

# Terveytemme – Termisanasto ja tilastolliset menetelmät

- Termisanasto
- Tilastolliset menetelmät
- Lähdeviitteet

## Termisanasto

<b>Elinajanodote</b>	Elinajanodote on keskimäärin jäljellä oleva elinaika. Vastasyntyneen elinajan odote, $e_0$ , on tietynä aikana syntyneiden henkilöiden eli syntymäkohortin kokonaiselinaika jaettuna näiden henkilöiden määrällä. Vastaavalla tavalla saadaan keskimäärin jäljellä oleva elinaika muun ikäisille laskemalla ikärajan saavuttaneiden yksilöiden jäljellä oleva kokonaiselinaika ja jakamalla se näiden henkilöiden määrällä. Käytännössä elinajan odote lasketaan yleisimmin periodilukuna jonkin ajanjakson ikäryhmittäisen kuolleisuuden perusteella. Tällöin luku ilmoittaa sen keskimääräisen elinajan pituuden, joka tietynikäisillä henkilöillä olisi odotettavissa, jos heidän kuolleisuutensa kussakin iässä olisi sama kuin ao. ikäryhmään kuuluneilla henkilöillä on ollut laskelman pohjana olevalla jaksolla.
<b>Esiintyvyys</b>	Katso vallitsevuus.
<b>Ilmaantuvuus</b>	Ilmaantuvuudella tarkoitetaan tietynä ajanjaksona ilmaantuvien uusien tapausten määrää jaettuna tarkasteltavan väestön koolla. Kuolleisuus (mortaliteetti) ja sairastuvuus (morbiditeetti) ovat yleisimmin käytetyt ilmaantuvuusluvut.
<b>Ikävakiointi</b>	Ikävakiointi on väestötutkimuksessa käytettävä menetelmä, jonka avulla eri väestöryhmien vertailu on mahdollista siten, että erot väestöryhmien ikärakenteissa eivät vääristä tuloksia. Yleisesti käytettyjä menetelmiä ovat ns. suora ikävakiointi ja epäsuora ikävakiointi.
<b>Kohortti</b>	Kohortti on jokin määritelty joukko henkilöitä. Yleensä termiä kohortti käytetään joukoille, joita seurataan ajassa eteenpäin. Syntymäkohortilla tarkoitetaan jonakin tietynä ajanjaksona syntyneiden henkilöiden joukkoa.
<b>Luottamusväli</b>	Tunnusluvun luottamusväli on sellainen havaintoaineistosta laskettu väli, että se sisältää tunnusluvun todellisen arvon valitulla todennäköisyydellä. Usein käytetty 95 %:n luottamusväli sisältää tunnusluvun todellisen arvon 95 %:n todennäköisyydellä. Mitä pienempi otos tai tutkittava joukko on, sitä suurempaa on myös sattuman aikaansaama vaihtelu. Leveä luottamusväli viittaa siihen, ettei tutkimuksen otoskoko riitä tutkittavan ilmiön luotettavaan kuvaamiseen. Katso tilastollinen merkitsevyys.
<b>Mallivakiointi</b>	Mallivakiointi korjaa erilaisten ikäjakaumien (ja/tai muiden sekoittavien tekijöiden) vaikutuksen käyttämällä hyväksi tilastollista regressiomallia. Mallin avulla lasketaan jokaiselle yksilölle ennustearvo, ja näiden ennustettujen arvojen keskiarvo on mallivakioitu keskiarvo, vallitsevuus tai muu tunnusluku. Katso myös vakiointi ja ikävakiointi.
<b>PYLL eli potentiaalisesti menetetyt elinvuodet</b>	PYLL-indeksi on kansainvälisesti laajalti käytetty hyvinvoinnin mittari. PYLL kuvaa tietyllä ikävälillä tapahtuneiden kuolemien takia menetettyjen elinvuosien lukumäärää väestössä 100 000 asukasta kohti. Se korostaa nuorella iällä tapahtuvien kuolemantapausten merkitystä. Laskettaessa PYLL-tunnusluku ikävälille 25–80 vuotta, esim. 25-vuotiaana kuolleen arvioidaan menettäneen 55 elinvuotta, kun taas 79-vuotiaana kuollut on menettänyt vain yhden vuoden ja yli 80-vuotiaana kuolleiden ei katsota menettäneen yhtään elinvuotta. Tässä tapauksessa ei huomioida alle 25-vuotiaana kuolleiden menetettyjä elinvuosia.
<b>Satunnaisotos</b>	Tutkimusta varten väestöstä voidaan ottaa satunnaisotos, eli otos johon jokaisen kyseiseen väestöön kuuluvan valituksi tuleminen todennäköisyys tunnetaan. Satunnaisotoksen pohjalta voidaan arvioida väestön tunnuslukuja ja niiden

