MAB05

Tilastot ja todennäköisyys

Iitin lukio

2023-2024

Aleksi Alenius

Sisältö

Tilaston kuvaaminen

Peruskäsitteitä

Havaintoaineisto

Määritelmä

- Perusjoukko on tutkimuksen kohteena oleva ryhmä.
- Havainto (tai havaintoarvo) on yksittäinen vastaus.

Esimerkki

Tutkitaan luokan oppilaiden pituuksia ja vaatekokoja. Tällöin perusjoukkona on luokka. Havaintoina ovat mittaus- ja kyselytulokset.

Tilastomuuttuja

Määritelmä

Tilastomuuttuja on tutkittava ominaisuus, joka voi olla:

- Kvantitatiivinen (luku)
  - erilliset arvot $\implies$ diskreetti
  - jatkuva
- Kvalitatiivinen (laatu)

Esimerkki

Pituus on jatkuva kvantitatiivinen muuttuja, kun taas vaatekoko on diskreetti. Esimerkiksi vaatteen merkki on kvalitatiivinen muuttuja.

Taulukointi (frekvenssijakauma)

Määritelmä (Frekvenssi)

Frekvenssi eli esiintymistiheys on havainnon esiintymiskertojen lukumäärä.

Esimerkki

20 opiskelijan luokalta kysytään heidän sisarustensa lukumäärää. Tulokset kootaan alla olevaan taulukkoon.

Sisarusta	f
0	4
1	7
2	8
3	1

Suhteellinen rekvenssi

Määritelmä (Suhteellinen frekvenssi)

Suhteellinen frekvenssi $f\%$ kertoo havainnon esiintymiskertojen %-osuuden kaikista tapahtumista.

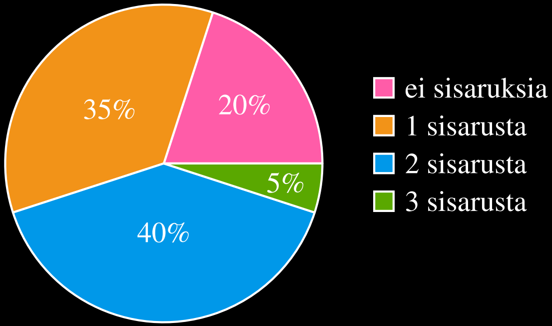
Esimerkki

20 opiskelijan luokalta kysytään heidän sisarustensa lukumäärää. Tulokset kootaan alla olevaan taulukkoon.

Sisarusta	f	$f\%$
0	4	20%
1	7	35%
2	8	40%
3	1	5%

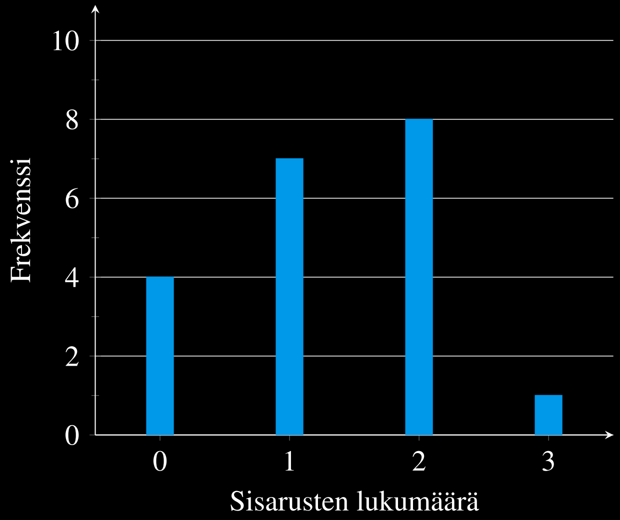
Ympyräkaavio

Ympyräkaaviosta eli sektoridiagrammista voi helposti arvioida muuttujien arvojen suhteellisia osuuksia.



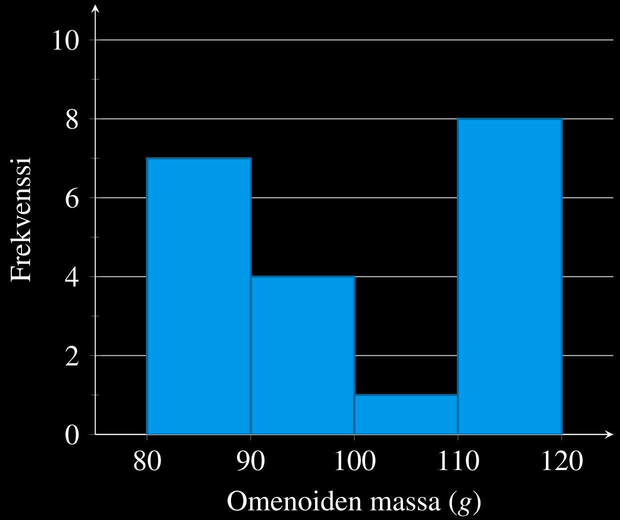
Pylväskuvaaja

Diskreettiä muuttujaa voi kuvata pylväsdiagrammilla.



