

1

Tilastotieteen peruskäsitteitä

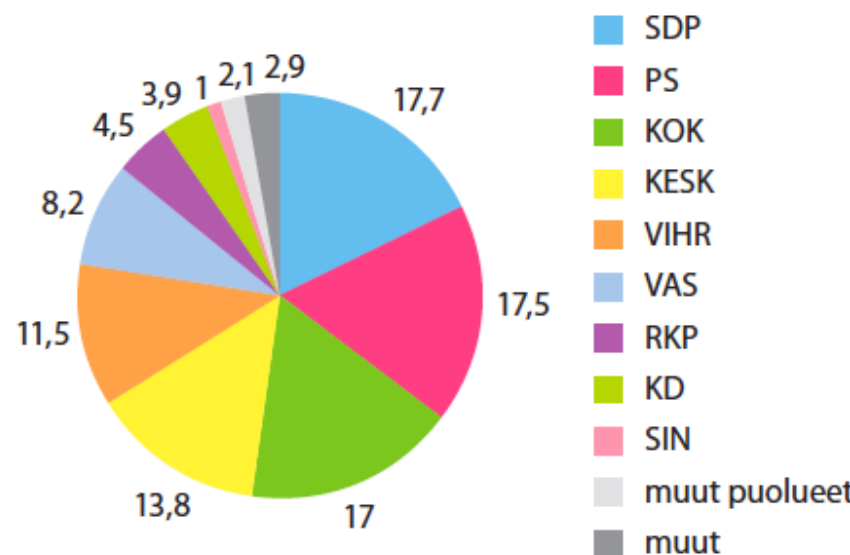
Tilastotiede

Taulukot ja kuvaajat ovat tapa tiivistää ja havainnollistaa tietoa ja niiden tehtävänä on kertoa asiasta tärkeimmät seikat yksinkertaisesti ja selkeästi.

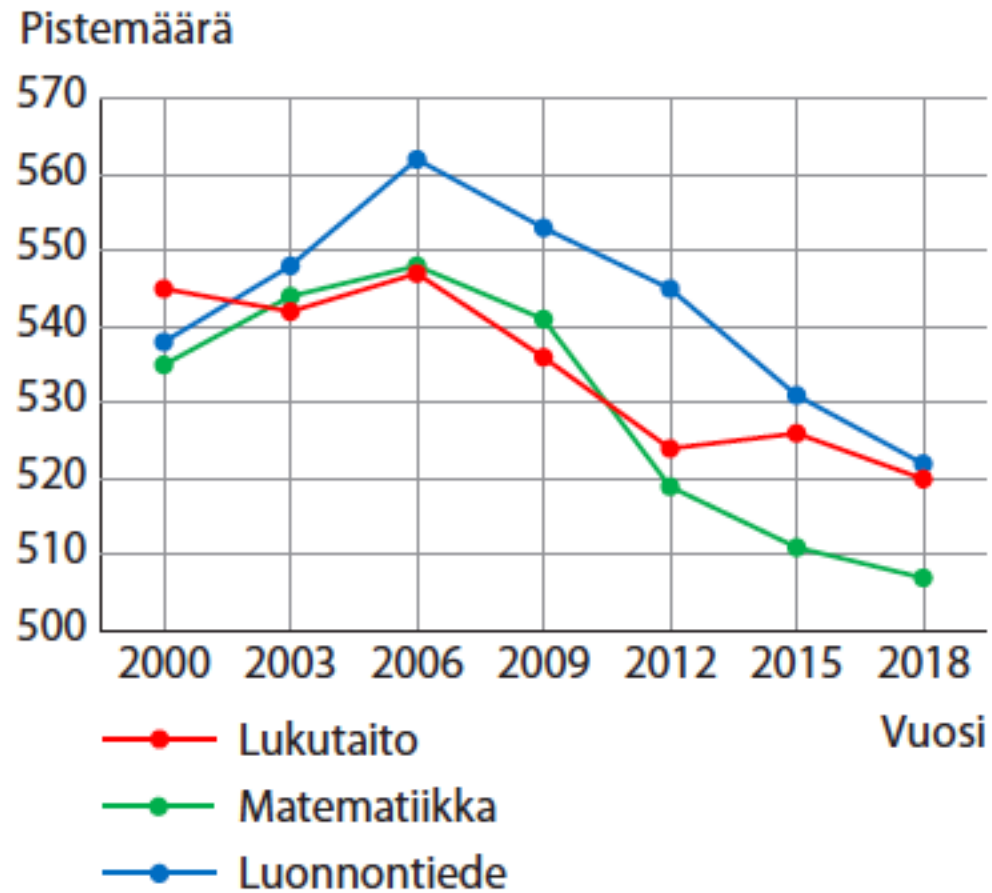
Puolue	Osuus (%)
SDP	17,7
PS	17,5
KOK	17
KESK	13,8
VIHR	11,5
VAS	8,2
RKP	4,5
KD	3,9
SIN	1
muut puolueet	2,1
muut	2,9

