



Norges teknisk-naturvitenskapelige universitet  
Institutt for matematiske fag

TMA4245 Statistikk  
Vår 2023

Skriftlig innlevering 6

### Oppgave 1

Baker Bollesen baker boller med rosiner og reklamerer med at forventet antall rosiner i hver bolle er 10. Olav er en trofast kunde hos Bollesen og kjøper flere boller hver dag. Olav er svært glad i rosiner, men har den siste tiden fått inntrykk av at det er blitt færre rosiner og sukater i bollene til Bollesen enn tidligere. Han bestemmer seg for å regne litt på situasjonen. Olav lar  $X$  betegne antall rosiner i en tilfeldig valgt bolle. Han antar at  $X$  er poissonfordelt med  $E(X) = \lambda_R$ .

For å undersøke om det virkelig er så mange rosiner i bollene som det reklameres med, kjøper Olav  $n = 50$  boller og teller omhyggelig opp antall rosiner i hver av disse bollene. La  $X_1, X_2, \dots, X_n$  betegne antall rosiner i hver av bollene og anta at antall rosiner i ulike boller er uavhengige.

- a) Vis at sannsynlighetmaksimeringsestimatoren (SME) for  $\lambda_R$  blir

$$\hat{\lambda}_R = \frac{1}{n} \sum_{i=1}^n X_i$$

Vis videre at

$$E(\hat{\lambda}_R) = \lambda_R \quad \text{og} \quad \text{Var}(\hat{\lambda}_R) = \frac{\lambda_R}{n}$$

Begrunn at  $\hat{\lambda}_R$  er tilnærmet normalfordelt med forventning og varians som gitt over.

I resten av oppgaven skal du benytte at  $\hat{\lambda}_R$  er tilnærmet normalfordelt. Olav ønsker å sjekke om hans observasjoner gir grunnlag for å påstå at bollene ikke inneholder så mange rosiner som reklamen lover.

- b) Formuler dette som et hypotesetestingsproblem og lag en test for dette formål med (tilnærmet) signifikansnivå  $\alpha$ .

Hva blir konklusjonen på testen når  $\alpha = 0.05$  og det totalt er 470 rosiner i de 50 bollene?

- c) Anta så at  $\lambda_R = 9$ , dvs. bollene inneholder færre rosiner enn hva reklamen hevder. I hvor mange boller,  $n$ , må da Olav sjekke rosinantallet for at en test med signifikansnivå  $\alpha = 0.05$  med sannsynlighet 0.9 skal avsløre at  $\lambda_R < 10$  (dvs. forkaste  $H_0$ )?

## Oppgave 2

Per er nylig ferdig med sine studier i Trondheim og på sin første arbeidsdag får han en interessant oppgave på laboratoriet. Når to væsker A og B blandes sammen, vil volumet av en ideell blanding av  $x$  deler A og  $1 - x$  deler B være lik volumet av  $x$  deler A pluss volumet av  $(1 - x)$  deler B. For en ikke-ideell blanding vil volumet avvike noe fra det ideelle tilfellet.

Per vil eksperimentelt bestemme avviket fra en ideell blanding når A og B blandes. La  $x_1, x_2, \dots, x_n$  være andeler av A som blandes med  $1 - x_1, 1 - x_2, \dots, 1 - x_n$  andeler B, og la  $Y_1, Y_2, \dots, Y_n$  være avviket i volum som måles fra en ideell blanding. Per antar at avviket er på formen

$$Y_i = ax_i(1 - x_i) + \epsilon_i, \quad i = 1, \dots, n,$$

der  $\epsilon_1, \dots, \epsilon_n$  er uavhengige normalfordelte støy-ledd med forventningsverdi lik null og kjent varians  $\sigma_0^2 = 0.025^2$  for alle  $i$ , slik at  $Y_i$  er normalfordelt med

$$E(Y_i|x_i) = ax_i(1 - x_i) \quad \text{og} \quad \text{Var}(Y_i|x_i) = \sigma_0^2, \quad i = 1, \dots, n$$

For en ideell blanding vil  $a$  være lik null.