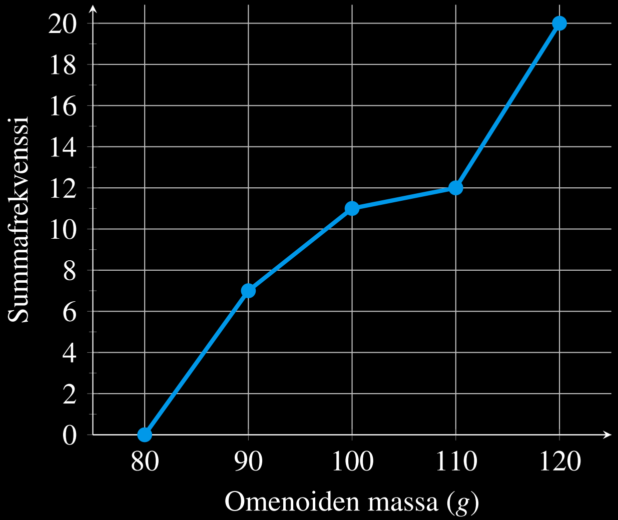
Histogrammi

Histogrammia käytetään vastaavasti luokitettaessa jatkuvia muuttujia.



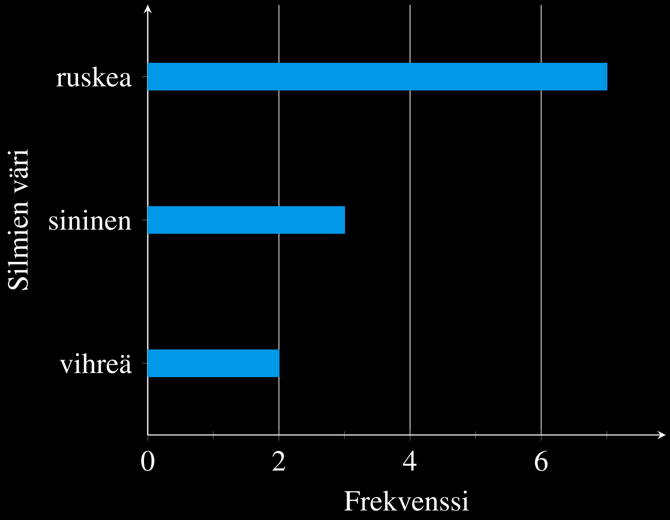
Viivadiagrammi

Viivadiagrammilla voidaan havainnollistaa muuttujan arvojen kehittymistä.



Vaakapylväät

Vaakapylväitä käytetään, kun muuttujan arvoilla ei ole luontevaa järjestystä.



Frekvenssijakauma

Summafrekvenssi

Määritelmä (Summafrekvenssi)

Summafrekvenssi sf eli kertyvä esiintymistiheys on yksittäisten havaintojen yhteenlaskettu arvo.

Esimerkki

20 opiskelijan luokalta kysytään heidän sisarustensa lukumäärää. Tulokset kootaan alla olevaan taulukkoon.

Sisarusta	f	$f\%$	sf
0	4	20%	4
1	7	35%	11
2	8	40%	19
3	1	5%	20

Suhteellinen summafrekvenssi

Määritelmä (Suhteellinen summafrekvenssi)

Suhteellinen summafrekvenssi $sf\%$ on summafrekvenssin sf osuus prosentteina.

Esimerkki

20 opiskelijan luokalta kysytään heidän sisarustensa lukumäärää. Tulokset kootaan alla olevaan taulukkoon.

Sisarusta	f	$f\%$	sf	$sf\%$
0	4	20%	4	20%
1	7	35%	11	55%
2	8	40%	19	95%
3	1	5%	20	100%

Tunnuslukuja

Määritelmä (Moodi)

Moodi Mo on aineistossa eniten esiintynyt havainto.

Määritelmä (Mediaani)

Mediaani Md on suuruusjärjestykseen lajitellun jakauman keskimmäinen luku, joka jakaa aineiston kahteen yhtä suureen osaan.

Esimerkki

Esimerkin taulukon perusteella sisarustutkimuksen moodi ja mediaani ovat

$$Mo = 2$$
$$Md = 1.$$

Sisarusta	f	$f\%$	sf	$sf\%$
0	4	20%	4	20%
1	7	35%	11	55%
2	8	40%	19	95%
3	1	5%	20	100%

Diskreetin muuttujan tilasto

Havaintoaineisto

Raakaa havaintoaineistoa voi olla hankala lukea sellaisenaan.

Esimerkki

Alla on taulukoitu joidenkin Suomen kaupunkien asukaslukuja (Tilastokeskus, 2022).

