

# Informe - Artículo Racial Stereotypes

Juan David Roa - Laura Camila Rodriguez

2025-11-15

## Introducción

Este informe presenta un análisis exploratorio de datos (EDA) basado en un estudio experimental que evalúa la presencia de sesgo racial implícito en la identificación rápida de objetos. Específicamente, se examina si los participantes muestran diferencias en sus tiempos de reacción al identificar armas (*guns*) versus herramientas (*tools*) cuando se les presenta previamente una cara de una persona negra (*Black prime*) o blanca (*White prime*). El objetivo es explorar si existe una interacción entre el tipo de *prime* racial y el tipo de objeto, lo cual sería indicativo de un sesgo estereotipado (por ejemplo, asociar más rápidamente armas con personas negras).

---

## Carga de librerías y estructura de los datos

```
library(dplyr)
library(ggplot2)
library(car)
library(nortest)
library(tseries)
library(ggpubr)
library(readr)
library(stringr)
library(afex)

data_raw <- read_delim("ANOVA beh RT.csv", delim = ",",
  col_names = TRUE)

data_limpia <- data_raw %>%
  mutate(
    rt_log = str_replace_all(rt_log, "\\\s+", ""),
    rt_log = str_replace(rt_log, "\\.(?=.*\\.\\.)", ""),
    rt_log = as.numeric(rt_log)
  )

data_limpia$id      <- as.factor(data_limpia$id)
data_limpia$prime   <- as.factor(data_limpia$prime)
data_limpia$target <- as.factor(data_limpia$target)
str(data_limpia)

## tibble [720 x 7] (S3:tbl_df/tbl/data.frame)
## $ ...1 : num [1:720] 1 2 3 4 5 6 7 8 9 10 ...
## $ id    : Factor w/ 30 levels "2","3","4","5",...: 1 1 1 1 1 1 1 1 1 ...
## $ run   : num [1:720] 1 1 1 1 2 2 2 2 3 3 ...
```

```

## $ prime : Factor w/ 2 levels "Black","White": 1 1 2 2 1 1 2 2 1 1 ...
## $ target: Factor w/ 2 levels "gun","tool": 1 2 1 2 1 2 1 2 1 2 ...
## $ rt_raw: num [1:720] 619 610 657 605 628 ...
## $ rt_log: num [1:720] 6.41 6.4 6.46 6.4 6.43 ...
summary(data_limpia)

##      .1          id        run     prime    target
## Min.   : 1.0   2   : 24   Min.   :1.0   Black:360   gun :360
## 1st Qu.:180.8  3   : 24   1st Qu.:2.0  White:360   tool:360
## Median :360.5  4   : 24   Median  :3.5
## Mean   :360.5  5   : 24   Mean   :3.5
## 3rd Qu.:540.2  6   : 24   3rd Qu.:5.0
## Max.   :720.0  7   : 24   Max.   :6.0
##                   (Other):576
##      rt_raw       rt_log
## Min.   :356.8   Min.   :5.868
## 1st Qu.:461.9   1st Qu.:6.125
## Median :508.3   Median :6.218
## Mean   :522.9   Mean   :6.233
## 3rd Qu.:569.4   3rd Qu.:6.326
## Max.   :836.2   Max.   :6.695
##

```

**Comentario:** Se verifican las variables y la ausencia de valores faltantes. El conjunto de datos contiene las columnas `prime`, `target`, `rt_raw` y `rt_log`, listas para el análisis.

---

## 1. Verificación del diseño experimental (balance)

```
table(data_limpia$prime, data_limpia$target)
```

```

##           gun tool
## Black 180 180
## White 180 180

```

**Comentario:** El diseño está balanceado: cada combinación de `prime` y `target` tiene el mismo número de observaciones, lo cual es necesario para un análisis de varianza válido.

