[회귀분석실습] 기말고사 (2014.12.17)

**\*\*\*\* 검정의 경우, 사용하는 모델이 무엇이고, 모델에서 검정하고자 하는 가설이 무엇인지, 사용하는 검정통계량에 대한 기술이 있어야 한다. 유의수준은 5%로 하고, 최종 답의 유효숫자는 소수점 아래 4자리로 하시오.**

#1. [자료 q1.csv] 어느 기업 내 직원들의 급여를 설명하기 위하여 다음과 같은 변수들이 자료가 주어져 있다. 다음 물음에 답하여라. (뒷장의 모델1~모델7을 이용하여 검정 시 모델을 제시하시오.)

x1 : 경력 (개월수)

e1 : 최종학력이 대졸 (1-대졸, 0- 그 이외)

e2 : 최종학력이 초대졸 (1-초대졸, 0- 그 이외)

s : 성별 (1- 남자, 0- 여자)

j : 직군 (1- 임원, 0- 일반사원)

1. 모델 4가 유의한지 검정하여라.(10점)
2. 직군과 성별이 같다면, 경력이 증가할 때 평균급여가 15만원보다 더 증가한다고 할 수 있는지 검정하여라. (10점)
3. 학력과 직군이 같을 때, 성별에 따라 급여에 차이가 있는지 검정하여라. (10점)
4. 경력이 같을 때 학력에 의해 생기는 연봉의 차이가 남자와 여자에서 다르게 나타나는지 검정하여라. (10점)
5. 모델 4 에서 를 검정하고, 그 의미를 설명하여라. (10점)
6. 모델 7을 적합시킨 결과를 이용하여 다음 각 경우의 값을 추정하라.
   1. 신입사원 대졸 남성의 평균 현재급여 (5점)
   2. 경력이 1년 증가할 때, 초대졸 임원급 여성의 평균 급여의 변화량 (5점)
7. 모델5 의 잔차플랏(표준화잔차(ti) 대 적합값)을 그린 후(답안지에 제시), 어떠한 문제가 있는지 서술하여라. 또한 그 문제를 어떻게 해결할 수 있는지 서술하여라. (10점)
8. 모델4, 모델5, 모델6, 모델7을 비교하고자 한다.
   1. 기준으로 어떤 모델이 가장 설명력이 좋은가? (모델별 제시) (5점)
   2. 기준으로 어떤 모델이 가장 설명력이 좋은가? (모델별 제시) (5점)

#2. [자료 q2.txt] 주어진 자료에 대한 다음 물음에 답하라.

1. 독립변수 모두를 포함하는 모델에서 VIF를 구하여 다중공선성의 문제가 있는지 판단하여라. (10점)
2. (1)에서 구한 VIF2(x2에 대한 VIF)가 무슨 뜻을 갖는지 설명하여라. (10점)
3. 후진제거법 (Backward elimination)을 이용하였을 때 선택되는 변수를 쓰시오. (10점)
4. (2)번에서 얻은 모델에서 관측치들 중 영향점(influential point)를 찾고자 한다.

다음 값들을 구하여, 영향점을 제시하여라. (ti(r-student residual), hii, dffits, cooks.distance) (10점)

#3. 같은 MSE을 갖는 두 모델이 있다. 하나는 bias가 크고, 또 다른 하나는 variance가 크다.

이때 두 모델의 차이를 서술하여라. (10점)

문제 1번 모델 제시