

Innovación en al aprendizaje estadístico: Inferencia bayesiana amigable en el lenguaje de programación R

Innovation in statistical learning: Friendly Bayesian inference in the R programming language

Marina Martínez Álvaro^a y Cristina Casto Rebollo^a

^aDepartamento de Ciencia Animal, Universitat Politècnica de València, Camino de Vera. 14, Valencia 46022, España, mamaral9@upv.es;  cricasre@upv.es; 

How to cite: Marina Martínez-Álvaro y Cristina Casto-Rebollo. 2023. Innovación en al aprendizaje estadístico: Inferencia bayesiana amigable en el lenguaje de programación R. En libro de actas: *IX Congreso de Innovación Educativa y Docencia en Red*. Valencia, 13 - 14 de julio de 2023.
Doi:<https://doi.org/10.4995/INRED2023.2023.16684>

Abstract

This paper presents the results of the first year of the implementation of the runRabbit program as an innovation tool in learning Bayesian inference applied to the subject of Quantitative Genetics III. The challenge faced by 14 students of the Animal Genetic Improvement Master of the Polytechnic University of Valencia consisted of solving a frequent statistical problem (calculating differences between groups by fitting a linear model) using runRabbit. This is didactic and interactive software programmed in the R language and designed to enhance the understanding of the bases of this branch of statistics. The results obtained by the students in the proposed practical exercise showed that the runRabbit tool helped them understand the subject. Regarding the value of runRabbit by students, the results of the surveys showed high satisfaction with the use of the program, as well as a desire to use Bayesian inference to solve new challenges. The experience of applying runRabbit as a learning tool has been very positive both from the point of view of understanding the theoretical bases of the subject, and from the potential of the software to solve new professional challenges.

Keywords: Bayesian Inference, runRabbit, animal breeding, graphics, R, training, methodology.

Resumen

En este trabajo se presentan los resultados del primer año de implementación del programa runRabbit como herramienta innovadora en la enseñanza de la inferencia bayesiana, aplicada a la asignatura de Genética Cuantitativa III para 14 alumnos del Máster de Mejora Genética Animal en la Universitat Politècnica de València. El objetivo de la actividad fue resolver un problema estadístico frecuente, calcular las diferencias entre grupos ajustando un modelo lineal, usando runRabbit, un software didáctico e interactivo programado en el lenguaje R. El programa está diseñado para potenciar la comprensión de las bases de esta rama de la estadística. Los resultados obtenidos en el ejercicio práctico demostraron que

runRabbit ayudó a los estudiantes a comprender mejor la asignatura. Además, los resultados de las encuestas indicaron que los estudiantes estuvieron altamente satisfechos con el uso de runRabbit y expresaron su deseo de aplicar la inferencia bayesiana en la resolución de nuevos desafíos. La experiencia de la aplicación de runRabbit como herramienta de aprendizaje fue muy positiva, tanto en términos de la comprensión de las bases teóricas de la asignatura, como en cuanto al potencial del software para resolver nuevos retos profesionales.

Palabras clave: *Inferencia Bayesiana, runRabbit, mejora genética animal, visualización, R, formación, metodología.*

Introducción

La asignatura de Genética Cuantitativa III introduce a los alumnos en el uso de la estadística bayesiana para resolver análisis genéticos. Aunque esta materia tiene un gran potencial de aplicación en campos que requieren análisis de datos, como la mejora genética animal, su contenido teórico y abstracto puede resultar complejo para muchos estudiantes. Para comprender la estadística bayesiana se requieren conocimientos básicos de estadística, álgebra y teoría probabilística, combinados con el Teorema de Bayes. El objetivo de la asignatura es formar profesionales especializados en mejora genética animal, dotándolos de las competencias necesarias para aplicar la estadística bayesiana en este ámbito, tal y como se describe en Blasco (2017). Sin embargo, muchos alumnos no están familiarizados con estos conceptos o se enfrentan por primera vez al análisis estadístico bayesiano, lo que hace que la asignatura pueda resultarles compleja.

