Linear Models II

May 1-6, 2008

Theorem 1. *Assume that* $\beta \in \mathbb{R}^k$ *and that* X *is a* $n \times k$ *matrix of rank* $k < n$. *Let* $Y_1, \ldots, Y_n$ *are independent normally distributed random variables with mean vector* $\mu = X\beta$. *Then, the likelihood ratio test of the hypothesis*

$$H_0 : A\beta = 0 \quad \text{versus} \quad H_1 : A\beta \neq 0.$$

where A *is a* $q \times k$ *matrix has critical region*

$$C = \{\mathbf{y}; F(\mathbf{y}) \geq F_0\}.$$

F *is given by*

$$F(\mathbf{y}) = \frac{\sum_{i=1}^n (y_i - \widehat{\widehat{\mu}}_i)^2 - \sum_{i=1}^n (y_i - \widehat{\mu}_i)^2}{\sum_{i=1}^n (y_i - \widehat{\mu}_i)^2} \frac{n-k}{q}. \tag{1}$$

For the expression

$$\sum_{i=1}^n (y_i - \mu_i)^2,$$

- *the vector* $\widehat{\mu}$ *is the minimum value under the restriction* $\mu = X\beta$, *and*
- *The vector* $\widehat{\widehat{\mu}}$ *is the minimum value under the pair of restrictions* $\mu = X\beta$ *and* $A\beta = 0$.

Proof. The likelihood function

$$L(\beta, \sigma^2|\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2} (\mathbf{y} - \mu)^T (\mathbf{y} - \mu)$$

The likelihood ratio

$$\Lambda(\mathbf{x}, \mathbf{y}) = \frac{\sup\{L(\beta, \sigma^2|\mathbf{x}, \mathbf{y}); \mathbf{y} = X\beta, A\beta = 0\}}{\sup\{L(\beta, \sigma^2|\mathbf{x}, \mathbf{y}); \mathbf{y} = X\beta\}}$$

For the numerator, let $\hat{\beta}$ be the maximum likelihood estimator for the parameter β and let $\widehat{\mu} = X\hat{\beta}$. Then, the the maximum likelihood estimator for σ^2 is

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{\mu}_i)^2 = \frac{1}{n} (\mathbf{y} - \widehat{\mu})^T (\mathbf{y} - \widehat{\mu})$$

Therefore

$$L(\hat{\beta}, \widehat{\sigma^2}|\mathbf{x}, \mathbf{y}) = \frac{\exp -\frac{n}{2}}{(2\pi\widehat{\sigma^2})^{n/2}}.$$

Similarly, for the denominator, let $\widehat{\widehat{\mu}} = X\hat{\hat{\beta}}$ and $\widehat{\widehat{\sigma^2}}$ be the corresponding maximum likelihood estimates when the null hypothesis is true. Then,

$$L(\hat{\hat{\beta}}, \widehat{\widehat{\sigma^2}}|\mathbf{x}, \mathbf{y}) = \frac{\exp -\frac{n}{2}}{(2\pi\widehat{\widehat{\sigma^2}})^{n/2}}.$$

Consequently, the likelihood ratio test,

$$\lambda_0 \geq \Lambda(\mathbf{x}, \mathbf{y}) = \left(\frac{\widehat{\sigma^2}}{\widehat{\widehat{\sigma^2}}}\right)^n = \left(\frac{(\mathbf{y} - \hat{\mu})^T(\mathbf{y} - \hat{\mu})}{(\mathbf{y} - \hat{\hat{\mu}})^T(\mathbf{y} - \hat{\hat{\mu}})}\right)^n.$$

Now, a little arithmetic gives the expression in (1). □

As we saw with the t-test, the major issue is the ability to compute the density function for $F(\mathbf{y})$ under the null hypothesis.

We will now show that the F statistic is a constant time the ratio of χ^2 random variables.

Property 1. Let $\xi_1, \ldots, \xi_k$ be the columns of X. Then these vectors are linearly independent. In other words, $\mathcal{L}$, the span of $\xi_1, \ldots, \xi_k$ has dimension k.

This follows from the assumption that X has rank k.

Property 2. $\mu \in \mathcal{L}$.

This follows from $\mu = X\beta = \beta_1\xi_1 + \cdots + \beta_n\xi_n$.

Property 3. When H_0 holds, $\mu = \tilde{\beta}_1\eta_1 + \cdots + \tilde{\beta}_{k-q}\eta_{k-q}$ where $\eta_i \in \mathcal{L}$ and $\tilde{\beta}_i$ is one of the components of β. Call $\mathcal{L}_\prime$ the linear span of the independent vectors $\eta_1, \ldots, \eta_{k-q}$.

