Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

https://doi.org/10.23913/ride.v14i28.1790

Scientific articles

Estimación de demanda estudiantil en una institución educativa para calcular capacidad docente y aulas utilizando modelo de pronósticos

Estimation of student demand in an educational institution to calculate teaching and classrooms capacity using a forecast model

Estimativa da demanda estudantil em uma instituição de ensino para cálculo da capacidade docente e de salas de aula por meio de modelo de previsão

Manuel Alonso Rodríguez Morachis
Tecnológico Nacional de México, Instituto Tecnológico de Ciudad Juárez, México
mmorachis@itcj.edu.mx
https://orcid.org/0000-0003-1581-7737

Luz Elena Terrazas Mata
Tecnológico Nacional de México, Instituto Tecnológico de Ciudad Juárez, México
lterrazas@itcj.edu.mx
https://orcid.org/0000-0003-1396-1751

Francisco Zorrilla Briones
Tecnológico Nacional de México, Instituto Tecnológico de Ciudad Juárez, México
fzorrilla@itcj.edu.mx
https://orcid.org/0000-0003-0553-9841

Resumen

El objetivo de este artículo es analizar el comportamiento histórico de la matrícula estudiantil en un bachillerato tecnológico de Ciudad Juárez, Chihuahua, con la finalidad de estimar la demanda futura. El propósito es determinar la capacidad requerida de infraestructura en la institución en lo referente a aulas y personal docente, y calcular las proporciones alumnos-maestro y alumnos-aula para compararlos con las proporciones medias internacionales y de México. Esto se debe a que, en los inicios de los ciclos escolares, se suelen presentar situaciones de falta de capacidad para atender al alumnado en tiempo, por lo que se tuvo que improvisar en la adecuación de las instalaciones y personal docente. Para este estudio, se revisaron diferentes métodos de pronósticos cuantitativos; a partir de esto, y de acuerdo con el comportamiento histórico de los datos, se utilizó el que más se apegaba a dicho comportamiento. En concreto, se colectaron datos de 15 periodos lectivos disponibles desde su inauguración. El tratamiento de los datos inició con pruebas de normalidad con el objetivo de establecer el estadístico que se utilizó. Una vez utilizado el modelo de pronósticos apropiado, se validó con 5 periodos para comprobar si los pronósticos calculados se apegan a los datos reales que se presentaron del periodo 16 al 20; asimismo, se realizaron las pruebas de hipótesis estadísticas para validación del modelo utilizado. Además, se efectuaron pruebas con diferentes factores de suavización de nivel y tendencia, y se empleó el que mostró el menor error por periodo, el cual varió desde 0.21 % a 8.88 %. Las proporciones alumnos-maestro y alumnos-aula encontradas fueron de 36 y 38, respectivamente.

Palabras clave: pronóstico, infraestructura, demanda escolar, suavizamiento exponencial, estadística paramétrica.

Abstract

The objective of this paper is to analyze the historical behavior of student enrollment in a technological high mid-level school, in Ciudad Juárez, Chihuahua, in order to estimate future demand, to identify the required infrastructure capacity in the institution, in terms of classrooms and teaching staff and calculate the student-teacher and student-classroom ratios and compare them with the international and Mexican average ratios. This is due to the fact that at the beginning of the school cycles, there were situations of lack of capacity to serve the students on time, having to improvise in the adequacy of the facilities and teaching staff. Different methods of quantitative forecasts were reviewed that, according to the historical behavior of the data, the one that most closely adheres to mentioned behavior was used. Data from 15 school periods available since its inauguration were collected. Data treatment begins with normality tests in order to establish the statistic that was used. Once the appropriate forecast model was used, it was validated with 5 periods to check if the calculated forecasts adhered to the real data that was presented from period 16 to 20, likewise, the statistical hypothesis tests were carried out for validation of the models used. Tests were performed with different level and trend smoothing factors, using the one that showed the lowest error per period, which ranged from 0.21% to 8.88%. The ratios student-teacher and student-class found were 36 and 38 respectively.

Keywords: Forecasting, infrastructure, school demand, exponential smoothing, parametric statistics.

Resumo

O objetivo deste artigo é analisar o comportamento histórico da matrícula de alunos em uma escola secundária tecnológica em Ciudad Juárez, Chihuahua, a fim de estimar a demanda futura. O objetivo é determinar a capacidade de infraestrutura necessária na instituição em termos de salas de aula e pessoal docente, e calcular as proporções aluno-professor e aluno-sala de aula para compará-las com as proporções médias internacionais e mexicanas. Isto porque, no início dos ciclos escolares, costumam surgir situações de falta de capacidade para atender os alunos atempadamente, razão pela qual foi necessário improvisar na adaptação das instalações e do corpo docente. Para este estudo, foram revisados diferentes métodos de previsão quantitativa; A partir disso, e de acordo com o comportamento histórico dos dados, foi utilizado aquele que mais aderiu a esse comportamento. Especificamente, foram coletados dados de 15 períodos letivos disponíveis desde a sua inauguração. O tratamento dos dados iniciou-se com testes de normalidade para estabelecer as estatísticas utilizadas. Uma vez utilizado o modelo de previsão adequado, foi validado com 5 períodos para verificar se as previsões calculadas estão de acordo com os dados reais que foram apresentados do período 16 ao 20; Da mesma forma, foram realizados testes de hipóteses estatísticas para validar o modelo utilizado. Além disso, foram realizados testes com diferentes níveis e fatores de suavização de tendência, sendo utilizado aquele que apresentou menor erro por período, que variou de 0,21% a 8,88%. As proporções aluno-professor e aluno-sala encontradas foram de 36 e 38, respectivamente.

Palavras-chave: previsão, infraestrutura, demanda escolar, suavização exponencial, estatística paramétrica.

Reception Date: June 2023 **Acceptance Date:** January 2024

Introduction

Education plays a very important role in the acquisition of knowledge, which has an impact on changing thinking and improving social conditions. This is especially evident in upper secondary education, since it is the educational period that prepares the student to decide their academic future towards the bachelor's level or to enter the workplace with a technical career. Likewise, it constitutes the stage in which young students begin to understand the social, economic, educational and political behavior of the environment in which they operate.

However, an important aspect to consider within the city environment is the insecurity that has prevailed during the last 15 years. Therefore, it is relevant to provide young people with the necessary spaces so that their studies after secondary education are not cut short, which includes addressing infrastructure needs in accordance with historical demand.

In this regard, Pacheco-Martínez (2021) highlights that the physical educational infrastructure must be taken into account to promote the development of students in a quality environment. For its part, according to the Organization for Economic Cooperation and Development (OECD) (2021), the teacher-student ratio for the secondary level has an average of 13 students per teacher, while in Mexico the ratio is 28 students per teacher. Regarding class size - that is, the number of students per classroom - in Mexico there are 27 students, while the OECD average (2021) is 23. Furthermore, in the private education sector, the ratio is 14 students per teacher.

