

MTTTP5, luento 4.12.2018

6.1.3 Kahden jakauman sijainnin vertailu (jatkoa)

Tutkimustilanteita

y = neliöhinta

x = sijainti (2 aluetta)

y = lepopulssi

x = sukupuoli

y = musikaalisuus

x = sukupuoli

y = kaksion koko

x = sijainti (keskusta/ei-keskusta)

Y = tehopisteet

x = pelipaikka (puolustaja/hyökkääjä)

y = lumilaudan hinta

x = kenelle tarkoitettu (miehille/naisille)

$$\boxed{H_0 : \mu_1 = \mu_2} \quad \text{tai} \quad \boxed{H_0 : \mu_1 - \mu_2 = 0}$$

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

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

Oletetaan, että varianssit tunnettuja ja satunnaisotokset riippumattomia. Kun H_0 on tosi, niin

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

Tätä käytetään testisuureena, päättely kuten aiemminkin normaalijakaumaa noudattavien testisuureiden tapauksessa.

Variansseja ei useinkaan tunneta. Jos oletetaan ne tuntemattomiksi mutta yhtä suuriksi, niin H_0 :n ollessa tosi

$$t = \frac{\bar{X} - \bar{Y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2),$$

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

Tätä käytetään testisuureena, päättely tehdään kuten aiemminkin t-jakaumaa noudattavien testisuureiden tapauksessa.

Oletusta varianssien yhtäsuuruudesta voidaan testata. SPSS tulostaa tähän liittyvän p-arvon.

Esim. 6.1.9 Lepopulssi

Group Statistics

| | Sukupuoli | N | Mean | Std. Deviation | Std. Error Mean |
|------------|-----------|----------|---------|------------------|-----------------|
| Lepopulssi | Mies | $n = 44$ | 70,6364 | 11,27684 = S_1 | 1,70005 |
| | Nainen | $m = 36$ | 77,5556 | 9,80800 = S_2 | 1,63467 |

Independent Samples Test

| | | Levene's Test for Equality of Variances | t-test for Equality of Means | | |
|------------|-----------------------------|---|------------------------------|--------|--------|
| | | F | Sig. | t | df |
| Lepopulssi | Equal variances assumed | ,125 | ,725 | -2,893 | 78 |
| | Equal variances not assumed | | | -2,934 | 77,685 |

$H_0: \sigma_1^2 = \sigma_2^2$

$H_0: \mu_1 = \mu_2$

Independent Samples Test

t-test for Equality of Means

| | | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
|------------|-----------------------------|-----------------|-----------------|-----------------------|
| Lepopulssi | Equal variances assumed | ,005 | -6,91919 | 2,39180 |
| | Equal variances not assumed | ,004 | -6,91919 | 2,35845 |

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

$$s^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

Analyze -> Compare Means -> Independent-Samples T Test -> Test Variable(s): Lepopulssi, Grouping Variable: Sukupuoli: Group 1: 1 (Mies), Group 2: 2 (Nainen)

Esim. 6.1.10 Testi lasten kehityshäiriön tunnistamiseen

Suoritusajat testissä ryhmittäin:

Normaali 204, 218, 197, 183, 227, 233, 191

Kehityshäiriö 243, 228, 261, 202, 343, 242, 220, 239

$$H_0 : \mu_N = \mu_K$$

$$H_1 : \mu_N < \mu_K$$

$$\sum x_i = 204 + \dots + 191 = 1453$$

$$\sum x_i^2 = 204^2 + \dots + 191^2 = 303737$$

$$SS_N = 303737 - 7 \cdot (1453/7)^2 = 2135,714$$

$$\sum y_i = 243 + \dots + 239 = 1978$$

$$\sum y_i^2 = 243^2 + \dots + 239^2 = 501692$$

$$SS_K = 501692 - 8 \cdot (1978/8)^2 = 12631,5$$

$$s^2 = \frac{2135,714 + 12631,5}{7 + 8 - 2} = 1135,94, s = 33,7$$

$$t_{hav.} = \frac{\frac{1453}{7} - \frac{1978}{8}}{33,7 \sqrt{\frac{1}{7} + \frac{1}{8}}} = -2,28$$

$$-t_{0,01;13} = -2,65 < t_{hav.} < -2,16 = -t_{0,025;13}$$

H_0 voidaan hylätä 2,5 %:n riskitasolla, mutta ei 1 %:n riskitasolla.

