

# Reeksamen i Statistik 2. år

Skriftlig prøve (4 timer)

24. august 2010 kl. 9.00–13.00

Eksamenssættet er på 3 sider.

Alle skriftlige hjælpemidler samt lommeregner er tilladt.

Vægtfordeling: Opgaverne vægtes ens.

## Opgave 1

Lad  $X_1, \dots, X_{10}$  være uafhængige, identisk fordelte 0–1 variable med fordeling givet ved  $P(X_i = 1) = p$  og  $P(X_i = 0) = 1 - p$  for et eller andet tal  $p$  mellem 0 og 1.

- (a) Opskriv sandsynlighedsfunktionen for  $Y = X_1 - X_2$ .
- (b) Beskriv fordelingen af

$$S = X_1X_2 + X_3X_4 + X_5X_6 + X_7X_8 + X_9X_{10}$$

- (c) Beskriv den betingede fordeling af  $X_1 + \dots + X_5$ , givet at  $X_1 + \dots + X_{10} = 7$ .

## Opgave 2

Lad  $U_1$  og  $U_2$  være uafhængige, normeret normalfordelte stokastiske variable.

- (a) Hvad er middelværdi og varians for  $2U_1 - U_2$ ? Og hvad er fordelingen af denne afledte variable?
- (b) Udregn  $P(4U_1 - 3 \in [0, 1])$ .
- (c) Sæt  $X = 2U_1 - U_2$  og  $Y = 2U_1 + U_2$ . Beregn korrelationen mellem  $X$  og  $Y$ .

### Opgave 3

Ved en spørgeskemaundersøgelse i 1991 blev 3095 personer spurgt om de var medlem af et politisk parti. Desuden spurgte man om deres alder. Resultatet (med passende gruppering i aldersintervaller) ses her:

Partimedlem	Ja	Nej	SUM
Ald.gr.	-----	-----	-----
16-24	21	469	490
25-34	21	561	582
35-54	108	959	1067
55-64	42	325	367
65-	69	520	589
	-----	-----	-----
SUM	261	2834	3095

Til analyse af disse tal betragter vi modellen hvor de fem rækkesummer er givne, og antal personer som er partimedlemmer i hver af aldersgrupperne er uafhængige, binomialfordelte med hver sin sandsynlighedsparameter og den tilsvarende rækkesum som antalsparameter.

- Estimer, med angivelse af 95% konfidensgrænser, modellens fem sandsynlighedsparametre.
- Undersøg ved et statistisk test, om der er forskel på aldersgruppernes tilbøjelighed til at være medlemmer af et parti. Hvis der er tydelig forskel, beskriv hvad denne forskel går ud på, og prøv på at forklare den.
- Ser man på de to yngste grupper er der meget lille forskel, og man fristes til at slå disse to aldersgrupper sammen. Foretag et statistisk test som kan underbygge (eller afkræfte) dette.

## Opgave 4

På 20 lige store og tilnærmelsesvis ens jordstykker er dyrket 4 sorter af byg, idet hver sort er dyrket på 5 jordstykker. Tallene i tabellen er udbytter i kg. på de enkelte jordstykker.

Sort	Maja	Freja	Carlsberg-II	Binder	
	452	335	435	356	
	549	464	504	411	
	413	352	445	391	
	504	406	496	448	
	552	486	518	401	
	-----				
Sum	2470	2043	2398	2007	8918
Kvadratsum	1234994	852457	1155606	810043	4053100

Bemærk at summer og kvadratsummer indenfor grupper står nederst i tabellen, og sum og kvadratsum for alle 20 observationer finder man nederst til højre.

- Vi betragter den ensidede variansanalysemodel, hvor de 20 udbytter opfattes som normalfordelte med samme varians og en middelværdi der afhænger af hvilken sort der er dyrket. Udfør Bartlett's test, og estimer parametrene i denne model.
- Udfør testet for homogenitet, altså testet for om de fire middelværdier kan antages at være ens.
- Erfaringsmæssigt giver sorterne Freja og Binder (2 og 4) omtrent samme udbytte. Foretag en parvis sammenligning af de to middelværdier.