STAT_1302; Lecture 21 ; April

Ch. 14 Multiple Linear Regression Model

An extension of the Simple Linear Regression (SLR):

$$Y = A + B_1 X_1 + B_2 X_2 + \cdots + B_p X_p + \mathcal{E}$$

where $\mathcal{E} \overset{ind}{\sim} N(0, \sigma^2)$.

- Must have have $p < n$ where

  p = # of explanatory variables and

  n = # of observations.

Facts:

1. Estimation of $A, B_1, \ldots, B_p$:

   Least Squares Estimation: Use Stat. Software; See STAT_3103.

2. Prediction:

For a given set of $x_1, \ldots, x_p$ values, can predict the response variable, y, using the estimated regression line

$$\hat{Y} = a + b_1 x_1 + b_2 x_2 + \dots + b_p x_p.$$

3. Interpretation of $A, B_1, \dots, B_p$:

(Similar to SLR model)

A = Estimated mean of Y ($\mu_{Y|X_1,\dots,x_p}$) when $x_1 = x_2 = \dots = x_p = 0$.

B_i = Change in the mean of Y ($\mu_{Y|X_1,\dots,x_p}$) when x_i is increased by one unit adjusting for the other x's in the model.

MCQ Ex. An estimated multiple linear regression model is given by

$$\hat{Y} = 15.07 + 0.17 x_1 - 0.13 x_2$$

Interpret the estimated regression Coefficients.

Sol'n:

a) 15.07 is the mean of Y, when $x_1 = x_2 = 0$.

b) 0.17 is the change in the mean of the response ($0.17 > 0$ so an increase in the mean of Y) when x_1 is increased by one unit when x_2 is held fixed. (or say, "adjusting for x_2 in the model").

c) -0.13 is the estimated change in the mean of Y ($-0.13 < 0$, so mean of Y is decreased) when x_2 is increased by one unit accounting for x_1 in the model.

End of Ch. 14.

Back to Testing:

Recall the following Sampling Distributions:

i) Let $X_1, \dots, X_n$ be a random Sample from a population with mean μ and Standard deviation, σ (known). Then,

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

(Exact if the pop. is normal but approximate if pop. not normal and yet $n > 30$ (CLT)).

ii) Let $X_1, \dots, X_n$ be a r.s. from a pop. with mean μ and Standard deviation σ (unknown).

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

iii) For large n,

$$\hat{p} \approx N\left(p, \sqrt{\frac{pq}{n}}\right)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \approx N(0,1) \quad \text{when}$$

$np > 5$ and $nq > 5$.

Hypothesis Testing:

Decision \ Truth	H_0 true	H_A true
Reject H_0	Type I Error	✓
Fail to reject H_0	✓	Type II Error

$$\alpha = P(\text{Type I Error})$$

$$\beta = P(\text{Type II Error})$$

Mnemonic:

$$\alpha = P(\text{Type I Error})$$

A R T → True

R → Reject

A → α

} in terms of H_0

$$\alpha = P(\text{Type I Error})$$

$$= P(\text{Reject } H_0 \mid H_0 \text{ true})$$

$\beta = P(\text{Type II Error})$

$$\left.\begin{array}{l} \text{B A F} \rightarrow \text{False} \\ \downarrow \quad \downarrow \\ \beta \quad \text{Accept} \end{array}\right\} \text{in terms of } H_0$$

$\beta = P(\text{Type II Error})$

$= P(\text{Accept } H_0 \mid H_0 \text{ False})$

$= P(\text{Accept } H_0 \mid H_A \text{ true}).$

$\text{Power} = 1 - \beta.$

$= 1 - P(\text{Accept } H_0 \mid H_A \text{ true})$

$= 1 - P(\text{fail to reject } H_0 \mid H_A \text{ true})$

$= P(\text{reject } H_0 \mid H_A \text{ true}).$

Power Computations are related to sample size determination.

Ex. For a given population with $\sigma = 8$, we want to test the null hypothesis $\mu = 65$

against the alternative $\mu > 65$, on the basis of a sample of size $n = 100$. The null hypothesis is rejected by $\bar{X} \geq 66$.

a) Compute α, the Prob. of Type I Error.

Given $\begin{cases} H_0: \mu = 65 \quad \text{vs.} \quad H_1: \mu > 65 \\ n = 100. \\ \text{Decision rule: } \bar{X} \geq 66 \text{ leads to rejection of } H_0 \end{cases}$

want: $\alpha = P(\text{Type I Error})$ ART

$= P(\text{Reject } H_0 \text{ when } H_0 \text{ true})$

$= P(\text{Reject } H_0 \mid \mu = 65)$

$= P(\bar{X} \geq 66 \mid \mu = 65)$

$= P\left(Z \geq \dfrac{66 - 65}{8/\sqrt{100}}\right)$ $(*)$

$= P(Z \geq 1.25)$

$= 1 - P(Z < 1.25) = 1 - 0.8944 = 0.1056$

$\approx 11\%$

(*) $\bar{X} \approx N\left(65, \frac{8}{\sqrt{100}}\right)$ by the CLT.

b) Find β, the probability of Type II Error when in fact $\mu = 65.50$.

$\beta = P(\text{Type II Error})$

BAF → False

β Accept

$= P(\text{Accept } H_0 \text{ when } H_0 \text{ False})$ $\underbrace{\beta \text{ Accept}}_{\text{in terms of } H_0}$

$= P(\bar{X} < 66 \mid \mu = 65.5)$

$= P\left(Z < \frac{66 - 65.5}{8/\sqrt{100}}\right) =$

$= P(Z < 0.63) = 0.7357 \approx 0.74$

c) What is the power of this test when in fact $\mu = 65.5$?

$\text{Power} = 1 - \beta = 1 - 0.74 = 0.26.$

That is, the likelihood of rejecting H_0 is 26% if the true mean is 65.5.