---

## 2. Exploración de la distribución de los tiempos de reacción

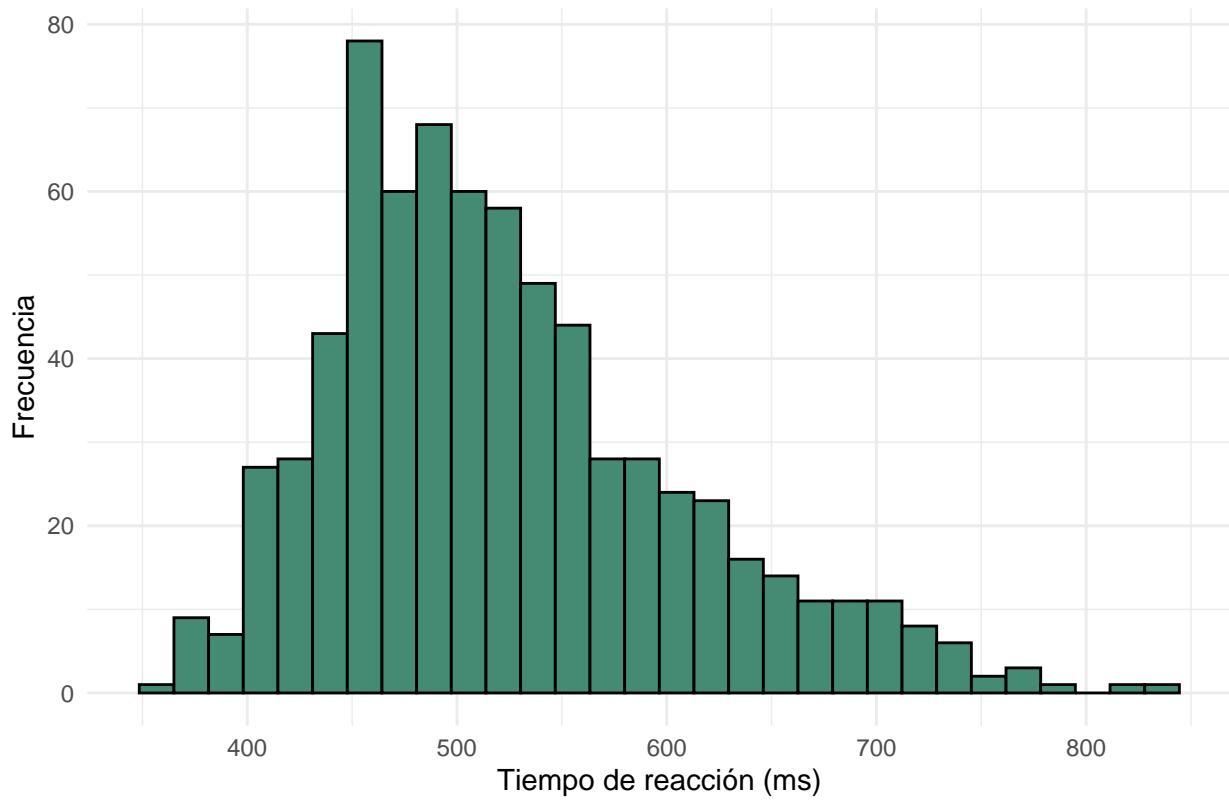
### 2.1. Datos brutos (`rt_raw`)

```

Histobrutos <- ggplot(data_limpia, aes(x = rt_raw)) +
  geom_histogram(bins = 30, fill = "#458B74", color = "black") +
  labs(title = "Histograma: Datos Brutos", x = "Tiempo de reacción (ms)", y = "Frecuencia") +
  theme_minimal()
print(Histobrutos)

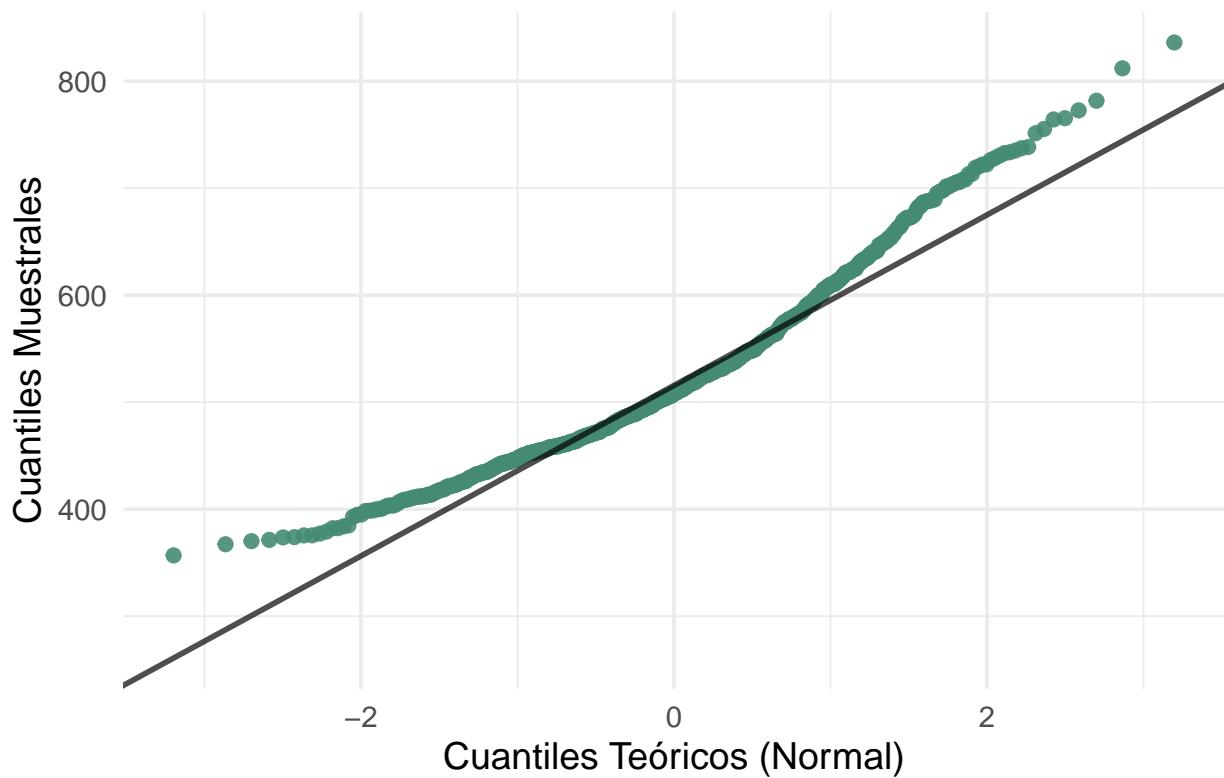
```

## Histograma: Datos Brutos



```
GGbrutos <- ggplot(data_limpia, aes(sample = rt_raw)) +
  geom_qq(color = "#458B74", size = 2, alpha = 0.9) +
  geom_qq_line(color = "black", linewidth = 1.0, alpha = 0.7) +
  labs(title = "Q-Q Plot: Datos Brutos",
       x = "Cuantiles Teóricos (Normal)",
       y = "Cuantiles Muestrales") +
  theme_minimal(base_size = 14)
print(GGbrutos)
```

