

제8장 공변량분석(ANCOVA)

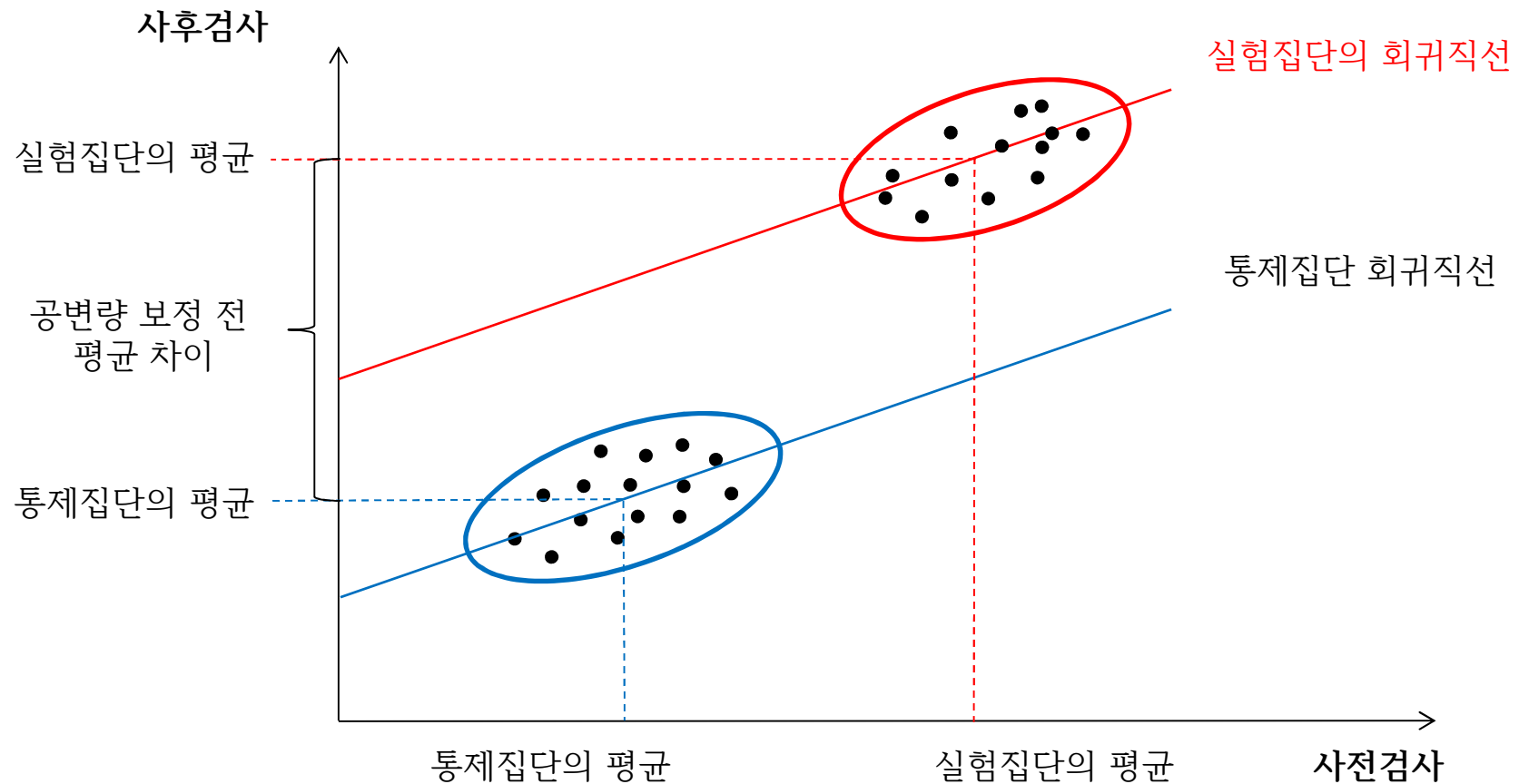
ANCOVA의 개요

- 실험집단간 차이를 검정
- ANCOVA = ANOVA+Regression
- 보정해야 할 공변량이 있는 경우
 - 실험 전후의 집단(실험집단과 통제집단)간 차이에 영향을 미치는 공변량의 효과를 통제 (보정, 제거)
 - 예: 새로운 교육기법의 효과를 알기 위해 실험대상자를 두 개의 집단(실험집단, 통제집단)으로 분류
 - 두 집단에 사전 시험 (Pre-test)을 치름
 - 실험집단에만 새로운 교육기법을 적용
 - 두 집단에 사후 시험 (Post-test)을 치름
 - 위의 실험에서 실험집단이나 통제집단 사전 시험 점수가 높아서 (공부를 더 잘하는 학생이 몰려서) 사후 검사의 결과에 영향을 준다면?
 - 두 집단의 사전 시험점수를 비슷하게 조정한 후 사후 시험점수를 비교한다면?
 - 일반적으로 편의상 **전체 사전 시험점수 평균에 일치하도록 조정**

