|  |  |
| --- | --- |
| A table with numbers and symbols  Description automatically generated  Probability & Statistics: Project 2 Final Report  A DEEP DIVE INTO NHL PLAYER STATISTICS | Abstract  Hello and welcome to my project, throughout this detailed report I dive into the world of National Hockey League (NHL) player statistics. From goals and assists to time on ice and advanced analytics, my goal with this project is to gain insight on player performance beyond the surface level. Whether you're a hockey fan, fantasy league player, or just a data science enthusiast, join me as I take a comprehensive look at the numbers that tell the story of NHL players and how they impact the game. This report is not just a statistical analysis but a journey into the heart of the game, celebrating the exceptional athletes who shape the National Hockey League (NHL).  Christopher Ward  Probability and Applied Statistics |

Top of Form

Contents

[**Chapter 1 – What Is Statistics?** 3](#_Toc184766420)

[Section 1.2 – Characterizing a Set of Measurements: Graphical Methods 3](#_Toc184766421)

[Section 1.3 – Characterizing a Set of Measurements: Numerical Methods 3](#_Toc184766422)

[**Chapter 2 – Probability** 4](#_Toc184766423)

[Section 2.3 – A Review of Set Notation 4](#_Toc184766424)

[Section 2.4 – A Probabilistic Model for an Experiment: The Discrete Case 5](#_Toc184766425)

[Table 1: 5](#_Toc184766426)

[Table 2: 5](#_Toc184766427)

[Section 2.5 – Calculating the Probability of an Event: The Sample-Point Method 6](#_Toc184766428)

[Section 2.6 – Tools for Counting Sample Points 6](#_Toc184766429)

[Section 2.7 – Conditional Probability and the Independence of Events 6](#_Toc184766430)

[Section 2.8 – Two Laws of Probability 7](#_Toc184766431)

[Section 2.10 – The Law of Total Probability and Bayes’ Rule 7](#_Toc184766432)

[**Chapter 3: Discrete Random Variables and Their Probability Distributions** 8](#_Toc184766433)

[Section 3.2 – The Probability Distributions for a Discrete Random Variable 8](#_Toc184766434)

[Section 3.4 – The Binomial Probability Distribution 9](#_Toc184766435)

[Section 3.5 – The Geometric Probability Distribution 9](#_Toc184766436)

[Section 3.6 – The Negative Binomial Probability Distribution 9](#_Toc184766437)

[Section 3.7- The Hypergeometric Probability Distribution 10](#_Toc184766438)

[Section 3.8 – The Poisson Probability Distribution 10](#_Toc184766439)

[Section 3.11 – Tchebysheff’s Theorem 11](#_Toc184766440)

[**Chapter 4: Continuous Variables and Their Probability Distributions** 11](#_Toc184766441)

[Section 4.4 – The Uniform Probability Distribution 11](#_Toc184766442)

[Section 4.5 – The Normal Probability Distribution 12](#_Toc184766443)

[Section 4.6 – The Gamma (Exponential) Probability 13](#_Toc184766444)

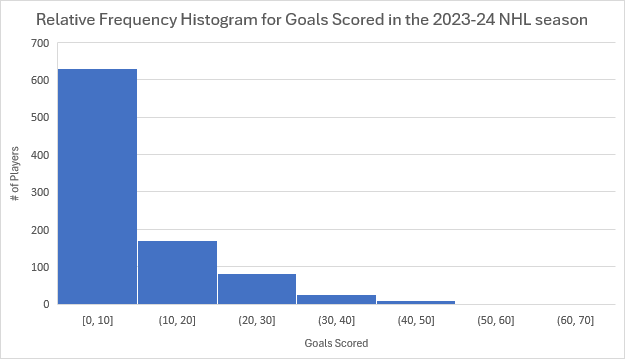
[**Chapter 5: Multivariate Probability Distributions** 13](#_Toc184766445)

[Section 5.2 – Bivariate and Multivariate Probability Distributions 13](#_Toc184766446)

# **Chapter 1 – What Is Statistics?**

## Section 1.2 – Characterizing a Set of Measurements: Graphical Methods

The Rocket Richard trophy winner (most regular season goals) for the 2023-24 NHL Season was Auston Matthews, who scored 69 goals in 81 games. To see how frequent players scored 60 or more goals we can create a relative frequency histogram for goals scored in the 2023-24 season.



