Frekvenssiaineistojen analyysi
Harjoitukset 8.11.2001

8.1. Tarkastellaan Taulukon 6.2 aineistoa, joka käsittelee syöpähoidon vaikutuksia (luentomoniste). Selitetään todennäköisyyttä, että tietty hoito tuo helpotusta potilaalle. Aineistoon sovitettu regressiomalli on

$$\text{logit}[\hat{\pi}(x)] = -3.777 + 0.145x, \qquad (6.8.1)$$

missä x on solujen kasvuaktiteettia mittaava LI-muuttuja (labeling index).

(a) Millä x:n arvolla ovat mahdollisuudet tasan saada hoidosta apua (ts. vastemuuttujan kummankin arvon todennäköisyydet ovat samat)?

(b) Millä x:n arvolla onnistumisen todennäköisyys kasvaa eniten?

(c) Millä x:n arvolla hoidosta on apua 90%:n todennäköisyydellä?

8.2. Tarkastellaan Taulukon 6.2 syöpäaineistoa ja siitä estimoitua mallia (6.8.1).

(a) Mitä on mahdollisuuksien suhde $\hat{\omega}(x)$? Montako prosenttia se kasvaa, kun x kasvaa kahdella yksiköllä?

(b) Lasketaan vedonlyöntisuhde vertaamalla suhteellisia mahdollisuuksia pisteissä x_1 ja x_2. Esitä vedonlyöntisuhde erotuksen $x_1 - x_2 = v$ funktiona.

8.3. Tutkitaan Taulukon 6.2 aineiston sopivuutta malliin (6.8.1).

(a) Laske empiiriset logitit ja vertaa niitä mallista estimoitujen logittien arvoihin.

(b) Yhdistä kaksi peräkkäistä LI-arvoa yhdeksi luokaksi ja käytä uusina LI-arvoina peräkkäisten LI-arvojen keskiarvoa. Estimoi logitmalli tästä aineistosta.

(c) Yhdistä luokat $1-5, 6-8, 9-11$ ja $12-14$ ja menettele muutoin niinkuin edellisessä kohdassa. Vertaile näillä kolmella eri luokituksella saatuja malleja ja niiden deviansseja toisiinsa.

8.4. Selitetään logistisella regressiomallilla todennäköisyyttä, että rapunaaraalla on seuralaisia, kun selittäjänä on ravun väri ($Y = 1$, kun seuralaisia; $Y = 0$, kun ei seuralaisia). Väri (`color`) on luokittelumuuttuja (1=vaalea, 2=keskivärinen, 3=melko tumma, 4=tumma).

(a) Testaa värin merkitsevyys selittäjänä. Kannattaisiko joitain luokkia yhdistää?

(b) Käytä selittäjinä väriä ja pituutta. Testaa muuttujien merkitsevyys ja tulkitse estimoitu malli.

(c) Mallin yhteensopivuus aineistoon a- ja b-kohdissa?

8.5. Tutkitaan epämuodostumien esiintymisen todennäköisyyttä vastasyntyneillä äidin alkoholinkäytön mukaan (Taulukko 6.7).

Taulukko 6.7 Äidin alkoholin käyttö ja lapsen epämuodostumat.

	Epämuodostuma		
Alkoholin kulutus	Ei	On	Yhteensä
0	17066	48	17114
<1	14464	38	14502
1-2	788	5	793
3-5	126	1	127
$>=6$	37	1	38

(a) Sovita aineistoon logistinen regressiomalli, kun alkoholin kulutuksen määrinä käytetään pisteitä $\{0, 0.5, 1.5, 4, 7\}$. Tarkista mallin sopivuus aineistoon residuaalien ja devianssin avulla.

(b) Sama kuin edellisessä kohdassa, mutta yhdistä kaksi viimeistä luokkaa. Olkoon uuden viimeisen luokan "luokkakeskipiste" 5.

(c) Sama kuin a-kohdassa, muutta yhdistetään kolme viimeistä luokkaa. Valitaan viimeisen luokan "luokkakeskipisteeksi" 3. Vertaile eri luokituksilla saatuja deviansseja.