

T8_P3

Guillermo Segura Gómez

2023-10-19

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Cargando las bibliotecas necesarias

Para este análisis, necesitaremos las bibliotecas **MASS** para la simulación de datos multivariados y **ggplot2** para la visualización.

```
{r libraries} library(MASS) library(ggplot2)
```

Establecer parámetros

Establecemos los parámetros de las distribuciones que vamos a simular. Aquí, μ representa la media y Σ la matriz de covarianza.

```
{r parameters} mu <- c(0, 0, 0, 0) Sigma <- matrix(c(1, 0.8, 0.6,
0.5,          0.8, 1, 0.7, 0.6,          0.6,
0.7, 1, 0.7,          0.5, 0.6, 0.7, 1),          ncol=4)
```

Generar datos

Generamos 1000 muestras a partir de una distribución normal multivariada con los parámetros establecidos anteriormente.

```
{r generate_data} set.seed(123) # Para reproducibilidad sample_data
<- mvrnorm(n=1000, mu=mu, Sigma=Sigma)
```

Convertir a data frame

Transformamos los datos a un formato de data frame para su fácil manipulación.

```
{r to_dataframe} df <- as.data.frame(sample_data)
```

Crear gráfica

Visualizamos la relación entre las variables V1 y V2 usando un gráfico de dispersión.

```
{r scatter_plot} ggplot(df, aes(x=V1, y=V2)) + geom_point(aes(color=V1),  
alpha=0.6) + labs(title="Scatter plot de V1 vs V2", x="V1",  
y="V2") + theme_minimal()
```

Conclusiones

A partir del gráfico, podemos observar que:

- Existe una relación positiva entre V1 y V2, lo que indica que cuando uno de ellos aumenta, el otro tiende a hacerlo también.
- La coloración en función de V1 permite visualizar la distribución y densidad de los puntos en el eje x.
- El uso de la transparencia (**alpha**) ayuda a identificar las áreas de mayor densidad de puntos.
- El diseño minimalista del gráfico facilita su interpretación y evita distracciones.

Es esencial realizar más análisis y visualizaciones para entender completamente las relaciones entre todas las variables en el conjunto de datos.