Portfolio Project Utilizing SAS Code

Ali Bruckner

Colorado State University – Global Campus

Dr. Dwight Davis

June 4, 2020

Portfolio Project Utilizing SAS Code

The primary goal of this Portfolio Project is to determine if there is a significant difference in healthcare expenses incurred by patients between the years of 2017 and 2012. A secondary goal will be to determine if there is a significant difference between three buckets of age. The age buckets will look at 0-18 year olds, 19-55 year olds, and 56 and older. The SAS code will include code for summary statistics, output creation of survey means, normality tests, and two sample t-tests.

The dataset has been sourced from the Medical Expenditure Panel Survey website that is maintained by the U.S. Department of Health & Human Services. The site provides data based on surveys of individuals, medical providers and employers within the United States in relation to health insurance coverage, the use of health care, and the cost of health care. Two datasets were used which are identical except one is for the year 2017 and the other is for 2012. There are many data points in the files with the focus being on the amount of overall healthcare expenses and age. Two hypotheses will be tested.

Hypothesis 1: The amount of healthcare expenses in 2017 did not change when

comparing them to healthcare expenses in 2012.

Null Hypothesis 1: H0:µd = Do

Alternate Hypotheses 1: μd ≠ Do

Hypothesis 2: The amount of healthcare expenses does not alter between the three age

buckets of 0-18, 19-55, and 56+ during the year 2017.

