RED THREAD®
by Ariadne Software

# Cross-Validation of Statistical Workflow for Cut Point Calculation using Red Thread (Python) Vs R

bdatasolutions.com

## INTRODUCTION:

In the traditional approach to cut point determination (here referred to as the outlier approach), analytical (intra-subject) and biological (inter-subject) outliers are removed from the dataset based on the Tukey Outlier test using the interquartile range. The new resulting distribution is then used to determine a cut point, using parametric method for normal/log-normal distributed data and nonparametric method for other types of distributions.

Red Thread's ADA module implements this approach in Python, an open-source programming language, using open-source Python packages. While Red Thread has been thoroughly tested to ensure its results are accurate, an additional level of validation can be achieved by implementing the same approach in another language such as R, a programming language used for statistical analysis. In addition to comparing the final cut point values, it is necessary to compare which outliers are removed and the results of the normality test to confirm equivalence of the entire statistical workflow and decision tree.

## PURPOSE:

The goal of this study is to validate the cut point values of Red Thread's ADA module through the duplication of the entire decision tree and workflow in a separate programming language, R, that is commonly used for statistical analysis. This comparison study used six anonymized datasets that cover a wide range of cases to ensure that the software is validated for all anticipated cases.

| DATASET | TEST CASE |
|---|---|
| Dataset 1 | Low SCP (< 1.10, S/N);<br>High CCP (> 50% inhibition);<br>Outliers > 15% (both SCP and CCP) |
| Dataset 2 | Neutralizing antibody (NAB) data |
| Dataset 3 | 1.1 < SCP < 2;<br>10% < CCP < 50%;<br>5% < Outliers < 15% |
| Dataset 4 | High SCP (> 2);<br>High CCP;<br>Outliers > 15% for both |
| Dataset 5 | Low CCP (< 10% inhibition) |
| Dataset 6 | Outliers < 5% |

Table 1: Dataset Summary. SCP – Screening Cut Point; CCP – Confirmatory Cut Point

## APPROACH:

The Python implementation (as used in Red Thread) leverages the modules ***numpy***, ***scipy***, and ***pandas*** in calculating cut points: ***numpy*** and ***scipy*** provide mathematical and statistical utilities for performing calculations on matrices and distributions; and the ***pandas*** module provides various data science functions including manipulating for efficient organization and computation on structured data. The R cut point module primarily leverages the R packages ***readxl***, ***dplyr***, and ***stringr***: ***readxl*** provides functionality of reading data from Excel files; ***dplyr*** is designed for data manipulation and transformation; and ***stringr*** provides functionality of string manipulation and pattern matching.

The Python and R programs have been built by different internal teams to implement the following approach to cut point determination (Figure 1):

