Universidad Cooperativa de Colombia

Proyecto Final

Andres Chacua

Juan Patiño

Gabriel Narvaez

San Juan De Pasto

2 de mayo de 2025

[Introducción 1](#_Toc963392681)

[Metodología 1](#_Toc178151584)

[Desarrollo del proyecto 1](#_Toc489901163)

[Gráficos 2](#_Toc1157657192)

[Histograma con curva de densidad 2](#_Toc1520512060)

[Diagrama de caja 2](#_Toc2017994540)

[Gráfico Q-Q 3](#_Toc1823572201)

[Género (Choose.your.gender.) 3](#_Toc769327507)

[Tabla de frecuencias 4](#_Toc882246744)

[Porcentajes de cada género respecto al total 4](#_Toc795834468)

[Gráficos 4](#_Toc420009411)

[Promedio Académico (What.is.your.CGPA.) 5](#_Toc1149422619)

[Tabla de frecuencias 5](#_Toc1601497714)

[Porcentajes para el rango de notas respecto al total 5](#_Toc663105135)

[Gráficos 6](#_Toc2128118172)

[Presenta depresión (Do.you.have.Depression.) 6](#_Toc2123718276)

[Tabla de frecuencias 6](#_Toc1275444599)

[Porcentaje respecto a si presenta o no presenta depresión 6](#_Toc1719348483)

[Gráficos 6](#_Toc842155161)

[Presenta Ansiedad (Do.you.have.Anxiety.) 7](#_Toc271364443)

[Porcentaje respecto a si presenta o no presenta ansiedad 8](#_Toc651451132)

[Gráficos 8](#_Toc633376866)

[Presenta ataques de Pánico (Do.you.have.Panic.atack.) 9](#_Toc2111867201)

# Introducción

Este proyecto tiene como objetivo aplicar los diferentes temas estudiados durante el curso de probabilidad y estadística, basándonos en un conjunto de datos relacionados con la salud mental de estudiantes de diferentes carreras universitarias.

El estudio de estos datos abarca diferentes variables como la edad, género, su promedio académico (CGPA), la presencia de diferentes condiciones como la ansiedad, depresión o ataques de pánico.

Desde diferentes herramientas estadísticas se analizarán los diferentes datos, basándonos en el comportamiento de cada uno, además identificaremos posibles distribuciones de probabilidad, sacar conclusiones sobre la población y examinar relaciones entre las variables más relevantes.

De esta manera, buscamos entender mejor los factores que podrían impactar en la salud mental dentro del contexto universitario.

## Metodología

Para poder llevar acabo el desarrollo de este proyecto utilizamos un conjunto de datos elegidos desde la página web de Kaggle, estos datos se basaron en encuestas hacia estudiantes universitarios, los cuales contienen variables como la edad, género, salud mental o promedio académico.

El análisis de estos datos se desarrolló a través del software Rstudio y diferentes librerías como readr, dplyr, para el manejo, análisis y la visualización de los datos a través de gráficos.

Las variables para para el análisis estadístico son las siguientes:

* Edad: variable cuantitativa continúa
* Género: variable cualitativa
* Promedio Académico (CGPA): variable cuantitativa discreta
* Presencia de ansiedad: variable cualitativa
* Presencia de depresión: variable cualitativa
* Presencia de ataques de pánico: variable cualitativa

Posterior a ello, se realizaron las siguientes actividades:

* Análisis descriptivo de las variables mediante medidas de tendencia central y dispersión.
* Análisis de distribución para determinar si las variables cuantitativas siguen una distribución conocida.
* Inferencia estadística mediante estimaciones de parámetros y pruebas de hipótesis.
* Análisis de correlación y regresión lineal simple.

## Desarrollo del proyecto

1. **Estadística descriptiva**

En este punto analizamos cada variable del estudio de manera individual, utilizando medidas de tendencia central, dispersión y gráficos para poder ver como se comportan estos datos.

**Edad (Age)**

**Media, mediana, mínimo, máximo:**

* Mínimo: 18 años
* 1er cuartil (Q1): 18 años
* Mediana (Q2): 19 años
* Media: 20.53
* 3cer cuartil (Q3): 23 años
* Máximo: 24 años

