Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
MAT2021-2732

Towards Generating Real-World Datasets for Teaching Statistics in Industrial Engineering

Farida Saïd
Associate Professor
University of South Brittany
France

Iehann Eveno
R&D Engineer
University of South Brittany
France

Jeanne Villaneau
Emeritus Associate Professor
University of South Brittany
France

ATINER CONFERENCE PAPER SERIES No: MAT2021-2732

An Introduction to ATINER's Conference Paper Series

Conference papers are research/policy papers written and presented by academics at one of ATINER's academic events. ATINER's association started to publish this conference paper series in 2012. All published conference papers go through an initial peer review aiming at disseminating and improving the ideas expressed in each work. Authors welcome comments.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Saïd, F., Eveno, I. and Villaneau, J. (2021). **"Towards Generating Real-World Datasets for Teaching Statistics in Industrial Engineering"**. Athens: ATINER's Conference Paper Series, No: **MAT2021-2732**.

Athens Institute for Education and Research
9 Chalkokondili Street, 10677 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm
ISSN: 2241-2891
10/11/2021

Towards Generating Real-World Datasets for Teaching Statistics in Industrial Engineering

Farida Saïd
Iehann Eveno
Jeanne Villaneau

Abstract

Driven by digital transformation, companies produce nowadays vast amounts of data. In the manufacturing industry, for example, the use of modern smart technology contributes to this data profusion. However, very few enterprise datasets are made freely available which results in a serious lack of open real data for research and education. In this paper, we present a discrete-events simulation tool that was developed to support undergraduate students in their Statistics and Data analysis course. It simulates scheduling scenarios in a manufacturing environment and the generated data may be used to put into practice Statistics concepts and methods to design cost-effective strategies for optimizing key performance indicators, such as reducing production time, improving quality, eliminating wastes, maximizing profit…

Keywords: industrial datasets, teaching statistics, discrete events simulation

Introduction

Industrial engineering (IE) is the branch of engineering that deals with improving processes, systems, or organizations and designing goods or services in the most efficient way possible, saving money, time, raw resources, labor, and energy while complying with safety standards and regulations. Industrial engineers use scientific and technical knowledge and skills to integrate and operate complex systems, and as such, their training programs have a significant scientific component. IE educators appear to agree on IE knowledge and curriculum structure while continually seeking innovations in content and instruction (Lang at al., 1999; Davies, 2001; Carrera, 2006; Eskandari et al., 2007; Lima et al., 2012; Sackey and Bester, 2016). In comparative studies (Kuo, 2001; Fraser and Teran, 2006; Fraser, 2015; Nguyen and Nguyen, 2018), there is consensus that statistics and data analysis are part of the core courses in IE programs, and contextualizing learnings by working with real-world industry data is highly recommended to help students better understand their future profession. For that, we could take advantage of the vast amounts of data produced by companies driven by digital transformation and the increasing use of connected devices and interconnected machines. However, because these data are central to the manufacturing systems, they are rarely shared or made freely available which results in a serious lack of open real data for research and education. To overcome this difficulty, various simulation tools have been developed, some of which are free[1].

To support IE undergraduate students at University of South Brittany (France) in their one-semester course on Statistics & Data analysis, we developed a simulation tool in the agri-food domain. It simulates the operation of a pastry factory based on discrete-events simulations (DES) (Elizandro and Taha, 2007). Simulation models are usually built to understand how systems behave over time and to compare their performance under different conditions. DES models are widely used for design and implementation tasks, operational analysis, advanced planning, resource allocation, and logistics management. They are also commonly used for scheduling and automation, which are at the heart of Industry 4.0 (Ram et al., 2018).

The main objective of our simulation tool was to create realistic industrial experiments and data to practice data analysis methods (sampling, confidence intervals, hypothesis testing, regression models...). The input parameters of the simulator are typical of ERP (Enterprise resource planning) data and the outputs are typical of SCADA (Supervisory control and data acquisition) feedback. The simulated data may be used, among others, to (1) identify the most significant key performance indicators (overall equipment effectiveness, capacity...) through the analysis of the production behavior, (2) to determine the critical phases of the production process and to understand the involvement of the physical environment in the quality of the production.

