

Material de estudio

OCW 2019: *Curso práctico para el análisis e inferencia estadística con Mathematica*

Tema 3. Tablas y representaciones gráficas

Equipo docente del curso

Arrospide Zabala, Eneko
Martín Yagüe, Luis
Unzueta Inchaurre, Aitziber
Soto Merino, Juan Carlos
Durana Apaolaza, Gaizka
Bikandi Irazabal, Iñaki

TEMA 3. TABLAS Y REPRESENTACIONES GRÁFICAS

Tablas de distribución de frecuencias

Definiciones

Considerando una muestra de k observaciones ordenadas de una variable, $\{x_i/i = 1, 2, \dots, k\}$:

- la frecuencia absoluta (n_i) del valor x_i es el número de veces que se repite el valor en el conjunto de las observaciones
- la frecuencia relativa (f_i) del valor x_i es el cociente entre la frecuencia absoluta y el número de observaciones de la muestra, k
- la frecuencia absoluta acumulada (N_i) hasta el valor x_i es el número de observaciones hasta la i inclusive; se obtiene sumando las frecuencias absolutas de los valores menores o iguales que x_i
- la frecuencia relativa acumulada (F_i) hasta el valor x_i es el cociente entre la frecuencia absoluta acumulada y el número de observaciones de la muestra, k

Elaboración

Una tabla de distribución de frecuencias presenta, por columnas, los valores que toma una variable en la muestra estudiada, junto con sus frecuencias (absoluta, relativa y sus acumuladas).

Funciones utilizadas en su construcción:

- **Tally[list]**. Devuelve una lista de listas en la que se muestran los elementos de una *lista* indicada como argumento junto con el número de veces que aparece cada uno
- **Sort[list]**. Ordena los elementos de una *lista* en forma canónica
- **Accumulate[list]**. Devuelve una lista con la suma acumulada de los elementos de una *lista* dada como argumento
- **All**. Especifica todas las partes de un determinado nivel (en una lista de listas, por ejemplo)

Ejemplo. Obtención de una tabla de frecuencias para la lista **serie**.

```

serie = {1, 3, 5, 2, 6, 7, 4, 5, 8, 3, 1, 9, 3,
        5, 4, 7, 2, 5, 9, 1, 4, 7, 5, 3, 8, 2, 0, 2, 0, 3, 6, 4, 9, 5, 2, 6};

datos = Tally[Sort[serie]]
{{0, 2}, {1, 3}, {2, 5}, {3, 5}, {4, 4}, {5, 6}, {6, 3}, {7, 3}, {8, 2}, {9, 3}}

n = Length[serie]
36

(* valores, xi: primeros elementos de todas las sublistas de datos *)
xi = datos[[All, 1]]
{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}

(* fr. absolutas, ni: segundos elementos de todas las sublistas de datos *)

```

```

ni = datos [ [All, 2] ]
{2, 3, 5, 5, 4, 6, 3, 3, 2, 3}

(* fr. absolutas acumuladas, Ni *)

Ni = Accumulate[ni]
{2, 5, 10, 15, 19, 25, 28, 31, 33, 36}

(* fr. relativas, fi *)

fi = ni / n
{ 1/18, 1/12, 5/36, 5/36, 1/9, 1/6, 1/12, 1/12, 1/18, 1/12 }

(* fr. relativas acumuladas, Fi *)

Fi = Accumulate[fi]
{ 1/18, 5/36, 5/18, 5/12, 19/36, 25/36, 7/9, 31/36, 11/12, 1 }

tabla = TableForm[
  {xi, ni, N[fi], Ni, N[Fi]},
  TableHeadings ->
    {"Valores: xi", "Frec. Absoluta", "Frec. Relativa", "Frec. Abs. Acumulada",
     "Frec. Relat. Acumulada"}, None}, TableDirections -> Row, TableAlignments -> Center]
Valores: xi Frec. Absoluta Frec. Relativa Frec. Abs. Acumulada Frec. Relat. Acumulada

```

0	2	0.0555556	2	0.0555556
1	3	0.0833333	5	0.138889
2	5	0.138889	10	0.277778
3	5	0.138889	15	0.416667
4	4	0.111111	19	0.527778
5	6	0.166667	25	0.694444
6	3	0.0833333	28	0.777778
7	3	0.0833333	31	0.861111
8	2	0.0555556	33	0.916667
9	3	0.0833333	36	1.

