

USOS DEL PROGRAMA R EN LA ENSEÑANZA DE LA ESTADÍSTICA

Albeiro Enrique López Cervantes

Institución Educativa Normal Superior de Sincelejo (Colombia)

albeirolopecervantes@gmail.com

Palabras clave: Tecnología, Programa R, Estadística Descriptiva y Situación Problema

Key words: Technology, R program, Descriptive Statistics and Problem Situation

RESUMEN: Es sabido por todos la trascendencia que tiene el aprendizaje de la estadística en estudiantes de bachillerato, en especial los estudiantes del nivel medio superior, en su proceso de formación preparatorio para el ingreso a la universidad. El programa estadístico R-Project se ha constituido en una herramienta básica y fundamental que posibilita enseñar temas de la estadística descriptiva y de la teoría de probabilidades de manera coherente con la estructura conceptual..

ABSTRACT: It is known to all the importance that the learning of statistics on high school students, especially students from high school, in the process of preparatory training for college entrance. The statistical program R -Project has become a basic and fundamental tool that enables teaching subjects descriptive statistics and probability theory consistent with the conceptual framework.

■ INTRODUCCIÓN

Hoy por hoy la enseñanza de la estadística se ha vuelto fundamental en los procesos formativos de los jóvenes que cursan los últimos años de bachillerato o lo que se conocen como preparatoria para los estudios universitarios. La enseñanza de esta disciplina generalmente ha sido abordada en este nivel de forma tradicional recurriendo a extensos cálculos manuales y dependiendo de la pericia de los profesores, recurriendo al uso de herramientas electrónicas como la calculadora científica y la hoja de cálculo Excel. Sin embargo, los procesos de conceptualización, comprensión y aplicación siguen siendo muy débiles.

En el campo de la estadística descriptiva es fundamental que los estudiantes reconozcan las técnicas de análisis adecuadas para la sistematización de información, el establecimiento de conclusiones y la toma de decisiones. Esto sólo se logra si son capaces de caracterizar los diferentes tipos de variables que existen, si organizan la información en tablas de distribución de frecuencias, si construyen gráficos apropiadas para presentar la información, si logran hacer los análisis de la información dada a partir de los cálculos numéricos de corte descriptivo (medidas de tendencia central, medidas de variabilidad y medidas de posición) y fundamentalmente si tiene la capacidad de redactar un informe coherente de la situación problema, con la información encontrada.

R es un software o un lenguaje de comandos de manipulación y análisis estadístico basado en el lenguaje estadístico S, además es un programa de código abierto y gratis, razón por la que debería usarse en diferentes ámbitos académicos. Atendiendo a sus características, R tiene un gran potencial para ser usado en el aula de clase, es una herramienta útil por su capacidad para hacer cálculos estadísticos, crear gráficos y sobre todo, por la posibilidad de trabajar desde la estadística elemental hasta las estadísticas más avanzadas. Una de esas temáticas que pueden enseñarse usando R como ayuda didáctica es la relativa a la estadística descriptiva, asunto que se abordará a interior de este taller.

■ MARCO TEÓRICO

Las variables estadísticas suelen clasificarse en cualitativas (nominales y ordinales) y cuantitativas (discretas y continuas), cada una de ellas permite la construcción de tablas de distribución de frecuencias, la construcción de gráficas y, en el caso de las variables cuantitativas, el cálculo de medidas de tendencia central (media aritmética, mediana y moda), medidas de dispersión (desviación media, varianza, desviación estándar) y medidas de posición (percentiles, deciles y cuartiles). A partir de estos conocimientos es posible hacer análisis de la información relativa a una situación problema de interés y tomar decisiones adecuadas que apunten a encontrar alternativas de solución a las mismas. Una forma particular de elaborar informes de corte descriptivo de un conjunto de datos es a través de los intervalos de confianza calculados a través de la regla empírica para distribuciones con comportamiento normal o aproximadamente normal o que tengan forma de campana.

