STA 201 Lecture Notes

Continuous Probability Distribution – Normal Distribution

Probability Distributions

Distribution of the probabilities among the different values of a random variable.

<u>Discrete probability distribution</u>- probability distribution of a discrete random variable

<u>Continuous probability distribution</u>- probability distribution of a continuous random variable

Examples:

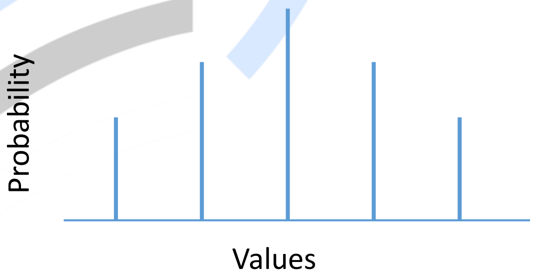
Continuous Random Variable:

1. X= Weight
2. X= Height
3. X=Temperature recorded by the meteorological office.
4. X= Time to occurring an event.
5. X= Age of any phenomenon
6. X= Length of any product etc.

Different types of probability distributions:

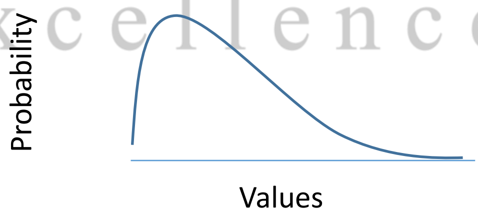
<u>Discrete probability distributions</u>-

1. Geometric Distribution
2. *Binomial Distribution*
3. *Poisson Distribution* etc.

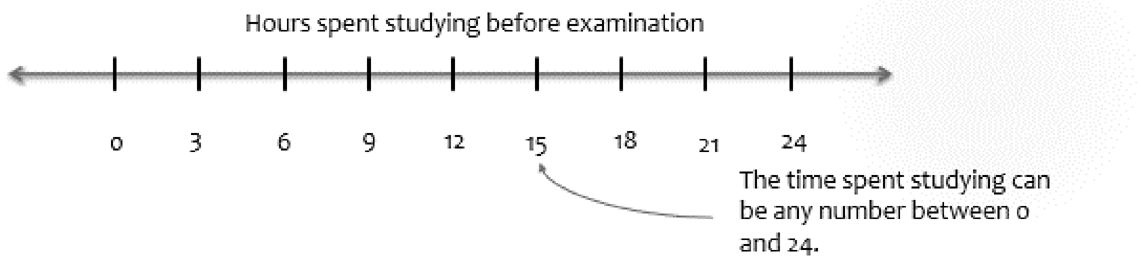


<u>Continuous probability distributions</u>-

1. Uniform Distribution
2. *Normal Distribution*
3. Exponential Distribution
4. t-distribution etc.



A Continuous random variable has an infinite number of possible values that can be represent by an interval of the number line.

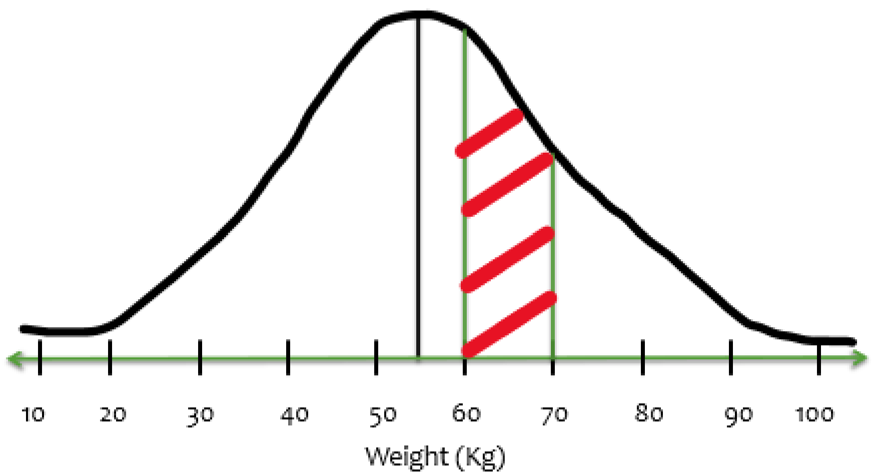


The probability distribution of a continuous random variable is called a Continuous Probability Distribution. Where, **X** = continuous random variable

What is the probability that the people weigh exactly 60 kg?

