$

Binomial Distribution

4.11 Which conditions for the binomial distribution, if any, fail to hold in the following situations?

(a) The number of persons having a cold at a family reunion attended by 30 persons.

Ans: Getting cold is not independent. So it is not binomial

(b) Among 8 projectors in the department office, 2 do not work properly but are not marked defective. Two are selected and the number that do not work properly will be recorded

Ans: Probability of success is not same in each trial. So this is not binomial

Binomial Distribution

You're a telemarketer selling service contracts for Macy's. You've sold 20 in your last 100 calls (**p = .20**). If you call **12** people tonight, what's the probability of

A. No sales?

B. Exactly 2 sales?

C. At most 2 sales?

D. At least 2 sales?

Binomial Distribution Solution*

***n* = 12, *p* = .20**

A. $p(0) =$ **.0687**

B. $p(2) =$ **.2835**

C. $p(\text{at most } 2) = p(0) + p(1) + p(2) = 0.0687 + 0.2062 + 0.2835$

$= 0.\mathbf{5584}$

D. $p(\text{at least } 2) = p(2) + p(3) \ldots + p(12)$

$= 1 - [p(0) + p(1)] = 1 - .0687 - 0.2062 = 0.\mathbf{7251}$

Understandings

Phrase	*Math Symbol*
"at least"	$\geq$
"more than" or "greater than"	$>$
"fewer than" or "less than"	$<$
"no more than"	$\leq$
"exactly"	$=$

Binomial Distribution

If a coin is tossed 6 times, what is the probability of getting 2 or fewer heads?

P(x ≤ 2) = ∑p(x) = P(0)+ P(1)+ P(2)

$$P(X = 0) = \binom{6}{0}(0.5)^0(0.5)^6 = \frac{6!}{6!\ 0!}(0.5)^6 = 0.015625$$

$$P(X = 1) = \binom{6}{1}(0.5)^1(0.5)^5 = \frac{6!}{5!\ 1!}(0.5)^6 = 0.09375$$

$$P(X = 2) = \binom{6}{2}(0.5)^1(0.5)^4 = \frac{6!}{4!\ 2!}(0.5)^6 = 0.078125$$

P(x ≤ 2) = ∑p(x) = 0.015625 + 0.09375 + 0.078125 = 0.1875

Example

The probability that a bomb dropped from a plane will strike the target is 1/5. if six such bombs are dropped.

Find the probability that

1) exactly two bombs hit the target

2) At least two will hit the target

Example

An unbiased dice is thrown 5 times and occurrence of 1 or 6 considered as success find,

1) Probability of exactly one success
2) Probability of at least 4 success
3) Probability of at least one success
4) Mean and variance

Example

- Find the binomial distribution if the mean is 4 and variance is 3

Additional problem

In a large number of parts manufactured by a machine , the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples , how many would be expected to contain atleast 3 defective parts.

Sol: Given n=20;

$$\text{mean} = \mu = 2, \text{ but } \mu = np \Rightarrow 2 = 20p \therefore p = 0.1$$

$$\therefore P(x) = 20c_x (0.1)^x (0.9)^{20-x}$$

- Therefore the probability that there are atleast 3 defective parts in a sample of 20 is

$$P(X \geq 3) = 1 - P(X < 3) = 1 - [P(0) + P(1) + P(2)]$$

$$\therefore P(X \geq 3) = 0.323$$

- Hence out of 1000 samples we get 0.323*1000=323 samples

Binomial Distribution Recall

A **binomial experiment** is a probability experiment that satisfies the following conditions.

1. The experiment is repeated for a fixed number of trials, where each trial is **independent** of other trials.
2. There are only two possible outcomes of interest for each trial. The outcomes can be classified as a success (*S*) or as a failure (*F*).
3. The probability of a success *P* (*S*) is the same for each trial.
4. The random variable *x* counts the number of successful trials.

Poisson Distribution

Definition

A random variable 'X' is said to have Poisson distribution if its probability mass function is given by

$$p(x) = \begin{cases} e^{-\lambda} \dfrac{\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, \text{ elsewhere} \end{cases}$$

- Poisson distribution is the discrete probability distribution of a discrete random variable X, which may not have upper bound. It is defined for non negative values of x as follows

Cont…

- Poisson distribution is suitable for rare events for which the probability of occurrence ‘p’ is very small and the number of trials ‘n’ is very large.

- Also binomial distribution can be approximated by Poisson distribution when $n \to \infty$ and $p \to 0$ such that $\lambda = np$ = constant

Example

If the probability of a bad reaction from a certain injection is 0.001.

- Determine the chance that out of 2000 individuals more than two will get a bad reaction.

Solution

Example

Police records show that number of accident victims died in road traffic accidents is 0.1%. What is the probability that among 500 randomly selected accident victims

(i) none have died?

