Rettevejledning til reeksamen i Økonometri A 2012-I, 2. år

Målbeskrivelse:

Kurset har som mål at introducere studerende til sandsynlighedsteori og statistik. Målet er, at de studerende efter at have gennemført faget kan:

- Forstå og benytte de vigtigste sandsynlighedsteoretiske begreber som: sandsynlighed, simultane-, marginale- og betingede- sandsynligheder, fordeling, tæthedsfunktion, uafhængighed, middelværdi, varians og kovarians samt at selvstændigt kunne anvende disse begreber på konkrete problemstillinger

- Kende resultatet fra den centrale grænseværdi sætning

- Kende og genkende de mest anvendte diskrete og kontinuerte fordelinger som: Bernoulli, Binomial, Poisson, multinomial, negative binomial fordeling, hypergeometrisk, geometrisk, lige-, normal-, Chi-i-anden-, eksponential, gamma-, t-, F-fordeling samt at selvstændigt kunne arbejde med disse fordelinger i konkrete problemstillinger

- Forstå de vigtigste statistiske begreber som: tilfældige udvælgelse, likelihood funtionen, sufficiens, stikprøvefunktion, egenskaber ved stikprøvefunktionen, estimation heruden maksimum likelihood og moment estimation, konsistens, konfidensinterval, hypoteseprøvning, teststørrelser, hypoteser, testsandsynlighed, signifikansniveau og type I og II fejl

- Være i stand til selvstændigt at gennemføre en simpel statistisk analyse, som involverer estimation, inferens og hypoteseprøvning, f.eks. sammenligning af middelværdien i to populationer eller uafhængighedstest for diskrete stokastiske variable.

- Indlæse og kombinere datasæt, lave nye variable, udtrække en stikprøve og udføre simple statistiske analyser ved hjælp af statistik-pakken SAS

- Beskrive resultatet af egne analyser og overvejelser i et klart og tydeligt sprog

Opgave 1

1. $Y_t\tilde{\ }Poi(\lambda t)$, t er målt i dage

$P(Y_{30} = 0) = \frac{(30\lambda)^0}{0!}e^{-30\lambda} = 0,1 => \lambda = 0,0768$

$E[Y_1] = 0,0768$

2. For hver af de 3 typer som i1:

$Y_t^1\tilde{\ }Poi(0,0768t)$
$Y_t^2\tilde{\ }Poi(0,0536t)$
$Y_t^3\tilde{\ }Poi(0,0999t)$

Lad $Y_t = Y_t^1 + Y_t^2 + Y_t^3$ antallet af jobtilbud i alt uanset type.

$Y_t\tilde{\ }Poi(0,2303t)$

$P(Y_{10} = 0) = e^{-0,2303\cdot 10} = 0.099959$

3. Vi skal beregne sandsynlighed for at de ikke har fået et jobtilbud

$P(Y_{60}^1 = 0) = e^{-0,0768\cdot 60} = 0.01$
$P(Y_{60}^2 = 0) = e^{-0,0536\cdot 60} = 0.04$
$P(Y_{60}^3 = 0) = e^{-0,0999\cdot 60} = 0.0025$

Dvs fordelingen mellem dem er 0.01 : 0.04 : 0.0025. Dvs. ca 19 pct. type 1, 76 pct. type 2 , og 5 pct. type 3.

Opgave 2

1. For den første $E[X_1] = \frac{9}{25}\cdot 90 + \frac{6}{10}\cdot 110 + \frac{1}{25}\cdot 120 = 103.2$ og den anden $E[X_2] = \frac{5}{18}\cdot 100 + \frac{1}{18}\cdot 125 + \frac{7}{18}\cdot 130 = 120$

2. $Z = Max(X_1, X_2)$ Først opstilles udfaldsrummet for Z og ssh. funktionen.

Z	$f(Z)$
100	$\frac{1}{10}$
110	$\frac{1}{6}$
120	$\frac{1}{90}$
125	$\frac{1}{3}$
130	$\frac{7}{18}$

$E[Z] = 121.89$

3. Først opstilles den simultane fordeling

	90	110	120
100	$\frac{1}{10}$	0	0
110	0	$\frac{1}{6}$	0
120	0	0	$\frac{1}{90}$
125	$\frac{3}{25}$	$\frac{1}{5}$	$\frac{1}{75}$
130	$\frac{7}{50}$	$\frac{7}{30}$	$\frac{7}{450}$

$E[X_1 Z] = 12594$
$Cov(X_1, Z) = 12594 - 121.89 \cdot 103.2 = 15.07$

Opgave 3

På en motorvejsstrækning nord for København registreres en tilfældig dag (i løbet af en time) antallet af køretøjer,

der kører markant over hastighedsbegrænsningen.

