

# Estimación por Intervalos de Probabilidad a Posteriori para la Proporción de Estudiantes Universitarios Desertores

Carlos J. Barrera-Causil<sup>1</sup>  
Juan de J. Sandoval<sup>2</sup>  
Fabio H. Sepúlveda-Murillo<sup>3</sup>

## Resumen

En este artículo se presenta la estimación por intervalos de probabilidad a posteriori de la proporción de estudiantes que desertan del Instituto Tecnológico Metropolitano (ITM) según algunos factores influyentes. Nuestro interés es determinar la proporción de estudiantes que desertan de dicha institución según los factores académico, laboral y personal. Se usó metodología bayesiana, implementando procesos de elicitación y método Delphi. Dentro de los resultados, se evidenció que, con una probabilidad del 95% se afirma que el factor laboral tiene un porcentaje a posteriori de incidencia en la deserción del ITM entre el 30,4% y 44,6%. Además, se notó que el factor laboral y el académico son los más influyentes en la deserción de la institución.

## Palabras clave

Intervalos de probabilidad, estimación, distribución, Bayesiana.

---

1 Facultad de Ciencias, Instituto Tecnológico Metropolitano, Medellín-Colombia, carlosbarrera@itm.edu.co

2 Facultad de Ciencias, Instituto Tecnológico Metropolitano, Medellín-Colombia, juansandoval@itm.edu.co

3 Facultad de Ciencias, Instituto Tecnológico Metropolitano, Medellín-Colombia, fabiosepulveda@itm.edu.co

**Abstract**

This paper presents the estimation by posterior probability intervals of the proportion of students who drop out of Instituto Tecnológico Metropolitano (ITM) considering some influential factors. Our interest is to determine the proportion of students who drop out of the institution by academic factor, occupational and personal factor. To implement the Bayesian methodology we use the elicitation technique and the Delphi method. Within the results, showed that, with a probability of 95% we can say that the occupational factor have a posterior percentage of incidence on drop out of ITM between 30,4% and 44,6%. Also, notice that the occupational factor and the academic factor are the most influential in the drop out of the institution.

**Keywords**

Probability intervals, estimation, distribution, Bayesian.

## 1. INTRODUCCIÓN

El problema de deserción estudiantil afecta a todas las instituciones de educación superior y en los estudios recientes se nota que la problemática va en aumento (Lopera-Oquendo, 2008). El estudio de las posibles causas de deserción estudiantil se ha realizado por parte de Ricoachury & Lara (1984), Tinto (1989), Hanushek (2004) y Sarmiento-Gómez (2006) entre otros.

En el Instituto Tecnológico Metropolitano (ITM) el problema de deserción es de mucho interés y por esto se deben implementar metodologías que permitan garantizar el aumento de la retención estudiantil manteniendo siempre en alto la calidad académica. Si se logra determinar la proporción de estudiantes que desertan por cada uno de los diferentes factores que más influyen en esta problemática, entonces, se podrán ejecutar procedimientos que logren mitigar el efecto de dichos factores en la deserción y así disminuir considerablemente la deserción estudiantil, es por esto que la estimación de dichos parámetros es de mucha importancia. En este artículo se pretende estimar a través de intervalos de probabilidad a posteriori la proporción de estudiantes que desertan por cada uno de los factores de riesgo más frecuentes, detectando así el factor más influyente para la deserción del ITM. Se hicieron uso de técnicas de elicitación (La palabra elicitación es un anglicismo que se definirá posteriormente) y del método Delphi para obtener las distribuciones a priori de interés.

Según Hill (2002) y Meyer & Booker (2002) la estadística bayesiana permite incorporar información no muestral (subjetiva) sobre características desconocidas del fenómeno bajo estudio con el fin de realizar dichas estimaciones. La cuantificación de la información subjetiva se realiza mediante un proceso conocido como elicitación. La cuantificación de las creencias del experto se ha utilizado en el desarrollo de los sistemas expertos probabilísticos para diagnósticos de enfermedades cardíacas congénitas (Spiegelhalter et al., 1994), en la proyección de la población (Kadane & Wolfson, 1998), etc. La elicitación de la creencia apriori también se ha utilizado con éxito al estimar los costos de mantenimiento futuros de plantas de tratamiento de aguas, en la determinación de la conductividad hidráulica de las rocas para el desarrollo del

depósito de los desechos nucleares (O'Hagan, 1997), y en el análisis de la incertidumbre para la protección radiológica (O'Hagan & Haylock, 1997), entre otros.

