THE ANALYSIS OF COVARIANCE

ERSH 8310

Keppel and Wickens Chapter 15

Today's Class

- Initial Considerations
- Covariance and Linear Regression
- The Linear Regression Equation
- The Analysis of Covariance
- Assumptions Underlying the Analysis of Covariance
- Example ANCOVA Analysis

The Analysis of Covariance

The Analysis of Covariance

- In a completely randomized design subjects are randomly assigned to the experimental treatments.
- Generally, the completely randomized design is relatively deficient in power.
- If there is a variable available before the start of experiment that is reasonably correlated with the dependent variable (i.e., control variable, concomitant variable, or covariate; e.g., intelligence, grade point average, etc.), we may either employ blocking or statistical adjustment.

The Analysis of Covariance

- The randomized-blocks design includes groups of homogeneous subjects drawn from respective blocks.
- Advantages of the randomized-blocks design are:
  - Blocking helps to equate the treatment groups before the start of the experiment more effectively than is accomplished in the completely randomized design.
  - The power is increased because smaller error term usually associated with the blocking design.
  - Interactions can be assessed.

The Analysis of Covariance

- Disadvantages may include:
  - There will be cost of introducing the blocking factor.
  - It may be difficult to find blocking factors that are highly correlated with the dependent variable.
  - Loss of power may occur if a poorly correlated blocking factor is used.

The Analysis of Covariance

- The analysis of covariance reduces experimental error by statistical, rather than experimental, means.
- Subjects are first measured on the concomitant variable called the covariate which consists of some relevant ability or characteristic.
- Subjects are then randomly assigned to the treatment group without regard for their scores on the covariate.

The Analysis of Covariance

- The analysis of covariance refines estimates of experimental error and uses the adjust treatment effects for any differences between the treatment groups that existed before the experimental treatments were administered.

Covariance and Linear Regression

Covariance and Linear Regression

- The correlation coefficient between two variables X and Y is:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y},$$

- The standard de

$$s_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}},$$

- Where the covariance between X and Y is:

$$s_{XY} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1},$$

- The standard deviation of Y is:

$$s_Y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}}.$$

Covariance and Linear Regression

- We may define the sum of products:

$$SP_{XY} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}),$$

- And the sums of squares:

$$SS_X = \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad SS_Y = \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

- Consequently:

$$r_{XY} = \frac{SP_{XY}}{\sqrt{SS_X \, SS_Y}}.$$

The Linear Regression Equation

The Linear Regression Equation

- The linear regression line relating the dependent variable Y to the covariate X is:

$$Y = b_0 + b_1 X$$

- And the prediction equation for i is:

$$\hat{Y}_i = b_0 + b_1 X_i,$$

- Where:

$$b_0 = \bar{Y} - b_1 \bar{X} \qquad b_1 = r_{XY} \frac{S_Y}{S_X} = \frac{SP_{XY}}{SS_X} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

Residual Variation and the Linear Model

- The sum of the squared deviation from the mean is:

$$SS_Y = \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

- The sum of th luals from the regression line is:

$$SS_{Y|X} = \sum_{i=1}^{n} (Y_i - Y')^2.$$

Residual Variation and the Linear Model

- The amount of the Y variability that can be attributed to the regression equation is:

$$SS_{regression} = SS_Y - SS_{Y|X}.$$

- The squared correlation coefficient is:

$$r_{XY}^2 = \frac{SS_{regression}}{SS_Y}.$$

The Analysis of Covariance

The Analysis of Covariance

- The analysis of covariance tests for differences between groups by comparing a description of the data based on a single regression line to one based on lines with the same slope and different intercepts for each group.
- For example, when a = 2, the same slope b_1 is obtained from (where (1) and (2) designate groups):

$$b_1 = \frac{SPxy(1) + SPxy(2)}{SSx(1) + SSx(2)}$$

The Analysis of Covariance and the General Linear Model

- For the analysis of covariance, the alternative-hypothesis model is:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \alpha_j + E_{ij}$$

- And the null-hypothesis model is:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + E_{ij}.$$

The Analysis of Covariance and the General Linear Model

Note that

$$SS_{effect} = SS_{unexp}^{H_0} - SS_{unexp}^{H_1}$$

and

$$df_A = df_{unexp}^{H_0} - df_{unexp}^{H_1}.$$

We may also obtain

$$MS_A = SS_A/df_A$$

and

$$MS_{error} = SS_{unexp}^{H_1}/df_{unexp}^{H_1}.$$

Hence

$$F = MS_A/MS_{error}$$

with the two respective degrees of freedoms.

