MTTTP5, luento 20.11.2018

Otossuureita ja niiden jakaumia (jatkuu)

4) Olkoot $X_1, X_2, \ldots, X_n$ satunnaisotos $N(\mu_1, \sigma_1^2)$:sta ja $Y_1, Y_2, \ldots, Y_m$ satunnaisotos $N(\mu_2, \sigma_2^2)$:sta sekä otokset riippumattomia.

Tällöin

$$\boxed{\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)}$$

Esim. 4.2.4 Koneiden A ja B pitäisi valmistaa keskimäärin samanmittaisia tankoja. Molempien koneiden tuottamien tankojen pituuksissa X ja Y (cm) on jonkin verran vaihtelua, jota voidaan luonnehtia normaalijakaumalla, jonka varianssi on 0,20 cm^2. Laadunvalvonnassa seurataan koneiden toimintaa ja valitaan satunnaisesti koneen A tuotannosta 20 ja koneen B tuotannosta 10 tankoa. Koneen A tuotannosta valittujen tankojen keskipituus on 40,0 cm ja koneelta B valittujen 39,5 cm. Tuottavatko koneet keskimäärin samanmittaisia tankoja?

Jos tuottaisivat keskimäärin samanmittaisia tankoja, niin

$$\bar{X} \sim N\left(\mu, \frac{0{,}20}{20}\right) \text{ ja } \bar{Y} \sim N\left(\mu, \frac{0{,}20}{10}\right), \text{ joten}$$

$$\bar{X} - \bar{Y} \sim N\left(0, \underbrace{\frac{0{,}2}{20} + \frac{0{,}2}{10}}_{0{,}03}\right).$$

Tällöin

$$P(\lfloor \overline{X} - \overline{Y} \rfloor > 0{,}5) = 1 - P\left(-0{,}5 \le \overline{X} - \overline{Y} \le 0{,}5\right)$$

$$= 1 - \left(\Phi\left(\frac{0{,}5 - 0}{\sqrt{0{,}03}}\right) - \Phi\left(\frac{-0{,}5 - 0}{\sqrt{0{,}03}}\right)\right)$$

$$= 1 - \left(\Phi(2{,}89) - \Phi(-2{,}89)\right) = 2 - 2\Phi(2{,}89)$$

$$= 0{,}0038$$

Eivät tuota keskimäärin samanmittaisia tankoja. Jos tuottaisivat, niin olisi harvinaista saada otokset, joiden keskiarvojen erotuksen itseisarvo olisi suurempi kuin 0,5 cm.

Luku 5
Parametrien estimointi

5.1 Piste-estimointi

- Estimointi

  populaation tuntemattoman parametrin arviointia otossuureen avulla

- Otossuure

  satunnaisotoksen avulla määritelty funktio

- Otosjakauma

  otossuureen todennäköisyysjakauma

- Estimaattori

  otossuure, jolla estimoidaan populaation tuntematonta parametria

- Estimaattorin keskivirhe

  estimaattorin keskihajonta

- Estimaatti

  estimaattorin arvo (tehdyn otoksen perusteella laskettu)

- Miten estimaattori valitaan?
- Mitä estimaattorista on tiedettävä?
- Mikä on hyvä estimaattori?
- Miksi $\bar{X}$ on hyvä μ:n estimaattori?
- Miksi p on hyvä π:n estimaattori?

Esim. Jos populaatiossa viallisia π %, niin viallisten prosenttiosuus otoksessa

$$p \sim N\left(\pi, \frac{\pi(100-\pi)}{n}\right), likimain.$$

Siis E(p) = π eli estimaattorin odotusarvo on estimoitava parametri. Tätä ominaisuutta kutsutaan harhattomuudeksi, p on π:n harhaton estimaattori.

Estimaattorin p:n hajonta eli keskivirhe on $\sqrt{\frac{\pi(100-\pi)}{n}}$.

Esim. 5.1.2 E($\bar{X}$) = µ, joten $\bar{X}$ on µ:n harhaton estimaattori, $\bar{X}$:n keskivirhe on $\frac{\sigma}{\sqrt{n}}$.

Estimaattorilta vaadittavia ominaisuuksia

- harhattomuus
- mahdollisimman pieni varianssi (tehokkain estimaattori)
- tarkentuvuus eli otoskoon kasvaessa rajatta estimaattorin varianssi lähenee nollaa

Esim. 5.1.2 Voidaan osoittaa, että normaalijakauman tapauksessa $\bar{X}$ on tehokkain μ:n estimaattori eli $\bar{X}$:lla on pienin varianssi harhattomien estimaattoreiden joukossa.

Esim. 5.1.3 $E(S^2) = \sigma^2$

5.2 Luottamusvälejä

Parametria väliestimoidaan nk. luottamusvälin avulla.

Olkoon $Z \sim N(0, 1)$. Määritellään z_α siten, että $P(Z \geq z_\alpha) = \alpha$. Vastaavalla tavalla $z_{\alpha/2}$ siten, että $P(Z \geq z_{\alpha/2}) = \alpha/2$.

