

Julio 20214

Evaluación I

ANalisis estadistico

Jorge Alejandro Cáceres Barrales

# 

Tabla de contenido

[Introducción 3](#_Toc169349040)

[Análisis 4](#_Toc169349041)

[Datos 4](#_Toc169349042)

[Carga Librerias 5](#_Toc169349043)

[Carga de Datos y Funciones 5](#_Toc169349044)

[Tipos de Variables 6](#_Toc169349045)

[Muestra de 150 registros. 6](#_Toc169349046)

[Normalización de las Variables 6](#_Toc169349047)

[Valores Nulos y reemplazar por la Mediana 7](#_Toc169349048)

[Trasformación de variables categóricas a Factor 7](#_Toc169349049)

[Selección de Variables Cuantitativas 8](#_Toc169349050)

[Selección de Variables Cualitativas 8](#_Toc169349051)

[Detalle por Variables Cuantitativas. 9](#_Toc169349052)

[Variable: Sexo 9](#_Toc169349053)

[Interpretación 9](#_Toc169349054)

[Variable: Colegio 10](#_Toc169349055)

[Interpretación 10](#_Toc169349056)

[Variable: Financiación 11](#_Toc169349057)

[Interpretación 11](#_Toc169349058)

[Variable: Clases Virtuales 12](#_Toc169349059)

[Interpretación 12](#_Toc169349060)

[Detalle por Variables Cualitativas 13](#_Toc169349061)

[Variable: Ingreso 13](#_Toc169349062)

[Interpretación 13](#_Toc169349063)

[Variable: Gastos 14](#_Toc169349064)

[Interpretación 14](#_Toc169349065)

[Variable: Puntaje 15](#_Toc169349066)

[Interpretación 15](#_Toc169349067)

[Relación entre Gastos e Ingresos 16](#_Toc169349068)

[Interpretación 16](#_Toc169349069)

[Análisis de variable: Financiamiento y Clases 17](#_Toc169349070)

[ Creación de una tabla de contingencia para las variables Financiacion y Clases 17](#_Toc169349071)

[ Determinar la probabilidad de elegir a un estudiante al azar, que tenga beca y prefiera clases virtuales 17](#_Toc169349072)

[ Si se elige a un estudiante al azar que prefiere clases presenciales, ¿Cuál es la probabilidad que el estudiante tenga beca? 17](#_Toc169349073)

[ Se sabe que el 36% de los estudiantes tiene crédito, el 31% de los estudiantes financia los estudios por otro medio y el resto tiene beca. Un asistente social, indica que hay estudiantes que presentan problemas con el pago de la mensualidad, de lo que deermina que son el 5% de los que tienen beca, el 2% de los que tienen crédito, y 3% de los estudiantes que utiliza otro medio de pago. 18](#_Toc169349074)

[o Determine la probabilidad de que un estudiante tenga problema con el pago. 18](#_Toc169349075)

[o Si el estudiante tiene problemas con el pago. ¿Cuál es la probabilidad que tenga beca? 19](#_Toc169349076)

[Análisis de variable: Fuma 20](#_Toc169349077)

[ Tabla de frecuencias relativas con la variable Fuma 20](#_Toc169349078)

[ Si se eligen 25 estudiantes al azar, ¿Cuál es la probabilidad de que 10 de ellos si fuma? 20](#_Toc169349079)

[ Si se eligen 30 estudiantes al azar, ¿Cuál es la probabilidad de que a lo más 15 de ellos si fuma? 20](#_Toc169349080)

[ Determine el número esperado de estudiante que no fuma. 20](#_Toc169349081)

[Análisis de variable: Ingresos 21](#_Toc169349082)

[ ¿Cuál es la media y la desviación estándar poblacional de la variable ingreso? 21](#_Toc169349083)

[ ¿Cuál es la probabilidad de que el ingreso sea a lo más de $1.500.000? 21](#_Toc169349084)

[ ¿Cuál es la probabilidad de que el ingreso sea de más de $1.000.000? 21](#_Toc169349085)

[Conclusión 22](#_Toc169349086)

# Introducción

En la presente evaluación se realiza un análisis estadístico basado en datos recolectados a través de una encuesta aplicada a una muestra de 400 estudiantes universitarios.