According to the General Directorate of Planning, Programming and Educational Statistics (DGPPEE) of the Ministry of Public Education (SEP) (2022), in the state of Chihuahua there are 6 schools of the same type as the institution under study, of which 5 of them are located in Cd. Juárez, Chihuahua. Due to the above, it is important to forecast student demand interested in this type of education to increase enrollment capacity, since in a city with 141 secondary education schools that have 79,835 students, it is critical to plan the capacity of high school educational institutions, especially in technological high schools, which are scarce in the city.

Now, the institution in which this research is carried out is part of a government program that seeks to facilitate access to education, both at the secondary and higher levels. This institution is located in the south of the city, in an area characterized by poor mobility and little in terms of transportation systems. Furthermore, it belongs to a sector of the city with high demographic growth, considered low-income and vulnerable. For this reason, the program was designed with the purpose of serving the student segment residing in this area, bringing the institutions closer to the students and avoiding mobility and possible desertion due to the remoteness of the facilities in the urban area.

In his analysis of the growth of enrollment and teaching staff in Mexico in high school (from the 1963-1964 cycle to 2011-2012), Olvera (2013) points out that the former has increased 30.9 times, while the latter has only grown 17.2 times. This shows a disproportion in the capacity necessary to meet the demand in terms of quality in attention to students. It should be noted that, according to INEGI (2023 a, 2023b), the increase in enrollment continues in similar proportions today, with increases of 19.5% and 8.1%, respectively. In other words, the difference between the ratio of students and teaching staff is approximately 55%, with a decrease due to the Covid-19 pandemic.

On the other hand, being a public sector educational institution, it is crucial to plan how to cover the needs required to provide quality care to high school students in the aforementioned sector. The process of hiring teaching staff is lengthy, since the authorization of resources depends on the annual budget authorized for the SEP.

In this regard, Armstrong and Kotler (2013) maintain that companies must first decide which customers they are going to serve, since customer satisfaction depends on perceived performance in relation to their expectations. Likewise, they consider demographic segmentation, which can range from the country, state, municipality, city and even neighborhoods. Another aspect that they define is segmentation by benefits, where they classify it according to the different benefits that customers seek to obtain. In this case, the company would be the educational authority, and the clients would be the students to be served.

In short, and following the perspective of Kotler and Keller (2011), who indicate that in the pattern of homogeneous preferences of market segmentation all consumers have the same preference and there are no natural segments, this research focuses on considering the behavior demand history.

Theoretical framework

The application of statistical forecasting models to estimate future demand in the institution under study has not been fully addressed. For example, similar studies have been carried out, such as the case of Tarango-Hernández *et al*. (2019), who carried out research at the Language Center of the National Technology of Mexico (Technological Institute of Ciudad Juárez), with the purpose of improving the service to the students of this center through the use of causal models.

Likewise, Sepúlveda-Silvestre (2020) used the Box-Jenkins methodology to determine student demand at the graduate level with the aim of guaranteeing sufficient infrastructure. Likewise, Kleiman (1975) points out that public higher education institutions base their forecasts on economic indices, for which they have as a reference the proportion of the student-age population by level, although he also highlights the existence of causal statistical methods and time series.

For its part, Tenjo-Galarza (2012) presents demographic and econometric methodologies to project student demand in Colombia. In the case of Miller *et al*. (2021) point out that determining student demand at a professional level in Mexico is not easy, since students consider several universities at the same time when starting their higher education, which is why they present a sociodemographic analysis to predict the demand for admission to a public educational institution.

James and Weese (2022) carry out an analysis of the application of neural networks based on the simple exponential smoothing model, with mean square errors of 0.27 and 0.24, respectively, a difference that, for practical purposes, is irrelevant. On the other hand, Silitonga *et al*. (2020) apply the double exponential smoothing model to predict the acceptance of new students at an Indonesian university, where they found a mean estimation error of 0.1172.

Chen (2022) presents a comparison between several models, including gray, simple exponential smoothing, neural networks, and ARIMA, applied in Chinese vocational schools. This author concludes that the exponential smoothing model yields the lowest mean absolute error, with 0.192. For the neural network model, estimate a mean absolute error of 0.216; for ARIMA, 0.196; and the gray model shows a mean absolute error of 0.237. However, neither James and Weese (2022) nor Chen (2022) consider the trends or cyclicalities of student demands.

Cirelli *et al*. (2018) use several analytical models to analyze the enrollment demand at an educational institution in the city of Washington, USA. The models used include neural networks, logistic regression, Bayesian networks, decision trees, CHAID, and SVM. On average, all these models show an accuracy of 72%. Truckman (1971) uses the least squares regression model to obtain the probabilities of graduating from vocational education in a sector of the minority population in the state of Florida, USA, for which he considers factors such as family income and the educational level of the parents.

Hill and Fildes (1984), Lusk and Neves (1984), Makridakis *et al*. (1982), Geurts and Kelly (1986), Clemen (1989), Filders *et al*. (1998), Koehler and Murphree (1988), Makridakis and Hibon (2000), among others, have conducted comparative investigations of different forecasting models, such as exponential smoothing, neural networks, machine learning, linear and nonlinear regression, etc. The general conclusion from these studies is that exponential smoothing models tend to provide more accurate forecasts.

As can be seen, the choice of forecast model depends on the behavior of historical data, time horizons and other factors. Likewise, it can be indicated that both quantitative and qualitative models have been developed to address this task. Within the quantitative models there are causal models and time series models, the latter being the ones addressed in this research due to what was mentioned above.

In the context of the public upper secondary education model in Mexico, most institutions present a pattern of cyclicality or seasonality in each school year, with higher enrollment in the second periods of each year. Although sometimes this pattern may not be very pronounced, it will be decided to use exponential smoothing models to carry out a predictive study of the student demand that will enter this public institution of higher secondary education. We will start with the simplest model, which does not consider seasonality or cyclicality, and is known as the Brown model. Subsequently, the Holt model will be presented, which considers level and trend and, finally, the Winters model will be addressed, which incorporates level, trend and seasonality.

Simple exponential smoothing (Brown)

The main time series models are exponential smoothing models, which are extensions of the one developed by Brown (1956) during World War II as an operations research analyst in 1944. This model does not consider trends or seasonality, according to with equation (1):

$$P_t = \propto D_{t-1} + (1-\propto) P_{t-1} \tag{1}$$

As:

P_t It is the forecast at time t

P_{t-1} It is the forecast at time $t-1$

$\propto$ It is the smoothing parameter for the series level ($0 \leq \propto \leq 1$)

D_{t-1} It is the demand or data at time $t-1$

Regrouping the terms of (1):

$$P_t = \propto D_{t-1} + P_{t-1} - \propto P_{t-1} \tag{2}$$

$$P_t = P_{t-1} + \propto (D_{t-1} - P_{t-1}) \tag{3}$$

Likewise, if it is considered that it starts at time t, and then equation (3) can be transformed to:

$$P_{t+1} = P_t + \propto (D_t - P_t) \tag{4}$$

In equations (3) and (4) the terms $(D_{t-1} - P_{t-1})$ and $(D_t - P_t)$ are the forecast errors corresponding to the periods $t - 1$ and t respectively, so:

$$(D_{t-1} - P_{t-1}) = e_{t-1} \tag{5}$$

$$(D_t - P_t) = e_t \tag{6}$$

Being equations (3) and (4) the most used, depending on the period considered as the first forecast (t or $t + 1$).