Hasta ahora, la metodología utilizada en esta materia se centraba principalmente en clases magistrales. Estas son esenciales para la enseñanza, pero sin un enfoque práctico adecuado pueden inducir a la pasividad en el alumno, dificultando la asimilación inmediata de los contenidos y limitando el pensamiento crítico y creativo (Atkins y Brown, 2002; Nichols, 2002). Por eso, es importante complementar las clases magistrales con sesiones prácticas que ayuden a clarificar y consolidar los conocimientos adquiridos (Coll y col, 2009). en las que los alumnos puedan resolver ejercicios y problemas reales. Sin embargo, en el caso de la asignatura de Genética Cuantitativa III, la falta de un software de inferencia bayesiana amigable e intuitivo limitaba la realización de sesiones prácticas. Aunque existen programas de este tipo, su uso suele ser poco intuitivo para usuarios principiantes, con salidas poco claras o programados en lenguajes obsoletos que ofrecen muy poca flexibilidad.

Innovación en el aprendizaje basado en nuevas herramientas

En el campo de la mejora genética animal, es común resolver problemas estadísticos utilizando software desarrollado en lenguajes de programación como R (por ejemplo, ASREML, <https://vsni.co.uk/software/asreml-r>). Sin embargo, la mayoría de los programas disponibles en este campo utilizan la estadística clásica frecuentista para resolver problemas estadísticos. Aunque la estadística clásica frecuentista está ampliamente extendida, presenta serios problemas de interpretación que a menudo conducen a la malinterpretación de los resultados (Amrhein y col., 2019). Como alternativa, la inferencia Bayesiana proporciona soluciones con interpretaciones más claras basadas en probabilidades obtenidas a partir de la información de los propios datos. En este campo, el Profesor Agustín Blasco de la UPV ha realizado una amplia labor docente a nivel nacional e internacional, desarrollando el uso didáctico de las inferencias bayesianas que facilitan la interpretación de los resultados (Blasco, 2017 y Blasco, 2021). A pesar de que existen programas de inferencia bayesiana para resolver modelos lineales, estos no suelen ofrecer salidas intuitivas ni flexibilidad en la inferencia de las salidas. Por lo tanto, es necesario desarrollar

herramientas amigables que permitan a los usuarios principiantes comprender y aplicar eficazmente la inferencia bayesiana en este campo.

Recientemente, las Dras. Marina Martínez-Álvaro y Cristina Casto Rebollo de la Universitat Politècnica de València han desarrollado un software llamado "runRabbit" para resolver modelos lineales utilizando inferencia bayesiana, que ofrece salidas altamente intuitivas y flexibles (disponible en <https://github.com/VLabUPV/runRabbit>). Estas herramientas interactivas de aprendizaje, como el uso de programas informáticos, obligan al usuario a reflexionar y previenen la situación común donde el usuario solo ejecuta el programa de manera mecánica (Cornejo y col., 2018). Este proyecto tiene como objetivo fomentar la comprensión de la inferencia bayesiana entre los estudiantes principiantes y promover su uso en el ámbito profesional de la mejora genética. Para evaluar el éxito del proyecto, se seguirán dos estrategias:

- 1) Se evaluará si el programa realmente mejora y facilita el aprendizaje y la comprensión de la inferencia bayesiana.
- 2) Se investigará si la herramienta propuesta, el programa runRabbit, ayuda a lograr el objetivo anterior, y si los estudiantes lo encuentran lo suficientemente amigable y fácil de usar como para emplearlo en la resolución de desafíos profesionales futuros.

Objetivos

El objetivo de este estudio es evaluar los resultados y la satisfacción de los estudiantes durante el primer año de implementación de las prácticas de Inferencia Bayesiana en el lenguaje de programación R, como parte de la asignatura de Genética Cuantitativa III del Máster de Mejora Genética Animal y Biotecnología de la Reproducción de la Universidad Politècnica de Valencia (UPV). Para llevar a cabo esta evaluación, se ha utilizado el programa *runRabbit* como herramienta de aprendizaje.

Desarrollo de la Innovación

En el primer año de implementación de prácticas de Inferencia Bayesiana en lenguaje de programación R, se ha diseñado el programa *runRabbit* en lenguaje de programación R que como herramienta de aprendizaje. Este programa es altamente interactivo y guía al usuario a través de una serie de preguntas que deben responderse para diseñar y reflexionar sobre el análisis (ver Figura 1).

```
=====
                RABBIT
                Version 1.0
=====

Enter the name of the datafile with its extension .csv, .xls or .xlsx
DataFixed.xlsx
Has the data file missing values? (Enter Yes=Y or No=N)
Y
Please enter the missing value. If its a blank enter a space
NA
The number of rows in the data file is 502
Enter the total number of traits
2
Help: the header of the datafile is  AE OP LG Sex LW pH IMF PFat
Do you want to enter the name of all Traits at once based on their columns number in the datafile? (Enter Yes=Y or No=N)
N
Enter the name of the Trait 1
IMF
```

Figura 1. Interface del programa runRabbit para la introducción y definición de la estructura de los datos.

