IX Jornadas de Enseñanza y Aprendizaje de la Estadística e Investigación Operativa

Programa / Comunicaciones

Universidad de Granada

Granada 4 y 5 de Abril de 2019

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

GENAEIO 2019

Programa

Comunicaciones

Granada, 4-5 abril 2019

.

Autores: Comité Organizador Jornadas GENAEIO 2019

Editores: @Los autores

ISBN: 978-84-09-18133-9

Lugar y fecha de edición: Granada, 30 de septiembre de 2020

Presentación

El Grupo de Trabajo "Enseñanza y Aprendizaje de la Estadística y la I.O." (GENAEIO) de la Sociedad de Estadística e Investigación Operativa (SEIO), ha venido organizando desde el año 2010 hasta el 2017 Jornadas sobre la Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa en diferentes universidades españolas y en la sede del Ministerio de Defensa. Concretamente, las I Jornadas celebradas en Melilla en 2010, las II Jornadas en Las Palmas de Gran Canaria en 2011, las III Jornadas en Pamplona en 2012, las IV Jornadas en Cádiz en 2013, las V Jornadas en Madrid en 2014, las VI Jornadas en Huelva en 2015, las VII Jornadas en La Rioja en 2016, y las VIII Jornadas en Gerona en 2017.

El éxito de las Jornadas celebradas hasta la actualidad y el compromiso de los miembros del grupo GENAEIO que pertenecen a la Universidad de Granada (UGR), llevaron a planificar las IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la I.O. (GENAEIO 2019) en Granada del 4 al 5 de Abril de 2019, organizadas por el Grupo GENAEIO y el Departamento de Estadística e I.O. de la Universidad de Granada, con sede en la Facultad de Ciencias y el Instituto de Matemáticas de la Universidad de Granada. Estas Jornadas suponen un punto de encuentro entre los profesores universitarios del área de Estadística e Investigación Operativa, así como profesores de secundaria, alumnos de Grado, Máster o Doctorado en Estadística y/o Matemáticas, profesionales del Ministerio de Defensa, además de otras personas interesadas en este ámbito.

La finalidad de las jornadas es intercambiar experiencias, de innovación educativa y de reflexión sobre las distintas problemáticas que afectan a la enseñanza y aprendizaje de la Estadística e Investigación Operativa en los diferentes niveles donde se imparten estas materias. Para conseguir este objetivo se planificaron dos talleres formativos de introducción a las técnicas de aprendizaje automático (machine learning) con el lenguaje de programación Python, y dos charlas coloquio sobre ideas y experiencias sobre la enseñanza de la estadística y otros temas relevantes como la prohibición de usar los p-valores. Además, se programó una sesión de pósteres en la que los participantes presentaron sus trabajos orientados a los recursos de innovación docente y experiencias didácticas desarrolladas para mejorar el proceso de enseñanza/aprendizaje. El nivel de acogida de las jornadas fue muy elevado con un total de 67 participantes repartidos entre las Universidades de Alcalá, Almería, Barcelona, Carlos III, Cartagena, Gerona, Granada, Huelva, Jaén, Politécnica de Cataluña, Politécnica de Valencia, Pública de Navarra, Rioja, Salamanca, Sevilla, y el CEIPSO "El Encinar" de Torrelodones en Madrid.

Hoy, más que nunca, con el auge del Big data y la Ciencia de Datos como herramientas potentísimas para la toma de decisiones en cualquier ámbito, no sólo el académico y científico, sino especialmente en el mundo empresarial, hacen que una formación en estadística de calidad sea fundamental, no sólo para los profesionales, sino para la sociedad en general que necesita tener una cultura estadística básica para interpretar correctamente la gran cantidad de datos que se generan diariamente. Por ello, el desarrollo de este tipo de eventos es fundamental para contribuir a la mejora de la formación en estadística en todos los ámbitos.

Finalmente, nos gustaría agradecer a todos los participantes en las jornadas y a las entidades colaboradoras, la colaboración y ayuda recibidas para que éstas pudieran celebrarse con éxito.

Granada, abril 2019
El Comité Organizador

Organizadores

- Departamento de Estadística e I.O. Universidad de Granada
- Instituto de Matemáticas. Universidad de Granada (IEMath-GR)
- Grupo de Trabajo GENAEIO de la SEIO

Comité Organizador

- Presidenta: Ana María Aguilera del Pino (Universidad de Granada)
- Secretario: Juan Eloy Ruiz Castro (Universidad de Granada)
- Coordinadora: Mónica Ortega Moreno (Universidad de Huelva)

Vocales:

- Christian José Acal González (Universidad de Granada)
- María del Carmen Aguilera Morillo (Universidad Carlos III de Madrid)
- Beatriz Cobo Rodríguez (Universidad de Granada)
- Ana María Lara Porras (Universidad de Granada)

Colaboradores

Queremos dar las gracias a los colaboradores de estas jornadas

- Sociedad de Estadística e Investigación Operativa
- Universidad de Granada
- Departamento de Estadística e Investigación Operativa
- Instituto de Matemáticas de la Universidad de Granada
- Facultad de Ciencias de la Universidad de Granada
- Academia de Ciencias Matemáticas, Físico-Químicas y Naturales de Granada

Índice

Índice VII

Programa 1

Talleres 5
Python como lenguaje para programación en estadística 7

Charlas-Coloquio 9
¿Están condenados los p-valores al destierro? 11
Ideas y experiencias sobre la enseñanza de la estadística 13

Sesión de pósteres 15
Visualización de datos con Tableau 17
Aprendiendo a modelizar datos de conteo a nivel de área con dependencia espacial y espacio-temporal mediante la aplicación web SSTCDapp 34
R Markdown como herramienta para la elaboración de materiales docentes . . 36
Apoyo al aprendizaje de la Estadística mediante el uso de datos reales 38
Dificultades de los estudiantes en la comprensión de la distribución muestral . . 47
Las infografías como estrategia de innovación docente para mejorar competencias en el Grado de Estadística 49
Introducción a la Estadística con SPSS y Statgraphics: ejercicios guiados y aplicaciones con datos reales 51
Test de Van Valen para la homocedasticidad multivariante 53
Actividad de Estadística Descriptiva: 'Reto: ¿eres capaz de descifrar la contraseña del candado?' 56
Software online como soporte docente e investigador 58
Uso de Geogebra en Programación Lineal 60
Generación de cuestionarios de autoevaluación de Moodle mediante el paquete exams de R 62
Intervalos de tolerancia en Grados en Ingenierías: un enfoque práctico 68
Producción automatizada de problemas de cálculo de probabilidades en Sweave 70
Realización de test de evaluación utilizando la plataforma ILIAS 72
Características que influyen en la competencia de los estudiantes españoles en PISA 2015 74

Técnicas Cuantitativas II: Motivación, utilidad y satisfacción 76
Aprendizaje cooperativo y gamificación: aplicación para la enseñanza en Estadística económica, empresarial y del sector turístico 78
Estimulando el autoaprendizaje de la Estadística: actualización de la aplicación 81
Aprendiendo Estadística investigando 84
Errores de estudiantes del grado en Educación Primaria y su tratamiento didáctico 90
Adaptación de las materias de Estadística Económica y Empresarial en los nuevos planes de estudio 92
Un método computacional para mejorar la interpretación de los intervalos de confianza para la proporción en Bachillerato y grados universitarios 94
Herramientas computacionales para el aprendizaje de las distribuciones Tipo Fase: Aplicación con datos reales de memorias resistivas 100
Propuesta didáctica basada en la metodología ABP para la asignatura de Investigación de Mercados del Grado en Administración y Dirección de Empresas 103
Software libre y representaciones gráficas – ggplot2 105
Enseñanza de la estadística en un entorno virtual para personas invidentes . . . 107
Idoneidad del software empleado en prácticas de Investigación Operativa 109
Estrategias de aprendizaje en las enseñanzas de Estadística y Matemáticas . . . 115

Lista de participantes **117**

Programa

Jueves 4 de abril:
Lugar: Sala de Conferencias del Instituto de Matemáticas de la Universidad de Granada.

- 09:00 - 9:30 Recepción de asistentes y entrega de material.
- 9:30 - 10:00 Inauguración de las jornadas.
- 10:00 - 12:00 Taller 1: **Introducción al lenguaje de programación Python** a cargo del profesor Miguel García Silvente del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.
- 12:00 - 12:30 Pausa - Café.
- 12:30 - 14:30 Taller 2: **Técnicas usuales de Machine Learning con Python** a cargo del profesor Miguel García Silvente del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.
- 14:30 - 16:00 Comida. Lugar: Cafetería de la Facultad de Ciencias.
- 17:30 - 20:00 Paseo matemático por Granada. Guía: Álvaro Martínez Sevilla. Punto de encuentro: Plaza de los Campos (Cuarto Real de Santo Domingo).
- 22:00 - 23.30 Visita nocturna Palacios Nazaríes de la Alhambra / Jardines y Palacio del Generalife. Punto de encuentro: Puerta de la Justicia.

Viernes 5 de abril
Lugar: Sala de Conferencias del Instituto de Matemáticas de la Universidad de Granada.

- 09:30 - 11:30 Charla-Coloquio: **¿Están condenados los p-valores al destierro?** a cargo del profesor Mathieu Kessler del Departamento de Matemática Aplicada y Estadística de la Universidad Politécnica de Cartagena.
- 11:30 - 12:00 Pausa - Café.
- 12:00 - 14:00 Charla-Coloquio: **Ideas y experiencias sobre la enseñanza de la Estadística** a cargo del profesor Pere Grima Cintas del Departamento de Estadística e I.O. de la Universitat Politécnica de Catalunya.
- 14:00 - 16:00 Comida. Lugar: Cafetería de la Facultad de Ciencias.
- 16:00 - 18:00 Sesión Póster - Café. Lugar: Hall Superior de la Facultad de Ciencias.
- 18:00 - 19:30 Reunión del grupo GENAEIO de la SEIO y clausura. Lugar: Sala de Claustros de la Facultad de Ciencias.
- 21:30 Cena de clausura en el Restaurante Asador de Castilla.

Talleres

Resumen

Python como lenguaje para programación en estadística

Taller 1: Introducción al lenguaje de programación Python
Taller 2: Técnicas usuales de Machine Learning con Python

Ponente invitado: **Miguel García Silvente**
m.garcia-silvente@decsai.ugr.es
Departamento de Ciencias de la Computación e Inteligencia Artificial,
Universidad de Granada.

Existe una gran cantidad de lenguajes de programación y su éxito depende de factores tales como antigüedad, facilidad de uso, potencia de las herramientas asociadas, facilidad de aprendizaje, etc. Los lenguajes que van surgiendo tienen que resultar sencillos y además deben poseer un conjunto de herramientas que faciliten una programación de propósisto general o para un campo determinado. Python cumple con esas dos características especialmente para el desarrollo de scripts dentro del ámbito científico. Para el caso concreto de la estadística incorpora varios potentes paquetes de manipulación de datos y para realizar machine learning.

Palabras clave: Programación científica, clasificación.
Clasificación AMS: 97U50.

Perfil del ponente Miguel García Silvente

Es Profesor Titular de Universidad del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada desde el año 2000. Desde entonces ha impartido numerosas asignaturas de programación en distintas titulaciones. Es Doctor en Informática y posee más de 20 publicaciones con alto índice de impacto JCR. Sus principales líneas de investigación son visión por ordenador (computer vision), reconocimiento de patrones, computación flexible e interacción hombre-robot. En la actualidad es el director del Secretariado Web de la Universidad de Granada.

Charlas-Coloquio

Resumen

¿Están condenados los p-valores al destierro?

Ponente invitado: **Mathieu Kessler**
mathieu.kessler@upct.es
Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena.

Los últimos años han sido testigos de un intenso debate en el mundo científico sobre el uso de tests de hipótesis nula, los p-valores y el famoso umbral de 0.05 para determinar la significancia de un efecto. Se enmarca en la más amplia crisis de reproductibilidad de la ciencia, que ha llegado a las portadas de los medios después de unos escándalos con enorme impacto. En esta crisis, algunos parecen echar toda la culpa a los métodos estadísticos y en particular a los p-valores, como una medida inadecuada y engañosa de la evidencia sobre efectos. ¿Están condenados los p-valores al destierro y la ignominia?
En esta charla, realizaremos un recorrido por este debate apasionante, abordando la dificultad de la interpretación de los p-valores, la crisis de la reproductibilidad en ciencia, la presión por publicar y el llamado "p-hacking", temas de enorme transcendencia en nuestra práctica y enseñanza de la estadística, especialmente en titulaciones aplicadas.

Palabras clave: P-valor, p-hacking, reproductibilidad, tests de hipótesis nula.
Clasificación AMS: 62A01, 62F03.

Perfil del ponente Mathieu Kessler

Es Doctor en Matemáticas por la Universidad de París 6, Francia, y Catedrático de Estadística e Investigación Operativa en el Departamento de Matemática Aplicada y Estadística de la Universidad Politécnica de Cartagena. Combina desde hace varios años investigación teórica en estadística para procesos estocásticos, en especial para ecuaciones diferencias estocásticas y procesos de Markov, con colaboraciones con científicos aplicados de distintas especialidades. Con ellos, ha publicado artículos de cristalografía, energías renovables, suelos, química, consumo eléctrico, cunicultura etc... Está interesado tanto por el enfoque bayesiano como por el frecuentista. Fruto de su intensa labor investigadora son numerosas publicaciones en revistas de alto impacto y ponencias en congresos relevantes en su área. Actualmente es Vicerrector de Tecnologías de la Información y la Comunicación de la Universidad de Murcia.

Resumen

Ideas y experiencias sobre la enseñanza de la estadística

Ponente invitado: **Pere Grima**
pere.grima@upc.edu
Departamento de Estadística e Investigación Operativa, Universitat Politècnica de Catalunya - BarcelonaTech.

La estadística está de moda. La facilidad de captación y manejo de grandes volúmenes de datos ofrece nuevas posibilidades en la obtención de información para la toma de decisiones. Ha aumentado la demanda de profesionales y la de estudiantes que quieren cursar nuestra disciplina. Y también han aparecido nuevos retos que se suman a los que ya teníamos. ¿Qué debemos enseñar? ¿Qué debemos hacer para concienciar a los estudiantes de la importancia y las posibilidades de la estadística? ¿Qué deben recordar nuestros estudiantes? ¿A qué dificultades nos enfrentamos?... No se pretende dar respuestas a todo, pero sí plantear algunas reflexiones y propuestas que faciliten un diálogo y cambio de impresiones enriquecedor para todos.

Palabras clave: Programación científica, clasificación.
Clasificación AMS: 97U50.

Perfil del ponente Pere Grima

Doctor Ingeniero Industrial. Profesor Titular en el Departamento de Estadística e Investigación Operativa de la Universidad Politécnica de Cataluña (UPC). Tras una primera etapa profesional en la industria, se incorporó a la UPC donde desarrolla sus tareas docentes en la Escuela Técnica Superior de Ingenieros Industriales y en la Facultad de Matemáticas y Estadística, donde ha sido Vicedecano Jefe de Estudios de Estadística.
Sus áreas de investigación son el diseño de experimentos y las aplicaciones de la estadística a la gestión de la calidad. Sobre estos temas ha dirigido tres tesis doctorales y ha codirigido otra y ha escrito artículos en revistas internacionales como *Q*uality Engineering, Quality and Reliability Engineering, Journal of Applied Statistics o Total Quality Management & Business Excellence.
Otro de sus temas de interés es la enseñanza de la estadística, tema sobre el que también la publicado varios artículos (revista *E*stadística Española, International Journal of Mathematical Education in Science and Technology y en *T*he American Statisitician entre otras). También ha publicado libros de divulgación como "La certeza absoluta y otras ficciones. Los secretos de la estadística", RBA, 2010 y sobre estadística industrial como "Industrial Statistics with Minitab" Wiley, 2012.

Sesión de pósteres

Visualización de datos con Tableau

Christian José Acal González[1], Juan Antonio Maldonado Jurado[1]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: chracal@ugr.es

Resumen

Hoy en día vivimos en la llamada *era de la información*, donde está disponible tanta cantidad de información que sin la debida atención, habilidad y recursos podríamos perdernos fácilmente en ella. Los gráficos proporcionan un excelente enfoque para la exploración de datos y son una herramienta esencial para presentar los resultados obtenidos al aplicar distintas técnicas estadísticas. Sin embargo, en discrepancia a los gráficos de apoyo estadístico que estamos acostumbrados a ver en todos los ámbitos y que presentan ciertas carencias, en los últimos años, y gracias al desarrollo computacional, ha surgido una nueva metodología para la representación visual de datos que hemos denominado Estadística Gráfica. Esta vertiente permite "jugar" con las visualizaciones de manera interactiva y es capaz de cambiar automáticamente conforme pudieran cambiar los datos. En la actualidad disponemos de una gran variedad de herramientas, donde Tableau es de las más pioneras y utilizadas en el mundo de la visualización de datos. Su funcionamiento es muy intuitivo y permite la creación de visualizaciones de alto nivel con tan solo arrastrar los datos.

Palabras claves: Tableau, gráficos dinámicos, visualización de datos, estadística

Introducción

Los gráficos proporcionan un excelente enfoque para la exploración de los datos y son una herramienta esencial para presentar los resultados obtenidos al aplicar distintas técnicas estadísticas. Aunque se han venido utilizando a lo largo de la historia y muchos son los autores que se atreven a dar algunas pautas de cómo debe desarrollarse un buen gráfico o qué debe contener una buena representación visual (1)(3)(13)(15), en la literatura no hay ninguna teoría aceptada universalmente sobre la misma. Muy lejos de la realidad, algunas personas piensan que realizar un buen gráfico es simplemente un problema de sentido común, mientras que otros creen que la preparación y desarrollo de los gráficos es una tarea de bajo nivel y que no es apropiado para la atención científica. Como dijo PlayFair "*un buen gráfico proporciona una explicación más adecuada de los hechos que una mera lista de datos o tablas. Sirve para simplificar lo complejo, permite al cerebro una mayor retención y es un instrumento visual de ayuda a hombres ocupados. Por último, los gráficos nos*

permiten ver relaciones aparentemente inexistentes entre variables, que suelen quedar ocultas entre la multitud de datos y cifras, de difícil comparación de otro modo".

Nuestros ancestros ya realizaban representaciones gráficas en forma de mapas pero no fue hasta el siglo XVIII cuando surgieron los primeros gráficos estadísticos modernos (3). Sin embargo, hasta el invento de los ordenadores y posteriormente el desarrollo de los programas estadísticos, todas estas visualizaciones se realizaban a mano con escuadra y cartabón, una tarea muy laboriosa y no exenta de sencillez. Todas estas representaciones, tanto las elaboradas a mano como aquellas obtenidas con los primeros programas estadísticos, aun siendo una herramienta de apoyo y eficaz para explorar los contenidos de un conjunto de datos, encontrar la estructura de los datos, comprobar suposiciones de modelos o comunicar los resultados obtenidos, etc. presentan una serie de carencias importantes para realizar una buena visualización de datos. Entre las carencias principales cabe destacar que dichos gráficos son estáticos, no aportan más información que la expuesta en el gráfico (por ejemplo, un gráfico de barras solo muestra la frecuencia de cada categoría) y cuando se alteran o se introducen nuevos datos, tenemos que repetir todo el proceso hasta alcanzar la representación gráfica modificada.

Hoy en día, estas características son de suma importancia ya que vivimos en la llamada *era de la información* donde está disponible tanta información que sin la debida atención, habilidad y recursos podemos ahogarnos fácilmente en ella (14). En los últimos años, principalmente gracias al desarrollo computacional, ha aparecido una nueva vertiente o metodología para la representación visual de los datos. Esta nueva metodología, la cual definiremos como *Estadística Gráfica*, nos permite "jugar" con el gráfico, ver cómo evoluciona el gráfico cuando se modifica o se introducen nuevas observaciones, transmitir una gran cantidad de información en un simple *Dashboard*, sugerir aplicar una u otra técnica mediante la búsqueda de patrones o tendencias, etc. En consecuencia, la Estadística Gráfica nos concede la posibilidad de ir eliminando toda esa morralla que no tiene ningún valor en nuestra base de datos hasta alcanzar el tesoro. Es por ello que dicha vertiente presenta un interés descomunal para el alumnado de cara a su futuro profesional (ya sea en el mundo empresarial o en el mundo de la investigación) y debe ser tratada como tal durante su formación académica.

Hoy en día, existe una gran diversidad de herramientas, tanto gratuitas como de pago, para realizar visualizaciones de datos (2)(5). Además, dichas técnicas pueden dividirse a su vez según si el analista posee conocimientos sobre programación o no, es decir, estas herramientas pueden enfocarse a personas que simplemente quieran realizar visualizaciones sin necesidad de programar o bien, pueden estar orientadas a analistas con un amplio conocimiento en el mundo de la programación que estén interesados en desarrollar programas algo más sofisticados. Una herramienta que está siendo muy pionera en la última década dentro de este mundo y de la que se dará pinceladas de ella en el presente trabajo, es el programa conocido como Tableau (12). Según sus creadores, es una herramienta que ayuda a las personas a ver y comprender sus datos de forma que puedan resolver los problemas de manera rápida, fácil, bella y útil.

Gráficos Estadísticos

Responder a la pregunta de cuándo surgieron los gráficos estadísticos no es nada sencillo. Hay indicios de que existían representaciones gráficas muchos años atrás en forma de mapas terrestres y celestes pero no fue hasta el siglo XVIII cuando los gráficos estadísticos modernos surgieron debido principalmente al avance y desarrollo de las distintas ramas científicas y sociales. En (7) recogen los principales acontecimientos y etapas más destacadas en la historia de los gráficos estadísticos.

Como ya se ha comentado en la introducción de este trabajo, un gráfico estadístico es una representación visual de una serie de datos estadísticos. Más genéricamente hablando, son una herramienta que es aplicada en diversas áreas para sus respectivos estudios (8), e incluso siguen generando interés desde el ámbito de la investigación (4). Entre sus principales objetivos destaca que permiten presentar la información de forma sencilla, clara y precisa, facilitan la comprensión de los datos y destacan las tendencias y diferencias de los mismos, exploran los contenidos de un conjunto de datos y permiten encontrar la estructura de ellos mismos (3)(6)(13).

Durante su formación, los alumnos ven y desarrollan distintos gráficos que sirven de apoyo, bien para facilitar la comprensión y análisis de los resultados

obtenidos a la hora de aplicar una cierta técnica, o bien para realizar un análisis descriptivo para comprobar cómo están distribuidos los datos que se tienen entre manos. A continuación, se muestra un recorrido ilustrativo por las

ASIGNATURA	GRÁFICO	TÉCNICA
Administración de empresas	•Gráfico piramidal	❖ División de puestos de trabajo.
Economía	•Diagrama de barras •Diagrama de sectores •Gráficos de puntos y líneas	❖ Demanda y consumo por año. Variación mensual del IPC, ... ❖ Ciudades con mayor porcentaje de paro, IPC de los países, ... ❖ Evolución de la oferta y demanda. Evolución del PIB. Evolución tasa del paro, ...
Cálculo de probabilidades I	•Diagrama de barras •Histograma •Diagrama de árbol •Gráfico de puntos y líneas • F. distribución y densidad	❖ Representación gráfica función masa prob. ❖ Simulación normal ❖ Combinaciones ❖ Simulación lanzamiento de una moneda
Cálculo de probabilidades II	•Diagrama de dispersión • F. distribución y densidad	❖ Regresión
Análisis matemático	•Gráficos geométricos	❖ Superficies en el espacio (integrales)
Informática	•Grafos	❖ Diagrama de flujos
Estadística descriptiva	•Diagrama de barras •Curva acumulativa •Histograma •Poligonal de frecuencias •Función de distribución •Diagrama de sectores •Pictograma •Diagrama de dispersión •Estereograma •Diagrama de caja	❖ Análisis descriptivo
Álgebra	•Gráficos de dispersión	

Tabla 1. Asignaturas de 1° curso del Grado en Estadística de la Universidad de Granada junto con los gráficos de apoyo en las distintas disciplinas.

ASIGNATURA	GRÁFICO	TÉCNICA
Inferencia estadística	•FD y fd •Gráfico de dispersión	❖ Contrastes de hipótesis e intervalos de confianza
Muestreo	•Mapa de calor •Gráfico de dispersión •Gráfico de puntos y líneas	❖ Muestreo por conglomerados
Investigación operativa	•Grafos	❖ Problema Ruta más corta, Modelo de reposición de equipo, Teoría de colas, ...
Métodos numéricos	•Gráficos de puntos y líneas	❖ Aproximaciones, ajustes y splines
Análisis de datos	•Gráfico de dispersión •Dendograma	❖ MDS, ACP, ACS ❖ Clúster
Estadística computacional I	•Diagrama de barras •Histograma •Poligonal de frecuencias • F. distribución y densidad •Diagrama de sectores •Diagrama de dispersión •Diagrama de caja •QQ-plot	❖ Ejercicios puntuales ❖ Simulación ❖ Simulación ❖ Ejercicios puntuales ❖ Regresión ❖ Comprobar puntos anómalos ❖ Comprobar normalidad
Modelos lineales	•Diagrama de dispersión •Gráficos de residuos •QQ-plot	❖ Regresión lineal univariante simple y múltiple

Tabla 2. Asignaturas de 2° curso del Grado en Estadística de la Universidad de Granada junto con los gráficos de apoyo en las distintas disciplinas.

ASIGNATURA	GRÁFICO	TÉCNICA
Estadística computacional II	•Mapa de calor •Los mismos gráficos de ECI	❖ Paquete ggplot de R
Procesos estocásticos	•Curva acumulativa •Gráfico de puntos y líneas •Grafos •Histograma	❖ Procesos de Poisson ❖ Simulación en trayectorias muestrales ❖ Procesos de Poisson, cadenas de Markov (transiciones entre estados)
Series temporales	•Diagrama de barras •Gráficos de puntos y líneas •QQ-plot	❖ Residuos y correlaciones de las series ❖ La serie en sí ❖ Comprobar normalidad
Diseño de experimentos	•Histograma •Función de densidad •Diagrama de dispersión •Gráfico de puntos y líneas •Caja de bigotes •QQ-plot	❖ Comprobar normalidad del modelo ❖ Contrastes ❖ Comprobar independencia entre los residuos ❖ Comprobar normalidad del modelo
Inteligencia artificial	•Grafos	❖ Teoría de grafos
Análisis de valores extremos	•Histograma •F. distribución y densidad •Diagrama de dispersión •QQ-plot	❖ Distribución de sucesos 'raros' ❖ Distribuciones Generalizadas ❖ Ejercicios puntuales ❖ Comprobar normalidad del modelo
Análisis de riesgos	• F. distribución y densidad	
Áreas de aplicación	•Biplot •Gráfico de dispersión •Dendograma	❖ Análisis de componentes principales (ACP) ❖ MDS, ACP, Análisis de correspondencias Simple ❖ Clúster
Diseño de encuestas	•Gráficos de puntos	❖ MDS

Tabla 3. Asignaturas de 3° curso del Grado en Estadística de la Universidad de Granada junto con los gráficos de apoyo en las distintas disciplinas.

ASIGNATURA	GRÁFICO	TÉCNICA
Bioestadística	•Diagrama de bosque •Gráfico de puntos y líneas	❖ Meta-Análisis ❖ Explicación de los test exactos (Barnard,...) Curvas de supervivencia, Modelo de Cox.
Fiabilidad	•Gráfico de puntos y líneas •Grafos	❖ Función de fiabilidad, función de razón de fallo ❖ Transiciones entre estados, tipos de sistemas
Análisis exploratorio	•Caja de bigotes •Bagplot •Histograma y f. densidad •QQ-plot •Gráfico de puntos y líneas	❖ Presencia de puntos anómalos y asímetria ❖ Presencia de puntos anómalos en multivariante ❖ Ver distribución de la muestra ❖ Comprobar normalidad ❖ Regresión
Técnicas básicas multivariante	•Gráfico de puntos y líneas	❖ Análisis de componentes principales y análisis factorial general.
Minería de datos	• Gráfico de puntos y líneas •Gráficos SSE (suma distancias al cuadrado) • Gráficos 3D •Dendograma •Diagrama de dispersión •Diagrama de sectores y árbol de decisión	❖ k-means ❖ Determinar el nº de dimensiones óptimo ❖ Observar la densidad de los datos ❖ Clúster ❖ Regresión y DBSCAN ❖ Técnica árbol de decisión
Técnicas avanzadas multivariante	•Gráficos de puntos •QQplot •Gráfico de dispersión	❖ Gráficos de residuos para el análisis del modelo en Regresión lineal múltiple multivariante ❖ Comprobar normalidad ❖ Análisis discriminante y Análisis de Correlación Canónica.
Modelización de datos categóricos	•Grafos	❖ Modelos gráficos/Modelos log-lineales.

Tabla 4. Asignaturas de 4° curso del Grado en Estadística de la Universidad de Granada junto con los gráficos de apoyo en las distintas disciplinas.

Normalmente, todos estos gráficos suelen ser representados por los estudiantes mediante los programas estadísticos SPSS o R, y en menor medida, mediante StatGraphics. Alguna de la bibliografía fundamental en la que se basa el alumnado es la siguiente (9)(10)(11). Sin embargo, todos estos gráficos estadísticos presentan una serie de carencias importantes para realizar una buena visualización de datos, como por ejemplo, que son estáticos, no aportan gran información y no cambian de manera automática cuando se introduce nuevas observaciones. En consecuencia, y en parte gracias al boom del *Big Data*, este tipo de visualización de datos se ha quedado obsoleta en el mundo actual y es por ello, por lo que los nuevos analistas se ven obligados a adaptarse a esta nueva situación y plantearse el enfoque de la visualización de datos desde una nueva perspectiva.

Visualización de datos. Estadística Gráfica

La visualización de datos es una disciplina que con el fin de facilitar la comprensión transforma las relaciones numéricas de los datos en impactos visuales, ya que para tomar decisiones es imprescindible conocer bien la situación. La visualización es una tendencia en los últimos años debido a que vivimos rodeados de datos. Cada individuo genera diariamente una inmensa cantidad de datos (tweets, mensajes de móviles, compras online, etc.). El análisis de los datos nos ayuda a sacar conclusiones, pero lo más importante es ser capaces de 'ver' estos datos. Su representación nos permite entender, detectar patrones o descubrir valores atípicos entre otras muchas cosas más. La visualización de datos no es una disciplina nueva, pero con el auge del marketing y la analítica digital en los últimos años, ha ganado gran visibilidad dentro de los departamentos digitales de las empresas, las agencias y consultoras que giran a su alrededor. Como se detalla en (16), este término surgió a raíz del nacimiento de la web 2.0, donde una gran cantidad de datos origina que la búsqueda e interpretación de los mismos sea una tarea bastante difícil y tediosa, hecho que provoca que surja la necesidad de encontrar una metodología que permita la comprensión y asimilación de la información de una manera más sencilla. Esta vertiente se ha convertido en una parte crítica de los análisis de negocios y un elemento fundamental para las comunicaciones empresariales y otros ámbitos.

Como ya se ha dicho, la visualización es el mejor aliado en el análisis de datos

por aportar una representación gráfica que amplifica el conocimiento de forma inmediata. Esta representación permite entender grandes volúmenes de información de un solo vistazo, para poder explorar y comprender su significado en cuestión de segundos. Debido a los problemas planteados en el apartado anterior, y gracias al desarrollo computacional, han aparecido nuevas formas de presentación de resultados que contienen una nueva metodología de visualización de datos, la cual hemos denominado *Estadística Gráfica*. Entre sus principales características destacan:

- Son dinámicos y permite al analista interactuar con el gráfico.
- Proporcionan bastante información con descripciones emergentes. Esto es, por ejemplo, que al situar el cursor sobre una barra de frecuencias detalla información extra de esa misma variable o incluso de otras.
- Las visualizaciones son más vistosas y con múltiples recursos que permiten sugerir aplicar una u otra técnica.
- Cambian automáticamente cuando se introducen nuevos datos, de manera que no haya que repetir todo el proceso hasta alcanzar la representación gráfica nueva.
- Permiten tomar decisiones en un plazo relativamente corto e incluso, a tiempo real, aspecto muy a tener en cuenta en el mundo empresarial donde una buena decisión puede suponer una ganancia importante.

En este ámbito se necesita una serie de virtudes, no sólo estadísticas, para poder llegar a alcanzar la meta. Puede ser por ello por lo que actualmente no existe un perfil determinado de visualizadores de datos, de hecho, muy pocas empresas dispone de un equipo encargado para dicha tarea, recayendo todo este trabajo sobre personas con formaciones muy diversas (psicólogos, estadísticos, ingenieros, periodistas, etc.). Si bien es cierto que es una rama altamente relacionada con la estadística, hay que ser conscientes de que no hace falta ser un gran estadístico o matemático para tener un gran dominio de esta técnica; simplemente uno debe tener una base de diseñador gráfico para conocer el lenguaje visual y poder explotar su potencial, ser un buen comunicador de manera que sea efectiva la visualización, ser técnico en el sentido de que sea capaz de darle una forma real a todo, tener conocimientos en el mundo del negocio, tener espíritu analítico para buscar las respuestas

necesarias, y disponer de un toque de artista para armonizar todos los elementos. No obstante, se debería concienciar que un estadístico/matemático es la persona con más capacidades para elaborar este trabajo y es quién debería realizarlo. De hecho, no tener conocimientos acerca de los datos puede dar lugar a una mala representación gráfica que engañe, distorsione y altera la realidad. Una representación puede ser mala porque los datos recogidos no son válidos para el estudio debido a que la resolución formal no es la idónea o por la forma en la que se combina todo. En consecuencia, hay que hacer una reflexión profunda y no perder la percepción sobre este tema.

Actualmente podemos encontrar por Internet diversas herramientas que facilitan la visualización de datos de manera mucho más clara y efectiva que si nos limitáramos a representarlos de una manera tradicional. Como es lógico, estas herramientas poseen sus propias características y funciones, pero por lo general, todas ellas están basadas en la idea principal de importar datos, elegir entre las distintas opciones de visualización y compartirlas posteriormente en futuras publicaciones o trabajos. Por otro lado, la mayoría de ellas permiten al usuario poder utilizarlas sin tener excesivo conocimiento de programación (o incluso ninguno en alguno de ellos) y que los gráficos que muestren ya estén optimizados para distintas plataformas con el fin de poder ser compartidos. Además, ofrecen versiones gratuitas y comunidades de usuarios, como pueden ser los foros, para compartir consejos y buenas prácticas a la hora de usarlas (2).

El cuadro que figura a continuación muestra un estudio comparativo entre los distintos programas de visualización de datos más utilizados en la práctica (columnas) y algunas características principales que contiene cada uno de los programas (filas). Notar que las características que aparecen son las principales o esenciales que debe poseer un programa de visualización de datos. Internamente cada programa atesora una serie de características propias que no se va a entrar a detallar. Si el usuario está interesado puede indagar en el tema dirigiéndose a las webs de cada uno de los programas. En el cuadro se puede observar una serie de colores: el color verde representa que el programa contiene la correspondiente característica; el color rojo representa que el programa contiene dicha característica pero no es el más adecuado ya sea por interfaz, rapidez de computación, capacidad de lectura de datos, etc.; el color azul representa que el programa contiene esa característica y además

es el programa ideal para llevar a cabo esa función; y la ausencia de color significa que el programa no posee la característica determinada. La tabla es muy útil principalmente por dos razones: por un lado, puede darse el caso de que el usuario que se dedique a este mundo esté interesado en adquirir un programa, pero no sabe por cuál decantarse. Pues bien, esta persona puede dirigirse a este cuadro por columnas y de manera lógica, se inclinará por aquel programa que contenga más características; o aquel que sea gratuito en caso de no querer gastarse dinero o bien, por aquel que no requiera tener unos conocimientos previos de programación en caso de que este sea un problema para el analista. Por otro lado, puede darse la situación de que el analista tenga un trabajo muy concreto determinado y esté buscando un programa que le resuelva dicho problema. Ante esta situación, el usuario se dirigirá a este cuadro por filas, buscaría la característica que desee y elegirá aquel programa que contenga dicha característica. En dicha tabla, destacar que el más completo es Tableau, el cual comentaremos en el siguiente apartado.

PROGRAMAS / CARACTERÍSTICAS	EXCEL	TAGXEDO	QLIKVIEW	MANY EYES	GOOGLE FUSION TABLE	GEPHI	DATA DRIVEN	NODEBOX	WEAVE	QUADIGRAM	VISUAL.LY	TABLEAU
Gratuito												
Tipo Web												
Código abierto												
Aplicación tlf												
Interactiva (dinámica)												
Nube de palabras												
Dashboards												
Compartir visualizaciones												
Gran variedad de visualizaciones												
Cartografía y mapeo												
Análisis de redes y grafos												
Soporta grandes bases de datos												
Orientación gráfica												
Predicciones y tendencias												
Tablas dinámicas												
Mapas 3D												
Descripción emergente												
Inserción de texto												
Filtrado rápido												
Creación de parámetros												

Figura 1. Cuadro comparativo entre los distintos programas de visualización de datos más utilizados en la práctica (columnas) y algunas características principales que contiene cada uno de los mismos (filas).

Tableau

Tableau es una de las herramientas líder en el mundo de la visualización. Se trata de una de las maneras más rápidas y sencillas que existen para crear visualizaciones de datos y obtener información. Su funcionamiento es muy intuitivo, permitiendo la creación de visualizaciones de alto nivel, informes y tableros de control con tan sólo arrastrar los datos y ver los cambios en tiempo real mientras éstos se van realizando. Tableau nos permite mostrar diferentes representaciones de los datos en un mismo dashboard y nos da la posibilidad de añadir información extra a través de la inserción de documentos o páginas

web que ampliarán la comprensión de los datos. El programa es muy fácil de manejar, lo que lo hace accesible a todo el mundo sin necesidad de dominio de programación. Como se dijo en la introducción de este informe, Tableau es una herramienta que está siendo muy pionera en la última década dentro del mundo de la visualización. De hecho, grandes compañías y empresas de cualquier sector (Audi, Barclays, Bank of America, Tune, Skype, Alcatel, Coca-cola, Universidades de EEUU, Gobierno de Canadá, Google, Microsoft, Yahoo y muchas más) utilizan esta técnica para elaborar informes y sacar conclusiones acerca de sus productos.

A modo de ejemplo, los autores del presente trabajo analizaron mediante este programa los datos correspondientes a la encuesta realizada al alumnado de la Facultad de Ciencias de Granada a lo largo de cuatro años, concretamente durante los cursos académicos de 2012-2013 hasta 2015-2016, por parte de los alumnos que cursaban la asignatura de 'Diseño de Encuestas' del Grado en Estadística en el respectivo año. El objetivo que se perseguía con esta encuesta era conocer la opinión del alumnado sobre varios aspectos relativos a sus estudios universitarios y, más concretamente, a la Facultad de Ciencias de Granada. Para conocer esta opinión se pidió a una muestra de dicho alumnado que cumplimentara un cuestionario con preguntas relacionadas con el ámbito académico, extra-académico y otros aspectos considerados de interés en la vida universitaria. Esta recogida de información permitía obtener un perfil del alumnado, así como tener una visión más precisa de qué opina éste de su Facultad, de sus compañeros, del profesorado, de las instalaciones y de los servicios ofertados por la Universidad que no están directamente relacionados con el estudio. También se estudiaron algunos datos extraídos de la página oficial de LaLiga acerca de la temporada 2016 y 2017. Algunas de las figuras que se extrajeron en estos análisis pueden verse a continuación. El objetivo es mostrar distintas representaciones visuales que se han obtenido en Tableau utilizando los datos recogidos, por lo que no se entrará a evaluar los resultados obtenidos ni cómo se han logrado dichas visualizaciones puesto que no es el fin de este trabajo. Simplemente, la meta que se persigue es mostrar al lector algunas de las infinidades de recursos que posee esta herramienta utilizando unos datos reales como los mencionados.

