DEFINITIONS

POPULATION : Entire aggregate of individuals or items from which SAMPLES are drawn.

SAMPLE : A set of individuals or items selected from a PARENT POPULATION so that properties or parametres of the population may be estimated.
Sampling distribution of the means : distribution of the mean of samples of particular size. **Sampling distribution of the variance** : distribution of the variance of samples.

STATISTIC : any function of sample data, containing no unknown parametres, such as mean, median, variance or standard deviation.
Population's statistics are generally indicated with Greek letters (μ, σ, etc.).
Sample's statistics are generally indicated with Latin letters (m, s, etc.).

RANDOM VARIABLE : a variable which takes values in certain range with probabilities specified by a PROBABILITY DENSITY FUNCTION or PROBABILITY MASS FUNCTION (ex. if we express head or tail as 0 or 1, the toss of a coin is a random variable).

EXPECTED VALUE : or **MEAN** of a random variable or a function of a variable E[X] or $\bar{X}$ is

1. $\sum_{i=1}^{n} x_i p(x_i)$ for a DISCRETE VARIABLE;
2. $\int xf(x)dx$ for a CONTINUOUS VARIABLE.

It has the following properties:
E[λ X] = λ E[X]
E[X $\pm$ Y] = E[X] $\pm$ E[Y]
E[X^2] = $E[X]^2$ + Var(X)
E[XY] = E[X]E[Y] + cov(X,Y)

VARIANCE : a measure of dispersion

- *Theoretical variance*

$$Var(X) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} = E[X^2] - E[X]^2$$

- *Sample variance*

$$s^2 = \frac{\sigma^2}{n}$$

STANDARD DEVIATION is the square root of the variation and has the same unit of the random variable

- *Theoretical standard deviation*

$$\sigma = \sqrt{Var(X)}$$

- *Sample standard deviation = Standard error*

$$s = \frac{\sigma}{\sqrt{n}}$$

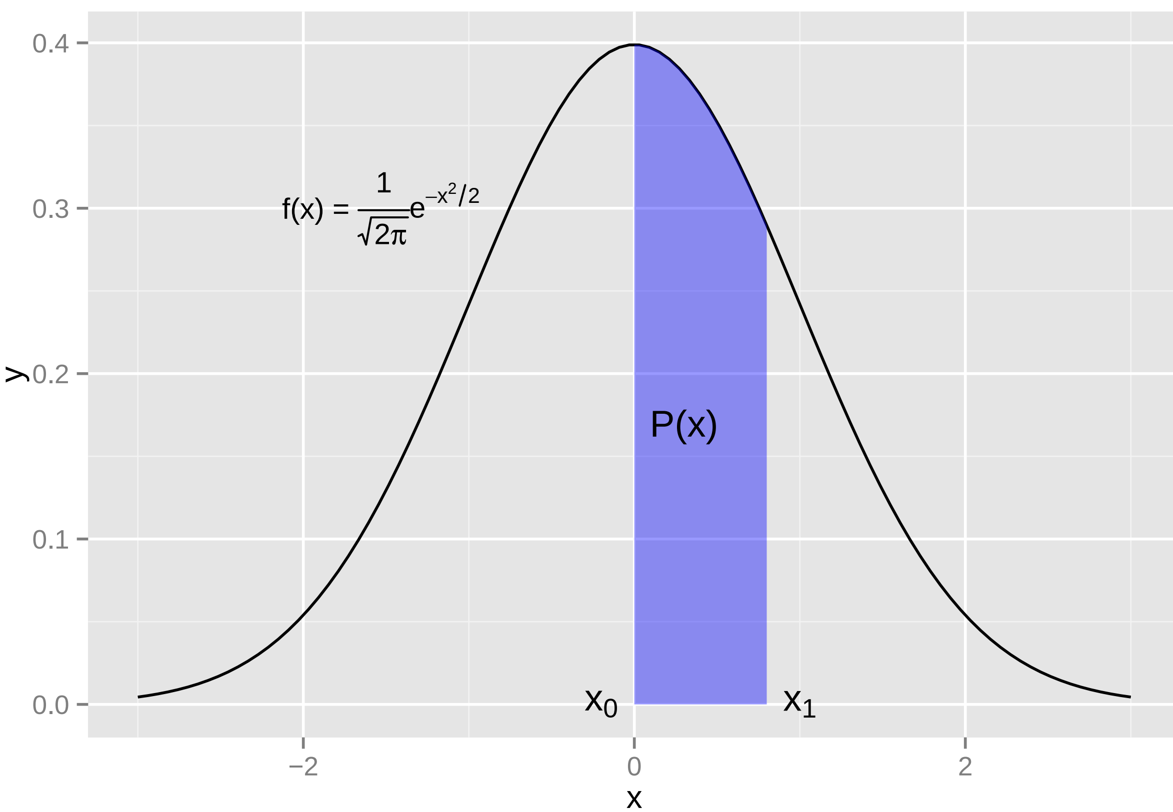
PROBABILITY MASS FUNCTION (PMF) : the distribution of the probability of the different values of a discrete random variable X. If the variable X takes values $x_1, ..., x_n$ with probabilities $p(x_1), ..., p(x_n)$, than:
1. $\sum_i p(x_i) = 1$
2. $p(x_i) \leq 0 \ \forall i$

PROBABILITY DENSITY FUNCTION (PDF) : the function f(x) of a continuous variable X such that:
1. the probability that X lies between x_0 and x_1 is $\int_{\mathrm{x_0}}^{\mathrm{x_1}} \mathrm{f(x)d}x$
2. the cumulative probability of the whole range of x is equal to 1 : $\int_{\mathrm{x}} \mathrm{f(x)d}x = 1$
3. the probability is always greater than 0 : $f(x) \leq 0 \forall x$

PROBABILITY : a measure of the relative frequency or likelyhood of occurrence of an event. Values are deived from a **theoretical distribution** or from **observations**.

- **P(x)** is the probability of the event x.
- $0 \leq P(x) \leq 1$
- **Discrete variables** : $\frac{number of required outcomes}{total number of possible outcomes}$
- **Continuous variables** : The relevant area under the graph of its probability density function f(x)

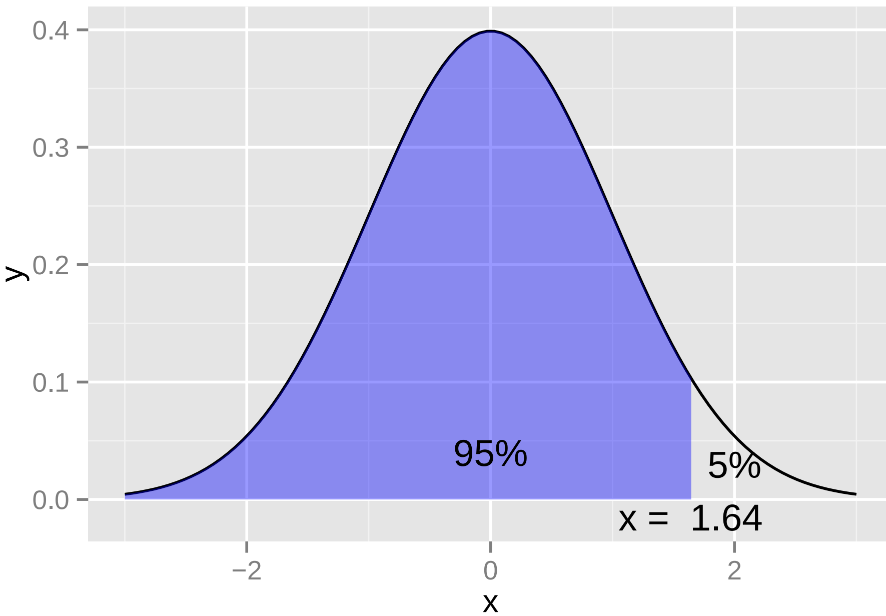
$$P(x) = \int_{\mathrm{x_0}}^{\mathrm{x_1}} \mathrm{f(x)d}x$$



QUANTILE : general name for the values of a variable which divides its distribution into equal groups

```
qnorm(.95)
```

```
## [1] 1.644854
```



CUMULATIVE DISTRIBUTION FUNCTION (CDF) : the function F(x) which gives the cumulative frequency

DISTRIBUTIONS

BERNOULLI DISTRIBUTION : a type of binomial distribution when the random variables take only values 0 or 1 with probability p and 1-p, respectively

- *Probability mass function* : $P(X = x_1) = p^{x_1}(1-p)^{1-x_1}$ with X = [0,1]
- *Expected value* : $\mu = p$
- *Variance* : $Var(X) = p(1-p)$

BINOMIAL : is obtained as the sum of a bunch of iid bernoulli random variables (ex. number of heads on a biased coin). Let $x_1, ..., X_n$ be an iid Bernoulli with probability p, then

$$X = \sum i = 1^n x_i$$

is a *binomial random variable* with mass function

$$P(X = x_i) = \binom{n}{k} p^{x_1}(1-p)^{1-x_1}$$

NORMAL DISTRIBUTION Gaussian distribution with mean μ and variance σ^2

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

- *Probability Mass Function* :

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

STANDARD NORMAL DISTRIBUTION : is the specific case of a normal distribution with $\mu = 0$ and $\sigma = 1$

$$X \sim N(0, 1)$$

with probability mass function

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