El marco ideal para trabajar la estadística descriptiva en el campo medio–superior escolar lo constituye el trabajo con situaciones problemas. En este taller debemos entenderla así:

Una situación problema la podemos interpretar como un contexto de participación colectiva para el aprendizaje, en el que los estudiantes, al interactuar entre ellos mismos, y con el profesor, a través del objeto de conocimiento, dinamizan su actividad matemática, generando procesos conducentes a la construcción de nuevos conocimientos. Así, ella debe permitir la acción, la exploración, la sistematización, la confrontación, el debate, la evaluación, la autoevaluación y la heteroevaluación. (Obando y Muñera, 2003, p.185)

En las últimas dos décadas se ha recomendado por diversos medios la aplicación de herramientas informáticas en la enseñanza de la estadística, en el ámbito educativo, "... Con respecto a la enseñanza de la probabilidad, la estadística y el análisis de datos, por su naturaleza disciplinar es indispensable usar tecnología computacional como coadyuvante en la formación académica". (Hernández y Cuevas, 2013, p.170).

Es necesario resaltar la diversidad de programas de cómputo que se usan como herramienta para enseñar estadística, algunos de ellos son programas de uso específico bajo licencia (Fathom, StatGraphics, SPSS, Minitab, Statistica, EViews, SAS y NCSS), otros que son de distribución gratuita (OpenStat, CAEST, StatDisk y PSPP) y los que conforman entornos de programación robustos como R.

El tipo de interacción que se propone en este taller es a través de herramientas computacionales, en particular con el programa R, como apoyo didáctico para la enseñanza y aprendizaje de conceptos básicos de la estadística descriptiva, buscando que "la interacción con el lenguaje estadístico pueda mejorar la comprensión y proporcionar un conocimiento más detallado y en profundidad de los métodos" (Ledesma, Valero-Mora y Molina, 2010, p.53).

El programa R-Project brinda mayores ventajas en el proceso de enseñanza y aprendizaje de la estadística en el nivel medio superior en comparación con los otros programas de uso libre, pues permite la construcción de conceptos (R puede utilizarse como un lenguaje de programación) y dispone de un conjunto de funciones matemáticas que complementan la exploración de los conceptos estadísticos. Por otra parte contiene funciones propias y subprogramas especializados que permiten verificar los resultados de manera automatizada como los otros programas.

El programa R dispone de un conjunto de funciones para calcular medidas de tendencia central [*mean()*, para la media y *median()*, para la mediana], para calcular las medidas de dispersión [*var()*, para la varianza y *sd()*, para la desviación estándar] y para calcular las medidas de posición [*quantile()*, para los cuartiles], pero no tendría ningún sentido didáctico usarlos sin tener claridad conceptual sobre los temas que involucran las situaciones problemas.

El propósito principal del taller llevado a cabo fue que los participantes usen el programa R para explorar los conceptos relativos a la estadística descriptiva en el nivel medio superior, utilizando funciones matemáticas básicas que les permitan la construcción de los conceptos. De igual manera, que esta experiencia contribuya al diseño de nuevas propuestas de intervención en el aula donde el programa R sirva como herramienta y como instrumento de enseñanza y de aprendizaje de la estadística.

■ PROPUESTA DE ACTIVIDAD

A continuación se detallan las etapas desarrolladas a partir de la siguiente situación problema:

El coordinador académico de la I.E. Normal Superior de Sincelejo desea evaluar los resultados obtenidos por los estudiantes de undécimo grado en las pruebas SABER 11 del año 2014, específicamente en el área de matemáticas. Para ello decide tomar una muestra significativa de los resultados obtenidos en esta área, estos son: 45, 45, 48, 51, 51, 48, 55, 51, 48, 48, 48, 51, 45, 45, 51, 48, 51, 51, 51, 55, 51, 55, 51, 51

El coordinador solicita a los estudiantes de grado undécimo que realicen un análisis de corte descriptivo con los puntajes dados, presenten un informe que incluya tablas de distribución de frecuencias (absolutas, relativas y porcentuales), gráficos estadísticos apropiados (diagrama de barras o diagrama circular) y medidas descriptivas numéricas (media, mediana, moda, desviación media, varianza, desviación estándar, cuartiles). Además la estimación de un intervalo de confianza para el verdadero valor de la media poblacional del puntaje obtenido por los estudiantes en la prueba de matemáticas, usando la regla empírica.

