

Categoría: Bachiller
Liceo Francés de Bilbao

Tutora: Sandrine Romera

¿ Por qué los intervalos de confianza nos ayudan a tomar decisiones ?

Dasha
Valdés

Gabriela
Espada

Gabriel
Rodríguez

Txaber
Treviño



Resumen

El proyecto presentado en este dossier se centra en la toma de decisiones basada en las estadísticas y los intervalos de confianza. Ya se trate de comparar un grupo de individuos con una población o de estimar la frecuencia de una característica en una población utilizando una muestra, los intervalos de confianza son una notable ayuda para la toma de decisiones. También hemos querido dar a este trabajo un aspecto práctico, por lo que, además de las ilustraciones y ejemplos, hemos construido nuestras propias simulaciones y hemos trabajado para disponer de nuestros propios datos empíricos a partir de un cuidadoso protocolo científico en el campo de la biología celular.

Abstract

The project presented in this dossier focuses on statistical decision-making and confidence intervals. Whether you're comparing a group of individuals with a population or estimating the frequency of a characteristic in a population using a sample, confidence intervals are a remarkable tool to decision-making. We have also wanted to give this work a practical aspect, so in addition to illustrations and examples, we have built our own simulations and worked to provide our own empirical data from a careful scientific protocol in the field of cell biology.

Índice:

Introducción

I. Estadística descriptiva

- 1. Fluctuación de muestras**
- 2. Propiedades**
 - a. Intervalo de predicción**
 - b. Caso de la distribución binomial**
 - c. Caso de la distribución normal**
 - d. Aspecto cualitativo de la muestra**
- 3. Ejemplo de toma de decisión : el control de calidad**

II. Estadística inferencial

- 1. Propiedades**
- 2. Ejemplo : el sondeo electoral**

III. Aplicación experimental en la biología celular: estimación del número de estomas en la hoja de hiedra

- 1. Elección de la muestra**
- 2. Protocolo**
- 3. Análisis y resultados**

Conclusión

Anexos

Introducción

En nuestra vida cotidiana nos encontramos a menudo frente a la toma de decisiones: ¿cómo pronunciarse sobre la extinción de una especie de aves?, ¿cómo se decide si un medicamento es eficaz o no?, ¿cómo saben los médicos si el crecimiento de un bebé es correcto o no? (*figura 1*). Para contestar a estas preguntas la estadística es una herramienta fundamental. Los intervalos de fluctuación y de confianza son esenciales para decidir o para asumir una postura.

En una primera parte de este trabajo, exploramos la estadística descriptiva, donde se conocen los parámetros de posición (como la media) y de dispersión (como la desviación media) de un carácter estudiado en una población. Evaluamos mediante una simulación programada con Python la fluctuación de dichos parámetros en muestras aleatorias extraídas de la población. Además, ilustramos cómo se puede validar una hipótesis o confirmar una decisión.

En una segunda parte del trabajo, exponemos las herramientas de la estadística inferencial y las ilustramos con una simulación en Python. Aplicamos después este método a la evaluación de la proporción de estomas abiertos en la epidermis de la hoja de hiedra, basándonos en una muestra aleatoria de 50 hojas analizadas en el microscopio.

Se trate de intervalos de confianza o de fluctuación, abordamos también la falta de certeza y la medición del error a la hora de tomar decisión u optar por una postura.

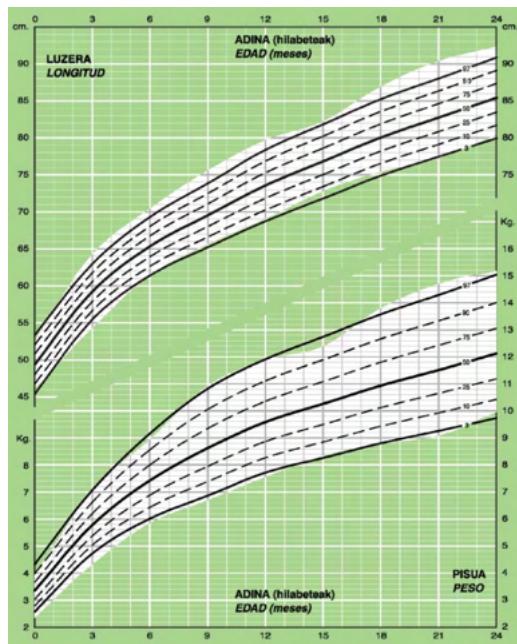


Figura 1: gráfico de la distribución en altura y peso de los niños entre 0 y 2 años.

Fuente: Cartilla de salud infantil de Osakidetza.

I. Estadística descriptiva

1. Fluctuación de las muestras

Cuando lanzamos un dado n veces, la frecuencia de aparición del 6 no es de $\frac{1}{6}$ como nos lo indica la probabilidad. Si lanzamos 10 veces, es bastante probable que la frecuencia de aparición del 6 esté más alejada del $\frac{1}{6}$ que si lo lanzamos 100 veces, y aún más que si lo lanzamos 1000 veces.

Para ilustrar este fenómeno, llamado la *fluctuación de las muestras*, hemos simulado con Python (anexo A) el lanzamiento de un dado de 6 caras (*figura 2*).

Desde muestras pequeñas de 10 lanzamientos hasta más grandes como de 5000, observamos la fluctuación de las frecuencias empíricas.

