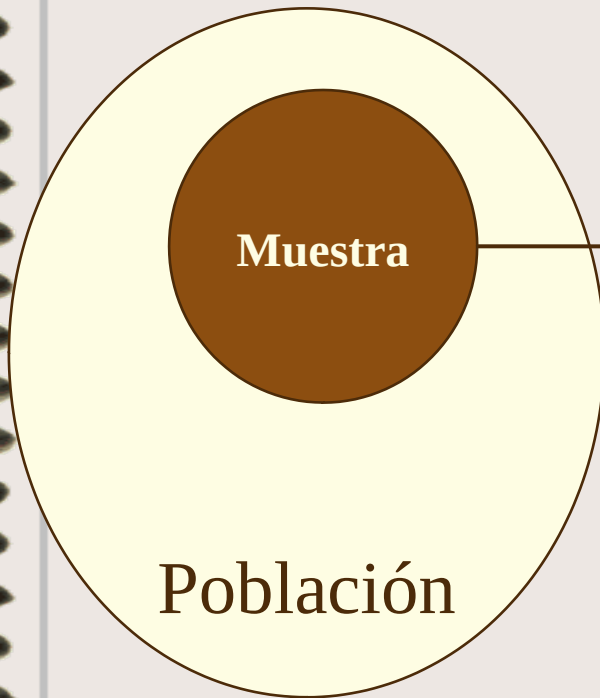


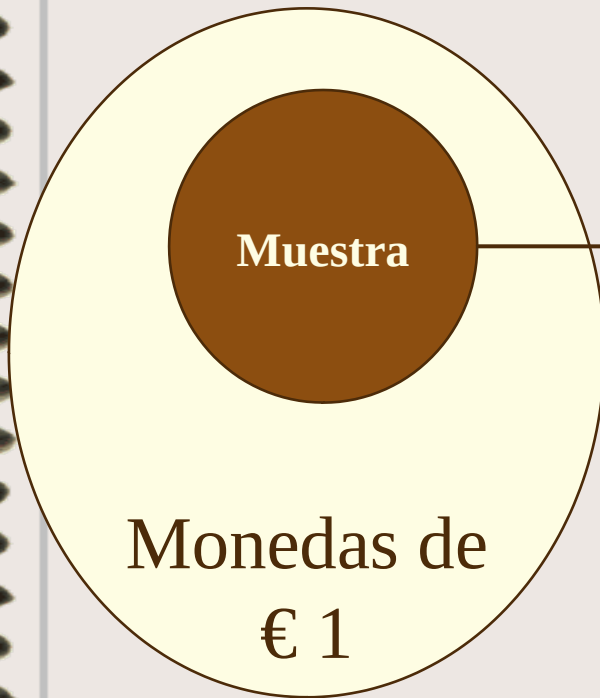
Estadística Descriptiva



Esperamos que sea *representativa*. Para ello, se toma al azar de entre toda la población.

$$\left\{ \mathbf{x}_i \right\}_{i=1}^n$$

Estadística Descriptiva

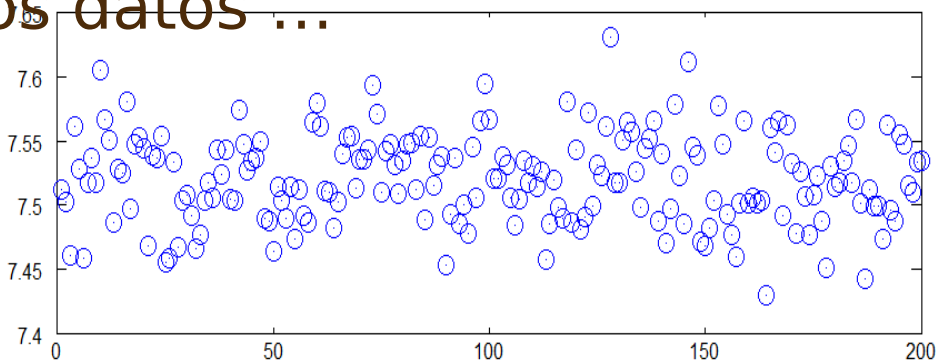


Se pesan 200 monedas...
¿Tomadas al azar?

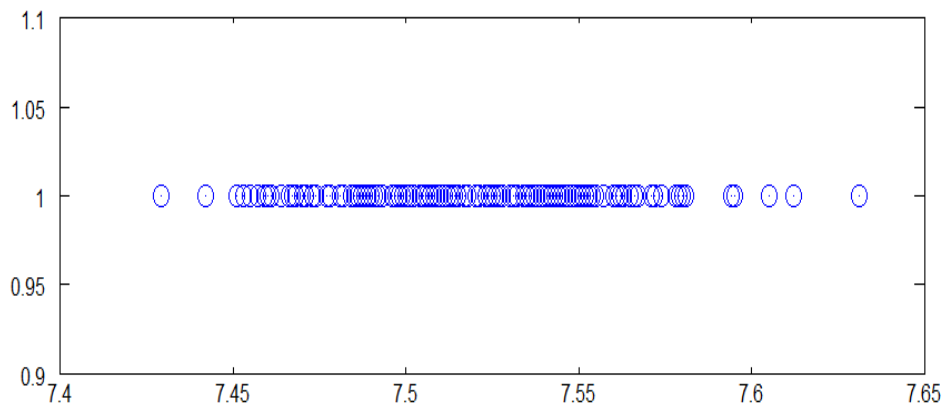
$$\left\{ \mathbf{x}_i \right\}_{i=1}^{200}$$

Estadística Descriptiva

Dos gráficos para dar un primer vistazo de los datos ...



*Muestra vs. Caso
o Scatterplot.*



*Constante vs.
Muestra.*

Estadística Descriptiva

Parámetros de tendencia central o posición

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

m

q₁ , q₃

Media muestral o promedio de los datos: informa sobre el comportamiento central, si todos los datos fuesen igual a la media entonces se conserva la suma. Si eso no se cumple entonces habrá datos mayores que la media y otros menores que ella.
Mediana: en un ordenamiento de menor a mayor es el término central, a diferencia de la media no es sensible a los valores alejados de la media.

Cuartiles primero y tercero: en un ordenamiento de menor a mayor son mediana de la primera mitad y de la segunda mitad de los datos.

Estadística Descriptiva

Parámetros de tendencia central o posición

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

**7.52
04**

m

**7.518
0**

q_1, q_3

**7.497
0
7.5445**

Estadística Descriptiva

Parámetros de variabilidad

$$R = M \acute{A} X - M \acute{I} N$$

$$IQR = q_3 - q_1$$

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

Rango muestral: la distancia entre el mínimo y el máximo de los datos muestrales (longitud del intervalo al que pertenece el 100% de los datos)

Rango intercuartil: la distancia entre los cuartiles primero y tercero de los datos muestrales (longitud del intervalo al que pertenece el 50% central de los datos)

Desvío estándar muestral: medida típica de variación de estos datos, es nula si y sólo si todos los datos son iguales y es tanto mayor cuanto más dispersos están los datos.

Estadística Descriptiva

Parámetros de variabilidad

$$R = \text{MÁX} - \text{MÍN}$$

0.202
00

$$IQR = q_3 - q_1$$

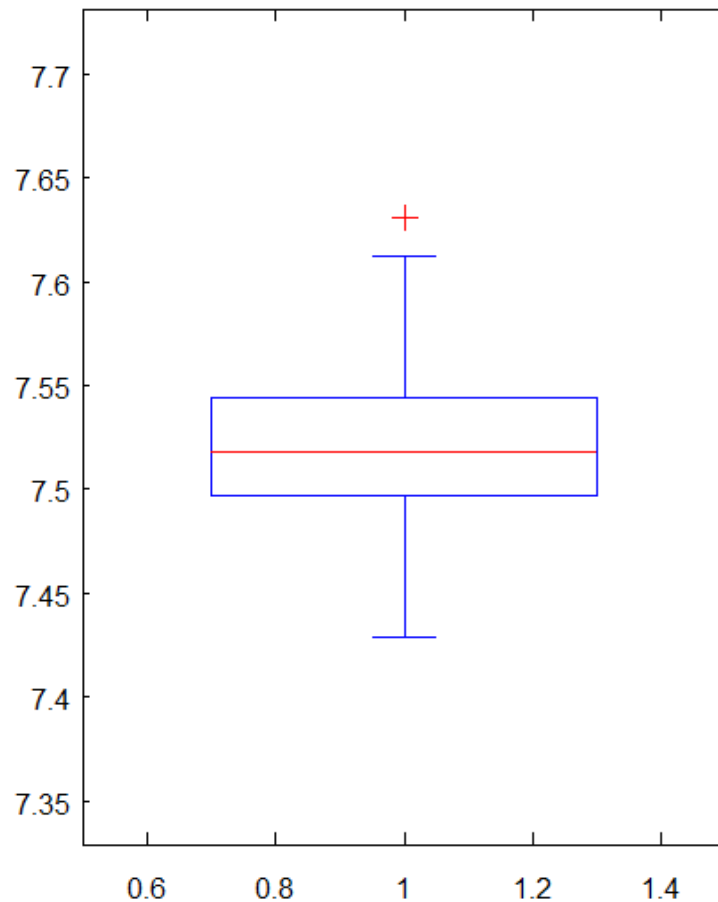
0.047
50

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

0.034
81

Estadística Descriptiva

Diagrama de *caja* o *boxplot*



- La caja central se determina con los cuartiles primero (Q_1), segundo (Q_2) y tercero (Q_3).
- Se representan con dos rayas (los bigotes) los datos inmediatos a los límites de Tuckey: $Q_1 - 1.5(IQR)$ y $Q_3 + 1.5(IQR)$.
- Se marcan como outliers moderados los datos fuera de esos límites.
- Se marcan como outliers severos los datos fuera de los límites: $Q_1 - 3(IQR)$ y $Q_3 + 3(IQR)$.

Estadística Descriptiva

Parámetros de forma

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

Coeficiente de simetría: medida típica de simetría de los datos respecto de la media. Es nulo si hay simetría. Si es positivo se dice que los datos tienen sesgo positivo y negativo en caso contrario.

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

Coeficiente de curtosis: mide la forma de la distribución de datos en torno del promedio. Es positivo si los datos tienen alta concentración en torno de la media y negativo en caso contrario. Es cercano a cero para el caso en que los datos provengan de una variable con distribución normal o Gaussiana.

Estadística Descriptiva

Parámetros de forma

$$\gamma = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3$$

0.131
74

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

-
0.1023
7

Estadística Descriptiva

Parámetros de forma - Corrección muestra pequeña

$$\tilde{\gamma} = \frac{\sqrt{n(n-1)}}{n-2} \gamma$$

La corrección evita un sesgo en la estimación (más sobre esto cuando hablemos de estimadores).

En el caso de la curtosis, la corrección sirve sólo para variables normales.

$$\tilde{\kappa} = \frac{(n+1)n^2}{(n-1)(n-2)(n-3)} (\kappa + 3) - \frac{(n-1)^2}{(n-2)(n-3)} 3$$

Estadística Descriptiva

Parámetros de forma - Corrección muestra pequeña

$$\tilde{\gamma} = \frac{\sqrt{n(n-1)}}{n-2} \gamma$$

0.132
74

$\tilde{\kappa}$

-
0.0444
2

Estadística Descriptiva

Frecuencias

Límite inferior	Límite superior					
7.405	7.435					
7.435	7.465					
7.465	7.495					
7.495	7.525					
7.525	7.555					
7.555	7.585					
7.585	7.615					
7.615	7.645					

Estadística Descriptiva

Frecuencias

Límite inferior	Límite superior	Marca	Frecuencia	Frec. Acum.	Frec. Rel.	Frec. Rel. Acum.
7.405	7.435	7.420	1	1	0.005	0.005
7.435	7.465	7.450	10	11	0.050	0.055
7.465	7.495	7.480	37	48	0.185	0.240
7.495	7.525	7.510	62	110	0.310	0.550
7.525	7.555	7.540	61	171	0.305	0.855
7.555	7.585	7.570	24	195	0.120	0.975
7.585	7.615	7.600	4	199	0.020	0.995
7.615	7.645	7.630	1	200	0.005	1.000
			200		1.000	

Estadística Descriptiva

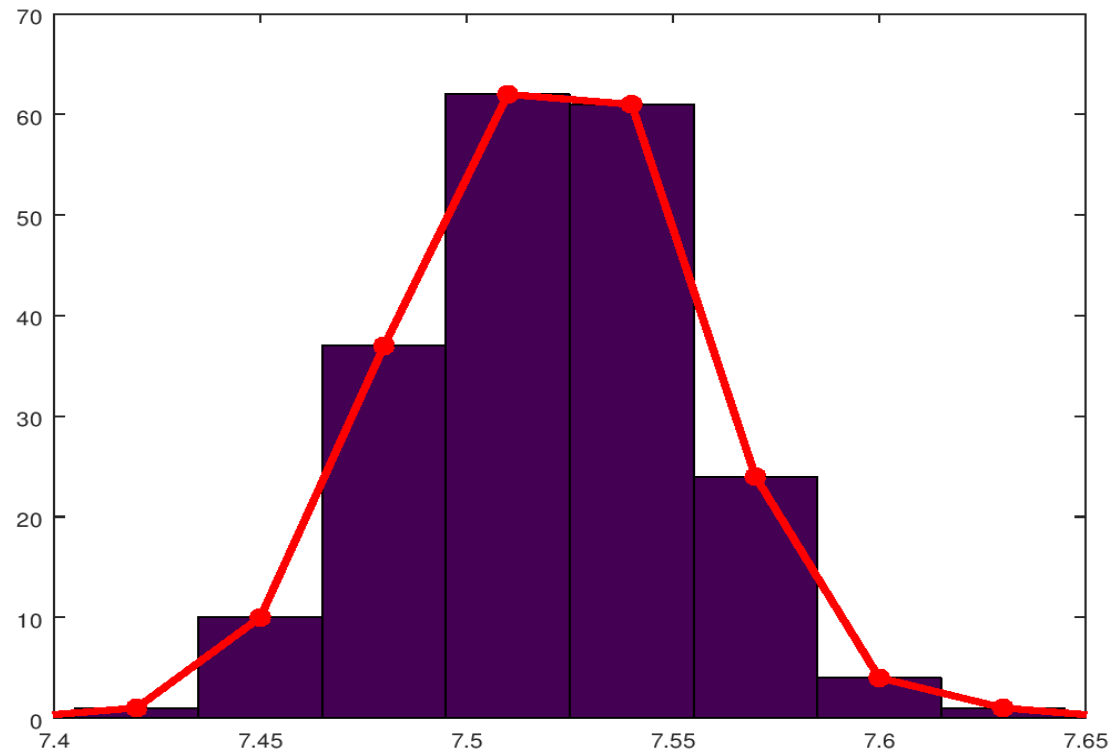
Histograma y Polígono de Frecuencias

Límite inferior	Límite superior	Marca	Frecuencia	Frec. Acum.	Frec. Rel.	Frec. Rel. Acum.
7.405	7.435	7.420	1	1	0.005	0.005
7.435	7.465	7.450	10	11	0.050	0.055
7.465	7.495	7.480	37	48	0.185	0.240
7.495	7.525	7.510	62	110	0.310	0.550
7.525	7.555	7.540	61	171	0.305	0.855
7.555	7.585	7.570	24	195	0.120	0.975
7.585	7.615	7.600	4	199	0.020	0.995
7.615	7.645	7.630	1	200	0.005	1.000
			200		1.000	

7.42	7.45	7.48	7.51	7.54	7.57	7.6	7.63
1	10	37	62	61	24	4	1

Estadística Descriptiva

Histograma y Polígono de Frecuencias



7.42	7.45	7.48	7.51	7.54	7.57	7.6	7.63
1	10	37	62	61	24	4	1

Estadística Descriptiva

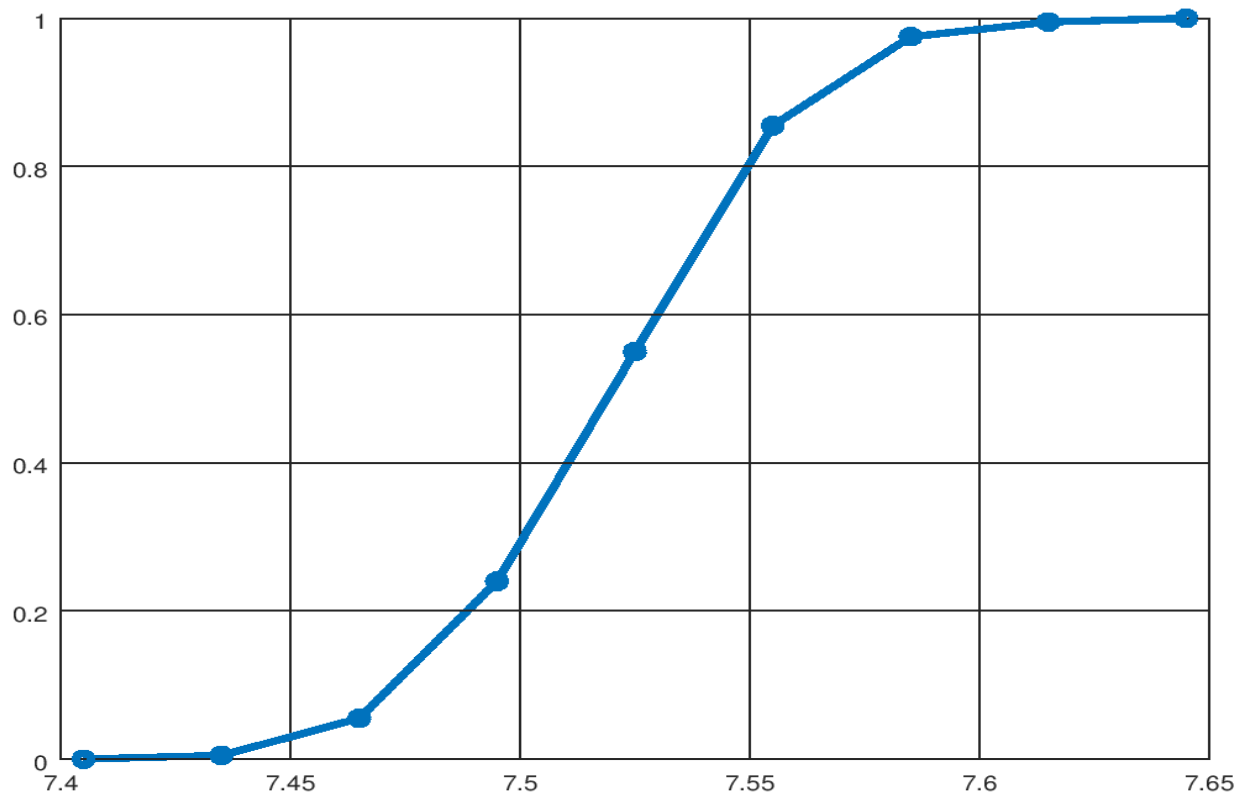
Polígono de Frecuencias Relativas Acumuladas

Límite inferior	Límite superior	Marca	Frecuencia	Frec. Acum.	Frec. Rel.	Frec. Rel. Acum.
7.405	7.435	7.420	1	1	0.005	0.005
7.435	7.465	7.450	10	11	0.050	0.055
7.465	7.495	7.480	37	48	0.185	0.240
7.495	7.525	7.510	62	110	0.310	0.550
7.525	7.555	7.540	61	171	0.305	0.855
7.555	7.585	7.570	24	195	0.120	0.975
7.585	7.615	7.600	4	199	0.020	0.995
7.615	7.645	7.630	1	200	0.005	1.000
			200		1.000	

7.405	7.435	7.465	7.495	7.525	7.555	7.585	7.615	7.645
0	0.005	0.055	0.240	0.550	0.855	0.975	0.995	1.000

Estadística Descriptiva

Polígono de Frecuencias Relativas Acumuladas



7.405	7.435	7.465	7.495	7.525	7.555	7.585	7.615	7.645
0	0.005	0.055	0.240	0.550	0.855	0.975	0.995	1.0

[illegible]

Estadística Descriptiva

Parámetros característicos con datos agrupados

$$\bar{x} = \frac{1}{n} \sum_{k=1}^m x_k f_k$$

$$\gamma = \frac{1}{n s^3} \sum_{k=1}^m (x_k - \bar{x})^3 f_k$$

$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^m (x_k - \bar{x})^2 f_k}$$

$$\kappa = \frac{1}{n s^4} \sum_{k=1}^m (x_k - \bar{x})^4 f_k - 3$$

Estadística Descriptiva

Parámetros característicos con datos agrupados

$$\bar{x} = \frac{1}{n} \sum_{k=1}^m x_k f_k$$

7.5198

$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^m (x_k - \bar{x})^2 f_k}$$

0.03519

Estadística Descriptiva

Parámetros característicos con datos agrupados

0.02201

$$\gamma = \frac{1}{ns^3} \sum_{k=1}^m (x_k - \bar{x})^3 f_k$$

-0.04880

$$\kappa = \frac{1}{ns^4} \sum_{k=1}^m (x_k - \bar{x})^4 f_k - 3$$

A vertical sequence of 20 frames showing a metal spring being compressed and then released. The frames are arranged vertically, with the top frame showing the spring in its fully extended state. As the frames progress downwards, the spring is gradually compressed, with the coils becoming more tightly packed. The bottom frame shows the spring in its fully compressed state. This sequence illustrates the concept of potential energy, as the work done to compress the spring is stored as potential energy, which can then be converted back into kinetic energy when the spring is released.

[illegible]

Estadística Descriptiva

Cuartiles con datos agrupados

Intervalos	Frecuencias	Frec. Relativas Acum. (%)

$$q_1 = \frac{25 - 24}{55 - 24} (7.525 - 7.495) + 7.495 = 7.4960$$

$$m = \frac{50 - 24}{55 - 24} (7.525 - 7.495) + 7.495 = 7.5202$$

$$q_3 = \frac{75 - 55}{85.5 - 55} (7.555 - 7.525) + 7.525 = 7.5447$$

Estadística Descriptiva

Datos Completos vs. Datos Agrupados

Estadístico	Completos	Agrupados
Media		
Desvío		
Coef. simetría		
Curtosis		
Primer cuartil		
Mediana		
Tercer Cuartil		