Power Simulation

Ryan Schneider, M.A.

2022-11-18

Table of Contents

## I. Generate practice sample

Using 2C:35-7 Drug Free School Zone as an example. This ranges from 12 to 63 months, for row E, across all time points.

# define the means and SD's  
mu\_struc=25  
sd\_struc=5  
  
mu\_UNstruc=35  
sd\_UNstruc=5  
   
# determine the group size  
n <- 50  
  
  
# simulate the data  
set.seed(1)  
  
d\_RS <-  
 tibble(group = rep(c("structured", "unstructured"), each = n)) %>%   
 mutate(treatment = ifelse(group == "structured", 0, 1),  
 y = ifelse(group == "structured",   
 rnorm(n, mean = mu\_struc, sd = sd\_struc),   
 rnorm(n, mean = mu\_UNstruc, sd = sd\_UNstruc)))

## II. Create practice model

fit\_lin=brm(data = d\_RS,   
 family = gaussian,  
 y ~ 1 + treatment,  
 prior = c(prior(normal(30, 5), class = Intercept),  
 prior(normal(0, 3), class = b),  
 prior(lognormal(0, 5), class = sigma)),  
 seed = 4,  
 file = here::here("fitted models", "test\_model"))  
  
  
model\_parameters(fit\_lin, ci\_method = "HDI") |> as\_tibble() |> select(1, 3, 5:6) |> format\_table()

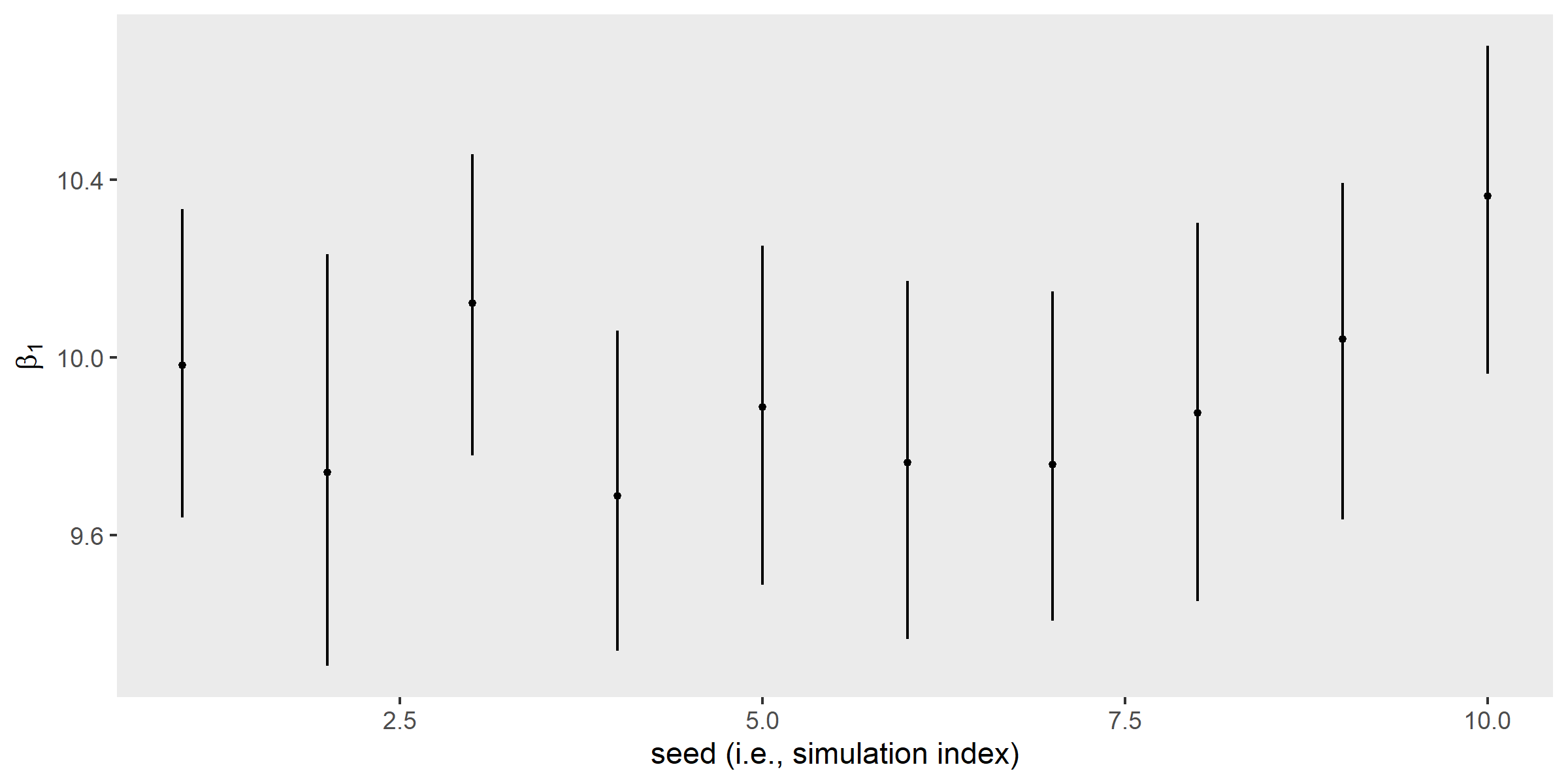
## Parameter Median 95% CI  
## 1 (Intercept) 25.90 [24.76, 27.23]  
## 2 treatment 9.24 [ 7.48, 10.96]  
## 3 sigma 4.54 [ 3.91, 5.19]

## III. Simulate

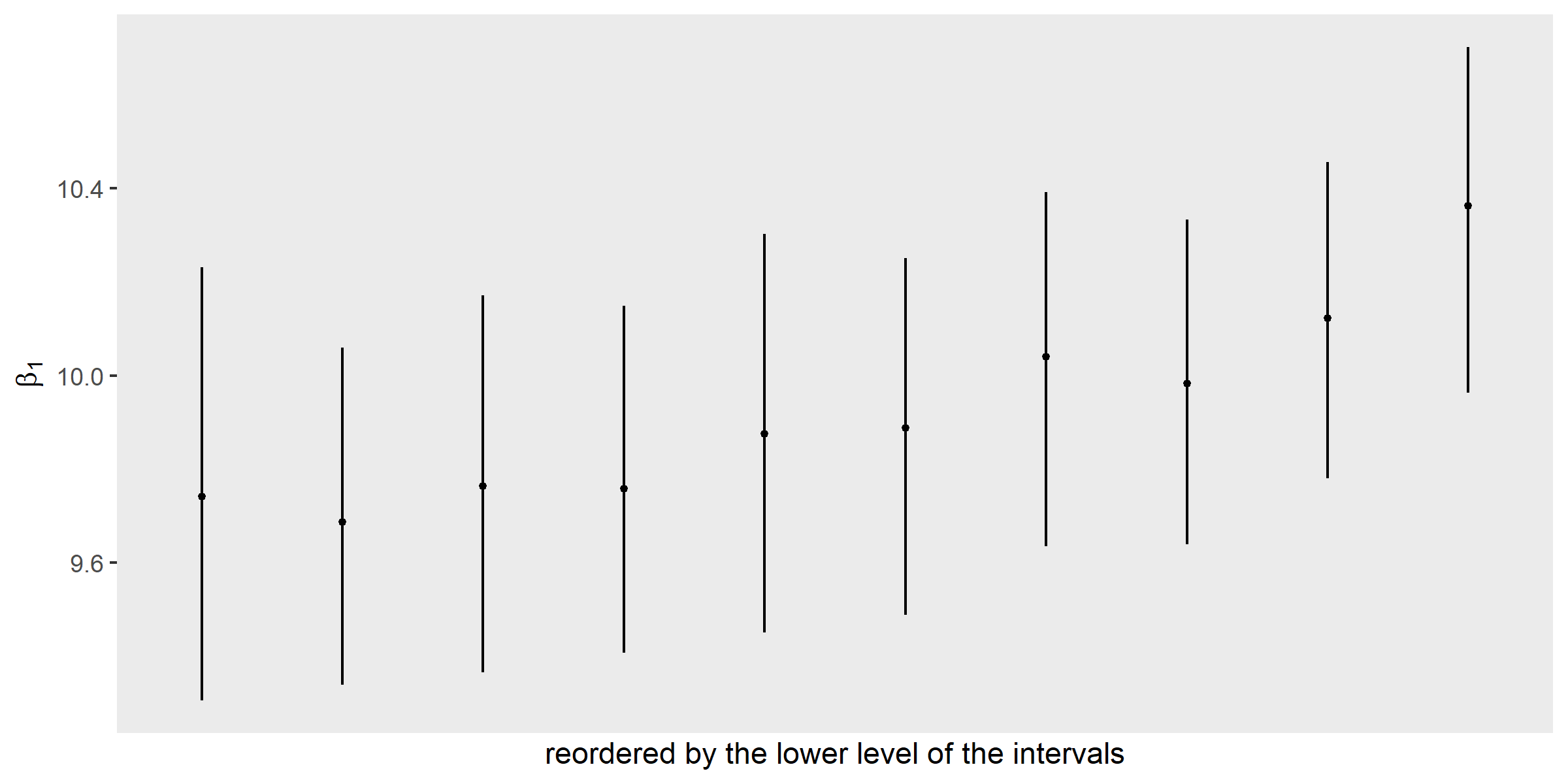
## IV. Check results

Plot the interval widths. Can we reject a value in the ROPE?

theme\_set(theme\_grey() +  
 theme(panel.grid = element\_blank()))  
  
  
sim |>   
 ggplot(aes(x = seed, y = Median, ymin = CI\_low, ymax = CI\_high)) +  
 #geom\_hline(yintercept = c(0, .5), color = "white") +  
 geom\_pointrange(fatten = 1/2) +  
 labs(x = "seed (i.e., simulation index)",  
 y = expression(beta[1]))



sim |>   
 ggplot(aes(x = reorder(seed, CI\_low), y = Median, ymin = CI\_low, ymax = CI\_high)) +  
 #geom\_hline(yintercept = c(0, .5), color = "white") +  
 geom\_pointrange(fatten = 1/2) +  
 scale\_x\_discrete("reordered by the lower level of the intervals", breaks = NULL) +  
 ylab(expression(beta[1]))

