2. Study Design

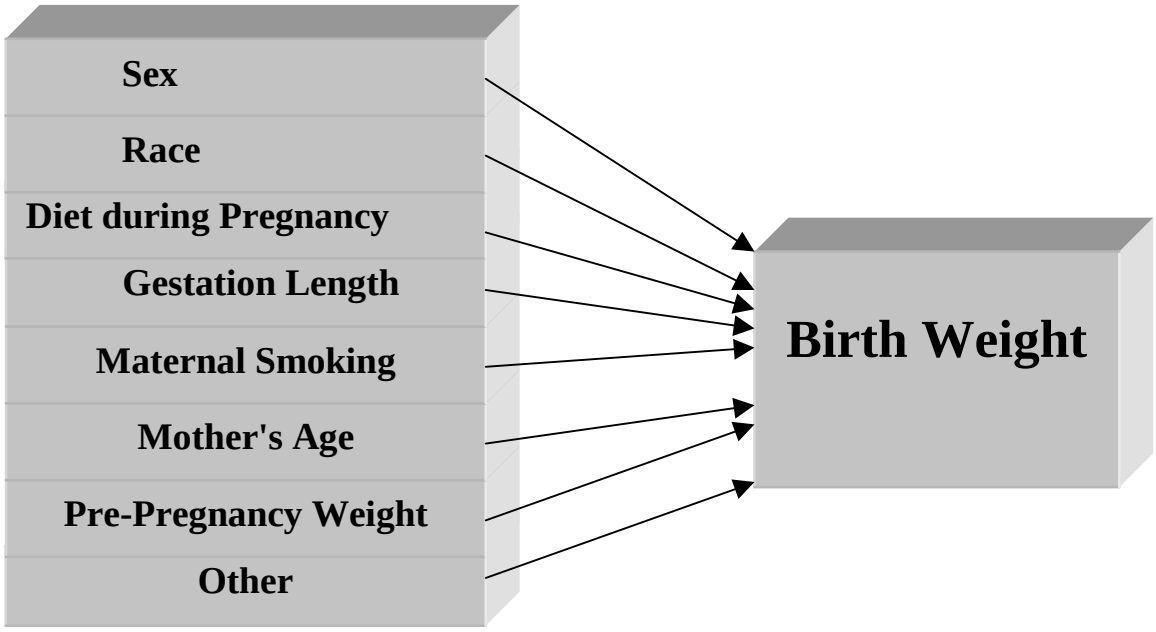
The inferences we may draw from the data depend crucially on the study's design and the way the data were collected. We will describe the study design by answering the following questions:

2.1 What factors affect birth weight?
2.2 What is the purpose of the study?
2.3 What is the scope of inferences in the study?

Before we will discuss the nature and scope of inferences we can make from the data, we have to identify important variables affecting the birth weight of a baby.

2.1 What factors affect birth weight?

Medical research shows that the birth weight is determined by several factors, among them sex, race, mother's diet during pregnancy, length of gestation period, mother's smoking during pregnancy, her age and pre-pregnancy weight. It is also believed that birth weight is affected by some father's characteristics such as weight, height, age, number of cigarettes smoked per day, and education level. In general, it would be difficult or even impossible to consider all variables affecting birth weight. Some variables may not be recognized or measured.



In our study a subset of data homogeneous with respect to sex (males) and race (white) will be employed. This removes sources of variation in birth weight caused by the two factors.

2.2 What is the purpose of the study?

The goal of the study is to describe the relationship between birth weight (BWT, response variable) and some quantitative variables that are believed to affect the response (predictor variables).

We used simple linear regression before (see *Child Development Study* in ***Simple Regression*** module) to develop two alternative models describing the relationship between birth weight and each of the two independent variables: length of gestation period and amount of maternal smoking.

In this module we will consider an extended set of nine predictor variables: length of gestation period (GESTWKS), amount of maternal smoking (MNOCIG), mother's age (MAGE), mother's height (MHEIGHT), mother's pre-pregnancy weight (MPPWT), father's age (FAGE), father's education (FEDYRS), amount of paternal smoking (FNOCIG), and father's height (FHEIGHT). We will apply a multiple linear regression model to examine the relationship between infant birth weight and the predictors in the following form:

$$BWT = \beta_0 + \beta_1 * GESTWKS + \beta_2 * MNOCIG + \beta_3 * MAGE + \beta_4 * MHEIGHT +$$
$$\beta_5 * MPPWT + \beta_6 * FNOCIG + \beta_7 * FAGE + \beta_8 * FHEIGHT + \beta_9 * FEDYRS + ERROR.$$

The random variable *ERROR* is assumed to follow a normal distribution with the mean of zero and an unknown standard deviation σ. The standard deviation is constant at all levels of the independent variables.

The above simple linear regression models assume a straight-line relationship between BWT and the predictors. This assumption is not very realistic for the relationship between birth weight and gestational age in the earlier weeks of pregnancy. We expect that the model will be able to provide better estimates of infant birth weight than the two simple linear regression models with gestational age amount of maternal smoking as predictors.

2.3 What is the scope of inferences in the study?

Live-born, white, male infants make up a study sample of 680 observations. The 680 infants were not selected randomly from any well-defined population. Therefore, the observed pattern cannot be inferred to hold in some general population, for example the population of white, male infants born in California, unless we assume that the infants are representative of the population.

What can then be inferred? As there is no random sample, the statistical results apply only to the infants actually measured. Any extrapolation of the pattern to other children comes from the assumption that the relationship between birth weight and gestation time (amount of maternal smoking) is similar for others. This is not necessarily a bad assumption. The point is that extending the inference to other infants is surely open to question.

The status of each infant (born to whom) is established beyond the control of the investigator. Thus the study is an observational study. That means that we cannot draw any causal conclusions from the statistical analysis alone. One cannot rule out the possibility that confounding variables are responsible for the differences in birth weights. For example, the result - that the infants born to smoking mothers tended to have lower birth weight relative to non-smoking mothers -does not prove that being born to a smoker is responsible for the difference. Diets during pregnancy, for example, may differ in the two groups and may be responsible for the different weights.

Nevertheless the study may be very useful to examine the factors affecting birth weight and to establish cause-and-effect relationship based on theory.