

“Stats and Bats”

A Multivariable Math Model for Predicting Win Probability in 12U Travel Baseball

INTRODUCTION

While baseball is packed with numbers, standard professional metrics don't always apply to youth sports. This project explores a simple question: Which game-level statistics actually predict whether a 12U baseball team wins? To find out, I built and trained a mathematical model using data from 82 youth baseball games collected from the *GameChanger* app box scores. Using a statistical method called multivariate linear regression, the model analyzed how different factors impact a team's win probability.

HYPOTHESIS

Teams would have a higher Win Probability (%) when they controlled the strike zone: earning more free bases, striking out less while hitting, getting some extra-base hits, and pitching with better control.

INDEPENDENT VARIABLES

Hitting / Offense (Base Team):
AB (At Bats), R (Runs), H (Hits), RBI (Runs Batted In), BB (Base on Balls), SO (Strikeouts), HBP (Hit by Pitch)
2B (Double), 3B (Triple), HR (Home Run)
Pitching / Defense (Base Team Pitchers):
BB-P (BB allowed), SO-P (Strikeouts), HBP-P (Hit batters)
P# (Total pitches), S# (Total strikes)

DERIVED VARIABLES

- WALKS = BB + HBP (free bases earned)
- WALKS/AB (% walk / free-base rate)
- AVG = H/AB (batting average)
- SO/AB (% strikeout rate)
- XBH/H (% of hits that are extra-base hits)
- Strike Rate (5%) = S#/P# (how often pitches are strikes)
- SO/W-P (Modified) = MIN(SO / (WALKS + 2), 2.36)
- This prevents “divide-by-zero” problems when WALKS = 0 and keeps the regression stable.

DEPENDENT VARIABLE

Win Probability (%) for the 12U team

RELEVANT REFERENCES

GameChanger. (2016, February 11). Baseball and softball analytics report: The 3 stats every player & coach should utilize to improve performance. PR Newswire. <https://www.prnewswire.com/news-releases/gamechanger-baseball-and-software-analytics-report-the-3-stats-every-player-coach-should-utilize-to-improve-performance-300218570.html>

MLB.com. (n.d.). Pythagorean winning percentage. MLB.com. <https://www.mlb.com/glossary/advanced-stats/pythagorean-winning-percentage>

Sanford Sports Analytics Club. (2022, April 27). MLB winning percentage breakdown: Which statistics help teams win more games? Sanford University. <https://www.sanford.edu/sports-analytics/fans/2022/MLB-Winning-Percentage-Breakdown-Which-Statistics-Help-Teams-Win-More-Games>

Ballou-Crawford, C. (2023). Which statistics in baseball are most important for winning? Fort Hays State University Scholars Repository. <https://scholars.fhsu.edu/sacad/vo12023/s23/>

Nayak, S. (2018, June 12). Winning at baseball with math and statistics. Oregon State University IMPACT. <https://science.oregonstate.edu/IMPACT/2018/06/winning-baseball-math-statistics>

METHODOLOGY

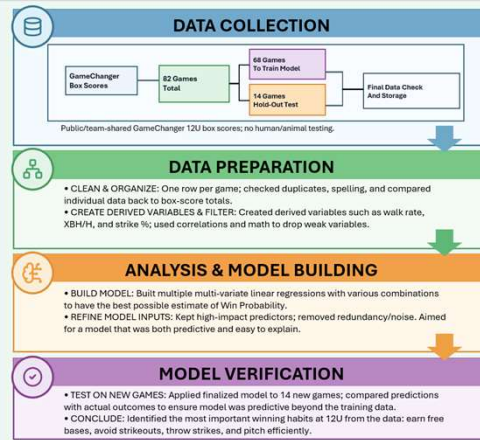


Figure 1: Project Methodology
Graphic created by finalist using MS PowerPoint, 2024.

DATA COLLECTION

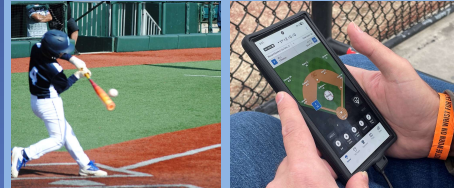


Figure 2: Finalist batting in 12U baseball tournament. Image taken by finalist's parent, 2025.

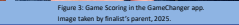


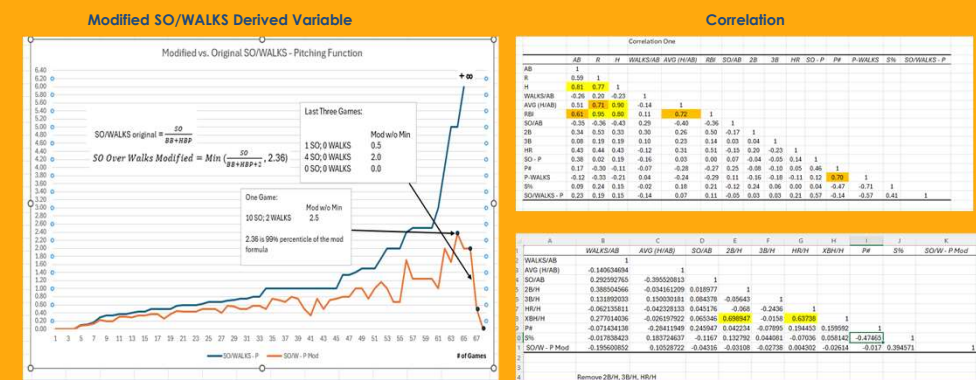
Figure 3: Game Scoring in the GameChanger app. Image taken by finalist's parent, 2025.

