# MAT 243 Project Two Summary Report

Nicholas Cleveland

nicholas.cleveland1@snhu.edu

Southern New Hampshire University

## Introduction: Problem Statement

The problem that I am going to solve in this report is the analysis of basketball teams from the NBA in a statistical manner. By doing this, performance patterns will be uncovered and interpreted. This will be done by using various hypothesis tests to validate all claims. These statistical analysis functions were used in the Python programming language with various packages.

## Introduction: Your Team and the Assigned Team

The team that I picked was the Cleveland Cavaliers, the same team that I chose for the last project, in order to keep consistency and discover more information about them. I was again given the years of 2013-2015. The team that I was assigned for a comparative study was again the Chicago Bulls for the range of 1996 – 1998.

We start with the same data frame that we calculated for both teams. These tables display the first five observations of each team:

**Cleveland Cavaliers (Our team):**

Graphical user interface, text, application

Description automatically generated

**Chicago Bulls (Assigned team):**

Graphical user interface, table

Description automatically generated with medium confidence

|  | **Name of Team** | **Years Picked** |
| --- | --- | --- |
| 1. Yours | Cleveland Cavaliers | 2013 – 2015 |
| 2. Assigned | Chicago Bulls | 1996- 1998 |

Table 1. Information on the Teams

## Hypothesis Test for the Population Mean (I)

*Suppose a relative skill level of 1342 represents a critically low skill level in the league. The management of your team has hypothesized that the average relative skill level of your team is greater than 1342. You tested this claim using a 5% level of significance. For this test, you assumed that the population standard deviation for relative skill level is unknown. Explain the steps you took to test this problem and interpret your results.*

Hypothesis testing is used to test claims about a population mean in that it processes a claim, and then wages that claim against the alternative by using a proof and comparative statistical method. The proof is that if we have a null hypothesis, we must have an alternative to it. It cannot be both. The proof usually sets the null hypothesis as the opposite of what the researcher wants, and the alternative hypothesis being what they do want. This is generally measured in one of three different ways. The first is when given a situation where we have a null hypothesis and an alternative hypothesis that is the opposite of the null hypothesis. An example of this is where the null hypothesis is true, denoted by H0: u = u0. The alternative of this hypothesis is then Ha: u ≠ u0 (two-tailed). The second and third situations is when we are given a null hypothesis that is assumed to be true, denoted as H0: u = u0 and the alternative hypothesis can be either u is greater than or less than u0. This is denoted as Ha: u > u0 (for right tailed) and Ha: u < u0 (for left tailed.)

These are generally described as H0: u = u0 where Ha: u ≠ u0, Ha: u > u0, or Ha: u < u0. One important factor here is that when we analyze this with a test for a p-value, the p-value can be a **one-tailed or two-tailed** p-value. This is decided by looking at our alternative hypothesis. If we are testing an Ha: of where u ≠ u0, we use a **two-tailed** p-value. If we are testing a Ha: where u > u0 or u < u0, we use a **one-tailed** p-value.

***Null Hypothesis***

The null hypothesis in this case is that the average relative skill of our team is 1342*.* This is described as H0: u = u0.

***Alternative Hypothesis***

The alternative hypothesis is that the average relative skill of our team is greater than 1342. This is described as Ha: u > u0.

***Level of Significance***

The level of significance is an alpha of 0.05 or 5%.

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 13.98 |
| P-value | 0.0000 |

Table 2: Hypothesis Test for the Population Mean (I)

The following data was generated from our Python script:

Text

Description automatically generated

From these findings, we calculated a test-statistic of 13.98, with a one-tailed p-value of about 0.0000. The p-value has been rounded to 4-decimal places for simplicity. Since our p-value is less than that of the alpha, 0.05, there is significant evidence to reject the null hypothesis and favor the alternative hypothesis. If we were to look at the calculated population mean value, we get a value of 1427.57, which favors the concluded alternate hypothesis of that our skill level was indeed greater than 1342

