

การวิเคราะห์ความแปรปรวนร่วม (Analysis of Covariance)

ดร.นำชัย ศุภฤกษ์ชัยสกุล
สถาบันวิจัยพฤติกรรมศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ



BSRI



หัวข้อการบรรยาย

- การสรุปความสัมพันธ์เชิงสาเหตุ
- ตรรกะของแบบแผนการวิจัยเชิงทดลอง
- การวิเคราะห์แบบแผนการวิจัยเชิงทดลอง
 - ANOVA
 - ANCOVA
 - MANOVA



ANCOVA

- ย่อมาจากคำว่า Analysis of Covariance
- เป็นการวิเคราะห์ที่ผสมผสานกันระหว่างการวิเคราะห์ ANOVA กับ Regression
- เพื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ย โดยที่ควบคุมความแตกต่างของกลุ่มตัวอย่างอื่นเนื่องมาจากตัวแปรแทรกซ้อนต่าง ๆ (Nuisance Variables)
- ตัวแปรที่นำมาควบคุมเรียกว่า “ตัวแปรร่วม” (Covariates)



ANCOVA

- ตัวแปรร่วม ที่นำมาใช้วิเคราะห์ ANCOVA จะเป็นตัวแปรที่มีระดับการวัดเชิงปริมาณ (Numerical Variable)
- ตัวแปรร่วมต้องเป็นตัวแปรที่มีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม
- อาจจะเป็น
 - ตัวแปรเดียวกันกับตัวแปรตาม แต่วัดคนละช่วงเวลา เช่น Pretest, Posttest
 - คนละตัวแปรกัน แต่มีความสัมพันธ์หรือส่งผลกระทบต่อตัวแปรตาม

ANCOVA

- ค่าเฉลี่ยที่นำมาเปรียบเทียบเป็นค่าเฉลี่ยที่ปรับ (Adjusted Means) ตามตัวแปรร่วม

$$H_0 : \mu'_1 = \mu'_2 = \mu'_3 = \dots = \mu'_k$$

H_1 : มีค่าเฉลี่ยอย่างน้อยหนึ่งคู่ที่แตกต่างกัน



Logic of ANOVA

$\frac{\text{Signal}}{\text{Noise}} \longrightarrow \frac{\text{Between Group Variation}}{\text{Within Group Variation}}$

$$F = \frac{\text{Between Group Variance}}{\text{Within Group Variance}}$$



Logic of ANOVA

$$F = \frac{\text{Mean Square Between Group}}{\text{Mean Square Within Group}}$$

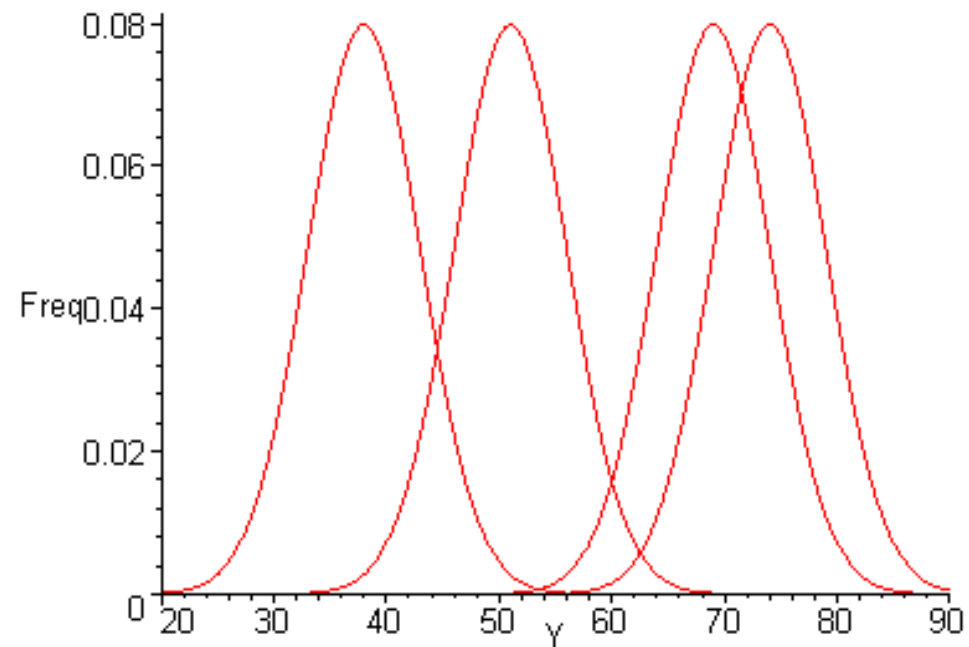
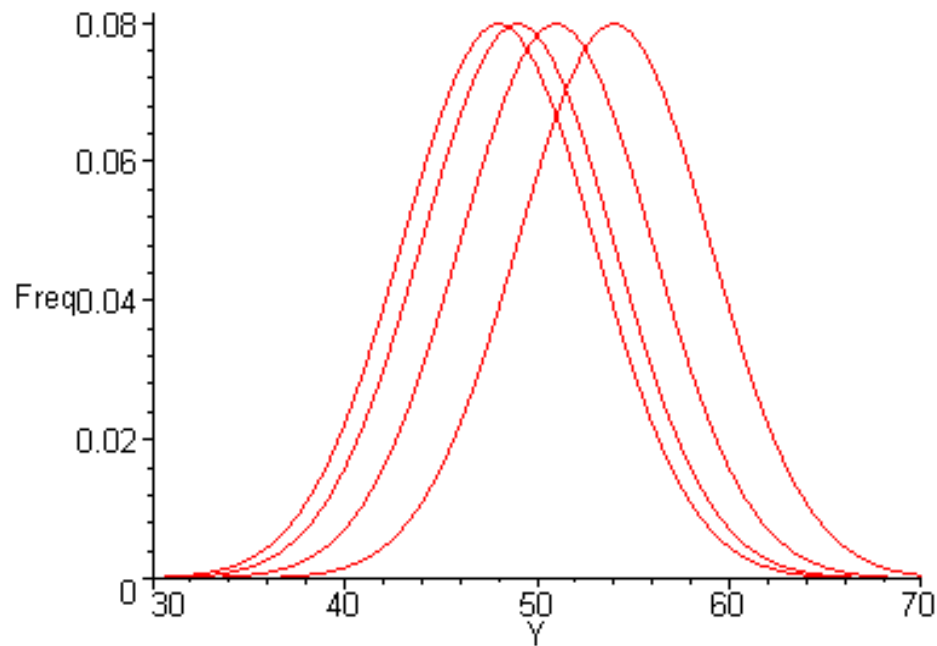
$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1}$$