	tilastollinen tarkkuus. Yksinkertaisessa satunaisotoksessa jokaisella on yhtä suuri valituksi tulemisen todennäköisyys.
<b>Terveysosoitin</b>	Terveysosoittimilla tarkoitetaan tilastollisia tunnuslukuja (esim. päivittäin tupakoivien osuus, alkoholikuolleisuus), jotka ovat kansanterveyden kannalta merkityksellisiä.
<b>Tilastollinen merkitsevyys</b>	<p>Tilastollinen merkitsevyys liittyy todennäköisyyteen tehdä virhepäätelmä, jos havaintoaineiston perusteella todetaan, että löydös poikkeaa ennalta asetetusta oletuksesta, niin sanotusta nollahypoteesista. Esim. verrattaessa kahta ryhmää tarkastellaan usein ryhmäkeskiarvoja, ja nollahypoteesina oletetaan, että ryhmien välillä ei ole eroa. Havaittujen ryhmäkeskiarvojen eroa pidetään tilastollisesti merkitsevänä 5%:n riskitasolla, jos nolla-hypoteesin vallitessa havaitun tai sitä suuremman keskiarvojen erotuksen todennäköisyys olisi alle 5%.</p> <p>Tilastollista merkitsevyyttä kuvataan usein p-arvolla, joka kertoo pienimmän merkitsevyydestason, jolla havaittu tulos on tilastollisesti merkitsevä. Mitä pienempi p-arvo on, sitä suurempi on tilastollinen merkitsevyys. Usein tulosta kutsutaan tilastollisesti merkitseväksi, jos p-arvo alittaa arvon 0,05.</p> <p>Se, että tulos ei ole tilastollisesti merkittävä, ei useinkaan tarkoita, että nollahypoteesi pitää paikkansa. Tilastollinen merkitsevyys nimittäin riippuu vaikutuksen suuruuden lisäksi myös havaintoaineiston koosta. Mitä pienempi otos tai tutkittava joukko, sitä suurempaa on sattuman aikaan saama vaihtelu. Tilastollinen merkitsevyys ei myöskään automaattisesti ole sama asia kuin ilmiön kansanterveydellinen merkittävyys. Esimerkiksi hyvin suurista tutkimusaineistoista saatetaan saada tilastollisesti merkitseviä tuloksia, joilla ei ole väestön terveyden kannalta merkitystä. Katso myös luottamusväli.</p>
<b>Sosioekonominen asema</b>	Sosioekonomisella asemalla viitataan yleensä sekä hyvinvoinnin aineellisiin ulottuvuuksiin (kuten tulot, omaisuus ja asumistaso) että niiden hankkimiseen tarvittaviin edellytyksiin (kuten koulutus, ammatti ja työllisyys). Sosioekonomisten väestöryhmien välillä on suuria terveyseroja.
<b>Terveyserot</b>	Väestöryhmien välisten terveyserojen syyt ovat monitasoisia. Ne kytkeytyvät väestöryhmien erilaisiin elinoloihin ja työoloihin, erilaiseen fyysiseen ja psykososiaaliseen kuormitukseen ja sukupolvienkin yli ulottuviin kulttuurisidonnaisiin käyttäytymismalleihin. Terveyseroja aiheuttaviin ja ylläpitäviin tekijöihin voidaan vaikuttaa yhteiskuntapolitiikan keinoin.
<b>Vakioiminen</b>	Tutkittaessa erilaisia väestöilmioita on nk. sekoittavien tekijöiden kuten iän tai sukupuolen vakioiminen on usein tarpeellista. Vakioinnilla kahden tai useamman väestön tunnusluvut muokataan niin, että iän tai sukupuolen (tai muiden sekoittavien tekijöiden) erilaiset jakaumat vertailtavissa väestöissä eivät vaikuta vertailun tulokseen. Esimerkiksi vakioidut kuolleisuusluvut osoittavat sen, minkälainen vertailtavien väestöjen kuolleisuus olisi, mikäli niiden väestöjen ikärakenteet olisivat samanlaiset. Katso ikä- ja mallivakiointi.
<b>Vallitsevuus</b>	Vallitsevuus eli esiintyvyys on suhdeluku, joka ilmaisee väestössä tietynä aikana niiden osuuden, joilla on määritelty ominaisuus (esim. sairaus). Vallitsevuus ilmaistaan yleensä tapauksina 10 000 tai 100 000 henkilöä kohti. Pistevallitsevuus tarkoittaa vallitsevuutta tietynä ajanhetkenä ja jaksovallitsevuus vallitsevuutta määritellyn ajanjakson aikana.

