vrije Universiteit amsterdam

School of Business and Economics

Exam:	**Sample Exam**
Code:	E_EOR1_DA1
Examinator:	-
Co-reader:	-
Date:	-
Time:	-
Duration:	2 hours
Calculator allowed:	Yes
Graphical calculator allowed:	No
Number of questions:	3
Type of questions:	Open
Answer in:	English

Credit score:	100 credits counts for a 10
Grades:	The grades will be made public within 10 working days
Inspection:	TBA
Number of pages:	4 (including front page)

Good luck!

(This page is intentionally left blank.)

Question 1 (30/100 points)

(a) Obtain the sample mean and median of the following data points

$$5.2, \quad ,2.7, \quad 2.2, \quad -0.7, \quad 0.3, \quad -1.8.$$

(b) State an advantage and a disadvantage of the median compared to the sample mean as a measure of centrality.

(c) Give the definition of a quantile of level α: q_α. For a certain variable you have obtained that the quantile of level 0.95 is $q_{0.95} = 23.5$ and the quantile of level 0.05 is $q_{0.05} = 12.7$. What can you say about the distribution of this variable?

(d) The following R code to create a function is given

```
> oper_vec <- function(x,y){
+   if(length(x)!=length(y)) {stop("incorrect argument")}
+   if(!is.numeric(x)) {stop("incorrect argument")}
+   if(!is.numeric(y)) {stop("incorrect argument")}
+   z <- ifelse(x>y, x, y)
+   return(z)
+ }
```

Explain what the function "oper_vec" does (input and output).

Question 2 (40/100 points)

(a) We have $n = 213$ observations for each of 2 variables. Consider a linear regression model of the form $y_i = \beta_0 + \beta_1 x_i + u_i$. The following sample statistics are given:

$$\bar{y} = 2.3, \quad \bar{x} = -1.3, \quad s_x^2 = 6.8, \quad s_y^2 = 11.6, \quad s_{xy} = 7.3.$$

(i) Obtain the OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

(ii) Obtain the R^2. How would you interpret the value of the R^2 you have obtained?

(iii) Obtain the TSS, RSS and ESS.

(b) Consider the following R code to perform a regression using the data vectors y and x

```
> x3 <- x^3
> reg <- lm(y~x+x3)
```

Write the regression model the R code is estimating. Is the intercept β_0 included or excluded from the model? How would you change the R code given above to include/exclude the intercept?

(c) Consider the regression model given by $y_i = \beta_0 + \beta_1 x_i + u_i$. Show that the OLS estimates of β_0 and β_1 are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

by setting the partial derivatives of the sum of squares to zero.

Question 3 (30/100 points)

(a) Discuss differences and similarities between the Maximum Likelihood (ML) rule and the ECM rule.

(b) We have observations that take values in the set $\{0, 1, 2, 3\}$, that is, $x \in \{0, 1, 2, 3\}$. Assume we have 2 groups Π_1 and Π_2. The distribution of Π_1 is $Bin(3, 0.5)$ and the distribution of Π_2 is $Bin(3, 0.4)$. Note that the probability mass function $f(x)$ of a binomial distribution $Bin(n, p)$ is

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for} \quad x \in \{0, 1, \ldots, n\}.$$

(i) Draw a plot of the densities $f_1(x)$, $f_2(x)$.

(ii) Find the sets R_1 and R_2 according to the Maximum Likelihood rule.

(iii) Obtain the Expected Cost of Misclassification (ECM) for the ML rule when the misclassification costs are $C(1|2) = 1$ and $C(2|1) = 2$.

(iv) Obtain the sets R_1 and R_2 of the ECM discriminant rule and the corresponding ECM.

End of the exam!