**DBA3701**

*Final Project*



**Leonard-Samuel Platzer**

**A0201276N**

Table of Contents

[1. Introduction 3](#_Toc100005149)

[1.1. Problem Statement 3](#_Toc100005150)

[1.2. Motivation for Project 3](#_Toc100005151)

[2. Problem Statement 3](#_Toc100005152)

[2.1. Objective 3](#_Toc100005153)

[2.2. Methodology 3](#_Toc100005154)

[*2.2.1. Competition Category* 3](#_Toc100005155)

[*2.2.2. Horse Type* 4](#_Toc100005156)

[*2.2.3. Drainage* 5](#_Toc100005157)

[3. Data 5](#_Toc100005158)

[3.1. Tournament Details 5](#_Toc100005159)

[3.2. Location 6](#_Toc100005160)

[3.3. Revenue / Entry Fee per Horse / Allowance 6](#_Toc100005161)

[4. Model 6](#_Toc100005162)

[4.1. Variables 6](#_Toc100005163)

[4.2. Objective 7](#_Toc100005164)

[4.3. Constraints & Assumptions 7](#_Toc100005165)

[*4.3.1. Constraint 1* 7](#_Toc100005166)

[*4.3.2. Constraint 2* 8](#_Toc100005167)

[*4.3.3. Constraint 3* 8](#_Toc100005168)

[*4.3.4. Constraint 4* 8](#_Toc100005169)

[*4.3.5. Constraint 5* 9](#_Toc100005170)

[*4.3.6. Constraint 6* 9](#_Toc100005171)

[*4.3.7. Constraint 7* 10](#_Toc100005172)

[*4.3.8. Constraint 8* 10](#_Toc100005173)

[4.4. Insights 10](#_Toc100005174)

[*4.4.1. Scenario 1* 11](#_Toc100005175)

[*4.4.2. Scenario 2* 11](#_Toc100005176)

[*4.4.3. Scenario 3* 11](#_Toc100005177)

[4.5. Model Limitations 12](#_Toc100005178)

[*4.5.1. What is the real energy drainage?* 12](#_Toc100005179)

[*4.5.2. The starting point does not have to be home* 12](#_Toc100005180)

[*4.5.3. What is the probability that a horse will win?* 12](#_Toc100005181)

[*4.5.4. Multi-Horse Purpose* 13](#_Toc100005182)

[*4.5.5. Rider-Horse Combination* 13](#_Toc100005183)

[*4.5.6. Parameter Freedom* 13](#_Toc100005184)

[5. Conclusion 13](#_Toc100005185)

[6. References 14](#_Toc100005186)

# **1. Introduction**

## 1.1. Problem Statement

Horse show-jumping is a niche sector with sincere enthusiasts for the sport (some people claim it is no less addictive than certain drugs (White, 2009)). However, given the centric interest in the sport itself, there are few providers and innovations for optimization of administrative tasks necessary for competing and allowing managers and riders to make cost-efficient decisions. This project focuses on one specific use case which is: **Which horses should a manager of a horse stable compete in which competitions every year?** For managers, it is very critical to find the optimal tournament plan, as this sport is prone to high costs (entry fees, transportation costs, time consumption). Also, different to other sports, there are two sides to the story, which is you do not only have the person as an athlete, but also the horse, which performance is impacted by a myriad of factors. Furthermore, if a manager knows at the beginning of the year and best-possible plan, he or she can start registering everywhere not having to worry about being rejected entrance because of late entry registration. Therefore, the following project aims to allow managers to make better decisions in terms of their planning by using a Linear-Optimization model which is driven by real data from the web and qualitative data assumed estimated based on experience.

## 1.2. Motivation for Project

My motivation to pursue this project is 1) one of my best friends manages a horse stable and has recently been struggling in optimizing his tournament plan and 2) I myself have been in the sport before entering university (Rooky of the Year 2015, four-time participant in European Championships, Part of the National Team for 4 consecutive years) and have encountered this problem in the past.

# **2. Problem Statement**

## 2.1. Objective

The objective of this project is to maximize the potential profit a manager can receive by choosing the optimal mix of competitions he or she sends their horses and riders to. The model also takes into consideration the type of horse the manager intends to send to competitions in a given year and simulates how their performance can vary from year to year as well as factors such as transportation costs all based on real data.

## 2.2. Methodology

### *2.2.1. Competition Category*

To understand the problem in depth, it is important to examine the types of competitions by using one data point as an example.

