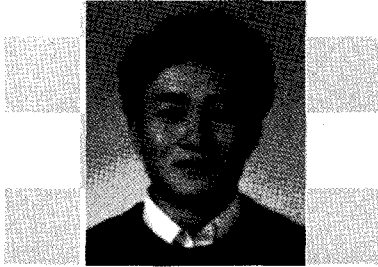


의학연구에서 실험계획의 원리와 교차계획법



박 선 일

서울시 관악구 신림9동 산 56-1

서울대학교 수의과대학 내과학교실

H.P: 016-208-2944

4. 실험 계획법

실험오차를 관리하는 두 가지 방법으로서 랜덤화의 개념을 이용한 완전랜덤화 계획법과 랜덤화와 함께 블록의 개념을 이용한 랜덤화 블록 계획법이 있다. 이러한 랜덤화 방법들과 여러 가지 처리의 형태에 따라 다양한 실험계획법이 만들어지는데 예를 들면, 두 요인 A와 B가 있을 때 요인 A의 모든 수준과 요인 B의 모든 수준의 조합에 대하여 실험하는 요인배치법 (factorial design), A인자의 주효과와 일차단위간의 차이가 교락되어 있고 이단계 랜덤화 (two-stage randomization)가 적용된 분할구법 (split-plot design), 실험횟수를 늘리지 않고 실험배치를 몇 개의 블록으로 나누어 배치시킴으로써 동일환경내의 실험횟수를 적게

하는 배치법인 교락법(confounding method), 불필요한 교호작용이나 고차의 교호작용을 구하지 않고 실험의 크기를 작게 할 수 있도록 인자의 조합 중에서 일부만을 실험하는 방법을 일부실험법 (fractional factorial design, fractional replication) 및 하나의 블록에 모든 처리가 완전히 포함되지 못하는 불완전블록에 의한 실험계획인 불완전 블록계획법 (incomplete block design) 등이 있다. 실험배치 측면에서는 일원배치 (one-way layout)와 이원배치(two-way layout) 등으로 구분하기도 하며 분석 측면에서 반복측정 (repeated measures)이라는 용어를 사용한다. 이러한 다양한 실험계획법들은 연구의 목적과 상황에 따라 적절하게 선택해야된다. 실험내용을 정확하게 계획하고 합리적으로 수

행하는 접근방법을 실험전략 (strategy of experiment)이라고 한다. 어떤 운동선수가 골프실력을 향상시키기 위한 전략을 수립한다고 할 때 타수를 줄이는데 관련되리라고 생각되는 요인들을 나열한 결과 다음과 같다고 하자 (1).

- 첫째, 사용하는 드라이버의 종류 (a, b)
- 둘째, 사용하는 공의 종류 (c, d)
- 셋째, 골프 클럽을 운반하는 형태 (e, f)
- 넷째, 게임을 하면서 마시는 음료의 종류 (g, h)

이 운동선수가 선택할 수 있는 실험전략에 대하여 생각하여 보자. 가장 간단한 방법은 4가지 요인 중 한가지 요인을 선택하여 게임에 임하여 점수의 변화를 관찰하는 것인데 이는 한번에 한가지 요인을 선택하였기 때문에 one-factor-at-a-time approach가 된다.

이 방법은 각 요인의 기저수준 (baseline level)을 선택한 후 그 수준의 해당 범위까지 변화시키는 과정으로 구성되는데 위의 4가지 요인에 대하여 그림 1과 같은 결과를 얻었다고 하자.

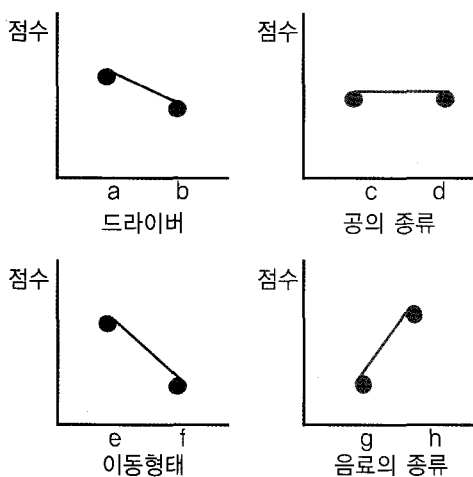


그림 1. 골프실험에서 one-factor-at-a-time approach의 결과

예를 들어 이동형태의 경우 이동수준을 f로 할 때 기록이 향상된다는 것을 의미한다. 이러한 방법은 4가지 요인에 대한 최적의 수준을 결정하는데 적합하지만 골프 점수가 어느 한가지 요인에 의해서만 결정되는것이 아니라고 할 때 여러 가지 요인들의 조합에 따른 최적수준은 알 수 없다. 즉 상호작용을 파악하지는 못한다는 것이다. 상호작용이란 어떤 한 요인이 다른 요인의 다른 수준에서 동일한 효과를 보이지 않는 경우를 의미한다. 예컨대, 드라이버의 종류와 음료의 종류 간 상호작용이 그림 2와 같다고 하자.

