

## REALIZACIÓN DE “EXPERIMENTOS VIRTUALES” DE INTERÉS BIOTECNOLÓGICO PARA LA APLICACIÓN CONTEXTUALIZADA DE PROBABILIDAD Y MÉTODOS ESTADÍSTICOS

Marta Ginovart – Mònica Blanco  
[marta.ginovart@upc.edu](mailto:marta.ginovart@upc.edu) – [monica.blanco@upc.edu](mailto:monica.blanco@upc.edu)

Universitat Politècnica de Catalunya, Barcelona, Spain

Núcleo temático: VI. Matemáticas y su integración con otras áreas

Modalidad: CB

Nivel educativo: No específico

Palabras clave: Modelización, Simulación, Estadística, “Experimento virtual”

### Resumen

*La simulación por ordenador permite generar “experimentos virtuales”, representando una alternativa efectiva a la experimentación real, inviable algunas veces en los laboratorios de los centros educativos. La simulación estocástica de sistemas compuestos por agentes autónomos y discretos (o por individuos) utiliza variables aleatorias, y se basa en la generación y uso de valores obtenidos bajo determinadas distribuciones de probabilidad para simular fenómenos reales. Esta comunicación se centra en la presentación de una serie de actividades que se realizaron en la asignatura de Estadística en los cuatro grados de Ingeniería de Biosistemas y Agroalimentaria de la Universitat Politècnica de Catalunya durante el primer cuatrimestre del curso académico 2016-2017. Los “experimentos virtuales” se realizaron con INDISIM-YEAST\_NL, un simulador implementado en la plataforma de acceso libre NetLogo para investigar las dinámicas de poblaciones de levaduras creciendo en distintas condiciones. Los datos de los experimentos generados con este simulador se analizaron estadísticamente a nivel descriptivo e inferencial en sesiones de laboratorio informático, complementadas después con un trabajo fuera del aula. Resultaron actividades de enseñanza-aprendizaje exitosas, integradas en el contexto de la Ingeniería de Biosistemas y Agroalimentaria, que combinaron conceptos probabilísticos y métodos estadísticos con procedimientos del ámbito de la matemática aplicada.*

### Introducción

Son diversas, y cada vez más numerosas, las propuestas centradas en la evaluación de alternativas para mejorar la calidad de la enseñanza de la estadística en las ciencias biológicas, de la salud, experimentales, tecnológicas, o sociales. Mayoritariamente estos estudios coinciden en la conveniencia de focalizar este proceso de enseñanza-aprendizaje, principalmente, en la resolución de problemas reales y en el desarrollo de proyectos

contextualizados (por ejemplo, Batanero et al. 2013, Serradó, 2013 y Walz, 2015, entre otros). La estadística aplicada es una ciencia que, entre otras posibilidades, puede ofrecer a otras ciencias un conjunto de instrumentos con los que resumir y comprender datos. Requiere de un tipo diferente de pensamiento del que se entrena muchas veces en asignaturas de matemáticas, ya que se trabaja con datos en un contexto. Si se define a la estadística como la ciencia de los datos (en el sentido general del término) es entonces crucial entrenar a los estudiantes en el razonamiento estadístico a partir de datos empíricos, insistiendo en la importancia que tiene el contexto para el análisis de estos datos. La estadística que se enseña y se practica en estudios posteriores a una enseñanza obligatoria necesita ser adaptada y orientada al entorno curricular en el cual se encuentra inmersa, debe ser insertada convenientemente en el plan de estudios del cual participa, con la finalidad de garantizar que pueda ser utilizada de forma inmediata para resolver situaciones propias del ámbito profesional que corresponda.

La simulación por ordenador permite generar “experimentos virtuales”, representando una alternativa efectiva a la experimentación real, inviable algunas veces en los laboratorios de los centros educativos. El uso de modelos computacionales en un entorno de ejecución amigable hace viable una “experimentación” controlada, generando “experimentos virtuales” que ilustran sobre el problema a estudiar, siendo una herramienta efectiva de entrenamiento para avanzar en la comprensión de conceptos y fenómenos. La simulación estocástica de sistemas compuestos por agentes (o individuos) autónomos y discretos utiliza variables aleatorias, tanto discretas como continuas, y se caracteriza por basarse en la generación y uso de valores obtenidos bajo determinadas distribuciones de probabilidad para simular procesos reales. La opción de realizar “experimentos” en un laboratorio virtual resulta conveniente y atractiva, ya que el estudiante puede emular la obtención de datos reales en un contexto relevante para él, y abordar su análisis durante la docencia de una asignatura de estadística.

