



Faglige kontakter under eksamen:

Mette Langaas 98847649

Ole Petter Løvdøen 99484158

EKSAMEN I FAG TMA4240 STATISTIKK

Torsdag 9. desember 2004

Tid: 09:00–13:00

Tillatte hjelpemidler:

Gult A5-ark med egne håndskrevne notater.

Tabeller og formler i statistikk (Tapir Forlag).

K. Rottmann: *Matematisk formelsamling*.

Kalkulator: HP30S.

BOKMÅL

Sensur: 6. januar 2005.

Oppgave 1 Bronsebolter

Bronse er en legering der kobber og tinn er hovedbestanddelene. Vi studerer kobberinnholdet i bronsebolter av en gitt dimensjon som er laget av en spesiell type bronselegering.

Ved bedriften *Bronsespesialisten* produseres det bronsebolter, og det er tatt stikkprøver av $n = 10$ bronsebolter fra produksjonen. Kobberinnholdet, X_i , $i = 1, \dots, n$, er målt. Vi antar at X_1, X_2, \dots, X_n er uavhengige og normalfordelte stokastiske variabler med $E(X_i) = \mu_x$ og $\text{Var}(X_i) = \sigma^2$.

- a) Vi antar at forventningen er $\mu_x = 85$ gram og variansen er $\sigma^2 = 1$ gram² (kun i dette punktet).

Hva er sannsynligheten for at kobberinnholdet i en tilfeldig valgt bronsebolt er mindre enn 84 gram?

Finn et tall, k , slik at sannsynligheten er 0.01 for at kobberinnholdet i en tilfeldig valgt bronsebolt er større enn k .

Vi ser på kobberinnholdet i to tilfeldig valgte og uavhengige bronsebolter. Hva er sannsynligheten for at kobberinnholdet i de to bronseboltene avviker med mer enn 1.5 gram?

Vi antar i resten av oppgaven at både μ_x og σ^2 er ukjente parametre. Vi ser først på estimatorer for σ^2 .

- b) Hvilke egenskaper kjennetegner en god estimator?

To aktuelle estimatorer for σ^2 er $\hat{\sigma}^2$ og S^2

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Finn forventningsverdien og variansen til de to estimatorene, og kommenter. (Hint: bruk relasjon til kji-kvadrat fordeling).

Ved bedriften *Metalleksperten* produseres det også bronsebolter av samme legering og dimensjon som hos Bronsespesialisten. Hos Metalleksperten er det tatt stikkprøver av $m = 10$ bronsebolter, og kobberinnholdet Y_j , $j = 1, \dots, m$, er målt. La Y_1, Y_2, \dots, Y_m være uavhengige og normalfordelte stokastiske variabler med $E(Y_j) = \mu_y$ og $\text{Var}(Y_j) = \sigma^2$, der både μ_y og σ^2 er ukjente parametre. Vi antar at målinger tatt av kobberinnhold i bronsebolter fra de to bedriftene er uavhengige. Videre antar vi at det er samme (ukjente) varians for målt kobberinnhold i bronsebolter fra de to bedriftene, dvs. $\text{Var}(X_i) = \text{Var}(Y_j) = \sigma^2$.

Data for stikkprøvene fra Bronsespesialisten og Metalleksperten er presentert i tabell 1.

Bronsespesialisten	84.44	84.77	86.56	85.07	85.13	86.72	85.46	83.73	84.31	84.55
Metalleksperten	85.22	84.36	84.40	84.11	83.44	85.79	84.50	82.03	84.70	83.53

Tabell 1: Kobberinnhold i bronselegering i stikkprøver fra Bronsespesialisten og Metalleksperten. Det oppgis at for Bronsespesialisten er $\sum_{i=1}^{10} x_i = 850.75$ og $\sum_{i=1}^{10} (x_i - \bar{x})^2 = 8.19$, og for Metalleksperten er $\sum_{j=1}^{10} y_j = 842.10$ og $\sum_{j=1}^{10} (y_j - \bar{y})^2 = 9.70$.

Ledelsen i Bronsespesialisten har i den siste tiden gjentatte ganger beskyldt Metalleksperten for at kobberinnholdet i bronseboltene fra Metalleksperten er lavere enn kobberinnholdet i bronseboltene fra Bronsespesialisten. Vi ønsker å undersøke om dette er tilfellet.

- c) Formulér dette som en hypotesetest ved å definere nullhypotese og alternativ hypotese.

Sett opp en testobservator og finn forkastningsområdet. Hva blir konklusjonen på testen, med data gitt i tabell 1, når signifikansnivået er $\alpha = 0.05$?

Regn ut p -verdien ved å bruke tabell 2.

t	1.82	1.86	1.90	1.94	1.98	2.02	2.06
$\nu = 18$	0.957	0.960	0.963	0.966	0.968	0.971	0.973
$\nu = 19$	0.958	0.961	0.964	0.966	0.969	0.971	0.973
$\nu = 20$	0.958	0.961	0.964	0.967	0.969	0.972	0.974

Tabell 2: Kumulativ sannsynlighet i t -fordelingen. For T t -fordelt med ν frihetsgrader, så viser tabellen $P(T \leq t)$ for ulike verdier av t .

Metalleksperten har blitt utsatt for hærverk og vinduet i bedriftens hovedkontor ble knust ved at en bronsebolt ble kastet igjennom vinduet. Politiet etterforsker saken, og er interessert i å finne ut om bronsebolten som knuste vinduet er produsert av Metalleksperten eller av Bronsespesialisten.

