



EKSAMENSOPPGAVE I STA-0001

Eksamen i : **STA-0001 Brukerkurs i statistikk 1**

Eksamensdato : **Torsdag 28. september 2006**

Tid : **Kl 0900-1300**

Sted : **Auditorium maximum**

Tillatte hjelpemidler : **Alle trykte og skrevne samt kalkulator**

Oppgavesettet er på 3 sider pluss forside

Kontaktperson under eksamen: Amanuensis Elinor Ytterstad, tlf 77644015

DET MATEMATISK-NATURVITENSKAPELIGE FAKULTET

Universitetet i Tromsø, N-9037 Tromsø, Telefon 77 64 40 01, Telefaks 77 64 47 65

Deloppgavene vil telle likt i bedømmelsen

Oppgave 1

Et bryggeri produserer halvliterflasker med vann. Fyllingsprosessen er forbundet med usikkerhet, slik at hver flaske ikke inneholder nøyaktig en halv liter vann. La X være vanninnholdet (i liter) av en tilfeldig flaske. Etter omfattende undersøkelser har man funnet at X er normalfordelt med forventning $\mu = 0.507$ og standardavvik $\sigma = 0.007$.

- a) Hva er sannsynligheten for at en tilfeldig flaske inneholder minst $\frac{1}{2}$ liter vann?
Hva er sannsynligheten for at 10 tilfeldig valgte flasker inneholder minst 5 liter vann til sammen?

Du kjøper 10 flasker. La Y være antall flasker som inneholder minst $\frac{1}{2}$ liter vann. Vi skal anta at innholdet, X , i en tilfeldig flaske er uavhengig av innholdet i de andre flaskene.

- b) Hvilken fordeling har Y ?
Hva er sannsynligheten for at alle de 10 flaskene inneholder minst $\frac{1}{2}$ liter vann?

Etter fornying av produksjonsprosessen ønsker bryggeriet å teste om forventningsverdien μ er blitt forskjellig fra verdien 0.507. Med basis i et utvalg på $n = 1000$ flasker finner de en middelværdi på $\bar{x} = 0.502$. Anta at $\sigma = 0.007$ som før.

- c) Test om det på nivå 5 % er grunnlag for å hevde at $\mu \neq 0.507$.
Spesifiser H_0 og H_1 som del av testen.

Oppgave 2

Seks personer går inn i heisen i 1. etasje i en bygning med 11 etasjer. Vi antar at sannsynlighetene for at en tilfeldig valgt person skal forlate heisen i hhv. 2., 3., ..., 11. etasje er like store, og at personene forlater heisen uavhengig av hverandre.

- a) Hva er sannsynligheten for at den personen som gikk først inn i heisen skal av i 11. etasje?
Hva er sannsynligheten for at de går av i hver sin etasje?
- b) La X være lik antall personer som kjører heisen helt til 11. etasje.
Hvilken sannsynlighetsfordeling har X ?
Begrunn svaret og skriv opp fordelingen innsatt riktige parameterverdier.

Oppgave 3

For å bestemme aldersfordelingen til reker, som er viktig i marin ressursforvaltning, måles lengden av ryggskjoldet manuelt ved hjelp av et elektronisk skyvelære. Typisk vil ryggskjoldlengden til ettårige reker kunne ha en fordeling med modal topp på ca. 9 mm, to-åringer har modal topp på ca. 14 mm og så videre, slik at aldersfordelingen bestemmes ved å bestemme modal toppene i lengdefordelingen. Manuell måling er en tidkrevende prosess, og for tiden utvikles en langt mer effektiv teknikk basert på bildeanalyse, der det har vist seg at det er en god lineær sammenheng mellom ryggskjoldlengde og kvadratrota av antall piksler som en reke opptar i et bilde.

