

# Recursos Cinegéticos y Piscícolas. 3º Grado en Ingeniería Forestal y del Medio Natural

## Práctica 1. Fundamentos de Bioestadística

### 1. Media y varianza

Considérese una muestra de  $n$  peces (Tabla 1), todos de la misma especie, capturados en un lance con pesca eléctrica, y sea  $x_i$  la talla del  $i$ -ésimo pez,  $i=1, 2, \dots, n$ . La talla media, se define como:

$$\bar{x} = (x_1 + x_2 + \dots + x_n)/n = \frac{1}{n} \sum_{i=1}^n x_i$$

La varianza muestral, que es una medida de la variabilidad en torno a la media, se define de la siguiente forma:

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

De esta forma, la varianza  $S^2$ , es la suma de los cuadrados de las desviaciones respecto a la media, dividida por  $n-1$ . La raíz cuadrada de la varianza se denomina desviación estándar. A menudo también es interesante determinar la varianza relativa al tamaño de la talla media, para ello,  $S$  es la cantidad apropiada ya que tiene la misma unidad que la media. Esto implica que la desviación estándar relativa o coeficiente de variación (CV) se pueda calcular como:

$$CV = \frac{S}{\bar{x}}$$

**1.1. Actividad 1.** Calcula la media, desviación estándar, varianza y coeficiente de variación de la muestra. Construye la tabla en MS Excel y comenta con el resto del grupo qué significa el valor de desviación estándar obtenido.

Para muchos propósitos es conveniente disponer de muestras en forma de tablas de frecuencia, para lo cual se divide el recorrido de las tallas en varios intervalos de longitud. En la Tabla 1 el recorrido de tallas va desde 11.2 a 19.0 cm.

**1.2. Actividad 2.** Construye una tabla con grupos de 1 cm de intervalo, tomando 10.5 cm como límite inferior y 19.5 cm como límite superior. Recalcula la talla media, la varianza, desviación estándar y el coeficiente de variación a partir del punto medio de cada intervalo (Tabla 2).

**Tabla 1. Talla de los 27 individuos capturados durante el lance**

Pez (n)	Talla (cm)	S	S <sup>2</sup>
<i>i</i>	$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	14.2		
2	16.3		
3	14.8		
4	13.2		
5	16.9		
6	12.2		
7	14.3		
8	15.7		
9	15.3		
10	11.2		
11	12.9		
12	13.5		
13	18.2		
14	11.6		
15	18.5		
16	16.3		
17	15.5		
18	15.8		
19	13.2		
20	19.0		
21	12.0		
22	17.1		
23	15.4		
24	14.6		
25	14.0		
26	18.1		
27=n			
Total			

**Tabla 2. Media y varianza de una muestra de frecuencias de tallas**

Índice ( <i>j</i> )	Intervalo (cm)	Punto medio $\bar{L}_j$	Frecuencia ( $F_j$ )	$F_j \bar{L}_j$	$(\bar{L}_j - \bar{x})$	$F_j (\bar{L}_j - \bar{x})^2$
1						
2						
3						
4						
5						
6						
7						
8						
9						
Total						
Talla media:						
Varianza:						
Desviación Estándar:						
Coefficiente de variación:						

**1.3. Actividad 3.** Construye un gráfico de frecuencias con MS Excel a partir de la columna 4 de la Tabla 2.

## 2. Distribución normal

En la Tabla 2 se muestra como ejemplo un pequeño conjunto de datos de frecuencias que se ajustan aproximadamente a una distribución normal. Ésta puede ser expresada como:

$$F(x) = \frac{n dL}{S \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2S^2}}$$

Donde,  $F(x)$  es la frecuencia calculada o frecuencia teórica;  $n$  es el número de observaciones;  $dL$  es el tamaño del intervalo;  $S$  es la desviación estándar;  $\bar{x}$  es la talla media y  $\pi$  tiene un valor de 3.1416.

