Linear models with applications in R

PUBHLTH 744: Handout 9(Testing)

Instructor: Andrea S. Foulkes

Division of Biostatistics and Epidemiology
UMass School of Public Health and Health Sciences

Fall 2007

Testing

- Nested linear models and the F test

Recall, we are generally interested in testing a linear model against a reduced model. We call our starting model the "full model" and consider whether a simpler, more parsimonious model is reasonable. That is, if our full model is given $Y = X\beta + \epsilon$ where $\epsilon \sim MVN_n(0, \sigma^2 I)$, we might consider the "reduced model" given by $Y = X_0\gamma_0 + \epsilon$ where $\epsilon \sim MVN_n(0, \sigma^2 I)$ and $C(X_0) \in C(X)$. For example in the regression setting, we may have the full model given by

$$E(Y) = X_1\beta_1 + X_2\beta_2$$

and the reduced model given by

$$E(Y) = X_2\beta_2$$

This corresponds to the null hypothesis $H_0 : \beta_1 = 0$. That is, our null hypothesis is $E(Y) \in C(X_0)$ and the (disjoint) alternative is $E(Y) \in C(X) \cap C(X_0)^C$. If we let M and M_0 be the orthogonal projection operators onto $C(X)$ and $C(X_0)$ respectively, then under the full model the UMVUE of $E(Y)$ is MY and under the reduced model the UMVUE of $E(Y)$ is M_0Y.
In the ANOVA setting, the full model is given by $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ where μ is the overall mean and α_i is the ith group (treatment) effect for $i = 1, \ldots, t$ and $j = 1, \ldots, n_i$. In general, we are interested in testing $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_t$. This corresponds to a reduced model given by $Y_{ij} = \mu + \epsilon_{ij}$.

Suppose the reduced model is true. This implies that MY and M_0Y are estimating the same quantity, $E(Y)$. On the other hand, if MY and M_0Y are different then they are not estimating the same quantity and the models must differ and so the reduced model is not correct. That is, the correctness of the reduced model depends on the quantity $(M - M_0)Y$. We measure this by the squared length given by $Y'(M - M_0)Y$.

Consider

$$\begin{aligned} E\left(\frac{Y'(M-M_0)Y}{r(M-M_0)}\right) &= tr\left(\frac{\sigma^2(M-M_0)}{r(M-M_0)}\right) + \frac{(X\beta)'(M-M_0)X\beta}{r(M-M_0)} \\ &= \sigma^2\left(\frac{r(M-M_0)}{r(M-M_0)}\right) + \frac{(X\beta)'(I-M_0)X\beta}{r(M-M_0)} \\ &= \sigma^2 + \frac{(X\beta)'(I-M_0)X\beta}{r(M-M_0)} \end{aligned}$$

If the reduced model is true, then we can replace $X\beta$ with $X_0\gamma$. But $(I-M_0)X_0\gamma = 0$ since $(I-M_0)$ is the orthogonal projection operator onto $C(X_0)^\perp$ and therefore $(I-M_0)X_0 = 0$.

This implies that under the reduced model,

$$E\left(\frac{Y'(M-M_0)Y)}{r(M-M_0)}\right) = \sigma^2$$

In other words, if the reduced model is correct, we expect

$$\left(\frac{Y'(M-M_0)Y)}{\sigma^2 r(M-M_0)}\right) = 1$$

Replacing σ^2 with the UMVUE of σ^2 from the full model (note: we always work under the assumption that the full model is true), $\widehat{\sigma}^2 = MSE = \frac{Y'(I-M)Y}{r(I-M)}$, gives us

$$\left(\frac{Y'(M - M_0)Y)}{MSEr(M - M_0)} \right) = 1$$

Suppose $r(M) = r$, $r(M_0) = r_0$ and $r(I) = n$, then this is

$$\left(\frac{Y'(M - M_0)Y)/(r - r_0)}{Y'(I - M)Y/(n - r)} \right)$$

Theorem

$$F = \left(\frac{Y'(M - M_0)Y)/(r - r_0)}{Y'(I - M)Y/(n - r)} \right) \sim F_{(r-r_0),(n-r),\gamma}$$

where $\gamma = (X\beta)'(I - M_0)X\beta/(2\sigma^2)$. If the reduced model is correct, then $\gamma = 0$.
Pf: Need to show (i) the quadratic form in the numerator has a non-central chi-square distribution (ii) the denominator has a central chi-square distribution and (iii) the two quadratic forms are independent.

