

MTTTA1 Tilastomenetelmien perusteet 5 op
Luento 10.1.2019

1 Kokonaisuudet johon opintojakso kuuluu

[https://www10.uta.fi/opas/opintojakso.htm?rid=14600
&idx=1&uiLang=fi&lang=fi&lvv=2018](https://www10.uta.fi/opas/opintojakso.htm?rid=14600&idx=1&uiLang=fi&lang=fi&lvv=2018)

2 Osaamistavoitteet

<https://www10.uta.fi/opas/opintojakso.htm?rid=14600&idx=1&uiLang=fi&lang=fi&lvv=2018>

- Opiskelija osaa käyttää opintojaksolla esiteltyjä tilastollisia menetelmiä sekä ymmärtää niihin liittyvän teorian.
- Hän osaa annetussa tutkimustilanteessa suorittaa tilastollisen päättelyn joko valmiiksi annettujen tai itse laskemiensa tulosten perusteella.
- Hän osaa valita asetettuun tutkimusongelmaan liittyen sopivan menetelmän, suorittaa tilanteeseen sopivalla ohjelmistolla kyseisen analyysin sekä tulkita saadut tulokset.

Esim. Tampereella keväällä 2006 myynnissä olleita kerrostalohuoneistoja, aineisto

http://www.sis.uta.fi/tilasto/tiltp_aineistoja/Asunnot_2006.sav sivulta

https://coursepages.uta.fi/mttp1/esimerkkiaineistoj_a/

Tutkimuskohteita

1) Vaikuttaako sijainti neliöhintaan?

y = neliöhinta

x = sijainti

SPSS-harj. 1 teht. 3a

2) Vaikuttaako huoneiden lukumäärä
neliöhintaan? Miten sijainti vaikuttaa tähän
riippuvuuteen?

y = neliöhinta

x = huoneiden lukumäärä (luokiteltuna)

z = sijainti

SPSS-harj. 1 teht. 3b

3) Vaikuttaako sijainti huoneiden lukumäärään?

y = huoneiden lukumäärä (luokiteltuna)

x = sijainti

SPSS-harj. 2 teht. 3

4) Miten huoneiston koko vaikuttaa hintaan?

Miten sijainti vaikuttaa tähän riippuvuuteen?

y = hinta

x = neliöt

z = sijainti

SPSS-harj. 3 teht. 2

3 Kurssin kotisivu

<https://coursepages.uta.fi/mttta1/kevat-2019/>

- Opetus
- Kurssi-info (sisältö, tentit, harjoitushyvyty)
- Luennot, luentorunko, kaavat, taulukot
- Harjoitukset, tehtävät, ohjeet (Moodle), ratkaisut
- Esimerkkiaineistoja
- Oheiskirjallisuutta
- Usein kysyttyä
- Linkkejä
- Palaute

4 Kertausta

Seuraaviin kohtiin 1) – 8) on koottu lyhyesti olennaisimmat asiat, jotka oletetaan opintojaksolla tunnetuiksi aiempien opintojen perusteella.

1) Empiiriset jakaumat

- yksiulotteiset
taulukot, graafiset esitykset, tunnusluvut
- kaksiulotteiset
ristiintaulukko, pisteparvi, korrelaatiokerroin
- ehdolliset jakaumat
riippuvuus, ehdolliset tunnusluvut, laatikko-
jana-kuvio
- toteutus SPSS:llä (tai muulla ohjelmistolla)

2) Satunnaismuuttuja X

- todennäköisyysjakauma, tiheysfunktio $f(x)$
- kertymäfunktio $F(x) = P(X \leq x)$
- $E(X) = \mu$, $\text{Var}(X) = \sigma^2$

3) Todennäköisyysjakaumia

- $X \sim N(\mu, \sigma^2)$, $Z = (X - \mu)/\sigma \sim N(0, 1)$.
Jos $Z \sim N(0, 1)$, niin merkitään
 $F(z) = P(Z \leq z) = \Phi(z)$.

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.

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

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

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

- Studentin t-jakauma

$$P(t_{df} \geq t_{\alpha,df}) = \alpha, \quad P(t_{df} \geq t_{\alpha/2,df}) = \alpha/2$$

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

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

4) X_1, X_2, \dots, X_n on satunnaisotos, jos X_i :t ovat riippumattomia ja noudattavat samaa jakaumaa.