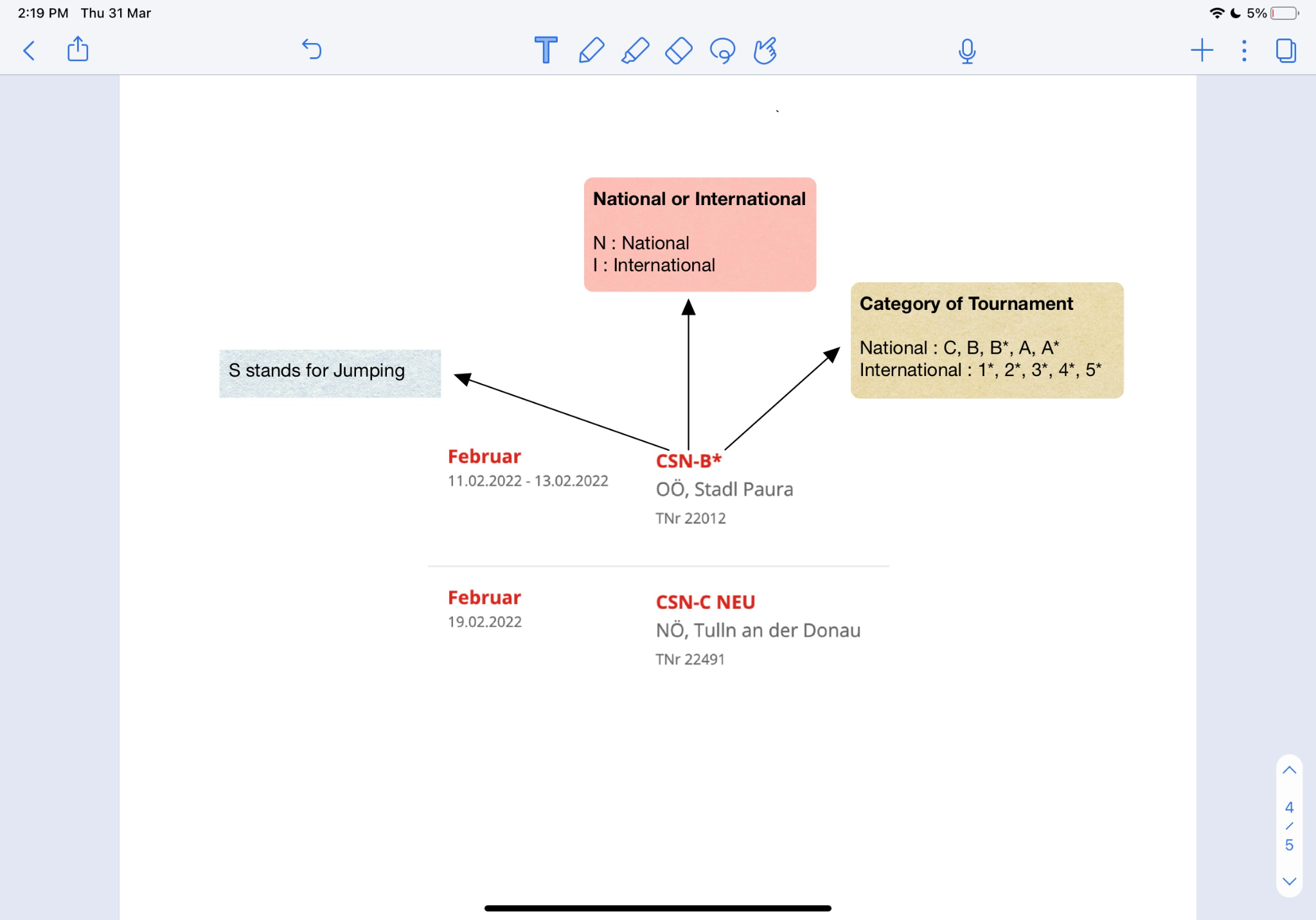


Figure 1 shows one competition on the official Tournament Website (OEPS, N.A.)

It is important to look at Figure 1 to understand what a competition category means, as it impacts the model. This project only focuses on show-jumping competition (which is the discipline where a horse jumps over obstacles) and these competitions are denoted by S on the second position of the keycode of the competition shown in Figure 1. There are two overarching types of competitions which are national and international ones, denoted by N and I respectively. The last part of the competition key is the sub-category. This defines the type of competition based on how difficult the courses (a sequence of fences to jump) are. The respective earnings and cost per horse increase proportionally to the difficulty or type of competition. For national tournaments the ordinality of difficulty is from C to A\* (where C is the easiest) and for international from 1\* to 5\* respectively.

### *2.2.2. Horse Type*

Because the model has to counter the assumption that every horse is the same, the user is able to choose the type of horses which will compete in the respective year. Given the rigid nature of tournaments, managers often decide which horse to send depending on the age of the horse. Therefore, age is used as a metric to determine the type.

|  |  |
| --- | --- |
| **Age (Years)** | **Type** |
| < 8 | 3 (Young) |
| [8, 10] | 2 (Runner-Up) |
| > 10 | 1 (Mature) |

Figure 2 shows the classification of each type of horse in the model

As shown in Figure 2, horses below the age of 8 are considered young because horses under that age have designated “Young horses” competitions classes, as well as young horse championships, are until the age of 7 (British Showjumping, N.A.). The best horses for show-jumping are between the age of 10 to 14, where they are considered mature (Worden, 2019). Everything between young and mature is considered a runner-up.

Based on which type the user chooses, the model will take into consideration how much each competition drains a horse when competing as well as which horse is allowed to compete in which competition category.

