Inference

1. Which of the following can cause the usual OLS *t* statistics to be invalid (that is, to not have *t* distributions under H_0)?

- Heteroskedasticity.
- A sample correlation coefficient of .95 between two independent variables that are in the model.
- Omitting an important explanatory variable.

Answer: Since the *t* statistic is a ratio of the form $\hat{\beta}_j/se(\hat{\beta}_j)$, both heteroskedasticity and omitted variables can render it invalid.

2. Are rent rates influenced by the student population in a college town? Let *rent* be the average monthly rent paid on rental units in a college town in the United States. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percentage of the total population. One model to test for a relationship is

$$\log(rent) = \beta_0 + \beta_1 \log(pop) + \beta_2 \log(avginc) + \beta_3 pctstu + u.$$

(i) State the null hypothesis that size of the student body relative to the population has no ceteris paribus effect on monthly rents. State the alternative that there is an effect.

Answer: H_0: $\beta_3 = 0$ and H_1: $\beta_3 \neq 0$.

(ii) State the null hypothesis that the elasticity of *rent* with respect to *pop* (i.e., the percentage change in *rent* for a one percent change in *pop*) is 1. State the alternative that the elasticity is not 1.

Answer: H_0: $\beta_1 = 1$ and H_1: $\beta_1 \neq 1$.

3. Consider the estimated equation below studying the effects of skipping class on college GPA:

$$\widehat{colGPA} = \underset{(0.33)}{1.39} + \underset{(0.094)}{0.412}hsGPA + \underset{(0.011)}{0.015}ACT - \underset{(0.026)}{0.083}skipped$$
$$n = 141, \;\; R^2 = 0.234.$$

Here, *colGPA* and *hsGPA* depict college and high school GPA, respectively; *skipped* depicts the average number of lectures missed per week; and *ACT* denotes the standardized college admission test's score. The standard errors are in parentheses.

(i) Using the standard normal approximation, find the 95% confidence interval for β_{hsGPA} .

Answer: Standard normal approximation for 95% CI implies the two-tailed critical value for 5% level of significance from Table G.2 with $df = \infty$. The interval is 0.412 ± 1.96(0.094), or about 0.228 to 0.596.

(ii) Calculate the p-value for the hypothesis H_0: $\beta_{hsGPA} = 0$ against the two-sided alternative.

Answer: The test statistic is given by 0.412/0.094, i.e., 4.38. Now, since *n*-*k*-1 is equal to 137, we can assume the test statistic to follow a standard normal distribution. The p-value is equal to 2 x (area to the right of 4.38 under a standard normal curve). From Table G.1, the p-value is less than 0.001.

(iii) Given the confidence interval in (i), can you reject the hypothesis H$_0$: $\beta_{hsGPA} = 0.4$ against the two-sided alternative at the 5% level?

Answer: No, because the value 0.4 is inside the 95% CI.

Alternatively, the test statistic is (0.412 – 0.4)/0.094 = 0.128. This is inside the interval [-1.96, 1.96].

Alternatively, while the test statistic is (0.412 – 0.4)/0.094 = 0.13, the p-value from Table G.1 is roughly 2(1 – 0.5517) = 0.9. This is well above 0.05.

4. Use the data in MLB1 and estimate the following model:

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + u.$$

(i) Test the significance of *hrunsyr* at the 5% level of significance.

Answer: The variable *hrunsyr* is significant.

```
. reg lsalary years gamesyr bavg hrunsyr

      Source |       SS           df       MS      Number of obs   =       353
-------------+----------------------------------   F(4, 348)       =    145.24
       Model |  307.800674         4  76.9501684   Prob > F        =    0.0000
    Residual |  184.374861       348   .52981282   R-squared       =    0.6254
-------------+----------------------------------   Adj R-squared   =    0.6211
       Total |  492.175535       352  1.39822595   Root MSE        =    .72788

------------------------------------------------------------------------------
     lsalary | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       years |   .0677325   .0121128     5.59   0.000     .0439089     .091556
     gamesyr |   .0157595   .0015636    10.08   0.000     .0126841    .0188348
        bavg |   .0014185   .0010658     1.33   0.184    -.0006776    .0035147
     hrunsyr |   .0359434   .0072408     4.96   0.000     .0217021    .0501847
       _cons |   11.02091   .2657191    41.48   0.000     10.49829    11.54353
------------------------------------------------------------------------------
```

(ii) State the null hypothesis that the effect of *hrunsyr* on log(salary) is 0.08. Test this null hypothesis against a two-sided alternative at the 5% significance level.

Answer: H$_0$: $\beta_4 = 0.08$.

```
. lincom hrunsyr - 0.08

 ( 1)  hrunsyr = .08

------------------------------------------------------------------------------
     lsalary | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -.0440566   .0072408    -6.08   0.000    -.0582979   -.0298153
------------------------------------------------------------------------------


. test hrunsyr = 0.08

 ( 1)  hrunsyr = .08

       F(  1,   348) =   37.02
            Prob > F =    0.0000
```

We reject H$_0$.