Portfolio Project

Charles W. Powers

Colorado State University – Global Campus

For the portfolio project I used a dataset from U.S Department of Health and Human Services. The project analyzed data to determine the relationships of dental care visits by income status. In the SAS code I tested a null hypothesis with a two sample T-Test.

The Medical Expenditure Panel Survey (MEPS) website had an abundance of data sets. I used data set H192 from MEPS. This data set has over 34,000 records with 1,900 variables. H192 is the 2016 full year consolidated data file from MEPS. The data set provides medical information collected and is a sample of the civilian noninstitutionalized population of the US (Agency for Healthcare Research and Quality).

The visualization tools I used for this project was within SAS. I have provided graphs to show relationships of the children’s dental care and their income status.

While comparing the health data from DHHS and using the SAS code to run statistical tests I was able to clearly present the comparison of dental visits with lower income children and higher income children. I believe based on the data analytics performed that lower income children have less access to dental care which would assume they have been seen by a dentist less than higher income children. The CDC website states that about 1 of 5 children in ages 5 to 11 years have not been treated by a dentist and have at least one decaying tooth (CDC 2019).

I modified SAS code from one of the exercises that is listed in the HH-SAHRQ MEPS GitHub site. The exercise file that I used from GitHub repository for MEPS was care1\_child\_dental.sas. This SAS file replicated the number and percentage of children with dental care, by poverty status.

I modified the code to keep only records with children of the ages of 2-17 with the Where statement in the data step. I also added a category in the procedure Format step. I added a code to categorize the lower income children (1-3) and higher income children (4-5). This allowed me to test the hypotheses using the categories of income rather than all five categories from the original SAS code.

I also tried to export the data to an excel file to verify the data that was being produced by the report and statistical summaries. I successfully wrote code to produce an excel document but because of the amount of records it took a very long time to produce the excel document. Once I had the excel document, I realized the importance of SAS running statistical test versus using an excel document to examine the data. There was to much data in the excel file to go through each record. I removed the code of exporting the dataset to an excel document because it was not useful to this project.

In order to provide the summary statistics the original code already produced these with the proc surveystep. I did not modify this part of the code. The step calculated estimates of the POVCAT16 which was the poverty category then split into child dental care category and then listed the frequency, weighted frequency, percent, standard error of percent, and the row percent or category percent.

**Charts, Summary Analysis and Statistical Analysis**

A screenshot of a cell phone

Description automatically generated

The above summary analysis shows the relationships between children’s dental care and their families’ income status. Among the ‘Negative or poor’ class, 55.75% have had no dental visits in the past year with only 45.25% having at least one visit. The trend is similar in the ‘Near-poor class’ and ‘Low Income’ class, with a ratio of 54.78% to 46.21% and 50.50% to 49.50% respectively. Among the lower poverty children, the proportion with no dental visit is greater than that the proportion with at least 1 dental visit. This trend reverses in the higher income groups where the proportion with no dental visits are lower than the proportion with at least one dental visit. Among the ‘Middle income’ class, only 46.67% have had no dental visits in the past year while 53.33 % having at least one dental visit. Similar to the ‘High income’ class where only 33.96% have had no dental visits in the past year while 66.04 % having at least one dental visit.

A screenshot of a cell phone

Description automatically generated

The bar plot shows the trend explained by the above summary analysis. The 1st three stacks for each group represent the Lower income children whereas the last two stacks represent the higher income children. The first three stacks in group 1, *No dental visits in the past year*, are relatively higher that the corresponding stacks in Group 2, *One or more dental visits*. The last two stacks in group 1 are relatively lower than the corresponding stacks in group 2.

**Two Sample T-Test**

**Null hypothesis:** Lower poverty children have equal or more dental visits than higher income children. **Alternate hypothesis:** Lower poverty children have fewer dental visits than higher income children.

