# TERVE!

## Lukion todennäköisyyslaskennan sovellukset ajankohtaisissa terveysilmiöissä

Annina Ruokonen

Pro Gradu-tutkielma
Kesäkuu 2022

MATEMATIIKAN JA TILASTOTIETEEN LAITOS
TURUN YLIOPISTO

TURUN YLIOPISTO
Matematiikan ja tilastotieteen laitos

RUOKONEN, ANNINA: TERVE! Lukion todennäköisyyslaskennan sovellukset ajankohtaisissa terveysilmiöissä
Pro gradu -tutkielma, 53 s., 67 liites.
Matematiikka
Kesäkuu 2022
Tarkastajat: Dosentti Petteri Harjulehto & Professori Peter Hästö

---

Tämän tutkielman tarkoitus on luoda terveysaiheinen tehtäväpaketti lukion pitkän matematiikan Tilastot ja todennäköisyys -opintojaksolle. Opetussuunnitelman perusteiden ja ajan didaktisen keskustelun hengessä tutkielman tehtävät ylittävät oppiaineiden välisiä kuiluja ja tuovat matematiikkaa lähemmäksi opiskelijan kokemusmaailmaa sekä ajankohtaisia yhteiskunnallisia ilmiöitä. Tehtävillä on kolme päätavoitetta: 1) lisätä opiskelijan ymmärrystä siitä, mitä kaikkea matematiikalla voidaan tehdä eli mihin opittua voidaan soveltaa ympäröivässä maailmassa, 2) innostaa opiskelijaa matematiikan opinnoissa sekä 3) tukea opiskelijaa lukion opetussuunnitelman yleisten ja matematiikan tavoitteiden, kuten ymmärtämisen, ongelmanratkaisun ja opitun soveltamisen toteutumisessa. Tutkielman tuloksena kehitetään ajankohtaisiin terveysaiheisiin pureutuva *Terve! Terveysaiheinen tehtäväpaketti opiskelijalle* –oppimateriaali, jonka matematiikanopettaja voi jakaa opiskelijoille käyttöön kurssikirjaa täydentämään tai josta hän voi poimia yksittäisiä esimerkkitehtäviä koti- tai tuntityöskentelyn tueksi.

Tutkielma on jaettu teoriaosioon sekä *Terve!*-materiaaliin, jonka opiskelija- ja opettajaversiot löytyvät tutkielman liitteistä. Tutkielman teoriaosio koostuu kahdesta luvusta, *Tilastot* ja *Todennäköisyyslaskenta*, joissa esitellään Tilastot ja todennäköisyys -opintojakson määritelmät ja tulokset todistuksineen. Näiden tulosten ja määritelmien esitysjärjestys noudattaa lukion opintojakson sekä useiden kurssikirjojen etenemistapaa. Terve!-opiskelijamateriaali koostuu 12:sta terveysaiheisesta tehtävästä ja sanallisista vihjeistä niiden ratkaisemiseksi. Terve!-opettajamateriaali on opiskelijamateriaalin kaltainen, mutta siinä esitetään eräät malliratkaisut tehtäville. Jokaisen tehtävän yhteydessä on kerrottu myös asiatietoa aiheesta sekä koottu linkkejä lisätietoa kaipaavalle opiskelijalle. Materiaalin tehtävät ovat sekalaisessa järjestyksessä. Näin opiskelija pääsee ratkaistavan ongelman määrittelystä ja sopivan matemaattisen ratkaisumenetelmän valinnasta alkaen kehittämään omaa ymmärrystä, kriittistä arviointia sekä ongelmanratkaisu- ja soveltamistaitoja.

Asiasanat: todennäköisyyslaskenta, tilastot, terveystieto, terveysilmiöt, oppimateriaali, kokemusmaailma, lukio

UNIVERSITY OF TURKU
Department of Mathematics and Statistics

RUOKONEN, ANNINA: TERVE! Health themed probability calculus applications to upper secondary school
Master's thesis, 53 p., 67 app.
Mathematics
June 2022
Supervisors: Docent Petteri Harjulehto & Professor Peter Hästö

---

The purpose of this thesis is to create exercises with health topics to Finnish upper secondary school advanced mathematics class *Probability calculus and statistics*. In the spirit of Finnish Upper Secondary School Curriculum the intention of these exercises is to bring down the figurative walls between subjects by integrating health sciences to mathematics and above all bring math closer to students' experiences and actual, current world health phenomena. There are three main goals with these exercises: 1) increase students' understanding of math applications in the real world and in that way hopefully 2) inspire them in their math studies and 3) support students with the fulfillment of the curriculum goals, such as improvement of (mathematical) understanding, problem solving skills and applying previously learned things to new contents. These exercises form a handy learning material that delves into current health topics. A class teacher may give the material to students as it is or use only some of the exercises as examples to enliven lessons or homework.

This thesis has both a theory part and the learning material part named *Terve!*, which is in Finnish and is to be found in the appendix. The theory part has two chapters, one for statistics and the other for probability, in which the definitions and theorems of the class are introduced. The order of those is similar to one in the probability class and many set books. Terve! -student material consists of 12 health themed exercises and verbal cues to solve them. The teacher material is very similar to the student material, but it contains answers and solutions. Besides every exercise there are some facts and additional information related to the theme of the exercise. The exercises are not in the same order as topics in the theory part, but in random order because it is important that the students have to figure out themselves how to define the problem and how to choose convenient methods to solve it. The aim of these kind of exercises is to guide students to reflect their own understanding and encourage them to think critically. Ideally, this way exercises are also helping students to improve their analytical problem solving skills and their ability to apply previously learned things to new contents.

Descriptors: probability calculus, statistics, health education, real-world problems, learning material, interest, upper secondary school

# Sisällys

# 1 Johdanto

Todennäköisyyslaskenta on matematiikan osa-alueena suhteellisen nuori: sen historian voidaan katsoa alkavan ranskalaismatemaatikkojen Blaise Pascalin ja Pierre Fermat'n uhkapeliaiheisesta kirjeenvaihdosta 1600-luvulla [1,2]. Perinteisesti todennäköisyyslaskenta onkin keskittynyt erilaisten rahapelien voiton mahdollisuuden laskemiseen ja nämä todennäköisyyslaskennan juuret ovat havaittavissa myös lukion todennäköisyyslaskennassa. Suuri osa lukion oppikirjojen tehtävistä liittyy pelisovelluksiin, kuten noppa- ja korttipeleihin sekä erilaisiin arvontoihin. Myös muut arkiset aiheet ovat suosittuja, mikä on harvinaisempaa muilla lukiomatematiikan opintojaksoilla. Harmillisesti aiheet jäävät usein irrallisiksi todellisista sovelluskohteista, sillä tehtävänannoissa tehdyt oletukset yksinkertaistavat rajusti tarkasteltavia ilmiöitä. Tällöin arkielämän teemat vaikuttavat hieman päälle liimatuilta. Lisäksi monet näistä aiheista, kuten rahapelit, jäävät etäisiksi useimpien opiskelijoiden kokemusmaailmasta. Tämä on sääli, sillä todennäköisyyslaskennan keinot taipuvat useiden nuoria kiinnostavien ja heidän kokemusmaailmaansa paremmin sopivien arjen, luonnon sekä yhteiskunnan ilmiöiden ja tapahtumien mallintamiseen.

Tämän tutkielman tarkoituksena on ollut tuottaa opiskelijalle mielenkiintoinen, terveysaiheinen tehtäväpaketti lukion pitkän matematiikan Tilastot ja todennäköisyys -opintojaksolle. Lukion opetussuunnitelman perusteiden 2019 mukaan matematiikan opetuksen lähtökohdat tulee valita opiskelijoita kiinnostavista aiheista, ilmiöistä ja ongelmista, joita voidaan matematiikan avulla ratkoa [3, s. 221]. Opiskelijoiden kiinnostuksen kohteiden huomioiminen on osa motivoivaa matematiikan opetusta, ja sitä korostetaan opetussuunnitelman lisäksi opetusalan keskusteluissa, kuten Beyranevandin blogikirjoituksessa [4]. Opetussuunnitelman perusteiden ja ajan didaktisen keskustelun hengessä tutkielman tehtävät ylittävät oppiaineiden välisiä kuiluja ja tuovat matematiikkaa lähemmäksi opiskelijan kokemusmaailmaa sekä ajankohtaisia yhteiskunnallisia ilmiöitä. Tehtävillä on kolme päätavoitetta: 1) lisätä opiskelijan ymmärrystä siitä, mitä kaikkea matematiikan avulla voidaan tehdä eli mihin opittua voidaan soveltaa ympäröivässä maailmassa, ja 2) innostaa opiskelijaa matematiikan opinnoissa sekä 3) tukea opiskelijaa lukion opetussuunnitelman yleisten ja matematiikan tavoitteiden, kuten ymmärtämisen, ongelmanratkaisun ja opitun soveltamisen toteutumisessa.

Tutkielma on jaettu teoriaosioon ja *Terve!*-materiaaliin, jonka opiskelija- ja opettajaversiot löytyvät tutkielman liitteistä. Tutkielman teoriaosio koostuu kahdesta luvusta, *Tilastot* (luku 2) ja *Todennäköisyyslaskenta* (luku 3), joissa esitellään lukion Tilastot ja todennäköisyys -opintojakson määritelmät ja tulokset todistuksineen. Näiden tulosten ja määritelmien esitysjärjestys noudattaa lukion opintojakson sekä useiden kurssikirjojen etenemisjärjestystä. Lukujen 2 ja 3 teorian esitystavat eroavat toisistaan melko paljon tilastotieteen ja matematiikan tieteenalojen kirjallisen ilmaisun erilaisuuden vuoksi: Todennäköisyyslaskenta –luvun esitysmuoto on matemaattinen, Tilastot –luvussa määritelmät ovat suurilta osin sanallisia ja niiden tukena käytetään esimerkkejä.

Terve!-opiskelijamateriaali koostuu kahdestatoista terveysaiheisesta tehtävästä ja sanallisista vihjeistä niiden ratkaisemiseksi. Opettajan materiaali on opiskelijamateriaalin kaltainen, mutta siinä esitetään eräät malliratkaisut tehtäville. Kaikki materiaalin tehtävät liittyvät ajankohtaisiin arkiympäristön ja yhteiskunnan terveysilmiöihin, kuten treenaamiseen, ensiaputaitoihin, rokotteisiin, punkkien levittämiin tauteihin ja väestöryhmien välisiin terveyseroihin. Kunkin tehtävän yhteydessä kerrotaan myös hieman tehtävän aiheesta ja annetaan vinkkejä asiallisen lisätiedon löytämiseen. Tehtävien tavoitteena on, että opiskelija tutustuu erilaisten terveysilmiöiden syihin ja seurauksiin, ymmärtää ja kehittää niihin liittyvää matemaattista ajattelua sekä oppii analysoimaan saatujen tulosten mielekkyyttä ja luotettavuutta. Tarkoituksena on harjoitella pohtimaan saatujen tulosten merkitystä terveydelle yksilö-, yhteisö- ja yhteiskuntatasolla. Materiaalin tehtävät eivät mukaile opintojakson etenemisjärjestystä, vaan ne ovat tarkoituksellisesti sekalaisessa järjestyksessä. Näin opiskelija pääsee ratkaistavan ongelman määrittelystä ja sopivan matemaattisen ratkaisumenetelmän valinnasta alkaen kehittämään omaa ymmärrystä, kriittistä arviointia sekä ongelmanratkaisu- ja soveltamistaitoja. Terve!-materiaalista kerrotaan tarkemmin luvussa 4 ja materiaalin käyttöön liittyviä pedagogisia teemoja käsitellään luvussa 5.

# 2 Tilastot

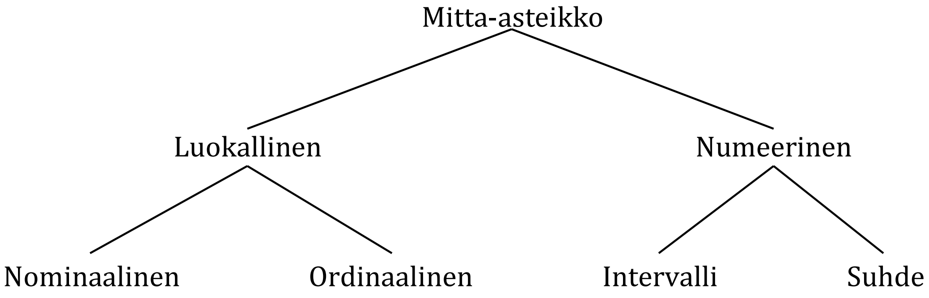
Tässä luvussa tutustutaan sellaisiin tilastotieteen peruskäsitteisiin ja tuloksiin, jotka kuuluvat lukion pitkän matematiikan Tilastot ja todennäköisyys -opintojakson sisältöihin. Luvun määritelmät ja tulokset perustuvat Henri Pesosen monisteeseen [5] ja Ilkka Mellinin esitykseen [6].

## 2.1 Tilastotieteen peruskäsitteitä

**Määritelmä 2.1** Tilastotieteellisessä tutkimuksessa tarkasteltavaa tutkimuskohteiden eli *tilastoyksiköiden* konkreettista tai hypoteettista joukkoa kutsutaan perusjoukoksi eli *populaatioksi*. Tarkastelun kohteena ovat tilastoyksiköiden havaittavat tai mitattavat ominaisuudet, joita kutsutaan *tilastollisiksi muuttujiksi* tai *tilastomuuttujiksi*.

Konkreettisessa populaatiossa jokainen yksilö voidaan tarvittaessa listata. Hypoteettisen populaation avulla pyritään ennustamaan tiettyjä populaation ominaisuuksia tulevaisuudessa.

Tilastomuuttujia voidaan jaotella monella tapaa. Ne voidaan jakaa havaittujen arvojen perusteella mm. diskreetteihin ja jatkuviin muuttujiin sen mukaan, voiko muuttuja saada mitä tahansa lukuarvoja (*jatkuva muuttuja*) vai vain yksittäisiä arvoja joltakin väliltä (*diskreetti muuttuja*). Lisäksi tilastomuuttujat voidaan jaotella numeerisiin ja luokallisiin muuttujiin sen mukaan, minkälaisilla mittaustasoilla niiden arvoja havaitaan (Kuva 2.1).



*Kuva 2.1* Tilastollisten muuttujien mitta-asteikkojen tyypit.

**Määritelmä 2.2** Jos tilastollista muuttujaa ei voi mitata diskreetillä tai jatkuvalla numeerisella asteikolla, ts. jos muuttujan havaittu arvo ei ole numeerinen luku, kyseessä on *luokallinen muuttuja*.

**Määritelmä 2.3** Jos tilastollisen muuttujan havaittu arvo on jokin lukuarvo, kyseessä on *numeerinen muuttuja*.

Luokallisia muuttujia ovat mm. henkilön sukupuoli (luokat nainen, mies ja muu) ja syöpäkasvaimen levinneisyysaste (luokat 0, I, II, III ja IV). Luokalliset muuttujat voidaan jaotella edelleen kahteen aliluokkaan sen mukaisesti, voidaanko ne laittaa mitta-asteikkonsa perusteella luonnollisesti johonkin järjestykseen. Jos muuttujan arvoja ei voida laittaa luonnolliseen järjestykseen, niin luokallisen muuttujan mitta-asteikko on laatuero- eli nominaaliasteikko. Esimerkiksi henkilön sukupuolen mitta-asteikko on nominaaliasteikollinen. Jos taas muuttujan arvoilla on luonnollinen järjestys, niin luokallisen muuttujan mitta-asteikko on järjestys- eli ordinaaliasteikollinen. Esimerkiksi syöpäkasvaimen levinneisyysluokat voidaan järjestellä paikallisesta, leviämättömästä kasvaimesta (0) laajalle etäpesäkkeisesti levinneisiin kasvaimiin (IV).

**Määritelmä 2.4** *Nominaaliasteikolla* mitattavat muuttujat voidaan jakaa ominaisuuksiensa perusteella ryhmiin sen perusteella, onko muuttujilla ominaisuutta vai ei.

**Määritelmä 2.5** *Ordinaaliasteikolla* mitattavat muuttujat voidaan järjestellä joidenkin ominaisuuksiensa perusteella ryhmiin, joilla on jokin luonnollinen järjestys.

Myös numeerisia, jatkuvia tai diskreettejä lukuarvoja saavia muuttujia voidaan jaotella aliluokkiin. Jos numeerinen muuttuja voi saada negatiivisia arvoja ja nollakohta on sopimuksenvarainen, niin numeerisen muuttujan mitta-asteikko on välimatka- eli intervalliasteikko. Esimerkiksi kristillinen ajanlasku vuosilukuina ja lämpötila Celsius-asteikolla ovat intervalliasteikollisia muuttujia. Intervalliasteikolla voidaan verrata mitta-asteikon välien pituuksia toisiinsa eli voidaan esimerkiksi sanoa marraskuisen vuorokauden minimi- ja maksimilämpötiloja kuvatessa Mumbain vaihteluvälin [20,5 °C, 33,4 °C] olevan noin 2,5 kertaa Lontoon vaihteluväliä [10,9 °C, 16,1 °C] pidempi. Jos taas numeerisen muuttujan mitta-asteikolla on intervalliasteikon ominaisuuksien lisäksi absoluuttinen nollakohta, kuten henkilön pituudella senttimetreissä, niin numeerisen muuttujan mitta-asteikko on suhde(luku)asteikko. Tämän mitta-asteikon nimitys viittaa suuruussuhteiden laskettavuuden mielekkyyteen; voidaan esimerkiksi sanoa että 208 cm pitkä koripalloilija on kaksi kertaa niin pitkä kuin 104 cm pitkä lapsi.

**Määritelmä 2.6** *Intervalliasteikolla* mitattavien muuttujien havainnoista voidaan laskea erotus.

**Määritelmä 2.7** *Suhdeasteikolla* mitattavilla muuttujilla on yksikäsitteinen nollapiste.

**Määritelmä 2.8** *Otos* on jollain satunnaismenetelmällä suuremmasta perusjoukosta valittu tilastoyksiköiden joukko.

Satunnaisotannassa jokaisella tilastoyksiköllä on tietty todennäköisyys tulla valituksi otokseen. Otos pyritään muodostamaan populaatiota mahdollisimman hyvin kuvaavaksi, mitä varten tilastotieteessä on kehitetty erilaisia otantamenetelmiä.

**Määritelmä 2.9** Muodostettu otos on *edustava*, jos sen ominaisuudet ovat samankaltaiset kuin populaation ominaisuudet. Jos taas otoksen ominaisuudet eivät vastaa populaation ominaisuuksia, on muodostettu otos *harhainen*.

Edustava otos on siis ikään kuin pienoismalli populaatiosta.

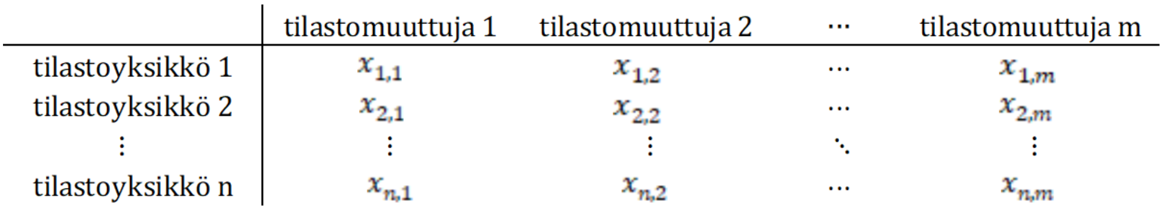
**Määritelmä 2.10** Kun tutkimuksen kohteena ovat kaikki populaation tilastoyksiköt, on kyse *kokonaistutkimuksesta*. Yleensä tutkimusaineistot ovat kuitenkin otospohjaisia eli ne koostuvat jollakin satunnaismenetelmällä valituista tilastoyksiköistä, jotka ovat vain pieni osa koko populaatiosta. Tällöin puhutaan *otantatutkimuksesta*.

## 2.2 Tilastoaineiston käsittely

Empiirisessä tilastollisessa tutkimuksessa havaintoaineiston keräämisen jälkeen aloitetaan aineiston analyysi, joka voidaan jakaa kahteen osaan: kuvailevaan tilastotieteeseen ja tilastolliseen päättelyyn. Kuvailevan eli deskriptiivisen tilastotieteen menetelmin pyritään esittämään havaintoaineiston pääpiirteet tiivistetyssä, informatiivisessa muodossa. Menetelmillä ei siis pyritä yleispätevien, koko populaatiota koskevien päätelmien tekemiseen. Tilastollinen päättely taas on induktiivista loogista päättelyä, jolla pyritään tekemään otoksesta koko populaatiota koskevia päätelmiä.

### 2.2.1 Tilaston muodostaminen ja luokittelu

Aluksi havaintoaineisto kootaan taulukoksi, johon listataan tilastoyksiköt riveille ja tilastomuuttujat sarakkeisiin. Jos havaintoaineisto koostuu $n$ tilastoyksiköstä, joista jokaisesta on kerätty $m$ tilastomuuttujasta havainnot, niin havainnot voidaan kirjoittaa taulukon muotoon



| | tilastomuuttuja 1 | tilastomuuttuja 2 | $\cdots$ | tilastomuuttuja m |
|---|---|---|---|---|
| tilastoyksikkö 1 | $x_{1,1}$ | $x_{1,2}$ | $\cdots$ | $x_{1,m}$ |
| tilastoyksikkö 2 | $x_{2,1}$ | $x_{2,2}$ | $\cdots$ | $x_{2,m}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| tilastoyksikkö n | $x_{n,1}$ | $x_{n,2}$ | $\cdots$ | $x_{n,m}$ |

,

jossa rivillä $i$ on $i$. tilastoyksikön havainto ja $j$ sarakkeessa on $j$. tilastollisesta muuttujasta havaitut arvot $x_{i,j}$. Siis yksi rivi on yhden tilastoyksikön tiedot kaikista tilastomuuttujista ja yksi sarake on kaikkien tilastoyksiköiden tiedot yhdestä tilastomuuttujasta.

Usein kerätyt havaintoaineistot ovat niin suuria, ettei havaintotaulukkoa suoraan tarkastelemalla näy aineiston pääpiirteitä. *Luokittelussa* on kysymys aineiston tiivistämisestä kohtuullisen kokoiseksi ja havainnollisempaan muotoon. Luokittelussa tilastomuuttujan arvot sijoitetaan eri luokkiin siten, että yhden tilastomuuttujan arvo voi kuulua vain yhteen luokkaan. *Luokka* ilmoitetaan yleensä luokkavälinä, kuten reaalilukuvälinä. Esimerkiksi henkilön ikä on tapana luokitella ikäjakauman kuvaamisessa 10-vuotisluokkiin (15-24, 25-34, ...), vaikka periaatteessa ikä voitaisiin ilmoittaa minuutinkin tarkkuudella.

Luokkien lukumäärään vaikuttavat muun muassa tilastomuuttujan arvojen vaihteluväli ja havaintoaineiston laajuus. Luokittelussa pyritään siihen, että luokkien lukumäärä saadaan tarvittaessa luokkia yhdistämällä kohtuulliseksi ja että luokat valitaan tasavälisesti eli siten, että kahden peräkkäisen luokan alarajojen erotus on vakio. Kun aineistoa luokitellaan, aineiston luettavuus paranee mutta toisaalta osa tiedoista menetetään eivätkä yksittäiset havaintoarvot ole enää tiedossa.

**Määritelmä 2.11** *Luokkakeskus* on luokan ala-ja ylärajan keskiarvo.

Yksi yksinkertainen tapa tiivistää aineistoa on muodostaa siitä *frekvenssijakauma*. Tilastoyksiköiden tilastollisilla muuttujilla on tietty mahdollisten arvojen joukko ja näillä arvoilla on jokin jakauma populaatiossa. Tilastollisen muuttujan (frekvenssi)jakauma ilmaisee, kuinka yleisiä muuttujan eri arvot ovat. Absoluuttinen jakauma sisältää havaintoarvot ja niiden esiintymiskertojen lukumäärät eli *frekvenssit f*. Suhteellinen jakauma sisältää havaintoarvot ja niiden esiintymiskertojen suhteelliset osuudet havaintojen kokonaismäärästä eli *suhteelliset frekvenssit* $f_{\%}$.

**Määritelmä 2.12** Oletetaan, että tilastomuuttujan mahdolliset arvot voidaan jakaa äärelliseen määrään ($k$ kpl) luokkia, joita merkitään $E_1, E_2, \dots, E_k$. Etsitään aineistosta niiden havaintojen määrät, *frekvenssit*, jotka kuuluvat näihin luokkiin ja merkitään näitä $f_1, f_2, \dots, f_k$. Tilastomuuttujan frekvenssijakauma on $(E_i, f_i), i = 1, \dots, k$. Havaintojen määrää ilmaisevaa frekvenssiä *f* kutsutaan myös *absoluuttisiksi frekvensseiksi*.

Frekvenssijakauma voidaan siis muodostaa tapauksissa, joissa tilastomuuttuja on joko luokallinen muuttuja tai diskreetti numeerinen muuttuja. Usein absoluuttisten frekvenssien sijaan on tarpeen selvittää eri luokkiin kuuluvien tilastomuuttujien suhteellinen osuus.

**Määritelmä 2.13** *Suhteellinen frekvenssi* $f_{\%}$ kuvaa kunkin frekvenssin osuutta havaintoaineiston tilastoyksiköiden kokonaismäärästä *n*. *Suhteellinen frekvenssijakauma* on $(E_i, \frac{f_i}{n}), i = 1, \dots, k$.

Suhteellinen frekvenssijakauma ilmaistaan usein prosentuaalisena frekvenssijakaumana $(E_i, \frac{f_i}{n} \cdot 100\%), i = 1, \dots, k$. Suhteellinen frekvenssijakauma ei ole riippuvainen otoskoosta *n*, minkä vuoksi jakaumien vertailu toisiinsa on helpompaa.

Kun tilastomuuttujan mitta-asteikko on luokallinen tai diskreetti numeerinen, on luokkajaon tekeminen selkeää. Jos luokkia on todella suuri määrä ja jokaiseen luokkaan kuuluu vain pieni määrä havaintoja tai niiden käsittely muodostuu muutoin hankalaksi, voidaan luokkia yhdistää. Jos taas muuttujan mitta-asteikko on jatkuva numeerinen, ei frekvenssijakaumaa voida käyttää suoraan.

**Määritelmä 2.14** (Jatkuvan) numeerisen tilastomuuttujan mahdollisten arvojen joukko voidaan jakaa toisensa poissulkeviin väleihin ($k$ kpl), jolloin jokainen tilastomuuttujan arvo kuuluu täsmälleen yhteen luokkaan $E_i, i = 1, \dots, k$. Luokista muodostettua frekvenssijakaumaa kutsutaan *luokitelluksi frekvenssijakaumaksi*.

Jos tilastomuuttuja on mitattu ordinaaliasteikolla, niin havaintoja voidaan kuvata myös summafrekvenssillä *F*.

**Määritelmä 2.15** *Summafrekvenssi* $F_i$ kuvaa, kuinka monta havaintoa kuuluu luokkaan $E_i$ ja sitä edeltäviin luokkiin. Summafrekvenssit saadaan laskettua frekvenssien avulla seuraavasti:

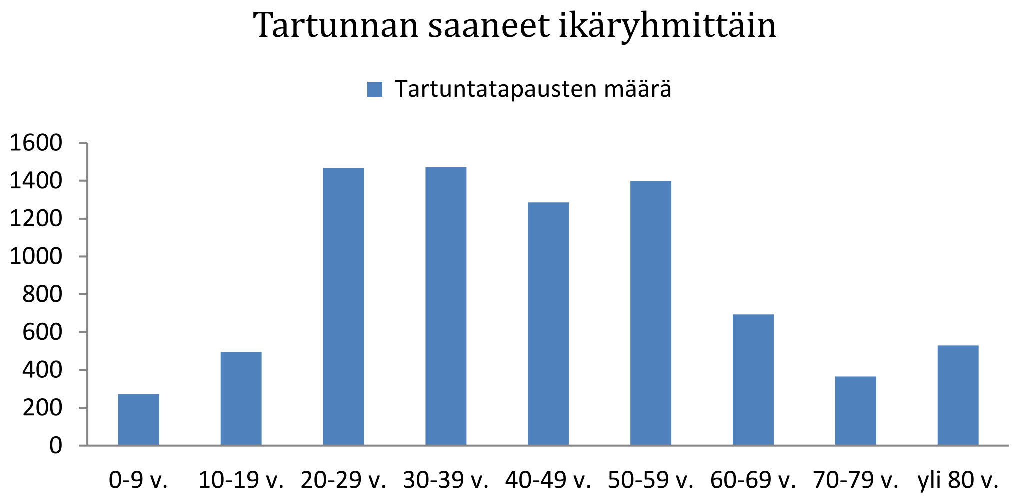
$$\begin{aligned}
&F_1 = f_1 && \\
&F_2 = f_1 + f_2 && = F_1 + f_2 \\
&F_3 = f_1 + f_2 + f_3 && = F_2 + f_3 \\
&\vdots && \\
&F_k = f_1 + f_2 + \cdots + f_k && = F_{k-1} + f_k.
\end{aligned}$$

Summafrekvenssijakauma on siis muotoa $(E_i, F_i), i = 1, \dots, k$. Kuten aiemmin, määritellään myös *suhteellinen summafrekvenssijakauma* $(E_i, \frac{F_i}{n}), i = 1, \dots, k$ ja *prosentuaalinen summafrekvenssijakauma* $(E_i, \frac{F_i}{n} \cdot 100\%), i = 1, \dots, k$.

### 2.2.2 Tilaston graafinen esitys

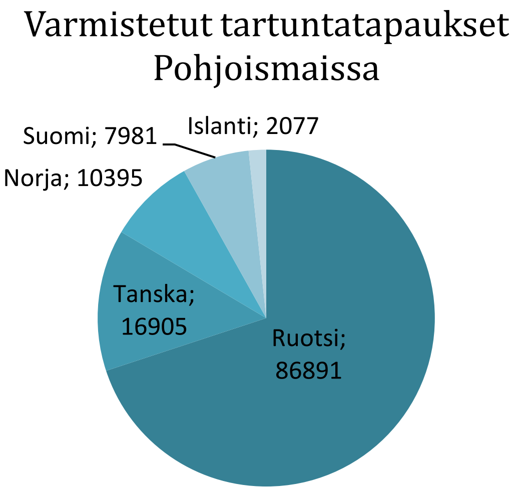
Graafiset menetelmät ovat tärkeä osa aineiston havainnollistamista, sillä kuvat helpottavat aineiston tulkitsemista ja usein myös antavat välitöntä palautetta, jos käytössä olevat tilastolliset mallit eivät ole riittäviä selittämään aineistoa. Kuviotyyppi valitaan sen mukaan, mikä aineistossa on huomionarvoista.

*Pylväskuvioita* eli pylväsdiagrammeja (Kuva 2.2) käytetään aineiston havainnollistamiseen, kun havainnot ovat luokallisia tai diskreettejä numeerisia. Niissä *pylvään korkeus* ilmaisee määrää, muuttujan arvojen frekvenssejä. Erityisesti silloin, kun kuvataan jonkin asian muuttumista ajassa vaaka-akselilla vasemmalta oikealle tai kun mitta-asteikko on ordinaalinen ja järjestys kasvaa vasemmalta oikealle, käytetään *pystypylväskuviota*. Myös *vaakapylväskuvioita* käytetään yleisesti. Niissä pylväiden nimet saadaan selvemmin kirjoitettua kuvaan. Silloin, kun luokilla ei ole luonnollista järjestystä, on hyvä piirtää pylväät suuruusjärjestyksessä joko vasemmalta oikealle tai ylhäältä alas. *Pylväsryhmäkuvioita* ja *summapylväskuvioita* voidaan käyttää aineiston havainnollistamisessa, kun mielenkiinnon kohteena on useita tilastomuuttujia ja halutaan tarkastella jakaumien mahdollisia eroja ryhmien välillä.



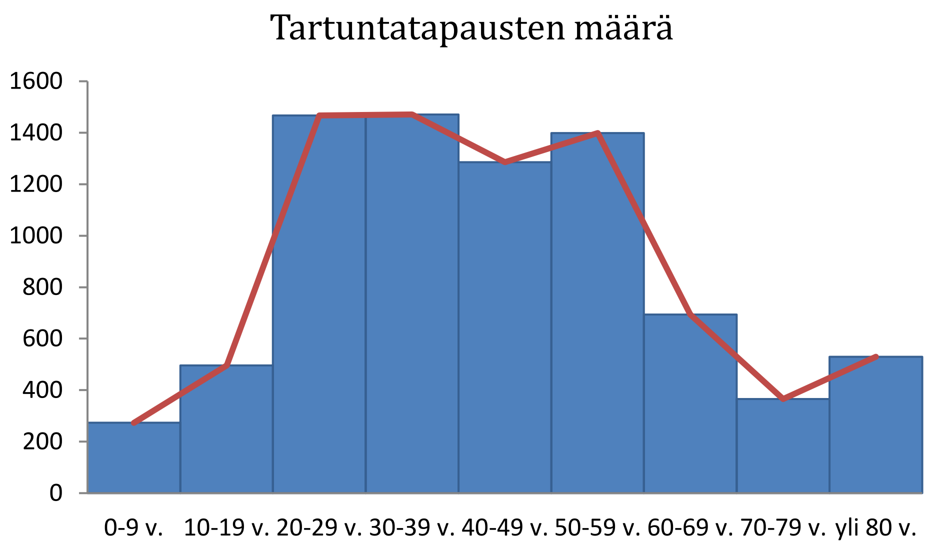
*Kuva 2.2* Esimerkki pylväsdiagrammista: pystypylväskuvio.

Kun halutaan havainnollistaa kokonaisuuden jakautumista osiin, sopii *ympyrädiagrammi* (Kuva 2.3) hyvin graafiseksi menetelmäksi. Siinä havaintoaineisto piirretään ympyränä, joka jaetaan luokkien mukaan sektoreihin siten, että sektorin koko vastaa luokan suhteellista osuutta jakaumassa. Sektorit suositellaan piirtämään suuruusjärjestyksessä myötäpäivään kiertäen.



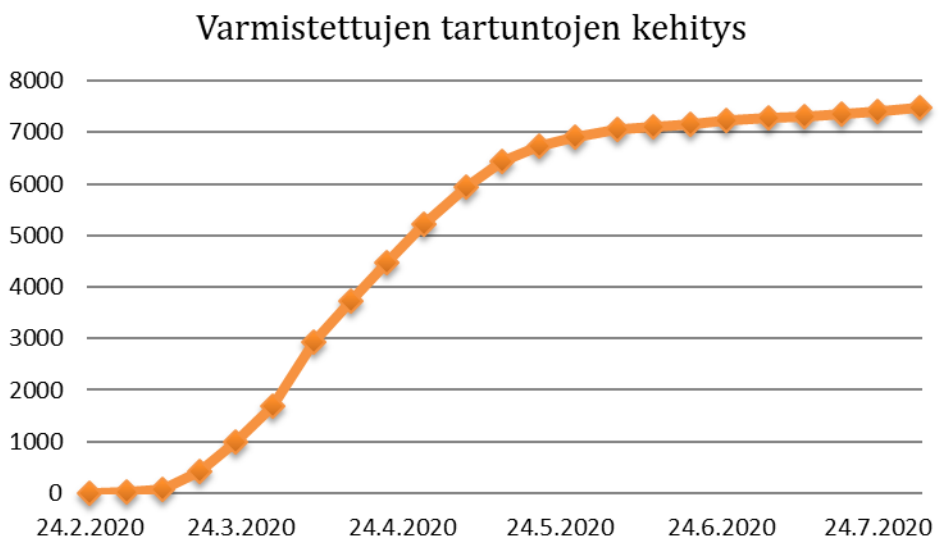
*Kuva 2.3* Esimerkki ympyrädiagrammista.

*Histogrammi* (Kuva 2.4) on pylväskuvio luokitellulle (suhteelliselle) frekvenssijakaumalle, jossa pylvään ala kuvaa luokkien frekvenssiä. Histogrammissa pylväät piirretään kiinni toisiinsa havainnollistamaan mitta-asteikon jatkuvuutta ja niiden reunat ovat luokkien todellisten luokkarajojen kohdalla. Usein luokittelu on tasavälinen, jolloin pylväät ovat yhtä leveitä. Tällöin myös pylväiden korkeudet kuvaavat frekvenssejä. Luokittelun ollessa tasavälinen histogrammin päälle voidaan muodostaa *frekvenssimonikulmio* (Kuva 2.4) siten, että yhdistetään murtoviivalla frekvenssiarvot kunkin luokan luokkakeskuksessa. Näin saadaan piirrettyä luokiteltu frekvenssijakauma, joka on eräänlainen approksimointi populaatiojakaumalle.



*Kuva 2.4* Esimerkki histogrammista ja frekvenssimonikulmiosta.

*Viivakaaviolla* eli viivadiagrammilla (Kuva 2.5) havainnollistetaan erityisesti ajan suhteen muuttuvien suureiden vaihtelua ja kehityssuuntia. Vaaka-akselilla on yleensä tasavälisiä aikavälejä ja pystyakselilla jokin muuttujan määrää kuvaava lukuarvo.



*Kuva 2.5* Esimerkki viivadiagrammista.

### 2.2.3 Tilastolliset tunnusluvut

Toisinaan on tarpeen esittää havaintoaineiston informaatio frekvenssijakaumaakin tiiviimmin. Tällöin käytetään tilastollisia tunnuslukuja, joita ovat keski- eli paikkaluvut ja hajontaluvut. Keskilukuja käytetään silloin, kun halutaan kuvata tilastoyksiköiden yleisimpiä ominaisuuksia. Riippuu tilastoyksikön mitta-asteikosta, mitä keskilukuja aineistolle voidaan määrittää.

**Määritelmä 2.16** Tilastomuuttujan *tyyppiarvo* eli *moodi*, merkitään Mo, on havaintoaineistossa yleisimmin esiintyvä tilastomuuttujan arvo eli tilastomuuttujan arvo, jonka frekvenssi on suurin.

Tyyppiarvon olemassaolo ei riipu tilastomuuttujan mitta-asteikosta vaan se voidaan aina määrittää. Tyyppiarvoja voi myös olla useita, jos useammalla tilastomuuttujan arvolla on sama frekvenssi. Tällöin tyyppiarvo ei ole yksikäsitteinen.

**Määritelmä 2.17** *Mediaani*, merkitään Md, on järjestetyn havaintoaineiston keskimmäinen alkio eli järjestykseen asetetuista havaintoarvoista keskimmäinen.

Mediaani kuvaa siis tilastomuuttujan arvojen keskimääräistä suuruutta. Se voidaan määrittää, jos tilastomuuttujan arvot voidaan mielekkäällä tavalla järjestää eli tilastomuuttuja on järjestys-, suhde- tai intervalliasteikollinen muuttuja. Mediaani on yksikäsitteinen, jos havaintoja on pariton määrä. Jos taas havaintojen määrä on parillinen ja kahdella keskimmäisellä järjestykseen asetetuista havaintoarvoista on eri arvo, mediaani on kumpikin näistä. Tilastomuuttujan mediaanin määrittämisessä voidaan käyttää apuna suhteellista summafrekvenssiä *F*. Mediaani on se arvo, jonka kohdalla suhteellinen summafrekvenssi saavuttaa arvon 50 % tai myös sitä seuraava arvo, jos suhteellinen summafrekvenssi on tasan 50 %.

Keskiluvuista yleisimmin käytetty on havaintoaineistosta intervalli- tai suhdeasteikollisille muuttujille laskettu *aritmeettinen keskiarvo* $\mu$. Aritmeettista keskiarvoa kutsutaan *otoskeskiarvoksi* $\bar{x}$ silloin, kun tarkastellaan otosta jostain populaatiosta. Tilastomuuttujan keskiarvo saadaan laskemalla yhteen muuttujan arvot ja jakamalla summa havaintojen (tilastoyksiköiden) lukumäärällä.

**Määritelmä 2.18** (Keskiarvo)

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

jossa *n* on havaintojen lukumäärä ja $x_1, x_2, \dots, x_n$ ovat tilastomuuttujan saamat arvot.

Jos numeerinen havaintoaineisto on luokiteltu $k$ luokkaan, voidaan frekvenssijakauman frekvensseistä $f_i$ ja luokkien $E_i$ luokkakeskipisteistä $x_i$ laskea aritmeettinen keskiarvo eli frekvenssijakauman otoskeskiarvo siten, että kerrotaan kukin luokkakeskipisteen arvo sen frekvenssillä ja jaetaan näiden tulojen summa havaintojen (tilastoyksiköiden) lukumäärällä.

**Määritelmä 2.19** (Frekvenssijakauman keskiarvo)

$$\mu = \frac{1}{n}\sum_{i=1}^{k} f_i x_i \,,$$

jossa *n* on havaintojen lukumäärä ja $f_i$ muuttujan arvon $x_i$ frekvenssi $(i = 1, 2, \ldots, k)$.

Jakauman hajonta on tärkeä jakaumien peruspiirre. Hajontaa mittaavat tunnusluvut ilmentävät havaintojen vaihtelun suuruutta havaintoaineistossa: sitä, kuinka laveasti tai tiiviisti mahdolliset arvot ovat jakautuneet. Mitä lähempänä hajonnan tunnusluvut ovat nollaa, sitä vähemmän arvot vaihtelevat.

**Määritelmä 2.20** *Vaihteluvälin pituus* on suurimman ja pienimmän havaintoarvon erotus.

Vaihteluvälin pituus kuvaa, kuinka suurelta väliltä kaikki aineiston arvot löytyvät. Se riippuu ainoastaan suurimmasta ja pienimmästä havainnosta eikä kerro mitään muiden havaintojen jakautumisesta tälle välille, minkä vuoksi se on tunnuslukuna käyttökelpoinen vain hyvin pienille, alle viiden havainnon havaintoaineistoille.

Tilastomuuttujan *varianssi* $\sigma^2$ on muuttujan arvojen keskiarvoista laskettujen poikkeamien neliöiden keskiarvo.

**Määritelmä 2.21** (Varianssi)

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2,$$

jossa *n* on havaintojen lukumäärä, $x_1, x_2, \ldots, x_n$ tilastomuuttujan saamat arvot ja $\mu$ havaintoarvojen $x_i$ $(i = 1, 2, \ldots, k)$ keskiarvo.

Jos numeerinen havaintoaineisto on luokiteltu *k* luokkaan ja luokan $E_i$ luokkakeskipisteen arvon $x_i$ frekvenssi on $f_i$ , voidaan varianssi laskea myös frekvenssien avulla, sillä jokainen termi $(x_i - \mu)^2$ toistuu summassa $f_i$ kertaa.

**Määritelmä 2.22** (Frekvenssijakauman varianssi)

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \mu)^2,$$

jossa $n$ on havaintojen lukumäärä, $f_i$ muuttujan arvon $x_i$ frekvenssi $(i = 1, 2, \dots, k)$ ja $\mu$ havaintoarvojen $x_i$ keskiarvo.

Varianssista voidaan laskea tilastomuuttujan *keskihajonta* $\sigma$, joka on varianssin neliöjuuri eli $\sigma = \sqrt{\sigma^2}$. Keskihajonta kuvaa, kuinka paljon havaintoaineiston arvot poikkeavat keskiarvosta.

**Määritelmä 2.23** (Keskihajonta) Havaintoarvojen $x_i, i = 1, 2, \dots, n$, keskihajonta $\sigma$ on

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \,,$$

jossa $n$ on havaintojen lukumäärä ja $\mu$ havaintoarvojen $x_i$ keskiarvo.

Keskihajonta voidaan laskea myös frekvenssien avulla.

**Määritelmä 2.24** (Frekvenssijakauman keskihajonta)

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}f_i(x_i - \mu)^2} \,,$$

jossa $n$ on havaintojen lukumäärä, $f_i$ muuttujan arvon $x_i$ frekvenssi $(i = 1, 2, \dots, n)$ ja $\mu$ havaintoarvojen $x_i$ keskiarvo.

Jos halutaan vertailla useasta aineistosta poimittuja havaintoarvoja, niin havaintoarvot tehdään ensin vertailukelpoisiksi normittamalla ne. Havaintoarvoa vastaava *normitettu arvo* on havaintoarvon poikkeama keskiarvosta jaettuna keskihajonnalla. Normitettu arvo siis kertoo, kuinka monen keskihajonnan päässä keskiarvosta tilastomuuttujan arvo on.

**Määritelmä 2.25** Tilastomuuttujan arvoa $x$ vastaava *normitettu arvo* $z$ on

$$z = \frac{x - \mu}{\sigma},$$

jossa $\mu$ on keskiarvo ja $\sigma$ keskihajonta.

Normitetun arvon avulla voidaan siis verrata, kumpi kahdesta eri aineistoihin kuuluvasta muuttujan arvosta on poikkeavampi.

### 2.2.4 Tilastollinen riippuvuus ja korrelaatio

Tyypillisesti tieteellisten tutkimusten mielenkiintoisimmat ja tärkeimmät kysymykset liittyvät tutkittavaa ilmiötä kuvaavien muuttujien välisiin riippuvuuksiin [6]. Terveysaiheisissa tutkimuksissa voidaan tarkastella esimerkiksi sitä, miten alkoholin kulutus riippuu alkoholijuomien hintatasosta, kuluttajan tulotasosta tai alkoholin saatavuudesta, tai sitä, miten todennäköisyys sairastua keuhkosyöpään riippuu tupakoinnin määrästä ja kestosta.

Kahden muuttujan välinen riippuvuus on *eksaktia*, jos toisen arvot voidaan ennustaa tarkasti toisen saamien arvojen perusteella. Jos muuttujien välillä ei ole eksaktia riippuvuutta, mutta toisen muuttujan arvoja voidaan käyttää apuna toisen muuttujan arvojen ennustamisessa, puhutaan *tilastollisesta riippuvuudesta*.