Definition: An α level F-test of the hypothesis

$$H_0 : E(Y) \in C(X_0)$$
$$H_A : E(Y) \in C(X) \cap C(X_0)^c$$

will reject H_0 if

$$F = \left(\frac{Y'(M - M_0)Y)/(r - r_0)}{Y'(I - M)Y/(n - r)} \right) > F_{1-\alpha,(r-r_0),(n-r)}$$

where $F_{1-\alpha,(r-r_0),(n-r)}$ is the $(1-\alpha) * 100\%$ percentile of a doubly central F distribution with $(r - r_0), (n - r)$ degrees of freedom.

- Linear parametric functions

Now consider tests of hypotheses with linear constraints on the parameters. Again consider the usual linear model $Y = X\beta + \epsilon$ where $\epsilon \sim MVN_n(0, \sigma^2 I)$. Suppose we are interested in testing a linear combination of the parameters $\Lambda'\beta = 0$ where $\Lambda' = P'X$. In this case we write the null and alternative hypotheses as follows:

$$H_0 : E(Y) \in C(X) \text{ and } P'X\beta = 0$$
$$H_A : E(Y) \in C(X) \text{ and } P'X\beta \neq 0$$

We need to find the reduced model design matrix X_0 that corresponds to the constraint $P'X\beta = 0$. Note that $P'X\beta = 0$ is equivalent to stating $X\beta \in N(P') = C(P)^{\perp}$. So we can write our null hypothesis as

$$H_0 : E(Y) \in C(X) \cap C(P)^{\perp}$$

Now we need to find a matrix X_0 such that $C(X_0) = C(X_0) \cap C(P)^{\perp}$ since then we could implement F-test as before. Note first that we can write $P = MP + (I - M)P$ where M is the orthogonal projection operator onto $C(X)$. Therefore, $P'X\beta = P'MX\beta + P'(I - M)X\beta = P'MX\beta$. This implies that $P'X\beta = 0$ if and only if $P'MX\beta = 0$ or equivalently, $E(Y) \perp C(P)$ if and only if $E(Y) \perp C(MP)$. So we can rewrite our null hypothesis as

$$H_0 : E(Y) \in C(X) \cap C(MP)^{\perp}$$

Theorem
$C(X) \cap C(MP)^{\perp} = C((I - M_{MP})X) = C((M - M_{MP})X)$
where M_{MP} is the orthogonal projection operator onto $C(MP)$ and is given by $M_{MP} = MP(P'MP)^{-}P'M$.
So in the construction of the F-test, we can replace X_0 with $M - M_{MP}$. This gives us
$Y'(M - M_0)Y = Y'(M - (M - M_{MP})Y = Y'M_{MP}Y$ and

$$F = \left(\frac{Y'M_{MP}Y)/r(M_{MP})}{Y'(I - M)Y/(n - r)} \right) \sim F_{r(M_{MP}),(n-r),\gamma}$$

where $\gamma = (X\beta)'M_{MP}X\beta/2\sigma^2$.

Note that this can also be written in terms of Λ:

$$\begin{aligned} Y' M_{MP} Y &= Y' MP(P'MP)^- P'MY \\ &= (X\widehat{\beta})' P(P'X(X'X)^- X'P)^- P'X\widehat{\beta} \\ &= \widehat{\beta}' \Lambda(\Lambda'(X'X)^- \Lambda)^- \Lambda' \widehat{\beta} \end{aligned}$$

and the rank of M_{MP} is equal to the rank of Λ.

Therefore, we have

$$F = \frac{\widehat{\beta}'\Lambda(\Lambda'(X'X)^{-}\Lambda)^{-}\Lambda'\widehat{\beta}/r(\Lambda)}{MSE}$$

and $\gamma = \beta'\Lambda(\Lambda'(X'X)^{-}\Lambda)^{-}\Lambda'\beta/2\sigma^2$.

