

**Data Contest Questions**

Address as many of the questions as you can. Don’t sweat it if you can’t answer all of these questions in the allotted time. Just focus on the easier ones to start off and see how far you can get. Use any resources you wish. Feel free to consult the internet and any books or notes. Also, feel free to use any previous code/syntax that you previously developed. No AI!

**Lower difficulty**

1. What is the most frequent type for blood?
2. Calculate the average skin score per blood group.
3. Create a histogram of skin, but we want the density on the y-axis. We also want to overlay a curve that represents a normal distribution. Create this plot and comment on skin’s distribution.
4. Are brain size and iq related?
5. Are brain size and thalamus related?
6. We need to know how many standard deviations away are males and females (sex) on height.
7. Build a linear model. The outcome will be mk and the predictors will be age and sex. Create a table of the estimates and the squared semi-partials.
8. Build a linear model. The outcome will be mk and the predictors will be age, job, sex, and iq. Create a table of the estimates and the squared semi-partials.
9. What is the predicted military knowledge for a person who is 20 years old, has indicated a job preference of miner, is a female, and has an IQ of 128?

**Higher difficulty**

1. Compare the two linear models that you made. Which one makes better out of sample predictions given the number of effective parameters?
2. Create a scatterplot with mk on the y-axis and age on the x-axis. Also incorporate iq and sex. Facet that by job. Also add a regression line for each faceted plot. Describe what you see.
3. Your Martian superiors are interested in if humans of different blood groups are more intelligent. Specifically, they want an index of the relative strength of evidence from the data about the hypotheses (i.e., the change from prior to posterior odds brought about by the data). The numerator would be the hypothesis that blood groups do have an effect on intelligence and the denominator would be that they don’t. Calculate that and summarize what you found.
4. You no doubt noticed that there is some missingness for the variable, continent. Can you figure out the cause of the missingness? Include a plot if you can.