

# Análisis descriptivo y exploratorio de la deserción estudiantil en la Universidad Nacional de Colombia sede Medellín \*

**Semillero R - FacCA**    *Universidad Nacional de Colombia - sede Medellín*

---

**Resumen:** En este documento se dan las instrucciones de presentación de trabajos para el Rday Medellín 2019. Este resumen debe contener i) el objetivo del trabajo; ii) el diseño metodológico usado en el trabajo; iii) hallazgos o principales resultados encontrados; y iv) las implicaciones prácticas. El resumen del trabajo debe tener máximo 250 palabras.

*Keywords:* pandoc, r markdown, knitr

---

## Introducción

La deserción estudiantil se define según el ministerio de educación nacional como “ Estado de un estudiante que de manera voluntaria o forzosa no registra matrícula por dos o más períodos académicos consecutivos del programa en el que se matriculó; y no se encuentra como graduado, o retirado por motivos disciplinarios” ([Ministerio de Educación Nacional, N.d.a](#)). Para hablar de deserción estudiantil y sus causas es pertinente conocer cómo se organiza el sistema de educación superior en Colombia, el cual esta dividido en dos niveles conocidos como pregrado y posgrado, el pregrado a su vez se divide en tres niveles de formación que son: Técnico profesional, Tecnológico y Profesional, y el posgrado se divide en Maestrías y doctorados ([Ministerio de educación Nacional, N.d.b](#)). La deserción estudiantil en educación superior en Colombia es documentada y consolidada por el ministerio de educación a través del sistema para la Prevención de la Deserción en las Instituciones de Educación Superior “SPADIES” ([Ministerio de educación Nacional, N.d.c](#)), el cual permite conocer y registrar aspectos socioeconómicos, académicos, institucionales e individuales de los estudiantes con el fin de estudiar este fenómeno. La deserción a través del SPADIES es analizada de dos maneras, por periodo académico “desertores un año después matriculados dos semestres atrás” y por cohorte; las tasas más altas registradas para los años 2015 y 2016 es para la técnica profesional ,mientras que la más baja es para el nivel universitario; Sin embargo, las cifras son alarmantes en todos los niveles de educación superior ya que la tasa de deserción por cohorte de todos los niveles de educación ronda entre un 45 % y un 57 % ([SPADIES, 2017](#)). Para el nivel universitario en Colombia se registra una tasa de deserción estudiantil por cohorte de 46,1% en el año 2015 y en Antioquia para el mismo año se registra una tasa de 47,1% ([SPADIES, 2016](#)). La tasa de deserción por perdido para el año 2015 en Colombia para el nivel universitario fue de 9,25 % y a nivel de Antioquia se registró una deserción por periodo en el nivel universitario de 9,55% ([Ministerio de Educación Nacional, 2016](#)). Antioquia es el segundo departamento con más instituciones de educación superior, contando a la fecha de 2015 con 52 instituciones, 15 de ellas de carácter oficial, 35 de carácter privado y 2 de régimen especial, oferta un total de 1.907 programas curriculares de los cuales 629 corresponden a programas universitarios, captura aproximadamente el 15% de la población de educación superior inscrita del país y

---

\* Autor de contacto: [jajaja@unal.edu.co](mailto:jajaja@unal.edu.co).

gradúa aproximadamente el 13% de estudiantes de educación superior de Colombia ([Ministerio de Educación Nacional, 2016](#)). Medellín siendo la capital de Antioquia cuenta con el Observatorio De Educación Superior De Medellín “ODES” el cual se encarga entre muchas cosas del monitoreo y seguimiento de indicadores de educación superior como lo es la deserción estudiantil ([Alcaldía de Medellín, 2019](#)). El ODES retomando información del SPADIES hace un análisis de la deserción estudiantil a nivel Medellín en el cual se registra una tasa de deserción por periodo del 11,9 % para el año 2015 ([SPADIES, 2017](#)). Dentro de la región antioqueña la Universidad Nacional de Colombia Sede Medellín ocupa el puesto 19 en tasas de deserción, registrando históricamente un índice de deserción por cohorte de 50,1 % al año 2010([Universidad Nacional de Colombia Sede Medellín. Oficina de Planeación, 2011](#)).

