

Recursos de acceso abierto como apoyo en la conceptualización de prueba de normalidad con estudiantes universitarios

José Armando Albert Huerta, Tecnológico de Monterrey, México, albert@tec.mx

Blanca Rosa Ruiz Hernández, Tecnológico de Monterrey, México, bruiz@tec.mx

Resumen

Una de las prácticas científicas y de ingeniería más sobresalientes, hoy día, es la disposición de recursos y repositorios de acceso abierto. Un caso ejemplar es el uso de R, un software estadístico con un amplio repositorio de aplicaciones en diversas áreas. Esto viene a cambiar no sólo la práctica científica y profesional, sino el modo de aprender desde la formación de estudiantes de ingeniería. Para ilustrar esto, se presenta un caso relativo a una de las condiciones más importantes para el uso de la estadística inferencial: la prueba de normalidad y de cómo los estudiantes, con el uso de R, pudieron ir más allá de la habitual suposición de normalidad.

Palabras clave: recursos de acceso abierto, estadística

Key words: open access resources, statistics

Abstract

One of the most outstanding scientific and engineering practices today is the provision of open access resources and repositories. An exemplary case is the use of R, a statistical software with a large repository of applications in various areas. This comes to change not only the scientific and professional practice, but the way of learning from the training of engineering students. To illustrate this, we present a case related to one of the most important conditions for the use of inferential statistics: the normality test and how students, with the use of R, could go beyond the usual assumption of normality.

1. Introducción

En una sociedad en constante cambio como la que vivimos y basada en el conocimiento, el acceso abierto a recursos a través de Internet da a los estudiantes y profesores nuevas oportunidades y retos para su educación (Ramírez, 2015). Un recurso notable de acceso libre que está cambiando la forma de aprender y desarrollar pensamiento estadístico a nivel universitario y profesional es R (<https://www.r-project.org/>). Se trata un software de acceso libre con un repositorio de 12,774 paquetes que le permiten a R tener herramientas estadísticas y gráficas para resolver un gran número de problemas de muy diversos contextos

de la ciencia o ingeniería. Y bajo ciertas condiciones, en nuestro caso, de Aprendizaje Invertido (Berman y Sams, 2012), es posible desarrollar pensamiento estadístico en estudiantes universitarios más allá de lo tradicionalmente esperado. El Aprendizaje Invertido resulta especialmente conveniente por su gran potencial para abordar contenidos disciplinares complejos, como es el caso de la prueba de normalidad. La condición de normalidad es muy frecuente para la aplicación de técnicas estadísticas y poco o nada abordada en un curso básico universitario de estadística. Nuestra propuesta da evidencia de que es posible su aprendizaje, con ayuda de recursos como R de acceso

abierto, bajo la perspectiva de aprendizaje invertido.

2. Desarrollo

2.1 Preliminares

Diversos estudios a nivel mundial reconocen que el *conocimiento* es actualmente un factor determinante de la producción, el desarrollo económico y el crecimiento social de los países. En esta sociedad del conocimiento los Recursos Educativos Abiertos (REA) son fundamentales para dar acceso al conocimiento a cada vez más personas. Entre esos recursos está el software y materiales que dan soporte al acceso de este conocimiento (Burgos, 2010). Tal es el caso de R y su gran diversidad de paquetes de aplicación de análisis cuantitativo en muy diversas áreas del conocimiento. En una sociedad del conocimiento que demanda una formación continua y desarrollo de competencias, este tipo de REA no solo da oportunidad al profesional de contar con herramientas cada vez más eficientes para el análisis de la información cuantitativa, sino que incluso durante la formación universitaria, les permite a los estudiantes ir más allá de lo propuesto por los planes y programas de estudio. García-Peñalvo, García de Figuerola y Merlo (2010) presentan al conocimiento abierto como la suma de cuatro dimensiones: software, contenido, ciencia e innovación. Tal es el caso de R que asocia software, materiales de acceso abierto para su aprendizaje tales como videos, manuales, cursos, sobre áreas específicas de contenido, así como también técnicas actualizadas de la Estadística para la aplicación y soluciones innovadoras de problemas de contexto real tanto en la etapa universitaria como laboral.

