**Chapter 13: Simple Linear Regression**

Linear Functions

Linear functions are straight-line upward or downward sloping functions. A linear function has one independent variable (x) and a dependent variable (y).

The format of a linear equation is $y=a\pm bx$

where $b$ is called the slope of the function. If $b>0$ then the function is upward sloping and if $b<0$ the function is downward sloping.

To sketch a linear function, you need to identify three things:

1. slope $(b)$
2. y-intercept: the value of y when x is equal to 0 $(a)$
3. x-intercept: the value of x when y is equal to 0

Join the x-intercept and y-intercept with an extended straight line. That will be your linear function.

Example: Sketch the following linear functions

$$\left(a\right) y=6+2x \left(b\right) y=21-7x \left(c\right) -9y=36+9x \left(d\right) 5y=-25-50x$$

A non-linear function is one where the relationship between the independent variable (x) and the dependent variable (y) cannot be represented by an upward/downward sloping straight line.

Some common **non-linear functions** are the quadratic, cubic, exponential and logarithmic functions.



Scatter diagram: The scatter diagram or scatterplot graphs pairs of numerical data, with one variable on each axis, to look for a relationship between them.

The local ice cream shop keeps track of how much ice cream they sell versus the noon temperature on that day. Here are their figures for the last 12 days:

****

From the scatterplot graph, it is easy to see that the relationship between temperature (x) and sales (y) is positive.

Suppose we now want to use an upward-sloping linear function to represent the relationship between temperature (x) and sales (y). This linear function is called a regression line.



Regression analysis is a statistical process for estimating the relationships among variables by applying a suitable function to represent that relationship. A regression model with only one independent variable is called a **Simple Regression**. A regression model that represents a straight-line relationship between the independent and dependent variables is called a **Simple Linear Regression**. A regression model with more than one independent variable is called a **Multiple Regression model**.

Examples: The diagram on the left is an example of Simple Non-Linear Regression while the diagram on the right is an example of Simple Linear Regression.



The image below is an example of Multiple Regression Model.



In this chapter we will solely concentrate on Simple Linear Regression Models.

The regression model is given as $y=\hat{y}+e$ where$\hat{y}=a\pm bx$ the linear regression line that is representing the relationship between the dependent variable and the independent variable. $e$ is the error term that represents the difference between the observed values $y$ and the predicted values from the regression line $\hat{y}$. **The error term contains the effects of other independent variables besides** $x$ **that affect the value of** $y$**.**



Population versus Sample Data



The population error term is given by the symbol $ϵ$ and the sample error/residual term is given by the symbol e.

**The Least Squares Regression Line: This is the best fit line that minimizes the total sum of (squared) errors.**



The least squares regression line has the equation $\hat{y}=a\pm bx$, such that the values of $a$ and $b$ minimize the Sum of Squared Errors (SSE).

To find the values of $a$ and $b$ that minimize SSE, we use the following formulas:

