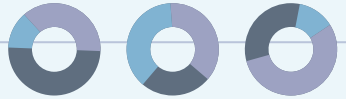


# Short Exact Confidence Intervals for the Parameters of the Negative Hypergeometric Distribution

**Rachel Roggenkemper**

Advisor: Dr. Bret Holladay

Committee Members: Dr. Jimmy Doi & Dr. Kevin Ross

Binomial

Negative Binomial

# The Forgotten Distribution

Hyper-geometric

# Relationships among Discrete Distributions

| | Sampling with Replacement | Sampling without Replacement |
|---|---|---|
| **# Trials Fixed** | Binomial | Hypergeometric |
| **# Successes Fixed** | Negative Binomial | **Negative Hypergeometric** |

# Negative Hypergeometric (NHG) Distribution

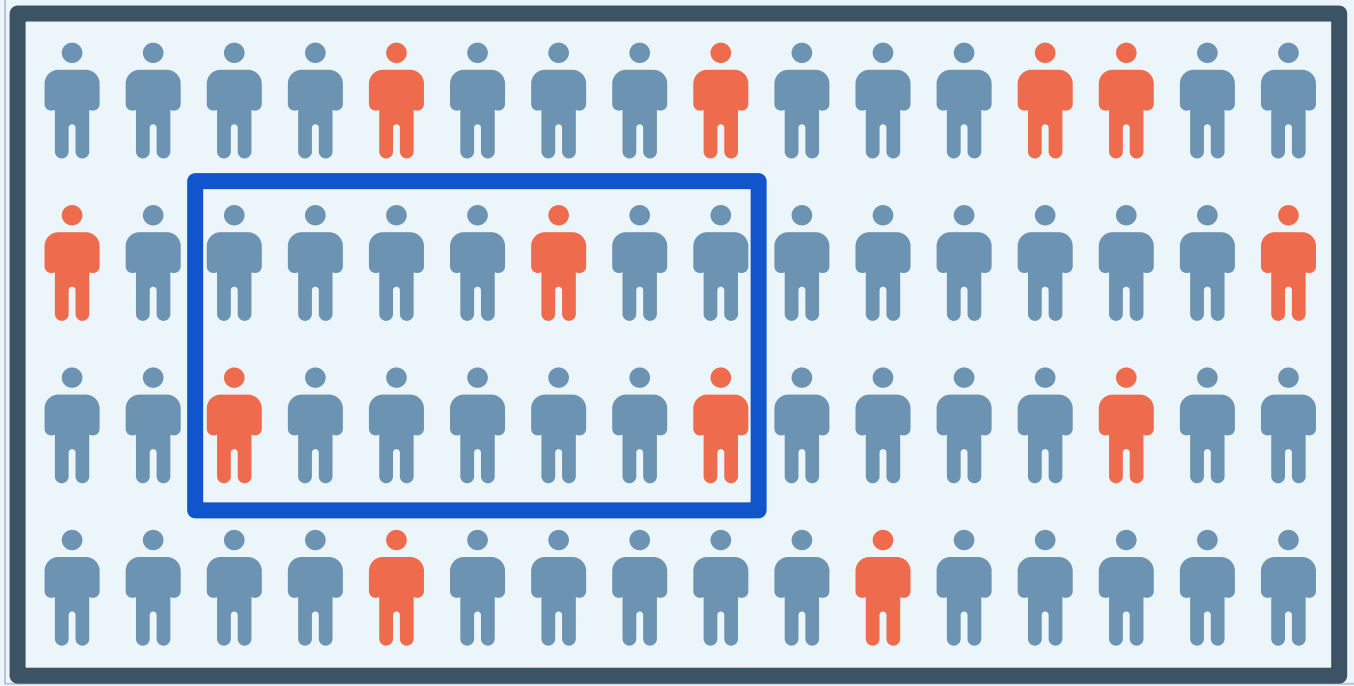
Used to model **the number of failures** needed to achieve a **fixed number of successes** when **sampling without replacement** from a **finite population**

**Notation:**

- $X \sim \text{NHG}(N, M, m)$ = number of failures sampled until m successes
- $N$ = total population size
- $M$ = number of successes in the population
- $m$ = fixed number of successes

# Estimation of *M*: Surveys

**Population:**

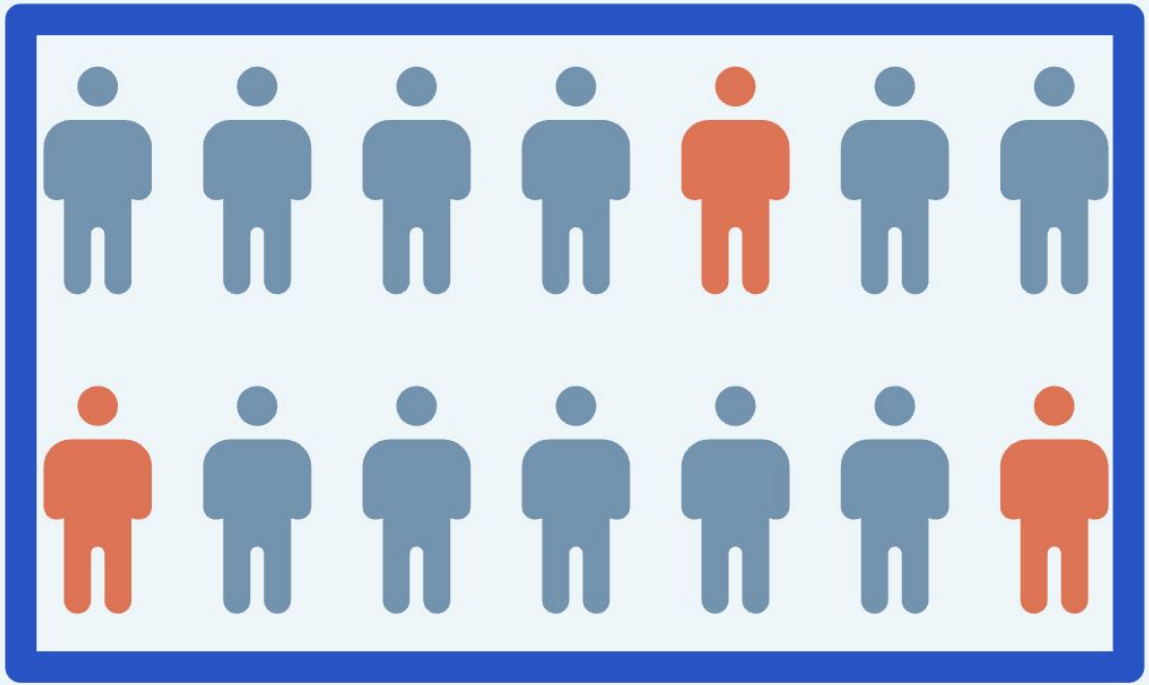


*N* = # Cal Poly students (known)

