

## 實驗計劃의 基本原理

張錫煥 \*

母集團에 대한 推論을 위해서는 資料(情報)를 수집해야 하는데 자료는 그 수집방법에 따라 標本調査(sample survey)에 의한 調査資料(survey data)와 計劃된 實驗에 의한 實驗資料(experimental data)로 생각할 수 있다. 어떤 자료를 수집해야 할 것인지는 研究 目的에 따라 다르겠지만 여기서는 實驗資料를 다루기로 한다.

實驗을 企劃할때 研究對象이 되는 因子(들)의 水準 또는 水準組合을 處理로 하여 實驗을 計劃하게 되며 Cox(1958)는 훌륭한 실험의 필요 요건으로 處理를 比較할때 (1) 體系的 誤差(systematic error)가 없어야 하고, (2) 精度가 充分히 높아야 하며, (3) 실험에서 도출된 결론은 폭넓은 妥當性이 있어야 하고, (4) 실험 배치는 가급적 간단할수록 좋고, (5) 결론의 불확실성을 평가할 수 있어야 한다고 하였다.

지금까지 統計學的 側面에서 여러가지 實驗計劃法이 研究 發展되고 있고 應用面에서도 폭넓게 強調되고 있으나 실제로 統計的 方法이나 實驗計劃法을 이용 또는 응용하는 일반연구자들이 실험을 企劃(planning)하고 計劃된 實驗을 遂行하는 과정에서 각 計劃法에 대한 理解不足으로 處理의 配列法, 確率化 또는 實驗遂行 과정에서의 기계적인 誤謬를 범하는 事例를 흔히 볼 수 있다.

실험계획에 있어서 가장 보편 타당성이 있고 공정한 처리의 효과와 이들 처리 효과를 검증하는데 필요한 실험오차를 공정하게 추정하기 위한 몇가지 원칙을 살펴봄으로써 실험을 계획하고 수행하는 이용자 또는 연구자들에게 이해를 돕고자 한다.

---

\* 계명대학교 자연과학대학 통계학과 교수

### 1. 確率化의 原理 (Principle of Randomization)

처리 이외의 외적인 영향이 실험 결과에 偏倚的으로 미치는 것을 배제하여 공정한 처리 효과와 실험오차를 추정하기 위하여 모든 처리가 각 실험단위에 배치되는 확률을 강제하여 배열하는 방법으로 R.A.Fisher(1926)가 그 理論을 정립하였다. 그는 뜻실험(Pot experiment)에서 2개의 씨앗을 뜻트에 심을때 씨앗을 심는 사람의 주관적인 영향을 배제하기 위하여 난수표에 의하여 씨앗을 배열할 만큼 세심한 주의를 기울였다고 한다.

처리를 실험단위에 배열할 때 있을 수도 있는 체계적오차(systematic error)를 확률화에 의하여 제거할 수 있으며 확률화 기법을 올바르게 적용한 실험의 결과에서만 처리간의 차에 대한 유의성을 검정하거나 신뢰구간을 추정할 수 있다.

### 2. 反復의 原理 (Principle of Replication)

실험의 규모(주로 처리수)가 아무리 커도 각 처리를 한번씩만 실험하면 실험오차를 구할 수 없기 때문에 가설 검정이 불가능하다. 실험오차는 동일한 처리를 되풀이하여 실험할 때 나타나는 차이로 정의되므로 실험오차를 구하기 위해서는 각 처리를 적어도 2회 이상 실험을 하여야 하며 이때 동일한 처리를 실험하는 순서는 완전히 확률화에 의하여 결정되어야 한다. 따라서 실험오차를 구하기 위하여 반복은 꼭 필요한 요건이다. 실험재료가 균일하거나 실험오차가 작은 특성에 대하여 실험할 경우에는 반복수를 적게할 수도 있고 경우에 따라서는 모든 처리를 반복하지 않고 精度 높은 추정을 요하는 처리만을 몇번 반복함으로써 이들에서 추정된 오차를 실험오차로 이용하기도 한다. 또 中心合成計劃(central composite design)에서와 같이 반복되는 中心點에서 실험오차를 구하여 필요한 검정을 할 수 있다. 추정된 실험오차를  $s^2$ 라고 할때  $r$ 번 반복 실험한 처리평균  $\bar{y}_i$ 의 표준오차는  $s(\bar{y}_i) = s / \sqrt{r_i}$  이고 두 처리평균차에 대한 표준오차는  $s(\bar{y}_i - \bar{y}_i') = s \sqrt{(1/r_i + 1/r_i')}$  이므로 반복수  $r_i$ 를 크게 하므로

