**Chap 11: Simple Linear Regression**

**Purpose**

Diagram

Description automatically generated

Comparison: ch9, ch10

Relationship: ch11

**Regression 🡪 Predict**

Example:

chiều cao 🡪 cân nặng

học vấn 🡪 thu nhập

chiều dài lá 🡪 diện tích lá

diện tích nhà 🡪 giá nhà

thu nhập 🡪 chi tiêu

**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.

**Simple linear regression 🡪 Find a linear equation between 2 variables. (Independent variable X and dependent variable Y)**

Example: chiều cao 🡪 cân nặng X: chiều cao Y: cân nặng

học vấn 🡪 thu nhập X: học vấn Y: thu nhập

chiều dài lá 🡪 diện tích lá X: chiều dài lá Y: diện tích lá

diện tích nhà 🡪 giá nhà X: diện tích nhà Y: giá nhà

thu nhập 🡪 chi tiêu X: thu nhập Y: chi tiêu

1. **Simple Linear Regression model:**

**Dependent variable** (Y): the variable we **wish to predict or explain**.

**Independent variable** (X): the variable used **to predict or explain the dependent variable.**

Only one independent variable, X.

Relationship between X and Y is described by a linear function.

a linear function y = ax + b

Changes in Y are assumed to be related to changes in X.

1. **Equation of linear regression**

: independent variable value (giá trị của biến độc lập)

: predicted value (kết quả dự đoán)

(Point estimate for y-intercept - hệ số chặn/hệ số tự do)

(Point estimate for slope -hệ số góc)

**Compute**

1st way hand calculates

2nd way CASIO

3rd Excel

**Compute**

1st way hand calculates

2nd way CASIO

3rd Excel

**Use linear regression to predict future values**

1. , 🡪 **Strength of a linear relationship (Strong/Weak)** & 🡪 **Positive/Negative correlation**

**Sample correlation coefficient**

R and have the same sign.

**Coefficient of determination (R Square)**

Meaning: the portion (%) of the total variation in the dependent variable that is explained by variation in the independent variable.

and both measure the strength of a linear relationship.

,

~ 0 → weak correlation  
 ~ 1 → strong correlation

1. **Compute sums of squares**Diagram

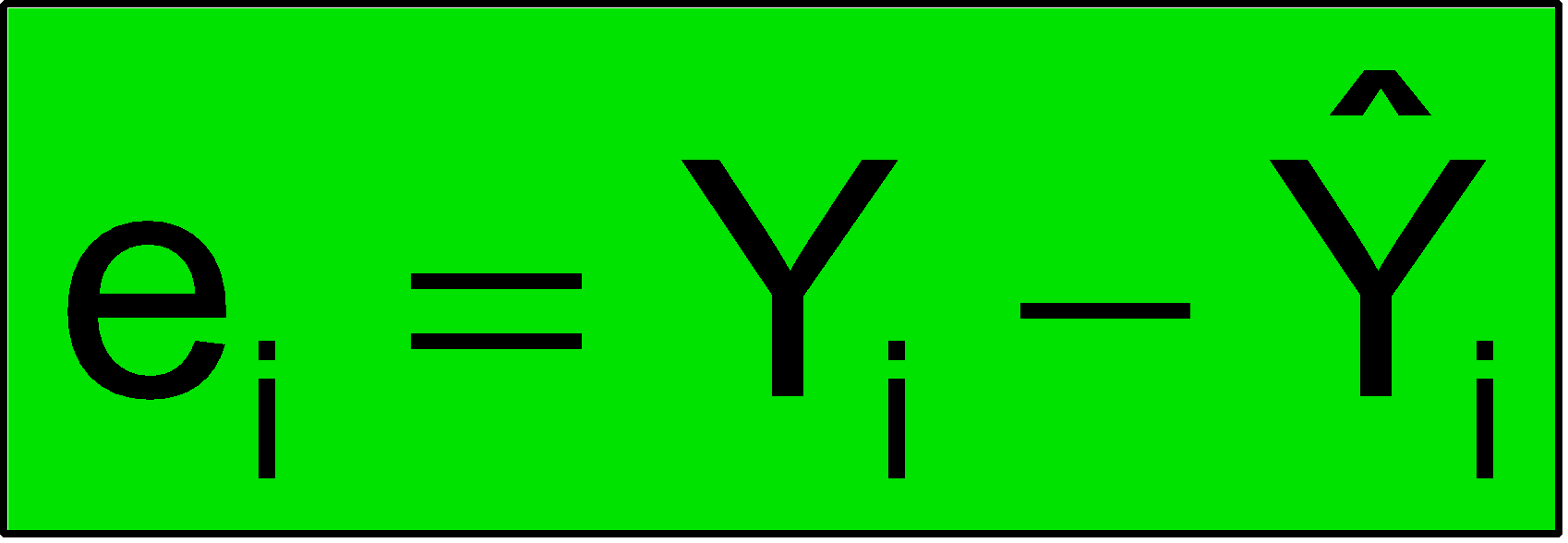
   Description automatically generated with medium confidence

Total sum of squares:

Regression sums of squares:

Error sum of squares:

SST = SSR + SSE

1. **Standard error of estimate**
2. **Residual Analysis** 

Meaning: is the difference between its observed and predicted value.

1. **Hypothesis Testing**

| **Test for the intercept** | **Test for the slope** | **Test for zero correlation** |
| --- | --- | --- |
|  | (No linear relationship)  (Linear relationship does exist) | (No correlation)  (Correlation exists) |
| **Test statistic** | **Test statistic** | **Test statistic** |
| Diagram  Description automatically generated | Diagram  Description automatically generated | Diagram  Description automatically generated |
| If the test statistic is in region of rejection 🡪 **Reject**  If the test statistic is in region of non-rejection 🡪 **Fail to Reject** | | |

**t distribution**

**ch8, ch9: one-sample df=n-1**

**ch10: two-sample df=n1-1+n2-1=n1+n2-2**

**ch11: regression df=n-2**