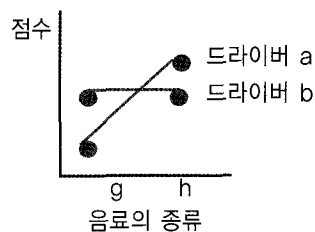


그림 2. 드라이버와 음료의 종류 간 상호작용

그림에서 보듯이 드라이버 b를 사용한다면 섭취한 음료의 종류는 골프점수에 거의 영향이 없는 반면, 드라이버 a를 사용할 경우 h음료를 섭취함으로써 훨씬 좋은 기록을 얻게 된다는 것을 알 수 있다. 이러한 one-factor-at-a-time 식의 접근은 실험계획에서 효율성이 가장 낮은 계획임에도 많은 연구자들은 이 방법을 흔히 사용한다. 두 번째 접근 방법은 경험이나 문헌고찰을 통하여 임의의 요인들을 선택하여 실험하는 것이다. 예를 들어 드라이버의 종류 a, 사용하는 공의 형태 c, 골프 클럽을 운반하는 형태 e, 게임을 하면서 마시는 음료의 종류 g를 선택하여 게임에 임하여 87점을 얻었다고 하자. 계

수 의 학 강 좌

입을 진행하면서 운동선수가 드라이버 종류 a를 사용할 때 자주 실수하는 것을 발견하였다고 하면 나머지 3개 요인은 그대로 두고 드라이버의 종류만 b로 바꿔 게임에 임하여 이전의 결과와 비교할 수 있을 것이다. 다음에는 음료의 종류에 대하여 위와 동일한 방법으로 수준을 바꾸어 가면서 게임의 결과를 비교하는 등의 과정을 무한히 반복할 수 있는데 이를 best-guess approach라고 한다. 그러나 이 방법은 두 가지 문제를 갖고 있다. 첫째, best-guess가 언제나 바람직한 결과를 낳는 것은 아니다. 즉 운동선수가 매번 best-guess라고 생각되는 요인들을 선택하여 무한히 반복적으로 게임에 임한다고 하여도 점수가 좋아지리라는 결과를 기대할 수 있는 것은 아니라는 것이다. 두 번째로 최초의 best-guess가 만족할만한 결과를 보인다고 할 때 다른 best-guess가 될 수 있는 방법들에 대해서는 검정하지 않으려 할 수 있다. 세번째 접근방법은 요인실험을 수행하는 것인데 이는 요인들에 대하여 한번에 하나씩 시행하는 것이 아니라 요인들을 동시에 변화시키는 방법이다. 예를 들어 드라이버의 종류와 공의 형태 두 가지 요인에 관심이 있다고 하자. 그림 3은 골프점수에 대한 이들 두 가지 요인의 합동효과를 나타내고 있다.

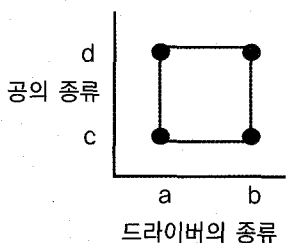


그림 3. 드라이버와 공의 종류를 이용한 2 요인 실험

이 실험은 두 가지 요인이 두 가지 수준을 가지고 있으며 두 요인의 모든 가능한 조합을 실험한다는 특징이 있다. 기하학적으로 볼 때 4회의 실험은 사각형을 이루는데 이러한 형태의 실험을 2²요인실험이라고 한다. 운동선수가 총 8라운드의 경기를 할 수 있다고 하면 각 요인 수준의 조합에 대해 2라운드씩 시행할 수 있는데 이를 실험계획에서 반복 (replication)이라고 한다. 이러한 실험계획은 연구자로 하여금 각 요인의 개별효과 (주효과, main effect)와 상호작용 효과를 파악할 수 있도록 해준다. 만일 드라이버, 공 및 음료의 종류 등 세 가지 요인에 관심이 있다고 하자. 이들 세 요인은 각각 2개의 수준이 있으므로 그림 4와 같이 도식화된다. 여기에서는 각각 2개의 수준을 갖는 3가지 요인에 대하여 총 8개의 조합이 만들어지는데 이를 2³요인 실험이라고 한다. 운동선수가 총 8회의 게임을 할 수 있다면 각 조합에 대하여 1회의 실험을 할 수 있는 셈이다. 한편, 위에서 나열한 4가지 요인 모두에 대하여 관심이 있다고 하면 2⁴요인 실험이 된다. 일반적으로 각각의 수준이 2인 k개의 요인이 있다면 총 2^k개의 실험횟수가 요구되며, 요인의 개수가 증가할수록 시행의 횟수는 급격히 증가한다. 2⁴의 경우 총 16회의 실험이 필요하고, 10개의 요인이 있다면 총 1024개의 실험이 필요하다.

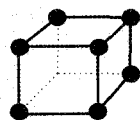


그림 4. 드라이버, 공 및 음료의 종류 등 세 가지 요인에 대한 3요인 실험

이 예에서는 총 16회의 실험이 필요한데 운동 선수는 8라운드 밖에는 게임을 할 수 없으므로 모든 요인의 조합에 대하여 실험을 수행할 수 없다. 어떻게 할 것인가? 다행히도 부분요인 실험으로 가능한데 이는 요인의 조합 중에서 그 일부에 대하여 시행하는 것으로 요인 실험의 변형이라고 생각하면 된다. 이를 도식화하면 그림 5와 같은데 이 경우 총 16회 중 8회만을 시행하므로 one-half fraction이라고 한다.

