

Assignment 3

Pernille Breili, Aleksander Josdal, Vemund Lien & Tonje Fredriksen

11/2/2021

Introduksjon

Simulering av to ulike studier

```
##
## Call:
## lm(formula = y ~ 1, data = samp1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5322 -1.2523 -0.0883  1.3540  4.8692
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.840      1.251    1.47   0.185
##
## Residual standard error: 3.539 on 7 degrees of freedom

##
## Call:
## lm(formula = y ~ 1, data = samp2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.6557 -2.2883  0.2636  2.2549  6.4212
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.5642      0.4774    3.276 0.00221 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.019 on 39 degrees of freedom
```

Spørsmål

1. Forklar følgende begreper:

Estimat: Gjennomsnittlige verdier

Standardfeil: Standardfeil til gjennomsnittet er et estimat av standardavviket for en populasjon. T-verdi:

Ratio mellom estimatet og standardfeilen. Jo lavere standardfeilen er, desto høyere høyere blir t-verdien.

P-verdi: P-verdi måler sannsynligheten for at de observerte forskjellene vi ser i et datasett bare har skjedd ved en tilfeldighet. Jo lavere p-verdi, desto større statistisk signifikans av den observerte forskjellen. P-verdi kan brukes om et alternativ til, eller i tillegg til forhåndsvalgte konfindensnivåer for hypotesetesting. Jo lavere p-verdi vi har, desto høyere t-verdi får vi. Dette vil også si høyere bevis for at nullhypotesen er feil.

2. Hva betyr forskjellene mellom resultatene i m1 og m2?

m2 har et større utvalg (40 deltakere) og vi vil derfor her ha en lavere standardfeil, som vil si at vi kan være sikrere på resultatene våre. Jo mindre standardfeil vi har, desto høyere t-verdi har vi, i tillegg til en lavere p-verdi. Dette til sammen øker beviset for at nullhypotesen er feil. Et mindre utvalg gir større usikkerhet fordi gjennomsnittsverdiene er mindre presise, det betyr at ekstreme verdier blir mindre overraskende. Dette vil igjen påvirke på vår p-verdi.

3. Hvorfor bruker vi de skraverte områdene i nedre og øvre hale av t-distribusjonen?

De skraverte områdene markerer forkastelsesområder. Dette vil si at vi forkaster nullhypotesen hvis verdien vår er innenfor dette område. Jo større utvalgsstørrelse, desto smalere vil halene i t-distribusjonen være. Da kan vi være mer sikre på at vi gjør det riktig i å forkaste nullhypotesen. Det vil si - mindre sannsynlig for å få type I- og type II-feil.

Simulering av 1000 studier

4. a. Kalkuler standardavvik av gjennomsnittet (estimat) og gjennomsnittet av standardfeilen for hver utvalgsstørrelsen?