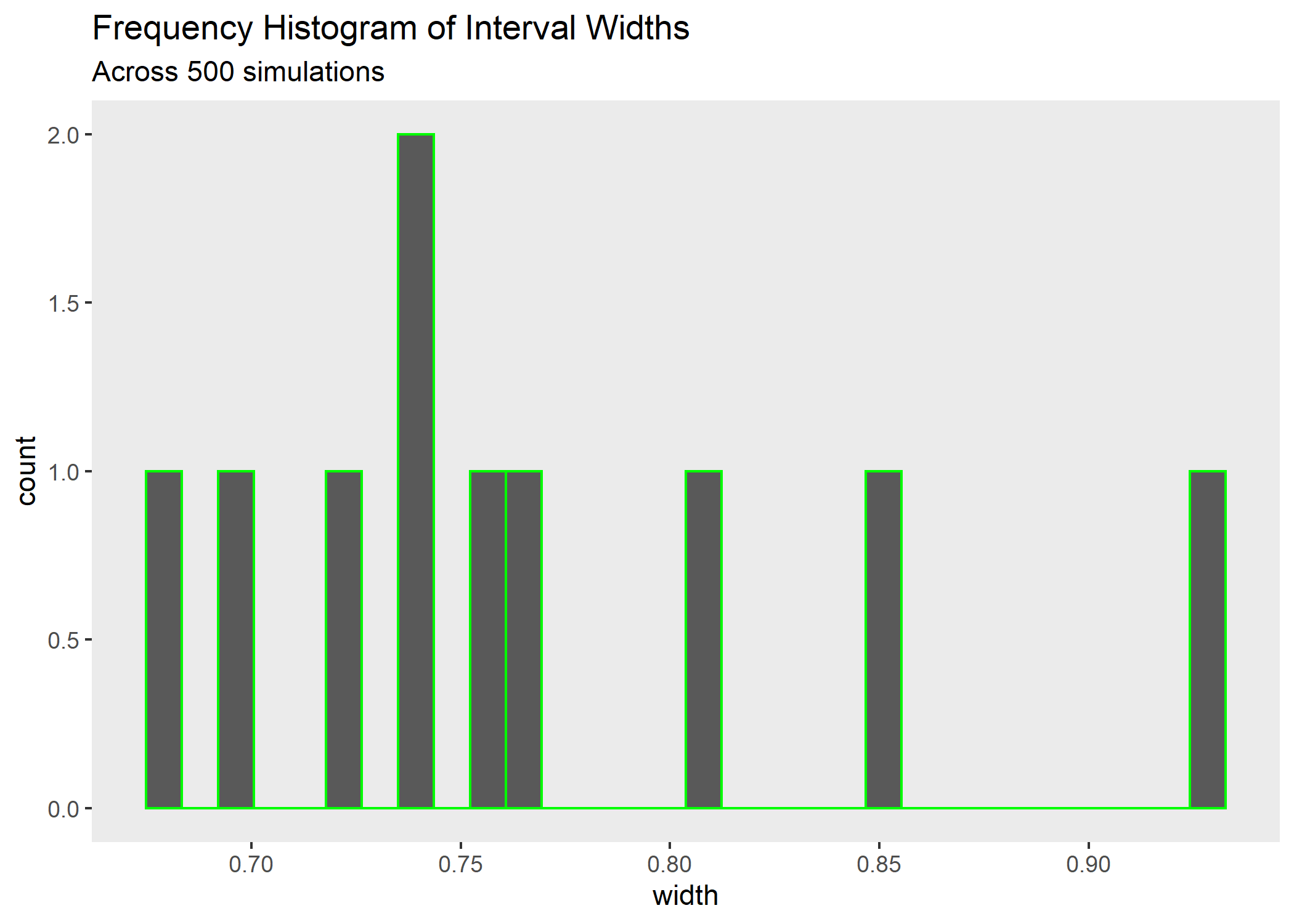


coord\_cartesian(ylim = c(-.5, 1.3))

## <ggproto object: Class CoordCartesian, Coord, gg>  
## aspect: function  
## backtransform\_range: function  
## clip: on  
## default: FALSE  
## distance: function  
## expand: TRUE  
## is\_free: function  
## is\_linear: function  
## labels: function  
## limits: list  
## modify\_scales: function  
## range: function  
## render\_axis\_h: function  
## render\_axis\_v: function  
## render\_bg: function  
## render\_fg: function  
## setup\_data: function  
## setup\_layout: function  
## setup\_panel\_guides: function  
## setup\_panel\_params: function  
## setup\_params: function  
## train\_panel\_guides: function  
## transform: function  
## super: <ggproto object: Class CoordCartesian, Coord, gg>

# compute exact HDI widths and summarize all 500 of them  
sim <-  
 sim |>   
 mutate(width = CI\_high - CI\_low)  
  
# view HDI widths as a histogram  
sim |>   
 ggplot(aes(x = width)) +  
 geom\_histogram(binwidth = NULL, color="green")+  
 labs(title = "Frequency Histogram of Interval Widths", subtitle = "Across 500 simulations")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



width\_descriptives=sim |> summarize('Median width'=median(width),  
 'Mean width'=mean(width),  
 'Narrowest width'=min(width),  
 'Widest width'=max(width)) |>   
 mutate(across(everything(), round, 2))   
  
  
width\_descriptives |> pivot\_longer(cols = everything(),  
 names\_to = "HDI Summary Stat",  
 values\_to = "Value") |>   
 flextable::flextable() |>   
 flextable::autofit()

| HDI Summary Stat | Value |
| --- | --- |
| Median width | 0.75 |
| Mean width | 0.77 |
| Narrowest width | 0.68 |
| Widest width | 0.93 |

The posterior shows that the most probable estimate for is about 0.6, and the intervals range from about [0.2, 1.0]. That’s wide. Lots of uncertainty in that estimate.

