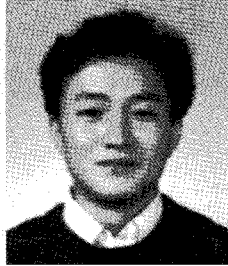


의학연구에서 실험계획의 원리와 교차계획법



박 선 일

서울시 관악구 신림9동 산 56-1

서울대학교 수의과대학 내과학교실

H.P: 016-208-2944

대한수의

실험은 과학적 사실을 규명하기 위한 적극적인 수단이다. 실험이란 문자 그대로 검정 (test)을 의미하는데 좀 더 공식적으로 표현하자면 “연구자가 어떤 체계(system) 혹은 과정 (process)속으로 의도하는 변화를 투입 변수 (input variable)로 넣었을 때 그 결과로 얻게되는 반응에서 어떠한 변화가 초래되었는지 그 원인을 찾아내는 일련의 검정(series of tests)” 이라고 정의할 수 있다. 넓은 의미에서 볼 때 실험의 형태는 발생조사 (outbreak investigation)와 같이 현장 (field)에서 진행되는 관찰연구 (observational study)와 실험실 연구로 대별된다. 어떠한 연구형태이든 실험의 객관성을 달성하기 위해서는 연구의 목적에 합당한 실험계획을 작성하는 것이 필수적이다. 아무리 완벽하게 실험을 수행했다고 하더라도 실험계획에서 심각한 문제가 있다면 그 결과에 대해서는 신뢰성을 부여할 수 없기 때문이다. 본 논문에서는 실험을 계획할 때 적용되는 원리와 분산분석에서의 다중비교 및 의학연구에서 활용빈도가 높은 교차계획법 (crossover design)에 대하여 소개하고자 한다.

1. 실험계획의 원리

실험계획 (experimental design)은 연구자가 관심을 두고 있는 처리방법 (treatment)간의 비교가 왜곡되지 않고 가능한 정확하고 높은 검정력 (statistical power)을 유지할 수 있도록 실험단위 (experimental unit)를 선정하고 처리방법을 효과적으로 할당하는 과정이라고 할 수 있다. 실험계획에서는 흔히 요인 (factor)과 수준 (level)이라는 용어를 사용한다. 요인이란 연구자가 관찰대상으로 선택한 변수로서 처리 혹은 인자라고도 한다. 예를 들어, 여러 품종의 동물들에 대하여 동일한 처리방법을 적용하여 그 반응을 연구한다고 할 때 품종이 요인이 되며, 돼지에게 먹이는 네 가지 사료의 증체효과를 비교하는 경우 사료가 요인이 된다.

수준은 요인이 취하는 값이다. 돼지 증체율 예제에서 요인인 사료가 취하는 네 가지 값이 수준이 된다. 어떤 마취제의 마취유도 시간을 비교하고자 두 집단에 대해 한 군에는 마취제 A를, 다른 한 군에는 마취제 B를 투여한 경우 요인은 마취제가 되며 요인수준은 마취제 A와 B로 2개의 질적 (qualitative) 수준을 갖는다. 어떤 약을 5mg/ml, 10mg/ml, 15mg/ml의 농도로 투여하는 경우 요인은 약이고, 요인의 수준은 세 가지 농도로 이 경우 3개의 양적 (quantitative) 수준을 갖게 된다.

한편, 변량인자 (random variable)라는 용어는 수준의 선택이 랜덤으로 이루어지며 각 수준이 기술적인 의미를 가지고 있지 못한 요인을 말한다. 예를 들면, 실험 기간을 3일 선택하여 수행한 경우 날짜라는 구분은 수준이 세 개인 변량인자가

