Reporte de Laboratorio 2

Coloque su nombre y apellido

fecha

# Ejercicio 1

1. Importa los datos al ambiente con el nombre **datos1** y verifica sus características y estructura.
2. Calcula la media (con la función mean) y desviación estándar (con la función sd) de la DBO en la muestra y obtén un histograma (hist) para ver su distribución. ¿Te parece que tienen una distribución normal? ¿cuál es el error estándar de la media? ¿Es un valor grande o pequeño?
3. Define la hipótesis alternativa y luego la Hipótesis nula. Aplica una prueba de *t* para una sola muestra, usando el comando t.test, pero antes de efectuar la prueba debes definir el alfa para rechazar la Hipótesis nula. Utiliza la ayuda help para escribir los argumentos correctos de t.test (tip: los mas importantes son *mu* y *alternative*)
4. A partir del resultado obtenido (output o salida) responde a las siguientes preguntas:
   1. ¿Cuál fue el valor de *t* que obtuviste?
   2. ¿Cuál es el valor de probabilidad asociado a ese valor de *t*?
   3. ¿Existen evidencias suficientes para considerar que DBO de las aguas redisuales superan los valores de la norma?
   4. ¿Cuáles son los límites inferior y superior del intervalo de confianza al 95%?
5. Vamos a calcular el intervalo de confianza paso por paso. Para ello requerimos contar con el valor de *t* que separa la curva en zonas de baja y alta probabilidad de ocurrir (dos colas, cada una de prob alfa/2). Este valor se obtiene con la función qt y requiere de tres argumentos esenciales (p = 0.025, df = 49, lower.tail = TRUE. Usa el help para tener más detalles de esta función).
   1. ¿Son un error los valores negativos de esta función? ¿Por qué?
   2. ¿Cómo puedes obtener un valor positivo?
   3. ¿Por qué razón usamos una probabilidade de 0.025, si el intervalo de confianza es de 95%?
   4. ¿Por qué usamos n-1 en vez de n?

* Ahora, multiplica el valor de t-student (absoluto) por el valor de error estándar calculado antes.
  1. ¿Qué obtuviste con esta operación?
  2. ¿Cuáles son las unidades de esta medida?
* Finalmente suma y resta el intervalo (al 95%) a la media para conocer los límites superior e inferior.
  1. Compara los valores obtenidos con el resultado de la función usada para hacer la prueba de hipótesis.

1. Calcula el intervalo de confianza al 99%?
   1. ¿En cuál de los dos intervalos (95% o 99%) tengo más confianza de encontrar la media si repito el experimento?
   2. ¿Cuál de los dos intervalos es más ancho?
   3. ¿Qué le sucedería al intervalo si en vez de 50 mediciones, tuviese solo 5 mediciones?
   4. Verifica tu conclusión (PISTA: usa la función qt)
2. ¿Qué sucedería con el IC al 95% si la DBO de los efluentes fuese mas variable? Obtén un histograma y verifica tu conclusión mediante una simulación usando la función rnorm (PISTA: mantén la muestra simulada en n=50, y la media en 40).

# Ejercicio 2

1. Importa los datos del archivo **datos2.csv** al ambiente R con el nombre **datos2** y verifica sus características y estructura. Asegúrate que el archivo esté en el ambiente de trabajo.
2. Obtén los nombres de las variables en el archivo y sigue las siguientes instrucciones:
   1. Identifica el vector que contiene la variable que se pretende analizar, y establece si ésta es contínua o discreta.
   2. Identifica el vector que contiene el factor ‘origen’. ¿Cuántos niveles tiene dicho factor?
   3. ¿Cuál es la unidad experimental en este estudio?
   4. ¿Cuántas réplicas fueron medidas en cada nivel? ¿Consideras este diseño uno balanceado?
3. Contruye un histograma con hist para para la variable DBO sin distinguir el momento, pero agregue dos líneas verticales para cada promedio. Luego obtén un gráfico de cajas-bigotes agrupando datos según el *Momento*. PISTA: usa la función boxplot, busca cómo se usa esta función. ¿cuál representación gráfica te parece más informativa? Explica la respuesta.
4. Describe las características de ambos momentos en términos de sus medias, dispersiones y la forma aproximada de sus distribuciones. Ayudate calculando estos estadísticos con las funciones ‘summary’, ‘mean’ y ‘var’. ¿Cuáles de estas características son parecidas y cuáles diferentes entre las dos muestras?
5. Formula la hipótesis del modelo y la hipótesis nula, y define si se trata de una prueba de **una** o **dos** colas. Obten el valor crítico de la distribución de *t*-student para rechazar la hipótesis nula. Considera un alfa = 0.05, y grados de libertad = n1+n2-2.
6. Aplica una prueba de *t*-student para dos muestras independientes (*S3 method for class formula*), y responde a las siguientes preguntas:
   1. ¿Cuál es el valor de t obtenido? ¿Es grande o pequeño?
   2. ¿Cuál es la probabilidad asociada al valor de t obtenido? ¿Es grande o pequeña?
   3. ¿Cuál es la conclusión de la prueba que aplicaste?
   4. ¿Podrías enlistar tres de las condiciones más importantes que deben cumplir los datos para interpretar adecuadamente e resultado de una prueba *t* clásica?
7. Considera que las varianzas de estas dos muestras son lo suficientemente similares como para asumirlas iguales, y aplica una prueba de *t*-student para varianzas iguales. (PISTA: revisa la informacion del help de t.test)
8. Compara los resultados de ambas pruebas y responde a las siguientes preguntas:
   1. ¿Qué ocurre con el valor de *t* cuando asumes homogeneidad de varianzas aún cuando no la hay?
   2. ¿Qué ocurre con el valor de probabilidad asociado al estadístico *t* cuando asumes homogeneidad de varianzas y no la hay?
   3. ¿Qué ocurre con los grado de libertad?
   4. ¿Cuál de las dos pruebas es más conservadora (i.e., tiene menos probabilidad de caer en errores tipo I)?
   5. ¿Confías más en el resultado de la primera prueba o en el de la segunda? Explica tu respuesta.