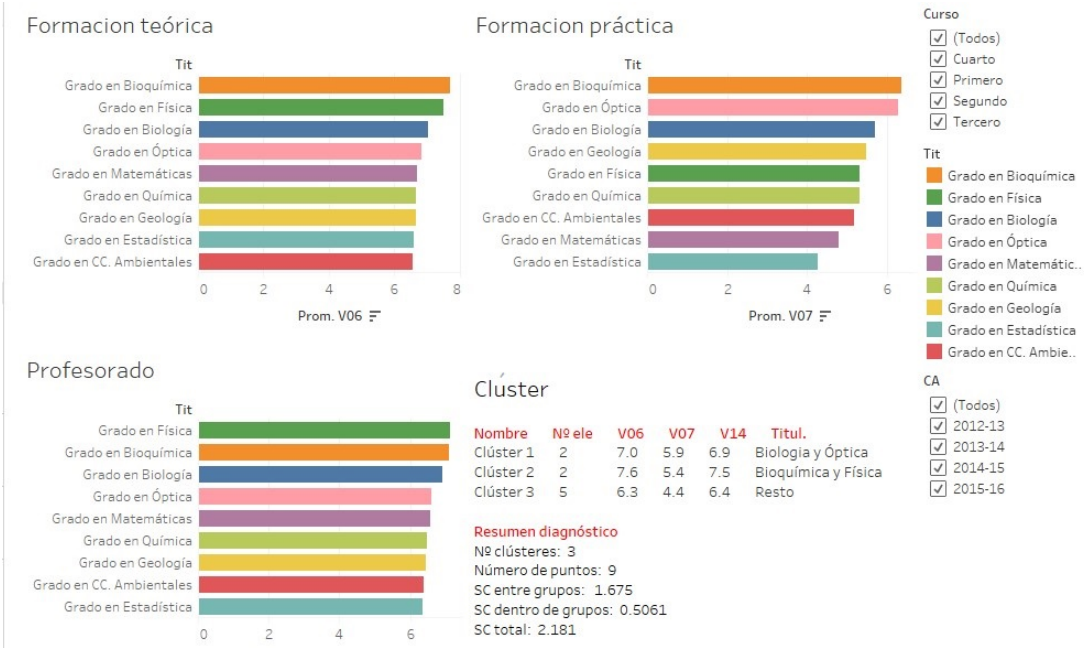


Figura 2. Resultados tras aplicar la técnica clúster con el fin de comprobar cómo se agrupan las distintas titulaciones tomando como referencia las variables que pertenecen a cuestiones sobre la titulación (se ha tomado el promedio). También se muestra algunos diagramas de barras de las calificaciones en ciertas preguntas. A la derecha, se puede filtrar por curso, año académico o titulación.

Distribucion muestral

	Curso / Sexo							
	Primero		Segundo		Tercero		Cuarto	
Tit	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Grado en Biología	42	56	46	57	54	64	49	64
Grado en Bioquímica	11	16	7	15	7	14	7	12
Grado en CC. Ambientales	36	28	30	18	32	17	35	28
Grado en Estadística	15	9	8	8	7	5	11	6
Grado en Física	41	13	28	9	27	8	24	9
Grado en Geología	19	11	14	8	13	8	16	7
Grado en Matemáticas	28	21	21	16	16	16	22	21
Grado en Óptica	19	36	15	30	16	36	16	33
Grado en Química	23	19	17	15	14	19	17	17

CNT(Sexo (copia))
5 64

Figura 3. Distribución muestral por titulaciones, sexo y curso. Los colores nos ayudan a identificar qué titulaciones tienen mayor peso en el estudio.

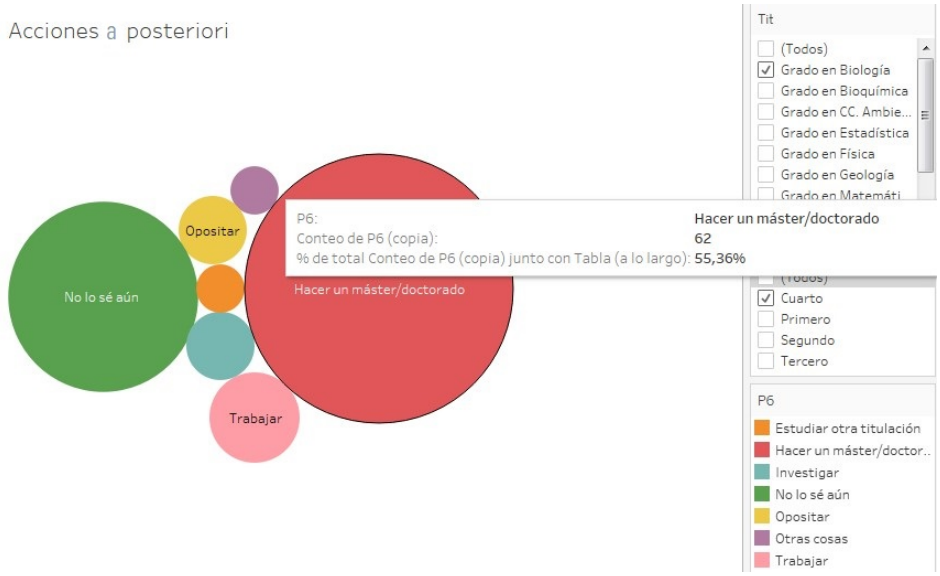


Figura 4. Gráfico de burbujas para detectar la intención del alumnado de Biología de cuarto curso (son los alumnos filtrados como se puede ver a la derecha del gráfico) una vez que terminen sus estudios. Resalta una ventana emergente proporcionando información extra de aquellos alumnos que quieren hacer un Máster/Doctorado.

Figura 5. A la izquierda aparecen los escudos de todos los equipos así como su puesto en la clasificación. A la derecha figura toda la información recogida donde cada variable está desglosada según sus categorías. Si se sitúa el cursor encima de alguna de las variables, se iluminan aquellos escudos de los equipos que cumplan dicha cualidad.

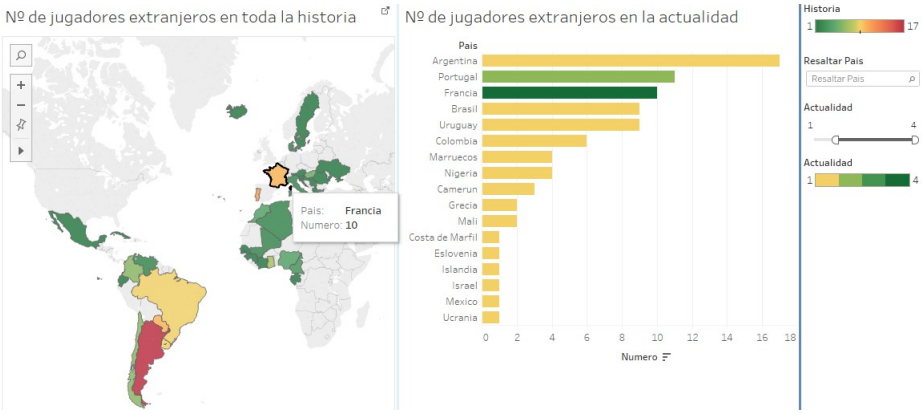


Figura 6. A la izquierda, mapeo sobre el n^o de jugadores extranjeros que han jugado en el Granada C.F. durante toda la historia en función de la nacionalidad. A la derecha, visualización del n^o de jugadores extranjeros en la actualidad filtrada de manera que sólo aparezcan aquellas nacionalidades de jugadores que han pertenecido dicho año al club (véase el filtro de la flecha Actualidad). Las barras representan el número de jugadores extranjeros del país correspondiente que han pertenecido al club durante toda su historia, mientras que el color simboliza el número de jugadores de cada país que han jugado dicho año en el Granada.

En dichas figuras queda patente la gran cantidad de información disponible que aparece en un mismo dashboard y la infinidad de recursos con los que cuenta para poder filtrar según las modalidades de la variable que se desee, cambiando de manera automática en función de las opciones que se seleccione (al igual que si se introducen nuevas observaciones). Además, Tableau cuenta con un módulo para crear historias (son como una especie de presentación PowerPoint pero desde Tableau) que permite explicar a los miembros del equipo por ejemplo, el proceso que se ha llevado a cabo para descubrir los patrones, tendencias, conclusiones, etc. y otro módulo que posibilita la opción de compartir vistas y datos con cualquiera que tenga acceso a internet, tanto de manera pública como privada.

Las comparaciones de estas visualizaciones frente a los gráficos estadísticos tradicionales son odiosas y ponen de manifiesto la gran ventaja que supone hoy en día utilizar este tipo de herramientas en el análisis de datos.

Conclusiones

En consecuencia de todo lo comentado en los apartados anteriores, parece evidente concluir que la visualización de datos es más importante de lo que

realmente parece. No se trata de una mera disciplina que se encargue únicamente de 'colorear' unos gráficos o de hacer representaciones llamativas y atractivas que llamen la atención del usuario. Esta metodología va mucho más allá y, aunque cabe recalcar que no es una herramienta sustituta de las técnicas analíticas, sí es un gran apoyo para ayudar al análisis matemático y estadístico que se fundamenta en teoremas, lemas y corolarios. Por otro lado, también es incuestionable dudar de la potencia que posee la Estadística Gráfica. Debido al boom de datos que generamos las personas hoy en día, es una herramienta necesaria e imprescindible para todo aquel que quiera extraer información y conocimiento acerca de los datos. A lo largo de este trabajo se ha intentado hacer hincapié en que los gráficos estadísticos que estamos acostumbrados a ver y a utilizar, se han quedado obsoletos tras la aparición del Big Data, ya que carecen de cierta información y rapidez conforme pudieran modificarse los datos. Actualmente, a las empresas y demás organismos les están llegando continuamente información y datos acerca de su negocio. Tener conocimiento de cómo evolucionan los datos en tiempo real permite a estas instituciones tomar decisiones que puede suponer una ganancia inmensa y considerable para ellos. Finalmente, notar que el principal objetivo del presente trabajo es hacer una pequeña reflexión sobre este tema y recalcar que la visualización y el tratamiento de datos es un tema que cada día va cogiendo más fuerza en nuestra sociedad y que desgraciadamente, no se le dedica el tiempo suficiente que se requiere durante la formación del alumnado.

Referencias

[1] I. Alcalde (2019) `https://ignasialcalde.es/4-pasos-clave-para-visualizar-datos/` [10 octubre 2019]

[2] Banco Bilbao Vizcaya Argentaria (2019) `https://www.bbva.com/es/10-herramientas-visualizacion-datos/` [07 octubre 2019]

[3] C.H. Chen, W. Hardle y A. Unwin (Editors) (2008) *Handbook of Data Visualization*. Berlín, Germany, Springer.

[4] L. Denby y C. Mallows (2009) Variations on the Histogram. *Journal of Computational and Graphical Statistics*, **18**(1), pp. 21-31.

[5] Doctor Metrics (2019) `http://www.doctormetrics.com/2012/08/29/herramientas-visualizacion-datos/#.WFcdZX1bj-l` [07 octubre 2019]

[6] A. Gelman y A. Unwin (2013) Infovis and Statistical Graphics: Different goals, Different Looks. *Journal of Computational and Graphical Statistics*, **22**(1), pp. 2-28.

[7] Instituto Nacional de Estadística `https://www.ine.es/expo_graficos2010/expogra_inicio.htm` [20 marzo 2017]

[8] W.G. Jacoby y S. K. Schneider (2010) *Graphical displays for political science journal articles.* In Visions in Methodology Conference, Iowa City, IA, March.

[9] K.J. Keen (2010) *Graphics for Statistics and Data Analysis with R*. Chapman & Hall/CRC.

[10] K. McCormick y J. Salcedo (2017) *SPSS statistics for data analysis and visualization.* John Wiley & Sons.

[11] D. Sarkar (2008) *Lattice: Multivariate Data Visualization with R*. New York, Springer Science & Business Media.

[12] Tableau (2017) `https://www.tableau.com/es-es` [21 febrero 2017]

[13] E.R. Tufte (2001) *The Visual Display of Quantitative Information*. Second ed., Cheshire, CT: Graphics Press.

[14] A. Unwin, M. Theus y H Hofmann (2006) *Graphics of Large Datasets: Visualizing a Million*. New York, Springer.

[15] H. Wickham (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York, Springer-Verlag.

[16] Wikipedia `https://es.wikipedia.org/wiki/Visualizaci%C3%B3n_de_datos` [11 octubre 2019]

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

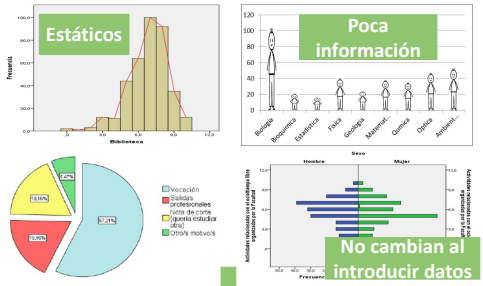
VISUALIZACIÓN DE DATOS CON TABLEAU

UNIVERSIDAD DE GRANADA

Christian J. Acal (chracal@ugr.es)
Juan Antonio Maldonado (jamaldo@ugr.es)

Dpto. de Estadística e Investigación Operativa
Universidad de Granada

Hoy en día vivimos en la llamada *Era de la información*, donde está disponible tanta cantidad de información que sin la debida atención, habilidad y recursos podríamos perdernos fácilmente en ella. Los gráficos proporcionan un excelente enfoque para la exploración de datos y son una herramienta esencial para presentar los resultados obtenidos al aplicar distintas técnicas estadísticas. Sin embargo, en discrepancia a los gráficos de apoyo estadístico que estamos acostumbrados a ver en todos los ámbitos y que presentan ciertas carencias, en los últimos años, y gracias al desarrollo computacional, ha surgido una nueva metodología para la representación visual de datos que hemos denominado *Estadística Gráfica*. Esta vertiente permite 'jugar' con las visualizaciones de manera interactiva y es capaz de cambiar automáticamente conforme pudieran cambiar los datos. En la actualidad disponemos de una gran variedad de herramientas, donde Tableau es de las más pioneras y utilizadas en el mundo de la visualización de datos. Su funcionamiento es muy intuitivo y permite la creación de visualizaciones de alto nivel con tan solo arrastrar los datos.



Dinámicos

Más vistosas

Con múltiples recursos

'Estadística Gráfica'

Toma de decisiones en un plazo corto

Amplia información con ventanas emergentes

Cambian solos al introducir nuevos datos

TABLEAU: El presente del análisis

PROGRAMAS / CARACTERÍSTICAS	Excel	Tagxedo	Qlikview	Many Eyes	Google Table Fusión	Gephi	Data Driven	Nodebox	Weave	Quadigram	Visual.ly	Tableau
Gratuito												
Tipo Web												
Código abierto												
Aplicación tlf												
Interactiva (dinámica)												
Nube de palabras												
Dashboards												
Compartir visualizaciones												
Gran variedad de visualizaciones												
Cartografía y mapeo												
Análisis de redes y grafos												
Soporta grandes bases de datos												
Orientación gráfica												
Predicciones y tendencias												
Tablas dinámicas												
Mapas 3D												
Descripción emergente												
Inserción de texto												
Filtrado rápido												
Creación de parámetros												

CONCLUSIONES

Es incuestionable dudar de la potencia que posee la *Estadística Gráfica*, siendo una herramienta necesaria e imprescindible para todo aquel que quiera extraer información y conocimiento acerca de los datos. Los gráficos estadísticos que estamos acostumbrados a ver y a utilizar, se han quedado obsoletos tras la aparición del Big Data, ya que carecen de cierta información y rapidez conforme pudieran modificarse los datos. Actualmente, a las empresas y demás organismos les están llegando continuamente información y datos acerca de su negocio. Tener conocimiento de cómo evolucionan los datos en tiempo real permite a estas instituciones tomar decisiones que puede suponer una ganancia inmensa y considerable para ellos. Esto hace esencial y de suma importancia conocer y saber cómo cambian los datos al instante y para ello se requiere de una herramienta potente y eficaz como puede ser Tableau.

Aprendiendo a modelizar datos de conteo a nivel de área con dependencia espacial y espacio-temporal mediante la aplicación web SSTCDapp

Aritz Adin[1], Tomás Goicoa[1], M.D. Ugarte[1]

[1]Departamento de Estadística, Informática y Matemáticas, Universidad Pública de Navarra

Contacto: tomas.goicoa@unavarra.es

Resumen

Los modelos espaciales y espacio-temporales para datos de conteo a nivel de área presentan una serie de dificultades difíciles de solucionar para alumnos y profesionales sin una amplia formación estadística. Uno de los problemas más importantes a la hora enseñar y de ajustar estos modelos tiene que ver con la identificación de los términos del modelo, que debe resolverse mediante el uso de restricciones o empleando una reparametrización del modelo, algo de evidente complejidad para alumnos con escasa formación en estadística matemática. En este trabajo proponemos el uso de la aplicación web SSTCDapp (accesible desde `https://emi-sstcdapp.unavarra.es/Login/`), desarrollada en shiny por nuestro grupo de investigación, para la docencia en las asignaturas "Análisis y Predicción Estadística de Datos Espaciales" y "Análisis y Predicción Estadística de Datos Espacio-Temporales" en el Máster Universitario de Información Geográfica y Teledetección de la Universidad Pública de Navarra. La aplicación permite el ajuste de modelos espaciales y espacio-temporales desde un punto de vista completamente Bayesiano. El ajuste de los modelos se realiza en un potente servidor remoto aunque desde el sitio web de la aplicación también puede obtenerse una versión de escritorio para su uso en local.

Palabras claves: aplicación web, datos de conteo, INLA, modelos espacio-temporales

Aprendiendo a modelizar datos de conteo a nivel de área con dependencia espacial y espacio-temporal mediante la aplicación web SSTCDapp

Adin, A.[1,2], Goicoa, T.[1,2] and Ugarte, MD.[1,2]

[1] Departamento de Estadística, Informática y Matemáticas. Universidad Pública de Navarra.
[2] InaMat. Universidad Pública de Navarra.

aritz.adin@unavarra.es; tomas.goicoa@unavarra.es; lola@unavarra.es

Introducción

Los modelos espaciales y espacio-temporales para datos de conteo a nivel de área presentan una serie de dificultades para alumnos y profesionales sin una amplia formación estadística. Uno de los problemas más importantes a la hora enseñar y de ajustar estos modelos tiene que ver con la identificación de los términos del modelo, que debe resolverse mediante el uso de restricciones o empleando una reparametrización del modelo (Goicoa et al., 2018), algo de evidente complejidad para alumnos con escasa formación en estadística matemática.

En este trabajo proponemos el uso de la aplicación web SSTCDapp (Adin et al., 2019) para la docencia en las asignaturas *"Análisis y Predicción Estadística de Datos Espaciales"* y *"Análisis y Predicción Estadística de Datos Espacio-Temporales"* en el **Máster Universitario de Información Geográfica y Teledetección** de la Universidad Pública de Navarra.

La aplicación web SSTCDapp

SSTCDapp es una aplicación web interactiva para el análisis espacial y espacio-temporal de datos de conteo, diseñada para:

1) Realizar análisis descriptivos en el espacio y el tiempo de riesgos o tasas de mortalidad/incidencia.
2) Ajustar un conjunto de modelos jerárquicos espacio-temporales ampliamente utilizados en disease mapping (Ugarte et al., 2014).

Características principales:

- Ha sido desarrollada en "shiny", una herramienta para crear aplicaciones web mediante el software estadístico `R`.
- Permite ajustar los modelos desde un punto de vista completamente Bayesiano utilizando la técnica INLA (integrated nested Laplace approximation) a través del paquete `R-INLA` (Rue et al., 2009).
- Accesible desde **`https://emi-sstcdapp.unavarra.es/Login/`** y no requiere ninguna instalación por parte del usuario, ya que el ajuste de los modelos se realiza en un potente servidor remoto. Desde el sitio web de la aplicación también puede obtenerse una versión de escritorio para su uso local.

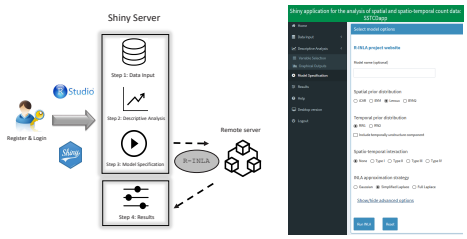


Figura 1: Diagrama de trabajo y captura de pantalla de la aplicación SSTCDapp.

Análisis espacio-temporal de mortalidad por cáncer de mama en las provincias de España

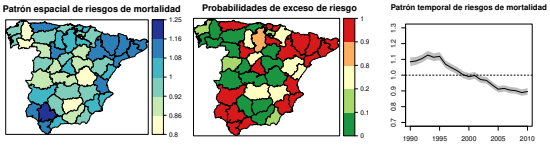


Figura 2: Patrones espaciales y temporales de los riesgos relativos de mortalidad por cáncer de mama en las provincias de España durante el periodo 1990-2010.

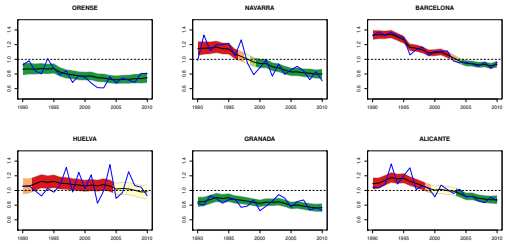


Figura 3: Evoluciones temporales del riesgo relativo de mortalidad por cáncer de mama en seis provincias de España e intervalos de credibilidad al 95%. Los colores utilizados en las bandas están asociados a las probabilidades de que los riesgos sean superiores a uno.

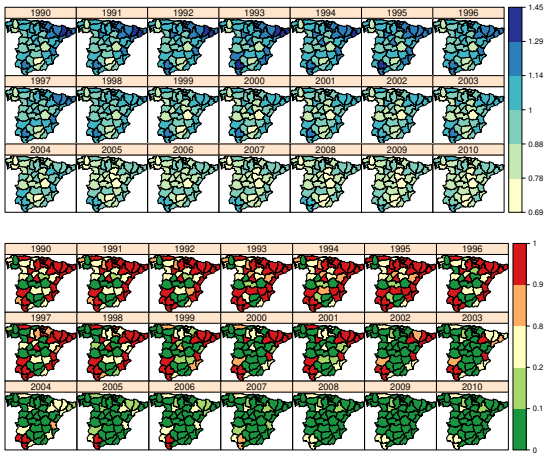


Figura 4: Mapas con las medias a posteriori de los riesgos relativos de mortalidad por cáncer de mama en las provincias de España (parte superior) y probabilidades de que los riesgos sean superiores a uno (parte inferior).

Referencias

Adin, A., Goicoa, T., and Ugarte, M. D. (2019). Online relative risks/rates estimation in spatial and spatio-temporal disease mapping. *Computer Methods and Programs in Biomedicine*, 172:103–116.

Goicoa, T., Adin, A., Ugarte, M. D., and Hodges, J. S. (2018). In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stochastic Environmental Research and Risk Assessment*, 32(3):749–770.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Ugarte, M. D., Adin, A., Goicoa, T., and Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research*, 23(6):507–530.

Desarrolladores y financiación

La aplicación SSTCDapp ha sido desarrollada por el grupo de investigación de Estadística Espacial de la Universidad Pública de Navarra, en el marco de investigación de los siguientes proyectos:

- Ministerio de Economía y Competitividad de España (proyecto MTM2014-51992-R).
- Departamento de Salud del Gobierno de Navarra (proyecto 113, Res.2186/2014).
- AEI/FEDER, UE (proyecto MTM2017-82553-R).

Para incidencias y soporte, utilice la siguiente cuenta de correo:
`app.estadistica@unavarra.es`

R Markdown como herramienta para la elaboración de materiales docentes

M. Carmen Aguilera-Morillo[1], Álvaro Méndez Civieta[1], Juan Carlos Laria[1]

[1]Departamento de Estadística, Universidad Carlos III de Madrid

Contacto: maguiler@est-econ.uc3m.es

Resumen

R Markdown es una herramienta que puede resultar de gran utilidad para los docentes, ya que permite elaborar presentaciones, guiones de prácticas, materiales didácticos en general, reproducibles y dinámicos con R. En este trabajo se mostrará su uso para la elaboración de materiales docentes de la Asignatura Functional Data Analysis del Máster en Statistics for Data Science de la Universidad Carlos III de Madrid. Lo más interesante de R Markdown es que permite combinar texto y códigos procedentes de diversos lenguajes de programación, como son R, Python y SQL. Además, ofrece diferentes formatos de salida estáticos y dinámicos, tales como HTML, PDF, MS Word, Beamer, MS PowerPoint, diapositivas HTML5, libros, artículos científicos, páginas web, entre otros. Existen diversos editores de texto que admiten este tipo de documentos Rmd. En este caso, la edición de todos los documentos se ha llevado a cabo con R Studio.

Palabras claves: R markdown, R Studio

R Markdown como herramienta para la elaboración de materiales docentes

M. Carmen Aguilera-Morillo, Juan Carlos Laria y Álvaro Méndez-Civieta

Departamento de Estadística, Universidad Carlos III de Madrid

Abstract

R Markdown es una herramienta que puede resultar de gran utilidad para los docentes, ya que permite elaborar presentaciones, guiones de prácticas, materiales didácticos en general, reproducibles y dinámicos con R. En este trabajo se mostrará su uso para la elaboración de materiales docentes de la Asignatura Functional Data Analysis del Máster en Statistics for Data Science de la Universidad Carlos III de Madrid. Lo más interesante de R Markdown es que permite combinar texto y códigos procedentes de diversos lenguajes de programación, como son R, Python y SQL. Además, ofrece diferentes formatos de salida estáticos y dinámicos, tales como HTML, PDF, MS Word, Beamer, MS PowerPoint, diapositivas HTML5, libros, artículos científicos, páginas web, entre otros. Existen diversos editores de texto que admiten este tipo de documentos Rmd. En este caso, la edición de todos los documentos se ha llevado a cabo con R Studio.

Formato de salida estático

```
---
title: 'Lesson 1: Introduction to functional data analysis'
author: "Prof. Carmen Aguilera"
date: ' Course: 2018/2019'
fontsize: 10pt
output: beamer_presentation
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE,cache=TRUE)
```


Real functional data

```{r, out.width = '90%', fig.align = "center", fig.cap="Twenty samples of handwriting", echo=FALSE,mesagge=FALSE}
library(fda)
matplot(handwrit[, 1:20, 1], handwrit[, 1:20, 2], type="l", lty=1, xlab="cm",ylab="cm", main="Handwriting data")
```
```

Presentación estática en Beamer - PDF

Figure 1: Twenty samples of handwriting

Formato de salida dinámico

```
---
title: "Practice 1. From vectors to functional data"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE, cache=TRUE)
```

Main R packages for FDA

```{r message=FALSE}
library(fda) # FDA functions -  Ramsay and Silverman book
library(fds) # Functional datasets
```

Exercise 1. Canadian's weather.

Let us consider the monthly average temperature collected during 12 months in 35 weather stations in Canada (raw data). The raw data is given by a 12×35 matrix, where each column represents the i-th sample path, i.e., i-th sample curve observed at 12 points $X_{i}=\left(X_{i1},\ldots,X_{i12} \right) ^{\prime}$.

From this information, the aim is to get the functional dataset (35 temperature curves, i.e., $\{X_{i}\left(t\right): t\in [0,12], i=1,\ldots, 35\}$) observed in a continuous interval of 12 months.

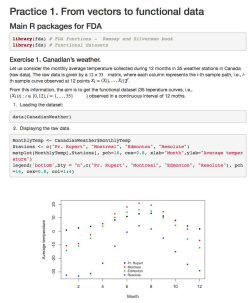
1. Loading the dataset:

```{r fda}
data(CanadianWeather)
```

2. Displaying the raw data

```{r, fig.align = "center"}
MonthlyTemp <- CanadianWeather$monthlyTemp
Stations <- c("Pr. Rupert", "Montreal", "Edmonton", "Resolute")
matplot(MonthlyTemp[,Stations], pch=16, cex=0.8, xlab="Month",ylab="Average temperature")
legend('bottom',bty = 'n',c("Pr. Rupert", "Montreal", "Edmonton", "Resolute"), pch=16, cex=0.8, col=1:4)
```
```

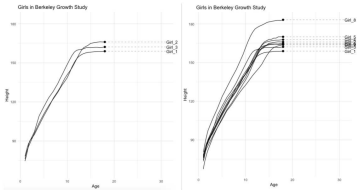
Guión de prácticas dinámico – HTML



Gráficos interactivos

```{r echo=FALSE}
library(fda)
library(ggplot2)
library(tidyverse)
library(gganimate)
data=data.frame(Age=rep(growth$age,ncol(growth$hgtf)),Height=as.vector(growth$hgtf),Girl=sort(rep(c(1:ncol(growth$hgtf)),length(growth$age))))
ggplot(data, aes(Age, Height, group = Girl)) +
geom_line() +
geom_segment(aes(xend = 31, yend = Height), linetype = 2, colour = 'grey') +
geom_point(size = 2) +
geom_text(aes(x = 31.1, label = paste("Girl",Girl,sep="_")), hjust = 0) +
transition_reveal(Girl) +
coord_cartesian(clip = 'off') +
labs(title = 'Girls in Berkeley Growth Study', y = 'Height') +
theme_minimal() +
theme(plot.margin = margin(5.5, 40, 5.5, 5.5))
```

ggplot y gganimate - HTML



Shiny en R Markdown

```
---
title: "Regression splines"
author: "Carmen Aguilera"
date: "14/3/2019"
output: html_document
runtime: shiny
---

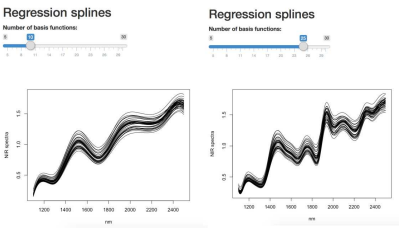
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Regression splines

```{r, echo=FALSE,message=FALSE}
library(fds)
library(fda)
sliderInput("basis", "Number of basis functions:", 30, min = 5, max = 30)

renderPlot({
t <- mircix # Observation points: (t1,..., t700)
Bspline.basis <- create.bspline.basis(rangeval=c(min(t),max(t)), nbasis=input$basis)
data_fd <- Data2fd(y=mircly, argvals=t, basisobj=Bspline.basis)
plot(data_fd, xlab="nm", ylab="NIR spectra",col=1,lty=1)
})
```
```

Shiny APP



Referencias

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. <https://rmarkdown.rstudio.com>.

JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2018). rmarkdown: Dynamic Documents for R. R package version 1.11. URL <https://rmarkdown.rstudio.com>.

J. O. Ramsay, Hadley Wickham, Spencer Graves and Giles Hooker (2018). fda: Functional Data Analysis. R package version 2.4.8. <https://CRAN.R-project.org/package=fda>

Apoyo al aprendizaje de la Estadística mediante el uso de datos reales

Inmaculada Barranco-Chamorro[1]

[1]Departamento Estadística e Investigación Operativa, Universidad de Sevilla

Contacto: chamorro@us.es

Resumen

En este trabajo se describe la metodología seguida como apoyo a la docencia en asignaturas del Grado en Estadística y Estadística Aplicada a la Bioquímica. Se propone analizar datos reales utilizando el software RStudio. Esta actividad supuso para los estudiantes adquirir soltura en las técnicas aprendidas, en la utilización de software estadístico y enfrentarse a nuevos retos. Se propone ir avanzando de forma progresiva en la complejidad de los casos reales a resolver. En una etapa inicial, se analizan conjuntos de datos disponibles en R. En ellos el alumno ha de realizar una tarea de investigación sobre las posibles soluciones para resolver nuevos problemas, por ejemplo, cómo tratar con datos perdidos. Se trata de que el alumno se vaya acercando a la realidad, y al mismo tiempo sea consciente de la técnica estadística que debe aplicar. En una etapa más avanzada se propone al alumno trabajar con datos reales disponibles en repositorios públicos. En particular, destacamos los análisis realizados para datos ómicos. A nivel docente, supuso un cambio de actitud hacia la materia. Aumentando el interés por la misma de forma significativa.

Palabras claves: Enseñanza de Estadística, estudio de casos reales, software RStudio, shinyR

Introducción

En el presente trabajo se describe la metodología seguida como apoyo a la docencia en asignaturas del Grado en Estadística y Estadística Aplicada a la Bioquímica. En ellas se ha propuesto, completar el estudio de la materia con la realización de prácticas en las que se analizan datos reales utilizando el software RStudio.

Para los estudiantes, esta actividad supuso:

- Adquirir soltura en las técnicas aprendidas, así como en la utilización de software estadístico.
- Enfrentarse a nuevos retos, similares a los que habrán de afrontar en su vida profesional.

En cuanto a la metodología seguida, se propone ir avanzando de forma progresiva en la complejidad de los casos reales a resolver.

En una etapa inicial, se realiza el análisis de conjuntos de datos disponibles en R. En ellos el alumno ha de realizar una tarea de investigación sobre las posibles soluciones para resolver los nuevos problemas que se les van planteando, por ejemplo, cómo tratar con datos perdidos. Se trata de que el alumno se vaya acercando a la realidad, y al mismo tiempo sea consciente de la técnica estadística que debe aplicar.

En una etapa más avanzada se propone al alumno trabajar con datos reales disponibles en repositorios públicos. En particular, destacamos los análisis realizados para datos ómicos en Muñoz Armayones (8), un resumen de los mismos puede verse en Barranco-Chamorro et al. (1).

Como logros alcanzados a nivel docente, destacamos que

- Supone un esfuerzo considerable, tanto por parte del profesor, como del alumnado.
- Se produce un cambio de actitud hacia la materia. Aumentando el interés por la misma de forma significativa.

Estudio básico de datasets en R

En la docencia de Estadística para primer curso del Grado en Bioquímica se utilizó el software RStudio para introducir al alumno en el manejo de estructuras de datos en R, tipos de variables, cálculo de medidas estadísticas, gráficos, estudio de relaciones entre las variables, datos perdidos, etc.

En esta etapa trabajamos con los conjunto de datos reales disponibles en R. Basta con escribir el comando data() para ver la lista de los conjuntos de datos disponibles y una breve descripción de ellos.

```
Data sets in package datasets:

AirPassengers           Monthly Airline Passenger Numbers 1949-1960
BJsales                 Sales Data with Leading Indicator
```

```
BJsales.lead (BJsales)
                        Sales Data with Leading Indicator
BOD                     Biochemical Oxygen Demand
CO2                     Carbon Dioxide Uptake in Grass Plants
ChickWeight             Weight versus age of chicks on different
                        diets
DNase                   Elisa assay of DNase
EuStockMarkets          Daily Closing Prices of Major European Stock
                        Indices, 1991-1998
Formaldehyde            Determination of Formaldehyde
HairEyeColor            Hair and Eye Color of Statistics Students
Harman23.cor            Harman Example 2.3

(se omite el resto de la salida)
```

Observamos que muchos de estos conjuntos de datos son de interés para las materias que han de cursar estos estudiantes en Bioquímica.

Como ejemplo de los análisis estadísticos realizados, podemos citar el realizado sobre el conjunto de datos airquality. Información detallada sobre este dataset puede verse en

https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/airquality.html

Este conjunto nos sirvió para ilustrar el uso de los gráficos boxplot, y las técnicas de regresión y correlación lineal. En su estudio se presentaron problemas reales, como es la presencia de datos perdidos, en el que los alumnos tuvieron que consultar bibliografía especializada para resolverlo. Como ilustración se recogen en las Figuras 1 y 2, algunos gráficos de interés realizados para las variables de este conjunto de datos.

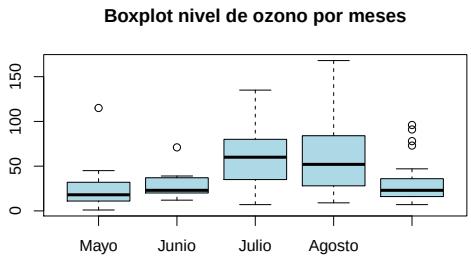


Figura 1. Boxplot nivel de ozono por meses

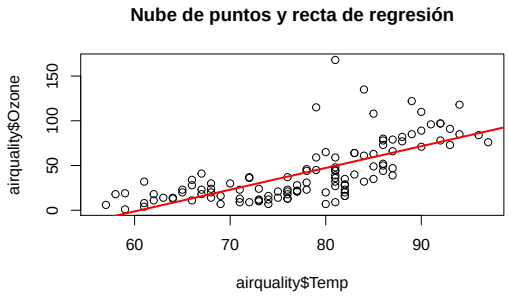


Figura 2. Nube de puntos y recta de regresión del nivel de ozono sobre la temperatura

Estudio de datos ómicos

En las últimas décadas se ha producido un gran desarrollo de las técnicas de biología molecular que permiten detectar y cuantificar características de las moléculas con una alta sensibilidad y especificidad. Estos avances han producido el desarrollo de las llamadas técnicas ómicas, y la generación continua de complejos y masivos datos de este tipo por los laboratorios de todo el mundo. Afortunadamente, muchos de estos conjuntos de datos se encuentran disponibles en repositorios públicos. Su estudio presenta un interés doble. Por un lado, el análisis estadístico de conjuntos de datos ómicos es de interés científico, pues permite contestar a importantes interrogantes de interés biológico y biomédico. Por otro lado, es de interés desde el punto de vista estadístico, pues se han de desarrollar nuevos métodos estadísticos y computacionales para analizarlos. Específicamente uno de los principales retos metodológicos que se plantea es la necesidad de desarrollar técnicas para tratar con un nú-

mero muy grande de variables recogidas sobre un número pequeño de casos. En este contexto, nos centramos en el estudio de técnicas multivariantes de proyección que permitan una reducción de la dimensionalidad del problema. Para tratar con este tipo de datos se consideraron los siguientes métodos de proyección multivariantes: Análisis de Componentes Principales (PCA), Análisis de Correspondencias (CA), y Análisis de Correlación Canónica (CCA). Detalles sobre el uso de estas técnicas para realizar un estudio descriptivo y análisis exploratorio, sobre ciertos conjuntos reales de datos ómicos, pueden verse en (8) y (1). Como software se han utilizado paquetes de R disponibles en Bioconductor. Adicionalmente, los estudiantes desarrollaron una página web interactiva utilizando el paquete de R shinyR. En ella implementaron una aplicación de PCA en la que podían modificar los parámetros del análisis estadístico y observar el efecto sobre los resultados obtenidos.

Como ilustración recogemos algunas pinceladas de la aplicación realizada para análisis de componentes principales. Se utilizó PCA como una técnica de análisis exploratorio de datos ómicos para el caso $n << p$. Mostramos cómo se logra reducir la dimensionalidad del problema, puesto que con unas pocas PC's podemos capturar un alto porcentaje de la variabilidad que hay en nuestros datos. Este uso se basa en que al aplicar PCA, las distancias entre los datos originales y los datos transformados coinciden. Para realizar el estudio se utilizó el paquete FactoMineR. Se analizaron diversos conjuntos de datos reales, entre ellos el denominado chicken , el cuál contiene datos de expresión génica de 7406 genes (variables) medidos en 43 pollos (muestras). También hay una variable categórica que contiene 6 dietas en función de distintos niveles. Se representaron las proyecciones, y se visualizaron las elipses de confianza de las clases de la variale categórica tras el PCA con el comando plotellipses(). Estas elipses son útiles para comparar categorías, decidir si una observación procede o no de una determinada población con distribución normal bivariante y además, detectar puntos anómalos. Para este conjunto de datos, esta salida se recoge en la Figura 3. Se observa que las elipses asociadas a las dietas J48R24 y J48 no se solapan, lo que significa que estas dietas son significativamente diferentes. Sin embargo, las elipses correspondientes a las 4 categorías restantes muestran que no existen diferencias significativas entre ellas.

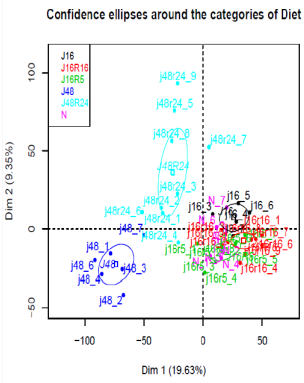


Figura 3. Gráfico realizado como ilustración del PCA

En las referencias (8) y (1) de la Bibliografía, pueden verse detalles sobre el análisis realizado en este caso y sobre otros conjuntos de datos reales. Se realizaron aplicaciones de técnicas de proyección multivariantes, obteniéndose gráficos novedosos y de interés para los estudiantes como el que se muestra en la Figura 4 y que corresponde a un estudio de Análisis de Correlación Canónica (CCA).