Double exponential smoothing (Holt)

The double exponential smoothing model is an extension of the Brown (1956) model, presented by Holt (2004), which includes the trend of historical data, so, in addition to the level parameter, it contains a trend smoothing parameter, and is expressed as equations (7), (8) and (9):

$$N_t = \propto D_{t-1} + (1-\propto)(N_{t-1} + T_{t-1}) \tag{7}$$

$$T_t = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1} \tag{8}$$

$$P_t = N_t + T_t \tag{9}$$

As:

P_t It is the forecast at time t

N_t It is the level smoothing at time t

T_t It is the smoothing of the trend at time t

D_{t-1} It is the demand or data at time $t-1$

$\propto$ It is the smoothing parameter for the series level $(0 \leq \propto \leq 1)$

β It is the trend smoothing parameter of the series $(0 \leq \beta \leq 1)$

The forecast for τ time periods is expressed by equation (10):

$$P_{t,t+\tau} = N_t + \tau T_t \tag{10}$$

Triple exponential smoothing (Winters)

This smoothing model developed by Winters (1960) considers, in addition to the level and trend parameters, cyclicality or seasonality.

With a periodicity *p*, at time *t*, and the calculation of level estimates N_t, trend T_t, and stationary factors $S_t, \dots, S_{t+p-1}$, the forecast for future states is determined by equations (11) and (12):

$$P_{t+1} = (N_t + T_t)S_t \tag{11}$$

$$P_{t+l} = (N_t + lT_t)S_{t+l} \tag{12}$$

So the demand for *t* +1 is expressed as:

$$N_{t+1} = \propto \left(\frac{D_{t+1}}{S_{t+1}}\right) + (1-\propto)(N_t + T_t) \tag{13}$$

$$T_{t+1} = \beta(N_{t+1} - N_t) + (1-\beta)T_t \tag{14}$$

$$S_{t+p+1} = \delta\left(\frac{D_{t+1}}{N_{t+1}}\right) + (1-\delta)S_{t+1} \tag{15}$$

As:

P_{t+1} It is the forecast at time $t+1$

P_{t+l} It is the forecast at time t plus a period of time l

N_t It is level smoothing at time t

T_t It is the smoothing of the trend at time t

D_{t-1} It is the demand or data at time $t-1$

S_t It is the seasonality factor at time t

$\propto$ It is the smoothing parameter for the series level $(0 \leq \propto \leq 1)$

β It is the trend smoothing parameter of the series $(0 \leq \beta \leq 1)$

δ It is the smoothing parameter for the seasonality of the series $(0 \leq \delta \leq 1)$.

Winters model is more elaborate when considering the three variants of the behavior of the time series, so it is generally composed of two steps (Chopra and Meindl, 2013):

1. Deseasonalize the series data and perform linear regressions in order to estimate the level and trend parameters.

2. Estimate seasonality factors.

The deseasonalization of the data is carried out using

$$\bar{D}_t = \begin{cases} \dfrac{D_{t-\left(\frac{p}{2}\right)} + D_{t+\left(\frac{p}{2}\right)} + 2\sum_{i=t+1-\left(\frac{p}{2}\right)}^{t-1+\left(\frac{p}{2}\right)} D_i}{2p} & ;\text{for } p \text{ pair} \\ \displaystyle\sum_{i=t-[(p-1)/2]}^{t+[(p-1)/2]} \frac{D_i}{p} & ;\ \text{for odd } p \end{cases} \tag{16}$$

Obtaining equation (17) as a linear relationship between deseasonalized demand and the time:

$$\bar{D}_t = N + T_t \tag{17}$$

When the deseasonalized demand is calculated, the seasonality factor is calculated using equation (18).

$$\bar{S}_t = \frac{D_i}{\bar{D}_t} \tag{18}$$

For a periodicity p, with r stationary cycles, for the period's pt+1, for $1 \leq i \leq p$ the seasonal factor is expressed as:

$$S_i = \frac{\sum_{j=0}^{r-1} S_{jp+i}}{r} \tag{19}$$

Forecast error

It is important to determine the accuracy of forecast models by comparing the predicted data against the observed or actual data, as expressed in equation (6). The three most commonly used measures to calculate forecast error are the *Mean Absolute Deviation (*MAD*), the Mean Squared Error* (MSE) and the *Mean Absolute Percentage Error* (MAPE). Equations (20) (21) and (22) are expressions for calculating forecast errors:

$$\text{MAD} = \frac{\sum_{i=1}^{n}|D_i - P_i|}{n} \tag{20}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n}(D_i - P_i)^2}{n} \tag{21}$$

$$\text{MAPE} = \frac{100\ \sum_{i=1}^{n}|D_i - P_i|/D_i}{n} \tag{22}$$

Where *n* is the amount of data in the series.

Materials and method

The research carried out is quantitative, transversal, experimental and exploratory. For this study, the entire population was considered, since the creation of the higher secondary education institution under investigation, and school enrollment information was collected since January 2010. With the objective of establishing the statistics to be used to verify that predicted data are statistically equal to the actual data, normality tests were performed.

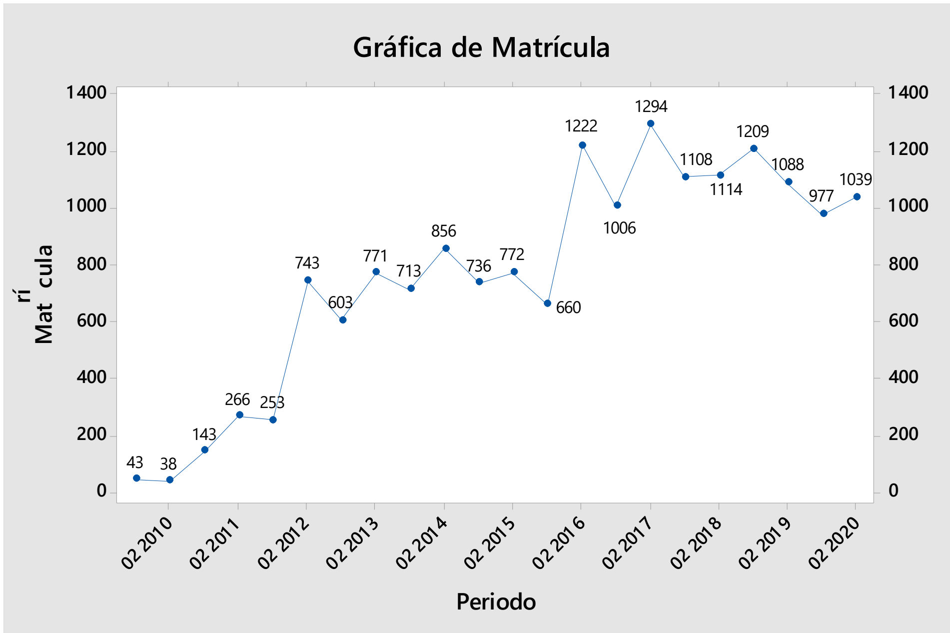
By observing the behavior of the data collected, it was searched to establish the forecast model that results in the greatest assertiveness for decision making. Additionally, data for the first 10 years were collected, as shown in Table 1.