공변인의 선택

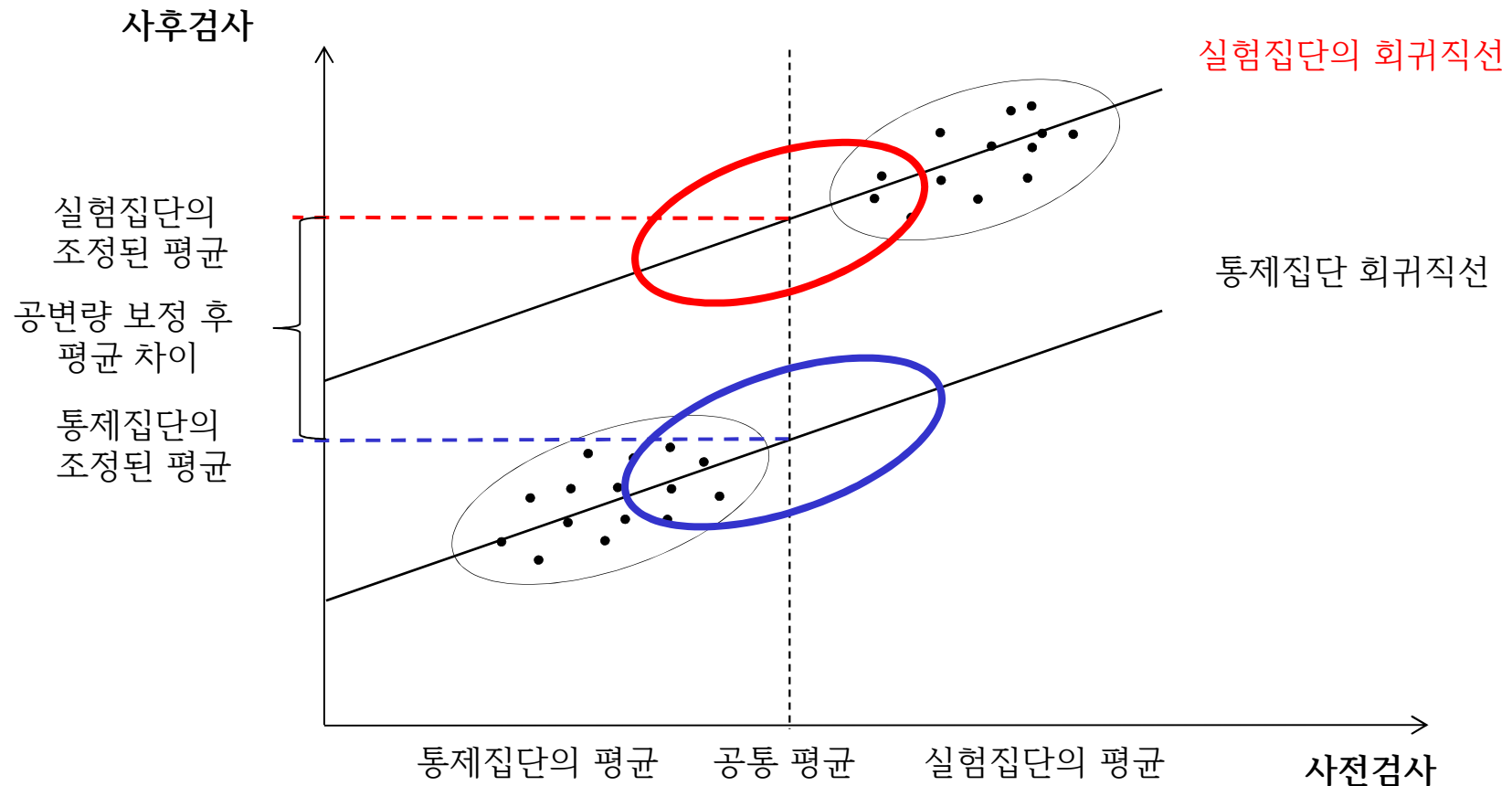
- 어떤 변수를 공변인(Covariate)으로 할 것인가?
 - 종속변수의 차이에 영향을 주는 변수
 - 연속형 변수
 - 대표적인 공분산 변수는 사전 시험결과
 - 사전 시험결과와 사후 시험결과는 상당한 수준의 상관관계가 존재
 - 사전 시험결과가 우수한 그룹이 사후시험결과도 우수할 가능성이 높음

ANCOVA의 원리 - 그림을 통해(1/2)



Question : 사전 검사의 차이가 사후 검사의 차이에 영향을 주지 않았을까?
사전 검사의 차이를 보정하면 어떻게 달라질까?

ANCOVA의 원리 - 그림을 통해(2/2)



Question : 공변량을 보정한 후에도 차이가 있다고 볼 수 있는가?

ANCOVA의 원리 - 회귀분석을 통해

- Raw data

통제집단=0
실험집단=1

사전점수

사후점수

	ID	Exp_Con	Pre	Post
1	1	0	18	12
2	2	0	20	13
3	3	0	14	8
4	4	0	19	12
5	5	0	16	10
6	6	0	15	8
7	7	0	17	11
8	8	0	16	9
9	9	1	23	20
10	10	1	26	24
11	11	1	19	16
12	12	1	22	20
13	13	1	24	21
14	14	1	28	26
15	15	1	20	17
16	16	1	26	24

회귀분석 (집단구분 변수만 사용)

- 집단간 평균차이 검정을 위해 Post를 DV, Exp_Con을 IV로 하는 회귀분석 시행

분산분석^b

모형	제곱합	자유도	평균 제곱	F	유의확률
1 회귀 모형	451.563	1	451.563	56.508	.000 ^a
잔차	111.875	14	7.991		
합계	563.438	15			

a. 예측값: (상수), Exp_Con

b. 종속변수: Post

계수^a

모형	비표준화 계수		표준화 계수	t	유의확률
	B	표준오차	베타		
1 (상수)	10.375	.999		10.381	.000
Exp_Con	10.625	1.413	.895	7.517	.000

a. 종속변수: Post

회귀식: $Post = 10.735 + 10.626 \times Exp_Con$

보정 전 평균 차이: 10.625

회귀분석 (집단구분 변수, 사전 점수사용)

