Rettevejledning til eksamen i økonometri A juni 2013

Spørgsmål 1

En fertilitetsklinik anbefaler på sin hjemmeside en behandling mod barnløshed, hvor sandsynligheden for succes per behandling er $0,15$, og de enkelte behandlinger er uafhængige. Lad Z være antallet af behandlinger indtil en succes.

1. Hvilken fordeling følger Z. Hvad er det ventede antal behandlinger?

Z er geometriskfordelt, hvor succesraten er 0,15. Det forventede antal behandlinger er $E[Z]+1=\frac{0{,}85}{0{,}15}+1=6\frac{2}{3}$.

Klinikken sælger forskellige pakker til sine kunder. I pakke A betales kr. 10.000 per behandling. I pakke B købes 5 behandlinger til kr. 40.000.

2. Hvad er den forventede pris for en succesfuld behandling, hvis kunden udelukkende benytter sig af henholdsvis pakke A og pakke B. (Hint: For at finde den forventede pris på en succesfuld behandling udelukkende ved at købe pakke B, skal sandsynligheden for en succesfuld behandling findes i en pakke B, inden det forventede antal pakker findes for en succesfuld behandling).

Den forventede pris for pakke A er givet ved $E[10.000Z]=10.000E[Z]\simeq$ 66000. Den forventede pris ved at bruge pakke B findes i to trin. Først findes sandsynligheden for mindst en succes ved 5 behandlinger. Lad X være antallet af successer ved 5 behandlinger, så er $X\tilde{}bin(5;0,15)$. $\Pr(X\geq 1)=1-\Pr(X=0)\simeq 0,56$. Dernæst skal antallet af pakker, der skal købes, beregnes. Som med de enkle behandlinger er denne geometrisk fordelt, da vi venter på en succesfuld behandling. Lad Y være antallet af pakke, $Y\tilde{}Geo(0,56)$. Den forventede pris bliver dermed $E[40.000Y]=40.000E[Y]=40.000(\frac{0{,}44}{0{,}56}+1)\simeq 71.000$.

Klinikken har i sin markedsføring ikke været 110 pct. ærlig, idet sandsynligheden for succes er mindre for kunder med en højere alder. Antag at hvis man er over 35 år er sandsynligheden for succes faldet til 10 pct.

3. Hvad er det forventede antal behandlinger for en tilfældig kunde, når halvdelen af kunderne er ældre end 35 år?

Her bruges loven om den totale sandsynlighed for forventninger. Lad Z være antallet af behandlinger til succes for en kunde. Lad D være en stokastisk variabel for om kunden er 35 eller derunder. Dvs.

$D = 1$ hvis kunden er 35 eller derunder

$D = 0$ hvis kunder er over 35.

$E[Z] = \Pr(D = 1)E[Z|D = 1] + \Pr(D = 0)E[Z|D = 0) = \frac{1}{2} \cdot 6\frac{2}{3} + \frac{1}{2} \cdot (\frac{0,9}{0,1} + 1) = \frac{1}{2} \cdot 6\frac{2}{3} + \frac{1}{2} \cdot 10 = 8\frac{1}{3}$

Spørgsmål 2

En landmand kan producere to afgrøder X og Y. Den simultane fordeling af afkastet er givet ved følgende tabel

$X\backslash Y$	100	110	120	130
60	$0,10$	$0,05$	$0,05$	0
90	$0,05$	$0,10$	$0,05$	$0,05$
120	$0,05$	$0,05$	$0,10$	$0,10$
150	0	$0,05$	0	$0,20$

1. Hvad er den marginale fordeling af X? Og hvad er den betingede fordeling af X givet Y er 110?

Den marginale fordeling af X angiver fordelingen af X uanset Y. Den fås ved at summere den simultate sandsynlighedsfunktion angivet ved tabellen hen over de enkle Y værdier.

x	60	90	120	150
f(x)	0,2	0,25	0,3	0,25

Den betingede fordeling af X givet Y er 110 angiver fordelingen af X, hvis vi ved at Y giver et afkast på 110. Den findes ved at tage den simultane sandsynlighedsfunktion angivet i søjlen for Y er 110 i tabellen og dividere den med sandsynligheden for at Y er 110.

x	60	90	120	150
f(x)	0,2	0,4	0,2	0,2

2. Hvad er $E[X]$ og $E[Y]$, og kovariansen mellem X og Y?

$E[X] = 108$ og $E[Y] = 117$. Kovariansen er $Cov(X, Y) = 204$. Dvs. større afkast af X er associeret med større afkast af Y.

Landmanden kan selv vælge hvor meget af X og Y han producerer. Lad θ være andelen af X og $1-\theta$ andelen af Y. Lad $Z = \theta X + (1-\theta)Y$ og $0 \leq \theta \leq 1$.

3. Hvad er $E[Z]$ og $Var(Z)$ som funktion af θ. Hvilken værdi af θ minimerer variansen?

$E[Z] = E[\theta X + (1-\theta)Y] = \theta E[X] + (1-\theta)E[Y]. Var(Z) = Var(\theta X + (1-\theta)Y) = \theta^2 Var(X) + (1-\theta)^2 Var(Y) + 2\theta(1-\theta)Cov(X,Y)$ (Corollary 4.6.1). For at finde minimum, finder vi først variansen på X og Y til at være hhv. Var(X)=1026 og Var(Y)=131. Så opskrives Var(Z) som et andengradspolynomium i $\theta : Var(Z) = (Var(X) + Var(Y) - 2Cov(X,Y))\theta^2 + 2(Cov(X,Y) - Var(Y))\theta + Var(Y)$.

