```
birthwt_hypothesis_test
   • Low: indicador de peso al nacer inferior a 2,5 kg.
   • Age: edad de la madre en años.
   • Race: raza de la madre (1 = blanca, 2 = negra, 3 = otra).
   • Smoke: estado de tabaquismo durante el embarazo.
```

• Ftv: número de visitas al médico durante el primer trimestre.

• Ht: historia de hipertensión.

Se busca:

data("birthwt")

install.packages("ggplot2")

install.packages("dplyr")

if (!requireNamespace("dplyr", quietly = TRUE)) {

No Fumadora

non smokers <- subset(birthwt, smoke==0)</pre>

data: smokers\$bwt and non smokers\$bwt

t = -2.7299, df = 170.1, p-value = 0.007003

print(t test result difference bwt)

Welch Two Sample t-test

95 percent confidence interval:

x = "Peso al Nacer (gramos)",

fill = "Estado de Tabaquismo"

color = "Estado de Tabaquismo",

y = "Densidad",

2e-04

0e+00

5000

4000

(gramos)

Peso al Nacer (ç

Peso al Nacer (gramos)

Wilcoxon.

No Fumadora

Imprimir los resultados del test de Wilcoxon

Wilcoxon rank sum test with continuity correction

alternative hypothesis: true location shift is not equal to 0

Df Sum Sq Mean Sq F value Pr(>F) 1 3625946 3625946 7.038 0.00867 **

Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

cat("\nLos datos de fumadoras NO siguen una distribución normal.\n")

cat("\nLos datos de fumadoras siguen una distribución normal.\n")

cat("Los datos de no fumadoras NO siguen una distribución normal.\n")

Los datos de fumadoras siguen una distribución normal.

if (shapiro_non_smokers\$p.value < 0.05) {</pre>

significativamente diferentes.\n")

distribuciones sean diferentes.\n")

cat("\nInterpretación del ANOVA:\n")

entre fumadoras y no fumadoras.

} else {

8. Concluir sobre la relación que tienen los hábitos de la madre y el peso del recién nacido.

data: smokers normality and non smokers normality

print(wilcoxon_test)

W = 3260.5, p-value = 0.006768

7. Realizar un test de ANOVA.

print(summary(anova_result))

fumadoras sea menor de 2500 gramos.

Residuals 187 96343710 515207

smoke

} else {

documento.

##

1000

2000

3000

4000

-488.97860 -78.57486

sample estimates:

##

t test result difference bwt <- t.test(smokers\$bwt, non smokers\$bwt)

alternative hypothesis: true difference in means is not equal to 0

title = "Distribución del Peso al Nacer según el Estado de Tabaquismo",

• Ui: presencia de irritabilidad uterina.

```
Contrastes de hipótesis en R: birthwt
Los datos con los que trabajararemos pertenecen al conjunto de datos birthwt, y están disponibles en MASS. Este conjunto de datos contiene
datos sobre los factores de riesgo asociados con el bajo peso en el nacimiento. Su estructura es la siguiente:
```

```
• Lwt: peso de la madre en libras en el último período menstrual.
• Ptl: número de trabajos prematuros previos.
```

 Bwt: peso al nacer en gramos. Se parte de las hipótesis: Ho: La media de los pesos de los recién nacidos no es menor a la media (2500) si las madres son fumadoras durante el embarazo. H1: La media de los pesos de los recién nacidos es menor a la media (2500) si las madres son fumadoras durante el embarazo.

```
1. Tratar los datos de manera que se puedan realizar test con ellos.
Primero de todo procedemos a cargar los datos y observar la estructura de los mismos.
 # Cargar los paquetes necesarios
 library(MASS)
```

```
library(ggplot2)
library(dplyr)
```

```
## Attaching package: 'dplyr'
## The following object is masked from 'package:MASS':
```

```
##
##
       select
## The following objects are masked from 'package:stats':
##
```

```
##
       filter, lag
## The following objects are masked from 'package:base':
##
##
       intersect, setdiff, setequal, union
# Cargar el conjunto de datos birthwt
```

```
# Estructura del conjunto de datos
str(birthwt)
## 'data.frame':
                 189 obs. of 10 variables:
   $ low : int 0 0 0 0 0 0 0 0 0 ...
   $ age : int 19 33 20 21 18 21 22 17 29 26 ...
  $ lwt : int 182 155 105 108 107 124 118 103 123 113 ...
  $ race : int 2 3 1 1 1 3 1 3 1 1 ...
  $ smoke: int 0 0 1 1 1 0 0 0 1 1 ...
  $ ptl : int 0 0 0 0 0 0 0 0 0 ...
## $ ht : int 0 0 0 0 0 0 0 0 0 ...
  $ ui : int 1 0 0 1 1 0 0 0 0 0 ...
```

```
## $ ftv : int 0 3 1 2 0 0 1 1 1 0 ...
    $ bwt : int 2523 2551 2557 2594 2600 2622 2637 2637 2663 2665 ...
   2. Realizar t-test con el peso de los recién nacidos, suponiendo que la media son 2500.
Considerando que la media del peso de los recién nacidos son 2500 gramos, realizamos un t-test con aquellos casos en los que la madre ha
sido fumadora durante el embarazo. Dicho test lo realizamos para observar con que hipotesis de las mencionadas anteriormente nos
quedaríamos.
menor de 2500 gramos.
```