- 사전점수의 효과를 검정하기 위해 Post를 DV, Exp_Con와 Pre를 IV로 하는 회귀분석 시행

분산분석^b

모형	제공합	자유도	평균 제공	F	유의확률
1 회귀 모형	559.750	2	279.875	986.613	.000 ^a
잔차	3.688	13	.284		
합계	563.438	15			

a. 예측값: (상수), Pre, Exp_Con

b. 종속변수: Post

계수^a

모형	비표준화 계수		표준화 계수	t	유의확률
	B	표준오차	베타		
1 (상수)	-7.458	.932		-7.999	.000
Exp_Con	3.624	.447	.305	8.115	.000
Pre	1.057	.054	.735	19.529	.000

a. 종속변수: Post

회귀식: $Post = -7.458 + 3.624 \times Exp_Con + 1.057 \times Pre$

보정 후 평균 차이: 3.624

회귀분석의 비교

- 공변량 보정 전

회귀식: $Post = 10.735 + 10.626 \times Exp_Con$

보정 전 평균 차이: 10.625

- 공변량 보정 후

회귀식: $Post = -7.458 + 3.624 \times Exp_Con + 1.057 \times Pre$

보정 후 평균 차이: 3.624

- 통제집단 회귀식 ($Exp_Con=0$)

회귀식: $Post = -7.458 + 1.057 \times Pre$

- 실험집단 회귀식 ($Exp_Con=1$)

회귀식: $Post = -7.458 + 3.624 + 1.057 \times Pre$

- 공변량 분석은 0, 1로 코딩된 집단 변수와 공변인을 IV로 하는 다중 회귀분석

조정된 평균의 계산

- 조정된 평균 공식

$$\bar{Y}_{adj_i} = \bar{Y}_i - \beta(\bar{X}_i - \bar{X})$$

\bar{Y}_{adj_i} 는 집단 i 의 조정 후 평균, \bar{Y}_i 는 집단 i 의 조정 전 평균

β 는 회귀선의 기울기

\bar{X}_i 는 집단 i 의 공변인 평균, \bar{X} 는 공변인 전체 평균

$$\bar{Y}_{adj_0} = \bar{Y}_0 - \beta(\bar{X}_0 - \bar{X}) = 10.375 - 1.057(16.875 - 20.188) = 13.876$$

$$\bar{Y}_{adj_1} = \bar{Y}_1 - \beta(\bar{X}_1 - \bar{X}) = 21.000 - 1.057(23.500 - 20.188) = 17.499$$

$$\bar{Y}_{adj_1} - \bar{Y}_{adj_0} = 3.623$$

집단간 평균 차이의 통계적 검정(1/2)

- 차이에 대한 검정 통계
 - 공변량분석 모형의 결정계수, 공변인 효과 모형의 결정계수를 이용
 - 공변량분석 모형: 모든 공변인과 IV를 포함하는 모델
 - $\hat{Y} = \beta_0 + \beta_1 \times \text{Exp_Con} + \beta_2 \times \text{Pre}$, 결정계수 $R^2_f = 0.993$
 - 공변인 효과 모형: 공변인만을 포함하는 모델 (공변인의 설명력)
 - $\hat{Y} = \beta_0 + \beta_2 \times \text{Pre}$, 결정계수 $R^2_c = 0.960$

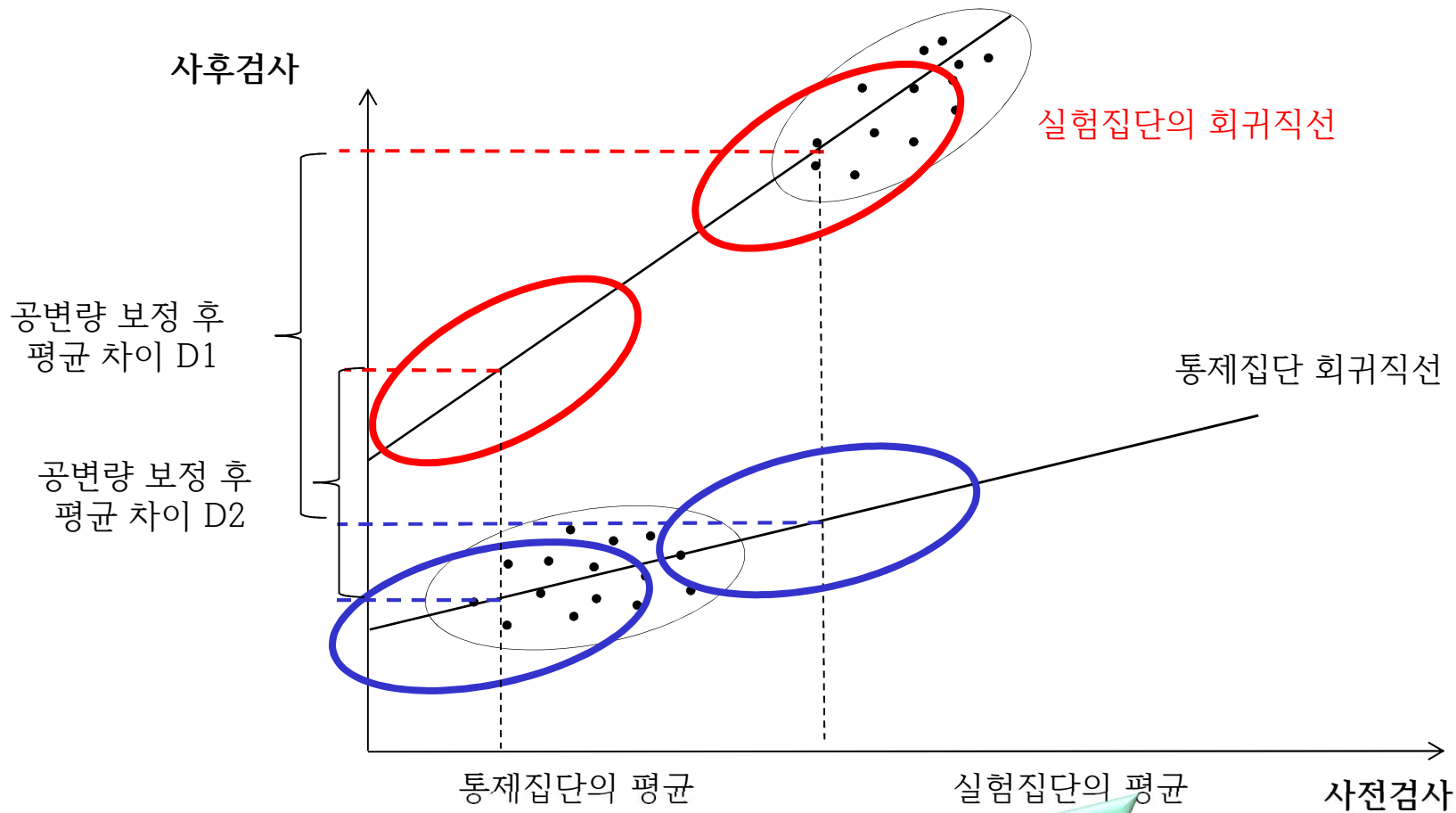
집단간 평균 차이의 통계적 검정(2/2)

