Properties of Least Square Estimators

The **least squares estimators (LSE)** in multiple linear regression have several important properties, especially under certain assumptions. Here's a concise breakdown:

1. Linearity

- The estimators are **linear functions** of the observed dependent variable y.
- That is, $\hat{\beta}=(X'X)^{-1}X'y$, which is linear in y.

2. Unbiasedness *(under Gauss-Markov assumptions)*

- $E[\hat{\beta}]=\beta$
- This means that, on average, the estimator hits the true parameter values.

3. Minimum Variance (Best Linear Unbiased Estimator - BLUE)

- Among all **linear** and **unbiased** estimators, LSE has the **smallest variance**.
- This is the **Gauss-Markov theorem**.

4. Consistency

- As sample size increases (n→∞n→∞), β^→ββ^→β in probability.
- This holds under mild conditions like no perfect multicollinearity and errors being uncorrelated with predictors.

5. Efficiency (under Normality)

- If the errors are normally distributed, LSE are also **maximum likelihood estimators (MLE)** and are **efficient**, meaning they have the lowest possible variance among all unbiased estimators (not just linear).

6. Sufficiency (under Normality)

- The LSEs are **sufficient statistics** for the regression coefficients if the errors are normally distributed.

7. Asymptotic Normality

- Even if the errors are not normally distributed, under large sample sizes, the distribution of β^β^ approximates normality due to the **Central Limit Theorem**.

In **multiple linear regression**, the variance of the error term (σ2) is **not directly observable**, so we estimate it using the **residuals** from the fitted model.

Estimator of σ2

The **unbiased estimator** of σ^2 is given by:

$$\hat{\sigma}^2 = \frac{RSS}{n-k}$$

Where:

- $RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ is the **residual sum of squares.**
- n = number of observations
- k = number of parameters estimated (including the intercept)
- $\hat{y}_i$ = predicted value from the regression

Matrix Form

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - k}$$

Why divide by $n - k$?

This is the **degrees of freedom** left after estimating k parameters.

It corrects for the loss of degrees of freedom due to estimation, making $\hat{\sigma}^2$ **unbiased**.

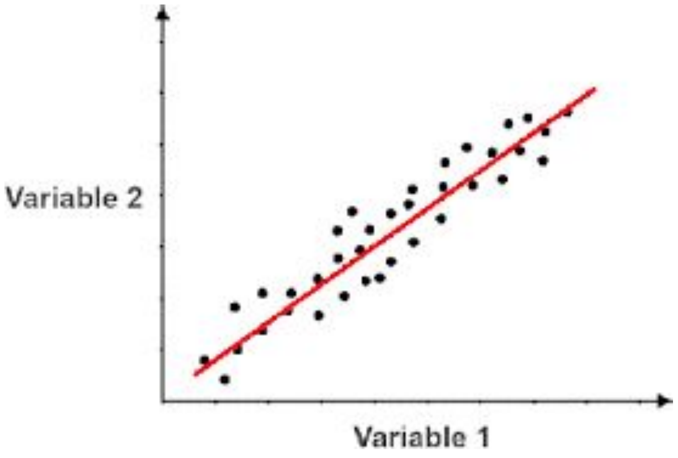
Use of $\hat{\sigma}^2$:

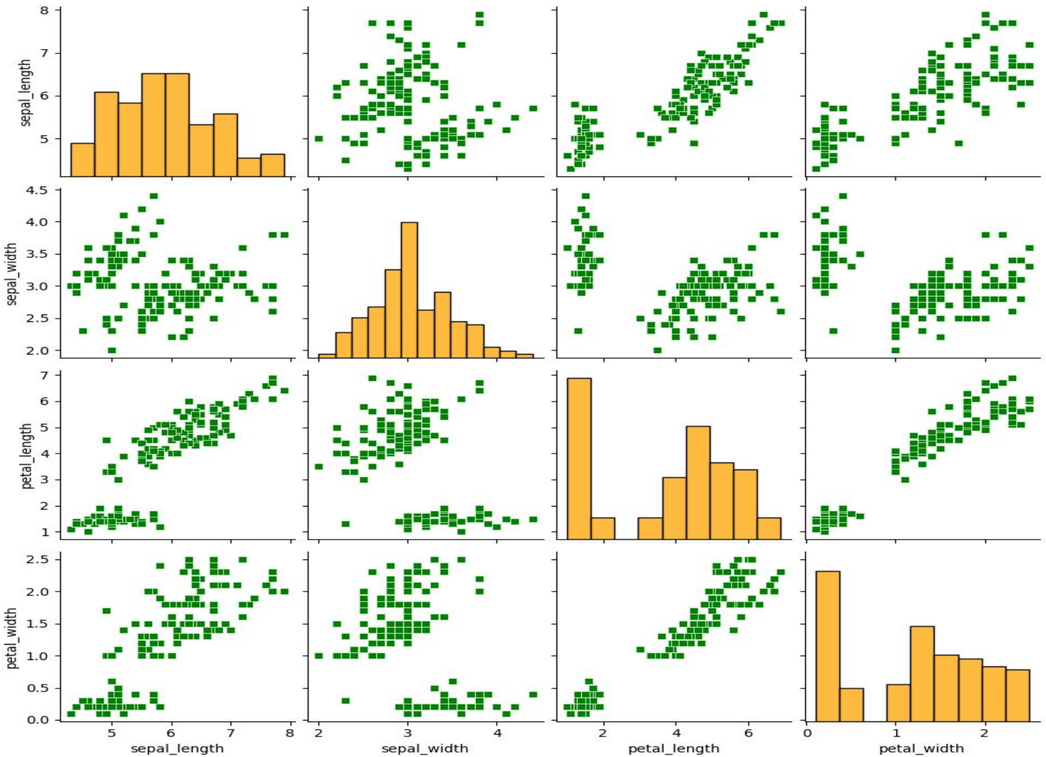
Estimating **standard errors** of regression coefficients

Constructing **confidence intervals** and **hypothesis tests**

Calculating R^2 and **Adjusted** R^2

Inadequacy of Scatter Diagrams in Multiple Regression:





Maximum Likelihood Estimation (MLE):

Maximum Likelihood Estimation (MLE) is a statistical technique used to estimate the parameters of a model by finding the values that make the observed data most probable. It works by calculating the likelihood, which measures how well the model explains the data for a given set of parameters. MLE seeks the parameter values that maximize this likelihood, effectively identifying the most plausible estimates based on the data. For example, if flipping a coin results in 7 heads out of 10 flips, MLE would estimate the probability of heads as 0.7, as this value best explains the observed outcome.

Parameter	**MLE Estimate**
β	$\hat{\beta} = (X^T X)^{-1} X^T Y$
σ^2	$\hat{\sigma}^2 = \frac{1}{n}(Y - X\hat{\beta})^T (Y - X\hat{\beta})$