AGM_Control_Chart_2024_02_12

Jeffrey Weidner

Summary

The basic functionality for the Control Charting web tool is complete. The code uses potency data from a single reference compound tested one or more times in each experiment across multiple independent experiments. After log transformation of the data, the qicharts2 package is used for statistical analysis with respect to the reproducibility of both the measured values and the variability between measurements within experiments. Additionally, if there are 6 or more experiments, MSR values are calculated to estimate the variability between experiments.

For individual data (no replicates) or data sets with replicates in $< 50\%$ of the run Individual Response (IR) and Moving Range (MR) charts are created. Average Potency (Xbar) and Standard Deviation (S) charts are created for data sets where all the experiments contain replicates and both sets of charts are created when $> 50\%$ of the experiments contain replicates. All charts and statistics are transformed back to the linear values for presentation to the user with log transformation of the y-axes on the charts, so that the values will appear to be normally distributed.

Scripts to generate pseudo-random data are also included to facilitate testing and development. The resulting data sets may be further modified to produce examples of specific error cases, such as random outliers, abrupt shifts in potency or variability, or trending changes.

User Input

Initially the user will asked to a name which will be used to label the output files. the section titles have been used as this name in the following examples to illustrate how this is used in the application. The user will also upload their data in a .csv file. This file must contain at least 2 columns labelled "Run" and "Data". The Run colum is an identifier for each Run and is a string of characters (preferably a date *most common date formats should be recognized). The Data column contains potency values for each replicate within a run. An optional column to identify the replicates within a run may also be supplied, but this is not used in the analysis.

Individual Data

Individual analysis is used when only a single potency value for each run is supplied, or if the majority of runs are individual values such as illustrated below:

```
## # A tibble: 30 x 2
##    Run          Data
##    <date>      <dbl>
##  1 2023-01-01 0.551
##  2 2023-01-08 0.447
##  3 2023-01-15 0.589
##  4 2023-01-22 1.03
##  5 2023-01-29 1.46
##  6 2023-02-05 1.58
##  7 2023-02-12 0.461
##  8 2023-02-19 0.900
##  9 2023-02-26 0.732
## 10 2023-03-05 1.26
## # i 20 more rows
```

This represents 30 individual runs of a compound with a potency of 1 and an MSR = 3.5. The application will return a table with all the data points used in the data and variability graphs including the Moving Range (MR) of the $\log_{10}$(Potency) which is transformed to Fold Moving Range (FMR)

```
## # A tibble: 30 x 5
##    Run          Data Log10Pot   MRLog    FMR
##    <date>      <dbl>    <dbl>   <dbl>  <dbl>
##  1 2023-01-01 0.551  -0.259   NA      NA
##  2 2023-01-08 0.447  -0.349    0.0903  1.23
##  3 2023-01-15 0.589  -0.230    0.119   1.32
##  4 2023-01-22 1.03    0.0142   0.244   1.76
##  5 2023-01-29 1.46    0.166    0.152   1.42
##  6 2023-02-05 1.58    0.199    0.0330  1.08
##  7 2023-02-12 0.461  -0.336    0.535   3.43
##  8 2023-02-19 0.900  -0.0458   0.291   1.95
##  9 2023-02-26 0.732  -0.135    0.0896  1.23
## 10 2023-03-05 1.26    0.100    0.236   1.72
## # i 20 more rows
```

