7.77 Støveksponering blant tunnelarbeidere

Sammenlignet $n_1 = 115$ bor- og sprengningsarbeidere med $n_2 = 220$ utendørs sementarbeidere.

Eksponeringen ble målt i $mg \cdot y / m^3$

$\bar{x}_1 = 18.0, \ sd_1 = 7.8$

$\bar{x}_2 = 6.5, \ sd_2 = 3.4$

a) Sample: Alle fra et firma, undersøkt i forbindelse med rutinemessig helsesjekk: OK, ingen skjevhet i utvalget fra dette firmaet. Hvis dette firmaet er representativt for alle slike firmaer, er utvalget omtrent SRS: Simple random sample og resultatene kan brukes om alle slike arbeidere. Men hvis dette firmaet er spesielt med beskyttelsesutstyr, vil disse resultatene ikke være generaliserbare til andre.

b) 95% KI for forskjell i eksponering

μ_1 = Forventet eksponering for bore/sprenge-arbeidere

μ_2 = ——"—— utendørs-arbeidere

$\mu_1 - \mu_2$ er forventet forskjell i eksponering

95% KI for $(\mu_1 - \mu_2)$:

Vi vet at

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}} \sim T_{df}$$

$df = 115 - 1 = 114$

eller, hvis $sd_1 = sd_2$:

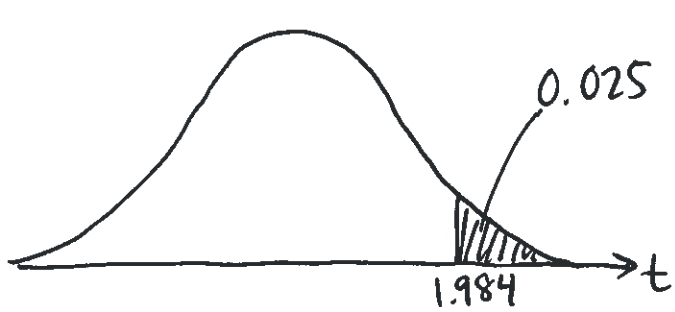
$df = 220 + 115 - 2 = 333$

Velger $df = 114$

T_{114} finnes ikke i Table D. Det nærmeste er T_{100}

I en T_{100}-fordeling er

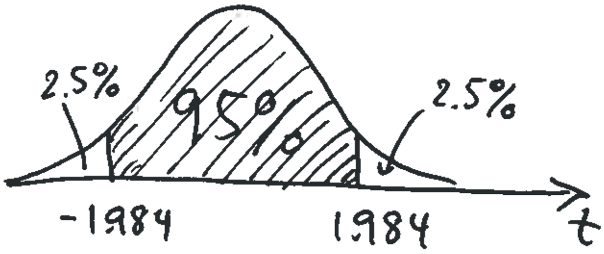
$P(t > 1.984) = 0.025$



Det betyr at

$$P(-1.984 < t < 1.984) = 0.95$$

(fordi alle T-fordelinger er symmetriske)



Dermed:

$$P\left(-1.984 < \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{sd_1^2}{n_1}+\frac{sd_2^2}{n_2}}} < 1.984\right) = 0.95$$

$$P\left(-1.984\cdot\sqrt{\frac{sd_1^2}{n_1}+\frac{sd_2^2}{n_2}} < (\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2) < 1.984\cdot\sqrt{\frac{sd_1^2}{n_1}+\frac{sd_2^2}{n_2}}\right) = 0.95$$

$$P\left((\bar{x}_1-\bar{x}_2)-1.984\cdot\sqrt{\frac{sd_1^2}{n_1}+\frac{sd_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2)+1.984\sqrt{\frac{sd_1^2}{n_1}+\frac{sd_2^2}{n_2}}\right) = 0.95$$

→ 95% KI for $\mu_1-\mu_2$:

$$(\bar{X}_1-\bar{X}_2) \pm 1.984\cdot\sqrt{\frac{sd_1^2}{n_1}+\frac{sd_2^2}{n_2}}$$

$$(18-6.5) \pm 1.984\cdot\sqrt{\frac{7.8^2}{115}+\frac{3.4^2}{220}} \qquad 18-6.5 = \underline{11.5}$$

$$\underline{[9.99\ ,\ 13.01]}$$

Tolkning: Vi estimere forskjellen til å være 11.5, altså at sprøytearbeiderne utsettes for 11.5 enheter mer støv enn utendørsarbeiderne. Vi kan ikke utelukke at det er enda større forskjell, inntil 13 enheter, eller litt mindre forskjell, 10 enheter, men med 95 % sikkerhet vil intervallet [9.99, 13.01] dekke den samme forskjellen i eksponering.

c) Test hypotesene:

$$\left.\begin{array}{l} H_0 \quad \mu_1 = \mu_2 \\ H_1 \quad \mu_1 > \mu_2 \end{array}\right\}$$

Ensidig:
Hvis vi kan bruke det vi vet til å anta at det må minst være likemye støv inne i tunnelen.

