ØAMET4100 · Spring 2019
Worksheet 3B

Instructor: Fenella Carpena

January 24, 2019

1 Hypothesis Testing in MRM: Single and Joint Tests

Suppose we have the following model that explains baseball players' salaries.

$$salary_i = \beta_0 + \beta_1 years_i + \beta_2 gamesyr_i + \beta_3 bavg_i + \beta_4 hrunsyr_i + \beta_5 rbisyr_i + u_i \tag{1}$$

where for each player i, *salary* is the salary in 1993, *years* is years in the league, *gamesyr* is average games played per year, *bavg* is the career batting average, *hrunsyr* is the number of home runs per year, *rbisyr* is runs batted in per year. Further, suppose that we estimated the above equation using data we have on hand, and that we obtained the following regression results

$$\widehat{salary} = \underset{(0.29)}{11.10} + \underset{(0.0121)}{0.0689} \cdot years + \underset{(0.0026)}{0.0126} \cdot gamesyr + \underset{(0.0010)}{0.00098} \cdot bavg + \underset{(0.0161)}{0.0144} \cdot hrunsyr + \underset{(0.0072)}{0.0108} \cdot rbisyr$$

$$N = 353,\ SSR = 183.186,\ R^2 = 0.6278$$

Exercise 1.1 What test statistic would we use to test the hypothesis $H_0 : \beta_3 = 0, H_1 : \beta_3 \neq 0$? Carry out this test at the 5% level.

Exercise 1.2 What test statistic would we use to test the hypothesis $H_0 : \beta_4 = 0, H_1 : \beta_4 \neq 0$? Carry out this test at the 5% level.

Exercise 1.3 What test statistic would we use to test the hypothesis $H_0 : \beta_5 = 0, H_1 : \beta_5 \neq 0$? Carry out this test at the 5% level.

Exercise 1.4 A sports analyst hypothesizes that once years in the league and games per year have been controlled for, the variables *bavg*, *hrunsyr*, and *rbisyr* (which we can think of as measure of performance) have no effect on salary.

(a) What is the null and alternative hypothesis?

(b) What test statistic would we use to test the hypothesis in part (a)? Assuming that the population errors are homoskedastic, what is the formula for this test statistic? Explain intuitively how this F-statistic works in terms of a hypothesis test.

(c) In the formula that you wrote down in part (b):

  (i) What is the restricted regression?

  (ii) What is the unrestricted regression?

  (iii) What is q?

  (iv) What is n?

  (v) What is k?

(d) Suppose that a regression of *salary* on *years* and *gamesyr* yielded an SSR of 198.311. Calculate the F-statistic under the assumption of homoskedasticity.

(e) Given a significance level of 10%, what is the critical value from the F-distribution.

(f) What is the conclusion of your hypothesis test?

Exercise 1.5 Consider again the joint hypothesis

$$H_0 : \beta_3 = 0,\ \beta_4 = 0,\ \beta_5 = 0\ vs.\ H_1 : \text{at least one of } \beta_3, \beta_4, \beta_5 \text{ is not equal to } 0$$

that we tested in the previous exercise using an F-test. Is it possible to carry out this joint hypothesis test using the three t-statistics from the following three separate tests: (1) $H_0 : \beta_3 = 0, H_1 : \beta_3 \neq 0$; (2) $H_0 : \beta_4 = 0, H_1 : \beta_4 \neq 0$; (3) $H_0 : \beta_5 = 0, H_1 : \beta_5 \neq 0$?

Exercise 1.6 We found that we rejected the joint hypothesis that *bavg*, *hrunsyr*, *rbisyr* have no effect on salary. But if we had considered a hypothesis test for each of these variables individually, we would have failed to reject each null hypothesis separately. What might explain the difference in these results?

2 Hypothesis Testing in MRM: Linear Restrictions

Exercise 2.1 Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$, and the hypothesis test $H_0 : \beta_1 = \beta_2$, $H_1 : \beta_1 \neq \beta_2$.

(a) What test statistics can we use to carry out the above hypothesis test?

(b) Describe how you would calculate the F-statistic (under the assumption of homoskedasticity). What is the restricted and unrestricted regression?

(c) Describe how you can use a t-statistic to test $H_0 : \beta_1 = \beta_2$, $H_1 : \beta_1 \neq \beta_2$. Specifically, transform the regression so that you can use a t-statistic to carry out the test.

Exercise 2.2 Consider again the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Transform the regression so that you can use a t-statistic to test $\beta_1 + 2\beta_2 = 0$.

3 Hypothesis Testing in MRM: Stata

Exercise 3.1 Suppose we have 1980 census data on the 50 states recording the population size in each state (`pop`), the median age (`medage`), the number of deaths (`death`), the number of marriages, (`marriage`), and the number of divorces (`divorce`). We estimate the following regression:

```
. reg pop medage death marriage divorce

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  4,    45) = 1299.46
       Model |  1.0800e+15     4  2.7000e+14           Prob > F      =  0.0000
    Residual |  9.3500e+12    45  2.0778e+11           R-squared     =  0.9914
-------------+------------------------------           Adj R-squared =  0.9907
       Total |  1.0893e+15    49  2.2232e+13           Root MSE      =  4.6e+05

------------------------------------------------------------------------------
         pop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      medage |  -181303.8   43749.97    -4.14   0.000    -269420.8   -93186.87
       death |   91.30243   4.137673    22.07   0.000     82.96873    99.63613
    marriage |    1.80206   4.303597     0.42   0.677    -6.865829    10.46995
     divorce |   39.80303   8.146704     4.89   0.000     23.39473    56.21134
       _cons |    5241295    1272002     4.12   0.000      2679350     7803239
------------------------------------------------------------------------------
```

What Stata commands would you use to test the following hypothesis?

(a) That the coefficient on `medage` is zero.

(b) That the coefficients on all four regressors are simultaneously zero.

(c) That the coefficient on `death` minus the coefficient on `marriage` is zero.

TABLE 5A Critical Values for the F_{n_1,n_2} Distribution—10% Significance Level

Denominator Degrees of Freedom (n_2)	Numerator Degrees of Freedom (n_1)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	39.86	49.50	53.59	55.83	57.24	58.20	58.90	59.44	59.86	60.20
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	**2.71**	**2.30**	**2.08**	**1.94**	**1.85**	**1.77**	**1.72**	**1.67**	**1.63**	**1.60**

This table contains the 90th percentile of the F_{n_1,n_2} distribution, which serves as the critical values for a test with a 10% significance level.