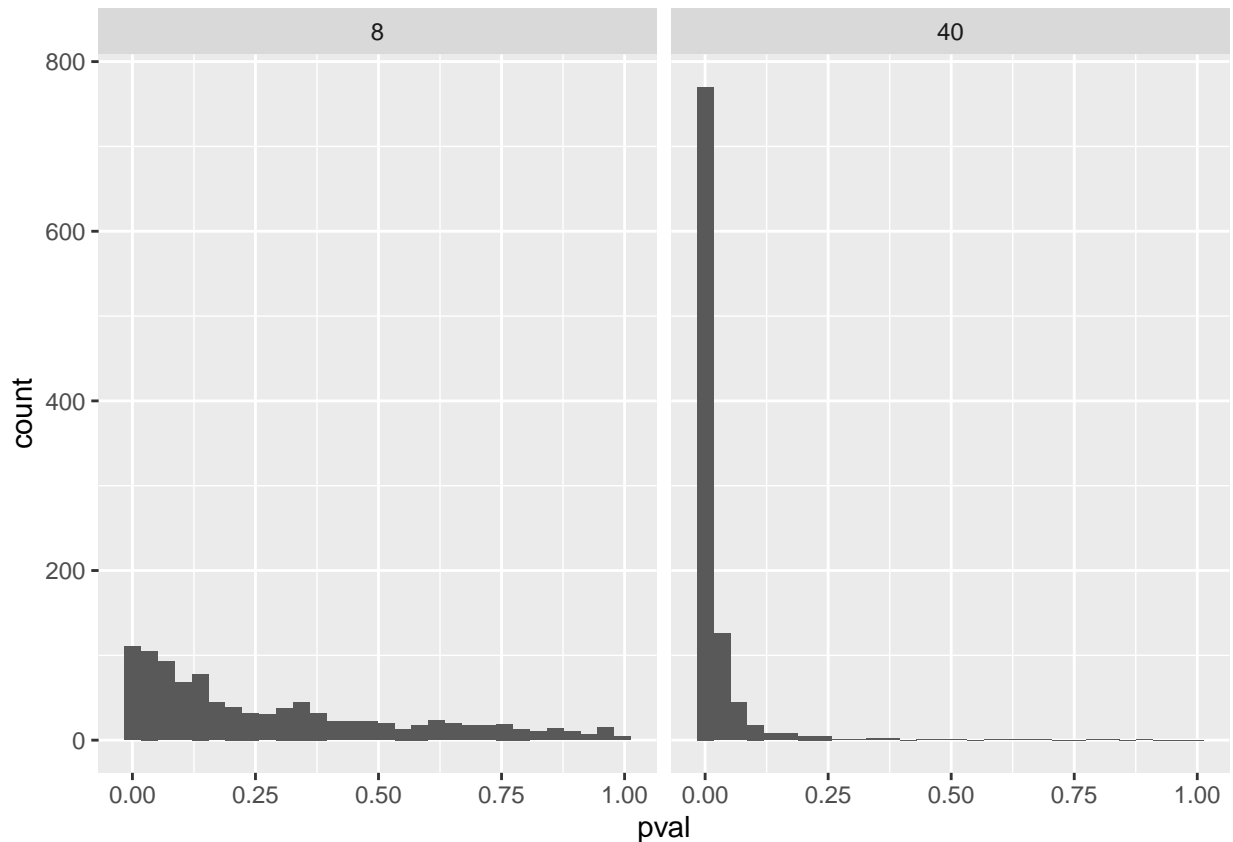
```
## # A tibble: 2 x 3
##       n estimate_sd se_mean
##   <dbl>      <dbl>   <dbl>
## 1     8        1.08     1.02
## 2    40        0.454    0.469
```

- b. Forklar hvorfor disse verdiene er veldig like? Hvordan kan vi definere standardfeil i lyset av disse kalkulasjonene? Standardavvik er variasjonen i de gjennomsnittlige verdiene i et utvalg. Standardfeil er gjennomsnittet til et estimat av standardavviket for en populasjon. Standardfeil beregnes ut ifra standardavviket og utvalgsstørrelsen.

$$SE = \frac{sd}{\sqrt{n}}$$

Standardavviket til utvalget samsvarer med standardfeilen som skal si noe om populasjonen. Vi gjorde forsøket 1000 ganger og dette jevner ut standard avviket til utvalget. Standardfeilen ble ganske lik standardavviket som vil si at utvalget er representativt for populasjonen.

5. Lag et histogram for p-verdiene i hvert av utvalgsstørrelsene. Hvordan kan vi tolke disse histogrammene, og hva forteller de om hvilken effekt utvalgsstørrelsen har på statistisk effekt?



Ut ifra histogrammene ser vi at utvalgsstørrelsen med 8 deltakere gir en større spredning i p-verdiene, og at dette er til forskjell fra en utvalgsstørrelse med 40 deltakere der de fleste studier får en statistisk signifikant p-verdi. Dette betyr at økt utvalgsstørrelse gir større sannsynlighet for at studier finner en sann effekt, og at utvalgsstørrelse dermed vil påvirke statistisk effekt.

6. Kalkuler antall studier fra hver utvalgsstørrelse som detekterer en statistisk signifikant effekt (spesifiser en grenseverdi for alpha)

```
## # A tibble: 2 x 2
##       n sig_results
##   <dbl>   <dbl>
## 1     8     0.213
## 2    40     0.895
```

For utvalgsstørrelse 8 viser 22,7 % av studiene en statistisk signifikant effekt, med grenseverdi for alpha på 0.05. Utvalgsstørrelse 40 finner derimot at 86,5 % av studiene har en statistisk signifikant effekt. Jo større utvalgsstørrelse, jo større effekt får vi.

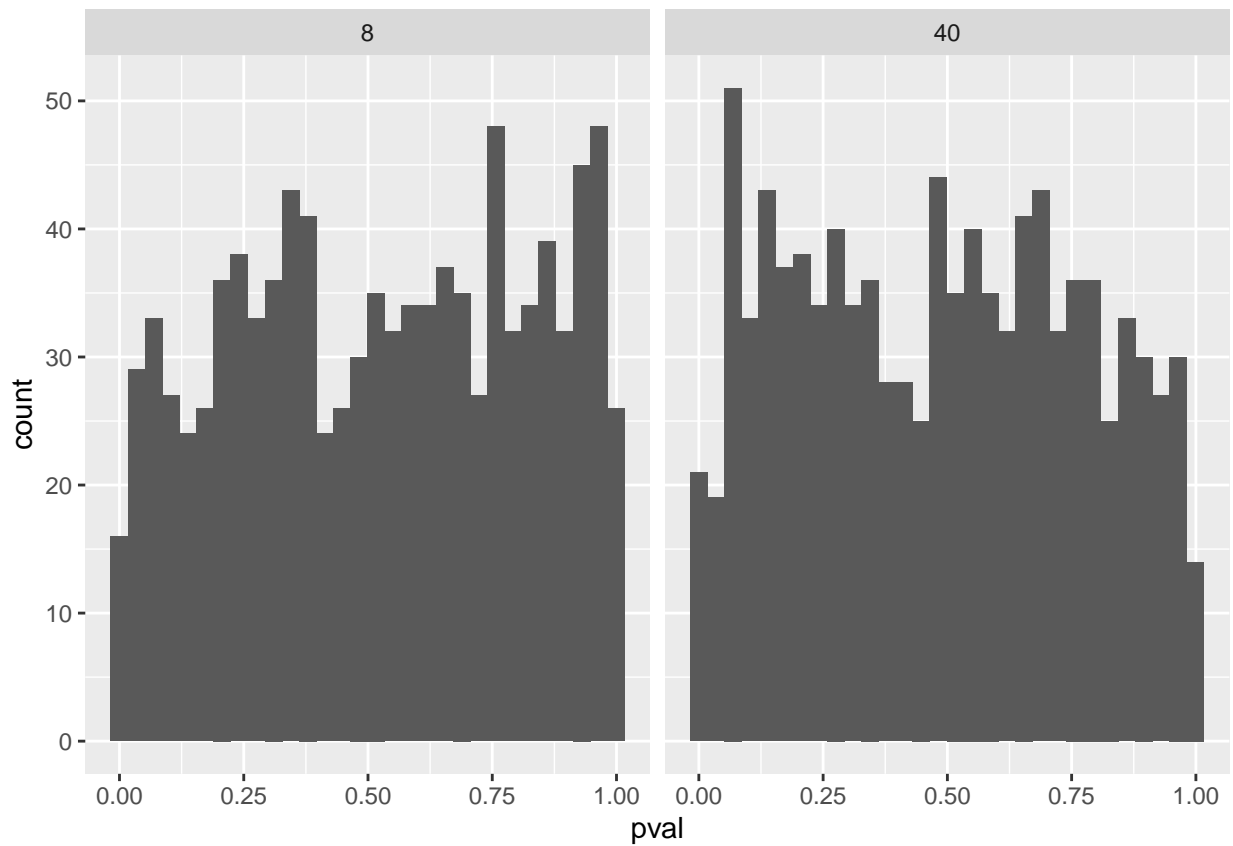
7. Bruk pwr package og kalkuler effekten av en ensidig t-test, med en effektstørrelse på 1.5/3, det spesifikke signifikansnivået og utvalgsstørrelse 8 og 40. Forklar resultatene i lyset av simuleringene.

```
##
##      One-sample t test power calculation
##
##              n = 8
##              d = 0.5
##      sig.level = 0.05
##      power     = 0.232077
##      alternative = two.sided

##
##      One-sample t test power calculation
##
##              n = 40
##              d = 0.5
##      sig.level = 0.05
##      power     = 0.8693981
##      alternative = two.sided
```

Ut ifra simuleringene vi har gjort kan vi si igjen at utvalgsstørrelse er med på å bestemme den statistiske effekten man finner. Utvalgsstørrelsen med 40 deltakere gir en høyere effekt enn utvalgsstørrelsen med 8 deltakere.

Mange studier uten populasjonseffekt



8. Med et signifikansnivå på 5 %, hvor mange studier gir oss et “falskt positiv” resultat hvis man repeterer studiene mange ganger?

```
## # A tibble: 2 x 2
##       n sig_results
##   <dbl>   <dbl>
## 1     8     0.045
## 2    40     0.039
```

4,5 % av studiene med utvalgsstørrelse 8 gir oss et “falskt positivt” resultat, mens med utvalgsstørrelsen på 40 vil 3,9 % av studiene gi oss et “falskt positivt” resultat.