



## World Series Projected Games

## Data Analysis Report

Prepared For:

Mrs. Wheeler,

President of CBS

Prepared By:

Evan Bennett,

Head of *Harborside's Awesome*

*Research and Analysis LLC.*

## Contents

<b>1. Summary</b> .....	<b>3</b>
<b>2. Frequency of World Series</b> .....	<b>4</b>
<b>3. Distribution of World Series</b> .....	<b>5</b>
3.1 Bar Chart.....	5
3.2 Description of distribution .....	5
<b>4. Measure of Central Tendency Deviation of World Series</b> .....	<b>6</b>
4.1 Central Tendency.....	6
4.2 Mean.....	6
4.3 World Series Variance and Deviation.....	7
<b>5. Accurate Representation</b> .....	<b>8</b>
5.1 Mean.....	8
5.2 Standard Deviation.....	8
<b>6. World Series Probability Model</b> .....	<b>9</b>
6.1 Probability Model.....	9
<b>7. Advertisement Bid</b> .....	<b>10</b>
<b>8. Conclusion</b> .....	<b>11</b>

## 1. Summary

---

### Introduction

As a major investor of the World Series and a supporter of the organization's success, CBS is determining the amount to bid for advertisement during each World Series Game. Based on data taken from every World Series since 1905 (with the exclusion of 1903, 1904, 1919, 1920, 1921 and 1994), the measure of central tendency can be used to determine the average game length. With a series length of four to seven games the mean can be identified along with the standard deviation. By doing so, CBS will better be able to make a bid on advertisement revenue. This report was conducted independently by *Harborside's Awesome Data Analysis Research (HADAR)* for a fair and efficient investment process. That benefits both the needs of private firms and the public good.

## 2. Frequency of the World Series

---

### Game Occurrence

To determine the frequency that a 4, 5, 6 or 7 game series occurs, data was plotted on the frequency chart below. From this group of information, a measure of central tendency can be used to find the most common series occurrence.

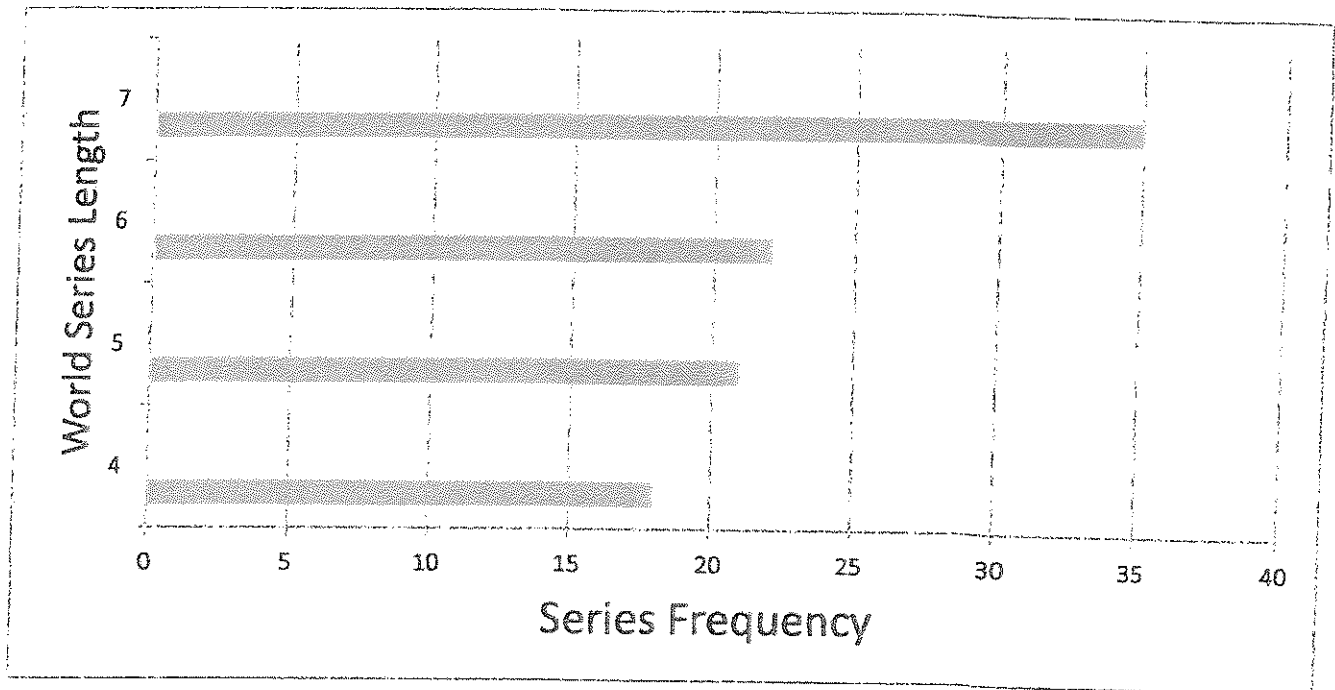
Series	Frequency	Occurrence
4	18	A series of 4 game occurred 18 times
5	21	A series of 5 games occurred 21 times
6	22	A series of 6 games occurred 22 times
7	35	A series of 7 games occurred 35 times

### 3. Distribution of World Series Games

---

#### 3.1 Bar Chart

By plotting the data in a bar chart, trends in World Series frequency can be determined. As series lengths progressed, the frequency of them increased. The chart showed a substantial difference in frequency for 7 games series compared to the others.



#### 3.2 Distribution

The distribution of data on the chart above showed a clear trend for 7 game series. There were 35 seven game series, 22 five game series, 21 five game series and 18 four game series. Although there was a trend favoring 7 game series, the margin between games 4, 5, 6 and 7 was minimal.