Sum of Square
Degree of Freedom

$$\text{Variance} = \frac{SS}{DF} = \text{Mean Square}$$

Partition of Variance

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}$$

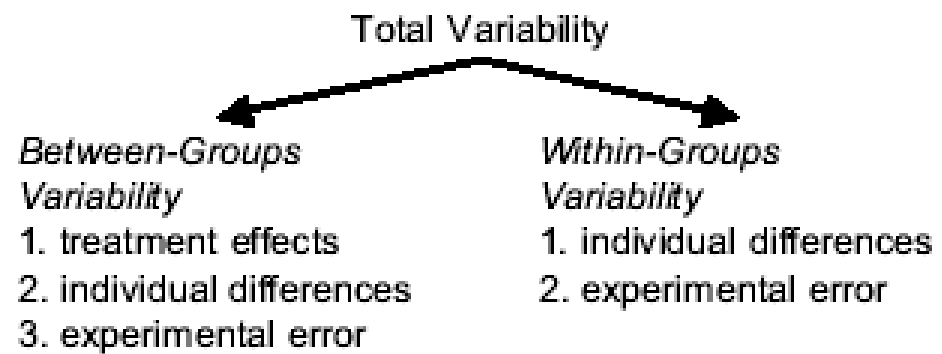


Partition of Variance

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}$$

$F = \frac{\text{variance between sample means}}{\text{variance expected by chance}}$

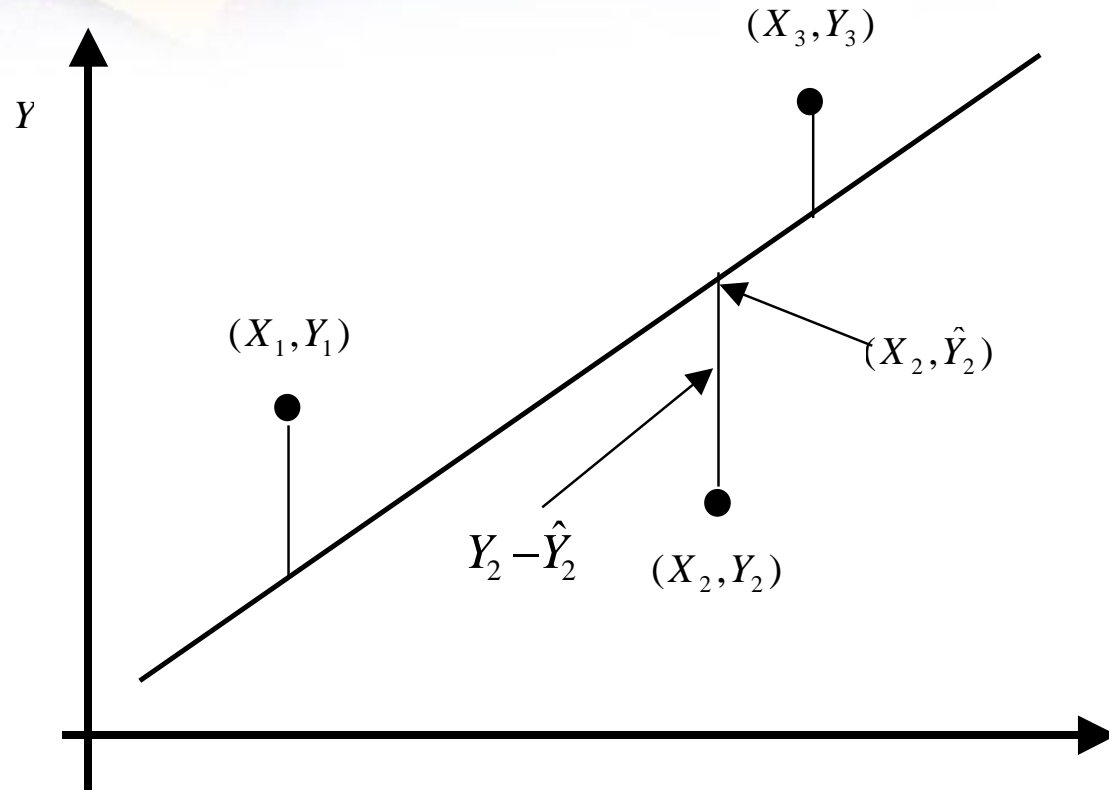
individual differences + experimental error = error (or unexplained variance)