## Metodología

### *Recolección de Información*

Los datos para realizar el análisis descriptivo de la percepción de deserción estudiantil fueron recolectados alrededor de todo el campus universitario utilizando un cuestionario de 35 preguntas, las cuales incluían variables continuas y categóricas. La pregunta objetivo en referencia a la interrupción de los estudios universitarios es una variable categórica lo que dirige el análisis a la estimación de una proporción. Para la estimación del parámetro de interés se usó un muestreo aleatorio simple con un nivel de confianza del 95%, un límite en el error de estimación de 0,05 y una varianza de 0,25 correspondiente a una proporción de 0,5. El tamaño de la población empleado fue obtenido de la información disponible en la oficina de planeación de la sede Medellín en el cual se reportan 10.594 estudiantes de pregrado al semestre 2018-1 ([Universidad Nacional de Colombia Sede Medellín. Oficina de Planeación, 2018](#)). Para el cálculo del tamaño de muestra necesario para realizar estimaciones correctas se utilizó la aplicación Shiny del semillero de R de la escuela de estadística de la Universidad Nacional de Colombia sede Medellín ([Semillero de R. Escuela de Estadística, 2019](#)), cuyo resultado fue un tamaño de muestra mínimo de 371 estudiantes a encuestar.

### *Análisis de datos*

Para el análisis de datos se utilizaron métodos descriptivos y exploratorios de las preguntas, se realizaron distintas agrupaciones y gráficos de apoyo utilizando el programa de libre acceso R ([R Core Team, 2019](#)). Para el manejo, selección y ajustes de los datos se utilizaron los paquetes “tidyr” ([Wickham, Henry and RStudio, 2019](#)) y “dplyr” ([Wickham et al., 2019](#)). Los gráficos de dispersión y las curvas ajustadas de regresión además de ser realizadas con el paquete base se elaboraron también con el paquete “ggplot2” ([Wickham, 2016](#)). Para visualizar y evidenciar la correlación de algunas de las variables se utilizó el paquete “corrplot” ([Wei et al., 2017](#)). En algunos casos se mezclaron gráficos en una misma pantalla utilizando el paquete “gridExtra” ([Auguie and Antonov, 2017](#)). Los árboles de decisión que se constituyen como una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas ([Berlanga Silvente, Rubio Hurtado and Vilà Baños, 2013](#)), fueron realizados con el paquete “rpart” ([Therneau, Atkinson and Ripley, 2019](#)), para graficarlos se utilizó el paquete “rpart.plot” ([Miliborrow, 2019](#)). Se generaron bosques aleatorios como una alternativa de regresión no paramétrica la cual se fundamenta en la inferencia condicional utilizando el paquete “party” ([Hothorn et al., 2019](#)).

## Resultados

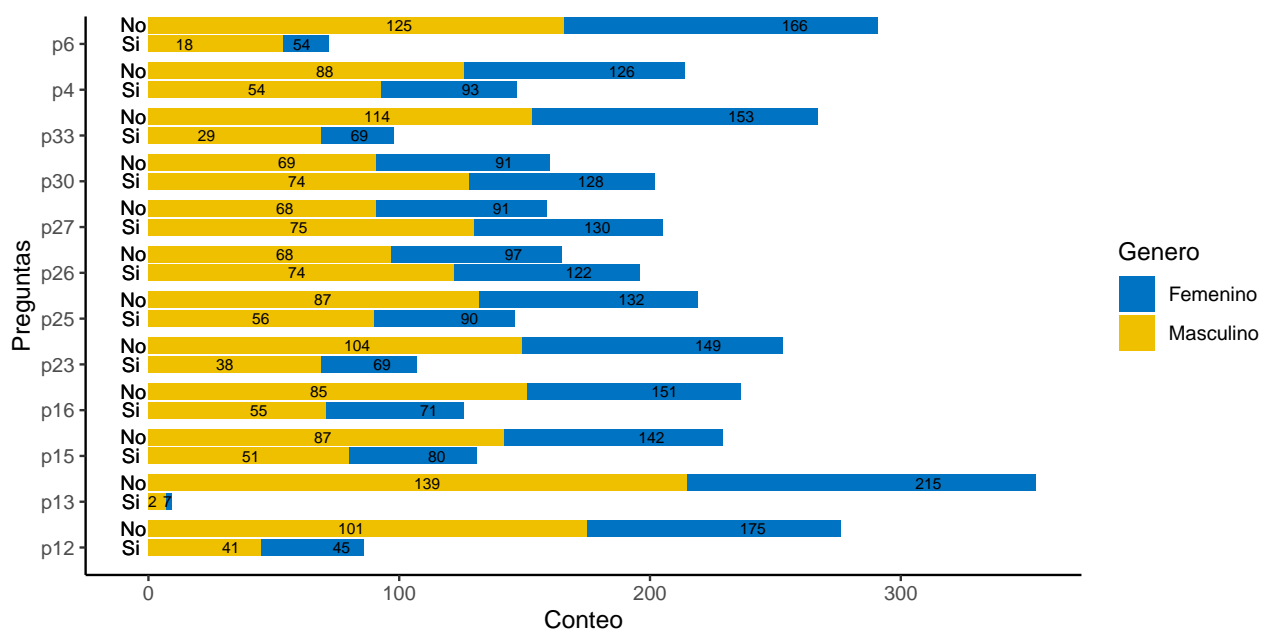


Figure 1: Frecuencia preguntas dicotómicas “si” y “no”, por genero del total de encuestados.

La figura 1 describe las preguntas dicotómicas basadas en las respuestas “si” y “no”, se realizó un conteo por respuesta y dentro de cada respuesta por genero de los encuestados.

La figura 2 describe las horas de dedicación académica por créditos inscritos, la línea verde representa las horas semanales dedicadas por el promedio de créditos calculado con los datos obtenidos, la línea amarilla representa el valor anteriormente calculado añadiendo las horas de ocio web en la semana y por ultimo la línea roja representa las horas en la semana dedicadas a estudio, ocio web y dormir para el promedio de créditos inscritos obtenido en el presente estudio.

## Elaboración de tablas

Para construir una tabla se deben tener en cuenta las siguientes recomendaciones.

- Los números van centrados siempre que tengan la misma cantidad de dígitos, de lo contrario deben ir alineados con el margen derecho del título.
- Cuando las tablas tienen datos con cifras decimales, el número de éstas debe ser igual dentro de la misma columna, pudiendo variar de columna a columna.
- Las tablas se deben nombrar en la parte superior.

La Tabla 1 mostrada a continuación se elaboró usando instrucciones usuales de LaTeX.

## Incluyendo código de R

Es posible incluir código de R en este documento para ilustrar a los lectores en la forma de usar R para realizar algún procedimiento. Se recomienda a los autores [visitar este enlace](#) para conocer

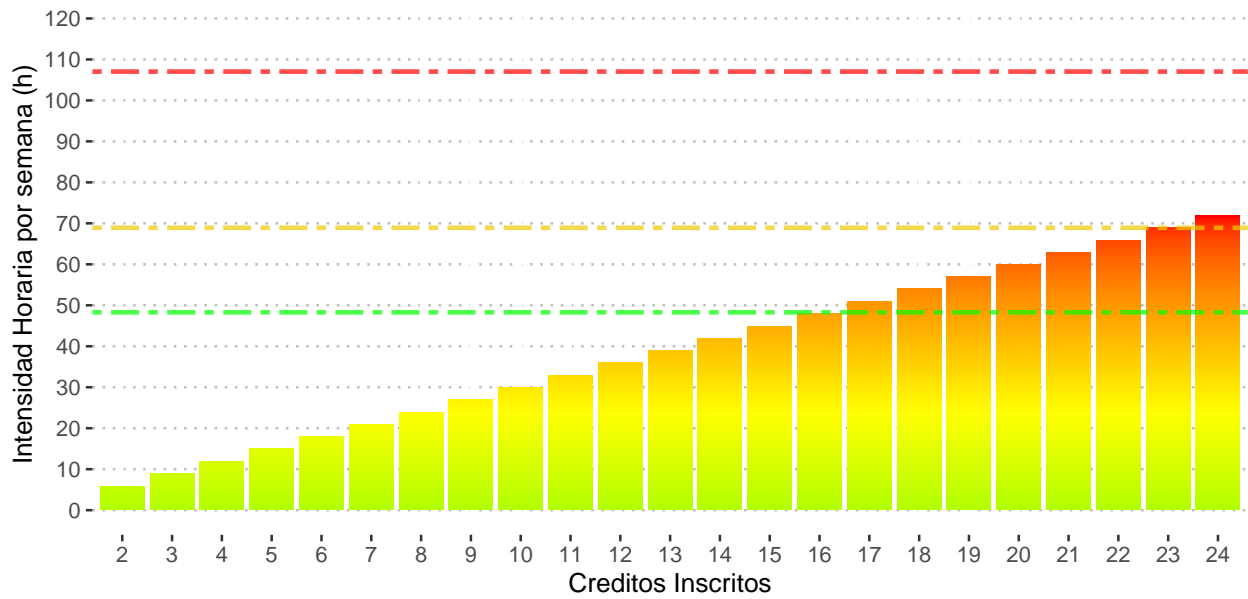


Figure 2: Intensidad horaria de creditos por semana, linea verde horas semanales del promedio de creditos inscritos, linea amarilla horas promedio de creditos mas horas de ocio web, linea roja horas promedio de creditos mas horas de ocio web y horas de sueño

Table 1: Nombre de la tabla completo.

Pais	Ciudad
Espana	Madrid
Espana	Sevilla
Francia	Paris

más detalles de como incluir código de R.

A continuación un muestra un código de R que genera cien observaciones aleatorias de una normal y luego calcula la media muestral  $\bar{x}$ .

```
x <- rnorm(n=100, mean=70, sd=5)
mean(x)
```

```
## [1] 70.60964
```

### Elaboración de figuras

Toda figura debe ir centrada y el título debe ir en la parte inferior

**No colocar el título aqui, colocarlo abajo**

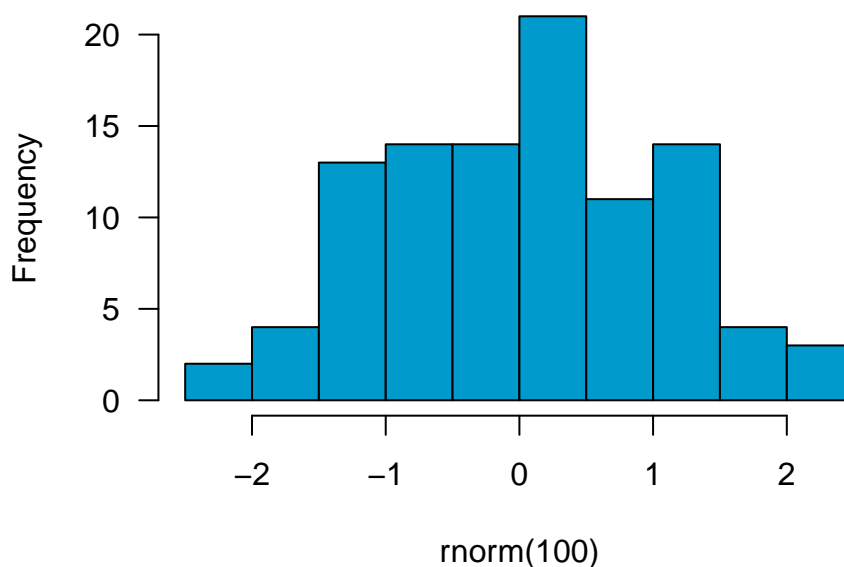


Figure 3: Histograma de valores simulados de una  $N(0, 1)$

### Ecuaciones

Para incluir ecuaciones dentro de un párrafo se usan  $\$$   $\$$  y dentro el símbolo deseado. Por ejemplo, para incluir la letra griega  $\mu$  se escribió  $\mu$ .

Para incluir ecuaciones se pueden usar  $\$$   $\$$  y dentro la ecuación. La siguiente ecuación fue hecha usando  $\mu = \frac{\theta}{\sigma}$ .

$$\mu = \frac{\theta}{\sigma}$$

Otra forma de incluir ecuaciones es de la forma usual como se hace en latex, a continuación un ejemplo usando  $\begin{equation}$  y  $\end{equation}$ . La ventaja de esta última opción es que la ecuación 1 sale numerada y se puede citar luego usando  $\ref{clave}$ .

$$E = mc^2 \quad (1)$$

### Incluyendo referencias

Todas las referencias deben estar en el archivo `master.bib` para que se puedan invocar en el trabajo. Existen dos formas de citar y son:

1. Como cita directa, es decir como [Xie \(2015\)](#). Para esto se debe escribir `@xie2013ddrk` y el archivo insertará automáticamente la referencia y la colocará en la última sección de Referencias.

## References

- Alcaldía de Medellín. 2019. "Observatorio de Educación Superior de Medellín."  
**URL:** <https://www.sapiencia.gov.co/observatorio/>
- Auguie, Baptiste and Anton Antonov. 2017. "gridExtra: Miscellaneous Functions for "Grid" Graphics." *R package gridExtra version 2.3 1*.  
**URL:** <https://cran.r-project.org/web/packages/gridExtra/index.html>
- Berlanga Silvente, Vanesa, María José Rubio Hurtado and Ruth Vilà Baños. 2013. "Cómo aplicar árboles de decisión en SPSS." *REIRE* 6:15.  
**URL:** <http://www.ub.edu/ice/reire.htm/Datosdelasautoras>
- Hothorn, Torsten, Kurt Hornik, Carolin Strobl and Achim Zeileis. 2019. "party: A Laboratory for Recursive Partytioning." *R package party version 1.3.3 1*.  
**URL:** <https://cran.r-project.org/web/packages/party/index.html>
- Milborrow, Stephen. 2019. "rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'." *R package rpart.plot version 3.0.8 1*.  
**URL:** <https://cran.r-project.org/web/packages/rpart.plot/index.html>
- Ministerio de Educación Nacional. 2016. Compendio Estadístico Educación Superior Colombiana. Technical report Ministerio de educación Nacional Bogotá, Colombia: .  
**URL:** [https://www.mineduacion.gov.co/1759/articles-360739\\_recurso.pdf](https://www.mineduacion.gov.co/1759/articles-360739_recurso.pdf)
- Ministerio de Educación Nacional. N.d.a. "Glosario - Sistemas información.".  
**URL:** <https://www.mineduacion.gov.co/sistemasdeinformacion/1735/w3-article-254707.html>
- Ministerio de educación Nacional. N.d.b. "Niveles de la Educación Superior - Ministerio de Educación Nacional de Colombia.".  
**URL:** [https://www.mineduacion.gov.co/1759/w3-article-231238.html?\\_noredirect=1](https://www.mineduacion.gov.co/1759/w3-article-231238.html?_noredirect=1)
- Ministerio de educación Nacional. N.d.c. "¿Qué es el SPADIES? - Sistemas información.".  
**URL:** <https://www.mineduacion.gov.co/sistemasdeinformacion/1735/w3-article-254648.html>
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.  
**URL:** <https://www.R-project.org/>
- Semillero de R. Escuela de Estadística. 2019. "Tamaños de muestra.".  
**URL:** <http://gauss.medellin.unal.edu.co:3838/fhernanb/samplesize/>
- SPADIES. 2016. ESTADÍSTICAS DESERCIÓN Y GRADUACIÓN 2015. Technical report Ministerio de educación Nacional Bogotá: .  
**URL:** [https://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-357549\\_recurso\\_3.pdf](https://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-357549_recurso_3.pdf)
- SPADIES. 2017. REPORTE SOBRE DESERCIÓN Y GRADUACIÓN EN EDUCACIÓN SUPERIOR AÑO 2016. Technical report Ministerio de Educación Nacional Bogotá, Colombia: .  
**URL:** [https://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-357549\\_recurso\\_5.pdf](https://www.mineduacion.gov.co/sistemasdeinformacion/1735/articles-357549_recurso_5.pdf)
- Therneau, Terry, Beth Atkinson and Brian Ripley. 2019. "rpart: Recursive Partitioning and Regression Trees." *R package rpart version 4.1-15 1*.  
**URL:** <https://cran.r-project.org/web/packages/rpart/index.html>

Universidad Nacional de Colombia Sede Medellín. Oficina de Planeación. 2011. LA DESERCIÓN ESTUDIANTEL DE LA UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN EN EL SISTEMA DE EDUCACIÓN SUPERIOR COLOMBIANO. Technical report Universidad Nacional de Colombia Medellín, Colombia: .

**URL:** <http://planeacion.medellin.unal.edu.co/images/documentos/DesercionComparadaUN-2011.pdf>

Universidad Nacional de Colombia Sede Medellín. Oficina de Planeación. 2018. "Reporte UN: Distribución de la población estudiantil por región (2018-01).".

**URL:** <https://public.tableau.com/profile/reportes.unmed#!/vizhome/Distribuciondelapoblacionestudiantilporregion2018-01/Distribuciondelapoblacionestudiantilporregion2018-01> <http://planeacion.medellin.unal.edu.co/distribucion-de-la-poblacion-estudiantil-por-region-201>

Wei, Taiyun, Viliam Simko, Michael Levy, Yihui Xie, Yan Jin and Jeff Zemla. 2017. "corrplot: Visualization of a Correlation Matrix." *R package corrplot version 0.84 1*.

**URL:** <https://cran.r-project.org/web/packages/corrplot/index.html>

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Second edi ed. Springer - Verlag New York.

Wickham, Hadley, Lionel Henry and RStudio. 2019. "tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions." *R package tidyr version 0.8.3 1*.

**URL:** <https://cran.r-project.org/web/packages/tidyr/index.html#targetText=tidyr%3A Easily Tidy Data with,with 'dplyr' data pipelines>.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller and RStudio. 2019. "dplyr: A Grammar of Data Manipulation." *R package dplyr version 0.8.3 1*.

**URL:** <https://cran.r-project.org/web/packages/dplyr/index.html>

Xie, Yihui. 2015. *Dynamic Documents with R and knitr*. 2nd ed. Boca Raton, Florida: Chapman and Hall/CRC. ISBN 978-1498716963.

**URL:** <https://yihui.name/knitr/>