Estadística I Final Exam, 26th May 2010.
Bachelor's Degrees in ADE, ADE-DER, ADE-INF, ECO, ECO-DER.

EXAM RULES: 1) Use separate booklets for each problem.
2) Perform the calculations with at least two significant decimal places.
3) You may not leave the exam during the first 30 minutes.
4) You are not allowed to leave the classroom without handing in the exam.

Problem 1 (10 points) An experiment is conducted whereby people who have been condemned for misdemeanors should make monthly payments to their victims, whose amount is determined by a judge. The following table contains the monthly salary (x) of 10 convicted people and the payments (in euros) to the victims (y):

Salaries	300	880	1000	1540	1560	1600	1600	2200	3200	6000
Payments	200	380	400	200	800	600	800	1000	1600	2700

Additionally, the following information is available: $\overline{x} = 1988$, $\overline{y} = 868$, $s_x^2 = 2594240$, $s_y^2 = 593351.1$ and $s_{xy} = 1206062$.

a) (2 points) Sketch the scatterplot (preserving the scale) showing the payments as a function of the salaries. Compute and interpret Pearson's correlation coefficient. Do you think that the relation between both variables is linear? Justify your comments.

b) (2 points) Compute the regression line for y on x. Determine its slope and interpret its value. Represent the regression line on the scatterplot of question a), explaining in detail how you did it.

c) (2 points) Compute the determination coefficient and the regression line model's percentage of unexplained variability. Justify the adequacy of this model.

d) (2 points) According to the regression line, what payment would correspond to a condemned person with a 1400 euros salary? Comment wether this prediction makes sense, and justify your comments.

e) (2 points) Represent a box-plot for the *Salaries* data and determine if there are any outliers. Justify your answer.

Problem 2 (10 points) The route transit time X (in minutes) for the buses of a certain city has been found to be a uniformly distributed random variable on the interval $(30, 40)$.

a) (1 point) Sketch the probability density function of X. Indicate clearly in the coordinate axes both the values of X with positive density and the values of the density function.

b) (2 points) What is the probability that the route transit time of a bus will be between 30 and 37.5?

c) (3 points) We select 100 buses at random and we are interested in the number of buses out of those 100 having a route transit time between 30 and 37.5 (we denote this random variable as Y). Give the name of the distribution of Y and its parameters' values. Calculate the expectation and standard deviation of Y.

d) (4 points) What is the probability that fewer than 64 buses will have route transit times between 30 and 37.5. Use an approximation.

Problem 3 (10 points) A study on consumption habits for the families of the Community of Madrid shows that the monthly expenditure on supermarket-branded products per family (X) is normally distributed with standard deviation 50 euros. The following table contains a simple random sample of $n = 10$ families:

Family	1	2	3	4	5	6	7	8	9	10
Expenditure	325	193	203	305	251	240	233	238	354	311

a) (2.5 points) Define what is an unbiased estimator and propose one for estimating the average monthly expenditure on supermarket-branded products for the families of the Community of Madrid. Justify your answer. What is the estimation for the average monthly expenditure based on your proposal?

b) (2.5 points) Compare the estimator proposed in a) with the following estimator:

$$T = \frac{X_1 + X_n}{2},$$

where $X_1, X_2, \ldots, X_n$ is a simple random sample of X. According to the criteria commented in class, which estimator is preferred? What is the estimation obtained using T?

c) (2.5 points) Compute the 95% confidence interval for the average monthly expenditure on supermarket-branded products for the families of the Community of Madrid, and specify the necessary hypotheses for this computation.

d) (2.5 points) If we want to work at a 90% confidence level, do we get a wider or a narrower interval than the one computed before? Justify your answer without computing the interval anew.