



# Jose Camacho Collados

1K Followers · About Follow

## The gender gap in top-level chess



Jose Camacho Collados 5 days ago · 10 min read

### *A critique and complement to “What gender gap in chess?”*

The gap between women and men in chess has been a long-standing debate subject. There have been many studies attempting to explain the seemingly large gap between men and women, both in their participation and strength at the top level. This old debate has reopened in the past weeks with an [article in Mint](#) attempting to explain the low presence of women at the top level, and a more recent Chessbase article titled [“What gender gap in chess?”](#) that has confronted their main claims with a new statistical perspective. While this statistical framing is not entirely new and is largely a replication of the study of [Bilalić et al. \(2008\)](#), its conclusions have had a significant impact on [social media](#).

These latter studies argue that the gap in strength (at least with regards to the top player of each country) is largely due to the low number of women chess players. The recent article of Prof. Ma in Chessbase explains this by performing a statistical experiment with Indian chess players, showing that the gap difference between the top players of the country was largely due to chance. This article is very well explained so I would recommend all readers to take a look at the [original piece](#) first.





Source: [ajedreznazari.es](http://ajedreznazari.es)

In general, I largely agree with the sentiment and reasoning of Prof. Ma's article, although not with its condescending tone. Nonetheless, in my opinion, not many conclusions can be drawn from their single experiment, which does not accurately reflect the global gender gap. Disregarding some methodological decisions and caveats about the experimental setting [1], in this article I provide extended experimentation to help answer the main question of whether women's lower strength at the top level is only due to their low participation. For this, I followed the same methodology as in the original article, considering only the non-junior players of the October 2020 FIDE list (including inactive players). In addition to replicating Prof. Ma's experiment for Indian players, I extended it to the top 20 federations in the world according to FIDE and introduced a different "top strength" measure that may be less affected by outliers. Additionally, our analysis provides interesting insights into the current state of each country with respect to the chess gender gap.

## The experiments

In the following, I present the experimental details. I have tried to make it as understandable as possible, but you can skip directly to the summary of the experiments at the end if you rather have a light reading.





Source: Unsplash, Randy Fath

The whole motivation of the experiment lies around the idea that we cannot simply compare the level of the strongest male and female players, as obviously the women are handicapped due to their lower participation numbers. In this case, it is much more meaningful to compare the actual gap between the top male and female players (what we will refer to as the “**observed**” value) and the gap we would expect statistically given their participation numbers (the “**expected**” value). In addition to comparing only the top player which can be sensitive to outliers, I have included the average of the top ten rated players of a federation as a compare the strength between men and women (this is the metric used by FIDE itself to compare the strength between different federations). To compute the expected value we simply run 100,000 random permutation tests (again I would recommend [Prof. Ma’s article](#) where all the details are clearly explained).





















In addition to computing the expected values, with the permutations we can compute the **p-value**, which is a statistical measure that can tell whether the “observed” difference is statistically significant or not. In our case, the p-value is computed as the



number of times the “observed” gap is lower than the “expected” gap, divided by the total number of permutations (i.e. 100,000). Usually, a p-value less than 0.05 is considered to be statistically significant.

### The results

The main results of our experiments are summarized in the following table.