- 공변량 분석표를 작성(집단 구분 변수의 유의성 검정)
 - ANOVA Table과 유사
- **SSAT(Sum of Square Adjusted Total)=(1- R²_c) SST**
- **SSAB(Sum of Square Adjusted Between Group)= (R²_f- R²_c) SST**
- **SSAW(Sum of Square Adjusted Within Group)= =(1- R²_f) SST**
- **SST=Σ(Y-Ȳ)²**

SS	D.F	SS/D.F	F	P-value
18.679(SSAB)	g-1	18.679(MSAB)	65.847 (MSAB/MSAW)	0.000
3.688(SSAW)	13 (N-g-C)	0.284(MSAW)		
22.536(SSAT)	14 (N-1-C)	1.610(MSAT)		

N: 사례 수, g: 집단의 수, C: 공변인의 수

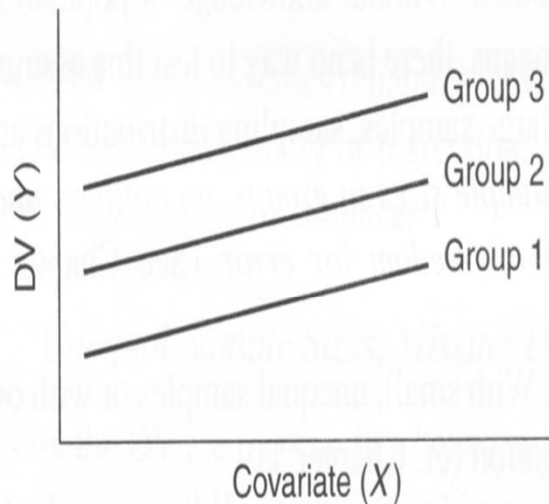
만일 회귀식의 기울기가 다르다면?



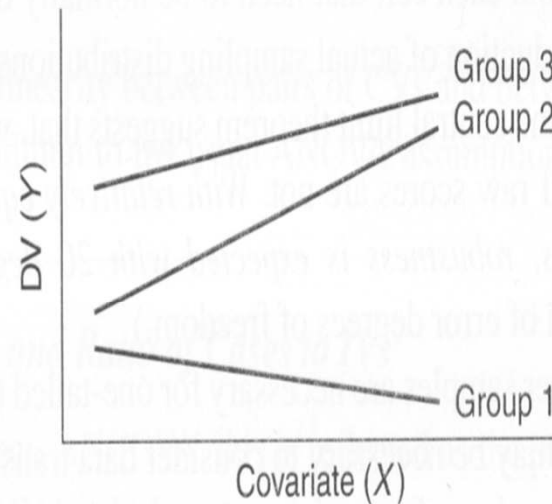
기울기가 다르면 기준에 따라
공변량 보정 후 평균이 다름
→ 차이 여부의 판단 결과가 달라질 수 있음

어떻게 기울기가 다른지 확인할 것인가?

- ANOVA 에서 반복이 있는 분산분석의 경우를 상기
 - IV간 상호작용이 있으면 그래프의 기울기가 달라짐
 - IV와 Covariate과 상호작용항을 회귀에 추가
 - 상호작용항이 유의하면, 그래프의 기울기가 다름
 - 공변량분석이 불가능 → 대응 : 공변량분석 절차 페이지 참고