Posteriormente, se diseñó una sesión de prácticas en la asignatura de Genética Cuantitativa III del Máster de Mejora Genética Animal de la UPV, en la que se resolvió un problema real utilizando el programa

runRabbit. Para esta sesión, se elaboró un manual explicativo del programa (se puede consultar en la documentación anexa al programa en la página <https://github.com/VLabUPV/runRabbit>), y un documento que detallaba los objetivos de la práctica. Los estudiantes tuvieron que estimar la diferencia en grasa intramuscular entre dos líneas de selección de conejo por grasa intramuscular utilizando un modelo lineal para corregir las variables que podían influir en la medición (sexo, orden de parto en el cuál nació en animal, estación del año dónde se realizó la medición y peso vivo del animal).

Además, tuvieron que establecer un umbral de valor relevante a partir del cual las diferencias entre las líneas llevarían a tomar una importante decisión económica (por ejemplo, dejar de usar una línea para empezar a usar otra), opción implementada en el propio programa. *RunRabbit* también fuerza a los estudiantes a diferenciar entre los efectos interesantes (tratamientos) y aquellos que no lo son (ruido), además de ofrecer la posibilidad de obtener muestras de las distribuciones de las estimaciones para calcular las probabilidades (inferencias) a mano o mediante el propio programa. En este ejercicio práctico, los estudiantes debían estimar las inferencias de las estimaciones por sí mismos, así como generar gráficos intuitivos que representaran las probabilidades estimadas (ver por ejemplo la Figura 2). Se les concedió a los alumnos una semana para entregar la práctica resuelta. Es importante destacar que la implementación de esta práctica permitió a los estudiantes aprender a aplicar técnicas de Inferencia Bayesiana a problemas reales y tomar decisiones basadas en evidencia probabilística.

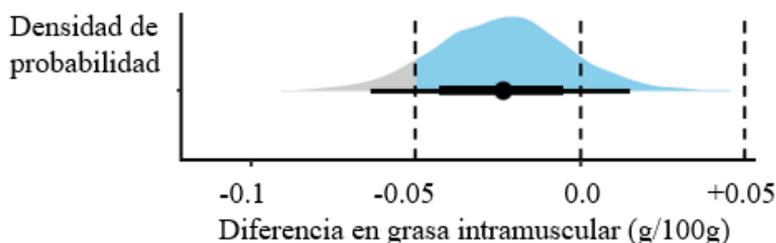


Figura 2. Salida gráfica del programa runRabbit donde se muestra la distribución de probabilidad marginal posterior de la diferencia entre líneas en Grasa intramuscular. Al asumir un valor relevavante de diferencia entre grasa de 0.05 g por 100g de carne, se representa en la probabilidad de similitud y en gris la probabilidad de relevancia.

Asimismo, al final de la asignatura y con la práctica entregda, se realizó una evaluación de la implementación de la práctica a través de una encuesta a los estudiantes (Tabla 1) y se evaluó la opinión de los estudiantes en cuanto a la implementación del programa runRabbit.

Tabla 1. Cuestionario realizado a los alumnos que llevaron a cabo el reto

Aspecto Valorado	Cuestionario	Respuestas
Evaluación del programa como herramienta de aprendizaje	¿Te ha resultado sencillo el uso del programa <i>runRabbit</i> ?	Muy sencillo – Sencillo – Ni sencillo ni difícil – Difícil –Muy difícil
	¿Resolver el reto planteado con el software <i>runRabbit</i> te ha	Muy de acuerdo – De acuerdo – En desacuerdo – Muy en desacuerdo

	ayudado a entender la Inferencia Bayesiana?	
	¿Crees que las salidas del programa <i>runRabbit</i> (gráficos, distribuciones posteriores e inferencias) son didácticas?	Muy satisfecho – Satisfecho – Insatisfecho – Muy insatisfecho
Manual	Evalúa el manual del programa <i>runRabbit</i> . ¿Se explica bien el funcionamiento del programa y las posibilidades que ofrece?	Muy de acuerdo – De acuerdo – En desacuerdo – Muy en desacuerdo
Implementación en nuevos retos profesionales	¿Usarías el programa <i>runRabbit</i> para resolver nuevos retos?	Seguro – Probablemente – No es probable – No lo recomiendo
	¿Recomendarías el uso del programa a otras personas interesadas en introducirse en la inferencia Bayesiana?	Seguro – Probablemente – No es probable – No lo recomiendo
Satisfacción	¿Qué mejorarías?	Respuesta corta
	Satisfacción general del programa <i>runRabbit</i>	Muy satisfecho – Satisfecho – Insatisfecho – Muy insatisfecho