T TEST SAS code

**proc** **ttest** data=meps alpha = **0.05** sides=L;

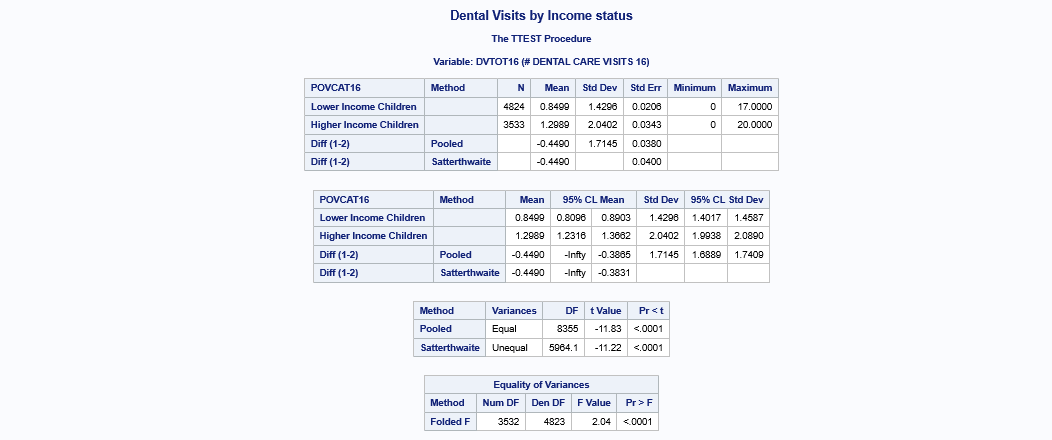
class POVCAT16;

format POVCAT16 POVCAT\_A.;

var DVTOT16;

