# Tolerance analysis tool

## Background

This tool performance a Monte Carlo simulation to analyze dimension distribution.

Three main types of objects make up the analysis:

1. **Dimension** – A single value representing the size of a dimension. For example, the size of a hole, the length of a shaft, the thickness of a plate.
2. **Stackup** – A summation of dimensions whose overall value is controlled. For example, a stack of plates might have the total height control. A shaft and sleeve might have a controlled overall radius so it can fit in a hole.
3. **Product** – A collection of stackups. If one stackup fails, the entire product fails.

When the simulation runs, **instances** of dimensions are created by randomly sampling from a probability distribution. These are summed to create instances of stackups, whose overall values are checked against its limits. If an instance of a stackup fails, then the corresponding product instance also fails.

## Running the Analysis

To begin, information is entered in the *Main* sheet. Only the white cells should be modified. Any gray cell is either locked or is filled in automatically by the program. Ensure that the rows are filled out continuously. Any data below an empty row will be ignored.

1. Define dimensions by filling out the columns D to I. Each dimension occupies one row. The numbers in the C column represent the ID number of each dimension.
   1. **Name** (text) – The name of the dimension
   2. **Nominal** (number) – The nominal value of the dimension. This value can be negative.
   3. **Tolerance** (number) – The bilateral tolerance value around the nominal value. A dimension with nominal 1 and tolerance 0.1 would be marked on a drawing as 1 ± 0.1.
   4. **Accuracy** (number) – The number of standard deviations between the mean and the upper or lower limit. For example, if the accuracy is 3 and the tolerance is 0.3 then the probability distribution is constructed such that the standard deviation is 0.1. If this field is left blank, a value of 3 is used.
   5. **Distribution** (text) – The name and parameters of the distribution to use. If this field is left blank, a normal distribution is used. See the *Distribution* section for more details.
   6. **Mean shift** (number) – This optional field adds a shift to the mean. This is used to account for variability in manufacturing processes. For example, a 3D printed hole tends to be undersized. If this field is left blank, zero is used.
2. Define stackups by filling out the columns J to M. Each stackup occupies one row. The numbers in the C column represent the ID number of each stackup.
   1. **Name** (text) – The name of the stackup.
   2. **Dimensions** (text) – A reference to the dimensions that make up the stackup. The format is the ID number of each dimension, separated by a comma. For example, if three dimensions are defined, “0,1,2” is a valid dimension reference.
   3. **Lower** (number) – The lower bound of the stacked dimension.
   4. **Upper** (number) – The upper bound of the stacked dimension.

The remaining fields are outputs from the program.

* 1. **Nominal** (number) – The sum of all the nominal values of the dimensions that make up the stackup.
  2. **Passed %** (number) – The percentage of stackup instances that are within the bounds.
  3. **Under %** (number) – The percentage of stackup instances that are below the lower bound.
  4. **Over %** (number) – The percentage of stackup instances that are above the upper bound.

1. Define products by filling out the columns R and S. If any of the stackups in a product fails, the entire product fails. Each product occupies one row.
   1. **Name** (number) – The name of the product.
   2. **Stackups** (text) – A reference to the stackups that make up the product. This has the same format as the Dimensions column but refer to stackups instead. If three stackups are defined, then “0,1,2” is a valid stackup reference.

Column T is an output from the program.

* 1. **Passed %** (number) – The percentage of product instances where all of its stackups have passed.

Once all desired dimensions, stackups, and products have been defined, the analysis is ready to be run. In cell B5, the number of samples in the Monte Carlo simulation can be specified. Typically, a sample size above 100,000 is enough for fairly accurate results. The checkbox in A6 can be checked if you want to save the generated samples in the *Dimension samples* sheet. This will slow down the program. Press *Generate Samples* to start the analysis. After the analysis is completed, columns N to Q and column T will be populated with results.

## Distributions

The program includes six types of common distributions. If you are using the Python version, you can edit the permitted distribution types in *distributions.py*. The accepted inputs for the **Distribution** column are listed below. Besides the normal and uniform distributions, every other distribution requires shape parameters to fully define the distribution.

1. **normal** – The default distribution. Also known as the Gaussian distribution.
2. **skew,a** – The skewed normal distribution with shape parameter a.
3. **beta,α,β** – The beta distribution with shape parameters α and β.
4. **t,df** – Student’s t-distribution with shape parameter df.
5. **triang,c** – The triangular distribution with shape parameter c.
6. **uniform** – The uniform distribution.

