

120 UNA ACTIVIDAD DE FORMACIÓN PRÁCTICA SOBRE CORRELACIÓN Y REGRESIÓN LINEAL SIMPLE CON RSTUDIO EN LA ENSEÑANZA DE ESTADÍSTICA

Belcastro Nilda Esther – Bogoni Gladys

Delegación Comodoro Rivadavia Facultad de Ciencias Económicas, Universidad Nacional de la Patagonia San Juan Bosco

nildaeb@gmail.com – gladysbogoni@gmail.com

Especialidad: Estadística aplicada

Palabras clave: Estadística, Rstudio, Regresión Lineal Simple, Formación Práctica, R

Resumen

La Ley de Educación Superior establece que los planes de estudio de carreras correspondientes a profesiones cuyo ejercicio pudiera comprometer el interés público, deben tener en cuenta la carga horaria mínima, los contenidos curriculares básicos y los criterios sobre intensidad de la formación práctica.

La resolución ministerial 3400-E/2017 establece que la intensidad de la formación práctica al igual que la carga horaria, deben ser criteriosamente establecidas, teniendo en cuenta cuánto puede aprender adecuadamente un alumno por la periodización de los procesos de enseñanza – aprendizaje que se pretenda establecer. Los programas de las asignaturas se han adecuado asignando un 40% del total de las horas, a prácticas profesionales.

En estadística planteamos distintas estrategias de formación práctica para algunas unidades, entre ellos, aprendizaje basado en problemas, proyecto de investigación de cátedra, proyecto de extensión de cátedra y método de casos.

El número de investigaciones sobre la didáctica de la estadística es aún muy escaso, en comparación con las existentes en otras ramas de las matemáticas, es por ello que el presente trabajo pretende mostrar el proceso de enseñanza de estadística en la Unidad de Correlación Lineal Simple y Regresión Lineal Simple, donde proponemos el uso del entorno R utilizando como interfaz el Rstudio, ejercicios prácticos tipo, actividades de formación práctica, en forma presencial y a través de aula virtual. Como así también conclusiones sobre la opinión de los alumnos frente a estas nuevas herramientas informáticas de aprendizaje y los rendimientos de los estudiantes es esta unidad comparando con años anteriores.

1. Introducción

La estadística no puede ser comprendida separada de su contexto de aplicación, ni aplicada únicamente a problemas abstractos que no se encuentran en la vida real. Ello implica que los conceptos y técnicas estadísticas deben ser presentadas contextualizadas, tal que permitan el desarrollo de las diferentes fases de un estudio estadístico: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos, obtención de conclusiones sobre el problema planteado, previsiones, toma de decisiones, etc.

Es innegable que la economía, la contabilidad y la administración, necesitan de la Estadística, ya que constituye un instrumento de suma importancia para que se conozca el comportamiento de las diferentes variables a diferentes niveles. En la era digital, la información es sinónimo de poder. Actualmente las grandes empresas manejan grandes conjuntos de datos, Big Data, que sólo pueden entenderse mediante su procesamiento a través de aplicaciones informáticas. Estas aplicaciones buscan patrones y coincidencias dentro de los datos, clasificándolos y organizándolos de forma que sean de utilidad y puedan comprenderse sencillamente.

Preparar a los estudiantes para aprender a lo largo de toda la vida aparece como fin de la formación básica.

Proporcionar a los jóvenes una educación que abarque los conocimientos y las competencias básicas que resultan necesarias en la sociedad actual, puede estimular en ellos el deseo de seguir aprendiendo desde la autonomía y capacidad de aprender por sí mismos.

En nuestra asignatura, como una implementación de las Ntics podemos mencionar la experiencia de la utilización de la plataforma virtual Moodle y sus facilidades, que contribuyen tanto como apoyo informático en cuanto a la facilidad de acercar a los estudiantes material digital, archivos zipeados que facilitan la instalación del Rstudio; videos instructivos de instalación, como así también, actividades de formación práctica, evaluar con posibilidad de retroalimentación estas actividades de formación práctica, foros, realizar autoevaluaciones y cuestionarios.

2. Fundamentación

Los estudiantes que adquieren habilidades de cálculo, pero que no comprenden ni pueden interpretar lo que están calculando en el contexto problemático, están desarrollando sólo una parte de la estadística, no están aprendiendo a manipular los conceptos, ni a descubrir relaciones entre ellos. Los softwares estadísticos y la tecnología son una herramienta que ayuda a lo anterior.

