

Tilastollisia tunnuslukuja

Esimerkki 1

Taulukossa on esitetty Lapin kuntien väkiluvut kesäkuussa 2017. Määritä väkilukujen mediaani ja keskiarvo. Kumpi keskiluku antaa paremman kuvan Lapin kuntien väkiluvusta?

Kunta	Väkiluku
Enontekiö	1 881
Inari	6 826
Kemi	21 408
Kemijärvi	7 604
Keminmaa	8 300
Kittilä	6 360
Kolari	3 794
Muonio	2 319
Pelkosenniemi	944
Pello	3 530
Posio	3 378

Ranua	4 010
Rovaniemi	61 854
Salla	3 638
Savukoski	1 034
Simo	3 149
Sodankylä	8 596
Tervola	3 076
Tornio	22 041
Utsjoki	1 238
Ylitornio	4 143

Avaa OpenOffice-tiedosto →

Ratkaisu

Tapa 1 ilman laskentaohjelmaa

Järjestetään kunnat väkiluvun mukaiseen järjestykseen.

Kunta	Väkiluku
Pelkosenniemi	944
Savukoski	1 034
Utsjoki	1 238
Enontekiö	1 881
Muonio	2 319
Tervola	3 076
Simo	3 149
Posio	3 378
Pello	3 530
Salla	3 638
Kolari	3 794

	Mediaani on väkiluvun mukaan järjestetyn aineiston keskimäinen havaintoarvo. Mediaani (Md) on 3 794.
Ranua	4 010
Ylitornio	4 143
Kittilä	6 360
Inari	6 826
Kemijärvi	7 604
Keminmaa	8 300
Sodankylä	8 596
Kemi	21 408
Tornio	22 041
Rovaniemi	61 854

Lasketaan väkilukujen keskiarvo.

$$\bar{x} = \frac{944 + 1\,034 + \dots + 22\,041 + 61\,854}{21} = 8\,529,66\dots \approx 8\,530$$

Tapa 2 laskentaohjelman avulla

Siirretään taulukon tiedot laskentaohjelmaan, ja määritetään väkilukujen keskiarvo ja mediaani. Keskiarvo on $\bar{x} \approx 8\,530$ ja mediaani on $Md = 3\,794$.

Jakaumaa vinouttaa Rovaniemen väkiluku, joka on suuri verrattuna muiden kuntien väkilukuihin. Mediaani antaa tällöin paremman kuvan Lapin kuntien väkiluvusta kuin keskiarvo, koska kaikista 21 kunnasta vain neljässä väkiluku on yli keskiarvon.

Vastaus

Keskiarvo on 8 530 ja mediaani 3 794. Mediaani kuvaa

Lapin kuntien väkilukua keskiarvoa paremmin.

Keskiluvut

Keskilukuja ovat keskiarvo, mediaani ja moodi.

Keskiarvo ($\bar{x}$) lasketaan jakamalla havaintoarvojen $x_1, x_2, \dots, x_n$ summa havaintojen lukumäärällä n .

Keskiarvo on

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Moodi (Mo) eli tyyppiarvo on aineiston yleisin tulos.

Moodeja voi olla useampi kuin yksi.

Mediaani (Md) on tutkittavan ominaisuuden mukaan järjestetyn aineiston keskimäinen havainto. Jos havaintoja on parillinen määrä, mediaani on kahden keskimäisen havainnon keskiarvo. Jos kahden keskimäisen havainnon keskiarvoa ei voi laskea, ne ovat molemmat mediaaneja.

Esimerkki 2

Taulukoissa on esitetty vuoden 2017 New Yorkin maratonin 12 parhaan naisen ja miehen loppuajat. Laske molemmista tulostaulukoista keskihajonnat ja vertaa niitä toisiinsa. Kummassa kisassa kärjen loppuajoissa oli keskimäärin vähemmän vaihtelua?

Naisten tulokset, 12 parasta

Sijoitus	Nimi	Maa	Loppuaika
1	Shalane Flanagan	USA	2:26:53

Miesten tulokset, 12 parasta

Sijoitus	Nimi	Maa	Loppuaika
1	Geoffrey Kamworor	KEN	2:10:53

2	Mary Keitany	KEN	2:27:54
3	Mamitu Daska	ETH	2:28:08
4	Edna Kiplagat	KEN	2:29:36
5	Allie Kieffer	USA	2:29:39
6	Sara Dossena	ITA	2:29:39
7	Eva Vrabcova	CZE	2:29:41
8	Kellyn Taylor	USA	2:29:56
9	Diane Nukuri	BDI	2:31:21
10	Stephanie Bruce	USA	2:31:44
11	Buzunesh Deba	ETH	2:32:01
12	Christelle Daunay	FRA	2:32:09

2	Wilson Kipsang	KEN	2:10:56
3	Lelisa Desisa	ETH	2:11:32
4	Lemi Berhanu	ETH	2:11:52
5	Tadesse Abraham	SUI	2:12:01
6	Michel Butter	NED	2:12:39
7	Abdi Abdirahman	USA	2:12:48
8	Koen Naert	BEL	2:13:21
9	Fikadu Girma Teferi	ETH	2:13:58
10	Shadrack Biwott	USA	2:14:57
11	Meb Keflezighi	USA	2:15:29
12	Jared Ward	USA	2:18:39

Avaa OpenOffice-tiedosto →

Ratkaisu

Määritetään aineistoista keskihajonnat laskentaohjelman avulla.

