

# Predicción de índices de reprobación escolar: un modelo de series de tiempo

Rosa-Irene Rojas-Rauda<sup>1</sup> y Omar-Jacobo Santos-Sánchez<sup>2</sup>

<sup>1</sup>Instituto Tecnológico de Pachuca,  
Pachuca, Hgo., México

<sup>2</sup>Universidad Autónoma del Estado de Hidalgo,  
Pachuca, Hgo., México

**Resumen** En este artículo se presenta un modelo de predicción basado en la técnica Box-Jenkins, para estimar el índice de reprobación escolar del programa de la Licenciatura en Administración del Instituto Tecnológico de Pachuca (ITP). El modelo empleado es una función polinomial en el tiempo cuyos coeficientes son hallados mediante la bien conocida técnica de mínimos cuadrados. El orden de la función polinomial es propuesto y reajustado de acuerdo a resultados parciales de predicción estimados mediante la técnica. Los errores finales de predicción obtenidos, muestran la factibilidad del método.

**Palabras clave:** modelo Box-Jenkins, serie de tiempo, predicción, índices de reprobación.

## 1. Introducción

La técnica de predicción Box-Jenkins ha sido utilizada en las áreas de ingeniería y económico administrativas para proponer modelos dinámicos complejos o difíciles de modelar. Ejemplo de ello es el uso de esta técnica en la predicción de la demanda escolar en México hecha por el ANUIES [1]. En [2] la técnica de predicción Box Jenkins es empleada en el área de la econometría para predecir volúmenes de producción agropecuaria, precios de estabilización de productos, etc. En [3] se propone un pronóstico de matrículas para asignaturas mediante la técnica Box-Jenkins. Además de propone un desarrollo de software para el análisis y presentación de resultados. Otro tipo de técnicas han sido empleadas para predecir índices de reprobación, tal como la propuesta expuesta en [4], en la cual se emplea inteligencia artificial para predecir índices de reprobación en la Universidad Tecnológica de León. En [5] utilizando técnicas probabilísticas y análisis de datos se predicen las trayectorias de los alumnos en la Universidad Veracruzana. En el presente trabajo, se estudia la factibilidad del uso de un modelo matemático descrito mediante una función polinomial [6] para la predicción de índices de reprobación escolar en las materias de un programa educativo específico. El orden del polinomio es propuesto de acuerdo a un valor de error preestablecido. Para ello se cuentan con los registros de los índices de reprobación

de 12 semestres de la totalidad de las materias impartidas en la Licenciatura en Administración del Instituto Tecnológico de Pachuca. Para establecer el orden de la función polinomial de predicción y el umbral de error, se toman en cuenta solo 11 semestres, y se predice el último dato (ya conocido). Una vez establecido el orden de la función polinomial tal que satisfaga el umbral de error, se predijo el semestre Agosto-Diciembre del 2012 (éste es el treceavo semestre, el cual se conoció hasta diciembre del 2012). Los resultados obtenidos muestran la factibilidad del método propuesto. Con ésta predicción la coordinación de la carrera puede tomar acciones de planeación como programación de horarios, salones, número de grupos y profesores, realizar acciones correctivas tales como programación de cursos de capacitación y actualización docente. Cabe señalar que en el presente trabajo, solo un modelo sencillo de predicción fué propuesto, éste puede ser mas complejo y usar un modelo que involucre entradas externas (modelos ARMAX) las cuales representarán profesores específicos, factores socioeconómicos de los alumnos, condiciones generales en la institución, etc. El orden del artículo es el siguiente: en el capítulo 2 se presentan los resultados preliminares para realizar la predicción, en el capítulo 3 se presentan el análisis de los resultados obtenidos y finalmente en el capítulo 4 se presentan los comentarios concluyentes y el trabajo a futuro.

