STAT 410 - Linear Regression

Lecture 8

Meng Li

Department of Statistics

Sep. 28, 2017

ANOVA in MLR

- We first obtain $SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$.
  - Facts: $\mathbf{e}'\mathbf{X} = \mathbf{0}$ and $\mathbf{e}'\hat{\mathbf{y}} = 0$ (Geometric interpretation of LS estimators)
- $SS_T = \mathbf{y}'\mathbf{y} - n\bar{y}^2 = \mathbf{y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{y}$, where $\mathbf{1} = (1, \ldots, 1)'$ is a $n \times 1$ vector.
- Thus, $SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2 = \mathbf{y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{y}$.

Method of extra-sum-of-squares

- The extra-sum-of-squares method allows to investigate the contribution of a single and a subset of the regressor variables to the model.
- Recall the multiple linear regression model with k regressors: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}$ is $p \times 1$ and $p = k + 1$.
- Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ where $\boldsymbol{\beta}_1$ is $(p-r) \times 1$ and $\boldsymbol{\beta}_2$ is $r \times 1$
- We wish to test

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}. \tag{1}$$

- Under H_0, the model reduces to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$ (this is the **reduced** model vs. the **full** model)
- In the ANOVA table, what are SS_R and SS_{Res} under both the full and reduced model?

- The regression sum of squares due to $\boldsymbol{\beta}_2$ given that $\boldsymbol{\beta}_1$ is already in the model

$$SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{y}.$$

- The degrees of freedom of $SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ is

$$k - (k - r) = r.$$

- Under the full model, we have

$$MS_{Res} = \frac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{n - p}.$$

- This leads to the F test:

$$F_0 = \frac{SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/r}{MS_{Res}} \sim F_{r,n-p}.$$

- It is sometimes called a **partial F test** because it measures the contribution of $\mathbf{X}_2$ given $\mathbf{X}_1$ were already in the model.

A special case when $r = 1$

- Consider three regressors: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, and the three sums of squares

$$SS_R(\beta_1|\beta_0, \beta_2, \beta_3), \quad SS_R(\beta_2|\beta_0, \beta_1, \beta_3), \quad SS_R(\beta_3|\beta_0, \beta_1, \beta_2).$$

  - Each measures the contribution of x_j as if it were the last variable added to the model.
  - Degrees of freedom: one
  - In general, we can assess the value of adding x_j to a model that did not include this regressor by using

$$SS_R(\beta_j|\beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k), \quad 1 \le j \le k.$$

- Partial F test provides a useful tool in **model building** when many regressors are available and we would like to find the best set of regressors for the model in use.
- We can show that this partial F test in this case is equivalent to the t test for $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.

More properties of SS

- Let $\boldsymbol{\beta}_2 = (\beta_1, \ldots, \beta_k)'$ and $\boldsymbol{\beta}_1 = \beta_0$, we then have
$$SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SS_R(\beta_1, \ldots, \beta_k|\beta_0) = SS_T - SS_{Res}.$$
- Sequential decomposition of SS:
$$\begin{aligned} & SS_R(\beta_1, \beta_2, \beta_3|\beta_0) \\ & = SS_R(\beta_1|\beta_0) + SS_R(\beta_2|\beta_1, \beta_0) + SS_R(\beta_3|\beta_1, \beta_2, \beta_0). \end{aligned}$$
- The decomposition above is invariant to a permutation of $(\beta_1, \beta_2, \beta_3)$, e.g.,
$$\begin{aligned} & SS_R(\beta_1, \beta_2, \beta_3|\beta_0) = SS_R(\beta_3, \beta_2, \beta_1|\beta_0) \\ & = SS_R(\beta_3|\beta_0) + SS_R(\beta_2|\beta_3, \beta_0) + SS_R(\beta_1|\beta_2, \beta_3, \beta_0). \end{aligned}$$
- But in general,
$$\begin{aligned} & SS_R(\beta_1, \beta_2, \beta_3|\beta_0) \\ & \neq SS_R(\beta_1|\beta_0, \beta_2, \beta_3) + SS_R(\beta_2|\beta_0, \beta_1, \beta_3) + SS_R(\beta_3|\beta_0, \beta_1, \beta_2). \end{aligned}$$

Testing General Linear Hypotheses

- Under the MLR model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, suppose we are interested in testing the hypotheses:

$$H_0 : \mathbf{T}\boldsymbol{\beta} = \mathbf{0} \text{ vs. } H_1 : \mathbf{T}\boldsymbol{\beta} \neq \mathbf{0},$$

  where $\mathbf{T}$ is an $r \times p$ matrix.
- Without loss of generality, we assume rows of $\mathbf{T}$ are linearly independent and $r \leq p$ (thus the rank of $\mathbf{T}$ is r).
- In the same spirit of sums of squares from ANOVA, we conduct a test statistic by

$$\begin{aligned} SS_H &= SS_R(\text{Full model}) - SS_R(\text{Reduced model}) \\ &= SS_{Res}(\text{Reduced model}) - SS_{Res}(\text{Full model}). \end{aligned}$$