Representaciones gráficas

Introducción

Hay diferentes formas de representar gráficamente los valores que toma cada tipo de variable estadística:

- cuantitativa discreta: gráfico de barras, polígono de frecuencias, diagrama de tallo y hojas
- cuantitativa continua: histograma, polígono de frecuencias
- cualitativa: diagrama de barras, diagrama de sectores

Diagrama de sectores

- **PieChart[{y1,y2, ...}]**. Traza un diagrama de sectores cuyos ángulos son proporcionales a los valores $y_1, y_2, \dots$

Ejemplo. Obtención de un diagrama de sectores para una lista introducida por teclado.

```
lista = {a, b, a, b, d, e, c, b, d, c, d, b, a, c, e, d, c, c, b, c, d, c, b, a};
```

```
datos = Tally[Sort[lista]]
```

```
{{a, 4}, {b, 6}, {c, 7}, {d, 5}, {e, 2}}
```

```
valores = datos[[All, 1]]
```

```
{a, b, c, d, e}
```

```
freq = datos[[All, 2]]
```

```
{4, 6, 7, 5, 2}
```

```
PieChart[freq, PlotLabel -> HoldForm[Diagrama de Sectores],  
ChartLabels -> freq, ChartLegends -> {"a", "b", "c", "d", "e"}]
```

Diagrama de Sectores

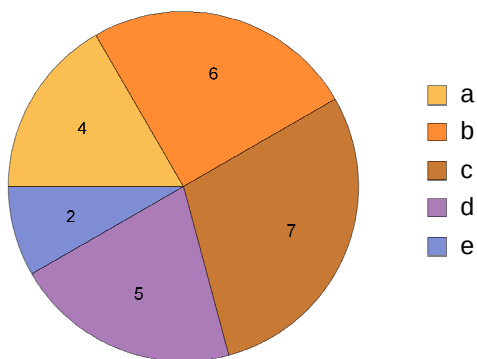


Diagrama de barras

- **BarChart[{y1,y2, ...}]**. Crea un diagrama de barras en el que las alturas de las barras toman los valores $y_1, y_2, \dots$ que se corresponden bien con las frecuencias absolutas bien con las frecuencias relativas de los valores de la variable.

Ejemplo. Obtención de un diagrama de barras para una lista introducida por teclado.

```
lista = {a, b, a, b, d, e, c, b, d, c, d, b, a, c, e, d, c, c, b, c, d, c, b, a};
```

```
datos = Tally[Sort[lista]]
```

```
{{a, 4}, {b, 6}, {c, 7}, {d, 5}, {e, 2}}
```

```
valores = datos[[All, 1]]
```

```
{a, b, c, d, e}
```

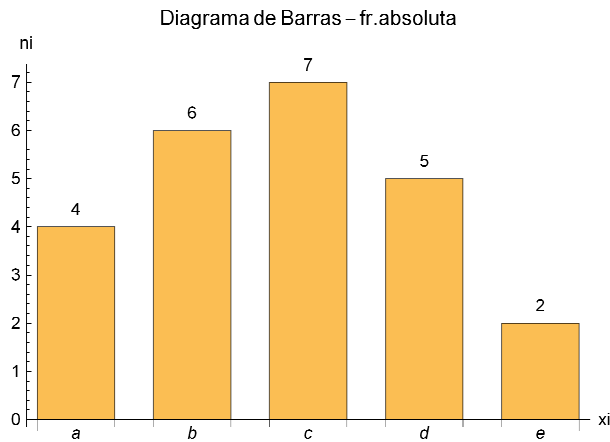
```
freq = datos[[All, 2]]
```

```
{4, 6, 7, 5, 2}
```

```

BarChart[freq, ChartLabels → valores, AxesLabel → {HoldForm[xi], HoldForm[ni]},
PlotLabel → HoldForm[Diagrama de Barras – fr. absoluta],
LabelingFunction → Above, BarSpacing → 0.5]

