**Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag**

Faglig kontakt under eksamen:
Turid Follestad 98066880

EKSAMEN I FAG TMA4240/TMA4245 STATISTIKK

19. august 2006

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: 11. september 2006.

Oppgave 1 Kvalitetskontroll av poteter

En grønnsaksprodusent produserer poteter, og er blitt godkjent som bruker av kvalitetsmerket PrimaVare. Dette innebærer blant annet at potetene skal tilhøre en bestemt vektklasse. Produsenten tar jevnlig stikkprøver for å kontrollere at potetene holder seg innenfor kvalitetsspesifikasjonen.

La X være vekten av en tilfeldig valgt potet, og anta at vekten er normalfordelt med forventning $\mu = 110\ g$ og varians σ^2. La kvalitetsspesifikasjonen for vekt være at vekten på potetene ikke skal avvike mer enn 20 g fra den forventede vekten på 110 g.

a) Anta, bare i dette punktet, at variansen er kjent, og lik $\sigma^2 = 100\ g^2$.

Hva er sannsynligheten for at en tilfeldig valgt potet veier mer enn 120 g?

Hva er sannsynligheten for at en tilfeldig valgt potet er innenfor spesifikasjonen, dvs. $P(|X - 110| \leq 20)$?

Hva er sannsynligheten for at en pose med 14 tilfeldig valgte poteter veier mindre enn 1500 g?

Anta at det blir tatt en stikkprøve med n tilfeldig valgte poteter fra produksjonen. La Y være antall poteter som *ikke* følger spesifikasjonen, blant de n i stikkprøven.

b) Forklar hvorfor vi kan anta at Y er binomisk fordelt med parametre n og $p = 1 - P(|X - 110| \leq 20)$, der X er definert som tidligere i oppgaven.

En estimator for p er $\hat{p} = \dfrac{Y}{n}$. Vis at dette er en forventningsrett estimator for p, og utled variansen til $\hat{p}$.

Bruk denne estimatoren til å finne et anslag for variansen σ^2 til vekten på en potet, når det i en stikkprøve på størrelse $n = 19$ blir observert at $y = 2$ poteter ikke følger spesifikasjonen.

Oppgave 2 Internettkafé

Torkel driver en internettkafé der han har tre datamaskiner tilgjengelig. Iblant er det mange kunder innom, og enkelte blir misfornøyde dersom det er lenge å vente på ledig maskin. Torkel bestemmer seg for å gjennomføre en statistisk analyse av påloggingstidene til kunder. Anta at påloggingstiden T til en tilfeldig valgt kunde har kumulativ sannsynlighetsfordeling

$$F(t) = 1 - \exp(-\lambda t), \qquad \text{for } t > 0 \text{ og } \lambda > 0.$$

a) Anta, kun i dette punktet, at $\lambda = 0.1$. Beregn $P(T < 15)$. Bruk definisjonen av betinget sannsynlighet til å beregne $P(T < 45|T > 30)$. Sammenlign de to resultatene og drøft i den sammenheng hvilken egenskap som gjelder for denne sannsynlighetsfordelingen.

For å gjennomføre sin analyse registrerer Torkel ti uavhengige pålogginger. La T_i være påloggingstiden til person $i = 1, \ldots, 10$, og anta at disse følger den kumulative sannsynlighetsfordelingen gitt over. Resultatet er følgende påloggingstider (i minutter): 11.1, 9.4, 40.4, 19.5, 6.9, 5.4, 14.5, 2.2, 8.0 og 7.7.

b) Vis at sannsynlighetstettheten til T_i, $i = 1, \ldots, 10$, er

$$f(t) = \begin{cases} \lambda \exp(-\lambda t) & \text{for } t > 0 \\ 0 & \text{ellers.} \end{cases}$$

Basert på tallmaterialet i undersøkelsen vil Torkel estimere λ. Sett opp rimelighetsfunksjonen (likelihoodfunksjonen) med tanke på dette formålet. Utled sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimatoren) for λ og regn ut estimatet.

Torkel vil benytte resultatet fra undersøkelsen til å informere kundene om ventetiden på ledig maskin.

c) En person kommer inn på kaféen, men alle tre maskinene er opptatt. Hva er sannsynligheten for at kunden får en ledig maskin innen 15 minutter?

