

Tietosuoja Eläketurvakeskuksen tilastoissa

Ohjeessa kuvataan Eläketurvakeskuksen (ETK) tilastoissa käytettävät tietosuojamenetelmät ja taulukkomerkinnot sekä kerrotaan jakauman äärihavainnoista luovutettavasta tilastotiedosta. Ohje perustuu Suomen virallisen tilaston (SVT) neuvottelukunnan ohjeisiin Ohje taulukkomuotoisen henkilötiedon suojaamiseksi ja Ohje taulukkomuotoisen yritystiedon suojaamiseksi.

Tietosuojaohje koskee ETK:n julkaisemaa ja luovuttamaa tilastotietoa kaikissa julkaisumuodoissa. Tilastotietoa esitetään eri muodoissa verkkojulkaisuissa, tilastotietokannassa, ulkopuoliseen verkkopalveluun toimitetuissa tilastotiedoissa tai asiakkaalle tilastotietopalveluna toimitetussa taulukossa. Ohjetta sovelletaan myös, kun tilastotietoja esitetään raporteissa ja esityksissä.

Tilastot perustuvat rekisteritietoihin tai erillisaineistoihin, jotka ovat työntekijän eläkelain (julkisuuslain) tai lain Eläketurvakeskuksesta (tilastolain) mukaan **salassa pidettäviä**. Vaikka tiedot ovat salassa pidettäviä, voidaan henkilön eläketietoja käsitellä lakisääteisten tehtävien hoitamisessa.

Tilastotiedoista ei saa suoraan tai välillisesti tunnistaa **tilastoyksiköitä** eli tilastoitavia henkilöitä tai yrityksiä. Tilastotaulukko tulee tietosuojata, koska tilastoyksikön salassa pidettävä tieto ei saa paljastua tilastoista. Taulukkoaineistojen sisältämät henkilöihin tai yrityksiin kohdistuvat tietosuojariskit on aina arvioitava ennen tietojen julkaisemista tai luovuttamista. Tilastotaulukot tulee laatia niin, että taulukon solut eivät sisällä tarpeettoman pieniä arvoja.

Paljastumisriskillä tarkoitetaan tilastoyksikön vaaraa paljastua tilastosta. Paljastumisriskiä arvioidaan erityisesti arkaluonteisten tietojen osalta. Tietosuojalain mukaisia arkaluonteisia tietoja ovat sairaustiedot, kuten työkyvyttömyyseläkkeen alkaminen tai saaminen tai siihen liittyvä diagnoosi. Muina arkaluonteisina tietoina pidetään henkilön kansalaisuutta, asuinmaata, sosioekonomista asemaa tai taloudellisia tietoja, kuten ansioiden ja eläkkeen suuruutta. Paljastumisriskiä lisäävät myös tilaston tuoreus ja tilastoaiheen ajankohtaisuus.

Tietosuojamenetelmät

Tilastotaulukon tietosuojamenetelmiä ovat luokituksen muuttaminen ja tilastotiedon peittäminen. Tietosuojamenetelmät on jaoteltu lukumäärätaulukoille eli frekvenssitaulukoille ja määrätaulukoille. Lukumäärätaulukossa soluarvo on soluun kuuluvien tilastoyksiköiden lukumäärä, esimerkiksi eläkkeensaajien tai eläkkeelle siirtyneiden lukumäärä. Määrätaulukossa soluarvo on taulukoitavan muuttujan tunnusluku. Tunnusluku on tyypillisesti keskiarvo tai summa, esimerkiksi keskieläke tai eläkemeno, joka on summa kuukausieläkeistä.

Luokituksen muuttamisessa vaihdetaan tilaston luokittelija tai karkeistetaan luokittelijan luokkia, jolloin vältetään liian pienet solufrekvenssit. Luokitusta muuttamalla yhdistetään taulukon pienet luokat, jotta paljastumisriskissä olevat solut eivät näy. Luokituksen muuttaminen tarkoittaa usein koko luokituksen karkeistamista, esimerkiksi muutetaan 1-vuotisikäluokat 5-vuotisikäluokkiin. Lisäksi voidaan yhdistää suurimpia ikäluokkia luokaksi ”80-”, vaikka muu aineisto on luokiteltu 5-vuotisikäluokittain.

Ensisijainen **peittäminen** tarkoittaa, että taulukon paljastumisriskissä olevien solujen sisältö korvataan puuttuvan tiedon merkillä. Tällöin lukumäärätaulukossa peitetään niiden solujen lukumäärät, jotka ovat pienempiä kuin kynnysarvo. Esimerkiksi kynnysarvolla kolme peitetään tilastotaulukon solut, joissa lukumäärä on alle 3 eli lukumäärät yksi ja kaksi peitetään. Määrätaulukossa tunnusluku peitetään niissä soluissa, joissa tunnusluvun laskemisessa tilastoyksiköitä on kynnysarvoa vähemmän, esimerkiksi kynnysarvolla 3 tunnusluku on laskettu yhden tai kahden tilastoyksikön tiedoista.

Arkaluonteista tietoa sisältävissä aluetaulukoissa peittämisen kynnysarvo on maakunnittain esitettävissä tiedoissa 3 ja kunnittain esitettävissä tiedoissa 7. Näistä isompaa kynnysarvoa 7 voidaan käyttää koko tilastotietokantataulukon peittämisessä, jos se sisältää sekä kunta- että maakuntaluokat.

Lisäksi voidaan käyttää toissijaista taulukon peittämistä, esimerkiksi peitetään kokonainen taulukkorivi, jos kaikki solufrekvenssit ovat pieniä. Näin voidaan peittää esimerkiksi lapseneläkkeensaajien tai ulkomailla asuvien keskieläkkeet ikäluokittain. Totaalisumman eli rivi- tai sarakesumman peittämistä käytetään tarvittaessa.

ETK ei tietosuojasyistä julkaise rekisteritiedon äärihavaintoja. Jakauman äärihavaintoja ovat sen pienet ja suuret arvot.

Taulukkomerkinnot

Tilastotaulukoiden merkinnöissä sovelletaan SVT-neuvottelukunnan julkaisuohjetta. Taulukon peitetyn solun merkintäsymbolina on kaksi pistettä.

Tilastotaulukoissa käytettävät merkinnot:

0 suure on pienempi kuin puolet käytetystä yksiköstä

- ei tapauksia

. loogisesti mahdoton

.. tietoa ei ole saatavissa tai sitä ei julkaista tapausten vähäisyyden vuoksi

* ennakkotieto tai arvio

Tilastotietokannassa käytettävät merkinnot:

0 suure on pienempi kuin puolet käytetystä yksiköstä

- ei tapauksia tai loogisesti mahdoton

.. tietoa ei ole saatavissa tai sitä ei julkaista tapausten vähäisyyden vuoksi