Testing in the multiple regression model

Susan Thomas
IGIDR, Bombay

11 November, 2008

Recap on jargon

- Level of significance (α)
- Confidence level ($1 - \alpha$)
- Critical value
- t-stats
- prob-value
- Standard null, $H_0 : \beta_1 = 0$; $H_A : \beta_1 \neq 0$ (two-tailed tests)
- Accepting the Null –> Parameter is not significant at the given levl of significance.
  Rejecting the Null –> Parameter is significant

Recap: Tests of hypothesis

- Test of single coefficients.
  $H_0 : \beta_i = 0$; $H_A : \beta_i \neq 0$ (also could do a one-tailed test.)
  Also $H_0 : \beta_i = C$; $H_A : \beta_i \neq C$
  Distribution of $\hat{\beta}_i$ is $t(n-k)$ or $\mathrm{N}(\beta_i, \sigma^2_{\beta_i})$.

Recap: Tests of hypothesis

- Test involving more than a single parameter.
  1. Multiple parameters, one restriction
     $H_0 : \beta_1 = \beta_2, H_A : \beta_1 \neq \beta_2$
     Reformulate the test as $H_0 : \beta_1 - \beta_2 = \gamma = 0; H_A : \gamma \neq 0$
     where $E(\gamma) = f(E(\beta_1), E(\beta_2)), var(\gamma) = f(var(\beta_1), var(\beta_2), cov(\beta_1, \beta_2))$
  2. Multiple parameters, multiple restrictions
     $H_0 : \beta_1 = \beta_2; \beta_1 + \beta_2 + \beta_3 = 1; H_A : \text{Not } H_0$
     H_A is accepted if either or both of the restrictions are rejected.
     Tests used: Reformulate the LR test after *reparameterising the model*.

$$\frac{(RSS_R - RSS_U)/M}{RSS_U/(N-K)} \sim F(M, N-K)$$

where M is the number of restrictions and K is the number of parameters estimated in the unrestricted model.

Recap: Tests of hypothesis

- Test whether the model is significant.

$$H_0 : \quad \beta_1 = 0; \beta_2 = 0; \ldots \beta_k = 0;$$
$$H_A : \quad \text{Not } H_0 = \text{ only intercept term}$$

- Tests used: Same as the last case above.
  With $K + 1$ parameters estimated (including the intercept term), the number of restrictions in the "test of the model" $M = K$.

$$\frac{(RSS_R - RSS_U)/K}{RSS_U/(N - K - 1)} \sim F(K, N - K - 1)$$

Recap: Tests of hypothesis

- Test whether the model is significant.
- The test can be reframed in terms of the R^2 of the model as:

$$
\begin{aligned}
RSS_R &= \sum_i (Y_i - \beta_0)^2 = TSS \\
RSS_U &= \sum_i (Y_i - \beta_0 - \beta_1 X_{1i} - \ldots - \beta_K X_{Ki})^2 \\
&= TSS - ESS_U \\
RSS_R - RSS_u &= ESS_U \\
\frac{RSS_R - RSS_u/K}{RSS_U/(N-K-1)} &= \frac{RSS_R - RSS_u/K \; TSS}{RSS_U/(N-K-1) \; TSS} \\
&= \frac{R^2/K}{(1-R^2)/(N-K-1)} \sim F_{(K,N-K-1)}
\end{aligned}
$$

- This statistic is reported as “the model F-test” by software.

Reading estimation results: an example

Model 1

	Value	Std. Err.	t-stats	p-value
Intercept	4.79000	0.12	38.4	0.000
Educ	-0.05000	0.02	-2.52	0.012
$Educ^2$	0.00515	0.00079	6.53	0.000
$\hat{\sigma}$ = 0.7212	RSS = 2015.49	R^2 = 0.083	$\hat{l}$ = -4233	

Model 2

	Value	Std. Err.	t-stats	p-value
Intercept	1.98000	0.14	13.9	0.000
Educ	-0.03800	0.018	-2.17	0.030
$Educ^2$	0.0046	0.00070	6.55	0.000
hours worked	0.783	0.025	31.6	0.000
$\hat{\sigma}$ = 0.643	RSS = 1601.53	R^2 = 0.271	$\hat{l}$ = -3787	

Testing hypothesis: an example

- Suggested model: $Y_i = \alpha + \beta X_i + \gamma Z_i + \epsilon_i$
- Y_i is log wages, X_i is education.
- We are told that the sample has N_1 men and N_2 women, where

$$\begin{aligned} \text{log wage}_{\text{men}} &\sim N(\alpha_M + \beta \text{educ}, \sigma^2) \\ \text{log wage}_{\text{women}} &\sim N(\alpha_W + \beta \text{educ}, \sigma^2) \end{aligned}$$

1. How would you use the specified model to test $\alpha_M = \alpha_W$?
2. Would you use a one-sided or a two-sided test?

Testing hypothesis: an example

- The model for log wages has three parameters: $\alpha, \beta, \sigma^2 \rightarrow$,with $\alpha = \alpha_M$ for men, $\alpha = \alpha_W$ for women. Same coefficients except for the intercept.
- $H_0 : \alpha_M = \alpha_W$ can be rewritten as $H_0 : \alpha_M - \alpha_W = \gamma = 0$
- This means modifying the given model as:

$$\begin{aligned} X_i &= \text{education} \\ Z_i &= \left[\begin{array}{ll} 1 & \text{for Men } i \\ 0 & \text{for Women } i \end{array} \right] \\ \gamma &= \alpha_M - \alpha_W \end{aligned}$$

- This gives: $Y_i = \alpha_w + \beta X_i + \gamma Z_i + \epsilon_i$

$$\begin{aligned} Y_i &= \alpha_W + \beta X_i + \gamma \times 0 + \epsilon_i \rightarrow \text{ for Women} \\ &= \alpha_W + \beta X_i + \epsilon_i \\ Y_i &= \alpha_W + \beta X_i + \gamma \times 1 + \epsilon_i \rightarrow \text{ for Men} \\ &= \alpha_W + \beta X_i + \alpha_M - \alpha_W + \epsilon_i = \alpha_M + \beta X_i + \epsilon_i \end{aligned}$$

- The test will be $H_0 : \gamma = 0$ with critical value set as $t(N - 3)$