회귀분석실습 기말고사

2015. 12. 17.

**※ 유의수준은 항상 5%로 계산한다.**

1. htwt.txt 파일 사용하여 work.htwt 데이터 셋을 생성하고, 다음 물음에 답하여라.

Sex : 성별 (남 : M, 여 : F)  
 age : 나이  
 height : 키 (cm)  
 weight : 몸무게 (kg)

(1) 모든 변수를 사용하여 weight를 반응변수로 하는 회귀모형을 적합하고 추정회귀식을 작성하여라. [5점]

(2) (1)의 모형으로부터 성별이 유의한지 검정하기 위한 가설과 Full Model, Reduced Model을 작성하고 검정한 결과에 대해서 해석하여라. [10점]

(3) (1)의 모형으로부터 sex는 M이고 age와 height가 각각 16, 160인 사람에 대한 predicted value의 C.I.와 expected value의 C.I.를 구하여라. [15점]

2. acetyl.txt 파일을 사용하여 work.acetyl 데이터 셋을 생성하고, 다음 물음에 답하여라.

temp : 원자로 온도 (celsius)  
 ratio : n-heptone에 포함된 h2의 비율  
 time : 접촉 시간 (sec)  
 conv : 전환율 (%)

(1) 모든 변수를 사용하여 conv를 반응변수로 하는 회귀모형을 적합하고 Correlation matrix와 Variance inflation factor를 구하여라. 또한, 그에 따른 결과에 대해서 해석하여라. [10점]

(2) temp와 ratio를 사용하여 conv를 예측하는 추정회귀식을 쓰고 temp와 ratio 사이에 interaction이 존재하는지 검정하고 가설과 Full Model, Reduced Model을 작성하여라. 만약, interaction이 존재한다면 해당 변수를 추가하여 모형을 재적합하여라. [10점]

(3) (2)의 모형으로부터 studentized residual plot을 그리고 어떤 추세가 존재하는지 설명하여라. [10점]

3. fitness.txt 파일 사용하여 work.fitness 데이터 셋을 생성하고, 다음 물음에 답하여라.

Age : 나이  
 Weight : 몸무게 (kg)  
 Oxygen : 산소흡수율  
 RunTime : 1.5 mile을 달리는데 걸린 시간 (min)  
 RestPulse : 쉬는 동안 심박수  
 RunPulse : 달리는 동안 심박수  
 MaxPulse : 달리는 동안 최대 심박수  
 Index : 관측치 번호

(1) Index를 제외한 모든 변수를 사용하여 Oxygen을 반응변수로 하는 회귀모형을 적합하고 추정회귀식을 작성하여라. [5점]

(2) (1)의 모형으로부터 Cook’s D, DFFITS, Hi를 사용하여 high leverage point와 outlier에 해당하는 관측치 번호를 구하고, 그때의 Cook’s D, DFFITS, Hi 값을 구하여라. [15점]

(3) (2)에서 구한 이상치를 제거한 후 회귀모형을 적합하고 해당 점이 influential point인지 서술하여라. [10점]

(4) (1)의 모형으로부터 forward, backward, stepwise selection을 각각 수행하고 최종 모형을 작성하여라. 변수 선택과 제거 시 유의수준은 5%로 한다. [10점]  
(유의수준 설정 옵션은 model y=x / SELECTION=option SLENTRY=0.05 로 설정하면 된다.)

한 학기 동안 수고하셨습니다^^