Project #4: Simple Linear Regression

For this assignment use the data set **ncbirth200.sav**. Recall the variables from the North Carolina birth data set are:

The variables examined are:

Variable Label	Description
sex	Sex of child (1=Male, 2=Female)
race	Race of child (0=other Nonwhite, 1=White, 2=Black, 3=American Indian, 4=Chinese, 5=Japanese, 6=Hawaiian, 7=Filipino, 8=Other Asian or Pacific Islander)
age	Age of mother
educ	Education level of mother
gest	Completed Weeks of Gestation
bwtgroup	Birthweight (grams) group (0=500 or less, 1=500-1000,2=1001-1500, 3=1501-2000, 4=2001-2500, 5=2501-3000, 6=3001-3500, 7=3501-4000, 8=4001-4500, 9=4501 and over)
marital	Marital status (1=married, 2=not married)
plural	Number of children born of the pregnancy
totounc	Weight of child in total ounces
low	Birthweight under 2500 grams (1=low, 0=not low)
smoke	Smoking status (1=mother smoked, 0=mother did not smoke)
drankalc	Drinking status (1=mother drank, 0=mother did not drink)

Answer the following for the variables **totounc** (response variable) and **age** (explanatory variable).

a. Make a scatterplot of this data. Fit the regression line. Report the parameter estimates (the estimates of the intercept and slope).

b. Is **age** useful in predicating **totounc**? Why? Report the level of significance (P-value).

c. What percentage of the variation in **totounc** is explained by **age**? Is that high or low?

d. What is the predicted value for **totounc** when **age** is 35? What if **age** is 17?

e. Make a residual plot. Comment on the fit of the model.

f. Determine if there are any outliers.

Answer the following for the variables **totounc** (response variable) and **educ** (explanatory variable).

a. Make a scatterplot of this data. Fit the regression line. Report the parameter estimates.
b. Is **educ** useful in predicating **totounc**? Why? Report the P-value.
c. What percentage of the variation in **totounc** is explained by **educ**? Is that high or low?
d. What is the predicted value for **totounc** when **educ** is 10? What if **educ** is 15?
e. Make a residual plot. Comment on the fit of the model.
f. Determine if there are any outliers.

Answer the following for the variables **totounc** (response variable) and **gest** (explanatory variable).

a. Make a scatterplot of this data. Fit the regression line. Report the parameter estimates.
b. Is **gest** useful in predicating **totounc**? Why? Report the P-value.
c. What percentage of the variation in **totounc** is explained by **gest**? Is that high or low?
d. What is the predicted value for **totounc** when **gest** is 35? What if **gest** is 40?
e. Make a residual plot. Comment on the fit of the model.
f. Determine if there are any outliers.

Write a paragraph summarizing your conclusions of the three analyses done above. Be sure to interpret the meaning of each. Which results are most useful, which are most surprising?

Write a summary of all your conclusions from the analysis of the birth weight data set. Reread your previous reports and comment on the findings that are most interesting and most surprising. Comment on what conclusions you are confident making, those you are not. Point out any contradictory results. Explicitly state at least one question you would ask mothers as they are giving birth. Then explain a test of hypotheses you would do or a relative risk calculation you would conduct with the new information you would collect. What would that test tell you? This summary should be at least two paragraphs.