Kaupunki	**Asukasluku**
Kouvola	79 434
Lahti	120 200
Kotka	50 640
Hamina	19 549
Lappeenranta	72 675
Mikkeli	51 980

Lajittelu

Taulukoitua havaintoaineistoa voi lajitella ohjelmistolla.

Esimerkki

Aiempi asukaslukutaulukko lajiteltuna asukasluvun perusteella laskevassa järjestyksessä:

Kaupunki	**Asukasluku**
Lahti	120 200
Kouvola	79 434
Lappeenranta	72 675
Mikkeli	51 980
Kotka	50 640
Hamina	19 549

Tilastokuvaaja

Kuvaaja on frekvenssitaulukkoa parempi esitystapa, jonka malli valitaan käyttötarkoituksen mukaan.

- Pylväskaavio
  - frekvenssien suora vertailu
  - vaakapylväät, jos muuttujan arvoilla ei ole luontevaa järjestystä
- Ympyräkaavio
  - korostaa havaintojen osuuksia
- Viivakaavio
  - havaintojen kehittyminen esim. ajan suhteen

Luokittelu

Omenoiden punnintaa

Esimerkki

Tuotannon tarkastelun yhteydessä punnittiin kahdenkymmenen omenan massat (g). Tulokset on lueteltu alla.

119,1	117,0	90,2	110,3
102,4	80,9	112,7	92,3
94,0	113,5	118,9	88,9
82,0	113,7	88,7	88,6
82,1	79,5	94,9	109,8

Miten tällaista havaintoaineistoa kannattaa kuvata?

Selvästikään ei kannata kuvata yksittäisten arvojen frekvenssejä.

$\Longrightarrow$ Jaetaan vaihteluväli pienempiin osiin eli **luokkiin**!

Luokkien muodostaminen

Esimerkki

119,1	117,0	90,2	110,3
102,4	80,9	112,7	92,3
94,0	113,5	118,9	88,9
82,0	113,7	88,7	88,6
82,1	79,5	94,9	109,8

Huomataan, että arvoista pienin on $79,5$ ja suurin $119,1$ grammaa. Vaihteluvälin pituus on tällöin

$$119,1 - 79,5 = 39,6,$$

joka on hieman epämukava luku käyttää. Sen sijaan 40 on hyvin lähellä ja jakautuu siististi neljään luokkaan. Muodostetaan nämä neljä luokkaa pyöristystääntöjä käyttäen, jolloin luokat ovat:

$$80 - 89, \quad 90 - 99, \quad 100 - 109 \quad \text{ja } 110 - 119$$

Luokitettu frekvenssijakauma

Esimerkki

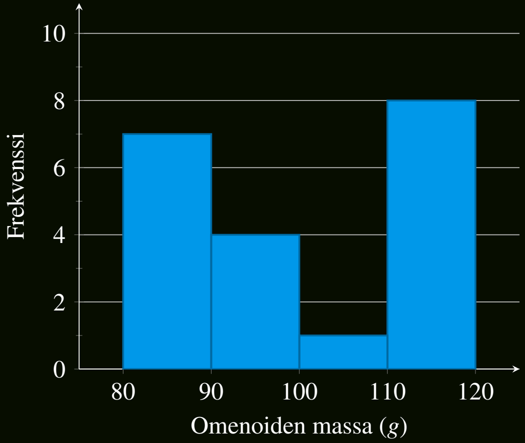
Nyt luokille voidaan muodostaa frekvenssijakauma

Massa (g)	f
$80 - 89$	7
$90 - 99$	4
$100 - 109$	1
$110 - 119$	8

jossa esimerkiksi luokan $80 - 89$ todelliset rajat ovat $79,5$ ja $89,5$. Vastaavasti luokan $80 - 89$ luokkakeskus on

$$\frac{89,5 - 79,5}{2} = 84,5$$

Saaduista frekvensseistä voidaan muodostaa histogrammi:



Tiivistelmä

Luokittelu lyhyesti

1. Etsitään havaintoaineiston arvojen pienin ja suurin arvo.
2. Määritellään näiden arvojen avulla sopiva vaihteluväli ala- ja ylärajoineen.
3. Jaetaan vaihteluväli yhtä suuriin osiin (luokkiin)
4. Muodostetaan luokat ja niiden avulla frekvenssijakauma.

Tilaston analysoiminen

Tunnuslukuja

Kertausta

Määritetään 20 aiemmin punnitun omenan massaluokkien moodi ja mediaani, kun

Massa (g)	f
80 – 89	7
90 – 99	4
100 – 109	1
110 – 119	8

Moodi on aineiston yleisin havainto, joten

$$Mo = 110 - 119.$$

Mediaani jakaa aineiston kahteen yhtä suureen osaan. Omenoita oli 20, joten puolivälinä on omena 10. Selvästi se kuuluu luokkaan 90 – 99, joten

$$Md = 90 - 99$$

Keskiarvo

Moodi ja mediaani ovat tunnuslukuja, joiden joukkoon luetaan myös keskiarvo.