써 精度 높은 처리의 효과를 검출할 수 있다. 그러나 필요한 반복수는 실험자가 원하는 精度와 費用에 따라 결정될 문제이다.

### 3. 實驗誤差를 줄이는 方法(Local control of error)

주어진 여건하에서 처리의 효과를 정확히 추정하고 처리 평균의 차(差)를 精度 높게 검정하기 위해서는 표준편차  $s$ 를 줄여야 하고 따라서 허용되는 범위 안에서 실험오차를 줄일 수 있다면 보다 정확하고 정도 높은 추론을 기대할 수 있을 것이다.

실험오차가 크면 처리의 효과를 올바르게 검정할 수 없는 것은 당연하지만 그렇다고 실험오차가 터무니 없이 작게 추정되었다고 해서 정확한 것은 아니다. 문제는 어떻게 하면 母數  $\sigma$ 에 가까운 표본표준편차  $s$ 를 추정하느냐 하는 것이다. 일반적으로  $s$ 가  $\sigma$ 의 불편추정량이라 하면  $s$ 가 작을수록 처리의 효과를 정도 높게 검정할 수 있다. 따라서 실험계획에 있어서 실험오차를 줄이는 방법은 중요한 요건이므로 이들 방법을 생각해 보기로 하자.

#### (1) 블럭화(Blocking)

실험재료를 균일한 것끼리 그룹(集團)을 짓거나 실험환경(또는 조건)을 균등하게 만드는 것을 블럭화(blocking)라고 하며 균일한 실험재료의 집단이나 조건이 균등한 실험환경을 블럭(block)이라고 한다. 블럭이란 말은 농업실험에서 肥沃度가 均一한 實驗園場의 한 區劃을 블럭이라 하고 이 블럭을 處理區로 세분하여 실험하는 계획법에서 유래되었다고 한다. 블럭내에서의 모든 조건은 가능한 한 균일하게 하고 블럭간에는 차이를 크게 하는 것이 일반적인 블럭화 방법이다.

따라서 블럭간의 차이는 처리의 효과에 영향을 미치지 않으므로 블럭내에서 처리의 효과를 정도 높게 검출할 수 있으며 이와 같이 동일한 블럭에 모든 처리가 한번씩 나타나도록 랜덤으로 배열하는 몇 개의 블럭에서 실험하는 방법을 亂塊法(Randomized Complete Block Design)이라 하고 블럭화를 두 방향 즉 行과 列方向으로 하여 각 처리

를 한번씩 나타나도록 배열하는 方法을 라틴방격법(Latin Square Design)이라고 한다.

