Data Exam One – Data Description

36-402 Advanced Methods for Data Analysis

Due Friday, April 2, 2021 at 3pm EDT
(grace period until Sunday, April 4, 2021 at 8pm EDT)

The AnAge Database of Animal Ageing and Longevity is a "curated database of ageing and life history in animals" that was "primarily developed for comparative biology studies". It contains information, compiled from hundreds of scientific papers, for over 4,200 species. It's mainly meant for studies of aging, so for each species, it includes variables such as lifespan, age of sexual maturity, adult body mass, typical body temperature, and metabolic rate. We will study this data to learn more about the relationships between each variable and the key outcome variable: lifespan.

More information about the AnAge database is available here: `https://genomics.senescence.info/species/index.html`

Relevant Variables

The data file `anage-chordata-data.csv` contains 347 observations of 14 variables. It contains all entries from AnAge for animals in the Chordata phylum, which, roughly speaking, includes animals with spinal cords. We have also removed entries with missing variables. The variables are:

HAGRID A unique ID for each entry (the Human Ageing Genomic Resources ID)

Kingdom Along with the following variables (phylum, class, order, family, genus, and species), this gives the taxonomic classification of the species. Species names are typically given as genus and species, e.g. *Anaxyrus americanus* is the species name for the American toad. See Wikipedia for a description of taxonomic ranking: `https://en.wikipedia.org/wiki/Taxonomic_rank`

Common.name Common name (i.e., the name used by ordinary people, not scientists) for the animal.

Maximum.longevity.yrs Maximum longevity (lifespan), in years.

Body.mass.g Typical adult body mass, in grams.

Metabolic.rate Typical resting metabolic rate (i.e., rate of energy use), in Watts.

Temperature Typical body temperature, in Kelvin.

Your Goals

You have been hired by an eccentric billionaire who is willing to do whatever it takes to become immortal. The billionaire, Preston Jorgensen, would like you to review the AnAge data to determine what factors appear to be related to lifespan, so he can fund research projects that try to use those factors to attain immortality.

Specifically, Jorgensen wants you to answer three questions:

(i) Some research has suggested that slowing the metabolic rate—for example, by intermittent fasting—may increase lifespan. Is this supported by the evidence? Develop a model for lifespan using metabolic rate, and interpret what it means.

(ii) Is the relationship between metabolic rate and lifespan nonlinear, even after transformations? Use a nonparametric model to determine if a nonlinear fit is more appropriate.

(iii) Jorgensen would like a pet to keep him company during his immortality. He thinks the crab-eating raccoon (*Procyon cancrivorus*, HAGRID = 1898) would be a good companion. If he pays for research to reduce its metabolic rate by 50%, what does your model estimate for the mean lifespan of an animal with the crab-eating raccoon's characteristics but a 50% smaller metabolic rate?

Report a confidence interval, and explain whether your model supports the conclusion that the metabolic rate treatment would *cause* this change in mean lifespan.

The template Rmd file **guides you through answering these questions** one step at a time, so you should follow the template as you work. It also gives more specific detail about what specific models and methods you should use.

Your Introduction and Conclusion should give your results in plain English, so Jorgensen can read your answers to the questions; the rest of your report should be written so it can be ready by someone moderately familiar with regression.

Note that our focus is inference, not prediction. You must address the questions in the R Markdown template—submissions that do not address these questions will not receive credit.