

Matkailijamittari datan ohje

Aineiston kuvaus (Matkailijamittari datan kuvaus) sisältää muuttujien kuvaukset välilehdellä "Muuttujat" sekä eri vastausvaihtoehtojen kuvaukset välilehdellä "Vastausvaihtoehdot".

Aineisto (Matkailijamittari aineisto) sisältää varsinaisen kohdejoukon matkat, eli siitä on poistettu Suomessa asuvat, vaihtomatrustajat ja kuljettajat/miehistön jäsenet ja yli vuoden kestävät matkat. Jokainen rivi datassa on yksi haastattelu eli yksi matka Suomeen.

Painokertoimien avulla yksittäisen vastaajan tiedot yleistetään koko perusjoukon tasolle. Painokertoimet lasketaan kuukauden tai neljänneksen ja lennon tai laivavuoron kohteen mukaan. Laskennan edellytyksenä on tuntee sekä otoksen koko (n) että perusjoukon koko (N). Otos muodostuu kuukauden tai neljänneksen aikana haastatelluista henkilöistä ja perusjoukko matkailijoista yhteensä kuukauden tai neljänneksen aikana.

Seuraava fiktiivinen esimerkki havainnollistaa painokertoimen laskentaperiaatetta

- Lähteitä lentomatrustajia Helsingistä Tukholmaan oli yhteensä 100 000 (N)
- Helsinki-Vantaan lentoasemalla tehtiin yhteensä 100 (n) sellaista matkailijoiden haastattelua, joiden kohdehenkilö oli menossa Tukholmaan suuntautuvalle lennolle
- Kohdehenkilöistä 50 oli Suomessa asuvia, jotka kuuluvat tutkimuksen perusjoukkoon mutta eivät kohdejoukkoon.
- 50 kohdehenkilöä oli kohdejoukkoon kuuluvia ulkomailla asuvia matkailijoita
- Kaikki em. kohdehenkilöt saavat aineistossa painokertoimen $N / n = 1\ 000$
- Vastaavanlainen painotus tehdään erikseen kaikille lento- ja laivavuorojen kohteille

PainoKk -muuttuja on lento- ja laivaliikenteen tilastojen perusteella laskettu kuukausipaino, joka siis kertoo kuinka montaa matrustajaa kyseinen henkilö edustaa. Toisin sanoen yhden kuukauden matkamäärän saa laskemalla painot yhteen. Aineistossa on vastaavasti myös muita Paino-alkuisia muuttujia, esimerkiksi PainoKkEURYht muuttujaan on valmiiksi kerrottu rahankäyttö painokertoimella. Esimerkiksi henkilö, jonka ID on 1007 on ilmoittanut rahankäytökseen (EURYht) 656,54 euroa ja matkan painokerroin on 696,88, eli henkilö edustaa noin 697 maaliskuun laivamatrustajaa. Kertomalla nämä yhteen saadaan yhteensä (PainoKkEURYht) 457 527 euroa, joka on kyseisen matkan osuus maaliskuun 2023 rahankäytöstä yhteensä (277 milj. euroa).

PainoQ -muuttuja on periaatteeltaan sama kuin PainoKk, mutta haastattelujen ja kaikkien matkailijoiden määrä on laskettu kyseisessä neljänneksessä.

Painokerrointa tulee aina käyttää, kun datasta lasketaan mitä tahansa asioita. Valitse PainoKk, jos lasket kuukausitasolla ja PainoQ, jos lasket neljänneksitasolla. Kuukausittain julkaistavan aineiston virhemarginaalit voivat olla suuria ja näin ollen ei ole järkevää laskea kovin yksityiskohtaista tietoa kuukausiaineistosta. Esimerkiksi matkailijoiden asuinmaiden jakaumaa ei tule laskea kuukausiaineistosta. Neljännesaineistosta voidaan laskea yksityiskohtaisempia tietoja, koska vastauksia on enemmän ja siitä johtuen virhemarginaali on pienempi.

Datassa joihinkin muuttujiin liittyy lisäksi "Flag"-loppuinen muuttuja, joista käy ilmi mitkä tiedot on täydennetty tilastollisin menetelmin. Esimerkiksi kaikki puuttuvat rahankäyttötiedot on täydennetty ja pakettimatrustojen menot on jaettu osatekijöihin (majoitus, ateriat, pääsyliput, muut). Tavoitteena on siis se, että aineistossa olisi mahdollisimman vähän sellaista puuttuvuutta, joka tekee analyysistä hankalaa.

Border Survey data guide

The description of the data (Border Survey data description) contains descriptions of the variables on the "Muuttujat" tab and descriptions of the different answer options on the "Vastausvaihtoehdot" tab.

The data (Border Survey data) contains the trips of the target group, that is people living in Finland, exchange passengers and drivers/crew members and trips lasting more than a year excluded. Each line in the data is one interview, i.e. one trip to Finland.

With the weighting factor one respondent's information is generalized to the population level. The weighting factors are calculated by destination based on sample (n) and population (N). The sample is the number of interviews and the population is all departing passengers.

The PainoKk variable is the monthly weight calculated based on air and ship traffic statistics, which indicates how many passengers the person in question represents. In other words, the number of trips for one month is obtained by adding the weights together. Correspondingly, the data also contains other variables beginning with Paino, for example, the variable PainoKkEURYht has already been multiplied by a weighting factor. For example, a person whose ID is 1007 has reported expenditures of 656.54 euros (EURYht) and the weighting factor for the trip is 696.88, meaning that the person represents roughly 697 ship passengers in March. By multiplying these together, we get a total of (PainoKkEURYht) 457,527 euros, which is the share of the trip in question from the total expenditure in March 2023 (277 million euros).

The variable PainoQ is in principal similar to the PainoKk, except the number of interviews and total passengers is calculated by quarter.

The weighting factor should always be used when calculating information from the data. Use the PainoKk with the time variable Kuukausi (month) and use the PainoQ with the time variable Neljannes (quarter). The margins of error for data published monthly can be large, and thus it is not reasonable to calculate very detailed information from monthly data. For example, the distribution of tourists' countries of residence should not be calculated from monthly data. The detailed information is recommended to be calculated from the quarterly data, because there are more answers and therefore the margin of error is smaller.

In the data, some variables are also associated with a "Flag" variable, which indicates which data have been supplemented with statistical methods. For example, all missing information on expenditure has been completed and package travel expenses have been divided into components (accommodation, meals, entrance tickets, others).