Esim. $z_{0,05} = 1,64$, koska $\Phi(1,64) = 0,9495$

$z_{0,025} = 1,96$, koska $\Phi(1,96) = 0,9750$

$z_{0,005} = 2,58$, koska $\Phi(2,58) = 0,9951$

Ks.

http://www.sis.uta.fi/tilasto/mtttp3/kevat2015/zalpha.pdf

5.2.1 Populaation odotusarvon luottamusväli

Olkoon $X_1, X_2, \ldots, X_n$ on satunnaisotos $N(\mu, \sigma^2)$:sta, missä σ^2 tunnettu.

Tällöin

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ ja}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

joten

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

Tästä saadaan

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Satunnaisväli $\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ sisältää μ:n todennäköisyydellä 1-α. Tätä väliä kutsutaan populaation odotusarvon μ 100(1 - α) %:n luottamusväliksi (kaava 4.1). Varmuus eli luottamustaso on 1 - α.

Esim. 5.2.2 Sokerin pussituskone tuottaa pusseja, joiden paino vaihtelee normaalijakauman mukaisesti keskihajontana 2,5 g. Koneeseen tehdään säätöjä ja punnitaan 20 pussia. Näiden keskipainoksi saadaan 1002 g. Voidaanko päätellä, että kone tuottaa keskimäärin kilon pusseja?

95 %:n luottamusväli μ:lle, kun σ tunnettu,

$$\bar{X} \pm z_{0,05/2} \frac{\sigma}{\sqrt{n}}.$$

Saadaan

$$1002 \pm 1{,}96 \cdot 2{,}5/\sqrt{20}$$

$$1002 \pm 1{,}1 \text{ eli } (1000{,}9,\ 1003{,}1)$$

Luottamusväli ei sisällä kiloa. Päätellään, että kone ei tuota keskimäärin kilon pusseja.

Sama päättely 99 %:n luottamusvälin

$$1002 \pm 2{,}58 \cdot 2{,}5/\sqrt{20}$$

$$(1000{,}6,\ 1003{,}4)$$

perusteella.

Olkoon $X_1, X_2, \ldots, X_n$ on satunnaisotos $N(\mu, \sigma^2)$:sta, missä σ^2 tuntematon. Tällöin

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}.$$

Olkoon t_{df} Studentin t-jakaumaa noudattava satunnaismuuttuja.

Määritellään $t_{\alpha,df}$ siten, että $P(t_{df} \geq t_{\alpha,df}) = \alpha$ ja $t_{\alpha/2,df}$ siten, että $P(t_{df} \geq t_{\alpha/2,df}) = \alpha/2$.

Sivut 19-24 seuraavalle luennolle

Esim. 5.2.4

$t_{0,05,\ 10}$ = 1,812, $t_{0,05,\ 30}$ = 1,697

$t_{0,01,\ 10}$ = 2,764, $t_{0,01,\ 30}$ = 2,457

Satunnaismuuttujan $t = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$ perusteella voidaan

johtaa (kuten kaava 4.1) populaation odotusarvon μ

100(1 - α) %:n luottamusväli, kun σ^2 tuntematon.

Saadaan

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}. \qquad \text{kaava (4.2)}$$

Esim. 5.2.5 Poikien keskimääräinen syntymäpituus (SAIDIT-aineisto) $\bar{x} = 50{,}95$, $s = 1{,}97$, $n = 65$, $t_{0,05/2,64} \approx 2{,}000$. Nyt 95 %:n luottamusväli on

$$50{,}95 \pm t_{\frac{0{,}05}{2};65-1} \frac{1{,}97}{\sqrt{65}}$$

$$50{,}95 \pm 2 \cdot \frac{1{,}97}{\sqrt{65}}.$$

Poikien keskipituuden arvellaan olevan välillä

(50,46, 51,44).

Esim. Cooperin testin tulos (CTESTI-aineisto), 15-vuotiaat, $\bar{x}$ = 2534, s = 255, n = 28, $t_{0,05/2,27}$ = 2,052, 95 %:n luottamusväli odotusarvolle

$$2534 \pm 2{,}052 \cdot \frac{255}{\sqrt{28}}$$ eli väli (2435, 2633).

SPSS-tulos

T-Test

luottamusväli μ:lle $\bar{x} \pm t_{\alpha/2;n-1} \cdot s/\sqrt{n}$

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
cooper	$n =$ 28	2533,93 $= \bar{x}$	255,460 $= s$	48,277 $= s/\sqrt{n}$

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
cooper	52,487	27	,000	2533,929	2434,87	2632,99

$t_{0{,}025,27} = 2{,}052$

$2533{,}93 - 2{,}052 \cdot 48{,}277$

$2533{,}93 + 2{,}052 \cdot 48{,}277$

Esim. 5.2.6 Keskimääräiset neliövuokrat Tampereen Hervannassa (2011), aineisto Tre_vuokra-asunnot_2011.sav sivulla https://coursepages.uta.fi/mtttp1/esimerkkiaineistoja/

$\bar{x}$ = 12,32, s = 2,25, n = 26,

95 %:n luottamusväli odotusarvolle (11,41, 13,23)

SPSS-tulos

Neliövuokra Hervannassa, aineisto Tre_vuokra-asunnot_2011.sav sivulla http://www.uta.fi/sis/mtt/mtttp1/aineistoja.html.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Neliövuokra	26	12,3190	2,25281	,44181

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Neliövuokra	27,883	25	,000	12,31904	11,4091	13,2290

$t_{0.025,25} = 2,060$, luottamusväli $12,3190 \pm 2,060 \cdot 2,25281/\sqrt{26}$