### *2.2.3. Drainage*

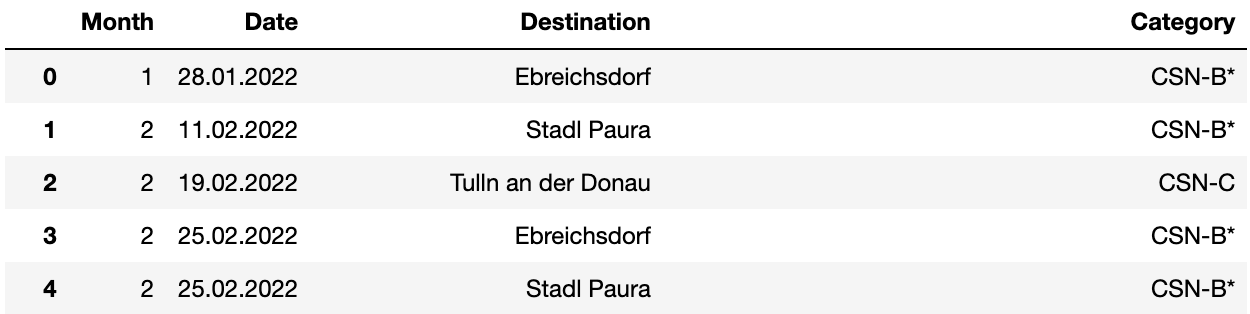
Drainage is understood as a factor that affects how many competitions a horse can do in a month, given that the model constraints the amount of drainage a horse can have every month. It is simply a number which represents how much physical energy a competition takes from a horse. The model assumes that for a young horse there is a higher deviation of drainage. Given that, for a horse which is not used to the routine, going on multiple competitions there are lots of other factors draining a horse besides competing such as transport, loud noise etc. As a result, the model simulates the greatest randomness for a young horse (60%), lesser for a “runner up” (40%), and the least randomness in drainage for a mature horse (10%). However, the base drainage is the same for all horses, given that a more difficult competition results in more drainage independent of which horse is competing.

# **3. Data**

The following will explain how the data which is fed into the model is composed and collected.

## 3.1. Tournament Details

I used BeautifulSoup4 and Selenium to scrap all possible tournaments in a given year from the Austrian Federation of Horse Sport. The selenium scrapper adjusts the settings to be only for show-jumping competitions and for the current year from January until December. The pandas Data frame will contain the Date, Destination and Category of the competition. As a result, the data will look like the following:



Data after Web-Scrapping from <https://www.oeps.at/de/termine>

## 3.2. Location

The location is used to estimate the fixed cost for going to a competition which is the transportation cost. For that, I first found the exact location in longitude and latitude using a Python API and then use a home location (which is my friend´s horse riding stable in Lower Austria) to map the driving distance for each competition. This value will then be calculated times two with the average gasoline price over the last 8 months in Austria (which is 1.59 Euros) to receive the final transportation cost per tournament, independent of how many horses will compete respectively.

## 3.3. Revenue / Entry Fee per Horse / Allowance

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Revenue/**  **Horse** | **Entry Fee/Horse** | **Drain** | **Mature** | **Runner-Up** | **Young** |
| CSN-C | 200 | 50 | 15 | Y | Y | Y |
| CSN-B | 300 | 100 | 20 | Y | Y | Y |
| CSN-A | 500 | 250 | 30 | Y | Y | Y |
| CS1\* | 500 | 350 | 25 | Y | Y | N |
| CSI2\* | 800 | 500 | 30 | Y | Y | N |
| CSI3\* | 1000 | 500 | 40 | Y | N | N |
| CSI4\* | 2000 | 700 | 70 | Y | N | N |

Figure 3 shows the assumed value taken from random samples of competitions

The data shown in Figure 3 is collected based on random samples of each competitions type, however, from my experience it is valid enough to use it as an approximation in a generalized model. Drain is a metric to define how much energy-draining each category is for a horse, given the difficulty it poses when competing. The data in blue represents the allowance for each horse to compete, which is considered by the model when choosing to compete or not.

# **4. Model**

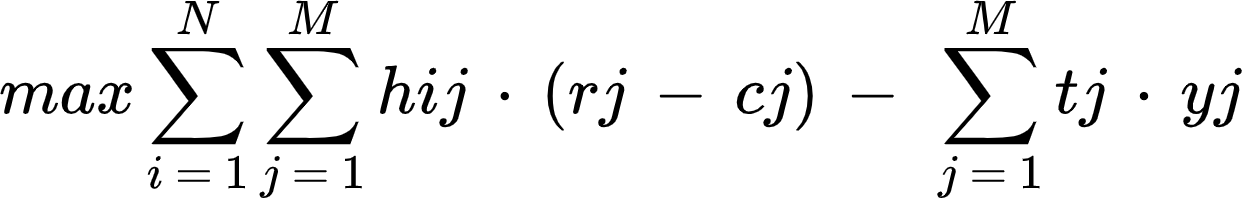
To use the model accurately, the data fed into it has to be sorted by date as some constraints assume it to be so and will only hold true as such.

