Estadística. Práctica 6

En esta práctica comprobaremos gráficamente algunos resultados teóricos del tema 4.

Ejemplo 1: Distribución T-Student. Aproximación por la N(0,1).

Mediante los menús de R-Commander, dibujamos la gráfica de la función de densidad de una distribución $N(0,1)$ y mantenemos abierta la ventana R-Graphics.

A continuación, con el comando `curve`, vamos a dibujar sobre los mismos ejes y con distintos colores las gráficas de las distribuciones t_5, t_{10} y t_{40}. Para ello escribimos y ejecutamos en la ventana de instrucciones (R Script) lo siguiente:

```
curve(dt(x,df=5),add=TRUE,col="green")
curve(dt(x,df=10),add=TRUE,col="red")
curve(dt(x,df=40),add=TRUE,col="blue")
```

Observamos que, a medida que aumentan los grados de libertad, la distribución T-Student se aproxima más a la $N(0,1)$.

Ejemplo 2: Caso particular del Teorema Central del Límite.[1]

Sea $X_1,\dots,X_{40}$ una muestra aleatoria simple de una variable $X \sim U(0,2)$. Por el Teorema Central del Límite, resulta que

$$\overline{X} = \frac{X_1+\cdots+X_{40}}{40} \simeq N\left(1,\frac{1}{120}\right)$$

Vamos a comprobar experimentalmente este resultado. Como no vamos a poder crear la variable $\overline{X}$ lo que haremos será generar una muestra representativa de ella. Para ello, a través de `Distribuciones` $\rightarrow$ `Distribuciones continuas` $\rightarrow$ `Distribución uniforme` $\rightarrow$ `Muestra de una distribución uniforme`, generamos, en un archivo de nombre **ejemplo2**, 1000 filas con 40 valores aleatorios de una distribución $U(0,2)$, calculando a la vez la media de cada fila. En la columna *mean* de dicho archivo tenemos una muestra de tamaño 1000 de la variable $\overline{X}$.

A continuación, a través de `Gráficas` $\rightarrow$ `Estimar densidad` dibujamos la función de densidad estimada para los datos de *mean*.

Finalmente dibujamos sobre los mismos ejes y con distinto color la función de densidad de la normal de media 1 y desviación típica $1/\sqrt{120}$ ejecutando en la ventana de instrucciones:

```
curve(dnorm(x,mean=1,sd=1/sqrt(120)),add=TRUE,col="red")
```

Podemos observar que las gráficas son similares.

[1] **Teorema Central del Límite:** Si $X_1,\dots,X_n$ es una muestra aleatoria simple de una variable aleatoria X, con media μ y varianza σ^2 finita, entonces $\overline{X} \simeq N\left(\mu,\frac{\sigma^2}{n}\right)$. En general, la aproximación es buena para $n \geq 30$.

Ejemplo 3.

Sabemos que si $X_1, \dots, X_n$ una muestra aleatoria simple procedente de una población $N(\mu, \sigma^2)$, entonces:

- $Y = \frac{(n-1)S_c^2}{\sigma^2} \sim \chi^2_{n-1}$
- $T = \frac{\overline{X}-\mu}{S_c}\sqrt{n} \sim t_{n-1}$

Caso particular:

Si $X_1, \dots, X_{10}$ una muestra aleatoria simple procedente de una población $N(1, 4)$, entonces:

- $Y = \frac{9S_c^2}{4} \sim \chi^2_9$
- $T = \frac{\overline{X}-1}{S_c}\sqrt{10} \sim t_9$

Para comprobar experimentalmente estos resultados, comenzamos, como en el ejemplo anterior, generando, en un archivo de nombre **ejemplo3**, 1000 filas con 10 valores de la distribución normal con $\mu = 1$ y $\sigma = 2$, calculando a la vez la media y la desviación típica de cada fila (se recuerda que la deviación típica que calcula R es S_c) . En las columnas *mean* y *sd* del archivo tenemos muestras de tamaño 1000 de $\overline{X}$ y de S_c, respectivamente.

A continuación calculamos las variables Y y T a través de `Datos` $\rightarrow$ `Modificar variables del conjunto de datos activo` $\rightarrow$ `Calcular una nueva variable`.

Para comprobar el primer resultado, dibujamos la función de densidad estimada de Y. A continuación, sobre los mismos ejes y con distinto color, dibujamos el gráfico de la función de densidad de la χ^2_9 ejecutando en la ventana de instrucciones:

```
curve(dchisq(x,df=9),add=TRUE,col="red")
```

Obsérvese que las gráficas son similares[2].

Para comprobar el segundo resultado, dibujamos la función de densidad estimada de T. A continuación, sobre los mismos ejes y con distinto color, dibujamos el gráfico de la función de densidad de la t_9:

```
curve(dt(x,df=9),add=TRUE,col="red")
```

Se observa que, en efecto, las gráficas son similares.

[2]Las gráficas no son iguales porque no hemos trabajado realmente con la variable Y sino con una muestra de ella. Lo mismo ocurre con la variable T.