(ii) at least 3 have died

(iii) between 2 and 6 have died

Example:

Example 6: A hospital switch board receives an average of 4 emergency calls in a 10 minute interval. What is the probability that

(i) there are at most 2 emergency calls in 10 minute interval

(ii) there are exactly 3 emergency calls in a 10 minute interval

Ans: Given Mean = 4calls. We have

$$P(x) = \frac{(4)^{x} e^{-4}}{x!}$$

(i) P(at most 2 calls) = P(x ≤ 2) = P(0)+P(1)+P(2)=0.2381

(ii) P(Exactly 3 calls)= P(x = 3) = 0.1954

Example:

Example 2: The mean number of power outages in the city of Bangalore follows a Poisson variate and is 4 per year. Find the probability that in a given year,

a) there are exactly 3 outages, b) there are more than 3 outages.

Mean and Variance

innovate achieve lead

Mean:

$$\boxed{\mu = \lambda}$$

For a Poisson random variable, the variance and mean are the same!

Variance and Standard Deviation

$$\boxed{\sigma^2 = \lambda}$$

$$\boxed{\sigma = \sqrt{\lambda}}$$

where λ= expected number of occurrences in a given experiment.

Example

4.55 In a factory, 8% of all machines break down at least once a year. Use the Poisson approximation to the binomial distribution to determine the probabilities that among 25 machines (randomly chosen in the factory):

(a) 5 will break down at least once a year;

(b) at least 4 will break down once a year;

(c) anywhere from 3 to 8, inclusive, will break down at least once a year.

Suggested Problems

Given that a switch board of a consultant's office receives on the average 0.6 calls per minute, find the probabilities that

(a) in a given minute there will be at least 1 call;

(b) in a 4-minute interval there will be at least 3 calls.

Solution:

a) here

$$\lambda = 0.6 \text{ and } t = 1 \therefore \quad p(x, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x\,!}$$

$$p(x, \lambda t) = \frac{e^{-0.6}(0.6)^x}{x\,!}$$

$$\mathrm{P}(\mathrm{X} \geq 1) = 1 - \mathrm{P}(\mathrm{X} < 1) = 1 - \mathrm{P}(0)$$

$$= 1 - \mathrm{e}^{-0.6} = 0.4512$$

b)

$$\lambda = 0.6 \text{ and } t = 4 \therefore \quad p(x, \lambda t) = \frac{e^{-2.4}(2.4)^x}{x\,!}$$

$$\mathrm{P}(\mathrm{X} \geq 3) = 1 - \mathrm{P}(\mathrm{X} < 3) = 1 - [\mathrm{P}(0) + \mathrm{P}(1) + \mathrm{P}(2)]$$

$$= 0.4303$$

Exercise:

The number of customers arriving at a cafeteria at an average rate of 0.3 per minute.

(a) Find the probability that exactly 2 customers arrive in a 10-minute span.

(b) Find the probability that 2 or more customers arrive in a 10-minute span.

(c) Find the probability that exactly one customer arrives in a 5-minute span and one customer arrives in the next 5-minute span.

Exercise:

The average number of trucks arriving on any one day at a truck depot in a certain city is known to be 12. what is the probability that on a given day fewer than nine trucks will arrive at this depot?

Exercise:

A certain kind of sheet metal has on the average, five defects per 10 square feet. If we assume a Poisson distribution, what is the probability that a 15-square-foot sheet of the metal will have at least six defects?

Exercise:

The number of flaws in a fiber optic cable follows a poisson process with an average of 0.6 per 100 feet.

a) Find the probability of exactly 2 flaws in 200 foot cable.

b) Find the probability of exactly 1 flaw in first 100 feet and exactly 1 flaw in the second 100 feet.

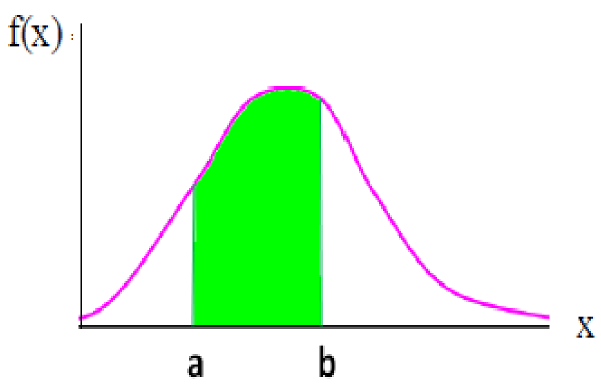
Exercise:

Suppose that on an average 1 out of 10000 houses catch fire in a year in a district. If there are 2000 houses in the district, find the probability that exactly 5 house will catch fire during that year.

RECALL: Continuous Random Variables

A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.

