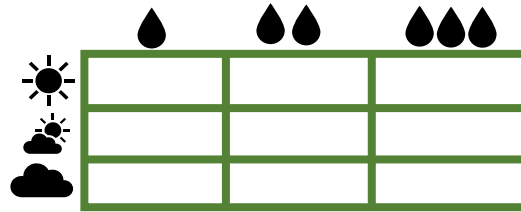


Ejercicio

Parte I.

Para hacer este ejercicio usaremos una base de datos de un floricultor, que está interesado en conocer qué combinaciones de agua y sombra dan una mayor masa de tulipanes (medida en gramos). Para ello, dividió una jardinera en 9 secciones, en las cuales uso tres combinaciones de riego y tres combinaciones de sombra. Este diseño lo replicó en dos jardineras adicionales, para un total de tres réplicas.

El diseño de su experimento para cada una de las réplicas se ve así:



De manera que en cada jardinera existen todas las combinaciones posibles de agua y sombra que quiere probar.

Objetivos:

0. ¿Cuáles son las tres posibles **variables predictivas** en el experimento del floricultor? ¿Cuál sería la variable respuesta en todos los casos?

Usando **sólo una** de las tres posibles variables predictivas:

1. Identificar una pregunta de investigación
2. Identificar la variable respuesta y la variable predictiva
3. Plantear la hipótesis nula y la hipótesis alternativa para la pregunta
4. Calcular la suma de cuadrados total a mano para todo el conjunto de datos
5. Explorar de manera gráfica los datos (histograma y boxplot) para contestar las siguientes preguntas:
 - a. ¿Qué puedes decir de la distribución de tu variable respuesta?
 - b. ¿Se nota algún patrón entre la variable respuesta y la variable predictiva?
6. Ajustar un GLM en R, según tu pregunta de investigación
7. Usar el comando `anova` y el comando `summary` para explorar los resultados del modelo
8. Calcular la suma de cuadrados total a partir de los resultados del punto anterior. ¿Son iguales a la estimada en el punto 4?
9. Identifica la estructura algebraica del modelo
10. Usando la estructura algebraica del modelo, calcula el valor estimado de la media de la variable predictiva para cada nivel de tu variable predictiva elegida.
11. Determinar si se rechaza o se acepta la hipótesis nula
12. Identificar la proporción de variación que explica el modelo
13. ¿Qué puedes concluir a partir de tu modelo?