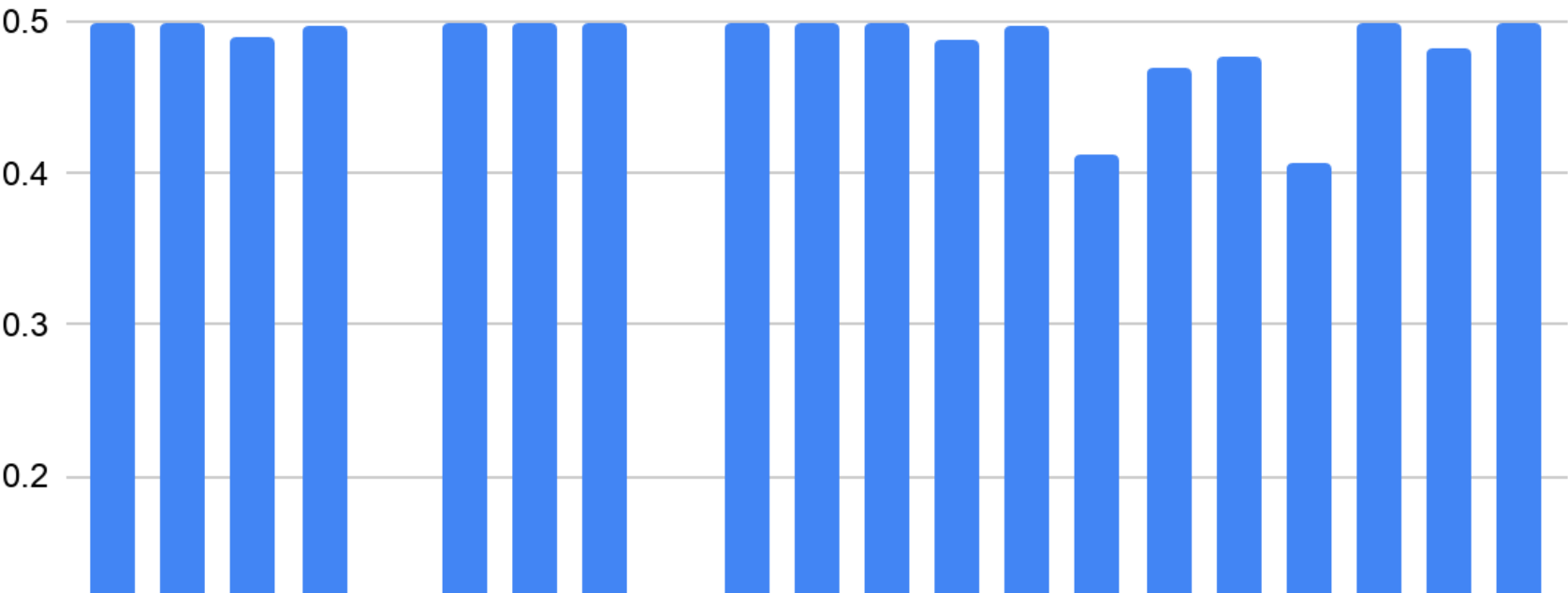
Federation	Number of players		Gap "top player"		Gap "top 10 players"	
	Men	Women	Expected	Observed	Expected	Observed
 RUS	17866 (87.5%)	2559 (12.5%)	72	<b>230*</b>	118	<b>259*</b>
 USA	4677 (95.8%)	204 (4.2%)	216	<b>399*</b>	261	<b>357*</b>
 CHN	393 (67.9%)	186 (32.1%)	34	<b>133*</b>	61	<b>180*</b>
 IND	17899 (93.9%)	1165 (6.1%)	153	<b>167</b>	264	<b>269</b>
 UKR	2872 (88.5%)	374 (11.5%)	36	<b>143*</b>	123	<b>245*</b>
 ARM	511 (85.0%)	90 (15.0%)	105	<b>360*</b>	197	<b>384*</b>
 AZE	614 (82.4%)	131 (17.6%)	106	<b>388*</b>	174	<b>352*</b>
 HUN	4865 (93.0%)	364 (7.0%)	<b>148</b>	84	229	<b>234</b>
 FRA	17310 (93.4%)	1228 (6.6%)	187	<b>346*</b>	225	<b>349*</b>
 POL	7049 (88.8%)	888 (11.2%)	135	<b>342*</b>	153	<b>276*</b>
 GER	22595 (94.8%)	1249 (5.2%)	86	<b>200*</b>	163	<b>296*</b>
 NED	3938 (95.3%)	193 (4.7%)	206	<b>381*</b>	254	<b>344*</b>
 ESP	19895 (96.3%)	774 (3.7%)	190	<b>300*</b>	222	<b>320*</b>
 ENG	2828 (95.2%)	142 (4.8%)	185	<b>302</b>	282	<b>345</b>
 ISR	1756 (95.4%)	85 (4.6%)	143	<b>349*</b>	260	<b>348*</b>
 CZE	6362 (95.7%)	288 (4.3%)	202	<b>305</b>	244	<b>322*</b>
 NOR	2667 (96.4%)	99 (3.6%)	413	<b>655*</b>	361	<b>439</b>
 CRO	2598 (94.5%)	152 (5.5%)	140	<b>306*</b>	234	<b>372*</b>
 GEO	531 (74.1%)	186 (25.9%)	52	<b>162*</b>	75	<b>153*</b>
 ARG	3395 (96.5%)	123 (3.5%)	154	<b>284*</b>	252	<b>375*</b>
WORLD	236138 (92.8%)	18401 (7.2%)	91	<b>188*</b>	111	<b>211*</b>