- Specifically, we use

$$F_0 = \frac{SS_H/r}{SS_{Res}(\text{Full model})/(n-p)} \sim F_{r,n-p}(\text{Under } H_0).$$

- Under the full model, we have $SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ which has $(n-p)$ degrees of freedom.
- Under the reduced model where $\mathbf{T}\boldsymbol{\beta} = \mathbf{0}$, we first solve for r regression coefficients in terms of the remaining $p-r$ regression coefficients, leading to

$$\mathbf{y} = \mathbf{Z}\gamma + \boldsymbol{\varepsilon},$$

  where $\mathbf{Z}$ is an $n \times (p-r)$ matrix and γ is a $(p-r) \times 1$ vector.
  - The estimate of γ is $\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$.
  - The residual SS is $SS_{Res}(\text{Reduced model}) = \mathbf{y}'\mathbf{y} - \hat{\gamma}'\mathbf{Z}'\mathbf{y}$ which has $(n-p+r)$ degrees of freedom.

Examples for obtain the reduced model

- Consider the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$.
- Example 1: we let $\mathbf{T} = (0, 1, 0, -1)$.
  - $\mathbf{T}\boldsymbol{\beta} = \mathbf{0}$ means that $\beta_1 - \beta_3 = 0$ or $\beta_3 = \beta_1$.
  - The reduced model becomes

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_1 x_3 + \varepsilon \\ &= \beta_0 + \beta_1 (x_1 + x_3) + \beta_2 x_2 + \varepsilon \\ &= \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon. \end{aligned}$$

- Example 2: we let

$$\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

  - $\mathbf{T}\boldsymbol{\beta} = \mathbf{0}$ means that $\beta_3 = \beta_1$ and $\beta_2 = 0$.
  - The reduced model becomes

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_1 x_3 + \varepsilon \\ &= \beta_0 + \beta_1 (x_1 + x_3) + \varepsilon \\ &= \gamma_0 + \gamma_1 z_1 + \varepsilon. \end{aligned}$$

An alternative approach

- Motivated by the t test, we consider

$$H_0 : \mathbf{T}\boldsymbol{\beta} = \mathbf{c} \quad \text{vs.} \quad H_1 : \mathbf{T}\boldsymbol{\beta} =\neq \mathbf{c}.$$

- Under H_0,
  - we have $\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c} \sim \text{MVN}(\mathbf{0}, \sigma^2 \mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}')$; (Why?)
  - It follows that

$$(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}')]^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c}) \sim \chi^2_r.$$

- Therefore, we propose to use the test statistic

$$F_0 = \frac{(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}')]^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c})/r}{SS_{Res}(\text{Full model})/(n-p)}, \tag{2}$$

  which follows $F_{r,n-p}$ under H_0.
- The numerator of Equation (2) measures the squared distance between $\mathbf{T}\boldsymbol{\beta}$ and $\mathbf{c}$ standardized by the covariance matrix of $\mathbf{T}\hat{\boldsymbol{\beta}}$.

Simultaneous Confidence Interval

- A 100(1 - α)% simultaneous confidence interval covers a set of parameters simultaneously with probability 1 - α. We usually refer to it as **joint confidence interval/region.**
- It is still derived based on a pivotal quantity.
  - A pivotal quantity depends on both parameters of interest and data;
  - The sampling distribution of a pivotal quantity does not depend on the parameters and is completely known.
- We here use

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/p}{MS_{Res}}$$

  as the pivotal quantity, which follows $F_{p,n-p}$.
- Thus a 100(1 - α)% joint confidence region for $\boldsymbol{\beta}$ is

$$\{\boldsymbol{\beta} : \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/p}{MS_{Res}} \leq F_{\alpha,p,n-p}\}. \quad (3)$$

- It is possible to combine individual confidence intervals to obtain a joint confidence region for $\boldsymbol{\beta}$ by using the Bonferroni method:

$$\{\boldsymbol{\beta} : \beta_j \in [\hat{\beta}_j - t_{\alpha/(2p),n-p} se(\hat{\beta}_j), \hat{\beta}_j + t_{\alpha/(2p),n-p} se(\hat{\beta}_j)]\},$$

  where the individual Bonferroni interval has a confidence coefficient $(1-\alpha/p)$ instead of $(1-\alpha)$.
- Similarly to Equation (3), a 100(1 - α)% joint confidence region for $\gamma = \mathbf{T}\boldsymbol{\beta}$ is

$$\left\{\gamma : \frac{(\mathbf{T}\hat{\boldsymbol{\beta}} - \gamma)'[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}')]^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \gamma)/r}{MS_{Res}} \leq F_{\alpha,r,n-p}\right\}.$$