Journal of Student Research

Siméon Poisson and the Stanley Cup Finals

Cate R. Wagner[1], Erica J. Barr[1], Joseph W. Spada[1], Cole M. Joslin[1] and
Paul M. Sommers[1]

[1]Middlebury College, Middlebury, VT, USA

ABSTRACT

The National Hockey League is a professional ice hockey league in North America currently comprised of 31 teams. Their seasons culminate with the Stanley Cup Playoffs. The top sixteen teams (eight in each conference) qualify for the playoffs. The conference champions face off in the final round, known as the Stanley Cup Finals. The authors show that goals scored per game in the Stanley Cup Finals follow a Poisson distribution. Using the results of all 438 Stanley Cup Final games played since 1939 (when the Finals became a best-of-seven series), chi-squared goodness-of-fit tests show that the observed distribution of goals scored per game by series winners, series losers, and game losers closely approximate a Poisson theoretical model. The combined number of goals scored by both finalists and goals scored by game winners do not.

Siméon Denis Poisson (1781 – 1840), French mathematician, engineer, and physicist, is no stranger to the National Hockey League (NHL). Mullet [1], using data from the 1973-74 season, found that for each of the sixteen teams then in the NHL, goals scored for and goals scored against, both home and away, followed a Poisson probability distribution.

In this paper, the authors hypothesize that a Poisson probability model can also describe the number of goals scored per game by the series winner as well as by the series loser in the Stanley Cup Finals.

The Data

In 1939, the Stanley Cup Finals (which matches the two conference champions) became a best-of-seven series. A season-long lockout in 2004-05 cancelled all games including the Stanley Cup Finals (for the first time since 1919, when completion of the Finals after five games had been played were cancelled due to an outbreak of the Spanish flu). Apart from the 2004-05 season, between 1939 and 2019, a combined 438 games have been played in eighty Stanley Cup Finals. Twenty of the series were 4-game sweeps, nineteen lasted five games, twenty-four ended in six games, and seventeen were decided in seven games. Game-by-game results for each year's Stanley Cup Finals best-of-seven series are from hockey-reference.com [2].

The number of goals scored in a game is a discrete random variable. The probability that a team scores a goal is small and assumed to be the same throughout a game. Moreover, the number of goals scored per game by either team in any one game in the Finals is assumed to be independent of the number of goals scored per game by either team in any other game in the same series. Under these assumptions, is the distribution of goals scored by either

the series winner or the series loser (regardless of the number of games in the best-of-seven series) Poisson? Surprisingly, the answer is yes.

Methodology

The Poisson distribution is a discrete probability distribution which has the following formula:

(1) $$p(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, 2, \ldots$$

where λ denotes the mean number of goals scored per game or the expected value of the Poisson distribution. In Table 1, a Poisson distribution is shown to be an appropriate model for the number of goals scored by a Stanley Cup series winner for all 438 games.

The expected value of the Poisson distribution is estimated by

(2) $$\hat{\lambda} = \frac{\sum x_i o_i}{n} = 3.1895$$

where $n = 438$. Small expected values (E_i), less than 1.0, can cause problems in a χ^2 goodness-of-fit test. In Table 1, all expected frequencies are greater than 1.0 and hence obviated the need to pool adjacent classes until all expected frequencies were at least 1.0 as suggested in [3].

How well does the observed frequency distribution conform to the Poisson distribution? The null hypothesis is that the observed or actual distribution can be represented by the theoretical (Poisson) distribution, and that the discrepancies between them are due to chance. The value of the test statistic is

(3) $$\chi^2_{calculated} = \sum \frac{(O_i - E_i)^2}{E_i} = 7.201$$

The number of degrees of freedom is, in general, equal to k – the number of estimated parameters – 1, where k is equal to the number of classes or categories. We used $\bar{x}$ from the actual distribution to estimate λ in the Poisson distribution. Hence, for the distribution shown in Table 1, $k = 10$ and the number of degrees of freedom is therefore $10 - 1 - 1$ or 8. Since $\chi^2_{.05,8} = 15.507$ and exceeds the value of the test statistic, it follows that we cannot reject the null hypothesis. In other words, the probability is greater than .05 (in particular, .515) that the observed discrepancies between the actual distribution and the Poisson distribution could be due to chance.

Since there is no general agreement on the minimum expected frequency allowed, when we combined the two classes in Table 1 for "8" or "9" goals scored per game so that expected values were at least 5, the calculated χ^2 value was 6.609 and the p-value of the test was still greater than .05 (in particular, .471).