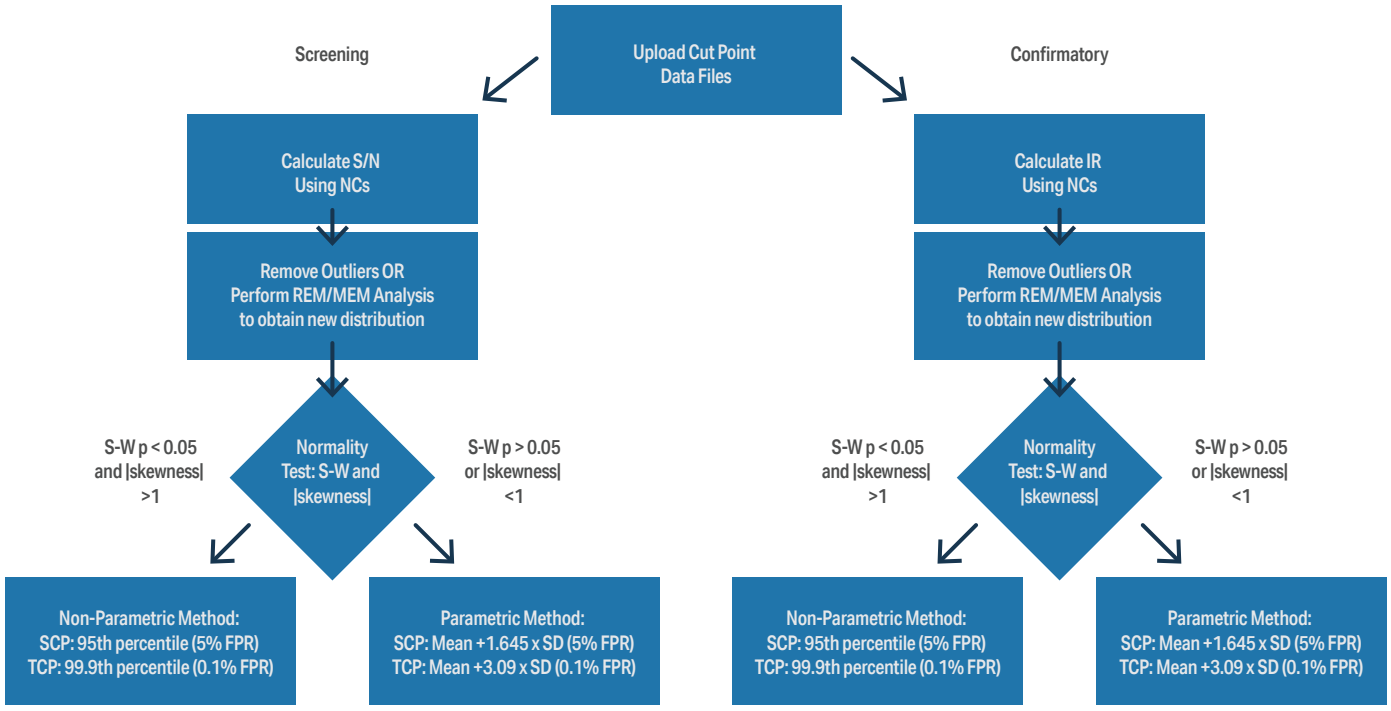


Figure 1: Red Thread ADA module workflow/decision tree

In summary, first analytical (intra-subject) outliers are removed based on the Tukey Outlier Test. The median values for all observations across each subject are then recalculated, and biological (inter-subject) outliers are subsequently removed from the dataset. The distribution/skewness assumption is evaluated subsequently, and if the assumption fails, a log-transformation is performed on the original data, and the outlier removal process is re-evaluated. Finally, depending on the distribution/skewness of the data with outliers removed, a parametric/nonparametric cutpoint is calculated.

## RESULTS:

The results from Red Thread's ADA module can be replicated by implementing the outlier approach for cut point determination in R across six datasets. All screening, confirmatory, and titer cut point values are equivalent across all six datasets and out to five decimal places, as limited by Red Thread's maximum output. Additionally, the outliers removed by both implementations are the same across two IQR levels each (IQR of 1.5 and 3), and the normality test results (parametric vs. non-parametric) are equivalent.

