

# Introductory Statistics Project Assignment

## Introduction

The National Hockey League (NHL) is by far the most competitive and highest-quality hockey league in the world today, and has been so for decades. With new young players being recruited from all over the world, a question that is always essential is where to look for the best players – the future stars of the league. Do they come from the many college teams across the United States; or are they found playing junior hockey in Canada or in other countries?

One way to answer this question is to analyze the average playing time of rookie players, and thereby determine, on a Time On Ice (TOI) basis, whether there is a noticeable difference between players coming from the different leagues. The average playing time shows, among many things, how the rookies were adopted by their coaches and how well they were able to adapt on the ice to their new teammates, tempo, and tactics of the NHL, compared to their old leagues.

Consequently, the purpose of our statistics project was to look at first-year players in the National Hockey League (NHL), and see if they came from a college league or a junior league - in terms of where they played previous to their rookie year in the NHL, and then determine if there was a difference in average playing time between the two groups of players.

## Methods

We found our data in the official NHL data base at the website [www.nhl.com](http://www.nhl.com). There, we could isolate the first-year players and gather all the available information about them in our own MS Excel database. The number of rookies was not extensive enough for us to be forced to make random samples, but instead we could use the whole group of players for our study. Further, we isolated the variables of interest, which were their previous league, and average TOI.

We used the program R for making our calculations, which included a Levene's test, a Two-Sample T-Test, and a box plot to illustrate the results.

## Results

We found that the total number of first-year players in the NHL during the 2006-2007 season was 207. These players were distributed as follows:

College background: 39 or 18.84 % out of the total of 207 players.

Junior or international background: 168 or 81.16% out of the total of 207 players

We then performed a hypothesis test as follows:

1. A quantitative variable (TOI) was measured on two independent populations (junior and college), thus, a two-sample t-test is required.
2. The hypothesis will be about comparing the differences of two means.

Thus,  $H_0: \mu_j = \mu_c$ , where j represents Juniors and C represents College. The alternative hypothesis is  $H_a: \mu_j \neq \mu_c$

3. We set alpha to 5%, or 0.05
4. The data is part of an observational study (see explanation above under "Method") and was not randomly selected.
5. The two populations are independent. The sample size is greater than 40. Therefore, the test statistic should reasonably follow a t-distribution with 205 DF ( $n_J + n_C - 2$ ). The equal variances formula will be used because the p-value for the Levene's test of the homogeneity of variance is "large" (i.e.  $> \alpha = 0.05$ )
6.  $\bar{x}_J - \bar{x}_C = 87,69643 - 107,41026 = -19,7138$
7. The t-test statistic is -2.1491 with 205 df.
8. P-value is 0.03280
9. Reject  $H_0$  because alpha is greater than p-value.
10. The mean TOI appears to differ among the two groups of players. In fact, it appears that the mean TOI is greater for college players than players from junior leagues.
11. A two-sided 95% confidence interval is warranted for this situation. Thus, one is 95% confident that players from Junior leagues are get between 1.628465 and 37.799191 less minutes of playing time, on average, per season than do players from college teams.

## Discussion

From our study we can conclude that the average playing time (TOI) differed between the two groups of players, and that it was the rookies with a college background that got the most playing time during the 2006-2007 NHL season.

We think it was fairly easy to find good data from the data base we chose for our study, and we are satisfied with the results. We were actually counting on a different result, as we have heard that players from junior and international leagues are more highly rated than college players, so this survey gives an interesting pointer in the opposite direction. However, there are more factors to a player's success than the playing time of his first year, but it is in general a very important figure.

For future surveys we would not like to do anything different; we got good, reliable data from our source, and we believe that the calculations were easy to perform with the help of R.

Official Website of the National Hockey League.

<http://www.nhl.com/nhlstats/app?service=page&page=Stats&fetchKey=20072ALLSRSAll&viewName=timeOnce&sort=timeOnce&pg=1>

Site viewed on the 12/8/2007.

Site last updated: after the season of 2006/2007

## Data from R

```
> hockey<-read.table("C:/Documents and Settings/Administratör/Mina  
dokument/statistik/statsprojecttoi.txt",header=T)
```

```
> hockey
```

```
> hockey$JC<-factor(hockey$Junior.College,levels=c("j","c"))
```

```
> hockey$JC
```

```
> hockey$toi.g<-c(hockey$TOI.G)
```

```
> attach(hockey)
```

```
> Levenes(toi.g~JC)
```

Levene's Test for Homogeneity of Variance

	Df	F	value	Pr(>F)
group	1	2.3584	0.1261	
	205			

```
> t.test(toi.g~JC,alt="two.sided",var.equal=T, conf.level=0.95)
```

Two Sample t-test

data: toi.g by JC

t = -2.1491, df = 205, p-value = 0.03280

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-37.799191 -1.628465

sample estimates:

mean in group j mean in group c

87.69643	107.41026
----------	-----------

```
> tapply(toi.g, JC, FUN=Summary)
```

```
$j
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev
1.00000	41.00000	84.50000	87.70000	132.20000	183.00000	52.88727

```
$c
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev
20.00000	74.50000	113.00000	107.40000	143.50000	182.00000	45.55663

```
> boxplot(toi.g~JC, ylab="Time on ice", xlab="Junior/College")
```

