

OPTIMIZACIÓN DE BANCOS DE PREGUNTAS MEDIANTE R

Pablo Ortiz y Martín Saino

Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

pablo.ortiz@unc.edu.ar

Como consecuencia de la pandemia del COVID-19, en 2020 se presentó la apremiante necesidad de virtualizar nuestras asignaturas. Esto resultó en un considerable volumen de tareas para las docentes concentradas en muy poco tiempo y, consecuentemente, exigió repensar nuestra manera de trabajar y organizarnos. Muchas de estas tareas, principalmente las que implican más tiempo, son proclives a ser automatizadas, por ejemplo, el desarrollo de instrumentos de examen, una vez trabajados y valorados qué temas, conceptos y competencias se quieren evaluar. El presente trabajo describe nuestra experiencia en el uso, adaptación y difusión de herramientas para la automatización de la generación de bancos de preguntas en plataformas educativas de cursos masivos en la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba.

INTRODUCCIÓN

La necesidad de virtualizar nuestros cursos tras la pandemia exigió un considerable volumen de tareas a desarrollar en poco tiempo: clases para su dictado vía *streaming* y videos, materiales y actividades adaptados a la modalidad, métodos adecuados de evaluación, comunicación efectiva con los estudiantes, entre otros. Particularmente, en los cursos de Estadística de las carreras de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba (FCE-UNC), hasta el advenimiento de la pandemia, nuestros métodos de evaluación consistieron en exámenes presenciales escritos. Trasladar de manera adecuada estas instancias evaluativas a la virtualidad, mediante el empleo de plataformas educativas (i.e., Moodle), constituyó un gran desafío, particularmente en un contexto de cursos con una gran cantidad de estudiantes.

En este trabajo compartimos la experiencia de la solución implementada en un curso masivo (con alrededor de 300 estudiantes) de Estadística II de las carreras de Contador Público y Licenciatura en Administración de la FCE-UNC. Tras adoptar la modalidad virtual, el método de evaluación de esta asignatura dos exámenes parciales administrados de manera *on-line* mediante cuestionarios en el aula virtual Moodle (con la posibilidad de que los estudiantes dispusieran de material bibliográfico durante su realización), complementados con dos trabajos prácticos de aplicación a desarrollar con consignas abiertas; al examen *on-line* y al trabajo se le asigna un puntaje de 80% y 20%, respectivamente, para conformar las notas definitivas y obtener la regularidad. (Para ganar regularidad y avanzar, los estudiantes deben asistir a las actividades del curso y aprobar los exámenes.) Luego, quienes acceden a la regularidad rinden un examen final de dos instancias: escrito primero y oral seguidamente.

De este modo, fue necesario generar un banco de preguntas amplio y variado que cumpliera con los objetivos de evaluación y se adecuara a la modalidad y a las condiciones establecidas. En el año 2020, el primero con las restricciones derivadas de la pandemia, las distintas preguntas que conformaban los exámenes *on-line* eran desarrolladas y cargadas de manera manual. De este modo, las tareas operativas de carga y posterior control de preguntas para evitar errores llevaban un tiempo considerablemente alto en relación al tiempo destinado al diseño adecuado a los aspectos que se pretendían evaluar. Asimismo, como consecuencia de esta limitación de tiempo y el acotado equipo docente, las preguntas realizadas eran mucho más directas, en general sin la inclusión de incisos que permitan evaluar aspectos conceptuales; los bancos de preguntas generados fueron de un tamaño reducido, entre 40 y 60 preguntas para cada examen, en relación a las necesidades por la cantidad de estudiantes y las condiciones en las que rendían (en sus hogares y con bibliografía a disposición).

Desde el año 2021, trabajamos sobre la automatización de la generación de preguntas atendiendo a que los enunciados se componen, en general, de características cualitativas y datos, valores de parámetros o formulaciones matemáticas susceptibles de ser descompuestas, elementos donde intervienen diferentes conceptos, y que claramente pueden ser aleatorizados para generar variedad y cantidad de alternativas. De este modo, el equipo de la cátedra pudo abocarse en mayor medida al diseño pormenorizado de enunciados y preguntas, de manera tal que incluyesen la

identificación de elementos conceptuales claves implícitos en la resolución de los mismos y que suelen ser evaluados a modo de justificación de las respuestas en los exámenes escritos. Mediante la programación del proceso en software R y la utilización de librerías específicas para la exportación a Moodle pudimos crear bancos de preguntas muchos más amplios y diversos, evitando errores que suelen surgir de la carga manual. Adicionalmente, adquirimos nuevas habilidades y capacidades que luego transmitimos a otros docentes de la FCE y otras universidades del país mediante talleres y cursos.