**run**;

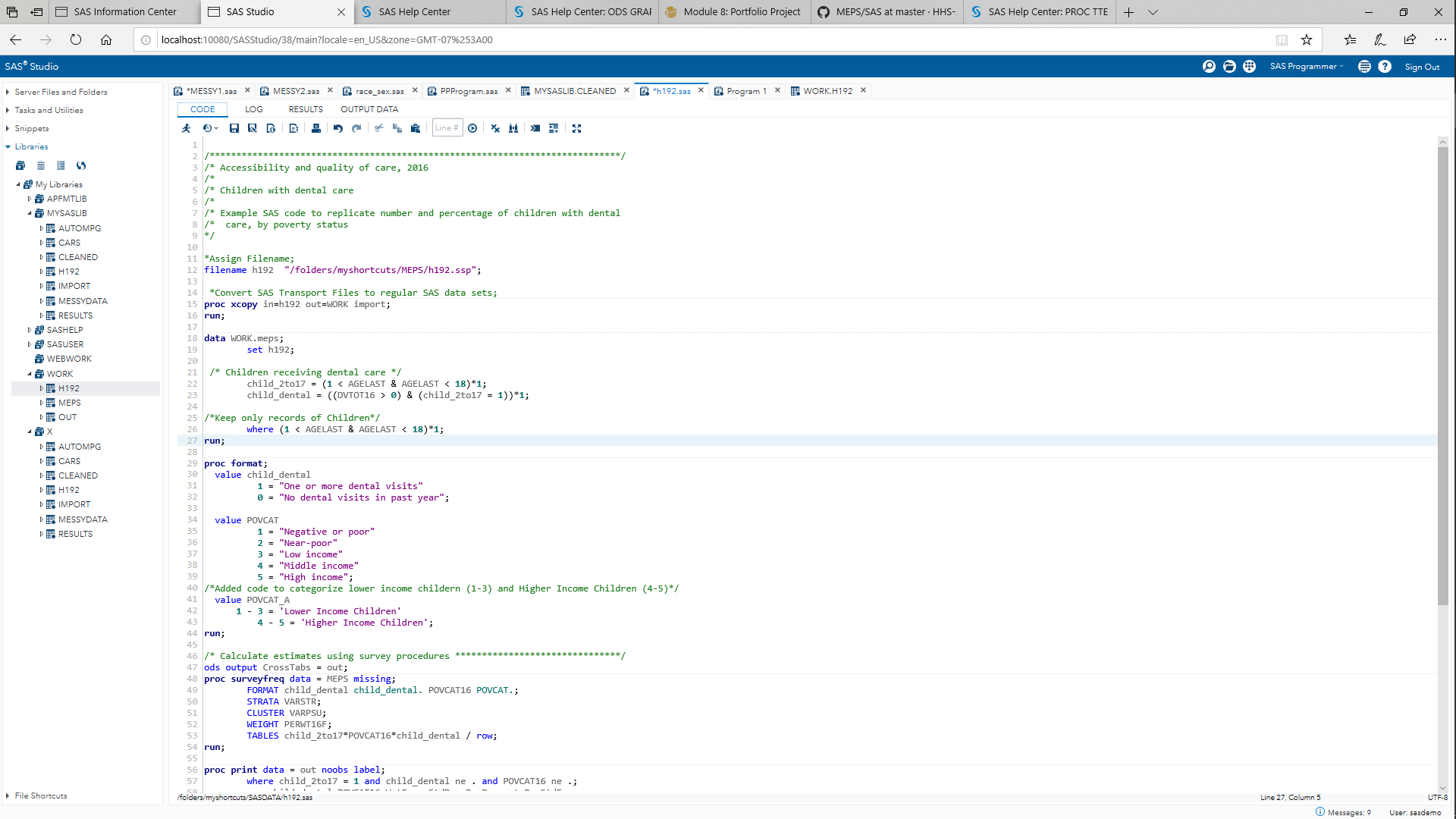
T Execution Output

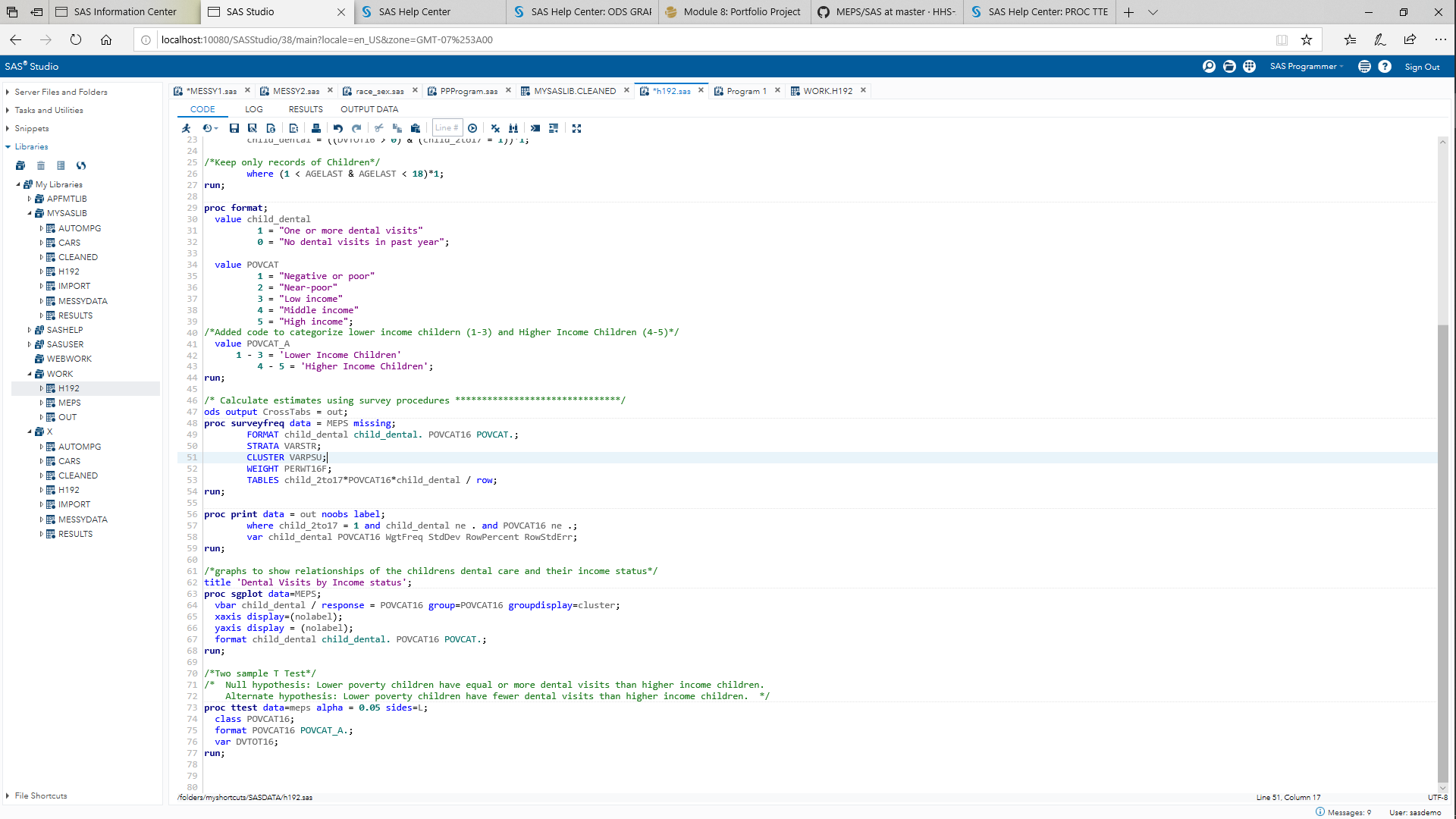


The p-value of the TTEST is less than our alpha significance value of 0.05. We therefore reject the null hypothesis and confirm our hypothesis that lower poverty children ranging from ages 2 – 17 have fewer dental visits than higher income children.

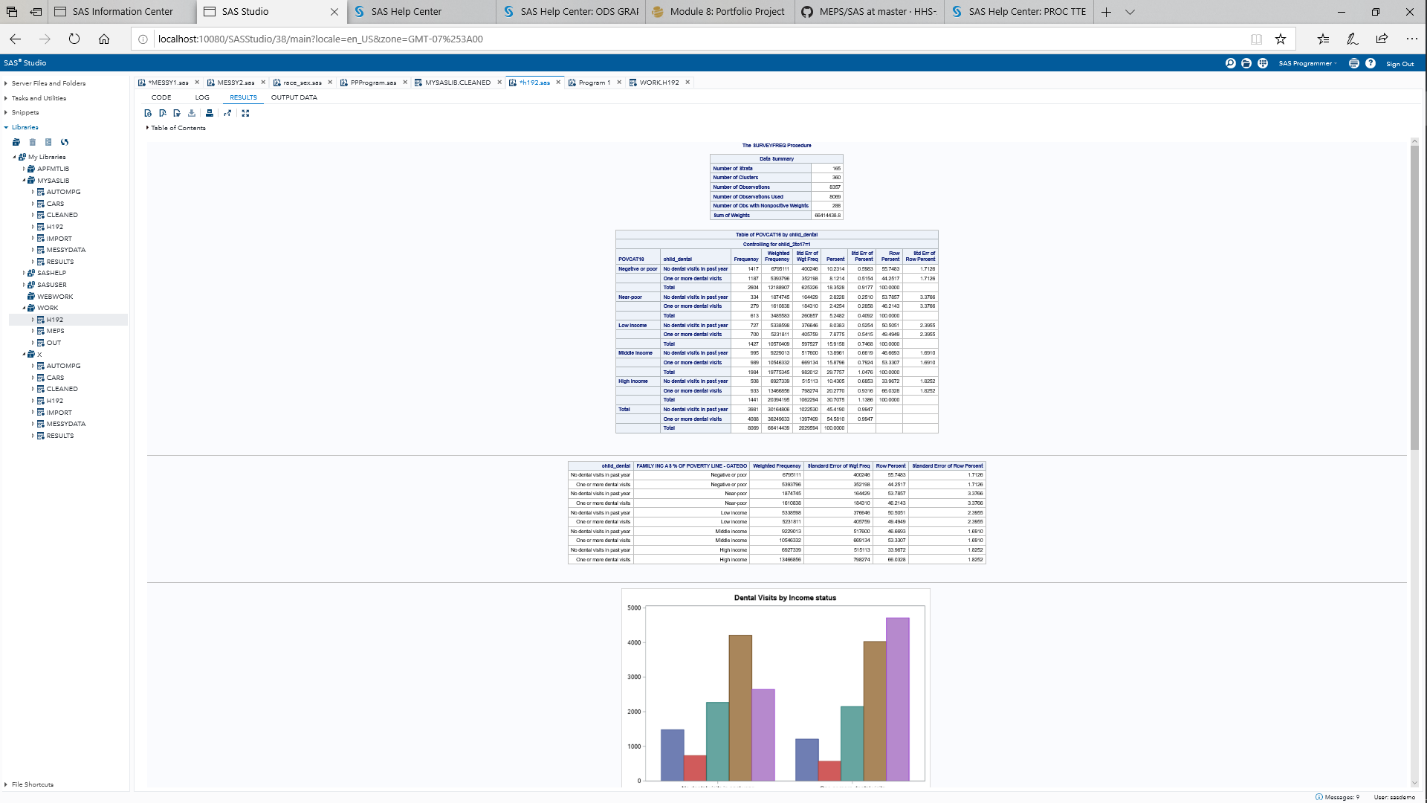
This portfolio project analyzed a data set from the Department of Health and Human services. In this project the SAS code analyzed the data to pull the information on children of the age of 2 – 17 and it also pulled their income level, also known as their poverty level. The code was modified to test a hypotheses of if lower poverty children had less access to health care by analyzing the mean of lower and higher poverty children’s dental visits. I was able to reject the null hypothesis and access the alternative hypotheses that children have less dental visits if they are lower income.