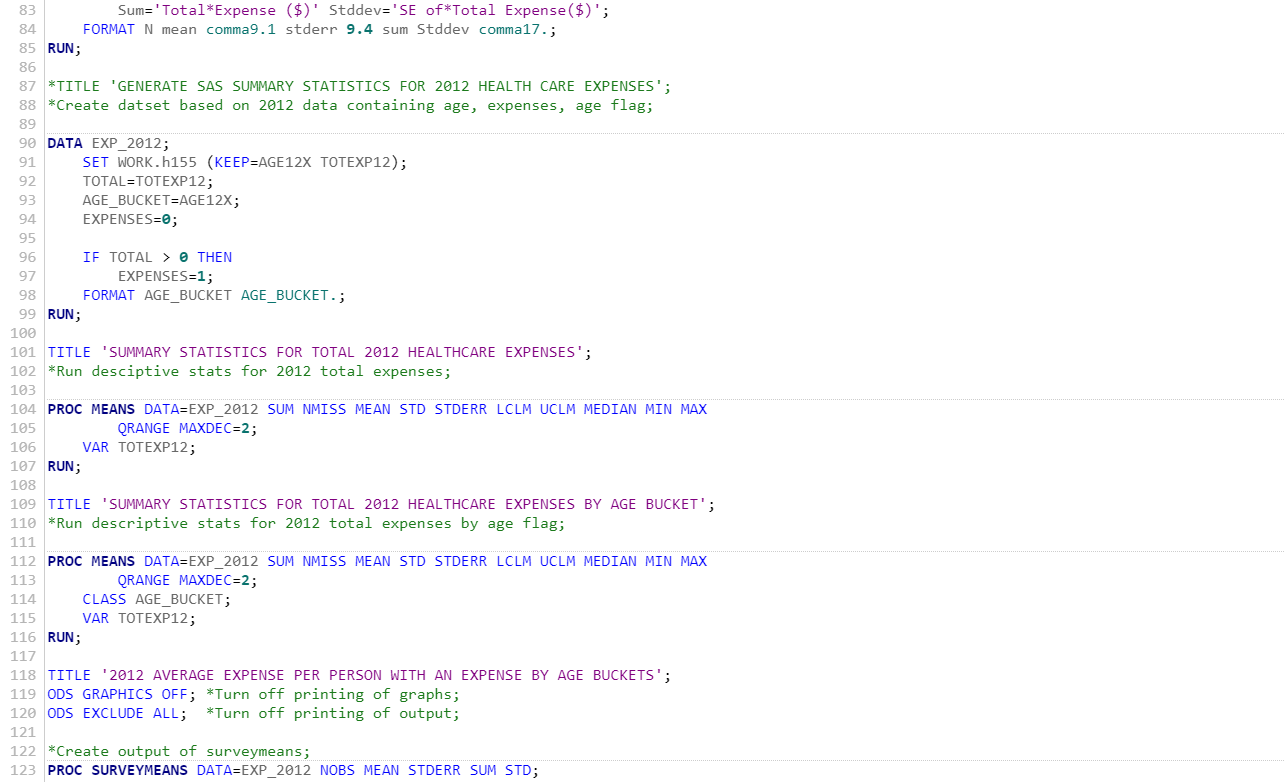
Null Hypothesis 2: H0:µ1 = µ2 = µ3= µ4

Alternate Hypotheses 2: µ1 ≠ µ2 ≠ µ3 ≠ µ4

**SAS Code**

**

**

**

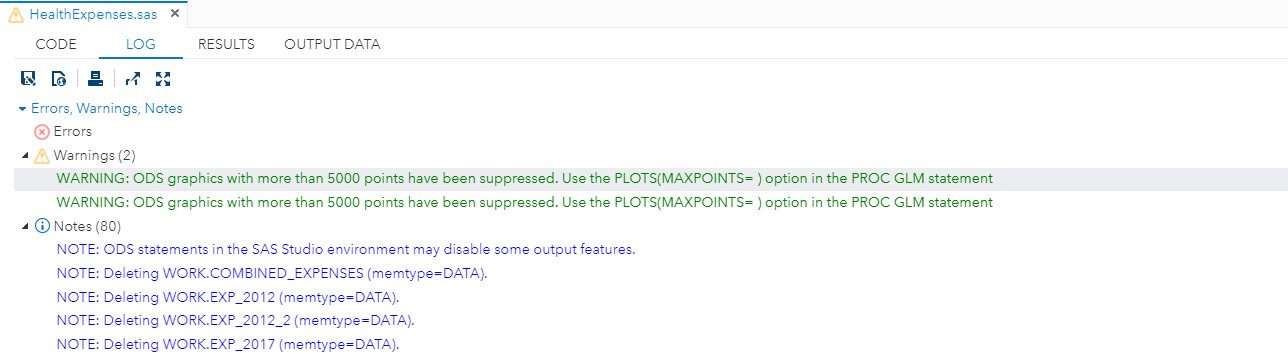
**

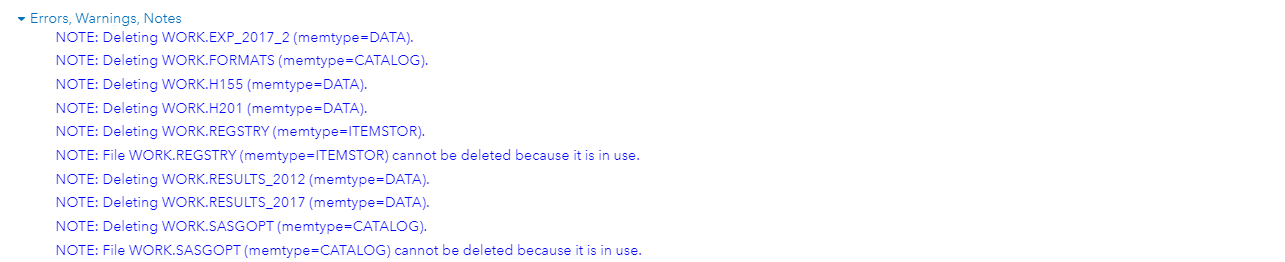
**

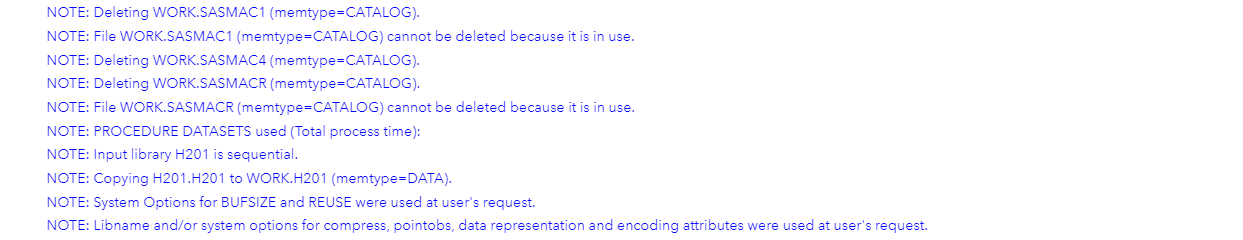
**

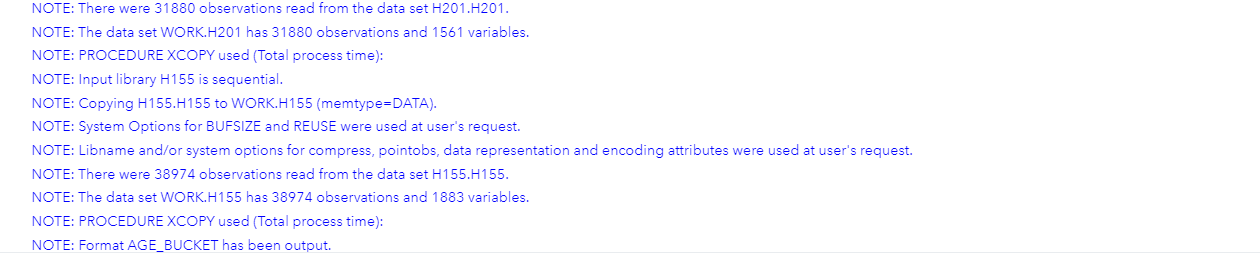
