더미변수 생성 및 Sample Size와의 연관성

❑ 더미변수의 특성

- 특정변수에서 더미변수를 생성할 경우, 더미변수가 여러 개 생성될 수 있다.
- 더미변수는 반드시 0과 1로 Value를 지정하게 된다.
- 더미변수의 개수는 (특정변수의 범주의 개수 – 1) (개)를 생성하게 된다.
- 성별의 경우, 남자는 1, 여자는 2로 코딩이 되어 있을 때, 성별(남자여부)라는 더미변수는 남자=1, 여자=0로 Recode하게 된다.
- 학년의 경우, 1학년=1, 2학년=2, 3학년=3로 코딩되어 있을 때, 더미변수의 개수는 2개여야 한다.
  - → 학년(1학년 여부)라는 변수는 1학년=1, 2~3학년=0
  - → 학년(2학년 여부)라는 변수는 2학년=1, 1학년과 3학년은 0
- 성별(남자여부)라는 더미변수는 남자=1, 여자=0로 Recode하게 될 경우,
  - → 여자는 참조범주가 된다.
  - → 남자는 비교범주가 된다.
  - → 여자 대비 남자는…
- 더미변수 생성의 의미는 참조범주를 정의하는 데 의의가 있다.

❑ 회귀모형 상에서의 적정 더미변수 개수

- 많은 연구에서 더미변수를 무분별하게 적용하는 사례가 있음. 회귀모형 상에서 **더미변수가 많아지는 것은 문제가 되지 않으나, 많은 더미변수를 투입하기 위해서는 그 만큼 큰 데이터 Set 을 기반으로 해야 함.**
- 만약, 성별(남녀), 학년(1~4 학년), 교육참여여부(참여 및 비참여)를 더미처리하여 모형에 투입하고자 할 때, **성별은 1 개의 더미변수, 학년은 3 개의 더미변수, 교육참여여부는 1 개의 더미변수를 생성하게 된다. 이럴 경우, 총 5 개의 더미변수가 생성되는데 각각의 경우의 수를 따지게 되면 2^5=32 개의 경우의 수가 도출된다. 즉, 32 개의 모형이 도출될 수 있는데, 각 모형 당 30 개의 데이터가 보장이 된다고 하더라도 약 900 개 이상의 데이터가 필요로 하게 된다.**

#1422, Officetel 3 Danji, Gyeonghuigungeui Achim
72 Naesu-dong, Jongno-gu, Seoul 110-872, South Korea
Tel_+82-2-6271-7140~1
URL_ www.stateer.com