```



```

freqrel = N[datos[[All, 2]] / Length[lista], 4]
{0.1667, 0.2500, 0.2917, 0.2083, 0.08333}

```

```

BarChart[freqrel, ChartLabels → valores, AxesLabel → {HoldForm[xi], HoldForm[fi]},
PlotLabel → HoldForm[Diagrama de Barras – fr. relativa],
LabelingFunction → Above, BarSpacing → 0.5]

```

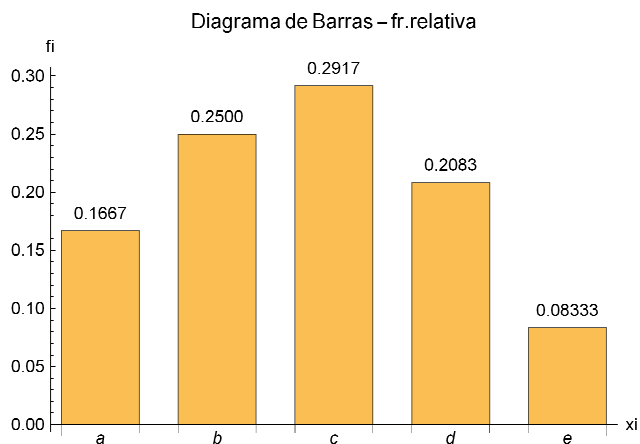


Diagrama de barras acumulado

- **BarChart[{y1,y2, ...}]**. Crea un diagrama de barras en el que las alturas de las barras toman los valores $y_1, y_2, \dots$. En este caso, se indica como argumento la lista de frecuencias acumuladas (absolutas o relativas)

Ejemplo. Obtención de un diagrama de barras para la lista del ejemplo anterior.

```
lista = {a, b, a, b, d, e, c, b, d, c, d, b, a, c, e, d, c, c, b, c, d, c, b, a};
```

```
datos = Tally[Sort[lista]];
```

```
valores = datos[[All, 1]];
```

```
freq = datos[[All, 2]];
```

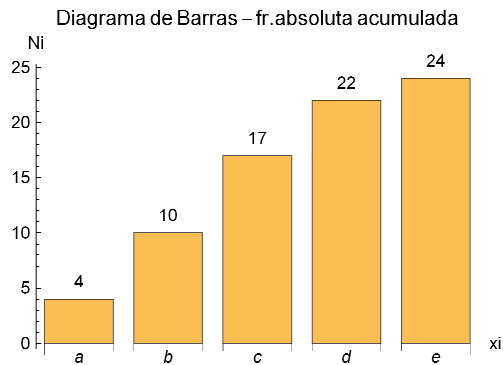
```
freqAc = Accumulate[freq]
```

```
{4, 10, 17, 22, 24}
```

```

BarChart[freqAc, ChartLabels → valores, AxesLabel → {HoldForm[xi], HoldForm[Ni]},
PlotLabel → HoldForm[Diagrama de Barras - fr. absoluta acumulada],
LabelingFunction → Above, BarSpacing → 0.3]

```



```
freqrel = N[datos[[All, 2]] / Length[lista], 4];
```

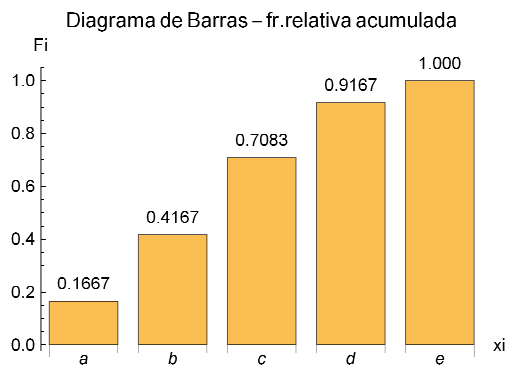
```
freqrelAc = Accumulate[freqrel]
```

```
{0.1667, 0.4167, 0.7083, 0.9167, 1.000}
```

```

BarChart[freqrelAc, ChartLabels → valores, AxesLabel → {HoldForm[xi], HoldForm[Fi]},
PlotLabel → HoldForm[Diagrama de Barras - fr. relativa acumulada],
LabelingFunction → Above, BarSpacing → 0.3]