### 1.1 Distribución A Priori y A Posteriori

El conocimiento previo acerca de los parámetros de una distribución de intereses expresado en una distribución de probabilidad, la cual se conoce como distribución apriori. La determinación de la distribución apriori es un problema fundamental en la estadística bayesiana. En muchos casos, se hace necesario utilizar procesos de elicitación para obtener estas distribuciones.

Los métodos bayesianos permiten incorporar opiniones subjetivas acerca de incertidumbres con respecto al parámetro o vector de parámetros de interés. La incertidumbre acerca del verdadero valor de un parámetro de interés  $\theta$  en la población, es modelado por la función de densidad apriori  $\xi(\theta)$ , ( $\theta \in \Theta$ ). La distribución a posteriori, la cual representa una actualización de la información a priori usando los datos muestrales, se calcula mediante el teorema de Bayes como:

$$\xi(\theta | \text{Datos}) \propto \xi(\theta)L(\theta | \text{Datos}) \quad (1)$$

donde  $L(\theta | \text{Datos})$  es la función de verosimilitud, (Hill, 2002).

### 1.2 Probabilidad Subjetiva y Técnica de Elicitación

La probabilidad subjetiva es definida en términos generales como una medida de la confianza que se tiene acerca de una proposición en particular (Savage, 1971). Existen dos aproximaciones a la probabilidad subjetiva. La primera es la de Savage (1971), la cual se basa en la maximización de una función de utilidad, de tal manera que las decisiones tomadas revelen un conjunto consistente de probabilidades. La segunda aproximación llamada intuitiva, fue desarrollada por Good (1962). Esta aproximación usa el concepto de relaciones ordenadas.

Adams (2005), define la elicitación como el proceso de capturar el conocimiento de una persona y la creencia acerca de una o más

cantidades desconocidas dentro de una distribución de probabilidad. Un problema adicional y muy complejo al que se enfrenta el investigador es cuantificar la información obtenida por el experto. Es importante distinguir entre la calidad de un conocimiento de expertos y la precisión con la cual el conocimiento es trasladado a forma probabilística. Una elicitación es considerada como buena, si la distribución que es obtenida con precisión, representa de manera adecuada el conocimiento de los expertos (Garthwaite et al., 2005).

El proceso de elicitación lo divide Garthwaite et al. (2005) en cuatro etapas. La primera etapa consiste en la preparación para la elicitación seleccionando al experto, entrenándolo, identificando los aspectos del problema a elicitar y de esa manera comenzar. En la segunda etapa se especifica el resumen del experto y la distribución para esos aspectos. La tercera es el ajuste de la distribución de probabilidad a esos resúmenes. Por último, se involucra la evaluación adecuada de la elicitación, con la opción de volver a la segunda etapa y continuar el proceso, puesto que el proceso se puede volver iterativo.

### 1.3 Método Delphi

El método Delphi se define como un método de estructuración de un proceso de comunicación grupal que es efectivo a la hora de extraer información de manera efectiva a un grupo de individuos (Listone & Turoff, 2002). Una Delphi consiste en la selección de un grupo de expertos a los que se les pregunta su opinión sobre cuestiones referidas a acontecimientos del futuro. Las estimaciones de los expertos se realizan en sucesivas rondas anónimas con el propósito de tratar de conseguir consenso entre los participantes, pero con la máxima autonomía por parte de los participantes para evitar sesgos.

Las preguntas se refieren, por ejemplo, a las probabilidades de realización de hipótesis o de acontecimientos con relación al tema de estudio, en este caso se cuestiona a los expertos acerca de la deserción de los estudiantes, dependiendo de posibles factores causales. La calidad de los resultados depende, sobre todo, del

cuidado que se ponga en la elaboración del cuestionario y en la elección de los expertos consultados (Listone & Turoff, 2002).

## 2. METODOLOGÍA

Se realizó una entrevista estructurada a dos expertos (participantes) en el área de bienestar universitario del ITM, los cuales poseían gran experiencia acerca de la deserción universitaria, específicamente deserción en el ITM. En el proceso de elicitación usando el método Delphi los expertos definieron los tres factores más influyentes en el problema de deserción en el ITM, los cuales se describen a continuación.

*Factor 1 (factor académico):* bajo rendimiento, deficiencias en formación académica, dificultades cognitivas, dificultades con el docente, orientación vocacional, etc. *Factor 2 (factor laboral):* cambio de turno del empleador, desempleo, incompatibilidad horaria, nuevo trabajo, etc. *Factor 3 (factor personal):* problemas familiares, baja autoestima, motivación, problemas de salud, problemas afectivos, etc.