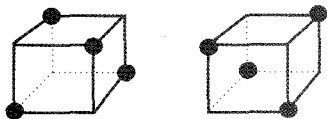


그림 5. 드라이버, 공, 음료 및 이동형태에 대한 4요인 부분실험의 예

이상과 같이 실험계획을 구성하는 전략에 대하여 간단히 살펴보았다. 연구내용과 목적이 동일하다고 하더라도 연구상황에 따라 실험계획은 얼마든지 바뀔 수 있으므로 실험전략을 충분히 검토하여 계획하는 것이 바람직하다.

다음으로 의학연구에서 많이 사용되는 교차계획법에 대하여 소개하고자 한다. 교차계획법(crossover 또는 changeover design)에 의한 자료는 반복측정 자료이면서도 반복측정 자료와는 서로 다른 점이 있다. 반복측정 자료는 처리 또는 치료가 모든 개체에게 고정된 순서(fixed order)로 적용되는 반면에, 교차계획법은 치료의 순서가 랜덤하게 적용되는 차이가 있다 (2). 가장 단순한 2x2 교차계획법은 동일한 개체에게 두 가지 치료, 즉 A와 B 치료를 모두 적용하게 되며 두 치료의 효과를 비교할 때 치료효과가

치료의 순서에 의해 달라질 수 있다는 것을 감안하여 연구대상자의 절반의 개체들에게 치료 A를 먼저 적용한 후에 치료 B를 적용하고 나머지 절반의 개체들에게는 치료 B를 적용한 후에 치료 A를 적용한다. 이 때 어떤 개체가 치료순서 AB군에 속하고 어떤 개체가 치료순서 BA군에 속하게 될지를 랜덤하게 결정한다. 이와 같이 서로 다른 처리방법을 동일한 개체에게 교차시켜 적용하는 실험계획법을 교차계획법이라 한다.

교차계획법의 자료분석

먼저 전형적인 교차계획에 의한 실험의 예를 들어보자. 말과 개에서 근골격계 질환에 사용되는 NSAID (non-steroidal anti-inflammatory drug)인 phenylbutazone을 투여할 경우 혈청 T₄수준이 저하되어 갑상선 질환의 감별에 문제가 있는 것으로 보고되었다. 따라서 연구자는 이를 평가하고자 10두의 말을 랜덤하게 두 군으로 할당한 후 첫 번째 군에는 0.9% NaCl을 투여하고 40일 후 phenylbutazone을 투여하여 각 시점에서 경정맥으로부터 혈액을 채취하여 T₄농도를 측정하였다.

다른 군에는 phenylbutazone을 먼저 투여한 후 0.9% NaCl을 투여하여 동일한 방법으로 T₄ 농도를 측정하였다.

(3) 다른 예로 건강한 말에서 laxative로 사용되는 기존의 안전한 mineral oil과 linseed oil의 독성을 비교하고자 6두의 말을 두 군으로 나누어 첫째 군은 linseed oil을 비간 튜브를 이용하여 12시간 간격으로 kg당 2.5ml을 투여하고 14일 후 mineral oil을 동일한 방법으로 kg당 10ml을

투여하였다. 둘째군은 mineral oil을 먼저 투여한 후에 linseed oil을 투여하였다 (4) 두 예제 모두 두 가지 치료법을 두 시기에 걸쳐 시행한 것이므로 이를 two-treatment, two-period 교차시험이라고 한다. 흔히 교차시험은 균형잡힌 (balanced) 일련의 라틴방격으로 계획하는 경우가 많다 (5-7).

분석시 고려사항

교차시험에서 고려해야할 사항은 첫째, 이월효과 [carry-over effect, 잔류효과 (residual effect), 수반효과 (sequence effect)]로서 이는 연속적인 두 치료 기간에서 시기 1에서의 치료 A의 영향이 시기 2에까지 미치고, 시기 1에서의 치료 B의 영향이 시기 2에까지 미치는 것을 뜻하며, 이 두 영향의 크기가 서로 다를 때 이월효과가 있다고 말한다. 만일 두 효과의 크기가 동일한 경우 단순한 2x2 교차계획법에서 이월효과는 시기효과로부터 구분해낼 수 없다. (8). 두 시기 사이에 첫 번째 치료효과가 충분히 씻어지는 기간 (wash-out period)을 두는 것이 이월효과를 제거하는 한가지 방법이다. 위에서 제시된 두 예제에서는 이월효과를 제거하는 기간이 충분히 고려되었다고 할 수 있다. 둘째, 단순한 2x2 교차계획법의 또 다른 단점은 이월효과와 시기와 치료의 교호작용을 구분할 수 없다는 것이다. 따라서 시기와 치료의 교호작용이 없다고 가정한 후 이월효과만을 고려하여 분석하게 된다. 여기에서 시기와 치료의 교호작용이 뜻하는 바는 시기와 치료의 여러 조합에서 측정된 반응값이 시기효과와 치료효과와 단순한 가산적 (additive) 관계로서 설명될

수 없다는 것이다. 셋째, 시험대상의 일부가 시험이 종료되기 이전에 폐사나 기타 다른 사유로 중단되는 탈락 (drop-out)이 발생할 수 있다. 교차시험에서는 두 군의 수가 동일할 필요는 없지만 탈락이 많을 경우 분석에 문제가 초래될 수 있다.