## 4.1. Variables

Let

1. N be the number of horses which are intended to compete
2. M be the number of competitions one can possibly compete in during a given year
3. J,F,..., D be the indexes of data points in the model of January, February, …, December
4. X numbers of weekends with multiple competitions (multiple competitions on one weekend)
5. Ai, Bi, Ci,... where i = 1,2 until X numbers is reached to denote the indexes of the weekends with multiple competitions, where i=1 represents the first index in the data frame with a competition on a specific weekend and i=2 represents the last index in the dataframe of that respective weekend
6. hij be the decision variable of whether a horse is sent or not where i = 1, .., N and j = 1, …, M
7. dij be the drain for each specific horse for each competition where i = 1, .., N and j = 1, …, M
8. aij be the allowance for each specific horse for each competition where i = 1, .., N and j = 1, …, M
9. rj be the revenue per competition where j = 1, …, M
10. cj be the cost per competition per horse where j = 1, …, M
11. tj be the transportation cost per competition where j = 1, …, M
12. yj be the binary variable forcing transportation cost to apply when at least one horse competes
13. Horses, YearCompetitions, MonthDrain, MonthCompetitons are constants which are the maximum amounts of horses to compete in a competition, competitions the manager can compete, drain a horse can accumulate within a month and the number of competitions a horse can compete within a given month

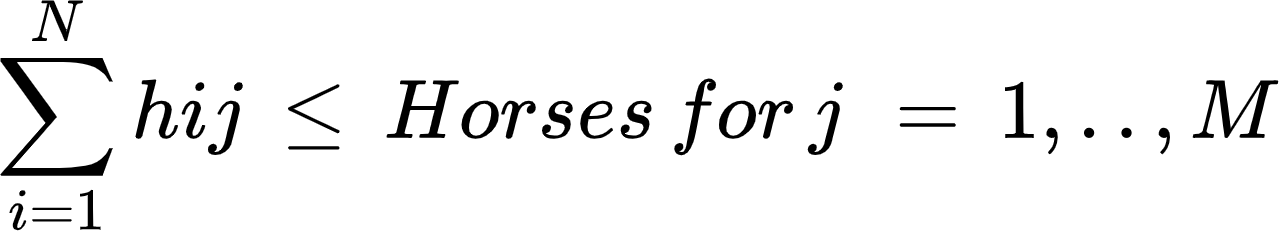
## 4.2. Objective



## 4.3. Constraints & Assumptions

### *4.3.1. Constraint 1*

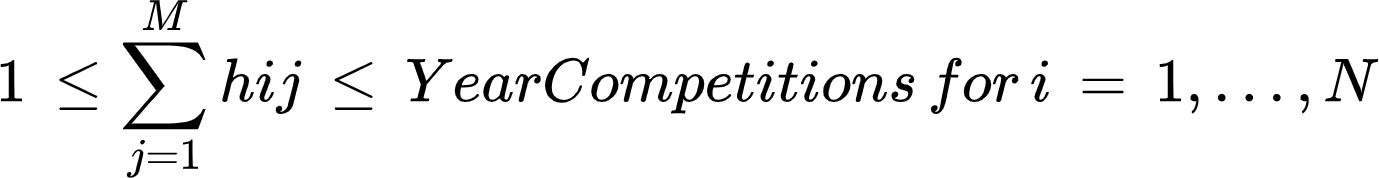
The first constraint is concerned with the number of horses that can compete in each competition. “Horses” is given by the user and is any number between 1 and N.



The assumption here is that the manager may not have an interest in sending all his horses to the same competition, because besides his interest in profit he would like to be present across more competitions, which could benefit his brand. If the manager chooses to put a lower limit of horses than N to compete in one competition the model will decide to compete in more competitions to maximize profit.

### *4.3.2. Constraint 2*

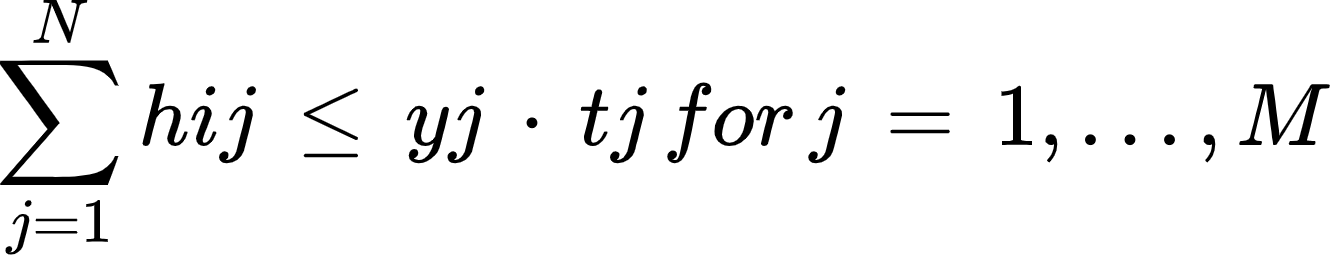
The second constraint limits the number of competitions a horse can compete in within a year. Here the model is currently assuming that the maximum is 24, or on average 2 competitions per month with a minimum of 1 competition every year. 24 is a very realistic number, as tournaments are hyperbolic in number throughout the year, where fewer tournaments are at the beginning and end of the year and many tournaments during the summer.



It is assumed that when the user inputs a horse, then he or she wants to compete with the horse. Secondly, if the horse on average competes twice a month in connection with the following constraints, it is a responsible and considerate policy to implement, as the manager also does not want to risk unfriendly behaviour towards the animal which could impact the performance of the horse as well as his company´s image.

