Preámbulo:
medidas de tendencia central y dispersión

Medidas resumen de los datos

- Medidas de tendencia central:
  - Media muestral, mediana, moda.
- Medidas de dispersión:
  - Min, max, varianza, desviación estándar, rango, cuantiles, coeficiente de variación.

Media muestral

Sea $x_1, x_2, \ldots, x_n$ las observaciones de una muestra. Se define a la media muestral, $\bar{x}$, como

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

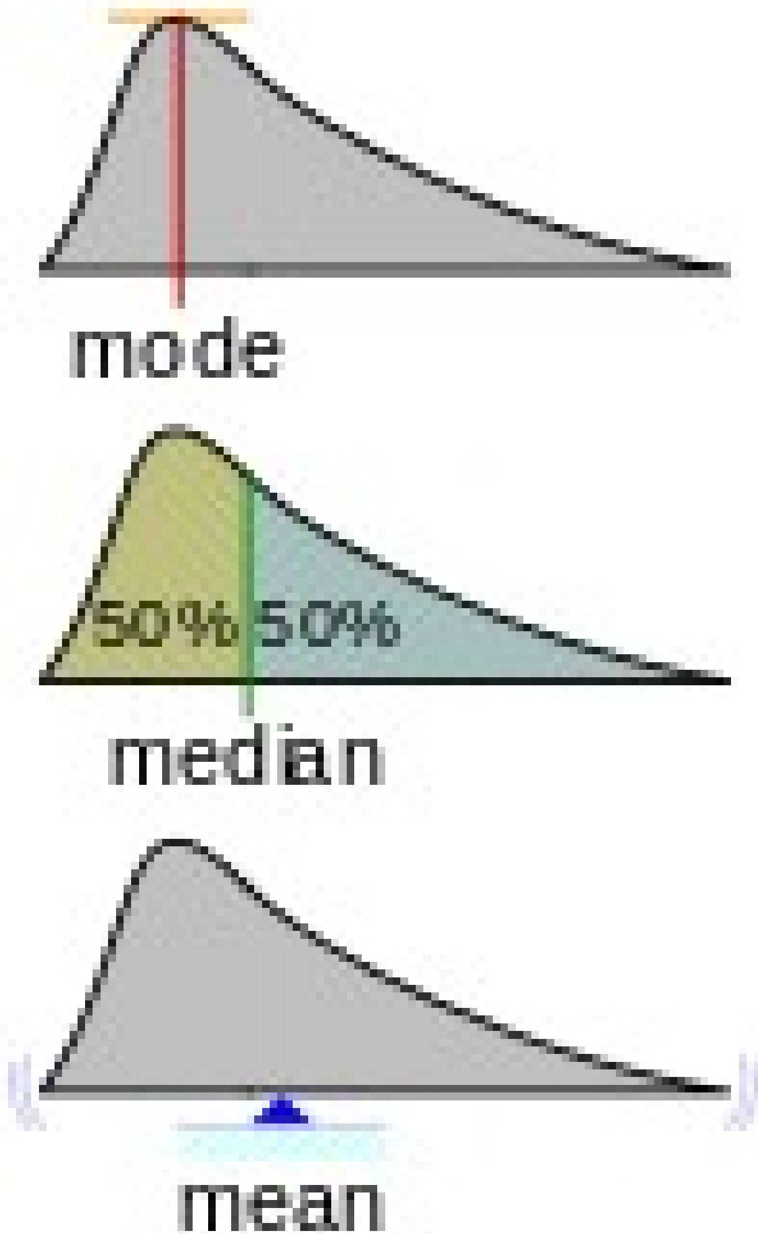
Moda

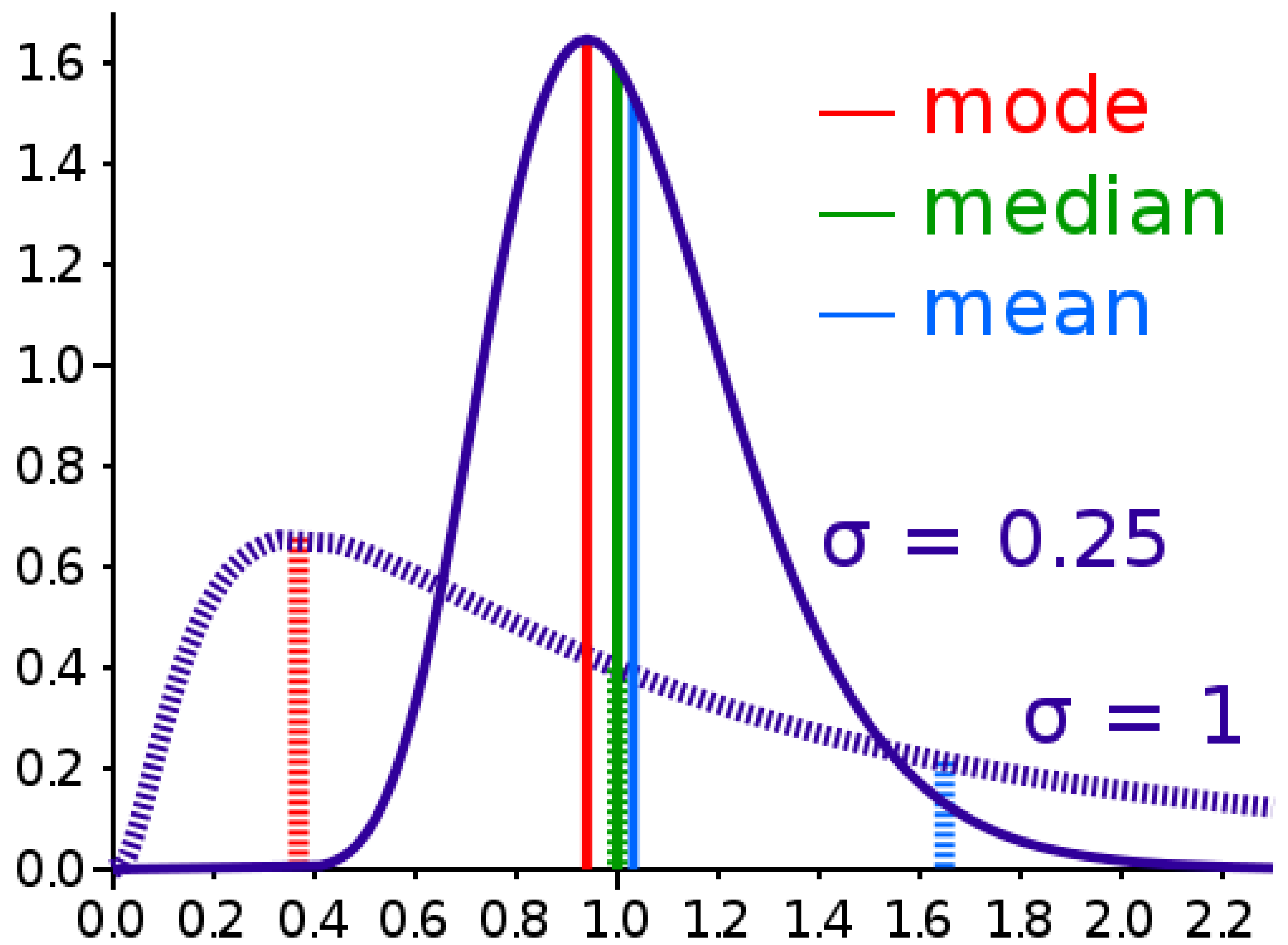
Es es valor más frecuente de una muestra, es decir, aquel con mayor cantidad de observaciones

Mediana

Sean $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ las observaciones de una muestra ordenadas de menor a mayor. Se define a la mediana, $\tilde{x}$, como

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & si\ n\ es\ impar \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & si\ n\ es\ par \end{cases}$$


mode
50%
50%
median
mean


mode
median
mean
σ = 0.25
σ = 1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2

Mínimo, Máximo, Rango, Quantil, Quartil

Sean $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ las observaciones de una muestra ordenadas de menor a mayor. Se define:

1. Mínimo $= x_{(1)}$
2. Máximo $= x_{(n)}$
3. Rango = máximo - mínimo $= x_{(n)} - x_{(1)}$
4. *Percentil*$_{\alpha\,\%}$ es el valor para el cual una fracción $\alpha\,\%$ se encuentra por debajo de dicha fracción.
5. Cuartil $Q_{1,\ 2,\ 3\ o\ 4}$: dividen a la muestra en 4. Así $Q_{1,\ 2,\ 3\ y\ 4}$ son el *Percentil*$_{25\,\%}$, *Percentil*$_{50\,\%}$, *Percentil*$_{75\,\%}$ y *Percentil*$_{100\,\%}$ respectivamente.

Varianza

La varianza muestral, s^2, se define como

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = \frac{1}{n(n-1)} \left[n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right]$$

Desviación estándar

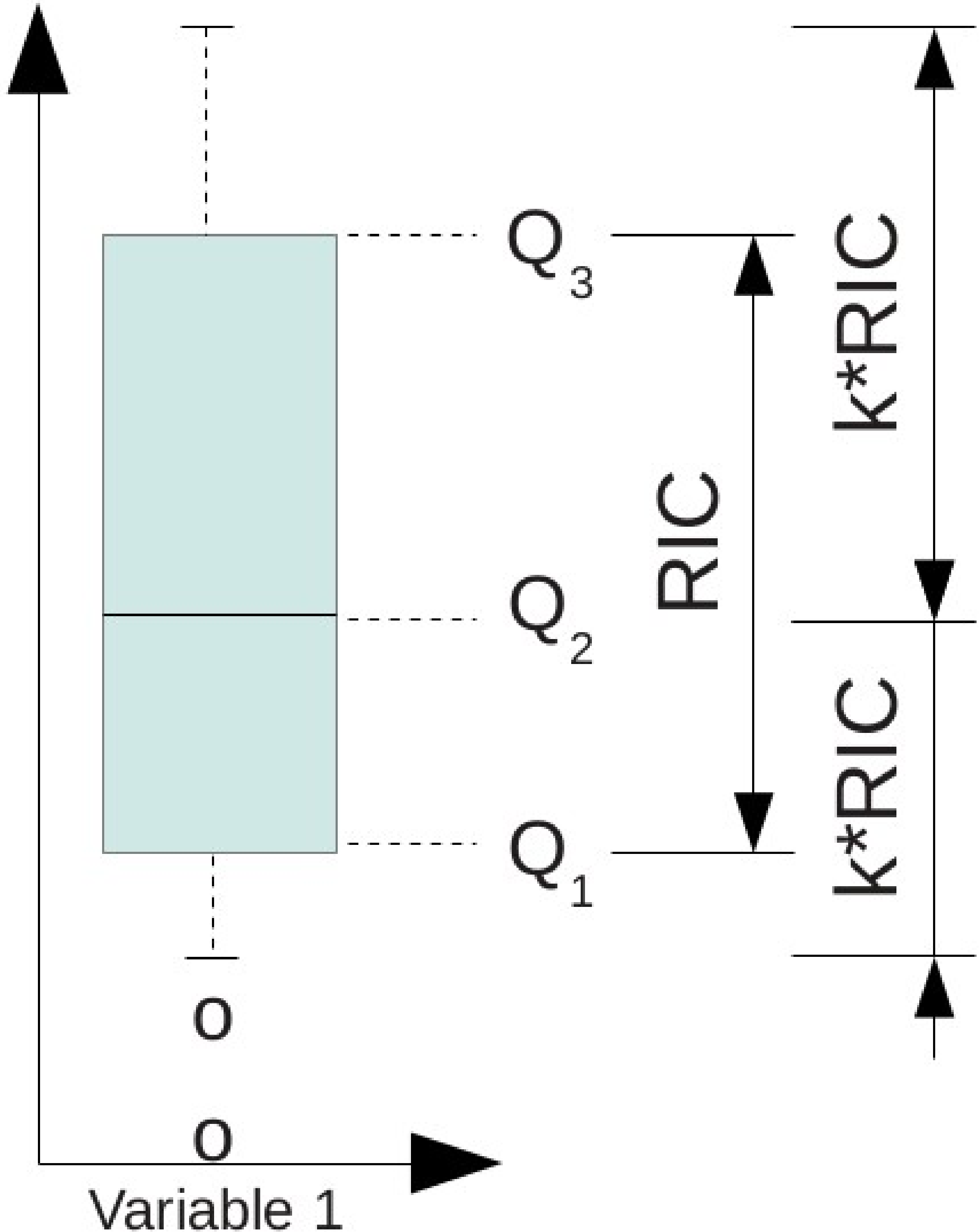
La desviación estándar muestral, s, es la raíz cuadrada positiva de la varianza, es decir

$$s = \sqrt{s^2}$$

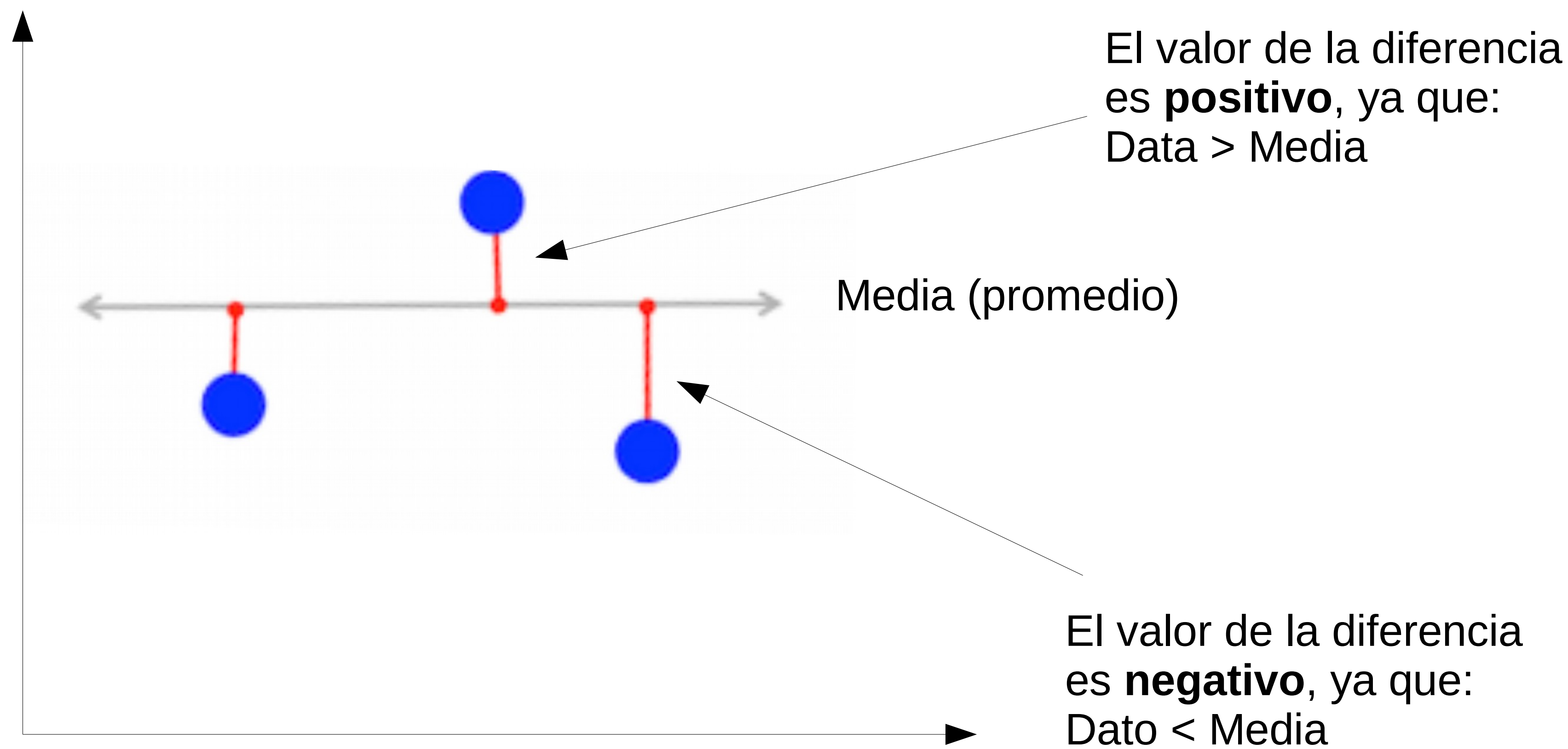
Coeficiente de variación

El coeficiente de variación de una muestra, $C.V.$, se define:

$$C.V. = 100\,\% \frac{s}{\bar{x}}$$

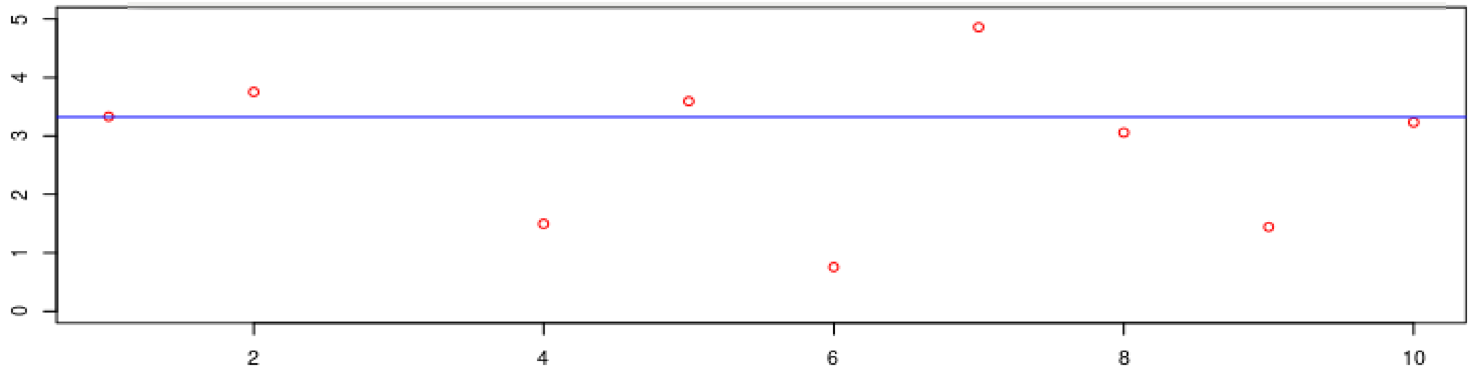


Q_3
Q_2
Q_1
RIC
k*RIC
k*RIC
o
o
Variable 1

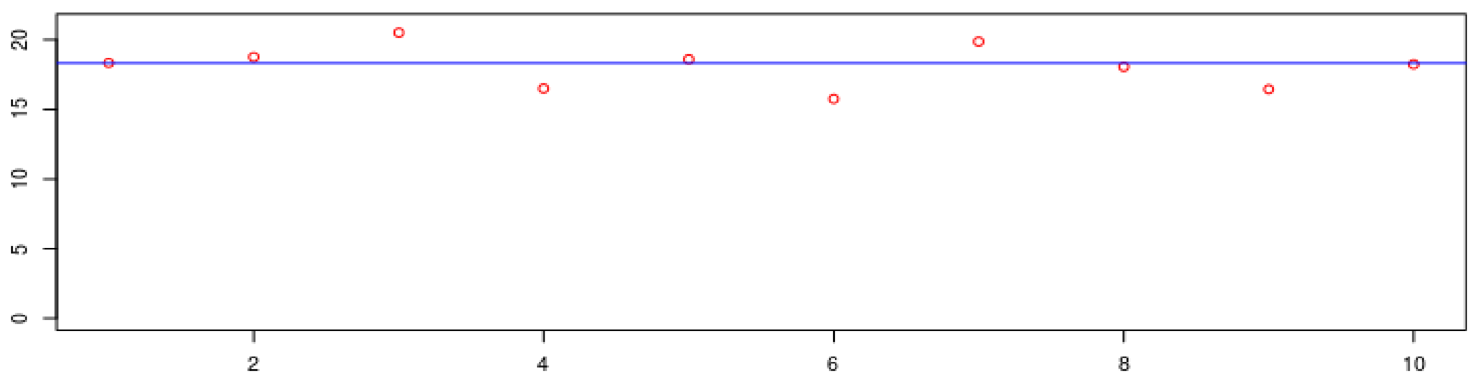

El valor de la diferencia
es **positivo**, ya que:
Data > Media
Media (promedio)
El valor de la diferencia
es **negativo**, ya que:
Dato < Media

```
b
[1] 3.3291997 3.7508973 5.5040235 1.4946863 3.5912733 0.7543911 4.8588845
[8] 3.0564333 1.4405746 3.2305013
b<-rnorm(10,3,1.5)
```

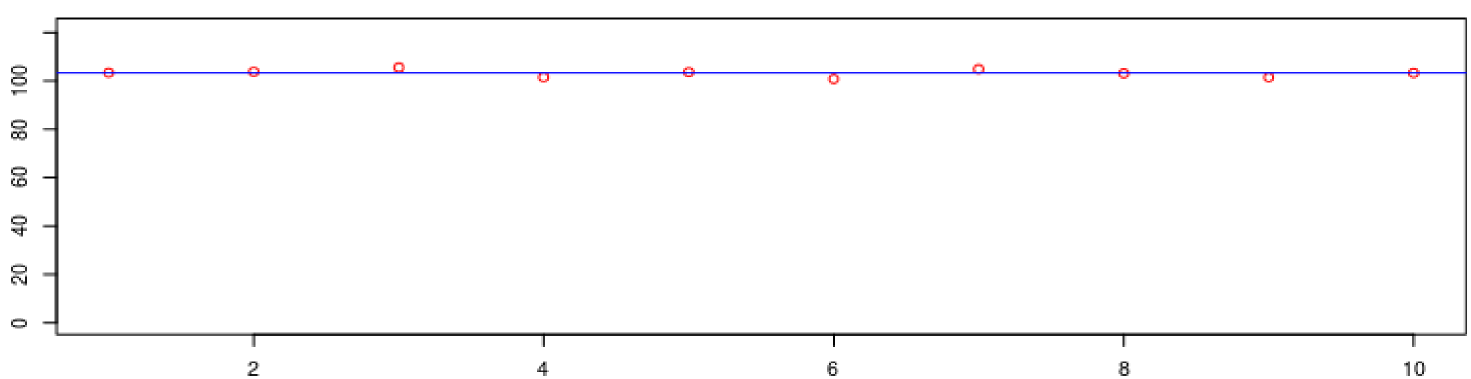
$$C.V. = 100\,\%\,\frac{s}{\bar{x}}$$



B
Mean = 3.101086
Sd = 1.507097
CV = 48.599



B + 15
Mean = 18.101086
Sd = 1.507097
CV = 8.326004



B + 100
Mean = 103.101086
Sd = 1.507097
CV = 1.461766

Ejercicio

Caso de estudio:

Se desea investigar si el nitrógeno es transferido por una bacteria al tallo de un árbol (ejemplo 1.2 [1]).

10 plantines con bacteria sin presencia de **nitrógeno**

10 plantines con bacteria con presencia de **nitrógeno**

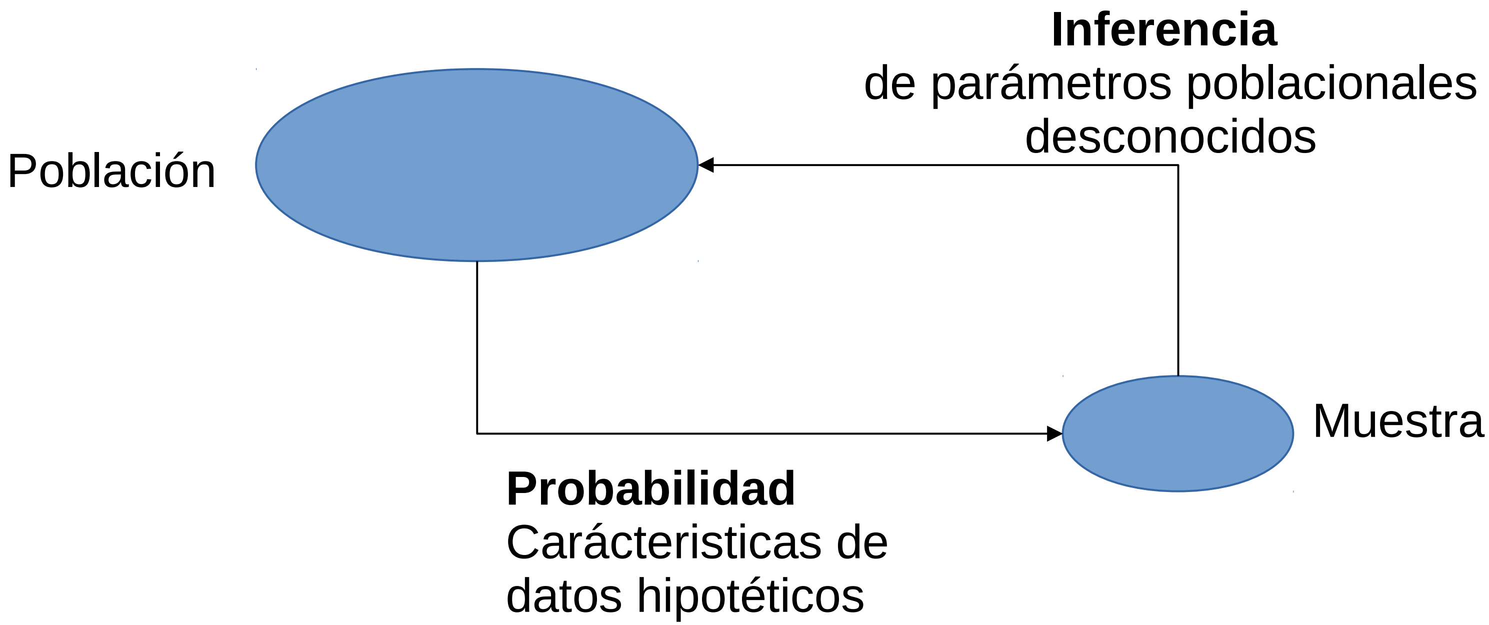
Se recolecta el **peso seco** en kg. del peso del tallo.

Obs.	Sin Nitrógeno	Con Nitrógeno
1	0.35	0.26
2	0.53	0.43
3	0.28	0.47
4	0.37	0.49
5	0.47	0.52
6	0.43	0.75
7	0.36	0.79
8	0.42	0.86
9	0.38	0.62
10	0.43	0.46

Pruebas de hipótesis

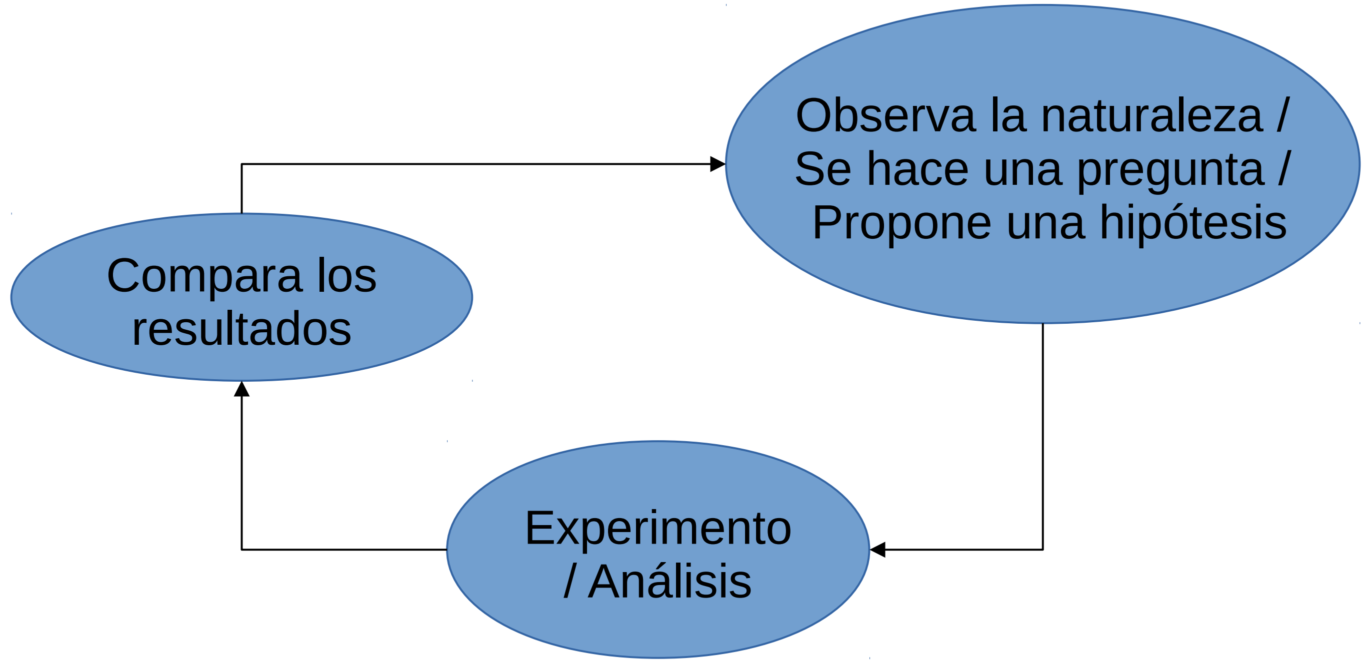
Objetivos

- Comprender el fundamento de cálculo de probabilidades
- Identificar los problemas donde sean aplicables técnicas estádisticas
- Aplicar la estadística y la probabilidad en diferentes ámbitos de la práctica médica.



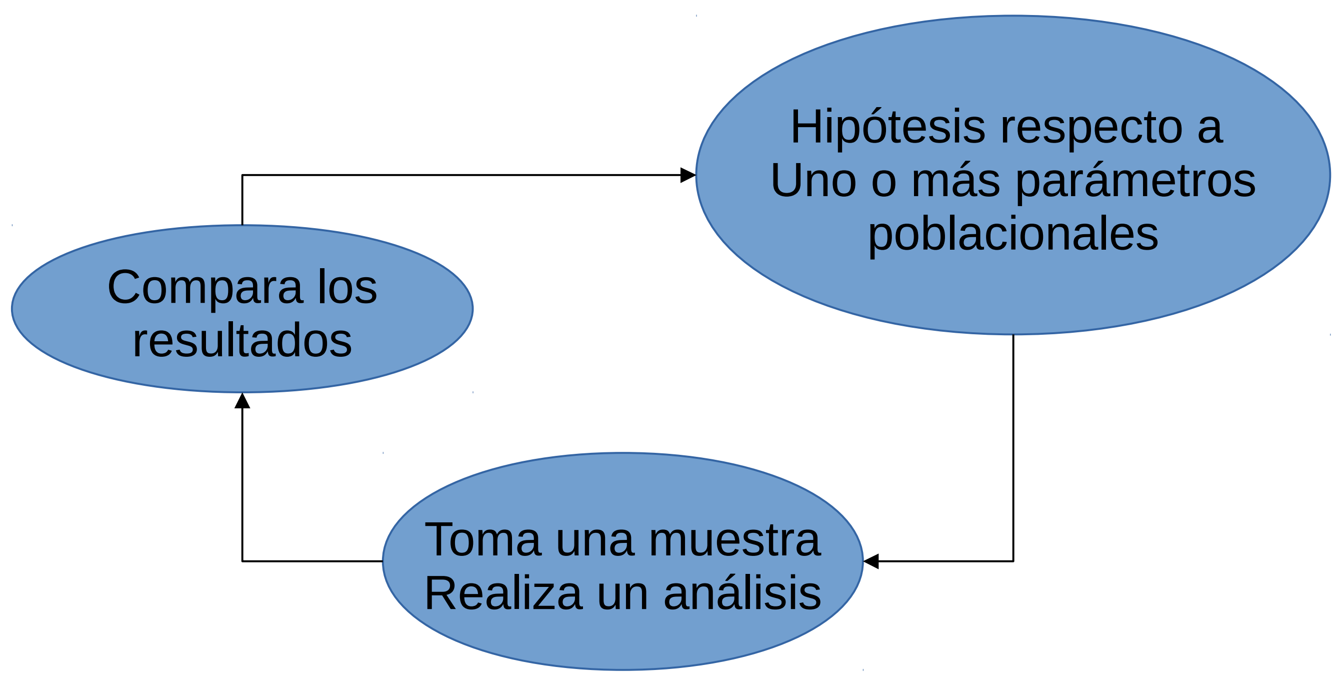
- Inferencia:
  - A) Estimaciones de los parámetros respectivos.
  - B) Pruebas de hipótesis acerca de sus valores.

Similitud al Método científico



Si las observaciones **NO** concuerdan con la Hipotésis => la rechaza

De lo contrario concluye que la hipótesis es verdadera o Que la prueba no detectó diferencia Entre los valores real e hipotético de los parámetros



Ejemplo

- Supongamos que el candidato Juan dice que obtendrá más del 50% de los votos, por lo tanto ganará.
- Si no creemos en lo que dice Juan podríamos tener la **hipótesis alternativa** (Ha) que no será favorecido por más del 50%.
- El apoyo a Ha se obtiene usando los datos muestrales que lo contrario a Ha, llamado **hipótesis nula** (H0) es falso (muy poco probable).
- Entonces una teoría se comprueba demostrando que no hay evidencia que sustente la teoría opuesta, “una prueba por contradicción”.

- Nuestra hipótesis alternativa es que $V < 0.5$. Si podemos encontrar que los datos apoyan el rechazo de H0 ($V >= 0.5$) hemos alcanzado nuestro objetivo.
- Suponga que se seleccionan n (= 15) votantes.
- Y = el número de votantes a favor de Juan.
- Si Y = 0, ¿Qué se concluiría de los dicho por Juan?.
  - Si en realidad $V >= 0.5$ no es imposible observar Y = 0, pero es altamente improbable.
- ¿ Si fuera Y = 1?

Elementos de una prueba estadística

- 1. Hipótesis nula. H_0
- 2. Hipótesis alternativa. H_a
- 3. Estadístico de prueba.
  - Es una función de las mediciones muestrales, en las que la desición estádistica estará basada.
- 4. Región de rechazo (RR).
  - Específica los valores del estadístico para el cual H_0 ha de ser rechazada en favor de H_a . Si el estadístico no cae en RR aceptamos H_0

Regresando a nuestro ejemplo de Juan

- Pequeños valores de Y nos llevarían a rechazar H0. Una región de rechazo que pudieramos considerar seria RR = { Y <= 2 }.
  - Hayar una buena región de rechazo es un problema no trivial que ameritará más de nuestra atención.
- Para cualquier región de rechazo fija, dos tipos de errores se pueden cometer:
  - Decidir a favor de Ha cuando H0 es verdadera (**error tipo I**), la probabilidad de error tipo I se denota por **α** (alfa).
  - Decidir a favor de H0 cuando Ha es verdadera (**error tipo II**) la probabilidad de error tipo II se denota por **β** (Beta).

- En nuestro ejemplo:
- Error tipo I sería rechazar H0 : V >= 0.5 (y por lo tanto aceptar Ha: p < 0.5) cuando H0 es verdadera.
  - Concluir que Juan perderá cuando en realidad va a ganar.
- Error tipo II sería aceptar H0 : V >= 0.5 cuando V < 0.5
  - Concluir que Jones ganará cuando en realidad va a perder.
- Las decisiones incorrectas cuestan dinero, prestigio, tiempo / pérdida.
  - Entonces **α** y **β** miden los riesgos relacionados con las dos posibles decisiones erróneas que podrían resultar.

Ejercicio 1

- Para la encuesta de Juan se muestrearon n = 15 votantes. Deseamos probar H0: p = 0.5 contra la alternativa Ha: p < 0.5 . El estadístico de prueba es Y, Calcule **α** si RR = { y <= 2 }.
- **α** = P (error tipo I) = P (rechazar H0 cuando H0 es verdadera).
- = P (estadístico de prueba esta en RR cuando H0 es verdadera).
- = P (Y >= 2 cuando p = 0.5).
- **α** $= \sum_{0}^{2} \binom{15}{y} (0.5^{y})(0.5^{15-y}) = \binom{15}{0} (0.5^{0})(0.5^{15-0}) + \binom{15}{0} (0.5^{1})(0.5^{15-1}) + \binom{15}{0} (0.5^{2})(0.5^{15-2})$
- **α = .004**

Entonces si decidimos usar R = { y <= 2} asumimos un riesgo
Muy pequeño (alfa = 0.004) de concluir que Jones perderá sin en
Realidad es ganador

Ejercicio 2

- ¿Nuestra pruba es tan buena como para evitar concluir que Jones va a ganar cuando en realidad perderá? (error tipo II → **β**) Suponga que él recibirá 30% de los votos (p = 0.3). ¿Cuál es la probabilidad **β** de que la muestra erróneamente nos lleve a concluir que H0 es verdadera y Jones va a ganar?.
- **β** = P (error tipo II) = P (aceptar H0 cuando Ha es verdadera).
- = P (estadístico de prueba **NO** esta en RR cuando Ha es verdadera).
- = P (Y >= 2 cuando p = 0.5).
- **β** $= \sum_{y=3}^{15} \binom{15}{y} (0.3^y)(0.7^{15-y})$
- **β = .873**

Entonces si decidimos usar R = { y <= 2} por lo gneral nos
Llevará a concluir que Jones es ganador (con una probabilidad = 0.873)
aún cuando p es tan bajo como 0.3

El valor de Beta depende del verdadero valor del parámetro p. Cuanto
Mayor sea la diferencia entre p real y el valor hipotético (nulo) p, menor es la
Probabilidad de que no rechazemos la hipotésis nula.

Ejercicio 3

- Calcule el valor de Beta si Juan recibirá sólo 10 % de los votos ($p = 0.1$).

Ejercicio 3

- Calcule el valor de Beta si Juan recibirá sólo 10 % de los votos (p = 0.1).
- **β** = P (error tipo II) = P (aceptar H0 cuando Ha es verdadera (P = 0.1)).
- = P (estadístico de prueba **NO** esta en RR cuando Ha es verdadera).
- = P (Y >= 2 cuando p = 0.1).
- **β** $= \sum_{y=3}^{15} \binom{15}{y} (0.1^{y})(0.9^{15-y})$
- = . 184

El valor de Beta cuando p =.1 es menor que el valor de Beta cuando p = .3 (.184 vs .873) No obstante usando esta RR todavía tenemos una Probabilidad bastante grande de decir que Juan ganará aún con p = 0.1

- Usando RR = { y >= 2} garantiza poco riesgo de cometer error tipo I (alfa = 0.004) pero no ofrece protección adecuada contra error tipo II.
- ¿Cómo podemos mejorar nuestra prueba?.
- Si agrandamos RR, esto nos llevará a rechazar H0 con más frecuencia entonces alfa se agrandará, al mismo tiempo una RR mayor nos llevará a aceptar RR con menos frecuencia por lo tanto Beta disminuirá.
- **En consecuencia α y β están relacionadas de manera inversa**.

Ejercicio 4

- Recordando el ejercicio 1 (p = .5) ahora suponga que RR = { y <= 5 }. Calcule el nivel α de la prueba y calcule β si p = .3. Compare los resultados con lo obtenido anteriormente.

Ejercicio 4

- Recordando el ejercicio 1 (p = .5) ahora suponga que RR = { y <= 5 }. Calcule el nivel **α** de la prueba y calcule **β** si p = .3. Compare los resultados con lo obtenido anteriormente.
- Alfa = P (estadístico está en RR cuando H0 es verdadera)
- = P (Y <= 5 cuando p = 0.5) $= \sum_{y=0}^{5} \binom{15}{y} (0.5^{y})(0.5^{15-y})$
- = .151
- Cuando p = 0.3
- Beta = P (estadístico no esta en RR cuando Ha es verdadera)
- = P (Y > 5 cuando p = 0.3) $= \sum_{y=6}^{15} \binom{15}{y} (0.3^{y})(0.7^{15-y})$
- = 0.278

	RR	
	Y < = 2	Y < = 5
α	.004	.151
β	.873	.278

- **α** de la prueba y calcule **β** siguen siendo muy grandes ¿Cómo podemos reducirlas?
  - Debemos obtener más información (aumentar el tamaño de la muestral)
- Para casi todas las pruebas estadísticas si **α** esta fija en algún valor aceptablemente pequeño **β** disminuye cuando el tamaño muestral aumenta.

Otra forma de presentar los resultados de una prueba estadística:

- Niveles de significancia alcanzados o **valores de p.**

- Aún cuando se recomienden pequeños valores de alfa, el valor real de alfa para usar en un análisis es un tanto arbitrario.

- Un experimentador puede escoger alfa = 0.5 y otro alfa = 0.1 . Por lo tanto es posible que dos personas que analicen la misma información lleguen a conclusiones opuestas.

- Los valores 0.5 o 0.1 a menudo se emplean por costumbre o comodidad más que como resultado de una cuidadosa evaluación de las consecuencias de cometer un error tipo I.

- Si W * es un estadístico de prueba , el valor p , o nivel de significancia alcanzado , es el nivel más pequeño de alfa para el cual la información observada indica que la hipotesis nula debe ser rechazada.
- Cuanto más pequeño sea el valor de p, es más fuerte la evidencia de que una hipotésis nula debe ser rechazada.

Referencias

Wackerly Dennis, Mendenhall William, Scheaffer Richard, Estadística Matemática con Aplicaciones, CENGAGE Learning, séptima edición, págs 488 a 513.

Agradecimiento a las notas en beamer del Dr. Cristobal Fresno Ródrigez (INMEGEN).