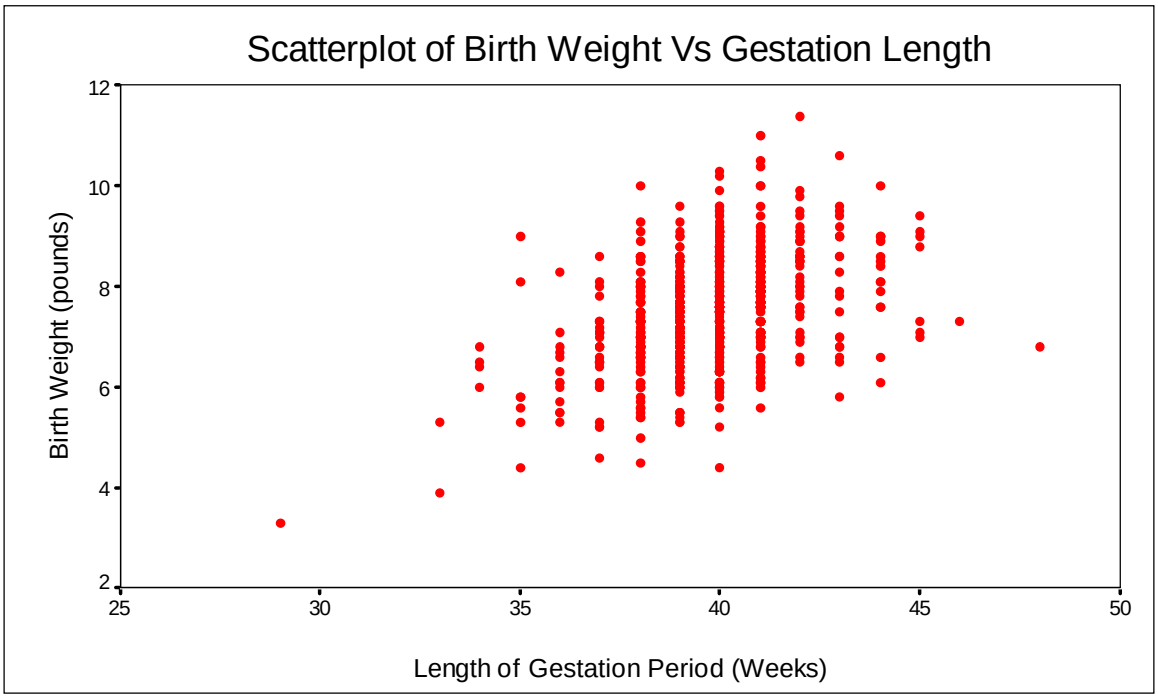
4. Displaying the Data

In this section we will visualize the data by obtaining the scatterplots of birth weight versus each of the nine independent variables. We will discuss the form of the relationship (linear, nonlinear), its strength, and the spread of the birth weight over the range of the predictor values.

4.1 The boxplot of birth weight over the range of values of gestational age (GESTWKS) obtained in **Section 4** in *Simple Regression* module allows us to describe the change in birth weight as the predictor variable changes. In particular, it allows us to examine the change in the spread and shape of the empirical distributions over the range.