El objetivo es comprender las características y comportamientos de los estudiantes en relación a diversas variables demográficas, socioeconómicas y académicas.

# Análisis

Comienzo del análisis de los datos.

## Datos

Los datos fueron recogidos aplicando una encuesta a una muestra de estudiantes universitarios.

Es un data frame con 400 observaciones y 18 variables, las cuales se describen abajo.

| **Variable** | **Significado** | **Codificación** |
| --- | --- | --- |
| ID | Código de identificación | Código de identificación |
| Sexo | Género | Masculino, femenino |
| Edad | Edad | Numérica de 12.0 a 30.0 |
| Fuma | Fuma | Si=Fuma; No=No fuma |
| Estatura | Estatura | Alta, Baja, Media |
| Colegio | Tipo de colegio de procedencia | Privado; Público |
| Estrato | Estrato socio-económico de su vivienda | 1=Bajo; 2=Medio; 3=Alto |
| Financiación | ¿Cuál fue tu medio de financiación para estudiar en la universidad? | Beca; Crédito; Otro |
| Acumulado | Promedio acumulado al finalizar el semestre anterior | Numérica de 0.0 a 5.0 |
| Gastos | Promedio de sus gastos mensuales (en 10 miles de $) | Numérica de 23.0 a 90.0 |
| Ingreso | Ingreso mensual de su padre (en millones de $) | Numérica de 1.0 a 3.0 |
| Clases | ¿Qué tipo de clases prefiere? | Virtual; Presencial |
| Pandemia | Cree usted que la pandemia afectó económicamente a su familia | 1= Totalmente de acuerdo; 2= En desacuerdo; 3= Ni de acuerdo, ni en desacuerdo; 4= De acuerdo; 5= Totalmente de acuerdo |
| Clases\_virtuales | Me preocupa que tal vez no pueda entender los contenidos de mis asignaturas como me gustaría | 1= Totalmente de acuerdo; 2= En desacuerdo; 3= Ni de acuerdo, ni en desacuerdo; 4= De acuerdo; 5= Totalmente de acuerdo |
| Estadística | Me gusta la estadística | 1= Totalmente de acuerdo; 2= En desacuerdo; 3= Indeciso; 4= De acuerdo; 5= Totalmente de acuerdo |
| Inseguridad | Me siento inseguro cuando hago problemas de estadística | 1= Totalmente de acuerdo; 2= En desacuerdo; 3= Indeciso; 4= De acuerdo; 5= Totalmente de acuerdo |
| Vida\_cotidiana | Utilizo la estadística en la vida cotidiana | 1= Totalmente de acuerdo; 2= En desacuerdo; 3= Indeciso; 4= De acuerdo; 5= Totalmente de acuerdo |
| Puntaje | Porcentaje de acierto en una determinada prueba | Numérica de 0 a 100% |

## 

## Carga Librerias

Carga de las librerias necesarias para realizar el análisis de datos.

cat("\014")

# Cargar librerías necesarias  
library(dplyr)

library(stringr)  
library(kableExtra)

library(readxl)  
library(summarytools)

## 

## Carga de Datos y Funciones

Importar la base de datos a utilizar y las funciones que permitirán realizar el análisis estadístico.

# Enlace raw al archivo de funciones en GitHub  
source("https://raw.githubusercontent.com/jkcrs1/Code/main/Analisis/funciones.R")

# Carga Archivo xlsx  
file\_path <- "/Users/jorge/Documents/GitHub/Code/Analisis/Estudiantes.xlsx"  
estudiantes\_total <- read\_excel(file\_path, sheet = "Estudiant")

## Tipos de Variables

Utilizando el paquete summarytools, se genera un resumen detallado de todas las variables del dataframe. Este informe incluye estadísticas descriptivas, frecuencias, gráficos de barras para variables categóricas y otras visualizaciones útiles para el análisis de datos.

