Analysis of Covariance Models

• **A simple situation suited for ANCOVA is when we have two independent variables affecting the response: one is a factor, and the other is a continuous variable.**

• **ANCOVA combines the one-way ANOVA model and the SLR model:**

$$Y_{ij} = \beta_0 + \tau_i + \beta_1 X_{ij} + \varepsilon_{ij}, i = 1, \ldots, t, j = 1, \ldots, n_i.$$

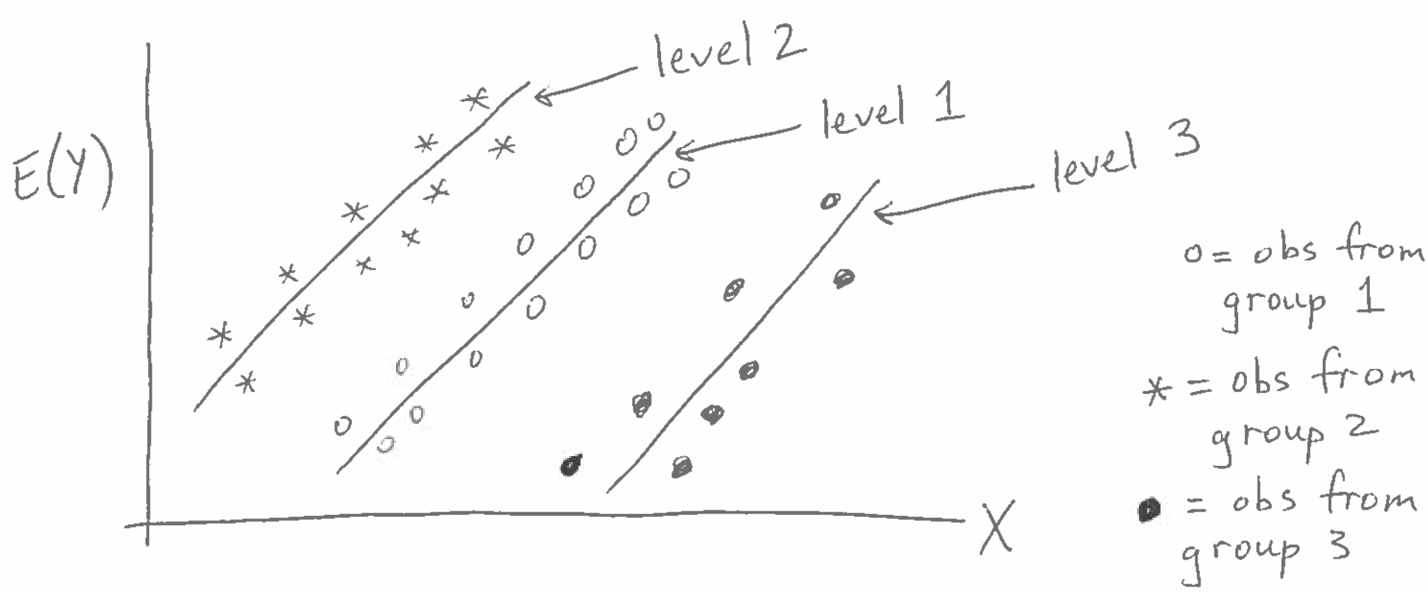
• **For the *t* levels of the factor (*i* = 1, …, *t*), define:**

$$\beta_{0i} = \beta_0 + \tau_i$$

$$\rightarrow Y_{ij} = \beta_{0i} + \beta_1 X_{ij} + \varepsilon_{ij} \quad (i = 1, \ldots, t)$$

• **This shows a set of *t* SLR lines having equal slopes, but having different intercepts for each of the *t* levels of the factor.**

Picture (parallel lines relating E(*Y*) and *X*):



Example: A study analyzing blood pressure reduction (Y) in patients. The factor is type of drug (3 different drugs). However, the weight of the patient (a continuous variable) will also affect blood pressure reduction.