Määritelmä (keskiarvo)

Keskiarvo $\overline{x}$ on havaintoarvojen summan $\sum x$ ja lukumäärän n osamäärä eli:

$$\overline{x} = \frac{\sum x}{n}.$$

Esimerkki

Aiemmin tarkastellun sisarustutkimuksen havaintojen keskiarvo on:

$$\overline{x} = \frac{0 \cdot 4 + 1 \cdot 7 + 2 \cdot 8 + 3 \cdot 1}{4 + 7 + 8 + 1} = \frac{26}{20} = 1{,}3$$

Sisarusta	f
0	4
1	7
2	8
3	1

Luokitetun jakauman keskiarvo

Esimerkki

Luokitetun jakauman keskiarvon laskemiseen on käytettävä luokkien luokkakeskuksia.

Massa (g)	**Luokkakeskus**	f
80-89	84,5	7
90-99	94,5	4
100-109	104,5	1
110-119	114,5	8

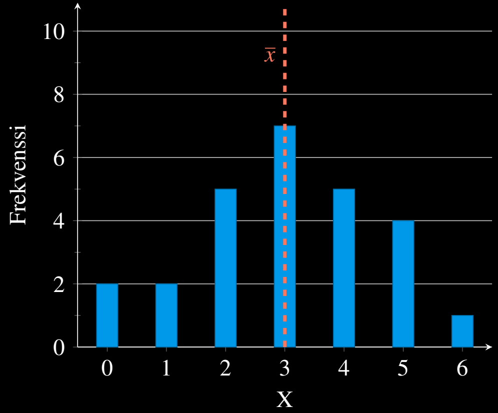
$$\bar{x} = \frac{84,5\cdot 7 + 94,5\cdot 4 + 104,5\cdot 1 + 114,5\cdot 8}{7+4+1+8} = \frac{1990}{20} = 99,5,$$

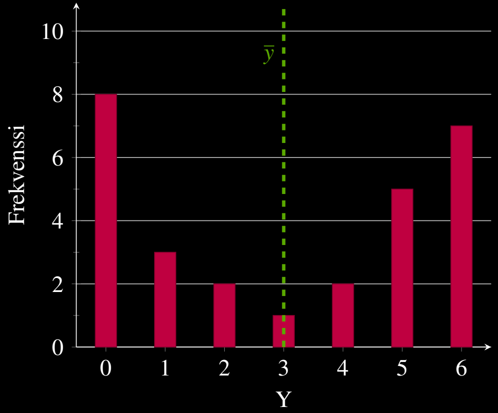
joten omenoiden massojen keskiarvo oli noin 99,5 grammaa.

Yhden muuttujan tilasto

Vertailu

Vertaa tilastomuuttujien X ja Y keskiarvoja Geogebralla. Ne ovat likimain samat. Mitä eroa jakaumilla on?





Hajontaluvut

Määritelmä

Satunnaismuuttujan X keskihajonta

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}},$$

missä n on havaintojen lukumäärä.

Määritelmä

Satunnaismuuttujan X otoskeskihajonta on

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}},$$

Huom.

Kirjan tehtävissä keskihajonnalla tarkoitetaan nimenomaan otoskeskihajontaa s!

Poikkeava havainto

Sivussa on arvosanojen frekvenssijakauma, jonka avulla voidaan havaita, että

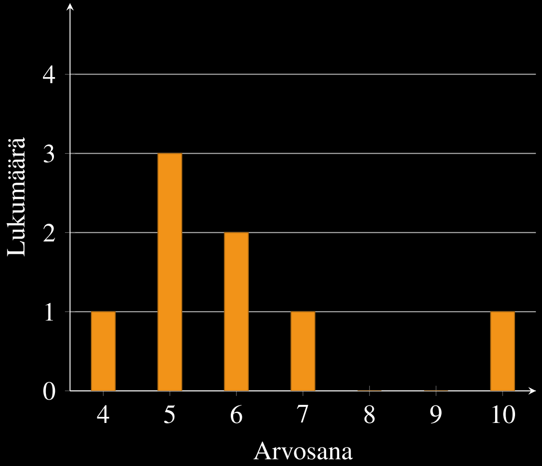
$$Md_0 = 5$$
$$\bar{x}_0 \approx 5,4$$
$$s_0 \approx 0,98$$

Mitä tapahtuu, jos havaintoaineistoon lisätään yksi 10?

$$Md = 5$$
$$\bar{x} = 6$$
$$s \approx 1,85$$

Määritelmä

Havaintoa sanotaan poikkeavaksi, jos sen etäisyys keskiarvosta on yli kaksi keskihajontaa.



Kahden muuttujan riippuvuus

Muuttujien riippuvuus

Usein tutkitaan kahta (tai useampaa) tilastomuuttujaa ja halutaan selvittää, onko niiden välillä yhteys.