[1]https://en.wikipedia.org/wiki/List_of_discrete_event_simulation_software.

The Manufacturing System

The manufacturing system we are concerned with includes the 4 main activities depicted in Figure 1. Various raw ingredients (eggs, butter, sugar, flour) are mixed and kneaded to form a dough that is shaped into pastries that are baked by lots in an oven. Once cooked, the pastries are bagged and palletized. Quality control occurs after the bagging phase; it consists of testing a random sample of pastries from a lot and deciding whether to accept or reject the entire lot based on the quality of the random sample.

Figure 1. *Main Activities of the Manufacturing Process*

Many factors are involved in the completion of each activity and any disruption in any one of them affects the rest of the process. For example, a stoppage during baking results in undercooking or overcooking of pastries and their subsequent disposal. This in turn has an impact on the number of products to be bagged, palletized, and sold.

According to Schruben and Schruben (2001), the rules or factors that govern the interaction of entities in a system that can be controlled are called parameters, while those that cannot be controlled are called laws. Figure 2 shows some parameters and laws involved in the manufacturing system; one can refer to Table 1 for a list of the main factors.

Figure 2. *Manufacturing System*

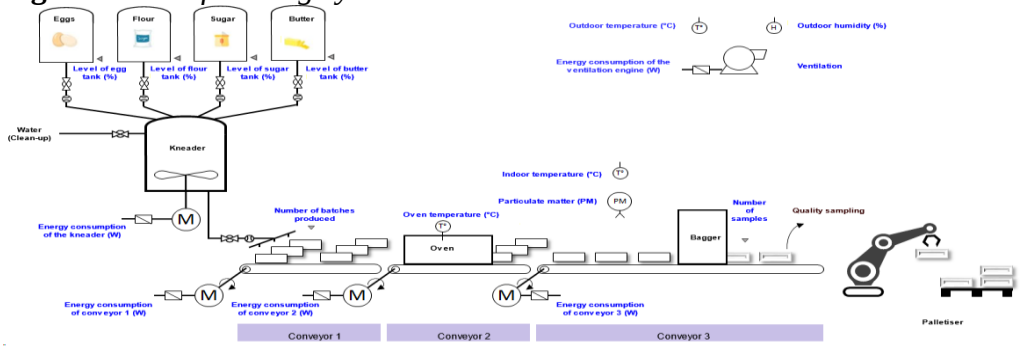


Table 1. *Main Manufacturing System Factors*

Equipment shutdown by equipment (0, 1)
Equipment failure by equipment (0, 1)
Oven temperature (°C)
Amount of each ingredient needed to make a lot of pastries (% of the tank size)
Number of rejects during a run

Level of each ingredient tank during production (%)
Maximal level of each ingredient tank
Maximal refill time for each ingredient tank (seconds)
Minimal refill time for each ingredient tank (seconds)
Refill time for each ingredient tank during production (seconds)
Maximal number of staffers
Minimal number of staffers
The actual number of staffers
Time spent in the oven (min)
Time spent on the conveyor 3 (min)
Quality of the sampled items
Disposal thresholds
Random disposal thresholds

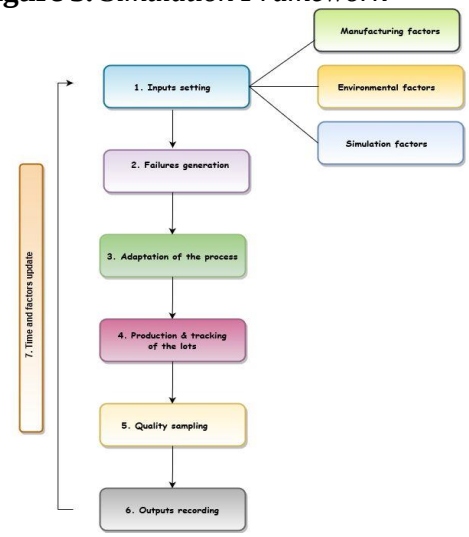
The Simulation Tool

A manufacturing system is a combination of resources (machines, people, raw materials...), planning, organizational structures, information flows, and IT-systems that aim to achieve the manufacture of an economic product cost-effectively. To understand how systems behave over time and to compare their performance under different conditions, two types of simulation models can be built: (1) discrete-event dynamic system models where the operation of the system is represented as a chronological sequence of events, and (2) continuous-event dynamic systems that track systems responses over time according to a set of equations involving usually differential equations.

