

# MD004 - Trabajo Practico Nro. 3

## Parte I

### 1. Plantea un diseño de test de hipótesis que te gustaría aplicar (1pt)

#### i. Definición de hipótesis nula y alternativa

Objetivo del estudio: Determinar si existe una diferencia significativa en la velocidad de las hormigas antes de un fenómeno climático.

Hipótesis nula ( $H_0$ ): No hay diferencia en la velocidad de las hormigas antes de un fenómeno climático.

1. Hipótesis nula:  $H_0 : \mu_{prueba} - \mu_{inicial} = 0$  (No hay un cambio de comportamiento significativo)

Hipótesis alternativa ( $H_1$ ): Existe una diferencia en la velocidad de las hormigas antes de un fenómeno climático.

2. Hipótesis alternativa:  $H_1 : \mu_{prueba} - \mu_{inicial} \neq 0$  (Hay un cambio de comportamiento significativo)

Diseño del estudio: Se realizaría un estudio de campo. Las hormigas serían observadas y su velocidad sería medida en dos momentos: antes de un fenómeno climático (por ejemplo, una tormenta) y durante un período de tiempo de control sin ningún fenómeno climático inminente.

#### ii. Selección del estadísticos de prueba

Análisis estadístico: Se utilizaría una prueba t de Student para comparar las medias de las velocidades de las hormigas en los dos grupos y determinar si hay una diferencia estadísticamente significativa entre ellos, ya que no podemos asegurar que ambos desvíos estándar (necesarios para el cálculo de una Gaussiana) sean representativos de los poblacionales.

#### iii. Selección del nivel de significancia (alpha)

Vamos a utilizar el nivel de significancia utilizado por convención,  $\alpha = 0,05$ .

#### iv. Cálculo de la regla de decisión

A- Si el valor  $p \leq \alpha$ , entonces rechazas la hipótesis nula. Esto significa que los resultados son estadísticamente significativos y hay suficiente evidencia para apoyar la hipótesis

alternativa de que las hormigas cambian su comportamiento medido por su velocidad previo a un fenómeno climático.

B- Si el valor  $p > \alpha$ , entonces no puedes rechazar la hipótesis nula. Esto significa que los resultados no son estadísticamente significativos y no hay suficiente evidencia para apoyar la hipótesis alternativa.

Es importante recordar que la prueba de hipótesis no prueba la hipótesis nula o alternativa directamente, te da una forma de inferir si los resultados de tus estudios proporcionan evidencia fuerte contra la hipótesis nula.

## 2. Indica la respuesta correcta de este pequeño test y justifica tu respuesta (1pt):

A- ¿Cuál es el propósito del valor crítico en un test de hipótesis?

**3. Marcar el límite entre rechazar y no rechazar la hipótesis nula:** El valor crítico en un test de hipótesis es un punto en la escala de la estadística de prueba que divide la región donde rechazaríamos la hipótesis nula de la región donde no la rechazaríamos. Su estimación se ve influenciada por el valor de  $\alpha$  que seleccionemos, como así también el tipo de hipótesis que planteemos: una cola a derecha, una cola a izquierda y doble cola. Si el valor de la estadística de prueba cae en la región de rechazo (más allá del valor crítico), rechazamos la hipótesis nula en favor de la hipótesis alternativa. Si el valor del estadístico de prueba cae en la región de no rechazo (un valor menor que el valor crítico), no rechazamos la hipótesis nula. Por lo tanto, el valor crítico es un umbral que nos ayuda a decidir si debemos o no rechazar la hipótesis nula.

B- ¿Qué representa el Error Tipo II?

**2. No rechazar incorrectamente la hipótesis nula:** El Error Tipo II ocurre cuando no rechazamos la hipótesis nula cuando en realidad es falsa. En otras palabras, es el error de no detectar un efecto o diferencia cuando realmente existe uno. Esto puede suceder por varias razones, como un tamaño de muestra insuficiente o una variabilidad excesiva en los datos. La probabilidad de cometer un Error Tipo II se denota comúnmente como  $\beta$ .

```
In [6]: # Instala el paquete ggplot2 si no está instalado
# install.packages("ggplot2")

# Cargar la biblioteca ggplot2
library(ggplot2)

# Parámetros de las distribuciones normales
media_n1 <- 0
varianza_n1 <- 1

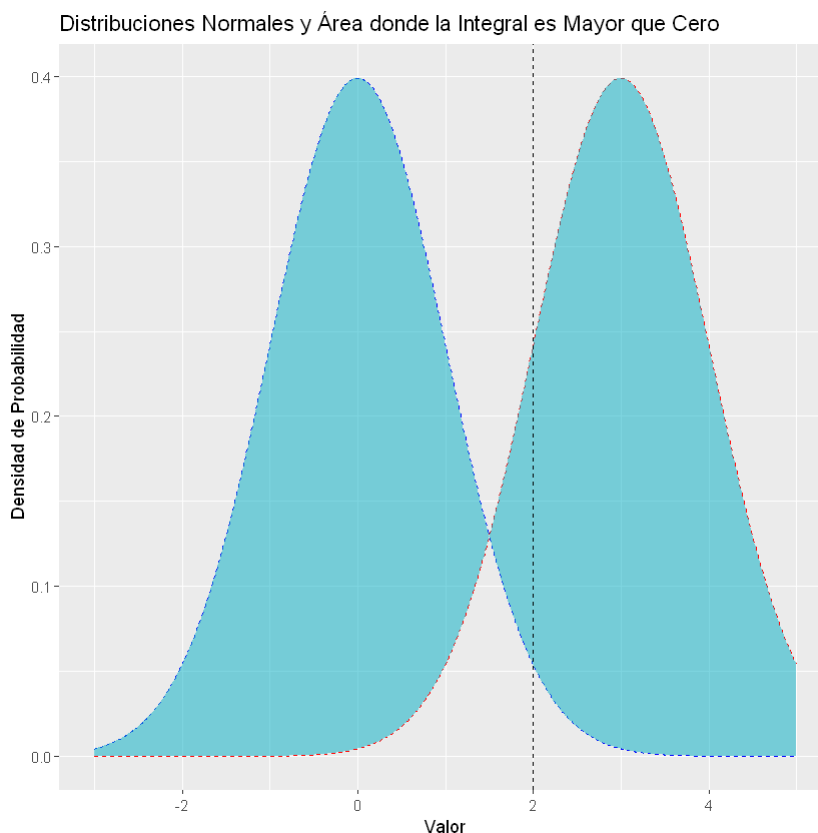
media_n2 <- 3
varianza_n2 <- 1
```

```
# Crear un conjunto de datos con valores de x y las dos distribuciones normales
x <- seq(-3, 5, length.out = 1000)
y_n1 <- dnorm(x, mean = media_n1, sd = sqrt(varianza_n1))
y_n2 <- dnorm(x, mean = media_n2, sd = sqrt(varianza_n2))

# Crear un dataframe para ggplot2
df <- data.frame(x, y_n1, y_n2)

# Crear el gráfico utilizando ggplot2
grafico <- ggplot(df, aes(x)) +
  geom_line(aes(y = y_n1), color = "blue", linetype = 2) +
  geom_line(aes(y = y_n2), color = "red", linetype = 2) +
  geom_ribbon(data = df[df$y_n1 + df$y_n2 > 0, ], aes(ymin = pmin(y_n1, y_n2), y
  geom_vline(xintercept = 2, linetype = "dashed", color = "black") + # Línea ve
  labs(title = "Distribuciones Normales y Área donde la Integral es Mayor que Ce

# Mostrar el gráfico
print(grafico)
```



