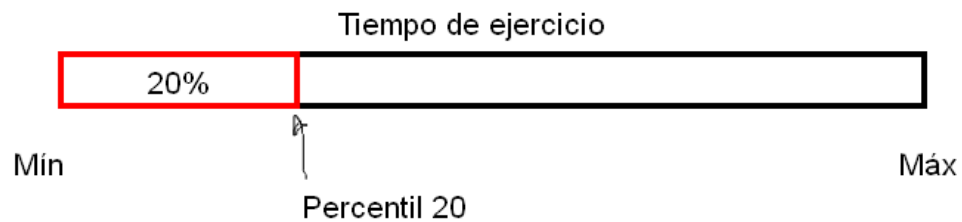


Medidas de posición

Mediana = Percentil 50 = Cuartil 2 = Decil 5

Ejemplo 11

¿Cuál es el tiempo máximo de ejercicio semanal que realiza un paciente joven (menor de 30 años) para estar dentro del 20% que menos ejercicio realiza?



Medidas de dispersión

Rango

3, 5, 4, 1, 8, 9 → rango = 9 - 1 = 8

3, 5, 4, 1, 8, 49 → rango = 49 - 1 = 48

sensible a valores extremos u outliers

3, 5, 4, 1, 8, 9 → rango = 8

3, 3, 3, 1, 3, 9 → rango = 8

el primer conjunto de datos es más heterogéneo, pero el rango **no evalúa la dispersión interna.**

Rango intercuartil



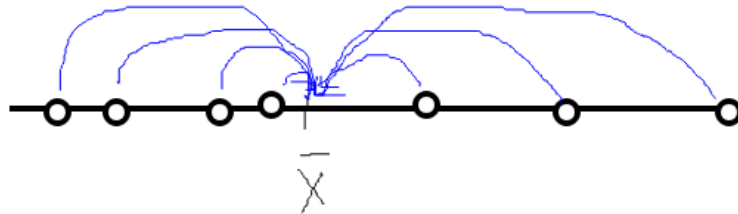
3, 5, 4, 1, 8, 9 → Promedio = 5

-2, 0, -1, -4, 3, 4 son los desvíos respecto a la media

La suma de desvíos es siempre 0

$$\sum_{i=1}^n x_i - \bar{x} = 0$$

Varianza



3, 5, 4, 1, 8, 9 → Promedio = 5

-2, 0, -1, -4, 3, 4 son los desvíos respecto a la media

$(-2)^2 + 0^2 + (-1)^2 + (-4)^2 + 3^2 + 4^2 = 4 + 0 + 1 + 16 + 9 + 16 = 46$

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

	Ingreso	Número de tardanzas
Desviación estándar	300 soles	3 tardanzas
Media	3000 soles	1 tardanza
Coeficiente de variación	10%	300%

El ingreso tiene mayor variabilidad que el número de tardanzas. Incorrecto porque las unidades son distintas (y además las magnitudes también)

$$CV = \frac{s}{|\bar{x}|} \times 100\%$$

En promedio, los ingresos se alejan en 300 soles de la media (3000 soles).

En promedio, las tardanzas se alejan en 3 tardanzas de la media (1)

Por ejemplo: 1, 0, 0, 0, 0, 7 tiene un valor extremo (7), lo que hace que se incremente la variación, y el cero también.

(medias cercanas a cero, el cv tiende a crecer)

```
> sd(c(1, 0, 0, 0, 0, 7))
[1] 2.804758
> mean(c(1, 0, 0, 0, 0, 7))
[1] 1.333333
```

Medias más altas tienden a tener mayor s.
Debe compararse con cv

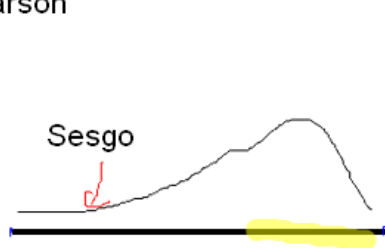
```
# A tibble: 3 x 4
  Grupo_Edad      s      m      cv
  <chr>      <dbl> <dbl> <dbl>
1 Adulto      89.6  167.  0.536
2 Adulto mayor 84.4  114.  0.740
3 Joven      91.8  156.  0.588
```

Rango	Unidades de la variable
Rango intercuartil	Unidades de la variable
Varianza	Unidades ² de la variable
Desviación estándar	Unidades de la variable
Coefficiente de variación	Sin unidades (%)

