09-031

DESIGN OF EXPERIMENTS, DESIGN THINKING AND DATA SCIENCE FOR LEARNING PRECISION AGRICULTURE

Saavedra Gastélum, Verónica (1); González Almaguer, Carlos Alberto (1); Gonzalez de Cosio Barrón, Arturo (1); Zubieta Ramírez, Claudia (1); Frías Reid, Natalia (1)

(1) Tecnológico de Monterrey

The Experimental Agricultural Field of Tec de Monterrey (CAETEC) implements precision agriculture to obtain better cultivation techniques, to increase productivity, and to have higher quality crops optimizing resources. The information obtained is used by students for their learning in statistics and by farmers in the Bajío region to improve their crops. Currently satellites, sensors and drones are used to obtain sensitive information, this generates a lot of information. It is necessary to build databases and then, through the design of experiments and design thinking, to elaborate statistical experiments to obtain the appropriate parameters to maximize the production of agricultural products with the highest nutritional value and quality. This research essay presents the results in three learning blocks during the August to December 2021 semester at Tecnologico de Monterrey, namely: "IN2004 Generation of Value with Data Analytics," "IN2005 Evaluation of Organisational Competitiveness," and "IN2006 Analysis of the Viability of Projects with a Systemic Perspective." These blocks focused on the operation of the robotic barn of the university's CAETEC, where design thinking, regression analysis, design of experiments, and the concepts of precision agriculture were introduced to solve the challenges posed in each project.

Keywords: Design_of_experiments; design_thinking; Regression_Analysis; Precision_agriculture; Higher_Education

DISEÑO DE EXPERIMENTOS, DESIGN THINKING Y CIENCIA DE DATOS PARA EL APRENDIZAJE DE LA AGRICULTURA DE PRECISIÓN

El Campo Agropecuario Experimental del Tec de Monterrey (CAETEC) implementa la agricultura de precisión con el fin de obtener mejores técnicas de cultivo, incrementar la productividad y tener mayor calidad en las cosechas optimizando los recursos. La información obtenida es utilizada por los estudiantes para su aprendizaje en estadística y por los agricultores de la región del bajío para la mejora de sus cosechas. Actualmente se utilizan satélites, sensores y drones para obtener información sensible, esto genera una gran cantidad de información. Es necesario construir bases de datos y posteriormente, a través del diseño de experimentos y design thinking, elaborar los experimentos estadísticos para la obtención de los parámetros adecuados para maximizar la producción de productos agrícolas con el mayor valor nutricional y calidad. En el presente trabajo se presentan los resultados obtenidos en el semestre Agosto diciembre 2021 en los Bloques de IN2004 Generación de valor con analítica de datos, IN2005 Evaluación de la competitividad Organizacional e IN2006 Análisis de la viabilidad de proyectos con perspectiva sistémica, aplicados a los datos del CAETEC en donde se utilizaron Design Thinking, Análisis de Regresión, Diseño de Experimentos y agricultura de precisión para resolver los retos planteados en cada proyecto.

Palabras clave: Diseño_de_Experimentos; Design_thinking; Análisis_de_Regresón; Agricultura_de_Presición; Educación_Superior.

Correspondencia: Verónica Saavedra Gastélum veronica.saavedra@tec.mx; Carlos Alberto González Almaguer cgonzalz@tec.mx

Agradecimientos: The authors acknowledge the technical and financial support of the Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico, in the production of this work.

1. Introduction

In the Bajío region of Mexico, work has been done to promote the importance of protecting the environment and making the most of natural resources. The awareness has led farmers in the region to seek for new technologies that help determine the treatments that maximise nutrients in cows' diets, where it has been observed that the primary food is corn. Therefore, this project arises as a need of CAETEC (Spanish acronym for Experimental Agricultural Field of Tecnologico de Monterrey) to maximise the production of litres of milk from cows based on their diet.

In precision agriculture, new monitoring tools include satellites, sensors, and drones. These allow us to create databases with all the parameters that are important to maximise productivity, quality, nutritional value, and to reduce waste by-products; hence, Mexico must invest in these technologies and promote projects in this sector.