It is impossible to see, but Auston Matthew with his 69 goals is the only player in the [60,70] range, and there are barely any players in the [50,60] goal range. Based on this data it would be fair to conclude that scoring 69 goals is an extremely hard feat in the NHL, but what is the actual percentage?   
Well, we answered it right there, only 1 player in the 60-70 range out of 924 players.  
So, or 0.1082% percent of skaters in the NHL scored between [60,70] goals.

## Section 1.3 – Characterizing a Set of Measurements: Numerical Methods

In the NHL, a player is awarded a ‘point’ to their individual stats if they record either a goal, an assist, or a secondary assist. Now let’s evaluate the average number of points scored by any player in the 2023-24 NHL Season.   
To calculate this, you would sum the total number of points scored by all skaters and divide that by the total number of skaters that hit the ice last year, or: . Using my java stats library for computation, we can see that the mean for points scored by any skater in the 2023-24 season is: **23.49 points**.

A black screen with white text

Description automatically generated

Now let’s focus on the top point scorer in the 2023-24 NHL Season, Nikita Kucherov, who scored 144 points last year. That’s extremely high compared to the mean we just saw for the whole league. However, there is a very good explanation for this. Many players only play a few games each season due to circumstances such as injury, which also forces teams to pull up minor league players to play a limited number of games. To counteract this, we filter the data to only players that played at least half the regular season (41+ games). But what if instead we wanted just the top 100 point-scorers that season, then what would the mean be?



With the proper filters set, we can see the average within only the top 100-point scorers was **70.98 points,** compared to the total mean which was **23.49 points**. Also, the average for players who played at least half the regular season (41+ games) was **34 points**. So, the top 100 players in the league are averaging about 71 points each, which is slightly under a point per game, or ~ .866-point percentage (82-game regular season). Meanwhile, players who have played most of the season average around **34 points**.

# **Chapter 2 – Probability**

## Section 2.3 – A Review of Set Notation

Next let’s look at individual players. Last season both Nikita Kucherov and Connor McDavid had 100 assists. This is an incredible feat considering just hitting the 100-point mark is a huge milestone for many players, so to hit it with only one stat is incredible. Suppose you wanted to know just how difficult it is for a player to have over 100 assists in a season. Well among the 924 players that played at least one game in the 2022-23 season, we know only 2 players had over 100 assists. What about within players who have over 100 points?

A screenshot of a computer program

Description automatically generated

(n = # of players, a100 = players with 100+ assists, p­100 = players with 100+ points)

Find the number of players who:

1. Had over 100 assists.  
   A: We need to solve for the intersection of players (n) and players with 100+ A’s (a100)

1. Had over 100 points.  
   A:
2. Both had over 100 assists and over 100 points.  
   A:

## Section 2.4 – A Probabilistic Model for an Experiment: The Discrete Case

Auston Matthews of the Toronto Maple Leafs scored 69 goals on 369 shots in the 23-24 season. That means he scores on ~18.7% of his shots. If he takes two shots, the four possible simple events are given by the table below. Find the probability that he will:

### Table 1:

|  |  |  |  |
| --- | --- | --- | --- |
| Shot # | Outcome of 1st Shot | Outcome of 2nd Shot | Probability |
| S­1 | Goal | Goal | .035 |
| S2 | Goal | Save | .152 |
| S3 | Save | Goal | .152 |
| S4 | Save | Save | .661 |

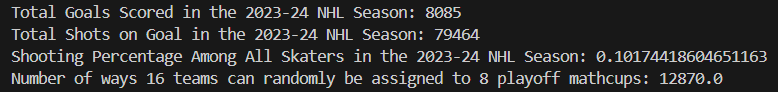
1. Score both shots.  
   A:
2. Score on the first shot but not the second.  
   A:
3. Miss the first shot and score the second.  
   A:
4. Miss both shots.  
   A:
5. Score on at least one of the two shots.  
   A:

What does this information tell us? Well, there’s a 3.5% chance that Auston Matthews scores twice on consecutive shots; a 15.2% chance he scores the first shot but misses the second; 15.2% chance he misses the first shot but scores the second; a 66.1% he misses both shots; and a 33.9% chance he scores on at least one of the two shots.

Now that we’ve seen Auston’s stats, what about the probabilities of the next two shots going in, in any random NHL game? The table below has the four possible events and their probabilities:

### Table 2:

|  |  |  |  |
| --- | --- | --- | --- |
| Shot # | Outcome of 1st Shot | Outcome of 2nd Shot | Probability |
| S­1 | Goal | Goal | .01 |
| S2 | Goal | Save | .09 |
| S3 | Save | Goal | .09 |
| S4 | Save | Save | .81 |



Using what my program calculated for the total goals, shots, and the shooting percentage for all skaters, I filled in the table. From this we can see that Auston Matthews h and every player in the league: ~1% chance of goals on consecutive shots, and >75% chance of saves on consecutive shots.

