Estadística Descriptiva

Observamos datos provenientes de una o varias muestras de la población bajo estudio.

El objetivo es obtener conclusiones sobre toda la población a partir de la muestra observada.

<u>Población</u>: todos los individuos que poseen la(s) caractrística(s) de interés.

<u>Muestra</u>: subconjunto de la población

Población **----------->** Muestra
<-----------
Inferencia

Factores necesarios para un buen análisis estadístico:

- Diseño del Experimento o Investigación
- Calidad de los Datos

Examinaremos los datos en forma descriptiva para:

- Organizar la información
- Sintetizar la información
- Ver sus características más relevantes
- Presentar la información

→ Métodos de Estadística de Inferencia*:*

- Estimación Puntual
- Estimación por Intervalos
- Tests de Hipótesis

Las técnicas del análisis exploratorio nos ayudan a organizar la información que nos dan los datos, de manera de detectar algún patrón de comportamiento así como también apartamientos importantes al modelo subyacente. Nos guían a la estructura subyacente en los datos de manera rápida y simple.

¿Qué observamos?

Cualquier característica de un individuo u objeto que nos resulte de interés y la expresaremos numéricamente

VARIABLE
Variable **----------->** Dato sujeto

Ejemplo:

Sexo:	Hombre	(0)
	Mujer	(1)
Edad:		
Peso		

Sexo es una variable categórica, mientras que **Edad** y **Peso** son cuantitativas.

Gráficos de Tallo y Hoja

Nos dan una primera aproximación rápida a la distribución de los datos sin perder de vista las observaciones.

1. Separamos a cada observación en dos partes: **tallo** y **hoja**

2. Listamos en forma vertical y creciente los tallos y agregamos las hojas a la derecha del tallo correspondiente.

```
 8 | 8 9
 9 | 3 4 4 5 6 6 7
10 | 3 3 4 6 7 8
11 | 1 2 2 3 3 7 7 8 9
12 | 0 0 4 5 5 5 7 7
13 | 2 4 5 6 8 9 9
14 | 2 3 8
15 | 5 5 6
```

Ejemplo: La siguiente tabla muestra los datos de la fuerza compresión de 45 muestras de aleación de Aluminio-Litio.

96	93	88	117	127	95	113	96
108	94	148	156	139	142	94	107
125	155	155	103	112	127	117	120
112	135	132	111	125	104	106	139
134	119	97	89	118	136	125	143
120	103	113	124	138			

Ordenamos los datos

88	89	93	94	94	95	96	96
97	103	103	104	106	107	108	111
112	112	113	113	117	117	118	119
120	120	124	125	125	125	127	127
132	134	135	136	138	139	139	142
143	148	155	155	156			

1. Elegimos un número de dígitos: 2 en este caso
2. Separamos los dígitos de los restantes, nos quedan 8 tallos de 8 a 15.
3. Obtenemos las hojas

Ejemplo. Consideremos el segundo dato:

8 9

TALLO HOJA

4. Elegimos cuantos dígitos a la derecha de cada número que corresponderán a las hojas: 1 en este caso.
5. Separamos esos dígitos, que constituirán los tallos. En este caso obtendremos 8 tallos, de 8 a 15.

8	8 9
9	3 4 4 5 6 6 7
10	3 3 4 6 7 8
11	1 2 2 3 3 7 7 8 9
12	0 0 4 5 5 5 7 7
13	2 4 5 6 8 9 9
14	2 3 8
15	5 5 6

¿Qué podemos ver en este diagrama?

- Forma de la distribución: simetría, asimetría a derecha, asimetría a izquierda
- Posición del centro de la distribución y concentración de los datos
- Desviaciones marcadas respecto al comportamiento general: outliers

Ejemplo:
Los siguientes datos corresponden a tiempos de falla de de cables Kevlar 49/epoxy sometidos a una presión del 90%:

TIEMPOS DE FALLA
0.01 0.01 0.02 0.02 0.02 0.03 0.03 0.04 0.05 0.06 0.07 0.07 0.08 0.09 0.09 0.10 0.10 0.11 0.11 0.12 0.13 0.18 0.19 0.20 0.23 0.80 0.80 0.83 0.85 0.90 0.92 0.95 0.99 1.00 1.01 1.02 1.03 1.05 1.10 1.10 1.11 1.15 1.18 1.20 1.29 1.31 1.33 1.34 1.40 1.43 1.45 1.50 1.51 1.52 1.53 1.54 1.54 1.55 1.58 1.60 1.63 1.64 1.80 1.80 1.81 2.02 2.05 2.14 2.17 2.33 3.03 3.03 3.24 4.20 4.69 7.89

El esquema de tallo y hoja que nos da el SX resulta:

```
STEM AND LEAF PLOT OF PER90

   LEAF DIGIT UNIT = 0.1           MINIMUM  0.0100
   7  8  REPRESENTS 7.8            MEDIAN   1.0750
                                   MAXIMUM  7.8900
          STEM  LEAVES
      25    0  0000000000000001111111122
      33    0  88889999
     (18)   1  000001111122333444
      25    1  55555555666888
      11    2  00113
       6    2
       6    3  002
       3    3
       3    4  2
       2    4  6
       1    5
       1    5
       1    6
       1    6
       1    7
       1    7  8
```

En este caso cada tallo ha sido dividido en 2 líneas.

* 0, 1, 2, 3, 4
• 5, 6, 7, 8, 9

Se observa asimetría a derecha y un valor alejado del resto: 7.8

¿Qué significan los números en la columna de la izquierda?
Es la Profundidad

A cada dato le podemos asignar un valor de ranking o rango contando desde cada extremo de la muestra ordenada. La **profundidad** es el menor de los dos valores.

En el *stem and leaf plot* el número en la columna de la izquierda es la mayor profundidad de la línea, excepto en aquella en la que el número está entre paréntesis, pues en ese caso el número que figura es la cantidad de hojas que hay en dicha línea.

Ejemplo:

Concentración de Inmunoglobulina en 298 niños sanos entre 6 meses y 6 años de edad.
Veamos el esquema de tallo y hoja que construye el SX.

```
0 | 111
0 | 22222223333333333333333333
0 | 4444444444444444444444444445555555555555555555555555*
0 | 6666666666666666666666666666666666677777777777777777*
0 | 8888888888888888888888888888888888888899999999999999*
1 | 00000000000000001111111111111111
1 | 2222223333333
1 | 444444444555555
1 | 66777
1 | 888
2 | 00011
2 | 2
2 | 5
2 | 7
2 |
3 |
3 |
3 |
3 |
3 |
4 |
4 |
4 |  5
```

En este caso cada tallo ha sido dividido en 5 líneas.

*	0 y 1
t	2 y 3
f	4 y 5
s	6 y 7
•	8 y 9

Hay reglas heurísticas para elegir el número de tallos. En general estos van entre 8 y 20.

Cuando el volumen de datos es muy grande conviene usar otro tipo de gráficos que también son de fácil interpretación.

Otra forma de representación de los datos es la de los histogramas.

Histogramas

- Dividimos el rango donde yacen los **n** datos en **intervalos o clases**, que no se superponen. Las clases deben ser **excluyentes** y **exhaustivas**.

- Contamos la cantidad de datos en cada intervalo o clase, es decir la **frecuencia**. También podemos usar para cada intervalo la **frecuencia relativa**

$$\boxed{fr_i = \frac{f_i}{n}}$$

- Graficamos el histograma en un par de ejes coordenados representando en las abscisas los intervalos y sobre cada uno de ellos un rectángulo cuya área es proporcional a la frecuencia relativa de dicho intervalo.

Obs.:

- No existen criterios óptimos para elegir la cantidad de intervalos. En general, entre 8 y 15 intervalos deberían ser suficientes. Muchos o muy pocos intervalos puede ser poco informativo.
  Se busca un equilibrio entre un histograma muy irregular y uno demasiado suavizado.

- No es necesario que todos los intervalos tengan la misma longitud, pero es recomendable que así sea. Esto facilita la lectura.
- El histograma representa la frecuencia o la frecuencia relativa a través del **área** y no a través de la altura.
- Es recomendable tomar

$$\textbf{Altura del rectángulo} = \frac{\textbf{frecuencia relativa}}{\textbf{Long. del intervalo}}$$

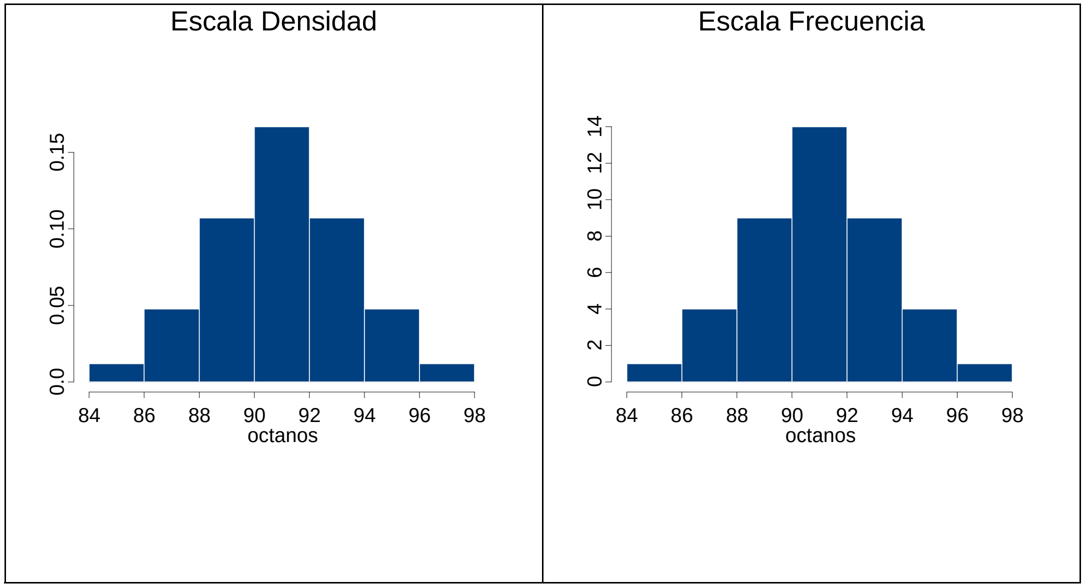
De esta manera el área es 1 y dos histogramas son fácilmente comparables independientemente de la cantidad de observaciones en las que se basa cada uno.

Ejemplo: Porcentajes de octanos para mezclas de naftas.

85.3	87.5	87.8	88.5	89.9	90.4	91.8	92.7
86.7	87.8	88.2	88.6	90.3	91.0	91.8	93.2
88.3	88.3	89.0	89.2	90.4	91.0	92.3	93.3
89.9	90.1	90.1	90.8	90.9	91.1	92.7	93.4
91.2	91.5	92.6	92.7	93.3	94.2	94.7	94.2
95.6	96.1						

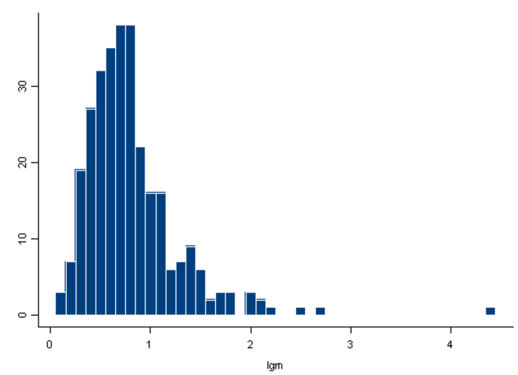
Clase	**Frecuencia f_i**	**Frecuencia relativa fr_i**
[84, 86]	1	0.02380952
(86, 88]	4	0.09523810
(88, 90]	9	0.21428571
(90,92]	14	0.33333333
(92,94]	9	0.21428571
(94,96]	4	0.09523810
(96,98]	1	0.02380952
Total	**42**	**1**

Histogramas para datos de OCTANOS

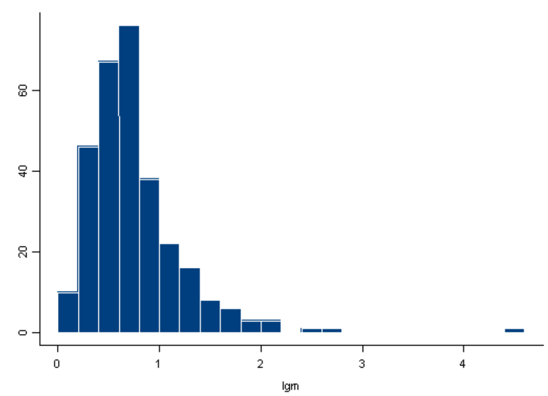


Ejemplo: Concentración de Inmunoglobulina

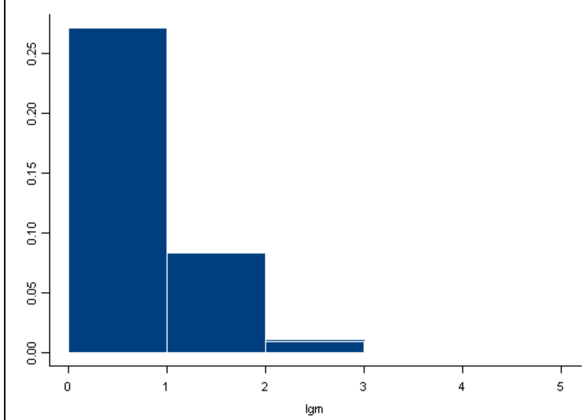
En este ejemplo vemos cómo la elección del ancho de las clases afecta el gráfico.



Longitud de Clase= 0.1 g/l



Longitud de Clase= 0.2 g/l



Longitud de Clase=1g/l

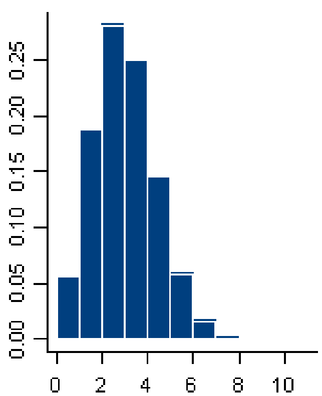
¿Qué podemos ver en un histograma?

- Rango de variación de los datos (Mínimo – Máximo)
- Intervalos o intervalos más frecuentes
- Simetría o Asimetría

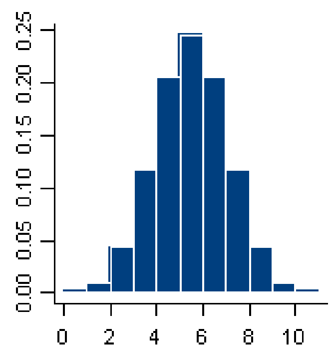
Asimetría

Un conjunto de datos que no se distribuye simétricamente, se llama **asimétrico**.
La asimetría puede verse en el esquema de Tallo y Hoja o en el Histograma.
También se puede apreciar a través de la posición relativa entre media y mediana.
En un boxplot lo haremos a través de la posición relativa entre la mediana y los cuartiles.
Un histograma tendería a tener las siguientes formas según cada caso:

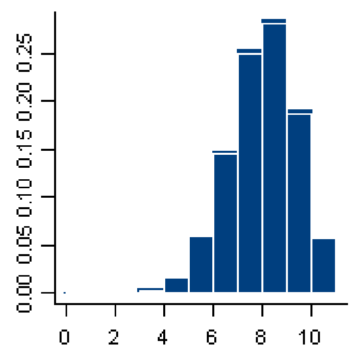
Asimétrica a derecha



Simétrica



Asimétrica a Izquierda



Medidas de Resumen

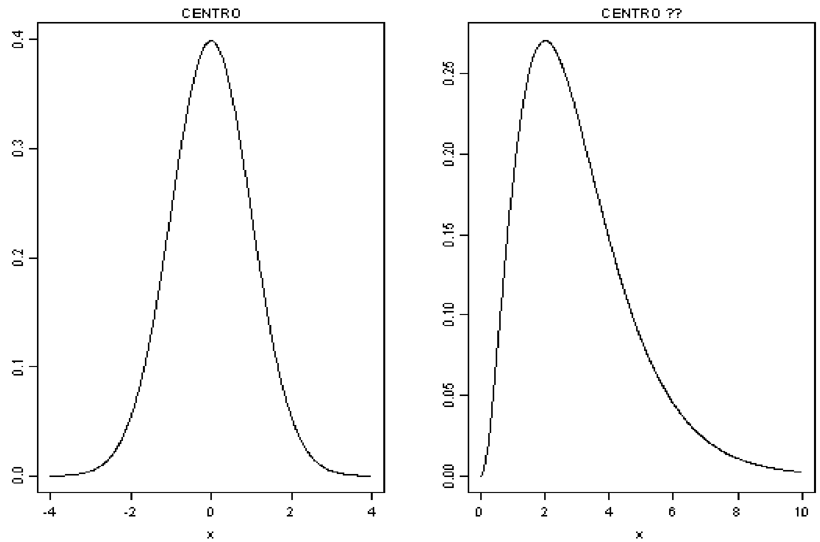
Resumiremos la información de los datos mediante medidas de fácil interpretación que reflejen sus características más relevantes.

La mediada a elegir dependerá de cada problema.

Medidas de Posición o Centrado

¿Cuál es el valor central o que mejor representa a los datos?
Buscamos un valor típico que represente a los datos.

Si la distribución es simétrica diferentes medidas darán resultados similares y hay un claro valor de centrado. Si es asimétrica no existe un centro evidente y diferentes criterios para resumir los datos pueden diferir considerablemente, en tanto tratan de captar diferentes aspectos de los mismos.



Promedio o Media Muestral

- Sumamos todas las observaciones y dividimos por el número total datos.

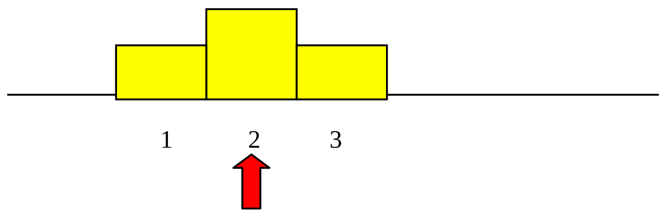
$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$ Promedio o Media Muestral

Ejemplo: Fuerza de compresión de muestras de Aleación de Aluminio-Litio

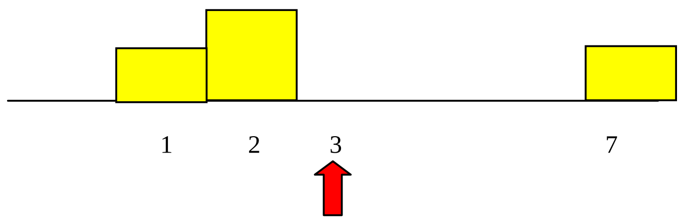
$$\bar{x} = \frac{\sum_{i=1}^{45} x_i}{45} = \frac{5350}{45} = 118.89$$

Es el punto de equilibrio del conjunto de datos.

X´s: 1, 2, 2, 3



X´s: 1, 2, 2, 7



Es una medida muy sensible a la presencia de datos anómalos (outliers).

Mediana Muestral

Es una medida del centro de los datos en tanto divide a la muestra ordenada en dos partes de igual tamaño. “Deja la mitad de los datos a cada lado”.
Sean los estadísticos de orden muestrales:

$$x_{(1)} \leq x_{(2)} \leq \ldots\ldots \leq x_{(n)}$$

definamos como mediana

$$\tilde{x} = \begin{cases} x_{(k+1)} & \text{si} \quad n = 2k+1 \\ \dfrac{x_{(k)} + x_{(k+1)}}{2} & \text{si} \quad n = 2k \end{cases}$$

Si la distribución es simétrica la mediana y la media identifican al mismo punto.

La mediana es resistente a la presencia de datos atípicos.

Si tenemos:

X´s: 1,2,2,3 $\bar{x} = 2$ $\tilde{x} = 2$

X´s: 1,2,2,7 $\bar{x} = 3$ $\tilde{x} = 2$

¿Qué pasa si tenemos un 70 en lugar de 7?

$$\bar{x} = 18.75 \qquad \tilde{x} = 2$$

Si tenemos una muestra de salarios de una población dada, ¿sería más adecuado tomar la media o la mediana muestral para representarlos?

Medias α-Podadas

Es un promedio calculado sobre los datos una vez que se han eliminado □ **%** de los datos más pequeños y **un** α **%** de los datos más grandes. Formalmente podemos definirla como:

$$\boxed{\bar{x}_{\alpha} = \frac{x_{([n\alpha]+1)} + \ldots + x_{(n-[n\alpha])}}{n - 2[n\alpha]}}$$

Ejemplo: Sea el siguiente conjunto de 10 observaciones, ya ordenadas

X's: 2 5 8 10 14 17 21 25 28 40

y calculemos la media 0.10-podada. Como el 10% de 10 es 1, debemos podar 1 dato en cada extremo y calcular el promedio de los 8 datos restantes, es decir

$$\bar{x}_{0.10} = \frac{5+8+10+14+17+21+25+28}{8} = \frac{128}{8} = 16$$

- La mediana puede ser vista como una 50% media podada.
- Es una medida intermedia entre la media y la mediana
- Es más resistente a datos atípicos que la media.

¿Cómo elegimos α?

Depende de cuantos outliers se pretende excluir y de cuán robusta queremos que sea la medida de posición.

Cuando seleccionamos $\alpha = 0$ tenemos la media, si elegimos el máximo valor posible para α (lo más cercano posible a 0.5) tenemos la mediana. Cualquier poda intermedia representa un compromiso entre ambas. Una elección bastante común es $\alpha = 0.10$, que excluye un 20% de los datos.

Ejemplo: calculamos las tres medidas de posición.

Los datos en la siguiente tabla corresponden al número de pulsaciones por minuto en pacientes con asma durante un espasmo:

Ordenamos los datos:

40 120 120 125 136 150 150 150 150 167

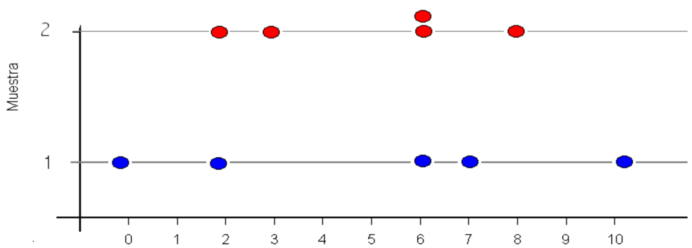
$\bar{x} = 130.8$ $\tilde{x} = 143$ $\bar{x}_{10} = 137.625$

Medidas de Dispersión o Variabilidad:

¿Cuán dispersos están los datos? ¿Cuán cercanos son los datos al valor típico?

Supongamos que tenemos datos $x_1, x_2, \ldots\ldots, x_n$

X's: 0 2 6 7 10
Y's: 2 3 6 6 8



$$\overline{X} = \overline{Y} = 5$$
$$\tilde{X} = \tilde{Y} = 6$$

¿Cómo medir la diferencia que se observa entre ambas muestras?

Rango Muestral

Se define como la diferencia entre el valor más grande y el pequeño de los datos:

$$\text{Rango} = \text{máx}(X_i) - \text{mín}(X_i)$$

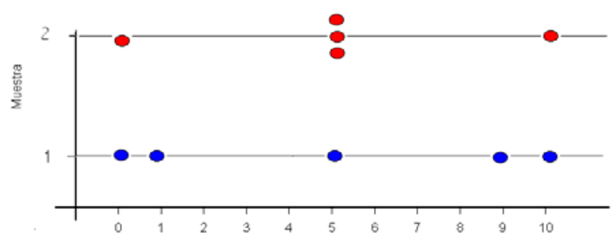
Ejemplo: en nuestros conjuntos de datos:

$$\text{Rango}(X) = 10 \qquad \text{Rango}(Y) = 6$$

- Esta medida es muy sensible a la presencia de outliers.

Veamos otro ejemplo:

X's: 0 1 5 9 10
Y's: 0 0 5 5 10



$$\overline{X} = \overline{Y} \qquad \tilde{X} = \tilde{Y} \qquad Rango(X) = Rango(Y)$$

Varianza Muestral

Es una medida de la variabilidad de los datos alrededor de la media muestral.

Varianza muestral : $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$

Desvío estándar muestral : $S = \sqrt{S^2}$

Ejemplo: en los dos ejemplos anteriores obtenemos

$S^2_x = 20.5 \quad S_x = 4.258$

$S^2_y = 12.5 \quad S_y = 3.536$

- El desvío estándar tiene las mismas unidades que los datos, mientras que la varianza no.

- Al basarse en promedios, estas medidas son sensibles a la presencia de datos atípicos. Por ejemplo, si en la muestra de los Y´s cambiamos el 10 por un 15 obtenemos $\mathbf{S^2_Y = 30}$ **y** $\mathbf{S_Y = 5.477}$, mientras que si lo cambiamos por un 20 obtenemos $\mathbf{S^2_Y = 57.5}$ **y** $\mathbf{S_Y = 7.583}$.

Distancia Intercuartil

Es una medida basada en el rango de los datos centrales de la muestra y más resistente que el desvío estándar.

Comenzaremos por definir los **percentiles.** El percentil $\alpha \cdot 100$ **%** de la muestra es el valor por debajo del cual se encuentra el $\alpha \cdot 100$ **%** de los datos en la muestra ordenada.

Para calcularlo:

- Ordenamos la muestra de menor a mayor

- Buscamos el dato que ocupa la posición $\alpha \cdot (n+1)$ en la muestra ordenada. Si este número no es entero se interpolan los dos adyacentes.

Ejemplo: Tenemos 19 datos que ordenados son

1 1 2 2 3 4 4 5 5 6 7 7 8 8 9 9 10 10 11

Percentil	Posición	Valor	
10%	0.10 (19+1) = **2**	1	
25%	0.25 (19+1) = **5**	3	**Cuartil Inferior**
50%	0.50 (19+1) = **10**	6	**Mediana**
75%	0.75(19+1) = **15**	9	**Cuartil Superior**
95%	0.95(19+1) = **19**	11	

Notemos que el percentil 50% (o segundo cuartil) coincide con la mediana. Llamaremos cuartil inferior (o primer cuartil) al percentil 25% y cuartil superior (o tercer cuartil) al percentil 75%.

Los cuartiles y la mediana dividen a la muestra ordenada en cuatro partes igualmente pobladas (aproximadamente un 25 % de los datos en cada una de ellas). Entre los cuartiles se hallan aproximadamente el 50% central de los datos y el rango de éstos es:

$$\boxed{\mathbf{d_I = distancia\ intercuartil = cuartil\ superior - cuartil\ inferior}}$$

Observación: Si en ejemplo cambiáramos el último dato por 110, la distancia intercuartil no cambiaría, mientras que el desvío pasaría de 3.2 a 24.13!!!!

Cuartos y Distancia entre Cuartos

Una medida muy cercana a los cuartiles inferior y superior son el cuarto inferior y el cuarto superior. Se calculan de la siguiente manera:

- Se ordena la muestra y se calcula la mediana de los datos.
- Dividimos a la muestra ordenada en dos partes: la primera corresponde a los datos más pequeños que la mediana y la segunda parte a la los datos más grandes que la mediana
- Si el tamaño de la muestra es par, el cuarto inferior es la mediana de la primera mitad, mientras que el cuarto superior es la mediana de la segunda mitad.
- Si el tamaño de la muestra es impar, a la primera y a la segunda parte se las expande agregándoseles a cada una de ellas la mediana de todos los datos. El cuarto inferior es la mediana de la primera parte expandida y el cuarto superior es la mediana de la segunda parte expandida. Es decir, en el caso impar, la mediana interviene en el cómputo de los dos cuartos.

Definimos la distancia entre cuartos como:

$$\boxed{d_C = \text{distancia entre cuartos} = \text{cuarto superior} - \text{cuarto inferior}}$$

Ejemplo: Consideremos las siguientes muestras ordenadas

Cuarto inferior=3

2 3 5 6 8 9

Cuarto superior=8

Cuarto inferior=4

2 3 5 6 7 8 9

Cuarto superior=7.5

Desvío Absoluto Mediano (Desviación absoluta respecto de la Mediana) MAD

Es una versión robusta del desvío estándar basada en la mediana. Definimos la MAD como:

$$MAD = \text{mediana}\left(\left| x_i - \tilde{x} \right| \right)$$

¿Cómo calculamos la MAD?

- Ordenamos los datos de menor a mayor.
- Calculamos la mediana.
- Calculamos la distancia de cada dato a la mediana.
- Despreciamos el signo de las distancias y las ordenamos de menor a mayor.
- Buscamos la mediana de las distancias sin signo.

Observación: Si deseamos comparar la distancia intercuartil y la MAD con el desvío standard es conveniente dividirlas por constantes adecuadas. En ese caso se compara a S mediante

$$\frac{MAD}{0.675} \qquad \frac{d_I}{1.35}$$

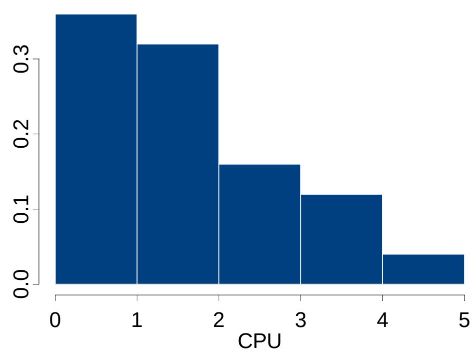
5 Números de Resumen

Los 5 números de resumen de la distribución de un conjunto de datos consisten en:
el **mínimo**, el **cuartil inferior**, la **mediana**, el **cuartil superior** y el **máximo**.

Ejemplo: Los siguientes datos corresponden a tiempos de CPU (en segundos) de 25 trabajos enviados a un server tomados al azar.

CPU
1.17 1.23 0.15 0.19 0.92 1.61 3.76 2.41 0.82
0.75 1.16 1.94 0.71 0.47 2.59 1.38 0.96 0.02
2.16 3.07 3.53 4.75 1.59 2.01 1.40

Calculamos los 5 números resumen y la media muestral para este conjunto de datos



stem(CPU)

N = 25 Mediana = 1.38
Quartiles = 0.82, 2.16

Decimal point is at the colon

```
0 : 01257789
1 : 022244669
2 : 0246
3 : 158
4 : 7
```

Boxplots

Con las medidas anteriores podemos construir un gráfico de fácil realización y lectura.

¿Cómo lo hacemos? Vamos a dar una versión, pero vale la pena advertir que hay variaciones de un programa a otro.

1. Representamos una escala vertical u horizontal

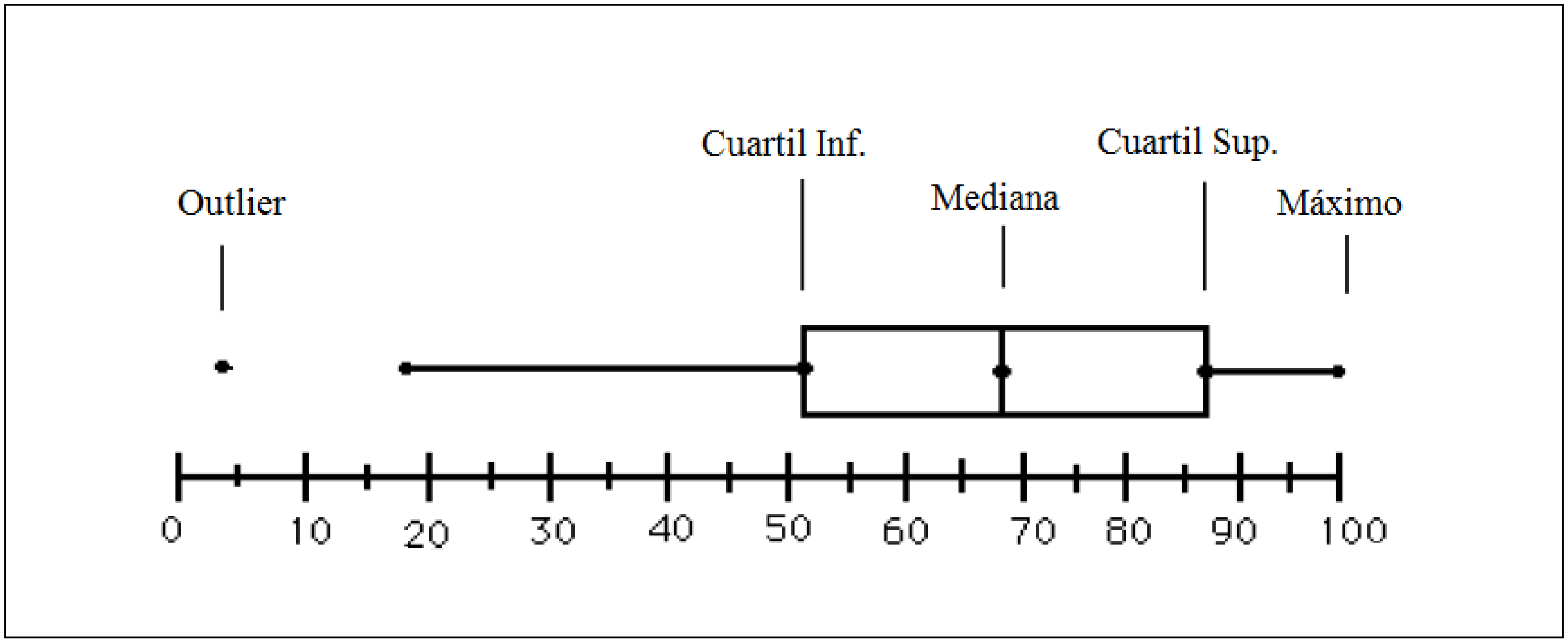
2. Dibujamos una caja cuyos extremos son los cuartiles y dentro de ella un segmento que corresponde a la mediana.

3. A partir de cada extremo dibujamos un segmento hasta el dato más alejado que está a lo sumo 1.5 $\mathbf{d_I}$ del extremo de la caja. Estos segmentos se llaman bigotes.

4. Marcamos con ***** a aquellos datos que están entre 1.5 $\mathbf{d_I}$ y 3 $\mathbf{d_I}$ de cada extremo y con **o** a aquellos que están a más de 3 $\mathbf{d_I}$ de cada extremo. Algunos paquetes indican a todos los outliers de la misma forma.

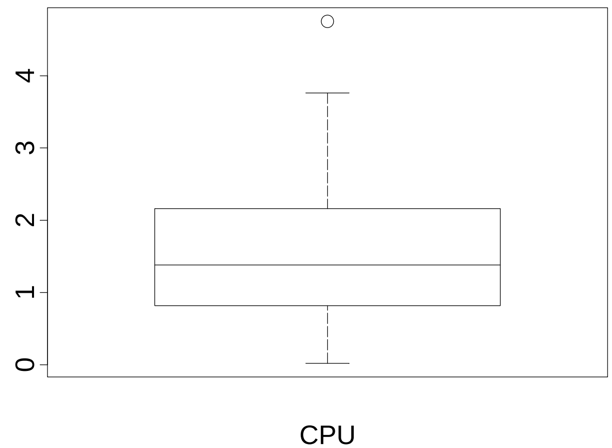
Observación: Muchos paquetes estadísticos realizan el boxplot usando los cuartos y la distancia entre cuartos en lugar de la distancia intercuartil. Como estas medidas son muy próximas, en general los resultados son análogos. Lo importante es que entre los cuartos o entre los cuartiles yace aproximadamente el 50% central de los datos.

ESQUEMA de BOXPLOT



Ejemplo:

El box-plot correspondiente a los tiempos de CPU es el siguiente



Es interesante observar que en el boxplot se indica a uno de los datos como outlier, mientras que en el análisis anterior esto no parecía evidente.

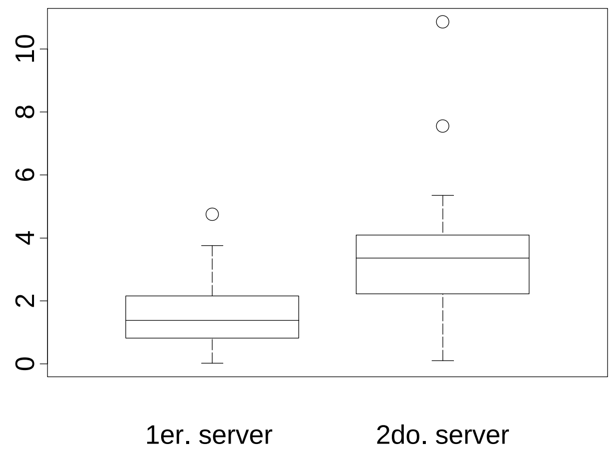
A partir de un box-plot podemos podemos apreciar los siguientes aspectos de la distribución de un conjunto de datos:

* posición
* dispersión
* asimetría
* longitud de las colas
* puntos anómalos o outliers.

Los box-plots son especialmente útiles para comparar varios conjuntos de datos, pues nos dan una rápida impresión visual de sus características.

Outliers: Los métodos que hemos visto nos permiten identificar puntos atípicos, que pueden aparecer en una o más variables. Su detección es importante pues pueden determinar o influenciar fuertemente los resultados de un análisis estadístico clásico, pues muchas de las técnicas habitualmente usadas son muy sensibles a la presencia de datos atípicos.

Ejemplo: Supongamos que se dispone de otros 25 datos correspondientes a tiempos de CPU enviados a otro server. Si realizamos boxplots paralelos para ambos conjuntos de datos obtenemos el siguiente gráfico.

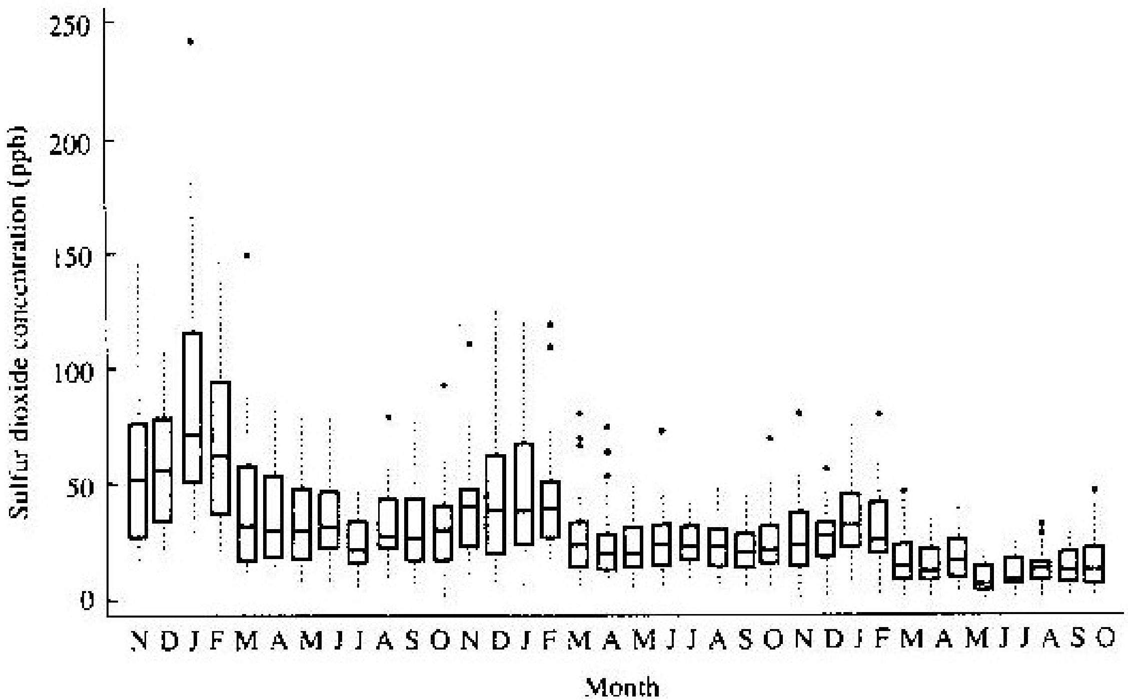


La simple comparación de los boxplots obtenidos revela que los trabajos enviados al segundo server son más largos. De hecho, el 75% de los trabajos muestreados en el segundo server tienen tiempos de CPU mayores que el cuartil superior de los trabajos muestreados en el primer server.

Ejemplo: Los siguientes boxplots corresponden a datos de concentración máxima diaria en partes por mil millones de dióxido de azufre en Bayonne, en el estado de Nueva Jersey, desde noviembre de 1969 hasta octubre de 1972 agrupados por meses. Hay 36 grupos de datos, cada uno de tamaño aproximadamente 30.

Los boxplots muestran algunas características de estos datos en forma muy rápida.

Hay una reducción general de la concentración de dióxido de azufre a lo largo del tiempo debida a la conversión gradual en la zona al uso de combustibles con baja concentración de azufre. Esta disminución es más fuerte para los cuartiles superiores. También se muestran concentraciones más elevadas para los meses de invierno debido al uso de calderas a petróleo. Claramente se ve un efecto cíclico y amortiguado. Los boxplots muestran una distribución asimétrica a derecha, con presencia de outliers en algunos meses, y que la dispersión de la distribución es mayor cuando el nivel general de la concentración es más alto.


250
200
150
100
50
0
Sulfur dioxide concentration (ppb)
N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O
Month

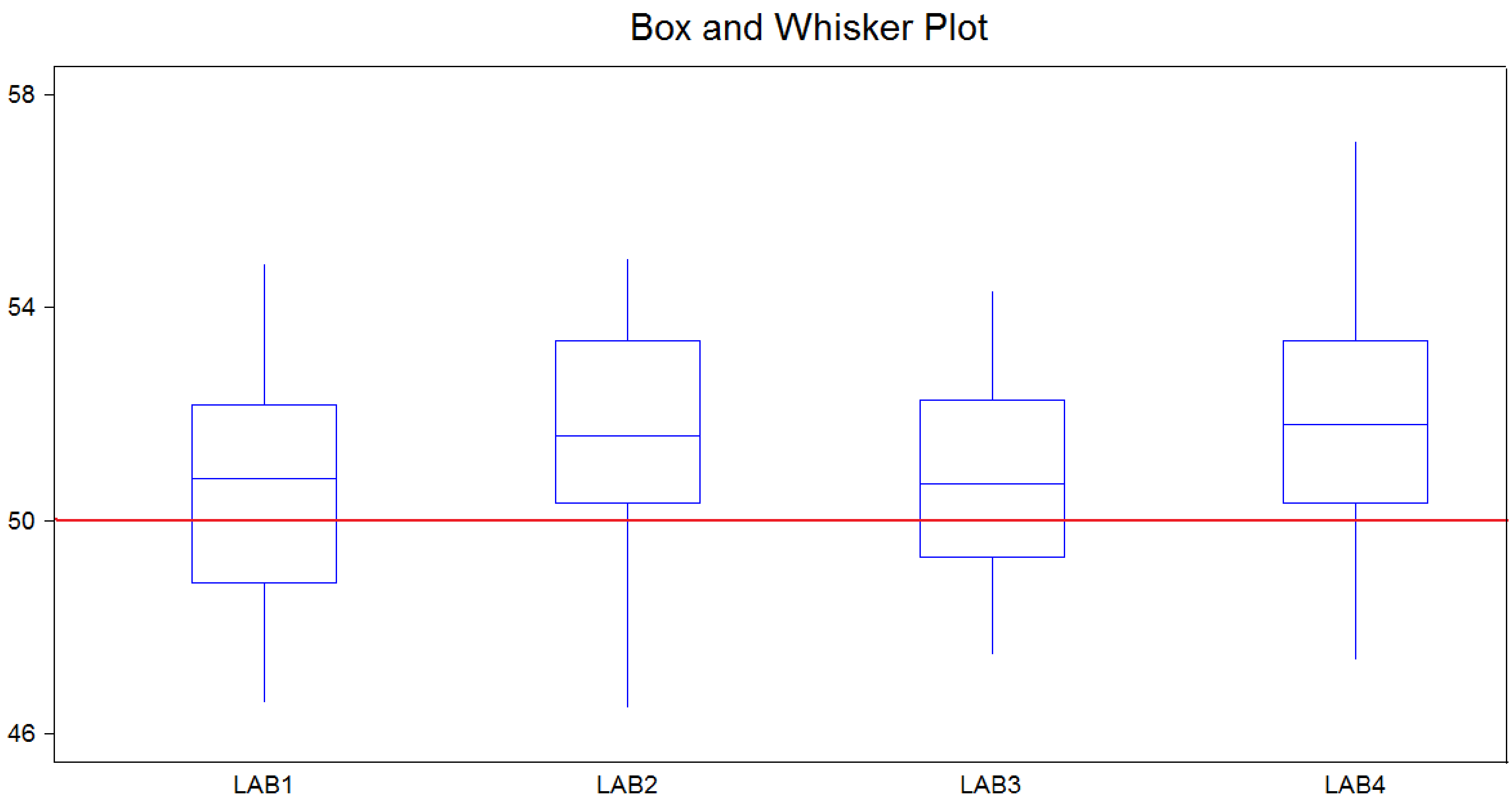
Ejemplo: Con el fin de estudiar las diferencias entre 4 laboratorios se miden 25 muestras con una concentración de analito de 50mg kg^{-1}.

Se analizan los datos correspondientes a 25 mediciones realizadas en 4 laboratorios. Veamos que da este análisis.

STATISTIX 7.0

DESCRIPTIVE STATISTICS

	LAB1	LAB2	LAB3	LAB4
N	25	25	25	25
MEAN	50.824	51.444	50.892	51.720
SD	2.5195	2.4316	2.0619	2.7217
MINIMUM	46.600	46.500	47.500	47.400
1ST QUARTI	48.700	49.850	49.300	49.850
MEDIAN	50.800	51.600	50.700	51.800
3RD QUARTI	52.600	53.600	52.500	53.650
MAXIMUM	54.800	54.900	54.300	57.100
MAD	2.0000	1.8000	1.6000	1.6000


Box and Whisker Plot
58
54
50
46
LAB1
LAB2
LAB3
LAB4