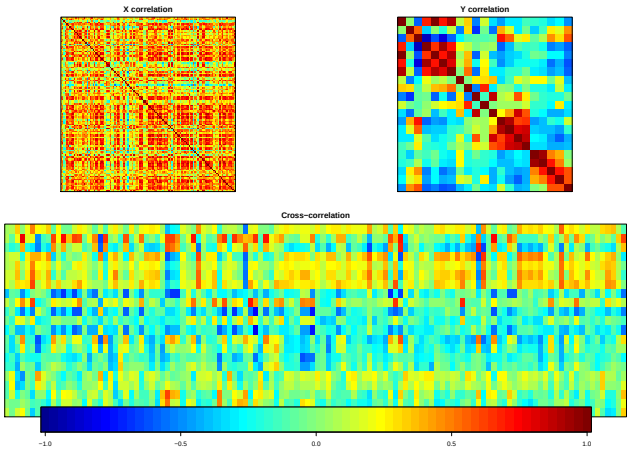


Figura 4. Gráfico realizado como ilustración del CCA

Conclusiones

Para evaluar el impacto del método seguido, utilizamos cuestionarios. En ellos se aprecia un mayor interés y motivación por la materia, así como posibles

deficiencias del método, que hay que ir solventando.

Referencias

[1] I. Barranco-Chamorro, S. Muñoz-Armayones, A. Romero-Losada y F. Romero-Campero (2019) Multivariate Projection Techniques to Reduce Dimensionality in Large Datasets. In *Smart Data. State-of-the-Art Perspectives in Computing and Applications*, 133-160. Taylor & Francis.

[2] A. Culhane (2016) *Introduction to Multivariate Analysis of Microarray Gene Expression Data using MADE4*. https://bioconductor.riken.jp/packages/3.3/bioc/vignettes/made4/inst/doc/introduction.pdf [03 mayo 2016]

[3] A. Culhane, J. Thioulouse, G. Perriere y DG. Higgins (2005) MADE4:an R package for multivariate analysis of gene expression data. *Bioinformatics*, **21(11)**, pp. 2789-2790.

[4] B. Everitt y T. Hothorn (2015) *An Introduction to Applied Multivariate Analysis with R*. Springer.

[5] I. González y S. Déjean (2012) *CCA: Canonical correlation analysis. R package version 1.2.* http://CRAN.R-project.org/package=CCA

[6] W.K. Hardle y L. Simar (2015) *Applied Multivariate Statistical Analysis*. Springer-Verlag.

[7] F. Husson, J. Josse, S. Le y J. Mazet (2015) *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package version 1.31.4*. http://CRAN.Rproject. org/package=FactoMineR

[8] S. Muñoz Armayones (2016) *Técnicas Multivariantes para el Análisis de Datos Ómicos* . TFG disponible en Depósito de Investigación de la Universidad de Sevilla, idus.

[9] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

[10] R Studio URL: http://www.rstudio.org/

[11] A. Sánchez. Multivariate Methods for the Integration and Visualization of Omics Data. URL: http://eib.stat.ub.edu/Integrative+Analysis+of+Omics+Data.

Rstudio como apoyo al aprendizaje de la Estadística mediante el uso de datos reales

I. Barranco-Chamorro (chamorro@us.es)

Dpto. De Estadística e I.O. Universidad de Sevilla

Descripción

En el presente trabajo se describe la metodología seguida como apoyo a la docencia en asignaturas del Grado en Estadística y Estadística Aplicada a la Bioquímica. En ellas se ha propuesto, completar el estudio de la materia con la realización de prácticas en las que se analizan datos reales utilizando el software RStudio.

Para los estudiantes, esta actividad supuso:

- Adquirir soltura en las técnicas aprendidas, así como en la utilización de software estadístico.
- Enfrentarse a nuevos retos, similares a los que habrán de afrontar en su vida profesional.

En cuanto a la **metodología seguida**, se propone ir avanzando de forma progresiva en la complejidad de los casos reales a resolver.

En una etapa inicial, se realiza el **análisis de conjuntos de datos disponibles en R.** En ellos el alumno ha de realizar una tarea de investigación sobre las posibles soluciones para resolver los nuevos problemas que se les van planteando, por ejemplo, cómo tratar con datos perdidos. Se trata de que el alumno se vaya acercando a la realidad, y al mismo tiempo sea consciente de la técnica estadística que debe aplicar.

En una etapa más avanzada se propone al alumno trabajar con datos reales disponibles en repositorios públicos. En particular, destacamos los análisis realizados para datos ómicos en Muñoz Armayones (2016), un resumen de los mismos puede verse en Barranco-Chamorro et al. (2019).

Como logros alcanzados a nivel docente, destacamos que

- Supone un esfuerzo considerable, tanto por parte del profesor, como del alumnado.

- Se produce un **cambio de actitud hacia la materia**. Aumentando el interés por la misma de forma significativa.

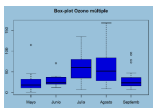
Palabras clave: enseñanza de Estadística, estudio de casos reales, software RStudio

Clasificación AMS:Clasificación AMS: 9702, 97U30, 97U40

- Estudio básico de datasets en R

Ejemplo: Trabajaremos con el conjunto de datos airquality. Información puede verse en https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/airquality.html

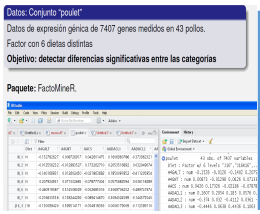
Se utilizó para introducir al alumno en el manejo de estructuras de datos en R, tipos de variables, cálculo de medidas estadísticas, gráficos, estudio de relaciones entre las variables, datos perdidos, etc.



Aplicación a datos ómicos

Análisis exploratorio de datos ómicos (n<<p). Reducción de la dimensionalidad.

- Con unas pocas PC's podemos **capturar un alto porcentaje de la variabilidad** que hay en nuestros datos.
- Las **distancias** entre los datos originales y los datos transformados coinciden.



Paquete: FactoMineR.
Funciones:
PCA(poulet; quali:sup = 1; graph = F)
summary()
plot(PCA; choix = ind; habillage = 1)
plotellipses()

Se supone normalidad multivariante.

Se representan individuos, categorías y sus elipses de confianza al 95 %.

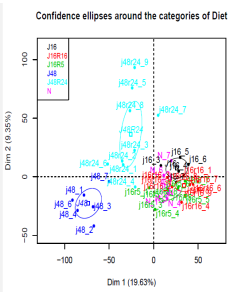
Elipses de Confianza

Utilidad:

- Comparación de categorías
- Decidir qué observaciones pertenecen a cada población
- Detectar puntos anómalos

Interpretación

- Elipses que **se solapan** muestran categorías **sin diferencias significativas** entre ellas.



En 1. y 2. de la Bibliografía, pueden verse estudios similares, incluyendo links a los conjuntos de datos, paquetes de R y Bioconductor utilizados. Así como una aplicación para PCA, en shiny R, que permite modificar los parámetros del studio

https://frannetworks.shinyapps.io/web_app/

Conclusiones. Para evaluar el impacto del método seguido, utilizamos **cuestionarios. En ellos se aprecia un mayor interés y motivación por la materia**, así como posibles deficiencias del método, que hay que ir solventando.

Bibliografía

1. Muñoz Armayones, S. (2016). Técnicas Multivariantes para el Análisis de Datos Ómicos. (TFG disponible en Depósito de Investigación de la Universidad de Sevilla, idus)

2. Barranco-Chamorro, I., Muñoz-Armayones, S., Romero-Losada, A., Romero-Campero, F. Chapter 7: Multivariate Projection Techniques to Reduce Dimensionality in Large Datasets (in Smart Data) Taylor&Francis (se publicará en 2019).

Dificultades de los estudiantes en la comprensión de la distribución muestral

Nuria Begué[1], María Magdalena Gea[2]

[1]Universidad de Zaragoza
[2]Departamento de Didáctica de la Matemática, Universidad de Granada

Contacto: nbegue@correo.ugr.es

Resumen

En el trabajo con la inferencia estadística, los estudiantes han de trabajar con tres distribuciones diferentes: a) La distribución teórica en la población, definida por uno o varios parámetros desconocidos; b) la distribución de datos obtenida en una muestra aleatoria de la población, cuyos resúmenes estadísticos se utilizan para estimar los parámetros de la población; c) la distribución muestral teórica de los estadísticos, que surge al considerar las infinitas muestras que pueden tomarse en las mismas condiciones, de la población dada. El objetivo de este trabajo es describir las principales dificultades de los estudiantes para diferenciar y comprender estas tres distribuciones. Entre ellas citamos la falta de comprensión de la variabilidad de la distribución muestral y del efecto del tamaño de la muestra sobre la misma, el considerar las diferentes muestras como conjuntos sin elementos comunes o aplicar el teorema central del límite incorrectamente.

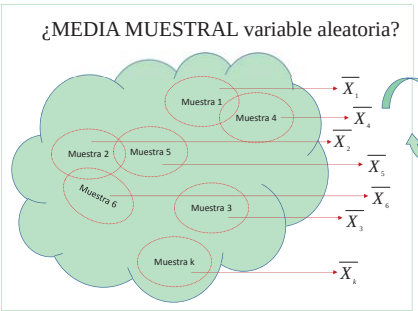
Palabras claves: Distribuciones muestrales, comprensión, dificultades

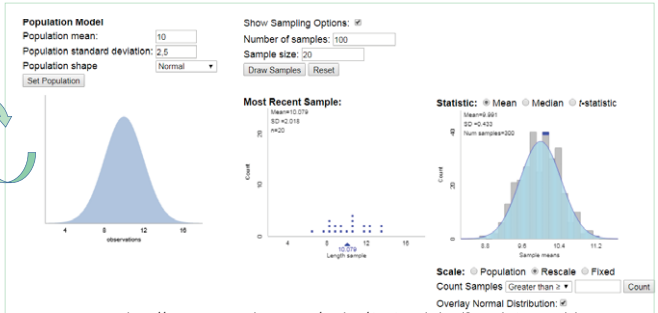
DIFICULTADES DE LOS ESTUDIANTES EN LA COMPRENSIÓN DE LA DISTRIBUCIÓN MUESTRAL

Nuria Begué[1], María M. Gea[2]
[1]Universidad de Zaragoza, [2]Universidad de Granada
nbegue@unizar.es, **mmgea@ugr.es**

En inferencia, los estudiantes han de trabajar con tres distribuciones diferentes (Harradine, Batanero y Rossman, 2011):

- ✓ La distribución teórica en la población, definida por uno o varios parámetros desconocidos;
- ✓ La distribución de datos, obtenida en una muestra aleatoria de la población, cuyos resúmenes estadísticos se utilizan para estimar los parámetros de la población;
- ✓ La distribución muestral teórica de los estadísticos, que surge al considerar las infinitas muestras que pueden tomarse en las mismas condiciones, de la población dada. El objetivo de este trabajo es describir las principales dificultades de los estudiantes para diferenciar y comprender estas tres distribuciones. Entre ellas citamos la falta de comprensión de la variabilidad de la distribución muestral y del efecto del tamaño de la muestra sobre la misma, el considerar las diferentes muestras como conjuntos sin elementos comunes o aplicar el teorema central del límite incorrectamente.





http://www.rossmanchance.com/applets/OneSample.html?population=model

La necesidad de trabajar en el mismo problema con las tres distribuciones origina una dificultad para los estudiantes pues, en estadística descriptiva, la unidad de análisis es cada dato (referido a una persona u objeto) y en inferencia, el interés está en conocer información sobre la población, por lo que este nuevo enfoque exige la obtención de distintas muestras de una población particular, donde la unidad de análisis es cada una de las posibles muestras de un mismo tamaño. Los estudiantes confunden con frecuencia estos diferentes niveles de concreción del concepto de distribución en estadística descriptiva e inferencia, lo que es un problema relacionado con la enseñanza y aprendizaje del muestreo (Schuyten, 1991).

En el programa de investigación heurísticas y sesgos (Kahneman, Slovic y Tversky, 1982), en relación al muestreo, destacamos:

La *heurística de representatividad*, donde la probabilidad de obtener una muestra se estima comparando la similitud que guarda la muestra con la población a la que pertenece o bien con el proceso de muestreo que la ha generado. En la insensibilidad al tamaño de la muestra, la persona asume que una muestra, aunque sea pequeña e independientemente de como se ha seleccionado, siempre representa a la población. Cuando la persona cree en la "ley de los pequeños números", generaliza indebidamente la ley de los grandes números. En la falacia del jugador, se considera que el resultado de un experimento aleatorio afectará en sucesos futuros (recencia positiva o negativa).

La *heurística de la disponibilidad* consiste en estimar la verosimilitud de un suceso basándose en la información previa de la persona sobre la ocurrencia del suceso de interés. La aplicación de dicha heurística supone la estimación de la probabilidad de un suceso en función de la facilidad por encontrar ejemplos de situaciones similares (Díaz, 2003). Se identifican tres sesgos en consecuencia: el sesgo de equiprobabilidad (Lecoutre, 1992), que ha sido recientemente explicado por Chernoff y Russsel (2012) en lo que llaman falacia de la composición); el enfoque del resultado (Konold, 1989), que implica interpretar el resultado de un experimento desde el enfoque subjetivo de la probabilidad (no asume la variabilidad del muestreo); y la ilusión de control (Langer, 1982).

Acknowledgement: Projects EDU2016-74848-P (AEI, FEDER), and FCT-16-10974 (FECYT-MEIC).

Referencias

Chernoff, E. J. y Russell, G. L. (2012). The fallacy of composition: Prospective mathematics teachers' use of logical fallacies. *Canadian Journal for Science, Mathematics and Technology Education*, *12*(3), 259-271. http://dx.doi.org/10.1080/14926156.2012.704128.

Díaz, C. (2003). Heurísticas y sesgos en el razonamiento probabilístico. Implicaciones para la enseñanza de la estadística. *Actas del 27 Congreso Nacional de Estadística e Investigación Operativa*. Lérida: Sociedad de Estadística e Investigación Operativa. [CD- ROM].

Harradine, A., Batanero, C. y Rossman, A. (2011).Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill, C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education*. (pp. 235-246). Dordrecht, The Netherlands: Springer.

Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*, 59-98.

Langer, E. J. (1982). The illusion of control. En D. Kahneman, P. Slovic, y A. Tversky (Eds.), *Judgment under uncertainty: Heuristic and biases* (pp. 231-238). New York: Cambridge University Press.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, *23*, 557-568.

Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere-Jones (Ed.) *Proceeding of the Third ICOTS* (pp. 486-490). Otago, Nueva Zelanda: International Statistical Institute.

Las infografías como estrategia de innovación docente para mejorar competencias en el Grado de Estadística

Eva Boj[1], Ana Castañer[1], Mª Mercè Claramunt[1], Mª Teresa Costa[1], Javier Martínez de Albéniz[1]

[1]Departament de Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona

Contacto: evaboj@ub.edu

Resumen

Una infografía permite presentar contenido informativo complejo y denso de una manera que apoya el procesamiento cognitivo y el aprendizaje. Se combinan visualizaciones de datos, ilustraciones, texto e imágenes en un formato que cuenta una historia. En este trabajo se presenta la experiencia de innovación docente del uso de infografías en dos asignaturas del Grado de Estadística de la UB-UPC. Los alumnos deben adquirir como competencia la capacidad para ordenar y representar la información proporcionada por un conjunto de datos y en los planes docentes se incide en la importancia de saber presentar los resultados obtenidos en público o en un entorno profesional. En las asignaturas implicadas se utilizan infografías publicadas para reforzar el aprendizaje y se facilita material e información a los alumnos sobre el proceso de creación de una infografía. En la evaluación continua se incluye la elaboración de una infografía y se evalúa mediante rúbricas. Se han recogido las opiniones de los estudiantes sobre el uso de infografías mediante una encuesta. La mayoría considera que ha sido útil en su proceso de aprendizaje y ha mejorado sus habilidades comunicativas.

Palabras claves: infografía, estadística, competencias transversales, innovación docente, rúbrica

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa
Granada, 4-5 abril 2019

Las infografías como estrategia de innovación docente para mejorar competencias en el Grado de Estadística

Boj, E.; Boncompte, M.; Castañer, A.; Claramunt, M.M.; Costa, T.; Martínez de Albéniz, J.
Dept. Matemàtica Econòmica, Financera i Actuarial (UB)

Resumen

Una infografía permite presentar contenido informativo complejo y denso de una manera que apoya el procesamiento cognitivo y el aprendizaje. Se combinan visualizaciones de datos, ilustraciones, texto e imágenes en un formato que cuenta una historia.

En este trabajo se presenta la experiencia de innovación docente del uso de infografías en dos asignaturas del Grado de Estadística de la UB-UPC.

Los alumnos deben adquirir como competencia la capacidad para ordenar y representar la información proporcionada por un conjunto de datos y en los planes docentes se incide en la importancia de saber presentar los resultados obtenidos en público o en un entorno profesional.

En las asignaturas implicadas se utilizan infografías publicadas para reforzar el aprendizaje y se facilita material e información a los alumnos sobre el proceso de creación de una infografía. En la evaluación continua se incluye la elaboración de una infografía y se evalúa mediante rúbricas.

Se han recogido las opiniones de los estudiantes sobre el uso de infografías mediante una encuesta. La mayoría considera que ha sido útil en su proceso de aprendizaje y ha mejorado sus habilidades comunicativas.

Asignaturas implicadas de Grado

Grado de Estadística UB-UPC

240 créditos

Análisis de supervivencia

Asignatura optativa de 6 créditos

Demografía

Asignatura optativa de 6 créditos

Competencias del Grado:

- Capacidad comunicativa, capacidad para entenderse y expresarse oralmente y por escrito y capacidad para buscar, usar e integrar información.
- Capacidad para ordenar, representar y agregar, con criterios objetivos, la información proporcionada por un conjunto de datos.
- Capacidad para aplicar los conocimientos adquiridos y capacidad para analizar y resolver problemas en contextos académicos y profesionales.

Objetivos de aprendizaje (Planes docentes):

- Desarrollar habilidades de comunicación.
- Adquirir nociones de hablar en público, de comunicarse y de cómo se hace una presentación oral.
- Presentar los resultados en un entorno profesional para la toma de decisiones.
- Resumir la información de las variables cualitativas y cuantitativas por medio de las estadísticas apropiadas y en forma de tablas y gráficos.

Plan de trabajo

1. Sesión explicativa sobre qué son las infografías y su elaboración
2. Búsqueda de información de manera autónoma por parte de los alumnos
3. Elaboración de una infografía y de un documento explicativo
4. Evaluación de la infografía mediante rúbricas
5. Encuesta de satisfacción

INFOGRAFÍA

Combinación de imágenes sintéticas y textos con la finalidad de comunicar información de manera visual para facilitar su transmisión

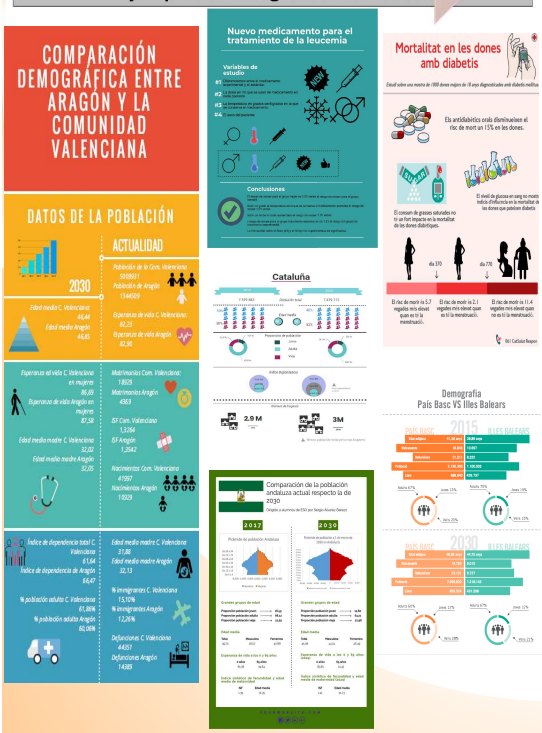
PROCESO PARA CREAR UNA INFOGRAFÍA

Elegir un tema
Recopilar información
Filtrar la información
Elegir la tipología
Elaborar borrador
Programas de diseño
Revisar elementos

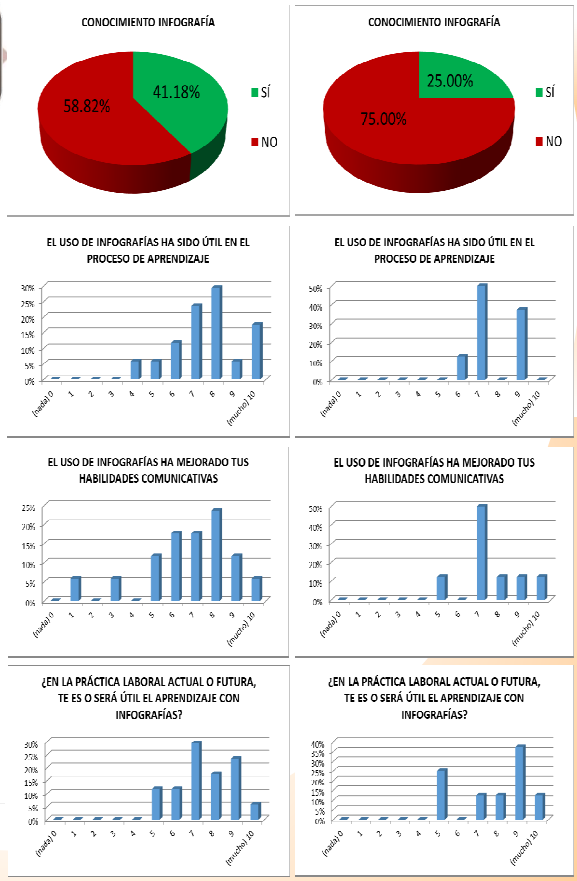
WEBS PARA LA ELABORACIÓN DE INFOGRAFÍAS

http://www.coolinfographics.com/
http://www.bestinfographics.co/
https://es.pinterest.com/mashable/infographics/
http://blog.hubspot.com/marketing/best-infographics-2015
http://infographicpics.com/

Ejemplos de infografías realizadas



Resultados encuesta de valoración en Demografía y Análisis de Supervivencia



Introducción a la Estadística con SPSS y Statgraphics: ejercicios guiados y aplicaciones con datos reales

Pedro Antonio Cano Chica[1], Valentina Cueva López[1], Antonio Jesús López Montoya[1], Irene García Garrido[1]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Jaén

Contacto: pacano@ujaen.es

Resumen

En diversas asignaturas correspondientes al grado de Estadística se instruye al alumnado en el manejo de distintos paquetes estadísticos, dos de los más utilizados en la Universidad de Jaén son SPSS y Statgraphics. El objetivo principal del uso de dichos paquetes es completar la formación recibida en las clases teóricas, así como resolver problemas que se puedan plantear en el ámbito profesional. Para ello, se le facilita al alumnado guiones de prácticas, los cuales están debidamente estructurados y son lo bastante precisos como para facilitar su correcta comprensión. Los referidos guiones se componen de varios ejercicios resueltos, de forma que cada instrucción lleva aparejada su correspondiente resultado gráfico, lo cual permite una inmediata correspondencia entre ambas que facilita enormemente la resolución de cada cuestión planteada. Como consecuencia, el alumnado es capaz de alcanzar los conocimientos requeridos de manera más atractiva con la ayuda de dichos paquetes estadísticos.

Palabras claves: Estadística, guiones, herramientas gráficas, software, SPSS, Statgraphics

Introducción a la Estadística con SPSS y Statgraphics: ejercicios guiados y aplicaciones con datos reales

Pedro Antonio Cano-Chica, Valentina Cueva-López, Antonio Jesús López-Montoya, Irene García-Garrido

Departamento de Estadística e Investigación Operativa, Universidad de Jaén

RESUMEN

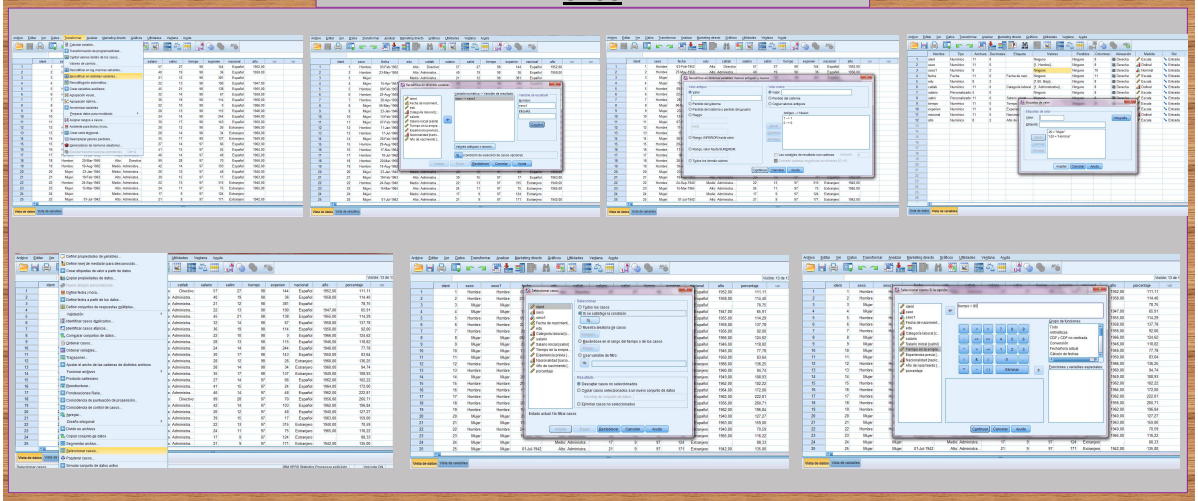
En diversas asignaturas correspondientes al grado de Estadística se instruye al alumnado en el manejo de distintos paquetes estadísticos, dos de los más utilizados en la Universidad de Jaén son SPSS y Statgraphics. El objetivo principal del uso de dichos paquetes es completar la formación recibida en las clases teóricas, así como resolver problemas que se puedan plantear en el ámbito profesional.

Para ello, se le facilita al alumnado guiones de prácticas, los cuales están debidamente estructurados y son lo bastante precisos como para facilitar su correcta comprensión. Los referidos guiones se componen de varios ejercicios resueltos, de forma que cada instrucción lleva aparejada su correspondiente resultado gráfico, lo cual permite una inmediata correspondencia entre ambas que facilita enormemente la resolución de cada cuestión planteada. Como consecuencia, el alumnado es capaz de alcanzar los conocimientos requeridos de manera más atractiva con la ayuda de dichos paquetes estadísticos.

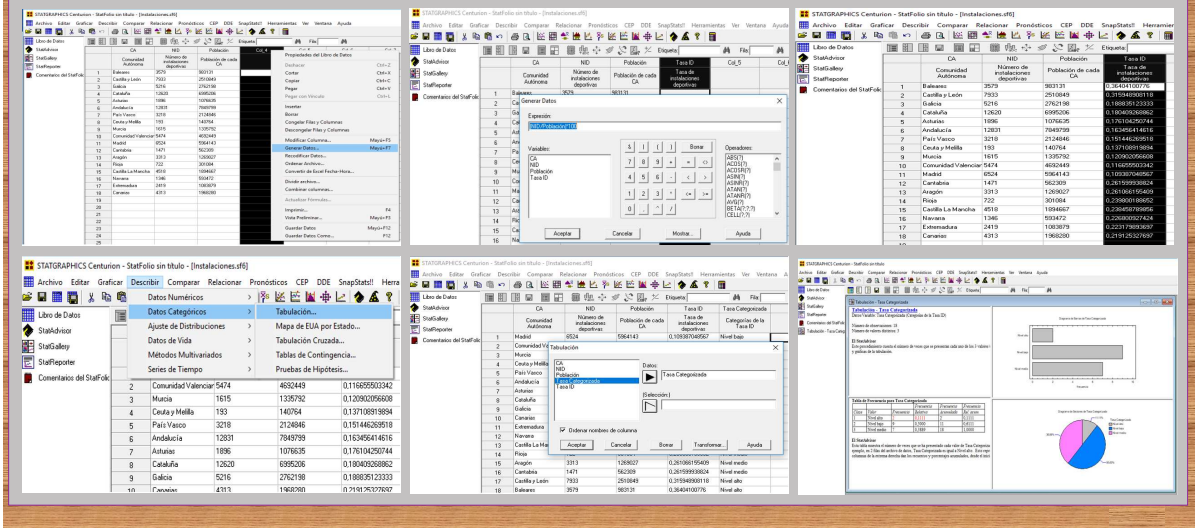
OBJETIVOS DE LA PRÁCTICA

- Familiarizar al alumnado con el entorno de SPSS y Statgraphics, haciendo más autónomo el aprendizaje de los mismos.
- Facilitar al alumnado la realización de técnicas estadísticas avanzadas mediante los paquetes estadísticos SPSS y Statgraphics.
- Generar ficheros y gráficos adecuados para visualización y tratamiento de datos.

SPSS



STATGRAPHICS



CONCLUSIONES

- **Se verifica el aprendizaje autónomo mediante el aumento en la comprensión de los conceptos objetivos de la asignatura en relación a años anteriores.**
- **Se observa un porcentaje mayor del alumnado que realiza correctamente todos los ejercicios propuestos en relación a años anteriores.**
- **Facilita tanto al alumnado como al profesorado, el estudio/impartición de la asignatura de Estadística en cuestión**

Test de Van Valen para la homocedasticidad multivariante

Francesc Carmona Pontaque[1]

[1]Departamento de Genética, Microbiología y Estadística, Universidad de Barcelona

Contacto: fcarmona@ub.edu

Resumen

Muchos paquetes de diversos programas estadísticos usan el test *M* de Box para comparar la variabilidad de dos o más muestras multivariantes. Sin embargo, es de sobras conocido que este test es muy sensible a la falta de ajuste a una distribución normal multivariante. Así pues, necesitamos un procedimiento alternativo que debería ser más robusto y que podemos construir con las ideas del test de Levene univariante. Van Valen (1978) sugirió el cálculo de las distancias euclídeas al vector de medianas muestrales. Las medias de esas distancias se pueden comparar con un test t de Student ($p = 2$ grupos), un test F ($p > 2$) o un test no paramétrico como el de Kruskall-Wallis. En el marco de la asignatura Análisis multivariante del máster de Bioinformática y Bioestadística de la UOC-UB se propone un taller para construir un programa o función en R que proporcione el test de Van Valen en sus diferentes versiones. Los resultados se exponen en la plataforma GitHub para su mejora y utilización pública.

Palabras claves: Análisis de datos, Estadística multivariante, Inferencia

Agradecimientos. El autor agradece a los alumnos de la asignatura de Análisis multivariante del máster de Bioinformática y Bioestadística de la UOC-UB su colaboración en el proceso de debate y concreción de este trabajo. También al grupo de investigación GRBIO.

Referencias

[1] E. Acuna, *Data Pre-Processing and Visualization Functions for Classification, Package 'dprep' was remove from the CRAN repository*.

[2] B.F.J., Manly y J.A., Navarro, (2017) *Multivariate Statistical Methods: A Primer*. Fourth Edition CRC Press.

[3] B.F.J., Manly y R.I.C.C., Francis. (2002) Testing for mean and variance differences with samples from distributions that may be non-normal with

unequal variances. *Journal of Statistical Computation and Simulation*, **72**, pp. 633-646.

[4] L., Van Valen. (1978) The statistics of variation. *Evolutionary Theory*, **4**, pp. 33-43. (*Erratum Evolutionary Theory*, **4**, pp. 202.)

Test de Van Valen para la homocedasticidad multivariante

Francesc Carmona Pontaque (fcarmona@ub.edu)
Facultad de Biología, Universidad de Barcelona

Introducción

- Para comparar la variabilidad de dos o más muestras multivariantes muchos programas estadísticos usan el test *M* de Box.
- Este test es muy sensible a la no normalidad multivariante de las muestras.
- Un resultado significativo del test puede deberse a la no normalidad, en vez de a las diferencias entre las matrices de covarianzas.

- Necesitamos un procedimiento alternativo que debe ser más robusto y que podemos construir con las ideas del test de Levene univariante.
- En ese test los datos se transforman en el valor absoluto de las desviaciones respecto a la media o la mediana muestral para cada grupo.
- La cuestión de si hay diferencias entre las varianzas se convierte en si hay diferencias significativas entre las medias de los valores transformados.
- El contraste se puede hacer con un test *t* de Student, para dos grupos, o con un ANOVA, más de dos.

- **Propuesta de Van Valen (1978)**:
  (1) Calcular

$$d_{ij}=\sqrt{\sum_{k=1}^{p}\left(x_{ijk}-M_{jk}\right)^2}$$

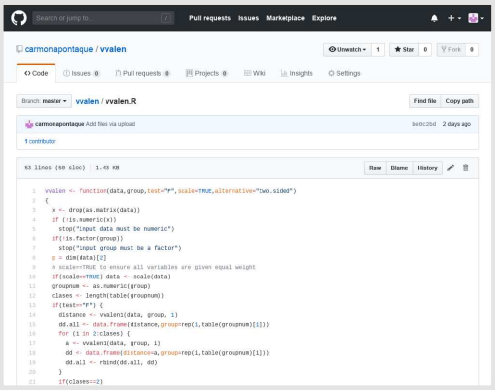
  donde p es el número de grupos, x_{ijk} es el valor de la variable para el individuo i en la muestra j y M_{jk} es la mediana de la misma variable en la muestra.
  (2) Las medias de los valores se comparan con un test *t* de Student (p=2 grupos), un test *F* (p>2) o un test no paramétrico como el de Kruskall-Wallis.

Metodología

- En el marco de la asignatura *Análisis multivariante* del máster interuniversitario de *Bioinformática y Bioestadística* de la UOC-UB se propone un taller para construir un programa o función en R que proporcione el test de Van Valen en sus diferentes versiones.

- Los resultados se exponen en la plataforma GitHub para su mejora y utilización pública.

- Más adelante se puede crear un paquete de R para facilitar su uso.



Resultados

```
library(devtools)
source_url("https://raw.githubusercontent.com/carmonapontaque/
vvalen/master/vvalen.R")

> # Ejemplo 1
> sparrows <- read.csv("Bumpus_sparrows.csv", header=TRUE)
> vvalen(sparrows[,-1], sparrows[,1], test = "F",
                        alternative = "greater")

   Two Sample t-test

data:  distance by factor(group)
t = 1.9241, df = 47, p-value = 0.0302
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.06392549        Inf
sample estimates:
mean in group 1 mean in group 2
       2.241557        1.741959

> # Ejemplo 2
> skulls <- read.csv("Egyptian skulls.csv", header=TRUE)
> vvalen(skulls[,-1], skulls[,1], test = "F")
Analysis of Variance Table

Response: distance
                Df Sum Sq Mean Sq F value Pr(>F)
factor(group)   4  2.091 0.52285  1.0525 0.3824
Residuals     145 72.031 0.49677
> vvalen(skulls[,-1], skulls[,1], test = "kruskall")

   Kruskal-Wallis rank sum test

data:  distance
Kruskal-Wallis chi-squared = 5.1876, df = 4, p-value = 0.2686
```

Conclusión

- El taller ha sido útil para debatir y mejorar las propuestas de los alumnos que han llegado vía e-mail.

- La práctica ha servido tanto para aprender algo más de programación, como para comprender y utilizar algunas funciones de R.

- Las aportaciones de todo tipo también han servido en la evaluación continua del alumnado.

Bibliografía

[1] Edgar Acuna, Data Pre-Processing and Visualization Functions for Classification, Package 'dprep' was removed from the CRAN repository.

[2] Manly, B.F.J. and Navarro, J.A. (2017) Multivariate Statistical Methods: A Primer. Fourth Edition CRC Press.

[3] Manly, B.F.J. and Francis, R.I.C.C. (2002) Testing for mean and variance differences with samples from distributions that may be non-normal with unequal variances. *Journal of Statistical Computation and Simulation* 72: 633-646

[4] Van Valen, L. (1978) The statistics of variation. *Evolutionary Theory* 4: 33-43. (Erratum *Evolutionary Theory* 4: 202.)

GRUP DE RECERCA EN ESTADÍSTICA I BIOINFORMÀTICA

GRBIO

Actividad de Estadística Descriptiva: "Reto: ¿eres capaz de descifrar la contraseña del candado?"

Sonia Castillo-Gutiérrez[1], María Dolores Estudillo-Martíŋnez[1], Nuria Ruiz-Fuentes[1], María del Pilar Fríŋas-Bustamante[1]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Jaén

Contacto: socasti@ujaen.es

Resumen

En este trabajo se presenta una actividad sobre contenidos de Estadística Descriptiva realizada en el aula de informática para el alumnado del Grado en Trabajo Social de la Universidad de Jaén. Esta es una de las actividades desarrolladas en un Proyecto de Innovación Docente a través de talleres motivacionales para la adquisición de competencias en Estadística. Esta práctica se presenta como un reto para el alumnado al que se le enseña una caja con un candado con contraseña. Para poder abrir dicha caja deben descubrir una contraseña de 3 dígitos, a través del cálculo de medidas descriptivas mediante Microsoft Excel. El objetivo principal de la actividad es consolidar los conocimientos teóricos aprendidos, despertando el interés del alumnado en la sesión de dos horas de duración. El alumnado expresó su satisfacción por el nuevo método empleado, puesto que el afán por conseguir la contraseña, como parte de un juego, hizo más interesante el desarrollo de la práctica.

Palabras claves: Estadíŋstica Descriptiva, innovación docente, reto

Agradecimientos. Los autores agradecen el apoyo de la Universidad de Jaén en el Proyecto de Innovación Docente con código PID49_201617, denominado "Talleres Motivacionales para la adquisición de competencias en Estadística".

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

Actividad de Estadística Descriptiva: "Reto: ¿eres capaz de descifrar la contraseña del candado?"

Sonia Castillo-Gutiérrez, María Dolores Estudillo-Martínez, Nuria Ruiz-Fuentes y María del Pilar Frías-Bustamante

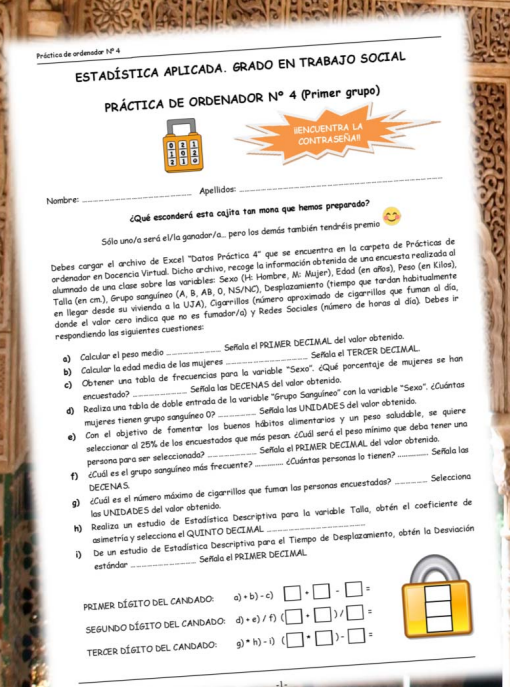
Departamento de Estadística e Investigación Operativa. Universidad de Jaén

Contacto: socasti@ujaen.es

RESUMEN

En este trabajo se presenta una actividad sobre contenidos de Estadística Descriptiva realizada en el aula de informática para el alumnado del Grado en Trabajo Social de la Universidad de Jaén. Esta es una de las actividades desarrolladas en un Proyecto de Innovación Docente a través de talleres motivacionales para la adquisición de competencias en Estadística. Esta práctica se presenta como un reto para el alumnado al que se le enseña una caja con un candado con contraseña. Para poder abrir dicha caja deben descubrir una contraseña de 3 dígitos, a través del cálculo de medidas descriptivas mediante Microsoft Excel. El objetivo principal de la actividad es consolidar los conocimientos teóricos aprendidos, despertando el interés del alumnado en la sesión de dos horas de duración. El alumnado expresó su satisfacción por el nuevo método empleado, puesto que el afán por conseguir la contraseña, como parte de un juego, hizo más interesante el desarrollo de la práctica.

PRÁCTICA



Práctica de ordenador Nº 4

ESTADÍSTICA APLICADA. GRADO EN TRABAJO SOCIAL

PRÁCTICA DE ORDENADOR Nº 4 (Primer grupo)

¡ENCUENTRA LA CONTRASEÑA!