### (2) 交絡法(Confounding)

要因 實驗에서 要因數와 水準數가 늘어남에 따라 처리 조합수는 기하급수적으로 증가하고 이들 처리를 한 블록 또는 동일한 실험조건하에서 실험을 할 수 없는 경우가 허다하다. 또 비록 한 블록에서 실험을 할 수 있다 하더라도 블록의 크기가 커지므로 같은 블록내의 실험환경이 균일하지 못하게 된다. 따라서 처리의 효과를 정확하게 추정할 수 없게 된다. 이러한 경우에는 블록화의 원리에 따라 실험재료나 실험환경이 균일한 몇개의 블록으로 만든후에 요인처리 조합을 적당한 방법에 의해 각 블록에 배열하여 실험하므로써 정도 높은처리 효과를 추정할 수 있다. 이때 요인 처리 조합을 작은 블록에 배열함에 따라 情報의 一部를 희생시키는 대가를 치르지 않으면 안된다. 즉 情報의 一部 또는 全部를 블록의 효과와 함께 처리조합을 블록에 배열하는 방법으로 Yates(1937)가 처음 제안하였다. 일반적으로 요인실험에서 가장 고차적인 교호작용을 블록과 交絡시키지만 중요하지 않은 情報(주로 2次이상의 交互作用)을 블록과 交絡시키는것이 보통이다. 예를 들면 2<sup>n</sup> 계열 요인 실험에서 블록의 크기를 2<sup>m</sup>라하면 블록수는 2<sup>n-m</sup>이고 블록의 자유도 2<sup>n-m</sup> - 1에 해당하는 교호작용이 블록과 交絡되어야 하므로 因子의 主效果, 1次교호작용 또는 중요하다고 생각되는 交互作用이 블록과 交絡되지 않도록 잘 선정하여야 한다.

### (3) 直交性(Orthogonality)

處理計劃(treatment design)을 수립할 때 直交性을 갖도록 計劃하면 같은 횟수의 실험으로 檢出力(power of test)을 높일 수 있고 정도 높은 추정을 기대할 수 있다. 몇 개의 因子가 포함되는 실험에서는 完全要因組합을 처리로 하거나 또는 一部實施法(fractional replication)을 이용하므로써 直交性을 유지하고 因子의 水準을 설정할 때 수준간격을 같게 하므로써 直交性을 유지할 수 있다.

(4) 기타 方法

실험오차를 줄이는 방법으로 실험계획법 또는 처리의 배열 방법 이외에 자료분석방법과 실험수행의 기술적인 면을 생각할 수 있다.

(i) 共分散分析法(Covariance analysis)

처리의 영향을 받는 반응변수(response variable)와 밀접한 관계를 갖는 補助變數(auxiliary variables)를 조사하여 실험을 완료한 후에 이들 보조변수를 이용하여 共分散分析을 하므로써 보조변수의 영향을 제거하고 순수한 처리의 효과를 검정할 수 있다. 이때 보조변수는 실험하기 전에 각 실험 단위별로 조사되어야 한다.

(ii) 精巧한 실험수행

실험을 수행하는 것은 실험계획과는 별개의 문제이나 실험 수행 과정에서 실험오차를 증폭시키는 여러가지 요인을 생각할 수 있다. 즉 한 블럭 또는 한 반복의 처리를 여러사람이 실험하거나 몇 일에 걸쳐서 실험을 하면 실험하는 사람에서 오는 차이, 실험일이 다르기 때문에 온도, 습도 또는 기압등의 기후조건 그리고 실험하는 사람의 기분등 여러가지 요인이 처리의 효과와 실험오차에 미치는 영향은 무시할 수 없을 정도로 클 수 있다. 따라서 실험을 수행할 때는 가급적 한사람이 하루에 실험하는 본량을 블럭 또는 반복단위로 계획하여 실험하므로써 이들 요인을 제거할 수 있을 것이다.

結論적으로 實驗計劃에 있어서 가장 중요한 것은 公正한 處理의 效果와 實驗誤差를 推定하는데 있으나 一般적으로 實驗材料의 均一性, 豫算, 人力 또는 時間的인 制約을 받기 때문에 精度만을 생각할 수 없게 된다. 그러나 주어진 與件下에서 앞에서 說明한 諸原則과 誤差를 줄이는 方法으로 최소의 資源을 利用하여 最大의 情報를 얻을 수 있도록 實驗計劃을 수립해야 하며 통계적인 지식 특히 實驗計劃法에 대한 知識이 없는 실험자는 반드시 實驗을 수행하기 전에 統計專門家와 相談을 하는 것이 바람직하다.