**2.1. Actividad 4.** Conocidos los valores de  $n$ ,  $dL$ ,  $S$  y  $\bar{x}$ , representa la curva de distribución normal con MS Excel.

La distribución normal se observa en una gran variedad de casos diferentes y de ahí su nombre. Hay otros tipos de distribuciones de probabilidad (log-normal, binomial negativa, delta, entre otras). Una diferencia notoria entre éstas y la distribución normal es que son asimétricas, en tanto que la distribución normal es simétrica.

La característica más importante de la distribución normal es la que tiene que ver con las medias. Si se toman, por ejemplo, 50 muestras aleatorias de una determinada población, cada una de ellas con 25 observaciones, los cincuenta valores medios se distribuirán aproximadamente de forma normal.

Si en la ecuación de la distribución normal hacemos  $x = \bar{L}_j$  y dividimos ambos miembros de la ecuación entre  $n$ , se obtiene:

$$\frac{F(\bar{L}_j)}{n} = \frac{dL}{S \sqrt{2\pi}} e^{-\frac{(\bar{L}_j - \bar{x})^2}{2S^2}}$$

$$j=1, 2, \dots, 9 \text{ y } dL=1.$$

Los nuevos valores encontrados de  $\frac{F(\bar{L}_j)}{n}$  sumarán casi 1. Es decir, cada valor indica la probabilidad de que un pez extraído aleatoriamente pertenezca al intervalo de tallas correspondiente.

**2.2. Actividad 5.** Calcula los valores de  $\frac{F(\bar{L}_j)}{n}$  en la Tabla 3 para cada intervalo de la Tabla 2. Supón ahora que se lleva a cabo un nuevo muestreo de 1000 ejemplares. ¿Cuántos ejemplares teóricamente deberían encontrarse en el intervalo 14.5-15.5?

**Tabla 3. Probabilidad para cada intervalo**

<i>j</i>	Intervalo	Probabilidad
1		
2		
3		
4		
5		
6		
7		
8		
9		

### 3. Límites de confianza

Generalmente la talla media verdadera de una población se desconoce. Si se tratara de una población de peces cultivados en un estanque se podría medir la talla de la población, pero en el caso de peces en libertad es imposible. Por lo tanto es necesario conocer el grado de precisión de la talla media muestral, es decir, de la probable magnitud de la desviación entre la estimación y la media verdadera. Esta incertidumbre acerca de la media verdadera se expresa a través de los límites de confianza. En el caso de una distribución normal, los límites de confianza están dados por:

$$\text{Límite inferior: } \bar{x} - t_{n-1} \frac{S}{\sqrt{n}}; \text{ Límite superior: } \bar{x} + t_{n-1} \frac{S}{\sqrt{n}}$$

Donde  $n$  es el tamaño de la muestra,  $S$  la desviación estándar y  $t_{n-1}$  son los llamados percentiles en la distribución  $t$  de Student (Tabla 4). El argumento g.l. en la Tabla 4 se denomina grados de libertad. En general el número de grados de libertad es el número de observaciones menos el número de parámetros. Como en este caso  $\bar{x}$  es el único parámetro, el número de g.l. será igual a 1.

Los límites de confianza pueden ser calculados con diferentes niveles de precisión, normalmente al 95% o 99%. Mientras más alto sea el nivel, mayores serán los cuantiles y por lo tanto los intervalos serán más anchos entre los límites superiores e inferiores.

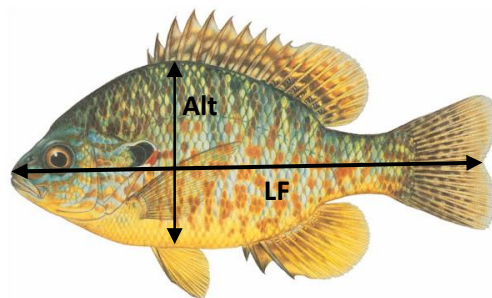
**3.1. Actividad 6.** Calcula los límites de confianza al 95% para la talla media de los peces que se muestran en la Tabla 1. Suponiendo que el muestreo se repite 100 veces bajo las mismas condiciones, ¿qué significan estos límites de confianza? ¿Qué sucede si se utiliza el cuantil del 99%?