Naisten kilpailussa 12 ensimmäisen loppuajan keskihajonta oli $\sigma \approx 0:01:37 = 1 \text{ min } 37 \text{ s}$ ja miesten kilpailussa se oli $\sigma \approx 0:02:09 = 2 \text{ min } 9 \text{ s}$.

Vaikka naisilla kului maratoniin enemmän aikaa kuin miehillä, naisten loppuaikojen keskihajonta oli kuitenkin pienempi. Naisten kilpailun 12 ensimmäisen juoksijan loppuajoissa oli keskimäärin vähemmän vaihtelua kuin miesten loppuajoissa.

Vastaus

Naisten kilpailun parhaiden loppuaikojen keskihajonta oli 1

Osa laskentaohjelmista pystyy käsittelemään ajat tunteina, minuutteina ja sekunteina. Jos ohjelma ei osaa tulkita näin syötettyjä aikoja, ajat muutetaan ensin sekunneiksi.

min 37 s ja miesten kilpailun 2 min 9 s. Naisten kilpailun loppuajoissa oli keskimäärin vähemmän vaihtelua kuin miesten.

Keskihajonta

Keskihajonta mittaa havaintoarvojen etäisyyttä jakauman keskiarvosta. Keskihajonta lasketaan yleensä laskentaohjelmalla. Keskihajonnan symboli on ohjelmasta riippuen σ , σ_n tai s_n .



Keskihajonta on

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}},$$

missä

$x_1, x_2, x_3, \dots, x_n$ ovat havaintoarvoja, joiden hajontaa mitataan

$\bar{x}$ on havaintoarvojen keskiarvo

n on havaintoarvojen lukumäärä.

Otos ja otoskeskihajonta

Usein ei ole mahdollista eikä tarkoituksenmukaistakaan tutkia kaikkia perusjoukon havaintoyksiköitä, vaan tulokset perustuvat perusjoukosta satunnaisesti valikoituneeseen otokseen. Jos on kyse otoksesta, keskihajonta lasketaan otoskeskihajontana.

Otoskeskihajonnan symboli on ohjelmasta riippuen s , s_{n-1} tai σ_{n-1} .

Kun laskentaohjelmat laskevat tilastollisia tunnuslukuja, ne antavat yleensä sekä keskihajonnan että otoskeskihajonnan. Otoskeskihajonta on näistä hajontaluvuista suurempi.

Otoskeskihajonta on

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}},$$

missä

$x_1, x_2, x_3, \dots, x_n$ ovat havaintoarvoja, joiden hajontaa mitataan

$\bar{x}$ on havaintoarvojen keskiarvo

n on havaintoarvojen lukumäärä.

Otoskeskihajonnan kaavassa jakajana on havaintojen määrä vähennettynä yhdellä. Mitä suurempi aineisto on, sitä pienempi ero on otoskeskihajonnan ja keskihajonnan välillä.

Esimerkki 3

Konsertin järjestäjä halusi selvittää yleisön tyytyväisyyttä konsertin järjestelyihin. Joka kymmenennelle istuimelle laitettiin kyselykaavake, jossa oli kolme kysymystä. Vastausprosentti oli 90.



Arvioi tämän illan konsertin järjestelyitä kouluarvosanoilla 4–10.

Salin äänentoisto ☐

Salin viihtyisyys ☐

Henkilökunnan ystävällisyys ☐

KIITOS VASTAUKSESTANNE!

Kyselyn tulokset olivat seuraavat:

Arvosana	Äänentoisto, frekvenssi	Viihtyisyys, frekvenssi	Ystävällisyys, frekvenssi
4	4	2	1
5	4	4	2

6	10	8	4
7	0	10	5
8	5	6	8
9	9	6	11
10	4	0	5

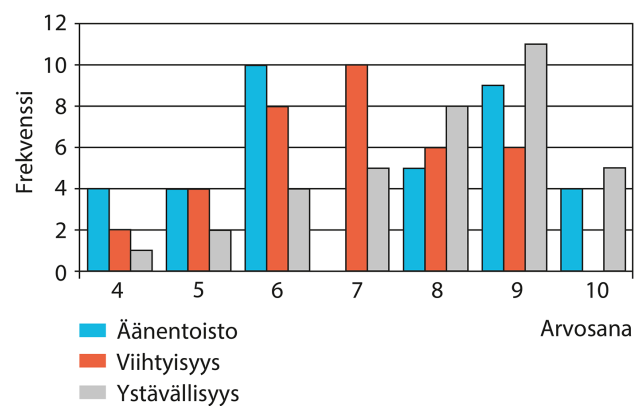
Avaa OpenOffice-tiedosto →

- a) Havainnollista kolmea jakaumaa pylväskuviolla.
 b) Laske jakaumien keskiarvo ja keskihajonta. Tulkitse saatuja tuloksia.

