

# Major League Baseball 2009

Team	Wins	Runs Scored	Runs Allowed	Margin
Arizona	70	720	782	
Atlanta	86	735	641	
Baltimore	64	741	876	
Boston	95	872	736	
Chicago (NL)	83	707	672	
Chicago (AL)	79	724	732	
Cincinnati	78	673	723	
Cleveland	65	773	865	
Colorado	92	804	715	
Detroit	86	743	745	
Florida	87	772	766	
Houston	74	643	770	
Kansas City	65	686	842	
Los Angeles (AL)	97	883	761	
Los Angeles (NL)	95	780	611	
Milwaukee	80	785	818	
Minnesota	87	817	765	
New York (NL)	70	671	757	
New York (AL)	103	915	753	
Oakland	75	759	761	
Philadelphia	93	820	709	
Pittsburgh	62	636	768	
San Diego	75	638	769	
San Francisco	88	657	611	
Seattle	85	640	692	
St. Louis	91	730	640	
Tampa Bay	84	803	754	
Texas	87	784	740	
Toronto	75	798	771	
Washington	59	710	874	

For more statistics and other information, go to [www.mlb.com](http://www.mlb.com)



## Scoring and Winning: Baseball

1. Identify the independent and dependent quantities. Do wins depend upon runs scored/allowed or do runs scored/allowed depend upon wins? What other data (players' performance) contribute to success?

### Part 1:

2. Graph the data (runs scored/allowed, wins) as a scatter plot and describe the scatter plot.
3. Determine a line of best-fit and graph with the scatter plot. Describe how well the line models the data.
4. Plot the residuals and determine the correlation coefficient for the line's fit. According to the residual plot and correlation coefficient, how well does the line model the data? Explain.
5. Assuming the line generated by the calculator is a good model, describe the slope in the context of the data graphed.
6. The New York Yankees had the most wins this season, 103 in 162 games played. According to the line of best-fit, what should be the total runs scored/allowed for a team with 103 wins? 162 wins (perfect season)?
7. How accurately does the model predict/reflect Atlanta's performance?



### Part 2:

8. Calculate the Scoring Margin (runs scored - runs allowed) for each team. What does a negative result mean?
9. Graph the data (scoring margin, wins) and describe the scatter plot.
10. Determine a line of best-fit and graph with the scatter plot. Describe how well the line models the data.
11. Plot the residuals and determine the correlation coefficient for the line's fit. According to the residual plot and correlation coefficient, how well does the line model the data? Explain.
12. Assuming the line generated by the calculator is a good model, describe the slope in the context of the data graphed.
13. The New York Yankees had the most wins this season, 103 in 162 games played. According to the line of best-fit, what should be the scoring margin for a team with 103 wins? 162 wins (perfect season)?
14. How accurately does the model predict/reflect Atlanta's performance?

# Teacher Guide

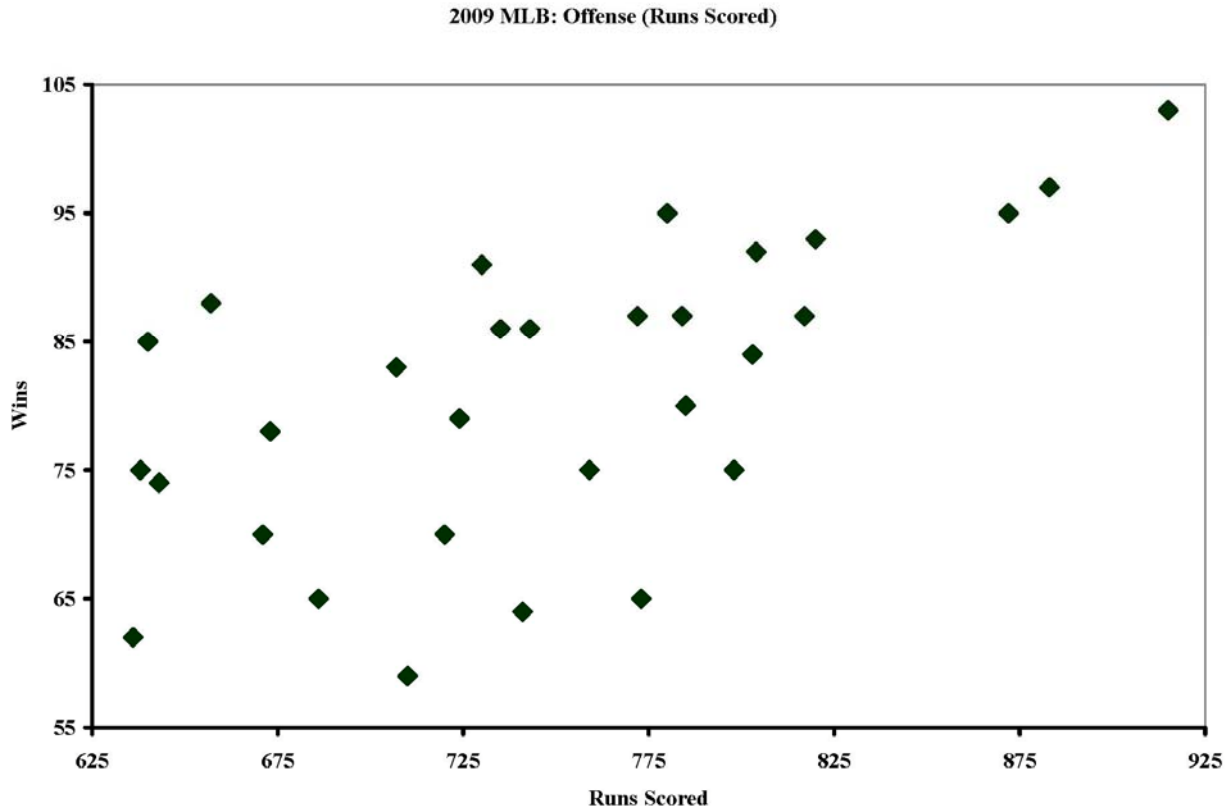
1. Identify the independent and dependent quantities. Do wins depend upon runs scored/allowed or do runs scored/allowed depend upon wins? What other data (players' performance) contribute to success?