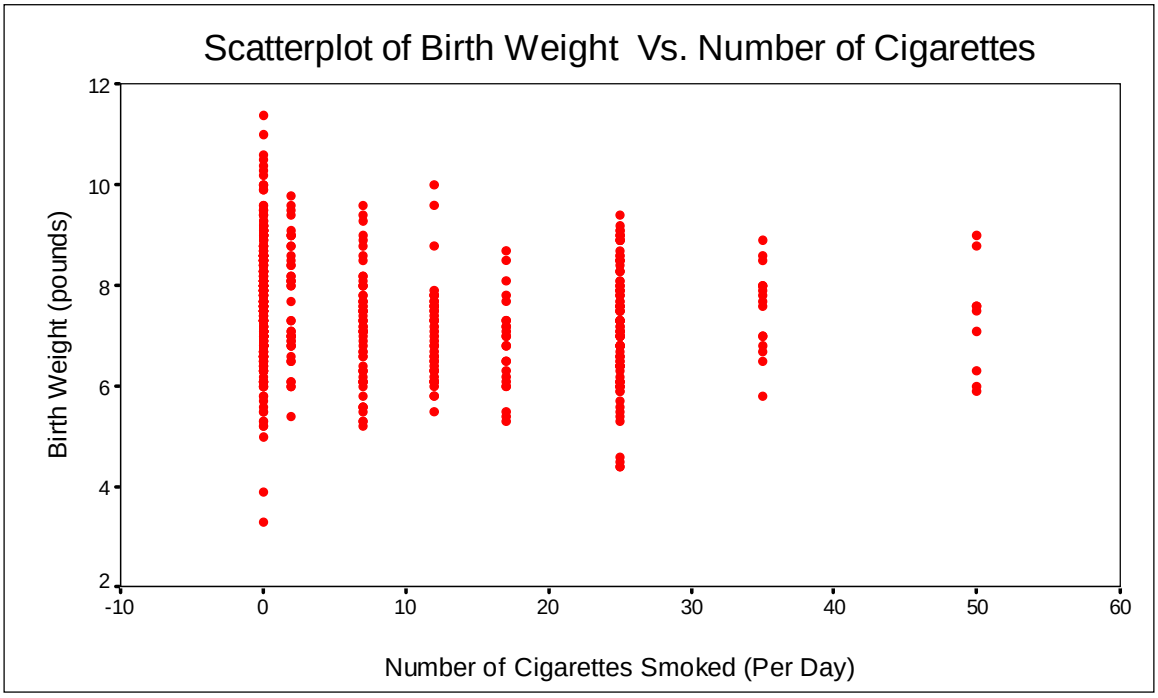
However, the presence or lack of a linear pattern in the data can be better evaluated by obtaining a scatterplot of birth weight versus the predictor variable.



The above scatterplot reveals a linear positive pattern in the data. The birth weight increases as the gestation time elapses. However, the spread of the scatter is large indicating the presence of other important predictor variables affecting birth

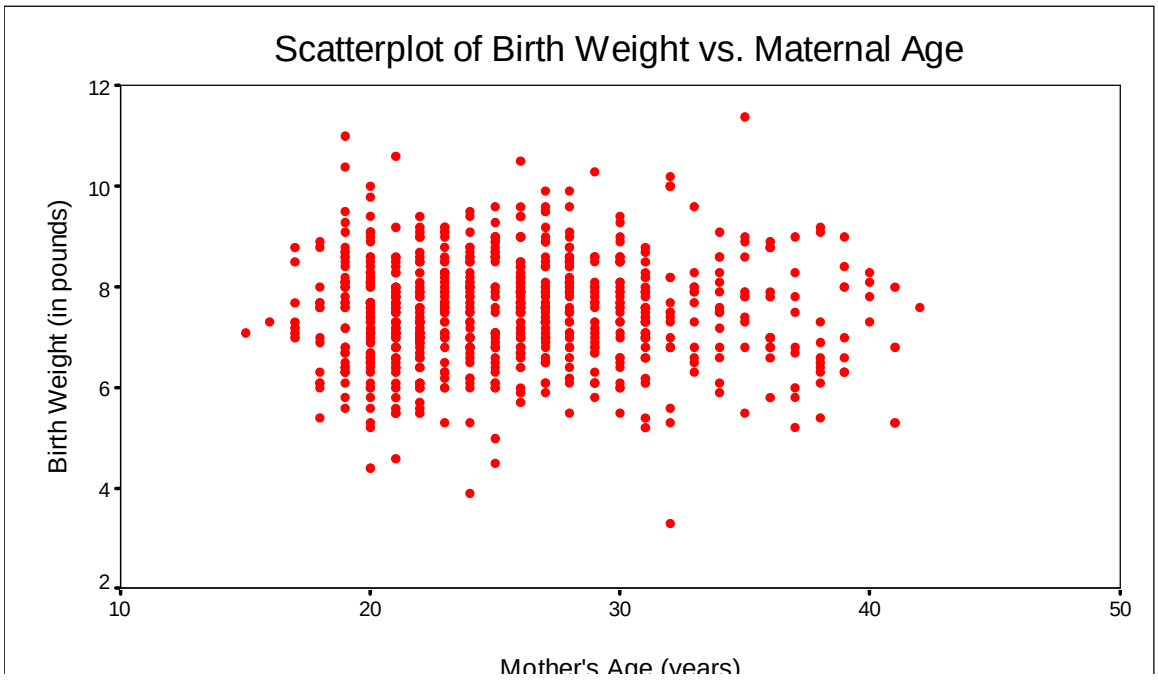
weight. It is obvious that a simple linear regression model is not able to provide very accurate predictions.

