EXERCISE

1. Introduction

It has been known for a long time that the final result of a pregnancy is affected by multiple factors. Typical risk behaviours such as smoking or alcohol consumption can negatively affect the characteristics of the newborns.

In a study abut tobacco and alcohol consumption during pregnancy, a total of 411 pregnant women were interviewed to collect information about their smoking habits, jointly with sociodemographic information. Mothers and newborn basic physical parameters were also collected. The data collected for this study have been stored in an Excel file "pesonac.xls", also available as STATA (pesonac.dta) or text (pesonac.csv) format.
Table 1 below contains the codebook defining the variables in the database.

N of variables: 12
N of observations: 411

Variable	Type	Format	Description	Label	Valid cases Missing
nserie	int	%8.0g	Number of mother identification		411 (100.00%) 0 (0.00%)
edad	byte	%8.0g	Mother age		411 (100.00%) 0 (0.00%)
tabco2	byte	%8.0g	Number of daily cigarettes consumed during pregnancy		411 (100.00%) 0 (0.00%)
socon2	byte	%18.0g	Education level	1: No studies 2: Primary studies 3: Secondary studies 4: University studies	411 (100.00%) 0 (0.00%)
socon3	byte	%10.0g	Labour situation	0: Unemployed 1: Employed 2: 2	411 (100.00%) 0 (0.00%)
tgr	byte	%8.0g	Gestational age at delivery(weeks)		411 (100.00%) 0 (0.00%)
sex	byte	%8.0g	Newborn sex	1: Male 2: Female	409 (99.51%) 2 (0.49%)
pn	int	%8.0g	Newborn weight		411 (100.00%) 0 (0.00%)
pm	float	%9.0g	Mother weight		411 (100.00%) 0 (0.00%)
tm	int	%8.0g	Mother height		411 (100.00%) 0 (0.00%)
tp	int	%8.0g	Father height		411 (100.00%) 0 (0.00%)
tn	int	%8.0g	Newborth height		411 (100.00%) 0 (0.00%)

Table 1 Labelbook describing the variables in the database collected for this study

The main objective of this study is to evaluate the influence of smoking as well as of the mother physical and sociological characteristics in the newborn's ones. Specifically the "Newborn weight" and the "Gestational age at delivery" can be considered the main outcomes while other variables can be seen as explanatory or as factors to adjust for.

Your work will consist of
- Defining as accurately as possible the questions to be answered
- Take a sample from the main database
- Proceed with those analysis that you think need to be done
- Write a report where you describe, the study goals, the methodology used for the analysis (briefly), the main results and the limitations of the study.

2. The analysis

You may proceed along the following steps.

2.2 The Analyses

In order to reach your objectives you may proceed in different ways. The steps below illustrate one posible way to proceed, not the only way to do this. If you wish to introduce changes, feel free to do so.

1. **Do a descriptive analysis of the variables and answer the following questions**
   a. What are the mean, median, standard deviation and IQR of birthweight? Interpret the results.
   b. Which is the best descriptive measure for `birthweight`, `gestational age` and the three heights (`mother, father and newborn`)
   c. Describe graphically the relationship between `birthweight` and `sex` of the newborn.
   d. Describe graphically the relationship between `birthweight` and `educational level` of the mother

3. Final report

After doing the analysis you must write a report (you can give a try to Rmarkdown!)