MEDIDAS DE RESUMEN

Jorge Galbiati Riesco

Las medidas de resumen sirven para describir en forma resumida un conjunto de datos que constituyen una muestra tomada de alguna población.

Podemos distinguir cuatro grupos de medidas de resumen: las medidas de centro, las medidas de posición, (las de centro son casos especiales de estas últimas), las medidas de dispersión y las medidas de forma.

Supóngase que se dispone de una muestra de observaciones x_1, x_2, ... x_n . Con estas observaciones se efectuarán los cálculos de todas las medidas de resumen que se presentan a continuación.

A modo de ejemplo, se dispone de tres muestras de datos con las que se obtendrán las medidas de resumen. Los tres están ordenados de menor a mayor, por columnas.

50	140	175	270	430
50	150	180	280	450
80	150	185	285	460
80	150	190	290	500
90	150	190	295	510
90	150	195	350	
90	150	250	350	
95	150	250	365	
130	160	250	370	
140	170	250	395	

Muestra 1. Ingresos de 45 empleados de una firma (miles de pesos).

8	14	17	21
8	15	17	22
9	15	18	23
10	15	18	23
12	16	18	23
12	16	19	24
13	16	19	25
14	17	19	27
14	17	20	
14	17	20	

Muestra 2. Pesos de bultos transportados por un correo (kgs).

2	9	11	12	14
4	9	11	12	14
4	9	11	12	
7	9	11	12	
7	9	12	12	
8	9	12	12	
8	9	12	13	
8	9	12	13	
8	9	12	14	
8	10	12	14	

Muestra 3. Escolaridad de los habitantes adultos de un condominio (años).

A continuación se presentan tablas de frecuencia por intervalos e histogramas de los tres conjuntos de datos. Se puede observar que el primer conjunto tiene sesgo (la cola) hacia la derecha; el segundo es bastante simétrico; y el tercero tiene sesgo hacia la izquierda.

Clase	Frecuenc.	Frec acumul.	Frec. relativa %
1 a 100	8	8	17.8
101 a 200	18	26	40.0
201 a 300	9	35	20.0
301 a 400	5	40	11.1
401 a 500	4	44	8.9
501 a 600	1	45	2.2
total	45	--	100

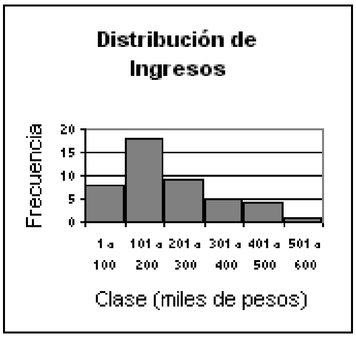


Figura 1. Ingresos de empleados. Tabla de frecuencias e histograma.

Clase	Frecuenc.	Frec acumul.	Frec. relativa %
7 a 9	3	3	7.9
10 a 12	3	6	7.9
13 a 15	8	14	21.0
16 a 18	11	25	28.9
19 a 21	6	31	15.8
22 a 24	5	36	13.2
25 a 27	2	38	5.3
total	38	--	100

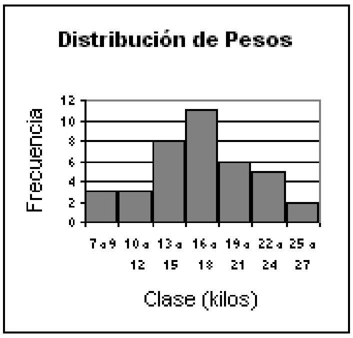


Figura 2. Pesos de bultos. Tabla de frecuencias e histograma.

Clase	Frecuenc.	Frec acumul.	Frec. relativa %
1 a 2	1	1	2.4
3 a 4	2	3	4.7
5 a 6	0	3	0
7 a 8	7	10	16.7
9 a 10	10	20	23.8
11 a 12	16	36	38.1
13 a 14	6	42	14.3
total	42	--	100

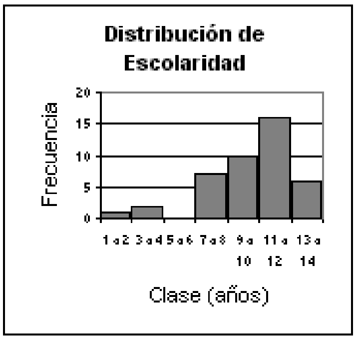


Figura 3. Escolaridad. Tabla de frecuencias e histograma.

Medidas de Centro

Son medidas que pretenden indicar dónde está lo que se podría considerar como el centro de la masa de datos. Promedio o media. Es la suma de todas las observaciones, dividida por el número de ellas. Las más conocidas son las siguientes:

Promedio o Media.

Es igual a la suma de todas las observaciones, dividida por el número de observaciones. Se usa el símbolo $\overline{x}$ para representar la media.

El promedio está dado por la fórmula $$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

El promedio es una medida muy influenciada por valores extremos. Por lo tanto, si los datos presentan mucha asimetría, el promedio resulta distorsionado.