Sanonta

" X_1, X_2, \dots, X_n on satunnaisotos $N(\mu, \sigma^2)$:sta" tarkoittaa, että jokainen $X_i \sim N(\mu, \sigma^2)$ ja X_i :t ovat riippumattomia.

5) Otossuure, otosjakauma

Esim. Otossuure $\bar{X} \sim N(\mu, \sigma^2/n)$, jos satunnaisotos normaalijakaumasta.

6) Estimointi

- Estimointi on populaation tuntemattoman parametrin arviointia otossuureen avulla. Voidaan myös muodostaa väli (luottamusväli), jolla arvioidaan tuntemattoman parametrin olevan.
- Estimaattori otossuure, jolla estimoidaan tuntematonta parametria.
- Estimaatti on estimaattorin arvo.
- Harhaton estimaattori
- Estimaattorin keskivirhe (= estimaattorin keskihajonta)

7) Testaus

- Tilastollinen hypoteesi väite populaatiosta, usein populaation jakauman parametrissa.
- Hypoteesin testaus on väitteen tutkimista otoksen perusteella.
- Asetetaan nollahypoteesi (H_0) ja vaihtoehtoinen hypoteesi (H_1).
- Testisuure on otossuure, jota käytetään hypoteesin tutkimisessa.

- Testisuureen jakauma tunnetaan nollahypoteesin ollessa tosi.
- Otoksesta lasketun testisuureen arvon perusteella nollahypoteesi hyväksytään tai hylätään kiinnitetyllä riskitasolla.
- p-arvo on pienin riskitaso, jolla H_0 voidaan hylätä.

8) Joitain testausilanteita

- $H_0 : \pi = \pi_0$

Prosenttiosuuden tutkiminen Z-testillä,
kaava (5.3) MTTTP5

Esim. Ystäväsi väittää, että suomalaisista on 10% vasenkätisiä. Tutkit asiaa ja valitset satunnaisesti 400 suomalaista, joista vasenkätisiä on 47. Uskotko ystäväsi väitteen?

$$H_0 : \pi = 10$$

$$H_1 : \pi > 10$$

Jos H_0 tosi,

$$Z = \frac{p - 10}{\sqrt{10(100 - 10)/n}} \sim N(0, 1), \text{ likimain.}$$

Otoksesta laskettu z :n arvo on

$$z = \frac{11,75 - 10}{\sqrt{10(100 - 10)/400}} = 1,17$$

Pienin riskitaso, jolla H_0 voidaan hylätä yksisuuntaisessa testissä, on $P(Z > 1,17) = 1 - \Phi(1,17) = 1 - 0,8790 = 0,121$. Uskotaan siis ystävän väite.

Jos valitaan 5 %:n riskitaso, niin yksisuuntaisessa testissä kriittinen arvo on $z_{0,05} = 1,64$ (koska $\Phi(1,64) = 0,9495$) ja kaksisuuntaisessa testissä $z_{0,05/2} = 1,96$ (koska $\Phi(1,96) = 0,975$).

- $H_0 : \mu_1 = \mu_2$

Riippumattomien otosten t-testi odotusarvojen yhtäsuuruuden testaamiseksi,
kaava (5.5) MTTTP5

$$t = \frac{\bar{X} - \bar{Y}}{s\sqrt{1/n + 1/m}} \sim t(n + m - 2), \quad \text{kun } H_0 \text{ tosi,}$$

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

Esim 1.4.3. 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 hypv. | -2,387 | 9,923 |

$H_0: \sigma_N^2 = \sigma_K^2$ (Levene's Test)
 $H_0: \mu_N = \mu_K$ (t-test)

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 |

$$H_1: \mu_N < \mu_K$$

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

Luentorungon luvussa 1

<http://www.sis.uta.fi/tilasto/mttta1/kevat2019/luentorunko.pdf#page=3> lyhyt kertaus olennaisimmista asioista, jotka oletetaan opintojaksolla tunnetuiksi aiempien opintojen perusteella.

Tarvittaessa kertaukseen ja tietojensa täydentämiseen voi käyttää kurssien

- MTTTP1

(<https://coursepages.uta.fi/mtttp1/syksy-2018/>)

- MTTTP5

(<https://coursepages.uta.fi/mtttp5/syksy-2018/>)

materiaaleja.