**C- Si realizas un test de una cola y obtienes un p-valor de 0.02 con un nivel de significancia de 0.05, ¿qué decisión tomarías?**

**1. Rechazar la hipótesis nula:** En un test de hipótesis, si el p-valor es menor que el nivel de significancia ( $\alpha$ ), entonces rechazamos la hipótesis nula. En este caso, dado que el p-valor (0.02) es menor que el nivel de significancia (0.05), rechazaríamos la hipótesis nula. Esto sugiere que hay evidencia suficiente para apoyar la hipótesis alternativa a un nivel de significancia del 5%, o con un nivel de confianza del 95%.

**D- ¿Cuál es la relación entre nivel de significancia y la probabilidad de cometer un Error Tipo I?**

El nivel de significancia, denotado comúnmente por la letra griega  $\alpha$  (alfa), es la probabilidad de cometer un Error Tipo I en un test de hipótesis estadísticas. El Error Tipo I ocurre cuando se rechaza incorrectamente una hipótesis nula que es verdadera.

Entonces, la relación es directa: a medida que aumenta el nivel de significancia ( $\alpha$ ), la probabilidad de cometer un Error Tipo I también aumenta. Esto se debe a que un nivel de significancia más alto implica que estás siendo más permisivo al rechazar la hipótesis nula. Al fijar un nivel de significancia, estás estableciendo el umbral para considerar un resultado como estadísticamente significativo. Un nivel de significancia comúnmente utilizado es 0.05, lo que significa que estás dispuesto a aceptar un 5% de probabilidad de cometer un Error Tipo I.

Es importante encontrar un equilibrio al elegir el nivel de significancia, ya que un nivel demasiado alto puede llevar a conclusiones incorrectas basadas en resultados no significativos, mientras que un nivel demasiado bajo puede hacer que sea difícil encontrar efectos reales. La elección del nivel de significancia a menudo depende del contexto específico y de la tolerancia hacia los errores estadísticos en un estudio partic.

Por lo tanto, **ninguna de las opciones es correcta.**

## Parte II

### 1. Exploración de los datos: análisis descriptivos de los datos, graficación y conclusiones ¿siguen una distribución normal?

#### Cargamos las librerías

```
In [7]: library(dplyr)
library(ggplot2)
library(gridExtra)
library(cowplot)
library(nortest)
library(fmsb)
library(pastecs)
```

#### Cargamos el data set

```
In [8]: data <- read.csv(file='ejercicio2.csv', sep=",")
head(data)
tail(data)
```

A data.frame: 6 × 3

	<b>ï..Reference</b>	<b>Order.Value</b>	<b>Group</b>
	<b>&lt;int&gt;</b>	<b>&lt;dbl&gt;</b>	<b>&lt;int&gt;</b>
<b>1</b>	570554	0.38	1
<b>2</b>	567869	0.40	0
<b>3</b>	539441	0.42	0
<b>4</b>	542736	0.55	0
<b>5</b>	573589	0.55	0
<b>6</b>	540833	0.65	1

A data.frame: 6 × 3

	<b>ï..Reference</b>	<b>Order.Value</b>	<b>Group</b>
	<b>&lt;int&gt;</b>	<b>&lt;dbl&gt;</b>	<b>&lt;int&gt;</b>
<b>11781</b>	578315	9467.23	0
<b>11782</b>	551698	9468.60	1
<b>11783</b>	556090	9471.27	0
<b>11784</b>	537253	9473.16	1
<b>11785</b>	554094	9475.20	1
<b>11786</b>	557888	9476.80	1

