Speed-cubing Champion versus Prodigy, Investigated with Statistics

Tyler Jordannio Krisanto

April 2026

Abstract

Hello, fellow reader! Thank you for taking the time to tune in and have a look at my little passion project. This is the convergence of two passions that have defined who I am: mathematics and a niche hobby called *speed-cubing.* In essence, speed-cubing is the act of solving the Rubik's Cube in the least possible amount of time. Official competitions are those under the *World Cube Association(WCA)*; only about 280,000 people in the world have competed in one before, myself included! Out of all competitions, there is one that everyone looks out for: the Rubik's WCA World Championship, held every 2 years. This paper aims to use statistics to investigate who will claim the renowned title of World Champion next.

1 Introduction

On one side sits **Yiheng Wang,** the reigning World Champion, who won the most recent World Championship in 2025 with an incredible *average* time of 4.23 seconds. On the other side sits **Xuanyi Geng,** who came second place with an average time of 4.49 seconds, but now going on an unbelievable trajectory and showing no sign of stopping. Could he steal the crown away from Yiheng? Let's find out. [1]

average: In an official WCA average, each competitor gets five solves. The slowest and fastest times are ignored and the mean of the middle three is taken to get the average.

2 How will we extract this data?

We will consider all of Yiheng and Xuanyi's averages, going from September 9th 2023, the date of Xuanyi's first competition (Xuanyi began competing first), to March 30th 2026, the time the writing of the paper and calculation of statistics began. (These kids were literally competing as I wrote, the data used in the paper is bound to be outdated by now!) (Yiheng's Results: [2] Xuanyi's Results: [3])

3 The General Outline

3.1 Determining a Trend

We start by determining and working out an appropriate trend for the averages of each competitor, relative to the number of official rounds they have competed in from the starting point (September 9th 2023), to get an estimate of the future mean and standard deviation of their averages. For the mean, these are the most sensible two choices:

- **Exponential Decay Model** ($y = a + be^{-cx}$): This model shows a slowdown in decrease in averages which is realistic for a skill like speed-cubing. Its decrease rate is fast early on but not too slow later, which is why it is preferable over models such as the logarithmic model. In addition, there is an asymptote (denoted as a in the equation) that y approaches as x approaches infinity. This is to represent the human limit of the 'stable' average for Yiheng and Xuanyi. Let's estimate it as **3.5 seconds**; the world record average now (3.84) is just slightly slower.

- **Power Series Model** ($y = a+bx^{-c}$): This model, just like the previous one, shows a diminishing rate of change and has an asymptote denoted as a. However, its decrease in rate of change is not as aggressive as the previous model. To illustrate this, we will differentiate both equations:

$$\frac{d}{dx}(a+bx^{-c}) = -cbx^{-c-1} = -\frac{c}{x}bx^{-c}$$
$$\frac{d}{dx}(a+be^{-cx}) = -cbe^{-cx}$$

Notice that the power series model has a '1/x' component to its derivative multiplied with the initial term (ignoring a) unlike the exponential decay model whose derivative is just a constant multiple of the initial term (without the a). This makes the power series model decay much more slowly than the exponential decay model, with a very large value of x (in the hundreds)

They both reflect diminishing returns and improvement over time while having an element representing a theoretical limit, but I will select **exponential decay** for the faster (and more realistic) rate of improvement it shows.

Then, there is the standard deviation (SD). This metric will normally increase or decrease with mean, but in the very top level, inconsistencies are decided by the same factors regardless of how fast they specifically are: possible turning errors, solution errors, *+2 penalties*, etc. Hence, the SD, in this case, does not follow a true trend and we will just take the average of its *moving averages* over a certain time period.

+2 penalties: In the WCA, +2 penalties can occur for a competitor breaking certain rules, but it is most commonly caused by the cube being one move away from being solved.

moving average: An n-point moving average is an average of a selection of n consecutive values in a dataset.

3.2 Testing for a Distribution

We will then attempt to reasonably estimate the averages Xuanyi and Yiheng get in one occasion as a distribution for each competitor, particularly the normal distribution for simplicity purposes. We will test whether the normal distribution is a reasonable fit using the **Shapiro-Wilk Method.** It is regarded as one of the strongest tests for normality, and can be used for samples up to size 2000 with Patrick Royston's approximations in 1992. The whole dataset for each competitor is bound to be far from a normal distribution because they will have a frequency of high anomalies (likely their first averages) but no low anomalies at the same extremity. Hence, we shall only use this test for the averages over the **last six months**. The process below shows an overview of the Shapiro-Wilk test.

