

DIAGRAMA DE TALLO Y HOJAS

Un procedimiento semi-gráfico de presentar la información para variables cuantitativas, que es especialmente útil cuando el número total de datos es pequeño (menor que 50), es el diagrama de tallo y hojas de Tukey. Los principios para constituirlo son:

- Redondear los datos a dos o tres cifras significativas, expresándolos en unidades convenientes.
- Disponerlos en una tabla con dos columnas separadas por una línea como sigue:
 - a) Para datos con dos dígitos, escribir a la izquierda de la línea los dígitos de las decenas (forma el tallo), y a la derecha las unidades (hojas).
 - b) Para datos con tres dígitos el tallo estará formado por los dígitos de las centenas y decenas, que se escribirán a la izquierda, separados de las unidades.
- Cada tallo define una clase, y se escribe sólo una vez. El número de hojas representa la frecuencia de dicha clase.

Ejemplo

1. Datos recogidos en cm:

11,357; 12,542; 11,384; 12,431; 14,212; 15,213; 13,300; 11,300; 17,206; 12,710; 13,455; 16,143; 12,162; 12,721; 13,420; 14,698.

2. Datos redondeados expresados en mm:

114; 125; 114; 124; 142; 152; 133; 113; 172; 127; 135; 161; 122, 127; 134; 147.

3. Diagrama de tallo y hojas, datos en mm:

11		443
12		54727
13		354
14		27
15		2
16		1
17		2
decenas		unidades

Cuando el primer dígito de la clasificación varía poco, la mayoría de los datos tienden agruparse alrededor de un tallo y el diagrama resultante tiene poco detalle. En ese caso

es conveniente subdividir cada tallo en dos o más partes introduciendo algún signo arbitrario, como se indica:

- Las pulsaciones por minuto de un grupo de 40 personas se han representado en el diagrama de tallo y hojas siguiente:

5	2 6
6	0 0 0 0 0 0 4 4 4 4 4 4 8 8 8 8 8 8 8
7	2 2 2 2 2 2 2 2 6 6 6 6 6
8	0 0 4 4 8 8
9	2

- Podemos obtener más detalle subdividiendo cada tallo en dos partes iguales: en una colocaremos las hojas 0 a 4 y lo representamos por (*) y en la otra las hojas de 5 a 9 y lo representaremos por (.), obteniendo el diagrama:

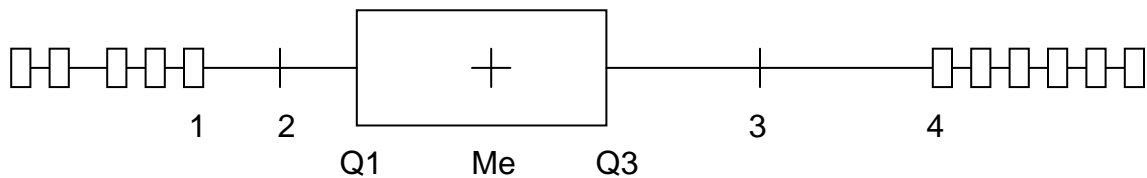
5	*	2
	.	6
6	*	0 0 0 0 0 0 4 4 4 4 4 4
	.	8 8 8 8 8 8 8 8
7	*	2 2 2 2 2 2 2 2
	.	6 6 6 6 6
8	*	0 0 4 4
	.	8 8
9	*	2
	.	

Observemos que todos los datos son múltiplos de 4, lo que hace sospechar que se han obtenido midiendo la pulsaciones cada 15 segundos y multiplicando por cuatro.

CHEQUEO DE PUNTOS EXTREMOS Y RECOMENDACIONES

Si la distribución es aproximadamente normal, se declaran puntos extremos (outliers) aquellos que caen por fuera del intervalo $\bar{X} \pm 2.7*S$ ya que $P(\mu - 2.7*S < X < \mu + 2.7*S) = 0.993$

Si la distribución es asimétrica, se acostumbra dividir la serie en áreas o segmentos como se muestra en la siguiente gráfica, llamada gráfica de Box and Whisker o caja esquemática o diagrama de bigotes,



Donde:

- 1: $Q_1 - 3.0R_q$
- 2: $Q_1 - 1.50R_q$
- 3: $Q_3 + 1.5R_q$
- 4: $Q_3 + 3.0R_q$

Obsérvese que los puntos a distancias menores de la representación 1 o mayores de la representación 4, son puntos extremos.

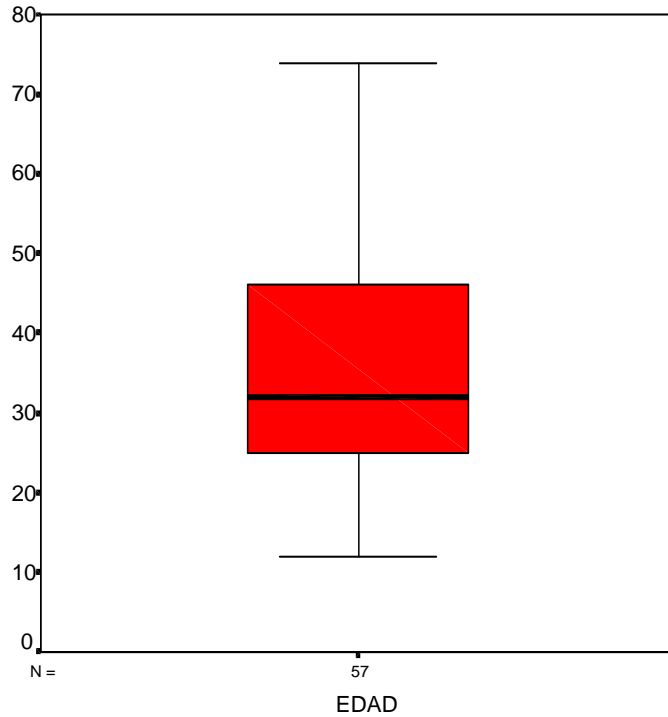
Los puntos entre la representación 3 y 4 se consideran como puntos de advertencia; los puntos entre la representación 2 y 3 se consideran como puntos normales.

DIAGRAMA DE BOX – WHISKERS

También llamado Caja y bigotes o Box–Plot, utilizado para averiguar la simetría de una variable y la presencia de datos atípicos. Consta de una caja rectangular cuyo largo es proporcional al rango intercuartílico, cuyo bigote izquierdo o inferior es proporcional a la diferencia entre el primer cuartil y el mínimo, cuyo bigote derecho o superior es proporcional a la diferencia entre el máximo y el tercer cuartil y en la cual se han dibujado la media con una x y la mediana con una línea paralela a la base y que atraviesa la caja a lo ancho. Cuando la variable examinada es simétrica la caja también lo es y entonces media y mediana coinciden. Cualquier asimetría se refleja en una asimetría más o menos pronunciada en la caja.

Ejemplo: La tabla siguiente muestra la edad en años cumplidos de 57 personas que participaron en un estudio económico.

68	65	12	22	30	51	32	74	27
63	43	32	43	36	36	28	31	22
42	25	49	27	42	28	27	31	28
27	74	38	49	28	42	23	19	50
46	23	25	24	45	38	30	25	12
21	43	44	57	16	49	51	24	12
69	47	23						



Al observar la gráfica se puede determinar que la línea mediana está mas cerca del bigote izquierdo que del derecho, por lo tanto hay una asimetría positiva o a la derecha, es decir los datos están sesgados a la derecha.