

World Cup Fever: **Study of FIFA World Cup 2010**

Project #1
MATH 10
Online

Project Team: Rohit Behl & Elizabeth Oing

Project Summary:

Given the popularity evidenced by its recent coverage in the news and in sports bars across the country, we decided to study the FIFA matches in the recent World Cup 2010 soccer competition. Our parameter of interest is the total number of goals scored by either team per match. For example, in entry 63 between Uruguay and Germany, the final result was 2-3 (2 goals Uruguay, 3 goals Germany). For our survey, that is counted as 5 goals (2+3=5). The statistic we expect to calculate is the distribution of goals scored in a World Cup soccer match. The data for this project are quantitative and discrete as they are numbers resulting from counting, not measuring. To gather our data, we collected the figures for all group state and Stage 2 matches from the 2010 FIFA World Cup website (located at <http://www.fifa.com/worldcup/matches/index.html>). The population total is 64 matches from which we took a random sample of 30 matches (see chart 1 “Sample Data” in “Charts & Graphs” appendix).

To choose our sample, we decided upon a simple random. We assigned each match in our population table a number, then entered **randint(1,64,30)** on the TI calculator to output 30 random numbers. Then we found the matches correlating with those numbers on the population grid and created our sample. We chose this method because we felt it offered the best probability of providing a truly random sample.

Having our sample, we created a chart to display the data concisely. Then, we entered the values (total points scored per match) into the Stats menu of our TI calculator in the L₁ column and the frequencies in L₂. Using **1-Var Stats L₁,L₂**, we were able to calculate the following:

Sample Mean (\bar{x})	2.4333
Standard Deviation (s)	1.9779
Q1 (25 th percentile)	1
Median (50 th percentile)	2
Q3 (75 th percentile)	3

In order to find the 70th percentile, we consulted our frequency chart. In examining the cumulative relative frequency column, we see that the data value 2 corresponds to the 0.6333 or the 63rd percentile. As a result, we know the 70th percentile must be above 2. For data value 3, the corresponding cumulative relative frequency is 0.8000 (80th percentile) so we know that 3 covers the range of the 64th percentile to the 80th percentile, which includes the 70th. From this, we conclude the 70th percentile is 3.

Using the values given by our calculators, we were able to determine the values above and below the mean.

$$2(1.9779)+2.433=6.3891$$

$$1.5(1.9779)-2.443=-0.53355$$

The value that is 2 standard deviations above the mean is 6.3891.

The value that is 1.5 standard deviations below the mean is -.053355

From these values, we were able to construct a histogram illustrating the data. The shape of our graph appears to be normal distribution, one mode centered around 0.5-3.0, and skewed to the right.

Looking at the graph itself, we see what appear to be outliers on the far right. To test this, we used the outlier formula. First we determined the interquartile range (IQR) with the formula $IQR=Q3-Q1$.

$$IQR=3-1=2$$

Next, we use the potential outlier formula $(1.5)(IQR)$

$$(1.5)(2)=3$$

And add to Q3, and subtract from Q1

$$3+3=6$$

$$1-3=-2$$

There are no data at or below -2, but there are data above 6. In 2 of the matches, 8 goals were scored. Being above 6, that makes those 2 matches outliers.

Next, we used our data to construct a boxplot. The middle 50% of our data appear to be concentrated together. In addition to the appearance as given on the boxplot, we calculated the range of that area by subtracting the first quartile from the third quartile.

$$3-1=2$$

The data are concentrated equally compared to the last quarter, which has a range of $8-3=5$.

After considering both the histogram and the boxplot, our results appear to be supported by both. The box is set more to the left with a longer whisker marking the last 25% from 3 to 8. Similarly, on the histogram, the last quartile is sparse, with the two adjacent modes peaking to the left. Our distribution is bimodal, with the modes being the values 1 and 2 as evidenced by table 2 (in the appendix "Charts & Graphs"). The median appears in the center of the box, indicating the data are fairly evenly distributed from quartile 1 to quartile 2, and quartile 2 to quartile 3. In both of the charts, the data spike around the values of 1 and 2, showing those to be the most common number of total goals scored per soccer match in our sample.

