MTTTP5, luento 22.11.2018

Luottamusväli, määritelmä

Olkoot A ja B satunnaisotoksen perusteella määriteltyjä satunnaismuuttujia. Väli (A, B) on parametrin θ 100(1 - α) %:n luottamusväli, jos

$P(A \leq \theta \leq B) = 1 - \alpha$.

5.2.1 Populaation odotusarvon luottamusväli (jatkoa)

Olkoon $X_1, X_2, \ldots, X_n$ on satunnaisotos $N(\mu, \sigma^2)$:sta, missä σ^2 tunnettu. Tällöin populaation odotusarvon μ 100(1 - α) %:n luottamusväli on

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}. \qquad \text{Kaava 4.1.}$$

Olkoon $X_1, X_2, \ldots, X_n$ on satunnaisotos $N(\mu, \sigma^2)$:sta, missä σ^2 <u>tuntematon</u>. 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 $P(t_{df} \geq t_{\alpha/2,df}) = \alpha/2$.

Esim. 5.2.4

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

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

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, $t_{0,05/2,64}$ ≈ 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. 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 $12,3190 \pm 2,060 \cdot 2,25281/\sqrt{26}$

SPSS-ohjeet:

Neliövuokra: Transform -> Compute -> Neliövuokra = Vuokra/Neliöt

Vain Hervanta analyyseihin: Data -> Select Cases -> If condition is satisfied -> Alue=8 (tai Kaupunginosa='Hervanta')

Luottamusväli: Analyze -> Compare Means -> One-Sample T Test -> Test Variable Neliövuokra

Esim. Eräs yritys tarjoaa valmennuskurssia yliopistoon pyrkijöille. Yritys haluaa tutkia kurssinsa tehokkuutta. Tutkitaan pareja, joilla on samanlaiset lähtötiedot. Toinen osallistuu valmennuskurssille, toinen ei. Saadaan aineisto, jossa pyrkijöiden valintakoepisteet.

	Osallistui	Ei osallistunut	Erotus D
Pari 1	82	75	7
Pari 2	73	71	2
Pari 3	59	52	7
Pari 4	48	46	2
Pari 5	69	70	-1
Pari 6	93	83	10

Tarkastellaan erotusta D, muodostetaan sen odotusarvolle luottamusväli, joka on

$$\bar{D} \pm t_{\alpha/2,n-1} \frac{s_D}{\sqrt{n}}$$

$\bar{d}$=(7+2+7+2-1+10)/6=4,5

$$s_D^2 = \frac{(7^2 + 2^2 + 7^2 + 2^2 + (-1)^2 + 10^2) - 6 \cdot 4{,}5^2}{6-1} = 17{,}1$$

95 %:n luottamusväli on $4{,}5 \pm 2{,}571 \cdot \frac{\sqrt{17{,}1}}{\sqrt{6}}$ eli väli (0,16, 8,84). Erotuksen odotusarvon ei siis ajatella olevan nolla, joten valmennuskurssilla on vaikutusta.

http://vassarstats.net/-> t-Tests & Procedures

VassarStats Printable Report
0.95 and 0.99 Confidence Intervals for the Estimated Mean of a Population
Thu Jan 15 2015 16:07:22 GMT+0200 (FLE Standard Time)

Values entered:
X = {7,2,7,2,-1,10}

Summary Values:

N	= 6
ΣX	= 27
ΣX^2	= 207
mean	= 4.5
variance	= 17.1
std. dev.	= 4.1352
std. error	= 1.6882
df	= 5
$t_{crit(.05)}$	= 2.57
$t_{crit(.01)}$	= 4.03

Confidence Intervals for Estimated Mean of Population
For .95 CI: 4.5±4.3387
For .99 CI: 4.5±6.8034

SPSS-tulosteet

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Erotus	$n =$ 6	$\bar{x} =$ 4,5000	$s =$ 4,13521	1,68819 $= s/\sqrt{n}$

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper
Erotus	2,666	5	,045	4,50000	,1604	8,8396

