

Pitkittäistutkimuksen tilastolliset menetelmät
Loppukoe 7.3.2006

1. (a) Mitä tarkoittaa "borrowing strength" -käsite? (2p) (b) Miten satunnaistekijöiden tarpeellisuus mallinnuksessa voidaan havaita pitkittäisen aineiston alustavassa (eksploratiivisessa) tarkastelussa? (2p) (c) Selitä lyhyesti, mitä REML-estimointi tarkoittaa ja milloin se on tarpeen. (2p)

2. (a) Miten pitkittäisaineiston logistisen regressiomallin parametri β_L tulkitaan kolmessa eri tapauksessa: marginaalimalli, satunnaistekijämalli, transitiomalli? (3p) (b) Miksi täyden todennäköisyyssmallin rakentaminen on olennaisesti hankalampaa kategoriselle kuin jatkuvalle vastemuuttujalle? (3p)

3. Oletetaan, että $Y_j = \mu_j + \epsilon_j$, $j = 1, \dots, n$, missä μ_j on j :nnen havainnon odotusarvo, $\epsilon_j = \alpha\epsilon_{j-1} + Z_j$ ($j > 1$, $0 < \alpha < 1$) ja muuttujat Z_j ovat toisistaan riippumattomia ja noudattavat $N(0, \tau^2)$ -jakaumaa. Lisäksi oletetaan, että ϵ_1 noudattaa $N(0, \sigma^2)$ -jakaumaa. Miten τ^2 on määriteltävä, jotta $\text{Var}(Y_j) = \sigma^2$ kaikilla $j = 1, \dots, n$? Perustele vastauksesi laskemalla. (6p)

4. Oletetaan, että jatkuva vastemuuttuja noudattaa seuraava mallia:

$$Y_{ij} = \beta_0 + U_i + Z_{ij}, \quad U_i \sim N(0, \nu^2), \quad Z_{ij} \sim N(0, \tau^2), \quad i = 1, \dots, n, j = 1, \dots, m.$$

Lisäksi U_i :t ja Z_{ij} :t ovat kaikki toisistaan riippumattomia. Yksilö i jää pois seurannasta hetkellä j (eli havainto Y_{ij} jää jo puuttumaan, 'drop-out') todennäköisyydellä p_{ij} , joka riippuu vastemuuttujasta seuraavasti:

$$\text{logit}(p_{ij}) = -1 + \alpha y_{i,j-1}, \quad \alpha > 0.$$

(a) Millainen havaintojen puuttumismekanismi on kyseessä? (2p) (b) Kuinka havainnoista laskettu empiirinen keskiarvo $m_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ keskimäärin käyttäytyy mittausajankohdan j funktiona? Selitä, miksi. (Yllä n_j on hetkellä j seurannassa vielä olevien yksilöiden lukumäärä.) (4p)

5. Tarkastellaan frekvenssiaineiston Poisson-mallia, missä

$$\log \{ E(Y_{ij}) \} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2} + \log(t_{ij}), \quad i = 1, \dots, n, j = 1, 2.$$

Tässä x_{i1} on 0, jossa yksilö i kuuluu lumelääkittyjen ryhmään, ja 1, jos aktiivilääkettä saavien ryhmään. Muuttuja x_{ij2} on 0 jaksolla $j = 0$ (ennen lumentai aktiivilääkityksen aloittamista) ja 1 jaksolla $j = 1$ (lume- tai aktiivilääkityksen aloittamisen jälkeen). Havaintovälien pituudet ovat t_{ij} ($j = 1, 2$). Tulkitse parametrit β_0 , β_1 , β_2 ja β_3 . Perustele. (6p)

Pitkittäistutkimuksen tilastolliset menetelmät
Loppukoe 3.4.2006

1. (a) Mitä tarkoittaa "borrowing strength" -käsite? (2p) (b) Miten satunnaistekijöiden tarpeellisuus mallinnuksessa voidaan havaita pitkittäisen aineiston alustavassa (eksploratiivisessa) tarkastelussa? (2p) (c) Selitä lyhyesti, mitä REML-estimointi tarkoittaa ja milloin se on tarpeen. (2p)

2. (a) Miten pitkittäisaineiston logistisen regressiomallin parametri β_L tulkitaan kolmessa eri tapauksessa: marginaalimalli, satunnaistekijämalli, transitiomalli? (3p) (b) Miksi täyden todennäköisyysmallin rakentaminen on olennaisesti hankalampaa kategoriselle kuin jatkuvalla vastemuuttujalle? (3p)

3. Oletetaan, että $Y_j = \mu_j + \epsilon_j$, $j = 1, \dots, n$, missä μ_j on j :nnen havainnon odotusarvo, $\epsilon_j = \alpha\epsilon_{j-1} + Z_j$ ($j > 1$, $0 < \alpha < 1$) ja muuttujat Z_j ovat toisistaan riippumattomia ja noudattavat $N(0, \tau^2)$ -jakaumaa. Lisäksi oletetaan, että ϵ_1 noudattaa $N(0, \sigma^2)$ -jakaumaa. Miten τ^2 on määriteltävä, jotta $\text{Var}(Y_j) = \sigma^2$ kaikilla $j = 1, \dots, n$? Perustele vastauksesi laskemalla. (6p)

4. Oletetaan, että jatkuva vastemuuttuja noudattaa seuraava mallia:

$$Y_{ij} = \beta_0 + U_i + Z_{ij}, \quad U_i \sim N(0, \nu^2), \quad Z_{ij} \sim N(0, \tau^2), \quad i = 1, \dots, n, j = 1, \dots, m.$$

Lisäksi U_i :t ja Z_{ij} :t ovat kaikki toisistaan riippumattomia. Yksilö i jää pois seurannasta hetkellä j (eli havainto Y_{ij} jää jo puuttumaan, 'drop-out') todennäköisyydellä p_{ij} , joka riippuu vastemuuttujasta seuraavasti:

$$\text{logit}(p_{ij}) = -1 + \alpha y_{i,j-1}, \quad \alpha > 0.$$

(a) Millainen havaintojen puuttumismekanismi on kyseessä? (2p) (b) Kuinka havainnoista laskettu empiirinen keskiarvo $m_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ keskimäärin käyttäytyy mittausajankohdan j funktiona? Selitä, miksi. (Yllä n_j on hetkellä j seurannassa vielä olevien yksilöiden lukumäärä.) (4p)

5. Tarkastellaan frekvenssiaineiston Poisson-mallia, missä

$$\log \{ E(Y_{ij}) \} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2} + \log(t_{ij}), \quad i = 1, \dots, n, j = 1, 2.$$

Tässä x_{i1} on 0, jossa yksilö i kuuluu lumelääkittyjen ryhmään, ja 1, jos aktiivilääkettä saavien ryhmään. Muuttuja x_{ij2} on 0 jaksolla $j = 0$ (ennen lumentai aktiivilääkityksen aloittamista) ja 1 jaksolla $j = 1$ (lumentai aktiivilääkityksen aloittamisen jälkeen). Havaintovälien pituudet ovat t_{ij} ($j = 1, 2$). Tulkitse parametrit β_0 , β_1 , β_2 ja β_3 . Perustele. (6p)

Pitkittäistutkimuksen tilastolliset menetelmät
Loppukoe 18.5.2006

1. (a) Mitä tarkoittaa "borrowing strength" -käsite? (2p) (b) Miten satunnaistekijöiden tarpeellisuus mallinnuksessa voidaan havaita pitkittäisen aineiston alustavassa (eksploratiivisessa) tarkastelussa? (2p) (c) Selitä lyhyesti, mitä REML-estimointi tarkoittaa ja milloin se on tarpeen. (2p)

2. (a) Miten pitkittäisaineiston logistisen regressiomallin parametri β_L tulkitaan kolmessa eri tapauksessa: marginaalimalli, satunnaistekijämalli, transitiomalli? (3p) (b) Miksi täyden todennäköisyysmallin rakentaminen on olennaisesti hankalampaa kategoriselle kuin jatkuvalle vastemuuttujalle? (3p)

3. Oletetaan, että $Y_j = \mu_j + \epsilon_j$, $j = 1, \dots, n$, missä μ_j on j :nnen havainnon odotusarvo, $\epsilon_j = \alpha\epsilon_{j-1} + Z_j$ ($j > 1$, $0 < \alpha < 1$) ja muuttujat Z_j ovat toisistaan riippumattomia ja noudattavat $N(0, \tau^2)$ -jakaumaa. Lisäksi oletetaan, että ϵ_1 noudattaa $N(0, \sigma^2)$ -jakaumaa. Miten τ^2 on määriteltävä, jotta $\text{Var}(Y_j) = \sigma^2$ kaikilla $j = 1, \dots, n$? Perustele vastauksesi laskemalla. (6p)

4. Oletetaan, että jatkuva vastemuuttuja noudattaa seuraava mallia:

$$Y_{ij} = \beta_0 + U_i + Z_{ij}, \quad U_i \sim N(0, \nu^2), \quad Z_{ij} \sim N(0, \tau^2), \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

Lisäksi U_i :t ja Z_{ij} :t ovat kaikki toisistaan riippumattomia. Yksilö i jää pois seurannasta hetkellä j (eli havainto Y_{ij} jää jo puuttumaan, 'drop-out') todennäköisyydellä p_{ij} , joka riippuu vastemuuttujasta seuraavasti:

$$\text{logit}(p_{ij}) = -1 + \alpha y_{i,j-1}, \quad \alpha > 0.$$

(a) Millainen havaintojen puuttumismekanismi on kyseessä? (2p) (b) Kuinka havainnoista laskettu empiirinen keskiarvo $m_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ keskimäärin käyttäytyy mittausajankohdan j funktiona? Selitä, miksi. (Yllä n_j on hetkellä j seurannassa vielä olevien yksilöiden lukumäärä.) (4p)

5. Tarkastellaan frekvenssiaineiston Poisson-mallia, missä

$$\log \{ E(Y_{ij}) \} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2} + \log(t_{ij}), \quad i = 1, \dots, n, \quad j = 1, 2.$$

Tässä x_{i1} on 0, jossa yksilö i kuuluu lumelääkittyjen ryhmään, ja 1, jos aktiivilääkettä saavien ryhmään. Muuttuja x_{ij2} on 0 jaksolla $j = 0$ (ennen lumentai aktiivilääkityksen aloittamista) ja 1 jaksolla $j = 1$ (lumentai aktiivilääkityksen aloittamisen jälkeen). Havaintovälien pituudet ovat t_{ij} ($j = 1, 2$). Tulkitse parametrit β_0 , β_1 , β_2 ja β_3 . Perustele. (6p)