## Q–Q Plot: Datos Brutos

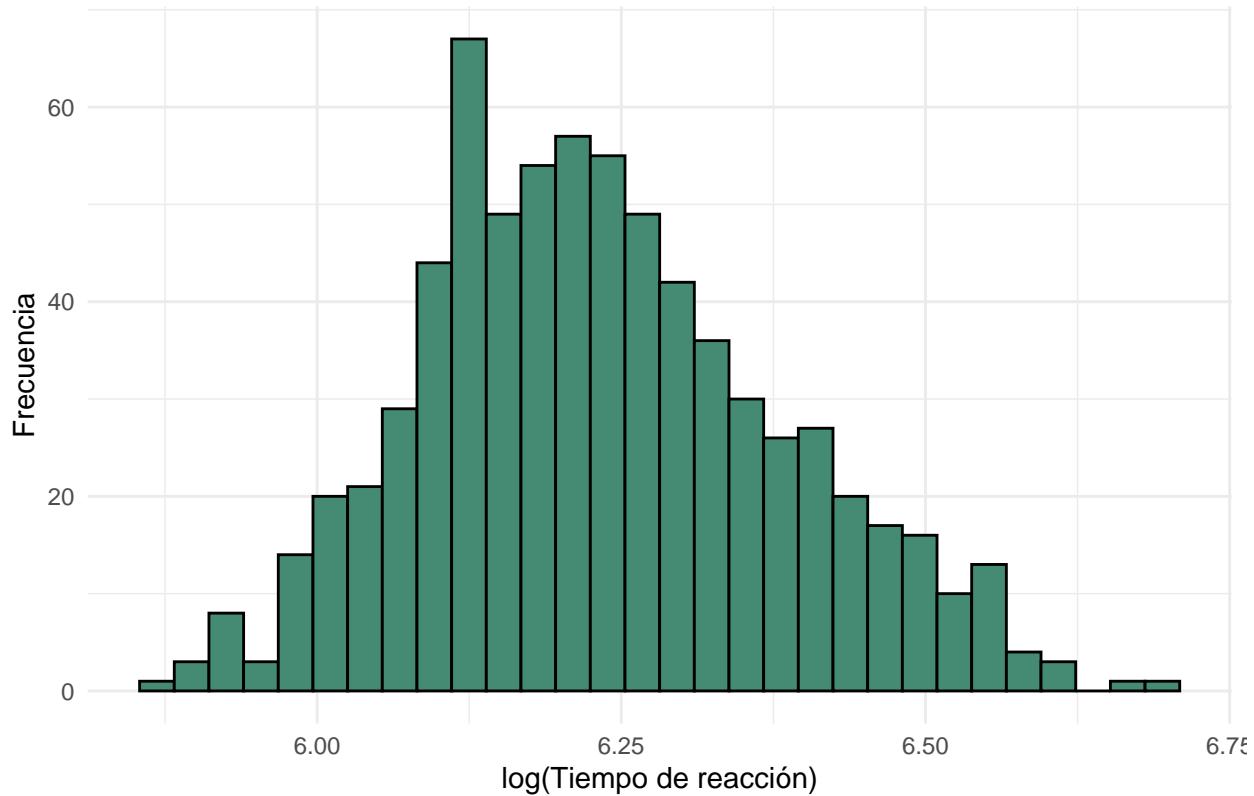


**Comentario:** Los tiempos de reacción brutos muestran una fuerte asimetría positiva, lo que viola el supuesto de normalidad. Se justifica la transformación logarítmica.

### 2.2. Datos transformados (rt\_log)

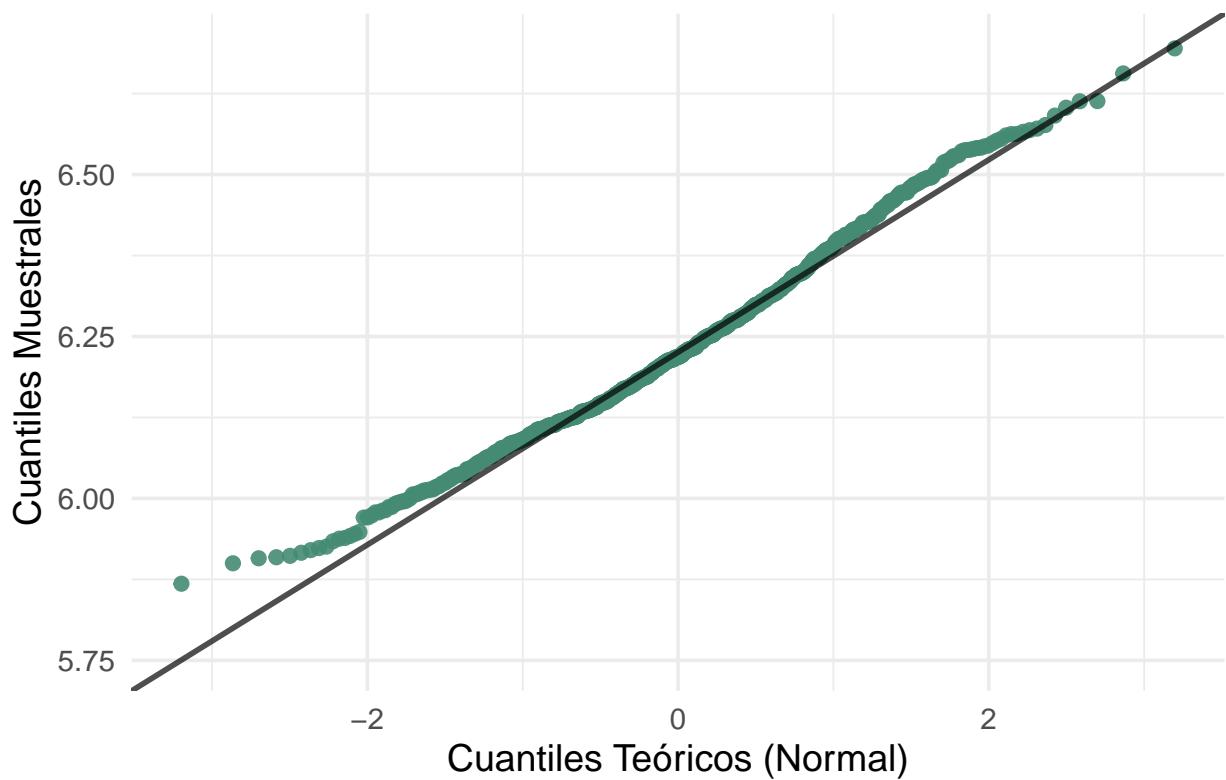
```
HistoTrans <- ggplot(data_limpia, aes(x = rt_log)) +  
  geom_histogram(bins = 30, fill = "#458B74", color = "black") +  
  labs(title = "Histograma: Datos Transformados", x = "log(Tiempo de reacción)", y = "Frecuencia") +  
  theme_minimal()  
print(HistoTrans)
```

## Histograma: Datos Transformados



```
GGTrans <- ggplot(data_limpia, aes(sample = rt_log)) +  
  geom_qq(color = "#458B74", size = 2, alpha = 0.9) +  
  geom_qq_line(color = "black", linewidth = 1.0, alpha = 0.7) +  
  labs(title = "Q-Q Plot: Datos Transformados",  
       x = "Cuantiles Teóricos (Normal)",  
       y = "Cuantiles Muestrales") +  
  theme_minimal(base_size = 14)  
print(GGTrans)
```

## Q-Q Plot: Datos Transformados



**Comentario:** La transformación logarítmica mejora significativamente la normalidad de los datos, haciendo que la distribución se aproxime más a una normal. Es adecuada para análisis paramétricos posteriores.