R es un ambiente computacional, de acceso abierto, para el análisis estadístico y gráfico desarrollado por R. Gentleman y R. Ihaka de la Universidad de Auckland, Nueva Zelanda, en la década de los 90's a partir de S, un lenguaje de programación desarrollado por los Laboratorios Bell en 1976 y administrado por *The R Foundation for Statistical Computing*, con sede en Austria (Ripley, 2001). R está extendido por 48 países y con el apoyo de más de 150 universidades. Cuenta toda clase de recursos disponibles para su uso y aprendizaje, así como una revista propia *The R Journal* (<https://journal.r-project.org/>). Es posible tener acceso desde las plataformas Windows, Mac y Linux (<https://www.r-project.org/>). R está siendo aún más popu-

lar que *Python*, *Minitab* y *SAS*, y es superado solamente, por el momento, en citas de artículos de investigación, por *SPSS* (que no es de acceso abierto).

2.2 Descripción de la actividad

Aunque el uso de las hojas electrónicas como Excel es muy recomendable y hasta necesario, se les ofreció la posibilidad a estudiantes de ingeniería y ciencias del Tecnológico de Monterrey, Campus Monterrey, del tercer semestre, del curso de probabilidad y Estadística honors, enero-mayo de 2018, usar R en lugar de Excel, o como una ampliación de sus recursos tecnológicos y de conocimiento.

Su preferencia de usar R en lugar de Excel se presentó en distintos momentos, particularmente para graficar diagramas de caja y bigote o hacer una tabla de múltiples gráficos. En este reporte se presenta una actividad de aprendizaje invertido referida a los supuestos de normalidad dentro de la inferencia estadística, en particular, intervalos de confianza y pruebas de hipótesis.

Antes de la clase

	Grupo testigo (sin marcapasos)		Grupo piloto (con marcapasos)	
	Pulso (segundos)	Intensidad del pulso (mvolts)	Pulso (segundos)	Intensidad del pulso (mvolts)
9 Num pers	x1	x2	y1	y2
10 1	1.2	0.131	0.94	0.14
11 2	0.9	0.303	0.81	0.296
12 3	0.9	0.297	0.82	0.281
13 4	0.8	0.416	0.73	0.355
14 5	0.7	0.595	0.60	0.444

Se parte de una situación de decisión de si un marcapasos resulta benéfico o no. Las variables decisivas son el tiempo de entre pulsos del corazón (segundos) y la intensidad de pulso (mili voltios). También se requiere hacer estimaciones confiables del valor esperado en población, para cada variable. Cada estudiante explora los datos correspondientes a dos grupos de 25 personas cada uno, con el propósito de dar argumentos gráficos y numéricos de la posible normalidad (o no) de cada población de donde fueron extraídos.

Posteriormente los estudiantes estudian un video que les ayudará a comparar linealmente los percentiles de los datos con los de la normal unitaria. Se trata de hacer un gráfico de dispersión por variable entre las frecuencias relativas acumuladas de la variable versus su valor Z equivalente en la distribución Normal unitaria, para luego calcular el coeficiente de correlación de Pearson y hacer

una interpretación. Se argumenta sobre los alcances y limitaciones de este procedimiento.

En la clase

El profesor y los alumnos discuten sobre las distintas pruebas de normalidad que hay a disposición tales como la prueba de Ryan-Joiner, Kolmogorov-Smirnov, Shapiro-Wilk, Shapiro-Francia, Anderson – Darling, Jarque-Bera, entre otras, y, con el uso de los recursos abiertos para R, ellos aplican al menos dos pruebas de normalidad a sus datos, e interpretan el *valor p* resultante.

Después de la clase

A partir de los argumentos de normalidad, los estudiantes ensayan las hipótesis estadísticas correspondientes para tomar de decisión de la funcionalidad del marcapasos.

2.3 Evaluación de resultados

En el módulo de Estadística descriptiva los estudiantes tuvieron la oportunidad de comparar la eficiencia de R sobre Excel con relación a agrupar y graficar datos. Sus opiniones estuvieron divididas sobre la naturaleza normal de los datos: para algunos tenían sesgo (48%) y para otros eran histogramas aproximadamente simétricos (52%) de un total de 25 alumnos.

```
> par(mfrow=c(2,2))
> hist(M$x1)
> hist(M$x2)
> hist(M$y1)
> hist(M$y2)
```

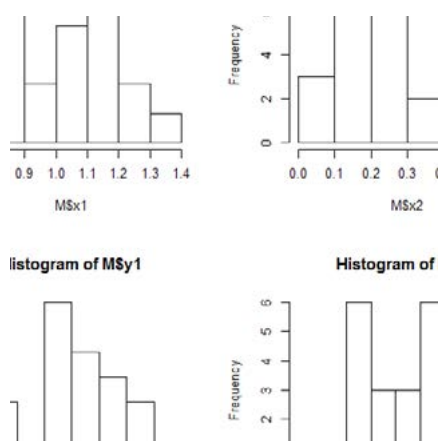


Gráfico 1. Una muestra de gráficos con R hecha por los alumnos de los cuatro grupos de datos

