Comparing x- and y-axis Consistencies on a CNC Router

Jiahong Long

23 May 2019. Revised 29 September 2019

1 Abstract

In this study, we seek to determine whether or not there is a statistically significant difference between the x- and y-axes on a CNC router, which is an advanced computer-controlled metal machining device. To do so, we will machine a set of features with a fixed distance, and we will compare the distributions of the differences between the theoretical fixed distance and the actual distance across both axes using a test performed on the difference of sample means. A two-sample-mean t-test revealed that there was no statistically significant difference between the two axes.

2 Introduction and Motivation

We begin with a brief description of the machine from which we are collecting data. A CNC router is essentially an table with a vertically-adjustable rotating tool bit (eg. drill bit) mounted on computer-adjusted x- and y-axes. They may be used in conjunction with patterns that tell the machine to follow a certain 3-dimensional path, thereby cutting the patterns into whatever material has been mounted on the table. This pattern generally represents a physical part or an object.

The CNC machine from which this data will be collected is one that my robotics team uses in the manufacture of robot components, and is notable for frequently miscutting (cutting in the wrong location) due to a lack of overall rigidity in the motorised x- and y-axes. A lack of rigidity means the location where the machine actually cuts may be slightly offset from where the tool path dictates the cut must be, simply due to the forces exerted on the axes by the cutting motion itself.

In this experiment we seek to determine whether there is a difference between the x- and y-movement axes when it comes to inaccuracy. For instance, if the x- axis has a higher mean inaccuracy as compared to the y-axis, we may conclude that the x-axis is less rigid than the y-.

3 Collection of Data

To compare two distributions, we must have two datasets, one for each axis that we are examining. There are a number of criteria that the collected data must meet.

Any given dataset must:

1. Isolate a single axis. y-axis movements should not impact the x-axis dataset and vice versa.

2. Be obtained from easy-to-obtain measurements. This is to say, measurements should not need to be made using any particular special equipment, and if they are to be made from finished parts, the parts they are obtained from should not be overly complex.

With this in mind, the data collection method ultimately selected was the following:

1. Along the x-axis, drill a series of fifty-one holes using a reference-level precision 0.250000" tungsten carbide drill bit. Space each of the holes 0.5000" apart in the tool path.

2. Repeat the above along the y-axis.

3. Using a set of high-precision calipers, measure the distances between the holes. Because it is impossible to measure directly between the centres of the holes, we will instead measure the shortest distance between the inner faces between the two holes. Simple geometry informs us that the *expected* distance between these two faces is 0.2500".

4. Subtract 0.2500" (the expected distance) from each collected data point to centre the distribution around zero.
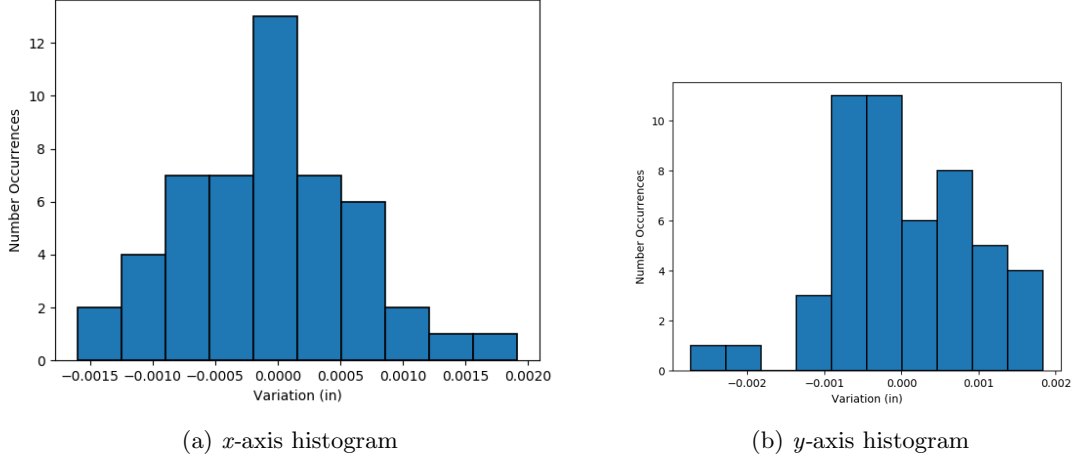
The raw data is as follows:

(a) *x*-axis histogram

(b) *y*-axis histogram

Figure 1: Histograms of the Deviations

4.277e-04	-4.106e-04	-1.205e-04	7.911e-04	-1.454e-03
-8.163e-04	-5.876e-04	8.623e-05	-4.329e-04	-5.936e-04
1.811e-04	1.917e-03	-4.792e-06	8.693e-05	-1.179e-04
8.631e-04	-3.820e-04	6.648e-04	-9.184e-04	-3.069e-04
-1.790e-04	1.104e-03	-6.383e-04	-6.624e-04	-4.448e-05
5.427e-04	6.318e-04	-1.002e-03	5.895e-04	1.450e-04
3.803e-04	-9.752e-04	-1.188e-03	3.219e-04	5.091e-04
-1.711e-04	2.272e-04	-4.260e-05	2.394e-04	1.346e-04
-1.278e-04	-4.862e-04	-1.602e-03	-2.274e-04	-6.903e-04
-8.437e-04	1.379e-03	7.358e-04	3.037e-05	-3.287e-04

Table 1: *x*-axis raw data

1.614e-04	6.021e-04	-3.644e-04	1.750e-04	-5.949e-04
-2.465e-04	1.064e-03	-4.886e-04	-6.145e-04	1.267e-03
1.428e-03	9.676e-04	-1.344e-04	4.930e-04	7.445e-04
1.544e-03	-9.612e-04	-8.544e-04	1.829e-03	-6.245e-04
7.460e-04	-7.391e-04	2.090e-04	-3.638e-05	-1.011e-03
-2.727e-04	7.792e-04	7.032e-04	-8.062e-04	8.657e-05
-6.168e-04	-3.800e-04	1.837e-03	-3.594e-05	-7.495e-05
1.709e-04	6.313e-04	-2.767e-04	1.220e-03	-7.665e-04
-8.848e-04	-4.235e-04	2.643e-04	-6.473e-04	-2.737e-03
-1.969e-04	8.763e-04	1.750e-04	-9.915e-04	-2.034e-03

Table 2: *y*-axis raw data

For convenience, it has also been hosted at the following URL: `https://gist.github.com/sushisharkjl/0cf1123e83491caf36c193f36c9d3f2c`.