Posteriormente, los expertos proporcionaron una distribución a priori a mano alzada para cada uno de los factores. Este procedimiento se realizó de manera separada para evitar sesgo. Cabe notar que para la construcción de la distribución a priori a mano alzada el investigador realizó una serie de preguntas al experto con el fin de validar la distribución obtenida. Luego de chequear que las a priori representen el verdadero conocimiento acerca del tema por parte de los participantes, se procedió a iniciar el método Delphi con el fin de obtener por consenso una distribución a priori a mano alzada para cada factor en estudio.

Luego de confrontar la información suministrada usando el método Delphi con las distribuciones a mano alzada se procedió a generar 1000 muestras de la distribución a priori para cada uno de los factores de riesgo mencionados. Las muestras obtenidas para cada factor se obtuvieron usando el método de aceptación-rechazo, (Ripley, 1987). Este proceso se describe a continuación.

Supóngase que  $g_i(x)$  es la distribución a priori elicitada a mano alzada obtenida por consenso para el factor  $i = 1,2,3$ . Esta

distribución es discretizada con el fin de calcular su media, llámese  $\bar{X}_{g_i}$ . Con el fin de generar las muestras, se encontró para cada caso (factor), una distribución “cobija” (es decir, esta distribución debe cubrir en todos sus puntos y ser la más cercana a la a priori elicitada usando el método Delphi, para el factor de interés), las cuales resultaron ser normales truncadas entre el mínimo y el máximo fijado por los expertos en consenso, con media  $\bar{X}_{g_i}$  y una varianza suficiente mente grande. Posteriormente, en cada caso, se generó un valor  $x_c$  de la distribución “cobija”, luego se evaluó  $x_c$  en la ecuación de la recta construida por los dos puntos (es decir, puntos resultantes del proceso de discretización), que contengan y sean más cercanos a  $x_c$  para calcular la altura y en dicho punto de la curva a priori elicitada con el método Delphi. Luego, se generó un valor  $u_c$  de una distribución uniforme que se encuentre entre 0 y la altura correspondiente a  $x_c$  evaluado en la distribución “cobija”. A partir de este valor se aceptó el valor  $x_c$  si se cumplía la condición de que  $u_c$  sea menor que  $y$ . Este procedimiento fue repetido hasta encontrar los primeros 1000 valores que cumplieran dicha condición.

Por otro lado, de una población objeto de estudio constituida por los estudiantes que desertaron el primer semestre de 2009 en el ITM, para la ejecución de nuestro estudio, adicionalmente al proceso de elicitación, se tomaron datos reales de la información correspondiente al número de estudiantes que desertaron por los diferentes factores causales, mediante muestreo aleatorio sistemático. El tamaño de muestra ajustado debido a errores de muestreo, y teniendo en cuenta que en el ITM desertaron aproximadamente 3.000 estudiantes el primer semestre de 2009 es de 400 estudiantes, ajustado por encima para poblaciones no normales. Para la estimación por intervalos de la proporción a posteriori de estudiantes que desertan según estos tres factores, se mezcla la información muestral con la información obtenida en el proceso de elicitación. El análisis de toda la información se realizó con el software estadístico R, versión 2.11.1.

### 3. RESULTADOS Y DISCUSIÓN

Es importante notar que los expertos elicitados tenían conocimiento antes de iniciar el estudio acerca de resultados obtenidos en una encuesta de deserción realizada en el año 2008 dentro de la institución en mención, por tal motivo, los expertos tenían como valor de referencia debido al estudio del 2008, que el porcentaje de incidencia del factor académico en la deserción de los estudiantes del ITM era de un 47,81%, para el factor laboral se tenía un 25,90%, y para el factor personal el porcentaje de incidencia fue de 16,41%.

La encuesta elaborada en el presente trabajo a los 400 estudiantes arrojó un porcentaje de incidencia en la deserción de los estudiantes del ITM por parte del factor académico de 32,3%, el factor laboral incidió en un 42,2% y el personal en 14,4%. Los porcentajes promedio a priori estimados para cada uno de los factores de riesgo asociados a la deserción en el ITM luego de utilizar el método Delphi y el muestreo de aceptación-rechazo se muestran en la Tabla 1.