### 3. Visualización de la interacción esperada

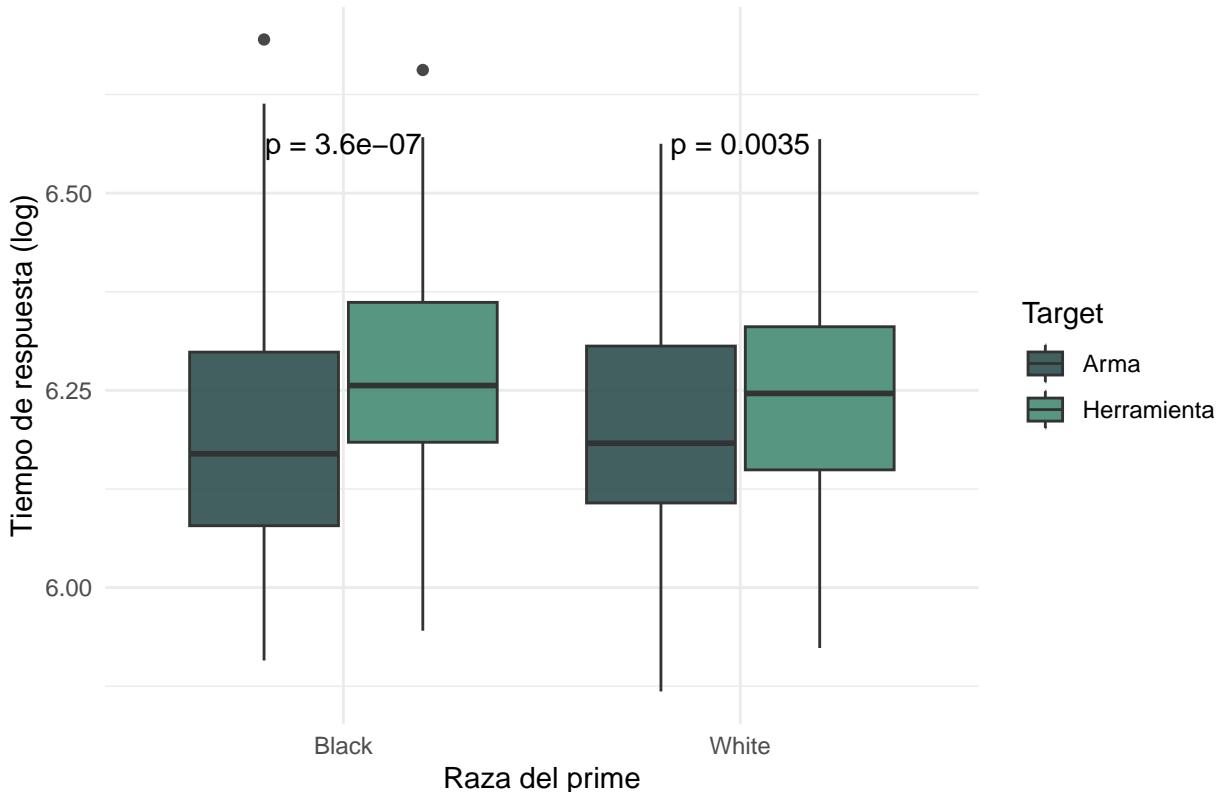
```
CajasObj <- ggplot(data_limpia, aes(x = prime, y = rt_log, fill = target)) +  
  geom_boxplot(alpha = 0.9, position = position_dodge(0.8)) +  
  labs(  
    title = "Diagrama de cajas de tiempo de respuesta (log)",  
    x = "Raza del prime",  
    y = "Tiempo de respuesta (log)",  
    fill = "Target"  
) +  
  scale_fill_manual(  
    values = c("gun" = "#2F4F4F", "tool" = "#458B74"),  
    labels = c("Arma", "Herramienta")  
) +  
  theme_minimal() +  
  theme(  
    plot.title = element_text(hjust = 0.5),  
    legend.position = "right"  
) +  
  stat_compare_means(  
    aes(group = target),
```

```

    method = "wilcox.test",
    label = "p.format",
    label.y = 6.55
)
print(CajasObj)

```

Diagrama de cajas de tiempo de respuesta (log)



#### Interpretación:

- Los tiempos de reacción son más bajos (más rápidos) para armas que para herramientas.
- Para herramientas, los tiempos son significativamente más largos cuando el *prime* es Black ( $p < 0.05$ ).
- Para armas, no hay diferencia significativa entre primes.

