UNIVERSIDAD CARLOS III DE MADRID
ECONOMETRICS I
January, 30, 2007

PROBLEM TEXT

Very important: Notice that some results have been omitted from the tables.

PROBLEM 1: EFFECT OF A GARBAGE INCINERATOR ON HOUSING PRICES

After 1978, rumors about a garbage incinerator being built in North Andover (Massachusetts, USA) began. The construction was begun in 1981, and it was believed that the incinerator would be operating soon (although it did not start to operate until 1985).

The hypothesis is that the price of those homes closer to the incinerator fell with respect to the ones that were farther.

We have a sample of 321 observations about homes sold in 1978 (the first 179 observations) and in 1981 (the last 142 observations) for which we observe their selling prices, distance to the incinerator, and other characteristics. We can consider that the building and operation of the incinerator is a completely exogenous event with respect to the determination of the observed prices.

Using observations for the year 1981, the following model is estimated:

$$\text{RPRICE} = \gamma_0 + \gamma_1 \text{NEARINC} + \varepsilon$$

where:

RPRICE = home price in real terms (1978 dollars);

NEARINC = binary variable that equals 1 if the home is closer to the incinerator (less than 3 miles apart) and 0 otherwise.

OUTPUT 1: OLS estimates using the 142 observations $180 - 321$
Dependent variable: RPRICE

Variable	Coefficient	Std. Error	t-statistic	p-value
C	101308	3093	32.75	0.0000
NEARINC		5828	-5.27	0.0000

Mean of dependent variable	92663
S.D. of dependent variable	34071
Sum of squared residuals	1.36614e+11
Standard error of residuals ($\hat{\sigma}$)	31238
Unadjusted R^2	0.1653
Adjusted $\bar{R}^2$	0.1594

The same model is also estimated with observations for the year 1978:

OUTPUT 2: OLS estimates using the 179 observations $1-179$
Dependent variable: RPRICE

Variable	Coefficient	Std. Error	t-statistic	p-value
C	82517	2654	31.09	0.0000
NEARINC	−18824	4745		0.0001

Mean of dependent variable	76628
S.D. of dependent variable	30626
Sum of squared residuals	1.53324e+11
Standard error of residuals ($\hat{\sigma}$)	29432
Unadjusted R^2	0.0817
Adjusted $\bar{R}^2$	0.0765

Furthermore, the following model is estimated with observations for both years 1978 and 1981:

$$\text{RPRICE} = \alpha_0 + \alpha_1\text{NEARINC} + \alpha_2\text{Y81} + \alpha_3\text{Y81NEARINC} + \varepsilon$$

where:
Y81 = binary variable that equals 1 if the home was sold in 1981 and 0 otherwise;
Y81NEARINC =Y81 × NEARINC.

OUTPUT 3: OLS estimates using the 321 observations $1-321$
Dependent variable: RPRICE

Variable	Coefficient	Std. Error	t-statistic	p-value
C	82517	2727	30.26	0.0000
NEARINC	−18824	4875	−3.86	0.0001
Y81	18790	4050	4.64	0.0000
Y81NEARINC	−11890	7457		0.1126

Mean of dependent variable	83721
S.D. of dependent variable	33119
Sum of squared residuals	2.89939e+11
Standard error of residuals ($\hat{\sigma}$)	30243
Unadjusted R^2	0.1739
Adjusted $\bar{R}^2$	0.1661
$F(3, 317)$	22.2511

Finally, the following model is estimated with observations for both years 1978 and 1981:

$$\text{LRPRICE} = \beta_0 + \beta_1\text{LDIST} + \beta_2\text{Y81} + \beta_3\text{Y81LDIST} + \varepsilon$$

where:
LRPRICE = log(RPRICE)
LDIST= log(DIST) = natural logarithm of the distance (in miles) of the home to the incinerator;
Y81LDIST = Y81 $\times$ log(DIST).

OUTPUT 4: OLS estimates using the 321 observations $1 - 321$
Dependent variable: LRPRICE

Variable	Coefficient	Std. Error	t-statistic	p-value
C	8.0585	0.5084	15.85	0.0000
LDIST	0.3167	0.0515	6.15	0.0000
Y81	−0.2752	0.8051	−0.34	0.7327
Y81LDIST	0.0482	0.0818	0.59	0.5562

Mean of dependent variable	11.26
S.D. of dependent variable	0.3879
Sum of squared residuals	37.1217
Standard error of residuals ($\hat{\sigma}$)	0.3422
Unadjusted R^2	0.2290
Adjusted $\bar{R}^2$	0.2217

PROBLEMA 2: FERTILITY IMPACT ON CHILDREN YEARS EDUCATION

(**Note**: The data has been changed for the purpose of this problem and it does not reflect the real data)