El objetivo de este trabajo es presentar el diseño, y la implementación en el aula, de una serie de actividades ligadas a la generación de “experimentos virtuales” de un sistema biológico relevante para los estudiantes en el contexto de sus estudios, para proceder al análisis de los datos obtenidos con estas simulaciones, de forma que los estudiantes tengan la oportunidad de usar conceptos y métodos estadísticos presentados y practicados en una asignatura de estadística. El ordenador será utilizado como recurso tanto para la generación de datos

contextualizados en un escenario biotecnológico de interés para estos estudiantes como para el análisis estadístico de estos datos.

### **Material y métodos**

Los participantes en este estudio fueron un grupo de 140 estudiantes de segundo año de la Escuela Superior de Agricultura de Barcelona de la Universitat Politècnica de Catalunya (UPC) de cuatro grados del ámbito de la Ingeniería de Biosistemas y Agroalimentaria. Estos estudiantes cursaron en el primer año asignaturas obligatorias de las áreas de matemáticas, física, química y biología. En particular, la asignatura Matemáticas II (segundo cuatrimestre) se centró en el cálculo integral, la resolución analítica y numérica de ecuaciones diferenciales ordinarias básicas, y en el reconocimiento de algunos modelos computacionales (Ginovart 2014). Esta preparación previa garantizó el conocimiento de algunos sistemas biológicos y de un conjunto de herramientas e ideas matemáticas básicas para abordar la modelización, que fueron complementadas en el decurso de la asignatura obligatoria de Estadística que se impartió en el primer cuatrimestre del segundo año, donde se desarrollaron las actividades objeto de este estudio. Desde el principio de esta asignatura, los estudiantes tuvieron la oportunidad de practicar con la descripción numérica y gráfica de conjuntos de datos, y con diversos métodos básicos de inferencia estadística. Las actividades diseñadas y presentadas en este estudio se implementaron durante las dos últimas sesiones de laboratorio informático (de dos horas de duración cada una), y formaron parte de un trabajo complementario realizado fuera del aula y entregado posteriormente para su evaluación. Así pues, los estudiantes realizaron todas estas actividades durante las tres últimas semanas del primer cuatrimestre del curso académico 2016/2017 en la asignatura de Estadística.

La colección de tareas secuenciales, agrupadas en distintos módulos, para guiar la descripción de estos datos simulados obtenidos por el propio estudiante con un modelo computacional se proporcionó y gestionó a través del campus virtual Atenea, la plataforma de soporte a la docencia de la UPC. Cada estudiante tuvo un ordenador con acceso a una hoja de cálculo (Excel), a dos programas estadísticos utilizados previamente en sesiones de laboratorio en la asignatura de Estadística (Minitab y R), y a una plataforma de programación, modelización y simulación presentada en la asignatura de Matemáticas II llamada NetLogo. NetLogo es una plataforma desarrollada y mantenida por el Northwestern's Center for Connected Learning and Computer-Based Modeling (EEUU), libre y accesible desde la web

(<http://ccl.northwestern.edu/netlogo/>), siendo su instalación inmediata y compatible con todos los sistemas operativos desde su propia web.

Las actividades fueron diseñadas para ser llevadas a cabo de forma individual en la primera sesión de laboratorio, y por parejas en la segunda sesión y en la elaboración del trabajo complementario. Las respuestas de los estudiantes con respecto a los análisis de los datos simulados con el modelo computacional fueron elaboradas y recogidas a través de documentos tipo texto dando respuestas a cuestiones formuladas previamente.

Considerando que esta asignatura de Estadística se encuentra en los planes de estudio de grados del ámbito de la Ingeniería de Biosistemas y Agroalimentaria, para la generación de datos a analizar se buscó un modelo computacional afín a esta área de conocimiento, centrado en levaduras, microorganismos muy importantes desde el punto de vista biotecnológico. Este tipo de microorganismos se ha utilizado extensamente a lo largo de la historia. Por ejemplo, *Saccharomyces cerevisiae*, conocida como levadura de cerveza o levadura del pan debido a su empleo como fermento, es una de las levaduras con mayor impacto económico y social, ya que es usada en las industrias agroalimentarias y en la producción de productos químicos, así como en la producción de etanol. Esta levadura también posee prácticamente todas las características que la configuran como un sistema modelo en el mundo microbiano. Por esto se consideró interesante realizar distintos “experimentos virtuales” con levaduras creciendo en un medio de cultivo, y realizar el estudio estadístico pertinente con los datos obtenidos, un estudio que, si se hubiera tenido que realizar “en vivo” en el centro académico, hubiera sido inviable por el coste económico, el tiempo de realización y los recursos diversos que hubiera requerido.

