**User Manual for Confidence Interval Analysis Program Using Cox Method**

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. The Cox confidence interval calculation method assumes that the sample data follows a lognormal distribution.

Key points:

* Application - benthic (bottom-dwelling organism) studies
* Functionality - predicts required sample sizes for new locations
* Precision - allows determining needed accuracy levels
* Statistical assumption - data must be lognormally distributed for the Cox method

**Program Description**

This program analyzes the relationship between measurement precision (confidence interval width) and sample size for three significance levels (70%, 90%, 95%). The primary output is the calculation of required sample sizes to achieve specified precision levels.

**Requirements**

* R (version 4.0+)
* R packages:
  + ggplot2
  + dplyr

Install packages:

R

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

**Usage Instructions**

1. **Data Preparation**:
   * Create a sample.csv file in the working directory
   * Data should be in a single column with header "x"
   * Example format:

text

x

1.5

2.3

0.8

...

1. **Running the Program**:
   * the code to analysis\_script.R
   * Execute in R:

R

source("analysis\_script.R")

**Algorithm Overview**

1. **Data Loading**:
   * Reads sample.csv file
   * Calculates mean value
2. **Data Processing**:
   * Creates 1000 permutations of original data
   * Performs logarithmic transformation (with zero-value handling)
3. **Confidence Interval Calculation**:
   * Uses Cox method for interval calculation
   * Three significance levels: 70%, 90%, 95%
4. **Result Analysis**:
   * Calculates average interval widths
   * Performs power regression modeling
   * Visualizes results
5. **Output Generation**:
   * Regression equations (nls\_regression\_equations)
   * Sample size prediction table (predicted\_number\_of\_samples.txt)
   * Dependency plot (ggplot2)

**Output Interpretation**

**Key Output Files**

1. predicted\_number\_of\_samples.txt:
   * Table showing required sample sizes for target precision
   * Example:

text

Int.% 70% 90% 95%

20 50 85 110

30 30 55 70

...

* + Where:
    - "Int.%" - target precision (interval width as % of mean)
    - Other columns - required samples for each significance level

1. nls\_regression\_equations:
   * Power regression equations for each significance level
   * Example:

text

mean\_interval70: y = 120.5 \* x^-0.85 (R² = 0.98, AIC = 1500)

...

1. **Dependency Plot**:
   * Visualizes sample size vs. interval width relationship
   * Points represent actual data, lines show model predictions

**Parameter Customization**

1. **Changing Significance Levels**:
   * Modify z-values in calculate\_interval() calls:

R

interval70 <- calculate\_interval(log\_data, 1.04, data) # 70%

interval90 <- calculate\_interval(log\_data, 1.64, data) # 90%

interval95 <- calculate\_interval(log\_data, 1.96, data) # 95%

1. **Adjusting Precision Range**:
   * Change prediction values at script end:

R

x\_values <- c(20, 30, 40, 50, 60) # Target precision values

**Sample Output**

text

Regression Equations:

mean\_interval70: y = 120.5 \* x^-0.85 (R² = 0.98, AIC = 1500)

mean\_interval90: y = 200.3 \* x^-0.92 (R² = 0.97, AIC = 1600)

mean\_interval95: y = 250.1 \* x^-0.95 (R² = 0.96, AIC = 1700)

Predicted Sample Sizes:

Int.% 70% 90% 95%

20 50 85 110

30 30 55 70

40 20 40 50

50 15 30 40

60 10 25 30

**Conclusion**

This program enables users to:

1. Determine required sample sizes for target precision
2. Compare confidence interval behavior across significance levels
3. Visualize the precision-sample size relationship