MOCK MID TERM

Remark: this is not a complete mid-term exam, but different exam type questions from previous mid-terms to practice.

QUESTION 1. Using the Household expenditure survey for 2017, you get the following information for 96 Spanish households:

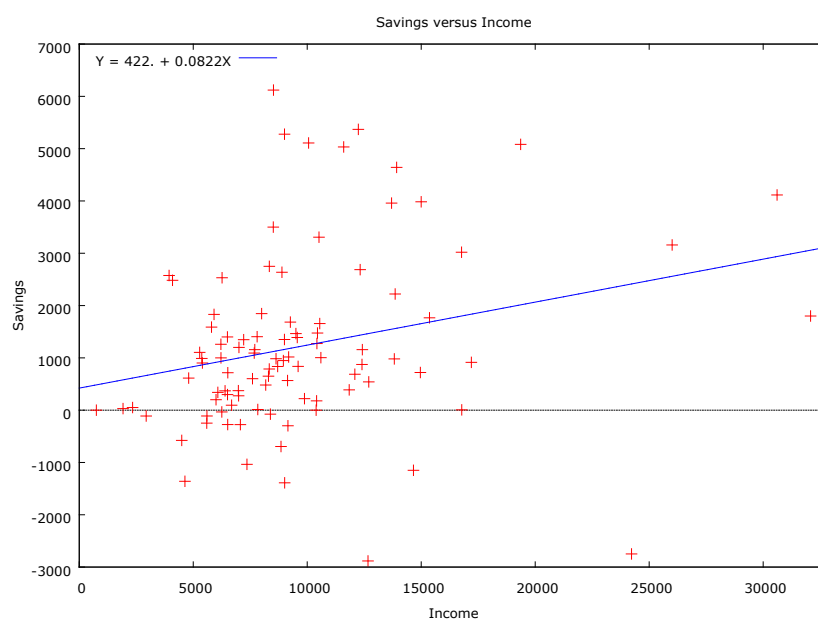|	
Savings	Annual savings (euros)
Income	Annual income (euros)

Source: INE, 2017.

Figure 1

OLS regression line: $\widehat{savings}_i = 422 + 0.0822\,income_i$

Answer the following questions:

a) According to **Figure 1**, *by how much* do average annual savings increase when annual income increases by 1000 euros? *Explain* why some actual values for savings fall below zero.

b) According to **Figure 1**, in a SLRM of savings on income, which of the Gauss Markov assumptions is very likely to be violated? *Explain*

c) Given the OLS regression, *compute* the residual for a household with annual income= 5000 euros and annual savings = 863 euros.

QUESTION 2 A group of researchers has conducted a survey that contains information on smoking behavior and other variables for a random sample of 807 single adults from the United States. The following information is available:

Variable	Description
cigs	Average number of cigarettes smoked per day
cigprice	State cigarette price, cents per pack
educ	Years of schooling
age	Age in years
income	Annual income, (thousand dollars)

Three different models have been estimated using **cigs** as dependent variable.

	Model 1	Model 2	Model 3
const	10.67	9.565	14.74
	(6.173)	(6.210)	(6.543)
cigprice	**-0.032**	**-0.032**	**-0.032**
	(0.102)	**(0.102)**	**(0.102)**
income		0.081	0.118
		(0.053)	(0.056)
educ			-0.376
			(0.169)
age			-0.041
			(0.028)
n	807	807	807
Adj. R²	0.005	0.006	0.010
SSR	151734	151294	150157

Note: Standard errors in parentheses

Answer the following questions:

a) <u>Which</u> is the unit of analysis in this dataset and the sample size?

b) Postulate the population model of the SLRM

c) <u>Write</u> the **OLS line** of the SLRM.

d) Does the **price** of cigarettes explain a large fraction of the variability in cigarettes smoked across individuals? <u>Explain</u>.

e) <u>Interpret</u> the estimated coefficients from **Model 2**.

f) <u>Compare</u> the estimated slope parameters associated to **cigprice** between the three models. Could you give an <u>explanation</u> for this happening?

g) Robert is 25 years-old and has completed 14 years of schooling. His annual income is around 26 thousand dollars and lives in Iowa where the price of the pack of

cigarettes is 58 cents. Predict Robert's number of cigarettes smoked per day (integer number) using **Model 3.**

h) It has been argued that, controlling for other factors, individuals tend to give up smoking as they get older. Is this result consistent with the regression in **Model 3**? Explain

i) Which would be your preferred Model? Explain

QUESTION 3

A group of researchers from the World Bank Development Department conducted a study in 2019 on the main **determinants of human development across countries.** For that purpose, they collected information of an Index provided by the United Nations called **Human Development Index (HDI)** ranging the level of development from 0 (the worst) to 1 (the best). They ended up with a sample of **169 countries** around the world. All the information they gathered for the different countries is described in Table 1.

Table 1. Variables Description

Name	Description
hdi	Human Development Index (being 0 the worst and 1 the best)
educ	Mean years of schooling
gdppc	GDP per capita (in thousand dollars)
imr	Infant Mortality rate (per 1,000 live births)
unem	Total Unemployment rate (% of the labor force)

The descriptive statistics are presented in Table 2.

Table 2. Summary Statistics

Variable	Mean	Coeff of Variation	Minimum	Maximum
hdi	0.702	0.214	0.377	0.954
educ	8.498	0.364	1.586	14.132
gdppc	18.49	1.068	0.660	112.53
imr	22.49	0.895	1.600	87.600
unem	7.239	0.786	0.100	30.200

1) Consider the following OLS regression line:

$$\widehat{hdi}_i = \hat{\beta}_0 + 0.045educ_i$$

Choose the correct answer for the OLS estimator of the intercept, $\hat{\beta}_0$

   a. 0.319
   b. 1.091
   c. 0.045
   d. 8.466

2) Consider the OLS regression line given below.

$$\widehat{hdi}_i = \hat{\beta}_0 + 0.045educ_i$$

If the Total Sum of Squares of educ equals 160.62 and the unbiased estimator of the error variance equals 2.89, the standard error of the slope is equal to

   a. 0.134
   b. 0.018
   c. 55.43
   d. 7.44

3) Consider the following OLS regression line

$$\widehat{hdi}_i = 0.368 + 0.035educ_i + 0.002gdppc_i$$

$$n = 169 \quad R^2 = 0.852$$

The fitted value of hdi for a country with mean values for both educ and gdppc is

   a. 0.368
   b. 0.702
   c. 0.954
   d. 0.377

4) Consider the following OLS regression line

$$\widehat{hdi}_i = 0.368 + 0.035educ_i + 0.002gdppc_i$$

$$n = 169 \quad R^2 = 0.852$$

Qatar is the richest country in the sample. The mean years of schooling in Qatar is 9.67 and it presents a high level of human development index (its hdi equals 0.85). The OLS residual for Qatar is:

   a. -0.08
   b. 0.08

    c. 0.15
    d. -0.15

5) Consider the following OLS model

$$\log\widehat{(hdi)}_i = 0.108 + 0.032 educ_i + 0.117\log(gdppc)_i$$
$$n = 169 \quad SSR = 0.46 \quad SST = 5$$

The adjusted R-squared is equal to

    a. 0.908
    b. 0.906
    c. 0.46
    d. 0