

Tilastot, ennakkojäsentäjä.

Tilastotiede on tieteenala, joka tutkii, miten tietoa voidaan kerätä, käsitellä ja tulkita. Se auttaa meitä ymmärtämään suuria tietomääriä ja tekemään johtopäätöksiä. Yhdeksäsluokkalaisille tilastotieteessä keskeisiä asioita ovat tunnusluvut ja kaaviot.

1. Tunnusluvut

Tunnusluvut ovat tilastollisia lukuja, jotka kuvaavat aineiston ominaisuuksia. Tärkeimpiä tunnuslukuja ovat:

1. Keskilukuja:

- a. **Keskiarvo:** Lasketaan lisäämällä kaikki havainnot yhteen ja jakamalla summa havaintojen määrällä. Esimerkiksi, jos kokeen pisteet ovat 8, 7, 9 ja 6, niin keskiarvo on $(8+7+9+6):4=7,5$.
- b. **Mediaani:** Aineiston keskimäinen arvo, kun havainnot ovat järjestyksessä. Jos aineistossa on pariton määrä havaintoja, mediaani on keskimäinen arvo. Jos havaintoja on parillinen määrä, mediaani on kahden keskimäisen arvon keskiarvo.
- c. **Moodi** eli tyypiarvo: Aineiston yleisin arvo. Jos vaikka kokeessa eniten saatu arvosana on 8, moodi on 8.

2. Hajontalukuja:

- a. **Vaihteluväli:** Suurimman ja pienimmän arvon erotus. Jos arvosanat ovat välillä 5 ja 10, vaihteluväli on $10-5=5$.
- b. **Keskihajonta:** Kuva, kuinka paljon havainnot keskimäärin poikkeavat keskiarvosta. Se kertoo, kuinka "levällään" aineisto on keskiarvon ympärillä. Jos arvot ovat lähellä toisiaan, keskihajonta on pieni, ja jos ne vaihtelevat paljon, keskihajonta on suurempi.

3. Kaaviot

Kaaviot ovat visuaalisia esityksiä, jotka helpottavat tietojen tulkitsemista ja vertailua. Yleisimpiä kaavioita ovat:

1. **Pylväsdiagrammi:** Käytetään, kun halutaan esittää eri luokkien välistä vertailua. Esimerkiksi eri luokkien kokeiden keskiarvot voidaan esittää pylväsdiagrammilla, jossa jokaisella luokalla on oma pylväs.
2. **Piirakkadiagrammi:** Käytetään, kun halutaan kuvata, kuinka suuri osa tietyistä kokonaisuudesta kukin osa muodostaa. Piirakkadiagrammi on erityisen hyvä prosenttijakaumien esittämiseen, esimerkiksi eri vastausvaihtoehtojen osuudet kyselyssä.
3. **Viivadiagrammi:** Hyödyllinen, kun halutaan esittää muutosta ajan kuluessa. Esimerkiksi oppilaan arvosanojen kehittymistä lukuvuoden aikana voi havainnollistaa viivadiagrammilla.
4. **Histogrammi:** Näyttää, miten havainnot jakautuvat eri arvoalueille. Tämä kaavio soveltuu erityisesti suuremmille aineistoille. Histogrammissa aineisto jaetaan ryhmiin (esim. pistemäärien alueisiin), ja pylväiden korkeus kertoo, kuinka monta havaintoa kuuluu kuhunkin ryhmään.

5. Frekvenssi ja sen käyttö

Frekvenssi tarkoittaa havaintojen lukumäärää eli sitä, kuinka monta kertaa jokin arvo esiintyy aineistossa. Se auttaa meitä hahmottamaan, kuinka yleisiä tietyt arvot tai arvovälit ovat, ja se on tärkeä osa tilastollista analyysia.

Frekvenssiä voidaan käyttää monissa tilanteissa:

- **Absoluuttinen frekvenssi** kertoo tarkasti, kuinka monta kertaa tietty arvo esiintyy aineistossa. Esimerkiksi, jos aineistossa on viisi oppilasta, joilla on arvosanana 8, niin arvosanan 8 absoluuttinen frekvenssi on 5.
- **Suhteellinen frekvenssi** taas kertoo, kuinka suuri osa kaikista havainnoista tietty arvo on prosentteina tai murtolukuna. Jos arvosana 8 esiintyy 20 oppilaan aineistossa viisi kertaa, niin suhteellinen frekvenssi on $5/20 = 0,25$ eli 25 %.