The reproducibility of the potency values is shown in the Individual Response (IR) chart which plots the measured values and reference lines for the overall median and control limits (99% CI).
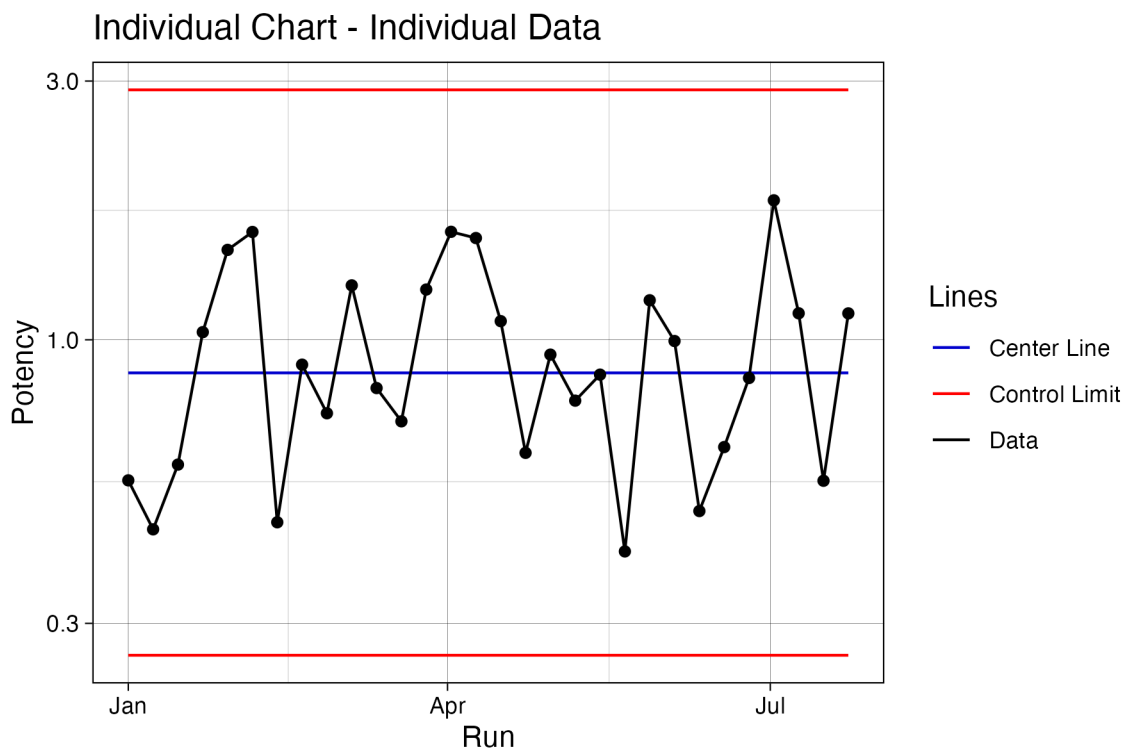

The y-axis for all graphs has been transformed to the $\log_{10}$ scale, so that the users original data points can be used directly and still appear to be normally distributed.

The statistics for the center and control limit lines are calculated by qicharts2 using the log potency data and then transformed back to the linear scale for the charts and returned to the user. Currently contains 16 values. This will ultimately be significantly reduced for export to the user, but at this time it contains useful information to customize the generated reports if values are observed outside of the control limits, or there are fewer crossings of the center line than would be expected from random variance. This will be addressed during development and testing of the webtool.

The within-run variability is monitored with a Moving Range chart for Individual data, however when this is transformed from the log scale the difference becomes a ratio, Fold Moving Range (FMR).