## 2. Resultados preliminares

En este capítulo se recuerdan los principales resultados técnicos útiles para resolver el problema de predicción planteado. Si se considera al proceso de cursar una materia por parte de un grupo específico, como una función que depende exclusivamente del tiempo, entonces es posible obtener un modelo como una función polinomial de la siguiente forma:

$$s(t) = d_0 + d_1 t + \dots + d_n t^n, \quad (1)$$

donde los coeficientes  $d_i$ ,  $i = 0, \dots, n$  son desconocidos y deben calcularse,  $s(t)$  es el índice de reprobación a predecir. Este modelo es propuesto en [6]. El orden  $n$  es ajustado para satisfacer un umbral de error, el cual se calcula mediante una predicción parcial de resultados. Aunque este modelo es exclusivamente temporal, como se ve más adelante, es suficientemente útil como para predecir los índices de reprobación en forma aceptable. La ecuación (1) puede escribirse en forma matricial como:

$$s(t) = x^T(t)\Theta + e(t), \quad (2)$$

where  $x^T(t) = [1 \ t \dots t^n]$ ,  $\Theta = [d_0 \ d_1 \dots d_n]^T$  y  $e(t)$  representa el error de estimación. Ahora sean los tiempos específicos  $t_1, t_2, \dots, t_m$ , los cuales representan los tiempos específicos en los que se obtuvieron los índices de reprobación en semestres anteriores. Así se obtienen los siguientes vectores:

$$\mathbf{S}(t) = \mathbf{X}(t)\Theta + \mathbf{e}(t), \quad (3)$$

donde

$$\mathbf{S}(t) = \begin{bmatrix} s(t_1) \\ \vdots \\ s(t_m) \end{bmatrix}, \mathbf{X}(t) = \begin{bmatrix} x^T(t_1) \\ \vdots \\ x^T(t_m) \end{bmatrix}, \mathbf{e}(t) = \begin{bmatrix} e(t_1) \\ \vdots \\ e(t_m) \end{bmatrix} \quad (4)$$

Defina el siguiente índice de error:

$$J = \mathbf{e}^T(t)\mathbf{e}(t), \quad (5)$$

y de (3) se tiene que

$$\mathbf{e}(t) = \mathbf{S}(t) - \mathbf{X}(t)\Theta \quad (6)$$

por lo que

$$\begin{aligned} J &= (\mathbf{S}(t) - \mathbf{X}(t)\Theta)^T (\mathbf{S}(t) - \mathbf{X}(t)\Theta) \\ &= \mathbf{S}^T(t)\mathbf{S}(t) - \Theta^T \mathbf{X}^T(t)\mathbf{S}(t) - \mathbf{S}^T(t)\mathbf{X}(t)\Theta + \Theta^T \mathbf{X}^T(t)\mathbf{X}(t)\Theta \end{aligned}$$

Para minimizar  $J$  con respecto al vector  $\Theta$ , se deriva con respecto a  $\Theta$  y se iguala a cero (condición necesaria para la optimalidad [6]) se obtiene la siguiente solución para el vector de parámetros  $\Theta$ :

$$\Theta = (\mathbf{X}^T(t)\mathbf{X}(t))^{-1} \mathbf{X}^T(t)\mathbf{S}(t), \quad (7)$$

Si  $\mathbf{X}^T(t)\mathbf{X}(t)$  es definida positiva se asegura que la solución exista y que haga que  $J$  alcance un mínimo [6]. Esta solución es el bien conocido algoritmo de mínimos cuadrados [6], brevemente recordada aquí.