SPSS-tulos

Group Statistics

| | ryhmä | N | Mean | Std. Deviation | Std. Error Mean |
|-------|---------------|-----------|----------|----------------------|-----------------|
| testi | normaali | $n_N = 7$ | 207,5714 | \bar{x}_N 18,86670 | S_N 7,13094 |
| | kehityshäiriö | $n_K = 8$ | 247,2500 | \bar{x}_K 42,47941 | S_K 15,01874 |

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | |
|-------|-----------------------------|---|------------|------------------------------|-------|
| | | F | Sig. | t | df |
| testi | Equal variances assumed | ,926 | ,353 | -2,275 | 13 |
| | Equal variances not assumed | | H_0 hyp. | -2,387 | 9,923 |

$H_0: \sigma_N^2 = \sigma_K^2$

$H_0: \mu_N = \mu_K$

Independent Samples Test

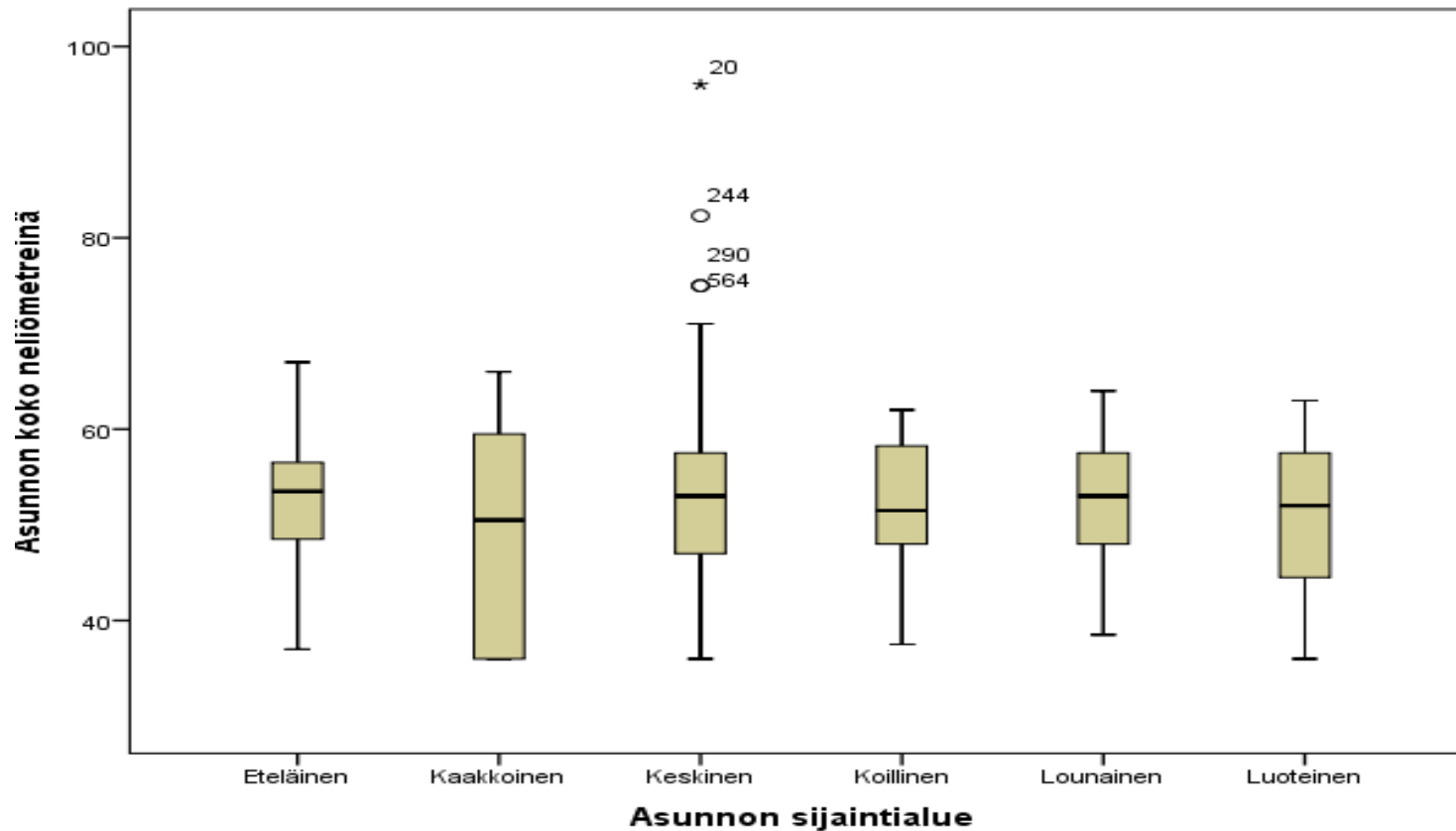
t-test for Equality of Means

| | | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
|-------|-----------------------------|-----------------|-----------------|-----------------------|
| testi | Equal variances assumed | ,041 | -39,67857 | 17,44332 |
| | Equal variances not assumed | ,038 | -39,67857 | 16,62567 |

$$H_1: \mu_N < \mu_K$$

$$p\text{-arvo } 0,041/2 = 0,0205$$

Esim. Ovatko kaksiot keskimäärin erikokoisia Tampereen eri alueilla? Aineisto: Tre_myydyt_kaksiot_2016.sav sivulla <https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>



Ei eroja:

| Group Statistics | | | | |
|-------------------------------|----------------------|----|-------|----------------|
| | Asunnon sijaintialue | N | Mean | Std. Deviation |
| Asunnon koko neliömetreinä | Eteläinen | 53 | 52,46 | 5,965 |
| | Koillinen | 44 | 52,09 | 6,379 |

| Independent Samples Test | | | | | | |
|-------------------------------|-----------------------------|---|------|------|--------|-----------------|