Independent: runs scored/allowed      Dependent: wins

Wins depend upon runs scored or allowed. Student answers will vary for the second question.

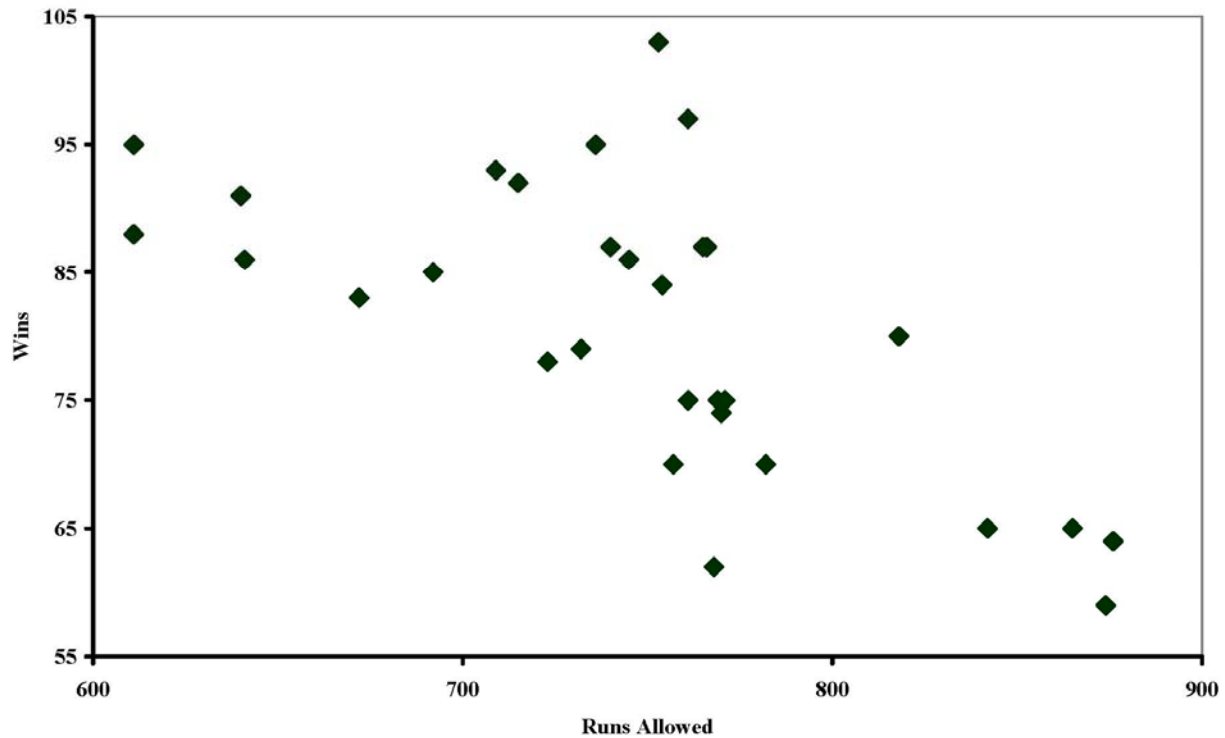
## Part 1:

2. Graph the data (runs scored/allowed, wins) as a scatter plot and describe the scatter plot.



The scatter plot for Runs Scored shows a positive relationship.