Ejemplo 1. Cálculo del promedio con los datos presentados al inicio:

	Muestra 1	Muestra 2	Muestra 3
Suma de los datos	10130	645	425
Número de datos	n = 45	38	42
Promedio	10130/45 = 225.1 M$	645/38 = 17.0 kilos	425/42 = 10.1 años

Mediana.

Es un número tal que al menos el 50% de las observaciones son menores o iguales a él, y al menos el 50% son mayores o iguales a él. La mediana es muy resistente a valores extremos. La representamos por el símbolo *Mn*.

Se calcula de la siguiente forma:

1 - Se ordenan las observaciones, de menor a mayor.

2 - Si el número n de observaciones es impar, la mediana es la que queda exactamente al centro.

3 - Si el número de observaciones es par, la mediana es el promedio de las dos observaciones centrales.

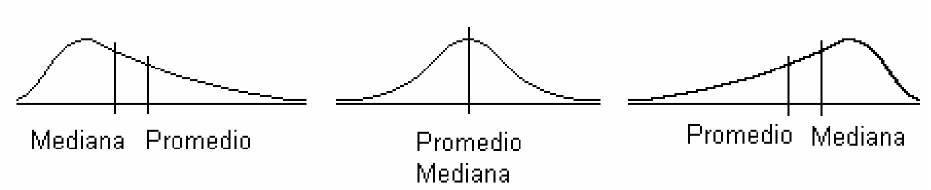


Figura 4. Comparación entre promedios y medianas en distintos casos.

Ejemplo 2. Cálculo de la mediana:

	Muestra 1	Muestra 2	Muestra 3
Número de datos	n = 45	38	42
Ubicación de la mediana	La mediana es la observación de orden 23	La mediana es el promedio de las observaciones 19 y 20	La mediana es el promedio de las observaciones 21 y 22
Mediana	Mn = 185 M$	Mn = (17+17)/2 = 17 kilos	Mn = (11+11)/2 = 11 años

Comparando con el ejmplo 1, se puede ver que el promedio es mayor que la mediana cuando hay sesgo hacia la derecha. Las observaciones extremas influyen más sobre el promedio que sobre la mediana, y lo desplazan a la derecha.
En el caso simétrico, ambas medidas coinciden.
Y cuando hay sesgo a la izquierda, el promedio está más a la izquierda que la mediana.

Promedio Recortado o **Truncado en α%.**
Es el promedio del conjunto de observaciones, al cual se le ha eliminado un porcentaje predeterminado α de las observaciones más pequeñas y el mismo porcentaje de las más grandes.
Para calcularlo se deben ordenar las observaciones, eliminar un porcentaje α%, previamente definido, en ambos extremos. Luego se obtiene el promedio de las
(100-2α)% observaciones centrales. Valores típicos de α son 5 o 10 %.
El promedio recortado es más resistente que la media, a valores extremos, precisamente porque se eliminaron los extremos. Pero no tan insensible como la mediana, en casos en que hay muchos valores que se alejan significativamente del promedio.

Ejemplo 3: Promedio recortado al 5%:

	Muestra 1	Muestra 2	Muestra 3
Número de datos	n = 45	38	42
Número de datos recortados por lado	2.25 que lo aproximamos a 2	1.9, aproximado a 2	2.1, aproximado a 2
Cálculo del promedio recortado	El promedio recortado al 5% es el promedio de las 41 observaciones centrales	El promedio recortado al 5% es el promedio de las 34 observaciones centrales	El promedio recortado al 5% es el promedio de las 38 observaciones centrales
Promedio recortado	(80+80+90+... ..+430+450+ 460)/41 = 220.0 M$	17.0 kilos	10.3 años

Ejemplo 4: Promedio recortado al 20%:

	Muestra 1	Muestra 2	Muestra 3
Número de datos recortados por lado	9 datos por lado	7.6 datos, que lo aproximamos a 8	8.4 datos, aproximado a 8
Cálculo del promedio recortado	El promedio recortado al 20% es el promedio de las 27 observaciones centrales	El promedio recortado al 20% es el promedio de las 22 observaciones centrales	El promedio recortado al 20% es el promedio de las 26 observaciones centrales
Promedio recortado	(140+140+150+... +350)/27=205.4 M$	16.9 kilos	10.4 años

Se observa que si se se recorta poco, el promedio recortado se parece al promedio simple. Sin embargo, si se recorta más, se parece más a la mediana.

Medidas de Posición

Señalan otras posiciones, aparte del centro, dentro de la masa de datos. Por ejemplo, a partir de qué valores está el 10% mayor. Las medidas de centro son casos especiales de medidas de posición.

Percentil q, en que q es un número entero entre 1 y 99.
El percentil q es un número P_q tal que:
1 - Al menos q% de las observaciones son menores o iguales que él.
2 - Al menos (100-q)% de las observaciones son mayores o iguales que él.