| File Name | CP type | Total Observation | Analytical Outliers | Biological Outliers | Accepted Samples | *p-value* | *Skew* | Normality Test | CP Value | Confidence Level |
|---|---|---|---|---|---|---|---|---|---|---|
| Dataset_1 IQR 1.5 | Screening -Python | 159 | 14 | 16 | 129 | *p<0.05* | *<1* | Pass | 1.09500 | 95 |
| | Screening -R | 159 | 14 | 16 | 129 | *p<0.05* | *<1* | Pass | 1.09500 | 95 |
| | Titer-Python | 159 | 14 | 16 | 129 | *p<0.05* | *<1* | Pass | 1.28540 | 99.9 |
| | Titer-R | 159 | 14 | 16 | 129 | *p<0.05* | *<1* | Pass | 1.28540 | 99.9 |
| | Confirmatory-Python | 156 | 22 | 15 | 119 | *p>0.05* | *<1* | Pass | 53.44733% | 99 |
| | Confirmatory-R | 156 | 22 | 15 | 119 | *p>0.05* | *<1* | Pass | 53.44733% | 99 |
| Dataset 1 IQR 3 | Screening -Python | 159 | 7 | 16 | 136 | *p<0.05* | *<1* | Pass | 1.09272 | 95 |
| | Screening -R | 159 | 7 | 16 | 136 | *p<0.05* | *<1* | Pass | 1.09272 | 95 |
| | Titer-Python | 159 | 7 | 16 | 136 | *p<0.05* | *<1* | Pass | 1.28584 | 99.9 |
| | Titer-R | 159 | 7 | 16 | 136 | *p<0.05* | *<1* | Pass | 1.28584 | 99.9 |
| | Confirmatory-Python | 156 | 13 | 10 | 133 | *p<0.05* | *>1* | Fail | 66.45868% | 99 |
| | Confirmatory-R | 156 | 13 | 10 | 133 | *p<0.05* | *>1* | Fail | 66.45868% | 99 |
| Dataset_2 IQR 1.5 | Screening -Python | 220 | 14 | 22 | 184 | *p<0.05* | *<1* | Pass | 0.88263 | 95 |
| | Screening -R | 220 | 14 | 22 | 184 | *p<0.05* | *<1* | Pass | 0.88263 | 95 |
| | Screening -Python | 220 | 14 | 22 | 184 | *p<0.05* | *<1* | Pass | 0.84795 | 99 |
| | Screening -R | 220 | 14 | 22 | 184 | *p<0.05* | *<1* | Pass | 0.84795 | 99 |
| | Screening -Python | 220 | 14 | 22 | 184 | *p<0.05* | *<1* | Pass | 0.80909 | 99.9 |
| | Screening -R | 220 | 14 | 22 | 184 | *p<0.05* | *<1* | Pass | 0.80909 | 99.9 |
| Dataset_2 IQR 3 | Screening -Python | 220 | 3 | 22 | 195 | *p>0.05* | *<1* | Pass | 0.87802 | 95 |
| | Screening -R | 220 | 3 | 22 | 195 | *p>0.05* | *<1* | Pass | 0.87802 | 95 |
| | Screening -Python | 220 | 3 | 22 | 195 | *p>0.05* | *<1* | Pass | 0.84110 | 99 |
| | Screening -R | 220 | 3 | 22 | 195 | *p>0.05* | *<1* | Pass | 0.84110 | 99 |
| | Screening -Python | 220 | 3 | 22 | 195 | *p>0.05* | *<1* | Pass | 0.79972 | 99.9 |
| | Screening -R | 220 | 3 | 22 | 195 | *p>0.05* | *<1* | Pass | 0.79972 | 99.9 |
| Dataset_3 IQR 1.5 | Screening -Python | 297 | 41 | 6 | 250 | *p<0.05* | *<1* | Pass | 1.03434 | 95 |
| | Screening -R | 297 | 41 | 6 | 250 | *p<0.05* | *<1* | Pass | 1.03434 | 95 |
| | Screening -Python | 297 | 41 | 6 | 250 | *p<0.05* | *<1* | Pass | 1.08903 | 99.9 |
| | Screening -R | 297 | 41 | 6 | 250 | *p<0.05* | *<1* | Pass | 1.08903 | 99.9 |
| | Confirmatory-Python | 294 | 15 | 6 | 273 | *p>0.05* | *<1* | Pass | 14.84036% | 99 |
| | Confirmatory-R | 294 | 15 | 6 | 273 | *p>0.05* | *<1* | Pass | 14.84036% | 99 |
| Dataset_3 IQR 3 | Screening -Python | 297 | 18 | 6 | 273 | *p<0.05* | *>1* | Failed | 1.03456 | 95 |
| | Screening -R | 297 | 18 | 6 | 273 | *p<0.05* | *>1* | Fail | 1.03456 | 95 |
| | Screening -Python | 297 | 18 | 6 | 273 | *p<0.05* | *>1* | Failed | 1.14247 | 99.9 |
| | Screening -R | 297 | 18 | 6 | 273 | *p<0.05* | *>1* | Fail | 1.14247 | 99.9 |
| | Confirmatory-Python | 294 | 9 | 6 | 279 | *p<0.05* | *<1* | Pass | 15.45937% | 99 |
| | Confirmatory-R | 294 | 9 | 6 | 279 | *p<0.05* | *<1* | Pass | 15.45937% | 99 |
| Dataset_4 IQR 1.5 | Screening -Python | 597 | 71 | 112 | 414 | *p<0.05* | *>1* | Fail | 4.01757 | 95 |
| | Screening -R | 597 | 71 | 112 | 414 | *p<0.05* | *>1* | Fail | 4.01757 | 95 |
| | Titer-Python | 597 | 71 | 112 | 414 | *p<0.05* | *>1* | Fail | 6.58263 | 99.9 |
| | Titer-R | 597 | 71 | 112 | 414 | *p<0.05* | *>1* | Fail | 6.58263 | 99.9 |
| | Confirmatory-Python | 594 | 41 | 0 | 553 | *p<0.05* | *<1* | Pass | 88.59740% | 99 |
| | Confirmatory-R | 594 | 41 | 0 | 553 | *p<0.05* | *<1* | Pass | 88.59740% | 99 |
| Dataset_4 IQR 3 | Screening -Python | 597 | 38 | 108 | 451 | *p<0.05* | *>1* | Fail | 5.12445 | 95 |
| | Screening -R | 597 | 38 | 108 | 451 | *p<0.05* | *>1* | Fail | 5.12445 | 95 |
| | Titer-Python | 597 | 38 | 108 | 451 | *p<0.05* | *>1* | Fail | 12.36959 | 99.9 |
| | Titer-R | 597 | 38 | 108 | 451 | *p<0.05* | *>1* | Fail | 12.36959 | 99.9 |
| | Confirmatory-Python | 594 | 22 | 0 | 572 | *p<0.05* | *>1* | Fail | 86.41012% | 99 |
| | Confirmatory-R | 594 | 22 | 0 | 572 | *p<0.05* | *>1* | Fail | 86.41012% | 99 |
| Dataset_5 IQR 1.5 | Confirmatory-Python | 282 | 30 | 22 | 230 | *p<0.05* | *<1* | Pass | 8.15490% | 99 |
| | Confirmatory-R | 282 | 30 | 22 | 230 | *p<0.05* | *<1* | Pass | 8.15490% | 99 |
| Dataset_5 IQR 3 | Confirmatory-Python | 282 | 10 | 24 | 248 | *p<0.05* | *<1* | Pass | 8.87391% | 99 |
| | Confirmatory-R | 282 | 10 | 24 | 248 | *p<0.05* | *<1* | Pass | 8.87391% | 99 |
| Dataset_6 IQR 1.5 | Screening -Python | 884 | 0 | 23 | 861 | *p<0.05* | *<1* | Pass | 1.12877 | 95 |
| | Screening -R | 884 | 0 | 23 | 861 | *p<0.05* | *<1* | Pass | 1.12877 | 95 |
| | Titer-Python | 884 | 0 | 23 | 861 | *p<0.05* | *<1* | Pass | 1.45038 | 99.9 |
| | Titer-R | 884 | 0 | 23 | 861 | *p<0.05* | *<1* | Pass | 1.45038 | 99.9 |
| Dataset_6 IQR 3 | Screening -Python | 884 | 0 | 6 | 878 | *p<0.05* | *<1* | Pass | 1.18048 | 95 |
| | Screening -R | 884 | 0 | 6 | 878 | *p<0.05* | *<1* | Pass | 1.18048 | 95 |
| | Titer-Python | 884 | 0 | 6 | 878 | *p<0.05* | *<1* | Pass | 1.53449 | 99.9 |
| | Titer-R | 884 | 0 | 6 | 878 | *p<0.05* | *<1* | Pass | 1.53449 | 99.9 |