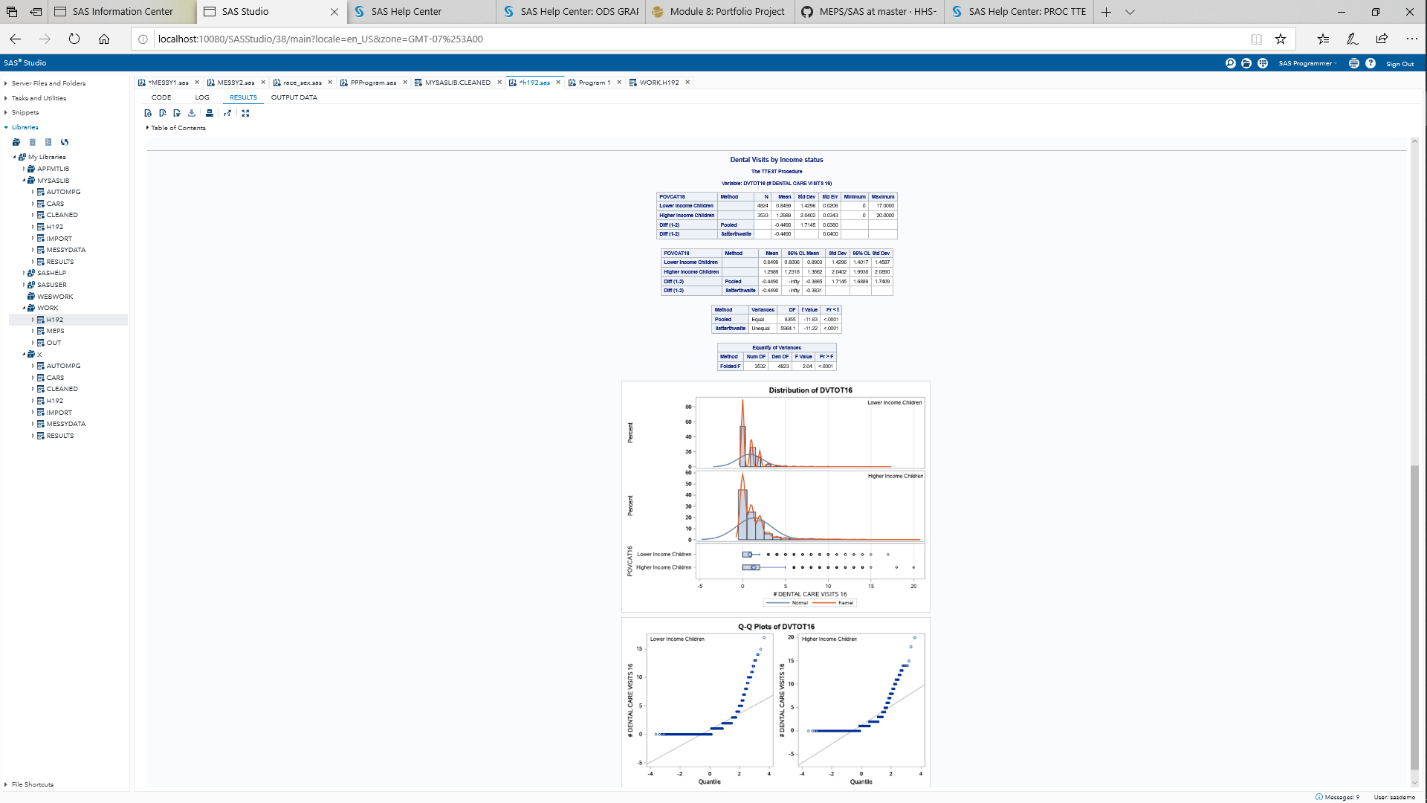
The SAS code used for this project and the data set has been upload to my personal GitHub.com account at <https://github.com/cwp982/PortfolioProject>. In the next two pages I have provided screen shots of the SAS code and the results of the code. The T-Test results are shown on the second part of the SAS results.