```



■ **ListStepPlot**[{ $\{x_1, y_1\}, \{x_2, y_2\}, \dots\}$]. Representa los valores y_i en los intervalos $[x_i, x_{i+1}]$.

```
valoresnum = Join[{0}, valores /. {a → 1, b → 2, c → 3, d → 4, e → 5}];
```

```
freqAc = Join[{0}, Accumulate[freq]];
```

```

ListStepPlot[Transpose[{valoresnum, freqAc}], AxesLabel → {"xi", "Ni"}, PlotRange → {0, 28},
PlotLabel → "Gráfico de barras acumuladas", LabelingFunction → Above, Joined → True]

```

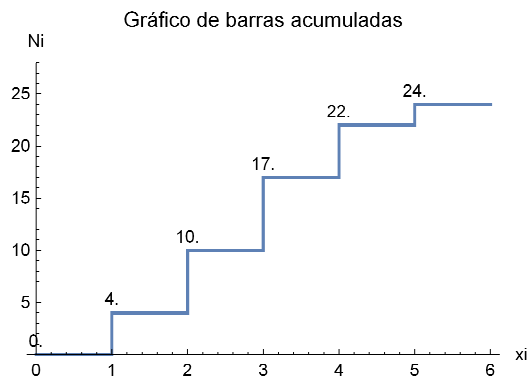


Diagrama de tallo y hojas

- **StemLeafPlot[list]**. Traza un diagrama de tallo y hojas con los datos reales de la lista dada como argumento. Requiere la carga previa del paquete "StatisticalPlots"

Ejemplo. Obtención de un diagrama de tallo y hojas para una lista introducida por teclado.

`Needs["StatisticalPlots`"]` (*carga del paquete StatisticalPlots*)

`pesos = {68, 71, 70, 65, 81, 80, 79, 77, 77, 74, 83, 68, 67, 76, 77, 81, 90, 69, 62, 76, 79, 85};`

`StemLeafPlot[pesos, StemExponent → 1]`

Stem	Leaves
6	257889
7	0146677799
8	01135
9	0

Stem units: 10

Histograma

- **Histogram[{x1,x2,...},bspec]**. Dibuja un histograma con los valores de la lista {x1,x2,...} y las especificaciones para la anchura de los intervalos indicadas en *bspec*.

Ejemplo. Obtención de un histograma para una lista generada de forma pseudoaleatoria.

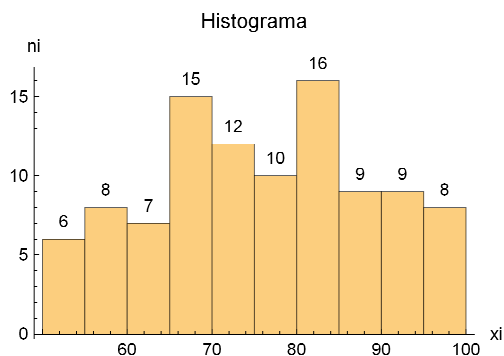
(* generación pseudoaleatoria de 100 números reales del intervalo [50,100] *)

`SeedRandom[0]`

`datos = RandomReal[{50, 100}, 100];`

(* histograma con diez intervalos de amplitud A=5 *)

`hist = Histogram[datos, {50, 100, 5}, AxesLabel → {HoldForm[xi], HoldForm[ni]},
PlotLabel → HoldForm[Histograma], LabelingFunction → Above]`



Histograma de frecuencias acumuladas

- **Histogram[{x1,x2,...},bspec]**. La función traza un histograma de frecuencias acumuladas sin más que añadir la opción "CumulativeCount".