*M* = # students who have COVID (unknown)

# Estimation of *M*: Surveys

**Sample:**



Continue sampling until *m* COVID positive students sampled

*m* = # students who have COVID to sample (fixed)

# Estimation of *N*: Capture-Recapture Methods

Before sampling: all *N* penguins ***(unknown)*** are **unmarked**

# Estimation of *N*: Capture-Recapture Methods

***M* penguins *(known)*** are **captured** and **marked**

# Estimation of *N*: Capture-Recapture Methods

***M* marked penguins** are **released** back into the **population**

# Estimation of *N*: Capture-Recapture Methods

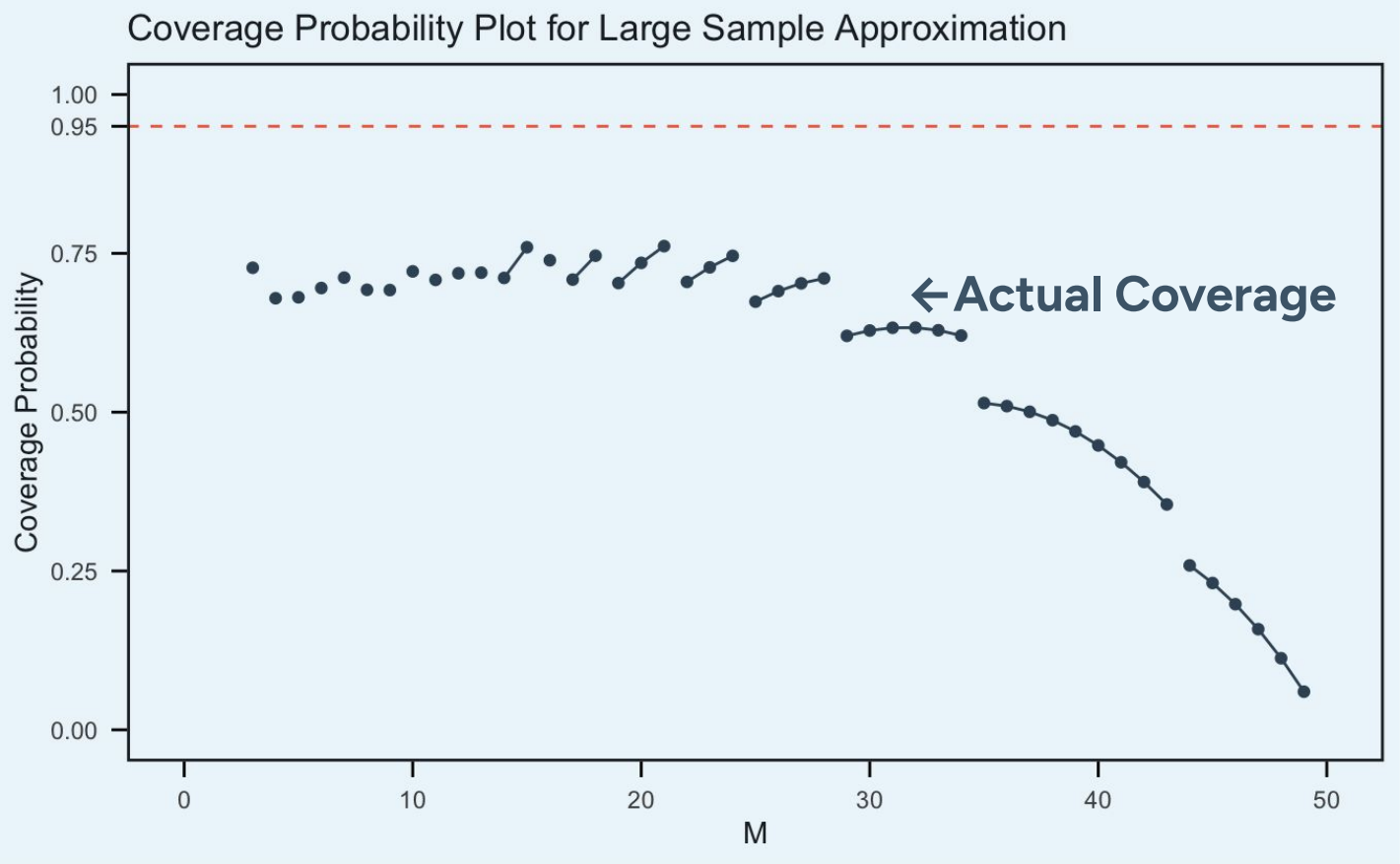
Sample ***until*** **m *(fixed)* marked penguins** are recaptured

# Confidence Intervals for Parameters of the Negative Hypergeometric:

(i) Estimating the population size ($N$ unknown)

(ii) Estimating the number of successes in a population ($M$ unknown)

