

Hinar fjórar gerðir breyta og myndræn framsetning þeirra

Sigrún Helga Lund

prófessor í tölfræði við Háskóla Íslands



Ómissandi grundvallaratriði allrar tölfræðiúrvinnslu er að átta sig á gerð breytanna sem gögnin geyma. Alls eru fjórar ólíkar gerðir breyta og mun ég í þessum pistli lýsa hverri og einni þeirra og tiltaka hvers konar myndræn framsetning er við hæfi fyrir hana.

Óraðaðar flokkabreytur skipta mælingum upp í flokka sem hafa enga innbyrðis röðun. Blóðflokkarnir (O, A, B, AB) eru gott dæmi um óraðaða flokkabreytu. Óröðuðum flokkabreytum má lýsa með **stöplarit** eða **kökuriti**. Stöplarit sýna fjölda mælinga í hverjum flokki á meðan kökuriti sýna hlutfallslega skiptingu þeirra. Mannsaugað á auðveldara með að bera saman hæð stöpla en gráður hringgeira og því eru tölfræðingar síður hrifnir af kökuritum. Á mynd 1 eru sömu gögn teiknuð með stöplarit annars vegar og kökuriti hins vegar. Á stöplaritinu er auðvelt að raða flokkunum frá þeim algengasta til þess fátíðasta á meðan það er ekki augljóst á kökuritinu.

Raðaðar flokkabreytur skipta mælingum upp í flokka sem hafa innbyrðis röðun, t.d. stigun krabbameina (I, II, III og IV). Röðuðum flokkabreytum er best lýst með **stuðlariti** þar sem flokkunum er raðað

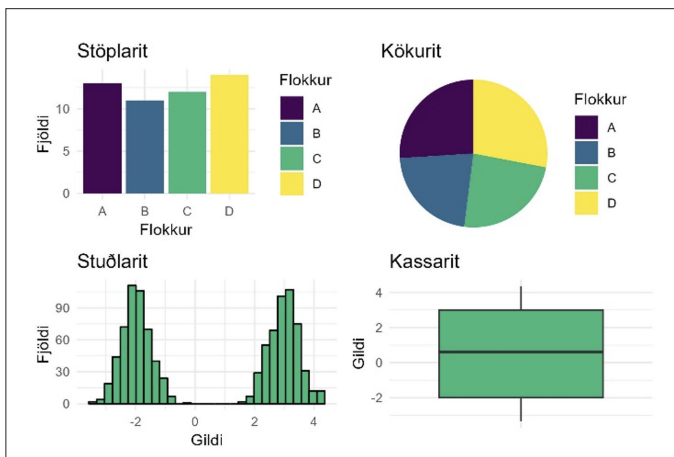
eftir innbyrðis röð, frá þeim minnsta til hins stærsta. Þannig væri flokkum stigunar raðað I, II, III og IV. Ekki ætti að nota kökuriti til að lýsa röðuðum flokkabreytum þar sem þau sýna ekki innbyrðis röð flokkanna.

Samfelldar talnabreytur eru mældar í tiltekinni einingu og geta tekið hvaða gildi sem er á einhverju bili. Dæmi er magn kólesteróls í blóði (mmól/L). Þær er hægt að setja fram með **stuðlariti** annars vegar og **kassariti** hins vegar. Þegar stuðlarit eru teiknuð er kvarðanum sem breytan er mæld á skipt upp í jafnbreið, samliggjandi bil og fjöldi mælinga sem lendar á hverju bili er talinn. Mikilvægt er að breidd bilanna hæfi mælingunum, ef bilin eru of stutt er breytileiki mælinganna ýktur en ef þau eru of löng er hann jafnaður út. Kassarit sýna miðgildi, fjórðungamörk, dreifð og útlaga breytu og gefa því einfalda en grófa mynd af úrtaksdreifingu hennar. Stuðlarit hafa þann kost fram yfir kassarit að þau gefa betri mynd af úrtaksdreifingu breytunnar á meðan kassarit eru algengari þegar samfelldrar talnabreytur eru bornar saman milli hópa. Á mynd 1 má sjá stuðlarit og kassarit af sömu gögnum. Stuðlaritið sýnir skýrt tvítoppa úrtaks-

dreifingu sem ekki er hægt að sjá út frá kassaritinu.

Strjálur talnabreytur taka ólík aðskilin gildi og lýsa yfirleitt fjölda einhvers. Dæmi eru fjöldi innlagna á sjúkrahús á síðasta ári, fjöldi fæðinga eða fjöldi ávísaðra lyfja á tiltekinni stundu. Strjálum talnabreytum er best lýst með **stöplarit** sem sýnir fjölda mælinga sem tilheyrar hverjum og einum fjölda. Mikilvægt er hafa skýrt bil milli allra stöpla til aðgreiningar frá stuðlariti. Stundum getur strjál talnabreyta tekið svo ótalmörg gildi að munurinn á einni einingu til og frá er hverfandi, eins og t.d. fjöldi blóðflagna í blóðsýni. Þá er eðlilegra að meðhöndla breytuna eins og hún væri samfelld í allri tölfræðiúrvinnslu og lýsa henni með **stuðlariti** eða **kassariti**.

Að lokum vil ég minna á nokkrar góðar reglur sem gilda um öll gröf. Merkið ása og flokka nákvæmt en skorinort, hafið liti aðgreinanlega og gætið að samræmi í litavali og merkingum milli grafa og annars texta. Forðist óþarfa flúr og flækjur, markmiðið er að sýna niðurstöðurnar ykkar á aðgengilegan og auðskiljanlegan hátt. Oft er það einfalda jafnframt það besta!



Mynd 1. Þær fjórar gerðir grafa sem lýsa úrtaksdreifingu einnar breytu. Sömu gögn eru notuð til að teikna stöplarit og kökuritið annars vegar og stuðlaritið og kassaritið hins vegar.