The mysterious beauty of the analysis of covariance[1]

STA305 Winter 2014

[1] See last slide for copyright information.

Background Reading

Optional

- Chapter 5 in *Data analysis with SAS* presents some important parts of this material as a special case of regression.

Basic idea

- Lots of things influence the response other than the treatment.
- Because of random assignment, they are independent of the treatment.
- They all go into the error (background noise) term ϵ_{ij}.
- $\sigma^2 = Var(\epsilon_{ij})$ is the loudness of the background noise.
- Reduce loudness of background noise by measuring important influences and including them in the model.
- Make sure that the treatment is not influencing the covariate.

It's just another regression model

The $d_{i,j}$ are dummy variables for the treatments

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 d_{i,1} + \cdots + \beta_{p-1} d_{i,p-1} + \epsilon_i \\
&= \beta_0' + \beta_1 d_{i,1} + \cdots + \beta_{p-1} d_{i,p-1} + (\alpha_1 X_{i1} + \cdots + \alpha_k X_{ik} + e_i) \\
&= \mathbf{X}_i' \boldsymbol{\alpha} + \mathbf{d}_i' \boldsymbol{\beta} + e_i
\end{aligned}
$$

- $Var(e_i) < Var(\epsilon_i)$.
- The $X_{i,j}$ are called *covariates*.
- They are random variables, but treat them as fixed.
- This is the usual conditional regression model.
- The assumption of unit-treatment additivity implies parallel regression planes.

Technical issues with the model $Y_i = \mathbf{x}_i'\boldsymbol{\alpha} + \mathbf{d}_i'\boldsymbol{\beta} + e_i$

- Assume this model is conditional on $\mathbf{X}_i = \mathbf{x}_i$.
- Error terms e_i are identically distributed given $\mathbf{X}_i = \mathbf{x}_i$.
- So the model assumes e_i and $\mathbf{X}_i$ are independent.
- Thus any other omitted variables that influence Y_i must be *independent of the covariates*.
- Impossible to believe, and a well-known recipe for trouble.
- Also, covariates are surely measured with error, another recipe for trouble.

Does it still work?

A simple example

The true model (e_i is different now)

- Binary dummy variable for experimental treatment.
- One covariate measured with error.
- One omitted variable, correlated with the (true) covariate.

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 d_i + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \epsilon_i \\
W_i &= \lambda_0 + \lambda_1 X_{i1} + e_i
\end{aligned}
$$

- Observe (d_i, W_i, Y_i).
- Fit $Y_i = \beta_0^* + \beta_1^* d_i + \beta_2^* w_i + \delta_i$
- Interest is in $\beta_1 = \Delta$.

A simulation study

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 d_i + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \epsilon_i \\
W_i &= \lambda_0 + \lambda_1 X_{i1} + e_i
\end{aligned}
$$

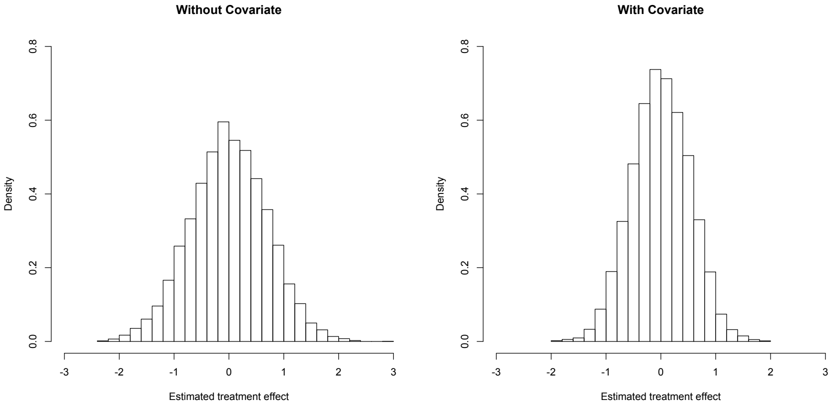
- X_1 and X_2 are both strongly related to Y.
- X_1 and X_2 are strongly correlated.
- Lots of measurement error.
- $n_1 = n_2 = 64$
- Fit $Y_i = \beta_0^* + \beta_1^* d_i + \beta_2^* w_i + \delta_i$
- Test $H_0 : \beta_1^* = 0$ ten thousand times when $\beta_1 = 0$ is true, and there is no treatment effect.

No inflation of Type I error probability

- Did it both ways, with and without the (corrupted) covariate W_i.
- Without covariate: $p \approx 0.0464$
- With covariate: $p \approx 0.0537$
- These are typical results.

Sampling distribution of $\widehat{\Delta}$

Based on ten thousand simulated data sets



$Var(\widehat{\Delta})$ is smaller with the covariate

- Without covariate, exactly $\sigma^{2\prime}\left(\frac{1}{n_1}+\frac{1}{n_2}\right) = 0.48125$
- With covariate, approximately 0.2769367 based on the sample variance of 10,000 estimates.
- Had $n_1 = n_2 = 64$. Keeping equal sample sizes, what sample size is needed to achieve this precision without the covariate?

$$\begin{aligned} & 15.4\left(\frac{1}{n_1}+\frac{1}{n_1}\right) = 0.2769367 \\ \Leftrightarrow \quad & n_1 = 111.2 \end{aligned}$$

- Need about 111+111=222 experimental units to get the same precision without the covariate.
- The covariate is worth about 222-128=94 experimental units.
- An estimator with lower variance is said to be more *efficient*.

