This statistics worksheet examines the goals and shots in two prominent lacrosse leagues: the Premier Lacrosse League (PLL) and the National Lacrosse League (NLL). The PLL and NLL are highly regarded professional lacrosse leagues that feature top-tier athletes from around the world. The NLL is a box lacrosse league that is played in an indoor, smaller field while the PLL is a field lacrosse league that plays on an outdoor field with larger dimensions. This worksheet will provide an analysis of goal-scoring within these leagues. The data is collected from the PLL and NLL website for all games in the 2021-2022 season and is stored in the file **lacrosse\_pll\_nll\_2021-2022.csv**. The *League* variable codes whether each game was played in the NLL or PLL and the total goals scored (both teams) is in *goals*.

**Descriptive Statistics for Goals:**

|  |  |  |  |
| --- | --- | --- | --- |
| **League** | **N** | **Mean** | **StDev** |
| NLL | 126 | 22.262 | 4.847 |
| PLL | 36 | 23.889 | 4.496 |

1. Identify both the variables in this data. For any numerical variables, denote the units and for categorical variables, list the levels.

*League – Levels: NLL/PLL– Categorical*

*Number of Goals Scored- Units: Goals per game - Numerical*

1. Suppose you are interested in using these data to determine if there is statistically discernible evidence that, on average, goals scored in the PLL and NLL differ.
	1. Identify the statistical inference procedure that would be appropriate to answer this research question.

*Depending on the instructor, several methods may be chosen. e,g., Welch’s t-test, Pooled t-test, or difference in means randomization test*

*These solutions will be based on the Welch’s t-test.*

* 1. Define appropriate parameters of interest that could be used to address the research question based on the procedure you identified in the previous part.

 *and represent the average number of goals scored in the PLL and NLL leagues, respectively.*

* 1. List any assumptions that must be made about these data and inference procedure you will use. Assess them and explain any weaknesses in them.

*The key assumptions are*

*Independence: The two samples must be independent of each other. This means that the data collected from one sample does not influence the data collected from the other sample.*

* *Note that this assumption is potentially suspect given that many of the PLL players also play in the NLL. However, the current structure of the data doesn’t allow us to investigate the impact of this.*

*Normality: We need a sufficiently large sample size for the Central Limit Theorem to work. In this case we have a reasonably large sample size for each group – especially given the approximately symmetric distributions for goals in each league.*

* 1. Record the null and alternative hypotheses associated with the research question.

Ho: Ha:

* 1. Calculate the appropriate test statistic and p-value.

p-value = P(|T| > 1.88) = 0.065 with df = 60

* 1. Using the p-value, provide an appropriate conclusion for the research question. Be sure to provide full context of the application.

*At best, there is weak evidence to suggest that the PLL and NLL average goals differ from one another (.*

Upon further analysis of the leagues, you discover that the NLL and the PLL play their games for different amounts of time. The PLL plays games for 48 minutes, while the NLL’s games are 60 minutes long. In an effort to correct this mistake, we will “scale down” the NLL goals to a 48 minute rate.

1. Briefly explain how you could convert the data for the NLL games from *Goals* into a “scaled down” version of *Goals per 48 mins*.

*Although these columns are provided with the data, students should still think about how data can be easily re-expressed. In this case, we could simply multiple each of the PLL goals by 0.8 (i.e., 48/60)*

1. The output below show the numerical and graphical summaries of the Goals per 48 Minutes for NLL and PLL. Use this information to help answer the following questions.

**Descriptive Statistics for Goals (per 48 minutes):**

|  |  |  |  |
| --- | --- | --- | --- |
| **League** | **N** | **Mean** | **StDev** |
| NLL | 126 | 17.810 | 3.877 |
| PLL | 36 | 23.889 | 4.496 |

* 1. Do the data provide evidence that, on average, goals scored per 48 minutes differ between the PLL and NLL? Include all the pieces of the appropriate hypothesis test. (Tip: Use a similar process as in Question 2)

*Here are the updated pieces.*

 *mean goals per 48 minutes for all NLL/PLL lacrosse leagues*

*Ho: Ha:*

There is very strong evidence to suggest that the PLL and NLL average goals (in 48 minute games) differ from one another (

Test Statistic

p-value =P(|T|>7.367) = 0.000 with df = 50

* 1. In many scenarios, it is useful to extend a conclusion of a hypothesis test by including a confidence interval with the results. Calculate and interpret a 95% confidence interval for the difference in average goals per 48 minutes for NLL and PLL lacrosse games. (Tip: Your interpretation should discuss how much more, on average, PLL scores.)

**Estimation for Difference**

|  |  |
| --- | --- |
| **Difference** | **95% CI forDifference** |
| 6.079 | (4.422, 7.737) |

*With 95% confidence, on average, the PLL is expected to have between 4.404 and 7.754 more goals per 48 minutes than the NLL.*

*Student may also notice that* ***zero*** *is not contained in the confidence interval, which would logically agree with the result from the hypothesis test.*

*Instructors may wish to combine the conclusion for these two parts together to produce a better flow. For example,*

*There is very strong evidence to suggest that the PLL and NLL average goals (in 48 minute games) differ from one another ( Further, on average (and with 95% confidence), the PLL is expected to have between 4.404 and 7.754 more goals per 48 minutes than the NLL.*