Regression Models for Qualitative/Quantitative Predictors and ANOVA

C. Andy Tsao

Dept. of Applied Math
National Dong Hwa University

April 7, 2011

Outline

Chapter 8, 16.

Variations on regression models

- Polynomial regression models: e.g. $Y_i = \beta_0 + \beta_1 x_i^{'} + \beta_2 x_i^{'2} + \epsilon_i, i = 1, \cdot, n$. where $x_i^{'} = x_i - \bar{x}$ or $x_i^{'} = (x_i - \bar{x})/sd(x)$ etc.
- Interaction Regression models: e.g. $E(Y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$. Graphical Illustrations.
- General form: $E(Y|x) = f_1(x_1) + f_2(x_2) + f_3(x_1, x_2)$
- Interpretation of parameters
- Numerical stable, practically interpretable and flexible

Interaction with indicator

Two simple linear regression models with the normal errors with common variance.

Interaction with indicator

Two simple linear regression models with the normal errors with common variance.

$$Y_i = a_1 + b_1 X_i + \epsilon_i, \quad i = 1, \cdots, n \tag{1}$$

$$Y_j = a_2 + b_2 X_j + \epsilon_j, \quad j = n+1, \cdots, n+m \tag{2}$$

Trend changes

Interaction with indicator

Two simple linear regression models with the normal errors with common variance.

$$Y_i = a_1 + b_1 X_i + \epsilon_i, \quad i = 1, \cdots, n \tag{1}$$

$$Y_j = a_2 + b_2 X_j + \epsilon_j, \quad j = n+1, \cdots, n+m \tag{2}$$

Trend changes

$$Y = a_1 + b_1 X + \epsilon, \quad \text{for } X < x_0$$
$$Y = a_2 + b_2 X + \epsilon, \quad \text{for } X \geq x_0$$

Some problems call for alternative models than regression.

- Which grad school is the best?
- Which treatment is better? (Program A, B, C)
- What dosage level (low, medium, high) is most effective?
- What is the best treatment combination to manufacture a product?

Math form

- $E(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- $E(Y|\tau, B) = \mu + \tau + B$

Qualitative vs. Quantitative

- Picture (Figure 16.1, KNNL)
- Factor, Factor Level ("Value" of the factor)
- Spectrum from Quantitative–Qualitative variables. Categorical Variables.
- Single factor versus Multifactor

Single Factor ANOVA

- Cell Means Model
  $Y_{ij} = \mu_i + \epsilon_{ij}, i = 1, \cdots, r; j = 1, \cdots, n_i,$
  $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$
- Factor Effects Model
  $Y_{ij} = \mu_. + \tau_i + \epsilon_{ij}, i = 1, \cdots, r; j = 1, \cdots, n_i,$
  $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$
- Connection
  $\mu_. = \sum_{i=1}^{r} \mu_i / r; \sum \tau_i = 0$
- Both models are GLM satisfying GM conditions.

Connection with two-sample t-tests

Regression vs. ANOVA

- Design Matrices for GLM
- Typical Question of interest:
  $\beta =, \geq, \leq 0?$ vs $\sum_i \tau_i^2 = 0?$ and ordering in τ.
- Calculation glm vs. lm

Choice of models

- Quantitative predictor: Resolution, Precision vs. Robustness
- Qualitative predictor: type of ordering. scoring.