El percentil q se obtiene de la siguiente forma:
1 - Se ordenan las observaciones, de menor a mayor.
2 - Se calcula el valor de r=q x n/100, en que n es el número de observaciones.
3 - Si el resultado de r es entero, el percentil P_q es el promedio de las observaciones que ocupan los lugares r y r+1.
4.- Si el resultado no es entero, se aproxima al entero <u>superior</u> y el percentil P_q es la observación que ocupa ese lugar.

Casos Particulares de percentiles:

La Mediana. Es el percentil 50.
Cuartiles. Son los percentiles de orden 25, 50 y 75, representados por Q_1, Q_2 y Q_3, respectivamente. Dividen la muestra ordenada en cuatro grupos que contienen la cuarta parte de las observaciones cada uno.
Quintiles. Son los percentiles de orden 20, 40, 60 y 80. Dividen la muestra ordenada en cinco grupos de igual tamaño.
Deciles. Son los percentiles de orden 10, 20, 30, 40, 50, 60,
70, 80 y 90. Dividen la muestra ordenada en diez grupos de igual tamaño.

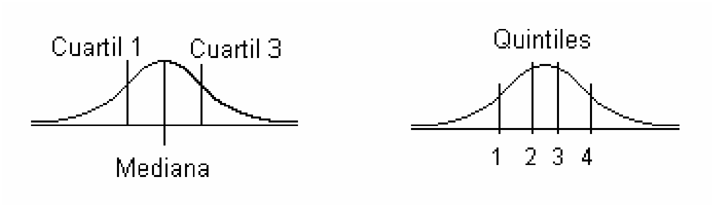


Figura 5. Representación de Cuartiles y Quintiles en un conjunto de datos simétrico.

Ejemplo 5. Cálculo del cuartil 1.

	Muestra 1	Muestra 2	Muestra 3
núm. de datos	n = 45	38	42
25*n/100	11.25	9.5	10.5
ubicación del cuartil Q_1	El cuartil 1 es la observación de orden 12	El cuartil 1 es la observación de orden 10	El cuartil 1 es la observación de orden 11
Cuartil Q_1	150 M$	14 kilos	9 años

Ejemplo 6. Cálculo del cuartil 2 o mediana.

	Muestra 1	Muestra 2	Muestra 3
50*n/100	22.5	19	21
ubicación del cuartil Q_2	El cuartil 2 es la observación de orden 23	El cuartil 2 es el promedio de las obs. de lugares 19 y 20	El cuartil 2 es el promedio de las obs. de lugares 21 y 22
Cuartil Q_2	185 M$	17 kilos	11 años

Observar que el cuartil 2 es igual a la mediana, del ejemplo 2, en todos los casos.

Ejemplo 7. Cálculo del cuartil 3.

	Muestra 1	Muestra 2	Muestra 3
75*n/100	33.75	28.5	31.5
ubicación del cuartil Q_3	El cuartil 3 es la observación de orden 34	El cuartil 3 es la observación de orden 29	El cuartil 3 es la observación de orden 32
Cuartil Q_3	290 M$	20 kilos	12 años

Medidas de Dispersión

Describen el grado de dispersión de los datos, es decir, cuán separados se encuentran, como opuesto a datos que estan muy concentrados o cercanos entre sí. Las más conocidas son las siguientes:

Rango.
Es la diferencia entre el mayor valor y el menor.
Depende sólo de dos observaciones, y justamente de las más extremas,
por lo que, en general es una muy mala medida de dispersión.

$$R = x_n - x_i$$

Ejemplo 8. Cálculo del rango.

	Muestra 1 (Ingresos)	Muestra 2 (Pesos)	Muestra 3 (Escolaridad)
Mínimo	50	8	2
Máximo	510	27	14
Rango	510-50=460 M$	27-8=19 kgs.	14-2=12 años

Desviación Media.
Es el promedio de las desviaciones absolutas (en valor absoluto) respecto de la media. Se mide en las mismas unidades que las observaciones originales.

$$DMd = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$\bar{x}$ es la media o promedio

Ejemplo 9. Cálculo desviación media.

	Muestra 1 (Ingresos)	Muestra 2 (Pesos)	Muestra 3 (Escolaridad)
Número de datos	45	38	42
Media	225.1	17.0	10.1
desviaciones absolutas	175.1, 175.1, 145.1, .., 274.9, 284.9	9.0, 9.0, 8.0,... .., 8.0, 10.0	8.1, 6.1, 6.1, ... 3.9, 3.9
suma	4645.8	135.1	94.8
promedio	4645.7/45=103.2 M$	135.1/38=3.6 kg	94.8/42=2.3 años

Desviación Mediana.

Es parecida a la desviación media. Se define como el promedio de las desviaciones absolutas (en valor absoluto) respecto de la mediana. También se mide en las mismas unidades que las observaciones originales. Tiene la característica de ser poco sensible a observaciones extremas.

$$DMn = \frac{\sum_{i=1}^{n} |x_i - Mn|}{n}$$

en que *Mn* es la mediana

Ejemplo 10. Cálculo desviación mediana.

