

The Effects of Climate in Major League Baseball

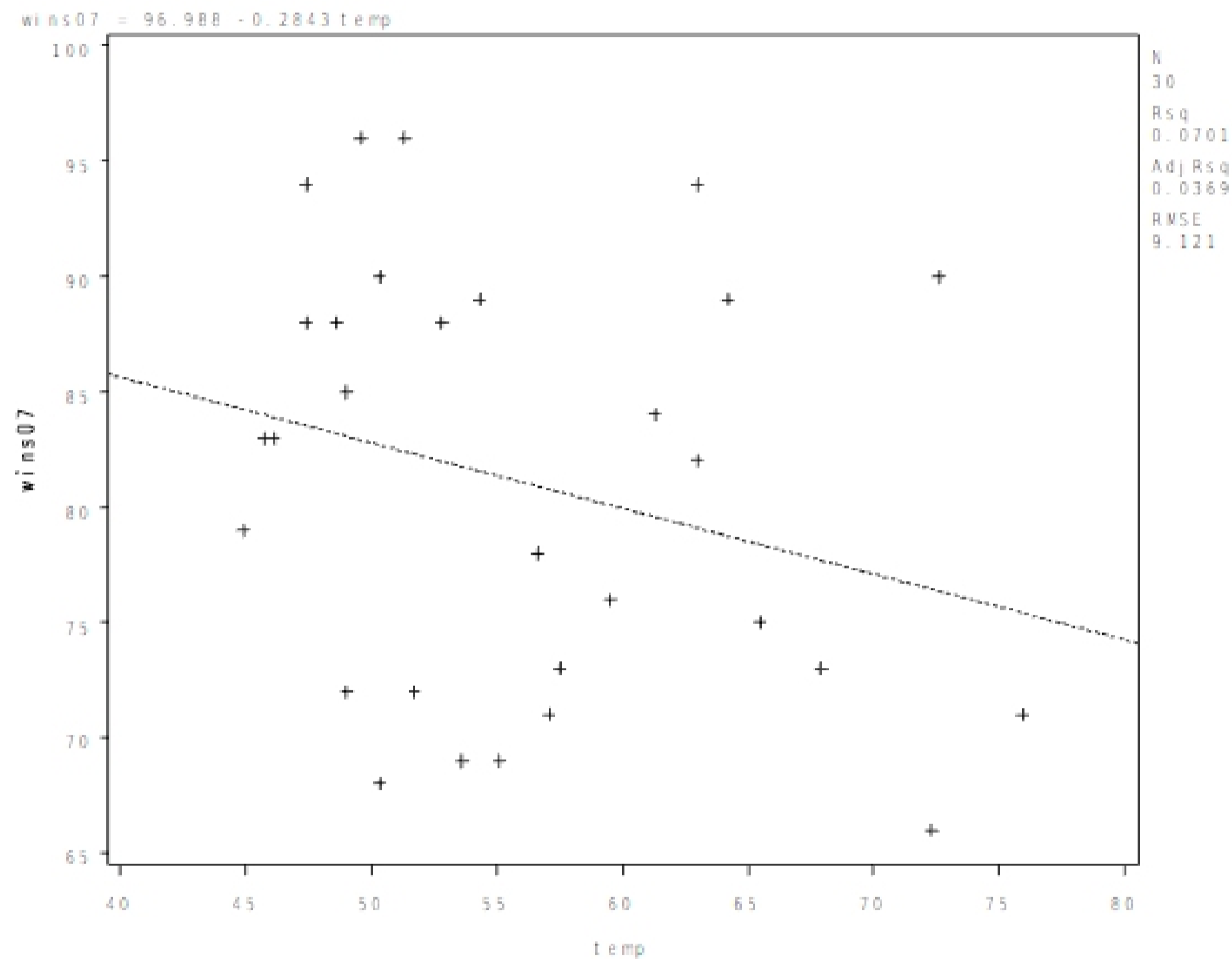
Baseball has been pegged as a seasonal sport since its invention in the late 18th century. The MLB regular season is 162 games long and runs from April through September. How does each team's respective city's climate affect their performance on the field? We wanted to know how, if at all, climate affected teams in Major League Baseball. We hypothesized that generally, the warmer the climate, the more wins the team has.

We gathered data from the internet concerning all 30 teams in Major League Baseball, including: each team's mean batting average in 2007, wins in 2007, wins in 2006, and the mean temperature of each team's home city over the past 30 years, shown below.

.297	94	89	63.0	Los Angeles Angels
.285	96	86	51.3	Boston Red Sox
.279	83	87	45.7	Toronto Blue Jays
.277	78	83	56.6	St. Louis Cardinals
.275	82	88	63.0	Los Angeles Dodgers
.273	90	76	72.6	Arizona Diamondbacks
.272	94	97	47.4	New York Yankees
.272	69	62	53.6	Kansas City Royals
.269	88	97	47.4	New York Mets
.267	85	66	49.0	Chicago Cubs
.266	84	79	61.3	Atlanta Braves
.265	71	78	75.9	Florida Marlins
.264	88	78	52.8	Seattle Mariners
.264	89	88	64.2	San Diego Padres
.262	79	96	44.9	Minnesota Twins
.262	88	95	48.6	Detroit Tigers
.258	75	80	65.5	Texas Rangers
.256	66	61	72.3	Tampa Bay Rays
.252	72	80	51.7	Cincinnati Reds
.252	69	70	55.1	Baltimore Orioles
.249	89	85	54.3	Philadelphia Phillies
.248	68	67	50.3	Pittsburgh Pirates
.247	90	76	50.3	Colorado Rockies
.246	72	90	49.0	Chicago White Sox
.243	83	75	46.1	Milwaukee Brewers
.241	76	93	59.5	Oakland Athletics
.235	71	76	57.1	San Francisco Giants
.233	73	82	67.9	Houston Astros
.232	96	78	49.6	Cleveland Indians
.227	73	71	57.5	Washington Nationals

Using SAS, we produced scatter-plots of the data using temperature as the explanatory variable and wins as the response variable. We started with 2007, shown below with a best fitting line.

It seems like the points are all over the place, and that there is absolutely no relationship between climate and wins. In fact, the best fitting line would suggest that a colder climate results in more wins. However, this alone does not provide enough to determine if there is a relationship so this is merely an inference.



Below is the Corr Procedure output in SAS. Displayed at the bottom is the correlation coefficient or r-value which will help determine how significantly our two variables have a linear relationship. To be considered a strong relationship, the r-value would have to be close to either 1 or -1. Here our r-value is -0.26476 , much closer to 0 than either 1 or -1, meaning a very weak linear relationship. It is enough to say that the two variables don't seem to affect each other at all. Our earlier inference is backed up.

```

The CORR Procedure
2 Variables:  wins07  temp

Simple Statistics
Variable      N      Mean      Std Dev      Sum      Minimum      Maximum
wins07        30     81.03333     9.29399     2431     66.00000     96.00000
temp          30     56.11667     8.65492     1684     44.90000     75.90000

Pearson Correlation Coefficients, N = 30
Prob > |r| under H0: Rho=0

           wins07      temp
wins07    1.00000    -0.26476
           0.1574
temp      -0.26476    1.00000
           0.1574

```

Further, we looked at the residuals to see just how far each actual value strayed its predicted value, this again backed up our inference.