$\bar{x} \pm t_{0,05/2;5} \cdot s/\sqrt{n}$

$t_{0,025;5} = 2,571$

5.2.2 Prosentuaalisen osuuden luottamusväli

Jos populaatiossa viallisia π %, niin viallisten prosenttiosuus otoksessa

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

Tällöin

$$Z = \frac{p - \pi}{\sqrt{\pi(100 - \pi)/n}} \sim N(0,1), likimain.$$

Tämän perustella saadaan π:n 100(1 - α) %:n luottamusväli

$$p \pm z_{\alpha/2} \sqrt{\frac{p(100-p)}{n}}.$$

Kaava 4.3.

Esim. Ruletissa 37 numeroa, joista pyöritettäessä jokaisella pitäisi olla sama todennäköisyys tulla tulokseksi. Pelipaikka voittaa numerolla nolla. Rulettia pyöritetään 3700 kertaa. Saadaan nollia 140 eli 3,78 %. Toimiiko ruletti oikein?

Lasketaan 99 %:n luottamusväli nollien % -osuudelle. Nyt α = 0,01, $z_{0,01/2} = z_{0,005} = 2,57$, koska $\Phi(2,57) = 0,9949$, p = (140/3700)·100 = 3,78, luottamusväli

$$3{,}78 \pm 2{,}57\sqrt{\frac{3{,}78(100-3{,}78)}{3700}}.$$

Nollien prosenttiosuuden arvellaan olevan välillä 2,97–4,59.

Jos ruletti toimisi oikein, niin nollia pitäisi tulla (1/37)·100 % = 2,70 %. Tämä ei kuulu luottamusvälille, joten päätellään ruletin toimivan väärin.

Jos laskettaisiin 95 %:n luottamusväli, saataisiin väli (3,17, 4,39).

Esim. Hyväkuntoisten osuus myydyistä kolmioista, aineisto Tre_myydyt_kolmiot_2010.sav sivulla https://coursepages.uta.fi/mtttp1/esimerkkiaineistoja/

Kunto

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Hyvä	120	69,0	**69,0**	69,0
	Tyydyttävä	51	29,3	29,3	98,3
	Huono	3	1,7	1,7	100,0
	Total	**174**	100,0	100,0	

95 %:n luottamusväli on $69 \pm 1{,}96\sqrt{\frac{69(100-69)}{174}}$

Hyväkuntoisten % -osuuden arvellaan olevan välillä 62,1 – 75,9.

Esim. Hyväkuntoisten osuus myydyistä kolmioista sijainnin mukaan tarkasteltuna

Kunto * Sijainti Crosstabulation

			Sijainti				
			Keskusta	Kaleva, Amuri, Pyynikki	Hatanpää, Nekala, Epilä, Kissanmaa	Lentävänniemi	Hervanta
Kunto	Hyvä	Count	23	21	25	27	24
		% within Sijainti	**76,7%**	**56,8%**	75,8%	71,1%	66,7%
	Tyydyttävä	Count	7	14	7	11	12
		% within Sijainti	23,3%	37,8%	21,2%	28,9%	33,3%
	Huono	Count	0	2	1	0	0
		% within Sijainti	0,0%	5,4%	3,0%	0,0%	0,0%
Total		Count	**30**	**37**	33	38	36
		% within Sijainti	100,0%	100,0%	100,0%	100,0%	100,0%

95 %:n luottamusväli

- Keskusta: $76{,}7 \pm 1{,}96\sqrt{\frac{76{,}7(100-76{,}7)}{30}}$ eli 61,5 – 91,9
- Kaleva, Amuri, Pyynikki: $56{,}8 \pm 1{,}96\sqrt{\frac{56{,}8(100-56{,}8)}{37}}$

  eli 40,8 – 72,8.

