

Data Snooping : Testing은 다중하게 당색한 결과, 실제 해하지 않은 Pattern을 발견하여 잘못된 결론 도출하는 경우) Multiple testing에서

ex) Sample-size = 800, P-value = 1% 라면,
8개의 유의미한 결과가 무의미하게 발생할 수 있다.

→ 이를 Multiple testing fallacy 또는 Look-elsewhere effect라고 한다.

Data Snooping은 Replicability의 가능성으로 보인다.

- Replicability ↑

- 다른 Sample, 학, 분석 방법도 사용해도 비슷한 결과가 나오는 것

- Reproducibility ↓

- 같은 데이터와 같은 방법을 사용했을 때 같은 결과가 나오지 않는다.

위 문제를 해결해보기 : Multiple testing에서 Type I (False-Positive, 귀가설 기각 오류) 해결
H₀의 레이트를 H₀의 레이트로 잘못 분류 (100개 중 각 test가 관련된 가설 선택)

1. Bonferroni Correction

a. Multiple test = 1 개가 m개일 때, 기존 P-value를 m배해서 보정한다.

$$P_{adjusted} = P \times m$$

b. 새로운 유의수준 설정

$$\alpha' = \frac{\alpha}{m}$$

c. $P_{adjusted} \leq \alpha'$ 라면 reject Null hypothesis

2. False Discovery Proportion (FDP)

$$FDP = \frac{\text{number of false discoveries}}{\text{Total number of discoveries}}$$

Discovery

: Alternative hypothesis를 선택한 경우

ex) 1000개의 testing 중 Discovery가 121개, False Discovery (Type I error)가 41개 있는 경우

→ $FDP = 0.34$ FDP를 최대한 낮추는 것이 좋다. ex) Sample size ↑

3. False Discovery Rate (FDR)

Benjamini-Hochberg procedure : FDR을 조절

1. m개 testing에서 P-values를 오름차순으로 정렬

$$P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$$

2. $P_{(m)} \leq \frac{k}{m} \alpha$ 를 만족하는 k 찾기

→ 해당 test가 k번 정도가 유의하다

3. $i = 1$ to $i = k$ 까지의 test를 유의미한 결과로 선택