Probability And Distribution

Let's have a look at the things we are going to discuss in this article:

- Probability and the basic concepts.
- Additive theorem of probability.
- Multiplicative theorem of probability
- Bayes theorem
- Random Variables-Discrete/Continuous

Probability

Probability implies 'likelihood' or 'chance'. When an event is certain to happen then the probability of occurrence of that event is 1 and when it is certain that the event cannot happen then the probability of that event is 0.Hence the value of probability ranges from 0 to 1.

Classical Definition of Probability

As the name suggests the classical approach to defining probability is the oldest approach. It states that if there are n exhaustive, mutually exclusive and equally likely cases out of which m cases are favourable to the happening of event A,Then the probabilities of event A is defined as given by the following probability function:

$$P(A) = \frac{\text{Number of favorable outcomes to A}}{\text{Total number of outcomes}}$$

Example:

Problem Statement:

A coin is tossed. What is the probability of getting a head?

Solution:

Number of outcomes favourable to head (m) = 1

Total number of outcomes (n) = 2 (i.e. head or tail)

Probability - Basic Concepts

Random Experiment

An experiment is said to be a random experiment, if it's out-come can't be predicted with certainty.
Example
If a coin is tossed, we can't say whether head or tail will appear. So it is a random experiment.

Sample Space

The set of all possible out-comes of an experiment is called the sample space. It is denoted by 'S' and its number of elements are n(s).
Example
In throwing a dice, the number that appears at top is any one of 1,2,3,4,5,6.
So here:
S ={1,2,3,4,5,6} and n(s) = 6
Similarly in the case of a coin, S={Head,Tail} or {H,T} and n(s)=2.

Event

Every subset of a sample space is an event. It is denoted by 'E'.
Example
In throwing a dice S={1,2,3,4,5,6}, the appearance of an even number will be the event E={2,4,6}.
Clearly E is a subset of S.

Equally likely events

Events are said to be equally likely, if the probability of occurrence of the events are the same.
Example
When a dice is thrown, all the six faces {1,2,3,4,5,6} are equally likely to come up.

Exhaustive events

When every possible outcome of an experiment is considered.
Example
A dice is thrown, cases 1,2,3,4,5,6 form an exhaustive set of events.

Mutually exclusive or Disjoint event

If two or more events can't occur simultaneously,
that is no two of them can occur together.
Example
When a coin is tossed, the event of occurrence of a head and the event of occurrence of a tail are mutually exclusive events.

Independent or Mutually independent events

Two or more events are said to be independent if occurrence or non-occurrence of any of them does not affect the probability of occurrence or non-occurrence of the other event.
Example
When a coin is tossed twice, the event of occurrence of head in the first throw and the event of occurrence of head in the second throw are independent events.

Difference between mutually exclusive and mutually independent events

Mutually exclusiveness is used when the events are taken from the same experiment, whereas independence is used when the events are taken from different experiments.

Additive Theorem Of Probability

For Non Mutually Exclusive Events

Statement: If A and B are not mutually exclusive events, the probability of the occurrence of either A or B or both is equal to the probability that event A occurs, plus the probability that event B occurs minus the probability of occurrence of the events common to both A and B. In other words the probability of occurrence of at least one of them is given by

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

For Mutually Exclusive Events

Statement: If A and B are two mutually exclusive events, then the probability of occurrence of either A or B is the sum of the individual probabilities of A and B. Symbolically

$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A \cup B) = P(A) + P(B)$$

Additive Theorem Of Probability - Examples

For Non Mutually Exclusive Events

1. A shooter is known to hit a target 3 out of 7 shots; whereas another shooter is known to hit the target 2 out of 5 shots. Find the probability of the target being hit at all when both of them try.
2. In a math class of 30 students, 17 are boys and 13 are girls. On a unit test, 4 boys and 5 girls made an A grade. If a student is chosen at random from the class, what is the probability of choosing a girl or an A student?

For Mutually Exclusive Events

1. A card is drawn from a pack of 52, what is the probability that it is a king or a queen?
2. A single 6-sided die is rolled. What is the probability of rolling a 2 or a 5?

