Today

- Inference algorithms
- Recap & questions from homework
- Results for bootstrapped confidence intervals
  - sharing, questions
- Designing and coding a bootstrapped p-value algorithm

Homework check in

- Quantity
  - 30% too much
  - 70% just right
- Difficulty
  - 40% too hard
  - 60% just right
- Completion
  - 60% have completed >75%
- Bottleneck: learning to program

Miscellaneous

- Big ideas in data science
- Coding style
  - Working together
  - Agreeing on a style

Git skills

- Branching & merging

Sampling distribution algorithm

repeat very many times
    sample data from the population
    fit the model
    estimate the parameters
plot sampling distribution (histogram) of the parameter estimates

Bootstrap algorithm

repeat very many times
    generate data based on the sample ← plug in
    fit the model
    estimate the parameters
plot sampling distribution (histogram) of the parameter estimates

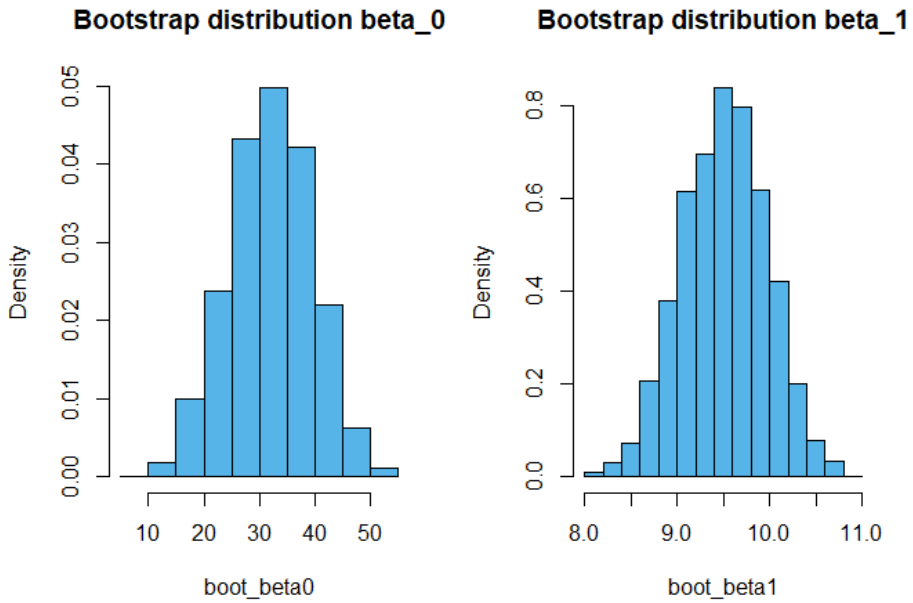
Bootstrap algorithms

- Non-parametric bootstrap
  - resample the data
- Empirical bootstrap
  - resample the residuals
- Parametric bootstrap
  - generate data from a distribution
  - use estimated parameters of the distribution

Code (e.g. empirical bootstrap)

```
for ( i in 1:10000 ) {
    e_boot <- sample(e_fit, replace=TRUE)
    df_boot$y <- coef(fit)[1] + coef(fit)[2]*df_boot$x + e_boot
    fit_boot <- lm(y ~ x, data=df_boot)
    boot_beta0[i] <- coef(fit_boot)[1]
    boot_beta1[i] <- coef(fit_boot)[2]
}
```

plug in



Pseudocode

For many times

- Resample errors from model fit (with replacement)
- Create new y-values at original x values
- Fit the model
- Keep parameter estimates

Bootstrapped CI

- For β_0, β_1
- For y|x
- Sharing and questions

Bootstrapped p-value

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