7. Tentative Multiple Linear Regression Model

The linear regression model discussed in Section 6 estimated the joint and separate effects of gestational age and amount of maternal smoking on infant birth weight. However, infant birth weight is also affected by some other variables.

In this section we will examine the relationship between infant birth weight (BWT) and the nine independent variables with the following multiple regression model:

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

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 response variable *BWT* under a range of settings of the nine independent variables *GESTWKS*, *MNOCIG*, *MAGE*, *MHEIGHT*, *MPPWT*, *FNOCIG*, *FAGE*, *FHEIGHT*, and *FEDYRS*.

The multiple linear regression model can be stated equivalently as follows:

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

The above model with the nine predictors is useful only if at least one slope β_i is different from zero. The hypothesis that the model is useful can be tested using F test.

The regression of birth weight (BWT) can now be done using the nine predictor variables. If the model explains a large portion of the variation in infant birth weight, it would be expected that at least some regression coefficients would not be significantly different from zero.

The following table displays the initial regression results for this data set.

MULTIPLE LINEAR REGRESSION

Multiple R	.51180
R Square	.26194
Adjusted R Square	.25203
Standard Error	.94472

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	9	212.22479	23.58053
Residual	670	597.97074	.89249

F = 26.42095 Signif F = .0000

The squared multiple correlation coefficient R^2 (0.26194) says that a significant portion (over 26 %) of the variation in infant birth weights is explained by these nine predictors. The adjusted squared multiple correlation coefficient (0.25203) is not very different from the unadjusted value of 0.26194 because the number of independent variables (k=9) is much less than the number of observations (n=680).

We analyze the ANOVA table associated with the multiple linear regression. The sum of squares due to the regression model is reported as 212.22479, and the sum of squares due to error (residual sum of squares) is 597.97074. The residual mean square is an estimate of the variance σ^2 and is equal to 0.89249.

The value of the F statistic is equal to 26.42095 with the corresponding p-value of 0 provides very strong evidence of the utility of the model.

Now we analyze the part of the output providing the estimates of the regression parameters.

---------- **Variables in the Equation** ----------

Variable	**B**	**SE B**	**Beta**	**VIF**	**T**	**Sig T**
GESTWKS	.233447	.019471	.400802	1.014	11.989	.0000
MNOCIG	-.015221	.003357	-.157066	1.089	-4.534	.0000
MAGE	-.001308	.011634	-.006544	3.073	-.112	.9105
MHEIGHT	.039756	.017511	.090378	1.439	2.270	.0235
MPPWT	.008421	.002374	.137819	1.370	3.548	.0004
FNOCIG	.001864	.002718	.024178	1.129	.686	.4931
FAGE	.000066	.010447	.000373	3.123	.006	.9949
FHEIGHT	.039018	.014742	.094239	1.151	2.647	.0083
FEDYRS	.004157	.017668	.008383	1.152	.235	.8140
(Constant)	-8.090991	1.43988			-5.619	.0000

According to the output, the estimated regression line of birth weights on the nine predictors is

$$\mu\{BWT\} = .233 * GESTWKS - .015 * MNOCIG - .001 * MAGE + .040 * MHEIGHT +$$
$$.008 * MPPWT + .002 * FNOCIG + .00007 * FAGE + .039 * FHEIGHT + .004 * FEDYRS - 8.091.$$

The comparison of the nine independent variables by means of individual t-statistics indicates the relative magnitude of the unique contribution of each variable to the overall variability in birth weight. According to the above SPSS output, gestation time is the largest contributor to the explained variation in birth weight. The regression coefficient associated with gestation time is .233447 with a corresponding t ratio of 11.989, indicating a very strong effect of gestation time on infant birth weight after accounting for the effect of maternal and paternal variables.

Maternal smoking, maternal pre-pregnancy weight, paternal height, and maternal height are the next four most important contributors. The remaining four variables (MAGE, FAGE, FNOCIG, and FEDYRS) have significance probabilities greater than .49, and therefore can be considered nonsignificant contributors.