- Pituus ja paino
- Tupakointi ja syöpäriski
- Asunnon ikä ja hinta
- Liikunnan määrä ja onnellisuus (Harvard Study of Adult Development, $\approx$ 80 vuotta)

Muuttujia tutkittaessa on muistettava seuraava

Riippuvuus ei aina tarkoita aiheuttamista, mutta aiheuttaminen tarkoittaa aina riippuvuutta.

Esimerkki

Kaikki vedenjuojat kuolevat.

Regressiosuora

Määritelmä

Regressiosuora on havaintoarvoparin (x, y) pistejoukkoon sovitettu suora.

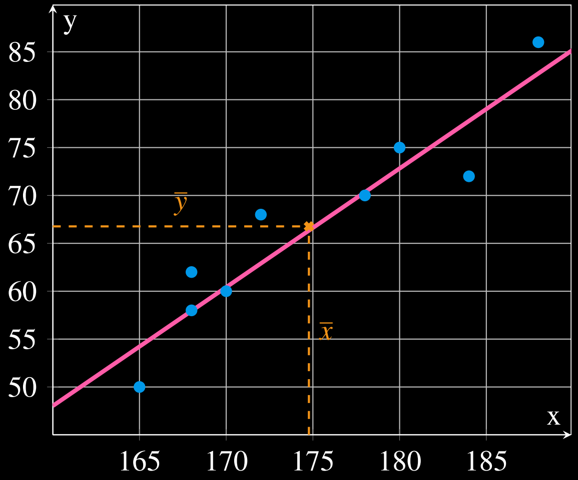
Seuraus

- *Regressiosuoran ylä- ja alapuolella on suunnilleen yhtä monta havaintopistettä.*
- *Regressiosuora kulkee aina muuttujien keskiarvopisteen $(\overline{x}, \overline{y})$ kautta*

Pituus ja paino

Muodostetaan regressiosuora seuraaville havaintopareille.

Pituus (cm)	Paino (kg)
165	50
168	62
168	58
170	60
172	68
178	79
180	75
184	72
188	86



Geogebrasta: $y = 1.2987x - 159,2073$

Korrelaatio

Määritelmä (Korrelaatiokerroin)

Korrelaatiokerroin r kertoo, kuinka voimakkaasti kaksi lineaarisesti riippuvaa muuttujaa riippuvat toisistaan.

- $-1 < r < 1$
- Mitä lähempänä r on arvoa 1 tai -1, sitä voimakkaampi lineaarinen riippuvuus.

Esimerkki

Geogebrasta voidaan lukea, että aiemman pituus-paino -esimerkin korrelaatio oli

$$r = 0.9172,$$

joten pituus ja paino korreloivat positiivisesti eli suora on kasvava.
Vastaavasti:

- Jos $r < 0$, niin suora on laskeva.
- Jos $r \approx 0$, niin lineaarista riippuvuutta ei ole.

Riippuvuuden tulkinta

Korrelaatiokerrointa tulkitaan asteikolla merkityksetön-voimakas:

- Merkityksetön, jos $0 \leq |r| \leq 0,3$
- Kohtalainen, jos $0,3 < |r| \leq 0,6$
- Huomattava, jos $0,6 < |r| \leq 0,8$
- Voimakas, jos $|r| \geq 0,8$

Määritelmä (Selitysaste)

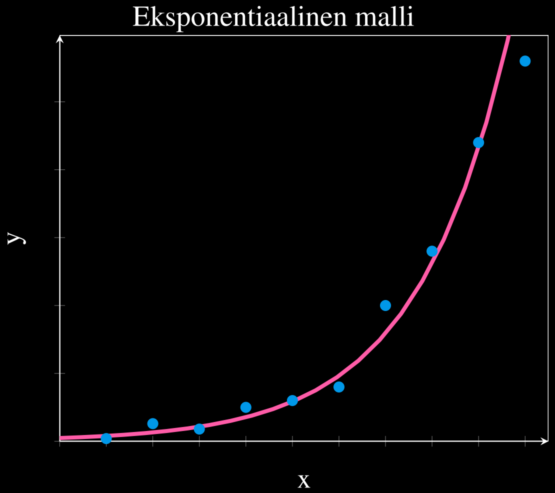
Korrelaatiokertoimen neliötä r^2 kutsutaan selitysasteeksi. Se kertoo, kuinka suuren osan kahden muuttujan korrelaatiosta voi selittää lineaarisella riippuvuudella.

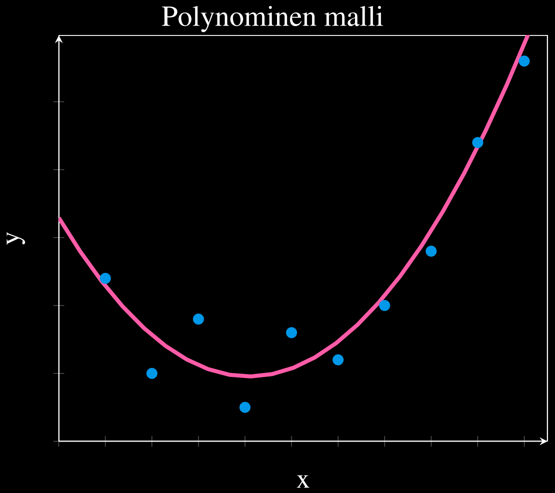
Esimerkki

Verrataan tulkintarajojen selitysasteita:

$$0,3^2 = 0,09 = 9\%$$
$$0,6^2 = 0,36 = 36\%$$
$$0,8^2 = 0,64 = 64\%$$

Epälineaarisia regressiomalleja





Todennäköisyys

Tilastoista todennäköisyyteen

Sisaruksia

Esimerkki

Palautetaan mieleen aiemman sisarustutkimuksen frekvenssijakauma. Jos tutkitusta joukosta valitaan sattumanvaraisesti yksi, niin millä todennäköisyydellä kyseisellä henkilöllä

Sisarusta	**f**	**f%**	**sf%**	**sf%**
0	4	20 %	4	20 %
1	7	35 %	11	55 %
2	8	40 %	19	95 %
3	1	5 %	20	100 %

- ei ole sisaruksia?

$$P(\text{ei sisaruksia}) = 20\%$$

- on sisaruksia?

$$P(\text{ainakin 1 sisarus}) = 100\% - 20\% = 80\%$$

- on ainakin 2 sisarusta?

$$P(\text{ainakin 2 sisarusta}) = 40\% + 5\% = 45\%$$

Tilastollinen todennäköisyys

Määritelmä

Havaintoaineiston tapahtuman A todennäköisyys on

$$P(A) = \frac{\text{esiintymiskertojen määrä}}{\text{kaikkien havaintojen määrä}}$$

Muistetaan, että $0,01 = 1\%$ ja $1 = 100\%$.

Seuraus

Tapahtuman A tilastolliselle todennäkäisyydelle pätee aina

$$0 \leq P(A) \leq 1$$

Seuraus

Tapahtuman A vastatapahtuman $\overline{A}$ todennäköisyys on

$$P(\overline{A}) = 1 - P(A)$$

Klassinen todennäköisyys

Klassinen todennäköisyys

Määritelmä

Satunnaisilmiön tulosvaihtoehtoja kutsutaan alkeistapauksiksi. Tapahtuman A tapahtumisen todennäköisyys $P(A)$ on

$$P(A) = \frac{\text{suotuisten alkeistapausten määrä}}{\text{kaikkien alkeistapausten määrä}}$$

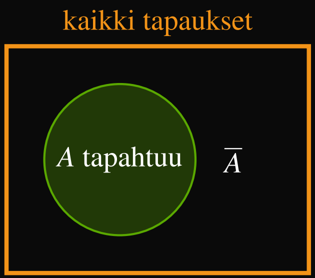
Esimerkki

- Kolikkoa heitettäessä alkeistapauksina ovat kruuna ja klaava.
- Noppaa heitettäessä alkeistapauksia ovat silmäluvut 1, 2, 3, 4, 5 ja 6.
- Nostettaessa korttia pakasta alkeistapaukset koostuvat jokaisesta yksittäisestä kortista (yht. 52 vaihtoehtoa).

Komplementtisääntö

Määritelmä (Vastatapahtuma)

Tapahtuman A vastatapahtuman eli komplementin $\overline{A}$ suotuisia alkeistapaukset ovat kaikki ne, jotka eivät ole suotuisia tapahtumalle A.



Seuraus (Komplementtisääntö)

Tapahtuman A vastatapahtuman $\overline{A}$ todennäköisyys on

$$P(\overline{A}) = 1 - P(A) \iff P(A) + P(\overline{A}) = 1$$

Malleja

Luettelo

Todennäköisyyksiä laskiessa on tarpeen selvittää sekä kaikkien alkeistapausten että suotuisten alkeistapausten lukumäärä. Yksinkertaisin tapa on luetteleminen.

Esimerkki

Kolikkoa heitetään neljä kertaa. Luetellaan kaikki mahdolliset tulokset. Merkitään $R :=$ kruuna ja $L :=$ klaava.

RRRR	LRRR	RLLR	LRLL
RRRL	LRRL	RLRL	LLRL
RRLR	LRLR	RRLL	LLLR
RLRR	LLRR	RLLL	LLLL

Huomataan, että alkeistapauksia on yhteensä 16.

Taulukointi

Esimerkki

Taulukoidaan aiemmat kolikoiden heittojen tulokset kruunien lukumäärän mukaan.

Ei kruunia	**1 kruuna**	**2 kruunaa**	**3 kruunaa**	**Kaikki kruunia**
LLLL	RLLL	RRLL	RRRL	RRRR
	LRLL	RLRL	RRLR	
	LLRL	RLLR	RLRR	
	LLLR	LRRL	LRRR	
		LRLR		
		LLRR		

Nyt voidaan helpommin lukea, että esimerkiksi

$$P(\text{tasan 1 kruuna}) = \frac{4}{16} = \frac{1}{4} = 25\%$$

Venn-diagrammi

Joskus on mahdollista, että kaksi tapahtumaa voivat tapahtua yhtä aikaa.

Esimerkki

Vuosiluokalla on yhteensä 30 oppilasta, joista 5 lukee vain ranskaa, 6 vain saksaa ja 3 sekä ranskaa että saksaa.

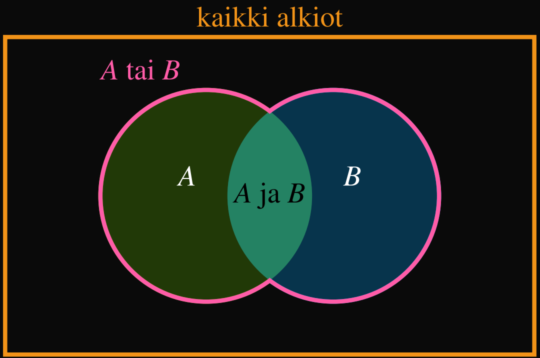
Tilanteesta voidaan piirtää oheinen Venn-diagrammi, josta on helppo lukea, että

- ranskaa lukee $5 + 3 = 8$ oppilasta.
- Kumpaakaan ei lue $30 - 5 - 6 - 3 = 16$ oppilasta.

Venn-diagrammi

Määritelmä

Venn-diagrammi on tiedonkuvaustapa, jossa suhteita eri joukkojen välillä kuvataan toisiaan leikkaavilla ympyröillä. Olkoot A ja B joukkoja, jolloin niitä kuvaava Venn-diagrammi olisi seuraava:



Koordinaatisto

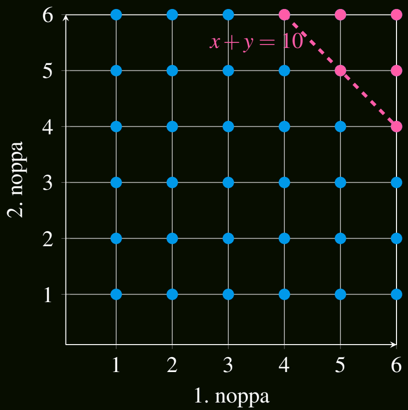
Jos yksi alkeistapaus koostuu kahdesta toisistaan riippumattomasta arvosta, niin tapauksia voi kuvata koordinaatiston avulla.

Esimerkki

Heitetään kahta noppaa. Millä todennäköisyydellä noppien summaksi tulee vähintään 10?

Merkitään alkeistapaukset koordinaatistoon. Nyt esimerkiksi piste $(1,2)$ kuvaa tapausta, jossa 1. nopasta tuli silmäluvuksi 1 ja 2. nopasta 2. Tapauksia on $6 \cdot 6 = 36$.
Tapauksia joiden koordinaattien summa on ainakin 10 löytyy 6 kappaletta, joten

$$P(X \geq 10) = \frac{6}{36} = \frac{1}{6} \approx 17\%$$



Todennäköisyyden laskusääntöjä

Kertolaskusääntö

Riippumattomuus

Määritelmä

Tapahtumia A ja B kutsutaan riippumattomiksi, jos yhden tapahtuminen ei vaikuta toisen tapahtumisen todennäköisyyteen. Tällöin ne toteuttavat kertolaskusäännön

$$P(A \text{ ja } B) = P(A) \cdot P(B)$$

Esimerkki

Heitetään kahta noppaa yhtä aikaa. Selvästi eri noppien silmälukemat ovat toisistaan riippumattomat. Lisäksi $P(6) = \frac{1}{6}$, joten todennäköisyys heittää kaksi kuutosta on

$$P(6 \text{ ja } 6) = P(6) \cdot P(6) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Riippuvat tapahtuvat

Esimerkki

Tavallisessa pelikorttipakassa on 52 korttia, jotka on jaettu tasan 4 maahan. Millä todennäköisyydellä kaksi peräkkäin nostettua korttia ovat molemmat herttoja?
Yhteen maahan kuuluu $\frac{52}{4} = 13$ korttia, joten

$$P(1.\text{ on hertta}) = \frac{13}{52} = \frac{1}{4}$$

Jos pakasta on jo nostettu yksi kortti, niin kortteja on jäljellä 51. Koska nostettu kortti oli hertta, niitä on jäljellä 12. Tällöin

$$P(2.\text{ on hertta}) = \frac{12}{51}$$

Nyt voidaan käyttää kertolaskusääntöä, jolloin

$$P(\text{kaksi herttaa}) = \frac{1}{4} \cdot \frac{12}{51} = \frac{3}{51} \approx 0{,}059 = 5{,}9\%$$

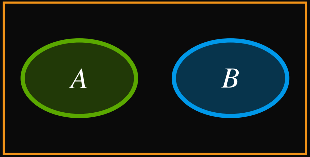
Yhteenlaskusääntö

Erilliset tapahtumat

Määritelmä

Jos tapahtumat A ja B eivät voi tapahtua yhtä aikaa, niitä kutsutaan erillisiksi ja tällöin

$$P(A \text{ tai } B) = P(A) + P(B)$$



Esimerkki

Korttipakasta nostetaan yksi kortti. Millä todennäköisyydellä kortti on joko pata tai risti?
Pakassa on 52 korttia, joista 13 on patoja ja 13 ristejä. Kortti ei voi olla sekä pata että risti, joten

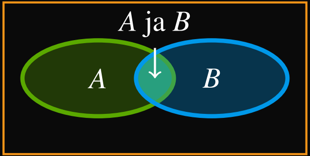
$$P(\text{Pata tai risti}) = P(\text{Pata}) + P(\text{Risti}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = 50\%$$

Yleinen tapaus

Määritelmä

Jos tapahtumat A ja B voivat tapahtua yhtä aikaa, niin

$$P(A \text{ tai } B) = P(A) + P(B) - P(A \text{ ja } B)$$



Esimerkki

Pakasta nostetaan kortti. Millä todennäköisyydellä se on pata tai ässä?
Pakassa on 52 korttia, joista 13 on patoja, 4 ässiä ja vain yksi pataässä. Yhteenlaskusäännön mukaan

$$P(\text{Pata tai ässä}) = P(\text{Pata}) + P(\text{Ässä}) - P(\text{Pataässä}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13} \approx 30{,}8\%$$

Tuloperiaate ja kombinaatiot

Tuloperiaate

Lause

Jos tilanne on jaettu eri järjestyksessä tapahtuviin vaiheisiin A_1 ja A_2, joilla on n_1 ja n_2 vaihtoehtoa, niin mahdollisten eri tilanteiden lukumäärä on

$$n_1 \cdot n_2$$

Esimerkki

Ravintolan ruokalistalla on 4 alkuruokaa, 3 pääruokaa ja 5 jälkiruokaa. Kuinka monta erilaista ateriakokonaisuutta vaihtoehdoista voi muodostaa?
Ateriakokonaisuuksia on yhteensä

$$4 \cdot 3 \cdot 5 = 60$$

Jonotusta

Esimerkki

Kurssilla on 10 opiskelijaa. Kuinka moneen järjestykseltään ainutlaatuiseen jonoon heidät voi järjestää?

Paikka	1	2	3	4	5	6	7	8	9	10
Vaihtoehtoja	10	9	8	7	6	5	4	3	2	1

Nyt voidaan käyttää tuloperiaatetta, jolloin mahdollisia jonoja on

$$10 \cdot 9 \cdot \ldots \cdot 2 \cdot 1 = 3\,628\,800$$

Kertoma

Määritelmä

Luonnollisen luvun n kertoma $n!$ on

$$n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$$

Kertoma kuvaa, kuinka moneen järjestykseltään ainutlaatuiseen jonoon n alkiota voidaan järjestää.

Seuraus

$$0! = 1,$$

koska 0 *alkiota voi jakaa vain yhteen järjestykseen.*

Ryhmäytymistä

Esimerkki

Kurssin 10 opiskelijasta muodostetaan sattumanvarainen 3 henkilön ryhmä. Montako ainutlaatuista ryhmää voidaan muodostaa?
Jos kurssilaisista muodostettaisiin jono, niin jonon kolme ensimmäistä henkilöä olisivat ryhmän jäseniä. Tällöin eri 3 hengen jonoja on

$$10 \cdot 9 \cdot 8 = 720$$

Järjestyksellä ei kuitenkaan ole väliä ryhmän kannalta. Kuinka moneen eri järjestykseen voidaan 3 henkilöä jakaa?

$$3! = 3 \cdot 2 \cdot 1 = 6,$$

joten ainutlaatuisia ryhmiä on

$$\frac{720}{6} = 120$$

Kombinaatio

Määritelmä

Kombinaatiot ovat erilaisia ainutlaatuisia joukkoja, jotka voi muodostaa isomman joukon alkioista.

Lause

Olkoot joukon A alkioiden lukumäärä n. Joukon A alkioista muodostettavien k alkioita sisältävien joukkojen lukumäärä saadaan binomikertoimella

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Luetaan "n yli k:n."

Huom.

Binomikertoimen kaavaa ei kurssilla tarvitse muistaa ulkoa. Laskimessa binomikerrointa vastaa komento "nCr". Käyttö:

$$nCr = \binom{n}{r} \text{ eli esim. } 3C2 = \binom{3}{2} = 3$$