IMPLEMENTACIÓN

El tipo de preguntas que admite el aula virtual Moodle admite como tipo de respuestas, en términos generales, la selección de alternativas (simples o múltiples), numéricas y abiertas; además algunas variantes o combinaciones de estas (Moodle, s.f.). Los exámenes con este tipo de preguntas, con ítems de respuesta cerrada, tienen la ventaja de reducir los tiempos y recursos necesarios para la corrección, así como una retroalimentación inmediata para los estudiantes; por lo tanto, resultan adecuados para un contexto de masividad y virtualidad (Basabe y Amantea, 2020), como la transitada durante las restricciones de la pandemia. Sin embargo, tiene las desventajas de servir para evaluar principalmente la capacidad de evocación o recuerdo y que los estudiantes respondan correctamente por un conocimiento parcial (no total) de la respuesta (Pardo Adames y Rocha Gaona, 2009). Estas desventajas pueden ser superadas con la elaboración de proposiciones que involucren procesos cognitivos relativamente más exigentes, estableciendo diversas preguntas combinadas en el mismo problema y con los distractores adecuados como alternativas incorrectas que conduzcan a los estudiantes a una formulación de su respuesta de manera análoga a una justificación en un examen escrito. El diseño de preguntas y exámenes con estas características implican una muy buena planificación y un considerable esfuerzo y dedicación; particularmente cuando el equipo docente es reducido y el grupo de estudiantes es numeroso, siendo necesario un importante volumen de preguntas para evitar prácticas no deseadas como la comunicación de respuestas entre los estudiantes (Basabe y Amantea, 2020).

Dadas las características de los contenidos de la asignatura Estadística II, los problemas planteados en los distintos exámenes parciales se conforman de enunciados núcleo compuestos de elementos cualitativos (que determinan las técnicas o herramientas estadísticas apropiadas en cada caso) y datos o valores de parámetros para su aplicación y la consecuente obtención de resultados y conclusiones. Así, aprovechando estas características, utilizamos el software R para codificar funciones *ad-hoc* que nos permitieron la definición aleatoria de esos elementos, datos y parámetros (en base a modelos probabilísticos más o menos complejos), vinculándolo con el texto que conforman los enunciados y las preguntas en LaTeX o Markdown mediante la librería *knitr* (Xie, 2015). De esta manera fue posible generar problemas y preguntas dinámicas de manera masiva, con respuestas cerradas, pero con alternativas correctas e incorrectas con distractores de forma tal que permiten evaluar la identificación precisa de los conceptos estadísticos que se pretenden evaluar. Esto se logró desglosando adecuadamente las preguntas en incisos de forma tal que sean equivalentes a la justificación de la resolución en un examen escrito.

La Figura 1 muestra un ejemplo de pregunta relacionada con prueba de hipótesis para la media generada con el método descrito, tal como es visualizada por los estudiantes en el aula virtual. Los elementos que se resaltan, tanto en el enunciado como las respuestas alternativas a las diferentes preguntas, son generados de manera aleatoria bajo las especificaciones necesarias para mantener la consistencia del problema. En efecto, en el inciso *a* no solo se debe determinar el sentido la prueba, sino también identificar la estimación de la media poblacional y su valor bajo hipótesis nula; la distribución del estadístico adecuado que se solicita en el inciso *b* depende de diferentes aspectos que el estudiante debe evaluar, como el tamaño de la muestra; para determinar la expresión del estadístico empleado en la prueba de hipótesis, que se pide en el inciso *c*, el estudiante debe identificar el valor del parámetro bajo hipótesis nula y la estimación de la media muestral, así como su error estándar y su cálculo; en el inciso *d* se debe completar con el valor del calculado del estadístico; y en el inciso *e* se debe concluir en el contexto de problema atendiendo al resultado obtenido de la prueba mientras que en el *f* identificar el tipo de error que dependerá de la conclusión de la prueba.

Fuente: elaboración propia

Una empresa que administra máquinas de expendedoras (vending) de bebidas a nivel nacional ha registrado que las ventas diarias promedio de sus máquinas es, hasta el momento, de 1800 pesos. Tras una política de relocalización de las máquinas, el gerente de marketing sostiene que se produjo un incremento del promedio de ventas diarias. Para evaluar esta situación se seleccionó una muestra aleatoria de 38 máquinas, a partir de la cual se estimó una media de 1918,44 pesos y una desviación estándar de 438,83 pesos.

Realice el procedimiento adecuado para definir si la empresa debe continuar con la relocalización de las máquinas considerando un nivel de significación $\alpha = 0,10$. (Asuma que el valor de ventas de bebidas distribuye normal).

Ayuda: el valor crítico del estadístico en valor absoluto es igual a 1,30. (El signo -o los signos- del valor crítico dependen del tipo de prueba)

RESPONDA LOS SIGUIENTES PUNTOS:

(Inciso a) ¿Cuál de las siguientes hipótesis corresponden al problema planteado?

- (i) $H_0 : \mu \leq 1800$ versus $H_1 : \mu > 1800$
- (ii) $H_0 : \mu \leq 1918,44$ versus $H_1 : \mu > 1918,44$
- (iii) $H_0 : \mu \geq 1800$ versus $H_1 : \mu < 1800$
- (iv) $H_0 : \mu \geq 1918,44$ versus $H_1 : \mu < 1918,44$

Seleccionar la opción correcta:

- Opción (i)
- Opción (ii)
- Opción (iii)
- Opción (iv)

(Inciso b) ¿Cuál es la distribución del estadístico que se debe emplear?

Seleccionar la opción correcta:

- normal con media 0 y varianza 1
- t de Student con 37 grados de libertad
- t de Student con 38 grados de libertad
- t de Student con 39 grados de libertad

(Inciso c) ¿Cuál expresión corresponde al valor del estadístico observado?

Seleccionar la opción correcta:

- $(1800 - 1918,44) / 438,83$
- $(1800 - 1918,44) / 11,55$
- $(1918,44 - 1800) / 71,19$
- $(1918,44 - 1800) / 438,83$
- $(1800 - 1918,44) / 71,19$
- $(1918,44 - 1800) / 11,55$

(Inciso d) ¿Cuál es el valor observado del estadístico? (Utilice 4 decimales en el resultado)

Valor observado del estadístico:

(Inciso e) ¿A qué conclusión arriba?

Seleccionar la opción correcta:

- Debe continuar con la relocalización
- No continuar con la relocalización
- La localización es indiferente

(Inciso f) De acuerdo a la conclusión arribada, ¿qué tipo de error podría cometerse?

Seleccionar la opción correcta:

- Error de tipo I
- Error de tipo II
- Error de tipo I y Error de tipo II
- Error de tipo I o Error de tipo II

Figura 1. Ejemplo ejercicio “Prueba de Hipótesis” y “Regresión Lineal” los campos señalados con rojo son campos dinámicos

Mediante este procedimiento se definió la estructura de los problemas y preguntas para los distintos contenidos a evaluar en cada parcial, programando su implementación masiva y posterior exportación en conjunto a un archivo de formato XML de Moodle mediante la librería (Zeileis, et al., 2014), evitando la carga manual al aula virtual y el posterior tiempo de control para evitar errores que son proclives de ocurrir.

RESULTADOS DE LA EXPERIENCIA Y NUEVAS COMPETENCIAS DOCENTES

Desde el año 2021, mediante el uso R y Markdown para la automatización del proceso de generación de preguntas pudimos mejorar la calidad de las mismas, en cuanto a los aspectos evaluados, y optimizar el uso del tiempo. Los docentes que integramos la cátedra nos abocamos con mayor dedicación al diseño pormenorizado de enunciados y preguntas, incorporando de manera

cuidadosa incisos que se corresponden a justificación de las respuestas en los exámenes escritos y la identificación de conceptos claves. Esta mejora en los enunciados y preguntas en conjunción con la posibilidad de aleatorización de ciertos parámetros, datos y aspectos cualitativos nos permitió generar un volumen amplio de preguntas con la misma estructura pero de similar dificultad. Así, fue posible generar un banco de preguntas de mucho mayor tamaño, entre 800 y 1000 preguntas, obteniendo en la práctica un examen diferente por estudiante y garantizando un mismo nivel de dificultad.

Estas mejoras se reflejaron tanto en los resultados de los exámenes como en la opinión de los estudiantes en las encuestas realizadas por la FCE al final del curso. En el año 2021 el porcentaje de aprobados en el primer y segundo parcial resultaron similares a los de la prepandemia (65 % y 74 %, respectivamente), donde los exámenes eran escritos y presenciales; en el año 2020, el primero con exámenes *on-line* los resultados fueron un poco más bajos (con 52 % de aprobados en los parciales receptados). Estos resultados tienen consonancia con la evolución de opinión de los estudiantes en las encuestas: en 2021 el 68 % valoró como muy bueno o excelente los instrumentos de evaluación (contra el 56 % de 2020) y el 78 % consideró que el tiempo para resolverlos fue suficiente en la mayoría de los casos o siempre (contra el 71 % en 2020).

Por otro lado, el incremento sustancial en el volumen en las preguntas y líneas códigos (*scripts*) creados para su generación demandó establecer una muy buena organización del flujo de trabajo, más aún si pretendíamos asegurar la reproducibilidad del proceso (Gandrud, 2015; Xie, 2015). Se adoptaron así, desde el inicio, algunas medidas de buenas prácticas que vienen de la programación: desde seguir una guía de estilo en la codificación y documentación de los procesos hasta el uso de Github (<https://github.com/>) como repositorio en la nube con control de versiones. Si bien esta organización requirió de una importante inversión de tiempo inicial, a la postre fue sumamente redituable en cuanto la posibilidad de incrementar el potencial de las herramientas empleadas, así como en la optimización de los tiempos de trabajo al garantizar la reproducibilidad de los procesos.

Por último, si bien el software R es muy conocido y utilizado, no lo es su versatilidad de integración con plataformas como Moodle. Por esto, buscamos difundir su potencialidad a otros docentes a través de diferentes capacitaciones. La primera de ellas, en modalidad de taller, a la FCE-UNC a través del Área de Formación Docente y Producción Educativa (FyPE), donde participaron docentes de distintas disciplinas, no solo cuantitativas. También llevamos adelante un curso dirigido a docentes de todo el país a través de la Sociedad Argentina de Estadística. Todas estas instancias de socialización nos permitieron enriquecer la propuesta, no solo desde lo técnico, sino también por la posibilidad de discernir mejores diseños de preguntas para los instrumentos de evaluación.

CONCLUSIÓN

La programación del flujo de trabajo implicado en el diseño de exámenes nos permitió automatizar la creación masiva de ejercicios. Si bien esto nos demandó mucho esfuerzo inicial, pudimos destinar una mayor proporción de tiempo al diseño de los instrumentos de evaluación, entendida como una etapa más creativa, que a lo operativo.

Disponer de bancos de preguntas cuidadosamente pensados, amplios y variados nos permitió contar con un examen por estudiante con un nivel de dificultad similar para cada estudiante en un curso masivo; conteniendo todos los aspectos que se pretendían evaluar, incluyendo aquellos conceptuales que son de más difícil implementación en la modalidad *on-line*. De esta manera, pudimos evitar el empleo de modalidades de evaluación con restricciones en la navegación de los cuestionarios y/o tiempo, medidas que suelen ser empleadas para evitar prácticas no deseadas, pero que al mismo tiempo impacta negativamente sobre la experiencia de los estudiantes. Esto fue muy valorado por el grupo de estudiantes en distintos relevamientos que se realizan al final de cada semestre.

Por último, mediante la socialización de este *know how* con otros docentes logramos que otras cátedras, de asignaturas de distinta índole, pudieran conocer y emplear estas herramientas. La interacción con docentes de otras áreas nos permitió identificar mejoras en las preguntas que podíamos realizar, sorteando las limitaciones por el tipo de preguntas que pueden emplearse en las plataformas educativas. Además, estas capacitaciones fueron el puntapié para trabajar sinérgicamente con otras divisiones de la misma materia en semestres posteriores.

REFERENCIAS

- Basabe, L., y Amantea, A. (2020). *Diseño de exámenes con ítems de respuesta cerrada*. Centro de Innovación en Tecnología y Pedagogía (CITEP). <http://citep.rec.uba.ar/disenio-de-examenes-con-tems-de-respuesta-cerrada/>
- Gandrud, C. (2015). *Reproducible research with R and RStudio* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315382548>
- Moodle. (s.f.). *Documentación: Tipos de preguntas*. Recuperado Agosto 5, 2022, de https://docs.moodle.org/all/es/29/Tipos_de_preguntas
- Pardo Adames, C., y Rocha Gaona, M. (2009). *Reglas para elaborar ítems de formato de selección y producción*. UNESCO.
- Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315382487>
- Zeileis, A., Umlauf, N., & Leisch, F. (2014). Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond. *Journal of Statistical Software*, 58(1). <https://doi.org/10.18637/jss.v058.i01>