To personer kommer samtidig inn på kaféen, men alle tre maskinene er opptatt. Hva er forventet tid til den siste av de to får en ledig maskin?

Uttrykk løsningene først som funksjon av λ, og sett deretter inn din estimerte verdi for λ fra forrige deloppgave for å gi tallsvar.

Oppgave 3 Kneoperasjon

Ved operasjon av kneleddet er det viktig at vinkelutslaget mellom leggbeinet og lårbeinet blir nær 0 grader. Vinkelen regnes som negativ hvis personen er kalvbeint og positiv hvis personen er hjulbeint.

I en medisinsk studie ble $n = 25$ pasienter kneoperert med konvensjonell kirurgi, og $m = 27$ pasienter ble operert med en ny teknikk som kalles computer-assistert kirurgi.

La X være vinkelutslaget mellom leggbeinet og lårbeinet etter operasjon ved konvensjonell kirurgi, og anta at X er normalfordelt med forventningsverdi $\mathrm{E}(X) = \mu$ og standardavvik $\mathrm{SD}(X) = \sigma$. Vi har $n = 25$ uavhengige målinger av vinkelutslag med konvensjonell kirurgi, og disse noteres $x_1, ..., x_{25}$. Det opplyses at $\bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 2.3$ og $\sum_{i=1}^{25}(x_i - \bar{x})^2 = 294$.

Tilsvarende, la Y være vinkelutslaget mellom leggbeinet og lårbeinet etter operasjon ved computer-assisert kirurgi, og anta at Y er normalfordelt med forventningsverdi $\mathrm{E}(Y) = \eta$ og standardavvik $\mathrm{SD}(Y) = \sigma$. Merk at standardavviket til de to operasjonsmetodene antas å være det samme. Vi har $m = 27$ uavhengige målinger av vinkelutslag med computer-assistert kirurgi, og disse noteres $y_1, ..., y_{27}$. Det opplyses at $\bar{y} = \frac{1}{27}\sum_{j=1}^{27} y_j = 1.5$ og $\sum_{j=1}^{27}(y_j - \bar{y})^2 = 115$.

Vi antar at målingene som er gjort fra de to operasjonsmetodene er uavhengige av hverandre.

a) Foreslå en god estimator for variansen σ^2 basert på alle målingene, dvs. fra både konvensjonell og computer-assistert kirurgi. Finn forventningsverdien til estimatoren du har foreslått. Er estimatoren forventningsrett?

Utled et 95% konfidensintervall for variansen σ^2 basert på målingene fra både konvensjonell og computer-assistert kirurgi. Regn også ut intervallet numerisk.

b) Vi ønsker å undersøke om forventet vinkelutslag mellom leggbeinet og lårbeinet er forskjellig for de to operasjonsmetodene.

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

Sett opp en testobservator og finn forkastningsområdet. Hva blir konklusjonen på testen, når signifikansnivået er $\alpha = 0.05$ og du bruker dataene som er gitt i oppgaven?

Regn ut p-verdien ved å bruke tabell 1.

t	1.05	1.03	1.01	0.99	0.97	0.95	0.75	0.55	0.35	0.15
$\nu = 25$	0.848	0.844	0.839	0.834	0.829	0.824	0.770	0.706	0.635	0.559
$\nu = 27$	0.848	0.844	0.839	0.835	0.830	0.825	0.770	0.707	0.635	0.559
$\nu = 50$	0.851	0.846	0.841	0.837	0.832	0.827	0.772	0.708	0.636	0.559
$\nu = 52$	0.851	0.846	0.841	0.837	0.832	0.827	0.772	0.708	0.636	0.559

Tabell 1: Kumulativ sannsynlighet i t-fordelingen. For T t-fordelt med ν frihetsgrader, så viser tabellen $P(T \leq t)$ for ulike verdier av t.

Oppgave 4 Simultanfordeling

Simultanfordelingen, $f(x, y)$, til de to diskrete stokastiske variablene X og Y er gitt i følgende tabell:

	y=0	y=1	y=2
x=-1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
x=0	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
x=1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Finn marginalfordelingen til X og til Y, og beregn forventning og varians til X og til Y.

Beregn kovariansen mellom X og Y, $\mathrm{Cov}(X, Y)$. Er X og Y uavhengige? Svaret skal begrunnes.