Water striders mating

Water striders are common bugs that skate across the surface of water. Water striders have different personalities and some of the males are hyper-aggressive, meaning they jump on and wrestle with any other water strider near them. Individually, because hyper-aggressive males are much more active, they tend to have better mating success than more inactive striders. This study\(^1\) examined the effect they have on a group. Four males and three females were put in each of ten pools of water. Half of the groups had a hyper-aggressive male as one of the males and half did not. The proportion of time females are in hiding was measured for each of the 10 groups, and a measure of mean mating activity was also measured with higher numbers meaning more mating. The data, WaterStriders, can be downloaded from [http://lock5stat.com/datapage.html](http://lock5stat.com/datapage.html).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AggressiveMale</td>
<td>Hyper-aggressive male in group? No or Yes</td>
</tr>
<tr>
<td>FemalesHiding</td>
<td>Proportion of time the female water striders were in hiding</td>
</tr>
<tr>
<td>MatingActivity</td>
<td>Measure of mean mating activity</td>
</tr>
</tbody>
</table>

\(^1\) Sih, A. and Watters, J., "The mix matters: behavioural types and group dynamics in..."
1) Do Stat -> Regression -> Fitted Line Plot to create a model predicting MatingActivity from FemalesHiding.

   a) Describe the scatterplot. Does there appear to be a linear association? If so, does it appear to be positive or negative? Are there any other relevant features? Does simple linear regression appear to be appropriate for this data?

   b) What is the regression equation?

   c) Interpret the slope in context.

   d) Interpret the intercept in context. Does the intercept make sense here?

   e) What is the actual response value for case #4? Calculate the predicted value for case #4, and use these to calculate the residual.

   f) What is $R^2$? Interpret this value in context.

   g) What is $r$, the sample correlation between these two variables?

2) Do Stat -> Regression -> Regression -> Fit Regression Model to perform inference.

   a) Look at the scatterplot. Does the constant variability condition for inference appear to be satisfied?

   b) State the hypotheses for testing for a nonzero slope.

   c) Use the information under “Coefficients” to make a conclusion to test.

   d) Use the estimated slope and standard error to verify the $T$ statistic given under “Coefficients”.

   e) If you were to do a test for correlation, what $p$-value would you get? (Feel free to do this by hand if you want extra practice, but you don’t have to.)

   f) Rerun this regression output, but under Graphs (just before clicking OK to get the output), check the box next to histogram of residuals to get a histogram of the residuals. Does the normality of the residuals condition look okay, or is it blatantly not satisfied?
3) The data was collected to investigate the presence or lack thereof of an aggressive male.

a) Create a scatterplot, but choose “with groups” and add AggressiveMale as a grouping variable. Comment on what you see.

b) Test whether FemalesHiding differs by whether or not there is an AggressiveMale. You can do this by hand if you want practice for the exam (get the needed summary statistics from Stat -> Basic Statistics -> Display Descriptive Statistics), or can just use Minitab if you want to check your work or do it faster. (Check the Minitab guide online if you don’t remember how to do this).

c) If you want to get really fancy, you can actually add AggressiveMale as an additional predictor to your regression model by adding it as a categorical variable. This will create a model with two explanatory variables! How does the coefficient for FemalesHiding change? Notice that now under regression equation you now get two different equations, depending on whether there is an aggressive male or not. How do they differ?

Regression is actually a very flexible modeling framework that can incorporate any number of explanatory variables, both quantitative and categorical. Call myself/TA/LA over if you get here and want to talk more about it.