

Utilización del software R para la enseñanza de la estadística: experiencia en un curso de ingeniería.

AUTORAS:

Prof. Florencia Gambetta.

Prof. Cintia Goitía.

Prof. Verónica San Román.

Prof. Carolina Zeppa.

DIRECCIONES ELECTRÓNICAS:

florigambetta@hotmail.com

cintiagoitia@hotmail.com

vsanroman@gmail.com

carolinazeppa@hotmail.com

INSTITUCION: Universidad Nacional del Sur (UNS)

Bahía Blanca. Provincia de Buenos Aires.

NIVEL: Universitario.

RESUMEN

A lo largo de las clases teóricas y prácticas en la cátedra “Estadística A” observamos ciertos conflictos para la comprensión de los conocimientos relacionados con la estadística descriptiva. Esto nos motivó a elaborar una herramienta facilitadora para la aprehensión, el desarrollo y el posterior análisis comparativo de los contenidos desarrollados.

En este trabajo presentaremos la implementación de un curso con modalidad semi-presencial utilizando el software libre R como una herramienta didáctica; el mismo se llevó a cabo con un grupo de estudiantes de las carreras de Ingeniería Mecánica e Ingeniería Industrial dictado por docentes del Departamento de Matemática del área de Estadística de la Universidad Nacional del Sur.

La experiencia se diagramó en dos etapas: **presencial** y **no-presencial**.

La **primera etapa** se llevó a cabo en algunas clases prácticas de la materia que sirvieron para la presentación e implementación del software. El espacio de trabajo

donde se desarrolló fue el laboratorio de informática del Departamento de Matemática de la Universidad Nacional del Sur.

La **segunda etapa** consistió en que los alumnos resolvieran los ejercicios propuestos en los trabajos prácticos de la cátedra utilizando R, comparando los resultados con los obtenidos previamente en papel, finalizando con una puesta en común donde se analizaron las ventajas y desventajas de R. En esta instancia, también, realizamos una devolución individual de los trabajos entregados.

Finalmente, como corolario de esta experiencia, podemos inferir que es posible seguir profundizando en los beneficios que brinda el software R en relación a la enseñanza de la estadística.

INTRODUCCION

Las aplicaciones de la estadística en la ingeniería han tomado un rápido y sostenido incremento, debido al poder de cálculo de la computación desde la segunda mitad del siglo XX, modificando así, notablemente la enseñanza de la Estadística en el aula y el análisis de la información en los lugares de trabajo.

El uso de programas estadísticos es parte integral del proceso de enseñanza-aprendizaje de la Estadística.

En este sentido, coincidimos con Batanero, C. (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.

Teniendo en cuenta que las computadoras pueden cambiar la relación entre estudiantes y profesores, ser un recurso didáctico del profesor y permitir al estudiante ser cognitivamente activo en estadística, surgió la idea de explorar el uso de la informática en relación a la enseñanza de la estadística, en particular con el software R.

Braslavsky y otros (1992) señalan que: *“Al enseñar cada profesor pone en acto una epistemología subyacente que le es propia (...) Esta epistemología subyacente se pone en evidencia, por ejemplo, (...), al seleccionar ciertos contenidos, al diseñar actividades...”*. Siguiendo estos conceptos, esta propuesta de trabajo está basada en una perspectiva constructivista permitiendo que el desarrollo de los contenidos no sea una

simple secuencia lineal sino que dé lugar a conceptualizaciones provisorias y a conocimientos no acabados; donde el rol del docente es muy importante para acompañar, servir de mediador y propiciador en esta construcción reflexiva del conocimiento.

Queremos enfatizar que uno de los objetivos del curso fue valorar el adecuado uso con el software R como un recurso complementario del aprendizaje, sin ahondar en complejas cuestiones informáticas. Considerando que la computadora como un mero objeto no mejora el proceso de enseñanza-aprendizaje, el desafío fue puesto en explorar los beneficios de sus utilidades para optimizar los resultados educativos.

¿QUÉ ES LA ESTADÍSTICA?

La estadística es la ciencia de la sistematización, recopilación, ordenación y presentación de los datos referentes a un fenómeno que presenta variabilidad o incertidumbre para su estudio metódico, con el objeto de deducir las leyes que rigen esos fenómenos, y poder de esa forma hacer previsiones sobre los mismos, tomar decisiones u obtener conclusiones. Está conformada por dos ramas: la estadística descriptiva y la inferencial.

La estadística descriptiva es la encargada de la organización, condensación, presentación de los datos en tablas y gráficos y del cálculo de medidas estadísticas que permiten estudiar los aspectos más importantes de los datos. La capacidad que brinda la estadística descriptiva de sintetizar la información de grandes volúmenes de datos es lo que la hace especialmente relevante.

La estadística inferencial está definida por un conjunto de técnicas mediante las cuales se hacen generalizaciones o se toman decisiones en base a la información obtenida de una población.

BREVE DESCRIPCIÓN DE LA EXPERIENCIA DEL CURSO

En este relato se presenta la implementación de un curso semi-presencial en el cual, a partir del software R, se trabajaron conocimientos afines a la estadística descriptiva.

La experiencia fue desarrollada por las autoras, integrantes de la cátedra de “Estadística A”, materia correspondiente al primer cuatrimestre del segundo año de las carreras “Ingeniería Mecánica e Industrial”.

La misma se llevó a cabo en forma simultánea al desarrollo del tema abordado por el cuerpo docente de la cátedra con la participación voluntaria de un grupo de alumnos, en el espacio correspondiente a la práctica de la materia en cuestión, utilizando el Laboratorio de Matemática de la Universidad Nacional del Sur.

OBJETIVOS

- Incorporar el uso y las aplicaciones del software *R* desarrollando nuevas estrategias y habilidades.
- Valorar la precisión de los gráficos obtenidos a través de *R*, posibilitando interpretaciones más precisas de las distintas situaciones problemáticas.
- Desarrollar en los alumnos una actitud crítica de los contenidos trabajados en *R*.
- Recuperar conceptos y procedimientos aprendidos con anterioridad, utilizándolos convenientemente y valorándolos como insumo importante para la gestación de nuevos aprendizajes.
- Brindar una herramienta para perfeccionar cálculos, propiciando el análisis de los datos.

¿POR QUÉ FUE ELEGIDO EL SOFTWARE R?

El software *R* es uno de los más flexibles, potentes y profesionales que existen actualmente para realizar tareas estadísticas de todo tipo, desde las más elementales, hasta las más avanzadas. Fue creado y oficialmente presentado en 1997 por Ross Ihaka y Robert Gentleman del departamento de estadística de la Universidad de Auckland (Nueva Zelanda).

Este software es un lenguaje de programación de uso libre, de distribución gratuita y de código abierto (es decir, el código fuente del programa está disponible para los usuarios). Se rige por la licencia general pública (“General Public License” o GPL) de la fundación de software libre (“Free Software Foundation” o GNU) su desarrollo y distribución son llevados a cabo por varios estadísticos conocidos como el *Grupo Nuclear de Desarrollo de R*. Los archivos necesarios para instalar *R*, se distribuyen desde el sitio de internet *Comprehensive R Archive Network* (CRAN) <http://cran.es.r-project.org>.

Este programa fue elegido por su doble función: brinda la posibilidad de programar o utilizar sus funciones propias ya que es un programa basado sobre

comandos, en el que se puede acceder a todos los procedimientos y opciones a través de una sintaxis computacional.

R es un lenguaje Orientado a Objetos: bajo este complejo término se esconde la simplicidad y flexibilidad de R. El hecho de que R sea un lenguaje de programación puede desanimar a muchos usuarios que no tienen formación en programación, esto no es así por distintas razones:

- R es un lenguaje interpretado (como Java) y no compilado (como Fortran, Pascal, etc.), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir sentencias ejecutables.
- la sintaxis de R es simple e intuitiva para el principiante pues adquiriendo las nociones básicas puede avanzar progresivamente.
- la gran versatilidad de los procedimientos estadísticos disponibles, la capacidad de producir gráficos de calidad; estos últimos pueden ser visualizados de manera inmediata en su propia ventana y ser guardados en varios formatos (jpg, png, bmp, ps, pdf, emf, pictex, xfig) y la amplia documentación gratuita.

Todas estas características hacen de R un excelente programa para ser utilizado en docencia e investigación.

MARCO ASIMILADOR PARA EL DESARROLLO DE LA EXPERIENCIA

Los conocimientos previos de los alumnos son:

- Conceptos teóricos básicos: muestra, experimento, unidad experimental, variables, parámetro estadístico.
- Clasificación de las variables: Cualitativa- Cuantitativa; Discreta-Continua.
- Representación de datos: tablas y gráficos. Diagrama de Caja. Histograma. Polígono de frecuencia. Ojiva.
- Medidas de centralización: media aritmética, mediana y moda. Cuartiles.
- Medidas de variabilidad: rango, rango intercuartil, varianza y desvío estándar muestral. Coeficiente de variación.
- Manejo de Windows.

METODOLOGÍA

La experiencia se dividió en dos etapas: presencial y no-presencial.

La **primera etapa** se llevó a cabo en algunas clases prácticas de la materia. El espacio donde se desarrolló fue el laboratorio de computación del Departamento de

Matemática de la UNS, que cuenta con veinte computadoras y equipamiento audiovisual (cañón-proyector, pantalla desplegable) utilizado para mostrar el manejo de los comandos empleados en R.

En esta primera etapa la experiencia se centró principalmente en:

- Entrada de datos de muestras en R.
- Primeros comandos de R para el tratamiento de datos.
- Comandos para la realización de gráficos.

Para una mejor visualización, a continuación uno de los ejemplos abordados en la misma y su implementación con el software R. Queremos destacar que este es un mero ejemplo ilustrativo de los trabajos durante la experiencia, dado que también se hizo hincapié en el manejo de otros comandos

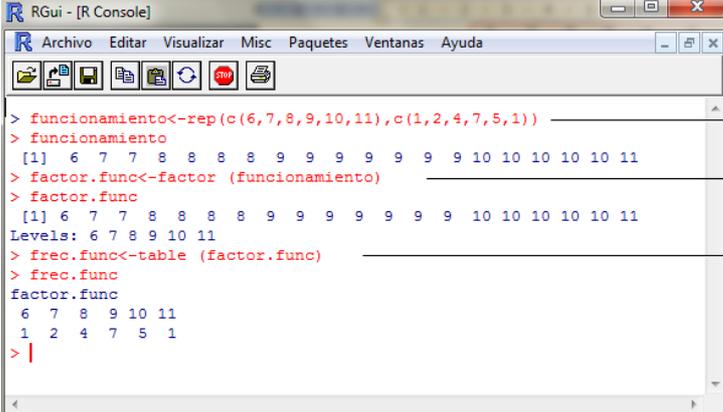
Ejemplo:

Se analiza el funcionamiento de las maquinarias de una industria. Para ello, se toma una muestra aleatoria de 20 máquinas, anotando el tiempo de trabajo (en horas) de cada una de ellas. Los resultados fueron los siguientes:

6	7	7	8	8	8	8	9	9	9
9	9	9	9	10	10	10	10	10	11

Una vez presentado el enunciado introductorio, las docentes a cargo brindaron las explicaciones necesarias para el manejo de comandos en R con la finalidad de que los alumnos realicen simultáneamente el ejemplo propuesto en las computadoras.

Entrada de datos de muestras en R



The screenshot shows the RGui console with the following code and output:

```
> funcionamiento<-rep(c(6,7,8,9,10,11),c(1,2,4,7,5,1))
> funcionamiento
[1] 6 7 7 8 8 8 8 8 9 9 9 9 9 9 9 10 10 10 10 10 11
> factor.func<-factor (funcionamiento)
> factor.func
[1] 6 7 7 8 8 8 8 8 9 9 9 9 9 9 9 9 10 10 10 10 11
Levels: 6 7 8 9 10 11
> frec.func<-table (factor.func)
> frec.func
factor.func
 6  7  8  9 10 11
 1  2  4  7  5  1
```

Annotations on the right side of the screenshot:

- Ingreso de datos en un vector
- Reordenamiento de los elementos de la muestra por medio del comando "factor"
- Elaboración de tabla de frecuencias absolutas

Primeros comandos de R para el tratamiento de datos



```
> mean(funcionamiento)
[1] 8.8
> median(funcionamiento)
[1] 9
> sd(funcionamiento)
[1] 1.239694
> var(funcionamiento)
[1] 1.536842
> |
>
```

→ **Media**
→ **Mediana**
→ **Desvío estándar**
→ **Varianza**

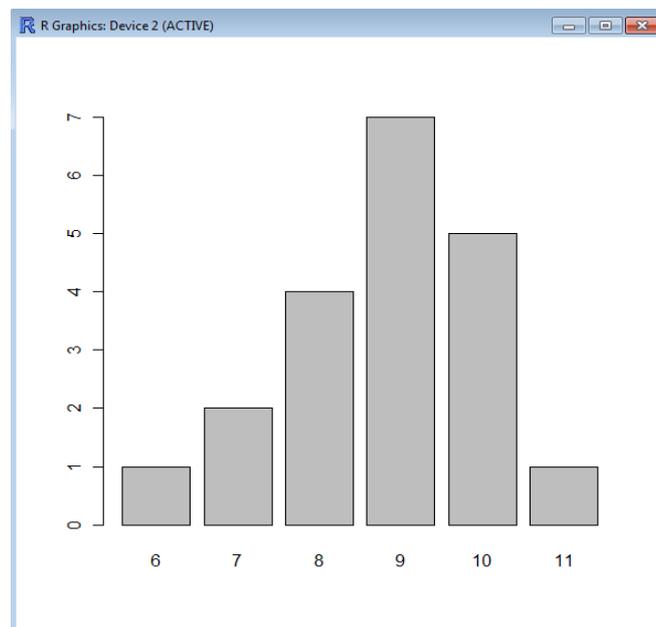
Los comandos anteriormente detallados están predefinidos en el software R y resultaron de fácil manejo para los alumnos.

Comandos para la realización de gráficos

Al ejecutarse una función gráfica, R abre una nueva ventana para mostrar el gráfico.

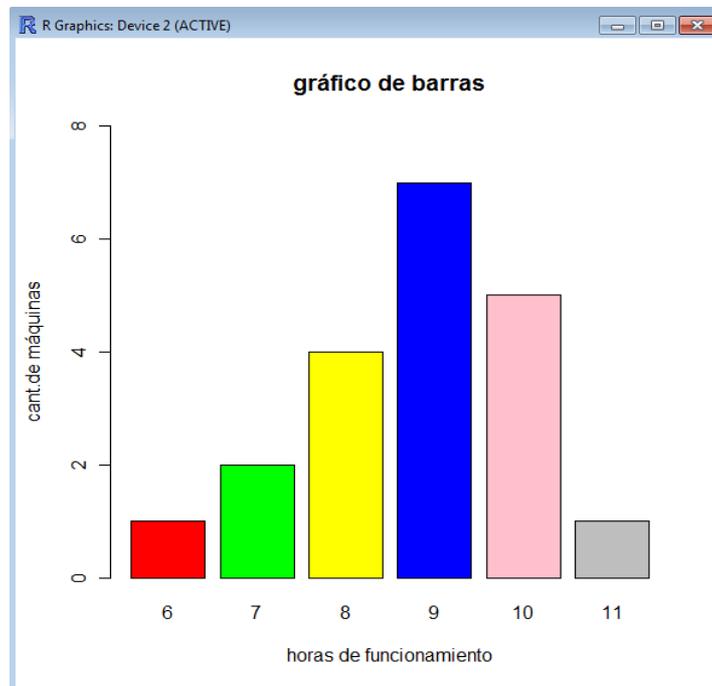
Representación de la información conocida mediante un gráfico de barras.

> **barplot (frec.func)**



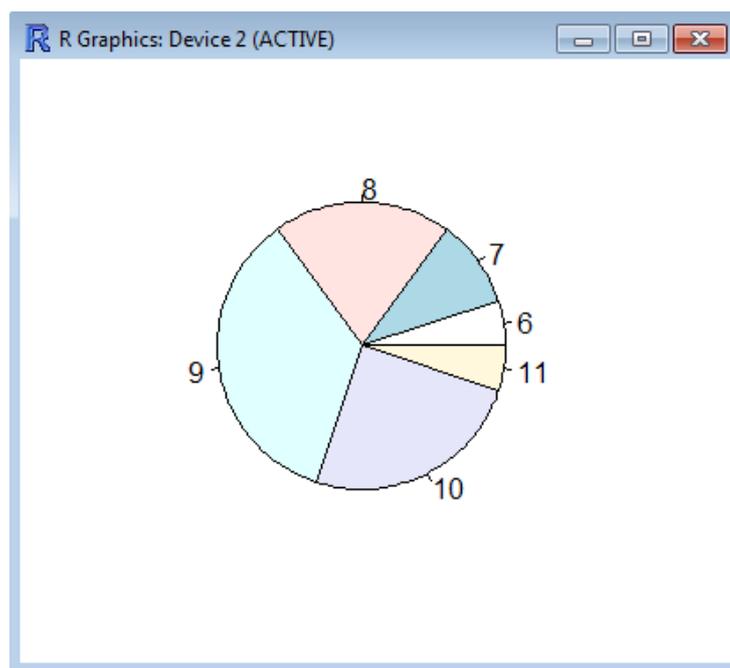
Representación de la información conocida mediante un gráfico de barras modificado según las exigencias del usuario.

> barplot (frec.func, main="gráfico de barras", ylim=c(0,8), xlab="horas de funcionamiento", ylab="cant.de máquinas", col=c("red", "green", "yellow", "blue", "pink", "grey"))



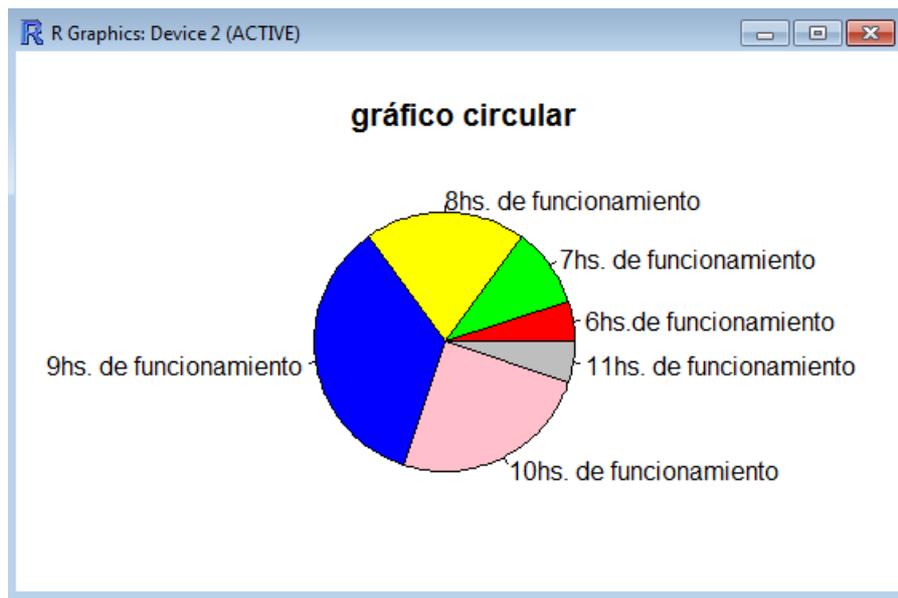
Representación de la información conocida en un gráfico circular

> pie (frec.func)



Representación de la información conocida mediante un gráfico circular modificado según las exigencias del usuario.

> pie (frec.func, main="gráfico circular", init.angle=0, labels=c("6hs.de funcionamiento","7hs. de funcionamiento","8hs. de funcionamiento","9hs. de funcionamiento","10hs. de funcionamiento","11hs. de funcionamiento"), col=c("red", "green", "yellow", "blue", "pink", "grey"))



Valores de los principales cuartiles y el rango

```
RGui - [R Console]
Archivo Editar Visualizar Misc Paquetes Ventanas Ayuda
[Icons]
> quantile(funcionamiento)
0% 25% 50% 75% 100%
6 8 9 10 11
> quantile(funcionamiento, probs=c(0.25,0.75,0.25))
25% 75% 25%
8 10 8
> diff(range(funcionamiento))
[1] 5
>
```

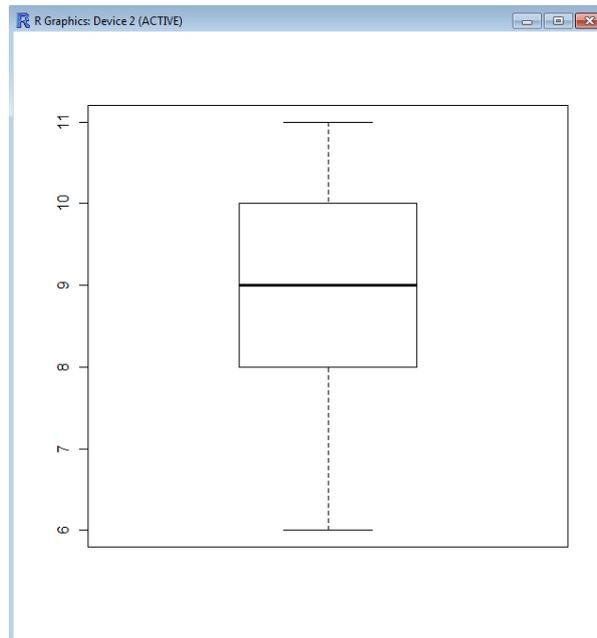
Principales cuartiles

Cuartiles buscados por el usuario

Rango

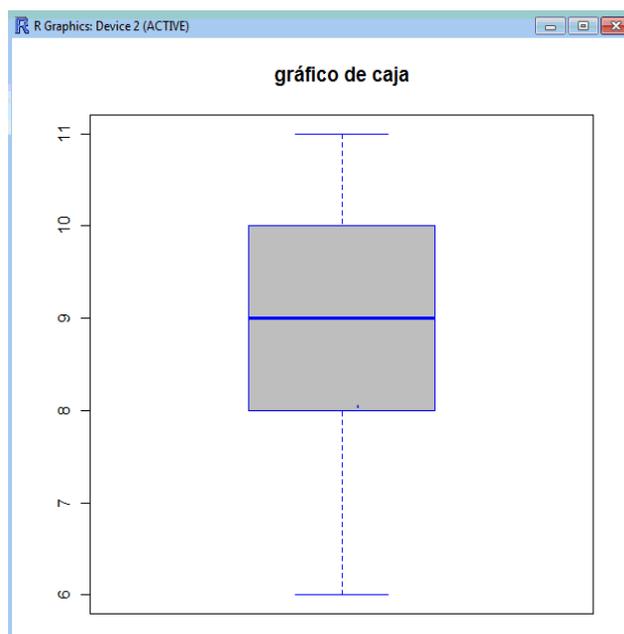
Representación de la información conocida mediante un diagrama de caja

>boxplot(funcionamiento)



Representación de la información conocida mediante un diagrama de caja modificado según las exigencias del usuario.

>boxplot (funcionamiento, main="gráfico de caja", names=c("funcionamiento de máquinas (en horas)"), border="blue", col="grey")



Los alumnos, con sus inquietudes e interrogantes, generaron un espacio de discusión donde, tanto el alumnado como las docentes, enriquecieron sus conocimientos al indagar más en el manejo de R.

Como expresa Verónica Edwards (1997), *El contenido no es independiente de la forma en la cual es presentado (...) El contenido se transforma en la forma (...) la presentación del conocimiento en formas distintas le da significaciones distintas y lo altera como tal*. En este sentido, fue tal buena la predisposición, el compromiso y la apropiación del nuevo contenido por parte de los alumnos, que propusieron nuevas alternativas para trabajar con los datos del problema seleccionando el gráfico más adecuado, según el carácter cualitativo o cuantitativo de los datos y también las medidas de centralización y dispersión que resultaban más representativas en cada una de las actividades propuestas.

Esto dio lugar a un proceso de producción matemática, emergente de un trabajo del pensamiento. Así fue como se generaron debates entre ellos, surgiendo interesantes conclusiones.

Creemos que fue muy enriquecedor dejar de ser, el plantel docente, la figura central de la escena y cooperar guiando el aprendizaje. *“La asesoría propuesta se basa en el criterio de no hacer por el grupo lo que éste pueda hacer para su aprendizaje por sí mismo”* (Santoyo, 1981).

La **segunda etapa** consistió en que los alumnos resolvieran los ejercicios propuestos en los trabajos prácticos de la cátedra utilizando R, comparando así los resultados con los ya obtenidos previamente en papel. Para dicho desarrollo se brindó un instructivo como guía, facilitando la resolución de los ejercicios abordados en la unidad Estadística Descriptiva, el cual estuvo disponible en la página web del curso.

Los alumnos presentaron sus respectivos trabajos al plantel docente, quien realizó una devolución individual de los mismos.

El rol de las docentes que preponderó en esta experiencia fue el de coordinadoras, pues, como afirma Santoyo (1981), *el coordinador no enseña sino que propicia el aprendizaje sin asumir el papel de líder o de director; intentando en todo momento que no exista la independencia sino la interdependencia entre pares*.

VENTAJAS OBSERVADAS

- Obtención rápida y precisa de resultados en forma analítica y gráfica.
- Eficiencia en actividades donde se manejan grandes volúmenes de datos favoreciendo a una perspectiva más realista del manejo de información.
- Modificación de los parámetros de las distintas actividades planteadas utilizando el software **R**.
- Resignificación de los conocimientos previos.
- Aprendizaje grupal y colaborativo entre los integrantes de la situación de enseñanza –aprendizaje propuesta.

DIFICULTADES OBSERVADAS

- Algunos alumnos tuvieron dificultades para descargar el software **R** en sus hogares.
- Escaso tiempo, por parte de los alumnos, para el desarrollo de las actividades propuestas debido a la carga horaria obligatoria en sus carreras.
- El uso no adecuado del software, puede conducir a conclusiones erróneas en el análisis de los problemas.

CONCLUSION

Comenzar el desarrollo de la experiencia con el software **R** a partir de un problema de la vida real, como se detalla en el relato de la experiencia, facilitó el aprendizaje significativo, pues los alumnos pudieron establecer relaciones trascendentes entre los contenidos trabajados con anterioridad y la información nueva que se generaba.

Teniendo como premisa que el aprendizaje grupal es, ante todo, un proceso de elaboración conjunta en el que el conocimiento no se da como algo acabado, organizamos esta experiencia de manera tal que los estudiantes fueran los principales protagonistas de ellas. Las actividades propuestas permitieron el desarrollo no sólo del pensamiento convergente y del razonamiento deductivo, sino también el inductivo.

Los trabajos recibidos reflejaron tanto el compromiso y el entusiasmo de los alumnos frente a la experiencia como el aprovechamiento del software **R**.

Explorando en la potencialidad de los recursos informáticos, en particular en los beneficios del software R, esta experiencia favoreció los cambios de registros de representación semiótica, de la expresión algebraica a la representación gráfica y viceversa, en el aprendizaje de la estadística.

Finalmente, como corolario de esta experiencia, podemos inferir que es posible seguir profundizando en los beneficios que brinda el software R en relación a la estadística.

El impacto observado a corto plazo fue brindarles otro instructivo con ejemplos del tema Distribución de Variables Aleatorias Discretas debido a la motivación del grupo.

El impacto esperado a largo plazo es que la implementación del software R coopere en la toma de decisiones, teniendo en cuenta que las principales incumbencias de la Ingeniería Industrial e Ingeniería Mecánica son realizar estudios de factibilidad, proyectar, dirigir, implementar, operar y evaluar el proceso de producción de bienes industrializados y la administración de los recursos destinados a la producción de dichos bienes.

AGRADECIMIENTOS

Agradecemos al cuerpo docente de la cátedra de Estadística “A” por el espacio cedido para llevar a cabo esta experiencia. Hacemos una mención especial para la Mg. Beatriz Marrón y la Dra. Ana Tablar por las sugerencias que hicieron a este trabajo, así como a los revisores anónimos de la UMA.

Cabe destacar que sin la participación y la buena predisposición del grupo de alumnos, esta experiencia no hubiese sido posible.

BIBLIOGRAFIA

Ausubel, D.,-Novak,J.,Hanesian, H.(1983) *Psicología educativa: Un punto de vista cognoscitivo*. 2 da edición. Trillas. México

Batanero, C. (2000). *¿Hacia dónde va la educación estadística?* *Blaix*, 15, 2-13.
<http://www.ugr.es/~batanero>

Braslavsky, C. y otros (1992). *Formación profesores*. Impacto, pasado y presente. Miño y Dávila. Buenos Aires. Cap. III.

Castillo, S. (2008). *Propuesta pedagógica basada en el Constructivismo para el uso óptimo de las TIC en la Enseñanza y aprendizaje de la Matemática*. *Revista Latinoamericana de Investigación en Matemática Educativa*, Vol 11/002, pp. 171-194.

Correa, J.C y González, N. (2002). *Gráficos estadísticos con R*. <http://cran.r-project.org>

Duval, R. (1998). *Investigaciones en Matemática educativa II. Registros de representación semiótica y funcionamiento cognitivo del pensamiento*. Université Louis Pasteur de Strasbourg. Francia. Ed. Hitt, F. Grupo editorial iberoamericana. Impreso en México.

Edwards, V. (1997). *La forma del conocimiento en el aula*. En Rockwell. *La escuela cotidiana*. Méjico. F.C.E.

Ortega Sánchez, J. (2004). *Introducción al Análisis Exploratorio de Datos*. Tercer Taller sobre enseñanza de la probabilidad y estadística. Guanajuato. CIMAT,AC, México.

Santoyo, R. (1981). *Algunas reflexiones sobre la coordinación en los grupos de aprendizaje*. Perfiles educativos.

Walpole, R. (2007). *Probabilidad y estadística para ingeniería y ciencias*. 8va. edición, Pearson Educación. Editorial Mc Graw-Hill.