Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Henning Omre (909 37848)
Mette Langaas (988 47649)

EKSAMEN I TMA4240 Statistikk

18. desember 2010
Tid: 9:00–13:00

Antall studiepoeng: 7.5

Hjelpemidler:

- Kalkulator HP30S eller Citizen SR-270X med tomt minne.
- *Statistiske tabeller og formler*, Tapir forlag.
- K. Rottman: *Matematisk formelsamling* eller *Matematische Formelsammlung*.
- Ett gult ark (A5 med stempel) med egne håndskrevne formler og notater.

Sensurfrist: 18. januar 2011.
Eksamensresultatene blir annonsert fra `http://studweb.ntnu.no/`.

BOKMÅL

Oppgave 1 Høysnue og eksem

Vi betrakter en populasjon av 11-årige barn. I denne populasjonen har man sett på forekomsten av høysnue og eksem.

Definer to hendelser:

- E: et tilfeldig valgt barn fra populasjonen har eksem.
- H: et tilfeldig valgt barn fra populasjonen har høysnue.

La oss anta at vi i denne populasjonen har følgende sannsynligheter:

$$\begin{aligned} P(E) &= 0.04 \\ P(H) &= 0.07 \\ P(E \cap H) &= 0.009 \end{aligned}$$

a) Tegn et Venn-diagram med de to hendelsene.
Er hendelsene E og H uavhengige? Begrunn svaret.
Blant de barna i populasjonen som ikke har eksem, velger vi tilfeldig ut ett barn. Hva er sannsynligheten for at dette barnet ikke har høysnue?

Oppgave 2 Avissalg

Vi ser på salg av lokalavisa ved en liten aviskiosk i løpet av en dag. Vi antar at vi har et ubegrenset antall aviser tilgjengelig for salg slik at avisa ikke blir utsolgt.

La X være antallet eksemplarer av avisa som blir solgt i løpet av en tilfeldig valgt dag, og anta at X er Poisson-fordelt med forventningsverdi $\mathrm{E}(X) = \mu$.

a) Anta i dette punktet at forventet salg er $\mu = 10$ eksemplarer, og at vi ser på salget en tilfeldig valgt dag.

Hva er sannsynligheten for at det denne dagen blir solgt akkurat 10 eksemplarer?
Hva er sannsynligheten for at det denne dagen blir solgt 13 eller flere eksemplarer?

Man vet fra tidligere år at forventet avissalg er omtrent det samme for alle hverdager i løpet av høsten.

La $X_1, X_2, \ldots, X_n$ være uavhengige Poisson-fordelte stokastiske variabler som angir salg for n tilfeldig valgte hverdager om høsten. Anta at forventet salg er μ for alle hverdagene, men at denne parameteren ikke er kjent. En estimator for μ er $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

b) Finn forventningsverdi og varians til $\bar{X}$.
Bruk sentralgrenseteoremet til å lage et tilnærmet 95% konfidensintervall for μ.
Beregn konfidensintervallet numerisk når $n = 30$ og $\frac{1}{30}\sum_{i=1}^{30} x_i = 10.75$.

Eieren av aviskiosken ønsker å bestemme hvor mange aviser som skal kjøpes inn fra utgiveren en valgt dag. Anta at forventet antall aviser som kan selges den dagen, μ, er kjent. Deretter er det to motstridende hensyn som må tas:

- Hvis man ved aviskiosken blir utsolgt for avisen mister man potensiell fortjeneste.
- Hvis man ved aviskiosken ikke selger alle avisene som man kjøper inn, vil man tape penger fordi det ikke gis full kompensasjon for aviser som returneres til utgiver.

La fortsatt X betegne salget hvis man har et ubegrenset antall aviser tilgjengelig. Anta videre at $Y = \min(X, a)$ er salget hvis man har kjøpt inn a aviser for salg. Vi antar som tidligere at X er Poisson-fordelt, og lar forventningsverdien være $\mu = 10$.

c) Finn sannsynlighetsfordelingsfunksjonen til Y, $P(Y = y)$ for $y = 0, 1, \ldots, a$.
Vis at $E(Y) = a - \sum_{y=0}^{a-1}(a - y)P(X = y)$.