*Figure 1.* SAS Code creating dataset, statistics and visuals

**Code Execution Log**

**

**

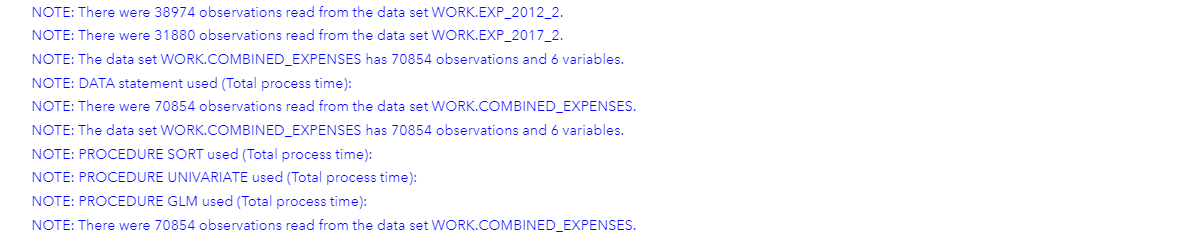
**

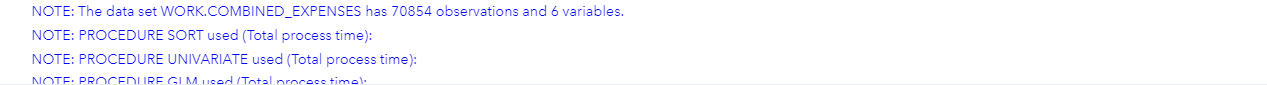
**

**

**

**

**

**

*Figure 2.* Screen shot of SAS Log errors, warnings, and notes

**Results**

Table 1

Summary statistics for 2017 by age bucket (includes records with no expenses)

