TMA4245 Statistikk
Vår 2020

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Innlevering 6

Dette er den siste skriftlige innleveringen. Denne øvingen dekker hele pensum. Spesielt er det i denne innleveringen fokus på hypotesetesting og lineær regresjon. For å få godkjent innleveringen kreves det at minimum 40% av svarene er riktige, og at det kommer fram at det er gjort et ærlig forsøk på 75% av oppgavene. Alle deloppgaver teller like mye. Husk: Vi har statistikklab hvis du trenger hjelp med øvingen. Se fagets nettsider for mer info.

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 $\mathrm{E}(X) = \lambda_R$.

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 $X_1, X_2, \ldots, X_n$ betegne antall rosiner i hver av bollene og anta at antall rosiner i ulike boller er uavhengige.

a) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for λ_R blir

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

Vis videre at

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

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

I resten av oppgaven skal du benytte at $\widehat{\lambda}_R$ 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 hypotetesetestingsproblem og lag en test for dette formål med (tilnærmet) signifikansnivå α.

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

$$Y_i = ax_i(1-x_i) + \epsilon_i, \qquad i = 1, \ldots, n,$$

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

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

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	$-.0561$	$-.0369$	$-.0651$	$-.0578$	$-.0941$	$-.0684$	$-.0546$	$-.0247$	$-.0072$

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 α. Hva blir konklusjonen på hypotesetesten for dataene over og $\alpha = 5\%$?

Oppgave 3

I situasjoner der det er uklart hvem som er den biologiske faren til et barn kan farskapet avklares ved å sammenligne DNA-prøver fra barnet med mulige fedre. For en mulig far gjøres dette ved å sammenligne n ulike deler av DNA-strukturen til mannen med de samme n deler av DNA-strukturen hos barnet. De n undersøkte delene av DNA-strukturen antas uavhengige.

Hos et barn og en tilfeldig valgt mann (som ikke er biologisk far) er det for hver enkel del av DNA-strukturen som undersøkes en sannsynlighet $p = 0.15$ for at delen er sammenfallende hos barnet og mannen. Anta videre at en biologisk far alltid har alle de undersøkte delene av DNA-strukturen sammenfallende med barnets (dvs. vi ser bort fra mutasjoner o.l.), slik

at hver undersøkte del av DNA-strukturen hos biologisk far og barn er sammenfallende med sannsynlighet $p = 1$.

La X være antall sammenfallende deler i DNA-strukturen hos et barn og en tilfeldig valgt mann (som ikke er biologisk far).

a) Begrunn at X er binomisk fordelt med parametre n og $p = 0.15$.

Dersom $n = 5$, beregn sannsynlighetene $P(X = 2)$, $P(X \geq 2)$ og $P(X = 2|X \geq 2)$.

I en farsskapssak blir en mann erklært å være biologisk far dersom alle undersøkte deler av DNA-strukturen er sammenfallende hos mannen og barnet. Dette kan vi se på som en hypotesetest der vi tester

$$H_0: \quad p = 0.15 \text{ (ikke far)} \qquad \text{mot} \qquad H_1: \quad p = 1.0 \text{ (far)}$$

der H_0 forkastes (dvs. mannen erklæres som far til barnet) dersom $X = n$.

b) For $n = 5$, finn sannsynligheten for å begå type 1 feil i testen over.

For $n = 5$, finn sannsynligheten for å begå type 2 feil i testen over.

Hvor mange ulike deler, n, av DNA-strukturen må man minst sammenligne dersom man ønsker at sannsynligheten for feilaktig å erklære en mann som far skal være mindre enn 0.000001?

Fasit

1. **b**) Ikke forkast H_0 **c**) Olav må sjekke 82 boller

2. **c**) Forkast H_0

3. **a**) $0.138, 0.165, 0.836$ **b**) $0.000076, 0, n = 8$