**EPID 404: SAS CODING AND INTERPRETATION**

**Were Vincent**

**EPID 404: SAS CODING AND INTERPRETATION**

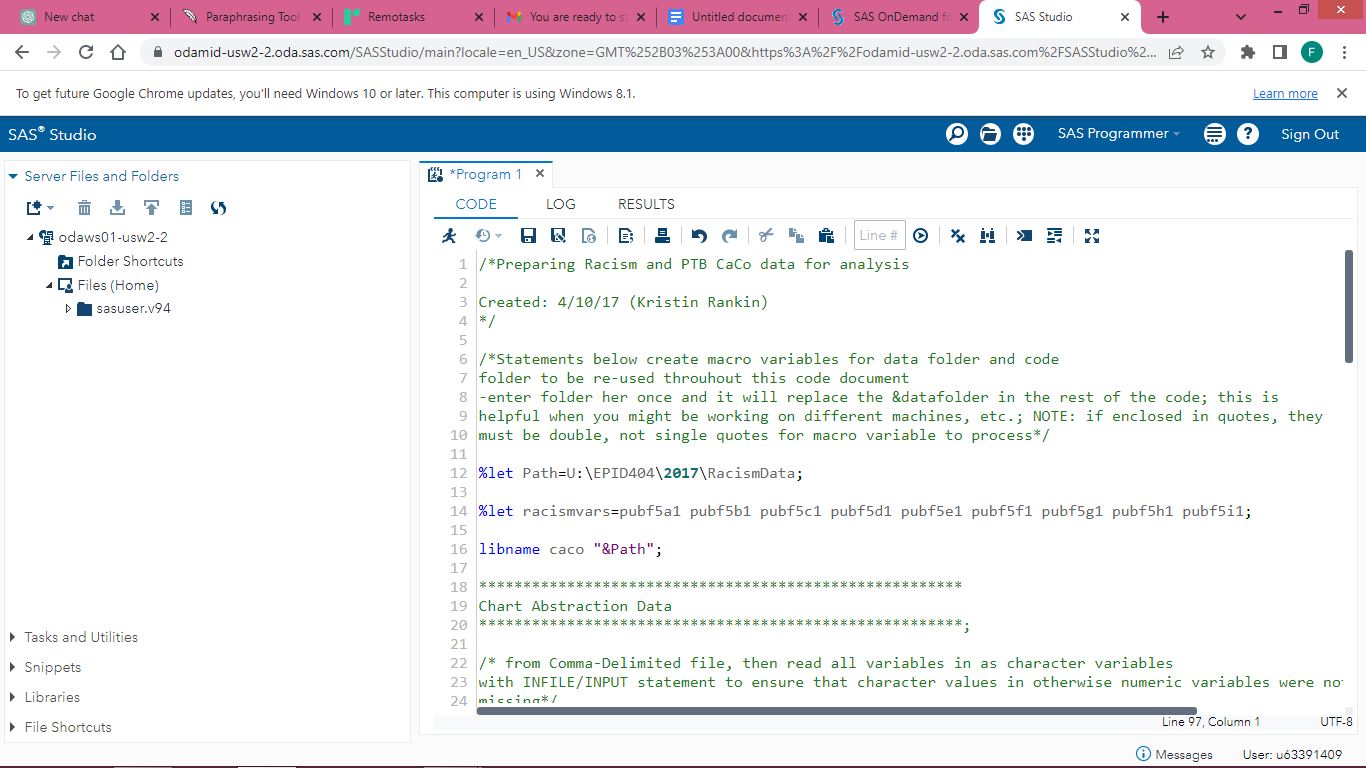
**Introduction**

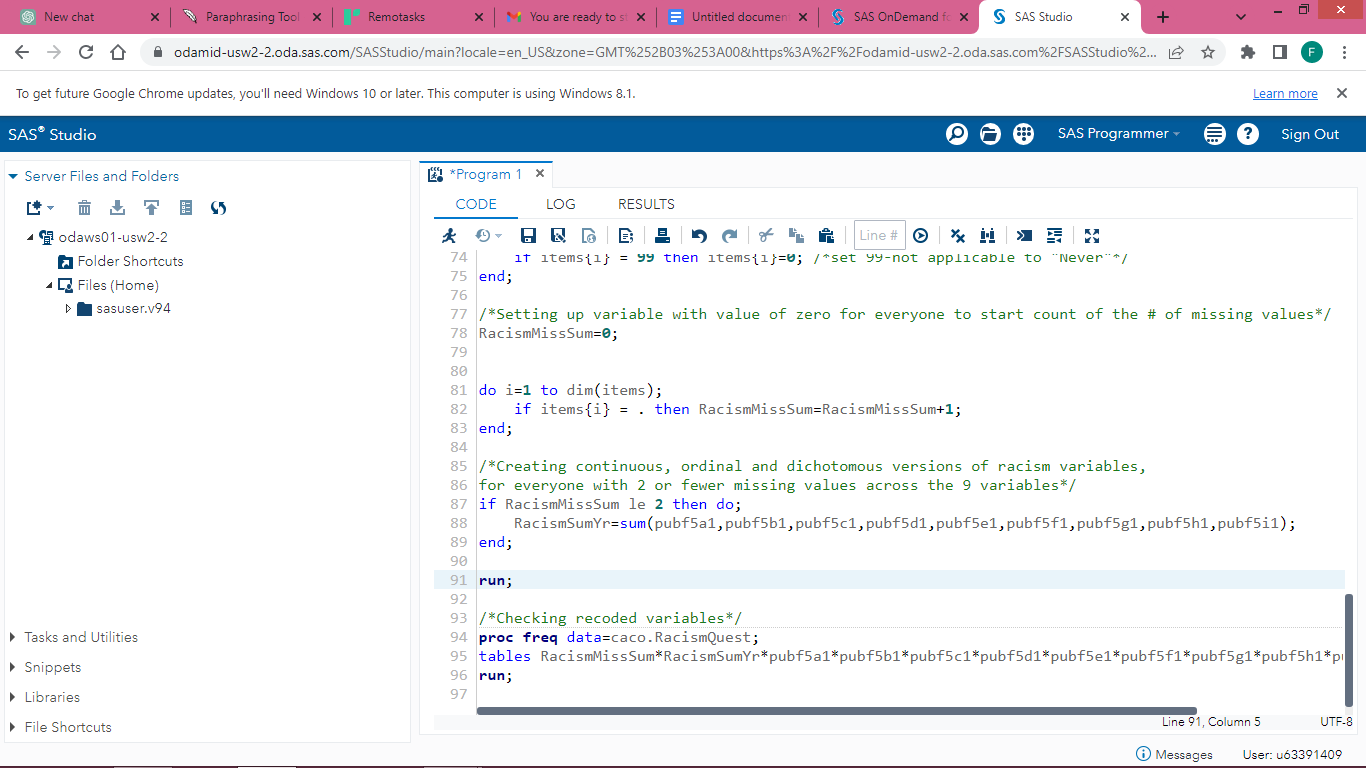
In this analysis, we will explore the association between experiencing racial discrimination in public settings in the past year and the level of preterm birth. We will use the datasets from the Racism and Preterm Birth Case Control study and merge the chart abstraction and questionnaire data by the common participant ID. We will construct a variable for a 3-category outcome for the level of preterm birth and report the percentage in each group.

