

Opgavesættet indeholder fire opgaver med i alt 11 delopgaver.  
Held og lykke!

### Opgave 1

Hydroxymethylfuran (HMF) er en substans, der opstår ved nedbrydning af fruktose. HMF findes blandt andet i honning; i frisk bihonning er koncentrationen forholdsvis lav. Erfaringen siger, at HMF koncentrationen stiger med lagring, og når honning varmes. Kunstig honning indeholder store mængder af HMF, derfor kan HMF koncentrationen bruges både som indikator af honningskvalitet og friskhed, og for at opspore honning, der sælges under falsk betegnelse.

Forsker i fødevarer kemi vil gerne finde ud af, om man allerede efter to måneder ser en stigning i HMF koncentrationen. Derfor målte de HMF koncentrationen i fjorten glas honning, frisk fra biavlaren, og igen efter to måneders opbevaring ved stuetemperatur.

Data er i filen *honning.csv* med søjlerne *honeysample*, der betegner honningglasset, *HMFstart*: koncentrationen i mg/kg ved første måling og *HMFend*: koncentrationen efter to måneder.

- (a) Lav et passende permutationstest for nulhypotesen, at den gennemsnitlige HMF koncentration ikke ændrer sig. Angiv verbalt den alternative hypotese, du tester imod. Husk at angive den faglige konklusion.

### Opgave 2

I en biologisk undersøgelse i det nordvestlige Grønland har man på 5 tidspunkter hen over sommeren indsamlet i alt 25 muslinger (5 muslinger til hvert tidspunkt) og målt indholdet af protein (mg per gram tørvægt). Data er i filen *musling.csv*, der har tre søjler: *Dato* der angiver de fem tidspunkter, *Tid* der angiver tidspunkt som antal dage fra det første tidspunkt, og *Protein*, der angiver proteinindholdet.

- (a) Opskriv modellen  $M_0$ , hvor proteinindholdet er normalfordelt, og hver gruppe (hvert tidspunkt) har sin egen middelværdi og varians på proteinindholdet. Undersøg, om det kan antages, at varianserne er ens.
- (b) Lav grafiske undersøgelser for at se på forholdet mellem de fem grupper og for at se, om det er rimeligt at beskrive proteinindholdet med en normalfordeling, hvor hver gruppe har sin egen middelværdi af proteinindholdet og varianserne er ens.

Forklar dit valg af grafiske metoder.

- (c) Opskriv modellen  $M_1$ , hvor proteinindholdet er normalfordelt, hver gruppe har sin egen middelværdi og alle grupperne har samme varians. Lav, under model  $M_1$ , et konfidensinterval for forskel i middelværdi mellem gruppe 1 og gruppe 5 (svarende til de to tidspunkter 6/7 og 9/9).
- (d) Opskriv modellen  $M_2$ , hvor proteinindhold er normalfordelt og middelværdien afhænger lineært af tiden. Lav et  $F$ -test for reduktion fra model  $M_1$  til model  $M_2$ .
- (e) Lav, under model  $M_2$ , et konfidensinterval for forskel i middelværdi mellem to tidspunkter, der ligger 65 dage fra hinanden. Sammenlign med konfidensintervallet i spørgsmål (c).

### Opgave 3

Data i tabellen nedenfor viser for 200 kvinder, hvor mange der er blevet gravide efter første, andet og tredje forsøg med ivf-behandling, og hvor mange der ikke er blevet gravide i de tre forsøg (data er opdigtede).

|       | Første | Andet | Tredje | Ikke-gravide |
|-------|--------|-------|--------|--------------|
| Antal | 29     | 21    | 21     | 129          |

- (a) Opskriv en statistisk model for antallet af kvinder i de fire kategorier *Første*, *Andet*, *Tredje* og *Ikke-gravide*.
- (b) Betragt hypotesen (fast-rate-hypotesen), at sandsynligheden for at blive gravid er den samme i hvert forsøg. Hvis  $\theta$  er sandsynligheden for at blive gravid i det enkelte forsøg, er sandsynligheden for at blive gravid i det første forsøg  $\theta$ , gravid i det andet forsøg  $(1 - \theta)\theta$ , gravid i det tredje forsøg  $(1 - \theta)^2\theta$ , og sandsynligheden for ikke at blive gravid  $(1 - \theta)^3$ .

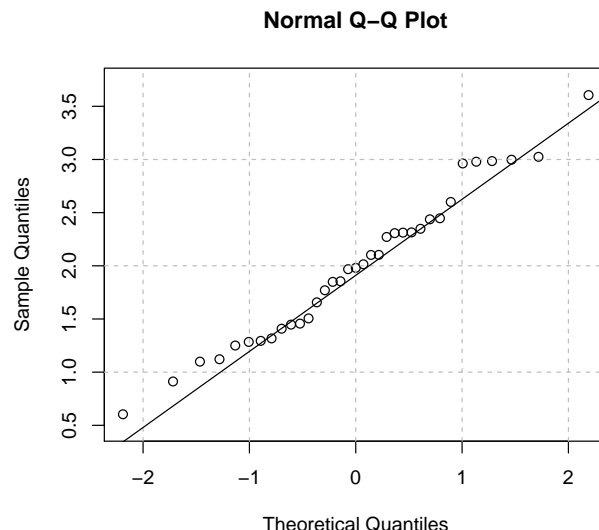
Opstil likelihoodfunktionen under hypotesen og find et udtryk for  $\hat{\theta}$ . Eftervis, at for data i tabellen ovenfor er  $\hat{\theta} = 0.1363$ .

- (c) Lav et test for fast-rate-hypotesen.

### Opgave 4

- (a) Figuren nedenunder viser normalfraktilplot af en stikprøve  $(x_1, \dots, x_{35})$  fra en normalfordeling  $N(\mu, \sigma^2)$ , samt indtegnet linje via `qqline` i R. Forklar kort, hvordan (eller om) du kan finde variansen  $\sigma^2$  ud fra linjen i figuren. Angiv, på denne baggrund, hvilket udsagn bedst passer til plottet:

- (A)  $\sigma^2 = 1.0$ ,  
(B)  $\sigma^2 = 0.7$ ,  
(C)  $\sigma^2 = 0.5$ ,  
(D)  $\sigma^2 = 2.0$ ,  
(E)  $\sigma^2 = 4.0$ ,  
(F) ingen af de ovenstående værdier passer,  
(G) det er umuligt at skønne variansen ud fra linjen i figuren.



- (b) En statistiker undersøger et kompliceret hypotesetest, baseret på teststørrelsen  $T$ , ved hjælp af simulation. Testet forkaster nulhypotesen når  $T$  er mindre end eller lig med en kritisk værdi  $C$ . Statistikerens har gemt `Nsim` udfald af  $T$ , som var simuleret under nulmodellen, i en vektor i R med navn `T0`. Dernæst har statistikerens simuleret lige så mange udfald under en bestemt alternativhypotese  $H_A$  og gemt disse i vektoren `T1`. Variablen `crit` indeholder den kritiske værdi.

Angiv R-programlinjer der beregner skøn over sandsynligheden for type 1 fejl og type 2 fejl fra de simulerede data, og gemmer værdierne henholdsvis som `type1fejl` og `type2fejl`.