## Section 2.5 – Calculating the Probability of an Event: The Sample-Point Method

In the NHL, if a game is not decided in regulation time, there is 5 minutes of 3 on 3 overtime to decide a winner, shootout to follow if still undecided. Suppose an NHL team randomly selected the 3 skaters for overtime out of the 18 skaters on the roster that night. It’s most common for teams to dress 12 forwards and 6 defensemen for each game.

1. Describe one sample point. Assume that you need describe only the three chosen and not the order in which they are selected.  
   A: One sample point in this set would be the formation most NHL teams typically play in overtime, 2 forwards with 1 defenseman.
2. List the sample space associated with this experiment.  
   A: The sample space (F=forward, D=defense)  
   [F,F,F], [F,F,D], [F,D,D], [D,D,D] (order here doesn’t matter)
3. What is the probability that all three skaters selected are forwards?  
   A: The number of ways to select 3 forwards is the number of ways we can select 3 forwards divided by the number of ways we can select 3 skaters from the roster. So,  
      
   So, there is a 26.9% that a team randomly chooses three forwards to play in overtime.

## Section 2.6 – Tools for Counting Sample Points

In 1979-80, the NHL expanded its playoff structure to include 16 teams. Originally, all 16 teams were pooled together and ranked 1-16 based on regular-season record ([NHL Records](https://records.nhl.com/history/playoff-formats)). Suppose the NHL instated a new rule that in the first round the teams were now randomly assigned to games 1-16, instead of the traditional seeding based on regular season standings. With this new style of seeding, how many ways can the 16 teams be assigned to the 8 matchups?

A: Typically, we could just use our combination formula: . However, since we are selecting teams in groups of 2, we need to alter our numerator to choose 2 teams each time. Remember we are solving for how many ways you can choose 8, 1v1 matchups out of 16 teams which is how we get: . Of those 16 teams we need to choose 2 at a time, not just one which is what 16! would calculate if we just plugged into the combination formula. So, we end up with:

So, there are over 2 million ways to randomly assign these teams to a 16-team playoff.

## Section 2.7 – Conditional Probability and the Independence of Events

Referring to Section 2.4, the shooting percentage among all skaters in the 2023-24 NHL season was ~0.10 percent. So, consider two events a Goal and a Save, are given that:

, , what is ? Are the events independent or dependent? Explain.

A: The intersection of two events A and B is:

If A and B are independent, then the intersection is:

To figure out which formula to use, we must ask whether events A and B are independent or dependent of one another. Well, if a shot is taken and scored, that shot cannot be a save. Likewise, if a shot is saved, that shot cannot be a goal. The two events, A and B, are mutually exclusive. So, we can conclude that the events are dependent on each other’s outcome and the correct formula is: .

## Section 2.8 – Two Laws of Probability

Referring to Section 2.7, there is a ~0.1 percent chance of a player scoring a goal on any single shot attempt. Now, A friend claims if there is a 10% chance (1/10) of scoring on a single shot, then there must be a 100% of scoring after 10 shots. Is your friend correct? Why or why not?  
A: We determined in the last section that the outcome of any given shot is dependent. However, here we are looking at a sequence of shots [S1, S2, …, S10]. So, we can say the shots are independent of one another since a shot does not make the following shot in a game any more likely to go in. Since the shots are independent, every shot has a 1/10 chance of going in, or 9/10 chance of not going in, and can calculated by doing the probability that you score at least once in 10 shots or:   
This means there is about a 65.1% chance that there will be at least one goal in 10 shots, so your friend is would be wrong.

## Section 2.10 – The Law of Total Probability and Bayes’ Rule

Part of the NHL’s new rules implemented in the 2023-24 season includes Rule 75.3: (Unsportsmanlike Conduct- Player Sitting on Boards) “The referee now will provide the offending team (coach and players) with one warning regarding players sitting on the boards (and will so advise the other team). After one warning in a game, the team precipitating the warning will be issued a bench minor penalty for future violations”([NHL](https://www.nhl.com/news/nhl-announces-rule-changes-for-next-season)).

Suppose 70% of the forwards react positively to this change, but only 30% of defensemen react positively. In the group of 924 NHL players to play at least 1 game in the 2022-23 season, 609 were forwards and 315 were defensemen. One player publicly responded about the rule change from the 924 players. What is the probability that player was a defenseman?