X= Weight of People

Ans: Zero, can't find the area here it only has the height no width at all



What is the probability that the people weigh between exactly 60- 70 kg?

Ans: P (60< X < 70)

The Normal Probability Distribution (Gaussian Distribution):

The normal distribution is the most important probability distribution in statistics because it fits many natural phenomena. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution. It is also known as the Gaussian distribution and the bell curve.

The normal distribution is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distribution are similarly unlikely.

As with any probability distribution, the parameters for the normal distribution define its shape and probabilities entirely. The normal distribution has two parameters, the mean and standard deviation. The normal distribution does not have just one form. Instead, the shape changes based on the parameter values.

Mean

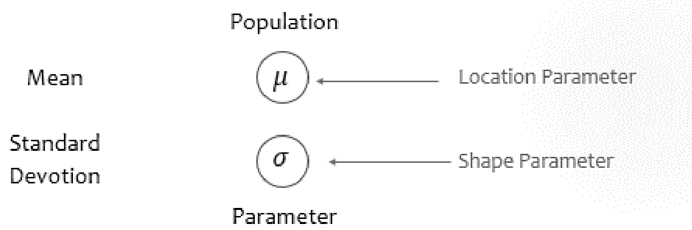
The mean is the central tendency of the distribution. It defines the location of the peak for normal distributions. Most values cluster around the mean. On a graph, changing the mean shifts the entire curve left or right on the X-axis.

Standard deviation

The standard deviation is a measure of variability. It defines the width of the normal distribution. The standard deviation determines how far away from the mean the values tend to fall. It represents the typical distance between the observations and the average.

On a graph, changing the standard deviation either tightens or spreads out the width of the distribution along the X-axis. Larger standard deviations produce distributions that are more spread out.

- The mean and standard deviation are parameter values that apply to entire populations. For the normal distribution, statisticians signify the parameters by using the Greek symbol μ (mu) for the population mean and σ (sigma) for the population standard deviation.
- Unfortunately, population parameters are usually unknown because it's generally impossible to measure an entire population. However, you can use random samples to calculate estimates of these parameters. Statisticians represent sample estimates of these parameters using $\bar{x}$ for the sample mean and s for the sample standard deviation.

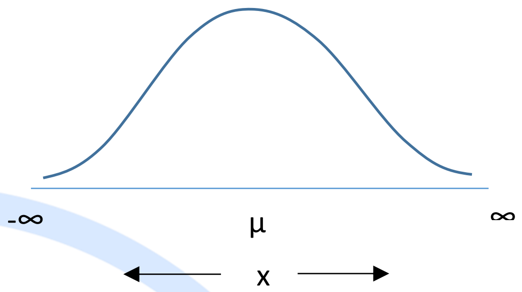


Let,

X is a continuous random variable

Then, if X has a distribution function (pdf),

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



We write it as, $X \sim N\,(\mu,\ \sigma^2)$

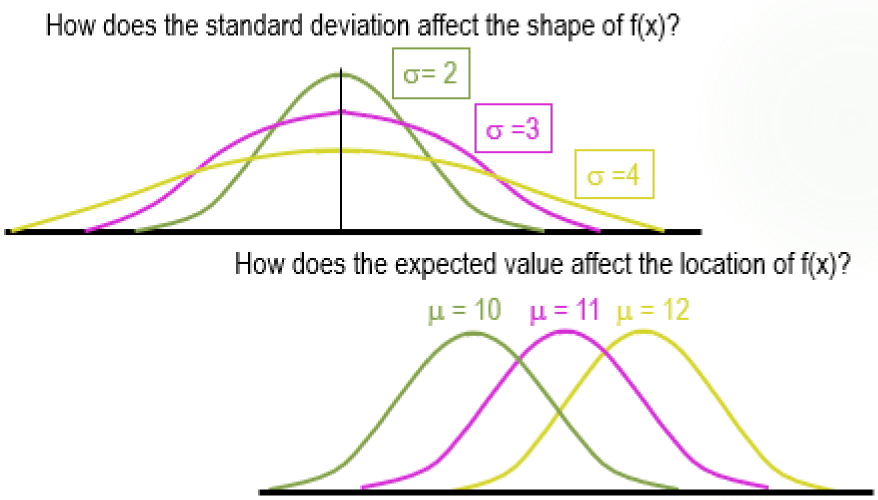
Mean, $E(X) = \mu$

Variance, $V(X) = \sigma^2$

.

