

**Universidad de Puerto Rico  
Recinto de Río Piedras  
Facultad de Estudios Generales  
Departamento de Ciencias Biológicas**

## **La estadística en la investigación Científica**

**Cursos para los que se recomienda el ejercicio de laboratorio:** CIBI 3005, CIBI 3006, CIBI 3015, CIBI 3026 y CIBI 3027.

**Temas del prontuario del curso relacionados con el ejercicio de laboratorio:** método científico.

**Competencias de investigación que se desarrollarán:** observación, planteamiento de un problema, formular una hipótesis, recopilación y análisis de datos cualitativos, y llegar a conclusiones.

**Duración:** uno o dos periodos de laboratorio, a discreción del profesor

### **Introducción:**

Siempre que se aplica el método científico, se debe de comenzar con el diseño apropiado del experimento, continuar con la recopilación y lectura exacta de los datos, y con el análisis apropiado de estos. Estos tres (3) pasos: diseño, recopilación y análisis de los datos, están relacionados estrechamente unos con otros. Si alguno de los pasos es tratado incorrectamente, el resultado estadístico carecerá de valor. Por ejemplo, es importante que, cuando se tome una muestra, cada elemento tenga la misma probabilidad de ser seleccionado para que los resultados sean confiables y la muestra sea representativa de la población.

La aplicación de la estadística en el campo de la biología facilita el análisis de las medidas tomadas (datos) a diversas características de los organismos. La biometría es la ciencia que utiliza los datos provenientes de los organismos vivos. La biometría se descompone en bio = vida y metría = medición.

La estadística es uno de los instrumentos con que cuenta el científico para recolectar, organizar, analizar, sintetizar, interpretar y evaluar conjuntos de datos o variables. El conjunto de observaciones o medidas que hace el investigador al realizar una investigación se conocen por datos. Estos nos permiten describir cualitativa o cuantitativamente el objeto bajo estudio. Algunos de los propósitos, en el uso de la estadística, son el determinar la confiabilidad de los datos y ayudar en predecir futuros resultados. Cuando se hacen observaciones, se recopilan datos para luego ser organizados y analizados para llegar a conclusiones.

El análisis estadístico se basa principalmente en la descripción cuantitativa de los datos u observaciones. El proceso mediante el cual se recopilan los datos se conoce por el **diseño experimental o experimento**. La estadística nos permite reducir (sintetizar) los datos a cifras que puedan ser entendidas y comparadas para luego evaluar los mismos. De otra parte, nos permite determinar la validez y confiabilidad de los datos, para poder llegar a conclusiones sobre la muestra bajo estudio. Es por ello que la estadística se divide en dos áreas de estudio: la estadística descriptiva y la estadística inferencial.

**La estadística descriptiva** describe aquello que se conoce, y se divide a su vez en medidas de tendencia central y medidas de dispersión. **Las medidas de tendencia central** nos dicen cuán cerca están los datos de su centro. Esta se compone de la moda, la mediana y la media o promedio aritmético ( $\bar{x}$ ). **La moda** se define como el valor o los valores que más se repiten dentro de un conjunto de datos, **la mediana** es el valor central luego de organizados los datos y **la media o promedio aritmético** se define como la suma de todas las muestras individuales dividido entre el total de muestra  $\bar{x} = \sum(x_i)/n$ . Este es un valor representativo de todos los datos

Por otro lado, **las medidas de dispersión** nos dicen cuán alejados están los datos del centro. Esta se compone del Rango, la varianza, la desviación estándar y el error estándar. **El rango (R)** es la diferencia entre el valor máximo y el valor mínimo dentro del conjunto de datos  $R = \text{Valor}_{\text{máximo}} - \text{Valor}_{\text{mínimo}}$ . **La varianza ( $s^2$ )** es la suma de las diferencias al cuadrado entre los valores individuales y el promedio dividido entre el total de muestra menos uno  $s^2 = \sum(x_i - \bar{x})^2 / n - 1$ . Por su parte, **la desviación estándar (s)** es la raíz cuadrada de la varianza  $s = \sqrt{s^2}$ . Por otro lado, **el error estándar (SE)** es el resultado entre el valor de la desviación estándar dividida entre la raíz cuadrada del total de muestra  $SE = s / \sqrt{n - 1}$ .