Solutions - 1

Probability of first shooter hitting the target P (A) = 3/7

Probability of second shooter hitting the target P (B) = 2/5

Event A and B are not mutually exclusive as both the shooters may hit the target. Hence the additive rule applicable is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{3}{7} + \frac{2}{5} - \left(\frac{3}{7} \times \frac{2}{5}\right)$$

$$= \frac{29}{35} - \frac{6}{35}$$

$$= \frac{23}{35}$$

Solutions - 2

Let Event (A) = Draw of a card of king

Event (B) Draw of a card of queen

P (card draw is king or queen) = P (card is king) + P (card is queen)

$$P(A \cup B) = P(A) + P(B)$$
$$= \frac{4}{52} + \frac{4}{52}$$
$$= \frac{1}{13} + \frac{1}{13}$$
$$= \frac{2}{13}$$

Multiplicative Theorem Of Probability

For Independent Events

Statement: The theorem states that the probability of the simultaneous occurrence of two events that are independent is given by the product of their individual probabilities.

$$P(A \text{ and } B) = P(A) \times P(B)$$
$$P(AB) = P(A) \times P(B)$$

For Dependent Events (Conditional Probability)

If we recall dependent event(), the earlier stated multiplicative theorem is not applicable for dependent events. For dependent event, we have an another theorem called the conditional probability which is given as:
The probability of event B given event A equals the probability of event A and event B divided by the probability of event A

$$P(B/A) = \frac{P(AB)}{P(A)} \text{ or } \frac{P(A \cap B)}{P(A)}$$

Learnvista Pvt Ltd.
2nd Floor, 147, 5th Main Rd, Rajiv Gandhi Nagar HSR Sector 7,Near Salarpuria Serenity, Bengaluru, Karnataka 560102
Mob:- +91 779568798, Email:- contacts@learnbay.co

Multiplicative Theorem/Conditional Probability - Examples

Independent Event: You have a cowboy hat, a top hat, and an Indonesian hat called a songkok. You also have four shirts: white, black, green, and pink. If you choose one hat and one shirt at random, what is the probability that you choose the songkok and the black shirt?

The two events are independent events; the choice of hat has no effect on the choice of shirt.
There are three different hats, so the probability of choosing the songkok is 1/3 .
There are four different shirts, so the probability of choosing the black shirt is 1/4.

So, by the Multiplication Rule:
P(songkok and black shirt)=(1/3) ⋅ (1/4)=1/12

Dependent Event: An urn contains 20 red and 10 blue balls. Two balls are drawn from a bag one after the other without replacement. What is the probability that both the balls drawn are red?

Solution: Let A and B denote the events that first and second ball drawn are red balls. We have to find P(A∩B) or P(AB).

P(A) = P(red balls in first draw) = 20/30

Now, only 19 red balls and 10 blue balls are left in the bag. Probability of drawing a red ball in second draw too is an example of conditional probability where drawing of second ball depends on the drawing of first ball.

Hence Conditional probability of B on A will be,

P(B|A) = 19/29

By multiplication rule of probability,

P(A∩B) = P(A) × P(B|A)

$$P(A \cap B) = \frac{20}{30} \times \frac{19}{29} = \frac{38}{87}$$

Bayes Theorem

- Named after Thomas Bayes
- Bayes' Theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.
- Note: This conditional probability is known as a hypothesis. This hypothesis is calculated through previous evidence or knowledge. This conditional probability is the probability of the occurrence of an event, given that some other event has already happened.

Example:
Suppose the weather of the day is cloudy. Now, you need to know whether it will rain today, given the cloudiness of the day. Therefore, you are supposed to calculate the probability of rainfall, given the evidence of cloudiness.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Theorem – Where Does It Come From?

