

**Universidad Rey Juan Carlos**

MÁSTER UNIVERSITARIO EN INGENIERÍA DE SISTEMAS DE DECISIÓN

# CASO PRÁCTICO 1: ESTADÍSTICA DESCRIPTIVA

**Autor:**

Elisa Cascudo

**Asignatura:**

Modelización y Tratamiento de la Incertidumbre

13 Octubre 2025

## Índice general

<b>1</b>	<b>Introducción</b>	<b>3</b>
1.1	Base de datos 1: origen y descripción . . . . .	3
1.1.1	Descripción de las variables . . . . .	4
1.2	Base de datos 2: origen y descripción . . . . .	4
1.2.1	Descripción de las variables . . . . .	5
<b>2</b>	<b>Tratamiento estadístico descriptivo</b>	<b>6</b>
2.1	Variables cuantitativas discretas . . . . .	6
2.1.1	Base de datos 1: <i>Hábitos de estudiantes vs Rendimiento académico</i> . . . . .	6
2.1.2	Base de datos 2: <i>Admisión en Máster 2</i> . . . . .	8
2.2	Variables cuantitativas continuas . . . . .	10
<b>3</b>	<b>Conclusiones</b>	<b>13</b>

## Índice de figuras

2.1	Gráficos de barras que muestran la distribución de frecuencia de las tres variables cuantitativas discretas presentes en la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> . Elaboración propia. . . . .	7
2.2	Diagrama de cajas de las variables continuas de la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> : edad, frecuencia de deporte e índice de salud mental. Elaboración propia. . . . .	7
2.3	Gráfico de barras que muestra la distribución de frecuencia de la variable cuantitativa discreta <i>Ranking Universidad</i> la base de datos <i>Admisión en Máster 2</i> . Elaboración propia. . . . .	9
2.4	Diagrama de cajas de la variable cuantitativa discreta <i>Ranking Universidad</i> la base de datos <i>Admisión en Máster 2</i> . Elaboración propia. . . . .	9
2.5	Histogramas que muestran la distribución de frecuencia de las cinco variables cuantitativas continuas presentes en la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> . Elaboración propia. . . . .	11
2.6	Diagrama de cajas de las variables cuantitativas continuas presentes en la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> , concretamente horas de estudio al día, horas en redes sociales al día, horas en Netflix diarias. Elaboración propia. . . . .	12
2.7	Diagrama de cajas de las variables cuantitativas continuas presentes en la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> , concretamente porcentaje de presencia- lidad y puntuación en examen . . . . .	12

## Índice de cuadros

1.1	Variables del conjunto de datos: <i>Hábitos de estudiantes vs Rendimiento Académico</i> . Elaboración propia. . . . .	4
1.2	Variables del conjunto de datos: <i>Admisión en Máster 2</i> . Elaboración propia. . . . .	5
2.1	Estadísticos descriptivos de variables cuantitativas discretas de la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> . Elaboración propia. . . . .	7
2.2	Estadísticos descriptivos de la variable cuantitativa discreta <i>Ranking Universidad</i> de la base de datos <i>Admisión en Máster 2</i> . Elaboración propia. . . . .	9
2.3	Estadísticos descriptivos de variables cuantitativas continuas de la base de datos <i>Hábitos de estudiantes vs Rendimiento académico</i> . Elaboración propia. . . . .	11

# Capítulo 1

## Introducción

El objetivo de este caso práctico es ilustrar el procedimiento de un análisis estadístico descriptivo, el cual permite extraer, sintetizar y ordenar la información de una base de datos de manera significativa. De esta manera se pueden obtener conclusiones iniciales sobre los datos disponibles, proporcionando una visión general de los datos y una base para el resto de análisis que se lleven a cabo, en los cuales no nos adentraremos en este informe.

Para ello, emplearemos dos bases de datos diferentes que se describirán con más detalle en el apartado siguiente. La primera y principal, *Hábitos de los Estudiantes vs Rendimiento Académico*, contiene información sobre la relación entre los hábitos de estudiantes y su rendimiento académico. En ella se analizarán tanto variables discretas como continuas. Sin embargo, al examinar los datos cuantitativos discretos, como observaremos más adelante, hay una excesiva uniformidad de los resultados que no nos permite sacar conclusiones relevantes. Por esta razón, y aunque su uso podría resultar relevante en un análisis descriptivo real, se ha decidido incorporar una segunda base de datos con variables cuantitativas discretas que resultan en un análisis más relevante para este caso de estudio. Esta segunda fuente, *Graduate Admission 2*, permitirá ilustrar con mayor representatividad la utilidad e importancia del análisis estadístico descriptivo en variables discretas.