**Määritelmä 2.26** Jos toinen muuttuja riippuu toisesta lineaarisesti, puhutaan lineaarisesta riippuvuudesta eli *lineaarisesta regressiosta*. Tätä lineaarista tilastollista riippuvuutta kutsutaan myös *korrelaatioksi*. Riippuvuutta kuvaavaa suoraa sanotaan *regressiosuoraksi*.

Jos regressiosuora on kasvava, eli toisen muuttujan arvon kasvaessa toisenkin muuttujan arvo kasvaa, niin sanotaan muuttujien välillä olevan positiivinen korrelaatio. Kun regressiosuora on vähenevä, niin muuttujien välinen korrelaatio on negatiivinen. Huomattavaa on, että kaksi muuttujaa voivat olla toisistaan riippuvia, vaikka niiden välillä ei olisi korrelaatiota (ei-lineaarinen riippuvuus). Toisaalta kahden muuttujan välinen voimakas korrelaatio ei välttämättä todista muuttujien välistä syy-seuraussuhdetta.

Seuraavaksi määritellään Pearsonin korrelaatiokerroin. Sen määritelmässä tulee tuntea myöhemmin luvussa 3 esiteltävä odotusarvon käsite (Määritelmät 3.23, 3.24 ja 3.29). Korrelaatiokertoimen määrittäminen matemaattisesti ohitetaan lukiomatematiikassa ja se selvitetään laskinohjelmistojen avulla.

**Määritelmä 2.27** *Korrelaatiokerroin* ($r$) on tilastollinen tunnusluku, joka kuvaa lineaarisen tilastollisen riippuvuuden voimakkuutta. Pearsonin korrelaatiokerroin satunnaismuuttujien X ja Y väliselle lineaariselle riippuvuudelle on

$$\rho_{X,Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y},$$

missä $\mu_X$ ja $\mu_Y$ ovat satunnaismuuttujien X ja Y odotusarvot sekä $\sigma_X$ ja $\sigma_Y$ ovat satunnaismuuttujien keskihajonnat . Korrelaatiokerroin on määritelty vain, kun keskihajonnat $\sigma_X$ ja $\sigma_Y$ ovat äärellisiä ja nollasta eroavia.

Korrelaatiokerroin kuuluu välille $[-1,1]$. Mitä lähempänä se on välin päätepisteitä, sitä lähempänä havainnot ovat suoraa $y = kx + b$, jossa kulmakertoimen $k$ etumerkki määrää korrelaatiokertoimen etumerkin.

**Määritelmä 2.28** Korrelaatiokertoimen toista potenssia $r^2$ kutsutaan *selitysasteeksi*. Se ilmaisee, kuinka suuren osuuden muuttujat selittävät toistensa arvojen vaihtelusta.

Korrelaatiot muodostavat perustan muuttujien välisten riippuvuuksien ymmärtämiselle, mutta tavallisesti näitä riippuvuuksia halutaan analysoida myös tarkemmin. *Regressioanalyysi* on tilastollinen menetelmä, jossa yhden muuttujan (nk. selitettävän muuttujan) tilastollista riippuvuutta muista muuttujista (ns. selittävistä muuttujista) pyritään mallintamaan ja ennustamaan regressiomalliksi kutsutulla tilastollisella mallilla.

# 3 Todennäköisyyslaskenta

Tässä luvussa esitellään sellaiset todennäköisyyslaskennan käsitteet ja tulokset, jotka ovat lukion pitkän matematiikan Tilastot ja todennäköisyys -opintojakson ydinainesta ja joita tarvitaan oppimateriaalissa esiteltävien tehtävien ratkaisemiseksi. Määritelmät ja tulokset perustuvat Heikki Ruskeepään [7], Jyrki Lahtosen [8] ja Iiro Honkalan monisteisiin [9]. Ensiksi tutustutaan todennäköisyyden perusominaisuuksiin: todennäköisyyskäsitteeseen, joukko-opin perusmerkintöihin ja niiden todennäköisyystulkintoihin sekä todennäköisyydeltä vaadittaviin ominaisuuksiin. Tämän jälkeen tutustutaan muutamiin kombinatoriikan tuloksiin, jotka helpottavat tapausten lukumäärän tutkimista ja laskemista. Sitten syvennytään todennäköisyyden laskusääntöihin sekä huomioihin niiden käytettävyyden taustalla. Lopuksi käsitellään diskreettejä ja jatkuvia satunnaismuuttujia sekä yleisimpiä todennäköisyysjakaumia tunnuslukuineen.

## 3.1 Todennäköisyyden perusominaisuuksia

### 3.1.1 Todennäköisyys

Todennäköisyyslaskenta tutkii ja mallintaa matemaattisesti stokastisia ilmiöitä eli satunnaisilmiöitä. Satunnaisilmiöt ovat sellaisia ilmiöitä, joiden lopputuloksen määrää sattuma ja joiden tulosta ei näin ollen pystytä tarkasti ennustamaan. Satunnaiskokeen mallintamista varten tarvitaan kuitenkin tieto mahdollisista lopputulosvaihtoehdoista.

**Määritelmä 3.1** Kokeen erilaisia tuloksia kutsutaan *alkeistapauksiksi*. Mikä tahansa alkeistapausten joukko on *tapaus (tapahtuma)*, mutta kaikkien alkeistapausten joukkoa kutsutaan *perusjoukoksi* eli *otosavaruudeksi* $\Omega$. Tapaus *sattuu*, kun kokeen tulokseksi saadaan alkeistapaus, joka kuuluu tapaukseen.

Klassisesta todennäköisyydestä puhutaan, kun alkeistapauksia on äärellinen määrä ja ne ovat kaikki yhtä todennäköisiä. Yksinkertainen oppikirjaesimerkki tällaisesta satunnaiskokeesta on nopan heitto, jonka yhtä todennäköiset alkeistapaukset ovat nopan silmäluvut 1, 2, 3, 4, 5 ja 6.

**Määritelmä 3.2** (*Todennäköisyyden klassinen määritelmä*) Merkitään tapauksen $A$ todennäköisyyttä $P(A)$. Olkoon $|A|$ tapauksen $A$ alkeistapausten lukumäärä ja $|\Omega|$ otosavaruuden alkeistapausten lukumäärä. Tällöin, jos kokeen kaikki alkeistapaukset ovat yhtä todennäköisiä, tapauksen $A$ todennäköisyys on $P(A) = \frac{|A|}{|\Omega|}$.

Tapauksen $A$ alkeistapauksia sanotaan myös *suotuisiksi* alkeistapauksiksi. Todennäköisyyden klassinen määritelmä voidaan täten kirjoittaa myös seuraavasti:

$$P(A) = \frac{\text{suotuisien alkeistapauksien lukumäärä}}{\text{kaikkien alkeistapauksien lukumäärä}}.$$

Todennäköisyyksiä voidaan laskea myös kokeellisten havaintojen pohjalta. Esimerkiksi nopan heiton tuloksia voitaisiin tilastoida ja laskea kunkin silmäluvun tilastollinen todennäköisyys. Havaintoaineistosta laskettavan tapauksen todennäköisyys on vastaavan tilastomuuttujan arvon suhteellinen frekvenssi.

**Määritelmä 3.3** (*Tilastollinen todennäköisyys*) Jos kaikkien havaintojen lukumäärä on $n$ ja tapaukselle $A$ suotuisten tapahtumien lukumäärä on $k$, niin tapauksen $A$ todennäköisyys on

$$P(A) = \frac{\text{suotuisien havaintojen lukumäärä}}{\text{kaikkien havaintojen lukumäärä}} = \frac{k}{n}.$$

Kun alkeistapauksia on ylinumeroituva määrä, niin todennäköisyyttä ei voida määrittää lukumääriä vertaamalla. Joskus alkeistapausten joukkoa voidaan silti kuvata geometristen mittojen, kuten pituuden, pinta-alan, tilavuuden ja kulman suuruuden avulla ja tällöin todennäköisyyksiä voidaan laskea mittojen suhteina. Esimerkiksi tikanheitossa, olettaen kaikkien taulun pisteiden olevan yhtä todennäköisiä ja heiton aina osuvan tauluun, saataisiin napakymppiheiton todennäköisyys keskiympyrän ja tikkataulun pinta-alojen suhteena.

**Määritelmä 3.4** (*Geometrinen todennäköisyys*) Tapahtuman $A$ todennäköisyys

$$P(A) = \frac{\text{suotuisien osien geometrinen mitta}}{\text{koko kuvion geometrinen mitta}}.$$

### 3.1.2 Joukko-oppia

Monet todennäköisyyslaskennan käsitteistä ja tuloksista perustuvat joukko-oppiin, minkä vuoksi joukko-opin peruskäsitteet ja -merkinnät sekä niiden todennäköisyystulkinnat ovat perusteltuja esitellä ennen siirtymistä todennäköisyyden ominaisuuksiin ja laskusääntöihin.

Joukko-opin peruskäsitteitä ovat *joukko* ja *alkio*. Jos alkio $a$ on joukon $A$ alkio (eli alkio $a$ kuuluu joukkoon $A$), merkitään $a \in A$. Vastaavasti jos alkio $a$ ei kuulu joukkoon $A$, merkitään $a \notin A$. Silloin, kun joukon alkioiden lukumäärä voidaan laskea, joukko on *äärellinen*. Muulloin joukko on *ääretön*.

**Määritelmä 3.5** Joukko $A$ on *sama joukko* kuin joukko $B$, merkitään $A = B$, jos joukot $A$ ja $B$ muodostuvat tarkalleen samoista alkioista. Muussa tapauksessa merkitään $A \neq B$. Jos jokainen joukon $A$ alkio kuuluu myös joukkoon $B$, joukko $A$ on joukon $B$ *osajoukko*, merkitään $A \subseteq B$. Jos lisäksi $A \neq B$, niin joukko $A$ on joukon $B$ *aito osajoukko*, merkitään $A \subset B$.

**Huomautus 3.1** Todennäköisyyslaskennassa perusjoukkoa nimitetään usein otosavaruudeksi, joukoista puhutaan tapahtumina tai tapauksina ja joukon alkioita nimitetään alkeistapauksiksi.

**Määritelmä 3.6** Joukkojen $A$ ja $B$ *leikkaus* $A \cap B$ sisältää ne alkeistapaukset, jotka ovat sekä joukossa $A$ että joukossa $B$. Tapaus $A \cap B$ siis sattuu, jos molemmat tapauksista $A$ ja $B$ sattuvat.

**Määritelmä 3.7** Joukkojen $A$ ja $B$ *unioni* (*yhdiste*) $A \cup B$ sisältää ne alkeistapaukset, jotka ovat ainakin toisessa joukoista $A$ ja $B$. Tapaus $A \cup B$ siis sattuu, jos ainakin toinen tapauksista $A$ ja $B$ sattuu.

*Tyhjään joukkoon*, merkitään $\emptyset$, ei kuulu yhtään alkiota. Tapauksena tyhjä joukko $\emptyset$ ei koskaan satu ja sitä sanotaan *mahdottomaksi* tapaukseksi. Usein joukkoja tarkastellaan jonkin yhteisen laajemman joukon osajoukkoina. Tällaista joukkoa sanotaan perusjoukoksi $\Omega$. Koko perusjoukko on *varma* tapaus eli se sattuu aina. Nyt joukon $A$ komplementti on joukko $A^C = \Omega - A$.

**Määritelmä 3.8** Tapaus $A^C$ on tapauksen $A$ *komplementtitapaus (vastatapahtuma)* eli se sisältää ne alkeistapaukset, jotka eivät sisälly tapaukseen $A$. Tapaus $A^C$ siis sattuu, jos tapaus $A$ ei satu.

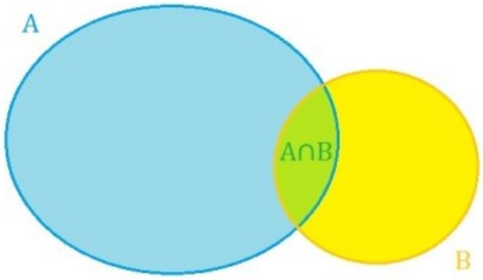
**Määritelmä 3.9** Joukkojen $A$ ja $B$ *erotus* $A - B = A \cap B^C$ sisältää ne joukon $A$ alkeistapaukset, jotka eivät kuulu joukkoon $B$. Tapaus $A - B$ siis sattuu, jos tapaus $A$ sattuu ja tapaus $B$ ei satu.

Seuraava määritelmä on todennäköisyyslaskennassa erittäin tärkeä.

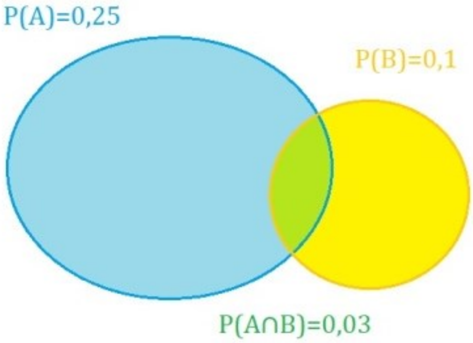
**Määritelmä 3.10** Tapaukset A ja B ovat *toisensa poissulkevat* eli *erilliset* tapahtumat, jos $A \cap B = \emptyset$.

Toisin sanoen, jos $A$ ja $B$ ovat toisensa poissulkevia tapauksia, niin tapahtuman $A$ tapahtuessa tapahtuma $B$ ei voi tapahtua ja toisaalta tapahtuman $B$ tapahtuessa tapahtuma $A$ ei voi tapahtua. Yleisemmin tapaukset $A_i$ ovat toisensa poissulkevat, jos $A_i \cap A_j = \emptyset$ kaikilla $i \neq j$.

Tapauksia (joukkoja) havainnollistetaan usein nk. Venn-diagrammeina (Kuva 3.1a ja Kuva 3.1b). Venn-diagrammiin voidaan myös merkitä tapaukseen kuuluvien alkeistapauksien lukumäärä tai tapauksen todennäköisyys.



*Kuva 3.1a* Venn-diagrammi



*Kuva 3.1b* Venn-diagrammi, johon merkitty tapausten todennäköisyydet.

Seuraavia, unionille ja leikkaukselle päteviä *distributiivilakeja* sekä komplementille päteviä *de Morganin lakeja* ei esitellä lukion todennäköisyyslaskennan opintojaksolla. Ne ovat kuitenkin välttämättömiä tuntea useiden todennäköisyyslaskennan lauseiden ja niiden todistusten ymmärtämiseksi, minkä vuoksi opettajan on hyvä tietää nämä joukkojen algebran laskusäännöt.

**Lause 3.1** (*Distributiivilait*)

a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cup (\bigcap_{i=1}^{n} B_i) = \bigcap_{i=1}^{n}(A \cup B_i)$

b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \quad A \cap (\bigcup_{i=1}^{n} B_i) = \bigcup_{i=1}^{n}(A \cap B_i)$

*Todistus.*

a) $x \in A \cup (B \cap C) \quad \Leftrightarrow$

$x \in A\ tai\ x \in\ B \cap C \quad \Leftrightarrow$

$x \in A\ tai\ (x \in\ B\ ja\ x \in\ C) \quad \Leftrightarrow$

$(x \in A\ tai\ x \in B)\ ja\ (x \in A\ tai\ x \in C) \quad \Leftrightarrow$

$x \in A \cup B\ ja\ x \in A \cup C \quad \Leftrightarrow$

$x \in (A \cup B) \cap (A \cup C).$

Yleinen tapaus todistettaisiin vastaavasti induktiolla.

b) $x \in A \cap (B \cup C) \quad \Leftrightarrow$

$x \in A\ ja\ x \in\ B \cup C \quad \Leftrightarrow$

$x \in A\ ja\ (x \in\ B\ tai\ x \in\ C) \quad \Leftrightarrow$

$(x \in A\ ja\ x \in B)\ tai\ (x \in A\ ja\ x \in C) \quad \Leftrightarrow$

$x \in A \cap B\ tai\ x \in A \cap C \quad \Leftrightarrow$

$x \in (A \cap B) \cup (A \cap C).$

Yleinen tapaus todistettaisiin vastaavasti induktiolla. ■

**Lause 3.2** (*de Morganin lait*)

a) $(A \cup B)^C = A^C \cap B^C, \quad (\bigcup_{i=1}^{n} A_i)^C = \bigcap_{i=1}^{n} {A_i}^C$

b) $(A \cap B)^C = A^C \cup B^C. \quad (\bigcap_{i=1}^{n} A_i)^C = \bigcup_{i=1}^{n} {A_i}^C$

*Todistus.*

a) $x \in (A^C \cap B^C)^C \quad \Leftrightarrow$

$x \notin A^C \cap B^C \quad \Leftrightarrow$

$x \notin A^C \; ja \; x \notin B^C \quad \Leftrightarrow$

$x \in A \; tai \; x \in B \quad \Leftrightarrow$

$x \in A \cup B.$

Yleinen tapaus todistettaisiin vastaavasti induktiolla.

b) $x \in (A^C \cup B^C)^C \quad \Leftrightarrow$

$x \notin A^C \cup B^C \quad \Leftrightarrow$

$x \notin A^C \; tai \; x \notin B^C \quad \Leftrightarrow$

$x \in A \; ja \; x \in B \quad \Leftrightarrow$

$x \in A \cap B.$

Yleinen tapaus todistettaisiin vastaavasti induktiolla. ■

De Morganin lakien mukaan siis unioni voidaan määritellä leikkauksen ja komplementin avulla ja joukkojen leikkaus unionin ja komplementin avulla.

### 3.1.3 Todennäköisyyden aksioomat ja ominaisuudet

Seuraavaksi esitetään todennäköisyydeltä vaadittavia, perustavanlaatuisia ominaisuuksia eli *aksioomia*, joista saadaan myöhemmin johdettua todennäköisyyden laskusääntöjä.

**Todennäköisyyden aksioomat**

**A1.** Todennäköisyys $P(A) \geq 0$ kaikille tapauksille $A$.

**A2.** $P(\Omega) = 1$.

**A3.** Jos tapaukset $A_i$ , jossa $i = 1,2, \dots$ , ovat toisensa poissulkevat,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

On luontevaa olettaa aksiooman A1 tapaan todennäköisyyden olevan ei-negatiivinen. Aksiooma A2 normeeraa todennäköisyydet siten, että suurin mahdollinen todennäköisyys on 1. Aksioomassa A3 yhtälön vasemmalla puolella esiintyvä tapaus $\bigcup_{i=1}^{\infty} A_i$ sattuu, jos ainakin jokin tapauksista $A_i$ sattuu. Kuitenkin, tapausten $A_i$ ollessa toisensa poissulkevia, tapaus $\bigcup_{i=1}^{\infty} A_i$ sattuu vain, jos tarkalleen yksi tapauksista $A_i$ sattuu. Täten on luontevaa, että tapauksien $A_i$ todennäköisyydet lasketaan yhteen eli $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. Näiden kolmen aksiooman avulla voidaan todistaa seuraavassa lauseessa esiteltävät todennäköisyyden ominaisuudet.

**Lause 3.3** Todennäköisyydellä on seuraavat ominaisuudet:

a) $P(\emptyset) = 0.$

b) Jos $A \cap B = \emptyset$, niin $P(A \cup B) = P(A) + P(B).$

c) $P(A) = 1 - P(A^C)$ $(komplementtikaava).$

d) Jos $A$ sisältyy tapaukseen $B$, niin $P(A) \leq P(B).$

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

f) $P(A - B) = P(A) - P(A \cap B).$

*Todistus.*

a) Valitaan aksioomassa $A3$, että $A_i = \emptyset$. Tällöin
$P(\bigcup_{i=1}^{\infty} \emptyset) = \sum_{i=1}^{\infty} P(\emptyset)$, eli
$P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset)$, josta seuraa välttämättä, että
$P(\emptyset) = 0.$

b) Valitaan aksioomassa A3, että $A_1 = A,\ A_2 = B$ ja $A_i = \emptyset$, kun $i \geq 3$.
$P(A \cup B \cup \emptyset \cup \dots) = P(A) + P(B) + P(\emptyset) + \cdots$, josta a-kohdan nojalla saadaan
$P(A \cup B) = P(A) + P(B).$

c) Koska $\Omega = A \cup A^C$, niin $P(\Omega) = P(A \cup A^C)$. Aksiooman A2 mukaan $P(\Omega) = 1$, joten
$P(A \cup A^C) = 1$. Tiedetään myös, että $A \cup A^C = \emptyset$, josta b-kohdan nojalla saadaan
$P(A \cup A^C) = P(A) + P(A^C)$. Näin ollen $1 = P(A) + P(A^C).$

d) Jos A sisältyy tapaukseen $B$, niin $B = A \cup (B - A)$, joten $P(B) = P[A \cup (B - A)]$
Koska $A \cap (B - A) = \emptyset$, niin b-kohdan mukaan
$P[A \cup (B - A)] = P(A) + P(B - A).$
Aksiooman A1 nojalla $P(B - A) \geq 0$, joten $P(A) + P(B - A) \geq P(A).$
Siis $P(B) \geq P(A).$

e) Aksiooman A1 mukaan $P(A) \geq 0$. Koska tapaus $A$ sisältyy otosavaruuteen $\Omega$, niin d-kohdan mukaan $P(A) \leq P(\Omega)$. Aksiooma A2 normittaa otosavaruuden todennäköisyyden $P(\Omega) = 1$, joten on oltava $P(A) \leq 1.$

f) Koska $A = (A \cap B) \cup (A - B)$, missä $(A \cap B) \cap (A - B) = \emptyset$, niin b-kohdan mukaan $P(A) = P(A \cap B) + P(A - B)$. ■

**Huomautus 3.1**

a) Tapaus $\emptyset$ on mahdoton tapaus ja sen todennäköisyys $P(\emptyset) = 0$ aina. On kuitenkin hyvä huomata, että vaikka joillekin tapauksille $A$ olisi $P(A) = 0$, niin siitä ei automaattisesti seuraa, että $A = \emptyset$. On olemassa ei-tyhjiä tapauksia, joiden todennäköisyys on nolla. Jos siis $P(A) = 0$, niin sanotaan tapauksen $A$ olevan *melkein mahdoton* tapaus.

b) Tapaus $\Omega$ on varma tapaus ja $P(\Omega) = 1$. Kuten edeltävässä tapauksessa, on myös tässä huomattava, että jos jollekin tapaukselle pätee $P(A) = 1$, niin siitä ei välttämättä seuraa, että $A = \Omega$. Sanotaan, että $P(A) = 1$ on *melkein varma* tapaus.

Seuraavan lauseen a-kohta yleistää Lauseen 3.3 b-kohdan useamman tapauksen unionille.

**Lause 3.4** Oletetaan, että tapaukset $A_i$ , jossa *i = 1, …, n*, ovat toisensa poissulkevat.

a) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ *(summakaava).*

b) Jos lisäksi $\bigcup_{i=1}^{n} A_i = \Omega$ , niin $\sum_{i=1}^{n} P(A_i) = 1$.

*Todistus.*

a) Valitaan aksioomassa A3 $A_i = \emptyset$, kun $i \geq n + 1$. Tällöin

$$P(\bigcup_{i=1}^{n} A_i) = P[(\bigcup_{i=1}^{n} A_i) \cup (\bigcup_{i=n+1}^{\infty} \emptyset)] \text{ (josta A3: n nojalla saadaan)} \quad .$$
$$= \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset)$$
$$= \sum_{i=1}^{n} P(A_i).$$

b) Aluksi a-kohdan nojalla saadaan

$$\sum_{i=1}^{n} P(A_i) = P(\bigcup_{i=1}^{n} A_i) \text{ (jolloin oletuksesta seuraa)}$$

$$= P(\Omega) \text{ (josta A2: n nojalla saadaan)}$$

$$= 1. \blacksquare$$

On huomionarvoista, että Lauseen 3.4 b-kohdan antamaa tulosta voidaan käyttää tarkastuskeinona silloin, kun kaikkien tapausten todennäköisyydet on laskettu. Tämän vuoksi on usein hyödyllistä laskea kaikkien tapauksien todennäköisyydet, vaikka tarvittaisiin vain tietyn tapauksen todennäköisyys. Jos tapausavaruus voidaan jakaa toistensa poissulkevien tapausten unioniksi ja kaikkien tapauksien todennäköisyyksien summa on erisuuri kuin 1, on laskuissa varmasti jokin virhe. Kyseistä tulosta voidaan käyttää myös yhden todennäköisyyden $P(A_i)$ ratkaisemiseksi yhtälöstä $\sum_{i=1}^{n} P(A_i) = 1$, kun muiden tapausten todennäköisyydet tunnetaan. Tämän käyttötavan varjopuolena on hyvän tarkistuskeinon menettäminen.

Myös Lauseen 3.3 e-kohta on erittäin tärkeä todennäköisyyksien tarkistamisessa, sillä jos laskujen tuloksena saatu todennäköisyys on negatiivinen tai suurempi kuin 1, on laskuvirhe varma tapaus.

## 3.2 Kombinatoriikkaa

Käytettäessä klassisen todennäköisyyden kaavaa $P(A) = \frac{|A|}{|\Omega|}$ joudutaan selvittämään tapauksen $A$ ja otosavaruuden $\Omega$ alkeistapauksien lukumäärät. Välillä lukumäärät saadaan selville helposti yksinkertaisen päättelyn avulla tai luettelemalla eri mahdollisuudet, mutta usein on sujuvampaa käyttää kombinatoriikan tuloksia.

**Määritelmä 3.11** Joukosta voidaan ottaa alkioita, eli suorittaa *otanta*, kahdella tavalla. Otanta tehdään *palauttamatta*, jos joukosta otetaan pois yksi alkio kerrallaan. Otanta tehdään *palauttaen*, jos jokaisen alkion poisoton jälkeen alkio palautetaan joukkoon.

**Määritelmä 3.12** Joukko on *järjestämätön*, jos sen alkioiden järjestyksellä ei ole väliä. Joukko on *järjestetty*, jos alkioiden järjestyksen muuttuessa myös joukko muuttuu.

Kun joukosta valitaan alkioita kiinnittämättä huomiota niiden järjestykseen, saadaan osajoukkoja eli *kombinaatioita*. Kun taas joukosta valitut alkiot järjestetään jonoon, jossa alkioiden keskinäinen järjestys on merkitsevä, puhutaan *permutaatioista*.

**Määritelmä 3.13** Joukon alkioiden *k-kombinaatio* muodostetaan ottamalla joukosta $k$:n alkion otos (palauttamatta tai palauttaen) ja muodostamalla siitä järjestämätön joukko. Joukon alkioiden *k-permutaatio* muodostetaan ottamalla joukosta $k$:n alkion otos (palauttamatta tai palauttaen) ja muodostamalla siitä järjestetty joukko.

Toisin sanoen $k$-kombinaatio on järjestämätön $k$-alkioinen joukko ja $k$-permutaatio on järjestetty $k$-alkioinen joukko. Jos kahdessa k-alkioisessa joukossa on täsmälleen samat alkiot, pidetään näitä joukkoja samoina. Esimerkiksi Lotto-arvonnan tulos on yksi joukon $\{1, 2, \ldots, 39\}$ 7-kombinaatioista palauttamatta ja Jokeri-arvonnan tulos on yksi joukon $\{1, 2, \ldots, 9\}$ 7-permutaatioista palauttaen.

**Lause 3.5** (*Tuloperiaate*) Jos operaatio $A_i$ voidaan tehdä $n_i$:llä eri tavalla, $i = 1, \ldots, k$, niin jono (operaatio $A_1, \ldots,$ operaatio $A_k$) voidaan tehdä $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ eri tavalla.

Lauseessa 3.5 oletetaan siis, että kokonaisuuden valitsemisessa on $k$ peräkkäistä vaihetta. Tällöin, jos ensimmäisessä vaiheessa on $n_1$ vaihtoehtoa joista valita, toisessa vaiheessa on $n_2$ vaihtoehtoa joista valita ja niin edelleen kunnes $k$. vaiheessa on $n_k$ vaihtoehtoa joista valita, niin kokonaisuus voidaan valita $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ eri tavalla.

**Lause 3.6** (*Järjestetty otanta palauttamatta*)

a) Joukosta, jossa on $n$ erilaista alkiota, voidaan ottaa $k$:n alkion järjestetty otos palauttamatta $n(n-1)(n-2)\cdots(n-k+1)$ eli $\frac{n!}{(n-k)!}$ eri tavalla. Tästä

   k-permutaatioiden lukumäärästä käytetään merkintää $P(n,k)$.

b) Jos joukossa on $n$ erilaista alkiota, niin alkiot voidaan järjestää jonoon $n!$ eri tavalla.

*Todistus.*

a) Otoksen ensimmäinen alkio voidaan valita $n$ eri tavalla, toinen alkio $n-1$ eri tavalla, kolmas alkio $n-2$ eri tavalla ja niin edelleen kunnes $k$. alkio voidaan valita $n-(k-1)$ eri tavalla. Täten tuloperiaatteen (Lause 3.5) nojalla $k$:n alkion otos voidaan valita $n(n-1)(n-2)\cdots(n-k+1)$ eri tavalla.

b) Jonon ensimmäinen alkio voidaan valita n:llä eri tavalla, toinen $n-1$ eri tavalla, kolmas $n-2$ eri tavalla ja niin edelleen kunnes jonon toiseksi viimeinen alkio voidaan valita kahdella eri tavalla ja viimeinen vain yhdellä tapaa. Näin ollen tuloperiaatteen nojalla $n$ alkiota voidaan järjestää jonoon $n(n-1)(n-2)\cdots 2\cdot 1 = n!$ eri tavalla.■

**Lause 3.7** (*Järjestetty otanta palauttaen*) Joukosta, jossa on $n$ erilaista alkiota, voidaan ottaa $k$:n alkion järjestetty otos palauttaen $n^k$ eri tavalla.

*Todistus.*

Otoksen ensimmäinen alkio voidaan valita $n$ eri tavalla, toinen alkio $n$ eri tavalla, kolmas alkio $n$ eri tavalla ja niin edelleen kunnes $k$. alkio voidaan valita $n$ eri tavalla. Täten tuloperiaatteen (Lause 3.5) nojalla $k$:n alkion otos voidaan valita $n^k$ eri tavalla. ■

**Lause 3.8** (*Järjestämätön otanta palauttamatta, binomikertoimen otantatulkinta*) Joukosta, jossa on $n$ erilaista alkiota, voidaan ottaa järjestämätön otos palauttamatta $\frac{n!}{k!(n-k)!}$ eri tavalla. Tästä $k$-kombinaatioiden lukumäärästä käytetään merkintöjä $\binom{n}{k}$ ja $C(n,k)$.

*Todistus.*

Joukon $k$-permutaatioita on $\frac{n!}{(n-k)!}$ kappaletta (Lause 3.6 a). Merkitään $k$-kombinaatioiden lukumäärää muuttujalla $x$. Tällöin $k$-permutaatioita on $k!\,x$ kappaletta. Siis $k!\,x = \frac{n!}{(n-k)!}$ jos ja vain jos $x = \frac{n!}{k!(n-k)!}$.

**Huomautus 3.1** Binomikertoimella $\binom{n}{k}$ on Lauseen 3.8 otantatukinnan lisäksi seuraavat hyödylliset tulkinnat:

a) *(Binomikertoimen järjestystulkinta)* $\binom{n}{k}$ on niiden tapojen lukumäärä, joilla $k$ kappaletta tyypin 1 ja $n-k$ kappaletta tyypin 2 alkioita voidaan järjestää jonoon.

b) *(Binomikertoimen lokerotulkinta)* $\binom{n}{k}$ on niiden tapojen lukumäärä, joilla $n$ erilaista alkiota voidaan jakaa kahteen lokeroon siten, että lokeroihin tulee $k$ ja $n-k$ alkiota.

## 3.3 Todennäköisyyden laskusääntöjä

Seuraavaksi tutustutaan tarkemmin lukion todennäköisyyslaskennan opintojaksolla esiteltäviin laskusääntöihin, joita ovat komplementtisääntö (Lause 3.3 c), yhteenlaskusääntö, kertolaskusääntö sekä toistokokeeseen liittyvän binomitodennäköisyyden laskukaava. Näiden laskusääntöjen oikea käyttö vaatii ehdollisen todennäköisyyden ja riippumattomuuden käsitteiden ymmärtämisen.

### 3.3.1 Yhteenlaskusääntö

Kun tapahtumat $A$ ja $B$ eivät ole erillisiä, tapahtuma $A \cup B$ eli "$A$ tai $B$" tarkoittaa, että ainakin toinen tapauksista $A$ tai $B$ sattuu. Kun lasketaan todennäköisyys $P(A \cup B)$ on summasta $P(A) + P(B)$ vähennettävä tapahtuman $A \cap B$ eli "$A$ ja $B$" todennäköisyys, jottei tapahtumille $A$ ja $B$ yhteiset alkeistapaukset tule mukaan kahteen kertaan.

Usein lukion pitkän matematiikan opintojaksolla esitellään seuraavasta lauseesta vain ensimmäinen tulos (a-kohta) tai kaksi ensimmäistä tulosta (a- ja b-kohdat).

**Lause 3.9** *(Mukaanlukemis-poissulkemis-periaate, yleinen yhteenlaskusääntö)*

a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

b) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

c) 
$$\begin{aligned} P(A \cup B \cup C \cup D) = {} & P(A) + P(B) + P(C) + P(D) \\ & -P(A \cap B) - P(A \cap C) - P(A \cap D) - P(B \cap C) - P(B \cap D) - P(C \cap D) \\ & +P(A \cap B \cap C) + P(A \cap B \cap D) + P(A \cap C \cap D) + P(B \cap C \cap D) \\ & -P(A \cap B \cap C \cap D) \end{aligned}$$

d) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) - + \cdots + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i)$

*Todistus.*

a) Koska $A \cup B = (A - B) \cup B$, missä $(A - B) \cap B = \emptyset$, niin Lauseen 3.3 b-kohdan mukaan $P(A \cup B) = P(A - B) + P(B)$, josta Lauseen 3.3 f-kohdan mukaan saadaan

$$P(A \cup B) = P(A) - P(A \cap B) + P(B).$$

b) $P(A \cup B \cup C) = P[A \cup (B \cup C)]$,

$$\overset{a-kohta}{\overset{\frown}{=}} P(A) + P(B \cup C) - P[A \cap (B \cup C)]$$

$$\overset{a-kohta}{\overset{\frown}{=}} P(A) + [P(B) + P(C) - P(B \cap C)] - P[A \cap (B \cup C)]$$

$$\overset{\substack{distributiivisuus \\ (Lause\ 3.1)}}{\overset{\frown}{=}} P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)]$$

$$\overset{a)}{\overset{\frown}{=}} P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cup B \cup C).$$

c) b-kohdan todistuksen tavoin.

d) yleinen tapaus todistettaisiin vastaavasti induktiolla. ■

Kun tapaukset A ja B ovat erilliset eli toisensa poissulkevat, Lauseen 3.9 d-kohta yksinkertaistuu muotoon $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$, joka on itse asiassa Lauseen 3.4 a-kohdan tulos eli toisensa poissulkevien tapausten yhteenlaskusääntö eli nk. summakaava.

### 3.3.2 Ehdollinen todennäköisyys ja kertolaskusääntö

Ehdolliseen todennäköisyyteen liittyvästä ongelmasta on kyse silloin, kun halutaan laskea jonkin tapauksen *A* todennäköisyys silloin, kun tapaus *B* on jo sattunut. Ehdollista todennäköisyyttä merkitään *P(A|B)*, jonka voi lukea esimerkiksi seuraavasti: "todennäköisyys, että *A* tapahtuu, kun *B* on tapahtunut". Ehdollinen todennäköisyys *P(A|B)* voidaan määritellä kahdella tavalla; joko *A*:n todennäköisyytenä *supistetussa otosavaruudessa B* tai *alkuperäisen otosavaruuden Ω* todennäköisyyksien avulla.

**Määritelmä 3.14a** Jos *P(B) > 0*, niin *ehdollinen todennäköisyys P(A|B)* on tapauksen *A* todennäköisyys otosavaruudessa *B*.

**Määritelmä 3.14b** Jos *P(B) > 0*, niin *ehdollinen todennäköisyys P(A|B)* on

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Jälkimmäisen määrittelytavan avulla voidaan johtaa seuraavaksi esiteltävät ehdolliseen todennäköisyyteen liittyvät tulokset.

**Lause 3.10** *(Kertolaskusääntö)* Tapausten A ja B leikkauksen todennäköisyys toteuttaa kaavat

$P(A \cap B) = P(A|B)P(B)$ ja

$P(A \cap B) = P(B|A)P(A).$

*Todistus.*

Leikkauksen todennäköisyys saadaan ratkaisemalla ehdollisen todennäköisyyden *P(A|B)* kaavasta leikkauksen P(A∩B) todennäköisyys. Vastaavasti leikkauksen P(A∩B) todennäköisyys saadaan myös ratkaistua ehdollisen todennäköisyyden *P(B|A)* kaavasta seuraavasti

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{jos } P(A) > 0$$

jolloin

$$P(A \cap B) = P(B|A)P(A). \ ■$$

Seuraavaksi esiteltävät Määritelmä 3.15, Lause 3.11 ja Lause 3.12 eivät yleensä kuulu lukion pitkän matematiikan Tilastot ja todennäköisyys -opintojakson sisältöön. Ne ovat kuitenkin ehdollisen todennäköisyyden yleistyksiä, joiden tunteminen lisää aihepiirin soveltamisen mahdollisuuksia opetusta ylöspäin eriyttäessä. Lauseiden tuntemista tarvitaan myös yhdessä opiskelijamateriaalin harjoituksessa, minkä vuoksi niihin tutustuminen on tässä yhteydessä oleellista.

**Määritelmä 3.15** Tapaukset $A_1, \dots, A_n$ muodostavat otosavaruuden Ω osituksen, jos tapaukset ovat

- toisensa poissulkevat
- $\Omega = \bigcup_{i=1}^{n} A_i$ ja
- $P(A_i) > 0$ *kaikilla i.*

**Lause 3.11** (*Kokonaistodennäköisyyslause*) Jos tapaukset $A_1, \dots, A_n$ muodostavat otosavaruuden osituksen, niin

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\, P(A_i).$$

*Todistus.*

Osituksen määritelmää (Määritelmä 3.15) ja summakaavaa (Lause 3.4 a) apuna käyttäen saadaan:

$$P(B) = P(B \cap \Omega)$$

$$= \mathrm{P}\big(B \cap (\bigcup_{i=1}^{n} A_i)\big)$$

$$= P(\bigcup_{i=1}^{n}(B \cap A_i))$$

$$= \sum_{i=1}^{n} P(B \cap A_i)$$

$$= \sum_{i=1}^{n} P(B|A_i)\, P(A_i)\,. \blacksquare$$

Lauseen 2.11 kokonaistodennäköisyyskaavalla kokonaistodennäköisyys *P(B)* saadaan koottua ehdollisista todennäköisyyksistä painottamalla kutakin itse ehdon todennäköisyydellä.

Keksijänsä mukaan nimetty, 1700-luvulla eläneen matemaatikon Thomas Bayesin keksimä kaava on usein hyödyllinen ehdollisia todennäköisyyksiä laskettaessa.

**Lause 3.12** (*Bayesin lause*) Jos tapaukset $A_1, \dots, A_n$ muodostavat otosavaruuden osituksen ja $P(B) > 0$

$$P(A|B) = \frac{P(B|A_k)\, P(A_k)}{P(B)} = \frac{P(B|A_k)\, P(A_k)}{\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}\,, k = 1, \dots, n.$$

*Todistus.*

Ehdollisen todennäköisyyden määritelmästä (Määritelmä 3.14 b) ja Lauseesta 3.10 seuraa

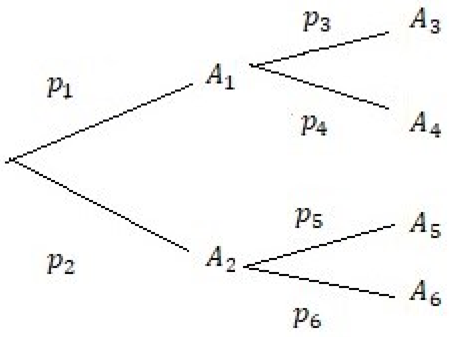
$$P(A|B) = \frac{P(B \cap A_k)}{P(B)} = \frac{P(B|A_k)\, P(A_k)}{P(B)}\,.$$

Koska tapaukset muodostavat otosavaruuden osituksen, niin *kokonaistodennäköisyyslauseen* (Lause 3.11) mukaan *P(B)* voidaan koota ehdollisista todennäköisyyksistä *P(B* $|A_i)$ painottamalla kutakin ehdollista todennäköisyyttä ehdon todennäköisyydellä *P(*$A_i$*)*. Siis

$$P(A|B) = \frac{P(B|A_k)\, P(A_k)}{P(B)} = \frac{P(B|A_k)\, P(A_k)}{\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}\,. \blacksquare$$

Monessa todennäköisyyslaskennan tehtävän ratkaisussa voidaan käyttää luontevasti apuna ns. *puukaavioita* (Kuva 3.2), joilla havainnollistetaan satunnaisilmiön vaihtoehtoisia tapahtumajonoja.

Puu koostuu oksista ja niiden haarautumiskohdista eli pisteistä. Puun piste kuvaa tapausta $A_i, i = 1, 2, \ldots, k, k \in \mathrm{Z}_+$ ja siitä lähtevät oksat tapahtumavaihtoehtoja, jotka ovat mahdollisia aiemmin tapahtuneen tapauksen jälkeen. Oksan viereen kirjataan myös kyseisen tapahtumavaihtoehdon todennäköisyys $p_i, i = 1, 2, \ldots, k, k \in \mathrm{Z}_+$. Mistä tahansa puun pisteestä lähtevien oksien todennäköisyyksien summa on 1. Tavallisesti puu piirretään ylösalaisin (alkupiste ylhäällä, loppupisteet alhaalla) tai horisontaalisesti (alkupiste vasemmalla, loppupisteet oikealla). Mielivaltaisen puun pisteen todennäköisyys saadaan laskemalla alkupisteestä kyseiseen pisteeseen vievän reitin todennäköisyys. Reitti on tapahtumajono, joka sattuu, jos jokainen jonon tapahtumista sattuu, joten reitin todennäköisyys saadaan tulosääntöä (Lause 3.5) puun oksiin soveltamalla. Koska puun eri pisteisiin vievät reitit ovat toisensa poissulkevia, niin useista puun pisteistä yhdistetyn tapahtuman todennäköisyys saadaan yhteenlaskusääntöä (Lause 3.4 a) käyttäen.



*Kuva 3.2.* Puukaavio.

### 3.3.3 Riippumattomuus

Yleisesti ehdollinen todennäköisyys *P(A|B)* riippuu tapauksesta *B*, jolloin *P(A|B)* on eri kuin *P(A)*. Toisinaan kuitenkin *P(A|B)=P(A)*, jolloin sanotaan tapauksen *A* olevan *riippumaton* tapauksesta *B*. Jos *P(A|B)=P(A)*, niin tällöin myös *P(B|A)=P(B)* eli tapaus *B* on myös riippumaton tapauksesta *A*, sillä jos $P(A|B) = P(A)$, niin

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B).$$

Sanotaan siis, että jos *P(A|B)=P(A)*, niin tapaukset *A* ja *B* ovat *riippumattomat*. Kaava *P(A|B)=P(A)* voitaisiin ottaa riippumattomuuden määritelmäksi, mutta yleisesti käytetään määritelmää $P(A \cap B) = P(A)P(B)$, joka on aiemmin mainitun kanssa ekvivalentti, sillä jos $P(A|B) = P(A)$, niin

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B),$$

ja jos $P(A \cap B) = P(A)P(B)$, niin

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

**Määritelmä 3.16** Tapaukset $A$ ja $B$ ovat *riippumattomat*, jos $P(A \cap B) = P(A)P(B)$. Jos tapaukset eivät ole riippumattomat, ne ovat *riippuvat.*

Tapausten $A$ ja $B$ riippumattomuutta voidaan merkitä $A\perp B$.

**Lause 3.13** Jos tapaukset $A$ ja $B$ ovat riippumattomat, niin riippumattomia ovat myös tapaukset $A$ ja $B^C$, $A^C$ ja $B$ sekä $A^C$ ja $B^C$.

$Todistus.$

i. Koska $A = (A \cap B^C) \cup (A \cap B)$, niin Lauseen 3.3 b-kohdan nojalla

$$\begin{aligned} P(A) &= P(A \cap B^C) + P(A \cap B) \\ &= P(A \cap B^C) + P(A)P(B). \end{aligned}$$

Täten $P(A \cap B^C) = P(A)\big(1 - P(B)\big) = P(A) \cdot P(B^C)$ eli tapaukset A ja $\mathrm{B^C}$ ovat riippumattomat.

ii. Vastaavasti, koska $B = (B \cap A^C) \cup (B \cap A)$, niin Lauseen 3.3 b-kohdan nojalla

$$\begin{aligned} P(B) &= P(B \cap A^C) + P(B \cap A) \\ &= P(B \cap A^C) + P(B)P(A). \end{aligned}$$

Nyt $P(B \cap A^C) = P(B)\big(1 - P(A)\big) = P(B) \cdot P(A^C)$ eli tapaukset $\mathrm{A^C}$ ja B ovat myös riippumattomat.

iii. Samalla lailla, koska $A^C = (A^C \cap B) \cup (A^C \cap B^C)$, niin Lauseen 3.3 b-kohdan nojalla

$P(A^C) = P(A^C \cap B) + P(A^C \cap B^C)$ ja koska ii-kohdan nojalla $A^C\perp B$, niin

$= P(A^C)P(B) + P(A^C \cap B^C)$.