# Configurar el entorno de summarytools para HTML  
st\_options(style = "rmarkdown", plain.ascii = FALSE)  
  
# Obtener un resumen de los datos utilizando summarytools y renderizar en HTML  
dfSummary(estudiantes\_total) %>%  
 print(method = 'render')

## 

## Muestra de 150 registros.

Seleccionar una muestra aleatoria de 150 registros basada en la desviación estándard de la variable puntaje. Esta muestra se utilizará para realizar un análisis sobre un conjunto de datos reducido pero que sigue siendo representativo del conjunto de datos original.

# Calcular la muestra aleatoria según Desviación Estándar  
cant <- nrow(estudiantes\_total)  
sd <- sd(estudiantes\_total$Puntaje)  
n <- tam.muestra(alfa = 0.05, epsilon = 2.4, s = sd, N = cant)  
set.seed(2)  
cant <- sample(nrow(estudiantes\_total), n)  
estudiantes\_data <- estudiantes\_total[cant, ]  
  
cat("La cantidad de registros de muestra es:", nrow(estudiantes\_data))

## La cantidad de registros de muestra es: 150

## 

## Normalización de las Variables

Se procede a normalizar y limpiar el conjunto de datos, asegurando que las variables estén en el formato correcto y que los valores nulos se manejen adecuadamente.

# Aplicar todas las transformaciones y normalizaciones  
estudiantes\_data <- estudiantes\_data %>%  
 mutate(  
 across(everything(), reemplazar\_nulos),  
 across(c(Sexo, Estatura, Fuma, Colegio, Financiacion, Clases, Pandemia, Clases\_virtuales, Estadistica, inseguridad, vida\_cotidiana), str\_to\_title)  
 ) %>%  
 mutate(  
 Edad = as.integer(Edad),  
 Acumulado = as.numeric(Acumulado),  
 Gastos = as.numeric(Gastos),  
 Ingreso = as.numeric(Ingreso)  
 ) %>%  
 distinct()

## 

## Valores Nulos y reemplazar por la Mediana

En caso que se encuentren valores, se reemplazan por el valor de la mediana de cada variable.

#Reemplazar valores nulos por la mediana  
estudiantes\_data <- estudiantes\_data %>%  
 mutate(across(everything(), reemplazar\_por\_mediana))

## 

## Trasformación de variables categóricas a Factor

Para asegurar un análisis e interpretación adecuados de los datos, las variables categóricas se transforman a factores.

# Definir el orden de las categorías para las variables ordinales  
niveles\_uno <- c("Totalmente En Desacuerdo", "En Desacuerdo", "Ni De Acuerdo, Ni En Desacuerdo",  
 "De Acuerdo", "Totalmente De Acuerdo")  
niveles\_dos <- c("Totalmente En Desacuerdo", "En Desacuerdo", "Indeciso",  
 "De Acuerdo", "Totalmente De Acuerdo")  
niveles\_tres <- c("En Desacuerdo", "De Acuerdo", "Ni De Acuerdo, Ni En Desacuerdo")  
niveles\_estatura <- c("Baja", "Media", "Alta")  
niveles\_sexo <- c("Masculino", "Femenino")  
niveles\_colegio <- c("Privado", "Publico")  
niveles\_financiacion <- c("Beca", "Credito", "Otro")  
  
# Lista de variables a transformar  
variables\_ordinales\_uno <- c("Clases\_virtuales", "Pandemia")  
variables\_ordinales\_dos <- c("Estadistica", "inseguridad", "vida\_cotidiana")  
variables\_ordinales\_tres <- c("Pandemia")  
variables\_estatura <- c("Estatura")  
variables\_sexo <- c("Sexo")  
variables\_colegio <- c("Colegio")  
variables\_financiacion <- c("Financiacion")  
  
