

Lab $n^{\circ}2$

Enunciado:

Carguen el dataset `iris` usando la función `data()`. Este dataset contiene 150 muestras, de tres especies de flores del iris, con los valores de cuatro variables (largo y ancho del sépalo y pétalo). Pueden seleccionar un subconjunto de los datos usando `[]`. Por ejemplo, el comando

```
> iris[iris$Petal.Length >= 6 & iris$Species == "virginica",]
```

retorna las filas de datos donde el pétalo tiene largo mayor o igual a 6, y es de la especie “virginica” (11 filas).

1. Usando todos los datos de `iris`, calculen la distribución de frecuencia y frecuencia relativa según **especie** de flor.
2. Ahora usen `tapply()` para encontrar el promedio del **largo de pétalo** por **especie** de flor.
3. Calculen la distribución de frecuencia del **largo del sépalo**, usando `cut()` y `table()`. Recomendación: si se tienen n datos, traten de dividir los datos en $\log_2(n) + 1$ intervalos. Generen un histograma de la distribución del **largo del sépalo**: primero usando el resultado de `table()`, y después usando `hist()`.
4. Generen un gráfico de la frecuencia acumulada del **ancho del pétalo**.
5. Generen los siguientes gráficos de dispersión:
 - (a) largo del sépalo vs. ancho del sépalo
 - (b) largo del pétalo vs. ancho del pétalo

Rotulen los ejes de sus gráficos. Calculen también la covarianza y coeficiente de correlación para los dos gráficos de dispersión. ¿Existe una relación lineal entre alguna de estas variables?

6. Generen un gráfico de caja mostrando la distribución del **largo del sépalo** por especie.
7. Seleccionen las filas correspondientes a “virginica”, asígnenlas a una variable llamada `virginica`. Calculen el promedio, mediana, varianza y desviación estándar para todas las columnas numéricas de la especie “virginica”. Pueden calcular los valores uno a uno, pero es un buen momento para ejercitar lo que aprendieron ayer.

Ejemplo: aquí genere un `data.frame` con los datos solicitados.

	promedios	medianas	varianzas	desviaciones
Sepal.Length	7	6	0.40	0.6
Sepal.Width	3	3	0.10	0.3
Petal.Length	6	6	0.30	0.6
Petal.Width	2	2	0.08	0.3

Pueden usar la función `colnames()` y `row.names()` para obtener y cambiar los nombres de las columnas y filas de un `data.frame`, respectivamente.

1	> f <- data.frame(c(34, 22, 56), c(11, 99, 56))	9	> colnames(f) <- c("columna 1", "columna 2")
2	> f	10	> row.names(f) <- c("fila 1", "fila 2", "fila 3")
3	c.34..22..56. c.11..99..56.	11	> f
4	1 34 11	12	columna 1 columna 2
5	2 22 99	13	fila 1 34 11
6	3 56 56	14	fila 2 22 99
7	> colnames(f)	15	fila 3 56 56
8	[1] "c.34..22..56." "c.11..99..56."		

Entrega: suban un archivo `.R` a UCursos con el código que escribieron para hacer los cálculos y generar los gráficos, usando la opción “Tareas” → “Laboratorio 02”. Incluyan su respuesta a la pregunta 5 como comentario en su código.