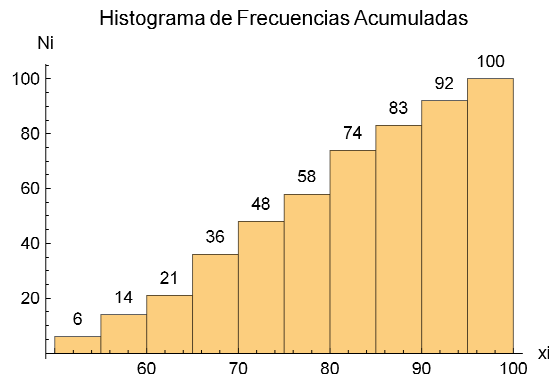
Ejemplo. Obtención de un histograma de frecuencias acumuladas para la lista **datos** del ejemplo anterior generada de forma pseudoaleatoria.

(* histograma con diez intervalos de amplitud A=5 *)

```

histAc = Histogram[datos, {50, 100, 5},
  "CumulativeCount", AxesLabel → {HoldForm[xi], HoldForm[Ni]},
  PlotLabel → HoldForm[Histograma de Frecuencias Acumuladas], LabelingFunction → Above]

```



- `ListStepPlot[{{x1,y1},{x2,y2}, ...}]`. Representa los valores y_i en los intervalos $[x_i, x_{i+1}]$.

```

escalón = Table[45 + 5 i, {i, 0, 10}];

```

```

frecAbs = Join[{0}, BinCounts[datos, {50, 100, 5}]] (* frecuencia de cada intervalo *)
{0, 6, 8, 7, 15, 12, 10, 16, 9, 9, 8}

```

```

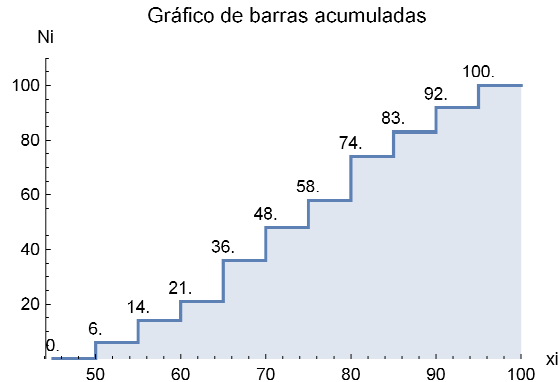
freqAc = Accumulate[frecAbs]
{0, 6, 14, 21, 36, 48, 58, 74, 83, 92, 100}

```

```

ListStepPlot[Transpose[{escalón, freqAc}], AxesLabel → {"xi", "Ni"}, PlotRange → {0, 110},
  PlotLabel → "Gráfico de barras acumuladas", LabelingFunction → Above, Filling → Axis]

```



Polígono de frecuencias absolutas

- `ListLinePlot[{{x1,y1},{x2,y2}, ...}]`. Traza una línea que une los puntos de la lista indicada como argumento.

Ejemplo. Obtención de un polígono de frecuencias absolutas para la lista **datos** del ejemplo anterior generada de forma pseudoaleatoria.

```

(* generación pseudoaleatoria de 100 números reales del intervalo [50,100] *)

```

```

SeedRandom[0]; datos = RandomReal[{50, 100}, 100];

```

```

(* diez intervalos de amplitud A=5 *)

```

```

bins = BinCounts[datos, {50, 100, 5}] (* frecuencia de cada intervalo *)