## Realizamos estadística descriptiva de nuestras variables

```
In [9]: data$Order.Value <- as.integer(data$Order.Value)
stat.desc(data)
media = round(mean(data$Order.Value),0)
cat("La media de euros que se gasta en cada compra es de aproximadamente: $", me
```

A data.frame: 14 × 3

	i..Reference	Order.Value	Group
	<dbl>	<dbl>	<dbl>
<b>nbr.val</b>	1.178600e+04	1.178600e+04	1.178600e+04
<b>nbr.null</b>	0.000000e+00	1.600000e+01	5.502000e+03
<b>nbr.na</b>	0.000000e+00	0.000000e+00	0.000000e+00
<b>min</b>	5.363650e+05	0.000000e+00	0.000000e+00
<b>max</b>	5.815870e+05	9.476000e+03	1.000000e+00
<b>range</b>	4.522200e+04	9.476000e+03	1.000000e+00
<b>sum</b>	6.586996e+09	3.658161e+07	6.284000e+03
<b>median</b>	5.586265e+05	2.376500e+03	1.000000e+00
<b>mean</b>	5.588830e+05	3.103819e+03	5.331750e-01
<b>SE.mean</b>	1.212238e+02	2.585259e+01	4.595652e-03
<b>CI.mean.0.95</b>	2.376186e+02	5.067535e+01	9.008238e-03
<b>var</b>	1.731977e+08	7.877247e+06	2.489205e-01
<b>std.dev</b>	1.316046e+04	2.806643e+03	4.989194e-01
<b>coef.var</b>	2.354779e-02	9.042550e-01	9.357517e-01

La media de euros que se se gasta en cada compra es de aproximadamente: \$ 3104

## Descripción de los Datos

En nuestro conjunto de datos, contamos con 1,785 observaciones. Cada línea observada corresponde a una orden de compra única identificada por la variable "i..Reference".

Al realizar una descripción, enfocamos nuestra atención en la variable de interés "Order.Value", la cual determinará nuestro test de hipótesis. Esta variable tiene una **media** de aproximadamente 3,104 euros por compra. La **mediana**, en cambio, es de aproximadamente 2,376. Podemos suponer que la distribución de "Order.Value" tiene una **asimetría positiva** o hacia la derecha, indicando una cola derecha más larga, ya que el valor de la **media** es superior al de la **mediana**.

El **rango** de la distribución es considerablemente amplio, aproximadamente de 9,476, y la **varianza** es elevada, alrededor de 7,877. Estos valores sugieren una distribución con variabilidad significativa, lo que podrían indicar una distribución **ancha** de nuestra variable.

Luego de estas primeras descripciones, procederemos a graficar la distribución para una mejor visualización. zación.

