SMDE Assignment

1st Carles Matoses
carles.matoses@estudiantat.upc.edu

2nd Ignasi Granell
ignasi.granell@estudiantat.upc.edu

3rd Isabel Castañeda
isabel.castaneda@estudiantat.upc.edu

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

Introduction

This document is a model and instructions for LaTeX. Please observe the conference page limits.

I. Problem description

We are provided with three marathon results, from a Kaggle dataset, on the years 2015, 2016 and 2017. The data is structured in a way that we have the runners' information, the time they took to get to interest points and the time they took to finish the race. Some of the most relevant variables we are provided are: the age, the gender and the city of origin. We believe this characteristics have a direct impact on the performance of the runners.

Reading the paper [1], we concluded that, effectively, environmental conditions have a direct impact on the performance of runners. The paper states that the temperature has a positive impact (longer races) and high humidity and high wind speed have a negative impact on the performance of the runners (faster races).

II. System description, introduction

The system to be modelled is the "Barcelona Marathon". This marathon involves runners of different skill levels, ages, and genders, coming from various cities. The event is influenced by environmental factors such as rainfall and temperature, all of which can impact runner performance.

Runners rely on essential supplies, including water, food, and medical assistance, to complete the marathon safely. The course is structured as a linear system, where participants progress from one checkpoint to the next. Along the way, they encounter resupply points, which have limited capacities. If demand exceeds supply, queues may form, potentially delaying runners. Environmental conditions also play a significant role in the system, affecting both individual runner performance and the overall flow of the event.

In summary, the system comprises the marathon course, checkpoints, and resupply points, along with the interactions between runners, resources, and environmental conditions.

III. Model specification

For this model, the entities and attributes are the following:

- **Runners**: The runners are the main entities in the system. They have attributes such as age, genre, and city of origin. They interact with the system by running the marathon.
- **Rainfall**: The rainfall is an environmental attribute that affects the performance of the runners.
- **Temperature**: The temperature is an environmental attribute that affects the performance of the runners.
- **Wind**: The wind is an environmental attribute .
- **Humidity**: The humidity is an environmental attribute.
- **Level**: The level of the runner is determined by the time it takes to finish the first 5. The standard deviation of runners dependens on they're level.

The operations of the model will be reduced to runners arriving at the end line. The processes of the model will be the following:

- **Generate runner**: The model will generate a runner .
- **Begins running**: The runner starts at the starting line and begins running.
- **Ends running**: The runner finishes the race.

In this process, the model will use a normal distributed random variable to determine the time it takes a runner to finish. We created different time ranges to decide the level of a runner. The "Elite" runners tend to run 5 km in less than 18 minutes, the "hobby" runners take more than 18 minutes and less than 25 and "new" runners take more than 25 minutes.

The model purpose differs from the system purpose. We aim to predict the consequences of the environmental conditions on the performance of the runners. The model will help us understand how the rainfall, heat, and other variables affects the performance of the runners.

A. Systemic Structural, Systemic Data and Simplifying Hypotheses

- **SH_01**: The runners will keep a constant speed over the race. Only the required time to finish the race will matter.
- **SH_02**: Heat, wind and rain will be combined into a single value computed by a formula.
- **SH_03**: Temperature wind and humidity change over the day, usually there is more wind later on the day than at morning, temperatures rise and humidity drops. The model will ignore this fact to eventually decreese as the sun gets down. The model will not account for this behaviour and will only take the average values of the day.

- **SS_01**: The runners will be split in three “level” groups based on the performance of the first 5 km.
- **SD_01**: The runners will not be affected by ethnics or age. An Elite runner with 70 years will be classified based on the time it takes to finish the first 5 km and not his status.
- **SD_02**: The time is transformed with a logarithm function to make it more normal. The results of the shapiro-wilk test and the figures 1 and 2 show that only elite runners follow a normal distribution. The model will ignore the outliers.
- **SD_03**: Temperature humidity and wind speed will be combined into three categorical variables (low, medium, high) based on the paper [1] shown on table I.

