Noregs teknisk–naturvitskaplege universitet
Institutt for matematiske fag

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

EKSAMEN I TMA4240 Statistikk

18. desember 2010
Tid: 9:00–13:00

Tal på studiepoeng: 7.5

Hjelpemiddel:

- Kalkulator HP30S eller Citizen SR-270X med tomt minne.
- *Statistiske tabeller og formler*, Tapir forlag.
- K. Rottman: *Matematisk formelsamling* eller *Matematische Formelsammlung*.
- Eit gult, stempla A5-ark med eigne handskrivne formlar og notat.

Sensurfirst: 18. januar 2011.
Eksamensresulta blir annonsert frå `http://studweb.ntnu.no/`.

NYNORSK

Oppgåve 1 Høysnue og eksem

Vi studerer ein populasjon av 11-årige barn. I denne populasjonen har ein sett på førekomsten av høysnue og eksem.

Definer to hendingar:

- E: eit tilfeldig valt barn frå populasjonen har eksem.
- H: eit tilfeldig valt barn frå populasjonen har høysnue.

La oss anta at vi i denne populasjonen har følgjande sannsyn:

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

a) Teikn eit Venn-diagram med dei to hendingane.
Er hendingane E og H uavhengige? Grunngje svaret.
Blant dei barna i populasjonen som ikkje har eksem, vel vi tilfeldig ut eitt barn. Kva er sannsynet for at dette barnet ikkje har høysnue?

Oppgåve 2 Avissal

Vi ser på sal av lokalavisa ved ein liten aviskiosk i løpet av ein dag. Vi går ut ifrå at vi har eit uavgrensa tal på aviser tilgjengeleg for sal slik at avisa ikkje blir utseld.

La X vere talet på eksemplar av avisa som blir selt i løpet av ein tilfeldig valt dag, og anta at X er Poisson-fordelt med forventningsverdi $\mathrm{E}(X) = \mu$.

a) Anta i dette punktet at forventa sal er $\mu = 10$ eksemplar, og at vi ser på salet ein tilfeldig valt dag.

Kva er sannsynet for at det denne dagen blir selt akkurat 10 eksemplar?
Kva er sannsynet for at det denne dagen blir selt 13 eller fleire eksemplar?

Ein veit frå tidlegare år at forventa avissal er omtrent det same for alle kvardagar i løpet av hausten.

La $X_1, X_2, \ldots, X_n$ vere uavhengige Poisson-fordelte stokastiske variablar som angir sal for n tilfeldig valde kvardagar om hausten. Anta at forventa sal er μ for alle kvardagane, men at denne parameteren ikkje er kjend. Ein 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 eit tilnærma 95% konfidensintervall for μ.
Rekn ut konfidensintervallet numerisk når $n = 30$ og $\frac{1}{30}\sum_{i=1}^{30} x_i = 10.75$.

Eigaren av aviskiosken ynskjer å bestemme kor mange aviser som skal kjøpast inn frå utgjevaren ein valt dag. Anta at forventa tal på aviser som kan selgjast den dagen, μ, er kjent. Deretter er det to motstridande omsyn som må takast:

- Dersom ein blir utselt for avisa mister ein potensiell forteneste.
- Dersom ein ikkje sel alle avisene ein kjøper inn til aviskiosken, vil ein tape pengar fordi ein ikkje får full kompensasjon for aviser som blir returnerte til utgjevaren.

La framleis X vere salet viss ein har eit uavgrensa tal på aviser tilgjengeleg. Anta vidare at $Y = \min(X, a)$ er salet viss ein har kjøpt inn a aviser for sal. Vi antek som i tidlegare at X er Poisson-fordelt, og lar forventningsverdien vere $\mu = 10$.

c) Finn sannsynsfordelingsfunksjonen 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 velje a ut ifrå betrakningar om total forteneste for aviskiosken. Aviskiosken betaler 5 kr for kvar avis og seljer avisa for 12 kr. Dei uselde avisene får aviskiosken returnert og får 3 kr for kvar avis.

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

Hint: Kall den forventa totale fortenesta for $h(a)$. Undersøk for kva verdiar a differansen $h(a) - h(a-1)$ er positiv og negativ.

Oppgåve 3 Kovarians

Vi har to tilfeldige (stokastiske) variablar 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 $\mathrm{Var}(Y) = 9$. Vidare er kovariansen mellom X og Y gitt som $\mathrm{Cov}(X, Y) = 5$.

a) Rekn ut talverdiane for uttrykka:
$\mathrm{E}(2X - Y)$
$\mathrm{Var}(2X - Y)$
$\mathrm{E}\left((X - 3)(Y - 5)\right)$

Oppgåve 4 Forsøksgarden

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

a) Rekn ut følgjande tre sannsyn:
$P(Y > 6)$
$P(4 < Y \leq 6)$
$P(Y > 6 \mid Y > 4)$

Anta no at biomassen Y av ein bestemt plante er avhengig av kultiveringsperioden x. For planten blir kultiveringsperioden definert som tida frå planten kan observerast over jordsmonnet til tidspunktet for biomassemåling.

Anta vidare at sammanhangen mellom biomasse Y og ein gitt kultiveringsperiode x kan modellerast som en lineær regresjon, utan 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. Biomassemålinga blir berre gjord for positive kultiveringsperiodar, $x > 0$.

Modellparametrane β og τ ser vi på som ukjente.

Eit forsøksopplegg med $n = 5$ uavhengige målingar ved kultiveringsperiodar $x_1, x_2, \ldots, x_5$ og tilhørande biomassar $Y_1, Y_2, \ldots, Y_5$ resulterte i følgjande observasjonar:

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 er oppgitt 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) Kva for sannsynsfordeling har Y_i gitt x_i?
Utlei uttrykk for estimatorar for β og τ^2, for eksempel ved å bruke sannsynsmaksimeringsmetoden (maximum likelihood method).
Nytt talverdiar frå tabellen over til å rekne ut estimat for β og τ^2.

Følgjande estimator for β skal nyttast i resten av oppgåva:

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

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

Vi ynskjer å utføre hypotesetesten

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

Utlei ein forkastningsregel med signifikansnivå 0.05.
Nytt talverdiar frå tabellen over til å utføre testen.
Utlei eit uttrykk for styrken til testen over når $\beta = 0.7$, og rekn ut talsvar når $n = 5$.