Requirements for the Module statistics_lib.data_classes

Conventions

Requirements listed in this document are constructed according to the following structure:

Requirement ID: REQ-UVW-XYZ

Title: Title / name of the requirement

Description: Description / definition of the requirement

Verification Method: I / A / T / D

The requirement ID starts with the fixed prefix 'REQ'. The prefix is followed by 3 letters abbreviation (in here 'UVW'), which defines the requiement type - e.g. 'FUN' for a functional and capability requirement, 'AWM' for an alarm, warnings and operator messages, etc. The last part of the ID is a 3-digits *hexadecimal* number (0..9|A..F), with the first digit identifing the module, the second digit identifing a class / function, and the last digit - the requirement ordering number for this object. E.g. 'REQ-FUN-112'. Each requirement type has its own counter, thus 'REQ-FUN-112' and 'REQ-AWN-112' requirements are different entities, but they refer to the same object (class or function) within the same module.

The verification method for a requirement is given by a single letter according to the table below:

Term	Definition
Inspection (I)	Control or visual verification
Analysis (A)	Verification based upon analytical evidences
Test (T)	Verification of quantitative characteristics with quantitative measurement
Demonstration (D)	Verification of operational characteristics without quantitative measurement

Functional and capability requirements

Requirement ID: REQ-FUN-300

Title: Functionality implemented by the module (scope)

Description: The module should implement 2 classes: for storing 1D data set, and for storing 2D data set (paired 1D data sets of the same length), which should be stored / treated as an immutable data type. These classes should provide attributes or properties or methods to obtain the statistical properties of the data, sufficient to perform the basic descriptive statistical analysis.

Verification Method: A

Requirement ID: REQ-FUN-310

Title: Instantiation of 1D statistics class

Description: The 1D statistics class' instantiation method should require and accept a single positional argument, which can be any flat sequence (e.g. list or tuple) of real numbers (int or float) or instances of a data type class imlementing a 'real life measurement' with the associated uncertainty of the measurement. Any real number in such sequence should be treated to have zero measurement uncertainty. Data sanity check and conversion (if required) should be applied during instantiation, which should quarantee that no errors / exceptions could occur when requesting any statistical property of this data sample, unless a specific method requires additional arguments.

Verification Method: T

Requirement ID: REQ-FUN-311

Title: 1D statistics class - access to the stored data

Description: The 1D statistics class should provide access to the stored 'mean / most probable' values of the measured data as well as to the associated measurements errors via two separate read-only attributes, returning the respective data as immutable sequences.

Verification Method: T

Requirement ID: REQ-FUN-312

Title: 1D statistics class - functionality ignoring the measurement uncertainties

Description: The 1D statistics class should provide the following properites:

- *N* - int > 0, length of the data set (number of data points)
- *Mean* - int OR float, aritmetic mean
- *Var* - int OR float, variance
- *Sigma* - int OR float, standard deviation
- *SE* - int OR float, standard error of the mean
- *Skew* - int OR float, skewness of the distribution
- *Kurt* - int OR float, excess kurtosis of distribution
- *Median* - int OR float, median of the distibution
- *Q1* - int OR float, the first quartile of the distribution
- *Q3* - int OR float, the third quartile of the distribution
- *Min* - int OR float, the minimum value within the distribution
- *Max* - int OR float, the maximum value within the distribution

The Bessel correction should not be applied, the data sample should be treated as the whole population.

In addition, the following methods should be implemented:

- *getQuantile*(k, m) - 0<= int k <= int m -> int OR float, generic k-th of m-quantile (m > 0)
- *getHistogram*(, NBins = None, BinSize = None) - /, int > 0 OR None, int > 0 OR float > 0 OR None/ -> tuple(tuple(int OR float, int >= 0))

Verification Method: T

Requirement ID: REQ-FUN-313

Title: 1D statistics class - functionality acccounting for the measurement uncertainties

Description: The 1D statistics class should provide the following properites:

- *FullVar* - int OR float, variance, including the contribution of the measurements uncertainties
- *FullSigma* - int OR float, standard deviation, including the contribution of the measurements uncertainties
- *FullSE* - int OR float, standard error of the mean, including the contribution of the measurements uncertainties

The Bessel correction should not be applied, the data sample should be treated as the whole population.

Verification Method: T

Requirement ID: REQ-FUN-314

Title: 1D statistics class - (optional) name of the data set

Description: The 1D statistics class should provide a getter + setter (full access) property *Name* to set and read-out an arbitrary string identificator (name) of the data set, with the default value being **None**.

Verification Method: T

Requirement ID: REQ-FUN-315

Title: 1D statistics class - full summary on the statistical properties of the distribution.

Description: The 1D statistics class should provide a read-only property *Summary* returning the following data as a multi-line string with TSV format tabulated report, with each line constructed in the form '{name of property}: TAB {value of property}'. The properties to be included into the report are:

- Name (unless **None**, in which case not to be included)
- N
- Mean
- Median
- Q1
- Q2
- Min
- Max
- Var
- FullVar
- Skew
- Kurt

The report may be encapsulated into visual separator lines (the first and the last).

