Parents of children who speak at a young age like to believe that this bodes well for the child exhibiting high intelligence later in life. To investigate this possibility, researchers collected data on the age of first speaking (in months) and score on the Gesell aptitude test taken later in life for a sample of 21 children. The data can be found in the file gesell.mtw, available from our course webpage.

a) Which is the explanatory variable, and which is the response?

b) Examine (and submit) a scatterplot of Gesell score vs. age of first speaking, with the response variable on the vertical axis. Comment on whether these two variables appear to be associated.

c) Determine (and report) the correlation coefficient between Gesell score and age of first speaking.

d) Determine (and report) the equation of the least squares line for predicting Gesell score from age of first speaking.

e) Identify and interpret the value of the slope coefficient.

f) What proportion of the variability in Gesell scores is explained by the least squares line with age of first speaking?

g) Do any of the children appear to be outliers in the age variable? If so, how long did it take him/her to speak, and what is his/her Gesell score?

h) Remove this outlier (child) from the analysis. Then reproduce a scatterplot and recalculate the correlation coefficient. Comment on how these have changed.

i) Now also remove the child who took the next longest time to speak, again look at a scatterplot, and recalculate the correlation coefficient. Comment again on how these have changed.

j) Write a paragraph explaining (as if to someone with no formal knowledge of statistics) why the scatterplot and correlation change so much. Also summarize what your analysis of these data reveals concerning the relationship between age of first speaking and aptitude.