2009 MLB Defense (Runs Allowed)



The scatter plot for Runs Allowed shows a negative relationship.

- Determine a line of best-fit and graph with the scatter plot. Describe how well the line models the data.

Runs Scored:  $Y = 0.09377X + 10.9286$ ; not a statistically sound fit, too much scatter.

Runs Allowed:  $Y = -0.1096X + 162.904$ ; not a statistically sound fit, too much scatter.

- Determine the correlation coefficient for the line's fit. According to the correlation coefficient, how well does the line model the data? Explain.

Runs Scored:  $r = 0.6113$ , not a very good fit.

Runs Allowed:  $r = -0.6589$ , not a very good fit.

- Assuming the line generated by the calculator is a good model, describe the slope in the context of the data graphed.

Runs Scored: slope = 0.09377. After rounding, think of the slope as  $\frac{9}{100}$ , for every 100 runs scored, nine more wins occur.

Runs Allowed: slope = -0.1096. After rounding, think of the slope as  $-\frac{1}{10}$ , for every 10 runs allowed, one fewer wins occur.

- The Los Angeles Angels had the most wins this season, 103 in 162 games played. According to the lines of best-fit, what should be the total runs scored/allowed for a team with 103 wins? 162 wins (perfect season)?

Runs Scored model: 981 runs should be scored to get 103 wins. 1611 runs should be scored to get 162 wins.

Runs Allowed model: 547 runs should be allowed to get 103 wins. Eight runs should be allowed to get 162 wins.

- How accurately do the models predict/reflect Atlanta's performance (86 wins)?

Runs Scored model: Missed by 66 runs (9.0%).

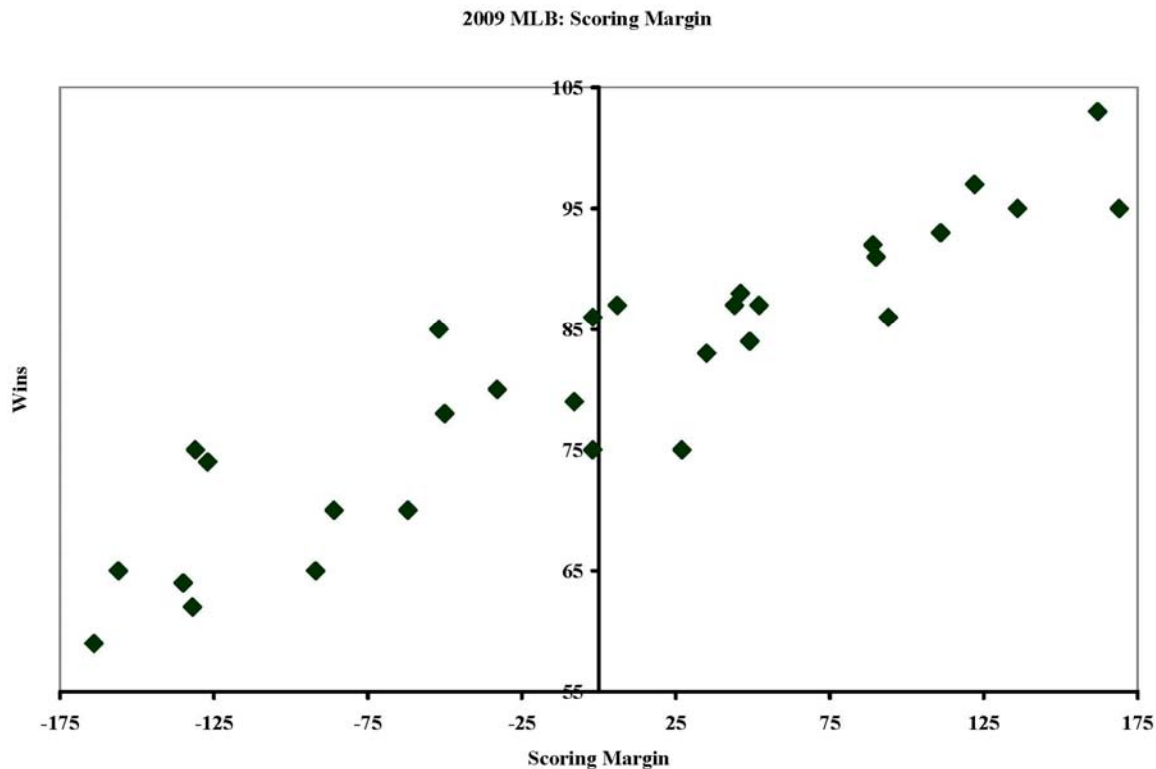
Runs Allowed model: Missed by 61 runs (9.5%).

