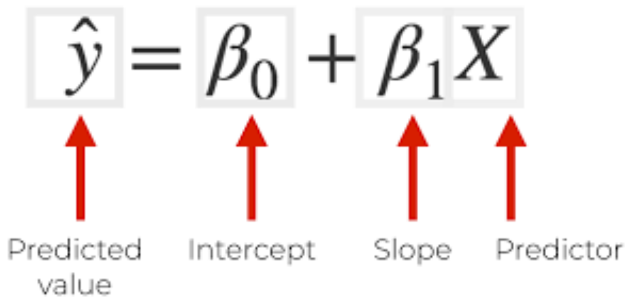
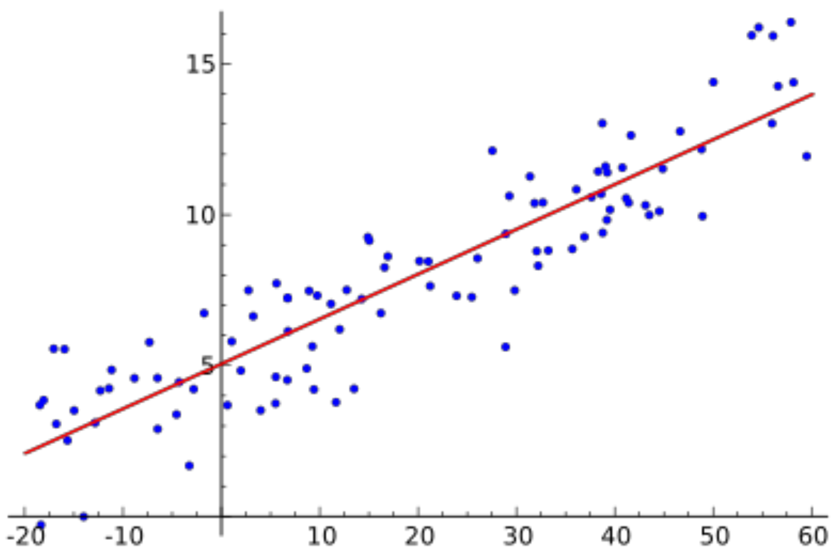
Examples of Linear Regression

What is Linear Regression?

- Linear regression is a way to explain the relationship between a **dependent variable** and **one or more explanatory variables** using a straight line.
- It is a special case of **regression analysis**.



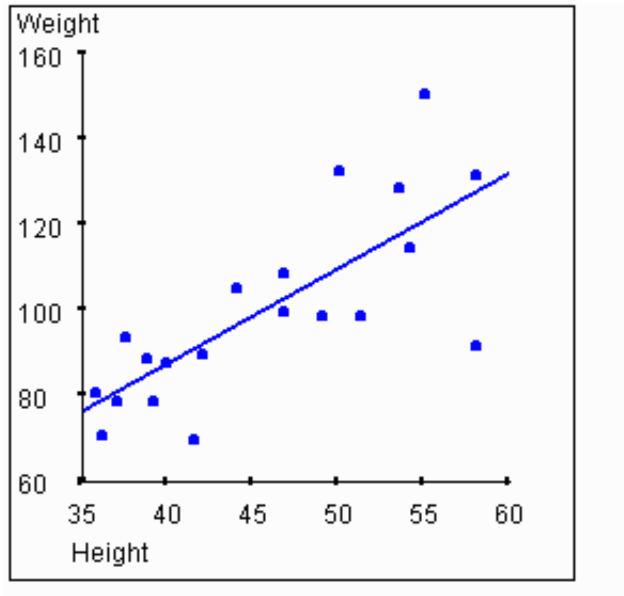
- Linear regression was the **first type** of regression analysis to be studied **rigorously**.
- This is because models which **depend linearly** on their unknown parameters are easier to fit than models which are **non-linearly related** to their parameters.