A: The probability of the player being a defenseman, given they reacted negatively can be found  
by: , where:  
B = Probability the player chosen reacts negatively =   
A = or 34% that a player is a defensemen  
*probability that a player has reacted negatively given they are a defenseman.*   
So, or 54.5% chance that the player is a defenseman

It makes sense that more than 50% of the time a defenseman would be the one complaining about the new rule, since this rule tends to be favorable for the forwards. As players get ready to come onto the ice, sometimes they sit on top of the boards for quicker changes. This is effectively banned due to the new rule change, which may favor forwards since zone entries are easier if the defensemen take longer to get on the ice. (However, 70% and 30% are made up statistics, this is just a ‘what if’ scenario).

# **Chapter 3: Discrete Random Variables and Their Probability Distributions**

## Section 3.2 – The Probability Distributions for a Discrete Random Variable

The top 5 point scorers of the 2023-24 NHL season had point totals of 107, 110, 120, 140, and 144 points. Two of the players are randomly selected from the five, and their point totals noted. Find the probability distribution for the following:

1. The *largest* of the two sampled numbers

First, we must list all possible pairs of point totals, and we have , without replacement since order of the pairs doesn’t matter here:

Next, we must calculate the largest for each pair:

|  |  |
| --- | --- |
| Pair | Largest |
|  | 110 |
|  | 120 |
|  | 140 |
|  | 144 |
|  | 120 |
|  | 140 |
|  | 144 |
|  | 140 |
|  | 144 |
|  | 144 |

Finally count the frequency of each largest number:  
, so .

1. The *sum* of the two sampled numbers

Repeating the same steps as part a except for the sum of points, creating our table first:

|  |  |
| --- | --- |
| Pair | Sum |
|  | 227 |
|  | 227 |
|  | 247 |
|  | 251 |
|  | 220 |
|  | 250 |
|  | 254 |
|  | 260 |
|  | 264 |
|  | 284 |

Again, counting the frequency of each sum:

## Section 3.4 – The Binomial Probability Distribution

Among the 924 players, 11.8% did not have a point on the 2023-24 regular season. If a sample of 5 players were randomly selected, what is the probability of finding at least one player who did not score a point the entire season?  
A: Here we can use the binomial distribution function: and since we are only trying to solve to finding one player, we can use the complement of 1 minus the probability that no player in the set of 5 scored zero points that season.

46.6% of the time you will find a player without a point, among the 5 that are randomly picked.

A black background with white text

Description automatically generated

I thought I made a miscalculation at first because I put this problem through my stats library binomial distribution solver, the result was .533, or 53.3% chance you’d find a player without any points. That didn’t seem right to me, but I quickly realized I forgot to take the complement of my binomial distribution in the code and got the correct answer the second time around.

## Section 3.5 – The Geometric Probability Distribution

Referring to section 2.4: Table 2, the probability a shot on goal will go into the net in any given NHL game during the 2023-24 season is .1 or 10%.

1. What is the probability that the third shot on goal is the first to go in?  
   A: We can find the geometric distribution by using:

I imported my geometric distribution solver and got an 8.1% chance the third shot is first to go in the net.

1. If there were 20 shots in a period, what is the probability that all of them were saved?  
   A: I imported my geometric distribution solver and got an 11.11% chance that out of 20 shots in a period, every shot is saved.  
   A black background with white text

   Description automatically generated

## Section 3.6 – The Negative Binomial Probability Distribution

Referring to Section 3.4, 11.8% of players did not record a point in the 2022-23 NHL season. If players are selected randomly one at a time, what is the probability that the first player with a point will be found on:

1. the second trial?

A:

Or, 10.4% chance the first player with a point is found on the second trial.

1. On or before the third?  
   A:

Or, 31.38% chance the first player with a point is found on or before the fifth trial.

  
My program confirmed my results, returning ~10.4% and ~31.38% respectively.

## Section 3.7- The Hypergeometric Probability Distribution

The NHL decides this year, the players going to the All-Star games are going to be decided completely random. From the group of 924 players, 20 are randomly selected to go to the All-Star games. What is the probability that the 20 selected include all 5 top point scorers in the group of 924?

A: In this example, N = 924, n = 20, r = 5, and y = 5 since we want all 5 top point scorers.

Or there is a .0000279% chance that all top 5-point scorers are chosen for the All-Star Games, if the 20 players that get to go were chosen randomly and not voted in.