Mexico's agricultural industry is one of its main economic sectors, hence scientific research, and innovation to improve technological capacity and product quality are vastly important. Thus, production processes must be continuously improved, reducing resources consumed and waste. Precision agriculture involves implementing data technologies, processing techniques, and statistical analysis to improve cultivation techniques and increase productivity and product quality.

This research essay presents the experiences in three learning blocks during the August to December 2021 semester at Tecnologico de Monterrey, namely: "IN2004 Generation of Value with Data Analytics," "IN2005 Evaluation of Organisational Competitiveness," and "IN2006 Analysis of the Viability of Projects with a Systemic Perspective." These blocks are focused on the operation of the robotic barn of the university's CAETEC, where design thinking, regression analysis, design of experiments, and the concepts of precision agriculture were introduced to solve the challenges presented for each project. The formal educational model of the university, TEC21, (IFE; 2021) is a challenge-based learning model where students, professors and training partners work together to solve a specific real-world problem of the latter's organisation. In each course, the challenge is presented for resolution.

This learning sequence is for students in the 5th semester of Industrial Engineering. It is defined as a requirement that the project must be within the same organisation; this, in the academic language of the Tecnologico de Monterrey, is called *Training Partner* in order to have a learning sequence.

During the training units (UF, Spanish acronym of Unidades de Formación) IN2005, students analyse the problem to be solved in the organisation and make a diagnosis of it applying soft systems methodologies.

The next phase consists of data science procedures during IN2004, where they apply quality control and statistical techniques and tools; students are able to define the problem, and through the design of experiments, they can find the correct parameters to solve the problem.

To obtain these parameters, it is necessary to propose a project, which can range from a quality task force to the purchase of machinery and equipment, where they close with the last

IN2006 project feasibility block, where they ultimately design the comprehensive solution proposal.

In other words, this learning block translates the needs and findings in project proposals, using the methodology that is presented in this congress.

2. Methodology

This research's work methodology is divided into two parts: the statistical treatment of the data and the statistical inference for the design of projects that add value to the cows' milking process throughout the value chain, from acquiring the seed to milking. CAETEC provided data collection.

2.1. Statistical analysis.

A regression analysis was performed with the data obtained from the CAETEC automated barn to know the factors affecting the production of milk litres. The factors analysed were the amount of rolled corn, soybean paste, cottonseed, molasses, dairy mixture, Distiller's dried grains (DDG), corn silo, triticale silage, water, total whole diet, and days. Table 1 shows the result of the ANOVA Table for the multiple linear regression (Anderson, et. al., 2017).

Table 1. Analysis of Variance

Source of variation	DF	ADj SS	Adj MS	F-value	P-value
Regression	10	1209855	120986	2.27	0.031
Rolled corn	1	82653	82653	1.55	0.220
Soybean paste	1	10316	10316	0.19	0.662
Cottonseed	1	1829	1829	0.03	0.854
Molasses	1	43467	43467	0.81	0.372
Dairy mixture	1	62583	62583	1.17	0.285
Distillers dried grains (DDG)	1	83094	83094	1.56	60.219
Corn silo	1	71	71	0	0.971
Triticale silage	1	18017	18017	0.34	0.564
Water	1	74440	74440	1.40	0.244
Days	1	89289	89289	1.67	0.203
Error	3	2293702	53342		

Total	53	3503557

As can be seen, although the regression analysis tells us that at least one of the factors serves to forecast the litres of milk, when analysing each of the factors separately, we observe that none of them serves to forecast. This behaviour is observed when the principle of independence of the explanatory variables is broken; that is, there is multicollinearity. To verify the presence of multicollinearity, we calculated Pearson correlation coefficients between the explanatory variables, as shown in Table 2 and 3.

Table 2. Pearson correlation coefficients

	Rolled corn	Soybean paste	Cotton seed	Molasses	Dairy mixture	Distillers dried grains (DDG)
Soybean paste	0.292					
Cottonseed	0.893	0.633				
Molasses	0.923	0.629	0.979			
Dairy mixture	0.905	0.653	0.971	0.987		
Distillers dried grains (DDG)	0.833	-0.276	0.540	0.568	0.526	
Corn silo	0.944	0.575	0.980	0.992	0.990	0.627
Triticale silage	-0.502	0.663	-0.120	-0.143	-0.099	-0.886
Water	0.868	-0.210	0.597	0.619	0.588	0.993
Total whole diet	0.996	0.290	0.902	0.921	0.905	0.835
Days	-0.414	-0.021	-0.482	-0.381	-0.339	-0.462

Table 2. Pearson correlation coefficients cont.

	Corn silo	Triticale silage	Water	Total whole diet
Soybean paste				
Cottonseed				
Molasses				
Dairy mixture				
Distillers dried grains (DDG)				
Corn silo				
Triticale silage	-0.207			
Water	0.673	-0.589		
Total whole diet	0.948	-0.505	0.873	
Days	-0.419	0.307	-0.454	-0.466

The Pearson correlation coefficients analysis shows mostly linear correlation between the variables because the values are higher than 0.7 or less than -0.7. Examples include cottonseed and rolled corn with a Pearson correlation coefficient of 0.893 or molasses with rolled corn having a value of 0.923.

The students who conducted this work do not have the necessary preparation to analyse factors or main components. So we worked with them using different analysis alternatives according to their academic level.

2.2 Variables analysis

As a first step, we calculated the correlations of all the explanatory variables related to the litres of milk to find those variables with a high level of linear relationship. Pearson's correlation

coefficients are shown in Table 4. All the correlations are very close to zero; there is no linear relationship between the variables, or the linear relationship is minimal.

Table 4. Pearson correlation coefficients vs. milk

	Litres of milk
Rolled corn	0.089
Soybean paste	0.055
Cottonseed	0.133
Molasses	0.104
Dairy mixture	0.044
Distillers dried grains (DDG)	0.067
Corn silo	0.067
Triticale silage	-0.069
Water	0.071
Total of integral diet	0.079
Days	-0.123

Due to the small values of the correlations between the variables with the litres per milk, we decided to analyse scatter plots to find the reason. Cottonseed, Milk Mixture, Molasses, Rolled Corn, Soybean Paste, DDG, and Triticale Silage have a too discretized behaviour; this

generates problems when running a regression analysis due to the assumption of normality. Figures 1 and 2 show an example of this behaviour.

Figure 1: Scatterplot of Lt vs cottonseed

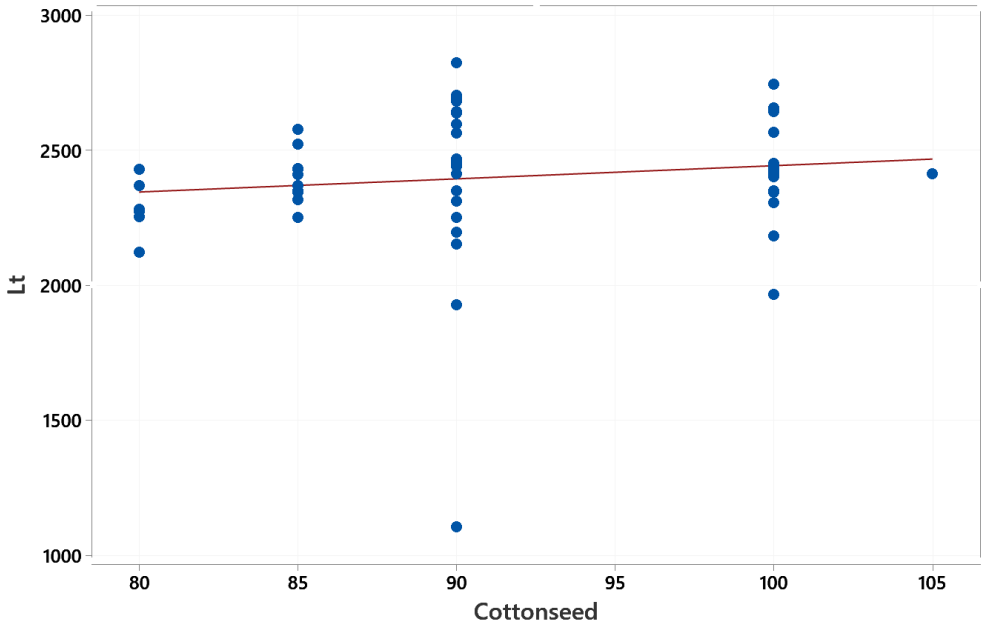
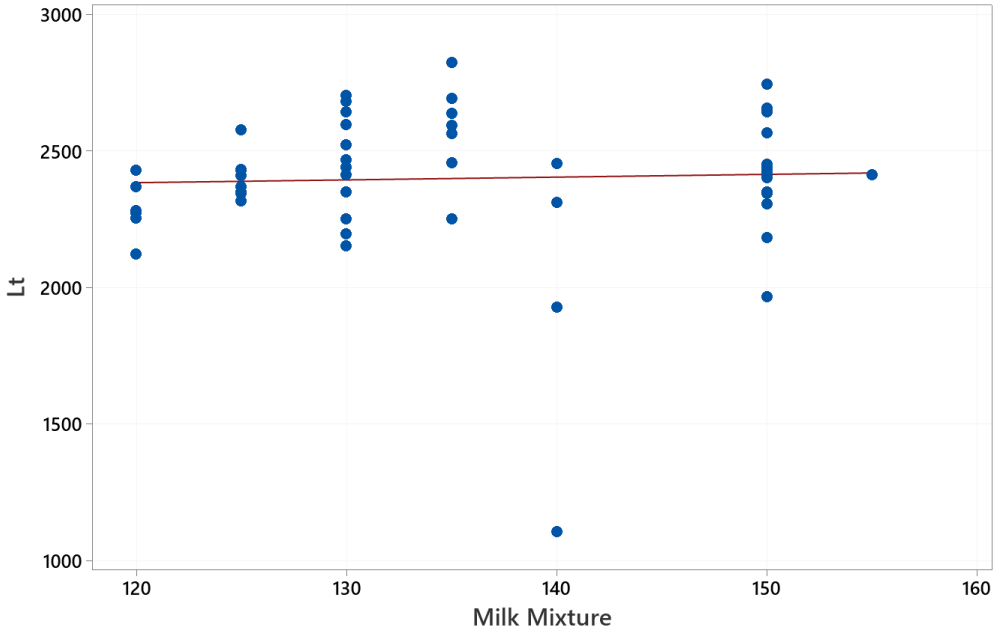
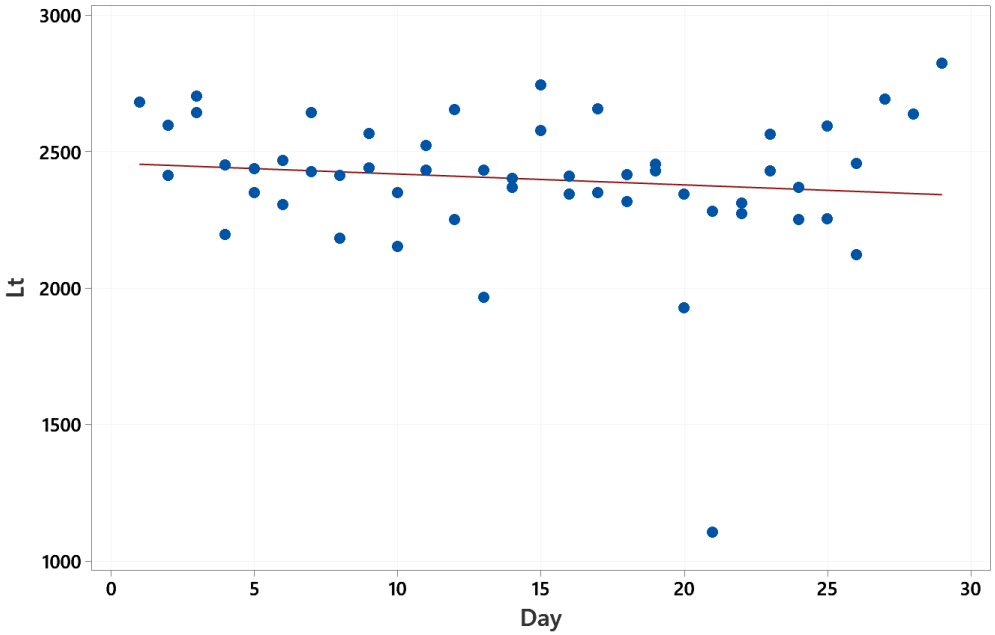


Figure 2: Scatterplot of Lt vs Milk Mixture



On the other hand, Corn Silo, Water, Total of Integral Diet, and Day may be considered as continuous variables. As can be observed an example in Figure 3.