The probability of the random variable assuming a value within some given interval from x1 to x2 is defined to be the area under the graph of the **probability density function** between x1 and x2



Example:

- Height of students in a class
- Amount of ice tea in a glass
- Change in temperature throughout a day
- Price of a car in next year

Continuous Probability Distributions:

Normal Distribution:

Normal distribution is a general distribution. Any information such as heights, weights, lengths, widths, sizes, profits, prices, wages, time etc. containing sufficiently large data with mean and variance follows normal distribution. The distribution of errors of repeated measurements follows normal distribution.

Normal Distribution

A continuous random variable X which assumes all possible values in the entire real space, i.e., -∞<x<∞ is said to follow normal distribution with two parameters such as mean μ and variance σ^2 if it has the probability density function given by

$$f(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$-\infty < x < \infty$$

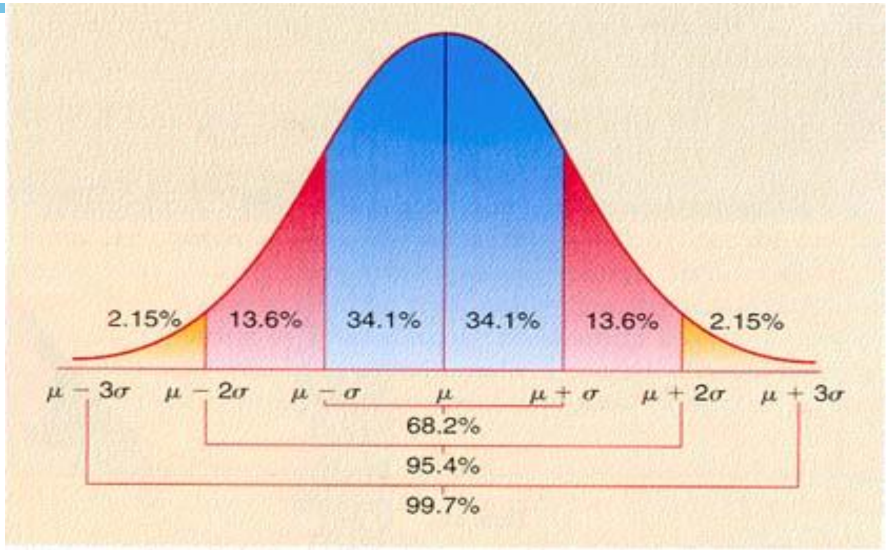
$$-\infty < \mu < \infty\,,\sigma > 0$$

Normal Distribution

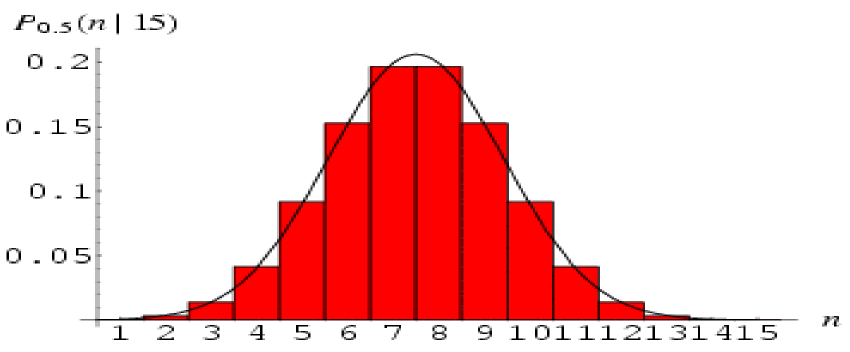
Change in Mean determines the shift in the distribution
Change in the deviation determines the spread of the data points

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

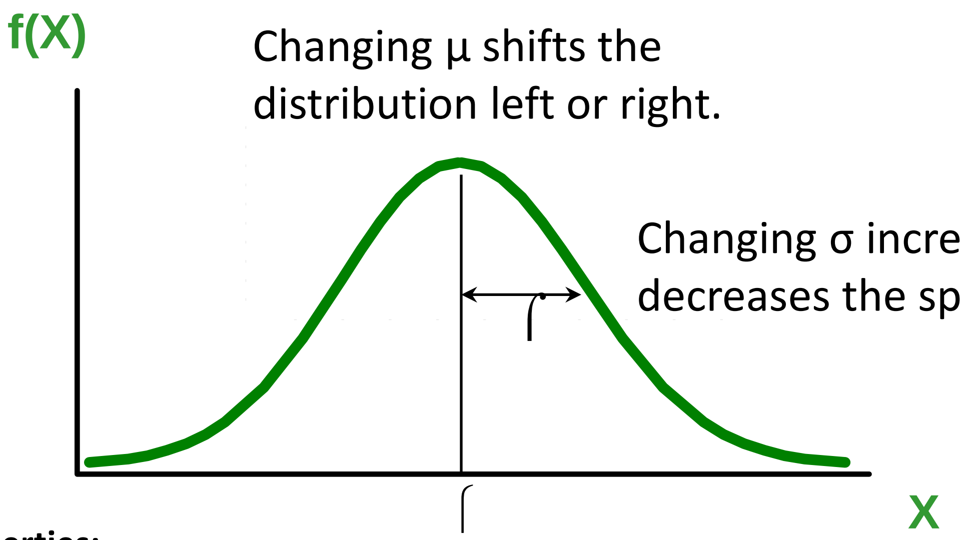
μ = mean
σ = standard deviation
π = 3.14159
e = 2.71828