There is a theory that stablishes a negative relationship between number of children in the household and their educational achievement. In order to evaluate such theory, we have a sample of families with 2 or more children from the Chilean Census data for the year 2002. We want to study the impact of number of children on years of education **for the oldest two siblings** at home. In order to evaluate the impact of fertility (measured by the number of children) on children's usual years of education, we concentrate on the following specification:

$$\begin{aligned} \text{LYEDU} &= \beta_0 + \beta_1 \text{AGE} + \beta_2 \text{AGE2} + \beta_3 \text{MOMAGE} + \beta_4 \text{MOMAGE2} \qquad \text{(E.1)} \\ &\quad + \beta_5 \text{MOMEDU} + \beta_6 \text{MOMEDU2} + \beta_7 \text{FEMALE} \qquad (1) \\ &\quad + \beta_8 \text{SECOND} + \beta_9 \text{URBAN} + \beta_{10} \text{NCHILD} + \varepsilon \end{aligned}$$

where, for each child:

LYEDU = log of child's years of education;
AGE = child's age in years;
AGE2 = square of child's age;
MOMAGE = mother's age in years;
MOMAGE2 = mother's square of age;
MOMEDU = mother's years of education;
MOMEDU2 = mother's square of years of education;
FEMALE = dummy variable that equals 1 if the child is female and zero otherwise;
SECOND = dummy variable that equals 1 if the child is second born and zero otherwise;
URBAN = dummy variable that equals 1 if the child lives in an urban area and zero otherwise;
NCHILD = number of children younger than 18 years old that live at woman's home.

Furthermore, we know that fertility decisions are correlated with unobserved characteristics that simultaneously affect child's educational attainment. Then, we would expect that

$$C(\varepsilon, \text{NCHILD}) \neq 0,$$

whereas the rest of the right-hand-side variables in (E.1) **are not correlated with** such unobserved variables (ε).

Besides the variables mentioned above, we have information whether or not a family has faced a multiple birth (MB). We understand that a woman had faced a multiple birth if she had twins, triplet, quadruple or quintuple children in the second birth. We also have information about the gender composition in the family. In general, families prefer to have children of different gender.

Then, we can define the variables Multiple Births (MB), and Same Sex (SSEX) as:
MB = dummy variable that equals 1 if the woman has faced a multiple birth and zero otherwise;
SSEX = dummy variable that equals 1 if the oldest two siblings are of the same gender and zero otherwise.
We have obtained the following estimates:

OUTPUT 1: OLS estimates using the 43972 observations 1 − 43972
Dependent variable: LYEDU

Variable	Coefficient	Std. Error	t-statistic	p-value
C	−3.8964	0.0285	−136.63	0.0000
AGE	0.6660	0.0026	254.38	0.0000
AGE2	−0.0185	0.0001	−188.05	0.0000
MOMAGE	0.0045	0.0012	3.84	0.0001
MOMAGE2	−0.00004	0.000014	−2.89	0.0038
MOMEDU	0.0221	0.0011	20.31	0.0000
MOMEDU2	−0.0008	0.00005	−15.12	0.0000
FEMALE	0.0174	0.0025	7.00	0.0000
SECOND	−0.0102	0.0023	−4.35	0.0000
URBAN	0.0219	0.0032	6.74	0.0000
NCHILD	−0.0266	0.0016	−16.50	0.0000

Mean of dependent variable	1.66418
Sum of squared residuals	2037.25
Unadjusted R^2	0.8868
Adjusted $\bar{R}^2$	0.8867

OUTPUT 2: OLS estimates using the 43972 observations 1 – 43972
Dependent variable: NCHILD

Variable	Coefficient	Std. Error	t-statistic	p-value
C	3.9502	0.0816	48.41	0.0000
AGE	0.0664	0.0077	8.65	0.0000
AGE2	−0.0010	0.0003	−3.51	0.0004
MOMAGE	−0.0498	0.0035	−14.40	0.0000
MOMAGE2	0.00048	0.00004	11.99	0.0000
MOMEDU	−0.0356	0.0032	−11.18	0.0000
MOMEDU2	0.0016	0.00016	9.81	0.0000
FEMALE	−0.0052	0.0075	−0.69	0.4877
SECOND	0.1359	0.0068	19.87	0.0000
URBAN	−0.0399	0.0095	−4.18	0.0000
MB	0.9280	0.0341	27.24	0.0000
SSEX	0.0252	0.0098	2.58	0.0099

Sum of squared residuals	17544.1
Unadjusted R^2	0.0480
Adjusted $\bar{R}^2$	0.0477

NOTE: In an OLS regression similar to OUTPUT 2
but that omits MB, SSEX: $R^2 = 0.0319$

