**Data Analysis**

**Development Documentation**

Program Designer: Effend Wang

Program Version: v0.7

Release Time: 2020/

**Contents**

[Version Information 2](#_Toc48505672)

[Introduction 3](#_Toc48505673)

[Develop Environments 3](#_Toc48505674)

[Program Frame Design 4](#_Toc48505675)

[Program Begin 4](#_Toc48505676)

[CPK Analysis 4](#_Toc48505677)

[Test Coverage Analysis 5](#_Toc48505678)

# Version Information

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Release Date** | **Add Info** | **Change Info** |
| v0.1 | / | File read & write test.  Calculate AVG & STD of data. | / |
| v0.2 | 2019/08/15 | Add logging function, output message and CPK calculate. | Optimize excel file read process. |
| v0.3 | 2019/08/19 | Add new data in result file.  Add module name in log file. | Optimize code into modular code.  Optimize STD algorithm to avoid computational error. |
| v0.4 | 2019/08/23 | Add tip at the end of program.  Add warning tips in log file; Add update message file. | Fix wrong description in log file.  Use openpyxl instead of xlwt to ensure large data processing. |
| v0.5 | 2019/08/29 | Add test coverage analysis function. | Optimize function call.  Optimize data processing. |
| v0.6 | 2019/12/15 | Add CPK data distribution image output function. | Fix wrong description in log file and optimize messages.  Fix wrong output in test coverage analysis. |
| v0.7 | 2020/08/16 | / | Fix wrong output of result excel file.  Move result files into independent result folder. |

# Introduction

Data Analysis program is used to analyze data from an excel form document and output analyze result, the result will be saved in a result folder. The program is designed to calculate the statistics of data.

This program is designed by Effend Wang.

Program Github page: <https://github.com/Effend-Wang/Data_Analysis>

Author’s email: effend\_wang@outlook.com

**Warning:** The program could only run at Vista/Win7/Win10 system.

# Develop Environments

1. **Python**: v3.8.1
2. **Python Library**: sys, datetime, os, shutil, xlrd, logging, math, openpyxl, numpy, matplotlib

# Program Frame Design

# Program Begin

At the beginning of program, create Result folder which will save result files. The Result folder name contains the time when program begin. In this way, it will be easier for user to distinguish results after several times use. Also, older results will not be auto deleted by program. The path must be defined first in program because log file will be written in Result folder.

# CPK Analysis

In CPK Analysis function, the program can output contents includes:

1. CPK result of data.
2. Distribution image of data.

When analyze CPK result, program’s processing flow is as below:

1. User input the absolute path of data file (This way is not very safe, should consider a better way).
2. Program will load data from file and output row & col numbers.
3. User input the beginning of data, input beginning row & col numbers separately.
4. User input the way to output distribution of data, includes: No output; Output <1.33 distribution; Output all distribution.
5. Program starts to analyze data.
6. Output analyze result, includes: CPK result file (.xlsx); Distribution image if chosen (.png).

This function has program defects includes:

1. The data file should only be .xlsx file, not support .csv file yet.
2. It still need user to input data beginning row & col numbers. Will consider a better way to ensure no more input.
3. Data cannot contain any blank column or program will report error. This will affect the efficiency of work. For this, user must check data first and ensure there is no blank column.
4. The distribution image settings are not good. Still need more change.

To see excel file example at path: example/cpk\_example.xlsx

# Test Coverage Analysis

In Test Coverage Analysis function, the program will compare two different data files and analyze the differences between two file’s test items. You will get results as below:

1. Same parameter name and limits.
2. Same parameter name but with different limits.
3. New parameter.
4. Deleted parameter.

In this analyze process, user only need to input the path of two files separately. Program will automatically find out data and calculate.

This function has program defects includes:

1. The data file should only be .xlsx file, not support .csv file yet.
2. Data cannot contain any blank row in data or program will report error. For this, user must check data first and ensure there is no blank row.

In result excel file, results will only show if it contains related item.

To see excel file example at path: example/test\_coverage\_new\_example.xlsx and test\_coverage\_old\_example.xlsx