Homework Assignment 7
Total points: 60
Due: Friday November 4, 2022 at 5PM

Please email your answer (compiled pdf file from R markdown) and R code to Yen-Yi Ho (hoyen@stat.sc.edu).

Instructions: feel free to discuss the homework with other students. However, each student must conduct their own analyses and write-up their own solutions. Write as if for a scientific journal. Be brief and accurate.

Use the WHO Child Growth Standards (IGROWUP) data for child age **1-6 year** to study the dependence of weight on age with and without adjustment for height. IGROWUP data is available http://people.stat.sc.edu/hoyen/Stat704/Data/survey.csv. More information about IGROWUP data can be found in http://www.who.int/childgrowth/en/

Added Variables Plot (AVPLOT)

Use the IGROWUP dataset for children older than 12 months to do the following.

1. Plot arm circumference (AC: variable named **MUAC** in the dataset) against agemons and briefly describe the nature of their association. (3 points)

2. Regress AC on age and add the least square line to the plot. Comment on the nature of the association and on any apparent departure from linearity. (3 points)

3. Now regress AC on age and height. Compare the age coefficient with and without adjustment for height. Explain in simple terms why the coefficient for age has changed once we added height to the model. (6 points)

The remainder of this problem addresses the question: what scatter plot can we make so that the best fitting line corresponds to the MLR coefficient.

4. Make an "added variables plot" of AC on age, adjusted for height; specifically: (10 points)
    a. Regress AC on "the other variables besides age" (here height) and save the residuals (R(AC|H). These residuals can be thought of as "AC adjusted for height" or as "the par of AC that is not linearly predictable

by height" or the "deviation in AC from what you would have expected given the child's height."

b. Regress the "predictor of interest", (here agemons) on the other variables (here height) and save the residauls (R(agemons|H)). These are the "height-adjusted ages."

c. Now plot R(AC|H) against R(age|H)

5. Regression R(AC|H) against R(age|H) using simple linear regression (SLR). Note that the intercept is 0 (in R it might be $a \times 10^{-17}$ due to numerical rounding error). Explain why in a sentence or two. Compare the slope and standard error with the MLR coefficient for agemons from Step 3. (10 points)

6. Now write a sentence or two that explains the meaning of the MLR coefficient in terms of the AVPLOT idea. (5 points)

7. Inspect the age AVPLOT and decide whether a linear relationship is reasonable. Comment in a sentence or two. (5 points)

8. Now make an AVPLOT for height and inspect it to see whether the linear height assumption given age is reasonable. (5 points)

9. Give an algebraic proof that the coefficient from an SLR using the adjusted response and adjusted predictor of interest is equal to the MLR coefficient for that variable. (13 points). [Hint: see the property for the inverse of a partitioned matrix on the next page]

A.68. Find the Cholesky factorization for the positive semidefinite matrix

$$\begin{bmatrix} 4 & 2 & 2 & 2 \\ 2 & 5 & -3 & 7 \\ 2 & -3 & 5 & -5 \\ 2 & 7 & -5 & 19 \end{bmatrix}$$

A.69. * Find the inverse of the matrix $\mathbf{I} + \mathbf{ab}^T$. Hint: Try the form $c\mathbf{I} + d\mathbf{ab}^T$ and find c and d. What happens if $\mathbf{a}^T\mathbf{b} = -1$?

A.70. Let $\mathbf{D}$ be a diagonal matrix; find the inverse of $\mathbf{D} + \mathbf{11}^T$.

A.71. * Let $\mathbf{D}$ be a diagonal matrix; find the inverse of $\mathbf{D} + \mathbf{ab}^T$.

A.72. Verify these results for the inverse of a partitioned matrix:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{BE}^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{BE}^{-1} \\ -\mathbf{E}^{-1}\mathbf{CA}^{-1} & \mathbf{E}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{F}^{-1} & -\mathbf{F}^{-1}\mathbf{BD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{CF}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{CF}^{-1}\mathbf{BD}^{-1} \end{bmatrix}$$

where $\mathbf{E} = \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}$ and $\mathbf{F} = \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}$.

A.73. Let $\mathbf{V}$ be a symmetric $p \times p$ matrix with eigenvalues $\lambda_1, \cdots, \lambda_p$; show the following:

a. $|\mathbf{I}_p + t\mathbf{V}| = \prod_{i=1}^{p}(1 + t\lambda_i)$.

b. The derivative of $|\mathbf{I}_p + t\mathbf{V}|$ with respect to t is equal to $trace(\mathbf{V})$.

A.74. Prove Result A.19.

A.75. (Binomial inverse theorem) Verify the following for $\mathbf{A}$, $\mathbf{B}$ nonsingular:

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}^{-1}(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{BVA}^{-1}$$