Nombre: ..... Apellidos: .....

¿Qué esconderá esta cajita tan mona que hemos preparado?

Sólo uno/a será el/la ganador/a... pero los demás también tendréis premio

Debes cargar el archivo de Excel "Datos Práctica 4" que se encuentra en la carpeta de Prácticas de ordenador en Docencia Virtual. Dicho archivo, recoge la información obtenida de una encuesta realizada al alumnado de una clase sobre las variables: Sexo (H: Hombre, M: Mujer), Edad (en años), Peso (en Kilos), Talla (en cm.), Grupo sanguíneo (A, B, AB, O, NS/NC), Desplazamiento (tiempo que tardan habitualmente en llegar desde su vivienda a la UJA), Cigarrillos (número aproximado de cigarrillos que fuman al día, donde el valor cero indica que no es fumador/a) y Redes Sociales (número de horas al día). Debes ir respondiendo las siguientes cuestiones:

a) Calcular el peso medio ..... Señala el PRIMER DECIMAL del valor obtenido.
b) Calcular la edad media de las mujeres ..... Señala el TERCER DECIMAL.
c) Obtener una tabla de frecuencias para la variable "Sexo". ¿Qué porcentaje de mujeres se han encuestado? ..... Señala las DECENAS del valor obtenido.
d) Realiza una tabla de doble entrada de la variable "Grupo Sanguíneo" con la variable "Sexo". ¿Cuántas mujeres tienen grupo sanguíneo O? ..... Señala las UNIDADES del valor obtenido.
e) Con el objetivo de fomentar los buenos hábitos alimentarios y un peso saludable, se quiere seleccionar al 25% de los encuestados que más pesan. ¿Cuál será el peso mínimo que deba tener una persona para ser seleccionado? ..... Señala el PRIMER DECIMAL del valor obtenido.
f) ¿Cuál es el grupo sanguíneo más frecuente? ..... ¿Cuántas personas lo tienen? ..... Señala las DECENAS.
g) ¿Cuál es el número máximo de cigarrillos que fuman las personas encuestadas? ..... Selecciona las UNIDADES del valor obtenido.
h) Realiza un estudio de Estadística Descriptiva para la variable Talla, obtén el coeficiente de asimetría y selecciona el QUINTO DECIMAL .....
i) De un estudio de Estadística Descriptiva para el Tiempo de Desplazamiento, obtén la Desviación estándar ..... Señala el PRIMER DECIMAL.

PRIMER DÍGITO DEL CANDADO: a) + b) - c) [ ] + [ ] - [ ] =

SEGUNDO DÍGITO DEL CANDADO: d) + e) / f) ([ ] + [ ]) / [ ] =

TERCER DÍGITO DEL CANDADO: g) * h) - i) ([ ] * [ ]) - [ ] =

-1-

DESARROLLO

EVALUACIÓN

El/la ganador/a es el/la estudiante que en primer lugar consigue descifrar la contraseña y abrir el candado. Se encuentra un "VALE" por 0,5 de calificación "extra" sobre el 100% de la calificación de la asignatura.

Para el resto de estudiantes, se calificará desde 0 a 0,3, dependiendo de los dígitos correctos de la contraseña.

En todo caso se trata de calificación "extra".

RESULTADOS

Trabajo financiado a través del Proyecto de Innovación Docente con código PID49_201617, denominado "Talleres Motivacionales para la adquisición de competencias en Estadística", concedido por la Universidad de Jaén

Software online como soporte docente e investigador

Beatriz Cobo Rodríguez[1], David Molina Muñoz[2], Ramón Ferri-García[1]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Granada
[2]Departamento de Didáctica de la Matemática, Universidad de Granada

Contacto: beacr@ugr.es

Resumen

Hoy en día los avances tecnológicos nos ayudan a la hora de llevar a cabo nuestra docencia e investigación. En este trabajo vamos a ver algunos programas online que nos ayudarán en nuestra labor diaria, podemos encontrar desde programas matemáticos, por ejemplo, Wolfram Alpha, editores de texto, ShareLaTeX, compiladores de código R, Python, C, PHP, entre otros. Wolfram Alpha tiene un amplio conocimiento y un profundo poder computacional cuando se trata de matemáticas, por lo que nos permite realizar cálculos y también nos ayuda gracias a los ejemplos que tiene sobre una gran variedad de temas. Sharelatex es un editor de Latex fácil de usar, online y colaborativo, con el cual podrás realizar tus trabajos, escribir artículos técnicos, realizar pósteres y slides utilizando LateX en la nube. En cuanto a la programación, actualmente es algo muy habitual en nuestro trabajo y podemos encontrar numerosas páginas web que nos permiten la compilación de sus códigos sin necesidad de tener instalado ningún programa en nuestro ordenador.

Palabras claves: software online, compilador online, software gratuito

Agradecimientos. Los autores agradecen el apoyo del Ministerio de Economía y Competitividad de España en el proyecto MTM2015-63609-R.

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

Software online como soporte docente e investigador

Beatriz Cobo Rodríguez[1], David Molina Muñoz[2], Ramón Ferri-García[3]

[1,3]Departamento de Estadística e Investigación Operativa, Universidad de Granada.

[2]Departamento de Didáctica de la Matemática, Campus de Ceuta, Universidad de Granada

1. Resumen

Hoy en día los avances tecnológicos nos ayudan a la hora de llevar a cabo nuestra docencia e investigación.
Son muchos los programas que nos facilitan las tareas, pero usualmente requieren de instalación en un ordenador y son de uso personal.
En este trabajo vamos a ver algunos programas online que nos ayudarán en nuestra labor diaria, podemos encontrar desde programas matemáticos, por ejemplo, Wolfram Alpha, editores de texto, ShareLaTeX, compiladores de código R, Python, C, PHP,...., incluso Telegram es capaz de ejecutar código en varios lenguajes de programación.

2. Ventajas y Desventajas

Ventajas

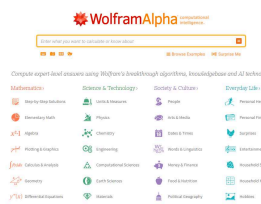
- No depende del sistema operativo, Windows, Mac, Linux, ni de las distintas versiones de este, ya que no necesitan instalaciones.
- No tienes que instalarlo, ni configurarlo, ni actualizarlo, en definitiva, no tienes que hacer nada más que registrarte y empezar a usarlo.
- Podemos acceder, tanto al programa como a los datos que guardas, desde cualquier equipo.
- El consumo de recursos lo realizan terceros por lo que lo único que nos hace falta es un navegador web, ya que la mayor parte de consumo de ciclos de procesador, memoria,...., no depende de nosotros.

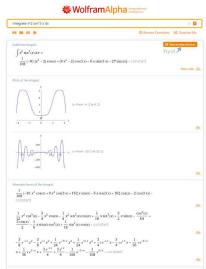
Desventajas

- Dependencia máxima de un tercero: si se cae la red o su servidor, no te sirve.
- ¿Privacidad?: todos tus datos y tu perfil de usuario junto a todas las acciones que realizas están almacenados en equipos que pueden estar comprometidos, eso pasa tanto en tu escritorio como en servidores pero en el caso del escritorio al menos puedes darte cuenta. No sabes qué intenciones pueden tener o cómo almacenan esos datos.
- Limitaciones de red y de procesamiento: por muy potentes que sean los equipos nunca te darán la misma capacidad de procesamiento que un ordenador entero para ti.

3. Wolfram Alpha

Wolfram Alpha tiene un amplio conocimiento y un profundo poder computacional cuando se trata de matemáticas, por lo que nos permite realizar cálculos y también nos ayuda gracias a los ejemplos que tiene sobre una gran variedad de temas.

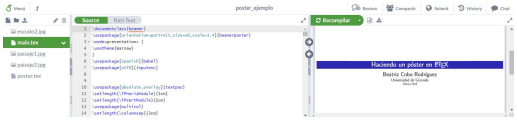




Este programa online, tiene su correspondencia en el programa Wolfram Mathematica, software propietario que se ha utilizado durante décadas en la enseñanza y aprendizaje de estudiantes de ingeniería, matemáticas,... Actualmente también cuenta con una versión online.

4. ShareLaTex

Sharelatex es un editor de Latex fácil de usar, online y colaborativo, con el cual podrás realizar tus trabajos, escribir artículos técnicos, realizar posters y slides utilizando LateX en la nube.
En 2017, ShareLaTeX fue adquirida por Overleaf, la cual planea continuar con la funcionalidad anterior, incluyendo además servicios Premium (revisión de artículos por personas de lengua inglesa, peer review de artículos antes de ser enviados a revistas, servicios de edición para libros y publicación directa en varios revistas Open Access).



5. Compiladores online

En cuanto a la programación, actualmente es algo muy habitual en nuestro trabajo y podemos encontrar numerosas páginas web que nos permiten la compilación de sus códigos sin necesidad de tener instalado ningún programa en nuestro ordenador.
Podemos encontrar páginas:

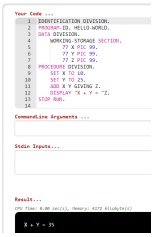
- multicódigo, como por ejemplo, paiza.io, rextester.com, joodle.com, que nos permites trabajar con diversos lenguajes de programación (C, C, C++, Cobol, D, Java, JavaScript, MySQL, Perl, PHP, Python2, Python3, R, VB,...).
- código R, por ejemplo, rdrr.io.

5.1 Multicódigo

paiza.io

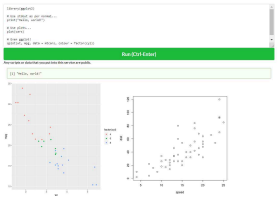
rextester.com

joodle.com



5.2 Lenguaje R

rdrr.io



5.3 Telegram

TeleMath le permite enviar comandos a los servidores en la nube, ejecutando un potente software matemático, y luego recuperar los resultados de cómputo, con mensajería instantánea.
TeleMath utiliza Telegram para enviar y recibir mensajes. Funciona en su teléfono, escritorio o navegador.
Entornos de código libre disponibles:

- GNU Octave
- Python + NumPy + Matplotlib
- The R Project for Statistical Computing

(usuario público @R_uji_bot)

Referencias

- https://paiza.io/es/projects/new
- https://rextester.com
- https://jdoodle.com
- https://rdrr.io/snippets/
- https://sites.google.com/uji.es/telemath/inici?authuser=0

Uso de Geogebra en Programación Lineal

Miguel Angel Daza[1], Aurora Llin Pérez[2]

[1]Departamento de Estadística, Universidad Carlos III de Madrid
[2]CEIPSO, El Encinar

Contacto: mdaza@est-econ.uc3m.es

Resumen

A la hora de impartir docencia cada vez se hace más necesario disponer de herramientas que ayuden en dicho proceso. Una de esas herramientas es Geogebra[1], que dispone de múltiples y variadas opciones. De entre todas las que dispone, en este caso, se va a hacer uso de su gran poder de representación gráfica, así como de la facilidad con que es capaz de situar intersecciones de rectas, planos, definir puntos y, al final, aplicar funciones a esos puntos. Geogebra, como software, dispone de versiones instalables en PCs pero, donde se ha podido comprobar su potencia ha sido con la versión on-line disponible. Como principal ventaja tiene el ser accesible desde cualquier ordenador conectado a Internet, lo que facilita el trabajo en aulas/laboratorios de informática en los que diariamente se reinician sus equipos. Además, todos los alumnos acceden a la misma versión del software. Otra ventaja, es que existe la posibilidad de crear grupos. Un grupo es un espacio donde sus miembros podrán acceder a las distintas aplicaciones Geogebra que se hayan ido creando y compartiendo, estando siempre disponibles en ese espacio común. Como ejemplo de uso se mostrará la aplicación de esta herramienta a la hora de enseñar la forma de resolver (gráficamente) problemas de programación lineal en la asignatura de Optimización y Simulación en la Empresa del Grado de Finanzas y Contabilidad (Universidad Carlos III de Madrid).

Palabras claves: Geogebra, Programación Lineal, LP

[1]GeoGebra es un software de matemáticas dinámicas para todos los niveles educativos que reúne geometría, álgebra, hoja de cálculo, gráficos, estadística y cálculo en un solo programa fácil de usar. GeoGebra es también una comunidad en rápida expansión, con millones de usuarios en casi todos los países. GeoGebra se ha convertido en el proveedor líder de software de matemática dinámica, apoyando la educación en ciencias, tecnología, ingeniería y matemáticas (STEM: Science Technology Engineering & Mathematics) y la innovación en la enseñanza y el aprendizaje en todo el mundo. Fuente: https://www.geogebra.org/about consultado el 5 de marzo de 2019

Uso de GeoGebra en Programación Lineal

Universidad Carlos III

Miguel Ángel Daza Arbolí- mdaza@est-econ.uc3m.es, UC3m

María Aurora Llin Pérez- aurora.llinperez@educa.madrid.org,

CEIPSO "El Encinar" Torrelodones

1.- INTRODUCCIÓN

A la hora de impartir docencia cada vez se hace más necesario disponer de herramientas que ayuden en dicho proceso. Una de esas herramientas es "**GeoGebra**", que dispone de múltiples y variadas opciones. De entre todas las que dispone, en este caso, se va a hacer uso de su gran poder de **representación gráfica**, así como de la facilidad con que es capaz de situar intersecciones de rectas, planos, definir puntos y, al final, aplicar funciones a esos puntos.

En este trabajo se quiere trasladar a los alumnos universitarios la herramienta utilizada en el CEIPSO "El Encinar" para la resolución de inecuaciones gráficamente.

Fig. 1 Logo GeoGebra

2.- REQUISITOS

GeoGebra, dispone de versiones instalables en PCs pero, donde se ha podido comprobar su potencia ha sido con la **versión "on-line"** disponible.

https://www.geogebra.org/graphing?lang=es

Además de ser accesible desde cualquier ordenador conectado a Internet, permite la posibilidad de crear grupos. Un grupo es un espacio donde sus miembros podrán acceder a las distintas aplicaciones GeoGebra.

3.- CASO DE USO

En este caso se va a solucionar, con la ayuda de GeoGebra, el ejemplo 3.1 "Planning Desserts" del libro "*Practical management science*", libro de texto de la asignatura de *Optimización y Simulación en la Empresa* del *grado de Finanzas y Contabilidad* de la *Universidad Carlos III*.

* Winston, W. L., Albright, S. C., Broadie, M. N., Lapin, L. L., & Whisler, W. D. (2007). Practical management science. Thomson South-Western.

4.- PROCEDIMIENTO

La formulación matemática del problema es:

$$\text{Maximize } 3145x + 6175y \quad \textbf{Función objetivo}$$

$$\begin{aligned} &s.a.: \\ &120x + 160 \leq 450 && \textbf{Restricción 1} \\ &5x + 10 \leq 25 && \textbf{Restricción 2} \\ &37x + 65y \geq 120 && \textbf{Restricción 3} \\ &x, y \geq 0 && \textbf{Restricciones 4 y 5} \end{aligned}$$

Fig. 2 Página inicial GeoGebra Calculadora Gráfica

Paso 1. Restricción 1

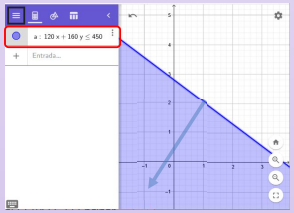


Fig. 3 Imagen restricción 1 color azul

Paso 2. Restricción 2

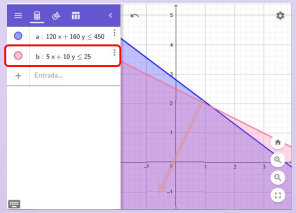


Fig. 4 Imagen restricción 2 color rosa

Paso 3. Restricción 3

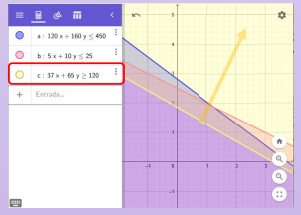


Fig. 5 Imagen restricción 3 color amarillo

Paso 4. Restricciones 4 y5

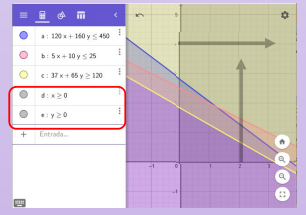


Fig. 6 Imagen restricciones 4 y 5 color gris

5.- RESULTADOS

Paso 5. Región factible

GeoGebra ha ido nombrado cada una de la restricciones con una letra minúscula. Ahora se le pide la intersección de todas esas regiones, para ello se usa **&&** que genera un ***y lógico***

p: a && b && c && d && e

Como resultado se muestra la región factible en un tono verde

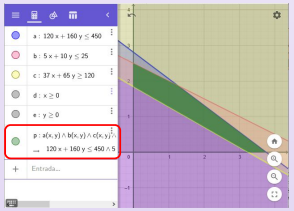


Fig. 7 región factible (color verde)

Paso 6. Calcular los Vértices

La región factible al ser un polígono se le pueden calcular los vértices con GeoGebra.

Vértices(p)

GeoGebra calcula todos los vértices, por lo que habrás que desactivar los que no pertenezcan a la región factible.

En este caso se han quedado los vértices A,D,F,H,I

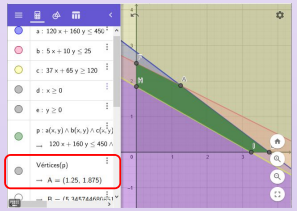


Fig. 8 vértices región factible (color gris)

Paso 7. Función Objetivo y Vértices

En GeoGebra se puede definir la función objetivo como una función matemática y pasarle como parámetros los vértices previamente calculados.

f(x,y)=3145x+6175y

Luego, para pasar el valor de los vértices f(A), f(D), ...

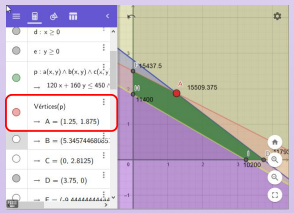


Fig. 9 vértices con valor función objetivo. En rojo, el vértice óptimo

6.- CONCLUSIONES

Una vez más se pone de manifiesto la versatilidad de la aplicación GeoGebra. Pero, además, todo el ejercicio queda grabado en un espacio en la nube, con posibilidad de ser compartido y a disposición de todos los alumnos.

Ver enlace QR

Fig. 10 código QR con acceso directo al ejercicio en el servidor de GeoGebra

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa (Granada, 4-5 abril 2019)

2019

Generación de cuestionarios de autoevaluación de Moodle mediante el paquete exams de R

Manuel Escabias[1], Ana M. Aguilera[1]

[1]Departamento de Estadístiaca e I.O., Universidad de Granada

Contacto: escabias@ugr.es

Resumen

En los últimos años está tomando mucha relevancia en nuestras universidades la docencia virtual. Cada vez son más las universidades con estudios oficiales de grado y posgrado impartidos en modalidad virtual. La docencia propiamente dicha tiene superados los retos que supone la enseñanza virtual con el uso de materiales impresos (apuntes en pdf, libros electrónicos,...), vídeos tutoriales, clases o tutorías virtuales en streaming con plataformas como adobe connect. Sin embargo, es en la evaluación donde la enseñanza virtual tiene todavía muchos retos que resolver. En este trabajo se muestra una iniciativa utilizada en la asignatura Modelos de respuesta discreta del Máster Oficial en Estadística Aplicada de la UGR, para la evaluación del alumnado utilizando cuestionarios de moodle auto-evaluables, generados con el paquete exams de R que permiten tanto selección aleatoria de preguntas con un mismo conjunto de datos, como fijación de preguntas con generación aleatoria de conjuntos de datos.

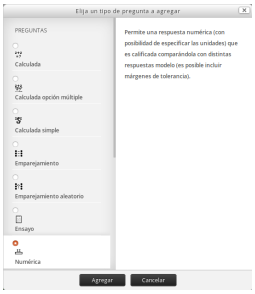
Palabras claves: Moodle, cuestionarios de autoevaluación, paquete exams de R

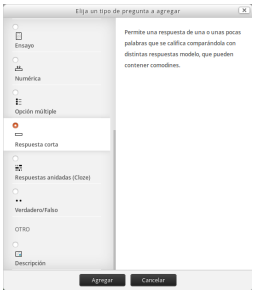
Bancos de preguntas de Moodle

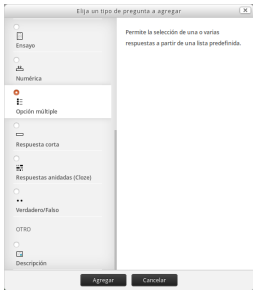
Las plataformas moodle permiten crear bancos de preguntas de evaluación, bien para exámenes, bien para test de autoevaluación. Algunos tipos de preguntas posibles son:

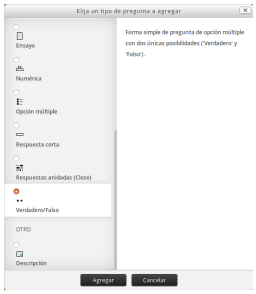
- Numérica, con posibilidad de especificar las unidades. Se califica comparándola con distintas respuestas modelo. Es posible incluir márgenes de tolerancia. Existen otras versiones de esta: calculada, calculada opción múltiple y calculada simple.
- Respuesta corta. Permite una respuesta de una o pocas palabras. Se califica comparándola con distintas respuestas modelo, que pueden tener comodines. Versiones más extensas se permiten con las de tipo ensayo, que permite una respuesta de unas pocas frases o párrafos y se califica manualmente.

- Opción múltiple. Permite la selección de una o varias respuestas a partir de una lista predefinida. Otras alternativas son emparejamiento y emparejamiento aleatorio
- Verdadero/Falso.









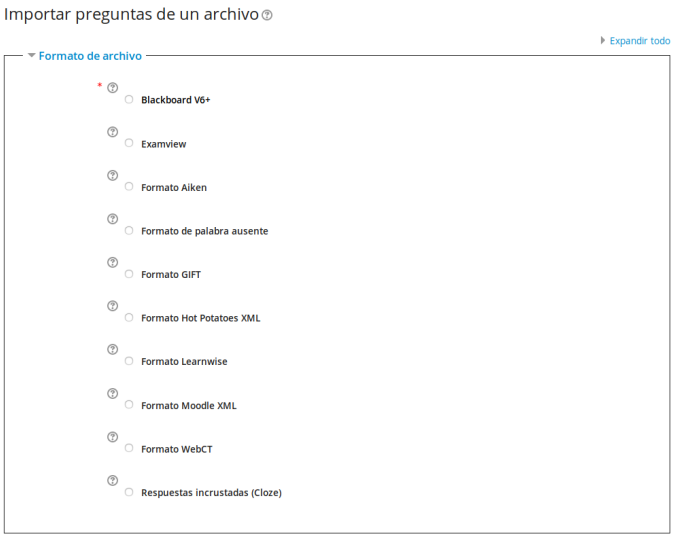
Cuestionarios de Moodle

Permite diseñar y plantear cuestionarios con preguntas escogidas del banco de preguntas.

- Se pueden permitir varios intentos o intentos únicos.
- Las preguntas pueden estar ordenadas o seleccionadas al azar.
- Se puede establecer un tiempo límite para la respuesta.
- Cada intento se califica automáticamente y el resultado se guarda en el libro de calificaciones.
- Cuando se permiten múltiples intentos, hay diferentes maneras en las que puede usar las distintas calificaciones obtenidas para calcular la puntuación final: calificación más alta, puntuación media, primera calificación, última calificación.
- Se puede decidir cuándo se muestra al usuario los resultados, los comentarios de retroalimentación y las respuestas correctas.

El paquete `exams` de R

El paquete `exams` de R permite generar preguntas en el formato de Moodle para su exportación a partir de documentos Markdown que permiten insertar código R (números aleatorios, gráficos, conjuntos de datos, etc.)



Máster en Estadística aplicada: Modelos de respuesta discreta. Aplicaciones biosanitarias

Descripción de la materia y contenido teórico.

- Asignatura 100 % virtual.
- Contenido: modelos de regresión logística y de respuesta múltiple nominal y ordinal.
- Material: proporcionar material teórico en formato de apuntes y guías prácticas explicadas de ajuste de modelos con R.
- Metodología docente: secuenciación de conceptos por grado de dificultad, participación en foros y tutorías virtuales mediante plataforma moodle.
- Evaluación: cuestionarios en plataforma moodle.

Reto: que cada alumno conteste unas cuestiones diferentes para evitar la copia y que sean autoevaluables.

Propuesta: utilizar cuestionarios generados escogiendo aleatoriamente preguntas de un banco de preguntas previamente almacenado en la plataforma Moodle.

Dificultad: generar gran cantidad de preguntas.

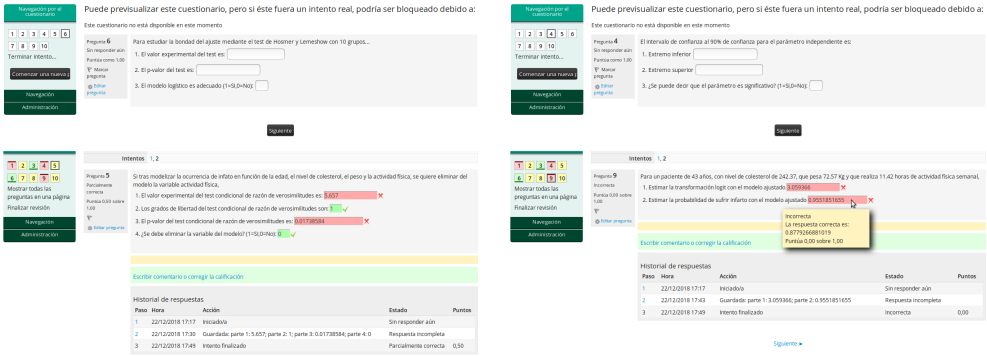
Solución: El paquete `exams` de R permite mediante aleatorización generar de manera rápida gran cantidad de preguntas.

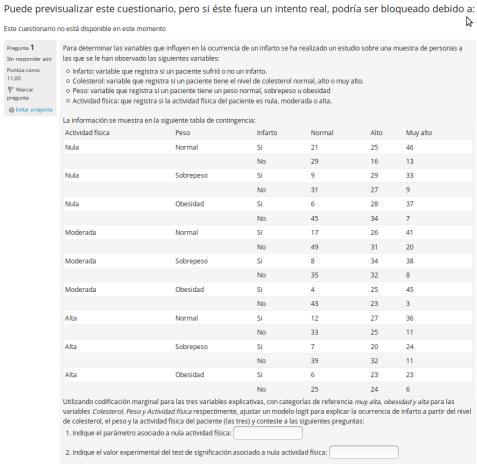
Tipos de cuestionarios y dificultades:

- Un mismo conjunto de datos para todas las preguntas, con suficientes variables que permitan generar gran cantidad de preguntas del mismo tipo para distintas variables.

- Distinto conjunto de datos y un número fijo de preguntas.

- Al permitir varios intentos las cuestiones o datos de segundos intentos varía.





Puede previsualizar este cuestionario, pero si éste fuera un intento real, podría ser bloqueado debido a:

Este cuestionario no está disponible en este momento

Pregunta 1
Sin responder aún
Puntúa como 11,00
Marcar pregunta
Editar pregunta

Para determinar las variables que influyen en la ocurrencia de un infarto se ha realizado un estudio sobre una muestra de personas a las que se le han observado las siguientes variables:

- Infarto: variable que registra si un paciente sufrió o no un infarto.
- Colesterol: variable que registra si un paciente tiene el nivel de colesterol normal, alto o muy alto.
- Peso: variable que registra si un paciente tiene un peso normal, sobrepeso u obesidad
- Actividad física: que registra si la actividad física del paciente es nula, moderada o alta.

La información se muestra en la siguiente tabla de contingencia:

Actividad física	Peso	Infarto	Normal	Alto	Muy alto
Nula	Normal	Si	21	25	46
		No	29	16	13
Nula	Sobrepeso	Si	9	29	33
		No	31	27	9
Nula	Obesidad	Si	6	28	37
		No	45	34	7
Moderada	Normal	Si	17	26	41
		No	49	31	20
Moderada	Sobrepeso	Si	8	34	38
		No	35	32	8
Moderada	Obesidad	Si	4	25	45
		No	43	23	3
Alta	Normal	Si	12	27	36
		No	33	25	11
Alta	Sobrepeso	Si	7	20	24
		No	39	32	11
Alta	Obesidad	Si	6	23	23
		No	25	24	6

Utilizando codificación marginal para las tres variables explicativas, con categorías de referencia *muy alta, obesidad y alta* para las variables *Colesterol, Peso y Actividad física* respectivamente, ajustar un modelo logit para explicar la ocurrencia de infarto a partir del nivel de colesterol, el peso y la actividad física del paciente (las tres) y conteste a las siguientes preguntas:

1. Indique el parámetro asociado a nula actividad física:
2. Indique el valor experimental del test de significación asociado a nula actividad física:

Alta	Sobrepeso	Si	14	37	47
		No	29	24	9
Alta	Obesidad	Si	11	42	43
		No	34	33	8

Utilizando codificación marginal para las tres variables explicativas, con categorías de referencia *muy alta, obesidad y alta* para las variables *Colesterol, Peso y Actividad física* respectivamente, ajustar un modelo logit para explicar la ocurrencia de infarto a partir del nivel de colesterol, el peso y la actividad física del paciente (las tres) y conteste a las siguientes preguntas:

1. Indique el parámetro asociado a un nivel de colesterol alto: -0,005853
2. Indique el valor experimental del test de significación asociado a sobrepeso: -0,764
3. Indique el p-valor del test de significación asociado a un nivel de colesterol normal: 2e-16
4. Por cuánto se multiplica la ventaja de sufrir un infarto si se tiene sobrepeso en lugar de peso normal permaneciendo el resto de variables constantes: 0,893878
5. Por cuánto se multiplica la ventaja de sufrir un infarto si se tiene actividad física alta en lugar de moderada permaneciendo el resto de variables constantes: 6,711819
6. Indique el valor experimental del test chi-cuadrado de bondad de ajuste: 327,9864
7. Indique el p-valor del test chi-cuadrado de bondad de ajuste: 0,01776848
8. ¿Sería adecuado el modelo logístico para explicar la ocurrencia de infarto a partir del nivel de colesterol, el peso y la actividad física? (1=Sí,0=No): 0
9. En la selección stepwise hacia delante se ha seleccionado la variable nivel de colesterol (1=Sí,0=No): 1
10. En la selección stepwise hacia delante se ha seleccionado la variable peso (1=Sí,0=No): 0
11. En la selección stepwise hacia delante se ha seleccionado la variable actividad física (1=Sí,0=No): 1

Escribir comentario o corregir la calificación

Historial de respuestas

Paso	Hora	Acción	Estado	Puntos
1	2/02/2019 13:17	Iniciado/a	Sin responder aún	
2	2/02/2019 14:15	Guardada: parte 1: -0,005853; parte 2: -0,764; parte 3: 2e-16; parte 4: 0,893878; parte 5: 6,711819; parte 6: 327,9864; parte 7: 0,01776848; parte 8: 0; parte 9: 1; parte 10: 0; parte 11: 1	Respuesta guardada	
3	2/02/2019 14:16	Intento finalizado	Parcialmente correcta	9,00

Referencias

[1] Blackboard, Inc. (2016). *Blackboard Help: Question types*. http://en-us.help.blackboard.com/Learn/Instructor/Tests_Pools_Surveys/040_Question_Types

[2] IMS Global Learning Consortium, Inc. (2012). *IMS Question & Test Interoperability: ASI XML Binding Specification Final Specification Version 1.2*. http://www.imsglobal.org/question/qtiv1p2/imsqti_asi_bindv1p2.html

[3] Zeileis A, Umlauf N y Leisch F (2014). Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond. *Journal of Statistical Software*, **58(1)**, pp. 1-36.

[4] MoodleDocs (2015). *Moodle XML Format*. http://docs.moodle.org/en/Moodle_XML

Generación de cuestionarios de autoevaluación de Moodle mediante el paquete exams de R

Manuel Escabias Machuca y Ana M. Aguilera Del Pino
Universidad de Granada
Abril, 2019

Introducción

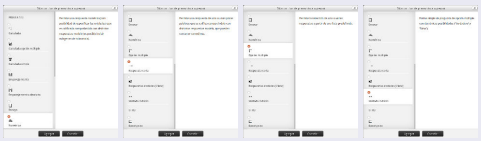
En los últimos años está tomando mucha relevancia en nuestras universidades la docencia virtual. Cada vez son más las universidades con estudios oficiales de grado y posgrado impartidos en modalidad virtual. La docencia propiamente dicha tiene superados los retos que supone la enseñanza virtual con el uso de materiales âimpresosâ (apuntes en pdf, libros electrónicos,...), vídeos tutoriales, clases o tutorías virtuales en streaming con plataformas como adobe connect. Sin embargo, es en la evaluación donde la enseñanza virtual tiene todavía muchos retos que resolver. En este trabajo se muestra una iniciativa utilizada en la asignatura Modelos de respuesta discreta del Máster Oficial en Estadística Aplicada de la UGR, para la evaluación del alumnado utilizando cuestionarios de moodle auto-evaluables, generados con el paquete exams de R que permiten tanto selección aleatoria de preguntas con un mismo conjunto de datos, como fijación de preguntas con generación aleatoria de conjuntos de datos.

Palabras clave: Moodle, cuestionarios de autoevaluación, paquete exams de R

Bacos de preguntas de Moodle

Bancos de preguntas. Las plataformas moodle permiten crear bancos de preguntas de evaluación, bien para exámenes, bien para test de autoevaluación. Algunos tipos de preguntas posibles son:

- Numérica, con posibilidad de especificar las unidades. Se califica comparándola con distintas respuestas modelo. Es posible incluir márgenes de tolerancia. Existen otras versiones de esta: calculada, calculada opción múltiple y calculada simple.
- Respuesta corta. Permite una respuesta de una o pocas palabras. Se califica comparándola con distintas respuestas modelo, que pueden tener comodines. Versiones más extensas se permiten con las de tipo ensayo, que permite una respuesta de unas pocas frases o párrafos y se califica manualmente.
- Opción múltiple. Permite la selección de una o varias respuestas a partir de una lista predefinida. Otras alternativas son emparejamiento y emparejamiento aleatorio
- Verdadero/Falso.



Cuestionarios de Moodle

Permite diseñar y plantear cuestionarios con preguntas escogidas del banco de preguntas.

- Se pueden permitir varios intentos o intentos únicos
- Las preguntas pueden estar ordenadas o seleccionadas al azar.
- Se puede establecer un tiempo límite para la respuesta.
- Cada intento se califica automáticamente y el resultado se guarda en el libro de calificaciones.
- Cuando se permiten múltiples intentos, hay diferentes maneras en las que puede usar las distintas calificaciones obtenidas para calcular la puntuación final: calificación más alta, puntuación media, primera calificación, última calificación.
- Se puede decidir cuándo se muestra al usuario los resultados, los comentarios de retroalimentación y las respuestas correctas.

El paquete exams de R

El paquete exams de R permite generar preguntas en el formato de Moodle para su exportación a partir de documentos Markdown que permiten insertar código R (números aleatorios, gráficos, conjuntos de datos, etc.)

Máster en Estadística aplicada: Modelos de respuesta discreta. Aplicaciones biosanitarias

Descripción de la materia y contenido teórico

- Asignatura 100% virtual
- Contenido: modelos de regresión logística y de respuesta múltiple nominal y ordinal
- Material: proporcionar material teórico en formato de apuntes y guías prácticas explicadas de ajuste de modelos con R
- Metodología docente: secuenciación de conceptos por grado de dificultad, participación en foros y tutorías virtuales mediante plataforma moodle.
- Evaluación: cuestionarios en plataforma moodle.

Reto: que cada alumno conteste unas cuestiones diferentes para evitar la copia y que sean autoevaluables.

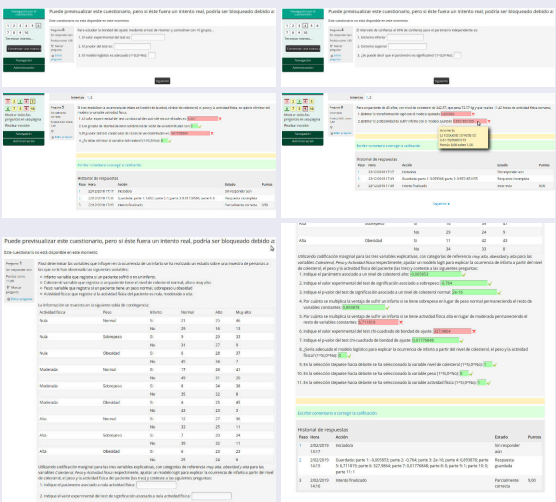
Propuesta: utilizar cuestionarios generados escogiendo aleatoriamente preguntas de un banco de preguntas previamente almacenado en la plataforma Moodle

Dificultad: generar gran cantidad de preguntas

Solución: El paquete exams de R permite mediante aleatorización generar de manera rápida gran cantidad de preguntas

Tipos de cuestionarios y dificultades:

- Un mismo conjunto de datos para todas las preguntas, con suficientes variables que permitan generar gran cantidad de preguntas del mismo tipo para distintas variables
- Distinto conjunto de datos y un número fijo de preguntas.
- Al permitir varios intentos las cuestiones o datos de segundos intentos varía.



Bibliografía

- Blackboard, Inc. (2016). Blackboard Help: Question types. http://en-us.help.blackboard.com/Learn/Instructor/Tests_Pools_Surveys/040_Question_Types
- IMS Global Learning Consortium, Inc. (2012). IMS Question & Test Interoperability: ASI XML Binding Specification Final Specification Version 1.2. http://www.imsglobal.org/question/qtiv1p2/imsqti_asi_bindv1p2.html
- 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), 1-36.
- MoodleDocs (2015). Moodle XML Format. http://docs.moodle.org/en/Moodle_XML
- 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), 1-36.

Intervalos de tolerancia en Grados en Ingenierías: un enfoque práctico

Jaione Etxeberria[1], Tomás Goicoa[1], M.D., Paula Camelia Trandafir[1], Aritz Adin[1], Guzmán Santafé[1], M.D. Ugarte[1]

[1]Departamento de Estadística, Informática y Matemáticas, Universidad Pública de Navarra

Contacto: tomas.goicoa@unavarra.es

Resumen

En este trabajo se presenta la propuesta de una práctica de ordenador de intervalos distintos a los clásicos intervalos de confianza (IC). Se trata de los denominados intervalos de tolerancia (IT) que son muy útiles en el ámbito de la ingeniería. Mientras que los IC proporcionan una estimación por intervalo del parámetro poblacional de interés, los IT permiten estimar entre qué valores se encuentra un determinado porcentaje de la población con un nivel de confianza preestablecido. Los intervalos de tolerancia se aplican dentro de la industria, en especial donde existen procesos de fabricación en serie. Debido a inexactitudes que puedan existir en los procesos de fabricación, puede haber cierta variación en las medidas de las piezas. Cuando éstas vayan a ser empleadas en un montaje posterior, es necesario tener un control de sus dimensiones. Es decir, debemos tener cierta confianza de que la mayoría de nuestras piezas siguen las especificaciones requeridas. En este trabajo se propone una práctica novedosa para alumnos/as de los grados en ingenierías con distintos problemas reales relacionados con intervalos de tolerancia y su resolución en R mediante la librería tolerance.

Palabras claves: Estimación por Intervalos, Intervalos de Tolerancia, Estadística en Ingenierías

Intervalos de tolerancia en Grados en Ingenierías: un enfoque práctico

Etxeberria J, Goicoa T, Trandafir PC, Adin A, Santanfé G, y Ugarte, MD

Departamento de Estadística, Informática y Matemáticas, InaMat, Universidad Pública de Navarra.
jaione.etxeberria@unavarra.es, tomas.goicoa@unavarra.es, camelia@unavarra.es,
aritz.adin@unavarra.es, guzman.santafe@unavarra.es, lola@unavarra.es

Introducción a los Intervalos de Tolerancia y sus Aplicaciones en Ingeniería

En este trabajo se presenta la propuesta de una práctica de ordenador de intervalos distintos a los clásicos intervalos de confianza (IC). Se trata de los denominados intervalos de tolerancia (IT) que son muy útiles en el ámbito de la ingeniería.

