

# Myndræn framsetning

(STATS205.1: Grunnhugtök, tilraunahögun og lýsandi tölfræði)

Anna Helga Jonsdottir og Sigrun Helga Lund

January 7, 2013

# Myndræn framsetning

- Fyrsta skref í tölfræðiúrvinnslu ætti ætíð að vera að skoða gögnin myndrænt
- Kjarni tölfræðiúrvinnslu er að átta sig sem best á eðli mælinganna sem skoðaðar eru
- Breytileiki gagnanna lykilatriði - hvernig er dreifing gagnanna?
  - Hversu mikinn mun sjáum við á útkomum viðfangsefnanna okkar?
  - Hvernig dreifast útkomurnar?
- Myndræn framsetning er ein besta leiðin til að átta sig á dreifingu mælinganna
- Gröfin sem við skoðum nú sýna eingöngu eina breyту í einu og gerum við greinarmun á hvort breytan sem um ræðir er strjál eða samfelld

## Myndræn framsetning á strjálum breytum

- Algengustu tegundir grafa fyrir strjálur breytur eru **stöplarit** (bar chart) og **kökurit** (pie chart)
- Kökurit eru mikið notuð í viðskiptaheiminum og í fjölmiðlum en eru sjaldséð í tímaritum og bókum um raunvísindi
- Stöplarit má sjá víðsvegar og eru þau í nánast öllum tilfellum betur til þess fallin að sýna gildi strjálra breyta myndrænt en kökuritin

## Stöplarit (bar chart)

Stöplarit samanstanda af tveimur eða fleiri súlum. Fjöldi súlna ræðst af fjölda flokka/gilda breytunnar. Hver súla stendur fyrir einn flokk og mega þær ekki liggja hvor að annarri. Hæð súlnanna sýnir tíðni eða hlutfall fyrir viðkomandi flokk. Raða skal súlunum svo auðvelt sé að greina upplýsingarnar, oft er þeim raðað upp eftir stærð.

Áður en hafist er handa við að teikna stöplarit er gott að setja upp litla töflu sem sýnir flokka breytunar og hversu mörg viðfangsefni eru í hverjum flokki.

## Kökurit (pie chart)

Pegar búa á til kökurit er mikilvægt að allir flokkar/gildi breytunnar sem verið er að skoða séu með á myndinni. Fjöldi sneiða í kökuritinu ræðst af fjölda flokka/gilda breytunnar. Stærð sneiðarinnar ræðst af hlutfallslegum fjölda í viðkomandi flokki af heildinni. Gætið að hlutföllin séu samanlögð 100%.

Áður en hafist er handa við að teikna kökurit er gott að gera litla töflu sem sýnir flokka/gildi breytunnar, fjölda í hverjum flokki og hlutfall viðfangsefna í hverjum flokki af heildinni.

VIÐ GERUM EKKI KÖKURIT Í HÖNDUNUM!

## Myndræn framsetning á samfelldum breytum

- Algengasta aðferðin til að skoða samfellda breytu myndrænt er stuðlarit.
- Kassarit eru einnig góð aðferð til að skoða samfelldar breytur og munum við kynnst þeim í kaflanum um lýsandi tölfræði.
- Við einnig skoða punktarit í kaflanum um aðhverfsgreiningu en þau eru notuð til að kanna samband tveggja samfelldra breyta.

- Stuðlarit er svipað stöplariti í útliti en helsti munur á útliti þeirra er að ekkert bil er á milli súlnanna í stuðlariti
  - Gott er að hugsa sér stuðlaberg til að muna hvort bil eigi að vera á milli súlnanna í stuðlariti!
- Snúnara að búa til stuðlarit en stöplarit þar sem samfelldar breytur innihalda ekki eiginlega flokka
- Byrja á að mynda flokka áður en talið er hversu margar mælingar falla í hvern flokk
- Þegar flokkarnir hafa verið myndaðir er gott að búa til töflu sem inniheldur flokkana og hversu margar mælingar lenda í hverjum flokki

# Stuðlarit

## Studlarit (histogram)

Stuðlarit samanstendur af súlum sem standa hvor upp að annari. Fjöldi súlna ræðst af fjölda flokka sem samfelldu breytunni er skipt upp í. Þegar flokkarnir eru myndaðir er gott að hafa eftirfarandi í huga.

- Neðri og efri mörk eiga að vera einföld og auðskilin
- Bilin mega ekki skarast og verða að ná yfir allar mælingar
- Bilin eiga að vera jafn breið
- Flokkarnir eiga að vera hæfilega margir. Engin ein rétt lausn er til en ágætt er að nota þumalputtaregluna að fjöldi flokka á að vera u.þ.b 5 sinnum logaritminn af fjölda mælinga,

$$\text{fjöldi flokka} = 5 \cdot \log(\text{fjöldi mælinga})$$

Þegar flokkarnir hafa verið myndaðir er teiknuð ein súla fyrir hvern flokk og ræðst hæð súlunnar af fjölda (eða hlutfalli) mælinga í þeim flokki.



# Lögun dreifinga

## Lögun dreifinga (Shape of distributions)

Eftirfarandi hugtök eru oft notuð til að lýsa dreifingum mælinga.

- Dreifingu minnstu mælinganna köllum við *vinstri hala* (left-tail) dreifingarinnar. Dreifingu stærstu mælinganna köllum við *hægri hala* (right-tail) dreifingarinnar.
- Dreifing er *samhverf* (symmetric) ef hægri hlið hennar dreifist eins og spegilmynd vinstri hliðarinnar.
- Dreifing sem ekki er samhverf er *skekkt* (skewed). Dreifing er *skekkt til hægri* (skewed to the right) ef hægri hali hennar er lengri en sá vinstri og *skekkt til vinstri* (skewed to the left) ef sá vinstri er lengri en sá hægri.
- Ef dreifingin hefur einn topp er talað um *einkryppudreifingu* (unimodal).
- Ef dreifingin hefur tvo toppa er talað um *tvíkryppudreifingu* (bimodal).
- Ef dreifing hefur fleiri en tvo toppa er talað um *fjölkrýppudreifingu* (multimodal).

## Utlagar (Outliers)

**Útlagar** eru mæligildi sem eru mjög ólík öðrum mæligildum í sama gagnasafni. Ýmsar ástæður geta verið fyrir útlögum og er mjög mikilvægt að skoða þá sérstaklega og hugleiða ástæðu þeirra.