

### Homework assignment in Week 4

- Due next Sunday (17:00, 24th, Oct.), submit electronically to Manaba+R.
- File name: YourStudentID\_W04\_n.py (ID without hyphen, e.g., 12345678901\_W04\_1.py).
- Your code must include your **own comments for all code. Go line-by-line.** Comments in your program must be full sentences and reflect your understanding of the code.

Q1: Refer to the final exercise in the slide for week 3, and add source code to achieve the following data manipulations.

(1): use one-line code to combine two transaction data

(2): use one-line code to combine two transaction\_detail data

(3): use one-line code to combine transaction and transaction\_detail data, the name of the newly generated data (by combining) should be "join\_data".

(4) use one-line code to add customer's information (customer\_name, registration\_date, email, gender, age, birth, pref) into each transaction (each row) in join\_data.

(5) use one-line code to add item's information (item\_name and item\_price) into each transaction (each row) of join\_data

(6) use two-line code to add the price column into join\_data. For each line, the value of "price"=item\_price\*quantity.

Q2: Refer to the source code in page 13 (CDF) of the slide for week 4, create your own program to calculate and visualize the CDF of the pregnancy length for first babies in the NSFG data (used in "Hypothesis testing", included in 2002FemPreg.csv file).

Q3: Suppose we toss a coin 250 times and see 140 heads and 110 tails (recorded in file tossCoin.csv, 1: heads, 0: tails). Based on this result, we might suspect that the coin is biased. Modify the source code provided in class (HT in python) to perform the Hypothesis Testing and calculate the p-value, give a conclusion for the Hypothesis Testing (by using the comments in source code).

Q1 → YourStudentID\_W04\_1.py file

Q2 → YourStudentID\_W04\_2.py file

Q3 → YourStudentID\_W04\_3.py file