

# Evaluación I

**ANALISIS ESTADISTICO** 

JORGE ALEJANDRO CÁCERES BARRALES

# Tabla de contenido

Introducción	3
Análisis	4
Datos	4
Carga Librerias	5
Carga de Datos y Funciones	5
Tipos de Variables	6
Muestra de 150 registros	6
Normalización de las Variables	6
Valores Nulos y reemplazar por la Mediana	7
Trasformación de variables categóricas a Factor	
Selección de Variables Cuantitativas	
Selección de Variables Cualitativas	8
Detalle por Variables Cuantitativas	9
Variable: Sexo	9
Interpretación	9
Variable: Colegio	
Interpretación	
Variable: Financiación	
Interpretación	
Variable: Clases VirtualesInterpretación	
Detalle por Variables Cualitativas	
Variable: Ingreso	
Interpretación	
Variable: Gastos	
InterpretaciónVariable: Puntaje	
Interpretación	
Relación entre Gastos e Ingresos	16
Interpretación	
Análisis de variable: Financiamiento y Clases	17
<ul> <li>Creación de una tabla de contingencia para las variables Financiacion y Clases</li> </ul>	
<ul> <li>Determinar la probabilidad de elegir a un estudiante al azar, que tenga beca y pre</li> </ul>	
virtuales	
• Si se elige a un estudiante al azar que prefiere clases presenciales, ¿Cuál es la pro	
<ul> <li>que el estudiante tenga beca?</li> <li>Se sabe que el 36% de los estudiantes tiene crédito, el 31% de los estudiantes fin</li> </ul>	
• Se sabe que el 36% de los estudiantes tiene credito, el 31% de los estudiantes fines estudios por otro medio y el resto tiene beca. Un asistente social, indica que hay estudi	
presentan problemas con el pago de la mensualidad, de lo que deermina que son el 5%	

tienen beca, el 2% de los que tienen crédito, y 3% de los estudiantes que utiliza otro medio de pago. 18	
<ul> <li>Determine la probabilidad de que un estudiante tenga problema con el pago</li> <li>Si el estudiante tiene problemas con el pago. ¿Cuál es la probabilidad que tenga beca?</li> </ul>	
Análisis de variable: Fuma	
Tabla de frecuencias relativas con la variable Fuma	20
<ul> <li>Si se eligen 25 estudiantes al azar, ¿Cuál es la probabilidad de que 10 de ellos si fuma?</li> <li>Si se eligen 30 estudiantes al azar, ¿Cuál es la probabilidad de que a lo más 15 de ellos si</li> </ul>	
fuma?	
Determine el número esperado de estudiante que no fuma	20
Análisis de variable: Ingresos	<b>.2</b> 1
• ¿Cuál es la media y la desviación estándar poblacional de la variable ingreso?	2
• ¿Cuál es la probabilidad de que el ingreso sea a lo más de \$1.500.000?	2
• ¿Cuál es la probabilidad de que el ingreso sea de más de \$1.000.000?	2
Conclusión	. 22

# 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

Variable	Significado	Codificación
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/funci
ones.R")
# Carga Archivo xlsx
file_path <- "/Users/jorge/Documents/GitHub/Code/Analisis/Estudiantes.xls
x"
estudiantes_total <- read_excel(file_path, sheet = "Estudiant")</pre>
```

#### **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</pre>
```

#### 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, Pandemi
a, Clases_virtuales, Estadistica, inseguridad, vida_cotidiana), str_to_ti
tle)
```

```
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 Acue</pre>
rdo, Ni En Desacuerdo",
               "De Acuerdo", "Totalmente De Acuerdo")
niveles_dos <- c("Totalmente En Desacuerdo", "En Desacuerdo", "Indeciso",</pre>
              "De Acuerdo", "Totalmente De Acuerdo")
niveles_tres <- c("En Desacuerdo", "De Acuerdo", "Ni De Acuerdo, Ni En De</pre>
sacuerdo")
niveles_estatura <- c("Baja", "Media", "Alta")
niveles_sexo <- c("Masculino", "Femenino")</pre>
niveles_colegio <- c("Privado", "Publico")
niveles financiacion <- c("Beca", "Credito", "Otro")</pre>
# Lista de variables a transformar
variables_ordinales_uno <- c("Clases_virtuales", "Pandemia")</pre>
variables_ordinales_dos <- c("Estadistica", "inseguridad", "vida_cotidian</pre>
variables ordinales tres <- c("Pandemia")</pre>
variables estatura <- c("Estatura")</pre>
variables_sexo <- c("Sexo")</pre>
variables colegio <- c("Colegio")</pre>
variables financiacion <- c("Financiacion")</pre>
```

