

Problem Set 5

계량경제학

남 준우 교수

1. 아래의 각 문장이 맞는지 틀리는지를 판단하고 그 이유를 정확히 그러나 간략히 설명하라.

- (1) 독립변수 X 의 분산이 크면 회귀계수 β_2 에 대한 신뢰구간은 넓어진다.
- (2) 오차항 ε_i 가 정규분포를 가지지 않으면 최소자승추정량은 BLUE가 아니다.
- (3) 예측에 있어서 만일 모집단 회귀함수의 절편과 기울기 β_1, β_2 의 값과 미래의 독립변수의 값을 정확히 알면 예측오차의 분산은 0과 같다.
- (4) 독립변수 X 의 분산이 크면 클수록 회귀계수 β_2 의 추정치로서 최소자승추정량 b_2 의 정확도는 떨어진다.
- (5) 독립변수 X 의 값들이 이의 표본평균 \bar{X} 에 가까울수록 최소자승법에 의한 기울기는 더욱 더 정확하게 추정된다.
- (6) 오차항의 분산이 크면 클수록 회귀계수의 최소자승추정량의 분산은 커진다.
- (7) 유의확률의 값이 크다는 것은 추정치가 유의적으로 0과 다르다는 것을 의미한다.
- (8) 오차항이 정규분포를 가진다는 가정은 최소자승추정량의 분산에는 아무런 영향이 없다.

2. 다음의 회귀분석 결과를 고려하자.

$$\hat{y}_i = 12 + 0.000045X_i, \quad R^2 = 0.40, \quad n = 12$$

(2.5) (0.000015)

()안 표준오차를 나타낸다.

위 회귀분석 결과에서 계량은 기울기의 추정계수의 크기가 0.000045로 아주 작으므로 독립변수의 종속변수에 대한 영향도는 무시할 정도이라고 평가하였다. 이 주장이 타당한지 여부와 그 이유를 설명하라.

3. 다음은 1993년 2월-1994년 12월 기간 동안의 월별 서울시 오존농도(O₃, 단위:ppm)을 상수항과 기온(TEM, 단위:섭씨 도)에 대하여 회귀분석한 결과이다.

$$\begin{aligned}\widehat{O_3}_i &= 0.0074 + 0.00060 \text{ TEM}_i, & R^2 &= 0.362 \\ se &= (0.0006) \quad (A) \\ t\text{-stat} &= (2.733) \quad (3.447) \\ p\text{-value} &= (0.0125) \quad (0.0024)\end{aligned}$$

여기서 se 는 회귀계수의 표준오차를 나타내며, $t\text{-stat}$ 는 t -통계치 값을, $p\text{-value}$ 는 유의확률을 각각 나타낸다.

- (1) $b_2 = 0.00060$ 에 대해 해석하라.
- (2) $R^2 = 0.362$ 에 대해 해석하라.
- (3) 괄호 안의 (A)는 얼마인가?
- (4) 위 결과에서 2-t 실험기준으로 판단했을 때 기온(TEM)은 오존농도(O3)의 발생에 중요한 변수로 판단되는가?
- (5) 기온(TEM)은 오존농도(O3)의 발생에 영향을 미치지 못한다는 $H_0: \beta_2 = 0$ 의 가설을 유의확률을 통해 5% 유의수준에서 검정하라.
- (6) $H_0: \beta_2 \geq 0$ 의 가설을 대립가설에 대해 5% 유의수준에서 검정하라.
- (7) 위 결과에 추가로 회귀식의 표준오차 s 의 값을 0.0074로 구하였다. 이러한 경우 잔차의 제곱합 값을 구하라.
- (8) 94년 12월 경우 (평균)기온은 -2 도 였다. 94년 12월의 오존농도(O3) 값을 예측하라.

3. (실증분석 문제) 첨부된 파일 USED CAR에는 중고차의 판매가격(PRICE)과 차량의 특징이 수록되어 있다. 중고차의 판매가격을 상수항 및 주행거리(MILEAGE)에 대해 다음의 모형을 추정하라.

$$PRICE_i = \beta_1 + \beta_2 MILEAGE_i + \varepsilon_i$$

- (1) 중고차의 판매가격과 주행거리의 평균, 표준편차 등의 통계치를 구하라.
- (2) 두 변수 판매가격과 주행거리에 대한 상관관계수 값은 얼마이며 그 해석은 어떠한가?
- (3) 위 모형에서 β_1 의 추정치는 얼마이며, 그 해석은 어떠한가?

- (4) 회귀계수 β_2 의 추정치는 얼마이며, 그 해석은 어떠한가?
- (5) 표본회귀선이 (1)에서 구한 두 변수의 평균값을 지나는가? 이는 왜 그러한가?
- (6) 회귀계수 β_2 에 대한 95% 신뢰구간을 구하라.
- (7) 결정계수 R^2 값은 얼마이며, 그 해석은 어떠한가?
- (8) (2)에서 구한 상관계수 값의 제곱은 (7)에서 구한 결정계수 값과 일치하는가? 이 결과는 어떻게 해석할 수 있는가?
- (9) 위 결과에서 20,000Km를 주행한 중고차의 가격은 얼마로 예측되는가?

4. (실증분석 문제) EPL_PLAYER 파일은 2017-2018 시즌 잉글리쉬 프리미어 리그 (EPL) 축구 선수들의 연봉과 성적에 관한 정보를 수록하고 있다. 축구 선수들의 연봉(WAGE)이 리그 출장 경기수(APPEARANCE)에 의해 결정된다고 생각하여 다음 식과 같은 회귀분석 모형을 설정하였다. 물음에 답하시오.

$$WAGE_i = \beta_1 + \beta_2 APPEARANCE_i + \varepsilon_i$$

- (1) 연봉(WAGE)과 출장경기수(APPEARANCE)에 대한 통계치를 구하고 해석하라.
- (2) 연봉(WAGE)과 출장경기수(APPEARANCE)에 대한 상관계수 및 점상관도를 구하고 해석하라.
- (3) 연봉(WAGE)과 출장경기수(APPEARANCE)에 대해 회귀분석하고 결과를 해석하라. EPL 선수들은 1 경기 더 출장함에 따라 연봉은 얼마나 증가하는가?
- (4) (3)의 회귀분석에서 $H_0: \beta_2 \geq 0$ 의 가설을 5% 유의수준에 검정하라.
- (5) (3)의 회귀분석에서 결정계수 값은 얼마이며 이는 어떻게 해석되는가?
- (6) 만일 어떤 선수가 시즌 기간 동안 40경기에 출장한다면 기대하는 연봉은 얼마인가?
- (7) STAT85 변수는 구단 외부에서 개별 선수들에 대하여 평가한 것으로 85점이상이면 1의 값을, 85점 이하이면 0의 값을 갖는 변수이다. 연봉(WAGE)을 상수항과 STAT85 변수에 대해 회귀분석하라. STAT85 변수의 유의성은 어떠한가?
- (8) (7)의 회귀식에서 잔차 제곱합 값은 얼마인가?

