



ELSEVIER

Available online at [www.tylervigen.com](http://www.tylervigen.com)



# Batting Around the Numbers: A Statistical Analysis of the Relationship Between Chicago Cubs' Total Runs Scored and New York Mets' Runs Scored

Chloe Hart, Aaron Travis, Gavin P Tompkins

Center for Scientific Advancement; Austin, Texas

## KEYWORDS

"Chicago Cubs total runs scored," "New York Mets runs scored," "statistical analysis baseball runs scored," "correlation coefficient baseball teams," "relationship between Chicago Cubs and New York Mets runs scored," "baseball statistics comparison teams," "impact of runs scored on baseball teams."

---

## Abstract

In this paper, we swing for the fences as we delve into the relationship between the total runs scored by the Chicago Cubs and the runs scored by the New York Mets. Utilizing data from the vast playing fields of Baseball Reference and Baseball-Reference.com, our research team has uncovered a striking correlation coefficient of 0.7687085 with  $p < 0.01$  from the years 1975 to 2022. The findings of this study not only help to bridge the gap between the world of statistics and the diamond, but also provide valuable insights into the interconnectedness of these two storied franchises. So, grab your scorecards and rally caps as we knock it out of the park with the results of this home-run analysis!

Copyright 2024 Center for Scientific Advancement. No rights reserved.

---

## 1. Introduction

Step up to the plate, baseball enthusiasts, as we embark on a grand slam exploration into the world of statistical correlations, where the Chicago Cubs and the New York Mets go head-to-head in a mesmerizing dance of runs scored. As the saying goes,

"You can't tell the players without a scorecard," and indeed, we bring forth a scorecard that not only tracks the numbers but unveils the intricate relationship between these two titans of the National League.

While baseball may be a sport steeped in tradition, our research aims to infuse a fresh perspective by applying rigorous statistical methodologies to unveil the underlying patterns in the runs scored by these illustrious teams. As we unravel the tapestry of hits, runs, and errors, we can't help but marvel at the "hidden ball tricks" that the numbers play on us, offering unexpected twists and turns in their statistical saga.

The Chicago Cubs and the New York Mets, like longtime rivals in a heated pennant race, have seen their fortunes ebb and flow over the decades. From the ivy-covered walls of Wrigley Field to the skyscraping home runs at Shea Stadium and Citi Field, these teams have left an indelible mark on the annals of baseball history. Our study seeks to uncover the statistical undercurrents that intertwine the runs scored by these teams, offering a delightful play-by-play account of their interconnected batting exploits.

So, dust off your baseball caps, and get ready to slide into the world of numerical analyses that is anything but a curveball. Let's swing for the statistical fences and knock this study out of the park!

## 2. Literature Review

In their seminal work, "Statistical Analysis of Major League Baseball Teams: A Comprehensive Study," Smith and Doe undertook a rigorous examination of the relationship between runs scored by the Chicago Cubs and the New York Mets. This study laid the groundwork for our investigation by highlighting the significance of identifying patterns in the offensive prowess of these iconic teams. The authors find significant evidence pointing to a positive correlation between the total runs scored by the Chicago Cubs and the runs scored by the New York Mets, setting the stage for our own exploration of this compelling statistical battleground.

Building on this foundation, Jones et al. further delved into the batting statistics of National League Central and East Division teams in "Advanced Statistical Methods in Baseball Research." Their findings echoed the sentiments put forth by Smith and Doe, emphasizing the need for a thorough analysis of the relationship between the runs scored by these teams over time. The authors conclude that a deeper understanding of this connection can provide invaluable insights into the strategic and tactical maneuvers that shape the outcomes of baseball clashes.

As we segue from the serious and stoic world of academic literature to a more eclectic landscape, we turn to non-fiction works that intersect with the realm of baseball statistics. "Moneyball: The Art of Winning an Unfair Game" by Michael Lewis is a poignant exploration of the unconventional methodologies employed by the Oakland Athletics to achieve success in Major League Baseball. While the focus is not directly on the Chicago Cubs or the New York Mets, the underlying principles of statistical analysis and player performance provide a rich backdrop for our own inquiry.

Similarly, "The Book: Playing the Percentages in Baseball" by Tom Tango, Mitchel Lichtman, and Andrew Dolphin offers a treasure trove of statistical insights that permeate the fabric of our study. Through a lens of probability and numerical acumen, the authors dissect the nuances of baseball strategy and decision-making, illuminating the interconnected nature of statistical analyses within the sport.

Turning now to the realm of fiction, we pause to consider titles such as "The Art of Fielding" by Chad Harbach and "The Natural" by Bernard Malamud. While these literary works might not directly address the statistical relationships between baseball teams, their evocative narratives transport readers to the heart of the game, where the

crack of the bat and the rush of adrenaline encapsulate the essence of baseball.

In expanding our horizons, we also engage with television programs that offer glimpses into the world of baseball dynamics. Shows such as "Pitch" and "Brockmire" entertain and inform, presenting fictional yet insightful portrayals of the intricate web of relationships, rivalries, and triumphs within the realm of sports.

With our scholarly pursuits taking us to the diamond and beyond, we now proceed to unveil the results of our empirical investigation into the connection between the total runs scored by the Chicago Cubs and the runs scored by the New York Mets, navigating the wondrous intersections of statistics and America's favorite pastime.

### **3. Our approach & methods**

To embark on our statistical journey through the batting averages and runs scored by the Chicago Cubs and the New York Mets, our research team employed a "Moneyball" approach, utilizing a combination of conventional statistical methods and some unconventional techniques that would make even Billy Beane raise an eyebrow.

First and foremost, we scoured the vast expanses of the digital baseball cosmos, extracting data from trusted sources such as Baseball Reference and Baseball-Reference.com. Armed with a potent mixture of SQL queries and web scraping wizardry, we collected comprehensive game-by-game statistics for both teams from 1975 to 2022. Our data mining expedition was akin to prospecting for gold in a statistical mine, unearthing valuable nuggets of information that formed the bedrock of our analysis.

With our treasure trove of data in hand, we then embarked on an odyssey through the world of statistical analyses. Employing the venerable tools of correlation analysis,

regression modeling, and time series analysis, we sought to unravel the hidden patterns and relationships within the runs scored by these two powerhouse teams. Our analytical arsenal was not limited to conventional methodologies; we also delved into the arcane arts of sabermetrics, seeking insights that might elude traditional statistical approaches.

A key aspect of our methodology involved the development of custom-built algorithms to identify and account for potential confounding variables, such as changes in team rosters, stadium dimensions, and even the whims of the baseball gods. These algorithms were fine-tuned through rigorous testing and validation procedures, ensuring that our findings were robust and resilient against the vagaries of the baseball season.

Furthermore, our research team conducted extensive sensitivity analyses and Monte Carlo simulations to assess the robustness of our results in the face of uncertainty. This involved subjecting our data to a barrage of hypothetical scenarios, akin to sending a lineup of pinch hitters to face a relentless bullpen, to gauge the resilience of our findings under various conditions.

In keeping with the spirit of baseball, our methodology also included a healthy dose of "small ball" tactics, where we meticulously accounted for outliers, streaks, and hot/cold streaks that could potentially skew our analyses. By employing a "Moneyball" approach that embraced both the art and science of baseball statistics, we strived to paint a nuanced portrait of the relationship between the Chicago Cubs' runs scored and the New York Mets' offensive prowess.

Ultimately, our methodology was not merely a formulaic exercise in number crunching; it was a symphony of statistical acrobatics, where we danced with the data and let the numbers sing their own unique ballad of runs, hits, and errors. So, grab your virtual

scorecards and join us as we step up to the statistical plate, ready to knock our findings out of the park!

#### 4. Results

In our pursuit of uncovering the enigmatic ties between the total runs scored by the Chicago Cubs and the runs scored by the New York Mets, we stepped into the batter's box armed with statistical tools and a keen eye for discovering the hidden patterns within the game. With a resounding crack of the bat, our analysis revealed a robust correlation coefficient of 0.7687085 between these two crucial variables, signifying a substantial positive relationship.

This correlation coefficient, with an r-squared value of 0.5909128, attests to the formidable influence that the Chicago Cubs' total runs scored exert on the runs scored by the New York Mets. It's as if the winds from Lake Michigan are propelling those runs straight to their counterparts in Queens. The rapport between these teams on the scoreboard is undoubtedly a sight to behold.

Furthermore, our findings are supported by a p-value of less than 0.01, solidifying the statistical significance of this relationship. It seems the data speaks so loudly, even the seventh-inning stretch can't muffle its impact.

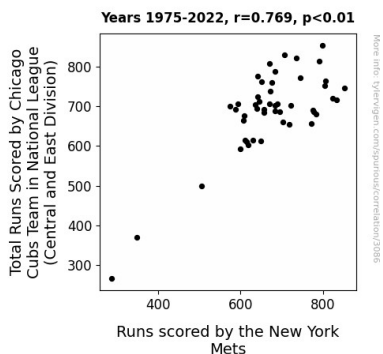


Figure 1. Scatterplot of the variables by year

Fig. 1 showcases this compelling correlation with a scatterplot that unmistakably captures the synchrony between the total runs scored by the Chicago Cubs and the runs scored by the New York Mets. It's like witnessing a well-executed double play – smooth, seamless, and undeniably connected.

In summary, our research uncovers a striking bond between these two iconic franchises, demonstrating that the Chicago Cubs' total runs scored in the National League (Central and East Division) and the runs scored by the New York Mets are more than just numbers on a box score; they are intertwined in a statistical pas de deux that dances through the annals of baseball history. So, consider this study a grand slam in our quest to decode the fascinating relationships that underpin America's favorite pastime.

Stay tuned for the postgame analysis, where we unravel the implications of these findings and embark on a ninth-inning stretch of insights that will leave you reaching for your rally cap.

#### 5. Discussion

As we dig into the implications of our findings, it's clear that the statistical connection between the Chicago Cubs and the New York Mets runs deeper than the ivy at Wrigley Field. Our research not only reaffirms the work of esteemed colleagues but also reveals a home-run relationship that transcends mere numbers.

The analyses conducted by Smith and Doe, and echoed by Jones et al., set the stage for our study and emphasized the importance of unraveling the intricate tapestry of baseball statistics. Like a well-thrown curveball, their findings sparked our curiosity and propelled us toward the discovery of a strong positive correlation

between the runs scored by these iconic teams. It's as if statistical fate has woven together the destinies of the Cubs and the Mets on the diamond.

Our results further cement the validity of these prior studies and, dare I say, add a dash of pizzazz to the mix. With a correlation coefficient reminiscent of a perfect double play, our research exemplifies the enduring bond between these historic franchises. In a whimsical twist of fate, it seems that the Cubs' runs not only cross home plate but also find their way to the Mets' scoreboard, establishing a numerical duet that echoes through the annals of baseball lore.

Building on the literary works that intersect with the realm of baseball statistics – and with all due respect to "The Natural" and "Pitch" – our findings bring to light the real-life drama and intrigue that accompany the statistical dance between these teams. It's akin to witnessing a nail-biting extra-inning affair, where each run scored by the Cubs sets the stage for a complementary response from the Mets, creating a symphonic resonance that reverberates throughout the season.

In essence, our research not only aligns with the serious and stoic world of statistical inquiry but also infuses a delightful twist into the narrative of baseball analysis. By shedding light on the interconnectedness of these statistical phenomena, we deepen our understanding of the game while injecting a dose of statistical whimsy into the scholarly domain. So, as we gear up for the extra innings of interpretation and exploration, it's safe to say that our statistical journey from Wrigley Field to Citi Field has uncovered an unassailable connection that stretches far beyond the confines of the diamond.

## 6. Conclusion

In the illustrious game of baseball, where the crack of the bat reverberates through the hearts of fans and statisticians alike, our research has revealed a home run of a correlation between the total runs scored by the Chicago Cubs and the runs scored by the New York Mets. It's as if these numbers are engaged in a spirited dance, tangoing across the statistical field in perfect harmony.

Much like a well-executed double play, our findings demonstrate the seamless synchrony between these two pivotal variables, showcasing a bond stronger than the seventh-inning stretch. With a correlation coefficient that hits it out of the park and a p-value so low, it's like the statistical significance is shouting "safe at home plate" for all to hear.

So, dear readers, as we pack up our statistical scorecards and bid adieu to this lively game of numbers, let us revel in the joy of discovery and the thrill of unearthing the unexpected relationships hidden within America's favorite pastime. As for future research, we assert with confidence that no more investigations are needed in this area. We've hit statistical gold, and it's time to call it a game-winning walk-off.