# Tilastolliset menetelmät

## Mallivakiointi

Mallivakiointi (Lee 1981, Graubard ja Korn 1999) korjaa erilaisten ikäjakaumien (ja/tai muiden sekoittavien tekijöiden) vaikutuksen käyttämällä hyväksi tilastollista regressiomallia. Terveystemme -raporteissa on käytetty lineaarista mallia jatkuvia muuttujia kuvaaville tunnusluvuille ja logistista mallia osuuksille. Aineistoon sovitetaan malli, jonka selittävinä muuttujina ovat ikä, sukupuoli ja muita tekijöitä, joiden sekoittava vaikutus halutaan eliminoida tai joiden arvojen välisiä vertailuja halutaan tehdä, kuten alue tai sosioekonominen luokka. Esim. verrattaessa sukupuolia naisten pitkäikäisyydestä johtuvat erilaiset ikäjakaumat vaikeuttavat sukupuolittaisten keskiarvojen tai vallitsevuuksien tulkintaa: vanhojen naisten suurempi sairastavuus voi aiheuttaa näennäisen eron miesten ja naisten välille. Mallivakioinnissa ajatellaan kaikkien henkilöiden olevan vuorollaan joko miehiä tai naisia, ja lasketaan mallin avulla jokaiselle yksilölle (tässä tapauksessa) kaksi ennustearvoa. Ikävakioidut tunnuslukujen arvot ovat näin saatujen ennustearvojen keskiarvo. Lopputuloksena saadut kaksi keskiarvoa tai vallitsevuutta ovat vertailukelpoisia keskenään, koska ikä- (ja muiden sekoittavien tekijöiden) jakaumat ovat samat.

## Suora ikävakiointi

Erilaiset ikäjakaumat vaikeuttavat usein ryhmien välistä vertailtavuutta. Suoran vakioinnin avulla voidaan poistaa erilaisten ikäjakaumien vaikutus verrattaessa jonkin ilmiön keskimääräistä tasoa (tai vallitsevuutta) eri vuosina tai eri väestöryhmissä. Tarkasteltavasta väestöstä tai väestöryhmästä lasketaan ikäluokakohtaiset keskiarvot (tai vallitsevuudet). Tämän jälkeen lasketaan painotettu keskiarvo ikäluokakohtaisista tuloksista käyttämällä painokertoimina ikäjakaumaa nk. standardiväestössä, joka voi olla esim. Suomen väestö vuonna 2007. Suuret ikäryhmät vaikuttavat tulokseen eniten, esim. ikäryhmä, jossa on 10% väestöstä, saa painoarvon 10% em. painotetussa keskiarvossa. Käytettävien painokertoimien summa on 100%.