SPSS-ohjeet:

Frekvenssijakauma:

Analyze -> Descriptive Statistics -> Frequencies -> Kunto

Ristiintaulukko:

Analyze -> Descriptive Statistics -> Crosstabs -> Row(s) -> Kunto, Column(s) -> Sijainti -> Cells... prosenttijakaumat

5.2.3 Kahden populaation odotusarvojen erotuksen luottamusväli

Esim. Kolmioiden keskineliöhinnat ja keskineliöhintojen luottamusvälit sijainnin mukaan tarkasteltuna

One-Sample Statistics

Sijainti		N	Mean	Std. Deviation	Std. Error Mean
Keskusta	Neliöhinta	**30**	**2598,7142**	517,63351	**94,50652**
Kaleva, Amuri, Pyynikki	Neliöhinta	37	2353,2405	368,52174	60,58460
Hatanpää, Nekala, Epilä, Kissanmaa	Neliöhinta	33	1963,3219	538,43268	93,72910
Lentävänniemi	Neliöhinta	38	1630,4808	429,16341	69,61950
Hervanta	Neliöhinta	36	1371,3906	136,52228	22,75371

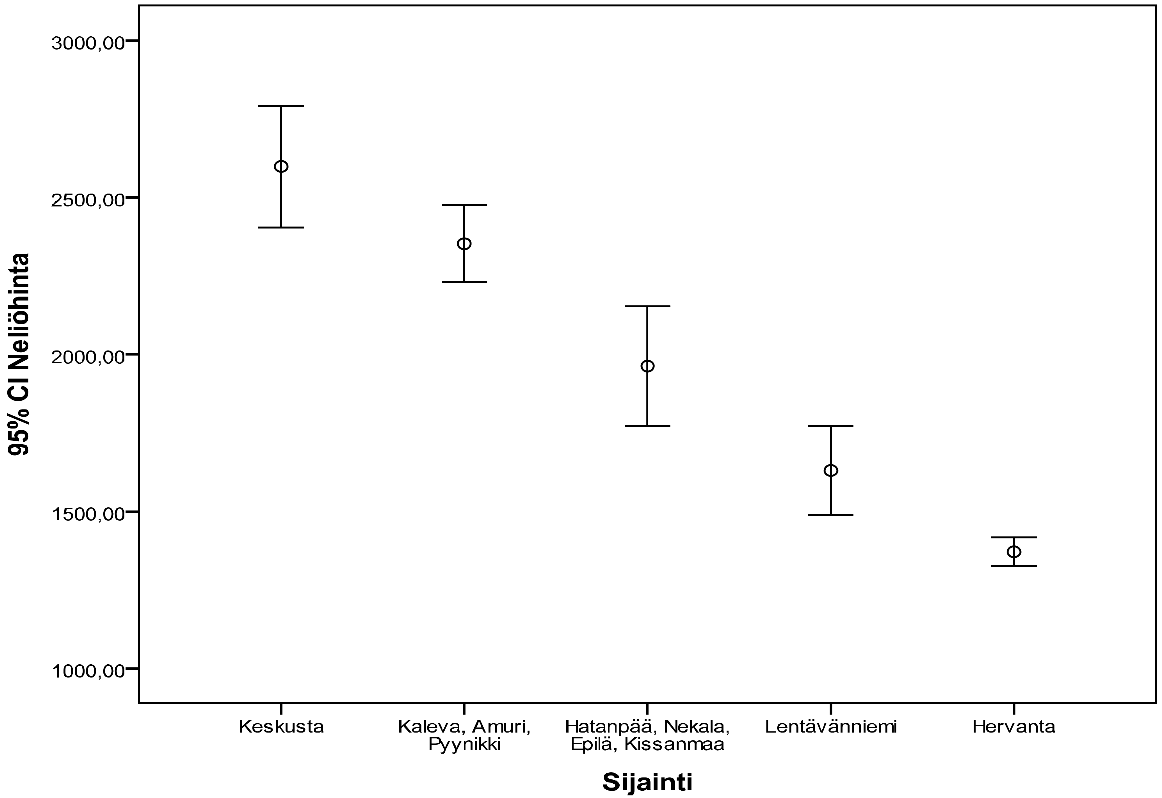
95 %:n luottamusvälit odotusarvolle

Sijainti		Lower	Upper
Keskusta	Neliöhinta	**2405,4267**	**2792,0017**
Kaleva, Amuri, Pyynikki	Neliöhinta	2230,3692	2476,1118
Hatanpää, Nekala, Epilä, Kissanmaa	Neliöhinta	1772,4019	2154,2418
Lentävänniemi	Neliöhinta	1489,4183	1771,5433
Hervanta	Neliöhinta	1325,1981	1417,5831