되는데, 이 날짜는 무수히 많은 날짜 중에서 임의로 선택된 3일이며 선택된 하루하루는 아무런 기술적인 의미를 가지고 있지 못하기 때문이다. 어떠한 형태의 실험에서도 통계적 해석을 가능하게 하는 세 가지 기본적인 원리는 랜덤화 혹은 무작위 (randomization), 반복 (replication) 및 블록화 (blocking)의 개념이다. 흔히 처리군간 비교의 목적으로 실험을 수행하는 경우가 많은데 비교실험 (comparative experiment)이란 문자 그대로 상호 비교를 목적으로 하는 실험으로 각 요인의 수준마다 동일한 조건하에서 실험한 후 요인수준에 따라 측정된 값들간에 통계적으로 유의한 차이가 있는지를 평가하는 것이다. 예컨대, 젖소의 소화기 질환에 대한 약물요법과 침술요법의 효과를 비교하는 실험, 마취제의 부작용을 평가하기 위하여 마취 후 경시적으로 혈청화학치의 변화를 평가하는 실험, 진균증을 치료하는데 사용되는 세 가지 약물의 효과를 비교하는 실험, 사료 급여 후 pulsed-Doppler ultrasonography를 이용하여 매 시간마다 장운동을 비교하는 실험 등은 이러한 예다. 실험의 목적이 두 가지 (혹은 여러 가지) 처리방법간의 비교인 만큼 측정결과에 영향을 미칠 수 있는 요인들을 처리군간에 모두 동일하게 유지해야 하는 것은 필수적이다. 예를 들어 돼지에서 네 가지 사료의 증체효율을 비교한다고 하자. 이 경우 실험의 대상이 되는 돼지들의 유전적 특성이나 실험전 체중과 같은 증체에 영향을 미칠 수 있는 요인들을 이론적으로 동일하게 유지해야 한다. 그렇지 않으면 실험결과 증체율에 차이가 있는 것으로 나타났을 때 과연 사료의 효과에 의한 차이인지 아니면 실험대상의 특성에 따른

차이인지 구분할 수 없기 때문이다. 이러한 문제에 대한 한가지 대안으로서 동일한 어미에서 태어난 자돈만을 대상으로 실험할 수도 있지만 수적인 측면에서 한계가 있어 처리군이 많아질수록 불가능해진다. 이러한 문제는 실험계획의 원리를 이용하면 비교적 간단하게 해결된다.

랜덤화의 원리란 실험대상 선택의 동질성(homogeneous)과 실험대상 배치의 랜덤화를 의미한다. 전자는 연구자가 관심을 두고 선발한 요인이외의 다른 원인에 의한 영향이 실험 결과에 영향을 미치지 않도록 실험의 순서를 시간적으로나 공간적으로 랜덤하게 배치하는 것을 의미한다. 앞에서 언급한 사료 증체율에 관한 예에서 네 가지 사료에 할당되는 돼지들의 특성들이 가능한 동질적인 개체들로 구성하는 것이다. 어느 정도의 동질성을 만족해야하는지는 연구자 스스로가 판단해야할 몫이다. 실험대상 배치의 랜덤화는 동질적인 실험대상을 선발한 후 네 종류의 사료에 어떻게 이들을 랜덤하게 배치할 것인가에 관한 것이다. 난수표(random number table)나 컴퓨터 패키지를 사용하면 랜덤배치를 얻을 수 있다.

예를 들어 각막 혼탁을 유발하는 각막 심부 간질의 창상에 대한 양막이식(amniotic membrane graft)의 효과를 알아보기 위하여 창상모형을 만든 후 항생제만 점안한 대조군, 안검봉합군(tarsorrhaphy), 콘택트렌즈군(contact lens) 및 양막이식군에 대한 상피치유 속도를 측정하는 실험에서 네 군에 토끼를 랜덤배치 후 표 1과 같은 결과를 얻었다고 하자.