## Gráficos

## Empecemos a hacer un analisis por grupo de variables. Los que usaron el identificador, y los que no lo han usado

```
In [11]: # Veamos una tabla con los estadísticos comparados por grupo
estadisticos <- data %>%
  group_by(Group) %>%
  summarize(
    mean = mean(Order.Value),
    sd = sd(Order.Value),
    median = median(Order.Value)
  )
print(estadisticos)

#Dividimos nuestro data set por grupos, calculamos sus estadísticos y los compar

estadisticas_inicial <- subset(data, Group == 0)
estadisticas_prueba1 <- subset(data, Group == 1)
estadisticas_inicial$media <- mean(estadisticas_inicial$Order.Value)
estadisticas_prueba1$media <- mean(estadisticas_prueba1$Order.Value)
estadisticas_inicial$varianza <- var(estadisticas_inicial$Order.Value)
estadisticas_prueba1$varianza <- var(estadisticas_prueba1$Order.Value)
estadisticas_prueba1$Group <- 1
estadisticas_inicial$Group <- 0
df_estadist_comb <- rbind(estadisticas_prueba1, estadisticas_inicial)

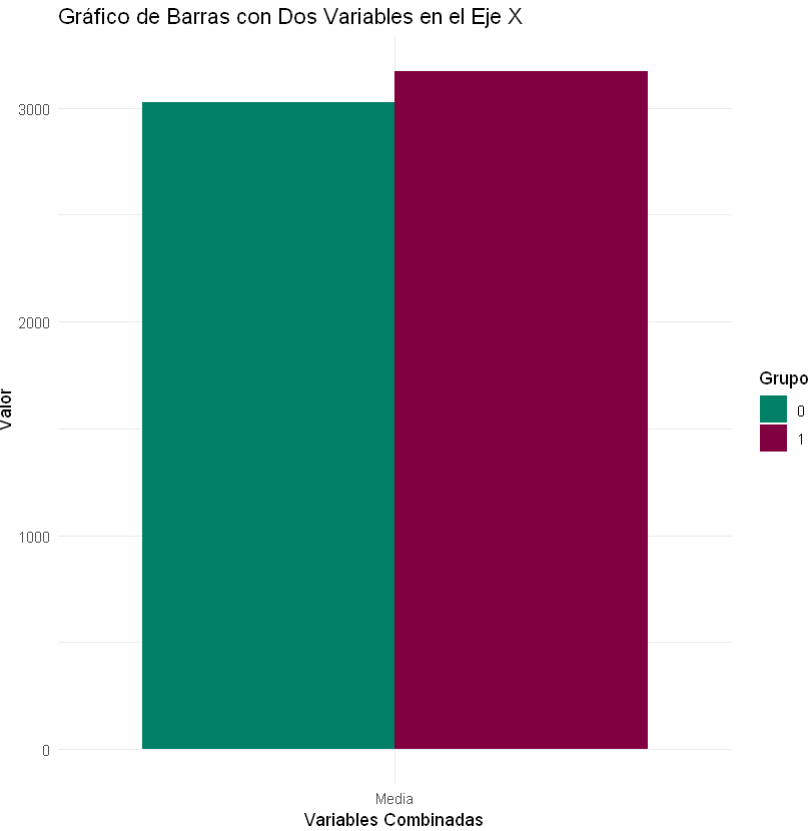
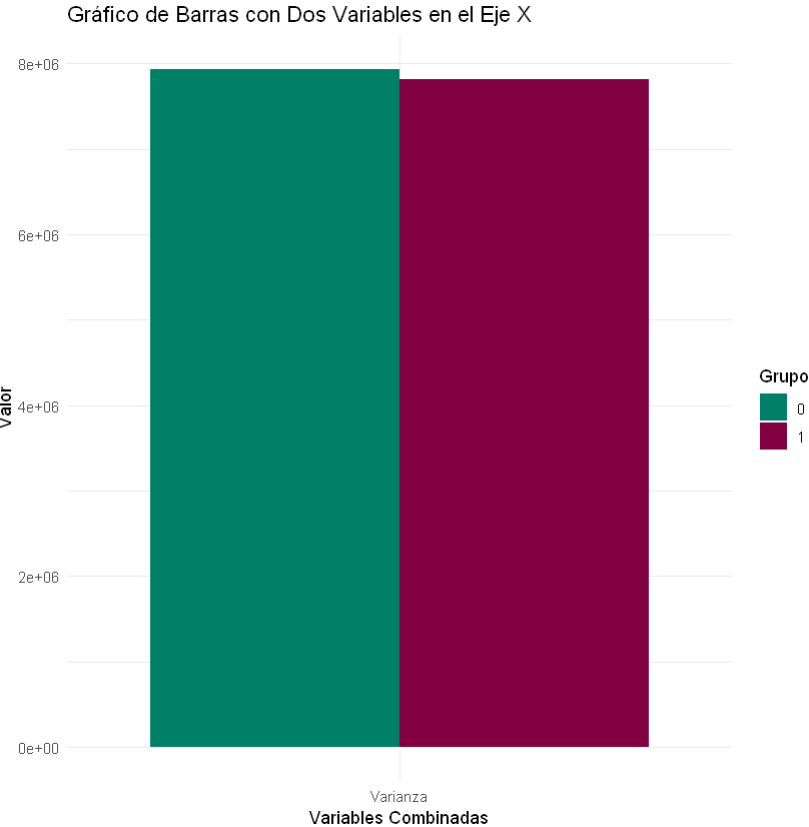
# Crear el gráfico de barras

colores = c("#008066", "#800040")

# Crear el gráfico de barras
a = ggplot(df_estadist_comb) +
  geom_bar(aes(x = "Varianza", y = varianza, fill = factor(Group)),
    position = "dodge", stat = "identity") +
  scale_y_continuous(name = "Valor") +
  scale_x_discrete(name = "Variables Combinadas") +
  labs(
    title = "Gráfico de Barras con Dos Variables en el Eje X",
    fill = "Grupo"
  ) +
  scale_fill_manual(values = colores) + # Personalizar colores
  theme_minimal()

# Crear el gráfico de barras
b = ggplot(df_estadist_comb) +
  geom_bar(aes(x = "Media", y = media, fill = factor(Group)),
    position = "dodge", stat = "identity") +
  scale_y_continuous(name = "Valor") +
  scale_x_discrete(name = "Variables Combinadas") +
  labs(
    title = "Gráfico de Barras con Dos Variables en el Eje X",
    fill = "Grupo"
  ) +
  scale_fill_manual(values = colores) + # Personalizar colores
  theme_minimal()
print(a)
print(b)
```

```
# A tibble: 2 x 4
  Group mean    sd median
<int> <dbl> <dbl> <dbl>
1     0 3024. 2818. 2170.
2     1 3173. 2795. 2526.
```



Descripción del grafico de estadisticos comparados



La idea de este gráfico es mostrar que a simple vista no podemos ver una diferencia significativa entre medias y varianzas entre ambos grupos. Si vemos que el grupo de prueba del sistema tiene una media mayor y una menor varianza, que en principio es un buen indicio, pero no puede ser concluyente, dado que necesitamos indagar con una mayor rigurosidad la significatividad de nuestro experimento.

