Course Title: Advanced Statistical Methods

Course code: MAT6001

Dr. G. Hannah Grace

Vellore Institute of Technology, Chennai

Module 2: Statistical Inference

FALL 2021

Prepared by Dr. G. Hannah Grace, Assistant Professor, Division of Mathematics, School of Advanced Sciences, VIT Chennai.
OFFERED TO : AI & ML , BIG DATA (M.Tech Engineering Program)

13/10/2021

Course content: Statistical Inference

Basic Concepts, Normal distribution-Area properties, Steps in tests of significance –large sample tests-Z tests for Means and Proportions, Small sample tests –t-test for Means, F test for Equality of Variances, Chi-square test for independence of Attributes.

Continuous Distributions

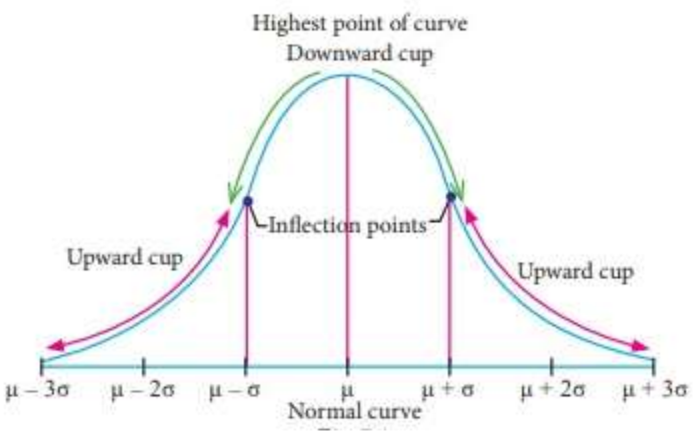
In order to have mathematical distributions suitable for dealing with quantities whose magnitudes vary continuously like weight, heights of individual, a continuous distribution is needed. Normal distribution is one of the most widely used continuous distribution.

Normal Distribution **(Gaussian distribution)**

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

$$f(x:\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad \begin{array}{c} -\infty < x < \infty, \\ -\infty < \mu < \infty, \\ \sigma > 0 \end{array}$$

Normal distribution is diagrammatically represented as follows :



Normal distribution is a limiting case of Binomial distribution under the following conditions:

(i) n, the number of trials is infinitely large, i.e. $n \to \infty$

(ii) neither p (or q) is very small,

The normal distribution of a variable when represented graphically, takes the shape of a symmetrical curve, known as the Normal Curve. The curve is asymptotic to x-axis on its either side.

Note:

(i) $B(X: n, p)$ when $n \to \infty$ and p, q are not small will become a Normal distribution.

(i) $P(X, \lambda)$ when $\lambda \longrightarrow \infty$ will become a Normal distribution.

(ii) When $X \sim N(\mu, \sigma^2)$ then $Z = \dfrac{x-\mu}{\sigma}$ then $Z \sim N(0, 1)$

i.e., $f(z) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \quad -\infty \le z \le \infty$

Z is known as a Standard Normal variate with mean 0 and variance 1.

To find probabilities of X we convert X into Z and then make use of standard normal table.

Standard Normal Distribution

A random variable Z = (X–μ)/σ follows the standard normal distribution. Z is called the standard normal variate with mean 0 and standard deviation 1 i.e Z ~ N(0,1). Its Probability density function is given by :

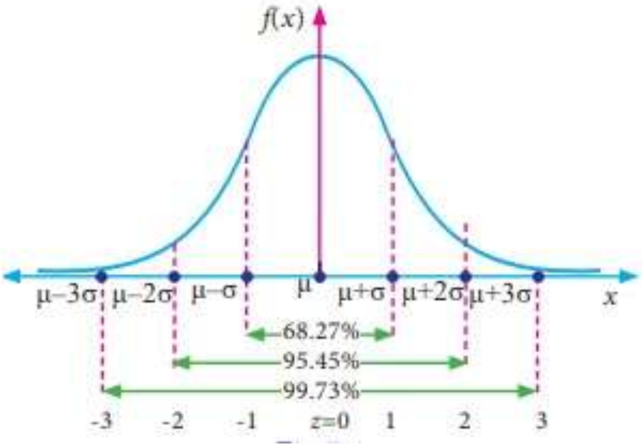
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2} / 2 \qquad -\infty < z < \infty$$

1. The area under the standard normal curve is equal to 1.

2. 68.26% of the area under the standard normal curve lies between z = –1 and Z =1

95.44% of the area lies between Z = –2 and Z = 2

99.74% of the area lies between Z = –3 and Z = 3



The normal distribution formula is based on two simple parameters—mean and standard deviation—which quantify the characteristics of a given dataset. While the mean indicates the "central" or average value of the entire dataset, the standard deviation indicates the "spread" or variation of data-points around that mean value.