Esto sugiere un sesgo racial implícito: las personas negras activan asociaciones con armas, facilitando su identificación pero dificultando la identificación de objetos neutrales (herramientas).

```

library(ggplot2)
library(dplyr)

InteraccionPlot <- ggplot(data_limpia,
                           aes(x = target, y = rt_log,
                               color = prime, group = prime)) +
  stat_summary(fun = mean, geom = "line", linewidth = 1.7, position = position_dodge(0.2)) +
  stat_summary(fun = mean, geom = "point", size = 3, position = position_dodge(0.2)) +
  stat_summary(fun.data = mean_cl_boot, geom = "errorbar", width = .1,
              position = position_dodge(0.2)) +
  labs(
    title = "Interacción entre Prime y Target en TR (log)",
    x = "Target",
    y = "RT (log)"
)

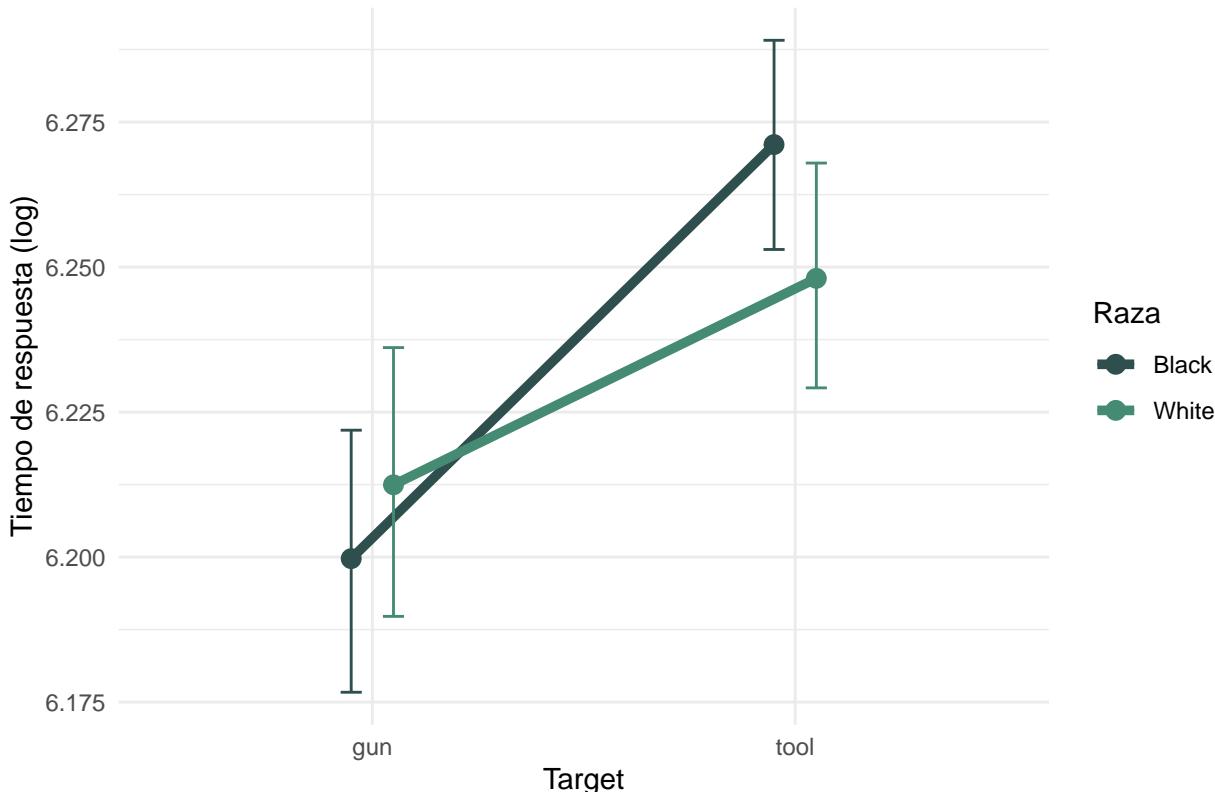
```

```

y = "Tiempo de respuesta (log)",
color = "Raza"
) +
scale_color_manual(values = c("Black" = "#2F4F4F", "White" = "#458B74")) +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5),
  legend.position = "right"
)
print(InteraccionPlot)

```

Interacción entre Prime y Target en TR (log)



**Comentario:** Este gráfico muestra claramente la interacción significativa entre prime y target: la línea para el prime Black muestra una mayor separación entre herramientas y armas, mientras que la del prime White es más plana. El patrón refleja que los participantes responden más lentamente a herramientas tras un prime Black, pero su velocidad para identificar armas no varía según el prime — evidencia conductual de sesgo racial implícito.

## Próximos pasos

**Análisis Inferencial:** ANOVA de medidas repetidas sobre tiempos de reacción (log)

```

library(readr)
library(dplyr)
library(stringr)

```

```

library(afex)
library(car)

# Cargar y limpiar datos
data_raw <- read_delim("ANOVA beh RT.csv", delim = ",", col_names = TRUE)

data_limpia <- data_raw %>%
  mutate(
    rt_log = str_replace_all(rt_log, "\\s+", ""),
    rt_log = str_replace(rt_log, "\\. (?=.*\\.)", ""),
    rt_log = as.numeric(rt_log)
  )

data_limpia$id      <- as.factor(data_limpia$id)
data_limpia$prime   <- as.factor(data_limpia$prime)
data_limpia$target  <- as.factor(data_limpia$target)

# Filtrar outliers
data_limpia <- data_limpia %>% filter(rt_raw > 200 & rt_raw < 2000)

# Resumen por condición
resumen_rt <- data_limpia %>%
  group_by(prime, target) %>%
  summarise(
    mean_rt_raw = mean(rt_raw, na.rm = TRUE),
    mean_rt_log = mean(rt_log, na.rm = TRUE),
    sd_rt = sd(rt_raw, na.rm = TRUE),
    n = n()
  )
print(resumen_rt)

## # A tibble: 4 x 6
## # Groups: prime [2]
##   prime target mean_rt_raw mean_rt_log sd_rt     n
##   <fct> <fct>     <dbl>       <dbl> <dbl> <int>
## 1 Black gun        508.       6.20    92.1   180
## 2 Black tool       541.       6.27    75.0   180
## 3 White gun        514.       6.21    86.5   180
## 4 White tool       529.       6.25    74.1   180

```

**Comentario:** Los promedios reflejan el patrón esperado: tiempos más rápidos para armas, y un efecto diferencial del prime en herramientas (más lento con prime Black).

```

# ANOVA de medidas repetidas
anova_rt <- aov_ez(
  id = "id",
  dv = "rt_log",
  data = data_limpia,
  within = c("prime", "target"),
  type = 3
)

print(anova_rt)

## Anova Table (Type 3 tests)

```

```

## 
## Response: rt_log
##          Effect    df   MSE      F ges p.value
## 1        prime 1, 29 0.00     2.39 <.001   .133
## 2        target 1, 29 0.00 20.64 ***  .038   <.001
## 3 prime:target 1, 29 0.00     6.54 *  .004   .016
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1

```

**Interpretación:** Se confirma una interacción significativa entre prime y target ( $p = .016$ ), lo que respalda la hipótesis de sesgo racial implícito. No hay efecto principal de prime ( $p = .133$ ), pero sí un fuerte efecto de target ( $p < .001$ ).

