# ****User Manual for Sample Size Determination Using Percentile-Based Confidence Intervals****

The script is designed for benthic research planning. Based on an existing dataset, it can predict how many samples need to be collected at a new site to achieve results with a specified precision level. This program analyzes how measurement precision (confidence interval width) changes with increasing sample size across three significance levels (70%, 90%, 95%). The primary output calculates the number of samples needed to achieve specified precision targets.

**❗ Important Method Limitation:**

This percentile-based method yields reliable estimates **only when the required number of sampling replicates does not exceed half of your original dataset's sampling effort**. For studies demanding more replicates than this threshold, the method will progressively **underestimate sampling requirements**.

### ****Key Algorithm Steps:****

1. Data loading from sample.csv
2. Generation of 1000 permuted datasets (pseudo-samples)
3. Calculation of cumulative means for each pseudo-sample
4. Determination of confidence intervals (70%, 90%, 95%) using percentiles
5. Normalization of interval widths relative to original sample mean
6. Hill equation regression modeling for sample size prediction
7. Visualization of results (graphs and tables)

## ****System Requirements****

* R (version 4.0+)
* Required R packages:
  + ggplot2 (visualization)
  + dplyr (data manipulation)
  + tidyr (data reshaping)

Install packages:

R

install.packages(c("ggplot2", "dplyr", "tidyr"))

## ****Usage Instructions****

### ****1. Prepare Input Data****

* Create sample.csv in your working directory
* Data format: Single column with header "x" (decimal point format)
* Example:

x

1.5

2.3

0.8

...

### ****2. Execute the Program****

* Save code as script.R
* Run in R:

R

source("script.R")

### ****3. Output Files****

| **File** | **Description** |
| --- | --- |
| Number\_of\_samples.txt | Sample size requirements for precision levels (20-60%) |
| hill\_models\_summary.txt | Hill equation parameters (tabular format) |
| hill\_models\_summary\_formatted.txt | Formatted model specifications |
| combined\_hill\_plot.png | Visualization of sample size vs. precision |

## ****Interpreting Results****

### ****1. Primary Output (****Number\_of\_samples.txt****)****

Example output:

Results (rounded to integers)

===================================

Date: 2024-03-15 12:30

70% 90% 95%

x=20 50 85 110

x=30 30 55 70

x=40 20 40 50

x=50 15 30 40

x=60 10 25 30

* **Rows**: Target precision (interval width as % of mean)
* **Columns**: Confidence levels
* **Values**: Required sample sizes

### ****2. Model Parameters (****hill\_models\_summary.txt****)****

Model top EC50 n R² AIC Equation

norm\_range\_70 150.2 45.3 1.2 0.98 1200 y = 150.2/(1+(45.3/x)^1.2)

* **top**: Maximum sample size (asymptote)
* **EC50**: Precision value at 50% of maximum
* **n**: Slope coefficient
* **R²**: Model fit (1 = perfect fit)

### ****3. Visualization (****combined\_hill\_plot.png****)****

* **X-axis**: Confidence interval width (%)
* **Y-axis**: Sample size
* **Lines**: Model predictions
* **Points**: Empirical data

## ****Customization Options****

### ****1. Adjusting Confidence Levels****

Modify probability thresholds in:

R

# For 70% CI:

percentil70 <- data.frame(

p15 = apply(... 0.15...),

p85 = apply(... 0.85...)

)

### ****2. Changing Precision Targets****

Edit prediction values:

R

x\_values <- c(15, 25, 35, 45, 55) # Custom precision levels

## ****Key Benefits****

1. **Precision Planning**: Determine optimal sample sizes for desired confidence levels
2. **Comparative Analysis**: Evaluate how different confidence levels affect requirements
3. **Visual Modeling**: Intuitive graphical representation of sample size-precision relationship