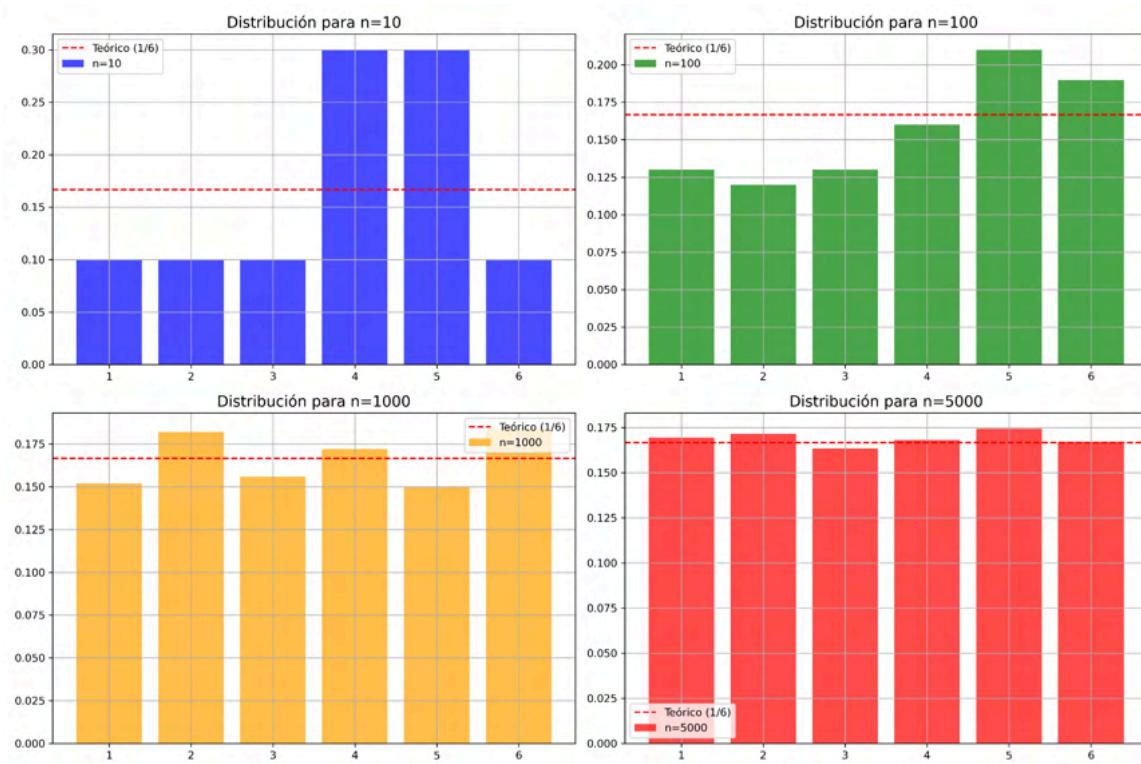


Figura 2

En la tabla de la *figura 3* hemos recogido las frecuencias de aparición de cada cara para poder calcular la desviación entre las frecuencias empíricas y la probabilidad teórica de

$\frac{1}{6}$ que aparece en la última columna. Vemos cómo ésta desviación se reduce al aumentar el tamaño de la muestra.

Tamaño	Frec. 1	Frec. 2	Frec. 3	Frec. 4	Frec. 5	Frec. 6	Desv.
10	0.1	0.1	0.1	0.3	0.3	0.1	0.05333
50	0.22	0.14	0.2	0.14	0.2	0.16	0.00653
100	0.13	0.12	0.13	0.16	0.21	0.19	0.00733
500	0.178	0.172	0.162	0.17	0.184	0.16	0.00053
1000	0.152	0.182	0.156	0.172	0.15	0.184	0.00117
5000	0.1694	0.1714	0.1634	0.168	0.1744	0.167	0.0001

Figura 3

Pero ¿cómo está formalizado matemáticamente este fenómeno y cómo se evalúa el hecho de que cuanto mayor sea el número de repeticiones de un experimento aleatorio, más se acercan las frecuencias a los valores teóricos dados por la probabilidad?

2. Propiedades

Cuando el número de repetición de un experimento aleatorio es alto, la ley de los grandes números permite interpretar la frecuencia de aparición de un carácter como su propia probabilidad de aparición. Precisamente, muestra que la media empírica, por ejemplo, converge hacia la esperanza de la variable aleatoria cuando el tamaño de la muestra tiende hacia el infinito.

a. Intervalo de predicción

La teoría asegura que, si un carácter tiene una proporción p incluida entre 0,2 y 0,8 en una población y que si tomamos una muestra de esta población de tamaño $n \geq 25$, entonces la frecuencia observada del carácter en la muestra estará en el intervalo $I_n = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$ con un nivel de confianza de 95%, es decir: $P(f \in I_n) \geq 0,95$.

b. Caso de la distribución binomial

Una experiencia de Bernoulli es una experiencia aleatoria con dos resultados: éxito y fracaso, siendo p y $(1 - p)$ las probabilidades respectivas de cada resultado. Tras la realización de n experiencias aleatorias idénticas e independientes de Bernoulli, una

variable aleatoria que cuenta el número de éxitos entre las n pruebas sigue la ley binomial de parámetros (n, p) y para todo k entre 0 y n tenemos:
 $P(X = k) = C(n, k) \cdot p^k \cdot (1 - p)^{n-k}$ donde $C(n, k)$ es el coeficiente binomial.

