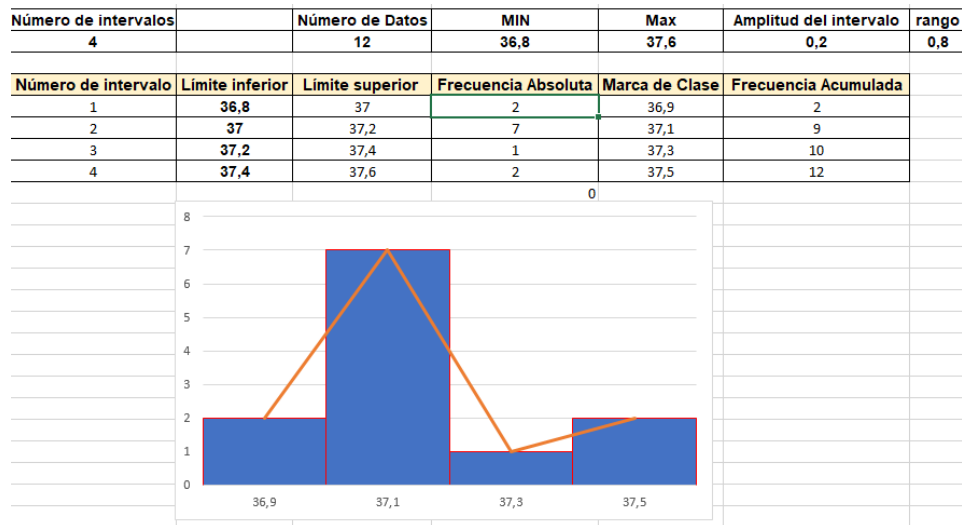


FIGURA 1.3. Ejemplo de un histograma construido a partir de los tiempos realizados.



- **Mediana:** Supongamos que las observaciones en una muestra son  $x_1, x_2, \dots, x_n$  ordenadas de forma creciente. La mediana de la muestra dependerá de si  $n$  es par o impar:

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{impar} \\ x_{\frac{n}{2}} & \text{par} \end{cases}$$

Ahumada [2] menciona otras medidas de posición que son importantes para los casos analizados en este taller que son los **cuartiles**:

- **Primer cuartil:** Se denomina primer cuartil ( $Q_1$ ) al número real tal que a lo sumo el 25 % de los datos son menores que él y a lo sumo el 75 % son mayores.
- **Segundo cuartil:** Coincide con la mediana  $\tilde{x}$ .
- **Tercer cuartil:** Se denomina tercer cuartil ( $Q_3$ ) al número real tal que a lo sumo el 75 % de los datos son menores que él y a lo sumo el 25 % son mayores.

También son importantes las medidas de dispersión:

- **Rango:** Es la diferencia entre el menor y el mayor valor de la variable.
- **Rango intercuartil:** Es la diferencia entre el primer cuartil y el tercer cuartil.
- **Varianza:** Se define como el promedio de los cuadrados de las desviaciones respecto de la media.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- **Desviación estándar:** Se define como la raíz cuadrada de la varianza. Mide la dispersión de los datos, es decir, mientras más dispersos se encuentren los datos mayor es su desviación estándar.

Ahora bien, describir un lote de datos significa hacer referencia, entre otras cosas, a la posición, dispersión, forma, que tal lote presenta para realizar un análisis y sacar conclusiones acerca del mismo. La técnica que utilizaremos para el análisis exploratorio de datos en este caso es el **Diagrama de Cajas** o comúnmente conocido como **box plot**.

Este diagrama consiste en una caja a lo largo del eje de la variable, donde se encuentra el 50 % central de los datos (o sea que incluye los dos cuartos centrales), y el resto constituyen las colas de la distribución (el primer cuarto, la cola izquierda; el cuarto, la cola derecha), representadas por segmentos a los costados de la caja. Si hay valores muy extremos, las colas no comienzan en los extremos sino que se destacan estos valores con una marca y la cola comienza en el dato inmediato siguiente. En la figura 1.4 se muestra un ejemplo de un *box plot* que muestra la información del lote de datos presentado en la tabla de la figura 1.2.

Como vemos, la simplicidad del dibujo hace que notemos rápidamente en él las características importantes, como por ejemplo:

- **Valores alejados:** se distinguen por las marcas especiales que lo separan del resto del diagrama.