4   Analysis of Data

We begin by stating our hypotheses:

$$H_0 : \bar{x}_{\text{x axis}} = \bar{x}_{\text{y axis}}$$
$$H_a : \bar{x}_{\text{x axis}} \neq \bar{x}_{\text{y axis}}$$

By common nomenclature, H_0 denotes the null hypothesis, stating that the sample means between the x- and y-axes are equal. H_a denotes the alternative hypothesis, stating that the sample means are *not* equal. For the two-sample-mean t-test, we must check the conditions for inference:

1. Data is from a SRS. Because the data was generated by running a machine, we may assume that the data is a representative sample of all possible hole spacings from any time.

2. Independence. Due to technical specifications, it is impossible for any single hole to impact the spacings of other holes. This is because the machine operates on the principle of relative positions.

3. Nonreplacement. There is an infinite population of such hole spacings, and as a result, the 50 samples represented in either dataset is far less than 10% of the population.

4. Sampling distribution is approximately normal, as seen in the histogram. Additionally, the sample size is 50, which is greater than 30.

Creating boxplots of both distributions allows us to check for outliers, of which there are very few:

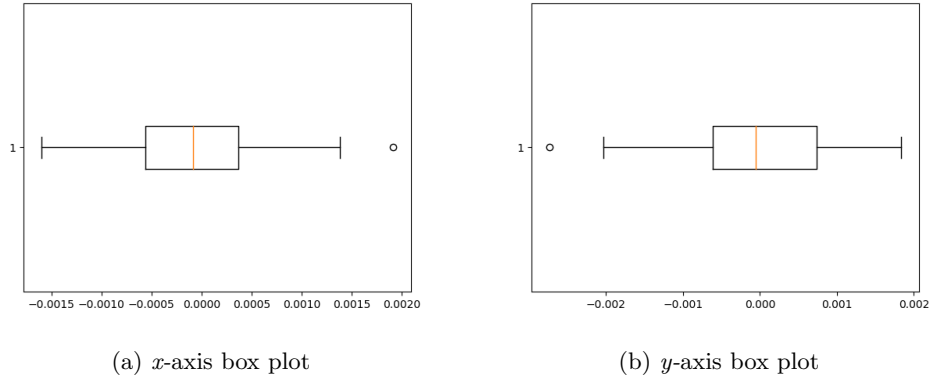(a) x-axis box plot   (b) y-axis box plot

Figure 2: Boxplots

The hypothesis test relies on the following test statistic:

$$t = \frac{\bar{x}_x - \bar{x}_y}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{n}}}$$

In order to analyse our data with a computer, we first format our data by placing one entry on each line of a new plaintext file. Then, we make use of `numpy`, a number-processing library of the Python programming language to obtain values for s_x, s_y, $\bar{x}_x$, and $\bar{x}_y$ with the following, assuming the x-axis data is in `xaxis.txt` and the y-axis data is in `yaxis.txt`:

```
import numpy as np
xData = np.loadtxt("xaxis.txt")
yData = np.loadtxt("yaxis.txt")

np.mean(xData)          # Returns mean of x-distribution
np.mean(yData)          # Returns mean of y-distribution
np.std(xData)           # Returns standard deviation of x-distribution
np.std(yData)           # Returns standard deviation of y-distribution
```

The obtained values are as follows:

$$\bar{x}_x = -0.0000671025$$
$$s_x = 0.000696156$$
$$\bar{x}_y = 0.0000186148$$
$$s_y = 0.000918202$$

Thus we find our test statistic by hand:

$$t = \frac{-0.0000671025 - 0.0000186148}{\sqrt{\frac{0.000696156^2}{50} + \frac{0.000918202^2}{50}}}$$

$$t = \frac{-0.000086}{\sqrt{0.00000000969266 + 0.0000000168619}}$$

$$t = -0.520730$$

We can then use `tcdf` on our calculators to find the p-value using the conservative degrees of freedom $n - 1 = 49$. On the TI-89 Titanium, this is done by using the Statistics application, and then inputting the key combination `F5 6` and entering the limits of the cdf calculation along with the appropriate degrees of freedom. This yields a $p = 0.302$.

The `scipy` statistics package in Python may corroborate our result with the following, if the x-axis data and y-axis data are in `xaxis.txt` and `yaxis.txt`, respectively:

```
import scipy as sp
import numpy as np

x = np.loadtxt("xaxis.txt")
y = np.loadtxt("yaxis.txt")

stats.ttest_ind(x, y, equal_var=0)      # Returns the test statistic
                                        # and two-tailed pvalue
```

5   Conclusion

From the previous section, we arrived at a p-value of $p = 0.302$. Because this p-value is greater than any of the commonly used confidence levels (namely 90%, 95%, and 99%, we do *not* have statistically sufficient evidence to reject H_0, the two population means are equal. As we fail to reject the null hypothesis, we can state that this dataset does not provide conclusive proof that the x- and y-axes have different amounts of variability.