Si la probabilidad de encontrar un criterio en una población es de p , evaluamos la frecuencia de dicho criterio en una muestra de talla n , y si se cumplen que: $n \geq 30$, $np \geq 5$ y $n(1 - p) \geq 5$, entonces la frecuencia observada f del criterio en la muestra estará en el intervalo siguiente con un nivel de confianza de $(1 - \alpha)$, siendo α un número de $]0; 1[$:

$$I_n = \left[p - z \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + z \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

de forma que $P(f \in I_n) \geq 1 - \alpha$, siendo z un coeficiente que varía según el grado de certeza, tal y como lo muestra la siguiente tabla:

Nivel de confianza $(1 - \alpha)$	75%	90%	95%	98%	99%	99,5%	99,9%
Coeficiente z	1,15	1,64	1,96	2,33	2,58	2,81	3,29

Observación: El intervalo de predicción al 95% visto en I.a. $I_n = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$ es un intervalo más grande pero más práctico según el nivel de matemáticas requerido.

c. Caso de la distribución normal

El teorema del límite central establece que la suma de una serie de variables aleatorias independientes e idénticas con media y varianza finitas converge en distribución hacia una distribución normal. Por lo tanto, según este teorema, la ley binomial converge en distribución hacia la ley normal cuando el número n crece al infinito (figura 4).

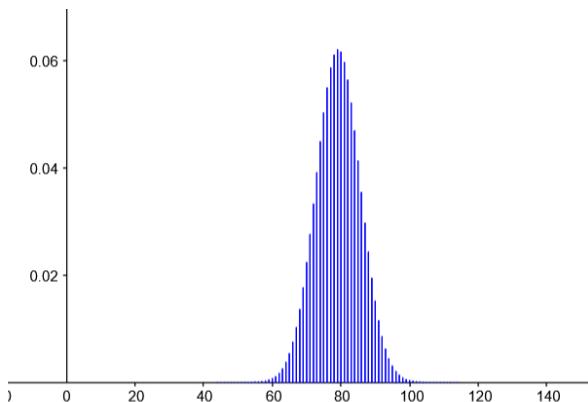


Figura 4: representación de la ley binomial ($n=165, p=0,48$)

Fuente: Geogebra.org

La distribución normal es una ley de probabilidad continua que depende de dos parámetros: su valor esperado μ y su desviación típica σ , siendo los dos números reales. Una variable aleatoria X que sigue una ley normal tiene la función de densidad siguiente:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

La curva de f se conoce como curva de Gauss o campana gaussiana (figura 5).

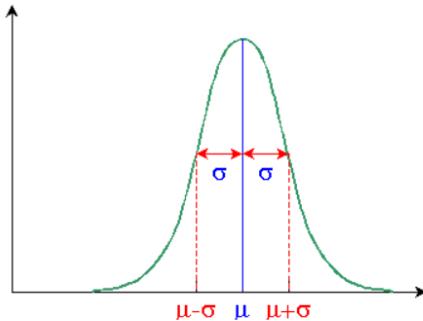


Figura 5: Curva de Gauss
Fuente: <https://www.astro.uliege.be>

Los tres intervalos de confianza más usados en la distribución normal son:

- $[\mu - \sigma, \mu + \sigma]$ que contiene aproximadamente el 68,26% de la distribución;
- $[\mu - 2\sigma, \mu + 2\sigma]$ que contiene aproximadamente el 95,44%;
- $[\mu - 3\sigma, \mu + 3\sigma]$ que contiene aproximadamente el 99,74%.

d. Aspecto cualitativo de la muestra

Sin embargo, la elección de la muestra también tiene un aspecto cualitativo que influye en la teoría. Si bien podemos basarnos en una muestra aleatoria para evaluar la calidad de piezas elaboradas por una máquina industrial en el sector automóvil, no siempre es el caso. Por ejemplo, en un sondeo de intención de voto existen numerosos factores que entran en juego para que la muestra sea representativa de la población: el barrio, la clase social, la edad o la profesión, que hacen que la muestra no pueda ser elegida de forma aleatoria.

Teniendo más claro qué son los intervalos relativos a la fluctuación de las muestras, veremos cómo intervienen en la toma de decisiones y cómo influyen los riesgos de error.

3. Ejemplo de toma de decisión: el control de calidad

Una empresa decide producir en Asia dos millones de camisetas y pide a su proveedor, tras un estudio de mercado, que el 30% de las camisetas sean azules y el 70% rojas. Cuando el pedido llega y queremos comprobar si el proveedor ha respetado el

contrato, tendríamos que contar todas las camisetas, lo cual es imposible. Para ello, se extrae una muestra de tamaño $n \geq 25$ donde se podrá calcular la proporción de camisetas rojas y azules. Supongamos que $n = 100$.

Si la frecuencia de camisetas rojas no está en el intervalo $[0,6; 0,8]$ podemos pensar que el proveedor no respetó el contrato. Pero como hay un margen del 5% de error posible, lo más prudente antes de desconfiar, sería extraer otra muestra de 100 camisetas o bien más grande aún y contar de nuevo. En caso de que el mal resultado se repita, comenzaremos a dudar de nuestro socio. Por el contrario, si los resultados son de nuestro agrado y parecen coincidir, confiaremos en el contrato establecido.

Para ilustrar que existe, no obstante, un 5% de probabilidad de cometer un error, hemos desarrollado una simulación en Python (anexo B) que sigue los siguientes pasos:

1. Generación de la población base:

Se crea una población de 2 millones de camisetas utilizando el código:

```
np.array([1]*int(2000000*0.7) + [0]*int(2000000*0.3)),
```

donde el 30% son camisetas azules (representadas por 0) y el 70% rojas (representadas por 1).