## Section 3.8 – The Poisson Probability Distribution



The New Jersey Devils scored a total of 242 goals in the 2023-24 NHL regular season, averaging 2.95 goals per game. Assuming the number of goals scored by the Devils in any game follows a Poisson distribution:

1. What is the probability the New Jersey Devils score exactly 3 goals in a game?  
   , or 22.3% chance the Devils score exactly 3 goals.
2. What is the probability they score at least 5 goals?

or a 17.6% chance they score at least 5 goals in a game.

1. In a standard length 82-game season, what is the number of games they are expected to score more than 4 goals?

The expected for a Poisson distribution is the same as the mean. We are looking for expected (so mean) of more than 4 goals (Y > 4), which we solved already in part b.  
 .  
So, we can multiply the probability they score more than 4 goals by the length of the season to get our expected number of games over 4 goals:

, or they are expected to score more than 4 goals in 14.432 games throughout an 82-game season.  
Realistically you’d round down and say 14 games, since you can’t play .432 of a game.

## Section 3.11 – Tchebysheff’s Theorem

A black background with white text

Description automatically generated  
During the 2023-24 NHL season, the avg number of shots per player and the standard deviation for average number of shots per player are: and .

Using Tchebysheff’s Theorem, find an interval that will include at least 1/3 (33%) of the shots per player in the league during the 2023-24 season.

Tchebysheff’s theorem says:

We need to find an interval with least of the values included, so we setup our ineqaulity as : , to solve for k we get:   
Referring to the theorem, the interval around our mean can be found by: .

Substituting in our mean, standard deviation, and newly found k, we get:

So, our interval that includes at least 33% of shots per player in the league during the 2023-24 NHL season is [-4.63, -176.63]. This is an unrealistic interval as no player can record negative shots on goal, but it gives us a general look at how shots on goal is distributed throughout the league.

# **Chapter 4: Continuous Variables and Their Probability Distributions**

## Section 4.4 – The Uniform Probability Distribution

There was a unique situation in the NHL during the 2023-34 season where a player played more games during the regular season than every other team player. But how is that possible? Well, this player was traded to a team who had played 1 less game than his former team and continued to play every game of the season. Assuming the games played by players in the NHL follows a uniform distribution between minimum (0) and maximum (83) games played.

1. Find . (1/2 regular season to typical full-length season)

The probability density function for uniform distribution is:   
Adapting that to our problem we get:

To get the cumulative distribution function we integrate f(y):

From this we can easily evaluate:  
   
There is a 39.4% chance that a randomly selected player had played between half a season (41 games) and a normal full season (82 games).

1. Calculate the expected variance and Standard Deviation of

, these are the formulas for E(Y) and V(Y) in a uniform distribution.

So, the average games played between 41 and 82 is 61.5, with a standard deviation of close to 12 games, if games played in the NHL followed a uniform distribuion.

1. Find .

To find this answer is simple since we know that from part a.   
There’s a 1.2% chance a randomly selected player from the 2023-24 season played more than 82 games. Realistically

## Section 4.5 – The Normal Probability Distribution

  
During the 2023-24 NHL season, the number of goals scored by players followed a normal distribution with a mean and standard deviation of goals goals. Using this information, find:

The Normal probability distribution is represented by the function:

, and that , so to solve for we normalize our bound 5 and 20.

Using Table 4 in the back of the textbook:

A table of numbers with a white background

Description automatically generated

,

, or a 50.27% chance a player scored between 5 and 20 goals in the 2023-24 NHL season.

## Section 4.6 – The Gamma (Exponential) Probability

Referring to Section 4.5, during the 2023-24 NHL season, the average number of goals per player is 8.75 goals. Assuming the number of goals scored by NHL players follows an exponential distribution, find:



, so , or 18% chance the number of goals an NHL player scored over 15 goals in the 2023-24 NHL season.

Following the same logic as part a, except this time we need since our bound is in the opposite direction as part a. So, , or a 43.5% chance the number of goals a player scored during the 2023-24 NHL season is under 5 goals.

# **Chapter 5: Multivariate Probability Distributions**

## Section 5.2 – Bivariate and Multivariate Probability Distributions

During the 2023-24 NHL season, player both players’ goals scored, and hits are tracked for every player. Let denote the number of players scoring 30 or more goals and let denote the number of players recording 80 or more hits. Suppose we randomly select a player, find the joint probability function for , and calculate the probability the randomly selected player:



1. Scored 50 or more goals ):

To do this we simply calculate

1. Recorded 80 or more hits :

Again for this we simple calculate

1. Both Scored 50 or more goals and recorded 80 or more hits :