Finalmente, se quiso evaluar si el Grado universitario que los estudiantes habían estudiado antes de matricularse en el máster tenía algún efecto en la nota obtenida en la práctica. Este análisis se formalizó usando el propio programa *runRabbit*.

Resultados

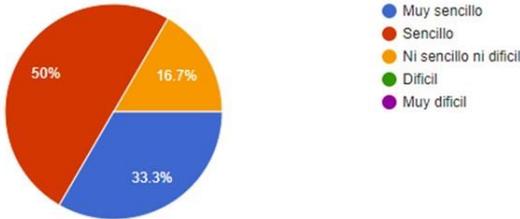
En total, 14 estudiantes participaron en el reto, de los cuales 6 estudiaron el grado de Biotecnología, 5 Biología, 2 Veterinaria y 1 Ingeniería Agronómica antes de matricularse en el Máster de Mejora Genética Animal de la UPV. En general, el alumnado no estaba familiarizado con la estadística bayesiana antes de comenzar la asignatura. El conocimiento previo sobre estadística bayesiana era el siguiente: nada familiar (66.7%), no muy familiar (25%) y algo familiar (8.3%). Tras resolver la práctica, la nota media de los estudiantes fue de 9.7 y ninguno de ellos tuvo una nota menor a 5. Se evaluó el efecto de la "Formación anterior" usando tres niveles: Biología, Biotecnología y Otros, y no hubo diferencias relevantes entre grupos (probabilidad de relevancia, asumiendo como valor relevante 0.5 puntos, fue de 0.12).

Los resultados de la encuesta de evaluación del programa, respondida por los 12 alumnos, se representan en la Figura 3. Los estudiantes evaluaron el programa como una herramienta de aprendizaje satisfactoria y de uso amigable. El 33.3% lo encontró muy sencillo, el 50% sencillo y el restante 16.7% lo encontró ni sencillo ni difícil. Además, los estudiantes consideraron que usar el programa *runRabbit* para resolver el ejercicio propuesto les había facilitado el entendimiento de la estadística bayesiana; con un 66.7% de alumnos estando muy de acuerdo y un 33.3% estando de acuerdo, lo cual era uno de los objetivos principales del proyecto. La mayoría de los estudiantes pensó que las salidas del programa (gráficos,

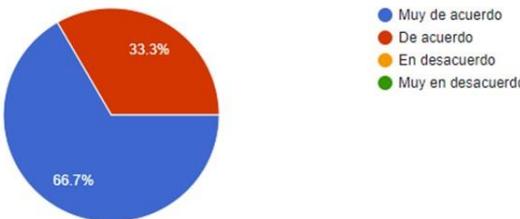
distribuciones posteriores e inferencias) fueron didácticas (58.3% quedaron muy satisfechos y 41.7% satisfechos).

Evaluación del programa como herramienta de aprendizaje

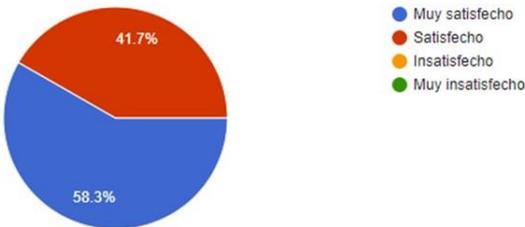
¿Te ha resultado sencillo el uso del programa runRabbit?



¿Resolver el reto planteado con el software runRabbit te ha ayudado a entender la Inferencia Bayesiana?

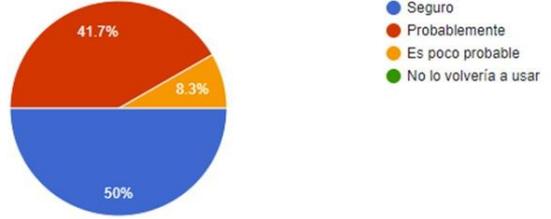


¿Crees que las salidas del programa runRabbit (gráficos, distribuciones posteriores e inferencias) son didácticas?