• If we're interested in the effect of each drug on BP reduction, we should account for the patients' weights as well.

• One way: Break weights into categories (levels) and make weight a blocking factor.

Problems: (1) There may not be enough people in some of the weight categories.
(2) We may not know weight is affecting BP reduction until after the experiment is ongoing.
(3) Weight is inherently continuous.
(4) In some studies, there may be several continuous variables affecting the response.

• ANCOVA achieves similar benefits to blocking, but is preferable when controlling for continuous covariates.

Example: Table 11.6 data (p. 593)

• Analyzing the effect of 3 types of classes on students' post-class test score in trigonometry.

Y **(POST) =** score on post-class trig test

Factor (CLASSTYPE) = type of class student is in

3 levels
1 = no computer math
2 = full semester of computer math
3 = partial use of computer math

- **We want to control for previous knowledge of trigonometry.**

covariate ***X*** **(PRE) =** score on pre-class trig test

Model equation: $Y_{ij} = \beta_0 + \tau_i + \beta_1 X_{ij} + \varepsilon_{ij}$

$i = 1, 2, 3$

$j = 1, \ldots, n_i$

n_i = # of students in the i-th class.

- **See example for ANCOVA data analysis (output similar to Table 11.7, p. 594).**

- **Important pieces of output: Overall F* = 8.46 (P-value near 0) → our model is useful overall.**

(1) Does the covariate (pre-class score) have a significant effect on post-class score?

$H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$

F-test for effect of PRE-class score: $F^* = 20.57$ (P-value ≈ 0)

t-test for effect of PRE-class score: $t^* = 4.54$ (P-value ≈ 0)

Reject H_0 in favor of $H_a: \beta_1 \neq 0$.

Yes, pre-class score is useful in this model.

(2) Does the factor (type of class) have a significant effect on post-class score? (Look at the Type III SS)

$H_0: \tau_1 = \tau_2 = \tau_3 = 0$

F-test for CLASSTYPE (Full vs. Reduced)

$F^* = 4.77$ (P-value = .0125)

df = (2,52) Note $F^* > F_{.05,2,52} \approx 3.17$

Using $\alpha = .05$: Yes, class type has a significant effect on post-class score, for any particular value of pre-class score. Here, we rejected H_0.

- **We also get least-squares estimates** $\hat{\beta}_0, \hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\beta}_1$ **for this model.** $\hat{\beta}_0 = 14.431$, $\hat{\tau}_1 = -4.058$, $\hat{\tau}_2 = -5.015$, $\hat{\tau}_3 = 0$, $\hat{\beta}_1 = 0.7732$

- **Interpreting** $\hat{\beta}_1 = 0.7732$**:**

We estimate that the expected post-class score increases by 0.7732 points for each one-point increase in pre-class score, holding constant the level of class type.

It appears that, on average, class type 3 produces highest mean post-class score, and class type 2 produces lowest mean post-class score (for any given preclass score)

• **Adjusted estimated mean post-class scores for each type of class (at any given value of pre-class score) are:**

Class type 1: $\hat{\beta}_0 + \hat{\tau}_1 + \hat{\beta}_1 X = 14.431 - 4.058 + 0.7732\,X$

Class type 2: $\hat{\beta}_0 + \hat{\tau}_2 + \hat{\beta}_1 X = 14.431 - 5.015 + 0.7732\,X$

Class type 3: $\hat{\beta}_0 + \hat{\tau}_3 + \hat{\beta}_1 X = 14.431 + 0 + 0.7732\,X$

• **We can extend the ANCOVA model to have more than one covariate.**

Example: Suppose we had two continuous covariates, pre-test score and IQ.