Figure 3: Scatterplot of Lt vs Day



Hence the students understood the need to use a general factorial design with several factors and different levels (Montgomery, 2013).

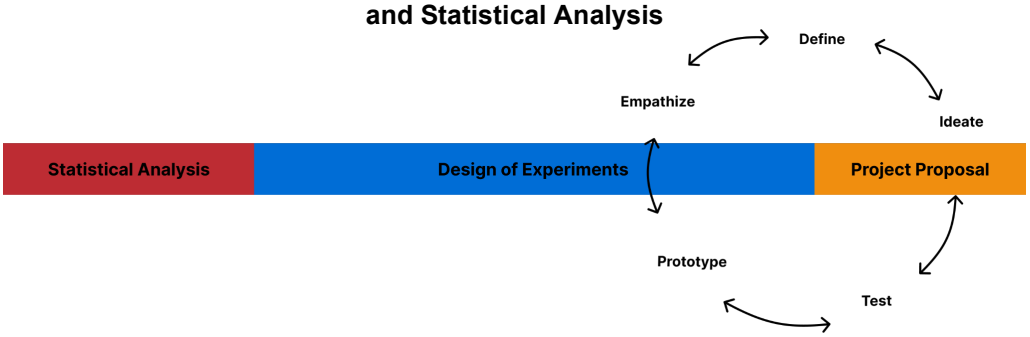
2.2 Design of solution proposals.

The design of experiments is essential for research, development, innovation and problem-solving because it allows using statistical modelling to propose projects that help the organisation improve its processes.

By combining design thinking with the design of experiments (Hernández-Gil and Nuñez-López, 2020) (González-Almaguer, et. al., 2021), one can focus the project through "thought engineering" to select the three most critical factors to modify parameters. This ideation and creativity optimise the process.

Figure 4 shows the transition from the statistical analysis stage through the design of experiments with design thinking to the project proposal stage.

Figure 4. Project Proposal derived from Design of Experiments, Design Thinking and Statistical Analysis



During the IN2005 block, the students conducted a statistical treatment of the data. In the IN2006 block, they make value propositions for projects that optimise the identified critical factors.

Table 5 shows the projects proposed by the students.

Table 5. Project Proposals

Project Name	Category
Resource management software	Software and data analytics
Design and implementation of an ERP Enterprise	ERP Software
Weighing and data collection system	System and data analytics
Food Mixing Quality Laboratory	Laboratory and data analytics
Alerts communication system	System and data analytics
System for saving and maintaining water	System and data analytics
Data analysis and forecast generation system	System and data analytics

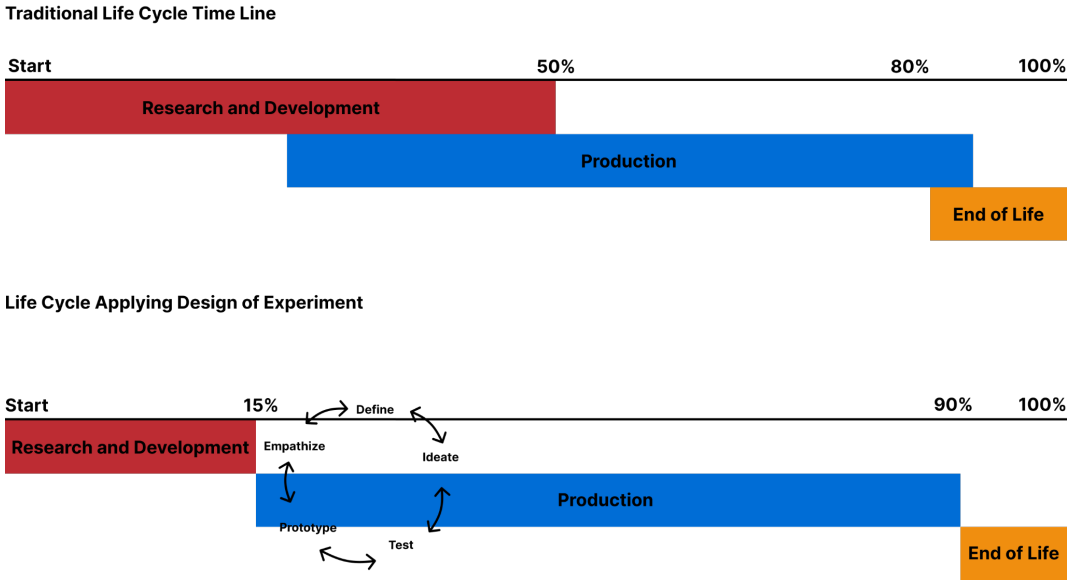
3. Results.