<u>The Standard Normal Probability Distribution</u>

There is a family of normal distribution. Each distribution may have a different mean (μ) or standard deviation (σ). The **<u>number</u>** of normal distributions is, therefore, unlimited. [It would be physically impossible to provide table of probabilities (such as for the Binomial and Poisson) for each combination of μ and σ].



Fortunately, one member of the family of normal distribution can be used for all cases where the normal distribution is applicable. **It has a mean of 0 and a standard deviation of 1 and is called the <u>Standard Normal Distribution</u>.** Any normal distribution can be converted into the "Standard Normal Distribution" by subtracting the mean from each observation and dividing by the standard deviation.

First it is necessary to convert, or standardize, the actual distribution to a standard normal distribution using a z value, also called a z score, a z statistic, the standard normal deviate, or just the normal deviate.

Z value: The distance between a selected value, designated X, and the mean, μ, divided by the standard deviation (S. D.), σ.

In terms of the formula, Standard Normal Value:

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

Where:

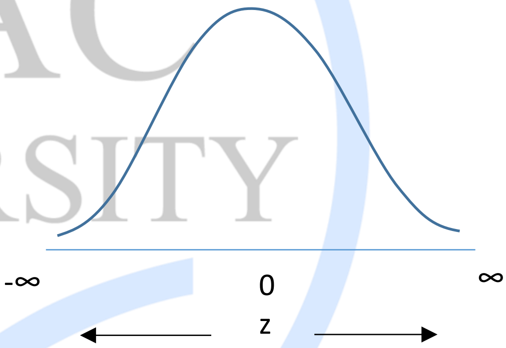
X is the value of any particular observation

μ is the mean of the distribution.

σ is the S.D. of the distribution.

Using the above formula, we can find the area or the probability under any normal curve.

Standard Normal Distribution



Let,

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

Then, Mean, $E(Z) = 0$

Variance, $V(Z) = 1$

And, if Z has a distribution function (pdf),

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \;\; ; \;\; -\infty < z < \infty$$

We write it as, $Z \sim N\,(0,\,1)$

Let, see how the mean $E\,(X) = \boldsymbol{\mu}$ of random variable X transform into mean $E\,(Z) = \mathbf{0}$ of standard normal variable Z.

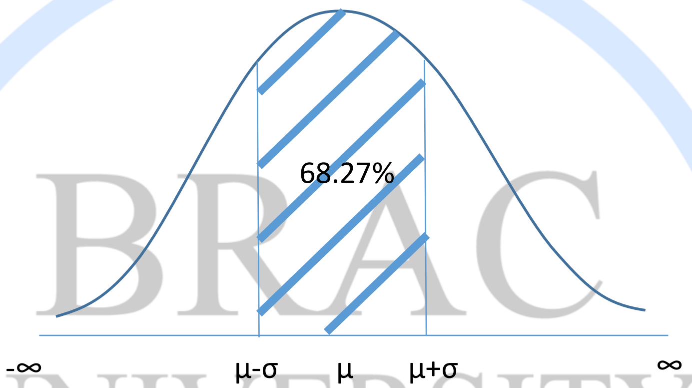
We know that, E (X + c) = E (x) + c and E (c X) = c E(x)

$E(Z) = E(\frac{X-\mu}{\sigma}) = E(\frac{X}{\sigma} - \frac{\mu}{\sigma}) = E(\frac{X}{\sigma}) - \frac{\mu}{\sigma} = \frac{1}{\sigma}E\,(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$ [As $E(X) = \mu$]

