

Eksamen i Statistik 2. år

Skriftlig prøve (4 timer)

14. juni 2001

Opgave 1

En reel funktion på $E = \{1, 2, \dots, 10\}$ defineres ved

$$p(x) = \frac{x}{55}.$$

(a) Gør rede for at p er en sandsynlighedsfunktion.

Lad X være en stokastisk variabel på E med denne sandsynlighedsfunktion.

(b) Udregn $E(X)$ og $E(1/X)$.

(c) Lad X_1 og X_2 være uafhængige med samme fordeling som X ovenfor. Hvad er $\text{cov}(X_1 - X_2, X_1 + X_2)$?

Opgave 2

Lad X være ligefordelt på intervallet $[-1, 1]$.

(a) Opskriv tætheden for X , og udregn $E(X)$ og $\text{var}(X)$.

En ny stokastisk variabel defineres ved $Y = e^X$.

(b) Udregn $E(Y)$.

(c) Opskriv tætheden for Y .

Opgave 3

Data til denne opgave er 10–15 år gamle og stammer fra Østerbrounder-søgelsen, som var en meget stor helbredsundersøgelse, foretaget primært i form af en spørgeskemaundersøgelse, hvor tilfældige udsnit af forskellige aldersgrupper i København blev spurgt om en lang række ting vedrørende helbred og livsstil. Blandt andet blev man spurgt om man røg. Svarene på dette spørgsmål, opdelt på passende aldersgrupper, er optalt i følgende tabel:

	Ryger	Ikke-ryger	I alt
Alder højst 40	1517	889	2406
Alder 41–50	2309	974	3283
Alder 51–60	3178	1708	4886
Alder 61–70	1702	1278	2980
Alder mindst 71	287	338	625
I alt	8993	5187	14180

Det samlede antal personer i de enkelte aldersgrupper (rækkesummerne i tabellen) bør her opfattes som faste tal, idet man i undersøgelsen har udtaget nogle ganske bestemte antal indenfor hver aldersgruppe. Disse tal har altså ikke direkte noget at gøre med aldersfordelingen i København.

- (a) Opstil en statistisk model til analyse af disse tal, og angiv estimater for de ukendte parametre med approksimative 95% sikkerhedsgrænser.
- (b) Gør rede for at der er signifikant forskel på rygevanerne i de fem aldersgrupper.
- (c) Beskriv hvordan hyppigheden af rygning afhænger af alderen, og diskuter rimeligheden af at reducere modellen ved at slå nogle aldersgrupper sammen.

Opgave 4

På en avisredaktion har man måned for måned over en periode på 2 år registreret samhørende værdier af

x : Tid anvendt på telefonisk markedsføring (enhed: 100 timer), og

y : Antal nytegnede abonnemeter (enhed: 100 stk.).

x	y	x	y
12.24	53.57	14.35	58.36
14.58	61.77	13.49	52.01
10.06	47.95	9.65	37.75
13.95	56.92	9.85	35.92
11.31	43.12	11.17	45.66
9.21	34.21	8.40	29.74
7.04	26.24	14.12	56.73
11.54	40.87	9.40	35.54
11.68	49.34	10.90	43.99
8.03	25.46	14.98	61.43
8.30	35.91	12.40	48.27
9.81	42.71	10.55	54.18

- (a) Estimer med angivelse af 95% sikkerhedsgrænser afskæring og hældning i en simpel regressionsmodel med y som den afhængige variabel

og x som den forklarende variabel, og angiv også estimater for varians og standardafvigelse i denne model. Til udregningerne kan følgende mellemregningsstørrelser benyttes:

$$12.24 + 14.58 + \dots + 10.55 = 267.01$$

$$12.24^2 + 14.58^2 + \dots + 10.55^2 = 3089.09$$

$$53.57 + 61.77 + \dots + 54.18 = 1077.65$$

$$53.57^2 + 61.77^2 + \dots + 54.18^2 = 51060.4$$

$$12.24 \times 53.57 + 14.58 \times 61.77 + \dots + 10.55 \times 54.18 = 12511.6$$

(b) Kontroller model og beregninger ved hjælp af et x - y plot med den estimerede regressionslinie indtegnet.

(c) Diskuter fortolkning og relevans af en model med afskæring 0, og foretag et test for den tilsvarende modelreduktion.