

STK1000: Løsningsforslag Uke 38

2022

Oppgave 3.16

- Denne metoden vil ikke nødvendigvis gi to grupper (alle kan havne i den ene gruppen). Man kan i stedet bruke denne metoden frem til vi har en gruppe på 10 og la resten være i den andre gruppen. Husk også at man ikke må ha noen ordning på individene (ikke start med alle kvinne).
- De 10 rottene i hver gruppe kan ha noen fellestrekk. For eksempel kan de ha blitt eksponert for noe under frakt som de andre gruppene ikke har blitt eksponert for. De er derfor bedre å behandle hver gruppe som en blokk og gjøre randomiseringen i hver gruppe.
- Å dele inn etter kjønn er ikke tilfeldig. Her ville det vært bedre med et blokk-design hvor man bruker kjønn som blokk-variabel.

Oppgave 3.17 (bruk R for å løse (c))

a)

Faktorene er dosene med kalsium og D-vitamin. Behandlingene er de 6 mulige kombinasjonene av kalsium og D-vitamin.

b)

Vi har 30 individer av hvert kjønn og 6 mulige behandlinger. Vi vil derfor bruke kjønn som blokkings variabel og tildele fem tilfeldige individer av hvert kjønn til hver behandling. Figur 1 viser oppsettet.

c)

Koden under viser hvordan vi utfører randomiseringen for menn. Man kan gjenta samme kode for å randomisere for kvinner også. Under skriver vi ut de 10 første mennene og hvilken behandling de skal få.

```
treatment = rep(1:6, 5) # Lager en liste over alle behandlingene til alle mennene
treatment = sample(treatment) # Gir listen en tilfeldig rekkefølge
treatment = data.frame(treatment) # For pen utskrift
head(treatment, 10)
```

```
##      treatment
## 1             1
## 2             1
## 3             4
## 4             2
## 5             6
## 6             1
## 7             2
## 8             4
## 9             3
## 10            5
```

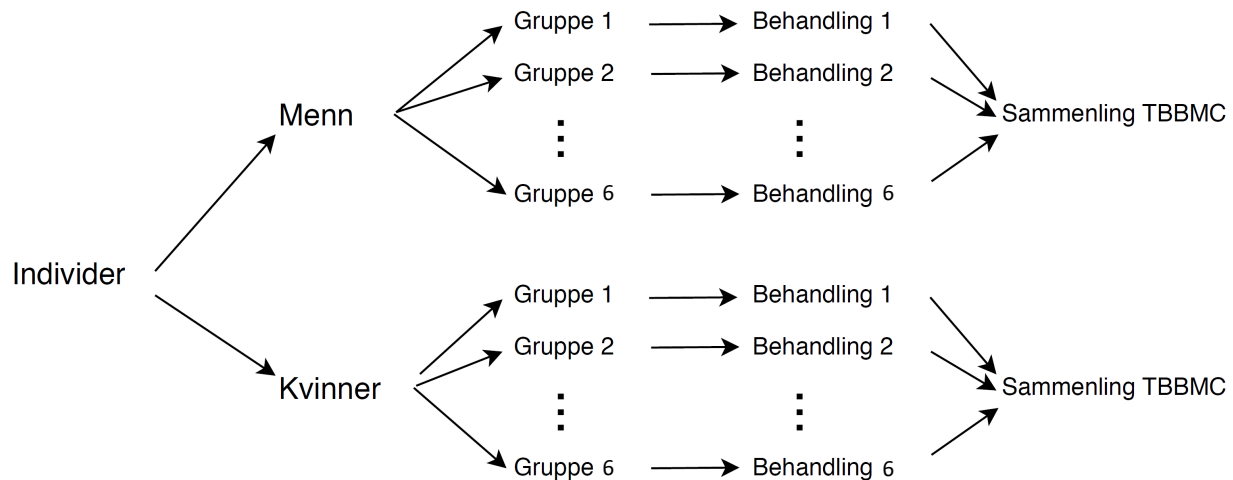


Figure 1: Oppsett av forsøk i Oppgave 3.17 b)

Extra: Listen over blir veldig lang hvis vi skal ta med alle 30 mennene. Vi kan derfor bryte det opp i en `data.frame` med en kolonne per behandling. Altså, i tabellen under viser vi ID til alle de 30 mennene (5 menn for hver av de 6 behandlingene).

```

id = 1:30 # lager ID'er
id = split(id, treatment) # splitter opp ID'ene etter behandling
id = data.frame(id) # Samler i data.frame for pen utskrift
names(id) = 1:6 # Gir navn på kolumnene
id

```

```

##   1  2  3  4  5  6
## 1  1  4  9  3 10  5
## 2  2  7 19  8 12 13
## 3  6 11 25 16 14 15
## 4 20 18 27 23 17 24
## 5 28 21 30 26 22 29

```

Sammenlign de to tabellene og se at du finner igjen de samme individene (id=1 får behandling 2 så han er i kolonne 2, etc.)

d)

Ja, gruppen som får 0 mg av både kalsium og D-vitamin kan regnes som en placebo gruppe.