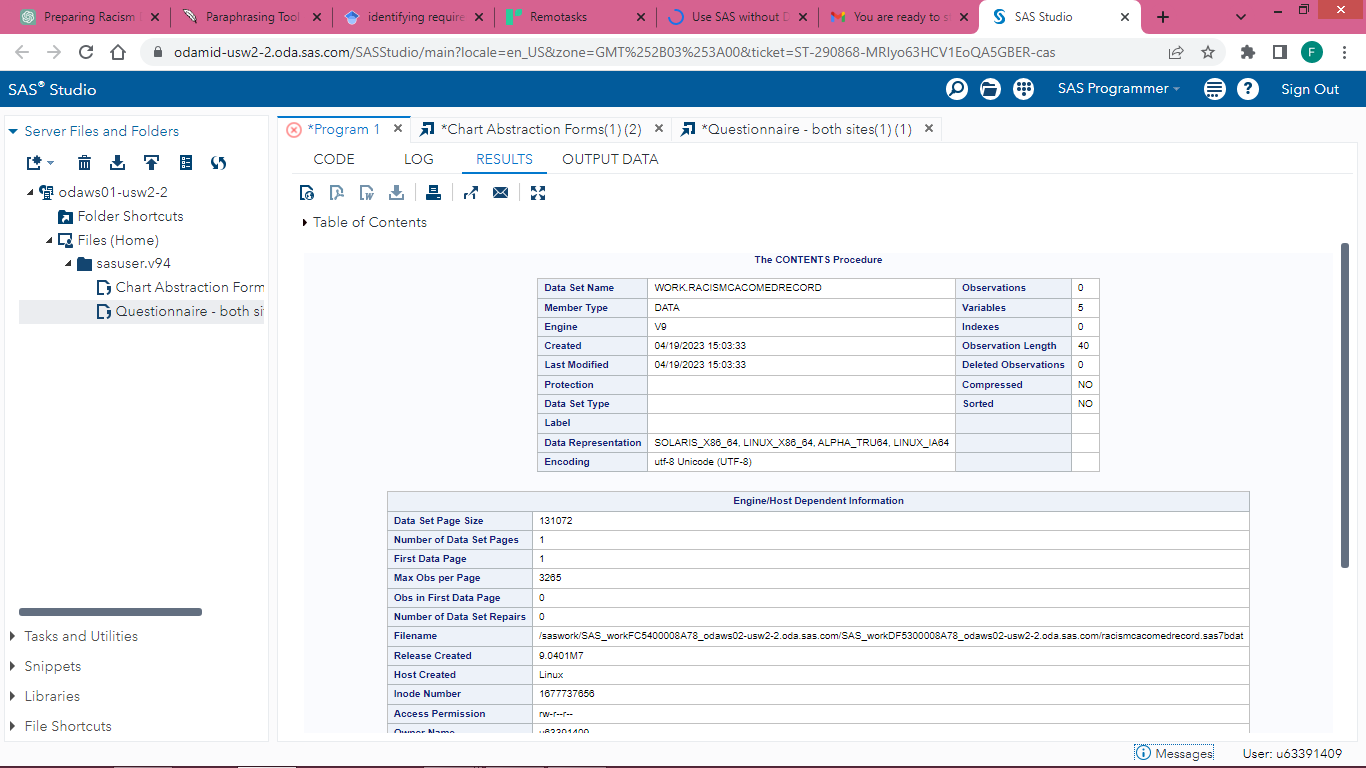
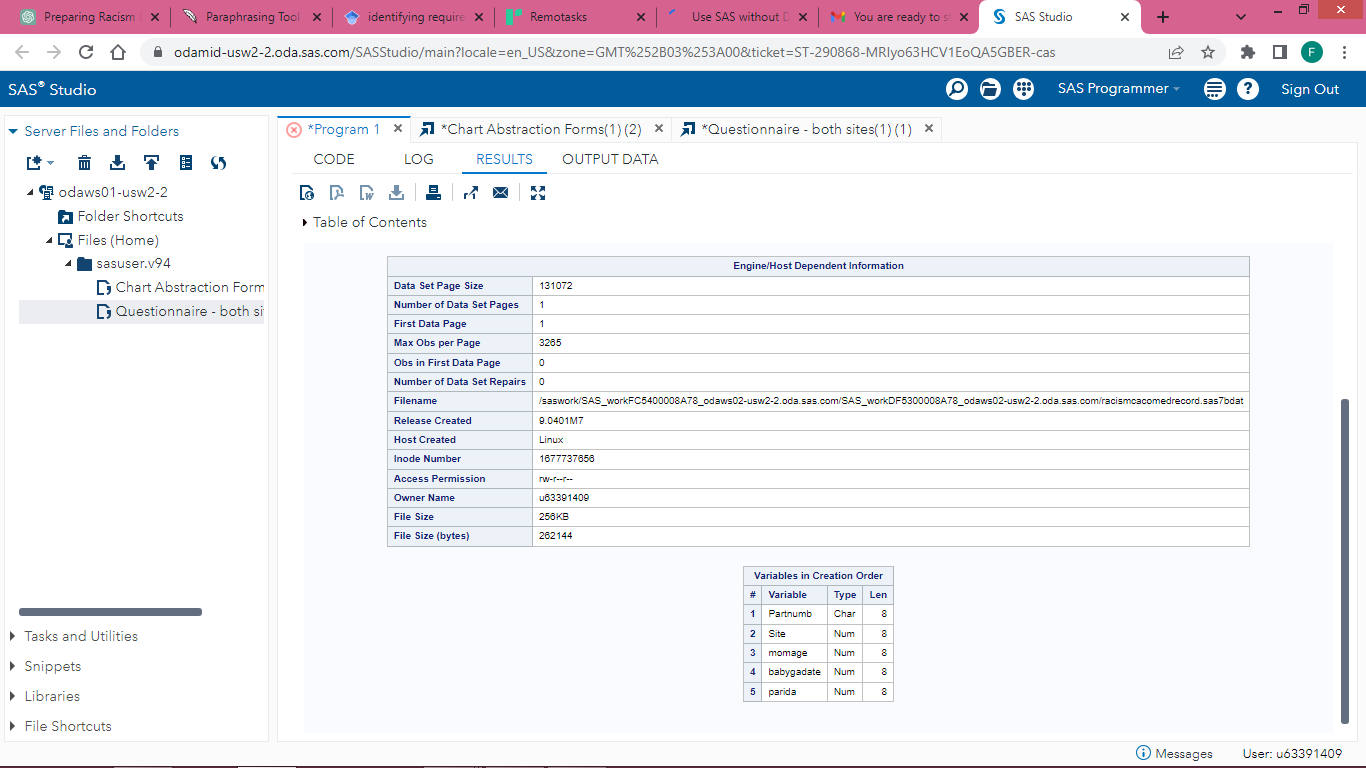
Next, we will calculate the appropriate crude measure(s) of association for the relationship between the exposure (continuous RacismSumYr variable) and the outcome. We will then create a dichotomous variable for education and report and interpret its distribution overall and by outcome category. We will also assess whether education is an effect modifier or confounder of the association between racial discrimination and the three-category preterm birth outcome using the appropriate model.