Todas las medidas de dispersión toman valores positivos

Medidas de asimetría

- En el 100% de los datos: coeficiente de asimetría de Fisher Pearson

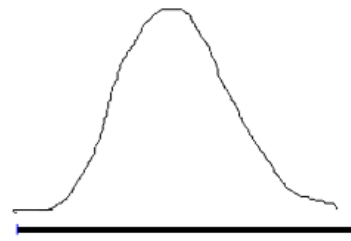


Distribución asimétrica
negativa o a la izquierda

$$as < 0$$

cuando es menor a -0.5

$$media < mediana$$

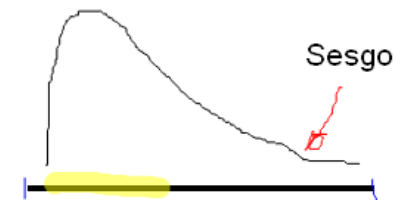


Distribución simétrica

$$as = 0$$

(-0.5 a 0.5) está lo
suficientemente **cercano a 0**

$$media = mediana$$



Distribución asimétrica
positiva o a la derecha

$$as > 0$$

cuando es > 0.5

$$media > mediana$$

Supongamos que medimos una variable en 3 grupos, y sus coeficiente de asimetría de Fisher Pearson son:

* $as1 = -0.12$

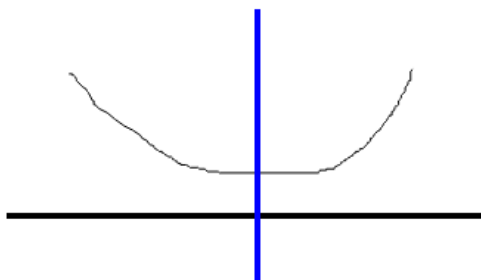
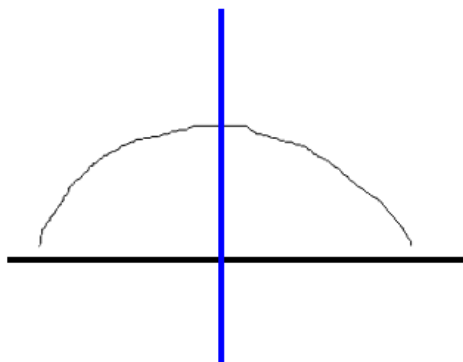
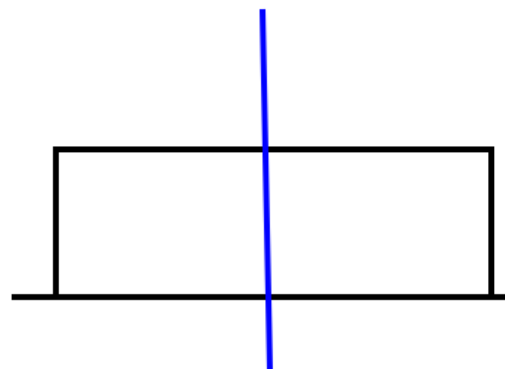
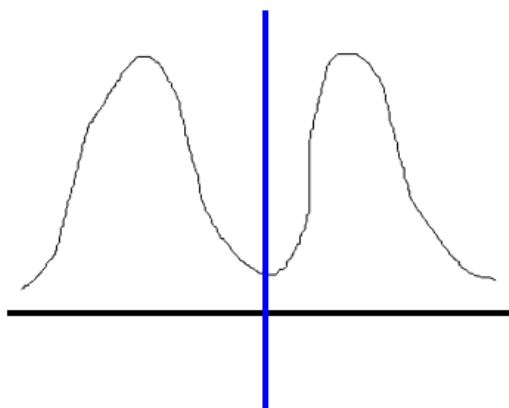
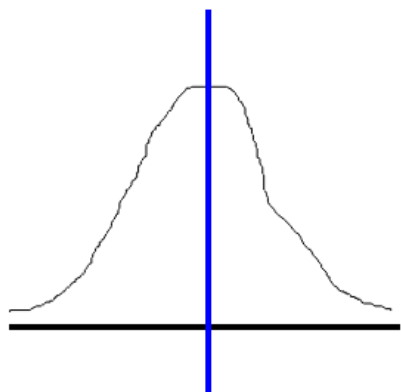
* $as2 = 0.94$

* $as3 = -1.35$

¿Cuál es el más simétrico? ¿Cuál es el más asimétrico?

El primer grupo presenta más simetría.