DATA PREPARATION

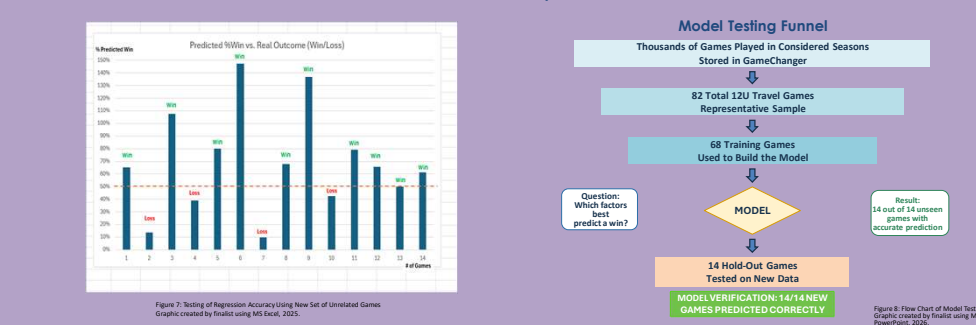


Figure 4: Finalist entering and organizing game data into Excel. Image taken by finalist's parent, 2025.

ANALYSIS & MODEL BUILDING



MODEL VERIFICATION – PREDICTED 14/14 NEW GAMES CORRECTLY



RESULTS / MULTIVARIABLE REGRESSION

Regression Statistics		Predicted Win Probability (W):	
Multiple R	0.64099	$W = f\left(\frac{WALKS}{AB}, \frac{SO}{AB}, \frac{XBH}{H}, P_AVG, S\%, \frac{SO}{W} - Mod\right)$	
R Square	0.41087		
Adjusted R Square	0.34214		
Standard Error	0.39563		
Observations	68		

Figure 9: Linear Regression Model Results
Graphic created by finalist using MS Excel, 2025.

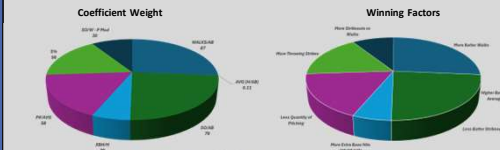
RESULTS / INTERPRETATION

At 12U, winning is more about controlling the strike zone than chasing batting average

What it means (in baseball terms):

- Free bases helped: higher WALKS/AB was one of the strongest “winning” signals.
- Strikeouts hurt: higher SO/AB strongly lowered win probability (balls in play matter at 12U).
- Extra-base hits help: more XBH/H increased win probability (doubles/triples/HR change games).
- Pitching control & efficiency matter: needing more pitches (P#/AVG) and a worse S% (more pitches needed per strike) lowered win probability.
- K-to-walk pitching skill matters: higher SO/W – P Mod increased win probability without divide-by-zero problems.
- Batting average alone didn't explain much (its coefficient was very small).

VARIABLES THAT AFFECT WIN PROBABILITY



LESSONS LEARNED

- Getting the data was harder than expected. GameChanger had many useful box-score stats, but some process stats I wanted, like first-pitch strike %, quality at-bats (QAB), and hard-hit balls, were not available.
- Real data is messy. Some games had zero walks, which broke the normal SO/WALKS ratio, so I created a modified version with a small cushion and cap to keep the model stable.
- Models can only learn from what is measured. Since I did not have reliable fielding, baserunning, catching, errors, or defensive-positioning data, my conclusions focus mostly on hitting & pitching stats.
- Prediction is not the same as causation. Regression showed strong patterns connected to winning, but baseball is still a team game, so one stat alone does not cause a win.
- Next time, I would expand the model. I would add more teams, more games, and factors like weather, field conditions, umpire strike zones, opponent strength, and more detailed GameChanger data.

TAKEAWAYS

- My hypothesis was mostly supported. The biggest winning habits were simple but powerful: take free bases, strike out less, hit for extra bases, and throw more strikes.
- Batting average alone did not explain winning well once stronger “process stats” were included.
- My model has limits because it only used GameChanger box-score data. Future versions should add defense, baserunning, opponent strength, weather, field conditions, and umpire strike zones.
- Regression shows strong patterns, but it does not prove that one stat alone causes winning. Baseball is still a team game with many connected parts.
- Support Tool: Turn the model into a Coach's Checklist, calculator, or AI Baseball Assistant that helps 12U coaches test “what-if” situations and plan smarter practices on the go.

WHAT COULD AN AI BASEBALL ASSISTANT DO?

What the assistant could help with:

- Game plan**: Turn team + opponent stats into a simple strategy.
- Practice plan**: Suggest drills tied to the strongest winning factors.
- What-if testing**: See how changing walk rate, strikeouts, or strike rate shifts the prediction.

AI Baseball Assistant

Coach: Last 8 games: WALKS/AB = .14, SO/AB = .29, XBH/H = .33, and 59% strikes. What should we improve?

AI Assistant: According to my model, the biggest opportunities are more walks, fewer hit strikeouts, and a higher strike rate. XBH/H is already a strength.

Coach: Their pitchers walk a lot of hitters. Game plan?

AI Assistant: Be patient. Force them into the zone and make them earn outs. Free bases are one of my strongest positive winning signals.

Coach: What should we practice today?

AI Assistant: Plate discipline, two-strike contact, and strike-percentaging pitching.

Coach: One message for the team?

AI Assistant: Take the freebies. Put the ball in play. Throw strikes.

Figure 11: Sample interaction between coach and AI Baseball Assistant based on the finalist's math model
Graphic created by finalist using MS PowerPoint, 2026.