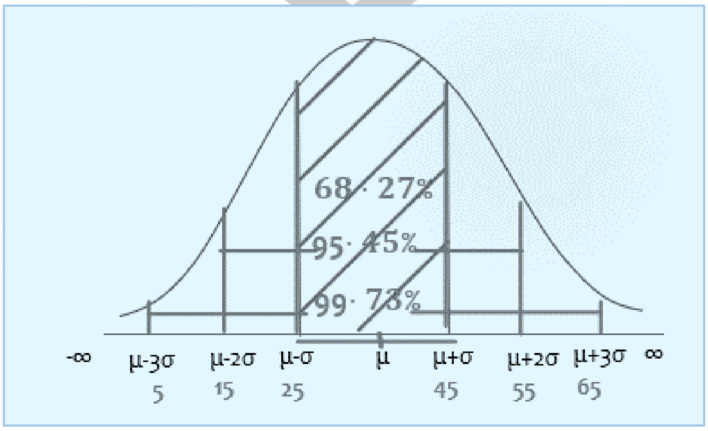
Similarly, we can show V (X) = 1

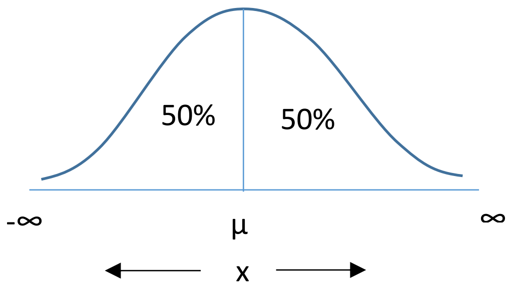
Characteristics of a Normal Distribution

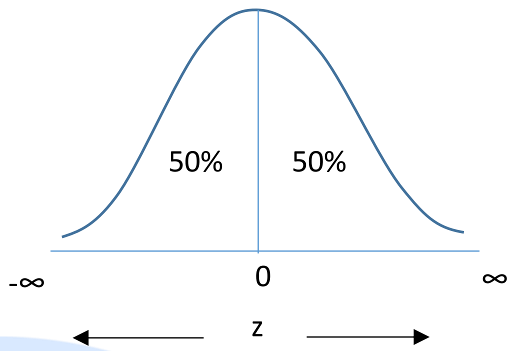
1. Mean= Median = Mode
2. Symmetric and Mesokurtic
3. Bell-shaped curve
4. The area under the curve lying between μ±σ is 68.27% of the total area
5. The area under the curve lying between μ±2σ is 95.45% of the total area
6. The area under the curve lying between μ±3σ is 99.73% of the total area



- Let’s say in any area the age of a population follows Normal distribution with Age ~ N (35, 10)
- 35 ±10 = 35-10, 35+10 = (25, 45)
- 35 ±2*10 = 35-20, 35+2 = (15, 55)
- 35 ±3*10 = 35-30, 35+30 = = (5, 65)







P[X<μ] = P[X> μ] = 0.5

P[X<-x] = P[X>x]

