

Reeksamen i Statistik 2. år

Skriftlig prøve (4 timer)

29. juli 2005 kl. 9.00–13.00

Eksamenssættet er på 3 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

Bemærkning. I det originale eksamenssæt var der en fejl, idet variabelbetegnelserne "x" og "y" var byttet om i indledningen til opgave 4. Denne meningsforstyrende fejl er rettet her.

Opgave 1

Fra en kasse, som indeholder 10 røde, 10 hvide og 10 blå kugler, udtages tilfældigt og uden tilbagelægning en stikprøve på 4 kugler.

- (a) Hvad er sandsynligheden for at ingen af de udtrukne kugler er røde?
- (b) Angiv middelværdi og varians for antallet af røde kugler i stikprøven.
- (c) Hvad er den betingede sandsynlighed for at der er præcis 2 blå kugler i stikprøven, givet at der ikke er nogle røde?

Opgave 2

- (a) Gør rede for at den ved

$$p(x, y) = \frac{1}{2}(x + y)e^{-(x+y)}$$

definerede funktion på mængden $]0, +\infty[\times]0, +\infty[$ (første kvadrant) er en sandsynlighedstæthed for en kontinuert fordeling på denne mængde. I det følgende betegner (X, Y) en stokastisk variabel med denne fordeling.

- (b) Opskriv tætheden for den marginale fordeling af Y , og udregn middelværdi og varians for Y .

- (c) Nye stokastiske variable $S \in]0, +\infty[$ og $R \in]0, 1[$ defineres ved

$$S = X + Y$$
$$R = \frac{X}{X + Y}$$

Beskriv fordelingen af (S, R) .

Opgave 3

Nedenstående tabel viser antal beståede (karakter ≥ 6) og ikke beståede (karakter ≤ 5) eksaminationer på studiet HA-almen ved Handelshøjskolen i København i årene 2001–2004, fordelt på eksamensform:

Eksamensform:	Bestået	Ikke bestået	I alt
SkrEks	14752	5564	20316
TagHjem	15974	531	16505
SkrMndlg	13428	373	13801
SkrMndli	2111	139	2250

hvor de fire eksamensformer er

SkrEks: Skriftlig prøve i eksamenslokale
TagHjem: Skriftlig tag-hjem prøve
SkrMndlg: Skriftlig med mundtlig gruppeeksamen
SkrMndli: Skriftlig med mundtlig individuel eksamen

Vi betragter modellen hvor tallene i første søjle (antal beståede under hver af de fire eksamensformer) opfattes som observationer af uafhængige binomialfordelte variable med hver sin sandsynlighedsparameter og antalsparametre som angivet i sidste søjle.

- (a) Estimer de fire sandsynlighedsparametre med angivelse af approksimative 95% konfidensgrænser.
- (b) Foretag et test for hypotesen om at sandsynligheden for at bestå ikke afhænger af eksamensformen.
- (c) Det kunne se ud som om sandsynligheden for at bestå er nogenlunde den samme under de to eksamensformer TagHjem og SkrMndlg. Undersøg dette ved hjælp af et passende test.

Opgave 4

Over en periode på ca. $2\frac{1}{2}$ år, fra juni 1994 til november 1996, indsamlede markedsanalyseinstituttet Millward Brown Denmark data vedrørende salg af kaffe i Danmark. Fra dette datasæt har vi udtrukket følgende tre oplysninger vedrørende 31 uger med start januar 1995:

y : Ugens samlede salg af Gevalia (enhed: 10.000 stk. pakker á 500g).

x : Ugens gennemsnitlige pris på Gevalia (kr. pr. pakke).

z : Ugens gennemsnitlige pris på det konkurrerende mærke Merrild (kr. pr. pakke).

Tallene er gengivet på næste side.

UGE	Y	X	Z
1	14.87	31.10	31.40
2	16.74	32.50	31.90
3	20.65	30.80	32.60
4	20.29	30.50	32.10
5	17.84	31.40	32.60
6	16.54	30.80	31.70
7	18.68	30.90	32.90
8	11.99	32.30	31.70
9	21.09	31.20	33.40
10	17.89	31.30	32.50
11	13.90	31.80	32.80
12	18.47	30.30	31.60
13	39.70	27.20	30.40
14	24.15	29.00	30.60
15	13.36	31.20	31.20
16	18.93	29.10	31.00
17	12.75	31.60	30.10
18	27.93	29.20	28.20
19	12.71	30.70	28.70
20	10.87	31.90	26.40
21	26.70	28.20	27.30
22	23.66	28.70	28.40
23	14.13	31.00	28.00
24	26.22	29.00	27.40
25	18.48	30.70	28.00
26	28.79	27.40	29.20
27	36.66	26.80	26.90
28	20.47	28.80	27.70
29	27.67	27.60	29.20
30	7.84	30.70	28.90
31	23.31	27.10	29.30

(a) I første omgang ser vi på en simpel regressionsmodel med salget af Gevalia (y) som respons og prisen på Gevalia (x) som forklarende variabel. Estimer parametrene i denne model. For hældningens vedkommende ønskes angivelse af 95% sikkerhedsgrænser. Følgende regnestørrelser kan benyttes:

$$14.87 + \dots + 23.31 = 623.28$$

$$31.10 + \dots + 27.10 = 930.80$$

$$14.87^2 + \dots + 23.31^2 = 14105.31$$

$$31.10^2 + \dots + 27.10^2 = 28030.42$$

$$14.87 \times 31.10 + \dots + 23.31 \times 27.10 = 18413.06$$

(b) Kontroller modellen og beregningerne ved hjælp af en tegning, hvor punkterne (x, y) er indtegnet sammen med den estimerede regressionslinje. Foretag et test for hypotesen om at salget ikke afhænger af prisen.

(c) I en multipel regressionsmodel hvor både prisen på Gevalia (x) og prisen på Merrild (z) er med som forklarende variable (stadig med y som den afhængige variable) fås et estimat på 0.730 for koefficienten til z , med en estimeret standardafvigelse på 0.3911. Hvad kan man slutte af det?