# Transformar las variables a factor con los niveles definidos  
estudiantes\_data <- estudiantes\_data %>%  
 mutate(across(all\_of(variables\_ordinales\_uno), ~factor(., levels = niveles\_uno, ordered = TRUE))) %>%  
 mutate(across(all\_of(variables\_ordinales\_dos), ~factor(., levels = niveles\_dos, ordered = TRUE))) %>%  
 mutate(across(all\_of(variables\_ordinales\_tres), ~factor(., levels = niveles\_tres, ordered = TRUE))) %>%  
 mutate(across(all\_of(variables\_estatura), ~factor(., levels = niveles\_estatura, ordered = TRUE))) %>%  
 mutate(across(all\_of(variables\_sexo), ~factor(., levels = niveles\_sexo, ordered = TRUE))) %>%  
 mutate(across(all\_of(variables\_colegio), ~factor(., levels = niveles\_colegio, ordered = TRUE))) %>%  
 mutate(across(all\_of(variables\_financiacion), ~factor(., levels = niveles\_financiacion, ordered = TRUE)))

## 

## Selección de Variables Cuantitativas

Selección y resumen detallado de tres variables cuantitativas.

# Seleccionar tres variables cuantitativas  
cuanti\_vars <- estudiantes\_data %>% select(Edad, Gastos, Ingreso)  
  
# Crear un resumen de variables cuantitativas utilizando summarytools  
print(dfSummary(cuanti\_vars), method = 'render')

## 

## Selección de Variables Cualitativas

Selección y resumen detallado de cinco variables cualiitativa.

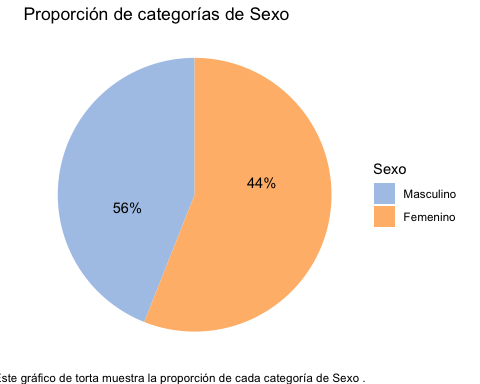
# Seleccionar cinco variables cualitativas  
cuali\_vars <- estudiantes\_data %>% select(Sexo, Colegio, Financiacion, Clases\_virtuales, Pandemia)  
  
# Crear un resumen de variables cualitativas utilizando summarytools  
print(dfSummary(cuali\_vars), method = 'render')

## Detalle por Variables Cuantitativas.

Se realiza un gráfico para cada variable cuantitativa con el fin de explicar su distribución en el conjunto de datos analizado.

### Variable: Sexo

var\_x <- "Sexo"  
var\_tipo\_grafico <- "torta"  
  
grafico(  
 data = estudiantes\_data,   
 var = var\_x,   
 tipo\_grafico = var\_tipo\_grafico  
)



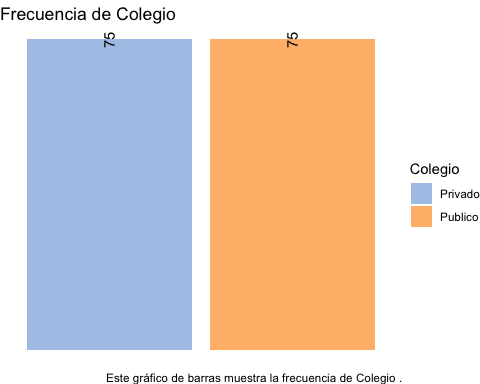
### Interpretación

Se observa que hay una mayor proporción de estudiantes masculinos (56%) en comparación con las estudiantes femeninos (44%).

### 

### Variable: Colegio

var\_x <- "Colegio"  
var\_tipo\_grafico <- "barra"  
  
grafico(  
 data = estudiantes\_data,   
 var = var\_x,   
 tipo\_grafico = var\_tipo\_grafico  
)



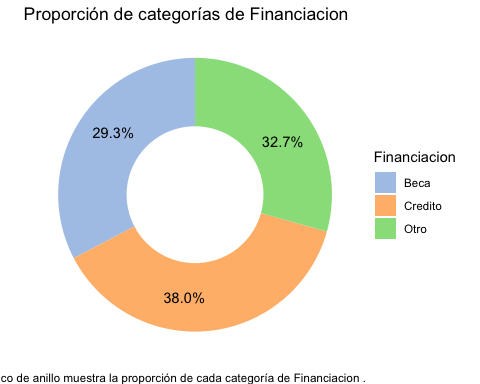
### Interpretación

Se observa una distribución equitativa de estudiantes entre colegios Privados y Públicos, con 75 estudiantes en cada tipo de colegio.

