

Problem Set 9

계량경제학

남 준우

1. 다음 추정식 (A)에서 Y_i 는 1일 커피소비량, X_i 는 흡연개비수, D_i 는 남자이면 1을, 여자이면 0을 나타내는 가변수이다.

$$\hat{Y}_i = -0.8 + 2D_i + 0.5X_i + 0.2(D_i \times X_i) \quad (A)$$

그런데 W_i 를 여자이면 1을 나타내고, 남자이면 0을 나타내는 가변수라 하여 추정한 다음 식

$$\hat{Y}_i = b_1 + b_2W_i + b_3X_i + b_4(W_i \times X_i) \quad (B)$$

의 결과에서 b_1, b_2, b_3, b_4 의 값을 구하라.

2. Problem Set 8에서는 회사원 표본에 대해 봉급수준(Y)을 상수항과 교육수준(ED; 교육년수), 경력월수(EXP)와 현직장근무월수(TEN)에 대해 회귀분석한 문제를 다루었다. 이에 추가로 남·여 간의 성별 구분에 따라 봉급수준에 차이를 검정하기 위해 다음의 모형을 고려하였다. 여기서 DUM변수는 남성이면 DUM=1, 여성이면 DUM=0의 값을 나타낸다.

$$y_i = \beta_1 + \beta_2 DUM_i + \beta_3 ED_i + \beta_4 EXP_i + \beta_5 TEN_i + \beta_6 (DUM_i \times TEN_i) + \varepsilon_i$$

위 모형을 93명의 회사원 표본에 대해 구한 결과는 다음과 같다.

$$y_i = 3551.39 + 660.0DUM_i + 89.53ED_i + 1.29EXP_i + 22.21TEN_i + 3.92(DUM_i \times TEN_i) + e_i$$

(337.30) (215.66) (24.86) (0.59) (0.31) (1.32)

$$R^2 = 0.512$$

() 안은 표준오차를 나타냄.

- (1) 위 모형의 조정된 결정계수 값을 구하라.
- (2) 위 모형의 적합도에 대해 5% 유의수준에서 검정하라.
- (3) 남·여 간에 절편의 차이가 없다는 가설을 5% 유의수준에서 검정하라.
- (4) 남·여 간에 현직장근무월수에 따라 봉급수준에 미치는 영향에 차이가 없다는 가설을 5% 유의수준에서 검정하라.
- (5) 제5장 연습문제 3번의 결과와 비교하여 남·여 간에 봉급수준에 대한

모형의 차이가 없다는 가설을 5% 유의수준에서 검정하라.

- (6) 위 결과로부터 남자의 봉급수준 추정식과 여자의 봉급수준 추정식을 각각 구하라.
- (7) 교육수준이 12이고, 경력월수가 52개월이며, 현직장근무월수가 3개월인 계량(여자)의 봉급수준을 예측하라.

3. **(Empirical Question)** ARTWORK 파일은 1998년 12월-2001년 12월의 기간 동안 서울옥션에서 낙찰된 미술품 중 서양화의 가격에 대한 자료이다.

이 자료에서 PRICE는 그림 가격(만원 단위)을, SIZE변수는 그림의 크기를 나타내는 호수를 나타내고 있으며, DIE는 작가의 생존여부를 나타내는 가변수로 DIE=0이면 생존, 1이면 사망의 경우를 나타내며, AGE는 작가의 나이를 나타낸다. 또한 MEDIUM은 그림의 소재를 나타내는 변수로 아크릴화의 경우에는 1, 유화의 경우에는 2, 수채화의 경우에는 3, 기타의 경우에는 0을 나타내며 SUPPORT는 그림의 바탕 화면을 나타내는 변수로 캔버스이면 1, 종이이면 2, 기타이면 3을 값을 나타낸다.

- (1) 그림가격, 작가의 나이, 그림의 크기에 대한 통계치를 구하고 설명하라.
 - (2) 공분산, 상관계수 혹은 점상관도 등 필요한 통계치(혹은 그래프)를 통해 그림 가격과 작가의 나이, 그림 가격과 그림의 크기간의 관계에 대하여 유추하여 보아라.
 - (3) 작가의 생존에 따라 그림 가격에 차이가 있는지를 통계치를 구하여 설명하라.
 - (4) 소재에 따라 그림 가격에 차이가 있는지를 통계치를 구하여 설명하라.
 - (5) 그림의 바탕화면의 재료에 따라 그림 가격에 차이가 있는지를 통계치를 구하여 설명하라.
 - (6) $\log(\text{그림 가격})$ 을 상수항과 작가의 나이(AGE), 작가의 나이의 제곱, 그림의 크기(SIZE), 생존여부(DIE)에 대해 회귀분석하고 각 변수의 유의성에 대하여 설명하고 회귀계수의 의미에 대하여 설명하라.
 - (7) 그림의 소재에 따라 그림 가격의 차이가 있다는 전제하에 MEDIUM변수로부터 아크릴화의 경우 1의 값을 갖는 ACRYLIC변수, 유화의 경우 1의 값을 갖는 OIL 변수, 수채화의 경우 1의 값을 갖는 WATERCOLOR변수를 만들어라.
- 또한 SUPPORT변수로부터 캔버스의 경우 1의 값을 갖는 CANVAS, 종이의 경우 1의 값을 갖는 PAPER 변수를 만들어라.

