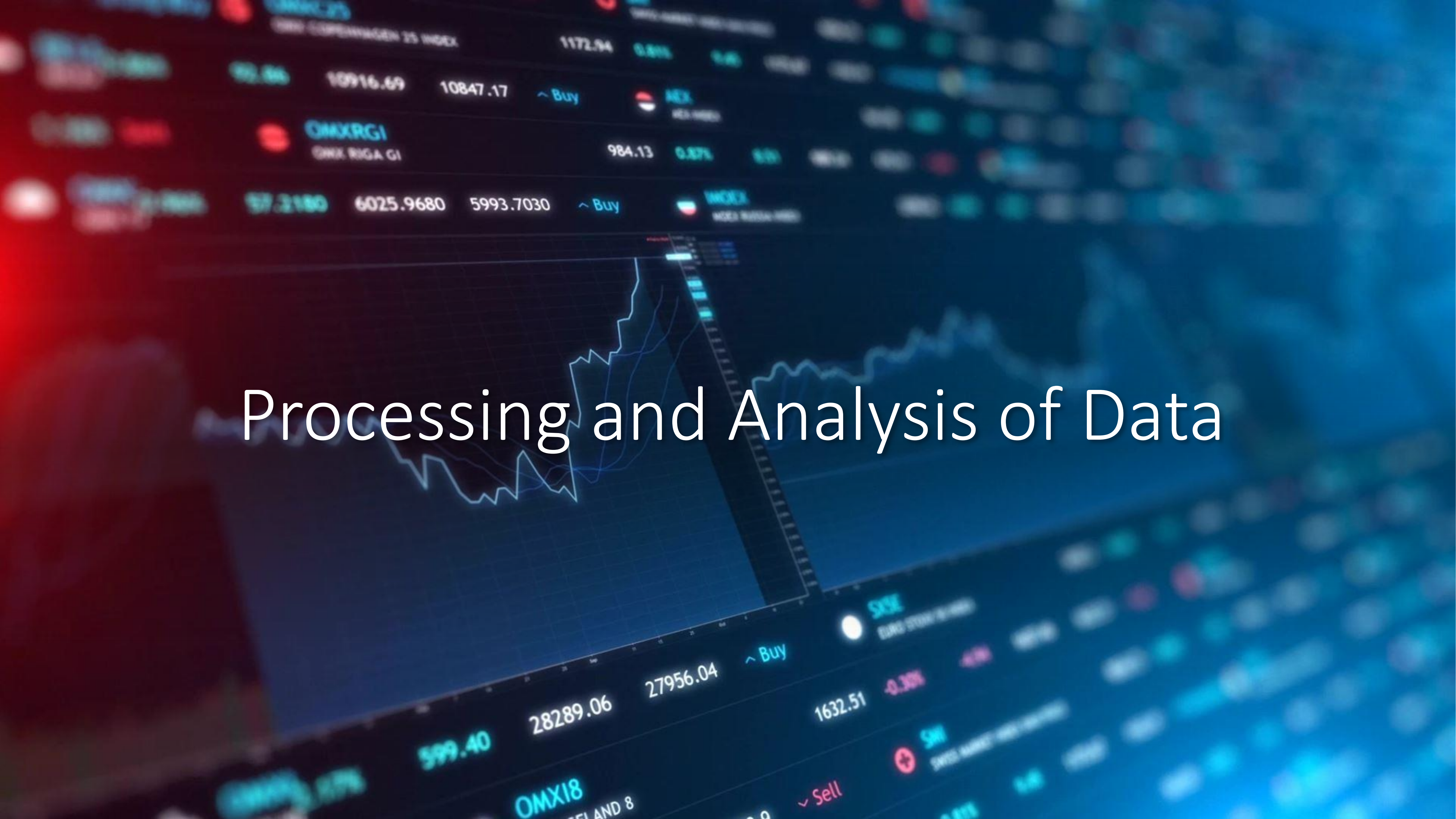


Processing and Analysis of Data

What can we find from this data?

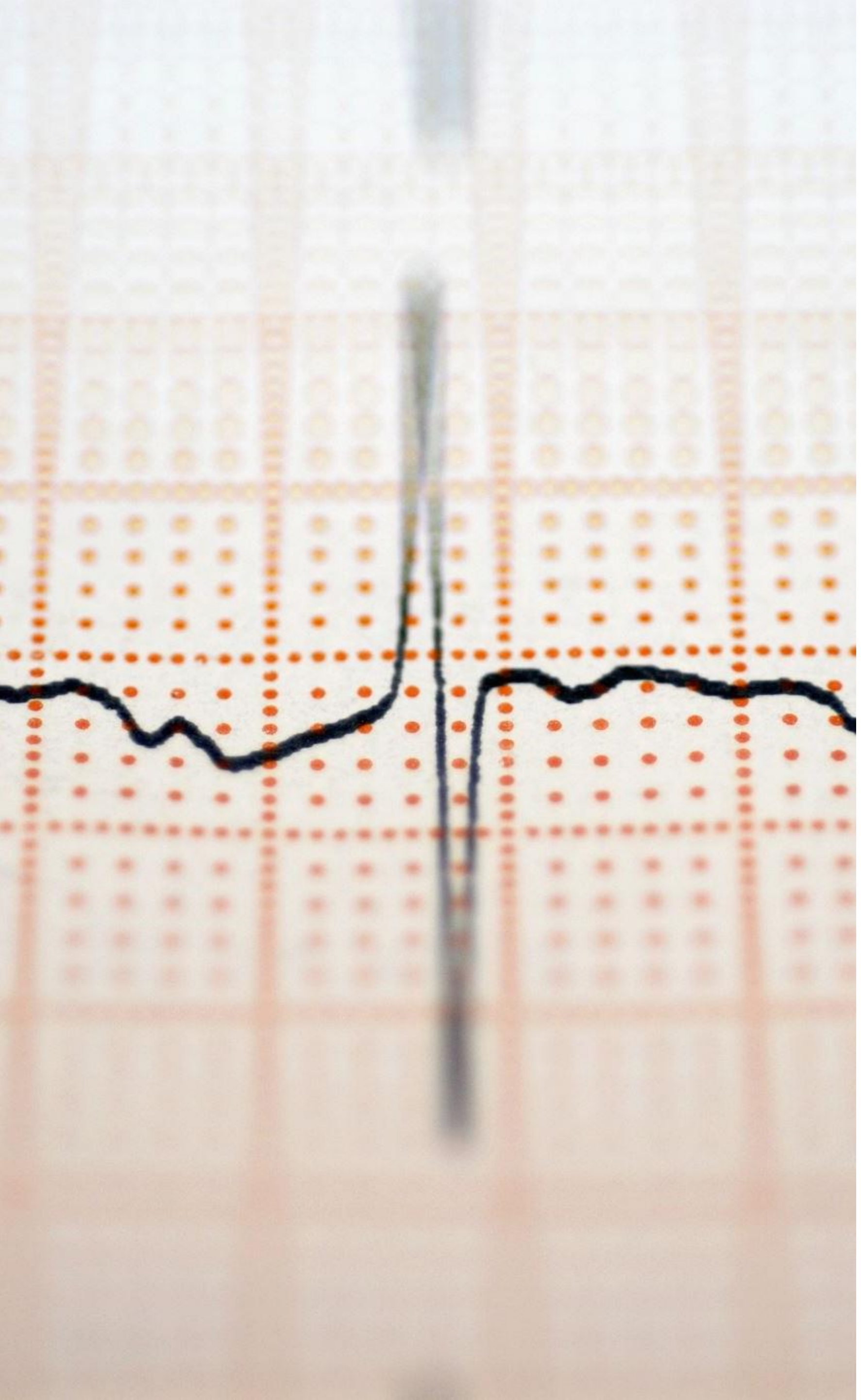
Respondent	Weight
1	72
2	75
3	70
4	60
5	70
6	86
7	78
8	69
9	75
10	72
11	62
12	66
13	67
14	72

