11주차 강의자료 및 과제

1. 다중회귀모형의 추가적 논의 : 모수에 대한 유의성 검정(교과서 204-211쪽)
모수 하나에 대한 가설검정시 t검정통계량으로, 모수 2개 이상에 대한 가설검정시 F검정통계량으로 가설 검정을 실시한다.

- 식(11-1)과 같이 모수 하나에 대해 가설검정을 할 경우 t검정통계량으로 귀무가설(단일귀무가설) 기각여부를 판별한다. 만일 t값이 임계치(t_c)보다 작을 경우 유의수준 α%하에서 귀무가설은 수락되고, 햄버거 가격이 매출액에 영향을 주지 않는다고 추론할 수 있다. 반면에 임계치보다 클경우 우리는 유의수준 α%하에서 귀무가설을 기각하고, 햄버거 가격이 매출액에 영향을 준다고 주장할 수 있다.

(11-1) $S_i = b_1 + b_2 P_i + \widehat{e_{i,R}}$

· 가설 설정 : $H_0 : \beta_2 = 0, \quad H_1 : \beta_2 \neq 0,$

· 검정통계량 : $t = \dfrac{b_2}{se(b_2)}$

· 추　　　론 : $t-value$, $t_{(\alpha/2, n-k)}$, α, $p-value$에 의해 귀무가설 기각여부 판별

- 다음은 식(11-2)와 같이 모수 두 개 이상에 대해 가설검정을 경우 F검정통계량으로 귀무가설(결합귀무가설) 기각여부를 판별한다.

(11-2) $S_i = b_1 + b_2 P_i + b_3 A_i + b_4 A_i^2 + \widehat{e_{i,U}}$

· 가설 설정 : $H_0 : \beta_3 = 0, \ \beta_4 = 0, \qquad H_1 : \beta_3 \ \beta_4$ 중 적어도 하나는 0이 아니다.,

· 검정통계량

(11-3) $F = \dfrac{(SSE_R - SSE_U)/J}{SSE_U \ / \ (n-k)}$

· 추　　　론 : $F-value$, $F_{(\alpha, J, n-k)}$,α, $p-value$에 의해 귀무가설 기각여부 판별

- 우리는 편의상 추정할 계수가 상대적으로 적은 식(11-1)을 '제한된 모형(restricted model)'이라고 하자. 그리고 추정할 계수가 상대적으로 많은 식(11-2)를 '제한되지 않은 모형(unrestricted model)'이라고 하자.
- 식(11-3)에서 SSE_R은 '제한된 모형'에서 추정한 잔차($\widehat{e_{i,R}}$)의 총변동'($\Sigma \widehat{e_{i,R}}^2$)이며, SSE_U은 '제한되지 않은 모형'에서 산출된 잔차($\widehat{e_{i,U}}$)의 총변동'($\Sigma \widehat{e_{i,U}}^2$)이다.

· 그리고 '제한되지 않은 모형'에서는 추정해야 할 계수가 4개이다. 따라서 자유도 $n-k$는

자유도($n-4$)가 된다. 또한 J는 '제한되지 않은 모형'의 계수 수(4)에서 '제한된 모형'에서 추정할 계수 수(2)를 뺀 것으로 여기서는 2가 된다.

· F검정통계량은 단측검정이다. 따라서 임계치(F_c)는 J 및 $n-k$ 자유도를 갖는 F분포의 오른쪽 끝부분에 있는 확률 α와 관련된다. 여기서 표본 관측치 수가 n개이고, 유의수준이 5%일 경우 임계치는 다음과 같다. $F_c = F_{(0.95, 2, n-4)}$

- 식(11-1)과 식(11-2)의 결합검정은 어떤 의미가 있을까? 이 검정은 β_3 ,β_4 둘 중 하나가 0인가 아닌가에 대한 검정이다. 따라서 F통계값이 임계치(F_c)보다 크다면 유의수준 α%하에서 귀무가설은 기각된다. 이 경우 β_3 β_4 중 최소 하나는 적어도 0이 아니다, 혹은 '광고효과는 있다'라고 주장할 수 있다(교과서 208-209쪽 참조)

2. 모의변수 또는 더미변수 (dummy variable)
(교과서 245-266쪽 참고)

- 그 동안 표본회귀모형에 투입된 종속변수와 독립변수 모두가 양적으로 관측가능한 변수(quantitative variable)였다. 그런데 독립변수가 질적변수(qualitative variable)도 있다. 이 경우 우리는 헤도닉 가격 모형에서 쉽게 발견할 수 있다..

