

¿Cómo medir y describir una variable aleatoria?

Análisis estadístico utilizando R



UNQ UNTreF CONICET

Pablo Etchemendy

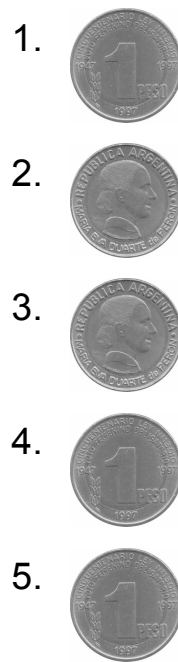
Ignacio Spiouzas

Agosto 2021

La variable aleatoria



Realización

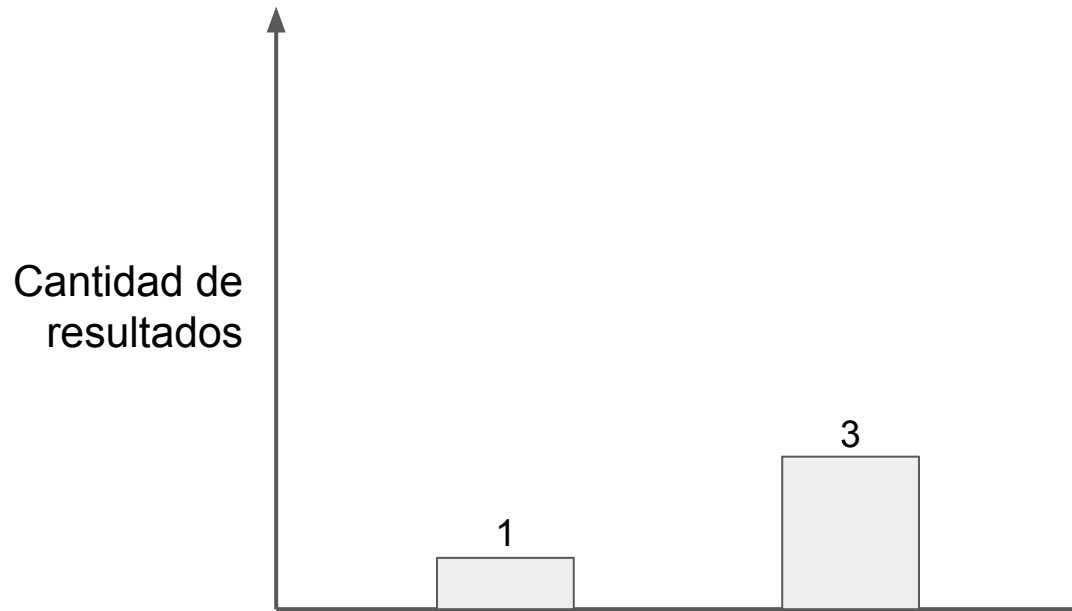


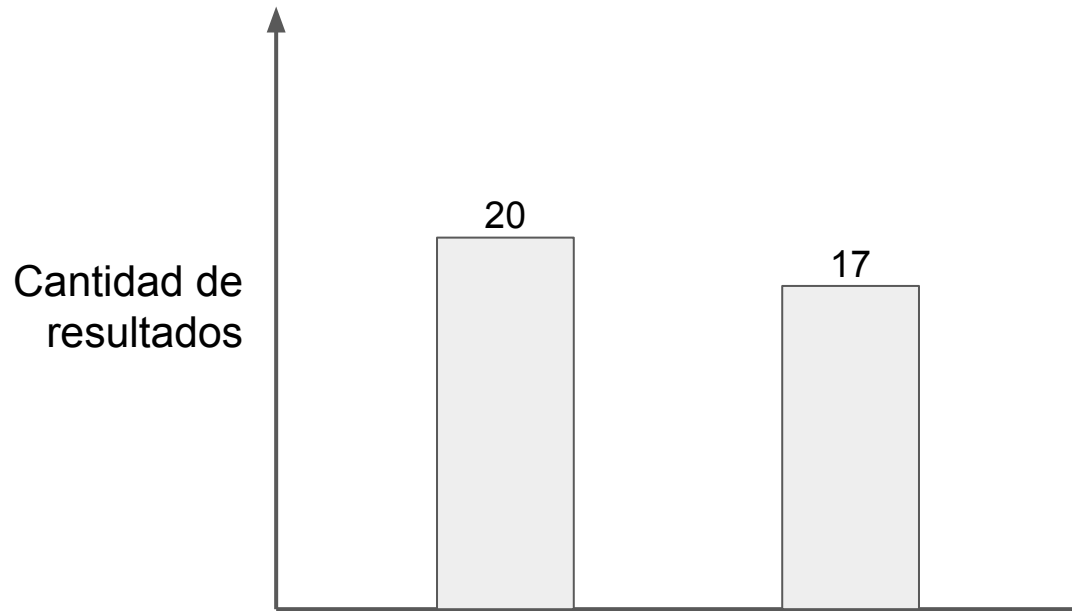
Histograma

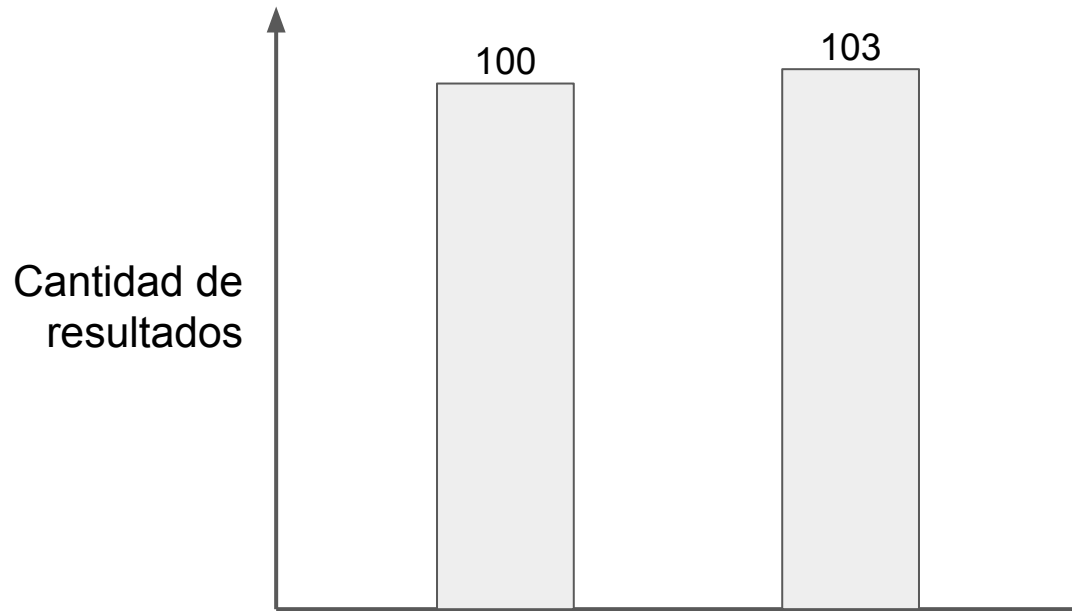
Gráfico que muestra la cantidad de apariciones de cada valor posible para una cantidad arbitraria N de realizaciones

Valores posibles









Frecuencia absoluta

Es la cantidad de veces que se obtiene cada valor posible

$$N_{\text{cara}}$$
$$N_{\text{ceca}}$$

Frecuencia relativa

Es la frecuencia absoluta normalizada por la cantidad de valores obtenidos

$$f_{\text{cara}} = \frac{N_{\text{cara}}}{N_{\text{cara}} + N_{\text{ceca}}}$$

$$f_{\text{ceca}} = \frac{N_{\text{ceca}}}{N_{\text{cara}} + N_{\text{ceca}}}$$

Caracterización de una variable aleatoria

- Conocer los valores posibles
- Obtener N realizaciones
- Calcular la cantidad de realizaciones que corresponden a cada valor posible
- Luego calcular la frecuencia relativa

En este ejemplo, para una cantidad **infinita** de realizaciones, esperamos obtener


$$f_{\text{cara}} = 0.5$$


$$f_{\text{ceca}} = 0.5$$

(si la moneda no está desbalanceada)

Para una cantidad **grande** (realizable):



$$f_{\text{cara}} \sim 0.5$$



$$f_{\text{ceca}} \sim 0.5$$

(la precisión dependerá del tamaño de la muestra)