We adopted the discrete-event simulation (DES) approach which is usually used to model workflow as a network of queues and activities where state changes occur at discrete and irregular time stamps. DES models are generally stochastic, and randomness is generated using statistical distributions.

The general framework of the simulation is depicted in Figure 3. It proceeds in the following steps for a single run: (1) the user selects a combination of input factors related to the manufacturing system, the environment, and the simulation. (2) Depending on the stated inputs, a variety of failures can be generated. These can be random or functional (they follow statistical distributions). In case of failure, an intervention request is launched. It feeds a knowledge base of failures and repair response times according to the type of failure, the number of available staffers, and the failure time occurrence (hour, day, month). (3) The laws of the industrial process are adapted to the inputs and, (4) the production process is launched. (5) An assurance sampling is carried out; it consists of selecting some items in a lot and deciding whether to accept or reject the entire lot based on the inspection of the sample. (6) The output data are stored in a spreadsheet (CSV file) for subsequent statistical analysis: one line by run. (7) The simulation tool updates the input settings for a new run.

Figure 3. *Simulation Framework*



The operation of the factory can be simulated over long periods - up to one year - with a time increment of one minute at least. Table 2 describes the main factors that are involved in the simulation process at different stages.

Table 2. *Main Simulation Factors*

Time	Start date of the simulation: parameter
	Simulation during weekends (yes, no)
	No simulation on Friday afternoons for clean-up (yes, no)
Failure generation	Failure simulation (yes, no)
	Time occurrence
	Number of runs before failure per equipment
	Repair response time
	Range of the response time per equipment
	Range of the random failures per equipment
Sampling	Sampling range per hour of operation
Outputs	Recording of the data (yes, no)
	Recording increment (day, minute)
Environment	Indoor and outdoor temperatures (°C)
	Indoor humidity (%)
	Particulate matter (ppm)
	Range of particulate matter when the fan is on (ppm)
	Thresholds that set the increase and decrease of the indoor temperature (°C)
Production	Theoretical time to produce a lot (seconds)
	Number of pastries in a lot
	Weight of a pastry (g)
	Number of lots produced during a run
	Theoretical power consumption per equipment (W)
	Thresholds for PID and engine power consumption (W)
	Refill time for raw material tanks (min)
	Temperature thresholds for fan operation (°C)
	Thresholds for increase and decrease of indoor temperature and particulate matter
	Thresholds for baking quality indicators, humidity index, ppm quality indicator, cooling quality indicator

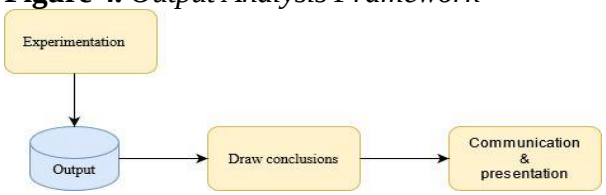
The simulation outputs are provided in Table 4. Among them is the production quality indicator which is calculated as a linear combination of 5 stochastic quality indicators: 1) a baking quality indicator which corresponds to the time spent by a lot of pastries in the oven. It follows a Gaussian distribution around a theoretical baking time with a given standard deviation; 2) an error function that ensures that products that have been in the oven for a long time are not systematically rejected if the oven is at low temperature; 3) a humidity index, namely humidex, which combines temperature and humidity in one computed value. The higher the humidex, the more mold can grow; 4) a ppm quality indicator which corresponds to an error function that reflects the quantity of particles suspended in air; 5) a cooling quality indicator which corresponds to an error function that reflects the time spent by a lot of pastries between the oven and the bagger; this is a critical time during which bacteria can grow. Production quality is a standardized metric with values ranging from 0 to 1. The closer the value is to 1, the higher is the production quality.

At last, the simulation tool was implemented in Python 3.6 with standard libraries.