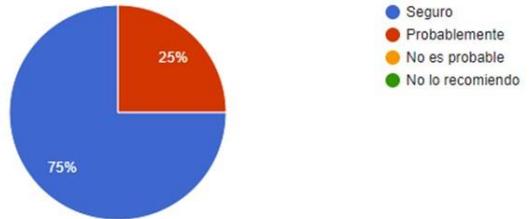


Implementación en nuevos retos profesionales

¿Usarías el programa runRabbit para resolver nuevos retos?

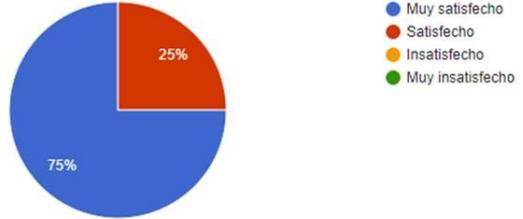


¿Recomendarías el uso del programa a otras personas interesadas en introducirse en la inferencia Bayesiana?



Satisfacción

Satisfacción general el programa runRabbit



Manual

Evalúa el manual del programa runRabbit. ¿Se explica bien el funcionamiento del programa y las posibilidades que ofrece?

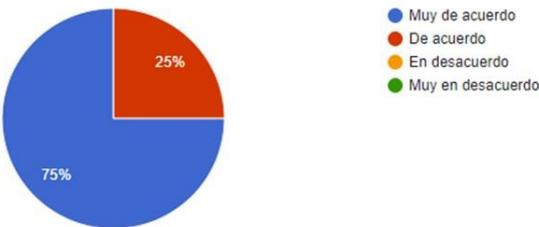


Figura 3. Resultados de las encuestas de valoración del programa runRabbit

En cuanto al manual de uso del programa, los alumnos estuvieron por lo general de acuerdo en que era explicativo y exponía de manera coherente las posibilidades que ofrece el programa (75% muy de acuerdo y 25% de acuerdo). Finalmente, los alumnos encontraron que el programa es una herramienta útil para enfrentar nuevos retos profesionales, con un 91.7% de los alumnos respondiendo que usarían de nuevo el programa con seguridad o alta probabilidad, y que además lo recomendarían a otros usuarios (75%). Cuando les preguntamos sobre las posibles mejoras que podríamos implementar, tanto en la práctica como en el software, un alumno respondió solicitando un docente auxiliar en el aula para que, cuando un alumno tenga una dificultad, no sea necesario parar la clase para atenderle; lo cual tendremos en cuenta en las próximas ediciones. La satisfacción general del programa fue del 75% muy satisfecho y del 25% satisfecho.

Conclusiones

En conclusión, la aplicación piloto del software runRabbit en el aprendizaje de Inferencia Bayesiana ha resultado en una experiencia altamente positiva para el alumnado, tanto en términos de aprendizaje como de satisfacción. Además, el uso de esta herramienta ha ayudado al desarrollo de una nueva capacidad de pensamiento de cálculo, y los estudiantes han expresado su disposición para utilizarla en futuros retos profesionales. En resumen, los resultados obtenidos sugieren que el software runRabbit puede ser una herramienta valiosa para mejorar la enseñanza de estadística bayesiana.

Referencias

- Amrhein, Valentin, Sander Greenland, and Blake McShane. "Scientists rise up against statistical significance." *Nature* 567.7748 (2019): 305-307.
- Atkins, Madeleine, and George Brown. *Effective teaching in higher education*. Routledge, 2002.
- Blasco, Agustín, and P. D. A. Blasco. *Bayesian data analysis for animal scientists*. Vol. 265. New York, NY, USA:: Springer, 2017.
- Blasco, Agustín. *Mejora genética animal*. Síntesis, 2021.
- Coll Serrano, V., & Blasco Blasco, O. M. (2009). Aprendizaje de la estadística económico-empresarial y uso de las TICs. *EduTec. Revista Electrónica De Tecnología Educativa*, (28), a109. <https://doi.org/10.21556/edutec.2009.28.457>
- Cornejo, María Alejandrina Nivela, et al. "Diseño de software interactivo en las matemáticas." *Journal of Science and Research: Revista Ciencia e Investigación* 3.1 (2018): 27-31.
- Nicholls, Gill. *Developing teaching and learning in higher education*. Routledge, 2002.