**Part 2:**

- Calculate the Scoring Margin (runs scored - runs allowed) for each team. What does a negative result mean?

A negative result means a team allowed more runs than it scored.

- Graph the data (scoring margin, wins) and describe the scatter plot.



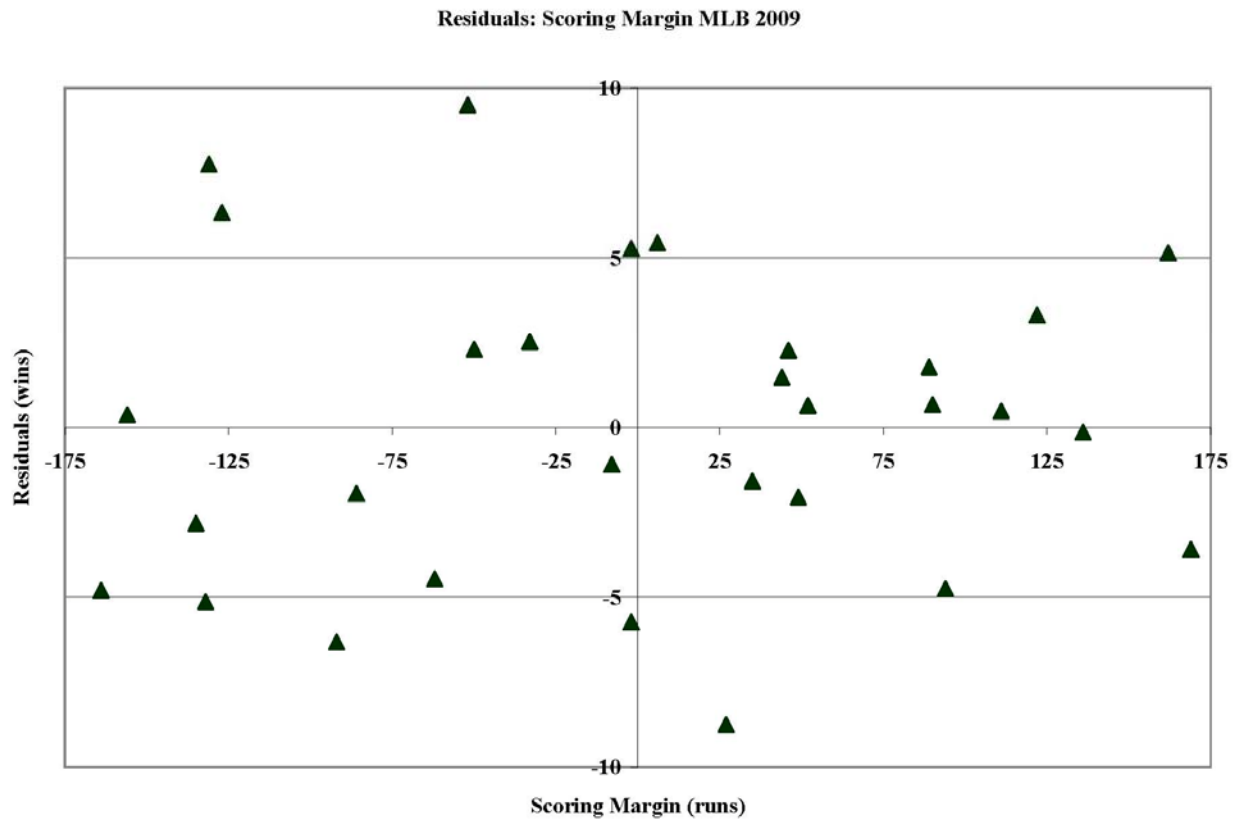
The scatter plot for Scoring Margin shows a positive relationship.

10. Determine a line of best-fit and graph with the scatter plot. Describe how well the line models the data.

$Y = 0.10656X + 81$ , appears to be a good fit with minimal scatter.

11. Plot the residuals and determine the correlation coefficient for the line's fit. According to the residual plot and correlation coefficient, how well does the line model the data? Explain.

$R = 0.92019$ , indicates a good fit since  $r$  is close to 1.



The scatter in the residual plot indicates that a linear model is a good fit for the data.

12. Assuming the line generated by the calculator is a good model, describe the slope in the context of the data graphed.

Slope = 0.10656. After rounding, think of the slope as  $\frac{1}{10}$ , for every 10 run difference, one more win occurs.

13. The Los Angeles Angels had the most wins this season, 103 in 162 games played. According to the line of best-fit, what should be the scoring margin for a team with 103 wins? 162 wins (perfect season)?

A team should score 206 more runs than its opponents to get 103 wins. A team should score 760 more runs than its opponents to get 162 wins.

14. How accurately does the model predict/reflect Atlanta's performance (86 wins)?

Missed by 47 runs (50%).