Results from software:

$$\hat{Y} = \hat{\beta}_0 + \hat{\tau}_i + \hat{\beta}_1(\text{PRE}) + \hat{\beta}_2(\text{IQ}) \qquad (i=1,2,3)$$

where $\hat{\beta}_0 = -9.889$, $\hat{\tau}_1 = -4.987$, $\hat{\tau}_2 = -6.390$, $\hat{\tau}_3 = 0$, $\hat{\beta}_1 = 0.780$, $\hat{\beta}_2 = 0.213$

• **Interpreting** $\hat{\beta}_2 = 0.213$ **:**

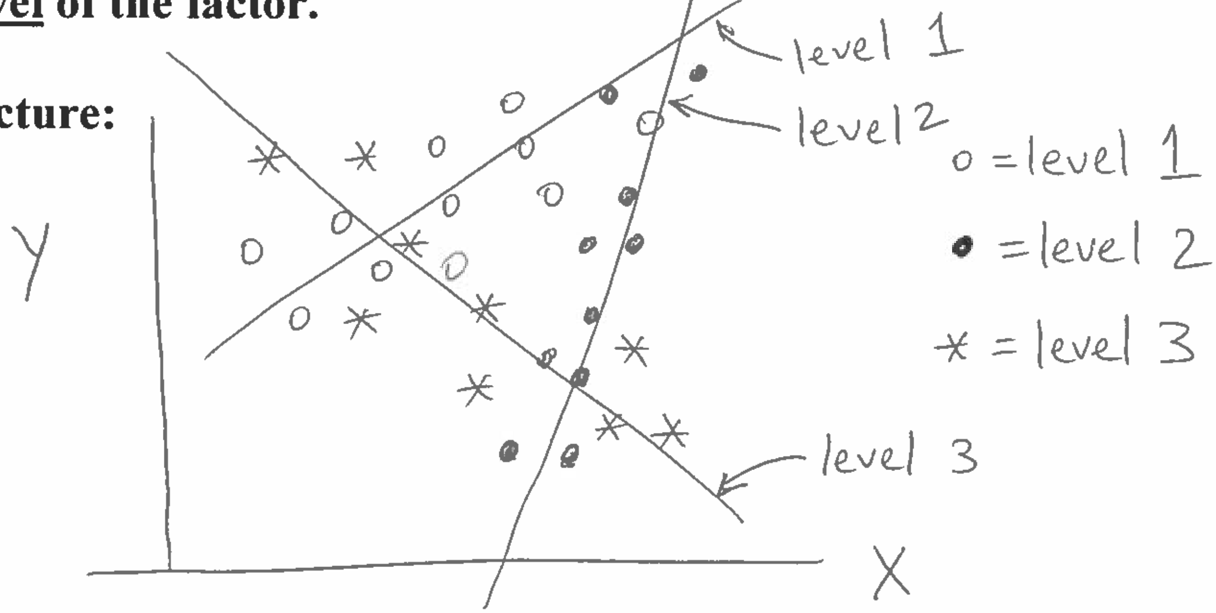
We estimate that the expected post-class score increases by 0.213 points for each one-point increase in IQ, holding constant type of class and pre-class score.

Inference results similar: both pre-class score and IQ are significant in model, and class type is a significant factor (P-value = .001)

Unequal Slopes Situation

- **Maybe the effect of the covariate is <u>different for each level</u> of the factor.**

Picture:



coefficient of X here is $\beta_1 + \beta_{2i} Z_i$ ↑ differs depending on the level of i

- **We can formally test whether this is true by including a term for the <u>interaction</u> between the factor and the covariate.** $E(Y) = \beta_0 + \tau_i Z_i + \beta_1 X + \beta_{2i} X Z_i$ $\quad i = 1, \dots, t$

↑ Z_i's = dummy variables

<u>Example</u>: Are the slopes unequal for the model with factor CLASSTYPE and covariate PRE?

- **Analysis: Include CLASSTYPE by PRE interaction term.**

- **Look at F-test for interaction term in output:**

F* = 0.33 **P-value =** 0.7235

Conclusion: Fail to reject H_0: "all β_{2i}'s $= 0$". Conclude there is no significant factor × covariate interaction. Conclude the equal-slopes ANCOVA model is sufficient.