2. Definición del intervalo de predicción teórico al 95% de confianza:

Se calcula usando la fórmula simplificada del intervalo de fluctuación $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$

3. Muestreos aleatorios:

Se realizan 1000 muestreos aleatorios de 100 camisetas cada uno, con reemplazo (`replace=True`). En cada muestra, se calcula la proporción de camisetas rojas y se comprueba si dicha proporción se encuentra dentro del intervalo de fluctuación teórico. En caso afirmativo, se incrementa una variable contadora que registra cuántas muestras cumplen esta condición.

4. Visualización de los resultados:

Finalmente, podemos visualizar los resultados mediante un gráfico de barras (*figura 6*). Para las 1000 muestras simuladas, el 96,40% se encuentran dentro del intervalo teórico, lo que concuerda con el nivel de confianza del 95%.

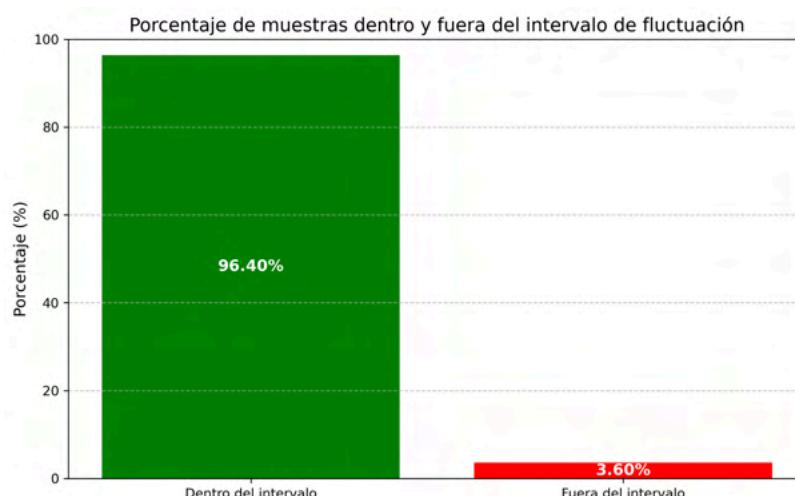


Figura 6: Porcentajes de muestras cuya frecuencia está o no en el intervalo a 95%

Esta simulación nos permite entender que, si sacamos 1000 muestras aleatorias e independientes de 100 camisetas dentro de nuestra población de 2 millones de camisetas, podemos equivocarnos en 5% de los casos a la hora de confiar o desconfiar del proveedor respecto a la frecuencia de productos acordada.

La estadística descriptiva permite evaluar la frecuencia de un carácter en una población donde se conoce la proporción de dicho carácter. Pero, ¿qué pasa cuando no conocemos esta proporción? ¿Cómo se puede evaluar? La estadística inferencial contesta a estas preguntas.

II. Estadística inferencial

1. Propiedades

La estadística inferencial permite obtener la estimación de un parámetro estadístico como la media o la frecuencia en una población entera a partir de una muestra, ya que estudiar la población entera puede resultar muy caro o imposible.

Si en una población sacamos una muestra de n individuos, y la frecuencia de un carácter en ésta muestra es de f , y si además se cumplen las condiciones: $n \geq 30$, $nf \geq 5$, $n(1 - f) \geq 5$ y la muestra es representativa (suficientemente grande y con independencia entre los individuos), entonces sabemos que la proporción de este carácter en la población total estará en el intervalo de confianza $[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}]$ con un grado de confianza del 95%.

De la misma manera que con los intervalos de predicción, este intervalo es una simplificación de la fórmula general:

$$I_n = \left[f - z \frac{\sqrt{f(1-f)}}{\sqrt{n}} ; f + z \frac{\sqrt{f(1-f)}}{\sqrt{n}} \right]$$

de forma que $P(p \in I_n) \geq 1 - \alpha$ siendo z el coeficiente que varía según el nivel de confianza indicado en la tabla del **I.1.b.**

Por otro lado, también se evalúa el tamaño n de la muestra cuando el tamaño N de la población es muy grande, es decir $N > 100\,000$: $n = \frac{z^2 \cdot p(1-p)}{e^2}$ siendo e el margen de error aceptado.

Cuando se conoce el tamaño de la población N y cuando $N < 100\,000$, se puede ajustar gracias a la fórmula: $n_{ajustada} = \frac{n}{1 + \frac{n-1}{N}}$.

Por ejemplo, para una población de 2000 personas, un nivel de confianza de 95% ($z = 1,96$), un margen de error de 5%, y una estimación de $p = 0,5$, obtenemos:

$$n \approx 384,16 \quad \text{y} \quad n_{ajustada} \approx 322$$

Si bien la diferencia entre n y $n_{ajustada}$ parece leve, esto es cada vez más notable a medida que el margen de error deseado disminuye, como se percibe en la tabla siguiente:

e (%)	1	2	3	4	5	6
n	9604	2401	1067	601	385	267
$n_{ajustada}$ (con $N = 2000$)	1322	741	614	442	322	228

2. Ejemplo : el sondeo electoral

Una aplicación corriente de la estadística inferencial es la previsión de resultados de elecciones, ya que para estimar la proporción de votantes de un partido, se hace una encuesta a una muestra de la población rigurosamente establecida.