```
# Transformar las variables a factor con los niveles definidos
estudiantes data <- estudiantes data %>%
 mutate(across(all_of(variables_ordinales_uno), ~factor(., levels = nive
les uno, ordered = TRUE))) %>%
 mutate(across(all of(variables ordinales dos), ~factor(., levels = nive
les dos, ordered = TRUE))) %>%
 mutate(across(all of(variables ordinales tres), ~factor(., levels = niv
eles tres, ordered = TRUE))) %>%
 mutate(across(all_of(variables_estatura), ~factor(., levels = niveles_e
statura, ordered = TRUE))) %>%
 mutate(across(all of(variables sexo), ~factor(., levels = niveles sexo,
ordered = TRUE))) %>%
 mutate(across(all of(variables colegio), ~factor(., levels = niveles co
legio, ordered = TRUE))) %>%
 mutate(across(all_of(variables_financiacion), ~factor(., levels = nivel
es 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.

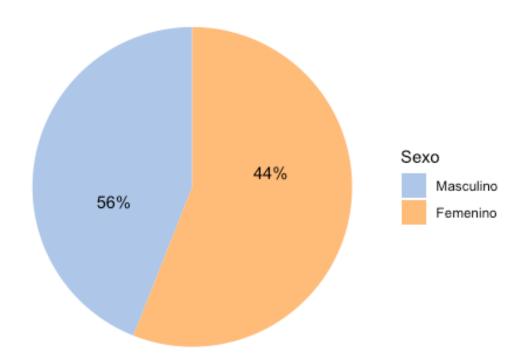
```
# Selectionar cinco variables cualitativas
cuali_vars <- estudiantes_data %>% select(Sexo, Colegio, Financiacion, Cl
ases_virtuales, Pandemia)
# Crear un resumen de variables cualitativas utilizando summarytools
print(dfSummary(cuali_vars), method = 'render')
```

### **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 )</pre>

# Proporción de categorías de Sexo



ste gráfico de torta muestra la proporción de cada categoría de Sexo .

#### Interpretación

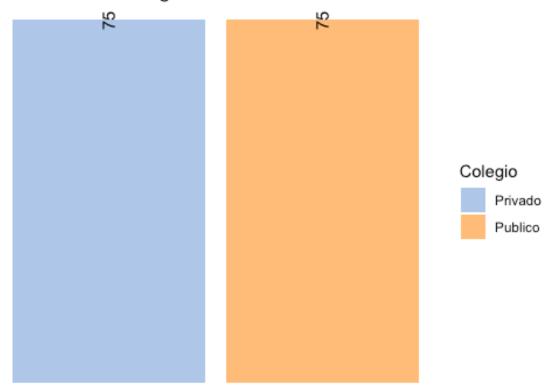
Se observa que hay una mayor proporción de estudiantes masculinos (56%) en comparación con las estudiantes femeninos (44%).

#### **Variable: Colegio**

```
var_x <- "Colegio"
var_tipo_grafico <- "barra"

grafico(
  data = estudiantes_data,
  var = var_x,
  tipo_grafico = var_tipo_grafico
)</pre>
```

# Frecuencia de Colegio



Este gráfico de barras muestra la frecuencia de Colegio .

#### Interpretación

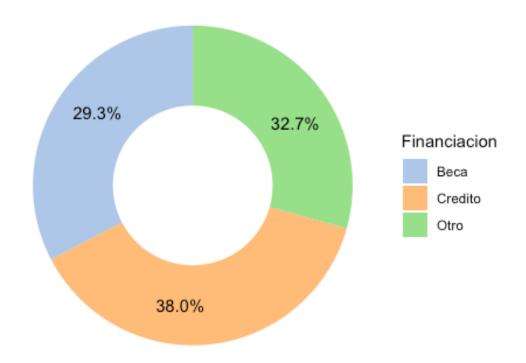
Se observa una distribución equitativa de estudiantes entre colegios Privados y Públicos, con 75 estudiantes en cada tipo de colegio.

#### Variable: Financiación

```
var_x <- "Financiacion"
var_tipo_grafico <- "anillo"

grafico(
  data = estudiantes_data,
  var = var_x,
  tipo_grafico = var_tipo_grafico)</pre>
```

# Proporción de categorías de Financiacion



co de anillo muestra la proporción de cada categoría de Financiacion .

#### Interpretación

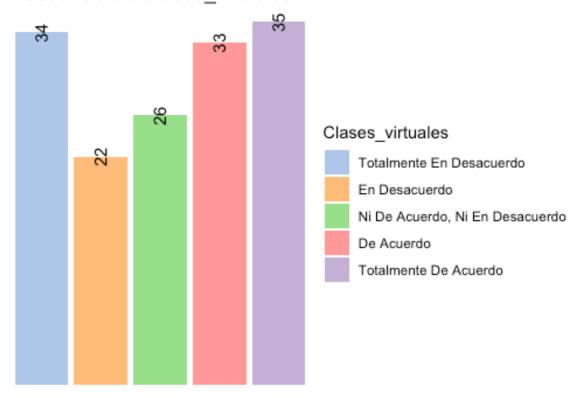
Se observa una mayor cantidad de estudiantes que optan por un financiamiento vía crédito (38%), seguido por otros métodos (32,7%) y en menor medida beca (29,3%).

#### **Variable: Clases Virtuales**

```
var_x <- "Clases_virtuales"
var_tipo_grafico <- "barra"

grafico(
  data = estudiantes_data,
  var = var_x,
  tipo_grafico = var_tipo_grafico)</pre>
```

# Frecuencia de Clases\_virtuales



barras muestra la frecuencia de Clases virtuales .

#### 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**

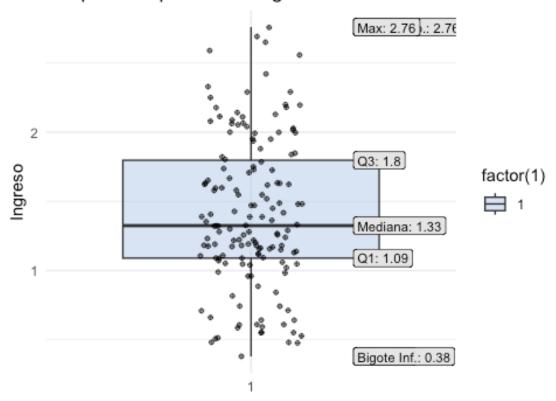
Se realiza un gráfico para cada variable cualitativa con el fin de explicar su distribución en el conjunto de datos analizado.

# Variable: Ingreso

```
var_x <- "Ingreso"
var_tipo_grafico <- "boxplot"

grafico(
  data = estudiantes_data,
  var = var_x,
  tipo_grafico = var_tipo_grafico
)</pre>
```

### Boxplot con puntos de Ingreso



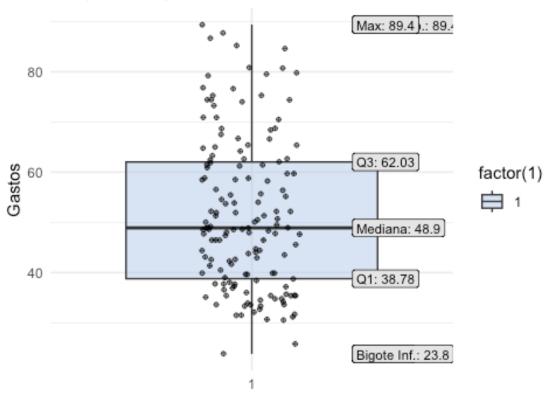
Este boxplot con puntos muestra la distribución de Ingreso .

#### 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 )</pre>

# Boxplot con puntos de Gastos



Este boxplot con puntos muestra la distribución de Gastos .

#### Interpretación

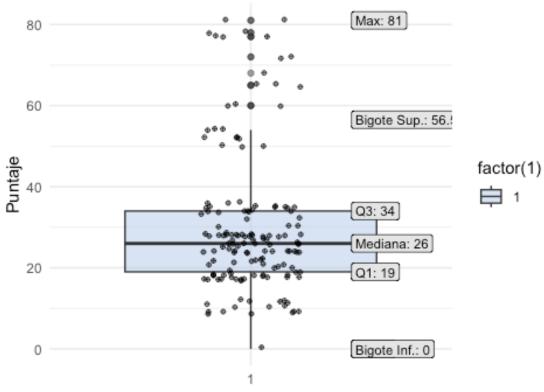
La mediana de los gastos de los estudiantes es de 48.9. Los valores para los gastos, están entre el siguiente rango 23.8 y 89.4. El rango intercuartil (IQR), que va de 38.78 a 62.03, refleja una alta variabilidad en los gastos. La mayoría de los gastos se encuentra entre 38.78 y 62.03, con una tendencia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

# Boxplot con puntos de Puntaje



Este boxplot con puntos muestra la distribución de Puntaje .

#### Interpretación

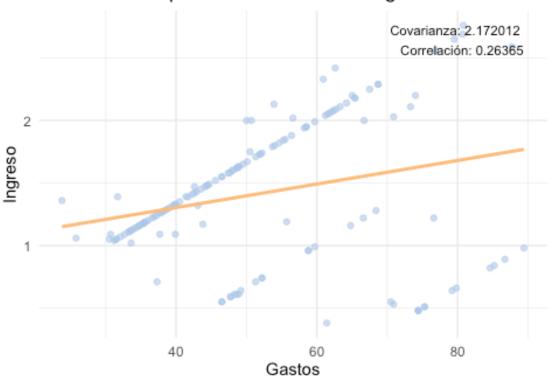
El gráfico muestra que la mediana de los puntajes de los estudiantes es 26, con un rango intercuartil (IQR) que varía de 19 a 34, indicando una variabilidad en los puntajes. Se observan valores atípicos por encima de 56.5, sugiriendo que algunos estudiantes tienen puntajes significativamente distintos 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
)</pre>
```

# Gráfico de Dispersión de Gastos vs Ingreso



Este gráfico de dispersión muestra la relación entre Gastos y Ingreso La relación es: Correlación débil con relación positiva (directa)

#### 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$Financiacion, estudiantes_da
ta$Clases)</pre>
```

• 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 clase
s virtuales: ", prob_beca_virtual)

## Probabilidad de elegir a un estudiante con beca y que prefiera clases
virtuales: 0.14</pre>
```

• 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$Financiacion == "Beca" & estudian
tes_data$Clases == "Presencial")

# Calcular la probabilidad condicional
prob_beca_dado_presencial <- beca_presencial / presencial_total
cat("Probabilidad de que un estudiante que prefiere clases presenciales t
enga beca: ", prob_beca_dado_presencial)

## Probabilidad de que un estudiante que prefiere clases presenciales ten
ga beca: 0.2641509</pre>
```

- 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.
  - o Diagrama de Árbol.

```
Financiamiento
                       Crédito
               Beca
                                    0tro
                        /
        Pagado No
                       Pagado No
                                     Pagado
                                                Pagado
                Pagado
                                Pagado
# Porcentajes de financiamiento
prob_beca <- 1 - 0.36 - 0.31</pre>
prob credito <- 0.36
prob otro <- 0.31
# Probabilidades de problemas con el pago
prob problema beca <- 0.05
prob_problema_credito <- 0.02</pre>
prob_problema_otro <- 0.03</pre>
```

• 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: ", pr
ob_problema_total)
## Probabilidad de que un estudiante tenga problemas con el pago: 0.033</pre>
```

• 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 (utiliz
ando el Teorema de Bayes)
prob_beca_dado_problema <- (prob_problema_beca * prob_beca) / prob_proble
ma_total
cat("Probabilidad de que un estudiante con problemas con el pago tenga be
ca: ", prob_beca_dado_problema)
## Probabilidad de que un estudiante con problemas con el pago tenga beca
: 0.5</pre>
```

#### 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</pre>
```

• 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</pre>
```

• 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_l
o_mas_15_de_30)
## Probabilidad de que a lo más 15 de 30 estudiantes fumen: 0.8580969</pre>
```

• 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</pre>
```

#### 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</pre>
```

• ¿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 = desviac
ion_estandar_ingreso)
cat("Probabilidad de que el ingreso sea a lo más de $1.500.000: ", prob_i
ngreso_1.5)
## Probabilidad de que el ingreso sea a lo más de $1.500.000: 0.5639279</pre>
```

• ¿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 = desviac
ion_estandar_ingreso)
cat("Probabilidad de que el ingreso sea de más de $1.000.000: ", prob_ing
reso_1)
### Probabilidad de que el ingreso sea de más de $1.000.000: 0.7765076</pre>
```

#### 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.