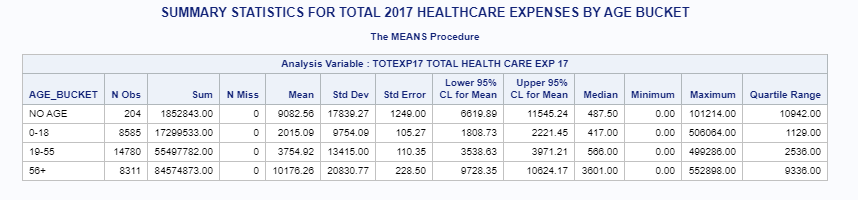
**

Table 2

Summary statistics for 2012 by age bucket (includes records with no expenses

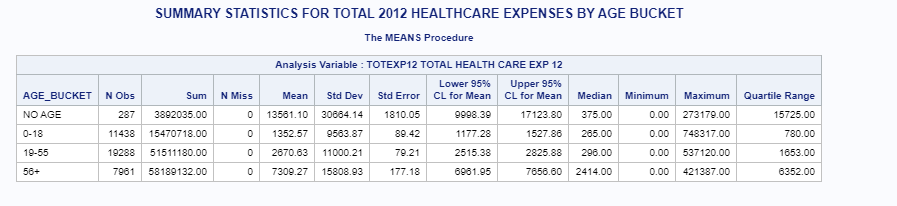


Table 3

Results of Surveymeans procedure for 2017 data (excludes records with no expenses)

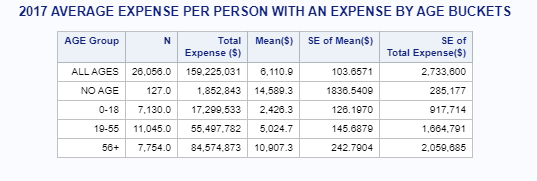


Table 4

Results of Surveymeans procedure for 2012 data (excludes records with no expenses)

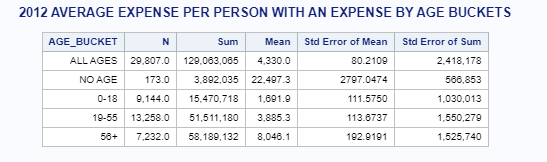
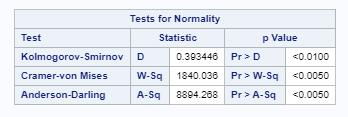
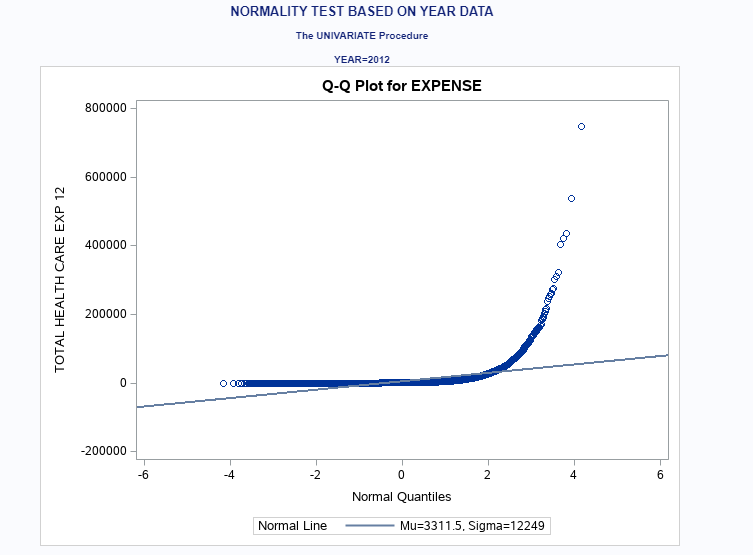


Table 5

Anderson-Darling normality test for 2012 indicate non-normal expense data as the p- value is less than 0.05



