The questions are practice for the quiz next week, and are not to be handed in. I would like you to bring in all of the code you used to run this assignment, and the output it generates. The best way to do this is just to run all of your code, and copy the console output to a .txt file and print it out. Clearly, your code should not print things that are not asked for, like an entire data frame. Even though you might want to look at the data frame while writing your code, your final version should only print the answers to the questions. Your console output (with code) should fit on two pages, double-sided, with two columns per side, at most. If your output is longer than 8 single pages (about 400 lines) you could probably tidy things up a bit. Remember, a whole lot of nonsense is not the same thing as concise, tidy code, even if the conclusions are the same. Also, please don’t write anything on your output in pen or pencil, once printed.

If you want to experiment with R Markdown files to make it look nicer, be my guest! I might show these a little later in the course.

Remember, the computer assignments in this course are not group projects. You are expected to do the work yourself. You may compare numerical answers but do not share R code!

---

*This assignment was prepared by Craig Burkett with contributions from Jerry Brunner, both of the Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely.*
1. For the usual fixed effects multiple regression model $Y = X\beta + \varepsilon$,

   (a) Show that $H$ is symmetric and idempotent (i.e. $H' = H$ and $H^2 = H$).
   
   (b) Show that $I - H$ also has these properties.
   
   (c) Show that $H(I - H) = 0$.
   
   (d) Show that $HX = X$. Since this is true, it’s also true for any column of $X$, including
   
   the first column. You can use this result for the next parts.
   
   (e) Show that $\Sigma e_i = 0$ using matrices.
   
   (f) Show that $\Sigma e_ix_{p,i} = 0$ for any vector $x_p$ from the design matrix, using matrices.
   
   (g) Show that $\Sigma Y_i = \Sigma \hat{Y}_i$ using matrices.
   
   (h) Show that the least squares hyper-plane passes through the centroid of the data.  

   Hint: Show that $\bar{Y} = b_0 + b_1\bar{X}_1 + \ldots + b_{p-1}\bar{X}_{p-1}$

2. For the usual fixed effects multiple regression model $Y = X\beta + \varepsilon$,

   (a) Show that the total sum of squares can be written as $Y'(I - \frac{1}{n}J)Y$ where $J$ is a
   
   matrix of 1s.
   
   (b) Show that the error sum of squares can be written as $Y'(I - H)Y$.
   
   (c) Show that the model sum of squares can be written as $Y'(H - \frac{1}{n}J)Y$.

3. For the usual fixed effects multiple regression model, let $W = (X'X)^{-1}X'e$.

   (a) Simplify this expression for $W$.
   
   (b) What is the probability distribution of $W$?
   
   (c) Now you know whether $V(e)$ has an inverse. Why?

4. In a study of math education in elementary school, equal numbers of boys and girls were

   randomly assigned to one of three training programmes designed to improve spatial

   reasoning. After five school days of training, the students were given a standardized

   test of spatial reasoning. Score on the spatial reasoning test is the response variable.

   You will define a regression model for this factorial analysis of variance. Don’t write

   the model yet.

   (a) In the table below, show how your dummy variables are defined. *Use effect coding.*

   That’s the scheme with an intercept and minus ones. Write the name of each

   dummy variable at the head of its column.

<table>
<thead>
<tr>
<th>Girls, Programme 1</th>
<th>Girls, Programme 2</th>
<th>Girls, Programme 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys, Programme 1</td>
<td>Boys, Programme 2</td>
<td>Boys, Programme 3</td>
</tr>
</tbody>
</table>
(b) Give $E[Y_i | X_i = x_i]$ for the full model. Include the interaction terms. Notice you are *not* being asked to write expected values in the table. They are too messy.

(c) Suppose you want to test whether, averaging across training programmes, there is a difference between girls and boys in their average performance on the spatial reasoning test. State the null hypothesis in terms of the $\beta$ values in your model.

(d) Suppose you want to test whether, averaging across boys and girls, there is a difference between training programmes in average performance on the spatial reasoning test. State the null hypothesis in terms of the $\beta$ values in your model.

(e) Suppose you want to test whether the sex difference in average performance depends on which training programme the children are in. State the null hypothesis in terms of the $\beta$ values in your model.

5. Steel is made by heating iron and adding some carbon. A steel company conducted an experiment in which knife blades were manufactured using two different amounts of carbon (Low and High), and three different temperatures (Low, Medium and High). Of course even the Low temperature was very hot. A sample of knife blades was manufactured at each combination of carbon and temperature levels, and then the breaking strength of each blade was measured by a specially designed machine. The response variable is breaking strength.

(a) In a table with one row for each treatment combination, please make columns giving the coefficients of the contrast or contrasts you would use to test for main effects of Temperature.

(b) In another table with one row for each treatment combination, please make columns giving the coefficients of the contrast or contrasts you would use to test the Temperature by Carbon Level interaction.

(c) In one last table with one row for each treatment combination, please make columns showing how you would set up dummy variables for both independent variables, using *effect coding* (that’s the scheme with the -1s).

(d) Write $E(Y | X = x)$ for the regression model, using the names from your table above. Include the interactions!

(e) Using the $\beta$ values from your answer to the preceding question, state the null hypothesis you’d use to test whether the effect of carbon level on breaking strength depends on the temperature.