Siis $P(A^C \cap B^C) = P(A^C)\big(1 - P(B)\big) = P(A^C) \cdot P(B^C)$ eli tapaukset $A^C$ ja $\mathrm{B^C}$ ovat riippumattomat. ■

Kahden tapauksen riippumattomuus voidaan yleistää myös usean tapauksen riippumattomuudeksi.

**Määritelmä 3.17** Tapaukset $A_i, i = 1, \dots, n$, ovat riippumattomat, jos seuraavat ehdot toteutuvat:

$P\big(A_i \cap A_j\big) = P(A_i)P\big(A_j\big)$ kaikilla $i \neq j$,

$P\big(A_i \cap A_j \cap A_k\big) = P(A_i)P\big(A_j\big)P(A_k)$ kaikilla erisuurilla $i, j$ ja $k$.

$$\vdots$$

$$P(A_i \cap A_j \cap \ldots \cap A_n) = P(A_i)P(A_j)\cdots P(A_n).$$

**Huomautus 3.2** Kertolaskusääntö (Lause 3.9) yksinkertaistuu siis riippumattomille tapauksille muotoon $P(A \cap B) = P(A)P(B)$, mikä on myös riippumattomuuden määritelmä. Kaavasta $P(A \cap B) = P(A)P(B)$ käytetään usein nimityksiä *riippumattomien tapahtumien kertolaskusääntö*, *kertolaskukaava* ja *(kahden tapauksen) tulokaava*. Tätä kaavaa sekä useamman tapauksen riippumattomuuden kaavoja voidaan käyttää leikkausten todennäköisyyksien laskemiseen, jos tiedetään satunnaiskokeen tapahtumien olevan toisistaan riippumattomia. Riippumattomuuden määritelmien (Määritelmät 3.16 ja 3.17) avulla voidaan myös testata, ovatko annetut tapaukset riippumattomat.

**Lause 3.14** (*Tulokaava l. yleinen riippumattomien tapausten kertolaskusääntö*) Jos tapaukset $A_i, i = 1, \ldots, n$, ovat riippumattomat, niin

a) $P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$.

b) $P(\bigcup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} P(A_i{}^C)$.

*Todistus.*

a) Saadaan suoraan usean tapauksen riippumattomuuden määritelmästä.

b) Oletetaan, että tapaukset $A_i, i = 1, \ldots, n$, ovat riippumattomat. Tällöin

$$\begin{aligned} P(\textstyle\bigcup_{i=1}^{n} A_i) &= 1 - P[(\textstyle\bigcup_{i=1}^{n} A_i)^C] \\ &= 1 - P(\textstyle\bigcap_{i=1}^{n} A_i{}^C) \\ &\overset{Lause\ 3.13}{\cong} 1 - \textstyle\prod_{i=1}^{n} P(A_i{}^C). \blacksquare \end{aligned}$$

### 3.3.4 Toistokoe

Yhdistettyä satunnaiskoetta, jossa täsmälleen samaa koetta toistetaan $n$ kertaa ja seurataan onnistuneiden kokeiden lukumäärää, kutsutaan *toistokokeeksi*. Toistokokeessa tuloksen $X$ esiintymistodennäköisyys $p$ on jokaisella toistolla sama ja riippumaton aiempien kierrosten tuloksista. Tarkastellaan seuraavaksi toistokokeen todennäköisyyttä (binomitodennäköisyyttä) eli sitä, millä todennäköisyydellä $n$-vaiheisessa toistokokeessa tapahtuma $X$ toteutuu tasan $k$ kertaa.

**Lause 3.15** (*Binomitodennäköisyys*) Koe toistetaan riippumattomasti $n$ kertaa. Kukin koe onnistuu todennäköisyydellä $p$; merkitään epäonnistumisen todennäköisyyttä $q = 1 - p$. Olkoon X onnistuneiden kokeiden lukumäärä. Tällöin k onnistumista saadaan todennäköisyydellä

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, 2, \dots, n.$$

*Todistus.*

Merkitään onnistunutta koetta kirjaimella $O$ ja epäonnistunutta koetta kirjaimella $E$. Yksi sellainen koesarja, jossa on $k$ onnistunutta ja $n - k$ epäonnistunutta koetta on $(O, O, \dots, O, E, E, \dots, E)$, missä onnistuneita kokeita $O$ on $k$ kappaletta ja epäonnistuneita kokeita $E$ $n - k$ kappaletta. Koska kokeet ovat toisistaan riippumattomia, niin kyseisen koesarjan todennäköisyys on $p^k q^{n-k}$. Binomikertoimen järjestystulkinnan (Huomautus 3.1 a) mukaan symbolit $O$ ja $E$ voidaan järjestää jonoon $\binom{n}{k}$ eri tavalla. Koska kunkin koesarjan todennäköisyys on $p^k q^{n-k}$ ja koesarjat ovat toisensa poissulkevat, niin $P(X = k) = \binom{n}{k} p^k q^{n-k}$. ■

## 3.4 Todennäköisyysjakauma

Luvussa 2 kerrottiin tilastoyksikköön liitetyistä ominaisuuksista ja tilastojakaumien tunnusluvuista. Tässä alaluvussa tutustutaan vastaavasti todennäköisyyslaskennassa tutkittavien satunnaismuuttujien ominaisuuksiin ja tutkitaan todennäköisyyden jakautumista tapahtuman eri vaihtoehtojen kesken sekä esitellään todennäköisyysjakaumia kuvaavia tunnuslukuja.

### 3.4.1 Satunnaismuuttuja

Satunnaismuuttuja on funktio, jonka arvot määrää jokin satunnaisilmiö. Funktion määrittelyjoukko on satunnaisilmiön alkeistapausten joukko ja sen arvot ovat reaalilukuja.

**Määritelmä 3.18** Satunnaismuuttuja $X$ otosavaruudessa $\Omega$ on funktio $X: \Omega \rightarrow \Re$.

Satunnaismuuttuja-funktio siis liittää jokaiseen satunnaisilmiön alkeistapaukseen jonkin reaaliluvun, satunnaismuuttujan arvon.

**Määritelmä 3.19** Satunnaismuuttuja $X$ on *diskreetti*, jos se voi saada vain äärellisen tai numeroituvasti äärettömän määrän arvoja.

Diskreetin satunnaismuuttujan saamat arvot $x_1, x_2, \dots$ ovat siis erillisiä ja ne voidaan täten luetella. Esimerkiksi 'COVID19-tartuntatapausten lukumäärä Suomessa' on diskreetti satunnaismuuttuja, joka saa yksittäisen (kokonaisluku)arvon joka kuukausi. Moni suure, kuten 'potilaan jonotusaika päivystyksessä', voi kuitenkin saada periaatteessa minkä tahansa arvon tietyltä reaalilukuväliltä. Näiden suureiden mallintamisessa käytetään jatkuvia satunnaismuuttujia.

**Määritelmä 3.20** Satunnaismuuttuja $X$ on *jatkuva*, jos se voi saada ylinumeroituvan määrän arvoja: tyypillisesti minkä tahansa arvon joltain reaalilukuväliltä.

Määritelmä 3.20 on lukion opintojaksolla riittävän tarkka määritelmä jatkuvalle satunnaismuuttujalle. Luvussa 3.4.3 esitellään myös matemaattisempi määrittely (Määritelmä 3.28) jatkuvalle satunnaismuuttujalle niin kutsutun tiheysfunktion avulla.

Satunnaismuuttujan *(todennäköisyys)jakauma* sisältää satunnaismuuttujan arvot ja niiden todennäköisyydet. Kuten tilastoyksiköiden tilastollisilla muuttujilla, myös satunnaismuuttujilla jakauma ilmaisee, kuinka yleisiä satunnaismuuttujan eri arvot ovat.

Diskreetin satunnaismuuttujan jakauma voidaan usein esittää taulukkomuodossa. Eri arvojen todennäköisyydet muodostavat kyseisen muuttujan todennäköisyysjakauman, jota voidaan havainnollistaa esimerkiksi pylväsdiagrammilla. Jatkuvan satunnaismuuttujan $X$ arvot muodostavat jonkin reaaliakselin välin, joka sisältää äärettömän määrän lukuja. Tämän vuoksi jatkuvan satunnaismuuttujan jakauman esittäminen taulukossa ei ole luontevaa, vaan jakauma esitetään yleensä satunnaismuuttujan tiheysfunktion avulla. Tähän palataan luvussa 3.4.5.

### 3.4.2 Diskreetin satunnaismuuttujan todennäköisyys- ja jakaumafunktio

Merkitään diskreetin satunnaismuuttujan saamien arvojen joukkoa kirjaimella $K$. Diskreeteillä satunnaismuuttujilla muotoa $P(X = k), k \in K$, olevat todennäköisyydet ovat tärkeimpiä ja näille todennäköisyyksille määritelään oma funktionsa:

**Määritelmä 3.21** Diskreetin satunnaismuuttujan $X$ t*odennäköisyysfunktio* on $p(k) = P(X = k)$, $k \in K$.

Todennäköisyysfunktiosta käytetään yleisesti myös nimityksiä pistetodennäköisyysfunktio ja tiheysfunktio. Todennäköisyyden ominaisuuksien mukaisesti todennäköisyysfunktiolle $p(k)$ $0 \leq p(k) \leq 1, k \in K$ (Lause 3.3 e) ja koska diskreetin satunnaismuuttujan $X$ saamat arvot $k$ ovat toisensa poissulkevia tapauksia, niin $\sum_{k \in K} p(k) = 1$ (Lause 3.4 b).

Todennäköisyyslaskennassa ollaan usein kiinnostuneita todennäköisyyksistä, jotka ovat muotoa $P(X \leq a), P(X > a)$ ja $P(a < X \leq b)$. Muotoa $P(X \leq a)$ olevat todennäköisyydet ovat keskeisiä, sillä kaikki muut muodot voidaan ilmaista muotoa $P(X \leq a)$ olevien todennäköisyyksien avulla seuraavasti:

$P(X > a) = 1 - P(X \leq a)$ ja

jos $a < b$, niin $P(X \leq b) = P(X \leq a) + P(a < X \leq b)$, joten

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a).$$

Todennäköisyyden $P(X \leq a)$ keskeisyyden vuoksi sille on määritelty oma funktio, joka kertoo millä todennäköisyydellä satunaismuuttujan X arvo on pienempi tai yhtä suuri kuin jokin reaaliluku $x$.

**Määritelmä 3.22** Satunnaismuuttujan *X jakaumafunktio* on $F(x) = P(X \leq x), x \in \mathfrak{R}$.

Jakaumafunktiosta käytetään myös nimityksiä kumulatiivinen jakaumafunktio ja kertymäfunktio. Sekä diskreetin että jatkuvan satunnaismuuttujan jakaumafunktio on ei-negatiivinen ja kasvava funktio, jonka arvojoukko on suljettu väli [0,1] eli $0 \leq F(x) \leq 1, x \in \mathfrak{R}$. Lisäksi diskreetin satunnaismuuttujan jakaumafunktiolla F(x) on seuraavat ominaisuudet:

- $F(x)$ on oikealta jatkuva porrasfunktio,
- $\lim_{x \to \infty} F(x) = 1$,
- $\lim_{x \to -\infty} F(x) = 0$.

Todennäköisyysfunktiosta saadaan jakaumafunktio ja jakaumafunktiosta todennäköisyysfunktio seuraavasti:

$$F(x) = \sum_{k \leq x} p(k),$$

$$p(k) = F(k) - F(k-1).$$

**Huomautus 3.3** Todennäköisyysfunktio $p(k)$ määritellään vain pisteissä $k \in K$, kun taas jakaumafunktio $F(x)$ kaikilla $x \in \mathfrak{R}$.

### 3.4.3 Diskreetin jakauman tunnusluvut

Tilastomuuttujan jakauman informaation tiivistämisessä käytettiin tilastollisia tunnuslukuja, kuten moodia ja aritmeettista keskiarvoa. Vastaavasti satunnaismuuttujan jakauman tietoa tiivistetään tunnusluvuin, joista keskeisimpiä ovat odotusarvo, keskihajonta ja varianssi.

Jos lasketaan sellainen satunnaiskokeen tulosten painotettu keskiarvo, jossa kunkin tuloksen painona on vastaavan tapauksen todennäköisyys, niin saatua arvoa sanotaan koetuloksen *odotusarvoksi E(X)*. Merkinnän *E(X)* käyttö juontaa juurensa englannin kielen sanoihin "odotus" *expectation* ja "odotusarvo" *expected value*.

**Määritelmä 3.23** Diskreetin satunnaismuuttujan $X$ *odotusarvo* on

$$E(X) = \sum_{k \in K} k\, p(k),$$

jos $\sum_{k \in K} |k| p(k) < \infty$.

Odotusarvon määritelmässä siis vaaditaan summan $\sum_{k \in K} |k| p(k)$ suppenemista itseisesti, jotta summan termit voidaan laskea yhteen missä järjestyksessä tahansa samaan tulokseen päätyen. Jos nimittäin summa $\sum_{k \in K} |k| p(k)$ ei suppenisi itseisesti, se voisi saada erilaisia arvoja riippuen siitä, missä järjestyksessä termit lasketaan yhteen. Näin ollen, jotta odotusarvo olisi hyvin määritelty, vaaditaan että summa suppenee itseisesti. Jos näin ei toisaalta ole, niin sanotaan, että odotusarvoa ei ole olemassa.

Jos diskreetin satunnaismuuttujan $X$ arvojoukko $K$ on äärellinen, niin myös odotusarvo $E(X)$ on äärellinen, kun taas arvojoukon ollessa numeroituvasti ääretön, voi odotusarvon määritelmä antaa äärettömän tuloksen.

Lukion opintojaksolla esitellään Määritelmän 3.23 sijaan seuraava diskreetin satunnaismuuttujan $X$ odotusarvolle yksinkertaisempi määritelmä:

**Määritelmä 3.24** Olkoot diskreetin satunnaismuuttujan $X$ saamat arvot $x_1, x_2, \dots, x_k$ ja näiden pistetodennäköisyydet $p_i = P(X = x_i), i = 1, \dots, n$. Tällöin satunnaismuuttujan $X$ *odotusarvo* on

$$E(X) = p_1 x_1 + p_2 x_2 + \cdots + p_k x_k = \sum_{i=1}^{k} p_i x_i.$$

Odotusarvolle käytetään usein myös merkintää $\mu$, jota käytettiin myös tilastoaineiston keskiarvosta. Itse asiassa satunnaismuuttujan odotusarvon kaava (Määritelmä 3.24) on sama kuin frekvenssijakauman keskiarvon kaava (Määritelmä 2.19), nimittäin:

$$\mu = \sum_{i=1}^{k} p_i x_i = \sum_{i=1}^{k} \frac{f_i}{n} x_i = \frac{1}{n} \sum_{i=1}^{k} f_i x_i \,.$$

Jos satunnaiskoetta toistetaan useita kertoja, niin koetulosten keskiarvo on lähellä odotusarvoa. Satunnaismuuttujan odotusarvo siis kuvaa sen arvojen keskiarvoa, kun koetta toistetaan useita kertoja.

Kun halutaan puolestaan kuvata satunnaismuuttujan arvojen vaihtelua, tutkitaan todennäköisyysjakauman *varianssia* $D^2(X)$ ja *keskihajontaa* $D(X)$. Mitä lähempänä nollaa keskihajonta ja varianssi ovat, sitä todennäköisempää on, että satunnaismuuttujan arvo on lähellä odotusarvoa. Merkintöjen $D(X)$ ja $D^2(X)$ taustalla on englannin kielen sana *deviation*, joka tarkoittaa poikkeamaa, hajontaa. Satunnaismuuttujan $X$ varianssista käytetään yleisesti myös merkintöjä $\mathrm{Var}(X)$ ja $\sigma^2$, ja keskihajonnasta merkintää $\sigma$.

**Määritelmä 3.25** (*Varianssi ja keskihajonta*) Olkoot diskreetin satunnaismuuttujan $X$ saamat arvot $x_1, x_2, \ldots, x_k$ ja näiden pistetodennäköisyydet $p_i = P(X = x_i), i = 1, \ldots, n$. Tällöin satunnaismuuttujan $X$ *varianssi* on

$$D^2(X) = p_1(x_1 - \mu)^2 + \cdots + p_k(x_k - \mu)^2 = \sum_{i=1}^{k} p_i(x_i - \mu)^2 = E[(X - \mu)^2].$$

Josta saadaan ratkaistua satunnaismuuttujan $X$ *keskihajonta*:

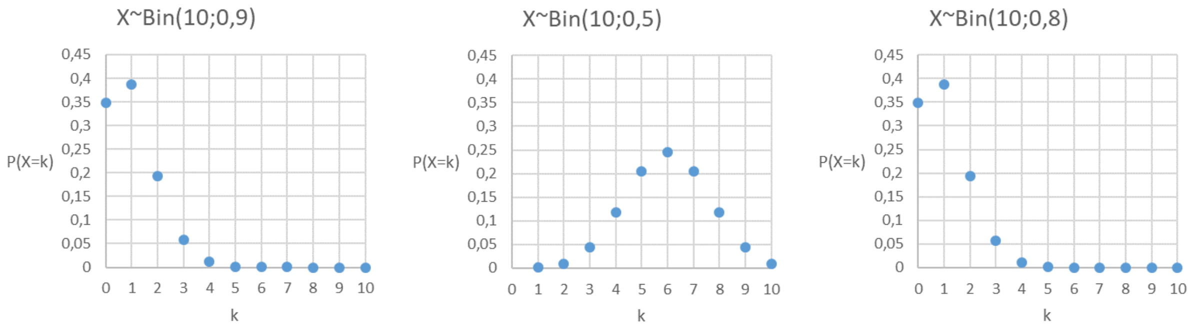
$$D(X) = \sqrt{\mathrm{Var}(X)}.$$

Määritelmän kaavaa ei itse asiassa todennäköisyyslaskennassa juuri koskaan käytetä, sillä varianssi saadaan mukavammin laskettua kaavoista $D^2(X) = E(X^2) - \mu^2$ ja $D^2(X) = E[X(X - 1)] + \mu - \mu^2$. Näiden kaavojen ymmärrys vaatisi kuitenkin satunnaismuuttujan X funktion g(X) odotusarvon $E[g(X)]$ sekä sen muunnosten $E[g(X)]$ ja $E[g(X)]$ johtamisen, minkä vuoksi lukion opintojaksolla määritelmän 3.25 kaavan tuntemus riittää.

### 3.4.4 Joitakin diskreettejä jakaumia

Luvussa 3.3.4 kerrottiin toistokokeesta ja siihen liittyvästä binomitodennäköisyydestä (Lause 3.15). Yksi tärkeimmistä diskreeteistä jakaumista on juuri toistokokeeseen liittyvä binomijakauma.

**Määritelmä 3.26** Merkitään yksittäisen kokeen onnistumistodennäköisyyttä kirjaimella $p$, epäonnistumisen todennäköisyyttä kirjaimella $q = 1 - p$ ja kokeen toistojen määrää kirjaimella $n$. Jos satunnaismuuttujan $X$ pistetodennäköisyydet ovat $P(X = k) = \binom{n}{k} p^k q^{n-k}$, jossa $k = 0, 1, 2, \ldots, n$, niin sanotaan satunnaismuuttujan $X$ noudattavan *binomijakaumaa* parametreilla $n$ ja $p$: merkitään $X \sim \mathrm{Bin}(n, p)$.



*Kuva 3.3* Eräiden binomijakaumien pistetodennäköisyydet koordinaatistossa.

**Lause 3.16** (*Binomijakauman odotusarvo*) Oletetaan, että $X \sim \mathrm{Bin}(n, p)$. Tällöin satunnaismuuttujan $X$ odotusarvo on

$$E(X) = n\,p.$$

*Todistus.*

Lasketaan odotusarvo määritelmän 3.24 mukaisesti:

$$E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}$$

$$= \sum_{k=1}^{n} k\,\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$= n\,p\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1} q^{n-k}$$

$$= n\,p\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1} q^{n-k}$$

$$= n\,p\sum_{k=0}^{n-1} \binom{n-1}{k}p^{k} q^{(n-1)-k}$$

$$\overset{Binomikaava}{\triangleq} \quad n\,p(p+q)^{n-1}. \blacksquare$$

**Lause 3.17** (*Binomijakauman varianssi ja keskihajonta*) Oletetaan, että $X \sim \mathrm{Bin}(n, p)$. Tällöin satunnaismuuttujan $X$ varianssi on

$$D^2(X) = n\,p\,q,$$

josta saadaan satunnaismuuttujan X keskihajonta:

$$D(X) = \sqrt{n\,p\,q}.$$

*Todistus.*

Satunnaismuuttujan $X$ varianssi voidaan laskea kaavalla $D^2(X) = E[X(X-1)] + \mu - \mu^2$ eli $D^2(X) = \sum_{k \in K} k(k-1)p(k) + \mu - \mu^2$, jota ei tässä yhteydessä todisteta.

Lasketaan odotusarvo:

$$E[X(X-1)] = \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k q^{n-k}$$

$$= \sum_{k=2}^{n} k(k-1)\,\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$= n(n-1)p^2\sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}p^{k-2} q^{n-k}$$

$$= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{n-k}$$

$$= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^{k} q^{(n-2)-k}$$

$$= n(n-1)p^2 (p+q)^{n-2}$$

$$= n(n-1)p^2.$$

Näin ollen

$$D^2(X) = E[X(X-1)] + \mu - \mu^2$$

$$\overset{Lause\ 3.15}{\triangleq} \quad n(n-1)p^2 + n\,p - (n\,p)^2$$

$$= n\,p - n\,p^2$$

$$= n\,p(1-p)$$

$$= n\,p\,q.$$

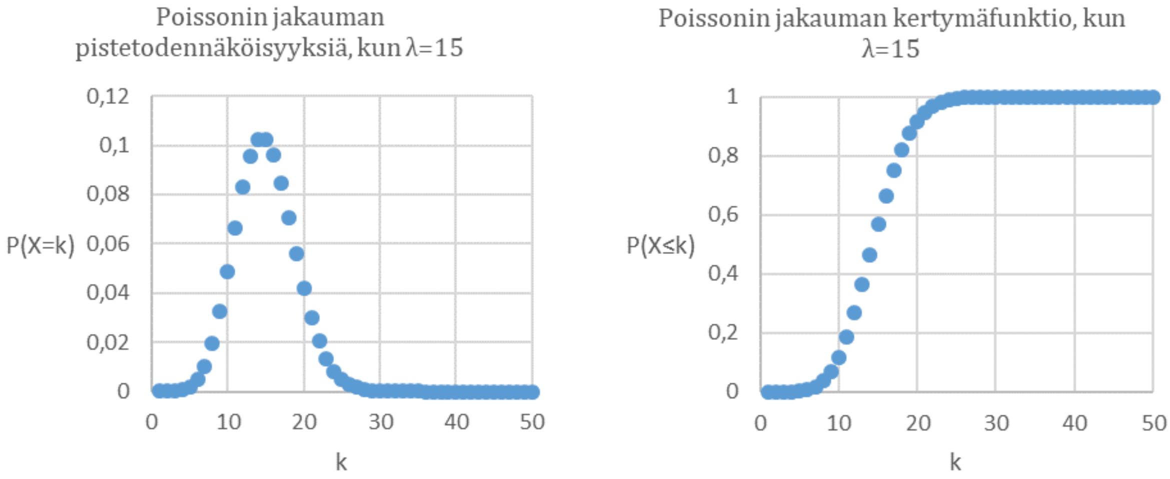
Tästä saadaan satunnaismuuttujan $X$ keskihajonnaksi $D(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{n\,p\,q}$. ■

Diskreetti satunnaismuuttuja $X$ voi saada myös äärettömän monta arvoa. Näin on esimerkiksi seuraavaksi esiteltävän Poissonin jakauman tapauksessa. Sanotaan että havaintojen lukumäärä tietyssä ajassa on Poisson-jakautunut parametrilla $\lambda$, jos jokin ilmiö havaitaan keskimäärin $\lambda$ kertaa kyseisessä ajassa ja havainnot ovat toisistaan riippumattomia. Poisson-jakauma (kuva 3.3) on binomijakauman ohella yksi sovelletuimmista diskreetin satunnaismuuttujan todennäköisyysjakaumista, mutta siitä huolimatta sitä ei kaikissa lukion todennäköisyyslaskennan oppikirjasarjoissa esitellä.

**Määritelmä 3.27** Jos satunnaismuuttujan $X$ pistetodennäköisyydet ovat $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, jossa $k = 0, 1, 2, \ldots$, niin satunnaismuuttuja $X$ noudattaa *Poisson-jakaumaa* parametrilla $\lambda$; merkitään $X \sim \mathrm{Po}(\lambda)$.

**Huomautus 3.4** Satunnaismuuttujan X todennäköisyysfunktio $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ on todella todennäköisyysfunktio, sillä

$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$



*Kuva 3.4* Poisson-jakauman pistetodennäköisyyksiä ja jakaumafunktio.

**Lause 3.18** (Poisson-jakauman odotusarvo, varianssi ja keskihajonta) Oletetaan, että $X \sim \mathrm{Po}(\lambda)$. Tällöin

$$E(X) = \lambda, \quad D^2(X) = \lambda \quad \text{ja} \quad D(X) = \sqrt{\lambda}.$$

*Todistus.*

Lasketaan odotusarvo $E(X)$:

$$E(X) = \textstyle\sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Jotta saadaan ratkaistua satunnaismuuttujan varianssi, lasketaan ensin odotusarvo $E[X(X-1)]$:

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}$$

$$= \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2.$$

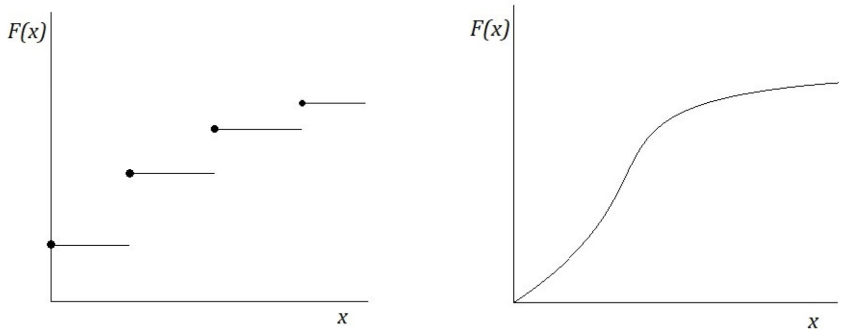
Lasketaan sitten satunnaismuuttujan $X$ varianssi kaavalla $D^2(X) = E[X(X-1)] + \mu - \mu^2$:

$$D^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Tästä saadaan satunnaismuuttujan $X$ keskihajonnaksi $D(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\lambda}$. ■

### 3.4.5 Jatkuvan satunnaismuuttujan jakauma- ja tiheysfunktio sekä todennäköisyyksien laskeminen

Jatkuvan satunnaismuuttuja voi saada ylinumeroituvan määrän arvoja ja on nimensä mukaisesti jatkuvasti jakautunut eli sen jakaumafunktio $F(x) = P(X \leq x), x \in \Re$, on jatkuva (vrt. diskreetin satunnaismuuttujan jakaumafunktio on porrasfunktio-tyyppinen, siis epäjatkuva; ks. kuva 3.5).



*Kuva 3.5* Vasemmalla diskreetin satunnaismuuttujan jakaumafunktio $F(x)$ ja oikealla jatkuvan satunnaismuuttujan jakaumafunktio $F(x)$, $F(x) = P(X \leq x), x \in \Re$.

Jatkuva satunnaismuuttuja määritellään kuitenkin ns. tiheysfunktion avulla:

**Määritelmä 3.28** Satunnaismuuttuja $X$ on *jatkuva*, jos on olemassa sellainen ei-negatiivinen funktio *f(x)*, että kaikille reaalilukujoukoille $A$ on $P(X \in A) = \int_A f(x)dx$ . Funktio $f$ on satunnaismuuttujan *X tiheysfunktio*.

Jatkuvan satunnaismuuttujan arvojen jakautumista ei voida kuvata pistetodennäköisyyksin *p(k)*, kuten voitiin tehdä diskreetin satunaismuuttujan tapauksessa, sillä jatkuvasti jakautuneen satunnaismuuttujan yksittäisen arvon todennäköisyys on nolla; nimittäin $P(X = x) = 0$, sillä $P(x \leq X \leq x) = \int_x^x f(t)dt = 0$. Tämän vuoksi jatkuvan satunnaismuuttujan arvojen jakautumisen kuvaamisessa käytetään tiheysfunktiota.

Jokainen tiheysfunktio on ei-negatiivinen ($f(x) \geq 0$, $x \in \Re$) ja tiheysfunktion kuvaajan ja x-akselin väliin jäävän alueen pinta-ala on 1 ($\int_{-\infty}^{\infty} f(x)dx = 1$). Itse asiassa voidaan myös osoittaa, että jos jokin funktio $f$ toteuttaa kyseiset ehdot, niin on olemassa satunnaismuuttuja, jonka tiheysfunktio on $f$. Näin ollen nämä ehdot ovat välttämättömät ja riittävät ehdot sille, että funktio $f$ on tiheysfunktio. On hyvä huomata, että jatkuvan satunnaismuuttujan tiheysfunktio voi hyvin saada ykköstä suurempia arvoja ja että $f(x)$ ei ole sama kuin $P(X = x)$. Vain tiheysfunktion määrätyillä integraaleilla on todennäköisyystulkinnat.

**Lause 3.19** Jatkuvan satunnaismuuttujan jakauma- ja tiheysfunktioilla on seuraavat ominaisuudet:

a) $\int_{-\infty}^{\infty} f(x)dx = 1$.

b) $F(x) = \int_{-\infty}^{x} f(t)dt$.

c) $f(x) = F'(x)$ pisteissä *x*, joissa *f(x)* on jatkuva.

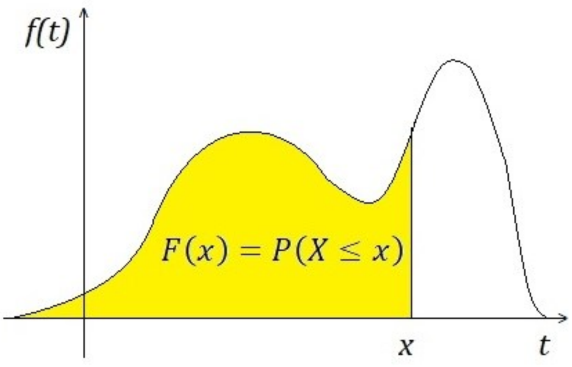
*Todistus.*

a) $P[X \in (-\infty, \infty)] = 1 = \int_{-\infty}^{\infty} f(x)dx$.

b) $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$.

c) Koska $F(x) = \int_{-\infty}^{x} f(t)dt$, niin differentiaali- ja integraalilaskennan peruslauseen mukaan siitä seuraa, että $F'(x) = f(x)$. ■

Jos siis tiheysfunktio on annettu, saadaan jakaumafunktio lauseen b-kohdan avulla. Tätä on havainnollistaa kuva 3.6. Jos taas jakaumafunktio on annettu, niin tiheysfunktio saadaan lauseen c-kohdan avulla; pisteissä, joissa $F'(x)$ ei ole olemassa, voidaan määritellä $f(x) = 0$. Koska $f(x) = F'(x)$, niin $f(x)$ ilmoittaa, kuinka voimakkaasti $F(x)$ kasvaa pisteessä *x*. Mitä voimakkaampi kasvu, sitä nopeammin todennäköisyyttä kertyy funktioon $F(x) = P(X \leq x)$, kun *x* kasvaa.



*Kuva 3.6* Todennäköisyyden laskeminen tiheysfunktion avulla; $F(x) = \int_{-\infty}^{x} f(t)dt$.

Lauseen 3.19 b-kohdasta nähdään, että jatkuvan satunnaismuuttujan jakaumafunktiolla on seuraavat ominaisuudet:

- $0 \leq F(x) \leq 1, x \in \mathfrak{R}$,
- $F(x)$ on jatkuva,

- $F(x)$ on kasvava,
- $\lim_{x \to \infty} F(x) = 1$ ja
- $\lim_{x \to -\infty} F(x) = 0$.

Sekä diskreetin että jatkuvan satunnaismuuttujan jakaumafunktion ominaisuudet ovat siis jatkuvuutta lukuun ottamatta samat.

Jatkuvan satunnaismuuttujan todennäköisyyksiä voidaan laskea sekä jakauma- että tiheysfunktion avulla seuraavasti:

| Todennäköisyys | *F(x)*:n avulla | *f(x)*:n avulla |
|---|---|---|
| $P(X \leq a)$ | $F(a)$ | $\int_{-\infty}^{a} f(t)d\,t.$ |
| $P(X > a)$ | $1 - F(a)$ | $\int_{a}^{\infty} f(t)d\,t.$ |
| $P(a \leq X \leq b)$ | $F(b) - F(a)$ | $\int_{a}^{b} f(t)d\,t.$ |

Jos jakaumafunktio on käytettävissä, niin sitä kannattaa käyttää, sillä tällöin vältytään integroinnilta. Jos todennäköisyys lasketaan tiheysfunktion avulla, niin silloin tulee itse asiassa lasketuksi tiheysfunktion $f$ ja tarkasteltavan $x$-akselin osan välisen alueen pinta-ala (ks. kuva 3.6). Kaavasta $P(X = a) = 0$ seuraa, että laskettaessa välien todennäköisyyksiä jatkuvalle satunnaismuuttujalle ei ole merkitystä, ovatko välin päätepisteet mukana vai eivät. Esimerkiksi $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$.

### 3.4.6 Jatkuvan jakauman tunnusluvut

Diskreetin satunnaismuuttujan tapaan myös jatkuvalle satunnaismuuttujalle määritellään jakaumasta saatavaa tietoa tiivistäviä tunnuslukuja, joita ovat muun muassa seuraavaksi esiteltävät odotusarvo, keskihajonta ja varianssi. On hyvä huomata, että jatkuvan jakauman tunnusluvut ovat lukion todennäköisyyslaskennan opintojaksolla syventävää tietoa, joita kaikki oppikirjat eivät käsittele ollenkaan.

**Määritelmä 3.29** (*Odotusarvo*) Jatkuvan satunnaismuuttujan *X* *odotusarvo* on

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)dx,$$

jos $\int_{-\infty}^{\infty} |x|\, f(x)dx < \infty$.

Diskreeteille satunnaismuuttujille esitetyt varianssin ja keskihajonnan määritelmät pätevät myös jatkuville satunnaismuuttujille:

**Määritelmä 3.30** (*Varianssi ja keskihajonta*) Jatkuvan satunnaismuuttujan $X$ *varianssi* on

$$D^2(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty}(x-\mu)^2\, f(x)dx,$$

josta saadaan ratkaistua satunnaismuuttujan $X$ *keskihajonta*:

$$D(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\int_{-\infty}^{\infty}(x-\mu)^2\, f(x)dx}.$$

Odotusarvon määritys edellyttää siis tiheysfunktion $f$ tuntemista. Varianssin ja keskihajonnan määritys taas edellyttää, että sekä tiheysfunktio $f$ että odotusarvo $E(X)$ ovat tiedossa.

Odotusarvo $E(X) = \mu$ on kaikkein käytetyin satunnaismuuttujan sijaintia ilmaiseva tunnusluku, mutta sijaintia voidaan mitata muillakin tunnusluvuilla. Muistetaan Tilastot-luvusta, että tilastomuuttujille moodi on havaintoaineiston yleisin arvo eli arvo, jonka frekvenssi on suurin. Vastaavasti satunnaismuuttujille moodi on satunnaismuuttujan todennäköisin arvo. Toki jatkuvan satunnaismuuttujan yksittäisen arvon todennäköisyys on nolla, mutta tällöinkin moodin ympärillä on eniten ns. todennäköisyysmassaa.

**Määritelmä 3.31** (*Moodi*) Diskreetin satunnaismuuttujan *moodi* on se muuttujan $k$ arvo (tai ne arvot), jossa (tai joissa) todennäköisyysfunktio $p(k)$ saa maksimiarvon. Jatkuvan satunnaismuuttujan *moodi* on se muuttujan $x$ arvo (tai ne arvot), jossa (tai joissa) tiheysfunktio $f$ saa maksimiarvon.

Tilastomuuttujilla arvojen keskimääräistä suuruutta kuvaa *mediaani*. Vastaavasti satunnaismuuttujilla mediaani jakaa populaation kahteen yhtä suureen osaan siten, että se ala- ja yläpuolella on yhtä paljon todennäköisyysmassaa.

**Määritelmä 3.32** (*Mediaani*) Diskreetin satunnaismuuttujan *mediaani* on se muuttujan $k$ arvo (tai ne arvot), jossa (tai joissa) jakaumafunktio $F$ toteuttaa ehdon $F(k-1) \leq \frac{1}{2} \leq F(k)$. Jatkuvan satunnaismuuttujan *mediaani* on se muuttujan $x$ arvo (tai ne arvot), jossa (tai joissa) jakaumafunktio $F$ toteuttaa ehdon $F(x) = \frac{1}{2}$.

### 3.4.7 Normaalijakauma ja tasainen jakauma

Normaalijakauma on käytetyin ja sovelletuin jatkuva jakauma ja keskeinen osa lukion nykymuotoisen todennäköisyyslaskennan opintojakson ydinainesta. Esimerkiksi monet ilmiöt ja niihin liittyvät satunnaismuuttujat luonnossa riippuvat useista satunnaistekijöistä ja noudattavat normaalijakaumaa.

Normaalijakauma tunnetaan myös nimillä Gaussin jakauma ja Gaussin kellokäyrä, jotka ovat saaneet nimensä 1800-luvun alussa saksalaisen matemaatikko Karl Fridrich Gaussin mukaan ja joista jälkimmäinen kuvaa normaalisti jakautuneen satunnaismuuttujan tiheysfunktion kuvaajan symmetrisen kellomaista muotoa (Kuva 3.7). Normaalijakauma-nimen käyttö vakiintui, kun huomattiin monien havaintoaineistojen noudattavan Gaussin jakaumaa ja alettiin ajatella tämän olevan ikään kuin "normaali" havaintojen jakauma. Normaalijakaumaa käytetään kuvamaan ilmiöitä, joissa nk. todennäköisyysmassa keskittyy ilmiön keskialueen arvoihin ja määrittelyalueen ääripäiden arvot ovat harvinaisia.

**Lause 3.20** (*Standardoitu normaalijakauma*) Jos satunnaismuuttujan X tiheysfunktio on

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty < x < \infty,$$

niin sanotaan, että satunnaismuuttujalla X on standardoitu normaalijakauma; merkitään $X \sim N(0,1)$. Tällöin odotusarvo $E(X) = 0$ ja varianssi $D^2(X) = 1$. Jakaumafunktiota merkitään

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt, \qquad -\infty < x < \infty.$$

On voimassa $\Phi(-x) = 1 - \Phi(x)$.

*Todistus.*

Funktio $\varphi(x)$ on todella tiheysfunktio sillä $\int_{-\infty}^{\infty} \varphi(x)\, dx = 1$.

Odotusarvoksi saadaan:

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}x^2} dx = -\frac{1}{\sqrt{2\pi}} /_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} = 0.$$

Varianssiksi saadaan:

$$D^2(X) = E(X^2) - \mu^2 = E(X^2)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx, \text{jossa } u = x, u' = 1, v' = xe^{-\frac{1}{2}x^2}, v = -e^{-\frac{1}{2}x^2}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}x^2} dx}_{=0,\text{koska } E(X)=0.} + \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx}_{=1,\text{koska } \int_{-\infty}^{\infty} \varphi(x)dx=1.} = 1.$$

Todennäköisyys $P(X \leq -x) = \Phi(-x)$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{1}{2}t^2} dt \text{, merkitään } s = -t$$

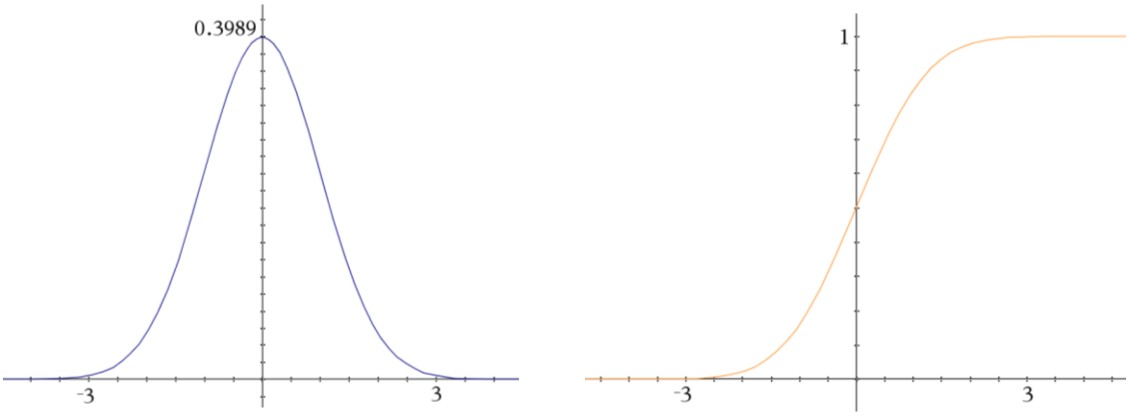
$$= \frac{-1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-\frac{1}{2}s^2} ds$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{1}{2}s^2} ds$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}s^2} ds - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}s^2} ds$$

$$= 1 - \Phi(x). \blacksquare$$

Standardoidun normaalijakauman eli standardinormaalijakauman jakaumafunktiota ei voida ilmaista alkeisfunktioiden avulla, minkä vuoksi jakaumafunktion (Kuva 3.7) arvoja on taulukoitu. Taulukoista saadaan yleensä todennäköisyyksiä $\Phi(x) = P(X \leq x)$, kun $x$ on positiivinen. Negatiivisilla muuttujan $x$ arvoilla voidaan käyttää lauseen 3.20 kaavaa $\Phi(-x) = 1 - \Phi(x)$.



*Kuva 3.7* Standardoidun normaalijakauman tiheysfunktio *f(x)* (vas.) ja jakaumafunktio *F(x)*.

Seuraavaksi esitellään standardoidun normaalijakauman lineaarinen muunnos, yleinen normaalijakauma.

**Lause 3.21** (*Yleinen normaalijakauma*) Jos $Y \sim N(0,1)$, niin sanotaan, että satunnaismuuttujalla $X = \mu + \sigma Y$, missä $\sigma > 0$, on normaalijakauma, $X \sim N(\mu, \sigma^2)$. Tällöin

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty.$$

$$E(X) = \mu, \qquad D^2(X) = \sigma^2.$$

Jos $X \sim N(\mu, \sigma^2)$ ja $Z \sim a + bX$, niin myös satunnaismuuttujalla Z on normaalijakauma; $Z \sim N(a + b\mu, b^2\sigma^2)$.

*Todistus.*

Tarkastellaan ensin jakaumafunktiota $F(x)$:

$$F(x) = P(X \leq x)$$

$$= P(\mu + \sigma Y \leq x)$$

$$= P\left(Y \leq \frac{x - \mu}{\sigma}\right)$$

$$= \Phi\left(Y \leq \frac{x - \mu}{\sigma}\right).$$

Tiheysfunktiolle saadaan:

$$f(x) = F'(x)$$

$$= \frac{1}{\sigma}\Phi\left(\frac{x - \mu}{\sigma}\right)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Odotusarvoksi $E(X)$saadaan:

$$E(X) = E(\mu + \sigma Y) = \mu + \sigma E(Y) = \mu.$$

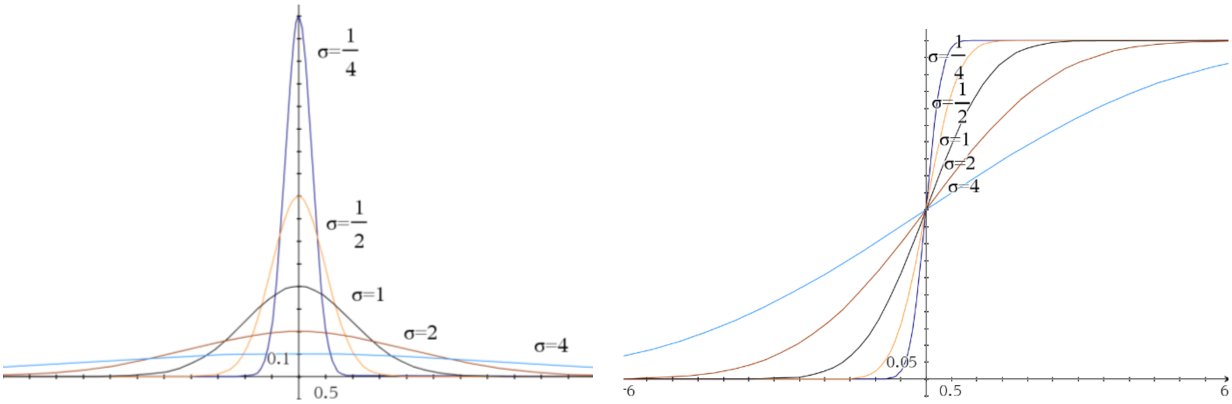
Varianssiksi $D^2(X)$ saadaan:

$$D^2(X) = D^2(\mu + \sigma Y) = \sigma^2 D^2(Y) = \sigma^2. \blacksquare$$

Lukion MAOL-taulukoissa normaalijakauman $N(\mu, \sigma^2)$ toinen parametri on hajonta $\sigma$ lauseessa 3.21 käytetyn varianssin $\sigma^2$ sijaan.

Yleisen normaalijakauman tiheysfunktion ja jakaumafunktion tarkan muodon määräävät satunnaismuuttujan $X$ odotusarvo ja varianssi (Kuva 3.8).

