1st hour: KKM Ch.15: #3, 6;

Other:

1. More additions to web page

2. Ch.11, Comp. Ex. #2 Wed. 10/29

3. PostScript printing at SPH lab

Ch.15: Analysis of Covariance

Popular procedure for regression modelling - abbrev. ANACOVA by KKM

Used to determine effects of study factors on the response variable adjusted for control (extraneous) variables, called *covariates*.

Study factors treated as nominal variables (via dummy variables), covariates can be on any scale.

Assumption in ANACOVA: no interaction of covariates and study (dummy) variables

ANACOVA model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + e$$

where X is the *covariate* and Z is the dummy variable, e.g. $X =$ age, $Z =$ sex (0 male/1 fem.)

The models for males and females, respectively, are:

$$Y_m = \beta_0 + \beta_1 X + e$$
$$Y_f = (\beta_0 + \beta_2) + \beta_1 X + e$$

How is this model different from "inclusive" dummy variable model in Ch.14?

Under this model, *adjusted mean scores* for males and females are defined to be the predicted values obtained by evaluating the model at $Z = 0$ and $Z = 1$ when X is set equal to the overall mean age for the two groups.

A partial F test of H_0: $\beta_2 = 0$ is used to determine whether these adjusted mean scores are significantly different

Potential drawback of ANACOVA: assumption of parallelism (i.e. $\beta_{1,m} = \beta_{1,f}$ assumed)

Should deal with this before using ANACOVA by testing H_0: $\beta_3 = 0$ in the inclusive dummy-variable model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + e$$

using a partial F test as in Ch.14

What if we reject H_0: $\beta_3 = 0$, i.e. lines are not parallel? Read KKM sec.15-4 p.302 (2nd paragraph)

In general, can use ANACOVA for multiple groups and covariates:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \beta_{p+1} Z_1 + \beta_{p+2} Z_2$$

$$+ \cdots + \beta_{p+k-1} Z_{k-1} + e$$

KKM Ch.15 #3:

8 rats in study, 2 diets, 4 rats given one diet, 4 the other diet. Want to compare the two diets (study variable Z), adjusting for the covariate initial weight (X), with respect to the absorption of a chemical (response Y)

a) State the ANACOVA regression model for comparing the two diets (set $Z = -1$ if diet I is used and $Z = 1$ if diet II is used).

Model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + e,$$

$$\text{where } Z = \begin{cases} -1, & \text{if diet I used} \\ +1, & \text{if diet II used} \end{cases}$$

b) Use the results in the table to determine the adjusted mean responses for each diet, controlling for each weight

The models for diet I and II, respectively, are:

$$\text{Diet I } (Z = -1): \quad \widehat{Y}_{\text{I}} = (\widehat{\beta}_0 - \widehat{\beta}_2) + \widehat{\beta}_1 X$$
$$\text{Diet II } (Z = +1): \quad \widehat{Y}_{\text{II}} = (\widehat{\beta}_0 + \widehat{\beta}_2) + \widehat{\beta}_1 X$$

From the given table, we have

$$\widehat{Y}_{\text{I}} = (13.0 - 0.5) + 0.5X = 12.5 + 0.5X$$
$$\widehat{Y}_{\text{II}} = (13.0 + 0.5) + 0.5X = 13.5 + 0.5X$$

so $=$

$$\overline{Y}_{\text{I}} \text{ (adj.)} = 12.5 + 0.5\overline{X} = 12.5 + 0.5(3) = 14$$
$$\overline{Y}_{\text{II}} \text{ (adj.)} = 13.5 + 0.5\overline{X} = 13.5 + 0.5(3) = 15$$

c) Test whether the two diets are significantly different usign the ANACOVA model defined in (a).

We perform a partial F test of H_0: $\beta_2 = 0$

$$\begin{aligned} F(Z|X) &= \frac{\{\text{SSR}(X,Z) - \text{SSR}(X)\}/1}{\{\text{SSE}(X,Z)\}/5} \\ &= \frac{2.0/1}{1.0/5} \\ &= 10 \end{aligned}$$

with 1 and 5 d.f. The p-value is $\approx .025$ so we reject H_0 at the $\alpha = .05$ level.

KKM Ch.15 #6:

Drugs A, B, C, D (Z_1, Z_2, Z_3); covariate initial total body weight (X); weight (W)

Model:

$$W = \beta_0 + \beta_1 X + \beta_2 Z_1 + \beta_3 Z_2 + \beta_4 Z_3 + e$$

Fitted ANACOVA model:

$$\widehat{W} = 0.1728 + 0.0032X - 0.3917Z_1 - 0.2257Z_2 - 0.1351Z_3$$

Adjusted means:

$$\overline{X} = 226.125,$$

$$\begin{aligned}
\overline{W}_{\text{A}}(\text{adj.}) &= 0.1728 + 0.0032(226.125) - 0.3917 \\
&= .5047 \\
\overline{W}_{\text{B}}(\text{adj.}) &= 0.1728 + 0.0032(226.125) - 0.2257 \\
&= .6707 \\
\overline{W}_{\text{C}}(\text{adj.}) &= 0.1728 + 0.0032(226.125) - 0.1351 \\
&= .7613 \\
\overline{W}_{\text{D}}(\text{adj.}) &= 0.1728 + 0.0032(226.125) \\
&= .8964
\end{aligned}$$

To determine whether these are significantly different, we test H_0: $\beta_2 = \beta_3 = \beta_4 = 0$ using the partial F test:

$$
\begin{aligned}
F &= \frac{\{\text{SSR}(X, Z_1, Z_2, Z_3) - \text{SSR}(X)\}/3}{\text{SSE}(X, Z_1, Z_2, Z_3)/11} \\
&= \frac{\{.6185 - .3202\}/3}{.1587/11} \\
&= 6.89, \text{ (3,11) df., } p\text{-value} < .05
\end{aligned}
$$