표 1. 양막이식의 효과를 알아보기 위한 창상모형에서 네 군의 치유속도

대조군	안검봉합군	콘택트렌즈군	양막이식군
42.5 μ m/hr	43.7 μ m/hr	33.4 μ m/hr	53.4 μ m/hr

이 결과에서 처리군에 따라 상피치유 속도에 차이가 있다고 할 수 있는가? 또한 양막이식군의 치유속도가 가장 빠르다는 결론을 내릴 수 있는가? 실험에는 언제나 오차(error)가 따르기 마련이다. 아마도 각 처리방법이 치유속도에 미치는 처리방법 고유의 효과(effect)가 있을 것이지만 이 효과가 어느 정도의 값일지는 아무도 모른다. 우리가 실제로 관측한 값들은 다소간의 오차가 덧붙여져 나온 값들이다. 만일 오차들이 ± 9 를 넘지 않는다면 위의 결과에서 처리방법간에 차이가 있고, 양막이식군의 치유속도가 가장 빠르다는 결론을 내릴 수 있을 것이다. 그러나 만일 오차가 ± 15 라면 양막이식군의 치유속도가 빠르기는 하지만 이러한 차이는 모두 오차 때문이라고 해도 과언이 아니며 별 차이가 없다는 결론을 얻게 된다. 따라서 적절한 분석 결과를 얻으려면 반드시 자신이 수행한 실험의 오차를 통계적으로 추정할 수 있어야 하며 오차는 가급적 작으면 작을수록 더 좋다. 통계학에서 오차의 추정값은 표본분산으로부터 산출되는데 본 예제의 경우 한 가지 처리방법에 한 마리를 배치하였다면 분산을 계산할 수 없다. 이러한 문제를 해결하는 방법은 각 요인의 수준마다 실험의 횟수를 적어도 둘 이상으로 늘리는 것이며, 반복 횟수가 늘어날수록 결과는 더 정확해진다. 실험계획에서 반복은 분석의

주의사항

정밀도를 향상시키는 주요한 도구로 각 수준의 조합에서 2회 이상 실험을 반복함으로써 실험 결과의 신뢰성을 증가시키는 방법이다. 반복의 개념은 두 가지 중요한 성질을 갖는데 첫째는 실험오차의 추정치를 제공해준다. 이 오차는 자료에서 관찰된 차이가 통계적으로 차이가 있는지를 판단하는데 기본단위가 된다. 두번째로 어떤 실험에서 요인의 효과를 추정하는데 표본 평균을 사용할 경우 반복은 연구자로 하여금 보다 정확한 효과의 추정치를 제공해 준다. 예를 들어 개별 관찰치의 분산을 σ^2 반복횟수를 n 이라고 하면 표본평균의 분산은 σ^2/n 이 된다 $n=1$ 이고, 관찰치의 평균을 각각 $y_1=145$, $y_2=147$ 이라고 하면 요인의 효과 즉 관찰된 차이가 실험오차에 의한 것인지에 대하여 만족할만한 추론이 불가능하다. 그러나 n 이 충분히 크고, 실험오차가 충분히 작다면 y_2 의 평균이 y_1 의 평균보다 크다는 론을 내리기가 용이해진다. 즉 반복은 오차항의 자유도 (degree of freedom)를 증가시켜 분산의 추정을 좀 더 정확하게 해주어 실험의 신뢰성을 증가시킬 수 있게 된다. 일반적으로 오차항의 자유도는 적게는 6, 많으면 20 정도로 해주는 것이 좋다고 한다. 위의 예제에서 각 처리군에 10두를 랜덤배치한다면 반복의 개념이 적용된 것이다. 만일 상피치유 속도가 토끼의 품종에 따라 다르다는 것을 경험적으로 알고 있다고 하자. 예를 들어 실험에 사용되는 품종이 2가지라고 할 때 랜덤화를 시행한 후 양막이식군에는 대부분 품종 1의 토끼만이 배치되었고, 나머지 군에는 품종 2의 토끼만이 배치되었다면 결과 해석을 그래도 믿을 수 있을 것인가? 이 문제는 블록화의 개념으로 해결된다.

블록화란 실험을 시간적 혹은 공간적으로 분할하여 그 내부에서 실험의 환경 (또는 조건)이 균일하도록 만들어 놓는 것을 의미한다.