Por ello este trabajo tiene como propósito formular una alternativa a la problemática del proceso de la enseñanza aprendizaje de la Estadística, en particular el tema Regresión y Correlación, en base al contexto real y el trabajo colaborativo, para lo cual utilizamos herramientas disponibles como el software estadístico R y la utilización de datos del mundo real. Consideramos estrategias que implican:

- Un aprendizaje grupal.
- Un aprendizaje basado en recursos (Ntics)
- Un aprendizaje basado en una situación específica más que un aprendizaje teórico.

2.1 . Ntics en Educación

“El auge y desarrollo de las Nuevas Tecnologías de la Información y Comunicación (Ntics) permite incorporar a la educación un recurso muy bien recibido por los alumnos, ya que ellos utilizan la computadora como elemento lúdico, de comunicación y de información, y el acercamiento al estudio, podría resultarles más atractivo, práctico y hasta económico en algunos casos”. Manacorda, Mc Cormack, Pezullo, Alvarez, Maimone, Villagra (2008)

En este sentido, coincidimos con Batanero (2000) cuando expresa que la ventaja de la informática es su naturaleza dinámica, su velocidad, y el creciente rango de software que permite a los estudiantes desde experimentar y explorar todos los aspectos de los procesos estadísticos, la planificación de la muestra o del diseño experimental hasta la recolección y el manejo de datos, la simulación y el análisis, para interpretar y comunicar los resultados.

2.2. R en Educación e Investigación

R está dedicado a cómputos estadísticos y financieros, es libre y como tal se distribuye gratuitamente bajo los términos GNU, General Public Licence. Se basa en una interfaz de línea de comandos, que requiere un entrenamiento básico para su utilización.

Coincidimos con Mirabal Sosa, Robaina García y Uranga Piña (2010) cuando mencionan que R es hoy día probablemente el entorno más usado por las universidades para investigaciones en estadística, lo cual ha garantizado su robustez.

2. 3. Rstudio: entorno de R

Rstudio es una interfaz de R, que facilita un entorno visual sencillo, especialmente a usuarios poco experimentados. Entre sus ventajas podemos mencionar su entorno unificado con cuatro ventanas como máximo, visor de gráficos e historial, visor de paquetes, autocompletado de código y variables, fácil instalación y actualización de paquetes. Los paquetes son las extensiones que el R necesita para poder realizar ciertas funciones. Otra de sus facilidades es la practicidad de poder cambiar fácilmente de una base de datos a otra para realizar análisis estadísticos y no perder de vista los datos, que se visualizan en diferentes lengüetas.

2. 4. La Formación Práctica aplicada en la unidad de Regresión y Correlación

Un tema estadístico de gran aplicación, tanto exploratorio como inferencial, es la asociación o dependencia estadística entre variables. En general, las técnicas de regresión y correlación se preocupan de la interrelación entre dos o más variables cuantitativas.

2.4.1 .Correlación Lineal Simple

Dos variables están asociadas, en el sentido de que el cambio en una de estas variables estará acompañado por un cambio en la otra variable. La medida de la intensidad de la relación entre dos variables, mediante un coeficiente adecuado, constituye el problema de correlación.

2.4.2. Regresión Lineal Simple

Se estudia el tipo más sencillo de análisis de regresión en el que interviene una variable independiente y una variable dependiente y en el que la relación entre estas variables es aproximada mediante una línea recta. Esta relación se representa por un modelo matemático, dado por la ecuación de regresión, a la que se añaden un conjunto de suposiciones básicas. Ya sea para obtener una descripción de la relación entre las variables, como una indicación de una posible causalidad o si se quiere predecir la variable dependiente, a partir de los valores de las variables independientes, lo cual es muy útil si la variable dependiente es costosa o difícil de medir.

3 Desarrollo

En esta sección comentamos sobre las actividades de formación práctica, las rutinas escritas para Rstudio, los datos de Salida del software Rstudio para Correlación Lineal Simple y para Regresión Lineal Simple. Y mostramos la

comparación de las distribuciones de las calificaciones de los estudiantes en los test de lectura, en las autoevaluaciones y en las calificaciones del ejercicio del parcial correspondiente a la unidad de Regresión y Correlación.

También hemos realizado análisis sobre el rendimiento de los alumnos y comparaciones sobre rendimientos con respecto a años anteriores.

Adjuntamos algunos gráficos sobre el análisis de la Encuesta Final de Cátedra realizada.