**Tabla 4. Valores de los cuantiles de la distribución t de Student**

g.l.	95% ( $t_f$ )	99% ( $t_f$ )
1	12.71	63.66
2	4.30	9.93
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.37	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3.17
11	2.20	3.11
12	2.18	3.06
13	2.16	3.01
14	2.15	2.98
15	2.13	2.95
16	2.12	2.92
17	2.11	2.90
18	2.10	2.88
19	2.09	2.86
20	2.09	2.85
25	2.06	2.79
30	2.04	2.75
40	2.02	2.70
50	2.01	2.68
60	2.00	2.66
80	1.99	2.64
100	1.98	2.63
$\infty$	1.96	2.58

#### 4. Análisis de regresión lineal simple

Este método se utiliza para describir la variación de una cantidad, por ejemplo, la altura corporal de un pez, como función lineal de otra cantidad, por ejemplo la talla. Supongamos que hemos medido la talla total y la altura corporal de una muestra de 7 individuos (Figura 1, Tabla 5).



**Figura 1.** Medidas tomadas en cada ejemplar

**Tabla 5. Ejemplo de mediciones de talla furcales (x) y las respectivas alturas (y)**

<b>I</b>	<b>LF=x<sub>i</sub></b>	<b>Alt=y<sub>i</sub></b>
<b>1</b>	11.2	3.0
<b>2</b>	12.4	3.2
<b>3</b>	13.5	4.0
<b>4</b>	15.7	4.8
<b>5</b>	17.1	4.8
<b>6</b>	18.5	4.9
<b>7</b>	19.0	5.6

Como es de suponer, la altura del cuerpo tiende a aumentar cuando la talla aumenta. Si las proporciones corporales de un pez permanecieran constantes para todos los tamaños, la altura sería proporcional a la talla, y podría describirse con un modelo del tipo:

$$y_i = a + bx_i$$

Donde  $a$  y  $b$  son constantes o parámetros denominados intercepto con el eje  $y$ , y pendiente respectivamente. Por otro lado,  $y$  se denomina variable dependiente y  $x$  variable independiente. Los parámetros  $a$  y  $b$  pueden ser estimados por el método de mínimos cuadrados, es decir, se buscan aquellos valores que minimicen:

$$\sum_{i=1}^n (y_i - a - bx_i)^2$$

Donde  $n$  es el número de pares de observaciones.

El valor de la pendiente puede ser obtenido mediante:

$$b = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n (x_i)^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}$$

Mientras que el valor del intercepto se puede calcular como:

$$a = \bar{y} - \bar{x}b$$

**4.1. Actividad 7.** Calcula los parámetros  $a$  y  $b$  para los pares de datos de la tabla 5. Representa la recta de regresión con MS Excel y comprueba que los cálculos son correctos.

## 5. El coeficiente de correlación ( $r$ )

El coeficiente de correlación ( $r$ ), es una medida de la asociación lineal entre dos cantidades, ambas sujetas a una variación aleatoria. Puede variar entre -1 y 1. Su interpretación es la siguiente: si  $r$  es negativo, entonces  $y$  tiende a disminuir cuando  $x$  aumenta. Esta afirmación también vale para la pendiente  $b$ . El coeficiente de correlación es fácilmente calculable a partir de la siguiente expresión:

$$r = b \left( \frac{S_x}{S_y} \right)$$

Donde  $b$  es la pendiente,  $S_x$  es la desviación estándar de  $x$  y  $S_y$  es la desviación estándar de  $y$ .

**5.1. Actividad 8.** Calcula el coeficiente de correlación asociado a los pares de datos de la Tabla 5.