

Reksamen august 2003, Opgave 3

(a)

Estimaterne for de fire sandsynlighedsparemetre bliver

$$\text{Afdeling 1: } 7/52 = 0.1346$$

$$\text{Afdeling 2: } 36/142 = 0.2535$$

$$\text{Afdeling 3: } 4/20 = 0.2000$$

$$\text{Afdeling 4: } 22/56 = 0.3929$$

Approksimative 95% sikkerhedsgrænser fås f.eks. for afdeling 1's vedkommende ved hjælp af formlen

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.1346 \pm 1.96 \sqrt{\frac{0.1346(1-0.1346)}{52}} = 0.1346 \pm 0.0928$$

svarende til sikkerhedsintervallet $[0.0418, 0.2274]$. Tilsvarende fås for alle afdelingerne nedenstående estimater med usikkerhed og 95% sikkerhedsgrænser:

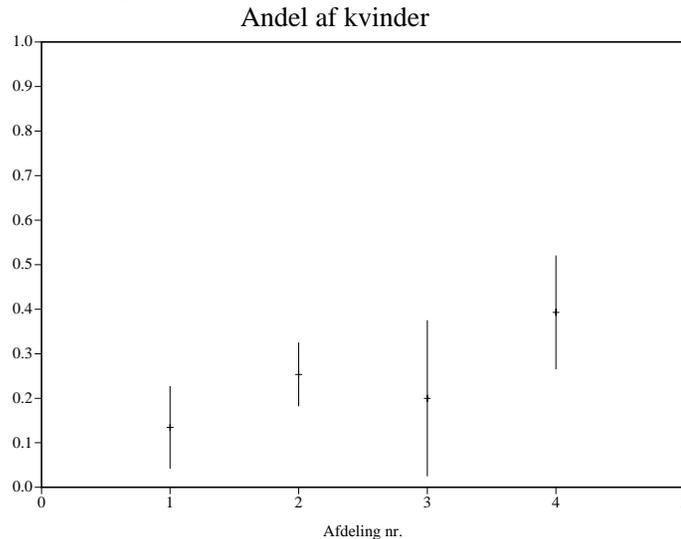
$$\text{Afdeling 1: } p_1 = 0.1346 \pm 0.0928, [0.0418, 0.2274]$$

$$\text{Afdeling 2: } p_2 = 0.2535 \pm 0.0716, [0.1819, 0.3251]$$

$$\text{Afdeling 3: } p_3 = 0.2000 \pm 0.1753, [0.0247, 0.3753]$$

$$\text{Afdeling 4: } p_4 = 0.3929 \pm 0.1279, [0.2650, 0.5208]$$

Grafisk ser det sådan ud:



Den mest markante forskel ser vi mellem afdeling 1 (udvikling), hvor der er ca. 13% kvinder, og afdeling 4 (administration og bogholderi) hvor andelen af kvinder er oppe på 39%. De to andre afdelinger ligger sådan cirka midt imellem. Da konfidensintervallerne for p_1 og p_4 faktisk er disjunkte, må vi forvente at finde en signifikant forskel mellem andelen af kvinder på de fire afdelinger.

(b)

Testet for homogenitet er ækvivalent med det sædvanlige test for uafhængighed i den givne 4×2 -tabel. Vi får $-2 \log q$ -størrelsen

$$2 \times (45 \log 45 + \dots + 22 \log 22 - 201 \log 201 - \dots - 56 \log 56 + 270 \log 270)$$

$= 9.9761$ (Pearsons teststørrelse bliver 9.8745). Denne størrelse skal vurderes i en χ^2 -fordeling med 3 frihedsgrader, hvor 95% fraktilen er 7.815 og 99% fraktilen er 11.345. Der er således en svagt signifikant forskel mellem kønsproportionerne på de fire afdelinger.

(c)

Hvis vi ser bort fra udviklingsafdelingen, og altså betragter 3×2 -tabellen bestående af de tre sidste rækker, er forskellen knapt så tydelig på tegningen. Testet for uafhængighed i denne model giver

$$-2 \log q = 4.4706$$

(Pearsons teststørrelse bliver 4.6020), som skal vurderes i en χ^2 -fordeling med 2 frihedsgrader. Her er 95% fraktilen 5.991, så vi kan godkende denne hypotese om delvis homogenitet. Det vil sige, at vi ikke på grundlag af datamaterialet kan afvise den formodning at de tre sidste afdelinger "i virkeligheden" har samme kønsproportion (der er bare ved et tilfælde havnet en masse kvinder i afdeling 4). Dette er dog ikke nødvendigvis en udtømmende beskrivelse, hvis den ikke understøttes af andre argumenter for at det skulle forholde sig sådan. Vi kunne formodentlig også få godkendt uafhængighed i tabellen bestående af de tre første rækker, svarende til at det er de relativt mange kvinder i administration og bogholderi som er "skyld i" forskellen mellem afdelingerne.