

Mat-1.2620 Sovellettu todennäköisyyslaskenta B**7. harjoitukset / Ratkaisut****Aiheet: Mitta-asteikot****Havaintoaineiston kuvaaminen ja otostunnusluvut****Avainsanat:**

Aritmeettinen keskiarvo, Frekvenssi, Frekvenssijakauma, Geometrinen keskiarvo, Harmoninen keskiarvo, Histogrammi, Intervalliasteikko, Järjestysasteikko, Järjestystunnusluvut, Keskiarvo, Keskihajonta, Kvalitatiiviset muuttujat, Kvantitatiiviset muuttujat, Laatueroasteikko, Luokiteltu frekvenssijakauma, Maksimi, Mediaani, Minimi, Mittaaminen, Mitta-asteikot, Mittarit, Nominaaliasteikko, Ordinaaliasteikko, Otos, Otoskeskihajonta, Otosvarianssi, Perusjoukko, Pylväsdiagrammi, Suhdeasteikko, Suhteellinen frekvenssi, t -jakauma Tilastollinen aineisto, Tilastollinen muuttuja, Vaihteluväli, Vaihteluvälin pituus, Välimatka-asteikko

Tilastollisten aineistojen kerääminen ja mittaaminen**Tilastolliset aineistot**

Tilastollisen tutkimuksen *kaikki mahdolliset kohteet* muodostavat tutkimuksen (*kohde-*) **perusjoukon**. Tutkimuksen kohteita tarkastellaan aina jonkin *perusjoukon muodostamassa kehikossa*.

Tutkimuksen kohteiksi valittuja *perusjoukon alkioita* kutsutaan **havaintoyksiköiksi**.

Tilastollinen aineisto koostuu havaintoyksiköiden *ominaisuuksia* ja *olosuhteita* kuvaavista *numeerisista* tai *kvantitatiivisista tiedoista*.

Havaintoyksiköitä koskevia numeerisia tai kvantitatiivisia tietoja kutsutaan **havaintoarvoiksi** tai **havainnoiksi**.

Tilastollisten aineistojen kerääminen

Muutetaanko tutkimuksessa tutkimuksen kohteiden olosuhteita *aktiivisesti*?

- (i) Tutkimus on **koe**, jos tutkimuksen tavoitteena on selvittää, miten kohteiden olosuhteiden *aktiivinen muuttaminen* vaikuttaa tutkimuksen kohteisiin.
- (ii) Tutkimus perustuu **suoriin havaintoihin**, jos tutkimuksen tavoitteena on *vain seurata*, miten kohteiden *olosuhteet* ja *niissä tapahtuvat muutokset* vaikuttavat kohteisiin.

Kohdistuuko tutkimus *kaikkiin* perusjoukon alkioihin vai johonkin perusjoukon *osaan*?

- (i) Tutkimusta kutsutaan **kokonaistutkimukseksi**, jos *kaikki perusjoukon alkiot tutkitaan*.
- (ii) Tutkimusta kutsutaan **otantatutkimukseksi**, jos *tutkimus kohdistuu johonkin perusjoukon osajoukkoon*.

Mittaaminen ja mittarit

Tilastollisen tutkimuksen *kohteiden ominaisuuksia* ja *olosuhteita* sekä niiden muutoksia kuvaavat *numeeriset* tai *kvantitatiiviset tiedot* saadaan selville *mittaamalla*. **Mittaaminen** tarkoittaa *numeeristen arvojen liittämistä* tutkimuksen kohteiden ominaisuuksiin ja olosuhteisiin.

Mittaria voidaan pitää *funktiona*, joka *liittää numeeriset arvot* tutkimuksen kohteiden ominaisuuksiin ja olosuhteisiin.

Mittauksen *tulos* voidaan aina ilmaista jonkin tutkimuksen kohteen ominaisuutta tai olosuhdetta kuvaavan *muuttujan arvona*. Siten tutkimuksen kohteiden ominaisuuksia ja olosuhteita kuvataan mittaustapahtumassa aina *numeerisilla muuttujilla*.

Mittarin validiteetti ja tarkkuus

Mittari on **validi** eli *oikea*, jos se esittää mittauksen kohteena olevaa ominaisuutta *oikein, merkityksellisesti ja tarkoituksenmukaisesti*.

Mittari on **tarkka**, jos se on *harhaton* ja *reliaabeli*:

- (i) Mittari on **harhaton**, jos se *ei systemaattisesti ali- tai yliarvioi* mitattavan ominaisuuden määrää.
- (ii) Mittari on **reliaabeli** eli *luotettava*, jos mittaustulos *ei muutu*, kun mittausta toistetaan.

Mitta-asteikot

Mittaus on tehty **nominaali-** eli **laatueroasteikolla**, jos mittaus kertoo *mihin luokkaan* mittauksen kohde kuuluu.

Mittaus on tehty **ordinaali-** eli **järjestysasteikolla**, jos mittaus kertoo onko mittauksen kohteella mitattavaa ominaisuutta *enemmän* tai *vähemmän* kuin jollakin toisella kohteella.

Mittaus on tehty **intervalli-** eli **välimatka-asteikolla**, jos mittaus kertoo *kuinka paljon* kahden mitattavan kohteen ominaisuudet *eroavat* toisistaan.

Mittaus on tehty **suhdeasteikolla**, jos mittaus kertoo *kuinka monta kertaa enemmän* tai *vähemmän* mittauksen kohteella on mitattavaa ominaisuutta kuin jollakin toisella kohteella.

Kvalitatiiviset ja kvantitatiiviset muuttujat

Ominaisuutta ja sitä kuvaavaa muuttujaa kutsutaan **kvalitatiiviseksi**, jos mittauksen kohteet voidaan *luokitella* mittauksen perusteella toisistaan eroaviin *kategorioihin* tai *luokkiin*. Kvalitatiivisia ominaisuuksia kuvataan *laatueroasteikollisilla muuttujilla*.

Ominaisuutta ja sitä kuvaavaa muuttujaa kutsutaan **kvantitatiiviseksi**, jos mittaus tuottaa ominaisuuden *määrällisen arvon*. Kvantitatiivisia ominaisuuksia kuvataan *välimatka-* tai *suhdeasteikollisilla muuttujilla*.

Diskreetit ja jatkuvat muuttujat

Mitattavaa ominaisuutta vastaava muuttuja on **diskreetti**, jos se voi saada vain *erillisiä arvoja*. Diskreettejä muuttujia ovat esimerkiksi kaikki laatueroasteikollisten ja järjestysasteikollisten muuttujien lisäksi myös sellaiset kvantitatiiviset muuttujat kuten lukumäärämuuttujat.

Mitattavaa ominaisuutta vastaava muuttuja on **jatkuva**, jos se voi saada *kaikki arvot joltakin väliltä*. Jatkuvia muuttujia ovat esimerkiksi useimmat fysikaaliset suureet kuten pituus, pinta-ala, tilavuus, paino, aika, nopeus ja paine sekä myös monet talouselämää kuvaavat suureet kuten rahamäärä ja korko.

Huomautus:

Muuttujien mitta-asteikollisilla ominaisuuksilla (kvalitatiivisuudella/kvantitatiivisuudella tai diskreettiydellä/jatkuvuudella) on syväallinen vaikutus siihen, mitä tilastollisia menetelmiä kyseisessä tilanteessa on luvallista soveltaa.