Eller

$$\left.\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array}\right\}$$

Tosidig:
Hvis det ikke er soleklare grunner til å tro at den ene gruppa er høyere. Dette er en mer konservativ tilnærming, og brukes nesten alltid innen medisin. Jeg velger denne.

Testobservator:

Når Ho er sann, er denne 0

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}} \sim T_{114}$$

$$t = \frac{11.5}{0.76} = \underline{15.1}$$

p-verdi = P(minst like ekstreme obs. som det vi har obs | Ho)

$$= P(|t| > 15.1 \mid t \sim T_{114}) \approx 0$$

Denne forskjellen er supersignifikant, og absolutt ikke i overensstemmelse med Ho. Ho forkastes. Det så vi også i oppgave b), der 95% KI absolutt ikke inneholdt Ho-verdien 0

d) Originaldata oppgis å være "litt skjevfordelte". T-testen er relativt robust for "litt skjevhet", og når n er såpass stor som her, burde ikke det være noe problem.

7.85 Åpne datasettet (fra boka) i Excel, og lagre som .CSV

RStudio : ▦ Import dataset
From local file
Finn csv-fila og trykk import

a)+b)

```
> Attach (ex07.85nspines)
> boxplot (dbh ~ ns)
> hist (dbh[ns == "n"])
> qqnorm( ——— )
> hist (dbh[ns == "s"])
> qqnorm( ——— )
```

kanskje ikke helt "relativt symmetrisk", men litt skjeve til hver sin side
→ t-test ok? siden skjevheten er såpass liten, og det ikke er noen superekstreme verdier.

Men her er jeg litt i tvil, og vil se om WRS-test gir samme konklusjon.

c) $H_0 : \mu_n = \mu_s$
$H_1 : \mu_n \neq \mu_s$ } hvis t-test

H_0 : like store trær n & s
H_1 : ikke like store n & s

Velger tosidig fordi jeg ikke vet noenting om trær.
(OBS: kan ikke se på data først og så velge ensidig ut i fra hva data sier. Det er juks.)

d) t-test:

```
> t.test (dbh ~ ns)
```

$t = -2.63$, $df = 55.7$

$p = 0.01$ Forkast H_0

$\bar{x}_n = 23.7$
$\bar{x}_s = 34.5$
95% KI : [-19.1, -2.6]

WRS-test:

```
> wilcox.test (dbh ~ ns)
```

$W = 281.5$

$p = 0.013$ Forkast H_0

Samme konklusjon.
Mer informasjon fra en t-test, derfor bruker jeg den.

7.89 Amming: $n_1 = 23$ $\bar{x}_1 = 13.3$, $sd_1 = 1.7$
Morsmelkserstatning: $n_2 = 19$ $\bar{x}_2 = 12.4$, $sd_2 = 1.8$

a) $H_1: \mu_1 > \mu_2$ Dette er påstanden vi skal undersøke

$H_0: \mu_1 \leq \mu_2$ Da blir H_0 slik.

c) Hva må antas for å gjøre en t-test?
Her er n_1 og n_2 små. Da må data være relativt normalfordelte i hver gruppe, og vi må ha

uavhengige observasjoner,

som er tilfeldige utvalg fra ammere og ikke-ammere.
Her antar jeg også at $\sigma_1 = \sigma_2$, så jeg kan bruke Spooled

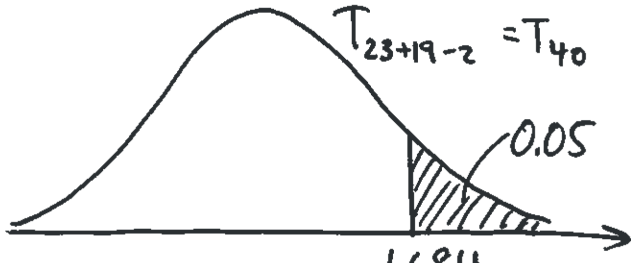
b)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot S_{pooled}} \sim T_{n_1+n_2-2}$$

