

Reeksamen i Statistik 2. år

Skriftlig prøve (4 timer)

10. august 2007 kl. 9.00–13.00

Eksamenssættet er på 3 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

Opgave 4 indeholdt i spørgsmål (c) en meningsforstyrrende fejl, som er rettet her. I modelspecifikationerne nederst side 2 stod ALDER før KØN, men variansanalyseeskemaet øverst side 3 var, som det ses her, KØN før ALDER.

Opgave 1

Lad X_1 og X_2 være uafhængige stokastiske variable på $\mathbf{N}_0 = \{0, 1, 2, \dots\}$ med samme sandsynlighedsfunktion

$$p(x) = \left(\frac{1}{2}\right)^{x+1}$$

(dvs. begge er geometrisk fordelte med parameter $\frac{1}{2}$).

(a) Udregn $P(X_1 + X_2 \leq 2)$.

(b) Udregn middelværdien af den afledte stokastiske variabel

$$Y = \left(\frac{1}{3}\right)^{X_1}$$

(c) Hvad er den betingede sandsynlighed for hændelsen $X_1 \leq 3$, givet at $X_1 + X_2 = 5$?

Opgave 2

Lad X_1 og X_2 være uafhængige stokastiske variable på $\mathbf{R}_+ =]0, +\infty[$ med samme tæthed $p(x) = xe^{-x}$ (dvs. begge er Γ -fordelte med formparameter 2).

(a) Udregn middelværdi og varians for summen $S = X_1 + X_2$.

(b) Opskriv tætheden for $Y = \sqrt{X_1}$.

(c) Gør rede for at X_1 har fordelingsfunktionen

$$F(x) = 1 - (x + 1)e^{-x}$$

og benyt dette til at udregne $P(X_1 > 2)$.

Opgave 3

Fra en større sundhedsvidenskabelig spørgeskemaundersøgelse, som fandt sted omkring 1976, er udtaget en tilfældig stikprøve bestående af 185 personer. Nedenstående tabel viser deres fordeling efter aldersgruppe og svaret på et spørgsmål om de var rygere:

Ryger?	Ja	Nej	I alt
Aldersgruppe:			
-49	49	21	70
50-59	32	30	62
60-	27	26	53
I alt	108	77	185

Vi betragter den statistiske model som er givet ved, at antal rygere i de tre aldersgrupper er uafhængige, binomialfordelte med aldersgruppernes størrelser som antalsparametre og hver sin sandsynlighedsparameter.

- (a) Estimer, for hver af de tre aldersgrupper, sandsynligheden for at en person fra denne gruppe er ryger, med angivelse af 95% sikkerhedsgrænser.
- (b) Foretag et test for hypotesen om, at sandsynligheden for at være ryger er den samme i de tre aldersgrupper.
- (c) Der er ikke megen forskel på rygefrekvensen i de to ældste grupper. Kan det antages, at sandsynligheden for at være ryger er den samme i de to grupper?

Opgave 4

For 100 personer har man registreret deres alder, køn og bmi (body mass index, vægten i kg divideret med kvadratet på højden i m). Data ses på næste side.

- (a) Estimer parametrene i en simpel regressionsmodel med bmi som responsvariabel og alder som forklarende variabel. (Hjælpestørrelser findes på næste side. Bemærk: Der kræves ikke angivelse af sikkerhedsgrænser i dette spørgsmål).
- (b) Estimer, med angivelse af 95% sikkerhedsgrænser, det forventede bmi for en person på 50 år.
- (c) For den lineære model beskrevet ved modelformlen

$$\text{BMI} = 1 + \text{KØN} + \text{ALDER} + \text{KØN} * \text{ALDER}$$

eller, på matematisk form,

$$E(\text{bmi}) = \alpha + \delta_{\text{køn}} + \beta \times \text{alder} + \gamma_{\text{køn}} \times \text{alder}$$

fås variansanalyseskemaet

Effect	D.F.	S.S.	M.S.	F	P
CONSTANT	1	65666.62508	65666.62508	4111.0571	0.000000
KOEN	1	92.85302	92.85302	6.1133	0.015139
ALDER	1	22.52344	22.52344	1.4903	0.225125
KOEN*ALDER	1	9.75849	9.75849	0.6433	0.424491
RESIDUAL	96	1456.20916	15.16885		
TOTAL	100	67247.96919			

Forklar konklusionerne af de tests, som kan aflæses af dette skema.

Data:

alder	køn	bmi	alder	køn	bmi	alder	køn	bmi	alder	køn	bmi
53	K	27.40	61	M	21.74	55	K	21.91	37	M	31.02
61	M	27.71	48	M	24.44	32	M	19.51	56	M	26.33
67	M	22.66	67	K	23.22	71	K	21.42	53	M	24.42
65	M	26.00	60	K	22.88	53	K	23.67	49	M	30.10
55	M	28.03	74	K	24.72	39	M	27.92	49	M	28.94
48	K	29.17	41	K	23.77	38	M	25.20	39	M	25.21
41	K	22.66	60	M	24.10	78	K	24.53	66	K	23.07
60	K	21.72	43	M	23.75	68	K	31.51	32	K	26.70
43	M	24.98	53	K	24.41	45	M	28.57	71	M	29.63
46	K	24.36	65	M	21.51	36	K	17.66	70	M	28.24
48	K	21.36	38	K	29.80	63	K	32.74	63	K	26.27
65	K	23.24	52	M	35.36	53	M	25.23	47	K	22.04
48	K	22.58	45	K	27.49	65	M	25.31	41	K	33.68
37	M	24.48	73	K	37.84	64	K	34.02	61	K	19.77
48	K	24.15	30	K	20.76	56	K	23.55	53	M	34.57
57	M	36.80	22	M	19.35	89	M	28.48	53	M	32.54
80	M	25.86	58	M	27.73	56	M	28.30	38	M	29.22
55	K	24.60	67	M	24.29	48	K	29.76	40	M	23.90
59	M	29.75	53	K	20.80	62	K	23.81	51	K	23.70
55	K	24.57	78	M	20.72	55	K	26.31	55	M	23.94
45	M	31.85	41	M	25.06	50	K	23.28	61	K	21.03
59	M	28.01	53	K	20.64	28	K	19.15	64	K	22.12
67	M	26.74	50	K	23.43	58	M	24.75	44	M	27.34
47	K	24.65	56	K	25.24	56	M	22.35	29	M	26.59
45	K	24.81	64	K	25.90	56	M	25.29	59	K	20.86

Hjælpestørrelser (med betegnelsen y for bmi, x for alderen):

$$S_x = 5361$$

$$S_y = 2562.55$$

$$SS_x = 302521$$

$$SS_y = 67247.9691$$

$$SP_{xy} = 137934.91$$