Lähde: Tilastokeskus

Puolueiden suhteelliset äänimäärät eduskuntavaaleissa 2019

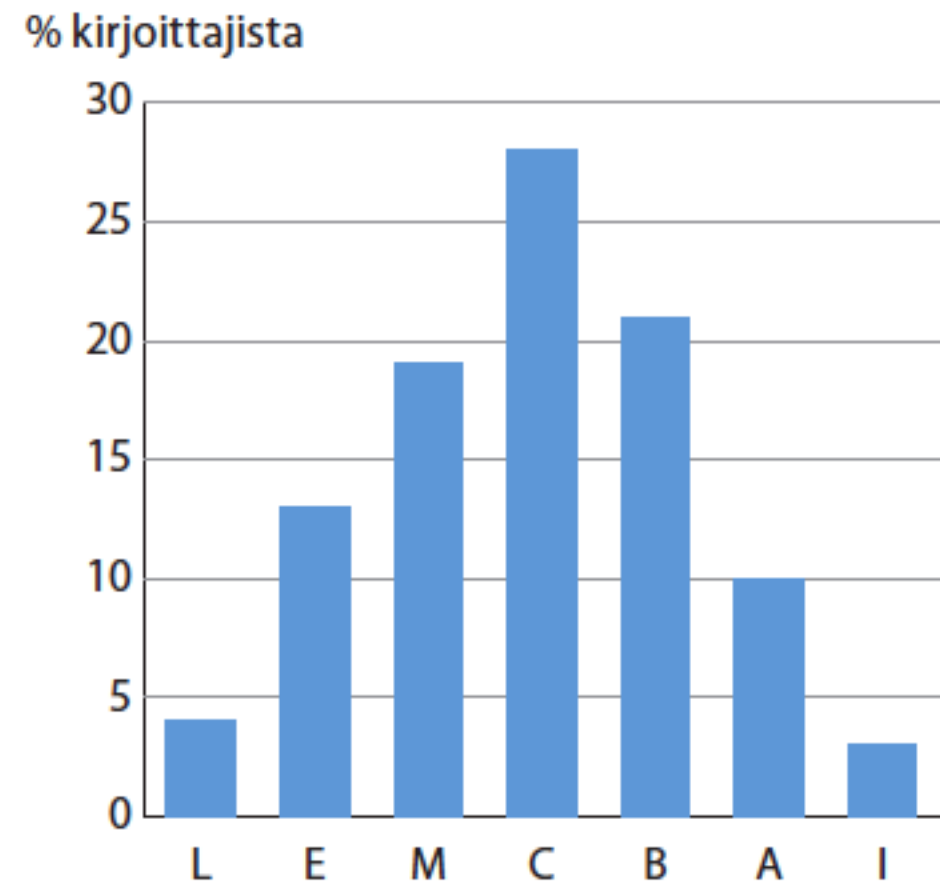


Suomen Pisa-historia



Lähde: OECD Pisa 2018

Äidinkielen yo-kokeen arvosanajakauma keväällä 2019



Lähde: YTL

Havaintoaineisto ja sen kerääminen

Havainto- eli **tilastoaineisto** voidaan kerätä esimerkiksi kyselylomakkeella.

Tutkimus kohdistuu joukkoon, jota kutsutaan **perusjoukoksi** eli **populaatioksi**.

Havaintoaineisto

eli tutkimusta varten kerätty tieto

Yksittäistä tutkimuksen kohdetta kutsutaan **havainto-** eli **tilastoyksiköksi**.

Jos tieto kerätään kaikilta perusjoukon havaintoyksiköiltä, havaintoaineistoa kutsutaan **kokonaisaineistoksi**.

Havaintoaineisto

eli tutkimusta varten kerätty tieto

Havaintoyksikkö

eli tutkimuksen kohde (esimerkiksi opiskelija)

Tutkittavia ominaisuuksia kutsutaan **muuttujiksi**.

Havaintoaineisto koostuu vastauksista eli **havaintoarvoista**.

Tilastolliset muuttujat voidaan luokitella **jatkuviin** ja **diskreetteihin**.

Havaintoaineisto

eli tutkimusta varten kerätty tieto

Havaintoyksikkö

eli tutkimuksen kohde (esimerkiksi opiskelija)

Tilastomuuttuja

eli tutkittava ominaisuus
(esimerkiksi käytettyjen sovellusten lukumäärä)

A

Esimerkki 1

Suomalaisten lukiolaisten on useissa tutkimuksissa todettu nukkuvan liian lyhyet yöunet. Tuhannelta lukiolaiselta kysyttiin, kuinka monta tuntia vuorokaudessa he arvioivat nukkuvansa. Vastaukset on koottu alla olevaan taulukkoon.

Yöunien pituus (h)	6	7	8	9	10
Lukiolaisten määrä	112	240	468	135	45

a) Mikä on tutkimuksen perusjoukko?

Perusjoukko

Tutkimuksen kohteena oleva joukko.

Ratkaisu

a) Koska tutkitaan suomalaisten lukiolaisten yönien pituutta, perusjoukkona ovat kaikki Suomen lukiolaiset.

A

Esimerkki 1

Suomalaisten lukiolaisten on useissa tutkimuksissa todettu nukkuvan liian lyhyet yöunet. Tuhannelta lukiolaiselta kysyttiin, kuinka monta tuntia vuorokaudessa he arvioivat nukkuvansa. Vastaukset on koottu alla olevaan taulukkoon.

Yöunien pituus (h)	6	7	8	9	10
Lukiolaisten määrä	112	240	468	135	45

b) Mitkä ovat tutkimuksen havaintoyksiköt?

Havaintoyksikkö

Yksittäinen tutkimuksen kohde.

Ratkaisu

b) Tutkimuksen kohteina eli havaintoyksikköinä ovat ne tuhat lukio-laista, jotka on valikoitu tutkimukseen.

A

Esimerkki 1

Suomalaisten lukiolaisten on useissa tutkimuksissa todettu nukkuvan liian lyhyet yöunet. Tuhannelta lukiolaiselta kysyttiin, **kuinka monta tuntia vuorokaudessa he arvioivat nukkuvansa**. Vastaukset on koottu alla olevaan taulukkoon.

Yöunien pituus (h)	6	7	8	9	10
Lukiolaisten määrä	112	240	468	135	45

c) Mikä on havaintoaineistossa muuttujana?

Muuttuja

Tutkittava ominaisuus.

Ratkaisu

c) Tutkittava ominaisuus eli muuttuja on yöunen pituus tunteina.

Suomalaisten lukiolaisten on useissa tutkimuksissa todettu nukkuvan liian lyhyet yöunet. Tuhannelta lukiolaiselta kysyttiin, kuinka monta tuntia vuorokaudessa he arvioivat nukkuvansa. Vastaukset on koottu alla olevaan taulukkoon.

Yöunien pituus (h)	6	7	8	9	10
Lukiolaisten määrä	112	240	468	135	45

Kokonaisaineisto

*Tieto kerätään kaikilta
perusjoukon havaintoyksiköiltä.*

Otos

*Perusjoukosta valitaan osa
havaintoyksiköistä.*

d) Onko kyseessä otos vai kokonaisaineisto?

Ratkaisu

d) Koska tutkimusta varten ei ole kysytty yöunen pituutta jokaiselta suomalaiselta lukiolaiselta, havaintoaineisto on otos.

Tee

**tehtävät 1.1
ja 1.2**

Havaintoaineiston kuvaaminen

Havaintoaineiston keräämisen jälkeen aineistoa käsitellään niin, että kaikki olennainen tieto saadaan näkyviin.

Taulukossa ovat muuttujan saamat arvot ja niiden esiintymiskerrat eli **frekvenssit** f .

Sisarusten lukumäärä on muuttuja.
Vastaajilla oli 0–3 sisarusta.



Sisarusten lukumäärä x	Frekvenssi f
0	4
1	12
2	6
3	3

Frekvenssien summa on
 $4 + 12 + 6 + 3 = 25$
eli vastaajien lukumäärä.

Suhteellinen frekvenssi $f\%$ kertoo esiintymiskertojen prosentuaalisen osuuden havaintoyksiköiden kokonaismäärästä.