Individual Chart - Individual Data
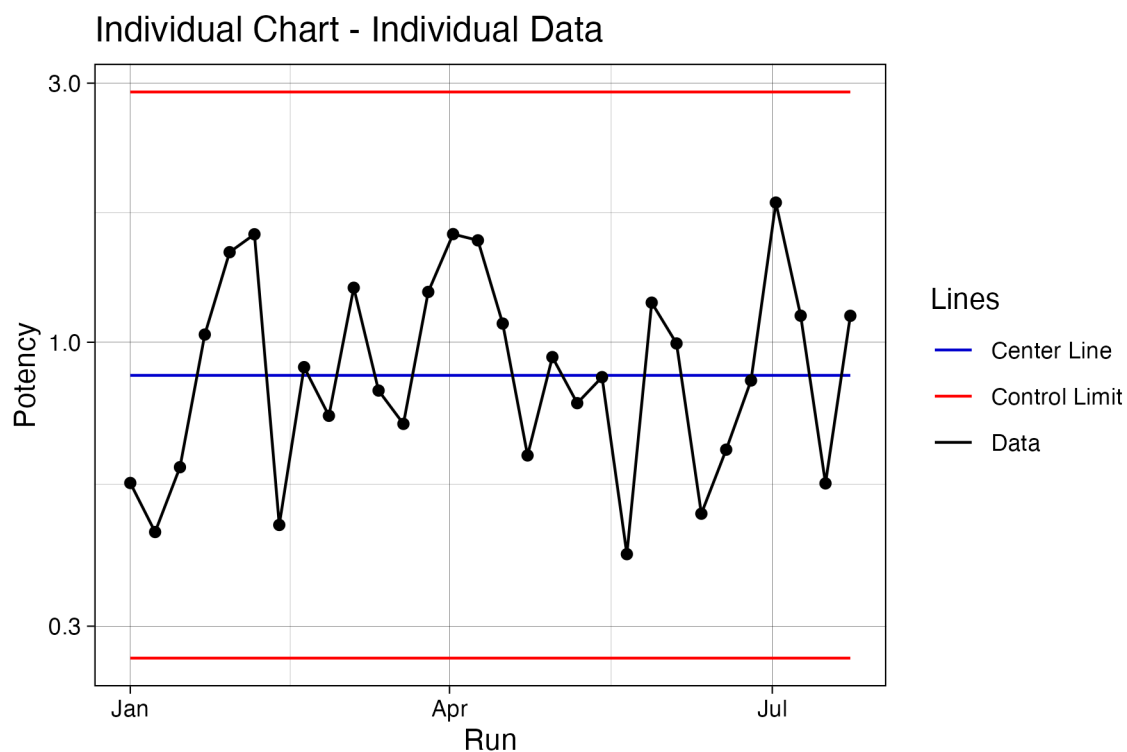
If there are at least 6 runs, a chart of the MSR values over a 6-run moving window will also be created. Additionally the cummulative MSR for all the previous data is also presented which should ultimately approach a stable value