## Minimum Degree of Precision

What if we could narrow the HDI from its current width of 0.8 to < .7? This code shows the probability of getting an interval width of less than the target X, for a given sample size

target\_width=.7  
  
sim |>   
 mutate(check = ifelse(width < target\_width, 1, 0)) |>   
 summarise(`width power` = mean(check))

## # A tibble: 1 x 1  
## `width power`  
## <dbl>  
## 1 0.2

report::cite\_packages(sessionInfo())

## Warning in utils::citation(pkg\_name): no date field in DESCRIPTION file of  
## package 'legaldmlab'

## Warning in utils::citation(pkg\_name): could not determine year for 'legaldmlab'  
## from package DESCRIPTION file

## - Ben-Shachar M, Lüdecke D, Makowski D (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. Journal of Open Source Software, 5(56), 2815. doi: 10.21105/joss.02815  
## - David Gohel (2021). flextable: Functions for Tabular Reporting. R package version 0.6.10. https://CRAN.R-project.org/package=flextable  
## - Dirk Eddelbuettel and Romain Francois (2011). Rcpp: Seamless R and C++ Integration. Journal of Statistical Software, 40(8), 1-18. URL https://www.jstatsoft.org/v40/i08/.  
## - Gabry J, Mahr T (2021). "bayesplot: Plotting for Bayesian Models." Rpackage version 1.8.1, <URL: https://mc-stan.org/bayesplot/>.  
## - H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.  
## - Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr  
## - Hadley Wickham (2021). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.1. https://CRAN.R-project.org/package=forcats  
## - Hadley Wickham (2021). tidyr: Tidy Messy Data. R package version 1.1.4. https://CRAN.R-project.org/package=tidyr  
## - Hadley Wickham and Evan Miller (2021). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.4.3. https://CRAN.R-project.org/package=haven  
## - Hadley Wickham, Jim Hester and Jennifer Bryan (2021). readr: Read Rectangular Text Data. R package version 2.1.1. https://CRAN.R-project.org/package=readr  
## - Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr  
## - Kirill Müller and Hadley Wickham (2021). tibble: Simple Data Frames. R package version 3.1.6. https://CRAN.R-project.org/package=tibble  
## - Lionel Henry and Hadley Wickham (2020). purrr: Functional Programming Tools. R package version 0.3.4. https://CRAN.R-project.org/package=purrr  
## - Lüdecke D, Ben-Shachar M, Patil I, Makowski D (2020). "Extracting,Computing and Exploring the Parameters of Statistical Models using R."\_Journal of Open Source Software\_, \*5\*(53), 2445. doi:10.21105/joss.02445 (URL: https://doi.org/10.21105/joss.02445).  
## - Lüdecke D, Waggoner P, Makowski D (2019). "insight: A Unified Interfaceto Access Information from Model Objects in R." \_Journal of Open SourceSoftware\_, \*4\*(38), 1412. doi: 10.21105/joss.01412 (URL:https://doi.org/10.21105/joss.01412).  
## - Lüdecke et al., (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software, 6(60), 3139. https://doi.org/10.21105/joss.03139  
## - Lüdecke et al., (2021). see: An R Package for Visualizing Statistical Models. Journal of Open Source Software, 6(64), 3393. https://doi.org/10.21105/joss.03393  
## - Makowski, D., Ben-Shachar, M. S. & Lüdecke, D. (2020). \*The {easystats} collection of R packages\*. GitHub.  
## - Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2019). Methods and Algorithms for Correlation Analysis in R. Journal of Open Source Software, 5(51), 2306. doi:10.21105/joss.02306  
## - Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). Estimation of Model-Based Predictions, Contrasts and Means. CRAN.  
## - Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40), 1541. doi:10.21105/joss.01541  
## - Makowski, D., Ben-Shachar, M.S., Patil, I. & Lüdecke, D. (2020). Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. CRAN. Available from https://github.com/easystats/report. doi: .  
## - Makowski, Lüdecke, Patil, Ben-Shachar, & Wiernik (2021). datawizard: Easy Data Wrangling. CRAN. Available from https://easystats.github.io/datawizard/  
## - Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01  
## - R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.  
## - Revelle, W. (2021) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.1.9,.  
## - Ryan Schneider (NA). legaldmlab: Legal Decision Making Lab Helper Functions for Common Tasks. R package version 0.0.0.9000.  
## - Sam Firke (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.1.0. https://CRAN.R-project.org/package=janitor  
## - Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P, Paananen T, Gelman A(2020). "loo: Efficient leave-one-out cross-validation and WAIC forBayesian models." R package version 2.4.1, <URL:https://mc-stan.org/loo/>.  
## - Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686