예를 들어, 15번의 실험을 시행할 때 1일에 3회씩 5일에 걸쳐 실험한다면 1일이 하나의 블록의 역할을 하는 것이 된다. 즉 실험의 결과가 온도, 습도 및 기타 기상조건 등 일일 간 변동에 영향을 받을 수 있다면 제 1일에 실험한 모든 개체는 동일한 조건에 놓이게 된다는 의미이다. 위의 예제의 경우에서와 같이 처리방법의 효과에 토끼의 품종에 따른 효과가 혼합되어 있을 가능성이 있을 경우 랜덤화를 수행 할 때 한 군에 특정 품종의 개체가 몰리는 것을 예방하기 위해서는 품종별로 별도로 랜덤화를 하여 배치하는 편이 나올 것이다.

즉 품종 1의 토끼가 12마리, 품종 2의 토끼가 12 마리 있다면 우선 품종 1의 토끼를 각 처리군 당 세 마리씩 랜덤배치하고, 그 다음 품종 2의 토끼 12마리를 각 처리군마다 세 마리씩 배치하는 것이다. 결국 품종에 따라 완전히 독립적인 랜덤화를 하게 되며 이 경우 품종을 블록화 (blocking)한다고 하며 각 블록내에서는 동질성이 유지된다 [참고로, 분석결과 처리간 효과가 있다고 할 때 순수한 처리에 의한 효과와 다른 효과가 혼합되어 있는 경우 효과들이 교락 (confounded)되었다고 표현한다. 효과들이 교락되어 있는 경우 각 효과들을 구분할 수 없는 문제가 있으나 실험횟수를 줄이려는 목적에서 의도적으로 효과들을 교락시키는 방법을 사용하기도 한다.]

이렇게 하면 품종에 기인한 상피 치유속도의 효과를 처리방법에 기인한 치유속도의 효과와

구별할 수 있게 되어 품종간의 차이와 무관한 순수한 처리방법에 의한 효과를 측정할 수 있게 되는 것이다. 블록의 개념을 도입하면 블록을 하나의 인자로 간주하여 총변동 (total variation)에서 블록간의 변동을 제외시키고 남는 것은 균일한 블록 내의 다른 인자의 변동만이 남게 되는 것이다.

이 원리를 이용한 대표적인 실험계획법이 난괴법 (randomized block design)이다. 결국 블록화 설계를 하게 되면 완전랜덤화에 비하여 실험 오차를 줄이게 되어 더욱 정확한 분석이 가능해지므로 블록화에 의한 설계를 오차조절 설계 (error control design)라고도 표현한다.

2. 실험계획의 순서

일반적으로 실험은 연구목적의 구체화, 반응 변수의 선택, 요인과 수준의 결정, 실험계획의 선택, 실험수행, 자료분석 및 결론의 단계를 거치게 되는데 각 단계별로 고려해야할 내용은 다음과 같다.

단계 1 : 실험목적의 설정

실험을 통하여 얻고자 하는 목적이 구체적이면서 명료하게 서술되어야 한다. 당연한 일인 것 같지만 흔히 다른 연구자의 실험계획을 그대로 인용하여 실험을 수행한 후 어떠한 분석이 가능한지를 역으로 찾는 경우가 많은데 이는 분명 잘못된 일이다. 연구 목적과 이용 가능한 표본의 크기에 따라 실험계획은 다를 수 있기 때문이다.

단계 2 : 반응변수의 선택

반응변수(response variable)는 실험목적에 비추어 볼 때 실험인자의 효과를 가장 잘 나타내는 특성치로 정해야 한다.

마취유도 시간을 측정하는 실험에서 반응 변수인 유도시간 (induction time)은 양적 (qualitative) 변수이고, 백신효과 실험에서 감염의 유·무로 판단하는 경우 질적 (qualitative) 변수가 된다.