### 3. Resultados obtenidos y discusión

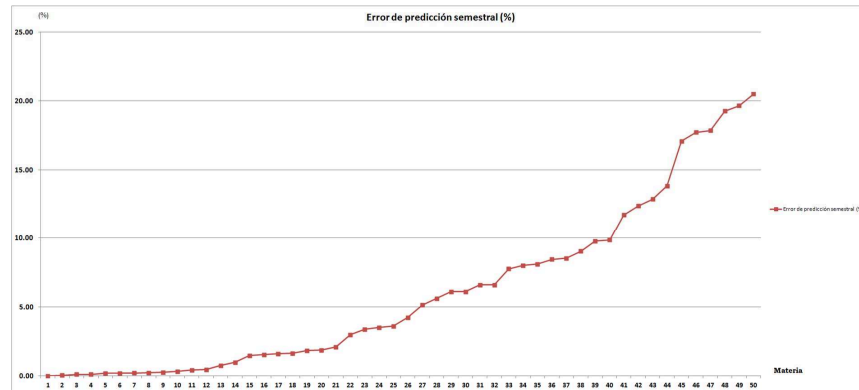
El modelo dado por (1) y la solución (7) se usan en este artículo para predecir el valor del índice de desempeño en el tiempo. Esto se hizo realizando un programa sencillo en Matlab (Math works) para calcular los coeficientes del modelo bajo estudio, proponiendo un polinomio de orden 5. El orden de dicho polinomio fué propuesto heurísticamente. Observe que un grado relativamente alto en el polinomio de predicción (1), hace que el ajuste a los índices de reprobación anteriores sea muy preciso, sin embargo, en general se observó que esto hace tener errores de predicción muy altos. Se estudiaron 50 materias del programa educativo de Licenciatura en Administración, del Instituto Tecnológico de Pachuca, Pachuca. Hidalgo. De éstas 50 materias, se tienen datos de índices de reprobación a partir del semestre Agosto-Diciembre del 2006, sin embargo, es necesario señalar que 5 materias (Sociología Organizacional, Derecho Laboral, Seguridad Social, Entorno Macroeconómico de las Organizaciones y Economía Empresarial) desaparecieron del mapa curricular debido a que en el año del 2010 se puso en marcha un nuevo plan de estudios. Con respecto a éstas materias se calculó el índice de reprobación usando hasta el penúltimo dato disponible y validándolo con el último dato disponible. Desde el año de 2006 a junio del 2012, 825 grupos fueron aperturados, correspondientes a las 50 materias bajo estudio.

En promedio se tuvieron grupos de 25 personas (existieron grupos de hasta 45 personas como máximo y de 1 persona como mínimo). El dato obtenido mediante la solución (1, 7) se validó con el índice de reprobación obtenido en el semestre Agosto-Diciembre del 2012 y los resultados se muestran la Figura 1, los errores están dados en porcentaje y ordenados en forma ascendente. El listado de materias bajo estudio es el siguiente

- |                        |                         |                                 |      |
|------------------------|-------------------------|---------------------------------|------|
| 1. Admon/Calidad       | 6. Derecho Mercantil    | 11. Sem.Admón.PM.ME.            |      |
| 2. Competitividad      | 7. Dis.Organizacional.  | 12. Admon/Suel/Sala.            |      |
| 3. Form/Eval/Proye.    | 8. Consultoría          | 13. Tall/Inf/Adm II             | (8)  |
| 4. Admon/Finan.II      | 9. Proc.Produc.         | 14. Tall/Inves. I               |      |
| 5. Admon/Estrateg.     | 10. Tall/ Inves. II     | 15. Comunica Corp.              |      |
| 16. Des./Sustentable.  | 21. Fund/Der.           | 26. Función Admva.II            |      |
| 17. Prob.Soc.Mex.      | 22. Ent/Macro/Org.      | 27. Mat.Aplic.Admón             |      |
| 18. Soc/Organ          | 23. Fundam.Econom       | 28. Taller Ética                |      |
| 19. Seg/Social.        | 24. Met/Cuant/Admon     | 29. Estad. para Admón.II        |      |
| 20. Inform/p/Admón     | 25. T.Gral.de Admón.    | 30. Tall.Desarr.Hum. Organ.M.A. | (9)  |
|                        | 31. Comport/Organ       | 36. Econ. Internac.             |      |
|                        | 32. Fund/Invest.        | 37. Costos Manuf.               |      |
|                        | 33. Sem/de Ética.       | 38. Neg.Gest.Contra.            | (10) |
|                        | 34. Dinámica social     | 39. Diag/Eval/Empre             |      |
|                        | 35. Antropol.Filos.     | 40. Función Admva.I             |      |
| 41. Fund/Mercadotec.   | 46. Der.Lab. Y Seg.Soc. |                                 |      |
| 42. Admón. Cap. Hum    | 47. Adm/Cap/Hum II      |                                 |      |
| 43. Conta.Gerencial.   | 48. Mezcla/Mdo.         |                                 | (11) |
| 44. Comun/Mercadología | 49. Sis/Inf/Mdo         |                                 |      |
| 45 Tall/Emprend.       | 50. Econ.Empresarial.   |                                 |      |