## 4. Measures of Central Tendency and Standard Deviation for the World Series

---

### 4.1 Central Tendency

In order to support the trend for 7 game series shown in the chart above, the frequency chart below was used to find the central tendency.

Series Length (x)	Frequency (f)	Xf	(x- $\bar{x}$ )	(x- $\bar{x}$ ) <sup>2</sup>	(x- $\bar{x}$ ) <sup>2</sup> f
4	18	72	-1.8	3.24	58.32
5	21	105	-0.8	0.64	13.44
6	22	132	0.2	0.04	0.88
7	35	245	1.2	1.44	50.4

### 4.2 Mean

To find the average number of World Series Games, the formula had to be adjusted to account for the how many times each series occurred. Therefore, the series length (x) was multiplied by the frequency (f) to get (xf). Then, by using the sum ( $\sum xf$ ) of the xf column and dividing it by the sum of the frequency, the average series length was determined. By finding the average length of the World Series, a better estimate of the most financially secure investment could be made.

$$\text{Mean} = \bar{x} = \frac{\sum xf}{n} = \frac{554}{96} = 5.8$$

### 4.3 World Series Variance and Deviation

The standard deviation showed a low spread of Series Lengths around the mean of 5.8. For determining the accuracy of the mean, the deviation is important for ensuring appropriate representation.

$$\text{Sample Variance} = s = \frac{\sum(x - \bar{x})}{n - 1} = \frac{123.04}{95} = 1.29$$

$$\text{Standard Deviation} = s = \sqrt{s} = \sqrt{1.29} = 1.13$$

## 5. Accurate Representation

---

### 5.1 Mean

The mean gives a balanced representation of the average amount of games in a World Series. If mode was used, there wouldn't be enough possible repeating series numbers to find the central measure. Range also wouldn't give an accurate representation because the highest number subtracted from the lowest one isn't a valid number for the analysis. In addition, the mean is a primary indication that the high or low numbers are out of proportion to the central tendency.

### 5.2 Standard Deviation

The standard deviation measured the spread of World Series game lengths. The deviation initially showed a low spread but to get a more accurate representation, the coefficient of variation needed to be determined. The coefficient of variation showed 19.48% indicating a higher percentage and a greater amount of spread.

#### 5.2.1 Coefficient of Variation

$$\text{Coefficient of variation} = cv = \frac{\text{Standard Deviation}}{\text{Mean}} \circ 100$$

$$cv = \frac{1.13}{5.8} \circ 100 = 19.48\%$$



## 6. Probability Model for the World Series

---

### 6.1 Probability Model

The most financial secure bid also needs to take into consideration the probability of each series length. By creating a probability model for 4 games, 5 games, 6 game and 7 games, the percentage of each was determined to calculate the odds of each occurrence. In correlation with the bar chart in section 3.1, a 7 games series occurred 36.45%, the highest out of the 4 series lengths.

Series of 4 Games

$$P(4 \text{ games}) = \frac{18}{96} = \frac{3}{16} = 0.1875 \text{ or } 18.75\%$$

Series of 5 Games

$$P(5 \text{ games}) = \frac{21}{96} = \frac{7}{32} = 0.2187 \text{ or } 21.87\%$$

Series of 6 Games

$$P(6 \text{ games}) = \frac{22}{96} = \frac{11}{48} = 0.2291 \text{ or } 22.91\%$$

Series of 7 Games

$$P(7 \text{ games}) = \frac{35}{96} = 0.3645 \text{ or } 36.45\%$$

## 7. Advertisement Bid

---

Based on the accumulative collection of all the data above, an advertisement bid could be made. Taking in to account the 2 million dollar advertisement revenue for each televised game, and the resulting net loss for each game not aired, 10 million would be the most financially secure and profitable bid. 7 game series showed a clear trend which would make a profit of fourteen million. However, the mean length showed just a game length of 5.8. By taking that, the 12 million dollars in profit and 2 million lost if the mean does occur into consideration, a slightly lower bid would be most practical. 10 million would be high enough to be profitable but also safe enough to not lose a large amount. The probability model showed that Series 5, 6 and 7 where most likely to occur, all of which would yield a profit or an equal return. If a Series ran 6 or 7 games as projected, CBS will have a net profit of 2 to 4 million.

Possible Series Game Lengths	Revenue	Gross Profit	Net Profit with 10,000,000 bid
4	2,000,000	8,000,000	-3,000,000
5	2,000,000	10,000,000	0
6	2,000,000	12,000,000	2,000,000
7	2,000,000	14,000,000	4,000,000

---

## 8. Conclusion

---

The World Series Data Analysis Report conducted by the non-profit and independent *Harborside's Awesome Data Analysis Research LLC. (HADAR)* used the data from all World Series games since 1905 (with the exclusion of 1903, 1904, 1919, 1920, 1921 and 1994) to find the likelihood of World Series Games. The purpose of the report is to determine the amount CBS should bid for advertisement during each World Series Game.

The results found that game seven showed a clear trend in frequency, occurring 35 times out of the 96 Series. However the mean for the World Series was 5.8 with a standard deviation of 1.13. To ensure the trends shown in the bar chart and mean, a probability model was used to demonstrate the exact percentage of each series. A 7 game series was 36.5%, 6 game series 22.91%, 5 game series 21.87% and 4 game series 18.75%.

From the data listed above, a bid of 10 million was found to be most logical. Every televised game earns 2 million dollars in advertised revenue. Therefore, adhering to the mean, probability and frequency of the length of the World Series, a bid of 10 million was submitted.