Tilastollisten aineistojen kuvaaminen

Frekvenssi

Olkoon muuttuja x *diskreetti* ja oletetaan, että sen *mahdolliset arvot* ovat

$$y_1, y_2, \dots, y_m$$

Olkoot

$$x_1, x_2, \dots, x_n$$

muuttujan x *havaitut arvot*. Muuttujan x mahdollisen arvon y_k , $k = 1, 2, \dots, m$ **frekvenssi**

$$f_k$$

kertoo *kuinka monta kertaa* y_k esiintyy havaintoarvojen $x_1, x_2, \dots, x_n$ joukossa.

Frekvenssijakauma

Muuttujan x *mahdolliset arvot*

$$y_1, y_2, \dots, y_m$$

yhdessä niiden *frekvenssien*

$$f_1, f_2, \dots, f_m$$

kanssa muodostavat muuttujan x *havaittujen arvojen* $x_1, x_2, \dots, x_n$ **frekvenssijakauman**.

Huomaa, että

$$f_1 + f_2 + \dots + f_m = n$$

jossa n on havaintojen kokonaislukumäärä.

Pylväsdiagrammi

Frekvenssijakaumaa

$$(y_k, f_k), k = 1, 2, \dots, m$$

voidaan kuvata graafisesti **pylväsdiagrammilla**, jossa muuttujan x mahdollisen arvon y_k havaintoarvojen $x_1, x_2, \dots, x_n$ joukossa esittää *pylväs*, jonka *korkeus vastaa frekvenssiä* f_k .

Huomautus:

Pylväsdiagrammin tulkinta on analoginen *diskreetin todennäköisyysjakauman pistetodennäköisyysfunktion* tulkinnan kanssa.

Luokkafrekvenssi

Olkoon muuttuja x *jatkuva* ja oletetaan, että sen *mahdolliset arvot* ovat *välillä*

$$(a, b)$$

jossa voi olla $a = -\infty$, $b = +\infty$. Jaetaan väli (a, b) pisteillä

$$a = a_0 < a_1 < a_2 < \dots < a_{m-1} < a_m = b$$

pistevieraisiin *osaväleihin*

$$(a_{k-1}, a_k], k = 1, 2, \dots, m$$

Olkoot

$$x_1, x_2, \dots, x_n$$

muuttujan x havaitut arvot. Muuttujan x havaittujen arvojen **frekvenssi**

$$f_k$$

luokassa k kertoo niiden havaintoarvojen $x_1, x_2, \dots, x_n$ lukumäärän, jotka kuuluvat väliin

$$(a_{k-1}, a_k], k = 1, 2, \dots, m$$

Luokiteltu frekvenssijakauma

Luokkavälit

$$(a_{k-1}, a_k], k = 1, 2, \dots, m$$

yhdessä vastaavien *luokkafrekvenssien*

$$f_1, f_2, \dots, f_m$$

kanssa muodostavat muuttujan x havaittujen arvojen $x_1, x_2, \dots, x_n$ **luokitellun frekvenssijakauman**.

Huomaa, että

$$f_1 + f_2 + \dots + f_m = n$$

jossa n on havaintojen kokonaislukumäärä.

Histogrammi

Luokiteltua frekvenssijakaumaa

$$((a_{k-1}, a_k], f_k), k = 1, 2, \dots, m$$

voidaan kuvata graafisesti **histogrammilla**, jossa muuttujan x havaittujen arvojen

$$x_1, x_2, \dots, x_n$$

frekvenssiä f_k luokassa $(a_{k-1}, a_k]$, esittää *suorakaide*, jonka *kantana* on väli

$$(a_{k-1}, a_k]$$

ja jonka *pinta-ala* vastaa luokkafrekvenssiä f_k .

Huomautus:

Histogrammin tulkinta on analoginen *jatkuvan todennäköisyysjakauman tiheysfunktion* tulkinnan kanssa.

Suhdeasteikollisten muuttujien tunnusluvut

Aritmeettinen keskiarvo

Olko

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen* muuttujan x havaittuja arvoja. Lukujen $x_1, x_2, \dots, x_n$

aritmeettinen keskiarvo saadaan kaavalla

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Aritmeettinen keskiarvo on havaintoarvojen *painopiste* ja kuvaa havaintoarvojen *keskimääräistä* arvoa.

Varianssi

Olko

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen* muuttujan x havaittuja arvoja. Lukujen $x_1, x_2, \dots, x_n$ (*otos-*)

varianssi saadaan kaavalla

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \end{aligned}$$

jossa

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

on lukujen $x_1, x_2, \dots, x_n$ *aritmeettinen keskiarvo*.

Otosvarianssi kuvaa havaintoarvojen *hajaantuneisuutta* (tai *keskittyneisyyttä*) niiden aritmeettisen keskiarvon (painopisteen) ympärillä.

Aritmeettisen keskiarvon ja varianssi laskeminen

Olko

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen* muuttujan x havaittuja arvoja. Jos havaintoarvojen $x_1, x_2, \dots, x_n$ aritmeettinen keskiarvo ja varianssi joudutaan laskemaan *käsin* tai *laskinta* käyttäen, kannattaa laskut järjestää alla olevan taulukon muotoon ja käyttää niiden vieressä esitettyjä kaavoja.

i	x_i	x_i^2
1	x_1	x_1^2
2	x_2	x_2^2
$\vdots$	$\vdots$	$\vdots$
n	x_n	x_n^2
Summa	$\sum_{i=1}^n x_i$	$\sum_{i=1}^n x_i^2$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s_x^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right)$$

Keskihajonta

Olkoot

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen* muuttujan x havaittuja arvoja. Lukujen $x_1, x_2, \dots, x_n$ (*otos-*) **keskihajonta** on

$$\begin{aligned} s_x &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right)} \\ &= \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)} \\ &= \sqrt{s_x^2} \end{aligned}$$

jossa

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

on lukujen $x_1, x_2, \dots, x_n$ *aritmeettinen keskiarvo* ja s_x^2 on lukujen $x_1, x_2, \dots, x_n$ (*otos-*) *varianssi*.

Otoskeskihajonta kuvaa (kuten otosvarianssi) havaintoarvojen *hajaantuneisuutta* (tai *keskittyneisyyttä*) niiden aritmeettisen keskiarvon (painopisteen) ympärillä.

Standardointi

Olko $\bar{x}$ *välimatka-* tai *suhdeasteikollisen muuttujan* x havaittujen arvojen $x_1, x_2, \dots, x_n$ *aritmeettinen keskiarvo* ja s_x^2 niiden *varianssi*. Tällöin **standardoitujen havaintoarvojen**

$$z_i = \frac{x_i - \bar{x}}{s_x}, i = 1, 2, \dots, n$$

aritmeettinen keskiarvo ja varianssi ovat

$$\begin{aligned} \bar{z} &= \frac{1}{n} \sum_{i=1}^n z_i = 0 \\ s_z^2 &= \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2 = 1 \end{aligned}$$

Tilastollinen etäisyys

Olkoot $\bar{x}$ *välimatka-* tai *suhdeasteikollisen muuttujan* x havaittujen arvojen $x_1, x_2, \dots, x_n$ *aritmeettinen keskiarvo* ja s_x^2 niiden *varianssi*.

Tällöin havaintoarvojen x_k ja x_l *tilastollinen etäisyys* on

$$d_{kl} = \frac{x_k - x_l}{s_x}$$

Origomomentit

Olkoot

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen muuttujan* x havaittuja arvoja. Lukujen $x_1, x_2, \dots, x_n$

k. origomomentti on

$$a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, k = 1, 2, \dots$$

Keskusmomentit

Olkoot

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen muuttujan* x havaittuja arvoja. Lukujen $x_1, x_2, \dots, x_n$

k. keskusmomentti on

$$m_x^k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

jossa

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

on lukujen $x_1, x_2, \dots, x_n$ *aritmeettinen keskiarvo*.

