Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen:

EKSAMEN I EMNE TMA4245 STATISTIKK

august 2008
Tid: 09:00–13:00

Hjelpemidler: *Tabeller og formler i statistikk*, Tapir Forlag
K. Rottmann: *Matematisk formelsamling*
Kalkulator HP30S
Gult, stemplet A5-ark med egne hndskrevne notater.

Sensuren faller: august 2008

Oppgave 1

Vi skal i denne oppgaven se på høydefordelingene til menn og kvinner. Anta at høyden til menn er normalfordelt med forventningsverdi $\mu_M = 179$ og varians $\sigma_M^2 = 6^2$, og at høyden til kvinner er normalfordelt med forventningsverdi μ_K og varians σ_K^2

a) Finn sannsynligheten for at en tilfeldig valgt mann er over 185 cm.

Finn sannsynligheten for at en mann er over 185 cm gitt at han er over 179 cm.

b) La $\mu_K = \beta \cdot \mu_M$, der β er en ukjent parameter vi ønsker å estimere. Anta at vi har høydedata fra et tilfeldig utvalg på $n = 5$ kvinner; $X_i \sim N(\mu_K, \sigma_K^2)$, $i = 1, 2, \ldots, n$.

Som estimator for β velges $\hat{\beta} = \dfrac{\sum_{i=1}^{n} X_i}{n\mu_M}$

Er $\hat{\beta}$ forventningsrett?

Utled et uttrykk for et 95% konfidensintervall for β, og finn tallsvar når vi fra data får $\bar{x} = \sum_{i=1}^{n} x_i/n = 167.5$ og $s^2 = \dfrac{1}{n-1} \sum_{i=1}^{n}(x_i - \bar{x})^2 = 5.1^2$.

Oppgave 2

En næringsmiddelbedrift har et laboratorium for testing av kvaliteten av bedriftens produkter. For testing av mengde pr. volumenhet av et bestemt sporstoff i et av produktene, bruker bedriften to måleapparater, som vi kan kalle A og B. A er det mest nøyaktige av de to, og det er også det raskeste. Ulempen med A er at det er dyrt i bruk. Bedriften har derfor etablert en testprosedyre som går ut på å ta ut to tilfeldige utvalg fra en gitt produksjonsserie, hvor vi kan anta at mengde pr. volumenhet, betegnet med μ, er konstant. Det ene utvalget, som er av størrelse n, testes i A. Etter denne testingen i A har en da et tilfeldig utvalg $x_1, \ldots, x_n$ med måleresultater. Det andre utvalget, som er av størrelse $2n$, testes i B. Det resulterer i et uavhengig tilfeldig utvalg $y_1, \ldots, y_{2n}$ av størrelse $2n$ med måleresultater etter testing i B.

Det tilfeldige utvalget $x_1, \ldots, x_n$ kan betraktes som et utfall av n stokastiske variabler $X_1, \ldots, X_n$, som er uavhengige og normalfordelte med forventningsverdi μ og varians σ^2, der σ^2 reflekterer A's nøyaktighet. Tilsvarende, $y_1, \ldots, y_{2n}$ kan betraktes som utfall av $2n$ stokastiske variabler $Y_1, \ldots, Y_{2n}$, som er uavhengige og normalfordelte med forventningsverdi μ og varians $4\sigma^2$, og de er uavhengige av X_i, $i = 1, \ldots, n$.

a) Mengde pr. volumenhet μ skal estimeres pågrunnlag av de tilsammen $3n$ målingene som er gjort, og en mulig estimator er

$$\hat{M} = \frac{1}{2}\left(\overline{X} + \overline{Y}\right) \qquad \left(\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad \overline{Y} = \frac{1}{2n}\sum_{j=1}^{2n} Y_j\right)$$

Hva er forventningsverdi og varians til denne estimatoren?

b) Vis at sannsynlighetsmaksimerings-estimatoren (SME) for μ er gitt som

$$M^* = \frac{1}{3}\left(2\overline{X} + \overline{Y}\right)\,,$$

og beregn dens forventningsverdi og varians.

