

# Reeksamen i Statistik 2. år

Skriftlig prøve (4 timer)

28. august 2008 kl. 9.00–13.00

Eksamenssættet er på 4 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

## Opgave 1

Lad  $X_1, \dots, X_{10}$  være uafhængige stokastiske variable, ligefordelte på mængden  $\{1, 2, \dots, 6\}$  (tænk på 10 kast med en terning), og lad  $S \in \{10, 11, \dots, 60\}$  betegne summen  $S = X_1 + \dots + X_{10}$  (= det samlede antal øjne).

- (a) Hvad er sandsynligheden for at  $X_i \leq 5$  for alle  $i$  (altså sandsynligheden for ikke at få en eneste sekser)?
- (b) Udregn  $P(S \leq 11)$  samt den betingede sandsynlighed  $P(S = 11 \mid S \leq 11)$ .
- (c) Udregn middelværdi og varians for  $S$ , og benyt den centrale grænseværdisætning til approksimativ beregning af  $P(S \leq 30)$ .

## Opgave 2

Lad  $X_1$  og  $X_2$  være uafhængige normalfordelte stokastiske variable med

$$EX_1 = 3, \quad \text{var}(X_1) = 16,$$

$$EX_2 = 0, \quad \text{var}(X_2) = 1.$$

- (a) Hvad er sandsynligheden for at  $X_1 \leq 5$ ?
- (b) Hvad er fordelingen af  $X_1 + X_2$ ?
- (c) Sæt  $X_3 = X_1 + 2X_2$ . Udregn kovarians og korrelation mellem  $X_1$  og  $X_3$ .

### Opgave 3

Følgende tabel viser antallet af personer, som er kommet til skade ved kørsel på knallert i Danmark i året 2002, fordelt på aldersgrupper og køn (tallene helt til højre ser vi bort fra i første omgang):

Køn:	M	K	I alt	
Aldersgruppe:				befolkning
-17	373	65	438	1176926
18-24	151	21	172	426795
25-44	230	33	263	1572208
45-64	136	30	166	1397841
65-	40	6	46	794584
I alt	930	155	1085	5368354

Som statistisk model benytter vi den der opfatter antallene i de 10 celler som uafhængige og Poissonfordelte, i første omgang med hver sin sandsynlighedsparameter  $\lambda_{ak}$  ( $a$ =aldersgruppe,  $k$ =køn).

(a) Foretag et test for hypotesen  $\lambda_{ak} = \alpha_a \beta_k$ , og forklar hvad hypotesen betyder.

(b) Det ser stærkt ud til, at der er færre kvinder end mænd som kommer til skade på knallert. Forklar hvad dette betyder for parametrene i modellen, og underbyg påstanden med et formelt test for hypotesen om at mænd og kvinder har samme risiko for at komme til skade på knallert.

(c) De store tal i søjlen til højre er aldersgruppernes størrelser i den danske befolkning. Det er nærliggende at forsøge med en model for rækkesummerne i antalstabellen, som går ud på at antal tilskadekomne i aldersgruppe  $a$  er Poissonfordelt med middelværdi af formen  $\beta n_a$ , hvor  $n_a$  er aldersgruppens størrelse. Gør rede for, at denne model ikke holder. Her kan man vælge at foretage et formelt test, men det er nok hvis man udregner modellens fittede værdier og gør rede for, at nogle af observationerne ligger alt for langt fra de tilhørende fittede værdier.

## Opgave 4

For 190 personer, udtaget tilfældigt fra et større socialmedicinsk datasæt fra 1976 (Østerbroundersøgelsen) er i udskriften på side 4 oplyst

```
VÆGT          Vægten i kilo
RYGER         Ja for ryger, Nej for ikke-ryger
KØN          M for mænd, K for kvinder
```

I forbindelse med udskriften under spørgsmål (c) får man brug for, at for faktoren RYGER er Ja niveau 1, Nej er niveau 2, medens for faktoren KØN er M niveau 1 og K niveau 2.

I første omgang ser vi bort fra faktoren KØN, og interesserer os kun for beskrivelsen af VÆGT som funktion af RYGER ved hjælp af en ensidet variansanalysemodel.

(a) Præciser modellens forudsætninger, foretag Bartlett's test og estimer modellens parametre.

(b) Afhænger folks vægt (ifølge dette datasæt) af om de er rygere?

(c) I en tosidet variansanalyse med VÆGT som respons og faktorerne RYGER og KØN som forklarende blev hypotesen om "ingen vekselvirkning" godkendt. Nedenstående skema viser parameterestimaterne i den additive model. Hvad kan man konkludere af dette angående betydningen af køn og rygning for vægten?

	Estimate	Std.dev.	T	P
CONSTANT	66.03	1.269	52.029	0.000000
RYGER[1]	-3.44	1.489	-2.308	0.022077
RYGER[2] set to zero				
KOEN[1]	11.93	1.479	8.066	0.000000
KOEN[2] set to zero				

Til udregningerne i forbindelse med den ensidede variansanalysemodel kan følgende hjælpestørrelser benyttes:

RYGER	Ja	Nej
Antal observationer	117	73
Sum	7931	5118
Kvadratsum	553819	367528

## Data til opgave 4.

RYGER	KØN	VÆGT									
Ja	M	96	Nej	K	66	Ja	K	49	Ja	K	56
Nej	K	67	Ja	M	66	Nej	M	89	Ja	K	56
Ja	K	59	Ja	M	84	Nej	K	65	Nej	M	74
Ja	K	75	Ja	M	72	Ja	M	70	Nej	M	67
Ja	M	79	Ja	M	72	Ja	K	56	Ja	K	65
Nej	M	84	Ja	K	54	Ja	M	59	Ja	M	83
Nej	K	62	Nej	K	62	Nej	K	66	Ja	K	57
Ja	K	61	Ja	K	70	Nej	K	71	Ja	K	50
Nej	K	60	Nej	K	77	Nej	K	65	Ja	M	68
Ja	M	64	Ja	K	57	Ja	K	74	Nej	K	63
Nej	M	78	Ja	M	76	Nej	K	67	Nej	K	59
Nej	M	71	Ja	M	69	Nej	K	60	Ja	K	56
Nej	M	78	Ja	K	58	Ja	K	58	Ja	M	73
Ja	K	72	Ja	M	88	Nej	K	55	Ja	K	53
Ja	K	62	Ja	K	55	Ja	K	55	Nej	K	77
Nej	K	57	Ja	M	54	Nej	K	59	Ja	K	78
Ja	K	62	Ja	M	64	Ja	M	77	Ja	K	72
Nej	K	60	Nej	K	67	Ja	K	66	Nej	K	57
Ja	K	48	Ja	M	64	Ja	K	57	Ja	K	68
Nej	K	51	Ja	K	61	Ja	M	79	Ja	M	78
Ja	K	66	Nej	K	74	Ja	K	59	Nej	M	88
Ja	K	61	Ja	K	86	Ja	M	66	Nej	K	100
Ja	K	55	Ja	K	74	Ja	K	59	Nej	K	67
Nej	K	77	Nej	K	72	Nej	M	67	Ja	K	70
Ja	M	105	Nej	M	63	Nej	K	65	Ja	M	78
Ja	M	84	Nej	K	57	Ja	M	83	Nej	K	59
Ja	K	60	Nej	K	85	Nej	K	78	Ja	K	71
Nej	K	61	Nej	K	80	Ja	K	63	Ja	M	61
Ja	M	98	Ja	M	61	Ja	M	94	Ja	K	66
Ja	M	87	Nej	K	57	Ja	M	84	Nej	K	69
Ja	K	65	Ja	K	51	Ja	M	85	Nej	K	68
Ja	K	67	Nej	M	77	Ja	K	64	Nej	M	72
Nej	M	71	Ja	M	78	Ja	K	54	Nej	M	99
Ja	M	78	Ja	M	59	Nej	K	60	Ja	M	64
Ja	K	59	Nej	K	67	Ja	M	85	Ja	M	66
Nej	K	77	Ja	K	64	Ja	M	69	Ja	M	81
Nej	K	74	Ja	K	81	Ja	K	66	Ja	M	79
Ja	K	62	Ja	K	73	Nej	K	54	Nej	M	79
Nej	M	73	Ja	M	68	Nej	K	54	Nej	M	72
Ja	M	85	Nej	K	81	Nej	K	74	Ja	K	59
Ja	K	60	Nej	M	106	Nej	M	71			
Ja	M	49	Ja	K	74	Ja	K	62			
Nej	M	70	Ja	M	60	Nej	M	82			
Ja	K	55	Nej	K	60	Nej	M	84			
Nej	M	71	Ja	K	56	Nej	M	74			
Nej	M	76	Ja	K	64	Ja	K	74			
Nej	K	61	Ja	K	58	Ja	M	87			
Ja	M	64	Nej	K	58	Ja	K	64			
Ja	K	61	Ja	M	57	Ja	M	98			
Ja	K	59	Ja	M	65	Ja	K	56			