### *4.3.3. Constraint 3*

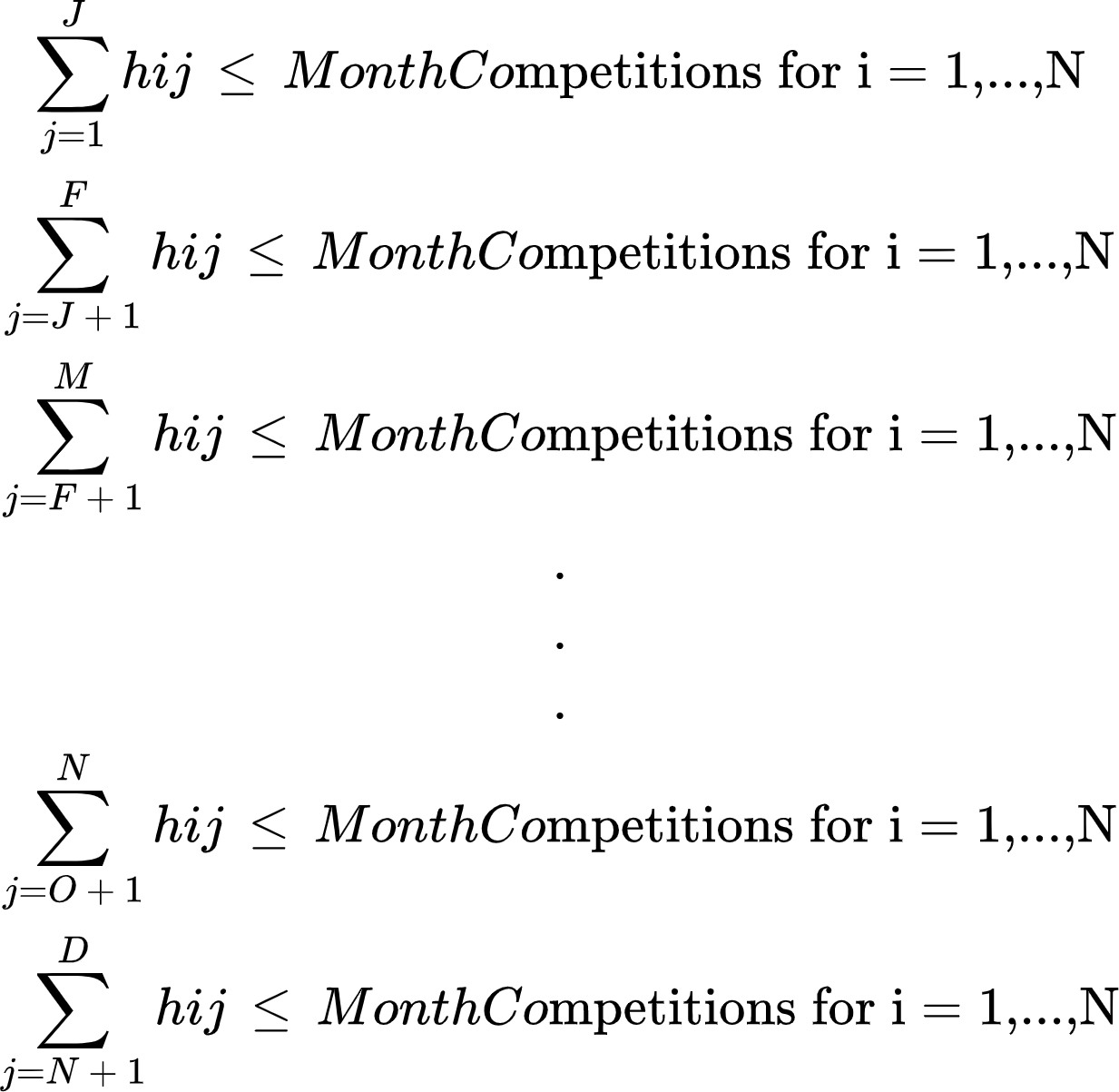
The third constraint is merely used to tell the model that when at least one horse is competing in the competition, then the transport cost applies to it as well.



For example, when the sum of hi1 is bigger than 1, then y1 has to be 1 forced by this constraint and as such t1 will be its value as well. Otherwise, given the objective function, the transport cost will be made zero given that the model aims to maximize profit.