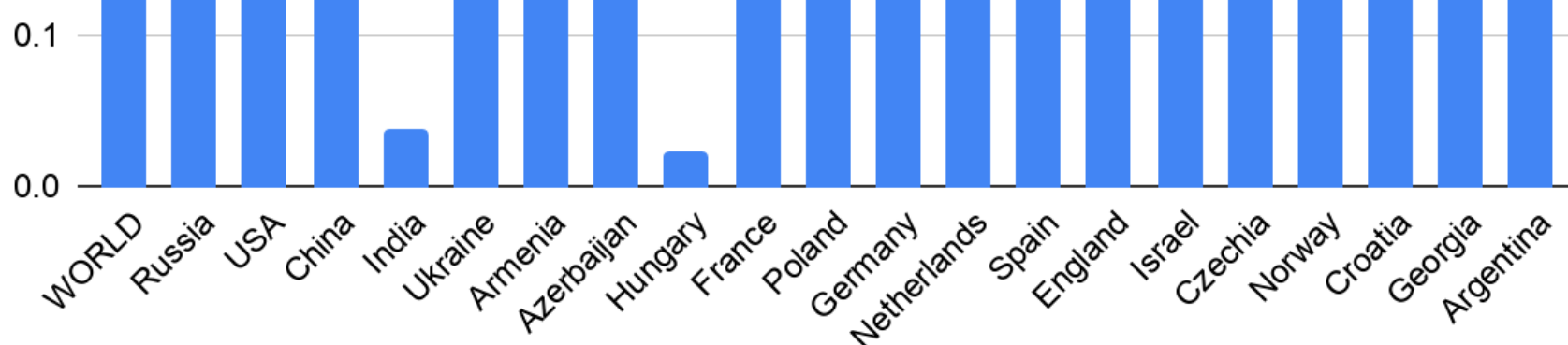
Gender gap in top-level chess according to our experiments. Bold numbers indicate the highest number being expected and observed values. \* is added to those results that are statistically significant with p-value<0.05.

If we check the top player metric as in Prof. Ma’s experiments, we can see how we could replicate the Indian experiment where the observed gender gap is indeed not statistically significant, as claimed in the original article. However, this gap is statistically significant in 16 of the 20 federations considered, and if we take the whole world as a reference.

Moreover, **according to the top 10 players, men are statistically stronger than women in all countries** (in all except Hungary if we only take the top player as reference). The difference is statistically significant in all twenty countries but India, Hungary, England and Norway (the latter two with a difference that is close to being statistically significant). By checking the whole world, the tendency is even more clear, as the difference is statistically significant with a p-value very close to 0 ( $\sim 0.01$ ). This may suggest that in federations with a chess tradition, women are relatively closer to men in strength. However, there does not seem to be any correlation between relative or absolute participation of women in a country and their top-level strength, so we may need to find the reasons elsewhere.

To make it more visual, the following bar chart indicates how big the gap is for the twenty federations analyzed and the world chess population as a whole (with 0 indicating no gender gap) [2]. **The only two countries that are close to full parity in top-level chess (statistically speaking) are Hungary and India.** As for when we consider all non-junior FIDE registered players in the world, this gap is even higher than in most of the twenty federations considered.