OUTPUT 3: TSLS Estimations using the 43972 observations 1 – 43972
Dependent variable: LYEDU
Instruments: MB SSEX

Variable	Coefficient	Std. Error	t-statistic	p-value
C	−3.9987	0.0567	−70.48	0.0000
AGE	0.6642	0.0027	254.38	0.0000
AGE2	−0.0184	0.0001	−185.65	0.0000
MOMAGE	0.0058	0.0013	4.36	0.0000
MOMAGE2	−0.00005	0.000015	−3.48	0.0001
MOMEDU	0.0230	0.0012	19.51	0.0000
MOMEDU2	−0.00086	0.00006	−14.89	0.0000
FEMALE	0.0175	0.0025	7.00	0.0000
SECOND	−0.0137	0.0029	−4.74	0.0000
URBAN	0.0230	0.0033	6.97	0.0000
NCHILD	−0.0007	0.0125	−0.05	0.9581

Mean of dependent variable	1.66418
Unadjusted R^2	0.8861

OUTPUT 4: OLS estimates using the 43972 observations 1 – 43972
Dependent variable: LYEDU

Variable	Coefficient	Std. Error	t-statistic	p-value
C	−3.9987	0.0567	−70.48	0.0000
AGE	0.6642	0.0027	254.38	0.0000
AGE2	−0.0184	0.0001	−185.65	0.0000
MOMAGE	0.0058	0.0013	4.36	0.0000
MOMAGE2	−0.00005	0.000015	−3.48	0.0001
MOMEDU	0.0230	0.0012	19.51	0.0000
MOMEDU2	−0.00086	0.00006	−14.89	0.0000
FEMALE	0.0175	0.0025	7.00	0.0000
SECOND	−0.0137	0.0029	−4.74	0.0000
URBAN	0.0230	0.0033	6.97	0.0000
NCHILD	−0.0007	0.0125	−0.05	0.9581
RES	−0.0264	0.0126	−2.10	0.0361

Sum of squared residuals 2037.04
Unadjusted R^2 0.8868
Adjusted $\bar{R}^2$ 0.8867
(NOTE: RES are the residuals from OUTPUT 2)

OUTPUT 5: OLS estimates using the 43972 observations 1 – 43972
Dependent variable: RES1

Variable	Coefficient	Std. Error	t-statistic	p-value
C	−0.0009	0.0279	−0.03	0.9744
AGE	0.000015	0.0026	0.01	0.9954
AGE2	−0.000001	0.0001	−0.01	0.9929
MOMAGE	0.00001	0.0012	0.01	0.9912
MOMAGE2	−0.0000001	0.000014	−0.01	0.9926
MOMEDU	−0.00001	0.0011	−0.01	0.9928
MOMEDU2	0.0000004	0.00005	0.01	0.9937
FEMALE	−0.0007	0.0025	−0.28	0.7767
SECOND	−0.00002	0.0023	−0.01	0.9937
URBAN	0.00007	0.0033	0.02	0.9837
MB	−0.0011	0.0116	−0.10	0.9226
SSEX	0.0045	0.0033	1.36	0.1741

Mean of dependent variable 0.000000
2049.15
Unadjusted R^2 0.00004
Adjusted $\bar{R}^2$ −0.0002
(NOTE: RES1 are the residuals from OUTPUT 3)

PROBLEM 3: EFFECT OF THE ETHNIC ORIGIN ON DEATH PENALTY

Some experts believe that in the United States the probability of being condemned to death is higher, *ceteris paribus*, when the accused is black. To evaluate such hypothesis, we analyze 679 in different States where the death penalty can be applied. In this sample, the proportion of white victims is 76%. We consider the following variables:

CONDENA = binary variable that equals 1 if the accused is condemned to death and 0 otherwise;
RAZA_ACUSADO = binary variable that equals 1 if the accused is black and 0 otherwise;
RAZA_VICTIMA = binary variable that equals 1 if the victim is white and 0 otherwise.

We obtained the following results:

OUTPUT 1: Logit estimates using the 679 observations 1 – 679
Dependent variable: CONDENA

Variable	Coefficient	Std. Error	t-statistic	Slope*
C	−2.08	0.14	−14.86	
RAZA_ACUSADO	−0.39	0.31	−1.26	−0.0356

*At mean

Mean of CONDENA = 0.102
Number of cases 'correctly predicted'= 610 (89.8 percent)
McFadden's Pseudo-R^2 = 0.0040
$f(\beta' x)$ at mean of independent variables = 0.090
Log-likelihood = −222.25
Likelihood ratio test: $\chi^2_1 = 1.770$

OUTPUT 2: Logit estimates using the 679 observations 1 – 679
Dependent variable: CONDENA

