26TH INTERNATIONAL CONFERENCE ON ENGINEERING AND PRODUCT DESIGN EDUCATION
5-6 SEPTEMBER 2024, ASTON UNIVERSITY, BIRMINGHAM, UNITED KINGDOM

TRAINING DESIGN FOR DISTANCE RUNNING; HOW TO APPLY STATISTICS, DESIGN OF EXPERIMENTS, AND DATA SCIENCE FOR ACTIVE LEARNING IN UNIVERSITY STUDENTS

Carlos Alberto GONZÁLEZ ALMAGUER, Verónica SAAVEDRA GASTÉLUM, Natalia FRIAS REID, Claudia RAMIREZ ZUBIETA and Arantza SANCHEZ GALLARDO
Tecnologico de Monterrey, Mexico

ABSTRACT
In recent years, long-distance races have boomed all over the world. Mexico is no exception, but just as the enthusiasm for races over 5k has increased, the rate of sudden heart attacks has also increased. This is due to not having personalised and adequate training. The following paper describes how, based on design and engineering techniques, young students at the Tecnologico de Monterrey have trained to increase their performance as athletes. Statistical methods allow the design of focused training to improve their time without affecting their physical and mental health. The introduction of data science, as well as the design of experiments, allows customising each training session according to information from each athlete, such as heart rate, VO_2 Max, weight, blood pressure, and the level of red and white blood cells, as well as running technique factors such as stride length, arm stroke technique, and stride. This allows students to actively learn about the application of statistical engineering, which will later allow them to transfer said knowledge to their professional field, making analogies between sport and a disciplinary competition. The results of more than five editions of the 5k and 10k race of the annual race of the Tecnologico de Monterrey Campus Querétaro are analysed, as well as of students who record daily data through recording devices which were processed in databases to calculate descriptive statistics and correlate efforts with the designed workouts.

Keywords: Distance training, data science, statistics, educational innovation, higher education

1 INTRODUCTION
"Design is everything; without design, science tends to be nothing."
Learning data science, as applied statistics is now known, is usually a bit traumatic for students. Antony & Antony [1] presents a UK study where students feel extremely comfortable solving book problems but cannot pose a real-life problem and solve it with statistical methods. In Latin America, the fear of learning statistical topics is constant [2]. In Mexico, this is a deeper problem that arises from the design of curricula of the Secretariat of Public Education (SEP), and which is evidenced in the evaluations of the Pisa test [3].
According to Gonzalez *et al.* [4], the design of a learning activity in the Tec21 educational model must be composed of three determining variables: student engagement, disciplinary learning, and competence development. Students who practice background races will likely have learn statistics concepts and apply them in data science to improving their time records on street racing circuits in Mexico or federated student competitions.
In recent years, there have been coaches who, although they have been good runners, lack the studies that support their ability to train other people who are initiating endurance training. This has led to an increase in injuries, physical problems attributed to poor diets, and an increase in deaths attributed to sudden death syndrome [5]. There have been studies that confirm this, an example is one presented by Bester *et al*. [6] where they show how an inadequate workout and training can lead to liver cancer due to training loads, volume, and intensity that are not specifically designed for each individual and their metabolism. Another example done by Nielsen *et al*. [7] in which they describe the negative effects of training without a scientific and technical foundation and the benefits of using statistics to improve health and competition times.

This research involves applying long-distance training to students from Tecnologico de Monterrey, using track tests of 5,000 metres and 10,000 metres and marathon routes of 5k, 10k, and 21k. It serves as a practical laboratory for learning data science and understanding inferential statistics through specially designed learning activities. The study has been extended to include alumni, who have developed a passion for long-distance races in Mexico and abroad. Moreover, the present research fills up a gap by utilising data science and statistics to achieves its objective of creating a comprehensive training program tailored for university students who want to participate in long-distance running doing it in safe and healthy way.