이로 부터 (6)의 독립변수에 추가하여 바탕화면을 나타내는 가변수와 소재를

나타내는 가변수를 삽입하여 회귀분석하고 개별 바탕화면의 가변수와 소재의 가변수에 대해 5% 유의수준에서 가설검정하라.

<gretl Tips: gretl에서 더미변수를 만드는 방법>

Add new variables에서

Acrylic=(medium=1)

을 통하여 기존 변수로부터 더미변수를 만들 수 있다.

혹은 Dummies for discrete variable 을 이용하겠.

gretl에서 help란의 manual을 참고하거나 google에서 질문을 해 보세요.

(8) (7)번 문제에서 바탕화면을 나타내는 가변수 전체의 유의성에 대하여 5% 유의수준에서 검정하라.

(9) (7)번 문제에서 소재를 나타내는 가변수 전체의 유의성에 대하여 5% 유의수준에서 검정하라.

4. **(Empirical Question)** MOVIE 파일은 2012년에 개봉된 한국영화 관객 수에 관한 정보를 담고 있다. 영화의 장르는 관객 동원에 관한 중요한 변수이다. GENRE 변수는 영화의 메인 장르에 관한 정보를 담고 있다. 이로부터 각 장르별 가변수를 생성하라. 또한 RATE 변수는 영화상영 등급에 관한 정보를 나타내고 있다. 이 변수로부터 등급을 나타내는 가변수를 생성하고 다음 질문에 답하시오.

(1) 관객 수(AUDIENCE)와 스크린 수(SCREEN) 간의 점상관도(scatter diagram)를 그려보라. 두 변수간의 관계는 어떠한가?

(2) 관객 수(AUDEINCE)를 상수항과 스크린 수(SCREEN)와 스크린 수 제곱 변수에 대해 회귀분석하라. 스크린 수의 관객 수에 대한 한계효과는 스크린 수의 증가에 대해 어떻게 변화하는가?

(3) 스크린 수(SCREEN)와 스크린 수 제곱 변수에 추가로 장르에 대한 가변수를 포함하여 회귀분석하라.

(4) (3)의 회귀분석에 있어서 스크린 수(SCREEN)와 스크린 수 제공 변수의 관객 수에 대한 영향은 영화의 유형, 즉 상업영화와 다양성 영화에 대해 다르다고 한다. 이러한 모형을 추정하라.

관객 수에 대한 스크린 수의 영향도는 영화의 유형에 따라 다르게 나타나는가?

(5) 관객 수는 학생들의 방학 여부에 따라 다르다고 한다. MONTH 변수로부터 겨울방학(1,2월), 여름방학(7,8월) 및 기타 시기를 나타내는 변수를 생성하고 (2)의 모형에 겨울방학, 여름방학, 기타 시기를 나타내는 가변수를 포함하여 회귀분석하고 해석하라.

(6) (2)와 (5)의 회귀분석 결과에서 관객 수는 상영 시기가 겨울방학, 여름방학, 기타시기 구분과 무관하다는 가설을 5% 유의수준에서 검정하라.

5. **(Empirical Question)** 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 기본 회귀분석 모형을 설정하였다.

$$\log(WAGE_i) = \beta_1 + \beta_2 APPEARANCE_i + \varepsilon_i$$

위 모형에 추가로 축구선수들의 포지션(POSITION)과 득점수(GOALS) 또한 연봉 결정에 중요한 요인으로 작용한다. 다음 물음에 답하시오.

(1) 연봉(WAGE)과 출장경기수(APPEARANCE)에 대한 통계치, 점상관도를 구하고 해석하라.

(2) 선수들의 포지션 별로 임금의 통계치를 구하라. 포지션 별로 임금 수준은 어떠한가?

(3) POSITION 변수로부터 각 포지션 별 가변수를 구하라. 선수들의 로그로 취환한 연봉(WAGE)을 상수항과 경기출장수(APPEARANCE), 득점 골 수(GOALS), 종합 평가점수(STAT85), 구단별 가치(TEAMVALUE) 변수와 포지션별 가변수에 대해 회귀분석하라.

(4) (3)의 회귀분석에서 개별 변수의 유의성에 대해 5% 유의수준에서 검정하라.

(5) (3)의 회귀분석에서 모형의 적합도에 대하여 5% 유의수준에서 검정하라.

(6) (3)의 회귀분석에서 득점 골 수(GOALS) 회귀계수의 추정치는 얼마이며 이에 대한 해석은 어떠한가>

(7) (3)의 회귀분석에서 종합 평가점수와 구단별 가치는 선수들의 기록에 직접적으로 영향을 미치지 못하는 요인으로 작용한다. 이 두 변수의 영향도가 0이라는 결합가설을 5% 유의수준에서 잔차제곱합을 통하여 검정하라.

(8) (7)의 가설을 결정계수 값을 통하여 검정하라.