

Section Note 05: Hypothesis Testing

JJ Chen^{*}

October 2, 2014

Bugs might exist. In the review session, we mainly talked about geometric interpretations of the Wald principle and the Likelihood principle. This note corresponds to materials covered in class in week 5 and week 6. Here I did not fill in relevant graphs I drew in the session (restricted LS triangle and likelihood contour) that delivers the geometric intuitions. If you did not come to the session but want to know the visualization, let me know and we can set up a time for meeting.

1 Review of the Weeks

These two weeks we have the following key results for hypothesis testing:

1. t statistic:

$$\begin{aligned} \text{Condition 1: } z_j &= \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\sigma^2 s_{jj}}} \sim N(0, 1) \\ + \text{Condition 2: } (n - k) \frac{s^2}{\sigma^2} &= \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2(n - k) \\ + \text{Condition 3: } z_j &\text{ is independent of } (n - k) \frac{s^2}{\sigma^2} \\ \implies t_j &= \frac{\hat{\beta}_j - \beta_j^0}{se(\hat{\beta}_j)} \sim t(n - k). \end{aligned} \tag{1}$$

^{*}jchen215@uic.edu

2. F statistic:

$$\begin{aligned}
 &\text{Condition 1: } W = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{\sigma^2} \sim \chi^2(J) \\
 &+ \text{Condition 2: } (n - k)\frac{s^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2(n - k) \\
 &+ \text{Condition 3: } W \text{ is independent of } (n - k)\frac{s^2}{\sigma^2} \\
 \implies F &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/J}{\mathbf{e}'\mathbf{e}/(n - k)} \sim F(J, n - k).
 \end{aligned} \tag{2}$$

Note that the W statistic comes from the Wald principle:

$$W = \mathbf{m}'[\text{Var}(\mathbf{m})]^{-1}\mathbf{m},$$

which measures the sum squared deviation relative to variance.

3. We also derive the F statistic for unknown σ in terms of residual sum of squares:

$$F = \frac{(RRSS - URSS)/\# \text{ restrictions}}{URSS/(n - k)}. \tag{3}$$

Note that the F statistic here can be considered as a result based on the Likelihood ratio principle. For known σ , we can write

$$F = \frac{(RRSS - URSS)}{\sigma^2}. \tag{4}$$

2 More on the Wald Principle

To simplify notations, we will tell both the Wald principle and the Likelihood ratio principle based on the model space $\text{col}(\mathbf{X})$, though every thing in the model space can be carried to the parameter space where $\boldsymbol{\beta}$ lives. We will also assume known σ .

Recall that the original GM model (with the normal assumption) assumes

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}),$$

which gives the LS model vectors

$$\hat{\mathbf{y}} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}).$$

Given a sample vector \mathbf{y} , we can always perform the LS method and get a particular model vector $\hat{\mathbf{y}}$ based on the sample. Hypothesis testing, in some sense, tries to assess whether the sample appears to favor or disfavor a particular model. This is done by partition the model space into a restricted model space and an unrestricted one.

In this sense, the W statistic (with known σ) measures the squared distance between the unrestricted model vector $\hat{\mathbf{y}}_U$ and the restricted model vector $\hat{\mathbf{y}}_R$, standardized by the squared “average distance” between any two model vectors σ^2 :

$$\frac{\|\hat{\mathbf{y}}_U - \hat{\mathbf{y}}_R\|^2}{\sigma^2}.$$

Note that this is identical to

$$\frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}^0\|^2}{\sigma^2}, \quad (5)$$

if we tie the restricted model vector to the null hypothesis.

If the standardized squared distance (the W value) is too large, we would consider the sample vector \mathbf{y} contains evidence against the restricted model (or the null hypothesis). (*Visualize it!*) Distribution theory helps us to quantify how large is too large.

We can Link the squared distance with the long expression:

$$\begin{aligned} & \|\hat{\mathbf{y}} - \hat{\mathbf{y}}^0\|^2 \\ &= \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_{\mathbf{X}'\mathbf{X}}^2 \\ &= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)' \mathbf{X}'\mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \\ &= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)' \mathbf{R}'(\mathbf{R}')^{-1} \mathbf{X}'\mathbf{X} \mathbf{R}^{-1} \mathbf{R} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \\ &= (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}^0)' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'] (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}^0) \\ &= (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'] (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}), \end{aligned} \quad (6)$$

which is the numerator of the W statistic.

3 More on the Likelihood Ratio Principle

While the Wald principle measures the distance between the null and the alternative (or the restricted and the unrestricted), the Likelihood ratio principle measures the distance between the likelihood of the null and the likelihood of the alternative. In the session we present the likelihood contour to visualize this distance. To complete that intuition, we derive the likelihood ratio test statistic here.

Begin with the likelihood function of the model vector

$$L(\hat{\mathbf{y}}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{2\sigma^2}\right],$$

its log-likelihood function is

$$\ell(\hat{\mathbf{y}}) = K - \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{2\sigma^2}.$$

The likelihood ratio test statistic can be constructed in the following way:

$$\begin{aligned} 2(L(\hat{\mathbf{y}})/L(\hat{\mathbf{y}}^0)) &= 2[\ell(\mathbf{y}) - \ell(\mathbf{y}^0)] \\ &= 2\left[K - \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{2\sigma^2} - K + \frac{\|\mathbf{y} - \hat{\mathbf{y}}^0\|^2}{2\sigma^2}\right] \\ &= \frac{\|\mathbf{y} - \hat{\mathbf{y}}^0\|^2 - \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\sigma^2}. \end{aligned} \tag{7}$$

Note that this is identical to equation (4). Moreover, it equals the same distance measured by the W statistic in equation (5). (*Visualize it!*)