Ratkaisu

- a) Kuvataan jakaumat samassa pylväskuviossa.

Yleisön tyytyväisyys konsertin järjestelyihin



- b) Koska kyseessä on otos, lasketaan jakaumien otoskeskihajonnat.

Siirretään frekvenssitaulukko laskentaohjelmaan, ja määritetään sen avulla keskiarvot sekä otoskeskihajonnat.

	Äänentoisto	Viihtyisyys	Ystävällisyys
Keskiarvo	7,14	6,89	7,94

Otoskeskihajonta	1,99	1,43	1,57
-------------------------	------	------	------

Keskiarvon mukaan asiakkaat olivat tyytyväisimpiä henkilökunnan ystävällisyyteen ja tyytymättömmimpiä salin viihtyisyyteen. Asiakkaiden mielipiteet vaihtelivat eniten arvioitaessa äänentoistoa ja vähiten arvioitaessa viihtyisyyttä.

Esimerkki 4

Taulukossa on esitetty Vaasassa ja Joensuussa 18.3.2018 myynnissä olleet kerrostaloasunnot huoneluvun mukaan.

Huoneluku	Vaasa	Joensuu
1 h	35	46
2 h	172	244
3 h	135	169
4 h	50	22
5 h	2	4
6 h	0	2
Yhteensä	394	487

Lähde: Oikotie

Avaa OpenOffice-tiedosto →

a) Laske eri huonelukujen suhteelliset frekvenssit Vaasassa ja Joensuussa.

b) Havainnollista jakaumia kuviolla.

Ratkaisu

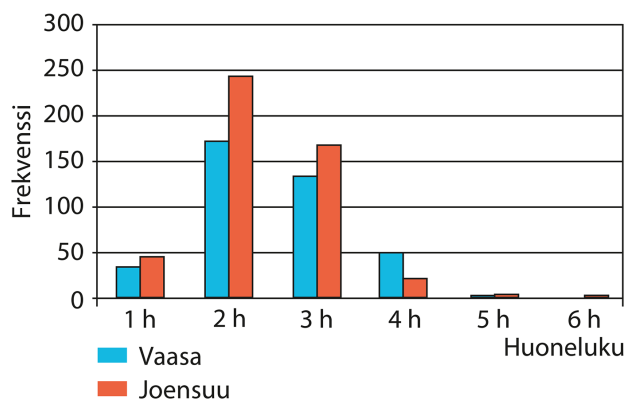
a) Siirretään taulukko laskentaohjelmaan, ja lasketaan prosenttiosuudet.

Huoneluku	Vaasa	Joensuu
1 h	$\frac{43}{89} \approx 8,9 \%$	$\frac{46}{487} \approx 9,4 \%$
2 h	43,7 %	50,1 %
3 h	34,3 %	34,7 %
4 h	12,7 %	4,5 %
5 h	0,5 %	0,8 %
6 h	0 %	0,4 %
Yhteensä	100 %	100 %

b) Huonelukujen jakaumia voidaan havainnollistaa esimerkiksi pylväskuviolla. Pylväskuvio voidaan laatia joko lukumääristä tai prosenttiosuuksista riippuen siitä, mitä halutaan tarkastella.

Laaditaan pylväskuvio lukumääristä.

Myynnissä olleet kerrostaloasunnot Vaasassa ja Joensuussa



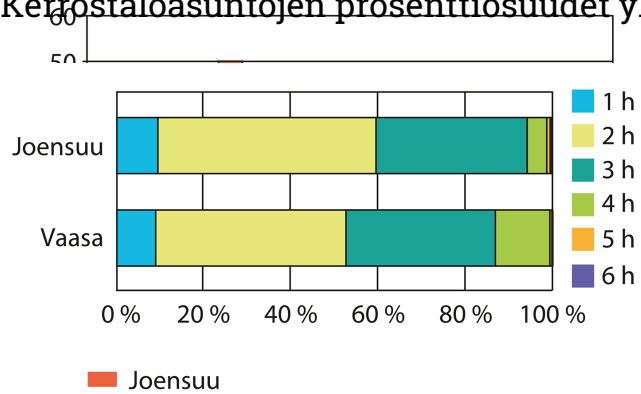
Kuvio havainnollistaa eri kokoisten asuntojen lukumääriä Vaasassa ja Joensuussa.

Laaditaan pylväskuvio prosenttiosuuksista.

Myynnissä olleiden kerrostaloasuntojen prosentuaalinen jakauma

Prosenttiosuuksia voidaan havainnollistaa myös palkkikuviolla.

Kerrostaloasuntojen prosenttiosuudet yhteensä



Kuvio havainnollistaa, kuinka suuri osa myynnissä olleista asunnoista oli huoneluvultaan tietyn kokoisia. Kuvioista ei voi lukea myynnissä olleiden asuntojen lukumääriä.

Avainkäsitteet

- keskiarvo
- mediaani
- moodi
- keskihajonta
- otoskeskihajonta