| | | Levene's Test for Equality of Variances | | | | |
| | | F | Sig. | t | df | Sig. (2-tailed) |
| Asunnon koko neliömetreinä | Equal variances assumed | ,248 | ,620 | ,296 | 95 | ,768 |
| | Equal variances not assumed | | | ,294 | 89,211 | ,769 |

On eroja:

Group Statistics

| | | N | Mean | Std. Deviation |
|-------------------------------|------------|-----|-------|----------------|
| Asunnon koko neliömetreinä | Kaakkoinen | 112 | 49,75 | 10,217 |
| | Keskinen | 277 | 52,56 | 7,888 |

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t | | |
|-------------------------------|-----------------------------|---|------|--------|---------|-----------------|
| | | F | Sig. | t | df | Sig. (2-tailed) |
| Asunnon koko neliömetreinä | Equal variances assumed | 25,812 | ,000 | -2,905 | 387 | ,004 |
| | Equal variances not assumed | | | -2,608 | 167,057 | ,010 |

Esim. Eräs yritys valmistaa 10 metrin teräskaapeleita tehtaissa X ja Y. Tarkasteltiin kaapeleiden murtolujuuksia (kilonewtoneina). Tehtaan X tuotannosta valittiin satunnaisesti 9 ja tehtaan Y tuotannosta 16 kaapelia, joiden murtolujuudet mitattiin. Saatiin tulokset:

$$\text{Tehtas X: } \bar{x} = 30,11, \sum(x_i - \bar{x})^2 = 0,8013$$

$$\text{Tehtas Y: } \bar{y} = 29,63, \sum(y_i - \bar{y})^2 = 3,0206$$

Asetetaan

$$H_0 : \mu_X = \mu_Y$$

$$H_1 : \mu_X \neq \mu_Y$$

Nyt

$$s^2 = \frac{0,8013 + 3,0206}{9 + 16 - 2} = 0,1662, s = 0,4076$$

$$t_{hav.} = \frac{30,11 - 29,63}{0,4076 \sqrt{\frac{1}{9} + \frac{1}{16}}} = 2,83$$

Pienin riskitaso, jolla H_0 voidaan hylätä, on $0,0047 \times 2 = 0,0094$, ks. <http://stattrek.com/online-calculator/t-distribution.aspx> tai http://onlinestatbook.com/2/calculators/t_dist.html

Keskimääräisissä murtolujuuksissa on siis eroja.