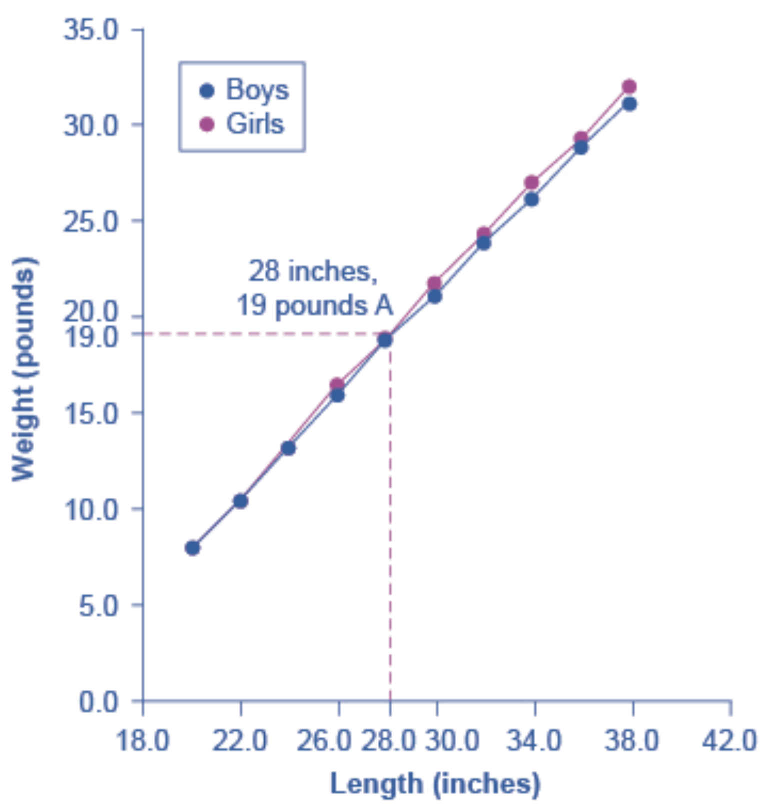
For Example

- Linear regression can be used to **fit a predictive model** to a set of observed values (data).

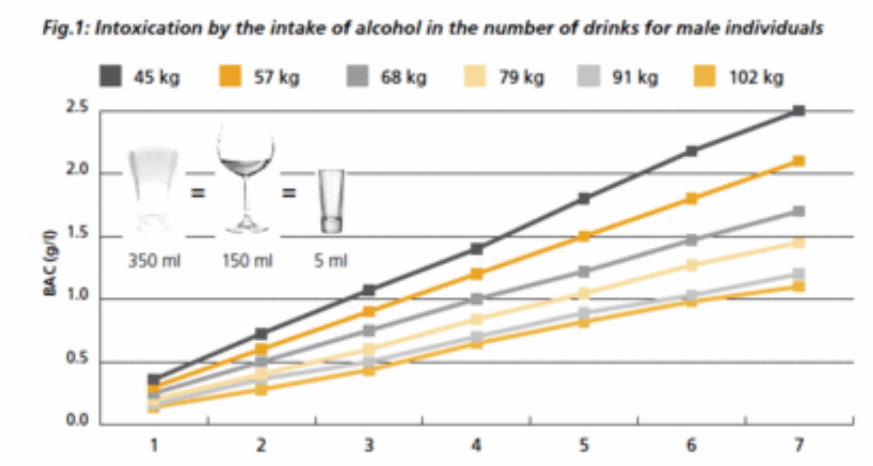
- This is useful, if the goal is **prediction, or forecasting, or reduction.**
- **Height and weight** as height increases, you'd expect weight to increase, but not perfectly.



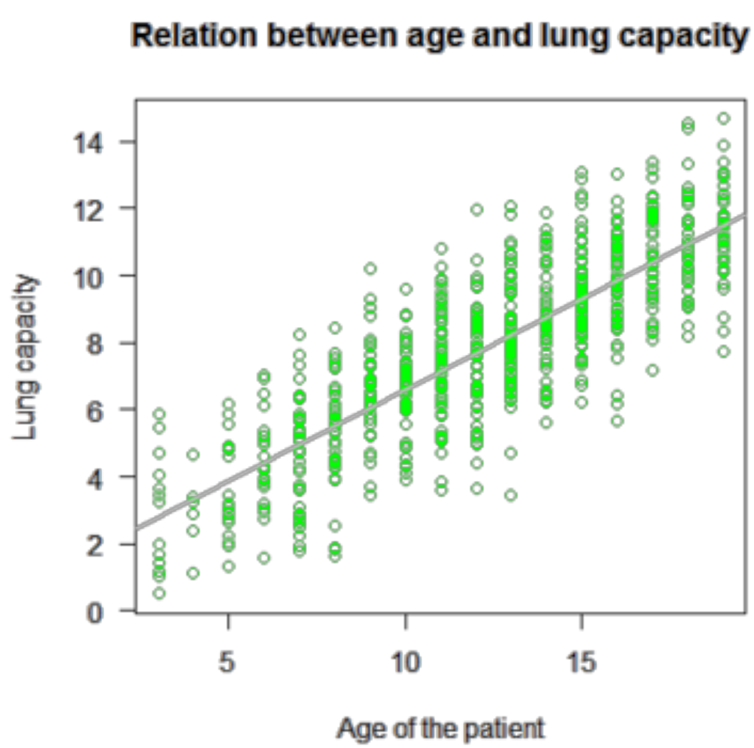
- Weight for **Age** as the baby grows older, the **weight increases.**