El siguiente paso será el análisis estadístico descriptivo de las variables cuantitativas continuas. En él se empleará exclusivamente la primera base de datos, *Hábitos de los Estudiantes vs Rendimiento Académico*. Una vez finalizado, presentaremos las conclusiones derivadas de aplicar un análisis estadístico descriptivos sobre los datos elegidos, lo cual ejemplificará la utilidad de estas herramientas.

### 1.1. Base de datos 1: origen y descripción

La base de datos principal de este caso de estudio se titula *Student Habits vs Academic Performance* [*Hábitos de los Estudiantes vs Rendimiento Académico*]. Esta base de datos se ha extraído de Kaggle, una plataforma en línea, actualmente propiedad de Google, que sirve como centro de encuentro para científicos de datos y profesionales del aprendizaje automático, donde pueden aprender, colaborar y competir en la resolución de desafíos reales de ciencia de datos. Ofrece un amplio repositorio de conjuntos de datos gratuitos, un entorno colaborativo para compartir código y proyectos en cuadernos basados en web, y una destacada comunidad donde los usuarios pueden aprender unos de otros a través de foros de discusión y cursos gratuitos.

Los datos se han extraído de [urlhttps://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance](https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance)<sup>1</sup>, descargado el 25 de septiembre de 2025 a las 11:00 horas.

---

<sup>1</sup>Copyright 2025 Jayanta Nath. Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0> Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Se trata de un conjunto de datos simulado con 1000 registros de estudiantes que explora la relación entre hábitos de estilo de vida (horas de estudio, sueño, uso de redes sociales, dieta, salud mental) y rendimiento académico.

### 1.1.1. Descripción de las variables

La base de datos original está en inglés. Para mayor claridad, se presentan las variables con su nomenclatura traducida al español, manteniendo entre paréntesis el nombre original cuando sea pertinente.

Cuadro 1.1: Variables del conjunto de datos: *Hábitos de estudiantes vs Rendimiento Académico*. Elaboración propia.

Variable original	Traducción	Descripción	Tipo
student_id	ID_estudiante	Identificador único por estudiante	Cuantitativa discreta
age	edad	Edad de cada estudiante	Cuantitativa discreta
gender	género	Género de cada estudiante	Cualitativa
study_hours_per_day	horas_estudio_día	Tiempo promedio diario de estudio en horas	Cuantitativa continua
social_media_hours	horas_rrss	Tiempo diario en redes sociales	Cuantitativa continua
netflix_hours	horas_netflix	Tiempo promedio diario en Netflix/series	Cuantitativa continua
part_time_job	trabajo_tiempo_parcial	Trabajo a tiempo parcial (Sí/No)	Booleana
attendance_percentage	porcentaje_presencial	Asistencia a clases (0–100 %)	Cuantitativa continua
sleep_hours	horas_sueño	Promedio diario de horas de sueño	Cuantitativa continua
diet_quality	calidad_dieta	Calidad de la dieta diaria de los estudiantes	Booleana
exercise_frequency	frecuencia_deporte	Sesiones promedio de ejercicio semanal	Cuantitativa discreta
parental_education_level	nivel_educación_padres	Nivel educativo de los padres	Cualitativa
internet_quality	calidad_internet	Calidad del internet en casa	Cualitativa
mental_health_rating	índice_salud_mental	Índice de salud mental (0 = más baja, 10 = más alta)	Cuantitativa discreta
extracurricular_participation	participación_extracurricular	Participación en actividades extracurriculares	Booleana
exam_score	puntuación_examen	Puntos en el examen final (0–100 %)	Cuantitativa continua

## 1.2. Base de datos 2: origen y descripción

La segunda base de datos, *Graduate Admission 2* [Admisión en Máster 2] se obtuvo igualmente de Kaggle, en <https://www.kaggle.com/datasets/mohansacharya/graduate-admissions/data><sup>2</sup>, descargada el 29 de septiembre de 2025 a las 18 horas. Esta base de datos está diseñada para analizar y predecir las probabilidades de admisión de estudiantes a programas de maestría, con enfoque en el contexto educativo de la India. Contiene información sobre los factores clave que las universidades consideran al evaluar solicitudes de posgrado, como calificaciones de exámenes estandarizados, promedios académicos y otros criterios relevantes. La base de datos se inspiró en el conjunto de datos de admisiones de posgrado de la Universidad de California en Los Ángeles (UCLA).

<sup>2</sup>Mohan S Acharya, Asfia Armaan, Aneeta S Antony : A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019

### 1.2.1. Descripción de las variables

Tal y como ocurre con la base de datos *Hábitos de estudiantes vs Rendimiento académico*, la base de datos original de *Admisión en Máster 2* está en inglés. Para mayor claridad, se presentan las variables con su nomenclatura traducida al español, manteniendo entre paréntesis el nombre original cuando sea pertinente.

Cuadro 1.2: Variables del conjunto de datos: *Admisión en Máster 2*. Elaboración propia.

Variable original	Traducción	Descripción	Tipo
Serial No.	ID	Número identificador de cada estudiante	Cuantitativa discreta
GRE Score	Puntuación GRE	Puntaje GRE (máximo 340)	Cuantitativa discreta
TOEFL Score	Puntuación TOEFL	Puntaje TOEFL (máximo 120)	Cuantitativa discreta
University Rating	Ranking Universidad	Valoración de la universidad (máximo 5)	Cuantitativa discreta
SOP	SOP	Calidad del Statement of Purpose (máximo 5)	Cuantitativa continua
LOR	LOR	Calidad de las cartas de recomendación (máximo 5)	Cuantitativa continua
CGPA	CGPA	Promedio de notas de licenciatura (máximo 10)	Cuantitativa continua
Research	Investigación	Experiencia en investigación (0 = No, 1 = Sí)	Booleana
Chance of Admit	Probabilidad de admisión	Probabilidad de ser admitido (rango de 0 a 1)	Cuantitativa continua

## Capítulo 2

# Tratamiento estadístico descriptivo

### 2.1. Variables cuantitativas discretas

#### 2.1.1. Base de datos 1: *Hábitos de estudiantes vs Rendimiento académico*

Como se observa en el Cuadro 2.1, las variables continuas discretas de la base de datos *Hábitos de estudiantes vs Rendimiento académico* presentan una asimetría casi nula y una curtosis mínima. Esta distribución se refleja también en los gráficos de barras de la Figura 2.1, donde se aprecia que la frecuencia de los datos es bastante uniforme. Esta uniformidad podría tener como consecuencia la falta de información significativa sobre la muestra analizada.

En el caso específico de la variable edad, los valores se encuentran entre los 17 y 24 años, un rango típico para estudiantes universitarios. En el gráfico de barras correspondiente de la Figura 2.1, se observa un pico en los 20 años, que coincide con el valor de la mediana presente en el Cuadro 2.1. Además, el rango intercuartílico (IQR) es compacto y la distribución estándar baja, mostrando homogeneidad entre edades. Su diagrama de cajas de la Figura 2.2 muestra una mayor dispersión en los valores por encima de la mediana. La base de datos que estamos analizando pretende representar datos sobre estudiantes universitarios. La dispersión asimétrica en la edad universitaria ocurre porque existe un límite mínimo natural de ingreso alrededor de los diecisiete o dieciocho años, pero no hay límite superior, lo que permite la presencia de estudiantes mayores que retoman estudios, cursan segundas carreras o tienen retrasos académicos. La mayoría de estudiantes se concentra en edades típicas de dieciocho a veinte años creando una distribución compacta, mientras que los estudiantes mayores muestran mayor variabilidad y extienden la cola superior del diagrama de cajas, reflejando una asimetría natural en las poblaciones estudiantiles.

Un patrón similar se observa en las variables frecuencia de práctica deportiva e índice de salud mental, las cuales presentan una distribución uniforme en sus respectivos gráficos de barras de la Figura 2.1. Esto se confirma con coeficientes de asimetría prácticamente nulos en el Cuadro 2.1, y diagramas de cajas en la Figura 2.2 que muestran un IQR compacto y desviación estándar baja. Esta uniformidad dificulta la extracción de conclusiones significativas sobre la muestra. Las frecuencias de los valores de la variable *frecuencia\_deporte* son similares entre sí, sin evidenciar una tendencia general clara dentro de la muestra. La misma observación aplica para el Índice de Salud Mental. Es posible que esta ausencia de variabilidad se deba a que los datos fueron generados de manera artificial.

Dado que los resultados obtenidos hasta este punto presentan limitaciones para evidenciar plenamente el potencial del análisis estadístico descriptivo aplicado a variables continuas, se procederá a examinar una de las variables continuas contenidas en una segunda base de datos denominada *Admisión en Máster 2*. Este análisis complementario permitirá demostrar con mayor profundidad de qué manera la integración de técnicas gráficas y estadísticas aplicadas a variables continuas puede revelar información sustantiva acerca de las características poblacionales.

Cuadro 2.1: Estadísticos descriptivos de variables cuantitativas discretas de la base de datos *Hábitos de estudiantes vs Rendimiento académico*. Elaboración propia.

Variable	Mediana	Media	Moda	IQR	D2	D9	Q25	Q75	Quasi Var	Var	SD	Asim.	Curt.
edad	20,00	20,50	20	4,25	18,00	24,00	18,75	23,00	5,327	5,321	2,308	0,008	-1,222
frecuencia_deporte	3,00	3,04	3,00	4,00	1,00	6,00	1,00	5,00	4,102	4,098	2,025	-0,031	-1,279
indice_salud_mental	5,00	5,44	4,00	5,00	3,00	9,00	3,00	8,00	8,108	8,100	2,847	0,038	-1,192

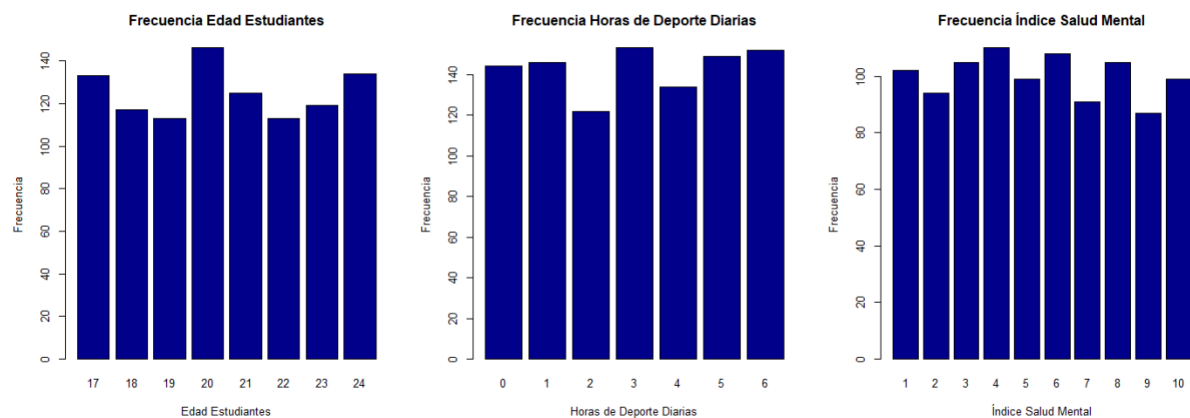


Figura 2.1: Gráficos de barras que muestran la distribución de frecuencia de las tres variables cuantitativas discretas presentes en la base de datos *Hábitos de estudiantes vs Rendimiento académico*. Elaboración propia.

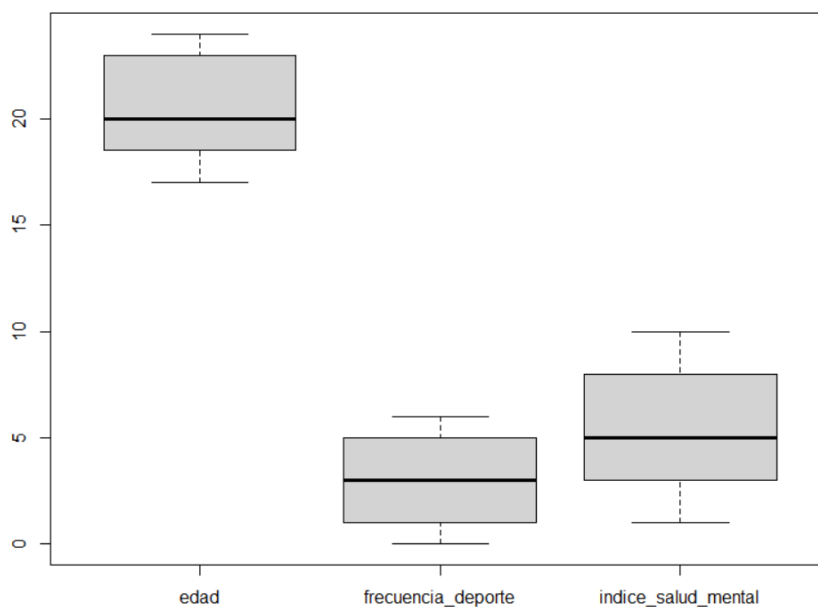


Figura 2.2: Diagrama de cajas de las variables continuas de la base de datos *Hábitos de estudiantes vs Rendimiento académico*: edad, frecuencia de deporte e índice de salud mental. Elaboración propia.



### 2.1.2. Base de datos 2: *Admisión en Máster 2*

Esta segunda base de datos, *Admisión en Máster 2*, está diseñada para analizar y predecir las probabilidades de admisión de estudiantes a programas de maestría, centrada en el contexto educativo de la India. Contiene información relevante sobre los factores que las universidades consideran al evaluar solicitudes de posgrado y se inspira en el conjunto de datos de admisiones de la Universidad de California en Los Ángeles (UCLA).

Uno de los aspectos más importantes es el ranking de la universidad de procedencia de los aspirantes, representado en este conjunto de datos por la variable *Ranking Universidades*, que está comprendido en una escala del 1 a 5, siendo 5 el nivel más alto de calidad. Para las universidades receptoras, conocer este dato permite comprender mejor el perfil de sus candidatos y diseñar estrategias de captación más efectivas. Además, la procedencia de estudiantes de universidades con un ranking más alto puede aportar prestigio a la institución. Para los estudiantes, conocer el ranking de su universidad es útil para evaluar sus posibilidades de admisión y entender el nivel de competencia al que se enfrentan. Esta información ayuda a identificar si su formación es comparable con la de otros solicitantes y a establecer expectativas realistas sobre el proceso de selección en universidades indias.

Observando el diagrama de cajas de la Figura 2.4, aunque sugiere una distribución equilibrada entre estudiantes de universidades de mayor y menor ranking, el gráfico de barras de la Figura 2.3 revela que la cantidad de aspirantes provenientes de universidades con ranking 1 es considerablemente menor que la de aquellos provenientes de universidades de ranking 2 y 3, siendo el ranking 3 el más frecuente. Esto último se ve reflejado también en el Cuadro 2.2, según el cual la moda del ranking de las universidades de origen se sitúa en la posición 3, lo que indica una procedencia intermedia de los aspirantes al posgrado.

Una posible explicación de estos resultados es que tanto las universidades de excelencia como las de bajo nivel son menos comunes que las que ofrecen un nivel medio de educación. Por lo tanto, la mayoría de los aspirantes proviene de instituciones con un ranking intermedio. Además, esto puede indicar que la universidad receptora tiene una reputación de dificultad y exigencia moderada, atrayendo a estudiantes de universidades con rankings más bajos.

Para un estudiante, esta información es fundamental para comprender su posición competitiva en el proceso de admisión. Por ejemplo, un alumno procedente de una universidad ubicada en el tercer nivel del ranking puede anticipar que sus probabilidades de ingresar a un programa de maestría en la universidad receptora son bajas debido al alto nivel de competencia. En cambio, un estudiante proveniente de una universidad con ranking más bajo, como el quinto nivel, podría tener mayores facilidades para ser admitido, ya que la competencia en ese segmento suele ser menos intensa.

Cuadro 2.2: Estadísticos descriptivos de la variable cuantitativa discreta *Ranking Universidad* de la base de datos *Admisión en Máster 2*. Elaboración propia.

Variable	Mediana	Media	Moda	IQR	D2	D9	Q25	Q75	Quasi Var	Var	SD	Asim.	Curt.
Ranking Universidad	3,00	3,087	3,00	2,00	2,00	5,00	2,00	4,00	1,308	1,305	1,144	0,169	-0,812

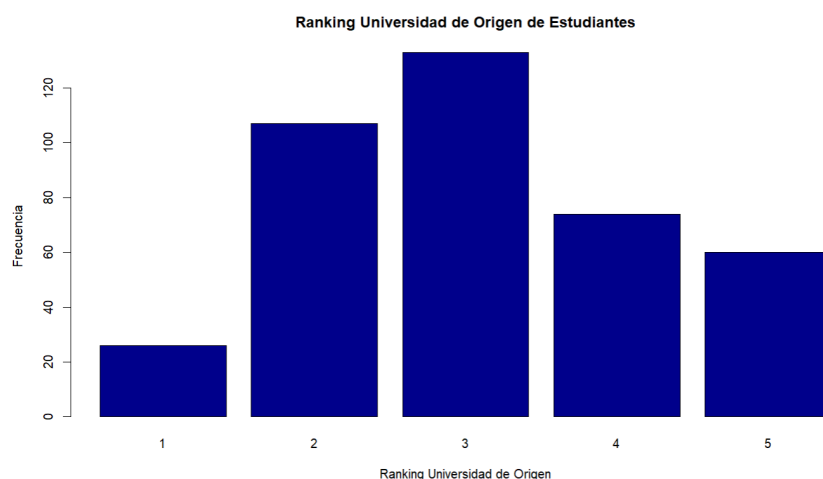


Figura 2.3: Gráfico de barras que muestra la distribución de frecuencia de la variable cuantitativa discreta *Ranking Universidad* la base de datos *Admisión en Máster 2*. Elaboración propia.

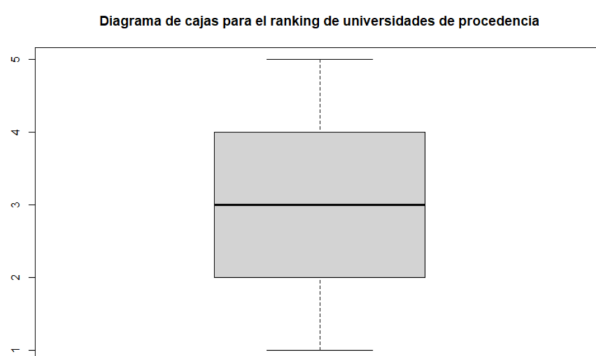


Figura 2.4: Diagrama de cajas de la variable cuantitativa discreta *Ranking Universidad* la base de datos *Admisión en Máster 2*. Elaboración propia.

## 2.2. Variables cuantitativas continuas

Procedemos ahora al análisis de las variables cuantitativas continuas de la primera base de datos, *Hábitos de estudiantes vs Rendimiento académico*.

Comencemos por el análisis de las variables *Horas en redes sociales* y *Horas en Netflix*. Estas dos son las que presentan la mayor asimetría positiva, como se observa en el Cuadro 2.3. En consecuencia, los gráficos correspondientes en la Figura 2.5 muestran una concentración significativa de los datos hacia la izquierda, es decir, un menor número de horas dedicado a las redes sociales y a Netflix, lo que podría indicar que los estudiantes, por lo general, dedican un tiempo moderado a estas actividades tradicionalmente consideradas perjudiciales para el rendimiento académico y la salud. Pero es necesario complementar esta información con las conclusiones que se derivadas de los diagramas de cajas de ambas variables, en la Figura 2.6. Vemos que hay una mayor dispersión del numero total de horas entre aquellos estudiantes que pasan un tiempo superior a la mediana tanto en redes sociales como en Netflix, llegando a aparecer valores atípicos por encima del limite superior del rango esperado. Esto indica que existen estudiantes con niveles preocupantes de adicción al uso de estas plataformas.

Para profundizar, es relevante comparar las medias de estas dos variables con *Horas de Estudio*, que según el Cuadro 2.3 son: 2.51 horas en redes sociales, 1.80 horas en Netflix y 3.55 horas diarias de estudio. Si sumamos tiempo medio dedicado a redes sociales y Netflix, 4.31 horas, se observa que los estudiantes pasan más tiempo en distracciones digitales que estudiando. Además, las *Horas de Estudio* presentan una mayor desviación estándar que las variables *Horas en redes sociales* y *Horas en Netflix*, lo que indican que todos dedican tiempo de forma similar en redes y Netflix, pero no todos estudian igual. En la Figura 2.6 se puede observar que también hay estudiantes estudiando excesivamente, con varios sobrepasando el limite superior superior del rango esperado. Es importante dedicarle tiempo el estudio, y no priorizar el entretenimiento digital sobre ello, pero un exceso de estudio puede derivar en fatiga, lo cual nos lleva a la siguiente variable a analizar, *Horas de Sueño*. En conclusión, los hábitos digitales de esta muestra se pueden considerar poco saludables, y perjudiciales al tiempo dedicado al estudio.

Según los datos obtenidos en el Cuadro 2.3, los estudiantes en general duermen menos de las siete horas diarias recomendadas, con un promedio de 6.47 horas por noche. Si nos fijamos en el diagrama de cajas de la Figura 2.6, un 50 % de lo estudiantes duerme menos horas que este promedio, que de por si es ya bastante bajo. La preferencia hacia pasar el tiempo en redes sociales y Netflix sobre el estudio, y la baja calidad de sueno promedio del estudiantado, indica que los hábitos de salud cotidianos de este grupo de estudiantes son de baja calidad, lo cual podrá tener consecuencias en su rendimiento académico.

A pesar de ello, los estudiantes muestran un porcentaje de presencialidad alto y una puntuación en el examen relativamente buena. Según el Cuadro 2.3, la media de presencialidad es 84.12 % y la media en el examen 70.50 %, ambos presentes en la base de datos como porcentajes. Estos valores no reflejan un nivel de excelencia académica, pero sí un desempeño positivo en términos generales. La ausencia de excelencia podría estar relacionada con el nivel bajo de hábitos saludables y de tiempo de estudio. Al observar los diagramas de cajas de la Figura 2.7, también se puede comprobar que la distribución del porcentaje de presencialidad está bastante concentrado, con un IQR de 13.03 según el Cuadro 2.3, indicando que existe consistencia en los niveles de asistencia del estudiantado. Sin embargo, si nos fijamos ahora en el IQR de la puntuación en el examen, de 22.85, vemos que es el doble de amplio, indicando gran heterogeneidad en el rendimiento académico, tal y como hay heterogeneidad en el número de horas de estudio, indicado anteriormente. La mayor variabilidad en las puntuaciones de examen, en comparación con porcentajes de asistencia relativamente altos y consistentes entre los estudiantes, deja entrever que otros factores además de la asistencia influyen en el rendimiento, como por ejemplo los hábitos negativos de los estudiantes.

Cuadro 2.3: Estadísticos descriptivos de variables cuantitativas continuas de la base de datos *Hábitos de estudiantes vs Rendimiento académico*. Elaboración propia.

Variable	Mediana	Media	IQR	D2	D9	Q25	Q75	Quasi Var	Var	SD	Asim.	Curt.
horas_estudio_dia	3,50	3,55	1,90	2,30	5,40	2,60	4,50	2,158	2,155	1,469	0,054	-0,067
horas_rrss	2,50	2,51	1,60	1,50	4,00	1,70	3,30	1,375	1,373	1,172	0,119	-0,105
horas_netflix	1,80	1,82	1,53	0,80	3,20	1,00	2,53	1,159	1,155	1,075	0,236	-0,442
porcentaje_presencial	84,40	84,13	13,03	76,40	97,30	78,00	91,03	88,346	88,257	9,399	-0,237	-0,399
horas_sueno	6,50	6,47	1,70	5,40	8,10	5,60	7,30	1,504	1,502	1,226	0,091	-0,225
puntuacion_examen	70,50	69,60	22,85	54,70	93,41	58,48	81,33	285,223	284,938	16,889	-0,156	-0,429

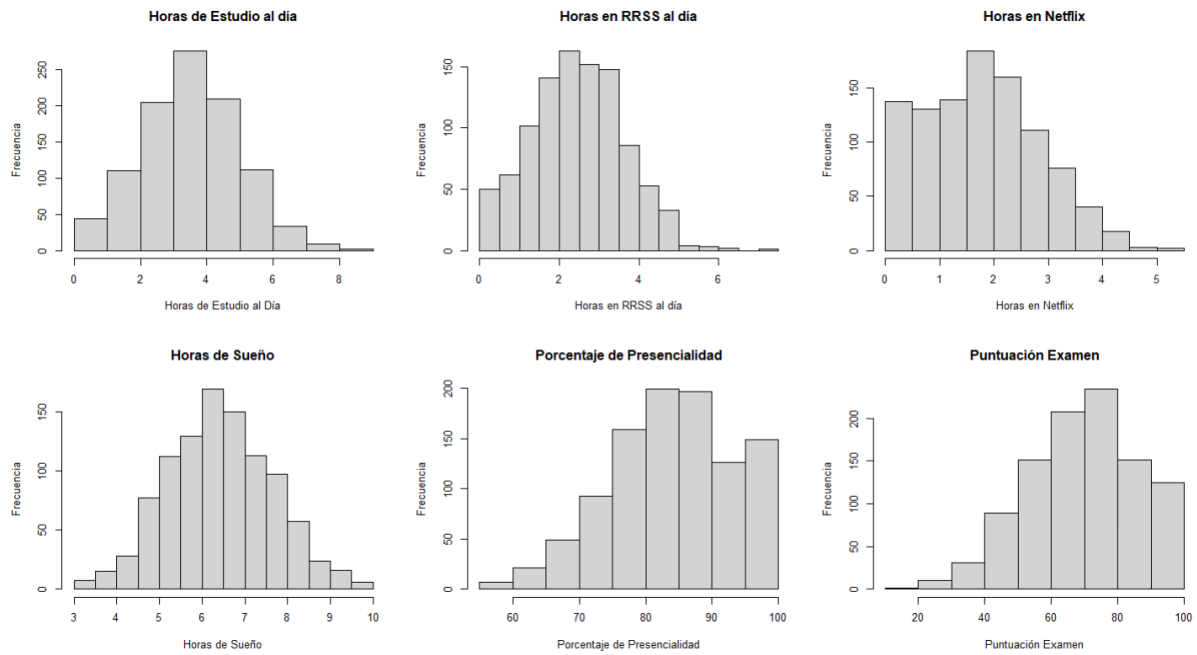


Figura 2.5: Histogramas que muestran la distribución de frecuencia de las cinco variables cuantitativas continuas presentes en la base de datos *Hábitos de estudiantes vs Rendimiento académico*. Elaboración propia.

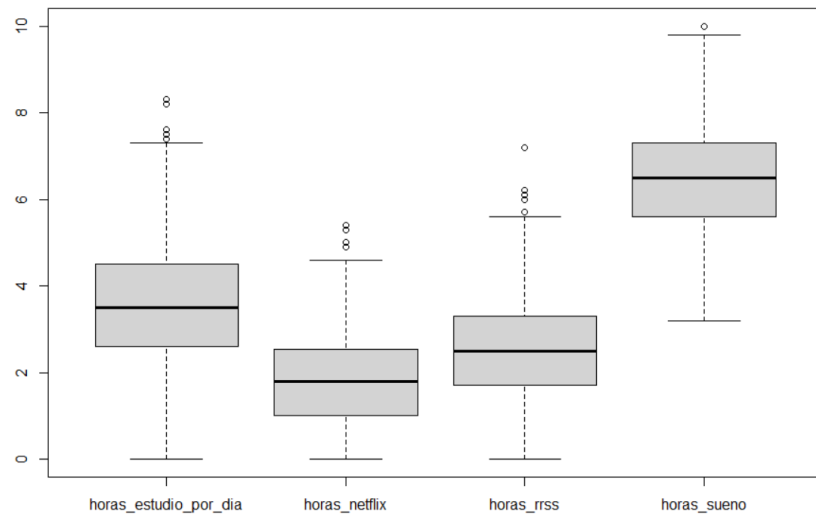


Figura 2.6: Diagrama de cajas de las variables cuantitativas continuas presentes en la base de datos *Hábitos de estudiantes vs Rendimiento académico*, concretamente horas de estudio al día, horas en redes sociales al día, horas en Netflix diarias. Elaboración propia.

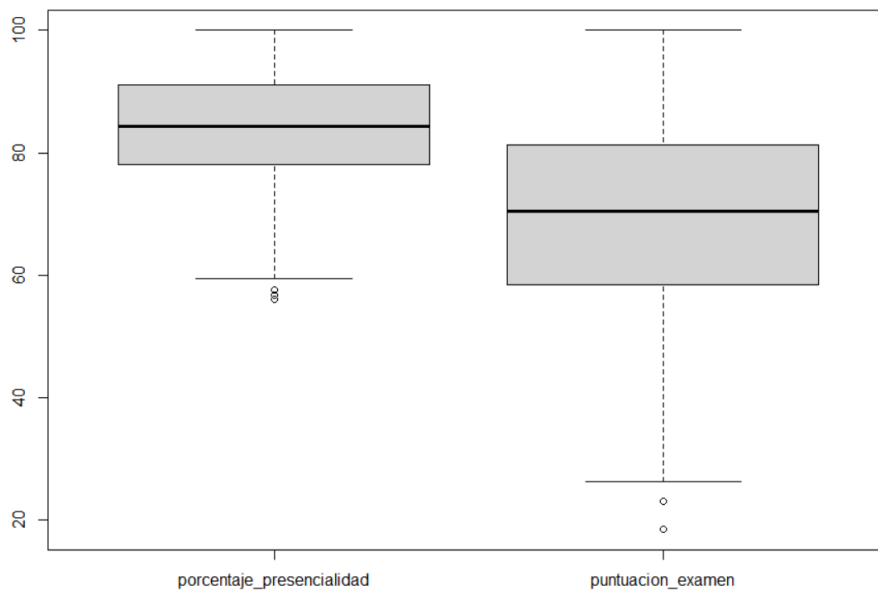


Figura 2.7: Diagrama de cajas de las variables cuantitativas continuas presentes en la base de datos *Hábitos de estudiantes vs Rendimiento académico*, concretamente porcentaje de presencialidad y puntuación en examen

## Capítulo 3

# Conclusiones

El presente caso práctico ha permitido ilustrar la utilidad del análisis estadístico descriptivo como herramienta fundamental para sintetizar y comprender los datos de una manera clara y significativa durante una primera fase del análisis. A través de la aplicación de este tipo de herramientas a dos bases de datos distintas, se han podido extraer conclusiones relevantes sobre cada una de ellas.

En primer lugar, el análisis de las variables cuantitativas discretas en la primera base de datos, *Hábitos de estudiantes vs Rendimiento académico*, reveló una alta uniformidad en las frecuencias, lo que limitó la capacidad de extraer información significativa sobre la muestra analizada, aunque sí se pudo concluir que el rango de edades representa una población de estudiantes universitarios realista. Por esta razón, se optó por continuar el estudio con una segunda base de datos, *Admisión en Máster 2*, que ofreciera mayor variabilidad y, por ende, un mayor potencial informativo.

El análisis reveló que la universidad receptora atrae mayoritariamente a estudiantes de instituciones con rankings intermedios o bajos (niveles 2 y 3 en una escala de 1 a 5, donde 5 representa mayor excelencia). Esto sugiere una reputación académica accesible, con implicaciones en la competitividad del proceso de admisión. Así, los aspirantes pueden evaluar sus posibilidades de ingreso y ajustar sus expectativas según su procedencia. Para quienes provienen de universidades de nivel 3, la competencia puede ser mayor, mientras que los de niveles 4 o 5 podrían encontrar menos barreras al competir en un grupo proveniente de universidades con una posición en el ranking más baja.

En cuanto al análisis de variables cuantitativas continuas, los resultados destacan hábitos de vida y estudio que reflejan niveles bajos en salud y compromiso académico. La media de sueño por debajo de las recomendaciones, el uso limitado de tiempo de estudio en comparación al tiempo dedicado a plataformas digitales como redes sociales y Netflix, apuntan hacia un estudiantado poco saludable. Los estudiantes presentan una asistencia alta y consistente (media del 84.12%) junto con un desempeño aceptable en el examen (media del 70.50%), aunque sin alcanzar la excelencia académica. La variabilidad en las calificaciones del examen es notablemente mayor que en la asistencia (IQR de 22.85 frente a 13.03), revelando una marcada heterogeneidad en el rendimiento académico. Esta disparidad sugiere que factores más allá de la presencialidad, como los hábitos de estudio y de salud, ejercen una influencia determinante en los resultados académicos.

En conjunto, estos resultados demuestran el potencial de un análisis estadístico descriptivo, y cómo permite no solo organizar y visualizar datos, sino también generar interpretaciones útiles que pueden informar decisiones, y guiar pasos futuros de análisis de los datos. Aunque este informe no profundiza en análisis estadísticos más complejos, queda demostrado el valor de este enfoque inicial.