Variable	Coefficient	Std. Error	t-statistic	Slope*
const	−4.43	0.61	−7.26	
RAZA_ACUSADO	0.83	0.36	2.31	0.0644
RAZA_VICTIMA	2.39	0.60	3.98	0.1861

*At mean

Number of cases 'correctly predicted'= 610 (89.8 percent)
McFadden's Pseudo-R^2 = 0.0496
$f(\beta' x)$ at mean of independent variables = 0.078
Log-likelihood = −212.07
Likelihood ratio test: $\chi^2_2 = 22.13$

Universidad Carlos III de Madrid
<u>ECONOMETRICS I</u>
Academic year 2006/07
FINAL EXAM
January 30, 2007

Exam's type: 1

TIME: 2 HOURS 30 MINUTES

Directions:

- BEFORE YOU START TO ANSWER THE EXAM:
  - Fill in your personal data in the optical reading form, which will be the only valid answering document. Remember that you must complete all your identifiying data (name and surname(s), and **NIU**) both in letters and in the corresponding optical reading boxes. **Very important**: The identification number that you must fill is your **NIU** (NOT the DNI or the Passport), which has 9 digit and always begins by 1000.
  - Fill in, both in letters and in the corresponding optical reading boxes, the course code (10188) and your group (65 or 75).
- **Check that this document contains 40 questions sequentially numbered.**
- Check that the number of exam type that appears in the questionnaire matches the number indicated in the optical reading form.
- Read the questions carefully.
  Whenever a question is referred to a Problem included in the enclosed document, the question will include within parentheses at the beginning of the question the corresponding problem number.
  It is advised to read carefully the text of the problem before answering its corresponding questions.
- For each row regarding the number of each question, fill the box which corresponds with your chosen option in the optical reading form (A, B, C or D).
- **Each question only has one correct answer.**
  Any question in which more than one answer is selected will be considered incorrect and its score will be zero.
- All the questions correctly answered has the same score. Any incorrect answer will score as zero. To obtain a grade of 5 over 10 you must correctly answer 24 questions.
- If you wish, you may use the answer table as a draft, although such table does not have any official validity.
- You can use the back side of the problem text as a draft (no additional sheets will be handed out).
- The relevant **statistical tables** are attached at the end of the Problem text document.
- **Any student who were found talking or sharing any sort of material during the exam will be expelled out inmediately and his/her overall score will be zero, independently of any other measure that could be undertaken.**

- **Date of grades publication**: Thursday, February, 1st.
- **Date of exam revision**:
  - Tuesday, February, 6 at 15 h in classrooms 15.0.05 y 15.0.06.

- **Rules for exam revision**:
  - Its only purpose will be to check that the number of correct answers is right.
  - To be entitled for revision, the student should bring a printed copy of the exam solutions, which will be available in Aula Global from the date of grades publication.

Draft of ANSWERS									
QUESTION	**(a)**	**(b)**	**(c)**	**(d)**	**QUESTION**	**(a)**	**(b)**	**(c)**	**(d)**
1.					**21.**				
2.					**22.**				
3.					**23.**				
4.					**24.**				
5.					**25.**				
6.					**26.**				
7.					**27.**				
8.					**28.**				
9.					**29.**				
10.					**30.**				
11.					**31.**				
12.					**32.**				
13.					**33.**				
14.					**34.**				
15.					**35.**				
16.					**36.**				
17.					**37.**				
18.					**38.**				
19.					**39.**				
20.					**40.**				

1. **(Problem 2)** Assuming that $C(\text{NCHILD}, \varepsilon) = 0$, and the model (E.1) satisfies all the further assumptions of the classical linear regression model except the assumption of conditional homoscedasticity:
   (i) The parameter estimates of OUTPUT 1 are not consistent.
   (ii) The standard errors of the parameters of OUTPUT 1 are not consistent.
   (iii) The R^2 of the model has no sense.

   (a) Only (ii) is true.
   (b) Only (ii) and (iii) are true.
   (c) The three assertions are true.
   (d) Only (i) and (ii) are true.

2. **Problem 2)** If from model (E.1) we want to test the null hypothesis that child's years of education is independent on mother's education.

   (a) The null hypothesis would be $H_0 : \beta_5 = 0$.
   (b) The null hypothesis would be $H_0 : \beta_0 = \beta_5 = \beta_6$.
   (c) The null hypothesis would be $H_0 : \beta_5 = \beta_6 = 1$.
   (d) The null hypothesis would be $H_0 : \beta_5 = \beta_6 = 0$.

3. **Problem 2)** We are concerned with obtaining consistent estimators for all the coefficients of equation (E.1).