La L betegne sann (ryggskjold)lengde til en tilfeldig reke, la x betegne målt lengde med elektronisk skyvelære og la Y betegne kvadratrota av antall piksler til bildet av den samme reka. Vi skal neglisjere måleusikkerheten til x , dvs. at vi antar at $x = L$. Videre skal vi anvende den lineære regresjonsmodellen

$$Y_i = \alpha + \beta \cdot x_i + e_i, \quad i = 1, \dots, n$$

der $e_1, \dots, e_n$ antas å være uavhengige $N(0, \sigma)$ -variable. I et pilotprosjekt der $n = 285$ tilfeldige reker ble målt med skyvelære og avbildet ble følgende data fremskaffet:

$$\sum x_i = 5740, \quad \sum (x_i - \bar{x})^2 = 3127, \quad \sum y_i = 69935, \quad \sum (y_i - \bar{y})^2 = 557770, \quad \sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) = 41297$$

- Bestem den estimerte regresjonslinja $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$ med basis i oppgitt datamateriale, der estimatorene $\hat{\alpha}$ og $\hat{\beta}$ skal være forventningsrette.
- Beregn empirisk korrelasjonskoeffisient, r , mellom x og y . Gir r -verdien støtte til den lineære regresjonsmodellen? Begrunn svaret.

Vi skal nå anta at etter en lang serie med forsøk kan vi benytte verdiene $\alpha = -20$, $\beta = 13$ og $\sigma = 0.5$ og neglisjere usikkerheten til disse parametrene. Anta nå at vi kun måler Y og ønsker å estimere lengden, L .

- Bestem en estimator for L som funksjon av Y med basis i regresjonsmodellen. Hva blir estimatet for L dersom $Y = 250$?

Oppgave 4

I en undersøkelse av blåkkeite ønsker en å undersøke om adferden ved sjøbunnen viser gruppe/klynge adferd eller om hver enkeltfisk er tilfeldig fordelt. Blåkkeita observeres med video montert på en bunntål som går med konstant hastighet, og tidspunktene hver gang en ny blåkkeite registreres blir notert. En tilfeldig fordeling av enkeltfisk er ensbetydende med at tidsdifferansen (og tilsvarende avstanden) mellom to påfølgende blåkkeiteregistreringer er eksponensialfordelt, mens dette ikke er tilfelle ved gruppe/klynge adferd.

I tabellen nedenfor er vist en frekvenstabell for tidsdifferansen y (i sek) mellom to og to påfølgende enkeltfisk (til sammen 101 fisk) samt annen informasjon. Betegnelsen $[0 \ 10 >$ betyr alle verdier mellom 0.00... sek og 9.99... sek, og midtpunktet i intervallet er 5.

Tidsintervall for y	[0-10>	[10-20 >	[20-30 >	[30-90 >	Sum
Observert antall, X	59	23	11	7	100
(Null)hypotesesannsynlighet, p	0.568	0.245	0.106	0.080	0.999
(Null)hypoteseforventning, E					
Avvikssum $(X - E)^2 / E$					

Tilleggsinformasjon: $\bar{y} = \frac{1}{100} \sum_{i=1}^{100} y_i = 11.88$ sek

- a) Beregn gruppert middelerdi, $\bar{y}_g$, med basis i de to første linjene i tabellen. Begrunn hvorfor gruppert middelerdi avviker fra $\bar{y}$.

For å undersøke mer kvantitativt om eksponensialfordelingen gir en god tilpasning til det relative frekvenshistogrammet skal du nå gjennomføre en kjikvadrat modelltest der H_0 er hypotesen at Y er eksponensialfordelt med parameter $\lambda = 1/11.88 = 0.084$.

- b) Fyll ut de uutfylte cellene i tabellen og test på nivå 5 % om H_0 kan forkastes eller ikke. Formuler konklusjonen i form av blåkkeitas adferd (tilfeldig fordelt kontra gruppe/klynge adferd).
- c) Hvilken sannsynlighetsfordeling har antall blåkkeite, N , som registreres iløpet av 20 minutters tråling dersom H_0 er riktig? Spesifiser både fordelingstype, parameterverdi og hvilke mulige verdier N kan ha.

----- FINE -----