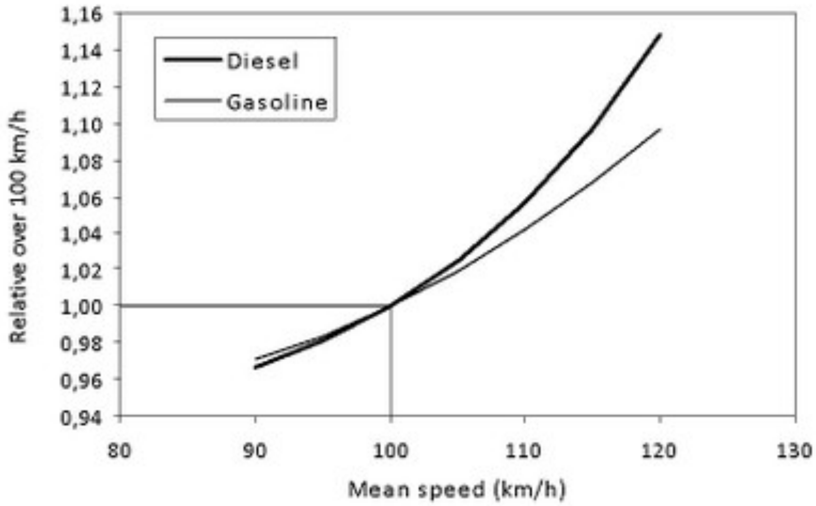
- **Alcohol consumed** and **blood alcohol** content as alcohol consumption increases, you'd expect one's blood alcohol content to increase, but not perfectly.



- **Vital lung capacity** and **age of a person** as age increases you'd expect lung function to decrease, but not perfectly.



- **Driving speed** and **gas mileage** as driving speed increases, you'd expect gas mileage to decrease, but not perfectly.