### *4.3.4. Constraint 4*

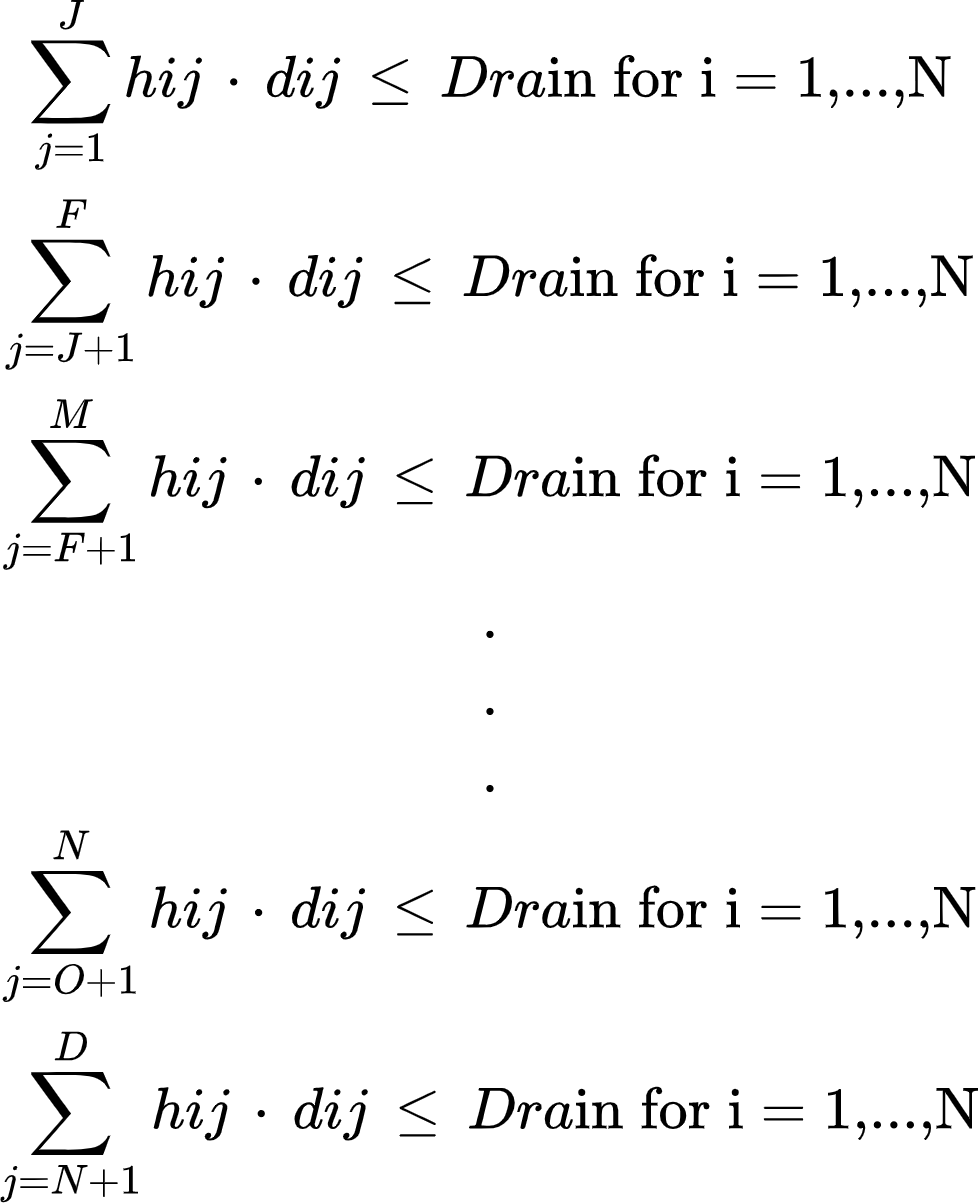
The fourth constraint limits the number of competitions a horse can compete in a month. Here J, F, M, …, D are indexes of the data points in the data which represent the last point of the month in a sorted data frame.



It is assumed that the model should prevent the model from putting a horse into e.g.: 24 consecutive tournaments in a year, which is unrealistic, given the stamina of the animal.

### *4.3.5. Constraint 5*

The fifth constraint concerns itself with how much drainage a horse can have each month. The drainage matrix is calculated based on the factor received from the type of horse as well as the randomness simulated throughout the year. However, “Drain” is constant.

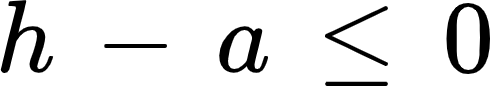


Currently, we assume “Drain” to be 100, as on average a horse should not compete in a 3\* and 4\* international competition within a month and in the case that it is competing in a 4\* star competition, it should at most compete in two C national competitions or one above.

### *4.3.6. Constraint 6*

The sixth constraint concerns itself with limiting a certain horse type to compete in a certain category at all. Programmatically the data is prepared in a manner so that there is an ordinality in all competition types. These integers are then mapped onto the data frame and according to the rule of the horse-type, the allowance is either 0 or 1 for this horse represented by aij.