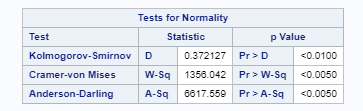


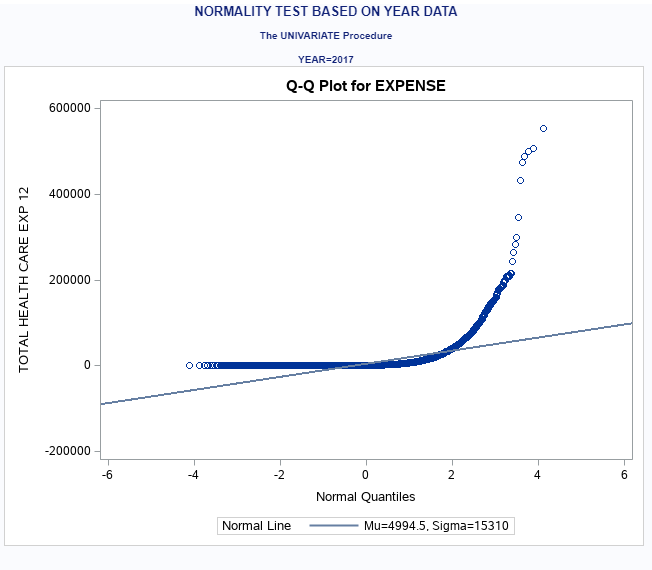
*Figure 3.* Q-Q plot indicating the non-normal data with the long tail at the end

Table 5

Anderson-Darling normality test for 2017 indicate non-normal expense data as the p-

value is less than 0.05





*Figure 4.* Q-Q plot indicating the non-normal data with the long tail at the end

Table 6

Welch’s ANOVA indicating a p-value less than 0.05 for year

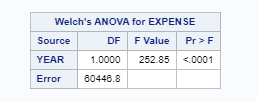


Table 7

Welch’s ANOVA indicating a p-value less than 0.05 for age bucket

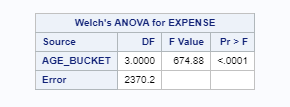
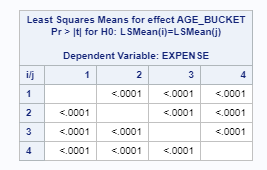
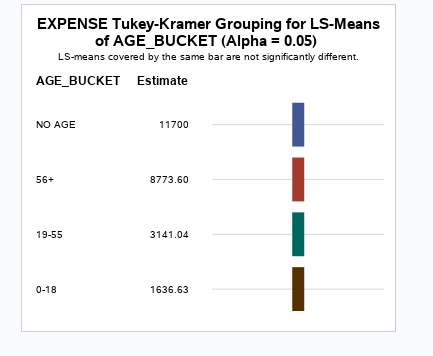


Table8

Least Squares Means indicates a p-value less than 0.05 in comparing all age buckets for 2017





*Figure 5.* Visual representation showing average expenses for each age bucket are

significantly different

**Conclusion**

Based on the results it can be concluded that the null hypothesis of average expenses having no variance for both the year and age bucket scenarios can be rejected. Therefore, it is determined that there is a significant difference between the average expenses between the years 2012 and 2017 as well as between the age buckets for the year 2017. In order to determine this, normality tests were generated in which the results showed that the data for both year and age buckets were not normal. The result of that test lead to using the Welch’s Anova T-test as the best test for non-normal data. Again, both the year and the age bucket results had a p-value less then 0.05 which meant the data is significantly different. Since the age buckets have 4 buckets, the Least Squares Means test was used to determine which age bucket was significantly different. With all buckets having a p-value less then 0.05 when comparing to each other, this shows that all age buckets were significantly different.

REFERENCES

Medical Expenditure Panel Survey. (n.d.). Full Year Consolidated Data File. Retrieved from <https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_results.jsp?cboDataYear=All&cboDataTypeY=1%2CHousehold+Full+Year+File&buttonYearandDataType=Search&cboPufNumber=All>