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

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

Standard Deviation

The *standard deviation* indicates

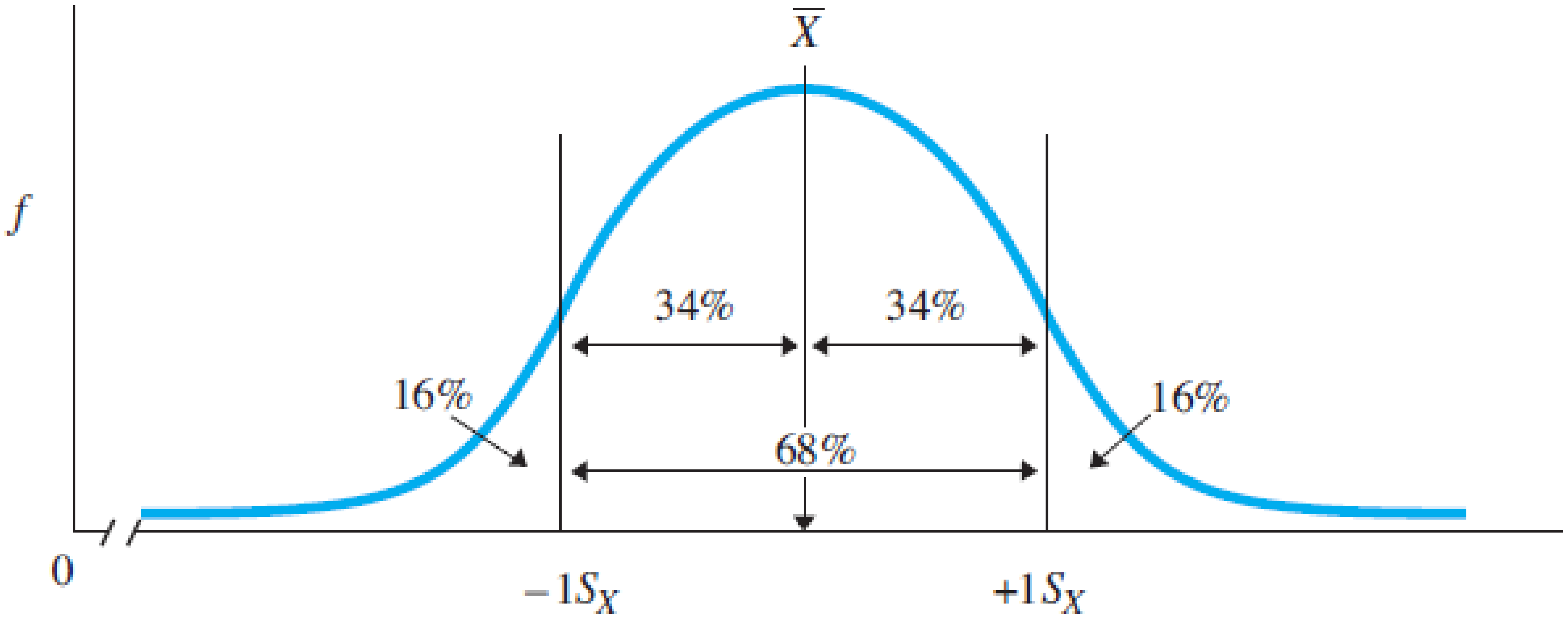
1. the “average deviation” from the mean,
2. the consistency in the scores, and
3. how far scores are spread out around the mean