Sisarusten lukumäärä x	Frekvenssi f	Suhteellinen frekvenssi $f\%$
0	4	$\frac{4}{25} \cdot 100 = 16$
1	12	$\frac{12}{25} \cdot 100 = 48$
2	6	$\frac{6}{25} \cdot 100 = 24$
3	3	$\frac{3}{25} \cdot 100 = 12$



Suhteellinen frekvenssi saadaan, kun muuttujan arvon frekvenssi (f) jaetaan frekvenssien summalla 25.

A Esimerkki 2

Lentävä hanhi -ravintolan ravintoloitsija tutki viikon ajan, kuinka suuria seurueita saapui syömään. Tulokset on koottu taulukkoon.

Seurueen henkilöiden lukumäärä x	Seurueiden lukumäärä f
1	62
2	78
3	63
4	55
5	31

a) Kuinka moni henkilö tuli syömään yksin?

Ratkaisu

a) Muuttujan arvon 1 frekvenssi on 62, joten yksin syömään tulevia henkilöitä oli viikon aikana 62 kappaletta.

Frekvenssi

Muuttujan arvon esiintymiskerrat.

A

Esimerkki 2

Lentävä hanhi -ravintolan ravintoloitsija tutki viikon ajan, kuinka suuria seurueita saapui syömään. Tulokset on koottu taulukkoon.

Seurueen henkilöiden lukumäärä x	Seurueiden lukumäärä f
1	62
2	78
3	63
4	55
5	31

Ratkaisu

- b) Viikon aikana ravintolassa käyneiden seurueiden määrä saadaan laskemalla frekvenssien summa.

$$62 + 78 + 63 + 55 + 31 = 289$$

Ravintolassa kävi siis 289 seuruetta.

- b) Kuinka monta seuruetta ravintolassa kävi viikon aikana?

A

Esimerkki 2

Lentävä hanhi -ravintolan ravintoloitsija tutki viikon ajan, kuinka suuria seurueita saapui syömään. Tulokset on koottu taulukkoon.

Seurueen henkilöiden lukumäärä x	Seurueiden lukumäärä f
1	62
2	78
3	63
4	55
5	31

- c) Laske muuttujan arvon 2 suhteellinen frekvenssi. Mitä tämä kuvaa?

Ratkaisu

- c) Suhteellinen frekvenssi saadaan jakamalla muuttujan arvon frekvenssi kaikkien frekvenssien summalla.

$$\frac{78}{289} = 0,2698... \approx 27 \%$$

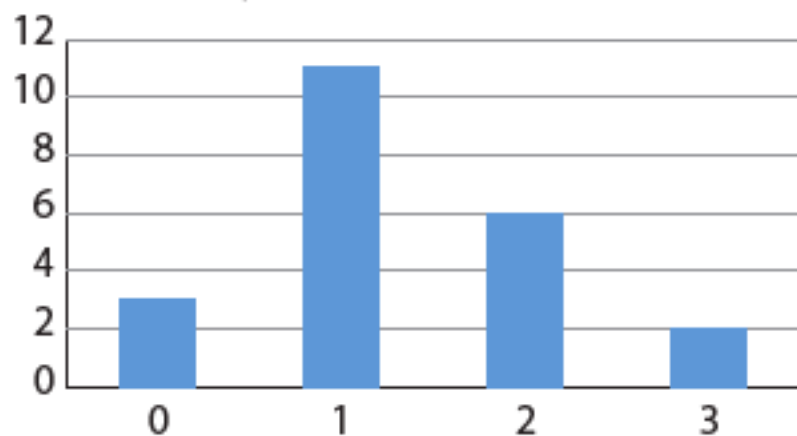
Desimaaliluku muutetaan prosenteiksi kertomalla 100 %:lla.

Tämä kuvaa kahden hengen seurueiden suhteellista osuutta eli kahden hengen seurueita oli 27 %.

Tilastoja voidaan havainnollistaa paitsi taulukoiden myös erilaisten kuvaajien avulla.

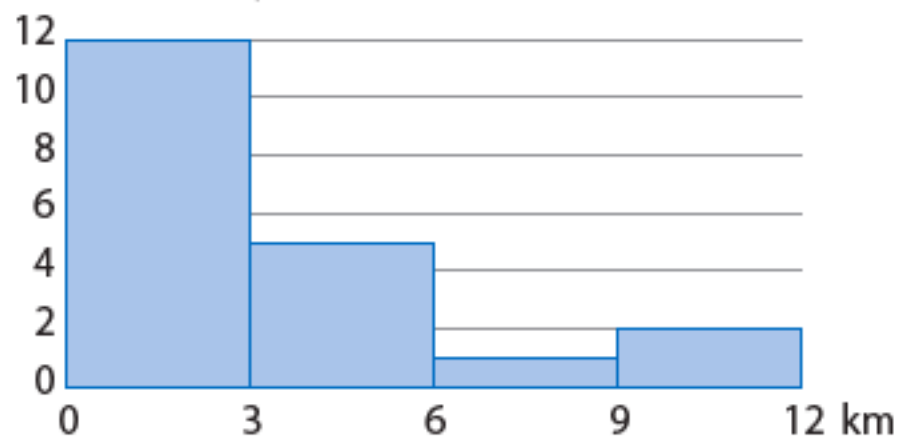
- **Pylväskuvaajalla tai histogrammilla** havainnollistetaan frekvenssi-jakaumia. Pylväiden korkeus määräytyy frekvenssin mukaan.

f Opiskelijoiden sisarusten lukumäärä



Diskreetin muuttujan jakaumaa kuvataan pylväsdiagrammilla, jossa pylväät on piirretty erilleen.

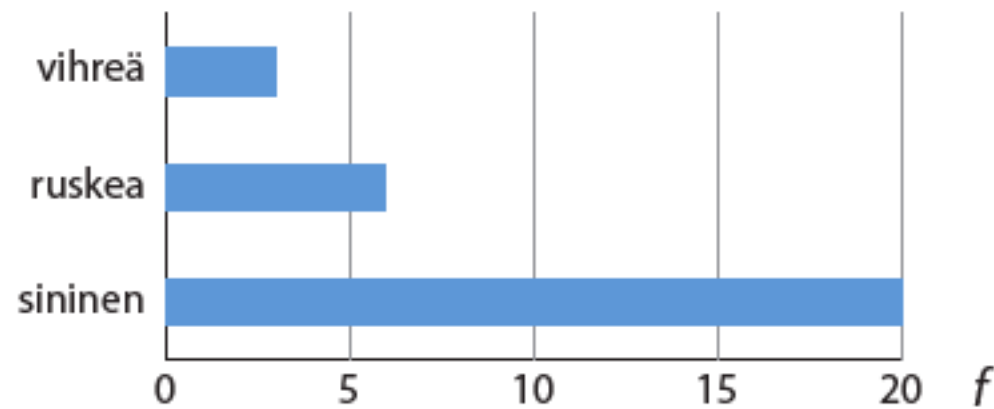
f Opiskelijoiden koulumatkan pituus



Jatkuvan muuttujan jakaumaa kuvataan histogrammilla, jossa pylväät on piirretty yhteen.