2 WHAT IS LONG-DISTANCE DESIGN TRAINING?

Long-distance training primarily focuses on the ability to run at the fastest pace with increased oxygen consumption, contrary to speed competitions that run without oxygen. Therefore, the variables studied in this research are VO_2 Max, which refers to the maximum amount of oxygen your body can absorb and use during exercise, in short words it measures your aerobic fitness levels, and heart rate per minute (bpm), which in turn affect three levels of activity that directly impact a runner's performance: maximum aerobic capacity, effort economy and anaerobic threshold.
Long-distance training is multifactorial. The following research is only focused on the three most crucial factors for reducing time versus distance travelled. In Figure 1, the variables under study, their relationships, and the constraints are described in a rich picture based on soft systems methodologies.

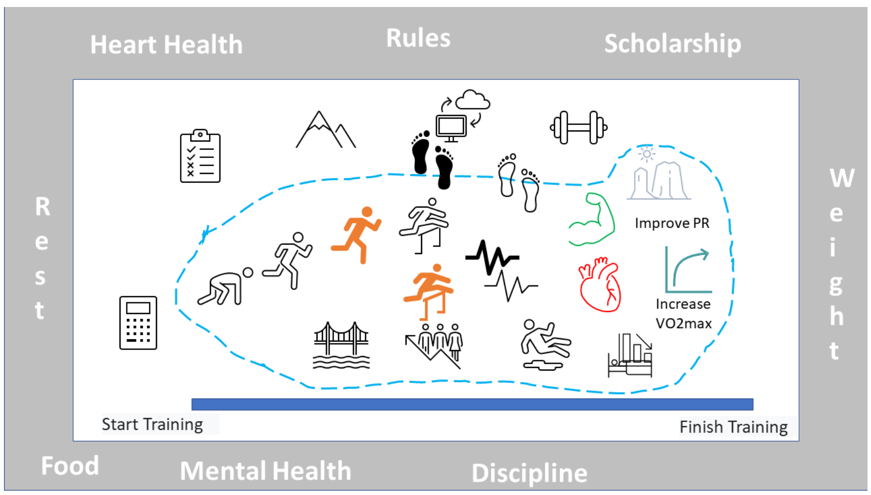

Figure 1. Soft System analysis with critical long-distance variables

3 TEACHING DATA SCIENCE, DESIGN OF EXPERIMENTS AND STATISTICS THROUGH RUNNING

All sports training starts with the diagnosis or analysis of the problem situation. Soft methodologies for problem-solving were used to design the analysis strategy. To do this, the Tec race database, which was held on August 25^{th}, 2019, was considered. In this race, fifty male and female students were invited to participate in the "Improve your times through data science and applied statistics" program to help them improve their competition times while learning statistics and data science. The other source of detailed information is from specific devices such as smart watches, phone apps, and cardiac monitoring bands. The first finding was that they conducted a situational analysis of how they were assessing their performance in the race. They were asked to analyse the race participants' results and perform data mining by discarding variables (columns) that did not add value. Then, by understanding the meaning of each column, they would transform it into highly significant variables. All this was done using Excel and Minitab software.
Through the construction of the database, they were able to apply descriptive statistics to all participants with the following variables:

- Age
- Starting time
- End time

- Partial time
- Average speed

Some invited students contributed by the phone app or with specific devices to monitor their physical performance, heart rate, altimetry, and average per kilometre. Before starting the program, a survey was carried out among them to know their level of learning statistics and data science through specific questions about their level of learning in certain concepts and a test with descriptive and inferential statistics exercises. For example, for the 10k Race, in the men's category, the media was 50 min and 32 seconds, with a standard deviation of 28 minutes, the winner's time was 37 min and 18 seconds, and the last runner to cross the finish line was of 1 hour, 36 minutes and 7 seconds. The most critical information to the training design is the heart rate. Figure 2 and Table 1 show the results of the Tec race of August 2019 with the statistic of 1352 participants. The students' interest in scientifically analysing the data was surprising since most did not belong to academic programs linked to engineering; from this moment, they achieved commitment and engagement in the activity. They had already realised that monitoring sensitive variables could enable them to improve their workouts.

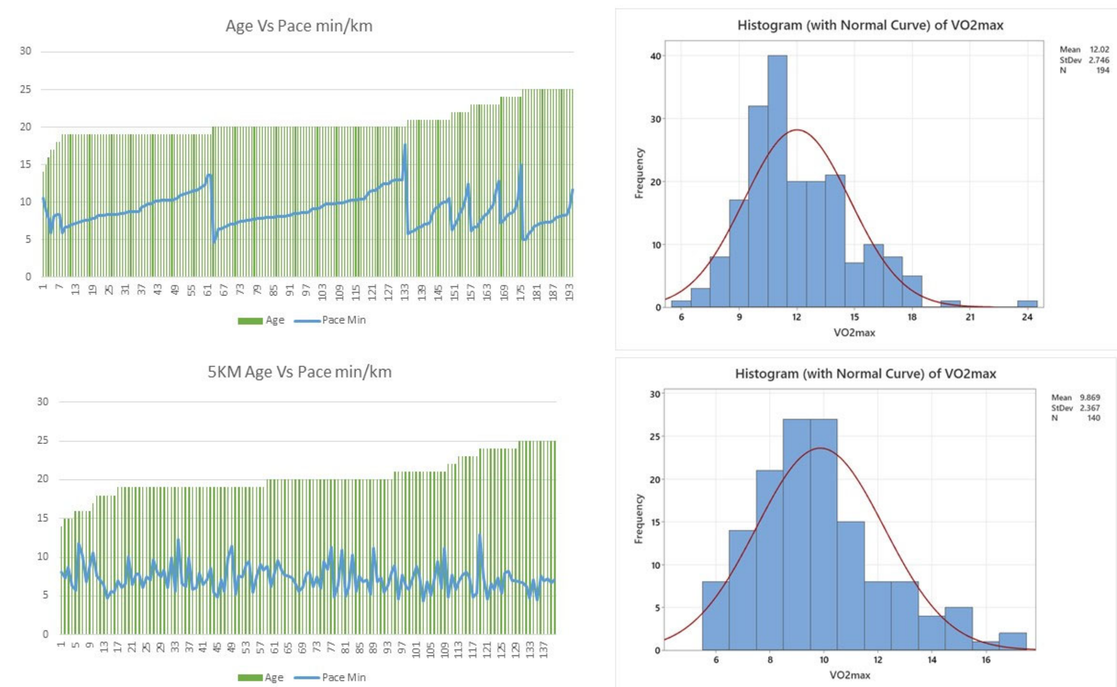

Figure 2. Comparison between Age and Pace (min/km) segmented by Woman (up) and Men (below)

Table 1. Statistics of 1,352 long distance runners to analyse VO_2 Max and Pace per minute; StDev: Standard deviation; Q1: first quartile; Q3: third quartile; FT Minute: finish time

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Mode	
Age	35.953	14.06	14	23	34	47	103	20	121
FT Minute	48.102	14.34	18.73	36.41	46.98	58.68	102.43	33.1	4
Pace Min	7.5067	2.15	3.73	6.02	7.03	8.47	18.35	7.51	13
VO_2 Max	10.134	2.90	5.04	8.12	9.50	11.43	24.77	10.14	13

4 DESIGN A TRAINING PLAN BASED ON SCIENCE

One mistake in long-distance training is following other athletes' workouts when the workout should be fully personally customised. Therefore, the goal was to design proper training for each student based on

personal diagnosis. The data was collected with manual chronometer and personal devices (smart watches).
Tec conducted a Cooper test on the track to make a better diagnosis regardless of the race results. This test consisted of running 2400 metres (six laps of 400 m) at the maximum possible speed, a flexibility test, and strength test. The test was done through three mixed groups integrated by the best times, either final or partial, in the 5k race.
Pat Bradley [9], former United States Olympic Committee physiologist, argues that most people can keep their speed at VO_2 Max level for eight to ten minutes, so a runner can do it for ten to twelve minutes. The test must do at two miles is 2200 m, equivalent to five and a half laps to the official track of four hundred metres, having a group not so homogeneous in athletic performance was considered better to follow the test Cooper.
Through these results, the VO_2 Max was calculated, which is the ability to use oxygen at maximum effort. In an analogy, the VO_2 Max is the engine, and the heart and lungs are those that pump fuel through the body, so having a high VO_2 Max equals having a powerful engine. It proceeded to calculate 65% of VO_2 Max, equation 1, for each student taking the average per kilometre in the Cooper test with the following formula where BT is the best time per kilometre in seconds, and TTP is the target training time in percent per kilometre.

$$BT*Km+(1-TTP*Km*BT*Km)) \qquad (1)$$

For example, if in the Cooper test a student had four min and 20 seconds on average per kilometre, this measurement is equivalent to 260 seconds, so to have a training at 65% of your VO_2 Max capacity, you must run every kilometre at 351 seconds, which is equivalent to five min with 51 seconds.
The intensity of training is correlated with possible cardiac affections, which can be related to comorbidity diseases such as overweight, diabetes, high blood pressure and which can cause sudden death if not well designed. In figure 3, the VO_2 Max calculation is shown with the part time of one kilometre during the Tec race for both categories, 5 and 10k. An acceptable VO_2 Max would be below six min/km, it can also be observed that most runners are well above with an average of 10 min, which statistically infers that they are not prepared for a competition for this distance and that in conditions of greater stress such as food, heat, humidity, sleep and consumption of alcohol or some drug this effort can be dangerous to health at the time of competition. Overall, VO_2 Max is a good measure of the ability of the heart to make a physical effort, in the data observed to the 1352 participants we can observe only a smaller percentage is suitable for this type of competition.

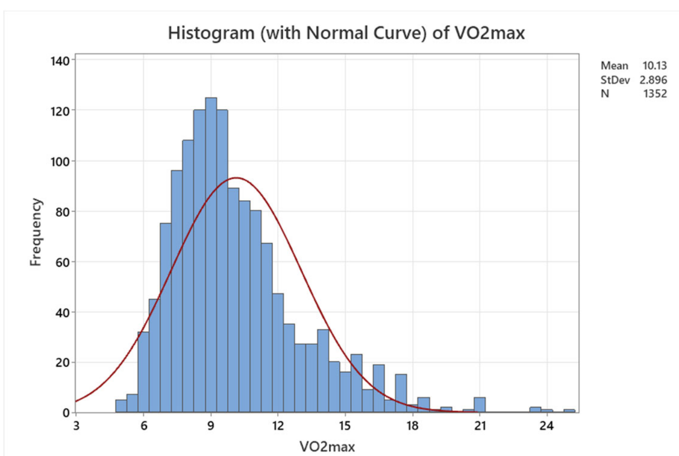

Figure 3. Calculation of VO_2 Max for all participants in the Tec race

It is also important to determine the training threshold by age, this is calculated in a quite simple way, and is to take as constant 220 beats per minute and subtract age. For example, a 19-year-old student's maximum ability to strain his heart is 201 beats. Furthermore, training must be individually designed to increase your VO_2 Max without reaching maximum beats per minute. This is achieved through training stages over a period of time. According to Tadeuz Kempka [8], who was one of the most brilliant long-distance coaches in Mexico, there are four stages to reach the maximum performance of a competition, this is shown in Table 2.

Table 2. The four training stages to reach the maximum performance

Stage 1	Stage 2	Stage 3	Stage 4
Development of aerobic resistance 1 to 6 Week	Anaerobic training 7 to 12 Week	Increased tolerance of lactic acid 13 to 18 Week	Advanced anaerobic training 19 to 24 Week

Through experiment design, the intensity parameters can be modified according to the reports of heart rate and average speed to improve the VO_2 Max, the variables that are under observation are distance, speed, and training ground time. Using Saavedra's [10] methodology of data science and applying concatenated experiment designs significant improvement in sports training can achieved.

5 DESIGN A TRAINING PLAN BASED ON SCIENCE

One common mistake made by long-distance runners is to copy the workouts of other athletes when they ought to be completely tailored to them. Designing appropriate instruction for each student based on their unique diagnoses was the aim. Thus, wrist band sensors were used to gather the data, which included heart rate, run pace, and height.

From October 2019 to February 2020, training and learning in statistics and data science could be carried out without any problem. However, COVID-19 confinement put a stop on the experiment, which later was resumed in October 2021 by performing a physical conditioning (weight, flexibility, and strength) stage to recover the lost form during confinement. In August 2022, the results from thirty-two students who resumed and followed the training program were received. Figure 4 shows a graph of the results obtained in a 10 km race in the sea level vs the Tec Race at 1885 metre over sea level, as average the rate per kilometre decrease from 5 minutes 54 seconds per km to 4 min 37 seconds per km, and the heart rate decrease from 182 ppm to 154 ppm, this is an excellent performance.

Figure 4. Comparison between race in altitude vs sea level analysing heart rate; using height, run pace, and heart rate variables

6 CONCLUSIONS

The design of long-distance racing training based on data science and statistics achieves substantial improvement on the athlete's performance. But in academic terms, it generates and increases learning by providing real-life scenarios and improves students' engagement. The student's performance in long-distance street tests was highly significant, but it was more remarkable in those who entered the athletics team and improved their times. In terms of academics, the learning of statistics was reflected in its training units for students of the Tec21 educational model and in the previous programs, and it offered students a hands-on approach to learning complex concepts, making the experience more relevant and engaging. In terms of educational design, such research can enhance engagement, provide practical learning experiences, improve outcomes, and lead to innovative curriculum development. Students say that this learning could be transferred to the professional sphere.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico, in this research study of applying and integrating statistical methods to student's real life.

REFERENCES

[1] Antony J. and Antony F. J. (2001) Teaching the Taguchi method to industrial engineers. MCB University Press. Vol. 50 No. 4, pp. 141-149.

[2] Ramos-Galarza C. et al. (2021). The Fear of Learning Statistics in Latin America: Computational Solutions. In: Russo, D., Ahram, T., Karwowski, W., Di Bucchianico, G., Taiar, R. (eds) *Intelligent Human Systems Integration* 2021. IHSI 2021. Advances in Intelligent Systems and Computing, vol 1322. Springer, Cham. https://doi.org/10.1007/978-3-030-68017-6_30

[3] https://www.oecd.org/pisa/publications/PISA2018_CN_MEX_Spanish.pdf

[4] González-Almaguer C., Saavedra V., Caballero E., Acuña A., Zubieta C., Barbosa E. and Lule M. (2021, September). Design Thinking and Design Experiments: The Fusion of the School of Design and Industrial Engineering to Create Learning Experiences in The Tec21 Educational Model. In *International Conference on Engineering and Product Design Education*, EPDE 2021, Herning.

[5] Valdés L. (2013) Tasa de recuperación de la frecuencia cardiaca al minuto de post-esfuerzo máximo en un equipo de segunda división profesional de fútbol asociación, como indicador de mortalidad. Universidad Autónoma del Estado de México.

[6] Bester R., Stander Z., Mason S., Keane K. M., Howatson G., Clifford T., Stevenson E. and Loots D. T. (2021, September). Characterising Marathon-Induced Metabolic Changes Using 1H-NMR Metabolomics, Published online 2021 Sep 27. doi: 10.3390/metabo11100656.

[7] Nielsen R. O., Buist I., Sørensen H., Lind M. and Rasmussen S. Training errors and running related injuries: a systematic review. *Int J Sports Phys Ther*. 2012 Feb;7(1):58-75. PMID: 22389869; PMCID: PMC3290924.

[8] Nielson D. E., George J. D., Vehrs P. R., Hager R. and Webb C. V. (2010) Predicting VO2max in College-Aged Participants Using Cycle Ergometry and Perceived Functional Ability.

[9] Castillo I. Discontinuous methods and their effect on indicators physiological and anthropometric in university athletes of 800 meters.

[10] Saavedra Gastelum V., Gonzalez Almaguer C. A., Gonzalez de Cosio Barron A., Zubieta Ramirez, C. et al. Use of design of experiments, design thinking and data science in precision agriculture. *DS 117: Proceedings of the 24th International Conference on Engineering and Product Design Education* (E&PDE 2022), London South Bank University in London, UK. 8th - 9th September 2022.