$$F = \frac{\text{Treatment Effect} + \text{Error}}{\text{Error}}$$

$$F = \frac{0 + \text{Error}}{\text{Error}} = 1$$

สมการ Regression



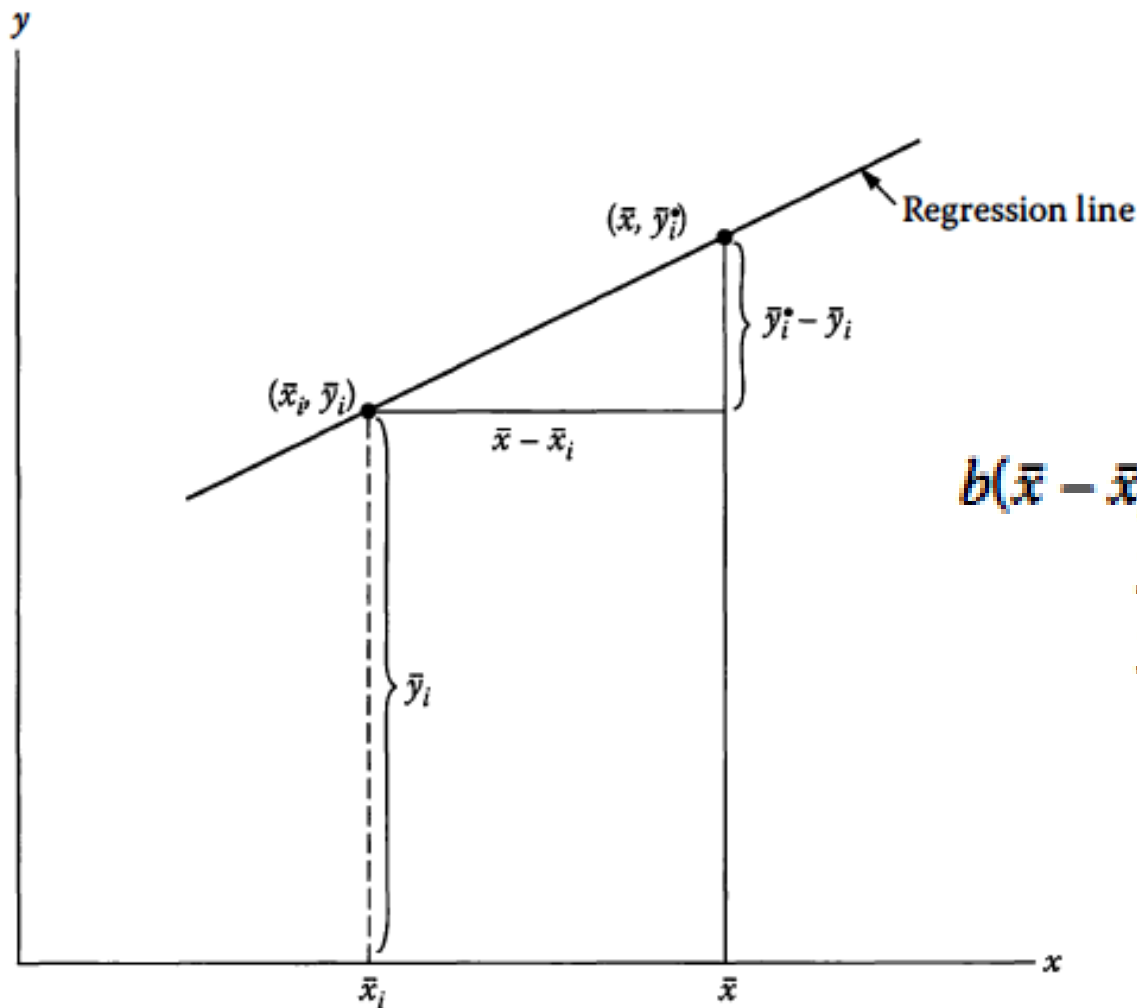
$$\Sigma(Y - \hat{Y})^2$$

การสร้างเส้นพยายามให้เกิด SS_{Error} น้อยที่สุด

Least-Squares estimates

Adjusted Means

$$\text{Slope of straight line} = b = \frac{\text{change in } y}{\text{change in } x}$$
$$b = \frac{\bar{y}_i^* - \bar{y}_i}{\bar{x} - \bar{x}_i}$$



$$b(\bar{x} - \bar{x}_i) = \bar{y}_i^* - \bar{y}_i$$
$$\bar{y}_i^* = \bar{y}_i + b(\bar{x} - \bar{x}_i)$$
$$\bar{y}_i^* = \bar{y}_i - b(\bar{x}_i - \bar{x})$$



วัตถุประสงค์ของ ANCOVA

- ขจัดความลำเอียงที่เกิดขึ้นอย่างเป็นระบบ (Systematic Bias)
 - Systematic Bias สำหรับการเปรียบเทียบค่าเฉลี่ยก็คือความแตกต่างระหว่างกลุ่มที่เกิดจากตัวแปรแทรกซ้อนต่าง ๆ
 - สำหรับงานวิจัยเชิงทดลอง ปัญหานี้มักเกิดจากการไม่ได้สุ่มหน่วยทดลองเข้ากลุ่ม (Random Assignment) หรือใช้กลุ่มที่มีอยู่แล้ว
 - แม้จะมีการสุ่มเข้ากลุ่มก็ตาม ปัญหานี้ก็เกิดขึ้นได้ถ้าแต่ละกลุ่มมีจำนวนหน่วยทดลองน้อย
 - หรือในกรณีที่กลุ่มมีความแตกต่างกันในคะแนนที่วัดก่อน (Pretest) การทดลอง