# **Large sample methods** commonly used in practice often exhibit **poor coverage** for any **discrete distribution**

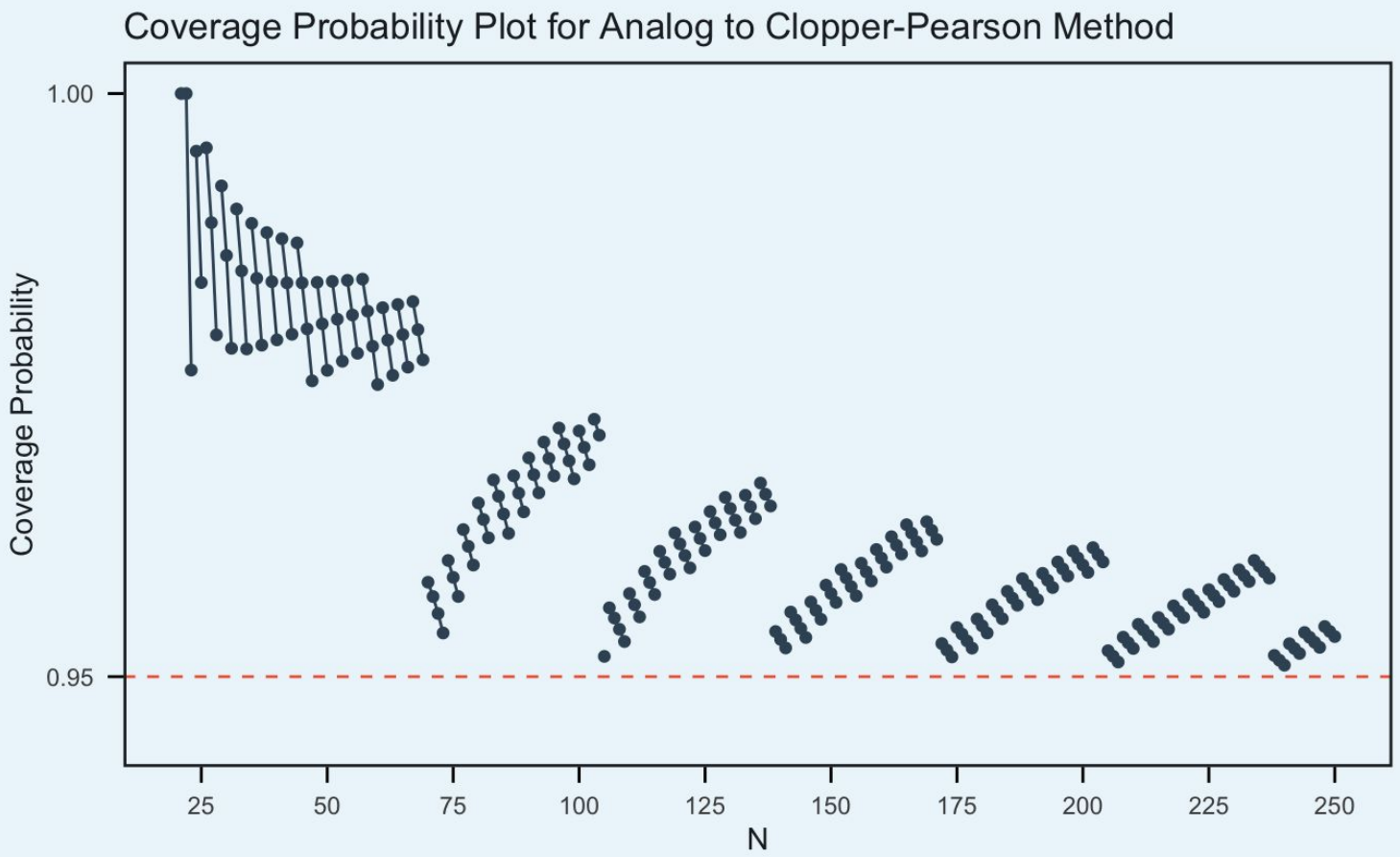



←Desired Coverage

Issue of
"**undercoverage**"

# Construct Intervals **Directly** from the **Exact Distribution**

Issue of "**overcoverage**" → results in **excessively wide intervals**, **diminishing precision**