3.1 Actividad de Formación Práctica

Presentamos al alumno material de lectura sobre el tema, y una clase expositiva-participativa para introducir los conceptos principales. Previamente a las actividades prácticas propuestas realizamos un Test de Lectura sobre el material digital entregado a través del Aula Virtual.

Para esta experiencia de formación práctica hemos utilizado el método de casos. Como inicio, planteamos al alumno un problema motivador con una situación problemática, los estudiantes deben analizar y generar una solución. Trabajan en forma individual en una primera instancia y luego en forma grupal hasta lograr definir la solución, siempre bajo la mediación de los docentes. Para esta actividad en el presente ciclo lectivo, utilizamos datos de setenta y nueve compañías multinacionales, los datos registrados de cada una de ellas son activos, ventas, valor de mercado, beneficio, flujo de fondos y empleados.

Para encontrar la solución al caso, los alumnos deben realizar un análisis de Correlación Lineal Simple, donde se pretende observar si las variables a analizar están o no relacionadas linealmente, mediante el uso de funciones gráficas del Rstudio para fijar conceptos básicos, como las gráficas de la matriz de Correlación. Dibujando el correlograma de los datos, como también la gráfica de la recta de regresión estimada, buscamos entrenar la visión del alumno hacia una primera presencia o no de una regresión lineal simple, que posteriormente se verifica con la prueba estadística correspondiente.

Por medio del aula virtual de la asignatura realizamos una autoevaluación sobre el tema. En la instancia de Parcial evaluamos con dos ejercicios los conceptos de Correlación y Regresión Lineal Simple.

Para valorar estas experiencias realizamos una encuesta final de Cátedra interrogando al alumno sobre algunos aspectos básicos de la contribución del Rstudio, del Aula Virtual y de las Actividades de Formación Práctica. También, una autoevaluación sobre el conocimiento adquirido al utilizar el software Rstudio.

3.2 Rutinas genéricas de código de Rstudio

Los archivos con los códigos simplificados del Rstudio son entregados a los estudiantes, ya que les permite realizar los engorrosos cálculos de este modelo en forma rápida y sencilla. Estos códigos siguen la consigna didáctica, de forma tal que puedan correr sobre el software la misma rutina con diferentes datos, facilitando así el análisis estadístico y la interpretación y comparación de resultados, con una rapidez y celeridad impensada años atrás. Esta forma de trabajo permite, además, que los estudiantes se animen a trabajar con el software tomando confianza en su utilización, sin la necesidad de convertirse en expertos en manejo del mismo.

Leemos los datos en Rstudio desde un archivo de Excel que tiene como nombre “companias”.

Ejemplo de código RSTUDIO

```

#Códigos Regresión y Correlación
dato=companias
# definición de variables
x=dato$valormercado
y=dato$ventas
# códigos para graficar los datos
# instalo paquete ggplot2
library("ggplot2", lib.loc=~R/win-library/3.4")
plot(x,y,xlab="Valor de mercado",ylab="Beneficio")
# código para graficar matriz de correlación
pairs(~dato$activos+dato$ventas+dato$valormercado+dato$beneficio+dato$fluj
ofondos+dato$empleados,cex.labels=1.2)
# instalo paquete corrplot
library("corrplot", lib.loc=~R/win-library/3.4")
correlacion=round(cor(d),2)
corrplot(correlacion, method="number", type="upper")
# códigos para regresión
regresion=lm(y ~ x)
summary(regresion)
# gráfica de la recta de regresión estimada
plot(x,y,xlab="Valor de mercado",ylab="Beneficio")
abline(regresion, col="2")

