Today

- Homework questions: bootstrapped CI
- Bootstrapped prediction intervals
- Designing and coding a bootstrapped p-value algorithm

Homework check in

- Quantity
  - 15% too much
  - 85% just right
- Difficulty
  - 15% too hard
  - 85% just right/too easy
- Completion
  - 70% have completed >75%

Ideas:

Thank you!

Reading links

Deadlines

Code feedback

Piazza

- Thanks for asking questions and sharing code!
- If you're stuck with code, post here. You can be anonymous.
- Conceptual questions are also great. I love it when you ask for clarification or refreshers, or basics you never covered!

LLM assistants

- Large language model (LLM)
  - code suggestions, general questions
  - GPT4 (commercial, openAi, cloud)
  - LLAMA (local, private, opensource)
- Warning!
  - Fluent and plausible replies but ...
  - often wrong (ca 15% in my experience)
  - Good enough to be helpful; need to check

LLM assistants

- Option (more technical to install)
  - https://mlverse.github.io/chattr/
  - delete tidyverse prompt in app options
  - prompt:
    - I am using base R
  - use with openAI developer API
    - cheap, PAYG
    - access to latest LLMs

Bootstrap prediction interval

- Prediction uncertainty for new y
- 05_7_bootstrap_prediction_interval.md
- Powerful idea: estimate uncertainty by
  - repeatedly
  - simulate fitting the model on a sample (parameter uncertainty)
  - simulate generating data from the fitted model (data generating process)

Bootstrap prediction interval

Algorithm

define a grid of new x values to predict y
repeat very many times
- sample from the error distribution of DGP
- simulate new y-values from original estimated parameters of model
- fit the model (estimate parameters: beta_0, beta_1, sigma_e)
- keep: simulate new data y|x using estimated parameters

calculate quantiles of the generated data distributions
plot quantiles

parametric

simulate generating data from the fitted model

simulate fitting the model on a sample

Bootstrapped p-value

- Learning goals
  - Understand p-values by understanding the underlying sampling algorithm
  - Further understand how the sampling distribution is the basis for frequentist inference
  - Understand how bootstrap algorithms mimic the sampling distribution algorithm
  - Formulate a bootstrap algorithm and translate it to R

Bootstrapped p-value

- Parametric bootstrap for H_0: $\beta_1 = 0$
  - what is the definition of a p-value?
  - what is the algorithm for parametric bootstrap?
  - combine these concepts
- Pseudocode first!
- R code from pseudocode

Definition of a p-value

The probability of a sample statistic as large or larger than the one observed given that some hypothesis is true

Basic parametric bootstrap algorithm

repeat very many times
    sample from the error distribution
    create new y-values from the estimated parameters and errors
    fit the linear model
    estimate the parameters
plot sampling distribution (histogram) of the parameter estimates

plug in: create simulated data from model