- 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

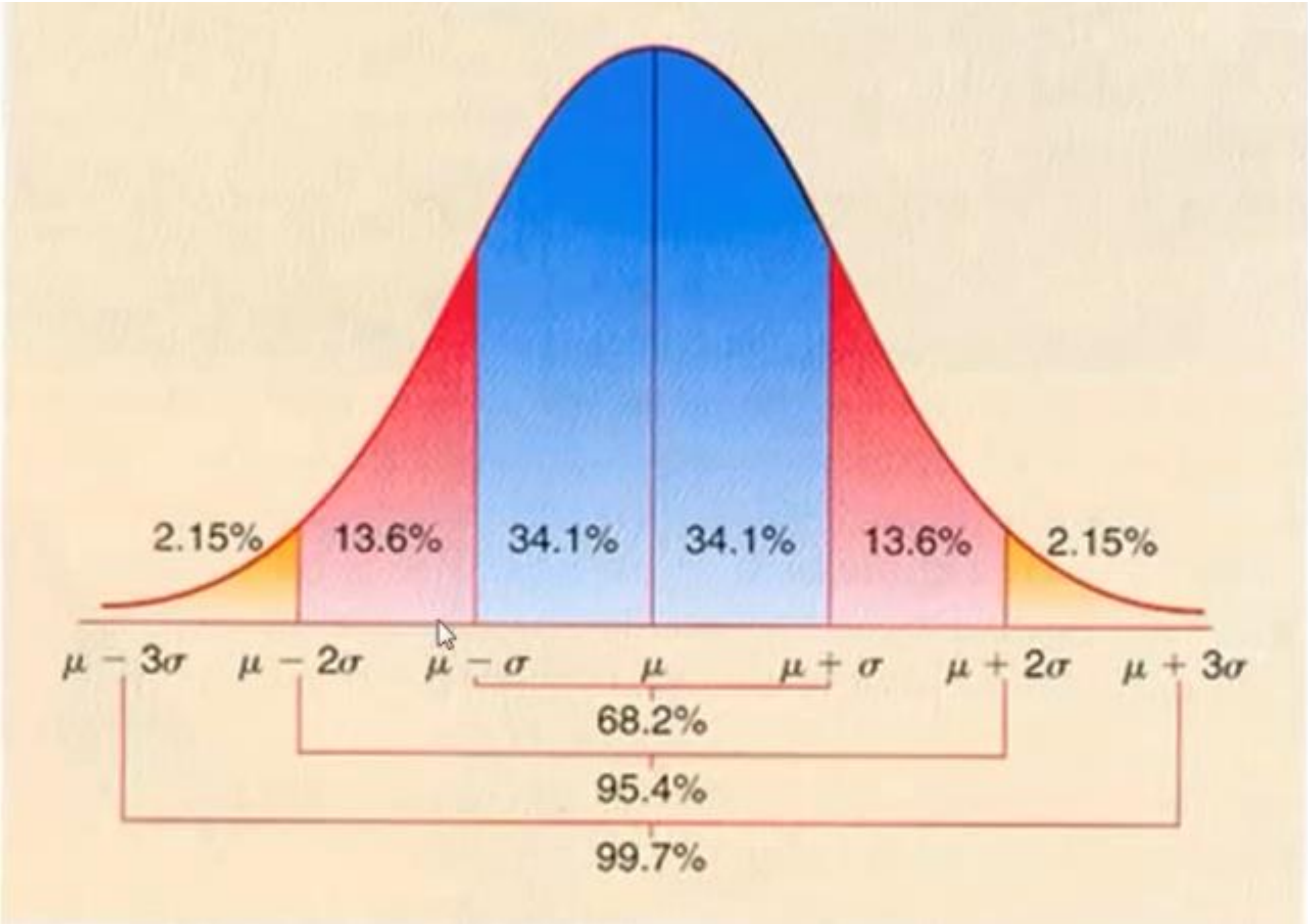
Sample Standard Deviation

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

Normal Distribution Curve



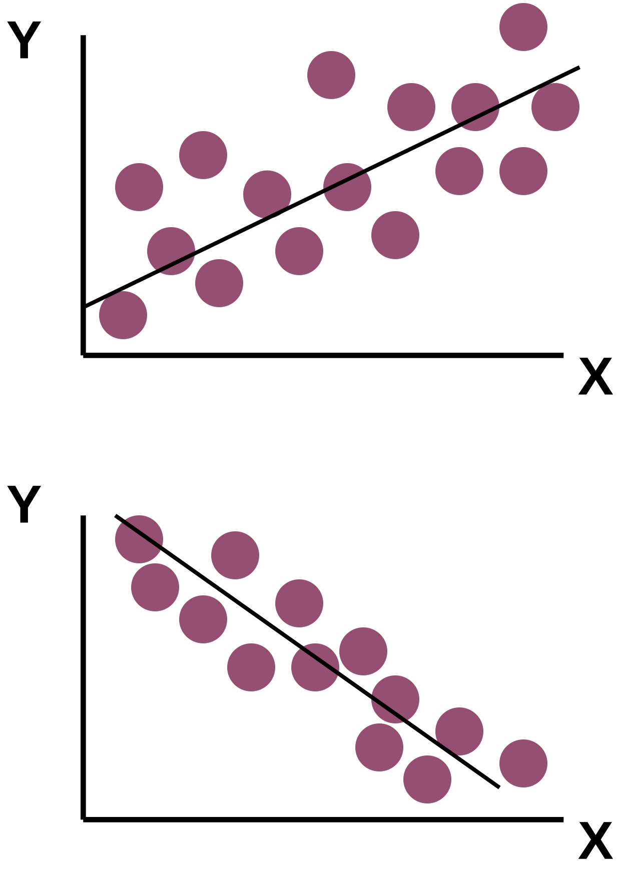
Coverage by Standard Deviation