	Muestra 1	Muestra 2	Muestra 3
Número de datos	45	38	42
Mediana	185	17	11
desviaciones absolutas	135, 135, 105,105... ...315, 325	9, 9, 8, 7... ...7,8,10	9, 7, 7, 4,... ...3, 3
suma	4405	135	93
DMd	4405/45= 97.9 M$	135/38=3.6 kg	93/42=2.2 años

Ambas medidas de dispersión, la desviación media y la desviación mediana dan resultados parecidos, y en un iguales.

Varianza.

Es un promedio de los cuadrados de las desviaciones respecto de la media, excepto que en lugar de dividir por n (el número de observaciones), se suele dividir por

$n-1$. Se usa el símbolo s^2 o bien *var* para representar la varianza.

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

Su unidad de medida es el cuadrado de unidades en que se midieron las observaciones originales. Hay una forma alternativa de calcularla, que da el mismo resultado:

$$s^2 = \text{var} = \frac{\sum_{i=1}^{n} x_i^{\,2} - n \cdot \bar{x}^2}{n-1}$$

Ejemplo 11. Cálculo varianza.

	Muestra 1	Muestra 2	Muestra 3
desviaciones al cuadrado (ver ejemplo 8)	$(-175.1)^2$, $(-175.1)^2$, $(-145.1)^2$,...274.9^2, 284.9^2= 30663.9, 30663.9, 21057.2,..., 75563.9, 81168.7	$(-9)^2$, $(-9)^2$, $(-8)^2$, $(-7)^2$...7^2,8^2, 10^2 = 81, 81, 64, 49, ..., 49.4, 64.4,100.5	$(-8.1)^2$, $(-6.1)^2$, $(-6.1)^2$,...3.9^2, 3.9^2= 65.6, 37.4, 37.4, ..., 15.21, 15.21
suma	680224.44	786.97	316.42
Divisor	44	37	41
s^2	15459.6 $(M\$)^2$	21.3 Kg^2	7.7 $años^2$

Observar que la unidad de medida es el cuadrado de las unidades originales.

Desviación estándar.

Es la raíz cuadrada de la varianza. Se mide en las mismas unidades que las observaciones originales. Se usa el símbolo *s* o bien *ds* para representar la desviación estándar.

$$s = ds = \sqrt{\text{var}}$$

Tanto la desviación media como la varianza y la desviación estándar se encuentran muy influidas por valores extremos. Por lo tanto, cuando la muestra presenta mucha asimetría, estos no son buenos indicadores de la dispersión, pues están sobrevalorados.

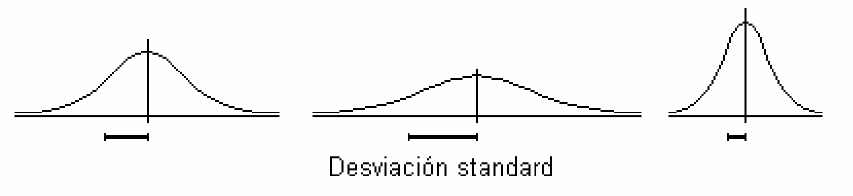


Figura 6. Dispersión y desviaciones estándar.

Ejemplo 12. Desviación estándar.

	Muestra 1	Muestra 2	Muestra 3
Varianza s^2	15459.6	21.3	7.7
desviación standard $\sqrt{s^2}$	124.3 M$	4.3 kg	2.8 años

Aunque mide el mismo concepto que las desviaciones media y mediana, y están expresados en la misma unidad de medida, los números están en escalas diferentes y no son comparables.

Coeficiente de Variación.

Es similar a la desviación estándar, pero dividido por la media. Con esto se logra que sea independiente de la unidad de medida con que se midieron las observaciones. El coeficiente de variación no tiene unidad de medida.

$$cv = \frac{ds}{\bar{x}}$$

Tiene una limitación, y es que sólo puede utilizarse cuando los datos se midieron en una escala que sólo admite valores positivos.

Ejemplo 13. Coeficiente de variación.

	Muestra 1	Muestra 2	Muestra 3
desviación standard	124.3 M$	4.6 kg	2.8 años
media	225.1 M$	17.0 kg	10.1 años
cv	124.3/225.1=0.552	4.6/17.0=0.274	2.8/10.1=0.272

El coeficiente de variación permite comparar dispersiones entre datos expresados en escas diferentes, como en este caso. Se puede concluir que la muestra 1 tiene mayor dispersión que las muestras 2 y 3, y estas últimas tienen similares dispersiones.

Desviación intercuartil o **Rango Intercuartil**.

Es la diferencia entre los cuartiles 3 y 1. Es decir, es el rango del 50\% de las observaciones centrales, las más representativas de la masa de datos. Tiene la propiedad de ser muy resistente a valores extremos.

$$DIC = Q_3 - Q_1$$

Ejemplo 14. Desviación inter cuartil.