for the assay. No control limits are provided for this chart.
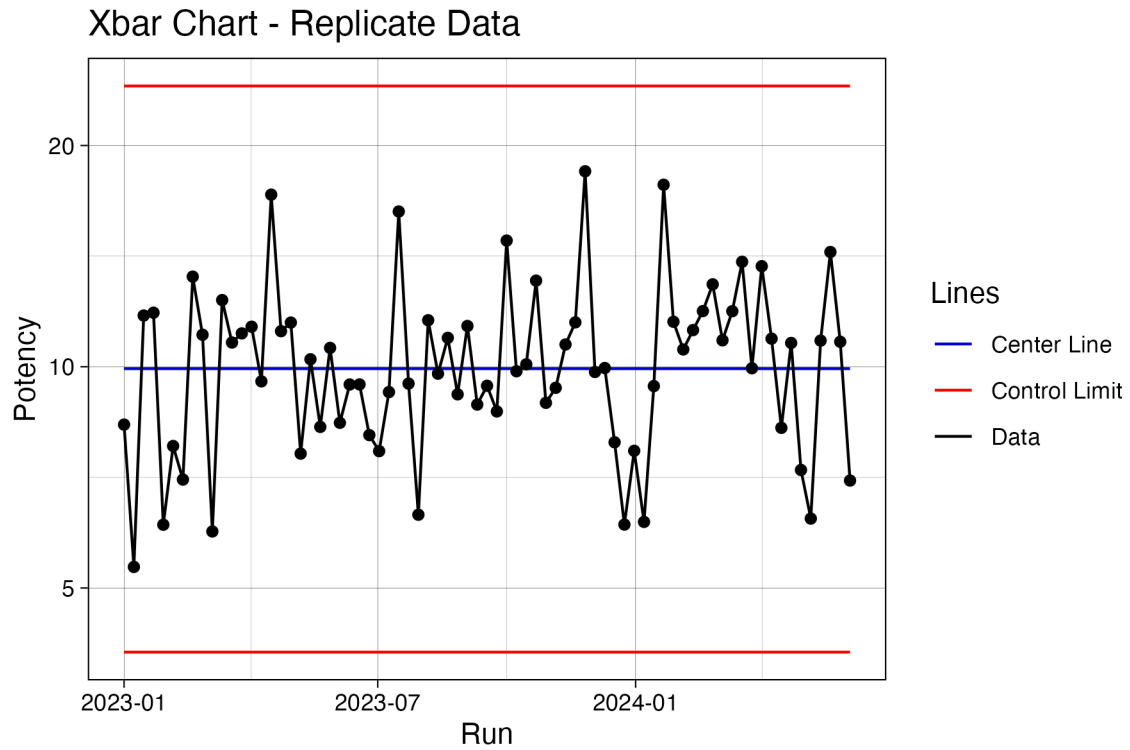
MSR Chart
Individual Data
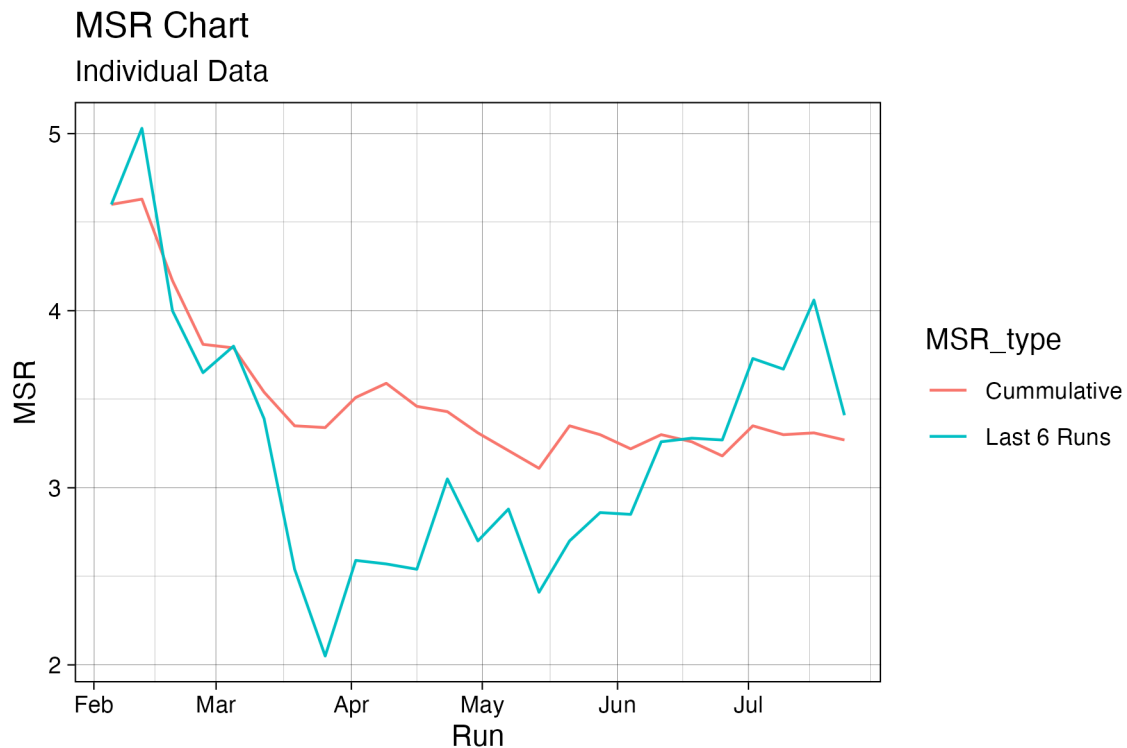
The individual moving MSR and cummulative MSR value are also provided as a table.
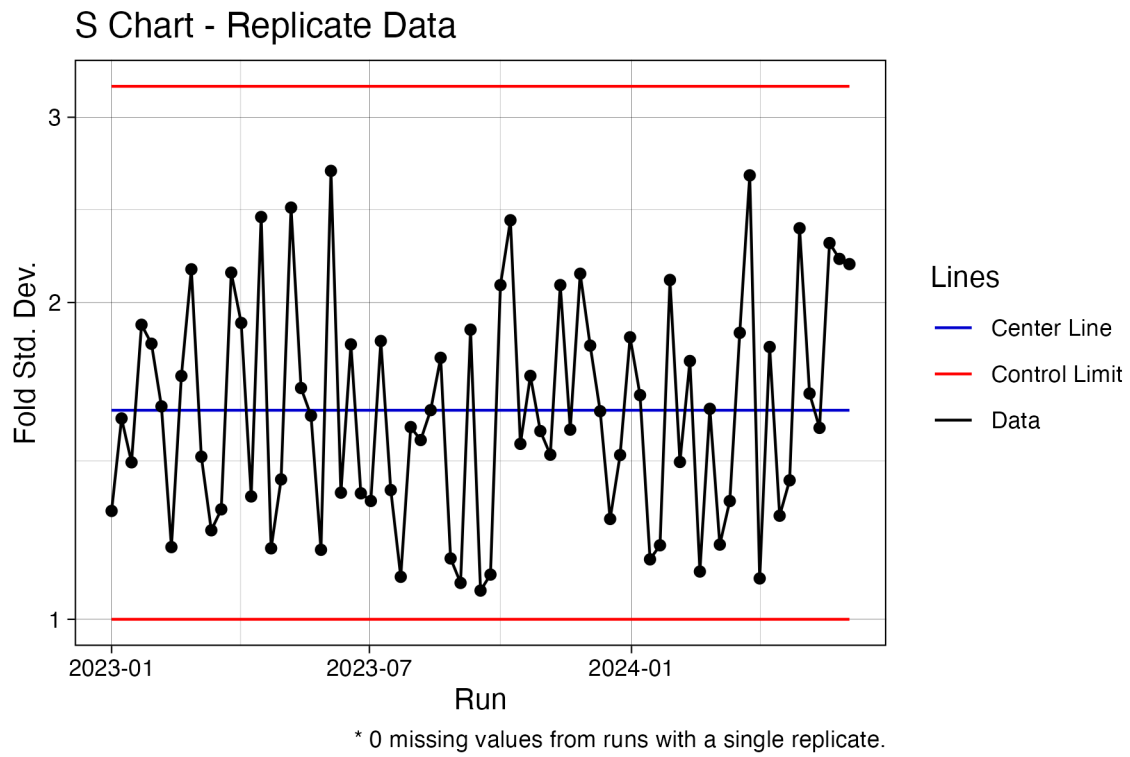
Replicate Data

