CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

Lecture 5
Regression Review, part 1

Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

Vocabulary (1)

- Experiment – explanatory variables are manipulated (controlled), all other inputs are held constant, and the response variable is measured
  - As opposed to an observation
- Model – a mathematical function that approximates the data and is useful for making predictions
- Regression – a method of finding the best fit of a model to a given set of data through the adjustment of model parameters (coefficients)
  - What do we mean by "best fit"?

Vocabulary (2)

- Residual: $\varepsilon_i = y_i - f(\boldsymbol{x_i}, \boldsymbol{\beta})$, $e_i = y_i - f(\boldsymbol{x_i}, \boldsymbol{b})$
- A "good" model has small residuals
  - Small mean, $|\bar{\varepsilon}|$
  - Small variance, σ_ε^2
- Much of our model and regression checking will involve studying and plotting residuals
  - Plot e_i $vs.$ $\hat{y}_i$ (works for multiple regression)

What Do We Mean by "Best Fit"?

- First, assume ε is a random variable that does not depend on x
  - There are no systematic errors
  - The model is perfect
- Desired properties of a "best fit" estimator
  - Unbiased, $\sum \varepsilon_i = \sum e_i = 0$ (and $E[b_k] = \beta_k$)
  - Efficient, minimum variance σ_ε^2 (and $var(b_k)$)
  - Maximum likelihood (for an assumed pdf of ε)
  - Robust (to be discussed later)

Maximum Likelihood Estimation

(for a review, see lectures 16 & 17 at www.lithoguru.com/scientist/statistics/review.html)

- Likelihood function: the probability of getting this exact data set given a known model and model parameters
  - To calculate the likelihood function, we need to know the joint probability distribution function (pdf) of ε
- Maximum Likelihood: what parameter values maximize the likelihood function?
  - Take the partial derivative of the likelihood function (or more commonly the log-likelihood function) with respect to each parameter, set it equal to zero
  - Solve resulting equations simultaneously

MLE Example – straight line

- Let $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma_\varepsilon)$, iid
- Since each ε_i is independent, the likelihood function for the entire data set is

$$L = \prod_{i=1}^{n} \mathbb{P}(y_i) = \prod_{i=1}^{n} \mathbb{P}(\varepsilon_i)$$

- Since the residuals are iid Normal,

$$L = \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^n exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_\varepsilon^2}\right]$$

MLE Example (2)

- Define chi-square as

$$\chi^2 = \sum_{i=1}^{n} \frac{\varepsilon_i^2}{\sigma_\varepsilon^2} \quad \text{(also called weighted SSE)}$$

- Maximizing L is the same as minimizing χ^2
- Solve these equations simultaneously

$$\frac{\partial \chi^2}{\partial \beta_k} = 0 \; for\ all\ k\ (all\ coefficients)$$

MLE Example (3)

- For our line model, $\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{\sigma_\varepsilon^2}$$

Intercept: $$\frac{\partial \chi^2}{\partial \beta_0} = 0 = \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{n} 2(y_i - (\beta_0 + \beta_1 x_i))(-1)$$

$$0 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i)) = n(\bar{y} - \beta_0 - \beta_1 \bar{x})$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

same as $\sum_{i=1}^{n} e_i = 0$

MLE Example (4)

- Substitute our estimate for β_0 into χ^2

$$\chi^2 = \sum_{i=1}^{n} \frac{\left((y_i - \bar{y}) - \beta_1 (x_i - \bar{x})\right)^2}{\sigma_\varepsilon^2}$$

Slope: $$\frac{\partial \chi^2}{\partial \beta_1} = 0 = \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{n} 2\left((y_i - \bar{y}) - \beta_1 (x_i - \bar{x})\right)(-(x_i - \bar{x}))$$

same as $\sum_{i=1}^{n} x_i e_i = 0$

$$b_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Assumptions for this MLE

- We assumed three things:
  - Each residual (and y_i) is independent
  - Each residual (and y_i) is identically distributed
  - Each $\varepsilon_i \sim N(0, \sigma_\varepsilon)$, and thus $y_i \sim N(\hat{y}, \sigma_\varepsilon)$
  - We call this “ordinary least squares” (OLS)
- If any of these assumptions are invalid, then our least-squares estimates will not be the maximum likelihood estimates

Properties of our OLS Solution

1. The parameters are unbiased estimators of the true parameters, with minimum variance compared to all other unbiased estimators (*if* our assumptions are correct)
2. $\sum_{i=1}^{n} e_i = 0$
3. $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$, *so that* $\bar{y} = \bar{\hat{y}}$
4. $\sum_{i=1}^{n} \hat{y}_i e_i = 0$
5. $\sum_{i=1}^{n} x_i e_i = 0$
6. The best fit line goes through the point $(\bar{x}, \bar{y})$

Lecture 5: What have we learned?

- What is a residual?
- What are the desired properties of a “best fit” parameter estimate?
- Explain MLE, maximum likelihood estimation.
- What are the assumptions of ordinary least squares (OLS) estimation?