This is a bell shaped curve with different centers and spreads depending on μ and σ



Normal Distribution



Properties:

1. Normal curve is bell shaped and symmetric about the mean
2. Mean = Mode = Median
3. Total area under normal curve is equal to 1
4. Normal curve approaches but never touches the x axis as it extends farther and farther away from the mean

Normal Distribution

E(X)=μ = $\int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$

Var(X)=σ² = $\int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx) - \mu^2$

Standard Deviation(X)=σ

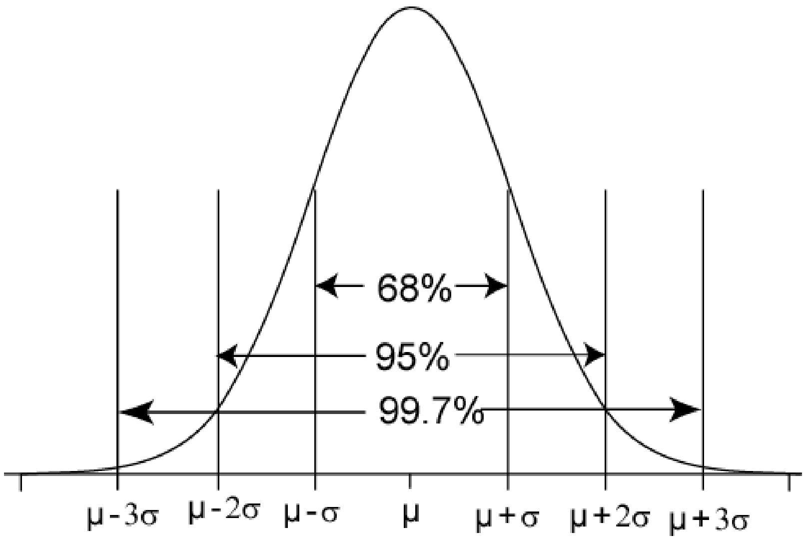
No matter what μ and σ are, the area between μ-σ and μ+σ is about 68%; the area between μ-2σ and μ+2σ is about 95%; and the area between μ-3σ and μ+3σ is about 99.7%. Almost all values fall within 3 standard deviations.

68-95-99.7 Rule for Normal Distributions

68% of the AUC(Area under curve) within ±1σ of μ

95% of the AUC within ±2σ of μ

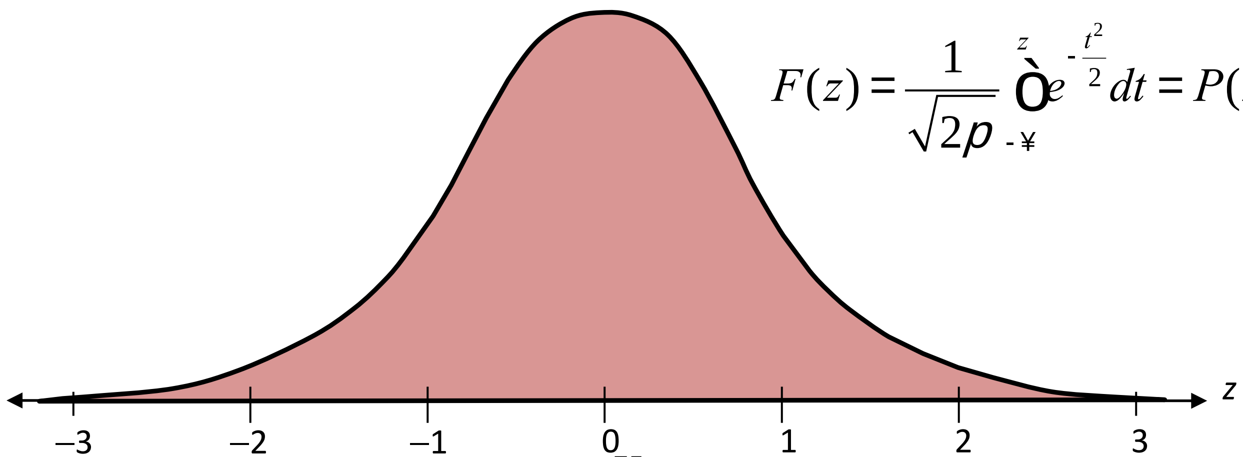
99.7% of the AUC within ±3σ of μ



Standard Normal Distribution

The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt = P(Z \le z)$$



$$Z = \frac{X - \mu}{\sigma}$$

All normal distributions can be converted into the standard normal curve by subtracting the mean and dividing by the standard deviation