### 

### Variable: Financiación

var\_x <- "Financiacion"  
var\_tipo\_grafico <- "anillo"  
  
grafico(  
 data = estudiantes\_data,   
 var = var\_x,   
 tipo\_grafico = var\_tipo\_grafico)



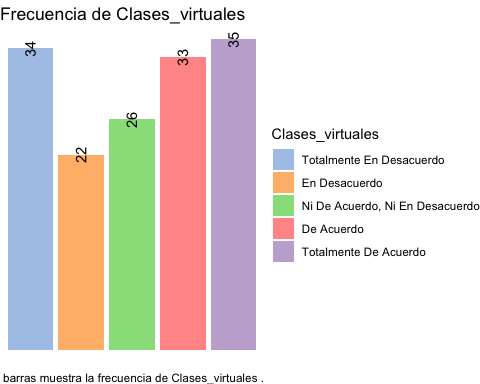
### Interpretación

Se observa una mayor cantidad de estudiantes que optan por un financiamiento vía crédito (38%), seguido por otros métodos (32,7%) y en menor medida beca (29,3%).

### 

### Variable: Clases Virtuales

var\_x <- "Clases\_virtuales"  
var\_tipo\_grafico <- "barra"  
  
grafico(  
 data = estudiantes\_data,   
 var = var\_x,   
 tipo\_grafico = var\_tipo\_grafico)



### Interpretación

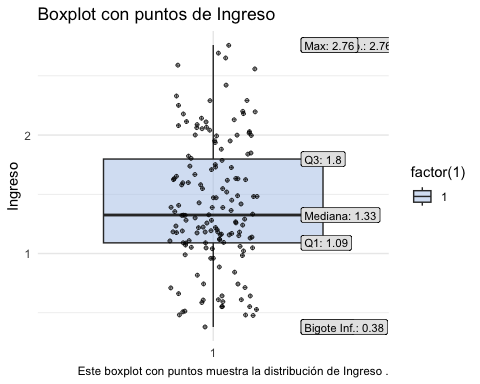
El gráfico revela una marcada polarización en las opiniones sobre las clases virtuales, con la mayoría de los estudiantes ubicándose en los extremos de “Totalmente En Desacuerdo” (34) y “Totalmente De Acuerdo” (35). Las respuestas intermedias son menos frecuentes, con 22 estudiantes en desacuerdo y 33 de acuerdo, y 26 estudiantes sin una opinión definida, lo que indica una clara división en las percepciones.

## Detalle por Variables Cualitativas

Se realiza un gráfico para cada variable cualitativa con el fin de explicar su distribución en el conjunto de datos analizado.

### Variable: Ingreso

var\_x <- "Ingreso"  
var\_tipo\_grafico <- "boxplot"  
  
grafico(  
 data = estudiantes\_data,   
 var = var\_x,   
 tipo\_grafico = var\_tipo\_grafico  
)

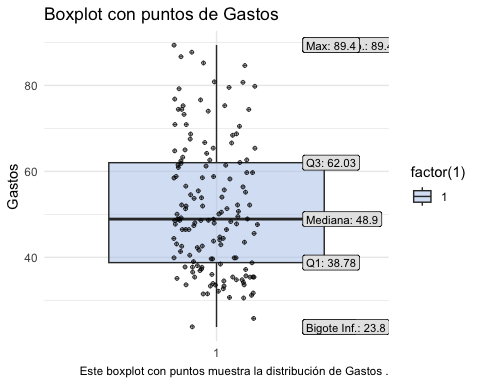


### Interpretación

La mediana de los ingresos de los estudiantes es de 1.33 millones de pesos. Los valores para los ingresos de los estudiantes, están entre los siguientes valores 0.38 y 2.76 millones de pesos. El rango intercuartil (IQR), que oscila entre 1.09 y 1.8 millones de pesos, refleja una variabilidad en los ingresos. La mayoría de los ingresos se sitúan entre 1.09 y 1.8 millones de pesos, con una tendencia general hacia valores más altos.