4.2 SPSS produces the following scatterplot of birth weight against amount of maternal smoking.



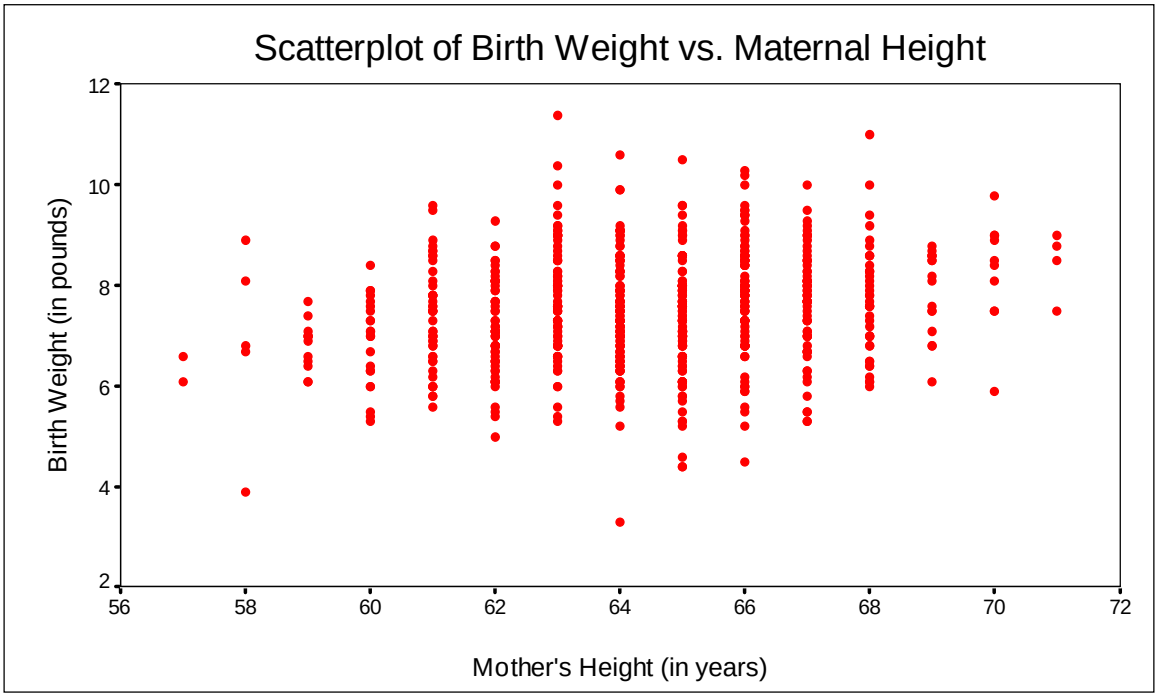
It is possible to question whether the above pattern can be described with a single straight line. Children born to smoking mothers tend to have lower birth weight than those born to non-smoking mothers, but it seems that there is no effect of the number of cigarettes smoked per day on infant birth weight. If we assume a linear relationship between birth weight and number of cigarettes smoked, the relationship is rather weak. The line describing the pattern is almost flat, although the negative slope is clear. The spread seems to decrease as the amount of smoking increases.