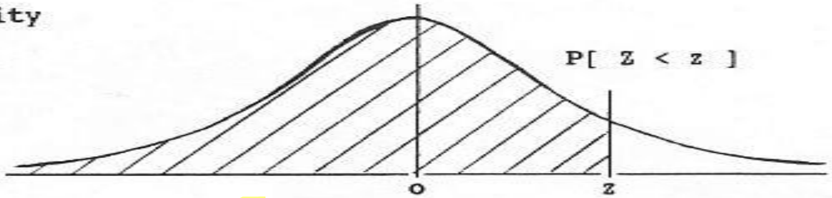
STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z

i.e.

$$P[\ Z < z\ ] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\ dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Problem

If a random variable has the standard normal distribution, find the probability that it will take on a value

a)less than 1.75 = $P(Z < 1.75) = F(1.75) = 0.9599$

b)less than -1.25 = $P(Z < -1.25) = F(-1.25) = 1 - F(1.25) = 0.1056$

c)greater than 2.06 = $P(Z > 2.06) = 1 - F(2.06) = 0.0197$

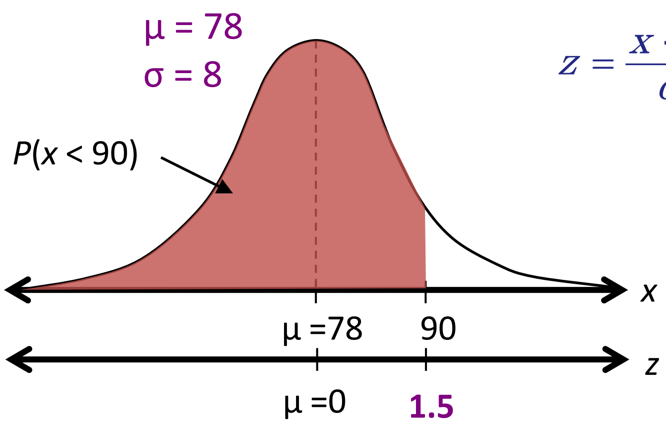
d)greater than -1.82 = $P(Z > -1.82) = 1 - F(-1.82) = 1 - 0.0344 = 0.9656$

$P(Z > -1.82) = 1 - F(-1.82) = 1 - (1 - F(1.82) = F(1.82) = 0.9656$

Probability and Normal Distributions

Example:

The average on a statistics test was 78 with a standard deviation of 8. If the test scores are normally distributed, find the probability that a student receives a test score less than 90.



$$z = \frac{x - \mu}{\sigma} = \frac{90 - 78}{8}$$

$$= 1.5$$

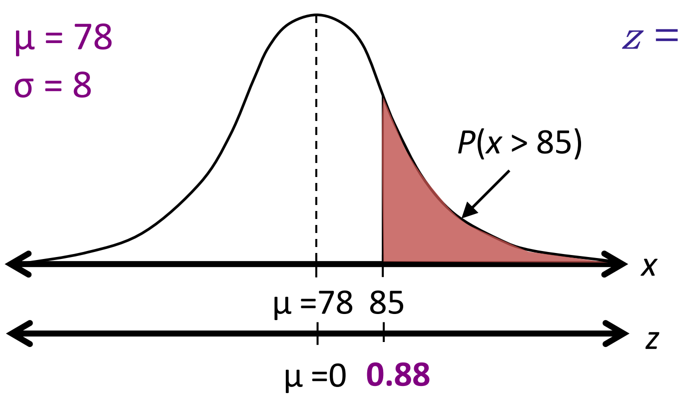
The probability that a student receives a test score less than 90 is 0.9332.

$P(x < 90) = P(z < 1.5) = 0.9332$

Probability and Normal Distributions

Example:

The average on a statistics test was 78 with a standard deviation of 8. If the test scores are normally distributed, find the probability that a student receives a test score greater than 85.



$$z = \frac{x - \mu}{\sigma} = \frac{85 - 78}{8}$$

$$= 0.875 \approx 0.88$$

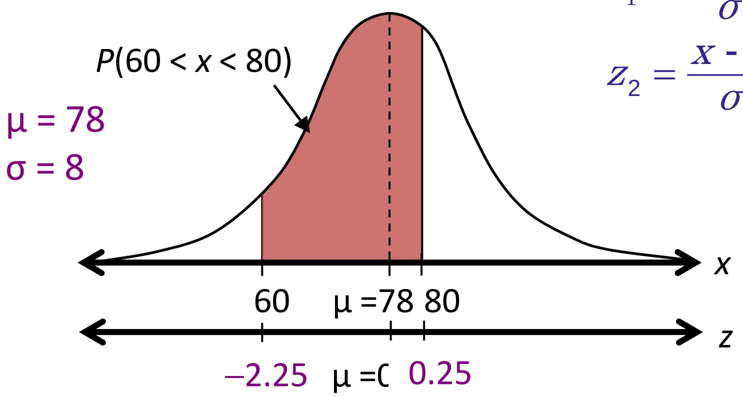
The probability that a student receives a test score greater than 85 is 0.1894.