We know from Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Rearranging Equation 1:

$$P(A \cap B) = P(A|B) \times P(B)$$

Similarly,

$$P(B \cap A) = P(B|A) \times P(A)$$

Since,

$$P(A \cap B) = P(B \cap A)$$

Hence,

$$P(B|A).P(A) = P(A|B).P(B)$$

Finally,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Theorem – Generalized Form?

$$p(H \mid E) = \frac{p(E \mid H)\ p(H)}{p(E)} \quad \Pr(H|E) = \frac{\Pr(E|H)\Pr(H)}{\Pr(E|H)\Pr(H) + \Pr(E|\text{not } H)\Pr(\text{not } H)}$$

Bayes Theorem - Examples

Epidemiologists claim that the probability of breast cancer among Caucasian women in their mid -50s is 0.005. An established test identified people who had breast cancer and those that were healthy. A new mammography test in clinical trials has a probability of 0.85 for detecting cancer correctly. In women without breast cancer, it has a chance of 0.925 for a negative result. If a 55-year-old Caucasian woman tests positive for breast cancer, what is the probability that she, in fact, has breast cancer?

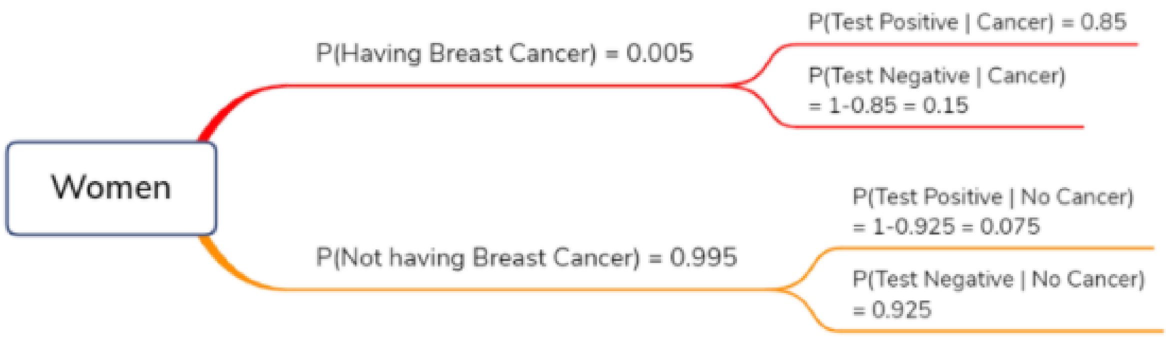
Solution:

P(Cancer) = 0.005

P(Test Positive | Cancer) = 0.85

P(Test -ve|No cancer) = 0.925

P(Cancer|Test +ve)= P(Cancer) * P(Test Positive | Cancer) / P(Test Positive)



	Probability of having Cancer or not	Test being Positive	Test being Negative
Cancer	0.005	0.005*0.85 = 0.00425	0.005*0.15 = 0.00075
No Cancer	0.995	0.995*0.075 = 0.074625	0.995*0.925 = 0.920375
Total	1.00	0.078875	0.921125

RANDOM VARIABLES – DISCRETE/CONTINUOUS

Random Variable

- A variable that can take multiple values with different probabilities.
- The mathematical function describing these possible values along with their associated probabilities is called a probability distribution.

Points scored per game	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Frequency, f	1	4	6	12	5	1	1

Total number of games – N= 30

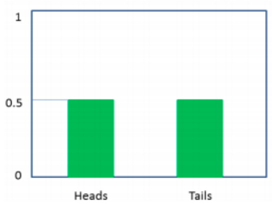
Points scored per game	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Probability *Recall the Frequentist (empirical) approach of assigning probabilities*	$\frac{1}{30}$	$\frac{4}{30}$	$\frac{6}{30}$	$\frac{12}{30}$	$\frac{5}{30}$	$\frac{1}{30}$	$\frac{1}{30}$

Leads to Descriptive Stats

Leads to Inferential Stats

RANDOM VARIABLE TYPES

Discrete and Continuous



Countable

Measurable

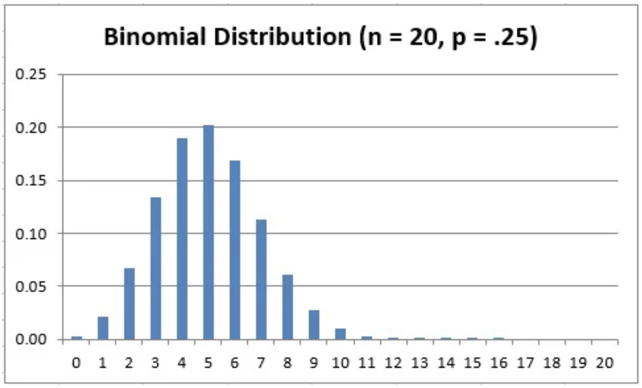
PROBABILITY MASS FUNCTION/PROBABILITY DENSITY FUNCTION