- Los intervalos de tolerancia se aplican dentro de la industria, en especial donde existe una **fabricación de piezas o productos en serie**
- Los procesos de fabricación no son perfectos y por tanto es esperable cierta variabilidad en las medidas de las piezas fabricadas
- Cuando las piezas vayan a ser empleadas en procesos posteriores, es necesario tener un control de las dimensiones. Es decir, **debemos tener cierta confianza de que la mayoría de nuestras piezas sigan las especificaciones requeridas**

IT vs IC

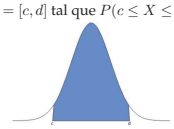
Sea X una población que sigue una distribución $N(\mu, \sigma)$

Intervalos de Confianza: Proporciona una estimación por intervalos del parámetro poblacional de interés.

$$\text{IC}_{1-\alpha}(\mu) = [\bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n}}]$$

Intervalos de Tolerancia: Permiten estimar entre qué valores se encuentra el $100 \times k\%$ de la población con una determinada confianza

$$\text{IT}_{1-\alpha} = [c, d] \text{ tal que } P(c \leq X \leq d) = k$$



Ejemplo Motivación

Supongamos que fabricamos piezas de forma cilíndrica. Nos interesa saber **entre qué valores se encuentra el diámetro del** 90% **de las piezas.**
Si denotamos por X al diámetro de las piezas, $X \sim N(\mu = 30, \sigma = 2)$, el 90% de las piezas tienen un diámetro comprendido entre 26.71 y 33.29 mm.

En la realidad, ¿qué sucede si se desconocen los valores de μ y σ?

- Podemos tomar una muestra de tamaño n y calcular $\bar{x}$ y s. Después construir $[\bar{x} - 1.645s, \bar{x} + 1.645s]$
- Debido a la variabilidad en las estimaciones de μ y σ, hay posibilidades de que el intervalo anterior incluya **menos del** 90% **de los valores de la población**.
- Para tener una confianza a priori del $100 \cdot (1-\alpha)\%$ de que el intervalo resultante incluya al menos el 90% de los valores de la población, se puede reemplazar 1.645 por un valor más grande, denotado por C.

Intervalo de Tolerancia

$$IT_{1-\alpha} = \bar{x} \pm C \times s$$

C= Valor crítico de tolerancia $= z_{(1+p)/2}\sqrt{\frac{(n-1)(1+\frac{1}{n})}{\chi^2_{\alpha;n-1}}}$ (Howe,1969); n =tamaño muestral; p = proporción de población que se desea cubrir; $1-\alpha$ = nivel de confianza

En el ejemplo de motivación, para una muestra de $n = 20$ y $1-\alpha = 0.95$, $C = z_{0.95}\sqrt{\frac{(20-1)(1+\frac{1}{20})}{\chi^2_{0.05;20-1}}} = 2.310$ y el intervalo de tolerancia resultante $\bar{x} \pm 2.310 \times s$

Problema propuesto a alumnos/as de ingeniería

- Una línea de montaje automatizada está programada para producir **rodamientos de bolas** con un diámetro nominal de 10 mm
- La sección de control de calidad quiere inspeccionar la precisión con la que se producen las piezas ya que, los rodamientos con un diámetro superior a 10 ± 0.3 mm serán descartados del proceso de producción posterior. El departamento de control de calidad estará satisfecho si al menos el 98% de la piezas son válidas para el proceso de producción posterior
- Con una confianza del 95%, construye el intervalo de tolerancia adecuado para dar respuesta a esta cuestión. Para ello, considera el fichero `Rodamientos` en el que se recoge el diámetro de 1000 rodamientos

SOLUCIÓN

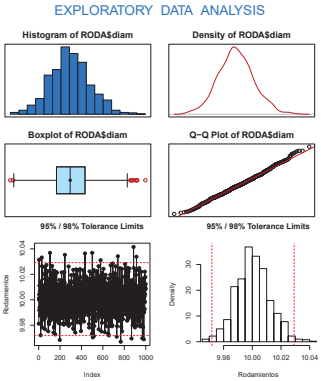
Paso 1: Cargamos en conjunto de datos `Rodamientos` y realizamos el estudio de la normalidad de los diámetros mediante la función `eda` de la librería `PASWR2` de R

```
library (PASWR2)
RODA=read.table(file="Rodamientos.txt")
eda(RODA$diam)
```

Paso 2: Calculamos el intervalo de Intervalo de Tolerancia al 95% utilizando la librería `tolerance` de R

```
library(tolerance)
normtol.int(x = RODA$diam, alpha = 0.05, P = 0.98, side = 2)
  alpha    P    x.bar    2-sided.lower  2-sided.upper
  0.05  0.98  10.0005     9.971842       10.02916
```

Paso 3. Conclusión: Tenemos una confianza del 95% de que al menos el 98% de los rodamientos tendrán un diámetro comprendido entre [9.971842, 10.02916]. Por tanto, tenemos una confianza del 95% de que al menos el 98% de la piezas serán válidas para el proceso de producción posterior



Referencias

Ugarte MD, Militino AF, and Arnholt AT. (2016). **Probability and Statistics with R**, Second Edition. Chapman and Hall, Boca Ratón.
Howe, WG(1969). **Two-sided tolerance limits for normal populations—some improvements**. Journal of the American Statistical Association, 64(326), 610-620.

Producción automatizada de problemas de cálculo de probabilidades en Sweave

Ramón Ferri-García[1], Beatriz Cobo Rodríguez[1], David Molina Muñoz[2]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Granada
[2]Departamento de Didáctica de la Matemática, Universidad de Granada

Contacto: rferri@ugr.es

Resumen

En este trabajo se propone la utilización de scripts automatizados escritos en la extensión Sweave de RStudio con el objetivo de combinar el procesamiento de textos de LaTeX con las herramientas estadísticas de R para la producción automatizada de problemas de cálculo de probabilidades. Estos scripts permiten producir, bajo una misma plantilla, una batería de ejercicios cuya solución es numéricamente distinta para cada uno de ellos. También es posible combinar diversas plantillas, eligiéndose una de ellas al azar para cada ejercicio, para producir así soluciones distintas en su naturaleza. Las implicaciones del uso de estos scripts son diversas. Por un lado, permite reducir parte de los problemas de plagio en la entrega de ejercicios propuestos ya que éstos dejan de tener una solución única conjunta. Por otra parte, la producción automatizada de sus soluciones permite que el alumnado tenga acceso a una ingente cantidad de problemas resueltos de un tipo específico.

Palabras claves: Sweave, cálculo de probabilidades, automatización

Producción automatizada de problemas de cálculo de probabilidades en Sweave

Ramón Ferri García[1], Beatriz Cobo Rodríguez[1], David Molina Muñoz[2]

[1] Departamento de Estadística e Investigación Operativa, Universidad de Granada

[2] Departamento de Didáctica de la Matemática, Facultad de Educación, Economía y Tecnología de Ceuta, Universidad de Granada

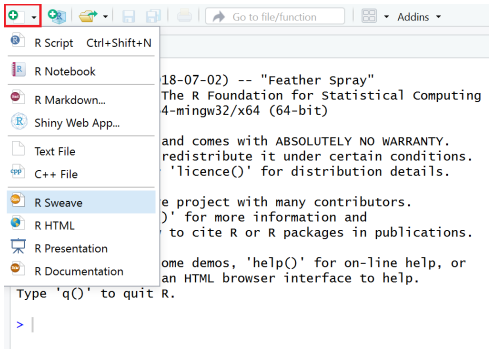
Resumen

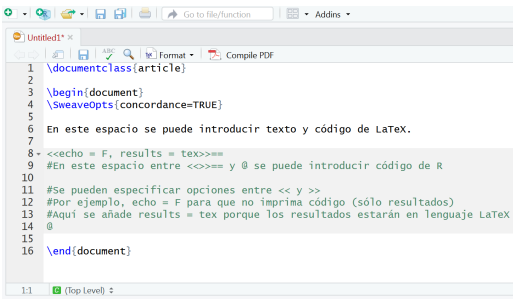
En este trabajo se propone la utilización de scripts automatizados escritos en la extensión Sweave de RStudio con el objetivo de combinar el procesamiento de textos de LaTeX con las herramientas estadísticas de R para la producción automatizada de problemas de cálculo de probabilidades. Estos scripts permiten producir, bajo una misma plantilla, una batería de ejercicios cuya solución es numéricamente distinta para cada uno de ellos. También es posible combinar diversas plantillas, eligiéndose una de ellas al azar para cada ejercicio, para producir así soluciones distintas en su naturaleza.

Las implicaciones del uso de estos scripts son diversas. Por un lado, permite reducir parte de los problemas de plagio en la entrega de ejercicios propuestos ya que éstos dejan de tener una solución única conjunta. Por otra parte, la producción automatizada de sus soluciones permite que el alumnado tenga acceso a una ingente cantidad de problemas resueltos de un tipo específico.

Sweave en RStudio

Apertura de nuevo documento de Sweave en RStudio





Los scripts para producir ejercicios de manera aleatorizada se implementan en los espacios dedicados al código con la ayuda de las funciones contenidas en el paquete ***xtable*** de R (Dahl et al., 2018). Por ejemplo, la impresión en código de LaTeX de las 6 primeras filas de la tabla *iris* de Ronald Fisher sería:

```
> install.packages("xtable")
> library(xtable)
> xtable( head(iris) )
% latex table generated in R 3.5.1 by xtable 1.8-3 package
\begin{table}[ht]
\centering
\begin{tabular}{rrrrrl}
 \hline
& Sepal.Length & Sepal.Width & Petal.Length & Petal.Width & Species
\\
 \hline
1 & 5.10 & 3.50 & 1.40 & 0.20 & setosa \\
2 & 4.90 & 3.00 & 1.40 & 0.20 & setosa \\
3 & 4.70 & 3.20 & 1.30 & 0.20 & setosa \\
4 & 4.60 & 3.10 & 1.50 & 0.20 & setosa \\
5 & 5.00 & 3.60 & 1.40 & 0.20 & setosa \\
6 & 5.40 & 3.90 & 1.70 & 0.40 & setosa \\
 \hline
\end{tabular}
\end{table}
```

	Sepal. Length	Sepal. Width	Petal. Length	Petal. Width	Species
1	5.10	3.50	1.40	0.20	setosa
2	4.90	3.00	1.40	0.20	setosa
3	4.70	3.20	1.30	0.20	setosa
4	4.60	3.10	1.50	0.20	setosa
5	5.00	3.60	1.40	0.20	setosa
6	5.40	3.90	1.70	0.40	setosa

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., Swinton, J. (2018). xtable: Export Tables to LaTeX or HTML. R package version 1.8-3. https://CRAN.R-project.org/package=xtable

Con la ayuda de la función *cat*, que en la práctica sirve para imprimir mensajes en pantalla en R, se imprime el código LaTeX directamente en el documento que luego compila Sweave.

Ejemplo práctico: cálculo de probabilidades en distribuciones

```
tablas <- list()
set.seed(1)
for(i in 1:length(dnis)){
  n <- sample(5:15,1)
  p <- sample(seq(0.3,0.6,by=0.05),1)
  r <- sample(2:5,1)
  N <- sample((n+5):30,1)
  percentiles <- sort(sample(seq(0.05,0.95,by=0.05),3,replace=F))

  c1<-sample(1:min(c(round(N*p,0)-1,n-1)),1)
  c2<-sample(3:min(c(round(N*p,0)-1,n-1)),1)
  c3<-sample(2:min(c(round(N*p,0)-1,n-1)),1)

  p2 <- round(1/c3,2)

  filas <- c(paste("B(" ,n,", ",p,")",sep=""),
             paste("P(",n*p,")", sep=""),
             paste("BN(",r,", ",p,")", sep=""),
             paste("G(",p2,")", sep=""),
             paste("H(",N,", ",round(N*p,0),", ",n,")", sep="") )

  columnas <- c(paste("$P(X = ",c1,")$", sep=""),
                paste("$P(X \\leq ",c2,")$", sep=""),
                paste("$P(X > ",c3,")$", sep=""),
                paste("Percentiles ",percentiles[1],", ",
                      percentiles[2],", y ",percentiles[3], sep=""),
                "Rango Intercuartílico")

  mat_ps<-data.frame(matrix(NA,ncol=5,nrow=5))
  colnames(mat_ps)<-columnas
  row.names(mat_ps)<-filas

  tablas[[i]] <- mat_ps
}
```

Bucle sobre DNIs de estudiantes con semilla para replicar las soluciones

Generación de parámetros aleatorios para el i-ésimo DNI

Edición:
- Filas en las que se plasmarán las distribuciones y sus parámetros
- Columnas donde se expresarán las medidas y probabilidades pedidas (aleatorias para cada DNI)

```
for(i in 1:length(dnis)){
  cat("\\textbf{DNI/PASAPORTE NÚMERO ",as.character(dnis[i]),"}")
  cat("Calcular, para cada una de las distribuciones especificadas en las filas,
       los valores de probabilidades, percentiles y rango intercuartílico
       solicitados en las columnas de la siguiente tabla")
  print(xtable(tablas[[i]], align="|c|c|c|c|c|c|"),
        floating = T, table.placement = "H", include.rownames=T,
        sanitize.text.function=identity, type="latex")

  if(i%%2==0){
    cat("\\newpage")
  }
}
```

Bucle para crear un enunciado con una tabla distinta para cada DNI.

Ejemplar de muestra:

DNI/PASAPORTE NÚMERO 0*******

Calcular, para cada una de las distribuciones especificadas en las filas, los valores de probabilidades, percentiles y rango intercuartílico solicitados en las columnas de la siguiente tabla

	$P(X=4)$	$P(X \leq 5)$	$P(X>2)$	Percentiles 0.2, 0.85 y 0.9	Rango Intercuartílico
B(7, 0.4)					
P(2.8)					
BN(4, 0.4)					
G(0.5)					
H(29, 12, 7)					

Programación de soluciones:

```
mat_ps <- data.frame(paste(round(c(dbinom(c1,n,p),
                  dpois(c1,n*p),
                  dnbinom(c1,r,p),
                  dgeom(c1,p2),
                  dhyper(c1,round(N*p,0),N-round(N*p,0),n)),4)),
                  paste(round(c(pbinom(c2,n,p),
                  ppois(c2,n*p),
                  pnbinom(c2,r,p),
                  pgeom(c2,p2),
                  phyper(c2,round(N*p,0),N-round(N*p,0),n)),4)),
                  paste(round(c(1-pbinom(c3,n,p),
                  1-ppois(c3,n*p),
                  1-pnbinom(c3,r,p),
                  1-pgeom(c3,p2),
                  1-phyper(c3,round(N*p,0),N-round(N*p,0),n)),4)),
                  c(paste(dbinom(percentiles,n,p),collapse=", "),
                  paste(dpois(percentiles,n*p),collapse=", "),
                  paste(dnbinom(percentiles,r,p),collapse=", "),
                  paste(dgeom(percentiles,p2),collapse=", "),
                  paste(dhyper(percentiles,round(N*p,0),N-round(N*p,0),n),collapse=",")),
                  c(qbinom(0.75,n,p)-qbinom(0.25,n,p),
                  qpois(0.75,n*p)-qpois(0.25,n*p),
                  qnbinom(0.75,r,p)-qnbinom(0.25,r,p),
                  qgeom(0.75,p2)-qgeom(0.25,p2),
                  qhyper(0.75,round(N*p,0),N-round(N*p,0),n)-qhyper(0.25,round(N*p,0),N-round(N*p,0),n)))
```

1ª columna

2ª columna

3ª columna

4ª columna

5ª columna

La matriz mat_ps, que para el enunciado se dejó vacía, se rellena con las soluciones de cada apartado

DNI/PASAPORTE NÚMERO 0*******

	$P(X=4)$	$P(X \leq 5)$	$P(X>2)$	Percentiles 0.2, 0.85 y 0.9	Rango Intercuartílico
B(7, 0.4)	0.1935	0.9812	0.5801	2, 4, 4	2.00
P(2.8)	0.1557	0.9349	0.5305	1, 5, 5	2.00
BN(4, 0.4)	0.1161	0.5174	0.8208	3, 10, 11	5.00
G(0.5)	0.0312	0.9844	0.125	0, 2, 3	1.00
H(29, 12, 7)	0.2157	0.9894	0.6307	2, 4, 4	2.00

Ramón Ferri García
rferri@ugr.es
Departamento de Estadística e Investigación Operativa
Universidad de Granada

Beatriz Cobo Rodríguez
beacr@ugr.es
Departamento de Estadística e Investigación Operativa
Universidad de Granada

David Molina Muñoz
dmolinam@ugr.es
Departamento de Didáctica de la Matemática
Facultad de Educación, Economía y Tecnología de Ceuta, Universidad de Granada

Realización de test de evaluación utilizando la plataforma ILIAS

Irene García-Garrido[1], Antonio Jesús López-Montoya[1], Pedro Antonio Cano-Chica[1], Valentina Cueva-López[1]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Jaén

Contacto: iggarrid@ujaen.es

Resumen

El Espacio Europeo de Educación Superior conlleva el proceso de evaluación continua para los estudiantes universitarios. Puesto que, en muchos casos, el grupo de alumnos es numeroso, es complicado realizar una evaluación personalizada y hay que acudir a herramientas virtuales de apoyo a la docencia presencial. En la asignatura de Estadística impartida en el primer curso de los grados en Industriales, se ha utilizado una de las herramientas proporcionadas por la plataforma de docencia virtual, ILIAS. Se han generado diversos bancos de preguntas a partir de los cuales se han creado diferentes test que han permitido la evaluación de los conocimientos de los alumnos en relación con el contenido de la asignatura. Este tipo de test posee herramientas estadísticas que permiten mostrar tanto de forma individual como global los resultados obtenidos en cada evaluación. Con estos datos, profesores y alumnos pueden realizar un diagnóstico de las necesidades, intereses y dificultades en cada caso.

Palabras claves: ILIAS, plataformas virtuales, test de evaluación

REALIZACIÓN DE TEST DE EVALUACIÓN UTILIZANDO LA PLATAFORMA ILIAS

I. García-Garrido, A. López-Montoya, P.A. Cano-Chica, V. Cueva-López

Dpto. Estadística e Investigación Operativa. Universidad de Jaén

El Espacio Europeo de Educación Superior conlleva el proceso de evaluación continua para los estudiantes universitarios. Puesto que, en muchos casos, el grupo de alumnos es numeroso, es complicado realizar una evaluación personalizada y hay que acudir a herramientas virtuales de apoyo a la docencia presencial. En la asignatura de Estadística impartida en el primer curso de los grados en Industriales, se ha utilizado una de las herramientas proporcionadas por la plataforma de docencia virtual, ILIAS. Se han generado diversos bancos de preguntas a partir de los cuales se han creado diferentes test que han permitido la evaluación de los conocimientos de los alumnos en relación con el contenido de la asignatura. Este tipo de test posee herramientas estadísticas que permiten mostrar tanto de forma individual como global los resultados obtenidos en cada evaluación. Con estos datos, profesores y alumnos pueden realizar un diagnóstico de las necesidades, intereses y dificultades en cada caso.

Guía Docente

En la Guía Docente de Estadística para el Plan de los diferentes Grados en Ingeniería de la Escuela Politécnica Superior de Jaén:

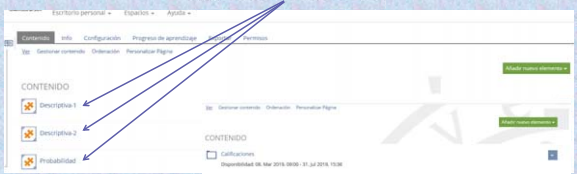
Realización de trabajos, casos o ejercicios	Correcta resolución de problemas	20.0%

Sistema elegido para el Curso 2018-2019

Docencia Virtual

Plataforma ILIAS

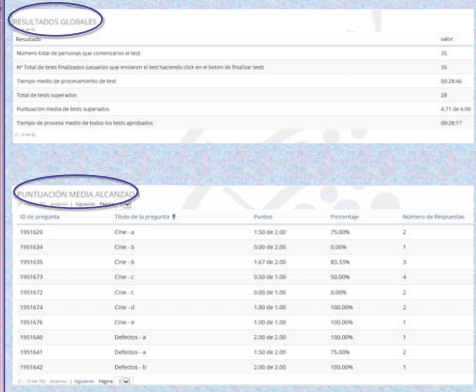
Bancos de Preguntas



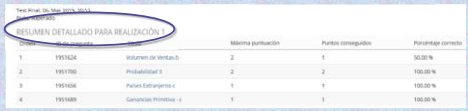
Test de Evaluación

Resultados obtenidos a partir del test

Resultados generales



Resultados por alumno



Corrección de cada ejercicio

Ventajas para el profesorado

- ✓ El banco de preguntas permite utilizar diferentes ejercicios para cada alumno.
- ✓ La evaluación del test es automática.
- ✓ No supone tiempo excesivo de corrección.
- ✓ Se tiene información detallada del conocimiento de cada alumno a lo largo de la asignatura.
- ✓ Se pueden reforzar los contenidos en los que hayan obtenido peores resultados globales.

Ventajas para el alumnado

- ✓ Les obliga a seguir un hábito de estudio.
- ✓ Asimilan los conocimientos estadísticos estudiados poco a poco.
- ✓ Les permite ver sus errores y aprender de ellos.
- ✓ Motivación para estudiar a diario.
- ✓ Les ayuda a familiarizarse con las herramientas informáticas.

Características que influyen en la competencia de los estudiantes españoles en PISA 2015

Ana María Lara[1], María del Mar Rueda[1], David Molina[2]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Granada
[2]Departamento de Didáctica de la Matemática, Universidad de Granada

Contacto: alara@ugr.es

Resumen

El objetivo de este trabajo es investigar qué factores influyen en el rendimiento de los estudiantes españoles en la competencia matemática de las pruebas PISA 2015. Analizamos variables relacionadas con los estudiantes y sus características familiares, las escuelas y las regiones donde se ubican las escuelas. Debido a su organización jerárquica, donde cada estudiante pertenece a una escuela y, a su vez, cada escuela está ubicada en una región, consideramos un modelo de regresión multinivel con tres niveles (estudiantes, escuelas y regiones). Nuestros resultados indican que la mayoría de las variables con una influencia significativa en el rendimiento matemático de los estudiantes son características personales, socioculturales y académicas del estudiante. Estas variables fueron el género femenino, la repetición de curso y el estatus de inmigrante (en un sentido negativo) y la educación preescolar y el estatus económico y sociocultural (en un sentido positivo). Algunas variables del nivel escolar, como el índice de responsabilidad escolar para el plan de estudios y la evaluación y la matriculación escolar total, resultaron significativas, ambas en un sentido positivo. Finalmente, la tasa de desempleo regional y el PIB de las regiones también influyeron en el rendimiento de los estudiantes.

Palabras claves: PISA, Análisis Multinivel, Competencia Matemática, Rendimiento

Características que influyen en la competencia matemática de los estudiantes españoles en PISA 2015

Ana María Lara Porras[1*], M. del Mar Rueda García[1] y David Molina Muñoz[2]

[1] Departamento de Estadística e Investigación Operativa, Universidad de Granada

[2] Departamento de Didáctica de la Matemática, Universidad de Granada

*alara@ugr.es

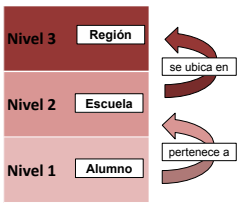
Introducción

El *Programa Internacional de Evaluación de Estudiantes* (PISA, por sus siglas en inglés), es un macro estudio comparativo realizado por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) de forma trienal desde el año 2000 y que presenta las siguientes características:

- Evalúa la **competencia en matemáticas, lectura y ciencias** de una muestra de estudiantes de 15 años residentes en alguno de los países miembros o asociados de a OCDE.
- Cada edición analiza en profundidad una de las áreas mencionadas en el punto anterior. La ediciónd e 2015 se centró en el estudio de la competencia científica.
- Se utilizan distintos tipos de cuestionarios para recoger información tanto de **los alumnos** como de **los centros educativos**. Esta información puede complementarse con información de la región en la que se localiza cada centro.

El **objetivo** de este trabajo es determinar los factores que condicionan el rendimiento en matemáticas de los estudiantes españoles a partir del análisis de los datos de PISA 2015.

Metodología



En este tipo de situaciones, los **modelos multinivel** constituyen una herramienta muy útil que ayuda a explicar una variable dependiente a partir de información recopilada en distintos niveles.

En la práctica, se parte de un modelo que no contiene ninguna variable independiente, denominado **modelo nulo**. Para este modelo, se contrasta la necesidad de incorporar variables independientes para explicar la dispersión de la variable respuesta. En tal caso, se añaden, progresivamente, variables de los distintos niveles.

Muestra y variables

La muestra se compone de un total de 32330 estudiantes de las 17 comunidades autónomas españolas que participaron en las pruebas de PISA 2015.

La **variable respuesta** usada en este trabajo se ha elaborado a partir de los diez valores plausibles en matemáticas que se han recogido en las pruebas PISA 2015, usando la metodología descrita por la OECD (2017).

Las **variables independientes** que se han considerado aparecen resumidas en la tabla 1. Estas variables son aquellas características de cada unos de los niveles que más influencia tienen en el rendimiento matemático de los estudiantes según los trabajos más relevantes en materia de educación.

Nivel 3	Nivel 2	Nivel 1
• Gasto en educación • Tasa de alumnado extranjero • Producto interior bruto • Tasa de desempleo	• Titularidad del centro: pública o privada. • Localización del centro: zona urbana o zona rural • Ratio profesor-alumnos • Número de alumnos matriculados • Implicación del centro en el diseño del currículo (RESPCUR) • Implicación del centro en la gestión de recursos (RESPRES)	• Género: masculino *o femenino. • ¿Ha repetido el alumno algún curso académico?: sí o no* • ¿Ha recibido el alumno educación infantil?: sí o no* • ¿Es el alumno inmigrante?: sí o no* • ¿Tiene estudios (nivel ISCED 3A o superior) el padre o la madre del alumno ?: sí o no* • ¿Dispone el alumno de una conexión a internet en casa?: sí o no* • ¿Dispone el alumno de libros de texto para su consulta en casa?: sí o no* • Índice socioeconómico y cultural (ESCS)

** Categoría de referencia*

Tabla 1: Variables independientes por niveles

Resultados

Tras calcular el modelo nulo y comprobar la necesidad de incorporar variables explicativas, se estimaron modelos considerando las variables de cada nivel de forma independiente y un modelo a partir de las variables de los tres niveles conjuntamente. Las estimaciones de los parámetros de los distintos modelos se muestran en las tablas 2 y 3. Entre paréntesis aparece el error estándar asociado a cada estimación.

Nivel 1		
Variables	**Intersección**	**Coeficientes**
Género	515,179 (2,946)	-27,617 (2,486)
Repetidor		-82,064 (0,951)
Escolarización infantil		10,686 (1,750)
Inmigrante		-25,094 (3,069)
Nivel educativo de los padres		3,200 (1,123)
Internet en casa		7,898 (1,960)
Libros en casa		3,370 (0,968)
ESCS		12,423 (0,466)
Género * Escolarización infantil		8,836 (2,601)
Repetidor * Inmigrante		11,762 (2,465)
Escolarización infantil * Inmigrante		6,724 (3,020)
Nivel 2		
Variables	**Intersección**	**Coeficientes**
Titularidad	493,130 (2,306)	-11,512 (2,212)
Número de alumnos matriculados		0,015 (0,002)
RESPCUR		3,563 (1,368)
RESPRES		8,776 (1,838)
Nivel 3		
Variables	**Intersección**	**Coeficientes**
Tasa de desempleo	567,510 (12,503)	-3,159 (0,346)
Producto Interior Bruto		-0,0009 (0,0003)
Tasa de desempleo * Tasa de estudiantes extranjeros		-0,377 (0,0179)

Tabla 2. Estimaciones usando variables de cada nivel. (El símbolo * representa interacción entre variables)

Niveles 1, 2 y 3		
Variables	**Intersección**	**Coeficientes**
Género	596,882 (8,181)	-27,560 (82,483)
Repetidor		-81,939 (0,948)
Escolarización infantil		11,071 (1,747)
Inmigrante		-25,610 (3,029)
Nivel educativo de los padres		3,976 (1,122)
Internet en casa		7,473 (1,958)
Libros en casa		3,388 (0,966)
ESCS		12,333 (0,465)
Género * Escolarización infantil		8,829 (2,590)
Repetidor * Inmigrante		11,304 (2,461)
Escolarización infantil * Inmigrante		6,770 (3,015)
Número de alumnos matriculados		0,007 (0,0014)
Tasa de desempleo		-3,067 (0,209)
Producto Interior Bruto		-0,0016 (0,0002)

Tabla 3. Estimaciones usando variables de los tres niveles. (El símbolo * representa interacción entre variables)

Conclusiones

- La mayoría de variables relevantes para explicar el rendimiento en matemáticas de los estudiantes son **características de los propios estudiantes**.
- La **condición de repetidor**, el **género femenino**, la **condición de inmigrante**, la **tasa de desempleo** de la región y **el producto interior bruto** influyen **negativamente** en el rendimiento académico en matemáticas del alumno.
- El **índice socioeconómico y cultural**, la **escolarización infantil** y la **posesión de internet en casa**, entre otras variables, influyen **positivamente** en el rendimiento académico en matemáticas del estudiante.
- Algunas interacciones de variables referidas al alumno también resultan significativas.
- La titularidad del centro deja de ser significativa cuando se consideran variables del nivel alumno y del nivel región.

Bibliografía

- Gaviria, J. y Castro, M. (2005). *Modelos jerárquicos lineales*, Cuadernos de Estadística, 29, Madrid, Editorial La Muralla S. A.
- Mislevy, R. J. (1991). Randomization-based inference about latent variable from complex samples. *Psychometrika* 56, 177 - 196.
- Mislevy, R. J. Beatin, A. E., Kaplan, B. y Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133 - 161.
- OECD (2017). PISA 2015 Technical Report. Recuperado de: http://www.oecd.org/pisa/data/2015-technical-report/
- Wu, M. y Adams, R. J. (2002). *Plausible values - why are they important*. International Objective Measurement Workshop, New Orleans.

Técnicas Cuantitativas II: Motivación, utilidad y satisfacción

Ana Eugenia Marín Jiménez, María del Pilar Fernández Sánchez, Rosaura Fernández Pascual

Departamento de Métodos Cuantitativos para la Economía y la Empresa, Universidad de Granada

Contacto: anamarin@ugr.es

Resumen

Las asignaturas en las que se enseña el uso de diferentes técnicas cuantitativas son fundamentales en los estudios de Ciencias Económicas y Empresariales. Se ha diseñado un cuestionario para medir la motivación, utilidad y satisfacción con la asignatura Técnicas Cuantitativas 2. Se obtiene una muestra de 194 estudiantes (49.4 % hombres) en cuatro grados de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Granada. Los resultados muestran altas puntuaciones en Satisfacción y puntuaciones bajas en Utilidad. Se ha comprobado que no existen diferencias de opinión en cuanto al género, y que existen diferencias significativas en la percepción entre los grados estudiados en la mayoría de las cuestiones planteadas. El Grado en Marketing e Investigación de Mercados da una puntuación más baja la motivación; el Grado en Economía da una puntuación más alta a las cuestiones sobre utilidad, ya sea para futuras asignaturas o para el futuro laboral, y, finalmente, los Grados en Finanzas y Contabilidad y Economía dan puntuaciones más altas a la satisfacción con las clases y con el sistema de evaluación.

Palabras claves: Técnicas Cuantitativas, motivación, utilidad, satisfacción

Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

UNIVERSIDAD DE GRANADA

TÉCNICAS CUANTITATIVAS II: MOTIVACIÓN, UTILIDAD Y SATISFACCIÓN

Marín-Jiménez, A.E., Fernández-Sánchez, M.P., Fernández-Pascual-, R
anamarin@ugr.es, pilarfs@ugr.es , rpascual@ugr.es
Departamento de Métodos Cuantitativos para la Economía y la Empresa. Universidad de Granada

INTRODUCCIÓN

Las asignaturas en las que se enseña el uso de diferentes técnicas cuantitativas son fundamentales en los estudios de Ciencias Económicas y Empresariales. Se ha diseñado un cuestionario para medir la motivación, utilidad y satisfacción con la asignatura Técnicas Cuantitativas 2. Esta asignatura se imparte en el segundo semestre del segundo curso en la mayoría de los grados impartidos en la Facultad de Ciencias Económicas y Empresariales de la Universidad de Granada.

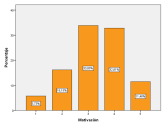
En este trabajo se han analizado 9 items medidos en una escala Likert 1-5, agrupados en tres dimensiones motivación, utilidad y satisfacción (con los contenidos, materiales, metodología y nuevas tecnologías utilizados durante el curso).

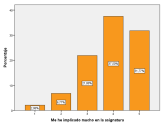
Los 194 participantes provienen de los grados en Administración y Dirección de Empresas (ADE) Económicas (ECO) Finanzas y Contabilidad (FICO) y Marketing e Investigación de Mercados (MIN), con edades entre 18 y 37 (media 21.55 y SD 2.82). 50.6% son mujeres.

Los resultados revelan una alta consistencia y confiabilidad del cuestionario con un alfa de Cronbach de 0.819.

MOTIVACIÓN

	N	Media	S.D.	Mo	Min	Max
Motivación	192	3.28	1.05	3	1	5
Implicación	192	3.90	0.10	4	1	5



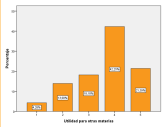


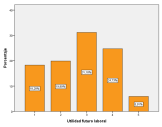
	Género	Media (SD)	U de Mann-Whitney (Sig)
Motivación	Masculino	3.32 (1.00)	0.744
	Femenino	3.26 (1.12)	
Implicación	Masculino	3.83 (0.10)	0.213
	Femenino	4.00 (1.01)	

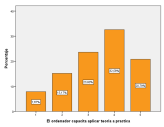
	Grado	Media (SD)	U de Mann-Whitney (Sig)
Motivación	GADE	3.00 (0.82)	0.000
	GECO	3.55 (0.97)	
	GFICO	3.14 (1.05)	
	GMIM	2.44 (1.15)	
Implicación	GADE	3.85 (1.35)	0.033
	GECO	3.74 (0.99)	
	GFICO	4.06 (0.90)	
	GMIM	4.33 (0.98)	

UTILIDAD

	N	Media	S.D.	Mo	Min	Max
Utilidad otras materias	187	3.63	0.10	4	1	5
Utilidad futuro laboral	186	2.80	1.18	3	1	5
Ordenador capacita para aplicar teoría a práctica	178	3.43	1.20	4	1	5





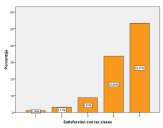


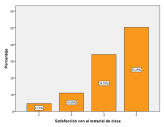
	Género	Media (SD)	U de Mann-Whitney (Sig)
Utilidad otras materias	Masculino	3.69 (0.10)	0.285
	Femenino	3.49 (1.16)	
Utilidad futuro laboral	Masculino	2.90 (1.11)	0.194
	Femenino	2.66 (1.19)	
El ordenador capacita para aplicar la teoría en la práctica	Masculino	3.62 (1.19)	0.061
	Femenino	3.27 (1.20)	

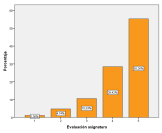
	Grado	Media (SD)	U de Mann-Whitney (Sig)
Utilidad otras materias	GADE	3.08 (0.76)	0.000
	GECO	4.03 (0.82)	
	GFICO	3.13 (1.22)	
	GMIM	3.47 (1.41)	
Utilidad futuro laboral	GADE	2.82 (0.98)	0.000
	GECO	3.17 (1.13)	
	GFICO	2.30 (1.13)	
	GMIM	2.56 (1.03)	
El ordenador capacita para aplicar la teoría en la práctica	GADE	3.31 (1.44)	0.148
	GECO	3.45 (1.10)	
	GFICO	3.52 (1.26)	
	GMIM	1.67 (1.16)	

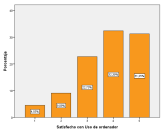
SATISFACCIÓN

	N	Media	S.D.	Mo	Min	Max
Satisfacción clases	193	4.35	0.85	5	1	5
Satisfacción material	191	4.30	0.85	5	2	5
Satisfacción evaluación	190	4.32	0.92	5	1	5
Satisfacción uso ordenador	176	3.77	1.13	4	1	5









	Género	Media (SD)	U de Mann-Whitney (Sig)
Satisfacción clases	Masculino	4.38 (0.80)	0.950
	Femenino	4.32 (0.91)	
Satisfacción material	Masculino	4.35 (0.70)	0.736
	Femenino	4.31 (0.90)	
Satisfacción evaluación	Masculino	4.23 (0.98)	0.301
	Femenino	4.40 (0.84)	
Satisfacción uso ordenador	Masculino	3.79 (1.07)	0.988
	Femenino	3.77 (1.15)	

	Grado	Media (SD)	U de Mann-Whitney (Sig)
Satisfacción clases	GADE	3.77 (1.30)	0.008
	GECO	4.49 (0.70)	
	GFICO	4.41 (0.77)	
	GMIM	3.75 (1.13)	
Satisfacción material	GADE	3.83 (1.19)	0.080
	GECO	4.40 (0.79)	
	GFICO	4.33 (0.78)	
	GMIM	3.87 (0.99)	
Satisfacción evaluación	GADE	3.75 (1.06)	0.009
	GECO	4.43 (0.79)	
	GFICO	4.44 (0.83)	
	GMIM	3.53 (1.41)	
Satisfacción uso ordenador	GADE	3.46 (1.45)	0.067
	GECO	3.66 (1.05)	
	GFICO	4.02 (1.14)	
	GMIM	3.00 (0)	

CONCLUSIONES

- Altas puntuaciones en Satisfacción y puntuaciones bajas en Utilidad.
- No existen diferencias de opinión en cuanto al género.
- El Grado en Marketing e Investigación de Mercados da una puntuación más baja a la motivación.
- El Grado en Economía da una puntuación más alta a las cuestiones sobre utilidad, ya sea para futuras asignaturas o para el futuro laboral.
- Los Grados en Finanzas y Contabilidad y Economía dan puntuaciones más altas a la satisfacción con las clases y con el sistema de evaluación.

Aprendizaje cooperativo y gamificación: aplicación para la enseñanza en Estadística económica, empresarial y del sector turístico

Sergio Martínez, Helena Martínez, Eva Mª Artés, Isabel Mª Ortiz, Mª Inmaculada López, Mª Dolores Illescas, Alfredo Martínez

Departamento de Matemáticas, Universidad de Almería

Contacto: spuertas@ual.es

Resumen

Una correcta asimilación de las técnicas estadísticas, que escape de su uso indiscriminado y sistemático, es un objetivo difícil de conseguir en las titulaciones de Ciencias Económicas y Empresariales, donde la labor docente está limitada por dos factores: la masificación de estudiantes por grupo y el perfil específico de este alumnado que carece de una base matemática sólida. Así, se observa que, por lo general, este alumnado presenta serias deficiencias a la hora de conectar adecuadamente los contenidos teóricos con aplicaciones reales. En esta aportación se recoge el plan de acción docente aplicando en diferentes asignaturas de Estadística adscritas a estudios de Ciencias Económicas y Empresariales, con el propósito de conseguir una implicación más activa del estudiante en su proceso de aprendizaje. En dicho plan se conjugan las estrategias siguientes: aprendizaje cooperativo, gamificación, cuestionarios de autoevaluación interactivos y elaboración de videos tutoriales.

Palabras claves: Estadística, aprendizaje cooperativo, juegos de aprendizaje interactivos, videos tutoriales

Agradecimientos. Los autores agradecen el apoyo de la Universidad de Almería en la creación del Grupo de Innovación Docente "Aprendizaje cooperativo y gamificación: aplicación para la enseñanza en Estadística económica, empresarial y del sector turístico", durante el bienio 2018-2019.

Referencias

[1] E. Bárcena, A. Fernández, J.L. Iranzo y B. Lacomba (2007) Una experiencia de aprendizaje colaborativo en estadística basado en la resolución de casos reales. *III Jornadas de Innovación educativa y enseñanza virtual en la Universidad de Málaga*.

[2] M.D. Jano y S. Ortiz (2007) Experiencia de innovación docente en estadística económica. *Revista de Docencia Universitaria REDU*, **2**, pp. 1-11.

[3] J.M. Pavía, R. Martínez, G. Morillas y B. García (2013) *Docencia en Estadística, experiencias de Innovación*. ADD Editorial, Madrid.