วัตถุประสงค์ของ ANCOVA

- เพื่อลดความแปรปรวนของความคลาดเคลื่อน (Error Variance) หรือความแตกต่างของคะแนนภายในกลุ่ม
 - ทำให้มีอำนาจการทดสอบ (Power) สูงขึ้น
 - R^2_{xy} = ความแปรปรวนของ y ที่ x สามารถอธิบายได้

$$MS_w - MS_w r_{xy}^2 = MS_w (1 - r_{xy}^2)$$

การเลือกตัวแปรร่วม

- ต้องมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม
- ถ้ามีหลายตัว ยิ่งตัวแปรร่วมมีความสัมพันธ์กันต่ำ ยิ่งบ่งบอกถึงความแปรปรวนของตัวแปรตามที่ตัวแปรร่วมแต่ละตัวอธิบายได้ ไม่ซ้ำซ้อนกัน

x_1 and x_2 Low correl.



x_1 and x_2 High correl.



จำนวนตัวแปรร่วม

- Huitema (1980) เสนอไว้ว่า

$$\frac{C + (J - 1)}{N} < .10$$

$$(C + 2)/60 < .10$$

$$C < 4$$

ข้อควรระวัง

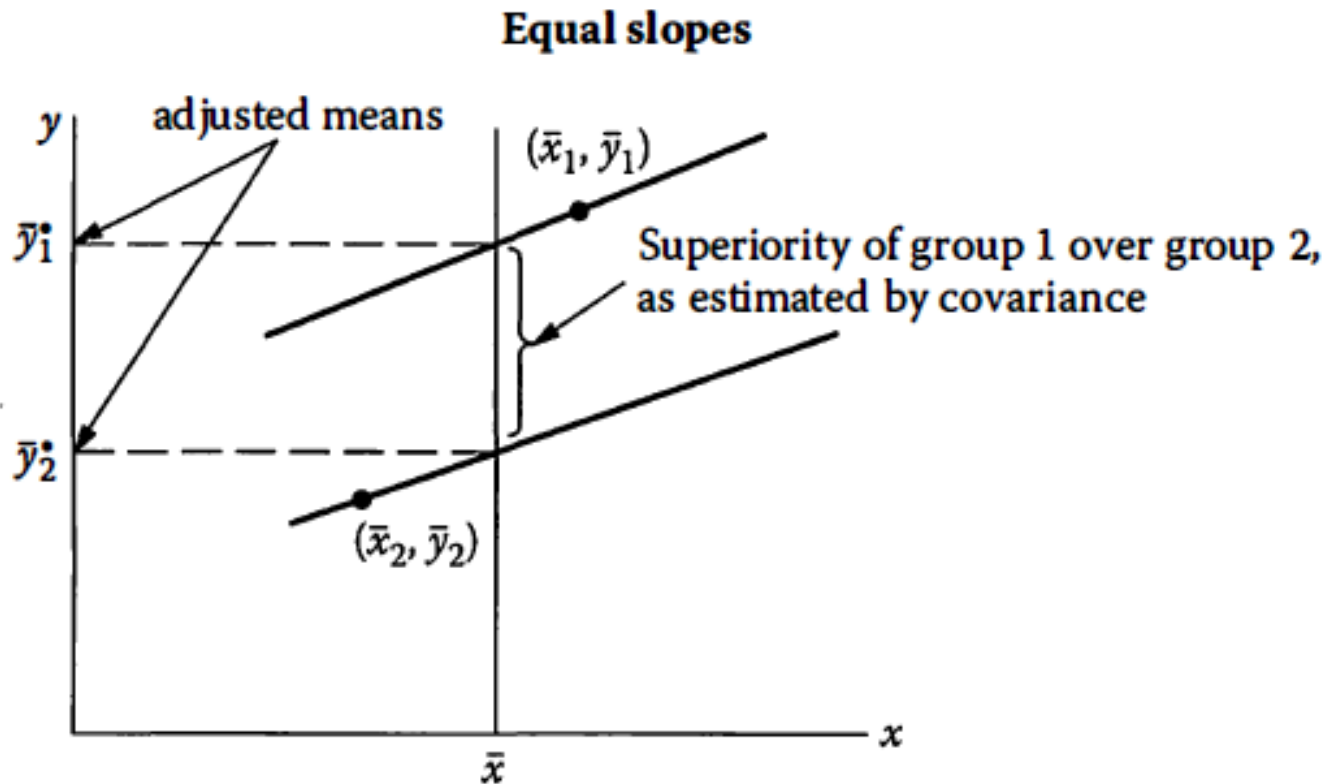
- สำหรับงานวิจัยเชิงทดลอง ตัวแปรร่วมต้องเป็นตัวแปรที่วัดก่อนที่จะให้ตัวแปรทดลอง
- สำหรับการวิจัยที่ไม่ใช่เชิงทดลอง แม้ว่าจะใช้ตัวแปรร่วมหลายตัว แต่ก็ไม่ได้ทำให้กลุ่มที่จะเปรียบเทียบมีความเท่าเทียมกัน
- ตัวแปรร่วมต้องปราศจากความคลาดเคลื่อนในการวัด (Measurement Errors)
- ถ้าตัวแปรร่วมมีความสัมพันธ์กับตัวแปรตามต่ำ ($r_{xy} < .4$) ให้นำตัวแปรร่วมมาใช้เป็นตัวแปรจัดกลุ่ม (Blocking) แทน



Assumptions of ANCOVA

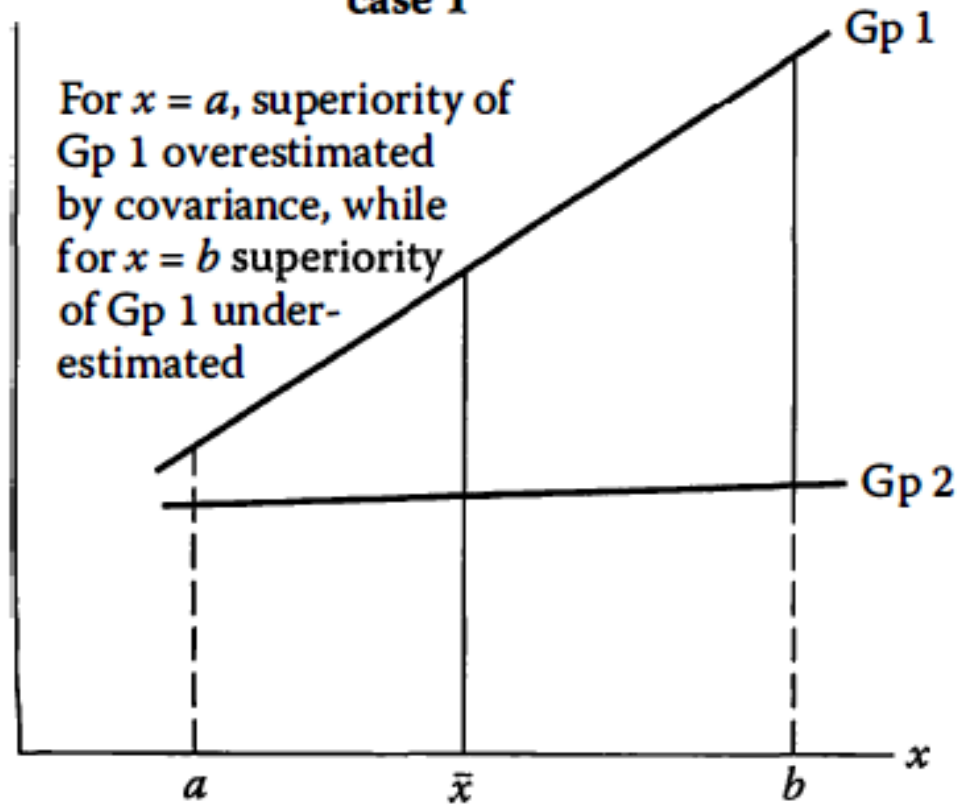
- Normality Distribution
- Homogeneity of Variance
- Independence Observation
- Linear Relationship between Covariate and Dependent Variable
- Homogeneity of Regression Slopes
- Covariates are measured without errors