Vinous

Olkoot

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen muuttujan* x havaittuja arvoja. Havaintoarvojen $x_1, x_2, \dots, x_n$ jakauman **vinoutta** voidaan kuvata otostunnusluvulla

$$c_1 = \frac{m_3}{m_2^{3/2}}$$

jossa

$$m_2 = 2. \text{ keskusmomentti luvuille } x_1, x_2, \dots, x_n$$

$$m_3 = 3. \text{ keskusmomentti luvuille } x_1, x_2, \dots, x_n$$

Huipukkuus

Olkoot

$$x_1, x_2, \dots, x_n$$

välimatka- tai *suhdeasteikollisen* muuttujan x havaittuja arvoja. Havaintoarvojen $x_1, x_2, \dots, x_n$ jakauman **huipukkuutta** voidaan kuvata otostunnusluvulla

$$c_2 = \frac{m_4}{m_2^2}$$

jossa

$m_2 = 2$. *keskusmomentti* luvuille $x_1, x_2, \dots, x_n$

$m_4 = 4$. *keskusmomentti* luvuille $x_1, x_2, \dots, x_n$

Geometrinen keskiarvo

Olkoot

$$x_1, x_2, \dots, x_n$$

positiivisia lukuja. Lukujen $x_1, x_2, \dots, x_n$ **geometrinen keskiarvo** on

$$G = \sqrt[n]{x_1 x_2 \cdots x_n}$$

Lukujen $x_1, x_2, \dots, x_n$ geometrisen keskiarvon logaritmi on lukujen $x_1, x_2, \dots, x_n$ logaritmien aritmeettinen keskiarvo:

$$\log(G) = \frac{\log(x_1) + \log(x_2) + \cdots + \log(x_n)}{n} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$$

Huomaa, että

$$G = \bar{x}$$

vain, jos

$$x_1 = x_2 = \cdots = x_n$$

Harmoninen keskiarvo

Olkoot

$$x_1, x_2, \dots, x_n$$

positiivisia lukuja. Lukujen $x_1, x_2, \dots, x_n$ **harmoninen keskiarvo** on

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}}$$

Lukujen $x_1, x_2, \dots, x_n$ harmonisen keskiarvon käänteisluku on lukujen $x_1, x_2, \dots, x_n$ käänteislukujen aritmeettinen keskiarvo:

$$\frac{1}{H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}$$

Huomaa, että

$$H = \bar{x}$$

vain, jos

$$x_1 = x_2 = \dots = x_n$$

Järjestysasteikollisten muuttujien tunnusluvut

Järjestystunnusluvut

Olkoot

$$x_1, x_2, \dots, x_n$$

järjestys-, välimatka- tai suhdeasteikollisen muuttujan x havaittuja arvoja. Järjestetään havaintoarvot $x_1, x_2, \dots, x_n$ suurusjärjestykseen pienimmästä suurimpaan ja olkoot

$$z_1, z_2, \dots, z_n$$

*järjestykseen asetetut havaintoarvot. Suurusjärjestyksessä k . havaintoarvoa z_k kutsutaan **k . järjestystunnusluvuksi.***

Minimi, maksimi, vaihteluväli

Olkoot

$$z_1, z_2, \dots, z_n$$

järjestys-, välimatka- tai suhdeasteikollisen muuttujan x havaitut arvot järjestettyinä suurusjärjestykseen pienimmästä suurimpaan.

Tällöin

$$z_1 = \text{minimiarvo}$$

$$z_n = \text{maksimiarvo}$$

$$(z_1, z_n) = \text{vaihteluväli}$$

$$z_n - z_1 = \text{vaihteluvälin pituus}$$

Prosenttipisteet

Olkoot

$$z_1, z_2, \dots, z_n$$

järjestys-, välimatka- tai suhdeasteikollisen muuttujan x havaitut arvot järjestettyinä suurusjärjestykseen pienimmästä suurimpaan.

Havaintoarvojen **p . prosenttipiste**

$$z_{(p)}, p = 1, 2, \dots, 99$$

on piste, joka jakaa havaintoaineiston kahteen osaan:

- (i) p % havaintoarvoista on lukua $z_{(p)}$ pienempiä tai korkeintaan yhtä suuria kuin $z_{(p)}$.
- (ii) $(100 - p)$ % havaintoarvoista on lukua $z_{(p)}$ suurempia.

Mediaani

Olkoot

$$z_1, z_2, \dots, z_n$$

järjestys-, välimatka- tai suhdeasteikollisen muuttujan x havaitut arvot järjestettyinä suuruusjärjestykseen pienimmästä suurimpaan.

Mediaani Me on havaintoarvojen 50. prosenttipiste:

$$Me = z_{(50)}$$

Mediaani jakaa havaintoaineiston *kahteen yhtä suureen osaan* niin, että toisessa *kaikki* havaintoarvot ovat mediaania *pienempiä*, toisessa *kaikki* havaintoarvot ovat mediaania *suurempia*.

Havaintoarvojen mediaani Me voidaan määrätä seuraavalla tavalla:

- (1) Järjestetään havaintoarvot *suuruusjärjestykseen* pienimmästä suurimpaan.
- (2a) Jos havaintoarvojen lukumäärä on *pariton*, mediaani on järjestetyistä havaintoarvoista *keskimmäinen*.
- (2b) Jos havaintoarvojen lukumäärä on *parillinen*, mediaani on järjestetyistä havaintoarvoista *kahden kesimmäisen aritmeettinen keskiarvo*.

Oletetaan, että *aritmeettinen keskiarvo M ja mediaani Me määrätään samasta jatkuvan muuttujan havaittujen arvojen luokitellusta frekvenssijakaumasta*. Jos havaintoarvojen jakauma on *yksihuippuinen*, pätee seuraava:

Vasemmalle vinoilla jakaumilla

$$M < Me$$

Symmetrisillä jakaumilla

$$M \approx Me$$

Oikealle vinoilla jakaumilla

$$Me < M$$

Kvartiilit

Olkoot

$$z_1, z_2, \dots, z_n$$

järjestys-, välimatka- tai suhdeasteikollisen muuttujan x havaitut arvot järjestettyinä suuruusjärjestykseen pienimmästä suurimpaan.

Tällöin

$$Q_1 = \text{Alakvartiili} = 25. \text{ prosenttipiste} = z_{(25)}$$

$$Q_2 = \text{Keskikvartiili} = 50. \text{ prosenttipiste} = z_{(50)}$$

$$Q_3 = \text{Yläkvartiili} = 75. \text{ prosenttipiste} = z_{(75)}$$

Kvartiilit Q_1, Q_2, Q_3 jakavat suuruusjärjestykseen asetetun havaintoaineiston *neljään yhtä suureen osaan*.

Erityisesti:

Alakvartiili Q_1 = Havaintoarvojen mediaania Me pienempien havaintoarvojen mediaani

Keskikvartiili Q_2 = Havaintoarvojen mediaani Me

Yläkvartiili Q_3 = Havaintoarvojen mediaania Me suurempien havaintoarvojen mediaani

Kvartiilit, kvartiiliväli, kvartiilipoikkeama

Olkoot havaintoarvojen kvartiilit Q_1, Q_2, Q_3 .

Tällöin

(Q_1, Q_3) = kvartiiliväli

$Q_3 - Q_1 = IQR$ = kvartiilivälin pituus

$(Q_3 - Q_1)/2 = IQR/2$ = kvartiilipoikkeama

Kvartiiliväliä, kvartiilivälin pituutta (IQR = interquartile range) ja kvartiilipoikkeamaa voidaan käyttää kuvaamaan havaintoarvojen *hajaantuneisuutta* (*keskittyneisyyttä*).

