

“본 강의 동영상 및 자료는 대한민국 저작권법을 준수합니다. 본 강의 동영상 및 자료는 상명대학교 재학생들의 수업목적으로 제작·배포되는 것이므로, 수업목적으로 내려받은 강의 동영상 및 자료는 수업목적 이외에 다른 용도로 사용할 수 없으며, 다른 장소 및 타인에게 복제, 전송하여 공유할 수 없습니다. 이를 위반해서 발생하는 모든 법적 책임은 행위 주체인 본인에게 있습니다.”



AI+X선도인재양성기초프로젝트

14. 회귀분석 참고

Acknowledgement:

한서경, 상관 및 회귀분석, 의학통계론, 서울대학교

Heenam Yoon

Department of
Human-Centered Artificial Intelligence

E-mail) h-yoon@smu.ac.kr
Room) 0112



회귀 분석

가변수 (Dummy) 설정

- 변수가 두 가지 이상의 범주를 가지며 순위형이 아닌 경우 가변수를 사용. e.g. 비만도: 저체중/정상/과체중
- 위와 같이 세 가지의 범주로 이루어지는 경우 두 개의 가변수를 설정

	D1	D2
정상	0	0
저체중	1	0
과체중	0	1

회귀 분석

가변수 (Dummy) 해석

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_3 D_2$
 - X_1 : 연속변수
 - D_1 : 가변수 1
 - D_2 : 가변수 2
- 상기 회귀모형으로부터 세개의 모형이 성립
 - 정상: $Y = \beta_0 + \beta_1 X_1$
 - 저체중: $Y = \beta_0 + \beta_1 X_1 + \beta_2$
 - 과체중: $Y = \beta_0 + \beta_1 X_1 + \beta_3$

회귀 분석

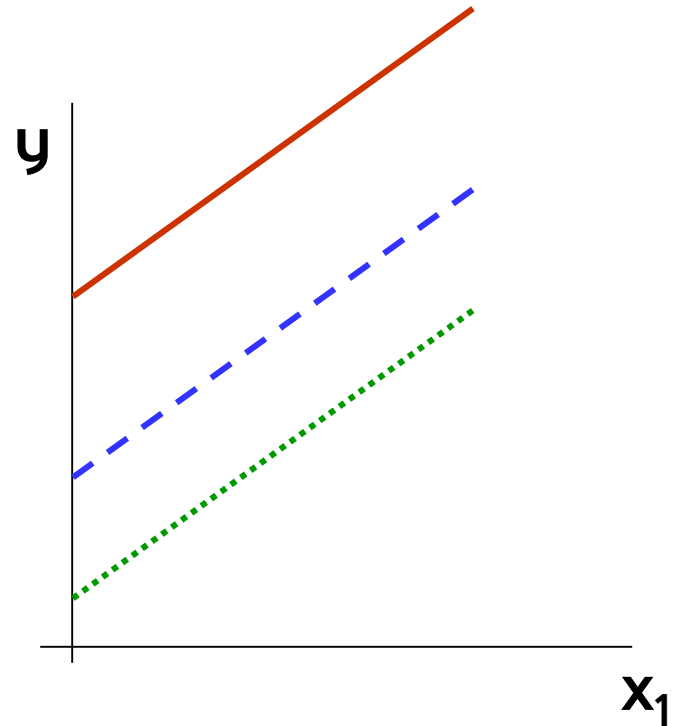
가변수 (Dummy) 해석

- 상기 회귀모형으로부터 세개의 모형이 성립

- 정상: $Y = \beta_1 X_1 + \beta_0$
- 저체중: $Y = \beta_1 X_1 + \beta_0 + \beta_2$
- 과체중: $Y = \beta_1 X_1 + \beta_0 + \beta_3$

기울기 같음

절편 다름



회귀 분석

예시: 교란효과

- 결과변수: BP

- X_1 : 연령

- X_2 : 성별을 표시하는 가변수 (male=1, female=0)

- 회귀식: $BP = 30 + 1.5 X_1 + 10 X_2$

- 여성: $30 + 1.5 \times \text{연령}$

- 남성: $30 + 1.5 \times \text{연령} + 10$

회귀 분석

예시: 교란효과 (해석)

- 나이와 혈압은 연관이 있으며, 연령이 한 해 높아질 때 마다 1.5 mmHg 씩 증가하게 되며, 이러한 경향은 남녀에 동등하게 적용되나 남성의 경우 연령당 10 mmHg 씩 여성에 비해 높음
- 1.5 mmHg의 상관성은 성별에 대해 '보정'이 된 후에 나타나는 연관성으로 해석, 즉, 남녀 상관없이 연령 1세 당 1.5 mmHg 증가

회귀 분석

교호작용 (Interaction)

- 설명변수와 독립변수 간의 관계가 한 변수의 그룹에 따라 달라질 수 있음
 - 나이와 심혈관 기능 간의 연관성이 남녀 간에 차이가 있음
- 남녀 간에 통계적으로 유의한 차이가 있는지 없는지를 검정하기 위해 모형에 교호작용에 대한 설명변수를 추가

회귀 분석

교호작용 (Interaction)

- 예. 성별에 따른 혈압의 증가 양상이 남녀에 따라 다르다면?

- 중회귀모형에 교호작용에 대한 효과를 추가

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \delta X_1 X_2$$

- 혈압에 대한 연령의 효과를 성별에 따라 다르게 추정

회귀 분석

교호작용 (Interaction)

- 혈압 = $30 + 1.5 \times \text{연령} + 10 \times \text{성별} + 0.5 \times \text{연령} \times \text{성별}$
= $30 + 1.5 \times \text{연령}$
+ $10 \times (1 \text{ for M and } 0 \text{ for F})$
+ $0.5 \times \text{연령} \times (1 \text{ for M and } 0 \text{ for F})$
= $30 (+10 \text{ for M}) + (2 \text{ for M and } 1.5 \text{ for F}) \times \text{연령}$
- 여성의 경우 한해 연령증가에 대한 혈압상승률이 1.5mmHg일 때 남성의 경우 $1.5 + 0.5 = 2\text{mmHg}$ 로 성별에 따른 연령에 대한 혈압상승률이 달라짐
- 남성에게서는 연간 2mmHg/yr 증가할 때 여성에게서는 연간 1.5 mmHg/yr가 증가