Oppgave 3.23 (bruk R for randomisering i (b))

a)

Del jentene tilfeldig i to grupper med 20 jenter i hver gruppe. Den ene gruppen vil få et høyt nivå av kalsium mens den andre gruppen vil få et lavt nivå av kalsium. Responsvariabelen er ikke definert i oppgaven, så vi får si at vi måler "hvordan kalsium er tatt opp i kroppen".

b)

Vi deler de 40 jentene inn i to grupper med 20 i hver gruppe. Det er forskjellige måter å gjøre dette på, men her er et forslag

```
treatment = c(rep('low', 20), rep('high', 20)) # Lager 15 'low' og 15 'high'  
treatment = sample(treatment) # Gjør rekkefølgen tilfeldig
```

c)

Under lager vi en tabell hvor radnummer gir ID til hver jente, og `treatment` gir dosering.

```
data.frame(treatment)
```

```
##      treatment  
## 1         low  
## 2         low  
## 3         low  
## 4         high  
## 5         low  
## 6         low  
## 7         high  
## 8         low  
## 9         high  
## 10        low  
## 11        high  
## 12        low  
## 13        low  
## 14        low  
## 15        high  
## 16        low  
## 17        low  
## 18        low  
## 19        high  
## 20        high  
## 21        low  
## 22        high  
## 23        high  
## 24        low  
## 25        high  
## 26        high  
## 27        low  
## 28        low  
## 29        high  
## 30        high  
## 31        low  
## 32        high  
## 33        high  
## 34        high  
## 35        low  
## 36        high  
## 37        high  
## 38        high  
## 39        low  
## 40        high
```

Extra: For penere output kan vi gjøre følgende:

```
id = data.frame(split(1:40, treatment))
id
```

```
##      high low
## 1         4  1
## 2         7  2
## 3         9  3
## 4        11  5
## 5        15  6
## 6        19  8
## 7        20 10
## 8        22 12
## 9        23 13
## 10       25 14
## 11       26 16
## 12       29 17
## 13       30 18
## 14       32 21
## 15       33 24
## 16       34 27
## 17       36 28
## 18       37 31
## 19       38 35
## 20       40 39
```

Oppgave 3.24

- Dette kan gjøres på flere måter, men her er et forslag: Vi kan starte med å utføre eksperimentet i Oppgave 3.23 for og så bytte om doseringen til alle jentene (de som hadde “høy” får nå “lav”). For hver jente regner vi så ut differansen i responsen for dosering “høy” og “lav”.
- Fordelen med dette designet over designet i Oppgave 3.23 er at vi nå tar hensyn til variasjonen til hver enkelt individ.
- Som forklart i a) må vi også i denne oppgaven starte med å dele individene inn i to grupper. Når vi bytter dosering (fra lav til høy og fra høy til lav) skal vi ikke randomisere på nytt. Vi bruker derfor samme randomisering som i Oppgave 3.23).
- Som forklart i a) og c) kan vi bruke samme kode som i Oppgave 3.23.

Oppgave 3.26

- Populasjonen er de 50 dagene.
- Utvalget er de 5 dagene du har tilfeldig valgt.

Oppgave 3.30

- Populasjonen (*population*) er alle individene og ikke et tilfeldig utvalg av individer.
- “Dihydrogen monoxide” er H_2O , altså vanlig vann. Folk som svarer at de er bekymret har trolig ikke forstått hva “dihydrogen monoxide” er, men velger å ikke innrømme dette.
- Det er typisk vanskelig å få ærlige svar på slike spørsmål når individene ikke er anonyme. Man kan derfor ikke stole på resultatene.

Oppgave 3.32

- Innholdet i et enkelt kapittel er trolig på samme nivå, og gir ikke et bra utgangspunkt for å evaluere hele teksten. Trekk heller tilfeldige seksjoner fra hele teksten.
- Studenter som har fag kl 7:30 på morgenen har kanskje noen felleskarakteristikk som skiller dem studenter som ikke har fag så tidlig (f.eks. hvis noen velger fag basert på at de ikke må stå opp før kl 10). Velg heller flere ulike tidspunkt i løpet av dagen.
- Alfabetisk ordning er ikke nødvendigvis tilfeldig. Det kan for eksempel inneholde individer med samme navn som søsken og ektepar. I tillegg vil det være noen etternavn knyttet til etniske grupper, noe som kan gi unaturlig representasjoner av disse gruppene. Trekk heller tilfeldig fra hele populasjonen.

