

Tutorial Business Analytics

R Tutorial 1

Exercise 1.1 Loading a data set and statistics.

- a) Read the CSV file "LaborSupply1988.csv" into a tibble `df`.
- b) How many attributes (columns) and observations (rows) does `df` have?
- c) Which attributes does the data set have?
- d) List the first rows of the data set.
- e) What is the value range of the attribute - `age`?
- f) Calculate the average of annual hours worked by the laborers with 0, 1, 2, ... 6 kids each.
- g) Calculate the average number of `kids` of a 40-year-old person.

Exercise 1.2 Plotting

- a) Plot a histogram of the attribute `age`. What is the most frequent age?
- b) Plot the average number of `kids` against the `age` and interpret the resulting graph. Underpin your observation using a statistical method.
- c) Plot the log of hourly wage (`lnwge`) against the `age`.
- d) Plot the mean of the log of hourly wage (`lnwge`) against the `age`. How are they correlated? Also compute the correlation.
- e) Plot `lnhr` against the `age` with different colors for `disab=0` and `disab=1`.
- f) Plot a boxplot of the log of annual hours worked (`lnhr`) against the number of `kids`. What could be observed regarding mean and variance? Is the observation meaningful for large values of `kids`?