Vi kan velge a ut fra betrakninger om total fortjeneste for aviskiosken. Aviskiosken betaler 5 kr for hver avis og selger avisa for 12 kr. De usolgte avisene får aviskiosken returnert og får 3 kr for hver avis.

Vis at total fortjeneste er $9Y - 2a$.
Hvor mange aviser a bør aviskiosken kjøpe inn slik at *forventet* total fortjeneste skal bli størst mulig?

Hint: Kall den forventede totale fortjenesten for $h(a)$. Undersøk for hvilke a differansen $h(a) - h(a - 1)$ er positiv og negativ.

Oppgave 3 Kovarians

Vi har to tilfeldige (stokastiske) variabler X og Y. La X ha forventningsverdi $\mathrm{E}(X) = 10$ og varians $\mathrm{Var}(X) = 4$, og Y ha forventningsverdi $\mathrm{E}(Y) = 8$ og varians $\mathrm{Var}(Y) = 9$. Videre er kovariansen mellom X og Y gitt som $\mathrm{Cov}(X, Y) = 5$.

a) Regn ut tallverdiene for uttrykkene:
$\mathrm{E}(2X - Y)$
$\mathrm{Var}(2X - Y)$
$\mathrm{E}\left((X - 3)(Y - 5)\right)$

Oppgave 4 Forsøksgården

På en forsøksgård utføres forsøk med produksjon av biomasse. Anta at biomassen, Y, av en plante er normalfordelt (gaussisk fordelt) med forventningsverdi $\mathrm{E}(Y) = 5$ og varians $\mathrm{Var}(Y) = 4$.

a) Regn ut følgende tre sannsynligheter:
$P(Y > 6)$
$P(4 < Y \leq 6)$
$P(Y > 6 \mid Y > 4)$

Anta nå at biomassen Y av en bestemt plante er avhengig av kultiveringsperioden x. For planten defineres kultiveringsperioden som tiden fra planten kan observeres over jordsmonnet til tidspunktet for biomassemåling.

Anta videre at sammenhengen mellom biomasse Y og en gitt kultiveringsperiode x kan modelleres som en lineær regresjon, uten konstantledd og med feilledd som er avhengig av kultiveringsperioden,

$$Y = \beta x + \varepsilon(x) \text{ for } x > 0,$$

der $\varepsilon(x)$ er normalfordelt (gaussisk fordelt) med forventningsverdi $\mathrm{E}(\varepsilon) = 0$ og varians $\mathrm{Var}(\varepsilon) = \tau^2 x^2$. Det betyr at standardavviket til feilleddet er proporsjonalt med kultiveringsperioden x. Biomassemålingen gjøres kun for positive kultiveringsperioder, $x > 0$.

Modellparameterene β og τ ansees som ukjente.

Et forsøksopplegg med $n = 5$ uavhengige målinger ved kultiveringsperioder $x_1, x_2, \ldots, x_5$ og tilhørende biomasser $Y_1, Y_2, \ldots, Y_5$ resulterte i følgende observasjoner:

i	1	2	3	4	5
x_i	3	6	7	10	14
y_i	1.0	5.0	3.0	3.0	10.0

Det oppgis at $\sum_{i=1}^{5} \frac{y_i}{x_i} = 2.61$ og $\sum_{i=1}^{5} \frac{y_i^2}{x_i^2} = 1.59$.

b) Hvilken sannsynlighetsfordeling har Y_i gitt x_i?
Utled uttrykk for estimatorer for β og τ^2, for eksempel ved å bruke sannsynlighets-maksimeringsmetoden (maximum likelihood method).
Bruk tallverdier fra tabellen over til å beregne estimater for β og τ^2.

Følgende estimator for β skal benyttes i resten av oppgaven:

$$B = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{x_i}$$

c) Anta i dette punktet at $\tau^2 = 0.04$ er kjent.

Vi ønsker å utføre hypotesetesten

$$H_0\colon \beta = 0.50 \text{ mot } H_1\colon \beta > 0.50$$

Utled en forkastningsregel med signifikansnivå 0.05.
Bruk tallverdiene i tabellen over til å utføre testen.
Utled et uttrykk for styrken til testen over når $\beta = 0.7$, og regn ut tallsvar når $n = 5$.