By observing the behavior of the data in Figure 1, and following what was mentioned in the theoretical framework, a periodicity and trend are identified to establish the model to use in calculating the forecast. The investigation begins at the end of period 2 of 2020, and 5 verification periods of the calculated forecast are carried out, one by one.

Initially, the forecast for period 1 of 2021 is calculated using the Winters smoothing model, with a multiplicative option, since the seasonality pattern increases or decreases as the magnitude of the data increases or decreases.

After applying the forecast model, a normality test is performed to determine the statistic to use (parametric or non-parametric) and to contrast the real data with the predicted data. Details of the presented methodology are found in the results section, where each step performed is shown and the findings of the corresponding calculations are presented.

Figure 1. Behavior of enrollment from period 1 of 2010 to period 2 of 2020



Source: own elaboration

As seen in Figure 1, enrollment presents a trend and the school cycles of the educational system in Mexico are repeated twice a year. Although the cyclicality is not very clear in the graph, an analysis was carried out with the Holt model. However, it was found that the forecast errors per period are considerably above those obtained with the Winters model. Furthermore, the period forecasts differ significantly from the actual data analyzed during the 5 validation periods. For this reason, this analysis is not presented in detail.

Table 1. Behavior of enrollment from period 1 of 2010 to period 2 of 2020

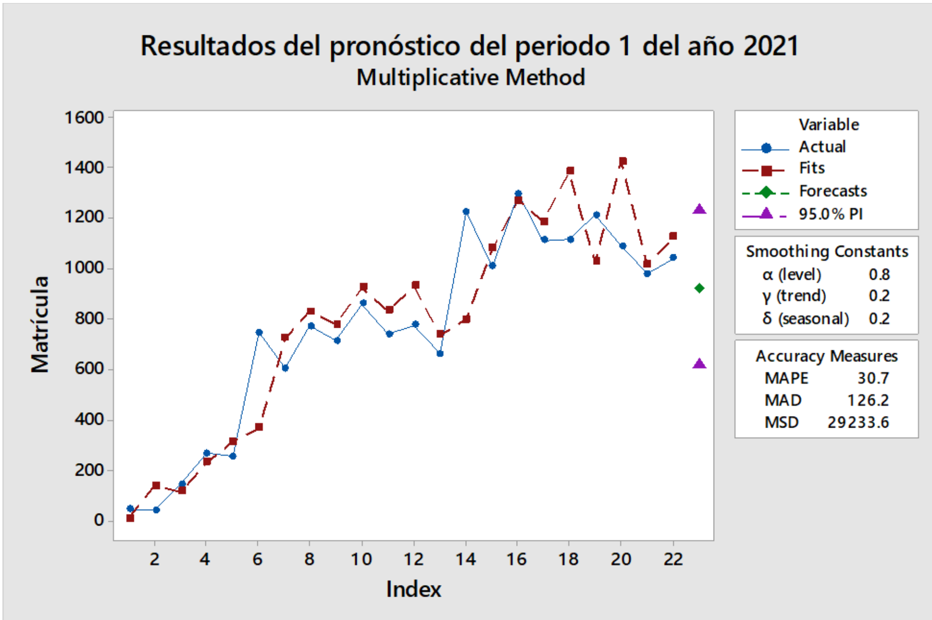
Year	Period	Tuition
2010	1	43
	2	38
2011	1	143
	2	266
2012	1	253
	2	743
2013	1	603
	2	771
2014	1	713
	2	856
2015	1	736
	2	772
2016	1	660
	2	1222
2017	1	1006
	2	1294
2018	1	1108
	2	1114
2019	1	1209
	2	1088
2020	1	944
	2	1039

Source: own elaboration

Results

After analyzing different values of α, β and δ , it was found that for the values of $\propto = 0.8$, β= 0.2 and δ= 0.2, the forecast of 919 with the lowest forecast error is obtained: MAD = 126.2, MSD = 29,233.6 and MAPE = 30.7 %. Figure 2 and Table 2 show the results. The calculated forecast is compared with the actual value of period 1 of the year 2021, which was 844. With this last data, the model is run again to forecast period 2 of the year 2021, and so on until period 2 of the year is forecast. 2023.

Figure 2. Analysis of enrollment behavior using the Winters model



Source: own elaboration

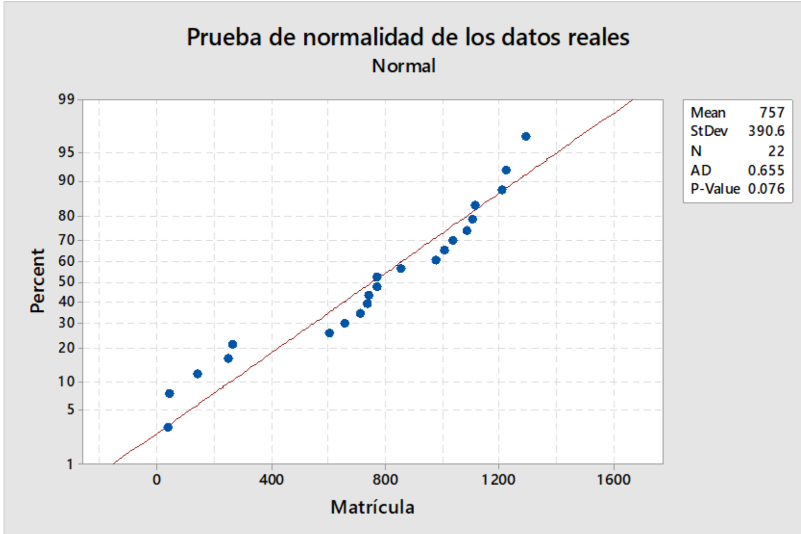
Table 2. Enrollment forecasts by period and for period 1 of 2021

Year	Period	Tuition	Forecast by period	Forecast for period 1 of 2021
2010	1	43	-63.55	919
	2	38	43.88	
2011	1	143	48.74	
	2	266	151.40	
2012	1	253	236.16	
	2	743	291.51	
2013	1	603	600.77	
	2	771	711.81	
2014	1	713	683.39	
	2	856	828.19	
2015	1	736	759.02	
	2	772	862.32	
2016	1	660	699.91	
	2	1222	768.76	
2017	1	1006	993.78	
	2	1294	1184.42	
2018	1	1108	1106.95	
	2	1114	1308.29	
2019	1	1209	999.44	
	2	1088	1357.07	
2020	1	944	1003.80	
	2	1039	1119.20	

Source: own elaboration

In order to establish the statistic to be used to test statistical equality that the predicted data are statistically equal to the actual data, the normality test was performed, which is shown in Figure 3.