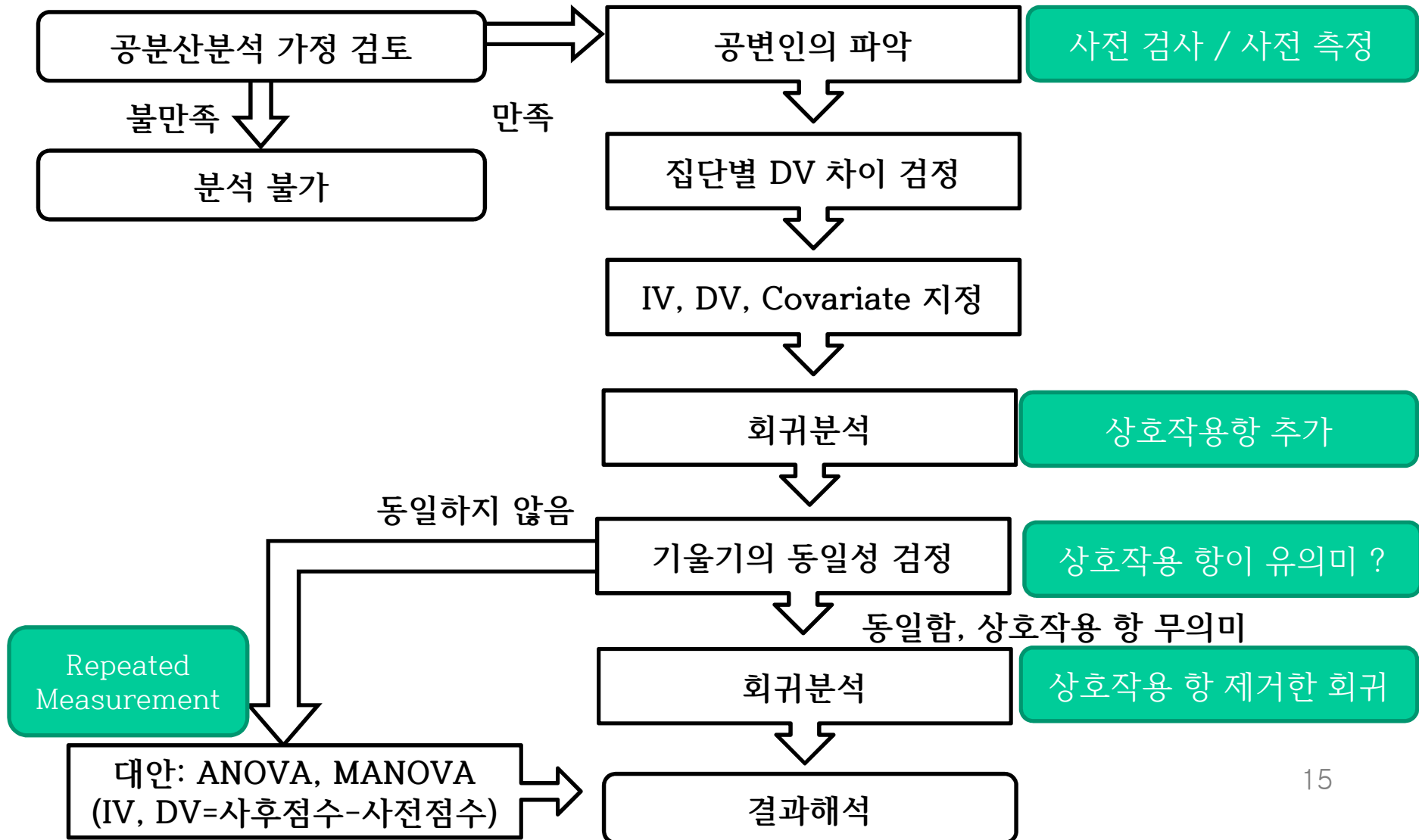


(a) Homogeneity of regression (slopes)



(b) Heterogeneity of regression (slopes)

공변량 분석절차



공변량분석의 가정

1. 표본의 정규성 → 확인이 어려움
 - Error(오차)의 D.F가 20이상 권장
 - Normal Probability Plot (45° line과 유사)
2. 변량-공변량 행렬의 동질성
 - Box's M, Ho : 공분산행렬이 동일하다.
3. DV 간의 선형성(MANCOVA): 종속변수간 일정한 선형관계 유지
 - Bartlett's test of Sphericity, Ho : 종속변수간의 상관행렬이 단위행렬이다.
4. 회귀계수의 동질성
 - 공변인과 DV의 상호작용 유의성 확인
 - 권장사항
 - 집단간 사례의 수를 동일하게
 - 집단간 사례수가 다르면 그룹평균의 산술평균이 전체평균이 되지 않음
 - Sum of Square 를 계산하는 방법이 달라져야 함 (Type III → Type I?)

실습1 - (T-test와 ANCOVA 비교)

- 새로운 학습기법의 적용효과를 알기 위해 80명의 학생을 두 집단으로 나누어 사전 시험을 치르고 40명에게 새로운 학습기법을 적용시켰다. 학습이후 80명에게 사후 시험을 치르게 하였다.
 - 학습.sav를 이용하여 다음 물음에 답하세요.
 - 귀무가설은?
 - 두 집단의 사후 시험결과에 차이가 있는가?
 - 새로운 학습기법의 효과가 있는가?

실습1 - 과정 및 결과 (1/7)

The screenshot shows the SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Analyze > Compare Means > Independent-Samples T Test' is highlighted. The 'Independent-Samples T Test' dialog box is open, showing the following configuration:

- Dependent Variable(s): 사전점수 [pretest]
- Test Variable(s): 사후점수 [posttest]
- Grouping Variable: group(0 1)
- Buttons: 확인 (OK), 불여넣기(Paste), 재설정(Restore Defaults), 취소(Cancel), 도움말(Help)

	group	pretest	post
1	0	8	
2	0	11	
3	0	10	
4	0	9	
5	0	13	
6	0	8	

실습1 - 과정 및 결과 (2/7)

집단통계량

집단 구분	N	평균	표준편차	평균의 표준오차
사후점수 대조그룹	40	12.48	3.088	.488
처리그룹	40	13.15	3.191	.505

독립표본 검정

	Levene의 등분산 검정		평균의 동일성에 대한 t-검정						
	F	유의확률	t	자유도	유의확률 (양쪽)	평균차	차이의 표준오차	차이의 95% 신뢰구간	
								하한	상한
사후점수 등분산이 가정됨	.040	.841	-.961	78	.339	-.675	.702	-2.073	.723
사후점수 등분산이 가정되지 않음			-.961	77.917	.339	-.675	.702	-2.073	.723

$$H_0: \sigma_0 - \sigma_1 = 0$$

$$H_1: \sigma_0 - \sigma_1 \neq 0$$

$$H_0: \mu_0 - \mu_1 = 0$$

$$H_1: \mu_0 - \mu_1 \neq 0$$

결론 : 두 집단의 사후점수가 동일하다는 귀무가설을 기각할 수 없음

의문: 진짜 그럴까? 사전점수 차이로 인해 그렇게 결론 난 것은 아닐까?