Para las elecciones del Parlamento Europeo de junio de 2024, se hizo una encuesta en España por la empresa Sigma Dos para la RTVE. Si aplicamos la fórmula para conocer la talla mínima de una muestra representativa, con un nivel de confianza del 95% y un margen de error de $\pm 1\%$, obtenemos: $n = \frac{z^2 \cdot p(1-p)}{e^2} = \frac{1,96^2 \cdot 0,5(1-0,5)}{0,01^2} \approx 9\,604$

Sin embargo, Sigma Dos realizó la encuesta en una muestra de 12.000 personas, lo que parece satisfactorio.

Después de haber entendido la teoría detrás de los intervalos de confianza y de la toma de decisión, decidimos aplicarla. A pesar de la gran cantidad de información disponible en línea, preferimos trabajar en el laboratorio del centro y reproducir una serie de experimentos para juntar datos propios y elaborar una muestra propia representativa.

III. Aplicación experimental en la biología celular: estimación del número de estomas en la hoja de hiedra

Un estoma es una estructura microscópica que se encuentra en la epidermis de las hojas (especialmente en la cara inferior). Su función es regular los intercambios gaseosos entre la planta y el ambiente.

Un estoma (*figura 7*) está formado por dos células oclusivas con forma de alubia. Cuando está abierto, se ve en su interior un poro donde pasa el oxígeno (O_2), el dióxido de carbono (CO_2) y el vapor de agua (H_2O).

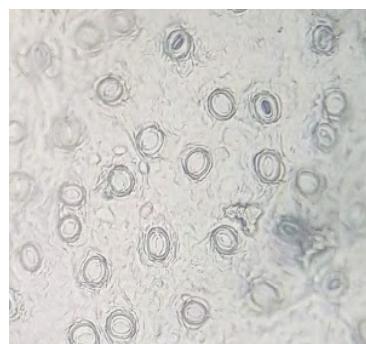


Figura 7: vista microscópica de estomas en una hoja de hiedra

1. Elección de la muestra

La hiedra de la que hemos extraído nuestra muestra es densa y trepadora. Se extiende en un rectángulo de 10 metros de largo y 0,8 metros de ancho, por lo que aproximadamente cubre $8\ m^2$. Con la ayuda de la inteligencia artificial, hemos podido estimar que el número de hojas de hiedra por m^2 es de 1 000 hojas. Por lo tanto, en nuestra superficie encontraríamos unas 8 000 hojas.

El cálculo de la talla de la muestra con un error de 5% nos da:

$$n = \frac{1,96^2 \cdot 0,5 \cdot (1-0,5)}{(0,05)^2} \simeq 384.$$

Sin embargo, como era complicado trabajar con un número tan alto de hojas, hemos decidido subir el margen de error al 14% y obtener una muestra de tamaño

$$n = \frac{1,96^2 \cdot 0,5 \cdot (1-0,5)}{(0,14)^2} \simeq 49$$

Con esto acordamos trabajar con una muestra de 50 hojas que nos pareció razonable.

2. Protocolo

→ Nuestras 50 hojas de hiedra (*figura 8*) se seleccionaron para ser cualitativamente representativas (algunas arriba de la planta, en el medio y más abajo, cerca de la periferia o más en el interior, a menudo al sol o más bien en sombra).



Figura 8: un individuo de la muestra de hojas de hiedra

→ Pasamos barniz de uñas transparente en la parte inferior de la hoja (donde se encuentran los estomas) para obtener una impresión de la epidermis foliar (*figura 9*).



Figura 9: fotografía de la impresión de la epidermis

→ Miramos las impresiones en el microscopio (*figura 10*), y fotografiamos las vistas para poder estimar el número de estomas abiertos por centímetros cuadrados.



Figura 10: fotografía de la visión microscópica

3. Análisis y resultados

Para empezar a contar, tuvimos que encontrar cómo establecer una escala en la visión ocular del microscopio. Entonces usamos una lámina de microscopio micrométrica de la marca Malassez (*figura 11*) donde en una visión x600 se ve una cuadrícula de dimensiones conocidas.

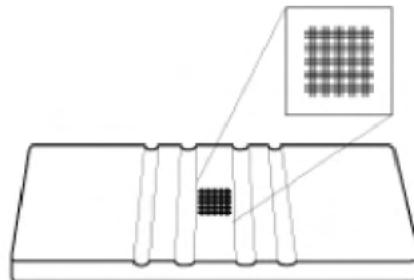


Figura 11: imagen de un portaobjetos de Malassez
(fuente: pedagogie.ac-toulouse.fr)

Gracias a la cuadrícula, pudimos estimar el ancho del retículo del microscopio, apoyándonos en el programa de medición gratuito y online Mesurim2 (*fuente: <https://www.pedagogie.ac-nice.fr/svt/productions/mesurim2/>*).

En la fotografía de la figura 10, siendo el aumento de x600 se mide el ancho del retículo de **27,53 micrómetros**.

