9주차 3차시 : 가변수모형(복수 가변수 및 응용)

1. 복수 가변수
2. 가변수의 응용(계절성 분석)

1.복수 가변수

(1)한 개 정성변수에 여러 범주의 경우
-한 개 정성변수에 범주가 n개(n≥3)인 경우 n-1개의 가변수를 생성

(예)학력을 중졸이하, 고졸, 대졸이상으로 구분할 경우 High, College 가변수를 생성하면 다음과 같이 3개의 범주로 구분이 됨
-중졸이하 : High=0, College=0
-고졸 : High=1, College=0
-대졸이상 : High=0, College=1

(2)여러 정성변수의 경우
-정성변수가 2개 이상인 경우 각각의 정성변수에 대해 가변수를 생성

(예)성별 및 학력별에 따른 임금구조를 알고자 할 경우 성별 가변수 Gender(남자=1, 여자=0)와 학력 가변수 High, College를 생성
-여자 : Gender=0
-남자 : Gender=1

-중졸이하 : High=0, College=0
-고졸 : High=1, College=0
-대졸이상 : High=0, College=1

(예)근무연수(나이, X) 외에 성별(Gender) 및 학력별(High, College) 임금구조를 분석하고자 할 경우 다음의 모형을 설정할 수 있음

$$Y_i = \alpha + \beta X_i + \gamma Gender_i + \delta High_i + \pi College_i + u_i$$

(해석)위 회귀식에서 다음과 같이 임금구조를 도출할 수 있음

중졸이하 여자의 경우 : $Y_i = \alpha + \beta X_i + u_i$

중졸이하 남자의 경우 : $Y_i = (\alpha + \gamma) + \beta X_i + u_i$

고졸 여자의 경우 : $Y_i = (\alpha + \delta) + \beta X_i + u_i$

고졸 남자의 경우 : $Y_i = (\alpha + \gamma + \delta) + \beta X_i + u_i$

대졸이상 여자의 경우 : $Y_i = (\alpha + \pi) + \beta X_i + u_i$

대졸이상 남자의 경우 : $Y_i = (\alpha + \gamma + \pi) + \beta X_i + u_i$

구분	여자	남자
중졸 이하	−1.363(reference)	(−1.363+0.658)=−0.705
고졸	(−1,363+0.389)	(−0.705+0.389)
대졸 이상	(−1,363+0.982)	(−0.705+0.982)

```
Regression Results of using Dummy Variable
================================================
                      Dependent variable:
                  ------------------------------
                            income
------------------------------------------------
age                        0.052***
                           (0.014)

gender                     0.658***
                           (0.209)

high                        0.389
                           (0.239)

college                    0.982***
                           (0.241)

Constant                  -1.363**
                           (0.596)

------------------------------------------------
Observations                  85
R2                          0.409
Adjusted R2                 0.379
Residual Std. Error    0.646 (df = 80)
F Statistic         13.820*** (df = 4; 80)
================================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

2.가변수의 응용(계절성 분석)

-일반적으로 경제시계열은 다음과 같은 네 가지 요인으로 구성되어 있음

$Y_t = T_t \times C_t \times S_t \times I_t$ 또는 $lnY_t = lnT_t + lnC_t + lnS_t + lnI_t$

- 추세요인(secular trend : T_t) : 기술향상 및 인구증가 등의 요인에 의해 체계적으로 변하는 부분
- 순환요인(cyclical factor : C_t) : 추세를 중심으로 순환변동의 현상을 보이는 부분
- 계절요인(seasonal factor : S_t) : 기후 및 시회적 관습 등에 의해 변하는 부분
- 불규칙요인(irregular factor : I_t) : 기상재해 및 파업 등에 의해 변하는 부분

(참고)가법모형과 승법모형
-가법모형: 계절성분의 진폭이 수준에 관계없이 일정할 때
-승법모형: 계절성분의 진폭이 시계열 수준에 따라 달라질 때
-실제로는 승법모형 또는 로그 가법모형이 많이 사용됨

–경제시계열에 계절적 요인이 있을 경우 계절적 요인을 설명할 수 있는 계절적 변수가 필요한데 이러한 계절성을 설명하는 변수로 가변수를 이용

–계절적 요인이 있는 분기별 자료의 경우 가변수를 이용하여 계절변동을 고려한 회귀식(모형의 절편이 분기별로 다르고 기울기는 같은 모형이라고 가정할 경우)은 다음과 같음

$$Y_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \beta X_t + u_t$$

단, $D_{2t} = \begin{Bmatrix} 1, t\text{가 } 2\text{분기이면} \\ 0, t\text{가 다른 분기 이면} \end{Bmatrix}$

$D_{3t} = \begin{Bmatrix} 1, t\text{가 } 3\text{분기이면} \\ 0, t\text{가 다른 분기 이면} \end{Bmatrix}$

$D_{4t} = \begin{Bmatrix} 1, t\text{가 } 4\text{분기이면} \\ 0, t\text{가 다른 분기 이면} \end{Bmatrix}$

(해석)
-각 분기의 해석은 다음과 같음

1분기 : $Y_t = \alpha_1 + \beta X_t + u_t$
2분기 : $Y_t = (\alpha_1 + \alpha_2) + \beta X_t + u_t$
3분기 : $Y_t = (\alpha_1 + \alpha_3) + \beta X_t + u_t$
4분기 : $Y_t = (\alpha_1 + \alpha_4) + \beta X_t + u_t$

-가변수가 모두 2,3,4,분기의 절편은 1분기에 비해 각각 $\alpha_2, \alpha_2, \alpha_4$만큼 차이가 나서 독립변수 X와 무관하게 Y의 값에서 발생하는 계절성을 설명할 수 있음
-각 분기별로 Y와 X의 관계를 별도로 추정할 수도 있지만 가변수를 포함한 식을 추정하면 이와 동일한 결과를 얻을 수 있음