**TEMA 3. ACTIVIDADES DE EVALUACIÓN CONTINUA**

**RELACIÓN ENTRE VARIABLES CUANTITATIVAS**

**EJERCICIO 1**. Importa los datos de la hoja *Ejemplo1* de la hoja de cálculo <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/ejercicios_regresion_2023.xlsx>.

1. Determina si existe una relación lineal entre las variables y si las variables cumplen con el supuesto de normalidad.
2. Calcula los coeficientes de correlación de Pearson y Spearman.
3. Determina si existe una relación estadísticamente significativa entre las variables.

**EJERCICIO 2**. Importa los datos de la hoja *Ejemplo2* de la hoja de cálculo <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/ejercicios_regresion_2023.xlsx>. A partir de ella:

1. Elabora una matriz de correlaciones entre las diferentes variables.
2. Exporta tanto la matriz de correlaciones como la matriz de p-values a Excel.
3. Representa esa matriz gráficamente con un correlograma, utilizando para ello los números correspondientes a los coeficientes de correlación de Pearson.

**EJERCICIO 3**. Importa los datos de la hoja *Ejemplo3* de la hoja de cálculo <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/ejercicios_regresion_2023.xlsx>. A partir de ella:

1. Elabora una matriz de correlaciones entre las diferentes variables.
2. Exporta tanto la matriz de correlaciones como la matriz de p-values a Excel.
3. Representa esa matriz gráficamente con un correlograma, utilizando para ello los números correspondientes a los coeficientes de correlación de Pearson.

**EJERCICIO 4**. A partir de la hoja *Ejemplo2* presente en la hoja de cálculo <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/ejercicios_regresion_2023.xlsx>, realiza los siguientes cálculos:

1. Representación gráfica de la relación entre cada una de las variables (peso, edad y grasas) con las restantes, para determinar si existe una relación lineal entre ellas.
2. Comprueba si esas variables cumplen en supuesto de normalidad.
3. Teniendo en cuenta que el contenido en grasas puede considerarse dependiente de la edad y el peso, identifica cuál está mejor correlacionada con las grasas y desarrolla un modelo de regresión lineal.

**EJERCICIO 5**. A partir de la hoja *difusión* presente en la hoja de cálculo <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/ejercicios_regresion_2023.xlsx>, realiza los siguientes cálculos:

1. Crear un data frame que almacene los datos del ejemplo propuesto.
2. Representa gráficamente la relación entre la variable ventas y las restantes variables, para determinar si existe una relación lineal entre ellas.
3. Comprueba si todas las variables cumplen en supuesto de normalidad.
4. Calcula una matriz de correlación entre esas variables y represéntalas gráficamente mediante un correlograma.
5. Si la variable ventas puede considerarse la variable dependiente, identifica con qué variable está mejor correlacionada y desarrolla un modelo de regresión lineal.
6. Interpreta los parámetros de esa recta ¿son significativos estos parámetros? ¿Qué puede decirse del ajuste del modelo a los datos?
7. Si para el mes que viene quisiéramos elevar nuestras ventas a 25 millones de euros, ¿cuál sería el valor de la variable independiente que nos permitiría alcanzar este valor?

**EJERCICIO 6**. A partir de la hoja *clima\_Madrid* presente en la hoja de cálculo <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/ejercicios_regresion_2023.xlsx>, realiza los siguientes cálculos:

1. Crear un data frame que almacene los datos del ejemplo propuesto.
2. Representa gráficamente la relación entre las variables alt, pp, tene y tjul, para determinar si existe una relación lineal entre ellas.
3. Comprueba si esas variables cumplen en supuesto de normalidad.
4. Calcula una matriz de correlación entre todas las variables y elabora al menos dos gráficos que te permitan observar de manera global el grado de relación entre todas esas variables.
5. Teniendo en cuenta que la altitud es una variable geográfica que genera alteraciones importantes en la distribución espacial de las variables climáticas, identifica cuál está mejor correlacionada con la altitud y desarrolla un modelo de regresión lineal.
6. Interpreta los parámetros de esa recta ¿son significativos estos parámetros? ¿Qué puede decirse del ajuste del modelo a los datos?
7. Queremos trasladarnos a una vivienda que estará situada en el Puerto de Somosierra a 1480 metros de altitud. ¿Qué temperatura registraremos en enero?

**EJERCICIO 6**

El dataset “Air quality” contiene los valores diarios de las siguientes variables ambientales entre el 1 de mayo y el 30 de septiembre de 1973. Se puede acceder al mismo desde R con los siguientes comandos:

data(airquality)

str(airquality)

Las variables de ese conjunto de datos son las siguientes:

* Month: mes.
* Day: día.
* Ozone: media diaria del ozono (en partes por millón) entre las 13 y las 15 horas
* Solar.R: radiación solar (en Langleys) entre las 08 y las 12 horas.
* Wind: velocidad media del viento (en millas por hora) entre las 07 y las 10.
* Temp: temperatura máxima diaria en ºF.

**TAREA 1**. Preparación de los datos.

1. Cambia el nombre de las variables originales por los siguientes nombres: mes, día, ozono, radiación.solar, viento, temperatura.máxima.
2. Transforma los valores de la temperatura máxima y viento en unidades del sistema métrico decimal. Para ello:
   1. Aplica a la temperatura del aire el siguiente factor de conversión: temperatura en ºF multiplicada por 0,5556 – 32.
   2. Aplica al viento el siguiente factor de conversión: viento en millas \* 0,44704.
3. Convierte en factores las variables día y mes.

**TAREA 2**. Elabora un análisis exploratorio de las variables ozono, radiación.solar, viento y temperatura.máxima en el que analiza cómo varían esos valores a lo largo de los meses analizados. Utiliza las herramientas que consideres más oportunas. Comenta brevemente los resultados.

**TAREA 3**. Analiza si las variables ozono, radiación.solar, viento y temperatura máxima tienen una distribución normal por los métodos que consideres oportunos o si presentan algún tipo de sesgo. Si existiera alguna variable que no sigue una distribución normal, transfórmala elevando sus valores originales al cuadrado y repite el análisis para comprobar si la distribución ha cambiado.

**TAREA 4**. Identifica que parejas de variables poseen un valor del coeficiente de correlación.

**TAREA 5**. Teniendo en cuenta que consideramos la variable ozono como variable dependiente, elabora un diagrama de puntos que muestre la relación entre el ozono y la variable independiente mejor correlacionada.

**TAREA 6**. Elabora un modelo de regresión entre el ozono como variable dependiente y la variable independiente mejor correlacionada.

**EJERCICIO 7**

El fichero <https://personales.unican.es/rasillad/docencia/G14/TEMA_3/datos_inmobiliaria_2023.csv> contiene información sobre las viviendas disponibles para el alquiler o la venta en una inmobiliaria.

**TAREA 1**. Elabora una matriz de correlación, utilizando la correlación de Spearman, entre las variables *precio*, *superficie*, *comunidad* y *valor*. Analiza la significación estadística de las correlaciones calculadas y determina cuál es la pareja de variables con una relación más estrecha, y cuál es la pareja con una relación más débil.

**TAREA 2**. Elabora sendos modelos de regresión que relacionen las siguientes variables

1. *Superficie* (variable independiente) con el *precio* (variable dependiente), y *precio* (variable independiente) con *valor* (variables dependientes).
2. Analiza inicialmente si dichas variables cumplen con el requisito de normalidad.
3. Representa gráficamente la relación mediante un gráfico de puntos que incluya la recta de regresión en color verde.
4. Exporta los valores originales, los valores reconstruidos y los residuos en una hoja de cálculo Excel.
5. De acuerdo con el modelo anterior, calcula qué precio y qué valor catastral tendrían los siguientes pisos de nueva construcción.

|  |
| --- |
| Superficie (m2) |
| 37,5 |
| 42,8 |
| 55,7 |
| 62,1 |
| 83,9 |
| 99,8 |
| 107,9 |
| 124,9 |