실습1 - 과정 및 결과 (3/7)

- 공변량 분석을 통해 사전 점수를 보정

The screenshot displays the PASW Statistics Data Editor interface. The 'Analyze' menu is open, with 'Covariance Analysis' selected. The 'Covariance Analysis' dialog box is also open, showing the following settings:

- 종속변수(D): 사후점수 [posttest] (highlighted with a red box)
- 모수요인(F): 집단 구분 [group] (highlighted with a red box)
- 공변량(C): 사전점수 [pretest] (highlighted with a red box)

The 'Analyze' menu options include: 보고서(P), 기술통계량(E), 표, 평균 비교(M), 일반선형모형(G) (highlighted with a red box), 일반화 선형 모형(Z), 혼합 모형(X), 상관분석(C), 회귀분석(R). The '일반선형모형(G)' sub-menu is open, showing: 일반량(U)... (highlighted with a red box), 다변량(M)..., 반복측정(R)..., 분산성분(V)...

실습1 - 과정 및 결과 (4/7)

일변량: 모형

모형 설정
○ 완전요인모형(A) ● 사용자 정의(C)

요인 및 공변량(F):
group
pretest

모형(M):
group
pretest
group*pretest

항 설정
유형(Y): 상호작용

제공함(O): 제 III 유형

모형에 절편 포함(I)

계속 취소 도움말

일변량: 옵션

주변평균 추정

요인 및 요인 상호작용(F):
(전체)
group

평균 출력 기준(M):
group

주효과 비교(C)

신뢰구간 조정
95% (지정됨)

표시

기술통계량(S)
 효과크기 추정값(E)
 관측 검정력(B)
 모수 추정값(T)
 대비 계수 행렬(C)

동질성 검정(H)
 평균-산포 도표(P)
 잔차도표(R)
 적합결여 검정(L)
 일반 추정가능 함수(G)

유의수준(V): .05 신뢰구간: 95.0 %의 케이스 추출

계속 취소 도움말

집단에 따른 평균 출력

Box's M (변량-공변량 행렬의 동질성 검정) 출력

Partial eta Squared 출력

실습1 - 과정 및 결과 (5/7)

개체-간 효과 검정

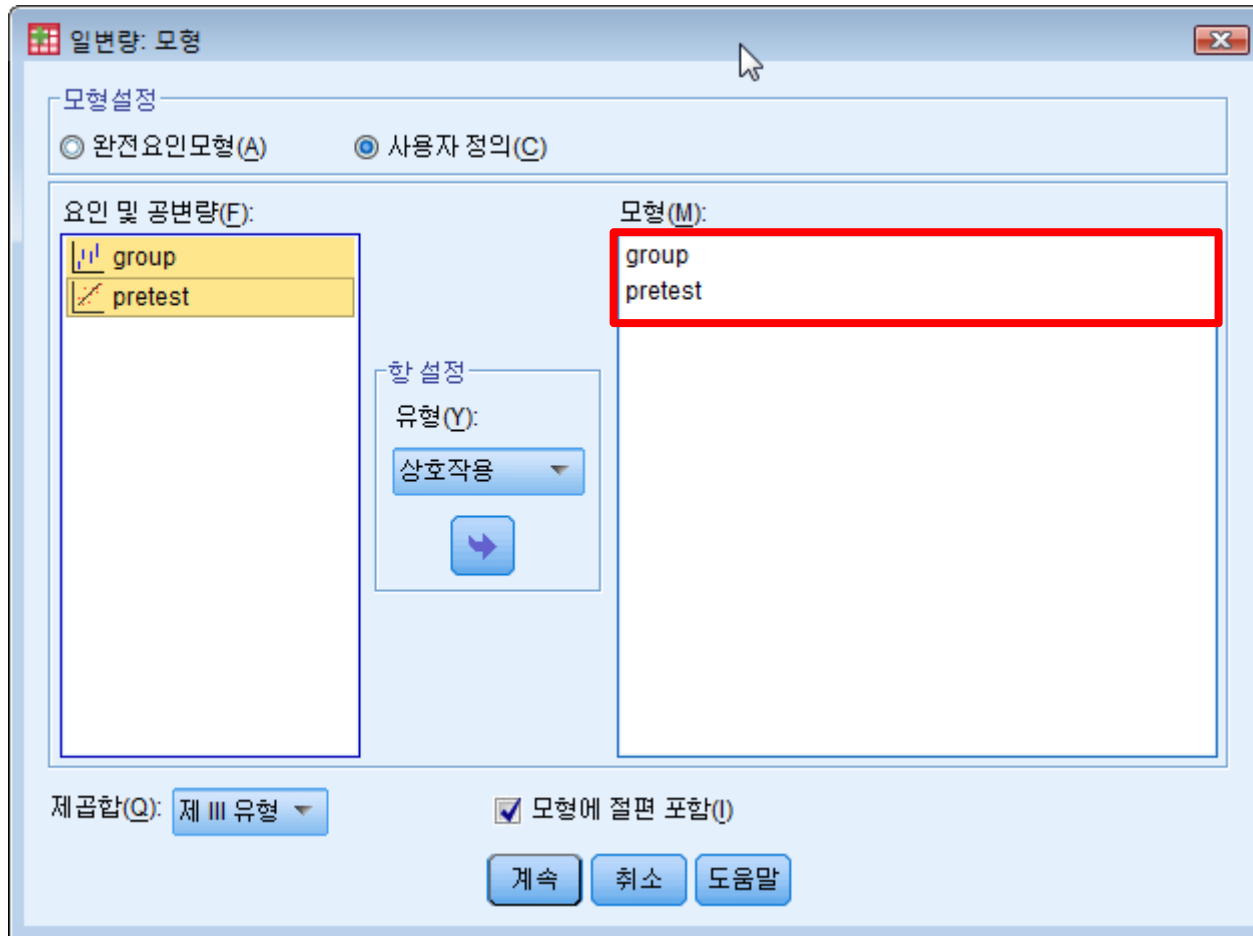
종속 변수: 사후점수

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률	부분 에타 제곱
수정 모형	437.415 ^a	3	145.805	32.518	.000	.562
절편	483.442	1	483.442	107.819	.000	.587
group	26.940	1	26.940	6.008	.017	.073
pretest	413.470	1	413.470	92.213	.000	.548
group * pretest	10.092	1	10.092	2.251	.138	.029
오차	340.772	76	4.484			
합계	13911.000	80				
수정 합계	778.188	79				

a. R 제곱 = .562 (수정된 R 제곱 = .545)