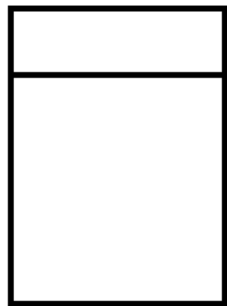
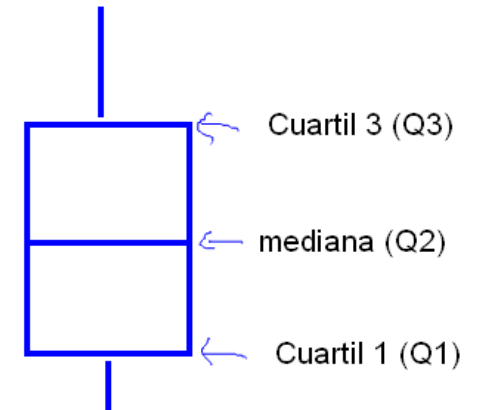
El tercer grupo presenta más asimetría



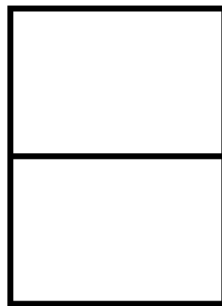
- En el 50% central de los datos: coeficiente de asimetría de Bowley

$$as_B = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

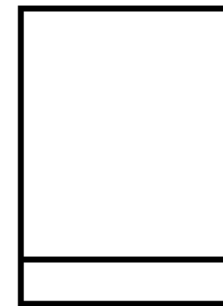
ric ($Q_3 - Q_1$) pequeño \rightarrow mayor asimetría en el 50% central de los datos



