



Universidad Simón Bolívar  
Departamento de Cómputo Científico y Estadística  
Estadística Para Ingenieros CO-3321  
Prof. Desiree Villalta

## Informe de Laboratorio: Laboratorio 2. Intervalos de Confianza

Estudiantes:

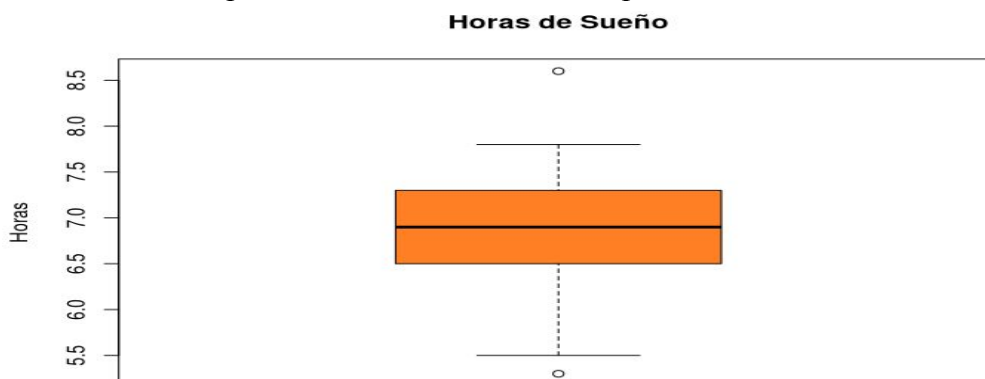
Carlos Sivira 15-11377

José Barrera 15-10123

### 1. Pregunta 1: Horas de Sueño

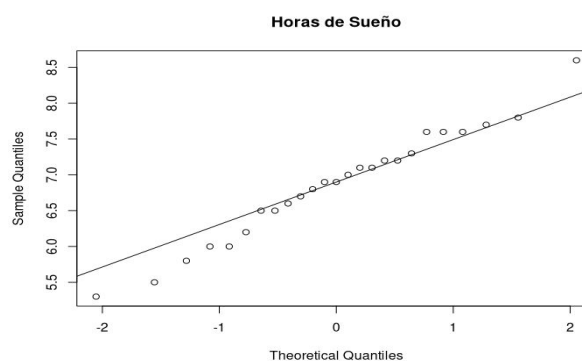
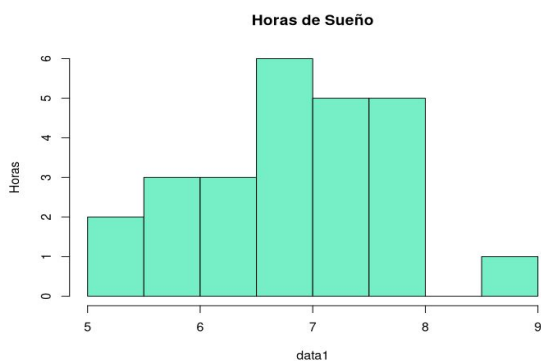
- a. Realice un análisis descriptivo de los datos. Apóyese en gráficos:

En primer lugar mediante la función `summary()` obtenemos que el valor mínimo del conjunto es 5.3 y el máximo 8.6, dichos valores representan valores atípicos como nos muestra el boxplot.



También se observa que el bigote inferior es más largo que el superior mientras que la caja es simétrica. Lo cual corresponde con el hecho de que la media y la mediana sean muy cercanas (6.90 y 6.86 respectivamente).

Por otro lado el histograma presenta cierta forma acampanada que nos recuerda a una normal, y mediante las funciones `qqnorm()` y `qqline()` se aprecia que efectivamente los datos tienden a una distribución normal.



Adicionalmente se calculó el coeficiente de variación de los datos, el resultado obtenido, 0.1133064, nos confirma que no existe una gran variabilidad en los datos.