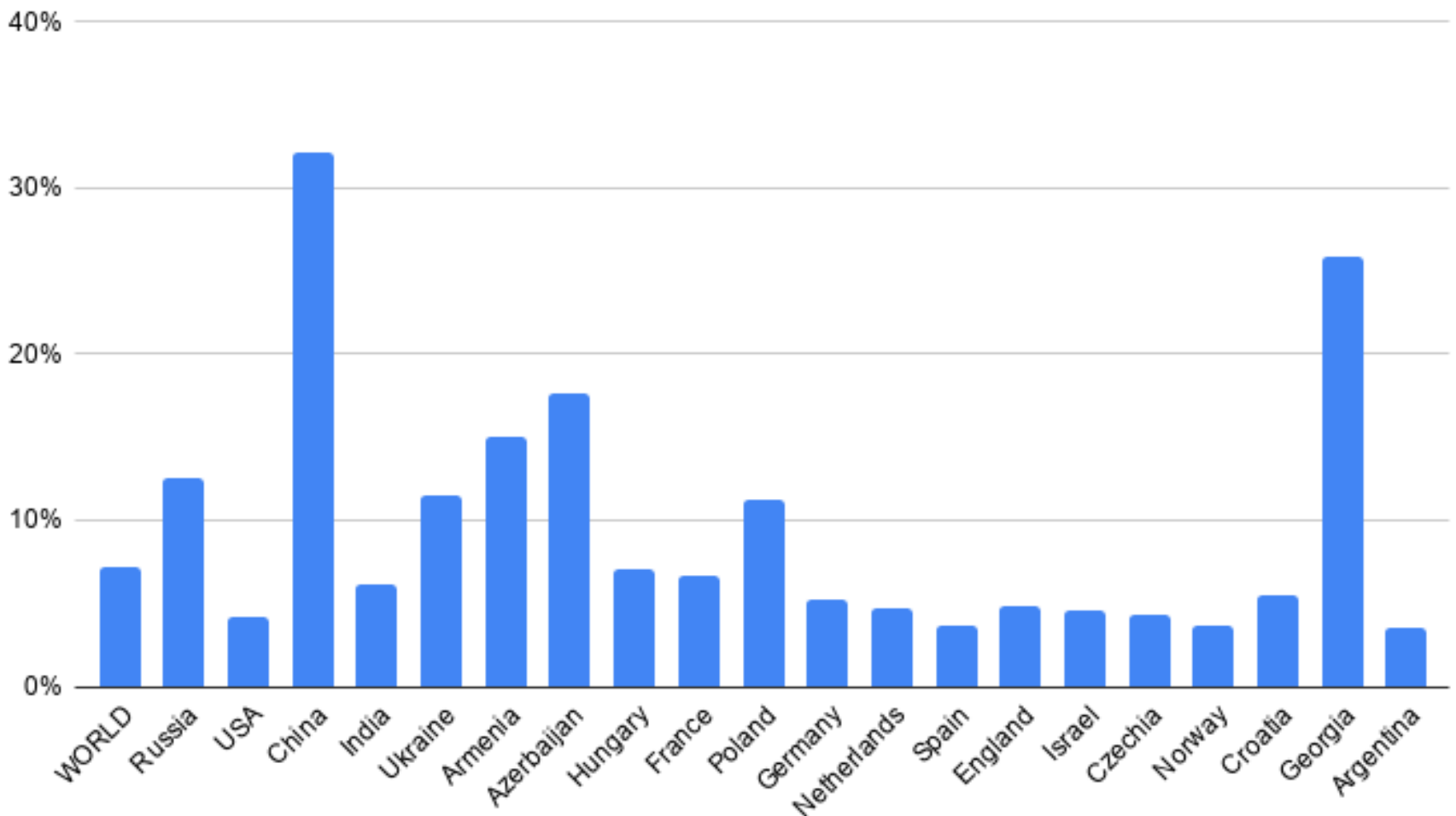
Statistical gender gap with respect to the average strength of the top 10 male and female players across the world and in the top 20 chess federations. Y-axis corresponds to 0.5 minus p-value, where the closer to 0.5, the more perceived difference between men and women strength, with 0.0 indicating gender equality.

### Some interesting findings:

- **Women and men in Hungary and India have a very close strength statistically**, both in terms of their top player (thanks to Judit Polgár, Hungary even surpasses men statistically) and the average rating among the top 10. This relates to what one could expect in Hungary with the Polgar sisters raising the level of women chess for many years (reminder that inactive players were also considered in the experiments), and in India with increased chess popularity and two women (Humpy Koneru and Dronavalli Harika) in the current world top 10.
- In addition to the experiments measuring the top-level strength, we also computed the overall Elo average of both men and women. In this case, men are still better than women in 17 of the 20 analyzed countries (all but Georgia, India and Azerbaijan), but also show that **Georgian and Indian women are (statistically) significantly better than men on average**. While this measure does not indicate the strength of the top players, it highlights some clear trends in these countries with respect to women's chess. This overall strength and high women participation numbers may also be the reason for the recent success of Georgia in women team competitions, winning medals in their last three major team events between 2017 and 2019, including the World Team Chess Championships and Chess Olympiad.
- We note that our **conclusions about the German gender gap differ from those of the study of Bilalić et al. (2008)** [3]. As mentioned at the beginning of the article, this can be due to several methodological factors such as the somewhat

arbitrariness of the Elo system and its fluctuations, or to some issues to their experimental setting, as pointed out by [Knapp \(2010\)](#).

- By giving some context and looking at the relative participation figures of each federation, we can note some marked differences across countries, as shown in the following chart.



Women participation percentage with respect to the total number of players across the top 20 chess federations and the world.

