

Arbeidskrav 7: Statistisk inferens

Gard, Henrik og Sebastian

- Beskriv hvordan vi kan si noe om en populasjon basert på et utvalg ved å bruke p-verdier og konfidensintervaller?

p-verdier kan si oss noe om hvor ofte vi vil se et slikt eller mere ekstremt resultat gitt at nullhypotesen er sann. Det vil si at hvis resultatet er veldig usannsynlig i forhold til nullhypotesen, kan vi anta at det faktisk er en effekt i populasjonen.

Konfidensintervaller kan si oss noe om en populasjon ved å gi oss et estimat på hvor populasjonsverdien befinner seg. Det brukes altså data fra et utvalg til å trekke en sluttning om populasjonen.

- Gi en definisjon av p-verdier.

En p-verdi er sannsynligheten for å få et resultat like ekstremt eller mer ekstremt enn det vi observerte, gitt at nullhypotesen er sann.

- Definere et konfidensintervall.

Et konfidensintervall er et intervall av verdier som med en bestemt sannsynlighet (ofte 95 %) vil dekke den sanne populasjonsparameteren dersom vi gjentar studien uendelig mange ganger.

- Når man planlegger en studie ønsker vi å kontrollere type 1 og type 2 feil, beskriv hva disse er og hvordan vi kan kontrollere de.

Type 1 feil er at vi forkaster nullhypotesen når den egentlig er sann. Type 2 feil er når vi beholder nullhypotesen når det faktisk finnes en effekt. Vi kan kontrollere de på to måter: hvis vi gjør alfa mindre reduserer vi sjansen for type 1 feil, øker vi power minsker sannsynligheten for type 2 feil.

- Hvorfor er det å beregne statistisk styrke i en enkelt studie ikke en god ide?

Fordi power sier oss sannsynligheten for å finne en effekt hvis studien blir repeteret uendelig mange ganger, og sier oss dermed lite om akkurat denne enkeltstudien. Power brukes da helst i planleggingsfasen av en studie, og gir oss ingen ny informasjon etter at studien er gjennomført fordi det bare er en omskalering av p-verdien.

- Beskriv noen hovedpunkter på forskjeller mellom frekventisme og en Bayesiansk tilnærming til statistisk inferens.

I frekventistisk statistikk tolker vi sannsynlighet på en måte hvor vi tar utgangspunkt i hvor ofte noe vil skje om vi gjentar det uendelig mange ganger. Det sier ikke noe om sannsynligheten for utfallet i akkurat denne studien.

Bayes handler om hvor sannsynlig hypotesen er gitt dataene. Sannsynlighet regnes som et mål på hvor usikre vi er om noe og vi bruker forhåndskunnskap for å si noe om sannsynligheten, altså hva vi allerede vet.

- Hva er en prior?

En prior er forhåndskunnskap eller forventning om effekten før dataen analyseres.

- Hva beskriver vi med en «posterior probability distribution»?

Dette er den oppdaterte sannsynligheten etter vi har sett dataen. Den beskriver altså hva vi nå tror er sannsynlig både ut i fra tidligere kunnskap og ny informasjon.

-Bruk data fra Arbeidskrav 3 (strengthvolume) og sett opp en analyse for frekvenstistisk tolkning av forskjell mellom 1 og tre sett. Sett også opp en analyse ved hjelp av en Bayesiansk modell. Beskriv forskjeller i fremgangsmåte og tolkning.

For å undersøke om tre sett gir større fremgang enn ett sett i leg press, ble endringen i belastning beregnet for hver deltaker (post – pre). Data ble delt i gruppene «single» og «multiple» sett. Gjennomsnittlig fremgang var 86,4 kg i «single» og 97,0 kg i «multiple».

Den frekventistiske analysen ble gjort med en to-utvalgs t-test. Resultatet var $t = 1,35$, $p = 0,18$ og 95 % konfidensintervall for gruppedifferansen [-5,08, 26,41]. Siden p-verdien er høyere enn 0,05 og konfidensintervallet inkluderer null, finner vi ikke en statistisk signifikant forskjell mellom ett og tre sett. Vi kan derfor ikke forkaste nullhypotesen om ingen forskjell i fremgang.

Den Bayesianske analysen ble gjort med en Bayes faktor-test, som ga $BF = 0,55$. Dette betyr at dataene er noe mer sannsynlige hvis det ikke er forskjell mellom gruppene enn hvis det er en reell forskjell. Evidensen er svak, men peker i retning av ingen klar forskjell. I motsetning til p-verdier kan Bayes faktor gi støtte til nullhypotesen, og her er det mer støtte for ingen forskjell enn for en forskjell.

Begge analysene peker dermed i samme retning: det er ikke tilstrekkelig evidens for at tre sett gir større fremgang enn ett sett i dette datasettet.

Welch Two Sample t-test

```
data: delta by sets
t = 1.3542, df = 62.495, p-value = 0.1805
```

```
alternative hypothesis: true difference in means between group multiple and group single is not equal to 0
95 percent confidence interval:
-5.076399 26.411626
sample estimates:
mean in group multiple   mean in group single
97.03125                 86.36364

Bayes factor analysis
-----
[1] Alt., r=0.707 : 0.5512485 ±0.01%

Against denominator:
Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```