Lab 12 *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention* `LastnameFirstinitial_Lab#.sas` *(e.g.,* `PileggiS_Lab12.sas`*).*

This US Surgeon General's warning was placed on the side of cigarette packages beginning in 1985. Prior to the placement of the warning, studies had to be conducted to investigate the effects of smoking during pregnancy. The data provided are part of the Child Health and Development Studies, which was a comprehensive investigation of all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San Francisco-East Bay area. Despite the warnings which went into effect in 1985, the National Center for Health Statistics found that 15% of women who gave birth in 1996 smoked during their pregnancy. (What year were you born?)

Why do we care about baby birth weight? Birth weight is a measure of a baby's maturity. Typically, smaller babies have lower survival rates than than larger babies. In this study, the rate at which babies died within 28 days of birth was 150 per thousand births for low birth weight babies, compared to 5 per thousand for babies of 'normal' weight. Babies that weigh under 5.5 pounds are considered of low birth weight.

`babies.csv`

`bwt`	baby's weight at birth in ounces
`gestation`	length of pregnancy in days
`parity`	0=first born, 1=otherwise
`age`	mother's age in years
`height`	mother's height in inches
`weight`	mother's pregnancy weight in pounds
`smoke`	smoking status of mother: 0=not now, 1=yes now

1. Assign the computer location of your STAT 330 data set to a macro variable called `path`.

2. Read the `babies.csv` data into SAS using your `path` macro variable. Write a couple of procedures to familiarize yourself with the data.

Follow the subsequent steps to export summary statistics used to investigate the relationship between mother's smoking status and baby birth weight.

3. Use one SAS procedure to obtain the sample size, minimum, maximum, sample mean, and sample standard deviation of baby birth weight separately for smoking and non-smoking mothers. Based on your descriptive statistics, does there appear to be a relationship between mother's smoking status and baby birth weight? Note your

observations in a comment in your SAS code. Your summary statistics should appear as shown below.

Analysis Variable : bwt						
smoke	N Obs	N	Mean	Std Dev	Minimum	Maximum
0	742	742	123.0471698	17.3986888	55.0000000	176.0000000
1	484	484	114.1095041	18.0989457	58.0000000	163.0000000

4. Modify your code from the previous question to create a SAS data set with the summary statistics. Print this data set - your results should match those shown below.

Obs	smoke	_TYPE_	_FREQ_	_STAT_	bwt
1	.	0	1226	N	1226
2	.	0	1226	MIN	55
3	.	0	1226	MAX	176
4	.	0	1226	MEAN	119.5187602
5	.	0	1226	STD	18.203567009
6	0	1	742	N	742
7	0	1	742	MIN	55
8	0	1	742	MAX	176
9	0	1	742	MEAN	123.04716981
10	0	1	742	STD	17.398688778
11	1	1	484	N	484
12	1	1	484	MIN	58
13	1	1	484	MAX	163
14	1	1	484	MEAN	114.10950413
15	1	1	484	STD	18.098945686

5. Rearrange the summary statistics data such that the columns correspond to the statistics and the rows correspond to mother smoking status. Note that a period (.) indicates the overall summary statistics, regardless of mother's smoking status. Exclude the overall summary statistics when you re-arrange the data. Your summary statistics data should appear as shown below.

Obs	smoke	_NAME_	N	MIN	MAX	MEAN	STD
1	0	bwt	742	55	176	123.04716981	17.398688778
2	1	bwt	484	58	163	114.10950413	18.098945686

6. Modify the summary statistics data.

   (a) Make sure that all means and standard deviations are formatted to one decimal place.

   (b) Create and apply a format such that 0 displays as "Non-smoking mothers" and 1 displays as "Smoking mothers".

   (c) Only retain relevant variables/observations.

   Your final summary statistics data set should appear as:

Obs	smoke	N	MIN	MAX	MEAN	STD
1	Non-smoking mothers	742	55	176	123.0	17.4
2	Smoking mothers	484	58	163	114.1	18.1

7. Export the summary statistics data to `BabiesSummaryStats.csv` using the `path` macro variable. Open the `BabiesSummaryStats.csv` file and verify that everything appears as intended. Upload the `BabiesSummaryStats.csv` file to PolyLearn in addition to your SAS program.

At some point in your career, it may be necessary to create a succinct table of essential statistical results. Follow the subsequent steps to get essential output from a two-sample t-test.

8. Perform a two-sample t-test to determine if we have evidence of a difference in population mean birth weight among babies born to smoking and non-smoking mothers.

9. Modify your code from the previous question to create two SAS data sets - one that contains the hypothesis test results and one that contains the confidence interval results. Your data sets should appear as those shown below.

Hypothesis test results

Obs	Variable	Method	Variances	tValue	DF	Probt
1	bwt	Pooled	Equal	8.65	1224	<.0001
2	bwt	Satterthwaite	Unequal	8.58	1003.2	<.0001

Confidence interval results

Obs	Variable	Class	Method	Variances	Mean	LowerCLMean	UpperCLMean	StdDev
1	bwt	0			123.0	121.8	124.3	17.3987
2	bwt	1			114.1	112.5	115.7	18.0989
3	bwt	Diff (1-2)	Pooled	Equal	8.9377	6.9112	10.9641	17.6783
4	bwt	Diff (1-2)	Satterthwaite	Unequal	8.9377	6.8939	10.9815	_

Obs	LowerCLStdDev	UpperCLStdDev	UMPULowerCLStdDev	UMPUUpperCLStdDev
1	16.5562	18.3321	16.5489	18.3237
2	17.0261	19.3172	17.0146	19.3036
3	17.0050	18.4076	17.0004	18.4025
4	_	_	_	_

10. Combine these two data sets such that you only keep results for the equal variances method. From the hypothesis test results, keep the degrees of freedom, test statistic, and p-value. From the confidence interval results, keep the lower and upper confidence interval limits. Print this data set - it should appear as shown below.

Obs	Variances	tValue	DF	Probt	LowerCLMean	UpperCLMean
1	Equal	8.65	1224	<.0001	6.9112	10.9641

11. In a comment in your SAS code, provide a brief interpretation of these results.