STATS 762 Assignment 5 Extra Question

The data used in this assignment comes from a paper "Heat as a Factor in the Penetration of Cloth Ballistic Panels by 0.22 Caliber Projectiles" by Prosser, Cohen and Segars that appeared in *Textile Research Journal* (2000). The data is in a file called **Ballistics Data** in "Data Sets" on the STATS 330 webpage.

This paper investigates the penetration of cloth panels (made of Spectra-1000) by 0.22 caliber bullets. The response, V_{50}, represents the velocity at which 50% of the shots are stopped by the cloth. Two explanatory variables were recorded: the type of projectile and the number of layers of cloth. Type of projectile is a qualitative variable where the three levels represent rounded (level 1), sharp (level 2) and intermediate (level 3) projectiles.

From theoretical considerations, it is thought that the relationship between $(V_{50})^2$ and the number of layers of cloth should be linear but that the intercepts and slopes may be different for the different types of projectiles. A regression analysis is to be done that will allow us to investigate this. **Note: throughout this assignment use $(V_{50})^2$ as the response**.

For a qualitative variable (such as projectile type) that has three levels we need to include two columns in the model matrix (in general the number of columns is the number of levels minus one). One way of designating these columns is the "baseline" approach:

projectile type	col-1	col-2
1-blunt	0	0
2-sharp	1	0
3-intermediate	0	1

1. First do an added variable F-test to test whether the slopes of the lines that describe the relationship between $(V_{50})^2$ and the number of layers of cloth are the same for the three different types of projectile. Use the following steps:

   (a) In *R* create the response vector $\mathbf{y}$, the model matrix for the full model $\mathbf{X}_F$ and the model matrix for the submodel $\mathbf{X}_S$ using the baseline model described above. Also create the projection matrix $\mathbf{P}_F$ for the full model and the projection matrix $\mathbf{P}_S$ for the submodel. Note: you don't need to print out these matrices but you do need to give the *R* commands you used to create them.

   (b) Find the squared length of the projection of $\mathbf{y}$ onto the model space for the full model and for the submodel.

   (c) Find the squared length of the projection of $\mathbf{y}$ onto the error space (for the full model).

   (d) Using your results from (b) and (c) find the value of the F-statistic for this test and the corresponding p-value. What do you conclude based on this F-test?

2. Now consider working with the common slope (i.e. no interaction) model for this data.

   (a) Find the fitted values and the residuals by projecting $\mathbf{y}$ onto the "model space" and onto the "error space". Use the vector of residuals to estimate the error variance σ^2.

   (b) Find (and print out) the vector of estimated coefficients $\hat{\boldsymbol{\beta}}$ and the estimated covariance matrix for $\hat{\boldsymbol{\beta}}$.

(c) Find a 95% confidence interval for the difference in the expected response between a sharp projectile with 20 layers and a intermediate projectile with 30 layers.

3. There are many (countless actually) ways that the indicator columns for the factor `type` could be created other than the baseline approach given above,

   Consider the following two possible ways of setting up these columns:

	Method A		Method B	
projectile type	col-1	col-2	col-1	col-2
1-blunt	0	0	0	1
2-sharp	1	0	1	0
3-intermediate	1	1	0	1

   (a) One of these methods does produce a valid set of indicator columns and one does not. Identify which is which and explain how you arrived at your answer.

   (b) For the method that does produce a valid set of indicator columns, explain what the coefficient for each of the columns represents in terms of the parallel lines model.

   (c) How do the coefficients for this new parameterisation relate to the coefficients for the baseline representation that was used in part B.