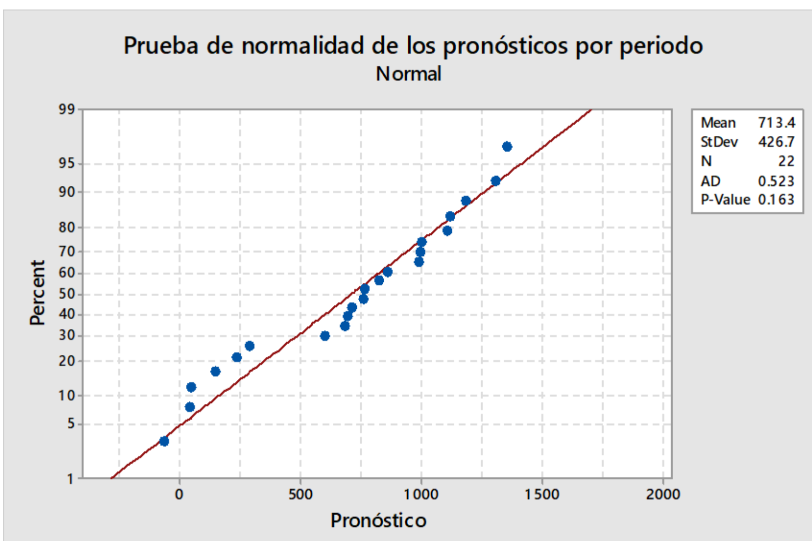
Figure 3. Normality test for real data from period 1 of 2010 to period 2 of 2020



Source: own elaboration

With the value of p = 0.076, it is concluded that the real data have normal behavior. Likewise, the normality test is carried out for the forecasts by period shown in Figure 4, whose value of p = 0.163, indicates that the forecasts of the periods also have a normal behavior, so the statistic to use is a comparison of means.

Figure 4. Test of normality for the forecasts from period 1 of 2010 to period 2 of 2020



Source: own elaboration

Once the normal behavior of both the actual values and the forecasts of the calculated periods has been verified, the null hypothesis that the means of both are statistically equal is tested against the alternative hypothesis that they are different. These hypotheses are shown in expressions (22) and (23):

$$H_0: \mu_{\text{Real data}} = \mu_{\text{Forecast}} \tag{22}$$

$$H_0: \mu_{\text{Real data}} \neq \mu_{\text{Forecast}} \tag{23}$$

The results of the hypothesis test are shown in Figure 5, while Figure 6 shows the box plot, where the similarity of means is observed. Consequently, there is not enough statistical evidence to not accept the null hypothesis, and it is established that the predicted data per period are statistically equal to the actual data with a p value = 0.726.

It is worth mentioning that starting from period 1 of 2022, the real data showed non-normal behavior, for which non-parametric statistics were used for hypothesis testing using the median as the test statistic, using Mann-Whitney . In other words, the data obtained per period and the actual data are statistically equal.

Figure 5. Result of the hypothesis test between averages of actual and forecast values by period

Two-Sample T-Test and CI: Matrícula, Pronóstico

Method

μ_1: mean of Matrícula
μ_2: mean of Pronóstico
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Matrícula	22	757	391	83
Pronóstico	22	713	427	91

Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
44	409	(-205, 292)

Test

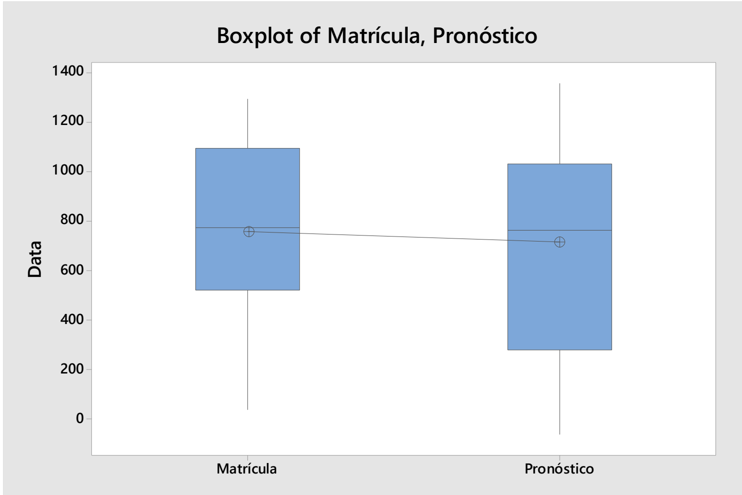
Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
0.35	42	0.726

Source: own elaboration

Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Figure 6. Boxplot comparing means



Source: own elaboration

On the other hand, it is desired that the variability of the data does not have heteroscedasticity, so the null hypothesis is tested that the variances of the real data. Furthermore, the forecasts per period are statistically equal against the alternative hypothesis that they are different. These hypotheses are shown in expressions (24) and (25):

$$H_0: \sigma^2_{\text{Real data}} = \sigma^2_{\text{Forecast}} \tag{24}$$

$$H_0: \sigma^2_{\text{Real data}} \neq \sigma^2_{\text{Forecast}} \tag{25}$$

The results are shown in Figure 7, where it is observed that for the Bonett and Levene tests the p values are 0.629 and 0.693 respectively. Consequently, there is not enough statistical evidence to do not accept the null hypothesis; In other words, the variances of the actual data and the forecasts per period are statistically equal, so there is no variability in both. Figure 8 shows the confidence interval between the variances, as well as the box plot of the comparison of variances.

Ride Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Figure 7. Result of hypothesis testing between variances of actual and forecast values by period

Test and CI for Two Variances: Matrícula, Pronóstico

Method

σ₁: standard deviation of Matrícula
σ₂: standard deviation of Pronóstico
Ratio: σ₁/σ₂
The Bonett and Levene's methods are valid for any continuous distribution.

