

7주차 2차시 : 다중회귀분석(가설검정 2)

1. 가설검정
 - (2) 결합검정
 - (3) 전체검정

(2) 결합 검정

(귀무가설) $H_0 : RB = r$

$$\text{단, } R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 \end{bmatrix} \quad r = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{즉, } H_0 : \beta_3 = \beta_4 = \beta_5 = 0 \quad (q=3)$$

(제약이 가해지지 않은 모형)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} \cdots + \beta_k X_{ki} + u_i \quad (i=1, 2, 3, \dots, n)$$

(제약(귀무가설이 사실임)이 가해진 모형)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_6 X_{6i} + \cdots + \beta_k X_{ki} + u_i \quad (i=1, 2, 3, \dots, n)$$

(검정통계량)
$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/n - k} \sim F_{n-k}^q$$

단, R_{ur}^2 은 제약이 가해지지 않은 모형의 결정계수
 R_r^2 은 제약이 가해진 모형의 결정계수

또는
$$F = \frac{(e'e_r - e'e_{ur})/q}{(e'e_{ur})/n - k} \sim F_{n-k}^q$$

단, $e'e_r$ 은 제약이 가해진 모형의 잔차 제곱의 합
 $e'e_{ur}$ 은 제약이 가해지지 않은 모형의 잔차 제곱의 합

(3) 전체 검정

(귀무가설) $H_0 : RB = r$

$$\text{단, } R = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad r = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

즉, $H_0 : \beta_2 = \beta_3 = \beta_4 = \cdots = \beta_k = 0$ ($q=k-1$)

(검정통계량)
$$F = \frac{R^2/k - 1}{(1 - R^2)/n - k} \sim F_{n-k}^{k-1}$$

(예제:전체검정)

-홍보비 지출액 및 연구개발 지출액이 동시에 연간매출액에 영향을 주지 않는다

$$H_0 : \beta_2 = \beta_3 = 0$$

$H_1 : H_0$ 가 사실이 아니다

-검정통계량 : $F = \frac{R^2/k-1}{(1-R^2)/n-k} = \frac{0.909/2}{(1-0.909)/1} = 5.0 \sim F_1^2=199.5$

-결론: 5% 유의수준에서 분자자유도 2, 분모자유도 1인 F-분포의 임계값은 $F_1^2 = 199.5$

5% 유의수준 하에서 귀무가설을 기각하지 못함
 모형설정이 완전히 잘못된 것으로 해석
 새로운 변수로 독립변수 모형을 설정

(예제) 다음의 홍보비 지출액 X_2 (단위:천만 원), 연구개발 지출액 X_3 (단위:천만 원)과 연간 매출액 Y (단위:억 원)에 관한 자료를 이용하여 전체검정을 하라.

Y	1	1	2	3
X_2	1	2	3	2
X_3	2	1	1	2

```
> ur.lm<-lm(y~x2+x3)
> summary(ur.lm)

call:
lm(formula = y ~ x2 + x3)

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

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.2500     1.9203  -2.213   0.270
x2             1.5000     0.5000   3.000   0.205
x3             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
```

(예제:결합검정) 앞의 예제에서 매출액 결정요인으로 달러환율 X_4 (단위: 원)을 추가한 다음의 모형을 추정하고 연구개발 지출액과 달러환율이 매출액에 영향을 주지 않는다는 귀무가설을 검정해 보자.

Y	1	1	2	3	3
X_2	1	2	3	2	4
X_3	2	1	1	2	3
X_4	2	3	2	4	2

$$H_0 : \beta_3 = \beta_4 = 0$$

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n - k)} = \frac{(0.8356 - 0.4327)/2}{(1 - 0.8356)/1} = 1.225 \sim F_1^2$$

5% 유의수준 하에서 F-분포의 임계값은 199.5이므로 귀무가설을 기각하지 못함

```
> ur.lm<-lm(y~x2+x3+x4)
> r2.ur<-summary(ur.lm)$r.squared
> summary(ur.lm)

Call:
lm(formula = y ~ x2 + x3 + x4)

Residuals:
    1         2         3         4         5
0.05978 -0.59783  0.41848  0.29891 -0.17935

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.8424    1.8744  -0.983   0.505
x2             0.5761    0.3966   1.453   0.384
x3             0.5109    0.5212   0.980   0.506
x4             0.5924    0.4745   1.248   0.430

Residual standard error: 0.8109 on 1 degrees of freedom
Multiple R-squared:  0.8356,    Adjusted R-squared:  0.3424
F-statistic: 1.694 on 3 and 1 DF,  p-value: 0.5017

> r.lm<-lm(y~x2)
> r2.r<-summary(r.lm)$r.squared
> summary(r.lm)

Call:
lm(formula = y ~ x2)

Residuals:
    1         2         3         4         5
-0.19231 -0.76923 -0.34615  1.23077  0.07692

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.6154    0.9946   0.619   0.580
x2            0.5769    0.3814   1.513   0.228

Residual standard error: 0.8697 on 3 degrees of freedom
Multiple R-squared:  0.4327,    Adjusted R-squared:  0.2436
F-statistic: 2.288 on 1 and 3 DF,  p-value: 0.2276

> fc<-((r2.ur-r2.r)/2)/((1-r2.ur)/(n-k))
> fc
[1] 1.225366
> pval<-1-pf(fc, 2, n-k)
> pval
[1] 0.5383249
```