Rettevejledning økonometri A reeksamen sommer 2014

Opgave 1

Et flyselskab overbooker sine fly, dvs. at de sælger flere billetter til en afgang, end der er pladser i flyet. Årsagen er, at en del passagerer aflyser deres rejse i sidste øjeblik, og hvis de ikke har overbooket flyet vil det flyve med flere tomme sæder end ellers. Sandsynligheden for at en passagerer aflyser sin rejse er $0,05$.

1. Opstil en model for antallet af passager, n, der aflyser deres rejse i sidste øjeblik under antagelse om uafhængighed mellem aflysningerne. Er uafhængighed en realistisk antagelse? Hvis $n = 200$, hvad er det forventede antal aflysninger? Hvad er da sandsynligheden for at mindre end 10 passagerer aflyser deres rejse?

Lad X være antallet af passagerer, der aflyser deres rejse. n er antallet af passagerer på flyet. $X\tilde{}BIN(n; 0,05)$. Antagelsen om uafhængighed kan tage mange former. Fornuftige besvarelse af denne giver point. Det forventede antal aflysninger er $n \cdot p = 10$. $P(X < 10) = 0,4547$. Bemærk at opgaven fejlagtigt har skrevet n er antallet af passagerer, der aflyser. Der skulle have stået X.Så alternative fornuftige modeller skal give fuld point.

Flyselskabet sælger 200 rejser til 190 sæder. Selskabet tjener ca. 10 pct. på hver rejse den sælger og en rejse koster kr. 2000. Det hænder, at flyselskabet skal afvise passagerer, hvis der dukker flere passagerer op, end der er pladser i flyet. Hvis flyselskabet skal afvise en passager, betaler den en gratis rejse og en kompensation på k kroner for ikke at miste sit omdømme.

2. Hvad må kompensation højest være hvis flyselskabet i gennemsnit ikke skal tabe penge på overbooking?

Lad Y være fortjenesten ved at sælge 200 rejser. $Y = 0,1 \cdot 2000 \cdot 200 = 40.000$. Lad V være antallet af passagerer, der bliver afvist. $V = \max(10 - X, 0)$, hvor X er antallet af passagerer, der ikke møder op. Fx hvis 11 passagerer ikke møder op er $V = \max(-1, 0) = 0$. Ingen passagerer skal afvises. Ialt betales en kompensation på $V \cdot (2000 + k)$.

Følgende tabel skal opstilles for at beregne k. $P(V = 0) = P(X \geq 10) = 1 - P(X < 10)$ (jf. opgave 1). $P(V = 1) = P(X = 9)$ etc.

V	p
0	$0,55$
1	$0,13$
2	$0,11$
3	$0,09$
4	$0,06$
5	$0,04$
6	$0,02$
7	$0,01$
8	$0,00$
9	$0,00$
10	$0,00$

Fra denne tabel kan det forventede antal afviste beregnes til $E[V] = 1,219$. k kan nu beregnes som $40.000 - 1,219 \cdot (2000 + k) = 0$. $k = 30803,15$. Dvs. der er rigelig rum for at betale kompensation, da sandsynligheden er meget lille for at afvise 1 eller flere passagerer på en afgang.

Antag nu at der er to typer af passagerer. Business class og economy class. Business class betaler mere for deres rejse end economy class og de får en bedre forplejning. Skulle det ske at business class ikke er fyldt, vælger flyselskabet at opgradere economy class passagerer tilfældigt for at give dem en oplevelse. Antag at flyselskabet gør det uanset om der er ledige pladser på economy class eller ej. Economy class passagerer aflyser deres rejser med sandsynligheden $0,05$, men business class passagerer aflyser med sandsynligheden $0,1$.

På en afgang er der 30 udsolgte billettere til business class og 170 solgte billetter til economy class. Lad p være en stokastisk variabel for sandsynligheden at blive opgraderet.

3. Hvad er den forventede sandsynlighed for at blive opgraderet? (Hint: Sandsynligheden for at blive opgraderet er givet ved antallet af ledige pladser på business class divideret med antallet af fremmødte economy class passagere)

$p = \frac{L}{F}$, hvor L er antallet af ledige pladser på business class og F er antallet af passager, der møder frem på economy class. For at beregningen skal give mening må $F > 0$.Dette kan ignoreres her, da sandsynligheden er ekstrem lille for a $F = 0$.Der skal beregnes følgende $E[p] = E[\frac{L}{F}] = E[L] \cdot E[\frac{1}{F}] = 0,1 \cdot 30 \cdot E[\frac{1}{F}] = 3 \cdot E[\frac{1}{F}] = 3 \cdot 0,00619 = 0,0186$. Dvs. $1,6\%$ chance for at blive opgraderet.

Opgave 2

En virksomhed har N medarbejdere, der skal på efteruddannelseskursus for at styrke holdånden og lære at elske virksomheden. Det sker ved at de laver nogle fællesaktiviteter og 1 gang i timen råber virksomhedens navn. Med sandsynligheden 90 pct. lykkes det at få succes og få medarbejderne til at holde mere af deres virksomhed. Antag at der er uafhængighed mellem udfaldet af kurset for medarbejderne. Lad Y være antallet som har fået styrket deres holdånd og lært at elske virksomheden.

1. Antag $N = 100$. Hvad er forventninger til Y og variansen på Y. Er uafhængighed en realistisk forudsætning?

$Y\tilde{}BIN(100; 0,9)$.Derfor er $E[Y] = 100 \cdot 0,9 = 90$ og $Var(Y) = 100 \cdot 0,9 \cdot 0,1 = 9$. Rimelige diskussioner af uafhængighed giver point.

Lad nu $N = 10.000$.

