

# Tutorial Notes 10 of MATH3424

## 1 Summary of course material

1. Consequences of variable selection
2. Criteria for evaluating equations:
  - (a) Residual mean square:  $\text{RMS}_p = \frac{\text{SSE}_p}{n-p}$
  - (b) Mallows  $C_p$ :  $C_p = \frac{\text{SSE}}{\hat{\sigma}^2} + (2p - n)$
  - (c) AIC:  $\text{AIC}_p = n \ln(\text{SSE}_p/n) + 2p$
  - (d) BIC:  $\text{AIC}_p = n \ln(\text{SSE}_p/n) + p \ln n$
3. Evaluating all possible equations: valid for both collinear and non-collinear data
4. Variable selection procedures
  - (a) Forward selection procedure
  - (b) Backward elimination procedure
  - (c) Stepwise method
5. The logit model
  - (a) Modeling qualitative data

$$\pi = \Pr(Y = 1 | X_1 = x_1, \dots, X_p = x_p) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

- (b) Maximum likelihood estimator

$$\hat{\beta} \leftarrow \max_{\beta} \sum_{i=1}^n (y_i \beta' \mathbf{x}_i - \log(1 + \exp(\beta' \mathbf{x}_i)))$$

- (c) Null deviance and Residual deviance: G-statistic for model testing
  - (d) Distribution of maximum likelihood estimator  $\hat{\beta}$ : approximately unbiased, follow normal distribution
  - (e) Statistical inference for  $\hat{\beta}$
6. Logistic regression diagnostic
  7. Multinomial logit model

## 2 Exercise

### 2.1

**Annual dues.** (The data is uploaded in Canvas under “Tutorial Slides” directory.)

The board of directors of a professional association conducted a random sample survey of 30 members to assess the effects of several possible amounts of dues increase. The sample results follow.  $X$  denotes the dollar increase in annual dues posited in the survey interview, and  $Y = 1$  if the interviewee indicated that the membership will not be renewed at that amount of dues increase and 0 if the membership will be renewed.

Logistic regression model is assumed to be appropriate.

1. Find the maximum likelihood estimates of  $\beta_0$  and  $\beta_1$ . State the fitted response function.
2. Obtain a scatter plot of the data with both the fitted logistic response function from part (a). Does the fitted logistic response function appear to fit well?
3. Obtain  $\exp(\hat{\beta}_1)$  and interpret this number.
4. What is the estimated probability that association members will not renew their membership if the dues are increased by \$40?
5. Estimate the amount of dues increase for which 75 percent of the members are expected not to renew their association membership.