STAT6038 Week 10 Lecture Notes

Rui Qiu

2017-05-10

1 Wednesday's Lecture

1.1 Orthogonal Contrasts

Predictor set 1 from the Multicolinearity example

x_1	10	10	10	10	15	15	15	15
x_2	10	10	15	15	10	10	15	15

$n = 2$ for each combinations of (x_1, x_2).

All the cross- classified category (treatment) sample sizes are equal $\implies$ this is a balanced experimental design.

$\implies$ orthogonal contrasts

(x_1, x_2 variables uncorrelated and order in the ANOVA table in a multiple regression will be unimportant)

for more info, see either STAT7030 GLMs or STAT7029 Design of Experiments & Surveys.

1.2 Variance Inflation Factor (for variable X_j)

In the set $X_1, X_2, \ldots, X_k$.

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R^2 for the regression of X_j on $X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$.

2 Thursday's Lecture

Nothing serious, seriously. Nothing.s

3 Friday's Lecture

Indicator Variables

$$\text{svi}_i = \begin{cases} 1 & \text{if "seminal vesicle invasion"} \\ 0 & \text{otherwise} \end{cases}$$

Model

$$\text{lcavol}_i = \beta_0 + \beta_1 \text{svi}_i + \epsilon_i, i = 1, \ldots, n, \epsilon \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

plot.

$\hat{\beta}_1$ = difference in the means = $\text{mean}_{(\text{lcavol—svi=1})} - \text{mean}_{(\text{lcavol—svi=0})}$

Indicator variables (continued) New Model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon; \epsilon \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

where Y = lcavol
$X_1 = \text{svi}_i = \begin{cases} 1 & \text{if svi=Yes for observation i} \\ 0 & \text{otherwise} \end{cases}$
We call X_1 factor or treatment variable.
X_2 = lpsa (a continuous covariate)
$\rightarrow$ "analysis of covariance" model
when svi = 0, $X_1 = 0$
fitted model

$$\hat{Y} = \hat{\beta}_0 + 0 + \hat{\beta}_2 X_2 = \hat{\beta}_0 + \hat{\beta}_1 X_1$$

when svi = 1, $X_1 = 1$

$$\hat{Y} =$$