Realizando el análisis de corte descriptivo

1. Abre el programa R-Project y simultáneamente un script para escribir los comandos básicos para construir una base de datos con la información dada y procede a realizar los cálculos de las frecuencias absolutas, relativas y porcentuales.
2. Dispón nuevamente del script para construir diagramas de barras y diagramas circulares para presentar la información dada en la base de datos.
3. ¿Has trabajado estas situaciones sin apoyo de herramientas electrónicas? ¿Qué ventajas o desventajas tiene trabajar estos conceptos utilizando este programa estadístico? ¿Te arriesgarías a usarlo en tus clases?
4. Escribe sobre el script las rutinas para calcular las medidas de tendencia central, las medidas de dispersión y las medidas de posición asociadas a los cuartiles.
5. Sobre el script escribe directamente los comandos `median()`, `mean()`, `var()`, `sd()` y `quantile()` para obtener directamente los resultados pedidos.
6. Comente como ha sido tu experiencia de trabajo con estos conceptos en el aula. ¿Qué ventajas o desventajas tiene explorar estos conceptos usando las rutinas de cálculo descritas anteriormente? ¿Te animarías a usar este programa en tus clases?

Pasando a la preparación del informe solicitado

1. Tome uno de los gráficos elaborados y analícelo a luz de la situación problema.
2. Estime los intervalos de confianza usando la media o promedio y la desviación estándar de los datos. Haz conjeturas sobre los resultados encontrados.

Buscando una extensión:

1. ¿Qué opinas sobre el desarrollo de esta actividad?
2. Propón una situación problema que nos permita, como docentes, poner en práctica las actividades desarrolladas.
3. ¿Qué situaciones problemas propondrías a tus alumnos para el proceso de enseñanza y aprendizaje de la estadística descriptiva?

Ensayo la construcción de algunas situaciones problemas que posibiliten la enseñanza y el aprendizaje de la estadística descriptiva, usando el programa R.

■ MÉTODO

A partir de la base de datos *Puntaje* construida en el programa R se obtienen las frecuencias absolutas usando la función *table()*. Como este programa permite la construcción de nuevas funciones a partir de las que ya tiene establecidas, se le llámela *fa* y se determinela así: $fa = table(puntaje)$. Se establece el tamaño de la muestra usando la función *length()* de la siguiente manera: $n = length(puntaje)$. A partir de estos resultados se calculan las frecuencias relativas $fr = fa/n$. Si las frecuencias relativas resultan con muchas cifras decimales, se aproxima el resultado a un número determinado de cifras decimales usando la función *round()*, así $fra = round(fr, 3)$, en este caso a tres cifras decimales. Con los resultados obtenidos para las frecuencias relativas se calculan las frecuencias porcentuales *P* usando la expresión $P = fra * 100\%$.

Para construir los diagramas de barras y los diagramas circulares se utilizan las funciones *barplot()* y *pie()* de la siguiente forma: para los diagramas de barras se usan *barplot(fa, ylab="Frecuencias absolutas", main="Diagrama de barras")* o *barplot(P, ylab="Porcentajes", main="Diagrama de barras")* y para los diagramas circulares se usan *pie(fa, main=c("Diagrama circular"))* o *pie(P, main=c("Diagrama circular"))*.

