Design of experiment

Lichen

December 28, 2021

1 실험설계 개요

실험을 통해 향상시키고자 하는 ***response*** y 를 통제 가능한 변수 x, 통제 불가능한 변수 z의 조합 그리고 오류항 ϵ으로 나타낼 수 있다.

$$\begin{bmatrix} y_0 \\ y_1 \\ \cdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} \beta_{i0} x_i \\ \sum_{i=0}^{n} \beta_{i1} x_i \\ \cdots \\ \sum_{i=0}^{n} \beta_{im} x_i \end{bmatrix} + \begin{bmatrix} \sum_{i=0}^{n} \sum_{j=0}^{n} \beta_{i0} x_i x_j \\ \sum_{i=0}^{n} \sum_{j=0}^{n} \beta_{i0} x_i x_j \\ \cdots \\ \sum_{i=0}^{n} \sum_{j=0}^{n} \beta_{i0} x_i x_j \end{bmatrix} + \begin{bmatrix} \sum_{i=0}^{l} \alpha_{i0} z_i + \epsilon_0 \\ \sum_{i=1}^{l} \alpha_{i1} z_i + \epsilon_1 \\ \cdots \\ \sum_{i=m}^{l} \alpha_{im} z_i + \epsilon_m \end{bmatrix} \tag{1}$$

실제 설계에서는 ***interation*** 을 나타내는 항 $\sum_{i=0}^{n} \sum_{j=0}^{n} \beta_i x_i x_j$ 을 구하기 위해 2^k 번의 실험을 진행해야 하기 때문에 모든 경우의 수를 다루는 것이 불가능하다.

통제 불가능한 요소가 있기 때문에 y는 확률 변수로 표현된다.

$$P(y = y_i) = p(y_i) \tag{2}$$

$$\int_{\infty}^{-\infty} p(y_i)\, dx = 1 \tag{3}$$

실험 설계는 다음 세가지 원칙을 따른다.

- **randomization** : 실험 순서를 무작위로 배치하는 것
- **replication** : 실험을 반복하는 것
- **blocking** : 통제 불가능한 요소가 동일한 조건에서 실험을 진행하여 해당 요소의 영향을 상쇄하는 것

2 rescaling

실험에 사용하는 요소의 갯수를 축소하는 기법 일반적인 경우 매개변수를 다음과 같이 축소할 수 있다.

$$\frac{dN}{dt} = \kappa(C_0 - \alpha N)N$$
$$\frac{dN}{dt} = ((\kappa C_0) - (\kappa\alpha)N) = (\widetilde{C_0} - \widetilde{\alpha}N)N, \quad (\widetilde{C_0} = \kappa C_0\, \widetilde{\alpha} = \kappa\alpha)$$

반응 변수와 시간을 조정하면 하나의 매개변수 방정식으로 바꿀 수 있다.

$$
\begin{aligned}
&N = \hat{N}N^*, t = \hat{t}t^* \\
&\frac{dN}{dt} = \kappa(C_0 - \alpha N)N \\
&\frac{\hat{N}N^*}{\hat{t}t^*} = \kappa(C_0 - \alpha\hat{N}N^*)\hat{N}N^* \\
&\Rightarrow \frac{dN^*}{dt^*} = (\kappa\hat{t}C_0 - \kappa\hat{t}\alpha\hat{N}N^*)N^* \\
&\Rightarrow \frac{dN^*}{dt} = (1 - N^*)N^*, \hat{t} := \frac{1}{\kappa C_0}, \hat{N} := C_0\alpha
\end{aligned}
$$

3 simple comparative experiment

한 ***factor***에 대해 두 ***treatment***가 있는 경우 ***treatment***가 ***response***에 미치는 영향을 비교.

- ***Mean :*** $\mu = \int_{-\infty}^{\infty} yf(y)\,dy = E\,(y)$
- ***Variance :*** $\sigma^2 = \int_{-\infty}^{\infty}(y-\mu)^2 f(y)\,dy = E\left[(y-\mu)^2\right] = V(y)$
- ***Cov*** : $Cov(y_1, y_2) = E\left[(y_1 - \mu_1)(y_2 - \mu_2)\right]$
- $\mu^2 + \sigma^2 = \sum_{i=1}^{n} y_i^2$

sample mean 과 ***sample variance***는 모집단에서 추출한 ***random sample*** 으로 가정.

$$
\begin{aligned}
\text{sample mean} = \bar{y} &= \frac{\sum_{i=1}^{n} y_i}{n} \\
\text{sample variance} = S^2 &= \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \\
E(\bar{y}) &= E\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) \\
&= \mu \\
E(S^2) &= E\left[\frac{\sum_{i=1}^{n} (y - \bar{y})^2}{n-1}\right] \\
&= \frac{1}{n-1} E(SS) \\
E(SS) &= E\left[\sum_{i=1}^{n} (y_i - \bar{y})^2\right] \\
&= E\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right] \\
&= (n-1)\mu^2 \\
E(S^2) &= \frac{1}{n-1} E(SS) = \mu^2
\end{aligned}
$$

Degree of Freedom은 독립인 요소의 수로 결정됨. $E(SS) = v\sigma^2$

- T-test
- ANOVA

4 randomized Block

factor에 대해 ***Block***에서 가능한 모든 ***treatment***의 조합을 실험하는 경우 블럭 b개 에서 a개의 처리가 가능한 경우

$$
y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad \begin{cases} i = 1, 2, 3, \cdots a \\ j = 1, 2, 3, \cdots b \end{cases}
$$

5 Latin square Design

2개의 ***factor***에 대해 ***Block***에서 가능한 모든 ***treatment***의 조합을 실험하는 경우 각 요소에 대해 한번씩 처리가 나타나도록 배치하여 두가지 변동요소를 제거

$$
y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk} \qquad \begin{cases} i = 1, 2, 3, \cdots p \\ j = 1, 2, 3, \cdots p \\ k = 1, 2, 3, \cdots p \end{cases}
$$

Latin square Design은 Greco-Latin square Design 으로 확장가능

6 Factorial Designs

6.1 Randomzied Complete Block Design RCBD

6.2 Two-factor Factorial Design

6.3 General Factorial Design

6.4 2^k Factorial Design

6.5 Two-Level Factional Factorial Design

7 Fitting Regression Model

8 Response Surface Methods and Designs

9 Nested Design

10 Split-plot Design