4.3 The scatterplot of infant birth weight versus maternal age is displayed below:



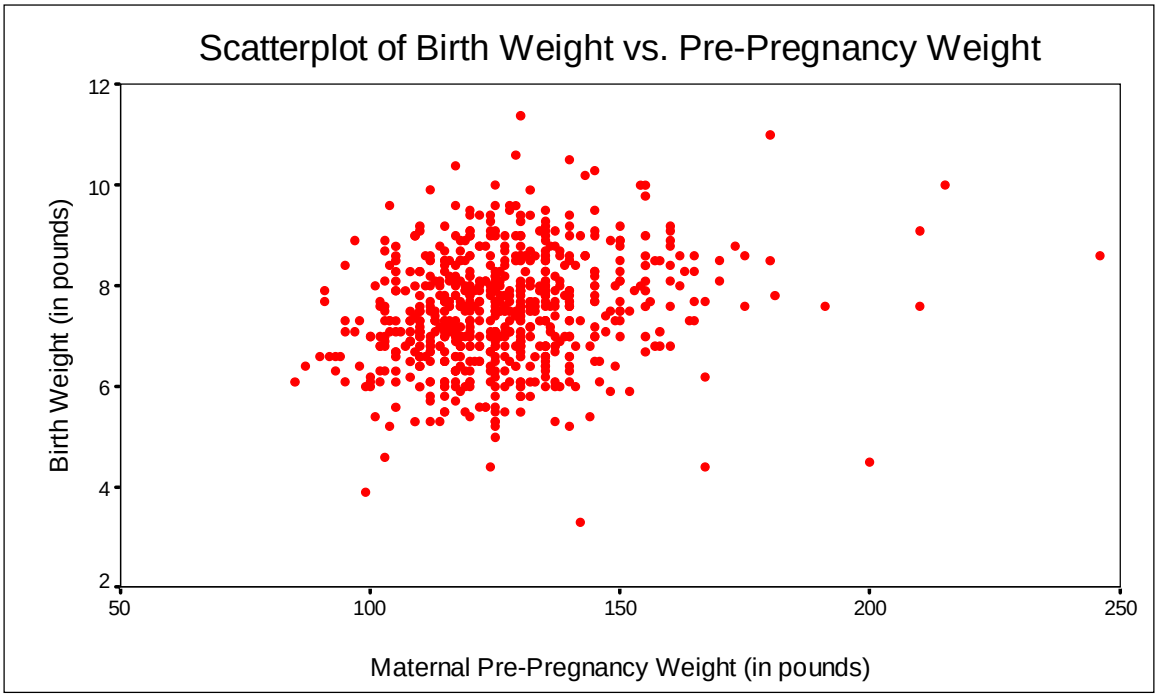
The above scatterplot reveals a very week linear pattern in the data. The line describing the pattern is almost flat. The spread seems to be approximately the same over the range of maternal age values.

4.4 The scatterplot of infant birth weight versus maternal height is displayed below:



The above scatterplot reveals a very weak positive linear pattern in the data. The spread seems to be smaller for the heights in the lower and upper ends of the range.

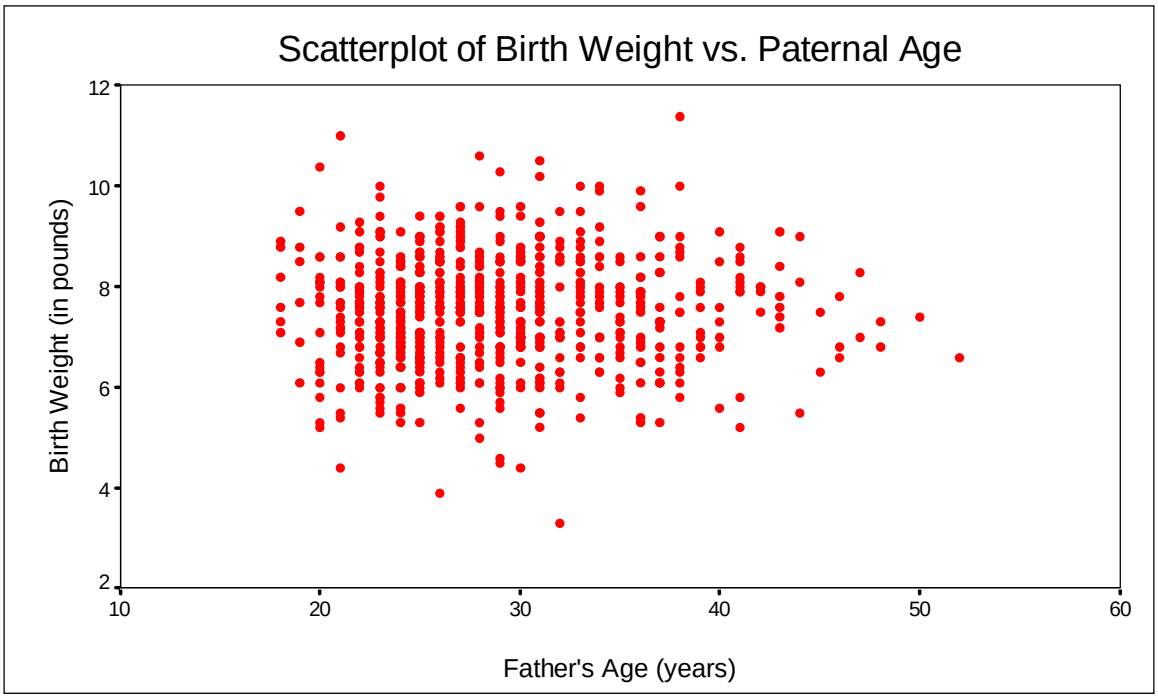
4.5 The scatterplot of infant birth weight versus maternal pre-pregnancy weight is displayed below:



The above scatterplot reveals a positive linear pattern in the data. The spread seems to be approximately the same over the range of maternal age values. The

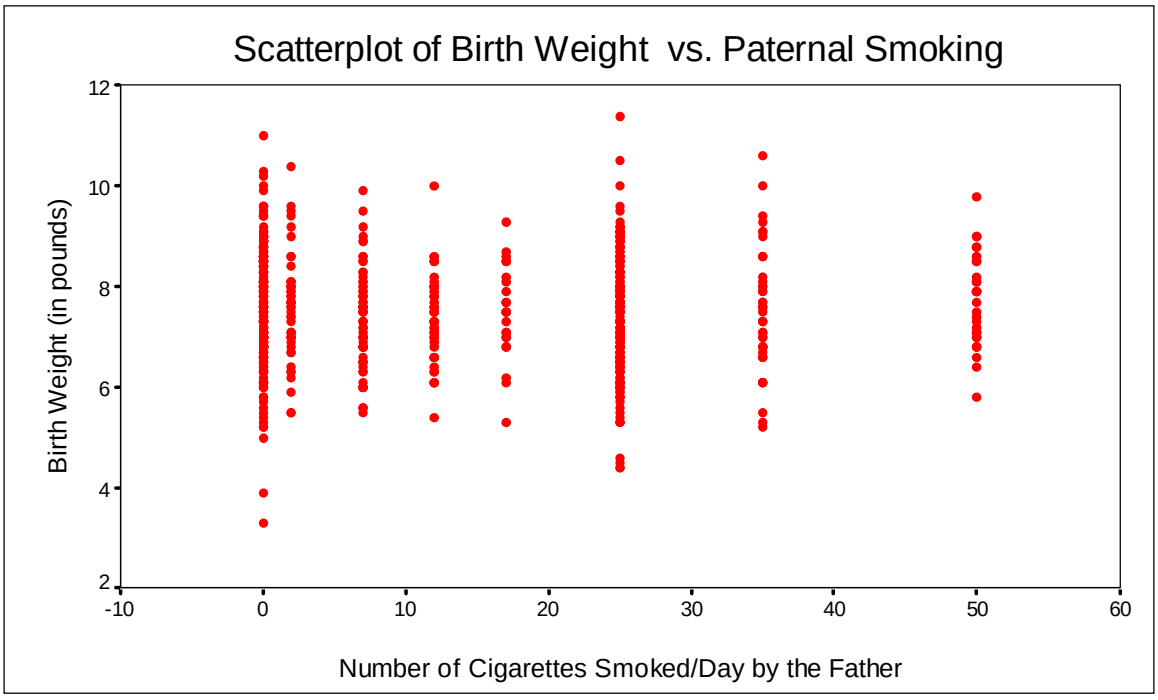
size of the spread indicates that maternal pre-pregnancy weight is not a strong predictor of infant birth weight.

4.6 The scatterplot of infant birth weight versus paternal age is displayed below:



The above scatterplot reveals a very negative weak linear pattern in the data. The line describing the pattern is almost flat. The spread seems to be approximately the same over the range of paternal age values. The size of the spread indicates that father's age is not a strong predictor of infant birth weight.

