

# Fundamentos de estadística

Juan F. Venegas Gutiérrez

2022-08-04



# Contents

<b>Motivación</b>	<b>5</b>
<b>1 Introducción</b>	<b>7</b>
1.1 Como esta organizado este libro . . . . .	7
1.2 Un breve recorrido por R y R Studio . . . . .	8
<b>2 Probabilidades</b>	<b>13</b>
2.1 Probabilidad 1 . . . . .	13
2.2 Probabilidad 2 . . . . .	13
2.3 Probabilidad n . . . . .	13
<b>3 Distribuciones de Probabilidad</b>	<b>15</b>
<b>4 Distribuciones de Probabilidad discretas</b>	<b>17</b>
<b>5 Distribuciones de Probabilidad Continuas</b>	<b>19</b>
<b>6 Distribuciones de Muestreo</b>	<b>21</b>
<b>7 Estimaciones con una y dos muestras</b>	<b>23</b>



# Motivación

El siguiente libro se encuentra enfocado en la aplicación de la estadística como herramienta transversal, ya que esta cuenta con técnicas transversales y que son utilizadas en distintas áreas de conocimiento. Además se contempla el uso del software R, ya que es un software de cómputo estadístico con alta demanda en distintas industrias así como en la academia, sin embargo se limita su uso a la lectura de datos e información, tratamiento y visualización de estos.

Hoy en día la automatización de procesos así como la llegada masiva del teletrabajo, a supuesto mayores desafíos para las actividades productivas, impulsando a las industrias así como a los centros de educación a adoptar más temprano que tarde una transformación tecnológica en diversas actividades que antes se realizaban de forma manual y presencial. Debido a lo anterior, la aplicación de la estadística en las distintas áreas del desarrollo humano requiere una perspectiva distinta. Es por esto que en el libro no solo encontrarás las técnicas estadísticas clásicas sino que se abordan algunas nuevas, mediante el uso del software R.



# Chapter 1

## Introducción

El término estadística hace referencia a la “Persona del estado” ampliamente utilizado desde hace años, debido a que durante décadas fue el estado el único que podía realizar estudios y análisis sociales, demográficos y económicos para el desarrollo de políticas públicas. Sin embargo desde los años 80 y de ahí en adelante la computación e informática han permitido que no solo el estado sea quien pueda desarrollar estos estudios, permitiendo además minimizar los tiempos de recolección de información, gastos y costos de almacenamiento así como de procesamiento.

De esta forma es que el término estadística ha evolucionado hacia un concepto más amplio acerca de las técnicas, métodos y herramientas que puede utilizar, incluso habilitando el considerar como objeto de estudio áreas que por su heterogeneidad décadas atrás habría sido imposible. En este contexto, este libro introduce al uso de aplicaciones estadísticas mediante la integración del software R y R Studio, ya que es uno de los con mayor demanda en el mundo así como Python, Julia, SPSS, entre otros pero que no se consideran en este libro.

### 1.1 Como esta organizado este libro

Cada de una de las secciones de este libro están desarrolladas de forma que el lector vaya conociendo desde las técnicas fundamentales hasta aquellas que son más avanzadas y que en algunos casos es necesario ir integrando lo leído para comprender aquellas que puedan tener una mayor complejidad, sin embargo es posible que si deseas volver a revisar alguna técnica en particular no sea estrictamente necesario que tengas que revisar cada una de las secciones.

- **Cuerpo Teórico:** En el libro se abordan cada una de las técnicas estadísticas inicialmente desde un punto de vista teórico pero pragmático, en este

sentido no se desarrollan todas las demostraciones necesarias para concluir algunas de las formulas que se expresan. En algunas secciones solo cuando se hace necesario se integran tanto formulas como el tratamiento de datos por software.

- **Visualización:** Si bien, la visualización de datos y/o estructuras de información es un campo bastante amplio para cubrir en este libro, se hace uso algunas librerías de visualización para la representación de diagramas de dispersión, de cajas, entre otros cuando sea necesario representar gráficamente aspectos o patrones que no son reconocibles al inspeccionar un set de datos.
- **Software:** A pesar de que este documento, no es un libro acerca de programación y/o estructura de datos. Se hace fundamental el uso de un software que permita resolver numericamente los problemas que se exponen en el libro, ya que este nos permite aplicar los conceptos y técnicas estadísticas en un entorno real, pudiendo incluso dentro de él realizar visualizaciones. Cabe destacar que el software R y R Studio incluso nos permite desarrollar informes, sin embargo es un tópico que no trataremos en este libro debido a la extensión del mismo.

## 1.2 Un breve recorrido por R y R Studio

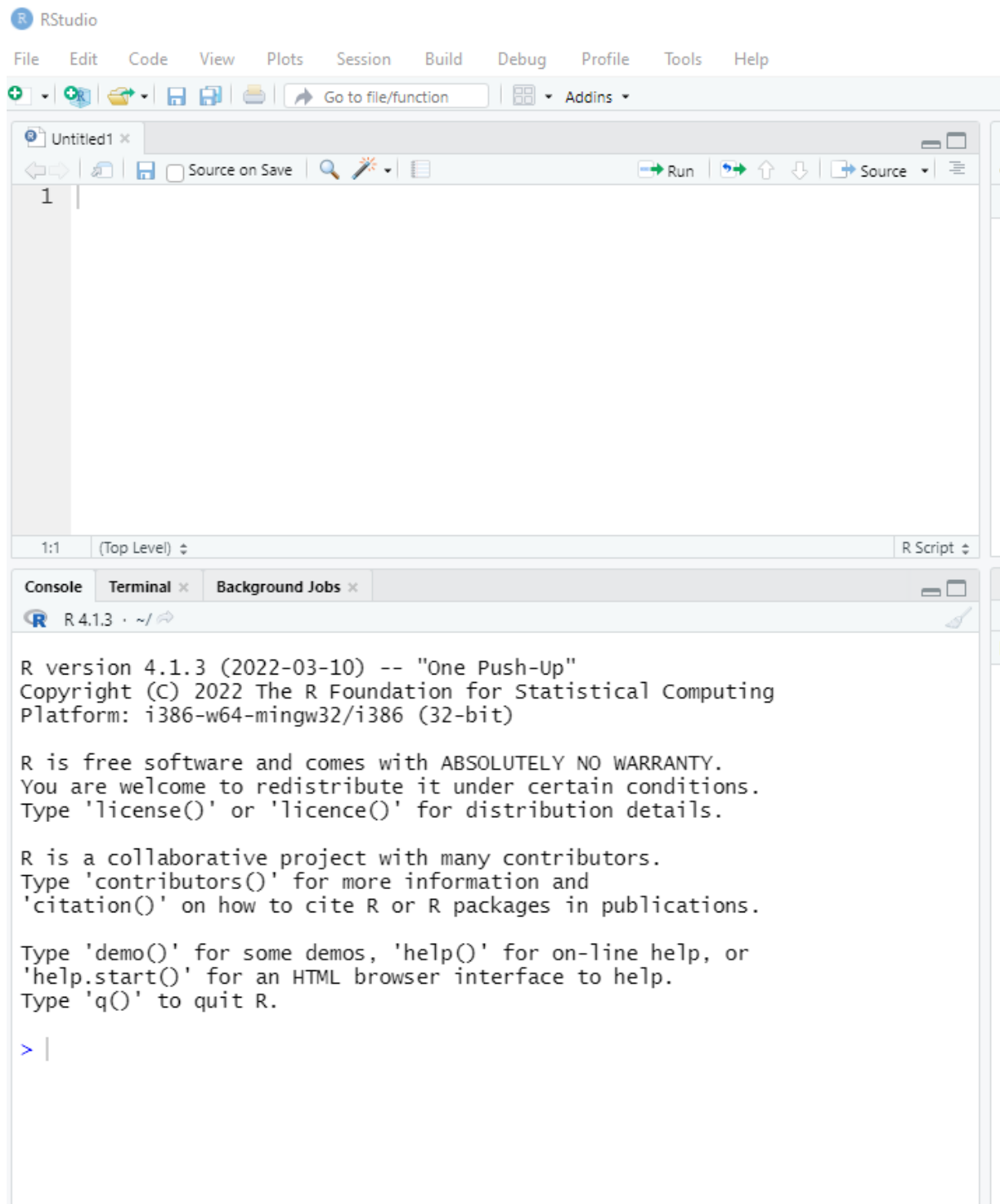
Tal vez es necesario una clara distinción entre los software R y R Studio, sobre todo porque en estos días la compañía R Studio a pasado a llamarse Posit, lo cual es mas que un simple cambio de nombre, mas adelante explicaremos brevemente el porque de la distinción.

R es un lenguaje de cómputo estadístico y visualización desarrollado en la Universidad de Auckland en 1993, sin embargo hasta el día de hoy se sigue utilizando y sobretodo en los ultimos años a tenido una mayor demanda tanto en la industria como en la academia debido a su carácter de Open Source, además de poseer una gran cantidad de librerías que permiten extender las capacidades básicas del mismo sin ningún costo. Es así que con R es posible realizar desde operaciones aritméticas sencillas, implementar técnicas estadísticas como modelos de regresión, análisis de series temporales, desarrollo de gráficos interactivos e incluso algoritmos de machine learning.

En cambio R Studio es un Entorno de Desarrollo Integrado (IDE), que brinda al software R de una interfaz visual en donde es posible gestionar en una única ventana los distintos documentos, scripts, set de datos y/o visualizaciones, esto debido a que la interfaz se compone de varios paneles que organizan cada uno de estos de forma óptima. Junto a lo anterior, R Studio también permite el uso de software de terceros como Github, Spark, conexiones a bases de datos, entre otros. Estos últimos aspectos no se tratarán en este libro, sin embargo son utilizados de forma intensiva en conjunto.



A continuación se muestra una imagen de la ventana principal de R Studio.



En el primer panel se ubica la ventana de script, en donde es posible escribir aquellas instrucciones secuenciales que queremos que se ejecuten sean operaciones aritméticas, definición de variables y gráficos. Luego el segundo panel se ubica la consola junto a otros sub-paneles en donde es posible ejecutar instrucciones una a una dentro de otras fuera del entorno de R, además es la ventana original que brinda el software R sin el uso del IDE de R Studio. En el tercer panel se ubica la ventana en donde se almacenan las variables así como otras estructuras de datos, a las cuales se puede acceder directamente desde aquí. Finalmente en la cuarta ventana se ubica el directorio de trabajo así como otros sub-paneles en donde se puede acceder a la ayuda de R, gráficos desarrollados, entre otros.



## Chapter 2

# Probabilidades

Aca comenzamos con probabilidad

### 2.1 Probabilidad 1

...

### 2.2 Probabilidad 2

### 2.3 Probabilidad n



## Chapter 3

# Distribuciones de Probabilidad

Now I'll teach you some crazy math, but I need to work it out first...





## Chapter 4

# Distribuciones de Probabilidad discretas

Now I'll teach you some crazy math, but I need to work it out first...



## Chapter 5

# Distribuciones de Probabilidad Continuas

Now I'll teach you some crazy math, but I need to work it out first...



## Chapter 6

# Distribuciones de Muestreo

Now I'll teach you some crazy math, but I need to work it out first...



## Chapter 7

# Estimaciones con una y dos muestras

Now I'll teach you some crazy math, but I need to work it out first