Jos havaintoarvojen jakaumaa kuvaavana *keskilukuna* on käytetty *mediaania*, *hajontalukuna* käytetään usein *kvartiilipoikkeamaa*.

Laatueroasteikollisten muuttujien tunnusluvut

Frekvenssi

Olkoon *otoskoko* eli kerättyjen havaintoarvojen lukumäärä n . Olkoon A jokin perusjoukon osajoukko ja olkoon f otokseen kuuluvien A -tyyppisten havaintoarvojen *frekvenssi* eli *lukumäärä*.

Tällöin A -tyyppisten havaintoarvojen **suhteellinen frekvenssi** eli **osuus** otoksessa on

$$\frac{f}{n}$$

Moodi

Frekvenssijakauman moodi eli *tyyppi-arvo* Mo on yleisin havaintoarvo.

Luokitellun frekvenssijakauman moodi eli *tyyppi-arvo* Mo on siinä luokassa, jossa luokiteltua frekvenssijakaumaa vastaava *histogrammi saavuttaa maksiminsa*.

Huomautuksia:

Jos käytetty luokitus on *tasavälinen*, luokitellun frekvenssijakauman moodi on siinä luokassa, jota vastaava frekvenssi on suurin.

Jos käytetty luokitus *ei ole tasavälinen*, luokitellun frekvenssijakauman moodi *ei välttämättä ole* siinä luokassa, jota vastaava frekvenssi on suurin.

Oletetaan, että aritmeettinen keskiarvo M , mediaani Me ja moodi Mo määrätään samasta jatkuvan muuttujan havaittujen arvojen luokitellusta frekvenssijakaumasta. Jos havaintoarvojen jakauma on yksihiippuinen, pätee seuraava:

Vasemmalle vinoilla jakaumilla

$$M < Me < Mo$$

Symmetrisillä jakaumilla

$$M \approx Me \approx Mo$$

Oikealle vinoilla jakaumilla

$$Mo < Me < M$$

Tehtävä 7.1.

Alla on lueteltu joukko tilastollisia muuttujia.

1. Mansikoiden C-vitamiinipitoisuus; yksikkö: mg/100 g
2. Alvarin aukiolta löydetyn kasvin laji
3. Paine, joka vaaditaan teräksisen säiliön murtumiseen; kg/cm^2
4. Henkilöiden reaktio väitteeseen
”Suomen on liityttävä NATO:on”
mitattuna asteikolla: täysin eri mieltä, yhden tekevää, täysin samaa mieltä
5. Jokereiden sijoitus jääkiekkoliigassa; asteikkona 1, 2, ...
6. Teekkarin koulutusohjelma
7. Teekkarin älykkyysosamäärä; yksikkö: äo-piste
8. Teekkarin pistemäärä kurssin 1. välikokeessa; asteikkona 0, 1, 2, ..., 30
9. Lentokoneen nopeus; yksikkö: km/h

- (a) Mitkä ovat muuttujien 1-9 mitta-asteikot?
- (b) Mitkä muuttujista 1-9 ovat kvalitatiivisia ja mitkä kvantitatiivisia?
- (c) Mitkä muuttujista 1-9 ovat diskreettejä ja mitkä jatkuvia?

Tehtävä 7.1. – Mitä opimme?

Tehtävässä tarkastellaan tilastollisten muuttujien *mitta-asteikollisia ominaisuuksia* sekä tilastollisten muuttujien luokittelua toisaalta *kvalitatiivisiin* ja *kvantitatiivisiin* muuttujiin ja toisaalta *diskreetteihin* ja *jatkuviin* muuttujiin.

Tehtävä 7.1. – Ratkaisu:

- (a) Laatueroasteikollisia muuttujia: 2, 6
Järjestysasteikollisia muuttujia: 4, 5, 7, 8
Suhdeasteikollisia muuttujia: 1, 3, 9
- (b) Kvalitatiivisia muuttujia: 2, (4), (5), 6
Kvantitatiivisia muuttujia: 1, 3, (4), (5), 7, 8, 9
Kvalitatiivisten ja kvantitatiivisten muuttujien välimaastossa olevat järjestysasteikolliset muuttujat on merkitty sulkuihin.
- (c) Diskreettejä muuttujia: 2, 4, 5, 6, 7, 8
Jatkuvia muuttujia: 1, 3, 9

Tehtävä 7.2.

Erään talon asukkailla on seuraavat kuukausitulot (€/kk):

20100	19400	10100	23000	24200	25100	8200	8900	10300
26000	11400	12900	13200	14300	15800	16100	17200	18900
5200	10100	12300	14000	15100	16000	11100	10800	9100
7200	4300	38000	51100	9600	10900	12000	13200	15100

Määää aineistosta seuraavat tunnusluvut:

- (a) minimi, maksimi
- (b) vaihteluväli, vaihteluvälin pituus
- (c) mediaani
- (d) kvartiiliväli, kvartiiliväli, kvartiilivälin pituus, kvartiilipoikkeama

Tehtävä 7.2. – Mitä opimme?

Tehtävässä tarkastellaan *järjestystunnuslukujen* määäämistä.

Tehtävä 7.2. – Ratkaisu:

Kaikki määrättäviksi pyydetty tunnusluvut ovat *järjestystunnuslukuja* tai niihin perustuvia tunnuslukuja.

Järjestetään havaintoarvot suuruusjärjestykseen pienimmästä suurimpaan järjestystunnuslukujen määäämistä varten:

4300	5200	7200	8200	8900	9100	9600	10100	10100
10300	10800	10900	11100	11400	12000	12300	12900	13200
13200	14000	14300	15100	15100	15800	16000	16100	17200
18900	19400	20100	23000	24200	25100	26000	38000	51100

- (a) *Minimi ja maksimi:*

$$Min = 4300, Max = 51100$$

- (b) *Vaihteluväli:*

$$(Min, Max) = (4300, 51100)$$

Vaihteluvälin pituus:

$$Max - Min = 51100 - 4300 = 46800$$

- (c) Etsitään havaintojen *mediaani* Me .

Mediaani Me jakaa havaintoaineiston *kahteen* yhtä suureen osaan siten, että puolet niistä havaintoarvoista, jotka eivät ole yhtä suuria kuin mediaani, ovat mediaania pienempiä, ja puolet niistä havaintoarvoista, jotka eivät ole yhtä suuria kuin mediaani, ovat mediaania suurempia.

Oletetaan, että n havaintoa on järjestetty suuruusjärjestykseen pienimmästä suurimpaan.

- (i) Jos n on pariton, niin mediaaniksi valitaan havaintoarvo, joka löytyy paikasta

$$(n + 1)/2$$

- (ii) Jos n on parillinen, mediaaniksi valitaan kahden keskimmäisen havainnon aritmeettinen keskiarvo.

Koska havaintojen lukumäärä on tässä parillinen, niin

$$Me = (13200 + 13200)/2 = 13200$$

- (d) Etsitään ensin havaintojen *kvartiilit* Q_1 , Q_2 , Q_3 .

Kvartiilit Q_1 , Q_2 , Q_3 jakavat suuruusjärjestykseen asetetun havaintoaineiston *neljään* yhtä suureen osaan. *Keskikvartiili* Q_2 on sama kuin *mediaani*. *Alakvartiili* Q_1 on mediaania *pienempien* havaintoarvojen mediaani ja *yläkvartiili* Q_3 on mediaania *suurempien* havaintoarvojen mediaani.