단계 3 : 요인과 요인수준의 선택

반응변수에 영향을 미치는 요인은 무수히 많지만 네 가지 요인으로 요약된다. 첫째, 고정인자군 (fixed factor, 모수인자)은 이미 반응변수에 대한 영향이 확인되어 연구대상으로 취급할 필요가 없는 인자로 각 수준이 기술적인 의미를 가지고 있는 인자이다. 실제 실험에서 이들 인자는 미리 정하여진 수준이 사용된다. 즉 실험시 이들 인자는 일정한 상태로 고정되는데 예를 들어 수술 방법이나 마취제 농도 등이다. 둘째, 랜덤인자군 (random factor)은 수준의 선택이 랜덤으로 이루어지며 각 수준이 기술적인 의미를 가지고 있지 못한 인자로서 흔히 실험대상 개체는 랜덤인자군으로 처리한다. 셋째, 실험인자군 (experimental factor)은 연구 대상으로 선정된 인자군으로 실험가설로 표현된다. 넷째, 오차인자군 (error factor)은 반응변수와 관련이 있을 것으로 생각되나 그 내용이 알려져 있지 않아 취급할 수 없는 인자군이다. 표2는 고정인자군과 랜덤인자군을 차이를 비교한 것이다.

표 2. 고정인자와 랜덤인자의 비교

고정인자 (모수인자)

1. 수준이 기술적인 의미를 가지며 실험인자에 의하여 미리 정하여 진다.
2. a_i 는 고정된 상수이다.
 $E(a_i) = a_i, \text{Var}(a_i) = 0$
3. a_i 들의 합은 0이다.
 $\sum_{i=1}^q a_i = 0, \bar{a} = 0$
4. a_i 들간의 산포성의 측정으로
 $\sigma^2_A = \frac{\sum_{i=1}^q a_i^2}{q-1}$ 을 사용한다.

랜덤인자 (변량인자)

1. 수준이 기술적인 의미를 갖지 못하며 수준의 선택이 랜덤으로 이루어진다.
2. a_i 는 랜덤으로 변하는 확률변수이다.
 $E(a_i) = 0, \text{Var}(a_i) = \sigma^2_A$
3. a_i 들의 합은 일반적으로 0이 아니다.
 $\sum_{i=1}^q a_i \neq 0, \bar{a} \neq 0$
4. a_i 들간의 분포의 분산율
 $\sigma^2_A = E \left[\frac{1}{q-1} \sum_{i=1}^q (a_i - \bar{a})^2 \right]$ 으로 정의한다.

* $a_i (a_i = \mu_i - \mu)$ 는 처리효과 ($\tau_i = \mu_i - \mu$)임

고정인자군을 어떠한 상태로 고정시키는가는 실험목적과 상황에 따라 결정하며 고정인자군의 고정상태에 따라 실험인자의 작용방식이 결정된다. 실험에서의 구체적인 문제로서 실험인자의 수준이 양적인 경우 특별한 문제가 없는 한 등간격을 유지하는 것이 좋으며 오차인자군은 우연성에 맡길 수밖에 없다. 또한 연구자가 관심을 가지고 있는 흥미영역(region of interest, 연구자가 관심을 가지고 있는 인자수준이 변할

수 있는 범위)에서만 수준을 잡아준다.

단계 4 : 실험계획의 선정

이 단계에서는 표본의 크기, 실험단위의 배치와 실험순서, 랜덤화, 블록화 및 반복의 원리 등을 어떻게 할 것인지를 결정한다. 표본의 크기는 통계적 검정력과 밀접한 관계가 있으므로 alpha 오류와 beta 오류를 적절히 선택해야 한다.

단계 5 : 실험수행

실험의 표준 protocol을 근거로 실험을 수행하며 전과정을 통하여 철저한 정도관리와 표준화가 이루어져야 한다. 실험수행에서 특히 중요한 것은 실험요인으로 선정된 조건들 이외의 다른 조건들은 언제나 동일하게 유지하는 것이다.