```

```

{6, 8, 7, 15, 12, 10, 16, 9, 9, 8}

```



```

limites = HistogramList[datos, 10][[1]] (* límites de los intervalos *)
{50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100}

intervalos = Partition[HistogramList[datos, 10][[1]], 2, 1] (* intervalos *)
{{50, 55}, {55, 60}, {60, 65}, {65, 70}, {70, 75}, {75, 80}, {80, 85}, {85, 90}, {90, 95}, {95, 100}}

mclase = Map[Mean, intervalos]; (* marcas de clase de cada intervalo *)

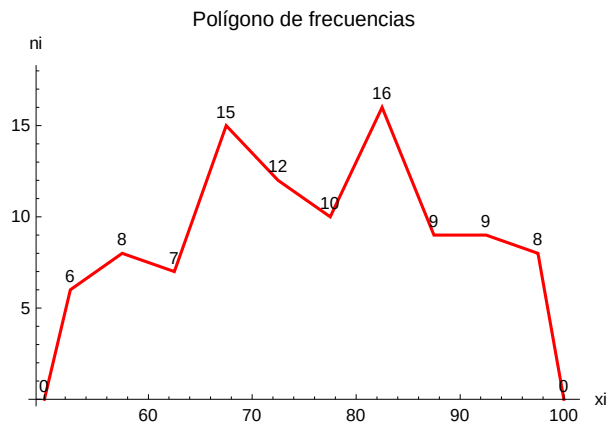
amp = 5;

marcas2 = Join[{mclase[[1]] - amp/2}, mclase, {mclase[[Length[mclase]]] + amp/2}]
{50,  $\frac{105}{2}$ ,  $\frac{115}{2}$ ,  $\frac{125}{2}$ ,  $\frac{135}{2}$ ,  $\frac{145}{2}$ ,  $\frac{155}{2}$ ,  $\frac{165}{2}$ ,  $\frac{175}{2}$ ,  $\frac{185}{2}$ ,  $\frac{195}{2}$ , 100}

ni = Join[{0}, bins, {0}]
{0, 6, 8, 7, 15, 12, 10, 16, 9, 9, 8, 0}
  
```

```

polfrec = ListLinePlot[Transpose[{marcas2, ni}], AxesLabel → {"xi", "ni"},
PlotLabel → "Polígono de frecuencias", LabelingFunction → Above, PlotStyle → Red]
  
```



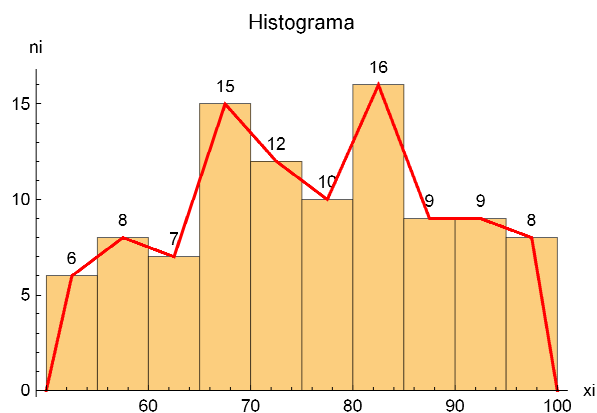
```

polfrecnolabel = ListLinePlot[Transpose[{marcas2, ni}], AxesLabel → {"xi", "ni"},
PlotLabel → "Polígono de frecuencias", LabelingFunction → None, PlotStyle → Red];
  
```

(* representación conjunta del histograma y el polígono de frecuencias absolutas *)

```

Show[hist, polfrecnolabel]
  
```



Polígono de frecuencias acumuladas

- `ListLinePlot[{x1,y1},{x2,y2}, ...]`. Traza una línea que une los puntos de la lista indicada como argumento.

Ejemplo. Obtención de un polígono de frecuencias acumuladas para la lista **datos** del ejemplo anterior generada de forma pseudoaleatoria.

(* generación pseudoaleatoria de 100 números reales del intervalo [50,100] *)

`SeedRandom[0]`

`datos = RandomReal[{50, 100}, 100];`

(* diez intervalos de amplitud A=5 *)

`bins = BinCounts[datos, {50, 100, 5}]` (* frecuencia de cada intervalo *)

`{6, 8, 7, 15, 12, 10, 16, 9, 9, 8}`

`limites = HistogramList[datos, 10][[1]]` (* límites de los intervalos *)

`{50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100}`

`intervalos = Partition[HistogramList[datos, 10][[1]], 2, 1]` (* intervalos *)

