Package ‘uncertainUCDP’

January 23, 2026

Type Package

Title Parametric Mixture Models for Uncertainty Estimation of Fatalities in UCDP Conflict Data

Version 0.7.6

Description Provides functions for estimating uncertainty in the number of fatalities in the Uppsala Conflict Data Program (UCDP) data. The package implements a parametric reported-value Gumbel mixture distribution that accounts for the uncertainty in the number of fatalities in the UCDP data. The model is based on information from a survey on UCDP coders and how they view the uncertainty of the number of fatalities from UCDP events. The package provides functions for making random draws of fatalities from the mixture distribution, as well as to estimate percentiles, quantiles, means, and other statistics of the distribution. Full details on the survey and estimation procedure can be found in Vesco et al (2024).

License GPL (>= 3)

Encoding UTF-8

LazyData true

Imports dplyr, lmtest, sandwich, foreach, rlang, tibble

Depends R (>= 3.5)

RoxygenNote 7.3.3

NeedsCompilation no

Author David Randahl [cre, aut]

Maintainer David Randahl <david.randahl@pcr.uu.se>

Repository CRAN

Date/Publication 2026-01-23 14:40:02 UTC

Contents

mean_uncertainUCDP *Mean, median, and quantiles of the parametric uncertainty distributions for UCDP events*

Description

Mean, median, and quantiles of the parametric uncertainty distributions for UCDP events. The parametric uncertainty distributions are based on the reported-value inflation Gumbel mixture distribution. The `median` and `quantile` functions are shortcuts for the `quncertainUCDP` function.

Usage

```
mean_uncertainUCDP(fatalities, tov = c("sb", "ns", "os", "any"))

median_uncertainUCDP(fatalities, tov = c("sb", "ns", "os", "any"))

quantile_uncertainUCDP(probs, fatalities, tov = c("sb", "ns", "os", "any"))
```

Arguments

`fatalities`	A vector of non-negative integers representing the number of fatalities of the UCDP events. Non-integer values are allowed but should be considered experimental.
`tov`	A character string representing the type of violence of the UCDP. Must be one of "sb", "ns", "os", or "any". The options are: * "sb" for state-based violence * "ns" for non-state violence * "os" for one-sided violence * "any" for parameters estimated across all type of violence. This is somewhat experimental and should be used with caution. This is possibly useful when the type of violence is unknown or when the user wants to combine all types of violence into a single category.
`probs`	A numeric vector of probabilities with values in [0,1]. The quantiles to calculate.

Value

A numeric vector of the same length as the input vector of fatalities representing the means, medians, and quantiles of the parametric uncertainty distribution for each UCDP event.

Examples

```
data(ucdpged)

# Calculate the mean for an arbitrary UCDP event
mean_uncertainUCDP(fatalities = 100, tov = 'sb')

# Calculate the mean for the first event in the UCDP GED sample
mean_uncertainUCDP(ucdpged$best[1], tov = ucdpged$type_of_violence[1])
```

```
# Calculate the median for an arbitrary UCDP event
median_uncertainUCDP(fatalities = 100, tov = 'sb')

# Calculate the median for the first event in the UCDP GED sample
median_uncertainUCDP(ucdpged$best[1], tov = ucdpged$type_of_violence[1])

# Calculate the 90th percentile for an arbitrary UCDP event
quantile_uncertainUCDP(probs = 0.9, fatalities = 100, tov = 'sb')

# Calculate the 90th percentile for the first event in the UCDP GED sample
quantile_uncertainUCDP(ucdpged$best[1], 0.9, tov = ucdpged$type_of_violence[1])
```

`models_rubin_rules` *Pool regression coefficients and standard errors via Rubin’s rules*

Description

Pools coefficient estimates and standard errors from multiple fitted models (typically one per dataset) using Rubin’s rules, returning a ‘coeftest‘-like matrix with pooled estimates, SEs, t-statistics, and p-values. Can be used when when pooling mulitple models with different draws from the UCDP uncertainty model.