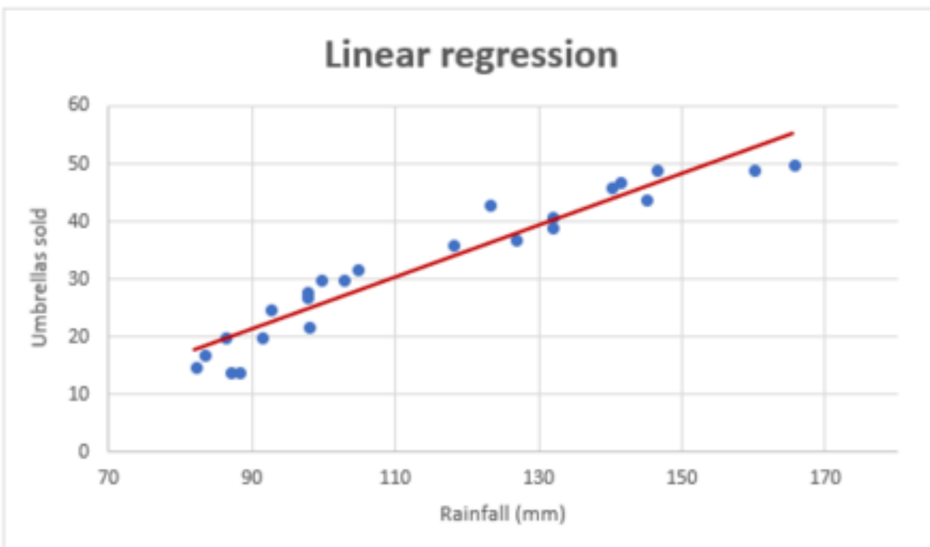
Features

- Simple linear regression is a statistical method that allows us to **summarize** and study relationships between **two continuous (quantitative) variables**.
- Linear regression models try to make the **vertical distance** between the **line** and **the data points** (e.g the residuals) as small as possible.
- This is called **fitting the line to the data**.
- Linear regression models try to **minimize** the **sum of the squares** of the residuals (least squares), but other ways of fitting exist.
- The **least squares** approach can also be used to fit models that are **not linear.**
- In simple linear regression, we **predict scores** on **one variable** from the scores on a **second variable**.
- The variable we are predicting is called the **criterion variable** and is referred to as **Y**.
- The **variable** we are basing our predictions on is called the **predictor variable** and is referred to as **X**.
- When there is only one predictor variable, the prediction method is called **simple regression**.
- In simple linear regression, the **predictions of Y** when plotted as a function of X form a **straight line**.

For example,

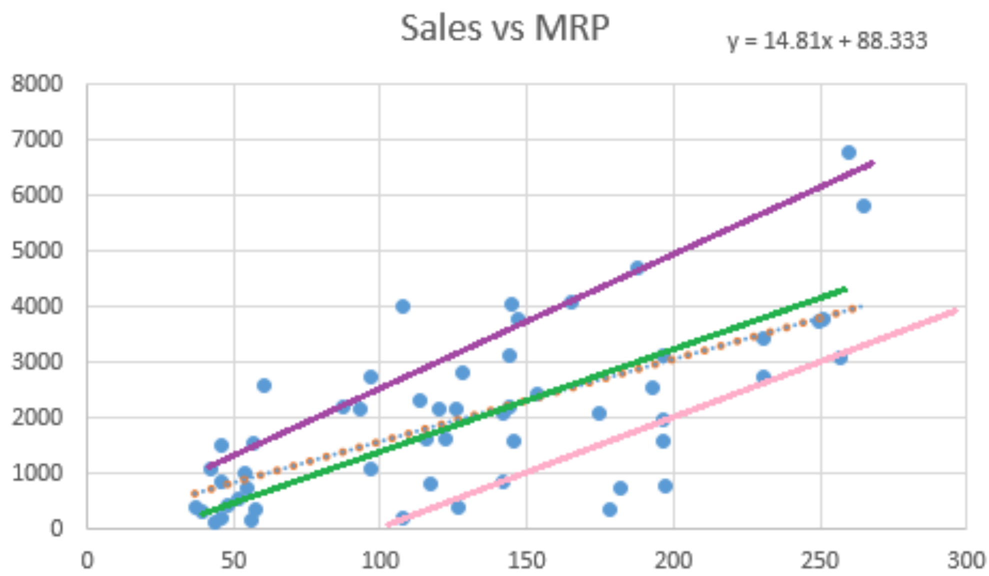
- Let’s take **sales numbers for umbrellas** for the last **24 months**

- Find out the **average monthly rainfall** for the same period.
- Plot this information on a chart, and the regression line will demonstrate the relationship between the **independent variable** (*rainfall*) and **dependent variable** (*umbrella sales*):

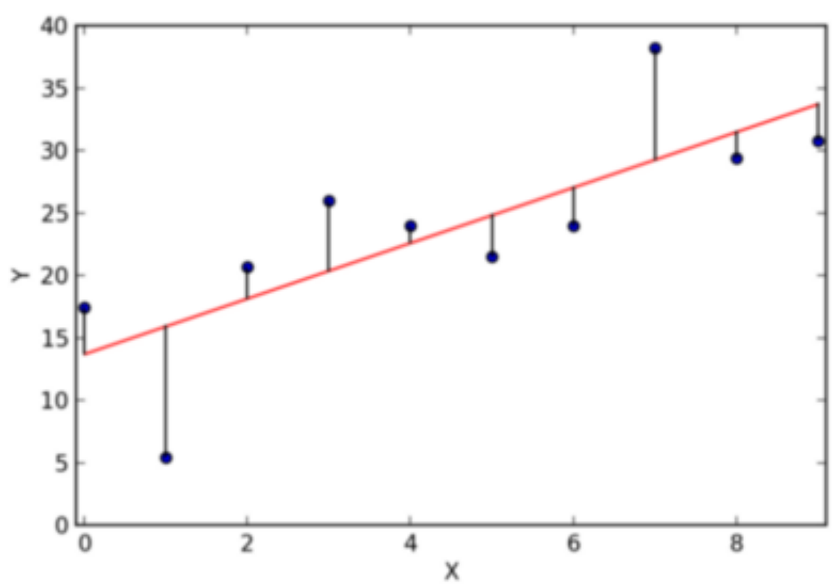


The Line of Best Fit

- There can be so many lines which can be used to **estimate Sales** according to their **MRP.**
- So how would you choose the **best fit line** or **the regression line**?