### **Modelo computacional para la generación de “experimentos virtuales”**

La plataforma NetLogo es especialmente adecuada para la modelización y simulación de sistemas complejos que evolucionan en el tiempo, proporcionando un entorno de programación que permite la simulación de fenómenos tanto naturales como sociales a partir de la modelización discreta y desde la perspectiva de los modelos basados en agentes (Wilensky, 1999). Fue diseñado para tratar con agentes (o “individuos”) móviles actuando al mismo tiempo en un espacio con características cambiantes, y con un comportamiento dominado por las interacciones locales durante un periodo de tiempo. Es bastante simple para aquellos estudiantes y profesores que quieran iniciarse en este tipo de modelización discreta,

y está suficientemente avanzado como para que pueda servir como una poderosa herramienta a investigadores de muchos campos científicos (Wilensky and Rand, 2105).

El modelo computacional utilizado para la realización de los “experimentos virtuales” representa una población de levaduras (individuos) evolucionando en un medio de cultivo, modelo que fue desarrollado principalmente durante un proyecto de investigación (Portell et al. 2014a) y readaptado posteriormente para finalidades de divulgación científica (Portell et al. 2014b). El simulador empleado es este segundo modelo computacional simplificado que se llama INDISIM-YEAST\_NL, implementado en la plataforma NetLogo y disponible para los estudiantes en el campus virtual Atenea de la UPC. Además de proporcionar salidas gráficas y numéricas de variables de interés para este contexto biotecnológico, este simulador facilita la modificación de los valores de los parámetros iniciales para fijar las condiciones de cada “experimento virtual”, lo que posibilita el diseño de experimentos distintos asumiendo características diferentes para las levaduras y el medio de cultivo (Figura 1). Por ser un modelo estocástico puede ser utilizado también para obtener diversas réplicas de un mismo experimento. Cabe mencionar que no era la primera vez que los estudiantes utilizaban un simulador de este tipo, pues ya anteriormente habían sido usuarios de modelos disponibles en la biblioteca que tiene NetLogo durante el desarrollo de Matemáticas II (Ginovart 2014). Con el simulador INDISIM-YEAST\_NL se realizaron distintos “experimentos virtuales” con diversas levaduras creciendo (fermentando) en un medio de cultivo con nutriente (glucosa) y produciendo un producto final (etanol). Las condiciones iniciales (o valores de los parámetros de entrada) de cada “experimento virtual” fueron fijadas de antemano por el profesor, y asignadas a los estudiantes, pudiendo diferir de un estudiante a otro según fuera la actividad. Para cada “experimento virtual” realizado se pudieron observar valores de salida de un conjunto de variables respuesta que se controlan durante el desarrollo del experimento (a lo largo del tiempo), o bien valores observados en el momento de finalizar el experimento, o en un determinado tiempo prefijado. Las variables que se estudiaron fueron, entre otras, el número de levaduras del cultivo (el número total, pudiéndose distinguir los microorganismos viables y los microorganismos no viables), el porcentaje de células gemando (la gema de una célula madre es la hija o célula nueva que aparecerá en el proceso de reproducción), la masa de las levaduras, el número de cicatrices (o edad) de las levaduras (cada gema deja una cicatriz en la célula madre debido a separación física que se produce), la glucosa del cultivo,

el etanol producido, la productividad y el rendimiento del proceso de fermentación (índices de interés en biotecnología). Al iniciar un experimento virtual o arrancar una simulación con los valores de los parámetros de entrada fijados, se puede ver cómo evolucionan las variables con los gráficos de la interface así como los resultados numéricos en ventanas de salidas de algunas variables de interés (Figura 1). También en la interface de INDISIM-YEAST\_NL se visualizan las “levaduras virtuales” en el medio líquido donde crecen y se reproducen, consumiendo glucosa y produciendo etanol. Un “experimento virtual” puede durar unos 4-5 minutos aproximadamente, y cuando finaliza se observa que el propio simulador ha generado un fichero de tipo texto (Simulation\_Results.txt) que contiene todos los datos controlados por el modelo, para que puedan ser analizados con otros programas.



Figura 1. Captura de pantalla de la interface del simulador INDISIM-YEAST\_NL.

## Resultados y discusión

Cada “experimento virtual” realizado, es decir, cada simulación ejecutada, dependía de valores de variables aleatorias con distribución normal (con un valor promedio y una desviación estándar) insertadas en instrucciones para acciones o características individuales, por lo que los resultados específicos de las diferentes evoluciones temporales no eran exactamente iguales. Cada realización del simulador con los mismos parámetros de entrada se consideró una réplica del experimento. El análisis estadístico de los datos simulados con diversos “experimentos virtuales”, o con distintas réplicas de un experimento, permitió trabajar con procedimientos propios tanto de la estadística descriptiva como de la estadística inferencial. Los estudiantes pudieron entrenarse en tareas habituales de un entorno de investigación o de actividad profesional como son la realización de experimentos diversos, la obtención de distintos tipos de información sobre cada uno de los experimentos, la automatización y el manejo de dicha información, el análisis de los datos conseguidos, y la comunicación de los resultados relevantes de dicho análisis.