Output Analysis Framework

After designing a simulation model and implementing a corresponding program, one must perform appropriate output analyses. As shown in Figure 4, the main activities related to the analyses of output data are experimentation, drawing reliable conclusions, communication, and presentation.

Figure 4. *Output Analysis Framework*



Since simulations of discrete event systems involve some randomness, the output of the simulation is effectively random variables. Thus, in principle, one can use the data analysis methods taught in statistics and data analysis courses to analyze the simulated data. However, in general, simulation data are not independent, and additional efforts may be needed to account for the dependency of simulation output.

Cases of Study

We simulated 10 months of operation of the factory, from 1 March to 31 December. The weekly production load during the simulated period is given in

Figure 5. For example, from 1 March to 21 March inclusive, the production load per week was 70%.

Figure 5. *Weekly Production Load (%)*

Quality sampling was carried out every two hours, starting at 00:00 for 13 minutes. In the following, we introduce the categorical variable Daytime which refers to the sampling periods (cf. Table 3).

Table 3. *Categorical Variable Daytime*

Sampling period	00:00 00:13	02:00 02:13	04:00 04:13	06:00 06:13	08:00 08:13	10:00 10:13
Daytime	0am	2am	4am	6am	8am	10am
Sampling period	12:00 12:13	14:00 14:13	16:00 16:13	18:00 18:13	20:00 20:13	22:00 22:13
Daytime	12am	2pm	4pm	6pm	8pm	10pm

We recorded 3,762 entries, one per sample, and for each sample, the variables in Table 4 were filled in.

Table 4. *Output Variables*

Categorical variables	Continuous variables	Discrete variables
Day and Daytime	Weekly production load: %	Number of staffers
Equipment Failure (oven, kneader, bagger, fan, conveyors): yes, no	Equipment downtime (oven, kneader, bagger, fan, conveyors): min	Number of lots produced
	Energy consumption per equipment (oven, kneader, bagger, fan, conveyors): W	Number of samples
	Production quality indicator	Number of rejects
	Tank level per ingredient (egg, flour, butter, sugar): %	
	Outdoor temperature: °C	
	Indoor temperature: °C	
	Outdoor humidity: %	
	Oven temperature: °C	
	Particulate matter: ppm	

In the following, we use some techniques from the Statistics and Data Analysis course to analyze the data collected. Two questions are posed to illustrate the use that can be made of the simulated data. All the calculations presented were carried out in R, free software for data analysis (R Core Team, 2012).

Question 1: What is the impact of fan failures on the factory's indoor temperature and particulate matter?

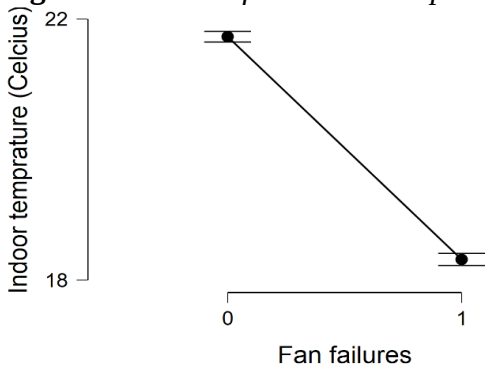
Temperature

We used a two-sided independent t-test to determine if there is a statistical difference between the average indoor temperatures in case of fan failure

(N_1=2148, m_1=21.73, s_1=1.93) versus no fan failure (N_2=1524, m_2=18.32, s_2=1.83). We found out that fan failures significantly increase the indoor temperature of the factory ($t(3670)$=53.95, p<0.001). Cohen's d (1.81) suggests that this is a large effect. The 95% confidence interval (CI) for the difference between temperature means is 3.29°C - 3.54°C and it suggests that the true increase in temperature means is likely to be within this range 95% of the time. Figure 6 depicts the 95% CIs of the indoor temperature by fan failure occurrence; the centers of the CIs are connected by segments for better graphical readability.

