



Processing and Analysis of Data

What can we find from this data?

Respondent	Weight
1	72
2	75
3	70
4	60
5	70
6	86
7	78
8	69
9	75
10	72
11	62
12	66
13	67
14	72

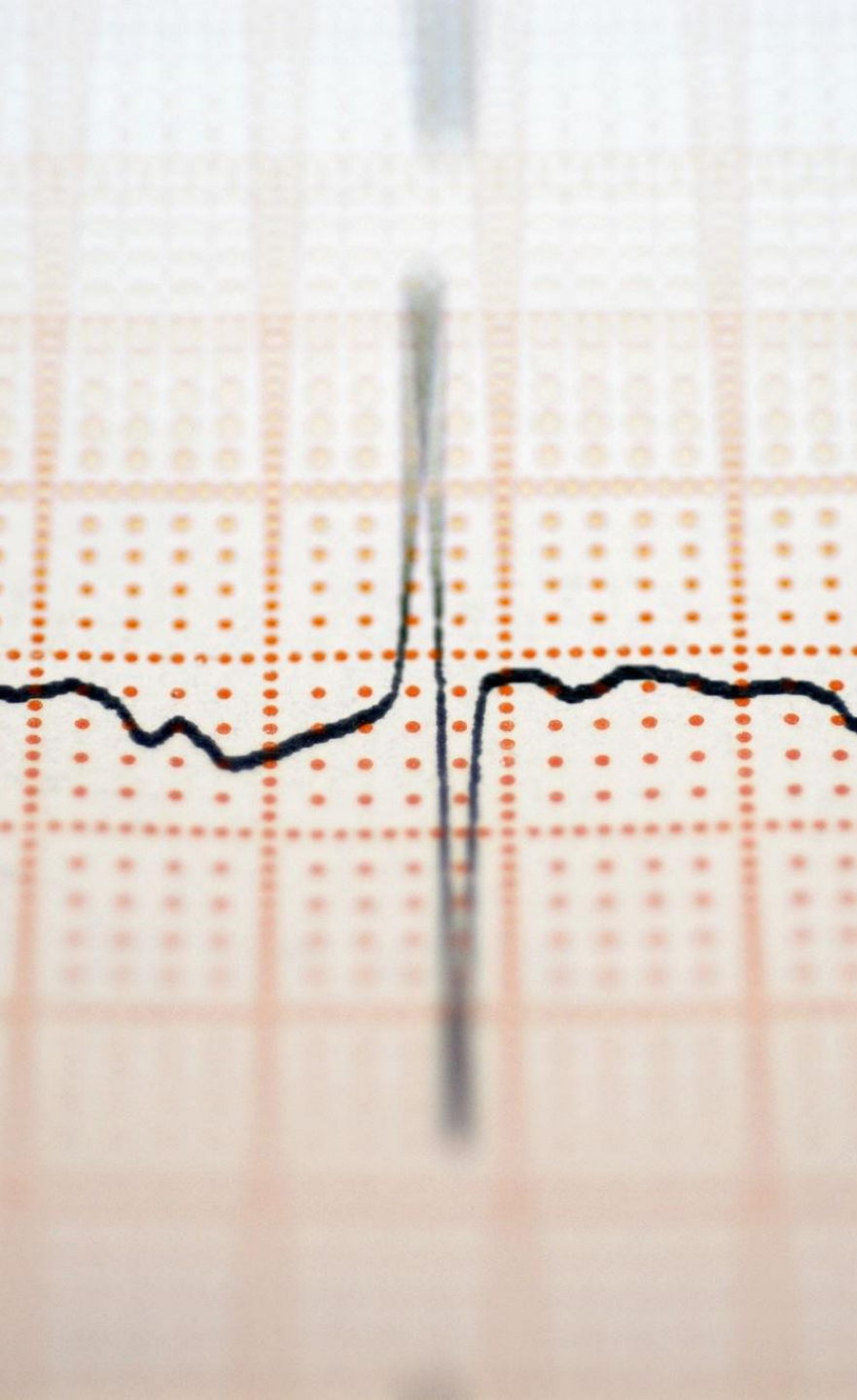
Measures of central tendency

- Measures used to describe the center or average of a distribution
- Mean
 - Sum of all values divided by the number of values
- Median
 - The middle value when the data is arranged in order
- Mode
 - Most frequently occurring value