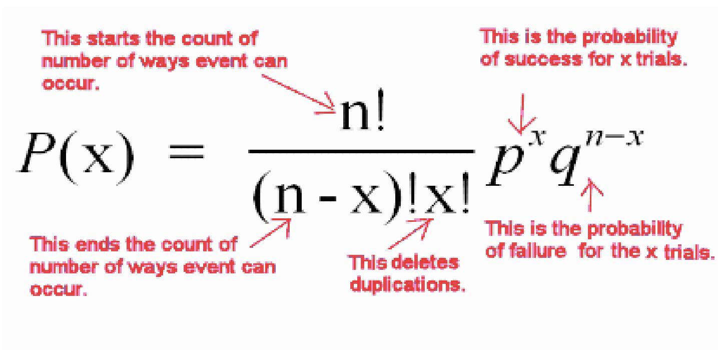
Discrete Distributions	Continuous Distributions
Probability that X can take a specific value x is $P(X = x) = p(x)$.	Probability that X is between two points a and b is $P(a \leq X \leq b) = \int_a^b f(x)dx$.
It is non-negative for all real x.	It is non-negative for all real x.
The sum of $p(x)$ over all possible values of x is 1, i.e., $\sum p(x) = 1$.	$\int_{-\infty}^{\infty} f(x)dx = 1$
Probability Mass Function	**Probability Density Function**

DISCRETE DISTRIBUTIONS - BINOMIAL DISTRIBUTION

- A distribution where only two outcomes are possible, such as success or failure, gain or loss, win or lose and where the probability of success and failure is same for all the trials is called a Binomial Distribution.
- The outcomes need not be equally likely.
- Each trial is independent.
- A total number of n identical trials are conducted.
- The probability of success and failure is the same for all trials. (Trials are identical.)

- Mathematical Representation





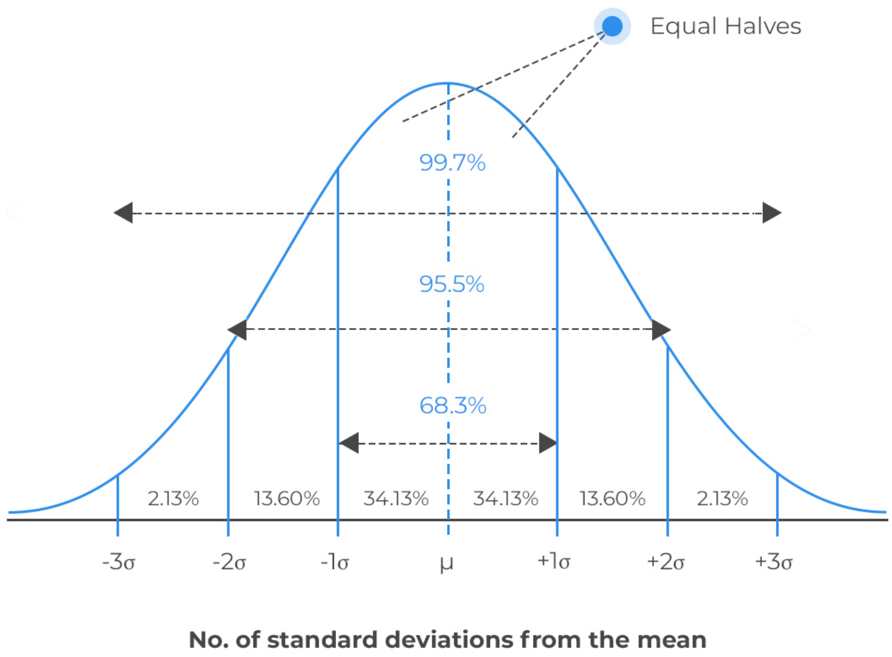
CONTINUOUS DISTRIBUTION – NORMAL DISTRIBUTION

- Normal distribution represents the behaviour of most of the situations in the universe (That is why it’s called a “normal” distribution.)
- The large sum of (small) random variables often turns out to be normally distributed, contributing to its widespread application.