넷째, 처리방법이 많아지면 순위형 결과 (예, 아주 나쁨, 보통, 양호함)에 대한 정확한 분석방법이 없다. 다섯째, 이월효과가 유의하게 다르다고 밝혀진 경우에는 두 번째 시기에 대한 측정값에는 첫 번째 치료의 이월효과가 함께 섞여 있으므로 (contaminated) 두 번째 시기의 측정값은 분석에 사용할 수 없다. 결과적으로 첫 번째 시기의 독립된 두 군의 측정값만으로 단순히 unpaired t검정하여 치료의 차이를 알아내게 되므로 치료의 이월효과가 있는 경우에는 두 시기의 자료를 모두 사용하는 교차계획법의 장점이 없어지는 결과를 초래한다 (9-11).

자료분석법

교차시험법은 임상시험 (clinical trial)에서 빈번하게 사용되고 있으며 치료순서로 AA와 BB를 모두 포함시킨 교차계획법이나 여러 시점의 측정까지로 연장된 교차계획법으로 확장할 수도 있다. 예를 들어 개에서 기관허탈을 치료하는 aminophylline 약제의 효과를 알아보기 위해 약제 (N)와 placebo (P)로써 총 13두의 환축을 각각 7두와 6두로 두 군으로 랜덤하게 나누어 2x2 교차계획법을 수행하였다고 하자 (표 3). 각 시기는 1주일로 구성되었고 첫 주의 복용이 끝난 후 3주일의 씻어내리는 기간을 가진 후에 두 번째 복용을 시작하였다.

표 3. 교차시험의 계획도

군	복용방법	시기 1	휴약기간	시기 2
1	PN	placebo (n ₁ = 6)	→	aminophylline
2	NP	aminophylline (n ₂ = 7)	→	placebo

표 4. 두 군에서의 분당 최대 호기량 측정 결과

순서	환측번호	시기		시기 차이	계
		시기 I	시기 II		
PN	2	370	385	-15	755
	3	310	400	-90	710
	5	380	410	-30	790
	9	290	320	-30	610
	12	260	340	-80	600
	13	90	220	-130	310
		$\bar{X}_{1,DIF} = -62.5$		$\bar{X}_{1,TOT} = 3775$	
				$\bar{X}_1 = -62.5$	
				$SD_1 = 173.97$	
NP	1	310	270	40	580
	4	310	260	50	570
	6	370	300	70	670
	7	410	390	20	800
	10	250	210	40	460
	11	380	350	30	730
14	330	365	-35	695	
		$\bar{X}_{2,DIF} = 30.71$		$\bar{X}_{2,TOT} = 4505$	
				$\bar{X}_2 = 643.6$	
				$SD_2 = 114.3$	

약제의 효과는 분당 최대 호기량(peak expiratory flow, liters/min)으로 판단하였다.

약제 후에 placebo를 복용한 첫째 군과 placebo 후에 약제를 복용한 둘째 군의 결과는 표4와 같다. 이 예제를 분석하기 앞서 수리적 모형에 대하여 살펴보면 두 가지 치료방법을 A와 B라 하고 각 시점에서의 시험대상수를 표 5와 같이 정의하자.

표 5. 기호정의

군	환측 수	시기		합		차이	
		1	2	계	평균 SD	계	평균 SD
AB	n ₁	A(x ₁)	B(x ₂)	T ₁ = x ₁ +x ₂	\bar{T}_1 S _{T1}	D ₁ =x ₁ -x ₂	\bar{D}_1 S _{D1}
BA	n ₂	B(y ₁)	A(y ₂)	T ₂ =y ₁ +y ₂	\bar{T}_2 S _{T2}	D ₂ = y ₁ -y ₂	\bar{D}_2 S _{D2}

1군과 2군, 즉 AB군 BA군을 합친 전체군의 합과 차이의 합동분산은 다음과 같고, 자유도는 모두 n₁ + n₂ -2이다.

전체군의 합에 대한 합동분산 :

$$S^2_T = \frac{(n_1 - 1) S^2_{T_1} + (n_2 - 1) S^2_{T_2}}{n_1 + n_2 - 2}$$

전체군의 차이에 대한 합동분산 :

$$S^2_D = \frac{(n_1 - 1) S^2_{D_1} + (n_2 - 1) S^2_{D_2}}{n_1 + n_2 - 2}$$

E(X), E(Y)는 위의 표에 제시된 X와 Y의 모평균이고, μ는 총평균, τ₁과 τ₂는 각각 치료 A와 B에 대한 직접효과 (direct effect), P₁과 P₂는 각각 시기 1과 시기 2에 대한 시기효과 (period effect)라 할 때 2x2 교차계획법에서 E(X), E(Y)는 다음과 같은 모형으로 가정한다.

