

1. Aineisto `laudatur`. Muuttujat:

`laud` = laudaturien lukumäärä

`lkm` = ryhmän koko

`sp` = 1, miehille ja 0 naisille,

`koulu` = 1,2,3,4; 4 koulua,

`ruotsi` = peruskoulun ruotsin arvosana

Sovita päävaikutusmalli, missä `sp` ja `koulu` ovat faktoreita ja `ruotsi` kvantitatiivinen. Muunna `ruotsi` sellaiseksi, että vakiolla on järkevä tulkinta. Tulkitse kertoimet. Onko malli residuaalidevianssin perusteella riittävä? Testaa, onko koulujen välillä eroa (logistisen mallin puitteissa).

2. Jatkoa. Tee monisteen kuvaa 6.4 vastaavat plotit.

3. Jatkoa. Laske odotetut frekvenssit laudatureille ja muille. Mitä arvelet residuaalidevianssin χ^2 -approksimaation hyvydestä. Tutki asiaa simuloimalla.

4. Aineisto `licence`. Tutkitaan, miten ajokokelaan sukupuoli, ikä ja harjoittelutausta vaikuttavat ajokokeessa hylätyksi tulemiseen.

`fail` = ajokokeen ensimmäisellä kerralla hylätyt (vaste),

`total` = ajokokeessa ensimmäisen kerran yrittäneet,

`sex`: female, male,

`age`: 18–19, 20–29, 30–39, 40+,

`training`: family = opetusluvalla harjoitelleet, school = autokoulun käyneet.

Sovita malli, jossa on vain muuttujien päävaikutukset. Voiko mallia pitää riittävänä. Tulkitse kertoimet, ja laske niiden luottamusvälit. Poikkevatko ikäryhmät toisistaan?

5. Jatkoa edelliseen. Laske mallin avulla todennäköisyydet tulla hylätyksi autokoulun käyneille miehille ja naisille eri ikäryhmissä (4 todennäköisyyttä sekä miehille että naisille). Laske näiden todennäköisyyksien osamäärät.

6. Osoita, että $Po(\lambda)$ -jakauman odotusarvo ja varianssi ovat yhtä suuret ja $= \lambda$.

7. Aineistossa `framingham` on tietoja pitkään sydäntautien seurantatutkimukseen osallistuneista. Muuttujat aineistossa ovat

- **chd** = sydäntautiin sairastuneet ko. seurantajakson aikana,
- **pyears** = henkilövuodet (=eletyt vuodet) seurantajakson aikana sydäntautiin sairastumiseen saakka,
- **period** = seurantajaksot 0–9 ja 9–18 vuotta tutkimuksen alkamisesta,
- **age** = ikäryhmät 45–49, 50–54, 55–59, 60–62 tutkimuksen alkaessa,
- **sex** = sukupuoli
- **sbp** = systolinen verenpaine, ryhmät < 165 ja ≥ 165 .

Aja Poisson-regressio osa-aineistossa, jossa **period** = 0–9. Vasteena on sydäntautiin sairastuneet ja selittäjinä **age**, **sex** ja **sbp**. Valitse offset-muuttuja niin, että sairastavuus lasketaan 1000 henkilövuotta kohden. Ota mukaan vain päävaikutukset. a) Miten paljon suurempi riski on miehillä sairastua kuin naisilla? b) Miten paljon suurempi riski on sairastua niillä, joilla on korkeampi verenpaine? c) Miten sairastavuuden riski kasvaa suhteessa nuorimpaan ikäryhmään.