Statistical Analysis of Goals in FIFA World Cups: Who Scores More, Men or Women?

TL;DR: We compared the number of goals in men's and women's international matches since 2002. The analyses show that, on average, women score more goals per match than men.

The question I asked myself

Are there significant differences in the number of goals scored per match between men's and women's teams in FIFA World Cups?

I wanted to answer this with statistical evidence using tests that compare two independent groups.

After years of closely following men's football, and with the strong rise of women's football at the professional level, my intuition suggested that women's matches tend to have more goals than men's. However, what was just a feeling needed to be confirmed with data and a valid statistical test.

To reduce variability across years and different tournaments, I selected only official FIFA World Cup matches from 2002 onward.

For this purpose, I used a dataset containing all official men's and women's international match results dating back to the 1900s. The question we want to answer is clear:

Do women's international matches have more goals scored per game than men's matches?

To approach this rigorously, I set a significance level of 10% and formulated the hypotheses:

H_0 : The average number of goals per match in women's matches is equal to that of men's matches.

H_A : The average number of goals per match in women's matches is greater than that of men's matches.

Step 1 - Preparing the data

Before jumping into the analysis, we need to understand how the data is organized, what information is available, and whether anything needs to be cleaned or adjusted. First, we import the essential libraries:

- `pandas` to manipulate the data and explore match results.
- `matplotlib.pyplot` to **visualize the data** through histograms and detect patterns at a glance.
- `pingouin` to run the **Mann–Whitney U** test, which helps us compare men's and women's goals without assuming a normal distribution.
- `scipy.stats.ttest_ind` to perform the **t-test**, useful if the distribution is sufficiently close to normal.
- We also define `alpha = 0.1` as our **significance level**, the rule that will help us determine whether there is statistical evidence supporting our intuition.

Next, we load the men’s and women’s results datasets into pandas DataFrames, ensuring that each column has the appropriate data type. Finally, we take an initial look at the information using `.info()`. Here we want to understand:

- What columns the dataset contains (teams, goals, tournament, date, etc.).
- Whether there are null or inconsistent values that could affect the analysis.
- The total number of records, giving us an idea of the sample size we will be working with.

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import pingouin
from scipy.stats import ttest_ind
alpha = 0.1

men_results = pd.read_csv('men_results.csv', parse_dates=['date'])
women_results = pd.read_csv('women_results.csv', parse_dates=['date'])

men_results.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44353 entries, 0 to 44352
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  44353 non-null  int64
 1   date        44353 non-null  datetime64[ns]
 2   home_team   44353 non-null  object
 3   away_team   44353 non-null  object
 4   home_score  44353 non-null  int64
 5   away_score  44353 non-null  int64
 6   tournament  44353 non-null  object
dtypes: datetime64[ns](1), int64(3), object(3)
memory usage: 2.4+ MB
```

Step 2 – Filtering the Relevant Data

Now that we understand how the datasets are structured, it’s time to focus on what truly matters for our analysis: FIFA World Cup matches from 2002 onward.To achieve this, we apply a two-step filter:

1. The match date must be later than January 1st, 2002, ensuring we analyze only recent and comparable data.
2. The tournament must be "FIFA World Cup", excluding friendlies, continental tournaments, and qualification phases.

This allows us to concentrate exclusively on important and representative matches, reducing noise that could distort our conclusions.

Below, we display the first 10 rows of the filtered datasets to visualize how the information looks after applying these filters.

In [ ]:
```
men_results = men_results[(men_results['date']> '2002-01-01') & (men_results['tournament'] == 'FIFA World Cup')]
women_results = women_results[(women_results['date']> '2002-01-01') & (women_results['tournament'] == 'FIFA World Cup')]

men_results.head(10)
```

Out[ ]:

	Unnamed: 0	date	home_team	away_team	home_score	away_score	tournament
25164	25164	2002-05-31	France	Senegal	0	1	FIFA World Cup
25165	25165	2002-06-01	Germany	Saudi Arabia	8	0	FIFA World Cup
25166	25166	2002-06-01	Republic of Ireland	Cameroon	1	1	FIFA World Cup
25167	25167	2002-06-01	Uruguay	Denmark	1	2	FIFA World Cup
25168	25168	2002-06-02	Argentina	Nigeria	1	0	FIFA World Cup
25169	25169	2002-06-02	England	Sweden	1	1	FIFA World Cup
25170	25170	2002-06-02	Paraguay	South Africa	2	2	FIFA World Cup
25171	25171	2002-06-02	Spain	Slovenia	3	1	FIFA World Cup
25172	25172	2002-06-03	Brazil	Turkey	2	1	FIFA World Cup
25173	25173	2002-06-03	Croatia	Mexico	0	1	FIFA World Cup

So far, we have the goals for each team separately: **home_score** and **away_score**.

However, our main question is **how many goals are scored in total per match**, regardless of which team scored them.

For this reason, we created a new column, **total_goals**, which sums the home and away goals in each game. This column will serve as the basis for all subsequent analyses, from histograms to statistical tests. To make sure everything worked correctly, we visualized the first rows of the filtered dataset with this new column.

In [ ]:
```
men_results['total_goals'] = men_results['home_score'] + men_results['away_score']
women_results['total_goals'] = women_results['home_score'] + women_results['away_score']

