PROBABILITY DISTRIBUTION

Probability theory developed from the study of games of chance like dice and cards. A process like flipping a coin, rolling a die or drawing a card from a deck are called probability experiments. An outcome is a specific result of a single trial of a probability experiment.

Probability distributions

Probability theory is the foundation for statistical inference. A probability distribution is a device for indicating the values that a random variable may have. There are two categories of random variables. These are discrete random variables and continuous random variables.

Discrete random variable

The probability distribution of a discrete random variable specifies all possible values of a discrete random variable along with their respective probabilities.

Examples can be
- Frequency distribution
- Probability distribution (relative frequency distribution)
- Cumulative frequency

Examples of discrete probability distributions are the binomial distribution and the Poisson distribution.

Binomial distribution

A binomial experiment is a probability experiment with the following properties.

1. Each trial can have only two outcomes which can be considered success or failure.
2. There must be a fixed number of trials.
3. The outcomes of each trial must be independent of each other.
4. The probability of success must remain the same in each trial.

The outcomes of a binomial experiment are called a binomial distribution.

Poisson distribution

The Poisson distribution is based on the Poisson process.

1. The occurrences of the events are independent in an interval.
2. An infinite number of occurrences of the event are possible in the interval.
3. The probability of a single event in the interval is proportional to the length of the interval.
4. In an infinitely small portion of the interval, the probability of more than one occurrence of the event is negligible.

Continuous probability distributions

A continuous variable can assume any value within a specified interval of values assumed by the variable. In a general case, with a large number of class intervals, the frequency polygon begins to resemble a smooth curve.

A continuous probability distribution is a *probability density function*. The area under the smooth curve is equal to 1 and the frequency of occurrence of values between any two points equals the total area under the curve between the two points and the x-axis.

The normal distribution

The normal distribution is the most important distribution in biostatistics. It is frequently called the Gaussian distribution. The two parameters of the normal distribution are the mean (μ) and the standard deviation (σ). The graph has a familiar bell-shaped curve.

Graph of a normal distribution
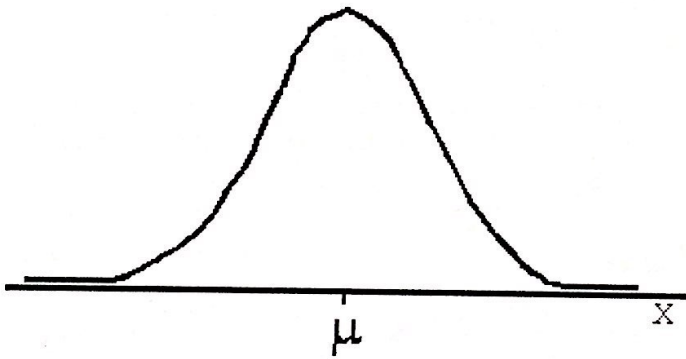
Characteristics of the normal distribution

1. It is symmetrical about μ.

2. The mean, median and mode are all equal.

3. The total area under the curve above the x-axis is 1 square unit. Therefore 50% is to the right of μ and 50% is to the left of μ.

4. Perpendiculars of:
   $\pm \sigma$ contain about 68%;
   $\pm 2\sigma$ contain about 95%;
   $\pm 3\sigma$ contain about 99.7%
of the area under the curve.

The standard normal distribution

A normal distribution is determined by μ and σ. This creates a family of distributions depending on whatever the values of μ and σ are. The standard normal distribution has $\mu = 0$ and $\sigma = 1$.

Finding normal curve areas

1. The table gives areas between $-\infty$ and the value of z_0.

2. Find the z value in tenths in the column at left margin and locate its row. Find the hundredths place in the appropriate column.

3. Read the value of the area (P) from the body of the table where the row and column intersect. Note that P is the probability that a given value of z is as large as it is in its location. Values of P are in the form of a decimal point and four places. This constitutes a decimal percent.
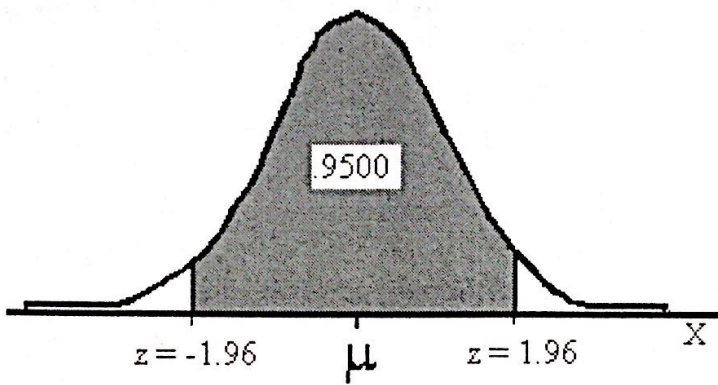
Finding probabilities

We find probabilities using the table and a four-step procedure as illustrated below.
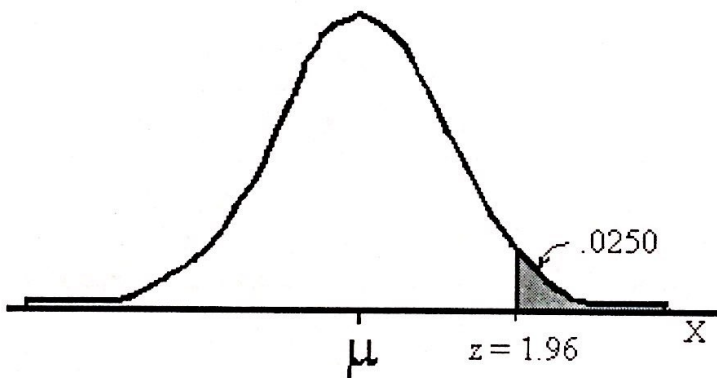
a) What is the probability that z < -1.96?
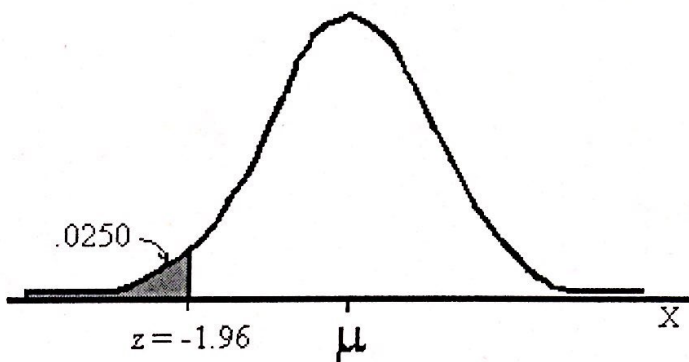
.0250

$z = -1.96$ μ X

(1) Sketch a normal curve
(2) Draw a line for z = -1.96
(3) Find the area in the table
(4) The answer is the area to the left of the line $P(z < -1.96) = .0250$

b) What is the probability that -1.96 < z < 1.96?

(1) Sketch a normal curve
(2) Draw lines for lower z = -1.96, and upper z = 1.96
(3) Find the area in the table corresponding to each value
(4) The answer is the area between the values--subtract lower from upper $P(-1.96 < z < 1.96) =$ .9750 - .0250 = .9500

c) What is the probability that z > 1.96?

(1) Sketch a normal curve
(2) Draw a line for z = 1.96
(3) Find the area in the table
(4) The answer is the area to the right of the line; found by subtracting table value from 1.0000; $P(z > 1.96) = 1.0000 - .9750 = .0250$

Applications of the normal distribution

The normal distribution is used as a model to study many different variables. We can use the normal distribution to answer probability questions about random variables. Some examples of variables that are normally distributed are human height and intelligence.