Siten

$$Q_2 = Me = 13200$$

$$Q_1 = (10100 + 10300)/2 = 10200$$

$$Q_3 = (17200 + 18900)/2 = 18050$$

Kvartiiliväli on

$$(Q_1, Q_3) = (10200, 18050)$$

Kvartiilivälin pituus on

$$IQR = Q_3 - Q_1 = 18050 - 10200 = 7850$$

Kvartiilipoikkeama on

$$IQR/2 = (Q_3 - Q_1)/2 = 7850/2 = 3925$$

Tehtävä 7.3.

Muodosta tehtävän 7.2. aineistosta luokiteltu frekvenssijakauma, jonka luokkaväleinä ovat

(4000,12000]

(1200, 28000]

(28000,60000]

Määrää myös frekvenssijakaumaa vastaavan histogrammikuvion suorakaiteiden korkeudet, kun luokkaväliä [4000,12000] vastaavan suorakaiteen korkeudeksi valitaan 15 yksikköä. Hahmottele myös ko. histogrammikuvi ruudulliselle paperille. Missä luokassa on jakauman moodi?

Tehtävä 7.3. – Mitä opimme?

Tehtävässä tarkastellaan *luokitellun frekvenssijakauman* ja sitä vastaavan graafisen esityksen eli *histogrammin* muodostamista.

Tehtävä 7.3. – Ratkaisu:

Histogrammikuvi muodostuu suorakaiteista, joiden *pinta-alat* suhtautuvat toisiinsa kuten vastaavat luokkafrekvenssit (tai suhteelliset luokkafrekvenssit).

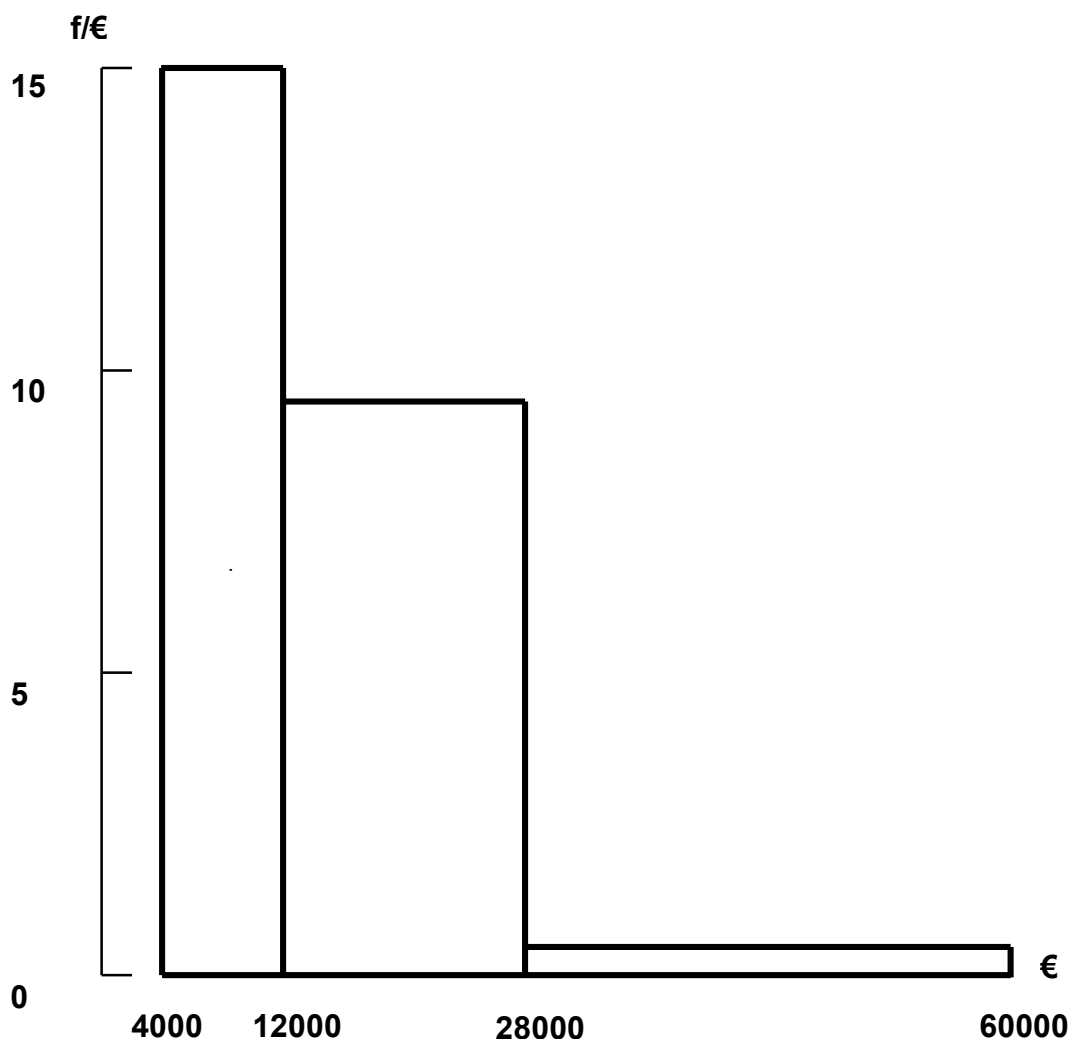
Tehtävän 7.2. aineistosta saadaan seuraava luokiteltu frekvenssijakauma, kun luokkaväleinä ovat (4000,12000] , (1200, 28000] , (28000,60000] :

Luokkaväli	Luokkafrekvenssi	Suorakaiteen korkeus (yksikköä)
(4000,12000]	15	15
(12000,28000]	19	$19/2 = 9.5$
(28000,60000]	2	$2/4 = 0.5$

Histogrammikuvion suorakaiteiden korkeuksien määrittäminen:

- (1) Valitaan luokkaväliin (4000,12000] liittyvän suorakaiteen korkeudeksi 15 yksikköä.
- (2) Luokkaväli (12000,28000] on kaksi kertaa pitempi kuin luokkaväli (4000,12000]. Siksi luokkaväliin (12000,28000] liittyvän suorakaiteen korkeus saadaan jakamalla luokkaväliä vastaava frekvenssi 19 luvulla 2.
- (3) Luokkaväli (28000,60000] on neljä kertaa pitempi kuin luokkaväli (4000,12000]. Siksi luokkaväliin (18000,60000] liittyvän suorakaiteen korkeus saadaan jakamalla luokkaväliä vastaava frekvenssi 2 luvulla 4.

Alla oleva kuvio esittää yo. luokiteltua frekvenssijakaumaa vastaavaa histogrammia.



Jakauman *moodi* on luokassa (4000,12000], koska siinä histogrammi saavuttaa maksiminsa. Huomaa, että moodi *ei ole* luokassa (12000,28000], vaikka sitä vastaava frekvenssi on suurin.

Huomautuksia:

- (i) Histogrammissa suorakaiteiden pinta-alat – eivät siis korkeudet – ovat suhteessa luokkafrekvensseihin.
- (ii) Histogrammissa suorakaiteiden korkeudet ovat suhteessa luokkafrekvensseihin vain, jos luokitus on tasavälinen.
- (iii) Oikea laatu pystyakselille on tehtävän 7.3. tapauksessa frekvenssi/€ :
 - Vaaka-akselin laatu: €
 - Pystyakselin laatu: frekvenssi/€
 - Suorakaiteen pinta-ala: € × frekvenssi/€ = frekvenssi

Tehtävä 7.4.