men_results.head()
```

Out[ ]:

	Unnamed: 0	date	home_team	away_team	home_score	away_score	tournament	total_goals
25164	25164	2002-05-31	France	Senegal	0	1	FIFA World Cup	1
25165	25165	2002-06-01	Germany	Saudi Arabia	8	0	FIFA World Cup	8
25166	25166	2002-06-01	Republic of Ireland	Cameroon	1	1	FIFA World Cup	2
25167	25167	2002-06-01	Uruguay	Denmark	1	2	FIFA World Cup	3
25168	25168	2002-06-02	Argentina	Nigeria	1	0	FIFA World Cup	1

Before visualizing goals per match, we need to understand the distribution of the data. Using `.describe()`, we obtain key statistics such as:

- The **maximum number of goals per match**, which will help us define the histogram `bins`. These are the intervals used to show how many matches had each number of goals.
- Averages, medians, and quartiles, which give us an idea of how goals generally behave.

This information will be important for choosing an appropriate range in the charts, so that the visualizations accurately reflect the match statistics without losing detail.

In [ ]:
```
men_results.describe()
women_results.describe()
```

Out[ ]:

	Unnamed: 0	home_score	away_score	total_goals
count	200.000000	200.000000	200.000000	200.000000
mean	3094.485000	1.805000	1.175000	2.980000
std	1010.682192	1.937977	1.289453	2.022387
min	1600.000000	0.000000	0.000000	0.000000
25%	2155.750000	1.000000	0.000000	2.000000
50%	3429.500000	1.000000	1.000000	3.000000
75%	4418.250000	2.000000	2.000000	4.000000
max	4469.000000	13.000000	7.000000	13.000000

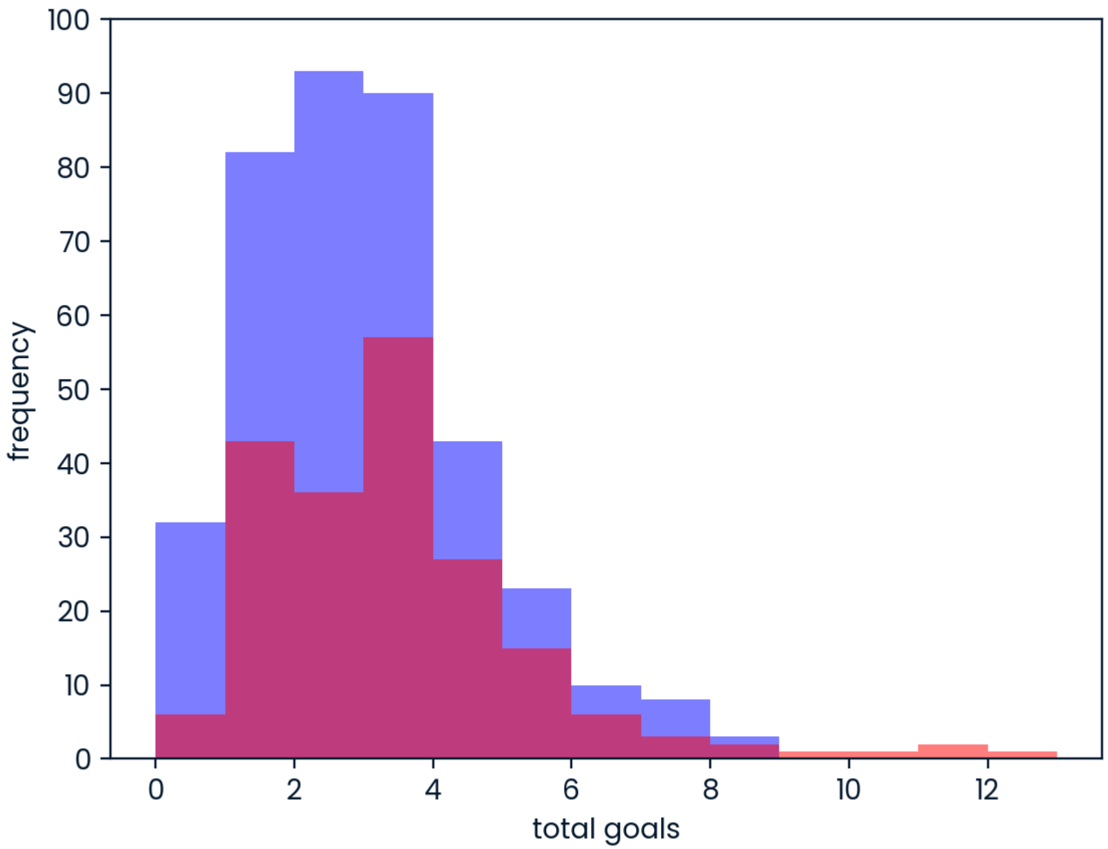
Step 3 – Distribution Analysis and Parametric vs. Non-Parametric Tests

To understand how goals are distributed, the next step was to visualize both datasets in a histogram. We chose a bin range from 0 to 13 goals—enough to cover all possible values—and overlaid the histograms for men and women to facilitate easy comparison.

Although the distributions are not perfectly symmetrical (which is expected for goal data, as goals tend to be low and discrete), the number of available matches is large enough (n≈200 for women and n≈380 for men) to consider that a parametric test could work robustly. Even so, this visualization reminds us that it is useful to complement the analysis with non-parametric tests.

In [ ]:

```
bins = range(0,14)
plt.hist(men_results['total_goals'], bins= bins, alpha=0.5, color='blue')
plt.hist(women_results['total_goals'], bins= bins, alpha=0.5, color='red')
plt.xlabel('total goals')
plt.ylabel('frequency')
plt.yticks(range(0,101,10))
plt.show()
```



Although the histogram shows a somewhat skewed distribution, we have a large sample size in both cases, which allows for a robust application of a t-test. However, since goals are discrete data and not strictly normal, we decided to complement the analysis with a non-parametric test, specifically the Mann–Whitney U (`.mwu()`), which does not assume normality and compares the distributions directly.

In [ ]:

```
results = pingouin.mwu(x=women_results['total_goals'], y=men_results['total_goals'], alternative='greater')
p_val_mwu = results['p-val'].iloc[0]
p_val_mwu = float(p_val_mwu)
print('p_val: ', p_val_mwu)
print('alpha: ', alpha)

if p_val_mwu < alpha:
    print('Reject the null hypothesis with an alpha of 0.1')
```

```
else:
    print('Fail to reject the null hypothesis with an alpha of 0.1')
```

```
p_val:  0.005106609825443641
alpha:  0.1
Reject the null hypothesis with an alpha of 0.1
```

Since the Mann–Whitney U test yielded a p-value < alpha, the null hypothesis is rejected at a 10% significance level.

Therefore, by rejecting the null hypothesis, we conclude that **women score more goals per match than men**, according to this dataset and with 90% confidence.

Next, we will analyze the same dataset under the same question, but using a t-test that assumes normality.

In [ ]:
```
t_stat, p_val_ttest = ttest_ind(
    women_results['total_goals'],
    men_results['total_goals'],
    equal_var=False, # Asumimos que las varianzas no son iguales (Welch's Test)
    alternative='greater')

print('p_val: ', p_val_ttest)
print('alpha: ', alpha)

if p_val_ttest < alpha:
    print('Reject the null hypothesis with an alpha of 0.1')
else:
    print('Fail to reject the null hypothesis with an alpha of 0.1')
```

```
p_val:  0.0025980724004871503
alpha:  0.1
Reject the null hypothesis with an alpha of 0.1
```

Since the Welch's t-test (normal distribution with unequal variances) yielded a p-value < alpha, the null hypothesis is rejected at a 10% significance level.

Therefore, by rejecting the null hypothesis, we conclude that **women score more goals per match than men**, according to this dataset and with 90% confidence.

Conclusion

The statistical analysis confirmed what had long seemed like little more than a journalistic hunch: international women's matches have, on average, more goals than men's matches. Both the parametric test (t-test) and the non-parametric test (Mann–Whitney U) agreed on the same verdict: the p-value was below our 10% significance level, so we rejected the null hypothesis.

In simple terms:

- According to FIFA World Cup data since 2002, women's matches tend to have a higher number of goals.

> 📌 Author of the Analysis
> Francisco Ara — Sports Data Analyst & Football Scout
>
> You can contact me via email at ffranciscoara@gmail.com ✉️
>
> Check out my profile on LinkedIn 🔗
>
> Or view my Portfolio Personal 💼