**Desviación estándar**: 2.49628

**Varianza**: 6.231414

**Rango**: 18, 24

## Gráficos

Histograma con curva de densidad

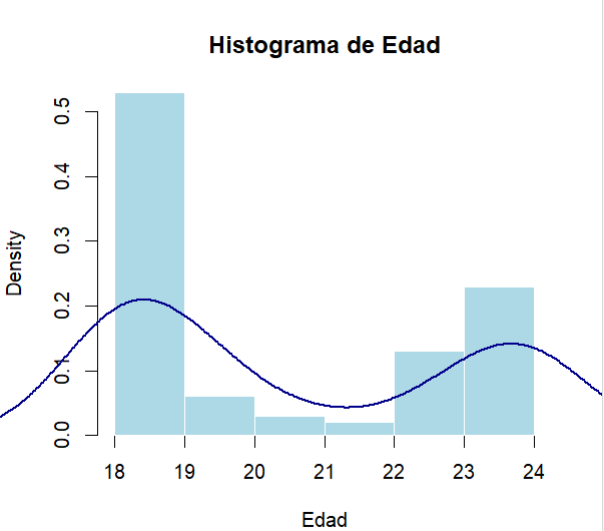
****

Diagrama de caja

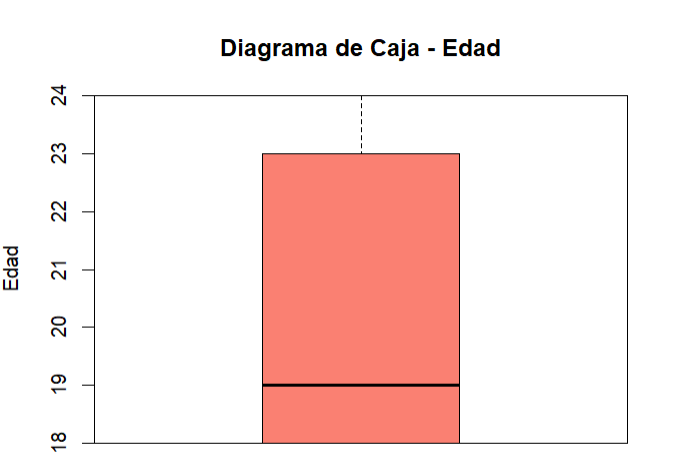
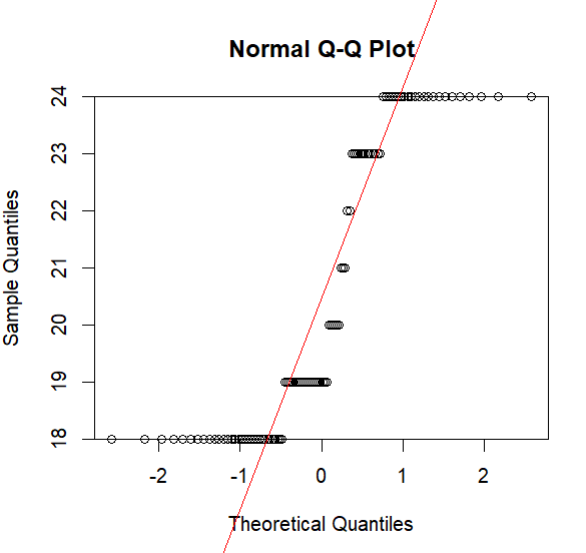


Gráfico de dispersión

****

## Género (Choose.your.gender.)

Tabla de frecuencias

|  |  |  |
| --- | --- | --- |
|  | Variable | Freq |
| 1. | Femenino | 75 |
| 2. | Masculino | 26 |

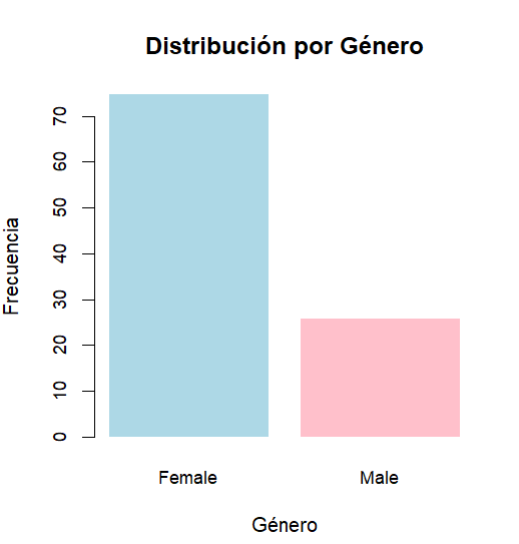
### Porcentajes de cada género respecto al total

Femenino: 74.25743%

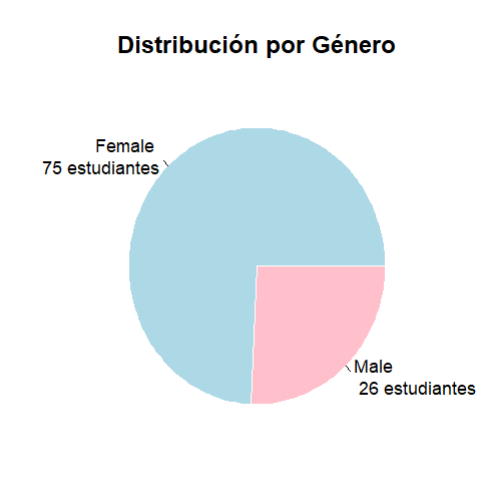
Masculino: 25.74257%

### Gráficos

1. Diagrama de barras



1. Gráfico circular



## Promedio Académico (What.is.your.CGPA.)

**Rangos más comunes**:

* 3.00 – 3.49: 43 estudiantes
* 3.50 – 4.00: 47 estudiantes

Tabla de frecuencias

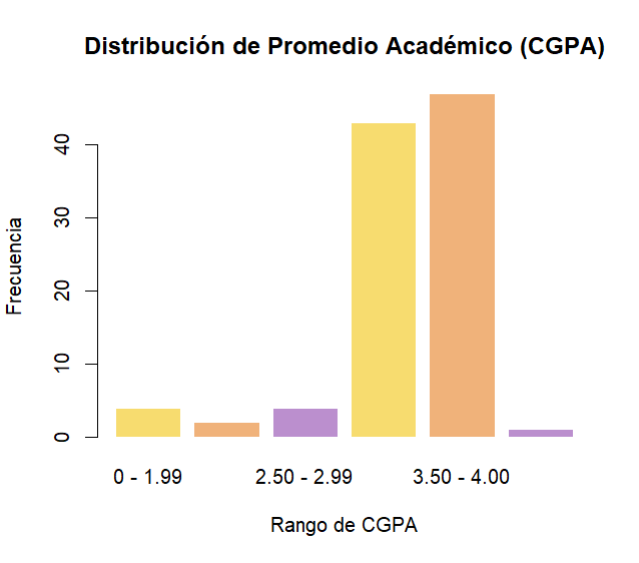


### Porcentajes para el rango de notas respecto al total

* + 0 – 1.99: 3.96039%
  + 2.00 – 2.49: 1.98019%
  + 2.50 – 2.99: 3.96039%
  + 3.00 – 3.49: 42.57425%
  + 3.50 – 4.00: 46.53465%
  + 3.50 – 4.00: 0.99009%

### Gráficos

1. Diagrama de barras



## Presenta depresión (Do.you.have.Depression.)

Tabla de frecuencias

|  |  |  |
| --- | --- | --- |
|  | Variable | Freq |
| 1 | No | 66 |
| 2 | Sí | 35 |

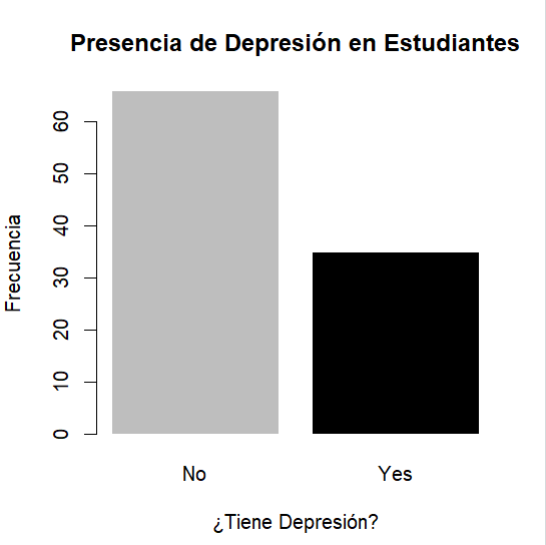
### Porcentaje respecto a si presenta o no presenta depresión

No: 65.34653%

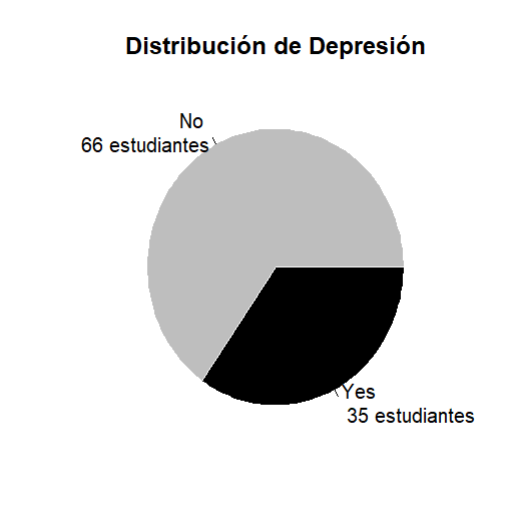
Sí: 34.65347%

### Gráficos

1. Diagrama de barras



1. Diagrama circular



## Presenta Ansiedad (Do.you.have.Anxiety.)

**Tabla de frecuencias**

|  |  |  |
| --- | --- | --- |
|  | Variable | Freq |
| 1 | No | 67 |
| 2 | Sí | 34 |

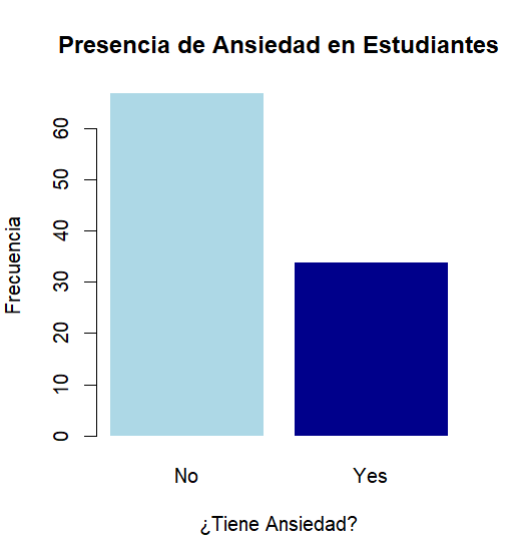
### Porcentaje respecto a si presenta o no presenta ansiedad

No: 66.33663 33.66337

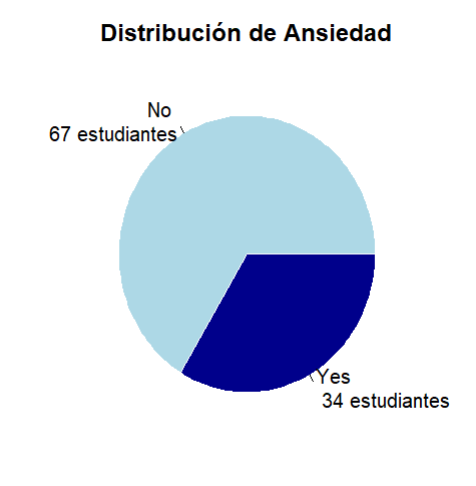
Sí: 33.66337

### Gráficos

1. Diagrama de barras



1. Diagrama circular

****

## Presenta ataques de Pánico (Do.you.have.Panic.atack.)

**Tabla de frecuencias**

|  |  |  |
| --- | --- | --- |
|  | Variable | Freq |
| 1 | No | 68 |
| 2 | Sí | 33 |

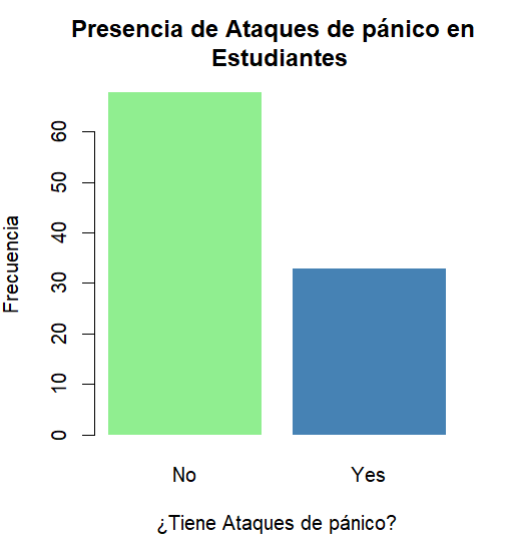
**Porcentaje respecto a si presenta o no presenta ataques de pánico**

No: 67.32673

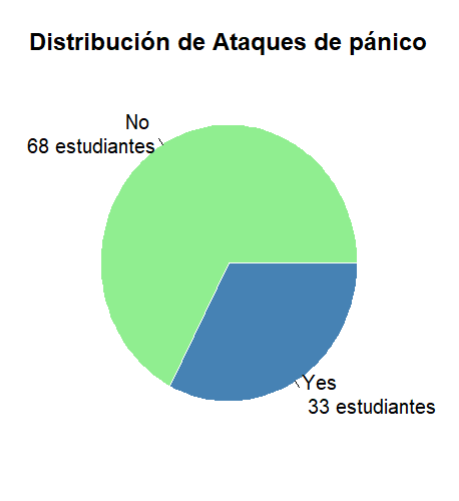
Sí: 32.67327

**Gráficos**

1. Diagrama de barras

****

1. Diagrama circular

****

1. **Distribuciones de probabilidad**

Para analizar las distribuciones se tomó la variable Edad (Age), ya que es una variable continúa numérica, esto, con el objetivo de determinar Age sigue una distribución normal. Para llevar a cabo esto aplicamos la prueba de Shapiro-Wilk, obteniendo los siguientes resultados:

Data: Age

W = 0.78617, p-value = 9.519e-11

Los datos nos muestran que p-valor es menor a 0.05, esto rechaza la hipótesis nula de que los datos se derivan de una distribución normal. Esto lo podemos confirmar visualmente a través del gráfico de histograma y el gráfico Q-Q mostrados anteriormente, en los cuales, estos muestran una distribución irregular y alejada a la simetría y alejada de la simetría característica de una distribución que suele ser normal.