```

3.3 Datos de Salida del software Rstudio

3.3.2 Correlación

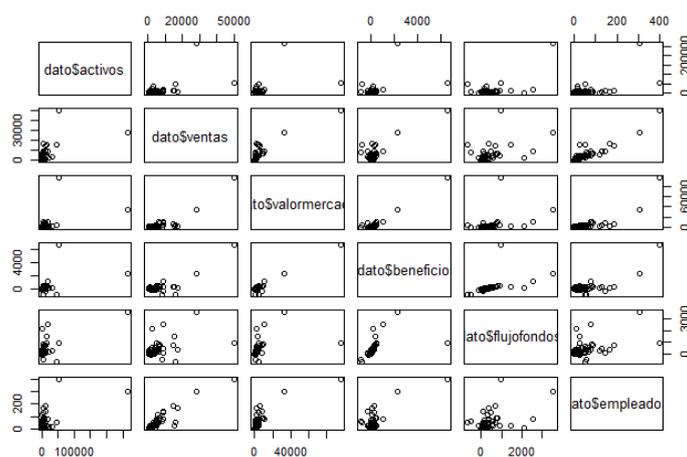
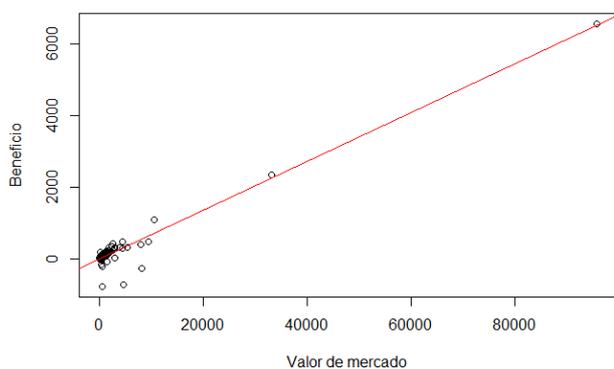


Gráfico 1. Correlogramas de las variables de la base "companias" con Rstudio.

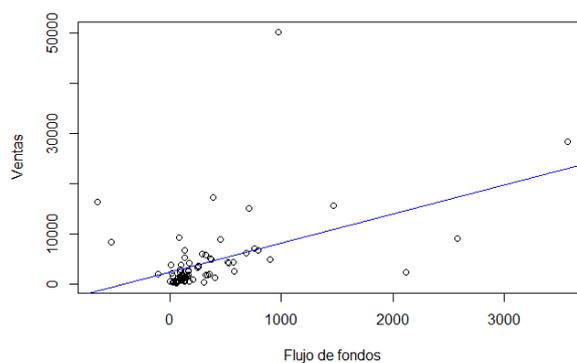


Gráfico 2. Matriz de Correlación de las variables de la base “compañias” con Rstudio.

3.3.2 Regresión Lineal Simple



Regresión Lineal Simple, con $r^2 = 0.9374$



Regresión lineal simple, con $r^2 = 0.2285$

Gráfico 3. Regresión Lineal Simple

3.3.3 Análisis de Calificaciones

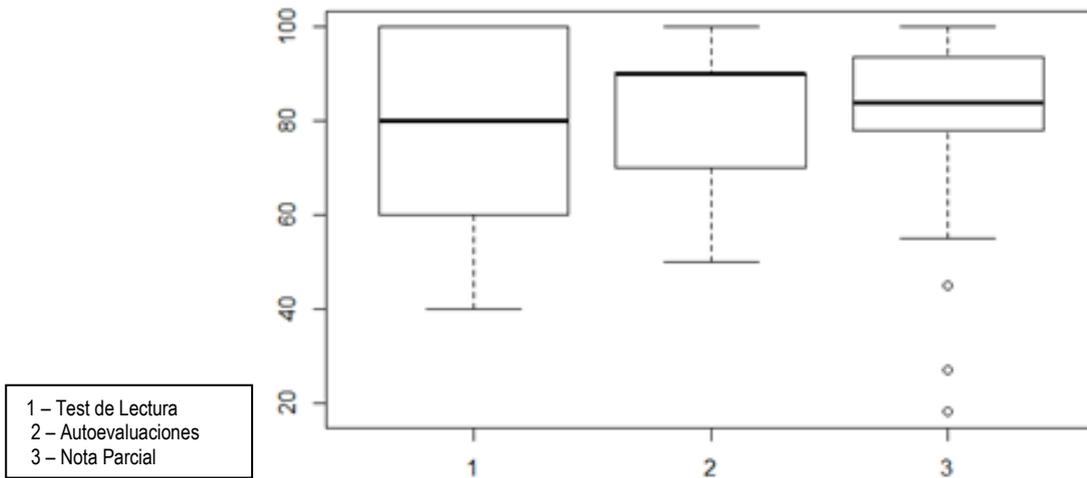


Gráfico 4. Calificaciones de Test de Lecturas, Autoevaluaciones y Notas de Parcial, sobre el tema Correlación y Regresión

