Practical on descriptive analyses

Daniela Zugna

The dataframe *births.csv* contains data from 500 singleton births in a London Hospital.

Variable	Code
id	Identity
bweight	Birth weight of baby (g)
lowbw	Indicator for birth weight less than 2500 g
gestwks	Gestation period (weeks)
preterm	Indicator for gestation period less than 37 weeks
matage	Maternal age
hyp	Indicator for maternal hypertension
sex	Sex of baby: 1:Male, 2:Female

1. Load these data and print the contents of the data frame to the screen.

2. Print all the data for subject 7.

3. Print the data on the variable *bweight* of the first 20 subjects.

4. Obtain a description of the data structure and summarize it.

5. Summarize the numeric variables *bweight, gestwks,matage* and make a note of the distribution of values.

6. Plot an histogram and a density plot for the numeric variables *bweight, gestwks,matage*.

7. Plot a box-plot for the numeric variables *bweight, gestwks,matage*.

8. Convert the variables *preterm, hyp, sex* into factors and label the levels of these variables.

9. Make a table of frequencies for variables *preterm, hyp, sex*

10. Plot a pie chart for variables *preterm, hyp, sex*.

11. Create a new factor *gest.cat* which cuts *gestwks* at 20, 35, 37, 39, and 45 weeks, including the left hand end, but not the right hand. Make a table of the frequencies for the levels of *gest.cat*.

12. Plot an histogram of *gest.cat*

13. Create a indicator variable called *early* according to whether gestwks is less than 30 or not and make a frequency table of *early*.

14. Print the id numbers of women with gestwks less than 30 weeks.

15. Count missing values in *gestwks*