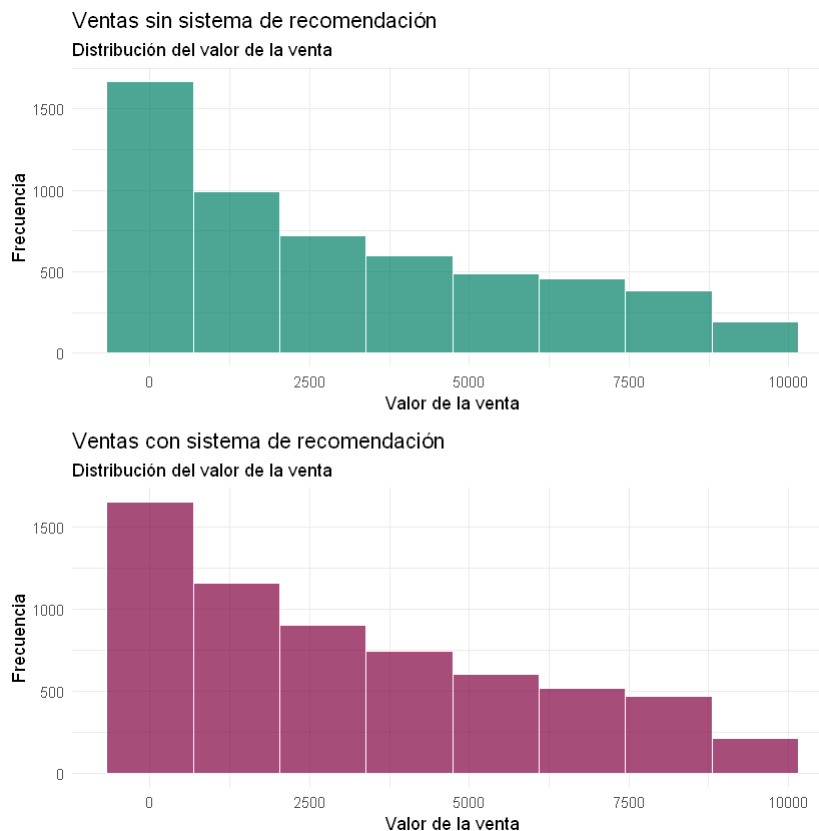
## Grupo de ventas sin asistencia REVISAR

```
In [12]: inicial <- subset(data, Group == 0)
prueba1 <- subset(data, Group == 1)

# Plot para 'inicial'
a = plot_inicial_histogram1 <- ggplot(inicial, aes(x = Order.Value)) +
  geom_histogram(bins = 8, fill = "#008066", color = "white", alpha = 0.7) +
  labs(title = "Ventas sin sistema de recomendación",
        subtitle = "Distribución del valor de la venta",
        x = "Valor de la venta",
        y = "Frecuencia") +
  theme_minimal()

b = plot_prueba_histogram1 <- ggplot(prueba1, aes(x = Order.Value)) +
  geom_histogram(bins = 8, fill = "#800040", color = "white", alpha = 0.7) +
  labs(title = "Ventas con sistema de recomendación",
        subtitle = "Distribución del valor de la venta",
        x = "Valor de la venta",
        y = "Frecuencia") +
  theme_minimal()

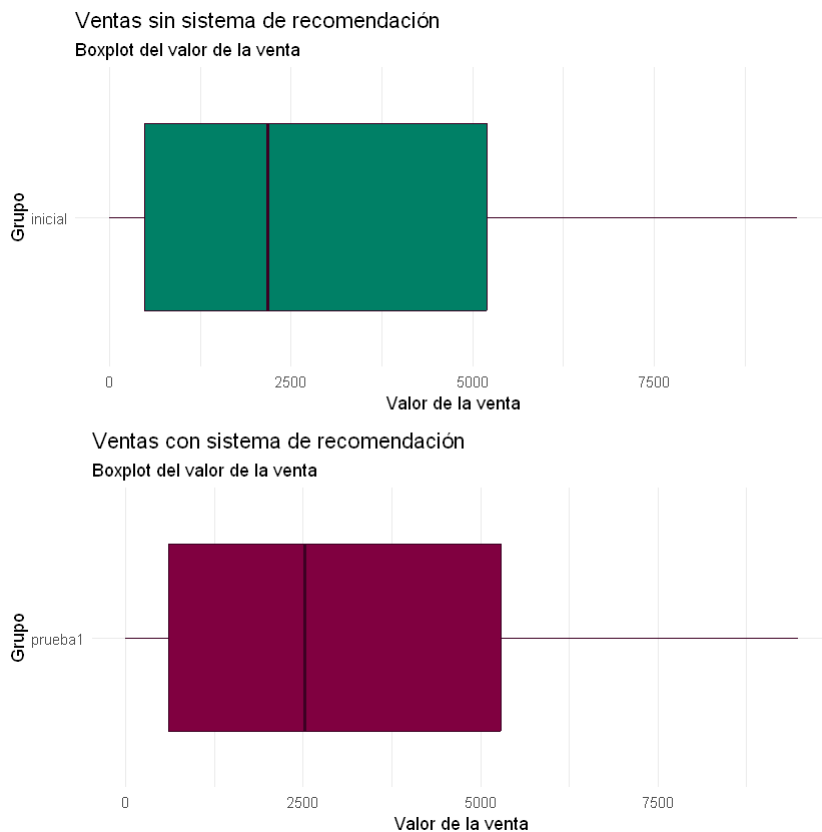
# Mostrar todos los gráficos en la misma salida de manera ordenada
grid.arrange(a, b)
```



```
In [13]: # Plot para 'inicial'
c = plot_inicial_boxplot1 <- ggplot(inicial, aes(x = "inicial", y = Order.Value))
  geom_boxplot(fill = "#008066", color = "#3c0023") +
  labs(title = "Ventas sin sistema de recomendación",
        subtitle = "Boxplot del valor de la venta",
        x = "Grupo",
        y = "Valor de la venta") +
  theme_minimal()+
  coord_flip()

# Plot para 'prueba1'
d = plot_prueba_boxplot1 <- ggplot(prueba1, aes(x = "prueba1", y = Order.Value))
  geom_boxplot(fill = "#800040", color = "#3c0023") +
  labs(title = "Ventas con sistema de recomendación",
        subtitle = "Boxplot del valor de la venta",
        x = "Grupo",
        y = "Valor de la venta") +
  theme_minimal()+
  coord_flip()

# Mostrar todos los gráficos en la misma salida de manera ordenada
grid.arrange(c, d)
```



Se puede ver una diferencia entre ambos historiogramas y box plots:

- Ambas variables tienen una distribución que en principio parece **normal**
- Ambos grupos presentan una **sesgo a derecha (asimetría positiva)**.
- No se observan valores atípicos.