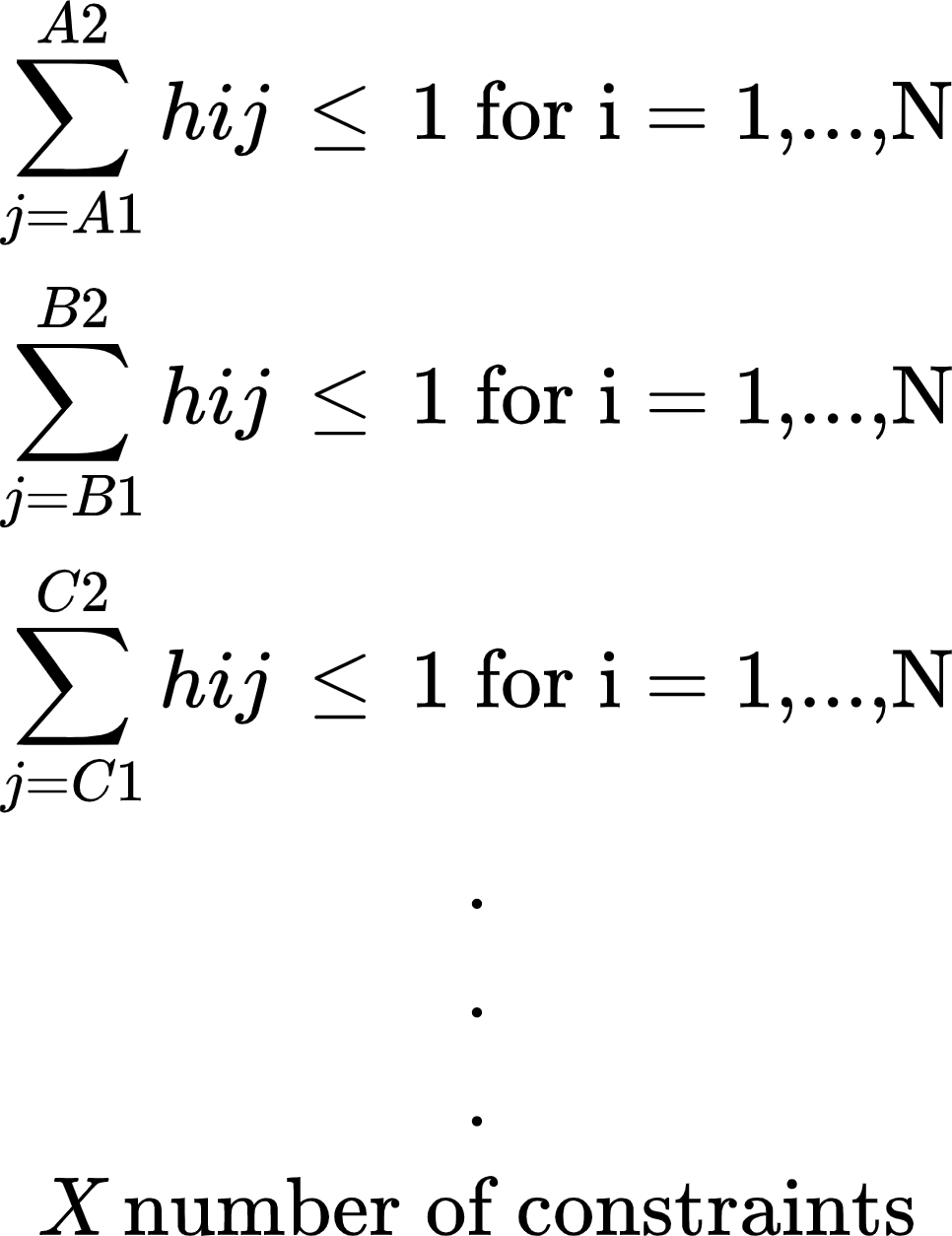
However, it is important to note that aij is not a decision variable, but is received by data pre-processing.



The model should allow h to be chosen freely if the allowance is 1, however, if the horse is not allowed to compete, h cannot be 1.

### *4.3.7. Constraint 7*

This constraint is highly essential to add to the realistic view of the model. If there are multiple competitions on one weekend, then one horse cannot compete in all of them. Therefore, the model has to constrain itself from choosing one horse to compete in multiple once at the same time. It is defined that competition is on the same weekend if they are three days apart.



### *4.3.8. Constraint 8*

Given that h is a decision variable choosing whether a horse should compete or not, I choose it to be binary. Furthermore, y is used to fix the transportation cost in the case at least one horse competes in the competition and hence is also binary.

*y, h are binaries*

## 4.4. Insights

To be able to view whether the model serves its purpose, I ran three different simulations.

I choose:

1. **6 mature horses**
2. **2 mature, 2 runner-ups and 2 young horses**
3. **6 young horses**

For all three scenarios we give the model complete freedom by allowing it to choose up to 6 horses to compete in each competition.

### *4.4.1. Scenario 1*

The expected output is that the model will recommend to compete in a variety of competition categories, but more importantly in the once with the highest payoff and furthermore has the resulting highest payoff from all three scenarios, given that there is the the least possible restriction by allowance as well as possible drainage.

|  |  |  |
| --- | --- | --- |
| Scenario 1 | | |
| Profit (Euros) | Competitions | Categories |
| 13156.99 | 12 | CSN-C, CSN-B\*, CSN-A\*, CSI2\*, CSI3\*, CSI4\* |

### *4.4.2. Scenario 2*

Comparing to Scenario 1 it is expected that the model will provide a lower payoff, given that there is more restriction on the allowance given the different horse-types. However, it is expected that in more competitions is competed in.

|  |  |  |
| --- | --- | --- |
| Scenario 2 | | |
| Profit (Euros) | Competitions | Categories |
| 11666.65 | 15 | CSN-C, CSN-B, CSN-B\*, CSN-A, CSN-A\*, CSI2\*, CSI3\*, CSI4\* |

What can be seen by the higher restriction, a greater variety of competitions is chosen, which is intended. The horses will compete on more national competitions, given that Runner Ups and Young horses are not as trained as mature horses. As a result, the model also shows a slightly lower payoff.