When there are replicates within each run, log potency values are summarized for each run as the mean and standard deviation. Summary tables of these values are also provided for the report.

A chart of the mean potency values is presented as an Xbar chart along with reference lines for the center and control limits.
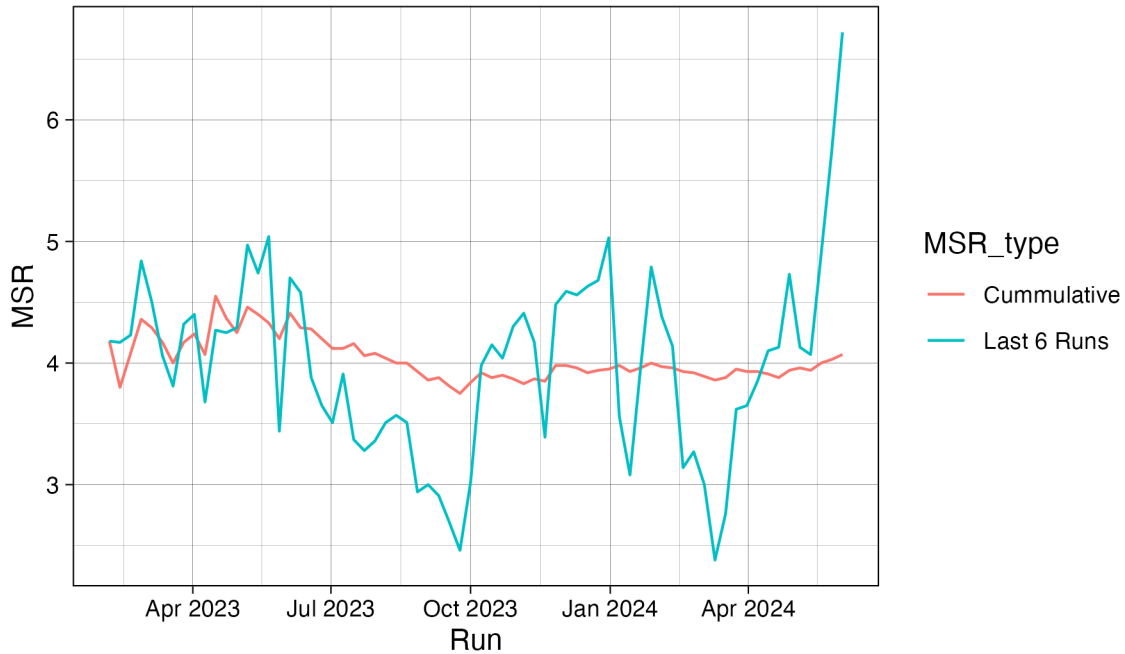
Variability is presented as a Standard Deviation chart (Schart) of the within run standard deviation transfomed to Fold Standard Deviation (FSD).

S Chart - Replicate Data

* 0 missing values from runs with a single replicate.

An MSR chart is also produced.

