Review

1. Notations

a. Parameters(unknown)

σ	Standard deviation
β_1, β_2	Coefficients

b. Varibles

$$Y\ vs.\ X_1, X_2, \cdots, X_k$$

Y	Response variable, Outcome variable, Dependent variable
X_k	Explanatory variable, Independent variable, Covariate, Predictor

c. Model

$$Y_i \quad vs. \quad X_{i1}, X_{i2}, X_{i3}$$

$$(Wage)\ vs.\ (degree), (gender), (age)$$

Data set:

$Subject$	Y	X_1	X_2	X_3
1	Y_1	X_{11}	X_{12}	X_{13}
2	Y_2	X_{21}	X_{22}	X_{23}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	Y_n	X_{n1}	X_{n2}	X_{n3}

d. Vector

$\underset{\sim}{Y}(\underset{\sim}{y})$	Vector of Y's(y's)
$\underset{\sim}{\beta}$	Vecror of coefficients

e. Matrix

$\mathbf{X}$	Design matrix
$\mathbf{W}$	Weight matrix

2. Distribution

$Y_i \sim \cdots$	Y_i followes/has $\cdots$ distribution
i.i.d.	Identically and independly distributed

a. Normal Distribution

$$Y \sim N(\mu, \sigma^2)$$

Once Y follows Normal distribution, let $Z = \frac{Y-\mu}{\sigma}$, then:

$$Z \sim N(0,1)$$

Interpretation of Z-score: The distance, in terms of standard deviation, that the variable is deviated from the center of the distribution.

b. χ^2-Distribution

Let $Z_1, \cdots, Z_n \overset{i.i.d}{\sim} N(0,1)$, then:

$$\sum_{i=1}^{n} Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n)$$

c. F-distribution

If:

$$\begin{cases} Y_1 \sim \chi^2(df_1) \\ Y_2 \sim \chi^2(df_2) \\ Y_1, Y_2 \ are \ independent \end{cases}$$

Then:

$$F = \frac{Y_1/df_1}{Y_2/df_2} \sim F(df_1, df_2)$$

d. T-distritbution

If:

$$\begin{cases} Z \sim N(0,1) \\ Y \sim \chi^2(df) \\ Z, Y \ are \ independent \end{cases}$$

Then:

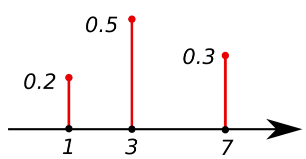
$$T = \frac{Z}{Y/df} \sim t(df)$$

3. Probability Functions

a. PMF(Probability Mass Function)

Example:

$$p_x(x) = \begin{cases} 0.2 & if \ x = 1 \\ 0.5 & if \ x = 3 \\ 0.3 & if \ x = 7 \end{cases}$$

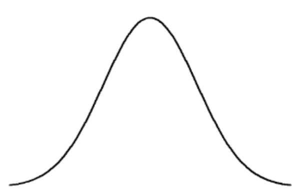


b. PDF(Probability Density Function)

Example:

$$Y \sim N(\mu_i, \sigma^2)$$

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{\sigma^2}}$$



c. Likelihood Function

Treat Y as constant/observed values, allow the parameters to vary.
Example:

$$Y_1, Y_2, \cdots, Y_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$$

$$L(\mu; \underset{\sim}{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{\sigma^2}}$$

As μ changes, $L(\mu)$ will change.

MLE: Find $\hat{\theta}$ to maximize $L(\theta)$. (In the example, find $\hat{\mu}$ to maximize $L(\mu)$)

$$\begin{cases} \text{Maximum Likelihood Estimation} & \rightarrow & Method \\ \text{Maximum Likelihood Estimator} & \rightarrow & \text{Function } (\hat{\mu} = \bar{x}) \\ \text{Maximum Likelihood Estimate} & \rightarrow & \text{Plug in the value from the data set} \end{cases}$$

4. Procedures in normal regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

$$\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$$

Estimate $\underset{\sim}{\hat{\beta}}(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ using OLS (Ordinary Least Squares) = MLE.

a. Inference

- CI for $\hat{\beta}_1$: $\hat{\beta}_1 \pm t_{(df=n-p,\ \alpha/2)} \cdot se(\hat{\beta}_1)$
- In general: (point est.) $\pm$ (critical value)$\cdot$(standard error of the point est.)
- Meaning of confidence level(95%): Repeat sampling process, each time get n independent sample, and every time $\hat{\beta}_1$ differs. In a run, 95% of the time , the interval would cover the true value.

b. Test hypothesis

$$\begin{aligned} \text{p-value} &\neq P(H_0) \\ &= P(\text{test-statistic} | H_0 \text{ is true}) \end{aligned}$$

- If p-value $< \alpha$

  reject H_0 $\begin{cases} Correct \\ TypeIerror \end{cases}$

  P(Type I error)$< \alpha$
- If p-value $> \alpha$

  do not reject H_0 $\begin{cases} Correct \\ TypeIerror \end{cases}$

  Power = 1 - P(Type II error)

i. One predictor (T-test)

$\begin{cases} H_0 : \beta_1 = 0 \\ H_A : \beta_1 \neq 0 \end{cases}$

- Interpretation of H_0: X_1 is not associated with y, after controling/ adjusting other parameters.
- Test statistics: $t = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)}$
- p-value $= 2 \cdot P(t_{(df=n-p)} > |t|)$

  Reject H_0 if p-value ¡ α

ii. Multiple predictors (F-test)

$\begin{cases} H_0 : \beta_1 = \beta_2 = 0 \\ H_A : ALOI0 \end{cases}$

- Global F-test if X_1 and X_2 are all the predictors in the model.
- Partial F-test (special case of global F-test) if X_1 and X_2 are part of the predictors in the model.

$$\begin{array}{lll} & \text{Full model} & \text{SSE(F)} \\ H_0 \longrightarrow & \Big\downarrow & \\ & \text{Reduced model} & \text{SSE(R)} \end{array}$$

c. Interpretation $\begin{cases} \hat{\beta} \\ R^2 \end{cases}$

d. Predictions $\begin{cases} \hat{y} \\ CI for \hat{y} at (X_1, X_2) PI for individual y at (X_1, X_2) \end{cases}$

e. Diagnostics $\begin{cases} \text{linear relationship} \\ \text{individual error} \\ \text{normally distributed error} \\ \text{constant variance} \end{cases}$