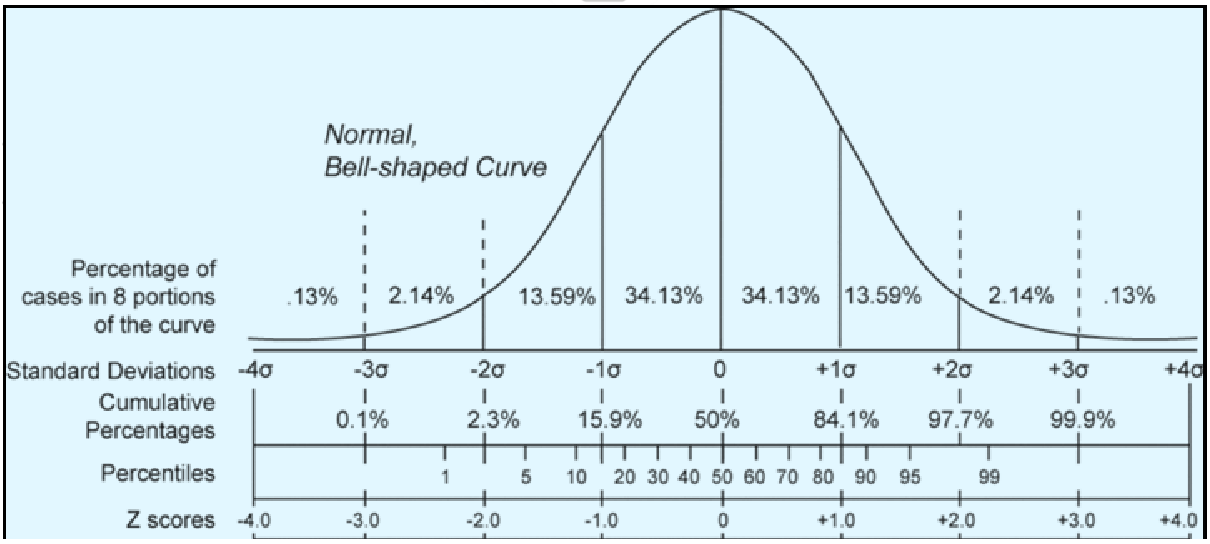
P[Z<0] = P[Z> 0] = 0.5

P[Z<-z] = P[Z>z]

Standard Normal Distribution Table (Z-table):

- Normal distribution table provides probabilities for N(0,1) i.e. for standard normal distribution
- Usually, normal table gives P[0 < Z < z] for positive values of Z.
- For other values, we can use the property of symmetry with median 0 of standard normal distribution
- To find probabilities for a normal random variable X, we can transform the probability statement about X in terms of probability statement for Z and then calculate the probability using the standard normal distribution table or Z-table

$$P[X < a] = P\left[\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right] = P\left[Z < \frac{a-\mu}{\sigma}\right]$$



Example 1:

The number of viewers of a TV show per week has a mean of 29 million with a standard deviation of 5 million. Assume that, the number of viewers of that show follows a normal distribution.

What is the probability that, next week's show will-

a. Have between 30 and 34 million viewers?

b. Have at least 23 million viewers?

c. Exceed 40 million viewers?

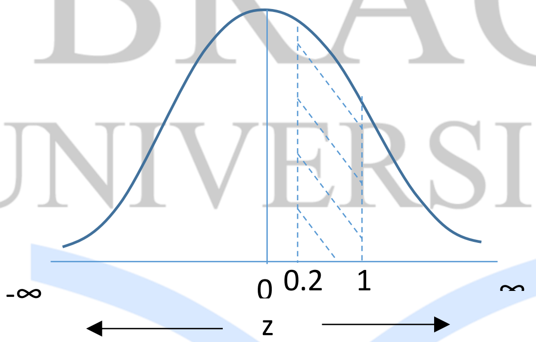
<u>Solution</u>**:**

Let, X= Number of viewers of the show per week (in million)

$$\therefore X \sim N\,(\mu,\ \sigma^2)$$

a. the probability that, next week's show will have between 30 and 34 million viewers-

$$P[30 \le X \le 34] = P\left[\frac{30-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{34-\mu}{\sigma}\right] = P\left[\frac{30-29}{5} \le \frac{X-\mu}{\sigma} \le \frac{34-29}{5}\right]$$
$$= P[0.20 \le Z \le 1] = P[0 \le Z \le 1] - P[0 \le Z \le 0.2] = 0.3413 - 0.0793$$
$$= 0.262$$

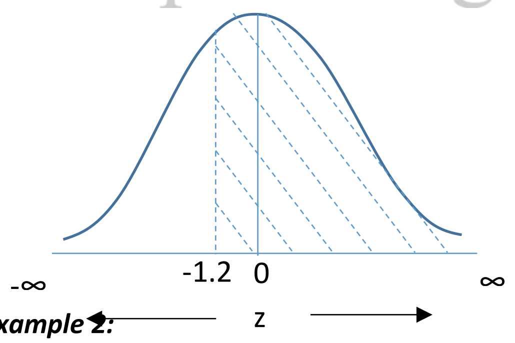


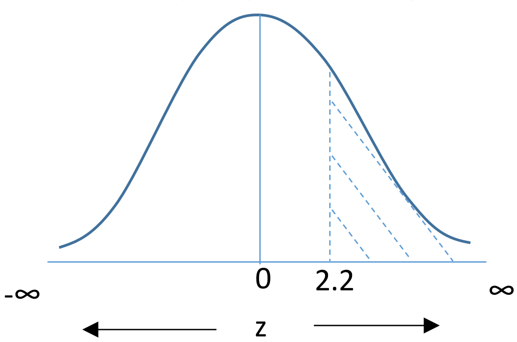
b. the probability that, next week's show will have at least 23 million viewers-

