

1. True model $y_i = 1 + x_{1i} + x_{2i} + u_i, i = 1, \dots, n, \quad n = 100,$

$x_i = (x_{1i}, x_{2i}) \text{ iid } N_2(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}), \quad u_i \text{ iid } N(0, 1), \quad \rho = 0.5$

추정 모형 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i}$ 으로부터 $\hat{\beta}_1$ 을 계산하는 함수를 아래와 같이 만들자. 즉 n, ρ 값을 주면 $\hat{\beta}_1 = \text{beta1.hat.x2.omitted}(n, \rho)$ 이다.

예를들면 $\text{beta1.hat.x2.omitted}(100, 0.5)$ 를 수행하면 1.4969 가 나온다.

```
library(MASS) #패키지불러오기에서 MASS 불러온 다음실행시킵
beta1.hat.x2.omitted <- function(n, rho) {
  set.seed(1); Sigma = matrix(c(1, rho, rho, 1), 2, 2); Mu = rep(0, 2); X = mvrnorm(n, Mu, Sigma); x1 = X[, 1]; x2 = X[, 2]
  u = rnorm(n); y = 1 + x1 + x2 + u; ols_misspecified = lm(y ~ x1)
  return(ols_misspecified$coef[2]) }
```

| n | 100 | 1000 | 10000 | 100000 | 1000000 |
|-----------------|----------|----------|----------|----------|----------|
| $\hat{\beta}_1$ | 1.496908 | 1.530526 | 1.517041 | 1.503238 | 1.500269 |

위표를 채우고 그의미를 설명하시오.

데이터 크기 n 이 무한히 커지면 beta1.hat 은 참값 beta1 에 가까워지면 beta1.hat 은 일치성을 갖는다고한다. 하지만 이 경우는 n 이 커지고 있는데 beta1.hat 은 참값 beta1 에 가까워지지 않으므로 일치성을 갖지 않는다고 볼 수 있다.

2. 교안 Chapter 4, 17쪽 까지 강의 동영상 본후에 과제를 푸세요.

교안 Chapter 5, 83쪽 담배의 수요탄력성 추정 결과

$\text{ivreg}(\log.q \sim \log.p + \log.inc | \text{sale.tax} + \log.inc) \quad \# -0.74$

를 개선하고자 한다. 이 추정모형에서는 각 주의 효과가 통제되지 않았음을 주목하자. 주의 효과를 통제하기위해 교안 Chapter 4, 17쪽 (식(8.2) 아래부분) 전후비교 방법을 적용하여 수요탄력성을 추정하고 (모든변수의 time difference 를 고려해야함) 95% 신뢰구간을 구하자.

> lm.85

Call:

$\text{lm}(\text{formula} = \text{income}[\text{year} == 1985] \sim \text{tax}[\text{year} == 1985])$

Coefficients:

(Intercept) $\text{tax}[\text{year} == 1985]$
49196423 722575

> lm.95

Call:

$\text{lm}(\text{formula} = \text{income}[\text{year} == 1995] \sim \text{tax}[\text{year} == 1995])$

Coefficients:

| | |
|-------------|-------------------|
| (Intercept) | tax[year == 1995] |
| -20189827 | 2753889 |

이 추정결과가 위 결과 -0.74 와 다른이유를 설명하시오.

담배의 수요탄력성에 유관한 다른 중요변수들이 누락되었기 때문이다.

전후 비교를 통하여 1995년과 1985년의 차이를 통해서 누락변수의 영향을 제거할 수 있다.

어떤 변수를 넣느냐에 따라 추정결과가 달라지기 때문이다.

포함되어있던 변수를 제외시키면서 좋은 추정치를 얻을 수도 있다.