$$E(X_1) = \mu + P_1 + \tau_1$$

$$E(X_2) = \mu + P_2 + \tau_2 + \rho_2$$

$$E(Y_1) = \mu + P_1 + \tau_2$$

$$E(Y_2) = \mu + P_2 + \tau_1 + \rho_2$$

추 의 역 강 직

시기효과란 비록 두 시기에 동일한 종류의 치료를 받았다 하더라도 두 번째 시기의 효과가 첫 번째 시기의 효과와 다르기 때문에 나타나는 일반적인 추세를 말한다. 또한 ρ_1 과 ρ_2 는 각각 치료 A와 B로 인한 이월효과를 나타내며, 치료순서 AB의 경우에는 A 치료로 인한 효과가 B 치료시에 아직 남아 있거나, 치료순서 BA의 경우에는 B 치료로 인한 효과가 A 치료시에 남아있을 수 있으며, 더욱 이 두 치료의 이월 효과의 크기가 서로 다를 수 있다. 2x2 교차계 획법에서 주요한 관심사인 치료 효과에 대한 검정은, 이월효과가 있으면 교차계획법으로 부터의 모든 자료를 사용하여 검정할 수가 없고 일부 자료만을 사용하여 분석하게 되므로 이월효과에 대해 먼저 검정하게 된다.

첫째, 이월효과 분석

실제로는 이월효과가 없음 (즉 $\rho_1 = \rho_2 = 0$)을 검정하기 원하지만 시기효과가 있는 경우에는 이월효과가 없음을 검정할 수 없다. 따라서 두 치료의 이월효과가 같다는 사실을 검정하며 귀무가설은 다음과 같다.

귀무가설 : $H_0 : \rho_1 = \rho_2$

이월효과에 대한 검정은 평균 T_1 와 T_2 의 비교에 근거한다. 다시 말하면 이월효과가 같다는 가설에 대한 검정은 1군과 2군의 개체들의 수치합의 평균인 T_1 와 T_2 를 비교하는 unpaired t 검정으로 가능하며 자유도가 n_1+n_2-2 인 검정통계량은 다음과 같다.

$$T = \frac{T_1 - T_2}{S_T} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

이 검정에서 유의하지 않다고 판단되는 경우에서야 다음의 치료효과에 대한 검정을 진행한다. 앞의 자료에 대하여 이월효과를 분석하면 표 6과 같다. 결과에서 보듯이 이월효과 (group effect 혹은 treatment by period interaction)는 없다는 결론을 얻을 수 있다.

표 6. 이월효과에 대한 분석 결과

	처치군	
	NP ($n_1 = 7$)	PN ($n_2 = 6$)
평균 (l /min)	643.571	629.167
합동분산	$\frac{643.571^2 + 629.167^2}{6+7-2} = 20886.96$	
표준오차	$\sqrt{\frac{20886.96}{6} + \frac{20886.96}{7}} = 80.41$	
검정통계량	$\frac{14.404}{80.41} = 0.18$	
자유도	6+7-2 = 11	
유의확률	0.00124	

둘째, 처리효과 분석

처리효과에 대한 분석은 두 표본 t 검정법과 Hills-Armitage 방법 (12, 13)이 있다. 두 방법의 차이는 전자의 경우 시기 I과 시기 II 각각의 평균의 차이에 대한 t-검정을 수행하며, 후자의 경우에는 두 시기효과의 평균을 사용한다. 즉 후자에서는 두 시기에서 시기효과가 평균적으로 작용한다는 것을 가정한 것이다. 시기효과를 보정하는 간단한 방법은 시기차이 (period difference)가 동일하다고 가정하고 두 표본 t-검정을 시행하는 방법으로 치료 효과가 동일하다는 검정에 대한 귀무가설은 다음과 같다.

$$H_0 : \tau_1 = \tau_2$$

이월효과가 동일하다는 가정하에서 다음 단계는 치료효과가 동일하다 ($\tau_1 = \tau_2$)는 검정을 수행하게 된다. 귀무가설에 대한 검정은 우선 각 개체에게서 측정된 두 수치의 차를 구한 후 1군과 2군의 이러한 차의 평균인 D_1 과 D_2 를 비교하는 unpaired t 검정으로 가능하며 자유도가 $n_1 + n_2 - 2$ 인 검정통계량을 갖는다.

$$T = \frac{\overline{D_1} - \overline{D_2}}{S_D} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

위의 공식에서 볼 수 있듯이 치료효과에 대한 검정은 시기효과가 어떠한 간에 D_1 와 D_2 의 비교로 가능하다. 앞의 자료에 대하여 치료효과를 분석하면 두 가지 방법 모두 치료효과는 있다고 판단할 수 있다 (표 7, 8).

