Logistinen regressio
Harjoitukset 1.3.2001

4.1. Selitä logistisella regressiomallilla mahdollisuutta sairastua sepelvaltimotautiin, kun selittäjänä on ikä ja aineistona `chddata.` Tarkastele mallin yhteensopivuutta devianssin sekä Hosmerin ja Lemeshowin testisuureen avulla. Vertaile tuloksia. Miten Hosmerin ja Lemeshowin testisuure lasketaan?

4.2. Laske edellisen tehtävän mallista selitysasteet R^2_{ss}, R^2 ja R^2_L sekä vertaile tuloksia.

4.3. Selitä keskosen syntymisen todennäköisyyttä äidin ominaisuuksilla ja raskaudenaikaisilla tapahtumilla. Aineisto on `lowbwt` ja samannimisessä aineistossa on myös selitetty muuttujat. Selitetään logistisella regressiomallilla alhaista syntymäpainoa ($< 2500g$) `low` äidin iällä `age`, painolla `lwt`, rodulla `race`, tupakoinnilla `smoke`, aikaisempiin raskauksiin liittyvillä ongelmilla `ptl` ja verenpaineella `ht`. Tutustu aineiston muuttujiin ja tarkastele mallin sopivuutta aineistoon ja tulkitse muuttujien vaikutus.

4.4. Tarkista edellisen tehtävän malli myös Pearsonin residualien ja devianssiresiduaalien avulla.

4.5. Estimoi Taulukon 3.1 (luentomoniste) aineistosta logitmalli, jossa työllistymistä selitetään koulutuksella. Korvataan sitten työllistyneiden sarakkeessa luvut 570 ja 360 vastaavasti luvuilla 57 ja 36 (10% edellisistä luvuista) ja estimoidaan jälleen malli, jossa työllistymistä selitetään koulutuksella. Vertaile mallien antamia kertoimia. Tulkitse tulos.