- b. Si se asume que la población se distribuye normal presente un intervalo de confianza del 78% para la media del número de horas de sueño por noche. Analice el resultado obtenido:

Se empleó la función `t.test()`, ya que buscamos un intervalo de confianza para la media de un conjunto de 25 datos (no grande), obteniendo: [6.66422, 7.05578]. Esto nos indica que, de cada 100 veces que se repita este experimento, 78 de ellas la media se encuentra entre 6.66422 y 7.05578. De aquí también se sabe que las personas en promedio duermen entre 7 y 6.7 horas.

## 2. Pregunta 2: Test de Rendimiento

- a. Presente en su estudio un análisis descriptivo, con apoyo de un diagrama de caja.

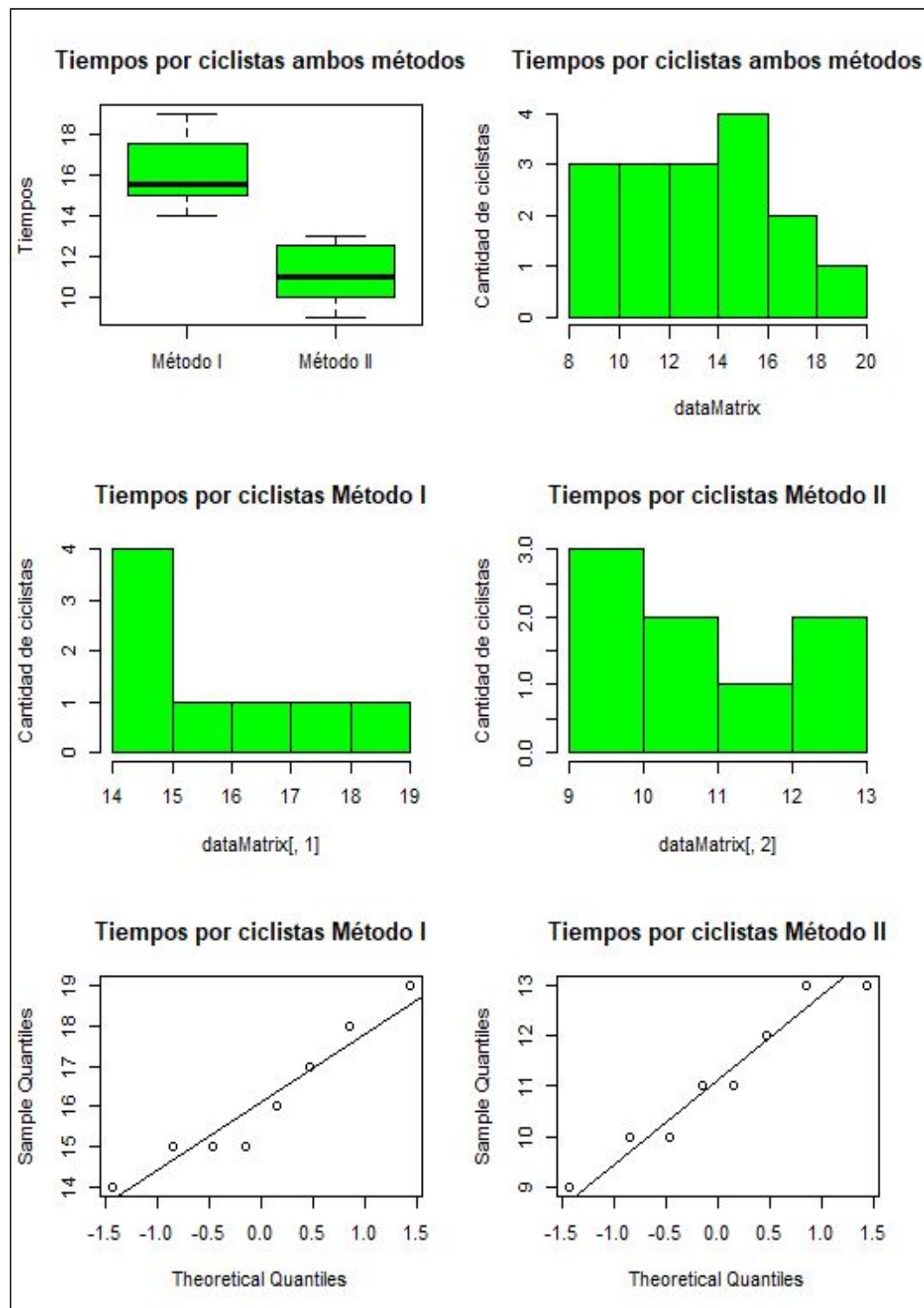
En primer lugar, se realiza el llamado de la función `summary()` para obtener la descripción de la distribución de los datos de la muestra.