- Generalized test procedure

Consider the null $H_0 : \Lambda'\beta = d$ where $\Lambda' = P'_{s\times n}X_{n\times p}$ and d is a known $s \times 1$ vector. We aim to derive the F-test for this null. For example, consider the ANOVA model, $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ where μ is the overall mean and α_i is the ith group (treatment) effect for $i = 1, \ldots, t$ and $j = 1, \ldots, n_i$. We may be interested in testing $H_0 : \mu + \alpha_1 = \alpha_0$.

Suppose b is any solution to the equation $\Lambda'\beta = d$ so that $P'Xb = d$. Our null can then be written

$$H_0 : Y = X\beta + \epsilon \text{ and } P'X\beta - P'Xb = 0$$

If we let $\beta^* = \beta - b$, this can be rewritten as

$$H_0 : Y^* = Y - Xb = X\beta^* + \epsilon \text{ and } P'X\beta^* = 0$$

But this is equivalent to
$H_0 : E(Y - Xb) \in C(X)$ and $X\beta^* \in C(P)^\perp$. Note that $X\beta^* = E(Y - Xb)$ so our null reduces to $H_0 : E(Y - Xb) \in C(X) \cap C(P)^\perp$. We described previously that this is equivalent to $H_0 : E(Y - Xb) \in C(X) \cap C(MP)^\perp$.
So we write the reduced model as $Y - Xb = X_0\gamma_0 + \epsilon$ where $X_0 = M - M_{MP}$ and the F test is given by

$$F = \frac{(Y - Xb)' M_{MP}(Y - Xb))/r(M_{MP})}{(Y - Xb)'(I - M)(Y - Xb)/(n - r)} \sim F_{r(M_{MP}),(n-r),\gamma}$$

Note that the denominator can be reduced as follows:

$$
\begin{aligned}
&(Y - Xb)'(I - M)(Y - Xb) \\
&= Y'(I - M)Y - b'X'(I - M)Y - Y'(I - M)Xb + b'X'(I - M)Xb \\
&= Y'(I - M)Y
\end{aligned}
$$

Furthermore, the numerator can be rewritten:

$$
\begin{aligned}
&(Y - Xb)'M_{MP}(Y - Xb)) \\
&= (Y - Xb)'MP(P'MP)^{-}P'M(Y - Xb) \\
&= (P'MY - P'Xb)'(P'MP)^{-}(P'MY - P'Xb) \\
&= (P'MY - P'Xb)'(P'X(X'X)^{-}X'P)^{-}(P'MY - P'Xb) \\
&= (\Lambda'\widehat{\beta} - P'Xb)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\widehat{\beta} - P'Xb) \\
&= (\Lambda'\widehat{\beta} - d)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\widehat{\beta} - d)
\end{aligned}
$$

Therefore, the F test can written equivalently as

$$F = \frac{(\Lambda'\widehat{\beta} - d)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\widehat{\beta} - d)/r(M_{MP})}{MSE} \sim F_{r(M_{MP}),(n-r),\gamma}$$

Notably, this statistic is invariant to choice of b.

- Confidence regions

Consider the estimable system of equations $\Lambda'\beta$ where $\Lambda' = P'X$. Using the results of the section on testing, we can construct a joint confidence region. For example, we may be interested in constructing a confidence interval around a single linear equation of the parameters such as $\lambda'\beta = \beta_1 - \beta_2$ where $\lambda' = (1\ -1\ 0\ \ldots\ 0)$. Or we may want to construct a region around the system of linear combinations $\Lambda'\beta = \begin{pmatrix} \beta_1 - \beta_2 \\ \beta_3 \end{pmatrix}$ where $\Lambda' = \begin{pmatrix} 1 & -1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}$.

Theorem: A $(1-\alpha)\times 100\%$ confidence region for $\Lambda'\beta$ is given by

$$\left\{\beta : \frac{(\Lambda'\widehat{\beta}-\Lambda'\beta)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\widehat{\beta}-\Lambda'\beta)/r(\Lambda)}{MSE} \le c_\alpha\right\}$$

where c_α is the upper $(1-\alpha)\times 100\%$ quantile of a central F distribution with $r(\Lambda)$ and $r(I-M)$ degrees of freedom.