Characteristics:

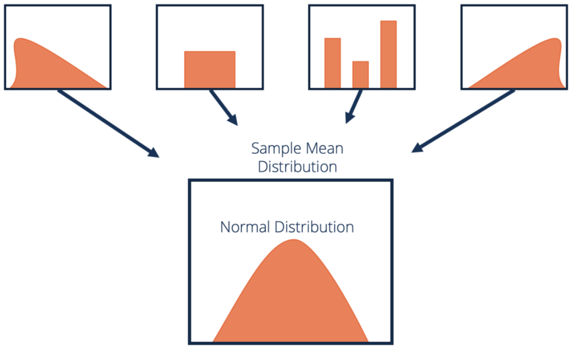
- The mean, median and mode of the distribution coincide.
- The curve of the distribution is bell-shaped and symmetrical about the line x=μ.
- The total area under the curve is 1.
- Exactly half of the values are to the left of the center and the other half to the right.
- A normal distribution is highly different from Binomial Distribution. However, if the number of trials approaches infinity then the shapes will be quite similar.

Shape of the normal distribution



NORMAL DISTRIBUTION – CENTRAL LIMIT THEOREM

- The central limit theorem in statistics states that, given a sufficiently large sample size, the sampling distribution of the mean for a variable will approximate a normal distribution regardless of that variable's distribution in the population.
- (In Layman's term – even if the data is not normally distributed, the mean of the distribution is normal distribution provided the sample size is large).
- Why is it useful: We can use mean's Normal Distributions to make confidence Intervals, perform hypothesis testing.



NORMAL DISTRIBUTION – CENTRAL LIMIT THEOREM

Central Limit Theorem (Key Takeaways)

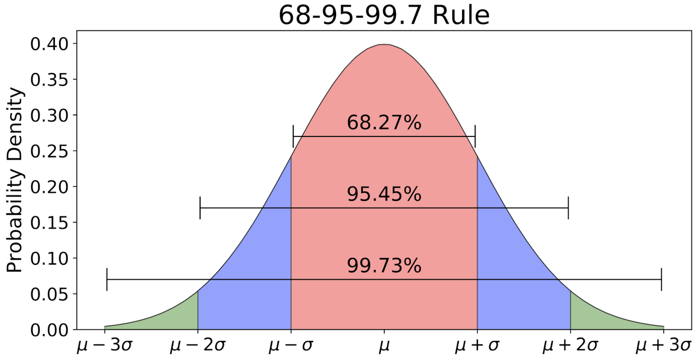
- The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger.
- Sample sizes equal to or greater than 30 are considered sufficient for the CLT to hold.
- A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
- A sufficiently large sample size can predict the characteristics of a population accurately.
- Expectation of Sample Mean as a random variable = Population Mean. Symbolically $E(\bar{X}) = \mu$.
- Standard Deviation $(\bar{X}) = \sigma / \sqrt{n}$ (where σ is standard deviation and n is sample size).

NORMAL DISTRIBUTION – EMPIRICAL RULE

The empirical rule states that for a normal distribution, nearly all of the data will fall within three standard deviations of the mean. The empirical rule can be broken down into three parts:

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

The rule is also called the 68-95-99.7 Rule or the Three Sigma Rule.



PDF OF A NORMAL DISTRIBUTION AND ORIGIN OF EMPIRICAL FORMULA

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

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

Note constants:
π=3.14159
e=2.71828

Empirical Formula of Normal Distribution: The empirical rule, also referred to as the three-sigma rule or 68-95-99.7 rule, is a statistical rule which states that for a normal distribution, almost all data falls within three standard deviations (denoted by σ) of the mean (denoted by μ). Broken down, the empirical rule shows that almost 68% falls within the first standard deviation ($\mu \pm \sigma$), almost 95% within the first two standard deviations ($\mu \pm 2\sigma$), and almost 99.7% within the first three standard deviations ($\mu \pm 3\sigma$).