Yhteenveto kurssin sisällöstä, joitain esimerkkejä

Todennäköisyyslaskentaa

- satunnaisilmiö ja tapahtuma
- klassinen todennäköisyys
- todennäköisyyslaskennan aksioomat ja laskusääntöjä
- kombinatoriikkaa

Todennäköisyysjakaumia

- satunnaismuuttuja ja todennäköisyysjakauma
- diskreetti satunnaismuuttuja
- jatkuva satunnaismuuttuja
- odotusarvon ja varianssin ominaisuuksia
- joitain todennäköisyysjakaumia

Bernoulli-jakauma

binomijakauma

diskreetti tasajakauma

jatkuva tasajakauma

normaalijakauma

t-jakauma

Satunnaisotos, otossuure ja otosjakauma

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

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

Esim.

Jos X_1, X_2, \dots, X_n on satunnaisotos $N(\mu, \sigma^2)$:sta, niin

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Esim.

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

Tällöin

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

Parametrien estimointi

- piste-estimointi
- estimaattori, estimaattorin ominaisuudet, estimaattorin keskivirhe
- luottamusvälit (väliestimointi)

µ: lle

Aineisto Lumilaudat.sav sivulla

<https://coursepages.uta.fi/mtt1/esimerkkiaineistoja/>

- Miesten lumilautojen keskihinta

One-Sample Statistics

| Kenelle lauta on tarkoitettu | | N | Mean | Std. Deviation | Std. Error Mean |
|------------------------------|----------------------|----|--------|----------------|-----------------|
| Miehet | Laudan hinta euroina | 51 | 522,84 | 136,219 | 19,074 |
| Naiset | Laudan hinta euroina | 18 | 511,06 | 102,737 | 24,215 |
| Lapset | Laudan hinta euroina | 11 | 278,18 | 34,005 | 10,253 |

One-Sample Test

Test Value = 0

95% Confidence Interval of the
Difference

| Kenelle lauta on tarkoitettu | | Lower | Upper |
|------------------------------|----------------------|--------|--------|
| Miehet | Laudan hinta euroina | 484,53 | 561,16 |
| Naiset | Laudan hinta euroina | 459,97 | 562,15 |
| Lapset | Laudan hinta euroina | 255,34 | 301,03 |

$$\bar{X} \pm t_{0,05/2;50} \cdot 136,219/\sqrt{51}$$

$$522,84 \pm 2 \cdot 19,074$$

π: lle

Aineisto Audi_A6.sav sivulla

<https://coursepages.uta.fi/mtt1/esimerkkiaineistoja/>

- Diesel-autojen osuus

Käyttövoima

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|----------|-----------|---------|---------------|--------------------|
| Valid | Bensiini | 99 | 74,4 | 74,4 | 74,4 |
| | Diesel | 34 | 25,6 | 25,6 | 100,0 |
| | Total | 133 | 100,0 | 100,0 | |

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

$$25,6 \pm 1,96 \sqrt{25,6 \cdot 74,4 / 133}$$

$$25,6 \pm 7,7$$

$$(18,2, 33,0)$$

$(\mu_1 - \mu_2)$:lle

Aineisto Plasma.sav sivulla

<https://coursepages.uta.fi/mtt1/esimerkkiaineistoja/>

- Onko keskimääräisissä kolesterolimäärissä eroja miehillä ja naisilla?

Group Statistics

| | sukupuoli | N | Mean | Std. Deviation | Std. Error Mean |
|----------------------------------|-----------|-----|----------|----------------|-----------------|
| päivässä saatujen kalorien määrä | mies | 42 | 2155,786 | 916,5653 | 141,4291 |
| | nainen | 273 | 1741,404 | 620,2701 | 37,5405 |

Independent Samples Test

| | | Levene's Test for Equality of Variances | t-test for Equality of |
|----------------------------------|-----------------------------|---|------------------------|
| | | F | Sig. |
| päivässä saatujen kalorien määrä | Equal variances assumed | 2,007 | 3,750. |
| | Equal variances not assumed | | 2,832 |

Handwritten notes: $H_0: \sigma_1^2 = \sigma_2^2$ and a circled 'Sig. ,158' in the table above.