Descriptive Statistics

Variable	N	StDev	Variance	95% CI for σ
Matrícula	22	390.592	152562.095	(305.509, 548.210)
Pronóstico	22	426.722	182091.600	(343.090, 582.648)

Ratio of Standard Deviations

Estimated Ratio	95% CI for Ratio using Bonett	95% CI for Ratio using Levene
0.915331	(0.605, 1.344)	(0.557, 1.493)

Test

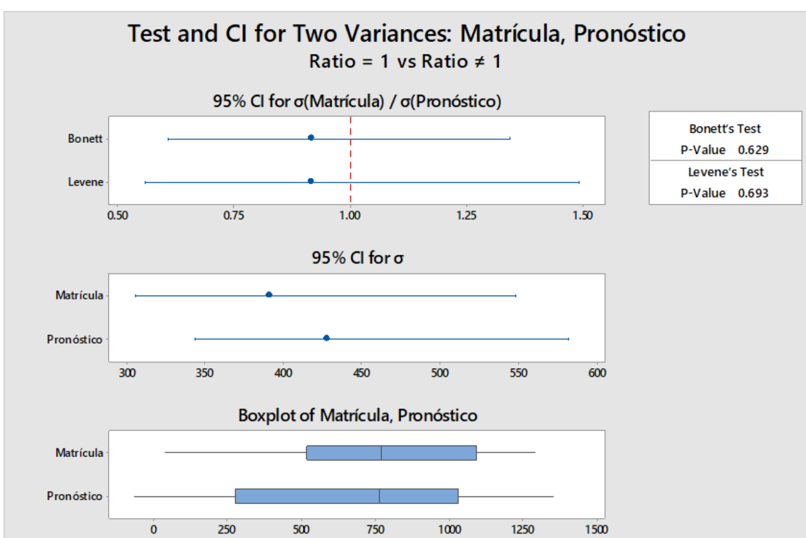
Null hypothesis $H_0: \sigma_1 / \sigma_2 = 1$
Alternative hypothesis $H_1: \sigma_1 / \sigma_2 \neq 1$
Significance level $\alpha = 0.05$

Method	Test Statistic	DF1	DF2	P-Value
Bonett	0.23	1		0.629
Levene	0.16	1	42	0.693

Test and CI for Two Variances: Matrícula, Pronóstico

Source: own elaboration

Figure 8. Graph of confidence intervals and hypothesis testing between variances of actual and forecast values by period



Source: own elaboration

The next part of the research focused on analyzing the required teachers and classrooms based on the demand for students enrolled for period 1 of the year 2023. The purpose was to examine the average ratio of students per teacher, as well as the student per classroom ratio shown by the OECD (2021) and SEP (2019). The data collected is shown in table 3.

According to the SEP (2019), the student-teacher and student-classroom ratios are determined according to equations (26) and (27):

$$\text{Student-teacher ratio} = \frac{\text{Total enrollment}}{\text{Total teachers}} \tag{26}$$

$$\text{Student-class room ratio} = \frac{\text{Total enrollment}}{\text{Total classrooms}} \tag{27}$$

Based on this consideration, the calculations shown in Table 3 are made, where it is observed that the student-teacher and student-classroom ratios are well above the OECD average (2021), both internationally as in Mexico, which can impact the quality of education.

Table 3. Proportions of students per teacher and students per classroom

Year	Period	Actual enrollment/students	Teachers	Classrooms	Student-teacher ratio	Student-classroom ratio
2010	1	43	5	1	8.60	43.00
	2	38	5	1	7.60	38.00
2011	1	143	twenty	4	7.15	35.75
	2	266	26	6	10.23	44.33
2012	1	253	26	6	9.73	42.17
	2	743	29	17	25.62	43.71
2013	1	603	28	17	21.54	35.47
	2	771	28	17	27.54	45.35
2014	1	713	30	17	23.77	41.94
	2	856	31	19	27.61	45.05
2015	1	736	30	19	24.53	38.74
	2	772	30	19	25.73	40.63
2016	1	660	32	19	20.63	34.74
	2	1222	32	27	38.19	45.26
2017	1	1006	32	27	31.44	37.26
	2	1294	32	29	40.44	44.62
2018	1	1108	3. 4	29	32.59	38.21
	2	1114	3. 4	29	32.76	38.41
2019	1	1209	3. 4	29	35.56	41.69
	2	1088	26	22	41.85	49.45
2020	1	977	24	22	40.71	44.41
	2	1039	28	22	37.11	47.23
2021	1	844	25	23	33.76	36.70
	2	950	28	23	33.93	41.30
2022	1	877	25	23	35.08	38.13
	2	942	28	23	33.64	40.96
2023	1	885	24	23	36.88	38.48
	2	?	?	?	?	?

Ride Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Source: own elaboration

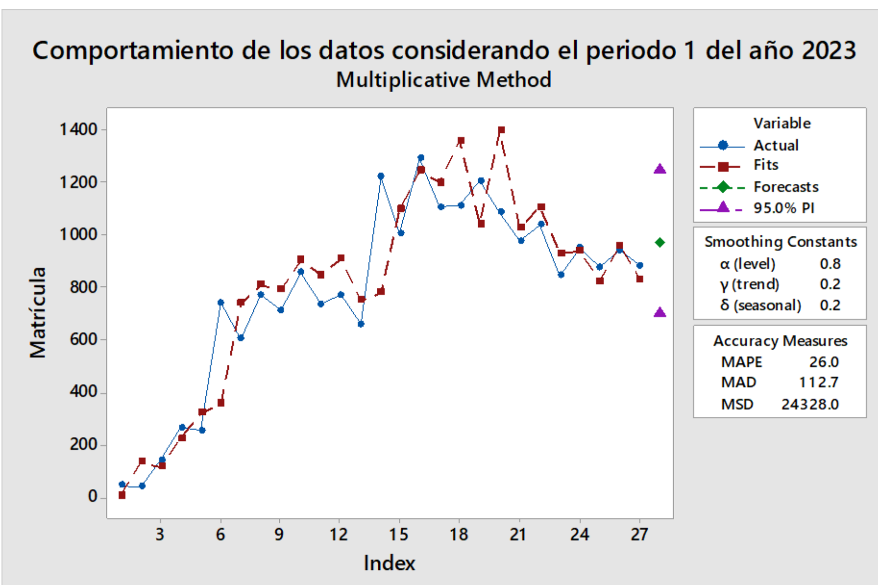
As indicated in the materials and methods section, the model used was validated by comparing the predicted results with the actual values. Initially, the forecast for period 1 of the year 2021 was calculated, and the corresponding actual value was collected. This last value was entered into the model to calculate the forecast for period 2 of the year 2021, and so on until period 1 of the year 2023. This process allowed forecasting period 2 of the year 2023, the results of which are presented in table 4. In Figure 9 shows the behavior of all the data considered, including the 5 validation periods of the Winters model.

Table 4. Model validation periods and results found.

Year	Period	Real data	Forecast	Mistake (%)
2021	1	844	919	8.88
	2	950	948	0.21
2022	1	877	817	6.84
	2	942	954	2.33
2023	1	855	826	3.39
	2	?	973	?

Source: own elaboration

Figure 9. Behavior of the data considered in the research



Source: own elaboration

According to the data presented in Table 4, it is observed that the model used to forecast enrollment has a relatively low forecast error in the forecasts by period. In this sense, it is important to highlight that the range to consider an acceptable forecast in this context, given that the data are not linear, is between 20% and 30%. The average demand forecast for period 2 of 2023 is 973, with a 95% confidence interval, which ranges between 697 and 1,249 students.

When considering the OECD average student-teacher ratio (2021), which is 13 students per teacher, it would be estimated that 75 teachers would be required to cover the expected enrollment. Likewise, taking into account the student-classroom ratio, 42 classrooms would be needed. However, if the current average proportions in Mexico are considered, the estimated number of teachers and classrooms required would be 35 teachers and 36 classrooms.