Kun satunnaismuuttujasta $X$ vähennetään sen odotusarvo $\mu$, tiheysfunktion kuvaaja siirtyy vaakasuunnassa $\mu$:n verran. Saadun satunnaismuuttujan $X - \mu$ odotusarvoksi tulee 0. Kun satunnaismuuttuja $X - \mu$ jaetaan luvulla $\sigma$, saadaan satunnaismuuttuja $\frac{X-\mu}{\sigma}$, jonka odotusarvo on myös 0 ja varianssi 1 eli $\frac{X-\mu}{\sigma} \sim N(0,1)$. Tällaisen muunnoksen tekemistä kutsutaan *normittamiseksi.*



*Kuva 3.8* Normaalijakauman $N(0,\sigma^2)$ tiheysfunktio (vas.) ja jakaumafunktio, kun $\sigma = \frac{1}{4}, \frac{1}{2}$, 1,2 ja 4.

Useille jakaumille voidaan johtaa normaalijakauma-approksimaatiot (keskeisen raja-arvolauseen avulla). Koska lukion opintojaksolla binomijakauman arvioiminen normaalijakaumalla on syventävää sisältöä, tyydytään tässä yhteydessä vain toteamaan hieman epätäsmällisesti:

- Jos $X{\sim}Bin(n,p)$, niin $X{\sim}N(np,npq)$, jos $n$ on suuri,
- Jos $X{\sim}Po(\lambda)$, niin $X{\sim}N(\lambda,\lambda)$, jos λ on suuri.

Normaalijakauma-approksimaatio on yleensä riittävän tarkka binomijakaumalle, kun $npq > 10$, ja Poisson-jakaumalle, kun $\lambda > 15$.

Normaalijakaumassa kaikki reaalilukuarvot ovat mahdollisia. Kuitenkin monesti reaalimaailman ilmiöitä tarkasteltaessa, kuten henkilön painoa suuressa ihmispopulaatiossa tutkittaessa, osa reaalilukuarvoista on käytännössä mahdottomia. Näin ollen normaalijakauman käyttäminen aiheuttaa malliin näiden arvojen kohdalle virheen. Virhe on kuitenkin merkityksettömän pieni tulosten luotettavuuden kannalta todennäköisyysmassan keskittyessä jakauman realististen arvojen alueelle.

Jos todennäköisyys on jakautunut tasaisesti jollekin välille $[a,b]$ eli tiheysfunktio on vakio kyseisellä välillä, niin kyseessä on tasaisesti jakautunut satunnaismuuttuja ja puhutaan *tasaisesta jakaumasta* (Kuva 3.9).

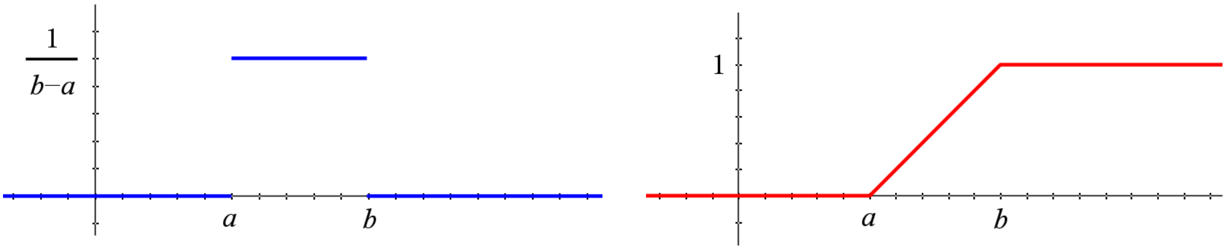
**Lause 3.22** (*Tasainen jakauma*) Jos satunnaismuuttujan $X$ tiheysfunktio on

$$f(x) = \begin{cases} 0, \text{kun } x \leq a, \\ \dfrac{1}{b-a}, \text{kun } a < x < b, \\ 0, \text{kun } x \geq b, \end{cases}$$

missä $a < b$, niin sanotaan, että satunnaismuuttujalla $X$ on *tasainen jakauma* välillä $[a, b]$; merkitään $X \sim U(a, b)$. Tällöin

$$F(x) = \begin{cases} 0, \text{kun } x \leq a, \\ \dfrac{x-a}{b-a}, \text{kun } a < x < b\,, \\ 0, \text{kun } x \geq b, \end{cases} \qquad E(X) = \frac{a+b}{2}, \qquad D^2(X) = \frac{(b-a)^2}{12}.$$

Todistus sivuutetaan.



*Kuva 3.9* Tasaisen jakauman tiheysfunktio *f(x)* (vas.) ja jakaumafunktio *F(x)*.

# 4 Terve! Terveysaiheinen tehtäväpaketti opiskelijalle -materiaali

*Terve! Terveysaiheinen tehtäväpaketti opiskelijalle* – oppimateriaali koostuu lukion matematiikan pitkän oppimäärän todennäköisyyslaskennan opintojaksolle tarkoitetuista tehtävistä, tehtävien aiheisiin liittyvästä asiatiedosta, sanallisista ratkaisuvihjeistä (vain opiskelijan materiaalissa) sekä tehtävien eräistä malliratkaisuista (vain opettajan materiaalissa). Kaikki tehtävät ovat terveysaiheisia, osa hyvin spesifiin jakson aiheeseen liittyviä perustehtäviä ja osa laajojen kokonaisuuksien hallintaa vaativia, soveltavia tehtäviä. Tehtävien järjestys on materiaalissa satunnainen eikä tehtävänanto kaikissa tehtävissä anna opiskelijalle ilmeistä vihjettä siitä, mitä laskennallisia keinoja hänen tulisi tehtävän ratkaisussa käyttää. Materiaalin kaksi viimeistä tehtävää ovat todennäköisyyslaskennan opintojakson tietoja syventäviä tehtäviä, joiden ratkaisemiseksi opiskelijan on tutustuttava kokonaistodennäköisyyslauseeseen ja Bayesin lauseeseen, jotka eivät kuulu lukion oppimäärään, mutta ovat esiteltyjä opiskelijan materiaalissa.

Sekä opiskelijamateriaali (Liite1) että malliratkaisut sisältävä opettajan materiaali (Liite 2) on käytettävyyden ja selkeyden vuoksi sijoitettu tutkielman liitteisiin. Opettaja voi käyttää materiaalia todennäköisyyslaskennan opintojakson opetuksessa monella tavalla: sen voi jakaa opiskelijoille itsenäiseksi kurssityöksi opintojakson alussa, sitä voi käyttää opintojakson kertauksessa tai yksittäisiä tehtäviä voi esittää elävöittävinä esimerkkeinä oppitunneilla. Koska tehtäviin on saatavilla malliratkaisut, niitä on helppo käyttää koetehtävinä tai opintojakson aikaisen formatiivisen arvioinnin tukena - esimerkiksi vertaisarvioitavina kotitehtävinä.

# 5 Pohdinta

*Terve! Terveysaiheinen tehtäväpaketti opiskelijalle* –materiaalin tehtävissä havainnollistetaan useita terveyteen liittyviä ilmiöitä todennäköisyyslaskennan keinoin. Terveysilmiöiden tutkiminen matemaattisesti luo opiskelijalle mahdollisuuden oivaltaa uutta tehtävän terveysaiheesta sekä matemaattisen ratkaisun rakentamisesta. Opiskelija saattaa oppia esimerkiksi *Rokotteen rooli taudin leviämisen ehkäisyssä* – tehtävästä (Tehtävä 11), että tehokaskaan rokote ei riitä suojaamaan rokottamattomia, jos rokotettuja ei ole tarpeeksi. Jotta tauti ei pääse leviämään populaatiossa eksponentiaalisesti, yksi sairastunut ei saa tartuttaa kahta tai useampaa ei-immuunia henkilöä. Näin kuitenkin pääsee tapahtumaan, jos rokotekattavuus ei ole riittävän suuri. Kansallisilla rokotusohjelmilla on siis terveydellisten, inhimillisten ja yhteiskunnallisten syiden lisäksi myös matemaattiset perusteet. Tautien ehkäisyn lisäksi Terve! -materiaalissa tutkitaan muun muassa taudin havaitsemista tautitestillä (Tehtävä 12). Intuitiivisesti luotettavalta vaikuttava HIV-testi, joka kertoo 95% varmuudella tartunnan saaneelle oikean tuloksen, on osoittautunut ehdollisia todennäköisyyksiä laskemalla yllättäen huonoksi tavaksi seuloa tartuntoja. Kun terveitä testin tekijöitä on paljon, syntyy suuri määrä virheellisiä positiivisia tuloksia. Monesti todennäköisyyslaskennassa onkin niin, että intuitiivinen kuvitelma tilanteesta ei vastaa todellisuutta. Ongelman ratkominen ohjaa sekä suhtautumaan kriittisen uteliaasti "liian hyvin" markkinoitujen testien luotettavuuteen että oivaltamaan todennäköisyyslaskennan jekuttavan perusluonteen.

Useimpia Terve! –materiaalin aiheita, kuten terveystottumuksia, kestävyysharjoittelua ja kansantauteja, käsitellään lukiossa terveystiedon oppiaineessa. Siinä terveyteen liittyviä ilmiöitä tarkastellaan terveysosaamisen osa-alueiden eli terveyteen liittyvien tietojen, taitojen, itsetuntemuksen, kriittisen ajattelun ja eettisen vastuullisuuden kautta [3, s. 328]. Terve! -materiaalin tehtävien aiheiden tarkastelu matemaattisesti voi auttaa opiskelijaa kehittämään omaa terveysosaamistaan. Toisaalta ajankohtaisen, reaalimaailman ongelman ratkominen elävöittää matematiikan opiskelua ja mielenkiintoinen tehtävä voi herättää opiskelijan kiinnostuksen matemaattiseen ongelmanratkaisuun. Opiskelijat pitävät kiinnostavina sekä ympäröivään maailmaan liittyviä että siihen liittymättömiä matematiikan tehtäviä [10]. Pelkät tehtävät eivät yksinään riitä innostamaan opiskelijaa matematiikan opiskeluun, vaan ne elävöityvät ja saavat merkityksen oppimisympäristön vuorovaikutussuhteissa [11]. Tehtävien innostavuuteen vaikuttavat mm. opettajan pedagogiset ratkaisut: tutkitaanko tehtäviä oppitunneilla vai kotona, yksin vai ryhmissä, osissa vai kokonaisuuksina, visualisoidaanko niitä tai käsitelläänkö niiden erilaisia ratkaisuvaihtoehtoja. Schukajlowin ja Krugin tutkimusten [12] mukaan opiskelijoiden rohkaisuun erilaisten ratkaisutapojen esiin tuomiseksi tulisi kiinnittää nykyistä enemmän huomiota erityisesti soveltavia tehtäviä tarkastettaessa. Tämä voi tarkoittaa opettajalle opiskelijoiden kanssa yhteistä keskustelua tai heidän ohjaamistaan keskustelemaan pienryhmissä tehtävien erilaisista ratkaisutavoista. Terve!-oppimateriaalin malliratkaisuissa esitellään useimmiten vain yksi tapa ratkaista kukin tehtävä. On toivottavaa, että opiskelijoille kerrotaan valmiin ratkaisun olevan yksi monista mahdollisista vaihtoehdoista lähestyä tehtävää ja että muiden ratkaisutapojen tutkimiselle varataan aikaa. Säännöllinen erilaisten ratkaisuvaihtoehtojen tutkiminen vahvistaa opiskelijan kokemusta omasta matemaattisesta kyvykkyydestään, mikä mahdollistaa kiinnostuksen kehittymisen matematiikkaa kohtaan.

Terve!-materiaalin malliratkaisut on rakennettu hyvän matemaattisen vastauksen piirteiden ja käytänteiden mukaisesti. Todennäköisyyden laskukaavojen käyttö perustellaan tarkasti ja laskujen välivaiheita on näkyvissä runsaasti. Vastauksen etenemistä kuvataan materiaalissa yksityiskohtaisesti, jotta ratkaisua olisi helppo lukea myös silloin, kun tehtävän ratkaisemisesta on kulunut paljon aikaa tai kun opiskelija on itse ratkaissut tehtävän eri tavalla ja haluaa ymmärtää vaihtoehtoisen laskutavan. Tehtävänannon tilannetta on useissa malliratkaisuissa pyritty visualisoimaan esimerkiksi Venn-diagrammin, taulukon, piirroksen tai sanallisen kuvailun avulla. Selvitettävän ongelman visualisointi on tutkimusten [13,14] mukaan avainasemassa matemaattisessa päättelyssä, ongelmanratkaisussa ja jopa todistamisessa.

Kansallisen koulutuksen arviointikeskuksen raportin [15] mukaan hyvin ja heikosti matematiikassa menestyneiden peruskoulun 9. luokkalaisten oppilaiden matematiikan osaamisessa on useiden vuosien taitotasollinen ero. Laskeva trendi matematiikan osaamisen tasossa yhdeksäsluokkalaisilla on jatkunut vuodesta 2000, mutta tulos on silti aiempiin vuosiin verrattuna poikkeava: osaaminen on polarisoitunut. Parhaat oppilaat pystyvät ratkaisemaan ylioppilaskokeen lyhyen oppimäärän tehtäviä, heikoimmilla oppilailla on vaikeuksia jopa alakoululuokkien tehtävissä. Syyt osaamisen eriytymisen taustalla saattavat liittyä koronapandemian aikaiseen pahoinvointiin perheissä, etäopetukseen sekä aiempaa suurempiin eroihin oppilaiden vastausteknisissä taidoissa [15]. Matematiikan osaamisen eriytyminen ja suuret koulujen väliset erot asettavat opiskelijat eriarvoiseen asemaan toisen asteen opintoihin hakeuduttaessa sekä opintojen aikana. Erilaiset lähtökohdat haastavat myös lukion matematiikan opettajia opetuksen suunnittelussa ja toteutuksessa.

Opiskelijoiden osaamisen hajonta oli 2021 julkaistun raportin [15] mukaan suurinta juuri tilastollisten tunnuslukujen ja todennäköisyyden sisältöalueella sekä tavoitteiden toteutumisessa todennäköisyys- ja prosenttilaskennassa. Eniten huolestuttavaa laskua kahdenkymmenen vuoden aikana on tapahtunut matematiikan ajattelun taidoissa ja menetelmissä. Minäpystyvyys, eli kokemus itsestä matematiikan osaajana sekä luottamus omiin kykyihin suoriutua erilaisissa tilanteissa ja tehtävissä, on heikentynyt erityisesti hyvin menestyneillä oppilailla [15]. Opettajan merkitys rohkaisijana ja kannustajana on ilmeinen opiskelijan minäpystyvyyden ja matemaattisen itseluottamuksen vahvistumiselle. Myönteinen käsitys omista matematiikan taidoista näkyy opiskelijan uskona pärjätä matematiikan tehtävissä myös tulevaisuudessa [16, s. 11]. Lukinin tutkimuksen [17, s. vi] mukaan opiskelijan menestymistä matematiikassa selittävätkin eniten vahva minäpystyvyys ja oppimisorientaatio. Lisäksi menestystä matematiikassa selittää opiskelijan ja opettajan kohtaaminen ilman ulkoiseen motivaatioon nojaavan suoritusorientaatioon liittyviä viestejä opettajalta opiskelijalle [17]. Opiskelijan sitoutuminen matematiikan opiskeluun pitkäjänteisesti tapahtuu matematiikan merkitykselliseksi kokemisen kautta, eikä suoritusorientaatio tue tätä kehityskulkua.

2020-luku on ollut tähän asti tartuntatautien uutta kulta-aikaa. Terveyteen ja sairauteen liittyvän asiatiedon määrä lisääntyy jatkuvasti - se on yhä helpommin ja yhä useamman saavutettavissa. Tästä huolimatta misinformaatio, huhut ja salaliittoteoriat tuntuvat leviävän oikeaa tietoa ja apinarokkoakin nopeammin. Ymmärrys terveysilmiöiden matematiikasta parantaa opiskelijan mahdollisuuksia navigoida turvallisesti ja lähdekriittisesti näillä haastavilla informaatiokentillä. Terve!-oppimateriaalin tehtävien ratkaisutapojen tutkiminen ja vastausten luotettavuuden, merkityksen ja seurausten pohdinta yhdessä vertaisten ja kannustavan opettajan kanssa vahvistavat opiskelijan matemaattista itseluottamusta ja oppimissuuntautunutta orientaatiota.

# 6 Lähteet

[1] Lehtinen, Matti. Todennäköisyyslaskenta: ajanvietteestä tiedettä. Matematiikan historia. Matematiikkalehti Solmu 3/2000-2001. https://matematiikkalehtisolmu.fi/2000/mathist/pdf/tnlask.pdf - *Luettu 30.5.2022*

[2] Lee, Peter M. Materials for the History of Statistics. University of York. Otteita Pascalin ja Fermat'n kirjeenvaihdosta. https://www.york.ac.uk/depts/maths/histstat/pascal.pdf - *Luettu 30.5.2022*

[3] Opetushallitus. Lukio-opetuksen opetussuunnitelman perusteet 2019. Määräykset ja ohjeet 2019:2a.16-20, 45, 58-65, 221-222 & 227.

[4] Beyranevand, Matthew L. 4 Ways to Increase Student Interest in Mathematics. TenMarks Education. 2016. https://tenmarks.typepad.com/tenmarks/2016/04/4-ways-to-increase-student-interest-in-mathematics.html

[5] Pesonen, Henri. Tilastollisen päättelyn peruskurssi. 2018. Turun yliopisto. 5-34.

[6] Mellin, Ilkka. Tilastolliset menetelmät. Osa 4: Lineaarinen regressioanalyysi. 2007. 3-11. https://math.tkk.fi/opetus/sovtoda/luennot/TILRI100.pdf

[7] Ruskeepää, Heikki. Todennäköisyyslaskenta I. 2011. Turun yliopisto. 3-22.

[8] Lahtonen, Jyrki. Analyysi I. 2010. Turun yliopisto. 10-11.

[9] Honkala, Iiro. Kombinatoriikka. 2009. Turun yliopisto. 3-4.

[10] Rellensmann, Johanna & Schukajlow, Stanislaw. Does students' interest in a mathematical problem depend on the problem's connection to reality? An analysis of students' interest and pre-service teachers' judgments of students' interest in problems with and without a connection to reality. ZDM Mathematics Education. 2017, 49. 367–378.

[11] Chapman, Olive. Mathematical-task knowledge for teaching. Journal of Mathematics Teacher Education. 2013, 16. 1–6.

[12] Schukajlow, Stanislaw & Krug, André. Do Multiple Solutions Matter? Prompting Multiple Solutions, Interest, Competence, and Autonomy. Journal for Research in Mathematics Education 2014, Vol. 45, No. 4. 497–533.

[13] Arcavi, Abraham. THE ROLE OF VISUAL REPRESENTATIONS IN THE LEARNING OF MATHEMATICS. Educational Studies in Mathematics. 2003, 52: 215–241. 235-236.

[14] Presmeg, Norma. Visualization and Learning in Mathematics Education. Teoksessa Lerman, S. (toim.) Encyclopedia of Mathematics Education. Springer, Cham. 2020. 900-904.

[15] Metsämuuronen, Jari & Nousiainen Saara. MATEMATIIKKAA COVID-19-PANDEMIAN VARJOSSA - Matematiikan osaaminen 9. luokan lopussa keväällä 2021. Kansallinen koulutuksen arviointikeskus. Julkaisut 27: 2021. 3-5, 61, 81 & 105-107.

[16] Samela-Aro, Katariina. Motivaatio ja oppiminen kulkevat käsi kädessä. PS-Kustannus. 2018. 9-22.

[17] Lukin, Tuija. Motivaatio matematiikan opiskelussa – seurantatutkimus motivaatiotekijöistä ja niiden välisistä yhteyksistä yläkoulun aikana. Itä-Suomen yliopisto. 2012. v-vi, 20-21 & 89.

# Liitteet

Liite 1 Materiaali opiskelijalle

Liite 2 Materiaali opettajalle

Lukion matematiikan pitkä oppimäärä

**Tilastot ja todennäköisyys**

# TERVE!

Terveysaiheinen tehtäväpaketti opiskelijalle

Annina Ruokonen

# Sisällys

# KÄYTTÄJÄLLE

Todennäköisyyslaskenta on suhteellisen nuori, noin 400-vuotias, matematiikan osa-alue, joka on perinteisesti keskittynyt erilaisten rahapelien voiton todennäköisyyden laskemiseen. Lukion todennäköisyyslaskennan opintojaksolla huomataankin todennäköisyyslaskennan olevan kätevä väline muun muassa useiden kortti- ja noppapelien mallinnuksessa, mutta laskennan keinot taipuvat oivallisesti myös useiden arjen, luonnon ja yhteiskunnan ilmiöiden ja tapahtumien mallintamiseen. Tämän materiaalin tehtävät liittyvät ajankohtaisiin arkiympäristön ja yhteiskunnan terveysilmiöihin, kuten treenaamiseen, ensiaputaitoihin, rokotteisiin, punkkien levittämiin tauteihin ja väestöryhmien välisiin terveyseroihin. Kunkin tehtävän yhteydessä kerrotaan hieman tehtävän aiheesta ja annetaan vinkkejä asiallisen lisätiedon löytämiseen. Tehtävissä on tarkoitus tutustua erilaisten terveysilmiöiden syihin ja seurauksiin, ymmärtää ja kehittää niihin liittyvää matemaattista ajattelua sekä oppia analysoimaan saatujen tulosten mielekkyyttä ja luotettavuutta. Tavoitteena on harjoitella pohtimaan saatujen tulosten merkitystä terveydelle yksilö-, yhteisö- ja yhteiskuntatasolla.

Tehtävissä on käytetty seuraavia symboleja, joiden merkitys on tässä selitetty:

| | |
|---|---|
| ? | Tökkääkö ratkaiseminen? Vihjeitä on saatavilla. |
| 🖩 | Laskinohjelmistoja saa käyttää tehtävän ratkaisemiseksi. |
| 🚫🖩 | Laskinohjelmistoja ei saa käyttää tehtävän ratkaisemiseksi. Funktiolaskin, kuten SpeedCrunch, KCalc ja Gnome, ovat käytössä. |

Vihjeet tehtävien ratkaisemiseksi löytyvät materiaalin takaosasta. Malliratkaisut saat opettajaltasi.

Innostavia hetkiä tehtävien ratkaisemiseen toivoo!

Annina Ruokonen

Tekijä

# Tehtävät

JOHDANTOTEHTÄVÄ:

## Terveyteen liittyvien ilmiöiden riippumattomuus

Todennäköisyyslaskennassa kahden tapahtuman riippumattomuus tarkoittaa, että jonkin tapahtuman todennäköisyys ei riipu toisen tapahtuman toteutumisesta ja toisin päin. Toisinaan tapahtumien riippumattomuus on ilmeistä tai asiayhteydestä helposti ymmärrettävissä. Näin on esimerkiksi silloin, kun heitetään kahta noppaa: toisen nopan silmäluku ei riipu siitä, mitä ensimmäisellä nopalla saatiin silmäluvuksi. Terveyteen liittyvät tapahtumat ovat kuitenkin usein monisyisiä, minkä vuoksi tapahtumien riippumattomuuden tarkastelu on tärkeää ja tehtävissä on syytä mainita, voidaanko tapahtumat olettaa riippumattomiksi vai ei.

Pohdi itsenäisesti tai keskustele parin kanssa, millä tavoin seuraavat asiat voivat liittyä toisiinsa ja ovatko ilmiöt toisistaan riippumattomia:

- ilmastonmuutos ja tartuntatautien leviäminen
- sosioekonominen asema ja terveydentila
- lapsityövoima ja tupakkatuotteet
- rokotteet ja kuolleisuus
- liikunta ja stressi
- ylipaino ja tyypin 2 diabetes
- ravitsemus ja uni
- mieliala ja ystävyyssuhteet
- tupakointi ja masennus
- aktiivisuusmittari ja unen laatu
- ilmaston lämpeneminen ja lonkkamurtumat
- pyöräilykypärän käyttäminen ja turvavyön käyttäminen

# Tehtävä 1: Lukiolaisen terveystottumuksia

Vuonna 2019 turkulaisista 1. ja 2. vuosikurssin lukio-opiskelijoista 33,8 % ei syönyt koululounasta päivittäin ja 85,4 % liikkui päivässä alle tunnin (Kouluterveyskysely 2019). Oletetaan, että neljäsosa opiskelijoista sekä jätti koululounaan väliin että liikkui alle tunnin päivän aikana. Millä todennäköisyydellä sattumanvaraisesti valittu turkulaisopiskelija syö koululounaan ja liikkuu vähintään tunnin joka päivä?

## PSST!

Kouluruokailu on niin opiskelijoiden kuin muun koulun henkilökunnan terveyttä ja hyvinvointia tukeva tekijä. Koululounas on suunniteltu monipuoliseksi ateriaksi, joka kokonaisuudessaan nautittuna on ravitsemuksellisesti täysipainoinen ja terveellinen. Kylläisenä ihminen jaksaa keskittyä paremmin opiskeluun ja sietää arjen kiireisyyttä. Kouluruokailu myös lisää yhteisiä ruokailuhetkiä ja ruoan arvostusta. Ruokailu antaa mahdollisuuden pieneen hengähdyshetkeen hektisen koulupäivän keskellä ja sillä on samalla sosiaalinen merkityksensä koulun arjessa.

Sosiaali- ja terveysministeriön asiantuntijakeskus UKK-instituutin laatimien, laajaan tutkimustietoon perustuvien liikkumissuositusten tarkoituksena on tiivistää terveyden kannalta tarvittavan viikoittaisen liikkumisen määrä sekä antaa esimerkkejä eri liikkumisen muodoista kaikenikäisille kansalaisille. Instituutin suosituksen mukaan 7-17-vuotiaiden lasten ja nuorten tulisi liikkua reippaasti ja rasittavasti tunti päivässä. Tunti voi kertyä useista liikuntahetkistä päivän aikana ja tärkeää olisi välttää pitkäkestoista paikallaanoloa, kuten istumista.

**Lisää aiheesta:**

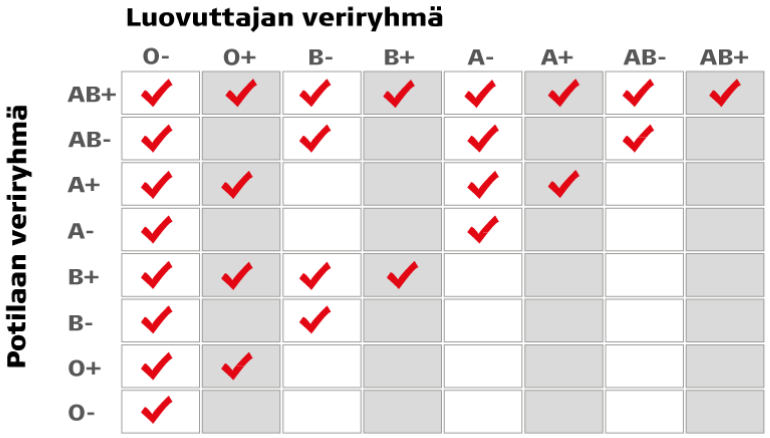
Kouluruokailun merkityksestä, Opetushallitus.

Lasten ja nuorten liikkumissuositus, UKK-instituutti.

Terveelliset välipalat voivat edistää jaksamista, Valtioneuvoston selvitys- ja tutkimustoimikunta.

# Tehtävä 2: Verenluovutus

Veriryhmät ovat tapa ryhmitellä verityyppejä ominaisuuksineen. Erilaisia veriryhmäluokitteluja on maailmassa yli 30, mutta yleisimmin käytössä ovat ABO- ja reesusveriryhmäjärjestelmät. Ihmisellä veri kuuluu yhteen neljästä ABO-veriryhmästä (A, B, AB ja O) ja on luokiteltavissa edelleen reesusnegatiivisiin ja reesuspositiivisiin: A-, A+, B-, B+, AB-, AB+, O- tai O+. Verenluovutuksessa potilaalle annetaan hänen veriryhmänsä kanssa yhteensopivaa verta alla olevan taulukon (Veripalvelu, SPR) mukaisesti. Verentarve noudattaa keskimäärin suomalaisten veriryhmäjakaumaa.



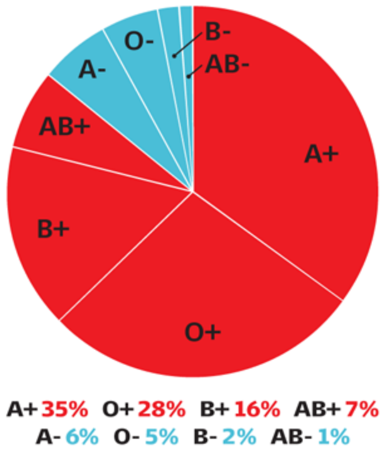
**Luovuttajan veriryhmä**

**Potilaan veriryhmä**

| | O- | O+ | B- | B+ | A- | A+ | AB- | AB+ |
|---|---|---|---|---|---|---|---|---|
| AB+ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| AB- | ✓ | | ✓ | | ✓ | | ✓ | |
| A+ | ✓ | ✓ | | | ✓ | ✓ | | |
| A- | ✓ | | | | ✓ | | | |
| B+ | ✓ | ✓ | ✓ | ✓ | | | | |
| B- | ✓ | | ✓ | | | | | |
| O+ | ✓ | ✓ | | | | | | |
| O- | ✓ | | | | | | | |

**Suomalaisten veriryhmäjakauma**

Blodgruppsfördelningen bland finländarna



Epäsopivan veren saaminen voi johtaa potilaan omien punasolujen rikkoutumiseen, millä voi olla vakavia seurauksia.

a) Kuinka suuri ryhmän suomalaisia tulee olla, jotta todennäköisyys, että ainakin yksi heistä kuuluu veriryhmään B+, ylittää 995 ‰?

b) Kuinka suuri satunnaisesti valituista henkilöistä koostuvan ryhmän tulee olla, jotta veriryhmään AB- kuuluvalle henkilölle löytyy sopiva luovuttaja yli 950 ‰:n todennäköisyydellä?

**Lisää aiheesta:**

Veriryhmät, niiden periytyvyys ja verenluovutus, Veripalvelu (Suomen Punainen Risti, SPR).

Mitä verenluovutuksessa tapahtuu? Veripalvelu.

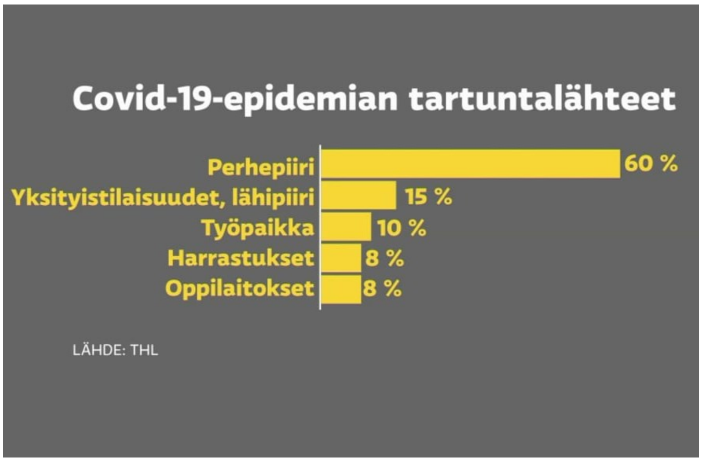
Voittoa tavoittelematon verenluovutus Euroopassa, EBA.

## PSST!

Tiesitkö, että verenluovutus on helppo tapa auttaa? Se vie vain noin 10 minuuttia. Luovutetusta verestä tehtyjä verivalmisteita tarvitaan muun muassa leikkauspotilaiden, onnettomuusuhrien, syöpää sairastavien ja keskoslasten hoidossa. Suomessa verenluovutustoiminta on perustunut aina vapaaehtoisuuteen ja maksuttomuuteen. Luovuttaja auttaa potilasta ilman korvausta, mikä on tärkeä asia verivalmisteen turvallisuuden kannalta. Joissain maissa verenluovutuksesta maksetaan korvaus, mikä voi aiheuttaa vakavia, jopa kuolemaan johtavia terveys- ja turvallisuusriskejä luovuttajan tavoitellessa palkkiota.

# TEHTÄVÄ 3: TARTUNTALÄHTEET

Yle Uutisten 5.11.2020 mukaan Covid-19-viruksen tartunnan jäljittäminen on aiempaan verrattuna helpottunut ja noin 60 % tartuntalähteistä selviää. Terveyden ja hyvinvoinnin laitoksen (THL) tietojen mukaan selvinneiden tartuntojen lähteet ovat seuraavat:



Covid-19, kuten moni virus- ja bakteeritauti, leviää hengitystie-eritteiden kautta pisaratartuntana suun, nenän ja silmien limakalvoilta. Koska Covid-19-tauti on erityisen herkästi tarttuva ja riskiryhmille jopa hengenvaarallinen, on suojatoimeksi pandemia-aikana vakiintunut tavanomaisen suojautumisen, kuten käsienpesun ja hyvän yskimishygienian lisäksi kasvomaskien käyttö julkisilla paikoilla.

Tarkastellaan viittä satunnaisesti valittua tartunnan saanutta henkilöä. Millä todennäköisyydellä

a) heillä kaikilla on eri tartuntalähde?

b) he kaikki ovat saaneet tartunnan perhepiiristä?

c) kaksi henkilöä on saanut tartunnan perhepiiristä ja kolme muualta?

## PSST!

Covid-19-taudin, kansanomaisesti koronan, aiheuttaja on koronaviruksiin kuuluva SARS-CoV-2-virus. Virus todettiin ensimmäisen kerran Kiinassa joulukuussa 2019 ja se levisi maailmanlaajuisesti alkukeväällä 2020. Maailman terveysjärjestö WHO julisti 11.3.2020 taudin levinneen pandemiaksi.

Koronaviruksia tunnetaan useita kymmeniä erilaisia, mutta vain pienehkö osa niistä voi tarttua ihmiseen ja aiheuttaa vakavia, jopa kuolemaan johtavia infektioita. Tällaisia ovat aiheuttaneet SARS-CoV-2-koronaviruksen lisäksi vain kaksi: SARS-CoV- ja MERS-CoV-koronavirukset, jotka levisivät kumpainenkin vakaviksi epidemiaksi asti. SARS vuonna 2003 ja MERS vuonna 2012. Tavallisesti koronavirukset aiheuttavat lieviä hengitystieinfektioita erityisesti syys- ja talvikaudella. Yleensä oireina on jonkinlainen yhdistelmä yskää, nuhaa sekä muita flunssan oireita.

**Lisää aiheesta:**

Koronavirus, Terveyden ja hyvinvoinnin laitos (THL).

Ajankohtaista tietoa Covid-19-taudista, THL.

# TEHTÄVÄ 4: SEKSUAALIVÄHEMMISTÖT

Vuoden 2017 kouluterveyskyselyn mukaan toisen asteen opiskelijoista noin 11 % (15 % tyttöoletetuista, 7 % poikaoletetuista) kokee kuuluvansa seksuaalivähemmistöihin. Suomalaisten lukiolaisten keskuudesta arvotaan 30 henkilön otos. Millä todennäköisyydellä näistä henkilöistä johonkin seksuaalivähemmistöön kuuluu:

a) puolet

b) ainakin 28

c) enintään yksi

d) ainakin yksi?

## PSST!

HLBTIQ on seksuaali- ja sukupuolivähemmistöistä käytetty lyhenne, joka tulee sanoista homo, lesbo, bi, trans, inter ja queer. Seksuaalinen suuntautuminen on ominaisuus, joka kertoo siitä, kehen ihminen ihastuu, rakastuu tai tuntee vetoa tunnetasolla tai eroottisessa mielessä. Monesti seksuaalinen suuntautuminen määritellään oman sukupuolen ja tunteiden kohteen sukupuolen perusteella, mikä on toisinaan ongelmallinen määrittelytapa niin seksuaalisen suuntautumisen kuin sukupuolien moninaisuuden kannalta. Puhuttiin sitten sukupuolesta tai seksuaalisesta suuntautumisesta, jokaisella ihmisellä on oikeus määritellä, kuka on ja kuinka haluaa tulla nähdyksi. Jokaisella on myös täysi oikeus olla tietämättä. Itsensä saa määritellä, olla määrittelemättä ja muuttaa mielensä määritelmien suhteen.

Kouluterveyskyselyssä kartoitetaan myös sukupuoli- ja seksuaalivähemmistöön kuuluvien nuorten hyvinvointia ja yhdenvertaisuuden toteutumista. Vuoden 2019 kyselyn tulokset osoittavat, että näihin vähemmistöihin kuuluvilla nuorilla oli muihin ikätovereihinsa verrattuna huomattavasti enemmän terveyteen ja hyvinvointiin sekä koulunkäyntiin ja kouluyhteisössä elämiseen liittyviä haasteita, kuten yksinäisyyden, kiusaamisen ja uupumuksen kokemuksia. Erityisen ongelmallista on sateenkaarinuorten kohtaaman väkivallan ja mielenterveyden ongelmien yleisyys sekä tarvitun avun ulkopuolelle jäämisen kokemukset muita nuoria useammin. Seksuaalivähemmistöjen syrjintäsuoja ja yhdenvertaisuuden edistämisvelvoite viranomaisille on kirjattu yhdenvertaisuuslakiin, mutta siitä huolimatta jokaisella meistä on myös yksilönä vastuu yhdenvertaisuuden toteutumisessa.

**Lisää aiheesta:**

Seksuaalinen suuntautuminen, Seta ry.

Itsemäärittelyoikeus ja normit, Seta ry.

Ensimmäinen HLBTIQ-henkilöiden tasa-arvoa koskeva EU:n strategia, Euroopan komissio.

Sukupuoli- ja seksuaalivähemmistöihin kuuluvien nuorten hyvinvointi, THL.

Terveyden ja hyvinvoinnin laitoksen joka toinen vuosi teettämä **Kouluterveyskysely** on perusopetuksen 4.-5. ja 8.-9. vuosiluokkalaisille, lukion 1.-2. vuoden opiskelijoille sekä ammatillisten oppilaitosten 1.-2. vuoden opiskelijoille suunnattu kysely. Siinä kartoitetaan eri-ikäisten lasten ja nuorten hyvinvointiin, terveyteen, koulunkäyntiin, osallisuuteen sekä avun ja tuen saamiseen liittyviä kokemuksia, tapoja ja tottumuksia sekä paikallisesti että maakuntatasolla. Kouluterveyskyselystä saatuja tietoja voidaan hyödyntää hyvinvoinnin edistämistyössä niin oppilaitoksissa, hyvinvointialueilla ja valtakunnallisesti.

# Tehtävä 5: Kansantaudit

Syöpäsairaudet ja diabetes ovat Suomessa kansantauteja. Yksilöllisten ja inhimillisten haittojen lisäksi kansantaudit vaikuttavat väestötasolla terveydentilaan, kuolleisuuteen ja työkykyyn sekä kuormittavat rajusti terveydenhuoltoa vaikuttaen samalla myös kansantalouteen.

Syöpäjärjestöjen tiedon mukaan joka kolmas suomalainen sairastuu jossakin elämänsä vaiheessa syöpään. Lähes kaksi kolmesta kuitenkin paranee. FinTerveys 2017-tutkimuksen mukaan tyypin 2 diabetekseen sairastuu elämänsä aikana n. 13% väestöstä (10 % naisista ja 15 % miehistä). Hiljattain todetusta diabeteksesta on mahdollista parantua täysin ja pitkään sairastanutkin voi saada sairauden remissiotilaan eli piiloon elämäntaparemontilla (terveellinen ravinto ja liikunnan lisääminen: eroon ylipainosta). Dataa remissioon päässeistä suomalaisista diabeetikoista ei ole kerätty, mutta brittitutkimuksessa* kahden vuoden seurannan jälkeen ilman painonhallintaohjelmaa remissioon pääsi 3 %, kun taas painonhallintaohjelman saaneista 36 % pääsi remissioon. Voidaan olettaa, että elintapaohjauksella on sama vaikutus myös suomalaisiin diabeetikoihin ja että sairaudesta parantuminen ei riipu sairastumisesta, vaan muista asioista (esim. hoidon onnistuminen ja yleiskunto).

Lääketieteellisissä tutkimuksissa** on havaittu voimakas yhteys syöpien ja tyypin 2 diabeteksen sairastavuuden välillä. Tämä johtuu erityisesti siitä, että molemmilla sairauksilla on yhteisiä riskitekijöitä. Tarkkaa arviota siitä, kuinka suuri osa potilaista sairastaa molempia samanaikaisesti, ei ole. Oletetaan tässä tehtävässä, että itse sairastumistapahtumat ovat toisistaan riippumattomia.

Laske, kuinka suurella todennäköisyydellä satunnainen suomalainen elämässään:

a) sairastuu sekä syöpäsairauteen että 2-tyypin diabetekseen

b) ei sairastu kumpaankaan

c) sairastuu ainakin toiseen

d) sairastuu enintään toiseen

e) sairastuu vain toiseen

f) sairastuu syöpään ja paranee siitä

g) sairastuu diabetekseen ja pääsee remissioon, kun hän on saanut painonhallintaohjelman.

Samanaikaissairastavuus eli komorbiditeetti tarkoittaa kahden tai useamman sairauden tai niin sanotun häiriön samanaikaista esiintymistä. Ilmiö on yleisempi sellaisilla sairauksilla, joilla on yhteisiä riskitekijöitä. Esimerkiksi ylipaino, tupakointi, korkea verenpaine ja veren korkea kolesterolipitoisuus ovat yhteisiä riskitekijöitä useille suomalaisille kansantaudeille, kuten syöville, diabetekselle sekä sydän- ja verisuonisairauksille. Myös erilaiset oppimisvaikeudet esim. hahmotusvaikeudet, lukivaikeus ja tarkkaavaisuuden vaikeudet esiintyvät usein päällekkäin.

* https://www.thelancet.com/journals/landia/article/PIIS2213-8587(19)30068-3/fulltext
** https://pubmed-ncbi-nlm-nih-gov.ezproxy.utu.fi/27219686/

**Lisää aiheesta**

Syöpä, Kaikki syövästä (Syöpäjärjestöt).

Diabeteksen ilmenemismuodot, Suomen Diabetesliitto ry.

Huonosti hoidettuna diabetes kuormittaa munuaisia
diabetes ja munuaiset & munuaisten tehtävät, Diabetesliitto ja Munuais- ja maksaliitto.

# Tehtävä 6: Puutiaisaivotulehdukset Saaristomeren alueella

Arviolta noin 500 000:ta suomalaista puree vuosittain punkki. Puutiaisaivokuumetta aiheuttavaa virusta kantaa noin 1-2 % riskialueiden* punkeista. Puutiaisaivotulehdus (TBE) on punkin kaikkien kehitysmuotojen (toukka, nymfi ja aikuinen) välityksellä tarttuva virustauti, jolle ei ole olemassa lääkehoitoa. Siihen on kuitenkin olemassa rokote, joka suojaa taudilta. Virus tarttuu punkista ihmiseen minuuteissa, minkä vuoksi ihon tarkastaminen punkkien varalta ei auta. Lähivuosina Suomessa on raportoitu vuosittain keskimäärin 60–80 puutiaisaivotulehdustapausta, mutta esimerkiksi vuonna 2020 diagnosoitiin 91 tapausta, mikä on enemmän kuin koskaan aikaisemmin. Määritä alla olevasta puutiaisaivotulehduksen esiintyvyys -taulukosta**

a) tapausaineiston moodi sekä

b) tapausten lukumäärien keskiarvo ja otoskeskihajonta

vuonna 2020 taulukossa nimetyillä paikkakunnilla. Selitä, kuinka hyvin keskiarvo mielestäsi kuvaa aineistoa ja mitä keskihajonnan arvo tarkoittaa tässä tilastossa.

* Eteläinen Suomi saaristoineen sekä Ahvenanmaan saaristo.

**Taulukko supistettu Varsinais-Suomeen ja Ahvenanmaahan *Puutiaisaivokuumeen esiintyvyys ja rokotussuositukset tartuntapaikkakunnittain* -taulukosta (THL, 2021).

| Kunta tai kaupunki | Tapausten lkm vuonna 2020 |
|---|---|
| Parainen | 7 |
| Ahvenanmaa | 15 |
| Kustavi | 0 |
| Naantali | 0 |
| Uusikaupunki | 2 |
| Kaarina | 2 |
| Turku | 1 |

THL ylläpitää valtakunnallista **tartuntatautirekisteriä** (tilastotietokanta) tartuntatautilakiin ja -asetukseen perustuen. Tartuntatautirekisteriin ilmoitetaan vuosittain noin 100 000 tartuntatautitapausta. Tietoja on kerätty vuodesta 1995 alkaen.

Tartuntatautirekisteriin voivat tehdä ilmoituksia lääkärit ja laboratoriot. Rekisterin lukuja tarkastellessa on hyvä pitää mielessä, että niihin vaikuttaa se, kuinka suuri osa tautiin sairastuneista hakeutuu terveydenhuoltoon sekä kuinka suurelle osalle tehdään mikrobiologisia tutkimuksia aiheuttajan toteamiseksi. Lisäksi lukuihin vaikuttavat muutokset diagnostisissa käytännöissä ja menetelmissä.

## PSST!

Puutiaiset, tuttavallisemmin punkit, ovat levinneet yhä laajemmalle alueelle Suomessa ja samalla punkkien välittämät borrelioosi- ja puutiaisaivotulehdustartuntojen määrät ovat kasvaneet. Yhtenä syynä siihen, että punkkeja tavataan yhä pohjoisemmilla leveysasteilla sekä yhä varhaisemmin keväällä ja myöhemmin syksyllä, pidetään ilmaston lämpenemistä. Uusia punkkiriskialueita on syntynyt jopa aivan kaupunkien puisto- ja viheralueille. Tämän vuoksi on hyödyllistä tietää, millä alueilla punkkeja on havaittu ja kuinka niiltä voidaan suojautua.

**Lisää aiheesta:**

Tietoa TBE-viruksesta: puutiaisaivotulehdus ja TBE:n esiintyvyys kartalla, THL

Tietoa borrelia-viruksesta ja oireista: borrelioosi, THL

Punkkien levinneisyys ja seuranta: punkkilive, Turun yliopisto

Punkkien levinneisyyttä kartoitetaan kansalaistutkimuksen avulla: mediatiedote, Turun yliopisto

# TEHTÄVÄ 7: KESTÄVYYSTREENI

Tutkimusnäyttöön perustuvan kansallisen Käypä hoito-suosituksen määritelmän mukaan kestävyysliikunta eli **aerobinen liikunta** on "suuria lihasryhmiä vähintään kohtalaisesti (suhteessa toteuttajansa suorituskykyyn) kuormittavaa, yleensä ainakin kymmeniä minuutteja (yhtäjaksoisesti tai jaksoittain) kestävää, aineenvaihduntaa ja hengitys- ja verenkiertoelimistöä kehittävää ja tällaisessa liikunnassa jaksamista ylläpitävää tai lisäävää liikuntaa".

Matalatehoisessa aerobisessa harjoittelussa liikutaan sykkeellä, joka on noin 60-70% maksimisykkeestä. Eräässä tutkimuksessa todettiin, että matalatehoisessa aerobisessa harjoittelussa koehenkilöiden sykkeen keskiarvo oli 130 bpm (beats per minute) ja keskihajonta 8 bpm. Syke on kutakuinkin normaalisti jakautunut.

a) Kuinka monella prosentilla koehenkilöistä syke oli alle 120 lyöntiä minuutissa?
b) Millä todennäköisyydellä satunnaisesti valitun koehenkilön syke oli välillä 120-140 lyöntiä minuutissa?
c) Millä sykevälillä syke oli 95,0 %:lla koehenkilöistä?

## PSST!

Ihmisen syke kertoo, montako kertaa sydän lyö yhden minuutin aikana. Syke vaihtelee ennen muuta kehon fyysisen aktiivisuuden ja rasituksen mukaan: mitä enemmän keho rasittuu, sitä enemmän sydämen on pumpattava verta työtä tekeville lihaksille. Syke on alhaisimmillaan levossa (nukkuessa) ja korkeimmillaan voimakkaassa fyysisessä rasituksessa esim. HIIT-harjoituksessa. Sykkeeseen vaikuttaa myös henkilön psyyke, kuten mielialan muutokset, stressi ja jännitys, sekä tupakka, alkoholi ja kofeiinipitoiset juomat.

Maksimisyke tarkoittaa sykettä, jonka henkilö voi saavuttaa kovassa fyysisessä suorituksessa ilman, että sen hetkellisestä saavuttamisesta on hänelle vaaraa. Maksimisyke on yksilöllinen ominaisuus, joka riippuu muun muassa henkilön iästä ja sukupuolesta. Vanhemmiten maksimisyke laskee. Yksi yleinen tapa arvioida henkilön maksimisykettä on vähentää henkilön ikä luvusta 220.

Leposyke tarkoittaa nimensä mukaisesti sykettä henkilön ollessa levossa, mutta kuitenkin hereillä. Tyypillisesti leposyke on aikuisella noin 50-90 lyöntiä minuutissa, mutta maksimisykkeen lailla leposyke on yksilöllinen ja myös geneettinen ominaisuus. Aktiiviliikkujilla leposyke voi hyvin olla alle viitearvon, sillä kestävyysliikunta ja –urheilu laskevat leposykettä. Jännittäviä huhuja liikkuu juoksijalegendojen historiallisen alhaisista leposykkeistä 25-30 lyönnin välillä. Kuitenkin esimerkiksi Paavo Nurmella oli nykytiedon valossa normaali sydän ja syke, vaikka aikanaan Amerikassa värikkäästi tokaistiin "alhaisin pulssi, korkein taksa".

**Lisää aiheesta**

Syke-käsitteistö kuntoilijalle, Polar.

Matalatehoinen peruskestävyysharjoittelu rakentaa kuntopohjaa ja auttaa palautumaan arjesta, Liikunta & Tiede 2-3/2019.

SYVENTÄVÄÄ TIETOA

# Myyttinen aamuaerobinen treeni

## FAKTOJA PUOLESTA JA VASTAAN

PSST!

Tämä kirjoitus on essee ja käsittelee aihetta useiden yliopistotutkimusten tuloksia tutkiskellen.

Aerobisella liikunnalla on tutkitusti monia terveyshyötyjä. Aerobisen harjoittelun seurauksena mm. kehon kyky vastustaa väsymystä paranee ja verenkiertoelimistön kyky kuljettaa happea paranee mm. hiussuoniston tilavuuden kasvaessa, happea sitovan hemoglobiinin määrän lisääntyessä verenkierrossa, punasolujen määrän lisääntyessä sekä sydämen iskutilavuuden laajentuessa. Toistuvassa kestävyystyyppisessä harjoittelussa keho oppii myös käyttämään rasvoja paremmin energiakseen silloin, kun hiilihydraatteja (glykogeenivarastot) ei ole käytössä. Aerobisella harjoittelulla on myös myönteisiä hermostollisia ja hormonaalisia vaikutuksia. Tunnetuimpia lienee endorfiinin lisääntynyt eritys, josta aiheutuu liikunnan jälkeinen hyvä olo ja mieli. Matalatehoiseen aerobiseen harjoitteluun liittyen puhutaan kansanomaisesti myös rasvanpolttoliikunnasta tai liikkumisesta rasvanpolttosykkeellä, jolla tarkoitetaan noin 60-70 %:lla maksimisykkeestä tehtävää harjoitusta. Tässä esseessä pureudutaan erityisesti aamulla ennen aamiaista tehtävään aerobiseen harjoitteluun liittyviin myytteihin ja totuuksiin.

### Aamulla tehtävän aerobisen harjoituksen vaikutuksia päivä- tai ilta-aerobiseen treeniin verrattuna

Yöunen aikana ihmisellä verenpaine alenee verrattuna diurnaaliseen eli päivänaikaiseen verenpaineeseen, mikä laskee verenkiertoelimistön stressiä. Sekä akuutista aerobisesta harjoituksesta aiheutuva madaltunut verenpaine että unen laatu voivat vaikuttaa tähän luontaiseen alenemisreaktioon. Kimberly Fairbrotherin ym. (2014) tutkimuksessa selvitettiin aerobisen harjoituksen ajoituksen vaikutusta elimistön sisäiseen vuorokausirytmiin liittyviin verenpaineen muutoksiin ja unen struktuuriin. Tutkimuksessa saatiin selville, että harjoitus aamuseitsemältä (klo 7) aiheutti suuremman pudotuksen öiseen systoliseen verenpaineeseen kuin jos sama harjoitus tehtiin iltapäivällä (klo 13) tai illalla (klo 19). Iltapäivän harjoitukseen verrattuna aamuharjoitus pidensi syvän unen vaihetta. Myös Yamanakan ym. (2015) tutkimuksessa aamulla ja illalla tehdyllä harjoituksella huomattiin olevan erilaisia vaikutuksia. Aamulla tehty harjoitus tehosti pitkällä aikavälillä parasympaattisen hermoston toimintaa, mitä osoitti sydämen sykkeen vaihteluvälin laajeneminen yöunien aikana. Iltaharjoittelu aktivoi sympaattisen hermoston toimintaa, mitä indikoi sydämen sykkeen nousu harjoitusta seuraavan yön unen aikana. Nämä aamu- ja iltaharjoittelun erilaiset vaikutukset heijastuvat sykemuutosten lisäksi mm. melatoniinin erityksen vuorokausirytmiin (iltaharjoitus siirsi melatoniinin erityksen aloitusvaihetta myöhemmäksi). Aamulla tehty harjoitus rauhoittaa sydäntä ja auttaa lihaksia palautumaan yön aikana, kun taas iltaharjoittelu kiihdyttää sydän- ja verisuonielimistön toimintaa yöunien aikana.

Harjoituksen jälkeinen verenpaineen lasku on suurempi ilta/iltapäivä- kuin aamuharjoituksen jälkeen. Toisaalta tämä harjoituksen jälkeinen laskuefekti voi aamuisin naamioitua verenpaineen vaihtelun normaaliin vuorokausirytmiin, jossa aamuisin verenpaine on iltaa korkeammalla. De Briton yms. (2015) tutkimuksessa selvisi, että aamuharjoituksessa systolisen verenpaineen laskuefekti on suhteessa suurempi huomioitaessa verenpaineen luontainen vuorokausivaihtelu. Tähän aamutreenin jälkeiseen laskuefektiin liittyy myös sydämen sykkeen maltillisempi nousu sekä pienempi nousumuutos sympaattisen hermoston ja vagus hermon (parasympaattisen hermoston kehoa rauhoittava ja stressiä vähentävä hermo) tasapainotilassa treenin jälkeen. Iltatreenin jälkeiseen verenpaineen laskuun liittyi verisuonien välitöntä laajenemista, minkä seurauksena aamutreeniä voidaan pitää iltatreeniä parempana vaihtoehtona stressin vähentämisessä.

Arjessa puhutaan usein aamuaerobisen treenin hyödyistä painonhallinnassa ja laihduttamisessa. Alizadehin yms. (2015) tutkimuksessa kuitenkin havaittiin, että aamulla tehtävästä aerobisesta treenistä ei ole iltapäivällä tehtyyn treeniin verrattuna merkittäviä painonhallinnallisia etuja ylipainoisilla. Harjoitusaika ei siis vaikuttanut ruokahaluun ja ruoan vetoisuuteen. Toisaalta kyseisessä tutkimuksessa selvisi, että aamuaerobinen treeni saattaa lisätä kylläisyydentunnetta hieman päivällä tehtävää harjoitusta enemmän. Tämä ei kuitenkaan vaikuttanut tutkittavien kokonaisenergiankulutukseen.

Toisessa hyvin spesifiin ryhmään, vaihdevuotisiin naisiin, kohdistuvassa Ammarin (2015) tutkimuksessa saatiin tuloksia, joiden mukaan aamulla tehdyn aerobisen harjoituksen myönteiset terveysvaikutukset eivät yllä myöhään iltapäivällä (klo 17) tehdyn aerobisen harjoituksen vaikutuksiin. Vaihdevuodet saattavat nostaa riskiä korkeaan verenpaineeseen ja verenpainetautiin sekä epätavalliseen lipidiprofiiliin (kokonaiskolesteroli-, SBP-, DBP-, HDL-, LDL- & TG-arvot). Tutkimuksessa tutkittiin pelkän lääkityksen, lääkityksen ja aamuaerobisen harjoittelun, sekä lääkityksen ja iltapäiväaerobisen harjoittelun vaikutuksia verenpaineeseen ja lipidikoostumukseen 49-60-vuotiailla naisilla. Tutkimuksen tulokset osoittivat tilastollisesti merkitsevät erot kaikkien em. ryhmien välillä. Eniten systolinen ja diastolinen verenpaine laskivat iltapäivällä harjoitelleilla naisilla. Myös lipiditasot olivat tilastollisesti merkittävän erilaiset kaikilla ryhmillä - jälleen iltapäivätreenaajien eduksi.

## Aamuaerobinen treeni - aamiaisella vai ilman

Ajatus tyhjällä vatsalla tehdyn aamuaerobisen harjoituksen tehokkuudesta rasvanpoltossa perustuu yön aikana tapahtuvaan elimistön hiilihydraattivarastojen tyhjenemiseen, jolloin aamulla tyhjällä vatsalla tehdyssä aerobisessa treenissä elimistön on otettava energiansa muualta eli muutettava rasvaa treenissä tarvitsemakseen energiaksi. Kuitenkin Hulmin (2013) mukaan rasvanpolttoa ajatellen on tyhjällä vatsalla tehdyn aamuaerobisen treenin etu varsin vähäinen. Seuraavassa esittelen Hulmin blogikirjoitusta mukaillen muutamia argumentteja tyhjällä vatsalla tehdyn aamuaerobisen treenaamisen puolesta ja vastaan.

Aamuaerobinen treeni ennen aamiaista voi olla miellyttävämpää kuin täydellä vatsalla aamiaisen jälkeen: olo on kevyt ja treenillä on virkistävämpi vaikutus. Treenaaminen ennen aamiaista saattaa olla myös ajankäytöllisesti kätevää, sillä aikaa ruoan sulatteluun ei työaamuina ennen harjoitusta välttämättä ole. Yöunien eli pienimuotoisen paaston jälkeen treenaaminen saattaa ainakin epäterveellisellä ruokavaliolla tehostaa glukoositoleranssia ja insuliiniherkkyyttä, mikä esim. Van Proeyenin (2011) tutkimuksen mukaan edistää terveyttä. Samassa tutkimuksessa havaittiin myös, että rasvojen hapetuskapasiteetti eli tuttavallisemmin rasvanpoltto tehostui ja harjoitusvaste parantui joillain kestävyysurheilijoilla tyhjällä vatsalla treenatessa. Nämä ilmiöt perustuvat pitkälti juuri lihasten glykogeenivarastojen keskimääräistä tyhjempään tilaan aamulla. Samanaikaisesti treenissä tarvittu energia kehossa saadaan tehostamalla ja lisäämällä rasvojen hapetusta. Gonzalesin ym. (2013) tutkimuksessa selvisi, että osalla tutkittavista ruoan kulutus oli pienempää päivinä, joina he olivat tehneet aamuaerobisen harjoituksen ilman aamiaista. Aamiaisen nauttineilla ei tällaista eroa ollut eli jonkinlaista yhteyttä näläntunteen vähenemiseen saattaa paastossa tehdyllä aerobisella aamutreenillä olla.

Tyhjällä vatsalla tehdyn aerobisen aamutreenin lisähyödyt aamiaisen jälkeen tehtyyn vastaavaan treeniin verrattuna ovat pääosin lähes olemattomia ja joissain tapauksissa tyhjällä vatsalla treenaamisesta voi olla jopa haittaa. Tyhjällä vatsalla treenatessa huono olo saattaa yllättää ja harjoituksen teho laskea. Pienikin energia- tai proteiinilataus ennen harjoitusta tai sen aikana mahdollistaa treenin tehostumisen. Aamuisin suorituskyky on alhaisempi kuin muina vuorokauden aikoina. Toisaalta aamuaerobisessa treenissä ei monesti tarkoituksena ole kovatehoisen treenin tekeminen. Paastotilassa tehdyn aamuaerobisen harjoituksen haittana on se, että lihaskatabolia ei lihasproteiinin käyttö energiaksi lisääntyy (Lemmon & Mullin 1980). Wilsonin ym. (2012) tutkimuksessa selvisi, että monta kertaa viikossa kovaa aerobisesti treenaavilla lihasmassa kärsii: usein, kovaa ja aerobisesti liikkuvilla voimaharjoittelijoilla lihasten kasvu vähenee.

Liikunnan aikainen rasvan hapetus energiaksi on lopputuloksen eli kehon rasvavarastojen vähenemisen kannalta varsin pikkuruisessa roolissa. Lopputulos eli harjoituksen jälkeisten päivien ja viikkojen aikana kehossa tapahtuneet muutokset rasvan määrässä ratkaisevat. Esimerkiksi korkeaintensiteettinen intervalliharjoittelu (HIIT-High Interval Training) näyttäisi Gibalan ym (2012) tutkimuksen mukaan toimivan elimistön rasvan vähentämisessä yhtä hyvin kuin pidempikestoinen rasvanpolttosyketreeni. Paastossa tehty aamuaerobinen harjoitus ei merkitsevissä määrin tehosta rasvanpolttoa pitkällä aikavälillä verrattuna harjoitukseen, jota edeltää ruokailu. Shimadan ym (2013) tutkimuksessa havaittiin, ettei 24 tunnin aikaisessa energiankulutuksessa ollut eroa aamiais- ja paastotreenaajilla. Tutkimuksessa havaittiin pieni ero harjoituksen aikaisessa rasvanpoltossa paastoajien eduksi. Toisaalta tutkimukset (Paoli ym. 2011 & Hulmi ym. 2005) osoittavat, että harjoitusta edeltävän ravinnon, etenkin proteiinin, nauttiminen lisää liikunnan aikaista ja sen jälkeistä energiankulutusta, sekä kohottaa proteiinisynteesiä, johon toisaalta myös kuluu energiaa. Van Proyenin (2011) kuusiviikkoisella tutkimusjaksolla osoittautui, ettei yön jälkeisessä paastossa tehty aamuaerobinen harjoitus vähentänyt kehon painoa enempää kuin sokeripitoisen aamiaisen jälkeen tehty. Tässä tutkimuksessa kehon koostumusta ei tutkittu.

Paastolla aamuaerobiseen treeniin lähtevien, painonpudotuksesta haaveilevien harjoittelijoiden on hyvä pitää mielessä, että pelkän liikunnan merkitys laihdutuksessa on vähäinen. Liikunnan lisäksi tarvitaan ravintosuositusten mukaista ravitsemusta ja itsekuria syömisen suhteen, jos mielessä on laihduttaminen. Ravinnon ja ruokailutottumusten muutokset ovat painonpudotuksessa merkittävässä roolissa. Liikunta on hyvä apukeino saavutettujen laihdutustulosten ylläpitämisessä ja itsessään terveyttä edistävää riippumatta siitä, harrastaako sitä syömisen jälkeen tai paastossa.

## Mitä tästä kaikesta pitäisi ajatella?

Kaikkiaan ei näyttäisi olevan merkittävää vaikutusta sillä, tekeekö aamuaerobisen harjoituksen tyhjällä vatsalla vai aamiaisen jälkeen. Painonhallinnan näkökulmasta kyse on kokonaiskulutuksesta eli siitä, että energian saanti ja kulutus ovat tasapainossa. Harjoittelun tavoitteet (esim. kunnon kohentuminen, kiinteytyminen tai painonpudotus) ratkaisevat, kannattaako ennen treeniä syödä vai ei. Tutkimuksissa havaittiin myös yksilöllisiä eroja rasvanpolttotehokkuudessa harjoituksen aikana sen mukaan, tehtiinkö treeni tyhjällä vai täydellä vatsalla. Oman kehon tuntemuksia on tärkeää treenatessa kuunnella. Rasvanpoltto on suhteellisesti tehokkainta, kun tehdään töitä sykkeen aerobisella alueella (ns. rasvanpolttosyke). Absoluuttisesti eniten rasvaa palaa silti kovatehoisessa liikunnassa.

Liikuntasuositusten mukaan (UKK-instituutin liikuntapiirakka 18-64-vuotiaille) aikuisen tulisi harrastaa kestävyyskuntoa parantavaa aerobista liikuntaa vähintään tunnin ja 15 minuuttia viikossa rasittavasti tai kaksi ja puoli tuntia reippaasti. Aerobinen harjoittelu on harjoitteluajankohdasta riippumatta terveyttä edistävää ja tekee hyvää niin keholle kuin mielelle. Aerobisen treenin ajankohdan valinnalla voi olla yksilöllinen merkitys ihmiselle riippuen siitä, miten hänen arkensa aikataulut rakentuvat. Harjoituksen tekeminen aamulla voi olla päivä- ja iltatreeniin verrattuna suositeltavaa matalamman stressitason, rauhoittumisen, nukahtamisen ja levollisen unen kannalta.

**Lähteet:**

Kimberly Fairbrother, Ben Cartner, Jessica R Alley, Chelsea D Curry, David L Dickinson, David M Morris and Scott R Collier: Effects of exercise timing on sleep architecture and nocturnal blood pressure in prehypertensives. Vasc Health Risk Manag. 2014; 10: 691–698.

Yamanaka Y, Hashimoto S, Takasu NN, Tanahashi Y, Nishide SY, Honma S, Honma K: Morning and evening physical exercise differentially regulate the autonomic nervous system during nocturnal sleep in humans. Am J Physiol Regul Integr Comp Physiol. 2015 Nov 1;309(9):R1112-21

Alizadeh Z, Mostafaee M, Mazaheri R, Younespour S.: Acute Effect of Morning and Afternoon Aerobic Exercise on Appetite of Overweight Women. Asian J Sports Med. 2015 Jun;6(2):e24222.

Ammar T.: Effects of aerobic exercise on blood pressure and lipids in overweight hypertensive postmenopausal women. J Exerc Rehabil. 2015 Jun 30;11(3):145-50.

de Brito LC, Rezende RA, da Silva Junior ND, Tinucci T, Casarini DE, Cipolla-Neto J, Forjaz CL.: Post-Exercise Hypotension and Its Mechanisms Differ after Morning and Evening Exercise: A Randomized Crossover Study. PLoS One. 2015 Jul 17;10(7):e0132458.

Juha Hulmi: Tyhjällä mahalla tehty aamuaerobinen – totuus? 2013. Blogikirjoitus. https://lihastohtori.wordpress.com/2013/02/17/tyhjalla-mahalla-aamuaerobinen-j/ *Luettu 16.10.2021*

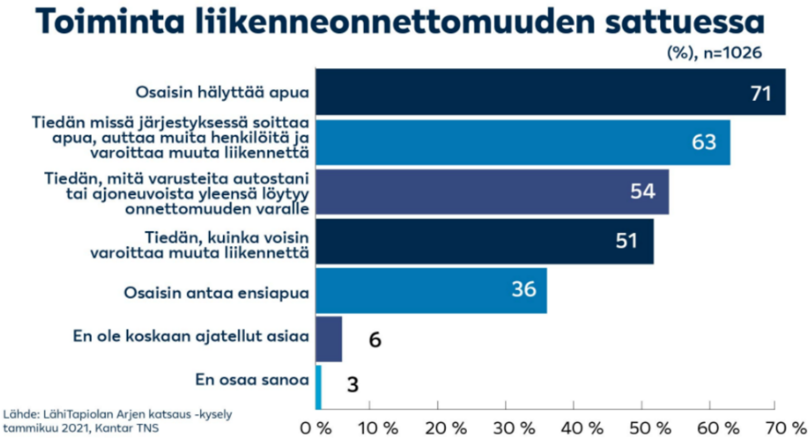
- Gibala MJ, Little JP, Macdonald MJ, Hawley JA: Physiological adaptations to low-volume, high-intensity interval training in health and disease. J Physiol. 2012 Mar 1;590(5):1077-84.
- Gonzalez JT, Veasey RC, Rumbold PL, Stevenson EJ: Breakfast and exercise contingently affect postprandial metabolism and energy balance in physically active males.Br J Nutr. 2013 Aug;110(4):721-32.
- Hulmi JJ[1], Volek JS, Selänne H, Mero AA: Protein ingestion prior to strength exercise affects blood hormones and metabolism. Med Sci Sports Exerc. 2005 Nov;37(11):1990-7.
- Lemon PW, Mullin JP: Effect of initial muscle glycogen levels on protein catabolism during exercise. J Appl Physiol Respir Environ Exerc Physiol. 1980 Apr;48(4):624-9.
- Paoli A, Marcolin G, Zonin F, Neri M, Sivieri A, Pacelli QF: Exercising fasting or fed to enhance fat loss? Influence of food intake on respiratory ratio and excess postexercise oxygen consumption after a bout of endurance training. Int J Sport Nutr Exerc Metab. 2011 Feb;21(1):48-54.
- Shimada K, Yamamoto Y, Iwayama K, Nakamura K, Yamaguchi S, Hibi M, Nabekura Y, Tokuyama K: Effects of post-absorptive and postprandial exercise on 24 h fat oxidation. Metabolism. 2013 Jun;62(6):793-800.
- Van Proeyen K, Szlufcik K, Nielens H, Ramaekers M, Hespel P: Beneficial metabolic adaptations due to endurance exercise training in the fasted state. J Appl Physiol (1985). 2011 Jan;110(1):236-45.
- Wilson JM, Marin PJ, Rhea MR, Wilson SM, Loenneke JP, Anderson JC: Concurrent training: a meta-analysis examining interference of aerobic and resistance exercises. J Strength Cond Res. 2012 Aug;26(8):2293-307.

# Tehtävä 8: Ensiaputaidot

Ensiapu on loukkaantuneelle tai sairastuneelle henkilölle tapahtumapaikalla usein nopeasti annettavaa apua, jolla pyritään turvaamaan autettavan peruselintoiminnot sekä estämään hänen tilansa paheneminen. Ensiavun käsite pitää sisällään monenlaisia auttamisen taitoja; se voi olla tilanteesta riippuen esimerkiksi hätäkeskukseen soittamista, tilannearvion tekemistä onnettomuuspaikalla, sairauskohtauksissa auttamisen taitoja, haavojen sidontaa, henkistä tukemista, tukehtumisen estämistä tai paineluelvytystä. Ensiapua ja hätäensiapua voidaan antaa, vaikka antajalla ei olisi erityistä ammattitaitoa tai välineitä. Usein maallikon antama apu parantaa huomattavasti apua saavan mahdollisuuksia selviytyä ja toipua.

LähiTapiolan Arjen katsaus -kyselyn (n=1026) tulosten mukaan noin 36% ihmisistä osaisi oman arvionsa mukaan antaa ensiapua liikenneonnettomuuden sattuessa. Valitaan viiden hengen satunnainen otos suomalaisia. Asetetaan satunnaismuuttujaksi X otokseen valittujen liikenneonnettomuustilanteessa ensiaputaitoisten henkilöiden lukumäärä.

a) Perustele, mitä jakaumaa satunnaismuuttuja X noudattaa ja miksi.
b) Muodosta satunnaismuuttujan X jakauma, esitä se taulukossa ja selvitä, mikä on sen todennäköisin arvo.



**Toiminta liikenneonnettomuuden sattuessa**

Lähde: LähiTapiolan Arjen katsaus -kysely tammikuu 2021, Kantar TNS

Auttamisen taitoja harjoitellaan suomalaiskouluissa jo alakoulusta lähtien ja moni suomalainen käy ensiapukurssin tai -kursseja. Suomen Punaisen Ristin (SPR) Ensiapukurssi EA 1® -koulutus on virallinen sertifioitu koulutus, josta saa Punaisen Ristin ensiapukortin eli pätevyystodistuksen. Kurssin suorittanut hallitsee ensiavun antamisen perusteet kurssin sisällön mukaisissa aiheissa. SPR Ensiapukurssi EA 1® -todistus on voimassa 3 vuotta.

PSST!

Bee Gees yhtyeen Stayin' Alive -kappaleessa on oikea rytmi (104 lyöntiä minuutissa) paineluelvytyksen antamiseen. Kappale on esimerkiksi Amerikan sydänliiton (the American Heart Association) suosittelema elvytyksen taustalle.

**Lisää aiheesta:**

Ensiapuohjeita eri tilanteisiin, Suomen Punainen Risti.

Maailman elvytyspäivä 16.10.(WorldRestartAHeart) ja biisejä elvytystilanteeseen, SPR ja Sydänliitto.

Arjen katsaus -kysely 2021: Miten suomalaiset toimivat liikenneonnettomuuden sattuessa, LähiTapiola.

# Tehtävä 9: Koulukiusaaminen

Koulukiusaaminen on koulussa tai muussa oppilaitoksessa tapahtuvaa psyykkistä, sosiaalista tai fyysistä väkivaltaa, joka on tarkoituksellista ja toista vahingoittavaa. Kiusaaminen on sen uhrille henkisesti vahingollista ja voi aiheuttaa traumoja, jotka saattavat jatkua läpi elämän. Kouluterveyskyselyssä 2021 yhteensä 91 560 8. ja 9. luokkalaiselta kysyttiin: "Kuinka usein sinua on kiusattu koulussa tämän lukukauden aikana?" Merkitään edellä mainittua kiusaamisen yleisyyttä kartoittavaa kysymystä diskreetillä satunnaismuuttujalla X. Alla olevan taulukon jakauma mukailee saatuja tuloksia.

a) Laadi jakaumasta pylväsdiagrammi.
b) Määritä suhteellinen frekvenssi f%, summafrekvenssi sf ja suhteellinen summafrekvenssi sf%.
c) Selvitä, millä todennäköisyydellä 8. tai 9. luokan opiskelija on joutunut kiusaamisen uhriksi lukukauden aikana.

| x | f |
|---|---:|
| useita kertoja viikossa | 2 106 |
| noin kerran viikossa | 3 388 |
| harvemmin | 17 213 |
| ei lainkaan | 68 853 |

Koulukiusaamisella on niin yksilö-, yhteisö- kuin väestötasonkin seurauksia. Kiusaaminen aiheuttaa turvattomuuden tunnetta sekä kiusatulle että hänen koululuokalleen. Se on vakava uhka hyvinvoinnille ja terveydelle, ja voi aiheuttaa mm. itsetunto-ongelmia, ahdistuneisuutta, yksinäisyyttä ja vaikeutta luottaa muihin ihmisiin. Kiusaaminen voi jättää jälkensä kiusattuun pysyvästi ja olla yhteydessä psyykkisen ja sosiaalisen terveyden sekä työelämässä menestymisen ongelmiin myös aikuisuudessa. Kiusaamisen kova hinta näkyy sekä sen inhimillisissä että kansanterveydellisissä kustannuksissa.

Alla olevat kuvat Unescon *Behind the numbers: Ending school violence and bullying* -raportista (2019).

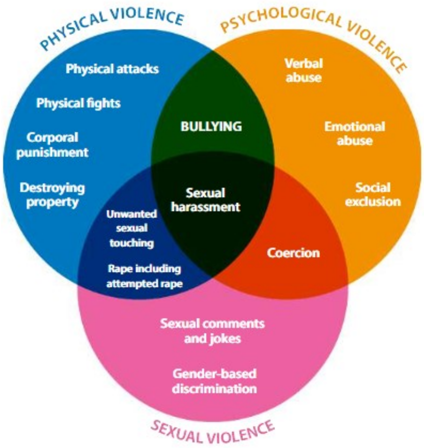
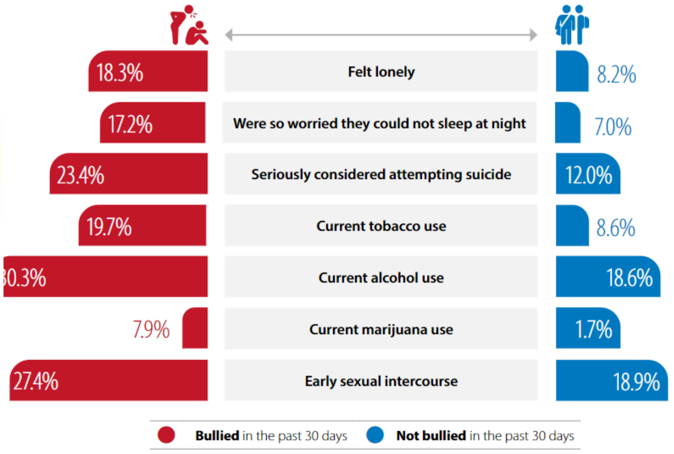


**Figure 21. Differences in mental health status and the prevalence of risk behaviours between students who were bullied and those who were not bullied**



Source: Secondary analysis calculations based on GSHS data.

**Lisää aiheesta:**

Tietoa koulukiusaamisesta, KiVa Koulu

Niko Tiuraniemen tarina kiusatuksi joutumisesta, Yle 20.2.2022

Kiusaamisen yleisyys yläkoulussa ja toisella asteella vuosina 2006-2021, THL Kouluterveyskysely

Kiusaamisen vastainen toimenpideohjelma KiVa Koulu, KiVA Koulu -ohjelma ja Turun yliopisto

Koulukiusaamisen vastaisten ohjelmien arviointi, Yle 20.10.2021

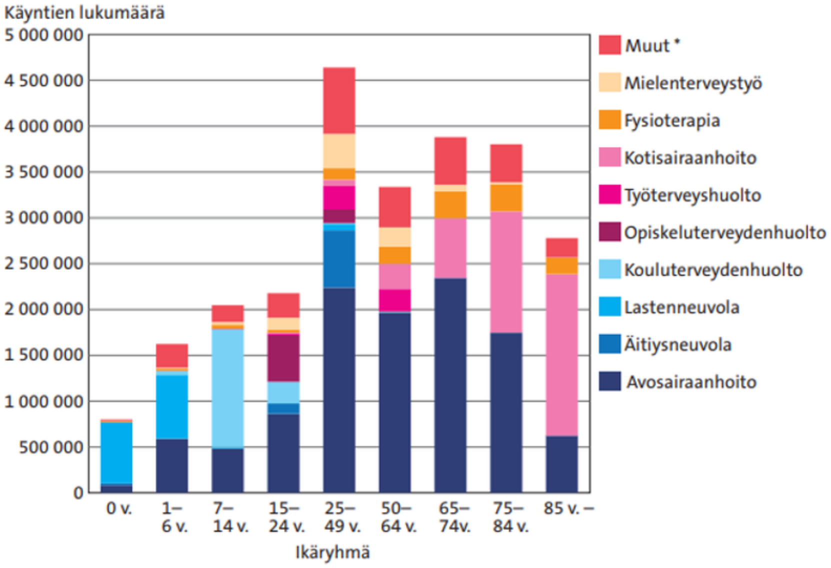
Harjoituksia mielenterveyden vahvistamiseen, Suomen Mielenterveys ry

Koulukiusaaminen on maailmalla ja sitä ehkäiseviä asioita, Unesco, The Global Education 2030 Agenda -hanke

# TEHTÄVÄ 10: TERVEYSPALVELUIDEN KÄYTTÖ ERI IKÄRYHMISSÄ

Tutkitaan iän ja terveyskeskuskäyntien välistä yhteyttä. Alla olevassa kuviossa on esitetty perusterveydenhuollon käynnit** palvelumuodoittain ja ikäryhmittäin vuonna 2017 (Lähde: Suomalaisten sosiaali- ja terveyspalveluiden käyttö tilastojen valossa).

a) Laadi pylväskuvion pohjalta taulukko perusterveydenhuollon käyntien lukumääristä eri ikäryhmissä 100 000 käynnin tarkkuudella. Muodosta taulukon pohjalta hajontakuvio.
b) Sovita aineistoon lineaarinen malli ja toisen asteen polynominen malli.
c) Määritä mallien selitysasteet ja lineaarisesta mallista iän ja terveyskeskuskäyntien välinen korrelaatiokerroin. Tulkitse riippuvuuden voimakkuutta ja pohdi tulosta mahdollisesti vääristäviä tekijöitä.



* Sisältää päihdetyön, perhesuunnittelu-, ehkäisy-, kasvatus- ja perheneuvolapalvelut sekä muut neuvolapalvelut, seulonnat ja joukkotarkastukset, apuvälinepalvelut, puhe- toiminta-, jalka- ja ravitsemusterapian, terveyssosiaalityön, muut kuntoutus- ja erityisterapiat, päivätoiminnan, päiväsairaalatoiminnan sekä muun palvelutoiminnan.

** Perusterveydenhuollon avohoidon käynnit vastaanotolla, kotikäynnit, työpaikkakäynnit ja sairaalakäynnit.

Alle kouluikäisillä lapsilla lastenneuvolapalvelut ovat yleisin palvelumuoto. Työikäiset käyttävät eniten avosairaanhoidon ja työterveyshuollon palveluja sekä äitiysneuvolapalveluja. Vanhimmissa ikäryhmissä avosairaanhoito ja kotisairaanhoito ovat merkittävimmät palvelumuodot.

## PSST!

Suomalaisten terveys on parantunut viime vuosikymmeninä paljon, mutta siitä huolimatta terveyserot eivät ole kaventuneet eri väestöryhmien välillä. Muun muassa henkilön ikä, sukupuoli, asuinalue, siviilisääty, äidinkieli ja sosioekonominen asema vaikuttavat kukin osaltaan yksilön terveyteen, mutta myös terveyteen väestötasolla luoden eriarvoisuutta väestöryhmien välille. Esimerkiksi terveys- ja toimintakyky heikkenevät iän myötä, miehet sairastavat naisia enemmän sydän- ja verisuonitauteja, asuinalueiden erilaiset elinolosuhteet (kuten työllisyystilanne, elämänmeno, sosiaali- ja terveyspalveluiden saavutettavuus) heijastuvat kansantautien sairastavuuteen, parisuhteessa elävien koettu terveys on parempi ja elinikä pidempi kuin yksineläjillä ja suomenruotsalaiset ovat suomenkielisiä terveempiä. Mitä parempi on henkilön sosioekonomisen asema, sitä terveempi ja hyvinvoivempi hän todennäköisesti on. Vähäinen koulutus, epävarma ja heikko ammatillinen asema (pienituloisuus) sekä vähäinen varallisuus ovat vahvasti yhteydessä toisiinsa sekä huonon terveyden ja pahoinvoinnin kasaantumiseen. Erot terveyskäyttäytymisessä, kuten tupakoinnissa ja alkoholinkäytössä, selittävät arviolta noin puolet terveyden sosioekonomisista eroista.

Sosioekonomiseen asemaan liittyvät terveyserot ovat monimutkainen toisiinsa linkittyvien syiden ja seurausten vyyhti. Yhteiskunnalliset rakenteet vaikuttavat sosioekonomisen aseman muotoutumiseen ja terveyttä heikentävän sosiaalisen eriarvoisuuden syntyyn. Kuitenkin yksi suomalaisen terveyspolitiikan perusperiaatteista on tarjota jokaiselle terveydentilan edellyttämät riittävät ja laadukkaat palvelut sosioekonomisesta asemasta, taloudellisista edellytyksistä tai asuinalueesta riippumatta. Käytännössä erot terveysosaamisessa sekä terveyteen liittyvissä tiedoissa ja taidoissa aiheuttavat sen, että eri sosioekonomisessa asemassa olevat henkilöt eivät ole yhdenvertaisia palveluihin hakeuduttaessa. Lisäksi esim. asuinalue ja ikä vaikuttavat palveluiden saatavuuteen ja saavutettavuuteen. Edellä mainituista syistä perusterveydenhuollon palveluiden käyttö vaihtelee eri väestöryhmissä.

TIETOTÖÖTTÄYS

# Rokote - miksi piikille?

Rokotteista, rokottautumisesta ja rokoteohjelmista on käyty eri medioissa kiivasta keskustelua viime vuosina. Huomiota ovat saaneet muun muassa 2009-2010 Pandemix-sikainfluenssarokotteesta aiheutuneet narkolepsiatapaukset ja puutiaisaivokuumerokote, johon keväästä 2018 alkaen riskialueilla asuvat yli kolmevuotiaat henkilöt ovat olleet oikeutettuja osana kansallista rokotusohjelmaa. Keväällä 2017 keskustelua herätti Keski-Euroopasta matkailijoiden mukana Suomeen paluun tehnyt erittäin tarttuva tuhkarokko, jolla on ikäviä jälkitauteja ja jota vastaan suomalaisia 60- ja 70- luvuilla syntyneitä on harvoin rokotettu. Maaliskuusta 2020 lähtien on medioissa hetki hetkeltä seurattu Covid-19-viruksen leviämistä maailmalla ja Suomessa.

Rokotteista saadut sivuoireet sekä yksittäisten henkilöiden tarinat niistä ovat saaneet julkisuutta mm. sosiaalisessa mediassa. Oireiden ja tarinoiden on arveltu vaikuttaneen rokotekriittisyyden lisääntymiseen ja rokotekattavuuden laskuun osassa Suomea. Rokotteista kieltäytymistä perustellaan sivuoireiden lisäksi mm. taipumuksella lieväoireiseen sairastamiseen. Osa tautitartunnan saaneista saa vain lieviä oireita, mutta toisilla taudin aiheuttamat oireet saattavat olla hengenvaarallisia esimerkiksi elimistön immuunijärjestelmän heikon tilan, korkean iän tai sairauden vuoksi. Rokottamisessa kyse ei ole vain omasta terveydestä ja mukavuudesta, vaan läheisten ja muiden ihmisten terveydestä ja toisinaan jopa elämästä. Rokotekriitikot ovatkin saaneet asiantuntijoilta tiukkaa palautetta rokottamatta jättäytymisen kansanterveydellisesti vakavista seurauksista.

Rokotekielteisyys on huolestuttava trendi, sillä rokotteen antama suoja perustuu paitsi tartunnan todennäköisyyden pienenemiseen rokotetulla, niin myös laumasuojaan eli laumaimmuniteettiin: riittävän monen rokottauduttua, myös rokottamattomat ovat suojassa taudin leviämisen estyessä populaatiossa. Päivä- ja iltapäivälehtien sekä somekanavien keskusteluista voi huomata monen kuvittelevan rokottautumisen estävän taudin tarttumisen kokonaan. Todellisuudessa on kyse taudin tarttumisen vaikeutumisesta eli tartunnan todennäköisyyden pienenemisestä. Ihmiselimistössä on satunnaistavia tekijöitä myös immuunijärjestelmän toiminnassa, joten rokotuskaan ei takaa täydellistä suojaa. Rokotekattavuuden eli rokotettujen osuuden koko populaatiosta on oltava riittävän laaja, jotta myös ne, joiden ei ole mahdollista ottaa rokotetta esimerkiksi iän, perussairauden tai heikon terveydentilan vuoksi, olisivat turvassa taudilta. Monissa infektiotaudeissa laumaimmuniteettiin riittävä rokotettujen osuus on noin 95 % populaatiosta, mutta osuus vaihtelee taudin tarttuvuuden mukaan.

Laumaimmuniteettiin liittyviä todennäköisyyslaskennan ongelmia tarkastellaan tehtävässä 11 (sivu 20). Tehtävä mukailee Mark Chu-Carrollin blogikirjoituksessaan (http://www.goodmath.org/blog/2015/02/05/quick-vaccine-math/ ) esittämää havainnollistusta.

"Vauva rokotettavana äitiysneuvolassa vuonna 1961", Museoviraston kuvakokoelmat (Museiverkets Bildsamlingar). Kuvakokoelman lisenssi: CC BY-NC-ND 2.0. Lisenssikopio saatavilla osoitteessa: https://creativecommons.org/licenses/by-nc-nd/2.0/

TEORIATAUSTAA TEHTÄVIIN 11 JA 12:

# EHDOLLINEN TODENNÄKÖISYYS JA KERTOLASKUSÄÄNTÖ

Ehdolliseen todennäköisyyteen liittyvästä ongelmasta on kyse, kun halutaan laskea jonkin tapauksen *A* todennäköisyys silloin, kun tapaus *B* on jo sattunut. Ehdollista todennäköisyyttä merkitään *P(A|B)*, jonka voi lukea esimerkiksi seuraavasti: "todennäköisyys, että *A* tapahtuu, kun *B* on tapahtunut". Ehdollinen todennäköisyys *P(A|B)* voidaan määritellä kahdella tavalla: joko tapahtuman *A* todennäköisyytenä *supistetussa otosavaruudessa B* tai *alkuperäisen otosavaruuden Ω* todennäköisyyksien avulla.

**Määritelmä 1** Jos *P(B) > 0*, niin *ehdollinen todennäköisyys P(A|B)* on tapauksen *A* todennäköisyys otosavaruudessa *B*.

**Määritelmä 2** Jos *P(B) > 0*, niin *ehdollinen todennäköisyys P(A|B)* on

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Jälkimmäisen määrittelytavan avulla voidaan johtaa seuraavaksi esiteltävät ehdolliseen todennäköisyyteen liittyvät tulokset.

**Lause 1** (*Kertolaskusääntö*) Tapausten A ja B leikkauksen todennäköisyys toteuttaa kaavat

$$P(A \cap B) = P(A|B)P(B), \text{jos } P(B) > 0 \text{ ja}$$

$$P(A \cap B) = P(B|A)P(A), \text{jos } P(A) > 0.$$

Todistus (SYVENTÄVÄÄ TIETOA)

Leikkauksen todennäköisyys saadaan ratkaisemalla ehdollisen todennäköisyyden *P(A|B)* kaavasta leikkauksen P(A∩B) todennäköisyys. Vastaavasti leikkauksen P(A∩B) todennäköisyys saadaan myös ratkaistua ehdollisen todennäköisyyden *P(B|A)* kaavasta seuraavasti

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{jos } P(A) > 0$$

jolloin

$$P(A \cap B) = P(B|A)P(A). \blacksquare$$

Seuraavaksi esiteltävät Määritelmä 3 ja Lauseet 2 ja 3 eivät yleensä kuulu lukion pitkän matematiikan Tilastot ja todennäköisyys -opintojakson sisältöön. Ne ovat kuitenkin ehdollisen todennäköisyyden yleistyksiä, joiden tuntemus lisää aihepiirin soveltamisen mahdollisuuksia.

**Määritelmä 3** Tapaukset $A_1, \dots, A_n$ muodostavat otosavaruuden Ω osituksen, jos tapaukset ovat

- toisensa poissulkevat
- $\Omega = \cup_{i=1}^{n} A_i$ ja
- $P(A_i) > 0$ kaikilla i

**Lause 2** (*Kokonaistodennäköisyyslause*) Jos tapaukset $A_1, \ldots, A_n$ muodostavat otosavaruuden osituksen, niin

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\, P(A_i).$$

Todistus (SYVENTÄVÄÄ TIETOA)

Osituksen määritelmää (Määritelmä 3) ja summakaavaa apuna käyttäen saadaan:

$$\begin{aligned} P(B) &= P(B \cap \Omega) \\ &= \mathrm{P}\big(B \cap (\cup_{i=1}^{n} A_i)\big) \\ &= P(\cup_{i=1}^{n}(B \cap A_i)) \\ &= \textstyle\sum_{i=1}^{n} P(B \cap A_i) \\ &= \textstyle\sum_{i=1}^{n} P(B|A_i)\, P(A_i)\,. \blacksquare \end{aligned}$$

Lauseen 2 kokonaistodennäköisyyskaavalla kokonaistodennäköisyys *P(B)* saadaan koottua ehdollisista todennäköisyyksistä painottamalla kutakin itse ehdon todennäköisyydellä.

Seuraava keksijänsä mukaan nimetty, 1700-luvulla eläneen matemaatikon Thomas Bayesin kehittämä kaava on monesti hyödyllinen ehdollisia todennäköisyyksiä laskettaessa.

**Lause 3** (*Bayesin lause*) Jos tapaukset $A_1, \ldots, A_n$ muodostavat otosavaruuden osituksen ja $P(B) > 0$

$$P(A|B) = \frac{P(B|A_k)\, P(A_k)}{P(B)} = \frac{P(B|A_k)\, P(A_k)}{\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}\,, k = 1, \ldots, n.$$

Todistus (SYVENTÄVÄÄ TIETOA)

Ehdollisen todennäköisyyden määritelmästä (Määritelmä 2) ja Lauseesta 1 seuraa

$$P(A|B) = \frac{P(B \cap A_k)}{P(B)} = \frac{P(B|A_k)\, P(A_k)}{P(B)}.$$

Koska tapaukset muodostavat otosavaruuden osituksen, niin *kokonaistodennäköisyyslauseen* (Lause 2) mukaan *P(B)* voidaan koota ehdollisista todennäköisyyksistä $P(B|A_i)$ painottamalla kutakin ehdollista todennäköisyyttä ehdon todennäköisyydellä $P(A_i)$. Siis

$$P(A|B) = \frac{P(B|A_k)\, P(A_k)}{P(B)} = \frac{P(B|A_k)\, P(A_k)}{\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}\,. \blacksquare$$

# Tehtävä 11: Rokotteen rooli taudin leviämisen ehkäisyssä

Yksittäiset tautitartunnat puhkeavat populaatiossa tautiaalloksi tartunnan saaneiden määrän kasvaa. Jos jokainen sairastunut tartuttaa enemmän kuin yhden ihmisen, tauti alkaa levitä populaatiossa eksponentiaalisesti. Kuinka suuresta tautiaallosta populaatiossa on kyse, riippuu siitä, kuinka monta kukin sairastunut tartuttaa.

Oletetaan, että tauti X on erittäin helposti tarttuva ja jokaista sataa ei-immuunia ihmistä kohden 96 sairastuu kohdatessaan tautia sairastavan henkilön. Oletetaan myös, että tautiin X kehitetty rokote on 95 %:n varma eli 95 ihmistä 100:sta ovat täysin immuuneja taudin tarttumiselle.

a) Oletetaan, että ensimmäinen tautiin sairastunut henkilö on tekemisissä kahdenkymmenen terveen kanssa. Tutki, montako ihmistä yksi sairastunut keskimäärin altistaa taudin tartunnalle. Pohdi, pääseekö tauti leviämään populaatiossa epidemiaksi jokaisen populaation jäsenen otettua rokotteen.

b) Tutki a-kohdan oletuksin tilannetta, jossa 5 % populaatiosta on jättänyt rokotteen ottamatta.

c) Tutkitaan taudin X leviämistä rokottamattomien, ei-immuunien ihmisten n-henkisessä ketjussa sairastuneelta taudinkantajalta S terveelle, ketjun viimeiselle henkilölle $T_n$, kun S tapaa ensin henkilön $T_1$, sitten $T_1$ tapaa henkilön $T_2$ ja niin edelleen, kunnes jonon toiseksi viimeinen, n-1:s henkilö $T_{n-1}$ kohtaa viimeisen henkilön $T_n$. Oletetaan kohtaamisaikavälien olevan tarttumisen kannalta suotuisat. Vertaile 100 %-rokottautuneen yhteisön tilannetta täysin rokottautumattoman yhteisön tilanteeseen tutkimalla henkilön $T_n$ tartunnan saamisen todennäköisyyttä eri jonon pituuksilla n=5, n=20 ja yleisessä tapauksessa n.

## PSST!

Taudin yleisellä leviämiskyvyllä kuvataan sitä, kuinka monta uutta henkilöä yksi sairastunut keskimäärin tartuttaa ympäristössään. Esimerkiksi tuberkuloosipotilas tartuttaa keskimäärin 0,5 henkilöä ja tuhkarokkopotilas 18. Tätä leviämiskyvyn ilmoittamisessa käytettyä lukua kutsutaan $R_0$-luvuksi, joka on sitä suurempi, mitä helpommin tauti leviää. $R_0$-luvun ollessa pienempi kuin 1 tauti ei yleensä aiheuta epidemioita.

Käytännössä siihen, kuinka paljon uusia tapauksia kukin sairastunut tartuttaa ja kuinka tehokkaasti tauti leviää yhteisössä, vaikuttaa myös ihmisyhteisön tiiviys, ihmisten käyttäytyminen ja minkä pituisesta tartuttavavasta vaiheesta on kyse. Tietyissä yhteisöissä, kuten perheissä, päiväkodeissa ja puolustusvoimien kasarmeissa, leviävät herkästi sellaisetkin taudit, joiden leviämiskyky muuten on vähäinen esim. kausi-influenssat.

**Lisää aiheesta:**

Infektioiden tartunta, taudin synty ja leviäminen, Duodecim.

# TEHTÄVÄ 12: TAUTITESTIN LUOTETTAVUUSTARKASTELU

Moni virus, kuten tässä tehtävässä tutkittava HI-virus, voidaan havaita pääosin terveyskeskuksissa ja sairaaloissa tehtävillä testeillä. Tällaisia ovat esimerkiksi verinäytteestä tehtävät vasta-aine- ja antigeenitestit.

Tämä tehtävä perustuu Richard Isaacin kirjassaan *The Pleasures of Probability* (s. 31-34) esittämään esimerkkiin HIV-testin luotettavuuden tutkimisesta. Esimerkissä testi on viruksen havaitsemiseksi sikäli hyvä, että jos henkilöllä on virus, sen havaitsemisen todennäköisyys verikokeella on korkea; suuressa ryhmässä, jonka tiedetään koostuvan vain tartunnan saaneista, testi antaa oikean, positiivisen tuloksen 95 prosentilla ryhmästä. Tässä yhteydessä puhutaan testin *sensitiivisyydestä*, joka edeltävässä tapauksessa on 0.95. Merkitään todennäköisyyttä, että testi on positiivinen henkilön ollessa taudin kantaja, seuraavasti

P(testi positiivinen | henkilöllä tauti) = 0.95.

Testi siis vaikuttaa intuitiivisesti hyvin luotettavalta. Seuraavaksi keskitytään kuitenkin tutkimaan mahdollisia virhetuloksia, jotka ovat

P(testi positiivinen | henkilöllä ei ole virusta) eli väärä positiivinen tulos ja

P(testi negatiivinen | henkilöllä on virus) eli väärä negatiivinen tulos.

Oletetaan, että HI-virusta kantaa väestössä kolme ihmistä tuhannesta. Jos ihmisellä on HI-virus, tietty testi ilmoittaa 95 % varmuudella, että ihmisellä on kyseinen virus. Jos taas ihmisellä ei ole virusta, testi ilmoittaa 95 % varmuudella, että ihmisellä ei ole virusta. Oletetaan myös sekä väärien positiivisten että väärien negatiivisten testituloksien osuuden olevan kummankin 5 %. Millä todennäköisyydellä

a) testitulos on positiivinen,

b) testitulos on negatiivinen,

c) ihmisellä on virus, jos tiedetään testituloksen olevan positiivinen,

d) ihmisellä ei ole virusta, jos tiedetään testituloksen olevan positiivinen? Pohdi testin luotettavuutta.

e) Tarkastellaan tilannetta Etelä-Karjalassa, jossa oli vuonna 2017 Suomen korkein HIV-ilmaantuvuus (6,1/100 000). Oletetaan muutoin tehtävänannon tiedot testistä oikeiksi. Millä todennäköisyydellä positiivisen testituloksen saaneella henkilöllä ei ole virusta?

## PSST!

HIV on krooninen autoimmunologinen sairaus, jonka aiheuttaa HI-virus. Viruksen aiheuttamille haitoille ei toistaiseksi ole löydetty parannuskeinoa, mutta lääkehoidoilla pystytään pitämään viruksen aiheuttamat vauriot hallinnassa. Taudin varhainen toteaminen on tärkeää, jotta hoito toteutuu parhaalla mahdollisella tavalla eikä infektoituneen elimistön puolustuskyky laske viruksen tuhotessa valkosoluja, jotka suojaavat elimistöä taudinaiheuttajilta. Hoitamattomana HIV-infektio voi edetä pahimmillaan henkeä uhkaavaan AIDS-vaiheeseen, jossa elimistön puolustusjärjestelmä heikentyy vakavasti. HI-viruksen aiheuttama tartunta voidaan havaita verinäytteestä tehtävällä vasta-ainetestillä.

**Lisää aiheesta:**

HIV sairautena, Hivpoint (Hiv-säätiö).

# VIHJEET JA VINKIT

JOHDANTOTEHTÄVÄ:

## TERVEYTEEN LIITTYVIEN ILMIÖIDEN RIIPPUMATTOMUUS

Harjoituksen tarkoituksena on huomata, että terveyteen liittyvät ilmiöt ovat usein monisyisiä ja voivat liittyä toisiinsa suoraan tai välillisesti. Terveysaiheisia tai muutoin ympäröivän maailman ilmiöihin linkittyviä todennäköisyyslaskennan tehtäviä ratkoessa on siis paikallaan pysähtyä hetkeksi pohtimaan aihetta itseään. Näin pystyy mahdollisesti välttämään tehtävän ratkaisun miinakohdat sekä lopuksi pohtimaan perustellusti saadun tuloksen järkevyyttä ja merkitystä.

## TEHTÄVÄ 1: LUKIOLAISTEN TERVEYSTOTTUMUKSIA

”Neljäsosa opiskelijoista sekä jätti koululounaan syömättä että liikkui alle tunnin päivän aikana”. *Neljäsosa* eli *yksi neljäsosa* voidaan tulkita todennäköisyydeksi $\frac{1}{4} = \frac{25}{100} = 0{,}25$.

Tilannetta kannattaa hahmotella kuvan (Venn-diagrammi) avulla.

Tehtävässä halutaan ratkaista, kuinka todennäköisesti satunnaisesti valittu opiskelija kuuluu joukkoon, joka koostuu niistä opiskelijoista, jotka sekä syövät koululounaan että liikkuvat vähintään tunnin joka päivä. Tähän joukkoon kuulumisen todennäköisyys voidaan ratkaista sen vastatapahtuman todennäköisyyden avulla. Tässä vastatapahtuma on tapaus, jossa opiskelija jättää koululounaan syömättä tai liikkuu alle tunnin päivässä. Kun vastatapahtuman todennäköisyys vähennetään koko otosavaruudesta, saadaan selville kysytyn tapahtuman todennäköisyys.

## TEHTÄVÄ 2: VERENLUOVUTUS

a) Promillet ovat tuhannesosia ($\frac{1}{1000}$) ja prosentit sadasosia ($\frac{1}{100}$). Yksi promille (merkitään 1 ‰) on siis prosentin kymmenesosa eli 1 ‰ = 0,1 %.

Todennäköisyys sille, että suomalainen kuuluu veriryhmään B+ saadaan suoraan ympyräkaaviosta.

Tapahtuman, jossa ainakin yksi suomalaisryhmästä kuuluu veriryhmään B+, todennäköisyyttä on luontevaa tarkastella sen vastatapahtuman todennäköisyyden avulla. Tapahtuman ”**ainakin yksi** suomalainen **kuuluu** veriryhmään B+” vastatapahtuma on tapahtuma, jossa **yksikään henkilö ei kuulu** veriryhmään B+. Kun vastatapahtuman todennäköisyys vähennetään koko otosavaruudesta, saadaan selville kysytyn tapahtuman todennäköisyys.

On hyvä huomata, että satunnaisesti valitussa suomalaisryhmässä henkilöiden kuuluminen tiettyyn veriryhmään **ei riipu** ryhmän muiden henkilöiden veriryhmistä.

b) Verenluovutuksessa potilaalle annetaan hänen veriryhmänsä kanssa yhteensopivaa verta tehtävänannon taulukon mukaisesti. Ympyräkaaviosta saadaan ratkaistua, paljonko sopivia luovuttajia AB- -ryhmäläiselle on väestössä **yhteensä** prosentuaalisesti.

Ratkaisussa voidaan a-kohdan tapaan hyödyntää **vastatapahtuma**n todennäköisyyttä.

# Tehtävä 3: Tartuntalähteet

Eri tartuntalähteet ovat toisensa **poissulkevia**, sillä tartunta on saatu tietyltä tartuttajalta tietyllä hetkellä. Esimerkiksi jos tartunta on saatu perhepiiristä, niin sitä ei ole voitu saada muista tartuntalähteistä. Lisäksi, koska viisi sairastunutta henkilöä on valittu satunnaisesti, voidaan näiden henkilöiden tartunnat lähteineen olettaa toisistaan **riippumattomiksi**.

P (”tartunnan saanut henkilö on saanut tartunnan jostain tartuntalähteestä”) = 1, sillä jokaisella tautitartunnalla on aina jokin lähde. Täten kuvan histogrammi on todennäköisyysjakauma ja eri tartuntalähteiden todennäköisyyksien summa on 1 eli 100%. Näin ollen kunkin tartuntalähteen todennäköisyys saadaan suoraan histogrammista.

## Huomautus!

Yle Uutisten grafiikassa eri tartuntalähteiden todennäköisyyksien pyöristys on tehty prosentin tarkkuudella, jonka vuoksi saadaan

$$0{,}6 + 0{,}15 + 0{,}10 + 0{,}08 + 0{,}08 = 1{,}01\ ,$$

mikä on ristiriidassa todennäköisyyden perusominaisuuksien kanssa; todennäköisyys kuuluu aina välille $[0{,}1]$. Tässä tehtävässä voidaan kuitenkin olettaa kyseessä olevan todennäköisyysjakauma. Edellä mainittu virhelähde on hyvä ottaa huomioon vastauksen mielekkyyttä ja pyöristystarkkuutta pohdittaessa.

# Tehtävä 4: Seksuaalivähemmistöt

a) Tilannetta voidaan ajatella toistokokeena, jossa täsmälleen samaa koetta toistetaan 30 kertaa ja seurataan onnistuneiden kokeiden lukumäärää. Tässä 30 henkilöä valitaan satunnaisesti ja tapahtuman ”henkilö kuuluu seksuaalivähemmistöön” todennäköisyys on jokaisella valinnalla sama ja riippumaton aiempien valintojen tuloksista. Kyseinen tapahtuma joko tapahtuu tai ei tapahdu. Näin ollen seksuaalivähemmistöön kuuluvien henkilöiden lukumäärä noudattaa siis binomijakaumaa.

b) Tapahtuma, jossa ainakin 28 henkilöä 30 henkilöstä kuuluu seksuaalivähemmistöihin, voidaan ajatella tapahtumana ”seksuaalivähemmistöihin kuuluu joko 28 tai 29 tai 30 henkilöä 30:stä”. Tapahtumat ”seksuaalivähemmistöihin kuuluu 28 henkilöä”, ”seksuaalivähemmistöihin kuuluu 29 henkilöä” ja ”seksuaalivähemmistöihin kuuluu 30 henkilöä” taas ovat toisensa poissulkevia eli erillisiä tapahtumia.

c) Tässä tapahtuma ”enintään yksi 30:stä henkilöstä kuuluu seksuaalivähemmistöön” voidaan ajatella tapahtumana ”seksuaalivähemmistöihin kuuluu yksi tai ei yksikään 30:stä henkilöstä”.

d) Tapahtuman, jossa ainakin yksi henkilö 30:stä kuuluu seksuaalivähemmistöön, vastatapahtuma on tapahtuma, jossa yksikään henkilö ei kuulu seksuaalivähemmistöön. Kun vastatapahtuman todennäköisyys vähennetään koko otosavaruudesta, saadaan selville kysytyn tapahtuman todennäköisyys.

# TEHTÄVÄ 5: KANSANTAUDIT

Tehtävänannosta löydetään todennäköisyydet tapahtumille ”henkilö sairastuu syöpään”, ”henkilö sairastuu diabetekseen”, ”syöpään sairastunut henkilö paranee syövästä” ja ”painonhallintaohjelmaan osallistunut diabeetikko pääsee remissioon”. Lisäksi **vastatapahtuman todennäköisyyden** avulla saadaan muotoiltua todennäköisyydet tapahtumille ”henkilö ei sairastu syöpään” ja ”henkilö ei sairastu diabetekseen”.

a-b) Koska tehtävässä oletetaan, että syöpään ja diabetekseen sairastuminen ovat toisistaan **riippumattomat tapahtumat**, saadaan a- ja b-kohdat laskettua **kertolaskusääntöä** käyttämällä.

c) Tapahtuma ”henkilö sairastuu ainakin toiseen kyseisistä sairauksista” tarkoittaa, että henkilö sairastuu syöpään tai diabetekseen tai molempiin. Tämän tapahtuman todennäköisyys saadaan ratkaistua mukavasti hyödyntämällä sen vastatapahtuman todennäköisyyttä.

d) Tapahtuma ”henkilö sairastuu enintään toiseen kyseisistä sairauksista” tarkoittaa sitä, että henkilö sairastuu syöpään, mutta ei tyypin 2 diabetekseen **tai** hän sairastuu tyypin 2 diabetekseen, mutta ei syöpään **tai** hän ei sairastu kumpaankaan”.

e) Tapahtuma ”henkilö sairastuu vain toiseen kyseisistä sairauksista”, tarkoittaa sitä, että hän sairastuu syöpään, mutta ei tyypin 2 diabetekseen **tai** hän sairastuu tyypin 2 diabetekseen, mutta ei syöpään.

f-g) Sairastuminen ja siitä paraneminen eivät ole toisistaan riippumattomia tapahtumia, sillä paraneminen voi tapahtua vain silloin, kun henkilö on ensin sairastunut. Tapahtuma ”henkilö paranee” tarkoittaa siis itse asiassa tapahtumaa ”henkilö paranee, kun hän on ensin sairastunut”. Näin ollen todennäköisyys tapahtumalle ”henkilö ensin sairastuu ja sitten paranee” saadaan laskettua yleistä kertolaskusääntöä käyttäen.

# TEHTÄVÄ 7: KESTÄVYYSTREENI

Merkitään sykettä satunnaismuuttujalla X. Satunnaismuuttuja X noudattaa normaalijakaumaa parametrein 130 (odotusarvo, joka on normaalijakaumalle sama kuin keskiarvo) ja 8 (keskihajonta). Koska odotusarvo poikkeaa nollasta ja keskihajonta on erisuuri kuin 1, tulee suorittaa normittaminen. Satunnaismuuttujan arvoa $x$ vastaava *normitettu arvo z* on

$$z = \frac{x - \mu}{\sigma},$$

jossa μ on keskiarvo ja $\sigma$ on keskihajonta.

Normaalijakauma on symmetrinen jakauma, jossa ns. todennäköisyysmassa on keskittynyt odotusarvon läheisyyteen. Todennäköisyyden laskemisessa käytetään apuna normaalijakauman kertymäfunktion taulukkoa, joka löytyy MAOL-taulukkokirjasta.

## Tehtävä 8: Ensiaputaidot

a) Jokainen viidestä henkilöstä joko on tai ei ole ensiaputaitoinen. Jokaisen henkilön valitseminen viiden joukosta on itsenäinen, muiden henkilöiden valinnoista riippumaton tapahtuma. Näin ollen tilannetta voidaan ajatella toistokokeena, jossa valintatilanteita eli toistoja on viisi. Ensiaputaitoisen henkilön löytymisen todennäköisyys on jokaisen viiden henkilön kohdalla sama 36%. Todennäköisyys, että valittu henkilö ei ole ensiaputaitoinen saadaan laskettua vastatapahtuman avulla.

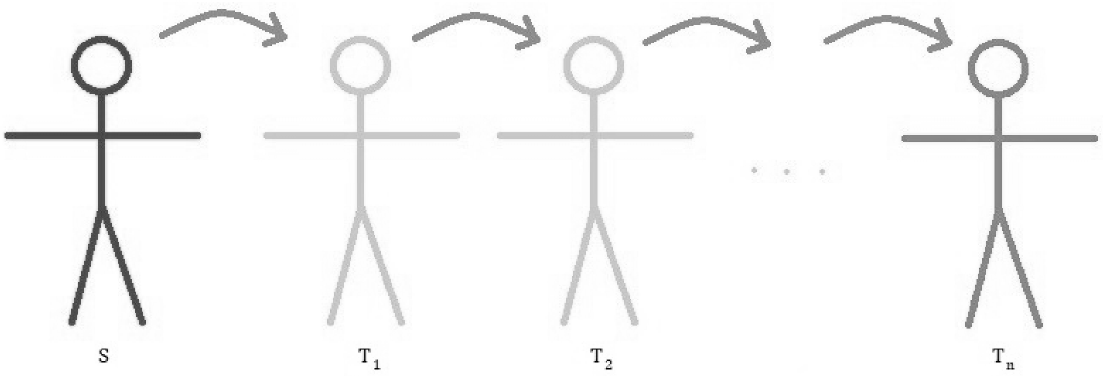
b) Ensiaputaitoisia henkilöitä voi olla viiden henkilön joukossa 0, 1, 2, 3, 4 tai 5. Kaikkien näiden tapausten todennäköisyydet on selvitettävä, jotta jakauma saadaan selvitettyä. Esimerkiksi jos joukossa on yksi ensiaputaitoinen henkilö, niin siinä on neljä ensiaputaidotonta henkilöä.

## Tehtävä 9: Koulukiusaaminen

c) Oletetaan, että 8. tai 9. luokan opiskelija on joutunut kiusaamisen uhriksi lukukauden aikana, jos hän on vastannut kysymykseen "Kuinka usein sinua on kiusattu koulussa tämän lukukauden aikana?" joko "useita kertoja viikossa", "noin kerran viikossa" tai "harvemmin".

## Tehtävä 11: Rokotteen rooli taudin leviämisen ehkäisyssä

a) Tauti voi tarttua vain rokotuksen jäljiltä ei-immuuneihin henkilöihin.
b) Tauti voi tarttua vain kahdenlaisiin ihmisiin: rokottamattomiin ja sellaisiin rokotettuihin, jotka eivät ole saaneet riittävää immuniteettia rokotteesta.
c) Jotta jonon viimeinen jäsen saisi taudin X, on taudin pitänyt tarttua kaikkiin häntä edeltäviin jäseniin.



# Tehtävä 12: Tautitestin luotettavuustarkastelu

a) Henkilön testitulos voi olla positiivinen, kun hänellä on HI-virus, mutta on myös mahdollista, että testitulos on positiivinen, kun henkilöllä ei ole virusta.

Tapaukset "henkilöllä on virus" ja "henkilöllä ei ole virusta" muodostavat otosavaruuden osituksen, sillä ne ovat toisensa poissulkevat, niiden summa muodostaa koko otosavaruuden $\Omega$ (nimittäin henkilöllä joko on tai ei ole virusta) ja kummankin tapauksen todennäköisyys on suurempi kuin 0 (annettu tehtävänannossa).

b) Henkilön testitulos voi olla negatiivinen, kun hänellä ei ole HI-virusta, mutta on myös mahdollista, että testitulos on negatiivinen, kun henkilöllä on virus.

Tapaukset "henkilöllä on virus" ja "henkilöllä ei ole virusta" muodostavat otosavaruuden osituksen ja ovat paitsi toisensa poissulkevat, niin myös toistensa vastatapahtumat.

c) Koska tapaukset "henkilöllä on virus" ja "henkilöllä ei ole virusta" muodostavat otosavaruuden osituksen ja tapahtuman "henkilö saa positiivisen testituloksen" todennäköisyys on suurempi kuin 0, niin saadaan ehdollisen tapahtuman "henkilöllä on virus, kun tiedetään testituloksen olevan positiivinen" todennäköisyys ratkaistua ehdollisen todennäköisyyden määritelmän ja kertolaskusäännön avulla.

d) Vastaavaan tapaan kuin c-kohdassa.

e) Vastaavaan tapaan kuin c-kohdassa.

Lukion matematiikan pitkä oppimäärä

**Tilastot ja todennäköisyys**

# TERVE!

Terveysaiheinen tehtäväpaketti opiskelijalle

OPETTAJAN MATERIAALI

Annina Ruokonen

# Sisällys

# KÄYTTÄJÄLLE

Todennäköisyyslaskenta on suhteellisen nuori, noin 400-vuotias, matematiikan osa-alue, joka perinteisesti on keskittynyt erilaisten rahapelien voiton todennäköisyyden laskemiseen. Myös useiden lukion todennäköisyyslaskennan oppikirjojen tehtävät ovat keskittyneet pelisovelluksiin, kuten noppa- ja korttipeleihin sekä erilaisiin arvontoihin. Tehtävät ovat monesti melko irrallaan todellisen maailman sovelluskohteista ja aiheiltaan opiskelijalle etäisiä. Tämä on sääli, sillä todennäköisyyslaskennan keinot taipuvat useiden arjen, luonnon ja yhteiskunnan ilmiöiden ja tapahtumien mallintamiseen, jotka kiinnostavat nuoria ja saattavat sopia paremmin heidän kokemusmaailmaansa. Mielenkiintoinen tehtävä voi innostaa opiskelijaa matemaattisen ongelmanratkaisuun.

Tämän materiaalin tehtävät liittyvät ajankohtaisiin arkiympäristön ja yhteiskunnan terveysilmiöihin, kuten treenaamiseen, ensiaputaitoihin, rokotteisiin, punkkien levittämiin tauteihin ja väestöryhmien välisiin terveyseroihin. Kunkin tehtävän yhteydessä kerrotaan myös hieman tehtävän aiheesta ja annetaan vinkkejä asiallisen lisätiedon löytämiseen. Tehtävien tavoitteena on, että opiskelija tutustuu erilaisten terveysilmiöiden syihin ja seurauksiin, ymmärtää ja kehittää niihin liittyvää matemaattista ajattelua sekä oppii analysoimaan saatujen tulosten mielekkyyttä ja luotettavuutta. Tarkoituksena on harjoitella pohtimaan saatujen tulosten merkitystä terveydelle yksilö-, yhteisö- ja yhteiskuntatasolla.

Tämä materiaali on tarkoitettu ensisijaisesti lukion pitkän matematiikan todennäköisyyslaskennan opintojaksolle (MAA8 Tilastot ja todennäköisyys, LOPS 2021), mutta osa tehtävistä* soveltuu myös matematiikan lyhyttä oppimäärää opiskeleville (MAB5 Tilastot ja todennäköisyys ja MAB9 Tilastolliset ja todennäköisyysjakaumat, LOPS2021).

Tehtävissä on käytetty seuraavia symboleja, joiden merkitys on tässä selitetty:

Tökkääkö ratkaiseminen? Vihjeitä on saatavilla.

Laskinohjelmistoja saa käyttää tehtävän ratkaisemiseksi.

Laskinohjelmistoja ei saa käyttää tehtävän ratkaisemiseksi. Funktiolaskin, kuten SpeedCrunch ja KCalc ovat käytössä.

Opiskelijan materiaalista löytyvät sanalliset vihjeet tehtävien ratkaisemiseksi. Tässä materiaalissa on esitetty jokaiselle tehtävälle eräs malliratkaisu.

Innostavia hetkiä tehtävien tutkailuun!

Annina Ruokonen

Tekijä

---

* MAB5: 1, 3, (5), 6, 9 ja 10 & MAB9: 2, 3, 4, 7 ja 8

# Tehtävät ja ratkaisut

JOHDANTOTEHTÄVÄ:

## Terveyteen liittyvien ilmiöiden riippumattomuus

Todennäköisyyslaskennassa kahden tapahtuman riippumattomuus tarkoittaa, että jonkin tapahtuman todennäköisyys ei riipu toisen tapahtuman toteutumisesta ja toisin päin. Toisinaan tapahtumien riippumattomuus on ilmeistä tai asiayhteydestä helposti ymmärrettävissä. Näin on esimerkiksi silloin, kun heitetään kahta noppaa: toisen nopan silmäluku ei riipu siitä, mitä ensimmäisellä nopalla saatiin silmäluvuksi. Terveyteen liittyvät tapahtumat ovat kuitenkin usein monisyisiä, minkä vuoksi tapahtumien riippumattomuuden tarkastelu on tärkeää ja tehtävissä on syytä mainita, voidaanko tapahtumat olettaa riippumattomiksi vai ei.

Pohdi itsenäisesti tai keskustele parin kanssa, millä tavoin seuraavat asiat voivat liittyä toisiinsa ja ovatko ilmiöt toisistaan riippumattomia:

- ilmastonmuutos ja tartuntatautien leviäminen
- sosioekonominen asema ja terveydentila
- lapsityövoima ja tupakkatuotteet
- rokotteet ja kuolleisuus
- liikunta ja stressi
- ylipaino ja tyypin 2 diabetes
- ravitsemus ja uni
- mieliala ja ystävyyssuhteet
- tupakointi ja masennus
- aktiivisuusmittari ja unen laatu
- ilmaston lämpeneminen ja lonkkamurtumat
- pyöräilykypärän käyttäminen ja turvavyön käyttäminen

### Ratkaisun idea:

Harjoituksen tarkoituksena on huomata, että terveyteen liittyvät ilmiöt ovat usein monisyisiä ja voivat liittyä toisiinsa suoraan tai välillisesti. Terveysaiheisia tai muutoin ympäröivän maailman ilmiöihin linkittyviä todennäköisyyslaskennan tehtäviä ratkoessa on siis paikallaan pysähtyä hetkeksi pohtimaan aihetta itseään. Näin pystyy mahdollisesti välttämään tehtävän ratkaisun miinakohdat sekä lopuksi pohtimaan perustellusti saadun tuloksen järkevyyttä ja merkitystä.

# Tehtävä 1: Lukiolaisen terveystottumuksia

Vuonna 2019 turkulaisista 1. ja 2. vuosikurssin lukio-opiskelijoista 33,8 % ei syönyt koululounasta päivittäin ja 85,4 % liikkui päivässä alle tunnin (Kouluterveyskysely 2019). Oletetaan, että neljäsosa opiskelijoista sekä jätti koululounaan väliin että liikkui alle tunnin päivän aikana. Millä todennäköisyydellä sattumanvaraisesti valittu turkulaisopiskelija syö koululounaan ja liikkuu vähintään tunnin joka päivä?

## PSST!

Kouluruokailu on niin opiskelijoiden kuin muun koulun henkilökunnan terveyttä ja hyvinvointia tukeva tekijä. Koululounas on suunniteltu monipuoliseksi ateriaksi, joka kokonaisuudessaan nautittuna on ravitsemuksellisesti täysipainoinen ja terveellinen. Kylläisenä ihminen jaksaa keskittyä paremmin opiskeluun ja sietää arjen kiireisyyttä. Kouluruokailu myös lisää yhteisiä ruokailuhetkiä ja ruoan arvostusta. Ruokailu antaa mahdollisuuden pieneen hengähdyshetkeen hektisen koulupäivän keskellä ja sillä on samalla sosiaalinen merkityksensä koulun arjessa.

Sosiaali- ja terveysministeriön asiantuntijakeskus UKK-instituutin laatimien, laajaan tutkimustietoon perustuvien liikkumissuositusten tarkoituksena on tiivistää terveyden kannalta tarvittavan viikoittaisen liikkumisen määrä sekä antaa esimerkkejä eri liikkumisen muodoista kaikenikäisille kansalaisille. Instituutin suosituksen mukaan 7-17-vuotiaiden lasten ja nuorten tulisi liikkua reippaasti ja rasittavasti tunti päivässä. Tunti voi kertyä useista liikuntahetkistä päivän aikana ja tärkeää olisi välttää pitkäkestoista paikallaanoloa, kuten istumista.

**Lisää aiheesta:**

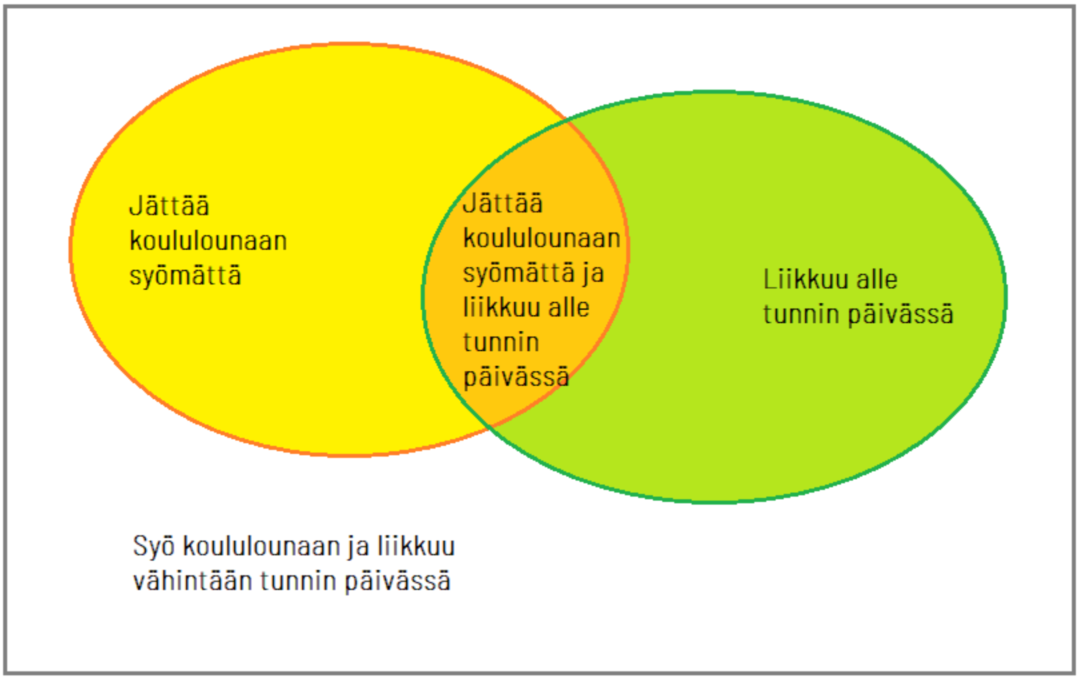
Kouluruokailun merkityksestä, Opetushallitus.

Lasten ja nuorten liikkumissuositus, UKK-instituutti.

Terveelliset välipalat voivat edistää jaksamista, Valtioneuvoston selvitys- ja tutkimustoimikunta.

# Eräs malliratkaisu

Hahmotellaan tilannetta Venn-diagrammin avulla:



Tehtävässä halutaan siis ratkaista, kuinka todennäköisesti satunnaisesti valittu opiskelija kuuluu joukkoon, joka koostuu niistä opiskelijoista, jotka sekä syövät koululounaan että liikkuvat vähintään tunnin joka päivä. Merkitään tätä joukkoa/tapahtumaa A:lla:

A = " opiskelija syö koululounaan ja liikkuu vähintään tunnin joka päivä".

Ratkaistaan tapahtuman A todennäköisyys sen vastatapahtuman avulla. Nimittäin $P(A) = 1 - P(B)$, jossa tapahtuma B on tapahtuman A vastatapahtuma. Venn-diagrammi auttaa hahmottamaan tapahtuman A vastatapahtuman B:

B = "opiskelija jättää koululounaan syömättä tai liikkuu alle tunnin päivässä".

Näin ollen tapahtuman A todennäköisyys P(A) saadaan seuraavasti

$$P(A) = 1 - P(B),$$

josta $P(B)$:n todennäköisyys saadaan mukaanlukemis-poissulkemis-periaatteen eli yleisen yhteenlaskusäännön mukaisesti:

$$P(A) = 1 - (0.338 + 0.854 - 0.25) = 1 - 0.942 = 0{,}058\ (5{,}8\%).$$

Siis todennäköisyys sille, että satunnaisesti valittu turkulaislukiolainen syö koululounaan ja liikkuu vähintään tunnin päivittäin on 0,058 eli 5,8 %.

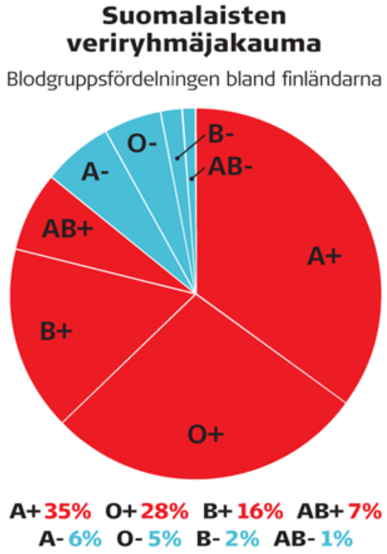
# Tehtävä 2: Verenluovutus

Veriryhmät ovat tapa ryhmitellä verityyppejä ominaisuuksineen. Erilaisia veriryhmäluokitteluja on maailmassa yli 30, mutta yleisimmin käytössä ovat ABO- ja reesusveriryhmäjärjestelmät. Ihmisellä veri kuuluu yhteen neljästä ABO-veriryhmästä (A, B, AB ja O) ja on luokiteltavissa edelleen reesusnegatiivisiin ja reesuspositiivisiin: A-, A+, B-, B+, AB-, AB+, O- tai O+. Verenluovutuksessa potilaalle annetaan hänen veriryhmänsä kanssa yhteensopivaa verta alla olevan taulukon (Veripalvelu, SPR) mukaisesti. Verentarve noudattaa keskimäärin suomalaisten veriryhmäjakaumaa.

**Luovuttajan veriryhmä**

**Potilaan veriryhmä**

| | O- | O+ | B- | B+ | A- | A+ | AB- | AB+ |
|---|---|---|---|---|---|---|---|---|
| AB+ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| AB- | ✓ | | ✓ | | ✓ | | ✓ | |
| A+ | ✓ | ✓ | | | ✓ | ✓ | | |
| A- | ✓ | | | | ✓ | | | |
| B+ | ✓ | ✓ | ✓ | ✓ | | | | |
| B- | ✓ | | ✓ | | | | | |
| O+ | ✓ | ✓ | | | | | | |
| O- | ✓ | | | | | | | |



Epäsopivan veren saaminen voi johtaa potilaan omien punasolujen rikkoutumiseen, millä voi olla vakavia seurauksia.

a) Kuinka suuri ryhmän suomalaisia tulee olla, jotta todennäköisyys, että ainakin yksi heistä kuuluu veriryhmään B+, ylittää 995 ‰?

b) Kuinka suuri satunnaisesti valituista henkilöistä koostuvan ryhmän tulee olla, jotta veriryhmään AB- kuuluvalle henkilölle löytyy sopiva luovuttaja yli 950 ‰:n todennäköisyydellä?

**Lisää aiheesta:**

Veriryhmät, niiden periytyvyys ja verenluovutus, Veripalvelu (Suomen Punainen Risti, SPR).

Mitä verenluovutuksessa tapahtuu? Veripalvelu.

Voittoa tavoittelematon verenluovutus Euroopassa, EBA.

PSST!

Tiesitkö, että verenluovutus on helppo tapa auttaa? Se vie vain noin 10 minuuttia. Luovutetusta verestä tehtyjä verivalmisteita tarvitaan muun muassa leikkauspotilaiden, onnettomuusuhrien, syöpää sairastavien ja keskoslasten hoidossa. Suomessa verenluovutustoiminta on perustunut aina vapaaehtoisuuteen ja maksuttomuuteen. Luovuttaja auttaa potilasta ilman korvausta, mikä on tärkeä asia verivalmisteen turvallisuuden kannalta. Joissain maissa verenluovutuksesta maksetaan korvaus, mikä voi aiheuttaa vakavia, jopa kuolemaan johtavia terveys- ja turvallisuusriskejä luovuttajan tavoitellessa palkkiota.

# Eräs malliratkaisu

a) Suomalaisten veriryhmäjakaumasta (ympyräkaavio) nähdään, että B+ veriryhmään kuuluu 16 % suomalaisista. Siis tapahtuman $A$ = "henkilö kuuluu veriryhmään B+" todennäköisyys $P(A) = 0{,}16$. Tapahtuman $A$ vastatapahtuman $A^C$ = "henkilö ei kuulu veriryhmään B+" todennäköisyys $P(A^C) = 1 - 0{,}16 = 0{,}84$.

Olkoon suomalaisryhmän koko $n$ henkilöä. Tapahtuman $B$ = "ainakin yksi henkilö ryhmästä kuuluu veriryhmään B+" todennäköisyys saadaan laskettua sen vastatapahtuman $B^C$ = "yksikään henkilö ryhmästä ei kuulu veriryhmään B+" todennäköisyyden avulla. Koska kunkin ryhmän henkilön kuuluminen tiettyyn veriryhmään ei riipu toinen toisistaan (erilliset tapahtumat), saadaan:

$$P(B) = 1 - P(B^C)$$

$$= 1 - \underbrace{P(A^C) \cdot P(A^C) \cdot \ldots \cdot P(A^C)}_{n\ kpl}$$

$$= 1 - 0{,}84^n.$$

Etsitään sitten pienin luku $n$, jolle todennäköisyys $P(B^C) > 0{,}995$ eli

$$1 - 0{,}84^n > 0{,}995$$

$$0{,}84^n < 0{,}005.$$

Ratkaistaan yhtälö $0{,}84^n = 0{,}005$ logaritmin avulla:

$$0{,}84^n = 0{,}005$$

$$n = log_{0{,}84} 0{,}005$$

$$n = 30{,}3883\ \ldots$$

Näin ollen 30 henkilöä ei aivan riitä ryhmäkooksi, vaan tarvitaan vähintään 31 henkilöä, jotta todennäköisyys sille, että ryhmässä on ainakin yksi veriryhmään B+ kuuluva henkilö, on yli 995 promillea. Mitä enemmän ryhmässä on henkilöitä, sitä todennäköisempää on, että heistä ainakin yksi kuuluu veriryhmään B+.

b) Verenluovutuksessa potilaalle annetaan hänen veriryhmänsä kanssa yhteensopivaa verta. Taulukon mukaan veriryhmään AB- kuuluvalle henkilölle verta voi luovuttaa veriryhmiin A-, B-, 0- ja AB- kuuluvat henkilöt. Näin ollen kaikki reesusnegatiiviset henkilöt kelpaavat luovuttajiksi. Heitä on veriryhmäjakauman mukaan väestöstä yhteensä 6 % + 5 % + 2 % + 1 % = 14 %. Tapahtuman $A$ = "henkilö on sopiva luovuttaja" todennäköisyys on siis $P(A) = 0{,}14$. Tapahtuman $A$ vastatapahtuman $A^C$ = "henkilö ei ole sopiva luovuttaja" todennäköisyys $P(A^C) = 1 - 0{,}14 = 0{,}86$.

Olkoon suomalaisryhmän koko $n$ henkilöä. Tapahtuman $B$ = "ainakin yksi henkilö ryhmästä on sopiva luovuttaja" todennäköisyys saadaan laskettua sen vastatapahtuman $B^C$ = "yksikään henkilö ryhmästä ei ole sopiva luovuttaja" todennäköisyyden avulla. Koska kunkin henkilön kuuluminen tiettyyn veriryhmään ei riipu toinen toisistaan (erilliset tapahtumat), saadaan:

$$P(B) = 1 - P(B^C)$$

$$= 1 - \underbrace{P(A^C) \cdot P(A^C) \cdot \ldots \cdot P(A^C)}_{n\ kpl}$$

$$= 1 - 0{,}86^n.$$

Etsitään sitten pienin luku $n$, jolle todennäköisyys $P(B^C) > 0{,}950$ eli

$$1 - 0{,}86^n > 0{,}950$$

$$0{,}86^n < 0{,}050.$$

Ratkaistaan yhtälö $0{,}84^n = 0{,}050$ logaritmin avulla:

$$0{,}86^n = 0{,}050$$

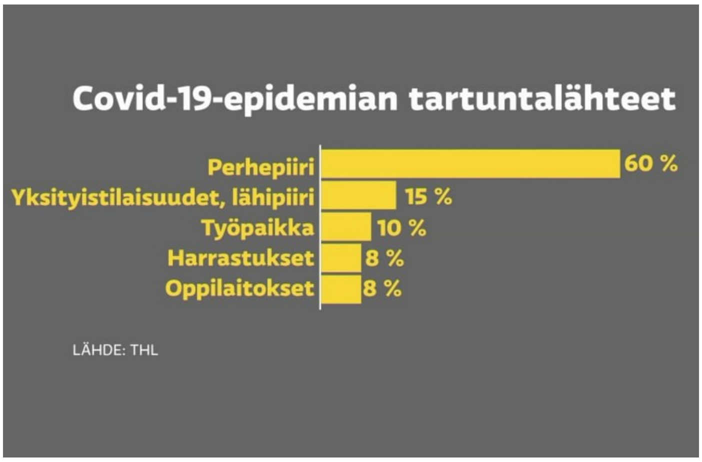
$$n = log_{0,86} 0{,}050$$

$$n = 19{,}8625 \ldots$$

Näin ollen, kun ryhmässä on vähintään 20 henkilöä, todennäköisyys sille, että ryhmässä on ainakin yksi sopiva luovuttaja veriryhmään AB- kuuluvalle henkilölle, on yli 950 promillea. Mitä enemmän ryhmässä on henkilöitä, sitä todennäköisempää on, että heistä löytyy ainakin yksi sopiva luovuttaja.

# Tehtävä 3: Tartuntalähteet

Yle Uutisten 5.11.2020 mukaan Covid-19-viruksen tartunnan jäljittäminen on aiempaan verrattuna helpottunut ja noin 60 % tartuntalähteistä selviää. Terveyden ja hyvinvoinnin laitoksen (THL) tietojen mukaan selvinneiden tartuntojen lähteet ovat seuraavat:



Covid-19, kuten moni virus- ja bakteeritauti, leviää hengitystie-eritteiden kautta pisaratartuntana suun, nenän ja silmien limakalvoilta. Koska Covid-19-tauti on erityisen herkästi tarttuva ja riskiryhmille jopa hengenvaarallinen, on suojatoimeksi pandemia-aikana vakiintunut tavanomaisen suojautumisen, kuten käsienpesun ja hyvän yskimishygienian lisäksi, kasvomaskien käyttö julkisilla paikoilla.

Tarkastellaan viittä satunnaisesti valittua tartunnan saanutta henkilöä. Millä todennäköisyydellä

a) heillä kaikilla on eri tartuntalähde?

b) he kaikki ovat saaneet tartunnan perhepiiristä?

c) kaksi henkilöä on saanut tartunnan perhepiiristä ja kolme muualta?

## PSST!

Covid-19-taudin, kansanomaisesti koronan, aiheuttaja on koronaviruksiin kuuluva SARS-CoV-2-virus. Virus todettiin ensimmäisen kerran Kiinassa joulukuussa 2019 ja se levisi maailmanlaajuisesti alkukeväällä 2020. Maailman terveysjärjestö WHO julisti 11.3.2020 taudin levinneen pandemiaksi.

Koronaviruksia tunnetaan useita kymmeniä erilaisia, mutta vain pienehkö osa niistä voi tarttua ihmiseen ja aiheuttaa vakavia, jopa kuolemaan johtavia infektioita. Tällaisia ovat aiheuttaneet SARS-CoV-2-koronaviruksen lisäksi vain kaksi: SARS-CoV- ja MERS-CoV-koronavirukset, jotka levisivät kumpainenkin vakaviksi epidemiaksi asti. SARS vuonna 2003 ja MERS vuonna 2012. Tavallisesti koronavirukset aiheuttavat lieviä hengitystieinfektioita erityisesti syys- ja talvikaudella. Yleensä oireina on jonkinlainen yhdistelmä yskää, nuhaa sekä muita flunssan oireita.

**Lisää aiheesta:**

Koronavirus, Terveyden ja hyvinvoinnin laitos (THL).

Ajankohtaista tietoa Covid-19-taudista, THL.

# Eräs malliratkaisu

a) Kuvan vaakapylväsdiagrammi (histogrammi) on todennäköisyysjakauma, jossa eri tartuntalähteiden todennäköisyyksien summa on 1. Merkitään tapahtumia seuraavasti:

A = ”Tartuntalähteenä on perhepiiri”

B = ”Tartuntalähteenä on yksityistilaisuus/lähipiiri”

C = ” Tartuntalähteenä on työpaikka”

D = ”Tartuntalähteenä on harrastus”

E = ”Tartuntalähteenä on oppilaitos”

Yle Uutisten grafiikassa kyseisten tapahtumien todennäköisyyksien pyöristys on tehty prosentin tarkkuuteen, jolloin satunnaismuuttujan X = ”tautilähde” jakauman muodostavien tapahtumien A, B, C, D ja E todennäköisyyksien summa onkin 101 % eli 1,01. Tämä virhelähde on hyvä ottaa huomioon vastauksen mielekkyyttä ja pyöristystarkkuutta pohdittaessa.

Tapahtumat A, B, C, D ja E ovat toisensa poissulkevia, sillä tartunta on saatu tietyltä tartuttajalta tietyllä hetkellä. Esimerkiksi jos tartunta on saatu perhepiiristä, niin sitä ei ole voitu saada muista tartuntalähteistä. Koska viisi sairastunutta henkilöä on valittu satunnaisesti, voidaan näiden henkilöiden tartunnat lähteineen olettaa toisistaan riippumattomiksi. Tapahtumien poissulkevuuden ja riippumattomuuden nojalla todennäköisyys, että kaikki viisi henkilöä ovat saaneet tartunnan eri tartuntalähteistä, voidaan laskea kertolaskusäännön avulla:

P (”kaikilla eri tartuntalähde”) = P (”yksi saanut tartunnan perhepiiristä **ja** toinen lähipiiristä **ja** kolmas työpaikalta **ja** neljäs harrastuksista **ja** viides oppilaitoksesta”)

$= P(A) \cdot P(B) \cdot P(C) \cdot P(D) \cdot P(E)$

$= 0{,}6 \cdot 0{,}15 \cdot 0{,}10 \cdot 0{,}08 \cdot 0{,}08$

$= 0{,}0000576$

$\approx 6 \cdot 10^{-5}.$

Tehtävänannon jakauman virhelähteen vuoksi tämä vastaus antaa vain suuruusluokan todennäköisyydestä, jolla viidellä satunnaisesti valitulla sairastuneella henkilöllä on eri tartuntalähde. Tämä todennäköisyys on hyvin pieni - suuruusluokaltaan noin kuusi sadastuhannesosaa.

b) Käytetään ratkaisussa a-kohdan merkintöjä. Koska satunnaisesti valittujen viiden henkilön tartunnat voidaan olettaa toisistaan riippumattomiksi, saadaan

P (”kaikki saaneet tartunnan perhepiiristä”) = P (”yksi saanut tartunnan perhepiiristä **ja** toinen **ja** kolmas **ja** neljäs **ja** viides myös”)

$= P(A) \cdot P(A) \cdot P(A) \cdot P(A) \cdot P(A)$

$$= P(A)^5 = 0{,}6^5 \quad = 0{,}07776 \approx 0{,}08.$$

Siis tehtävänannon histogrammin todennäköisyyksien pyöristyksen vuoksi todennäköisyys, jolla viidellä satunnaisesti valitulla sairastuneella henkilöllä on kaikilla tartuntalähteenä perhepiiri, on 0,08 eli 8 % (l. kahdeksan sadasosaa l. kaksi kahdeskymmenesviidesosaa) tai vastaavaa suuruusluokkaa (0,05–0,10). Tämä todennäköisyys on joka tapauksessa merkittävä.

c) Tapahtuma A = ”Tartuntalähteenä on perhepiiri” a-kohdan merkintöjen mukaisesti. Todennäköisyyttä, jolla kaksi henkilöä on saanut tartunnan perhepiiristä ja kolme muualta, voidaan tarkastella binomitodennäköisyytenä, jossa satunnaisen tartunnansaajan valinta toistetaan riippumattomasti viisi (5) kertaa. Kukin valinta onnistuu, eli sen seurauksena valitaan perhepiiristä tartunnan saanut henkilö, todennäköisyydellä p=P(A)=0,6. Merkitään epäonnistumisen todennäköisyyttä q=1-p=1-0,6=0,4. Todennäköisyys q on siis tapahtuman A vastatapahtuman ”tartunta saatu perhepiirin ulkopuolelta” todennäköisyys. Olkoon satunnaismuuttuja X perhepiiristä saatujen tartuntojen lukumäärä. Tällöin kaksi perhepiiritartuntaa saadaan todennäköisyydellä

$$P(X=2) = \binom{5}{2} \cdot 0{,}6^2 \cdot 0{,}4^3$$

$$= \frac{5!}{2!\,3!} \cdot 0{,}6^2 \cdot 0{,}4^3$$

$$= \frac{5 \cdot 4 \cdot \cancel{3 \cdot 2 \cdot 1}}{2 \cdot 1 \cdot \cancel{3 \cdot 2 \cdot 1}} \cdot 0{,}6^2 \cdot 0{,}4^3$$

$$= 10 \cdot 0{,}6^2 \cdot 0{,}4^3$$

$$= 0{,}2304$$

$$\approx 0{,}23.$$

Siis tehtävänannon histogrammin todennäköisyyksien pyöristyksen vuoksi todennäköisyys, jolla kaksi henkilöä on saanut tartunnan perhepiiristä ja kolme muualta, on 0,23 (eli 23 %) tai vastaavaa suuruusluokkaa (n. 0,20–0,25).

## HUOMAUTUS RATKAISUSTA

Toinen tapa ratkaista tehtävä jakauman virhe huomioiden on skaalata jakaumassa esitetyt todennäköisyydet:

P (”tartuntalähteenä perhepiiri”) = $\frac{60}{101}$,

P (”tartuntalähteenä yksityistilaisuudet/lähipiiri”) = $\frac{15}{101}$,

⋮

Käyttämällä skaalattuja todennäköisyyksiä myös vältetään virhemarginaalin pohdinta ratkaisun yhteydessä.

# Tehtävä 4: Seksuaalivähemmistöt

Vuoden 2017 kouluterveyskyselyn mukaan toisen asteen opiskelijoista noin 11 % (15 % tyttöoletetuista, 7 % poikaoletetuista) kokee kuuluvansa seksuaalivähemmistöihin. Suomalaisten lukiolaisten keskuudesta arvotaan 30 henkilön otos. Millä todennäköisyydellä näistä henkilöistä johonkin seksuaalivähemmistöön kuuluu:

a) puolet b) ainakin 28 c) enintään yksi d) ainakin yksi?

## PSST!

HLBTIQ on seksuaali- ja sukupuolivähemmistöistä käytetty lyhenne, joka tulee sanoista homo, lesbo, bi, trans, inter ja queer. Seksuaalinen suuntautuminen on ominaisuus, joka kertoo siitä, kehen ihminen ihastuu, rakastuu tai tuntee vetoa tunnetasolla tai eroottisessa mielessä. Monesti seksuaalinen suuntautuminen määritellään oman sukupuolen ja tunteiden kohteen sukupuolen perusteella, mikä on toisinaan ongelmallinen määrittelytapa niin seksuaalisen suuntautumisen kuin sukupuolien moninaisuuden kannalta. Puhuttiin sitten sukupuolesta tai seksuaalisesta suuntautumisesta, jokaisella ihmisellä on oikeus määritellä, kuka on ja kuinka haluaa tulla nähdyksi. Jokaisella on myös täysi oikeus olla tietämättä. Itsensä saa määritellä, olla määrittelemättä ja muuttaa mielensä määritelmien suhteen.

Kouluterveyskyselyssä kartoitetaan myös sukupuoli- ja seksuaalivähemmistöön kuuluvien nuorten hyvinvointia ja yhdenvertaisuuden toteutumista. Vuoden 2019 kyselyn tulokset osoittavat, että näihin vähemmistöihin kuuluvilla nuorilla oli muihin ikätovereihinsa verrattuna huomattavasti enemmän terveyteen ja hyvinvointiin sekä koulunkäyntiin ja kouluyhteisössä elämiseen liittyviä haasteita, kuten yksinäisyyden, kiusaamisen ja uupumuksen kokemuksia. Erityisen ongelmallista on sateenkaarinuorten kohtaaman väkivallan ja mielenterveyden ongelmien yleisyys sekä tarvitun avun ulkopuolelle jäämisen kokemukset muita nuoria useammin. Seksuaalivähemmistöjen syrjintäsuoja ja yhdenvertaisuuden edistämisvelvoite viranomaisille on kirjattu yhdenvertaisuuslakiin, mutta siitä huolimatta jokaisella meistä on myös yksilönä vastuu yhdenvertaisuuden toteutumisessa.

**Lisää aiheesta:**

Seksuaalinen suuntautuminen, Seta ry.

Itsemäärittelyoikeus ja normit, Seta ry.

Ensimmäinen HLBTIQ-henkilöiden tasa-arvoa koskeva EU:n strategia, Euroopan komissio.

Sukupuoli- ja seksuaalivähemmistöihin kuuluvien nuorten hyvinvointi, THL.

Terveyden ja hyvinvoinnin laitoksen joka toinen vuosi teettämä **Kouluterveyskysely** on perusopetuksen 4.-5. ja 8.-9. vuosiluokkalaisille, lukion 1.-2. vuoden opiskelijoille sekä ammatillisten oppilaitosten 1.-2. vuoden opiskelijoille suunnattu kysely. Siinä kartoitetaan eri-ikäisten lasten ja nuorten hyvinvointiin, terveyteen, koulunkäyntiin, osallisuuteen sekä avun ja tuen saamiseen liittyviä kokemuksia, tapoja ja tottumuksia sekä paikallisesti että maakuntatasolla. Kouluterveyskyselystä saatuja tietoja voidaan hyödyntää hyvinvoinnin edistämistyössä niin oppilaitoksissa, hyvinvointialueilla ja valtakunnallisesti.

# Eräs malliratkaisu

Tilannetta voidaan ajatella **toistokokeena**, jossa toistoja eli tässä yhteydessä henkilöitä on 30 ja onnistunut toisto on tapahtuma A:

A = "henkilö kuuluu seksuaalivähemmistöön".

Tapahtuman A todennäköisyys p on jokaisella toistolla 11% eli 0,11. Näin ollen tapahtuman A vastatapahtuman "henkilö ei kuulu seksuaalivähemmistöön" todennäköisyys q saadaan laskettua $q = 1 - p$ eli

$$q = 1 - 0{,}11 = 0{,}89.$$

Merkitään seksuaalivähemmistöihin kuuluvien henkilöiden lukumäärää 30 henkilön otoksessa satunnaismuuttujalla X. Seksuaalivähemmistöön kuuluvien henkilöiden lukumäärä X noudattaa **binomijakaumaa** Bin(30; 0,11), jolloin

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \text{missä } n = 30, p = 0{,}11, q = 0{,}89 \text{ ja } k = 0, 1, 2, \ldots, 30.$$

Kolmellakymmenellä toistolla k onnistumista saadaan siis todennäköisyydellä

$$P(X = k) = \binom{30}{k} \cdot 0{,}11^k \cdot 0{,}89^{30-k}, k = 0, 1, 2, \ldots, 30.$$

a) Binomitodennäköisyyden mukaisesti puolet 30:stä eli 15 henkilöä kuuluu seksuaalivähemmistöihin todennäköisyydellä:

$$P(X = 15) = \binom{30}{15} \cdot 0{,}11^{15} \cdot 0{,}89^{15} = 155\,117\,520 \cdot 0{,}11^{15} \cdot 0{,}89^{15}$$

$$= 0{,}0000001128\ldots$$

$$\approx 0{,}00000011.$$

Siis 30:stä satunnaisesti valitusta lukiolaisesta puolet kuuluu seksuaalivähemmistöön todennäköisyydellä 0,00000011 eli 0,000011 % (yksitoista miljoonasosaa).

b) Tapahtuma, jossa ainakin 28 henkilöä 30:stä kuuluu seksuaalivähemmistöihin, voidaan ajatella tapahtumana "seksuaalivähemmistöihin kuuluu joko 28 tai 29 tai 30 henkilöä 30:stä", joka taas koostuu kolmesta erillisestä tapahtumasta. Binomitodennäköisyyttä ja **erillisten tapausten yhteenlaskusääntöä** käyttäen saadaan laskettua todennäköisyys, jolla ainakin 28 kuuluu seksuaalivähemmistöihin.

$$P(X \geq 28) = P(X = 28) + P(X = 29) + P(X = 30)$$

$$= \binom{30}{28} \cdot 0{,}11^{28} \cdot 0{,}89^2 + \binom{30}{29} \cdot 0{,}11^{29} \cdot 0{,}89^1 + \binom{30}{30} \cdot 0{,}11^{30} \cdot 0{,}89^0$$

$$= 435 \cdot 0{,}11^{28} \cdot 0{,}89^2 + 30 \cdot 0{,}11^{29} \cdot 0{,}89 + 0{,}11^{30}$$

$$= 5{,}011 \ldots \cdot 10^{-25}$$

$$\approx 5{,}1 \cdot 10^{-25}.$$

Siis 30:stä satunnaisesti valitusta lukiolaisesta ainakin 28 kuuluu seksuaalivähemmistöön todennäköisyydellä $5{,}1 \cdot 10^{-25}$ eli 0,00000000000000000000000050 %. Tämä todennäköisyys on niin lähellä nollaa, että voidaan puhua melkein mahdottomasta tapahtumasta.

c) Samoin kuin b-kohdassa, saadaan binomitodennäköisyyden ja erillisten tapausten yhteenlaskusäännön avulla laskettua todennäköisyys, jolla enintään yksi kolmestakymmenestä kuuluu seksuaalivähemmistöihin. Tapahtuma "enintään yksi 30:stä henkilöstä kuuluu seksuaalivähemmistöön" voidaan ajatella tapahtumaksi "0 tai 1 henkilöä kolmestakymmenestä kuuluu seksuaalivähemmistöön". Täten

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{30}{0} \cdot 0{,}11^0 \cdot 0{,}89^{30} + \binom{30}{1} \cdot 0{,}11^1 \cdot 0{,}89^{29} \\ &= 0{,}89^{30} + 30 \cdot 0{,}11 \cdot 0{,}89^{29} \\ &= 0{,}1427 \ldots \\ &\approx 0{,}14. \end{aligned}$$

Siis kolmestakymmenestä satunnaisesti valitusta lukiolaisesta enintään yksi kuuluu seksuaalivähemmistöön todennäköisyydellä 0,14 eli 14 % (neljätoista sadasosaa).

d) Tapahtuman, jossa ainakin yksi henkilö 30:stä kuuluu seksuaalivähemmistöön, todennäköisyys saadaan laskettua sen vastatapahtuman "yksikään henkilö ei kuulu seksuaalivähemmistöön" avulla:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - 0{,}89^{30} \\ &= 0{,}9696 \ldots \\ &\approx 0{,}97. \end{aligned}$$

Siis kolmestakymmenestä satunnaisesti valitusta lukiolaisesta ainakin yksi kuuluu seksuaalivähemmistöön todennäköisyydellä 0,97 eli 97 %. On siis erittäin todennäköistä (lähes varmaa), että esimerkiksi 30 opiskelijan ryhmässä on ainakin yksi seksuaalivähemmistöön kuuluva opiskelija.

## HUOMAUTUS RATKAISUSTA

Edeltävien b- ja c-kohtien tavoin kysytty tapahtuma voitaisiin paloitella erillisiksi tapauksiksi: yksi tai kaksi tai kolme tai … tai 29 tai 30 henkilöä kolmestakymmenestä kuuluu seksuaalivähemmistöihin. Tämä tapa olisi kuitenkin esimerkiksi ylioppilaskokeen A-osiossa yllä esitettyä ratkaisua alttiimpi virheille, sillä laskettavaa on enemmän.

# Tehtävä 5: Kansantaudit

Syöpäsairaudet ja diabetes ovat Suomessa kansantauteja. Yksilöllisten ja inhimillisten haittojen lisäksi kansantaudit vaikuttavat väestötasolla terveydentilaan, kuolleisuuteen ja työkykyyn sekä kuormittavat rajusti terveydenhuoltoa vaikuttaen samalla myös kansantalouteen.

Syöpäjärjestöjen tiedon mukaan joka kolmas suomalainen sairastuu jossakin elämänsä vaiheessa syöpään. Lähes kaksi kolmesta kuitenkin paranee. FinTerveys 2017-tutkimuksen mukaan tyypin 2 diabetekseen sairastuu elämänsä aikana n. 13% väestöstä (10 % naisista ja 15 % miehistä). Hiljattain todetusta diabeteksesta on mahdollista parantua täysin ja pitkään sairastanutkin voi saada sairauden remissiotilaan eli piiloon elämäntaparemontilla (terveellinen ravinto ja liikunnan lisääminen: eroon ylipainosta). Dataa remissioon päässeistä suomalaisista diabeetikoista ei ole kerätty, mutta brittitutkimuksessa* kahden vuoden seurannan jälkeen ilman painonhallintaohjelmaa remissioon pääsi 3 %, kun taas painonhallintaohjelman saaneista 36 % pääsi remissioon. Voidaan olettaa, että elintapaohjauksella on sama vaikutus myös suomalaisiin diabeetikoihin ja että sairaudesta parantuminen ei riipu sairastumisesta, vaan muista asioista (esim. hoidon onnistuminen ja yleiskunto).

Lääketieteellisissä tutkimuksissa** on havaittu voimakas yhteys syöpien ja tyypin 2 diabeteksen sairastavuuden välillä. Tämä johtuu erityisesti siitä, että molemmilla sairauksilla on yhteisiä riskitekijöitä. Tarkkaa arviota siitä, kuinka suuri osa potilaista sairastaa molempia samanaikaisesti, ei ole. Oletetaan tässä tehtävässä, että itse sairastumistapahtumat ovat toisistaan riippumattomia.

Laske, kuinka suurella todennäköisyydellä satunnainen suomalainen elämässään:

a) sairastuu sekä syöpäsairauteen että 2-tyypin diabetekseen

b) ei sairastu kumpaankaan

c) sairastuu ainakin toiseen

d) sairastuu enintään toiseen

e) sairastuu vain toiseen

f) sairastuu syöpään ja paranee siitä

g) sairastuu diabetekseen ja pääsee remissioon, kun hän on saanut painonhallintaohjelman.

Samanaikaissairastavuus eli komorbiditeetti tarkoittaa kahden tai useamman sairauden tai niin sanotun häiriön samanaikaista esiintymistä. Ilmiö on yleisempi sellaisilla sairauksilla, joilla on yhteisiä riskitekijöitä. Esimerkiksi ylipaino, tupakointi, korkea verenpaine ja veren korkea kolesterolipitoisuus ovat yhteisiä riskitekijöitä useille suomalaisille kansantaudeille, kuten syöville, diabetekselle sekä sydän- ja verisuonisairauksille. Myös erilaiset oppimisvaikeudet esim. hahmotusvaikeudet, lukivaikeus ja tarkkaavaisuuden vaikeudet esiintyvät usein päällekkäin.

**Lisää aiheesta**

Syöpä, Kaikki syövästä (Syöpäjärjestöt).

Diabeteksen ilmenemismuodot, Suomen Diabetesliitto ry.

Huonosti hoidettuna diabetes kuormittaa munuaisia
diabetes ja munuaiset & munuaisten tehtävät, Diabetesliitto ja Munuais- ja maksaliitto.

---

* https://www.thelancet.com/journals/landia/article/PIIS2213-8587(19)30068-3/fulltext
** https://pubmed-ncbi-nlm-nih-gov.ezproxy.utu.fi/27219686/

# Eräs malliratkaisu

Merkitään tehtävänannon tapahtumia ja niitä vastaavia todennäköisyyksiä seuraavasti:

S = "Henkilö sairastuu syöpään", $P(S) = \frac{1}{3}$

E = $S^C$= "Henkilö ei sairastu syöpään", $P(E) = 1 - P(S) = \frac{2}{3}$

K = "Sairastunut paranee syövästä", $P(K|S) = \frac{2}{3}$

D = "Henkilö sairastuu diabetekseen", $P(D) = 0{,}13$

T = $D^C$= "Henkilö ei sairastu diabetekseen", $P(T) = 1 - P(D) = 0{,}87$

R = "Painonhallintaohjelmaan osallistunut diabeetikko pääsee remissioon", $P(R|D) = 0{,}36$.

Oletetaan, että syöpään ja diabetekseen sairastuminen ovat toisistaan riippumattomat tapahtumat.

a) Todennäköisyys, jolla henkilö sairastuu sekä syöpäsairauteen että 2-tyypin diabetekseen, saadaan laskettua seuraavasti:

$$P(S\ ja\ D) = P(S) \cdot P(D) = \frac{1}{3} \cdot 0{,}13 = 0{,}043333.. \approx 0{,}04.$$

Siis todennäköisyys, että henkilö sairastuu elämänsä aikana molempiin sairauksiin, on noin 4 %.

b) Todennäköisyys, että henkilö ei sairastu kumpaankaan, saadaan laskettua seuraavalla tavalla:

$$P(E\ ja\ T) = P(E) \cdot P(T) = \frac{2}{3} \cdot 0{,}87 = 0{,}58.$$

Siis todennäköisyys, että henkilö ei sairastu elämänsä aikana kumpaankaan sairauteen, on 58 %.

c) Todennäköisyys, että henkilö sairastuu ainakin toiseen kyseisistä sairauksista, voidaan laskea vastatapahtuman todennäköisyyttä hyödyntäen seuraavasti:

$$\begin{aligned} P(sairastuu\ ainakin\ toiseen) &= 1 - P(ei\ sairastu\ kumpaankaan) \\ &= 1 - 0{,}58 \\ &= 0{,}42. \end{aligned}$$

Siis todennäköisyys, että henkilö sairastuu elämänsä aikana ainakin toiseen kansantautiin, on 42 %.

d) Todennäköisyys, että henkilö sairastuu enintään toiseen kyseisistä sairauksista, tarkoittaa sitä, että henkilö sairastuu **joko** syöpään, mutta ei tyypin 2 diabetekseen **tai** tyypin 2 diabetekseen, mutta ei syöpään **tai** hän ei sairastu kumpaankaan. Saadaan täten

$$\begin{aligned} P(sairastuu\ enintään\ toiseen) &= P(S) \cdot P(T) + P(D) \cdot P(E) + P(E) \cdot P(T) \\ &= \frac{1}{3} \cdot 0{,}87 + 0{,}13 \cdot \frac{2}{3} + \frac{2}{3} \cdot 0{,}87 \end{aligned}$$

$$= 0{,}956666\ldots$$

$$\approx 0{,}96.$$

Siis todennäköisyys, että henkilö sairastuu elämänsä aikana enintään toiseen taudeista, on 96 %.

e) Todennäköisyys, että henkilö sairastuu vain toiseen kyseisistä sairauksista, tarkoittaa sitä, että hän sairastuu **joko** syöpään, mutta ei tyypin 2 diabetekseen **tai** tyypin 2 diabetekseen, mutta ei syöpään. Saadaan siis

$$P(sairastuu\ vain\ toiseen) = P(S) \cdot P(T) + P(D) \cdot P(E)$$

$$= \frac{1}{3} \cdot 0{,}87 + 0{,}13 \cdot \frac{2}{3}$$

$$= 0{,}376666\ldots$$

$$\approx 0{,}38.$$

Eli todennäköisyys, että henkilö sairastuu elämänsä aikana vain toiseen taudeista, on 37 %.

f) Syöpään sairastuminen ja siitä paraneminen eivät ole toisistaan riippumattomia tapahtumia, sillä paraneminen voi tapahtua vain silloin, kun henkilö on ensin sairastunut. Tapahtuma "henkilö paranee syövästä" tarkoittaa siis itse asiassa tapahtumaa "henkilö paranee syövästä, kun hän on sairastunut syöpään". Näin ollen tapahtuman "henkilö sairastuu syöpään ja paranee siitä" laskemisessa käytetään yleistä kertolaskusääntöä ja saadaan

$$P(hlö\ sairastuu\ syöpään\ ja\ paranee) = P(S) \cdot P(K|S)$$

$$= \frac{1}{3} \cdot \frac{2}{3}$$

$$= 0{,}222222\ldots$$

$$\approx 0{,}22.$$

Eli todennäköisyys, että henkilö sairastuu elämänsä aikana syöpään ja paranee siitä, on 22 %.

g) Kuten edeltävässä kohdassa, myös tässä on huomattava, että diabetekseen sairastuminen ja remissiotilaan pääseminen eivät ole toisistaan riippumattomia tapahtumia, sillä remissioon voi päästä vain silloin, kun henkilö on ensin sairastunut. Tapahtuma "henkilö pääsee remissioon" tarkoittaa siis itse asiassa tapahtumaa "henkilö pääsee remissioon, kun hän on ensin sairastunut diabetekseen". Näin ollen käyttämällä yleistä kertolaskusääntöä saadaan

$$P(hlö\ sairastuu\ diabetekseen\ ja\ pääsee\ remissioon) = P(D) \cdot P(R|D)$$

$$= 0{,}13 \cdot 0{,}36$$

$$= 0{,}0468$$

$$\approx 0.05$$

Eli todennäköisyys, että henkilö sairastuu elämänsä aikana diabetekseen ja tauti saadaan remissiotilaan, on 5 %.

# Tehtävä 6: Puutiaisaivotulehdukset Saaristomeren alueella

Arviolta noin 500 000:ta suomalaista puree vuosittain punkki. Puutiaisaivokuumetta aiheuttavaa virusta kantaa noin 1-2 % riskialueiden* punkeista. Puutiaisaivotulehdus (TBE) on punkin kaikkien kehitysmuotojen (toukka, nymfi ja aikuinen) välityksellä tarttuva virustauti, jolle ei ole olemassa lääkehoitoa. Siihen on kuitenkin olemassa rokote, joka suojaa taudilta. Virus tarttuu punkista ihmiseen minuuteissa, minkä vuoksi ihon tarkastaminen punkkien varalta ei auta. Lähivuosina Suomessa on raportoitu vuosittain keskimäärin 60–80 puutiaisaivotulehdustapausta, mutta esimerkiksi vuonna 2020 diagnosoitiin 91 tapausta, mikä on enemmän kuin koskaan aikaisemmin. Määritä alla olevasta puutiaisaivotulehduksen esiintyvyys -taulukosta**

a) tapausaineiston moodi sekä

b) tapausten lukumäärien keskiarvo ja otoskeskihajonta

vuonna 2020 taulukossa nimetyillä paikkakunnilla. Selitä, kuinka hyvin keskiarvo mielestäsi kuvaa aineistoa ja mitä keskihajonnan arvo tarkoittaa tässä tilastossa.

---

* Eteläinen Suomi saaristoineen sekä Ahvenanmaan saaristo.

**Taulukko supistettu Varsinais-Suomeen ja Ahvenanmaahan *Puutiaisaivokuumeen esiintyvyys ja rokotussuositukset tartuntapaikkakunnittain* -taulukosta (THL, 2021).

| Kunta tai kaupunki | Tapausten lkm vuonna 2020 |
|---|---|
| Parainen | 7 |
| Ahvenanmaa | 15 |
| Kustavi | 0 |
| Naantali | 0 |
| Uusikaupunki | 2 |
| Kaarina | 2 |
| Turku | 1 |

THL ylläpitää valtakunnallista **tartuntatautirekisteriä** (tilastotietokanta) tartuntatautilakiin ja -asetukseen perustuen. Tartuntatautirekisteriin ilmoitetaan vuosittain noin 100 000 tartuntatautitapausta. Tietoja on kerätty vuodesta 1995 alkaen.

Tartuntatautirekisteriin voivat tehdä ilmoituksia lääkärit ja laboratoriot. Rekisterin lukuja tarkastellessa on hyvä pitää mielessä, että niihin vaikuttaa se, kuinka suuri osa tautiin sairastuneista hakeutuu terveydenhuoltoon sekä kuinka suurelle osalle tehdään mikrobiologisia tutkimuksia aiheuttajan toteamiseksi. Lisäksi lukuihin vaikuttavat muutokset diagnostisissa käytännöissä ja menetelmissä.

## PSST!

Puutiaiset, tuttavallisemmin punkit, ovat levinneet yhä laajemmalle alueelle Suomessa ja samalla punkkien välittämät borrelioosi- ja puutiaisaivotulehdustartuntojen määrät ovat kasvaneet. Yhtenä syynä siihen, että punkkeja tavataan yhä pohjoisemmilla leveysasteilla sekä yhä varhaisemmin keväällä ja myöhemmin syksyllä, pidetään ilmaston lämpenemistä. Uusia punkkiriskialueita on syntynyt jopa aivan kaupunkien puisto- ja viheralueille. Tämän vuoksi on hyödyllistä tietää, millä alueilla punkkeja on havaittu ja kuinka niiltä voidaan suojautua.

**Lisää aiheesta:**

Tietoa TBE-viruksesta: puutiaisaivotulehdus ja TBE:n esiintyvyys kartalla, THL

Tietoa borrelia-viruksesta ja oireista: borrelioosi, THL

Punkkien levinneisyys ja seuranta: punkkilive, Turun yliopisto

Punkkien levinneisyyttä kartoitetaan kansalaistutkimuksen avulla: mediatiedote, Turun yliopisto

# Eräs malliratkaisu

a) Moodi eli tyyppiarvo on havaintoaineistossa yleisimmin esiintyvä tilastomuuttujan arvo eli se arvo, jolla on suurin frekvenssi. Tämän aineistoin tyyppiarvo on "Ahvenanmaa", sillä sen frekvenssi on 15 eli siellä puutiaisaivokuume- eli TBE-tapauksia on ollut 15 vuonna 2020.

b) Keskiarvo kuvaa, kuinka paljon puutiaisaivokuumetapauksia eri paikkakunnilla on keskimäärin. Tilastomuuttujan keskiarvo saadaan laskemalla yhteen muuttujan arvot ja jakamalla summa havaintojen lukumäärällä:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \ = \frac{1}{7}\cdot(7+15+0+0+2+2+1) = \frac{27}{7} = 3\frac{6}{7} \approx 4.$$

Taulukon paikkakuntien puutiaisaivokuumetapausten keskiarvo on siis 4 tapausta paikkakunnalla. Tämä keskiarvo ei yksistään riitä kuvamaan aineistoa, sillä se ei anna kuvaa TBE-tapausten todellisesta jakautumisesta eri paikkakunnille; esimerkiksi Kustavissa ja Naantalissa TBE-tapauksia ei ole vuonna 2020 ollut lainkaan, kun taas Ahvenanmaalla tapauksia on ollut lähes nelinkertainen määrä keskiarvoon verrattuna.

Jakauman hajontaa mittaavat tunnusluvut ilmentävät havaintojen eli tässä TBE-tapausten lukumäärän vaihtelun suuruutta havaintoaineistossa. Tutkitaan TBE-tapausten keskihajontaa, joka kuvaa havaintoaineiston arvojen poikkeamaa keskiarvosta. Keskihajonta $\sigma$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2} = \sqrt{\frac{1}{7}\sum_{i=1}^{7}(x_i-4)^2}$$

$$= \sqrt{\frac{(7-4)^2+(15-4)^2+(0-4)^2+(0-4)^2+(2-4)^2+(2-4)^2+(1-4)^2}{7}}$$

$$= \sqrt{\frac{(3)^2+(11)^2+(-4)^2+(-4)^2+(-2)^2+(-2)^2+(-3)^2}{7}}$$

$$= \sqrt{\frac{1253}{7}} = 5{,}0568\ldots \approx 5.$$

Keskihajonta on siis 5 eli TBE-tapausten lukumäärät eri paikkakunnilla ovat hajautuneet keskimäärin 5:n tautitapauksen päähän keskiarvosta.

# Tehtävä 7: Kestävyystreeni

Tutkimusnäyttöön perustuvan kansallisen Käypä hoito-suosituksen määritelmän mukaan kestävyysliikunta eli **aerobinen liikunta** on "suuria lihasryhmiä vähintään kohtalaisesti (suhteessa toteuttajansa suorituskykyyn) kuormittavaa, yleensä ainakin kymmeniä minuutteja (yhtäjaksoisesti tai jaksoittain) kestävää, aineenvaihduntaa ja hengitys- ja verenkiertoelimistöä kehittävää ja tällaisessa liikunnassa jaksamista ylläpitävää tai lisäävää liikuntaa."

Matalatehoisessa aerobisessa harjoittelussa liikutaan sykkeellä, joka on noin 60-70% maksimisykkeestä. Eräässä tutkimuksessa todettiin, että matalatehoisessa aerobisessa harjoittelussa koehenkilöiden sykkeen keskiarvo oli 130 bpm (beats per minute) ja keskihajonta 8 bpm. Syke on kutakuinkin normaalisti jakautunut.

a) Kuinka monella prosentilla koehenkilöistä syke oli alle 120 lyöntiä minuutissa?
b) Millä todennäköisyydellä satunnaisesti valitun koehenkilön syke oli välillä 120-140 lyöntiä minuutissa?
c) Millä sykevälillä syke oli 95,0 %:lla koehenkilöistä?

## PSST!

Ihmisen syke kertoo, montako kertaa sydän lyö yhden minuutin aikana. Syke vaihtelee ennen muuta kehon fyysisen aktiivisuuden ja rasituksen mukaan: mitä enemmän keho rasittuu, sitä enemmän sydämen on pumpattava verta työtä tekeville lihaksille. Syke on alhaisimmillaan levossa (nukkuessa) ja korkeimmillaan voimakkaassa fyysisessä rasituksessa esim. HIIT-harjoituksessa. Sykkeeseen vaikuttaa myös henkilön psyyke, kuten mielialan muutokset, stressi ja jännitys, sekä tupakka, alkoholi ja kofeiinipitoiset juomat.

Maksimisyke tarkoittaa sykettä, jonka henkilö voi saavuttaa kovassa fyysisessä suorituksessa ilman, että sen hetkellisestä saavuttamisesta on hänelle vaaraa. Maksimisyke on yksilöllinen ominaisuus, joka riippuu muun muassa henkilön iästä ja sukupuolesta. Vanhemmiten maksimisyke laskee. Yksi yleinen tapa arvioida henkilön maksimisykettä on vähentää henkilön ikä luvusta 220.

Leposyke tarkoittaa nimensä mukaisesti sykettä henkilön ollessa levossa, mutta kuitenkin hereillä. Tyypillisesti leposyke on aikuisella noin 50-90 lyöntiä minuutissa, mutta maksimisykkeen lailla leposyke on yksilöllinen ja myös geneettinen ominaisuus. Aktiiviliikkujilla leposyke voi hyvin olla alle viitearvon, sillä kestävyysliikunta ja –urheilu laskevat leposykettä. Jännittäviä huhuja liikkuu juoksijalegendojen historiallisen alhaisista leposykkeistä 25-30 lyönnin välillä. Kuitenkin esimerkiksi Paavo Nurmella oli nykytiedon valossa normaali sydän ja syke, vaikka aikanaan Amerikassa värikkäästi tokaistiin "alhaisin pulssi, korkein taksa".

**Lisää aiheesta**

Syke-käsitteistö kuntoilijalle, Polar.

Matalatehoinen peruskestävyysharjoittelu rakentaa kuntopohjaa ja auttaa palautumaan arjesta, Liikunta & Tiede 2-3/2019.

# Eräs malliratkaisu

Merkitään sykettä satunnaismuuttujalla X. Nyt $X \sim N(130{,}8)$. Koska odotusarvo poikkeaa nollasta ja keskihajonta on erisuuri kuin 1, tulee suorittaa normittaminen.

a) $P(X < 120) = P(Z < \frac{120-130}{8})$

$$= P(Z < -\frac{5}{4})$$

$$= P(Z < -1{,}25).$$

Tästä normaalijakauman symmetrian nojalla saadaan:

$$P(Z < -1{,}25) = 1 - P(Z \leq 1{,}25)$$

$$= 1 - \Phi(1{,}25)$$

$$= 1 - 0{,}8944$$

$$= 0{,}1056$$

$$\approx 10{,}6\ \%.$$

Siis noin 10,6 %:lla koehenkilöistä syke oli alle 120 lyöntiä minuutissa.

b) $P(120 < X < 140) = P(\frac{120-130}{8} < Z < \frac{140-130}{8})$

$$= P(-1{,}25 < Z < 1{,}25)$$

$$= \Phi(1{,}25) - \Phi(-1{,}25)$$

$$= \Phi(1{,}25) - [1 - \Phi(1{,}25)]$$

$$= \Phi(1{,}25) + \Phi(1{,}25) - 1$$

$$= 2 \cdot \Phi(1{,}25) - 1$$

$$= 2 \cdot 0{,}8944 - 1$$

$$= 0{,}7888$$

$$\approx 78{,}9\ \%.$$

Siis satunnaisesti valitun koehenkilön syke on välillä 120-140 lyöntiä minuutissa 78,9 %:n todennäköisyydellä.

c) Normaalijakauman symmetrian vuoksi oletetaan, että kysytyn välin minimi- ja maksimisyke ovat yhtä kaukana odotusarvosta (keskiarvosta) 130 bpm. Täten voidaan merkitä kysyttyä sykeväliä $(130 - a, 130 + a)$. Syke kuuluu kyseiselle välille todennäköisyydellä

$$P(130 - a < X < 130 + a) = P(\frac{130 - a - 130}{8} < Z < \frac{130 + a - 130}{8})$$

$$= P(-\frac{a}{8} < Z < \frac{a}{8})$$

$$= \Phi\left(\frac{a}{8}\right) - \Phi(-\frac{a}{8})$$

$$= \Phi\left(\frac{a}{8}\right) - [1 - \Phi\left(\frac{a}{8}\right)]$$

$$= 2 \cdot \Phi\left(\frac{a}{8}\right) - 1$$

Koska tehtävässä oltiin kiinnostuneita sykevälistä, johon koehenkilön syke kuuluu 95,0 %:n todennäköisyydellä, saadaan yhtälö

$$2 \cdot \Phi\left(\frac{a}{8}\right) - 1 = 0{,}950 \quad || + 1$$

$$2 \cdot \Phi\left(\frac{a}{8}\right) = 1{,}950 \quad ||: 2$$

$$\Phi\left(\frac{a}{8}\right) = 0{,}975$$

Taulukosta saadaan $\Phi(x) = 0{,}975$, kun $x = 1{,}96$. Näin ollen

$$\frac{a}{8} = 1{,}96 \quad || \cdot 8$$

$$a = 15{,}68.$$

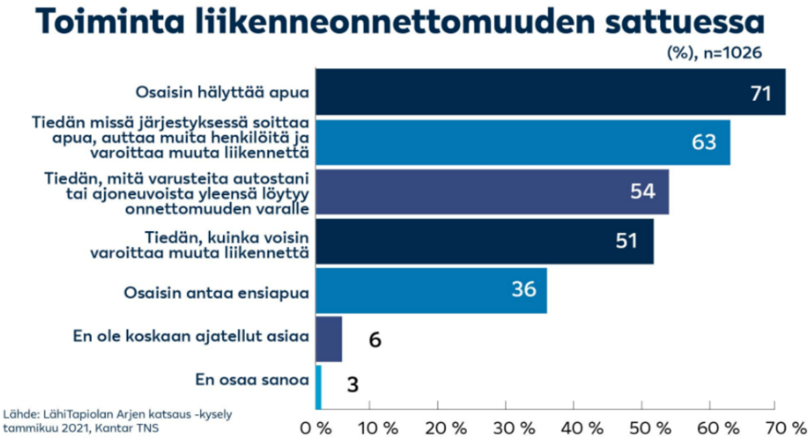
Sijoittamalla $a = 15{,}68 \approx 16$ sykevälille $(130 - a, 130 + a)$, saadaan sykeväliksi (114,146). Tämä on siis sykeväli, johon 95,0 %:n koehenkilöistä syke kuului heidän tehdessään matalatehoista aerobista harjoitusta.

# Tehtävä 8: Ensiaputaidot

Ensiapu on loukkaantuneelle tai sairastuneelle henkilölle tapahtumapaikalla usein nopeasti annettavaa apua, jolla pyritään turvaamaan autettavan peruselintoiminnot sekä estämään hänen tilansa paheneminen. Ensiavun käsite pitää sisällään monenlaisia auttamisen taitoja; se voi olla tilanteesta riippuen esimerkiksi hätäkeskukseen soittamista, tilannearvion tekemistä onnettomuuspaikalla, sairauskohtauksissa auttamisen taitoja, haavojen sidontaa, henkistä tukemista, tukehtumisen estämistä tai paineluelvytystä. Ensiapua ja hätäensiapua voidaan antaa, vaikka antajalla ei olisi erityistä ammattitaitoa tai välineitä. Usein maallikon antama apu parantaa huomattavasti apua saavan mahdollisuuksia selviytyä ja toipua.

LähiTapiolan Arjen katsaus -kyselyn (n=1026) tulosten mukaan noin 36% ihmisistä osaisi oman arvionsa mukaan antaa ensiapua liikenneonnettomuuden sattuessa. Valitaan viiden hengen satunnainen otos suomalaisia. Asetetaan satunnaismuuttujaksi X otokseen valittujen liikenneonnettomuustilanteessa ensiaputaitoisten henkilöiden lukumäärä.

a) Perustele, mitä jakaumaa satunnaismuuttuja X noudattaa ja miksi.
b) Muodosta satunnaismuuttujan X jakauma, esitä se taulukossa ja selvitä, mikä on sen todennäköisin arvo.



Lähde: LähiTapiolan Arjen katsaus -kysely tammikuu 2021, Kantar TNS

Auttamisen taitoja harjoitellaan suomalaiskouluissa jo alakoulusta lähtien ja moni suomalainen käy ensiapukurssin tai -kursseja. Suomen Punaisen Ristin (SPR) Ensiapukurssi EA 1® -koulutus on virallinen sertifioitu koulutus, josta saa Punaisen Ristin ensiapukortin eli pätevyystodistuksen. Kurssin suorittanut hallitsee ensiavun antamisen perusteet kurssin sisällön mukaisissa aiheissa. SPR Ensiapukurssi EA 1® -todistus on voimassa 3 vuotta.

PSST!

Bee Gees yhtyeen Stayin' Alive -kappaleessa on oikea rytmi (104 lyöntiä minuutissa) paineluelvytyksen antamiseen. Kappale on esimerkiksi Amerikan sydänliiton (the American Heart Association) suosittelema elvytyksen taustalle.

**Lisää aiheesta:**

Ensiapuohjeita eri tilanteisiin, Suomen Punainen Risti.

Maailman elvytyspäivä 16.10.(WorldRestartAHeart) ja biisejä elvytystilanteeseen, SPR ja Sydänliitto.

Arjen katsaus -kysely 2021: Miten suomalaiset toimivat liikenneonnettomuuden sattuessa, LähiTapiola.

# Eräs malliratkaisu

a) Tehtävän tilannetta voidaan ajatella toistokokeena, jossa toistoja on viisi. Niin sanotun onnistuneen toiston eli tässä ensiaputaitoisen henkilön löytymisen todennäköisyys on jokaisella toistolla sama 36% eli 0,36. Toistokokeessa onnistuneiden toistojen lukumäärä noudattaa aina binomijakaumaa eli $X \sim \mathrm{Bin}(5; 0{,}36)$, jossa satunnaismuuttuja X on otokseen valittujen liikenneonnettomuustilanteessa ensiaputaitoisten henkilöiden lukumäärä.

b) Ensiaputaitoisia henkilöitä voi olla viiden henkilön joukossa 0, 1, 2, 3, 4 tai 5. Näin ollen satunnaismuuttujan X mahdolliset arvot 0, 1, 2, 3, 4 tai 5. Lasketaan eri arvoja vastaavat pistetodennäköisyydet toistokokeen kaavalla

$P(X = k) = \binom{n}{k} p^k q^{n-k}$, jossa n on toistojen määrä.

Tässä n=5; k=0, 1, 2, 3, 4, 5; p=0,36 ja q=1-p=1-0,36.

Viiden joukossa on 0 ensiaputaitoista:

$$P(X = 0) = \binom{5}{0} 0{,}36^0 \cdot 0{,}64^5 = 0{,}1073\ldots \approx 0{,}11$$

Viiden joukossa on 1 ensiaputaitoinen:

$$P(X = 1) = \binom{5}{1} 0{,}36^1 \cdot 0{,}64^4 = 0{,}3019\ldots \approx 0{,}30$$

Viiden joukossa on 2 ensiaputaitoista:

$$P(X = 2) = \binom{5}{2} 0{,}36^2 \cdot 0{,}64^3 = 0{,}3397\ldots \approx 0{,}34$$

Viiden joukossa on 3 ensiaputaitoista:

$$P(X = 3) = \binom{5}{3} 0{,}36^3 \cdot 0{,}64^2 = 0{,}1911\ldots \approx 0{,}19$$

Viiden joukossa on 4 ensiaputaitoista:

$$P(X = 4) = \binom{5}{4} 0{,}36^4 \cdot 0{,}64^1 = 0{,}0537\ldots \approx 0{,}05$$

Viiden joukossa kaikki ovat ensiaputaitoisia:

$$P(X = 5) = \binom{5}{5} 0{,}36^5 \cdot 0{,}64^0 = 0.0060\ldots \approx 0.01$$

Tämä on satunnaismuuttujan X jakauma. Se voidaan esittää taulukkona

| k | P(X=k) |
|---|---|
| 0 | 0,11 |
| 1 | 0,30 |
| 2 | 0,34 |
| 3 | 0,19 |
| 4 | 0,05 |
| 5 | 0,01 |

Satunnaismuuttujan X todennäköisin arvo on 2, sillä sitä vastaava pistetodennäköisyys 0,34 on suurin. Täten siis on todennäköisintä, että viiden henkilön joukosta ensiaputaitoisia on kaksi.

# Tehtävä 9: Koulukiusaaminen

Koulukiusaaminen on koulussa tai muussa oppilaitoksessa tapahtuvaa psyykkistä, sosiaalista tai fyysistä väkivaltaa, joka on tarkoituksellista ja toista vahingoittavaa. Kiusaaminen on sen uhrille henkisesti vahingollista ja voi aiheuttaa traumoja, jotka saattavat jatkua läpi elämän. Kouluterveyskyselyssä 2021 yhteensä 91 560 8. ja 9. luokkalaiselta kysyttiin: "Kuinka usein sinua on kiusattu koulussa tämän lukukauden aikana?" Merkitään edellä mainittua kiusaamisen yleisyyttä kartoittavaa kysymystä diskreetillä satunnaismuuttujalla X. Alla olevan taulukon jakauma mukailee saatuja tuloksia.

a) Laadi jakaumasta pylväsdiagrammi.
b) Määritä suhteellinen frekvenssi f%, summafrekvenssi sf ja suhteellinen summafrekvenssi sf%.
c) Selvitä, millä todennäköisyydellä 8. tai 9. luokan opiskelija on joutunut kiusaamisen uhriksi lukukauden aikana.

| x | f |
|---|---|
| useita kertoja viikossa | 2 106 |
| noin kerran viikossa | 3 388 |
| harvemmin | 17 213 |
| ei lainkaan | 68 853 |

Koulukiusaamisella on niin yksilö-, yhteisö- kuin väestötasonkin seurauksia. Kiusaaminen aiheuttaa turvattomuuden tunnetta sekä kiusatulle että hänen koululuokalleen. Se on vakava uhka hyvinvoinnille ja terveydelle, ja voi aiheuttaa mm. itsetunto-ongelmia, ahdistuneisuutta, yksinäisyyttä ja vaikeutta luottaa muihin ihmisiin. Kiusaaminen voi jättää jälkensä kiusattuun pysyvästi ja olla yhteydessä psyykkisen ja sosiaalisen terveyden sekä työelämässä menestymisen ongelmiin myös aikuisuudessa. Kiusaamisen kova hinta näkyy sekä sen inhimillisissä että kansanterveydellisissä kustannuksissa.

Alla olevat kuvat Unescon *Behind the numbers: Ending school violence and bullying* -raportista (2019).

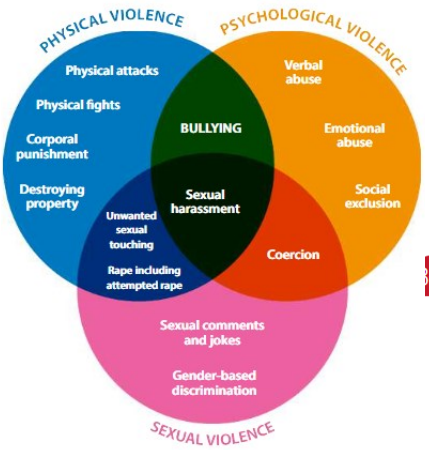
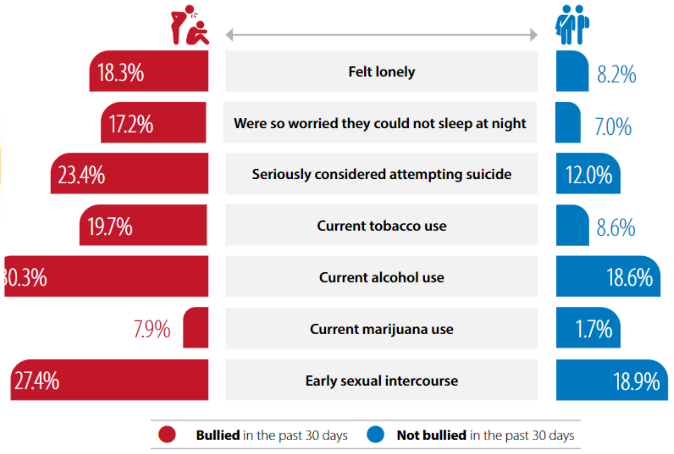


**Figure 21. Differences in mental health status and the prevalence of risk behaviours between students who were bullied and those who were not bullied**



Source: Secondary analysis calculations based on GSHS data.

**Lisää aiheesta:**

Tietoa koulukiusaamisesta, KiVa Koulu

Niko Tiuraniemen tarina kiusatuksi joutumisesta, Yle 20.2.2022

Kiusaamisen yleisyys yläkoulussa ja toisella asteella vuosina 2006-2021, THL Kouluterveyskysely

Kiusaamisen vastainen toimenpideohjelma KiVa Koulu, KiVA Koulu -ohjelma ja Turun yliopisto

Koulukiusaamisen vastaisten ohjelmien arviointi, Yle 20.10.2021

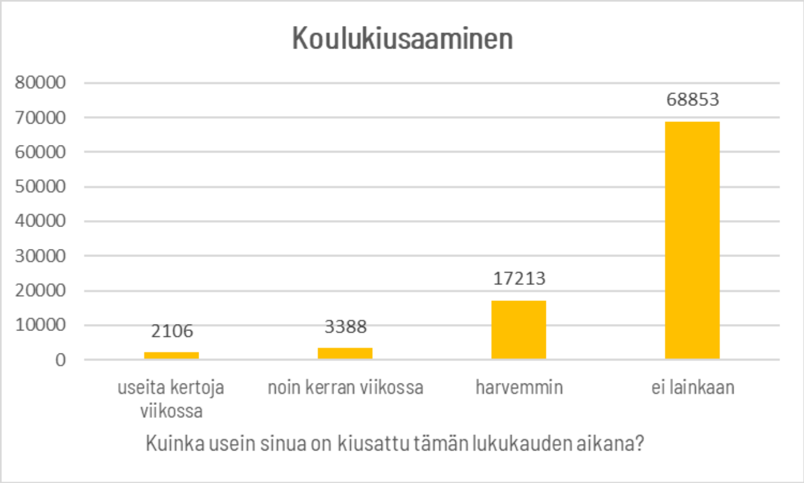
Harjoituksia mielenterveyden vahvistamiseen, Suomen Mielenterveys ry

Koulukiusaaminen on maailmalla ja sitä ehkäiseviä asioita, Unesco, The Global Education 2030 Agenda -hanke

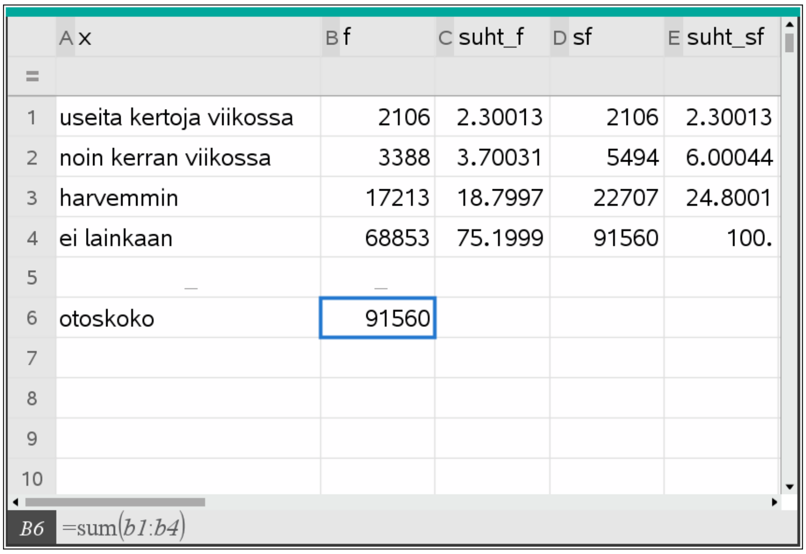
# ERÄS MALLIRATKAISU

Tässä ratkaisussa on käytetty MS Excel -ohjelmistoa ja TI-Nspire CX CAS -ohjelmistoa.

a) Laaditaan tehtävänannon taulukon pohjalta pylväsdiagrammi.



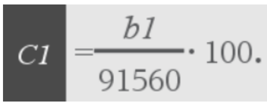
b) Määritetään suhteellinen frekvenssi f%, summafrekvenssi sf ja suhteellinen summafrekvenssi sf%. Käytetään apuna taulukkolaskentaa.



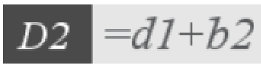
| | A x | B f | C suht_f | D sf | E suht_sf |
|---|---|---|---|---|---|
| = | | | | | |
| 1 | useita kertoja viikossa | 2106 | 2.30013 | 2106 | 2.30013 |
| 2 | noin kerran viikossa | 3388 | 3.70031 | 5494 | 6.00044 |
| 3 | harvemmin | 17213 | 18.7997 | 22707 | 24.8001 |
| 4 | ei lainkaan | 68853 | 75.1999 | 91560 | 100. |
| 5 | | | | | |
| 6 | otoskoko | 91560 | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |

B6 $=\mathrm{sum}(b1{:}b4)$

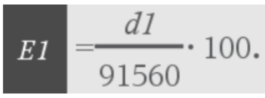
Sarakkeissa käytetyt laskukaavat (soluviittaukset):



C1 $=\dfrac{b1}{91560}\cdot 100.$



D2 $=d1+b2$



E1 $=\dfrac{d1}{91560}\cdot 100.$

c) Oletetaan, että 8. tai 9. luokan opiskelija on joutunut kiusaamisen uhriksi lukukauden aikana, jos hän on vastannut kysymykseen ”Kuinka usein sinua on kiusattu koulussa tämän lukukauden aikana?” joko ”useita kertoja viikossa”, ”noin kerran viikossa” tai ”harvemmin”. Näin ollen yhteenlaskusäännön nojalla:

$$P(opiskelijaa\ kiusattu) = 0{,}0230013 + 0{,}0370031 + 0{,}187997$$

$$= 24{,}8001 \approx 0{,}2480.$$

Siis opiskelija on joutunut lukuvuoden aikana kiusaamisen uhriksi todennäköisyydellä 0,2480. Toisin sanoen 24,80 %:a 8. ja 9. luokan opiskelijoista on kiusattu kuluneen lukuvuoden aikana.

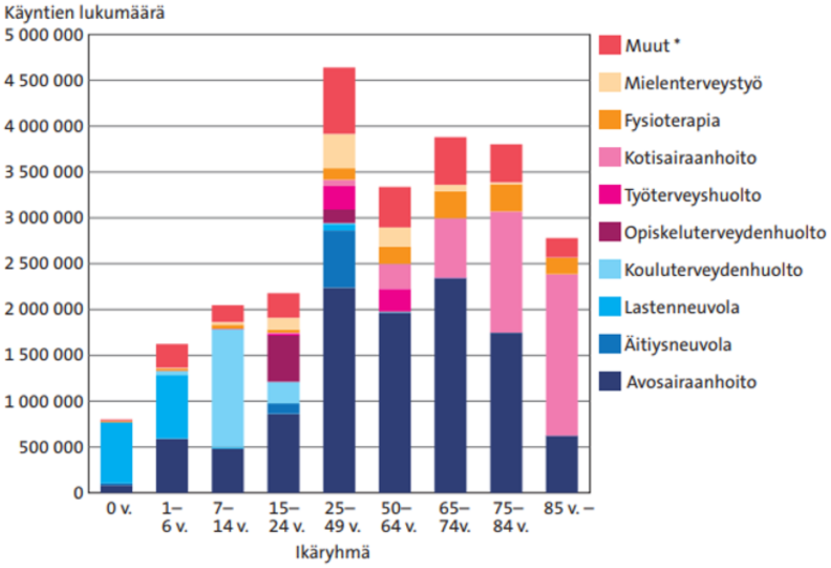
# TEHTÄVÄ 10: TERVEYSPALVELUIDEN KÄYTTÖ ERI IKÄRYHMISSÄ

Tutkitaan iän ja terveyskeskuskäyntien välistä yhteyttä. Alla olevassa kuviossa on esitetty perusterveydenhuollon käynnit** palvelumuodoittain ja ikäryhmittäin vuonna 2017 (Lähde: Suomalaisten sosiaali- ja terveyspalveluiden käyttö tilastojen valossa).

a) Laadi pylväskuvion pohjalta taulukko perusterveydenhuollon käyntien lukumääristä eri ikäryhmissä 100 000 käynnin tarkkuudella. Muodosta taulukon pohjalta hajontakuvio.

b) Sovita aineistoon lineaarinen malli ja toisen asteen polynominen malli.

c) Määritä mallien selitysasteet ja lineaarisesta mallista iän ja terveyskeskuskäyntien välinen korrelaatiokerroin. Tulkitse riippuvuuden voimakkuutta ja pohdi tulosta mahdollisesti vääristäviä tekijöitä.



* Sisältää päihdetyön, perhesuunnittelu-, ehkäisy-, kasvatus- ja perheneuvolapalvelut sekä muut neuvolapalvelut, seulonnat ja joukkotarkastukset, apuvälinepalvelut, puhe- toiminta-, jalka- ja ravitsemusterapian, terveyssosiaalityön, muut kuntoutus- ja erityisterapiat, päivätoiminnan, päiväsairaalatoiminnan sekä muun palvelutoiminnan.

** Perusterveydenhuollon avohoidon käynnit vastaanotolla, kotikäynnit, työpaikkakäynnit ja sairaalakäynnit.

Alle kouluikäisillä lapsilla lastenneuvolapalvelut ovat yleisin palvelumuoto. Työikäiset käyttävät eniten avosairaanhoidon ja työterveyshuollon palveluja sekä äitiysneuvolapalveluja. Vanhimmissa ikäryhmissä avosairaanhoito ja kotisairaanhoito ovat merkittävimmät palvelumuodot.

## PSST!

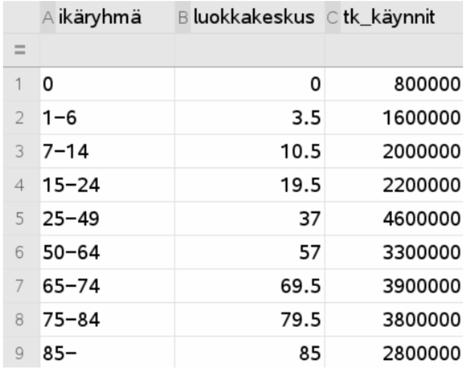
Suomalaisten terveys on parantunut viime vuosikymmeninä paljon, mutta siitä huolimatta terveyserot eivät ole kaventuneet eri väestöryhmien välillä. Muun muassa henkilön ikä, sukupuoli, asuinalue, siviilisääty, äidinkieli ja sosioekonominen asema vaikuttavat kukin osaltaan yksilön terveyteen, mutta myös terveyteen väestötasolla luoden eriarvoisuutta väestöryhmien välille. Esimerkiksi terveys- ja toimintakyky heikkenevät iän myötä, miehet sairastavat naisia enemmän sydän- ja verisuonitauteja, asuinalueiden erilaiset elinolosuhteet (kuten työllisyystilanne, elämänmeno, sosiaali- ja terveyspalveluiden saavutettavuus) heijastuvat kansantautien sairastavuuteen, parisuhteessa elävien koettu terveys on parempi ja elinikä pidempi kuin yksineläjillä ja suomenruotsalaiset ovat suomenkielisiä terveempiä. Mitä parempi on henkilön sosioekonomisen asema, sitä terveempi ja hyvinvoivempi hän todennäköisesti on. Vähäinen koulutus, epävarma ja heikko ammatillinen asema (pienituloisuus) sekä vähäinen varallisuus ovat vahvasti yhteydessä toisiinsa sekä huonon terveyden ja pahoinvoinnin kasaantumiseen. Erot terveyskäyttäytymisessä, kuten tupakoinnissa ja alkoholinkäytössä, selittävät arviolta noin puolet terveyden sosioekonomisista eroista.

Sosioekonomiseen asemaan liittyvät terveyserot ovat monimutkainen toisiinsa linkittyvien syiden ja seurausten vyyhti. Yhteiskunnalliset rakenteet vaikuttavat sosioekonomisen aseman muotoutumiseen ja terveyttä heikentävän sosiaalisen eriarvoisuuden syntyyn. Kuitenkin yksi suomalaisen terveyspolitiikan perusperiaatteista on tarjota jokaiselle terveydentilan edellyttämät riittävät ja laadukkaat palvelut sosioekonomisesta asemasta, taloudellisista edellytyksistä tai asuinalueesta riippumatta. Käytännössä erot terveysosaamisessa sekä terveyteen liittyvissä tiedoissa ja taidoissa aiheuttavat sen, että eri sosioekonomisessa asemassa olevat henkilöt eivät ole yhdenvertaisia palveluihin hakeuduttaessa. Lisäksi esim. asuinalue ja ikä vaikuttavat palveluiden saatavuuteen ja saavutettavuuteen. Edellä mainituista syistä perusterveydenhuollon palveluiden käyttö vaihtelee eri väestöryhmissä.

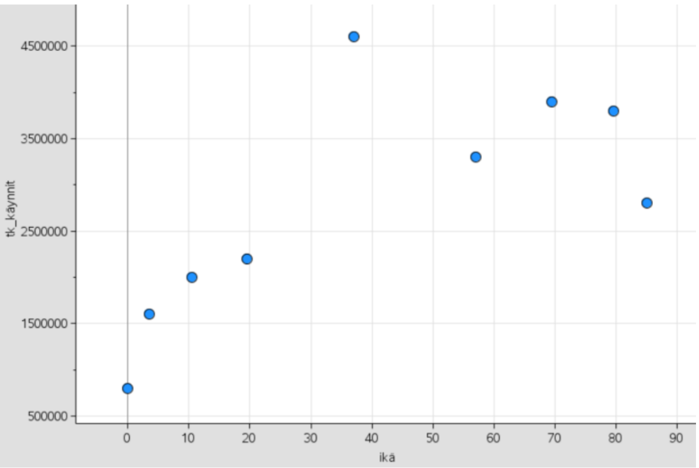
# Eräs malliratkaisu

Tässä ratkaisussa on käytetty TI-Nspire CX CAS -ohjelmistoa.

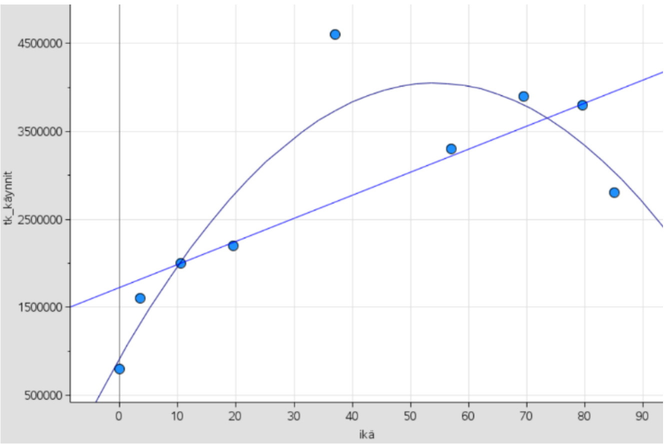
a) Laaditaan pylväsdiagrammin pohjalta taulukko, josta nähdään, paljonko käyntejä kussakin ikäryhmässä on. Ikäryhmät ovat itse asiassa luokkia, joiden luokkakeskukset saadaan laskettua luokan todellisen alarajan ja todellisen ylärajan keskiarvona. Merkitään siis taulukkoon ikäryhmät, niiden luokkakeskukset (ikä-sarake) hajontakuvion laatimista helpottamaan sekä terveyskeskuskäynnit (tk_käynnit-sarake) kyseisessä ikäryhmässä 100 000 käynnin tarkkuudella ja muodostetaan hajontakuvio:



| | A ikäryhmä | B luokkakeskus | C tk_käynnit |
|---|---|---|---|
| = | | | |
| 1 | 0 | 0 | 800000 |
| 2 | 1–6 | 3.5 | 1600000 |
| 3 | 7–14 | 10.5 | 2000000 |
| 4 | 15–24 | 19.5 | 2200000 |
| 5 | 25–49 | 37 | 4600000 |
| 6 | 50–64 | 57 | 3300000 |
| 7 | 65–74 | 69.5 | 3900000 |
| 8 | 75–84 | 79.5 | 3800000 |
| 9 | 85– | 85 | 2800000 |



b) Piirretään lineaarinen malli (regressiosuora) ja toisen asteen polynominen malli:



Lineaarinen malli (suora)

$$y = 6156{,}8x + 1{,}72715 \cdot 10^6$$

2. asteen polynominen malli (paraabeli)

$$y = -1063.49x^2 + 115371x + 920146$$

c) Iän ja terveyskeskuskäyntien väliseksi korrelaatiokertoimeksi saadaan laskinohjelmiston avulla r=0,711371. Näin ollen iän ja terveyskeskuskäyntien välinen korrelaatio on siis huomattava (ks. korrelaatiokertoimen tulkintataulukko, MAOL). Lisäksi koska korrelaatio on positiivinen, niin voidaan todeta, että iän kasvaessa myös terveyskeskuskäyntien määrä lisääntyy.

Selitysaste on korrelaatiokertoimen neliö eli $r^2$=0,506048 eli noin 50,6%. Voidaan siis sanoa lineaarisen mallin mukaan iän selittävän 50,6% terveyskeskuskäyntien määrän vaihtelusta. Laskinohjelmistolla saadaan toisen asteen polynomisen mallin selitysasteeksi $r^2$=0,837506 eli noin 83,8%, joka selittää jo yli neljä viidesosaa terveyskeskuskäyntien määrän vaihtelusta. Selitysasteiden perusteella toisen asteen polynominen malli sopii kuvaamaan aineistoa lineaarista mallia paremmin. Toisen asteen mallissa terveyskeskuskäyntien määrä on huipussaan noin 50-vuotiailla, jonka jälkeen käyntien määrä vähenee, kun taas lineaarisessa mallissa käyntien määrä kasvaa tasaisesti iän karttuessa.

Saatua tulosta saattaa vääristää tehtävänannossa asetettu terveyskeskuskäyntien määrän suurpiirteinen tarkkuus sekä etenkin aineiston ikäryhmä-luokkien eri laajuudet: luokkavälit ovat eripituisia, mikä on hyvien laskukäytänteiden vastaista. Tuntuu luonnolliselta, että luokkaväliltään pisimmällä ikävälillä 25-49-vuotiaissa myös terveyskeskuskäyntejä on eniten. Jos kyseinen luokka olisi jaettu kahdeksi ikäryhmäksi, on mahdollista, että lineaarinen malli olisi kuvannut aineistoa toisen asteen mallia paremmin, sillä muiden datapisteiden perusteella vaikuttaisi siltä, että terveyskeskuskäyntien määrä lisääntyy iän myötä.

# TEHTÄVÄ 11: ROKOTTEEN ROOLI TAUDIN LEVIÄMISEN EHKÄISYSSÄ

Yksittäiset tautitartunnat puhkeavat populaatiossa tautiaalloksi tartunnan saaneiden määrän kasvaa. Jos jokainen sairastunut tartuttaa enemmän kuin yhden ihmisen, tauti alkaa levitä populaatiossa eksponentiaalisesti. Kuinka suuresta tautiaallosta populaatiossa on kyse, riippuu siitä, kuinka monta kukin sairastunut tartuttaa.

Oletetaan, että tauti X on erittäin helposti tarttuva ja jokaista sataa ei-immuunia ihmistä kohden 96 sairastuu kohdatessaan tautia sairastavan henkilön. Oletetaan myös, että tautiin X kehitetty rokote on 95 %:n varma eli 95 ihmistä 100:sta ovat täysin immuuneja taudin tarttumiselle.

a) Oletetaan, että ensimmäinen tautiin sairastunut henkilö on tekemisissä kahdenkymmenen terveen kanssa. Tutki, montako ihmistä yksi sairastunut keskimäärin altistaa taudin tartunnalle. Pohdi, pääseekö tauti leviämään populaatiossa epidemiaksi jokaisen populaation jäsenen otettua rokotteen.

b) Tutki a-kohdan oletuksin tilannetta, jossa 5 % populaatiosta on jättänyt rokotteen ottamatta.

c) Tutkitaan taudin X leviämistä rokottamattomien, ei-immuunien ihmisten n-henkisessä ketjussa sairastuneelta taudinkantajalta S terveelle, ketjun viimeiselle henkilölle $T_n$, kun S tapaa ensin henkilön $T_1$, sitten $T_1$ tapaa henkilön $T_2$ ja niin edelleen, kunnes jonon toiseksi viimeinen, n-1:s henkilö $T_{n-1}$ kohtaa viimeisen henkilön $T_n$. Oletetaan kohtaamisaikavälien olevan tarttumisen kannalta suotuisat. Vertaile 100 %-rokottautuneen yhteisön tilannetta täysin rokottautumattoman yhteisön tilanteeseen tutkimalla henkilön $T_n$ tartunnan saamisen todennäköisyyttä eri jonon pituuksilla n=5, n=20 ja yleisessä tapauksessa n.

## PSST!

Taudin yleisellä leviämiskyvyllä kuvataan sitä, kuinka monta uutta henkilöä yksi sairastunut keskimäärin tartuttaa ympäristössään. Esimerkiksi tuberkuloosipotilas tartuttaa keskimäärin 0,5 henkilöä ja tuhkarokkopotilas 18. Tätä leviämiskyvyn ilmoittamisessa käytettyä lukua kutsutaan $R_0$-luvuksi, joka on sitä suurempi, mitä helpommin tauti leviää. $R_0$-luvun ollessa pienempi kuin 1 tauti ei yleensä aiheuta epidemioita.

Käytännössä siihen, kuinka paljon uusia tapauksia kukin sairastunut tartuttaa ja kuinka tehokkaasti tauti leviää yhteisössä, vaikuttaa myös ihmisyhteisön tiiviys, ihmisten käyttäytyminen ja minkä pituisesta tartuttavavasta vaiheesta on kyse. Tietyissä yhteisöissä, kuten perheissä, päiväkodeissa ja puolustusvoimien kasarmeissa, leviävät herkästi sellaisetkin taudit, joiden leviämiskyky muuten on vähäinen esim. kausi-influenssat.

**Lisää aiheesta:**

Infektioiden tartunta, taudin synty ja leviäminen, Duodecim.

# Eräs malliratkaisu

a) Kaikki yhteisön jäsenet ovat rokotettuja. Tautiin sairastunut henkilö tapaa siis 20 ihmistä. Rokotteen vuoksi 95 % heistä ovat immuuneja, joten tauti voi teoreettisesti tarttua vain 0,05·20=1 henkilöön. Todennäköisyys, että tauti tarttuu tähän ei-immuuniin henkilöön, on 0,96 eli taudin tarttumisen todennäköisyys on huomattavan suuri. Tauti X ei kuitenkaan pääse puhkeamaan tautiaalloksi ja häviää, sillä jokainen tartunnan saanut henkilö tartuttaa keskimäärin vain yhden ihmisen eikä aina sitäkään.

b) Tutkitaan tapausta, jossa 5 % populaatiosta ei ota rokotetta. Tauti voi potentiaalisesti tarttua rokotettuihin ei-immuuneihin henkilöihin, joita on populaatiosta

$$100\% - 95\% = 5\%$$

sekä rokottamattomiin ihmisiin, joita on populaatiosta 5 %, eli yhteensä

$$5\% + 5\% = 10\%$$

eli kymmenesosaan populaatiosta. Kun ensimmäisenä tartunnan saanut tapaa 20 ihmistä, voi tauti yhteensä siis tarttua

$$0{,}1 \cdot 20 = 2$$

eli kahteen henkilöön. Nyt todennäköisyys, että tauti tarttuu molempiin, on

$$0{,}96 \cdot 0{,}96 = 0{,}9216.$$

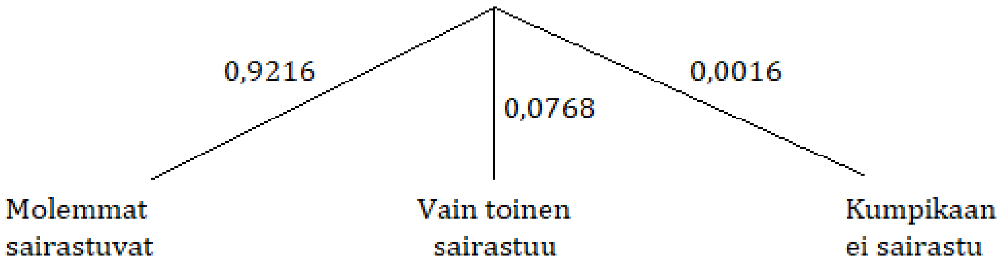
Voidaankin siis sanoa, että noin 92 %:n todennäköisyydellä sairastunut henkilö tartuttaa kaksi henkilöä. Todennäköisyys, että tauti tarttuu vain toiseen kahdesta, on jaettavissa kahteen tapaukseen. Joko tauti tarttuu henkilöön A eikä tartu henkilöön B tai tauti ei tartu henkilöön A ja tarttuu henkilöön B. Merkitään *A="tauti tarttuu vain henkilöön A"* ja *B="tauti tarttuu vain henkilöön B"*. Tapaukset A ja B ovat selvästi toisensa poissulkevat, joten saadaan

$$\begin{aligned} P\big((A \cap B^C) \cup (A^C \cap B)\big) &= P(A)P(B^C) + P(A^C)P(B) \\ &= 0{,}96 \cdot 0{,}04 + 0{,}04 \cdot 0{,}96 \\ &= 2 \cdot 0{,}04 \cdot 0{,}96 \\ &= 0{,}0768. \end{aligned}$$

Siis todennäköisyys sille, että vain toinen saisi taudin on noin 8 %. Todennäköisyys, ettei kumpikaan saa tautia on

$$0{,}04 \cdot 0{,}04 = 0{,}0016.$$

Täten todennäköisyys, ettei kumpikaan saisi tartuntaa on todella pieni, vain noin 2 ‰. Havainnollistetaan laskettuja tartunnan todennäköisyyksiä puukaaviolla:

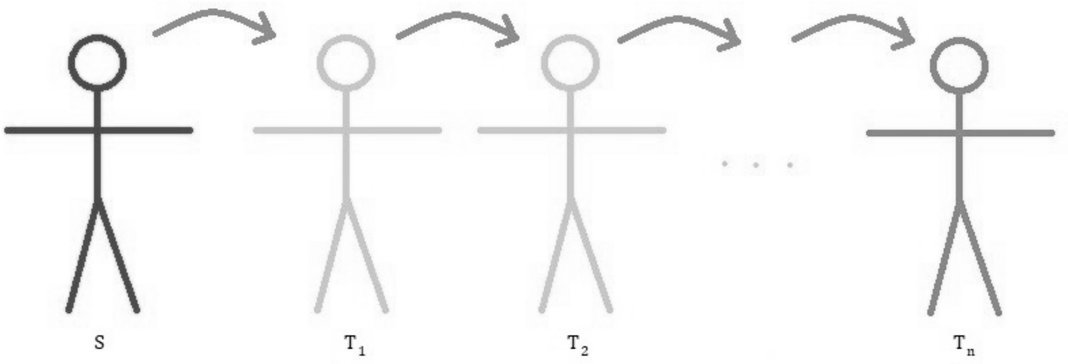


On siis selvästi todennäköisintä, että yksi sairastunut tartuttaa kaksi henkilöä ja tämä riittää tartunta-aallon puhkeamiseen. Nimittäin keskimäärin jokainen sairastunut tartuttaa 1,92 (0,96·2) ihmistä. Vaikka rokote tautiin X on tehokas suojaten 95 % ottajistaan, niin tämän tehtävän tilanteessa, taudin ollessa erittäin tarttuva, laumaimmuniteettia ei saavuteta vielä sillä, että 95 % populaatiosta ottaa rokotteen.

### HUOMAUTUS!

Tehtävän havainnollistus taudilla X on hieman kärjistetty. Todellisuudessa useimmilla tartuntataudeilla laumaimmuniteetti saavutetaan, kun noin 95 % populaatiosta ottaa rokotteen, sillä suurin osa tartuntataudeista ei ole tartuttavuudeltaan taudin X tasoa. Viruksen aiheuttamat rokot ovat taudeista helpoimmin tarttuvia, esimerkiksi tuhkarokon tarttuvuus on noin 90 %:n luokkaa.

c) Piirretään tehtävän havainnollistukseksi kuva tartuntaketjusta:



**Tutkitaan ensin tilannetta, jossa kaikki jonon henkilöt ovat rokottamattomia**. Jonon ensimmäiseen henkilöön tauti X tarttuu todennäköisyydellä 0,96. Jotta tauti tarttuisi jonon viimeiseen henkilöön $T_n$ (jonon n:s henkilö), sen täytyy tarttua hänen lisäkseen jokaiseen henkilöön kohtaamisten ketjussa ennen häntä. Siis todennäköisyys, että tauti tarttuu henkilöön $T_n$ on

$$P(T_n\ sairastuu) = P(T_1 sairastuu\ ja\ T_2\ sairastuu\ ja\ \ldots\ ja\ T_n\ sairastuu)$$

$$= \prod_1^n P(T_i\ sairastuu\ )$$

$$= 0{,}96 \cdot 0{,}96 \cdot \ldots \cdot 0{,}96$$
$$= 0{,}96^n.$$

Siis n-henkisessä tartuntaketjussa todennäköisyys, että tauti tarttuu jonon viimeiseen henkilöön on $0{,}96^n$. Näin ollen todennäköisyys on sitä pienempi, mitä pidempi jono on. Kun jono on viisihenkinen, todennäköisyys että tauti X tarttuu jonon viimeiseen henkilöön, on

$$P(T_5\ sairastuu) = 0{,}96^5 = 0{,}815372 \ldots \approx 0{,}8154.$$

Vastaavasti, kun jono on 20-henkinen, saadaan

$$P(T_{20}\ sairastuu) = 0{,}96^{20} = 0{,}442002 \ldots \approx 0{,}4420.$$

Sekä viisi- että 20-henkisessä jonossa tartunnan todennäköisyys on täten huomattavan suuri.

**Tarkastellaan sitten tilannetta, jossa kaikki jonon henkilöt ovatkin rokotettuja.** Tällöin

$$P(T_n\ sairastuu) = 0{,}05^n, \text{jolloin}$$

$$P(T_5\ sairastuu) = 0{,}05^5 = 0{,}000000312 \text{ ja}$$

$$P(T_{20}\ sairastuu) = 0{,}05^{20} = 9{,}5..\cdot 10^{-27} \approx 0.$$

Siis tartunnan leviäminen jonossa, jossa kaikki ovat rokotettuja, on hyvin epätodennäköistä viisihenkisessä jonossa ja lähes mahdotonta 20-henkisessä jonossa.

### HUOMAUTUS!

Tässä ketjuesimerkissä huomataan hyvin, että sairastunut henkilö tartuttaa muita myös epäsuorasti. Todellisuudessa tartuntaketjut muodostavat tartuntaverkkoja sairastuneen henkilön tavatessa useita ihmisiä ja toisaalta myös näiden henkilöiden taas tavatessa muita ihmisiä.

# Tehtävä 12: Tautitestin luotettavuustarkastelu

Moni virus, kuten tässä tehtävässä tutkittava HI-virus, voidaan havaita pääosin terveyskeskuksissa ja sairaaloissa tehtävillä testeillä. Tällaisia ovat esimerkiksi verinäytteestä tehtävät vasta-aine- ja antigeenitestit.

Tämä tehtävä perustuu Richard Isaacin kirjassaan *The Pleasures of Probability* (s. 31-34) esittämään esimerkkiin HIV-testin luotettavuuden tutkimisesta. Esimerkissä testi on viruksen havaitsemiseksi sikäli hyvä, että jos henkilöllä on virus, sen havaitsemisen todennäköisyys verikokeella on korkea; suuressa ryhmässä, jonka tiedetään koostuvan vain tartunnan saaneista, testi antaa oikean, positiivisen tuloksen 95 prosentilla ryhmästä. Tässä yhteydessä puhutaan testin *sensitiivisyydestä*, joka edeltävässä tapauksessa on 0.95. Merkitään todennäköisyyttä, että testi on positiivinen henkilön ollessa taudin kantaja, seuraavasti

P(testi positiivinen | henkilöllä tauti) = 0.95.

Testi siis vaikuttaa intuitiivisesti hyvin luotettavalta. Seuraavaksi keskitytään kuitenkin tutkimaan mahdollisia virhetuloksia, jotka ovat

P(testi positiivinen | henkilöllä ei ole virusta) eli väärä positiivinen tulos ja

P(testi negatiivinen | henkilöllä on virus) eli väärä negatiivinen tulos.

Oletetaan, että HI-virusta kantaa väestössä kolme ihmistä tuhannesta. Jos ihmisellä on HI-virus, tietty testi ilmoittaa 95 % varmuudella, että ihmisellä on kyseinen virus. Jos taas ihmisellä ei ole virusta, testi ilmoittaa 95 % varmuudella, että ihmisellä ei ole virusta. Oletetaan myös sekä väärien positiivisten että väärien negatiivisten testituloksien osuuden olevan kummankin 5 %. Millä todennäköisyydellä

a) testitulos on positiivinen,

b) testitulos on negatiivinen,

c) ihmisellä on virus, jos tiedetään testituloksen olevan positiivinen,

d) ihmisellä ei ole virusta, jos tiedetään testituloksen olevan positiivinen? Pohdi testin luotettavuutta.

e) Tarkastellaan tilannetta Etelä-Karjalassa, jossa oli vuonna 2017 Suomen korkein HIV-ilmaantuvuus (6,1/100 000). Oletetaan muutoin tehtävänannon tiedot testistä oikeiksi. Millä todennäköisyydellä positiivisen testituloksen saaneella henkilöllä ei ole virusta?

### PSST!

HIV on krooninen autoimmunologinen sairaus, jonka aiheuttaa HI-virus. Viruksen aiheuttamille haitoille ei toistaiseksi ole löydetty parannuskeinoa, mutta lääkehoidoilla pystytään pitämään viruksen aiheuttamat vauriot hallinnassa. Taudin varhainen toteaminen on tärkeää, jotta hoito toteutuu parhaalla mahdollisella tavalla eikä infektoituneen elimistön puolustuskyky laske viruksen tuhotessa valkosoluja, jotka suojaavat elimistöä taudinaiheuttajilta. Hoitamattomana HIV-infektio voi edetä pahimmillaan henkeä uhkaavaan AIDS-vaiheeseen, jossa elimistön puolustusjärjestelmä heikentyy vakavasti. HI-viruksen aiheuttama tartunta voidaan havaita verinäytteestä tehtävällä vasta-ainetestillä.

**Lisää aiheesta:**

HIV sairautena, Hivpoint (Hiv-säätiö).

# ERÄS MALLIRATKAISU

a) Merkitään seuraavasti

V = ”ihmisellä on virus”, E = $V^C$= ”ihmisellä ei ole virusta”,

P = ”testi on positiivinen”, N = $P^C$ = ”testi on negatiivinen”.

Halutaan siis selvittää $P(P)$.

Tunnetaan todennäköisyydet

$$P(V) = \frac{3}{1000} = 0{,}003$$

$$P(E) = 1 - 0{,}003 = 0{,}997$$

$$P(P|V) = P(N|E) = 0{,}95$$

$$P(N|V) = P(P|E) = 1 - 0{,}95 = 0{,}05.$$

Sijoittamalla tunnetut todennäköisyydet kokonaistodennäköisyyskaavaan (Lause 2) saadaan, että testitulos on positiivinen todennäköisyydellä

$$\begin{aligned} P(P) &= P(P|V)P(V) + P(P|E)P(E) \\ &= 0{,}95 \cdot 0{,}003 + 0{,}05 \cdot 0{,}997 \\ &= 0{,}0527. \end{aligned}$$

Koska todellisuudessa HIV-virusta kantaa 0,3 % väestöstä, huomataan testin antavan positiivisia tuloksia enemmän kuin taudinkantajia todellisuudessa on.

b) Tarkastellaan negatiivisen testituloksen todennäköisyyttä

$$P(N) = P(N|V)P(V) + P(N|E)P(E)$$

$$= 0{,}05 \cdot 0{,}003 + 0.95 \cdot 0{,}997$$

$$= 0{,}9473.$$

Tähän oltaisiin päästy yhtä lailla vastatapahtuman todennäköisyyden avulla (komplementtikaava). Tuloksesta huomataan testin antavan viruksen todelliseen esiintyvyyteen verrattuna liian vähän negatiivisia tuloksia.

c) Lasketaan seuraavaksi Bayesin kaavalla (Lause 3) ehdolliset todennäköisyydet käänteistapauksille P(V|P) ja P(E|P). Positiivisen testituloksen saanut henkilö on viruksen kantaja todennäköisyydellä

$$P(V|P) = \frac{P(P|V)\,P(V)}{P(P)} = \frac{0{,}95 \cdot 0{,}003}{0{,}0527} = 0{,}05407.. \approx 0{,}05.$$

Siis vain 5 % saaduista positiivisista tuloksista tarkoittaa, että testin tehneellä henkilö on todellisuudessa viruksen kantaja.

d) Vastaavasti positiivisen testituloksen saaneella henkilöllä ei ole virusta todennäköisyydellä

$$P(E|P)=\frac{P(P|E)\,P(E)}{P(P)}=\frac{0{,}05\cdot 0{,}997}{0{,}0527}=0{,}94592..\approx 0.95.$$

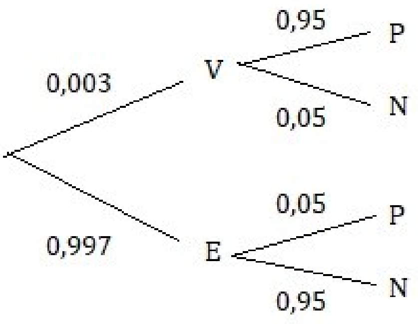
Samaan tulokseen päästäisiin myös komplementtikaavalla

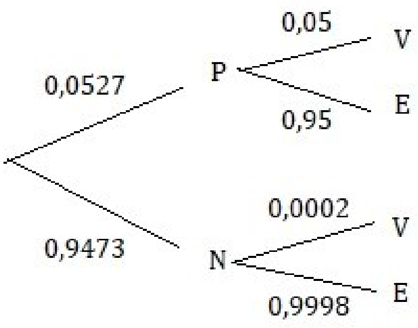
$$P(E|P)=1-P(V|P)=1-0{,}05=0{,}95.$$

Siis 95 % positiivisista testituloksista tulee virheellisesti terveille henkilöille. Huomataan, että intuitiivisesti luotettavalta vaikuttanut HIV-testi osoittautuikin huonoksi tavaksi seuloa tartuntoja. Testin huono luotettavuus johtuu virheellisten positiivisten tulosten suuresta määrästä, joka taas johtuu siitä, että vaikka virheellisten tulosten todennäköisyys on pieni (0.05), niin virheellisiä positiivisia tuloksia syntyy runsaasti, koska terveiden ihmisten osuus on suuri (0.997).

## HUOMAUTUS:

Vastaavasti saataisiin laskettua myös todennäköisyydet tapauksille, joissa negatiivisen testituloksen saanut henkilö on viruksen kantaja tai terve henkilö, ja saataisiin P (V|N) ≈ 0,0002 ja P (E|N) ≈ 0,9998. Havainnollistetaan laskettuja ehdollisia todennäköisyyksiä puukaaviolla:





e) Etelä-Karjalassa 2017 positiivisen testituloksen saaneella henkilöllä ei ole virusta todennäköisyydellä

$$P(E|P)=\frac{P(P|E)\,P(E)}{P(P)}=\frac{P(P|E)\,P(E)}{P(P|V)P(V)+P(P|E)P(E)}=\frac{0{,}05\cdot 0{,}999939}{0{,}95\cdot 0{,}000061+0{,}05\cdot 0{,}999939}\approx 0{,}99.$$

Tässä testi siis vaikuttaisi vieläkin epäluotettavammalta. Todellisuudessa HI-viruksen havaitsemisessa vasta-ainetestillä oleellinen luotettavuustekijä on se, kuinka kauan mahdollisesta tartuntahetkestä on kulunut aikaa. Jos HIV-vasta-ainetta ei ole kolmessa kuukaudessa tartunnasta kehittynyt, voi olla varma, ettei ole viruksen kantaja eli negatiivinenkin testitulos on vasta tuolloin varmasti luotettava. Yleensä vasta-aineet kehittyvät kuitenkin muutamassa viikossa tartunnan jälkeen.