H_0 : 상호작용 효과 = 0 H_0 : 두 회귀직선의 기울기가 동일하다.
 H_1 : 상호작용 효과 \neq 0 H_1 : 두 회귀직선의 기울기가 동일하지 않다.

실습1 - 과정 및 결과 (6/7)



실습1 - 과정 및 결과 (7/7)

개체-간 효과 검정

종속 변수: 사후점수

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률	부분 에타 제곱
수정 모형	427.323 ^a	2	213.661	46.6	.000	.549
절편	487.286	1	487.286	106.939	.000	.581
group	37.680	1	37.680	8.269	.005	.097
pretest	418.210	1	418.210	91.780	.000	.544
오차	350.865	77	4.557			
합계	13911.000	80				
수정 합계	778.188	79				

H₀: 집단간 차이 없다.
H₁: 집단간 차이 있다.

독립변수를 통해 종속
변수를 설명하는 정도

a. R 제곱 = .549 (수정된 R 제곱 = .537)

결론 : 두 집단의 사후 점수가 동일하다는 귀무가설을 기각할 수 있음

공변량 보정 전 후로 결론이 바뀜. 원인 : 사전점수 차이

실습2 - (ANOVA와 ANCOVA 비교)

- 3개의 회사에 근무중인 직원들의 월급여, 교육년수, 근무년수를 이용하여 회사에 따른 월급여의 차이가 있는지 알아보려고 한다.
- Salary.sav를 이용하여
 1. 회사에 따른 월급여의 차이가 있는지 분산분석을 실시하세요.
 - (IV =회사, DV=월급여)
 - 귀무가설은?
 - 회사별 월급여의 차이가 있다고 볼 수 있는지?
 2. 회사별 평균 교육년수와 평균 근무년수를 계산하세요.
 3. 교육년수와 근무년수를 보정하였을 때 회사에 따른 월급여의 차이가 있는지 분석하세요. (분산분석 x, 선형회귀 o)
 - (IV =회사, Covariate = 교육년수, 근무년수, DV=월급여)

실습2 - 과정 및 결과 (1/7)