Si en cambio la moneda no es equiprobable



$$f_{\text{cara}} = p$$



$$f_{\text{ceca}} = 1 - p$$

Otro ejemplo



- Valores posibles: 1, 2, 3, 4, 5 y 6
- Frecuencias relativas: equiprobables

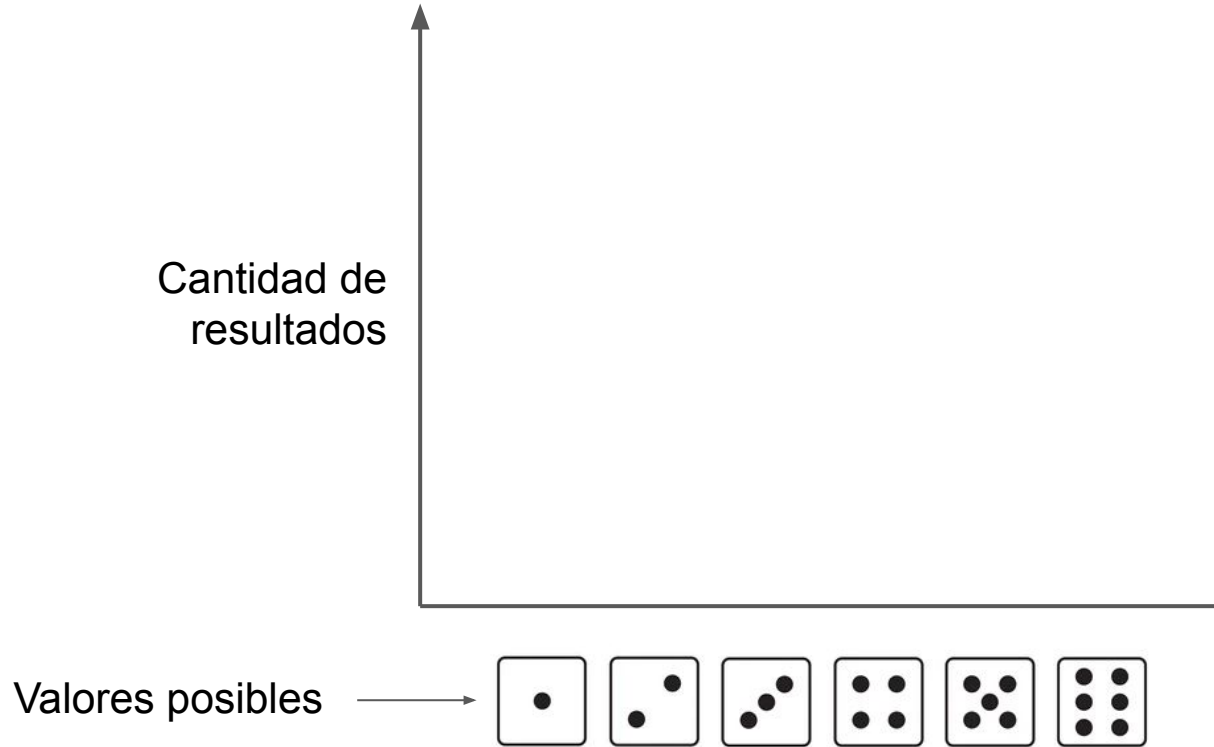
La variable aleatoria

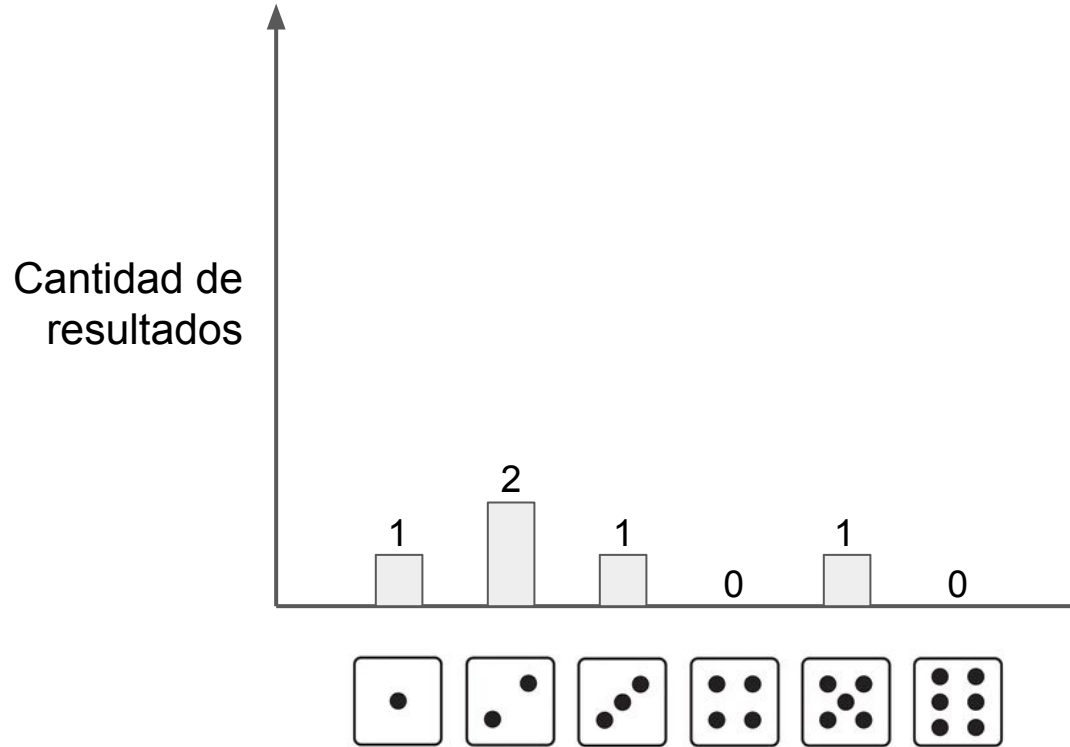


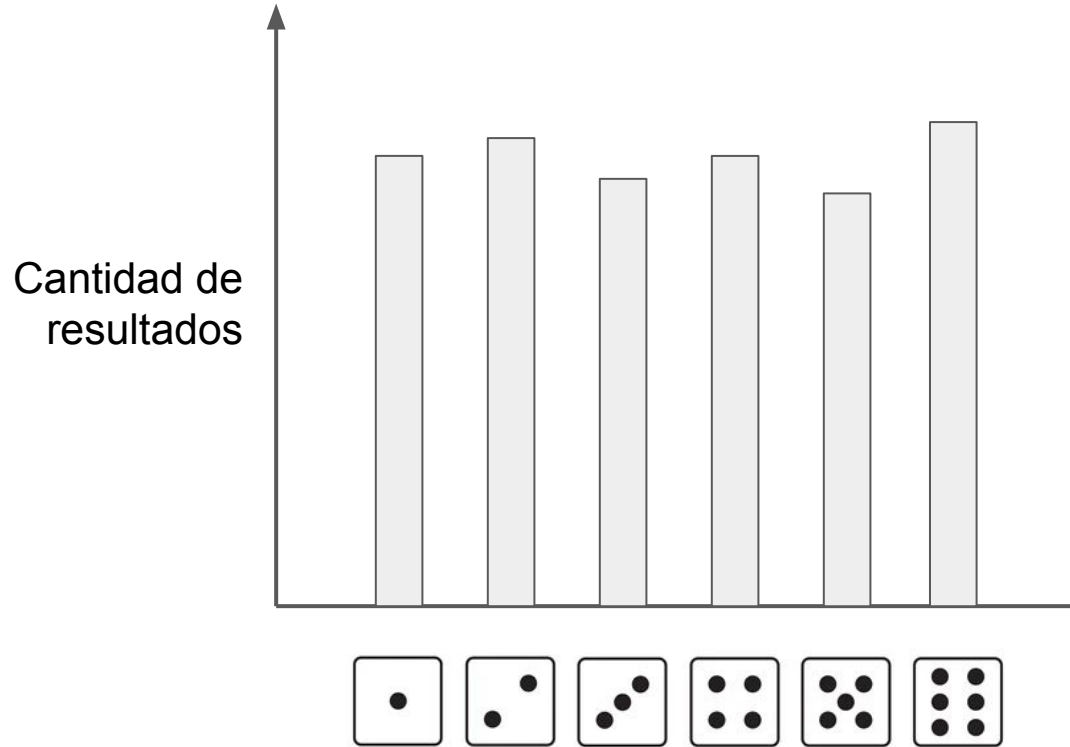
Realización



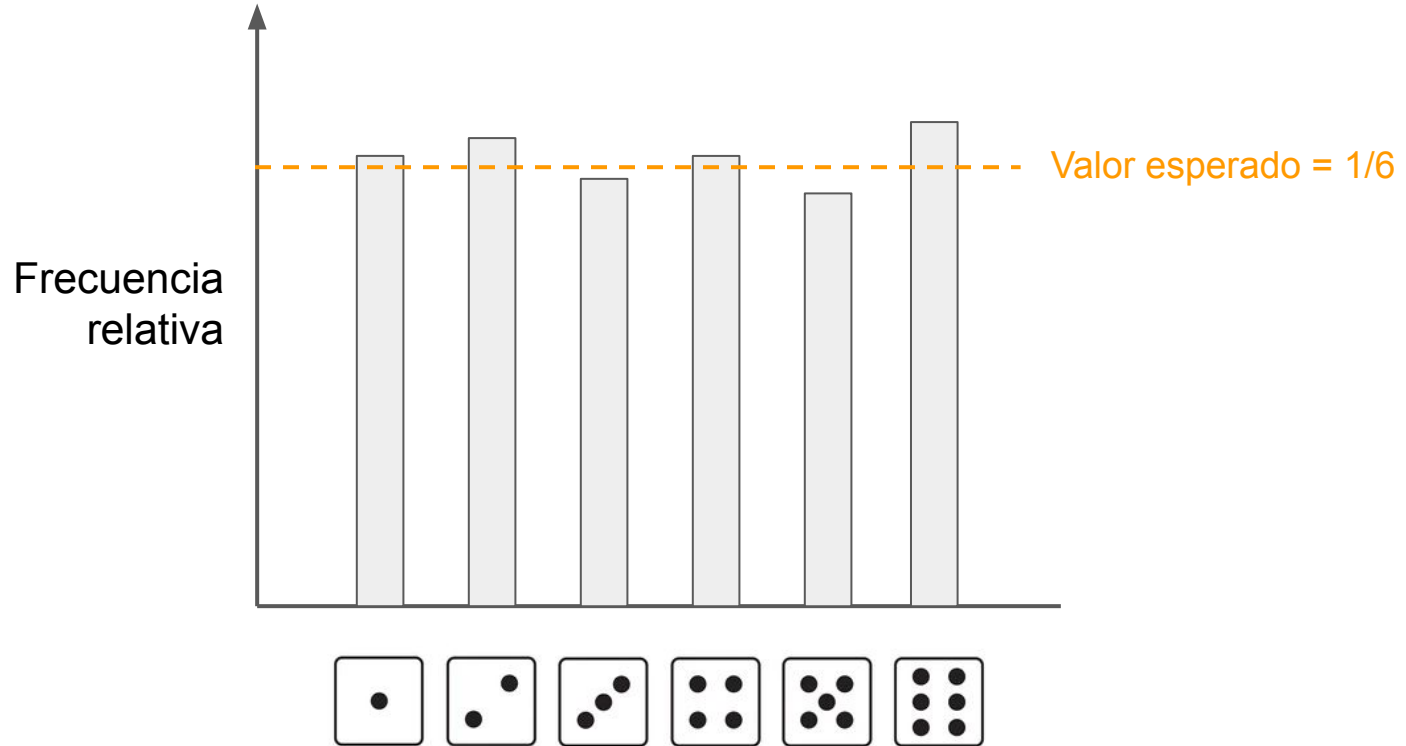
etc.



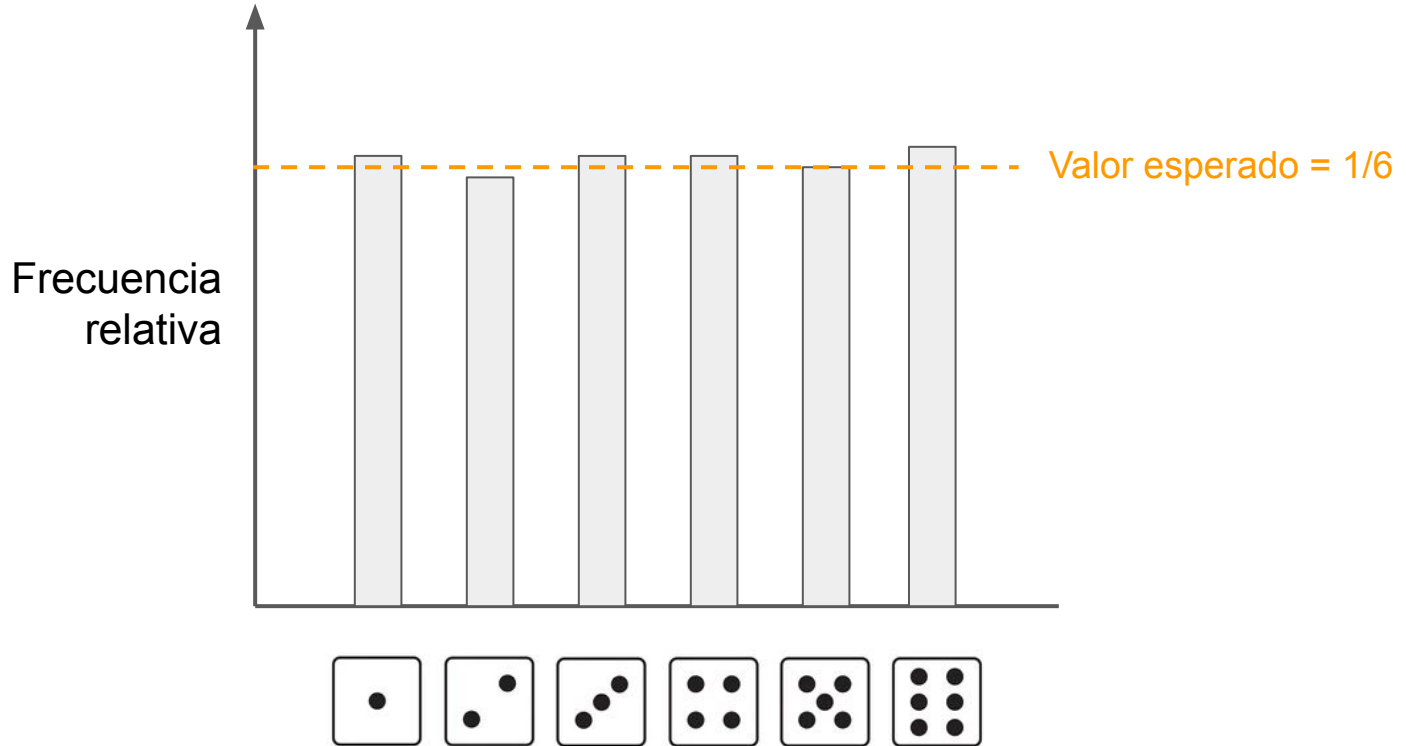




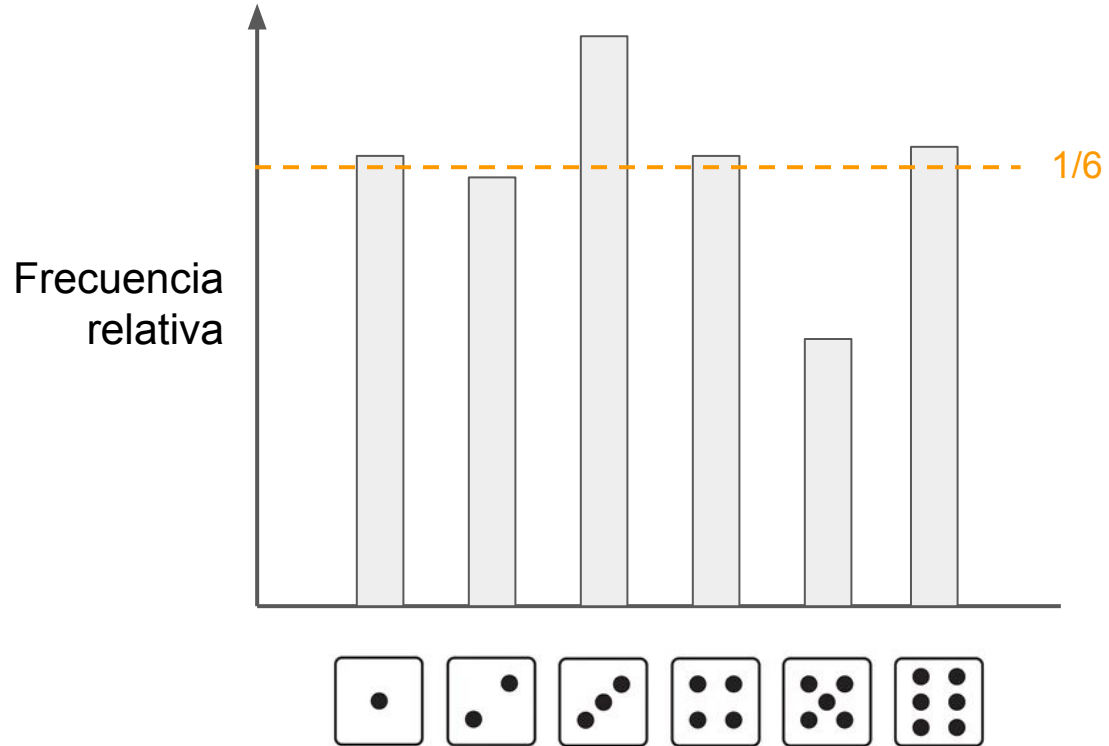
Veamos la frecuencia relativa...



A medida que agregamos realizaciones



Si el dado está desbalanceado



Los pasos son los mismos que ya vimos

- Conocer los valores posibles
- Obtener N realizaciones
- Calcular la cantidad de realizaciones que corresponden a cada valor posible
- Luego calcular la frecuencia relativa
- Necesitamos 6 valores: f_1, f_2, f_3, f_4, f_5 y f_6
- Y debe cumplirse que la suma es 1:

$$f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 1$$

Frecuencia relativa y probabilidad

- En cualquier experimento, la **frecuencia relativa** es una estimación de la **probabilidad** asociada a cada valor posible.
- Desde el punto de vista experimental, nunca es posible conocer perfectamente la probabilidad; pero podemos mejorar la **precisión de su estimación** aumentando el tamaño del experimento.
- Desde el punto de vista teórico, podemos definir variables aleatorias dándole un valor exacto a la probabilidad de ocurrencia de cada valor posible.

Variables aleatorias categóricas

- En el ejemplo de la moneda, la variable aleatoria es categórica: sus valores asociados no son numéricos.
- Otro ejemplos:
 - Color
 - Sexo
 - Etc.

Variables aleatorias discretas y continuas

- En el ejemplo del dado, las variables son numéricas **discretas**, ya que se trata de números enteros.
- A diferencia del caso categórico, los valores se pueden ordenar de menor a mayor
- Una variable aleatoria que toma valores reales será una variable **continua**.

Otros ejemplos

- Tirar N monedas y contar la cantidad de veces que sale cara.
 - Una realización consiste en tirar las N monedas.
 - Los valores posibles son: $1, 2, \dots, N$
 - Es una variable numérica discreta
 - Esperamos observar, en cada tirada, un valor alrededor de $N/2$ (asumiendo una moneda balanceada)
 - La probabilidad asociada a cada valor posible no será la misma

Otros ejemplos

- Tirar N dados y sumar sus valores
 - Una realización consiste en tirar los N dados.
 - Los valores posibles son: $1, 2, \dots, 6N$
 - Es una variable numérica discreta
 - Esperamos observar, en cada tirada, un valor alrededor de $3.5N$ (para pensar...)
 - Al igual que el caso anterior, la probabilidad asociada a cada valor posible no será la misma

Variables aleatorias continuas

- Los pasos que vamos a seguir son similares
- Conocer el *rango* de valores posibles:
 - Acotados entre dos valores finitos: entre a y b
 - Semi-acotados: 0 a infinito
 - No acotados: $-\infty$ a $+\infty$
- Obtener N realizaciones y luego...
- ...para cada valor, calcular las frecuencias absoluta y relativ...



Momento...

¿Para cada valor?

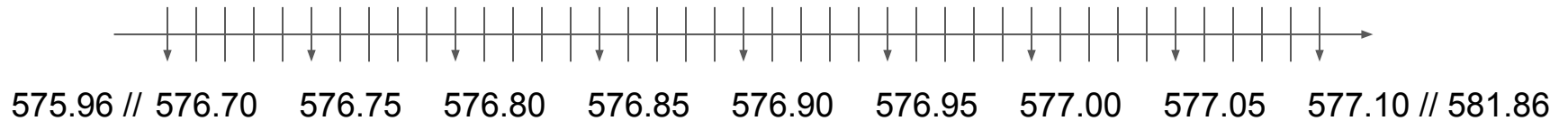
```
>> LakeHuron
```

```
[1] 580.38 581.86 580.97 580.80 579.79 580.39 580.42 580.82 581.40 581.32  
[11] 581.44 581.68 581.17 580.53 580.01 579.91 579.14 579.16 579.55 579.67  
[21] 578.44 578.24 579.10 579.09 579.35 578.82 579.32 579.01 579.00 579.80  
[31] 579.83 579.72 579.89 580.01 579.37 578.69 578.19 578.67 579.55 578.92  
[41] 578.09 579.37 580.13 580.14 579.51 579.24 578.66 578.86 578.05 577.79  
[51] 576.75 576.75 577.82 578.64 580.58 579.48 577.38 576.90 576.94 576.24  
[61] 576.84 576.85 576.90 577.79 578.18 577.51 577.23 578.42 579.61 579.05  
[71] 579.26 579.22 579.38 579.10 577.95 578.12 579.75 580.85 580.41 579.96  
[81] 579.61 578.76 578.18 577.21 577.13 579.10 578.25 577.91 576.89 575.96  
[91] 576.80 577.68 578.38 578.52 579.74 579.31 579.89 579.96
```

```

[1] 580.38 581.86 580.97 580.80 579.79 580.39 580.42 580.82 581.40 581.32
[11] 581.44 581.68 581.17 580.53 580.01 579.91 579.14 579.16 579.55 579.67
[21] 578.44 578.24 579.10 579.09 579.35 578.82 579.32 579.01 579.00 579.80
[31] 579.83 579.72 579.89 580.01 579.37 578.69 578.19 578.67 579.55 578.92
[41] 578.09 579.37 580.13 580.14 579.51 579.24 578.66 578.86 578.05 577.79
[51] 576.75 576.75 577.82 578.64 580.58 579.48 577.38 576.90 576.94 576.24
[61] 576.84 576.85 576.90 577.79 578.18 577.51 577.23 578.42 579.61 579.05
[71] 579.26 579.22 579.38 579.10 577.95 578.12 579.75 580.85 580.41 579.96
[81] 579.61 578.76 578.18 577.21 577.13 579.10 578.25 577.91 576.89 575.96
[91] 576.80 577.68 578.38 578.52 579.74 579.31 579.89 579.96

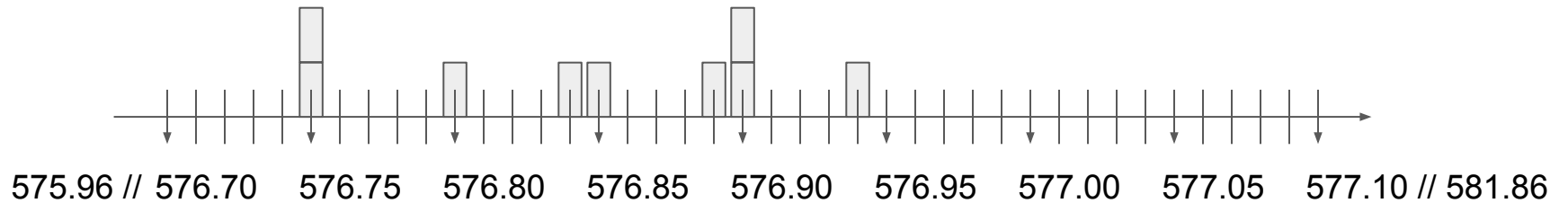
```



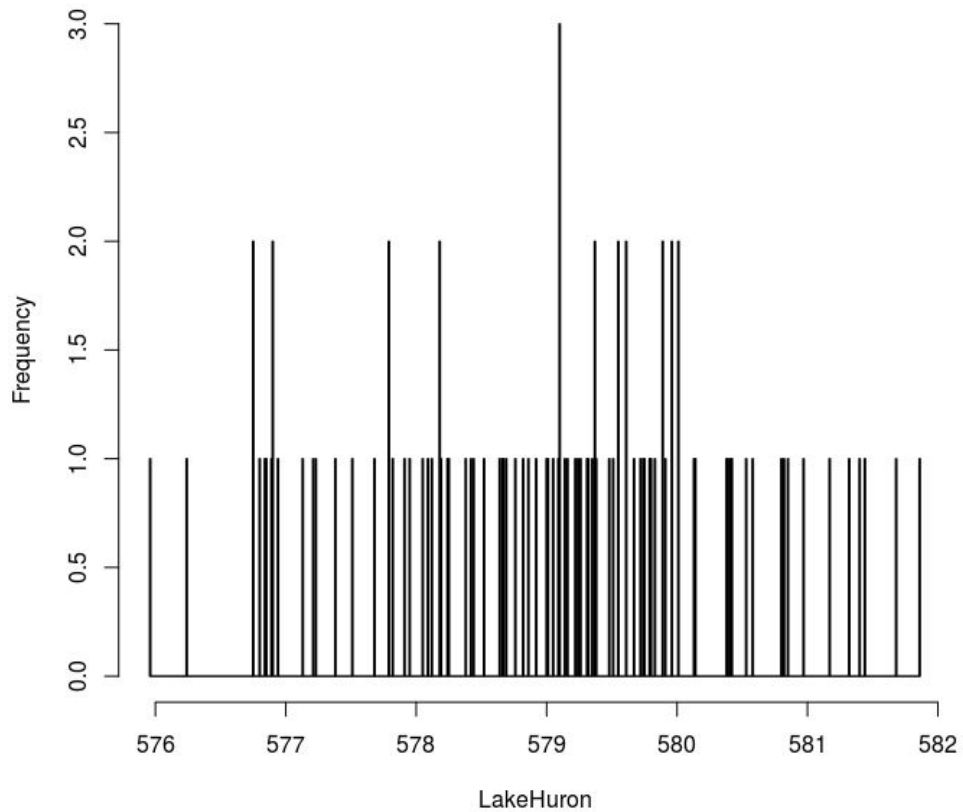
```

[1] 580.38 581.86 580.97 580.80 579.79 580.39 580.42 580.82 581.40 581.32
[11] 581.44 581.68 581.17 580.53 580.01 579.91 579.14 579.16 579.55 579.67
[21] 578.44 578.24 579.10 579.09 579.35 578.82 579.32 579.01 579.00 579.80
[31] 579.83 579.72 579.89 580.01 579.37 578.69 578.19 578.67 579.55 578.92
[41] 578.09 579.37 580.13 580.14 579.51 579.24 578.66 578.86 578.05 577.79
[51] 576.75 576.75 577.82 578.64 580.58 579.48 577.38 576.90 576.94 576.24
[61] 576.84 576.85 576.90 577.79 578.18 577.51 577.23 578.42 579.61 579.05
[71] 579.26 579.22 579.38 579.10 577.95 578.12 579.75 580.85 580.41 579.96
[81] 579.61 578.76 578.18 577.21 577.13 579.10 578.25 577.91 576.89 575.96
[91] 576.80 577.68 578.38 578.52 579.74 579.31 579.89 579.96

```



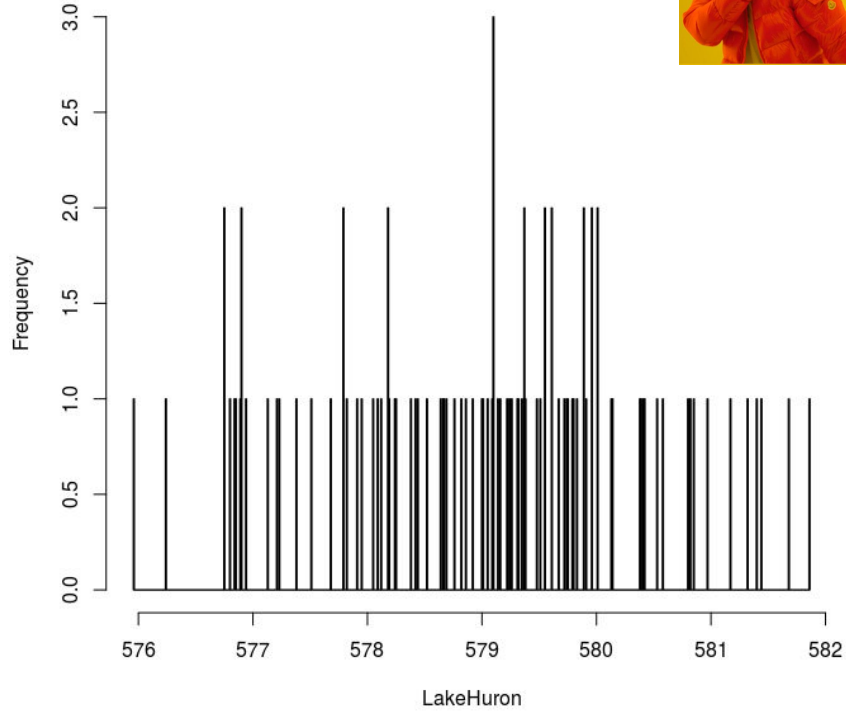
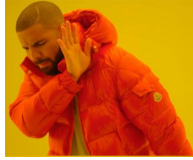
Histogram of LakeHuron



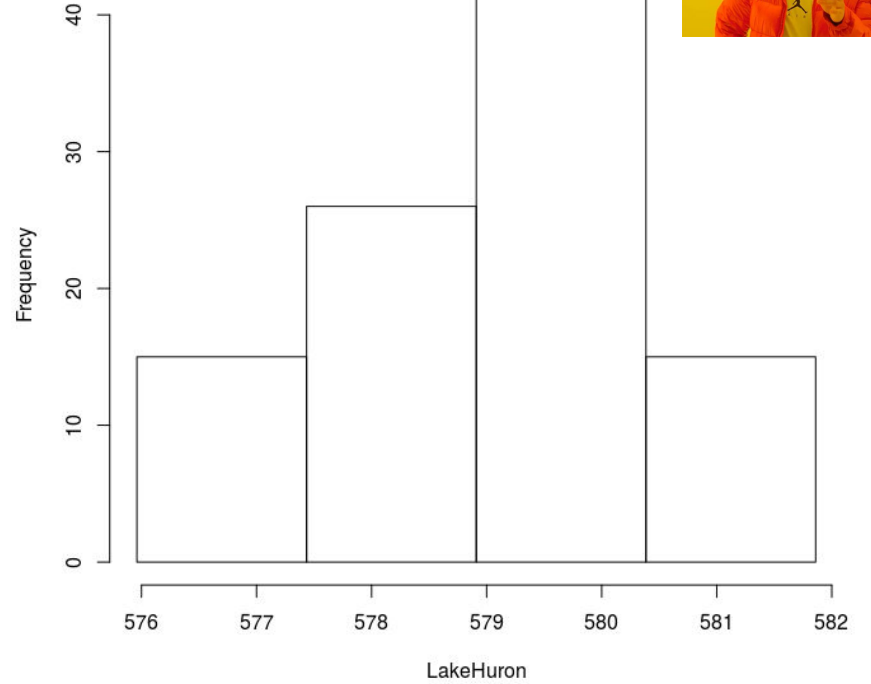
Histograma de una variable continua

- Determinar el rango de valores obtenidos
 - Es decir, los valores extremos: mínimo y máximo
- Dividir el rango en M intervalos arbitrarios de igual longitud
 - En nuestro ejemplo, el rango va desde 575.96 hasta 581.86
 - Si queremos 4 intervalos, calculamos la diferencia: 5.9
 - Luego calculamos $5.9 / 4 = 1.475$
 - Los intervalos serán:
 $[575.96, 577.435)$, $[577.435, 578.91)$, $[578.91, 580.385)$ y $[580.385, 581.86)$

Histogram of LakeHuron



Histogram of LakeHuron



- La cantidad de rangos que forman el intervalo es “arbitraria”: podemos elegirla (dentro de lo razonable)
- Existen algoritmos para determinar el rango óptimo para un conjunto de datos:
 - Sturges
 - Scott
 - Freedman-Diaconis
- Cada algoritmo está pensado para contemplar ciertos escenarios.