MSR Chart

Replicate Data

Mixed Data

Data sets which contain both n = 1 and n > 1 replicates present a challenge, since standard deviations can not be calculated for n = 1 runs. If the majority of the runs are individual values, then the entire data set is treated as individual data using the first value for each run and the analysis proceeds as with the Individual data above. If most runs contain replicates, then both the individual and replicate analysis paths are followed. The Xbar and S charts should be presented as the preferred interpretation, but the I charts and MR charts are available for the user as well. Alternatively this could be a user selected toggle.

Xbar Chart - Mixed Data
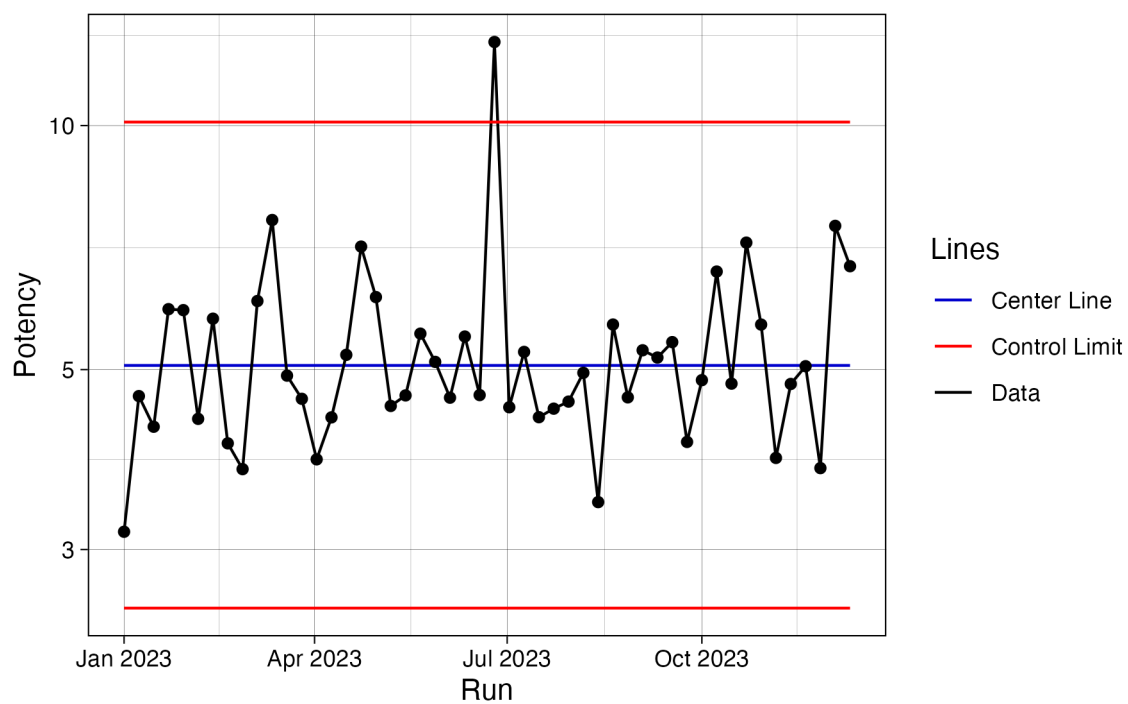
Individual Chart - Mixed Data
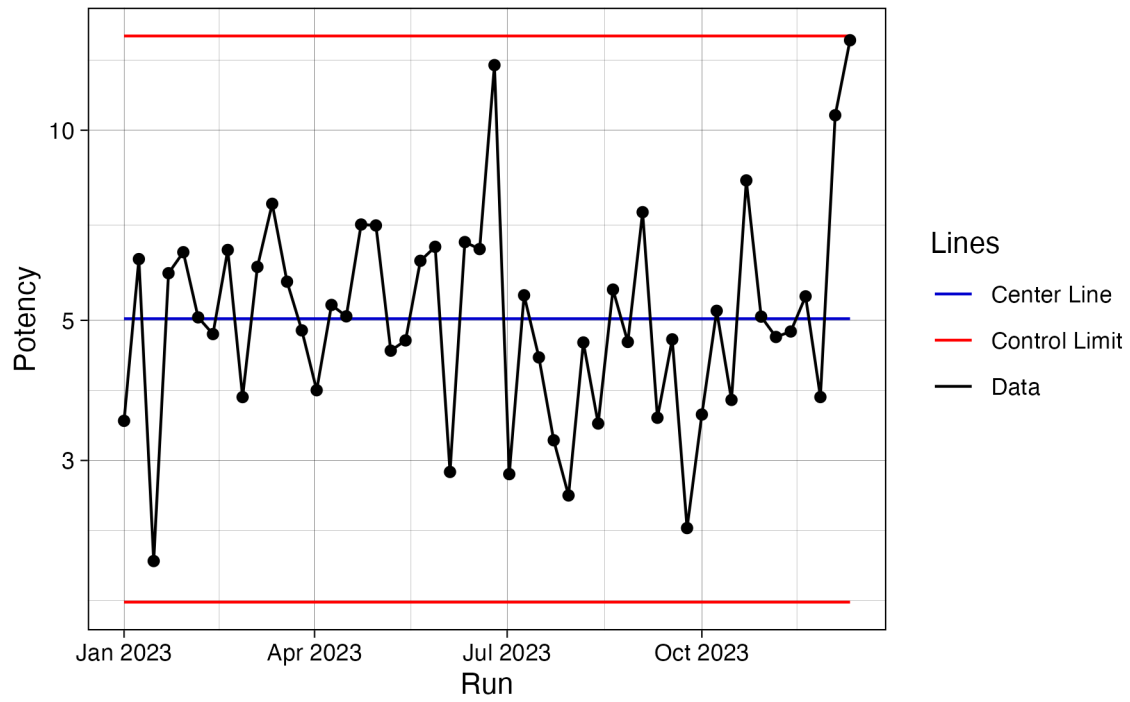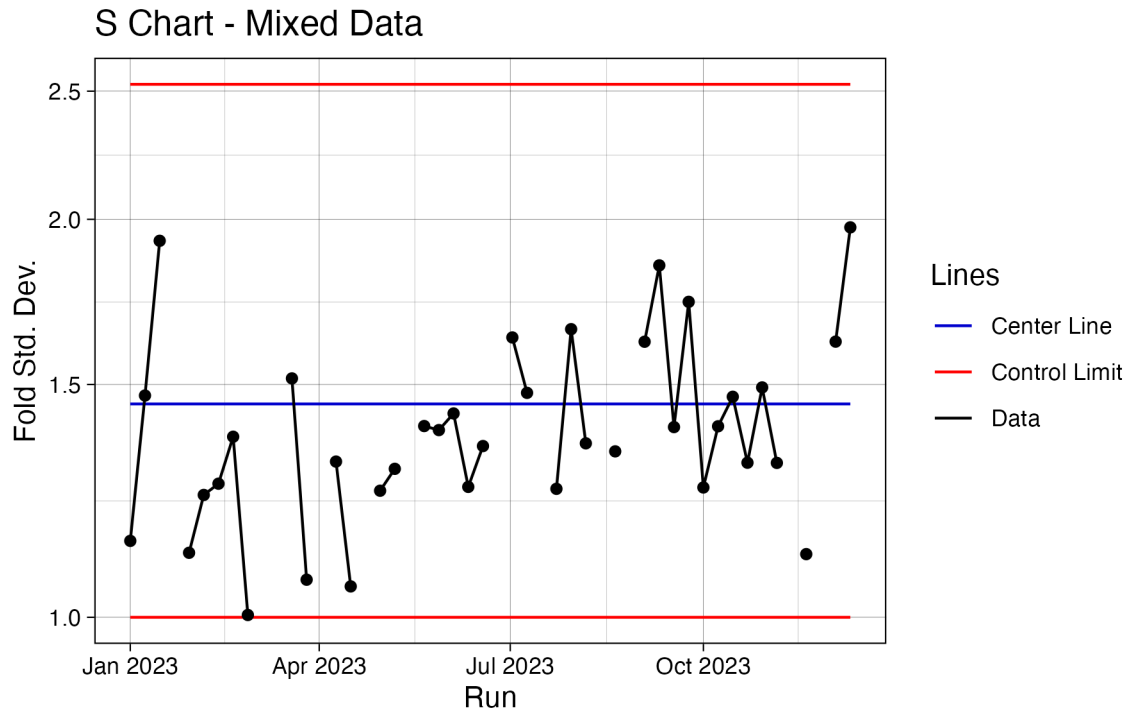
S Chart - Mixed Data

* 12 missing values from runs with a single replicate.

The gaps in the S chart occur where n = 1.

Fold Moving Range Chart - Mixed Data

MSR Chart

Mixed Data