Questions and Answers on Hypothesis Testing and Confidence Intervals

L. Magee Fall, 2008

---

1. Using 25 observations and 5 regressors, including the constant term, a researcher estimates a linear regression model by OLS and finds

$$b_2 = 4.21 \quad , \quad \text{st.err. of } b_2 = 1.80$$
$$b_3 = -0.116 \quad , \quad \text{st.err. of } b_3 = 0.0388$$

Let t_{20df} be a random variable having a t distribution with $n - K = 25 - 5 = 20$ d.f.. Use the following probabilities:

Prob$(t_{20df} > 1.75) = .05$
Prob$(t_{20df} > 2.086) = .025$

For each of the pairs of hypotheses given below, use a t-test to accept or reject H_0 at the 5% significance level. Show your reasoning.

(a) $H_0 : \beta_2 = 0 \;\; ; \;\; H_a : \beta_2 < 0$
(b) $H_0 : \beta_2 = 0 \;\; ; \;\; H_a : \beta_2 \neq 0$
(c) $H_0 : \beta_2 = 0 \;\; ; \;\; H_a : \beta_2 > 0$
(d) $H_0 : \beta_3 = 0 \;\; ; \;\; H_a : \beta_3 < 0$
(e) $H_0 : \beta_3 = 0 \;\; ; \;\; H_a : \beta_3 \neq 0$
(f) $H_0 : \beta_3 = 0 \;\; ; \;\; H_a : \beta_3 > 0$
(g) $H_0 : \beta_2 = 1 \;\; ; \;\; H_a : \beta_2 \neq 1$

2. Suppose that you estimate the model

$$y = \beta_1 \iota + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \tag{1}$$

using 13 observations ($n = 13$), and compute the sum of squared residuals to be $SSE_1 = 126.1$. Then the following smaller models are estimated using the same 13 observations.

$$y = \beta_1 \iota + \beta_2 x_2 + \beta_3 x_3 + \epsilon, \quad \text{with } SSE_2 = 166.1 \tag{2}$$
$$y = \beta_1 \iota + \beta_2 x_2 + \beta_4 x_4 + \epsilon, \quad \text{with } SSE_3 = 204.4 \tag{3}$$
$$y = \beta_1 \iota + \beta_2 x_2 + \epsilon, \quad \text{with } SSE_4 = 221.4 \tag{4}$$

(a) Compute F statistics for each of the following

(i) $H_0 : \beta_3 = 0$; $H_a : \beta_3 \neq 0$

(ii) $H_0 : \beta_4 = 0$; $H_a : \beta_4 \neq 0$

(iii) $H_0 : \beta_4 = 0$; $H_a : \beta_4 \neq 0$ as in part (ii), but this time impose the restriction $\beta_3 = 0$ in both H_0 and H_a.

(iv) $H_0 : \beta_3 = 0$ and $\beta_4 = 0$; $H_a : \beta_3 \neq 0$ and/or $\beta_4 \neq 0$

Parts (b) and (c) require the probabilities:

$\text{Prob}(t_{9df} > 2.262) = .025$
$\text{Prob}(t_{10df} > 2.228) = .025$
$\text{Prob}(t_{9df} > 1.383) = .10$
$\text{Prob}(t_{9df} > 1.833) = .05$

(b) What are the absolute values of the t statistics for testing the hypotheses in parts (i), (ii) and (iii)? What would the t-test accept/reject conclusions be at the 5% level?

(c) Suppose the OLS estimate of β_4 in equation (1) above is $b_4 = 6.13$.

(i) Use some of the above results to obtain a 90% confidence interval for β_4. (Use the t-statistic from part (b) to get the standard error of b_4.)

(ii) Use some of the above results to test $H_0 : \beta_4 = 4$ against $H_a : \beta_4 > 4$ at the 10% significance level.