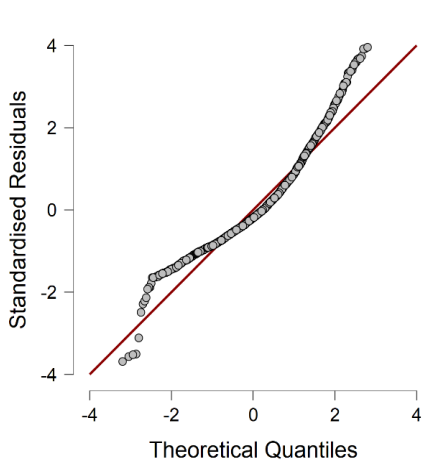
The assumptions of the independent t-test require: (1) independence of the two groups (with and without fan failures groups are independent); (2) the dependent variable should be approximately normally distributed in each group. The QQ-plots in Figure 7 show deviation from normality of the two distributions, which is confirmed by Shapiro-Wilk tests (without fan failures: W_1=0.93, p<0.001; with fan failures: W_2=0.92, p<0.001). However, we have very large sample sizes, and we can still use t-tests; (3) homogeneity of variance is tested using Levene's test of comparison of variances and it showed no difference between the variances of the two groups.

Figure 6. *95% CIs for Indoor Temperature*

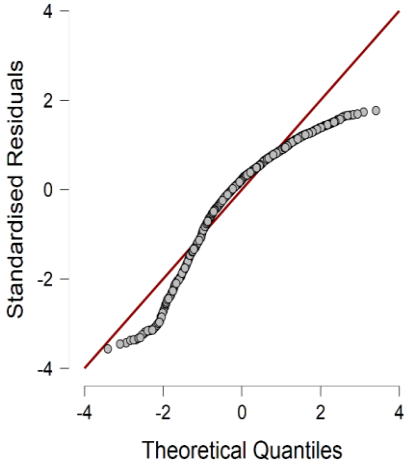


Note: 0 and 1 stand for 'without' and 'with' fan failures respectively.

Figure 7. *QQ-Plots of Indoor Temperature*



Note: without fan failures.



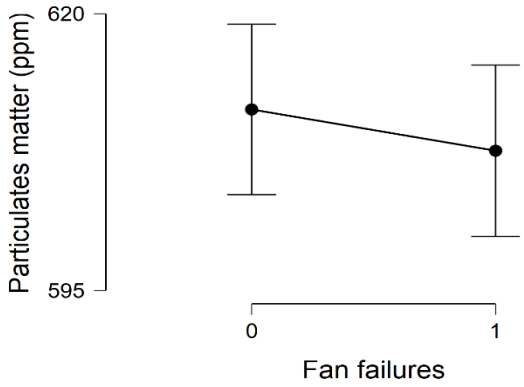
Note: with fan failures.

Particulate Matter

We used a two-sided independent t-test with Welch correction, to compare the average indoor temperatures in case of fan failure (N_1=2148, m_1=611.36, s_1=181.99) versus no fan failure (N_2=1524, m_2=607.63, s_2=154.19). There is no evidence of an effect of fan failures on the amount of particulate matter in the air (t(3556.7)=0.67, p=0.50 ns); Figure 8 shows the large overlap of the 95% confidence intervals of the two groups.

When checking t-test assumptions, we found out that the normality assumption was violated in both groups as assessed by Shapiro-Wilk tests (without fan failure: W_1=0.91, p<0.001; with fan failure: W_2=0.97, p<0.001), but this is not an obstacle to using t-tests as we have very large samples. On the other hand, Levene's test showed a violation of the equality of variances condition; we corrected for this violation by using an adjusted t-test statistic based on the Welch method.

Figure 8. *95% CIs for Particulate Matter*



Note: 0 and 1 stand for 'without' and 'with' fan failures respectively.

Question 2: Does the production quality indicator change over time and is it affected by the weekly charge load?

Tables 5 and 6 provide descriptive statistics of the production quality indicator by weekly charge load and daytime respectively. Differences in average quality scores can already be observed by weekly load and over time. It is yet to investigate whether these differences are significant and whether there is an interaction effect between the weekly production load and the sampling timetable.