- 가장 일반적인 예는 주택 가격($PRICE$)이다. 주택가격은 주택 크기, 위치, 방수, 건축연도, 등 처럼 주택 개별 특성에 의해 결정된다.
· 우선 주택가격($PRICE$)이 주택 크기($SQFT$)에 의해서 결정된다고 하자. 이 경우 식(11-4)와 같이 주택 가격은 주택 크기에 비례한다.

(11-4) $PRICE = b_1 + b_2 SQFT + \hat{e_i}$

하지만 주택 위치에 따라 동일 크기라도 다르다. 도시안에 있는 경우 주택가격은 높다. 반면에 도시외곽에 있는 경우 주택 크기가 동일할지라도 주택가격은 낮다. 따라서 위치 특성을 구분하는 모의변수를 새롭게 만들 필요가 있다. 주택가격 결정모형에 위치 특성을 설명하는 모의변수를 추가하면 통상적으로 표본회귀모형의 설명력은 더 높아진다.

. (11-5) $PRICE = b_1 + b_2 SQFT + \delta UTOWN + \hat{e_i}$

· 식 (11-5)에서 $UTOWN$는 위치 특성을 나타내는 모의변수이다. 편의상 도시안에 있는 경우 $UTOWN$=1, 도시외곽에 있는 경우 $UTOWN$=0으로 정하자[1] 참고적으로 모의변수 자료에 대해서는 교과서 251쪽에 나와 있는 표<7-2>를 참고하기 바란다. .

- 가령 주택가격과 크기, 그리고 위치를 조사하여 데이터 파일를 구축한 후 식 (11-5)의 계수를 최소자승법에 의해 추정한 결과를 식(11-6)과 같다고 하자. 이때 개별 변수의 측정

1) 이와 같이 주택 위치 특성을 구분하는 변수를 모의변수(Dummy variable)라 한다. 모의변수는 통상적으로 0 또는 1의 값을 갖게 된다. 모의변수는 이원변수 또는 이분변수라고도 한다.

단위는 주택가격의 경우 백만원, 크기의 경우 1평이다.

· 식(11-6)에 의하면 주택 크기가 1평씩 늘어나면 7.6백만원씩 주택가격이 늘어난다, 또 도시안에 위치하고 있으면 도시밖에 있는 경우보다 주택 가격이 27백만원 더 높다.

(11-6) $\widehat{PRICE} = 24.5 + 7.6SQFT + 27UTOWN \quad R^2 = 0.5$

- 식(11-5)~(11-6)을 종합하면 주택 위치에 따라 절편이 다르므로 우리는 이를 절편 모의변수라고 한다. 절편모의변수를 추가함으로써 동일한 주택 크기라 할지라도 주택 위치가 도시안에 비해 도시밖에 있는 경우 주택 가격이 27백만원 더 저렴한 것을 알 수 있다.

· 도시안과 밖으로 주택가격결정모형 추정결과를 구분하면 다음과 같다.

$UTOWN$=1인 경우 $\widehat{PRICE} = 51.5(= 24.5 + 27) + 7.6SQFT$
$UTOWN$=0인 경우 $\widehat{PRICE} = 24.5 + 7.6SQFT$

- 다음은 기울기 모의변수에 대해 알아 보자. 식(11-5)~(11-6)서 다음과 같은 의문을 제기할 수 있다.

· 도시안에 있는 주택의 경우 평당 가격이 도시밖에 있는 경우보다 더 크지 않을까? 예를 들어 밀양에 비해서 서울 강남 아파트의 경우 평당 가격이 더 큰 게 사실이다. 그런데 앞의 추정결과에서는 평당 가격이 똑 같은 게 문제이다.

- 도시안과 도시밖의 평당 가격이 차이가 나는 것을 모형에 반영하기 위해서는 식(11-7)과 같이 기울기 모의변수(상호작용변수)를 추가로 투입해야 한다.

(11-7) $PRICE = b_1 + b_2SQFT + \delta UTOWN + \gamma(SQFT \times UTOWN) + \hat{e_i}$

· 기울기 모의변수는 주택 크기와 위치를 나타내는 모의변수의 곱이다[2). 여기서 위치와 크기가 주택가격에 미치는 상호작용 효과를 나타낸다는 의미에서 상호작용변수라고도 한다.

· 식 (11-7)의 계수를 최소자승법에 의해 추정한 결과는 식(11-8)과 같다고 하자[3). 또한 도시안과 밖으로 주택가격결정모형 추정결과를 구분하면 다음과 같다.

(11-8) $\widehat{PRICE} = 24.2 + 7.2SQFT + 25UTOWN + 1.5(SQFT \times UTOWN) \quad R^2 = 0.65$

---

2) 다중회귀모형에 상호작용변수를 투입할 경우 반드시 새로운 변수로 치환해서 투입해야 한다. 여기서는 편의상 치환된 변수로 설명하기보다는 이해를 돕기 위해서 두 변수의 곱의 형태로 표시해서 설명하기로 하자.
3) 다중회귀모형에서 독립변수가 늘어날 경우 계수, 계수의 t값, 결정계수 값 등이 바뀐다. 또한 추정식에 나타나는 계수는 모두 통계적으로 유의하다는 전제하에서 논의를 전개하고 있다.

$UTOWN$=1인경우 $\widehat{PRICE}=(24.2+25)+(7.2+1.5)SQFT=49.2+8.7SQFT$
$UTOWN$=0인 경우 $\widehat{PRICE}=24.2+7.2SQFT$

- 식(11-7)~(11-8)을 종합하면 절편모의변수를 투입함으로써 동일한 주택 크기라 할지라도 주택 위치가 도시안에 비해 도시밖에 있는 경우 주택 가격이 25백만원 더 저렴하다는 점을 밝혀냈다. 또한 기울기모의변수를 추가함으로써 도시안에 있는 경우 평당 가격은 8.7백만원, 도시 밖에 있는 경우 평당 가격은 7.2백만원임을 추가로 밝혀냈다. .이에 해당하는 그림은 교과서 250쪽 그림 7.4를 참고하기 바란다.

11주차 과제

문제 1. 결합검정을 통해 광고효과가 있는지 검정하시오. 단, 아래식에서 모든 계수는 5%에서 통계적으로 유의한 것으로 이미 검정되었다. 이때 S_i는 햄버거체인점 판매액을 A_i는 광고비지출액이며, 이 두 변수 측정단위는 백만원이다. 그리고 P_i는 햄버거 판매가격(측정단위 천원), i는 체인점을 구분하는 하첨자이며, 여기서 표본 수는 24개이다.

식(1) $\hat{S_i}=100-7.0P_i \quad R^2$=0.4

식(2) $\hat{S_i}=90-0.8P_i+10.0A_i-2.0A_i^2 \quad R^2=0.8$

1.1 어느 것이 제한된 모형이며, 제한되지 않은 모형인가?

1.2. 매출액 평균에 대한 S_i의 총변동'(Total Sum of Squares; SST)은 100이다. SSE_R , SSE_U을 구하시오.

1.3. 유의수준 5%에서 임계치를 구하고, 결합검정을 통해 광고효과가 있는지 검정하시오.

1.4. 식(2)에서 무작정 광고비 지출액을 늘리면 매출액이 오르는가? 오른다면 또는 오르지 않는다면 그 이유를 설명하시오.

1.5, 광고지출이 증가함에 도 불구하고 매출액이 증가하지 않을 경우 이에 해당하는 광고비 지출 수준은?

1.6 광고 한 단위 증가함에 따른 한계편익과 한계비용이 같아질 때 광고의 최적 수준은?

문제 2. 무는 남부지방에서 봄철과 가을철에 출하되는 봄 무와 가을 무, 강원도에서 여름에 출하되는 고냉지 무, 그리고 겨울에 제주도에서 출하되는 월동 무가 있다. 따라서 무 도매시장 가격($PRICE_m$)은 출하물량( )에 영향을 받지만, 봄, 여름, 가울, 겨울 출하 시기에 따라서도 무 가격이 다르다. 이때 m은 월을 구분하는 하첨자이다.

2.1 무 출하량에 따라 가격이 하락한다고 가정하자. 이 경우 출하량에 따른 가격의 영향(신축성 계수=물량 변동률 대한 가격 변화율)을 계측하기 위해서 어떻게 계량경제모형을 구축할까? 표본회귀식으로 나타내시오.

2.2. 모의변수(dummy variable)에 대한 질문이다. 모의변수 함정을 설명하시오.

2.3. 무 출하량이외에도 계절에 따라 무 가격이 영향을 받는다고 할경우 어떻게 계량경제모형을 구축할까? 표본회귀식으로 나타내시오.