For example, in **China and Georgia women represent over 25% of all the chess players** in their respective countries. This reflects a tendency of the recent successes in women's chess for China (e.g. winning the last two Olympiads) and increased popularity of chess among women in Georgia. However, in other countries such as Argentina, Norway and Spain, female participation is lower than 4%.

## Conclusions

To sum up, in this piece we show that the **gender gap in top-level chess cannot be**

**statistically explained by the lower participation of women alone.** Even if this hypothesis cannot be proved statistically, it is clear that the women low participation comes with other associated disadvantages as well, but this was out of the scope of this article. So, what are then the main questions answered by this piece? Unfortunately, there are still more open questions than answers...

**Can we say that the gender gap in top-level chess can be attributed to the low number of women competing?**

No, or at least this is what the numbers say (Figure 1 is quite explicit in that respect). Nonetheless, it seems clear that the low participation of women in competitive chess agglutinates other related issues (e.g. lack of role models, lack of motivation, hostility, etc.) but these are not measured in our statistical experiment. Following Prof. Ma's setting, our experiments take into account the relatively low presence of women in chess, but this factor alone does not explain the low presence of women in the highest ranks of the countries analyzed.

**Can we say that men are “statistically” better than women in chess?**

Not at all. We can only say that, according to this very specific experimental setting, men appear to be better than women in top-level chess even if we account for relative participation figures. This is just a very limited simulation, with many caveats such as the usage of Elo as an absolute measure of strength, and many other factors that have not been taken into account in the model.

**But Jose, I'm a bit confused, could you please then clarify if the gender gap in top-level chess is due to participation numbers, genetics, or other social factors?**

Unfortunately, we cannot answer this question based on these experiments, and I'm definitely not an expert on the topic to even attempt to give an answer. I haven't been interested in the biological differences to explain this gap (I cannot say anything about it as I don't know about the area), but if I'm If I'm forced to give my opinion, this would be related to previous sociological studies published in the topic: there are certainly systemic societal factors that could easily explain this gap between men and women,



which may indeed be associated with the participation numbers. For this, I can also relate to my experience as a semi-professional player, where I have witnessed and am aware first-hand of some of the difficulties women face in a male-dominated chess world. That said, I would again recommend reading articles by experts (e.g. [Maass et al. \(2008\)](#); [Bilalić et al. \(2008\)](#); [Gerdes and Gränsmark, 2010](#); [Backus et al. 2016](#); [Cubel, 2017](#); [Dilmaghani, 2020](#)) [4] for a better understanding of the subject. These sources should be trusted more than my simple non-expert and biased opinion.

To sum up, the gender gap in chess is a quite complex issue and there have been multiple studies without any definite answer. This is again the case, as this piece raises more questions than answers. Finally, if I can give advice as a mathematician and as a researcher, I would be wary about drawing bold conclusions from simple data-driven experiments in a complex social issue — there are often many ways to manipulate the numbers to prove/disprove what we want using statistical methods [5]. Instead, to get closer to the answer and increase the participation of women in chess, we would need an interdisciplinary approach involving areas such as sociology, gender studies, psychology and data science. But please don't take me wrong, I believe these types of statistical experiments are extremely useful and add valuable data points into the discussion, only that we should also be careful of what we can (and cannot) conclude from them.

**Note:** This piece has been reproduced in chess24, both in [English](#) and [Spanish](#).

### [Footnotes]

*[1] One methodological concern is due to how “chess strength” is measured with the Elo rating system, which may appear to be often arbitrary and fluctuates over time. For example, less than twenty years ago the minimum Elo was 2000, while now it is 1000. This and other issues certainly affect the conclusions of any experiment based on the whole distribution of chess players.*

*[2] In this experiment we use the p-value as a comparison among different distributions. Note that this is often not an optimal strategy and that the effect size should be accounted for to provide a more accurate comparison. Nonetheless, the visualization is provided as a*

*further evidence of the marked differences across countries, and not to compare individual scores.*

*[3] I also tried the top 100 players as in the original paper of Bilalić et al. (2008) and the conclusions still differ.*

*[4] Disclaimer: Even this selection may be biased (since it was based on a quick literature search by me) and I may have missed more relevant articles on the subject.*

*[5] Note for example that even for this simple experiment we already took many seemingly small decisions (e.g. including inactive players, excluding junior), and assumptions (e.g. considering the whole distribution of FIDE-rated players, assuming random distributions in the samples). Had we taken other decisions/assumptions, the outcome of the results might have been different.*

Chess

Gender Gap



[About](#) [Help](#) [Legal](#)

Get the Medium app