$P(x > 85) = P(z > 0.88) = 1 - P(z < 0.88) = 1 - 0.8106 = 0.1894$

Probability and Normal Distributions

Example:

The average on a statistics test was 78 with a standard deviation of 8. If the test scores are normally distributed, find the probability that a student receives a test score between 60 and 80.



$$z_1 = \frac{x - \mu}{\sigma} = \frac{60 - 78}{8} = -2.25$$

$$z_2 = \frac{x - \mu}{\sigma} = \frac{80 - 78}{8} = 0.25$$

The probability that a student receives a test score between 60 and 80 is 0.5865.

$$P(60 < x < 80) = P(-2.25 < z < 0.25) = F(0.25) - F(-2.25) = 0.5987 - 0.0122 = 0.5865$$

Problem

The time for oil to percolate to all parts of an engine can be treated as a random variable having a normal distribution with mean 20 seconds. Find its standard deviation if the probability is 0.25 that will take a value greater than 31.5 seconds.

Given : Mean μ = 20 seconds, P(X > 31.5) = 0.25

σ = ?

$$z = \frac{x - m}{s} = \frac{31.5 - 20}{s}$$

$$s = \frac{11.5}{z} = \frac{11.5}{0.675} = 17.04$$

Ex.

Monthly Salary X in a big organisation is normally distributed with mean Rs 3000 and standard deviation of Rs.250 what should be the minimum salary of a worker in this organisation so that the probability that he belongs to top 5%workers?

Problem

Butterfly-style valves used in heating and ventilating industries have a high flow coefficient. Flow coefficient can be modeled by a normal distribution with mean 496 C_v and standard deviation 25C_v. Find the probability that a valve will have a flow coefficient of

a) atleast 450C_v

$P(X > 450) = P[(450 – 496) /25] = P(Z > -1.84) = 1 – F(-1.84) = 1- 00329 = .9671$

b. between 445.5 and 522C_v

$P(445.5 <X <522) = P(-2.02 <Z <1.04) = F(1.04) – F(-2.02)= 0.8508 – 0.0217 =$

0.8291

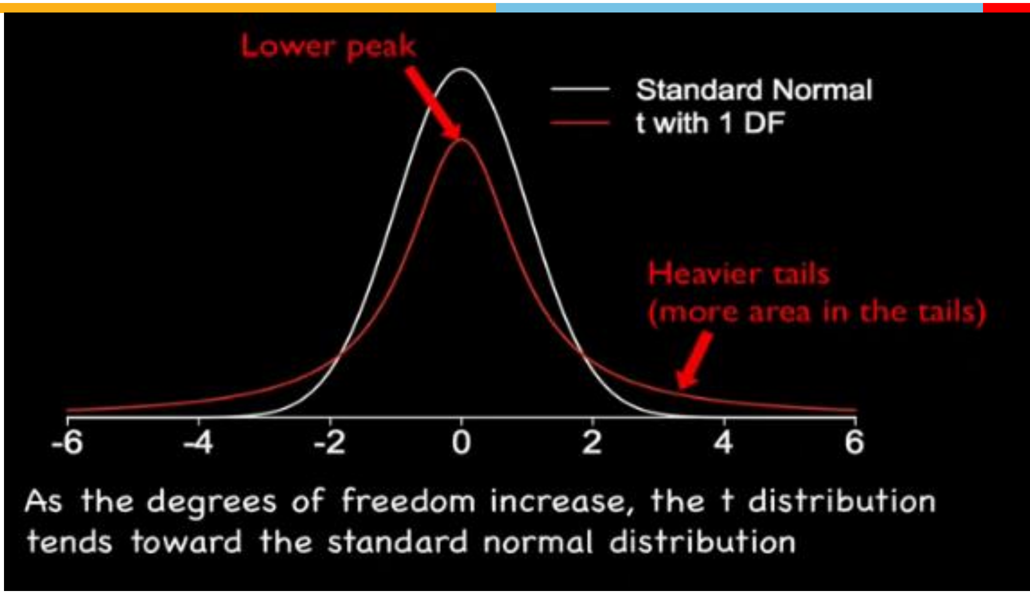
Normal Approximation to Binomial Distribution:

If n tends to infinity (preferably n>30) and neither p nor 1-p is so small, that is p and 1-p both are large preferably np>15 and n(1-p)>15 then binomial probabilities can be approximated by normal probabilities using continuity correction and taking

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}}$$

Where X is binomial random variable approximated as normal random variable.