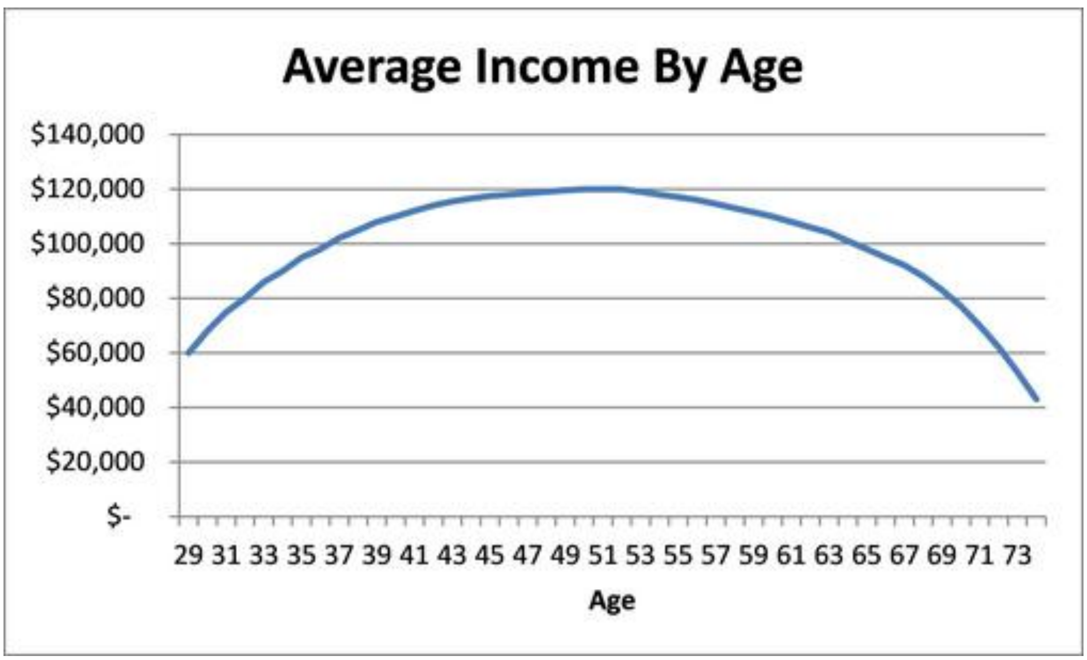
- The **main purpose** of the best fit line is that our **predicted values** should be **closer** to our actual or the **observed values**.
- There is no point in predicting values which are **far away** from the **real values**.
- In other words, we tend to **minimize** the difference between the **values predicted** by us and the **observed values**, and which is actually termed as **error**.
- Graphical representation of error is as shown below.
- These errors are also called as **residuals**.
- The **residuals** are indicated by the **vertical lines** showing the difference between **the predicted and actual value**.



Disadvantages

Linear Regression Is Limited to Linear Relationships

- Linear regression only looks at **linear relationships** between dependent and independent variables.
- That is, it assumes there is a **straight-line** relationship between them.
- Sometimes this is incorrect.
- For example:
    - The relationship between **income** and **age** is curved.



    - Income tends to rise in the **early parts of adulthood**, **flatten out** in later adulthood and **declin**e after people retire.
    - You can tell if this is a problem by looking at **graphical representations** of the relationships.

Linear Regression Only Looks at the Mean of the Dependent Variable

- Linear regression looks at a relationship between **the mean of the dependent variable** and the **independent variables**.
- **For example**
  - If you look at the relationship between the **birth weight of infants** and **maternal characteristics such as age**.
  - Linear regression will look at the **average weight of babies** born to **mothers of different ages**.
- However, sometimes you need to look at the **extremes** of the **dependent variable**.
- **For Example**
  - **Babies are at risk** when their weights are low, so you would want to look at the **extremes** in this example.
- Just as the mean **is not a complete** description of a **single variable**, linear regression is not **a complete description** of relationships among **variables**.