	Muestra 1	Muestra 2	Muestra 3
Cuartiles 1 y 3 (ejemplos 5 y 7)	150 y 290	14 y 20	9 y 12
DIC	140 M$	6 kg	3 años

<u>Medidas de Forma.</u>

Como un complemento a la posición y la dispersión de una muestra de datos, puede ser útil describir algunas características de su forma

Coeficiente de simetría.

Cuantifica el grado de asimetría que presenta la muestra. Se define como el promedio de los cubos de las desviaciones en torno a la media, dividido por la desviación standard elevada también al cubo. La fórmula es

$$cs = \frac{{}^{1}\!/_{n} \sum_{i=1}^{n} (x_i - \bar{x})^3}{s^3}$$

en que *s* es la desviación estándar

Si los datos presentan una cola larga hacia la derecha, el coeficiente de simetría es positivo. Si presentan una cola larga hacia la hacia la izquierda, el coeficiente de simetría es negativo. Si hay simetría, el coeficiente es cercano a cero.

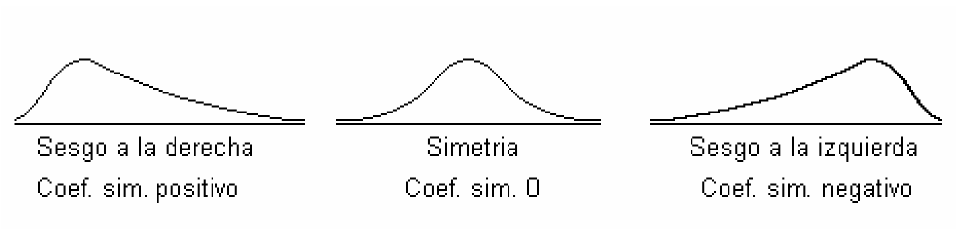


Figura 7. Coeficientes de simetría en distintos casos.

Ejemplo 15. Coeficiente de simetría

	Muestra 1	Muestra 2	Muestra 3
Suma de las desviaciones al cubo	60201633.46	73.1	-750.073
Promedio	1337814.1	1.9	-17.86
Desviación standard *ds*	124.3	4.6	2.8
ds al cubo	1922203.383	98.093	21.438
C. simetría *cs*	0.70	0.02	-0.83

El signo positivo del coeficiente de simetría de la muestra 1 indica que tiene sesgo hacia la derecha. El coeficiente de simetría de la muestra 2 indica que no tiene sesgo. La muestra 3 tiene sesgo hacia la izquierda.

Coeficiente de curtosis.

Cuantifica el hecho que la masa de datos presenta una forma de campana (mesocúrtica), una forma más bien puntiaguda en la parte central (leptocúrtica) o muy plana (platicúrtica). El coeficiente de curtosis se define como el promedio de las desviaciones elevadas a la cuarta potencia, respecto de la media, dividido por la desviación standard elevado a la cuarta. A todo esto se le resta el número 3. La fórmula es

$$k = \frac{\frac{1}{n}\sum_{i=1}(x_i - \bar{x})^4}{s^4} - 3$$

Los datos con forma de campana (mesocúrticos) tienen un coeficiente de curtosis cercano a cero. Si son leptocúrticos o con forma puntiaguda, el coeficiente es negativo. Si son planos o platicúrticos, su coeficiente de curtosis es positivo.

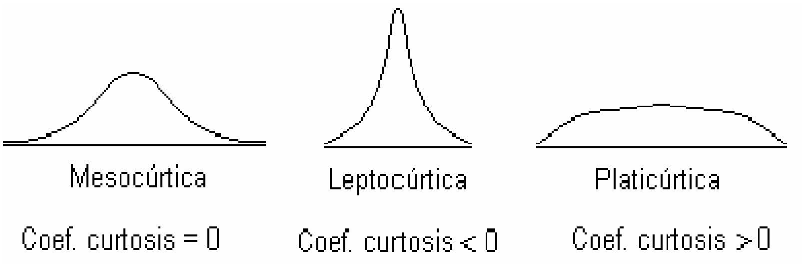


Figura 8. Curtosis para conjuntos de datos de diferentes formas.

Ejemplo 16. Coeficiente de curtosis.

	Muestra 1	Muestra 2	Muestra 3
Promedio de las desviaciones a la cuarta	586243699.0	1131.1	205.982
ds a la cuarta	239000668.81	452.394	59.555
Curtosis *k*	-0.55	-0.50	0.46

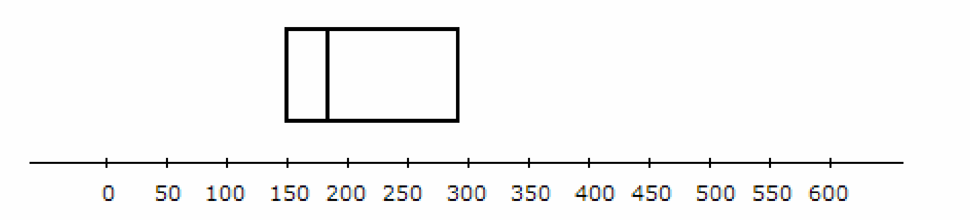
Los primeros dos conjuntos aparecen con forma lepticúrtica (puntiagudos), mientras el de la muestra 3 aparece con forma platicúrtica (más plano). Eso se puede apreciar por el hecho que las tres barras más grandes, en el histograma correspondiente a este tercer conjunto, tienen alturas similares. Si se comparan con los histogramas de los primeros dos conjuntos, hay más diferencia entre la barra más alta y las que le siguen.