Details for these distributions and their parameters can be found in the appendix.

## Fitting Empirical Data

If you have obtained actual measurements of a dimension, the tool can fit a distribution and generate more accurate samples. With more measurements, the better the fit.

1. Go to the row of the dimension you have measurements for. In column H, change the distribution to “**fit**”.
2. Go to the *Dimension Data* sheet. Fill in the experimental data under the columns corresponding to the correct dimensions, starting on row 6.
3. On row 5, fill in the number of bins to use when fitting the data. 200 seems to be a good value.
4. Go back to the *Main* sheet and press *Fit Data.* After this complete, rows 2 to 3 in the *Dimension Data* sheet should be populated.
5. Press *Generate Samples* to start the analysis.

Note that *Fit Data* will only need to be done when new empirical measurements are added. As long as rows 2 to 3 in the *Dimension Data* sheet is populated, the **fit** distribution can be used.

## Plotting Results

The tool can also plot results from the analysis as histograms. Though Excel can be used to create graphs, the tool’s plots are more interactive, and allow easy comparison of different dimensions. Note that to use the plotting functions, samples be saved. Make sure *Generate Samples* was run with cell A6 ticked so there are samples to plot.

1. Fill cell B9 with the number of bins to use in histogram plots.
2. Fill cell B11 and B16 with the dimensions and stackups you want to plot. The format is similar to columns K and S, ID numbers separated by commas.
3. On rows 12-14 and 17-19, select the desired plot type.
   1. **Single figure and plot** – All histograms are plotted in a single figure with shared axes. This allows easy comparison of distributions.
   2. **Single figure, multiple plots** – The histograms are plotted in a single figure but arranged vertically, each with its own y-axis. They share the same x-axis. This is useful when comparing distributions that would overlap if plotted with shared axes.
   3. **Multiple figures** – Each histogram is plotted in its own figure, with its own axes. If plotting many dimensions/stackups, then many windows will be opened.
4. Press *Generate Plots* to show the figures. Note that no other action can be taken until all windows have been closed.

# Appendix

## Normal Distribution

The normal distribution will be centered about the nominal dimension, with a standard deviation equal to the tolerance divided by the accuracy. For example, if a dimension is defined with nominal 1.5, tolerance 0.15, accuracy 3, and distribution “normal”, the dimension instances will be normally distributed with a mean of 1.5 and a standard deviation of 0.05.

## Skewed Normal Distribution

The skewed normal distribution is the normal distribution but with a non-zero skewedness **a**. If **a** is larger than zero, the distribution skews towards higher values. If **a** is smaller than zero, the distribution skews towards lower values. When **a** is zero, the distribution is normal.

## Beta Distribution

The beta distribution is a family of probability distributions parametrized by two positive shape parameters α and β. The mean of the distribution is shifted to fall on the desired nominal value. The location of the mean can be computed with the following equation:

Where *μ* is the nominal value, *μS* is the shifted mean, and *σ* is the standard deviation. For example, if α and β have the same value, the mean will be shifted by 0.5*σ.*

## Student’s T-distribution

The Student’s t-distribution is useful when estimating the mean of a normally distributed population but the sample size is small and population standard deviation is unknown. Essentially, it accounts for uncertainty by spreading out the normal distribution. As the number of degrees of freedom grows, the Student’s t-distribution approaches the normal distribution. The degrees of freedom **df** is calculated by subtracting 1 from the number of samples.

## Triangular Distribution

The triangular distribution takes on the shape of a triangle and is often used as a “lack of knowledge” distribution. By guessing the lower bound, the upper bound, and the most likely value, a triangular distribution can be created. The probability begins increasing starting at the lower bound, starts decreasing at the most likely value, and becomes zero at the upper bound. In this program, the nominal value minus the tolerance is used as the lower bound, the nominal value plus the tolerance is used as the upper bound, and the shape parameter **c** is the most likely value. For example, if the nominal value is 0, the tolerance is 1, and the shape parameter is 0.5, then a triangular distribution will be created with lower bound -1, upper bound 1, and the peak of the triangle will be perfectly centered at 0. This distribution ignores the **accuracy** parameter.

## Uniform Distribution

The probability is constant throughout the span of the distribution. Outside the span the probability is zero. This represents a situation where all values within the span are equally likely to occur. The mean of the distribution is at the nominal value, and the span is equal to double the tolerance. This distribution ignores the **accuracy** parameter.