LISBOA SCHOOL OF ECONOMICS & MANAGEMENT — ISEG SINCE 1911

Statistics II

Graduations in Management and Economics

2nd semester – 2016/2017
03/07/2017

Duration: 2 h 10 min

EN A

Name_____ **N^o**_____

Space for marks

Multiple choice questions: only one option is correct
each right answer scores **1**
each wrong answer scores **-0.25**

True an false questions: each right answer scores **0.25** points
each wrong answer scores **-0.25** points

Derive and justify all your answers.

If you need extra space, you can use page 8 of the paper, showing clearly which question are you answering.

The only page you can detach from the staple is the last one, which contains the regression output and can be used as draft paper.

1. (a) For each statement, indicate if it is true **(T)** or false **(F)**. (1.0)

	T	**F**
Let (a,b) and (c,d) be two confidence intervals at 90% and 95%, respectively, for the mean of a normal population, obtained through the usual procedure with the Pivotal and based on the same sample. Then, we can conclude that $(a,b) \subset (c,d)$.		
T_1 is a biased estimator for μ, the mean of a given population, whereas T_2 is an unbiased estimator for μ. However, T_1 is consistent but T_2 is not. Then, you prefer estimator T_1.		
If, in a statistic test, the p-value is 0.023, than we reject H_0 at the 5% significance level.		
The presence of heteroskedasticity makes the OLS estimators of the regression coefficients to be biased.		

(b) Justify your answer to the second statement in the previous question. (1.0)

2. The claim amount, in thousands of euros, of a given multirisk insurance follows the following distribution:

$$F_X(x) = 1 - \left(\frac{3}{x+3}\right)^2$$

(a) Given two claims, what is the probability that the size of the lower of the two is higher than three thousand euros. (1.0)

(b) Given a sample of 50 claims from such multirisk policies, what is the approximate probability that the sample average does not exceed two thousand euros. (1.5)

3. Let $X \sim Poi(\lambda)$ and let $T_1 = \bar{X}$ and $T_2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ be two estimators for λ.

(a) Show that both T_1 and T_2 are unbiased estimator for λ. (1.5)

(b) Which of the two estimators is more efficient? (1.5)

4. A risk credit analyst is willing to study the credit default risk in automobile loans. A random sample of 1000 automobile loans was analyzed, with 230 being in default.

(a) Build a 95% confidence interval for the proportion of automobile credit default. Is it plausible to state the that proportion is 10%. (1.5)

(b) Can the analist state that the proportion of automobile credit default is not higher than 20%? Answer based on an approriate test of size 5%. (1.5)

(c) The same risk credit analyst wants now to compare the credit default risk of automobile loans with holydays loans. Another random sample of 1500 holydays loans was analyzed, with 195 being in default. The analyst states that the proportion of defaults in holydays loans is not higher than that of automobile loans. Does the data supports the statment? Answer based on an approriate test of size 5%. (2.0)

5. Two shops sell three types of adventure packs. Based on a random sample of sales to 3000 customers, the following results were obtained. (2.0)

	Pack 1	**Pack 2**	**Pack 3**
Shop A	450	520	530
Shop B	390	540	570

Can we consider that the sales of the three packs depend on the shop? Answer performing an appropriate test.

6. In order to study the amount of student loans, the following multiple linear regression model was estimated based on a sample of 500 applications:

$$\text{LLOAN}_\text{i} = \beta_0 + \beta_1 \text{LFEES}_\text{i} + \beta_2 \text{YEARS}_\text{i} + \beta_3 \text{LINC}_\text{i} + \beta_4 \text{FEDUC}_\text{i} + u_i$$

where the variables are as follows:

- LLOAN: logarithm of the amount of a student's loan (€);
- LFEES: logarithm of the total amount of fees for the student's degree (€);
- YEARS: duration of the degree (years)
- LINC: logarithm of annual family income (€);
- FEDUC: father's years of schooling.

Taking into account the results of the estimation of the above model, and of other additional regressions presented in the output, answer the following questions.

(a) Interpret the estimates of parameters β_1 and β_2, and comment on the statistical significance of regressors LFEES_i and YEARS_i. (1.5)

(b) Can you consider the elasticity of the loan amount with respect to family income to be equal to 0.3? Answer based on an appropriate test of size 5%. (1.5)

(c) Test the hypothesis that the loan amount only depends on the characteristics of the degree. (1.5)

(d) A new regressor, EUR which is a dummy variable that takes the value 1 if the student is form Europe and zero otherwise, was introduced in the model. The coefficient of determination of the new model is: (1.0)

☐ $R^2 = 0.685334$ ☐ $R^2 = 0.031126$ ☐ $R^2 = 0.732901$ ☐ $R^2 = 0.345792$

OUTPUT 10/03/2017

Equation 1:

Dependent Variable: LLOAN
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.300113	0.348675	0.86073	0.390
LFEES	0.597644	0.031871	18.7520	0.000
YEARS	0.253055	0.043526	5.81385	0.000
LINC	0.326096	0.038106	8.55760	0.000
FEDUC	0.020294	0.005175	3.92154	0.000

R-squared	0.685334	Mean dependent var	10.3719
Adjusted R-squared	0.682791	S.D. dependent var	0.525674
S.E. of regression	0.296066	Sum squared resid	43.3893
F-statistic	269.524	Prob(F-statistic)	0.0000

Equation 2:

Dependent Variable: LLOAN
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.554210	0.268316	9.51940	0.0000
LFEES	0.771882	0.026889	28.7065	0.0000
YEARS	0.091151	0.041270	2.20867	0.0280

R-squared	0.631067	Mean dependent var	10.3719
Adjusted R-squared	0.629585	S.D. dependent var	0.525674
S.E. of regression	0.320587	Sum squared resid	51.1824
F-statistic	425.919	Prob(F-statistic)	0.0000