DUE: Friday, October 21, 11:59PM (64 points scaled to 100 percent)

You may discuss this assignment with whomever you wish, but please prepare and submit work in groups of ***ONE to THREE*** students, no more and no fewer. ***Each group*** will submit a copy of their group's completed assignment, via BbLearn, including the ***names and student ID numbers of all group members*** who participated on the assignment. If you discover a mistake, you may submit another version before the deadline. The last submitted (on-time) version will be graded, with all team members receiving the same score, which will be recorded in BbLearn.

GROUP MEMBERS WHO DO NOT CONTRIBUTE SUBSTANTIALLY TO AN ASSIGNMENT MAY BE REQUIRED TO WORK IN THEIR OWN GROUP OF ONE FOR THE REMAINDER OF THE SEMESTER.

While you are permitted to discuss the assignment with other groups, please prepare your own group's code/output and written answers. GROUPS WHOSE CODE AND SOLUTIONS APPEAR SUBSTANTIALLY SIMILAR MAY BE SUBJECT TO A 10% PENALTY.

Please prepare solutions in a ***neat, organized and concise fashion***! I prefer typeset presentations (e.g., cut and paste code/output into MS Word with added exposition when appropriate; knitr via EMACS and ESS; knitr or R Markdown via RStudio, the latter being the method preferred by students in recent years). At the very least, you need to ensure code and output are presented with a fixed-width font. Neatly handwritten presentations may also be appropriate for some problems. Sloppily prepared or disorganized solutions will not receive full credit.

To complete the items below, I expect you to find and use material in our lecture notes, including code/output, possibly after some modification. Some questions may be answered with code and output alone, but some exposition may be required beyond code and output for other questions. It's up to you to communicate concisely!

1. (a) Use `lm` to regress distance (`Dist`) (feet) on the four covariates of right leg strength (pounds), left leg strength (pounds), right hamstring flexibility (inches) and left hamstring flexibility (inches) in the high school boys `longjump` data in the the file `longjump.RDS` in BbLearn; the covariate names in the data frame are obvious. (Use `readRDS` to read this data frame into `R`.) Use `summary` output to determine if these four covariates *collectively* have a significant ($\alpha = 0.05$) relationship with distance. Report your hypotheses, p-value, conclusion, code and output. (5 points)

   (b) Can we infer a causal effect of covariates on the response? Explain. (3 points)

   (c) Can we infer an association of covariates with the response beyond the sample subjects represented in these data? Explain. (3 points)

   (d) Use the `anova` function, with an appropriate reduced model, to perform the same test as in item (a), above. Report your code and output. (No need to repeat other requested parts of (a) as results should be the same, of course!) (2 points)

(e) Use the `glh.test` function to perform the same test as in the previous problems. Report your code and output. (No need to repeat other items as results should be the same, of course!) (2 points)

(f) Continuing to use the model fit in part (a), above, compute a 2-sided level $\alpha = 0.05$ t-test of whether the right and left leg strengths have the same effects on (mean) distance jumped. You may use any method that we've discussed to do this. Report your hypotheses, p-value, conclusion, code and output. ***Also***, what are the units of these effects and their difference? (You might use different methods to check your results, but please report just one method in your code/output.) (6 points)

(g) Use `gmodels::estimable` to construct a 95% 2-sided confidence interval for the above difference of effects and report briefly your interval. (3 points)

(h) Similar to the test and interval for the difference in leg strength effects on distance jumped, above, test ($\alpha = 0.05$) whether the right and left hamstring flexibilities have the same effects on distance jumped and give an associated 95% interval. (6 points)

(i) Testing the previous two hypotheses simultaneously may be considered a test of the null hypothesis of "leg symmetry." That is, if the effects of leg strengths are equal and the effects of hamstring flexibility are equal, then legs are somehow "symmetric." Perform this simultaneous test for leg symmetry. You may use any method that we've discussed to do this. Report your code, output and conclusion. (8 points)

2. With so few observations, it is difficult to assess our model assumption of normality. Use the permutation procedure described in §3.3 (textbook or notes) to test the overall association of the inputs with distance jumped. Use the F statistic as your test statistic, but use the permutation distribution instead of an F distribution to compute a p-value. Actually, we use a Monte Carlo approximation as in the book and notes. Use `set.seed(8675309)` just before you randomly sample 5000 permutations. Compare to the normal theory results. Show your code and results and discuss appropriately. (6 points)

3. Use the bootstrap procedure as discussed in §3.6 (textbook or notes) to compute a 95% bootstrap confidence intervals, one for the difference in the effects of leg strengths on jump distance and another for the difference in the effects of leg flexibilities on jump distance. Use `set.seed(5551212)` and 5000 bootstrap samples. Show your code, output and report your results. How do these compare with the normal theory intervals, above? (8 points)

4. Here, we use the body fat data as in the example of Section 4.2 of our notes. (Fit the same model as fit in that section.)

(a) Compute both a 95% confidence interval for the population mean percent body fat (brozek score), $E(Y \mid \mathbf{x}_0)$, and a 95% prediction interval for an individual $Y \mid \mathbf{x}_0$, where $\mathbf{x}_0$ denotes the particular characteristics of a population of patients in which we are interested, in the first case, or denotes the characteristics of an individual, in the latter case. For this purpose, use the average (not the median) values of the covariates. (In practice, for an individual, we would use that individual's characteristics.) Report your code and output and give a brief summarizing statement for each interval..."we are 95% confident that..." similar to your previous interval reporting. (6 points)

(b) Which interval do you think a medical doctor may use to assess a patient? Answer briefly. (3 points)

(c) Which interval do you think an exercise science researcher may use to infer about the relationship of percent body fat to these characteristics? (3 points)