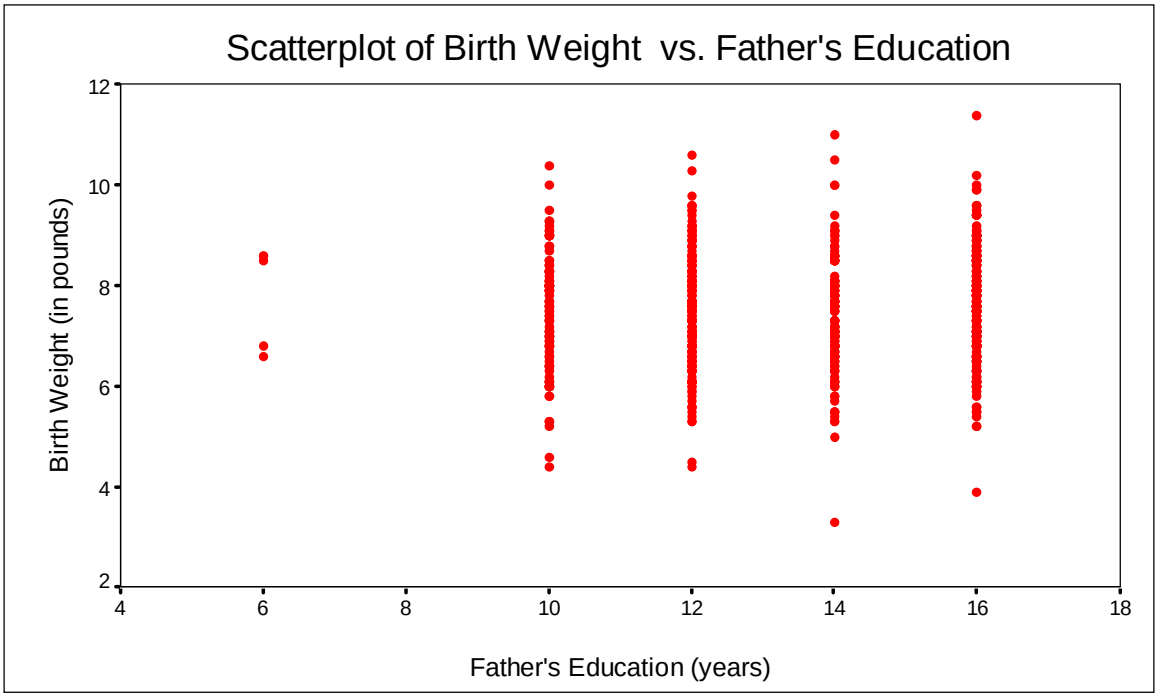
4.7 The scatterplot of infant birth weight versus amount of paternal smoking is displayed below:



It is possible to question whether the above pattern can be described with a single straight line. Children born to smoking fathers tend to have lower birth weight

