Assignment

Child weight at birth as a function of mother age, weight, smoking status and race during pregnancy (**dataset**: LowBirthWeight.xlsx).

In this study, we have the explanatory variable age of the mother (variable AGE in years), the weight of the mother (variable LWT in Kg), the smoking status of the mother (variable SMOKE: No/Yes), the race of the mother (variable RACE: White/Black/Other) and the child weight at birth (response variable BWT in grams).

a) Run a simple linear regression model with response variable BWT and explanatory variable LWT (model1). Interpret the results.

b) Run a multiple linear regression with response variable BWT and explanatory variables: AGE, LWT and SMOKE. Find the regression equation for this model (model 2).

c) For the model2 interpret the meaning of the coefficients (b1, b2) analytically and mention whether they are significant)?

d) Calculate 95% Confidence Intervals for the coefficients of the model2.

e) Find and interpret the adjusted R-squared of the model2.

f) How can you interpret the p-value of F-statistic of the model2 based on Ho and H1 for this test?

g) Test model2 for multicollinearity.

h) Remove the variable AGE from the model2. Compare this nested model (model3) with the full model (model2) (NOTE: use the anova function). What can you comment about the p-value of ANOVA test and the p-value of AGE produced by model2?

i) Run a multiple linear regression with response variable BWT and explanatory variables: LWT and RACE. Find the regression equation for

this model (model 4) setting as reference group White. Interpret the results.

j) Compare model 4 with model1 (NOTE: use anova or AIC function) and explain the results.

Important

Don't forget to convert the SMOKE and RACE into factor variables!

Keep on the good work!!