I alt blev der registreret 1014 personer der kørte for stærkt.

De 1014 personer var fordelt med 710 mænd og 304 kvinder.

1. Argumenter for at antallet af mænd der kørte for stærkt på strækningen kan beskrives med en binomialfordeling med sandsynlighedsparameter p.

Der er to udfald enten mand eller kvinde, det antages at være konstant sandsynlighed for at det er mand der kører for stærkt og endelig antages uafhængighed.

2. Estimer sandsynlighedsparameteren i denne binomialfordeling og angiv egenskaberne for denne estimator

$\hat{p} = \frac{710}{1014}$ =0,70 Som er middelret og hvor variansen vil gå mod nul, når antallet af forsøg går mod uendelig.

Dermed er estimatoren konsistent

3. udregn et 95% konfidensinterval for denne estimator

$\hat{p}\pm 1,96*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,70\pm 1,96*\sqrt{\frac{0,7(1-0,7)}{1014}}$ =0,70± 0,028 (s. 383 i B&L)

Der bruges large sample confidence limits.

På den omtalte motorvejsstrækning er det observeret at andelen af mænd er 2/3.

H0: $p = 2/3$ HA: p forskellig fra 2/3, men der må også gerne bruges ensidet alternativ

teststørrelse Z=$\frac{\hat{p}-2/3}{\sqrt{\frac{2/3(1-2/3)}{1014}}} = 2,26$ (se side 432 i B&L). Teststørrelsen er approksimativ normalfordelt med middelværdi 0 og varians 1. Signifikanssandsynligheden bliver da (dobbelsiddet test) = 2,4%. Så H0 skal forkastes, hvilket også kan konkluderes ud fra konfidensintervallet i forrige spørgsmål.

4. Test om antallet af mænd, der er blevet målt til at have kørt for stærkt svarer til antallet mænd, der kører på denne strækning.

I nedenstående er antallet af personer, der har kørt for stærkt, yderligere blevet inddelt efter alder.

	alder 18-50	alder 50+	ialt
mænd	538	172	710
kvinder	239	65	304
	777	237	1014

5. test om der er uafhængighed mellem køn og alder.

Et uafhængighedstest klares hurtigt i SAS med programmet

```
data a;
input sex $ alder $ antal;
cards;
m ung 538
m gam 172
k ung 239
```

```
k gam 65
;
proc freq data=a;
table sex*alder/norow nocol nopercent chisq;
weight antal;
run;
```

køres dette fås at teststørrelsen bliver 0,96 som er chi-i-anden fordelt med 1 frihedsgrad og signifikanssandsynligheden bliver 32%.

Der skal ikke forkastes og dermed er der tale om uafhængighed mellem køn og alder.

X_1, X_2 er to uafhængige poissonfordelinger med parametre λ_1 og λ_2.Her angiver X_1 antallet af mænd, der har overtrådt hastighedsbegrænsningen.

Tilsvarende med X_2 .

6. her skal man bruge at summen af to uafhængige poissonfordelinger igen er poisson fordelt. Når man betinger med summen får man at gøre med en binomialfordeling

hvor $n = 1014$ og $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$.Dette spørgsmål må betragtes som værende vanskeligt og dette bør indgå i debømmelsen.

$P(X=x) = (1/x!)\lambda^x e^{-x}$

$P(X_1 = x_1|X_1{+}X_2 = x_1{+}x_2) = \frac{P(X_1=x_1 og\ X_1=x_1)}{P(X_1+X_2=x_1+x_2)} = \frac{(1/x_1!)\lambda_1^{x_1}e^{-\lambda_1}(1/x_2!)\lambda_2^{x_2}e^{-\lambda_2}}{(1/(x_1+x_2)!)(\lambda_1+\lambda_2)^{(x_1+x_2)}e^{-(\lambda_1+\lambda_2)}} = \frac{(x_1+x_2)!{\lambda_1}^{x_1}{\lambda_2}^{x_2}}{(x_1)!(x_2)!(\lambda_1+\lambda_2)^{(x_1+x_2)}\ (\lambda_1+\lambda_2)^{(x_1+x_2)}}$

$= \binom{x_1+x_2}{x_1}p^{x_1}(1-p)^{x_2}$ hvor $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$