3. The model

$$y = \beta_1 \iota + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \tag{5}$$

is estimated by OLS using 15 observations ($n = 15$), and compute the sum of squared residuals to be $SSE_1 = 55$. Then the following smaller models are estimated using the same 15 observations.

$$y = \beta_1 \iota + \beta_2 x_2 + \beta_3 x_3 + \epsilon, \quad \text{with } SSE_2 = 75 \tag{6}$$
$$y = \beta_1 \iota + \beta_2 x_2 + \beta_4 x_4 + \epsilon, \quad \text{with } SSE_3 = 72 \tag{7}$$
$$y = \beta_1 \iota + \beta_2 x_2 + \epsilon, \quad \text{with } SSE_4 = 79 \tag{8}$$

(a) Using the above information, compute F statistics for each of the following hypotheses

(i) $H_0 : \beta_4 = 0$; $H_a : \beta_4 \neq 0$

(ii) $H_0 : \beta_4 = 0$; $H_a : \beta_4 \neq 0$, impose the restriction $\beta_3 = 0$ in both H_0 and H_a

(iii) $H_0 : \beta_3 = 0$ and $\beta_4 = 0$; $H_a : \beta_3 \neq 0$ and/or $\beta_4 \neq 0$

(b) Suppose you wish to test the restriction $\beta_2 + \beta_3 + \beta_4 = 1$.

(i) Define R and q in such a way that this restriction can be expressed in the form $R\beta - q = 0$.

(ii) Write a model that could be estimated by OLS to get the restricted sum of squared residuals, SSE_R, that could be used in an F test of this restriction.

4. Let $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_5 \end{bmatrix}$. State whether each of the following restrictions or sets of restrictions are linear or nonlinear. They are linear if an equivalent set of restrictions can be written in the form $R\beta - q = 0$ with matrix R and vector q being known constants. For the ones that are linear, give numerical values of R and q.

(a) $\beta_4^3 = 27$

(b) $\beta_2\beta_3 = 1$

(c) $\frac{\beta_2}{\beta_3} = 1$

(d) $\frac{\beta_2}{\beta_3} = \beta_4$

(e) $\beta_2 = \beta_3 = \beta_4$

5. (a) Explain what is meant when a coefficient estimate is said to be "statistically significant at the 5% significance level".

(b) Briefly describe a situation in which a coefficient estimate is statistically significant at the 5% level, but is not significant in a practical or economic sense.

6. A data set contains the following variables on individuals. They live in one of two cities, city #1 or city #2.

VARIABLE	DESCRIPTION
y	log of hourly wage
x	equals 1 if the individual is an immigrant and $x = 0$ otherwise
c	equals 1 if the individual lives in city #2, otherwise $c = 0$
$x \times c$	an interaction variable which equals one only if both x and c equal one.

The sums of squared residuals (SSEs) of four OLS regressions, each based on $n = 20$ observations, are

$$
\begin{aligned}
y &= b_1 + b_2 x + e_1, \qquad \text{with } \ SSE_1 = 59 & (9)\\
y &= \tilde{b}_1 + \tilde{b}_3 c + e_2, \qquad \text{with } \ SSE_2 = 54 & (10)\\
y &= b_1^* + b_2^* x + b_3^* c + e_3, \qquad \text{with } \ SSE_3 = 49 & (11)\\
y &= b_1^{**} + b_2^{**} x + b_3^{**} c + b_4^{**}(x \times c) + e_4, \quad \text{with } \ SSE_4 = 46 & (12)
\end{aligned}
$$

Compute F statistics that could be used for testing the following null hypotheses.

(a) City of residence has no effect on log wages, using model (11) as the unrestricted model.

(b) City of residence has no effect on log wages, using model (12) as the unrestricted model.

(c) Immigration status has no effect on log wages, using model (12) as the unrestricted model.

7. Using 100 observations and 10 regressors, including the constant term, a researcher estimates a linear regression model by OLS and finds

$$b_2 = 3.1 \quad , \quad \text{st.err. of } b_2 = 1.7$$

Let t_{90df} be a random variable having a t distribution with 90 d.f.. Use the following probabilities:

$\text{Prob}(t_{90df} > 1.66) = .05$
$\text{Prob}(t_{90df} > 1.99) = .025$

(a) For each of the pairs of hypotheses given below, use a t-test to accept or reject H_0 at the 5% significance level. Show your reasoning.