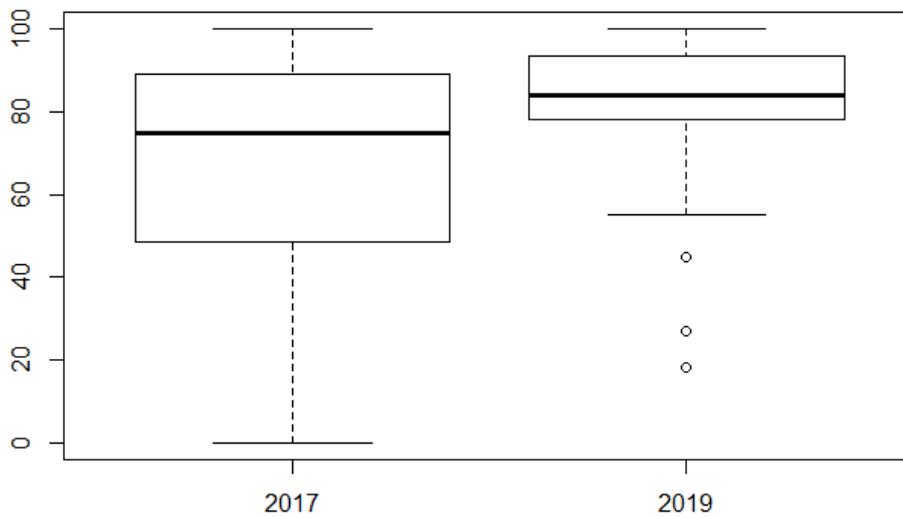


Gráfico 5. Comparación de calificaciones sobre Regresión y Correlación

3.3.4 Análisis de Encuesta Final de Cátedra



Gráfico 6. Contribución a cálculos matemáticos del Rstudio

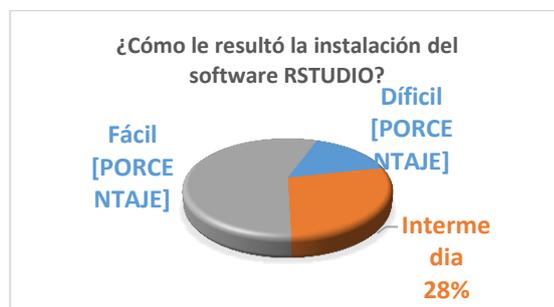


Gráfico 7. Instalación de Rstudio



Gráfico 8. Comprensión de Rstudio

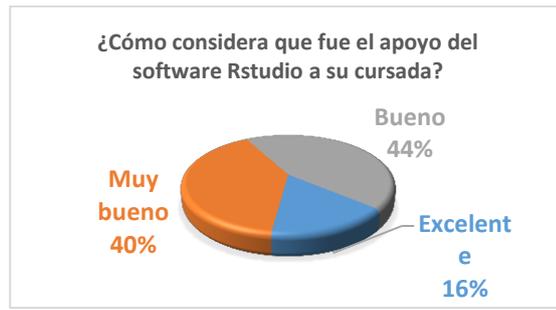


Gráfico 9. Apoyo de Rstudio al cursado



Gráfico 10. Aporte del aula virtual



Gráfico 11. Contribución Formación Práctica

4 Resultados

El Gráfico 1, obtenido a través del Rstudio, permite visualizar para todas las variables analizadas el correlograma de los datos, mostrando en la diagonal el nombre de las variables en el Rstudio. Observamos en el mismo una positiva y alta correlación entre ventas y valor de mercado, y una baja y positiva correlación entre los activos y el beneficio.

El Gráfico 2, también obtenido del Rstudio, muestra los valores del coeficiente de correlación incluyendo una escala de colores para facilitar su interpretación. Del mismo podemos concluir que entre las ventas y los empleados hay una fuerte y directa correlación, dado su coeficiente de correlación r de 0.9; y entre el valor de mercado y el flujo de fondos existe una directa y débil correlación, dada por su coeficiente de correlación r de 0.4.

Sobre Regresión Lineal Simple, observamos en el Gráfico 3 la posibilidad que brinda este software para graficar en el mismo diagrama tanto los puntos observados como la recta de regresión estimada. Mostramos a la derecha un análisis de regresión lineal simple con un coeficiente de determinación alto, como 0.9374; y a la izquierda una regresión lineal simple con un r^2 de 0.2285.

Análisis del rendimiento

Si bien sabemos que son muchas las variables que afectan al aprendizaje y al rendimiento académico de los alumnos, en este caso se analizó la calificación obtenida.