Observando el resultado obtenido, en lo primero que nos fijamos es en el p-valor el cual es mucho mayor que 0.05 y, por lo tanto, no tenemos la evidencia suficiente como para rechazar nuestra hipotesis nula (H0) y concluir que la media de los recién nacidos de las madres fumadoras sea Además del valor de p, dicha prueba nos muestra diferentes estadísticos que nos aportan mayor información para tomar la decisión anterior. El primero es el valor del t-estadístico el cual es positivo y ya nos sugiere que la media muestral estará por encima de la media supuesta. El segundo es el valor de la media muestral (2771,919) que ya podemos ver como está por encima del valor de mu. Y, por último, el intervalo de confianza del 95% nos marca que el valor de la media muestral está dentro de este rango. # Realizar un t-test con el peso de los recién nacidos, suponiendo que la media es 2500

Suponemos que las madres son fumadoras durante el embarazo smokers <- subset(birthwt, smoke == 1)</pre> t test result <- t.test(smokers\$bwt, mu = 2500, alternative = "less")</pre> # Mostrar el resultado del t-test print(t test result)

```
##
##
   One Sample t-test
##
## data: smokers$bwt
## t = 3.5461, df = 73, p-value = 0.9997
## alternative hypothesis: true mean is less than 2500
## 95 percent confidence interval:
        -Inf 2899.669
## sample estimates:
## mean of x
## 2771.919
 3. Crear un diagrama de caja que muestre cómo varía el peso de nacimiento según el estado de tabaquismo.
```

```
Si analizamos el diagrama obtenido a partir del filtrado de los datos en dos grupos, podemos observar ya una primera diferencia con el valor de
la mediana, que obtiene valores por encima de los 3000 gramos para las madres no fumadoras y, por debajo dee este mismo valor para las
madres fumadoras. Esto nos sugiere que los recién nacidos por parte de las madres no fumadoras tienden a nacer con un peso mayor. Esta
conclusión coge todavía una mayor fuerza si observamos la caja central, que nos representa el rango intercuartílico entre el 25% y 75% de los
datos. Como bien se observa a primera vista, presentan una distribución más amplia y un promedio más elevados respecto a la caja de las
madres fumadoras, que presentan una distribución más concentrada y una menor variabilidad.
Por lo tanto, este gráfico nos indica que el tabaquismo puede guardar relación con un menor peso en lo recién nacidos. Así que hay que
plantearse dicha hipotesis y tratar de demostrarla mediante pruebas estadísticas.
 # Instalar los paquetes necesarios si no están ya instalados
 if (!requireNamespace("ggplot2", quietly = TRUE)) {
```

boxplot(bwt ~ smoke, data = birthwt, main = "Peso al Nacer según el Estado de Tabaquismo", xlab = "Estado de Tabaquismo", ylab = "Peso al Nacer (gramos)", col = c("lightblue", "pink"), names = c("No Fumadora", "Fumadora"))

Crear el diagrama de caja que muestra cómo varía el peso de nacimiento según el estado de tabaquismo

```
Peso al Nacer según el Estado de Tabaquismo
      5000
      4000
Peso al Nacer (gramos)
      3000
      2000
      1000
                                                                                0
```

Fumadora

Cómo bien hemos comentado en el cierre del apartado anterior, con tal de demostrar esa hipotesis tenemos que evaluarlo con un t-test para

```
Esta prueba nos compara ambas medias de los dos grupos y evalua si dicha diferencia es significativa como para afirmar que existen diferencias
entre ambos grupos. A simple vista podemos ver como la media de los recién nacidos de madres no fumadoras es superior al de las madres
fumadoras. Partiendo de estas dos medias se formulan dos hipotesis: la primera, es la hipotesis nula (H0) que afirma que no existen diferencias
significativas en ambos grupos; la segunda, es la hipotesis alternativa (H1), que afirma que las medias entre ambos grupos es diferente. Con tal
de decidir que hipotesis descartar, tenemos diferentes valores estadísticos que nos dan información. El más evidente es el valor de p, inferior a
0,05 lo que nos indica que existe suficiente evidencia como para rechazar la hipótesis nula (H0), y por lo tanto afirmar que existen diferencias
significativas entre ambos grupos analizados. Si observamos los demás valores estadísticos, podemos ver como a diferencia del test realizado
anteriormente, aquí el valor de t es negativo y nos sugiere que el valor de la media del primer grupo es inferior a la del segundo. Y, el intervalo de
confianza obtenemos un rango que no contiene el 0, lo que también sugiere una diferencia estadísticamente significativa, ya que la difencia entre
ambas medias no es igual a 0, que es un supuesto de la hipotesis alternativa. Por lo tanto, podemos evidenciar como existen diferencias
significativas entre los dos grupos y, que, los recién nacidos de las madres no fumadoras presentan un peso mayor respecto al grupo de las
madres fumadoras.
Además de presentar dicho analisis del resultado del t-test, también se realiza un gráfico de densidades para comparar ambas distribuciones. Lo
que primero salta a la vista es el desplazamiento hacia la derecha de los datos del grupo de no fumadoras, y que nos indica que los pesos de
dicho grupo son mayores respecto del otro grupo. Un indicativo más que sigue dando fuerza a nuestra hipotesis es la ubicación de los picos de
ambas distribuciones, siendo el de fumadoras más alto pero desplazado hacia la izquierda respecto al otro grupo. También como hemos
observado en el boxplot anterior, podemos vere nuevamente como la distribución de las no fumadoras es más amplia, y nos indica una mayor
variabilidad entre los datos. En cambio, la distribución de los datos del grupo de fumadoras es más estrecha y concentrada alrededor de valores
más bajos del peso.
 # Evaluar si la diferencia en el peso de nacimiento es significativa utilizando un t-test
```

Estado de Tabaquismo

poder afirmarla con la suficiente evidencia. Para ello, utilizamos un t-test con dos muestras poblacionales.

4. Evaluar ahora si esta diferencia es significativa. Usamos para ello t-test y creamos los gráficos de densidad.

mean of x mean of y ## 2771.919 3055.696 # Crear un gráfico de densidad para comparar las distribuciones birthwt\$smoke <- factor(birthwt\$smoke, levels = c(0, 1), labels = c("No Fumadora", "Fumadora"))ggplot(birthwt, aes(x = bwt, color = smoke, fill = smoke)) +geom density(alpha = 0.5) + scale color manual(values = c("blue", "red"), labels = c("No Fumadora", "Fumadora")) + scale fill manual(values = c("blue", "red"), labels = c("No Fumadora", "Fumadora")) +

```
) +
  theme minimal()
       Distribución del Peso al Nacer según el Estado de Tabaquismo
                                                                     Estado de Tabaquismo
Densidad
                                                                         No Fumadora
                                                                         Fumadora
```

5000





Fumadora

6. Comprobar la normalidad de los datos, apoyándote en el test Shapiro. Comprobar la equidistribución de las poblaciones con el test de

siguen una distribución normal; la segunda, la hipotesis alternativa (H1) que afirma que los datos no siguen una distribución normal. Este test nos

Tanto para el grupo de la madres fumadoras como para el de las no fumadoras, observamos como el valor del estadístico W es muy cercano a 1

ya nos indica que los datos seguramente se van a distribuir de una forma normal, y así lo corrobora el p-valor el cual es mayor de 0,05. Por lo

tanto, no rechazamos la hipótesis nula y podemos afirmar que los datos de ambos grupos siguen una distribución normal.

Con tal de evaluar la normalidad de los datos de ambas poblaciones, nos hemos apoyado del test Shapiro-Wilk el cual mide que tan bien se

devuelve dos estadísticos, el valor de W que mide la correlación entre los valores de los datos a comparar respecto a los de una distribución

normal y, el p-valor que como en los otros tests es la probabilidad de obtener un resultado igual o más extremo del observado en los datos.

ajustan los datos a una distribución normal. Para ello, se plantean dos hipotesis: la primera, la hipotesis nula (H0) que afirma que los datos

Estado de Tabaquismo

Estado de Tabaquismo

Estado de Tabaquismo

No Fumadora

Fumadora

No Fumadora

Fumadora



```
la variabilidad dentro de los grupos, cuanto más grande es dicho valor mayor es la variabilidad entre grupos. El valor de p mide la probabilidad
del mismo modo que lo hace en los otros tests.
Para nuestro caso, el valor del estadístico F es positivo y por lo tanto como bien hemos mencionado anteriomente la variabilidad entre grupos es
más grande. El valor de p también es menor al valor de significación indicado en este caso por los dos asteriscos de 0,01. Lo que nos quiere
decir que existe una evidencia suficiente como para rechazar la hipótesis nula (H0) y afirmar que existen diferencias significativas entre grupos y
por lo tanto el estado de tabaquismo tiene un efecto significativo sobre el peso al nacer. También a través del valor estadístico F podemos
explicar buena parte de la variabilidad entre grupos, o lo que es lo mismo, que el hecho de que la madre haya fumado a lo largo del embarazo es
una de las variables más influyentes en el peso del recién nacido.
 # Realizar el test ANOVA
 anova_result <- aov(bwt ~ smoke, data = birthwt)</pre>
 # Imprimir los resultados del ANOVA
 cat("Resultados del ANOVA:\n")
 ## Resultados del ANOVA:
```

El último test que realizamos es un test de ANOVA, o lo que es lo mismo, un análisis de la varianza. Dicha prueba estadística se utiliza para

comparar las medias de dos o más grupos y determinar si existen diferencias significativas entre ellas. Dicha comparación se realiza analizando

si la variabilidad entre las medias de los grupos es mayor que la variabilidad dentro de los grupos formulando dos hipotesis: la primera, hipótesis

nula (H0) que afirma que las medias de los grupos son iguales; la segunda, hipótesis alternativa (H1) que afirma que almenos una de las medias es diferente. Los resultados que nos arroja este test son diferentes estadísticos, el valor de F mide la relación entre la variabilidad entre grupos y

```
madres fumadoras.
La tercera es que tanto en el grupo de las madres fumadoras como en el de las no fumadoras, los datos se ajustan a la normalidad.
Por último mediante línea de código se compara los valores de los tests realizados con los que supuestamente se deberían obtener para aprobar
o rechazar cada una de las hipótesis, y así mediante un output tomar la decisión adecuada.
 # Comprobación de normalidad
 if (shapiro smokers$p.value < 0.05) {</pre>
```

Con los datos obtenidos de dicho análisis se pueden extraer unas conclusiones que ya hemos ido comentando a lo largo de este

La segunda, es que se puede afirmar que existen diferencias significativas entre ambos grupos analizados, el de madres no fumadoras y el de

La primera de las conclusiones es que no tenemos la evidencia suficiente como afirmar que la media de los recién nacidos de las madres

```
} else {
cat("Los datos de no fumadoras siguen una distribución normal.\n")
## Los datos de no fumadoras siguen una distribución normal.
# Comprobación de equidistribución
if (wilcoxon_test$p.value < 0.05) {</pre>
cat("\nRechazamos la hipótesis nula: Las distribuciones de las dos poblaciones son
```

cat("\nNo se rechaza la hipótesis nula: No hay evidencia suficiente para concluir que las

```
## Rechazamos la hipótesis nula: Las distribuciones de las dos poblaciones son
## significativamente diferentes.
# ANOVA
anova_summary <- summary(anova_result)</pre>
p_value <- anova_summary[[1]][["Pr(>F)"]][1]
# Interpretar el resultado del ANOVA
```

```
## Interpretación del ANOVA:
if (p_value < 0.05) {
cat("El p-valor es", round(p_value, 4), ", que es menor a 0.05.\n")
cat("Rechazamos la hipótesis nula: Existe una diferencia significativa en el peso al nacer
entre fumadoras y no fumadoras.\n")
```

```
} else {
cat("El p-valor es", round(p_value, 4), ", que es mayor o igual a 0.05.\n")
cat("No se rechaza la hipótesis nula: No hay evidencia suficiente para afirmar que exista
una diferencia significativa en el peso al nacer entre fumadoras y no fumadoras.\n")
## El p-valor es 0.0087, que es menor a 0.05.
```

Rechazamos la hipótesis nula: Existe una diferencia significativa en el peso al nacer