

Forprosjekt – Statistikk

Gruppediskusjon om et stokastisk forsøk

Verdimengde

Verdimengden til en stokastisk variabel er mengden av alle mulige verdier variabelen kan anta. Skostørrelse er en diskret variabel, og basert på våre data blir verdimengden:

- $X = \{42, 43, 47\}$.

Høyde er i teorien en kontinuerlig variabel, men siden vi kun har en endelig mengde observerte verdier, blir verdimengden:

- $Y = \{173, 183, 186, 189, 193\}$

Diskret/kontinuerlig

I praksis er datasettet kontinuerlig, men i virkeligheten er det diskret. Dette er fordi vi ikke kan få et uendelig antall desimaltall i virkeligheten når vi jobber med sko størrelse og høyde. Så ettersom at det er et finitt datasett vil X og Y alltid være diskret.

Forventningsverdi

Teoretisk sett vil forventningsverdien være gjennomsnittlig verdi for skostørrelse/høyde for norsk person i aldersgruppe student. I kontekst av vårt datasett vil forventningsverdien bare bli gjennomsnittets verdi fra datasettet vårt.

Standardavvik

Standardavvik blir *kvadratroten av gjennomsnittet av de kvadrerte avvikene fra gjennomsnittet*. Dette gir oss et tall på hvor stor spredning vi har mellom dataene.

Hvis det er ett stort standardavvik vet vi at det er stor variasjon mellom dataene våre og gjennomsnittshøyden/skostørrelsen, og ett lite standardavvik vil si at alle høydene/skostørrelsene er veldig nære gjennomsnittet.

Kovarians Ettersom at høyde korrelerer med skostørrelse vil X og Y ha en kovarians. Kovariansen mellom skostørrelse og høyde viser hvordan de varierer sammen. Siden høyere personer ofte har større sko, forventes det en positiv kovarians. Det vil si at når X øker, øker Y også, punktene danner en stigende sky, og en enkel regresjonslinje vil ha positiv helning.

Form på sannsynlighetsfordelingene for X og Y

X (skostørrelse): er en diskret heltallsvariabel. I en blandet NTNU-populasjon forventer vi en svakt bimodal fordeling (to "klynger" for kvinner/menn). Innen hver delgruppe er den unimodal og nær normalform rundt en typisk størrelse.

Y (høyde): er en kontinuerlig variabel, men i praksis vil det bli en diskret heltallsvariabel fordi man oftest runder opp høyde til nærmeste cm. For voksne er høyde tilnærmet normalfordelt innen hver delgruppe (én symmetrisk klokke). Sammenslått gruppe kan gi to overlappende klokker (bimodal).