Del análisis del rendimiento de Regresión y Correlación (Gráfico 4) podemos indicar que, tanto los test de lectura como la autoevaluación y el parcial, indican que los estudiantes obtienen buenas calificaciones, se observa claramente que en el test de lectura los estudiantes tienen calificaciones más dispersas que en las otras instancias. Podemos destacar también que la menor dispersión se observa para las calificaciones del parcial. Las calificaciones del parcial son mejores a las obtenidas en la autoevaluación, si bien hay tres alumnos que presentan calificaciones outliers, pertenecientes a los alumnos que no aprobaron el tema. El 82 % de los alumnos aprobó el tema en el parcial.

De la comparación de las calificaciones obtenidas en la unidad de Regresión y Correlación entre los años 2017 y 2019 (Gráfico 5) podemos observar que las calificaciones del 2019 son superiores a la del 2017, y sobre todo presentan menor dispersión. Aparece la presencia de tres puntos outliers, que corresponden a aquellos alumnos que desaprobaron con muy bajo puntaje.

Análisis de la encuesta final

Del análisis de algunas preguntas de la Encuesta Final de Cátedra, relacionadas con el tema a tratar podemos concluir, observando los Gráficos 6 a 11 que el 76 % de los alumnos considera que el Rstudio contribuyó en mucho a los cálculos matemáticos, el 56 % considera que fue fácil la instalación del software, el 92 % considera que **comprendió la forma de utilización del Rstudio**, el 16% considera Excelente y el 40 % Muy bueno el aporte de este software a su cursada, el 92 % considera que fue alto el aporte del aula virtual y que el 68 % que fue mucho el aporte de las Actividades de Formación Práctica a su cursada.

5 Conclusiones y trabajos futuros

Las actividades de Formación Práctica, cuando incluyen situaciones en las que se contrastan hipótesis, se diseñan planes para resolver problemas, se analizan e interpretan resultados, y se obtienen conclusiones, tienen un valor formativo muy importante. Las actividades aplicadas a problemas reales han permitido que los estudiantes se apropien del razonamiento estadístico, aumenten la motivación y responsabilidad por su propio aprendizaje y desarrollen el razonamiento estadístico hacia la comprensión y capacidad de explicar e interpretar resultados de procesos estadísticos. Introducir a los estudiantes en el procesamiento y análisis de datos con el Rstudio apoya lo anterior, ya que el alumno no debe concentrarse en fórmulas ni cálculos.

Con esta propuesta pretendemos que los estudiantes logren el aprendizaje de la estadística de una manera básica, real y aplicada, utilizando recursos didácticos, con su propia iniciativa, orientados por el equipo docente.

La implementación de estas estrategias conjuntamente con actividades propuestas a través del Aula Virtual y del Rstudio supone una importante planificación y gestión por parte de los docentes, es por ello que nos importa la opinión y el rendimiento de los estudiantes.

La encuesta final a través del aula virtual, donde consultamos, entre otros temas, sobre la utilidad del Aula Virtual, sobre las Actividades de Formación práctica y sobre el aporte del software Rstudio como herramienta de cálculo, nos permitió evidenciar que se están logrando resultados positivos. Las encuestas de este tipo son una herramienta básica para detectar los puntos fuertes y débiles de la oferta formativa.

El análisis de la encuesta final y del rendimiento proporciona una retroalimentación muy importante al proceso de enseñanza - aprendizaje, por lo que nos permitirá ajustar las actividades y guiar a los alumnos a lograr los fines propuestos por la cátedra.

Referencias

BATANERO, C. (2010) Didáctica de la Estadística. GEEUG. Departamento de Didáctica de la Matemática. Universidad de Granada, de <https://www.ugr.es/~batanero/pages/ARTICULOS/didacticaestadistica.pdf>

MANACORDA, A. M., MC CORMACK, L., PEZZULLO, S. y ÁLVAREZ, A. (2008) Incorporación de las NTICs en clases de ciencias del nivel superior. de Congreso Virtual Iberoamericano de Calidad en Educación a Distancia Educ@2008 Ponencias. Recuperado en 19 de julio de 2019 http://eduqa2008.eduqa.net/eduqa2008/images/ponencias/eje_tematico_5/5_73_Utilizacion_de_la_plataforma_Manacorda_Cormack_Pezzullo_Alvarez_Maimone_Villagra_.pdf

MIRABAL SOSA, M., ROBAINA GARCÍA, M. y URANGA PIÑA, R. (2010). R: una herramienta poco difundida y muy útil para la investigación clínica. Revista Cubana de Investigaciones Biomédicas, 29(2), 302-308. Recuperado en 23 de julio de 2019, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-03002010000200012&lng=es&tlng=es.