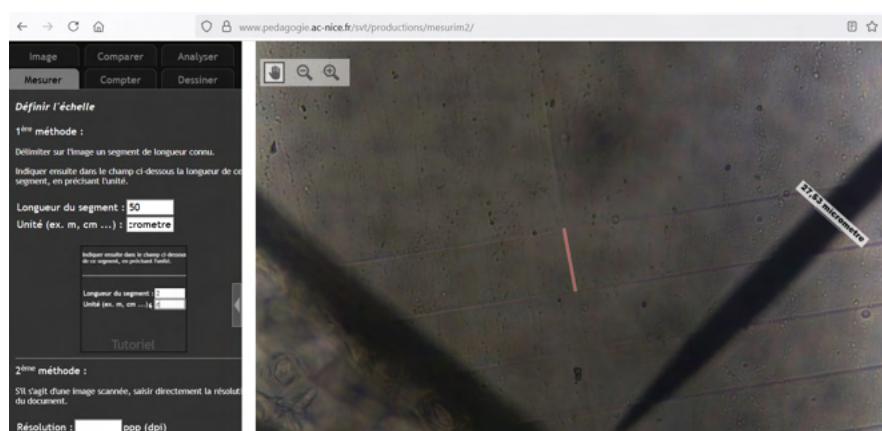


Figura 12: captura del programa Mesurim

Para cada fotografía pasamos a contar los estomas abiertos (*figura 13*).

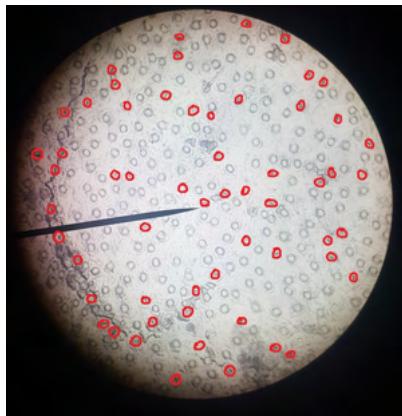


Figura 13: cálculo del número de estomas abiertos por fotografía.

Finalmente, con Mesurim2 y la escala dada por el retículo podemos obtener el área de nuestra visión microscópica (*figura 14*). En las fotografías de epidermis, el aumento es de x100; entonces, ponemos el ancho del retículo a 4,58 micrómetros.

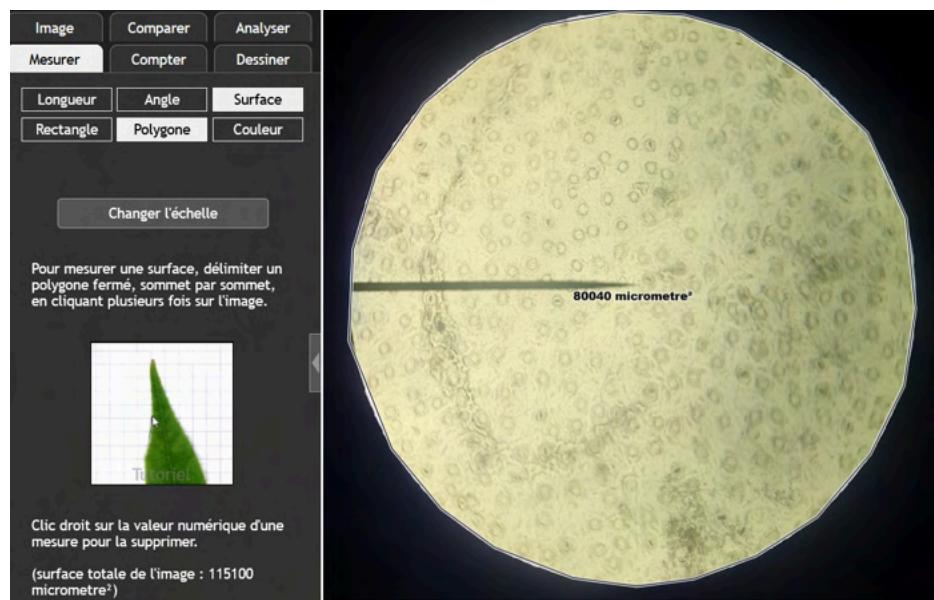


Figura 14: medición con Mesurim2 del área de la vista fotografiada

Calculamos el número de estomas abiertos por centímetros cuadrados y recolectamos los resultados en la tabla siguiente (*figura 15*):

Muestra Gabriela	65 147	121840	37 878	43 201	106 994	71 293	146 552	54 673	35 449	70 392
Muestra Gabriel	85 062	76 290	111 431	121 302	115 483	160 752	88 944	86 302	160 642	108 060
Muestra Txaber	112398	100611	84348	103376	122877	104592	97427	154896	122844	167555
Muestra Dasha	137 127	112 551	89 622	89 539	124 959	114 959	121 739	143 298	169 257	179 110
Muestra Sandrine	78 125	97 302	107 569	120 215	114 942	128 265	97 838	95 952	94 815	147 299

Figura 15: Número de estomas por centímetro cuadrado en la muestra de 50 hojas.

Modelizamos los resultados de este trabajo estadístico con una curva de Gauss, con parámetros la media de los datos $\mu = 101\,007$ y la desviación típica $\sigma = 33\,840$.

El intervalo de confianza al 95% es de [33 327, 168 687]

Una toma de decisión o de postura a raíz de este resultado sería lo siguiente: si se encontrara en una zona geográfica cercana a la de nuestra muestra unas hojas de hiedra cuyo número de estomas medio por centímetros cuadrados está fuera de este intervalo, convendría empezar a sospechar que influyen otros factores ambientales como presencia de pesticidas, mala calidad del agua o del aire, presencia de un insecto invasor, etc.