Määrä tehtävän 7.2. aineiston kahden ensimmäisen sarakkeen 8:sta luvusta aritmeettinen keskiarvo, otosvarianssi ja otoskeskihajonta.

Tehtävä 7.4. – Mitä opimme?

Tehtävässä tarkastellaan *aritmeettisen keskiarvon, otosvarianssin ja otoskeskihajonnan* määrittämistä.

Tehtävä 7.4. – Ratkaisu:

Laskutoimitukset voidaan suorittaa kahdella tavalla.

Tapa 1:

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ s_x &= \sqrt{s_x^2}\end{aligned}$$

Tapa 2:

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ s_x^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) \\ s_x &= \sqrt{s_x^2}\end{aligned}$$

Jos havaintoarvojen aritmeettisen keskiarvon ja varianssi laskemista varten laaditaan tietokoneohjelma, laskutoimitukset voidaan järjestää laskutavassa 2 niin, että havainnot käydään läpi vain kerran, kun taas laskutavassa 1 havainnot on käytävä läpi kaksi kertaa. Sen sijaan laskutavan 1 kaavat ovat numeerisesti vakaampia kuin laskutavassa 2.

Alla on kopio laskutoimitusten tekemisessä apuna käytetty Microsoft Excel –taulukosta.

Taulukosta saadaan:

$$\sum_{i=1}^8 x_i = 103700$$

$$\bar{x} = \frac{1}{8} \sum_{i=1}^8 x_i = 12962.5$$

$$\sum_{i=1}^8 x_i^2 = 1785710000$$

$$s_x^2 = \frac{1}{8-1} \left(\sum_{i=1}^8 x_i^2 - \frac{1}{8} \left(\sum_{i=1}^8 x_i \right)^2 \right) = 63071250$$

$$\sum_{i=1}^8 (x_i - \bar{x})^2 = 441498750$$

$$s_x^2 = \frac{1}{8-1} \sum_{i=1}^8 (x_i - \bar{x})^2 = 63071250$$

$$s_x = \sqrt{s_x^2} = 7941.741$$

	Palkka			
i	x	x-Ka	(x-Ka)^2	x^2
1	20100	7137.5	50943906.25	404010000
2	26000	13037.5	169976406.3	676000000
3	5200	-7762.5	60256406.25	27040000
4	7200	-5762.5	33206406.25	51840000
5	19400	6437.5	41441406.25	376360000
6	11400	-1562.5	2441406.25	129960000
7	10100	-2862.5	8193906.25	102010000
8	4300	-8662.5	75038906.25	18490000
Summa	103700	0	441498750	1785710000

$$Ka = 12962.5$$

Tapa 1: $Var = 63071250$

$Hajonta = 7941.741$

Tapa 2: $Var = 63071250$

$Hajonta = 7941.741$

Tehtävä 7.5.

Olkoon

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

lukujen $x_1, x_2, \dots, x_n$ aritmeettinen keskiarvo.

(a) Todista, että

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

(b) Todista, että $\bar{x}$ minimoi neliösumman

$$\sum_{i=1}^n (x_i - a)^2$$

parametrin a suhteen.

Tehtävä 7.5. – Mitä opimme?

Tehtävässä tutkitaan *aritmeettisen keskiarvon* karakteristisia ominaisuuksia.

Oikea keskiluku tehtävän ongelman ratkaisemiseen on tässä *geometrinen keskiarvo*.

Tehtävä 7.5. – Ratkaisu:

Olkoon

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

lukujen $x_1, x_2, \dots, x_n$ aritmeettinen keskiarvo.

(a) Tällöin

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}) &= \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} \\ &= \sum_{i=1}^n x_i - n\bar{x} \\ &= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i \\ &= 0 \end{aligned}$$

(b) Todistus 1:

$$\begin{aligned}
 & \sum_{i=1}^n (x_i - a)^2 \\
 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - a)^2 \\
 &= \sum_{i=1}^n [(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - a) + (\bar{x} - a)^2] \\
 &= \sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - a) \sum_{i=1}^n (x_i - \bar{x}) + \sum_{i=1}^n (\bar{x} - a)^2 \quad \left| \sum_{i=1}^n (x_i - \bar{x}) = 0 \right. \\
 & \hspace{15em} \text{(a)-kohdan mukaan} \\
 &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2 \\
 &\geq \sum_{i=1}^n (x_i - \bar{x})^2
 \end{aligned}$$

ja lisäksi alaraja

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

saavutetaan, kun

$$a = \bar{x}$$

Todistus 2:

Etsitään funktion

$$f(a) = \sum_{i=1}^n (x_i - a)^2$$

ääriarvot parametrin a suhteen derivoimalla funktiota $f(a)$.

Derivoidaan funktio $f(a)$, merkitään derivaatta nolaksi ja ratkaistaan saatu normaaliyhtälö parametrin a suhteen:

$$f'(a) = \frac{\partial}{\partial a} \sum_{i=1}^n (x_i - a)^2 = -2 \sum_{i=1}^n (x_i - a) = -2 \sum_{i=1}^n x_i + 2 \sum_{i=1}^n a = -2 \sum_{i=1}^n x_i + 2na = 0$$

Ratkaisuksi saadaan

$$a = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Ratkaisu vastaa funktion $f(a)$ minimiä, koska

$$f''(a) = \frac{\partial^2}{\partial a^2} \sum_{i=1}^n (x_i - a)^2 = 2n > 0$$

Lisäksi funktion $f(a)$ minimiarvoksi saadaan

$$f(\bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2 = (n-1)s^2$$

jossa s^2 on lukujen $x_1, x_2, \dots, x_n$ otosvarianssi.

Tehtävä 7.6.

Heitetään noppaa 30 kertaa. Olkoon tuloksena seuraava silmälukujen jono:

2, 4, 4, 1, 5, 4, 1, 1, 6, 5, 3, 3, 5, 4, 4, 1, 6, 1, 5, 4, 4, 3, 3, 3, 5, 5, 3, 4, 5, 4

- (a) Määrää silmälukujen frekvenssit ja suhteelliset frekvenssit.
 (b) Määrää silmälukujen frekvenssijakauman moodi.

Tehtävä 7.6. – Mitä opimme?

Tehtävässä tarkastellaan *laatueroasteikollisten muuttujien tunnuslukujen määräämistä*.

Tehtävä 7.6. – Ratkaisu:

Kaikki määrättäviksi pyydetty tunnusluvut ovat *laatueroasteikollisten muuttujien tunnuslukuja* tai niihin perustuvia tunnuslukuja.

Olemme heittäneet noppaa 30 kertaa ja tuloksena on seuraava silmälukujen jono:

2, 4, 4, 1, 5, 4, 1, 1, 6, 5, 3, 3, 5, 4, 4, 1, 6, 1, 5, 4, 4, 3, 3, 3, 5, 5, 3, 4, 5, 4

- (a) Määrätään silmälukujen *frekvenssit* f ja *suhteelliset frekvenssit* f/n . Tehtävää helpottaa, jos järjestämme jonon luvut suuruusjärjestykseen. Saamme tällöin jonon

1, 1, 1, 1, 1, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6

Siten silmälukujen frekvenssien ja suhteellisten frekvenssien jakaumat voidaan esittää alla olevan taulukon muodossa:

Silmäluku	1	2	3	4	5	6	Summa
Frekvenssi f	5	1	6	9	7	2	30
Suhteellinen frekvenssi f/n	5/30	1/30	6/30	9/30	7/30	2/30	1

- (b) Kohdan (a) frekvenssitaulukosta voimme lukea, että frekvenssijakauman *moodi* on

$$Mo = 4$$

koska silmäluvun 4 frekvenssi on *suurin*.