Método I	Método II
Min. :14.00	Min. : 9.00
1st Qu.:15.00	1st Qu.:10.00
Median :15.50	Median :11.00
Mean :16.12	Mean :11.12
3rd Qu.:17.25	3rd Qu.:12.25
Max. :19.00	Max. :13.00

Luego, se calcula el coeficiente de variación para cada método, dejando como resultado 0.1070938 para el método 1 y 0.1310326 para el método 2. Para ambos casos el coeficiente de variación es lo suficientemente pequeño para decir que los datos varían muy poco respecto a la media. Así mismo, el primer cuartil del método 1 es mayor que el tercer cuartil del método 2. Esto indica que los resultados del método 1 son en general mayores que en el método 2, es decir, son peores los tiempos del método 1 en el test de rendimiento.

En cuanto a la distribución de los datos, los histogramas no revelan información de relevancia. Al realizar los respectivos `qqplot()` de cada método, se evidencia una tendencia de los datos a seguir una distribución normal. Por lo tanto es de suponer que los datos de la muestra tienden a una distribución normal aun cuando los histogramas no revelan esa información. Así mismo, el histograma de los resultados de ambos métodos si muestra una semejanza a una distribución normal. Pero este resultado no es relevante para este problema.

Por otra parte, los datos del enunciado no presentan dependencia alguna, por lo tanto los cálculos de intervalos de confianza no requieren la consideración de muestras pareadas.



- b. A un nivel de confianza del 95% ¿Puede considerarse que los dos métodos en promedio producen resultados equivalentes?

Para determinar si los resultados en ambas pruebas son equivalentes, se procede a calcular el intervalo de confianza del 95% de la diferencia de los promedios de cada test de rendimiento. Para esto debemos saber si las varianzas de cada método son iguales. Usando `var.test()`, obtenemos el intervalo  $[0.2809583, 7.0096625]$  que incluye al número 1, por lo tanto se asume que las varianzas son iguales. Luego, calculamos el intervalo de confianza con `t.test()` y se obtiene el rango  $[3.286326, 6.713674]$  siempre positivo. Esto indica que la diferencia de los promedios de cada test de rendimiento siempre tiende a un mayor resultado en el método 1. Esto tiene sentido ya que la media de este método es mayor a la del método 2.

Podemos concluir entonces que los resultados obtenidos por ambos métodos no son equivalentes, ya que, el intervalo de confianza sólo muestra valores del método 1 lo que es consecuencia de que su media sea mayor a la media del método 2. Además, los datos entre métodos son distintos, esto se evidencia en el diagrama de caja y en el análisis descriptivo.

- c. La variabilidad del método I es mayor a la del método II?

Al calcular la diferencia de las varianzas de cada método podemos deducir cuál tiene mayor variabilidad. Si hacemos  $\text{var}(\text{método 1}) - \text{var}(\text{método 2})$  se obtiene como resultado 0.8571429 positivo, lo que indica que la variabilidad del primer método es mayor que la del segundo.

Aun así, al usar la función `var.test()` en la pregunta anterior, se obtiene que la razón de ambas varianzas incluye al número 1. Este resultado hace pensar que no se puede asegurar que la variabilidad de un método es mayor a la del otro siempre.