## Elinajanodote

Elinajanodote on keskimäärin jäljellä oleva elinaika (esim. Pitkänen ja Sihvonen 2003). Yleisin elinajan pituutta kuvaava mittari on vastasyntyneellä keskimäärin odotettavissa oleva elinaika, joka ilmoitetaan usein erikseen kummallekin sukupuolelle. Vastasyntyneen elinajan odote,  $e_0$ , on tiettyä aikana syntyneiden henkilöiden eli syntymäkohortin kokonaiselinaika jaettuna näiden henkilöiden määrällä. Vastaavalla tavalla saadaan keskimäärin jäljellä oleva elinaika muun ikäisille laskemalla ikärajan saavuttaneiden yksilöiden jäljellä oleva kokonaiselinaika ja jakamalla se näiden henkilöiden määrällä. Käytännössä elinajan odote lasketaan yleisimmin periodilukuna jonkin ajanjakson ikäryhmittäisen kuolleisuuden perusteella. Tällöin luku ilmoittaa sen keskimääräisen elinajan pituuden, joka tietynikäisillä henkilöillä olisi odotettavissa, jos heidän kuolleisuutensa kussakin iässä olisi sama kuin ao. ikäryhmään kuuluneilla henkilöillä on ollut laskelman pohjana olevalla jaksolla.

Elinajanodote saadaan nk. elinajantaulusta, joka kuvaa kuinka elossa olevien määrä vähenee kuolleisuuden johdosta iän karttuessa. Elinajantaulu muodostuu iän mukaan lasketuista kuolleisuusfunktioista. Kuolemanvaaraluvut  ${}_nq_x$  ilmoittavat iässä  $x$  olevien kohortin jäsenten todennäköisyyden kuolla ikärajojen  $x$  ja  $x+n$  välillä ja eloonjäämistodennäköisyydet puolestaan todennäköisyyden pysyä hengissä iästä  $x$  ikään  $x+n$ .

Terveystemme -raporteissa esitetyt elinajanodotteet ja potentiaalisesti menetetyt elinvuodet (PYLL) on muodostettu Elinolot ja kuolleisuus -tutkimusaineistosta (EKSY). Aineisto on Tilastokeskuksen ja Helsingin yliopiston Sosiologian laitoksen ylläpitämä. Se on muodostettu yhdistämällä työssäkäyntitilastoon tietoja kuolinsyyrekisteristä. Aineisto sisältää 80% kaikista kuolemantapauksista vuosilta 1999-2003.

Osoittimien laskemisessa käytetään ikäluokitusta 0-24, 25-39, 40-49, 50-59, ..., 80-89 (yhdistetty elinajan taulu, jossa iät ovat  $x \in \{0, 25, 40, 50, 60, 70, 80, 90\}$  ja ikäväljen pituudet

$n \in \{25, 15, 10, 10, 10, 10, 10, \infty\}$  vuotta (Koskinen ym. 2007). Näiden avulla lasketaan kuolemanvaaraluvut  ${}_n q_{x_k} = {}_n d_{x_k} / R_{x_k}$  jokaiselle ikävälille  $x_k - (x_k + n_k - 1)$ , jossa  ${}_n d_{x_k}$  on kuolleiden ja  $R_{x_k}$  seurantavuoden alussa elossa olevien henkilöiden lukumäärä. Henkilön, joka on  $x_i$ -vuotias, eloonjäämistodennäköisyys  $x_j$ -vuotiaana ja kuolemistodennäköisyys ikävälillä  $[x_{j-1}, x_j)$  ovat

$${}_{x_j-x_i} S_{x_i} = \prod_{k=i}^{j-1} (1 - {}_n q_{x_k}) \text{ ja } {}_{x_j-x_i} f_{x_i} = {}_{x_{j-1}-x_i} S_{x_i} {}_n q_{x_{j-1}}.$$

Elinajanodote  $x_i$ -vuotiaalle lasketaan summana

$$e_{x_i} = \sum_{k=i}^K (x_k + {}_n a_{x_k}) {}_{x_{j+1}-x_i} f_{x_k} + {}_{x_K-x_i} S_{x_k} e_{x_K} - x_i,$$

${}_n a_{x_k}$  on ikävälillä  $[x_k, x_{k+1})$  kuolleiden keskimääräinen elinaika iän  $x_k$  jälkeen.  ${}_n a_{x_k}$  lasketaan Tilastokeskuksen vuoden 2001 elinajantaulun perusteella.  $x_K$ -vuotiaiden jäljellä olevien elinvuosien määrä  $e_{x_K}$  poimitaan samasta Tilastokeskuksen taulukosta.

Luottamusvälien laskemisessa huomioidaan EKSU-aineiston ositettu otanta, jossa kuolleista on poimittu 80% ja eloonjääneistä 11%. Kuolemanvaaralukujen estimoinnissa käytetään osuuksien estimoinnissa käytettävää menetelmää (Lohr 1999, esimerkki sivulla 102).

