13 Analysis of covariance and other methods for adjusting continuous data

Applied regression analysis and other multivariate methods

Yi Zhou

2016-03-29

Preview

Three reasons for controlling for the covariates

- To access interaction
- To correct for confounding
- To increase the precision

The usual approach is to fit a model containing both exposure variables and control variables (and even the product terms).

Analysis of covariance (ANACOVA) is a multiple regresion model, in which

- the study factors of interest are all treated as nomial variables
- the covariates are measured on any scale
- the dependent variable is considered as continuous

A different perspective:

- the `nominal` covariates are now the main predictors
- the continous variables are included for control purpose
- no interaction of covariates

Adjustment problems

Example on pp.309 emanating from two questions:

- Q1: whether the relationship `Y~X` is the same for male and female
- Q2: whether `Y` for males differ from that for female controlling for `X`

Considering

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

where X is age, Y is blood pressure, and Z is the sex group.
For Q1,

- the lines are coincident, $\beta_2 = \beta_3 = 0$
- the lines are parellel but not coincident, $\beta_3 = 0$, but $\beta_2 \neq 0$
- the lines are not parallel, $\beta_3 \neq 0$

For Q2,

- adjusted mean difference and adjusted means for each group
- proper statistical test

ANACOVA

Considering

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

this model assumes that the regression lines for male and female are **parallel**.
the `adjusted means` are the the predicted values

- at $Z = 0$ and $Z = 1$
- X is the `overall mean age`

A `partial` F test is used to test $H_0 : \beta_2 = 0$ to determine the differences of between the adjusted means.

Assumption of parallel: a potential drawback

If the `parallelism` assumption is invalid ($\beta_3 \neq 0$)

- the covariance method of adjustment should be avoided.
- no adjustment should be made

To put it another way, if parallelism is found, adjusted means should be considered.

The extensions

Several groups and several covariates

Considering

$$\bar{Y}_j(adj) = (\hat{\beta}_0 + \hat{\beta}_{q+j}) + \hat{\beta}_1\bar{X}_1 + \cdots + \hat{\beta}_q\bar{X}_q \quad j = 1, \ldots, s-1$$
$$\bar{Y}_s(adj) = (\hat{\beta}_0) + \hat{\beta}_1\bar{X}_1 + \cdots + \hat{\beta}_q\bar{X}_q$$

$$H_0 : \beta_{q+1} = \beta_{q+2} = \cdots = \beta_{q+s-1} = 0$$

using $F_{s-1,n-q-s}$ test.

Several nominal independent variables

Considering the model under different combinations of categories to adjust the means.

Comments and cautions

- ANACOVA adjust the disparity in covariate distribution over groups by assuming **a common distribution**, not just a common set of means.
- The use of ANACOVA-adjusted mean is appropriate only when interaction effects are not present.
  - See eq.13.10 for a chunk test for parallelism
- Validity (adjusting for confounding) and precision (SE or CI)
- Predicted values based on the best regression model can be regarded as adjusted values
- If all variables are categorical, we can treat it as a two-way ANOVA with unequal cell numbers (see `Chapter 20`)

Summary

- ANACOVA is used to control for covariates
- The study variables of interest must be treated as categorical
- No interaction between covariates and study variables
- The interaction can be checked using the extended ANACOVA model
- If interaction is significant, no adjustment on the means