If $X_{n \times p}$ is full rank, then all of the β's are estimable. In other words, $\Lambda'\beta$ is estimable where $\Lambda' = I_{p \times p}$. In this case, the above confidence region reduces to:

$$\left\{ \beta : \frac{(\widehat{\beta} - \beta)'(X'X)(\widehat{\beta} - \beta)}{p \times MSE} \leq c_\alpha \right\}$$

- Examples in simple and multiple regression

First, some notation. In the multiple regression setting, our model is given by $Y = X\beta + \epsilon$ and we generally assume $X_{n \times p}$ is full rank. We have the following:

$$
\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'Y \\
Cov(\widehat{\beta}) &= \sigma^2(X'X)^{-1} \\
SSR(X) &= Y'MY = \widehat{\beta}'X'X\widehat{\beta} \\
SSE &= Y'(I - M)Y \\
dfE &= r(I - M) = n - p
\end{aligned}
$$

In the usual regression setting, the first column of X is given by a vector of 1's and the corresponding parameter, β_0 is the overall mean. The ANOVA table is then tabulated for testing the full model against the reduced model given by $Y = \beta_0 + \epsilon$. Note that the reduced model design matrix is $X_0 = J_n$ and so $M_0 = X_0(X_0'X_0)^{-1}X_0' = \frac{1}{n}J_n^n$.

The ANOVA table is given by:

ANOVA

Source	df	SS
β_0	1	$Y'(\frac{1}{n}J_n^n)Y$
Regression	$p-1$	$Y'(M-\frac{1}{n}J_n^n)Y$
Error	$n-p$	$Y'(I-M)Y$
Total	n	$Y'Y$

- Likelihood ratio test

Theorem: The F-test is equivalent to the Likelihood Ratio Test for the hypothesis $H_0 : E(Y) \in C(X_0)$ where $X_0 \in X$ under the model $Y = X\beta + \epsilon$ and $\epsilon \sim MVN(0, \sigma^2 I)$.

Proof:

$$\lambda(y) = \frac{sup_{\gamma,\sigma} L(\gamma, \sigma|Y)}{sup_{\beta,\sigma} L(\beta, \sigma|Y)}$$

$$= \frac{(2\pi)^{-n/2}\widehat{\sigma}_0^{-n} exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}(Y - X_0\widehat{\gamma})'(Y - X_0\widehat{\gamma})\right\}}{(2\pi)^{-n/2}\widehat{\sigma}^{-n} exp\left\{-\frac{1}{2\widehat{\sigma}^2}(Y - X\widehat{\beta})'(Y - X\widehat{\beta})\right\}}$$

$$= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2} \frac{exp\left\{-\frac{n}{2Y'(I-M_0)Y}(Y - M_0Y)'(Y - M_0Y)\right\}}{exp\left\{-\frac{n}{2Y'(I-M)Y}(Y - MY)'(Y - MY)\right\}}$$

$$= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2} \frac{exp\left\{-\frac{n}{2Y'(I-M_0)Y}Y'(I - M_0)Y\right\}}{exp\left\{-\frac{n}{2Y'(I-M)Y}Y'(I - M)Y\right\}}$$

$$= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2}$$

We reject the null hypothesis if this quantity, $\lambda(y)$ is less than or equal to a critical value, c:

$$\left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2} \leq c$$

$$\Leftrightarrow \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right) \leq c^{2/n} = c_1$$

$$\Leftrightarrow \left(\frac{Y'(I-M)Y}{Y'(I-M)Y + Y'(M-M_0)Y}\right) \leq c_1$$

$$\Leftrightarrow \left(\frac{Y'(I-M)Y + Y'(M-M_0)Y}{Y'(I-M)Y}\right) \geq 1/c_1 = c_2$$

$$\Leftrightarrow \left(1 + \frac{Y'(M-M_0)Y}{Y'(I-M)Y}\right) \geq c_2$$

$$\Leftrightarrow \left(\frac{Y'(M-M_0)Y}{Y'(I-M)Y}\right) \geq c_2 - 1 = c_3$$

$$\Leftrightarrow \left(\frac{Y'(M-M_0)Y/r(M-M_0)}{Y'(I-M)Y/r(I-M)}\right) \geq c_4$$

This final criterion is of the same form as the F test.