Adjusted Means

- The adjusted mean is:

$$\bar{Y}'_j = \bar{Y}_j - b_1(\bar{X}_i - \bar{X}_T).$$

Contrasts on the Adjusted Means

Use

$$\hat{\psi}' = \sum_{j=1}^{a} c_j \bar{Y}_j' ,$$

$$\hat{\sigma}_{\psi'} = \sqrt{\sum_{j=1}^{a} c_j^2 \, s_{M_j'}} ,$$

and

$$t = \hat{\psi}' \,/\, \hat{\sigma}_{\psi'}$$

with degrees of freedom equal to those of the error from the overall analysis.

Extensions of the Design

The design can be extended to include more factors and more covariates.

Assumptions Underlying the Analysis of Covariance

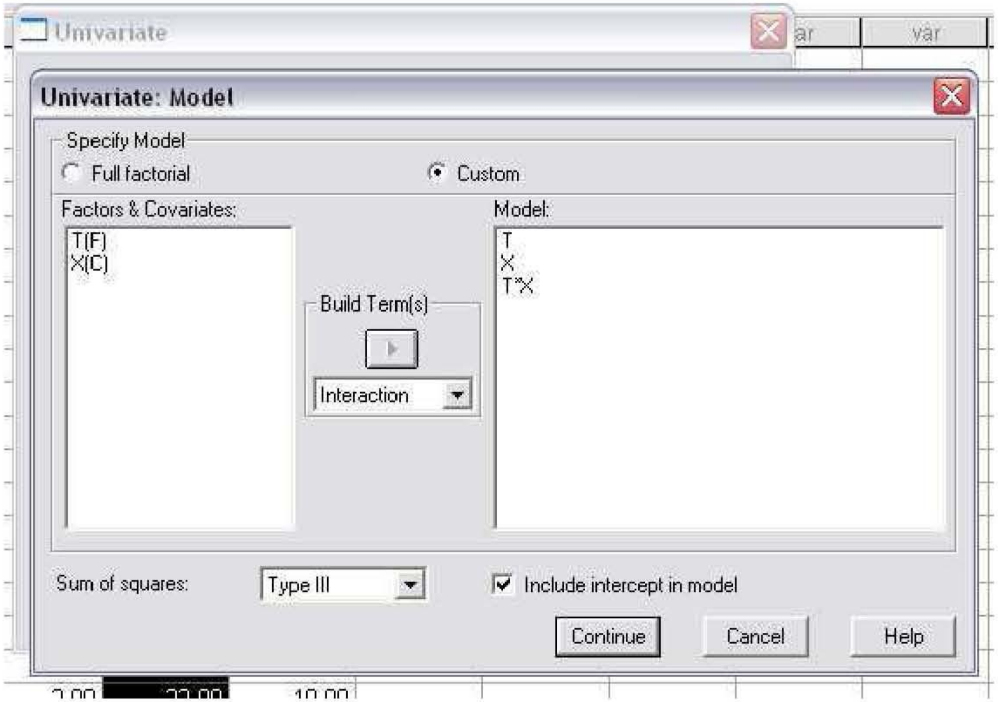
Assumptions Underlying the Analysis of Covariance

- Three assumptions in addition to the usual analysis of variance assumptions are:
  1. The assumption of linear regression: The deviations from regression are normally and independently distributed in the population, with means of zero and homogeneous variances.
  2. The assumption of homogeneous group regression coefficients: The within groups regression coefficient is actually an average of the regression coefficients for the respective treatment groups.
  3. The exact measurement of the covariate: The covariate is measured without error.

Example ancova analysis

In SPSS...

- Using Analyze...General Linear Model...Univariate
- First, test for presence of significant interaction.
- This must be done under the Model Box:



In SPSS...

