

## Project #4: Simple Linear Regression

For this assignment use the data set **ncbirth150.MTW**. Recall the variables from the North Carolina birth data set are:

The variables examined are:

Variable Label	Description
<b>PLURALITY</b>	Number of children born of the pregnancy
<b>SEX</b>	Sex of child (1=Male, 2=Female)
<b>MAGE</b>	Age of mother (years)
<b>WEEKS</b>	Completed Weeks of Gestation (weeks)
<b>MARITAL</b>	Marital status (1=married, 2=not married)
<b>RACEMOM</b>	Race of Mother (0=Other Non-white, 1=White, 2=Black 3=American Indian, 4=Chinese, 5=Japanese, 6=Hawaiian, 7=Filipino, 8=Other Asian or Pacific Islander)
<b>HISPMOM</b>	Mother of Hispanic origin (C=Cuban, M=Mexican, N=Non-Hispanic, O=Other and Unknown Hispanic, P=Puerto Rican, S=Central/South American, U=Not Classifiable)
<b>GAINED</b>	Weight gained during pregnancy (pounds)
<b>SMOKE</b>	0=mother did not smoke during pregnancy 1=mother did smoke during pregnancy
<b>DRINK</b>	0=mother did not consume alcohol during pregnancy 1=mother did consume alcohol during pregnancy
<b>POUNDS</b>	Weight of child (pounds)
<b>TGRAMS</b>	Weight of child (grams)
<b>LOW</b>	0=infant was not low birth weight 1=infant was low birth weight
<b>PREMIE</b>	0=infant was not premature 1=infant was premature premature defined at 36 weeks or sooner

Answer the following for the variables **pounds** (dependent variable) and **mage** (independent 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 **mage** useful in predicating **tounces**? Why? Report the level of significance (P-value).
- c. What percentage of the variation in **tounces** is explained by **mage**? Is that high or low?
- d. What is the predicted value for **tounces** when **mage** is 35? What if **mage** is 17?
- e. Make a residual plot. Comment on the fit of the model.
- f. Explain the implication of the  $R^2$  value.

Answer the following for the variables **pounds** (dependent variable) and **weeks** (independent variable).

- a. Make a scatterplot of this data. Fit the regression line. Report the parameter estimates.
- b. Is **weeks** useful in predicating **tounces**? Why? Report the P-value.
- c. What percentage of the variation in **tounces** is explained by **weeks**? Is that high or low?
- d. What is the predicted value for **tounces** when **weeks** is 35? What if **weeks** is 40?
- e. Make a residual plot. Comment on the fit of the model.
- f. Compare the  $R^2$  value to the one constructed with **mage** predicting **pounds**. What does this imply?

Fit a regression line using **tgrams** (dependent variable) and **weeks** (independent variable). Explain why the  $R^2$  is the same as when **weeks** predicted **pounds**. Explain to a non-statistician how much, on average, each week adds to the weight of the baby from this regression model.