Observemos que sucede si vemos las dos funciones de densidad en un mismo gráfico:

```
In [14]: # Creamos el gráfico con las líneas de densidad usando ggplot para ambas pruebas
combined_plot <- ggplot() +
```

```

geom_density(data = prueba1, aes(x = Order.Value, color = "Prueba 1"), size = 3)
geom_density(data = inicial, aes(x = Order.Value, color = "Inicial"), size = 3)
labs(
  title = "Líneas de Densidad de 'Order.Value'",
  x = "Valor de la Orden",
  y = "Densidad",
  caption = "Nota: Aquí podemos ver lo que se aproximó previamente observando"
) +
theme_minimal() +
theme(
  plot.title = element_text(size = 16, hjust = 0.5, face = "bold"),
  axis.title.x = element_text(size = 14),
  axis.title.y = element_text(size = 14),
  axis.text = element_text(size = 12),
  plot.caption = element_text(size = 11, color = "gray")
) +
scale_color_manual(values = c("#008066", "#800040"), name = "Variable")

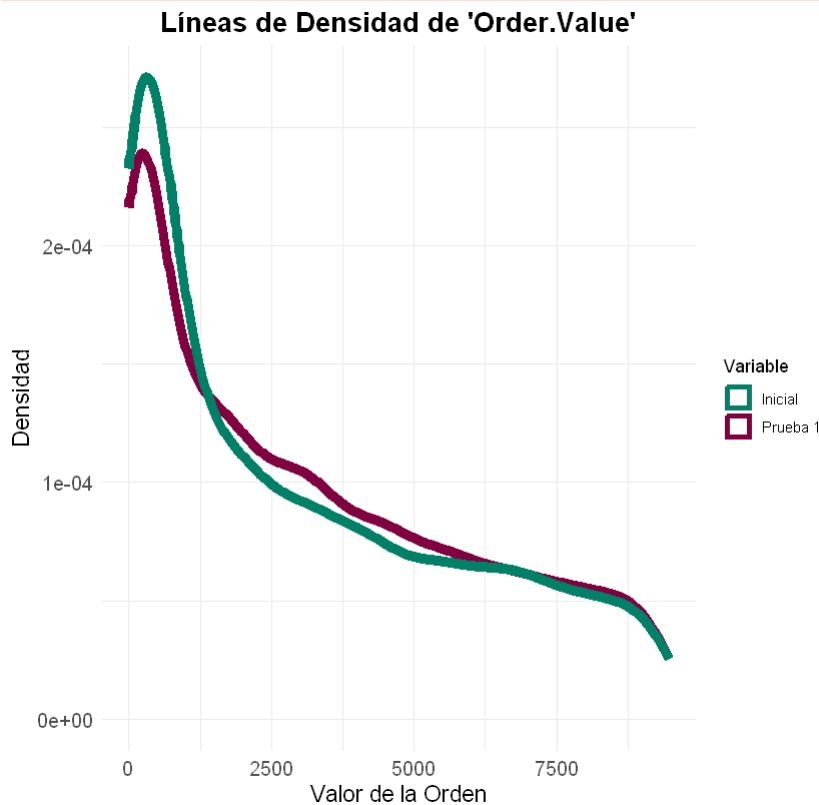
# Mostramos el gráfico combinado con Leyenda
print(combined_plot)

```

Warning message:

"Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.

[i](#) Please use `linewidth` instead."



## Descripción de la distribución

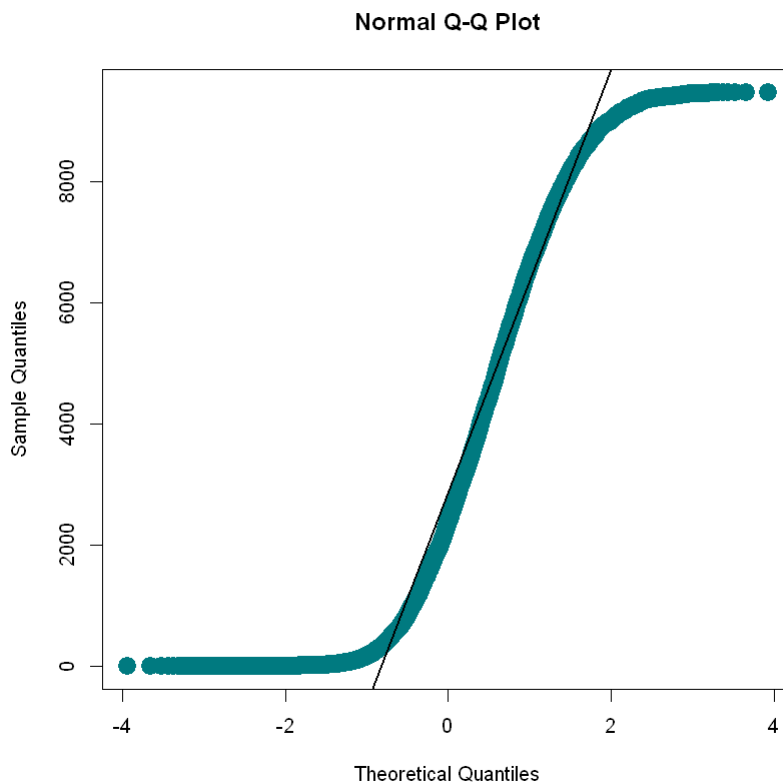
Luego de realizar la distribución de densidad de nuestra variable "Order.Value", para el grupo de ventas sin asistencia del sistema de recomendación, podemos ver que efectivamente cuentan con:

- **Asimetría positiva:** la cola derecha efectivamente es más larga que la izquierda, representado por una media mayor a la mediana.