Formula for
finding out
Sample Mean

$$\bar{X} = \frac{\Sigma X}{N}$$



Measures of Dispersion

- Measures used to describe the spread or variability of data.
- The range
 - the difference between the highest and lowest values.
- Variance
 - the average squared deviation from the mean.
- Standard deviation
 - the square root of the variance and provides a measure of the average distance from the mean

Sample Variance

$$S_X^2 = \frac{\Sigma(X - \bar{X})^2}{N}$$

Standard Deviation

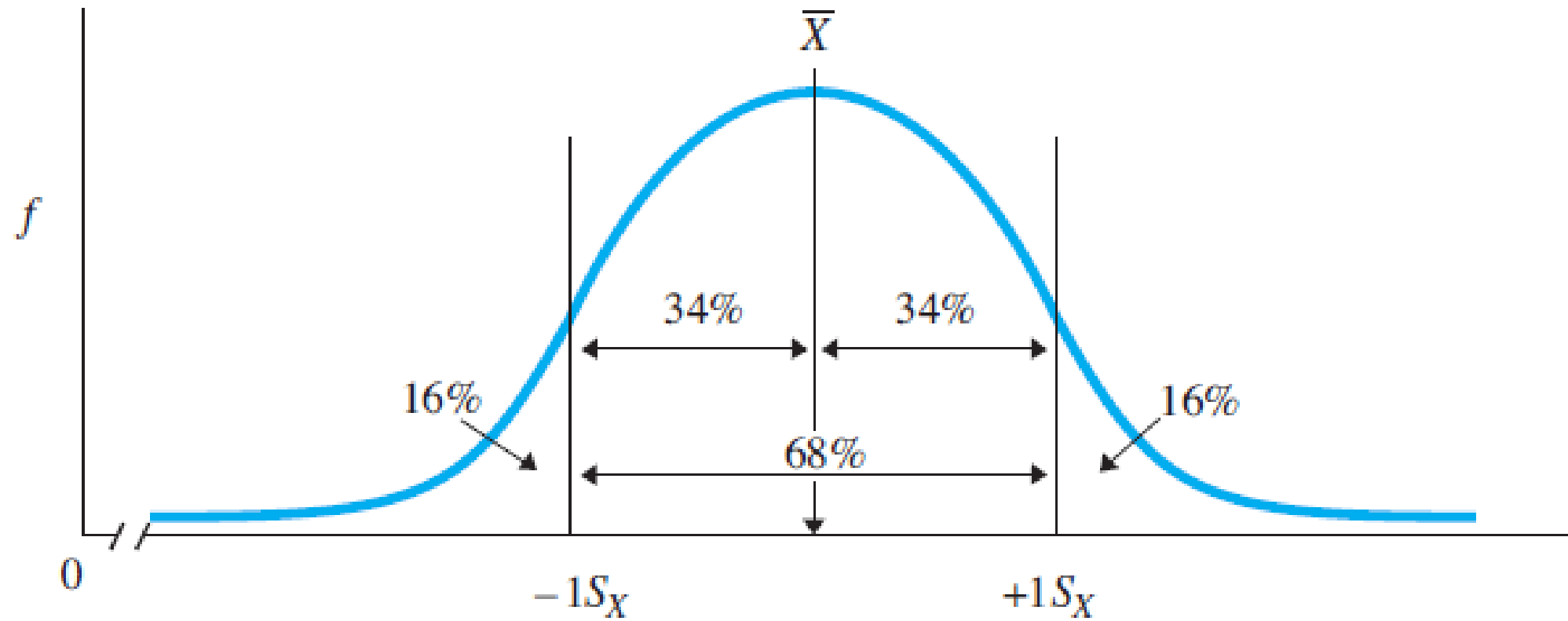
The *standard deviation* indicates

1. the “average deviation” from the mean,
 2. the consistency in the scores, and
 3. how far scores are spread out around the mean
- Approximately 34% of the scores in a normal distribution are between the mean and the score that is 1 standard deviation from the mean
 - Altogether, about 68% of the scores are between the scores at +1 SD and – 1 SD from the mean

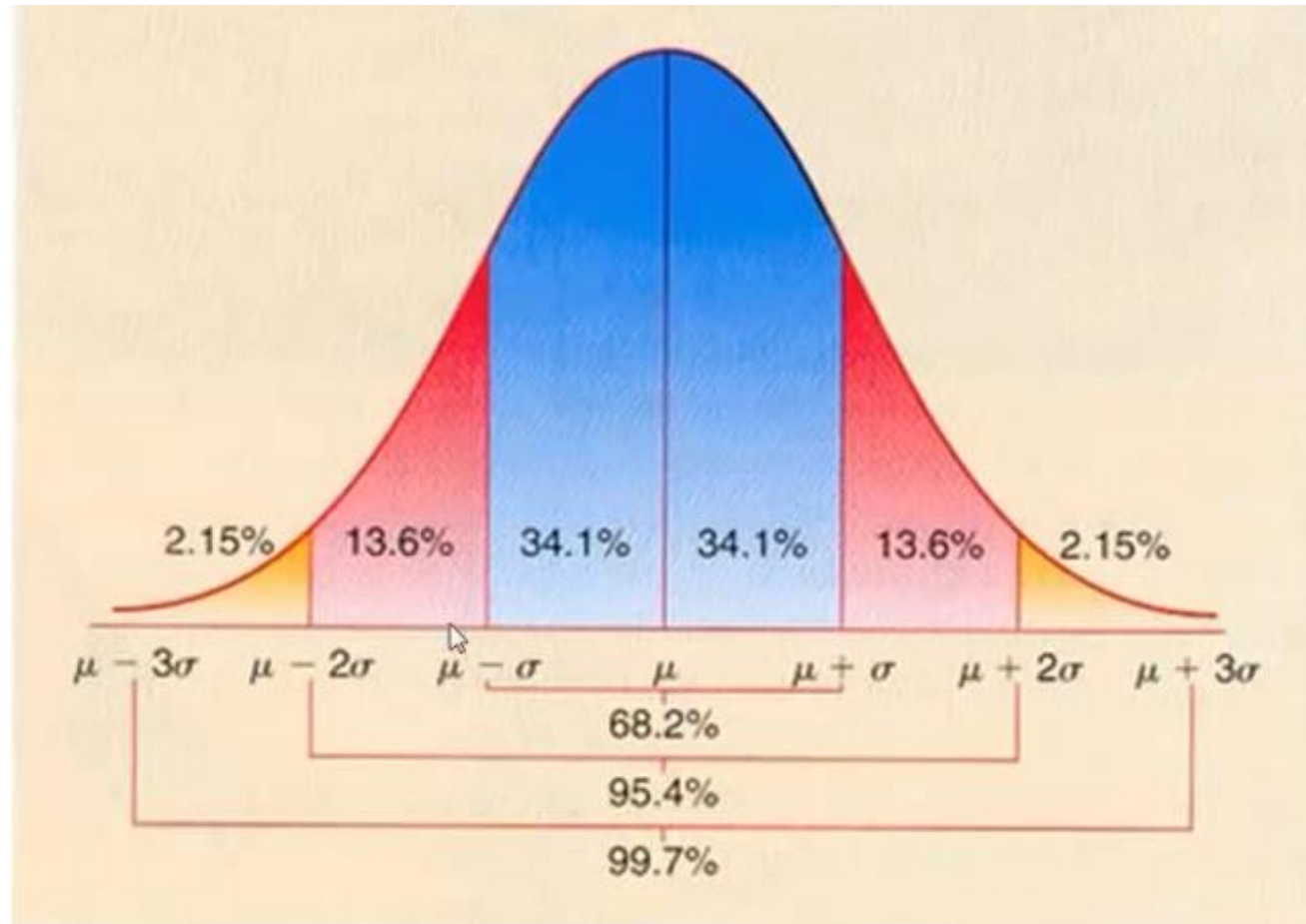
Sample Standard Deviation

$$S_X = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}}$$

Normal Distribution Curve



Coverage by Standard Deviation

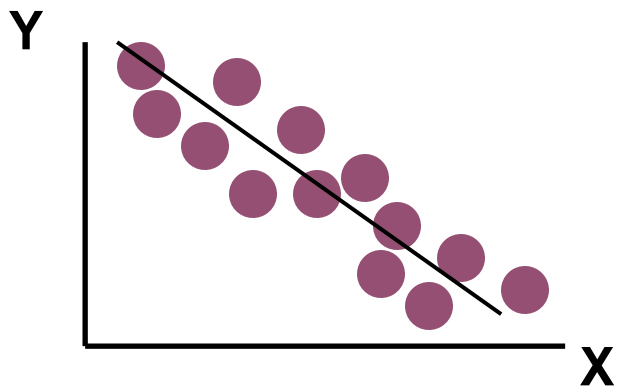
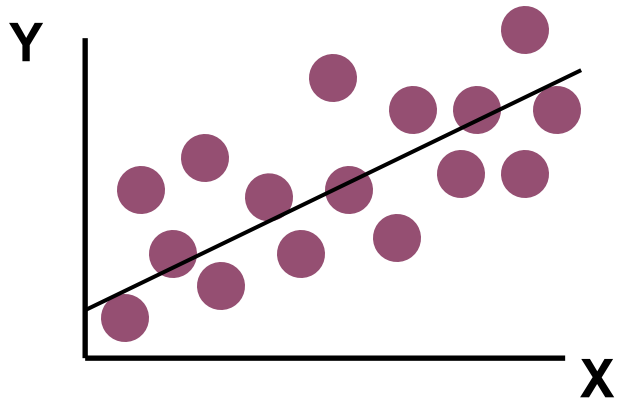