- d) Utled et intervall som med 95 % sannsynlighet inneholder en ny observasjon av kobberinnholdet i en bronsebolt fra Bronsespesialisten. Finn numeriske verdier for intervallet ved å bruke dataene fra tabell 1. Lag et tilsvarende intervall for Metalleksperten.

Bronsebolten som ble brukt til å knuse vinduet hos Metalleksperten ble målt til å ha et kobberinnhold på 86.30 gram. Kan du ut fra intervallene du har laget over si noe om hvilken produsent som kan ha laget bronsebolten?

Oppgave 2 Pyramidespillet

Ole Petter har blitt spurt om å bli med i et pyramidespill. Alt han trenger å gjøre, er å betale inn en viss sum penger, og deretter rekruttere fem nye personer, som skal plasseres i nivået under ham i pyramiden. Etter at det er gjort, forsikres Ole Petter om at pengene vil begynne å strømme inn. Ifølge personen som spurte Ole Petter, vil en person som blir spurt om å delta i pyramidespillet ha en sannsynlighet $p = 1/3$ for å bli med, så det å få fem personer til å bli med, skal ikke være noe problem.

Før Ole Petter bestemmer seg, vil han regne litt på sannsynligheter og forventningsverdier, for å finne ut hvor mye arbeid rekrutteringsprosessen vil medføre for ham. Han bestemmer seg for å se på den stokastiske variabelen X , som angir antall personer Ole Petter må spørre, inntil den *første personen* blir med i pyramidespillet.

- a) Under hvilke antagelser vil X være geometrisk fordelt? I resten av oppgaven kan du anta at X er geometrisk fordelt med punktsannsynlighet

$$f(x) = p(1-p)^{(x-1)} \quad \text{for } x = 1, 2, \dots$$

Dersom Ole Petter bestemmer seg for å delta i pyramidespillet, hva er forventet antall personer han må spørre for å få med en ny person, når $p = 1/3$?

Hva er sannsynligheten for at han må spørre flere enn fem personer for å få en person til å delta i pyramidespillet, når $p = 1/3$?

Ole Petter er skeptisk til informasjonen om at hver tredje person som blir spurt vil melde seg inn, og bestemmer seg for å estimere sannsynligheten p for at en person som blir spurt vil delta i pyramidespillet. Han bestemmer seg for å samle inn data, ved at han teller antall personer han må spørre inntil den første svarer ja. Dette vil han prøve å få noen kamerater til å hjelpe seg med. Den *ite* kameraten hans observerer at han må spørre X_i personer. Fra n undersøkelser sitter da Ole Petter da igjen med n uavhengige observasjoner X_1, X_2, \dots, X_n av den stokastiske variabelen X .

- b) Finn sannsynlighetsmaksimeringsestimatoren \hat{p} for p basert på observasjonene X_1, X_2, \dots, X_n .

Ingen av kameratene til Ole Petter er interessert i å hjelpe ham med undersøkelsen. Han sitter derfor igjen med kun den ene observasjonen som han har innhentet selv, X_1 . Er sannsynlighetsmaksimeringsestimatoren basert på kun denne ene observasjonen forventningsrett?

$$\text{Hint : } \sum_{n=1}^{\infty} \frac{1}{n} a^n = -\ln(1-a) ; 0 < a < 1$$

Oppgave 3 Test nasjonen

Lørdag 27. november 2004 ble TV-programmet “Test nasjonen” sendt på NRK1. I programmet ble 270 deltakere i studio stilt spørsmål innen ulike tema. Basert på alder og antall rette svar fikk hver deltaker tildelt en IQ-score.

I programmet ble det opplyst at testen var laget slik at man forventet at IQ-score til en tilfeldig valgt person skulle være normalfordelt med forventningsverdi 100 og standardavvik 15.

Den maksimale IQ-score som ble registrert i studio var 122, og det var to personer som hadde denne IQ-scoren.

- a) Hva er sannsynligheten for at en tilfeldig valgt person skal få en IQ-score på minst 122?

Hvis vi tester et representativt utvalg på 270 personer, hva er da forventet antall personer som får en IQ-score på minst 122?

Hva er sannsynligheten for at maksimal IQ-score i et tilfeldig utvalg av størrelse 270 vil være større enn 122?

Deltakerne i studio var delt inn i seks grupper, der en av gruppene bestod av 42 tidligere “Reality-deltakere”. Før programmet startet bestemte vi oss for at når det gjelder IQ anser vi de 42 tidligere Reality-deltakerene som et representativt utvalg fra befolkningen. Reality-deltakerene fikk en gjennomsnittlig IQ-score på 94.

- b) Vi antar at IQ-score til en tilfeldig valgt person er normalfordelt med ukjent forventningsverdi μ , men kjent standardavvik 15. Utled et 95% konfidensintervall for forventet IQ-score basert på data for de tidligere Reality-deltakerene.

Etter programmet har det blitt stilt spørsmål om oppgavene i IQ-testen var for vanskelige, slik at IQ-scorene som ble oppnådd var lavere enn man kunne forvente. Hvis vi antar at IQ-score til en tilfeldig valgt person er normalfordelt med forventningsverdi 100 og standardavvik 15, hva er da sannsynligheten for at man i et tilfeldig utvalg på 42 personer oppnår en gjennomsnittlig IQ-score på mindre eller lik 94?