$$\downarrow$$

$$S_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = 1.75$$

$$t = \frac{13.3 - 12.4}{0.54} = 1.66$$



Her måtte t-verdien vært > 1.684 for at vi skulle forkastet H_0

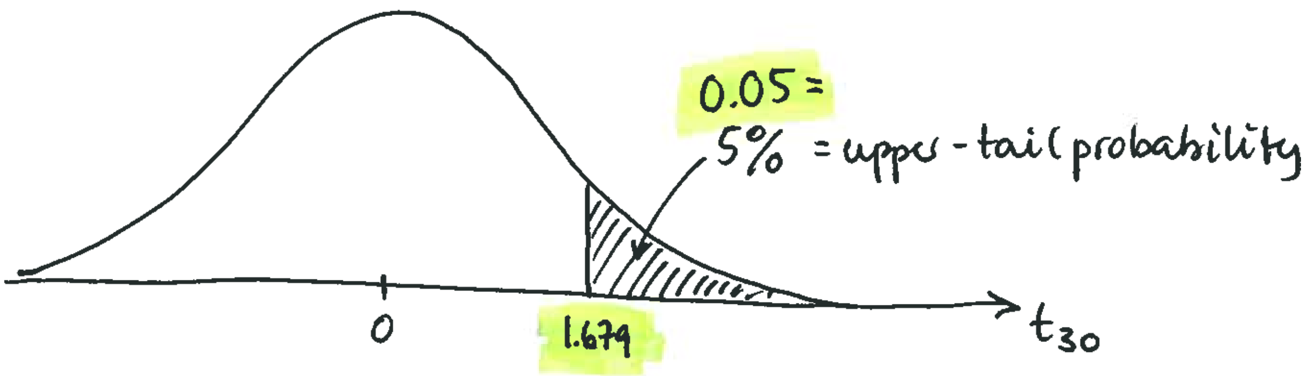
p-verdien er så vidt under 0.05.

95% KI for $\mu_1 - \mu_2$:

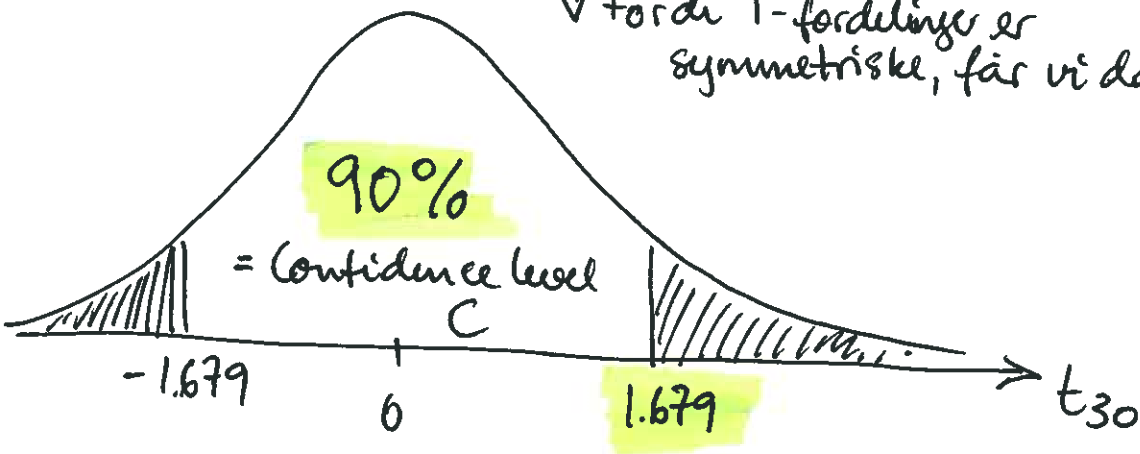
$$(\bar{x}_1 - \bar{x}_2) \pm \underbrace{2.021}_{\text{Fra tabell D, upper p = 0.025}} \cdot 0.54 \rightarrow [-0.19, 1.99]$$

Kan mao ikke utelukke at forskjellen er 0. (0 er i intervallet)

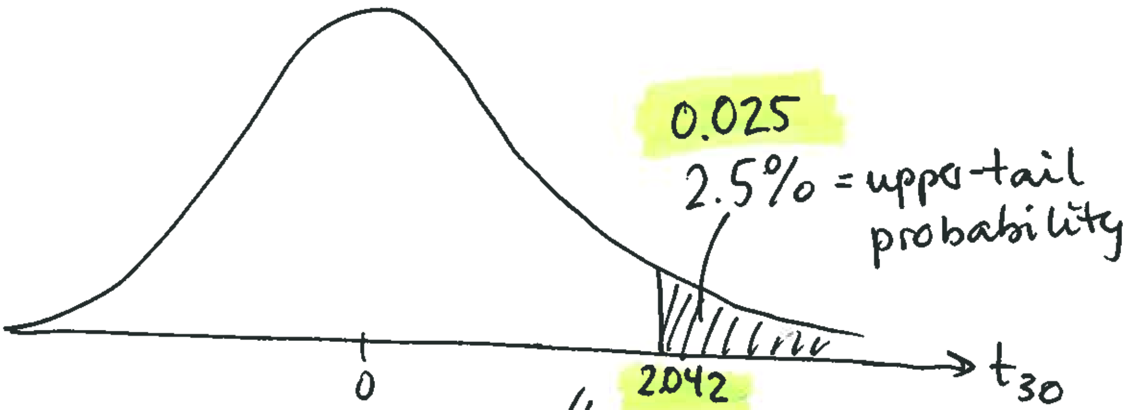
Disse tallene står i Table D



Fordi T-fordelinger er symmetriske, får vi da at



Tilsvarende:



Ergo