Simple linear regression

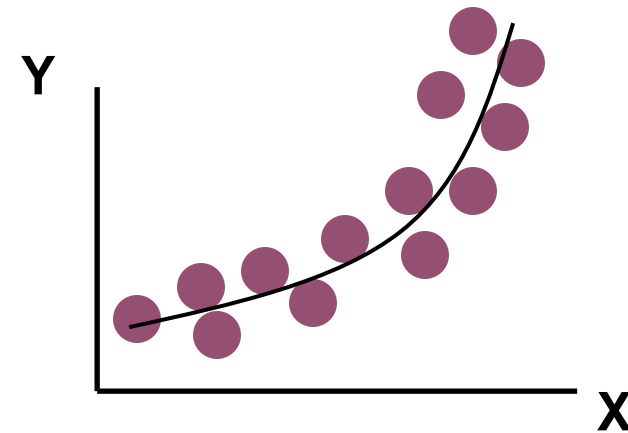
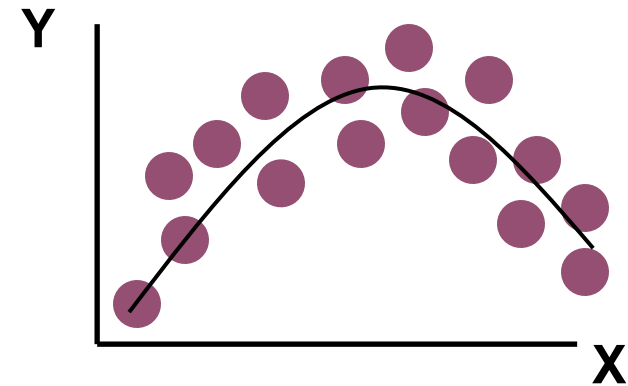
The background of the slide is black and features a series of white, wavy, horizontal lines that create a sense of motion and depth. These lines are more densely packed and varied in curvature on the right side, while they become more uniform and spread out towards the left.

Types of Relationships

Linear relationships

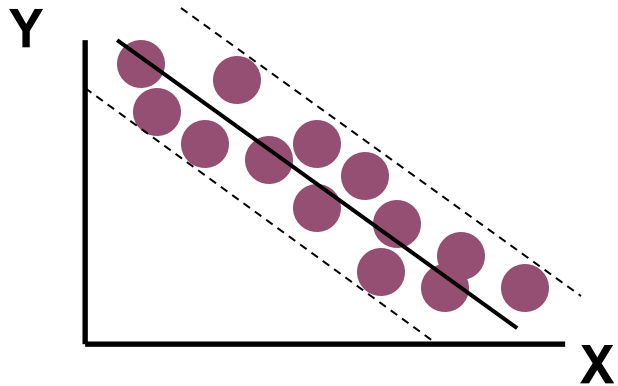
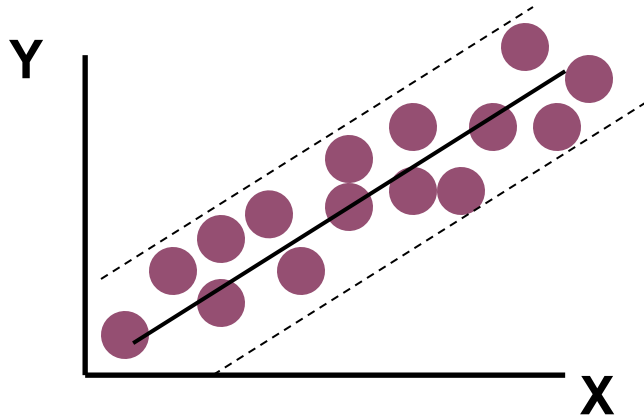