Tehtävä 7.7.

Olet ottanut pankista 10000 euron lainan, jota ei saa lyhentää kahden ensimmäisen vuoden aikana. Alkuperäisen sopimuksen mukainen korko on 1. vuotena 10 % ja 2. vuotena 20 %, jolloin takaisin maksettava lainapääoma kasvaa kahdessa vuodessa x %.

Oletetaan, että pankin vaatimuksesta sopimusta muutetaan niin, että kahden ensimmäisen vuoden aikana käytetään *samaa* korkoprosenttia, joka määrätään niin, että lainapääoma kasvaa tänä aikana samaksi kuin alkuperäisen sopimuksen mukaan.

- (a) Määrittää x .
 (b) Näytä, että uuden sopimuksen korkoa *ei saada* kaavalla

$$(10 + 20)/2 \%$$

- (c) Näytä, että uuden sopimuksen korko *saadaan* kaavalla

$$(\sqrt{1.1 \times 1.2} - 1) \times 100 \%$$

jossa

$$\sqrt{1.1 \times 1.2}$$

on lukujen 1.1 ja 1.2 geometrinen keskiarvo.

Tehtävä 7.7. – Mitä opimme?

Tehtävässä näytetään, että *aritmeettinen keskiarvo* ei ole aina käypä tunnusluku.

Oikea keskiluku tehtävän ongelman ratkaisemiseen on tässä *geometrinen keskiarvo*.

Tehtävä 7.7. – Ratkaisu:

- (a) Olkoon korko 1. vuotena 10 % ja 2. vuotena 20 %.

Lainapääoma 1. vuoden lopussa:

$$\begin{aligned} (1 + 10/100) \times 10000 \\ &= (1 + 0.1) \times 10000 \\ &= 1.1 \times 10000 \\ &= 11000 \end{aligned}$$

Lainapääoma 2. vuoden lopussa:

$$\begin{aligned} (1 + 20/100) \times 11000 \\ &= (1 + 0.2) \times 11000 \\ &= 1.2 \times 11000 \\ &= 13200 \end{aligned}$$

Siten lainapääoma kasvaa kahdessa vuodessa

$$100 \times (13200 - 10000) / 10000 \% = 32 \%$$

joten

$$x = 32$$

- (b) Määrätään 1. ja 2. vuoden korkoprosenttien aritmeettinen keskiarvo:

$$\frac{10 + 20}{2} \% = 15 \%$$

Olkoon korko siis molempina vuosina 15 %.

Lainapääoma 1. vuoden lopussa:

$$\begin{aligned}(1 + 15/100) \times 10000 \\ &= (1 + 0.15) \times 10000 \\ &= 1.15 \times 10000 \\ &= 11500\end{aligned}$$

Lainapääoma 2. vuoden lopussa:

$$\begin{aligned}(1 + 15/100) \times 11500 \\ &= (1 + 0.15) \times 11500 \\ &= 1.15 \times 11500 \\ &= 13225\end{aligned}$$

Siten lainapääoma kasvaa kahdessa vuodessa

$$100 \times (13225 - 10000) / 10000 \% = 32.25 \% > 32 \%$$

Huomaa, että oikea korkoprosentti ei ole myöskään

$$\frac{32}{2} \% = 16 \%$$

- (c) Määrätään korkoprosentti kaavalla

$$(\sqrt{1.1 \times 1.2} - 1) \times 100 \%$$

jossa

$$\sqrt{1.1 \times 1.2}$$

on lukujen 1.1 ja 1.2 geometrinen keskiarvo:

$$(\sqrt{1.1 \times 1.2} - 1) \times 100 \approx 14.8913$$

Olkoon korko siis molempina vuosina 14.8913 %.

Lainapääoma 1. vuoden lopussa:

$$\begin{aligned}(1 + 14.8913/100) \times 10000 \\ &= (1 + 0.148913) \times 10000 \\ &= 1.148913 \times 10000 \\ &= 11489.13\end{aligned}$$

Lainapääoma 2. vuoden lopussa:

$$\begin{aligned}(1 + 14.8913/100) \times 11489.13 \\&= (1 + 0.148913) \times 11489.13 \\&= 1.148913 \times 11489.13 \\&\approx 13200\end{aligned}$$

Siten lainapääoma kasvaa kahdessa vuodessa

$$100 \times (13200 - 10000) / 10000 \% = 32 \%$$

kuten pitääkin.

Huomautus:

Olkoon korko 1. vuotena p % ja toisena vuotena q %.

Yleisesti pätee:

$$\left[1 + \left(\sqrt{\left(1 + \frac{p}{100} \right) \left(1 + \frac{q}{100} \right)} - 1 \right) \right]^2 = \left(1 + \frac{p}{100} \right) \left(1 + \frac{q}{100} \right)$$

mutta

$$\left(1 + \frac{(p+q)/2}{100} \right) \left(1 + \frac{(p+q)/2}{100} \right) \neq \left(1 + \frac{p}{100} \right) \left(1 + \frac{q}{100} \right)$$

paitsi, jos

$$p = q$$

Tehtävä 7.8.

Paikkakuntien A ja B välimatka on 120 km. Henkilö ajaa A:sta B:hen keskinopeudella 60 km/h ja B:stä A:han keskinopeudella 120 km/h.

- (a) Määrää keskinopeus edestakaisella matkalla.
 (b) Näytä, että keskinopeutta edestakaisella matkalla *ei saada* kaavalla

$$(60 + 120)/2 = 90 \text{ km/h}$$

- (c) Näytä, että oikea keskinopeus *saadaan* määräämällä lukujen 60 ja 120 harmoninen keskiarvo

$$\frac{1}{\frac{1}{2} \left(\frac{1}{60} + \frac{1}{120} \right)}$$

Tehtävä 7.8. – Mitä opimme?

Tehtävässä näytetään, että *aritmeettinen keskiarvo* ei ole aina käypä tunnusluku.

Oikea keskiluku tehtävän ongelman ratkaisemiseen on tässä *harmoninen keskiarvo*.

Tehtävä 7.8. – Ratkaisu:

- (a) A:n ja B:n välimatka:

$$120 \text{ km}$$

Ajoaika A:sta B:hen (60 km/h):

$$120/60 = 2 \text{ h}$$

Ajoaika B:stä A:han (120 km/h):

$$120/120 = 1 \text{ h}$$

Matka edestakaisin:

$$240 \text{ km}$$

Ajoaika edestakaisin:

$$2 + 1 = 3 \text{ h}$$

Keskinopeus edestakaisella matkalla:

$$240/3 = 80 \text{ km/h}$$

- (a) Määrätään keskinopeuksien *aritmeettinen keskiarvo*:

$$\frac{1}{2} (60 + 120) \text{ km/h} = 90 \text{ km/h} \neq 80 \text{ km/h}$$

- (b) Määrätään keskinopeuksien *harmoninen keskiarvo*:

$$\frac{1}{\frac{1}{2} \left(\frac{1}{60} + \frac{1}{120} \right)} \text{ km/h} = 80 \text{ km/h}$$

Tehtävä 7.7.

- (a) Kone valmistaa kuulalaakerin kuulia, joiden halkaisijat vaihtelevat satunnaisesti noudattaen normaalijakaumaa parametrein

$$\mu = 10 \text{ mm}, \sigma = 0.01 \text{ mm}$$

Poimitaan kuulien joukosta yksinkertainen satunnaisotos, jonka koko $n = 10$.