NORMAL DISTRIBUTION/ORIGIN OF Z-SCORE

Z Score gives how many standard deviations away from mean a value is. However, to understand the probability associated with it, we need to refer to the Z-Table.
Let's solve some questions:
Type 1: Comparison of 2 different Normally Distributed values (Z-Score is enough)
Type 2: Finding the probability or percentage of values. (Need Z-table)

Type 1: Happy and Ekta are two students. Happy Scored 65 marks in Math Exam while Ekta scored 80 in English Exam. Given that both Math and English marks follow an approx. Normal Distribution, who performed better?
Math ~ N(60,4)
English ~ N(79,2)

Solution:
We have been given that
Math ~ ND(60,4)

English ~ ND(79,2)

From this we know that

$\mu = 60, \ \sigma = 4$ for Math and

$\mu = 79, \ \sigma = 2$ for English

From the z-score formula,

$$z - score = \frac{(x - \mu)}{\sigma}$$

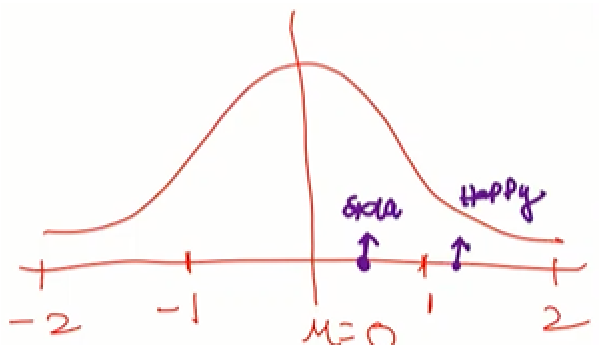
Let's find the individual z-scores,

$$(z - score)_{Happy} = \frac{x - \mu}{\sigma} = \frac{65 - 60}{4} = \frac{5}{4}$$

$$(z - score)_{Ekta} = \frac{x - \mu}{\sigma} = \frac{80 - 79}{2} = \frac{1}{2}$$

We already know that =1.25 and 1/2 is 0.5.

Let's plot the same on a graph,



Clearly, from the graph, Happy performed better than Ekta since the z-score is higher and the point is more outward than the other one.

Type 2: According to the Center for Disease Control, heights for U.S. adult females and males are approximately normal.

Females: mean of 64 inches and SD of 2 inches

Males: mean of 69 inches and SD of 3 inches

Find the probability of a randomly selected U.S. adult female being taller than 65 inches.

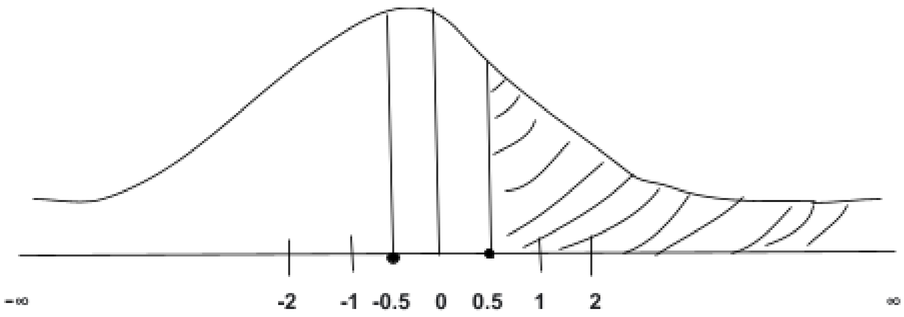
Solution:

Given, the females have a mean of 64 inches and SD of 2 inches

Let's find the z-score first,

$$z-score_{Females} = \frac{x-\mu}{\sigma} = \frac{65-64}{2} = \frac{1}{2} = 0.5$$

Now, to find the probability we must find the area under the curve to the right side of point 0.5. To find the area,



$$\int_{-\infty}^{0.5} - \int_{-\infty}^{-0.5}$$

From the z-table,

$$\int_{-\infty}^{0.5} = 0.69146 \ \ and \int_{-\infty}^{-0.5} = 0.30854$$

So,

$$\int_{-\infty}^{0.5} - \int_{-\infty}^{-0.5}$$

=0.69146-0.30854

=0.38292

=38%

But we need the positive value of the curve so 38/2 = 19%

Hence the probability of a randomly selected U.S. adult female being taller than 65 inches is 19%.