Oppgave 3.35 (bruk R for trekke tilfeldig utvalg)

Datasettet er ikke tilgjengelig fra boken, men det er en lenke på øvingsiden til datasettet. Vi trenger ikke å laste det ned ettersom `scan` (og `read.csv`) også virker på nettadresser. I koden under henter vi inn datasettet og trekker 8 tilfeldige leilighetskomplekser og skriver de ut fire og fire.

```
url = 'https://www.uio.no/studier/emner/matnat/math/STK1000/h18/rkoder/data_oppg_3-59.txt'
data = scan(url, what='character')
my_sample = sample(data, 8)
my_sample[1:4]
```

```
## [1] "Chauncey Village" "Country View"      "Sagamore Ridge"    "Cambridge"
```

```
my_sample[5:8]
```

```
## [1] "Salem Courthouse" "Country Square"    "Mayfair Village"  "Nobb Hill"
```

Oppgave 3.36 (bruk R for trekke tilfeldig utvalg)

Oppgaven her er å trekke 12 tilfeldige heltall fra 1 til 100. Vi kan gjøre det i R med få følgende måte

```
pixels = 1:100
sample(pixels, 12)
```

```
## [1] 35 40 62 24 98 53 7 97 59 8 16 78
```

Oppgave 4.6

Svarene her varierer etter utfallene fra eksperimentene.

Oppgave 4.10

- Sannsynligheter kan ikke være større enn 1.
- Sannsynligheter kan ikke være mindre enn 0. $P(\text{ikke } A) = 1 - P(A) = 0.45$.
- Sannsynligheten for at begge de to disjunkte hendelsene inntreffer er per definisjon 0.

Oppgave 4.17

- $P(O) = 1 - (P(A) + P(B) + P(AB)) = 1 - 0.42 - 0.11 - 0.03 = 0.44$.
- $P(O \cup A) = P(O) + P(A) = 0.42 + 0.44 = 0.86$ siden A og O er disjunkte.

Oppgave 4.18

Vi kaller det amerikanske individet X og det Irske individet Y . Siden hendelsene er uavhengige har vi at

$$P(\text{begge } O) = P(X = O)P(Y = O) = 0.44 \cdot 0.52 = 0.2288.$$

Videre har vi at sannsynligheten for at begge har samme type er

$$\begin{aligned} P(\text{begge samme}) &= P(\text{begge } A) + P(\text{begge } B) + P(\text{begge } AB) + P(\text{begge } O) \\ &= 0.1470 + 0.0110 + 0.0009 + 0.2288 \\ &= 0.3877. \end{aligned}$$

Oppgave 4.20

a) $P(?) = 1 - (0.56 + 0.21) = 0.23.$

b) $P(\text{not English}) = 1 - 0.56 = 0.44.$

Oppgave 4.23

La X betegne blodtype (A, B, O) og la Y betegne rhesus $+$ og $-$. Siden X og Y er uavhengige er $P(X \cap Y) = P(X) \cdot P(Y)$. Vi får derfor følgende sannsynligheter for blodtypene

Blodtype	A+	A-	B+	B-	AB+	AB-	O+	O-
Sannsynlighet	0.352	0.067	0.092	0.018	0.025	0.0048	0.370	0.070

Oppgave 4.27

Når man spør om minst én, bruker man stort sett den samme fremgangsmåten (så lær deg dette!). Vi begynner med å regne ut sannsynligheten for at en tilfeldig valgt person IKKE har blodtype O , altså $P(\text{ikke } O) = 1 - 0.07 = 0.93$. Siden vi har 6 uavhengige individer, har vi at sannsynligheten for at ingen av dem har blodtype O er

$$P(\text{ingen har } O) = 0.93^6 = 0.647$$

Vi kan nå svare på oppgaven

$$P(\text{minst én har } O) = 1 - P(\text{ingen har } O) = 1 - 0.647 = 0.353$$

Midtveiseksamen Våren 2006

10) d).

Midtveiseksamen Høsten 2008

1) d).

2) b).

3) Vi tar $Q_3 - Q_1$ som gir a).

4) $Q_3 + 1.5 IQR$ som gir c).

5) $P(X < 10) = P(Z < \frac{10-15}{5}) = P(Z < -1) = 0.1587$ som gir c).

6) Vi ønsker å finne $P(H \geq h) = 0.05$ og må begynne med en standard normalfordeling der $Z = \frac{H-\mu}{\sigma}$. Vi har da at $P(Z \geq z) = 0.05$ er det samme som $P(Z < z) = 0.95$. Da vet vi at $z = 1.6449$ (tabell eller R). Vi transformerer så dette til høydeskalaen $h = z \cdot 6.5 + 170 = 180.69$ som gir d).

7) b)

8) d)

9) b)

10) Predikert verdi er $\hat{y} = 97.20 + 0.03283 \cdot 190.7 = 103.46$. Residualet blir derfor $y - \hat{y} = 115 - 103.46 = 11.54$. Dette gir c).