Why does the analysis of covariance work so well?

When the model is so wrong

After a lot of work,

$$\begin{aligned}
\widehat{\Delta} &= \frac{\hat{\sigma}_w^2(\overline{Y}_1 - \overline{Y}_2) - \hat{\sigma}_{wy}(\overline{W}_1 - \overline{W}_2)}{\hat{\sigma}_w^2 + q(1-q)(\overline{W}_1 - \overline{W}_2)^2} \\
&= \left(\frac{\hat{\sigma}_w^2}{\hat{\sigma}_w^2 + q(1-q)(\overline{W}_1 - \overline{W}_2)^2}\right)(\overline{Y}_1 - \overline{Y}_2) \\
&\quad - \frac{\hat{\sigma}_{wy}(\overline{W}_1 - \overline{W}_2)}{\hat{\sigma}_w^2 + q(1-q)(\overline{W}_1 - \overline{W}_2)^2}
\end{aligned}$$

And $\overline{W}_1 - \overline{W}_2 \to 0$ as $n \to \infty$.

The real reason it works (Details omitted)

- If covariates were unrelated to omitted variables and measured without error, everything would be fine.
- Call this the “pretend model.”
- But actually, covariates are related to omitted variables and measured *with* error.
- Call this the “true model.”
- Data from the pretend model are indistinguishable from data from the true model.
- This does not always happen.

Parameters

- The true model has more parameters (13 versus 6 in the example).
- Parameters of the pretend model are *functions* of parameters of the true model.
- Regression coefficients of the dummy variables are the same under both models. This is the key.
- It happens only because of random assignment.
- Other parameters of the pretend model are crazy functions of the parameters of the true model.
- But estimation and inference about the *treatment effects* are excellent (as usual) under the pretend model.

For the little example

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 d_i + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \epsilon_i \\
W_i &= \lambda_0 + \lambda_1 X_{i1} + e_i
\end{aligned}
$$

$$
Y_i = \beta_0^* + \beta_1^* d_i + \beta_2^* W_i + \delta_i
$$

- $\beta_1^* = \beta_1$
- $Var(W_i) = \lambda_1^2 \phi_{11} + \omega$
- $\beta_2^* = \frac{\lambda_1(\alpha_1\phi_{11} + \alpha_2\phi_{12})}{\lambda_1^2\phi_{11} + \omega}$
- $Var(\delta_i)$ is breathtaking.

Moral of the story

- Analysis of covariance can greatly increase the precision of an analysis by reducing background noise.
- Precision of estimation translates directly into time and money.
- The covariates may be measured with error and related to other important but unknown variables that influence the dependent variable.
- As long as there is random assignment, it still works beautifully even though the model is wrong.
- Technically, the analysis of covariance model is "equivalent to a re-parameterization."
- Of course you must be sure that the treatment is not influencing the covariate.

Assumption of Unit-treatment additivity

- Without any treatment, the response is $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.
- Treatment j just adds Δ_j to the response, moving all the responses of the units in condition j up (or down) by Δ_j.
- Write it as a multiple regression model with dummy variables:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 d_{i,1} + \beta_3 d_{i,2} + \epsilon_i$$

- Make a table.

Equal slopes model

$$Y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \epsilon$$

Treatment	d_1	d_2	$E(Y\|\mathbf{x})$
1	1	0	$(\beta_0 + \beta_2) + \beta_1 x$
2	0	1	$(\beta_0 + \beta_3) + \beta_1 x$
3	0	0	$\beta_0 + \beta_1 x$

Look at the least squares means

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 d_1 + \hat{\beta}_3 d_2$$

Treatment	d_1	d_2	Estimated Response
1	1	0	$(\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 \overline{x}$
2	0	1	$(\hat{\beta}_0 + \hat{\beta}_3) + \hat{\beta}_1 \overline{x}$
3	0	0	$\hat{\beta}_0 + \hat{\beta}_1 \overline{x}$

- The least squares means are actually $\hat{Y}$ values.
- In plain language, call them "corrected means," or something like "average teaching evaluation, corrected for teacher's age."

Equal slopes assumption is testable

- Interaction means slopes are not equal: “It depends.”
- Form a product of quantitative variable by each dummy variable for the categorical variable.

$$Y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 x\, d_1 + \beta_5 x\, d_2 + \epsilon$$

- Make a table.

Unequal slopes model

$$Y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 x\, d_1 + \beta_5 x\, d_2 + \epsilon$$

Treatment	d_1	d_2	$E(Y\|\mathbf{x})$
1	1	0	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4)x$
2	0	1	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5)x$
3	0	0	$\beta_0 + \beta_1 x$

Sample questions

Group	d_1	d_2	$E(Y\|\mathbf{x})$
1	1	0	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4)x$
2	0	1	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5)x$
3	0	0	$\beta_0 + \beta_1 x$

What null hypothesis would you test?

- Are all the slopes equal?
- Compare slopes for group one vs three.
- Compare slopes for group one vs two.
- Is there an interaction between treatment and covariate?
- Test the null hypothesis of equal regressions.

Copyright Information