Observación:

Los científicos han utilizado cálculos de estomas por cm^2 para estudiar las variaciones de la concentración atmosférica en CO_2 y así comprender mejor los climas antiguos. Para ello utilizan experimentalmente el Ginkgo (un grupo muy antiguo que ha cambiado poco en los últimos 200 millones de años). El número de estomas es inversamente proporcional a la concentración de CO_2 atmosférico: cuanto más CO_2 menos estomas abiertos necesita la planta para la fotosíntesis (anexo C).

Conclusion

A través de este proyecto hemos estudiado la noción de fluctuación de las muestras y los intervalos de confianza, lo que nos ha permitido comprender cómo se usa la estadística en la vida cotidiana para tomar posiciones o tomar decisiones: ya sea en el ámbito de la salud, la seguridad o la política, los campos de aplicación son vastos y numerosos.

Además de los conceptos teóricos, aprovechamos los conocimientos informáticos de un miembro del grupo para ilustrar las nociones mediante simulaciones. También queríamos hacer un trabajo de laboratorio y utilizar nuestros propios datos empíricos en lugar de obtener datos ya preparados. Así pues, llevamos a cabo la tediosa pero necesaria tarea de repetir un protocolo científico para obtener una muestra representativa que nos permitiera, por inferencia, estimar la proporción de estomas por centímetro cuadrado de hojas de hiedra. Este trabajo de laboratorio y el análisis de los datos nos obligaron a movilizar nuestra capacidad de pensamiento crítico y a buscar soluciones a los problemas que se nos planteaban, lo que resultó especialmente estimulante, ya que cada uno de nosotros se embarcará en estudios científicos en septiembre.

Bibliografía y sitografía

- <https://www.meilleurenmaths.com/images/misyl/terminaleS/ProgrammeOfficiel/probabilites/loi-normale-cours.pdf>
- <https://www.omnicalculator.com/fr/statistiques/calculateur-intervalle-confiance>
- https://es.wikipedia.org/wiki/M%C3%ADnimos_cuadrados
- <https://riunet.upv.es/bitstream/handle/10251/7940/El%20teorema%20central%20del%20límite.pdf?sequence=3>
- <https://planet-terre.ens-lyon.fr/ressource/co2-depuis-4ga.xml>
- <https://www.rtve.es/noticias/20240609/sondeo-elecciones-europeas-espana-2024/16132491.shtml>
- <https://www.ehu.eus/eu/web/campusa/-/manual-para-saber-leer-las-encuestas-electorales>
- https://www.ine.es/prensa/elecgral_jul2023.pdf?utm_source=?

Anexos

Anexo A: simulación del lanzamiento de un dado

(o ver en el enlace [🔗 Simulación Dados.ipynb](#))

```
❶ import numpy as np, matplotlib.pyplot as plt

def simular_dados(n, num_sim=1):
    """Simula lanzamientos de dados y calcula las frecuencias de cada valor."""
    return np.array([np.mean(np.random.randint(1, 7, size=(num_sim, n)) == i, axis=1) for i in range(1, 7)]).T

def crear_tabla(datos, cols, filename):
    """Crea y guarda una tabla con matplotlib."""
    fig, ax = plt.subplots(figsize=(12, 6))
    ax.axis('off')
    tabla = ax.table(cellText=datos, colLabels=cols, loc='center', cellLoc='center')
    tabla.auto_set_font_size(False)
    tabla.set_fontsize(10)
    tabla.scale(1.2, 1.5)
    # Ajustar ancho y colorear encabezados
    for j in range(len(cols)):
        tabla.auto_set_column_width(j)
        tabla._cells[(0, j)].set_facecolor('yellow')
        tabla._cells[(0, j)].set_text_props(weight='bold')
    plt.savefig(filename, dpi=300, bbox_inches='tight')
    plt.close()

# Parámetros y simulaciones
tamanos, teorico = [10, 50, 100, 500, 1000, 5000], np.array([1/6] * 6)
muestras_viz, colores = [10, 100, 1000, 5000], ['blue', 'green', 'orange', 'red']

# Generar datos
sim_resultados = [simular_dados(n, 1000) for n in tamanos]
desviaciones = [np.std(sim) for sim in sim_resultados]
frecuencias_viz = {n: simular_dados(n)[0] for n in tamanos}

# Gráfico: Distribución de frecuencias
fig, axs = plt.subplots(2, 2, figsize=(15, 10))
for i, n in enumerate(muestras_viz):
    row, col = i // 2, i % 2
    freq = frecuencias_viz[n]
    axs[row, col].bar(range(1, 7), freq, color=colores[i], alpha=0.7, label=f'n={n}')
    axs[row, col].axhline(y=1/6, color='red', linestyle='--', linewidth=1.5, label='Teórico (1/6)')
    axs[row, col].set_title(f'Distribución para n={n}', fontsize=14)
    axs[row, col].set_xticks(range(1, 7))
    axs[row, col].legend()
    axs[row, col].grid(True)
plt.tight_layout()
plt.savefig('distribucion_todas_caras.png', dpi=300)
plt.close()

# Preparar datos para la tabla errores
tabla_errores = []
for i, n in enumerate(tamanos):
    freq = frecuencias_viz[n]
    error = round(np.sum((freq - teorico) ** 2), 5)
    tabla_errores.append([n] + [str(f) for f in freq] + [str(error)])

# Crear y guardar tablas
crear_tabla(tabla_errores,
            ['Tamaño'] + [f'Frec. {i}' for i in range(1, 7)] + ['Desv.'],
            'tabla_errores.png')

print("\nResultados guardados en: 'distribucion_todas_caras.png' y 'tabla_errores.png'")
```

