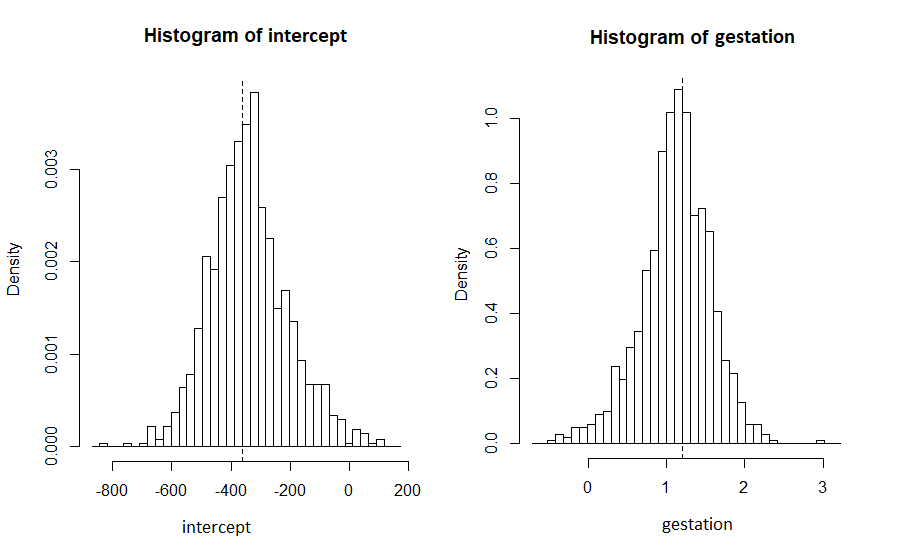
**Bootstrapping**

As mentioned previously, our final multiple linear model relies on assumptions of xxx. Any violations may lead to inaccurate standard error and therefore unreliable confidence intervals for the parameters. Hence, bootstrap was applied here to give empirical intervals.

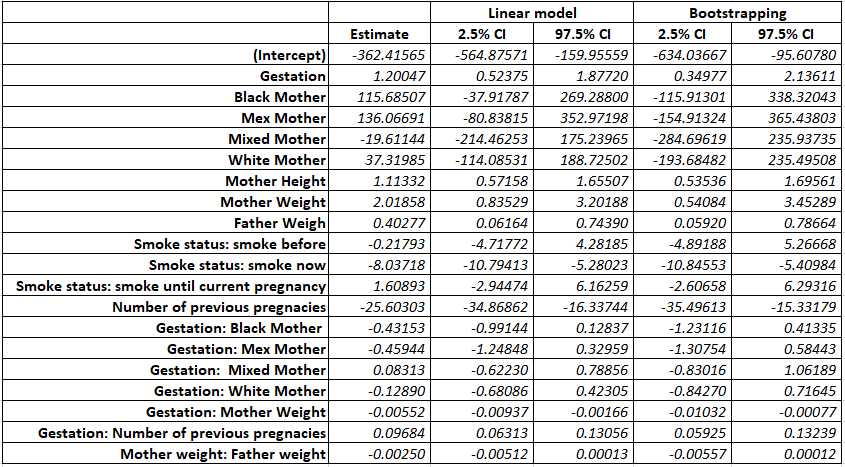
Boot package in R studio was employed to calculate standard error and confidence intervals for each variable, including factor variables. Formula x is our final model we settle on. It can expand as below:

Bootstrap resamples each parameter with replacement as original sample size in order to investigate the uncertainty in parameter estimate due to sampling uncertainty. Firstly, I replicated bootstrap on all parameters for 1000 times. Then histogram was plot for each item to generate approximate sampling distributions so as to get confidence intervals. As indicated in Figure k, estimates from the model are indicated by imaginary lines. 95% of resampling estimates were taken to get corresponding confidence intervals. This standard non-parametric bootstrap can assess standard errors (Diaconis, 1983) which does not rely on any assumptions as the linear model. As a result, it is likely to provide a more accurate confidence intervals of parameters than standard intervals (Efron, 1987).