Una vez recolectados y organizados los datos producto del experimento se pueden presentar los mismos en una tabla de distribución de frecuencia o **histograma**.

### **Objetivos específicos:**

Al finalizar el ejercicio de laboratorio el estudiante podrá:

1. Formular preguntas de investigación e hipótesis a partir de las observaciones de los fenómenos observados.
2. Distinguir entre datos cualitativos y datos cuantitativos.
3. Recolectar, organizar, analizar, sintetizar, interpretar y evaluar sets de datos.
4. Distinguir las medidas de tendencia central de las medidas de dispersión.
5. Reconocer y utilizar las medidas de tendencia central de: la moda, la mediana y la media o promedio aritmético.
6. Reconocer y utilizar las medidas de dispersión de: el Rango, la varianza, la desviación estándar y el error estándar.
7. Sostener, modificar o rechazar la hipótesis formulada.

## Equipo y materiales

Para el ejercicio de laboratorio se requiere lo siguiente:

1. 20 hojas de plantas, las hojas deben ser mayores de 10 centímetros de largo.
2. Reglas milimetradas o un calibrador Vernier.
3. Una balanza
4. Un bolígrafo permanente
5. Papel de gráfica

## Procedimiento

1. Se recolectará una muestra de 20 hojas provenientes todas de una misma planta o árbol.
2. Utilizando un marcador permanente se identificará cada hoja asignándole un número comenzando por el uno (1) y terminando con el veinte (20).
3. Medir el largo en centímetros (cm) y el peso en gramos (gr) de cada una de las hojas. Anotar los resultados en la Tabla 1, Tamaño vs peso de las hojas.
4. Hacer el análisis de datos utilizando la estadística descriptiva de tendencia central; moda, mediana, media o promedio aritmético. Anotar los resultados en la Tabla 2. Calcular los valores de la estadística descriptiva de dispersión; rango, varianza, desviación estándar y error estándar, para el conjunto de datos de peso y largo de las hojas. Anotar los resultados en la Tabla 3. Para determinar la varianza, se utilizará las Tablas 4 y 5. Determinación de la varianza de la muestra de hojas, para el largo y el peso de las hojas.
5. Construir el histograma (distribución de frecuencia) para el peso y el largo de las hojas.
6. Determinar el coeficiente de correlación de Spearman para el largo y el peso de las hojas. Para determinar el Coeficiente de Correlación de Spearman utilizará la Tabla 6.

**Tabla 1. Tamaño vs peso de las hojas**

Núm. de muestra	Peso de la hoja en gr ( $x_i$ )	Largo de la hoja en cm ( $y_i$ )	Observaciones Describa cualitativamente 1 hoja y haga un dibujo de la hoja
1			
2			
3			
4			
5			
6			
7			
8			
9			

10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

**Tabla 2. Estadística descriptiva de tendencia central de la muestra de hojas.**

<b>Análisis estadístico</b>	<b>Peso de las hojas (gr)</b>	<b>Largo de las hojas (cm)</b>
Moda		
Mediana		
Media o promedio aritmético		

**Tabla 3. Determinación de la varianza de la muestra de hojas para el largo en centímetros**

<b>Núm. de la muestra</b>	<b>Largo en cm (<math>x_i</math>)</b>	<b>Promedio (<math>\bar{x}</math>)</b>	<b><math>(x_i - \bar{x})</math></b>	<b><math>(x_i - \bar{x})^2</math></b>
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				

20				
----	--	--	--	--

**Tabla 4. Determinación de la varianza de la muestra de hojas para el peso en gramos**

Núm. de la muestra	Peso en gr ( $x_i$ )	Promedio ( $\bar{x}$ )	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

**Tabla 5: Estadística descriptiva de dispersión para la muestra de hojas.**

Análisis estadístico	Peso de las hojas (gr)	Largo de las hojas (cm)
Rango		
Varianza		
Desviación estándar		
Error estándar		

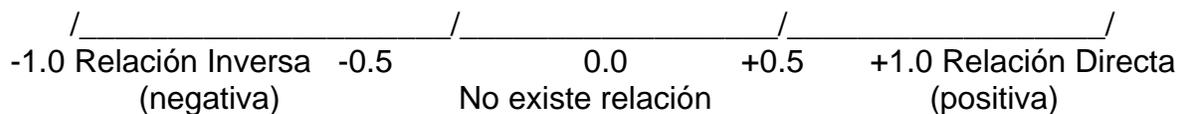
### Histograma:

Construya un histograma con los datos del tamaño y peso de las hojas en un papel de gráfica. Debe incluir el título de la gráfica, identificar los nombres del eje de X (variable independiente) y de Y (variable dependiente). Identificar los límites inferiores y superiores de cada clase. Para construir la gráfica debe seguir los siguientes pasos.

1. Organizar los datos obtenidos en orden ascendente o descendente.
2. Determinar el **rango o amplitud** de la muestra restando el valor mayor al valor menor.
3. Agrupe los datos en **intervalos o clases**. El número de intervalos o clases será arbitrario y dependerá del total de datos obtenidos. A mayor cantidad de datos, mayor número de intervalos o clases. Sin embargo, el número de intervalos o clases no debe de ser menor de 2 y mayor de diez.
4. Calcule el **tamaño del intervalo o clase** dividiendo el rango o amplitud entre el número de intervalos o clases que se ha seleccionado.
5. Determine los límites inferior y superior de cada intervalo. El **límite inferior** es donde comienza la clase y el **límite superior** es donde termina cada clase o frontera.
  - a. El límite inferior de la primera clase corresponde al valor mínimo de los datos, mientras el límite superior de la primera clase corresponde al valor del límite inferior más el valor del intervalo.
  - b. El límite inferior de la próxima clase corresponde al valor del límite superior de la clase anterior más 1 unidad (1, .01, .001, etc. dependiendo de las unidades de los datos) y así sucesivamente.
  - c. El límite superior de la segunda clase es el límite inferior de la clase mas el valor del intervalo, y así sucesivamente se obtienen los límites superiores e inferiores.
  - d. Se calcula la frecuencia determinando la cantidad de datos que se encuentran dentro de los límites inferior y superior de cada clase.

### **Coeficiente de Correlación de Rangos de Spearman (Sr):**

El Coeficiente de Correlación de Rangos de Spearman (Sr) es una prueba no paramétrica que se utiliza de manera particular para aquellas situaciones en la que el número de casos es menor de 30. Esta prueba nos dice que entre más se acerca a **+1 mayor será la relación directa o positiva** entre las variables bajo estudio. Por el contrario, a medida que nos acercamos a **-1 mayor será la relación inversa o negativa entre las variables**. Sí el valor del Coeficiente de Spearman (Sr) se encuentra entre **-0.5 y +0.5** entonces **no existe relación** alguna entre las variables



La ecuación para determinar el Coeficiente de Correlación de Rangos de Spearman (Sr) es la siguiente:

$$Sr = 1 - \frac{6 \sum (R_x - R_y)^2}{n(n^2 - 1)}$$

Donde  $\sum (R_x - R_y)^2$  es la suma de las diferencias al cuadrado entre los **rangos** de los valores de **X** (variable independiente) y **Y** (variable dependiente), mientras **n** es el total de sujetos o muestra.



**Preguntas autoevaluación**

1. ¿Qué relación existe entre el peso y el largo de las hojas?
2. ¿Los datos del peso y el largo de las hojas tienden a agruparse o dispersarse de su centro? Explique su contestación.
3. La muestra de hojas, ¿exhibe variabilidad? Explique su contestación.
4. ¿Cuál es la importancia de la estadística en la investigación científica?

**Actividades sugeridas para desarrollar las competencias de información:**

1. Haga un listado de tres (3) referencias sobre el tema discutido en este ejercicio de laboratorio. Use el formato APA.

**Referencias:**

- Bluman, Allan G. Elementary Statistics: 2004. A Step by Step Approach, 5ta ed. McGraw Hill: Higher Education.
- Clíford Blair, R. & Richard A. Taylor. 2008. Bioestadística. Pearson: Prentice Hall. México.
- Hernández Sampieri, R., Fernández Collado, C. y Batista Lucio, P. 2004. Metodología de la Investigación, 2da ed. McGraw-Hill,
- Hincle, D., E., Wiersma, W. and Jurs, S., G. 2003. Applied Statistics for Behavioral Sciences, 5ta ed. Houghton Mifflin Co.
- Sánchez Viera, J. A. Fundamentos del Razonamiento Estadístico, 3ra ed. 2001 Universidad Carlos Albizu. San Juan, Puerto Rico.