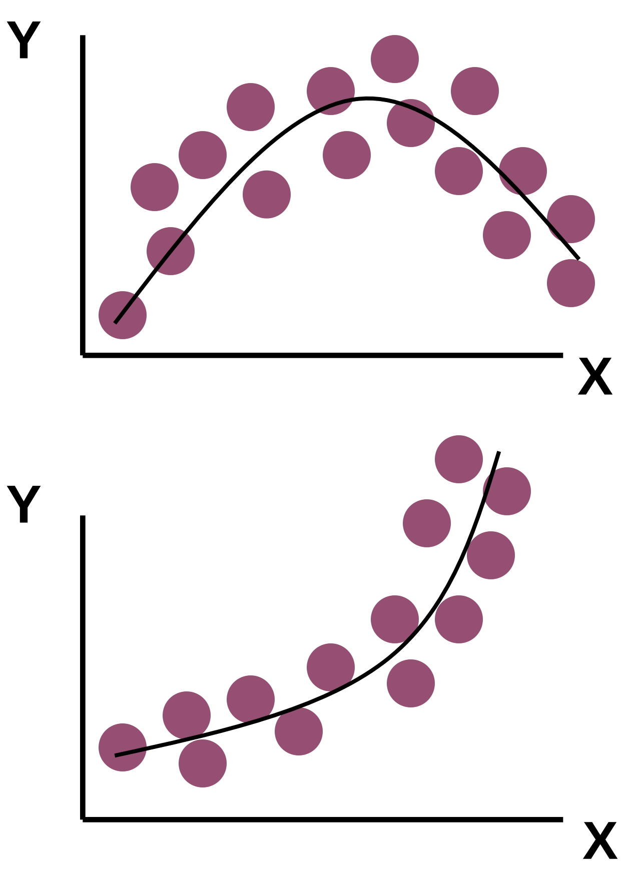
Simple linear regression

Types of Relationships

Linear relationships

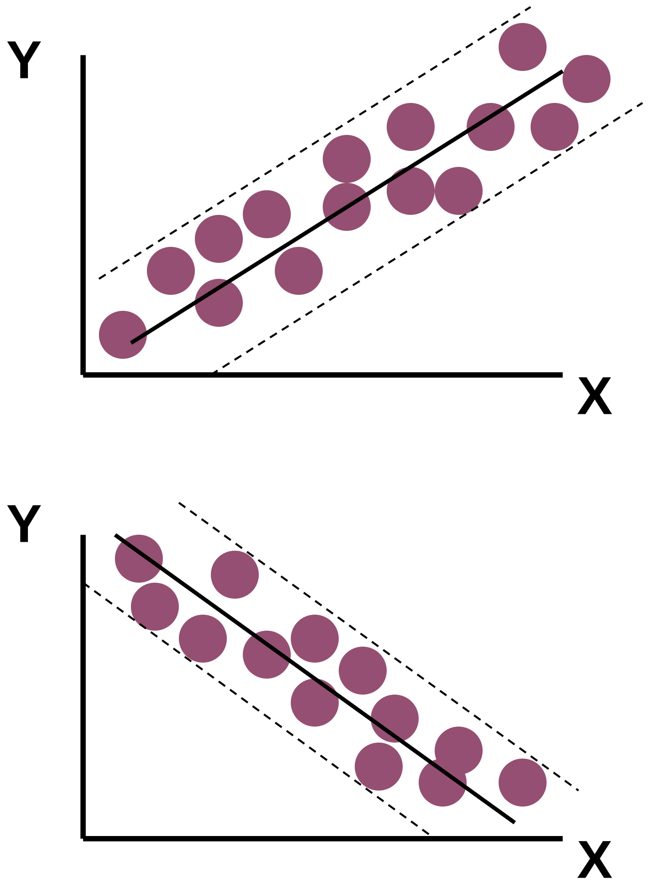


Curvilinear relationships

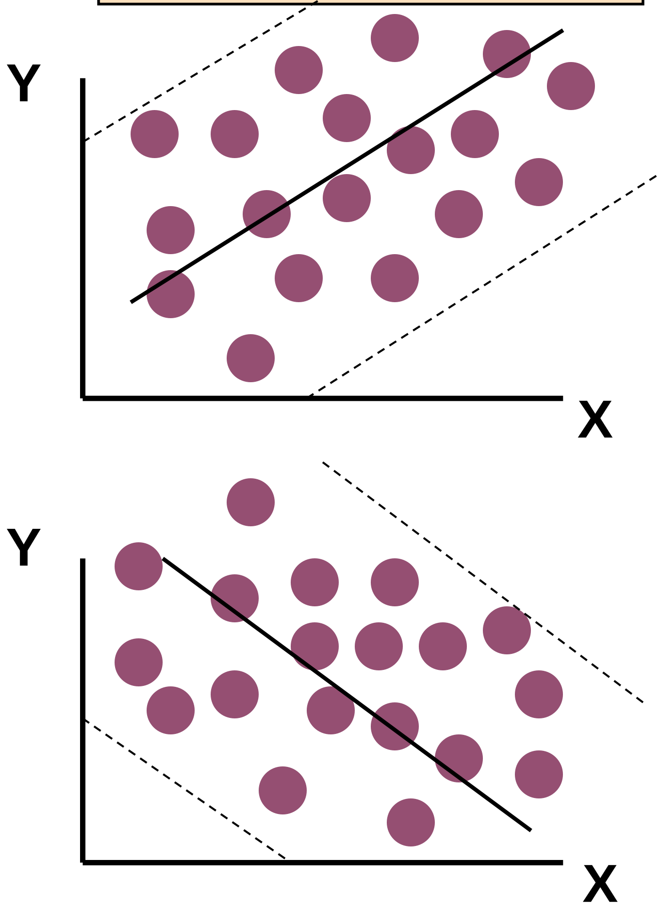


Types of Relationships

Strong relationships