표 7. 두 표본 t 검정에 의한 치료효과 분석

	처치군	
	NP ($n_1 = 7$)	PN ($n_2 = 6$)
기간 차이의 평균	30.71	-62.5
합동분산	$\frac{30.71^2 + (-62.5)^2}{6+7-2} = 1500.81$	
평균차이	$30.71 - (-62.5) = 93.21$	
표준오차	$\sqrt{\frac{1500.81}{6} + \frac{1500.81}{7}} = 21.55$	
검정통계량	$\frac{93.21}{21.55} = 4.33$	
자유도	$6+7-2 = 11$	
유의확률	0.00124	

표 8. Hills-Armitage 방법에 의한 치료효과 분석

	처치군	
	NP ($n_1 = 7$)	PN ($n_2 = 6$)
시기차이의 평균	30.71	-62.5
평균의 평균	46.61	
합동분산	$\frac{30.71^2 + (-62.5)^2}{6+7-2} = 1500.81$	
표준오차	$\sqrt{\frac{1500.81}{13}} = 10.74$	
검정통계량	$\frac{46.61}{10.74} = 4.33$	
자유도	$6+7-2 = 11$	
유의확률	0.00096	

셋째, 시기효과 분석

두 치료의 이월효과가 없을 때 시기효과가 동일하다는 검정에 대한 귀무가설은 다음과 같다.

$$H_0 : p_1 = p_2$$

단순한 2x2 교차계획법에서는 위의 두 가지 검정으로 그치는 것이 보통이지만 연구의 특성에 따라서는 이월효과가 전혀 개입되지 않아서 $\rho_1 = \rho_2 = 0$ 임을 가정하는 것에 무리가 없는 경우가 있다. 이 때 치료효과에 대한 검정과 더불어 다음에 제시하는 시기효과에 대한 검정이 가능하다. 시기효과에 대한 검정은 우선 각 개체에게서 측정된 두 수치의 차를 구한 후, 1군과 2군의 차의 평균인 D_1 와 D_2 를 비교하는 unpaired t 검정으로 가능하며, 자유도가 $n_1 + n_2 - 2$ 인 검정통계량은 다음과 같다.

$$T = \frac{\overline{D_1} + \overline{D_2}}{S_D} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

추 의 혁 강 죄

앞의 예제에 대하여 시기효과를 검정하면 표 9와 같고 시기효과는 유의하지 않다는 결론을 얻을 수 있다.

표 9. 시기효과에 대한 분석

	처치군	
	NP (n ₁ = 7)	PN (n ₂ = 6)
시기 차이의 평균	30.71	-62.5
시기 차이	$\frac{30.71-62.5}{2} = -15.895$	
표준오차	$\frac{\sqrt{1500.81}}{\sqrt{13}} = 10.74$	
검정통계량	$\frac{-15.895}{10} = -1.48$	
자유도	6+7-2 = 11	
유의확률	P>0.05	

지금까지의 설명을 SAS 코드로 작성하면 다음과 같다.

```
DATA CROSS;
INPUT GROUP PATIENT TREAT $ PERIOD PEF @@;
CARDS;
1 1 N 1 310 1 1 P 2 270
1 2 N 1 310 1 2 P 2 260
1 3 N 1 370 1 3 P 2 300
1 4 N 1 410 1 4 P 2 390
1 5 N 1 250 1 5 P 2 210
1 6 N 1 380 1 6 P 2 350
1 7 N 1 330 1 7 P 2 365
2 8 N 2 385 2 8 P 1 370
2 9 N 2 400 2 9 P 1 310
2 10 N 2 410 2 10 P 1 380
2 11 N 2 320 2 11 P 1 290
```

```
2 12 N 2 340 2 12 P 1 260
2 13 N 2 220 2 13 P 1 90
```

```
PROC FORMAT;
VALUE GROUP 1='N/P' 2='P/N';
VALUE TREAT 1='N' 2='P';
PROC GLM DATA=CROSS;
CLASS PATIENT TREAT;
MODEL PEF=PATIENT TREAT;
ESTIMATE 'N/P' TREAT 1 -1;
PROC GLM DATA=CROSS;
CLASS GROUP PATIENT PERIOD TREAT;
MODEL PEF=GROUP PATIENT(GROUP) PERIOD
TREAT;
TEST H=GROUP E=PATIENT(GROUP);
ESTIMATE 'N/P' TREAT -1 1;
RUN;
```

이분형 결과인 경우

한편, 측정결과가 이분형 (binary outcome) 으로 측정되는 경우가 있다. 임상시험에서 흔히 결과변수를 증상 호전 여부와 같이 이분적인 결과로 측정되는 자료에 대하여 살펴보자 (14).

첫째, 시기효과를 무시한 상태에서의 분석

예를 들어 총 24두의 개에 대한 AB/BA 교차 시험에서 두 처치에 시험대상을 무작위로 할당하였다.

약제투여군 (N)과 위약투여군 (P) 두 약물의 효과를 비교하는데 결과를 1=poor; 2=fair; 3=moderate; 4=good 등으로 판정하였다고 하자. 분석의 목적 상 1, 2, 3을 '-' (not good) 로, 4는 '+' (good) 로 가정하여 표 10과 같은 결과를 얻었다.

표 10. AB/BA 교차시험에서 이분형 결과의 예

순서	환측번호	약제효과	
		N	P
N/P	3	4+	4+
	4	3-	1-
	7	4+	1-
	8	4+	3-
	9	4+	4+
	11	4+	3-
	15	4+	3-
	16	4+	1-
	19	4+	3-
	20	4+	1-
	22	4+	3-
23	4+	2-	
P/N	1	4+	2-
	2	4+	3-
	5	4+	4+
	6	4+	4+
	10	4+	4+
	12	4+	4+
	13	4+	4+
	14	3-	4+
	17	4+	3-
	18	4+	2-
	21	4+	2-
24	4+	3-	

이 자료에서 N과 P의 선호결과 (preference)에 따라 3개의 군으로 분류하면 다음과 같다.