Pers målinger gir følgende resultat:

$i$	1	2	3	4	5	6	7	8	9
$x_i$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$y_i$	-0.0561	-0.0369	-0.0651	-0.0578	-0.0941	-0.0684	-0.0546	-0.0247	-0.0072

Det oppgis at  $\sum_{i=1}^9 x_i(1 - x_i)y_i = -0.0945$  og  $\sum_{i=1}^9 (x_i(1 - x_i))^2 = 0.3333$ .

a) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for  $a$ ,  $\hat{a}$ , er

$$\hat{a} = \frac{\sum_{i=1}^n x_i(1 - x_i)Y_i}{\sum_{i=1}^n (x_i(1 - x_i))^2}.$$

b) Finn forventningsverdi og varians til  $\hat{a}$ .  
Hvilken fordeling har  $\hat{a}$ ? (Begrunn svaret.)

Per vil undersøke om det er grunn til å påstå at han har å gjøre med en ikke-ideell blanding.

c) Formuler dette som et hypotesetestingsproblem og lag en test for dette formål med signifikansnivå  $\alpha$ . Hva blir konklusjonen på hypotesetesten for dataene over og  $\alpha = 5\%$ ?

## Oppgave 3

En viktig vitenskapelig oppdagelse fant sted i 1929 da Edwin Hubble oppdaget at universet er ekspanderende. Hubble's tallmateriale bestod blant annet av;  $x_i =$  avstanden til galakse  $i$  (målt i millioner lysår), og  $y_i =$  hastigheten til galakse  $i$  (målt i 1000 km/s). Verdiene Hubble benyttet i en av sine analyser er som følger:

Det oppgis her at  $\sum_{i=1}^{11} x_i = 4185$ ,  $\sum_{i=1}^{11} y_i = 237.7$ ,  $\sum_{i=1}^{11} x_i^2 = 2685141$  og  $\sum_{i=1}^{11} x_i y_i = 152224$ .

Navn	Avstand, $x_i$	Hastighet, $y_i$
Virgo	22	1.2
Pegasus	68	3.8
Perseus	108	5.1
Coma Berenices	137	7.5
Ursa Major 1	255	14.9
Leo	315	19.2
Corona Borealis	390	21.4
Gemini	405	23.0
Bootes	685	39.2
Ursa Major 2	700	41.6
Hydra	1100	60.8

Hubble foreslo en modell for hastighet som funksjon av avstand på formen  $y = \beta x$ , der  $\beta$  senere har blitt kalt Hubble's konstant. En statistisk versjon av ligningen kan gis ved:

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, \dots, 11, \quad (3.1)$$

der  $\varepsilon_i$ ,  $i = 1, \dots, 11$ , er uavhengige og normalfordelte stokastiske variabler med forventning 0 og varians  $\sigma^2$ .

- a) Vi vil i første omgang finne en estimator for  $\beta$ .

Bruk minste kvadraters metode (method of least squares) til å estimere  $\beta$  med utgangspunkt i ligning (3.1), og vis at estimatoren for  $\beta$  da blir gitt ved  $\hat{\beta} = \frac{\sum_{i=1}^{11} x_i Y_i}{\sum_{i=1}^{11} x_i^2}$ . Regn ut estimatet for  $\beta$  basert på dataene over.

Finn også forventning og varians til  $\hat{\beta}$ .

- b) Anta at en annen galakse befinner seg en avstand  $x_0 = 900$  millioner lysår borte.

Finn predikert hastighet,  $\hat{y}_0$ , til denne galaksen.

Utlede et 95% prediksjonsintervall for en måling av hastigheten til denne galaksen. Det oppgis at  $\sum_{i=1}^{11} (y_i - \hat{y}_i)^2 = 9.87$ , der  $\hat{y}_i = \hat{\beta} x_i$ .

## Fasit

- b) Ikke forkast  $H_0$  c) Olav må sjekke 82 boller
- c) Forkast  $H_0$
- a) 0.0567 b) 51.03, (48.5, 53.5)