Aprendizaje cooperativo y gamificación: aplicación en Estadística económica, empresarial y del sector turístico

S. Martínez Puertas, H. Martínez Puertas, E.M. Artés Rodríguez, I.M. Ortiz Rodríguez, M.I. López García, M.D. Illescas Manzano y A. Martínez Almécija

Facultad de Ciencias Económicas y Empresariales, Universidad de Almería

Motivación

Una correcta asimilación de las técnicas estadísticas, que escape de su uso indiscriminado y sistemático, es un objetivo difícil de conseguir en las titulaciones de Ciencias Económicas y Empresariales, donde la labor docente está limitada dos factores: la masificación de estudiantes por grupo ($\approx$ 100 matriculados) y el perfil específico de este alumnado que carece de una base matemática sólida.

¿Qué hemos observado de este alumnado en nuestra experiencia como docentes?

- Serias deficiencias a la hora de conectar adecuadamente los contenidos teóricos con aplicaciones reales.
- Grandes dificultades para afrontar la asignatura con éxito.

Ante este escenario, se hace necesario adoptar nuevas **metodologías activas** que supongan una implicación más profunda de los estudiantes en su proceso de aprendizaje.

Objetivos

- Fomentar el trabajo en equipo.
- Enfrentar a los estudiantes en la aplicación de técnicas estadísticas a problemas reales.
- Que el estudiante adquiera habilidades relacionadas con la comunicación oral.
- Hacer más lúdico el estudio de la materia a través del juego.

- Evaluar los conocimientos previos de Estadística del alumnado.
- Conocer las espectativas y grado de interés del alumnado respecto a la asignatura.
- Mejorar los resultados académicos de cursos anteriores.
- Conocer el grado de satisfacción de los estudiantes respecto al proyecto.

Casos de estudio

Curso académico 2018/2019:

- Estadística
  Carácter: Básica, Curso: 1º, Cuatrimestre: 2º
  Grado en: ADE, Derecho+ADE, Economía, FYCO, Marketing e Investigación de Mercados.

- Estadística aplicada al sector turístico
  Carácter: Básica, Curso: 2º, Cuatrimestre: 2º
  Grado en Turismo.

Instrumentos y Metodología

Autoevaluación inicial:

- 15 preguntas tipo test sobre nociones muy básicas de Estadística Descriptiva y Probabilidad y 10 preguntas sociodemográficas para conocer algunos aspectos del perfil de nuestro alumnado.

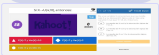
Aprendizaje cooperativo:

- Por parejas y con base de datos real con información recabada de una encuesta, realizar análisis descriptivos de datos unidimensionales y bidimensionales con SPSS y redactar el informe correspondiente.
- Preparación en grupos de 4 personas del tema Números Índices: prueba individual con nota final la media del grupo.
- Resolución y exposición de problemas del bloque de probabilidad (Probabilidad, Variable aleatoria, Modelos de distribuciones) al finalizar cada tema. Se propone un ejercicio para que resuelvan en clase en grupos de 4 personas. El primer equipo en resolver el problema sale a pizarra para exponerlo y explicarlo al resto de compañeros, siendo el profesor el encargado de elegir arbitrariamente qué miembro del equipo realiza la exposición de cada apartado. El equipo vencedor acumula 0.5 puntos positivos.

Tablero de discusión:

- Medio para puesta en común de ejercicios no resueltos en clase.
- El profesor interviene pasado un tiempo prudencial de debate.
- La participación activa y de calidad acumula puntos positivos.

Quizzes interactivos:



- Instrumento de Feed-back formativo: ¿Concepto asimilado correctamente?
- Repaso de conceptos al finalizar cada tema teórico. El vencedor acumula 0.25 puntos positivos.
- Herramientas utilizadas: Kahoot! y Socrative.

Videos tutoriales:

- Refuerzo en el proceso de aprendizaje: complemento a la lección magistral fuera del aula.
- Material elaborado con *OBS STUDIO* y *OpenShot* y enriquecido con *EDpuzzle* (vídeos interactivos).
- Contenido sobre manejo de *Excel* y *SPSS*.

Juegos interactivos:

- Repasar jugando. Material recabado de internet:
  https://blogs.ua.es/violeta/category/videos-y-videojuegos-estadisticaim/

Estado del proyecto

Proyecto puesto en práctica en el cuatrimestre actual.

Instrumentos que se utilizarán para la evaluación:

- Dimensión objetiva: asistencia, seguimiento de actividades propuestas y resultados académicos.
- Dimensión subjetiva:
  - Cuestionario inicial al alumnado (conocimientos previos y opinión sobre la materia). Primer día de clase.
  - Cuestionario final al alumnado (grado de satisfacción con el proyecto). Última semana de clase.
  - Intercambio de impresiones e incidencias del profesorado involucrado en el proyecto.

Conclusiones

- La metodología utilizada está teniendo una buena aceptación por parte del alumnado, que afirma sentirse más motivado a estudiar de forma continua por la recompensa de obtener puntos positivos al final de cada tema.
- Las impresiones del profesorado involucrado también están siendo muy positivas, ya que el sistema utilizado no está suponiendo una carga de trabajo difícil de conllevar, dada la gran cantidad de alumnos por grupo.

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

Granada, 4 y 5 de abril de 2019

Estimulando el autoaprendizaje de la Estadística: actualización de la aplicación

Emilia I. Martos Gálvez[1,2], Luis F. Rivera Galicia[1,2], Juana Domínguez Domínguez[1,2], Eva Senra Díaz[1,2], Fco. Javier Callealta Barroso[1,2]

[1]Departamento de Economía, Universidad de Alcalá
[2]Grupo de Innovación Docente para la Enseñanza de los Métodos Cuantitativos, Universidad de Alcalá

Contacto: emilia.martos@uah.es

Resumen

El seguimiento continuado del trabajo del estudiante incide positivamente en el resultado de su proceso de aprendizaje. Sin embargo, esta tarea es tanto más difícil cuanto mayor es el número de estudiantes. En la unidad de Estadística de nuestra Facultad se desarrolló, y se emplea desde el curso 2012-13, una herramienta informática que genera tareas con datos individualizados para cada alumno en las asignaturas Estadística Económica I del grado de ECO y Estadística Económica del grado de ENI. Además, permite la corrección automatizada de sucesivas entregas de los estudiantes para su supervisión, informando a cada alumno de sus errores y aciertos en tiempo real, favoreciendo así la evolución de su aprendizaje. Las necesidades docentes surgidas durante los últimos años han obligado a crear nuevas utilidades en esta herramienta, como son la posibilidad de generar las tareas en inglés, o la parametrización de las preguntas asignadas a cada alumno, de tal forma que incluso estas sean personalizadas. En este trabajo se describen las innovaciones desarrolladas para satisfacer dichas necesidades.

Palabras claves: actividades personalizadas, aprendizaje autónomo, seguimiento automatizado

Agradecimientos. Los autores agradecen el apoyo de la Universidad de Alcalá a través del proyecto UAH/EV1075: *Adaptación y modernización de los Contenidos de Métodos Cuantitativos para la Economía y la Empresa en la Facultad de Ciencias Económicas, Empresariales y Turismo*.

Referencias

[1] F.J. Callealta Barroso (2013) *Automatización de un Proceso de Generación y Seguimiento de Actividades Personalizadas para el Aprendizaje*

de la Estadística, en Entornos Masivos. IV Jornadas del Grupo de Enseñanza y Aprendizaje de la Estadística y la I.O. (GENAEIO). Actas de las Jornadas, pp. 85-122 (ISBN: 84-695-9457-5). Cádiz, España.

[2] F.J. Callealta Barroso (2014) *Estimulando el autoaprendizaje en entornos masivos: una aplicación en el ámbito de la Estadísitca*. X Jornadas sobre Docencia de Economía Aplicada. Madrid, España.

[3] F.J. Callealta Barroso, E.I. Martos Gálvez y E. Senra Díaz (2015) *Estimulando el autoaprendizaje en entornos masivos: una aplicación en el ámbito de la Estadística (versión 2)*. Jornadas de Intercanvi d'Experiències d'Innovació Educativa en Econometria i Mètodes Quantitatius. Valencia, España.

Estimulando el autoaprendizaje de la Estadística: actualización de la aplicación

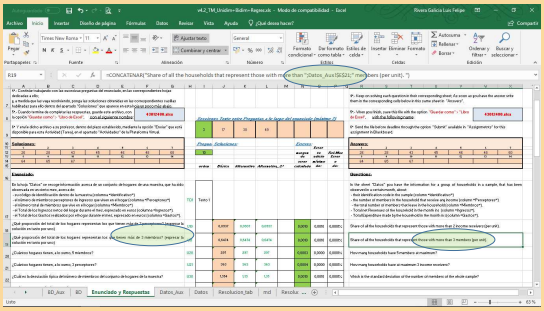
Universidad de Alcalá

Martos Gálvez, Emilia I., Rivera Galicia, Luis F., Domínguez Domínguez, Juana, Senra Díaz, Eva, Callealta Barroso, Fco. Javier
Departamento de Economía, Facultad de Ciencias Económicas, Empresariales y Turismo

Objetivos de la experiencia: Propiciar condiciones para que el alumno trabaje con continuidad y reciba información sobre el grado de conocimientos que va alcanzando. Dar al alumno un trabajo totalmente personalizado.
Metodología: Facilitar al alumno diversas pruebas prácticas personalizadas en el conjunto de datos y en las preguntas, que se supervisen de forma automática. De esta forma, se podrán hacer diferentes revisiones a lo largo del tiempo de modo que el alumno pueda alcanzar sus objetivos en cuanto a nota y aprendizaje.
En esta nueva versión de la aplicación se han perseguido dos objetivos:

- Que el alumno reciba la tarea en castellano o en inglés, según el idioma en el que recibe la docencia de la asignatura.
- Parametrizar las preguntas de las diferentes tareas con objeto de que los alumnos no sólo tengan diferentes conjuntos de datos para su resolución, sino que las preguntas a las que tienen que responder también estén personalizadas.

Actualización de la Aplicación

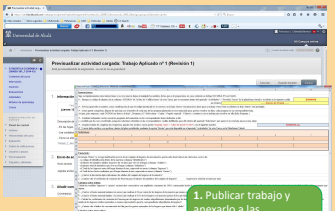


En la actualidad estamos trabajando en la versión 4 de la aplicación, en la que se ha conseguido, en primer lugar, la traducción de las tareas al inglés, dado que esta asignatura se oferta en dicho idioma en el grado de Economía y Negocios Internacionales.
Al definir los grupos de alumnos que tienen que realizar la tarea, se define el idioma en que la tiene que resolver cada alumno.
Con objeto de personalizar la tarea, además de que cada alumno recibe un conjunto de datos diferente, en función del código que se le asigna a principio de curso, muchas de las preguntas que el alumno tiene que resolver se encuentran parametrizadas, de acuerdo con dicho código.
De este modo, las tareas están más personalizadas para los alumnos.

Sistema Automatizado de asignación, revisión y seguimiento de las Tareas Prácticas

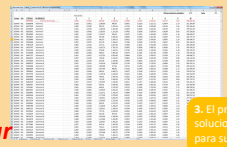
INICIO DEL CURSO - TAREAS DEL PROFESOR

- Preparar los trabajos desde el repositorio de Trabajos Maestros
- Preparar los listados de los alumnos y sus grupos y les asigna un código
- Preparar calendario de actividades y revisiones

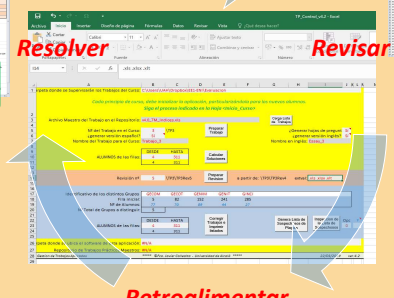


1. Publicar trabajo y anexarlo a las actividades del Aula Virtual

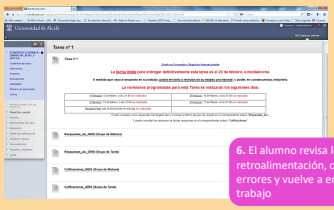
2. El alumno recibe el trabajo, comienza su resolución y lo envía para su revisión a través del Aula Virtual



3. El profesor prepara las soluciones de los trabajos para su revisión (sólo la primera vez)



4. El profesor ejecuta el proceso de revisión y se emiten los listados de retroalimentación



5. Dichos listados se publican en el Aula Virtual

6. El alumno revisa los listados de retroalimentación, corrige sus errores y vuelve a enviar el trabajo

Versiones Anteriores del programa:

- F.J. Callealta Barroso (2013) Automatización de un Proceso de Generación y Seguimiento de Actividades Personalizadas para el Aprendizaje de la Estadística, en Entornos Masivos. IV Jornadas del Grupo de Enseñanza y Aprendizaje de la Estadística y la I.O. (GENAEIO). Actas de las Jornadas, .pp: 85-122 (ISBN: 84-695-9457-5). Cádiz - (España)
- F.J. Callealta Barroso (2014) Estimulando el autoaprendizaje en entornos masivos: una aplicación en el ámbito de la Estadística. X Jornadas sobre Docencia de Economía Aplicada. Madrid - (España)
- F.J. Callealta Barroso; E.I. Martos Gálvez; E. Senra Díaz (2015) Estimulando el autoaprendizaje en entornos masivos: una aplicación en el ámbito de la Estadística (versión 2). Jornadas de Intercanvi d'Experiències d'Innovació Educativa en Econometria i Mètodes Quantitatius. Valencia - (España)

GIDEMECU - Grupo de Innovación Docente para la Enseñanza de los Métodos Cuantitativos
Proyecto UAH/EV1061: Parametrización de las pruebas de Evaluación Continua

Aprendiendo Estadística investigando

Felícita Doris Miranda Huaynalaya

Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: fdmirandah@correo.ugr.es

Resumen

Tras aplicar metodologías activas, realicé un proyecto en que los alumnos debían llevar a cabo una investigación estadística sobre un tema de libre elección. El objetivo era que el alumno aprendiera las herramientas de la estadística y su importancia, pudiendo responder a las preguntas: ¿Para qué?, ¿Por qué?, ¿Cuándo?, ¿Cómo?. Después de las clases, el alumnado debía aplicar los conocimientos aprendidos en su trabajo de investigación. El fin era reforzar los conocimientos adquiridos, facilitando el aprendizaje y generando conocimientos duraderos. Mi seguimiento continuo y mis orientaciones en el trabajo de cada alumno favorecieron el alcance del objetivo planteado. La calidad del trabajo final tuvo una relación directa con el rendimiento del alumno en el curso, su evolución y la calificación del examen final.

Palabras claves: Estadísticas, Metodologías activas, Proyecto de investigación

Introducción

En las últimas décadas la metodología de enseñanza en la universidad ha cambiado de las clases magistrales (tradicional) al uso de metodologías activas: Aprendizaje Basado en Problema (ABP o PBL), Estudio del Caso, Trabajo Colaborativo, aprendizajes por proyectos, etc. Usando estas metodologías activas he obtendio excelentes resultados en el aprendizaje de los estudiantes de diferentes especialidades.

- **Motivación:** Desarrollar el razonamiento estadístico en los alumnos sobre la importancia del uso de las herramientas estadísticas con el proyecto "Aprender investigando".
- **Proyecto:** Investigación estadística llevada a cabo por cada alumno sobre un tema de libre elección. Basado en metodología activas.
- **Objetivo:** Enseñar las herramientas de la estadística y su importancia. Poder responder ¿Para qué? ¿Por qué? ¿Cuándo? ¿Cómo?.
- **Finalidad:** Generar conocimientos duraderos y facilitar el aprendizaje. Despertar la motivación en el alumno por el contenido estadístico.

- **Resultado:** Relación directa entre calidad del trabajo final, rendimiento del alumnado, evaluación y el examen final.

Desarrollo del trabajo

Enfoque Temático

Se presenta al estudiante las técnicas estadísticas elementales, los métodos estadísticos descriptivos e inferenciales. Tiene la siguiente estructura:

- **Teoría:** las clases son expositivas. El enfoque es práctico, dándole énfasis a los siguientes temas generales: Análisis de Datos, Probabilidades, Modelos de probabilidad o Inferencia Estadística.
- **Prácticas:** consiste en talleres y prácticas calificadas. Los talleres tienen dos partes: un trabajo grupal y un trabajo individual.
- **Soporte informático:** Uso de software estadístico SPSS y Excel.

Proyecto

Planteamiento: Trabajo de investigación opcional que se desarrollará en cada semana según el sílabo de la asignatura de Estadística.

- **Inicio:** Cada alumno selecciona la variable de investigación de alguna actividad personal o de la lista de variables que se le proporciona.
- **Seguimiento:** Los alumnos deben aplicar el tema desarrollado en clase a su variable de investigación. Se presenta en cada semana.
- **Trabajo Final:** Los alumnos deben entregar una monografía según las indicaciones antes del examen final. El estudiante adquiere 2 puntos sobre la puntuación del examen final.

Proceso del Proyecto

La asignatura dirigida a estudiantes de diferentes especialidades de Estudios Generales de Letras de la Pontificia Universidad Católica del Perú (PUCP) y la proporción de estudiantes que presentaron el TRABAJO FINAL.

Figura 1. Distribución de los estudiantes que están matriculados en el curso

Figura 2. Porcentaje de los estudiantes que realizaron el trabajo final del proyecto

Resultados

Se relaciona el CRAEST (Coeficiente de Rendimiento Académico Estandarizado) de los estudiantes con las puntuaciones de la Práctica Calificada 1 (PC1). Para medir la efectividad de la metodología de aplicar el proyecto de investigación en la asignatura se relaciona las puntuaciones de PC1 y las puntuaciones del Examen Final.

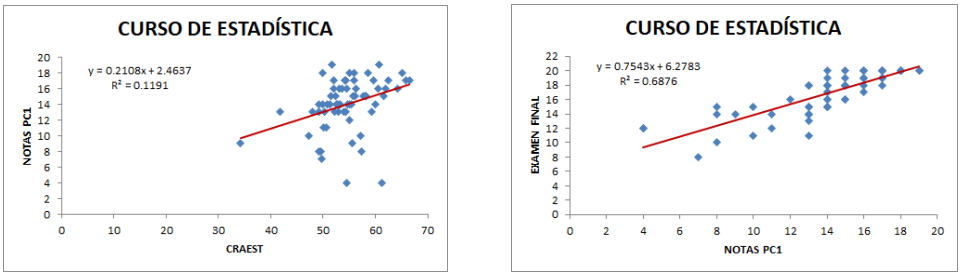


Figura 3. Relación entre el CRAEST del estudiante y la nota de la práctica calificada 1 (PC1) (lado izquierdo) y la relación entre la nota de la práctica calificada 1 y la nota del examen final (lado derecho)

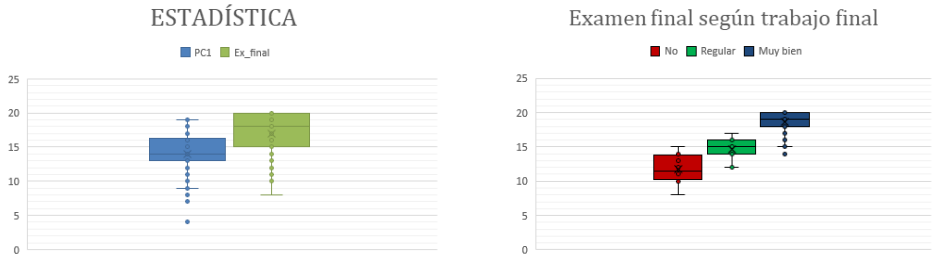


Figura 4. Comparación de la mediana entre nota de la práctica calificada 1 (PC1) y la nota del examen final (lado izquierdo) y comparación de la nota del examen final según el trabajo final del proyecto (lado derecho)

Conclusiones

El proyecto fue muy efectivo para el aprendizaje del estudiante, obteniendo resultados positivos:

- Cada avance del proyecto, reforzaba sus conocimientos sobre los contenidos de las Estadísticas.
- En cada etapa de los temas que aplicaban en el proyecto, el estudiante desarrollaba el razonamiento de la importancia y utilidad de las herramientas estadísticas.
- En la supervisión de cada trabajo he podido observar que se han logrado alcanzar los objetivos previstos del proyecto "Aprender investigando".

Reflexiones del estudiante:
Algunas reflexiones de los estudiantes que escribieron en el trabajo final, donde valoran el efecto que surgió en ellos y su motivación con cada avance logrado durante todo el desarrollo de la asignatura.

- Estas tareas, fueron muy efectivas para todos ya que nos permitió reforzar en el tema visto cada semana. Nos quitó de dudas y ayudó a practicar.
- Quería resaltar que este trabajo ha sido muy útil para terminar de entender los conceptos dados en clase, a través de todo el proceso de aprendizaje durante el semestre.

Aprendiendo Estadística investigando

Felícita Doris Miranda Huaynalaya
(fdmirandah@correo.ugr.es)
Universidad de Granada

Motivación y objetivos

Motivación: Desarrollar el razonamiento estadístico en los alumnos sobre la importancia del uso de las herramientas estadísticas con el proyecto "Aprender investigando".

Proyecto: Investigación estadística llevada a cabo por cada alumno sobre un tema de libre elección. Basado en metodología activas.

Objetivo: Enseñar las herramientas de la estadística y su importancia. Poder responder ¿Para qué? ¿Por qué? ¿Cuándo? ¿Cómo?.

Finalidad: Generar conocimientos duraderos y facilitar el aprendizaje. Despertar la motivación en el alumno por el contenido estadístico.

Resultado: Relación directa entre calidad del trabajo final, rendimiento del alumnado, evaluación y el examen final.

Enfoque temático

Se presenta al estudiante las técnicas estadísticas elementales, los métodos estadísticos descriptivos e inferenciales. Tiene la siguiente estructura:

1. **Teoría:** las clases son expositivas. El enfoque es práctico, dándole énfasis a los siguientes temas generales: Análisis de datos. Probabilidades. Modelos de probabilidad. Inferencia estadística.
2. **Prácticas:** consiste en talleres y prácticas calificadas. Los talleres tienen dos partes: un trabajo grupal y un trabajo individual.
3. **Soporte informático:** Uso de software estadístico SPSS y Excel.

Proyecto

Planteamiento: Trabajo de investigación opcional que se desarrollará en cada semana según el sílabo de la asignatura de Estadística.

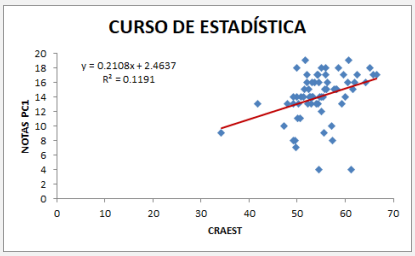
1. **Inicio:** Cada alumno selecciona la variable de investigación de alguna actividad personal o de la lista de variables que se le proporciona.
2. **Seguimiento:** Los alumnos deben aplicar tema desarrollado en clase a su variable de investigación. Se presenta en cada semana.
3. **Trabajo Final:** Los alumnos deben entregar una monografía según las indicaciones antes del examen final. El estudiante adquiere 2 puntos sobre el puntaje del examen final.

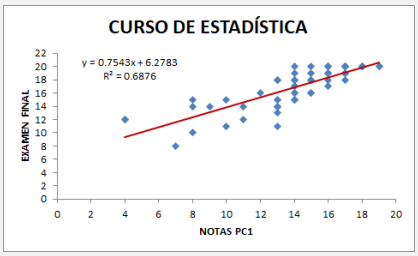
Proceso

La asignatura dirigida a estudiantes de diferentes especialidades de Estudios Generales de Letras de la PUCP y la proporción de estudiantes que presentaron el TRABAJO FINAL.

Resultados

Se relaciona el CRAEST (Coeficiente de Rendimiento Académico Estandarizado) de los estudiantes con las puntuaciones de la Práctica Calificada 1 (PC1). Para medir la efectividad de la metodología de aplicar el proyecto de investigación en la asignatura se relaciona las puntuaciones de PC1 y las puntuaciones del Examen Final.

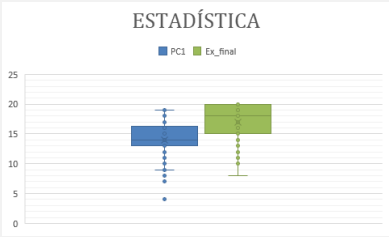


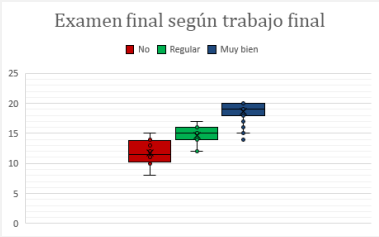


Conclusión

El proyecto fue muy efectivo para el aprendizaje del estudiante, obteniendo resultados positivos:

- Cada avance del proyecto, reforzaba sus conocimientos sobre los contenidos de las Estadísticas.
- En cada etapa de los temas que aplicaban en el proyecto, el estudiante desarrollaban el razonamiento de la importancia y utilidad de las herramientas estadísticas.
- En la supervisión de cada trabajo he podido observar que se han logrado alcanzar los objetivos previsto del proyecto "Aprender investigando".





Reflexiones del estudiante:

"Estas tareas, fueron muy efectivas para todos ya que nos permitió reforzar en el tema visto cada semana. Nos quitó de dudas y ayudó a practicar."

"Quería resaltar que este trabajo ha sido muy útil para terminar de entender los conceptos dados en clase, a través de todo el proceso de aprendizaje durante el semestre."

Errores de estudiantes del grado en Educación Primaria y su tratamiento didáctico

David Molina[1], Beatriz Cobo[2], Ramón Ferri[2]

[1]Departamento de Didáctica de la Matemática, Universidad de Granada
[2]Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: dmolinam@ugr.es

Resumen

Los contenidos relativos a la estadística y la probabilidad conforman uno de los cinco bloques en los que se divide la materia de Matemáticas a lo largo de la Educación Primaria. Consecuentemente, se espera que los maestros de Educación Primaria cuenten con un profundo conocimiento de estos contenidos que le permita un adecuado desempeño de sus tareas docentes. Sin embargo, ya desde su etapa como estudiantes, es frecuente encontrar carencias en la formación estadístico-probabilística de los futuros maestros que les llevan a incurrir en errores. En este trabajo se analizan las respuestas a varios ítems con contenido estadístico y probabilístico de un grupo de estudiantes del grado en Educación Primaria de la Universidad de Granada. Se pretende identificar cuáles son los errores más comunes y, a partir de ellos, proponer estrategias didácticas que sirvan para evitar que los estudiantes los cometan de nuevo en el futuro.

Palabras claves: educación primaria, errores, didáctica

Errores de estudiantes del grado en Educación Primaria y su tratamiento didáctico

David Molina Muñoz[1]*, Beatriz Cobo Rodríguez[2] y Ramón Ferri García[2]

[1] Departamento de Didáctica de la Matemática, Universidad de Granada

[2] Departamento de Estadística e Investigación Operativa, Universidad de Granada

*dmolinam@ugr.es

Introducción

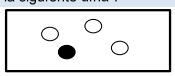
Los **errores** son datos **empíricos**, inherentes al proceso de aprendizaje, que ponen de manifiesto un desconocimiento o un conocimiento inadecuado sobre un concepto o un procedimiento por parte de un alumno o un grupo de alumnos (Rico, 2015). El carácter formal de las matemáticas hace que los errores aparezcan con mucha frecuencia en todas las áreas de esta materia. Sin embargo, pese a constituir un resultado indeseable, los errores **pueden tratarse como oportunidades de aprendizaje** a partir de las cuales reorganizar y completar el conocimiento. Tal y como indican Batanero et al. (1994), los los errores no se presentan de un modo aleatorio sino que, con frecuencia, es posible encontrar regularidades, ciertas asociaciones con variables propias de las tareas propuestas, de los sujetos o de las circunstancias presentes o pasadas. De hecho, existen líneas de investigación centradas en el estudio curricular de los errores de aprendizaje en matemáticas (Rico, 1995).

Los errores tienen una especial relevancia para los maestros en formación por dos motivos: por un lado, deben evitarlos para no transmitir a sus futuros alumnos un conocimiento incorrecto y, por otro lado, deben reconocerlos en sus estudiantes para corregirlos.

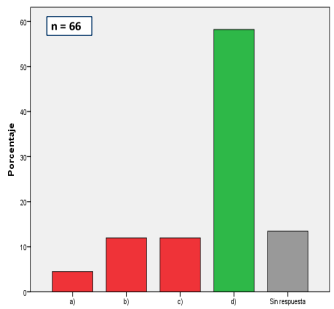
El **objetivo** de este trabajo es analizar las respuestas a varios ítems con contenido estadístico y probabilístico de un grupo de estudiantes del grado en Educación Primaria de la Universidad de Granada. Se pretende identificar algunos de los errores más comunes y, a partir de ellos, proponer aproximaciones didácticas que sirvan para evitar que los estudiantes los cometan de nuevo en el futuro.

Ítem 1

Ante el experimento aleatorio "Sacar una bola de la siguiente uma":



a) "Sacar una bola blanca" es un suceso seguro.
b) La probabilidad de sacar una bola blanca es el doble que la de sacar una bola negra.
c) La probabilidad de sacar una bola negra es 1/3.
d) El suceso complementario de "Sacar una bola blanca" es "Sacar una bola negra".



Errores

- No conocer el concepto de **suceso seguro** .
- No conocer o aplicar incorrectamente la **regla de Laplace**.

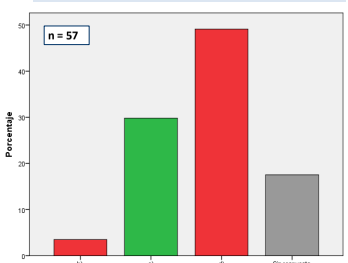
Tratamiento didáctico

- Enfocar el concepto de suceso seguro como **el único posible** en un experimento aleatorio: hacer ver que desde el momento en el que para un experimento aleatorio hay dos sucesos que pueden darse con una probabilidad mayor que 0, ninguno de ellos puede ser seguro.
- Enfatizar el **conteo** como medio para determinar el número de casos posibles y casos favorables a la hora de aplicar la regla de Laplace.

Ítem 2

La moda de un conjunto de datos…

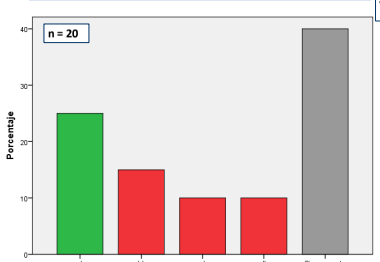
a) …coincide siempre con la mediana.
b) …solo puede calcularse para variable cuantitativas.
c) …puede no ser única.
d) …es la mayor frecuencia absoluta.



Ítem 3

¿Cuál de las siguientes medidas define mejor la tendencia central de los datos 5, 4, 42, 4, 6?

a) La mediana.
b) La media.
c) El rango.
d) La desviación típica.



Errores

- Confundir la **moda** con la **mayor frecuencia absoluta**.
- No distinguir entre **medidas de tendencia central** y **de dispersión**.
- No tener en cuenta la **sensibilidad** de la media **a los valores atípicos**.

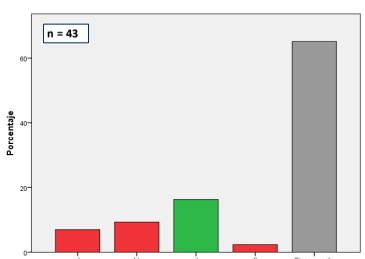
Tratamiento didáctico

- Calcular la moda para una **variable cualitativa** para hacer ver que coincide con una o varias modalidades (valores de la variable) y nunca con frecuencias.
- Calcular la media y la mediana para un conjunto de datos y, a continuación repetir el cálculo tras añadir un dato atípico para apreciar el efecto sobre ambas medidas.
- Enfatizar que la mediana se calcula a partir de la **posición** de los datos dentro del conjunto y no de su **valor**.

Ítem 4

¿Cuál de las siguientes es una variable cuantitativa discreta?

a) Distancia (en km.) recorrida en bicicleta el mes pasado por un grupo de ciclistas.
b) Calificación media del expediente de un grupo de alumnos.
c) Número de veces que llovió el año pasado en cada capital de provincia de España.
d) Temperatura del asfalto en las calles de la ciudad de Ceuta.



Errores

- No distinguir entre variables cuantitativas discretas y cuantitativas continuas.

Tratamiento didáctico

- Caracterizar las **variables cuantitativas continuas** como aquellas cuyos valores resultan de una **medición** y las **variables cuantitativas discretas** como aquellas cuyos valores resultan de un **conteo**.

Conclusiones

- **Notable tasa de error** (cercana al 50% en algún caso) en conceptos elementales sobre estadística y probabilidad en maestros en formación.
- **Importante tasa de no respuesta** en algunos ítems (por encima del 60% en algún caso).
- Los errores encontrados coinciden con los identificados por otros autores (como Carvalho (1998), citado por Batanero (2000) o Estrada, Batanero y Fortuny (2003).
- Es necesario fomentar el **razonamiento crítico** de los estudiantes como herramienta para detectar errores.
- Junto a los contenidos, podría ser aconsejable que el docente presentara una **lista con los errores más comunes**, a fin de evitarlos.
- Si ya se ha producido, **hacer útil el error** para, a partir de él, reorganizar el conocimiento: **la simple corrección o penalización no es recomendable**.

Bibliografía

- Batanero, C. (2000). Significado y comprensión de las medidas de posición central. UNO, 2000, 25, 41-58.
- Batanero, C., Díaz-Godino, J, Green, D. R., Holmes, P. y Vallecillos, A. (1994). Errores y dificultades en la comprensión de los conceptos estadísticos elementales. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547
- Estrada, Mª. A., Batanero, C. y Fortuny, J. (2003). Dificultades de los profesores en formación en conceptos estadísticos elementales. En E. Castro (ed.), Investigación en Educación Matemática: séptimo Simposio de la Sociedad Española de Investigación en Educación Matemática. (pp. 201-212). Granada: España: Universidad de Granada.
- Rico, L. (1995). Errores y dificultades en el aprendizaje de las Matemáticas. En Kilpatrik, J., Gómez, P. y Rico L. Educación Matemática. México: Grupo Editorial Iberoamérica.
- Rico, L. (2015). Matemáticas escolares y conocimiento didáctico. En Flores y Rico (Coords.) Enseñanza y aprendizaje de las matemáticas en Educación Primaria. Madrid: Editorial Pirámide.

Adaptación de las materias de Estadística Económica y Empresarial en los nuevos planes de estudio

Luis F. Rivera Galicia[1,2], Emilia I. Martos Gálvez[1,2], Juana Domínguez Domínguez[1,2], Eva Senra Díaz[1,2], Fco. Javier Callealta Barroso[1,2]

[1]Departamento de Economía, Universidad de Alcalá
[2]Grupo de Innovación Docente para la Enseñanza de los Métodos Cuantitativos, Universidad de Alcalá

Contacto: luisf.rivera@uah.es

Resumen

Durante el curso 2009-2010 se introdujeron en Alcalá los nuevos planes de estudios basados en el Espacio Europeo de Educación Superior (EEES). Debido a la acreditación de los Grados impartidos en la Facultad de Ciencias Económicas, Empresariales y Turismo de la Universidad de Alcalá durante el curso 2017-18, se ha producido un cambio en los planes de estudio de nuestras titulaciones y, consecuentemente, se ha modificado la estructura de las materias impartidas por la Unidad de Estadística del Departamento de Economía. En este trabajo, se presenta la configuración de todas las asignaturas básicas y obligatorias de Estadística en los distintos planes de estudio renovados, así como los contenidos en los que se ha decidido estructurar su impartición. Ante esta nueva situación, las asignaturas básicas y obligatorias de Estadística tienen como máximo 15 créditos, que se han divido en tres grandes bloques: Estadística Descriptiva, Probabilidad y modelos de distribución e Inferencia Estadística. Dependiendo del plan de estudios de cada Grado, estos bloques se agruparán según sus necesidades.

Palabras claves: programación, planes de estudio, organización docente

Agradecimientos. Los autores agradecen el apoyo de la Universidad de Alcalá a través del proyecto UAH/EV1075: *Adaptación y modernización de los Contenidos de Métodos Cuantitativos para la Economía y la Empresa en la Facultad de Ciencias Económicas, Empresariales y Turismo*.

Adaptación de las materias de Estadística Económica y Empresarial en los nuevos planes de estudio

Universidad de Alcalá

Rivera Galicia, Luis F., Martos Gálvez, Emilia I., Domínguez Domínguez, Juana, Senra Díaz, Eva, Callealta Barroso, Fco. Javier

Departamento de Economía, Facultad de Ciencias Económicas, Empresariales y Turismo

Motivación del trabajo

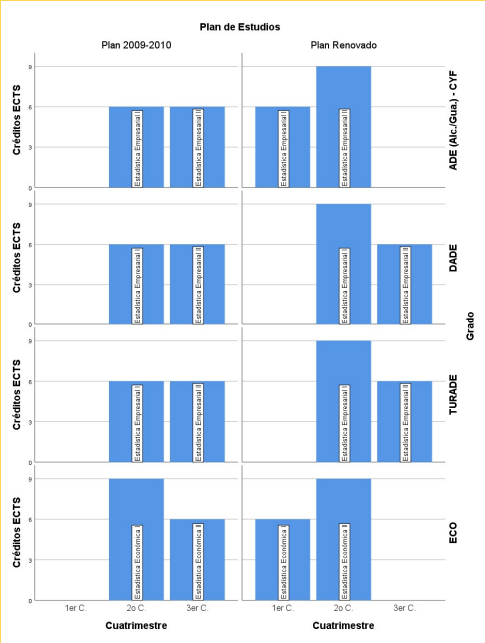
Durante el curso 2009-2010 se introdujeron en Alcalá los nuevos planes de estudios basados en el Espacio Europeo de Educación Superior (EEES). Debido a la acreditación de los Grados impartidos en la Facultad de Ciencias Económicas, Empresariales y Turismo de la Universidad de Alcalá durante el curso 2017-18, se ha producido un cambio en los planes de estudio de nuestras titulaciones y, consecuentemente, se ha modificado la estructura de las materias impartidas por la Unidad de Estadística del Departamento de Economía.

En este trabajo, se presenta la configuración todas las asignaturas básicas y obligatorias de Estadística en los distintitos planes de estudio renovados, así como los contenidos en los que se ha decidido estructurar su impartición.

En algunos casos se ha aumentado el número de créditos en las titulaciones, lo que hace más fácil la docencia de nuestra materia.

Características de los Grados en la UAH

- Modelo de enseñanza centrado en el alumno.
- Carga medida en ECTS: 1 ECTS =
  - 25-30 horas trabajo alumno.
  - 8 horas de clase presencial.
- Gran carga de trabajo para el alumno y pocas horas de contacto con el profesor.



* La configuración de las asignaturas en los Grados de ENI (Estadística Económica – 9 ECTS – 2º C.) y Turismo (Estadística Aplicada – 6 ECTS – 1er C.) no cambian

Estructuración de contenidos

Estadística Descriptiva — 6 ECTS

- Introducción y utilidad de la Estadística en la Economía y la Empresa
- Estadística Unidimensional
- Estadística Bidimensional e introducción a la regresión
- Análisis descriptivo de series temporales
- Números Índices: elaboración e interpretación

Cálculo de probabilidades — 3 ECTS

- Fenómenos aleatorios y al Cálculo de probabilidades
- Variables aleatorias y sus características
- Modelos de distribuciones de probabilidad

Inferencia Estadística — 6 ECTS

- Estimación puntual y por intervalos de confianza
- Contrastes de hipótesis paramétricas
- Análisis de la Varianza
- Contrastes de bondad de ajuste y tablas de contingencia
- Contrastes no paramétricos

GIDEMECU - Grupo de Innovación Docente para la Enseñanza de los MÉtodos CUantitativos
Proyecto UAH/EV1075: Adaptación y modernización de los Contenidos de Métodos Cuantitativos para la Economía y la Empresa en la Facultad de Ciencias Económicas, Empresariales y Turismo

Un método computacional para mejorar la interpretación de los intervalos de confianza para la proporción en Bachillerato y grados universitarios

Antonio Francisco Roldán López de Hierro

Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: aroldan@ugr.es

Resumen

Los intervalos de confianza suponen un procedimiento inferencial de gran interés con muchas aplicaciones en la vida real. Este tema se enseña en España a estudiantes de Bachillerato en la especialidad de Ciencias Sociales, y también en la mayoría de los grados universitarios. Como el tiempo disponible para la enseñanza es escaso, a los estudiantes se les enseña principalmente el procedimiento de cálculo, sin prestar mucha atención a la interpretación de los resultados. Olivo, Batanero y Díaz (2008) describieron algunas dificultades que aparecen en la interpretación de los intervalos de confianza de los estudiantes. En este trabajo proponemos un método computacional de cálculo del intervalo de confianza para la proporción siguiendo la metodología de Wilson (1927). Su utilización en Bachillerato (y en grados) contribuye a aclarar muchos de los errores que comete el alumnado a la hora de interpretar tanto el significado de los propios intervalos de confianza como del nivel de confianza.