## Hypothesis Test for the Population Mean (II)

*Your team’s coach has hypothesized that average number of points scored by your team in the team’s years is less than 110 points. For this test, you assumed that the population standard deviation for points scored is unknown. You tested the claim using a 1% level of significance. Explain the steps you took to test this problem and interpret your results.*

***Null Hypothesis***

The null hypothesis in this case is that the average number of points scored by our team in the teams’ years is 110 points. This is described as H0: u = u0

***Alternative Hypothesis***

The alternative hypothesis that we are looking at to study is that the average number of points scored by our team in their years is less than 110 points. This is described as Ha: u < u0

***Level of Significance***

The level of significance here is 0.01 or 1%.

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | -13.85 |
| P-value | 0.0000 |

Table 3: Hypothesis Test for the Population Mean (II)

The following data was generated from our Python script:

A picture containing text

Description automatically generated

We calculated our one-tailed p-value here to be approximately 0.0000, rounded to four decimal places. With this p-value, since it is less than our alpha of 0.01 or level of significance, we can say that there is significant evidence to reject the null hypothesis and to favor the alternative hypothesis. The alternative hypothesis is that we scored less than 110 points, to which if we compare this to our calculated mean score, we get a score of 99.28, which is indeed less than 110. We can say this result favored our accepted alternative hypothesis.

## Hypothesis Test for the Population Proportion

*Suppose the management claims that the proportion of games that your team wins when scoring 80 or more points is 0.50. You tested this claim using a 5% level of significance. Explain the steps you took to test this problem and interpret your results.*

In general, a hypothesis test used to test claims about a population proportion will calculate a test statistic as well as a p-value. This p-value will be a decimal the represents a proportion, that will be compared against another proportion, in this case our alpha.

***Null Hypothesis***

The null hypothesis in this case is that the proportion is 0.50, given the team scored 80 points or more. This is described as H0: u = u0.

***Alternative Hypothesis***

The alternative hypothesis is that the proportion is less than 0.50. This is described as Ha: u < u0.

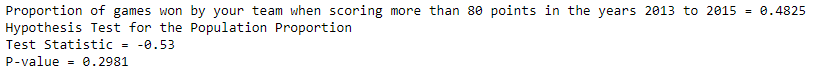
***Level of Significance***

The level of significance is 0.05 or 5%.

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | -0.53*.* |
| P-value | 0.2981 |

Table 4: Hypothesis Test for the Population Proportion

The following data was generated from our Python script:



After calculating our z-test, we were given back a test statistic of -0.53 and a two-tailed p-value of 0.5962. Since we are comparing if the proportion is less than the given value of 0.50, we need to take a one-tailed p-test, in which we cut this value in half and get a p-value of **0.2981**. Since the p-value, 0.2981 is greater than our level of significance, 0.05, we conclude that we fail to reject the null hypothesis. Our rounded proportion of games won by scoring 80 or more points during the 2013 to 2015 seasons was a rounded value of 0.4825, which is very close to the null hypothesis value of 0.50.

## Hypothesis Test for the Difference Between Two Population Means

*You were asked to compare your team’s skill level (from its years) with the assigned team’s skill level (from the assigned time frame). You tested the claim that the skill level of your team is the same as the skill level of the assigned team, using a 1% level of significance.*

In general, to test claims about the difference between two population means, you would calculate an independent t-test, as we did for our example. This gives the correct test-statistic and p-value. You would also then compare each of the calculate means to support the conclusion you found given the p-value compared to the alpha value.

***Null Hypothesis***

The null hypothesis in this case is that the skill level of our team, the Cleveland Cavaliers during 2013-2015, is **the same** as the skill level of the assigned team, the Chicago Bulls during 1996-1998. This is described as H0: u = u0.

***Alternative Hypothesis***

The alternative hypothesis is that the skill level of our team, the Cleveland Cavaliers during 2013-2015, is **not the same** as the skill level of the assigned team, the Chicago Bulls during 1996-1998. This is described as Ha: u ≠ u0.

***Level of Significance***

The level of significance is an alpha of 0.01 or 1%.

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 5.61 |
| P-value | 0.0000 |

Table 5: Hypothesis Test for the Difference Between Two Population Means

The following data was generated from our Python script:

Text

Description automatically generated

After calculating the t-test, and getting a test-statistic of 5.61, and a two-tailed p-value of 0.0000 (rounded), if we compare this p-value to the alpha or level of significance of 0.01, we see that it is less than it, justifying strong evidence to reject the null hypothesis and favor the alternative hypothesis of the **skill levels not being the same**. If we look at our calculated means of relative skill for both teams, we get rounded values of 1739.80 for the Chicago Bulls and 1427.57 for the Cleveland Cavaliers. These values are indeed not the same, which supports the conclusion of rejecting the null hypothesis and favoring the alternative hypothesis.

## Conclusion

In conclusion, there were multiple hypotheses tests that were conducted to explain the observances of the calculated data. We first looked at the statistics of our team, the Cleveland Cavaliers, and specifically looked at and compared the skill level of the team to the expected skill level. We then found the test-statistic and p-value. The p-value supported the claim that we could reject the null hypothesis of the relative skill level of team being 1342, and siding with the alternative hypothesis that our skill level was high than this. The mean data generated supported this claim, as it was indeed higher than 1342.

We then looked at the number of points that our team, the Cavaliers, was scoring. The hypothesized claim was that the team was scoring less than an average of 110 points. The null hypothesis here was set as our team scoring 110 points, and the alternative hypothesis being that our team scored less than 110 points. After writing our script, we observed an output with a test statistic of -13.85 and a p-value of approximately 0.0000. Since the alternative case is a less than case, we needed to take a one-tailed p-value from the default two-tailed p-value, but it is still approximated at 0.0000. Since our alpha value was 0.01 or 1%, our p-value was less than that, which rejects the null hypothesis and favors the alternative hypothesis. This alternative hypothesis was that the team scored less than 110 points, which lined up with the results from our mean calculation during those seasons of approximately 99.28, which is less than 110.

We then looked at population proportions and decided to look at how our team, the Cavaliers, proportioned in terms of score points to see it was 0.50, or scoring 80 points or more.

The null hypothesis being that our proportion was 0.50 and the alternative being that it was less than 0.50. For this case, we used a proportions z-test, that took in inputs of the number of games won with over 80 points, the total number of games when our team scored over 80 points, the claimed value, and the proportion variable. This gave us a test statistic of -0.53 and a two tailed p-value of 0.5962. Since our alternative hypothesis is comparing the value being less than 0.50, we took the one-tailed p-value of 0.2981 and used this. Our alpha was 0.05 and since 0.2981 is greater than this, we did not have sufficient evidence to reject the null hypothesis. The result from the mean that was calculated was 0.4825, which is very close to 0.50, which tells me that alternative hypothesis would have worked out favorably if we just compared the means and not the conclusion from the p-value comparison. This hints me into believing there could be some amount of error in the calculation that was done.

We then looked at the hypothesis test for the difference between population means for both basketball teams given their corresponding seasons. This was done with an in**dependent t-test** in our python code, as these results do not rely on one another. The null hypothesis here was that both teams had the same skill level and the alternative hypothesis being that they did not have the same skill level. After calculating our independent t-test, we received a test-statistic of 5.61 and a two-tailed p-value of approximately 0.000. This function gives us a two-tailed p-value, in which we want to keep since our alternative hypothesis is the direct opposite of our null hypothesis. H0: u = u0 and Ha: u ≠ u0. Our given alpha for this situation was 0.01 or 1%. Since our two-tailed p-value was indeed less than this, we chose to reject the null hypothesis and favor the alternative hypothesis. The alternative hypothesis was that the teams do not have the same mean skill level, which checks out with the mean results calculated for each team where we got 1739.8 for the Bull from 1996-1998 and 1427.57 for the Cavaliers from 2013-2015, in which these skill levels are indeed different and not the same.