Pilot Study on Female Athletes Age and Position in the FIFA Women's Soccer World Cup Using Power Analysis

Sarah Bender

MISERICORDIA UNIVERSITY.

PURPOSE OF ANALYSIS

The purpose of the analysis is to use this FIFA Women's soccer pilot study to determine the smallest sample size of players needed for a future study. In this study, the analysis that I am conducting is on professional women soccer players from various countries. This test of two sample means will determine whether the population mean of players ages are significantly different by the position that they play on the field. It is important to include these characteristics of this fast-paced sport of soccer to look at the difference between the age that these athletes are continuing to contend in their assigned positions. From this association of these two crucial variables, it is important to compare the mean of how old these elite players are who are deciding to continue their careers. The overall intention of this statistical analysis is to use this pilot study to find additional research on professional female athletes competing in this advanced soccer league and dominating additional championships. Changes in athletics and other aspects of life are required because "For years, players around the world have demanded better pay and more support from their federations. In some countries, serious progress has been made; in many others, the fight continues" (McCann & Syam, 2019). In other words, this analysis can be incorporated into this completed study to find additional information where the values will be determined theoretically.

DESCRIPTION OF PILOT STUDY DATA

In this pilot study, these professional female soccer players belong to different nations who compete against each other in the FIFA World Cup. Overall, there were 24 countries included in this study as they produced some of the most prominent athletes of this sport that compete at the highest national level. Equal samples of the larger population of females were collected by first determining if the player made an impact by playing on the offensive end or identified as the defensive spot. This data can be obtained freely from an online database source from FIFA so that all individuals can access these statistics to conduct different analyses. Within this data, the variables that will specifically be used is the assigned field position and the age of each player. The field position that is assigned is either designated as 'O' defined as offense and 'D' defined as defense. The offensive spot is considered as an athlete who plays in the Midfield of Forward field position. Additional players on the team's roster in the defensive field position are either the Goal Keeper or Defender.

ANALYSIS

A power analysis is computed below so that the effect size and the smallest sample size needed for future studies using this method will be determined. In this power analysis with the program R, a t-test of balanced groups was conducted which came from the experimental design of the total population. Given the specified power, type 1 error which is the confidence level, and effect size, we will overall determine the required sample size from the original data. The sample size for both variables that are being considered, age and a player's field position, was 552 different individuals. It will be known that the main method that I will be using is the built-in power function. Therefore, the primary goal will then be to find the sample size of intended players from all countries associated with their professional team. Specifically, this calculation is of independent samples with considering the sample size in each group, the t-test will look at two-samples, as this power calculation can look at one, two, or paired samples of data. First to examine the given data, the effect size known as Cohen's d, type of effect size between two means, which is the strength of the evidence is further computed with other set values. These set values include the significance level established for this design which is 0.05 as well as the power which will be set at 0.80. But overall this effect size will measure the strength of the differences between the means of two populations which will be based on the data from the sample. Once the effect size is determined, the value of d will either be labeled as a small, medium, or large value which will then be factored into finding the value of power for this t-test of means. A small value of effect size denoted as d is 0.2, the medium value is equal to 0.5, and the large size is 0.8. With this effect size, it also determines the 95% confidence interval that gives the lower and upper bounds. Then as the value of d will be equal the Cohen's d calculation of equal group sizes, the sample size needed can then be calculated.

$$Cohen's\ d = \frac{(M_2 - M_1)}{Pooled\ SD}$$

$$Pooled\ SD = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}$$

M_1 = The mean of group 1 (e.g. control group)
M_2 = The mean of group 2 (e.g. experimental group)
SD_1 = The standard deviation of group 1
SD_2 = The standard deviation of group 2

```
Cohen's d

d estimate: 0.2092877 (small)
95 percent confidence interval:
      lower      upper
0.04069122 0.37788426


        Two-sample t test power calculation

              n = 359.3472
              d = 0.2092877
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

A power calculation for a two-sample t-test was then composed that contains the estimated value of Cohen's d, the set significance level of 0.05, and the value of power. Lastly, the computing power of tests and a way to determine parameters by obtaining a target power was calculated for this study of professional athletes. The code displayed above identifies the sample size needed which must be rounded up and then considered for the number of groups, that is two, which are being tested within the analysis.

RESULTS

After analyzing the results from the pilot study, the result of this power analysis is that for one of the groups, the required sample size is 360 players. Then for the built-in power calculation, the d estimate of Cohen's d was designated as small and equal to the value 0.2093. This calculated effect size since it close to Cohen's set standard of being 0.2, it will then be considered as a small effect size. With this set confidence interval for this study, the lower bound was calculated to be 0.0407 and the upper bound of the data is 0.3779. Then to further address the power calculation of this two-sample t-test, it is concluded that the total number of sample size that is required for a future study would be 720 professional female soccer players.

CONCLUSIONS

Once the results were produced from the analysis from the original study, it can be explained that additional information on the sample sizes required for future studies can be found through this process. These two variables that were compared, age and field position, between the different players with various capabilities in the sport, are important to consider and understand how the team is needed to progress and win championships. In years to come, further studies will be needed to be conducted as the ages of each player will increase. Even then, it is also valuable to consider how over time some athletes may change their field position as the team faces significant changes and other challenges. As my initial question indicated that a sample size needed to be found with this analysis of the data, it concluded that more individuals would be required than the pilot study. With this data, other ways of finding the improvement and effectiveness of players for each nation can be found as issues are still prominent within the league. This analysis allows for additional research to be conducted so that further changes can be made to a roster and determine the age ranges of individuals that are included in these professional teams.

REFERENCES

- Bradburn, S. (2019). What is and how to calculate cohen's d? Retrieved from https://toptipbio.com/cohens-d/
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863
- McCann, A., & Syam, U. (2019, June 7). We asked, they answered: 108 women's world cup players on their jobs, money and sacrificing everything. Retrieved from https://www.nytimes.com/interactive/2019/06/07/sports/soccer/world-cup-survey.html