Luottamusväli keskustassa

$$2598{,}71 \pm t_{0{,}05/2;30-1} \frac{517{,}63}{\sqrt{30}} = 2598{,}71 \pm 2{,}045 \cdot 94{,}51.$$

Luottamusvälit graafisesti:



Esim. Kolmioiden keskineliöhinnat, kahden sijainnin vertailu, luottamusväli odotusarvojen erotukselle

Group Statistics

	Sijainti	N	Mean	Std. Deviation	Std. Error Mean
Neliöhinta	Keskusta	30	2598,7142	517,63351	94,50652
	Kaleva, Amuri, Pyynikki	37	2353,2405	368,52174	60,58460

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						95% Confidence Interval of the Difference	
		F	Sig.	Mean Difference	Std. Error Difference	Lower	Upper
Neliöhinta	Equal variances assumed	4,462	,039	245,47371	108,42446	**28,93510**	**462,01231**
	Equal variances not assumed			245,47371	112,25852	20,08589	470,86152

SPSS-ohjeet:

Neliöhinta: Transform -> Compute -> Neliöhinta = Hinta/Neliöt

Luottamusväli: Analyze -> Compare Means -> Independent-Samples T Test -> Test Variable(s): Neliöhinta, Grouping Variable: Sijainti: Group 1: 1 (Keskusta), Group 2: 2 (Kaleva, Amuri, Pyynikki)

$X_1, X_2, \ldots, X_n$ on satunnaisotos $N(\mu_1, \sigma_1^2)$:sta,

$Y_1, Y_2, \ldots, Y_m$ on satunnaisotos $N(\mu_2, \sigma_2^2)$:sta.

Oletetaan, että varianssit tunnettuja ja satunnaisotokset riippumattomia. Tällöin otoskeskiarvojen erotus

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

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1).$$

Tästä voidaan johtaa 100(1 - α) %:n luottamusväli erotukselle (μ_1- μ_2). Luottamusväliksi saadaan

$$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}. \qquad \text{Kaava 4.4.}$$

Jos varianssit tuntemattomia, mutta voidaan olettaa, että $\sigma_1^2 = \sigma_2^2$, niin 100(1 - α) %:n luottamusväli erotukselle (μ_1- μ_2) on

$$\overline{X} - \overline{Y} \pm t_{\frac{\alpha}{2},n+m-2}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{1}{m}} \qquad \text{Kaava 4.5.}$$

Tuntematonta varianssia estimoidaan otosvarianssien avulla

$$\hat{\sigma}^2 = \frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2} = s^2.$$

Esim. Kolmioiden keskineliöhinnat, kahden sijainnin vertailu

Group Statistics

	Sijainti	N	Mean	Std. Deviation	Std. Error Mean
Neliöhinta	Keskusta	$n =$ 30	2598,7142	$s_x =$ 517,63351	94,50652
	Kaleva, Amuri, Pyynikki	$m =$ 37	2353,2405	$s_y =$ 368,52174	60,58460

$$s^2 = \frac{(30-1)\,517{,}63351^2 + (37-1)\,368{,}52174^2}{30+37-2} \approx 441{,}32^2$$

$s\sqrt{\frac{1}{n}+\frac{1}{m}}$

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						95% Confidence Interval of the Difference	
		F	Sig.	Mean Difference	Std. Error Difference	Lower	Upper
Neliöhinta	Equal variances assumed	4,462	,039	245,47371	108,42446	**28,93510**	**462,01231**
	Equal variances not assumed			245,47371	112,25852	20,08589	470,86152

$t_{0,05/2,\,65} \approx 2$

luottamusväli:

$245{,}47371 \pm 2 \cdot 108{,}42446$