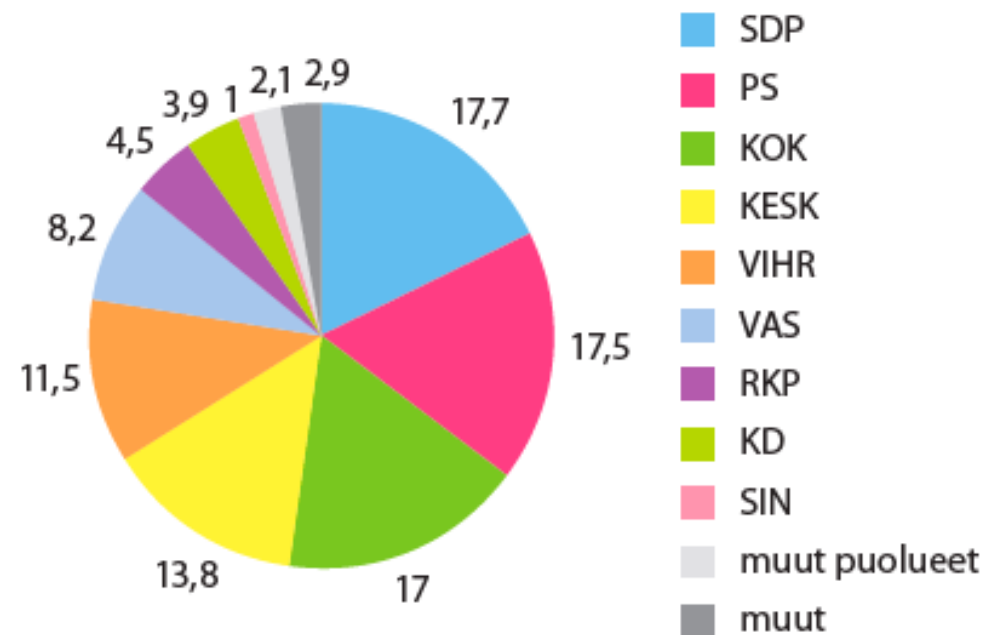
- **Vaakapylväitä** käytetään silloin kun muuttujan arvoilla ei ole luontevaa suuruusjärjestystä.

MAB5-ryhmän opiskelijoiden silmien väri



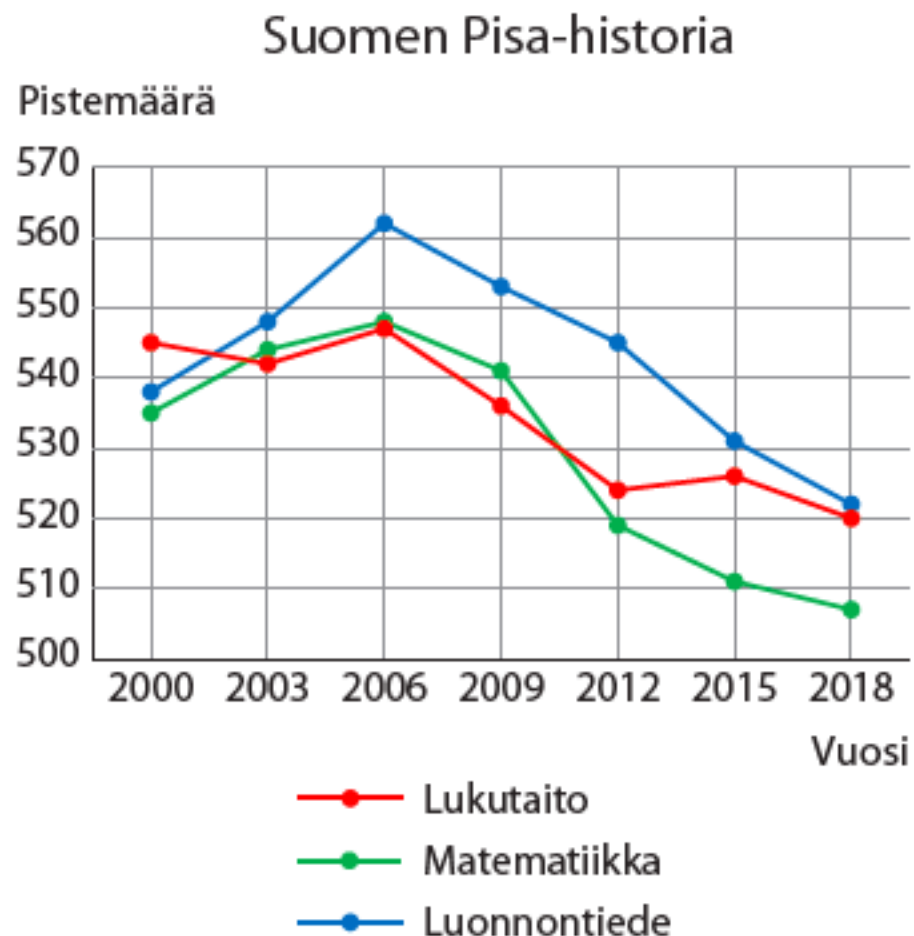
- **Sektoridiagrammilla** havainnollistetaan kokonaisuuden jakaantumista osiin. Sektorin keskuskulman suuruus määräytyy suhteellisen frekvenssin mukaan.

Puolueiden suhteelliset äänimäärät eduskuntavaaleissa 2019



- **Viivadiagrammia** käytetään erityisesti silloin, kun halutaan havainnollistaa jonkin asian kehittymistä tai vaihtelua ajan kuluessa.

Kun samasta tutkittavasta asiasta on havaintoja peräkkäisten, yhtä pitkien ajanjaksojen ajan, tilastoa kutsutaan **aikasarjaksi**.



Asukasyhdistys teki tutkimuksen alueellaan asuvien nuorten viikoittaisten harrastuskertojen määristä. Alla on tuloksista piirretty frekvenssi-jakauman kuvaaja.



- a) Kuinka moni nuori käy harrastuksissa neljä kertaa viikossa?

Ratkaisu

- a) Muuttujan arvoja kuvataan vaaka-akselilla.

Frekvenssejä kuvataan pystyakselilla.

Kunkin pylvään korkeus kertoo muuttujan arvon frekvenssin.

Tutkimukseen osallistuneista nuorista 10 käy harrastuksissa neljä kertaa viikossa.

Asukasyhdistys teki tutkimuksen alueellaan asuvien nuorten viikoittaisten harrastuskertojen määristä. Alla on tuloksista piirretty frekvenssi-jakauman kuvaaja.



Ratkaisu

b) Lasketaan frekvenssit yhteen.

$$35 + 10 + 10 + 5 = 60$$

60 nuorta käy harrastuksissa vähintään kolme kertaa viikossa.

b) Kuinka moni nuori käy harrastuksissa vähintään kolme kertaa viikossa?

A Esimerkki 3

Asukasyhdistys teki tutkimuksen alueellaan asuvien nuorten viikoittaisten harrastuskertojen määristä. Alla on tuloksista piirretty frekvenssi-jakauman kuvaaja.

**Tee
tehtävät
1.3 ja 1.4**



c) Kuinka monta nuorta osallistui tutkimukseen?

Ratkaisu

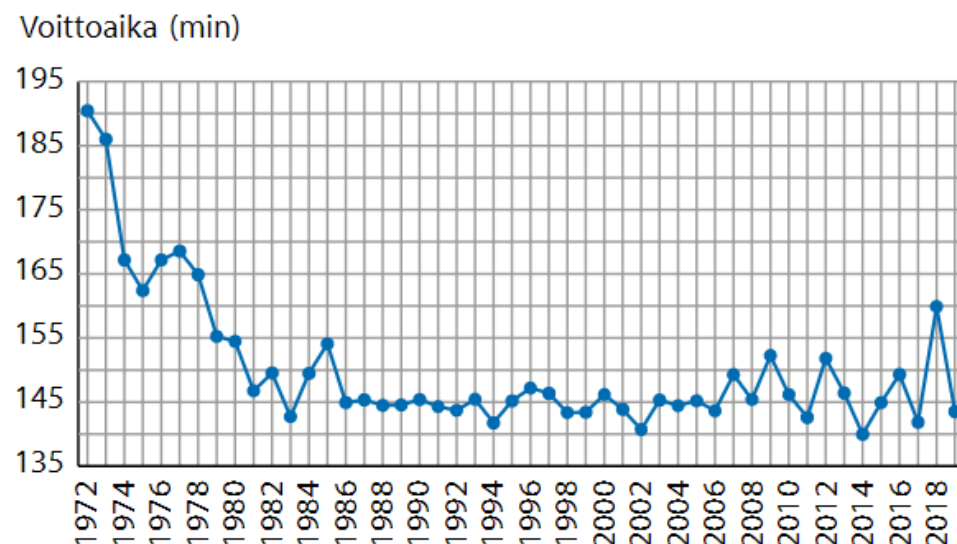
c) Kaikkien tutkimukseen osallistuneiden määrä saadaan, kun lasketaan kaikki frekvenssit yhteen.

$$5 + 15 + 20 + 35 + 10 + 10 + 5 = 100$$

Tutkimukseen osallistui
100 nuorta

A Esimerkki 4

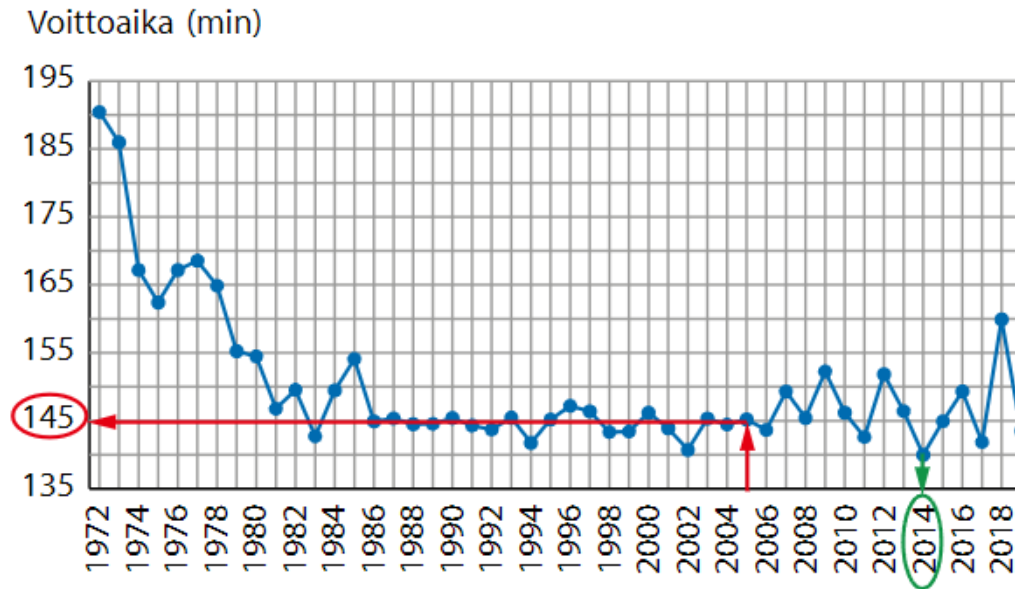
Bostonin maraton on vuosittain järjestettävä 26,2 mailin (eli 42,2 kilometrin) mittainen juoksukilpailu. Se on vanhin vuosittain järjestettävä maratontapahtuma, ja naiset saivat virallisen osallistumisoikeuden vuonna 1972. Kilpailu peruutettiin ensimmäistä kertaa historiansa aikana vuonna 2020 Covid-19-pandemian vuoksi. Oheinen viivakaavio esittää naisten voittoaikoja vuosina 1972–2019.



- a) Mikä oli voittoaika vuonna 2005?
- b) Miltä vuodelta hallitseva ennätys on?
- c) Kuinka monta prosenttia parempi vuoden 2019 voittoaika on verrattuna 1972 voittoaikaan?

Ratkaisu

- a) Voittoaika vuonna 2005 on **145** minuuttia.
- b) Ennätysaika on se, jolloin maratonin juoksemiseen on käytetty vähiten aikaa. Tämä löytyy viivakaavion alimmasta pisteestä. Hallitseva ennätys on vuodelta **2014**.

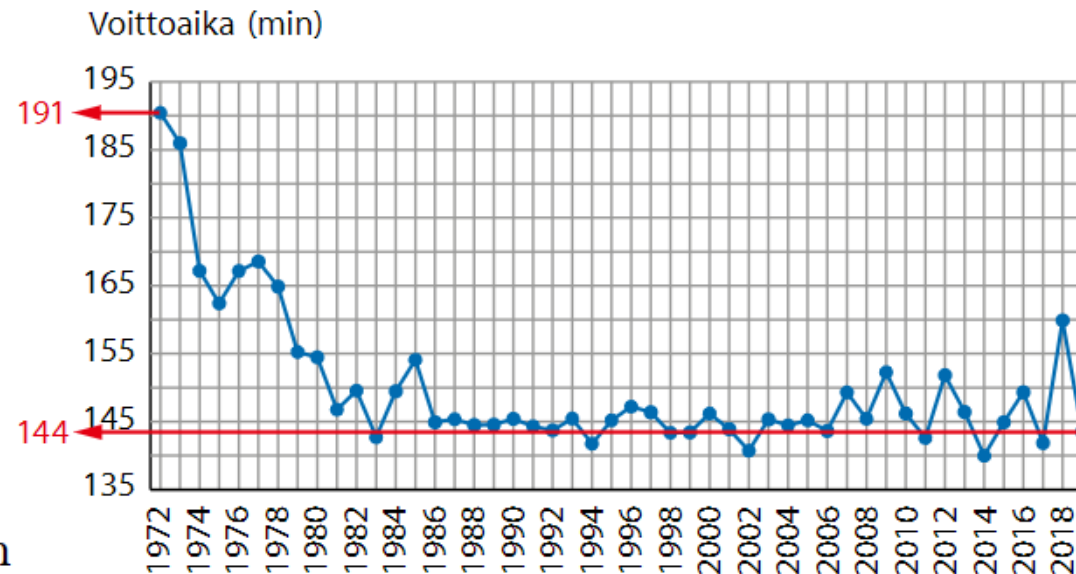


**Tee
tehtävät
1.5, 1.8 ja
1.20**

- c) Luetaan viivakaaviosta vuosien 1972 ja 2019 voittoajat. Lasketaan voittoaikojen erotus.

$$191 \text{ min} - 144 \text{ min} = 47 \text{ min}$$

Lasketaan, kuinka monta prosenttia tämä on vuoden 1972 ajasta.



2

Frekvenssijakauma

Tavallisimpia keinoja tiivistää havaintoaineisto on esittää taulukkomuodossa muuttujan eri arvot ja niiden lukumäärät eli frekvenssit aineistossa.

Tällaista taulukkoa kutsutaan **frekvenssijakaumaksi**.

Usein lasketaan myös suhteelliset frekvenssit eli frekvenssien prosenttiosuudet.

Sisarusten lukumäärä x	Opiskelijoiden lukumäärä f	$f\%$
0	4	$\frac{4}{25} \cdot 100 = 16$
1	12	$\frac{12}{25} \cdot 100 = 48$
2	6	$\frac{6}{25} \cdot 100 = 24$
3	3	$\frac{3}{25} \cdot 100 = 12$

Kun frekvenssejä lasketaan yhteen, saadaan **summafrekvenssijakauma**.

- **Summafrekvenssi** sf ilmaisee kertymän lukumäärän avulla.
- **Suhteellinen summafrekvenssi** $sf\%$ ilmaisee kertymän prosentteina.

Sisarusten lukumäärä x	Opiskelijoiden lukumäärä f	$f\%$	sf	$sf\%$
0	4	$\frac{4}{25} \cdot 100 = 16$	4	16
1	12	$\frac{12}{25} \cdot 100 = 48$	$4 + 12 = 16$	$16 + 48 = 64$
2	6	$\frac{6}{25} \cdot 100 = 24$	$16 + 6 = 22$	$64 + 24 = 88$
3	3	$\frac{3}{25} \cdot 100 = 12$	$22 + 3 = 25$	$88 + 12 = 100$

A Esimerkki 1

Koulussa tehdyssä kyselyssä selvitettiin, kuinka monta tuntia nuoret yleensä nukkuvat arkiyönä. Osa tiedoista on kuitenkin kadonnut.

x	f	$f\%$	sf	$sf\%$
6	40	20	40	20
7	100	50	C	E
8	35	B	D	F
9	A	12,5	200	100

- a) Määritä puuttuvat frekvenssit A , B , C , D , E ja F .
- b) Kuinka moni nuori nukkui korkeintaan 7 tuntia?
- c) Kuinka monta prosenttia nuorista nukkui korkeintaan 8 tuntia?

Ratkaisu

- a) Suurimman muuttujan arvon 9 summafrekvenssi sf on 200, joten havaintoarvoja on yhteensä 200.

x	f	sf
6	40	40
7	100	
8	35	
9	A	200

$$A = 200 - (40 + 100 + 35) = \mathbf{25}.$$

Suhteellisten frekvenssien $f\%$ summa on aina 100 %.

x	$f\%$	$sf\%$
6	20	20
7	50	
8	B	
9	12,5	100

$$B = 100\% - (20\% + 50\% + 12,5\%) = \mathbf{17,5\%}.$$

Tietyn arvon summa-
frekvenssi on summa siihen-
astisista frekvensseistä.



Summafrekvenssit sf ja suhteelliset summa-
frekvenssit $sf\%$ lasketaan frekvenssien f ja $f\%$ avulla.

0	D	
9	12,5	100

Suhteellinen summa-
frekvenssi voidaan laskea
myös summafrekvenssistä:



$$\frac{140}{200} \cdot 100 = 70$$

x	f	sf
6	40	40
7	100	C
8	35	D
9	25	200

$$C = 40 + 100 = \mathbf{140}$$

$$D = 140 + 35 = \mathbf{175}$$

x	$f\%$	$sf\%$
6	20	20
7	50	E
8	17,5	F
9	12,5	100

$$E = 20 + 50 = \mathbf{70}$$

$$F = 70 + 17,5 = \mathbf{87,5}$$

- b) Luetaan vastaus summafrekvenssijakaumasta (sf).
Korkeintaan **7** tuntia nukkui **140** vastaajaa.

x	f	$f\%$	sf	$sf\%$
6	40	20	40	20
7	100	50	140	70
8	35	17,5	175	87,5
9	25	12,5	200	100

- c) Luetaan vastaus suhteellisen summafrekvenssin jakaumasta ($sf\%$).
Korkeintaan **8** tuntia nukkui **87,5** prosenttia vastaajista.



Moodi ja mediaani

Frekvenssitaulukoiden lisäksi havaintoaineistoa voidaan tarkastella erilaisten **tunnuslukujen** avulla.

Tyyppiarvo eli **moodi** M_o on se muuttujan arvo, jota esiintyy eniten.

x	f	$f\%$
4	2	25
8	3	37,5
10	1	12,5
11	2	25

Muuttujan arvolla 8 on

- suurin frekvenssi (3)
- suurin suhteellinen frekvenssi (37,5 %).

$M_o = \mathbf{8}$

Moodeja voi olla useita, jos suurin frekvenssi on kahdella tai useammalla arvolla.

Mediaani M_d on suuruusjärjestykseen asetetun aineiston keskimmäinen arvo.

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran saavuttaa arvon 50 %.

x	f	$f\%$	$sf\%$
5	6	15	15
10	12	30	45
15	18	45	90
18	4	10	100

Muuttujan arvo 15 on se, johon mennessä kertymä (90 %) on ensimmäisen kerran yli 50 % havainnoista.
 $M_d = \mathbf{15}$

**Tee
tehtävät 2.1
ja 2.5**

Opiskelijoille suunnatussa kyselyssä selvitettiin, kuinka monta kertaa viikossa opiskelija harrasti liikuntaa koulun ulkopuolella. Vastausten perusteella muodostettiin oheinen frekvenssijakauma. Määritä liikuntakertojen moodi ja mediaani.

Liikuntakerrat	$f\%$
0	30
1	15
2	25
3	20
4	10

Ratkaisu

Moodi on se muuttujan arvo, jonka suhteellinen frekvenssi on suurin.

Liikuntakertojen moodi on $M_o = 0$.

Mediaani on se muuttujan arvo, jonka kohdalla suhteellinen summafrekvenssi ensimmäisen kerran saavuttaa arvon 50 %.

Liikuntakerrat	$f\%$	$sf\%$
0	30	30
1	15	$30 + 15 = 45$
2	25	$45 + 25 = 70$
3	20	$70 + 20 = 90$
4	10	$90 + 10 = 100$

Liikuntakertojen mediaani on $Md = 2$.

Vastaus

$Mo = 0$ (kertaa) ja $Md = 2$ (kertaa)

B Esimerkki 3

Katugallupissa kysyttiin, kuinka monta kertaa henkilö oli käynyt elokuvateatterissa kuluneen vuoden aikana. Vastaukset on koottu oheiseen taulukkoon.

- Laadi puuttuvat frekvenssijakaumat.
- Mikä on käyntikerran moodi? Mitä se kertoo vastaajien elokuvissakäynnistä?
- Mikä on käyntikerran mediaani? Mitä se kertoo vastaajien elokuvissakäynnistä?

Käyntikerrat	f
0	25
1	47
2	38
3	21
4	12
5	5
6	2

Ratkaisu

- Täydennetään frekvenssijakaumat taulukkolaskentaohjelmalla laske-
malla suhteelliset frekvenssit $f\%$, summafrekvenssit sf ja suhteelliset
summafrekvenssit $sf\%$ kukin omaan sarakkeeseen.

Aloitetaan laskemalla frekvenssien summa.

Frekvenssien summa

- Kaava alkaa merkillä =.
- Lasketaan frekvenssien summa funktiolla SUMMA().

	A	B
1	Käynti- kerrat	f
2	0	25
3	1	47
4	2	38
5	3	21
6	4	12
7	5	5
8	6	2
9		=SUMMA(B2:B8)



	A	B
1	Käynti- kerrat	f
2	0	25
3	1	47
4	2	38
5	3	21
6	4	12
7	5	5
8	6	2
9		150

Vastaajia oli yhteensä 150.

Laaditaan seuraavaksi suhteellinen frekvenssijakauma $f\%$, joka saadaan jakamalla kukin frekvenssi havaintoyksiköiden kokonaismäärällä eli luvulla 150.

Suhteellinen frekvenssi

- Kaava alkaa merkillä =.
- Kirjoitetaan laskukaava soluviittauksen avulla.

Laaditaan seuraavaksi suhteellinen frekvenssijakauma $f\%$, joka saadaan jakamalla kukin frekvenssi havaintoyksiköiden kokonaismäärällä eli luvulla 150.

	A	B	C
1	Käyntikerrat	f	f%
2	0	25	=B2/150
3	1	47	
4	2	38	
5	3	21	
6	4	12	
7	5	5	
8	6	2	
9		150	



	A	B	C
1	Käyntikerrat	f	f%
2	0	25	0,167
3	1	47	
4	2	38	
5	3	21	
6	4	12	
7	5	5	
8	6	2	
9		150	

Taulukkolaskennassa soluun kirjoitettu laskukaava voidaan kopioida sarakkeessa alaspäin. Osuudet voidaan muotoilla prosenttiluvuksi.

	A	B	C
1	Käynti-kerrat	f	f%
2	0	25	0,167
3	1	47	0,313
4	2	38	0,253
5	3	21	0,140
6	4	12	0,080
7	5	5	0,033
8	6	2	0,013
9	Yhteensä	150	

Muotoile prosentteiksi.

	A	B	C
1	Käynti-kerrat	f	f%
2	0	25	16,7%
3	1	47	31,3%
4	2	38	25,3%
5	3	21	14,0%
6	4	12	8,0%
7	5	5	3,3%
8	6	2	1,3%
9	Yhteensä	150	

Summafrekvenssijakauma sf saadaan laskemalla frekvenssejä f yhteen.

Summafrekvenssi

- Kaava alkaa merkillä =.
- Kirjoitetaan laskukaava soluviittausten avulla.

	A	B	C	D
1	Käynti-kerrat	f	f%	sf
2	0	25	16,7%	25
3	1	47	31,3%	=D2+B3
4	2	38	25,3%	
5	3	21	14,0%	
6	4	12	8,0%	
7	5	5	3,3%	
8	6	2	1,3%	
9		150		

Kopioi kaava.

D
sf
25
72
110
131
143
148
150

Suhteellinen summafrekvenssi $sf\%$ saadaan myös laskemalla suhteellisia frekvenssejä $f\%$ yhteen.

Suhteellinen summafrekvenssijakauma $sf\%$ saadaan jakamalla summafrekvenssin arvo havaintoyksiköiden kokonaismäärällä eli luvulla 150.

	A	B	C	D	E
1	Käynti- kerrat	f	f%	sf	
2	0	25	16,7%	25	=D2/150
3	1	47	31,3%	72	
4	2	38	25,3%	110	
5	3	21	14,0%	131	
6	4	12	8,0%	143	
7	5	5	3,3%	148	
8	6	2	1,3%	150	
9		150			

Kopioi
kaava.

E
sf%
0,167
0,480
0,733
0,873
0,953
0,987
1,000

Muotoile
prosen-
teiksi.

E
sf%
16,7%
48,0%
73,3%
87,3%
95,3%
98,7%
100,0%

- b) Moodi voidaan lukea suoraan frekvenssijakaumasta sen arvon kohdalta, jolla on suurin frekvenssi.

Käyntikerrat	f
0	25
1	47
2	38
3	21
4	12
5	5
6	2

$M_o = \mathbf{1}$

Tavallisin elokuvissa käyntien määrä vuoden aikana oli **yksi**.

- c) Mediaani luetaan sen muuttujan arvon kohdalta, jonka suhteellinen summafrekvenssi ensimmäisen kerran saavuttaa arvon 50 %.

A	E
Käyntikerrat	sf%
0	16,7%
1	48,0%
2	73,3%
3	87,3%
4	95,3%
5	98,7%
6	100,0%

Md = 2

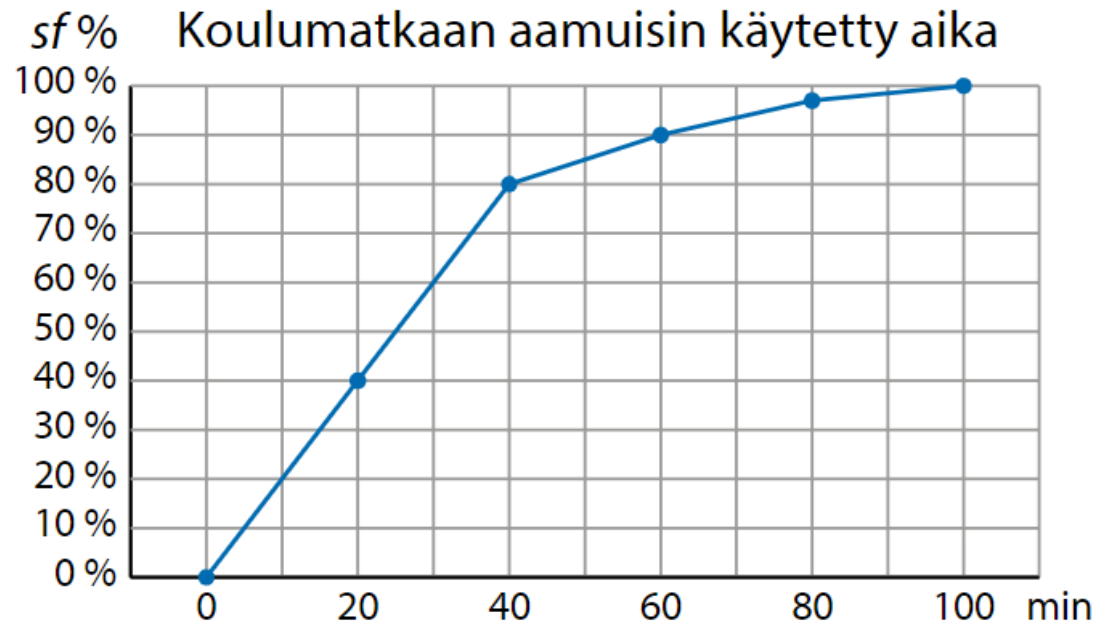
**Tee
tehtävät
2.2, 2.3 ja
2.8**

Puolet vastaajista oli kuluneen vuoden aikana käynyt elokuvissa korkeintaan **kaksi** kertaa ja puolet vastaajista vähintään **kaksi** kertaa.

A Esimerkki 4

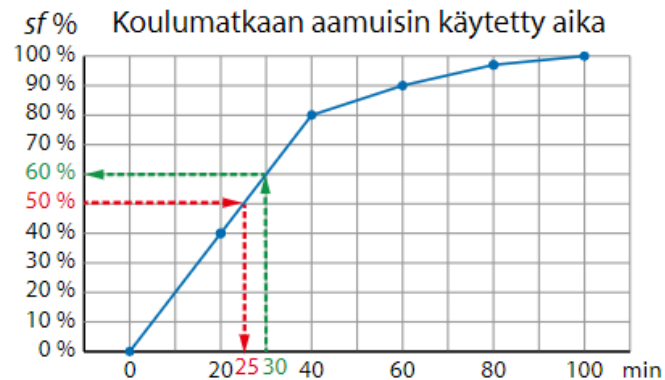
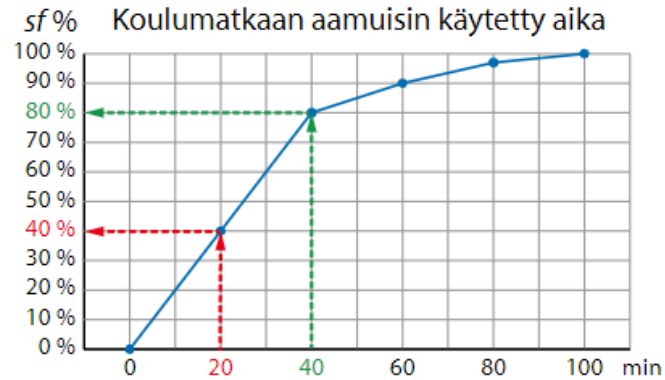
Oppilaitos järjesti opiskelijoilleen kyselyn, jossa selvitettiin koulumatkaan aamuisin käytettyä aikaa. Tulosten perusteella piirrettiin oheinen kuvaaja.

- a) Kuinka monta prosenttia opiskelijoista käyttää aamuisin koulumatkaan korkeintaan 20 minuuttia?
- b) Kuinka monta prosenttia opiskelijoista käyttää aamuisin koulumatkaan yli 40 minuuttia?
- c) Mikä on koulumatkan mediaani?
- d) Kyselyyn vastasi 200 opiskelijaa. Kuinka moni vastaajista käytti koulumatkaan aikaa korkeintaan puoli tuntia?



Kuvaajaa, joka havainnollistaa summafrekvenssiä, kutsutaan **summakäyräksi** eli **kertymäkuvaajaksi**.

Vaaka-akselilla ovat muuttujan arvot. Pystyakselilla ovat lukumäärät (*sf*) tai prosenttiluvut (*sf* %).



- d) Kyselyyn vastasi 200 opiskelijaa. Kuinka moni vastaajista käytti koulumatkaan aikaa korkeintaan puoli tuntia?

Ratkaisu

Diagrammin pystyakselilta nähdään, että se kuvaa suhteellista summafrekvenssijakaumaa (*sf* %) eli tiettyyn muuttujan arvoon mennessä kertyneiden havaintoyksiköiden määrää prosentteina.

- a) Koulumatkaan aamuisin käyttää korkeintaan **20** minuuttia **40 %** opiskelijoista.
- b) Koulumatkaan käyttää korkeintaan **40** minuuttia **80 %** opiskelijoista. Yli 40 minuuttia koulumatkaan käyttää $100 \% - 80 \% = 20 \%$ opiskelijoista.
- c) Mediaani luetaan sen muuttujan arvon kohdalta, jonka suhteellinen summafrekvenssi on 50 %.

$M_d = 25$ eli 50 % opiskelijoista käyttää koulumatkaan aikaa korkeintaan **25** minuuttia.

- d) Koulumatkaan kuluu korkeintaan **30** minuuttia **60 %**:lla opiskelijoista. Lasketaan, kuinka paljon on 60 % kyselyyn vastanneista 200 henkilöä.

$$0,6 \cdot 200 = 120$$

Kyselyyn vastanneista opiskelijoista 120 opiskelijalla kului koulumatkaan korkeintaan puoli tuntia.

**Tee
tehtävät
2.10 ja 2.18**