The hypotheses are set: H_0(null hypothesis): The data is normally distributed H_1(alternative hypothesis): The data is not normally distributed

All the data is sorted in ascending order (lowest takes index 1) and a W-value is calculated using the following formula:

$$W = \frac{(\sum_{i=1}^{n} x_i a_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where n is the sample size, x_i is the data item at index i, $\bar{x}$ is the mean of the dataset, and a_i is a specific Shapiro Wilk coefficient to be determined, which changes with sample size. Royston found approximations to determine the specific Shapiro Wilk coefficients, but they are too complicated to cover in this paper.

Next, Royston also discovered that $ln(1-W)$ approximately follows a normal distribution with mean

$$\mu = -1.5861 - 0.31082y - 0.083751y^2 + 0.0038915y^3$$

and standard deviation

$$\sigma = exp(-0.4803 - 0.082676y + 0.0030302y^2)$$

where $y = \ln n$

With this information, we can calculate our test z-value with the standardisation formula:

$$z = \frac{ln(1-W) - \mu}{\sigma}$$

Then, our p-value is the probability that a variable under a normal distribution with mean 0 and standard deviation 1 exceeds this value. In other words,

$$p = 1 - \phi(z)$$

where

$$\phi(z) = P(Z \leq z), Z \sim N(0, 1)$$

We will finally compare this to a significance level, commonly 0.05, which we choose. If the p-value is smaller than the significance level, we reject the null hypothesis, concluding that the data is very far from fitting a normal distribution. Otherwise, we accept the null hypothesis and conclude that the data can be reasonably modeled by a normal distribution.

This process is outlined in [HT15]. The approximate polynomials are outlined in [Exc24].

3.3 Comparing the Variables

If the variables can be reasonably modeled using a normal distribution, we can model the difference of these variables as a normal distribution as well, by the linear properties of normal random variables. If X and Y are independent and normally distributed, where

$$X \sim N(\mu_1, \sigma_1^2)$$

$$Y \sim N(\mu_2, \sigma_2^2)$$

Then we have:

$$X - Y = D \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

From here, through the standardisation formula, we can work out the probability that $X - Y$ is smaller or greater than 0, to obtain the probability that X or Y is smaller.

4 The Candidates

4.1 Yiheng Wang

We start by extrapolating for the number of rounds. Yiheng Wang will have competed in by the World Championship 2027. Over the last six months (30/9/25 - 30/3/26), he has competed in a total of 44 rounds. The World Championship always occurs at some point in July or August (See some past World Championships here: [4][5][6]), hence a prediction 15 months into the future (exactly 15 months from 30/3/26 is 30/6/27) is sufficient. He has already competed in 251 rounds from 9/9/23, so using linear extrapolation, the number of rounds he will have competed in is estimated to be about:

$$251 + 44 * \frac{15}{6} = 361$$

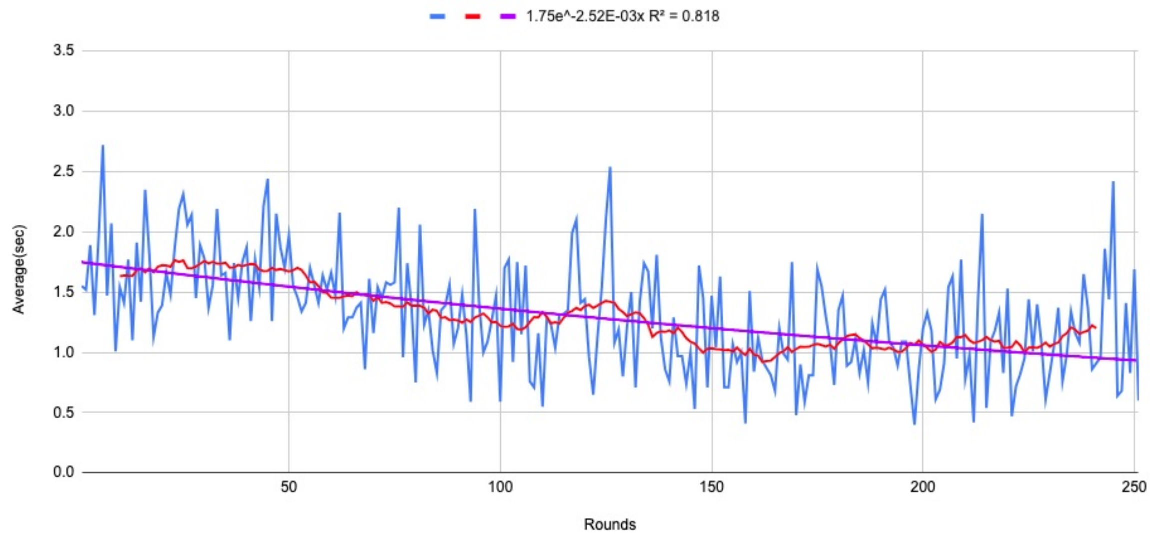

Figure 1: The blue line represents Yiheng's averages from Singapore Mini (9/9/23) to today (30/3/26) but I subtracted them all by 3.5 (the limit established) as we are using an exponential trend line. The red line represents the 20-point moving averages; this is done as individual averages can be very volatile. The purple line is the line of best fit, based on the red line. R^2 is the coefficient of determination; its value of 0.818 is high for many standards, this context included (A value of 0.7-0.8 is often considered sufficient [Kum16]), so the exponential model is a reasonably strong fit for this data.

By the World Championship 2027, we predict Yiheng's mean to be:

$$3.5 + 1.75 * e^{-0.00252*361} \approx 4.2046(4dp)$$

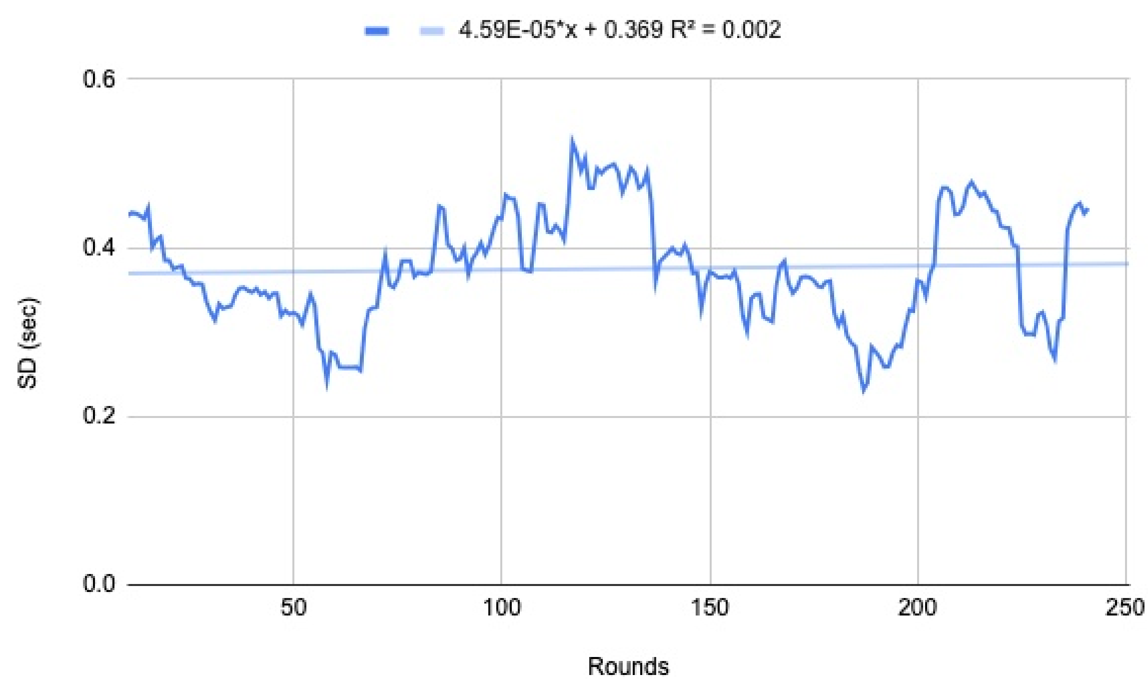

Figure 2: This is the graph of 20-point moving SDs over rounds.

Using the average moving SDs over the past year (the reason why one year is chosen will be explained later) yields a standard deviation of **0.3589** to 4 decimal places.

Checking for Normality
Beginning with the Shapiro-Wilk test, to obtain the values, I manually took them from the table of Royston's a-coefficients as an aid to the calculation. Manually calculating through, I obtained a p-value of around **0.0123** to 4.d.p, which **rejects** our null hypothesis that the data can be modeled by a normal distribution as it is smaller than 0.05. (The table of approximate coefficients is in [HT11]. NOTE: You might not be able to open the file as it is not set in the https protocol.) However, the a-coefficients used in the manual calculation are just approximations, and the p-value is bound to be very volatile to even small differences in the W test statistic; the calculation of z involves $ln(1-W)$ and the calculation of p from z is defined exponentially, as with the PDF of a normal distribution, which is defined as:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$$

for a normal distribution of mean 0 and SD 1.