   (a) The estimators of OUTPUT 1 are consistent.
   (b) The estimators of OUTPUT 3 are consistent, because the instruments (MB and SSEX) fulfills the two required conditions to be a valid instrument: being uncorrelated with ε (as it is pointed by OUTPUT 5) and being correlated with the endogenous variable NCHILD (as it can be seen in the first stage regression OUTPUT 2).
   (c) The estimators of OUTPUT 3 are consistent, because the instruments (MB and SSEX) fulfills the two required conditions to be a valid instrument: being uncorrelated with ε (as it is pointed by OUTPUT 4) and being correlated with the endogenous variable NCHILD (as it can be seen in the first stage regression OUTPUT 2).
   (d) The estimators of OUTPUT 3 are not consistent, because we would need the instruments (MB and SSEX) were not correlated with the endogenous variable NCHILD, what does not appear to be the case given OUTPUT 2.

4. **Problem 2)** Assume that $C(\text{NCHILD}, \varepsilon) = 0$, so that NCHILD is exogenous. A girl 10 years old, who is the oldest with one aditional sibling in the family whose mother is 25 years old and with 10 years of education, and living in a rural area, will have on average, approximately:

   (a) 1.1 years of education.
   (b) We need more information.
   (c) 2.8 years of education.
   (d) 3 years of education.

5. **Problem 2)** Assume that $C(\text{NCHILD}, \varepsilon) = 0$, so that NCHILD is exogenous. For a given gender, age, mother's age, mother's education, urban status and birth order, an additional child entails on average an estimated decrease on years of education of approximately:

   (a) 2.7% years of education.

   (b) 0.027% years of education.

   (c) It depends on the age of the individual.

   (d) 0.027 years of education.

6. **Problem 2)** Assume that SSEX is not a valid instrument. In this situation:
   (i) OUTPUT 3 does not give a consistent estimate of model (E.1).
   (ii) If MB is a valid instrument we still can have a consistent estimator of model (E.1).
   (iii) We can not test the validity of MB as instrument.

   (a) The three assertations are true.

   (b) Only (i) is true.

   (c) Only (i) and (ii) is true.

   (d) Only (ii) is true.

7. **Problem 2)** Assume for this question that $C(\text{NCHILD}, \varepsilon) \neq 0$, $C(\text{MB}, \varepsilon) \neq 0$ and $C(\text{SSEX}, \varepsilon) = 0$. Then:
   (i) The number of children coefficient in OUTPUT 1 is an inconsistent estimate of β_{10}.
   (ii) The number of children coefficient in OUTPUT 3 is an inconsistent estimate of β_{10}.
   (iii) The inconsistency bias of the coefficient of the number of children is always larger (in absolute value) in OUTPUT 1 that in OUTPUT 3.

   (a) The three assertions are true.

   (b) Only (i) and (ii) are true.

   (c) The three assertions are false.

   (d) Only (i) and (iii) are true.

8. **Problem 2)** If we want to evaluate whether the variable NCHILD is an endogenous variable:

   (a) We will test whether NCHILD is endogenous in the first stage equation by means of a test that the coefficient of MB and SSEX are jointly equal to zero.

   (b) We will test the joint significance of all the regressors in OUTPUT 2 (test of joint significance, or regression test).

   (c) We will test whether NCHILD is endogenous in the equation for LYEDU by means of a Hausman test.

   (d) We will get the residuals from OUTPUT 1 and run an auxiliary regression of such residuals on NCHILD and the rest of the explanatory variables in the model. A significant coefficient for NCHILD will provide evidence that the residuals and NCHILD are correlated.

9. **Problem 2)** Given the results:
(i) Given OUTPUT 4, we reject that NCHILD is exogenous at the 5% level (but not at the 1% level).
(ii) The test for the null hypothesis of no correlation between the instruments and the model error term, which is approximately distributed as a χ^2_1, has a value of $43972 \times 4 \times 10^{-5} \simeq 1.76$, so that we do not reject the null hypothesis at the 10% significance level.
(iii) The test for the null hypothesis that both MB and SSEX are uncorrelated with NCHILD, which is approximately distributed as a χ^2_2, has a value about 739, so that we reject such hypothesis at any reasonable significance level.

   (a) Only (i) and (ii) are true.
   (b) Only (i) and (iii) are true.
   (c) Only (ii) and (iii) are true.
   (d) The three assertions are true.

10. **Problem 2)** Using the appropriate estimates, we can conclude that for a given gender, age, mother's age, mother's education, gender, birth order and urban status, an additional child:

   (a) Entails an estimated decrease in approximately 2.7% years of education.
   (b) Entails an estimated decrease in approximately 0.07% years of education.
   (c) Does not decrease years of education.
   (d) Entails an estimated decrease in approximately 2.7 years of education.

