Requirements for the Module statistics_lib.stat_tests

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-700

Title: Functionality and purpose of the module

Description: The module should implement function performing the following statistical tests (see DE005 document):

- Z-test on sample's mean for the known population mean and standard deviation
- t-test on sample's mean for the known population mean but unknonw standard deviation
- Chi-squared test on the sample's variance for the known population standard deviation / variance
- unpaired t-test comparison of the means of two samples
- paired t-test comparison of the means of two samples
- Welch's t-test comparison of the means of two samples
- F-test comparison of the variances of two samples

- ANOVA homoscedasticity F-test comparison of the variances of two samples
- Levene test comparison of the variances of two samples
- Brown-Forsythe comparison of the variances of two samples

Verification Method: A

—

Requirement ID: REQ-FUN-710

Title: Z-test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for a single data sample, population's mean and standard deviation. Hence, the function must accept and expect the following arguments (in order):

- data sample
- population's mean
- population's standard deviation
- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-720

Title: Single sample Student's t-test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for a single data sample and population's mean. Hence, the function must accept and expect the following arguments (in order):

- data sample
- population's mean
- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-730

Title: Chi-squared test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for a single data sample and population's standard deviation. Hence, the function must accept and expect the following arguments (in order):

- data sample
- population's standard deviation

- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-740

Title: Unpaired Student's t-test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples. Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample
- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-750

Title: Paired Student's t-test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples and the expected difference between mean (optional, defaults to zero). Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample
- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - expected difference between means, defaults to zero
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-760

Title: Welch's t-test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples. Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample

- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-770

Title: F-test (generic)

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples and the expected populations variances ratio (optional, defaults to 1.0). Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample
- test type - 2-sided (not equal) or 1-sided (greater than or less than)
- optional, keyword - populations variances ratio, defaults to 1.0
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-780

Title: ANOVA homoscedasticity F-test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples. Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-790

Title: Levene homoscedasticity test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples. Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-7A0

Title: Brown-Forsythe homoscedasticity test

Description: The respective function should perform the statistical test, i.e. calculate the p-value and reject / fail to reject the null hypothesis at the given confidence level, for two data samples. Hence, the function must accept and expect the following arguments (in order):

- the first data sample
- the second data sample
- optional, keyword - confidence interval

Verification Method: T

—

Requirement ID: REQ-FUN-7B0

Title: Instantiation of the return value class

Description: The test result report class should be instantiated with the following arguments:

- string: test name and discription, including confidence level and used test's parameters
- string: data sets / samples names
- string: model distribution name and parameters
- integer or floating point: calculated test value
- floating point in range (0, 1): calculated CDF at test value
- 2-tuple of integer or floating point or None (both cannot be None): calculated critical values of the test

Verification Method: T

—

Requirement ID: REQ-FUN-7B1

Title: Base functionality of the return value class

Description: Based on the values of the arguments of the initialization method, which should be stored as the class instance's state, the respective instance should be able to infer the type of the test (2- sided or 1-sided greater / less), calculate the associated p-value and if the null hypothesis is rejected. Thus, the class should implement the following 3 read-only properties:

- boolean: rejection (True) or failure to reject (False) the null-hypothesis
- floating point in range [0, 1]: calculated p-value of the test
- string: human readable, multi-line test report

The format of the report should be comparable to the following examples.

Statistical test report. Name: Z-test at 95% confidence on sample's mean vs population mean = 1.2 and sigma = 0.3 Data: Some my measurement Type: 1-sided Model distribution: Z_Distribution() Null hypothesis: less then or equal to Alternative hypothesis: greater Critical value: 2.34 Test value: 2.1 p-value: 0.08 Is null hypothesis rejected?: No

Statistical test report. Name: Unpaired Student's t-test at 95% confidence on samples' means Data: Week_1 vs Week_3 Type: 2-sided Model distribution: Student(Degree = 47) Null hypothesis: equal Alternative hypothesis: not equal Critical values: (-2.45, 2.45) Test value: 3.1 p-value: 0.03 Is null hypothesis rejected?: Yes

Verification Method: T

Software system inputs and outputs

Requirement ID: REQ-SIO-700

Title: Test functions input data types

Description: The functions should calculate and use the proper statistical properties of the sample(s) distribution automatically, instead of relying on the user provided values, thus minimizing the chance of the misleading conclustions due to faulty input. Therefore, the following input data types should be expected by the test functions:

- The actual data sample(s) should be passed (by reference) as instances of **statistics_lib.data_classes.Statistics1D** class instead of the pre-calculated values like sample mean or variance, etc.
- The additional parameters of the test should be passed as native Python data types **int** or **float** values
- The type of the test (1- or 2-sided), greater or less (for 1-sided) should be unambiquously identified by using the enumeration data type
- The confidence level of the test should be passed as **float** type value in the range (0, 1)

Verification Method: T

—

Requirement ID: REQ-SIO-701

Title: Test functions arguments order

Description: The following convention should be followed in the signature of the test functions:

- The data sample(s) should be passed as the first mandatory positional argument(s)
- The mandatory test parameters (for Z-test, Chi-squared test and single sample Student's t-test) should be mandatory positional arguments following the data sample(s) arguments
- The test type (enumeration value) should be the last mandatory positional argument, except for the ANOVA, Levene and Brown-Forsythe tests, where it is not needed, since they are by definition single-sided (techincally, greater, although the test hypothesis is equal / unequal).
- The confidence level should be passed as keyword-only argument (with the default value or 0.95)
- The optional parameters (as delta in generic F-test and expected difference in paired t-test) should be passed as keyword-only arguments

Verification Method: T

—

Requirement ID: REQ-SIO-702

Title: Test functions return value

Description: The test functions should return an instance of a report class - see REQ-FUN-7B0 and REQ-FUN-7B1

Verification Method: T

Alarms, warnings and operator messages

Requirement ID: REQ-AWM-700

Title: Improper input data types

Description: An sub-class of **TypeError** exception should be raised if, at least, one of the arguments of the test function call is of the improper data type, specifically:

- Not an instance of **statistics_lib.data_classes.Statistics1D** for expected data sample parameter
- Not an instance of **int** or **float** for the expected mandatory or optional test parameter
- Not an enumeration value of the expected test type parameter
- Not a **float** value for the expected confidence level parameter

Verification Method: T

—

Requirement ID: REQ-AWM-701

Title: Improper input values of the proper type

Description: An sub-class of **ValueError** exception should be raised if, at least, one of the arguments have unacceptable value of the proper type, specifically:

- The confidence level is not in the interval (0, 1) - all tests
- Population standard deviation is not positive - Z-test and chi-squared test
- Delta parameter of the generic F-test is not positive
- Length of the data sample (any of two) is less than 2 elements

This exception should also be raised by the paired Student's t-test if the data samples differ in length.

Verification Method: T