*Figure k. Histogram of coefficient for intercept and gestation of the linear model.*

In this project, bias-corrected, accelerated method was employed to calculate confidence intervals which bootstrap percentile has been adjusted. Figure m demonstrates estimates for the parameters and 95% confidence intervals under standard linear model method and bootstrapping method. For most of continuous variables (like gestation, mother and father’s weight) confidence intervals from bootstrap are smaller than standard intervals. Nevertheless, most of factor variables (such as mother’s race) and interactions items (like gestation: mother’s race) tend to have larger bootstrapping intervals.

  
*Figure m. Confidence intervals for the parameters under standard method and bootstrapping.*

One of the causes may be the number of sample sizes. Figure n calculates the number of samples for the categories of two factor variables in the linear model. As suggested, Asia mother was the level of the baseline and other categories represent estimate difference between the baseline and their factor levels. Only 22 samples come from Asian mother compared to 414 samples from white mother. As a result, estimate of the baseline may be inaccurate due to small simple size and lead to unreliable standard error and confidence intervals. Furthermore, this might also result in inaccurate linear model.

|  |  |  |
| --- | --- | --- |
| **Factor Variable** | **Categories** | **Number of samples** |
| Mother’s Race | Asian (baseline) | 22 (0.037%) |
|  | Black | 130 (0.217%) |
|  | Mex | 18 (0.03%) |
|  | Mixed | 14 (0.023%) |
|  | White | 414 (0.692%) |
| Smoking status | Never (baseline) | 279 (0.467%) |
|  | Smoke before | 53 (0.089%) |
|  | Smoke now | 215 (0.36%) |
|  | Smoke until current pregnancy | 51 (0.085%) |

*Figure n. number of samples for two factor variables.*

**Practical and statistical Significance**

The existence of statistical significance can be determined by p value. If p value is small enough, normally below 0.05, it is fair to conclude that the observed data are significantly enough. However statistically significant may have little practical importance (Drasgow, 2004). Ellis also suggested (2003) that practical significance can be considered as the magnitude of an effect in practice.

This project investigates relationships between the birth weight of babies and various measured variables. Some variables (such as gestation, mother and father’s weight) are statistically significant therefore they were included in the regression model. On the other hand, they do not have practical significance due to uncontrolled cause. For instance, mother cannot decide days of gestation (assume there is no oxytocic available). Therefore even through we identify there is a statistical significance between baby weight and days of gestation, there is not much we can do to increase baby weight via this factor. Same situation also applies to mother’s race, mother’s weight and height and father’s weight.

There is one factor in our multiple linear model that has practical effect: smoking status. As mentioned before, babies’ weights are statistically significant (p-value 1.28 \* 10-7) different in various smoking conditions of mothers. It is also reasonable to suggest that it can have an impact in practice as mothers can control the number of cigarettes. According to the original data, average weight of babies from non-smoking mothers (including never smoked, smoked before and smoked until pregnancy mothers) are 8 ounces heavier than smoking mothers. This effect was also proved by Yerushalmy (1964).

**Reference**

Diaconis, P. and Efron, B. (1983) *Computer intensive methods in statistics*. Scientific American, Vol. 248, pp. 115-130

Drasgow, F. (2004) *Examining the Effects of Differential Item (Functioning and Differential) Test Functioning on Selection Decisions: When Are Statistically Significant Effects Practically Important?* Journal of Applied Psychology Vol. 89, No. 3, pp. 497-508

Efron, B. (1987) *Better Bootstrap Confidence Intervals*. Journal of the American Statistical Association Vol. 85, No. 397, pp. 171-185

Ellis, S.M. & Steyn, H.S. (2003) *Practical significance (effect sizes) versus or in combination with statistical significance (p-values)* Management Dynamics Vol. 12, No. 4, pp. 51-53

Yerushalmy, J. (1964) *Mother's cigarette smoking and survival of infant.* American Journal of Obstetrics & Gynecology. Vol. 88, Issue 4, pp. 505-518