Equal Slopes Assumption



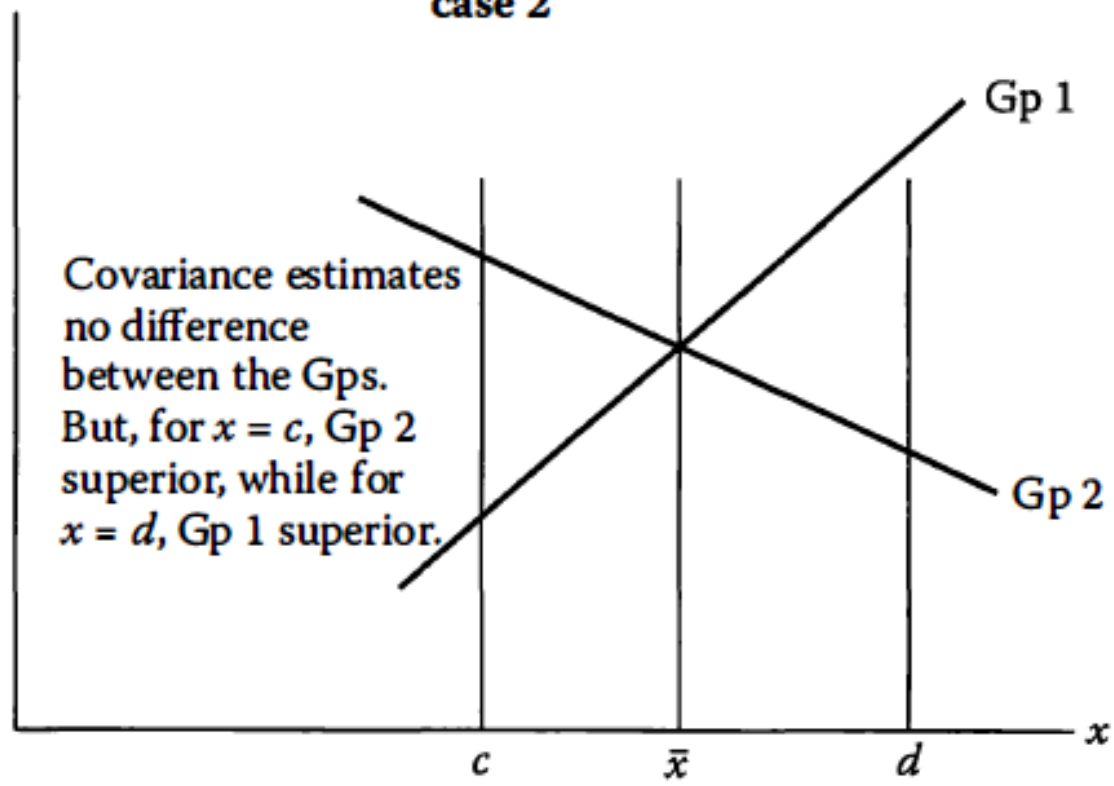
Equal Slopes Assumption

Heterogeneous slopes case 1



Equal Slopes Assumption

Heterogeneous slopes
case 2





กรณีตัวอย่าง

- นาย ก. ทำการวิจัยเพื่อทดสอบว่า วิธีการฝึกความจำวิธีใดจะมีประสิทธิภาพมากที่สุด ระหว่าง
 - วิธีที่ 1 ฝึกแบบใช้การสร้างความเชื่อมโยง (Control Group)
 - วิธีที่ 2 ฝึกแบบใช้เทคนิค Mindmap (Experiment Group)
 - วิธีที่ 3 ฝึกแบบใช้เทคนิคการนั่งสมาธิ (Experiment Group)
- ก่อนการฝึก นาย ก. วัดคะแนนความสามารถในการจำ (escore) หลังจากฝึกเสร็จแล้วก็วัดความสามารถในการจำหลังการฝึก (pscore)

Normality Assumption

Tests of Normality

วิธีการฝึก	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
pscore แบบที่ 1	.070	35	.200 [*]	.992	35	.995
แบบที่ 2	.108	35	.200 [*]	.958	35	.203
แบบที่ 3	.103	35	.200 [*]	.965	35	.320

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Homogeneity of Variance Assumption

Levene's Test of Equality of Error Variances

Dependent Variable: pscore

F	df1	df2	Sig.
.628	2	102	.536

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

1. Design: Intercept+group

ตัวอย่างการวิเคราะห์

Between-Subjects Factors

		Value Label	N
วิธีการฝึก	1	แบบที่ 1	35
	2	แบบที่ 2	35
	3	แบบที่ 3	35

Descriptive Statistics

Dependent Variable: pscore

วิธีการฝึก	Mean	Std. Deviation	N
แบบที่ 1	84.96	6.822	35
แบบที่ 2	83.54	6.407	35
แบบที่ 3	82.28	7.294	35
Total	83.60	6.873	105

ตัวอย่างการวิเคราะห์

ANOVA

pscore

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	125.894	2	62.947	1.341	.266
Within Groups	4786.625	102	46.928		
Total	4912.518	104			

ตัวอย่างการวิเคราะห์

Descriptives

escore

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					แบบที่ 1	35		
แบบที่ 2	35	81.73	8.486	1.434	78.82	84.65	62	96
แบบที่ 3	35	75.83	8.685	1.468	72.85	78.81	54	91
Total	105	81.97	9.258	.903	80.18	83.76	54	98

ตัวอย่างการวิเคราะห์

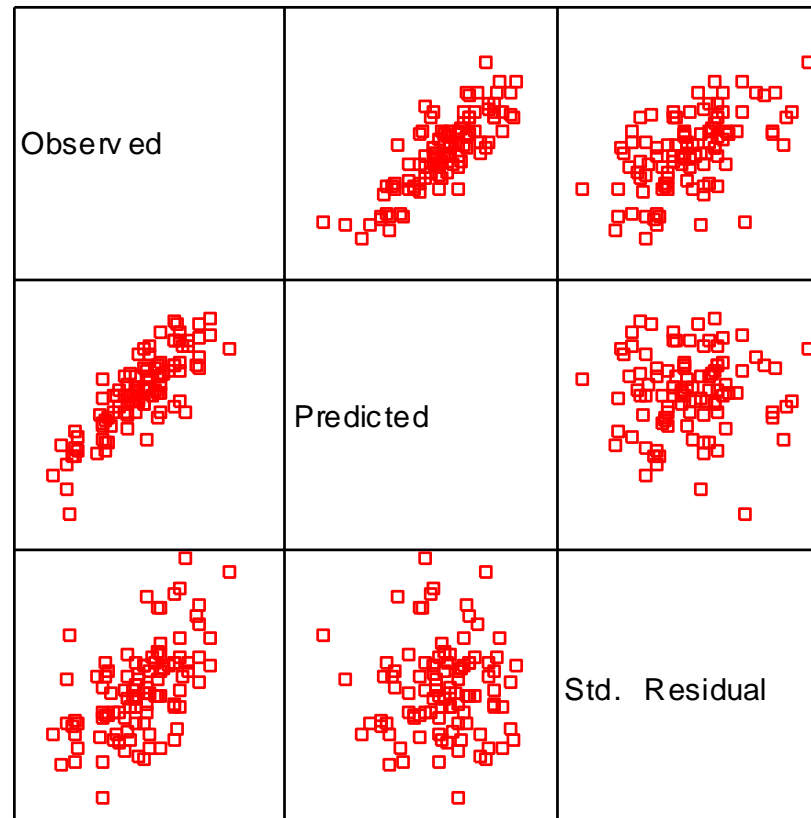
ANOVA

escore

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2743.517	2	1371.758	22.677	.000
Within Groups	6170.138	102	60.492		
Total	8913.655	104			

ตัวอย่างการวิเคราะห์

Dependent Variable: PSCORE(Posttest)



Model: Intercept + ESCORE

ตัวอย่างการวิเคราะห์

Correlations

		escore	pscore
escore	Pearson Correlation	1	.647 ^{**}
	Sig. (2-tailed)		.000
	N	105	105
pscore	Pearson Correlation	.647 ^{**}	1
	Sig. (2-tailed)	.000	
	N	105	105

^{**}. Correlation is significant at the 0.01 level (2-tailed).