Table 5. *Production Quality Indicator by Weekly Production Load*

	Weekly charge load (%)					
	50	60	70	80	90	100
Sample size	504	252	839	1008	672	393
Mean	0.87	0.77	0.89	0.85	0.85	0.78
Std. Deviation	0.07	0.05	0.08	0.09	0.09	0.07
Minimum	0.63	0.57	0.60	0.57	0.53	0.60
Maximum	0.98	0.91	0.98	0.98	0.97	0.96

Table 6. *Production Quality Indicator over Time*

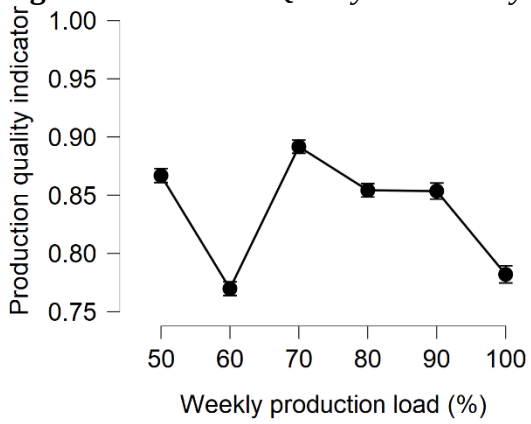
	Daytime											
	0am	2am	4am	6am	8am	10am	12am	2pm	4pm	6pm	8pm	10pm
Sample size	305	306	306	306	306	306	306	306	305	305	305	306
Mean	0.86	0.86	0.84	0.85	0.87	0.86	0.84	0.84	0.87	0.86	0.83	0.84
Std. Deviation	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
Minimum	0.63	0.63	0.60	0.64	0.60	0.61	0.61	0.57	0.60	0.57	0.63	0.53
Maximum	0.98	0.97	0.96	0.97	0.98	0.97	0.96	0.97	0.98	0.98	0.96	0.96

A two-way independent Anova (Analysis of variance) was conducted to examine the effect of the weekly production load and daytime on the production quality indicator. There were significant main effects for both weekly production load ($F(5,3595)=147.62$, $p<0.001$) and daytime ($F(5,3595)=6.51$, $p<0.001$). There was no evidence of an interaction effect ($F(55,3595)=0.26$, ns). Omega squared measure suggests a large main effect of the weekly production load ($\omega^2=0.17$) and a small effect of daytime ($\omega^2=0.014$). After checking Anova's assumptions, we carry out post hoc testing to go further.

Effect of the Weekly Production Load

Figure 9 depicts the behavior of the production quality according to the weekly production load; p-values and confidence intervals are adjusted for comparing a family of 6 estimates using Tukey's correction method. The centers of the confidence intervals are connected by segments for better graphical readability.

Figure 9. *Production Quality Indicator by Weekly Production Load*



We observe that the best quality scores are achieved for a weekly production load of 70% and the lowest for 60% and 100% loads. These findings are confirmed by Tukey's pairwise comparisons which are summed up in Table 7.

Table 7. *Tukey's Post Hoc Comparisons - Weekly Production Load (%)*

Weekly Production Load (%)		Mean Difference	95% CI for Mean Difference		*t*	p_{tukey}
			Lower	Upper		
60	50	-0.10	-0.12	-0.08	-15.20	< 0.001 ***
	70	-0.12	-0.14	-0.10	-20.49	< 0.001 ***
	80	-0.08	-0.10	-0.07	-14.47	< 0.001 ***
	90	-0.08	-0.10	-0.07	-13.70	< 0.001 ***
70	50	0.02	0.01	0.04	5.32	< 0.001 ***
	80	0.04	0.03	0.05	9.69	< 0.001 ***
	90	0.04	0.03	0.05	8.89	< 0.001 ***
	100	0.11	0.10	0.12	21.66	< 0.001 ***
100	50	-0.08	-0.10	-0.07	-15.22	< 0.001 ***
	80	-0.07	-0.09	-0.06	-14.65	< 0.001 ***
	90	-0.07	-0.09	-0.06	-13.60	< 0.001 ***