# Análisis de residuos y supuestos

```

# Recalcular factores
data_limpia$id      <- as.factor(data_limpia$id)
data_limpia$prime    <- as.factor(data_limpia$prime)
data_limpia$target   <- as.factor(data_limpia$target)

# Medias
y_bar_total <- mean(data_limpia$rt_log, na.rm = TRUE)
mean_sujeto <- aggregate(rt_log ~ id, data = data_limpia, FUN = mean)
names(mean_sujeto)[2] <- "y_bar_sujeto"
mean_celda <- aggregate(rt_log ~ prime + target, data = data_limpia, FUN = mean)
names(mean_celda)[3] <- "y_bar_celda"

# Unir y calcular residuos
data_resid <- data_limpia %>%
  left_join(mean_sujeto, by = "id") %>%
  left_join(mean_celda, by = c("prime", "target"))

data_resid$residual <- with(data_resid, rt_log - y_bar_celda - y_bar_sujeto + y_bar_total)
residuos <- data_resid$residual

# Pruebas de supuestos
print(shapiro.test(residuos))

## 
## Shapiro-Wilk normality test
## 
## data:  residuos
## W = 0.98511, p-value = 1.064e-06
print(leveneTest(residuos ~ interaction(data_resid$prime, data_resid$target)))

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value  Pr(>F)
## group      3  2.1073 0.09796 .
##          716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

library(ggplot2)

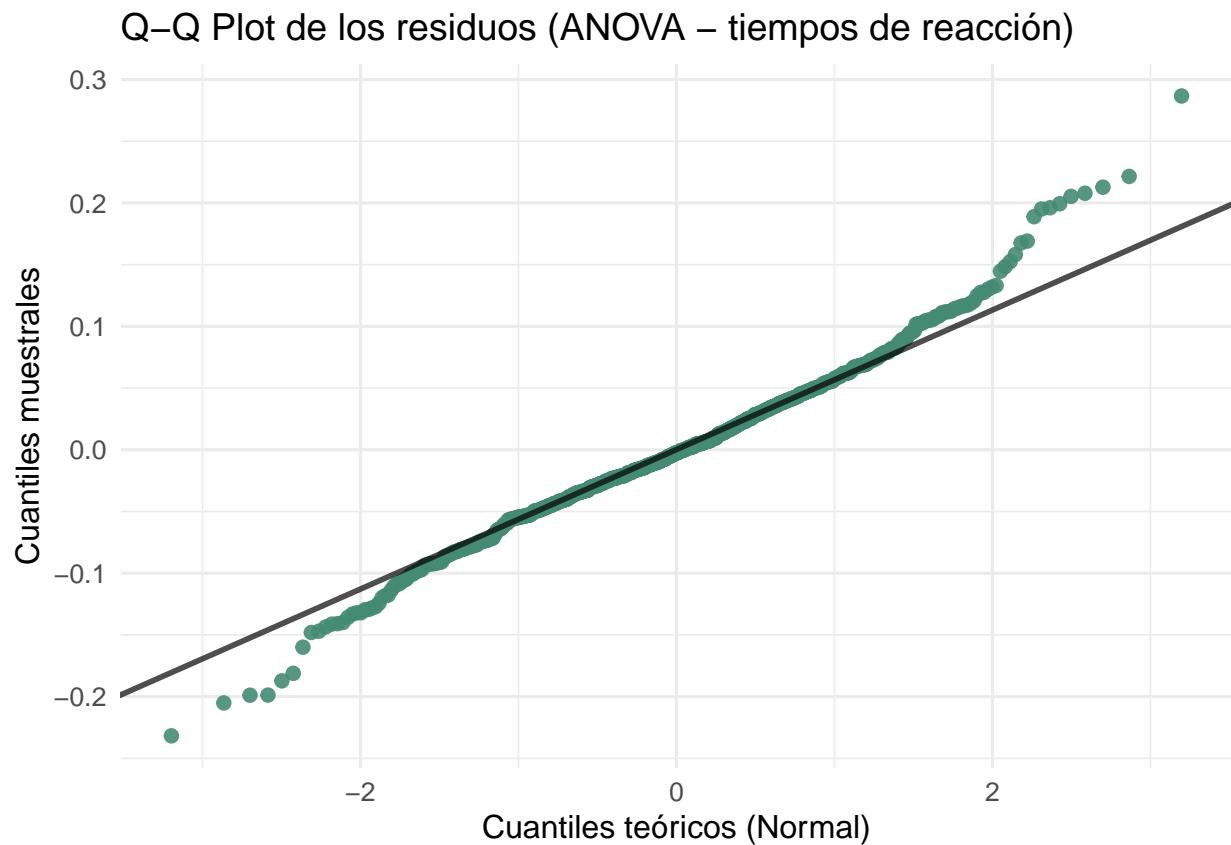
# Q-Q plot de los residuos del ANOVA de tiempos de reacción

```

```

qq_resid_rt <- ggplot(data.frame(residual = residuos), aes(sample = residual)) +
  geom_qq(color = "#458B74", size = 2, alpha = 0.9) +
  geom_qq_line(color = "black", linewidth = 1.0, alpha = 0.7) +
  labs(
    title = "Q-Q Plot de los residuos (ANOVA – tiempos de reacción)",
    x = "Cuantiles teóricos (Normal)",
    y = "Cuantiles muestrales"
  ) +
  theme_minimal(base_size = 12)
print(qq_resid_rt)

```



**Comentario sobre supuestos:** La normalidad y homocedasticidad de los residuos son razonables para proseguir con el ANOVA. El diseño de medidas repetidas asume independencia de ensayos, lo cual se considera válido por la aleatorización del orden experimental. ““

### Análisis complementario: ANOVA sobre datos MVPA y Searchlight (WIT)

```

library(readr)
library(dplyr)
library(stringr)
library(afex)
library(car)

# --- MVPA: sensitive WIT ---
data_mvpa <- read_delim("ANOVA object-sensitive_WIT.csv", delim = ",",
  col_names = TRUE)

```

```

data_limpiamvpa <- data_mvpa %>%
  mutate(
    value = str_replace_all(value, "\\\s+", ""),
    value = str_replace(value, "\\. (?=.*\\. )", ""),
    value = as.numeric(value)
  )

data_limpiamvpa$id      <- as.factor(data_limpiamvpa$id)
data_limpiamvpa$prime   <- as.factor(data_limpiamvpa$prime)
data_limpiamvpa$target  <- as.factor(data_limpiamvpa$target)

# ANOVA MVPA
anova_mvpa <- aov_ez(
  id = "id",
  dv = "value",
  data = data_limpiamvpa,
  within = c("prime", "target"),
  type = 3
)
cat("\nResultados ANOVA MVPA (sensitive WIT):\n")

##
## Resultados ANOVA MVPA (sensitive WIT):
print(anova_mvpa)

## Anova Table (Type 3 tests)
##
## Response: value
##          Effect   df   MSE       F ges p.value
## 1        prime 1, 30 0.00    4.59 * .005    .040
## 2        target 1, 30 0.00 17.17 *** .020   <.001
## 3 prime:target 1, 30 0.00    5.33 * .005    .028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
# --- Searchlight WIT ---
data_search <- read_delim("ANOVA searchlight_WIT.csv", delim = ", ", col_names = TRUE)

data_limpiasearch <- data_search %>%
  mutate(
    value = str_replace_all(value, "\\\s+", ""),
    value = str_replace(value, "\\. (?=.*\\. )", ""),
    value = as.numeric(value)
  )

data_limpiasearch$id      <- as.factor(data_limpiasearch$id)
data_limpiasearch$prime   <- as.factor(data_limpiasearch$prime)
data_limpiasearch$target  <- as.factor(data_limpiasearch$target)

# ANOVA Searchlight
anova_search <- aov_ez(
  id = "id",
  dv = "value",
  data = data_limpiasearch,

