Module 1B

The Dangers and Use of Aggregating Data

Module 1 : Descriptive Statistics

Measurements of *location* and *spread* of data

Agenda:

- ~~Mean, mode, median~~

- ~~Variability, variation, range~~

- Simpson's paradox

- Intuitions about uncertainty: Fermi Estimation

- Accuracy/Bias and Precision/Spread

Will Rogers Phenomenon

"When the Okies left Oklahoma and moved to California, they raised the average intelligence level in both states."

https://en.wikipedia.org/wiki/Will_Rogers_phenomenon

Actual medical phenomenon: "medical stage migration"

Example: There were more COVID deaths among the vaccinated than the unvaccinated. In September 2022, 12,593 COVID deaths occurred in the United States. Of those, 39% were unvaccinated, while 61% were vaccinated. **WHY?**

Another general example: "Genomic aggregation effects and Simpson's paradox" "(2014, Brimacombe)

Yet another example: "No Gene is an Island: The Flip-Flop Phenomenon" (Lin et al, 2007)

Simpson's Paradox: Activity vs Glucose12
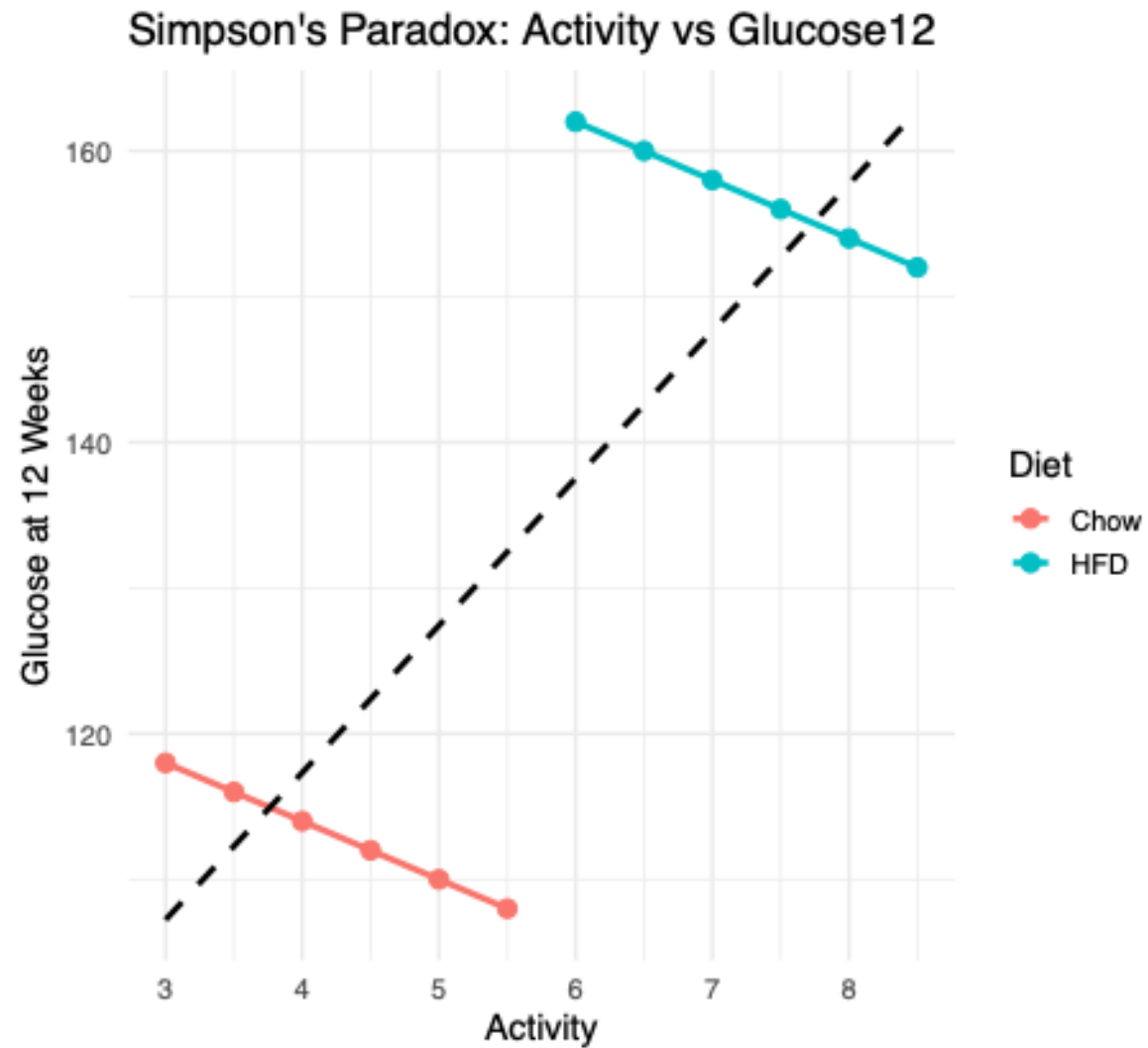
Within each diet: Activity decreases as Glucose decreases (negative slope)

Overall: Activity increases as Glucose increases (positive slope)

Fermi Estimation*

- A good way to practice so you aren't bamboozled so easily
- https://www.njaapt.org/resources/Documents/Physics Olympics -- All/Fermi Questions - Worksheet and Answers.pdf

- **Question:** Genes are composed of exons (and introns) and all the exons in the genome comprise the exome. Using a Fermi estimation, what is a reasonable estimate for the size of the human genome?

- **Question:** Using the 12-mouse colony dataset – if this was scaled up to a colony of 2000 mice, how many hyperglycemic mice would you expect under the same diet proportions?

* The link to how to improve/justify your estimation is here in case you want to use it as a party trick