

Eksamen i Statistik 2. år

Skriftlig prøve (4 timer)

16. maj 2007 kl. 9.00–13.00

Eksamenssættet er på 3 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

Opgave 1

Lad X_1 og X_2 være stokastiske variable på $\{0, 1, 2, 3\}$ som er uafhængige, identisk fordelte med sandsynlighedsfunktion

$$p(x) = \begin{cases} \frac{1}{8} & \text{for } x = 0 \text{ eller } 3, \\ \frac{3}{8} & \text{for } x = 1 \text{ eller } 2. \end{cases}$$

(a) Udregn middelværdi og varians for X_1 .

En ny stokastisk variabel på $\{0, 1, 2, 3, 4, 5, 6\}$ defineres ved $S = X_1 + X_2$.

(b) Udregn middelværdi og varians for S , samt $P(S = 2)$.

(c) Udregn den betingede sandsynlighed $P(X_1 \leq 1 \mid S = 3)$.

Opgave 2

Lad X_1 og X_2 være uafhængige, ligefordelte på enhedsintervallet.

(a) Udregn middelværdi og varians for $2X_1 - X_2$.

(b) Udregn $P(|X_1 - \frac{1}{2}| \leq \frac{1}{3} \text{ og } |X_2 - \frac{1}{2}| \leq \frac{1}{3})$.

(c) Beskriv fordelingen af $|X_1 - \frac{1}{2}|$.

Opgave 3

I 4×4 antalstabellen nedenfor er 285 studerende ved et amerikansk universitet grupperet efter deres eksamensresultater (rækkefaktoren, hvor -- betyder "meget dårlige" og ++ betyder "meget gode") og omfanget af deres erhvervsarbejde ved siden af studiet (søjlefaktoren, angivet i timer pr. uge).

	0	1–5	6–10	>10
--	10	8	12	21
-	31	28	21	34
+	18	20	12	8
++	32	14	11	5

(a) Opstil en tabel, hvor antallene er omregnet til procent af søjlesummerne, således at man for hver af de fire erhvervsarbejdsgrupper kan aflæse fordelingen på de fire eksamensresultatgrupper i procent. Hvilke tendenser ser man i tabellen, som kunne tale imod hypotesen om uafhængighed? Procenterne i denne tabel kan opfattes som estimater for sandsynlighedsparametre i en polynomialfordelingsmodel. Præciser denne model.

(b) Foretag et test for uafhængighed i antalstabellen. Opskriv en tabel, der indeholder de normerede residualer, og kommenter de afvigelser man evt. ser fra uafhængighedshypotesen.

(c) En rimelig hypotese kunne være, at erhvervsarbejde i moderat omfang ikke indvirker negativt på eksamensresultaterne. Hvordan ser det for eksempel ud, hvis man ser bort fra studerende med mere end 10 timers erhvervsarbejde?

Opgave 4

Tabellen på næste side er en lille del af et stort datasæt vedrørende forbrugeres præferencer for forskellige typer af svinekød, indsamlet ved et smagsprøvningseksperiment udført af Slagteriernes Forskningsinstitut. Tallene i tabellen vedrører kun almindelige svinekoteletter (adskillige andre kødtyper indgik i det oprindelige eksperiment), og kun data for 20–25-årige forbrugere er medtaget. Af dem var der 36. For hver af disse er oplyst

Y: Smagsvurdering på en kontinuert skala fra 0 til 15,

BOPAEL: Respondentens bopæl, 1 (Roskilde) eller 2 (Holsterbro),

KON: Respondentens køn, 1 (mand) eller 2 (kvinde).

Smagsvurderingen er fremkommet ved, at respondenterne har sat et mærke på en 15 cm lang skala, hvor venstre endepunkt (0) betyder “kan slet ikke lide ... ” medens højre endepunkt (15) betyder “kan meget godt lide ... ”.

Vi betragter i første omgang den ensidede variansanalysemodel, hvor de 36 smagsbedømmelser y_i opfattes som observationer af uafhængige normalfordelte variable med samme varians og en middelværdi, der kun afhænger af forbrugers køn.

Y	BOPAEL	KON	Y	BOPAEL	KON
4.0	1	2	8.0	1	1
7.6	1	2	8.4	1	1
10.6	1	1	11.7	1	1
5.8	1	1	3.5	1	1
11.6	1	1	11.1	1	1
9.0	1	2	10.6	2	1
6.2	1	1	11.5	2	1
8.7	1	1	11.9	2	1
11.8	1	1	8.4	2	2
7.1	1	2	4.8	2	1
14.2	1	1	0.5	2	2
2.5	1	2	3.4	2	2
7.5	1	1	1.8	2	2
14.7	1	1	5.3	2	2
11.0	1	1	5.8	2	2
15.0	1	2	14.3	2	1
12.6	1	1	10.7	2	1
9.9	1	1	7.4	1	1

Mellemregningstørrelser:

	antal	sum	kvadratsum
Observationer:			
Mænd	24	238.5	2575.59
Kvinder	12	70.4	583.76
Alle	36	308.9	3159.35

- (a) Undersøg, om modellens forudsætning om varianshomogenitet er opfyldt.
- (b) Estimer parametrene i den ensidede variansanalysemodel. Estimer desuden forskellen mellem middelværdierne i de to grupper, med angivelse af 95% konfidensgrænser.
- (c) Ved estimation af den lineære model givet ved modelformlen

$$Y=1+KON+BOPAEL+KON*BOPAEL$$

fås følgende variansanalysekema:

Effect	D.F.	S.S.	M.S.	F	P
CONSTANT	1	2650.533617	2650.533617	182.3225	0.000000
KON	1	132.573474	132.573474	11.9803	0.001469
BOPAEL	1	4.524083	4.524083	0.4016	0.530615
BOPAEL*KON	1	32.682722	32.682722	3.0848	0.088596
RESIDUAL	32	339.036110	10.594878		
TOTAL	36	3159.350006			

Hvad kan man konkludere af det?