Diagramas de cajón

El diagrama de cajón o cajagrama es una representación gráfica basada en medidas resistentes, como la mediana, los cuartiles y la desviación intercuartil.

Se construye dibujando una línea horizontal que con una escala que representa el rango de las observaciones. Se representa la mediana mediante un pequeño trazo vertical. A los lados se dibujan dos trazos iguales, que representan los cuartiles. Dos trazos horizontales cierran el rectángulo, denominado <u>cajón</u>, que tiene los cuartiles por lados, y contiene la mediana en el interior.

Recordamos que para la muestra 1, ingresos de 45 empleados de una firma, en miles de pesos, el cuartil 1 es Q1 = 150 M$, la mediana es M = 185 M$ y el cuartil 3 es Q_3 = 290 M$. El cajón se muestra a continuación.



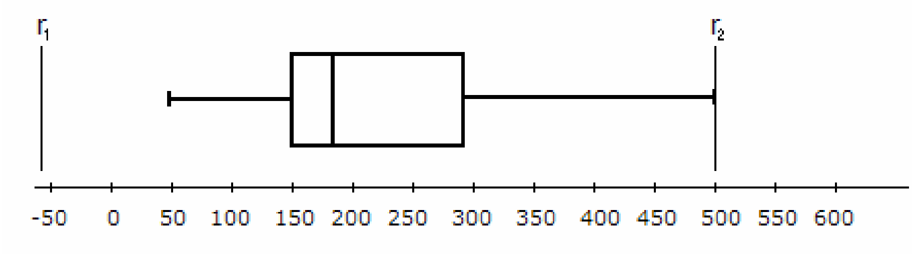
Luego se calcula la desviación intercuartil DIC = $Q_3 - Q_1$ = 290 – 150 = 140
Se dibujan dos trazos verticales, denominadas <u>rejas</u>, en forma provisoria, a distancias

$r_1 = Q_1 - 1.5*DIC = 150 - 1.5*140 = 150 - 210 = -60$

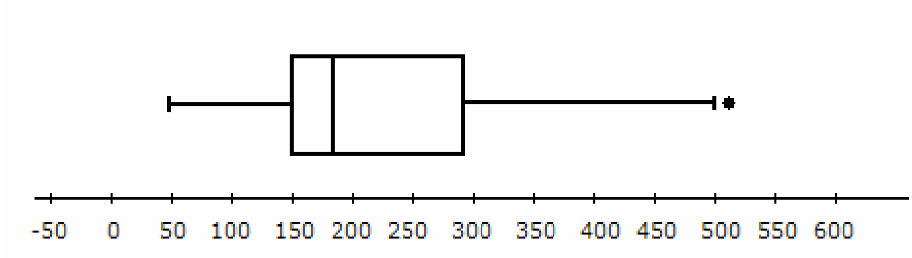
y $r_2 = Q_3 + 1.5*DIC = 290 + 1.5*140 = 290 + 210 = 500$

Las observaciones que quedan dentro de estos dos límites, se denominan observaciones interiores. La menor y la mayor de las observaciones interiores se denominan adyacentes. En nuestro ejemplo, son los valores 50 y 500.

Se dibujan dos líneas horizontales hacia ambos lados del cajón, hasta las respectivas observaciones adyacentes. Estas líneas se denominan bigotes. La figura siguiente representa el cajón con sus bigotes.



Finalmente, las observaciones que quedan fuera de las rejas se denominas observaciones exteriores, o extremas (outliers, en inglés). Estas se representan individualmente, mediante puntos o asteriscos. En el ejemplo, hay un valor extremo, que es el 510. El diagrama de cajón queda como en la siguiente figura:

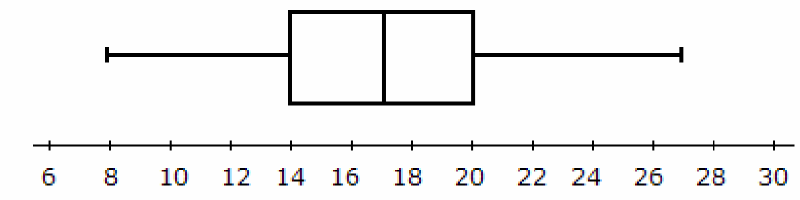


Se aprecia la forma asimétrica del conjunto de datos, con una cola hacia la derecha.

Veamos cómo son los cajagramas correspondientes a las otras dos muestras.