Tests of Between-Subjects Effects

Dependent Variable: Dependent Variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	222.234[a]	7	31.748	10.193	.000
Intercept	300.680	1	300.680	96.540	.000
T	1.914	3	.638	.205	.892
X	152.279	1	152.279	48.893	.000
T * X	1.006	3	.335	.108	.955
Error	99.666	32	3.115		
Total	12642.000	40			
Corrected Total	321.900	39			

a. R Squared = .690 (Adjusted R Squared = .623)

Omitting the Interaction

- Using Analyze...General Linear Model...Univariate
- Second, remove interaction term.
- Save the means for each group.

Tests of Between-Subjects Effects

Dependent Variable: Dependent Variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	221.227[a]	4	55.307	19.220	.000
Intercept	313.501	1	313.501	108.992	.000
X	165.127	1	165.127	57.408	.000
I	65.042	3	21.681	7.538	.001
Error	100.673	35	2.876		
Total	12642.000	40			
Corrected Total	321.900	39			

a. R Squared = .687 (Adjusted R Squared = .652)

Estimated Marginal Means

Treatment Condition

Dependent Variable: Dependent Variable

Treatment Condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	15.521[a]	.538	14.430	16.613
2.00	18.027[a]	.537	16.937	19.116
3.00	19.024[a]	.536	17.935	20.113
4.00	17.628[a]	.537	16.537	18.718

a. Covariates appearing in the model are evaluated at the following values: Covariate = 7.1250.

Omitting the Covariate

- Using Analyze...General Linear Model...Univariate
- Remove the covariate altogether.
- Save the means for each group.

Tests of Between-Subjects Effects

Dependent Variable: Dependent Variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	56.100[a]	3	18.700	2.533	.072
Intercept	12320.100	1	12320.100	1668.637	.000
T	56.100	3	18.700	2.533	.072
Error	265.800	36	7.383		
Total	12642.000	40			
Corrected Total	321.900	39			

a. R Squared = .174 (Adjusted R Squared = .105)

Estimated Marginal Means

Treatment Condition

Dependent Variable: Dependent Variable

Treatment Condition	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	15.800	.859	14.057	17.543
2.00	17.900	.859	16.157	19.643
3.00	19.100	.859	17.357	20.843
4.00	17.400	.859	15.657	19.143

Adjusted Means

- ANCOVA adjusts the mean for each treatment group by that of the mean deviation for the covariate:

$$\bar{Y}_{j(\text{adj})} = \bar{Y}_j - b(\bar{X}_j - \bar{X})$$

- Although not easily found in SPSS, consider the following (with $b = 1.013$):

Group	$\bar{Y}$	$\bar{X}$
1	15.8	7.4
2	17.9	7.0
3	19.1	7.2
4	17.4	6.9
Overall	17.55	7.125

Tests for Differences of Adjusted Means

- With the ANCOVA adjusted means, one can now test for differences between the means (using, for example, multiple planned comparisons).
- Using Analyze...General Linear Model...Univariate
- Use the covariate.
- Go under Options...

Pairwise Comparisons

Dependent Variable: Dependent Variable

(I) Treatment Condition	(J) Treatment Condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1.00	2.00	-2.505*	.760	.014	-4.632	-.378
	3.00	3.503*	.759	.000	5.629	1.380
	4.00	2.107	.761	.054	4.236	.023
2.00	1.00	2.505*	.760	.014	.379	4.632
	3.00	.997	.759	1.000	3.120	1.126
	4.00	.399	.759	1.000	-1.723	2.520
3.00	1.00	3.503*	.759	.000	1.380	5.626
	2.00	.997	.759	1.000	1.126	3.120
	4.00	1.396	.760	.447	-.720	3.520
4.00	1.00	2.107	.761	.054	-.023	4.236
	2.00	.399	.759	1.000	2.520	1.723
	3.00	1.396	.760	.447	3.520	.728

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Final Thought

- Today’s class covered a method for controlling for important variables in an experiment: ANCOVA.

- ANCOVA is a general technique that adds additional (continuous) variables to a model and adjusts for the values of such variables.
- Any ANOVA design can include such variables.

Next Time…

- Next class: 11/27.
- Chapter 16: Within Subject Designs.