Problems

1. What is the probability that a standard normal variate Z will be

(i) greater than 1.09

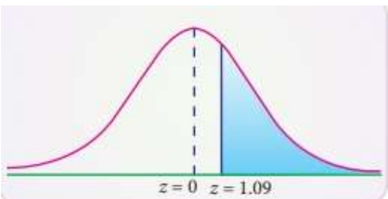
(ii) less than -1.65

(iii) lying between -1.00 and 1.96

(iv) lying between 1.25 and 2.75

(i) greater than 1.09

The total area under the curve is equal to 1 , so that the total area to the right Z = 0 is 0.5 (since the curve is symmetrical). The area between Z = 0 and 1.09 (from tables) is 0.3621

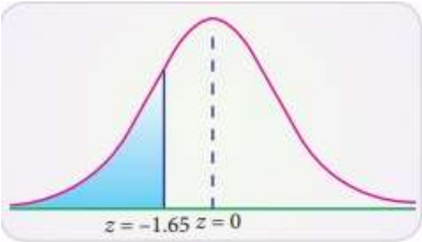


P(Z > 1.09) = 0.5000 - 0.3621 = 0.1379

The shaded area to the right of Z = 1.09 is the probability that Z will be greater than 1.09

(ii) less than –1.65

The area between -1.65 and 0 is the same as area between 0 and 1.65. In the table the area between zero and 1.65 is 0.4505 (from the table). Since the area to the left of zero is 0.5 , P(Z< 1.65) = 0.5000 – 0.4505 = 0.0495.



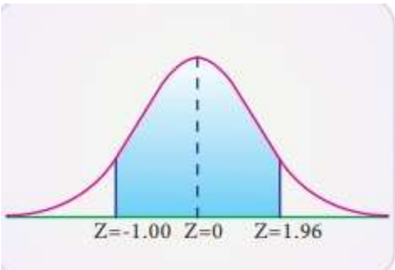
(iii) lying between -1.00 and 1.96

The probability that the random variable Z in between -1.00 and 1.96 is found by adding the corresponding areas :

Area between -1.00 and 1.96 = area between (-1.00 and 0) + area between (0 and 1.96)

$P(-1.00 < Z < 1.96) = P(-1.00 < Z < 0) + P(0 < Z < 1.96) = 0.3413 + 0.4750$ (using table value)

= 0.8163

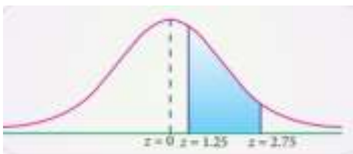


(iv) lying between 1.25 and 2.75

Area between Z = 1.25 and 2.75 = area betwn (z = 0 and z = 2.75) – area betwn (z=0 and z = 1.25)

$P(1.25 < Z < 2.75) = P(0 < Z < 2.75) - P(0 < Z < 1.25)$

= 0.4970 - 0.3944 = 0.1026



2. A sample of 125 dry battery cells tested to find the length of life produced the following resultd with mean 12 and sd 3 hours. Assuming that the data to be normal distributed , what percentage of battery cells are expected to have life

(i) more than 13 hours

(ii) less than 5 hours

(iii) between 9 and 14 hours

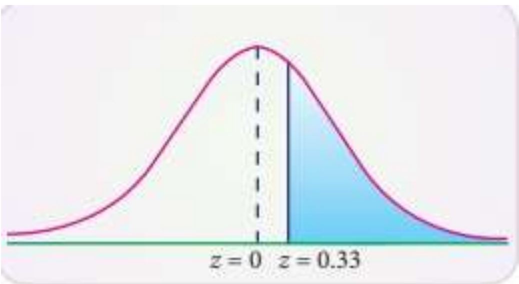
Solution :

Let X denote the length of life of dry battery cells follows normal distribution with mean 12 and sd 3 hours

(i) more than 13 hours

P(X > 13)

When X = 13



$$Z = \frac{X-\mu}{\sigma} = \frac{13-12}{3} = 0.333$$

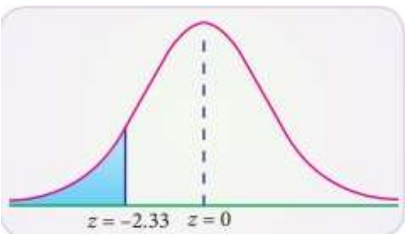
P(X > 13) = P(Z>0.333) = 0.5 – 0.1293 = 0.3707

The expected battery cells life to have more than 13 hours is

125 × 0.3707 =46.34%

(ii) less than 5 hours

P(X < 5)



