

Reeksamen i Statistik 2. år

Skriftlig prøve (4 timer)

28. august 2009 kl. 9.00–13.00

Eksamenssættet er på 3 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

Opgave 1

Lad (X, Y) være en stokastisk variabel på $\{1, 2, 3\} \times \{1, 2, 3\}$ med sandsynlighedsfunktion $p(x, y)$ givet ved følgende tabel:

	y	1	2	3
x				
1		0.1	0.3	0.0
2		0.1	0.1	0.1
3		0.0	0.2	0.1

- Udregn $P(\{X \leq 2\} \cup \{Y \leq 2\})$.
- Opskriv sandsynlighedsfunktionerne for de marginale fordelinger af X og Y . Er X og Y stokastisk uafhængige?
- Opskriv sandsynlighedsfunktionen for den betingede fordeling af Y , givet $X = 2$, og udregn middelværdi og varians i denne fordeling.

Opgave 2

Lad X være en stokastisk variabel med tæthed

$$p(x) = \begin{cases} \frac{x+1}{2} & \text{for } x \in]-1, 1[\\ 0 & \text{ellers.} \end{cases}$$

- Tegn funktionen $p(x)$ og gør rede for at den faktisk er en tæthed. Udregn middelværdi og varians for X .
- Opskriv fordelingsfunktionen for X . Bestem fordelings 25%-fraktil og dens 50%-fraktil.
- Lad X_1, \dots, X_{10} være uafhængige med samme fordeling som X . Definér nye stokastiske variable Y_1, \dots, Y_{10} ved

$$Y_i = \begin{cases} 1 & \text{hvis } X_i < 0 \\ 0 & \text{ellers.} \end{cases}$$

Beskriv fordelingen af $S = Y_1 + \dots + Y_{10}$, og udregn middelværdi og varians for S .

Opgave 3

Følgende tabel viser antal af dødsfald i i Vestaustralien 1982, opdelt på køn og fire hovedgrupper af dødsårsager.

Køn:	M	K	I alt
Dødsårsag:			
Kredsløb, hjertelateret	1317	854	2171
Kræft	1119	828	1947
Kredsløb, hjernerelateret	371	460	831
Ulykke	346	147	493
I alt	3153	2289	5442

(a) Opstil en tabel hvor antal dødsfald i hver enkelt celle angives i procent af den tilhørende søjlesum. Forklar disse tal som estimerede sandsynligheder i en passende model, og benyt dem til at beskrive de forskelle der tilsyneladende er mellem mænd og kvinder.

(b) Betragt den statistiske model der beskriver tallene i de 8 celler som uafhængige og Poissonfordelte, i første omgang med hver sin parameter. Foretag testet for multiplikativitet og forklar konklusionen.

(c) Givet at en person dør af en kredsløbssygdom, hvad er sandsynligheden for at den er hjertelateret? Dette spørgsmål må i første omgang besvares for hvert køn for sig, og når vi spørger hvad "sandsynligheden" er menes der naturligvis et passende estimat for denne sandsynlighed. Men derudover ønskes også udført et test for, om der overhovedet er forskel mellem de to køn hvad dette angår.

Opgave 4

Denne opgave handler om et ret gammelt datasæt vedrørende bilers bremselængder. Man har for 50 opbremsninger registreret

x Bilens hastighed i mph. (miles per hour)
y Bremselængden i ft. (feet, engelske fod)

Det gav anledning til følgende datasæt:

X	Y	X	Y	X	Y	X	Y	X	Y
4	10	11	17	14	80	17	32	20	52
4	2	12	24	14	26	18	76	20	32
7	22	12	28	14	36	18	84	20	48
7	4	12	14	15	54	18	42	22	66
8	16	12	20	15	26	18	56	23	54
9	10	13	34	15	20	19	68	24	93
10	34	13	46	16	40	19	46	24	120
10	26	13	26	16	32	19	36	24	70
10	18	13	34	17	50	20	56	24	92
11	28	14	60	17	40	20	64	25	85

(a) Estimer parametrene i en simpel regressionsmodel med x som forklarende variabel og y som respons. For hældningens vedkommende ønskes angivelse af 95% sikkerhedsgrænser. Følgende mellemregningsstørrelser kan benyttes:

$$4 + 4 + \dots + 25 = 770$$

$$10 + 2 + \dots + 85 = 2149$$

$$4^2 + 4^2 + \dots + 25^2 = 13228$$

$$10^2 + 2^2 + \dots + 85^2 = 124903$$

$$4 \times 10 + 4 \times 2 + \dots + 25 \times 85 = 38482$$

(b) Estimer, med angivelse af 95% sikkerhedsgrænser, den forventede bremselængde for en bil som kører 20 mph.

(c) En bremselængde er sum af to bidrag, hvor det første skyldes reaktionstiden og det andet er det stykke bilen kører medens den bremses ned. Ifølge almindelige fysiske regler skulle det første bidrag være proportionalt med hastigheden, medens det andet er proportionalt med kvadratet på hastigheden. Det er derfor nærliggende at forsøge med en model af formen

$$E y_i = \beta x_i + \gamma x_i^2.$$

Nedenfor er gengivet parameterestimer m.v. i denne model. Desuden kan det oplyses at variansen i denne model blev estimeret til 225.65. Hvad kan man konkludere?

	Estimate	Std.dev.	T	P
X	1.239	0.5600	2.213	0.031710
X*X	0.0901	0.02939	3.067	0.003546