ตัวอย่างการวิเคราะห์

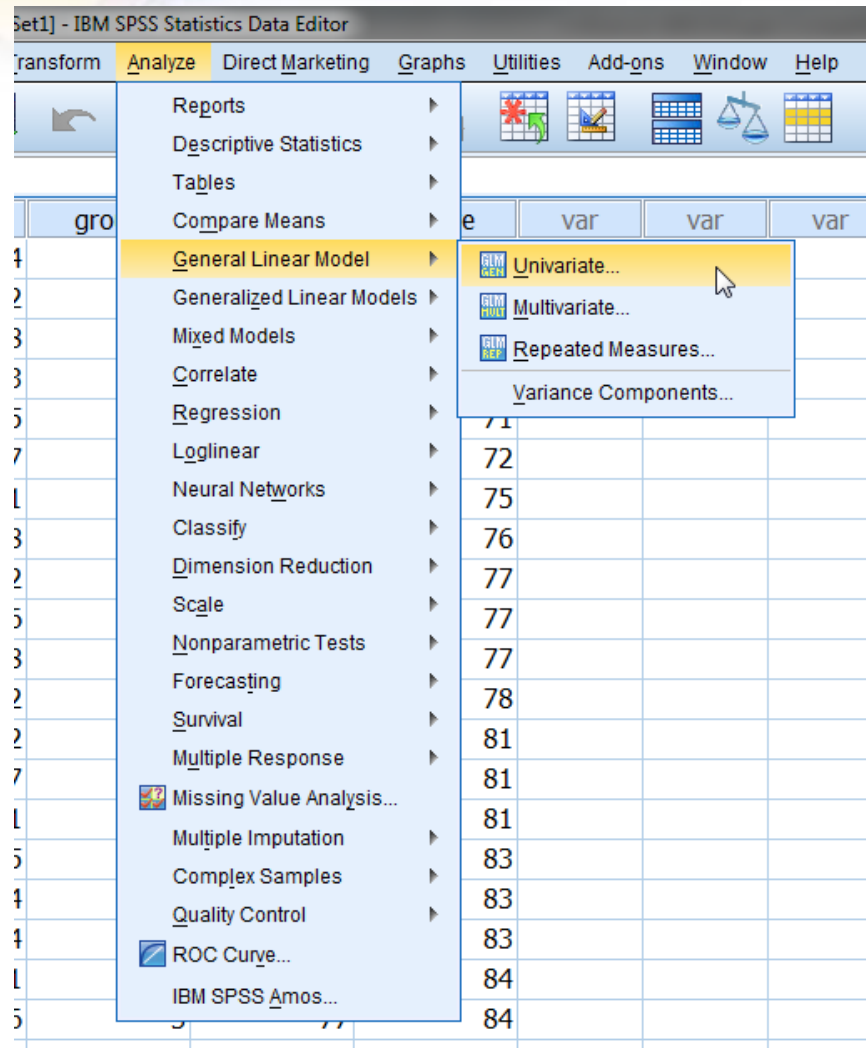
Tests of Between-Subjects Effects

Dependent Variable: pscore

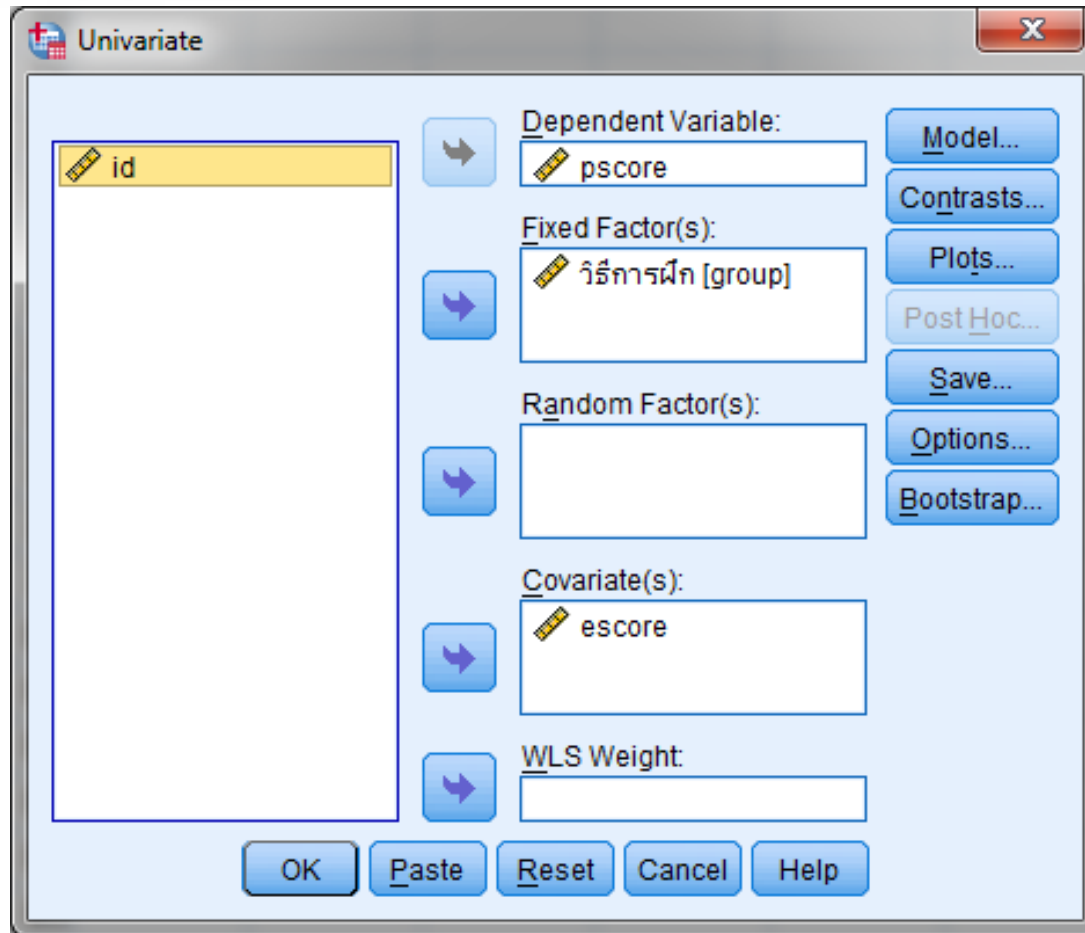
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2385.439 ^a	5	477.088	18.690	.000
Intercept	1062.587	1	1062.587	41.628	.000
group	28.571	2	14.285	.560	.573
escore	1762.979	1	1762.979	69.066	.000
group * escore	49.852	2	24.926	.976	.380
Error	2527.079	99	25.526		
Total	738671.828	105			
Corrected Total	4912.518	104			

a. R Squared = .486 (Adjusted R Squared = .460)

ตัวอย่างการวิเคราะห์



ตัวอย่างการวิเคราะห์



ตัวอย่างการวิเคราะห์

Univariate: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:

(OVERALL)
group

Display Means for:

group

Compare main effects

Confidence interval adjustment:
Bonferroni

Display

Descriptive statistics
 Estimates of effect size
 Observed power
 Parameter estimates
 Contrast coefficient matrix

Homogeneity tests
 Spread vs. level plot
 Residual plot
 Lack of fit
 General estimable function

Significance level: .05 Confidence intervals are 95.0 %

Continue Cancel Help

ตัวอย่างการวิเคราะห์

Tests of Between-Subjects Effects

Dependent Variable: pscore

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2335.587 ¹	3	778.529	30.514	.000
Intercept	1086.185	1	1086.185	42.572	.000
escore	2209.694	1	2209.694	86.607	.000
group	280.362	2	140.181	5.494	.005
Error	2576.931	101	25.514		
Total	738671.828	105			
Corrected Total	4912.518	104			

1. R Squared = .475 (Adjusted R Squared = .460)