Finally, the data does contain anomalies that naturally make the data appear skewed. (From the last 6 months, his worst results were 5.92 and 5.65, WAY off his projected (and current) mean.) Therefore, we, in this case, can roughly accept the null hypothesis, ignoring these anomalies.

4.2 Xuanyi Geng

We similarly extrapolate for the number of rounds Xuanyi Geng will have competed in by the World Championship in 2027. Over the last six months (30/9/25 - 30/3/26), he has competed in a total of 40 rounds. He has already competed in 205 rounds from 9/9/23, so the number of rounds he will have competed in is estimated to be about:

$$205 + 40 * \frac{15}{6} = 305$$

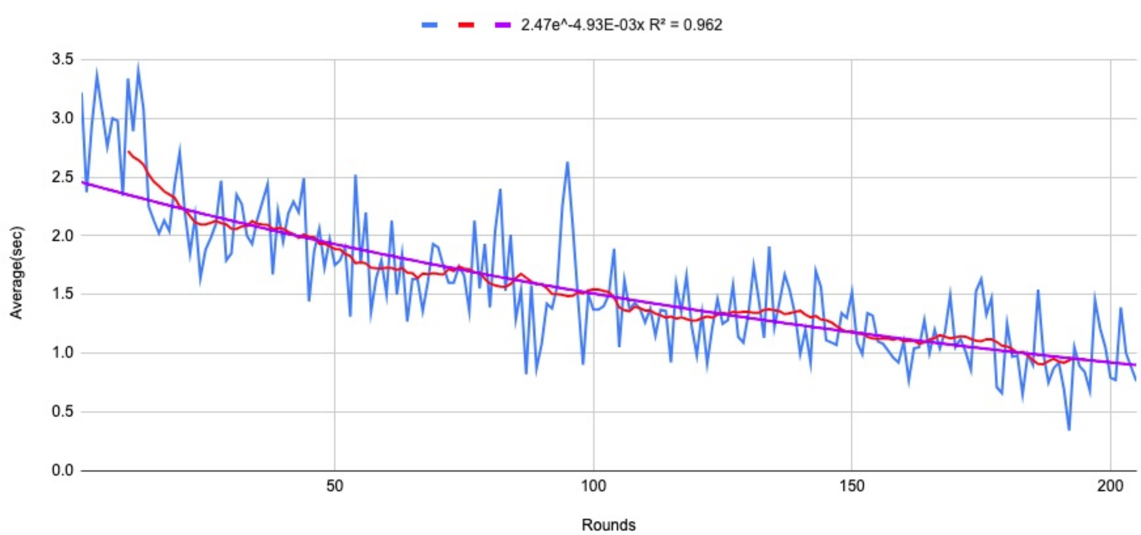

Figure 3: The graph follows the same format as with Yiheng's. The coefficient of determination of the trend line is much higher than with Yiheng's ($0.962 > 0.818$) so the exponential model extremely accurately fits Xuanyi's data.

We predict Xuanyi's mean to be:

$$3.5 + 2.47e^{-0.00493*305} \approx 4.0491(4dp)$$

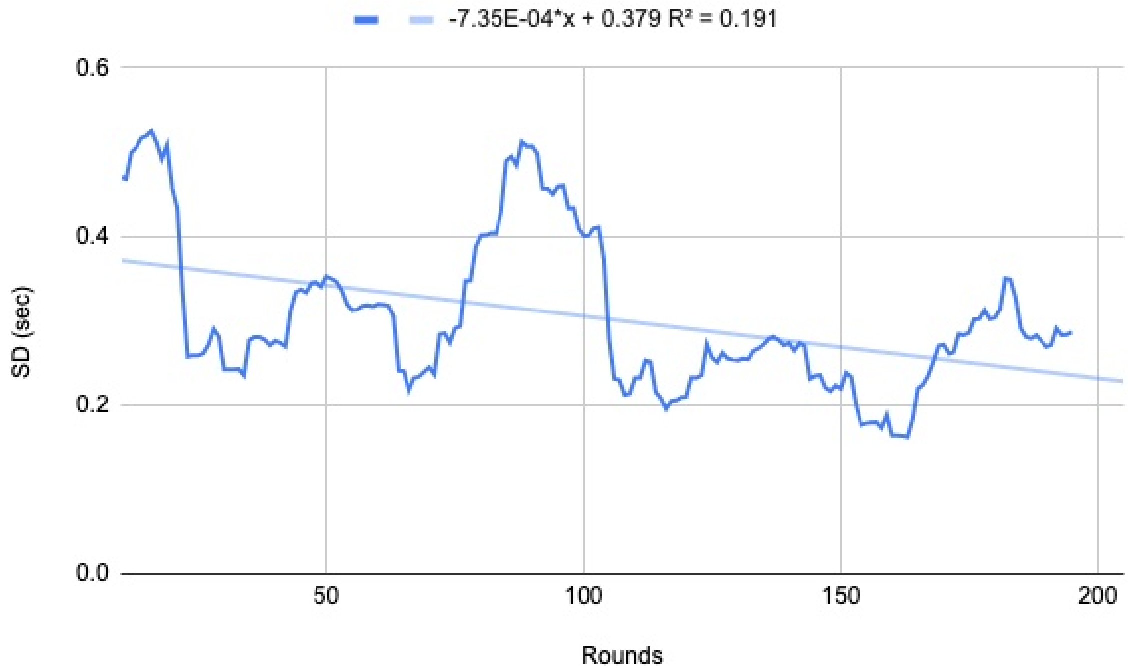

Figure 4: Again, this is the graph of 20-point moving SDs against rounds.

Notice that from around the 100 round mark, his peak moving SD drops to lower than its previous peak prior, so this is why we'll just take the average of the moving SDs over the past year; results from the past year start at round 107 (We also do it with Yiheng for consistency purposes.) Using this method, we predict Xuanyi's SD to be **0.2499** to 4 d.p.

Checking for Normality
Doing the Shapiro-Wilk test manually (same method as before), I obtained a p-value of **0.2832** to 4.d.p. Therefore, we can assert with strong confidence that the p-value is larger than the significance level of 0.05 and accept the null hypothesis.

5 Comparing the Two

Variable	Mean	Variance
X	4.0491	0.2499^2
Y	4.2046	0.3589^2
X-Y	-0.1555	$0.2499^2 + 0.3589^2$

Table 1: We shall model Xuanyi's averages as variable X and Yiheng's averages as variable Y. This is a recap of what we have predicted for their means and variances (square of the standard deviation)

The probability that Xuanyi wins is equal to:

$$P(X < Y) = P(X - Y < 0) = P(Z < \frac{0.1555}{\sqrt{0.2499^2 + 0.3589^2}}) = \phi(0.35556...) \approx 0.6389 = 63.89\%$$

6 Interpretation and an External Factor

Even though Xuanyi, statistically, is the favourite to win the championship, it is still not so obvious; Yiheng winning ($\approx 36\%$) would not necessitate some kind of miracle. Besides, this model is not perfect and there are still external factors to consider, particularly those affecting independence, since it is a crucial assumption made in the final probability calculation.

One of these factors is the **pressure and nerves** of such a large championship; the two are, after all, sharing the same stage, facing the same pressure and stakes. This is a major blow against independence, but this doesn't mean the predictive models I've established can't be relevant. This is mainly due to the fact that Yiheng and Xuanyi deal with pressure very similarly in large championships. For example, both participated in the Asian Championship 2024 and World Championship 2025; and importantly, both were totally unfazed under the pressure, getting results that were on par, or even above our expectations. Therefore, it will probably still come down to who is fundamentally faster.

7 Conclusion

The World Championship 2027 will determine whether Yiheng Wang will win and solidify himself as an unstoppable force, or Xuanyi Geng be crowned the new champion. It isn't obvious who's going to take the crown. But what's obvious is that with these two, and possibly more dark horses to win it all, the World Championship 2027 Final will be the most electric final in the history of the sport.

Thank you for reading until the very end:)

References

[1] 1. Rubik's WCA World Championship 2025 — World Cube Association.

[2] 2. Yiheng Wang — World Cube Association.

[3] 3. Xuanyi Geng — World Cube Association.

[4] 4. Rubik's WCA World Championship 2023 — World Cube Association.

[5] 5. WCA World Championship 2019 — World Cube Association.

[6] 6. World Rubik's Cube Championship 2017 — World Cube Association.

[Exc24] Real Statistics Using Excel. Shapiro-Wilk Royston Test — Real Statistics Using Excel, 9 2024.

[HT11] Zofia Hanusz and Joanna Tarasińska. TABLES FOR SHAPIRO-WILK w STATISTIC ACCORDING TO ROYSTON APPROXIMATION. *Colloquium Biometricum*, 41, 1 2011.

[HT15] Zofia Hanusz and Joanna Tarasińska. Normalization of the Kolmogorov–Smirnov and Shapiro–Wilk tests of normality. *Biometrical Letters*, 52(2):85–93, 12 2015.

[Kum16] Sarraf R. Kumar. What is the acceptable r-squared value? — ResearchGate, 9 2016.