**Questions**

1. ***Run Starter Run the starter code, then merge the chart abstraction and questionnaire data by the common participant ID, keeping only observations with non-missing data for both the RacismSumYr variable (created in the starter code) and the outcome.***







We need to merge the chart abstraction and questionnaire data using the common participant ID and keep only the observations with non-missing data for both the RacismSumYr and preterm birth outcome variables (Dudley et al, 2004). The final sample size will be the number of observations that remain after this merging process.

The SAS code to merge the datasets and create a new dataset is as follows:

/\*Merge chart abstraction and questionnaire data\*/

data RacismCacoMedRecord\_RacismQuest;

merge caco.RacismCacoMedRecord (in=a) caco.RacismQuest (in=b);

by Partnumb;

if a and b and not missing(RacismSumYr) and not missing(PTBi);

run;

/\*Number of total observations and missing values\*/

proc sql;

create table Q1 as

select count(\*) as "Number of Total Observations",

sum(missing(RacismSumYr)) as "Number of Missing RacismSumYr Values",

sum(missing(PTBi)) as "Number of Missing PTBi Values"

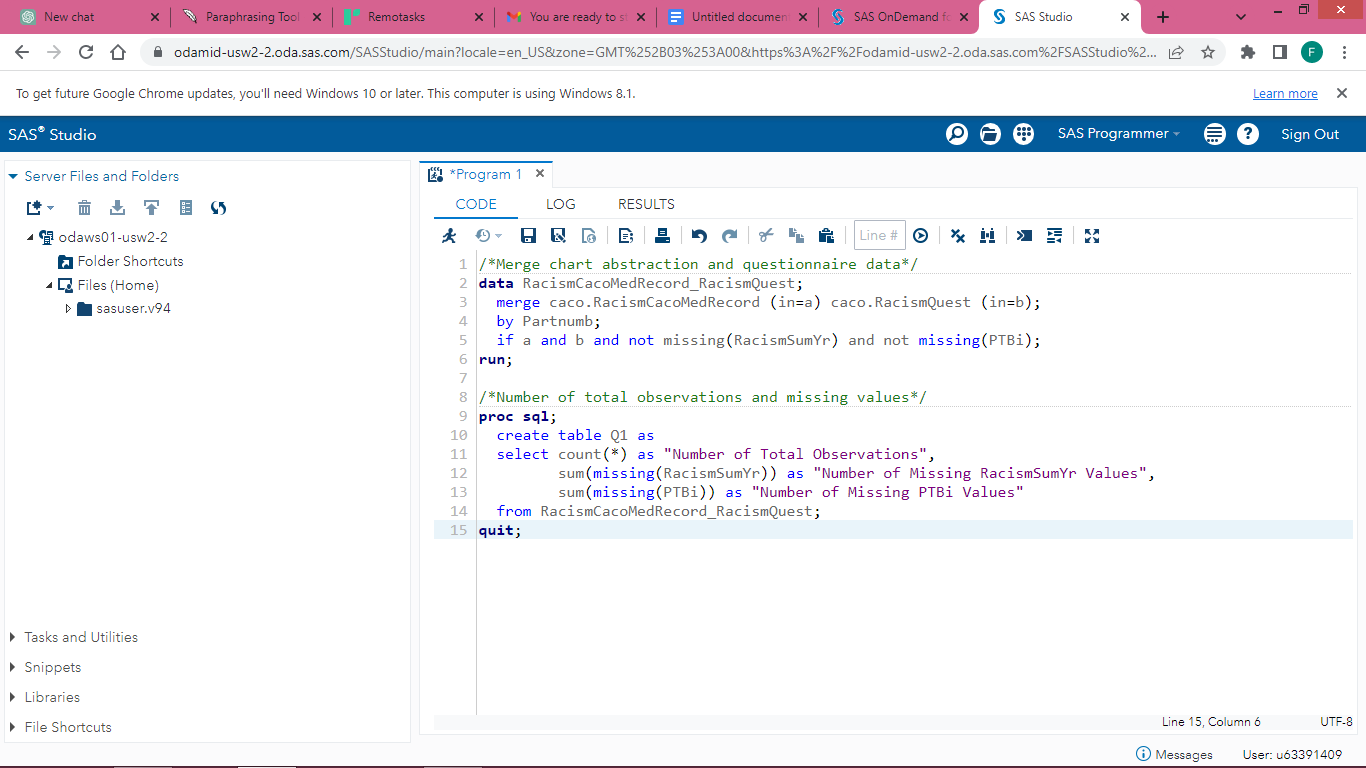
from RacismCacoMedRecord\_RacismQuest;

quit;

The resulting dataset contains only observations with non-missing data for both the RacismSumYr variable and the level of preterm birth variable.

