Normal Distribution - Review

Review Component

- **The Assumption of Normality** will be a learning outcome for this course, in almost every section.
- In preparation, a review of the **Normal Distribution** is advisable

Random Variables

A **random variable** is a numerical description of the outcome of an experiment.

1. Random variables are denoted by uppercase letters, **X** or **Y**.
2. The values that the random variable can take on are denoted by lowercase letters, **x** or **y**.
3. The values of the random variable together with the associated probabilities form a **probability distribution**.

Figure:

Random Variables

A **random variable** is a numerical description of the outcome of an experiment.

1. Random variables are denoted by uppercase letters, **X** or **Y**.
2. The values that the random variable can take on are denoted by lowercase letters, **x** or **y**.
3. The values of the random variable together with the associated probabilities form a **probability distribution**.

The probability distribution of a Random Variable

The **probability distribution** of a rand.var. **X** is a description of the probabilities associated with the possible values of X.

Notation: **P[X=** ***x*****]** the probability that the random variable X takes the value *x*.

Normal Distribution : Review

Continuous random variables

A **continuous random variable** can take any value in an interval.

Since any interval contains an infinite number of values, it is not possible to talk about the probability that the random variable **X** will assume a specific value ***x***.

In fact **P[X=x]=0** for any x.

Instead we calculate the probability that a continuous random variable will assume a value within a given interval
P[$a \leq X \leq b$**]** or **P[**$a \leq X$**]** or **P[**$X \leq b$**]**

Normal Distribution : Review

Example

Consider the experiment of picking a student at random from this class. We are interested in the blood pressure the student has. The outcome of the experiment can be **any** value $[0, +\infty]$.

The random variable: **X** = blood pressure.
Possible value of the rand.var.: **x**= 100mm/Hg, but P[X=100]=0

Probabilities of interest: $P[X \geq 100]$
Note: $P[X \geq 100] = P[X > 100]$

Normal Distribution : Review

Density Function

Let **X** be a continuous random variable.
Repeat the random experiment which produces **X** many times (∞) and draw a histogram using very small interval classes.
The histogram can then be approximated by a curve
f(x)=probability density function of X.
Properties of density curve:

- the total area under the curve = **1**
- the area under the curve between two points **a** and **b** is the probability that X lies between *a* and *b*, **P[**$a \leq X \leq b$**]**

Normal Distribution : Review

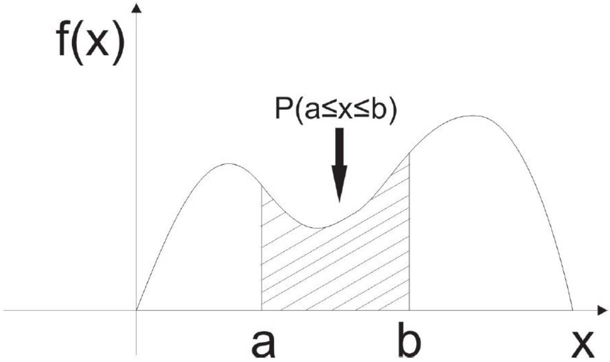


Figure: Probability density function.

Normal Distribution : Review

Mathematical properties of f(x)

For a continuous random variable X, a function f(x) is a probability density function if:

1. $f(x_i) \geq 0$ for all x_i $\rightarrow$ f(x) is always positive

2. $\int_{-\infty}^{\infty} f(x)dx$=1 $\rightarrow$ area under f(x)=1

3. $P(a \leq X \leq b) = \int_a^b f(x)dx$=area under $f(x)$ from a to b, for any a and b.

Normal Distribution

The most used model for the distribution of a random variable is a normal distribution $N(\mu, \sigma^2)$.
The density function a normal variable is

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

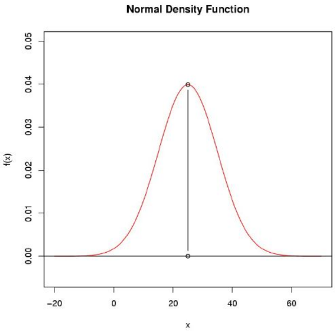
where

μ = expected value, or mean of the random variable X.
σ = standard deviation of the random variable X.

Normal Distribution : Review

Normal Density function

The mathematical shorthand for saying that a variable X is normally distributed with a mean of μ and a variance of σ^2 is $X \sim N(\mu, \sigma^2)$.



Normal Distribution : Review

Properties of $N(\mu, \sigma^2)$

- 1. Highest point on the normal curve is at the mean(μ), which is also the median and mode of the distribution.
- 2. The mean μ can be any numerical value: < 0, zero or > 0.
- 3. The curve is symmetric with respect to mean μ.
- 4. The standard deviation determines the width of the curve. Larger $\sigma \rightarrow$ in wider curves, showing more dispersion in the data. Smaller $\sigma \rightarrow$ taller curves.
- 5. The total area under the curve is 1.

Normal Distribution : Review

Properties of $N(\mu, \sigma^2)$ continued

Probabilities for some commonly used intervals are:

- 68% of the time, **X** assumes values within $\pm$ 1 σ from its mean μ.
- 95% of the time, **X** assumes values within $\pm$ 1.96 σ from its mean μ.
- 99% of the time, **X** assumes values within $\pm$ 2.58 σ from its mean μ.

Normal Distribution : Review

The Standard Normal Probability Distribution $N(0,1)$

A random variable $\mathbf{X} \sim N(\mu, \sigma^2)$ follows a normal distribution with a mean μ and standard deviation σ.

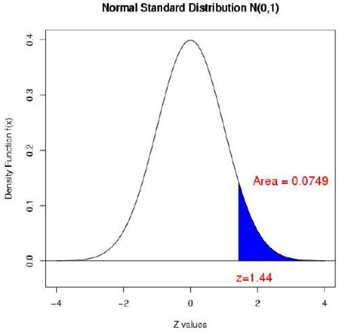
If $\mu = 0$ and $\sigma = 1 \rightarrow \mathbf{Z} \sim N(0,1)$ follows a **standard normal probability distribution**.

Convert $N(\mu, \sigma^2)$ to $N(0,1)$ with

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

Normal Distribution : Review

The probability that **Z** takes values greater than 1.44.



P[Z≥1.44]=0.0749 from Table 3.