TABLE I
ENVIRONMENTAL IMPACT ON MARATHON PERFORMANCE

	MAN: mean, std. dev.	**WOMAN: mean, std. dev.**
Humidity		
low	0, 0	0, 0
medium	-1.73, 0.22	-1.74, 0.28
high	-2.11, 0.2	-0.90, 0.34
Temperature		
low	0, 0	0, 0
medium	1.19, 0.19	1.38, 0.23
high	7.73, 0.18	7.78, 0.38
Wind Speed		
low	0, 0	0, 0
medium	-3.18, 0.2	-1.90, 0.32
high	-4.92, 0.2	-0.87, 0.3

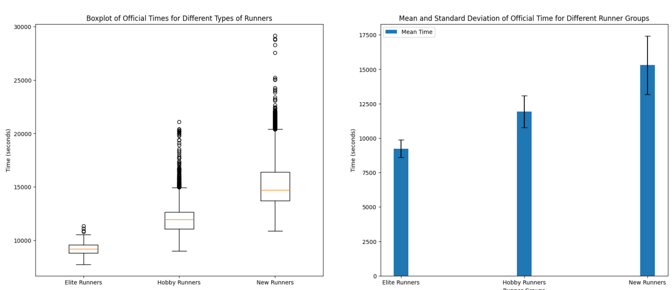

Fig. 1. Standard deviation of race times for elite, hobby, and new runners.

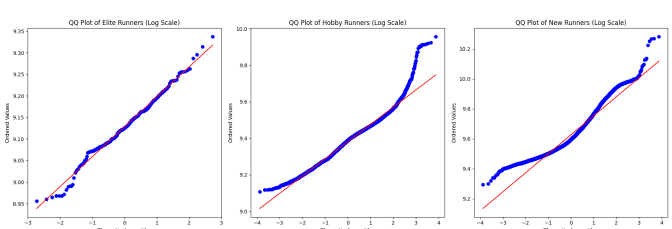

Fig. 2. Q-Q plot for new runners.

IV. CODING

A. Data

Temperature and precipitations are extracted from the years 1950 to 2019 from the meteorologic service of Barcelona. The wind speed and humidity are extracted from the same source but years 2007 to 2016. Race times are extracted from the marathon results of 2015. We calculated the mean and standard deviation of each group: elite, hobby, and new runners. The results are shown in table III.

TABLE II
GENDER PERCENTAGE IN RUNNER GROUPS

Group	**Male (%)**	**Female (%)**	**Total (%)**
Elite Runners	93.61	6.39	0.82
Hobby Runners	70.30	29.70	49.23
New Runners	38.74	61.26	49.38

TABLE III
MEAN AND STANDARD DEVIATION OF RACE TIMES FOR RUNNER GROUPS

Group	**Mean Time (min)**	**Standard Deviation (min)**
Elite Runners	153.92	10.66
Hobby Runners	198.63	19.24
New Runners	255.00	35.26

TABLE IV
MONTHLY TEMPERATURE AND WIND DATA

Month	**Max Temp (°C)**	**Min Temp (°C)**	**Wind (km/h)**
Jan	11.11	5.28	16.95
Feb	12.22	5.46	14.25
Mar	14.65	7.30	14.80
Apr	16.89	8.95	12.90
May	20.53	12.32	14.00
Jun	24.60	16.12	10.65
Jul	27.84	19.00	17.65
Aug	27.71	19.08	12.75
Sep	24.42	16.67	12.10
Oct	19.99	13.23	15.85
Nov	14.74	8.83	15.15
Dec	11.75	6.22	16.35
Avg	18.00	11.00	14.00

TABLE V
MONTHLY PRECIPITATION AND HUMIDITY DATA

Month	**Precipitation (mm)**	**Humidity (%)**
Jan	42.74	68
Feb	36.53	65
Mar	51.46	64
Apr	54.91	67
May	58.24	65
Jun	37.25	60
Jul	28.75	63
Aug	43.68	65
Sep	73.14	68
Oct	88.29	72
Nov	62.91	69
Dec	47.32	67
Avg	55.00	66

V. Definition of the experimental framework

VI. Model validation

VII. Results/Conclusions

References

[1] B. Knechtle, C. McGrath, O. Goncerz, E. Villiger, P. T. Nikolaidis, T. Marcin, and C. V. Sousa, “The Role of Environmental Conditions on Master Marathon Running Performance in 1,280,557 Finishers the ‘New York City Marathon’ From 1970 to 2019,” Frontiers in Physiology, vol. 12, 2021. [Online]. Available: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.665761. DOI: 10.3389/fphys.2021.665761. ISSN: 1664-042X.