

Eksamen i Statistik 2. år

Skriftlig prøve (4 timer)

10. juni 2004 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 , X_2 og X_3 være uafhængige stokastiske variable, ligefordelte på mængden $\{1, 2, 3, 4, 5, 6\}$ (tænk på et kast med tre terninger, eller tre kast med én). Med $X_{(1)}$, $X_{(2)}$ og $X_{(3)}$ betegnes som sædvanligt de ordnede variable (dvs. $X_{(1)}$ er det mindste, $X_{(3)}$ det største og $X_{(2)}$ det midterste af de tre antal øjne).

- (a) Udregn middelværdi og varians for $X_{(1)} + X_{(2)} + X_{(3)}$.
- (b) Opskriv sandsynlighedsfunktionen for $X_{(1)}$ (Vink: Begynd med at udregne størrelserne $P(X_{(1)} \geq x)$).
- (c) Beskriv den betingede fordeling af X_1 , givet $X_{(3)} = 2$

Opgave 2

- (a) Gør rede for at funktionen

$$p(x_1, x_2) = \begin{cases} \frac{9}{4}(x_1 x_2)^2 & \text{for } |x_1| \leq 1 \text{ og } |x_2| \leq 1 \\ 0 & \text{ellers} \end{cases}$$

er sandsynlighedstæthed for en kontinuert fordeling på $[-1, 1] \times [-1, 1]$.

I det følgende betegner (X_1, X_2) en stokastisk variabel med denne fordeling.

- (b) Udregn middelværdi og varians for X_1 og X_2 samt kovariansen mellem X_1 og X_2 .
- (c) Opskriv tætheden for $(Y_1, Y_2) = (2X_1, X_2^2)$

Opgave 3

Et markedsanalyseinstitut har foretaget samme spørgeskemaundersøgelse i to versioner: En netbaseret (formodentlig med henvendelse via e-mail) og en mere traditionel telefoninterview-version. For begge versioner henvendte man sig til et antal respondenter (formodentlig 1000, det præcise antal kendes ikke). 352 besvarede den netbaserede version, 301 telefonversionen. Fordelingen af aldrene for de respondenter der besvarede, så — med passende gruppering — sådan ud:

Alder:	–30	31–40	41–50	51–60	61–
Version:					
Net	125	100	61	46	20
Tlf	42	62	47	61	89

Vi betragter modellen hvor rækkesummerne i denne tabel er faste, således at tabellens rækker kan opfattes som observationer af to uafhængige, polynomialfordelte variable af orden 5, med antalsparametre henholdsvis 352 og 301 og hvert sit sæt af sandsynlighedsparametre.

- (a) Opstil en tabel der viser den procentvise fordeling på aldersgrupper inden for de to respondentgrupper (altså en tabel hvor rækkesummerne er 100), og benyt denne tabel til at beskrive forskellene mellem de to respondentgrupper.
- (b) Udfør et egentligt test for om de to polynomialfordelinger har samme sandsynlighedsparametre.
- (c) Estimer, med angivelse af 95% sikkerhedsgrænser, sandsynligheden for at en person der responderer på net-spørgeskemaet er under 31.

Opgave 4

Fra et datasæt bestående af driftsoplysninger for alle norske busselskaber i 1991 har man udtrukket de tal, som er gengivet på næste side. For hvert selskab over en vis størrelse (selskaber som har kørt under 2 mio km er udeladt) er følgende tre oplysninger registreret:

y : De samlede udgifter til brændstof (i mio norske kroner).

x : Antal km kørt (i mio km).

z : Det gennemsnitlige passagerantal i busserne.

(a) Estimer parametrene i en simpel regressionsmodel med x som forklarende variabel og y som afhængig variabel, altså modellen givet ved at y_i 'erne er uafhængige, normalfordelte med samme varians σ^2 og middelværdier

$$E y_i = \alpha + \beta x_i$$

For parametrene α og β ønskes angivelse af 95% sikkerhedsgrænser.

Til udregningerne kan benyttes følgende mellemregningsstørrelser:

$$n = 42 \quad S_x = 164.10 \quad S_y = 161.67$$

$$SS_x = 759.49 \quad SS_y = 786.23 \quad SP_{xy} = 763.46$$

(b) I modellen ovenfor, diskuter fortolkningen af det specialtilfælde hvor regressionslinien går gennem (0,0), og foretag et test for denne hypotese.

(c) Hvis modellen fra spørgsmål (a) udvides til

$$E y_i = \alpha + \beta x_i + \gamma z_i$$

ved tilføjelse af z som forklarende variabel, fås estimaterne

$$\hat{\alpha} = -1.068 \quad \hat{\beta} = 0.970 \quad \hat{\gamma} = 0.117$$

og residualkvadratsummen falder til 11.44. Foretag F-testet for hypotesen $\gamma = 0$. Prøv at fortolke resultatet, og diskuter herunder om $\hat{\gamma}$ har det fortegn man ville forvente.

Y	X	Z
3.178908	4.047293	8.316669
2.825300	2.930500	5.569016
2.366400	2.790000	5.871506
2.927500	3.547700	11.108437
3.984779	3.595000	9.444495
5.422800	4.809100	12.245742
4.451000	6.370000	7.884930
2.090000	2.313000	4.755729
2.737000	3.282000	8.309567
5.155300	4.900500	12.852934
4.629000	4.282000	9.804997
6.089000	6.996000	7.883505
3.281500	3.249491	9.403378
2.951941	2.445491	12.030257
7.927315	6.229555	15.014139
3.375000	3.350000	10.208956
4.795900	4.350000	10.336079
6.893500	6.656800	11.099292
4.368700	4.890600	6.805709
1.834892	2.204126	6.242384
2.237097	2.842394	5.162549
2.671000	3.287000	7.780652
2.365000	2.485145	4.446421
3.737700	4.175000	8.528934
2.340000	2.070000	9.661836
2.162574	2.670625	4.787930
9.775000	8.863117	12.639571
2.580689	2.864900	8.415651
3.663000	3.727100	7.016179
2.145944	2.004874	9.423900
2.356785	2.355100	8.151247
6.842000	5.450000	14.139468
6.468749	5.294173	14.041098
4.910609	5.259778	13.141239
3.040197	3.914453	5.371089
1.888600	2.113445	13.112950
2.781155	3.186445	8.486070
8.880985	8.054750	25.868536
2.022741	2.078230	11.709117
3.713200	3.373300	11.757271
2.297512	2.492862	8.867735
1.502200	2.296133	7.519120