# 70-511: Statistical Programming
# Programming Assignment 7 – Aggregating ACS PUMS Data

**Introduction**

For this assignment, you will work again with the same ACS PUMS dataset as for assignment 6 to produce several tables which aggregate the data.

**Requirements**

You are to create a program in Python that performs the following using the pandas packages:

1. Loads the ss13hil.csv file that contains the PUMS dataset (assume it's in the current directory) and create a DataFrame object from it.

2. Create 3 tables:

   TABLE 1: Statistics of HINCP - Household income (past 12 months), grouped by HHT - Household/family type
   - Table should use the HHT types (text descriptions) as the index
   - Columns should be: mean, std, count, min, max
   - Rows should be sorted by the mean column value in descending order

   TABLE 2: HHL - Household language vs. ACCESS - Access to the Internet (Frequency Table)
   - Table should use the HHL types (text descriptions) as the index
   - Columns should the text descriptions of ACCESS values
   - Each table entry is the sum of WGTP column for the given HHL/ACCESS combination, divided by the sum of WGTP values in the data. Entries need to be formatted as percentages.
   - Table should include marginal values ('All' row and column).
   - Any rows containing NA values in HHL, ACCESS, or WGTP columns should be excluded.

   TABLE 3: Quantile Analysis of HINCP - Household income (past 12 months)
   - Rows should correspond to different quantiles of HINCP: low (0-1/3), medium (1/3-2/3), high (2/3-1)
   - Columns displayed should be: min, max, mean, household_count
   - The household_count column contains entries with the sum of WGTP values for the corresponding range of HINCP values (low, medium, or high)

3. Display the tables to the screen as shown in the sample output on the last page.

**Additional Requirements**

1. The name of your source code file should be `tables.py`. All your code should be within a single file.
2. You need to use the pandas DataFrame object for storing and manipulating data.
3. Your code should follow good coding practices, including good use of whitespace and use of both inline and block comments.
4. You need to use meaningful identifier names that conform to standard naming conventions.
5. At the top of each file, you need to put in a block comment with the following information: your name, date, course name, semester, and assignment name.
6. The output should **exactly** match the sample output shown on the last page.

**What to Turn In**

You will turn in the single tables.py file using BlackBoard.

**HINTS**

- To get the right output, use the following functions to set pandas display parameters:
  ```
  pd.set_option('display.max_columns', 500)
  pd.set_option('display.width', 1000)
  ```
- To display entries as percentages, use the applymap method, giving it a string conversion function as input. The string conversion function should take a float value v as an input and output a string representing v as a percentage. To do this, you can use formatting strings or the format() method

### Sample Program Output

```
70-511, [semester] [year]
NAME: [put your name here]
PROGRAMMING ASSIGNMENT #7

*** Table 1 - Descriptive Statistics of HINCP, grouped by HHT ***
                                                                        mean            std  count    min      max
HHT - Household/family type
Married couple household                                       106790.565562  100888.917804  25495  -5100  1425000
Nonfamily household:Male householder:Not living alone           79659.567376   74734.380152   1410      0   625000
Nonfamily household:Female householder:Not living alone         69055.725901   63871.751863   1193      0   645000
Other family household:Male householder, no wife present        64023.122122   59398.970193   1998      0   610000
Other family household:Female householder, no husband present   49638.428821   48004.399101   5718  -5100   609000
Nonfamily household:Male householder:Living alone               48545.356298   60659.516163   5835  -5100   681000
Nonfamily household:Female householder:Living alone             37282.245015   44385.091076   8024 -11200   676000

*** Table 2 - HHL vs. ACCESS - Frequency Table ***
                                              sum
                                             WGTP
ACCESS                             Yes w/ Subsrc. Yes, wo/ Subsrc.      No      All
HHL - Household language
English only                               58.71%            2.93%  16.87%   78.51%
Spanish                                     7.83%            0.52%   2.60%   10.95%
Other Indo-European languages               5.11%            0.18%   1.19%    6.48%
Asian and Pacific Island languages          2.73%            0.06%   0.28%    3.08%
Other language                              0.80%            0.03%   0.14%    0.97%
All                                        75.19%            3.73%  21.08%  100.00%

*** Table 3 - Quantile Analysis of HINCP - Household income (past 12 months) ***
          min      max           mean  household_count
HINCP
low    -11200    37200   19599.486904          1629499
medium  37210    81500   57613.846298          1575481
high    81530  1425000  159047.588900          1578445
```