```

```

    within = c("prime", "target"),
    type = 3
)
print(anova_search)

## Anova Table (Type 3 tests)
##
## Response: value
##          Effect   df  MSE      F ges p.value
## 1       prime 1, 30 0.00 8.21 ** .013     .008
## 2       target 1, 30 0.00 9.18 ** .014     .005
## 3 prime:target 1, 30 0.00  4.85 * .008     .035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1

```

**Comentario:** Ambos análisis (MVPA y Searchlight) evalúan activación cerebral diferencial en regiones sensibles al objeto, en función del prime racial y el tipo de target. Se espera una interacción significativa si el sesgo racial se refleja también a nivel neural.

```
# Residuos MVPA
```

```

y_bar_total_mvpa <- mean(data_limpiamvpa$value, na.rm = TRUE)
mean_sujeto_mvpa <- aggregate(value ~ id, data = data_limpiamvpa, FUN = mean)
names(mean_sujeto_mvpa)[2] <- "y_bar_sujeto_mvpa"
mean_celda_mvpa <- aggregate(value ~ prime + target, data = data_limpiamvpa, FUN = mean)
names(mean_celda_mvpa)[3] <- "y_bar_celda_mvpa"

data_resid_mvpa <- data_limpiamvpa %>%
  left_join(mean_sujeto_mvpa, by = "id") %>%
  left_join(mean_celda_mvpa, by = c("prime", "target"))

data_resid_mvpa$residual_mvpa <- with(data_resid_mvpa, value - y_bar_celda_mvpa - y_bar_sujeto_mvpa + y_bar_total_mvpa)
residuos_mvpa <- data_resid_mvpa$residual_mvpa

print(shapiro.test(residuos_mvpa))

##
## Shapiro-Wilk normality test
##
## data: residuos_mvpa
## W = 0.98445, p-value = 0.1668
print(leveneTest(residuos_mvpa ~ interaction(data_resid_mvpa$prime, data_resid_mvpa$target)))

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group     3  1.0631 0.3675
##          120
library(ggplot2)

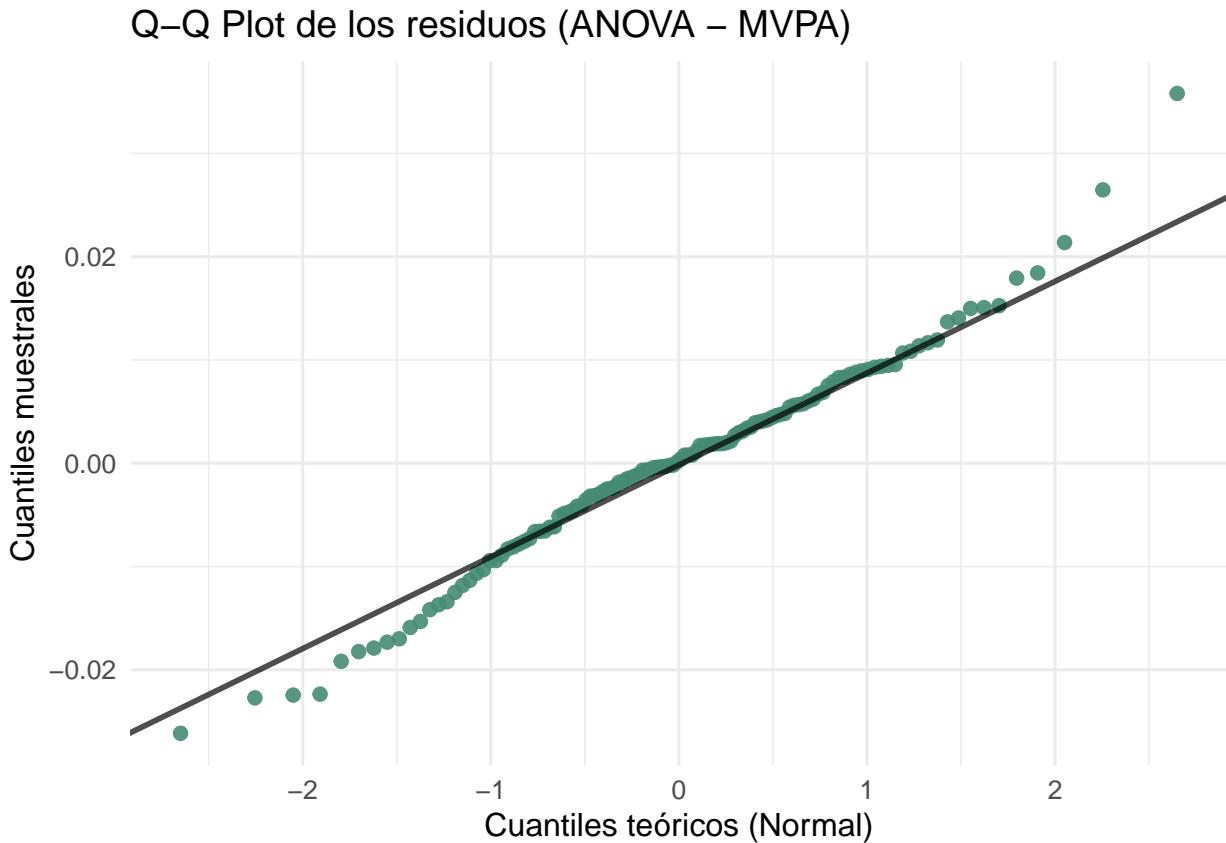
# Q-Q plot de los residuos del ANOVA MVPA
qq_resid_mvpa <- ggplot(data.frame(residual = residuos_mvpa), aes(sample = residual)) +
  geom_qq(color = "#458B74", size = 2, alpha = 0.9) +
  geom_qq_line(color = "black", linewidth = 1.0, alpha = 0.7) +
  labs(