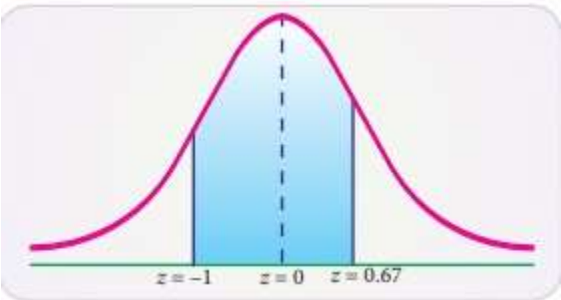
When $X = 5$ $\quad Z = \dfrac{X-\mu}{\sigma} = \dfrac{5-12}{3} = -2.333$

P(X < 5) = P(Z < –2.333) = P(Z > 2.333)

= 0.5 – 0.4901 = 0.0099

The expected battery cells life to have more than 13 hours is 125 × 0.0099 =1.23%

(iii) between 9 and 14 hours



When $X = 9$

$$Z = \frac{X-\mu}{\sigma} = \frac{9-12}{3} = -1$$

When $X = 14$

$$Z = \frac{X-\mu}{\sigma} = \frac{14-12}{3} = 0.667$$

P(9 < X < 14) = P(–1 < Z < 0.667)

= P(0 < Z < 1) + P(0 < Z < 0.667)

= 0.3413 + 0.2486

= 0.5899

18|10|21

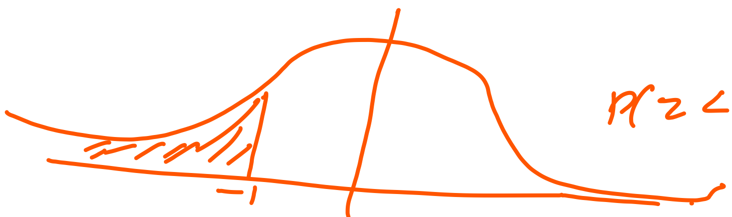
The expected battery cells life to have more than 13 hours is 125 x 0.5899 =73.73%

3. Weights of fish caught by a traveler are approximately normally distributed with a mean weight of 2.25 kg and a standard deviation of 0.25 kg. What percentage of fish weigh less than 2 kg?

μ σ

Solution :

We are given mean μ = 2.25 and standard deviation σ = 0.25. Probability that weight of fish is less than 2 kg is P(X < 2.0)



$P(Z < -1) = P(0 < Z < \infty) - P(0 < Z < -1)$

$= 0.5 - P(0 < Z < 1)$

$$Z = \frac{x-\mu}{\sigma} = \frac{2-2.25}{0.25} = \frac{-0.25}{0.25} = -1$$

When x = 20

Z = [X − μ] / σ = [2.0 − 2.25] / 0.25 = P(Z< -1.0) = P(Z > 1.0) (sym)

0.5 – 0.3413 = 0.1587

$P(x<2) = P(Z<-1)$

$= P(Z>1)$

Therefore 15.87% of fishes weigh less than 2 kg.

4. The average daily procurement of milk by village society in 800 litres with a standard deviation of 100 litres. Find out proportion of societies procuring milk between 800 litres to 1000 litres per day.

Solution :

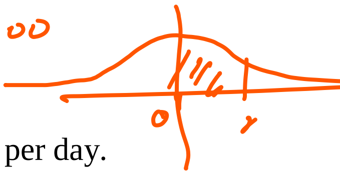
We are given mean μ = 800 and standard deviation σ = 100 . Probability that the procurement of milk between 800 litres to 1000 litres per day is

P(800 < X < 1000)

If $x = 800 \Rightarrow z = \frac{800-800}{100} = 0$

$$P\left(\frac{800-800}{100} < z < \frac{1000-800}{100}\right)$$

If $x = 1000 \Rightarrow z = \frac{1000-800}{100} = 2$



P(0 < Z < 2) = 0.4772 (table value)

Therefore 47.75 percent of societies procure milk between 800 litres to 1000 litres per day.

5. **If X is a Normal variate with mean30 and SD 5.Find P [26<X<40].**

P [26<X<40] =P $[-0.8 \le Z \le 2]$

where $Z = \frac{X-30}{5}$

$= P[0 \le Z \le 0.8] + P[0 \le Z \le 2]$

= 0.2881+0.4772

=0.7653.

6. *X* has Normal distribution with mean 2 and standard deviation 3. Find the value of the variable x such that the probability of the interval from the mean to that value is 0.4115.

$z = \frac{x-2}{3}$

Given $\mu = 2,\ \sigma = 3, z = \frac{x-\mu}{\sigma} = \frac{x-2}{3}$

Let $Z_1 = \frac{x_1 - 2}{3}$

We have $P(\mu < x < x_1) = 0.4115$

$P(0 < Z < z_1) = 0.4115$? z_1 $= 0.415$ table valu 1.35

But $P(0 < Z < 1.35) = 0.4115$ (from the Normal Probability table)

$\therefore\ Z_1 = 1.35 \therefore \frac{x_1 - 2}{3} = 1.35$ (or) $x_1 = (1.35) \times 3 + 2 = 6.05$

$\frac{x_1 - 2}{3} = 1.35 \Rightarrow x_1$

problems for practice

1. The amount of fuel consumed by the engines of a jetliner on a flight between two cities is a normally distributed random variable X with mean μ= 5.7 tons and standard derivation σ=0.5 tons. Carrying too much fuel is inefficient as it slows the plans. If, however, too little fuel is loaded on the plane, an emergency landing may be necessary. What should be the amount of fuel to load so that there is 0.99 probability that the plane will arrive at its destination without emergency landing?

Solution: Given that $X \sim N(5.7, 0.5^2)$,

We have to find the value x such that

$$P(X < x) = 0.99$$

or $$P\left(\frac{X-\mu}{\sigma} < z\right) = 0.99$$

or $$P(Z < z) = 0.99$$

$$= 0.5 + 0.49$$

$$= 0.5 + P(0 < Z < z)$$

From the table, value of z is 2.33

So $$x = \mu + z\sigma$$

$$x = 5.7 + 2.33x0.5$$

$$x = 6.865$$

1) 2000 electric lamps

average life of lamp 1000 burning hours

with SD $\rightarrow$ 200 hrs

(a) What no. of lamps might be expected to fail in 1^{st} 700 burning hours.

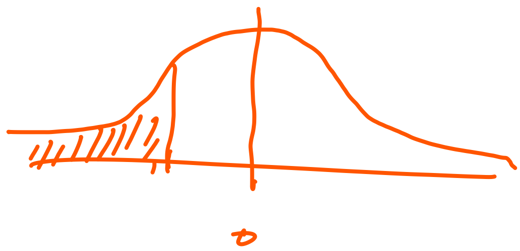
$\mu = 1000 \qquad \sigma = 200$

$P(X < 700) = ?$

If $X = 700$

$$z = \frac{700 - 1000}{200} = -1.5$$

$\rightarrow P(Z < -1.5)$



$= P(0 < z < \infty) - P(0 < z < -1.5)$

$= 0.5 - P(0 < z < 1.5)$

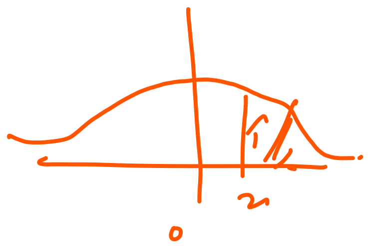
$= 0.5 - 0.4332 = \underline{0.0668}$

(b) After what period of burning hours would you expect that 10% of the lamps would have failed and would be still burning?

(assume that $\rightarrow$ normally dist!)

$X = x_1 \ni P(X > x_1) = 0.1$

$\Rightarrow P(x < x_1) = 0.9$



$\Rightarrow P(-\infty < z < z_1) = 0.9$

$\underline{P(-\infty < z < 0)} + P(0 < z < z_1) = 0.9$

$0.5 + P(0 < z < z_1) = 0.9$

$P(0 < z < z_1) = 0.9 - 0.5$

$= 0.4$

$\Rightarrow z_1 = 1.28$

$$\frac{X - \mu}{\sigma} = 1.28$$

$$\frac{X - 1000}{200} = 1.28 \Rightarrow X = 1256$$

let x_2

$P(X < x_2) = 0.1$

$\Rightarrow P(z < -z_2) = 0.1$

$0.5 - P(0 < z < z_2) = 0.1$

$P(0 < z < z_2) = 0.4 \Rightarrow z = -\underline{1.28}$

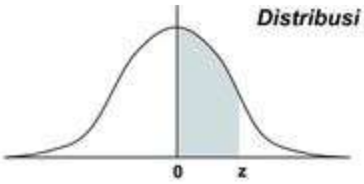
$$\frac{X - 1000}{200} = -1.28$$

$\Rightarrow X = 744$

after 744 hrs of burning, 10 % of lamps are expected to fail.

Normal Distribution Table:

Kumulatif sebaran frekuensi normal
(Area di bawah kurva normal baku dari 0 sampai z)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Dipergunakan untuk kepentingan Praktikum dan Kuliah Statistika Agrotek cit. Ade

1.00

2.1

1.3
0.05
1.35

1.2
0.08
1.28