

Faraldsfræði í dag

Að lýsa gögnum

FARALDSFRÆÐILEGAR rannsóknir, hvaða nafni sem þær nefnast, byggjast á gögnum um einstaklinga. Ýmsar aðferðir eru til að flokka mismunandi tegundir gagna en almennt má segja að þau skiptist í afmörkuð (discrete) og samfelld (continuous) gögn (1).

Afmörkuð gögn eru þau sem aðeins geta tekið á sig tiltekinn afmarkaðan fjölda gilda og engin gildi þar á milli (ekki brot eða hluta af gildum). Einfaldasta gerð afmarkaðra gagna eru þau sem aðeins geta tekið á sig tvö gildi, til dæmis karl, kona eða lifandi/látinn (hvoru tveggja oft táknað með tölustöfunum 1 og 2, 0 eða 1 eða einhverjum bókstöfum). Er þá talað um tvískipt eða tvískorin (dichotomous) gögn. Ef um fleiri flokka er að ræða er talað um margskipt (multitomous) gögn.

Margskipt gögn falla í þrjá flokka, nafntölugögn (nominal), raðtölugögn (ordinal) og töluleg afmörkuð gögn (numerical discrete data). Nafntölugögn og raðtölugögn eru á eigindlegu formi (qualitative) og eru oftast skráð með einhvers konar kóða sem getur verið tala, talnaröð, röð bókstafa eða orð. Töluleg afmörkuð gögn eru hins vegar meginleg og eru skráð sem stærð, það er gögnin fela í sér einhvers konar tölulega mælingu eða talningu, til dæmis fjöldi sjúkdómsgreininga.

Nafntölugögn fela ekki í sér neina sérstaka röð, gögnin hreinlega nefna fyrirbærin sem þau standa fyrir en fela ekki í sér samanburð eða uppröðun á borð við betra-verra eða stór-stærri-stærstur. Sem dæmi um slík gögn má taka búsetu eftir kjördæmum, sjúkdómsgreiningar eða kynþátt (til dæmis svartir, asískir, hvítir, aðrir). Raðtölugögn hins vegar fela í sér slíka uppröðun eða samanburð. Þar er um að ræða eigindleg gögn sem skipa sér í tiltekna röð með skilgreindum fjölda gilda (því þetta eru jú afmörkuð gögn). Slík gögn eru til dæmis stigun krabbameins (til dæmis samkvæmt TNM kerfi) þar sem er skilgreindur fjöldi flokka sem raðast eftir stigi meinsins en engin millistig. Eins má nefna einfaldari stigunarkerfi til dæmis í þvagræktun þar sem fjöldi sýkla er gefinn til kynna með +, ++ eða +++. Mikilvægt er að átta sig á að bilin milli flokkanna eru ekki endilega jöfn. Tökum dæmi um gögn um reykingavenjur sem stundum eru skráð sem þrjú flokkar: hættur, reykir, aldrei reykt. Augljóslega er ekki hægt að skilja þetta sem svo að munurinn á flokkunum „hættur“ og „reykir“ sé sá sami og munurinn á „reykir“ og „aldrei reykt“.

Töluleg afmörkuð gögn eru flóknasta gerð afmarkaðra gagna þar sem hinir afmörkuðu flokkar fela ekki í sér eigindleg gögn heldur meginleg eins og áður segir. Dæmi um slík gögn er til dæmis fjöldi lyfja sem sjúklingur er á, fjöldi miðeyrnasykinga hjá

barni á einu ári. Þessi gögn eru skilgreind sem afmörkuð þar sem, þó að um töluleg, meginleg gögn sé að ræða, þau geta aðeins tekið á sig ákveðin gildi og ekkert þar á milli, annaðhvort er sjúklingur á þremur eða fjórum lyfjum, en ekki þremur og hálfu.

Samfelld (continuous) gögn eru þau sem eru meginleg og geta tekið á sig hvaða gildi sem er á samfelldum skilgreindum skala með upphafi og endapunkti. Klínísk gögn eru gjarnan af þessu tagi, svo sem hæð, þyngd, blóðsykur, blóðþéttni lyfja og svo framvegis. Þessi gögn eru almennt nákvæmari en hin afmörkuðu og nákvæmni þeirra takmarkast fyrst og fremst af aðferðunum sem notaðar eru við mælingu þeirra. Því bjóða þau mun meiri og nákvæmari möguleika á lýsingu og greiningu en afmörkuð gögn.

Ofangreind flokkun gagnategunda er sú sem mest er notuð í faraldsfræði en til viðbótar, einkum í gagnavinnslu, eru meginleg gögn oft einnig flokkuð eftir því hvort þau eru mæld á bilkvarða (interval scale) eða hlutfallskvarða.

Gögn sem mæld eru á bilkvarða hafa það umfram raðtölugögn að vera í raun meginleg og jafnan langt er á milli gagnaflokka, gögnin liggja á skilgreindum jöfnum bilum. Dæmi um slíkt eru gögn sem mæld eru sem heilir tugir. Við vitum að á milli 30 og 40 er sama fjarlægð og milli 1010 og 1020. Slíkir bilkvarðar byggjast á því að kvarðinn hafi núllpunkt og einingar. Hvorutveggja er þó afstætt, einingin er skilgreiningaratriði og núllið sömuleiðis, núllpunktur þýðir ekki að það sem verið er að mæla sé ekki til staðar. Sem dæmi má taka hitamæla; mælieiningin er gráður og viðmiðunarpunkturinn er núll en núll gráðu hiti merkir ekki að hitastig sé ekki til staðar, aðeins að á þessum kvarða er hitastigið í núllpunkti.

Hlutfallskvarði hefur einnig núllpunkt og skilgreinda einingu en hins vegar eru þau ekki afstæð. Núll á slíkum kvarða þýðir í raun að það sem mælt er sé ekki til staðar. Fjölmargar klínískar upplýsingar eru á þessu formi, svo sem hæð, þyngd, blóðrauðagildi.

Í stuttu máli: samkvæmt almennri faraldsfræðilegri flokkun rannsóknargagna falla gögn í tvo meginflokka (afmörkuð og samfelld). Afmörkuð gögn geta verið tvískipt eða margskipt. Margskipt gögn geta verið á þrennu formi: nafntölur, raðtölur og töluleg afmörkuð gögn. Nafntölugögn og raðtölugögn eru eigindleg en hin síðastnefndu eru meginleg. Samfelld gögn (meiginleg) eru svo flokkuð frekar eftir því hvort um bilkvarða eða hlutfallskvarða er að ræða.

Þessi umfjöllun hefur ekki aðeins flokkunarfræðilegt gildi (!) heldur miklu fremur verulegt praktískt gildi þar sem hún er undirstaða þess hvers konar tölfræðilegum aðferðum má beita til úrvinnslu.



María
Heimisdóttir

mariahei@landspitali.is

Heimild

1. Við þessari skrif var stuðst við CH Hennekens, JE Buring. *Epidemiology in Medicine*. 1st edition 1987. Little Brown and Company, Boston/Toronto.