Polynomial regression

e.g. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + e$

e.g.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + e + \alpha_1 X_2 + \alpha_2 X_2^2 + \alpha_3 X_2^3 + \gamma_1 \underline{X_1 X_2} + \gamma_2 \underline{X_1^2 X_2} \cdots$$

interaction terms

Computation is same as multiple regression

Delta method

Given : $\hat{\theta} \sim N(\theta, \sigma^2 D)$

Want : distribution of $g(\hat{\theta})$, g is some function

Delta method : $g(\hat{\theta}) \sim N(g(\theta), \sigma\, \underline{g'(\theta)} D [\underline{g'(\theta)}]^T)$

proof : use Taylor's expansion

e.g. $\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \sigma^2 (X'X)^{-1}\right) \leftarrow y = \beta_1 X_1 + \beta_2 X_2 + e$

Q: Construct C.I. for $\hat{\beta}_1^2 + \hat{\beta}_2^2 \leftarrow g(\beta_1, \beta_2) = \beta_1^2 + \beta_2^2$

C.I. (95%)

$\hat{\beta}_1^2 + \hat{\beta}_2^2 \pm 1.96 \sqrt{\widehat{Var}}$

From Delta method:

$g' = \begin{pmatrix} 2\beta_1 \\ 2\beta_2 \end{pmatrix}$

$\hat{\beta}_1^2 + \hat{\beta}_2^2 \sim N\left(\beta_1^2 + \beta_2^2, \sigma^2 (2\beta_1 \; 2\beta_2)(X'X)^{-1} \begin{pmatrix} 2\beta_1 \\ 2\beta_2 \end{pmatrix}\right)$

$\hat{\sigma}^2 (2\hat{\beta}_1 \; 2\hat{\beta}_2)(X'X)^{-1} \begin{pmatrix} 2\hat{\beta}_1 \\ 2\hat{\beta}_2 \end{pmatrix}$

6.2 Factors

- Two types of variables
  - Quantitative variables (continuous)
    - Measured in a numeric scale.
    - Has unit
      m kg m/s person
    - e.g. Height, weight, speed, population
  - Qualitative/Categorical variables (discrete)/Factor
    - Divided into levels/categories
    - No unit
    - e.g. Gender, blood type, eye color.

      M/F A/B/O/AB

6.2 Factors

- Simple case: Factor with 2 levels
  - e.g. male/female, pass/fail
  - Use dummy variable (also called indicator)

$$U_i = \begin{cases} 0 -- & \text{the i-th observation is F} \\ 1 -- & \text{the i-th observation is M} \end{cases}$$

  - e.g. $\text{Height}_i = \beta_0 + \beta_1 U_i + e_i$
    - β_1 = the additional effect for one level over another.
      = the mean height difference between male and female

$$\text{Height} = \begin{cases} \beta_0 + e & \text{if } F \\ \beta_0 + \beta_1 + e & \text{if } M \end{cases}$$

6.2 Factors

- Not so simple case: Factor with >2 levels
  - e.g. blood type, eye color, college
  - e.g. GPA = $\beta_0+\beta_1 C+e$

$$C = College = \begin{cases} 0 -- & \text{CC} \\ 1 -- & \text{UC} \\ 2 -- & \text{NA} \\ 3 -- & \text{NC (New Colleges)} \end{cases}$$

- Is it reasonable?

If $\beta_1 > 0$

CC < UC < NA < NC

β_0 $\beta_0+\beta_1$ $\beta_0+2\beta_1$ $\beta_0+3\beta_1$

If $\beta_1 < 0$

CC > UC > NA > NC

6.2 Factors

- Factor with d levels
  - The factor rule:
    - A factor with d levels can be represented by (d-1) dummy variables and the intercept.
  - How to use dummy variable?
  - e.g. GPA = $\beta_0+\beta_1 C_{UC}+ \beta_2 C_{NA} + \beta_3 C_{NC}$ +e

$$C_{UC}=\begin{cases}0--\text{not UC}\\1--\text{UC}\end{cases},\quad C_{NA}=\begin{cases}0--\text{not NA}\\1--\text{NA}\end{cases},\quad C_{NC}=\begin{cases}0--\text{not NC}\\1--\text{NC}\end{cases}$$

- Essentially, we model each group with a separate mean.
- β_0=E(GPA|CC), $\beta_0+\beta_1$=E(GPA|UC), $\beta_0+\beta_2$=E(GPA|NA), $\beta_0+\beta_3$=E(GPA|NC)

Chapter6.pdf - Adobe Reader

File Edit View Window Help

16 / 21 100% Tools Sign Comment

- Factor with d levels
  - e.g. GPA = $\beta_0+\beta_1 C_{UC}+\beta_2 C_{NA}+\beta_3 C_{NC}+e$

$$C_{UC}=\begin{cases}0\text{ -- not UC}\\1\text{ -- UC}\end{cases},\quad C_{NA}=\begin{cases}0\text{ -- not NA}\\1\text{ -- NA}\end{cases},\quad C_{NC}=\begin{cases}0\text{ -- not NC}\\1\text{ -- NC}\end{cases}$$

  - Implementation

Raw data

Student	GPA	College	C_{UC}	C_{NA}	C_{NC}
Bun	3.6	UC	1	0	0
Wing	3.9	WS	0	0	1
KiKi	3.5	NA	0	1	0
Cherry	3.8	UC	1	0	0
Jeremy	3.6	CC	0	0	0
...	...	...	...	...	...

$\Rightarrow$ lm(GPA ~ $C_{UC}+C_{NA}+C_{NC}$)

$\hat{\beta}_0$ $\hat{\beta}_1$ $\hat{\beta}_2$ $\hat{\beta}_3$

  - R-Implementation

College=c("uc",“nc",“na",“uc","cc")

C.uc=as.numeric(College=="uc"); C.na=as.numeric(College=="na"); C.nc=as.numeric(College==“nc")

11.69 x 8.27 in

NEW PAGE TEMPLATES PREV. Page 6 of 11 NEXT PC MODE POWERPOINT BALL PEN HIGH LIGHT SIZE COLOR LINE RECT. CIRCLE SELECT ERASE ALL ERASE REC Recording 00:00:00 SAVE OPEN HELP EXIT

CH 2:11 PM 11/4/2014