Palabras claves: Proporción, nivel de confianza, intervalo de confianza, interpretación

Introducción

Los intervalos de confianza suponen un procedimiento inferencial muy importante con muchas aplicaciones en la vida real, como estudiar el porcentaje de votantes que finalmente votará por un partido político en particular o la proporción de la población que sufre de asma. Como el tiempo disponible para la enseñanza es escaso, a los estudiantes se les enseña principalmente el procedimiento de cálculo, sin prestar mucha atención a la interpretación de los resultados. Olivo, Batanero y Díaz (2008) describieron algunas dificultades que aparecen en la interpretación de los intervalos de confianza para la proporción.

El objetivo de este trabajo es presentar una metodología alternativa para calcular los intervalos de confianza para la proporción utilizando el enfoque

debido a Wilson (1927). Sería interesante para los estudiantes descubrir que no existe una forma única de calcular dicho intervalo. Con la ayuda de ordenadores, al comparar el procedimiento de Wilson con el clásico, los estudiantes podrían mejorar su interpretación del intervalo de confianza para la proporción y de su nivel de confianza.

Intervalos de confianza para la proporción

Supongamos que estamos interesados en estimar la proporción p de individuos de una población que satisfacen una cierta condición. Tomamos una muestral aleatoria de la población y observamos el número X de individuos de la muestra que satisfacen esa propiedad, la cual sigue una distribución Binomial de parámetros n y p. Cuando n es grande, aproximamos esta distribución por la distribución Normal $X \approx \widetilde{X} \hookrightarrow \mathcal{N}(np, npq)$, donde $q = 1 - p$, que puede tipificarse como $(\widehat{p} - p)/\sqrt{pq/n} \hookrightarrow \mathcal{N}(0,1)$ donde $\widehat{p}$ es la proporción muestral definida como el número de éxitos dividido entre el número total de observaciones de la muestra. Denotemos por $1-\alpha$ al nivel de confianza (usualmente $1-\alpha \in \{0.9, 0.95, 0.99\}$) y sea $z_{1-\alpha/2} > 0$ el único número real positivo tal que $P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1-\alpha$, donde Z es cualquier variable aleatoria con distribución Normal estándar $\mathcal{N}(0,1)$. Entonces podemos considerer la siguiente aproximación:

$$P\left(-z_{1-\alpha/2} < \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{1-\alpha/2}\right) = 1 - \alpha. \tag{1}$$

Dado que p es desconocido, lo reemplazamos por su estimador de máxima verosimilitud, $\widehat{p}$, de donde es sencillo deducir que el intervalo de confianza para p, al nivel de confianza $1-\alpha$, es:

$$IC(p) = \left[\ \widehat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\ \right]. \tag{2}$$

La ventaja de esta expresión es que es muy simple de usar en la práctica y produce buenos resultados. Su principal debilidad es que se ha obtenido a través de un proceso en el que, sin una justificación aparente, el verdadero valor de p ha sido reemplazado por su estimador $\widehat{p}$. Este reemplazo transforma un problema no lineal en un problema lineal. Uno de los inconvenientes más importantes de la expresión (2) es que esta fórmula no alcanza el nivel de confianza.

La metodología de Wilson

Esta metodología propone resolver la ecuación (de segundo grado tras elevar al cuadrado):

$$\frac{\widehat{p}-p}{\sqrt{\dfrac{p\,(1-p)}{n}}}=z_{1-\alpha/2}$$

antes de reemplazar p por $\widehat{p}$ en el denominador. Sus dos soluciones determinan los extremos del intervalo de confianza para la proporción, que es:

$$IC\,(p)=\left[\;\frac{2n\widehat{p}+z_{1-\alpha/2}^2\pm z_{1-\alpha/2}\sqrt{z_{1-\alpha/2}^2+4n\,\widehat{p}\,(1-\widehat{p}\,)}}{2(n+z_{1-\alpha/2}^2)}\;\right]. \qquad (3)$$

Una aproximación TIC

Dado que la fórmula (3) es más complicada que (2), proponemos utilizar una hoja de cálculo para generar muchas muestras aleatorias de la distribución binomial $\mathcal{B}\,(n,p)$ (por ejemplo, 1000 o más), cuando n y p son previamente fijados por los estudiantes. Utilizando esas muestras, calculamos los intervalos (2) y (3), y determinamos cuántos de ellos realmente contienen al verdadero valor de p. Esta comparación ayuda a los estudiantes a comprender que la interpretación correcta del nivel de confianza se basa en la metodología, pero no en el intervalo: cuando se generan aleatoriamente muchos intervalos, confiamos en que el $100\,(1-\alpha)\,\%$ de ellos contendrá al verdadero (pero desconocido) valor de la proporción p. Además, estamos interesados en corregir la afirmación errónea de los estudiantes de que, después de calcular el intervalo de confianza, la probabilidad de que p pertenezca a tal intervalo es el nivel de confianza. Este punto de vista novedoso podría ser de interés tanto para estudiantes universitarios como para estudiantes de Bachillerato.

Para comparar ambos métodos, presentamos los resultados de una simulación con 10000 y 1000000 datos tomando $n=30$ (ver Tabla 1). Destacamos que solo el 88 % de los intervalos construidos usando (2) contiene al verdadero valor de p y que la metodología de Wilson mejora este porcentaje.

Metodología clásica		Metodología de Wilson	
$N = 10000$	$N = 1000000$	$N = 10000$	$N = 1000000$
$87.19\,\%$	$87.54\,\%$	$94.88\,\%$	$95.23\,\%$

Tabla 1. Proporción de intervalos de confianza en la simulación que contienen al verdadero valor de p

Choose the following options					☑
Poblational proportion "p"	0,9	Check for 'p'	Adequate value for 'p'		Ok
Size of random sample "n"	150	Check for 'n'	Adequate value for 'n'		
Confidence level (%)	95%	Critical value	1,9600		

Number of computed intervals	1000	
	Classical	Wilson
Number of random samples containing the true value of "p"	927	953
Proportion of random samples containing the true value of "p"	92,70%	95,30%

Random samples			CLASSICAL confidence interval			WILSON confidence interval		
No.	Data	Sample proportion	Lower endpoint	Upper endpoint	Does "p" belong to CI?	Lower endpoint	Upper endpoint	Does "p" belong to CI?
1	138	0,920	87,66%	96,34%	Yes	86,54%	95,36%	Yes
2	136	0,907	86,01%	95,32%	Yes	84,94%	94,36%	Yes
3	132	0,880	82,80%	93,20%	Yes	81,83%	92,27%	Yes
4	136	0,907	86,01%	95,32%	Yes	84,94%	94,36%	Yes
5	134	0,893	84,39%	94,27%	Yes	83,38%	93,33%	Yes
6	141	0,940	90,20%	97,80%	No	88,99%	96,81%	Yes

Figura 1. Captura de pantalla de la simulación desarrollada

Conclusiones

1) En este trabajo se ha introducido un procedimiento alternativo (basado en la aproximación de Wilson) para determinar el intervalo de confianza para la proporción.

2) Sirve para que los estudiantes aprendan la interpretación correcta y el significado del nivel de confianza que se utiliza.

3) Al comparar ambos métodos, los estudiantes profundizan en la importancia del intervalo de confianza y en cómo ambos procedimientos implican suposiciones y aproximaciones adecuadas.

4) Además, pueden realizar sus propias simulaciones utilizando recursos tecnológicos (como hojas de cálculo).

Agradecimientos. El autor agradece el apoyo del Ministerio de Economía y Competitividad de España en el proyecto TIN2017-89517-P.

Referencias

[1] E.B. Wilson (1927) Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, pp. 209-212.

[2] E. Olivo, C. Batanero y C. Díaz (2008) Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación Matemática*, **20**(3), pp. 55-82.

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

Un método alternativo para calcular intervalos de confianza para la proporción en Bachillerato

Antonio Francisco Roldán López de Hierro
Universidad de Granada

1. Introducción

Los intervalos de confianza suponen un procedimiento inferencial muy importante con muchas aplicaciones en la vida real, como estudiar el porcentaje de votantes que finalmente votará por un partido político en particular o la proporción de la población que sufre de asma. Como el tiempo disponible para la enseñanza es escaso, a los estudiantes se les enseña principalmente el procedimiento de cálculo, sin prestar mucha atención a la interpretación de los resultados. Olivo, Batanero y Díaz (2008) describieron algunas dificultades que aparecen en la interpretación de los intervalos de confianza para la proporción.

El objetivo de este póster es presentar una metodología alternativa para calcular los intervalos de confianza para la proporción utilizando el enfoque debido a Wilson (1927). Sería interesante para los estudiantes descubrir que no existe una forma única de calcular dicho intervalo. Con la ayuda de ordenadores, al comparar el procedimiento de Wilson con el clásico, **los estudiantes podrían mejorar su interpretación del intervalo de confianza para la proporción y de su nivel de confianza**.

2. Intervalos de confianza para la proporción

Supongamos que estamos interesados en estimar la proporción p de individuos de una oblación que satisfacen una cierta condición. Tomamos una muestral aleatoria de la población y observamos el número X de individuos de la muestral que satisfacen esa propiedad, la cual sigue una distribución Binomial de parámetros n y p. Cuando n es grande, aproximamos esta distribución por la distribución Normal $X \approx \tilde{X} \mapsto N(np, npq)$, donde $q = 1 - p$, que puede tipificarse como $(\hat{p} - p)/\sqrt{pq/n} \mapsto N(0,1)$ donde $\hat{p}$ es la proporción muestral definida como el número de éxitos dividido entre el número total de observaciones de la muestra. Denotemos por $1 - \alpha$ al nivel de confianza (usualmente $1 - \alpha \in \{0.9, 0.95, 0.99\}$) y sea $z_{1-\alpha/2} > 0$ el único número real positive tal que $P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$, donde Z es cualquier variable aleatoria con distribución Normal estándar $N(0,1)$. Entonces podemos considerer la siguiente aproximación:

$$P\left(-z_{1-\alpha/2} < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < z_{1-\alpha/2}\right) = 1 - \alpha. \qquad (1)$$

Dado que p es desconocido, lo reemplazamos por su estimador de máxima verosimilitud, $\hat{p}$, de donde es sencillo deducir que que el interval de confianza para p, al nivel de confianza $1 - \alpha$, es:

$$IC(p) = \left[\ \hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}\,(1-\hat{p})/n}\ \right]. \qquad (2)$$

La ventaja de esta expresión es que es muy simple de usar en la práctica y produce buenos resultados. Su principal debilidad es que se ha obtenido a través de un proceso en el que, sin una justificación aparente, el verdadero valor de p ha sido reemplazado por su estimador $\hat{p}$. Este reemplazo transforma un problema no lineal en un problema lineal. Uno de los inconvenientes más importantes de la expresión (2) es que esta fórmula no alcanza el nivel de confianza.

3. La metodología de Wilson

Esta metodología propone resolver la ecuación (de segundo grado tras elevar al cuadrado):

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = z_{1-\alpha/2}$$

antes de reemplazar p por $\hat{p}$ en el denominador. Sus dos soluciones determinan los extremos del interval de confianza para la proporción, que es:

$$IC_W(p) = \left[\frac{2n\hat{p} + z_{1-\alpha/2}^2 \pm z_{1-\alpha/2}\sqrt{z_{1-\alpha/2}^2 + 4n\hat{p}(1-\hat{p})}}{2(n + z_{1-\alpha/2}^2)}\right]. \qquad (3)$$

4. Una aproximación TIC

Dado que la fórmula (3) es más complicada que (2), proponemos utilizar una hoja de cálculo para generar muchas muestras aleatorias de la distribución $B(n,p)$ (por ejemplo, 1000 o más), cuando n y p son previamente fijados por los estudiantes. Utilizando esas muestras, calculamos los intervalos (2) y (3), y determinamos cuántos de ellos realmente contienen al verdadero valor de p. Esta comparación ayuda a los estudiantes a comprender que **la interpretación correcta del nivel de confianza se basa en la metodología, pero no en el intervalo: cuando se generan aleatoriamente muchos intervalos, confiamos en que el $100(1-\alpha)\%$ de ellos contendrá al verdadero (pero desconocido) valor de la proporción p**. Además, estamos interesados en supercar la afirmación errónea de los estudiantes de que, después de calcular el interval de confianza, la probabilidad de que p pertenezca a tal interval es el nivel de confianza. Este punto de vista novedoso podría ser de interés tanto para estudiantes universitarios como para estudiantes de Bachillerato.

Para comparer ambos métodos, presentamos los resultados de una simulación con 10000 y 1000000 de datos tomando $n = 30$ (ver Tabla 1). Destacamos que solo el 88% de los intervalos construidos usando (2) contiene al verdadero valor de p y que la metodología de Wilson mejora este porcentaje.

Tabla 1. Proporción de intervalos de confianza en la simulación que contienen al verdadero valor de p.

Metodología clásica		Metodología de Wilson	
$N = 10000$	$N = 1000000$	$N = 10000$	$N = 1000000$
87.19%	87.54%	94.88 %	95.23 %

Choose the following options

Poblational proportion "p"	0,9	Check for 'p'	Adequate value for 'p'	Ok
Size of random sample "n"	150	Check for 'n'	Adequate value for 'n'	
Confidence level (%)	95%	Critical value	1,9600	

Number of computed intervals	1000	
	Classical	Wilson
Number of random samples containing the true value of "p"	927	953
Proportion of random samples containing the true value of "p"	92,70%	95,30%

Random samples			CLASSICAL confidence interval			WILSON confidence interval		
No.	Data	Sample proportion	Lower endpoint	Upper endpoint	Does "p" belong to CI?	Lower endpoint	Upper endpoint	Does "p" belong to CI?
1	138	0,920	87,66%	96,34%	Yes	86,54%	95,36%	Yes
2	136	0,907	86,01%	95,32%	Yes	84,94%	94,36%	Yes
3	132	0,880	82,80%	93,20%	Yes	81,83%	92,27%	Yes
4	136	0,907	86,01%	95,32%	Yes	84,94%	94,36%	Yes
5	134	0,893	84,39%	94,27%	Yes	83,38%	93,33%	Yes
6	141	0,940	90,20%	97,80%	No	88,99%	96,81%	Yes

5. Conclusiones

1) En este póster se ha introducido un procedimiento alternativo (basado en la aproximación de Wilson) para determinar el intervalo de confianza para la proporción.
2) Sirve para que los estudiantes aprendan la interpretación correcta y el significado del nivel de confianza que se utiliza.
3) Al comparar ambos métodos, los estudiantes profundizan en la importancia del intervalo de confianza y en cómo ambos procedimientos implican suposiciones y aproximaciones adecuadas.
4) Además, pueden realizar sus propias simulaciones utilizando recursos tecnológicos (como hojas de cálculo).

6. Referencias

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.

Olivo, E., Batanero, C. and Díaz, C. (2008). Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación Matemática*, 20(3), 55-82.

Agradecimientos: Grupo de investigación FQM268 y Proyecto TIN2017-89517-P

Departamento de Estadística e Investigación Operativa
aroldan@ugr.es

Herramientas computacionales para el aprendizaje de las distribuciones Tipo Fase: Aplicación con datos reales de memorias resistivas

Juan Eloy Ruiz Castro, Christian J. Acal, Ana M. Aguilera

Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: jeloy@ugr.es

Resumen

La fiabilidad es una rama de la Estadística altamente relacionada con la Ingeniería que estudia y analiza el comportamiento de sistemas sujetos a fallo donde la probabilidad juega un papel fundamental en la modelización, resolución y optimización de problemas. En el ámbito de la docencia, se suelen desarrollar metodologías que permiten un estudio detallado del comportamiento de dispositivos, introduciendo técnicas clásicas con algunas de las distribuciones más importantes en el campo de la fiabilidad. Para el uso de estas distribuciones es muy importante la estimación de los parámetros. Sin embargo, desde un punto de vista teórico y práctico, existen distribuciones con buenas propiedades algebraicas que generalizan otras distribuciones clásicas facilitando el desarrollo metodológico. Son las distribuciones Tipo Fase. En este trabajo se presenta una nueva metodología para el aprendizaje de la estimación en fiabilidad mediante programas propios en R y Matlab, y mediante EMpht. Estas técnicas se aplican a un conjunto de datos reales de memorias RRAM, demostrando que este enfoque funciona mejor que el análisis estadístico clásico empleado en esta área.

Palabras claves: Fiabilidad, modelización, probabilidad, distribuciones Tipo Fase, memorias resistivas

Agradecimientos. Los autores agradecen al Ministerio de Ciencia, Innovación y Universidad de España su apoyo bajo el proyecto MTM2017-87708-P, también respaldado por el programa FEDER y a la Junta de Andalucía por el grupo FQM-307.

Referencias

[1] C. Acal, J.E. Ruiz-Castro y A.M. Aguilera (2019) Distribuciones tipo fase en un estudio de fiabilidad. *TEMat*, **3**, pp. 63-74.

[2] C. Acal, J.E. Ruiz-Castro, A.M. Aguilera, F. Jimenez-Molinos y J.B. Roldan (2019) Phase-type distributions for studying variability in resistive memories. *Journal of Computational and Applied Mathematics*, **345**, pp. 23-32.

[3] S. Asmussen (2000) *Ruin probabilities*. World Scientific, Hong Kong.

[4] M. F. Neuts (1975) *Probability distributions of phase type*. Liber Amicorum Professor Emeritus Dr. H. Florin.

[5] E. Pérez, D. Maldonado, C. Acal, J.E. Ruiz-Castro, F.J. Alonso, A.M. Aguilera, F. Jiménez-Molinos, Ch. Wenger y J.B. Roldán (2019) Analysis of the statistics of device-to-device and cycle-to-cycle variability in TiN/Ti/Al:HfO2/TiN RRAMs. *Microelectronics Engineering*, **214**(1), pp. 104-109.

Herramientas computacionales para el aprendizaje de las distribuciones Tipo Fase: Aplicación con datos reales de memorias resistivas

UNIVERSIDAD DE GRANADA

Juan Eloy Ruiz Castro, Christian J. Acal González, Ana M. Aguilera del Pino

jeloy@ugr.es; chracal@ugr.es; aaguiler@ugr.es

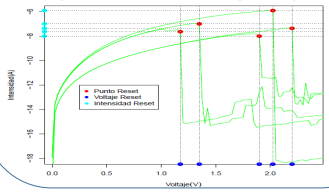
RESUMEN Y OBJETIVOS DEL ESTUDIO

La fiabilidad es una rama de la Estadística altamente relacionada con la Ingeniería que estudia y analiza el comportamiento de sistemas sujetos a fallo donde la probabilidad juega un papel fundamental en la modelización, resolución y optimización de problemas. En el ámbito de la docencia, se suelen desarrollar metodologías que permiten un estudio detallado del comportamiento de dispositivos, introduciendo técnicas clásicas con algunas de las distribuciones más importantes en el campo de la fiabilidad. Para el uso de estas distribuciones es muy importante la estimación de los parámetros.

Sin embargo, desde un punto de vista teórico y práctico, existen distribuciones con buenas propiedades algebraicas que generalizan otras distribuciones clásicas facilitando el desarrollo metodológico. Son las distribuciones Tipo Fase. En este trabajo se presenta una nueva metodología para el aprendizaje de la estimación en fiabilidad mediante programas propios en R y Matlab, y mediante EMpht. Estas técnicas se aplican a un conjunto de datos reales de memorias RRAM, demostrando que este enfoque funciona mejor que el análisis estadístico clásico empleado en esta área.

CURVAS RESET

```
num=c(5,29,101,575,1970) #CURVAS A PINTAR
set.seed(5);v=abs(rnorm(100));i=abs(rnorm(100))
plot(v,i,type="n",xlab=expression(Voltaje(V)),ylab="Intensidad(A)",ylim=c(-18,-6))
for(j in 1:length(num)){
  curva< - read.table(paste(dir,paste("Cycle_R",num[j],".txt",sep=""),sep="/"))
  v=abs(curva$V1);i=log(curva$V2)
  pp=sapply(2:length(curva$V2),
        function(j)(curva$V2[(j-1)]-curva$V2[j])>=(0.7*curva$V2[(j-1)]))
  p=min(which(pp==TRUE))
  lines(v,i,type="l",col="green",lwd=1.5); points(v[p],i[p],col="red",lwd=4)
  points(v[p],(par("usr")[3]),col="blue",lwd=4,xpd=TRUE)
  points((par("usr")[1]),i[p],col="cyan",lwd=4,xpd=TRUE)
  abline(v=v[p],col="black",lty=3);abline(h=i[p],col="black",lty=3)          }
legend(0.25,-12,c("Punto Reset","Voltaje Reset","Intensidad
      Reset"),pch=c(16,16,16),col=c("red","blue","cyan"),box.lty=1)
```



ASPECTOS TEORICOS

- **Funciones y distribuciones usuales en estudios de fiabilidad.**
  - **Función de distribución. $F(t)$**
  - **Función de fiabilidad. $R(t)=1-F(t)$**
- **Análisis gráfico.**
- **Distribuciones Tipo Fase**

EXPRESIONES

DISTRIBUCIÓN WEIBULL

$$R(t) = \exp\{-(\lambda t)^{\beta}\};\ \ t \geq 0;\ \ \lambda, \beta > 0$$

DISTRIBUCIONES TIPO FASE

$$R(t) = \alpha \exp(Tt)e;\ \ t \geq 0$$

CASO PARTICULAR DE LA ERLANG

$$\alpha = (1, 0, \ldots, 0) \quad , \quad T = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\lambda & \lambda \\ 0 & \cdots & \cdots & 0 & -\lambda \end{pmatrix}$$

LECTURA PUNTOS RESET

```
#-> DIRECTORIO QUE CONTIENE LOS ENSAYOS
dir="DIRECTORIO DONDE SE ENCUENTRE LOS DATOS"

#-> Vr,Ir e Tr VECTORES DONDE GUARDAREMOS LOS PUNTOS CRÍTICOS
Vr=vector();Ir=vector();Tr=vector()

#-> n NÚMERO DE CICLOS REGISTRADOS EN EL LABORATORIO
n=2818

#->FUNCIÓN PARA DETERMINAR LOS PUNTOS RESET
for(i in 1:n){
file <- read.table(paste(dir,paste("Cycle_R",i,".txt",sep=""),sep="/"))
        file$V1=abs(file$V1)
        vol=file$V1;int=file$V2
        a=sapply(2:length(int), function(j)(int[(j-1)]-int[j])>=(0.7*int[(j-1)]))
        punto=min(which(a==TRUE))
Vr=c(Vr,file[punto,1]); Ir=c(Ir,file[punto,2]); Tr=c(Tr,file[punto,3])
}
reset=matrix(c(Vr,Ir,Tr),ncol=3);
reset=as.data.frame(reset);names(reset)=c("Vreset","Ireset","Treset")

#-> ELIMINAMOS AQUELLOS CICLOS DONDE NO SE PRODUCEN PUNTOS RESET
na=which(is.na(Vr));reset=reset[-na,]

#-> ELIMINAMOS LOS PUNTOS ANÓMALOS ENCONTRADOS
eli=which(log(reset$Treset)<0);reset=reset[-eli,]
```

AJUSTE WEIBULL

```
mc=function(d,fig=TRUE,tipo){
    datos=sort(d)
    x=log(datos); pi=(1:length(x)-0.5)/length(x); y=log(-log(1-pi))  ## NUBE DE PUNTOS

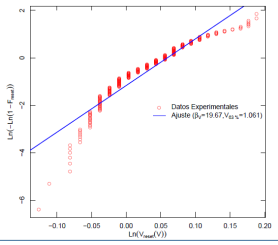
recta=lm(y~x)
    b=recta$coefficients[1]  ## TERMINO INDEPENDIENTE DE LA RECTA y=ax+b
    a=recta$coefficients[2]  ## PENDIENTE DE LA RECTA y=ax+b
    correlacion=round(summary(recta)$r.squared,2) ## BONDAD DEL AJUSTE
    beta=a; lambda=exp(b/beta) ## ESTIMACION PARAMETROS WEIBULL

#-> REPRESENTACIÓN GRÁFICA
if(fig==TRUE){
if(tipo=="RESET"){
plot(x,y,xlab=expression(Ln(V[reset](V))),mgp = c(1.75, 0.5, 0),tck=0.025,
    ylab=expression(Ln(-Ln(1-F[reset]))),main="Ajuste por Weibull",col="red")        }
if(tipo=="SET"){
plot(x,y,xlab=expression(Ln(V[set](V))),mgp = c(1.75, 0.5, 0),tck=0.025,
    ylab=expression(Ln(-Ln(1-F[set]))),main="Ajuste por Weibull",col="red")        }

abline(recta,col="blue",lwd=2)
axis(side=4,labels=F,tick=TRUE,tck=0.02)
axis(side=3,labels=F,tick=TRUE,tck=0.02)

#-> PARA LA LEYENDA
beta_b=round(beta,3)
v_63=round(1/lambda,3)
legend("right",c(expression("Experimental Data"),
   substitute(paste("Fitting (", beta[V], "=",b,",",V[63~'%'],"=",l,")"),list(b=beta_b,l=v_63))),
   lty=c(0,1),pch=c(1,NA),col=c("red","blue"),box.lty=0)    }

#-> SALIDA DE LA FUNCIÓN (RESULTADOS)
resultado=data.frame(correlacion,beta,lambda,b,a)
names(resultado)=c("Coef.Corr","beta","lambda","b: y=ax+b","a: y=ax+b")
return(resultado)                                                          }
```



AJUSTE ERLANG

```
library(expm); par(cex=1)

phd=function(d,tipo,fases){
    w=mc(d,fig=F);w=as.numeric(w)  ## AJUSTE WEIBULL CON EL FIN DE COMPARAR
    datos=sort(d)
    x=log(datos); pi=(1:length(x)-0.5)/length(x); y=log(-log(1-pi))

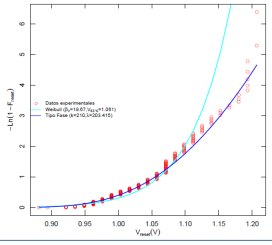
 #-> REPRESENTACIÓN GRÁFICA
if(tipo=="RESET"){
plot(exp(x),exp(y),xlab=expression(V[reset](V)),mgp = c(1.75, 0.5, 0),tck=0.025,
    ylab=expression(-Ln(1-F[reset])),main="Ajuste Weibull vs PHD",col="red")        }
if(tipo=="SET"){
plot(exp(x),exp(y),xlab=expression(V[set](V)),mgp = c(1.75, 0.5, 0),tck=0.025,
    ylab=expression(-Ln(1-F[set])),main="Ajuste Weibull vs PHD",col="red")        }

#-> REPRESENTAMOS AJUSTE WEIBULL (r=x*beta+log(lambda)*beta)
r=x*w[2]+w[4]

points(exp(x),exp(r),lwd=2,col="cyan",type="l")
axis(side=4,labels=F,tick=TRUE,tck=0.02)
axis(side=3,labels=F,tick=TRUE,tck=0.02)

#-> REPRESENTAMOS AJUSTE ERLANG CON ESTRUCTURA PHD
f=fases                    #NUMERO DE ESTADOS FIJOS
a=f/mean(datos)            #ESTIMACION M.V. DEL PARAMETRO DE LA MATRIZ T
vpi=c(1,rep(0,f-1)); uno=rep(1,f)   #VECTOR ALPHA Y VECTOR E
D0=diag(1,f)               #ESTIMACION MATRIZ T
    diag(D0)=-a; for(i in 1:f-1){ D0[i,i+1]=a}
kk=vector()                # -ln(R(t))
    for(i in 1:length(datos)){
    kk[i]=-log(vpi%*%expm(D0*datos[i])%*%uno)                                  }
points(datos,kk,type="l",col="blue",lwd=2)

#->PARA LA LEYENDA EXTRAPOLAMOS EL CODIGO EMPLEADO UTILIZADO EN WEIBULL
}
```



F. FIABILIDAD

```
rt=function(d,tipo,fases){

    w=mc(d,fig=F);w=as.numeric(w)
    datos=sort(d)
    pi=(1:length(datos)-0.5)/length(datos)

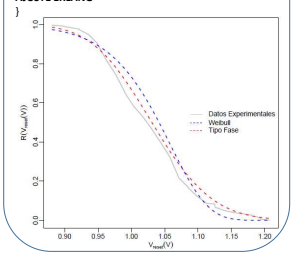
#-> CALCULO DE LA R(t) DE LA WEIBULL
beta=w[2];lambda=w[3]
w=exp(-(lambda *datos)^beta)   #R(t) WEIBULL

#-> CALCULO DE LA R(t) DE LA ERLANG
f=fases                    #NUMERO DE ESTADOS FIJOS
a=f/mean(datos)            #PARAMETRO DE LA MATRIZ T
vpi=c(1,rep(0,f-1))
uno=rep(1,f); D0=diag(1,f)  #VECTOR e Y ALPHA
diag(D0)=-a                #MATRIZ T
    for(i in 1:f-1){ D0[i,i+1]=a}
ph=vector()                #R(t) PHD
    for(i in 1:length(datos)){
    ph[i]=vpi%*%expm(D0*datos[i])%*%uno}

#-> REPRESENTACION GRÁFICA
if(tipo=="RESET"){
    plot(datos,1-pi,type="l",xlab=expression(V[reset](V)),
    ylab=expression(R(V[reset](V))),main="Comparacion
    R(T)",mgp = c(1.75, 0.5, 0),tck=0.025)          }
if(tipo=="SET"){
    plot(datos,1-pi,type="l",xlab=expression(V[set](V)),
    ylab=expression(R(V[set](V))),main="Comparacion
    R(T)",mgp = c(1.75, 0.5, 0),tck=0.025)          }

points(datos,w,type="l",col="blue",lwd=2)
points(datos,ph,type="l",col="red",lwd=2)

#-> PARA LA LEYENDA COPIAMOS EL CODIGO DEL AJUSTE ERLANG
}
```



- C. Acal, J.E. Ruiz-Castro, A.M. Aguilera, F. Jimenez-Molinos, J.B. Roldan, Phase-Type distributions for studying variability in resistive memories, Journal of Computational and Applied Mathematics. 345 (2019) 23-32.
- Keen, Kevin J. (2010) *Graphics for Statistics and Data Analysis with R*. Boca Raton, FL: Chapman & Hall/CRC.
- Sarkar, Deepayan. (2008) *Lattice: Multivariate Data Visualization with R*. New York: Springer.

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

MTM2017-87708-P

FEDER

Propuesta didáctica basada en la metodología ABP para la asignatura de Investigación de Mercados del Grado en Administración y Dirección de Empresas

Mª Carmen Segovia García, Silvia Mª Valenzuela Ruiz, Ana Esther Madrid García

Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: msegovia@ugr.es

Resumen

En este trabajo se presenta una propuesta didáctica basada en la metodología ABP (Aprendizaje Basado en Problemas) para la asignatura Investigación de Mercados del Grado en Administración y Dirección de Empresas, con la que intentamos que los alumnos comprendan y apliquen al mundo real los conocimientos estadísticos estudiados, adquieran autonomía para resolver problemas, y estén mejor preparados para su incorporación al mercado laboral. Se presentan distintas sesiones de trabajo agrupadas por bloques en función de las competencias y objetivos de aprendizaje que se pretende que alcancen los alumnos. Durante el proceso, los alumnos deberán seleccionar un tema de interés que requiera la elaboración de una encuesta, concretar la población objetivo, seleccionar el tipo de muestreo más adecuado, analizar los datos obtenidos y extraer las conclusiones finales. Todo esto deberá ser presentado por el alumno en forma de informe final. El objetivo de este trabajo es proponer una metodología de trabajo alternativa a las clases magistrales tradicionales.

Palabras claves: ABP, Educación Universitaria, Estadística

Propuesta didáctica basada en la metodología ABP para la asignatura de Investigación de Mercados del Grado en Administración y Dirección de Empresas

Mª Carmen Segovia García, Silvia Mª Valenzuela Ruiz, Ana Esther Madrid García
Departamento de Estadística e Investigación Operativa. Universidad de Granada

Problemática

Este trabajo presenta una propuesta didáctica basada en la metodología ABP (Aprendizaje Basado en Problemas) para la asignatura Investigación de Mercados del Grado en Administración y Dirección de Empresas, con la que intentamos que los alumnos comprendan y apliquen al mundo real los conocimientos estadísticos estudiados, adquieran autonomía para resolver problemas, y estén mejor preparados para su incorporación al mercado laboral. Se presentan distintas sesiones de trabajo agrupadas por bloques en función de las competencias y objetivos de aprendizaje que se pretende que alcancen los alumnos.

El objetivo de este trabajo es proponer una metodología de trabajo alternativa a las clases magistrales tradicionales.

El método ABP

Barrows (1986) define al ABP como "un método de aprendizaje basado en el principio de usar problemas como punto de partida para la adquisición e integración de los nuevos conocimientos". En este método los protagonistas del aprendizaje son los propios alumnos y el profesor actúa como orientador.

Este método trabaja diversas competencias, entre otras cabe destacar: Autonomía para resolver problemas y tomar decisiones, trabajo en equipo, mejora de la argumentación y presentación de la información, identificación de problemas relevantes en el propio contexto laboral.

Fases del método ABP

Existen numerosas propuestas para la aplicación del método (Restrepo Gómez (2005)). Los fases sugeridas en este trabajo para la aplicación del método son: 1. Introducción al método ABP, 2. Planteamiento del problema por parte del profesor, 3. Realización de lluvia de ideas. Definición de problemas de trabajo, 4. Identificación de las necesidades de aprendizaje (Conceptos y herramientas) 5. Obtención de la información, 6. Presentación de resultados.

El método se aplica en varias sesiones de trabajo alineadas con cada una de las fases del método que se describen a continuación:

Sesiones de trabajo

SESIÓN 1. INTRODUCCIÓN AL MÉTODO ABP

Presentación de la asignatura y del método docente a seguir.
- Protagonista principal: el profesor

Duración: 1 h.

SESIÓN 2. PLANTEAMIENTO DEL PROBLEMA POR PARTE DEL PROFESOR

Exposición de distintos trabajos de investigación de mercados seleccionados por el profesor en base a los contenidos de la asignatura con el fin de ofrecer diversas oportunidades de aprendizaje.
- Protagonista principal: el profesor
- Competencias: **E1**

Duración: 2 h.

SESIÓN 3. REALIZACIÓN DE LLUVIA DE IDEAS. DEFINICIÓN DE PROBLEMAS DE TRABAJO

Debate en grupos de temas de interés para el futuro profesional y selección de los que se van a investigar por cada grupo en el trascurso de la asignatura.
- Protagonistas principales: los alumnos, el profesor solo actuará como guía.
- Competencias: **E1, G5**

Duración: 2 h.

SESIÓN 4. IDENTIFICACIÓN DE LAS NECESIDADES DE APRENDIZAJE (CONCEPTOS Y HERRAMIENTAS)

Análisis de los trabajos presentados en la segunda sesión para identificar los principales aspectos relacionados con la planificación, diseño y desarrollo de la investigación. Determinación de los conceptos y herramientas para el análisis de datos que se necesitan conocer para desarrollar el trabajo (por ejemplo tipos de encuestas, tipos de preguntas, tipos de muestreo, población, etc.)
- Protagonistas principales: los alumnos aunque el profesor proporcionará información relevante al respecto
- Competencias: **E1, E2, G5**

Duración: 5 h.

SESIÓN 5. OBTENCIÓN DE INFORMACIÓN

Identificación de fuentes secundarias a consultar, valoración de su calidad, síntesis de la información obtenida y desarrollo de un cuestionario relacionado con el tema elegido. Para ello, será necesario identificar la población bajo estudio, el tipo de muestreo, etc. Obtención de información mediante la realización del cuestionario elaborado a una muestra de la población seleccionada.
- Protagonistas principales: los alumnos.
- Competencias: **E1, E2, G1, G2, G3, G5**

Duración: 8 h.

SESIÓN 6. PRESENTACIÓN DE RESULTADOS

Estructuración de la información recopilada, desarrollo de un informe final y presentación en el aula.
- Protagonistas principales: los alumnos.
- Competencias: **E1, E2, G2, G3, G4, G5**

Duración: 3 h.

Competencias a adquirir por parte de los alumnos

Específicas

- **E1:** Compresión de la importancia de los métodos estadísticos de análisis y recogida de datos en el área de Investigación de Mercados
- **E2:** Mejor asimilación de los conceptos estadísticos propuestos en la asignatura:
  - Medidas de tendencia central y dispersión
  - Técnicas de investigación cuantitativa y cualitativa
  - Tipos de muestreo y conceptos relacionados
  - Tipos de encuestas – tipos de preguntas

Generales

- **G1:** Habilidad para analizar y buscar información proveniente de fuentes diversas aplicables al ámbito de estudio
- **G2:** Capacidad de análisis y síntesis
- **G3:** Capacidad para la resolución de problemas y la toma de decisiones
- **G4:** Análisis y presentación de resultados de forma profesional – uso de software estadístico
- **G5:** Trabajo en equipo

REFERENCIAS

Aprendizaje basado en problemas. Guías rápidas sobre nuevas metodologías, 2008, Servicio de Innovación Educativa de la Universidad Politécnica de Madrid.

Barrows, H. S. A taxonomy of problem-based learning methods. *Medical Education*, 1986, 20, 481–486.

Restrepo Gómez, B. (2005). Aprendizaje Basado en Problemas (ABP): Una innovación didáctica para la enseñanza universitaria. *Educación y Educadores*, 2005, 8, 9-19.

Software libre y representaciones gráficas – ggplot2

Isabel Serrano Czaia, Mónica Ortega Moreno

Departamento de Economía, Universidad de Huelva

Contacto: iserrano@uhu.es

Resumen

El uso y manejo de herramientas gráficas es fundamental para que el alumno adquiera una perspectiva más completa y creativa de las posibilidades del análisis de datos estadísticos. Por otro lado, la programación basada en software libre está cada vez más presente en las investigaciones con datos reales en cualquier campo de interés. En esta línea, el paquete de R ggplot2 presenta altas prestaciones gráficas, dado que no se limita a representaciones específicas de forma cerrada, sino que permite elaborar un gráfico a partir de un conjunto de componentes independientes, capas, que se pueden combinar de diferentes formas para construir distintos tipos de visualización gráfica, con una gran flexibilidad para ampliar o transformar el contenido. El hecho de que se guarde el gráfico como objeto permite su manipulación. A pesar de las ventajas, también presenta algunas limitaciones; por ejemplo, no se pueden realizar gráficos tridimensionales o interactivos.

Palabras claves: Software libre, representación gráfica

Software libre y representaciones gráficas ggplot2

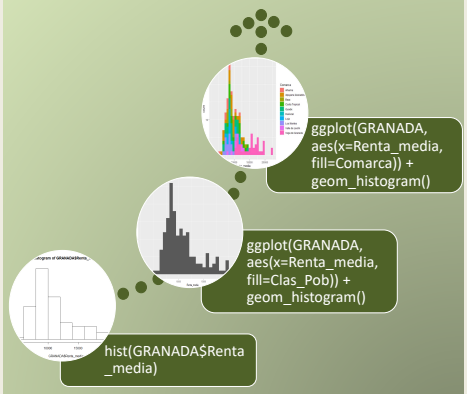
RESUMEN

El uso y manejo de herramientas gráficas es fundamental para que el alumno adquiera una perspectiva más completa y creativa de las posibilidades del análisis de datos estadísticos. Por otro lado, la programación basada en software libre está cada vez más presente en las investigaciones con datos reales en cualquier campo de interés.