$$P[X \ge 23] = P\left[\frac{X-\mu}{\sigma} \ge \frac{23-\mu}{\sigma}\right] = P\left[\frac{X-\mu}{\sigma} \ge \frac{23-29}{5}\right] = P[Z \ge -1.2]$$
$$= P[-1.2 \le Z \le 0] + P[Z \ge 0] = 0.3849 + 0.5 \ = 0.8849$$

c. the probability that, next week's show will exceed 40 million viewers-

$$P[X > 40] = P\left[\frac{X-\mu}{\sigma} > \frac{40-\mu}{\sigma}\right] = P\left[\frac{X-\mu}{\sigma} > \frac{40-29}{5}\right] = P[Z > 2.2]$$
$$= P[Z \ge 0] - P[0 \le Z \le 2.2] = 0.5 - 0.4861 = 0.0139$$





Example 2:

a. For what value of 'a', $P[Z \le a] = 0.95$?

b. For what value of 'a', $P[Z \ge a] = 0.05$?

c. For what value of 'a', $P[Z \le a] = 0.975$?

<u>Solution:</u>

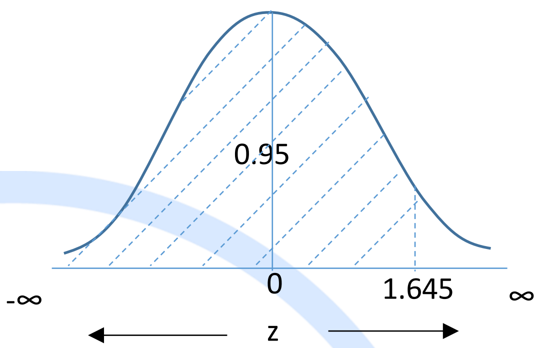
a. $P[Z \le a] = 0.95$

Or, $P[Z \le 0] + P[0<Z \le a] = 0.95$

Or, $0.5 + P[0<Z \le a] = 0.95$

Or, $P[0<Z \le a] = 0.95 - 0.5 = 0.45$

For $a = 1.645$, $P[0<Z \le a] = 0.45$



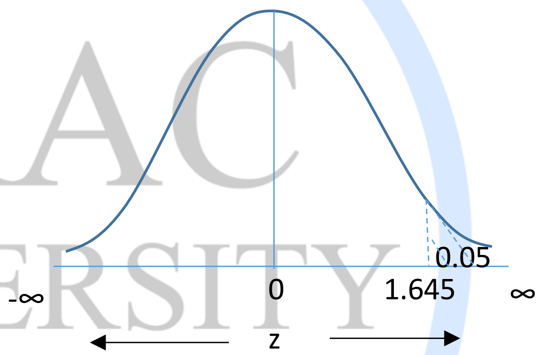
b. $P[Z \ge a] = 0.05$

Or, $P[Z \ge 0] - P[0<Z \le a] = 0.05$

Or, $0.5 - P[0<Z \le a] = 0.05$

Or, $P[0<Z \le a] = 0.5 - 0.05 = 0.45$

For $a = 1.645$, $P[0<Z \le a] = 0.45$



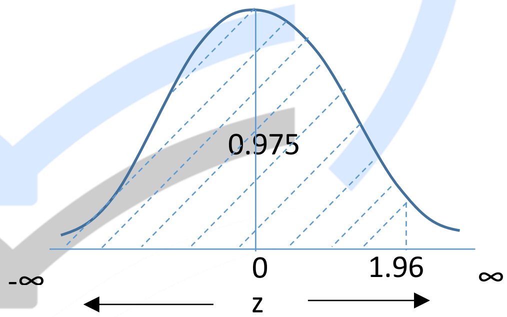
c. $P[Z \le a] = 0.975$

Or, $P[Z \le 0] + P[0<Z \le a] = 0.975$

Or, $0.5 + P[0<Z \le a] = 0.975$

Or, $P[0<Z \le a] = 0.975 - 0.5 = 0.475$

For $a = 1.96$, $P[0<Z \le a] = 0.475$



***Example 3**:* Suppose that the height of UCLA female students has normal distribution with mean 62 inches and standard deviation 8 inches.

a. Find the height below which is the shortest 30% of the female students.
b. Find the height above which is the tallest 5% of the female students.