ตัวอย่างการวิเคราะห์

Estimates

Dependent Variable: pscore

วิธีการฝึก	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
แบบที่ 1	81.149 ¹	.947	79.270	83.027
แบบที่ 2	83.680 ¹	.854	81.986	85.374
แบบที่ 3	85.957 ¹	.941	84.091	87.823

1. Covariates appearing in the model are evaluated at the following values: escore = 81.97.

ตัวอย่างการวิเคราะห์

Pairwise Comparisons

Dependent Variable: pscore

(I) วิธีการฝึก	(J) วิธีการฝึก	Mean Difference (I-J)	Std. Error	Sig. ¹	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
แบบที่ 1	แบบที่ 2	-2.532	1.280	.152	-5.648	.585
	แบบที่ 3	-4.809*	1.451	.004	-8.341	-1.276
แบบที่ 2	แบบที่ 1	2.532	1.280	.152	-.585	5.648
	แบบที่ 3	-2.277	1.266	.225	-5.358	.805
แบบที่ 3	แบบที่ 1	4.809*	1.451	.004	1.276	8.341
	แบบที่ 2	2.277	1.266	.225	-.805	5.358

Based on estimated marginal means

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

1. Adjustment for multiple comparisons: Bonferroni.



ANCOVA

- เป็นสถิติที่มีอำนาจทดสอบ (Power) สูงกว่า ANOVA เนื่องจากสามารถลดความแปรปรวนของความคลาดเคลื่อน (Error Variance) ที่เกิดจากตัวแปรพร้อมที่นำมาวิเคราะห์ได้
- การวิเคราะห์ความแปรปรวนร่วมจึงเป็นการควบคุมความคลาดเคลื่อนด้วยวิธีการสถิติ (Statistical Control)

ANCOVA

ความแปรปรวนที่อธิบายได้ โดยทรีทเมนต์ (125.878)
ความแปรปรวนของความ คลาดเคลื่อน (4785.307)

ความแปรปรวนทั้งหมด (4911.185)

ความแปรปรวนที่อธิบายได้ โดยทรีทเมนต์ (222.549)
ความแปรปรวนที่อธิบายได้ โดยตัวแปรร่วม (3473.874)
ความแปรปรวนของความ คลาดเคลื่อน (1311.433)



จบการบรรยาย