- ANOVA



개체-간 효과 검정

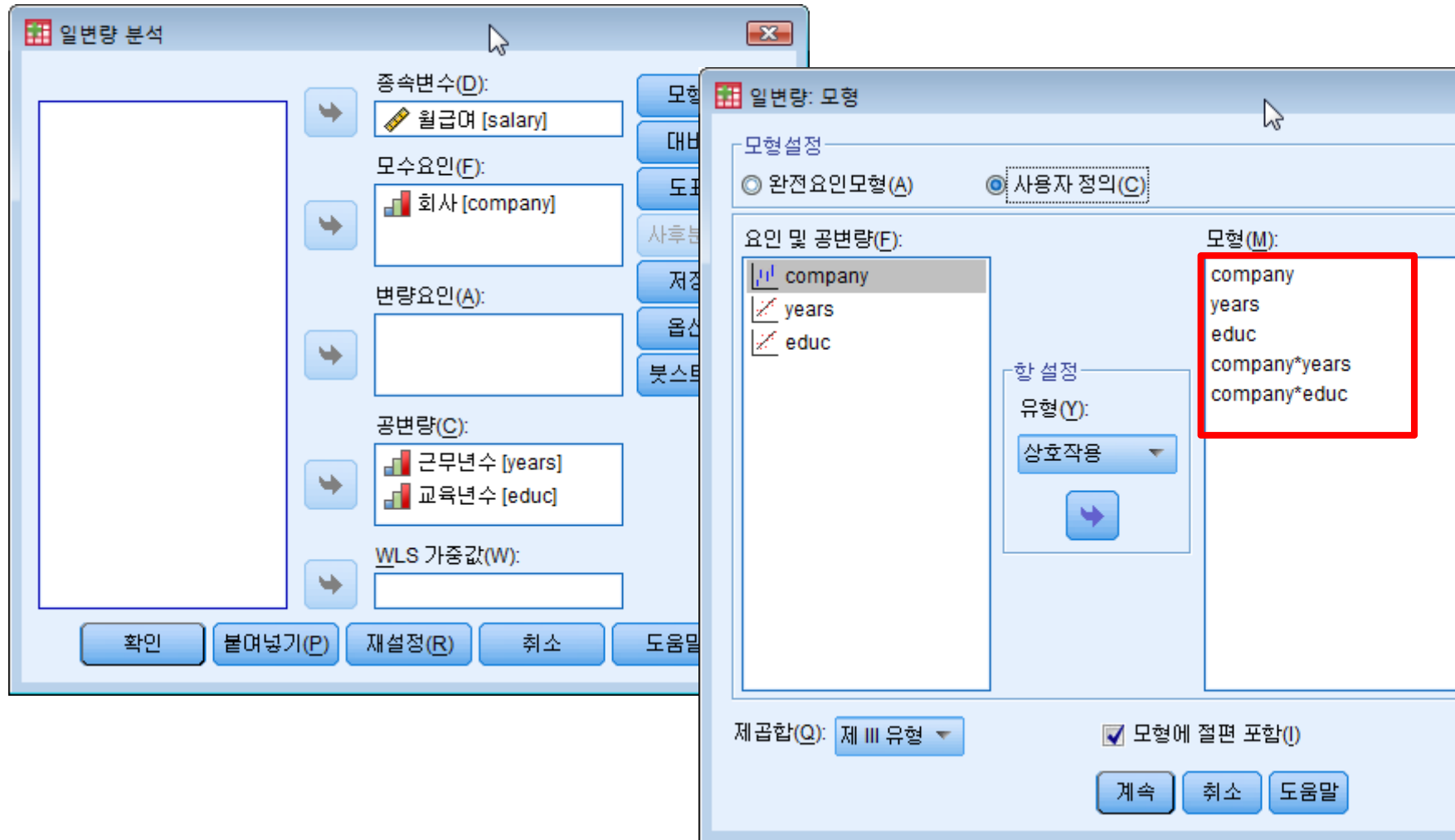
종속 변수: 월급여

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률	부분 에타 제곱
수정 모형	791.255 ^a	2	395.628	4.249	.020	.153
절편	2538916.637	1	2538916.637	27269.829	.000	.998
company	791.255	2	395.628	4.249	.020	.153
오차	4375.865	47	93.104			
합계	2564394.000	50				
수정 합계	5167.120	49				

a. R 제곱 = .153 (수정된 R 제곱 = .117)

실습2 - 과정 및 결과 (2/7)

- ANCOVA



실습2 - 과정 및 결과 (3/7)

개체-간 효과 검정

종속 변수:월급여

소스	제 III 유형 제공합	자유도	평균 제공	F	유의확률	부분 에타 제곱
수정 모형	2231.496 ^a	8	278.937	3.896	.002	.432
절편	34964.185	1	34964.185	488.323	.000	.923
company	17.753	2	8.877	.124	.884	.006
years	112.861	1	112.861	1.576	.216	.006
educ	812.480	1	812.480	11.347	.002	.217
company * years	60.519	2	30.259	.423	.658	.020
company * educ	38.282	2	19.141	.267	.767	.013
오차	2935.624	41	71.601			
합계	2564394.000	50				
수정 합계	5167.120	49				

상호작용 없음
회귀계수는 두 공변인에 대해
동일함

a. R 제곱 = .432 (수정된 R 제곱 = .321)

실습2 - 과정 및 결과 (4/7)

개체-간 효과 검정

종속 변수:월급여

소스	제 III 유형 제공합	자유도	평균 제공	F	유의확률	부분 에타 제곱
수정 모형	2231.496 ^a	8	278.937	3.896	.002	.432
절편	34964.185	1	34964.185	488.323	.000	.923
company	17.753	2	8.877	.124	.884	.006
years	112.861	1	112.861	1.576	.216	.006
educ	812.480	1	812.480	11.347	.002	.217
company * years	60.519	2	30.259	.423	.658	.020
company * educ	38.282	2	19.141	.267	.767	.013
오차	2935.624	41	71.601			
합계	2564394.000	50				
수정 합계	5167.120	49				

상호작용 없음
회귀계수는 두 공변인에 대해
동일함

a. R 제곱 = .432 (수정된 R 제곱 = .321)

실습2 - 과정 및 결과 (5/7)

일변량: 모형

모형설정

완전요인모형(A) 사용자 정의(C)

요인 및 공변량(F):

- company
- years
- educ

모형(M):

- company
- years
- educ

항 설정

유형(Y): 상호작용

제공합(Q): 제 III 유형 모형에 절편 포함(I)

계속 취소 도움말

일변량: 옵션

주변평균 추정

요인 및 요인 상호작용(F):

- (전체)
- company

평균 출력 기준(M):

- company

주효과 비교(C)

신뢰구간 조정(N): LSD(지정없음)

표시

- 기술통계량(D)
- 동질성 검정(H)
- 효과크기 추정값(E)
- 평균-산포 도표(P)
- 관측 검정력(B)
- 잔차도표(R)
- 모수 추정값(T)
- 적합결여 검정(L)
- 대비 계수 행렬(C)
- 일반 추정가능 함수(G)

유의수준(Y): .05 신뢰구간: 95.0%의 케이스 추출

계속 취소 도움말

실습2 - 과정 및 결과 (6/7)

개체-간 효과 검정

종속 변수:월급여

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률
수정 모형	2128.007 ^a	4	532.002	7.877	.000
절편	37623.010	1	37623.010	557.082	.000
company	468.208	2	234.104	3.466	.040
years	394.036	1	394.036	5.834	.020
educ	801.934	1	801.934	11.874	.001
오차	3039.113	45	67.536		
합계	2564394.000	50			
수정 합계					

회사에 따라 급여 차이 있음

회사1과 회사2는 차이 없으나
나머지 경우는 차이 있음

대응별 비교

공변인으로 보정함이 적절함

종속 변수:월급여

(I) 회사	(J) 회사	평균차(I-J)	표준오차	유의확률 ^a	차이에 대한 95% 신뢰구간 ^a	
					하한값	상한값
회사1	회사2	.264	2.933	.929	-5.644	6.172
	회사3	-6.322*	2.938	.037	-12.239	-.405
회사2	회사1	-.264	2.933	.929	-6.172	5.644
	회사3	-6.586*	2.791	.023	-12.207	-.964
회사3	회사1	6.322*	2.938	.037	.405	12.239
	회사2	6.586*	2.791	.023	.964	12.207

추정된 주변평균을 기준으로

a. 다중비교에 대한 조정: 최소유의차 (조정하지 않은 상태와 동일합니다.)

*. 평균차는 .05 수준에서 유의합니다.