### *4.4.3. Scenario 3*

Scenario 3 should give the lowest expected payoff as a portfolio of purely young horses will only allow the manager to compete in lower categories with less prize money. Furthermore, only national competitions should be chosen based on the allowance restriction.

|  |  |  |
| --- | --- | --- |
| Scenario 3 | | |
| Profit (Euros) | Competitions | Categories |
| 4867.98 | 12 | CSN-C, CSN-B\*, CSN-A, CSN-A\* |

What can be seen is that the horses will only compete in national competitions such as the model is intended to optimize upon as well as gives the lowest payoff of all three scenarios.

## 4.5. Model Limitations

### *4.5.1. What is the real energy drainage?*

As it can be seen the model currently chooses about half of the possible competitions it can possibly choose from in all three scenarios. As a result, there is a lot of slack on Constraint 2. This is partially due to the monthly constraint; however, it is mainly because of the drainage value which is currently set to 100. 100 is a reasonable number given the proposed drainage values for each specific category. However, in reality, none of these values is static. At each competition, for example, there is a different course builder and different conditions. For example, a CSI3\* could possibly be as draining as a CSI4\* for a horse or a CSN-B\* as energy taking as a CSN-A\*. A horse can be injured and the manager does not realize it right at the moment which will incur much greater drainage for the horse independent of which competition it competes in. Furthermore, drainage is not independent of each competition as it is currently considered in the model. If a horse competes in consecutive tournaments, it may incur additional drainage, even though each individual competition has very low drainage. That is why the model would have to be fed with even more accurate data which could possibly be received from monthly or weekly medical examinations of each individual horse which in reality is also done. As a result, the model can be more accurate in giving the perfect expected payoff.

### *4.5.2. The starting point does not have to be home*

The model is currently assuming that the manager is sending horses only from his home stable. The model could optimize the plan even better if it could consider a network where it has a matrix of transportation costs of each competition and recommend the manager to instead of sending his horses from Home to A to Home and then to B from Home to A and then to B. Nevertheless, in reality, it really is the case that the horses are transported back and forth and the reason for that is because competitions are normally only on weekends. As such, the model would have to also take into consideration the additional costs of housing and feeding the horses for the days between two competitions as well as extra costs (e.g.: paying for a place to keep the truck, etc.).

### *4.5.3. What is the probability that a horse will win?*

The model is currently taking into consideration the average prize money a horse can win independently of the past performance of the specific horse. Animals are beings with unique emotions and capabilities. The script should scrap the horse from the internal database of the Show-Jumping Federation. As a result, for each individual horse the user chooses, the script can calculate a personalized probability of success and map it onto the actual average price of each category. However, that may not be directly linked as a model limitation but rather a data limitation.

### *4.5.4. Multi-Horse Purpose*

The model is currently only made for working with more than one horse. Hence, someone with only one horse cannot use this model. However, given the nature of the problem, this model is purely intended for managers with more than one horse.

### *4.5.5. Rider-Horse Combination*

The model is currently considering the number of riders and horses a manager has is in equality. However, in reality, a manager may have two riders with three horses each. As a result, these horses can only compete with this one rider which would change the current model by adding additional constraints.

### *4.5.6. Parameter Freedom*

The user is right now able to choose values of for example how many competitions a horse can compete in per year and what the maximum number of horses is to compete in a competition. However, that may reduce the value of the model. Sometimes the user may not know the optimal value. Nevertheless, freedom of choice may protect the user from unrealistic recommendations by the model. Therefore, giving freedom to the user to choose can be a limitation and benefit at the same time.

# **5. Conclusion**

The proposed model is a great benefit to any horse-show-jumping manager in the world, as it reduces their time for administrative tasks and maximizes their potential payoff. By upgrading the data source in multiple streams which dynamically feed the model, it can very accurately recommend the best possible tournament plan. However, limitations are present which have to be taken into consideration when using the model, as they imply a certain idea of what is happening in the circumstance of the user, which may not always equalize to reality. Depending on the size and intricacies of the qualitative management system of the respective manager, the model will add value to their operations.

# **6. References**

* White, C. (2009, February 20). *Anger at ‘ridiculous’ claims that ecstasy risk is equal to horseriding*. Horse&Hound. Retrieved March 30, 2022, from <https://www.horseandhound.co.uk/news/anger-at-ridiculous-claims-that-ecstasy-risk-is-equal-to-horseriding-277450>
* Worden, T. (2019, October 31). *Big Grand Prix classes in 2018: The year of the mare*. World of Showjumping. Retrieved March 30, 2022, from https://www.worldofshowjumping.com/en/Exclusives/WoSJ-Focus/Big-Grand-Prix-classes-in-2018-The-year-of-the-mare.html#:~:text=These%20days%2C%20the%20horses%20start,10%20and%2014%20years%20old.
* British Showjumping. (N.A., N.A. N.A.). *Young Horse Championships*. British Showjumping. Retrieved March 29, 2022, from https://www.britishshowjumping.co.uk/page/Young-Horse-Championships-
* OEPS. (N.A., N.A. N.A.). *TURNIERKALENDER INLAND*. OEPS. Retrieved March 31, 2022, from https://www.oeps.at/de/termine