

Eksamen i Statistik 2. år

Skriftlig prøve (4 timer)

19. juni 2006 kl. 9.00–13.00

Eksamenssættet er på 4 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

Opgave 1

En symmetrisk mønt kastes 10 gange. Vi lader $(X_1, \dots, X_{10}) \in \{0, 1\}^{10}$ betegne udfaldet, idet vi som sædvanligt koder plat som 0 og krone som 1. Vi sætter $S = X_1 + \dots + X_{10}$ = det samlede antal krone.

- (a) Beregn $P(S = 3)$.
- (b) Beregn middelværdi og standardafvigelse for S , samt middelværdien af den stokastiske variabel 2^S .
- (c) Beskriv den betingede fordeling af S , givet at de to første kast giver plat.

Opgave 2

Lad X_1 og X_2 være uafhængige, eksponentialfordelte stokastiske variable med skalaparametre henholdsvis 1 og 2. Sagt på en anden måde, X_1 og $X_2/2$ er uafhængige, normeret eksponentialfordelte. Vi sætter $S = X_1 + X_2$.

- (a) Udregn middelværdi og varians for S .
- (b) Udregn tætheden for S .
- (c) Gør rede for at $Z = X_1 \wedge X_2 = \min\{X_1, X_2\}$ er eksponentialfordelt med skalaparameter $\frac{2}{3}$ (Vink: Udregn først $P(Z > z)$).

Opgave 3

I nedenstående tabel er angivet antallet af børn (dvs. personer under 18 år) der kom til skade i trafikken på cykel i år 2002, opdelt efter køn og region (Hovedstadsområdet inklusive Nordsjælland og Roskilde Amt, resten af øerne, og Jylland).

	Drenge	Piger	I alt
Hovedst.	56	32	88
Øerne	43	31	74
Jylland	89	59	148
I alt	188	122	310

(a) Idet antallene i tabellens seks celler opfattes som observationer af uafhængige Poissonfordelte variable X_{rk} (r for region, k for køn), foretag testet for hypotesen om multiplikativitet og forklar konklusionen.

(b) Er risikoen for tilskadekomst på cykel den samme for drenge og piger? Det er meningen, at man skal besvare dette spørgsmål *uden* at skele til de oplysninger der følger under spørgsmål (c). Til gengæld er man så nødt til at antage, at der i alle tre regioner er ca. lige mange drenge og piger.

(c) En multiplikativ Poisson (eller log-lineær) model for disse tal, som tager hensyn til størrelserne af de underliggende befolkningsgrupper, er

$$EX_{rk} = \lambda_{rk} = m_{rk} \exp(\gamma + \alpha_r + \beta_k)$$

hvor m_{rk} er antal børn af køn k med bopæl i region r pr. 1. juli 2002, som givet ved følgende tabel:

	Drenge	Piger	I alt
Hovedst.	192529	182389	374918
Øerne	119382	114220	233602
Jylland	291578	276828	568406
I alt	603489	573437	1176926

Det viser sig, at denne model bliver godkendt ved test imod den fulde model ($P=0.78$). Parameterestimaterne (med betegnelser der burde være til at gætte, standard konventioner for opnåelse af entydig parametrisering ved at sætte udvalgte parametre til 0, og i sidste søjle P -værdierne for de sædvanlige tests for om de enkelte parametre kan sættes lig med 0) ser sådan ud:

	estimat	std.afv.	U	P
CONSTANT	-8.467	0.1083	-78.171	0.000000
REGION[1]	-0.104	0.1346	-0.772	0.439974
REGION[2]	0.197	0.1424	1.382	0.166901
REGION[3]	sat lig med nul			
KØN[1]	0.382	0.1163	3.283	0.001025
KØN[2]	sat lig med nul			

Hvad kan man heraf slutte om indflydelsen af køn og bopælsregion på børns risiko for tilskadekomst på cykel?

Opgave 4

I april 2005 er fra “www.bilbasen.dk” hentet følgende oplysninger om 195 brugte varebiler af mærket Toyota Hiace, som var annonceret til salg af forhandlere:

x : Alderen (i år)

z : Antal km kørt (i 1000 km)

y : Prisen (i 1000 kr).

x	z	y	x	z	y	x	z	y	x	z	y
0	5	160	4	61	125	5	119	85	6	79	80
0	1	185	4	65	120	5	214	70	6	146	90
0	3	155	4	52	120	5	48	100	7	153	65
0	2	155	4	38	120	5	89	90	7	119	80
1	3	150	4	54	110	5	51	120	7	59	80
1	15	170	4	32	135	5	97	89	7	173	55
1	5	160	4	117	85	5	126	85	7	123	75
1	14	170	4	216	60	5	86	100	7	126	80
1	36	135	4	82	90	5	120	90	7	184	55
1	19	160	4	126	80	5	122	85	7	105	70
1	15	160	4	61	115	5	162	70	7	224	55
1	7	150	4	62	125	5	50	95	7	229	60
1	1	200	4	38	120	5	46	95	7	127	80
2	16	150	4	100	80	5	116	85	7	85	70
2	20	145	4	62	95	5	111	95	8	98	80
2	22	144	4	40	110	5	175	70	8	198	60
2	15	145	4	86	85	5	163	70	8	241	65
2	35	145	4	77	120	5	99	100	8	80	75
2	11	150	4	20	130	5	155	100	8	183	70
2	17	120	4	84	100	6	155	75	8	202	55
2	34	140	4	125	85	6	145	75	8	85	60
2	45	140	4	103	95	6	122	77	8	126	60
2	17	150	4	58	100	6	103	77	8	151	63
2	45	145	4	65	110	6	68	85	8	177	54
2	21	145	4	42	140	6	109	90	8	145	65
3	55	115	4	119	90	6	183	75	8	160	60
3	56	115	4	86	100	6	88	90	8	144	57
3	35	135	4	175	95	6	84	103	8	85	75
3	19	125	4	63	95	6	74	105	8	169	60
3	28	125	5	80	105	6	93	98	8	168	67
3	24	130	5	118	95	6	65	70	9	212	60
3	15	135	5	105	75	6	159	96	9	330	35
3	63	125	5	89	80	6	122	75	9	148	60
3	37	125	5	119	83	6	315	50	9	224	50
3	31	140	5	119	100	6	173	65	9	205	45
3	41	140	5	115	90	6	220	60	9	173	50
3	22	140	5	129	80	6	186	80	9	186	40
3	50	120	5	78	100	6	84	90	9	207	76
3	89	130	5	86	100	6	59	90	9	160	60
3	53	100	5	152	90	6	72	85	9	199	55
3	27	125	5	98	100	6	215	66	9	174	57
3	58	140	5	135	95	6	71	90	10	134	45
3	34	123	5	53	89	6	65	90	10	180	35
3	32	123	5	92	80	6	68	90	11	210	45
4	95	105	5	156	90	6	110	85	11	150	50
4	52	130	5	66	100	6	106	80	12	116	35
4	53	135	5	135	79	6	73	90	12	163	38
4	44	118	5	82	120	6	90	85	13	320	25
4	50	120	5	95	90	6	88	80			

(a) Estimer parametrene i en simpel regressionsmodel for prisen y som funktion af alderen x . For hældningen ønskes angivelse af 95% konfidensgrænser. Til udregningerne kan benyttes følgende hjælpstørrelser:

$$\begin{aligned} S_x &= 1003 & S_y &= 18767 \\ SS_x &= 6301 & SS_y &= 2018875 \\ SP_{xy} &= 82516 \end{aligned}$$

(b) Estimer med angivelse af 95% sikkerhedsgrænser den forventede pris for en bil som er 4 år gammel.

(c) I en multipel regressionsmodel, hvor priserne y_i antages uafhængige og normalfordelte med samme varians σ^2 og middelværdier

$$\mu_i = \alpha + \beta x_i + \gamma z_i$$

bliver estimaterne for middelværdiparametrene og disses estimerede standardafvigelse

Parameter	Estimat	Std.afv.
α	158.4	2.002
β	-8.34	0.533
γ	-0.194	0.0197

medens estimatet for σ^2 bliver 141.34. Hvilke konklusioner kan man drage af det?