The restriction $A\beta = 0$ results in q independent homogenous linear restrictions on the β_i's. Use this to eliminate q of the components of β and let the η_i be the linear combination of the ξ_i resulting from this elimination.

Property 4. $\mathcal{L}_0 \subset \mathcal{L}$. Choose an orthonormal basis $\alpha_1, \ldots, \alpha_n$ so that $\alpha_1, \ldots, \alpha_k$ is an orthonormal basis for for $\mathcal{L}$ so that $\alpha_1, \ldots, \alpha_{k-q}$ is an orthonormal basis for $\mathcal{L}_0$. Let P be a matrix whose i-th row is α_i. Then $PP^T = I$, the identity matrix.

$$(PP^T)_{ij} = \alpha_i^T\alpha_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Property 5. $\mathbf{y} = \sum_{j=1}^n z_j\alpha_j$ for some scalars $z_1, \ldots, z_n$ and $\mu = \sum_{j=1}^k \nu_j\alpha_j$.

The follows because $\alpha_1, \ldots, \alpha_n$ is a basis, $\mu \in \mathcal{L}$ and $\alpha_1, \ldots, \alpha_k$ is a basis for $\mathcal{L}$.

Property 6. $(\mathbf{y} - \mu)^T(\mathbf{y} - \mu) = \sum_{j=1}^k (z_j - \nu_j)^2 + \sum_{j=k+1}^n z_j^2$.

$$\mathbf{y} - \mu = \sum_{j=1}^k (z_j - \nu_j)\alpha_j + \sum_{j=k+1}^n z_j\alpha_j$$

and $(\mathbf{y} - \mu)^T(\mathbf{y} - \mu)$ is the square of the norm of this vector.

Property 7. $\hat{\mu} = \sum_{j=1}^{k} z_j \alpha_j$ and $(\mathbf{y} - \hat{\mu})^T(\mathbf{y} - \hat{\mu}) = \sum_{j=k+1}^{n} z_j^2$.

To minimize $(\mathbf{y} - \mu)^T(\mathbf{y} - \mu)$ over all $\mu \in \mathcal{L}$, make the choice $\nu_j = z_j$.

Property 8. $\hat{\hat{\mu}} = \sum_{j=1}^{k-q} z_j \alpha_j$ and $(\mathbf{y} - \hat{\mu})^T(\mathbf{y} - \hat{\mu}) = \sum_{j=k-q+1}^{n} z_j^2$

This uses an argument very similar to checking Property 7. Combining we find:

Property 9.

$$F = \frac{\sum_{j=k-q+1}^{n} z_j^2 - \sum_{j=k+1}^{n} z_j^2}{\sum_{j=k+1}^{n} z_j^2} \frac{n-k}{q} = \frac{\sum_{j=k-q+1}^{k} z_j^2}{\sum_{j=k+1}^{n} z_j^2} \frac{n-k}{q}.$$

Property 10. Let $Z = PY$, then $Z = (Z_1 \ldots, Z_n)$ are n independent normal random variables with mean $\nu = P\mu$ and variance σ^2.

For a *linear* transformation, the Jacobian is simply the determinant. Thus, the density

$$f_Z(\mathbf{z}) = |\det(P^{-1})| f_Y(P^{-1}\mathbf{z}).$$

Because $PP^T = I$, $P^{-1} = P^T$ and $|\det(P^{-1})| = |\det(P)| = 1$. Recall that $Y_1, \ldots, Y_n$ are n independent normal random variables with mean μ and variance σ^2. Therefore,

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2}(\mathbf{y} - \mu)^T(\mathbf{y} - \mu).$$

Also,

$$(\mathbf{y} - \mu)^T(\mathbf{y} - \mu) = (P^T\mathbf{z} - P^T\nu)^T(P^T\mathbf{z} - P^T\nu) = (\mathbf{z} - \nu)PP^T(\mathbf{z} - \nu) = (\mathbf{z} - \nu)(\mathbf{z} - \nu).$$

Property 11.

$$\frac{\sum_{j=k-q+1}^{k} Z_j^2}{\sum_{j=k+1}^{n} Z_j^2}$$

is the ratio of independent χ^2 random variables. The numerator has q degrees of freedom and the denominator has $n - k$ degrees of freedom.

Use Property 10 and divide both the numerator and denominator by σ^2. The F statistic is

$$\frac{\sum_{j=k-q+1}^{k} Z_j^2/q}{\sum_{j=k+1}^{n} Z_j^2/(n-k)}$$

Recall that a χ^2_ℓ random variable has mean ℓ. The the constant factor was chosen so that the numerator an denominator each has mean 1.