The Results

The results of nineteen different χ^2 goodness-of-fit tests are reported in Table 2. All degrees of freedom were adjusted to account for the estimated parameter ($\bar{x}$ in place of λ) for the Poisson hypothesis.

For all tests involving series winners or series losers, however many games were required to determine a Stanley Cup champion, the Poisson models provided an excellent fit. Only in the case where we combined the number of goals scored by the series winner *and* the series loser were we able to reject the Poisson hypothesis.

For all tests involving the number of goals scored by the *game* (not series) winners, we were able to reject the Poisson hypothesis, largely because the observed frequency of no goals scored by game winners was always zero and less than the expected frequency of no goals scored using the Poisson distribution. That is, a game winner always scored at least one goal. Pooling the two classes for "0" and "1" goal scored by game winners (however long the series) did not improve the fit.

For all tests involving the number of goals scored per game by the *game* loser, we could not reject the Poisson hypothesis. That is, the Poisson model provides a surprisingly excellent fit to the data on goals scored per game by the *game* loser.

For all but one of the nineteen tests reported in Table 2 (the exception being game losers in a 5-game series), the χ^2 tests produced the same results with pooling until expected values were at least 5.

Concluding Remarks

Using goals scored per game in the Stanley Cup Finals for the period 1939 through 2019, hypothesis tests indicate that the number of goals scored per game by series winners, series losers, and game losers follow a Poisson probability model. The results of our tests are robust for 4-game sweeps as well as for a best-of-seven series of any length.

Interested readers might want to know whether the Poisson model fits the number of runs scored in a best-of-seven World Series in Major League Baseball or the number of goals scored in the knockout stage of the FIFA World Cup.

Journal of Student Research

Table 1. Distribution of the Number of Goals Scored per Game, Series Winners, 1939 – 2019

Number of goals scored (x_i)	Observed number of games (O_i)	Poisson probability[a] (p_i)	Expected number of games ($E_i = 438 \times p_i$)
0	17	.0412	18.0423
1	52	.1314	57.5460
2	88	.2095	91.7714
3	113	.2228	97.5682
4	83	.1776	77.7984
5	38	.1133	49.6276
6	29	.0602	26.3812
7	11	.0274	12.0204
8	6	.0109	4.7924
9	1	.0039	1.6984

[a]The expected value of the Poisson distribution is estimated by $\bar{x} = 3.1895$ where n = 438.

Table 2. Stanley Cup Goals Scored per Game, Chi-squared Goodness-of-Fit Tests to a Poisson Probability Distribution, 1939 – 2019

Group	Number of games	Average number of goals per game	χ^2 calculated	p-value
All series winners	438	3.189	7.201	0.515
All series losers	438	2.123	5.962[*]	0.427
All series, winners and losers	438	5.313	41.151[*]	**< 0.001**
4-game sweeps, series winners[a]	80	3.763	8.990	0.253
4-game sweeps, series losers[a]	80	1.613	1.422	0.840
5-game series winners	95	3.474	9.041	0.250
5-game series losers	95	2.137	3.980	0.552
6-game series winners	144	3.063	9.758	0.203
6-game series losers	144	2.340	9.357[*]	0.154
7-game series winners	119	2.731	4.866[*]	0.561
7-game series losers	119	2.193	6.875[*]	0.230

Table 2. Stanley Cup Goals Scored per Game, Chi-squared Goodness-of-Fit Tests to a Poisson Probability Distribution, 1939 – 2019 (Continued)

Group	Number of games	Average number of goals per game	χ^2 calculated	*p*-value
All series, game winners	438	3.733	50.020	**< 0.001**
All series, game losers	438	1.580	8.366[*]	0.137
5-game series, game winners	95	3.874	22.663	**0.002**
5-game series, game losers	95	1.737	9.148	0.058
6-game series, game winners	144	3.757	23.249	**0.002**
6-game series, game losers	144	1.646	4.660[*]	0.324
7-game series, game winners	119	3.571	19.185[*]	**0.008**
7-game series, game losers	119	1.353	3.136[*]	0.371

[*]Classes were pooled until all expected values were at least 1.0 as suggested by Snedecor and Cochran (1967).
[a]The 4-game sweep series winner (loser) is the same as the 4-game sweep game winner (loser).

References

1. G. M. Mullet, "Siméon Poisson and the National Hockey League," *The American Statistician*, Vol. 31, No. 1 (February 1977), pp. 8-12.

2. Game-by-game results in the Stanley Cup Finals (1939 – 2019) retrieved from: https://www.hockey-reference.com/search/search.fcgi?hint=stanley+cup+winners&search=stanley+cup

3. G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 6th edition, Ames, Iowa: Iowa State University Press, 1967.