단계 6 : 자료분석, 해석 및 조치

실험결과로부터 얻은 자료는 실험계획에 합당한 모형을 근거로 분석하고 해석해야 한다. 간혹, 실험 후 정리된 자료를 보고 모형을 결정하는 경우가 있는데 이러한 접근은 금물이다. 실험계획에 따라 네 가지 종류의 제곱합 (sum of square) 중에서 어느 것을 사용할 것인지를 숙지하고 있어야 한다. 또한 관심을 두고 있는 요인의 특성이 모수인자인지 랜덤인자인지 정확히 파악하여 모형에 어떻게 투입할 것인지를 결정해야 한다.

실험결과와 해석은 실험에서 주어진 조건 내에서만 결론을 지어야 한다. 또한 실험에서 취급한 인자에 대한 결론은 그 인자수준의 범위 내에서만

얻어지는 결론이지 이 범위를 넘게되면 아무런 결론을 내릴 수 없다.

예를 들어 양막이식 효과의 예제에서 분석결과 유의한 차이가 있는 것으로 나타나면 네 가지 처리군간 상피 치유속도가 동일하다는 귀무가설을 기각하게 된다. 즉 네 집단 중 적어도 하나의 집단은 다른 집단과 상피 치유속도에 차이가 있다는 것을 의미한다. 따라서 어느 집단 간 차이가 있는지에 대하여 다중비교 (multiple comparison)를 시행하게 된다.

3. 다중비교

가설검정 결과 귀무가설을 채택하는 경우에는 처리효과가 있다고 주장할 수 없으며, 더 이상의 추가적인 분석은 필요없다. 반대로 귀무가설이 기각되면, 즉 처리간에 차이가 있다는 결론을 얻게 되면 어느 군간에 차이가 있는지를 검정하게 된다.

처리수준이 질적인 값일 경우 처리평균들간의 차이가 주된 관심사이며 이는 대비(contrast)로 표현된다. 대부분의 연구에서 밝히고자 하는 가설(대립가설)을 대비로 나타낼 경우 ‘대비들은 서로 직교하지 않는다’ 즉 ‘대비들이 서로 독립이 아니다’ 라고 표현하며 이는 곧 각각의 검정이 t로 독립이 아니라는 뜻이다. 바로 이러한 경우에 적용되는 검정방법이 다중비교다. 즉, 서로 독립이 아닌 검정을 여러번 계속할 경우 제 1종의 오류는 사전에 설정한 α 보다 더욱 크게되므로, 전체검정의 유의수준이 α 가 되도록 조정해주는 검정법을 적용해야 한다.

일반적으로 Bonferroni방법과 합교법칙 (union intersection principle)이 사용되는데 후자의 방법은 다중비교 방법 중에서 가장 보수적인 (conservative) Scheffe 방법을 사용하므로 p 값이 전자의 방법 보다 크게 나온다. 전자는 Bonferroni 부등식에서 비롯된 것으로 α 수준에서 k번의 통계검정을 시행할 경우 적어도 한번의 가양성 결과(α_T)를 얻을 확률은 $k\alpha$ 를 초과 하지않는다는 것이다.

$$\text{즉 } \alpha_T \leq k\alpha \Leftrightarrow \alpha \geq \frac{\alpha_T}{k}$$

서로 독립이 아닌 검정의 실시 횟수를 k라 할 때 $\alpha = \frac{\alpha_T}{k}$ 를 각 검정의 유의수준으로 하여 검정하는 방법이다. 대부분의 통계 패키지는 각 검정에 대한 p값이 제공되므로 출력물에 나타난 p값에 시험횟수인 k를 곱해주면 된다. 일반적으로 다중비교에서 귀무가설을 잘못하여 기각할 총 확률(α_T) 즉 유효 p값 (effective p-value)은 $1-(1-\alpha)^k$ 가 되며 간단히 $\alpha_T = k\alpha$ 로 정의된다. 예를 들어,

$$\begin{aligned} \alpha &= 0.05, k = 3 \text{ 일 경우 :} \\ 1 - (1 - 0.05)^3 &= 1 - 0.857 = 0.143 \\ \alpha_T &= 3 \times 0.05 = 0.15 \end{aligned}$$

$$\begin{aligned} \alpha &= 0.05, k = 6 \text{ 일 경우 :} \\ 1 - (1 - 0.05)^6 &= 1 - 0.735 = 0.265 \\ \alpha_T &= 6 \times 0.05 = 0.30 \end{aligned}$$