Koefficienten til θ^2 kan så beregnes til 749. Koefficienten til θ kan beregnes til 146. Dermed er minimum i $\frac{-146}{2 \cdot 749} = -0,097$, dvs. udenfor intervallet $[0,1]$.Minimum er dermed enten i 0 eller 1. Når $\theta = 1$ dyrkes udelukkende afgrøde X og variansen er 1026 på Z og $\theta = 0$ dyrkes udelukkende afgrøde Y og variansen er 131. Dermed er variansen mindst når $\theta = 0$.

Spørgsmål 3

I forbindelse med en større international undersøgelse kaldet European Social Survey (ESS) er 1576 personer i Danmark og 1728 personer i Frankrig udvalgt simpelt tilfældigt. De er alle blevet stillet spørgsmålet om hvor tilfredse de er med retsvæsenet i deres hjemland. Svarene er givet på en skala fra 0 til 10 hvor 0 er yderst utilfreds og 10 er absolut tilfreds.

I det følgende kan det antages at normalfordeling kan beskrive de indsamlede data. Således at Danmark: $X_1, ...X_{1576} \sim N(\mu_1, \sigma_1^2)$ uafhængige og Frankrig:

$Y_1, ...Y_{1728} \sim N(\mu_2, \sigma_2^2)$ uafhængige

Resultaterne af undersøgelsen er gengivet i nedenstående tabel

	antal	gennemsnit	spredning
	N	$\bar{X}$	s
Danmark	1.548	7,35	1,96
Frankrig	1.724	4,91	2,36
I alt	3.272		

1. Estimer de 4 ukendte parametre

   $\widehat{\mu_1} = 7,35 \qquad \widehat{\mu_2} = 4,91$

   $\widehat{\sigma_1} = 1,96 \qquad \widehat{\sigma_2} = 2,36$

2. Udregn et 95% konfidensinterval for μ_1 (Danmarks gennemsnit).

   $[\bar{X}-1,96\sqrt{\frac{s^2}{n}};\bar{X}+1,96\sqrt{\frac{s^2}{n}}]$ [7,35-1,96$\frac{1,96}{\sqrt{1548}}$;7,35+1,96$\frac{1,96}{\sqrt{1548}}$] [7,25;7,44]

3. Angiv teststørrelsen for hypotesen at $\sigma_1^2 = \sigma_2^2$. Antag at de to varianser er ens. Estimer den fælles varians.

   F=$\frac{\max(s_1^2,s_2^2)}{\min((s_1^2,s_2^2)} = \frac{2,36^2}{1,96^2} = 1,45$ og man vil bruge F-fordelingen med frihedgraderne: 1724-1,1548-1.

   s$^2_{pooled}$ =S^2_p=$\frac{1723*2,36^2+1547*1,96^2}{1724+1548-2} = 2,18^2$

4. Test hypotesen at de to middelværdier er ens. Dvs. $\mu_1 = \mu_2$.Alternativet er her at Danmarks middelværdi er større end Frankrigs.

   test1=$\frac{\bar{X}_1-\bar{X}_2}{s_p^2\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{7,35-4,91}{2,18\sqrt{\frac{1}{1548}+\frac{1}{1724}}} = 32,0$ som er t-fordelt med med 1724+1548-2= 3270 frihedsgrader så teststørrelsen er i praksis normalfordelt.

   Aternativt kunne man have brugt teststørrelsen

   test2=$\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{7,35-4,91}{\sqrt{\frac{1,96^2}{1548}+\frac{2,36^2}{1724}}} = 32,3$ som er approximativt normalfordelt.

   Sss=signifikanssandsynlighed=p-værdi=P(Z>32,3; hvor Z er standardiseret normalfordelt)= ekstemt lille. Så H$_0$ forkastes og vi konkluderer at Danmarks gennemsnit er større end Frankrigs.

Man er bekymret for om kønsfordelingen er repræsentativ. Man ser derfor i første omgang på landene Danmark, Norge og Sverige og vil undersøge om der er ensartet kønsfordeling i disse tre lande. I nedenstående tabel er fordelingen vist.

ESS køn	Danmark	Norge	Sverige	i alt
Mænd	797	800	708	2.305
Kvinder	751	733	753	2.237
I alt	1.548	1.533	1.461	4.542

5. Test om fordelingen for køn er ens i de tre lande

   teststørrelsen der bruges er $\sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}$ hvor e_{ij} =rækkesum*søjlesum/totalsum

   så teststørrelsen Q=4,66 som er $chi^2(2)$ DF=(2-1)*(3-1)

   Sss=signifikanssandsynlighed=p-værdi= P(Q>4,66)=9,7%>5% hypotesen accepteres (opretholdes). Så kønsfordelingen er ensartet fra land til land.

6. Udregn 95% konfidensintervallet for andelen af kvinder i de tre lande.

   Da hypotesen er accepteret er det "uproblematisk" at se på den samlede kønsfordeling for de tre lande under et.

   andelen for kvinder bliver $\frac{2237}{4542} = 0,49$.

   Vi bruger binomialfordelingen og beregner $[\widehat{p} - 1,96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}; \widehat{p} + 1,96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}]$ $=$

   [0,49-1,96$\sqrt{\frac{0,49*0,51}{4542}}$;0,49+1,96$\sqrt{\frac{0,49*0,51}{4542}}$]= [0,48;0,51]