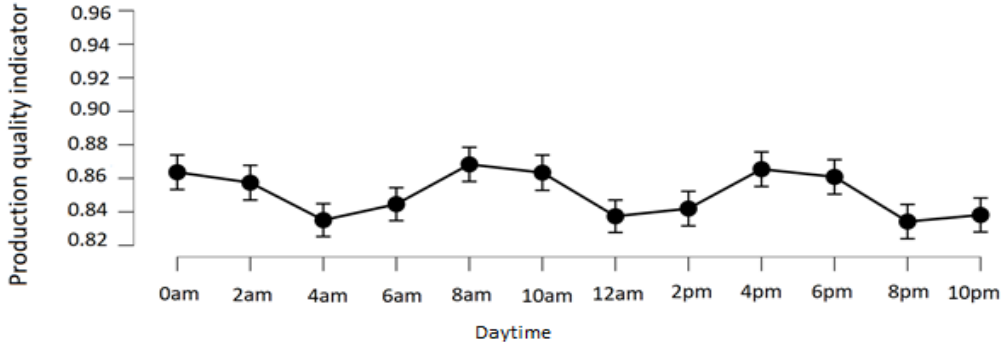
* p < 0.05, ** p < 0.01, *** p < 0.001.
Note: Results are averaged over the levels of Daytime.
Note: The pairwise comparisons that are not presented are non-significant.

A low-quality score for a 100% production load can be understood by the fact that higher workloads lead to higher equipment utilization and therefore higher risk of breakdowns and sub-quality. On the other hand, such a low score for a 60% workload should raise questions and call for investigations to find potential sources and correct them.

Effect of Daytime

Figure 10 depicts the evolution of the production quality over time; p-values and confidence intervals are adjusted for comparing a family of 12 estimates using Tukey's correction method.

Figure 10. *Production Quality Indicator over Time*



One can observe that the best quality scores are achieved for 0am, 8am, 10am, 4pm, and 6pm, and the lowest for 4am, 12am, 8pm, and 10pm. These findings are confirmed by Tukey's pairwise comparisons which are given in Table 8.

Table 8. *Tukey's Post Hoc Comparisons – Daytime*

Daytime		Mean difference	95% CI for mean difference		*t*	p_{tukey}	
			Lower	Upper			
4am	0am	-0.03	-0.05	-4.13e-3	-3.82	< 0.01	**
	8am	-0.03	-0.06	-7.79e -3	-4.32	< 0.001	***
	10am	-0.03	-0.05	-3.53e -3	-3.74	0.01	*
	4pm	-0.03	-0.05	-5.85e -3	-4.05	< 0.01	**
	6pm	-0.03	-0.05	-2.46e -3	-3.60	0.02	*
12am	0am	-0.03	-0.05	-2.10e -3	-3.55	0.02	*
	8am	-0.03	-0.05	-5.77e -3	-4.04	< 0.01	**
	10am	-0.03	-0.05	-1.51e -3	-3.47	0.03	*
	4pm	-0.03	-0.05	-3.83e -3	-3.78	< 0.01	**
	6pm	-0.02	-0.05	-4.37e -4	-3.33	0.04	*
8pm	0am	-0.03	-0.05	-5.72e -3	-4.04	< 0.01	**
	8am	-0.03	-0.06	-9.38e -3	-4.53	< 0.001	***
	10am	-0.03	-0.05	-5.13e -3	-3.96	< 0.01	**
	4pm	-0.03	-0.06	-7.44e -3	-4.27	< 0.01	**
	6pm	-0.03	-0.05	-4.05e -3	-3.81	< 0.01	**
10pm	0am	-0.02	-0.05	-2.90e -4	-3.31	0.04	*
	8am	-0.03	-0.05	-3.95e -3	-3.80	< 0.01	**
	4pm	-0.03	-0.05	-2.01e -3	-3.54	0.02	*
2pm	8am	-0.03	-0.05	-2.74e -3	-3.64	0.01	*
	4pm	-0.03	-0.05	-7.97e -4	-3.38	0.04	*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Note: Results are averaged over the levels of Weekly production load.
Note: The pairwise comparisons that are not presented are non-significant.