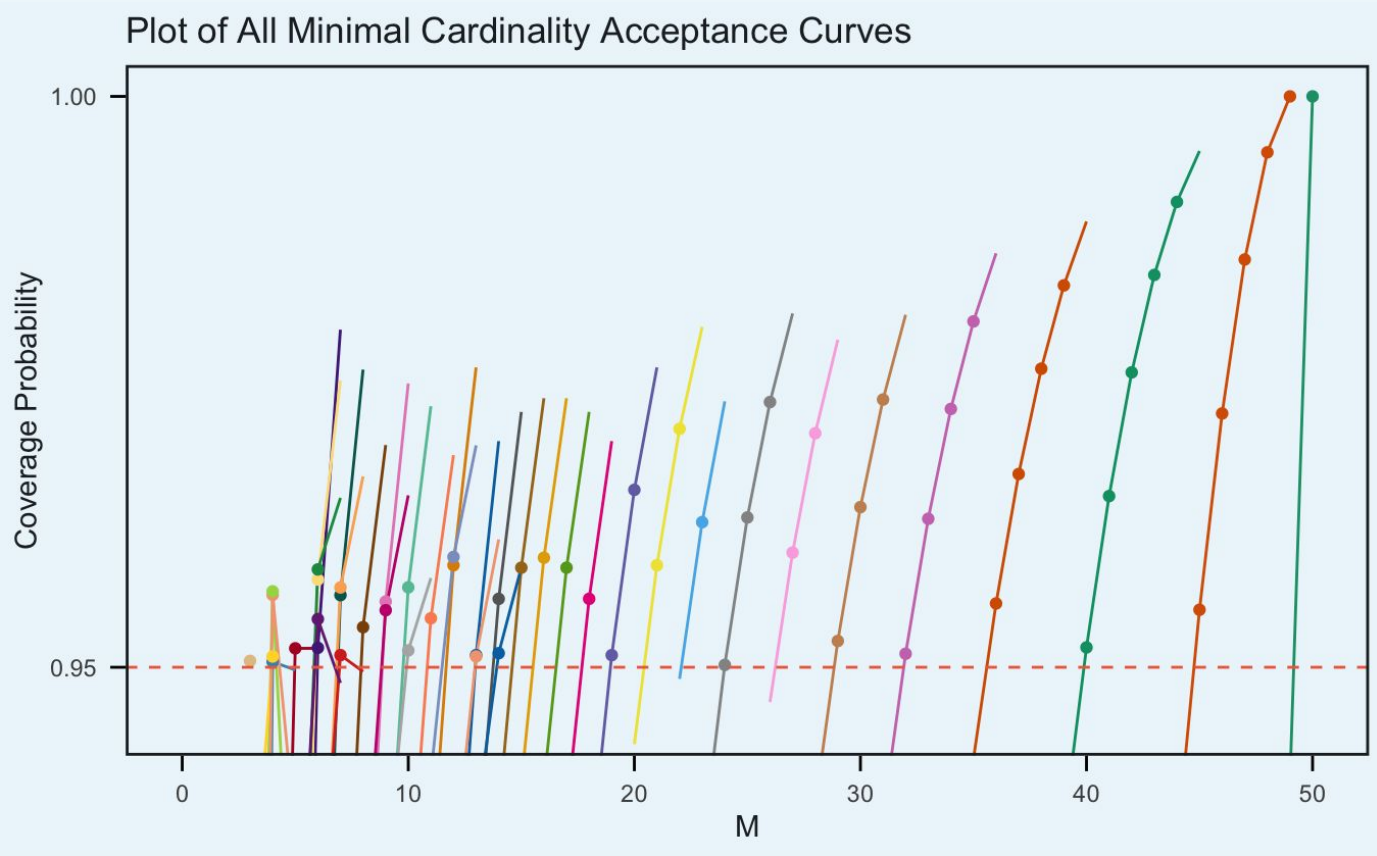


# Construct Intervals with **Coverage at or Above the Confidence Level** that are as **Narrow (Precise) as Possible**

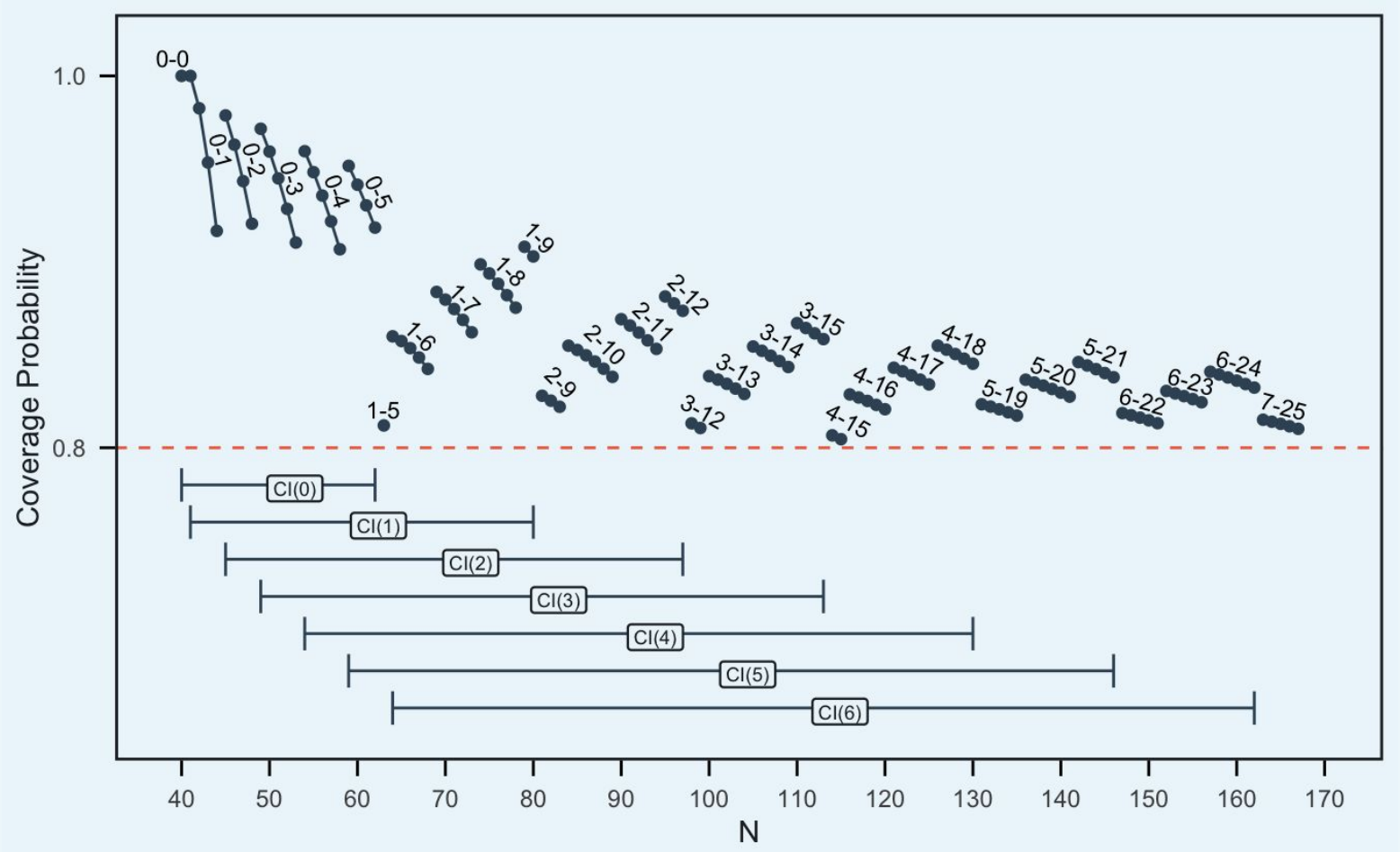
## Optimization Problem:

**Algorithm: reverse engineer the optimal CPF**

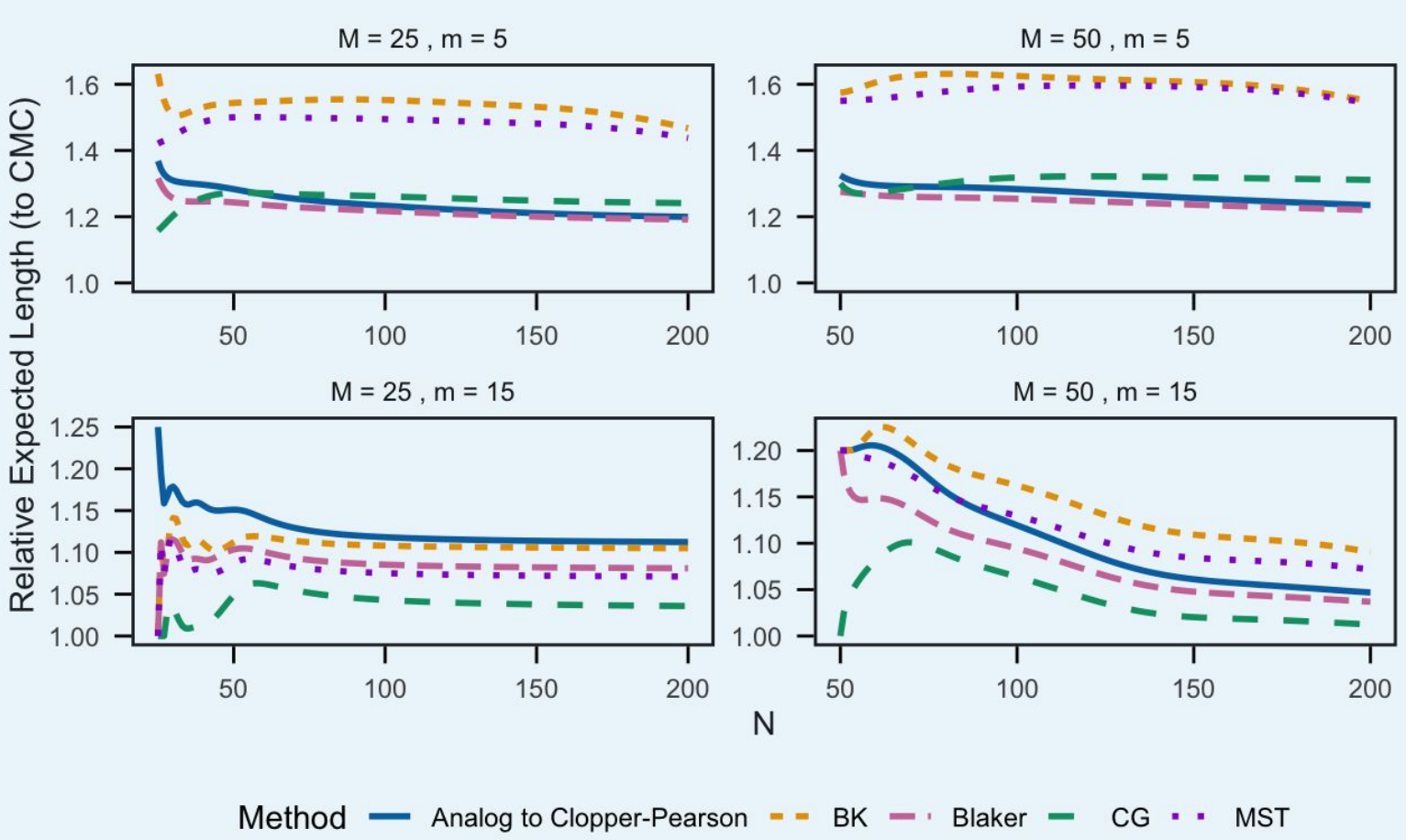
**Determine confidence intervals**



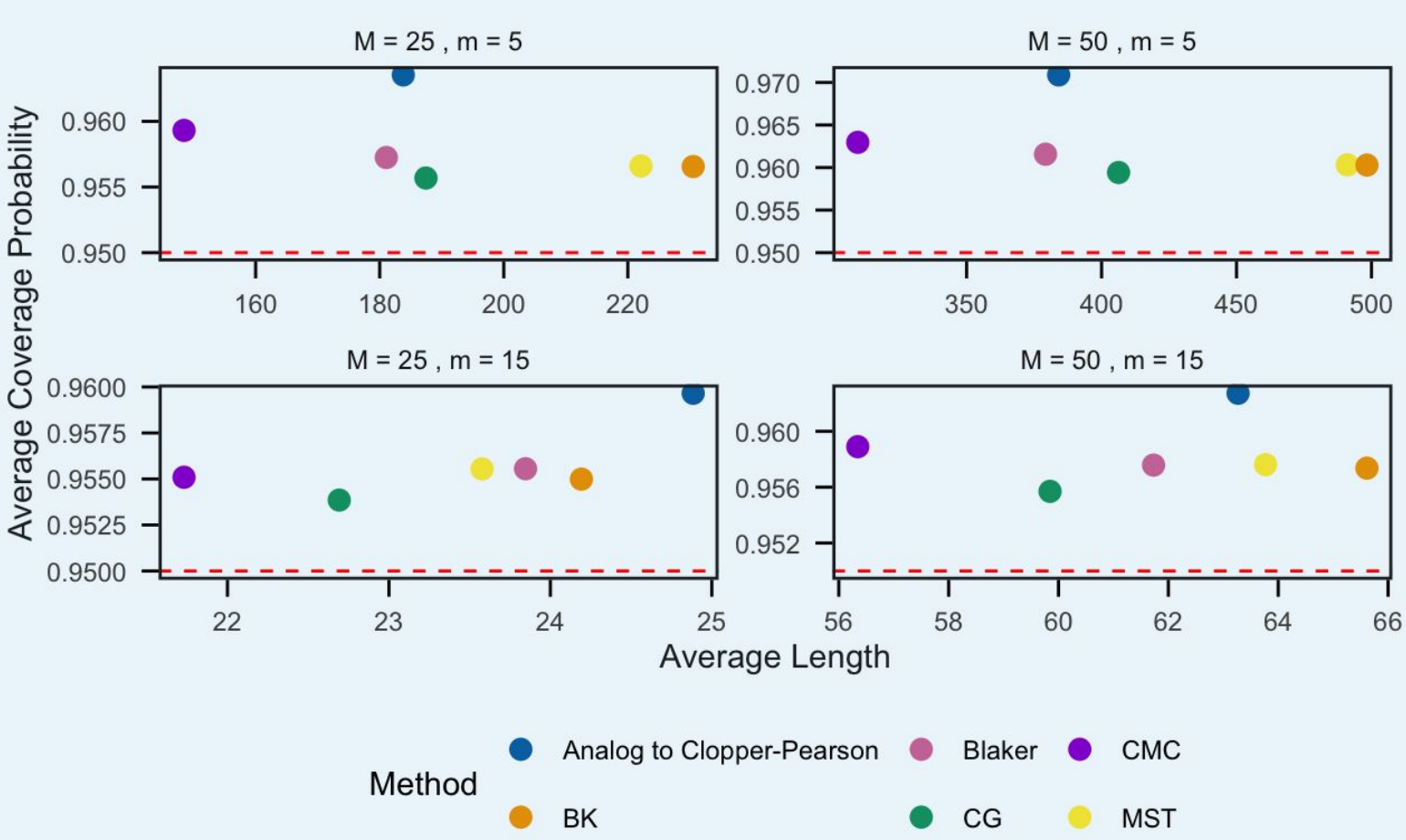
