Topic 10: Marginal Models for clustered responses

Nov19/revisited on Dec2

Problème

1. pp 69 correlation can be no larger than 0.25
2. pp76 working covariance matrix

Announcements: Final exam Dec.4th 2-5

Final Exam materials: Modeling

1. Binary logistic regression
2. generalized liinear models
3. multinomial logistic regression
4. proportional odds model
5. Log linear modeling
6. analysis of matched pairs
7. mariginal and conditional models for clustered data
8. need to be able to interpret SAS out from these types of model fits

Clarifications

1. **overlapping CI question:** Do not overlap _ > result is statistically significant
2. **DF 60** : Wrong
3. # of parameters in the saturated model = 64
4. # of parameters in the incercept only model = 4
5. Devicance df = 64 -4 = 60
6. λ in a log linear model
    1. LHS observed value VS RHS model fitting value
    2. $log\mu_{black,belief=Yes} = \lambda + \lambda_{black}^{RACE} + \lambda_{Yes}^{Belief}$
    3. **Observed**: Log260 = 5.56
    4. **Model fit value**: Lambda + lambda + lambda = 3.0003 + 1.05 + 1.49 = 5.55

Marginal Models for clustered responses

1. Many studies observe the response at several times/various conditions eg. **Longitudinal Studies**.
2. correlated observations: the response variables is observed for **matched sets** of subjects: **Clusters**
   1. Repeated measurement on subjects
   2. Ususally **positively** correlated
   3. Analyses ignore the correlation can estimate model parameters well but se estimators can be badly biased
3. **Mariginal Probabilities** $P(Y_1 = 1)...P(Y_T = 1)$

   Marginal Model: $logit[P(Y_i = 1)]$
4. GLM: choice of distribution for Y determines the relationshipbetween μ and $Var(Y)$.

   eg. Binary Case: $E(Y) = \pi. Var(Y) = \pi(1-\pi), which is \mu(1-\mu)$

   Con: ML method must assume a particular type of dist'n of Y

5. GEE: **Quasi-likelihood assumes only a relationship betweenμ and $Var(Y)$ rather than a specific probability**
   1. Which **only** links each marginal mean to a linear predictor and provides a guess for a variance-covariance structure of (Y1, Y2...YT)
   2. Generalized estimating equations
   3. **PART1**$g(\mu) = \eta_{ij} = X'_{ij}\beta$. The conditional expectation or mean of each response $E(Y_{ij}|X_{ij}) = \mu_{ij}$
   4. **PART2** $Var(Y_{ij}|X_{ij}) = \phi v(\mu_{ij})$ where phi is a scale parameter adjucst the var for over dispersion

      First two: no distributional assumptions about the responses
   5. **PART3** conditional with-subject association among the repeated responses, **Covariance** is assumed to be a functinon of an additiona set of parameters α

      MAIN extension of GLM to Longitudinal Data**
   6. Avoid term Correlation: since its for continuous responses with range from -1 to +1. Not the case for discrete responses.
6. **Marginal Model for binary response**:
   1. E(Yij | Xij) = Pr(Yij = 1 | Xij)
   2. $Log(\frac{\mu_{ij}}{1-\mu_{ij}}) = \eta_{ij} = X'_{ij}\beta$
   3. Var(Yij | Xij) = $\mu_{ij}(1-\mu_{ij})$
   4. LogOR(Yij, Yik| Xij, Xik) = $alpha + jk$
   5. OR(Yj, Yk). = Pr(Yj= 1 Yk = 1)Pr(Yj= 0 Yk = 0)/ (Pr(Yj= 1 Yk = 0)Pr(Yj= 0 Yk = 1))

7. $V_i = A_i^{1/2} Corr(Y_i) A_i^{1/2}$ where A_i is a diagonal matrix with $Var(Y_{ij}|X_{ij}) = \phi v(\mu_{ij})$.

   1/2 term is a diagonal matrix with standard deviations.

   Corr(Yi) is a correlation matrix, a function o fα

8. Vi is **working** covariance

9. Structures

   1. Independence correlation
   2. Exchangeable
   3. Autoregressive
   4. unstructured