`{{50, 55}, {55, 60}, {60, 65}, {65, 70},`

`{70, 75}, {75, 80}, {80, 85}, {85, 90}, {90, 95}, {95, 100}}`

`mclase = Map[Mean, intervalos];` (* marcas de clase de cada intervalo *)

`amp = 5;`

`marcas2 = Join[{mclase[[1]] - amp/2}, mclase, {mclase[[Length[mclase]]] + amp/2}]`

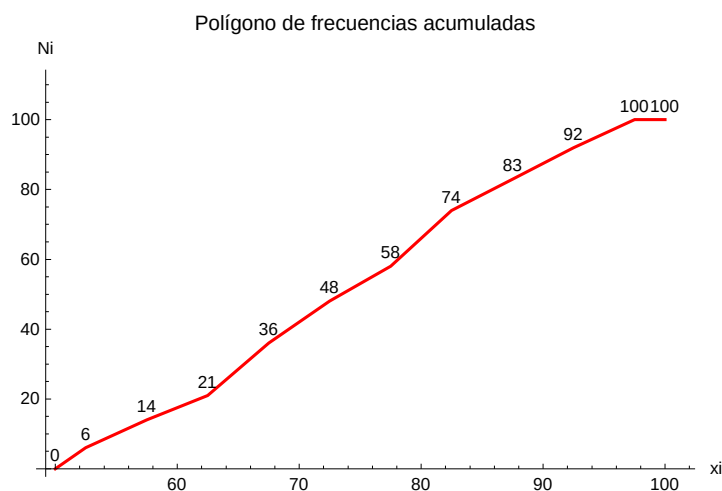
`{50,  $\frac{105}{2}$ ,  $\frac{115}{2}$ ,  $\frac{125}{2}$ ,  $\frac{135}{2}$ ,  $\frac{145}{2}$ ,  $\frac{155}{2}$ ,  $\frac{165}{2}$ ,  $\frac{175}{2}$ ,  $\frac{185}{2}$ ,  $\frac{195}{2}$ , 100}`

`Ni = Join[{0}, Accumulate[bins], {Accumulate[bins][[Length[bins]]}]`

`{0, 6, 14, 21, 36, 48, 58, 74, 83, 92, 100, 100}`

`polfreqAc = ListLinePlot[Transpose[{marcas2, Ni}], AxesLabel → {"xi", "Ni"},`

`PlotLabel → "Polígono de frecuencias acumuladas", LabelingFunction → Above, PlotStyle → Red]`



`polfreqAcnolabel = ListLinePlot[Transpose[{marcas2, Ni}], AxesLabel → {"xi", "Ni"},`

`PlotLabel → "Polígono de frecuencias acumuladas", LabelingFunction → None, PlotStyle → Red];`

(* representación conjunta del histograma y el polígono de frecuencias acumuladas *)

Show[histAc, polfrecAcnolabel]

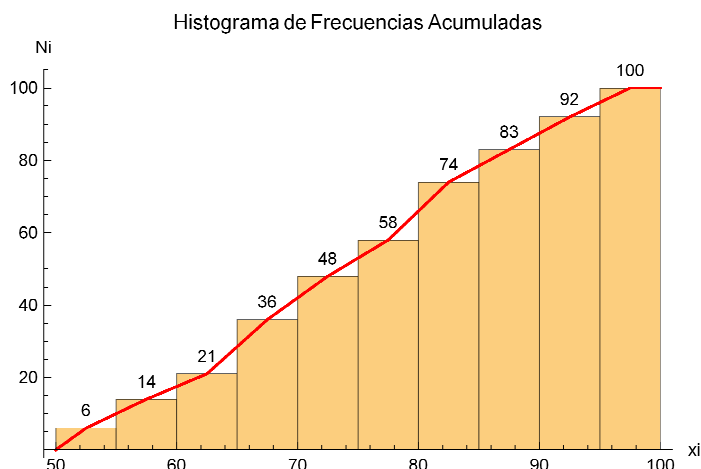


Diagrama de caja o *boxplot*

También, conocido como diagrama de caja y bigotes (de su nombre original, *box and whisker plot*).