Discussion

The application of exponential smoothing models, specifically the Winters model, offers a reliable forecast of student demand that can be considered in similar situations of enrollment calculations in areas with homogeneous characteristics in Mexico. In this regard, there are many regions that share similarities in terms of econometric, social and demographic aspects, where these variables are not significant in demand, since these regions are already defined as having a similar economic, social and demographic level. This is especially relevant in areas such as rural regions and marginalized areas of cities, where it is crucial to prepare young people academically to improve their opportunities in the labor market and raise their socioeconomic conditions in the future.

On the other hand, and regarding the literature presented in the theoretical framework, the model used in this research has demonstrated its effectiveness in various branches of basic and applied science. In fact, it has been successfully used in solving problems related to sales demands, inventory control, new product launches and evaluation of new customers in internationally renowned companies.

Some notable applications include the work of González-Luna and Rodríguez- Morachis (2017), who used single and double exponential smoothing models to establish product demands, thereby they were able to reduce inventory by $408,658.43 US dollars over an eleven weeks period. Likewise, Nevárez- Carrazco *et al*. (2018) used simple exponential smoothing to reduce inventories in two products, achieving an assertiveness of 98 % and a reduction of one million US dollars in the 9 months validation period.

For their part, Serrato-Córdova and Rodríguez- Morachis (2014) applied simple exponential smoothing for inventory control, with which they achieved a certainty of 92.3% between the real value of demand and the predicted value. In another case, Fernández-Muñoz *et al*. (2022) used the double exponential smoothing model to address the problem of on-time product deliveries, achieving an increase from 50% to 80% in on-time delivery.

In addition to the research presented in the theoretical framework, in the educational sector there are studies such as that of Garro-Bordonaro and Arcos-Calzonci (2019), who carried out an analysis of student demand at a professional level in Mexico using a multiple regression model. In their results, they highlight that the relative labor income of workers does not significantly influence the demand for professional education, while social variables do have an impact. On the other hand, Slim *et al*. (2018) used logistic regression models and support vector machines to forecast demand at the professional level student at the University of New Mexico in the United States, with which they achieved an accuracy of 89.41% and 91.25%, respectively. Hernández-Pérez *et al*. (2020) also began their student demand analysis using the double exponential smoothing model, which resulted in the lowest forecast error for the data collected up to that year.

Now, when reviewing the literature, it is evident that in Mexico the application of this type of models in the public sector and, especially, in the educational sector, has not been sufficient to use it as a tool for planning school cycles. This points to an important opportunity in this area, since there are many regions with a homogeneous student market, where sociodemographic and economic analyzes would not be so relevant, since there are no marked differences in these aspects. In other words is, in these regions the crucial variable to consider would be student demand.

Conclusions

The choice of exponential smoothing models in this research is based on comparisons made in the specialized forecasting literature in scientific journals such as *International Journal of Forecasting* and *Journal of Forecasting.* These comparisons have shown that the accuracy of exponential smoothing models with appropriate characteristics is superior. In this sense, the best practice with these models involves making forecasts for one period at a time and calculating the forecast error in each one. In this research, the forecasts per period show that the error varied from 0.21% to 8.88%, which is considered low, especially when compared to the acceptable range of errors, which varies between 20% and 30%.

On the other hand, the forecast for period 2 of the year 2023 is 973 students, which suggests that, conservatively and maintaining the average student-teacher and student-classroom ratios in Mexico, the teacher workforce should be increased by 11 and the number of classrooms in 13. This is necessary so that, according to OECD standards, it is considered that quality education is being provided to students.

Finally, although the calculation of the student demand forecast was carried out using double exponential smoothing, the forecast errors in the periods turned out to be larger, so they were not presented in detail in this research.

Future lines of research

The use of forecasting models in the educational sector has not been widely applied, so its implementation in a more formal way is recommended. Currently, the official enrollment estimate is mainly based on political decisions, which can affect proper resource planning.

In situations where the potential market is homogeneous, the use of exponential smoothing models offers reliable results with a relatively small forecast error. Therefore, in future research its application is suggested in cases such as the calculation of failure rates, dropout rates, estimation of equipment resources in workshops, laboratories, software licenses, among others.

In this research, the behavior of student demand shows stability, which could imply, in future occasions, the consideration of a change in the forecast model to one that does not take into account neither trend nor seasonality, such as simple exponential smoothing or simple or double moving average.

Finally, the application of quantitative forecasting models, specifically exponential smoothing, is highly recommended for resource planning in cases similar to this research, where the student population shares similar social, economic and demographic characteristics.

References

Armstrong, G. and Kotler, P. (2013). *Marketing. An Introduction* (11th [ed] .). Pearson Education.

Brown, G. R. (1956). *Exponential Smoothing for Predicting Demand.* Arthur D. Little, Inc. https://www.industrydocuments.ucsf.edu/docs/jzlc0130

Chen, Q. (2022). A Comparative Study on the Forecast Models of the Enrollment Proportion of General Education and Vocational Education. *International Education Studies, 15* (6), 109-126.

Chopra, S. and Meindl, P. (2013). *Supply chain management: strategy and operation* (5. [a] ed). Pearson.

Cirelli, J., Konkol, A.M., Aqlan , F. and Nwokeji, J C. (2018). Predictive analysis models for student admission and enrollment. *Proceedings of the International Conference on Industrial Engineering and Operations Management.* September 27-29.

Clemen, R. (1989). Combining forecast: a review and annotated bibliography with discussion. *International Journal of Forecasting* . *5* (4), 559-608. DOI: https://doi.org/10.1016/0169-2070(89)90012-5

General Directorate of Planning, Programming and Educational Statistics (DGPPEE) of the Ministry of Public Education (SEP) (2022). *Chihuahua Educacional Statistics*. https://www.seech.gob.mx/estadistica/#

Fernández-Muñoz, CA, Rodríguez- Morachis, MA, Terrazas-Mata, LE and Zorrilla-Briones, F. (2022). Determination of anticipated customer demands with the use of forecasting models for planning the production of medical products. *IPSUMTEC* , *5* (5), 82-89.

Filders , R., Hibon , M., Makridakis , S., & Meade, N. (1998). Generalizing about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting.14* (3), 339-358. DOI: https://doi.org/10.1016/S0169-2070(98)00009-0

Garro-Bordonaro, NN and Arcos-Calzonci, A. (2019). The demand for higher education in Mexico 2016. *Denarius, Revista de Economía y Administración, 36* (2019), 49-83.

Geurts , M.D. and Kelly, J.P. (1986). Forecasting demand for special services. *International Journal of Forecasting, 2* (3), 261-272. https://doi.org/10.1016/0169-2070(86)90046-4

González-Luna, M. and Rodríguez- Morachis, MA (2017). Analysis and Selection of a Quantitative Forecasting Model for an Enterprise in the Electronic Sector. *International Journal of Advance Research in Science and Engineering* , *6* (5), 171-180.