Eloojääneiden ositteessa kuolemanvaaraestimaatti on nolla ja kuolleiden ositteessa kuolleiden määrä jaettuna seurantavuoden alussa elossa olevien määrällä. Ositekohtaiset estimaatit

$\widehat{{}_n q_{x_k}}^h$  summataan yhteen painottamalla

$$\widehat{{}_n q_{x_k}} = \sum_{h=1}^2 \frac{N_h}{N} \widehat{{}_n q_{x_k}}^h.$$

Edellisessä  $N_h$  on ositteen  $h$  koko väestössä, jonka koko on  $N$ . Ositteesta  $h$  poimitaan  $n_h$  henkilöä. Varianssiestimaatit saadaan vastaavasti

$$\widehat{V}(\widehat{{}_n q_{x_k}}) = \sum_{h=1}^2 \left(1 - \frac{n_h}{N}\right) \left(\frac{N_h}{N}\right)^2 \frac{\widehat{{}_n q_{x_k}}^h (1 - \widehat{{}_n q_{x_k}}^h)}{n_h - 1}.$$

Summan sisällä ensimmäinen tulo termi kuvaa nk. äärellisen populaation korjausta, joka pienentää tässä tapauksessa varianssiestimaattia merkittävästi. Eri ikäryhmien kuolinvaaralukujen estimaatit voidaan olettaa riippumattomiksi, ja elinajanodotteen varianssiestimaatti voidaan johtaa helposti käyttämällä delta-menetelmää.

## PYLL (potentiaalisesti menetetyt elinvuodet)

PYLL-indeksi (esim. Vohlonen ym. 2007) on kansainvälisesti laajalti käytetty hyvinvoinnin mittari. PYLL kuvaa tietyllä ikävälillä tapahtuneiden kuolemien takia menetettyjen elinvuosien lukumäärää väestössä 100 000 asukasta kohti. Se korostaa nuorella iällä tapahtuvien kuolemantapausten merkitystä.

Laskettaessa PYLL-tunnusluku ikävälille 25–80 vuotta, esim. 25-vuotiaana kuolleen arvioidaan menettäneen 55 elinvuotta, kun taas 79-vuotiaana kuollut on menettänyt vain yhden vuoden ja yli 80-vuotiaana kuolleet eivät menetä yhtään elinvuotta. Tässä tapauksessa ei huomioida alle 25-vuotiaana kuolleiden menetettyjä elinvuosia. Usein käytetty ikärajaus 0-70 vuotta ei ole mielekäs tarkasteltaessa koulutusryhmiä, koska koulutusryhmä on selvä vasta yli 20 vuoden ikäisillä.

PYLL ikävälille  $[x_i, x_j)$  lasketaan samaan tapaan kuin elinajanodote:

$$\text{PYLL}_{(x_i, x_j)} = \sum_{k=i}^{j-1} (x_j - (x_k + n_k a_{x_k}))^{x_{j+1} - x_i} f_{x_k} .$$

## Lähdeviitteet

Dobson A, Kuulasmaa K, Eberle E, Scherer J. Confidence intervals for weighted sums of Poisson parameters. *Stat Med* 1991;10: 457-62.

Graubard, B. I. & Korn, E. L. (1999). Predictive Margins with Survey Data. *Biometrics*. Vol. 55. No. 2. 652-659.

Koskinen S. ym. (toim.) (2007). *Suomen väestö*. Gaudeamus.

Lee, J. (1981). Covariance Adjustment of Rates Based on the Multiple Logistic Regression Model. *Journal of Chronic Diseases*. Vol. 34. 415-426.

Lohr (1999) *Sampling: Design and Analysis*. Duxbury Press Pacific Grove, CA.

Pitkänen K., Sihvonen A-P (2003). Johdatus väestötieteeseen:  
<http://www.valt.helsinki.fi/sosio/vaesto/>.

Vohlonen I., Bäckmand H. , Korhonen J. (2007). Menetetyt elinvuodet PYLL-indeksi väestön hyvinvoinnin mittana. *Suomen Lääkärilehti* 305 – 309.