- 첫째, N + 이고 P - 인 경우 = N 군 선호
- 둘째, N - 이고 P + 인 경우 = P 군 선호
- 셋째, N - 이고 P - 인 경우 = 선호군이 없음

	①군	②군	③군
	15	8	1
	a	b	c
또는		NP	PN
	N 선호	9	0
	선호 없음	3	5
	P 선호	1	6

$$a=9+6=15 ; b=3+5=8 : c=1+0=1$$

위의 결과표 좌측에서는 짝지어진 쌍이며, 16개의 preference 쌍이 있으므로 McNemar 검정에 해당한다. 검정의 귀무가설은 preference를 보이는 처리 군이 특정한 약물을 선호할 확률이 0.5로 동일하다는 것이다. 따라서 p 값은 $n=16$, $p=0.5$ 인 이항분포로부터 계산된다. 두 치료로부터의 결과가 서로 동등하게 나타나면 (즉 preference가 없는 경우) 어떤 치료가 더욱 효과적임을 말해주지 못하므로 치료효과를 알아보는 데는 일반적으로 preference가 있는 자료만이 사용된다. 행과 열의 주변도수가 고정된 상태에서 정확확률 (exact probability)은 초기화 분포를 따르지만 이항분포를 이용할 경우 큰 표본 (large-sample) 가정에 부합되면 카이제곱 (chi-square) 분포에서 근사치를 계산한다. 시기효과가 없고 치료와 시기의 교호효과도 없다고 할 때 카이분포의 적합도 통계량의 일반형은 다음과 같다.

$$\frac{\sum (O_i - E_i)^2}{E_i}$$

여기에서 O_i 는 i 번째 셀의 관찰치, E_i 는 i 번째 셀의 기대값이다. 이 공식을 변형시키면 자유도가 1인 짝지어진 자료에 대한 McNemar 카이제곱 통계량이 만들어진다.

$$\frac{(a-c-1)^2}{a+c} = \frac{(15-11-1)^2}{16} = 10.56$$

따라서 카이분포표에서 임계값과 비교하면 $p=0.001$ 을 얻는다. 만일 치료와 시기의 교호 효과는 없으나 시기효과가 있다고 가정하면 위 결과표 (3x2 분할표)에 대하여 카이제곱으로 치료효과를 검정하면 된다.

시기효과를 고려한 상태에서의 분석

첫째, Mainland-Gart test (15)

앞의 분석에서는 시기효과를 고려하지 않았는데 만일 시기효과를 고려하여 분석한다면 첫째, 1차 시기에서 선호된 치료를 갖는 환자의 수 둘째, 2차 시기에서 선호된 치료를 갖는 환자의 수를 고려하게 되며 이를 정리하면 표 11과 같다.

표 11. 앞의 예제에서 이분형 결과의 분류

순서	1차시기 선호	선호도		선호도 합계
		없음	선호	
N/P	9	3	0	12
	d	e	f	γ_1
P/N	1	5	6	12
	g	h	k	γ_2
계	10	8	6	24
	s_1	s_2	s_3	N

처리효과가 없다는 귀무가설하에서 시간 경과에 따른 preference pattern은 시간효과를 반영하지만 처리효과는 특정한 기간에서 두 순서 (sequence) 군간의 차이에 반영된다. McNemar 검정에서와 같이 아무런 정보를 갖지 못하는 no-preference 쌍을 무시하면 표 12와 같다.

표 12. 앞의 예제에서 시기 선호도 (period preference) 별 이분형 결과의 분류

순서	선호도		
	1차시기	2차시기	계
N/P	9	0	9
	d	f	d+f
P/N	1	6	7
	g	k	g+k
total	10	6	16
	d+g	f+k	d+f+g+k

Mainland-Gart 검정은 위와 같은 표에서 시기와 순서 (sequence) 간의 연관성을 검토한다. 처리간의 차이는 처리간의 연관성을 나타낸다. 예를 들어 N/P 군에서 1차시기 값이 선호되는 반면, P/N 군에서는 2차시기 값이 선호된다는 사실은 P 보다는 N에 대한 선호경향이 있음을 시사한다. 표 12에서 연관성에 대한 검정은 Fisher's exact 검정이나 큰 표본에 대한 카이제곱분포의 근사성을 이용한다. 이분형 결과를 preference로 줄여야 하는 Mainland-Gart 검정에서는 동일한 값 (tied-value)은 분석에서 제외된다. 카이제곱분포 근사성을 적용하면 다음과 같은 공식이 만들어지며, 본 예제에

대하여 계산하면 $X^2 = 8.96$ 을 얻는다. ($p=0.003$).

$$\frac{M\{|dk-fg|-M/2\}^2}{(d+g)(f+k)(d+f)(g+k)} \quad [M=d+f+g+k]$$

둘째, Prescott's test

이 방법은 tied preference 자료를 무시하는 Mainland-Gart 검정의 대안으로 제시된 방법으로 자세한 내용은 전문서적을 참고하기 바란다 (16). 한편, 앞에서 순위형 자료를 설명의 편의상 이분형으로 간주하여 처리하였는데 순위형을 그대로 유지하면서 logistic 회귀분석을 이용한 방법도 흔히 사용되고 있다 (17, 18).