Distribución asimétrica
negativa o a la izquierda
en el 50% central de los
datos

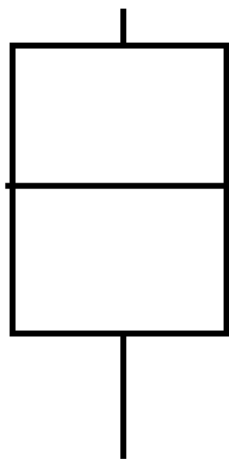


Distribución simétrica
en el 50% central de
los datos

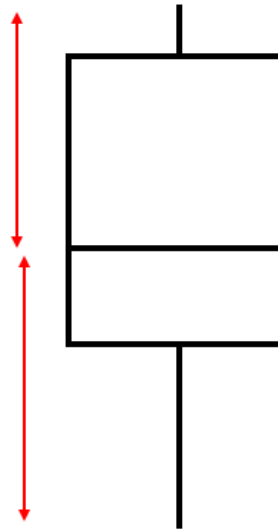


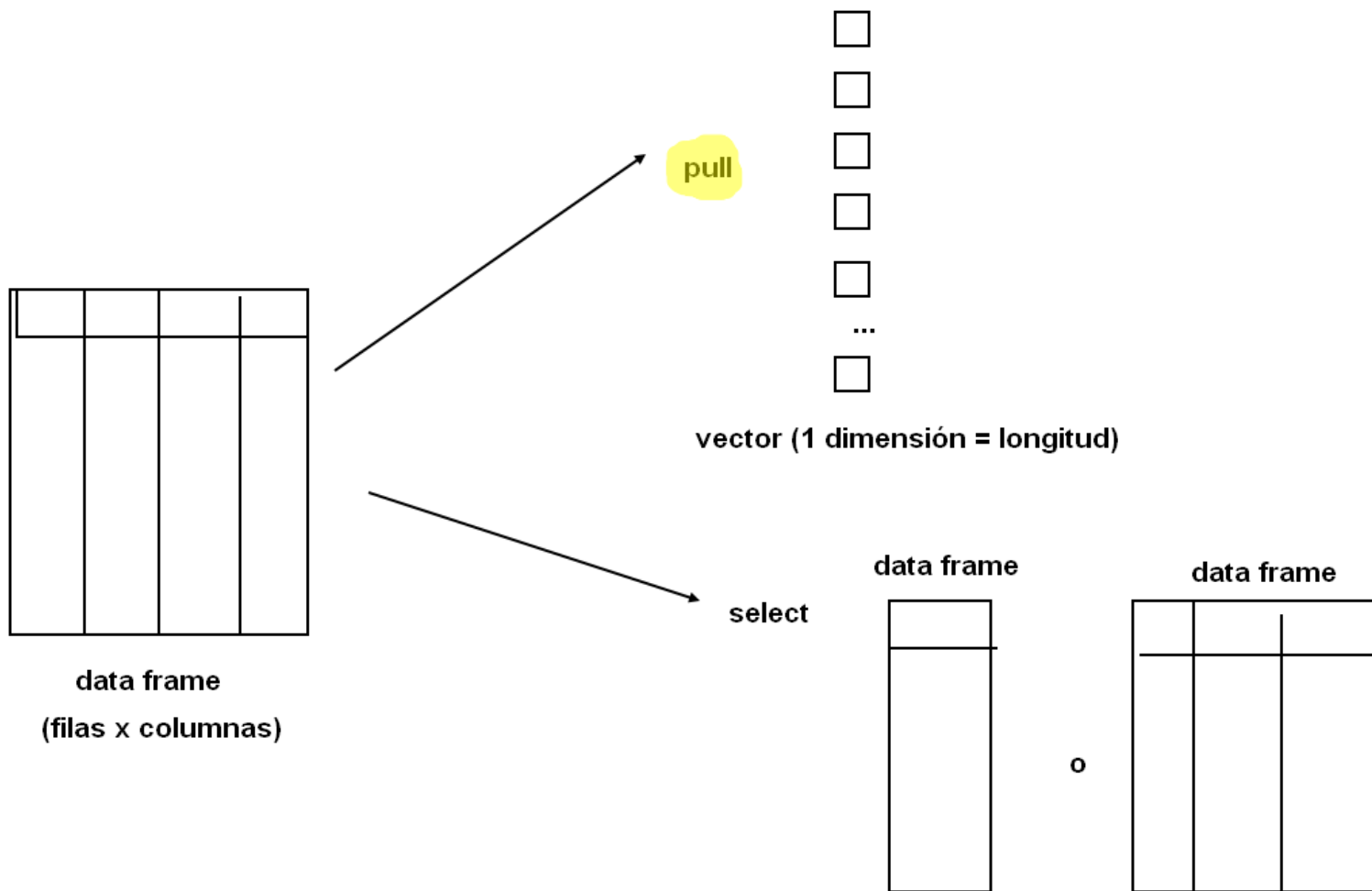
Distribución asimétrica
positiva o a la derecha
en el 50% central de los
datos

IMC



Edad





Tablas de frecuencia para variables cuantitativas continuas

```
> datos |>
+ pull(Edad) |>
+ DescTools::Freq()
```

	level	freq	perc	cumfreq	cumperc
1	[20,25]	8	8.0%	8	8.0%
2	(25,30]	10	10.0%	18	18.0%
3	(30,35]	6	6.0%	24	24.0%
4	(35,40]	14	14.0%	38	38.0%
5	(40,45]	13	13.0%	51	51.0%
6	(45,50]	7	7.0%	58	58.0%
7	(50,55]	11	11.0%	69	69.0%
8	(55,60]	11	11.0%	80	80.0%
9	(60,65]	13	13.0%	93	93.0%
10	(65,70]	7	7.0%	100	100.0%

- * 10 personas tienen una edad mayor a 25 años y como máximo 30.
- * El 14% de las personas tienen más de 35 años y a lo más 40.
- * 58 personas tienen una edad entre 20 y 50 años, inclusive.
- * El 24% de las personas tienen una edad mayor o igual a 20 y menor o igual a 35 años.

```
> datos |>
+ pull(Edad) |>
+ DescTools::Freq(breaks = c(20,30,40,50,60,70))
```

	level	freq	perc	cumfreq	cumperc
1	[20,30]	18	18.0%	18	18.0%
2	(30,40]	20	20.0%	38	38.0%
3	(40,50]	20	20.0%	58	58.0%
4	(50,60]	22	22.0%	80	80.0%
5	(60,70]	20	20.0%	100	100.0%

```
> datos |>
+ pull(Edad) |>
+ DescTools::Freq(breaks = c(20,30,40,50,60,70)) |>
+ data.frame() |>
+ rename(Edad = 1,
+         Cantidad = 2,
+         Porcentaje = 3,
+         Cant_Acum = 4,
+         Porc_Acum = 5)
```

	Edad	Cantidad	Porcentaje	Cant_Acum	Porc_Acum
1	[20,30]	18	0.18	18	0.18
2	(30,40]	20	0.20	38	0.38
3	(40,50]	20	0.20	58	0.58
4	(50,60]	22	0.22	80	0.80
5	(60,70]	20	0.20	100	1.00