

# Ekonomietrian tilastolliset menetelmät

## Laskuharjoitus 3

2014, Tammikuu 30

### 1 Laskutehtäviä

1. Olkoon  $Y$  reaaliarvoinen satunnaismuuttuja, jolle  $EY^2 < \infty$ . Okoon  $Y_1, \dots, Y_N$  i.i.d. otos  $Y$ :n jakaumasta. Osoita, että

$$s_y^2 \xrightarrow{p} \text{Var}(Y),$$

kun  $N \rightarrow \infty$ , missä  $s_y^2 = N^{-1} \sum_{i=1}^N Y_i^2 - \bar{Y}^2$  ja  $\bar{Y} = N^{-1} \sum_{i=1}^N Y_i$ .

2. Olkoon

$$Y_i = \alpha + \beta X_i + u_i, \quad i = 1, \dots, N,$$

missä  $Y_i$ ,  $X_i$  ja  $\epsilon_i$  ovat reaaliarvoisia satunnaismuuttujia. Pienimmän neliösumman estimaattorit ovat

$$\hat{\beta} = \frac{s_{xy}}{s_x^2}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X},$$

missä  $s_{xy} = N^{-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$ ,  $\bar{X} = N^{-1} \sum_{i=1}^N X_i$ ,  $\bar{Y} = N^{-1} \sum_{i=1}^N Y_i$  ja  $s_x^2 = N^{-1} \sum_{i=1}^N (X_i - \bar{X})^2$ .

Osoita, että jos  $(X_i, Y_i)$ ,  $i = 1, \dots, N$ , ovat i.i.d,  $EX_i^2 < \infty$  ja  $EY_i^2 < \infty$ , niin

$$\hat{\beta} \xrightarrow{p} \frac{\text{Cov}(X, Y)}{\text{Var}(X)},$$

ja

$$\hat{\alpha} \xrightarrow{p} EY - \frac{\text{Cov}(X, Y)}{\text{Var}(X)} EX,$$

kun  $N \rightarrow \infty$ .

3. (a) Olkoon  $X_n \sim t_n$  ( $X_n$  noudattaa  $t$ -jakaumaa vapausastein  $n$ ). Perustelee miksi

$$X_n \xrightarrow{d} N(0, 1),$$

kun  $n \rightarrow \infty$ .

Ohje: Koska  $X_n \sim t_n$ , niin

$$X_n = \frac{Z}{\sqrt{U_n/n}},$$

missä  $Z \sim N(0, 1)$ ,  $U_n \sim \chi_n^2$  ( $U_n$  noudattaa  $\chi^2$ -jakaumaa vapausastein  $n$ ) ja  $Z$  ja  $U_n$  ovat riippumattomia. Koska  $U_n \sim \chi_n^2$ , niin

$$U_n = \sum_{i=1}^n u_i^2,$$

missä  $u_1, \dots, u_n$  ovat i.i.d. satunnaismuuttujia jakaumalla  $N(0, 1)$ . Lisäksi voidaan käyttää ominaisuutta, että  $Y_n \xrightarrow{p} c$  implikoi  $Y_n \xrightarrow{d} c$ .

- (b) Olkoon  $X_{m,n} \sim F(m, n)$  ( $X_{m,n}$  noudattaa  $F$ -jakaumaa vapausastein  $m$  ja  $n$ ) ja  $mY_m \sim \chi_m^2$ . Perustelee, että

$$X_{m,n} \xrightarrow{d} Y_m,$$

kun  $n \rightarrow \infty$ .

Ohje: Koska  $X_{m,n} \sim F(m, n)$ , niin

$$X_{m,n} = \frac{U_m/m}{V_n/n},$$

missä  $U_m \sim \chi_m^2$ ,  $V_n \sim \chi_n^2$  ja  $U_m$  ja  $U_n$  ovat riippumattomia.

4. Olkoon

$$Y = \beta_1 X_1 + \dots + \beta_K X_K + u.$$

Määritä matriisi  $R$  ja vektori  $r$  siten, että alla olevat rajoitukset voidaan kirjoittaa matriisimuodossa

$$R\beta = r,$$

missä  $\beta = (\beta_1, \dots, \beta_K)'$  on  $K \times 1$ -vektori,  $R$  on  $Q \times K$ -matriisi,  $1 \leq Q \leq K - 1$  ja  $r$  on  $Q \times 1$ -vektori.

