STA 303 / 1002

Note Title 4/2/2012

Exam: covers everything. Bring a calculator

Last page: posted (see Announcements on Blackboard)

Pre-exam office hours: Wed April 11 10-12 am
Mon April 16 10-12 am
Wed April 18 3-5 pm

Today - course evaluations at break

Diet - Diabetes Example

Subjects randomly assigned to 1 of 3 diets
Measured HDL at 3 time points

We've fit 2 models:

For both: Fixed effects
- diet (2 indicator variables)
- time (2 indicator variables)
- diet * time interaction (4 terms)

Covariance structure

Model 1 Compound symmetry, same for all subjects (same for every diet)

↳ same variance at each time point and same covariance on any pair of observations on a subject

⇒ 2 covariance parameters

Model 2 Unstructured, different for every diet

↳ different variances for each time point & all covariances different

⇒ 18 covariance parameters

Now Compare three models for how well they fit the data

Since covariance parameters estimated by maximum likelihood, can compare models 1 & 2 using AIC, BIC or LRT

Model 1

AIC = -105.8

BIC = -101.3

Model 2:

AIC = -103.4

BIC = -62.7 ← BIG penalty for # of parameters (Too big)

AIC picks Model 1

_______.

LRT H_0: fit of models ① & ② is equivalant

OR

16 restrictions on variances & covariances to turn model ② into model ① hold.

H_a: more complicated fits data better (model ②)

OR

at least one of 16 restrictions doesn't hold

Test statistic $-2\log L_R - (-2\log L_F)$

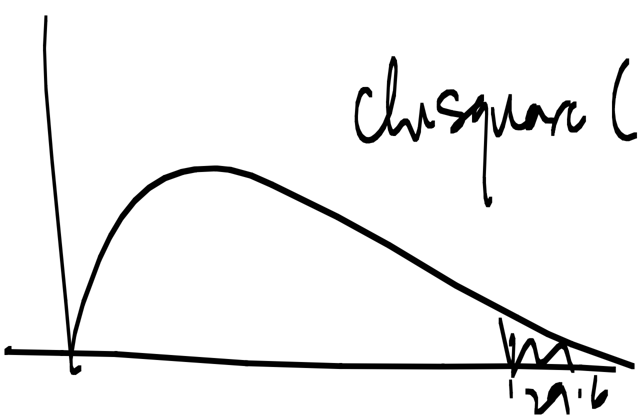
Restricted log-likelihoods

$$= -109.8 - (-139.4)$$

$$= 29.6$$

Under H_0, this is an observation from a chi-square distribution with $18-2=16$ df

Chisquare(16)



p-value = .02 (R)

Tables: $.01 < p < .025$

Moderate ~~Some~~ evidence in favour of unstructured cov. structure with different values for each diet.

(Different decision than looking at AIC. Perhaps AIC's penalty on number of parameters is also too strong)

A possibly better model for variance-covariance structure (Model 3)

Compound Symmetry within diets, but different estimates for each diet

$$D_{HG} = \begin{pmatrix} \sigma^2_{u,HG} + \sigma^2_{e,HG} & \sigma^2_{u,HG} & \sigma^2_{u,HG} \\ & \sigma^2_{u,HG} + \sigma^2_{e,HG} & \sigma^2_{u,HG} \\ & & \sigma^2_{u,HG} + \sigma^2_{e,HG} \end{pmatrix}$$

$+ D_{HM} + D_{LG}$

$\Rightarrow 2 \times 3 = 6$ covariance parameters

Output $-2\log L = -120.6$

$AIC = -108.6$ (for model ①: $AIC = -105.8$) model ② AIC is bigger)

AIC picks model ③

LRT: Compare models ③ to ②

Test **stat**: $G^2 = -120.6 - (-139.4)$

$= 18.8$

Under H_0, this is an observation from a chi-square distribution with 12 df

$p = .09$

Weak evidence that model 2 fits data better

Since evidence is only weak and model 3 is simpler, I'm going to proceed with model 3.

Tests of fixed effects

Analogous to Type III tests in proc glm

diet × time interaction $p = .0291$

$$\left(H_0: \underbrace{\beta_5 = \beta_6 = \beta_7 = \beta_8}_{\text{coefficients of interaction terms}} = 0 \quad \text{vs} \quad H_a: \text{at least } \beta_5, \beta_6, \beta_7, \beta_8 \text{ is not } 0 \right)$$

Follow-up significant F-test for interaction by investigate the least square means

LSmeans estimates of $E(Y)$ calculated from fitted equation

For example,

diet = HG, time = 1 $\quad \hat{E}(Y) = .9435 + .1323 + .02769 + 0.0232 = 1.1267$

Q2:

lsmeans diet*time / pdiff ;

t-tests for pairwise differences of means

Adjusting for Type I error rate using Tukey's or Bonferroni's method would be conservative since I'm only interested in comparisons between same diet at different times or different diets at same time.

Significant Results of interest (ignoring $0.05 < p < .01$ since I didn't adjust for Type I error)

HM time 1 $\neq$ time 3 $p = .0179$

HM $\neq$ LG at time 1 $p = .0235$

HM $\neq$ LG at time 2 $p = .0028$

HM $\neq$ LG at time 3 $p = .0006$

Conclude:

HM is the only that differs over time

(means at time 3 for HM is ____ and at time 1 is ____ and this difference is statistically significant

________________________ .

<u>Model</u> assumptions:

- Need right form of model
  - include right explanatory variables and don't exclude useful explanatory variables
  - If an explanatory variable is continuous, need linear relationship with Y
- Need subjects to be independent (but <u>not</u> independent observations on the same subject)

- Need variance-covariance structure to be as modelled
- Need normality of error terms – so no outliers

Can check some of these with residuals.