README

SOFTWARE for SEER Cancer Data Analysis

Incidence-Based Mortality (IBM) Rate and Rate Ratio

This repository provides R functions to compute **Incidence-Based Mortality (IBM) rates** and **rate ratios**.
The methods include adjustments for small-count bias and variance estimation, making them suitable for rare cancers and small populations.

IBM Rate

The **IBM rate** links deaths to incident cancer cases in the registry.
Adjustments for small counts are available:

- **Fay–Feuer method** (recommended for rare events)
- **Tiwari's modification** (alternative adjustment)

Rates can be age-adjusted if needed.

Rate Ratios

Rate ratios compare IBM rates across groups (e.g., sex, race, calendar period).

- Variances are estimated with the **Delta method**, which approximates the standard error of the log rate ratio.
- Confidence intervals are computed on the log scale and then exponentiated.

Resources

- Age-adjusted Rate Confidence Intervals (SEER Documentation)
- **Rate Ratios**
  - Confidence interval formula:
    Fay MP. *Approximate confidence intervals for rate ratios from directly standardized rates with sparse data.*
    Communications in Statistics: Theory and Methods. 1999; 28(9):2141–2160.
  - P-value formula:
    Fay MP, Tiwari RC, Feuer EJ, Zou Z. *Estimating average annual percent change for disease rates without assuming constant change.*
    Biometrics. 2006; 62(3):847–854.
- R code

Example Usage

```r
# Load functions
source("IBM.R")

# Install package if needed
install.packages("readxl")

# Load the library
library(readxl)

# Read the Excel file
df <- read_excel("age_adjusted_data_grace.xlsx")


# Define index sets for comparison:
# Example: ER-negative, Non-Hispanic White vs Non-Hispanic Black, age 30–54
idx1 <- which(df$ER == "Negative" &
                df$Race == "Non-Hispanic White" &
                df$age_group_strata == "30 - 54")

idx2 <- which(df$ER == "Negative" &
                df$Race == "Non-Hispanic Black" &
                df$age_group_strata == "30 - 54")

# Compute DSRs and rate ratio with Fay–Feuer CI
compute_dsr_and_rr_for_subset(df, idx1, idx2,
                              "ER- & NHW & 30-54",
                              "ER- & NHB & 30-54",
                              "Subset",
                              ci_method = "fayfeuer")

# Compute DSRs and rate ratio with Tiwari CI
compute_dsr_and_rr_for_subset(df, idx1, idx2,
                              "ER- & NHW & 30-54",
                              "ER- & NHB & 30-54",
                              "Subset",
                              ci_method = "tiwari")



################################################
# Output interpretation
################################################

# The function returns a list with two elements:
#
# 1) $rates : Group-specific age-adjusted incidence-based mortality (IBM) rates
#    - DSR (Directly Standardized Rate) is computed per unit population,
#      then scaled to "Rate_per1e5" (per 100,000 people).
#    - CI_low and CI_high give the 95% confidence interval for the rate.
#
# Interpretation of Rates:
#   - The "Rate_per1e5" column represents the IBM rate per 100,000 people,
#     adjusted for age using standard weights.
#   - Higher rates indicate a greater burden of mortality linked to cancer incidence.
#   - The confidence interval (CI_low, CI_high) reflects statistical uncertainty.
```