Como ya se mencionó durante 2010, la currícula de la carrera tuvo un rediseño, por lo que algunas materias cambiaron de nombre, sin embargo, en general el contenido temático se mantuvo, por lo que se consideraron los datos de la materia con el nombre anterior y el presente como una sola materia. En la lista dada se da el nombre actual de la materia. El menor error porcentual obtenido fué de 0 % y el mayor fué de 20.52 %. Esta diferencia, puede explicarse debido a que los índices de reprobación son afectados por cambios en los profesores que imparten las materias. En ese sentido, el error de predicción puede minimizarse aún más, mejorando el modelo. En efecto, un modelo que considere a los profesores como entradas externas puede mejorar los resultados obtenidos. adicionalmente pueden considerarse como perturbaciones factores socioeconómicos, malos hábitos de estudio, etc.

*Remark 1.* La correlación de los datos estaba bien condicionada debido a que en todos los casos la matriz  $(\mathbf{X}^T(t)\mathbf{X}(t))$  estuvo bien condicionada y su inversa siempre existió. Esto da evidencia de la identificabilidad del proceso bajo estudio.



**Figura 1.** Resultados de predicción obtenidos.

*Remark 2.* El modelo propuesto (1) considera las siguientes supocisiones:

1. La totalidad (o la mayoría) de los profesores evalúan a los alumnos apegados al reglamento escolar.
2. El índice de deserción escolar es relativamente bajo.

#### 4. Comentarios concluyentes

En el presente artículo se presentó un modelo polinomial temporal sencillo para la predicción de índices de reprobación de materias en una carrera específica. Los resultados de predicción obtenidos son útiles para una correcta planeación de requerimientos de espacios físicos y profesores. El modelo considera al fenómeno bajo estudio como una serie de tiempos. De acuerdo a los resultados obtenidos, el modelo de predicción brindó un buen desempeño. Dicho modelo puede mejorarse considerando un modelo ARX (AutoRegresive eXogenous) que considere dos tipos de factores externos: profesores y perturbaciones tales como factores socioeconómicos, entre otros, esto será objeto de trabajo a futuro.

## Referencias

1. Kleiman, A. La previsin de la demanda de eduacin superior y los recursos necesarios para satisfacerla, Presentado a la XVI Asamblea General Ordinaria de la ANUIES celebrada en marzo de 1975, México.
2. Rosales Álvarez, R. A., et al. Fundamentos de econometría intermedia: Teoría y aplicaciones, Apuntes de clase, ISSN 1909-4442, enero 2010, Universidad de los Andes, Venezuela.
3. González, G., M. L. Pronstico de matrícula de asignaturas sin prerequisites mediante modelos basados en series de tiempo, Tesis de licenciatura, junio 2005, Caracas, Venezuela.
4. Reyes, A., N. G. Análisis de datos para la predicción del índice de reprobación en el área de tecnologías de la información y comunicaciones de la UTL, CITIBB 2012, México.
5. Chaín, R. R. Alumnos y trayectorias. Procesos de análisis de información para diagnóstico y predicción. ANUIES (ED.), deserción, rezago y eficiencia terminal en las IES. Propuesta metodológica para su estudio. Serie Investigaciones. 2001, México: ANUIES.
6. Wellstead, P.E. and Zarrop, M.B. Self tuning systemas: Control and signal processing, Jhon Wiley and Sons, 1991, England.