Charts & Graphs*Sample Data*

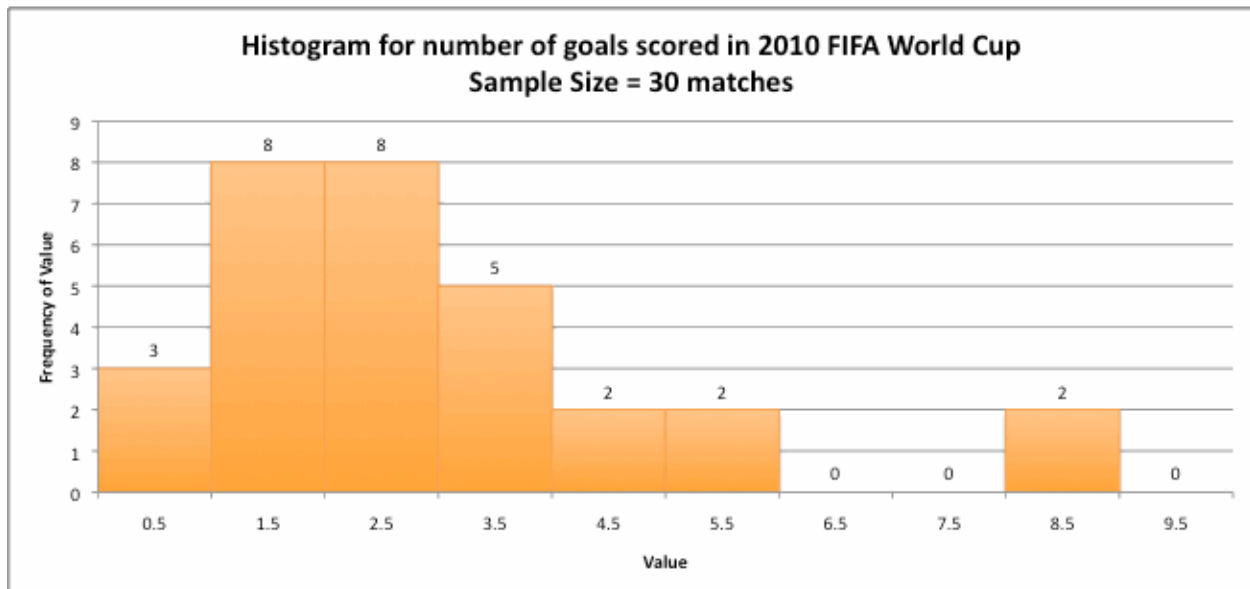
#	Date - Time	Venue	Home	Results	Away	Total Goals
21	6/18/10 1:30 PM	Nelson Mandela Bay/Port Elizabeth	Germany	0-1	Serbia	1
31	6/21/10 4:00 PM	Nelson Mandela Bay/Port Elizabeth	Chile	1-0	Switzerland	1
48	6/25/10 8:30 PM	Mangaung / Bloemfontein	Switzerland	0-0	Honduras	0
36	6/22/10 8:30 PM	Polokwane	Greece	0-2	Argentina	2
21	6/18/10 1:30 PM	Nelson Mandela Bay/Port Elizabeth	Germany	0-1	Serbia	1
27	6/20/10 1:30 PM	Mangaung / Bloemfontein	Slovakia	0-2	Paraguay	2
6	6/13/10 1:30 PM	Polokwane	Algeria	0-1	Slovenia	1
29	6/20/10 8:30 PM	Johannesburg - JSC	Brazil	3-1	Côte d'Ivoire	4
4	6/12/10 1:30 PM	Nelson Mandela Bay/Port Elizabeth	Korea Republic	2-0	Greece	2
32	6/21/10 8:30 PM	Johannesburg - JEP	Spain	2-0	Honduras	2
36	6/22/10 8:30 PM	Polokwane	Greece	0-2	Argentina	2
21	6/18/10 1:30 PM	Nelson Mandela Bay/Port Elizabeth	Germany	0-1	Serbia	1
45	6/25/10 4:00 PM	Durban	Portugal	0-0	Brazil	0
63	7/10/10 8:30 PM	Nelson Mandela Bay/Port Elizabeth	Uruguay	2-3	Germany	5
11	6/14/10 8:30 PM	Cape Town	Italy	1-1	Paraguay	2
52	6/27/10 8:30 PM	Johannesburg	Argentina	3-1	Mexico	4
49	6/26/10 4:00 PM	Nelson Mandela Bay/Port Elizabeth	Uruguay	2-1	South Korea	3
55	6/29/10 4:00 PM	Tshwane/Pretoria	Paraguay	5-3	Japan	8
14	6/15/10 8:30 PM	Johannesburg - JEP	Brazil	2-1	Korea DPR	3
49	6/26/10 4:00 PM	Nelson Mandela Bay/Port Elizabeth	Uruguay	2-1	South Korea	3
64	7/11/10 8:30 PM	Johannesburg	Netherlands	0-1	Spain	1
31	6/21/10 4:00 PM	Nelson Mandela Bay/Port Elizabeth	Chile	1-0	Switzerland	1
32	6/21/10 8:30 PM	Johannesburg - JEP	Spain	2-0	Honduras	2
50	6/26/10 8:30 PM	Rustenburg	USA	1-2	Ghana	3

25	6/19/10 1:30 PM	Durban	Netherlands	1-0	Japan	1
49	6/26/10 4:00 PM	Nelson Mandela Bay/Port Elizabeth	Uruguay	2-1	South Korea	3
13	6/15/10 4:00 PM	Nelson Mandela Bay/Port Elizabeth	Côte d'Ivoire	0-0	Portugal	0
12	6/15/10 1:30 PM	Rustenburg	New Zealand	1-1	Slovakia	2
55	6/29/10 4:00 PM	Tshwane/Pretoria	Paraguay	5-3	Japan	8
41	6/24/10 4:00 PM	Johannesburg - JEP	Slovakia	3-2	Italy	5

Data Chart

Data Value	Frequency	Relative Frequency	Cumulative Relative Frequency
0	3	0.1000	0.1000
1	8	0.2667	0.3667
2	8	0.2667	0.6333
3	5	0.1667	0.8000
4	2	0.0667	0.8667
5	2	0.0667	0.9333
8	2	0.0667	1.0000

Histogram



Boxplot