Anexo B: simulación camisetas (o ver en el enlace [Simulación Camisetas.ipynb](#))

```
❶ import numpy as np
import matplotlib.pyplot as plt

# Configuración inicial
np.random.seed(41) # Para poder obtener la misma figura que en el ejemplo
p_teorica = 0.7 # Proporción teórica (70% rojas)
n = 100 # Tamaño de muestra
num_muestreos = 1000

# Crear población (70% camisetas rojas=1, 30% azules=0) usando numpy directamente
poblacion = np.array([1] * int(2000000 * 0.7) + [0] * int(2000000 * 0.3))

# Calcular intervalo de fluctuación al 95%
margen_error = 1/np.sqrt(n) # Uso de la fórmula simplificada
limites = [p_teorica - margen_error, p_teorica + margen_error] # Intervalo de fluctuación

# Realizar muestreos
corresponde = 0
for muestra in range(num_muestreos): # Bucle para cada muestra
    p_muestra = np.mean(np.random.choice(poblacion, size=n, replace=True)) # Tomar muestra y calcular proporción
    if limites[0] <= p_muestra <= limites[1]: # Comprobar si la muestra está dentro del intervalo
        corresponde += 1

# Calcular resultados
porcentajes = [corresponde * 100 / num_muestreos, (num_muestreos - corresponde) * 100 / num_muestreos] # Porcentaje empírico

# Crear gráfico de barras
categorias = ['Dentro del intervalo', 'Fuera del intervalo']
plt.figure(figsize=(10, 6))
bars = plt.bar(categorias, porcentajes, color=['green', 'red'])

# Añadir etiquetas con porcentajes
for bar in bars:
    plt.text(bar.get_x() + bar.get_width()/2., bar.get_height()/2,
             f'{bar.get_height():.2f}%', ha='center', va='center',
             color='white', fontweight='bold', fontsize=12)

# Configurar gráfico
plt.title('Porcentaje de muestras dentro y fuera del intervalo de fluctuación', fontsize=14)
plt.ylabel('Porcentaje (%)', fontsize=12)
plt.ylim(0, 100)
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Guardar gráfico en 'resultados_simulacion.png'
plt.savefig('resultados_simulación.png', dpi=300, bbox_inches='tight')
```

Anexo C: Estudios científicos basados en los estomas para estudiar el clima

Fuente : <https://cordis.europa.eu/article/id/84194-effect-of-elevated-co2-levels-on-poplar-stomata/es>

Las plantaciones de árboles tienen una gran capacidad para disminuir los niveles de CO₂ y pueden contribuir a la reducción de las emisiones de dióxido de carbono en Europa. Se llevaron a cabo estudios multidisciplinarios para analizar el papel de las plantaciones de chopos en la captura de CO₂ y apoyar el compromiso de Europa con el Protocolo de Kyoto. El trabajo llevado a cabo en el marco del proyecto EUROFACE consistió en cuantificar el factor de absorción de altos niveles de dióxido de carbono atmosférico en las condiciones climáticas actuales y futuras. Los resultados proporcionaron datos a los responsables de las políticas medioambientales referentes a la exposición a largo plazo de los árboles a niveles elevados de CO₂ y sus efectos sobre la fijación del carbono por parte de la vegetación leñosa.

El estudio de los estomas fue decisivo para el proyecto. Su importancia radica en que estos poros diminutos del envés de las hojas son los encargados de regular el intercambio de gases y vapor de agua con la atmósfera. Se han llevado a cabo estudios en el pasado para medir la respuesta de los estomas ante incrementos en los niveles de dióxido de carbono. Sin embargo, el proyecto EUROFACE fue el primero en realizar estudios a largo plazo, de varios años, acerca de los efectos de los niveles elevados de CO₂ en los estomas. La importancia de averiguar la respuesta de los estomas residía en que ésta puede afectar a los niveles de fotosíntesis, transpiración y uso eficiente del agua de las hojas.

Los resultados indicaron que los niveles elevados de CO₂ causaban una reducción de los estomas de los árboles jóvenes, mientras que en el caso de los árboles viejos, tras una exposición prolongada, no se observó una reducción significativa. A pesar de que no se conocen en profundidad los mecanismos que controlan la respuesta de los estomas al CO₂, está claro que tal sensibilidad no está provocada por los tejidos internos de las hojas. Parece que estos mecanismos están controlados por las células oclusivas que rodean los estomas y regulan su apertura, o por la cutícula, es decir, la capa cerosa que recubre las hojas.

Fuente : https://acces.ens-lyon.fr/acces/thematiques/CCIC/ressources/bio_point2

Principio: Las hojas de las plantas vasculares modernas muestran una correlación inversa entre el porcentaje de células epidérmicas que son estomatas, llamadas índice estomacal, y la presión parcial atmosférica de CO₂.

- Ejemplo de dos especies conservadas: Ginkgo y Metasequoia

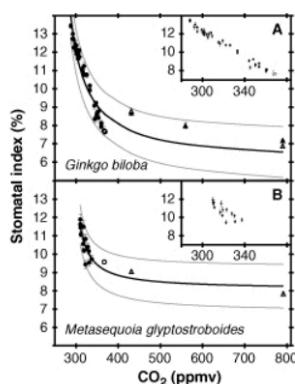


Figura 2 : Vinculación entre la presión parcial de CO₂ y el índice estomacal para dos especies actuales.

Según Royer et al (2001, Ciencia, 292: 2310-2313)