Giuseppe Ragusa
Department of Economics
Luiss Guido Carli

Econometrics

Midterm

Name: ____

(Please type clearly)

- Read carefully the questions and think before writing.
- When you do write, please do so in a clear way.
- Always, always, **explain** the reasoning behind your answer.
- If you need more space than the one allotted after the questions, please use the back pages.

1. Attendance at sports events depends on various factors. Teams typically do not change ticket prices from game to game to attract more spectators to less attractive games. However, there are other marketing tools used, such as fireworks, free hats, etc., for this purpose. You work as a consultant for a sports team, the Los Angeles Dodgers, to help them forecast attendance, so that they can potentially devise strategies for price discrimination. After collecting data over two years for every one of the 162 home games of the 2000 and 2001 season, you run the following regression:

$$
\begin{aligned}
\widehat{Attend} = &\underset{(8770)}{15005} + \underset{(121)}{201} \times Temperat + \underset{(169)}{465} \times DodgNetWin \\
&+ \underset{(26)}{82} \times OppNetWin + \underset{(1505)}{9647} \times DFSaSu + \underset{(3355)}{1328} \times Drain \\
&+ \underset{(1819)}{1609} \times D150m + \underset{(1184)}{271} \times DDiv - \underset{(1143)}{978} \times D2001; \\
R^2 = .416, \quad &SER = 6983
\end{aligned}
$$

where `Attend` is announced stadium attendance, `Temperat` it the average temperature on game day in Farenheit[1], `DodgNetWin` are the net wins of the Dodgers before the game (wins-losses), `OppNetWin` is the opposing teams net wins at the end of the previous season, and `DFSaSu`, `Drain`, `D150m`, `Ddiv`, and `D2001` are binary variables, taking a value of 1 if the game was played on a weekend, it rained during that day, the opposing team was within a 150 mile radius, the opposing team plays in the same division as the Dodgers, and the game was played during 2001, respectively. Numbers in parentheses are heteroskedasticity-robust standard errors.

(a) **What is the fraction of variance of attendance explained by the linear model?**

(b) **How many more fans will attend the game if temperature increases by 3 degrees Celsius? And if the game is played on week days?**

[1]If C denotes the temperature in Celsius and F in Farenheit, $F = 9/5 \times C + 32$.

(c) **Your employer wants to know whether attendance responds to the opponent win record. What is the statistic evidence you will report to your employer? What is the "economic" evidence?**

(d) **To test whether the effect of the last four binary variables is significant, you have your regression program calculate the relevant F-statistic, whose p-value is 0.079. Can you reject the null hypothesis at the 10% significance level? Can you reject it at the 1% significance level?**

(e) **(*) Construct a 90% confidence interval for the effect of a 17 degrees Celsius increase on attendance rate. Do you find this finding reasonable? How would you fix the model?**

2. The European Union Statistics on Income and Living Conditions (EU-SILC) collects income and demographic data on individuals of the European Union. You have obtained a sample of 13,596 observations for Italy and you intend to use it to study the determinant of wage in this country.

You postulate the following model:

$$hwage = \beta_0 + \beta_1 hischool_i + \beta_3 college_i + \beta_4 female_i + \beta_5 age_i + \beta_6 age_i^2 + u_i, \tag{1}$$

where:

- `hwage` is the hourly wage of individual i in Euro;
- `female` is 1 if the individual 1 is a female, 0 otherwise;
- `age` is the age of individual i;
- `lowhscl`, `hischool`, and `college` are dummy variables that are 1 if the maximum level of education individual i has achieved is: lower than high school diploma, high school diploma and/or some college, college degree and higher, respectively.

The estimated model is given by

$$\begin{aligned} hwage = \underset{(0.6137)}{2.3128} + \underset{(0.0916)}{3.3294} \times hischool + \underset{(0.1909)}{8.4807} \times college - \underset{(0.0953)}{1.4916} \times female \\ + \underset{(0.0328)}{0.3960} \times age - \underset{(0.0004)}{0.0020} \times I(age^2) \\ R^2 = 0.307 \end{aligned}$$

(a) **What is the effect (on the expected hourly wage) of earning an high school diploma?**

(b) **What is the interpretation of β_6 in this model?**

(c) **What is the predicted hourly wage for a 34 years old male with college degree?**

(d) **Can you test whether females and males earn the same hourly wage? (If you can, do it!)**

(e) **Suppose you also estimated the following model:**

$$lwage = \underset{(0.6736)}{0.2194} - \underset{(0.1049)}{0.6906} \times female + \underset{(0.0370)}{0.4318} \times age - \underset{(0.0005)}{0.0026} \times I(age^2), \quad R^2 = 0.132$$

Use this new information to test the following null hypothesis that "Education does not have an effect on hourly wages". (Be precise and explain what assumptions you are making).

(f) (*) Carefully explain whether your analysis in part d) can be taken as *serious* evidence that women are discriminated in Italy.

3. A random sample of 546 owners of single-family homes is drawn from the population of a city. Let the random variable Y denote annual household income, in dollars, and the random variable X denote the value of the house, also in dollars. The following information is available

$$n = 546$$

$$\sum_{i=1}^{n} Y_i = 37194394 \qquad \sum_{i=1}^{n} X_i = 7578578.3$$

$$\sqrt{\frac{1}{546}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 26702.67 \qquad \sqrt{\frac{1}{546}\sum_{i=1}^{n}(X_i - \bar{X})^2} = 2253.046$$

$$\rho_{xy} = \frac{\frac{1}{546}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{546}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\sqrt{\frac{1}{546}\sum_{i=1}^{n}(X_i - \bar{X})^2}} = 0.64$$

(a) **Construct a** 90% **confidence interval for the mean value of household income.**

(b) **Using a** 5% **significance level, test the hypothesis that household income is equal to** $\$12,000$ **against the alternative that household income is greater than** $\$12,000$**.**

(c) **Calculate the OLS coefficients of the regression of the value of the house on the household income.**