Curvilinear relationships

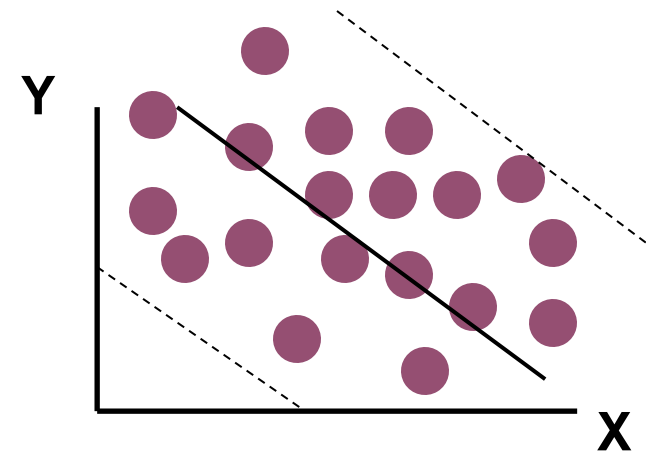
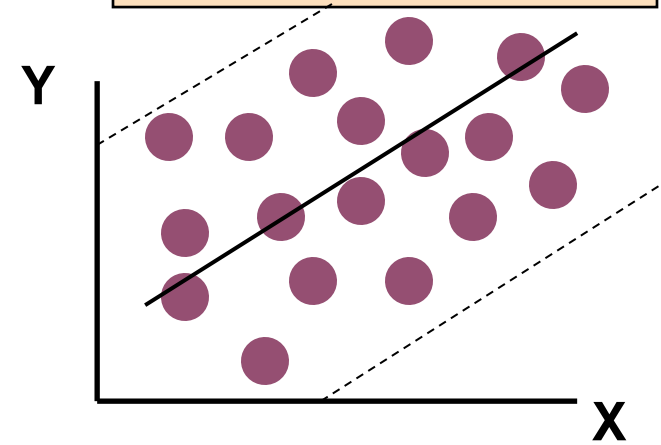


Types of Relationships

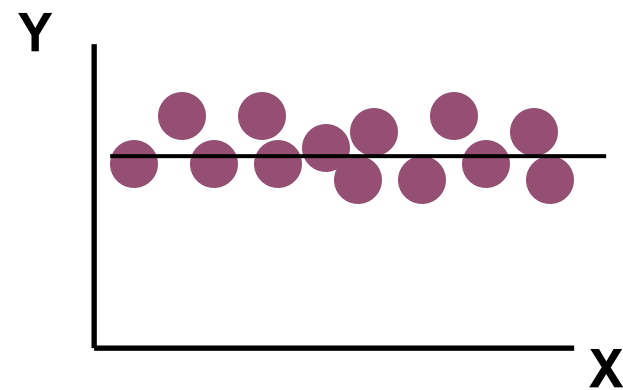
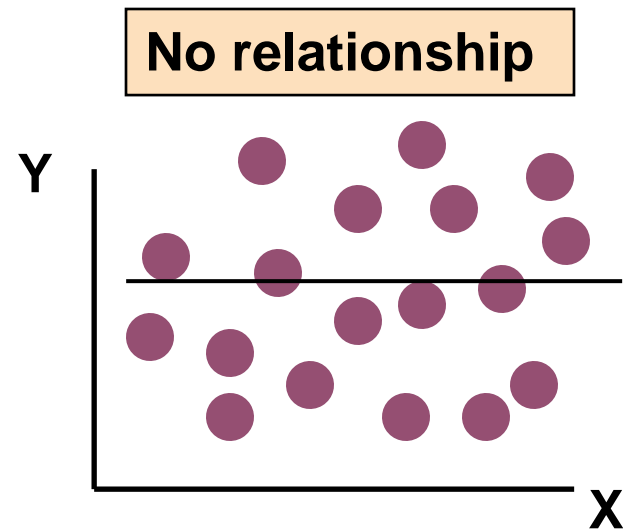
Strong relationships