In conclusion, the two case studies presented illustrate some practical uses of the simulation tool in a statistics and data analysis course. We have covered some descriptive and inferential statistics topics (confidence intervals, normality tests, tests of equality of means and variances, Anovas, etc.), however the diversity of the output variables (continuous, discrete, categorical) allows for a wider variety of analyses (linear regression, logistic regression, multiple component analysis, etc.). An advantage of this tool is also the ability to generate customized datasets, to focus on a particular technique and explore its multiple facets.

Discussion and Conclusion

We have proposed a simulation tool to generate realistic industrial datasets and we have presented two case studies to illustrate their utilization in the context of a statistics and data analysis course for IE undergraduate students. The examples treated involve common concepts of descriptive and inferential statistics. However, the varied nature of the output variables allows for a wider range of analysis techniques.

The simulated data can serve the analysis of the production behavior of manufacturing systems and the identification of the most significant key

performance indicators (overall equipment effectiveness, capacity...). They can also be used to identify the critical phases of the production process and to understand the involvement of the physical environment in production quality.

As for future evolutions of the simulation tool, randomness is currently generated via uniform and Gaussian distributions, and we plan to introduce other statistical distributions for specific events (Exponential and Erlang for inter-arrival times, triangular, Beta, Normal and LogNormal for service times, Weibull for inter-failure times…).

It also remains to develop a user interface that allows students to design their datasets and their quality indicators for subsequent statistical analyses. We believe that when students are actively involved in an experiment, they internalize better the material being taught and mobilize more inner resources for learning. The industrial context could help to get more commitment of the students by projecting them in their future profession and following that, to adopt a professional posture in the analysis of the problem, the choice of the data analyses to carry out, their rigorous application, and the restitution of the results in a form that is clear, concise, and adapted to the recipient.

References

Carrera, E. 2006. Engineering statistical needs an engineering curriculum: an analysis. *ICOTS7*, Brazil.

Davies, N. 2001. A new way to teach statistics to engineers. *MSOR Connections*, 1(2), 7-9.

Elizandro, D., Taha, H. 2007. Simulation of Industrial Systems: Discrete Event Simulation Using Excel/VBA (1st. ed.). *Auerbach Publications*, USA.

Eskandari, H., Sala-Diakanda, S., Furterer, S., Rabelo, L., Crumpton-Young, L. and Williams, K. 2007. Enhancing the undergraduate industrial engineering curriculum: Defining desired characteristics and emerging topics, *Education + Training*, Vol. 49 No. 1, pp. 45-55.

Fraser, J. 2015. Benchmarking IE programs: 2005-2015. *2015ASEE Annual Conference and Exposition*, 14-17 June, Seattle, WA.

Fraser, J. and Teran, A. 2006. Benchmarking international industrial engineering programs. *2006 ASEE Annual Conference & Exposition*, 18-21 June, Chicago, Illinois.

Kuo, W. 2001. Educational programs for the industrial engineer. *Maynard's Industrial Engineering Handbook, 5th edition*, McGraw-Hill, New York, pp. 1.39-1.53.

Lang, J.D., Cruse, S., McVey, F.D. and McMasters, J. 1999. Industry Expectations of New Engineers: A Survey to Assist Curriculum Designers. *Journal of Engineering Education*, 88: 43-51.

Lima, R.M., Mesquita, D., Amorim, M., Jonker, G. and Flores, M. A. 2012. An analysis of knowledge areas in industrial engineering and management curriculum. *International Journal of Industrial Engineering and Management*, Vol. 3 No. 2, pp. 75-82.

Nguyen, A.T. and Nguyen, N.V.-P. 2018. Benchmarking industrial engineering programs. *Benchmarking: An International Journal*, Vol. 25 No. 4, pp. 1194-1212.

R Core Team. 2012. R: A language and environment for statistical computing. *Foundation for statistical Computing*. Vienna, Austria.

Ram, Mangey & Davim, J. 2018. Modeling and Simulation in Industrial Engineering. 10.1007/978-3-319-60432-9.

Sackey, S. M., Bester, A. 2016. Industrial engineering curriculum in industry 4.0 in a South African context. *South African Journal of Industrial Engineering,* December 2016, Vol 27(4), pp 101-114.

Schruben, D.L. and Schruben, L.W. 2001. Event Graph Modeling with SIGMA, *4th Ed. Custom Simulations*.