11. **Problem 2)** Using the appropriate estimates, we can conclude that for a given gender, age, mother's age, mother's education, gender, number of siblings and urban status, the oldest child in the family has approximately:

   (a) 1.4 more years of education than the second sibling.
   (b) 1% less years of education than the second sibling.
   (c) 1.4% more years of education than the second sibling.
   (d) 1 less years of education than the second sibling.

12. **Problem 2)** Using the appropriate estimates, a girl 17 years old, who is the second born with 5 additional siblings in the family whose mom is 43 years old and with 12 years of education, and living in a rural area will have on average, approximately:

   (a) 9.8 years of education.
   (b) 9 years of education.
   (c) 9.2 years of education.
   (d) 2.3 years of education.

13. **Problem 2)** Using the appropriate estimates, the mother's age threshold over which mother's age has a negative impact on children education is:

   (a) We need more information.
   (b) 60 years old.
   (c) 58 years old.
   (d) 56 years old.

14. **(Problem 3)** Using the most appropriate available estimations, we can say:
   (a) If the accused is black, it is more likely that he is condemned.
   (b) The race of the accused has an influence only if the victim is black.
   (c) The race of the accused does not have an influence on the probability of being condemned.
   (d) If the accused is white, it is more likely that he is condemned.

15. **Problem 3)** According to the results, we can say that:
   (a) If the victim is white, it is more likely that the accused is condemned.
   (b) The race of the victim only has an influence if the accused is black.
   (c) The race of the victim does not influence the probability of being condemned.
   (d) If the victim is black, it is more likely that the accused is condemned.

16. **Problem 3)** The predicted probability that an accused black person is condemned when the victim is white is approximately equal to:
   (a) 0.23
   (b) 0.96
   (c) 0.08
   (d) 0.83

17. **Problem 3)** If the specified model in the OUTPUT 2 had been estimated by a linear probability model instead of using the logit model:
   (i) The error term would present heteroscedasticity (conditional on the explanatory variables).
   (ii) The error term would follow a normal distribution (conditional on the explanatory variables).
   (iii) The predicted probabilities could be bigger than 1 or smaller than 0.
   (a) The three statements are true.
   (b) Only (i) and (ii) are true.
   (c) Only (i) and (iii) are true.
   (d) Only (ii) and (iii) are true.

18. **Problem 3)** Consider the following statements:
   (i) The model of OUTPUT 1 presents a omitted variable problem.
   (ii) An explanation for differences in the estimated coefficient of the race of the accused between OUTPUT 1 and OUTPUT 2 is the negative correlation between the race of the accused and the race of the victim.
   (iii) We can conclude that the majority of the accused persons as well as the majority of the condemned persons are black.
   (a) Only (i) is true.
   (b) Only (i) and (iii) are true.
   (c) Only (ii) and (iii) are true.
   (d) Only (i) and (ii) are true.

19. **Problem 3)** Consider the following statements:
(i) If the victim is white, the estimated effect of being an accused black person on the probability of conviction is approximately equal to 0.11.
(ii) The estimated effect of being an accused black person on the probability of conviction is approximately equal to 0.22, independently of the race of the victim.
(iii) The race of the victim has a higher influence on the probability of conviction than the race of the accused.

    (a) The three statements are true.
    (b) Only (i) and (iii) are true.
    (c) Only (iii) is true.
    (d) Only (i) and (ii) are true.

20. **Problem 3)** Consider the following statements related to OUTPUT 2:
(i) The model is linear in parameters.
(ii) The model is linear in variables.
(iii) The predicted probabilities could be bigger than 1 or smaller than 0.

    (a) Only (i) and (iii) are true.
    (b) Only (i) is true.
    (c) Only (ii) and (iii) are true.
    (d) The three statements are false.

21. **Problem 3)** In OUTPUT 2, the hypothesis that all the coefficients, except for the constant, are zero:
(i) Is not rejected at the significance level of 1%.
(ii) Is not rejected at the significance level of 5%.
(iii) Is not rejected at the significance level of 10%.

    (a) Only (i) and (ii) are true.
    (b) Only (i) is true.
    (c) We do not have enough information to evaluate any of the three statements.
    (d) The three statements are false.

22. **Problem 3)** The predicted probability that an accused black person is condemned, evaluated in the sample proportion of white victims, is approximately equal to:

    (a) 0.14.
    (b) 0.07.
    (c) 0.08.
    (d) 0.93.

23. **Problem 3)** If we want to test if the race of the victim is a relevant variable to explain the probability that an accused person is condemned to death:
(i) Since the value of the corresponding t-statistic is approximately 3.98 , we reject at the 1% level the null-hypothesis that the coefficient of that variable is zero.
(ii) Since the value of the corresponding likelihood-ratio test is approximately 20.36, we reject at the 1% level the null-hypothesis that the coefficient of that variable is zero.
(iii) Comparing the restricted model with the unrestricted one, the number of correct predictions is identical, which gives evidence that that variable is irrelevant.