En esta línea, el paquete de R ggplot2 presenta altas prestaciones gráficas, dado que no se limita a representaciones específicas de forma cerrada, sino que permite elaborar un gráfico a partir de un conjunto de componentes independientes, capas, que se pueden combinar de diferentes formas para construir distintos tipos de visualización gráfica, con una gran flexibilidad para ampliar o transformar el contenido. El hecho de que se guarde el gráfico como objeto permite su manipulación.

A pesar de las ventajas, también presenta algunas limitaciones; por ejemplo, no se pueden realizar gráficos tridimensionales o interactivos.

HISTOGRAMA



La principal característica de la estructura ggplot2 es que se especifican de forma independiente los bloques de construcción y se combinan para crear prácticamente cualquier tipo de visualización gráfica que se desee.

Los bloques de construcción de un gráfico incluyen:

- Datos: dataframe
- Configuración (**aes**): ejes X e Y, color, tamaño, forma y trazo.
- Capas (**geom**): capas que se pueden ir agregando una encima de otra.
- Etiquetas (**labs**): ayudan a una mejor visualización. Las más comunes: título principal, títulos de ejes y leyendas.
- Tema (**theme**): ajusta el tamaño de etiquetas y leyenda.
- Aspecto (**facet**): permite ver un gráfico separado por grupos.

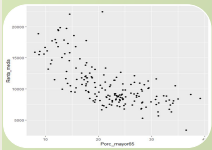
Algunos gráficos:

- ***Nube de puntos***: geom_point
- ***Curvas de ajuste***: geom_smooth
- ***Histograma***: geom_histogram
- ***Diagramas de cajas***: geom_boxplot
- ***Gráficos de barras***: geom_bar
- …

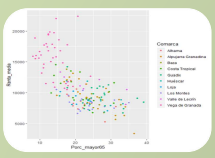
También existe la posibilidad de dibujar gráficos por subgrupos (facet_grid) o unir varios gráficos en la misma figura.

Limitación: no se pueden realizar gráficos tridimensionales o interactivos.

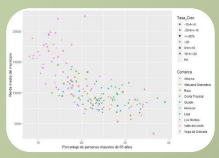
GRÁFICOS



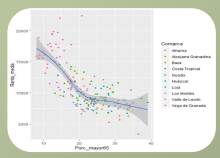
NUBE DE PUNTOS
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media)) + geom_point()



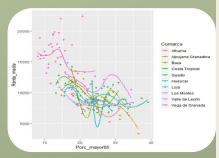
NUBE DE PUNTOS POR COMARCA
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media, color=Comarca)) + geom_point()



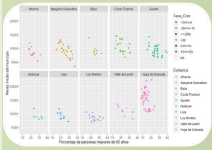
NUBE DE PUNTOS CON GRUPOS Y ETIQUETAS
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media, color=Comarca, shape=Tasa_Crec)) + geom_point()+labs(x="et1", y="et2")



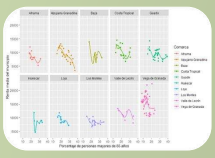
NUBE DE PUNTOS CON CAPAS
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media)) + geom_point(aes(color=Comarca)) + geom_smooth()



NUBE DE PUNTOS CON CAPAS
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media, color=Comarca)) + geom_point() + geom_smooth(se=FALSE)



NUBE DE PUNTOS POR GRUPOS
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media, color=Comarca)) + geom_point() + facet_wrap(~ Comarca, ncol=5)



NUBE DE PUNTOS POR GRUPOS CON CAPAS
ggplot(GRANADA, aes(x=Porc_mayor65, y=Renta_media, color=Comarca)) + geom_point()+geom_smooth(se=FALSE) + facet_wrap(~ Comarca, ncol=5)

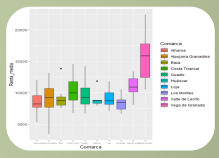


DIAGRAMA DE CAJAS POR COMARCA
ggplot(GRANADA,aes(Comarca,Renta_media)) + geom_boxplot(aes(fill=Comarca)) + theme(axis.text.x = element_text(size = 4))

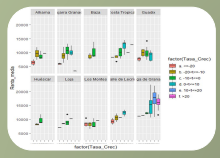


DIAGRAMA DE CAJAS POR COMARCA Y GRUPOS ggplot(GRANADA, aes(factor(Tasa_Crec),Renta_media)) + geom_boxplot((aes(fill=factor(Tasa_Crec)))+ facet_grid(.~Comarca) + facet_wrap(~ Comarca, ncol=5) + theme(axis.text.x = element_text(size = 2))

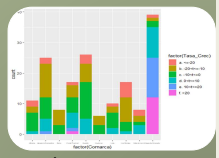


GRÁFICO DE BARRAS APILADO
ggplot(GRANADA, aes(x=factor(Comarca),fill=factor(Tasa_Crec))) + geom_bar() + theme(axis.text.x = element_text(size=4))

BIBLIOGRAFÍA

- Chang, Winston. 2012. R Graphics Cookbook. Sebastopol, CA: O'Reilly Media.
- Wickham, Hadley. 2009. Ggplot2: elegant graphics for data analysis. New York: Springer.

Isabel Serrano Czaia (iserrano@uhu.es)
Mónica Ortega Moreno (ortegamo@uhu.es)
Dpto. Economía.
Universidad de Huelva

Enseñanza de la estadística en un entorno virtual para personas invidentes

Gabriel Sierra Valdivieso, Beatriz Cobo Rodríguez[1], Ana María Lara Porras[1]

[1]Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: gabrielsierra@correo.ugr.es

Resumen

El objetivo de este trabajo es el diseño de un entorno virtual de Estadística Accesible y Universal donde no existan barreras para aprender Estadística, utilizando los programas SPSS, R, R Commander y RStudio. A través de un proyecto de innovación docente, estamos desarrollando dicho entorno web (http://wdb.ugr.es/bioestad), a través del cual impartimos docencia práctica en diversos grados de la UGR. Respecto a la Accesibilidad, nuestro objetivo es la discapacidad visual, para lo cual el paquete estadístico R es accesible para cualquier persona independientemente del sistema operativo, la elección del lector de pantalla, pantalla braille. La capacidad de R ha abierto las puertas para la accesibilidad de este colectivo a los análisis estadísticos. La extensibilidad de R hace esto posible a través de añadirle funcionalidad que está disponible en un paquete adicional llamado RBrailler. La finalidad de las funciones de este paquete es que la información gráfica esté disponible en forma de texto. Para diseñar esta Web Accesible nos hemos puesto en contacto con la ONCE de Granada para que nos apoyen y ayuden en la realización de este proyecto.

Palabras claves: Accesibilidad, invidente, estadística, R

UNIVERSIDAD DE GRANADA

Gabriel Sierra Valdivieso; gabrielsierra@correo.ugr.es
Beatriz Cobo Rodríguez; beacr@ugr.es
Ana María Lara Porras; alara@ugr.es

2019

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

ENSEÑANZA DE LA ESTADÍSTICA EN UN ENTORNO VIRTUAL PARA PERSONAS INVIDENTES

Capaz de ser capaces

La accesibilidad se define como una condición necesaria para la participación de personas con diversidad funcional en tareas laborales y cotidianas. En este aspecto las tecnologías de la información deben tener un nuevo rol desde su inicio para una completa integración de las TICs en la accesibilidad global

ENTORNO WEB ACESIBLE
http://wdb.ugr.es/~bioestad

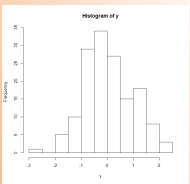
Pocos entornos web toman conciencia del problema de la discapacidad visual. A través del proyecto de innovación docente se intenta plantear nuevas formas de autoprendizaje de la estadística accesibles para invidentes

BrailleR

Acerca el uso de R a las personas con discapacidad visual mediante la descripción en forma de texto de los gráficos y de las salidas de consola

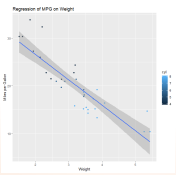
Descripción de gráficos mediante BrailleR

R Base



This is a histogram, with the title: Histogram of y
"y" is marked on the x-axis.
Tick marks for the x-axis are at: -3, -2, -1, 0, 1, and 2
There are a total of 150 elements for this variable.
Tick marks for the y-axis are at: 0, 5, 10, 15, 20, 25, 30, and 35
It has 11 bins with equal widths, starting at -3 and ending at 2.5 .
The mids and counts for the bins are:
mid = -2.75 count = 1
mid = -2.25 count = 0
mid = -1.75 count = 5
mid = -1.25 count = 10
mid = -0.75 count = 29
mid = -0.25 count = 34
mid = 0.25 count = 27
mid = 0.75 count = 15
mid = 1.25 count = 18
mid = 1.75 count = 8
mid = 2.25 count = 3

ggplot2



This chart has title 'Regression of MPG on Weight'.
It has x-axis 'Weight' with labels 2, 3, 4 and 5.
It has y-axis 'Miles per Gallon' with labels 10, 20 and 30.
There is a legend indicating that colour is used to represent cyl, ranging from 4 represented by colour dark purplish blue to 8 represented by colour brilliant blue.
It has 2 layers.
Layer 1 is a set of 32 points.
Layer 2 is a smoothed curve using method 'lm' with confidence intervals.

Conclusiones

- Las características de lenguaje interpretado de R permiten una gran adaptabilidad para el uso accesible y se espera que pueda evolucionar aún más.
- La gramática de las gráficas plantea un amplio abanico de interpretabilidad para los invidentes.
- Mediante la utilización de BrailleR podemos ampliar la accesibilidad de la web a las personas con discapacidad visual.
- El apoyo del software y el contenido de la web permitirá hacer más accesible la docencia teórica y práctica de la estadística.

Idoneidad del software empleado en prácticas de Investigación Operativa

Úrsula Torres Parejo

Departamento de Estadística e Investigación Operativa, Universidad de Granada

Contacto: ursula@ugr.es

Resumen

Este estudio se llevó a cabo durante el curso académico 16/17 entre estudiantes de diversos Grados de la Universidad de Cádiz y consistió en la evaluación del software utilizado en prácticas de asignaturas de investigación operativa. Las aplicaciones evaluadas fueron: Lingo (software privativo para la resolución de programas lineales, no lineales y enteros), LP-Solve (software libre para programación lineal), PHP-Simplex (herramienta online, libre y gratuita, para programación lineal) y el complemento Solver de Excel (software privativo). El objetivo es que los mismos estudiantes evaluaran dichos programas y los clasificaran según sus características y funcionalidad, fomentando el trabajo autónomo y el pensamiento crítico por parte del alumnado. Las conclusiones más destacables fueron: la conveniencia de trabajar con varios programas simultáneamente y la predilección de los estudiantes por el PHP-Simplex, a pesar de ser el Excel el que más motivación les produjo por pensar que sería el más útil para su futuro profesional.

Palabras claves: optimización, software, investigación operativa, prácticas

Introducción

Son numerosos los títulos impartidos en las distintas Universidades que cuentan con una o más asignaturas de investigación operativa básica, sin que exista un estudio sobre la idoneidad del software utilizado en prácticas.

Partiendo de la idea de que lo más beneficioso para el alumnado es darle a conocer el mayor número posible de aplicaciones informáticas para un mismo o para distintos propósitos, se pretende hacer un análisis comparativo de software para cada uno de los principales campos que abarca la investigación operativa y más concretamente, la optimización.

Este estudio se ha llevado a cabo durante el curso académico 2016/17 entre los estudiantes de diversos Grados de la Universidad de Cádiz, en concreto, del

Grado en Ingeniería en Diseño Industrial y Desarrollo del Producto, asignatura Desarrollo Óptimo del Producto y Diseño de Experimentos de 4° curso, en el Grado en Ingeniería Química, asignatura Estadística y Optimización de 1° curso y en el Grado en Administración y Dirección de Empresas, asignatura Métodos Cuantitativos para la Toma de Decisiones Empresariales de 2° curso.

Hasta el momento, los principales programas que se habían utilizado en dichas asignaturas son:

- Lingo (1): Software privativo para la resolución de programas lineales, no lineales y enteros.
- LP-Solve (2): Software libre utilizado en programación lineal, con un lenguaje similar al de Lingo.
- PHP-Simplex (3): Es una herramienta online, cuyo uso es libre y gratuito, para resolver problemas de programación lineal.
- Scientific Workplace (4): Es un programa propietario para la resolución de operaciones en el ámbito del álgebra y el cálculo. Se utiliza principalmente en programación no lineal, para aspectos muy concretos como el cálculo de operaciones matriciales.
- Excel (5): Software privativo que permite resolver problemas de optimización mediante el complemento solver.

Lo que se persigue, es que los mismos estudiantes evalúen dichos programas y los clasifiquen según sus características y funcionalidad, fomentando el trabajo autónomo y el pensamiento crítico por parte del alumnado.

Desarrollo del trabajo

El proyecto se aplica en las clases prácticas de las asignaturas de Investigación Operativa, más concretamente de optimización, donde se dota al alumno/a del mayor número de herramientas a su alcance para la resolución de los supuestos prácticos que se le plantean, de modo que sea el alumno/a quien trabaje de forma autónoma, siendo crítico/a a la hora de utilizar una u otra herramienta para un determinado problema, investigando y descubriendo las limitaciones de cada programa y justificando la utilización de uno u otro en los informes finales de prácticas, comentando sus ventajas y desventajas.

Resultados

Terminadas las prácticas en las distintas asignaturas, se pidió a los estudiantes que emitieran su grado de acuerdo con distintas afirmaciones, quedando de esta forma reflejada la valoración para cada programa. Como el Scientific Workplace es más un software de ayuda que un programa concreto para optimización, sólo se evaluaron los siguientes softwares: PHPSimplex, LP-Solve, Lingo y Excel (complemento solver).

A continuación se resumen los resultados recogidos para cada uno de los aspectos evaluados:

1. Calidad del entorno visual, la velocidad de cálculo y la presentación de la información. La mayoría de los encuestados manifiestan estar de acuerdo con la adecuación de estos aspectos para el programa PHP-Simplex, que es el mejor valorado, seguido de Lingo. LP-Solve y Excel no obtienen tan buena valoración en este aspecto.

2. Promoción de la iniciativa y el proceso de aprendizaje autónomo. De nuevo para este aspecto, el PHP-Simplex es la aplicación mejor valorada, seguida en este caso de Excel , obteniendo los programas Lingo y LP-Solve peores valoraciones.

3. Fomento del desarrollo de un perfil reflexivo, crítico e innovador. Claramente el PHP-Simplex vuelve a ser el mejor valorado para este aspecto, seguido de Excel.

4. Posibilidad de continuación con el uso del programa. En este caso, la mayoría de los encuestados opina que seguirá utilizando el Excel, no teniéndolo tan claro con el resto de aplicaciones evaluadas.

5. Motivación con el programa y creencia en su utilidad en el futuro profesional. Igual que en el apartado anterior, el Excel vuelve a ser el programa mejor valorado, en este caso seguido del PHP-Simplex y no estando tan claro para los softwares Lingo y LP-Sove.

6. Profundización en su conocimiento. La mayoría de los encuestados manifiestan que les hubiera gustado profundizar más en el conocimiento de Excel y Lingo.

7. Utilización del programa según el tema tratado en la asignatura. El aspecto que se recoge en esta valoración es de vital importancia, ya que nos permite establecer qué programa prefieren los estudiantes para distintos temas de la asignatura. Aquí es donde realmente se refleja el conocimiento que tienen sobre el software, si son capaces de ver en qué aspectos es más ventajoso utilizar una u otra herramienta. Así vemos que la mayoría utilizaría el PHP-Simplex para problemas relacionados con la aplicación del método Simplex, el LP-Solve para el Análisis de sensibilidad, y el Lingo para problemas donde haya que emplear el método de las ponderaciones, así como para programación no lineal. Aquí lo más destacable es que pocos utilizarían el Excel como primera opción para ninguno de estos temas, siendo el mejor valorado en los aspectos 4 y 5.

8. Conveniencia de trabajar con varios programas simultáneamente. La gran mayoría de los estudiantes considera que es conveniente trabajar en clase con varios programas de forma simultánea.

Conclusiones

Realizando un análisis crítico de los resultados obtenidos con la encuesta resaltamos las siguientes conclusiones:

En primer lugar queda de manifiesto la conveniencia de trabajar con distintos softwares simultáneamente, así es más fácil ver las carencias y ventajas de unos y otros.

La predilección de los estudiantes apunta hacia la aplicación PHP-Simplex, esto puede deberse a que es la que presenta los resultados de forma más clara y detallada, mostrando los métodos aplicados paso a paso.

Sin embargo, esta aplicación está restringida a aspectos básicos de la optimización, motivo por el cual los estudiantes no creen que sea útil para su futuro profesional, sintiendo mayor motivación en clase cuando usaron el Excel, a pesar de resultarles menos intuitivo y de que no lo consideren el más apropiado para ninguno de los temas tratados en la asignatura.

De forma global, los programas Lingo y LP-Solve obtienen valoraciones similares, si bien el LP-Solve no resuelve problemas de programación no lineal ni aquellos en los que se aplica el método de las ponderaciones, por lo que utilizarían Lingo en esos casos. Curiosamente, hay una mayoría de estudiantes que utilizaría LP-Solve en análisis de sensibilidad, a pesar de que el programa Lingo ofrece más claramente los resultados sobre este análisis, este hecho se debe a que siempre que sea posible, los estudiantes prefieren el uso del software libre sobre el uso del software privativo.

Agradecimientos. A la convocatoria de proyectos para la innovación y mejora docente del curso 16/17 de la Universidad de Cádiz.

Referencias

[1] Lindo Systems INC. (1981) *Descarga e Información del Programa Lingo*. https://www.lindo.com/index.php/products/lingo-and-optimization-modeling.

[2] SourceForge LP-Solve (2016) *Descarga e Información del Programa LP-Solve*. https://sourceforge.net/projects/lpsolve/.

[3] D. Izquierdo-Granja y J.J. Ruiz-Ruiz (2006) *PHP-Simplex. Optimizando recursos con programación lineal*. http://www.phpsimplex.com/.

[4] Mackichan Software (1981) *Descarga del Software Scientific Workplace*. https://www.mackichan.com/.

[5] Microsoft Office (1985) *Software de hojas de cálculo: Microsoft Excel*. https://products.office.com/es-es/excel.

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa (Granada, 4-5 abril 2019)

IDONEIDAD DEL SOFTWARE EMPLEADO EN PRÁCTICAS DE INVESTIGACIÓN OPERATIVA.

UNIVERSIDAD DE GRANADA

Úrsula Torres Parejo
Departamento de Estadística e Investigación Operativa
Universidad de Granada

Introducción

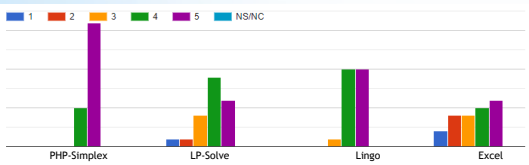
Partiendo de la idea de que lo más beneficioso para el alumnado es darle a conocer el mayor número posible de aplicaciones informáticas para un mismo o para distintos propósitos, se pretende hacer un **análisis comparativo de software** para cada uno de los principales campos que abarca la investigación operativa y más concretamente, la optimización.
Este estudio se ha llevado a cabo durante el curso académico 2016/17 entre los estudiantes de diversos Grados de la Universidad de Cádiz.
Lo que se persigue, es que los mismos estudiantes evalúen dichos programas y los clasifiquen según sus características y funcionalidad, fomentando el trabajo autónomo y el pensamiento crítico por parte del alumnado.
Los principales programas utilizados, objeto de evaluación son:
- **Lingo:** Software privativo para la resolución de programas lineales, no-lineales y enteros.
- **LP-Solve[1]:** Software libre utilizado en programación lineal, con un lenguaje similar al de Lingo.
- **PHP-Simplex[2]:** Es una herramienta online, cuyo uso es libre y gratuito, para resolver problemas de programación lineal.
- **Excel:** Software privativo que permite resolver problemas de optimización mediante el complemento **"solver"**.

Desarrollo del proyecto

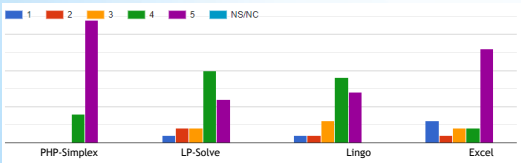
El proyecto se aplica en las clases prácticas de las asignaturas de Investigación Operativa, más concretamente de optimización, donde se dota al alumno/a del mayor número de herramientas a su alcance para la resolución de los supuestos prácticos que se le plantean, de modo que sea el alumno/a quien trabaje de forma autónoma, siendo crítico/a a la hora de utilizar una u otra herramienta para un determinado problema, investigando y descubriendo las limitaciones de cada programa y justificando la utilización de uno u otro en los informes finales de prácticas, comentando sus ventajas y desventajas.

Resultados[3]

1. **Calidad del entorno visual, la velocidad de cálculo y la presentación de la información.** La mayoría de los encuestados manifiestan estar de acuerdo con la adecuación de estos aspectos para el programa PHP-Simplex, que es el mejor valorado, seguido de Lingo. LP-Solve y Excel no obtienen tan buena valoración en este aspecto.



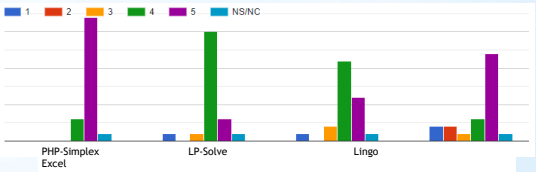
2. **Promoción de la iniciativa y el proceso de aprendizaje autónomo.** De nuevo para este aspecto, el PHP-Simplex es la aplicación mejor valorada, seguida en este caso de Excel , obteniendo los programas Lingo y LP-Solve peores valoraciones.



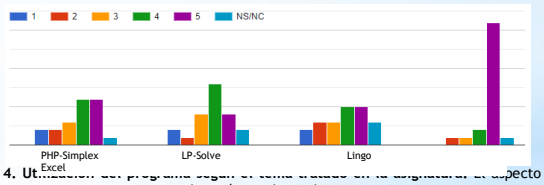
1. http://lpsolve.sourceforge.net/
2. http://www.phpsimplex.com/
3. https://docs.google.com/forms/d/e/1FAIpQLSe1a9ZYwgYYfmS7HpMIgFI8quDPylkz0Jde5KtTaP_OjivT5Q/viewform

Resultados

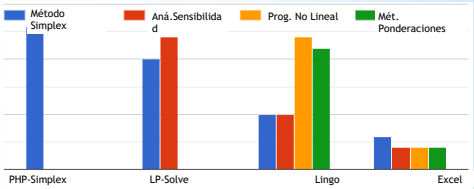
3. **Fomento del desarrollo de un perfil reflexivo, crítico e innovador.** Claramente el PHP-Simplex vuelve a ser el mejor valorado para este aspecto, seguido de Excel.



3. **Motivación con el programa y creencia en su utilidad en el futuro profesional.** En este caso, la mayoría de los encuestados sintió mayor motivación en clase cuando usaron el Excel porque creen que les será más útil en su futuro profesional.



4. **Ut**[illegible] **aspecto** que se recoge en esta valoración es de vital importancia, ya que nos permite establecer que programa prefieren los estudiantes para distintos temas de la asignatura. Así vemos que la mayoría utilizaría el PHP-Simplex para problemas relacionados con la aplicación del método Simplex, el LP-Solve para el Análisis de sensibilidad, y el Lingo para problemas donde haya que emplear el método de las ponderaciones, así como para programación no lineal. Aquí lo más destacable es que pocos utilizarían el Excel como primera opción para ninguno de estos temas, siendo el mejor valorado en el aspecto anterior.



Conclusiones

La predilección de los estudiantes apunta hacia la aplicación PHP-Simplex, esto puede deberse a que es la que presenta los resultados de forma más clara y detallada, mostrando los métodos aplicados paso a paso. Sin embargo, esta aplicación está restringida a aspectos básicos de la optimización, motivo por el cual los estudiantes no creen que sea útil para su futuro profesional, sintiendo mayor motivación en clase cuando usaron el Excel, a pesar de resultarles menos intuitivo y de que no lo consideren el más apropiado para ninguno de los temas tratados en la asignatura.
De forma global, los programas Lingo y LP-Solve obtienen valoraciones similares, si bien el LP-Solve no resuelve problemas de programación no lineal ni aquellos en los que se aplica el método de las ponderaciones, por lo que utilizarían Lingo en esos casos. Curiosamente, hay una mayoría de estudiantes que utilizaría LP-Solve en análisis de sensibilidad, a pesar de que el programa Lingo ofrece más claramente los resultados sobre este análisis, este hecho se debe a que siempre que sea posible, los estudiantes prefieren el uso del software libre sobre el uso del software privativo.

Estrategias de aprendizaje en las enseñanzas de Estadística y Matemáticas

María Concepción Vega Hernández[1], María Carmen Patino Alonso[1,2], María Purificación Galindo Villardón[1,2]

[1]Departamento de Estadística, Universidad de Salamanca
[2]Instituto de Investigación Biomédica de Salamanca (IBSAL)

Contacto: mvegahdz@usal.es

Resumen

El planteamiento óptimo de la enseñanza de la Estadística parte del conocimiento que los profesores tienen de sus alumnos. El objetivo de esta investigación se centra en identificar las estrategias de aprendizaje predominantes en los estudiantes de los Grados en Estadística y Matemáticas de la Universidad de Salamanca. Para realizar la evaluación de dichas estrategias de aprendizaje utilizadas por los universitarios se empleó la Escala de Estrategias de Aprendizaje ACRA-Abreviada (ACRA-A) (De la Fuente y Justicia, 2003). Los resultados pusieron de manifiesto que la estrategia de aprendizaje más utilizada por los estudiantes que cursaban el Grado en Matemáticas fueron los hábitos de estudio, mientras que los alumnos de Estadística se inclinaron más por las estrategias cognitivas y de control del aprendizaje. Conocer las estrategias de aprendizaje que más aplican los universitarios cuando se enfrentan al aprendizaje permite mejorar la enseñanza y ayudarles en su proceso académico.

Palabras claves: estrategias de aprendizaje, estadística, universitarios, ACRA-A

IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa

Dpto. de Estadística
Universidad de Salamanca

ESTRATEGIAS DE APRENDIZAJE EN LAS ENSEÑANZAS DE ESTADÍSTICA Y MATEMÁTICAS

María Concepción Vega-Hernández[1], María Carmen Patino-Alonso[2], María Purificación Galindo-Villardón[3]

[1] mvegahdz@usal.es , Universidad de Salamanca; [2] carpatino@usal.es, Universidad de Salamanca, IBSAL; [3]pgalindo@usal.es, Universidad de Salamanca, IBSAL

RESUMEN

El planteamiento óptimo de la enseñanza de la Estadística parte del conocimiento que los profesores tienen de sus alumnos. El objetivo de esta investigación se centra en identificar las estrategias de aprendizaje predominantes en los estudiantes de los Grados en Estadística y Matemáticas de la Universidad de Salamanca. Para realizar la evaluación de dichas estrategias de aprendizaje utilizadas por los universitarios se empleó la Escala de Estrategias de Aprendizaje ACRA-Abreviada (ACRA-A) (De la Fuente y Justicia, 2003). Los resultados pusieron de manifiesto que la estrategia de aprendizaje más utilizada por los estudiantes que cursaban el Grado en Matemáticas fueron los hábitos de estudio, mientras que los alumnos de Estadística se inclinaron más por las estrategias cognitivas y de control del aprendizaje. Conocer las estrategias de aprendizaje que más aplican los universitarios cuando se enfrentan al aprendizaje permite mejorar la enseñanza y ayudarles en su proceso académico.

Palabras clave: estrategias de aprendizaje, estadística, universitarios, ACRA-A.

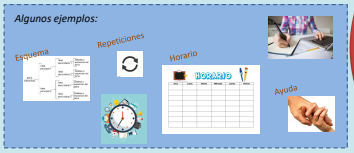
INTRODUCCIÓN

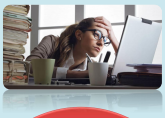
Conocer la forma en que los alumnos capta la información es esencial pues, cada estudiante aprenden de manera única, a pesar de tener la misma edad, las mismas motivaciones, etc.

Ellos cuentan con estrategias y técnicas de aprendizaje; es decir, secuencias de procedimientos o planes orientados hacia la consecución de metas de aprendizaje. Dentro de cada secuencia, los procedimientos específicos se denominan tácticas de aprendizaje. En este caso, las estrategias serían procedimientos de nivel superior que incluirían diferentes tácticas o técnicas de aprendizaje.

Las estrategias de aprendizaje son el conjunto de procedimientos cognitivos y recursos que se ponen en práctica cuando se enfrentan al aprendizaje (Valle, González, Cuevas y Fernández, 1998). Son directa o indirectamente manipulables y tienen carácter intencional o propositivo (Beltran, 1993).

Algunos ejemplos:





OBJETIVO

El objetivo de esta investigación se centra en identificar las estrategias de aprendizaje predominantes en los estudiantes de los Grados en Estadística y Matemáticas.

MÉTODO

Participantes

El estudio se llevó a cabo sobre **estudiantes** de los Grados en Estadística y Matemáticas de una muestra de 254 alumnos de la **Universidad de Salamanca.**

Instrumento

Para realizar la evaluación de las estrategias de aprendizaje utilizadas por los alumnos se empleó la **Escala de Estrategias de Aprendizaje ACRA-Abreviada** (De la Fuente y Justicia, 2003) que es una simplificación de la Escala de Estrategias de Aprendizaje ACRA de Román y Gallego (1994).

Está formada por 44 ítems inspirados en los principios cognitivos del procesamiento de la información. La respuesta de los ítems consiste en una escala Likert de cuatro opciones (A, B, C y D), aunque en este caso se agruparon para dar una respuesta dicotómica (Sí/No).

Presenta una estructura tridimensional:

- La dimensión de las estrategias cognitivas y de control del aprendizaje incluye 25 ítems que se refieren a la selección, organización, resaltado o conocimiento de la funcionalidad de las estrategias (es decir, "Elaboro resúmenes ayudándome de palabras o frases anteriormente subrayadas")
- La dimensión de estrategias de apoyo al aprendizaje contiene 14 ítems motivacionales y afectivos (es decir, "Estudio para ampliar mis conocimientos, para saber más, para ser más experto")
- Y la dimensión de los hábitos de estudio incluye 5 ítems que implican comprensión y hábitos de estudio (es decir, "Intento expresar lo aprendido con mis propias palabras, en vez de repetir literalmente o al pie de la letra lo que dice el libro o profesor").

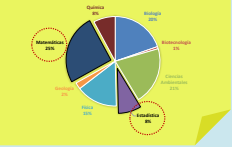
Análisis

Para analizar el uso de estrategias de aprendizaje se utilizó la metodología establecida por Honey y Mumford (1986) basada en los rangos. Este método proporciona un esquema de interpretación denominado *baremo*, basado en las puntuaciones obtenidas y la experiencia de los test. La categorización se realiza por terciles, creando tres grados de interpretación (alto, moderado y bajo). Y, si existiera algún empate se consideró la puntuación media más alta en el estudiante como estrategia de aprendizaje predominante.

RESULTADOS DEL ESTUDIO

Descriptiva

La muestra consta estudiantes pertenecientes al área de Ciencias, donde un 33,5% eran de los Grados en Estadística y Matemáticas.



Baremo

Se clasificó a los alumnos según su preferencia por las estrategias cognitivas y de control del aprendizaje (técnicas metacognitivas), las estrategias de apoyo al aprendizaje (técnicas motivacionales y afectivas) y los hábitos de estudio (rutinas de estudio) cuando uno de estos tipos predomina por encima de los demás. Sin embargo, puede darse el caso de que un estudiante no muestre preferencia por una de las estrategias de aprendizaje mencionadas, denominando a este modo como *Mixto*.

Niveles de preferencia de cada una de las subescalas del cuestionario ACRA-A.

DIMENSIONES DE ESTRATEGIAS DE APRENDIZAJE	Bajo	Moderado	Alto
Estrategias cognitivas y de control del aprendizaje	[< 18] 36,22%	[18 - 21] 31,10%	[> 21] 32,68%
Estrategias de apoyo al aprendizaje	[< 9] 35,83%	[9 - 12] 40,94%	[> 12] 23,23%
Hábitos de estudio	[< 3] 39,37%	[3 - 4] 24,80%	[> 4] 35,83%

ESTRATEGIAS DE APRENDIZAJE PREPONDERANTES

Los resultados pusieron de manifiesto que la estrategia de aprendizaje más utilizada por los estudiantes que cursaban el Grado en Matemáticas fueron los hábitos de estudio, mientras que los alumnos de Estadística se inclinaron más por las estrategias cognitivas y de control del aprendizaje.

ESTRATEGIAS DE APRENDIZAJE		TITULACIONES	
		Estadística	Matemáticas
Estrategias cognitivas y de control del aprendizaje	n	8	19
	% dentro de Estrategia	10,3%	24,4%
	% dentro de Titulación	38,1%	29,7%
Estrategias de apoyo al aprendizaje	n	5	16
	% dentro de Estrategia	10,2%	32,7%
	% dentro de Titulación	23,8%	25,0%
Hábitos de estudio	n	7	23
	% dentro de Estrategia	7,2%	23,7%
	% dentro de Titulación	33,3%	35,9%
Mixto	n	1	6
	% dentro de Estrategia	3,3%	20,0%
	% dentro de Titulación	4,8%	9,4%
Total	n	21	64
	% dentro de Estrategia	8,3%	25,2%
	% dentro de Titulación	100,0%	100,0%

Preferencia de estrategias de aprendizaje

REFERENCIAS

Beltrán, J. (1993). *Procesos, estrategias y técnicas de aprendizaje*. Madrid: SÍntesis.

De la Fuente, J., & Justicia, F. (2003). Escala de estrategias de aprendizaje ACRA-Abreviada para alumnos universitarios. *Revista Electrónica de Investigación Psicoeducativa y Psicopedagógica*, 2(1), 139-158.

Honey, P., & Mumford, A. (1986). *Using your learning styles*. Berkshire: Ardingly House.

Román, J., & Gallego, S. (1994). *ACRA Escalas de estrategias de aprendizaje*. Madrid: TEA Ediciones S. A.

Valle, A., González, R., Cuevas, L. M., & Fernández, A. P. (1998). Las estrategias de aprendizaje: características básicas y su relevancia en el contexto escolar. *Revista de Psicodidáctica*, (6), 53-68. Recuperado a partir de http://www.redalyc.org/html/175/17514484006/

CONCLUSIONES

Conocer las estrategias de aprendizaje que más aplican los universitarios cuando se enfrentan al aprendizaje permite mejorar la enseñanza y ayudarles en su proceso académico.

Lista de participantes

ACAL GONZALEZ	CHRISTIAN JOSÉ	chracal@ugr.es
AGUILERA DEL PINO	ANA MARÍA	aaguiler@ugr.es
AGUILERA MORILLO	M. CARMEN	mariacarmen.aguilera@uc3m.es
ALONSO MORALES	FRANCISCO JAVIER	falonso@ugr.es
ÁLVAREZ ARROYO	ROCÍO	rocioaarroyo@ugr.es
ARTÉS RODRÍGUEZ	EVA MARIA	eartes@ual.es
BARRANCO CHAMORRO	INMACULADA	chamorro@us.es
BOJ DEL VAL	EVA	evaboj@ub.edu
CALDUCH LOSA	ÁNGELES	mcalduch@eio.upv.es
CANO CHICA	PEDRO ANTONIO	pacano@ujaen.es
CAÑADAS DE LA FUENTE	GUSTAVO R.	grcanadas@ugr.es
CARMONA PONTAQUE	FRANCESC	fcarmona@ub.edu
CASTILLO GUTIÉRREZ	SONIA	socasti@ujaen.es
COBO RODRÍGUEZ	BEATRIZ	beacr@ugr.es
CUEVA LÓPEZ	VALENTINA	vcueva@ujaen.es
DAUNIS I ESTADELLA	PEPUS	pepus@imae.udg.edu
DAZA ARBOLÍ	MIGUEL ÁNGEL	mdaza@est-econ.uc3m.es
DUEÑAS ALCALDE	MARIA ELENA	tick@correo.ugr.es
ESCABIAS MACHUCA	MANUEL	escabias@ugr.es
FERRI GARCÍA	RAMÓN	rferri@ugr.es
GÁMIZ PÉREZ	MARÍA LUZ	mgamiz@ugr.es
GARCÍA GARRIDO	IRENE	iggarrid@ujaen.es
GARCÍA PÉREZ	LOURDES	lourdesgarcia_18@hotmail.com
GARCÍA-LÍGERO RAMÍREZ	MARÍA JESÚS	mjgarcia@ugr.es
GARCÍA SILVENTE	MIGUEL	m.garcia-silvente@decsai.ugr.es
GEA SERRANO	MARÍA MAGDALENA	mmgea@ugr.es
GOICOA MANGADO	TOMÁS	tomas.goicoa@unavarra.es
GRIMA	PERE	pere.grima@upc.edu
HERNÁNDEZ MARTÍN	ZENAIDA	zenaida.hernandez@unirioja.es
HUETE MORALES	MARÍA DOLORES	mdhuete@ugr.es
KESSLER	MATHIEU	mathieu.kessler@upct.es
LARA PORRAS	ANA MARÍA	alara@ugr.es
LARIA	JUAN CARLOS	jlaria@est-econ.uc3m.es
LLIN PÉREZ	MARÍA AURORA	aurorallinp@gmail.com
LÓPEZ MONTOYA	ANTONIO JESÚS	amontoya@ujaen.es
MADRID GARCÍA	ANA ESTHER	anaesther@ugr.es
MALDONADO JURADO	JUAN ANTONIO	jamaldo@ugr.es
MARÍN JIMÉNEZ	ANA EUGENÍA	anamarin@ugr.es
MARTÍNEZ ÁLVAREZ	FERNANDO	falvarez@ugr.es
MARTÍNEZ ÁLVAREZ	Mª CARMEN	malvarez@ugr.es

MARTÍNEZ PUERTAS	HELENA	hmartinez@ual.es
MARTOS GÁLVEZ	EMILIA ISABEL	emilia.martos@uah.es
MÉNDEZ CIVIETA	ÁLVARO	almendez@est-econ.uc3m.es
MIRANDA HUAYNALAYA	FELÍCITA DORIS	fdmirandah@correo.ugr.es
MOLINA MUÑOZ	DAVID	dmolinam@ugr.es
NAVARRETE ÁLVAREZ	ESTEBAN	estebang@ugr.es
OCAÑA LARA	FRANCISCO A.	focana@ugr.es
ORTEGA MORENO	MÓNICA	ortegamo@uhu.es
QUESADA RUBIO	JOSÉ MANUEL	quesada@ugr.es
RAMOS ÁBALOS	EVA MARÍA	ramosa@ugr.es
RAYA MIRANDA	ROCÍO	rraya@ugr.es
RIVERA GALICIA	LUIS F.	luisf.rivera@uah.es
ROLDÁN LÓPEZ DE HIERRO	ANTONIO FRANCISCO	aroldan@ugr.es
ROMÁN ROMÁN	PATRICIA	proman@ugr.es
RUEDA GARCÍA	MARÍA DEL MAR	mrueda@ugr.es
RUIZ CASTRO	JUAN ELOY	jeloy@ugr.es
SAN MARTÍN	MONTSERRAT	momartin@ugr.es
SÁNCHEZ BORREGO	ISMAEL RAMÓN	ismasb@ugr.es
SEGOVIA GARCÍA	MARI CARMEN	msegovia@ugr.es
SERRANO CZAIA	ISABEL	iserrano@uhu.es
SIERRA VALDIVIESO	GABRIEL	gabriel.s.valdivieso@gmail.com
TORRES PAREJO	ÚRSULA	ursula@ugr.es
TORRES RUÍZ	FRANCISCO DE ASÍS	fdeasis@ugr.es
VALDERRAMA BONNET	MARIANO JOSÉ	valderra@ugr.es
VALENZUELA RUÍZ	SILVIA MARÍA	svalenzuela@ugr.es
VEGA HERNÁNDEZ	MARÍA CONCEPCIÓN	mvegahdz@usal.es

UNIVERSIDAD DE GRANADA

SEIO

IEMATH-GR
Instituto de Matemáticas de la Universidad de Granada

Facultad de Ciencias
Universidad de Granada

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA