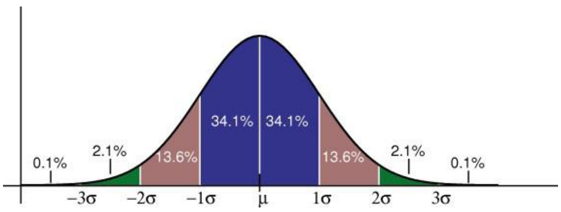
Normal Distribution

A. THE NORMAL DISTRIBUTION

A normal distribution, sometimes called the bell curve, is a distribution that occurs naturally in many situations. This distribution resembles a **symmetrical** bell curve. Half of the data will fall to the left of the mean; half will fall to the right.
Many groups follow this type of pattern:

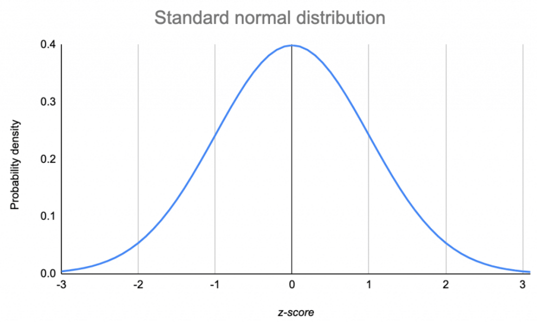
- Heights of people.
- Measurement errors.
- Blood pressure.
- Points on a test.
- IQ scores.
- Salaries



The empirical rule tells you what percentage of your data falls within a certain number of standard deviations from the mean:

- 68% of the data falls within one standard deviation of the mean.
- 95% of the data falls within two standard deviations of the mean.
- 99.7% of the data falls within three standard deviations of the mean.

Normal distribution of different variables will have different locations and spread. In order to specify the distribution, you need to give the mean, μ, and the variance, σ^2 . The notation $\boldsymbol{X \sim N(\mu, \sigma^2)}$ is used to denote a continuous variable which is normally distributed with mean, μ, and variance, σ^2.



The curves of the distribution $X \sim N(\mu, \sigma^2)$ can be described mathematically by the function

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

$for\ all\ real\ values\ of\ x.$

B. THE STANDARD NORMAL DISTRIBUTION

One table is sufficient for all normal distributions if the variable is **standardised**. The standardised value, Z, is calculated from the value X by

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

and

If $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0,1)$

To find the area between two values, $z = a$ and $z = b$, use the fact that the area between $z = a$ and $z = b$ can be written as

(area between $z = a$ and $z = b$) = (area up to $z = b$) – (area up to $z = a$)

or

$$\text{P}(a \le Z \le b) = \text{P}(Z \le b) - \text{P}(Z \le a) = \phi(b) - \phi(a)$$

C. EXERCISE

1. **[9709_w15_qp_61_007]**

The faces of a biased die are numbered 1, 2, 3, 4, 5 and 6. The probabilities of throwing odd numbers are all the same. The probabilities of throwing even numbers are all the same. The probability of throwing an odd number is twice the probability of throwing an even number.

(i) Find the probability of throwing a 3.

(ii) The die is thrown three times. Find the probability of throwing two 5s and one 4.

(iii) The die is thrown 100 times. Use an approximation to find the probability that an even number is thrown at most 37 times.

Answer :

(i) Let $\text{P}(2,4,6)$ all $= p$ then $\text{P}(1,3,5)$ all $= 2p$

$3p + 6p = 1$

$p = \frac{1}{9}$ so prob (3) $= \frac{2}{9}(0.222)$

(ii) $\text{P}(5,5,6) = \frac{2}{9} \times \frac{2}{9} \times \frac{1}{9} \times 3C2$

$= \frac{4}{243}\ (0.0165)$

(iii) $\mu = 100 \times \frac{1}{3}, \sigma = 100 \times \frac{1}{3} \times \frac{2}{3} = 22.2$

$$\text{P}(x \le 37) = \text{P}\left(z \le \frac{37.5 - \frac{100}{3}}{\sqrt{\frac{200}{9}}}\right)$$

$= \text{P}(z \le 0.8839)$

$= 0.812$

2. **[9709_w13_qp_62_005]**

On trains in the morning rush hour, each person is either a student with probability 0.36, or an office worker with probability 0.22, or a shop assistant with probability 0.29 or none of these.

(i) 8 people on a morning rush hour train are chosen at random. Find the probability that between 4 and 6 inclusive are office workers.

(ii) 300 people on a morning rush hour train are chosen at random. Find the probability that between 31 and 49 inclusive are neither students nor office workers nor shop assistants.

Answer:

(i) $\text{P}(4,5,6) = (0.22)^4(0.78)^4 8C4 + 9(0.22)^5(0.78)^3 8C5 + (0.22)^6(0.78)^2 8C6$

$= 0.0763$

(ii) $\text{prob} = 0.13$

$\text{mean} = 300 \times 0.13 = 39$

$\text{var} = 300 \times 0.13 \times 0.87 = 33.93$

$\text{P}(30 < x < 50) = \text{P}\left(\frac{30.5 - 39}{\sqrt{33.93}} < z < \frac{49.5 - 39}{\sqrt{33.93}}\right)$

$= \text{P}(-1.4592 < \text{z} < 1.8026)$

$= \Phi(1.8026) + \Phi(1.4592) - 1$

$= 0.9643 + 0.9278 - 1 = 0.892$

3. **[9709_s14_qp_63_002]**

There is a probability of $\frac{1}{7}$ that Wenjie goes out with her friends on any particular day. 252 days are chosen at random.

(i) Use a normal approximation to find the probability that the number of days on which Wenjie goes out with her friends is less than than 30 or more than 44.

(ii) Give a reason why the use of a normal approximation is justified.

Answer:

(i) $np = 252 \times \frac{1}{7} = 36$

$npq = 252 \times \frac{1}{7} \times \frac{6}{7} = 30.857$

$\text{P}\left(z < \left(\frac{29.5 - 36}{\sqrt{30.857}}\right)\right) + \text{P}\left(z > \left(\frac{44.5 - 36}{\sqrt{30.857}}\right)\right)$

$= \text{P}(z < -1.170) + \text{P}(z > 1.530)$

$= 1 - 0.8790 + 1 - 0.9370 = 0.184$

(ii) np and nq are both > 5