1. Estadísticos de la muestra ordenada

En esta sección identificamos algunos valores característicos de una muestra de observaciones que han sido ordenadas de menor a mayor. Tales características de la muestra se denominan *estadísticas de orden*. En general, las estadísticas se calculan a partir de observaciones y se usan para hacer inferencias acerca de las características de la población de donde se extrajo la muestra.
Sea $X_1, X_2, \ldots, X_n$ los valores observados de alguna variable aleatoria, obtenidos mediante un proceso de muestreo aleatorio. Por ejemplo, consideremos los siguientes diez valores de la resistencia al corte de soldaduras de acero inoxidable (*lb/soldadura*): 2385, 2400, 2285, 2765, 2410, 2360, 2750, 2200, 2500, 2550. ¿Qué podemos hacer para caracterizar la variabilidad y la ubicación de estos valores?

El primer paso es **ordenar** los valores de la muestra en orden creciente. Es decir, reescribimos la lista de valores muestrales como: 2200, 2285, 2360, 2385, 2400, 2410, 2500, 2550, 2750, 2765. A estos valores ordenados los denotamos por $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, donde $X_{(1)} = 2200$ es el valor más pequeño de la muestra, $X_{(2)} = 2285$ es el segundo valor más pequeño, y así sucesivamente. Llamamos a $X_{(i)}$ *el i-ésimo estadístico de orden* de la muestra. Para mayor comodidad, también podemos denotar el promedio de dos estadísticos de orden consecutivos mediante

$$X_{(i,5)} \;=\; \frac{X_{(i)} + X_{(i+1)}}{2} \;=\; X_{(i)} \;+\; 0{,}5\big(X_{(i+1)} - X_{(i)}\big). \tag{2.1}$$

Por ejemplo, $X_{(2,5)} = \big(X_{(2)} + X_{(3)}\big)/2$. Ahora identificamos algunos valores característicos que dependen de estos estadísticos de orden, a saber: el *mínimo muestral*, el *máximo muestral*, el *rango muestral*, la *mediana muestral* y los *cuartiles muestrales*.

El mínimo muestral es $X_{(1)}$ y el máximo muestral es $X_{(n)}$. En nuestro ejemplo, $X_{(1)} = 2200$ y $X_{(10)} = 2765$. El *rango muestral* es la diferencia

$$R \;=\; X_{(n)} \;-\; X_{(1)} \;=\; 2765 \;-\; 2200 \;=\; 565.$$

El valor "medio" en la muestra ordenada se llama *mediana muestral*, denotada por M_e. La mediana muestral se define como $M_e = X_{(m)}$, donde $m = (n+1)/2$. En nuestro ejemplo $n = 10$, así que $m = (10+1)/2 = 5{,}5$. Por lo tanto,

$$M_e \;=\; X_{(5,5)} \;=\; \frac{X_{(5)} + X_{(6)}}{2} \;=\; X_{(5)} \;+\; 0{,}5\big(X_{(6)} - X_{(5)}\big) \;=\; \frac{2400 + 2410}{2} \;=\; 2405.$$

La mediana caracteriza el centro de dispersión de los valores de la muestra y, por lo tanto, se denomina estadístico de tendencia central o estadístico de localización. Aproximadamente el 50 % de los valores muestrales son menores que la mediana. Los **cuartiles muestrales** se definen como

$$Q_1 \;=\; X_{(q_1)} \quad \text{y} \quad Q_3 \;=\; X_{(q_3)},$$

donde

$$q_1 \;=\; \frac{n+1}{4} \quad \text{y} \quad q_3 \;=\; \frac{3(n+1)}{4}. \tag{2.2}$$

A Q_1 se le llama *cuartil inferior* y a Q_3 se le llama *cuartil superior*. Estos cuartiles dividen la muestra de modo que aproximadamente una cuarta parte de los valores son menores que Q_1, la mitad están entre Q_1 y Q_3, y una cuarta parte es mayor que Q_3.

En nuestro ejemplo $n = 10$, por lo que

$$q_1 = \frac{11}{4} = 2{,}75 \quad \text{y} \quad q_3 = \frac{33}{4} = 8{,}25.$$

Por lo tanto,

$$Q_1 = X_{(2{,}75)} = X_{(2)} + 0{,}75\left(X_{(3)} - X_{(2)}\right) = 2341{,}25,$$

$$Q_3 = X_{(8{,}25)} = X_{(8)} + 0{,}25\left(X_{(9)} - X_{(8)}\right) = 2600.$$

La mediana muestral y los cuartiles son casos específicos de una clase de estadísticas conocidas como *cuantiles muestrales*. El cuantil muestral p-ésimo es un número que supera exactamente el $100p\,\%$ de los valores de la muestra. Por lo tanto, la mediana es el cuantil muestral de $p = 0{,}5$, Q_1 es el cuantil de $p = 0{,}25$ y Q_3 es el cuantil de $p = 0{,}75$. [1]

[1] Tomado de: Estadística industrial moderna : diseño y control de la calidad y la confiabilidad Kenett, Ron; Zacks, Shelemyahu.