Package: uwedragon (via r-universe)

August 24, 2024

Type Package

Title Data Research, Access, Governance Network : Statistical Disclosure Control

Version 0.1.0

Author Ben Derrick

Maintainer Ben Derrick <ben.derrick@uwe.ac.uk>

Description A tool for checking how much information is disclosed when reporting summary statistics.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.2.9000

Imports gtools

NeedsCompilation no

Repository CRAN

Date/Publication 2022-03-02 19:40:01 UTC

Contents

`disguise` *Disguise the sample mean and sample deviation*

Description

Disguises the sample mean and standard deviation via a choice of methods.

Usage

```
disguise(usersample, method = 2)
```

Arguments

usersample	A vector of all individual sample values.
method	Approach for disguising mean and standard deviation. (default = 1)

Details

Method 1

Randomly split the sample into two (approx. equal size) samples A, and B. For sample A calculate and report mean. For sample B calculate and standard deviation.

Method 2 (default)

Take a sample of size N with replacement; calculate and report mean. Repeat to calculate and report standard deviation.

Method 3

Generate a random number (RN1) between N/2 and N. Sample with replacement a sample size of RN1; calculate and report mean. Generate a random number (RN2) between N/2 and N. Sample with replacement a sample size of RN2; calculate and report standard deviation.

Method 4

As Method 3, but sampling without replacement.

Value

Outputs disguised mean and disguised standard deviation.

References

Derrick, B., Green, L., Kember, K., Ritchie, F. & White P, 2022, Safety in numbers: Minimum thresholding, Maximum bounds, and Little White Lies. Scottish Economic Society Annual Conference, University of Glasgow, 25th-27th April 2022

Examples

```
usersample<-c(1,1,2,3,4,4,5)

disguise(usersample,method=1)
disguise(usersample,method=2)
disguise(usersample,method=3)
disguise(usersample,method=4)
```

SDCdragon — *Statistical Data Control. Data Research, Access, Governance Network.*

Description

A tool for checking how much information is disclosed when reporting summary statistics

solutions — *Find individual sample values from the sample mean and standard deviation*

Description

For integer based scales, finds possible solutions for each value within a sample. This is revealed upon providing sample size, minimum possible value, maximum possible value, mean, standard deviation (and optionally median).

Usage

```
solutions(
  n,
  min_poss,
  max_poss,
  usermean,
  usersd,
  meandp = NULL,
  sddp = NULL,
  usermed = NULL
)
```

Arguments

n	Sample size.
min_poss	Minimum possible value. If sample minimum is disclosed, this can be inserted here, otherwise use the theoretical minimum. If there is no theoretical maximum 'Inf' can be inserted.
max_poss	Maximum possible value. If sample maximum is disclosed, this can be inserted here, otherwise use the theoretical maximum. If there is no theoretical minimum '-Inf' can be inserted.
usermean	Sample mean.
usersd	Sample standard deviation, i.e. n-1 denominator.
meandp	(optional, default=NULL) Number of decimal places mean is reported to, only required if including trailing zeroes.

sddp	(optional, default=NULL) Number of decimal places standard deviation is reported to, only required if including trailing zeroes.
usermed	(optional, default=NULL) Sample median.

Details

For use with data measured on a scale with 1 unit increments. Samuelson's inequality [1] used to further restrict the minimum and maximum. All possible combinations within this inequality are calculated [2] for factorial(n+k-1)/(factorial(k)*factorial(n-1))<65,000,000.

No restriction on number of decimal places input. Reporting less than two decimal places will reduce the chances of unique solution to all sample values being uncovered [3]

Additional options to specify number of digits following the decimal place that are reported, required for trailing zeroes.

Value

Outputs possible combinations of original integer sample values.

References

[1] Samuelson, P.A, 1968, How deviant can you be? Journal of the American Statistical Association, Vol 63, 1522-1525.

[2] Allenby, R.B. and Slomson, A., 2010. How to count: An introduction to combinatorics. Chapman and Hall/CRC.

[3] Derrick, B., Green, L., Kember, K., Ritchie, F. & White P, 2022, Safety in numbers: Minimum thresholding, Maximum bounds, and Little White Lies. Scottish Economic Society Annual Conference, University of Glasgow, 25th-27th April 2022

Examples

```
# EXAMPLE 1
# Seven observations are taken from a five-point Likert scale (coded 1 to 5).
# The reported mean is 2.857 and the reported standard deviation is 1.574.

solutions(7,1,5,2.857,1.574)

# For this mean and standard deviation there are two possible distributions:
# 1  1  2  3  4  4  5
# 1  2  2  2  3  5  5

# Optionally adding median value of 3.

solutions(7,1,5,2.857,1.574, usermed=3)

# uniquely reveals the raw sample values:
# 1  1  2  3  4  4  5


# EXAMPLE 2
```

```
# The mean is '4.00'.
# The standard deviation is '2.00'.
# Narrower set of solutions found specifying 2dp including trailing zeroes.

solutions(3,-Inf,Inf,4.00,2.00,2,2)

# uniquely reveals the raw sample values:
# 2  4  6
```

Index