Verification Method: D

Requirement ID: REQ-FUN-320

Title: Instantiation of 2D statistics class

Description: The 2D statistics class' instantiation method should require and accept two positional arguments, each of which can be any flat sequence (e.g. list or tuple) of real numbers (int or float) or instances of a data type class imlementing a 'real life measurement' with the associated uncertainty of the measurement. Any real number in such sequence should be treated to have zero measurement uncertainty. The both sequences must be of the same length. Data sanity check and conversion (if required) should be applied during instantiation, which should guarantee that no errors / exceptions could occur when requesting any statistical property of this data sample, unless a specific method requires additional arguments.

Verification Method: T

Requirement ID: REQ-FUN-321

Title: 2D statistics class - access to the stored data

Description: The 2D statistics class should provide access to the stored X and Y data sets via two separate read-only attributes, returning the instances of 1D statistics class.

Verification Method: T

Requirement ID: REQ-FUN-322

Title: 2D statistics class - functionality

Description: The 2D statistics class should provide the following properites:

- *N* - int > 0, length of the 2D data set (number of paired X-Y data points)
- *Cov* - int OR float, covariance
- *Pearson* - int OR float, Pearson's correlation coefficient
- *Spearman* - int OR float, Spearman rank correlation coefficient
- *Kendall* - int OR float, Kendall τ-b rank correlation coefficient

The measurement uncertainties should not be taken into account

Verification Method: T

Requirement ID: REQ-FUN-323

Title: 2D statistics class - (optional) name of the data set

Description: The 2D statistics class should provide a getter + setter (full access) property *Name* to set and read-out an arbitrary string identificator (name) of the data set, with the default value being **None**.

Verification Method: T

Requirement ID: REQ-FUN-324

Title: 2D statistics class - full summary on the statistical properties of the distribution.

Description: The 2D statistics class should provide a read-only property *Summary* returning the following data as a multi-line string with TSV format tabulated report, with each line constructed in the form '{name of property}: TAB {value of property}'. The properties to be included into the report are:

- Name (unless **None**, in which case not to be included)
- Cov
- Pearson
- Spearman
- Kendall

These lines should be followed by the respective reports on 1D sub-sets with a clear indication on which a *X* and *Y* set.

The report may be encapsulated into visual separator lines (the first and the last).

Verification Method: D

Alarms, warnings and operator messages

Requirement ID: REQ-AWM-300

Title: Instantiation - unacceptable type of the input data

Description: The **TypeError** or its sub-class should be raised in response to the unacceptable input (argument(s) data type(s)) of the instantiation method of 1D statistics class or one of arguments of the 2D data set, which is not a flat sequence of real numbers of 'real life measurements' with the associated uncertainties, meaning:

- The respective argument is not a sequence, OR
- The respective argument is a sequence, but, at least, one element of it is neither integer, or floating point number, or an instance of a measurement with the associated uncertainty data type class

Verification Method: T

Requirement ID: REQ-AWM-301

Title: Instantiation - unacceptable value of the input data

Description: The **ValueError** or its sub-class should be raised in response to the unacceptable input value of the proper data type of the instantiation method:

- 1D statistics class
    - The argument is an empty sequence
- 2D statistics class
    - At least, one of the arguments is an empty sequence
    - The lengths of the arguments (sequences) are not equal

Verification Method: T

Requirement ID: REQ-AWM-302

Title: Read-only attribute access - AttributeError

Description: The **AttributeError** or its sub-class should be raised in response to

- Attempt to assign a value to a read-only attribute of 1D or 2D statistics classes
- Attempt to delete a read-only attribute of 1D or 2D statistics classes

Verification Method: T

Requirement ID: REQ-AWM-310

Title: Protection of data points

Description: The **TypeError** or its sub-class should be raised in response to

- Attempt to modify a value of the 'mean' or measurement uncertainty of a data point stored in 1D statistics class
- Attempt to delete a value of the 'mean' or measurement uncertainty of a data point stored in 1D statistics class

Verification Method: T

Requirement ID: REQ-AWM-311

Title: 1D statistics class methods - improper argument data type

Description: The **TypeError** or its sub-class should be raised in response to

- *getQuantile*() method:
    - Total number of quantiles is not an integer number
    - Requested quantile index is not an integer number
- *getHistogram*() method:
    - Number of bins requested is not an integer number
    - Number of bins is not requested (OR None) and the requested bin size is neither integer nor floating point number

Verification Method: T

Requirement ID: REQ-AWM-312

Title: 1D statistics class methods - improper argument value of the proper data type

Description: The **ValueError** or its sub-class should be raised in response to

- *getQuantile*() method:
    - Total number of quantiles is an integer number, but not positive

- Requested quantile index is an integer number, but either negative or greater than the total number of quantiles
- Length of the stored data sequence is less than 2
- *getHistogram*() method:
    - Number of bins requested is an integer number, but not positive
    - Number of bins is not requested (OR None) and the requested bin size is integer or floating point number, but not positive

Verification Method: T