La extensión y diversidad de los resultados conseguidos por los estudiantes fue notable. En un contexto de interés innegable para sus estudios de grado, los estudiantes: i) Discutieron diversos tipos de representaciones gráficas, como evoluciones temporales de variables, histogramas de distribución temporal de masas de las levaduras, observaciones atípicas mediante diagramas de caja, y diagramas de dispersión para valorar la relación entre parejas de variables cuantitativas; ii) Analizaron diferentes modelos de crecimiento poblacional como el lineal, exponencial y logístico; iii) Utilizaron transformaciones no lineales de datos como la transformación logarítmica, y estimaron parámetros cinéticos propios de un crecimiento microbiano como la densidad de población inicial y de población máxima mantenida por el ambiente, el tiempo en que finaliza la fase de adaptación de los microorganismos al entorno, el tiempo en el que se alcanza la máxima densidad de población antes de empezar a decrecer, y la constante específica de velocidad de crecimiento ( $\mu$ ) del microorganismo, parámetro relevante en aplicaciones biotecnológicas (Buchanan et al. 1997); iv) Hicieron uso de regresiones lineales simples para la determinación de parámetros cinéticos del crecimiento microbiano; v) Utilizaron muestras aleatorias simples para la construcción de intervalos de confianza y para el contraste de hipótesis de medias y varianzas

de parámetros y variables de interés; vi) Exploraron distintas opciones para una regresión no lineal y discutieron su idoneidad para los datos de las evoluciones temporales de las variables que tenían; vii) Utilizaron el análisis de la variancia unifactorial y el método de Tukey para la separación de medias con el fin de investigar el efecto de la resistencia al etanol de distintas cepas de levadura sobre la variable rendimiento del proceso; viii) Propusieron posibles trabajos de investigación y el diseño del experimento correspondiente para obtener y analizar los datos simulados obtenidos.

Durante la realización de estas actividades la mayoría de los estudiantes mostró una actitud positiva y proactiva. En general, los estudiantes fueron capaces de seleccionar, asimilar, procesar e interpretar los datos conseguidos con los “experimentos virtuales”. Las cuestiones tratadas resultaron atractivas para los estudiantes, especialmente porque reconocieron temáticas estudiadas en otras asignaturas del currículo, por lo que conectar dichas cuestiones con procedimientos estadísticos presentados durante el desarrollo de la asignatura de Estadística evidenció la aplicabilidad inmediata de estas técnicas estadísticas y ofreció una perspectiva interdisciplinar. Como en cualquier otro ámbito matemático, para que el alumno participe activamente del proceso de enseñanza-aprendizaje de la estadística debe ser estimulado de manera adecuada con actividades o tareas efectivas centradas en el área específica en la que se encuentren sus estudios, requiriendo del estudiante reflexión, comprensión y acción que permita cimentar sus progresos en la adquisición de conceptos y métodos estadísticos. Estamos convencidas de que estas actividades sintonizan con esta forma de hacer y favorecen el proceso de enseñanza-aprendizaje en el que cual se encuentran inmersas.

### **Referencias bibliográficas**

- Batanero, C., Díaz, C., Contreras, J. M., & Roa, R. (2013) El sentido estadístico y su desarrollo. *Números*, 83, 7-18.
- Buchanan, R. L., Whiting, R.C., & Damert, W.C. (1997). When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiology*, 14, 313-326.
- Ginovart, M. (2014). Discovering the power of individual-based modelling in teaching and learning: the study of a predator-prey system. *Journal of Science Education and Technology*, 23, 496-513.
- Portell, X., Gras, A., & Ginovart, M. (2014a). INDISIM-Saccha, an individual-based model to tackle *Saccharomyces cerevisiae* fermentations. *Ecological Modelling*, 279, 12-23.



- Portell, X., Gras, A., Prats, C., & Ginovart, M. (2014b). The microbial individual-based model INDISIM-YEAST ready to be used in the free access NetLogo modelling environment. En A. Méndez-Vilas (Ed.), *Industrial, medical and environmental applications of microorganisms: current status and trends*, pp. 686-691. Badajoz: Wageningen Academic Publishers.
- Serradó, A. (2013). El proyecto internacional de alfabetización estadística. *Números*, 83, 19-33.
- Walz, M. F. (2015). Programa de estadística aplicada a la biología: una propuesta. *Números*, 88, 17-29.
- Wilensky, U. (1999). "NetLogo". Center for Connected Learning and Computer-Based Modelling, Northwestern University. Evanston, IL. <http://ccl.northwestern.edu/netlogo/>.
- Wilensky, U., & Rand W. (2015). *An introduction to agent-based modelling - Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. London: The MIT Press.