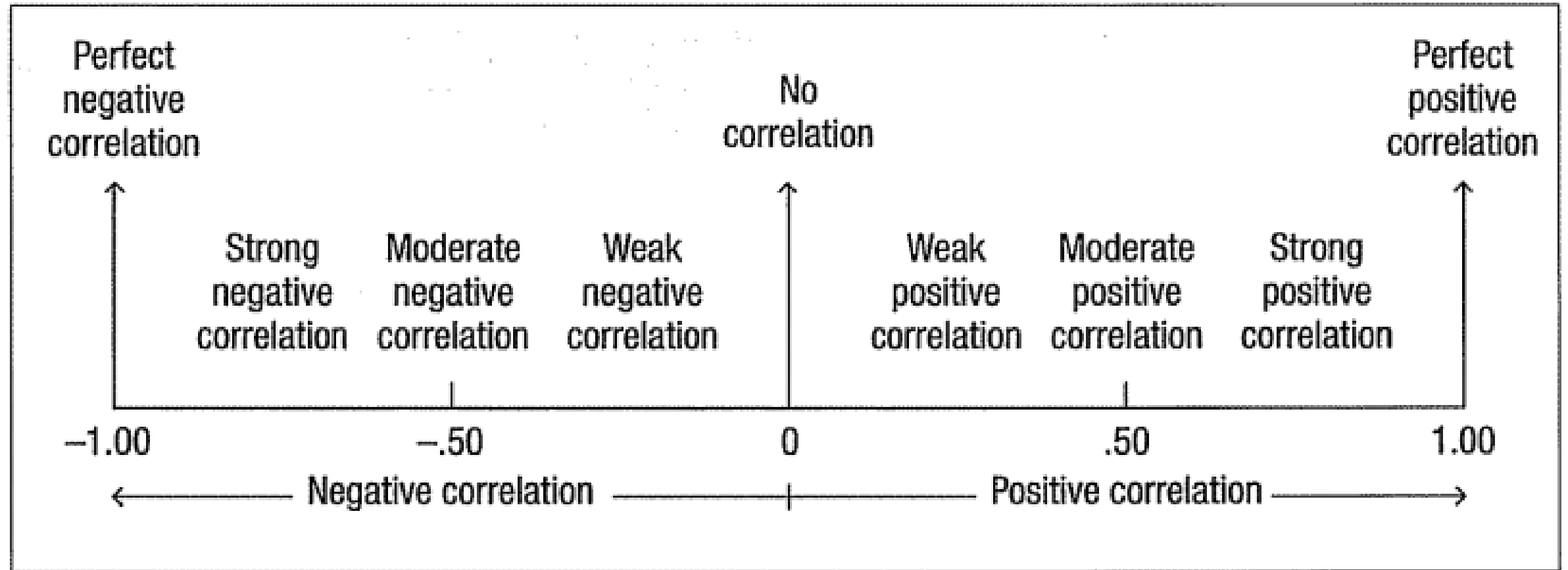


Weak relationships



Types of Relationships





Correlation

- Correlation measures the strength and direction of the relationship between two variables.
- Correlation coefficients range from -1 to +1, where -1 indicates a perfect negative correlation, +1 indicates a perfect positive correlation, and 0 indicates no correlation.