Tabla 1. Resumen estadístico a priori, porcentajes incidencia de la deserción.  
Fuente: Autores

	Factores		
	Académico	Laboral	Personal
Máximo porcentaje posible	60,00%	30,00%	25,00%
Mínimo porcentaje posible	30,00%	10,00%	5,00%
Porcentaje promedio	43,86%	18,39%	14,69%
Desviación estándar	6,69%	4,93%	5,21%

Se puede notar que los porcentajes a priori promedios de deserción por factores, tienen un comportamiento similar a los valores de referencia dados en el estudio de deserción del 2008. Debido al comportamiento funcional de las 1000 muestras generadas para cada factor, y como éstas no provienen de una función con forma distribucional conocida, se tomaron (2), (3) y (4) como distribuciones a priori para el factor académico, laboral y personal, respectivamente.

$$\xi(p) \propto \frac{1}{\sqrt{2\pi}(0,0669)} e^{-\frac{1}{2}\left(\frac{p-0,4386}{0,0669}\right)^2}; \quad \text{para } 0 \leq p \leq 1 \quad (2)$$

$$\xi(p) \propto \frac{1}{\sqrt{2\pi}(0,0493)} e^{-\frac{1}{2}\left(\frac{p-0,1839}{0,0493}\right)^2}; \quad \text{para } 0 \leq p \leq 1 \quad (3)$$

$$\xi(p) \propto \frac{1}{\sqrt{2\pi}(0,0521)} e^{-\frac{1}{2}\left(\frac{p-0,1469}{0,0521}\right)^2}; \quad \text{para } 0 \leq p \leq 1 \quad (4)$$

Se puede notar que éstas son normales truncadas en (0,1) y tienen en cuenta la media y la varianza obtenida en el proceso de muestreo aceptación-rechazo. Las distribuciones de verosimilitud para el factor académico, laboral y personal se muestran en (5), (6) y (7), respectivamente.

$$L(p | \text{Datos}) = \binom{400}{129} p^{129}(1-p)^{271} \quad (5)$$

$$L(p | \text{Datos}) = \binom{400}{169} p^{169}(1-p)^{231} \quad (6)$$

$$L(p | \text{Datos}) = \binom{400}{58} p^{58}(1-p)^{342} \quad (7)$$

Ahora, de (1) se tiene que, las distribuciones a posteriori para porcentaje de incidencia del factor académico, laboral y personal están dadas por (8), (9) y (10), respectivamente.

$$\xi(p | \text{Datos}) \propto e^{-\frac{1}{2}\left(\frac{p-0,4386}{0,0669}\right)^2} \times p^{129}(1-p)^{271}; \text{ para } 0 \leq p \leq 1 \quad (8)$$

$$\xi(p | \text{Datos}) \propto e^{-\frac{1}{2}\left(\frac{p-0,1839}{0,0493}\right)^2} \times p^{169}(1-p)^{231}; \text{ para } 0 \leq p \leq 1 \quad (9)$$

$$\xi(p | \text{Datos}) \propto e^{-\frac{1}{2}\left(\frac{p-0,1469}{0,0521}\right)^2} \times p^{58}(1-p)^{342}; \text{ para } 0 \leq p \leq 1 \quad (10)$$

Las distribuciones a posteriori se muestran en la Fig.1. Se nota que la proporción de estudiantes que desertan del ITM debido al

factor laboral es mayor que el porcentaje de estudiantes que desertan debido al factor académico, siendo esto muy importante puesto que inicialmente con la información a priori se pensaba lo contrario.

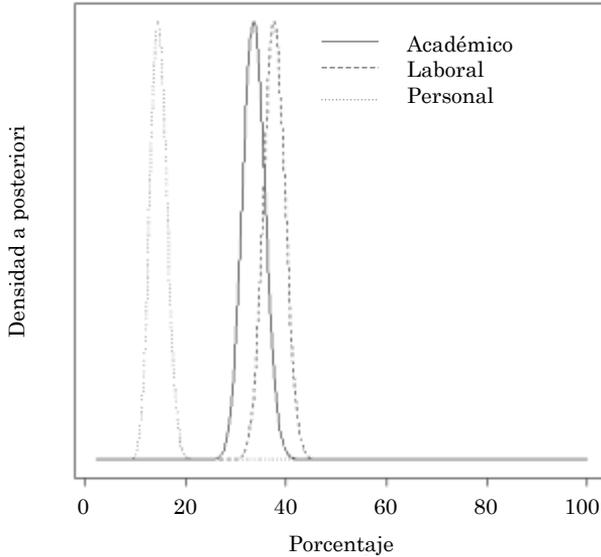


Fig. 1. Distribuciones a posteriori del porcentaje de estudiantes que desertan según factores. Fuente: Autores

En la Tabla 2 se presentan algunas medidas descriptivas para cada factor y los respectivos intervalos de probabilidad correspondientes a la distribución a posteriori.