Solution:

We are given $X \sim N(62, 8)$.

a. We want to find the height h such that $P(X < h) = 0.30$.

From the standard normal table this corresponds to $z = -0.52$

Therefore $-0.52 = (h-62)/8 \Rightarrow h = 57.84$ inches.

b. We want to find the height h such that P (X > h) = 0.05.

From the standard normal table this corresponds to z = 1.64.

Therefore 1.64 = (h−62)/ 8 ⇒ h = 75.12 inches.

More Examples:

Example: Given a normal distribution with $\mu = 50, \sigma = 10,$ find the probability that X assumes a value between 45 and 62

Solution: The Z values corresponding to x_1=45 and x_2=62

$$\therefore Z_1 = \frac{45-50}{10} = -0.5 \qquad Z_2 = \frac{62-50}{10} = 1.2$$

P(45<x<62) = P(-0.5<Z<1.2)

= P(Z<1.2) – P(Z<-0.5)

= P(Z<1.2) –{1- P(Z<0.5)}

= ϕ(1.2) - {1- ϕ(0.5)}

= 0.8849 – {1 – 0.6915} [USING TABLE] = 0.5764

__**Example:**__ The weekly incomes of the bankers of a bank follow normal distribution with a mean of \$ 1,000 and std. of \$100.

What is the livelihood of selecting a banker whose weekly income is between \$1000 and \$100?

$$P(1{,}000<X<1{,}100) = P\left(\frac{1000-1000}{100} < Z < \frac{1100-1000}{100}\right)$$

= P(0<Z<1)

= 0.3413

P(X<1,100) = P(Z<1) = 0.5 + 0.3413 = 0.8413

P(790<X<1,000) = (-2.10<Z<0) = 0.4821

P(X<790) = P(Z<-2.10) = 0.5 – 0.4821 = 0.0179

P(X>790) = P(Z>-2.10) = 0.4821 + 0.5 = 0.9821

P(X>482) = P[Z> $\frac{482-400}{50}$] = P(Z>1.64) = 0.5 – 0.4495 = 0.0505

Practice Problem:

1. The weekly incomes of some employees in an industry are normally distributed with a mean of \$1,000 and standard deviation of \$100. What is the likelihood of selecting an employee whose weekly income is -

a. between \$1,000 and \$1,100?

b. between \$790 and \$1,000?

c. Less than \$790?

d. Between \$840 and \$1,200?

e. Between \$1,150 and \$1,250?

[**Answer:** a. 0.3413; b.0.4821; c. 0.179; d. 0.9224; e. 0.0606]

2. A large group of students took a test in Physics and the final grades have a mean of 70 and a standard deviation of 10. If we can approximate the distribution of these grades by a normal distribution, what percent of the students

a. scored higher than 80?

b. should pass the test (grades≥60)?

c. should fail the test (grades<60)?

[**Answer:** a. 0.1586; b. 0.8413; c. 0.1586]

3. The annual salaries of employees in a large company are approximately normally distributed with a mean of \$50,000 and a standard deviation of \$20,000.

a. What percent of people earn less than \$40,000?

b. What percent of people earn between \$45,000 and \$65,000?

c. What percent of people earn more than \$70,000?

[**Answer:** a. 0.3085; b. 0.3720; c. 0.1586]

4. A tire manufacturer wishes to set a minimum mileage guarantee on its new MX100 tire. Tests reveal the mean mileage is 67,900 with a standard deviation of 2,050 miles and a normal distribution. The manufacturer wants to set the minimum guaranteed mileage so that no more than 4 percent of the tires will have to be replaced. What minimum guaranteed mileage should the manufacturer announce?

[Answer: 64,312**]**