Lab 2

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_Lab2.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. Read the `babies.csv` data into SAS using `PROC IMPORT`. Include the `PROC IMPORT` code in your SAS program. Note that you will be referring to the data set as `work.babies` (not `sashelp.babies`).

2. Use a SAS procedure to identify the number of variables and observations in the data set. Write your findings in a comment in your SAS code.

3. Use a single SAS procedure to print the data set following these guidelines. A figure showing how your output should look is on the next page.

   - Only print the first 20 observations.

   - Only print the birth weight and gestation period variables.

- Do not print the observation number.

- Include the sum of the birth weights.

- Insert a blank line after every 5 observations.

4. Apply the `MEANS` procedure to the `babies` data set. Do all of the numeric summaries make sense? Explain why or why not.

5. Use a SAS procedure to obtain summary statistics of baby birth weight *separately* for smoking and non-smoking mothers as follows (output shown on next page):

   - Only obtain N, the sample mean, and the sample standard deviation.

   - Round the statistics to one decimal place.

   - Suppress the double printing of the number of observations.

   What is the relationship you observe between maternal smoking status and babies' birth weight? Note your observations in a comment in your SAS code.

6. Copy and paste the SAS code from your previous question. Now modify it to separate the summary statistics for birth weight by parity in addition to mother's smoking status (see output on next page). Identify the correct responses in the places underlined in this sentence:

   (First born / Non-first born) babies born to mothers who (did / did not) smoke

   had the *largest* average birth weight. Within each smoking status, the effect of

   parity is that first born babies tend to weigh about (X) ounces (more / less)

   than non-first born babies.

7. Review the SAS output that you used to answer question 6. If you could modify the display to make the output easier to interpret, what would you change?

Question 3 Output

bwt	gestation
120	284
113	282
128	279
123	.
108	282
136	286
138	244
132	245
120	289
143	299
140	351
144	282
141	279
110	281
114	273
115	285
92	255
115	261
144	261
119	288
2495	

Question 5 Output

The MEANS Procedure

Analysis Variable : bwt			
smoke	N	Mean	Std Dev
0	742	123.0	17.4
1	484	114.1	18.1

Question 6 Output

The MEANS Procedure

Analysis Variable : bwt				
smoke	parity	N	Mean	Std Dev
0	0	548	123.7	17.6
	1	194	121.1	16.8
1	0	363	114.4	18.6
	1	121	113.3	16.6