Table Number:_____ Group Name: _____

Group Members:_____ _____ _____

Activity: The Regression Line

The data set, which is stored in the file North Carolina birth data, is a random sample of 1000 birth records taken by the North Carolina State Center for Health and Environmental Statistics in 2005. Of particular interest will be incidents of low infant birth weight. Low birth weight is defined as less than 2500 grams. Then goal of this activity will be to summarize the variables in this data set both graphically and numerically. The variables in the study are:

Variable Label	Description of variable
Plurality	Refers to the number of children associated with the birth
Sex	Sex of child (Gender 1=Male 2=Female)
Fage	Age of father (years)
Mage	Age of mother (years)
Weeks	Completed weeks of gestation
Visits	Number of pre-natal medical visits
Marital	Marital status (1=married 2=unmarried)
Racemom	Race of Mother (0=Other Non-white 1=White 2=Black 3=American Indian 4=C hinese 5=Japanese 6=Hawaiian 7=Filipino 8=Other Asian or Pacific Islander)
Hispmom	Whether mother is 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)
Gained	Weight gain during pregnancy (pounds)
Lowbw	If birth weight is 2500 grams or lower, 0=infant was not low birth weight, 1=infant was low birth weight
Tpounds	Birth weight in pounds
Smoke	0=no 1=yes for mother admitted to smoking
Mature	0=no for mother is 34 or younger 1-yes for mother is 35 or older
Premie	0=no 1=yes to being born 36 weeks or sooner.

1. Answer the following for the variables Tpounds (response variable, *y*) regressed on Weeks (Explanatory Variable, *x*).
   A. Make a scatterplot of the data with the regression line. Report the parameter estimates (estimates of the slope and intercept). Copy and paste your information here:

B. Interpret the slope and the intercept

C. Use the coefficient of determination to determine the percentage of the variation in Tpounds that is explained by Weeks?

D. What is the predicted value for Tpounds when Weeks is 35? What if Weeks is 40?

E. Use the correlation coefficient to comment on the fit of the model.

2. Answer the following for the Fage (response variable) regressed on Mage (Explanatory Variable).
   A. Make a scatterplot of the data with the regression line. Report the parameter estimates (estimates of the slope and intercept). Copy and paste your information from StatCrunch here:

B. Interpret the slope and the intercept in context of the “real world” data.

C. Use the coefficient of determination to determine the percentage of the variation in Fage that is explained by Mage?

D. What is the predicted value for Fage when Mage is 35? What if Mage is 17?

E. Use the correlation coefficient to comment on the fit of the model.

Answer the following using the information calculated

3. A friend of yours is pregnant with her first child. She is nervous that she might have her baby at 34 weeks. What can you tell her about the predicted weight of the baby?

4. Your cousin is 40 years old. Predict the age of the father of the child. Using the scatterplot, give a range for the suspected ages of the father.

5. Of the two regression models, which has the best fit and why?

If time permits try one more set:

Answer the following for the Tpounds (response variable, *y*) regressed on Mage (Explanatory Variable, *x*).

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

B. Interpret the slope and the intercept

C. Use the coefficient of determination to determine the percentage of the variation in Tpounds that is explained by Mage?

D. What is the predicted value for Tpounds when Mage is 35? What if Mage is 17?