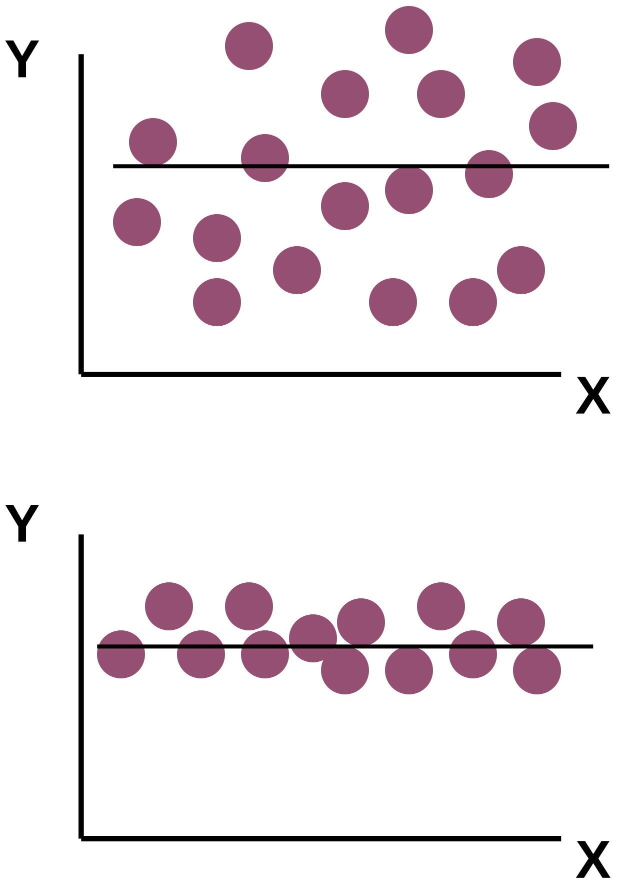


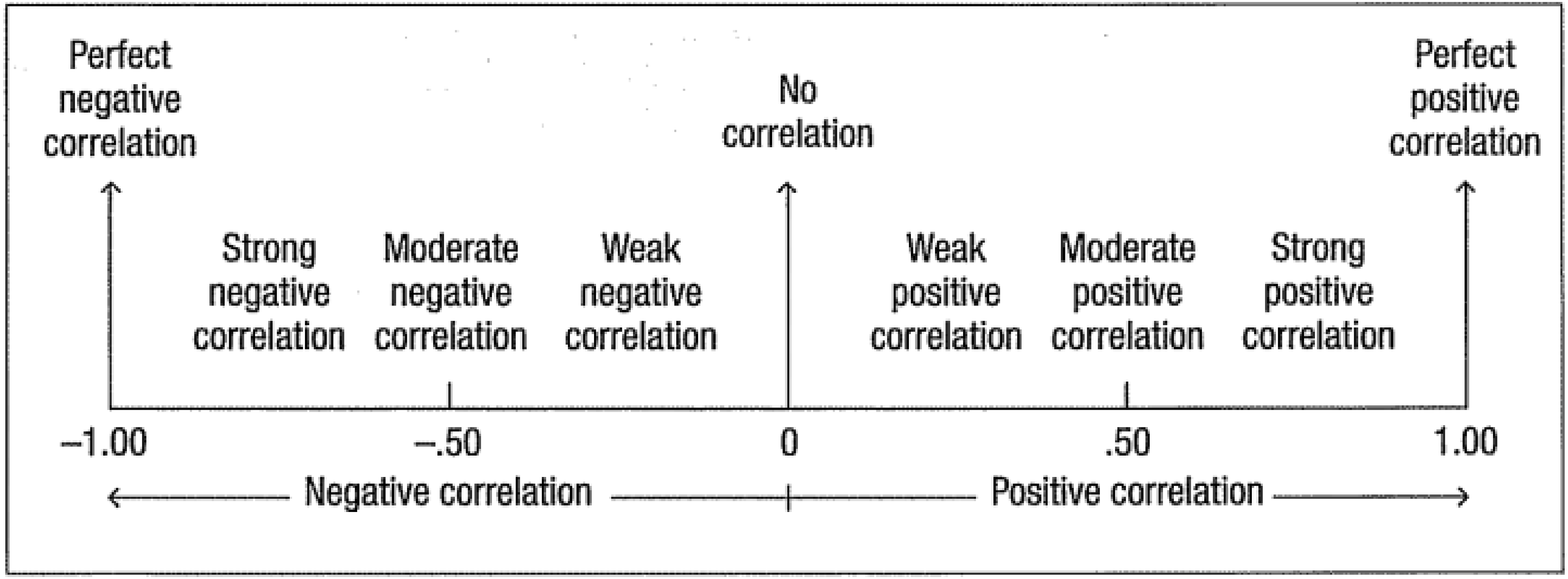
Weak relationships

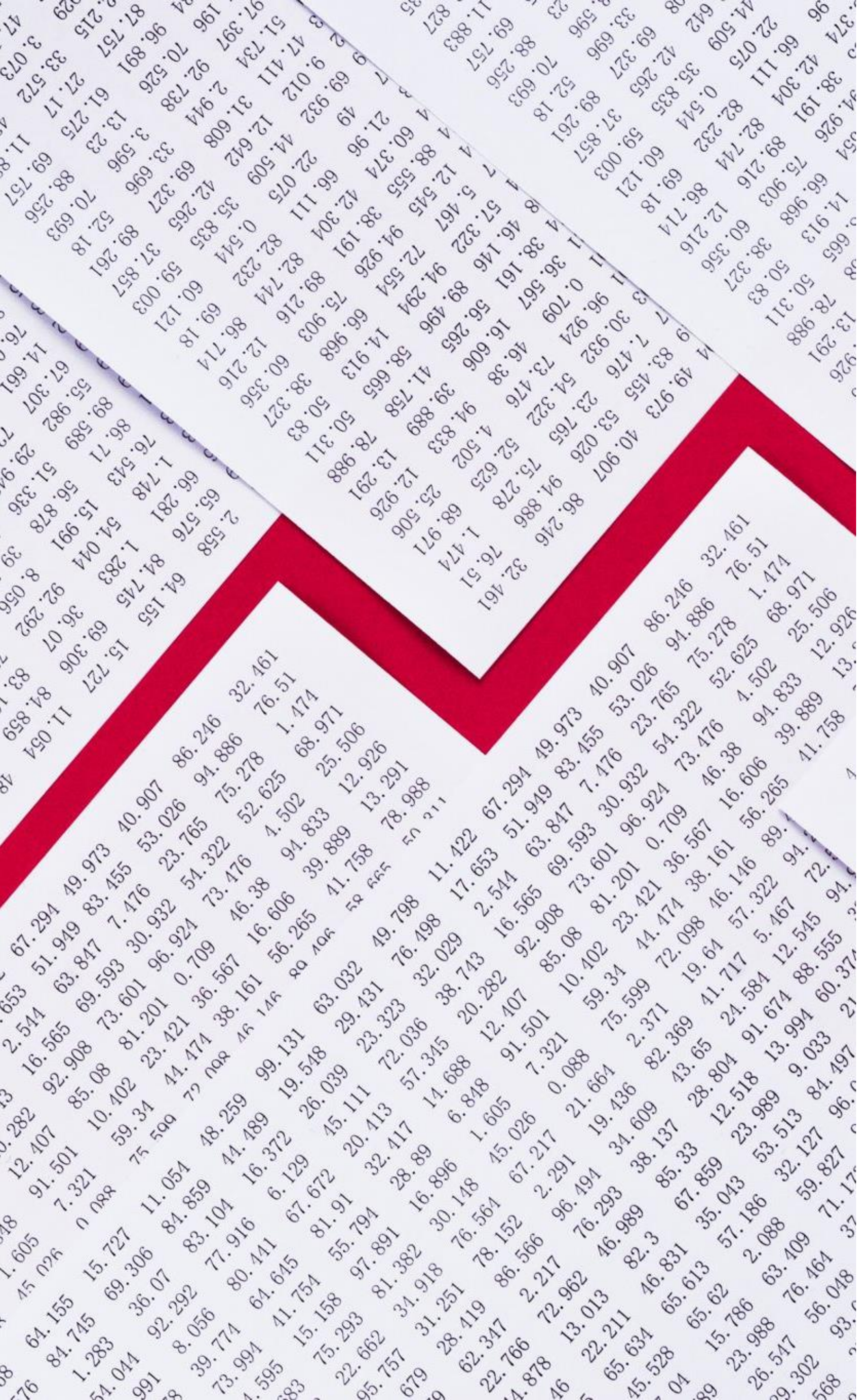


Types of Relationships

No relationship

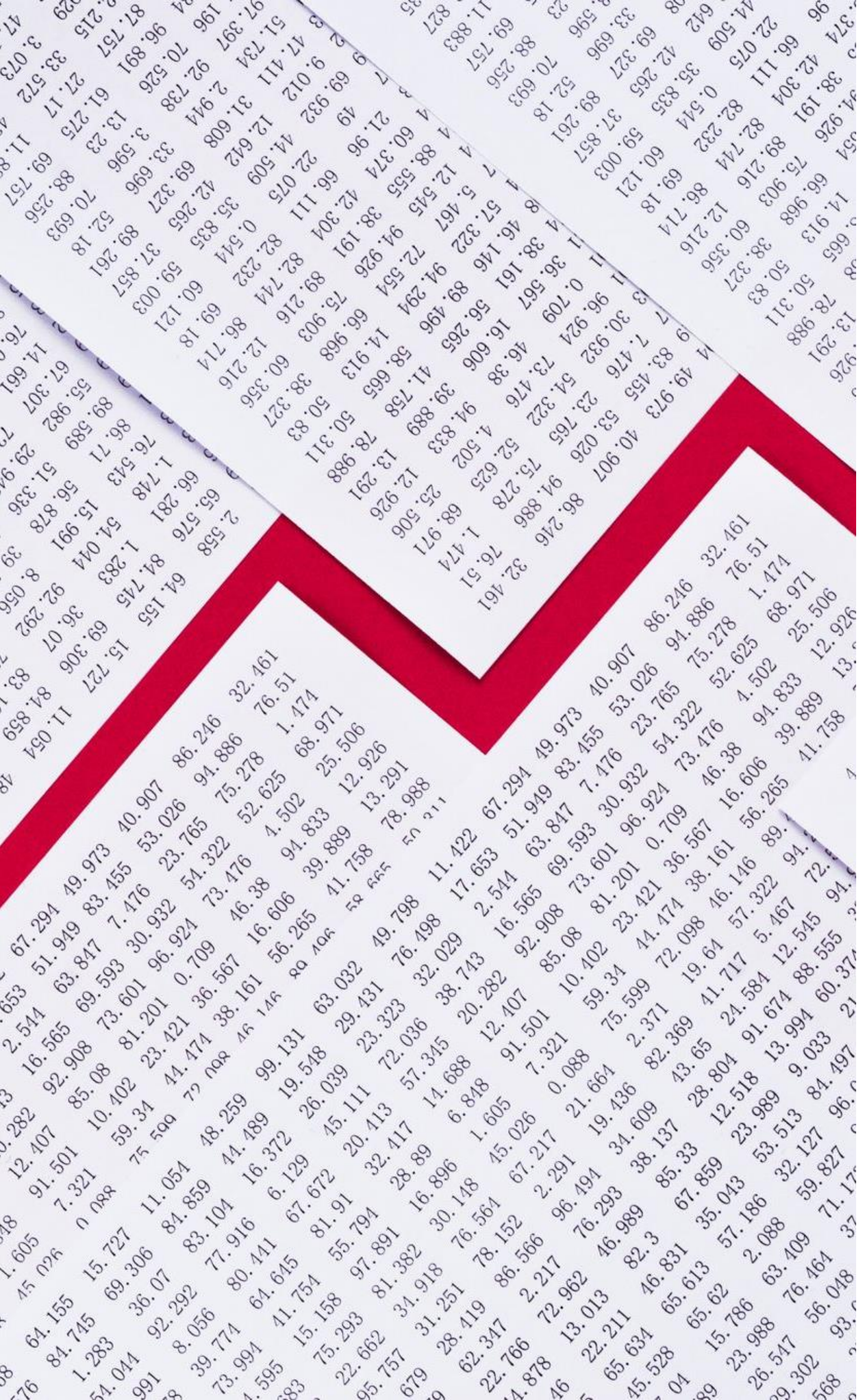



Perfect negative correlation
No correlation
Perfect positive correlation
Strong negative correlation
Moderate negative correlation
Weak negative correlation
Weak positive correlation
Moderate positive correlation
Strong positive correlation
–1.00
–.50
0
.50
1.00
Negative correlation
Positive correlation



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

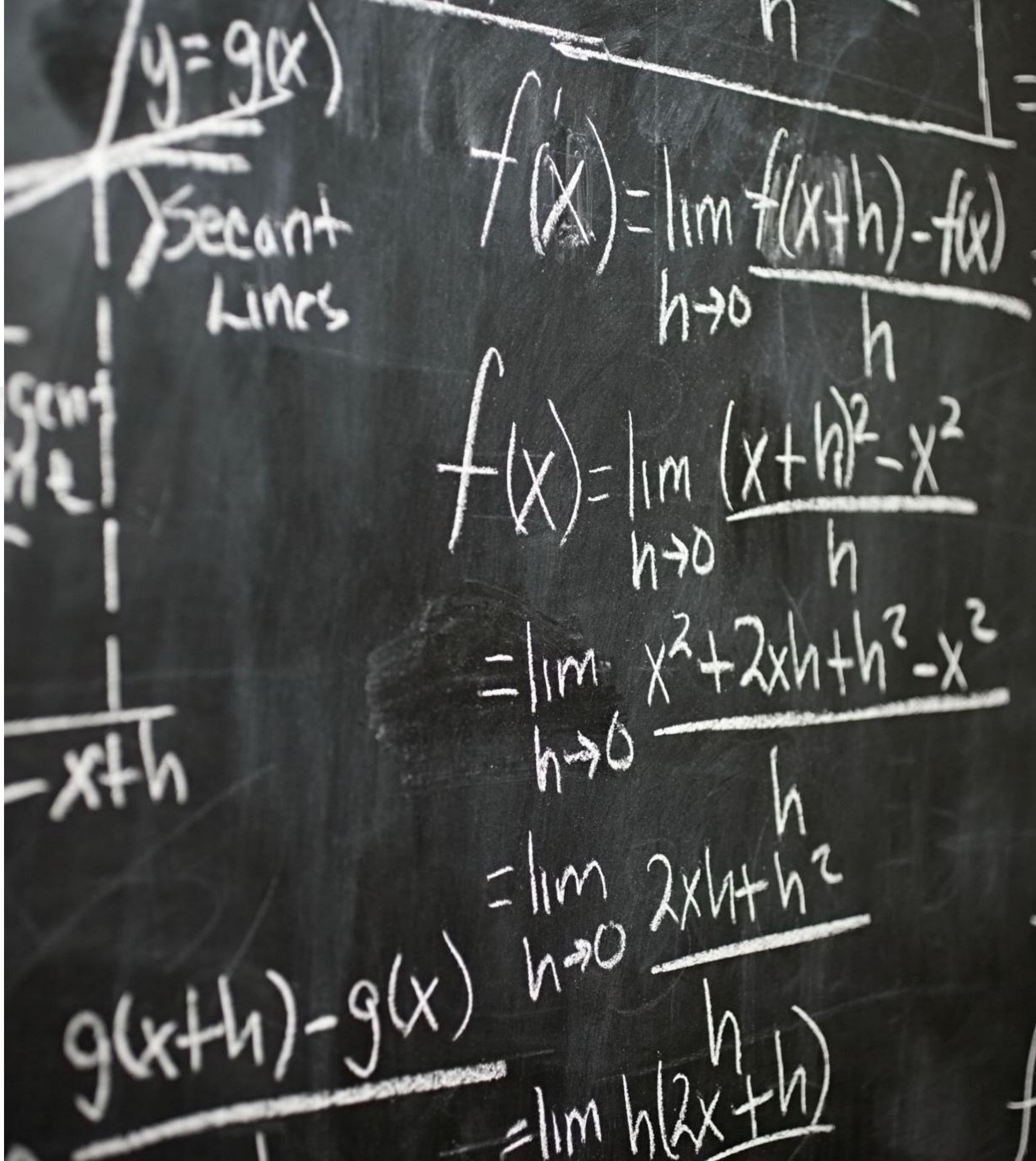
- Regression analysis helps predict or estimate the value of the dependent variable based on the independent variable(s).
- Regression analysis examines the relationship between a dependent variable and one or more independent variables.

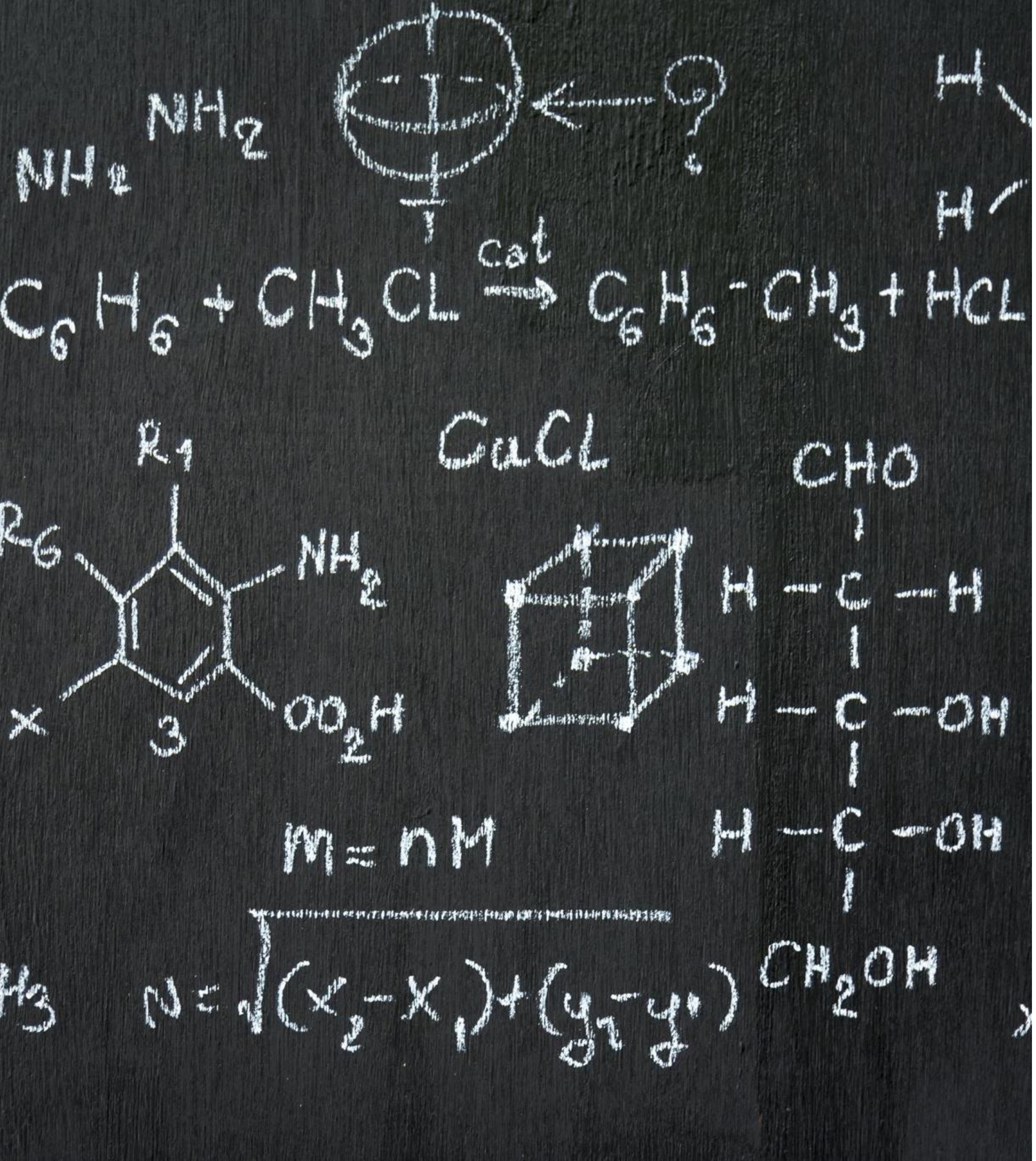
Introduction to Regression Analysis

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





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

Population
Y intercept

Population
Slope
Coefficient

Independent
Variable

Dependent
Variable

$$Y = \underbrace{a + bX}_{\text{Linear component}}$$

Simple Linear Regression Model

Population Y intercept

Population Slope Coefficient

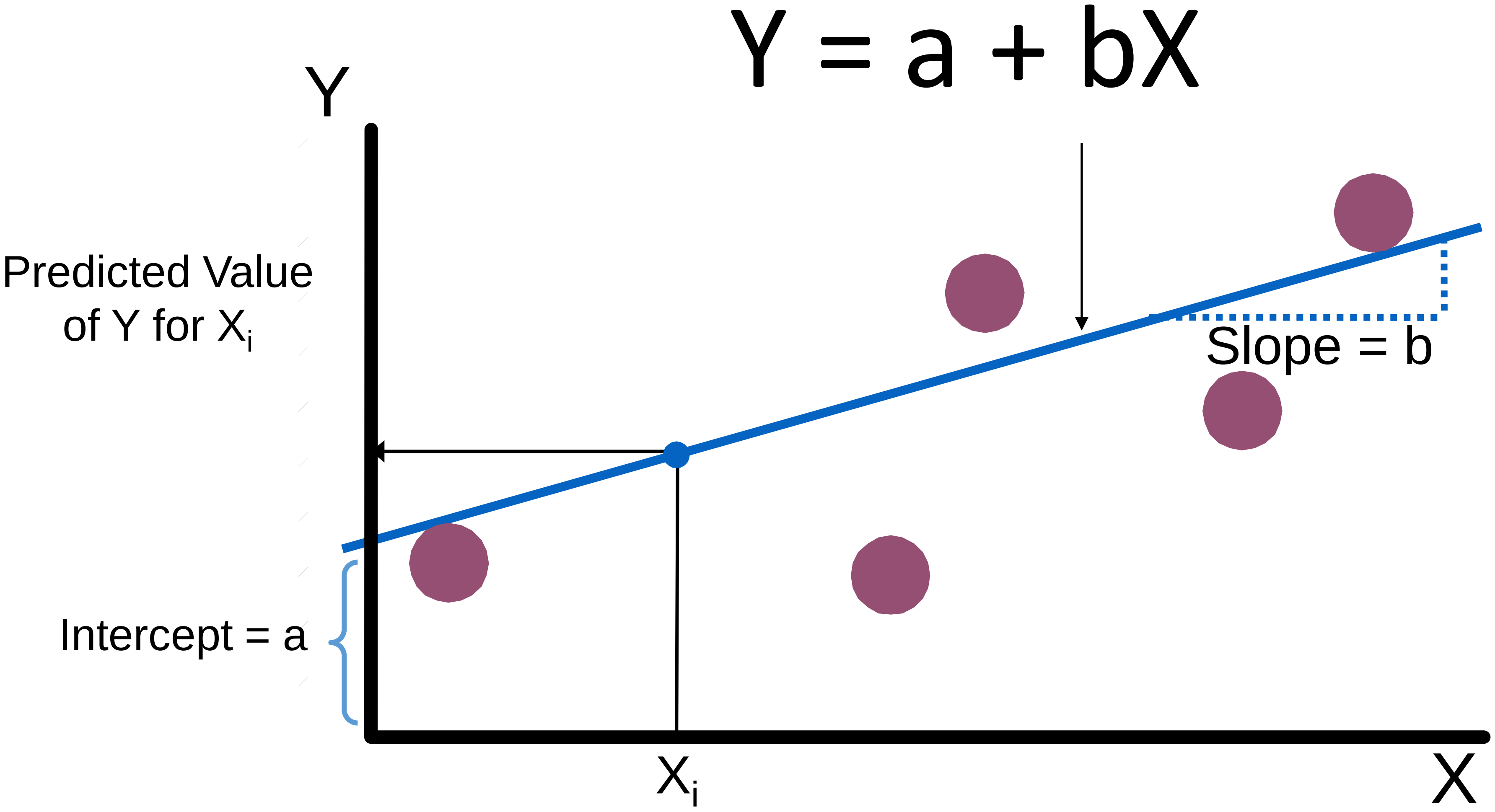
Independent Variable

Random Error term

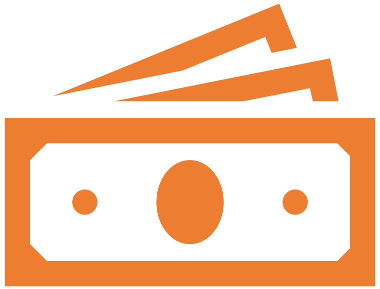
Dependent Variable

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random Error component}}$$

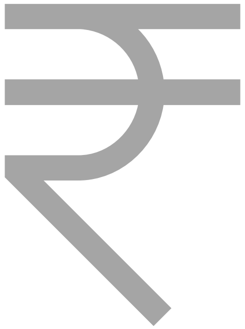
Simple Linear Regression Model



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