Exercise getStartedR

Alessio Crippa

October 13, 2016

Get started with R, Exercises

The dataset *lowbwt* is about a study that aims to identify risk factors associated with giving birth to a low birth weight baby (weighing less than 2500 grams). Variables that were thought to be of importance were age, weight of the subject at her last menstrual period, smoking during pregnancy and race.

Questions

Data inspection

1. Load the **lowbwt** available at http://alecri.github.io/downloads/data/ (full address for Rdata file http://alecri.github.io/downloads/data/lowbwt.Rdata)

2. How many observations and variables are in the dataset?

3. Sort the data by (increasing) age

4. Categorize age in two groups (< 30, $>= 30$ years). Attach the proper labels to the new factor variable.

5. Select and print white subjects whose child's birth weight is less than 1.5 kg

Univariate statistics

6. Summarize the continuous response variable birth weight. What is its mean and standard deviation?

7. Provide a graphical presentation of its distribution

8. Categorize birth weight in two groups: <2500 g and $>= 2500$ g (same as the *lwt* variable)

9. What is the percentage of women who had a baby weighting less than 2.5 kg?

10. Provide a graphical presentation for this binary variable

Bivariate association

11. What is the mean and standard deviation of mother's age among white, black, and other races?

12. Present graphically the distribution of mother's age in the races subgroups

13. What is the percentage of smoking mothers among white, black, and other races?

14. What is the difference in the mean birth weight comparing smoker vs non-smoker women? Test the hypothesis of equality of means. What do you conclude?

15. What is the risk of low birth weight among smoker and non-smoker women? Test the hypothesis of equality of proportions (no association). What do you conclude?