# CPF → Confidence Intervals



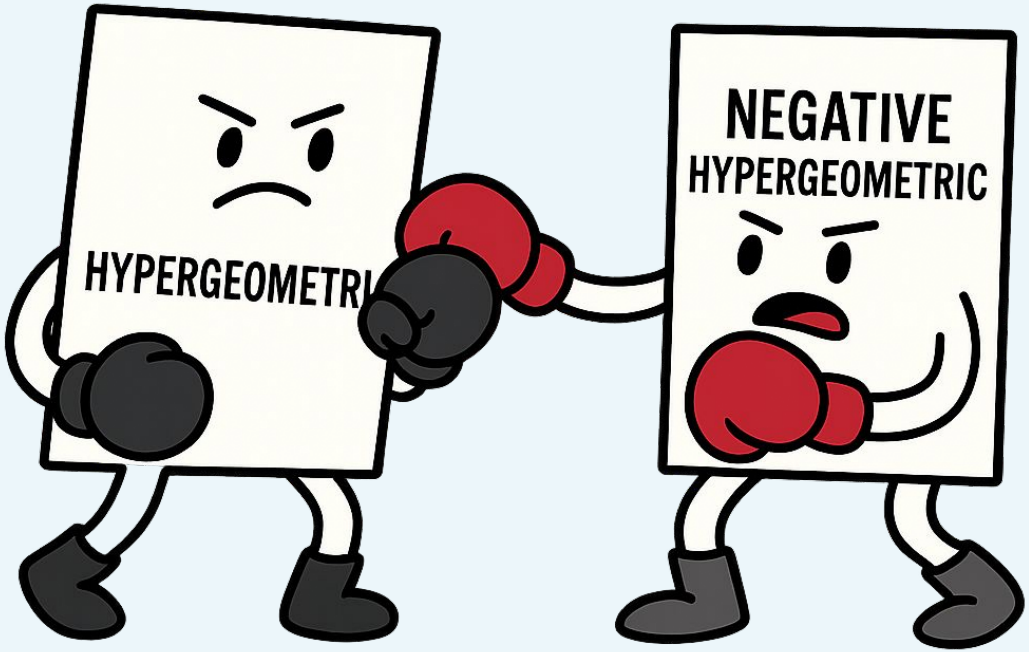
# Comparative Analysis: Relative Expected Length