**Screenshots from Execution**

****

**Screenshots from Output Result**

****

****

SAS Code with Modification comments:

/\* Accessibility and quality of care, 2016

/\* Children with dental care

/\* Example SAS code to replicate number and percentage of children with dental

/\* care, by poverty status

\*/

\*Assign Filename;

filename h192 "/folders/myshortcuts/MEPS/h192.ssp";

\*Convert SAS Transport Files to regular SAS data sets;

proc xcopy in=h192 out=WORK import;

run;

data WORK.meps;

set h192;

/\* Children receiving dental care \*/

child\_2to17 = (1 < AGELAST & AGELAST < 18)\*1;

child\_dental = ((DVTOT16 > 0) & (child\_2to17 = 1))\*1;

/\*Keep only records of Children\*/

where (1 < AGELAST & AGELAST < 18)\*1;

run;

proc format;

value child\_dental

1 = "One or more dental visits"

0 = "No dental visits in past year";

value POVCAT

1 = "Negative or poor"

2 = "Near-poor"

3 = "Low income"

4 = "Middle income"

5 = "High income";

/\*Added code to categorize lower income childern (1-3) and Higher Income Children (4-5)\*/

value POVCAT\_A

1 - 3 = 'Lower Income Children'

4 - 5 = 'Higher Income Children';

run;

/\* Calculate estimates using survey procedures \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

ods output CrossTabs = out;

proc surveyfreq data = MEPS missing;

FORMAT child\_dental child\_dental. POVCAT16 POVCAT.;

STRATA VARSTR;

CLUSTER VARPSU;

WEIGHT PERWT16F;

TABLES child\_2to17\*POVCAT16\*child\_dental / row;

run;

proc print data = out noobs label;

where child\_2to17 = 1 and child\_dental ne . and POVCAT16 ne .;

var child\_dental POVCAT16 WgtFreq StdDev RowPercent RowStdErr;

run;

/\*graphs to show relationships of the childrens dental care and their income status\*/

title 'Dental Visits by Income status';

proc sgplot data=MEPS;

vbar child\_dental / response = POVCAT16 group=POVCAT16 groupdisplay=cluster;

xaxis display=(nolabel);

yaxis display = (nolabel);

format child\_dental child\_dental. POVCAT16 POVCAT.;

run;

/\*Two sample T Test\*/

/\* Null hypothesis: Lower poverty children have equal or more dental visits than higher income children.

Alternate hypothesis: Lower poverty children have fewer dental visits than higher income children. \*/

proc ttest data=meps alpha = 0.05 sides=L;

class POVCAT16;

format POVCAT16 POVCAT\_A.;

var DVTOT16;

run;

**References:**

Agency for Healthcare Research and Quality, Retrieved from https://meps.ahrq.gov/mepsweb/data\_stats/download\_data\_files.jsp

Center for Disease Control and Prevention (CDC) Children's Oral Health. (2019, May 14). Retrieved from https://www.cdc.gov/oralhealth/basics/childrens-oral-health/index.html

Elliot, A., & Woodward, W. (2016). SAS Essentials: Mastering SAS for data analytics (2 nd ed.). Hoboken, NJ: Wiley Publishing. ISBN-13: 978-1119042167