(i) $H_0 : \beta_2 = 0 \ ; \ H_a : \beta_2 < 0$

(ii) $H_0 : \beta_2 = 0 \ ; \ H_a : \beta_2 \neq 0$

(iii) $H_0 : \beta_2 = 0 \ ; \ H_a : \beta_2 > 0$

(iv) $H_0 : \beta_2 = 1 \ ; \ H_a : \beta_2 \neq 1$

(b) Which one of the four tests in part (a) has the smallest P-value?

(c) Compute the 95% confidence interval for β_2.

8. (13 marks: 9 for a, 4 for b)
Four regression models are estimated by OLS using 20 observations. The equations and their

sums of squared residuals are

$$
\begin{aligned}
y_i &= \beta_1 + \beta_2 w_i + \epsilon_i, & SSE_1 = 400 \qquad (13)\\
y_i &= \beta_1 + \beta_2 w_i + \beta_3 g_i + \epsilon_i, & SSE_2 = 360 \qquad (14)\\
y_i &= \beta_1 + \beta_2 w_i + \beta_3 g_i + \beta_4 (w_i g_i) + \epsilon_i, & SSE_3 = 350 \qquad (15)\\
y_i &= \beta_1 + \beta_2 w_i + \beta_3 h_i + \epsilon_i, & SSE_4 = 340 \qquad (16)
\end{aligned}
$$

In equations (2) and (3), $g_i = 1$ if $w_i > 10$ and $g_i = 0$ if $w_i \leq 10$.
In equation (4), $h_i = 0$ if $w_i < 10$ and $h_i = w_i - 10$ if $w_i \geq 10$. The β values can differ across models.

(a) Compute F statistics for testing the following pairs of hypotheses.

(i) H_0: The same linear relationship between y_i and w_i applies when $w_i \leq 10$ as when $w_i > 10$

H_a: The linear relationship between y_i and w_i jumps to a different level at $w_i = 10$, but the slope of the linear relationship is the same, regardless of whether $w_i \leq 10$ or $w_i > 10$

(ii) H_0: The same linear relationship between y_i and w_i applies when $w_i \leq 10$ as when $w_i > 10$

H_a: The linear relationship between y_i and w_i jumps to a different level at $w_i = 10$, and/or the slope of the linear relationship depends on whether $w_i \leq 10$ or $w_i > 10$

(iii) H_0: The same linear relationship between y_i and w_i applies when $w_i \leq 10$ as when $w_i > 10$

H_a: The slope of the linear relationship between y_i and w_i depends on whether $w_i \leq 10$ or $w_i > 10$, but these two lines join at $w_i = 10$

(b) Sketch four diagrams, one for each model, each with w_i on the horizontal axis and $E(y_i|w_i)$ on the vertical axis. On each one, draw an example of the type of relation between y_i and w_i that is allowed by that model.

Answers

1. (a) t = (4.21 - 0)/1.80 = 2.34

Reject if $t < -1.725$. Since $t > -1.725$, then <u>accept H_0</u> at the 5% significance level

(b) $t = 2.34$

Reject if $|t| > 2.086$. Since $t = 2.34$, then $|t| > 2.086$, so <u>reject H_0</u> at the 5% level

(c) $t = 2.34$

Reject if $t > 1.725$. Since $t > 1.725$, then <u>reject H_0</u> at the 5% level

(d) $t = (-0.116 - 0)/.0388 = -2.99$

Reject if $t < -1.725$. Since $t < -1.725$, then reject H_0 at the 5% level

(e) t = -2.99

Reject if $|t| > 2.086$. Since $t = -2.99$, then $|t| > 2.086$, so reject H_0 at 5% level

(f) t = -2.99

Reject if $t > 1.725$. Therefore accept H_0 at the 5% level

(g) $t = (4.21 - 1)/1.80 = 1.78$

Reject if $|t| > 2.086$. Since $t = 1.78$, then $|t| < 2.086$, so accept H_0 at the 5% level

2. (a) (i) $F = \frac{(SSE_3 - SSE_1)/1}{SSE_1/(13-4)} = \frac{(204.4-126.1)/1}{126.1/9} = 5.59$

(ii) $F = \frac{(SSE_2 - SSE_1)/1}{SSE_1/(13-4)} = \frac{(166.1-126.1)/1}{126.1/9} = 2.86$