    (a) Only (i) is true.
    (b) Only (i) and (ii) are true.
    (c) Only (iii) is true.
    (d) Only (ii) is true.

24. **Problem 3)** If the OUTPUT 2 includes all relevant explanatory variables and the assumption of the logistic distribution is correct, the model of OUTPUT 2 characterizes:
(i) The probability that an accused is condemned to death, conditional on the race of the accused and on the race of the victim.
(ii) The conditional expectation of the event that an accused is condemned to death, conditional on the race of the accused and on the race of the victim.
(iii) The linear projection of the event that an accused is condemned to death, conditional on the race of the accused and on the race of the victim.

    (a) The three statements are true.
    (b) Only (i) is true.
    (c) Only (i) and (iii) are true.
    (d) Only (i) and (ii) are true.

25. **Problem 3)** Given the model of OUTPUT 2:
(i) We would have obtained similar effects of the explanatory variables (race of the accused, race of the victim) if we had assumed a normal distribution instead of a logistic one and if we had estimated by Maximum Likelihood (whenever there are not too many extreme values in the sample).
(ii) We would have obtained consistent estimators of the effects of the explanatory variables (race of the accused, race of the victim) if we had mantained the assumption of a logistic distribution and if we had estimated by non-linear least squares.
(iii) If, mantaining the assumption of a logistic distribution, we had estimated by non-linear least squares, the conventional standard errors would be inappropriate because of the existence of conditional heteroscedasticity.

    (a) The three statements are true.
    (b) Only (ii) and (iii) are true.
    (c) Only (i) and (ii) are true.
    (d) Only (i) is true.

26. **Problem 3)** Given the model of OUTPUT 2:
(i) If we mantain the assumption of a logistic distribution, the estimations of the parameters would be identical if we estimate by Maximum Likelihood as well as if we apply non-linear least squares.
(ii) If we estimated by Maximum Likelihood assuming a normal distribution instead of a logistic one (and assuming that there are not too many extreme values in the sample) the magnitudes of the estimated coefficients would be bigger in the case of the normal distribution.
(iii) If, mantaining the assumption of a logistic distribution, we estimate by non-linear least squares instead of by Maximum Likelihood, the standard errors of estimation by non-linear least squares of the estimated coefficients will be smaller or equal to the corresponding standard errors when we estimate by Maximum Likelihood.

    (a) The three statements are true.
    (b) Only (i) and (iii) are true.
    (c) Only (i) and (ii) are true.
    (d) The three statements are false.

27. **Problem 3)** According to the estimations of OUTPUT 2, the average of the estimated effect of being an accused black person on the probability of condemnation is approximately equal to:

    (a) 0.22.
    (b) We do not have enough information to calculate this average effect.
    (c) 0.09.
    (d) 0.11.

28. **(Problem 1)** Given the results in OUTPUT 1, consider the following statements:
(i) the constant term can be interpreted as the average selling price for the houses that are not close to garbage incinerator in the year 1981.
(ii) The results imply that the cause of lower selling price for house located close to the garbage incinerator is the own garbage unit; the coefficient for the variable NEARINC is negative and statistically significant.
(iii) The results allows us to measure the average selling price for house close to the garbage incinerator in 1981.

    (a) Only (i) and (ii) are true.
    (b) The three statements are true.
    (c) Only (i) and (iii) are true.
    (d) Only (ii) and (iii) are true.

29. **Problem 1)** Identify which of the following statements is true:

    (a) The average estimated selling price for houses in 1981 is approximately 70594 dollars.
    (b) None of the previous answers is true.
    (c) The average estimated selling price for houses that are closer to the garbage incinerator is approximately 30714 dollars higher than houses located farther away.
    (d) The average estimated selling price for houses located farther away from the garbage incinerator in 1981 is approximately 30714 dollars higher than the ones for houses that are closer.

30. **Problem 1)** The results in OUTPUT 2 show that:

   (a) The garbage incinerator was constructed in a zone with houses whose selling price in average is lower than those for houses located father away.

   (b) The garbage incinerator did not produce a significant drop in house selling price.

   (c) If the garbage incinerator would have not been constructed, the selling prices for the houses would be 82517 dollars.

   (d) The garbage incinerator caused a drop in selling prices in a magnitude of 18824 dollars.

31. **Problem 1)** The estimated average difference between the selling price for houses located farther away from the garbage incinerator and those located closer to the unit in 1981:
(i) is approximately 30714 dollars.
(ii) is approximately 70594 dollars.
(iii) is a good measure of the impact associated to the construction of the garbage incinerator on the average house selling price.