따라서 유의수준이 0.05이고 k=3인 경우 유의수준의 판단기준을 0.15로 결정하거나 아니면 계산된 p 값에 3을 곱하여 얻은 p 값을 0.05와 비교하면 된다.

앞의 예제에서 10개의 쌍별 비교에 대한 귀무

가설에서 두 표본 t-검정과 같이 각각의 유의수준을 고정시키는 방식으로 오차를 조절하는 방법은 비교오류율 혹은 개별오류율 (comparison-wise error rate, individual error rate)을 조절한다고 말한다. 한편, 10개의 귀무가설에 대한 동시검정에 대한 유의수준을 정하는 방식으로 오차를 조절하는 것을 실험 오류율 또는 제 1종 모입오류율 (experiment-wise error rate, family error rate)을 조절한다고 말한다. 다중 비교와 관련된 가설의 형태에 따라 오차조절 방법이 달라진다. 예를 들어 μ_1 부터 μ_5 까지의 5개의 모평균 모두에 대한 비교에는 관심이 없고 그 중 일부의 모평균들에 대한 비교만을 고려하는 경우 예컨대, 위의 예제에서 5개 전체 모평균에 대한 귀무가설은 $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ 으로 이러한 귀무가설을 완전 귀무가설 (complete null hypothesis)이라 한다. 그러나 경우에 따라서 완전 귀무가설보다는 다음과 같은 가설에만 관심이 있을 수 있다.

$$H_0: \mu_1 = \mu_3 = \mu_5$$

이러한 귀무가설을 부분 귀무가설 (partial null hypothesis)이라 한다. SAS에서 부분 및 완전 귀무가설을 고려한 오차조절과 관련하여 CER (비교오류율 조절), EERC (완전 귀무가설하의 실험오류율 조절), EERP (부분 귀무가설하의 실험오류율 조절) 및 MEER(최대 실험오류율 조절, maximum experiment-wise error rate) 등의 방법이 제시되어 있다. 이들 중 MEER은 귀무가설의 종류에 무관하게 결정된다. 두 표본 t-검정으로 다중비교를 하는 것은 CER을 조절하는 것이며, 분산분석표를 이용한 F 검정은 다중비교는 아니지만 EERC를 조절하는 예이다.

다중비교는 오차조절뿐만 아니라 비교방식의 차이에 의하여 다음의 세 가지로 대별된다.

첫째, 쌍별 비교 (pair-wise comparison)

이 비교법은 한번에 두 개씩의 모평균들이 동일 한지 다른지를 검정하는 방법으로 두 표본 t-검정이 기초적인 예이다. 특히 두 표본 t-검정에서 비교하려는 두 모평균에 대한 표본의 크기들이 같은 경우를 Fisher의 최소유의차 (least square difference, LSD) 방법이라 한다. 이 방법은 CER을 조절한다. 쌍별비교에서 MEER을 조절하는 방법으로는 Bonferroni의 t 검정, Sidak의 t 검정, Scheffe의 다중비교법 (multiple comparison method), Tukey의 스튜던트화 범위 검정 (Studentized range test) 또는 정직 유의차 (honestly significant difference, HSD) 검정, 기타 Hochberg와 Gabriel 등의 비교법이 있다. 이들 중 Tukey와 Scheffe의 방법이 가장 보수적인 방법이며 Scheffe의 다중비교법은 분산분석표와 항상 일치된 결과를 보인다. 쌍별 비교에서 추천할만한 방법은 Tukey와 Sidak의 두 가지 방법이다.