Olkoot $\bar{X}$ ja s^2 kuulien halkaisijoiden aritmeettinen keskiarvo ja otosvarianssi otoksessa.

Mitkä ovat aritmeettisen keskiarvon $\bar{X}$ ja otosvarianssin s^2 muunnoksen $(n-1)s^2/\sigma^2$ jakaumat otoksessa?

- (b) Äänestäjistä 25 % kannattaa puoluetta ABC. Poimitaan äänestäjien joukosta yksinkertainen satunnaisotos, jonka koko $n = 1000$.

Mikä on puolueen ABC kannattajien suhteellisen osuuden f/n approksimatiivinen jakauma otoksessa?

Tehtävä 7.7. – Mitä opimme?

Tehtävän (a)-kohdassa tarkastellaan *aritmeettisen keskiarvon ja otosvarianssin otosjakaumia*.

Tehtävän (b)-kohdassa tarkastellaan *suhteellisen osuuden (approksimatiivista) otosjakaumaa*.

Tehtävä 7.7. – Ratkaisu:

- (a) Oletuksen mukaan havainnot $X_1, X_2, \dots, X_n$ muodostavat *yksinkertaisen satunnaisotoksen* normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$n = 10$$

$$\mu = 10 \text{ mm}$$

$$\sigma^2 = 0.01^2 \text{ mm} = 0.0001 \text{ mm}^2$$

Siten kuulien halkaisijoiden aritmeettinen keskiarvo $\bar{X}$ noudattaa otoksessa *normaalijakaumaa* $N(\mu, \sigma^2/n)$, jossa

$$\mu = E(\bar{X}) = 10 \text{ mm}$$

$$\frac{\sigma^2}{n} = \text{Var}(\bar{X}) = D^2(\bar{X}) = \frac{0.0001}{10} = 0.00001 \text{ mm}^2$$

Olkoon s^2 kuulien halkaisijoiden varianssi otoksessa. Tällöin satunnaismuuttuja $(n-1)s^2/\sigma^2$ noudattaa otoksessa χ^2 -jakaumaa vapausastein

$$n-1 = 10-1 = 9$$

(b) Olkoon

A = satunnaisesti valittu äänestäjä kannattaa puoluetta ABC

Oletuksen mukaan

$$\Pr(A) = p = 0.25$$

Poimitaan äänestäjien joukosta *yksinkertainen satunnaisotos*, jonka koko on $n = 1000$.

Puoluetta ABC kannattavien äänestäjien suhteellinen frekvenssi $\hat{p} = f/n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa*:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

jossa siis

$$p = \Pr(A) = 0.25$$

$$q = \Pr(A^c) = 1 - \Pr(A) = 1 - p = 0.75$$

Siten puolueen ABC kannattajien suhteellinen frekvenssi $\hat{p} = f/n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa* parametrein

$$E(\hat{p}) = p = 0.25$$

$$\text{Var}(\hat{p}) = D^2(\hat{p}) = \frac{pq}{n} = \frac{0.25 \times 0.75}{1000} = \frac{0.1875}{1000} = 0.0001875$$

Tehtävä 7.8.

- (a) Miesten pituus eräässä maassa vaihtelee satunnaisesti noudattaen normaalijakaumaa parametrein

$$\mu = 180 \text{ cm}, \sigma = 5 \text{ cm}$$

Poimitaan miesten joukosta yksinkertainen satunnaisotos, jonka koko $n = 100$.

Olkoot $\bar{X}$ ja s^2 pituuksien aritmeettinen keskiarvo ja otosvarianssi otoksessa.

Mitkä ovat aritmeettisen keskiarvon $\bar{X}$ ja otosvarianssin s^2 muunnoksen $(n-1)s^2/\sigma^2$ jakaumat otoksessa?

- (b) Koneen valmistamista muttereista 5 % on viallisia. Poimitaan muttereiden joukosta yksinkertainen satunnaisotos, jonka koko $n = 100$.

Mikä on viallisten muttereiden suhteellisen osuuden f/n approksimatiivinen jakauma otoksessa?

Tehtävä 7.8. – Mitä opimme?

Tehtävän (a)-kohdassa tarkastellaan *aritmeettisen keskiarvon ja otosvarianssin otosjakaumia*.

Tehtävän (b)-kohdassa tarkastellaan *suhteellisen osuuden (approksimatiivista) otosjakaumaa*.

Tehtävä 7.8. – Ratkaisu:

- (a) Oletuksen mukaan havainnot $X_1, X_2, \dots, X_n$ muodostavat *yksinkertaisen satunnaisotoksen* normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$n = 100$$

$$\mu = 185 \text{ cm}$$

$$\sigma^2 = 5^2 \text{ cm} = 25 \text{ cm}^2$$

Siten miesten pituuksien aritmeettinen keskiarvo $\bar{X}$ noudattaa otoksessa *normaalijakaumaa* $N(\mu, \sigma^2/n)$, jossa

$$\mu = E(\bar{X}) = 185 \text{ cm}$$

$$\frac{\sigma^2}{n} = \text{Var}(\bar{X}) = D^2(\bar{X}) = \frac{25}{100} = 0.25 \text{ cm}^2$$

Olkoon s^2 miesten pituuksien varianssi otoksessa. Tällöin satunnaismuuttuja $(n-1)s^2/\sigma^2$ noudattaa otoksessa χ^2 -jakaumaa vapausastein

$$n-1 = 100-1 = 99$$

(b) Olkoon

A = satunnaisesti valittu mutteri on viallinen

Oletuksen mukaan

$$\Pr(A) = p = 0.05$$

Poimitaan muttereiden joukosta *yksinkertainen satunnaisotos*, jonka koko on $n = 100$.

Viallisten muttereiden suhteellinen frekvenssi $\hat{p} = f / n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa*:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

jossa siis

$$p = \Pr(A) = 0.05$$

$$q = \Pr(A^c) = 1 - \Pr(A) = 1 - p = 0.95$$

Siten viallisten muttereiden suhteellinen frekvenssi $\hat{p} = f / n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa* parametrein

$$E(\hat{p}) = p = 0.05$$

$$\text{Var}(\hat{p}) = D^2(\hat{p}) = \frac{pq}{n} = \frac{0.05 \times 0.95}{100} = \frac{0.0475}{100} = 0.000475$$

kasvaa.

Huomautuksia tehtäviin 7.7. ja 7.8.:

- (1) Tehtävien 7.7. ja 7.8. ideana on kertoa siitä, *millaisia ovat tavanomaisten havainnoista laskettavien otostunnuslukujen jakaumat perusjoukossa, jos havaintojen jakauma perusjoukossa tunnetaan.*
- (2) Otostunnuslukujen jakaumia koskevat tulokset ovat kuitenkin *epäoperationaalisia*, koska *jakaumien parametreja ei yleensä tunneta.*
- (3) Jos havaintojen jakauman parametreja ei tunneta, ne voidaan pyrkiä *estimoimaan* eli arvioimaan otoksesta saatujen tietojen perusteella; ks. lukua **Tilastollisten mallien parametrien estimointi**.
- (4) Perusjoukon parametrien arvoista tehtyjä oletuksia voidaan pyrkiä *testaamaan tilastollisesti* otoksesta saatujen tietojen perusteella; ks. lukua **Tilastollisten hypoteesien testaus**.
- (5) Myös perusjoukon jakauman tyyppiä koskevia oletuksia voidaan pyrkiä *testaamaan tilastollisesti* otoksesta saatujen tietojen perusteella; ks. lukua **Yhteensopivuuden, homogeneisuuden ja riippumattomuuden testaaminen**.