CHILD HEALTH AND DEVELOPMENT STUDY

1. Problem Formulation

The problem was already discussed in *Simple Regression* module in *STAT 252 Laboratories* Web site. We used simple linear regression to develop two alternative models describing the relationship between birth weight of male infants and each of the two independent variables: length of gestation period and amount of maternal smoking. We compared both models in terms of their utility and validity of linear regression assumptions.

We found that the interrelationship between gestation time and maternal smoking makes interpretation of the two separate simple linear regression analyses unsatisfactory because neither analysis accounts for the possible influence of the other variable. In this module we will overcome the obstacle by using a multiple linear regression model. We will demonstrate that multiple regression is able to take into account and assess simultaneous influences of the two independent variables on birth weight.

The data in the case study were collected for 680 live-born, white, male infants born to members of the Kaiser Foundation Health Plan who reside in the San Francisco-East Bay area. We will use the parental observations as independent variables, and assess their relationship to the characteristics of the infants (principally birth weight) with multiple linear regression. In particular, we will determine the influence of parental observations on infant birth weight and estimate the relative impact of maternal and paternal variables on birth weight.

The data from the study are available in the SPSS file *child.sav* located in the STAT 252 directory on the FTP server.

There are 12 variables in the data file describing each infant-mother-father set. They are organized in three blocks: *Infant Measurements*, *Maternal Measurements*, and *Paternal Measurements*. The following is a description of the variables in the data file:

	COLUMN	VARIABLE CODE	VARIABLE DESCRIPTION
INFANT MEASUREMENTS	1	ID	ID Number
	2	HEADCIR	Head circumference (inches)
	3	LENGTH	Length (inches)
	4	BWT	Birth Weight (pounds)
MATERNAL MEASUREMENTS	5	GESTWKS	Gestation (weeks)
	6	MAGE	Maternal Age (years)
	7	MNOCIG	Cigarettes (number smoked/day)
	8	MHEIGHT	Maternal Height (inches)
	9	MPPWT	Pre-pregnancy Weight (pounds)
PATERNAL MEASUREMENTS	10	FAGE	Father's Age
	11	FEDYRS	Father's education (years)
	12	FNOCIG	Cigarettes (number smoked/day)
	13	FHEIGHT	Father's Height (inches)

In this module we will consider a set of nine predictor variables: GESTWKS, MNOCIG, MAGE, MHEIGHT, MPPWT, FAGE, FEDYRS, FNOCIG, FHEIGHT. We will apply a multiple linear regression model to examine the relationship between infant birth weight and the predictors in the following form:

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

The results of the study are described by J. Yerushalmy in the report "The California Child Health and Development Studies-Study Design, and Some Illustrative Findings on Congenital Heart Disease" published in *Congenital Malformations, Proceedings of The Third International Conference, pp. 299-306, International Congress Series No.204, New York, 1970.*