(iii) Because of the restriction that $\beta_3 = 0$, the unrestricted model is now (11) and the restricted model is (12), so $F = \frac{(SSE_4 - SSE_3)/1}{SSE_3/(13-3)} = \frac{(221.4-204.4)/1}{204.4/10} = 0.83$

(iv) $F = \frac{(SSE_4 - SSE_1)/2}{SSE_1/(13-4)} = \frac{(221.4-126.1)/2}{126.1/9} = 3.40$

(b) When testing one restriction, the t statistic and the F statistic are related by $t^2 = F$. So from the F statistics in (a), we know: in (i) $|t| = 2.36$; in (ii) $|t| = 1.69$; and in (iii) $|t| = 0.91$.

Reject H_0 when $|t| > 2.262$ at the 5% level for a two-tailed test with 9 d.f. Reject H_0 when $|t| > 2.228$ at the 5% level for a two-tailed test with 10 d.f. So the test in (i) rejects H_0, and the tests in (ii) and (iii) accept H_0.

(c) (i) From (b)(ii), $|t| = 1.69$ for testing $H_0 : \beta_4 = 0$. Since $b_4 = 6.13$ and $t = (b_4 - 0)/[\text{st.err}(b_4)]$ then $1.69 = 6.13/[\text{st.err}(b_4)]$ so that $\text{st.err}(b_4) = 3.63$.

The 90% c.i. is $6.13 \pm t_{.05} \times 3.63 = 6.13 \pm 1.833 \times 3.63 = 6.13 \pm 6.65 = (-0.52, 12.78)$

(ii) $t = (b_4 - 4)/\text{st.err}(b_4) = (6.13 - 4)/3.63 = 0.59$.

Reject H_0 if $t > t_{.10}$ $\Rightarrow$ Reject if $t > 1.383$. Accept H_0 at the 10% level, since $t = 0.59 < 1.383$.

3. (a) Use $F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-K)}$

(i) Compare (5) and (6). $F = \frac{(75-55)/1}{55/(15-4)} = 4$

(ii) Compare (7) and (8). $F = \frac{(79-72)/1}{72/(15-3)} = 7/6$

(iii) Compare (5) and (8). $F = \frac{(79-55)/2}{55/(15-4)} = 2.4$

(b) (i) $R = \begin{bmatrix} 0 & 1 & 1 & 1 \end{bmatrix}$; $q = 1$

(ii) Substitute out one of the β's from the original model, then re-arrange the variables. There should be one less regressor since there is one restriction imposed. For example, replacing β_4 with $1-\beta_2-\beta_3$ gives

$$\begin{aligned} y &= \beta_1\iota + \beta_2 x_2 + \beta_3 x_3 + (1-\beta_2-\beta_3)x_4 + \epsilon \\ y - x_4 &= \beta_1\iota + \beta_2(x_2 - x_4) + \beta_3(x_3 - x_4) + \epsilon \end{aligned}$$

4. (a) linear. $R = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix}$; $q = 3$

(b) nonlinear

(c) linear. $R = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix}$; $q = 0$

(d) nonlinear

(e) linear. $R = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{bmatrix}$; $q = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

5. (a) Let the coefficient be β and its estimate be $\hat{\beta}$. "Statistically significant at the 5% significance level" means "The null hypothesis $H_0 : \beta = 0$ is rejected at the 5% significance level." Usually this means that $|\frac{\hat{\beta}}{\text{st.err.}(\hat{\beta})}| >$ (critical value), or P-value $< .05$, where P-value $=$ Prob$(|t| > |\frac{\hat{\beta}}{\text{st.err.}(\hat{\beta})}|)$ and t represents a random variable with a distribution that one hopes is close to the one that the test statistic would have when H_0 is true. (This is the "null distribution".)

(b) Suppose that if β really was equal to $\hat{\beta}$ in the model, the economic implications would be much the same as if $\beta = 0$. In that case, $\hat{\beta}$ is not significant in an economic sense. For example, suppose $\hat{\beta}$ is estimating an income elasticity and $\hat{\beta} = .01$. Such a $\hat{\beta}$ still could be statistically significant, because its standard error might also be very small. In other words, $|\frac{\hat{\beta}}{\text{st.err.}(\hat{\beta})}|$ might be large enough for statistical significance, even when $\hat{\beta}$ is small enough to be economically insignificant.