   (a) Only (iii) is true.

   (b) Only (i) and (iii) are true.

   (c) Only (i) is true.

   (d) Only (ii) and (iii) are true.

32. **Problem 1)** The estimated average difference in the selling price for houses located close to the garbage incinerator between 1981 and 1978 is approximately:

   (a) 70594 dollars.

   (b) 11890 dollars.

   (c) 6901 dollars.

   (d) 63693 dollars.

33. **Problem 1)** The estimated impact of the garbage incinerator on the average house selling price for the houses located close to the garbage unit in relationship those located farther away is approximately:

   (a) 18791 dollars.

   (b) −11890 dollars.

   (c) 63693 dollars.

   (d) 70594 dollars.

34. **Problem 1)** Under the evidence of OUTPUT 1 and 2:
(i) We are able to measure the causal impact associated to the construction of the garbage incinerator on house selling price
(ii) We are not able to know rather or not the causal impact associated to the construction of the garbage incinerator on house selling price is statistically different from zero.
(iii) We are able to know the estimated difference in selling price for houses located farther away from the incinerator between the years 1981 and 1978.

   (a) Only (i) and (iii) are true.

   (b) The three statements are true.

   (c) Only (i) and (ii) are true.

   (d) Only (ii) and (iii) are true.

35. **Problem 1)** Under the evidence of OUTPUT 3:
(i) The coefficient for the variable Y81 captures the change in the average selling price for all houses between 1978 and 1981.
(ii) The coefficient for the constant term captures the average selling price for houses located in 1978 independently on the location of the house.
(iii) The coefficient for the variable NEARINC captures the impact on average selling price associated to the construction of the garbage incinerator.

 (a) Only (i) and (ii) are true.
 (b) The three statements are false.
 (c) Only (ii) and (iii) are true.
 (d) Only (i) and (iii) are true.

36. **Problem 1)** Given the results in OUTPUT 3:

 (a) The estimation suggests that the garbage incinerator reduces the average selling price of the houses in 18824 dollars.
 (b) None of the previous statements is correct.
 (c) We cannot know if the reduction in the selling price of the houses due to the garbage incinerator is statistically different from zero.
 (d) The estimation suggests that the average difference between houses located closer to garbage incinerator and those located farther away is not statistically significant at the 10% level.

37. **Problem 1)** Given the results in OUTPUT 4:

 (a) Due to the construction of the garbage incinerator, there was an approximately 31% devaluation in the house selling price.
 (b) None of the previous answers is correct.
 (c) In 1981 the house value was approximately 0.27% lower that the one in 1978.
 (d) Due to the construction of the garbage incinerator, there was an approximately 0.31% devaluation in the house selling price.

38. **Problem 1)** Given the results in OUTPUT 4:
(i) We cannot reject the null hypothesis that the construction of the garbage incinerator does not have an effect on the house selling price.
(ii) We cannot reject the null hypothesis that the garbage incinerator was constructed close to houses whose value is not different from the one of houses located farther away from the garbage incinerator.
(iii) We cannot reject the null hypothesis that the average selling price of houses in 1981 is not different from the one in 1978.

 (a) Only (i) and (iii) are true.
 (b) The three statements are true.
 (c) Only (ii) and (iii) are true.
 (d) Only (i) and (ii) are true.

39. **Problem 1)** The model in OUTPUT 4:
(i) As well as with previous outputs, this one does not allow us to get an estimate of the causal impact of the construction of the garbage incinerator on house selling price because the included variables explain just an approximately 23% of the selling price variation.
(ii) the model does not allow us to get a precise estimation of the impact of the garbage incinerator on the house selling price, since the coefficient of the variable Y81LDIST is not statistically significant at 5%.
(iii) is a better model to get an estimate of the causal impact of the construction of the garbage incinerator than the model in OUTPUT 3; OUTPUT3 has a considerable higher determination coefficient.

(a) The three statements are true.
(b) Only (i) and (iii) are true.
(c) Only (i) and (ii) are true.
(d) The three statements are false.

40. **Problem 1)** Given the results in OUTPUT 1, analyze the following statements:
(i) The coefficient for the constant term is understood as the average selling price of houses conditional of not being close to the incinerator in 1978.
(ii) The results imply that the cause of the lower house selling price for houses located close to the garbage incinerator is the own construction of this unit; the coefficient for the variable NEARINC is negative and significant.
(iii) The results allow us to measure the average price for houses located close to the garbage incinerator in 1981.

(a) The three statements are true.
(b) The three statements are false.
(c) Only (iii) is true.
(d) Only (i) is true.