

Project #1: Data Summary

The data which is stored in the file **ncbirth1450.sav** is a random sample of 1450 birth records taken by the North Carolina State Center for Health and Environmental Statistics in 2001. Of particular interest will be incidents of low infant birth weight. Low birth weight is commonly defined as less than 2500 grams. Over the course of the semester we will investigate the relationship of several variables with low birth weight and each other. The goal of this assignment will be to summarize the variables in this data set both graphically and numerically. The variables in this study are:

| Variable Label | Description |
|------------------|--|
| plurality | Number of children born of the pregnancy |
| sex | Sex of child (1=Male, 2=Female) |
| mage | Age of mother (years) |
| weeks | Completed Weeks of Gestation (weeks) |
| marital | Marital status (1=married, 2=not married) |
| racemom | Race of Mother (0=Other Non-white, 1=White, 2=Black 3=American Indian, 4=Chinese, 5=Japanese, 6=Hawaiian, 7=Filipino, 8=Other Asian or Pacific Islander) |
| hispmom | Mother of Hispanic origin (C=Cuban, M=Mexican, N=Non-Hispanic, O=Other and Unknown Hispanic, P=Puerto Rican, S=Central/South American, U=Not Classifiable) |
| gained | Weight gained during pregnancy (pounds) |
| smoke | 0=mother did not smoke during pregnancy 1=mother did smoke during pregnancy |
| drink | 0=mother did not consume alcohol during pregnancy 1=mother did consume alcohol during pregnancy |
| tounces | Weight of child (ounces) |
| tgrams | Weight of child (grams) |
| low | 0=infant was not low birth weight 1=infant was low birth weight |
| Premie | 0=infant was not premature 1=infant was premature premature defined at 36 weeks or sooner |

The goal of this assignment is to obtain summary statistics for the variables in the data set. This is an important activity of most statistical studies. In your report, clearly label all tables and when appropriate give the units of measure. The components of the assignment are given below. Be sure your presentation is clear and organized. The use of tables is required.

First identify the categorical variables and the continuous variables. For the categorical variables construct a frequency table that gives the counts and percentages of observations in each category. For the continuous variables, create a histogram, comment on the shape of the histogram, and determine if the mean or median is the most appropriate measure of center. For each continuous variable, create a table that gives the mean, median, standard deviation, minimum and maximum value.

Suppose a friend of yours has given birth to a 10 pound 3 ounce baby. Is this unusual? Why or why not? In writing explain your reasoning. Another friend of yours had a baby and gained approximately 30 pounds during her pregnancy. Explain to her why she should not be too depressed over this occurrence. Comment on the percentage of women who reported smoking and drinking during pregnancy.

Lastly, propose three other variables you would like to investigate in regard to weight of the mother. Give the explicit question you would ask the mother prior to delivery and explain why you want to know that information.