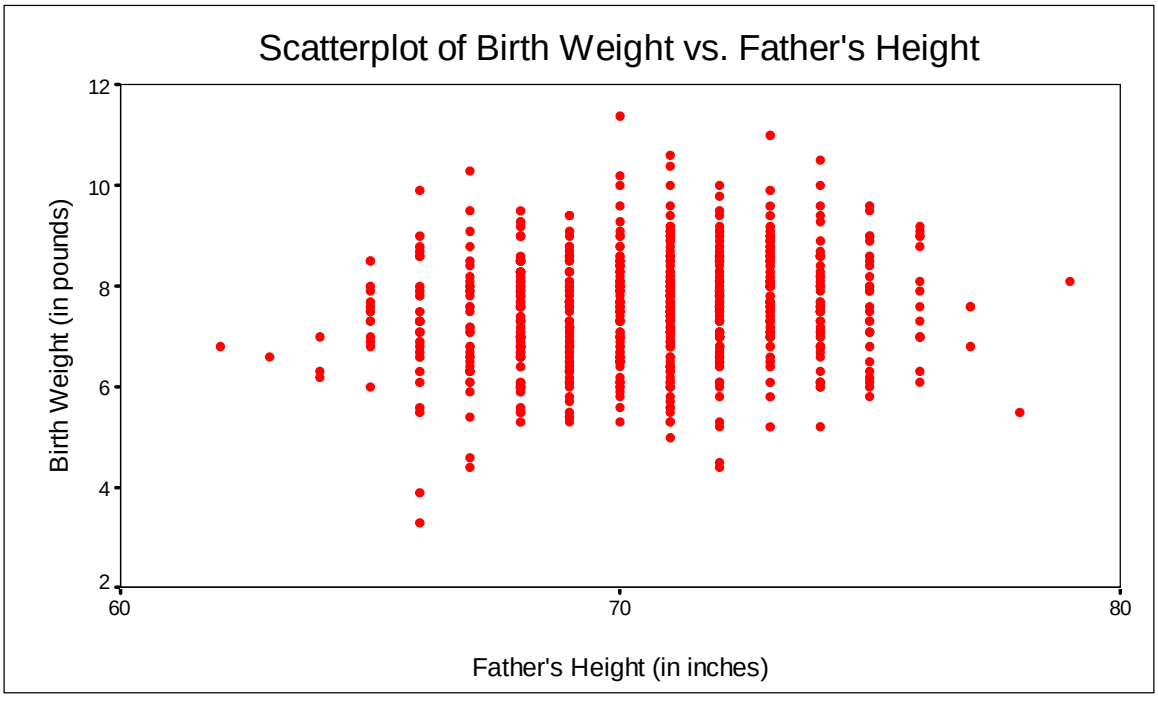
than those born to non-smoking fathers, but it seems that there is a very weak effect (if any) of the number of cigarettes smoked per day by father on infant birth weight. If we assume a linear relationship between birth weight and number of cigarettes smoked, the relationship is very weak. The spread seems to decrease as the amount of smoking increases.

4.8 SPSS produces the following scatterplot of infant birth weight versus father's education:



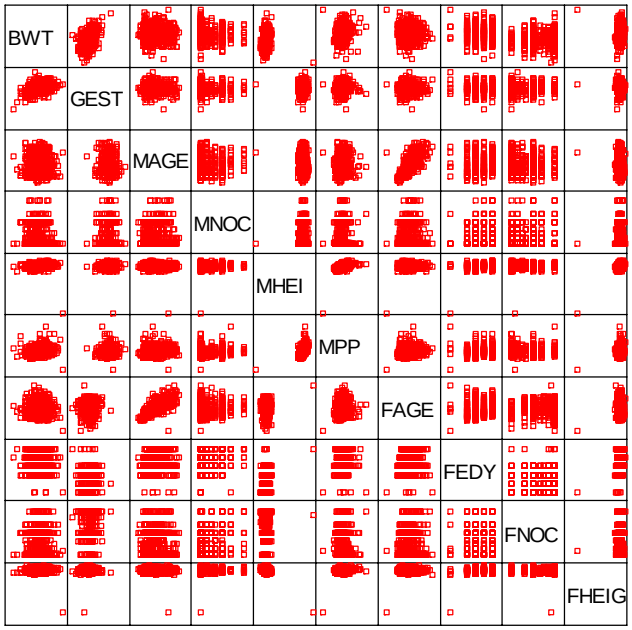
There is no linear pattern in the plot. It seems there is no relationship between infant birth and father's education. The spread in infant birth weight is approximately the same over the range of the predictor values.

4.9 The scatterplot of infant birth weight versus father's height is displayed below:



The plot is similar to the plot of birth weight against mother's height. The above scatterplot reveals a weak positive linear pattern in the data. The spread seems to be smaller for the heights in the lower and upper ends of the range.

4.10 The following scatterplot matrix allows you to visualize the relationships between any two variables in the data.



In the scatterplot matrix, for each plot, the names of variables are indicated in the corresponding row and the column of the matrix. For example, the middle plot in the first column of the matrix is the scatterplot of birth weight (BWT) versus gestation time (GESTWKS).

The first row of the matrix displays the relationship between the dependent variable (birth weight) and each of the nine independent variables. The degree of linearity of the relationship between birth weight and each of the nine independent predictors varies from very strong for GESTWKS (gestation time) to very weak for mother's age or father's age. The matrix also reveals high positive correlation between father's age and mother's age.