

(MRC/Wits Rural Public Health & Health Transitions Research Unit (Agincourt))

Objectives of the exercise:

The questions are designed to assess candidates familiarity with data management problems, and ability to formulate solution and present them in response to set of requirements.

Data sets: (provided)

- *FertilityData.xlsx: This dataset contains record of individuals, and pregnancy outcomes*

Exercises:

Complete the exercise and return your responses/well documented code, including an appropriate & well written report.

Part A:

1. Please study the following data and comment on anything strange about it.

Name	Year Of Birth	Father	Mother
William	1959	Bright Mabunda	Busi Khosa
Walter	1962	Bright Mabunda	Busi Khosa
Valies	1963	Bens Mabunda	Thandeka Nyalungu
Suzan	1962	Bens Mabunda	Busi Khosa

2. Everyone in a survey is given a unique 6 digit identity number and check digit separated by a dash (-). Please study the following list and comment on anything unusual.

IndividualId	YearOfBirth	FatherId	MotherId
101010-1	1926	172463-1	176464-2
120963-3	1933	193741-2	120963-3
131044-5	1930	1722A3-4	197651-2
137249-7	1936	101010-1	139724-3
134572-9	1948	129374-2	172463-1
149323-2	1938	101010-2	147263-5
155421 -3	1995	14726-5	139724-3

Part B:

1. Import the data in the attached excel file ("FertilityData.xlsx") into a database management system of your choice. The file contains two tables (Individuals and PregnanciesAndBirths). The description and domain values of the variables in the tables are included in the "Codebook" table. Send to us a description of the steps you used to complete the task .
2. Write an SQL script that extracts the following information from the data you imported into your database, and graphically present the table using appropriate plots.

	Total Population	Number of Births	Number of Male Births	Number of Female Births
1996				
1997				
1998				
1999				
2000				

3. Using the same dataset, compute the average birth weight for males and for females born during the period 1997 to 2000? Exclude birth weights which are bad data from your calculation.
4. What is the total number of women who experienced at least two pregnancies over the period 1996 to 2000?