

Løsningsforslag re-eksamen vinter 2012-2013 Økonometri A 2. Årsprøve

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 funktionen, sufficiens, stikprøvefunktion, egenskaber ved stikprøvefunktion, estimation herud af 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

Spørgsmål 1

En producent sælger to produkter, X_1 og X_2 . Salget af de to produkter varierer stokastisk som vist i følgende tabel:

	X_2			
X_1	0	1	2	3
0	0,02	0,02	0,08	0,18
1	0,02	0,06	0,10	0,10
2	0,04	0,08	0,04	0,08
3	0,08	0,04	0,02	0,04

Profitmargin på X_1 og X_2 er henholdsvis kr. 5000 og kr. 10000. Profitten, Y , er dermed: $Y = 5000X_1 + 10000X_2$.

1. Opstil udfaldsrum for Y og tilhørende sandsynligheder for Y . Hvad er sandsynligheden for at profitten er højere end kr. 30000?

y	0	5000	10000	15000	20000	25000	30000	35000	40000	45000
f(y)	0,02	0,02	0,06	0,14	0,16	0,14	0,22	0,12	0,08	0,04

$$\Pr(Y > 30000) = 0,24$$

2. Hvad er den forventede profit givet der ikke er noget salg af X_1 ?

$$E[Y|X_1 = 0] = \sum_y yf(y|X_1 = 0) = \sum_{x_2} 10000x_2f(x_2|X_1 = 0) = 0 \cdot \frac{0,02}{0,4} + 10000\frac{0,02}{0,4} + 20000\frac{0,08}{0,4} + 30000\frac{0,18}{0,4} = 18000$$

I et anfald af godhed beslutter ejeren at forære en del af profitten til velgørenhed. Han beslutter at give den del væk han tjener mindst på, dvs $Z = \min(5000X_1, 10000X_2)$.

3. Opstil udfaldsrummet og de tilhørende sandsynligheder for Z . Hvad er det forventede beløb, som gives væk til velgørenhed?

z	0	5000	10000	15000
f(z)	0,44	0,26	0,24	0,06

$$E[Z] = 0,54 \cdot 0 + 0,26 \cdot 5000 + 0,24 \cdot 10000 + 0,06 \cdot 15000 = 4600$$

0.1 Spørgsmål 2

En butik sælger pc'ere. Antag at salget følger en Poisson process med 4 solgte om dagen.

1. Hvad er sandsynligheden for at sælge mindst 30 pc'er i løbet af 6 dage?

Lad X være antallet af solgte pc'ere på 1 dage.

$$X_6 \sim \text{Pois}(4 \cdot 6)$$

$$\Pr(X_6 \geq 30) = 1 - \Pr(X_6 < 30) = 0,13$$

For at minimere omkostningerne holder butikken et lille lager af pc'ere. Der leveres varer hver 14. dag, og butikken holder lukket en dag om ugen.

2. Hvis sandsynligheden for at lageret tømmes højst må være 5 pct., hvor mange pc'ere skal butikken da have stående på lager?

Der er 12 dage åbne, dvs. vi skal finde lagerstørrelsen, S , der opfylder $\Pr(X_{12} > S) = 1 - \Pr(X_{12} \leq S) \leq 0,05$

Denne findes ved at lave en lille tabel over $1 - \Pr(X_{12} \leq S)$ for forskellige værdier af S .

S	58	59	60	61
$1 - \Pr(X_{12} \leq S)$	0,068	0,052	0,040	0,029

Når $S = 60$ er sandsynligheden mindre end 5 pct.

Du kommer en dag ligefør lukketid forbi butikken og går ind og spørger efter en pc. Til din overraskelse er der ingen på lager.

3. Hvad er den forventede ventetid (målt i åbningsdage) til der kommer pc'ere på lager. Antag at sandsynligheden for at vente mere end 4 dage er 0.

Først skal sandsynligheden findes for at ventetiden er 1, 2, 3 eller 4 dage. Lad T være antal dage du venter. Så er X_{12-T} antal solgte enheder til dag 12-T. Nu kan sandsynligheden for at der er udsolgt 1,2,3 eller 4 dage før tiden beregnes. Denne er sandsynlighed (søjle 2) er den ubetingede sandsynlighed. Beting nu på at $T > 0$ (søjle 3),

T	$1 - P(X_{12-T} \leq 60)$	$P(T T > 0)$
1	0,00875	0,8710
2	0,001201	0,1196
3	$9,13E - 5$	0,0091
4	$3,31E - 6$	0,0003

Den forventede ventetid kan nu bestemmes af $E[T] = 1 \cdot 0,8710 + 2 \cdot 0,1196 + 3 \cdot 0,0091 + 4 \cdot 0,0003 = 1,1387$

0.2 Spørgsmål 3

I den sidste offentliggjorte PISA undersøgelse som blev foretaget i foråret 2009, blev der stillet følgende spørgsmål til de 5.924 elever: "School has been a waste of time".

Spørgsmålet kan besvares på en skala fra 1 til 4.

1="meget uenig"

2="uenig"

3="enig"

4="meget enig"

	antal	gennemsnit	spredning
	N	\bar{X}	s
dreng	2.886	1,68	1,38
piger	3.038	1,80	1,41
ialt	5.924		

1. Argumenter for at gennemsnittet for henholdsvis drenge og piger kan beskrives med en normalfordeling.

Her bruges den centrale grænseværdisætning. med en vis rimelighed kan det antages at middelværdi

og varians for henholdsvis drenge og piger er ens. Antallet af uafhængige gentagelser er henholdsvis 2.886 og 3.038.

Dermed kan det antages at gennemsnittene er asymptotiske normalfordelte. dvs.

$$X_d \sim N(\mu_d, \delta_d^2) \quad d = 1, \dots, 2.886 \quad X_p \sim N(\mu_p, \delta_p^2) \quad p = 1, \dots, 3.038$$

2. Estimer parametrene for henholdsvis drenge og piger

	antal	gennemsnit	spredning
	N	$\hat{\mu} = \bar{X}$	$\hat{\delta} = s$
dreng	2.886	1,68	1,38
piger	3.038	1,80	1,41
ialt	5.924		

3. Udregn et 95% konfidensinterval for pigernes middelværdi.

Der bruges $\bar{X} \pm t_{0,975}(n-1) \frac{s}{\sqrt{n}}$ her er antallet af frihedsgrader stort så $t_{0,975}(n-1) = 1,96$

$$1,80 \pm 1,96 \frac{1,41}{\sqrt{3.038}} = [1,75; 1,85]$$

4. Test hypotesen at drenge og piger er ens mht. at betragte skolen som "waste of time". Kommenter resultatet

herunder stikprøvestørrelsen.

$H_0 : \mu_d = \mu_p$ mod $H_A : \mu_d <> \mu_p$ (dobbeltest)

På grund af det store antal observationer kan vi bruge følgende teststørrelse

$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$ som vil være normalfordelt med middelværdi 0 og varians 1 givet H_0 er sand.

$$U = \frac{1,68 - 1,80}{\sqrt{1,38^2/2.886 + 1,41^2/3.038}} = -3,31. \text{ Så p-værdien bliver meget lille (0,0009).}$$

Det er værd at notere, at der er mange observationer. Hypotesen forkastes og dermed konkluderes at der er forskel på pigers og drenges opfattelse.

Uanset hvad du har konkluderet i ovenstående test så

antag nu at den samlede population på 5.924 kan beskrives med en normalfordeling

med parametre μ og δ^2 . dvs $X \sim N(\mu, \delta^2)$

5. Estimer parametrene μ og δ .

$$\hat{\mu} = \frac{2.886 \cdot 1,68 + 3.038 \cdot 1,8}{2.886 + 3.038} = 1,74$$

Hvis man bruger nedenstående formel til løsning af estimation af δ^2 er det OK. Dette er faktisk ikke helt korrekt da vi nu antager, at middelværdierne er ens.

$$\hat{\delta}^2 = s^2 = \frac{(2.886-1)*1,38^2 + (3.038-1)*1,41^2}{2.886+3.038} = (1,40)^2$$

I nedenstående tabel er vist fordelingen for de to køn, mht. besvarelsen af ovenstående spørgsmål.

I tabellen er svaret "enig" og "meget enig" slået til et svar "enig"

6. test om der er uafhængighed mellem køn og opfattelsen af skolen.

	meget uenig	uenig	enig	i alt
dreng	1499	1053	334	2886
piger	1781	1026	231	3038
i alt	3280	2079	565	5924

En mulig model kunne være

$(X_1, X_2, X_3) \sim \text{multinomisk}(2.886, p_1, p_2, p_3,)$ og $(Y_1, Y_2, Y_3) \sim \text{multinomisk}(3.038, q_1, q_2, q_3,)$

$H_0 : p_1 = q_1, p_2 = q_2, p_3 = q_3$ $H_A : \text{mindst et sted er der forskel}$

testtørrelsen bliver χ^2 fordelt med $(3-1)*(2-1) = 2$ frihedsgrader

$\chi^2 = 39,5$ og dermed bliver p-værdien meget lille $p < 0,001$. Hypotesen skal forkastes, der er forskel på drenge og piger mht. til besvarelse af dette spørgsmål.