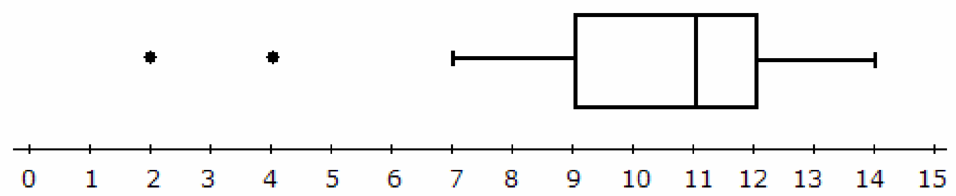
Pesos de bultos transportados por un correo:

La mediana es 17, los cuartiles 1 y 3 son 14 y 20, respectivamente. La desviación intercuartil es 20-14=6, las rejas son 14 - 1.5*6 = 5 y 20 + 1.5*6 = 29. Por lo tanto los valores adyacentes son 8 y 27. No hay valores. El cajagrama es el siguiente, que muestra la forma simétrica de este conjunto de datos:



Escolaridad de los habitantes de un condominio:

En este caso la mediana es 11, los cuartiles 1 y 3 son 9 y 12, respectivamente. La desviación intercuartil es 3, las rejas son 4.5 y 12.5. Por lo tanto los valores adyacentes son 7 y 14. Hay tres valores extremos en el lado izquierdo, 2,4 y 4. El cajagrama, entonces, es



Se ve claramente la asimetría con una cola larga hacia los valores pequeños.

<u>PREGUNTAS</u>

1. ¿Qué ventajas tiene la mediana sobre el promedio, como medida de centro?

2. ¿Cuándo se prefiere usar la mediana en lugar del promedio, como medida de centro?

3. ¿Qué ventajas tiene el promedio sobre la mediana, como medida de centro?

4. ¿Cuándo se prefiere usar el promedio en lugar de la mediana, como medida de centro?

5. ¿Qué tiene de malo el rango, como medida de dispersión?

6. ¿En qué unidad se expresan las siguientes medidas de dispersión, en relación a la unidad de medida de los datos originales?

Rango, desviación media, varianza, desviación stándar, desviación intercuartil.

7. Un conjunto de 50 datos presenta las siguientes medidas descriptivas:

El cuartil 1 es 16, la mediama es 25, el cuartil 3 es 43, el promedio es 32.5.

Explique qué forma tiene el conjunto de datos (puede ayudarse con un dibujo).

8. Se tiene un conjunto de datos, y resulta que la mediana está a la izquierda del promedio. ¿Qué características presentan esos datos?

9. Se tiene un conjunto de datos. El valor más grande es 23, pero por un error de digitación, se registró como 203. ¿En qué afecta este error a la desviación intercuartil? ¿En qué afecta a la mediana?

10. Un conjunto de 40 datos presenta las siguientes medidas descriptivas:
El rango es 100, la desviación standard es 28.4, el rango inter cuartil es 19.3.
Explique qué forma tiene el conjunto de datos (puede ayudarse con un dibujo).

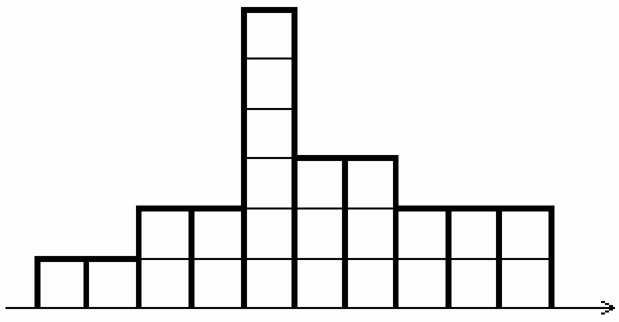
11. Luis y Felipe son alumnos de cuarto medio, de un total de 100 alumnos del colegio. Dieron una prueba especial, y Luis estuvo en el percentil 85, mientras que Felipe en el percentil 62. ¿Cuál de los dos estuvo mejor, y aproximadamente cuántos alumnos obtuvieron puntajes entre los de Luis y Felipe?

12. Diga qué característica de una muestra de datos mide cada uno de los indicadores:
La mediana, la media recortada, La desviación estándar, el coeficiente de variación, el cuartil 3.

13. Un conjunto de 30 datos presenta las siguientes medidas descriptivas:
El promedio es 23.6, el promedio recortado al 10% es 16.9.
Explique qué característica tiene el conjunto de datos.