Para el cálculo de las medidas de tendencia central se utilizan las funciones *sort()* para ordenar los datos en forma ascendente, *sum()* para sumarlos y las siguientes expresiones funcionales para determinar la media y la mediana:

$Me = sum(puntaje)/n$ y $Md = (D_{(n/2)} + D_{(n/2+1)})/2$, donde $D_{(n/2)}$ y $D_{(n/2+1)}$ corresponden respectivamente a los datos centrales de la distribución ordenada, cuando el número de datos es par, si es impar $Md = D_{(n/2)}$.

Con relación al cálculo de las medidas de dispersión se utilizan las funciones *abs()* para determinar valores absolutos de diferencias, *sqrt()* para extraer raíz cuadrada, *round()* para aproximar el resultado a un número de cifras decimales determinado y las siguientes expresiones funcionales para determinar la desviación media (*dm*), la varianza (*va*) y desviación estándar (*de*), teniendo en cuenta que X_1, X_2, \dots, X_k son los valores que asume la variable en la tabla de distribución de frecuencias absolutas y f_1, f_2, \dots, f_k son las frecuencias absolutas correspondientes a cada valor:

$$dm = (abs(X_1 - Me) * f_1 + abs(X_2 - Me) * f_2 + \dots + abs(X_k - Me) * f_k) / n$$

$$va = ((X_1 - Me)^2 * f_1 + (X_2 - Me)^2 * f_2 + \dots + (X_k - Me)^2 * f_k) / (n - 1)$$

$$de = sqrt(va) \text{ y si requiere aproximación } de = round(de, 2).$$

Para calcular algunas medidas de posición de interés utilizamos las funciones *sort()*, *round()* y las expresiones funcionales siguientes:

$Q_1 = (n+1)/4$ para calcular la posición del primer cuartil, $Q_{1a} = round(Q_1, 0)$ para identificar la posición aproximada al entero más cercano y $X_{Q_{1a}}$ para determinar el valor del primer cuartil.

$Q_3 = 3*(n+1)/4$ para calcular la posición del tercer cuartil, $Q_{3a} = round(Q_3, 0)$ para identificar la posición aproximada al entero más cercano y $X_{Q_{3a}}$ para determinar el valor del tercer cuartil.

Finalmente se usan las siguientes expresiones funcionales para la estimación de intervalos de confianza para distribuciones de datos normales o aproximadamente normales, usando la regla empírica, así:

Para aproximadamente el 68% de las mediciones:

$$Li=Me-de; Ls=Me+de; IC=round(c(Li,Ls), 1)$$

Para aproximadamente el 95% de las mediciones:

$$Li=Me-2*de; Ls=Me+2*de; IC=round(c(Li,Ls), 1)$$

Para aproximadamente el 100% de las mediciones:

$$Li=Me-3*de; Ls=Me+3*de; IC=round(c(Li,Ls), 1)$$

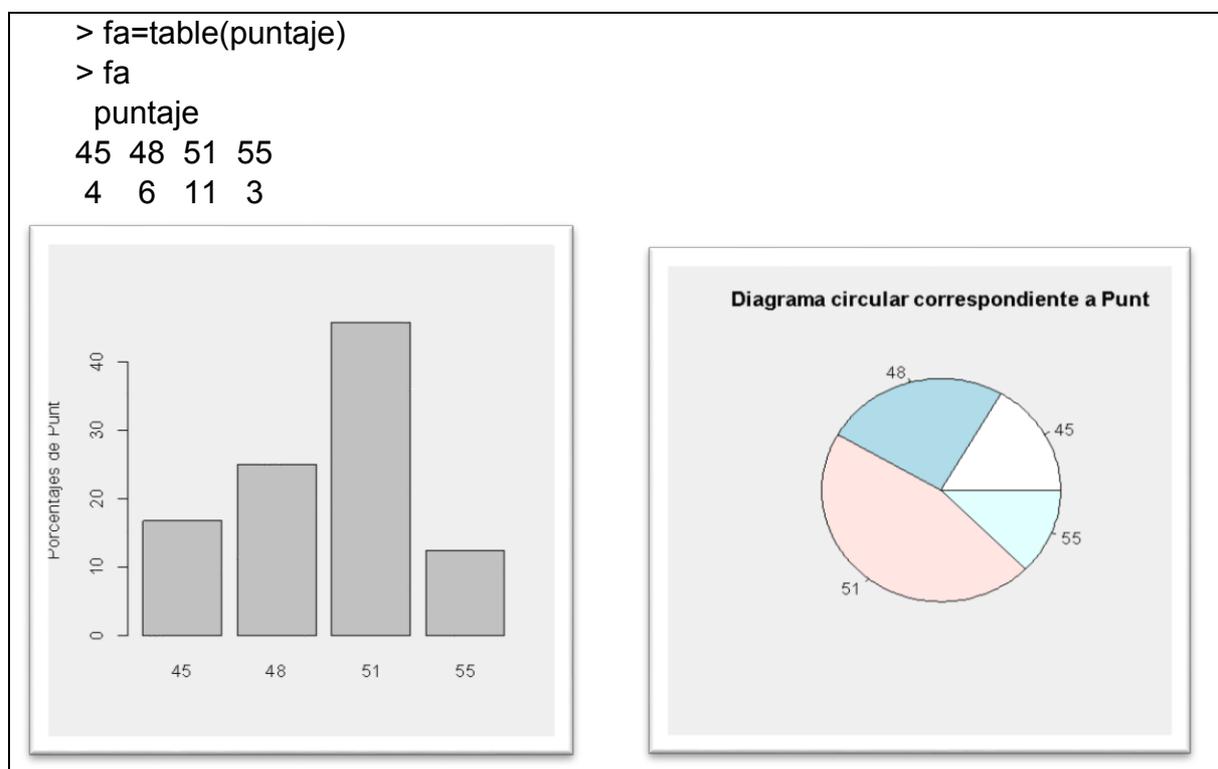
Se solicita a los participantes que elaboren sus informes a partir de la información obtenida y se discute con ellos las ventajas que tiene el uso del programa R para explorar conceptos asociados con la estadística descriptiva, en comparación con los métodos tradicionales donde se emplean calculadoras científicas y la hoja de cálculo Excel.

■ RESULTADOS DE LA IMPLEMENTACIÓN DEL PROGRAMA R

A continuación se comparten las principales apreciaciones, en relación al impacto de la propuesta, sobre la implementación de ésta con un grupo de educadores matemáticos latinoamericanos que asistieron al taller, en el marco de la XXIX Reunión Latinoamericana de Matemática Educativa.

Se hicieron las construcciones de las tablas de distribuciones de frecuencia absolutas y de gráficos apropiados para presentar la información dada, los cuales pueden apreciarse en la Figura 1.

Figura 1. Tabla de frecuencias absolutas, diagrama de barra y diagrama circular.



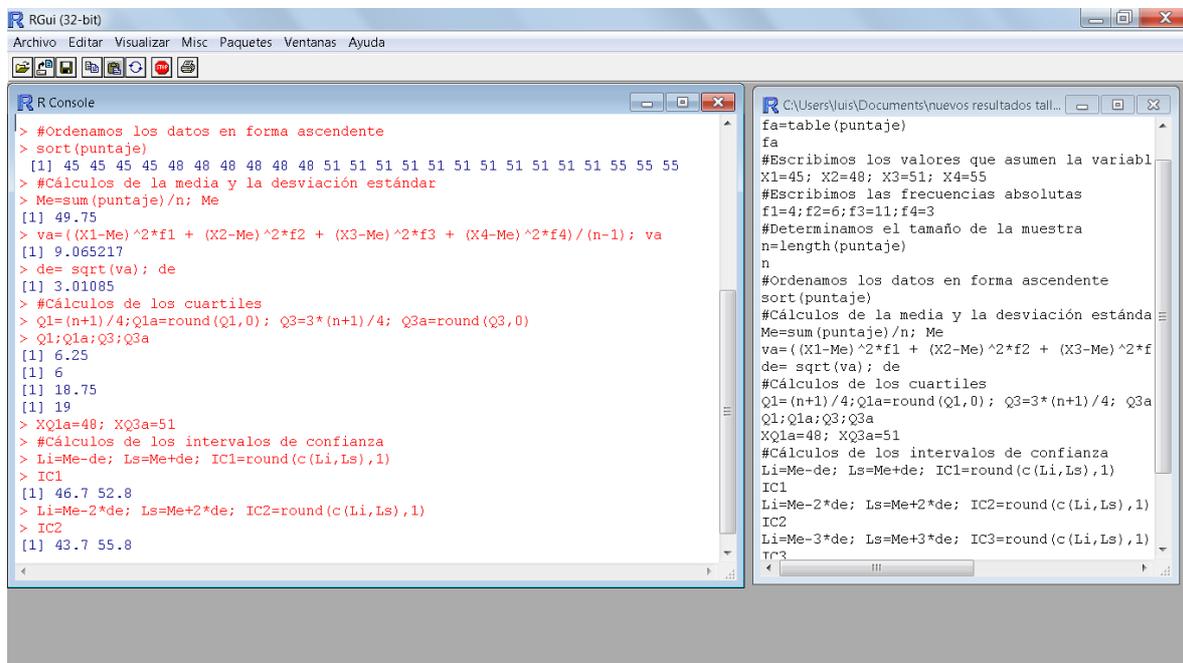
Se discutió con los participantes la explicación estadística de las construcciones, comparando las formas tradicionales de trabajar estas temáticas con la implementada en la propuesta.

Otros resultados obtenidos en el programa se muestran en la figura 2.

Las rutinas de cálculo en R se discutieron con los participantes, al igual que los informes que cada uno construyó a partir de las medidas descriptivas encontradas en el proceso.

En relación a la implementación del taller, consideramos necesario el uso de herramientas informáticas como apoyo al trabajo de aula en el proceso de enseñanza y aprendizaje de la estadística en estudiantes del nivel medio superior.

Figura 2. Salidas de resultados en el programa R-Project.



Los participantes del taller manifestaron tener poco contacto con programas estadísticos que sirvieran de apoyo para el trabajo en el aula con estas características. Sin embargo, consideraron y discutieron aspectos centrados en la incidencia de esta propuesta en los procesos y resultados del aprendizaje de los temas relacionados con la estadística descriptiva.

Resaltaron la fortaleza del programa R al facilitar a los estudiantes una aproximación a los conceptos con relativa facilidad, además de permitir realizar estimaciones con cualquier tipo de datos numéricos, con alto nivel de confiabilidad. Se mencionaron las bondades del programa, al permitir visualizar los resultados que obtienen al manipular formulas y funciones propias del programa R.

Todos los participantes en el taller reconocieron en el programa su utilidad como herramienta y, cuando fueron llevados a utilizarlo, como instrumento (Moreno y Santos, 2002) y realizaron comentarios muy favorables sobre sus potencialidades didácticas. Esto pone en evidencia que su escaso uso en el aula se debe más bien a la falta de interacción con este tipo de programas, más

que por apatía o temor a usarlos. Hicieron los siguientes comentarios, con relación al uso del programa como instrumento: brinda posibilidades para interactuar el docente, los estudiantes y el objeto de conocimiento, ya que la construcción de gráficos personalizados y la programación de los cálculos de estadísticos da lugar a una nueva metodología de enseñanza de la estadística descriptiva; y el uso del programa R permite alcanzar los mismos resultados que se obtienen usando métodos convencionales de enseñanza de los temas relativos a la estadística descriptiva, pero con mayor rapidez, y permite utilizar todas las rutinas para resolver nuevas situaciones problemas modificando solo la base datos.