Hernández-Pérez, H., Rodríguez- Morachis, MA, Rodríguez-Medina, MA and Tarango-Hernández, LE (2020). *School demand forecast to determine required capacity in upper secondary education schools*. International Research Congress Puebla 2020, pp. 447-452.

Hill, G. and Fildes, R. (1984). The accuracy of extrapolations methods: an automatic Box - Jenkins package SIFT. *Journal of Forecasting*, *3* , 319-323.

Holt, C. C. (2004). Forecasting Seasonal and Trend by Exponentially Weighted moving Averages. *International Journal of Forecasting*, *20* (2004), 5-10. DOI: doi:10.1016/j/ijforecast.2003.09.015

INEGI (2023a). *School enrollment by federal entity according to educational level, selected school cycles 2000/2001 to 2022/2023.* https://www.inegi.org.mx/app/tabulados/interactivos/?px=Educacion_06&bd=Educacion

INEGI (2023b). *Teachers and schools by federal entity according to educational level, selected school cycles 2000/2001 to 2022/2023.* https://www.inegi.org.mx/app/tabulados/interactivos/?pxq=8c29ddc6-eeca-4dcc-8def-6c3254029f19

James, F.E. and Weese, J.L. (2022). Neural Network-Based Forecasting of Student Enrollment with Exponential Smoothing Baseline and Performance Analysis. *ASSE 2022 Annual Conference. Excellence Through Diversity*. https://www.neural-network-based-forecasting-of-student-enrollment-with-exponential-smoothing-baseline-and-performance-analysis.pdf

Kleiman, A. (1975). Forecasting the Demand for Higher Education and the Resources Needed to Satisfy It. *XVI Ordinary General Assembly of the ANUIES*. https://blicaciones.anuies.mx/acervo/revsup/res017/txt1.htm

Koehler, A.B. and Murphree, E.S. (1988). A comparison of results from state space for casting with forecasts from the Makridakis competition. *International Journal of Forecasting*, *4* (1), 45-55. DOI: https://doi.org/10.1016/0169-2070(88)90009-X

Kotler, P.T., & Keller, K.L. (2011). *Marketing Management* (14th ed .). Pearson.

Lusk, E.J. and Neves, J.S. (1984). A comparative ARIMA analysis of the 111 series of Makridakis competition. *Journal of Forecasting. 3* (3), 329-332. https://doi.org/10.1002/for.3980030311

Makridakis , S., Andersen, A., Carbone, R., Fildes, R., Hibon , M., Lawandowski , R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*. 111-153.

https://doi.org/10.1002/for.3980010202

Makridakis , S. and Hibon , M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*. *16* (2000), 451-476. https://doi.org/10.1016/S0169-2070(00)00057-1

Miller, D., De Garay, A. and Montoya-Zepeda, I. (2021). Discount crossing. Superior demand and characteristics of applicants for admission to the Metropolitan University. *Mexican Journal of Educational Research, 26* (88), 253-282.

Nevárez- Carrazco , CU, Rodríguez- Morachis, MA, Zorrilla-Briones, F. and Esparza-Marín, A. (2018). *Reduction and control of the cost of printing supplies inventory with the use of forecasting models.* International Research Congress Chetumal 2018, pp. 1691-1697.

Organization for Economic Cooperation and Development (OECD) (2021). *Education Overview. OECD Indicators 2021*. https:// www.educacionyfp.gob.es/inee/dam/jcr:3922aacd-04c0-45ac-b8d4- 4aebb9b96ab5/panorama-2021-papel.pdf. p.p. 117-120

Olvera, A. (2013). The last five decades of the Mexican educational system. *Latin American Journal of Educational Studies*, *43* (3), 73-97.

Pacheco-Martínez, NJ (2021). Follow-up to the analysis of buildings for school use: their evaluation and impact on primary and secondary education. *Ibero-American Magazine for Educational Research and Development, 12* (23). https://doi.org/10.23913/ride.v12i23.1082

Ministry of Public Education (2019). *Guidelines for the Formulation of Educational Indicators* . https://www.planeacion.sep.gob.mx/Doc/estadistica_e_indicadores/lineamientos_formulacion_de_indicadores.pdf

Sepúlveda-Silvestre, GA (2020). *Forecast for the demand for enrollment in higher education studies at the undergraduate level at the University of La Amazonia*. Los Libertadores, University Foundation. https://repository.libertadores.edu.co/bitstream/handle/11371/3585/Sepulveda_Gener_2020.pdf?sequence=1&isAllowed=y

Serrato-Córdova, J. and Rodríguez- Morachis, MA (2014). *Analysis and application of forecasting techniques for inventory planning and control within a company in the electronic industry* , pp. 4533-4538. International Research Congress Celaya 2014.

Silitonga , PDP, Himawan , H. and Damanik , R. (2020). Forecasting Acceptance of New Students Using Double Exponential Smoothing Method. *Journal of Critical Reviews, 7* (1), 300-305.

Slim, A., Hush, D., Ojah , T. and Babbit, T. (July 16-20, 2018). *Predicting Student Enrollment Based on Student and Collage Characteristics.* Proceedings of the 11 th . International Conference on Educational Data Mining, pp. 383-389.

Tarango-Hernández, LE, Rodríguez- Morachis, MA, Fraustro, Y., Rojas, EJ and Lucero-Gaytán, M. (2019). Organizational Development in Improving Operations of a Language Center: Impact on Development of Students. In A. Realyvázquez , A. Maldonado-Macías and K. Arredondo (eds.), *Advanced Macroergonomics and Sociotechnical Approaches for Optimal Organizational Performance* (pp. 203-230). IGI Global. 10.4018/978-1-5225-7192-6.ch011.

Tenjo-Galarza, J. (2012). Demand for higher education: Projections until 2025. In YO Pérez-Alemán (ed.). *Collection, institutional analysis 2* (pp. 11-50). Pontifical Javeriana University. https://www.javeriana.edu.co/documents/15838/273636/SPEDES30112012.pdf/731df021-acd0- 4d39-9ed4- fc3704eaa6f0

Truckman, H. P. (1971). The Use of Predictive Models in Forecasting Student Choice. Paper presented at the Annual Meeting of the American Educational Research Association, February 1971. www. files.eric.ed.gov/fulltext/ED050169.pdf

Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science, 6* (3), 324-342. https://doi.org/10.1287/mns.6.3.324 .

Contribution Role	Author(s)
Conceptualization	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>
Methodology	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<supports>>
Software	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<supports>>
Validation	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>
Formal Analysis	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<supports>>
Investigation	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>
Resources	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>
Data curation	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<supports>>
Writing - Preparation of the original draft	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<supports>> Luz Elena Terrazas Mata <<supports>>
Writing - Review and editing	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>
Display	Manuel Alonso Rodríguez Morachis <<main>> Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>
Supervision	Manuel Alonso Rodríguez Morachis <<main>>
Project management	Manuel Alonso Rodríguez Morachis <<main>>
Fund acquisition	Manuel Alonso Rodríguez Morachis <<main>>

	Francisco Zorrilla Briones <<same>> Luz Elena Terrazas Mata <<same>>