### Variable: Gastos

var\_x <- "Gastos"  
var\_tipo\_grafico <- "boxplot"  
  
grafico\_cuantitativo(  
 data = estudiantes\_data,   
 var\_cuant\_x = var\_x,   
 tipo\_grafico = var\_tipo\_grafico  
)



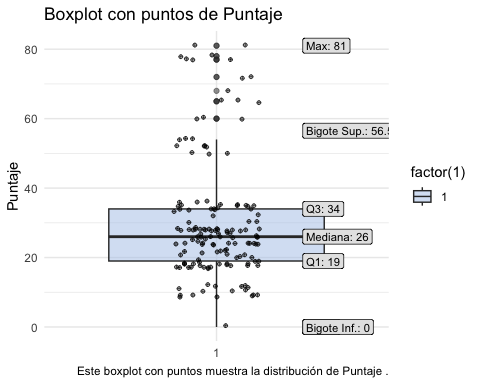
### Interpretación

La mediana de los gastos de los estudiantes es de 48.9. Los valores para los gastos, están entre el siguiente rango 23.8 y 89.4. El rango intercuartil (IQR), que va de 38.78 a 62.03, refleja una alta variabilidad en los gastos. La mayoría de los gastos se encuentra entre 38.78 y 62.03, con una tendencia a extenderse hacia valores más altos.

### 

### Variable: Puntaje

var\_x <- "Puntaje"  
var\_tipo\_grafico <- "boxplot"  
  
grafico(  
 data = estudiantes\_data,   
 var = var\_x,   
 tipo\_grafico = var\_tipo\_grafico  
)

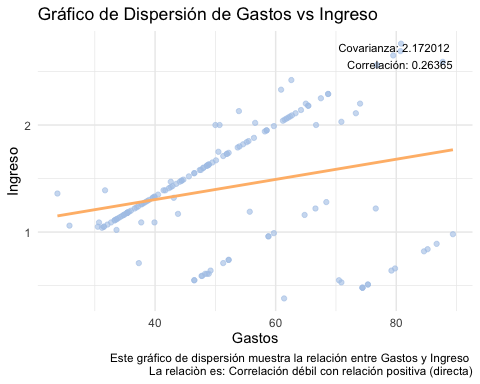


### Interpretación

El gráfico muestra que la mediana de los puntajes de los estudiantes es 26, con un rango intercuartil (IQR) que varía de 19 a 34, indicando una variabilidad en los puntajes. Se observan valores atípicos por encima de 56.5, sugiriendo que algunos estudiantes tienen puntajes significativamente distintos del promedio, alcanzando un máximo de 81 puntos.. La mayoría de los puntajes se encuentran entre 19 y 34, aunque hay una dispersión hacia valores más altos.

## Relación entre Gastos e Ingresos

var\_x <- "Gastos"  
var\_y <- "Ingreso"  
  
grafico\_dispersion(  
 data = estudiantes\_data,   
 var\_x = var\_x,   
 var\_y = var\_y  
)



### Interpretación

La covarianza de 2.172012 indica una relación directa entre las dos variables. El coeficiente de correlación de 0.26365 revela una correlación positiva débil, sugiriendo que, aunque los ingresos tienden a aumentar con los gastos, esta relación no es muy fuerte.

El gráfico muestra que los estudiantes con mayores gastos suelen tener ingresos más altos, aunque esta tendencia es moderada. La dispersión de los puntos y el bajo coeficiente de correlación indican una considerable variabilidad en la relación entre gastos e ingresos.

## Análisis de variable: Financiamiento y Clases

* Creación de una tabla de contingencia para las variables Financiacion y Clases

tabla\_contingencia <- table(estudiantes\_data$Financiacion, estudiantes\_data$Clases)

* Determinar la probabilidad de elegir a un estudiante al azar, que tenga beca y prefiera clases virtuales

# Número total de estudiantes  
total\_estudiantes <- nrow(estudiantes\_data)  
  
# Número de estudiantes con beca y que prefieren clases virtuales  
beca\_virtual <- sum(estudiantes\_data$Financiacion == "Beca" & estudiantes\_data$Clases == "Virtual")  
  
# Calcular la probabilidad  
prob\_beca\_virtual <- beca\_virtual / total\_estudiantes  
  
cat("Probabilidad de elegir a un estudiante con beca y que prefiera clases virtuales: ", prob\_beca\_virtual)

**## Probabilidad de elegir a un estudiante con beca y que prefiera clases virtuales: 0.14**

* Si se elige a un estudiante al azar que prefiere clases presenciales, ¿Cuál es la probabilidad que el estudiante tenga beca?

# Número de estudiantes que prefieren clases presenciales  
presencial\_total <- sum(estudiantes\_data$Clases == "Presencial")  
  
# Número de estudiantes con beca y que prefieren clases presenciales  
beca\_presencial <- sum(estudiantes\_data$Financiacion == "Beca" & estudiantes\_data$Clases == "Presencial")  
  
# Calcular la probabilidad condicional  
prob\_beca\_dado\_presencial <- beca\_presencial / presencial\_total  
cat("Probabilidad de que un estudiante que prefiere clases presenciales tenga beca: ", prob\_beca\_dado\_presencial)

**## Probabilidad de que un estudiante que prefiere clases presenciales tenga beca: 0.2641509**

* Se sabe que el 36% de los estudiantes tiene crédito, el 31% de los estudiantes financia los estudios por otro medio y el resto tiene beca. Un asistente social, indica que hay estudiantes que presentan problemas con el pago de la mensualidad, de lo que determina que son el 5% de los que tienen beca, el 2% de los que tienen crédito, y 3% de los estudiantes que utiliza otro medio de pago.
  + Diagrama de Árbol.

Financiamiento  
 / | \  
 Beca Crédito Otro  
 / \ / \ / \  
 Pagado No Pagado No Pagado No  
 Pagado Pagado Pagado

# Porcentajes de financiamiento  
prob\_beca <- 1 - 0.36 - 0.31  
prob\_credito <- 0.36  
prob\_otro <- 0.31  
  
# Probabilidades de problemas con el pago  
prob\_problema\_beca <- 0.05  
prob\_problema\_credito <- 0.02  
prob\_problema\_otro <- 0.03

* + Determine la probabilidad de que un estudiante tenga problema con el pago.

# Probabilidad total de tener problemas con el pago  
prob\_problema\_total <- (prob\_problema\_beca \* prob\_beca) + (prob\_problema\_credito \* prob\_credito) + (prob\_problema\_otro \* prob\_otro)  
cat("Probabilidad de que un estudiante tenga problemas con el pago: ", prob\_problema\_total)

**## Probabilidad de que un estudiante tenga problemas con el pago: 0.033**

* + Si el estudiante tiene problemas con el pago. ¿Cuál es la probabilidad que tenga beca?

# Probabilidad de tener beca dado que tiene problemas con el pago (utilizando el Teorema de Bayes)  
prob\_beca\_dado\_problema <- (prob\_problema\_beca \* prob\_beca) / prob\_problema\_total  
cat("Probabilidad de que un estudiante con problemas con el pago tenga beca: ", prob\_beca\_dado\_problema)

**## Probabilidad de que un estudiante con problemas con el pago tenga beca: 0.5**

## 

## Análisis de variable: Fuma

* Tabla de frecuencias relativas con la variable Fuma

# Tabla de Frecuencia  
frecuencias\_absolutas <- table(estudiantes\_data$Fuma)  
frecuencias\_relativas <- prop.table(frecuencias\_absolutas)  
print(frecuencias\_relativas)

**## No Si   
## 0.58 0.42**

* Si se eligen 25 estudiantes al azar, ¿Cuál es la probabilidad de que 10 de ellos si fuma?

# Probabilidad de que 10 de 25 estudiantes fumen  
prob\_fuma <- frecuencias\_relativas["Si"]  
prob\_10\_de\_25 <- dbinom(10, size = 25, prob = prob\_fuma)  
cat("Probabilidad de que 10 de 25 estudiantes fumen: ", prob\_10\_de\_25)

**## Probabilidad de que 10 de 25 estudiantes fumen: 0.1578686**

* Si se eligen 30 estudiantes al azar, ¿Cuál es la probabilidad de que a lo más 15 de ellos si fuma?

# Probabilidad de que a lo más 15 de 30 estudiantes fumen  
prob\_a\_lo\_mas\_15\_de\_30 <- pbinom(15, size = 30, prob = prob\_fuma)  
cat("Probabilidad de que a lo más 15 de 30 estudiantes fumen: ", prob\_a\_lo\_mas\_15\_de\_30)

**## Probabilidad de que a lo más 15 de 30 estudiantes fumen: 0.8580969**

* Determine el número esperado de estudiante que no fuma.

# IV. Número esperado de estudiantes que no fuman  
total\_estudiantes <- nrow(estudiantes\_data)  
prob\_no\_fuma <- frecuencias\_relativas["No"]  
esperado\_no\_fuma <- total\_estudiantes \* prob\_no\_fuma  
cat("Número esperado de estudiantes que no fuman: ", esperado\_no\_fuma)

**## Número esperado de estudiantes que no fuman: 87**

## Análisis de variable: Ingresos

* ¿Cuál es la media y la desviación estándar poblacional de la variable ingreso?

# Media y desviación estándar poblacional de la variable ingreso  
media\_ingreso <- mean(estudiantes\_data$Ingreso)  
desviacion\_estandar\_ingreso <- sd(estudiantes\_data$Ingreso)  
  
cat("Media de Ingreso: ", media\_ingreso)

**## Media de Ingreso: 1.412667**

cat("Desviación Estándar de Ingreso: ", desviacion\_estandar\_ingreso)

**## Desviación Estándar de Ingreso: 0.5426602**

* ¿Cuál es la probabilidad de que el ingreso sea a lo más de $1.500.000?

# Probabilidad de que el ingreso sea a lo más de $1.500.000  
ingreso\_1.5 <- 1.5  
prob\_ingreso\_1.5 <- pnorm(ingreso\_1.5, mean = media\_ingreso, sd = desviacion\_estandar\_ingreso)  
cat("Probabilidad de que el ingreso sea a lo más de $1.500.000: ", prob\_ingreso\_1.5)

**## Probabilidad de que el ingreso sea a lo más de $1.500.000: 0.5639279**

* ¿Cuál es la probabilidad de que el ingreso sea de más de $1.000.000?

ingreso\_1 <- 1.0  
prob\_ingreso\_1 <- 1 - pnorm(ingreso\_1, mean = media\_ingreso, sd = desviacion\_estandar\_ingreso)  
cat("Probabilidad de que el ingreso sea de más de $1.000.000: ", prob\_ingreso\_1)

**## Probabilidad de que el ingreso sea de más de $1.000.000: 0.7765076**

# 

# Conclusión

El análisis estadístico de 400 estudiantes universitarios reveló importantes aspectos sobre sus características y comportamientos:

* Género y Colegio de Procedencia: La muestra está equilibrada en cuanto al género y el tipo de colegio (privado o público) de procedencia.
* Financiación: Los estudiantes se financian mayormente a través de créditos (38%), seguidos por otros métodos (32.7%) y becas (29.3%).
* Clases Virtuales: Existe una polarización marcada en las opiniones sobre las clases virtuales, indicando la necesidad de estrategias de enseñanza adaptativas.
* Ingresos y Gastos: La mediana de ingresos de los padres es de 1.33 millones de pesos y la mediana de gastos mensuales de los estudiantes es de 48.9 miles de pesos, con una relación débil entre ambas variables.
* Problemas de Pago: El 3.3% de los estudiantes tiene problemas de pago. La mitad de estos estudiantes tiene becas, sugiriendo dificultades financieras entre los becados.
* Hábitos de Fumar: El 42% de los estudiantes fuma, y hay una alta probabilidad de que un número significativo de estudiantes fume si se seleccionan al azar.
* Desempeño Académico: La mediana de los puntajes es 26, con una variabilidad notable, indicando la necesidad de intervenciones académicas.

En resumen, estos hallazgos subrayan la diversidad de antecedentes y necesidades entre los estudiantes, destacando áreas clave para la implementación de políticas y programas universitarios más efectivos.