Tabla 1. Resumen estadístico para las distribuciones a posteriori. Fuente: Autores

	Factores		
	Académico	Laboral	Personal
Porcentaje promedio	30,00%	37,50%	14,50%
Desviación estándar	5,78%	4,34%	3,18%
Intervalo de probabilidad al 95%	20,5%-39,5%	30,4%-44,6%	9,3%-19,7%

Con una probabilidad del 95% se puede decir que el factor académico tiene un porcentaje a posteriori de incidencia en la deserción de los estudiantes del ITM entre 20,5% y 39,5%. Además, con la misma probabilidad se dice que el porcentaje a posteriori de incidencia del factor laboral en la deserción del ITM se encuentra entre 30,4% y 44,6%. Por último, se nota que el factor personal incide en la deserción de los estudiantes del ITM con un porcentaje a priori entre el 9,3% y 19,7% con un nivel de credibilidad del 95%.

#### 4. CONCLUSIONES

Los resultados muestran que es posible construir intervalos de probabilidad para estimar el porcentaje de incidencia a posteriori de los factores más influyentes en la deserción del ITM. Con estos intervalos se notó que el factor que tiende a ser más influyente en la deserción de los estudiantes del ITM es el factor laboral, caso contrario a lo que evidenciaba la información del estudio previo.

Esto va en sentido inverso a lo que se pensaba respecto a la deserción, que ésta estaba más influenciada por factores académicos tales como el bajo rendimiento o la no elección de la carrera adecuada. Se concluye que esta no influye notablemente en las decisiones de deserción por parte de los estudiantes. Adicionalmente, factores personales como enfermedad o problemas personales entre estudiantes y docentes poco influyen para tal decisión.

Se llega a la conclusión que el factor laboral, como es, la búsqueda incesante de oportunidades de empleo mientras se estudia, es un factor altamente responsable a la hora de tomar una decisión de dejar la universidad. En otras palabras, el estudiante, bajo una grata oferta de empleo prefiere el empleo a seguir estudiando en la mayoría de los casos en el ITM. Desde el sentido metodológico, es recomendable realizar el método Delphi a un grupo más grande de expertos para realizar las estimaciones.

## 5. REFERENCIAS

- Adams, F., (2005); Expert elicitation and Bayesian analysis of construction contract risks: an investigation, *Construction Management and Economics*, 24: 81–96.
- Garthwaite, P., Kadane, J., & O'Hagan, A., (2005); Statistical Methods for Eliciting Probability Distributions, *Journal of the American Statistical Association*, 100(470): 680-701.
- Good, I., (1962); Subjective Probability as the Measure of a Non-Measurable Set, En E. Nagel, P. Suppes, & A. Tarski, *Logic, Methodology and Philosophy of Science* (pp. 319-29), Stanford University Press.
- Hanushek, E., (2004); *Economic Analysis of School Quality, Education for All Global Monitoring*.
- Hill, G., (2002); *Bayesian Methods*, Chapman and Hall.
- Kadane, J., & Wolfson, L., (1998); Experiences in Elicitation (with discussion), *The Statistician*, 47, 1–20.
- Listone, H., & Turoff, M., (2002); *The Delphi method, techniques and applications*. California: Addison Wesley publishing.
- Lopera-Oquendo, C., (2008); Determinantes de la Deserción Universitaria en la Facultad de Economía Universidad del Rosario, *Economía* (95), 1-25.
- Meyer, M., & Booker, J., (2002); *Eliciting and Analyzing Expert Judgment*, Chapman and Hall.
- O'Hagan, A., (1997); The ABLE story: Bayesian asset management in the water industry, En S. French, & Q. Smith, *The Practice of Bayesian Analysis* (173-198), United Kingdom: Arnold.
- O'Hagan, A., & Haylock, R., (1997); Bayesian uncertainty analysis and radiological protection, En V. Barnett, & K. Turkman, *Statistics for the Environment, Pollution Assessment and Control* (3 ed., pp. 109-128), New York: Wiley, Chichester.
- Ricoachury, H., & Lara, L., (1984); Los Desertores, En U. P. Nacional, *Cuadernos de Planeación*, Bogotá.
- Ripley, B., (1987); *Stochastic Simulation*, New York: John Wiley & Sons.

- Sarmiento-Gómez, A., (2006); Una Estrategia para Aumentar la Retención de los Estudiantes, Departamento Nacional de Planeación, Bogotá: Ministerio de Educación Nacional.
- Savage, L., (1971); Elicitation of Personal Probabilities and Expectations, *Journal of the American Statistical Association*, 66(336), 783-801.
- Spiegelhalter, D., Freedman, L., & Parmar, M., (1994); Bayesian approaches to randomized trials (with discussion), *J. R. Statist. Soc. A*, (157), 357-416.
- Tinto, V., (1989); Definir la deserción: una cuestión de perspectiva, *Revista de Educación Superior*, XVIII (3) (71).