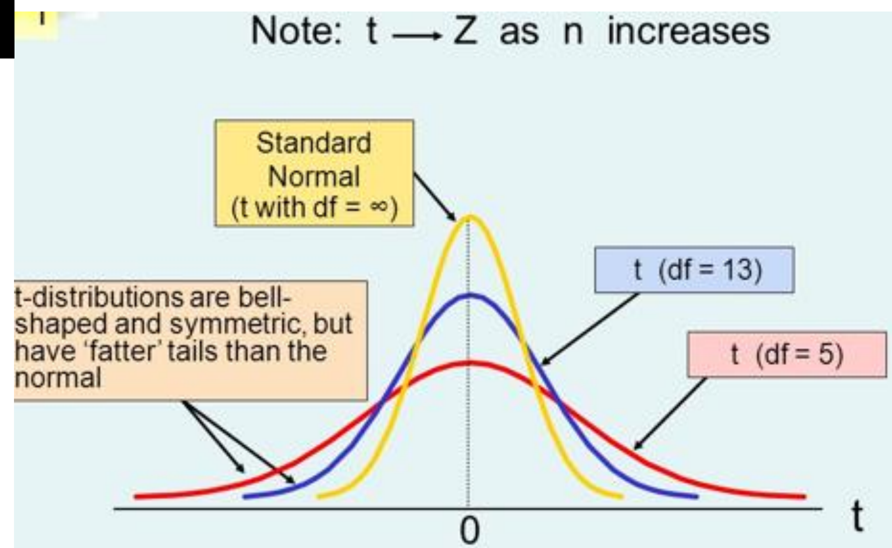
Introduction to t-distribution



The pdf of the t distribution with ν degrees of freedom:

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

for $-\infty < t < \infty$



Introduction to t-distribution (Sampling distribution of mean when σ unknown)

innovate achieve lead

The t-distribution, also known as the **Student's t-distribution**, is a type of probability distribution that is similar to the normal distribution with its bell shape but has heavier tails.

It is used for estimating population parameters for **small sample sizes when**

t - Distribution When σ is unknown and n < 30.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad with\ (n-1)\ d.o.f.$$

Where sample S.D. is calculated by formula $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

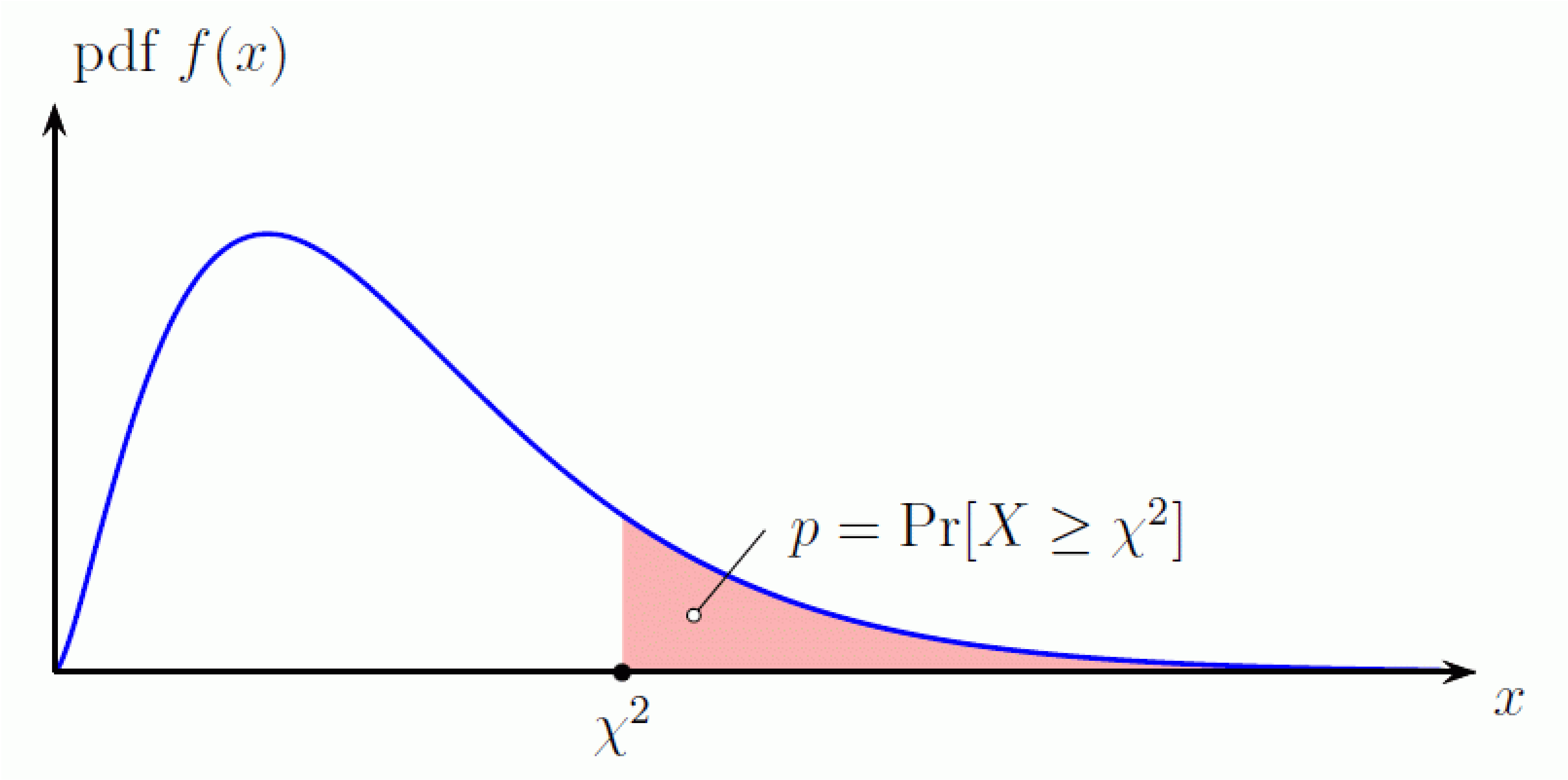
Introduction to Chi- Square(χ^2) distribution

- Chi-Square distribution pdf is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\vartheta/2}\,\Gamma(\frac{\vartheta}{2})} x^{\frac{\vartheta}{2}-1} e^{-\frac{x}{2}} & for\ x > 0 \\ 0 & otherwise \end{cases}$$

where ϑ is the parameter of distribution, also known as degrees of freedom.

- Introduction to Chi- Square(χ^2) distribution



Introduction to Chi- Square(χ^2) distribution

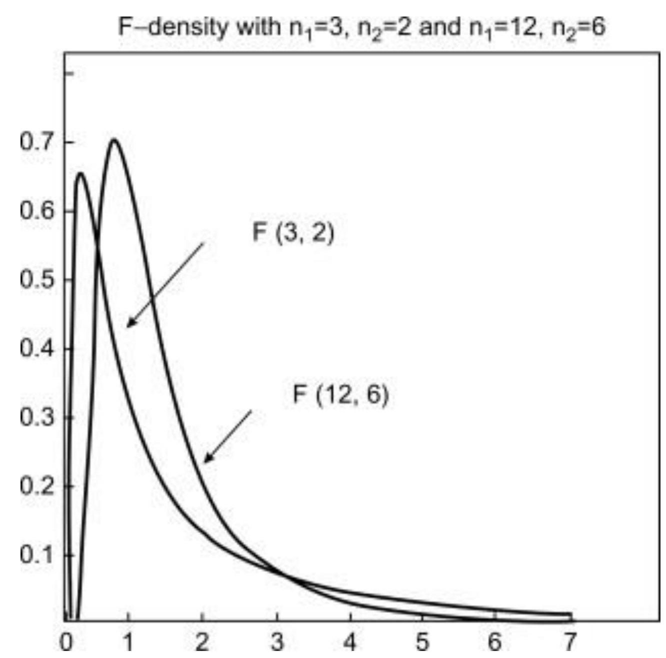
- Chi-Square distribution curve is not symmetrical, and hence not a normal curve.
- Chi-Square varies from 0 to ∞(Curve lies entirely in first quadrant).
- It depends only on degrees of freedom ϑ.
- It is very important in estimation and hypothesis testing.
- It is used in sampling distributions of the sample variance, analysis of variance.
- It is used as a measure of goodness of fit.

Introduction to F-distribution (Sampling distribution of the ratio of two sample variances)

- The pdf for a random variable $X \sim F(n_1, n_2)$ is given by

$$f(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} \quad , x > 0$$

- F- distribution curve lies entirely in first quadrant.
- The F- curve depends not only on the two parameters ϑ_1 and ϑ_2 but also on the order in which they are stated.



Introduction to F-distribution
(Sampling distribution of the ratio of two sample variances)

- The F-distribution was **developed by Fisher to study the behavior of two variances from random samples taken from two independent normal populations**.
- In applied problems we may be interested in knowing whether the population variances are equal or not, based on the response of the random samples.
- If S_1^2 and S_2^2 are variances of independent random sample of size n_1 and n_2 from a normal populations with variances σ_1^2 and σ_2^2, then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

which follows F –distribution with $\vartheta_1 = n_1 - 1$ and $\vartheta_2 = n_2 - 1$ d.o.f.

Thanks