**TEMA 3. ACTIVIDADES DE EVALUACIÓN CONTINUA**

**RELACIÓN ENTRE UNA VARIABLE CUANTITATIVA Y UNA VARIABLE CUALITATIVA**

**EJERCICIO 1**.

El dataset “mtcars” contiene los detalles técnicos de los conches disponibles en un concesionario de automóviles. Se puede acceder al mismo desde R con los siguientes comandos:

data(mtcars)

str(mtcars)

Aunque en España no son habituales los conches de transmisión automática, se considera que son más cómodos y tienen una menor propensión a sufrir daños mecánicos. En este apartado analizaremos además si tienen ventaja sobre los coches de transmisión manual en los apartados consumo (`mpg`) y aceleración (`qsec`)

1. En primer lugar, carga el fichero `mtcars` y crea con él un objeto denominado coches.
2. A continuación, transforma en factor la variable `am`.
3. Elabora un gráfico de caja y bigotes que relacione las variables `am`y `mpg`. Incluye una línea que corresponda al valor mediano de `mpg`
4. Construye un gráfico de violín que represente el consumo en función del tipo de transmisión. Para ello, deberás crear dos subconjuntos en función del tipo de transmisión. Posteriormente, elabora el gráfico de violín. Incluye una leyenda a tu gusto.
5. Aplica la prueba de la t de Student para determinar si hay diferencias significativas en el consumo de los vehículos en función del tipo de transmisión. ¿Es estadísticamente significativo el consumo medio en función del tipo de transmisión? En consecuencia, ¿qué vehículo recomendarías?
6. Para estar seguro de los resultados, compruébalo mediante la aplicación de la prueba de la suma de rangos de Wilcoxon o prueba de Mann-Whitney.
7. Comprueba que la variable `mpg` sigue una distribución normal aplicando la prueba de Shapiro. Dibuja igualmente un gráfico Q-Q. De acuerdo con lo antes expuesto, ¿sigue la variable `mpg` una distribución normal?
8. Analiza igualmente si la varianza de `mpg` en los dos grupos creados por el factor `am` es similar.
9. Repite la aplicación de la prueba t de Student para verificar si el tipo de transmisión (variable `am`) influye también en la aceleración del vehículo (variable `qsec`). A partir de los resultados de la prueba, ¿influye o no?

**EJERCICIO 2**.

El conjunto de datos `airquality` muestra los valores diarios de ozono y de diferentes variables meteorológicas en Nueva York, entre mayo y septiembre de 1973. Se puede acceder al mismo desde R con los siguientes comandos:

data(airquality)

str(airquality)

El objetivo de esta práctica es comprobar si existen diferencias significativas entre esas variables en función del mes en el que se obtuvieron.

1. Para ello, comenzamos cargando el dataset y creando un objeto denominado `aq`.
2. Transforma en factor la variables `Month`.
3. Calcula el valor de cada una de las variables cuantitativas en función del mes y crea con todos los resultados un dataframe denomando `tabla.resumen`. En dicha tabla, transforma los registros de temperatura de ºF a ºC y procura que este dataframe no repita varias columnas idénticas con los meses.
4. Dibuja un gráfico de caja y bigotes que muestre los valores mensuales de la variable `Ozone` y su valor medio durante el periodo de análisis.
5. Para comprobar si existe alguna diferencia significativa entre el valor promedio de ozono en cada mes se debe aplicar un análisis ANOVA. Realiza este análisis e interpreta su resultado. ¿Qué meses muestran una diferencia significativa con los restantes?
6. Comprueba si los residuos del modelo ANOVA se distribuyen según una curva normal.
7. Las conclusiones pueden reforzarse aplicando la prueba de Shapiro-Wilk test a esos mismos residuos. ¿Qué criterio sirve para establecer la significación estadística de esta prueba?

**EJERCICIO 3.**

El dataset “Ozone”, incluido en el paquete `mlbench`, consiste en un dataframe con 366 observaciones de 13 variables realizadas en la ciudad de Los Ángeles, en el año 1976, correspondiendo cada observación a un día. Se puede acceder al mismo desde R con los siguientes comandos:

library(mlbench)

data(Ozone)

str(Ozone)

De ese dataframe extrae un subconjunto con las variables

* [4] Daily maximum one-hour-average ozone reading.
* [5] 500 millibar pressure height (m) measured at Vandenberg AFB.
* [6] Wind speed (mph) at Los Angeles International Airport (LAX).
* [8] Temperature (degrees F) measured at Sandburg, CA.
* [10] Inversion base height (feet) at LAX.
* [11] Pressure gradient (mm Hg) from LAX to Daggett, CA.
* [12] Inversion base temperature (degrees F) at LAX.
* [13] Visibility (miles) measured at LAX.

Calcula el valor de la mediana de la variable ` Daily maximum one-hour-average ozone reading` y crea una nueva variable en la que el valor 1 corresponda a registros inferiores a la mediana, y el valor 2 consista en los registros iguales o superiores a la mediana.

Calcula si existen diferencias significativas, en función de las categorías de esa nueva variable, en los valores medios de las siguientes variables:

* [5] 500 millibar pressure height (m) measured at Vandenberg AFB.
* [10] Inversion base height (feet) at LAX.
* [13] Visibility (miles) measured at LAX.

Representa gráficamente esas diferencias mediante gráficos de caja y bigotes.

Repite el procedimiento anterior con la variable [10] `Inversion base height (feet) at LAX`, pero en vez de usar la mediana, utiliza los cuartiles, de tal manera que la nueva variable tendrá 4 niveles (1, 2, 3, 4). Calcula existen diferencias significativas, en función de las categorías de esa nueva variable, en los valores medios y en la varianza de:

* [6] Wind speed (mph) at Los Angeles International Airport (LAX).
* [8] Temperature (degrees F) measured at Sandburg, CA.
* [12] Inversion base temperature (degrees F) at LAX.