Usage

```
models_rubin_rules(list, df = NULL, cluster = NULL)
```

Arguments

`list` A list of fitted model objects (e.g., ‘lm‘, ‘glm‘, etc.) for which ‘lmtest::coeftest()‘ methods exist.

`df` Optional. Degrees of freedom for t-distribution when computing p-values. If ‘NULL‘, uses the degrees of freedom from the first model’s ‘coeftest‘ output, i.e. the df from the original model.

`cluster` Optional. A clustering variable passed to ‘sandwich::vcovCL()‘. Typically a one-sided formula specficying clusters, e.g., ‘~ cluster_id‘

Details

If ‘cluster‘ is provided, cluster-robust covariance matrices are computed for each model via ‘sandwich::vcovCL()‘ and passed to ‘lmtest::coeftest()‘.

The pooling follows the usual within/between decomposition: within-imputation variance is the mean of squared SEs, and between-imputation variance is the sample variance of point estimates across imputations. The total variance is $W + (1 + 1/m)B$ where m is the number of models.

Value

A numeric matrix with the same layout as ‘lmtest::coeftest()‘, containing pooled coefficients, pooled standard errors, t-statistics, and two-sided p-values. The returned object uses the first element’s ‘coeftest‘ matrix as a template (including row/column names and attributes).

Examples

```
library(foreach)
library(dplyr)
# Draw 100 datasets from UCDP uncertainty model and return as a list
data('ucdpged')
ucdp_sb <- ucdpged %>% filter(type_of_violence == 1)
list_of_models <- foreach(i = 1:100) %do%{
 lm(ged_sb_draw ~ year + region, data =
ucdp_sb %>% mutate(ged_sb_draw = runcertainUCDP(n = n(), fatalities = best, tov = 'sb')))
}
#' # Pool model coefficients and standard errors via Rubin's rules
pooled_results <- models_rubin_rules(list_of_models)
```

rubin_b_and_ses *Pool coefficients and standard errors from multiple ‘coeftest‘ objects*

Description

Internal helper that takes a list of ‘lmtest::coeftest()‘ matrices and returns a pooled ‘coeftest‘-like matrix using Rubin’s rules.

Usage

```
rubin_b_and_ses(coeftests)
```

Arguments

coeftests	A list of objects returned by ‘lmtest::coeftest()‘. Each element must have matching coefficient rows/columns, where column 1 is the estimate and column 2 is the standard error.

Value

A ‘coeftest‘-like numeric matrix with pooled estimates (col 1), pooled standard errors (col 2), pooled t-statistics (col 3), and two-sided p-values (col 4). Row/column names and attributes are inherited from the first element of ‘coeftests‘.

runcertainUCDP *Parametric uncertainty distributions for UCDP events*

Description

Density, distribution, quantile and random number generation functions for the parametric reported-value inflated Gumbel mixture distribution for UCDP events. The functions estimate the parameters of the distribution based on the number of fatalities and the type of violence of the UCDP event.

Usage

```
runcertainUCDP(n, fatalities, tov = c("sb", "ns", "os", "any"))

puncertainUCDP(q, fatalities, tov = c("sb", "ns", "os", "any"))

duncertainUCDP(x, fatalities, tov = c("sb", "ns", "os", "any"))

quncertainUCDP(p, fatalities, tov = c("sb", "ns", "os", "any"))
```

Arguments

`n`	Number of observations to generate random values for
`fatalities`	A vector of non-negative integers representing the number of fatalities of the UCDP events. Non-integer values are allowed but should be considered experimental.
`tov`	A character string representing the type of violence of the UCDP. Must be one of "sb", "ns", "os", or "any". The options are: * "sb" for state-based violence * "ns" for non-state violence * "os" for one-sided violence * "any" for parameters estimated across all type of violence. This is somewhat experimental and should be used with caution. This is possibly useful when the type of violence is unknown or when the user wants to combine all types of violence into a single category.
`x, q`	Vector of quantiles
`p`	Vector of probabilities