둘째, 하나의 대조군과의 비교

다중비교의 특별한 경우로 어느 하나의 모평균과 나머지 모평균들 간의 비교할 때가 종종 있다. 예를 들어 어떤 요인의 한 가지 수준 (즉 control의 의미를 갖음)과 다른 수준들과의 비교를 목적으로 실험하는 경우이다. Dunnett의 다중비교법은 이러한 목적으로만 개발된 기법으로 MEER을 조절한다.

셋째, 다단계 검정

다단계 검정 (multiple-stage test)은 쌍별 비교의

설명에서 5개 평균간의 차이, 4개 평균간의 차이 식으로 단계별로 검정하는 비교법의 한 가지 형태로 쌍별 비교와의 차이점은 각 단계마다 유의수준을 다르게 놓는다는 것이다. 이 방법들은 대체로 강력하기는 하나, 주로 표본의 크기가 동일할 때를 가정하고 제안된 것들이라 표본의 크기에 많은 차이가 있을 경우 결과 해석이 어려워지는 문제가 있다. 다단계 검정 중 많이 알려진 방법은 Duncan의 다중범위 검정 (multiple range test), SNK 검정 (Student-Newman-Keuls test)이다. Duncan의 방법은 CER을 조절하며, SNK 방법은 EERC를 조절한다. 다단계 검정법 중에서 추천할만한 방법은 Ryan, Einot, Gabriel, Welsch 등이 고안한 방법으로 SAS에서는 GLM 절차

에서 REGWQ나 REGWF를 MEANS 문장에 지정하면 출력된다. 이 두 방법들은 지금까지 개발된 다단계 검정법 중에서 가장 강력한 방법으로 분산분석의 F 검정값과 일치하는 특성을 갖는다. 이 두 가지 방법은 모두 MEER을 조절한다. 이상의 설명을 요약하면 MEER을 조절하고 싶을 경우 표본의 크기들이 동일하면 REGWF나 REGWQ 방법 혹은 Tukey의 방법, 표본의 크기가 동일하지 않으면 Tukey의 방법을 사용하는 것이 좋다. CER을 조절하고 싶을 경우 Fisher의 LSD 방법이 쓸만하다. Ryan-Einot-Gabriel-Welsch의 다중범위 검정법은 Tukey의 표준화 범위 검정법을 개선하여 제 1종 모임오류율을 유지하며 검정력을 높인 방법이다.

농경애니텍 수의학 신간안내

New Vet Books!!

농경애니텍 TEL. 02) 3461-0295

- Saunder's Manual of small animal practice 2nd/2000/Birchard
- Atlas of dog&cat cytology/2000/Raskin
- Fluid Theraphy in small animal practice 23d/2nd/2000/Dibatola
- Diagnostic Radiology&ultrasonography of dog&cat 3rd/2000/kelay
- H/B of Vet Procedures&Emergency Treatmeat 7th/2000/Bistner
- Kirk's current Vet Therapy 13ed/2000/Bonagura
- T/B of Vet internal medicine 5ed/2000/Ettinger
- Small anamal Ear diseases/2000/Gotthelf
- Vet Dental Techniques/2000/Holmstrom
- Vet ophthalmology 3ed/99/Gelatt
- Comprehensive Vet dictionary/99/Blood
- Canine&Feline cardiology/99/Phillip
- Atlas of Small anamal Reconstructive Surgery/Michael/99
- Diseases of Swine/99/
- Color atlas of Small animal dermatology/99/Wilkinson
- Fish diseases and disorders/99/kwoo
- Zoo&wild animal medicine 4ed/99/Fowler
- Clinical Vet microbiology/99/Quinn
- Diagnostie Cytology&Hematology of dog.cat/99/Cowell
- Color Atlas of Vetanatomy I Dog&Cat/99/Done
- Color Atlas of Vet anatomy I Ruminants/99/Ashdwn