2. Brug normalfordelingen til at approksimere Y. Hvad er medianen, 1. og 3. kvartil. Giv en fortolkning.

$Y\tilde{}N(np, np(1-p))$.Medianen er lig middelværdien i normalfordelingen, da den er symmetrisk, så medianen er $10.000 \cdot 0,9 = 9000$. 1. og 3. kvartil er givet ved 8980 og 9020. medianen svarer til den værdi, hvor halvdelen af antallet, som har styrket deres holdånd, er over og halvdelen af antallet, som har styrket deres holdånd, er under. Tilsvarende er 1. kvartil den værdi, hvor 75 pct. af antallet, som har styrket deres holdånd, er over og 25 pct. af antallet, som har styrket deres holdånd, er under. Endelig er 3. kvartil den værdi, hvor 25 pct. af antallet, som har styrket deres holdånd, er over, og 75 pct. af antallet, som har styrket deres holdånd, er under.

Lad W_1 være en normalfordelingsapproksimation til Y. Virksomheden sender

resten af virksomhedens medarbejdere på et andet efteruddannelseskursus med samme formål. Antag at antallet af succeser for disse kan approksimeres ved $W_2 \sim N(11000, 4000)$. Antag der er uafhængighed mellem de to grupper af medarbejdere, dvs. W_1 og W_2.

3. Er det realistisk at antage uafhængighed mellem de to grupper? Hvad er fordelingen af summen, $W_1 + W_2$, af de to stokastiske variable?

Uafhængighed skal diskuteres og alle rimelige argumenter giver point. Summen af to normalfordelte stokastiske variable er også normal fordelte. Dvs. $W_1 + W_2\tilde{}N(11.000 + 9000, 4000 + 900)$

Opgave 3

Den følgende tabel viser målinger af højden på 75 værnepligtige fra by A og 50 værnepligtige fra by B.

v grænse	h grænse	A	B
152,5	157,5	1	0
157,5	162,5	3	1
162,5	167,5	10	5
167,5	172,5	17	10
172,5	177,5	25	18
177,5	182,5	14	13
182,5	187,5	5	2
187,5	192,5	0	1
	I alt	75	50
	Gennemsnit	$173,3$	$174,7$
	SAK	$3124,7$	$1720,5$

Der opstilles nu den model at
X_i er normalfordelt med middelværdi μ_1 og varians σ_1^2, $i = 1, \ldots, 75$
og
Y_i er normalfordelte med middelværdi μ_2 og varians σ_2^2, $i = 1, \ldots, 50$

1. Estimer de 4 parametre og angiv deres egenskaber

   $\widehat{\mu}_1 = \overline{x} = 173,3 \quad \widehat{\mu}_2 = \overline{y} = 174,7$

   $\widehat{\sigma}_1^2 = s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{3124,7}{75-1} = 42,22 \qquad \widehat{\sigma}_2^2 = s_2^2 = \frac{1720,5}{50-1} = 35,11$

   Estimatorerne er konsistente (og dermed centrale)

2. Test om $\sigma_1^2 = \sigma_2^2$ mod alternativet at $\sigma_1^2 \neq \sigma_2^2$

   F-test$=\frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)} = \frac{42,22}{35,11} = 1,2$

   her skal F-fordelingen med frihedsgrader 74,49 bruges.

   Sss=p-værdi=2*P(F(74,49))>1,2)=2*[1-P(F(74,49)<1,2)]=2*(1-0,75)=50%.

   Hypotesen om ens varianser opretholdes. og dermed fås at den fælles varians bliver
   $s_p^2 = \frac{(75-1)*s_1^2+(50-1)*s_2^2}{75+50-2} = 39,39$

Det påstås, at værnepligtige fra B har en større gennemsnitshøjde end værnepligtige fra by A.

3. Test hypotesen at $\mu_1 = \mu_2$ begrund valg af alternativ hypotese

   $\mathrm{H}_0 : \mu_1 = \mu_2 \quad \mathrm{H}_A : \mu_1 <> \mu_2$ Der er ikke nogen begrundelse for at vælge ensidet t= $\frac{173{,}3-174{,}7}{\sqrt{39{,}39*(\frac{1}{75}+\frac{1}{50})}} = -1,22$ som er t-fordelt med 75+50-2=123 frihedsgrader

Sss=p-værdi=2*P(t(123)<-1,22)= 22,4% større end de 5% så fortsat ingen forskel

4. Udregn sandsynligheden for at en tilfældigt udvalgt værnepligtig fra by A har en højde der er mindre end 170 cm.

   I og med at der er testet for at de to gennemsnit er ens er det naturligt at estimere det fælles gennemsnit

   $\mu_f = \frac{75*173{,}3+50*174{,}7}{75+50} = 173,9$ man kan så diskutere om man skulle reestimere den fælles varians. Men lad nu det være

   P(X<170) når X er N(173,9; 6,28^2)=27%

I en ældre opgørelse over levevilkårene i Danmark, finder man nedenstående tabel. Tabellen viser arbejdsløse fordelt efter alder og arbejdsløshedens varighed indenfor de sidste 5 år.

Aldersgruppe	under 5 mdr.	6-11 mdr.	12+ mdr.	I alt
20 – 29	37	18	30	85
30 – 49	13	13	28	54
50 – 69	9	8	20	37
I alt	59	39	78	176

5. Opstil en statistisk model for ovenstående tabel

   Det mest "naturlige" er vel en multinomisk fordeling hvor n=176 og der er 6 kasser

6. Er der sammenhæng mellem alder og arbejdsløshedens længde?

   dette vil så være et uafhængighedstest. teststørrelsen vil blive 8,09 som er χ^2 med 4 frihedsgrader.

   Sss=p-værdi bliver 8,8% $>$ 5% og dermed opretholder vi hypotesen om uafhængighed.