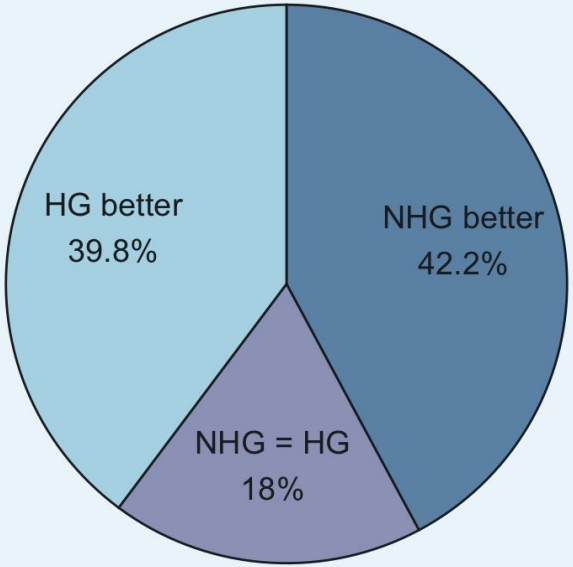
# Comparative Analysis: Average Length

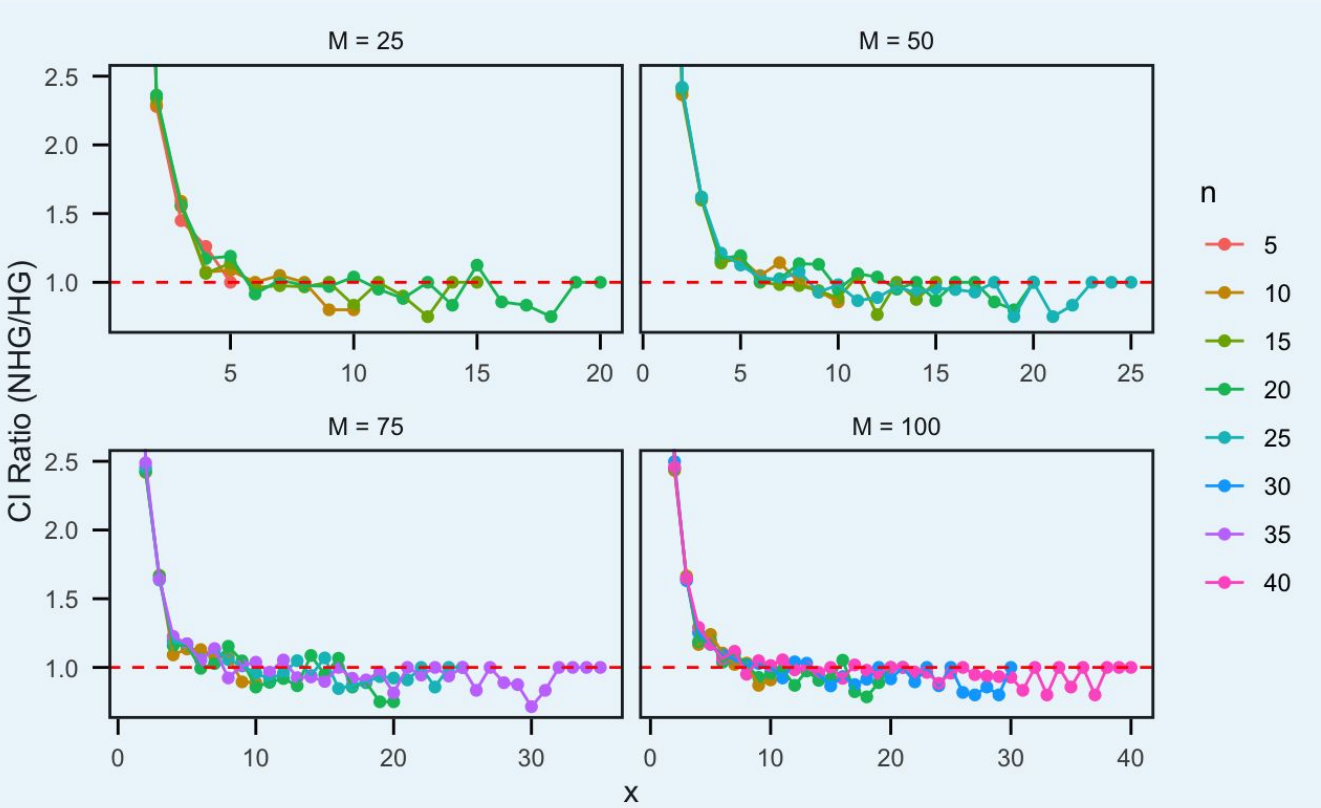


# Inverse Sampling versus Sampling with Fixed Sample Size

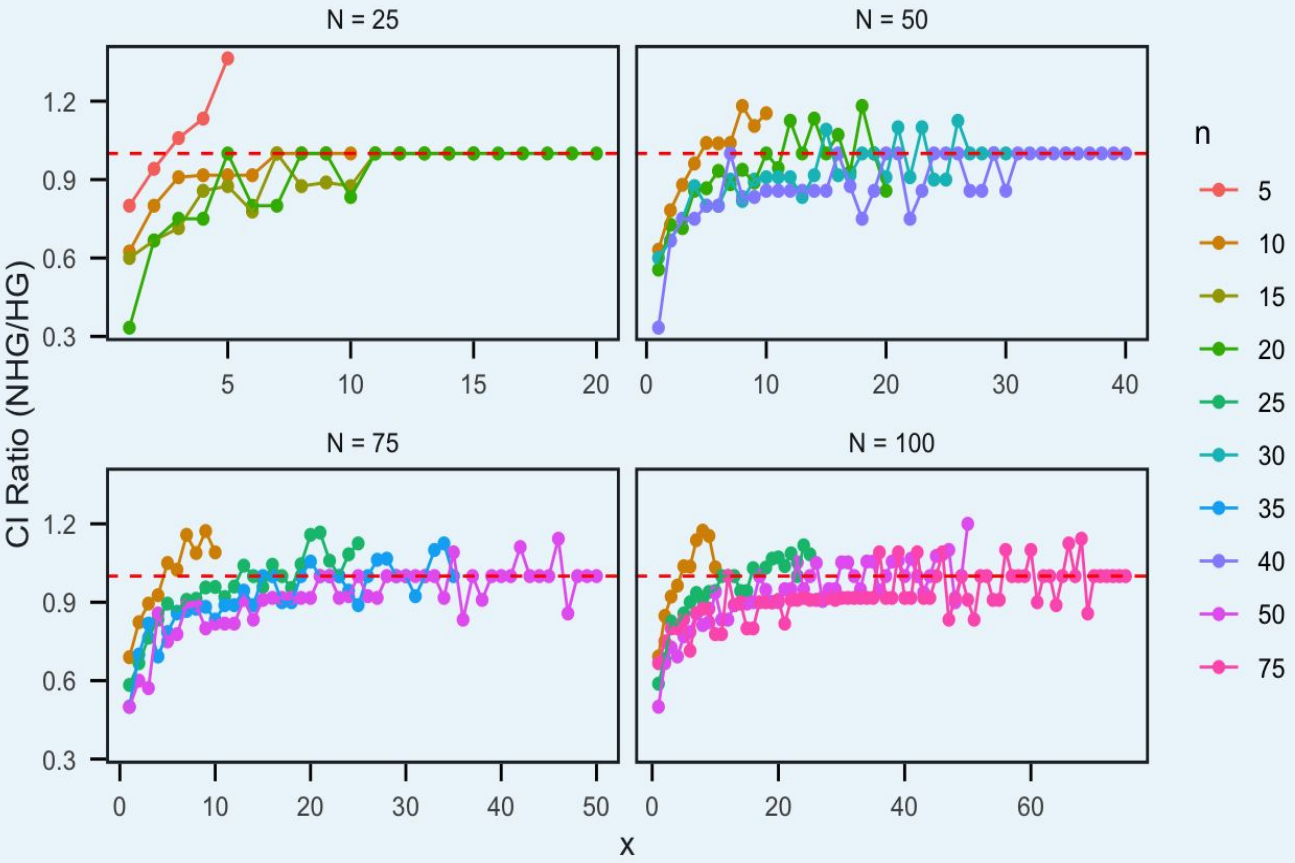


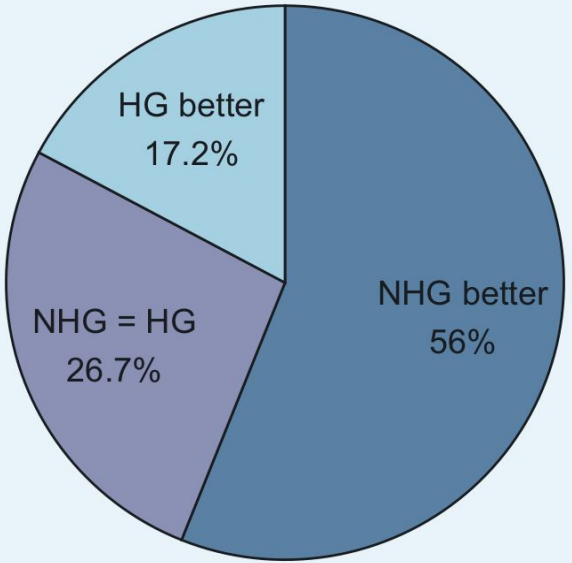
# Estimating Population Size (*N*)