Por otra parte, un gráfico representativo de diagrama de caja y bigote que hicieron es el siguiente:

```
> boxplot(M)
```

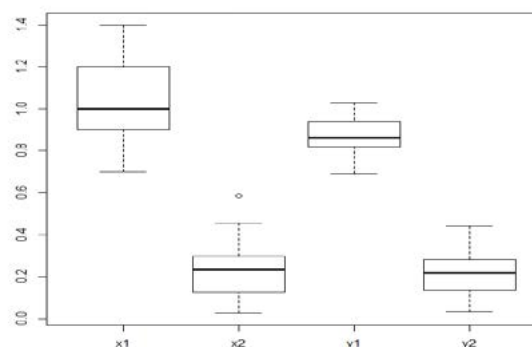
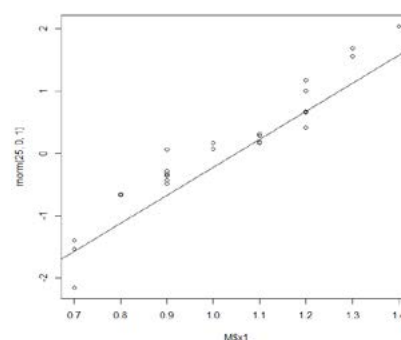


Gráfico 2. Una muestra de gráficos, caja y bigote, con R de los estudiantes

En esta etapa, el 76% de los estudiantes identificó las similitudes en valor de mediana entre x2 y y2, así como entre x1 y y1, así como la no concordancia en variabilidad con relación a x1, y1. Sin embargo, la mayoría tuvo problemas para interpretar los resultados adecuadamente en función del contexto del problema.

Con relación a la segunda actividad donde se les pedía graficar un diagrama de dispersión entre las variables Z (normal unitaria) y X (la variable en cuestión que podía ser x1, x2, y1, y2) a través de las equivalencias de sus respectivos percentiles, obtuvieron alrededor del 50% de ellos un gráfico similar al de la Gráfica 3. El resto también lo hicieron, pero con los valores en los ejes correspondiente a la proporción de los percentiles.



Gráfica 3. Gráfica que muestra una aparente alta correlación entre Z y x1

En un tercer momento de la actividad, dentro de la clase.

Los estudiantes hallaron en recursos abiertos de R, la biblioteca especializada en pruebas de normalidad nortest así como las instrucciones para ejecutar la correspondiente prueba en R. A modo de ejemplo de lo que hicieron se reporta lo siguiente:

```
> shapiro.test(M$x1)

      Shapiro-Wilk normality test

data:  M$x1
W = 0.94544, p-value = 0.1974
```

El resultado muestra que suficiente evidencia para afirmar que el grupo de datos x1 proviene de una distribución de población normal debido a que el valor p de 0.19 es muy elevado (comparativamente al nivel de significación de 0.05).

3. Conclusiones

En general se pudo observar que los estudiantes tuvieron gran facilidad para utilizar R para aplicar las Pruebas de Ryan-Joiner y Shapiro-Wilk y argumentar sobre la normalidad de los datos y quedaron muy impresionados de la facilidad con que se pueden hacer estas pruebas usando R. Este caso pudo dar evidencia de que los REA son un poderoso recurso para que los estudiantes vayan más allá, según sus propios intereses, de lo que les ofrece un curso formal.

Si bien esto es válido para nuestros cursos ordinarios, para el caso de los cursos honors, es una exigencia ofrecer este tipo de recursos de acceso abierto a nuestros estudiantes para que potencien sus propias capacidades e iniciativas de aprender por propia cuenta.

Referencias

- Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. International Society for Technology in Education. Vol. 20.
- García-Penalvo, F. J., García de Figuerola, C., & Merlo, J. A. (2010). Open knowledge: Challenges and facts. Online Information Review, 34(4), 520-539.
- Ramírez, M. S. (2015). Acceso abierto y su repercusión en la Sociedad del Conocimiento: Reflexiones de casos prácticos en Latinoamérica. *Education in the Knowledge Society*, 16(1), 103-118.

- Ripley, B. D. (2001). The R project in statistical computing. MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network, 1(1), 23-25.