

**In-Class Practice (HW) 7**

Due: March 30, 2018

- Asking questions to TAs and collaborating with classmates are encouraged, but copying, sharing, or distributing any material is strictly prohibited. Homework should be students' original work.
- Please submit
  - 1) SAS code (.SAS) with detailed comments
  - 2) PDF document with relevant output and interpretations
- Late homework will not be accepted.

**Mid-Atlantic Wage Data**

Dataset 'wage.xlsx' contains wage and other information for a group of 300 workers in the Mid-Atlantic region. Following is the list of variables included in the dataset:

Variable	Description
ID	ID
Year	Year that wage information was recorded.
Age	Age
Sex	Gender
Marital	Marital status: 5 categories Never Married / Married / Widowed / Divorced / Separated
Race	Race: 4 categories White / Black / Asian / Other
Education	Education level: 5 categories < HS Grad / HS Grad / Some College / College Grad / Advanced Degree
Jobclass	Type of job: 2 categories Industrial / Information
Health	Health level: 2 categories Good / Very Good
Insurance	Whether the worker has health insurance Yes / No
Wage	Raw wage

The dataset was manually assembled by Steve Miller of Open BI, from the March 2011 Supplement to Current Population Survey data.

**a) Import the dataset and name it 'wage'.**

**b) Descriptive statistics:** Use the dataset 'wage' to provide the following tables and plots. Describe the distribution (e.g. missing values, symmetry, skewness, association between variables, location (mean, median), dispersion (range, standard deviation, outliers) of variables displayed in those tables and plots.

- i. Frequency table and bar chart of race
- ii. Distribution of age
  - 1) Descriptive statistics (n, mean, median, standard deviation, min, max)  
Use two decimal points.
  - 2) Histogram (binwidth = 10)
- iii. Distribution of wage depending on education level
  - 1) Descriptive statistics (n, mean, median, standard deviation, min, max)  
Use two decimal points.
  - 2) Boxplots

**c) Produce the following report.**

		Jobclass							
		Industrial				Information			
		Age		Wage		Age		Wage	
Education	N	mean	std	mean	std	mean	std	mean	std
1. < HS Grad	268	40.57	12.41	82.7	21.4	44.78	12.68	87.5	21.8
2. HS Grad	971	41.04	11.81	94.9	28.4	44.46	12.13	97.4	28.8
3. Some College	650	39.87	11.43	106.3	30.3	42.02	11.54	109.4	34.7
4. College Grad	685	43.28	11.09	121.2	40.3	42.43	10.78	126.6	41.7
5. Advanced Degree	426	45.27	10.75	136.0	46.9	44.92	10.12	155.6	55.2

Hint: Use ACROSS and GROUP appropriately.

The number of observation N, mean, and standard deviation are calculated.

Be careful with the decimal points.

d) Produce the following table.

Job class and education level		Marital						
		Divorced	Married	Never Married	Separated	Widowed	Total	
		Freq	Freq	Freq	Freq	Freq	Freq	
Industrial	1. < HS Grad	9	124	46	10	1	190	
	2. HS Grad	44	423	152	10	7	636	
	3. Some College	29	218	92	2	1	342	
	4. College Grad	15	199	53	4	3	274	
	5. Advanced Degree	6	82	14	NONE	NONE	102	
	Total	103	1046	357	26	12	1544	
Information	1. < HS Grad	7	50	16	4	1	78	
	2. HS Grad	29	228	67	10	1	335	
	3. Some College	23	203	72	9	1	308	
	4. College Grad	26	288	90	5	2	411	
	5. Advanced Degree	16	259	46	1	2	324	
	Total	101	1028	291	29	7	1456	
Total	1. < HS Grad	16	174	62	14	2	268	
	2. HS Grad	73	651	219	20	8	971	
	3. Some College	52	421	164	11	2	650	
	4. College Grad	41	487	143	9	5	685	
	5. Advanced Degree	22	341	60	1	2	426	
	Total	204	2074	648	55	19	3000	

Hint: Use ALL option appropriately. Rename N and ALL.

Variable names ‘jobclass’ and ‘education’ are not displayed.

In empty cells, print ‘NONE’.

Print ‘Job class and education level’ in the upper left corner of the table.

**e) Hypothesis testing:** The researchers have several research questions that they want to investigate by using this dataset. For each question,

- 1) Clarify the null and alternative hypotheses.
- 2) Determine an appropriate statistical test.
- 3) Check the assumptions.
- 4) Report your conclusion based on the test result. Test at the significance level of 0.05.
  - i. Is the mean age different from 40?
  - ii. Is there a difference in mean wage depending on education level?
  - iii. Is the proportion of having health insurance less than 70%?
  - iv. Among those who were never married, is job class independent of education level?