- **Distribucion Ancha:** tenemos una distribucion con gran variabilidad, representada por un elevado rango y varianza

En conclusion, siempre es sumamente util realizar un analisis detallado de nuestras estadísticas descriptiva, para intentar ver a priori como se debería comportar nuestra funcion de distribucion. Luego, realizamos un grafico para ver si nuestra lectura es correcta o si esta pasando algo distinto que no supimos ver con los estadísticos descriptos.

```
In [15]: # Crear el QQ-plot
qqnorm(data$Order.Value, pch = 19, col = "#007a80", cex=2)
qqline(data$Order.Value, col = "black", lwd = 2)
```



## Conclusión

### Descripción del grafico de estadísticos comparados con funciones de densidad

Luego de analizar los estadísticos descriptivos de cada grupo, de ver una simple representacion grafica de las diferencias entre ambos grupos, y de armar el grafico conjunto de las funciones de densidades para cada grupo, podemos ver dos cosas:

1- Existe una **diferencia** respecto de la distribucion sin sistema y con sistema. Sin embargo, esto no es suficiente para poder concluir respecto a nuestra hipotesis, necesitamos indagar con una mayor profundidad. Para ello, se procedera a realizar un test de hipotesis de diferencia de medias.

2- Segun nuestro QQ-Plot, pareciera que nuestra distribucion es **normal**. Sin embargo, debemos adentrarnos en el analisis de nuestra distribucion, viendo si aplica o no el

Teorema del Limite Central y haciendo algun test de normalidad.

## Normalidad

Vamos a evaluar la normalidad de nuestra distribucion:

```
In [16]: resultados_test <- ad.test(data$Order.Value)
print(resultados_test)
```

Anderson-Darling normality test

data: data\$Order.Value  
A = 388.16, p-value < 2.2e-16

```
In [17]: k_test = ks.test(data$Order.Value, "pnorm", mean = mean(data$Order.Value), sd =
print(k_test)
```

Warning message in ks.test(data\$Order.Value, "pnorm", mean = mean(data\$Order.Value), :  
"ties should not be present for the Kolmogorov-Smirnov test"

One-sample Kolmogorov-Smirnov test

data: data\$Order.Value  
D = 0.13439, p-value < 2.2e-16  
alternative hypothesis: two-sided

### Resultado del test:

Realizamos el test de normalidad de Anderson-Darling, ya que para mas de 5000 observaciones no podemos utilizar el test de Shapiro. Dado que el valor p es prácticamente cero, rechazaríamos la hipótesis nula en favor de la hipótesis alternativa. En otras palabras, hay evidencia significativa para afirmar que los datos no siguen una distribución normal. Agregamos el test de Kolmogorov-Smirnov para asegurar nuestro resultado.

### Teorema del Limite Central (TCL):

Sin embargo, través del Teorema del Límite Central, podemos suponer que cualquier distribución que sea una suma de variables independientes con media y varianza finita tiene una distribución normal. En este caso, tenemos dos bases de datos de compras que podemos suponer, a priori, que son independientes una de la otra. Al asumir esto como correcto, podemos tratar las distribuciones como gaussianas.

## 2. Diseña un test de hipótesis: ¿cuál es nuestro objetivo? ¿se trataría de un test de 2 colas? ¿Qué estadístico debemos usar para este Test?(2pts)

### Objetivo y definicion del test:

Para poder realizar una inferencia de manera critica debemos realizar un test de hipotesis, que nos permita conocer, dadas las muestras que tenemos y los estandares que fijemos, si estamos en condiciones o no de decir que es conveniente utilizar un sistema de recomendacion en el ecommerce. Para este test debemos optar por una diferencia de medias, para conocer precisamente, si la media del valor de las ventas con la participacion del sistema de recomendacion es estrictamente mayor a la media del valor de las ventas sin utilizar este sistema. Este test de diferencias es de una sola cola, es decir, vamos a plantear si la diferencia entre la media con sistema de recomendacion y la media sin sistema de recomendacion es significativamente distinta de cero. Sin embargo, se podria plantear un test a dos colas, viendo si hay alguna diferencia entre aplicar el sistema o no. Es decir, si el sistema mejoro o empeoro significativamente el valor de las ventas. En este ejercicio, decidimos plantear usar un test de hipotesis de una cola.

### Definicion del estadistico:

- **Repasemos nuestro escenario:** Tenemos 11786 observaciones, luego de asumir la normalidad por el TCL y viendo la cantidad de datos que tenemos, podemos construir un estadistico **normal**. Como en este caso optamos por hacer un test de diferencia de medias, debemos realizar el calculo de nuestro estadistico de la siguiente manera:

$$Z = \frac{\bar{x}_{H1} - \bar{x}_{H0}}{\sqrt{\frac{S_{H1}^2}{n_{H1}} + \frac{S_{H0}^2}{n_{H0}}}}$$

Utilizamos una T Test de muestra independiente de una cola a izquierda para las medias:

1. Hipótesis nula:  $H_0 : \mu_{inicial} - \mu_{prueba} \geq 0$  (Hay un impacto significativo al usar el sistema de recomendacion)
2. Hipótesis alternativa:  $H_1 : \mu_{inicial} - \mu_{prueba} < 0$  (No hay un impacto significativo al usar el sistema de recomendacion)

D Con lo cual debemos proceder a hacer algunos calculos auxiliares:

```
In [26]: media_inicial <- mean(inicial$Order.Value)
media_prueba <- mean(prueba1$Order.Value)
varianza_inicial <- var(inicial$Order.Value)
varianza_prueba <- var(prueba1$Order.Value)
sd_inicial <- sd(inicial$Order.Value)
sd_prueba <- sd(prueba1$Order.Value)
observaciones_inicial <- nrow(inicial)
observaciones_prueba <- nrow(prueba1)
```

```

denominador_prueba_z <- (varianza_prueba)/(observaciones_prueba)
denominador_inicial_z <- (varianza_inicial)/(observaciones_inicial)
denominador_z <- sqrt(denominador_prueba_z + denominador_inicial_z)
numerador_z <- (media_inicial) - (media_prueba)
estadistico_z <- (numerador_z)/(denominador_z)
valor_p = pnorm(estadistico_z)

print("Estadístico Zc:")
print(round(estadistico_z, 2))
print("P-valor")
print(round(valor_p, 3))

