6주차 1차시 : 다중회귀분석(표준화 회귀모형)

1.회귀계수 추정 비교(단순회귀 vs. 다중회귀)

2.표준화된 회귀모형

1. 회귀계수 추정 비교(단순회귀 vs. 다중회귀)

단순회귀 : $\hat{\beta} = \dfrac{Cov(X,Y)}{Var(X)}$

다중회귀 : $\hat{\beta}_k = \dfrac{Cov(\tilde{X}_k,Y)}{Var(\tilde{X}_k)}$, k=2,⋯k

```
> x2<-c(1,2,3,2)
> x3<-c(2,1,1,2)
> y0<-c(1,1,2,3)
> n<-length(y0)
> olsx2<-lm(x2~x3)
> resx2<-resid(olsx2)
> ols2<-lm(y0~resx2+x3)
> summary(ols2)

Call:
lm(formula = y0 ~ resx2 + x3)

Residuals:
    1     2     3     4 
-0.25  0.25 -0.25  0.25 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0000     0.7906   1.265    0.426
resx2         1.5000     0.5000   3.000    0.205
x3            0.5000     0.5000   1.000    0.500

Residual standard error: 0.5 on 1 degrees of freedom
Multiple R-squared:  0.9091,	Adjusted R-squared:  0.7273 
F-statistic:     5 on 2 and 1 DF,  p-value: 0.3015
```

(예) $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

$$\hat{\beta}_2 = \frac{Cov(\tilde{X}_2,Y)}{Var(\tilde{X}_2)}$$

단, $X_{2i} = \alpha_1 + \alpha_2 X_{3i} + \tilde{X}_{2i}$

$$\hat{\beta}_3 = \frac{Cov(\tilde{X}_3,Y)}{Var(\tilde{X}_3)}$$

단, $X_{3i} = \alpha_1 + \alpha_2 X_{2i} + \tilde{X}_{3i}$

```
> olsx3<-lm(x3~x2)
> resx3<-resid(olsx3)
> ols3<-lm(y0~x2+resx3)
> summary(ols3)

Call:
lm(formula = y0 ~ x2 + resx3)

Residuals:
    1     2     3     4 
-0.25  0.25 -0.25  0.25 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.7500     0.7500   1.000    0.500
x2            0.5000     0.3536   1.414    0.392
resx3         2.0000     0.7071   2.828    0.216

Residual standard error: 0.5 on 1 degrees of freedom
Multiple R-squared:  0.9091,	Adjusted R-squared:  0.7273 
F-statistic:     5 on 2 and 1 DF,  p-value: 0.3015
```

2. 표준화된 회귀모형

-회귀모형에서 독립변수의 상대적인 중요도를 알고자 할 경우 독립변수 및 종속변수를 표준화한 다음 표준화된 변수로 설정한 다음 식과 같은 회귀모형(이를 표준화된 회귀모형이라고 함)을 추정해야 함

$$\frac{Y_i - \bar{Y}}{s_Y} = \beta_2^* \frac{X_{2i} - \bar{X}_2}{s_{X_2}} + \cdots + \beta_k^* \frac{X_{ki} - \bar{X}_k}{s_{X_k}} + u_i$$

-단, 위 모형은 상수항을 포함하지 않음

-이때 추정된 회귀계수를 베타회귀계수(Beta coefficients) 또는 표준화된 회귀계수(standardized regression coefficients) 라고 함

-베타회귀계수 $\widehat{\beta^*}$ 와 회귀계수 $\hat{\beta}$은 다음과 같은 관계가 있음

$$\hat{\beta}_j^* = \hat{\beta}_j \frac{s_{X_j}}{s_Y}, (j = 2,3, \ldots k)$$

-추정된 베타회귀계수의 값이 예를 들어 0.7이면 독립변수가 표준편차 하나 크기(one standard deviation change in the independent variable)만큼 변할 때 종속변수의 표준편차는 0.7크기만큼 변한다고 해석함

```
> y0bar<-mean(y0)
> x2bar<-mean(x2)
> x3bar<-mean(x3)
> y0sd<-sqrt(var(y0))
> x2sd<-sqrt(var(x2))
> x3sd<-sqrt(var(x3))
> sy0<-(y0-y0bar)/y0sd
> sx2<-(x2-x2bar)/x2sd
> sx3<-(x3-x3bar)/x3sd
> sols<-lm(sy0~sx2+sx3-1)
> summary(sols)

Call:
lm(formula = sy0 ~ sx2 + sx3 - 1)

Residuals:
      1       2       3       4 
-0.2611  0.2611 -0.2611  0.2611 

Coefficients:
    Estimate Std. Error t value Pr(>|t|)  
sx2   1.2792     0.3015   4.243   0.0513 .
sx3   1.2060     0.3015   4.000   0.0572 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3693 on 2 degrees of freedom
Multiple R-squared:  0.9091,    Adjusted R-squared:  0.8182 
F-statistic:    10 on 2 and 2 DF,  p-value: 0.09091

> y0sd
[1] 0.9574271
> x2sd
[1] 0.8164966
> x3sd
[1] 0.5773503
> beta2hat<-summary(sols)$coef[1,1]*(y0sd/x2sd)
> beta3hat<-summary(sols)$coef[2,1]*(y0sd/x3sd)
> beta2hat
[1] 1.5
> beta3hat
[1] 2
```