Hvilken av de to estimatorene $\hat{M}$ og M^* ville du anbefale? Begrunn svaret.

c) Anta i dette punktet at σ^2 er kjent, og lik 1.0. En skal nå teste hypotesen

$$H_0\,:\,\mu = \mu_0 = 100\,,$$

mot

$$H_1 \,:\, \mu < \mu_0\,.$$

på signifikansnivå $\alpha = 0.05$. Foreslå en testobservator, og angi testens kritiske område.

Hva blir konklusjonen når $n = 4$ og måleresultatene er:

X_i: 100.3, 100.8, 99.5, 98.8; som gir $\overline{X} = 100.4$

Y_j: 100.1, 98.7, 99.2, 102.3, 103.0, 97.5, 96.6, 103.1; som gir $\overline{Y} = 100.5$

d) Anta i dette punktet at σ^2 er ukjent. Følgende to estimatorer innføres,

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \qquad \left(\mathrm{E}[S_1^2] = \sigma^2\right)$$

og

$$S_2^2 = \frac{1}{2n-1}\sum_{j=1}^{2n}\left(Y_j - \overline{Y}\right)^2 \qquad \left(\mathrm{E}[S_2^2] = 4\sigma^2\right)$$

Disse to kombineres i en ny stokastisk variabel,

$$V = \frac{(n-1)S_1^2}{\sigma^2} + \frac{(2n-1)S_2^2}{4\sigma^2}\,.$$

Begrunn hvorfor V blir χ^2-fordelt. Hvor mange frihetsgrader har V?

Foreslå en testobservator for μ basert på M^* og V og angi fordeling.

Oppgave 3

I semifinalen mellom Frankrike og Tyskland i et verdensmesterskap i fotball er resultatet uavgjort etter ekstraomganger.

Vi definerer en runde i straffesparkkonkuransen til å være at hvert lag skyter en straffe hver.

I første del av straffesparkkonkurransen er det 5 runder, 5 straffespark fra hvert lag, og laget som skårer flest ganger vinner.

Dersom lagene skårer like mange ganger i første del, går man over til andre del av straffesparkkonkurransen. Nå spilles det en og en runde: Hvert lag har en straffe hver. Dersom det ene laget skårer og det andre bommer, har vi en vinner. Ellers spilles en ny runde (hvert lag får en straffe hver) inntil vi har en vinner.

Anta at de tyske spillerene har sannsynlighet $p_T = 0.80$ for å skåre på straffe, at de franske spillerene har sannsynlighet $p_F = 0.70$ for å skåre, og at utfallene av straffesparkene er uavhengige av hverandre.

a) Hva er sannsynligheten for at stillingen blir 5-5 etter første del?

Hva er sannsynligheten for at stillingen blir 3-3 etter første del?

Sett opp uttrykk for / algoritme for hvordan man kan finne sannsynligheten p_{D1-lik} for at lagene står likt etter første del?

Du trenger ikke å finne tallsvar, og kan senere i oppgaven benytte at $p_{D1-lik} = 0.27$

b) Anta i dette punktet at vi er kommet til del 2 av straffesparkkonkuransen.

Hva er sannsynligheten $p_{V1|D2}$ for at vi har en vinner etter første runde i del to?

Gitt at vi har en vinner etter første runde i del 2, hva er sannsynligheten $p_{T|V1,D2}$ for at dette er Tyskland?

Hva er sannsynligheten $p_{T|D2}$ for at Tyskland vinner konkurransen (gitt at vi er kommet til del 2)?

c) La X være antall runder i del 2 til og med runden det blir kåret en vinner i.

Hvilken fordeling har X? Begrunn svaret og oppgi parameter/re.

Hva er forventningsverdi og varians i antall runder (i del 2) til det blir kåret en vinner?

d) Vi antar at det alltid blir spilt 5 runder i del 1. Hva er forventet totalt antall runder i straffesparkkonkurransen (del 1 og evt. del 2) til det blir kåret en vinner?

Finn også variansen av det totale antall runder i straffesparkkonkurransen (del 1 og evt. del 2) inntil det blir kåret en vinner.