Tutorial Business Analytics

R Tutorial 1 - Solution

Exercise 1.1 Loading and describing a data set

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 labourers with 0, 1, 2, ... 6 kids each.
g) Calculate the average number of `kids` of the 40 year old.

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 (`lnwg`) against the `age`.
d) Plot the mean of the log of hourly wage (`lnwg`) 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`?

Tutorial Business Analytics

Homework 1 - Solution

Exercise 2.1: Describing the beer consumption on the Oktoberfest

a) Read the provides CSV file ("Oktoberfest.csv") and store it in a tibble named *oct.*
b) Which attributes does the data set have?
c) What was the price of a beer in 1995?
d) Based on the data set, when did the city of Munich first recorded the beer price?
e) What is the value range of the attribute – *Visitors_Total* describing the total number of visitors in million in the corresponding year?
f) Plot and describe the beer consumption over the years
g) The number of visitors could provide an explaination to this observation. Create a scatter-plot that shows the number of visitors per year. Subsequently, calculate a statistic to validate or reject this explanation.

Exercise 2.2: Describing the beer price on the Oktoberfest

a) What was the average beer price from 2000 to 2007?
b) What was the variance of the beer price within this time frame?
c) Add a new variable *difference* using the *mutate* function that describes the difference between the beer price of a year and the previous year.
d) Plot these differences per year using ggplot2.