실험오차의 크기를 계량화할 수 있다. 블록화는 일종의 제한된 랜덤화로서 실험의 정밀성을 높이는 기법이다. 모든 실험설계에서 반드시 블록화를 사용해야하는 것은 아니지만 실험의 특성상 필요하다고 판단되면 블록화를 하는 것이 유리하다. 블록화를 하면 블록내에서는 동질성이 유지되고 블록간에는 이질적이 되므로 블록간의 차이가 크면 클수록 블록화의 효율이 증대된다. 예컨대, 사료의 종류에 따른 돼지 증체율 비교실험에서 품종간 증체율의 차이를 클수록 블록화의 가치가 높아지는 셈이다. 결과를 해석할 때 반드시 통계적 분석에 의존할 필요는 없으며 통계적 유의성과 생물학적 가능성 (biological plausibility)이 반드시 일치하는 것은 아니므로 결과에 대한 신중한 검토가 요구된다. 마지막으로, 모든 실험을 종료하여 자료를 정리한 후 역으로 분석모형을 결정하려는 태도는 버려야 한다. 이는 자신의 몸과 취향에 맞는 옷을 구입하는 것이 아니라 옷에 몸과 취향을 맞추려는 시도와 다를 바 없기 때문이다.

결론

모든 실험자료에 대하여 통계분석이 가능한 것은 아니다. 실험의 목적이 요인의 수준을 조절하였을 때 반응이 어떻게 변하는지를 비교·분석하려는 것이라 할 때 실험 결과에 대한 객관적이고 타당한 해석이 가능하기 위해서는 올바른 실험설계를 계획하는 것이 필수적이다. 이러한 실험계획의 원리를 따르는 일차적인 이유는 실험에 개재될 수 있는 가능한 주관적인 요인을 배제하기 위한 것이다. 실험설계를 계획할 때 몇 가지 주의사항을 정리하면 다음과 같다. 측정값은 연구자가 임의로 조절할 수도 없고 바라지도 않는 실험오차와 관찰오차가 개입되어 나타난 결과이므로 오차를 가능한 줄이는 것이 중요하다. 실험계획이 최소한의 비용으로 최대의 효과를 얻기 위한 수단이므로 적정 수준의 반복은

참고문헌

1. Montgomery DC. Design and analysis of experiments. 4th ed, John Wiley & Sons, New York, 1997.
2. Daya S. Differences between crossover and parallel study designs-debate. Fertil Steril, 1999;71:771-772.
3. Ramirez S, Wolfsheimer KJ, Moore RM, et

- al. Duration of effects of phenylbutazone on serum total thyroxine and free thyroxine concentrations in horses. *J Vet Intern Med*, 1997;11:371-374.
4. Schumacher J, DeGrave FJ, Spano JS. Clinical and clinicopathologic effects of large doses of raw linseed oil as compared to mineral oil in healthy horses. *J Vet Intern Med*, 1997;11:296-299.
 5. Burton SA, Lemke KA, Ihle SL, et al. Effects of medetomidine on serum insulin and plasma glucose concentrations in clinically normal dogs. *Am J Vet Res*, 1997; 58:1440-1442.
 6. Papich MG, Alcorn J. Absorption of diazepam after its rectal administration in dogs. *Am J Vet Res*, 1995;56: 1629-1636.
 7. Michels GM, Boudinot FD, Ferguson DC, et al. Pharmacokinetics of the antihyperglycemic agent metformin in cats. *Am J Vet Res*, 1999;60: 738-742.
 8. Grender JM, Johnson WD. Analysis of crossover designs with multivariate response. *Stat Med*, 1993;12:69-89.
 9. Freeman PR. The performance of the two-stage analysis of two-treatment, two-period cross-over trials. *Stat Med*, 1989; 8:1421-1432.
 10. Fleiss JL. The design and analysis of clinical experiments. Wiley, New York, 1986.
 11. Senn S, Hildebrand H. Cross-over trials, degree of freedom, the carry-over effect and its dual. *Stat Med*, 1991;10:1361-1374.
 12. Hills M, Armitage P. The two-period cross-over clinical trial, *British J Clin Pharmacol*, 1979;8:7-20.
 13. Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics*, 1955;11:375-386.
 14. Senn S. Cross-over trials in clinical research. John Wiley & Sons, New York, 1993.
 15. Jones G, Kenward MG. Design and analysis of cross-over trials. Chapman and Hall, London, 1989.
 16. Prescott RJ. The comparison of success rates in cross-over trials in the presence of an order effect. *Appl Stat*, 1981;30: 9-15.
 17. Grizzle JE. The two-period change-over design and its use in clinical trials. *Biometrics*, 1965;21:467-480.
 18. Gart JJ. An exact test for comparing matched proportions in cross-over designs. *Biometrika*, 1969;56:75-80. 