Regression

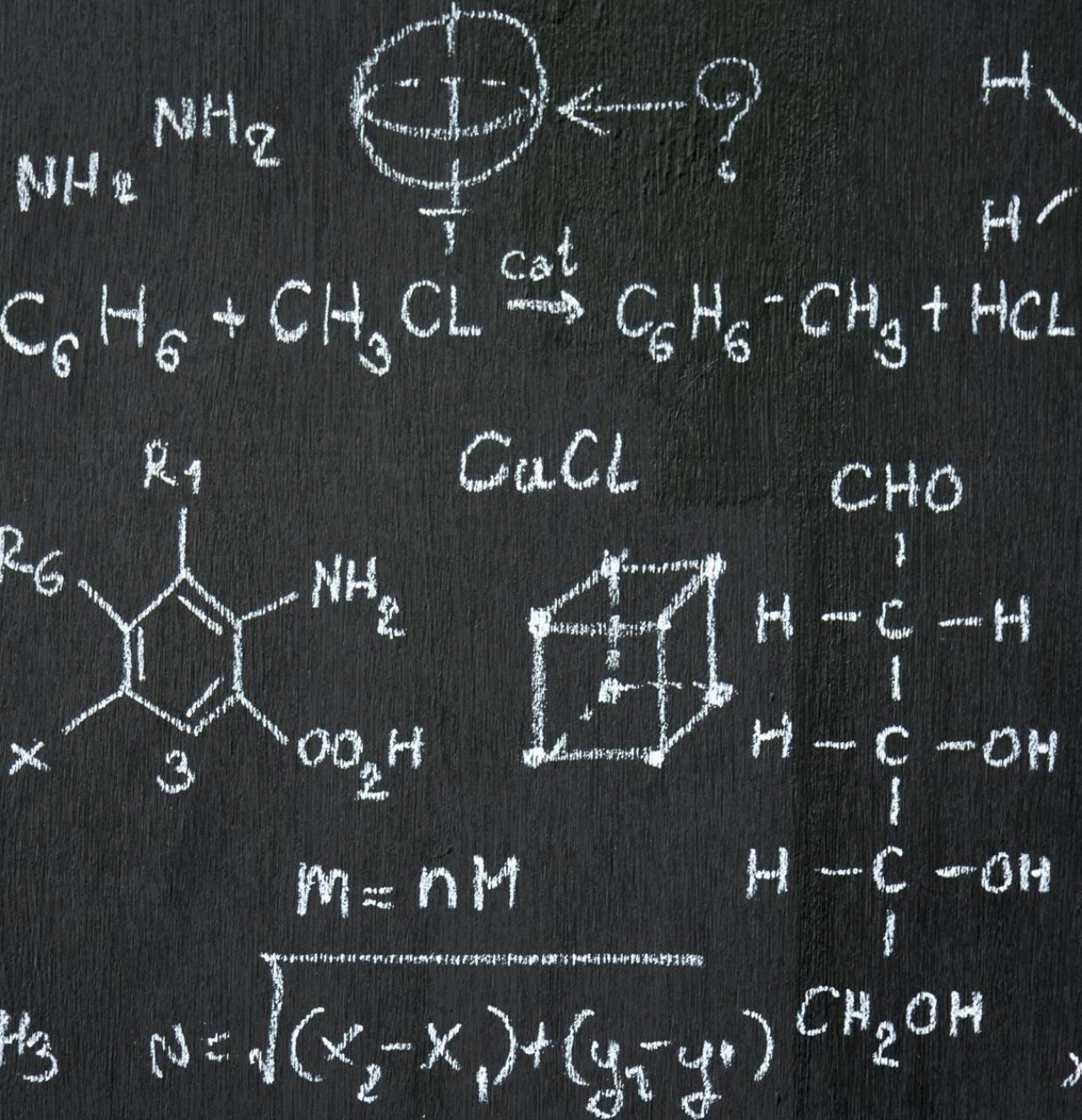
- Regression analysis helps predict or estimate the value of the dependent variable based on the independent variable(s).
- Regression analysis examines the relationship between a dependent variable and one or more independent variables.

Introduction to Regression Analysis

Regression analysis is used to:

- Predict the value of a dependent variable based on the value of at least one independent variable
- Explain the impact of changes in an independent variable on the dependent variable
- Dependent variable: the variable we wish to predict or explain
- Independent variable: the variable used to predict or explain the dependent variable

The image shows a chalkboard with handwritten mathematical derivations. At the top left, there is a graph of a function $y = g(x)$ with a secant line drawn through two points on the curve, labeled "Secant Lines". To the right, the definition of the derivative is written as $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$. Below this, the derivative of $f(x) = x^2$ is calculated using the difference quotient: $f'(x) = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h}$. This is then expanded to $= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h}$, which simplifies to $= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h}$. Finally, it is simplified to $= \lim_{h \rightarrow 0} h(2x + h)$. On the left side of the board, there are additional notes including $y = g(x)$, "Secant Lines", and the expression $x+h$.



Simple Linear Regression Model

- Only **one** independent variable, X
- Relationship between X and Y is described by a linear function
- Changes in Y are assumed to be related to changes in X

Simple Linear Regression Model

The diagram illustrates the Simple Linear Regression Model equation $Y = a + bX$. The components are labeled as follows:

- Dependent Variable**: Points to Y .
- Population Y intercept**: Points to a .
- Population Slope Coefficient**: Points to b .
- Independent Variable**: Points to X .

The equation is $Y = a + bX$. A bracket under the terms $a + bX$ is labeled "Linear component".

Simple Linear Regression Model

The diagram illustrates the Simple Linear Regression Model equation, $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, with labels and annotations:

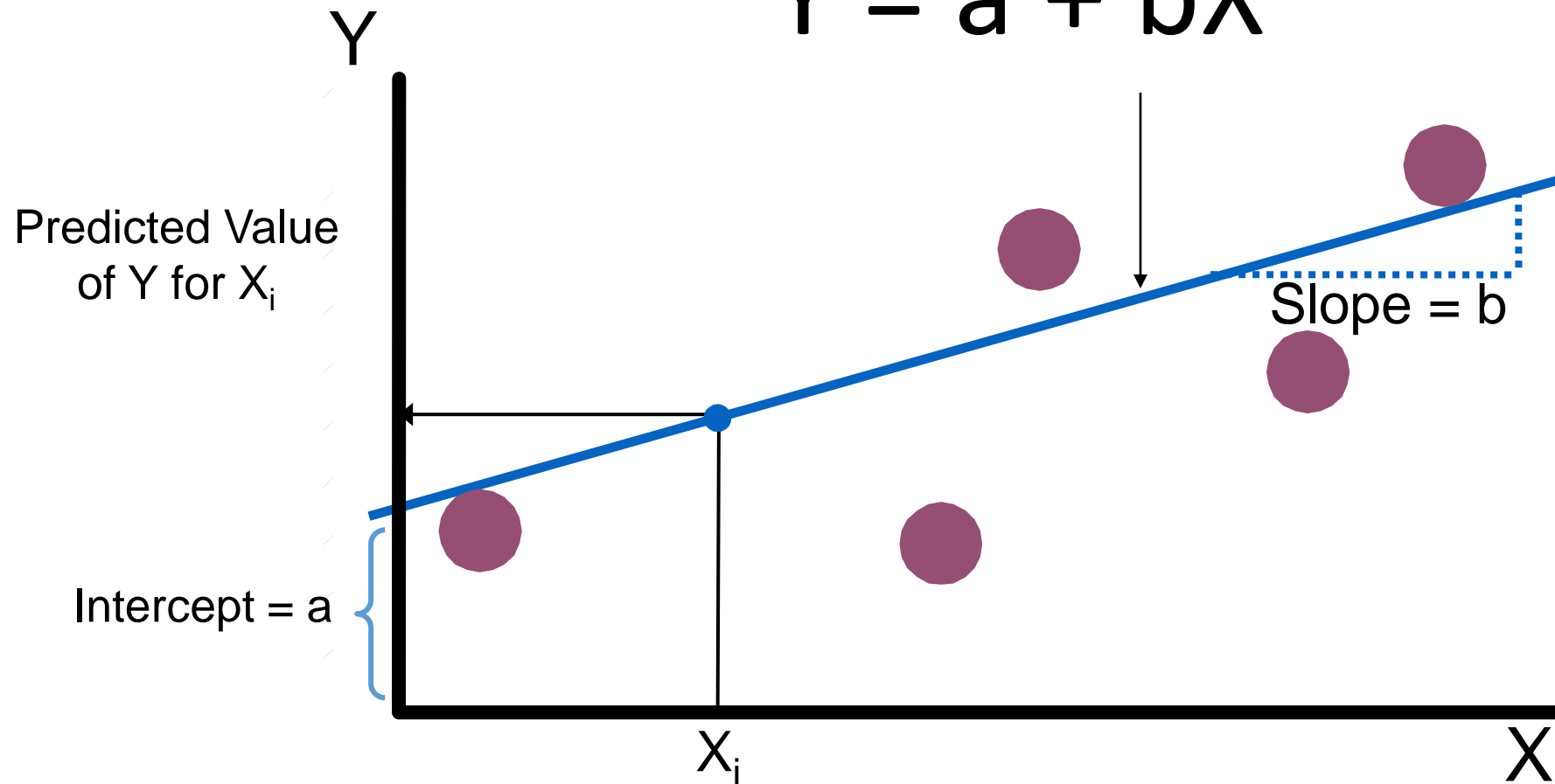
- Dependent Variable:** Points to Y_i .
- Population Y intercept:** Points to β_0 .
- Population Slope Coefficient:** Points to β_1 .
- Independent Variable:** Points to X_i .
- Random Error term:** Points to ϵ_i .

The equation is presented within a light orange rectangular box. Below the box, two purple curly braces group the terms:

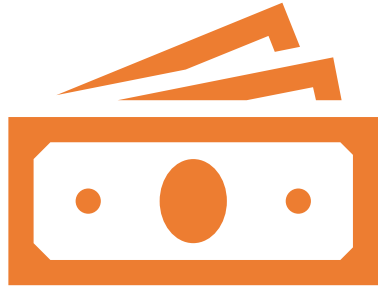
- Linear component:** Groups $\beta_0 + \beta_1 X_i$.
- Random Error component:** Groups ϵ_i .

Simple Linear Regression Model

$$Y = a + bX$$



Interpretation of the Slope and the Intercept



a is the estimated mean value of Y when
the value of X is zero



b is the estimated change in the mean value
of Y as a result of a one-unit increase in X

Correlation vs. Regression

- A scatter plot can be used to show the relationship between two variables
- Correlation analysis is used to measure the strength of the association (linear relationship) between two variables
 - Correlation is only concerned with strength of the relationship
 - No causal effect is implied with correlation