Property 12. The F statistic can be written in the more compact form

$$F \frac{\sum_{i=1}^{b} (\hat{\mu}_i - \hat{\hat{\mu}}_i)^2}{\sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2} \frac{n-k}{q}.$$

In this form, we see that the F statistic is the ratio of the between group variance and the within group variance. If this is significantly large, then the between group variance dominates and we reject the null hypothesis.

The F distribution is said to have two degrees of freedom. for the numerator and n_2 for the denominator. The density is found using very much the same strategy in finding the density for the t-distribution

$$f_{n_1,n_2}(x) = \frac{1}{\mathrm{B}(n_1/2, n_2/2)} \left(\frac{n_1\, x}{n_1\, x + n_2}\right)^{n_1/2} \left(1 - \frac{n_1\, x}{n_1\, x + n_2}\right)^{n_2/2} x^{-1}.$$

The beta function

$$\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt$$

is related to the gamma function through the identity,

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}.$$

1 One way analysis of variance

One of the most common uses for the general linear model set up is analysis of variance (ANOVA). In the case the matrix X consists entirely of 0's and 1's. A 1 in position ij indicates that the observation belongs to group j. With a change in notation, we can examine the one variable case.

The data $\{y_{ij}, 1 \le i \le n_j, 1 \le j \le q\}$ represents that we have n_j observation for the j-th group and that we ℓ is one of k groups. The model is

$$y_{ij} = \mu_j + \epsilon_{ij}.$$

where ϵ_{ij} are independent $N(0, \sigma^2)$ random variables with σ^2 unknown. The total number of observations $n = n_1 + \cdots + n_q$. The hypothesis is

$$H_0 : \text{all } \mu_i \text{ are equal} \quad \text{versus} \quad H_1 : \text{not all } \mu_i \text{ are equal}.$$

The within group mean

$$\bar{y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}.$$

is the maximum likelihood estimator $\widehat{\mu}_i$ for μ_i.

The overall mean

$$\bar{\bar{y}}_{\cdot\cdot} = \frac{1}{n} \sum_{j=1}^{q} \sum_{i=1}^{n_j} y_{ij} = \frac{1}{n} \sum_{j=1}^{q} n_j y_{\cdot j}$$

is the maximum likelihood estimator $\widehat{\widehat{\mu}}$

Write the total **sums of squares**

$$s^2_{\text{total}} = \sum_{j=1}^{q} \sum_{i=1}^{n_j} (y_{ij} - \bar{\bar{y}}_{\cdot\cdot})^2$$

then

$$\widehat{\sigma^2} = \frac{1}{n} s^2_{\text{total}}.$$

The interior sum can be written

$$\sum_{i=1}^{n_j} (y_{ij} - \overline{\overline{y}}_{\cdot\cdot})^2 = \sum_{i=1}^{n_j} (y_{ij} - \overline{y}_{\cdot j})^2 + n_j(\overline{y}_{\cdot j} - \overline{\overline{y}}_{\cdot\cdot})^2$$

yielding

$$s^2_{\text{total}} = s^2_{\text{residual}} + s^2_{\text{between}}$$

with

$$s^2_{\text{residual}} = \sum_{j=1}^{q} \sum_{i=1}^{n_j} (y_{ij} - \overline{y}_{\cdot j})^2 \quad \text{and} \quad s^2_{\text{between}} = \sum_{j=1}^{q} n_j(\overline{y}_{\cdot j} - \overline{\overline{y}}_{\cdot\cdot})^2.$$

This gives the general form for one-way analysis of variance.

source of variation	degrees of freedom	sums of squares	mean square
between samples	$q-1$	s^2_{between}	$s^2_{\text{between}}/(q-1)$
residuals	$n-q$	s^2_{residual}	$s^2_{\text{residual}}/(n-q)$
total	$n-1$	s^2_{total}	

The test statistic

$$F = \frac{s^2_{\text{between}}/(q-1)}{s^2_{\text{residual}}/(n-q)}.$$

has, under the null hypothesis, an F distribution with $q-1$ numerator degrees of freedom and $n-q$ denominator degrees of freedom.

Example 2 (Crash test data)**.** *The data on 20 cars from 4 different categories is summarized in the table below. The* **head impact criterion** *(hic) is the measurement*

	subcompact	compact	midsize	full size
n	5	5	5	5
$\bar{x}$	668.8	555.8	486.8	537.8
s	242.0	91.0	167.7	154.6

The analysis of variance

source of variation	degrees of freedom	sums of squares	mean square
between samples	3	88425	29475
residuals	16	475323	29708
total	19	563748	

The value of the F statistic is 0.992 and the P-value is 0.4216. The critical value for a 0.05 level test is 3.2389. So, we do not reject the null hypothesis that mean head impact criterion depends on car size.

```
> pf(0.992,3,16)
[1] 0.5783587
> qf(0.95,3,16)
[1] 3.238872
```