- (a)  $\beta_3 = 0$ ,
- (b)  $\beta_2 = \dots = \beta_K = 0$ ,
- (c)  $\beta_3 = \beta_4$ ,
- (d)  $\beta_1 = 0$ ,  $\beta_3 + \beta_4 = 1$ .

## 2 Tietokonetehtäviä

US-seatbelt data is a balanced panel from 50 U.S. States, plus the District of Columbia, for the years 1983-1997. These data were provided by Professor Liran Einav of Stanford University and were used in his paper with Alma Cohen "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," The Review of Economics and Statistics, 2003, Vol. 85, No. 4, pp 828-843. See

<http://www.stanford.edu/~leinav/pubs/RESTAT2003.pdf>

Datan lukeminen R:ään:

```
file<-"http://cc.oulu.fi/~jklemela/econometrics/SeatBelts.csv"
data<-read.table(file,skip=1,sep=",")
```

Datan lukeminen SAS:iin:

```
FILENAME myurl URL 'http://cc.oulu.fi/~jklemela/econometrics/SeatBelts.txt';
```

```
DATA SeatBelts;
  INFILE myurl firstobs=2;
  INPUT year fips vmt fatalityrate sb_usage speed65 speed70
  drinkage21 ba08 income age primary secondary;
RUN;
```

Muuttujat:

1. state = State (csv-tiedostossa, mutta ei txt-tiedostossa)
2. year = Year
3. fips = State ID Code
4. vmt = Millions of traffic miles per year. (Note: Number of fatalities = fatalityrate  $\times$  vmt)
5. fatalityrate = Number of fatalities per million of traffic miles
6. sb\_usage = Seat belt usage rate
7. speed65 = Binary variable for 65 mile per hour speed limit
8. speed70 = Binary variable for 70 or higher mile per hour speed limit
9. drinkage21 = Binary variable for age 21 drinking age

10. ba08 = Binary variable for blood alcohol limit  $\leq .08\%$
  11. income = Per capita income
  12. age = Mean age
  13. primary = Binary variable for primary enforcement of seat belt laws
  14. secondary = Binary variable for secondary enforcement of seat belt laws
5. (a) Valitse FatalityRate y-muuttujaksi ja sb\_usage, speed65, speed70, drinkage21, ba08, log(income) ja age x-muuttujiksi. Suorita OLS-regressio ja luettele kertoimien pienimmän neliösumman estimaatit. Ohje: R:ssä voi käyttää ohjelmaa "lm" ja SAS:ssa ohjelmaa PROC REG.
  - (b) Suorita t-testit ja F-testi. Ohje: R:ssä voi käyttää komentoa "summary".
  - (c) t-testisuure testille  $H_0 : \beta_k = 0$ ,  $H_1 : \beta_k \neq 0$  on

$$T = \frac{\hat{\beta}_k}{\widehat{\text{sd}}(\beta_k)}.$$

Olkoon  $t$  testisuureen  $T$  havaittu arvo. Määritellään  $p$ -arvo kaavalla  $P(|T| > t)$ . Mikäli  $t$ -testisuure noudattaa  $t$ -jakaumaa vapausastein  $n$ , niin

$$P(|T| > t) = 2(1 - F_n(t)),$$

missä  $F_n$  on  $t_n$ -jakauman kertymäfunktio. Mikäli  $t$ -testisuure noudattaa standardia normaalijakaumaa, niin

$$P(|T| > t) = 2(1 - \Phi(t)),$$

missä  $\Phi$  on standardin normaalijakauman kertymäfunktio. Vertaile arvoja  $2(1 - F_n(t))$  ja  $2(1 - \Phi(t))$ .

### 3 Kertaustehtäviä (eivät kuulu laskuharjoitukseen)

1. Osoita, että

$$\text{Var}(Y) = E\text{Var}(Y | X) + \text{Var}E(Y | X).$$

2. Olkoon  $Z \sim N(0, V)$ , missä  $Z$  on  $K \times 1$ -matriisi ja  $V$  on kääntyvä  $K \times K$ -kovarianssimatriisi. Osoita, että  $Z'V^{-1}Z \sim \chi_K^2$ , missä  $\chi_K^2$  tarkoittaa  $\chi^2$ -jakaumaa vapausastein  $K$ .
3. Muotoile heikko suurten lukujen laki
4. Muotoile keskeinen raja-arvolause.