Details

The reported-value inflated Gumbel mixture distribution is a parametric distribution for modeling the uncertainty in the number of fatalities of UCDP events. The distribution is a mixture of a Gumbel distribution and a point mass at the reported number of fatalities. The distribution is estimated based on the number of fatalities and the type of violence of the UCDP event. The distribution is estimated using a set of regression models that estimate the location, scale, and weight parameters of the distribution based on the number of fatalities and the type of violence of the UCDP event.

Value

* duncertainUCDP gives the density function * puncertainUCDP gives the distribution function * quncertainUCDP gives the quantile function * runcertainUCDP generates random values as a vector of length n

Examples

```
data(ucdpged)

# Generate 10 random values for an arbitrary UCDP event
runcertainUCDP(n = 10, fatalities = 100, tov = 'sb')

# Generate 10 random values for the first event in the GED sample
runcertainUCDP(n = 10, fatalities = ucdpged$best[1], tov = ucdpged$type_of_violence[1])

# Obtaining the probability that an arbitrary UCDP event has at least 150 fatalities
puncertainUCDP(q = 150, fatalities = 100, tov = 'ns')

# Obtaining the probability that the for the first event in the GED sample has at least 5 fatalities
puncertainUCDP(q = 5, fatalities = ucdpged$best[1], tov = ucdpged$type_of_violence[1])

# Obtaining the 90th percentile for an arbitrary UCDP event and one-sided violence
quncertainUCDP(p = 0.9, fatalities = 100, tov = 'os')

# Obtaining the 90th percentile for the first event in the GED sample
quncertainUCDP(p = 0.9, fatalities = ucdpged$best[1], tov = ucdpged$type_of_violence[1])

# Obtaining the density for an arbitrary UCDP event and state-based violence
duncertainUCDP(x = seq(from = 0, to = 500), fatalities = 100, tov = 'sb')

# Obtaining the density for the first event in the GED sample
duncertainUCDP(x = seq(0, 50), fatalities = ucdpged$best[1], tov = ucdpged$type_of_violence[1])
```

---

ucdpged *UCDP Georeferenced Event Dataset (GED) sample*

Description

A sample of the UCDP Georeferenced Event Dataset (GED) from the 2023 data release. The data contains information about the date, location, and type of conflict events. The data is a sample of the full dataset, which can be downloaded from the UCDP website <https://ucdp.uu.se/downloads/>.

Usage

```
ucdpged
```

Format

a tibble with 1000 rows and 49 columns

Source

<https://ucdp.uu.se/downloads/>

`uncertainUCDP_parameters`

Parameter extraction for uncertainUCDP-functions

Description

Extracting parameters for the reported-value inflated Gumbel mixture distribution for UCDP events. Primarily intended for internal use by the uncertainUCDP-functions, but can be used to extract parameters for the distribution manually.

Usage

```
uncertainUCDP_parameters(fatalities, tov)
```

Arguments

`fatalities`	A vector of non-negative integers representing the number of fatalities of the UCDP event. Non-integer values are allowed but should be considered experimental
`tov`	A character string or integer value representing the type of violence of the UCDP. Must be one of "sb", "ns", "os", or "any" or their numeric equivalent The options are: * "sb" or 1 for state-based violence * "ns" or 2 for non-state violence * "os" or 3 for one-sided violence * "any" or 4 for parameters estimated across all type of violence. This is somewhat experimental and should be used with caution. This is possibly useful when the type of violence is unknown or when the user wants to combine all types of violence into a single category.

Value

A list with three elements: loc, scale, and w. loc and scale are the location and scale parameters of the Gumbel distribution, respectively. w is the weight parameter for the reported-value inflation

Index