Es un gráfico representativo de las distribuciones de un conjunto de datos. Se obtiene a partir de los cinco números resumen de una serie: mínimo (L), máximo (H) y cuartiles (Q_1 , Q_2 y Q_3).

- **BoxWhiskerChart**[*list*]. Traza un diagrama de caja para el conjunto de datos de la lista especificada como argumento.

Ejemplo. Obtención de un diagrama de caja para la lista **datos** del ejemplo anterior generada de forma pseudoaleatoria.

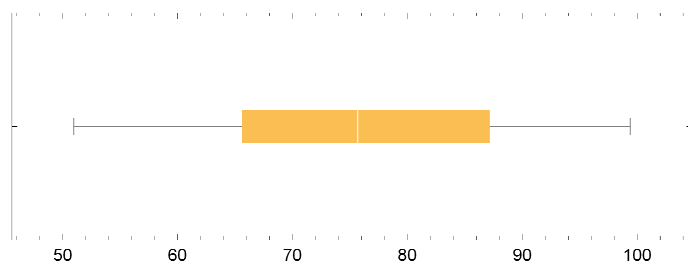
(* generación pseudoaleatoria de 100 números reales del intervalo [50,100] *)

SeedRandom[0]

datos = RandomReal[{50, 100}, 100];

(* boxplot sin datos atípicos*)

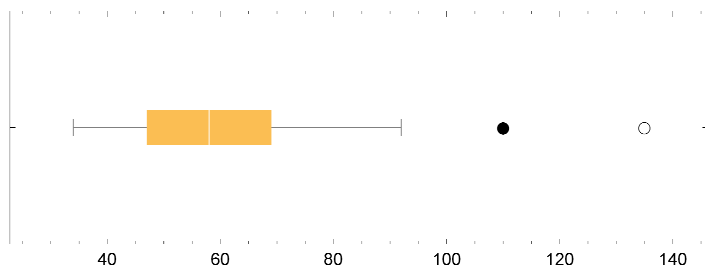
BoxWhiskerChart[datos, {"Outliers", {"Outliers", "●"}, {"FarOutliers", "○"}}, AspectRatio → 1 / 3, BarOrigin → Left]



Ejemplo. Trazado de un diagrama de caja para la lista **datos2** que contiene datos atípicos.

```
datos = {39, 78, 56, 45, 92, 70, 59, 64, 86, 43, 62, 74, 52, 82, 43, 68, 42, 73,
  41, 67, 45, 53, 87, 71, 55, 54, 135, 74, 48, 60, 65, 67, 47, 48, 86, 110, 45, 61,
  78, 49, 67, 44, 45, 69, 40, 56, 58, 58, 63, 79, 67, 58, 62, 91, 67, 44, 41, 69, 73,
  40, 72, 46, 68, 56, 46, 49, 63, 39, 54, 48, 43, 34, 42, 68, 70, 70, 58, 78, 69, 55,
  47, 50, 64, 50, 64, 51, 48, 46, 45, 60, 41, 40, 49, 58, 70, 64, 55, 68, 54, 82};
```

```
BoxWhiskerChart[datos, {"Outliers", {"Outliers", "●"}, {"FarOutliers", "○"}],
  AspectRatio → 1 / 3, BarOrigin → Left]
```



Ejemplo. Obtención de diagramas de caja para dos listas que representan valores de una variable recogidos en dos muestras diferentes.

(* generación pseudoaleatoria de los valores de la muestra 1 *)

```
SeedRandom[0]
```

```
altura1 = Join[RandomReal[{1.65, 1.88}, 100], {1.98}];
```

(* generación pseudoaleatoria de los valores de la muestra 2 *)

```
SeedRandom[1]
```

```
altura2 = Join[{1.59}, RandomReal[{1.69, 1.91}, 120]];
```

(* diagramas de caja *)

```
BoxWhiskerChart[{altura1, altura2},
  {"Outliers", {"Outliers", "●"}, {"FarOutliers", "○"}], AspectRatio → 3 / 5]
```