Table 2: Comparison of Statistical Workflow: Python vs R

## CONCLUSION:

The Red Thread workflow, implemented in Python, was cross validated in the identical workflow written in the statistical language R as demonstrated by the equivalency of outlier selection, normality results and calculated cut point values for all three tiers across the six tested datasets.

## REFERENCES:

Food and Drug Administration. Guidance for Industry. "Immunogenicity Testing of Therapeutic Protein Products – Developing and Validating Assays for Anti-Drug Antibody Detection. Jan. 2019.

Shankar, Gopi, et al. "Recommendations for the Validation of Immunoassays Used for Detection of Host Antibodies Against Biotechnology Products." *Journal of Pharmaceutical and Biomedical Analysis*, vol. 48, no. 5, Elsevier BV, Dec. 2008, pp. 1267–81. https://doi.org/10.1016/j.jpba.2008.09.020.

Devanarayan, Viswanath, Wendell C. Smith, Rocco L. Brunelle, Mary E. Seger, Kim Krug, and Ronald R. Bowsher. "Recommendations for systematic statistical computation of immunogenicity cut points." *The AAPS journal* 19, no. 5 (2017): 1487-1498.

Kubiak, Robert J., Jianchun Zhang, Pin Ren, Harry Yang, and Lorin K. Roskos. "Excessive outlier removal may result in cut points that are not suitable for immunogenicity assessments." *Journal of immunological methods* 463 (2018): 105-111.