Independent Samples Test

| | | t-test for Equality of Means | | |
|----------------------------------|-----------------------------|------------------------------|-----------------|-----------------|
| | | df | Sig. (2-tailed) | Mean Difference |
| päivässä saatujen kalorien määrä | Equal variances assumed | 313 | ,000 | 414,3821 |
| | Equal variances not assumed | 46,946 | ,007 | 414,3821 |

Independent Samples Test

| | | t-test for Equality of Means | | |
|----------------------------------|-----------------------------|------------------------------|---|----------|
| | | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | Lower | Upper |
| päivässä saatujen kalorien määrä | Equal variances assumed | 110,4912 | 196,9827 | 631,7815 |
| | Equal variances not assumed | 146,3266 | 120,0019 | 708,7622 |

$$s^2 = \frac{(n-1)s_x^2 + (n-1)s_y^2}{n+n-2}$$

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

$$\bar{x} - \bar{y} \pm t_{\alpha/2, n+m-2} \cdot s\sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$t_{0.05/2; 313} = 1,96$$

$$414,3821 \pm 1,96 \cdot 110,4912$$

Hypoteesien testaus

$$H_0: \mu = \mu_0$$

- Onko kynttilöiden keskimääräinen palamisaika 24 h

20 kynttilän palamisajat minuutteina:

1426, 1438, 1425, 1435, 1432, 1434,
1427, 1441, 1439, 1427, 1435, 1435,
1427, 1432, 1445, 1439, 1429, 1435,
1434, 1447

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|-------------|----|---------|----------------|--------------------|
| Palamisaika | 20 | 1434,01 | S = 6,149 | 1,375 s/\sqrt{n} |

One-Sample Test

Test Value = 1440

| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
|-------------|--------|----|-----------------|-----------------|---|-------|
| | | | | | Lower | Upper |
| Palamisaika | -4,354 | 19 | ,000 | -5,987 | -8,86 | -3,11 |

$$H_0: \mu = 1440$$

$$H_1: \mu < 1440$$

$$t_{0,005,19} = 2,861$$

$$< 0,001$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}, \text{ kun } H_0 \text{ tosi.}$$

$$H_0: \pi = \pi_0$$

Aineisto Saidit.sav sivulla

https://coursepages.uta.fi/mtttp1/esimerkkia_ineistoja/

Syntyykö tyttöjä ja poikia yhtä paljon?

| | | sex | | | |
|-------|-------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | poika | 65 | 54,2 | 54,2 = p | 54,2 |
| | tyttö | 55 | 45,8 | 45,8 | 100,0 |
| | Total | 120 | 100,0 | 100,0 | |

$$H_0: \pi = 50$$

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(100 - \pi_0)/n}}$$

$$Z_{hav.} = \frac{54,2 - 50}{\sqrt{50 \cdot 50 / 120}} = 0,92$$

Yksisuuntaisessa testissä

$$p\text{-arvo} = P(Z > 0,92) = 1 - \Phi(0,92) = 0,1788$$

Poikia ja tyttöjä syntyy yhtä paljon

$$H_0: \mu_1 = \mu_2$$

Aineisto Plasma.sav sivulla

<https://coursepages.uta.fi/mtt1/esimerkkia/aineistoja/>

- Onko keskimääräisissä kolesterolimäärissä eroja miehillä ja naisilla?

$t = 3,750$, $p\text{-arvo} < 0,001$, on eroja

Lopuksi

- Osaamistavoitteet

http://www.sis.uta.fi/tilasto/mtttp5/syksy2018/luento_1.pdf

- Tentit

ks. <https://coursepages.uta.fi/mtttp5/syksy-2018/kurssi-info/>

- Jatkoksi: Tilastomenetelmien perusteet MTTTA1

<https://www10.uta.fi/opas/opetusohjelma/marjapuu.ro.htm?id=38653>