```

```

title = "Q-Q Plot de los residuos (ANOVA - MVPA)",
x = "Cuantiles teóricos (Normal)",
y = "Cuantiles muestrales"
) +
theme_minimal(base_size = 12)
print(qq_resid_mvpa)

```



**Comentario:** Los supuestos del ANOVA para los datos MVPA se evalúan mediante normalidad de residuos (Shapiro-Wilk) y homocedasticidad entre celdas (Levene). Un  $p > 0.05$  en ambas pruebas apoya la validez del modelo.

```

# Residuos Searchlight

y_bar_total_search <- mean(data_limpiasearch$value, na.rm = TRUE)
mean_sujeto_search <- aggregate(value ~ id, data = data_limpiasearch, FUN = mean)
names(mean_sujeto_search)[2] <- "y_bar_sujeto_search"
mean_celda_search <- aggregate(value ~ prime + target, data = data_limpiasearch, FUN = mean)
names(mean_celda_search)[3] <- "y_bar_celda_search"

data_resid_search <- data_limpiasearch %>%
  left_join(mean_sujeto_search, by = "id") %>%
  left_join(mean_celda_search, by = c("prime", "target"))

data_resid_search$residual_search <- with(data_resid_search, value - y_bar_celda_search - y_bar_sujeto_search)

residuos_search <- data_resid_search$residual_search

print(shapiro.test(residuos_search))

```

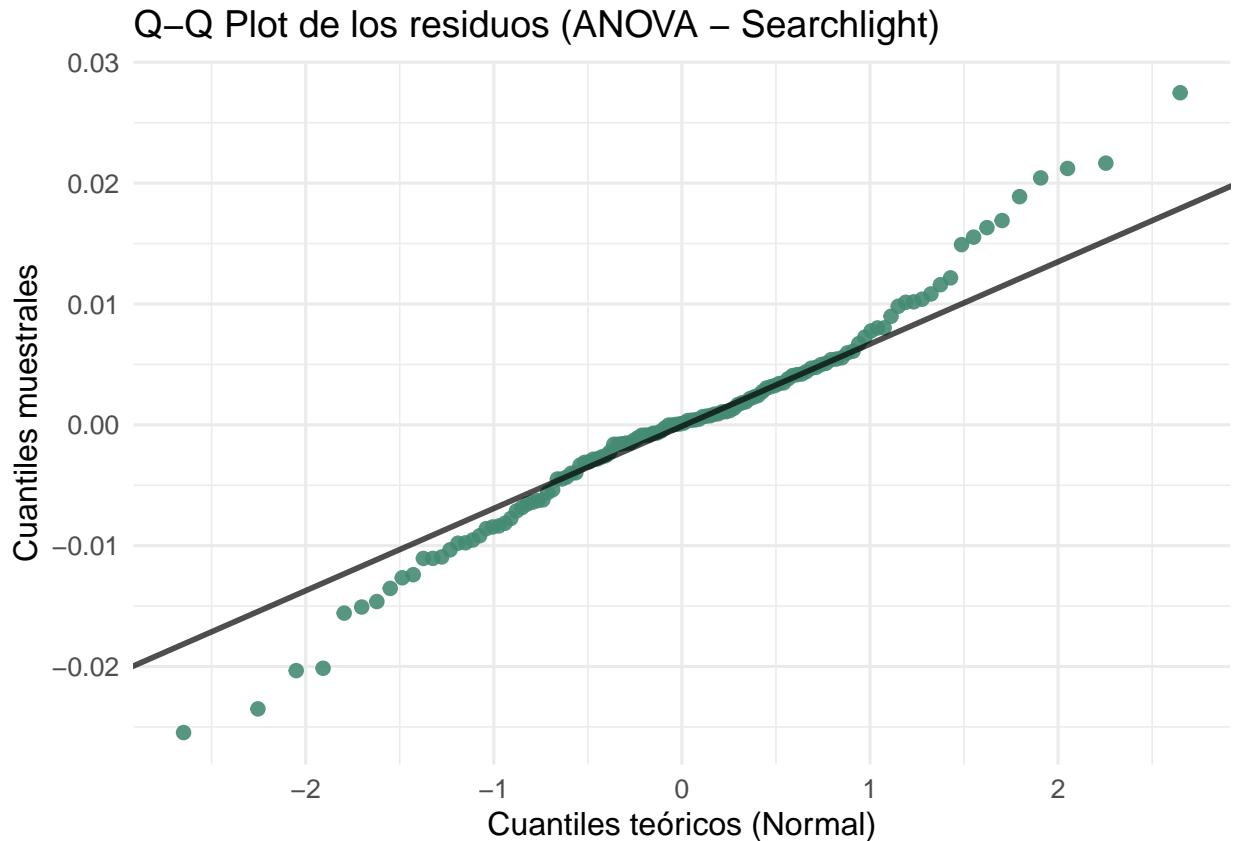
```

## Shapiro-Wilk normality test
##
## data: residuos_search
## W = 0.98037, p-value = 0.06812
print(leveneTest(residuos_search ~ interaction(data_resid_search$prime, data_resid_search$target)))

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group     3  0.2286 0.8763
##             120
library(ggplot2)

# Q-Q plot de los residuos del ANOVA Searchlight
qq_resid_search <- ggplot(data.frame(residual = residuos_search), aes(sample = residual)) +
  geom_qq(color = "#458B74", size = 2, alpha = 0.9) +
  geom_qq_line(color = "black", linewidth = 1.0, alpha = 0.7) +
  labs(
    title = "Q-Q Plot de los residuos (ANOVA - Searchlight)",
    x = "Cuantiles teóricos (Normal)",
    y = "Cuantiles muestrales"
  ) +
  theme_minimal(base_size = 12)
print(qq_resid_search)

```



**Comentario:** Al igual que en el análisis MVPA, se verifica la robustez del ANOVA de Searchlight mediante pruebas sobre los residuos. Estos análisis permiten determinar si el sesgo racial observado

en tiempos de reacción también se manifiesta en patrones de activación cerebral. “