14. Describa las medidas descriptivas que se representan en un diagrama de caja. ¿Cuál es su característica común?

15. El siguiente histograma describe un conjunto de datos. Dibuje el diagrama de caja, a la misma escala.



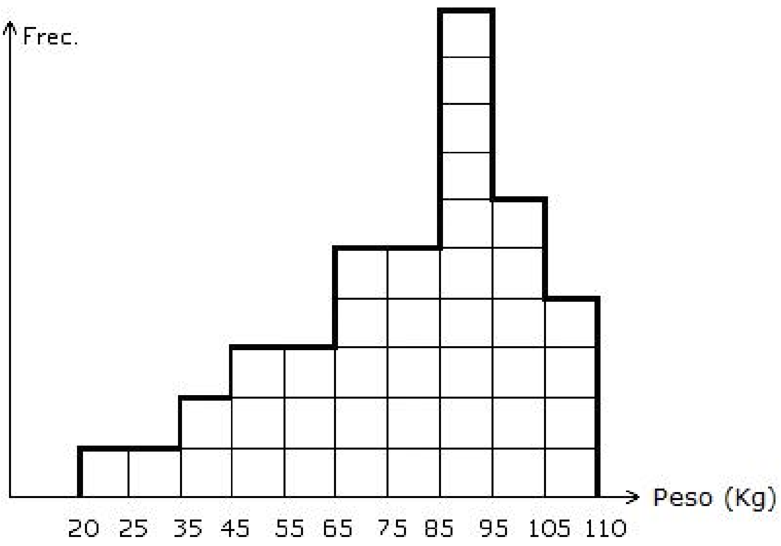
16. Explique qué miden los cuartiles.

17. Un conjunto de 40 datos contiene cinco valores mucho más grandes que todo el resto. Indique cuáles de estas medidas de centro estarán afectadas por los cinco valores y cuáles no:
Promedio, mediana, promedio recortado al 5%, promedio recortado al 20%

18. Alberto y Rubén son alumnos de cuarto medio, de un total de 200 alumnos del colegio. Dieron una prueba especial, y Alberto estuvo en el percentil 76, mientras que Rubén en el percentil 82. ¿Cuál de los dos estuvo mejor, y aproximadamente cuántos alumnos obtuvieron puntajes mejores que los de Alberto?

19. Se miden los días de ausencia anuales de los 35 empleados de una empresa. Aparecen tres valores mucho mayores que el resto. Indique cuáles de estas medidas de dispersión estarán afectadas por los tres valores y cuáles no:
Rango, varianza, desviación intercuartil, desviación estándar, coeficiente de variación.

20. En el siguiente histograma, ubique, los cuartiles y la mediana. Indique, en forma aproximada, dónde estaría el promedio.



21. Ordene de menos a más con respecto a su sensibilidad a puntos extremos, las siguientes medidas de dispersión: La desviación estándar, el rango, la desviación intercuartil, la varianza, la desviación media.

22. Se miden los pesos de un grupo de pacientes, en kilos. ¿En qué unidad se expresan las siguientes medidas de dispersión?

El rango, la desviación media, la varianza, la desviación standard, la desviación intercuartil.

23. ¿Se mide el peso de la producción de una muestra de industrias, en pesos. En qué unidad se expresan las siguientes medidas de resumen?

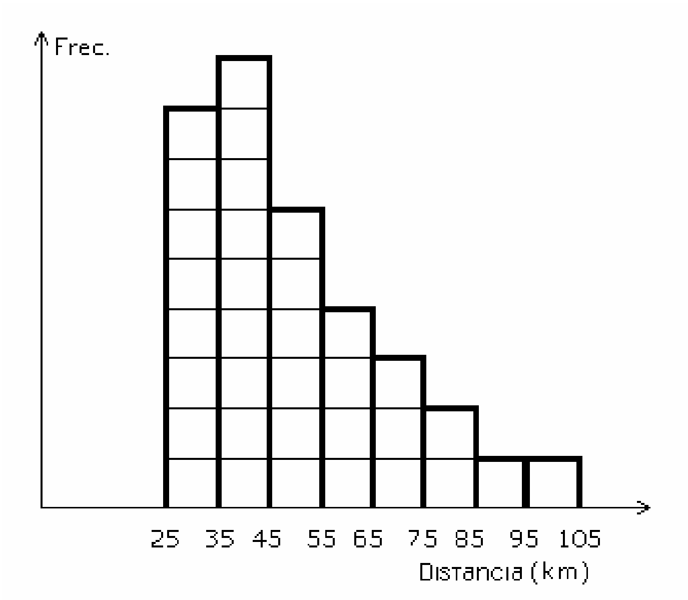
Promedio, mediana, desviación media, desviación estándar, promedio recortado al 15% desviación intercuartil, varianza.

24. Explique las siguientes afirmaciones:

El promedio de notas de Juan está sobre el percentil 90.

La desviación intercuartil de las edades de los consumidores de este producto es 25 años.

25. En el siguiente histograma, ubique, en forma aproximada, el promedio y la mediana.



26. Se hizo un curso de capacitación sobre uso de Excel a un grupo de 50 secretarias de una institución. Rosario y Consuelo estuvieron entre ellas. Rosario estuvo en el percentil 62, mientras que Consuelo estuvo en el percentil 76. ¿Cuál de las dos estuvo mejor, y aproximadamente cuántas secretarias obtuvieron puntajes entre los de Rosario y Consuelo?