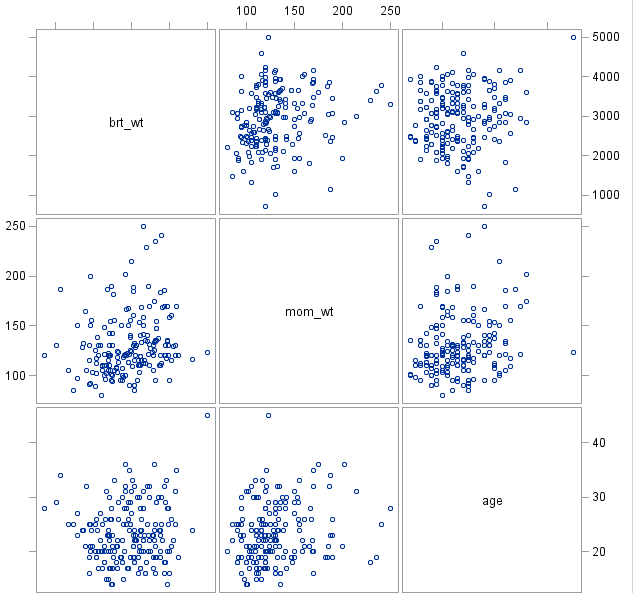
Remy Lagrois

Section 402

HW 13 and 14

Low Birthweight Prediction

1.

brt\_wt = Child’s birth weight

mom\_wt = Weight of mother prior to pregnancy

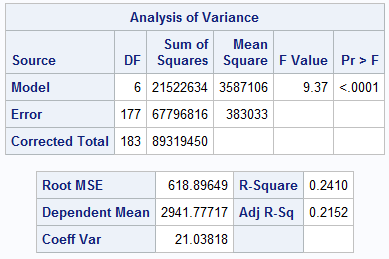
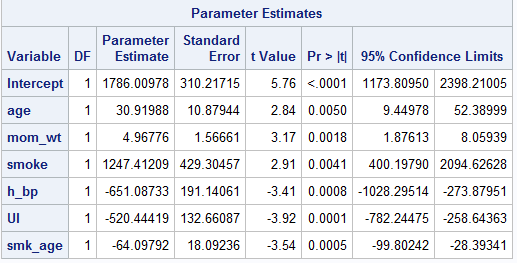
age = age of mother

The linear trends clearly are not very strong but it does seem that there’s a slight trend for mothers who weigh more to give birth the heavier children as well. Most of the variables in the model though are categorical and as will be seen later all are significantly different than zero.

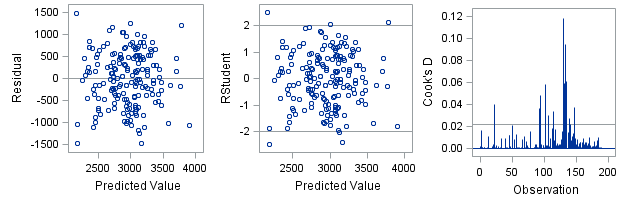
2. The estimated birthweight for a child is equal to:

β0 + β1(age) + β2(MW) + β3(smoker) + β4(HBP) + β5(UI) + β6(smoker\*age)

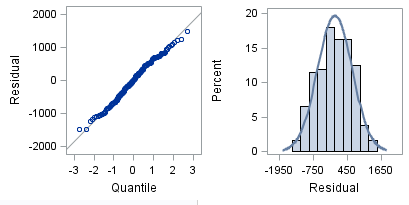
Age is the mother’s age in years, MW is the mother’s weight prior to pregnancy, smoker is whether or not the mother smokes, HBP is if the mother has a history of hypertension, and UI is if the mother has uterine irritability. Of these variables smoker, HBP, and UI are represented by either 1 or 0 where 1 is a yes and 0 is a no. The mothers weight is given in pounds while the baby’s birthweight is in grams.

3.

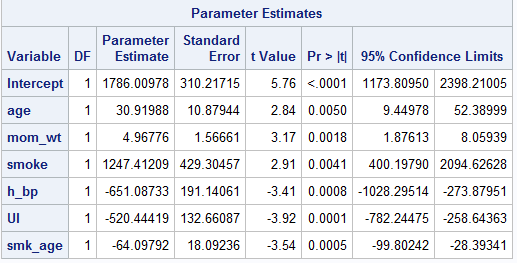
The p-value of the F test shows that our test does have significance overall. All of our parameters have a p-value below 0.05 and thus are significantly different from zero. The R2 is 0.241 and the adjusted R2 is 0.2152. Using the adjusted R2 may be more appropriate due to the number of factors used in the model. Note: These results are from after removing outliers.

4. 

The residual plots show that a linear model is appropriate as it is a random cloud with no real pattern and is evenly spread across the Y axis. There are some high leverage values still in the data, however it was found removing these in addition to the ones already removed resulted in a lower R2 so they were kept in the model.



The residual QQ plot and histogram show that the data are normally distributed which further confirms that a linear regression model is appropriate as none of the assumptions appear to be violated.

5.

Our intercept is 1786g which makes sense as a baby which is born can only be so light and will certainly have a non-zero/non-negative weight. Age and the mother’s weight have a positive slope, for each year older the mother is about 30.9 grams is added to the baby’s weight (95% CI between 9.44g and 52.39g) while each pound heavier the mother is about 4.97 grams heavier the baby is (95% CI between 1.88g and 8.06g). Somewhat surprisingly being a smoker adds a significant amount of weight to the baby, about 1247 grams (95% CI between 400g and 2095g). Having high blood pressure or uterine irritation tends to decrease the baby’s weight by 651g and 520g respectively (95% CI for high blood pressure between -1028 and -274 grams, for uterine irritation -782 and -258 grams). Interestingly while being a smoker and being older both increase a baby’s weight, the older the smoker is the less the effect. For every year older a smoker the baby will weigh about 64 grams less (95% CI between -99.8 and -28.4 grams). Given the adjusted R2 of 0.21 there clearly are more factors than provided by the model that influence the weight of a baby.

6. The MSE is 383033 and was estimated using 117 degrees of freedom.

7. a) For a 29 year old mother who is a smoker and weighs 123 lbs, the 95% confidence interval for her baby’s weight is 2455 to 2909 grams which means the mean weight of babies with that type of mother will fall within that interval 95% of the time

b) For the same 123 lb 29 year old smoker the 95% prediction interval is 1440 to 3925 grams meaning any given baby whose mother matches the above description will have a weight within that range 95% of the time