■ COMENTARIOS FINALES

De Oliveira, Da silva, Rodríguez y Ribeiro (2014) afirman que en los días de hoy la Estadística es una herramienta indispensable para el ejercicio de la ciudadanía, puesto que ella permite analizar informaciones y subsidiar la toma de decisiones, ya sea de la vida personal o laboral. De allí la necesidad de incluirlas en las estructuras curriculares de la educación básica y de crear escenarios que posibiliten el uso de herramientas tecnológicas en su enseñanza, "...se sabe que para el profesor es un reto cambiar la forma de dar clases, pero la tecnología está cambiando todo en el entorno y los docentes también deben actualizarse por el bienestar de sus educandos" (Sánchez, 2014, p.9).

El programa R-Project al igual que otros programas de distribución libre, sirven de base para diseñar actividades en el aula que guíen el proceso de enseñanza y aprendizaje de la estadística descriptiva, pero requieren de la orientación del docente para lograr los resultados esperados, precisamente al tener que utilizarlos como herramienta de enseñanza. Por ello debemos tener presente que las herramientas tecnológicas producirán un efecto en el aprendizaje de los estudiantes siempre que el docente adapte las mismas a las necesidades educativas.

Por último, actividades como las desarrolladas en este taller, que involucran el uso de programas estadísticos, van más allá de la acción técnica de aprender a usar una herramienta electrónica, puesto que involucra otro conjunto de habilidades ampliamente relacionadas con el desarrollo del pensamiento crítico. En este sentido

"Es interesante animar a los chicos a escribir un informe sobre su análisis, ya que la habilidad para producir informes comprensivos y estructurados donde la información estadística se incorpore y presente adecuadamente para apoyar la argumentación será sin duda útil en su futura vida profesional, sea cual fuere y es un medio también para el aprendizaje de los procesadores de texto". (Batanero, Díaz, Contreras y Arteaga, 2010, p.29).

Se trata de familiarizar a los docentes en el uso de programas estadísticos de distribución libre y de dar elementos para, a futuro, potenciar los aprendizajes estadísticos de los estudiantes en el nivel medio superior, presentando una actividad concreta a realizar bajo el programa R.

■ REFERENCIAS BIBLIOGRÁFICAS

Batanero, C., Díaz, C., Contreras, M. y Arteaga, P. (2010). Enseñanza de la estadística con proyectos. C. Batanero y C. Díaz (Editoras). Estadística con proyectos (pp.29-46). Universidad de Granada. España.

<http://www.ugr.es/~batanero/pages/ARTICULOS/Libroproyectos.pdf>.

De Oliveira, A., Da silva, B, Rodríguez, O. y Ribeiro, Vanderleia. La historia de la estadística en el caso de la enseñanza y el aprendizaje de la estadística en la escuela secundaria. Memorias del IV Encuentro sobre Didáctica de la Estadística, la Probabilidad y el Análisis de Datos. Costa Rica 2014

Hernández, S. y Cuevas, J. (2013). *Programas informáticos de uso libre y su aplicación en la enseñanza de la estadística*. Revista investigación operacional. Universidad Veracruzana, México.

Ledesma, R., Valero-Mora, P. y Molina, J.. (2010). *Un Software para la Enseñanza de la Estadística y la Psicometría*. Revista Argentina de Ciencias del Comportamiento. Volumen II No 2.

Moreno, L. y Santos, M. (2002). El proceso de transformación del uso de la tecnología en una herramienta para la solución de problemas de matemáticas por parte de los estudiantes. Memorias del Seminario Nacional de formación de docentes sobre el uso de nuevas tecnologías en el aula de Matemáticas (pp.263-268). Bogotá: Ministerio de Educación Nacional.

Obando, Gilberto y Muñera, Jairo. *Las situaciones problema como estrategia para la conceptualización matemática*. Revista educación y pedagogía. Volumen XV No 35. Colombia 2003.

Sánchez, S. "Principios básicos para la creación de animaciones interactivas con el Software Geogebra". II Encuentro Centroamericano de Matemática Educativa (2014).