Frekvenssitaulukko

Frekvenssit voidaan esittää **frekvenssitaulukossa**, joka tekee aineiston tarkastelusta helpompaa. Frekvenssitaulukkoon merkitään jokainen arvo, sen absoluuttinen frekvenssi ja usein myös suhteellinen frekvenssi. Taulukko voi sisältää esimerkiksi arvosanojen esiintymistiheyksiä.

Esimerkki frekvenssitaulukosta, jossa esitetään erään luokan matematiikan kokeen arvosanojen frekvenssit:

Arvosana	Absoluuttinen frekvenssi	Suhteellinen frekvenssi
5	2	10 %
6	4	20 %
7	3	15 %
8	5	25 %
9	4	20 %
10	2	10 %

Tässä taulukossa voimme nähdä heti, kuinka monta kertaa kukin arvosana esiintyy ja kuinka suuri osuus se on koko aineistosta.

Frekvenssin käyttö kaavioissa

Frekvenssit voidaan esittää kaavioina, jotka helpottavat tietojen tulkintaa:

- **Pylväsdiagrammi:** Frekvenssit esitetään pylväinä, jolloin korkeampi pylväs tarkoittaa suurempaa frekvenssiä.

3. Otokoko ja otantatapa

Otoksen koko ja otantatapa vaikuttavat merkittävästi tilastojen luotettavuuteen. Tässä muutama keskeinen huomio:

- **Riittävä otoskoko:** Jos otos (valittu joukko) on liian pieni, tilastollisten johtopäätösten tekeminen on vaikeaa, koska pieni otos ei välttämättä kuvaa koko perusjoukkoa. Esimerkiksi 10 henkilön otos ei todennäköisesti kuvaa hyvin koko kaupungin väestöä, kun taas 1000 henkilön otos on todennäköisesti tarkempi.
- **Otantatapa:** Otoksen tulee edustaa koko tutkimusjoukkoa. **Satunnaisotanta** on yleinen tapa varmistaa, että jokaisella perusjoukon jäsenellä on yhtä suuri todennäköisyys tulla valituksi otokseen, jolloin tulokset ovat todennäköisesti luotettavampia. Jos esimerkiksi tutkitaan koulun oppilaiden keskiarvoja, otoksen tulee sisältää oppilaita eri luokilta ja taustoista.

4. Aineiston käsittely ja tulkinta

Aineiston käsittelyvaiheessa tilastojen luotettavuuteen vaikuttavat seuraavat asiat:

- **Virheiden tarkistaminen:** Aineistosta tulee tarkistaa mahdolliset poikkeavat arvot tai virheet ennen analyysin tekemistä, sillä ne voivat vääristää tuloksia.
- **Sopiva analyysimenetelmä:** Valittu analyysimenetelmä vaikuttaa lopputulokseen. Jos käytetään väärää analyysimenetelmää, tulokset eivät ole luotettavia. Esimerkiksi keskiarvon käyttö voi olla harhaanjohtavaa, jos aineistossa on suuri hajonta.
- **Tulkinnan puolueettomuus:** On tärkeää, että tuloksia tulkitaan ilman ennako-oletuksia tai puolueellisuutta. Tilastolliset tulokset voidaan tulkita eri tavoin, ja joskus saman aineiston pohjalta voidaan tehdä erilaisia johtopäätöksiä riippuen siitä, mitä näkökulmaa korostetaan.

5. Luotettavuuden parantaminen

Tilastojen luotettavuutta voidaan parantaa useilla tavoilla:

- **Kattava otos ja tarkka otantasuunnitelma:** Käyttämällä riittävän suurta ja edustavaa otosta voidaan parantaa tulosten yleistettävyyttä.
- **Tietojen tarkka kerääminen ja mittaaminen:** Käyttämällä tarkkoja ja toistettavia mittausmenetelmiä voidaan vähentää mittausvirheitä.
- **Useiden tutkimusmenetelmien käyttö:** Vahvistamalla tuloksia useiden eri tutkimusmenetelmien avulla (esimerkiksi kyselytutkimukset ja havainnointi) voidaan varmistua tulosten luotettavuudesta. Tilastojen luotettavuuden tarkastelu on keskeistä, koska väärät johtopäätökset voivat johtaa virheellisiin päätöksiin tai käsityksiin. Luotettavuuden arviointi auttaa saamaan mahdollisimman tarkkaa ja puolueetonta tietoa, jota voidaan käyttää päätöksenteon tukena.