```

```

[1] "Estadístico Zc:"
[1] -2.88
[1] "P-valor"
[1] 0.002

```

```

In [27]: # Estadístico Z y valor p
nivel_significancia = 0.05

# Valor crítico superior (cola derecha)
z_critico_superior = qnorm(1 - nivel_significancia)
z_critico_inferior = qnorm(nivel_significancia)
print(z_critico_superior)

# Generamos y dibujamos una curva Z (distribución normal estándar)
x = seq(-4, 4, length.out = 1000)
y = dnorm(x)
plot(x, y, type = "l", lwd = 6, col = "008066", xlab = "Valor Z", ylab = "Densi

# Pintamos el área bajo la curva (zona de rechazo)
x_area_rechazo = x[x <= z_critico_inferior]
y_area_rechazo = dnorm(x_area_rechazo)
# Escalar y_area_rechazo para que esté por debajo de la curva
y_max = max(y)
y_area_rechazo = dnorm(x_area_rechazo)
polygon(c(z_critico_superior, x_area_rechazo, max(x_area_rechazo)), c(0, y_area_

# Línea vertical en el estadístico Z
abline(v = estadistico_z, col = "#800040", lty = 1, lwd = 4)

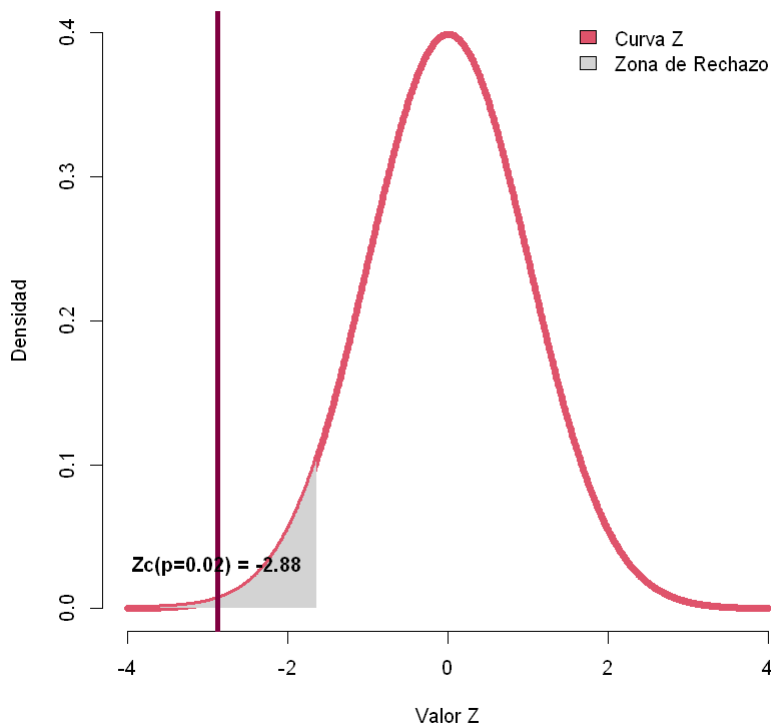
# Añadimos valor p
text(estadistico_z, 0.03, paste("Zc(p=0.02) =", round(estadistico_z, 2)), col =

# Título y Leyenda
title("Distribución Normal Estándar con Zona de Rechazo", cex.main = 1.5)
legend("topright", legend = c("Curva Z", "Zona de Rechazo"), fill = c("008066",

[1] 1.644854

```

## Distribución Normal Estándar con Zona de Rechazo



## Resultado

Al calcular `pnorm(estadístico_z)`, obtenemos la probabilidad acumulativa asociada al estadístico Z calculado (-2.88). Este valor representa la probabilidad de observar un valor igual o menor al estadístico Z en una distribución normal estándar.

El estadístico Z está bastante lejano del centro de la distribución y se encuentra en la cola izquierda de la distribución normal estándar.

Dado un nivel de significatividad de 0,05 podríamos considerar rechazar la hipótesis nula en favor de la alternativa. Sin embargo, la interpretación completa de los resultados depende del contexto y del diseño de la prueba.

Es relevante ahora estimar si el número de pruebas realizado es suficiente para realizar inferencias con mayor confianza.

```
In [31]: power.t.test(power = .95,          # Potencia deseada
                    delta = sqrt(varianza_inicial), #Tamano esperado al menos una dif
                    sd = varianza_inicial - varianza_prueba, # Desviación estándar de
                    type = "t",              # Tipo de test (t-test)
                    alternative = "one.sided", # Dirección de la alternativa (una col
                    sig.level = .05)          # Nivel de significancia
```



Two-sample t test power calculation

```
n = 43194.14
delta = 2817.712
sd = 125873.2
sig.level = 0.05
power = 0.95
alternative = one.sided
```

NOTE: n is number in *each* group

## Conclusión general

A partir de los datos proporcionados, se puede concluir que, estadísticamente, **hay un cambio significativo en el valor promedio de las ordenes utilizando el sistema de recomendacion.**

Sin embargo, debemos remarcar que la prueba se llevo a cabo con un tamaño de muestra menor al requerido. Es importante recordar que la prueba de hipótesis no prueba la hipótesis nula o alternativa directamente, te da una forma de inferir si los resultados de tus estudios proporcionan evidencia fuerte contra la hipótesis nula.

In [ ]: