M Tech(Data Science & Engineering)

Introduction to Statistical Methods [ISM]

BITS Pilani

Pilani|Dubai|Goa|Hyderabad

innovate achieve lead

BITS Pilani

Session No 5

Random Variables & Probability Distributions

(Session 5: 4th /5th June ,2022)

Session No 5 Course Handout

Contact Session	List of Topic Title	Reference
CS - 5	Probability Distributions – Binomial, Poisson and Normal Distributions	T1:Chapter 3 & 4
HW	Problems on probability distributions	T1:Chapter 3 & 4
Lab		

Agenda

- Probability Distributions
  - Bernoulli
  - Binomial
  - Poisson
  - Normal

innovate achieve lead

BITS Pilani

Probability 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

innovate achieve lead

Definition

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

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

Definition

A random variable X is said to follow binomial distribution if it assumes only non negative values and its probability mass function is given by

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

$$= 0 \quad ; \text{ otherwise}$$

n and p are two parameters of the distribution and q = 1-p

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.

Binomial Distribution

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

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

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

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

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

Example

The probability that a man aged 60 years will remain alive till 70 is 0.65.

- What is the probability that out of 10 such men at least 7 would be alive at 70?

Solution

Example

Nationalized bank sanctions short term loan for its customers. The probability of sanctioning loan by the bank is 0.53. If 5 persons are selected at random what is the probability that

a) none were sanctioned loan?

b) between 1 to 4 were sanctioned the loan?

c) more than 3 were sanctioned the loan?

Mean & variance

Binomial Distribution

All probability distributions are characterized by an expected value and a variance

If X follows a binomial distribution with parameters n and p then we write X ~ B(n, p)

$$\mu = E(X) = np$$

$$\sigma^2 = Var(X) = npq,\ q = 1 - p$$

$$\sigma = SD(X) = \sqrt{npq}$$

Note: the variance will always lie between

0*n - 0.25 *n

p(1 - p) reaches maximum at p=0.5

p(1 - p)=.025

Practice Problems

- 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

- 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

Practice Problems

1. With usual notation find p of binomial distribution if n=6 $9P(x=4)=P(x=2)$

2.

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

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 has no 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 poission distribution when $n \rightarrow \infty$ and $p \rightarrow 0$ such that $\lambda = np =$ constant

Poisson distribution

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

Poisson Distribution

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

Poisson Distribution

A car hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as Poisson variate with mean is 1.5. calculate the probability that

1) Neither a car is used
2) Some demand is refused

Problem

Average number of accidents on any day on a national highway is 1.8 . Determine the probability that the number of accidents are 1) at least one

2) at most one.

Practice Problems

- A hospital switch board receives an average of 4 emergency calls in a 10 mins interval, what is the probability that

1) There are 2 emergency calls
2) There are 3 emergency calls in an interval of 10 mins

- If a random variable x follows Poisson distribution such that

$P(x=1)=2P(x=2)$, find the mean and variance of the distribution

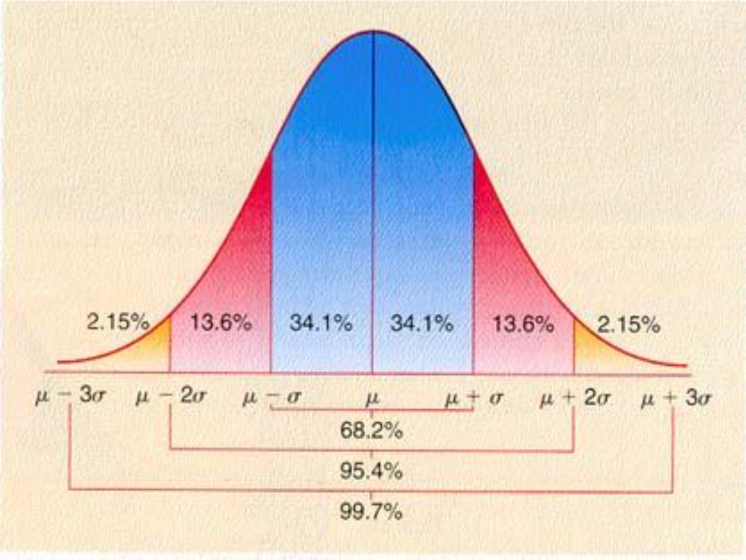
Normal Distribution

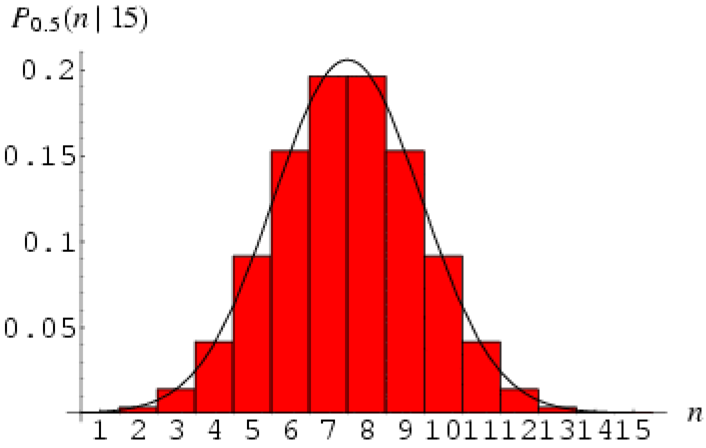
Change in Mean determines the shift in the distribution
Change in the deviation determines the spread of the data points

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

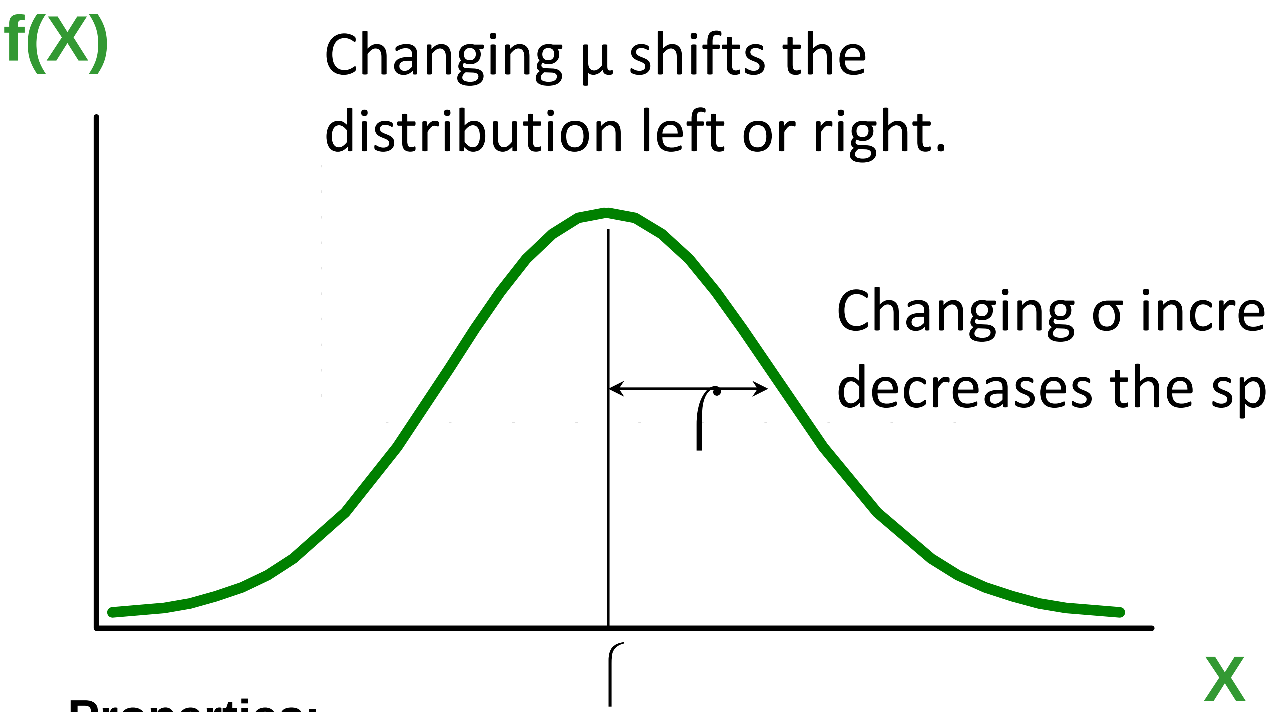
μ = mean
σ = standard deviation
π = 3.14159
e = 2.71828

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

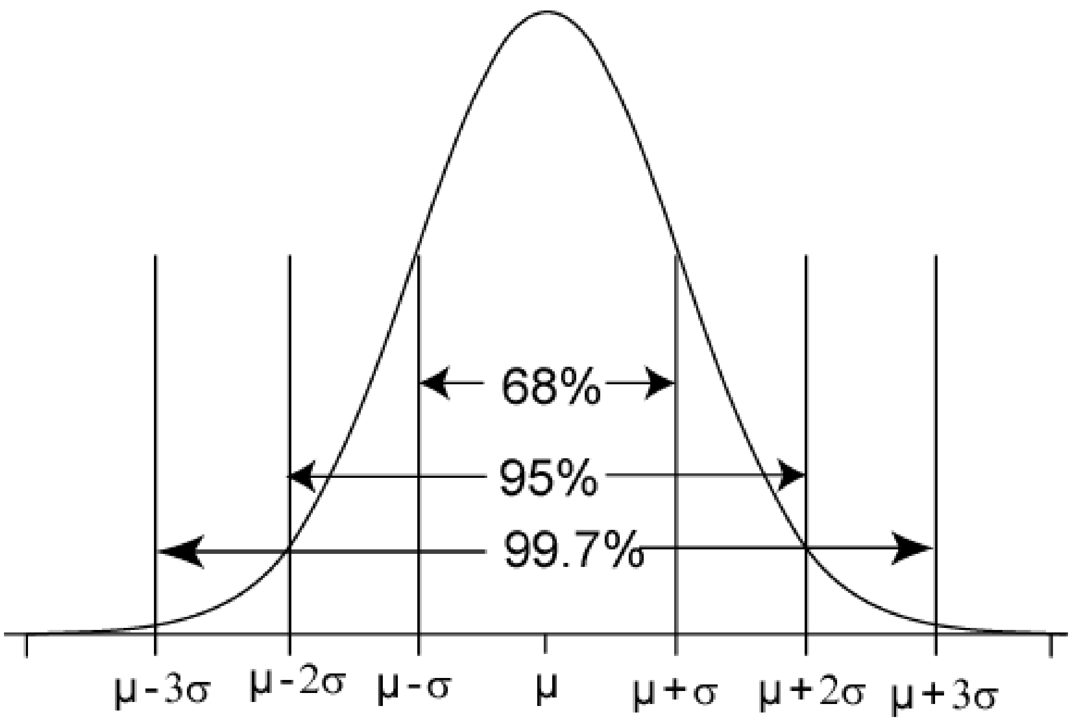
No matter what μ and σ are, the area between μ-σ and μ+σ is about 68%; the area between μ-2σ and μ+2σ is about 95%; and the area between μ-3σ and μ+3σ is about 99.7%. Almost all values fall within 3 standard deviations.

68-95-99.7 Rule for Normal Distributions

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

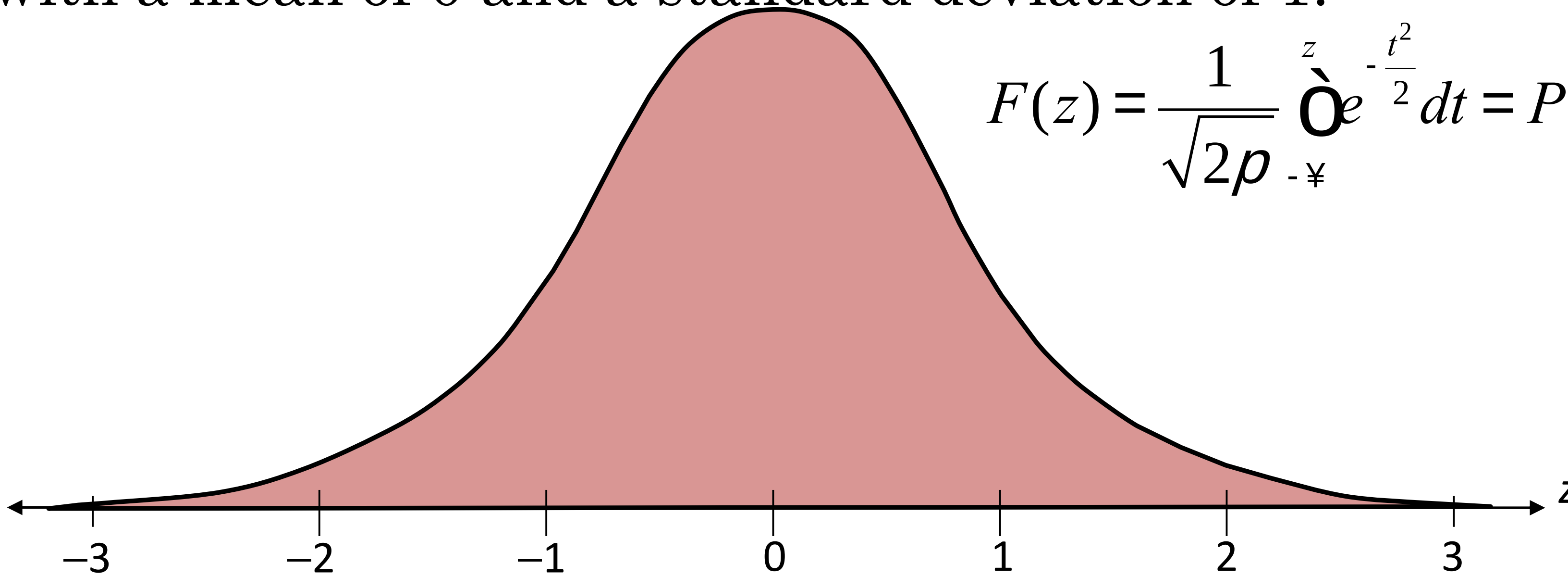
95% of the AUC within ±2σ of μ

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.



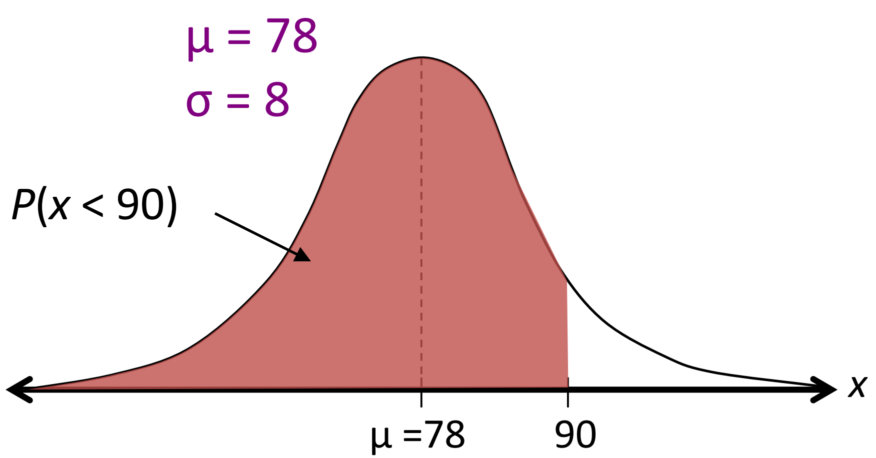
$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt = P(Z \leq z)$$

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

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

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.



$$z = \frac{x - \mu}{\sigma} = \frac{90 - 78}{8} = 1.5$$

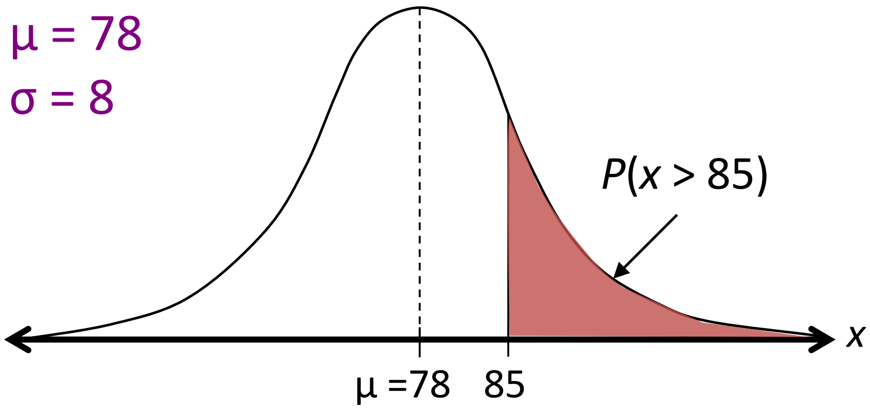
The probability that a student receives a test score less than 90 is 0.9332.

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

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 than 85.

$\mu = 78$
$\sigma = 8$



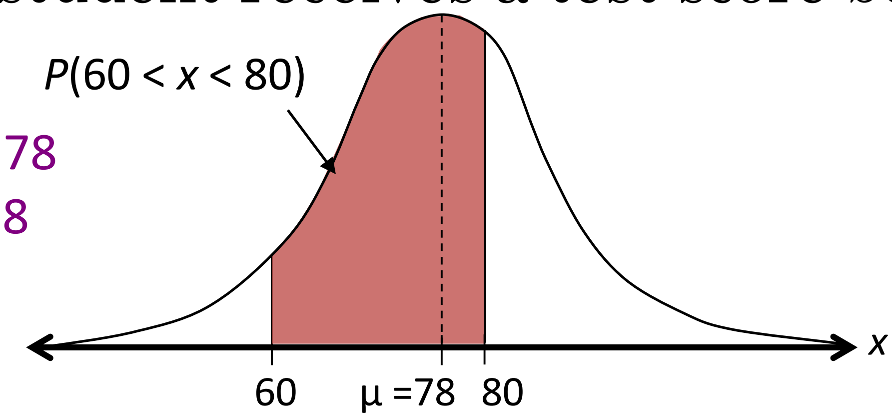
$$z = \frac{x - \mu}{\sigma} = \frac{85 - 78}{8}$$
$$= 0.875 \approx 0.88$$

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

$P(x > 85) = P(z > 0.88) = 0.1894$

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.



$\mu = 78$
$\sigma = 8$

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

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

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

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

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361

Problem 5.25(Page 150)

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 = ?

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

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

Problem 5.26(Page 150)

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

$P(X > 450) = P[(450 – 496) /25] = P(Z > -1.84) =$

b) between 445.5 and 522C_v

$P(445.5 < X < 522) = P(-2.02 < Z < 1.04) =$

Practice Problems

In an intelligence test administered to 1000 students, the average was 42 and standard deviation was 24. find the number of students

i) Exceeding the score 50

ii) Between 30 and 54

Practice Problems

Monthly salary X in a big organization is normally distributed with mean Rs 3000 and standard deviation of Rs 250.

What should be the minimum salary of a worker in this organization so that the probability that he belongs to top 5% workers?

Thanks