



실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

<http://jupiter.hallym.ac.kr>

실험계획과 분산분석

Data 의 종류

- ➡ Experimental Data (실험계획법)
 - ➡ Observational Data (회귀분석, 시계열, 다변량, etc)
 - ➡ Survey Data (표본조사론)
-
- 실험단위 (Experimental Unit) : 실험조건이 행해지는 최소단위 (e.g. 기계, 논, 환자, ...)
 - 실험오차 (Experimental Error): 같은 실험조건을 받은 다른 실험단위 사이에 발생하는 자연스런 오차
 - 처리군 (treatment group) : 관심있는 실험조건에 노출된 집단
 - 대조군 (control group) : 처리군과 달리, 관심있는 실험조건에 노출되지 않은 집단

실험계획과 분산분석

척도(scale)의 종류

- 연속형 척도 (Continuous Scale)
 - ✓구간척도(interval scale) : 온도
 - ✓비율척도(ratio scale) : 길이, 무게
- 이산형 척도 (Discrete Scale)
 - ✓명목척도 (nominal scale): 성별
 - ✓순서척도 (order scale): 5점 척도의 만족도 등

또는

- 양적 척도 (quantitative Scale)
 - ✓연속형(continuous) : 온도, 길이, 무게
 - ✓이산형(discrete scale) : 불량품의 수, 자동차 통행량
- 질적 척도 (qualitative Scale)
 - ✓명목척도 (nominal scale): 성별
 - ✓순서척도 (order scale): 5점 척도의 만족도 등

실험계획과 분산분석

변수의 종류

- Response Variable(반응변수) : 실험을 통해 개체(subject)로부터 반응을 측정하고자 하는 변수
- Explanatory Variable(설명변수) : 반응변수에 영향을 줄 것으로 예상되는 실험에 주관심이 되는 변수
- Extraneous Variable(외생변수) : 실험의 주관심은 아니지만 반응변수에 영향을 다소 미칠 것으로 우려되는 노이즈(noise)변수

외생변수의 영향을 최소화 시키면서, 반응변수에 끼치는 설명변수의 영향 조사

실험계획과 분산분석

실험(Experiment)

- 실험(Experiment) vs. 준실험(Quasi-Experiment)
- 선택편의(Sampling Selection Bias) : 표본조사에서 모집단을 정상적으로 대표하지 못하고 특정계층만 대표 하는 선택에 의해 발생한 편의(bias)
- 준실험: 선택편의가 존재하는 실험
- 선택편의가 없기 위한 방법 = 랜덤배치 (random allocation)

실험계획과 분산분석

랜덤배치에 의한 실험계획

- 위약효과 (Placebo Effect) : 가짜약을 처방받은 환자가 심리적 요인으로 실제와 비슷한 효과가 생기는 효과
- 맹검 (blind test) : 약을 처방받은 환자가 진짜약인지 가짜약인지 모르게 하는 시험
- 이중맹검 (double blind test) : 약을 처방받은 환자뿐만 아니라 약을 주고, 측정치를 관찰하는 사람도 환자가 처방받은 약이 진짜인지 가짜인지 모르게 하는 시험

실험계획과 분산분석

Vitamin-C effect on flu ?

- Response Variable: $y=???$
- Explanatory Variable: $x=??$
- Extraneous Variable : $???$

외생변수의 영향을 최소로 하면서,
반응변수에 끼치는 설명변수의 영향 조사

실험계획과 분산분석

실험계획의 단계

1. What do you want from your data?
2. Define (response variable : 반응변수) & unit
3. Define (explanatory variables: 설명변수) & unit
4. List the possible (extraneous variables: 외생변수) ?
5. What is the method of statistical analysis?

실험계획의 기본원리

- 반복의 원리 (replication)
- 랜덤화의 원리 (randomization)
- 블록화의 원리 (blocking)

실험계획과 분산분석

서로 다른 두 개의 다트제품을 서로 비교하기 위한 실험계획



- 같은 실험조건에 개체수를 증가시킴으로써 개체가 가지고 있는 외생변수의 영향을 상쇄시킨다
- 실험단위를 처리군/대조군에 랜덤배치함으로써 개체가 가지고 있는 외생변수의 영향을 상쇄시킨다
- 같은 외생변수 값을 지닌 실험단위를 블록화시킴으로써 블록내 실험단위를 비교함으로써 외생변수의 영향을 상쇄시킨다.

실험계획과 분산분석

비교할 다트가 2개가 아니라 3개라면?



Tungsten



Nickel Silver



Brass