The design of experiments and analyses allowed the students to propose different solutions projects throughout the value chain of the milk-producing process in the automated barn, from strategies to optimise the processes of planting and irrigation through laboratories to measure the quality of the feed to generate the database management systems.

Figure 5 shows how a product's life span increases when enriched by creativity and the design of experiments. For this reason, it was proposed to let the students apply creativity and the

design of experiments to improve the cow feeding process and increase milk production in the framework of sustainable development and precision agriculture.

Figure 5: Design of experiments and Creativity applied to processes



Together with the CAETEC specialists, we saw the need to propose a concatenated design of experiments, in which the factors do not remain fixed throughout the process but seek to maximise at each stage; the different factors should vary to achieve precision agriculture that generates more milk based on all the nutrients that cows need to consume inside the stables. Thus, future research should involve the concatenated design of experiments applied to CAETEC, where students actively collaborate to select the factors and corresponding designs.

4. Conclusions

In their analyses, the students used design thinking to find solutions to the problems indicated in the data provided by CAETEC. Its discretized data meant that classical regression models would not provide the correct analysis of the response variable. It was necessary to make students aware of the problems in professional practice when the necessary assumptions to perform the regression analysis are not necessarily met.

Also, the measuring instruments used tend to discretize data that, by nature, are continuous. This creates a problem in the assumption of normality of the response variable given the explanatory variable. Again, it creates the need to search for new statistical techniques to solve the problem. It contributes to the awareness that CAETEC must search for new measurement systems that allow more accurate values for the variables analysed. The students learned about multicollinearity in the explanatory variables by finding this behaviour in the analysis of the variables.

The primary objective of the blocks is to enrich students' learning through real-world cases and problems for the students to solve using design thinking and statistical methods.

4. References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2017). Multiple Regression. Statistics for Business & Economics, (pp. 683 - 727). USA: Cengage Learning.

González-Almaguer, C., Saavedra, V., Caballero, E., Acuña, A., Zubieta, C., Barbosa, E. & 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.

Hernández-Gil, C. & Núñez-López, J.A. (2020). Design thinking aplicado al mejoramiento de las competencias ciudadanas en universitarios: voto popular. Revista de Investigación Desarrollo e Innovación, 11(1), 85-98. doi: https://doi.org/10.19053/20278306.v11.n1.2020.11685

Montgomery, D.C. (2013). Introducción a los diseños factoriales. (8a ed). Diseño y análisis de experimentos, (pp. 183 - 224). México: Limusa Wiley.

Observatory IFE. (2020, Febrero 7). World Bank shows worldwide Tecnológico de Monterrey's Educational Model. Institute for the Future Education. Accessed from https://observatory.tec.mx/edu-news/world-bank-tec21-tec-de-monterrey-educational-model

United Nations. (n.d). Objetivo 9: Construir infraestructuras resilientes, promover la industrialización sostenible y fomentar la innovación. Accessed April 25, 2022, from https://www.un.org/sustainabledevelopment/es/infrastructure/

United Nations. (n.d.). Objetivo 12: Garantizar modalidades de consumo y producción sostenibles. Accessed April 25, 2022, from https://www.un.org/sustainabledevelopment/es/sustainable-consumption-production/

Communication aligned with the Sustainable Development Goals