

Reeksamen i Statistik 2. år

Skriftlig prøve (4 timer)

30. juli 2004 kl. 9.00–13.00

Eksamenssættet er på 3 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

Opgave 1

Definer $p : \mathbf{R} \rightarrow \mathbf{R}$ ved

$$p(x) = \begin{cases} 0.4 & \text{for } x = -1 \\ 0.6 & \text{for } x = 1 \\ 0 & \text{ellers.} \end{cases}$$

(a) Gør rede for at p er en sandsynlighedsfunktion. Beskriv den tilhørende sandsynlighedsfordeling P , og udregn middelværdi og varians for en stokastisk variabel X med denne fordeling.

Lad X_1 , X_2 og X_3 være uafhængige stokastiske variable med den ovenfor beskrevne fordeling.

(b) Udregn middelværdi og varians for $X_1 + X_2 + X_3$. Udregn endvidere

$$E(2^{X_1+X_2+X_3}).$$

(c) Beskriv den betingede fordeling af X_1 , givet $X_1 + X_2 + X_3 = 3$.

Opgave 2

(a) Gør rede for at den ved

$$p(x, y) = x + y$$

definerede funktion

$$p : [0, 1] \times [0, 1] \rightarrow \mathbf{R}$$

er en sandsynlighedstæthed for en kontinuert fordeling på enhedskvadratet $[0, 1] \times [0, 1]$.

I det følgende betegner (X, Y) en stokastisk variabel med denne fordeling.

(b) Beskriv den marginale fordeling af X .

(c) Udregn $\text{var}(X)$ og $\text{cov}(X, Y)$.

Opgave 3

Nedenstående tabel viser antal dræbte ved trafikulykker i Danmark i 2002, fordelt på amt (dvs. det amt hvor ulykken fandt sted). Desuden er befolkningens størrelse i amterne oplyst.

| AMT | | BEF | ANTAL |
|-------|------------|---------|-------|
| 1 | Kbh.+Frb. | 591853 | 20 |
| 2 | Kbh. Amt | 617336 | 24 |
| 3 | Fr.borg | 370555 | 39 |
| 4 | Roskilde | 234820 | 24 |
| 5 | Vestsjæll. | 298731 | 34 |
| 6 | Storstrøm | 260498 | 37 |
| 7 | Bornholm | 44197 | 3 |
| 8 | Fyn | 472504 | 39 |
| 9 | Sønderjyl. | 253166 | 25 |
| 10 | Ribe | 224444 | 20 |
| 11 | Vejle | 351328 | 32 |
| 12 | Ringkøb. | 274385 | 41 |
| 13 | Århus | 644666 | 56 |
| 14 | Viborg | 234323 | 20 |
| 15 | Nordjyl. | 495548 | 48 |
| ----- | | 5368354 | 462 |

Lad y_i og x_i betegne henholdsvis antal trafikdræbte og befolkningens størrelse i amt i ($= 1, \dots, 15$). Vi betragter “proportionalitetsmodellen”, hvor y_i ’erne antages at være observationer af uafhængige, Poissonfordelte variable med parametre af formen βx_i .

(a) Estimer β .

(b) Foretag et test for denne model imod den fulde model, hvor y_i ’erne har hver sin frit varierende parameter (bemærk: Man er velkommen til at udnytte tabellen under spørgsmål (c) i udregningerne).

(c) Nedenfor ses en tabel med observationer, fittede værdier og normerede residualer i proportionalitetsmodellen. Kommenter denne, med særligt henblik på at udpege de amter som i højest grad afviger fra modellens proportionalitetshypotese.

| AMT | Y | FIT | NRES |
|-----|----|------|-------|
| 1 | 20 | 50.9 | -4.33 |
| 2 | 24 | 53.1 | -4.00 |
| 3 | 39 | 31.9 | 1.26 |
| 4 | 24 | 20.2 | 0.84 |
| 5 | 34 | 25.7 | 1.64 |
| 6 | 37 | 22.4 | 3.08 |
| 7 | 3 | 3.8 | -0.41 |
| 8 | 39 | 40.7 | -0.26 |
| 9 | 25 | 21.8 | 0.69 |
| 10 | 20 | 19.3 | 0.16 |
| 11 | 32 | 30.2 | 0.32 |
| 12 | 41 | 23.6 | 3.58 |
| 13 | 56 | 55.5 | 0.07 |
| 14 | 20 | 20.2 | -0.04 |
| 15 | 48 | 42.6 | 0.82 |

Opgave 4

Data til denne opgave stammer oprindelig fra et annoncetillæg, bestilt af Ejendomsmæglerkæden Home til indlæg i Berlingske Tidende 19. september 1999. Herfra er følgende oplysninger udtaget fra de 12 annoncer, der vedrørte villaer beliggende i Lyngby:

| Pris | Areal | Sovevær. |
|------|-------|----------|
| 3176 | 228 | 4 |
| 1850 | 114 | 3 |
| 2159 | 145 | 3 |
| 1500 | 91 | 2 |
| 2032 | 106 | 3 |
| 3101 | 170 | 4 |
| 2700 | 198 | 3 |
| 1550 | 87 | 3 |
| 1800 | 126 | 3 |
| 3354 | 166 | 4 |
| 1575 | 102 | 3 |
| 1673 | 106 | 3 |

Kilde: Data til hjemmeopgave på HA (Jens Overø).

Pris er kontantprisen i 1000 kr., **Areal** er boligarealet i m² og **Sovevær.** er antal soveværelser (2, 3 eller 4). I det følgende betegner y_i prisen og x_i arealet for bolig nr. i .

(a) I første omgang betragtes en simpel regressionsanalyse med prisen som respons og boligareal som forklarende variabel, altså modellen der antager y_i 'erne uafhængige, normalfordelte med samme varians σ^2 og middelværdier

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

Estimer parametrene i denne model. For hældning og afskæring ønskes angivelse af 95% sikkerhedsgrænser. Følgende regnestørrelser kan benyttes:

$$228 + \dots + 106 = 1639$$

$$3176 + \dots + 1673 = 26470$$

$$228^2 + \dots + 106^2 = 246267$$

$$3176^2 + \dots + 1673^2 = 63627352$$

$$228 \times 3176 + \dots + 106 \times 1673 = 3918147$$

(b) Gør rede for at hypotesen $\alpha = 0$ kan godkendes, og diskuter fortolkningen af denne hypotese.

(c) Hvis antal soveværelser tilføjes til modellen som en faktor på 3 niveauer, dvs. hvis modellen udvides til

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

hvor $v = 2, 3$ eller 4 betegner antal soveværelser, falder residualkvadratsummen til 397019. Foretag et test for den simple regressionsmodel imod denne udvidede model, og forklar hvad konklusionen af dette test betyder.