STK1000: Løsningsforslag Uke 38

H2023

Oppgave 3.16

a) 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).

b) 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.

c) 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 3.23 (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

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

```
 [1] "low"  "high" "high" "high" "high" "low"  "low"  "low"  "low"  "high"
[11] "low"  "high" "high" "low"  "low"  "low"  "low"  "high" "low"  "high"
[21] "low"  "low"  "high" "low"  "low"  "high" "high" "high" "high" "high"
[31] "high" "high" "high" "low"  "low"  "low"  "low"  "high" "low"  "high"
```

c)

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

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

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

Oppgave 3.24

a) 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".

b) Fordelen med dette designet over designet i Oppgave 3.23 er at vi nå tar hensyn til variasjonen til hver enkelt individ.

c) 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).

d) Som forklart i a) og c) kan vi bruke samme kode som i Oppgave 3.23.

Oppgave 3.26

a) Populasjonen er de 50 dagene.

b) Utvalget er de 5 dagene du har tilfeldig valgt.

Oppgave 3.32

a) 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.

b) Studenter som har fag kl 7:30 på morgenen har kanskje noen felleskarakteristikker 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.

c) Alfabetisk ordning er ikke nødvendigvis tilfeldig. Det kan for eksempel inneholde individer med samme navn som søsken og ektepar. I tillegg er noen etternavn knyttet til etniske grupper, noe som kan gi unaturlig representasjoner av disse gruppene. Trekk heller tilfeldig fra hele populasjonen.

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] 51 25 27 75 73 17 99 77 23 44 49 32
```

Oppgave 4.6

Svarene her varierer etter utfallene fra eksperimentene.

Oppgave 4.17

a) $P(O) = 1 - (P(A) + P(B) + P(AB)) = 1 - 0.42 - 0.11 - 0.03 = 0.44.$

b) $P(O \cup A) = P(O) + P(A) = 0.42 + 0.44 = 0.86$ 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 4.23

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 viktig å lære seg 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) Se neste uke.
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).