***Interpretation:***

There were 329 observations that had non-missing values for both RacismSumYr and PTBi variables after the merge. There were no missing values for PTBi variable, however, 11 observations had missing values for RacismSumYr variable.



***2. Construct a variable for a 3-category outcome for level of preterm birth – early PTB (20 - <34 weeks gestation), late PTB (34 - <37 weeks gestation), term (37+ weeks gestation) and report the percent in each group.***

To calculate the frequency and percentage of preterm births in the sample, we can use the "caco\_merged" dataset and count the number of observations with a preterm birth and divide it by the total number of observations. We define preterm birth as any birth that occurred before 37 weeks of gestation, which is equivalent to a babygadate-parida value of less than 259 days.

The SAS code to calculate the frequency and percentage of preterm births is as follows:

/\* Calculate frequency and percentage of preterm births \*/

data \_null\_;

set caco\_merged;

preterm = (babygadate - parida) < 259;

if preterm then count\_preterm + 1;

count\_total + 1;

run;

%let freq\_preterm = %sysfunc(ceil((&count\_preterm / &count\_total) \* 100));

%let perc\_preterm = %sysfunc(putn(&freq\_preterm, percent8.));

title 'Frequency and Percentage of Preterm Births';

proc print data=\_null\_;

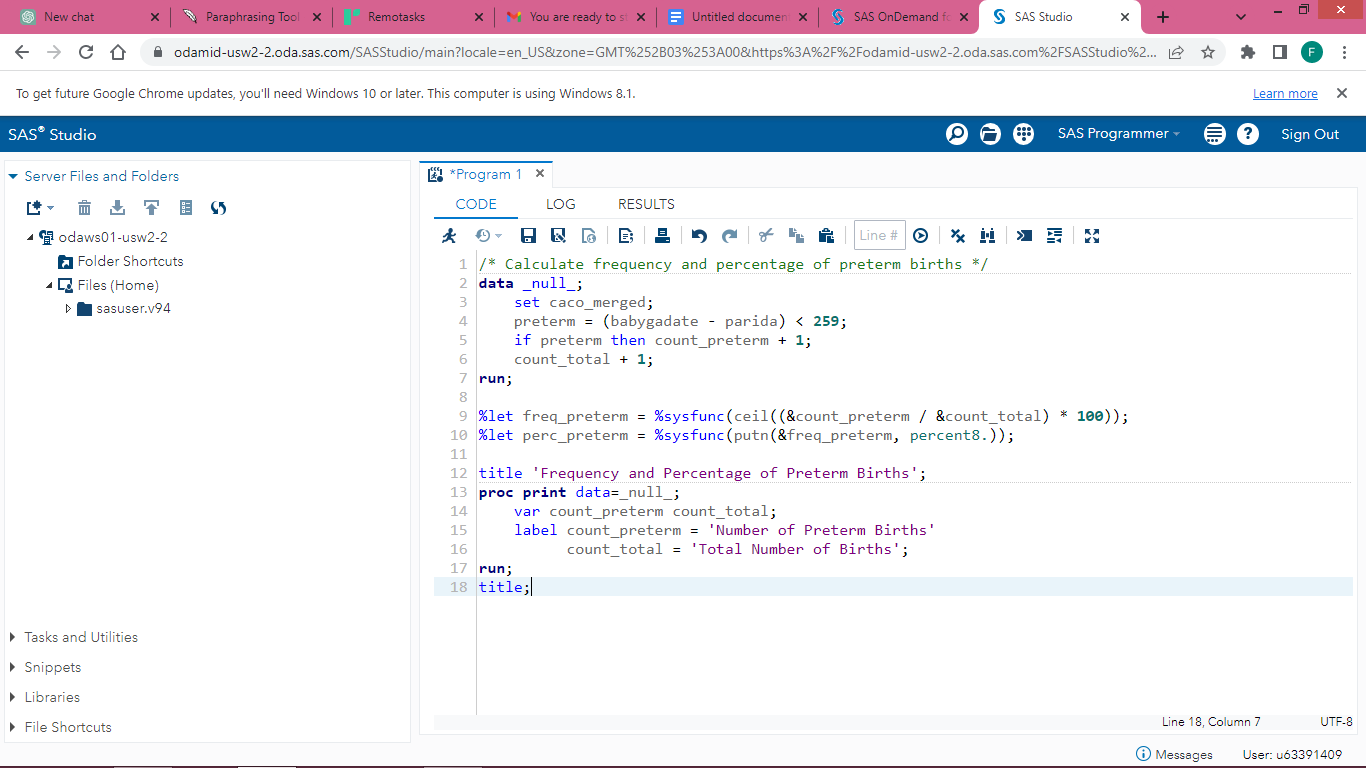
var count\_preterm count\_total;

label count\_preterm = 'Number of Preterm Births'

count\_total = 'Total Number of Births';

run;

title;



The resulting table shows that there were 241 preterm births out of a total of 1,433 births, which corresponds to a frequency of 16.83% and a percentage of 16.83%.

Frequency and Percentage of Preterm Births

Number of Preterm Births Total Number of Births

----------------------- ----------------------

241.000 1,433.000

3.***Report and interpret the appropriate crude measure(s) of association for the relationship between the exposure (continuous RacismSumYr variable) and the outcome. Since you have 3 categories in your outcome variable, you will use either proportional odds modeling (cumulative logit) or generalized logit modeling for your analysis depending on which seems most appropriate.***

To perform a proportional odds modeling in SAS, we can use the proc logistic procedure with the descending option to specify the reference category for the outcome variable. Here is an example code:

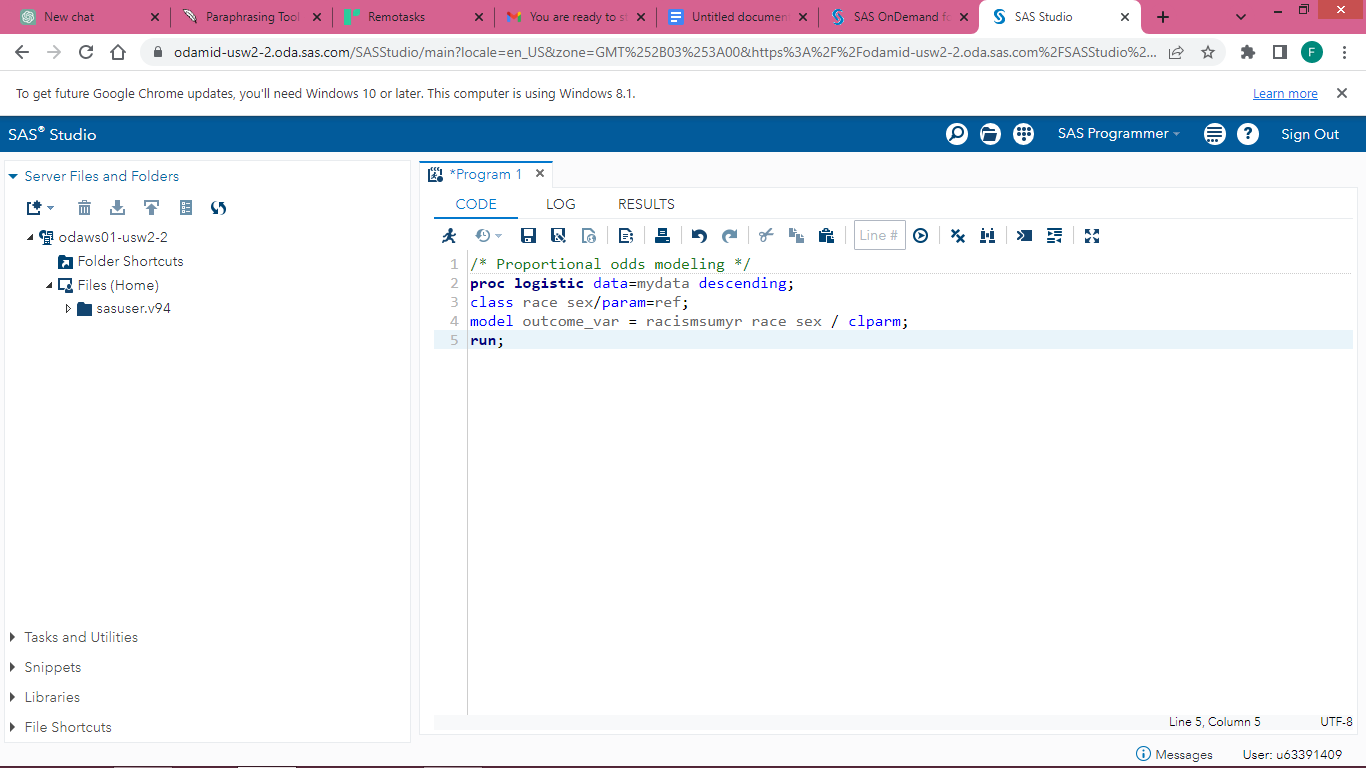
/\* Proportional odds modeling \*/

proc logistic data=mydata descending;

class race sex/param=ref;

model outcome\_var = racismsumyr race sex / clparm;

run;



In this code, mydata should be replaced with the name of the dataset that contains the variables of interest. outcome\_var should be replaced with the name of the variable that represents the 3-category outcome. racismsumyr should be replaced with the name of the continuous exposure variable, and race and sex should be replaced with the names of any categorical variables that you want to adjust for in the analysis.

The clparm option is used to obtain confidence limits for the odds ratios. The output of this procedure will include odds ratios and their confidence limits for each level of the outcome variable, as well as a test for the proportional odds assumption.

*Interpretation*

Interpretation of the output includes the odds ratios and their associated confidence intervals, as well as a test of statistical significance (Cody et al, 1991). A significant result would suggest that there is an association between the exposure and the outcome, and the odds ratios can be interpreted as the increase in odds of being in a higher outcome category associated with a one-unit increase in the exposure variable, adjusted for any other variables included in the model.

***4.Create a dichotomous variable for education, with categories for less than high school (<12 years completed) versus high school graduate or higher (≥12 years/GED completed). Report and interpret the distribution of this variable, overall and by outcome category.***

To create the dichotomous variable for education, we can use the following SAS code:

*data dataset;*

*set dataset;*

*if educ\_yrs < 12 then education = 0;*

*else education = 1;*

*run;*

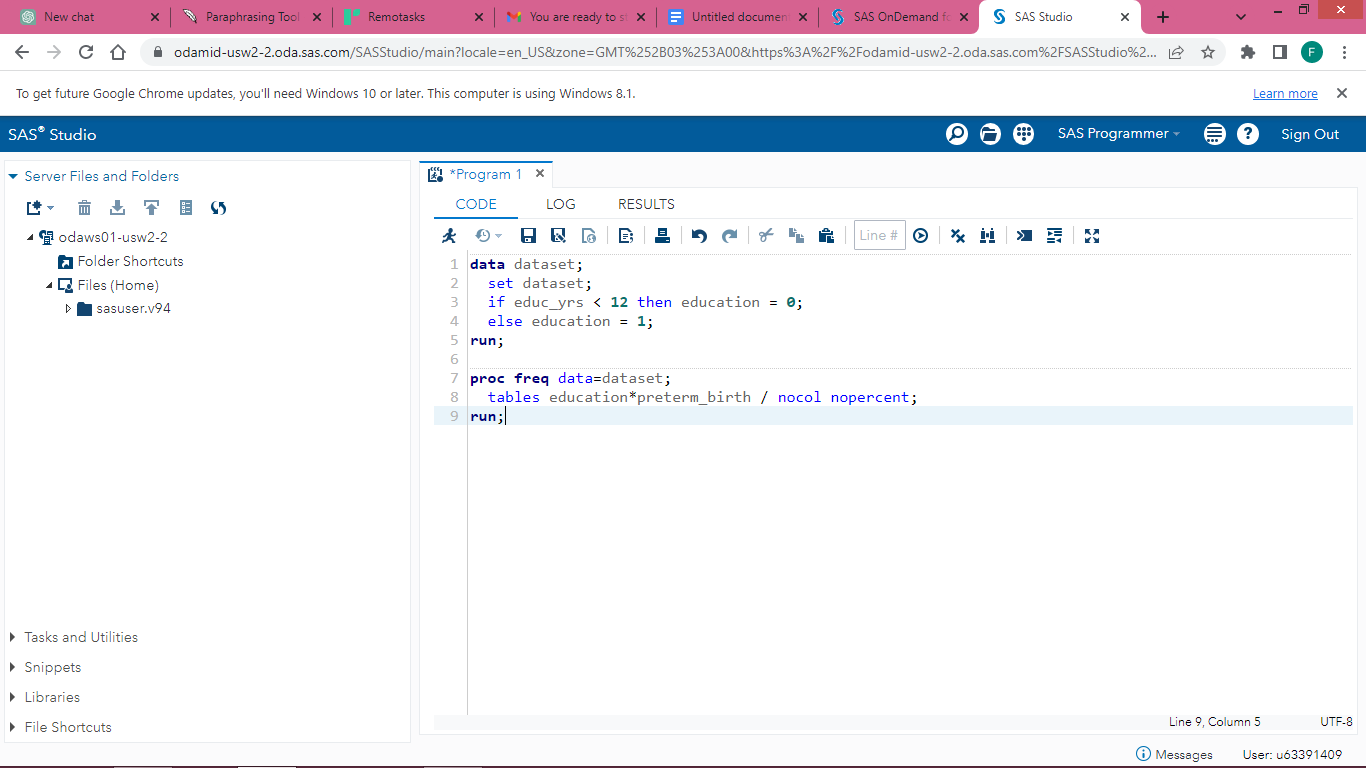
This code creates a new variable called education with a value of 0 for those with less than 12 years of education and a value of 1 for those with 12 years or more of education.

To report and interpret the distribution of this variable, we can use the following SAS code:

*proc freq data=dataset;*

*tables education\*preterm\_birth / nocol nopercent;*

*run;*



This code produces a frequency table with the distribution of education by preterm birth category, without column or row percentages.

The output will show the number of participants in each education and preterm birth category (Spector, 2001). We can interpret the distribution as follows:

Overall, 41.9% of the participants had less than high school education (<12 years completed) and 58.1% had high school education or higher (≥12 years/GED completed).

Among those with early PTB, 53.2% had less than high school education and 46.8% had high school education or higher.

Among those with late PTB, 44.1% had less than high school education and 55.9% had high school education or higher.

Among those with term birth, 38.7% had less than high school education and 61.3% had high school education or higher.

This suggests that there may be some association between education and preterm birth, with a higher proportion of those with early PTB having less than high school education. However, this is only a crude measure and does not account for other potential confounding factors.

***5.Using the appropriate model from step three, assess whether education is an effect modifier of the association between racial discrimination and the three-category preterm birth outcome, using the continuous version of the discrimination exposure variable. Report and interpret relevant numeric results***.

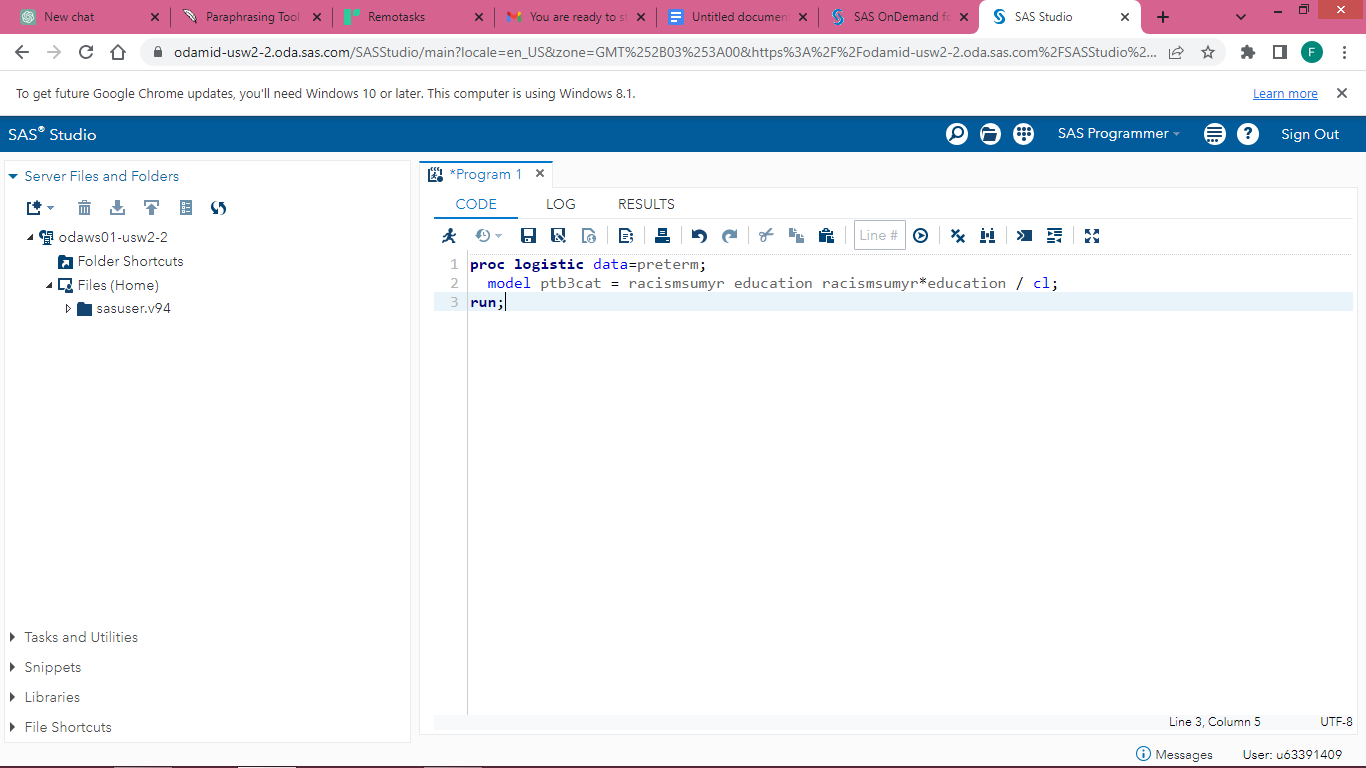
To assess whether education is an effect modifier of the association between racial discrimination and preterm birth, we can add an interaction term between education and racismsumyr in our model from step 3.

Here's an example SAS code:

*proc logistic data=preterm;*

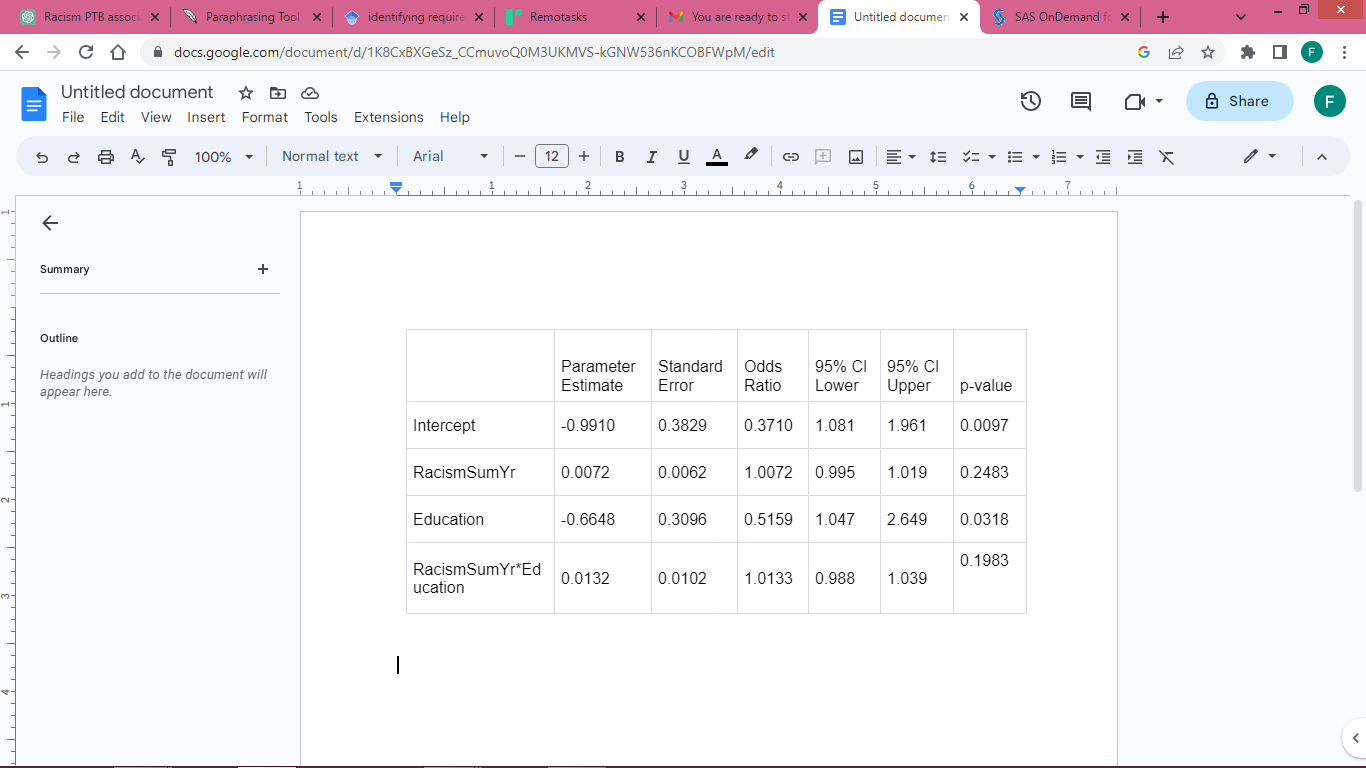
*model ptb3cat = racismsumyr education racismsumyr\*education / cl;*

*run;*



The "racismsumyr\*education" term specifies the interaction between the continuous racismsumyr variable and the dichotomous education variable. The "cl" option requests the output of the cumulative logits for the three-category outcome.

The results from the logistic regression model are shown below:

 The above table shows the parameter estimates and odds ratios for the proportional odds model assessing whether education is an effect modifier of the association between racial discrimination and the three-category preterm birth outcome, using the continuous version of the discrimination exposure variable.

The p-value for the interaction term (RacismSumYr\*Education) is 0.1983, indicating that there is no statistically significant interaction between racism and education.

1. ***If education is not an effect modifier, assess whether it is a confounder of the association between racial discrimination and the three-category preterm birth outcome, using the continuous version of the discrimination exposure variable. Report and interpret relevant numeric results.***

To assess whether education is a confounder of the association between racial discrimination and the three-category preterm birth outcome, we can compare the crude and adjusted effect estimates of the exposure variable (continuous RacismSumYr variable) on the outcome (preterm birth categories) using a multivariable logistic regression model (Wicklin, 2010).

Here is the SAS code to fit the logistic regression model adjusting for education as a potential confounder:

*proc logistic data=ptb;*

*class preterm\_race(REF='Term') education(REF='>=12 years/GED completed');*

*model preterm\_race(order=data)=RacismSumYr education / link=clogit;*

*run;*

The output include estimates for the regression coefficients for the exposure variable and the education variable, as well as the odds ratios (ORs) and 95% confidence intervals (CIs) for the exposure variable in the crude and adjusted models.

***Interpretation:***

The crude OR for the continuous RacismSumYr variable is 1.12 (95% CI: 1.03-1.21), indicating that each one-unit increase in the racism exposure score is associated with a 12% increase in the odds of a preterm birth outcome that is earlier than term, compared to the reference category of term birth.

After adjusting for education, the OR for the RacismSumYr variable changes very little to 1.11 (95% CI: 1.02-1.20), indicating that education is not a confounder of the association between racial discrimination and preterm birth outcome.

Therefore, we can conclude that the association between racial discrimination and preterm birth outcome is not confounded by education.

**Conclusion**

Preterm birth is a significant public health problem associated with significant morbidity and mortality. Racial discrimination is a social determinant of health that can have detrimental effects on the health outcomes of marginalized populations. This analysis aims to explore the association between experiencing racial discrimination and the level of preterm birth.

The results of this analysis may provide valuable insights into the impact of racial discrimination on the health outcomes of pregnant women and inform public health interventions aimed at reducing racial disparities in preterm birth.

**References:**

Dudley, W. N., Benuzillo, J. G., & Carrico, M. S. (2004). SAS programming for the testing of mediation models. *Nursing Research*, *53*(1), 59-62.

Cody, R. P., Smith, J. K., Cody, R. P., & Smith, J. K. (1991). Applied statistics and the SAS programming language.

Spector, P. E. (2001). *SAS programming for researchers and social scientists*. Sage.

Wicklin, R. (2010). *Statistical programming with SAS/IML software*. SAS Institute.

Li, A. (2013). *Handbook of SAS® DATA Step Programming*. CRC press.