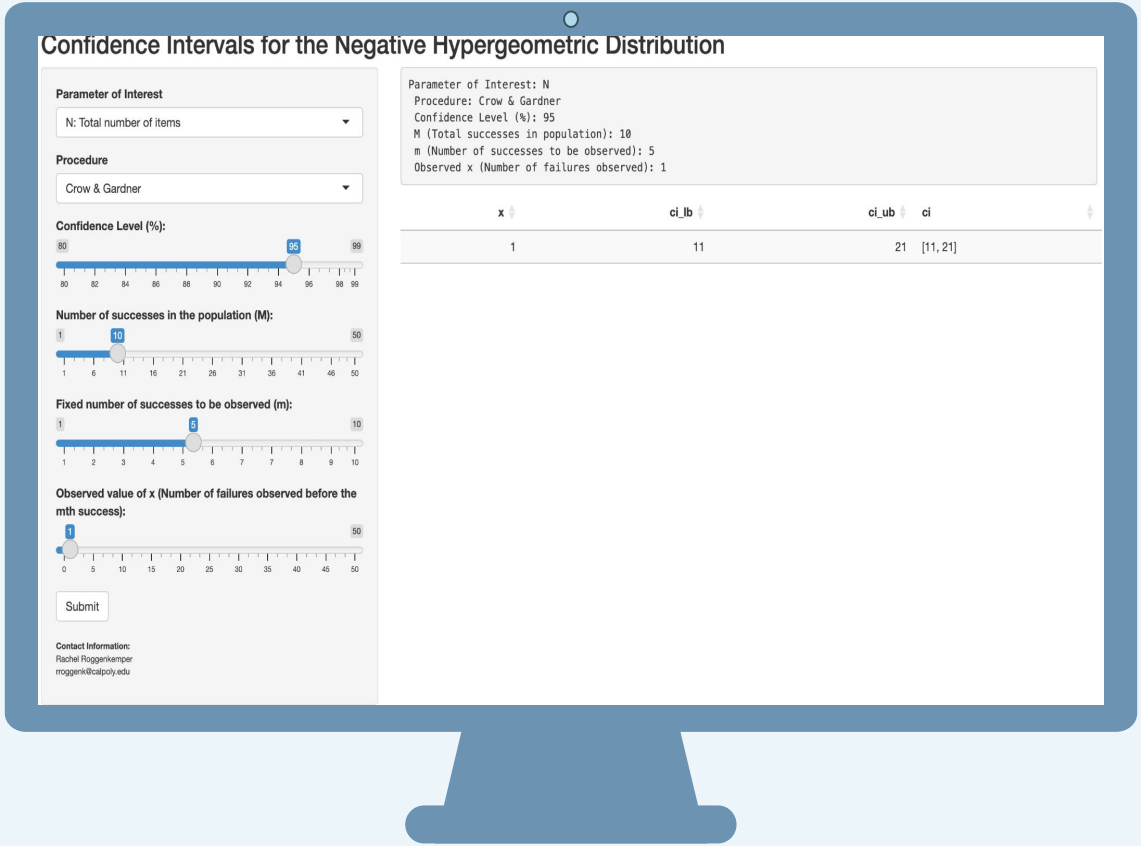



# Estimating Population Successes (*M*)

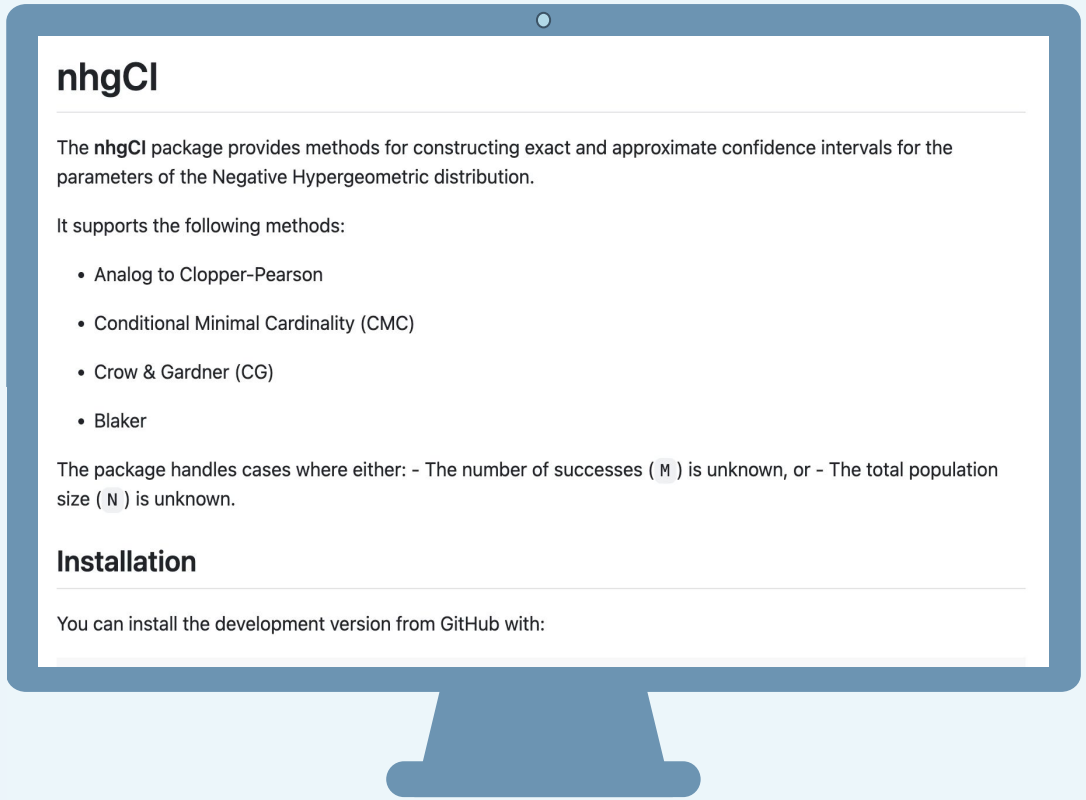


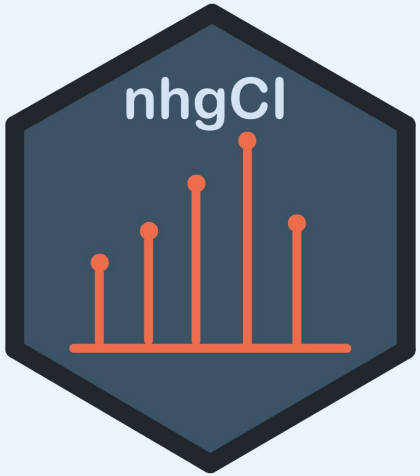



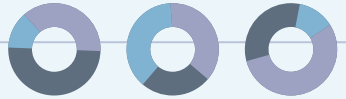
# Shiny App


Confidence Intervals for the Negative Hypergeometric Distribution
Parameter of Interest
N: Total number of items
Procedure
Crow & Gardner
Confidence Level (%):
Number of successes in the population (M):
Fixed number of successes to be observed (m):
Observed value of x (Number of failures observed before the mth success):
Submit
Contact Information:
Rachel Roggenkemper
rroggenk@calpoly.edu
Parameter of Interest: N
Procedure: Crow & Gardner
Confidence Level (%): 95
M (Total successes in population): 10
m (Number of successes to be observed): 5
Observed x (Number of failures observed): 1
x
ci_lb
ci_ub
ci
1
11
21
[11, 21]

# R Package


nhgCI
The nhgCI package provides methods for constructing exact and approximate confidence intervals for the parameters of the Negative Hypergeometric distribution.
It supports the following methods:
Analog to Clopper-Pearson
Conditional Minimal Cardinality (CMC)
Crow & Gardner (CG)
Blaker
The package handles cases where either: - The number of successes ( M ) is unknown, or - The total population size ( N ) is unknown.
Installation
You can install the development version from GitHub with:


nhgCI



# Thank you!

## Any questions?