6. (a) Compare (11) with (9).
$F = \frac{(59-49)/1}{49/(20-3)} = 3.47$

(b) Compare (12) with (9).
$F = \frac{(59-46)/2}{46/(20-4)} = 2.26$

(c) Compare (12) with (10).
$F = \frac{(54-46)/2}{46/(20-4)} = 1.39$

7. (a) $t = 3.1/1.7 = 1.82$

(i) Reject if $t < -1.66$. Therefore, do not reject H_0.

(ii) Reject if $|t| > 1.99$. Therefore, do not reject H_0.

(iii) Reject if $t > 1.66$. Therefore, reject H_0.

(iv) Now $t = (3.1-1)/1.7 = 1.24$. Reject if $|t| > 1.99$. Therefore, do not reject H_0.

(b) The only test that rejected H_0 at the 5% level was the one in part (iii). So it is the only one with a P-value less than .05, and so it must have the smallest P-value of the four tests.

(c) $b_2 \pm t_{.025} \times \text{st.err.}(b_2) = 3.1 \pm 1.99 \times 1.7 = 3.1 \pm 3.383 = (-0.283, 6.483)$

8. (a) (i) Compare (1) and (2)

$$F = \frac{(400 - 360)/1}{360/(20 - 3)} = 1.89$$

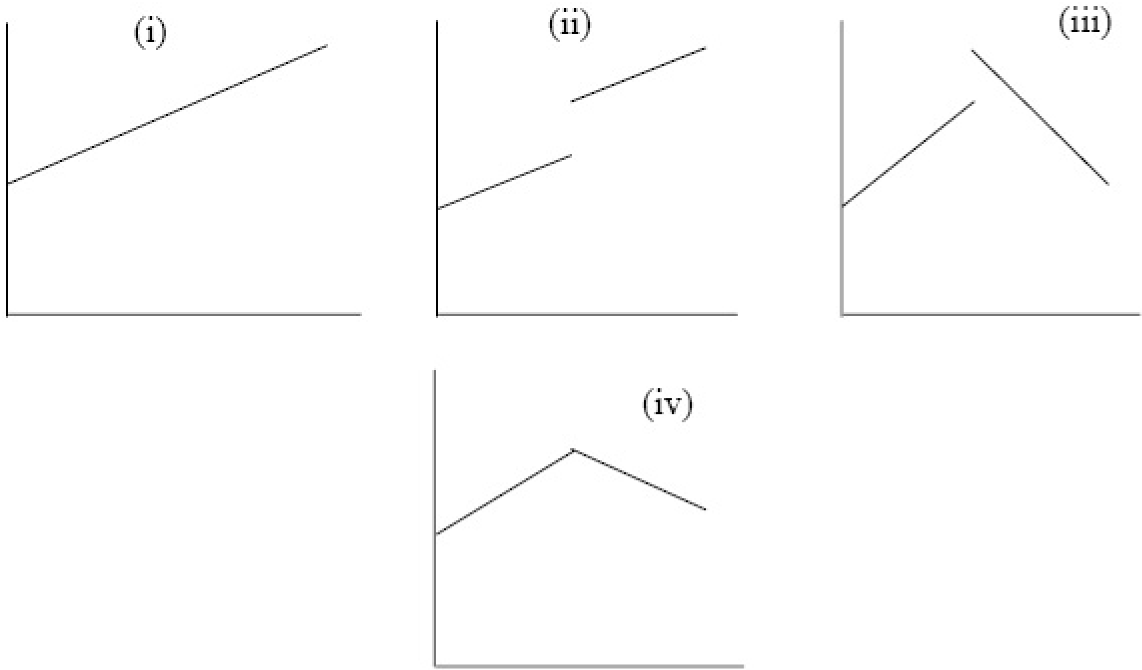
(ii) Compare (1) and (3)

$$F = \frac{(400 - 350)/2}{350/(20 - 4)} = 1.14$$

(iii) Compare (1) and (4)

$$F = \frac{(400 - 340)/1}{340/(20 - 3)} = 3.00$$

(b) see next page


(i)
(ii)
(iii)
(iv)